JOHANNES KEPLER
UNIVERSITÄT LINZ

Eingereicht von
DI Vaishali Dhanoa

Angefertigt am
Institut für Computergrafik

Erstbeurteiler
Univ.-Prof. DI Dr. Marc Streit

Zweitbeurteiler
FH-Prof. Priv.-Doz. DI Dr. Wolfgang Aigner, MSc

Mitbetreuer
Univ.-Prof. DI Dr. Eduard Gröller

September 2024

From Analysis to Communication: Supporting Users in Understanding Complex Spreadsheets and Dashboards

Dissertation
zur Erlangung des akademischen Grads
Doktorin der Technischen Wissenschaften
im Doktoratsstudium der
Technischen Wissenschaften

JOHANNES KEPLER
UNIVERSITÄT LINZ
Altenberger Straße 69
4040 Linz, Österreich
www.jku.at

To my family and friends

Abstract

Advancements in big data processing and interactive visualization tools have led to significant changes in how users analyze and explore their data. This thesis aims to address the challenges resulting from these changes through a two-step approach to support users. We first address the issues at the spreadsheet level before moving on to more complex visual representations in a dashboard environment. We use the Fuzzy Spreadsheet approach at the spreadsheet level to include uncertain information in the decision-making process. Our approach augments traditional spreadsheets with uncertain information where a cell can hold and display a distribution of values, in addition to other contextually relevant information, such as impact and relationship between cells, to convey sensitivity and robustness information to the user. When users transition from spreadsheet representations to advanced visualization tools such as interactive dashboards, they often face challenges related to their use that can lead them to revert to their old, familiar static analysis tools. With the help of dashboard onboarding, authors can communicate the intended use and purpose of their dashboards, along with the workings of visualizations present on the dashboards, to fill the user's knowledge gap. We created a process model for dashboard onboarding that formalizes and unifies different onboarding strategies for dashboards and facilitates the design and implementation of new onboarding approaches. Using this process model as a base and drawing inspiration from the fields of data storytelling and open-world game design, we developed an approach for crafting semi-automated interactive dashboard tours (D-Tours) to produce an onboarding experience tailored to individual users while preserving their agency. We implemented this concept in a tool called D-Tour Prototype which allows authors to create D-Tours from scratch or using automatic templates. Finally, we provide future directions based on the insights from this thesis to explore the role of AI in the design and development of dashboard onboarding.

Kurzfassung

Fortschritte bei der Verarbeitung großer Datenmengen und interaktiven Visualisierungswerkzeugen haben dazu geführt, dass sich die Art und Weise, wie Nutzer ihre Daten analysieren und erforschen, erheblich verändert hat. Diese Arbeit zielt darauf ab, die Herausforderungen, die sich aus diesen Veränderungen ergeben, durch einen zweistufigen Ansatz zur Unterstützung der Nutzer anzugehen. Zunächst werden die Probleme auf der Ebene der Tabellenkalkulation behandelt, bevor wir uns komplexeren visuellen Darstellungen in einer Dashboard-Umgebung zuwenden. Wir verwenden den Fuzzy Spreadsheet-Ansatz auf Tabellenkalkulationsebene, um unsichere Informationen in den Entscheidungsprozess einzubeziehen. Unser Ansatz erweitert herkömmliche Tabellenkalkulationen mit unsicherer Informationen, wobei eine Zelle eine Verteilung von Werten enthalten und anzeigen kann. Weiters werden zusätzliche kontextrelevanten Informationen integriert, wie z.B. Auswirkungen und Beziehungen zwischen Zellen, um dem Benutzer Sensitivitäts- und Robustheitsinformationen zu vermitteln. Wenn Benutzer von Tabellenkalkulationen zu komplexeren Visualisierungen übergehen, wie z.B. interaktiven Dashboards, die Daten mit Hilfe von miteinander verknüpften Visualisierungen darstellen, ergeben sich neue Herausforderungen. Ohne effiziente Einschulung könnten sich die Benutzer veranlasst sehen, sich auf ihre alten, vertrauten statischen Analysewerkzeuge zurückzuziehen. Mit Hilfe von Dashboard-Onboarding können die Autoren ihren Nutzern den Verwendungszweck und die Funktionsweise der Dashboards sowie der darin enthaltenen Visualisierungen vermitteln und Wissenslücken schließen. Wir haben ein Prozessmodell für das Onboarding von Dashboards entwickelt, das die verschiedenen Onboarding-Strategien für Dashboards formalisiert und vereinheitlicht. Die Entwicklung und Implementierung neuer Onboarding-Ansätze wird dadurch erleichtert. Auf der Grundlage dieses Prozessmodells und inspiriert von den Bereichen Data-Storytelling und Open-World-Game-Design haben wir einen Ansatz zur Erstellung halbautomatischer interaktiver Dashboard-Touren (D-Tours) entwickelt, um ein auf den einzelnen Nutzer zugeschnittenes Onboarding-Erlebnis zu schaffen und gleichzeitig seine Handlungsfähigkeit zu erhalten. Wir haben dieses Konzept in einem Werkzeug namens D-Tour Prototype umgesetzt, mit dem Autoren D-Touren von Grund auf oder mit Hilfe automatischer Vorlagen erstellen können. Abschließend geben wir auf der Grundlage der Erkenntnisse aus dieser Arbeit Hinweise für zukünftige Arbeiten, um die Rolle der KI bei der Gestaltung und Entwicklung von Dashboard-Onboarding zu beleuchten.

Acknowledgements

To my supervisor and mentor, Marc Streit, thank you so much for giving me the opportunity to pursue my PhD. Marc made me feel welcome and helped me integrate into the Austrian culture. He offered me guidance beyond the academic world and carefully nurtured me in the industrial projects as well. My time in his lab was extremely productive and filled with valuable learning experiences.

I am equally grateful to my co-supervisor, Eduard Gröller. Despite his busy schedule, he found time to mentor me and provided extremely useful advice. His ability to balance seriousness and fun made the project meetings interesting and enjoyable.

I would also like to thank my manager, Belgin Mutlu, who supported me and helped me grow, especially in the industrial project setting. I would also like to thank the entire team at Pro2Future GmbH for their support. Special thanks to Gerd Hribernig, Alois Ferscha, Sandra Neuhold-Pauer, Christina Pichler, Thomas Martetschläger, Michael Haslgrübler, Ouijdane Guiza, Markus Brillinger, Michael Krisper, Josef Suschnigg, Matej Vukovic, Jakob Zethofer, Georgios Sopidis, Behrooz Azadi, and Muaaz Abdul Hadi.

Special thanks to Prof. Wolfgang Aigner for evaluating my thesis and thank you to all the anonymous reviewers who made my research better.

I am also thankful to "Dr." Kresimir Matković and Rainer Splechtna for being the best travel companions for the after-conference vacations. I am also grateful to Niklas Elmqvist for his mentorship and for hosting me during my three-month stay at Aarhus University, Denmark. It was one of the most intense and wonderful times during my PhD. Thank you to the Österreichische Forschungsgemeinschaft and the Auslandsbüro at Johannes Kepler University for supporting me financially during this stay.

A special thank you to my industrial collaborators for their intriguing use cases that piqued my curiosity. It was a pleasure working with you.

I would also like to thank my fellow PhD colleagues: Klaus Eckelt, Andreas Hinterreiter, Christina Humer, Christian Steinparz, Conny Walchshofer, and Patrick Adelberger. Our "short" lunch and coffee-breaks helped me unwind and provided me with the motivation to continue my work. Also, our brainstorming sessions and late night paper submission moments will stay with me forever.

I am extremely thankful to my greatest inspiration, Prof. Sanjay Goel, who supervised me during my Bachelor's in Technology and has remained a mentor to me over the years. His visionary approach, passion for teaching, and genuine kindness in mentoring and encouraging students, is truly admirable.

To all my physical, emotional, and mental coaches and gurus: a big thank you! Without your training, I would not have made it this far.

Next, I want to thank my sisters, Ashneet and Raymon, whose love and support kept me motivated throughout this journey. I am also thankful to all my friends who brought a sense of home to my life in a "foreign" country. Special thanks to Vera, Uma, Edit, Sukhmani, Jonny, Magdalena, Neelam, Shaista, and many others who were by my side through all phases of my life. A heartfelt thank you to Matthias Preuner for his constant

support and for always being the sunshine on my rainy days.

I am forever indebted to my parents, Raj and Ajit, for their love, support, and encouragement. Without their support, none of this would have been possible. Thank you so much!

This work was supported by the Austrian Science Fund (FWF DFH 23–N), Villum Fonden (Villum Investigator grant VL-54492, Denmark), and the Austrian Research Promotion Agency, FFG 881844: "Pro^{2}Future" and FFG 879730: "VRVis". Additional support was granted by the State of Upper Austria and the Austrian Federal Ministry of Education, Science and Research via the LIT – Linz Institute of Technology (LIT-2019-7-SEE-117), and by the State of Upper Austria (Human-Interpretable Machine Learning).

Contents

List of Figures

List of Tables

1 Introduction

"The purpose of visualization is insight, not pictures." Ben Shneiderman.

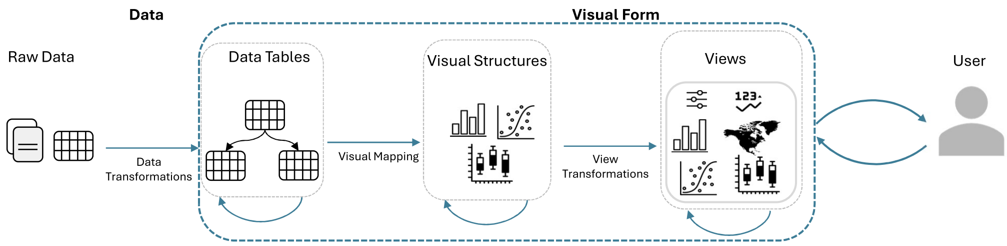


Fig. 1.1: Information visualization pipeline reproduced after Stuart et al. [1]

In 1999, Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman [1] proposed a reference model for visualization to show a conceptual overview of the visualization process. Figure 1.1 illustrates a repurposed model based on Stuart et al.'s [1] model of the visualization process, which shows how raw data can be transformed into interactive visualizations that enable users to perform tasks and extract insights from the visualizations.

The core of this thesis is to support users in analyzing their tabular data and assisting them as they move on to engaging with more complex analysis processes using interactive visualizations.

During my research, I collaborated with two industrial partners, gaining valuable insights into their analytical processes and the critical role visualization plays in them. My thesis is driven by the need to address real-world challenges, which gave me the opportunity to contribute and build upon existing research in the area of visual analytics. To explain the contribution of this thesis, it is important first to discuss how the visualisation process has evolved since 1999 and highlight the areas where my work contributes to this evolution (Section 1.1 and Section 1.1.1). The research contribution is then specified in Section 1.2.

1.1 Using Visualizations for Analysis and Communication

In 2005, Thomas and Cook [2] discussed the role of visual display for analytic reasoning, paving the path to the research and development agenda for the visual analytics field. They defined visual analytics as the science of analytical reasoning facilitated by interactive visual interfaces. In 2008, Keim et al. [3] enhanced this definition to specify that visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision-making on the basis of very large and complex data sets. They also proposed a visual analytics model, which integrated visual and automatic data analysis methods for scalable interactive decision support [3].

The rise in the field of artificial intelligence (AI) and machine learning (ML) further led to significant improvements in the data preparation and analysis steps in the visual analytics model. With large amounts of data, the need for scalable, interactive and real-time data analysis grew. To address this, several visual analytics tools and solutions that employed interactive visualizations became pervasive in the workflow of analysts from different domains. Additionally, the technical challenges of maintaining these tools and communicating the findings from the analysis process came along as new challenges for both the industry and research fields. We represent the current process of visual analysis using Figure 1.2 and explain it in detail in the next Section 1.1.1.

1.1.1 Extended visual analytics model

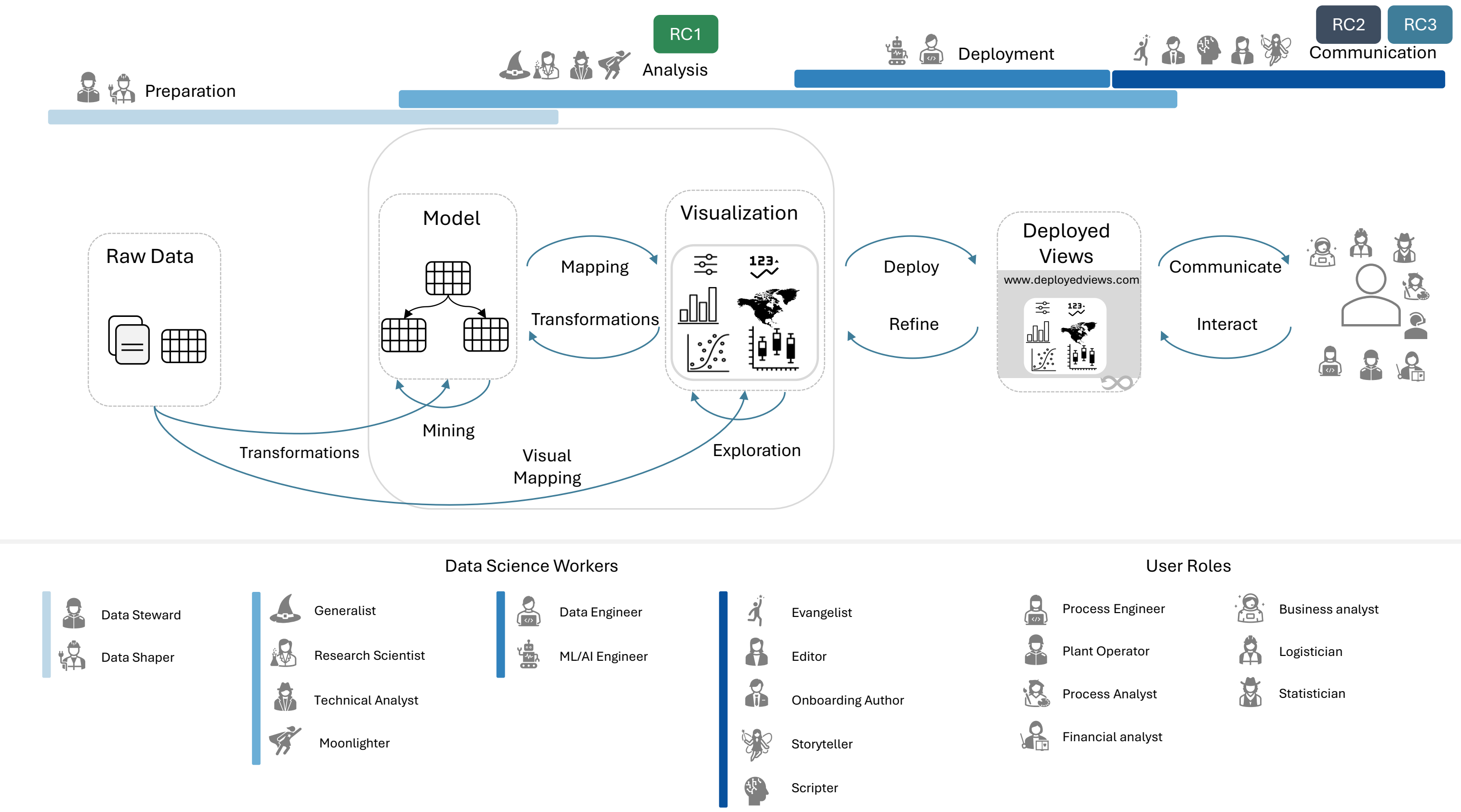


Fig. 1.2: Extended visual analytics pipeline from Keim et al. [3] comprising of additional steps such as deployment and communication [4]. The data science workers and user roles are taken from the literature and our experience with industrial collaborators. RC1 is the first paper contribution: Fuzzy Spreadsheet [5]. RC2 is the second paper contribution: a process model for dashboard onboarding [6]. RC3 is the third paper contribution: D-Tour: semi-automatic dashboard tours for dashboard onboarding [7].

Figure 1.2 shows an extended version of the visual analytics model based on Stuart K. Card et al.'s [1] information visualization pipeline and Daniel Keim et al's [3] visual analytics model. It is further enhanced based on the literature on data science [4, 8, 9], storytelling [10], and our experience with industrial collaborators.

The four higher-order processes in the extended model, namely, preparation, analysis, deployment, and communication, are based on the work of Crisan et al. [4]. We list them below with a focus on visual analytics aspects and outline additional user roles from our experience with the industrial partners, literature from storytelling [11] and onboarding [12]:

- **Preparation**: Data preparation is the prerequisite for doing analysis and requires data gathering, profiling, and wrangling [4]. Keim et al. [3] discuss the transformation of raw data into a data model, which is represented in Figure 1.2. Crisan et al. [4] describe two main user roles involved in this step: data stewards, who are domain experts responsible for accessing and using the data, and data shapers, who are responsible for curating and preparing the data for the analysis. The job roles of data stewards and data shapers can vary greatly across different sectors of the economy. Brehmer et al. [9] briefly discuss these roles based on interviews with participants from several industries, including software, manufacturing, education, private equity, retail consulting, law, and healthcare.

- **Analysis**: Depending on the volume and complexity of the data, the analysis process can range from simple to highly sophisticated. This can involve using basic statistical methods to advanced computational techniques to extract meaningful insights from the data. Visualizations can help in data exploration and extracting insights from them as they can be used for both simple and complex datastes [13]. The user roles involved in this step could be domain experts who understand the domain and the data, referred to as research scientists by Crisan et al. [4]. Additional roles include generalists who focus solely on data science, technical analysts who perform data science tasks occasionally, and moonlighters who are non-technical but are tasked with technical duties. In collaboration with our industry partners, we examine the challenges these roles encounter during the transition to commercial dashboarding systems [8].

- **Deployment**: Elmqvist [14] discusses how data is collected everywhere and can be accessed anywhere. With this growing need for anytime, anywhere access, the prepared data, as well as the visualizations, need to be deployed for continuous access. Typically the data engineers and ML/AI engineers who are proficient in developing and deploying ML/AI methods are assigned to these tasks. While commercial tools, such as Microsoft Power BI [15], offer support in deploying their dashboards to production, engineers are still needed to monitor and manage the process.

- **Communication**: The deployed visualizations need to be presented to the users [9]. This can include anything from static documentation that highlights the findings from the data to dashboards, which allow the users to extract the insights themselves. Brehmer et al. [9] discuss both static and interactive choices made by the users to present their findings within organizations. In addition to the role of evangelist described by Crisan et al. [4], we integrate the roles of scriptor, editor, storyteller, and onboarding author from data-driven storytelling [10] and onboarding literature [6]. We include these additional roles as they are mainly responsible for disseminating the findings from the visualizations to the users.

Based on our observations, the preparation and deployment phases require the most engineering effort, particularly in the domain of data analytics and software engineering. While the thesis focuses on the research related to the **analysis** and **communication** aspects of the extended visual analytics model, it also addresses other components of this model.

To address the analysis aspect, the first part of my research focuses on creating an interactive visualization that highlights the underlying uncertainty in the data for better decision-making. It allows for the direct manipulation of tabular data within spreadsheets to explore what-if scenarios and offers users a variety of modeling alternatives, such as best, average, and worst cases.

For a more in-depth investigation, analysts typically switch from spreadsheet representation to more advanced analytics tools. These tools range from low-code solutions such as Power BI, Tableau, and other commercial tools to medium—and high-code solutions such as D3, Vegalite, Jupyter lab notebooks, and Matlab.

As a secondary contribution, we study the transition to a low-code visual analytics system, i.e., Microsoft Power BI, in a large manufacturing company. We report on the socio-technical challenges experienced by workers in creating and using visualization dashboards [8]. Our findings highlight the importance of providing adequate training and support for employees–dashboard authors and users–during this transition period.

Although dashboard authors have access to numerous training programs and our research community has produced several works on dashboard creation [16] and design patterns [17], we observed a gap in assisting users in familiarizing themselves with new dashboards. New users may find the interactive nature of dashboards challenging, which could hinder their adoption of this visual analytics tool [18, 19, 20]. One approach to overcome this difficulty is to onboard new users to the dashboard's visualizations, data, and interactions. We call this process "dashboard onboarding", which is the focus of the rest of my research. For the sake of clarity, we would like to explain what we mean by "dashboard", as we use it frequently in this thesis.

1.1.2 Dashboards

Dashboards are one of the most prevalent applications for visualizing data to facilitate data-driven decision-making, communication, and learning [21]. Several attempts have been made to define a dashboard [28, 29, 30, 31, 21, 32, 33, 34]. Some of the frequently used definitions from dashboard-related literature include:

- Few (2006) [28, 29]: "*a visual display of the most important information needed to achieve one or more objectives; consolidated and arranged on a single screen so the information can be monitored at a glance*".

- Kitchin et al. (2015) [31]: "*a dashboard seeks to act as a translator, not simply a mirror, setting the forms and parameters for how data are communicated and thus what the user can see and engage with.*"

- Wexler et al. (2017) [30]: "*a visual display of data used to monitor conditions and/or facilitate understanding*," and can include infographic elements or textual descriptions for narrative visualizations.

- Sarikaya et al.(2018) [21]: They identify two major design perspectives: the visual genre of dashboards and the functional genre. The visual genre provides "*a visual data representation structured as a tiled layout of simple charts and/or large numbers*". Meanwhile, the functional genre displays "*an interactive display that*

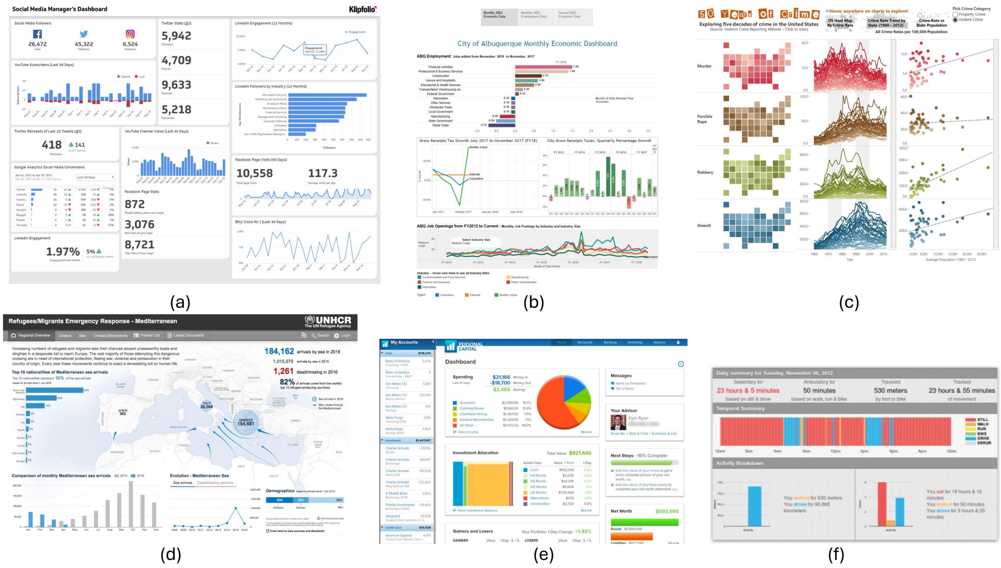


Fig. 1.3: Collection of dashboards surveyed by Sarikaya et al. [21] and extended by Bach et al. [17]. (a) shows Klipfolio's Social Media Manager Dashboard: a traditional dashboard with key performance indicators and visualizations with real-time data [22]. (b) represents a dashboard for communicating the economic situation for strategic decision-making [23]. (c) displays a dashboard depicting 50 years of crime for general-purpose communication [24]. (d) is a dashboard for raising awareness about migrants' emergency response (infographic-style) [25]. (e) is a dashboard for learning about finances [26]. (f) represents a quantified-self dashboard which shows daily activity logs using static visualizations [27].

> *enables real-time monitoring of dynamically updating data". They acknowledge the tension between these two genres and do not provide a "single authoritative definition of dashboards*".

We use the term dashboard to refer to a combination of multiple data visualizations and textual displays that are often interlinked/interactive and typically arranged in a single-page layout [6]. This is not an attempt to provide a clear factual description of a dashboard; rather, it is intended to serve as a working definition for this thesis.

Sarikaya et al. [21] grouped the dashboards into interactive dashboards for decision-making, as shown in Figure 1.3(a), (b), and (c) (further divided into operational, strategic, and tactical), static dashboards for raising awareness, Figure 1.3(d), dashboards for motivation and learning, Figure 1.3(e), and personal analytics, Figure 1.3(f). This grouping was based on their extensive survey of 83 dashboards which they analyzed from the perspectives of purpose, audience, visual & interactive features, and data semantics. Bach et al. [17] further extend this corpus and provide seven genres of dashboards, similar to the storytelling genres by Segel and Heer [35]. These genres are static dashboards, analytic dashboards, magazine dashboards, infographic dashboards, repository dashboards, and embedded mini dashboards. The dashboards used in our research, including exemplary and our collaborator's dashboards, encompass most of the above-mentioned groups except personal analytics. We did not categorize them specifically, as this was outside the scope of our work.

In this thesis, we first address the *analysis* phase at the spreadsheet level before moving on to *communicating* complex visual representations in a dashboard environment. We describe our research questions and contributions in the next Section 1.2.

1.2 Research Questions and Research Methodology

This thesis helps users understand complex spreadsheets and visualization dashboards. This leads to the following research questions:

- **RQ1**: How can we support users in comprehending and exploring uncertainty during the analysis process?
- **RQ2**: How can we support users in understanding complex, interlinked visualizations, such as in a dashboard?
- **RQ3**: How can we support dashboard authors in developing onboarding approaches tailored to their end users while maintaining the users' autonomy?

Figure 1.4 shows the three main contributions that answer the above-mentioned research questions. These research contributions use well-established qualitative and quantitative methods to evaluate the resulting approaches. The main contributions are described in Section 1.2.1, and the secondary contributions are listed in Section 1.2.2

1.2.1 Main Contributions

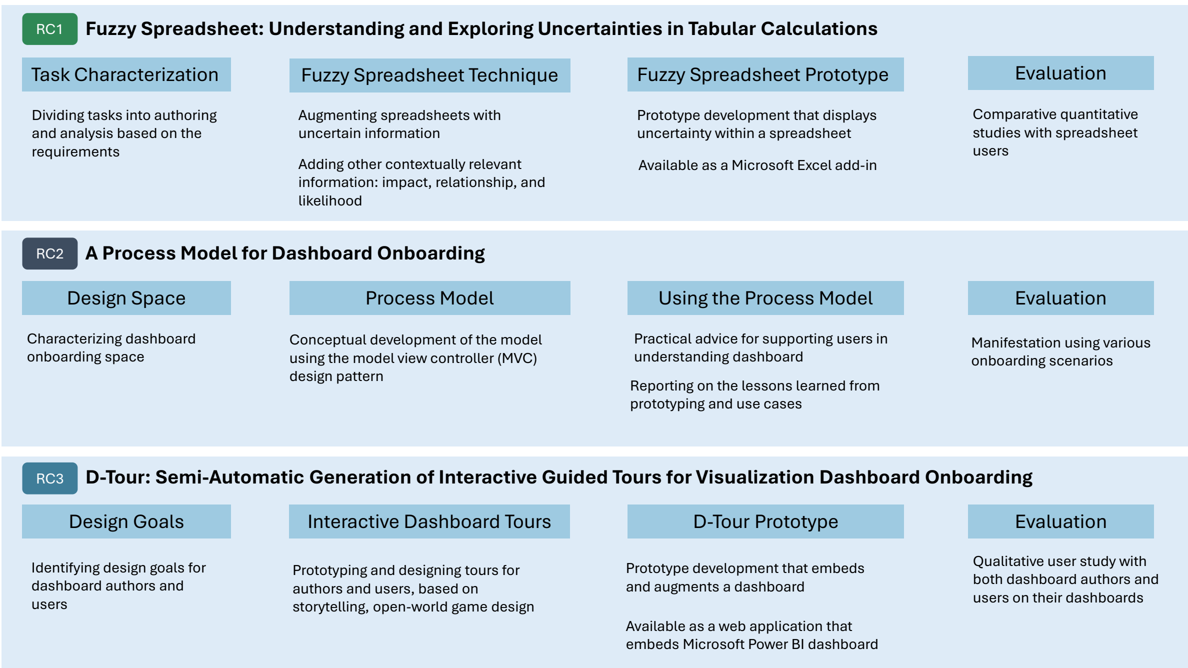


Fig. 1.4: RC1 is the first paper contribution: Fuzzy Spreadsheet [5]. RC2 is the second paper contribution: a process model for dashboard onboarding [6]. RC3 is the third paper contribution: D-Tour: semi-automatic dashboard tours for dashboard onboarding [7].

The thesis includes and explains the following three main publications. Each publication refers to a chapter of the thesis. I specify each publication on one or more of the processes shown in Figure 1.2, namely, preparation, analysis, deployment, and communication. The publications are described in the same sequence as they appear in the thesis. Due to varying review times, their publication dates might not match the order in which they are listed.

Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Eduard Gröller, and Marc Streit. "Fuzzy Spreadsheet: Understanding and Exploring Uncertainties in Tabular Calculations". In: *IEEE Transactions on Visualization and Computer Graphics* 29.2 (2023), pp. 1463–1477. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2021.3119212.

Fuzzy Spreadsheet is an approach that augments traditional spreadsheets with uncertain information using in-cell visualizations. With these compact visualizations, users can assess probability distributions and trace computational relationships within a spreadsheet. The key contribution of this work primarily lies in the analysis phase of the extended visual analytics pipeline, as shown in Figure 1.2. Certain aspects also extend to other pipeline segments, such as deployment and communication. *Analysis*: Our approach allows users to analyze how uncertainty propagates through the cells and explore the effects of hypothetical changes. These changes can be explored by directly manipulating the data inside the spreadsheet cells, which updates the in-cell visual encodings. *Deployment*: The Fuzzy Spreadsheet is available as an add-in for both the desktop and the web version of Microsoft Excel. *Communication*: The spreadsheet can be created by all data science professionals seeking to analyze the uncertainties present in their data. After exploring the uncertainties, they can communicate insights to other stakeholders using this information to make informed decisions.

My contributions: preparation of the manuscript; development of the Fuzzy Spreadsheet extension; task analysis; study of related work; evaluation.

Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Holger Stitz, Eduard Gröller, and Marc Streit. "A Process Model for Dashboard Onboarding". In: *Computer Graphics Forum* 41 (2022), pp. 501–513. DOI: 10.1111/cgf.14558.

The process of onboarding visualization dashboards to users is complex and requires careful consideration of various aspects such as what needs to be onboarded, how to onboard it, where to provide such an onboarding, when, and why. We formalize these questions based on existing literature as well as our experience with industrial collaborators and create a process model that addresses them to explain existing onboarding strategies and potentially help develop new techniques of onboarding. Our work introduces an onboarding loop alongside the dashboard usage loop. We discuss the most important building blocks of dashboards and the onboarding process, which touches upon all aspects of the extended visual analytics pipeline, as shown in Figure 1.2. *Preparation*, *Analysis*, and *Deployment*: The work discusses the raw data and their transformations that result in the data model. This data model can be represented by visual components with which users can interact. The interactions can then manipulate the underlying data model resulting in the usage loop. *Communication*: The main contribution of the work lies in the communication phase, where the author presents a dashboard to the users and onboards them to it. Within the onboarding loop, we specify the content, how it is presented to the user, and ways in which the user can interact with it. To incorporate various roles in the onboarding process and make them explicit, a highly complex structure was formed. We discuss this structure and the reason for obscuring the various roles of the involved actors in our work

My contributions: preparation of the manuscript; development of the process model; manifestation of onboarding scenarios; study of related work.

Vaishali Dhanoa, Andreas Hinterreiter, Vanessa Fediuk, Niklas Elmqvist, Eduard Gröller, and Marc Streit. "D-Tour: Semi-Automatic Generation of Interactive Guided Tours for Visualization Dashboard Onboarding". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3456347.

D-Tour is a semi-automatic approach to designing interactive dashboard onboarding experiences. It is based on the concepts of data-driven storytelling and open-world games and aims to preserve user agency. We demonstrate the applicability of our work using the D-TOUR Prototype, which allows the authors to craft interactive onboarding experiences for their end users. The main contribution is in the communication phase, as the approach allows authors to communicate information about visualizations using semi-automatically created onboarding narratives. *Deployment* and *Communication*: We use dashboards that are deployed in the Microsoft Power BI workspace and embed them in a custom web application which provides an interface for the authors to create their onboarding narratives. These narratives are published and shared with the end users, who can then interact with these prepared narratives to follow and choose their onboarding path. We discuss two roles: the author and the user of the onboarding material. Based on the extended visual analytics pipeline, any data science worker involved in the communication phase can assume the author's role.

My contributions: preparation of the manuscript; design and development of the D-TOUR Prototype; user study and its evaluation; scenarios; study of related work.

1.2.2 Secondary Contributions

Matej Vukovic, **Vaishali Dhanoa**, Markus Jäger, Conny Walchshofer, Josef Küng, Petra Krahwinkler, Belgin Mutlu, and Stefan Thalmann. "A Forecasting Model-Based Discovery of Causal Links of Key Influencing Performance Quality Indicators for Sinter Production Improvement". In: *AISTech2020 Proceedings of the Iron and Steel Technology Conference*. AIST, 2020, pp. 2028–2038. ISBN: 978-1-935117-88-9. DOI: 10.33313/380/218.

In this work, we built a forecasting model for the sintering process (used in steel production) through AI-based models. We also used visual analytics to display and analyze the relationship between the variables in the sinter production. We extended Ordino to show the correlations between given data and time-adjusted (time shift) data from the physical and chemical analysis to support the complex and time-consuming ML approaches.

My contributions: conceptual input; preparation of the manuscript.

Klaus Eckelt, Andreas Hinterreiter, Patrick Adelberger, Conny Walchshofer, **Vaishali**

Dhanoa, Christina Humer, Moritz Heckmann, Christian Steinparz, and Marc Streit. "Visual Exploration of Relationships and Structure in Low-Dimensional Embeddings". In: *IEEE Transactions on Visualization and Computer Graphics* 29.7 (2023), pp. 3312–3326. DOI: 10.1109/TVCG.2022.3156760.

Here we propose an interactive visual approach for exploring structural relationships in embeddings of high-dimensional data. Most existing techniques for the visual exploration of embeddings do not prioritize the structural relationships inherent in them. Therefore, we prioritize these relationships, such as item sequences, associations of items with groups, and hierarchies between groups of items, and display them using summary and difference visualizations.

My contributions: conceptual input; preparation of the manuscript.

Conny Walchshofer, **Vaishali Dhanoa**, Marc Streit, and Miriah Meyer. "Transitioning to a Commercial Dashboarding System: Socio-technical Observations and Opportunities". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2023), pp. 381–391. DOI: 10.1109/TVCG.2023.3326525.

In this work, we study and report on the socio-technical challenges as observed in a large manufacturing company where the data science workers transitioned to a commercial dashboard system. This work reports on the opinions of these workers and various user roles to discuss the challenges in training, creating, and using dashboards.

My contributions: conceptual input; study of the related work; design of the interview study; pilot interviews; participation in the interviews and survey; discussion of transcripts; discussion of the results; preparation of the manuscript.

We outline the main contributions in the next subsections: Sections 1.2.3, 1.2.4, 1.2.5 and elaborate upon them in individual chapters (Chapters 3, 4, 5).

1.2.3 Uncertainty Analysis Using Spreadsheets

Due to their ease of use and simple yet effective way of calculating values, spreadsheet-based tools are the number one choice for building, formalizing, and analysing simple models for budget planning and many other applications. A cell in a spreadsheet holds one specific value and gives a discrete, overprecise view of the underlying computational model. Therefore, spreadsheets are of limited use when investigating the inherent uncertainties of such models and answering what-if questions. Existing extensions [38, 39, 40, 41] typically require a complex modeling process that cannot easily be embedded in a tabular layout. Other forms of uncertainty exploration may require users to leave the familiar layout of spreadsheet-based tools and switch to a web-based application, for instance. Our analysis of uncertainty visualization tools, literature, and an interview with a principal contributor to the web-based uncertainty visualisation tool Guesstimate, revealed that users are reluctant to model their use cases in unfamiliar environments and tend to prefer the traditional spreadsheet layout. This motivated us to develop an approach that retains the familiar spreadsheet layout and augments it with compact visualizations to communicate the underlying uncertainty and support the user in the analysis process (Figure 1.2 (analysis)). We call this approach Fuzzy Spreadsheet [5].

In a Fuzzy Spreadsheet, a cell can hold and display a distribution of values. This integrated uncertainty handling immediately conveys sensitivity and robustness information. The fuzzification of the cells enables calculations not only with precise values but also with distributions and probabilities. We conservatively added and carefully crafted visuals to maintain the look and feel of a traditional spreadsheet. In addition to uncertainty visualization, users can also view, explore, and understand the impact of a cell on another cell and the relationship between them, which can be useful in modeling various scenarios to facilitate what-if analyses. Fuzzy Spreadsheet automatically extracts and visualizes this contextually relevant information based on a user-specified reference cell.

We evaluated its usability and the perceived mental effort required to complete different tasks through a user study. The results show that our approach outperforms traditional spreadsheets in terms of answer correctness, response time, and perceived mental effort in almost all tasks tested.

Spreadsheets are still widely used for performing analysis on data and using charts to view the analysis results. However, these charts are mainly static and are of limited use when performing complex analysis, which requires interactions to discover new insights. Therefore, for a comprehensive analysis, users move to other VA tools to identify patterns and enhance their decision-making process. Dashboards are commonly used for such interactive analysis [21]. Based on our experience with the collaborators and the literature on dashboards and their usage [8, 21, 42], we observed that end users subdivided into various user roles in Figure 1.2, have difficulty interpreting and interacting with the interactive visualizations present in the dashboards. This inherent interactivity of dashboards confused and overwhelmed many end users. In the absence of sufficient onboarding tailored to individual needs, many users returned to their traditional, static reports, which were easy to use, well-explained, and familiar to them [8, 20]. Dashboard onboarding emerged as one of the key approaches for addressing user concerns and supporting them as they familiarize themselves with the dashboards

In our subsequent works, we focused on (i) a process model for dashboard onboarding [6] (Section 1.2.4), and (ii) creating custom onboarding experiences that preserve the user's agency [7] (Section 1.2.5).

1.2.4 Dashboard Onboarding

A dashboard author initially creates most dashboards, which are then used by a diverse group of users. These users can have different skills and expertise and may range from the general public to analysts. Several studies [18, 19, 20, 8] have shown that non-expert users find it challenging to understand and use potentially complex datasets or visualizations. Dashboard authors typically provide supplementary materials to help dashboard users fill their knowledge gaps. These materials could include videos, documentation, interactive tutorials, or in-person meetings to answer user questions.

Onboarding a user to a visualization dashboard entails explaining its various components, including the chart types used, the data loaded, and the interactions available. The process of introducing a new dashboard to the user is what we refer to as dashboard onboarding. Depending on the end users, different onboarding strategies are employed. For instance, users with low visualization literacy may need in-depth training on visualizations and their interrelationships, while users with high visualization literacy may only need a high-level explanation of the visualizations and the data behind them. In addition to what needs to be onboarded to whom, it is equally important to realize how, when, and where the onboarding should occur. An example of an onboarding that occurs outside of the dashboard interface could be a video tutorial on a web-based platform. The dashboard users have the option to view this video either before or during their interaction with the dashboard. Alternatively, an automated, programmed tour, for instance, might be integrated and overlaid directly on the dashboard. The user can navigate through the tour in a step-by-step manner to gain a basic understanding of the dashboard before using it. The onboarding author decides the *what*, *how*, *where*, and *when* of the onboarding based on the overarching question of *why* a user needs to be onboarded in the first place. We created a process model for dashboard onboarding that formalizes and unifies diverse onboarding strategies.

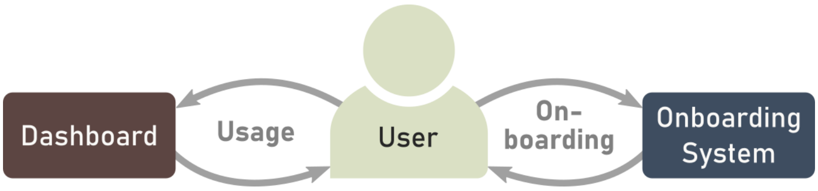


Fig. 1.5: Introducing the dashboard onboarding loop alongside the usage loop.

An abstract version of our model, as shown in Figure 1.5, introduces the onboarding loop alongside the dashboard usage loop. We discuss this process model in detail in Chapter 4, where both the usage and onboarding loop are explored in-depth. This helps us to understand existing techniques and potentially leads to the design and development of new onboarding approaches.

The onboarding loop describes how each onboarding strategy combines selected dashboard building blocks with an onboarding narrative. This narrative, in the form of text, video, audio, or human narration, is presented to the user through an interface. Using our process model, we demonstrate how real-world onboarding approaches can be described in Section 4.4.

We discuss four different real-world onboarding examples, i.e., video tutorials, programmed tours, documentation, and human narrators, and show how they can be derived from our process model. Finally, proposing a hypothetical AI-based approach, we exemplify how our model can serve as an actionable blueprint for developing new onboarding systems.

It is interesting to note that the conceptual AI-based scenario has now become a reality, with many commercial dashboard systems offering conversational interfaces embedded directly into the dashboards. For instance, Microsoft Copilot [43] provides users with a chat-based AI assistant that allows them to ask queries about specific dashboards and onboard themselves with the help of textual explanations provided by the assistant.

Our model puts the users in the centre and discusses the dashboard and onboarding loop from their perspective. We also attempted to include various roles similar to the ones described in Figure 1.2 and realized that the model became quite complex. To avoid this complexity, we obscured the various roles and depicted the user as the sole actor to illustrate the potential applications of the model in the description of onboarding processes.

1.2.5 Strategies for Dashboard Onboarding

The authoring of an onboarding experience is time-consuming and requires significant knowledge. In addition to the workings of the dashboard, the author must also understand their target audience, the content they need to deliver, how to deliver it, and various other factors that influence the onboarding process. Little guidance exists on how best to complete this authoring task and create an onboarding experience for the users.

Depending on their levels of expertise, end users being onboarded to a new dashboard can be either confused and overwhelmed or disinterested and disengaged. For the authors, it can be tedious and time-consuming to create an onboarding experience that matches the needs of their end users. We propose *interactive dashboard tours* (*D-Tours*) as semi-automated onboarding experiences. They preserve the agency of users with various levels of expertise to keep them interested and engaged. Our interactive tours concept draws from open-world game design to give the user freedom in choosing their path through onboarding. We have implemented the concept in a tool called D-TOUR PROTOTYPE, which allows authors to craft custom interactive dashboard tours from scratch or using ready-made templates. Automatically generated tours can still be customized to utilize different media (e.g., video, audio, and highlighting) or new narratives to produce an onboarding experience tailored to an individual user.

We demonstrate the usefulness of interactive dashboard tours through use cases and expert interviews. Our evaluation shows that authors found the automation in the D-Tour Prototype helpful and time-saving, and users found the created tours engaging and intuitive.

So far, we have primarily discussed the roles of dashboard authors and users, however, in practice, many other roles exist [4, 9]. In the next subsections, we outline the roles and responsibilities of various data science professionals, based on both research and our experience [4, 9]. We also highlight these roles within the extended visual analytics model, briefly explaining the model and describing how our research addresses its key aspects.

1.3 Structure

The structure of the thesis is based on the publications, i.e., main contributions. Chapter 2 summarizes the existing research work related to the main publications of the thesis. Chapter 3 focuses on Fuzzy Spreadsheet. Chapter 4 describes the process model for dashboard onboarding. The manifestation of the onboarding scenario as a programmed interface is described in Chapter 4 and becomes the basis of the implementation of the D-Tour work described in Chapter 5. Chapter 6 concludes the thesis with a focus on future work on the topics of storytelling, artificial intelligence (AI), and accessibility in the context of dashboards.

2 Related Work

This chapter summarizes the related work on the research goal of this thesis. This includes research on the analysis and communication aspects of the visual analytics(VA) pipeline, for spreadsheet-based analysis and analysis and communication using VA tools. The topics that are closely related to these works but do not fit into the overall frame of this thesis are discussed separately within their respective chapters.

2.1 Spreadsheet-based Analysis

Due to the ease of data entry and storage, spreadsheets are still a widely used software tool for analysis and visualization [44]. Their ability to process numerical data, such as by applying algebraic operations and exploring "what-if" scenarios, has proven them effective in businesses for budgeting, financial modeling, planning and managing projects [45, 46]. Several research works have demonstrated the use of spreadsheets for displaying and exploring visualizations with large, abstract, multidimensional datasets [46, 47, 48]. Bendre et al. [49, 50] developed DataSpread to integrate spreadsheets with databases for addressing the scalability issues inherent within spreadsheets. A spreadsheet program, such as Microsoft Excel, can now combine multiple data sources using Power Query [51], making them more powerful in terms of data handling.

In 2016, Hermans et al. [52] referred to spreadsheets as an end-user programming language and discussed the goals and challenges similar to that of any other software tool. They highlighted three main factors that contributed to the success of spreadsheets (i) live programming: immediate result on formula execution, with no compilation time, (ii) directness: the ability to directly manipulate data, metadata and calculations together in one view with easy access, and (iii) one click deployment: the universal availability on almost all computers that makes spreadsheets easy to access as an executable package consisting of data and calculations packed together. In 2022, Bartram et al. [53], also reported on the effectiveness of data tables, especially in spreadsheets, which are still widely used by data workers as part of their job. Their findings were similar to the ones echoed by Hermans et al. [52], where the directness and live programming nature of spreadsheets make them ubiquitous in analysis workflows to this day. They also emphasized the need to bring these capabilities to the visual analytics systems of tomorrow. Birch et al. [48] discuss the future of spreadsheets in the big data era, detailing the challenges and research opportunities in the field of spreadsheet-based analysis.

In our research question (RQ1), we use spreadsheets to convey uncertain information to support users in their decision-making process [6]. We investigated various techniques, both in the literature [54, 55] and commercially available tools [38, 39, 40] that display uncertainty within spreadsheets. A detailed investigation of these techniques can be found in Chapter 3.

2.2 Analysis Using VA Tools

While spreadsheets provide an effective way of displaying tabular information [53], they are of limited use when finding patterns or discovering insights using their static charts [56, 57]. For a more comprehensive and interactive analysis process, analysts move to different tools, such as Python, R, Matlab, Tableau, and Microsoft Power BI, among others, based on their expertise and needs [8].

The rest of this chapter focuses on the research on the empirical studies on the analysis process using both spreadsheet-based and VA tools in the real-world setting (Section 2.2.1), the rise of commercial dashboarding systems to address the challenges in the analysis process (Section 2.2.2), and ways to improve their adoption (Sections 2.2.2, 2.2.3, 2.2.4, 2.2.5).

2.2.1 Empirical Studies on the Data Analysis Process

In 2012, Kandel et al. [58] interviewed 35 analysts from 25 different organizations, reporting on their tasks and responsibilities, the tools they use for the analysis, their challenges, and their barriers to adopting VA tools. They identified five high-level tasks in the analysis efforts: (i) discovery: acquiring the data required to complete their tasks, (ii) wrangling: manipulating the data to use it for analysis, (iii) profiling: diagnosing data quality issues and inspecting data for outliers or errors, (iv) modeling: for summarizing or finding useful patterns, and (v) reporting: communicating assumptions and insights from the data. Each of these tasks comes with its set of challenges which require analysts to switch to different tools based on the task at hand. Visualizations are often applied at the late stages of their workflow. Therefore, they emphasized the need for integration of all these tasks within the VA tools to support the users in the analysis process. Their findings were echoed by Fisher et al. [59], who interviewed 16 analysts at Microsoft to uncover the pain points for big data analysis and the opportunity for better user experience. They discuss the challenges analysts face with acquiring data, choosing the right infrastructure to model and analyse them, and highlighting the role of visualization for inspecting data at multiple scales. Similarly, Kim et al. [60] interviewed data scientists in software development teams from several product groups at Microsoft. They reported on their working styles and described how software teams embrace data-driven insights and decisions. Batch and Elmqvist [42] reported on the visualization gap during the initial exploratory analysis. Based on their conversation with eight data scientists, they outlined ways to close the gap between visualization and data science, similar to Kendal et al. [58]. Example strategies include using the same tools for visualization as for data analysis, designing self-contained visualization components, and educating data scientists on various visualization techniques to improve visualization adoption. Another interview study with 30 professional data analysts by Alspaugh et al. [61] reported on the role of exploration in analysis, common challenges, and how analysts use software tools. Almost half of the participants in their interview reported using VA tools as their primary analysis software, which includes Tableau, SAS, Splunk, Stata, Alteryx, Periscope, and others. Many respondents expressed a desire for tool integration, similar to Batch and Elmqvist' [42] observations so that the burden on their focus is reduced.

Crisan et al. [4] describe the process of analysis with data in organizations and discuss the process of deployment and communication of these tools. For data science work, deployment is a necessary step to monitor and refine data and models. The same is true for visualizations, especially for visual analytics applications, which are routinely supplied with new data to explore and identify various trends [62, 63, 64, 65].

2.2.2 Commercial Dashboarding Systems

As discussed above, the visualization research community has extensively studied the use of visual analytics (VA) tools and highlighted the gaps needed for the effective adoption of VA tools [66, 58]. In an attempt to integrate analysis tasks in one tool and make visualization a part of every analysis step, commercial tools rooted in research were launched. These include Tableau from Polaris [67], Spotfire [68], and Advizor [69]. Their success led to other companies investing in building commercial VA systems, including Microsoft, SAP, Jaspersoft, and QlikView.

Due to their ability to cover the whole data analysis pipeline, from data preparation to dissemination [4], commercial dashboarding tools are widely used within organizations [8]. In 2012, Zhang et al. [70] conducted a comparative review of the state-of-the-art commercial systems. They identified challenges such as the need for supporting semi- and unstructured data, the use of advanced and customizable visualizations, and the opportunity for real-time analysis. In 2019, Behrisch et al. [71]extended this survey by reporting on the achievements of commercial VA systems since 2012 and proposed novel ideas to improve their integration within the analysis workflows. They highlighted the need for automated mechanisms to create visualizations and predictions and make VA tools available to many business users. With the rise in AI and its integration in current VA tools, such as Copilot for Microsoft [43], automated dashboards from a given data support the authors in the creation process.

Even though both the visualization research community and the commercial VA landscape have significantly improved the field of VA tools over the last few years, there are still barriers to their adoption in large-scale organizations [8]. As a secondary contribution, we report on these challenges that hinder the use of VA tools within a large organization [8]. Through interviews with 17 participants, we observed the difficulties in training, using, and creating dashboards from the perspectives of users with different jobs, skills, expertise, and roles. Based on these findings, we highlight opportunities for both companies and researchers to support the users in their transition to commercial visual analytics tools. One of the challenges faced by the users points to data and visualization literacy, especially among novice users, also reported by Tory et al. [20]. Similarly, Sarikaya et al. [18] highlight a critical gap in the current dashboards, questioning whether users are properly supported to use dashboards and whether dashboards can teach literacy skills. To address these challenges, we investigated the areas of storytelling, guidance, and onboarding to find strategies that can support users in their dashboard usage.

2.2.3 Data-driven Storytelling

One of the main objectives of visualization is the effective communication of insight [72]. Thus, by extension, data-driven storytelling is a key mechanism supporting this objective by providing a narrative communicating the intended message encapsulated within the data visualization [72] [73]. Effectively leveraging varied storytelling modalities, from annotated visuals to immersive presentations, enhances the narrative when delivered with a clearly defined purpose to a targeted audience. Additionally, storytelling can help in enhancing visual and data literacy [74], promoting user engagement [75] [76] [77] [72] [78] [79], and enhancing memorability [77] [80] [81] [82].

SketchStory [11] allows the authors to create visualizations with annotations to promote user engagement. Narvis [83] also allows for authoring narrative slideshows for introducing data visualizations. There are works that focus on the order in which the story events are told in the narration [84, 85, 86]. Apart from the order, the type of narration, from fixed to free-form, can also be authored, determining how the story is perceived. Fixed narration, commonly used in author-driven stories, can be authored using tools such as [83] [87] [88], where a narrative is created once, and the reader follows this sequence to consume the story. Free-form narration, commonly used in reader-driven stories, supports a more open-ended exploration, such as in [75] [89] [90]. Typically, data-driven stories are not just narrated, but can also be conveyed through video [87] [91], audio [91], text, annotations, slideshows [92], images [91] [93], webpages [74] [91].

These concepts from data-driven storytelling can be applied to dashboards to generate narrative dashboards, as highlighted in our secondary contribution [8] to enrich dashboards with insights and important context.

2.2.4 Guidance

Guidance refers to assisting a user in achieving their goal [94]. In the field of visual analytics, Ceneda et al. [94] characterized guidance in visual analytics in terms of knowledge gap, input and output, and guidance degree. They define guidance in VA as a *"dynamic, iterative, and forward-oriented process that aims to help users in carrying out analytical work using VA methods"*. Stoiber et al. [19] examine the perspectives of visualization onboarding and guidance in VA. They proposed a descriptive model on user assistance for VA that integrates the process of visualization onboarding and guidance.

We focus on onboarding as opposed to guidance, as our focus is to help users to understand, interpret, and use the applied VA methods. [95, 96] In our research, we focus on onboarding rather than guidance, as onboarding can be a useful first step for users to familiarize themselves with the dashboards before moving on to seeking guidance. Our research is motivated by storytelling concepts to improve the onboarding process.

2.2.5 Visualization Onboarding

Onboarding is a commonly used term in organizations where it refers to the process of integrating and acclimating new employees into an organization [97, 98]. The term gained traction in the user experience practitioner community, where they view onboarding as a process which can potentially increase the success rate of adopting a product [99]. In the field of visualization, Stoiber et al. [96] characterized the onboarding space for *single* visualizations, listing online guides [100] and cheat sheets [101] as well as more recent approaches, such as step-by-step guides and scrollytelling [102].

We delve into the concepts of single and multiple visualization onboarding in Chapter 4 (RQ2) and Chapter 5 (RQ3). Our research shows that the research in the field of onboarding for dashboards has recently gained momentum and is inspired by the research on single visualization onboarding, which is well established [19, 12].

Our research work focuses broadly on the analysis and communication of data using spreadsheets and dashboards. The rest of the thesis delves into these topics, describes our contribution, addresses the research gap in these areas, and explains how our solutions can help advance the field of spreadsheet and dashboard research.

3 Fuzzy Spreadsheet: Understanding and Exploring Uncertainties in Tabular Calculations

Spreadsheets, due to their intuitive and direct approach to computation, are one of the most widely used tools for building and formalizing models in areas as diverse as science, finance, and business [103, 104]. However, the precision implicit in traditional spreadsheet cells [105] often hinders users in performing calculations with uncertain values. Established spreadsheet tools—such as Microsoft Excel [106], Google Sheets [107], and Apache OpenOffice Calc [108] do not take uncertainty into account at all or lack features for exploring how uncertainties propagate and combine throughout calculations.

Integrated handling of variability would be helpful, for instance, to model estimated future values [54] or to plan a budget. To cope with the lack of built-in support for working with uncertain information, users often simplify their problems by calculating alternative scenarios, for instance, covering the worst, best, and average cases. This oversimplification of the problem results in a limited understanding of the complete scenario. To better support users who want to properly include uncertainty in their calculations, a tool must communicate effectively how individual components influence the final result of the calculations— in terms of both expected value and of uncertainty. To be integrated into existing workflows, such a tool should retain the familiar spreadsheet interface and enhance it with additional information about uncertainty.

Our **primary contribution** is the Fuzzy Spreadsheet approach, which augments well-established spreadsheets with compact in-cell visualizations. These visualizations allow users to assess probability distributions and trace computational relationships directly within the spreadsheet cells. Users can analyze how uncertainty propagates through the values and explore the effects of hypothetical changes. As a **secondary contribution** we present the results of a user study, which indicates that Fuzzy Spreadsheet is effective in working with uncertain information in tabular calculations.

3.1 Characterizing Uncertainty in Spreadsheets

To discuss the limitations of traditional spreadsheets and to introduce our terminology, we present an example of a spreadsheet that models the resources required for the maintenance of cars (see Figure 3.1). We use this spreadsheet as a guiding example throughout the paper. Let us assume that a classic car enthusiast wants to manage the predicted costs for two of her cars. For the first car, she sums the cost of the checkup and the cost of replacing its expendable parts. For the second car, she plans to upgrade the engine. To reflect this, she enters the cost of buying a new engine and subtracts the selling price of the old one, which results in the total cost for Car 2. To obtain the Grand Total Cost, she combines the summed costs of each car.

Car Maintenance

	Cost (in 1000 €)	Probability	Time Spent (in hours)	Standard Deviation
Car 1				
Checkup	3	1	5	1.5
Spare Part A	2	0.75	2	1
Spare Part B	1.5	0.5	3	0.5
Total	6.5	0.38	10	1.87
Car 2				
New Engine (SP)	3	1	3	1
Old Engine (CP)	1	0.8	1	0.5
Total	2	0.80	4	1.12
Grand Total	**8.5**	0.30	**14**	2.18

Fig. 3.1: A simplified spreadsheet for planning the maintenance of two classic cars, containing costs and time required for the maintenance of each car and uncertainty values provided in the adjacent columns.

To make the underlying network of cells more transparent and to understand which cells influence other cells, we use a computational graph [109]. Figure 3.2 shows the computational graph of the example above that models the relationships between cells and highlights the functional expressions that combine them.

Based on experience with Car 1 in the previous years, she knows that each spare part has a different chance of failing within a year. This makes the cost uncertain, as it will only be incurred if the spare part fails. She models such costs with a Bernoulli distribution, indicating the cost expected in the case of failure along with the probability of failure. For the second car, her friend shows an interest in buying the engine for the

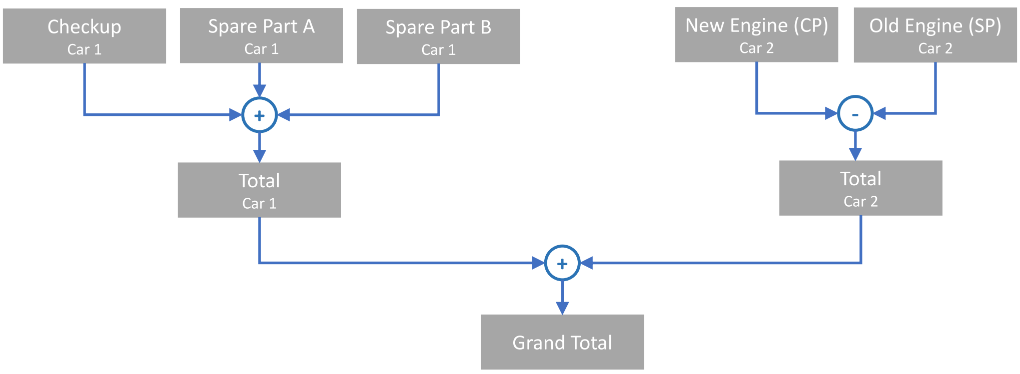


Fig. 3.2: Computational graph of the example in Figure 3.1, showing the relationships between cells.

Term	Definition
Bernoulli probability	Probability $p = P(X = A)$ that a variable X modeled via a Bernoulli distribution assumes the value A
Relationship	Relationship between a cell and other cells based on the direction of influence (input/output) and the degree of neighborhood
Degree of Neighborhood (DoN)	Distance between two cells in the computational graph
Impact	Contribution of a cell value to the value of another cell, expressed as percentage

Table 3.1: Terminology used for describing uncertainty and other relevant concepts.

price specified. As the deal has not yet been closed, she captures this uncertainty in the same way. She computes the total cost by summing the input costs and multiplying the result with the product of the individual probabilities. She observes that, according to the spreadsheet (shown in Figure 3.1), the Total Cost (Car 1) is 6.5 k Euros with a probability of 38 %. However, since such a calculation covers only one of many possible paths in a probability tree, she receives no indication of the value(s) for the remaining 62 % of the cases. In order to gain a complete overview of the actual cost, all the influencing costs must be combined correctly into a distribution. Furthermore, the contribution of each influencing cost on the distribution of the Total Cost (Car 1) can be analyzed if each influencing cost is treated as a distribution rather than a scalar value.

As a traditional spreadsheet lacks this ability to treat a cell as a distribution, she must somehow keep track of the underlying distributions herself. The distribution of an independent cell, such as Spare Part A (Car 1), which is not influenced by other cells, may still be easy to understand. However, as more of these cells are combined by using formulas to produce cells such as Total Cost (Car 1) and Grand Total, the calculation becomes increasingly complex. The functional expressions in these cells, in this case "sum", no longer perform computations on scalar values, but on distributions. Hence, the Total Cost (Car 1) is then obtained as the sum of individual distributions. In the absence of additional visual cues, it becomes difficult to estimate how the uncertainty propagates from independent cells (Spare Parts A & B (Car 1)) to the cells which depend on them (Total Cost (Car 1)). Furthermore, the car owner decides to include the time it would take her to inspect and later fix or remove each part to avoid paying a professional car mechanic. She enters the time required for each part. However, the time spent on each part varies. Since she needs to spend time on inspecting parts regardless of a potential failure, she wants to model the time spent on each part with normal distributions (as opposed to the costs, which she modeled with Bernoulli distributions). She does so by specifying a standard deviation for each part in a column adjacent to the estimated mean. The Total Time spent is then computed by adding the influencing distributions. She now has to monitor not only the uncertainty in cost but also in time. Using only the basic functionality provided by typical spreadsheet programs, it becomes difficult to keep track of the distributions and how they propagate throughout the spreadsheet. The Fuzzy Spreadsheet approach supports users in achieving these tasks.

In some cases, it may be relevant to analyze the impact a cell has on another one, for instance, to identify highly influential cells such as Total Cost (Car 1) and its impact on the Grand Total. Conceptually, a cell's impact can be defined as how much its value contributes to the value of another cell. In some cases, it is sufficient to simply look at the values of the cells to identify those with the highest impact. However, determining the exact impact a cell has on another one requires calculation. We compute this impact as a percentage of how much the value x_i of cell i contributes to the value x_j of cell j. If i is a first-degree influencing neighbor of j, i.e., $i \in \mathcal{N}_1(j)$, then we define its impact as:

$$\mathscr{I}(i \rightarrow j) = \frac{\sigma \odot x_i}{\sum_{k \in \mathcal{N}_1(j)} x_k}. \tag{3.1}$$

Here, $\sigma = +1$ if i contributes to j via a summation, and $\sigma = -1$ if i contributes to j via a difference. The total relative sum of the impact percentages of each contributing cell c_i must be 100 %. This direct computation involving simple addition or subtraction is straightforward, but longer calculations involving intermediate results, such as a difference of sums, can make assessing the impact challenging. We categorize the nature of this impact as positive or negative based on the sign and relationship of cell j with i. In a spreadsheet, an additional column may be created to compute the impact that value x_i in cell i would have on value x_j of cell j. For a cell i, as its impact value influences j, j must be re-evaluated every time i changes.

Finally, it can often be helpful to explore the results of hypothetical changes to the spreadsheet to answer questions such as "What would happen to the Grand Total Cost if I were to change the chance of failure of Spare Part A of Car 1 from 0.75 to 0.5?" The model must be able to incorporate the new uncertainty for Spare Part A of Car 1 and propagate the change through the computational graph. The analytical concept of a *what-if analysis* often involves changing the certainty of values to create and analyze different scenarios. Such an analysis requires considerable mental effort if intermediate changes are not tracked and no visual clues are provided. With Fuzzy Spreadsheet, we aim to support users in maintaining their mental models during calculations involving uncertainty. Brief descriptions of the key terms introduced in this section are given in Table 3.1.

3.2 Related Work

Uncertainty is inherent in data collection, processing, and sensemaking. A recent survey showed a consensus among visualization researchers on the importance of using visualization to communicate not only the data itself, but also the underlying uncertainty [110]. Exploring and understanding variabilities may have a considerable impact on the outcome of decision-making [111, 112]. Research into how uncertain values can be visualized effectively has been ongoing for centuries [113, 114]. Various uncertainty visualization techniques have been applied in a diverse range of fields, such as medicine [115], geoscience [116], business intelligence [117], and fishery [118]. As technology develops rapidly, tools are proliferating, but current applications focus on showing complex information and do not make use of the familiar spreadsheet layout to present different scenarios—such as worst, best, and expected case—in a way that is easy to understand and visualize [119, 120]. With our work, we aim to deepen the understanding of uncertain values and their use in tabular calculations.

3.2.1 Uncertainty Visualization

Uncertainty can be visualized for zero-, one-, or higher- dimensional data [120]. Sanyal et al. [121] introduced a framework for data dimensionality (0D, 1D, 2D, 3D), visualization approaches (scalar, vector, tensor), and uncertainty visualization techniques (e.g., blurring, transparency, noise). Typical techniques for showing zero-dimensional data (after decoupling of the temporal dimension) include changes in glyph size to indicate a variation in the data points [122]. As with one-dimensional data, line blurring and transparency changes are common techniques to indicate the level of certainty [121]. Further, probability density functions [123, 124] are used to represent uncertainty, for instance, by line and bar charts [125, 126], dotplots [127], violin/boxplots [128], and heatmaps [129].

3.2.2 Spreadsheet-based Tools

We investigated multiple applications, such as WebCharts [130] and Google Visualizations [131], which allow users to create and use web applications in spreadsheets. However, we decided to limit our discussion to spreadsheet-based tools that address the encoding and propagation of uncertainty information.

Streit et al. [54] proposed a technique for augmenting a spreadsheet with uncertain information. To introduce uncertainty in the calculations, users can specify an interval or range in a cell. The underlying theoretical concept is known as interval arithmetic. The uncertainty information is then propagated to other cells that use these ranges as inputs. To convey to the user that—as an effect of the propagation—these cells now contain intervals, affected cells are highlighted by shading. Finally, these intervals are plotted in a chart that represents the uncertainty.

The FuziCalc tool [55], intended for modeling under uncertainty, followed a similar approach. FuziCalc allowed the user to enter fuzzy inputs in the form of shapes, such as triangles representing best, worst, and expected cases. It then generated a fuzzy output from fuzzy input cells. This way, it reduced the mental effort for users by not exposing them to the complex underlying computations. Since FuziCalc was developed as an independent spreadsheet tool, it missed fundamental features of conventional general-purpose spreadsheet tools. In most cases, it was used in addition to conventional spreadsheet tools and not as a replacement, which resulted in a cumbersome workflow. Since this commercial software is no longer available, it could not be tested.

More recent tools for uncertainty visualization in spreadsheets are Palisade @Risk [38], Oracle Crystal Ball [39], and SIPmath [40]. These tools use Monte Carlo simulation to compute possible outcomes of uncertain events [132]. They are commercially available as extensions for Excel, and allow users to explore different scenarios and provide results in the familiar Excel charts. They share the concept that the user can input a distribution for the cells by using either a formula or a graphical user interface. The user can choose distributions for cells that are not influenced by other cells and compute the model for a *reference cell* (which is usually a cell that is influenced by input cells). In the case of @Risk and Oracle Crystal Ball, the cells are then colored to indicate that they no longer contain exact but fuzzy values based on probability distributions. More information on a distribution appears upon mouse-over action and click-based interaction. The user can interact with the distribution to analyze the probability of finding a value in a particular range. However, these tools lack an overview of the underlying uncertainty across the spreadsheet, and require a complex modeling process of the distributions for each cell. Furthermore, for users who are not familiar with probability distributions, it is challenging to define and interpret them [133]. In the case of SIPmath, histograms in the form of sparklines are added in the cells to represent the underlying distribution. However, the histograms are stretched in both directions to fill out the cell, which makes it difficult to compare distributions across cells, especially if the values on the horizontal axis are not the same.

A tool that avoids the complexity of the modeling process is Guesstimate [41]. It was developed as an independent web-based spreadsheet tool that allows the user to easily input a fuzzy number as a range, an interval, or a distribution, and immediately provides an output based on the relationship. Relationships between the cells are shown as links connecting the related cells. Similar to Palisade @Risk, Oracle Crystal Ball, and SIPmath, it uses Monte Carlo simulation to compute the propagation of uncertainty. In the course of our investigation, we contacted one of the main contributors of Guesstimate. He informed us that users generally prefer staying close to traditional spreadsheet-based approaches and primarily rely on normal distributions due to a lack of knowledge about applying other distributions, even though further ones are supported. This confirmed our findings in the context of the FuziCalc software that users are less likely to model their use cases in an unfamiliar environment. We found that existing tools for uncertainty exploration with *what-if analyses* are either too complex for a novice user or lack the wide range of features available in general-purpose spreadsheet tools. This inspired us to develop an approach that (1) retains the layout of a traditional spreadsheet and (2) is augmented with compact visualizations to communicate uncertainty.

3.3 Task Characterization

We collected the most prevalent user tasks that are supported by various related tools (see Section 3.2 for a survey) and put these tasks into the context of existing analysis task frameworks [134, 135]. For the purpose of discussing Fuzzy Spreadsheet, we categorize user tasks in two different ways.

First, we assign the tasks to one of two *phases* of the typical spreadsheet workflow: *Authoring* and *Analysis*. In the Authoring phase, users set up the content and structure of the spreadsheets. In the Analysis phase, users seek to understand the results of calculations and draw conclusions for their applications. Complex problems may require users to go through multiple iterated Authoring and Analysis phases.

Second, we label user tasks by the *requirements* they impose on Fuzzy Spreadsheet. In order to facilitate decision-making in the face of uncertainty, Fuzzy Spreadsheet must extend traditional spreadsheets in two ways. To reflect on the extensions necessary, we define the following three task requirements:

- **Basic spreadsheet functionality** is required for all tasks that can be readily performed in a typical spreadsheet tool without any extensions, such as inputting numbers and referencing cells in formulas.

- **Computational graph parsing** is required for all tasks that involve extracting relationship information from the computational graph underlying the spreadsheet.

Tasks requiring basic spreadsheet functionality
Tasks requiring parsing of the computational graph
Tasks requiring fuzzification

Authoring		Tabulate Numbers
		Set Up Relations
		Specify Uncertainties & Introduce Alternatives
Analysis	T1	Look up Values
	T2	Trace Relations
	T3	Assess Impact
	T4	Expose Uncertainty
	T5	Formulate Cause and Effect
	T6	Compare Probability Distributions

Table 3.2: User tasks for Fuzzy Spreadsheet. Tasks are divided into Authoring and Analysis phases, and labeled depending on the requirements.

Fuzzification of the spreadsheet is required whenever tasks cause a transition from exact numbers to probability distributions.

Not all user tasks need to be assignable to exactly one of these requirements. Some involve only basic spreadsheet functionality, but may lead to fuzzification or enable improved analysis of the computational graph. Other tasks may be equivalent to those performed in non-augmented spreadsheets, but may acquire a new meaning in the context of uncertainty. Table 3.2 summarizes the user tasks we identified for spreadsheets with uncertainty propagation and lists their requirements and to which phase we assigned them. The following sections describe the user tasks in more detail.

3.3.1 Authoring Tasks

In the authoring phase, users set up the initial content and structure of the spreadsheets. We extracted three tasks that are essential in this phase. First, users **Tabulate Numbers**, that is, they input their data as exact numbers.

Each value is a potential node in the spreadsheet's underlying computational graph. Second, users **Set Up Relations** by inputting formulas, which introduce links between nodes in the computational graph. Third, users **Specify Uncertainties & Introduce Alternatives**. In this step, essential for the fuzzification of spreadsheets, users specify which cells should no longer be treated as exact, but as fuzzy values. To the users, this task can be equivalent to a *Tabulate Numbers* task, with the important conceptual difference that the input numbers are given a special meaning in the subsequent analysis phase. In particular, input values adjacent to "fuzzified" cells are interpreted as the parameters of underlying probability distributions (see Section 3.4.1).

3.3.2 Analysis Tasks

While the authoring tasks are important for setting up content and structure of the spreadsheet, Fuzzy Spreadsheet focuses mainly on enabling new tasks in the analysis phase. The analysis tasks are extracted from the rationale-based tasks identified by Amar et al. [135] and extended to incorporate tasks that played a role in the tools we surveyed. Additionally, we also cover the four search tasks (Lookup, Locate, Browse, Explore) and two query tasks (Identify and Compare) from the Multi-Level Task Typology framework of Brehmer and Munzner [134]. Our discussion of the visual encodings used in Fuzzy Spreadsheet and of the user study results are based on the following analysis tasks:

T1 **Look up Values**

The most basic analysis task when working with spreadsheets is to look up values in cells. In the same way that input tasks can acquire a special meaning if they introduce uncertainty to the spreadsheet, so do lookup tasks that retrieve these special values.

T2 **Trace Relations**

Tracing relationships between cells in a spreadsheet is important for understanding the "flow" of computations. As discussed in Section 3.1, this flow is described by the spreadsheet's underlying computational graph. In classic spreadsheet tools, the computational graph is, for the most part, hidden from the user. Given a selected cell, only the immediate inputs used in formulas in that cell can be readily retrieved. However, *Trace Relations* is a more general, bidirectional task that also includes finding cells that are influenced by other cells. Neither *influencing* nor *influenced by* cells are limited to direct neighbors in the computational graph, but can be extended to a higher degree of neighborhood (DoN).

T3 **Assess Impact**

Having located cells that influence another cell, users often need to gain a better understanding of the individual contributions of each of these influencing cells. In Section 3.1, we introduced the *impact* as a direct measure of how strongly the value of one cell contributes to the result in another cell. In simple cases, such as the addition of multiple values, assessing the impact can be as straightforward as looking up several values and comparing them. If more elaborate formulas are used, and/or in the case of tracing longer paths in the computational graph, assessing impacts is challenging in standard spreadsheet tools.

T4 **Expose Uncertainty**

In the authoring phase, users *Specify Uncertainties & Introduce Alternatives*. As a result, computations in the spreadsheet can no longer be based on exact numbers but require probability distributions. Users must be able to assess these distributions, preferably directly in the spreadsheet cells without auxiliary calculations. Depending on the level of detail with which they want to analyze these probability distributions, users must perform a number of low-level visualization tasks, such as assessing shapes, finding extrema, and estimating areas under the curve.

T5 **Formulate Cause and Effect**

While adding uncertainty to spreadsheets allows users to make more informed decisions, it also makes interpreting the relationships between cells in a spreadsheet more challenging. As described in Section 3.1, a *what-if analysis* is a powerful approach that allows users to safely explore the results of hypothetical changes in an attempt to better understand which values are important. To enable this, users must be able to change the spreadsheet reversibly. When users end their exploratory analysis, they can decide to either keep the new results or revert back to the initial values. Hypothetical changes typically include changing the parameters of distributions while keeping the topology of the computational graph unaltered. Given the density of data in typical spreadsheets, making reversible changes might not be enough for an effective *what-if analysis*. Users must be able to see how a change in one cell propagates through the computational graph and how each change leads to additional, implicit changes in other cells. Explicitly addressing this subtask by visualizing changes prevents users from having to *Trace Relations* (T2) repeatedly in the course of a *what-if analysis*.

T6 **Compare Probability Distributions**

In later stages of the analysis, most of the intermediate and final results of calculations in the spreadsheet are modeled by probability distributions. This means that comparing two or more cells requires comparing two or more probability distributions. We refer to this as *between-cell* comparison. Additionally, hypothetical changes in a cell during a *what-if analysis* may entail comparing old and new results within a single cell. We refer to this as *within-cell* comparison.

Fuzzy Spreadsheet addresses most of these analysis tasks by introducing compact visualizations directly in the spreadsheet cells, with additional detailed information in a separate side panel. The *between-cell* comparison is supported by in-cell encodings. Fuzzy Spreadsheet allows users to compare probability distributions of two or more cells by means of individual encodings presented within the cells. Detailed information about each cell, such as the probability distribution with computed mean and standard deviation, can be viewed in the side panel upon selecting a cell. During a *what-if analysis*, the encodings of the old and the new probability distributions are stacked within a cell, which facilitates a *within-cell* comparison. The within-cell changes can also be viewed in the side panel.

3.4 Fuzzy Spreadsheet Technique

We designed Fuzzy Spreadsheet as an extension to traditional spreadsheet programs, users can incorporate our solution into their familiar workflows. Fuzzy Spreadsheet comprises three parts: (1) functionalities for parsing the spreadsheet to extract the presence and propagation of uncertain information; (2) carefully chosen, compact visualizations that are embedded directly in the spreadsheet cells; and (3) a side panel with the user interface and additional information. Figure 3.3 shows an example spreadsheet with some of the Fuzzy Spreadsheet encodings (the traditional spreadsheet version is shown in Figure 3.1). We first discuss the pre-processing needed for calculating the propagation of uncertainty information, and how it relates to the authoring tasks. We then describe how users can control the in-cell visualizations from the side panel, and explain in detail the design of the visual components.

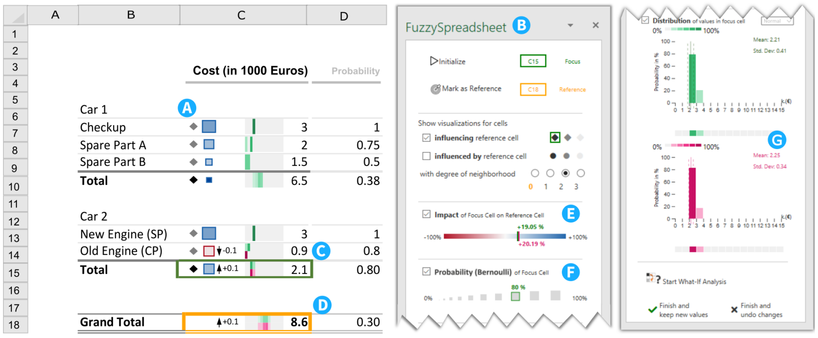


Fig. 3.3: Fuzzy Spreadsheet applied to a subset of the car-maintenance example in Figure 3.1. Visualizations are embedded directly in the cells (A) to indicate impact, Bernoulli probability, relationship, and probability distributions. The side panel (B) allows users to control the visualizations. Users select a cell (C) and explore its influence on a reference cell (D). The active legends (E & F) indicate the original values encoded by the impact and Bernoulli visualizations (green markers) and new values (dark pink markers) obtained during a *what-if analysis*. Here the user reduced the value of *Old Engine* by 0.1. The detailed distribution charts (G) for the original and the new values of (C) are also displayed, and the probability ranges can be viewed by hovering over them.

3.4.1 Computational Graph Parsing and Uncertainty Calculations

Before relational and uncertainty information can be encoded and interacted with, the relevant content must be extracted from the spreadsheet. To this end, Fuzzy Spreadsheet parses the spreadsheet and its underlying computational graph. Fuzzy Spreadsheet automatically treats cells as fuzzy if they are the result of an "average" calculation. Other cells can be made fuzzy by wrapping their content with Excel's "average" function. Fuzzy Spreadsheet then looks for additional uncertainty information in the adjacent cells. By default, the cell itself is treated as the mean of a normal distribution, and the cell to the right as the standard deviation. In addition, Fuzzy Spreadsheet provides a shortcut for multiplying a value with the result of a Bernoulli distribution by looking for the corresponding probability in the next but one cell. We focused on these two distributions due to a discussion with the creator of the Guesstimate software [41] in which he mentioned that typical users almost exclusively choose normal and Bernoulli distributions for their models. Advanced users can also switch to other distributions via the side panel, which results in the adjacent cells to be interpreted differently depending on the distribution chosen. In addition to normal and Bernoulli distribution, Fuzzy Spreadsheet also supports uniform and Poisson distributions. We discuss the advantages and limitations of letting users *Specify Uncertainties & Introduce Alternatives* via a simple *Tabulate Numbers* task in Section 3.8.

In order to perform an analysis task (see Section 3.3.2), a user needs to choose a *reference cell*—typically a cell of interest whose value depends on other cells. The Fuzzy Spreadsheet extension stores the address of this cell and uses it as a reference for all future computations. This gives the user the possibility of selecting another cell and viewing its details with respect to the reference cell. Thus, Fuzzy Spreadsheet allows the user to focus on two cells at any given time: a *reference cell* and a *selected cell*.

For the analysis tasks, the extension performs the computations required for obtaining impact values and probability distributions, as explained in Section 3.1. The tool then performs a Monte Carlo simulation to obtain the distributions. Due to computational limitations, we use 100 samples for each simulation. We briefly discuss the scalability of Fuzzy Spreadsheet in Section 3.8. During a *what-if analysis*, after a cell value is changed, an automatic parsing of the computational graph is performed to derive updated results. If the newly obtained results differ from those of the initial parsing process, the cell encoding for the probability distribution and the side-panel visualization for all selected options is updated.

3.4.2 Side Panel

The *side panel*, as shown in Figure 3.3(B), provides the control interface for our extension. To start working with Fuzzy Spreadsheet, users need to trigger the parsing and the creation of the computational graph by clicking on the *Initialize* button. Users must then select a cell in the spreadsheet and mark it as the reference cell. The reference cell is automatically highlighted with an orange border in the spreadsheet (see Figure 3.3(D)), and its address is shown in the same color in the upper-right corner of the side panel.

Next, users select which kind of cells they are interested in: cells *influencing* the reference cell or cells that are *influenced by* the reference cell. Subsequently, they choose the degree of neighborhood (DoN) up to which they want to explore related cells. A DoN of zero refers to the reference cell itself. A DoN of one corresponds to the direct neighbors (connected by links in the computational graph) of the reference cell, excluding the reference cell itself. A DoN of two corresponds to the direct neighbors of the direct neighbors, including the cells with a DoN of one. Up to this point (DoN of zero), no visual changes have appeared in the spreadsheet. To indicate that a DoN of zero limits the selection to the reference cell itself, the zero option is colored in the same orange hue that is used for the borders of the reference cell. This neighborhood-selection dialog simultaneously serves as a legend for the relationship encoding (see Section 3.4.3). We refer to this dialog/legend hybrid as *active legend*. In our Excel extension prototype (see Section 3.5), we use a dark green shade for all information related to the selected cell to keep the design consistent with Excel's default color setting for the border of the selected cell (see Figure 3.3 (C)).

Based on this information, users can display visualizations of impact, Bernoulli probability, and distribution for the selected and related cells on demand. At any point, they can change the DoN, which controls the overall amount of information displayed. Furthermore, users can start a *what-if analysis* by activating a button at the bottom of the side panel. We explain for each visualization separately how this analysis mode affects the encoding.

3.4.3 Compact Visual Encoding

Given the small cell size in traditional spreadsheets, we had to evaluate encodings that work effectively in this constrained space. Sparklines and glyph-based encodings [122] are one option. Nobre et al. [136] introduced an aggregated layout in tabular cells that could be adapted to show fuzzy values within a cell. Box plots, heatmaps, sparklines, histograms, and dotplots used in Taggle by Furmanova et al. [137] are also possible choices for in-cell encodings. All these visualizations inspired our final encodings. The cell encodings are designed to give overview information about the reference cell and its related cells—serving as an initial guide to identifying critical cells. We justify our final choice of encoding for each of the suggested visualizations in individual sections below.

Relationship Encoding

To enable Trace Relations tasks (T2), each cell of interest is equipped with a relationship marker. The shape of this marker encodes the direction of influence with respect to the reference cell. A diamond marker indicates a cell that *influences* the reference cell, and a disk indicates a cell that is *influenced by* the reference cell. We chose these shapes to avoid cross-talk with the impact and Bernoulli encodings. The side panel also emphasizes this information by outlining the marker in dark green if a related cell is selected. Additionally, the DoN is encoded in the brightness of the marker.

In the design process, we considered two alternatives to the categorical relationship markers. First, we considered drawing lines that connect related cells, which would have represented the underlying node-link structure of the computational graph. In Excel, users already have the option to turn on this kind of encoding as an overlay. Jannach et al. [138] also mention research works which used a similar representation for the relationship encoding, such as the work of Chen et al. [139], where differently colored arrows indicated the varying degree of neighborhood. Hermans et al. [140] used a flow diagram to represent the underlying relationship structure directly in the spreadsheet. However, we found that this encoding caused visual clutter and was prone to drastic changes depending on the layout (i.e., positioning) of the cells chosen by the user in the Authoring phase. Second, we considered small arrow-like glyphs or stubs [141] pointing towards or away from the reference cell. While this encoding was easy to understand in simple cases, we discarded it because identical glyphs would have acquired different meanings depending on their locations in the spreadsheet. The current encoding has the advantage that it is fully independent of the spreadsheet layout chosen by the user in the Authoring phase, and the markers make it easy to quickly spot the cells of interest.

Impact Encoding

Impact markers enable users to *Assess Impact* (T3). We add boxes of fixed size on the left side of a cell and color them according to how strongly the value in the cell influences the reference cell (see Figure 3.3(A)). We use a diverging color scale from dark red (strong negative impact) to white (no impact) and dark blue (strong positive impact) to help users to quickly identify the sign of a cell's contribution to the reference cell. Furthermore, the dark colors at each end of the scale make high-impact cells more salient, while cells with an (almost) white impact marker can be safely disregarded by users when gaining an overview.

The side panel also serves as an active legend for the impact. The active legend consists of a red/blue color scale (see Figure 3.3(E)). Using a dark green marker, the legend shows—for a selected cell—either its impact on the reference cell or the impact of the reference cell on it. Using the same dark green which is used as the border color of the selected cell, the legend also shows the value of the impact in percent. The position of this marker changes dynamically when the selected cell is changed. To support *Formulate Cause and Effect* tasks (T5) during a *what-if analysis*, a second, dark pink marker is shown in the active legend if any of the hypothetical changes affect its value, as indicated in the legend in Figure 3.3(E).

Bernoulli Encoding

Based on our decision to facilitate the use of Bernoulli distributions (see also Section 3.4.1), we decided to include an additional indicator to highlight values in cells that are the results of "binary scenarios". The area of this marker encodes the probability that all binary outcomes leading to this result have an outcome of one rather than zero. If only the Bernoulli but not the impact visualization is switched on, we show gray markers with corresponding areas. If both Bernoulli and impact visualizations are switched on, we encode the Bernoulli probability in the size of the impact marker and use the color from the impact encoding.

The side panel also functions as an active legend for the Bernoulli probability. The active legend for the Bernoulli probability consists of grey squares placed in ascending order of their area. For a selected cell, we show the probability value (in percent) in the legend and draw a dark green frame around the corresponding square, as seen in Figure 3.3(F). As for the impact encoding, a second, dark pink marker in the active legend indicates values that changed during *what-if analyses*.

Probability Distribution Encoding

Users can switch on the probability visualization in the side panel to show small, in-cell distribution heatmaps attached to the cells of interest (see Figure 3.3). The heatmaps serve as a first visual aid for *Expose Uncertainty* tasks (T4). Each heatmap is based on a binned histogram of samples drawn from a continuous probability distribution $p(x)$. The value x increases from left to right, and the heatmap is divided into a fixed number of intervals I (i.e., the bins). The range of the heatmap is based on the minimum and maximum cell values present in the spreadsheet, and is the same for all cells to ensure an unbiased comparison. Due to the limited space, the range cannot be shown numerically or quantitatively within a cell. However, as the vertical space in the side panel is not limited, we indicate the range of the heatmap on a shared x-axis with a bar chart for the detailed distribution plot, as shown in Figure 3.3(G). The horizontal orientation of the distribution heatmaps facilitates comparison between values of cells in the same column (T6) and makes optimal use of the space provided by typically sized spreadsheet cells, which is an advantage over SIPmath [40]. The color scale for the heatmaps is based on the color selected (by default dark green in Excel), and each interval is colored according to the binned probability $P(x)$, where $x \in I$. Thus, higher probabilities are represented by darker shades of the base color. Zero is represented by a neutral grey, which lets users distinguish easily between improbable (light-colored) and impossible (gray) values.

Similarly, we color each bar of the detailed distribution plot in the side panel with the same shade of the base color as in the heatmap, and align a copy of the heatmap with the bottom of the bars. This allows users to easily match the heatmap encoding in the spreadsheet with the bar-chart encoding in the detailed view. Mouse-over interactions let users assess exact probability values for each interval. Additionally, we add the mean and standard deviation of the shown samples as numbers and indicate these values by vertical lines in the bar chart. The distribution bar chart in the side panel enables *Expose Uncertainty* tasks (T4), for which the heatmap encoding cannot provide sufficient detail.

In the course of a *what-if analysis*, changes made by the users often affect the probability distributions of cells. To let users assess these changes at a glance, we replace each heatmap with a stacked heatmap (see Figure 3.3(C) or thumbnail at the beginning of this section). The stacked heatmaps allow users to *Compare Probability Distributions* (T6) and *Formulate Cause and Effect* (T5) directly in the spreadsheet. We show the original distribution heatmap in green in the top half of the cell, and the updated distribution in pink in the bottom half. Both heatmaps have the same value range as the original heatmap. Similarly, we change the detailed view in the side panel from a single bar chart to two bar charts that represent the original and the updated distributions (see Figure 3.3(G)). The base hue used in the updated distribution heatmap and bar chart is the same shade of pink that indicates updated values in the active legends for the impact and Bernoulli markers.

In the design process, we also considered using sparklines to visualize the probability distributions, as done in SIPmath [40]. We chose the heatmap encoding instead for two reasons. First, the vertical space in the cells is constrained. As heatmaps encode the values in the color channel, they require little vertical space compared to bar charts or sparklines, which rely on the vertical position for encoding the values. Additionally, during a *what-if analysis* two distributions must often be shown for a cell at the same time. We found that a stacked heatmap encoding uses the space more effectively and with less visual clutter than overlaid sparklines. Second, heatmaps are particularly suited to comparing distributions (T6) across vertically neighboring cells [142].

Change Indicators

▲+21.8 ▼-18

Not all *Formulate Cause and Effect* tasks (T5) necessarily involve tracking the changes of probability distributions in detail. In many cases, it might be sufficient to simply see how the (mean) value of a cell is affected by changes made elsewhere. During a *what-if analysis*, we display small arrows to indicate increases (upward arrow) or decreases (downward arrow) as a result of the changes (see Figure 3.3(C) or thumbnail at the beginning of this section). Additionally, the value difference is specified next to the arrow.

3.5 Implementation

For the implementation, we investigated the most commonly used spreadsheet tools and evaluated them according to their official API extension support and feature list. Microsoft Excel emerged as our tool of choice because it provides the Office JavaScript API. We used the Host Specific Office API model for Excel [143], which allows our extension to be loaded as a so-called *add-in* in both the desktop and the web version. The add-in currently supports Office 365 and the Microsoft Edge Web Browser. We implemented the prototype in TypeScript, HTML, and CSS. The source code is available at `https://github.com/jku-vds-lab/fuzzy-spreadsheet`. A demo version can be accessed at `https://jku-vds-lab.at/fuzzy-spreadsheet/`. Our prototype uses external libraries, the most notable ones for sampling discrete probability distributions [144], jstat [145] for performing advanced statistical operations, maths.js [146] for flexible expression parsing, and D3.js [147] for creating customized visualizations.

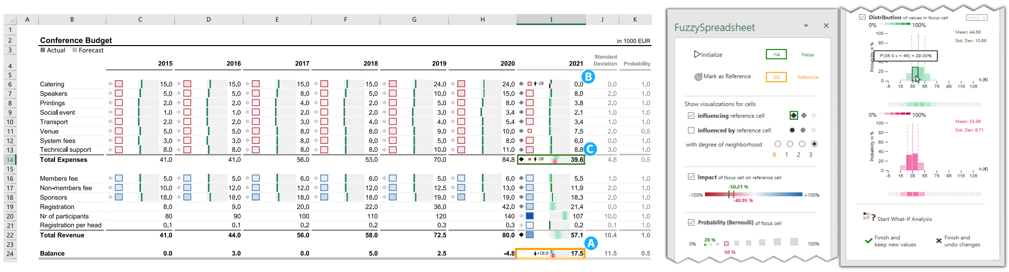


Fig. 3.4: Conference planning usage scenario with Fuzzy Spreadsheet encodings. The user selects a reference cell (A) and controls the visualization using the side panel. In the *what-if analysis* mode, they reduce the value of Catering 2021 to zero (B) and analyze its effect on the focus cell (C). More detailed information about the focus cell is shown in the side panel (right).

3.6 Usage Scenario

To demonstrate the applicability of Fuzzy Spreadsheet, let us assume that conference organizers have been assigned the task of planning the budget for an annual conference. Using a simple model, the organizers predict the revenues, expenses, and the balance for the next year as the average of the previous six conferences (see Figure 3.4). Since planning is prone to uncertainty, decisions on the variation of all uncertain values must be made and conveyed in the spreadsheet. Since the spreadsheet itself does not offer much support for dealing with uncertain data, the organizers decide to use the Fuzzy Spreadsheet extension. Due to its easy modeling requirements, they simply create two additional columns to enter the uncertainty parameters, namely the standard deviation for normally distributed values and the Bernoulli probability to model binary events. The organizers expect a probability of one (100 %) for all uncertain values except for Venue and Catering. These two costs will only occur if the conference takes place on site. Since unexpected worldwide events have recently led to travel restrictions, they estimate the probability for an on-site conference—and therefore for the Venue and Catering costs—as 50 %.

To visually analyze the effect of uncertainty on the balance and the intermediate calculations, they load the Fuzzy Spreadsheet extension. Conference organizers tend to aim for a slightly positive balance, using any surplus for the next conference or better catering. The organizers start the analysis by clicking on the *Initialize* button, which automatically parses the computational graph of the spreadsheet, and mark the summed balance as the reference cell (see Figure 3.4(A)).

The organizers can now decide which visual analysis support they want to activate in the side panel: relationship, neighborhood, impact, Bernoulli probability, and/or distribution. To determine which cells influence the uncertainty of the balance (T4), they first display the influencing cells (T2) and change the DoN to two. The glyphs that appear within the cells indicate the direction of influence by shape and the DoN by color (◆). Subsequently, to analyze the probability ranges of expenses, the organizers activate the distribution visualizations. Related cells then show the heatmaps of the distribution for the expenses. Upon analyzing the distribution, they realize that the value of the balance as given in the spreadsheet is not very likely, and the computed mean—based on correct uncertainty propagation—is negative.

To evaluate how changes in the probabilities for the expenses affect the balance, they enter the *what-if* analysis mode and choose a DoN of three to view the changes in all influencing cells. Their goal is to compare the scenario of an on-site conference with the alternative of a virtual event.

To reflect the expected budget changes in the case of a virtual conference, the organizers start by reducing the expenses of venue and catering to almost zero with a high probability (see Figure 3.4, where the value of Catering is reduced to zero). This new value causes changes in the distributions of Catering, Total Expenses, and Balance. They inspect the new distribution in the lower stack of the heatmap present in each of these cells and keep an eye on the side panel, as shown in Figure 3.4(C). To reflect the increased technical support needed for running the event virtually, they increase the corresponding cost. Since no local advertising is necessary and venue/catering expenses are reduced, conference sponsors will also pay less. Based on this information, the organizers receive a more reliable (i.e., narrower) range for the balance. Additionally, registration costs per person are also to be reduced. After decreasing sponsorship amount and registration cost, the organizers try to increase the number of participants to make sure that the balance lies within a positive range, while being insensitive to minor fluctuations. Once the organizers are satisfied with their adapted planning, they decide to keep the updated values and save the changes in a separate file. The in-cell visualizations are preserved in the new file. They may choose to share this file with their colleagues, who can locally view the encodings even in the absence of the Fuzzy Spreadsheet add-on.

Once familiar with these encodings, such as the relationship encodings for the influencing cells, it becomes easy to grasp the underlying computational graph. The impact encoding helps to identify the extent and nature of the influence a cell has on the reference cell. The changes made during the *what-if analysis* present in the form of change indicators provide an overview of what has changed and by how much. The stacked heatmap encodings in changed cells indicate how a value has become more certain. After sharing the findings, the organizers continue the analysis of the on-site conference scenario in the original file. In this case, the probability for expenses such as Venue and Catering to occur can be increased, and the deviation in the amount of sponsorship can be decreased. This changes the probability distribution of the balance, which in turn makes the balance number more certain.

This usage scenario highlights the benefits of using Fuzzy Spreadsheet: hypothetical scenarios can be analyzed based on immediate visual and numerical feedback. This is especially important in the case of high-stakes investments, where changes are crucial to ensuring a positive balance. The *what-if analysis* helps to determine sensitive values and to obtain a stable range of values for the final outcome without limiting the number of possible scenarios that can be investigated. Thus, Fuzzy Spreadsheet can improve the outcome of the decision-making process.

3.7 Evaluation

To evaluate the usefulness of Fuzzy Spreadsheet, we conducted a user study. The goal of this study was to compare the performance of a standard spreadsheet tool with that of the Fuzzy Spreadsheet approach for the six analysis tasks summarized in Table 3.2. We chose Microsoft Excel as a representative spreadsheet tool, assuming that alternatives such as Google Sheets and Apache OpenOffice Calc would perform similarly. As Fuzzy Spreadsheet is built for casual users and is not meant to compete with specialized tools for expert analysis, such as Oracle Crystal Ball and @Risk, we did not choose them for the user study. Our main objective is to keep the familiar spreadsheet layout and to add uncertainty exploration. Therefore, we also dropped Guesstimate as a comparison tool, as it lacks spreadsheet features and focuses on the uncertainty analysis only. To avoid a learning effect, we used a between-subject design for comparing the two conditions (i.e., Excel with and without our extension). We conducted three pilot studies to test the functionality of the Fuzzy Spreadsheet, to fine-tune the levels of difficulty of the tasks, and to estimate the time needed for each participant. We incorporated the feedback regarding functionality and usability from the pilot into our prototype. To reduce the time per subject, we decided to reduce the number of questions to two per task, with additional two questions for the more high-level task T5. A dataset was chosen that could easily be understood by all participants. The context of the dataset was conference planning, as for the dataset described in Section 3.6. It contained three columns of forecast values with their corresponding Bernoulli probabilities and standard deviation values and fewer parameters than described in Section 3.6. We conducted the main study remotely with 14 participants (P1–P14; Gender: m = 10, f = 4; Age: $M = 28.21$, $SD = 4.32$), seven of whom were assigned to Excel and seven to the Fuzzy Spreadsheet prototype. Two instructors were present throughout—one moderator guiding the participant through the experiment and a second one for taking notes. Note that these studies were performed with some slightly different choices for the visual encodings. The initial design of the change indicators consisted of red and green glyphs that—despite of their different shapes—could be hard to distinguish for people with color vision deficiency, which is why we changed them for the current version. Additionally, we fixed a minor error in the color schemes of the heatmaps.

We used a mixed-method approach to compare how both tools perform at each task in terms of three quantitative measures: (1) the ratio of correct answers in percent; (2) the response time in seconds, and (3) the mental effort required [148], measured on a seven-point Likert scale ranging from one (very low mental effort) to seven (very high mental effort). In addition, we assessed the System Usability Scale (SUS) score based on ten questions developed by Brooke et al. [149, p. 189ff]. Furthermore, we analyzed the thinking-aloud protocol, the feedback from questionnaires, and the follow-up interviews.

To assess the usability of Fuzzy Spreadsheet, we formulated four hypotheses, as listed in Table 3.3. H1 investigated the differences between a traditional spreadsheet in Excel and a Fuzzy Spreadsheet with regard to answer correctness. We expected Fuzzy Spreadsheet—with its in-cell visualizations and the active legend—to show a higher answer correctness than Excel (H1). H2 evaluated the influence of our prototype on the response time. We presumed that, with the additional support provided by Fuzzy Spreadsheet, a shorter response time can be achieved. With H3, we tested whether Fuzzy Spreadsheet or traditional Excel requires a higher mental effort. Given that our approach uniquely combines visual support and calculation of changes (*what-if analyses*), we hypothesized that the mental effort required is lower with Fuzzy Spreadsheet than with Excel. With H4, we examined statistical differences between the two systems regarding the SUS score. Based on these hypotheses, we expected Fuzzy Spreadsheet to achieve a higher SUS score than Excel.

#	Hypothesis	Accept / Reject
H1	Fuzzy Spreadsheet shows a statistically significantly higher answer correctness than the traditional spreadsheet.	✓
H2	Fuzzy Spreadsheet shows a statistically significantly lower response time than the traditional spreadsheet.	✓
H3	Fuzzy Spreadsheet requires a statistically significantly lower mental effort than the traditional spreadsheet.	✓
H4	Fuzzy Spreadsheet shows a statistically significantly higher SUS score than the traditional spreadsheet.	✓

Table 3.3: Overview of the hypotheses tested in our user study, with ✓ and ✗ indicating accepted and rejected hypotheses, respectively.

3.7.1 Procedure

To make the results comparable, we followed a similar procedure for both conditions, as illustrated in Figure 3.5. On enrollment in the study, each participant received an explanation of the study goals and the procedures, and was given access to the dataset loaded into the spreadsheet. This introduction also incorporated the nomenclature and the formulas used. Note that for the user study we referred to the probability parameter of the Bernoulli distribution as "likelihood", because we found that the name "Bernoulli" was not commonly known (despite participants being familiar with the concept of modeling binary scenarios with probabilities). Participants from both conditions received similar introductions, including, for instance, how to perform a second DoN task. The slides we used for onboarding are part of the supplementary material [150]. For the Fuzzy Spreadsheet condition, we provided an additional explanation of the specific visual encodings of Fuzzy Spreadsheet in the side panel and the in-cell visualizations for relationship, impact, Bernoulli probability, and distribution. Further, for the Fuzzy Spreadsheet participants, we installed our extension on the participants' computers and verified its functionality by means of a test file. After successful set-up and introduction, we asked the participants to perform the six analysis tasks. To avoid a selection bias, tasks were assigned in random order. In the course of the study, we recorded (1) externalized knowledge in a thinking-aloud protocol and (2) the mental effort required for each task. After completion of the tasks, we gathered data for the SUS score and demographics in a survey, and performed semi-structured interviews to gain further insights into the usage of uncertain calculations. We closed each session with a debriefing.

3.7.2 Setup

As both conditions were based on Excel, we created one worksheet per question in a single file. For a fair comparison, we developed a Visual Basic for Applications (VBA) script that recorded the answers, checked for correctness, and recorded the time until answer submission. Response-time tracking started with the activation of a sheet and ended when a participant submitted an answer. In addition, the script stored the Likert scale value for the mental effort required. Since this was a remote laboratory experiment, we performed and recorded each session using the video conference platform Zoom [151]. The moderator shared the screen to introduce the prototype features to the participants. During the study, we asked the participants to share their screens so we could observe their actions. If the installation of the extension did not work properly

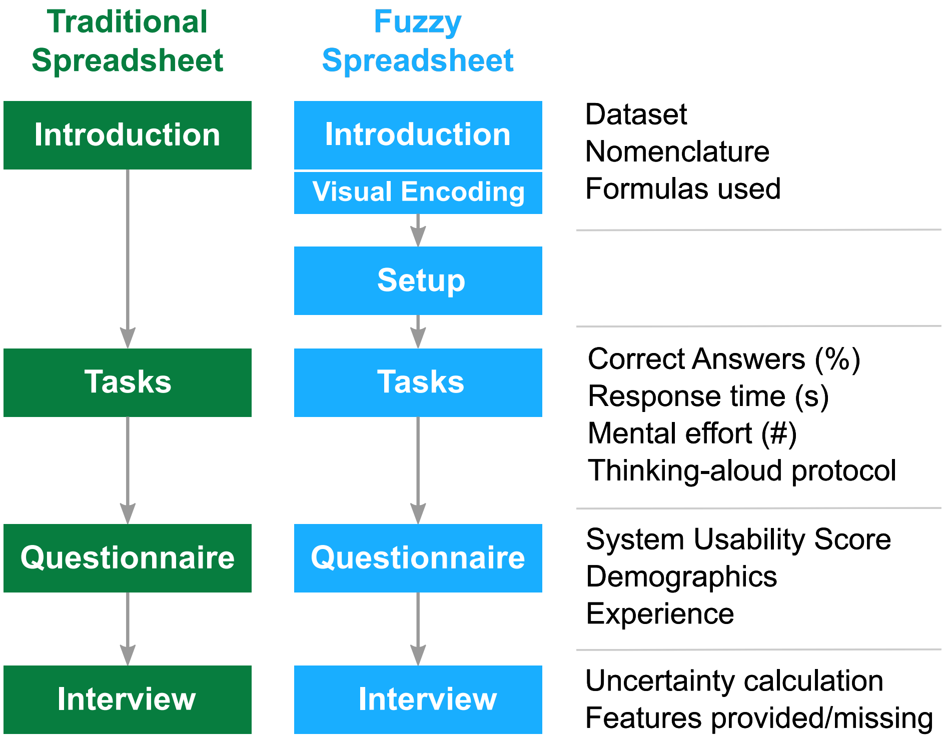


Fig. 3.5: Study procedure for both systems, Excel and Fuzzy Spreadsheet, starting with an introduction followed by task completion, a questionnaire, and an interview.

due to incompatible system requirements, participants were asked to connect to the moderator's computer via TeamViewer [152]. We employed the online and open-source survey tool LimeSurvey [153] for scoring the ten items from the System Usability Scale, for capturing the demographics, and for checking the participants' levels of experience with spreadsheet tools, data visualization, and data science.

3.7.3 Study Results

To assess differences between the two conditions, we used a χ^2 independence test for our dichotomous variable answer correctness. The effect size was ascertained using Cramér's V. Further, to analyze differences in the response time, in the mental effort captured with a seven-point Likert scale and in the SUS scores, we applied a Student's t-test and use eta-squared (η_p^2) for the effect size. The results at task level were examined using a Mann–Whitney U test for answer correctness and a one-way Analysis of Variance (ANOVA) for both response time and mental effort required. Further, we analyzed behavioral observations from the user studies and the thinking-aloud protocol.

Tool Comparison

First, we analyzed the significance of differences between the two tools in terms of answer correctness. We ascertained that answer correctness with Fuzzy Spreadsheet was statistically significantly higher than that of the standard Excel condition ($\chi^2.1/ = 21.448$, $p = .000$, $V = .334$); as a result we can accept H1. Based on our analysis of response times, we can also accept H2. With Fuzzy Spreadsheet, the mean response time was

102.21 s ($SD = 57.90$ s), which is significantly shorter than the response time for the Excel tool ($M = 170.60$ s, $SD = 142.54$ s, $t.184/ = *4.286$, $p = .000$, $\eta_p^2 = .091$).

Subsequently, we took a closer look at the subjective measure of mental effort required and investigated the difference between the two tools. There was a significant difference in mental effort required for the Fuzzy Spreadsheet ($M = 2.94$, $SD = 1.23$) and the traditional spreadsheet ($M = 3.44$, $SD = 1.77$) conditions ($t.188/ = *2.260$, $p = .025$, $\eta_p^2 = .026$), leading us to an acceptance of H3. These results coincide with our observation that Fuzzy Spreadsheet was also used as an affirmative support tool. Participants first sought to calculate the results in their heads before confirming them with the extension (e.g., P2 for T6; P4 for T1). Thus, the mental effort was perceived to be lower with Fuzzy Spreadsheet than with Excel. In general, participants in both groups tried to rely on prior knowledge to answer questions about the financial model without recourse to the extension. Further, participants from the Excel control group made supplementary calculations, either directly in Excel or with external tools, such as calculators or Wolfram Alpha [154]. Participants using Fuzzy Spreadsheet scored statistically significantly higher on the System Usability Scale ($M = 78.57$, $SD = 15.27$) than participants using a standard spreadsheet ($M = 48.57$, $SD = 15.93$, $t.12/ = 3.597$, $p = .004$, Cohens $d = 0.19$). Hence, we can also accept H4. According to the SUS scores, Fuzzy Spreadsheet was classified as a *good* and the traditional Excel as an *awful* tool for the given tasks.

In summary, Fuzzy Spreadsheet had a higher response accuracy (with a large effect size), shorter response time, required less mental effort overall, and was rated with a higher SUS score than Excel.

Task Comparison

We noticed that some tasks were performed more easily with one tool than with the other one. Thus, we decided to assess the applicability of our spreadsheet augmentation in a fine-grained analysis. Based on our quantitative measures, we determined differences between the six tasks.

As indicated in Figure 3.6, Fuzzy Spreadsheet outperformed the traditional spreadsheet in all tasks. Especially the low response accuracy below 50 % for T2 and T4 in the Excel group is noticeable. On closer examination, we found that participants took statistically significantly longer ($M = 351.47$ s, $SD = 196.84$ s) for T3 with the traditional spreadsheet than for the other tasks ($F.25, 13, 597/ = 18, 540$, $p = .000$, $\eta_p^2 = .329$). This is consistent with our observations, the thinking-aloud protocol, and the large effect size. Participants using Excel without Fuzzy Spreadsheet stated that an impact in percent "is difficult to compute, especially in the case of higher DoNs" (P6–P10). We also determined—by means of a post-hoc Tukey HSD test—that *Assess Impact* (T3) was not only the task that showed the longest response time, but also the task which required the highest mental effort for Excel users ($M = 5.46$, $SD = 1.61$). All participants in the control group used auxiliary calculations to answer questions related to this task, while participants from the Fuzzy Spreadsheet condition used the side panel as an additional confirmatory tool. In contrast, Look up Value (T1) tasks were perceived as the least demanding by Excel users.

Overall, Fuzzy Spreadsheet led to higher response accuracies, shorter response times, and required lower mental effort (except for T1).

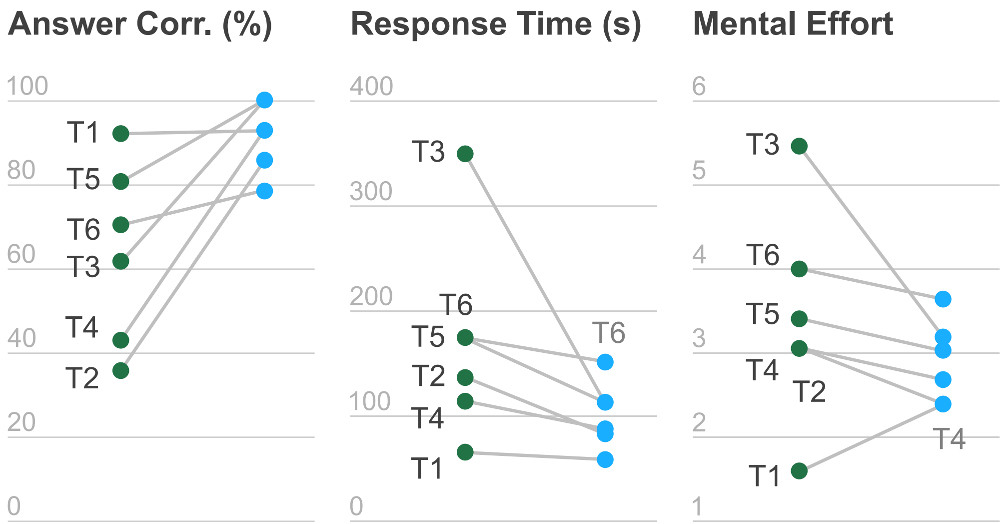


Fig. 3.6: Differences in answer correctness, response time, and mental effort between Excel and Fuzzy Spreadsheet for each analysis task (T1–T6).

Behavioral Patterns

To further describe differences between the Excel and the Fuzzy Spreadsheet condition, we analyzed the thinking-aloud protocols and the interviews. We discuss behavioral patterns that we observed while participants performed the tasks.

Observations of the control group showed behavioral patterns with the traditional spreadsheet, which we had anticipated and addressed with Fuzzy Spreadsheet. For instance, we observed that all participants from the Fuzzy Spreadsheet condition activated the relationship and neighborhood encoding for each task by default. In contrast, participants who used the Excel tool had "difficulties in incorporating the probability and further determining the impact" (translated from German) (T3, but also true for T4 and T6). In particular, P10 mentioned that it was unclear "how one can determine the impact of a neighbor by taking both probability and standard deviation into account". This can be traced back to a generally reduced understanding of probability calculations (P8, P9) [133] or understanding problems of uncertainty propagation (P1, P4, P5, P6, P9, P11) [109]. "If the likelihood rises, the value will be more certain [...] but how about the mean value? If I think about the residual probability being equal" to zero, I assume that the mean will increase (P11 for T6). The encoding selected to indicate *influencing* and *influenced by* cells—glyphs with a gray value for the DoN—was received particularly well and considered to be very helpful. However, some participants (P1, P2, P14) were confused by the placement of the glyphs on the far left edges of the cells. This behaviour was observed especially for T1 and T2. If only the relationship glyphs were displayed and no other visual support was enabled, participants tended to associate the glyphs with the cell to the left rather than the selected cell itself. Interestingly, observations from the Excel control group revealed that two participants (P8 and P11) introduced similar highlighting for related cells themselves. They used different shades for the DoN to facilitate fast detection of coherent cells. P11 from the Excel group even colored reference and selected cells in a consistent color scheme across all tasks, which is reminiscent of our highlighting approach. Participants who did not mark important cells needed longer to complete the task. These observations support our design choice of introducing distinct colors for the different cell types.

We further gained insight into how Excel users performed *Formulate Cause and Effect* tasks (T5). In the course of the analysis, two participants (P8 and P10) realized that they forgot to write down the initial values. "Well, now I should know what was there before I changed a value" (P10). Participants typically took notes next to the financial dashboard to reliably track and compare value changes (P8, P9, P10) and to summarize their insights (P8, P9). This way, they were able to compare an initial value with a newly calculated one. Furthermore, in most cases participants entered initial values in hard-coded form (P6, P7, P8), regardless of the level of familiarity with Excel. This static note-taking approach made it difficult for participants to compare and track changes, particularly for tasks that included secondary calculations or even an increased degree of neighborhood.

3.7.4 Summary of Findings

Fuzzy Spreadsheet makes it easier to understand the information about the underlying relationships between cells and to explore the propagation of uncertainty. Most importantly, the user study indicates that the correctness of answers can be influenced significantly by computational and visual support. Necessary functionalities that we anticipated in our Fuzzy Spreadsheet approach and that are available only to a limited extent in traditional spreadsheets were well received. In particular, the interviewees considered the glyph encodings applied to the impact (P1, P2, P4, P5), directed relationships with neighborhood information (P2, P4), and the *what-if analyses* for assessing variations (P3, P5) as the key features of Fuzzy Spreadsheet. Thus, information is immediately and transparently shown to the user, which otherwise remains hidden. In summary, learning from our results, we recommend holding on to the familiar tools of the spreadsheet and augmenting these established designs judiciously and with great care to support the encodings of visuals.

3.8 Discussion

In this section, we discuss conceptual limitations of the overall approach and technical limitations of the prototype implementation.

Supported Operations For this work, we focused on making the visual encoding and the uncertainty propagation easy to understand. We concentrated on supporting the basic mathematical operations of addition, subtraction, multiplication, and average, and we plan to extend this set of functions in future work. For certain functions, such as division or the square root, a "fuzzy" implementation requires treatment of potentially diverging or undefined results caused by individual samples. More elaborate Excel functions, such as those used in regression modelling, would require completely new implementations to correctly account for uncertainty. Additionally, Fuzzy Spreadsheet currently does not support users in tracing relations (T2) across worksheets (i.e., when a calculation depends on cells from a different tab within the same file).

Uncertainty Authoring As described in Section 3.3.1, the current extension assigns specific meaning to the cells to categorize them as uncertain cells, with some default settings that facilitate the use of normal distributions and Bernoulli distributions. One advantage of this input scheme is that relevant uncertainty parameters are always visible to the users and can easily be edited [103]. Furthermore, fuzzified spreadsheets can easily be shared between users without having to resort to additional configuration files or automatically created helper sheets. However, this simple input scheme restricts the users' freedom during the authoring phase and may necessitate considerable edits when they want to fuzzify existing spreadsheets. We did not address these shortcomings, since we first decided to focus on the analysis phase with given pre-authored spreadsheets. An obvious future improvement of Fuzzy Spreadsheet would be a more advanced authoring scheme in which the uncertainty parameters are directly attached to a single cell and stored in the background by the extension (as it is the case in Oracle Crystal Ball, @Risk, and Guesstimate). In this case, retrieval and editing of parameters could be incorporated into the side panel.

Scalability The fact that Fuzzy Spreadsheet is implemented as a Microsoft Excel extension leads to particular limitations regarding its scalability. The main concern in terms of computational complexity is the number of samples used to estimate the probability distributions. The computation time per fuzzy cell increases linearly from 0.5 ms for 100 samples to 6 ms for 10,000 samples. This means that for large spreadsheets with several thousands of cells, even a rough estimate of the distributions may take several seconds. Another performance issue is related to the way Excel draws the visualizations. If the distribution visualization is switched on for several hundred fuzzy cells at once, Excel can become unresponsive until all visualizations have been drawn. This can take up to several seconds, but can typically be avoided by setting the DoN properly. Both the computational performance and the drawing issues become worse in the online version, where the auto-save feature interferes with the extension. The parsing step is typically of no concern: extracting all fuzzy values and relationships from a spreadsheet with 1400 cells takes approximately half a second using the Excel desktop version. All timing experiments were performed on a standard laptop computer.

Fuzzy Spreadsheet currently provides a responsive and smooth experience for small to medium spreadsheets with up to several hundreds of cells. The usability for larger spreadsheets would not only benefit from computational optimization, but perhaps also from additional visual encodings. Identifying suitable strategies to quickly provide users with a visual overview of uncertainty remains an important challenge for future work.

3.9 Conclusion

Tracing uncertainty in spreadsheets is a challenging task that arises in many application contexts. In this paper, we have presented Fuzzy Spreadsheet, an augmentation approach that adds in-cell visualizations to communicate sensitivity and robustness information while staying as close to the familiar spreadsheet layout as possible. Fuzzy Spreadsheet allows users to track the propagation of uncertainty information through a spreadsheet and to compare alternative scenarios as part of *what-if analyses*.

To evaluate the efficacy of our solution, we performed a small-scale user study that compared the Fuzzy Spreadsheet approach with traditional spreadsheets in terms of answer correctness, response time, mental effort, and usability. In summary, the results indicate that Fuzzy Spreadsheet outperforms traditional spreadsheets and empowers users to carry out tasks related to tracking and exploring uncertain information more effectively.

4 A Process Model for Dashboard Onboarding

Data dashboards with visualizations are routinely used across diverse domains [18], such as health care, education and assembly lines. Dashboards have been defined as visual information displays for monitoring conditions [155, 30], but recent works suggest that they have a much broader scope [18]. We use the term *dashboard* to refer to a combination of multiple data visualizations and textual displays that are often interlinked/ interactive and typically arranged in a single-page layout. Most dashboards are initially created by a dashboard author and may then be used by a diverse group of users, ranging from the general public [156] to analysts [157]. While analysts often have deep domain expertise, they may lack visualization literacy to interpret the data and understand the interactions with and between visualizations in a dashboard. Therefore, *dashboard onboarding* is required to fill the users' knowledge gaps [158, 94, 159]. Typical methods for onboarding users to dashboards include textual descriptions, human narration, or programmed "guided tours". While there is a substantial amount of previous work on dashboard recommendation and the presentation of insights [160, 161], there is a lack of research into formalizing and unifying different onboarding strategies for dashboards. We argue that a process model for dashboard onboarding can be beneficial to understanding existing techniques and facilitate the design and implementation of new onboarding approaches.

Based on the literature [4], our own experience from previous projects related to onboarding [96, 12], and discussions with collaboration partners from the steel and pharmaceutical industries who create and use dashboards, we know that the onboarding process can vary based on the target groups (characterized by differences in dashboard literacy). Some users may require in-depth onboarding to a specific visualization type with an explanation of the data sources and the model, while others require only a high-level understanding of the visualizations and how they are linked. Furthermore, onboarding techniques vary in terms of when and where they are applied. Thus, a process model for dashboard onboarding must be sufficiently flexible to correctly describe a diverse range of onboarding approaches. At the same time, it must be specific enough to help developers and researchers to better understand the building blocks of each approach.

The main contribution of our work is a process model for onboarding users to dashboards that fulfills these criteria. In our model, we introduce the onboarding loop alongside the dashboard usage loop (Figure 4.1). Inspired by the model-view-controller (MVC) software design pattern, we discuss the most important building blocks of the onboarding process. Our model gives rise to an onboarding creation pipeline in which a structured dashboard representation is combined with an onboarding narrative and a set of onboarding means (e.g., human narration, screen recording). The resulting onboarding artifacts, such as oral or video explanations, are presented to the user. We illustrate both

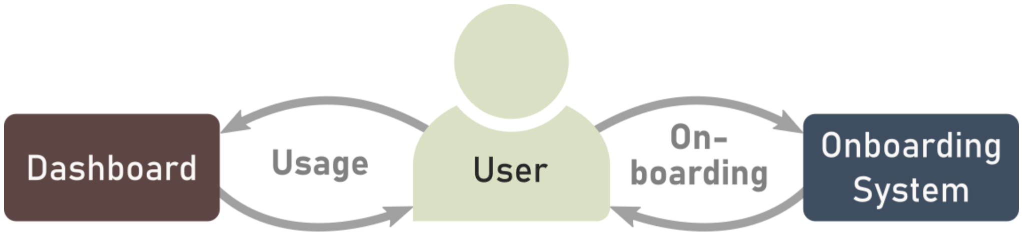


Fig. 4.1: Introducing the dashboard onboarding loop alongside the usage loop.

the generalizability and specificity of our model by showing three different onboarding approaches that can be viewed as manifestations of our model. We also demonstrate an example of onboarding performed using our process model. Additionally, based on our model, we conceptualize a hypothetical interactive and adaptive dashboard onboarding system using an AI model.

4.1 Related Work

The literature on onboarding is broad, covering multiple domains such as gaming [162, 163] and software [164]. While onboarding strategies from these domains could be adapted for dashboard onboarding, we limit our discussion of related work to visualization onboarding. Note, that we focus on onboarding as opposed to guidance. The aim of the guidance is to support users in performing specific tasks with visual analytic tools [94], such as orienting them during the analysis or providing *suggestions* about how to arrive at certain insights. In contrast, onboarding helps users to understand, interpret, and use the applied visual analytics methods, providing *explanations* about key features and capabilities [95, 96].

For onboarding users to a single visualization, Stoiber et al. [96] applied the five Ws and H framework to characterize the visualization onboarding space. We use a similar approach to characterize the onboarding space for dashboards in Section 4.2. In relation to the how, where, and when onboarding is provided, we can further differentiate between *active* and *passive* learning approaches [158]. Passive approaches for visualization onboarding include cheat sheets [101], textual explanations [165, 166], and tutorial videos [166]. *Sticky notes* [167] are textual callouts that are placed directly within dashboards. They may be passive or active, depending on whether they include a call to action. Fully active approaches onboard users while interacting with a visualization, for instance, by enabling two-way communication [168, 42], or through context-aware annotations [160, 58]. Active approaches have been shown to be more effective than passive ones in specific cases [169, 170].

Regardless of the exact means chosen, visualization onboarding is often based on a narrative. Stoiber et al. [12], for example, separated their onboarding instructions for single visualizations into *reading* the chart, *interacting* with the chart, and *using* the chart. In the course of conceptual work for our process model, we realized that an *onboarding narrative* is particularly important for dashboards, as they include multiple visualizations and potentially complex interactions. This relates dashboard onboarding to (visual) storytelling, a common technique for communicating insights into data to users [32]. Segel and Heer [171] differentiated between *author-driven* and *user-driven* stories. In the context of onboarding, we found that author-driven stories created before the onboarding process [10, 75] are used mainly in passive approaches. Active onboarding may also include user-driven elements, for example, when a predefined story is adapted according to the users' questions and needs. This can be handled via various storytelling techniques, such as Martini glass structure [171] and Slide-Show [172]. Also, encoding the interaction history in the visualization guides the users to insight-driven exploration [173].

Several attempts have been made to automate the storytelling process [84, 174, 175]. In Section 4.3.1 we introduce a structured representation of dashboards that may form the basis for automated narrative selection. This representation is closely related to the dashboard meta-models introduced by Vázquez-Ingelmo et al. [176], augmented with a visualization grammar such as Vega-Lite [177]. While the dashboard meta-model formalizes aspects of the dashboard itself, we noticed a lack of work on formalizing the onboarding process. We sought to fill this gap with our process model for dashboard onboarding, for which we drew inspiration from other process models related to visualization [178, 179, 96], and from the MVC software design pattern [180, 181].

4.2 Characterization of the Dashboard Onboarding Space

To characterize the dashboard onboarding space, we adapt and refine the Ws and H questions proposed by Stoiber et al. [96]. Their characterization focuses solely on literacy with respect to a single visualization and does not consider the interplay between the multiple different and linked visual components that form the dashboard. For the purpose of dashboard onboarding, we adjust the Ws and H as follows: (i) WHO is the target user, (ii) WHY is a dashboard onboarding needed, (iii) WHAT needs to be onboarded, (iv) HOW is dashboard onboarding provided, (v) WHERE is dashboard onboarding provided, and (vi) WHEN is dashboard onboarding used. In all of these questions, we refer to both the visual and the functional genre of dashboards, as discussed by Sarikaya et al. [18].

WHO needs to be onboarded is referred to as the *dashboard user* [20] in our process model. Stoiber et al. [96] describe the characteristics of such a user based on their domain, data, visual encoding, interaction and analytical knowledge. While the same user traits are relevant for dashboard onboarding, we do not make them explicit in our process model as we focus on formalizing the overall process of onboarding. We also highlight the role of the *onboarding author*, who takes over the tasks related to the creation of the onboarding [10]. The onboarding author decides the *what*, *how*, *where*, and *when* of the onboarding process based on the overarching question of *why* a user needs to be onboarded.

WHY a dashboard requires onboarding depends on various factors, such as its data, complexity, purpose, and dashboard user characteristics. The *why* is a driving question behind every choice made in the onboarding phase. We consider this question to be asked before deciding on all the other Ws and H. Therefore, we couple it with the following Ws and H questions to highlight its significance.

WHAT can be onboarded in a dashboard includes, but is not limited to, visualizations, global filters, and text elements that provide context [176]. The visualizations may also have specific controls that not only affect themselves but also other visualizations, through filtering or highlighting. These interactions may also have to be explained to users. To specify a dashboard's component in more detail, we further break them down into their low-level characteristics (such as marks, axis, etc.) using the Vega-lite grammar [177]. Although the list of components that can be onboarded is extensive, we use the *what* to refer to components that are actually explained to the user, which is a subset of all the components. The choice of *why* a subset is explained is made by the onboarding author depending on the purpose of onboarding, data complexity, and dashboard user characteristics, among other factors. Our interpretation of the *what* deviates considerably from Stoiber et al.'s definition as they use this W to define the onboarding space itself (i.e., what *is* onboarding?) [96].

HOW a dashboard onboarding is provided to the user refers to the means chosen. We use the same principles as described by Stoiber et al. [96], namely, *type*, *context sensitivity* and *interaction*, to describe how a dashboard onboarding is provided. For example, a human narration with a participatory dialogue serves as a means to an active onboarding approach. However, if the participatory dialogue is removed, the same narration will serve as a means to a passive onboarding approach. Other onboarding means include sticky notes, in-place annotations, narrations, video sequences, among others. Therefore, *why* certain means are chosen depends on the level of user engagement planned during the onboarding. Another reason can be the time and effort required to author these means, as highlighted by the interview summary by Brehmer et al. [9].

WHERE a dashboard onboarding is provided can vary between being integrated directly into a dashboard or being presented in an external source [96]. When embedded directly in the dashboard, it can be displayed as overlay information by means of, for example, sticky notes that explain the visual components [167] or a step-by-step guide positioned in a side panel next to the dashboard and can be opened on demand [182]. In contrast, external sources are accessible to the user at any time and are independent of the direct use of the dashboard. This, for example, includes screen recordings and textual documentation. The interviews by Brehmer et al. [9] indicate that onboarding authors consider aspects like the type and purpose of content consumption to assess *why* a specific placement may be favorable.

WHEN a dashboard onboarding is provided depends on whether the user needs to be onboarded before or during interaction with a dashboard (same as in [96]). Textual onboarding or video tutorials can, for instance, be presented before interacting with a dashboard. In the case of an integrated dashboard onboarding, the onboarding artifacts are shown to the user while interacting with a dashboard. *Why* an onboarding author chooses a synchronous versus an asynchronous strategy may depend on the possibility of in-person meetings with the users or whether onboarding material should be available for later usage [9].

4.3 Process Model

Dashboard onboarding is a multifaceted problem. While several of its aspects have been touched upon in related work, we found a lack of a solid "scaffold" to support, sustain, and consolidate the discussion of dashboard onboarding in the visualization community. We designed our process model to provide such a scaffold. In order to inspire future

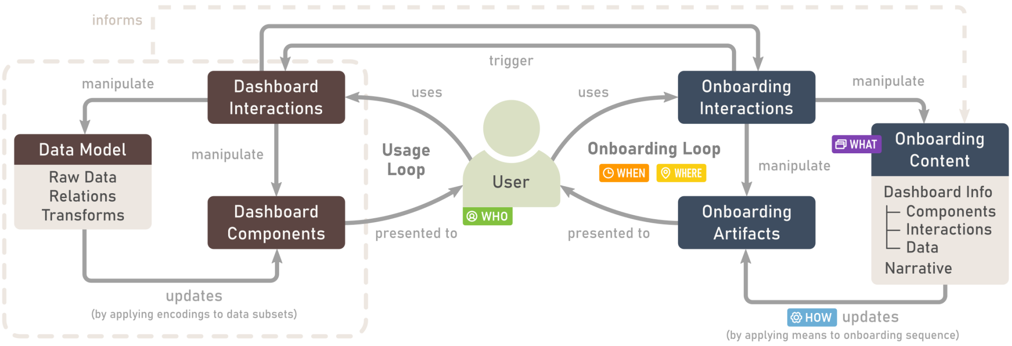


Fig. 4.2: Process model for onboarding users to dashboards. In the usage loop, the user manipulates the underlying data model and/or the visual dashboard components through interactions. In the onboarding loop, the user likewise manipulates onboarding artifacts and/or the underlying onboarding content through interactions. The onboarding content is informed by a structured representation of the inner workings of the dashboard.

work on onboarding authoring, implementation, and evaluation over a wide range of different onboarding strategies, our design process for the model was guided by three requirements. We wanted the model to be:

- *general* enough to cover all existing onboarding strategies without having to introduce new concepts for each particular case;
- *specific* enough to allow detailed description of individual strategies, for instance, by mapping particular elements of existing strategies to general concepts; and
- *actionable*, which means that adhering to it should facilitate the design of new onboarding strategies.

While these requirements may seem obvious, it was not clear how to arrive at a model that would meet all of them simultaneously. We started the design process for the model with an analysis of existing dashboard onboarding strategies. By extracting commonalities and differences of various strategies, and based on related work, we identified fundamental steps and building blocks in the creation and application of the onboarding that we wanted to be reflected in the model. We then iteratively created models and checked whether they met the requirements.

We created several "bottom-up" models based on specific strategies, however, the models generalized poorly. We further created "top-down" models from broad principles but found that in certain cases they require too many modifications, or are too abstract to be actionable. In early versions of our model, we struggled with how to reflect various user roles (see discussion in Section 4.6.2). Sketches for discarded models can be found in the supplementary material [183].

After many iterations, we finally converged on a process model that met all requirements to our satisfaction. This model introduces the *onboarding loop* as an equally important counterpart to the dashboard *usage loop* (see Figure 4.1). We based the onboarding loop's internal structures on the MVC software design pattern [181]. Our choice to model the onboarding loop in a similar way was motivated not only by aesthetic considerations but also by the actionability requirement.

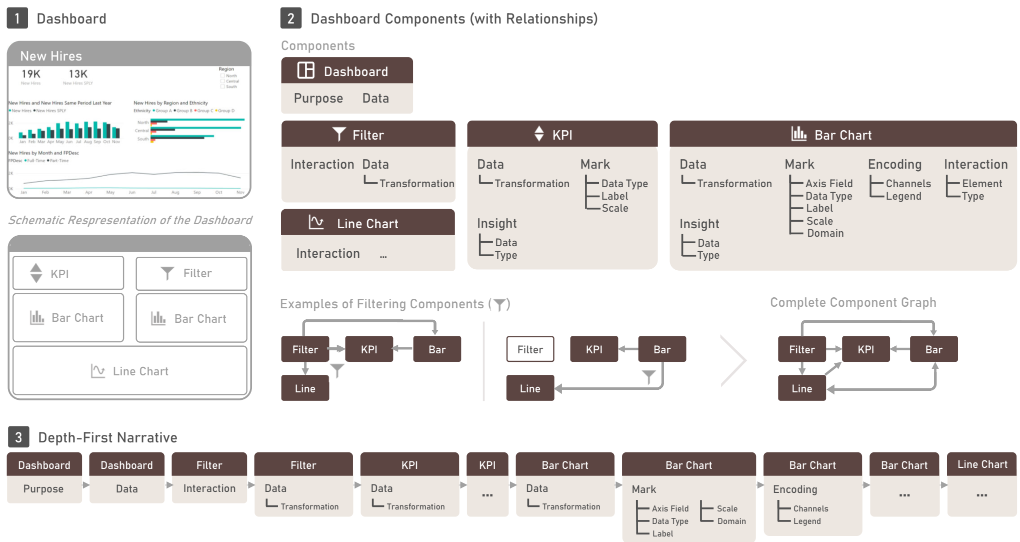


Fig. 4.3: 1 Example dashboard from Microsoft Power BI using six visual components: two KPIs, one filter component, two bar charts, and one line chart. 2 List of all visual components, and simplified representations of interactions through filtering using either the filter or the bar chart component. The complete component graph shows the inter-dependencies between all four components. A fully-fledged component graph with relationships at the lowest component level can be retrieved from the supplementary material [183]. 3 Traversal strategy based on a depth-first narrative to explain all dashboard components one by one.

In the following sections, we first look at the usage loop in more detail and explain how it leads to a structured representation of a dashboard with its components (Section 4.3.1). We then unpack the onboarding loop, discuss its structural similarity to the usage loop, and derive a "recipe" that describes which parts need to be specified when creating an onboarding system/strategy (Section 4.3.2).

4.3.1 Dashboard Usage Loop

Dashboards can be described as a collection of typically linked visualizations of a dataset that are controlled by a user through a set of interactions. To formalize this description, we make use of the MVC software design pattern [181], which allows us to break the dashboard down into three conceptual building blocks (see left part of Figure 4.2): (i) the underlying data model, (ii) the dashboard's visual components, and (iii) the control elements or interactions. Typically, dashboards do not reveal the underlying data model and logic to the end users, but let them manipulate the data model indirectly through interactions. Such an indirect manipulation of the data model frequently results in an update of the visual components presented to the users. Consequently, it is also possible that interactions directly affect the visual components without altering the data model. We explain the roles of the three building blocks in the onboarding loop in more detail, making use of a simplified version of the *New Hires* dashboard provided by Microsoft Power BI [184], shown in Figure 4.3.

The **Data Model** broadly consists of the raw data, relations between data subsets, and the transformations applied to the raw data [185]. The data model outputs transformed subsets of the data that can later be presented by visual components. For the given example dashboard (Figure 4.3), the data consists of the business unit, region, ethnicity, and time.

The **Dashboard Components** are all visual components that are presented to the user. In particular, this includes visualizations such as bar and line charts. Most dashboard components are constructed by applying specific encodings to data subsets which are output by the data model. Details about these components can be listed with the aid of (i) the dashboard meta-model presented by Ingelmo et. al [176] combined with (ii) a visualization grammar, such as Vega-Lite [177]. The meta-model for dashboards first enlists all the components required to create a dashboard. Combining the information with the Vega-Lite grammar breaks down individual components into low-level concepts such as axes, marks, and channels. In our guiding example, the main visual components are two bar charts, a line chart, a filter, and the key performance indicator (KPI) at the top left (see schematic in Figure 4.3 1).

The **Dashboard Interactions** are all interactive components in the dashboard that can be controlled by a user through input devices (e.g., mouse and/or keyboard). In typical dashboards, interactions are closely tied to visual components to create an immersive dashboard experience [186]. This close relationship has led to ambiguities concerning the view and controller parts of the MVC pattern [181]. Some interactions, such as filtering or selecting, manipulate the data model. Other interactions, such as zooming, panning, or highlighting, may directly influence the visual components without altering the data model. For instance, filtering through checkboxes is one of the main interactions technique used in our guiding example.

The three building blocks of the dashboard usage loop lend themselves well to a structured representation of a dashboard that can be later used in the onboarding process. The first step is to create a list with all components. For each component, we note the data subset required for its creation and list all its characteristics (e.g., mark type and channels). Figure 4.3 2 shows such a representation for the bar chart in our guiding example.

In the next step, we list interactions tied to the visual components. For linked and coordinated views, interactions result in edges between components. Thus, we arrive at a *component graph* that reflects the relationships between the dashboard's visual components. In our example, the selection of a bar in the respective chart triggers the filtering of the data used in both the line chart and the KPI. The selection also leads to highlighting in the stacked bar chart. In the component graph, we represent this information as directed edges from the node of the bar chart to the nodes of the other components (see simplified example of filtering components in Figure 4.3 2).

To make the component graph useful for the onboarding process, it needs to be enriched with explanations. The explanations can be descriptions of the purpose of the dashboard, information about specific tasks addressed by the components [187], or exemplary insights that can be gained by using them. Explanations can also include information about the low-level characteristics of a component, for instance, why a certain mark type or color scheme is used. Such explanations can be provided by the dashboard/onboarding author or inferred automatically. In our sample dashboard, one example of insight is that the grouped bar chart shows the highest number of new hires in the region *North* and *Group A*.

Note that the enriched component graph is not always constructed explicitly. In many cases, the onboarding author has only a mental model of the dashboard's components, interactions, and insights that they want to explain to the user. Regardless of whether the structured representation of the dashboard is made explicit as an enriched component graph or whether the onboarding author internalized it as a mental model, this representation forms the basis of the onboarding content in the onboarding loop (dashed arrow in Figure 4.2).

4.3.2 Onboarding Loop

The onboarding loop makes use of the information contained in the structured representation explained above and, unpacking it describes how this information is processed to finally present it to the user. Similar to the dashboard usage loop, the onboarding loop comprises three building blocks: (i) the onboarding content, (ii) the onboarding artifacts, and (iii) the onboarding interactions.

The **Onboarding Content** consists of the structured dashboard representation together with an onboarding narrative. This narrative drives the traversal of the component graph for the purpose of storytelling, and results in a sequence of components and/or interactions, enriched with their explanations (WHAT). The narrative is specified directly by the onboarding author or determined by an algorithm. Examples of simple traversal strategies are depth-first traversal (give all or most details for each component immediately) and layout-based strategies (explaining dashboard components in the same order as they appear in the dashboard, e.g., from left to right or top to bottom). More insight-focused narratives may lead to more complex traversals and usually have to be hand-crafted by the onboarding author. A simple depth-first traversal for our guiding example is shown in Figure 4.3 3.

In the usage loop, carefully selected encodings are applied to data subsets that are output by the data model to construct the visual dashboard components. In a conceptually similar way, carefully selected onboarding *means* (HOW) are applied to the sequences derived from the narrative, which result in **Onboarding Artifacts**. As outlined in Section 4.2, onboarding means can range from (human) narrations to tooltip information and sticky notes [160]. Depending on the means, the artifacts can be visual, auditory, or textual, for instance.

The onboarding artifacts can be either embedded directly within the dashboard or presented to the user elsewhere. WHERE the artifacts are presented depends largely on the means chosen (HOW). For example, in the case of video onboarding, the onboarding interface is a video player and the means of onboarding may be a screen recording of the dashboard along with human narration.

Furthermore, the presenter can decide WHEN to onboard the user: either before or while interacting with the dashboard. In the case of a programmed *guided tour* embedded within the dashboard, the onboarding can be retrieved on demand during the dashboard usage; see example in Section 4.4.3.

The **Onboarding Interactions** provide the user with functionality to control the onboarding process. Depending on the type of onboarding, the interactions manipulate different aspects. In static onboarding, interactions merely affect the onboarding artifacts. For instance, in a video tutorial, the script has already been written and the screen has been recorded prior to the onboarding. Only limited controls are given to the user, such as pausing and continuing a video or going back to a specific timestamp/sequence (see

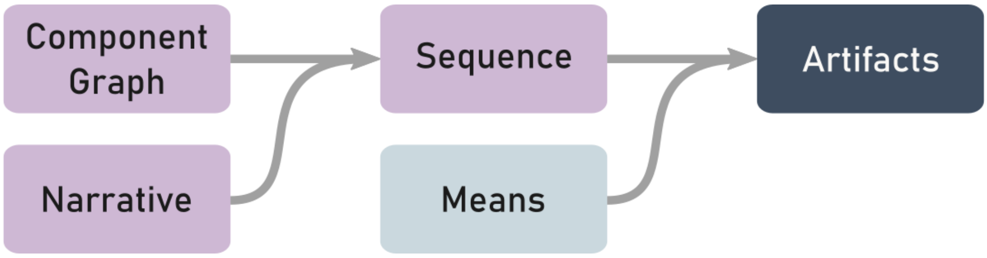


Fig. 4.4: "Recipe" for constructing onboarding artifacts. The component graph is traversed based on a narrative, resulting in an ordered sequence to which onboarding means are applied.

example in Section 4.4.1). In an adaptive onboarding, the interactions trigger updates of the underlying onboarding content. This means that a human presenter may come up with a new onboarding narrative based on a question asked by the user (see example in Section 4.4.4). In the exceptional case in which interface and dashboard interactions are directly connected, a user input in the onboarding interface may trigger a change within the dashboard. Conversely, an active onboarding system may listen to events in the dashboard, which may result in updated onboarding artifacts or an updated narrative.

Each block and the connections between them in the general onboarding process model shown in Figure 4.2 depends on the specific use case.

4.4 Manifestation of the Process Model in Various Onboarding Scenarios

In this section, we show how real-world onboarding scenarios can be described using our process model. We present four different onboarding strategies—a video tutorial, static textual documentation, a programmed guided tour, and an interactive onboarding session with a human narrator—and discuss the role of each part of the process model in each case. In the first three examples, we analyzed existing techniques and mapped them to our process model. In these cases, we also hypothesize that the component graph represents the mental model of the onboarding author. The fourth example (Section 4.4.4) describes an onboarding based on our process model, which we carried out with a collaborator. Finally, we describe a hypothetical AI-powered adaptive onboarding system based on our process model. The detailed onboarding process for these examples can be found in the supplementary material [183].

4.4.1 Onboarding Video

Video tutorials allow the onboarding author to prepare a prescribed story that enables any user to play and recall it sequence by sequence. We screened multiple video tutorials (e.g., on YouTube) that introduce interactive dashboards created for visual analytics tools and chose one that required little domain knowledge. Below we describe a video tutorial using a sales dashboard developed in Microsoft Power BI [188] and show the corresponding process model in Figure 4.5.

Before recording the tutorial, we assume that the onboarding author chose a narrative based on an example usage scenario of the dashboard. To come up with this narrative, the onboarding author assumed the role of the dashboard user and completed several iterations of the dashboard usage loop. This example usage informed a narrative that covered dashboard overview, filter selection, bar chart explanation with drill-down, and KPIs information (WHAT). In the case of the video tutorial, the onboarding author decided to record the screen and use an oral narration as the onboarding means (HOW). The resulting onboarding artifact is the video, which is presented to the user on the video platform YouTube (WHERE).

As shown in Figure 4.5, during the onboarding, the user has the ability to navigate through the video using typical controls such as pausing, fast-forwarding, or jumping to specific sections. While these interactions affect the state of the onboarding artifact that is being shown to the user (i.e., the current video frame), there is no way of triggering an update of the onboarding content. In this sense, onboarding using a typical video tutorial is *static*, which is reflected by the greyed-out edge between the onboarding interactions and the onboarding content in Figure 4.5.

The users can choose whether they want to watch the video prior to using the dashboard or whether they want to try to recreate the presented interactions in the dashboard, for instance, while viewing the video (WHEN) on a second screen. There is no explicit connection between the onboarding and the usage loops, which is indicated by greyed-out edges between the user and the dashboard usage loop (in Figure 4.5).

4.4.2 Textual Explanation

For the second example, we used the regional information dashboard about climate change projections from the IPCC WGI Interactive Atlas [189, 190]. Here, the onboarding author enabled two different ways of user onboarding: textual documentation and a programmed guided tour embedded in the dashboard. This section focuses on the documentation, and the next one (Section 4.4.3) on the guided tour.

The textual documentation includes information about the dashboard components and their corresponding interactions. For each component, the document describes which data subset is visualized, which encodings are applied, and which interactions are available (WHAT). Based on this description, the traversal of the component graph can be considered depth-first.

The onboarding author decided to use a textual format as onboarding means (HOW) and

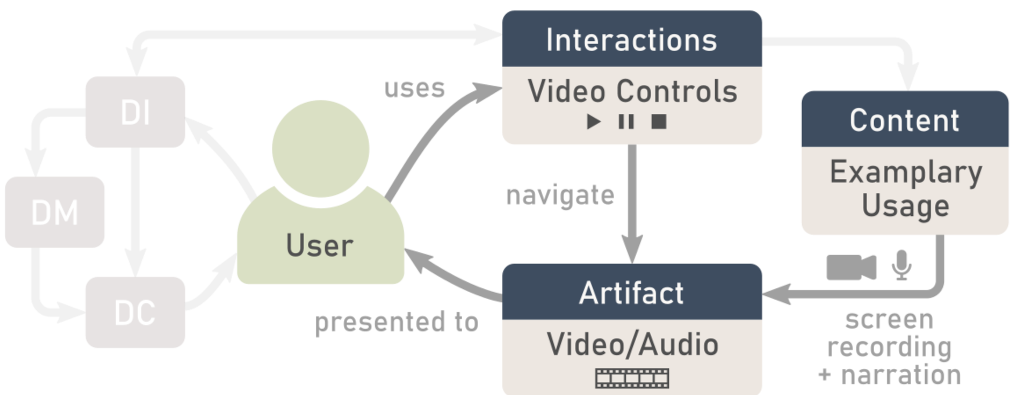


Fig. 4.5: Onboarding with a video tutorial, described using our process model.

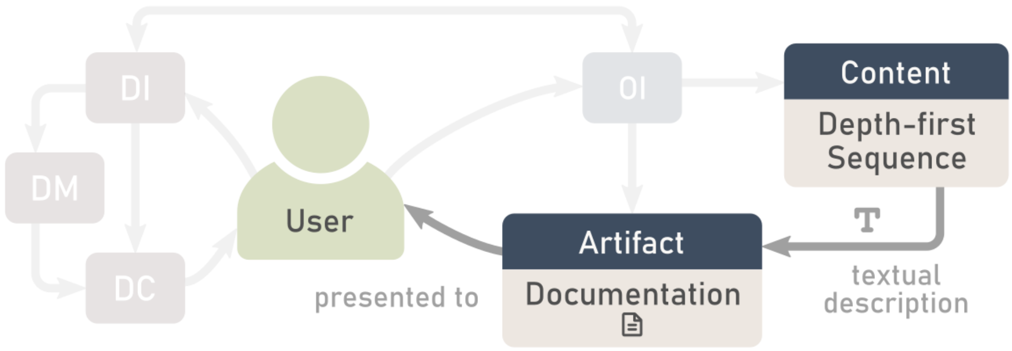


Fig. 4.6: Onboarding with a text document, described using our process model.

the documentation as the resulting onboarding artifact, which can be accessed via the web (WHERE) [189] (see Figure 4.6).

Although the document includes hyperlinks to external resources, we do not consider them to be controls. Instead, we assume that the brief documentation is read from top to bottom and that clicking on external links is equivalent to stopping the onboarding. Without proper onboarding controls, the process model (Figure 4.6) no longer includes an onboarding loop. Instead, the one-time consumption of the external source, in this case a static document, is reflected in the fact that the user has no connection to any other location during onboarding. The document can be accessed by the user at any time before, during, or after the dashboard usage (WHEN). As in the video case, there is no explicit connection between usage and onboarding in this case.

4.4.3 Programmed Guided Tour

As introduced in the previous section, the programmed guided tour example uses the same data as in the textual explanation. Compared to the textual explanation, it provides slightly fewer details about the components for the onboarding.

In this case, we infer that the program follows a breadth-first traversal of each component, as the sequence with its required components (WHAT)is fixed and programmed in a way that cannot be adapted. Onboarding is represented using the means of highlighting and callouts or tooltips (HOW), which are overlaid on the dashboard itself (WHERE). The programmed guided tour is presented as a simple stepper interface when accessing the dashboard for the first time (WHEN). It can be accessed at any time when interacting with the dashboard by clicking on the information icon.

The stepper interface lets the user navigate the sequence of explanations, allowing

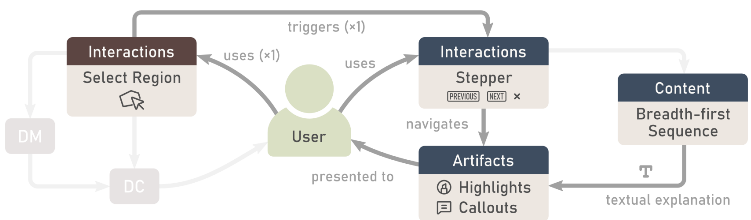


Fig. 4.7: Onboarding with a programmed "guided tour", described using our process model.

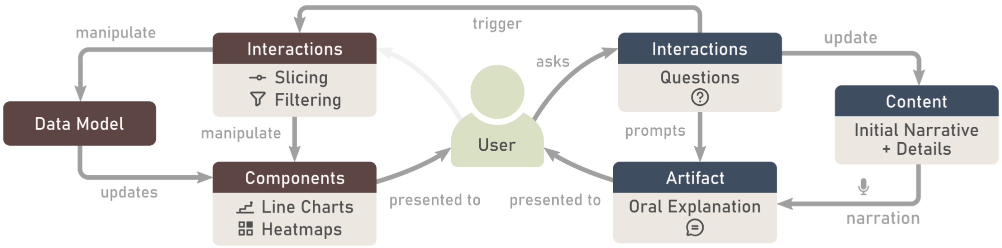


Fig. 4.8: Onboarding in an interactive session with a human presenter, described using our process model.

them to either go back to the previous step, proceed to the next one, or end the onboarding process. This predefined onboarding narrative and the resulting sequence cannot be customized by the user. It is reflected by a grayed-out edge between the onboarding interactions and the content in the process model (Figure 4.7). During the onboarding process, the usage of the dashboard is either blocked or limited and has no effect on the onboarding narrative. For the most part, this leads to disconnected usage and onboarding loops. In only one case does the interface explicitly encourage the user to interact with the dashboard by selecting a region the map to continue the onboarding process. This one-time interaction is indicated in the process model (Figure 4.7) by the edge from user to *select region* in the dashboard. This selection triggers the onboarding *stepper* interface once and the onboarding process continues. The subsequent onboarding steps are again controlled through the overlaid stepper interface.

Upon restarting the onboarding process during dashboard usage, the user has to start the predefined onboarding story again from the first component, regardless of the already introduced components. The system does not save any past interactions with the dashboard or the onboarding, and the traversal order of the component graph is fully fixed.

4.4.4 Interactive Onboarding with Human Presenter

For this example, we authored an onboarding for an interactive dashboard based on our process model. We presented it to a dashboard user in the course of a collaboration with the steel industry.

In the previous examples, onboarding was designed for heterogeneous groups of users. In this case, we present an onboarding that was tailored to specific users who had comprehensive data and domain knowledge. They were not new to interactive dashboards overall, however, a beginner in the field of visualization. For this purpose, we created a list of components that required onboarding (WHAT) from our experience. We then listed two possible predefined sequences for explanation and chose the one which increased the complexity of onboarding gradually. The onboarding took place virtually using screen sharing via an internal communication platform for teleconferences (WHERE). The narrative was provided to the user in the form of oral explanations (HOW). As we made use of visualizations in which the user had little to no prior experience, we provided the onboarding before the first-time use of the dashboard (WHEN). This scenario is reflected in the process model (Figure 4.8) by the lack of a "use" connection from the user to the dashboard, while the "presented to" connection from the dashboard components to the user still exists.

After our explanation, the user posed questions to us, which is represented by the *Interactions* block in Figure 4.8. They also used the shared screen to annotate the visuals for clarifying the explanations provided by us. The user's questions either (i) *prompted* a change in the oral explanation, which led us to rephrase what was said without changing the onboarding strategy; (ii) caused us to *update* the underlying onboarding model, for example, by changing the narrative to explain components of interest; or (iii) *triggered* an interaction in the dashboard, potentially requiring an update to the narrative to reflect the resulting changes.

After one hour of onboarding, the user was done posing questions on various visualization types and interactions. At this point the user had gained an understanding of how to derive actionable insights from the dashboard, indicating a successful session. We discuss the extent to which the process model helped us in the authoring and evaluation of our onboarding in Section 4.5.

4.4.5 AI-based Onboarding

Finally, we applied our process model to a hypothetical onboarding scenario, which tries to overcome the existing flaws of human-narrated onboarding scenarios described in Section 4.4.4. One of its obvious flaws is that, once the onboarding has been completed, it cannot be repeated or accessed on-demand. Further questions can be asked by sending requests to the onboarding author, but the answers may not be provided immediately. Thus, we imagine an AI-based solution to this problem that can be represented by our process model. In contrast to the human narration, such an AI-powered onboarding can be accessed by the user flexibly at any time.

In our proposed automated onboarding solution, the user can initiate the onboarding process by typing in a query such as "Onboard me to the dashboard". This triggers an AI model to derive a component graph from the dashboard with some example insights. With the help of this component graph and a predefined, initial traversal strategy (e.g., depth-first), the AI model can create an onboarding narrative. This narrative may be presented using callouts or tooltips (HOW), overlaid directly onto the relevant dashboard components (WHERE). The user can then choose to follow the predefined narrative, skip steps, or stop it by interacting with the callouts and/or the dashboard. When the user poses a question to the interface, an AI model could respond in ways that resemble a human reaction by (i) rephrasing the current explanation, (ii) changing the narrative, or (iii) initiating changes directly in the dashboard. Possible strategies for updating the narrative could be based on the user's previous interactions with the dashboard and/or the onboarding interface. For parsing the questions and relating them to individual dashboard components, we envision a natural language processing model similar to LILY [191], which was developed for creating dashboards from written prompts.

The process model for such an AI-based onboarding looks like the one for an interactive session with a human presenter (Figure 4.8), except for two additional edges. First, since such an onboarding would take place synchronously with the dashboard usage, the edge from the user to the dashboard is present. Second, as the AI model can update the onboarding based on interactions between the user and the dashboard, the second, left-to-right "trigger" edge between the interactions would be present. As a result, this hypothetical AI-powered onboarding process would make use of all connections in the general model shown in Figure 4.2.

4.5 Using the Process Model

In a data-driven world, the use of dashboards is inevitable. As highlighted by Sarikaya et. al [18] and Tory et. al [20], it is important to support the end users in dashboard literacy. Answering their call to action, we provide practical advice to those who want to support users in understanding dashboards. Accordingly, in this section, we discuss ideas and insights gained from onboarding in Section 4.4.4 by applying our process model for dashboard onboarding in practice. We group these ideas into three phases that reflect the onboarding process: authoring, implementation, and evaluation.

4.5.1 Onboarding Authoring

The authoring phase of a dashboard requires the author to choose subsets of data to be visualized and presented to the user. Similarly, the authoring phase of the onboarding requires the author to select specific visual components and interactions that should be explained to the user.

Previous approaches aimed to assist the author in the dashboard authoring phase by selecting the needed data subsets and/or visualizations [192]. However, there is insufficient literature on how to author the onboarding phase. In this section, we describe how our contribution can potentially facilitate various aspects of the onboarding authoring.

The structured representation, introduced in Section 4.3.1, can be made explicit by the onboarding authors to identify dashboard components, interactions, or pieces of data-related information that may be vital to the end user. For the onboarding case in Section 4.4.4, we used the structured representation to gain the first idea of narration sequences that arose "naturally" from the dashboard structure. It indicated that the time slicer had the most relations to the other components. This helped us in deciding two possible narrative sequences, where the slicer could either be introduced first or towards the end. We decided to go with the latter to gradually increase the complexity of onboarding. Additionally, attaching small explanations to individual nodes in the component graph allows the author to break down a potentially complex onboarding narrative into manageable parts. This may also enable the author to reuse individual explanations, regardless of the specific narrative chosen or onboarding means applied later in the process.

Finally, the onboarding *recipe*, shown in Figure 4.4, describes how a specific set of onboarding artifacts can result from choices made during the authoring phase. Even though the choices of narrative and means may be interdependent and strongly coupled to the overarching question of WHY an onboarding is needed, the recipe reveals that one onboarding strategy may be transformed into another by choosing alternative means. For instance, using the onboarding recipe, it should be straightforward to create a programmed guided tour based on textual documentation.

4.5.2 Onboarding Implementation

The structure we chose for both the usage and the onboarding loops, is closely based on the MVC software design pattern. Thus, our process model can serve as a guide in the implementation of new onboarding systems. For example, it can be used to infer when the dashboard needs to send information to the onboarding interface or vice versa. More importantly, the process model can help onboarding authors identify when an update of the underlying narrative is necessary and when updating only the onboarding artifacts is sufficient (e.g., choosing a different explanation strategy for the same visual component).

Since the structured representation is the foundation of the onboarding loop, we present our thoughts on how to automatically extract this representation. The first step is to extract the components of the dashboard and their relationships. For an embedded Microsoft Power BI dashboard [193], Embedded Report APIs, such as Power BI Client API [194], can be used to this end. We assume that similar workflows are viable for other tools such as Tableau [195]. In the second step, the enrichment of the structured representation with explanations must be formalized. This can easily be specified through an approach similar to *Encodable* [196], a configurable grammar for visualization. Further thoughts on how to make use of Encodable can be found in the supplementary material [183]. After the creation of the fully enriched graph, any graph traversal strategy can be implemented as a default strategy for the onboarding. The resulting sequence can then be shown as a textual description using callouts, overlaid directly on the dashboard for the user. In the case of adaptive onboarding, the onboarding controls can allow the user to skip or access parts of onboarding based on their needs. Note that this is only one way of implementing the process model, and the capabilities of the implemented model can be enhanced in various ways, for instance, by inferring user preferences.

4.5.3 Onboarding Evaluation

Section 4.4 lists several examples of real-world onboarding processes described with our process model. We believe that taking existing onboarding sessions and phrasing them in terms of the process model can help presenters of dashboards to identify pain points, either regarding the dashboard usage or the quality of the onboarding. In case of Section 4.4.4, we recorded the session to analyze when it would have been more appropriate to show something directly in the dashboard or when it would have been more effective to update the narrative based on the user's needs. This helped us to improve our narrative for the next sessions, for instance, by adding the explanation of the line chart again after the interaction with the time slicer.

4.6 Discussion

In this section, we describe the extent to which our process model can be used for informing and evaluating onboarding techniques and how our model can, in turn, be evaluated by such a technique. We also discuss future strategies for evaluating the process model. Finally, we share the lessons learned by highlighting some of these techniques during the creation of our process model.

4.6.1 Evaluation

As described in Section 4.5.1, our model can only inform the authoring and evaluation process to the extent of providing a *scaffold* for onboarding a user. It is up to the onboarding author to make the final choices required for onboarding, weighing additional factors that are beyond the scope of this process model. Some of these factors are the author's presentation, engagement and language skills, the amount of user interaction planned, and the skill set of the user, which all have a high impact on the onboarding process.

Failing to consider these factors and making incorrect choices for the questions raised by our model, can lead to an ineffective onboarding. Therefore, evaluating a specific onboarding—even if it was created based on our model—does not directly evaluate the model but rather the multitude of choices made by the onboarding author along the way.

One way to evaluate the process model is to test it against new onboarding strategies (as done in Section 4.4) to determine where the process model fails to describe them. Although we covered diverse cases in Section 4.4, we acknowledge that the proof of generalizability is limited and can only be solidified by future applications of the model. Therefore, as a next step, we plan to inform and evaluate future onboarding with our collaborators using our process model to establish its validity.

4.6.2 Lessons Learned

We share the lessons learned during the course of our work. We briefly discussed the main reason behind discarding most of the initial versions of our process model in Section 4.3. Here, we describe additional complications arising from the interplay among different user roles as indicated in Figure 4.9. Even without unpacking the usage and the onboarding loops, making all roles explicit resulted in a highly complex structure.

In order to reduce the complexity of the model, we decided to remove the dashboard creation loop as it is always asynchronous with the usage and onboarding loops. After all, the dashboard has already been created at the time of onboarding.

Based on this decision, we reconsidered the onboarding authoring process. Here, we realized that the onboarding authoring could either be done once, like in the case of tutorial videos or textual documentation or repeatedly, like in the case of an interactive onboarding with a human presenter. In the latter, the presenter may need to come up with a new narrative on the fly, temporarily assuming the role of the author. Thus, we faced two challenges. One, we could no longer clearly distinguish between the onboarding author and the onboarding presenter. Two, our model had to allow dynamic updates of the onboarding content. We addressed both of these challenges by making all roles (apart from the user) implicit in the model. Rather than focusing on *who* updates the onboarding content, we focused on the actions that trigger updates.

After multiple iterations of process models, we also realized two additional advantages of this viewpoint. First, it allowed us to draw parallels between the interactively updated views of the dashboard with the dynamically updated content of the onboarding, leading to a symmetric process model. Second, it naturally led to the user being placed at the center of the usage and onboarding process as the sole explicit role.

4.7 Conclusion

Supporting users in their dashboard literacy is a major challenge [18]. We have presented a process model for dashboard onboarding that formalizes and unifies various onboarding strategies. We, therefore, introduced the *usage* loop, where users interact with the dashboard which results in the manipulation of either the underlying data model or the dashboard components. Alongside the dashboard usage loop, we introduced the *onboarding* loop, where the user manipulates onboarding artifacts or the underlying content through interactions. We demonstrate the generalizability of our model by applying it to four real-world examples (video, textual explanation, programmed guided tour, and interactive onboarding with a human presenter) and proposed a hypothetical usage scenario using an AI-based onboarding that sought to solve the flaw in interactive onboarding with a human presenter. Furthermore, we provided actionable advice for developing new dashboard onboarding systems systematically and efficiently. We are confident that our process model will help to create a structured dashboard onboarding and to improve user understanding of dashboards.

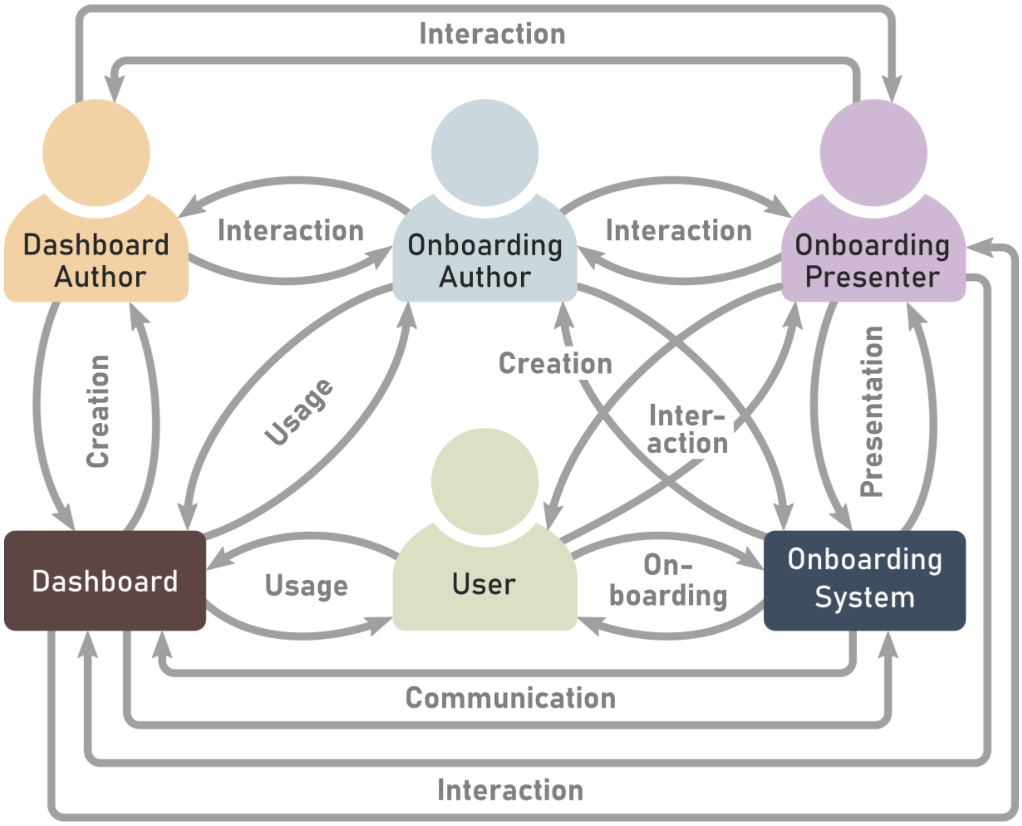


Fig. 4.9: Initial process model showing four roles involved in the dashboard onboarding pipeline.

5 D-Tour: Semi-Automatic Generation of Interactive Guided Tours for Visualization Dashboard Onboarding

Visualization dashboards—collections of charts, graphs, and other visual elements that provide users with a comprehensive overview of information—have become one of the most popular forms of visualization used in business [17, 18]. Introducing an end user to a dashboard they have not seen before is known as *onboarding* [6, 96], during which the dashboard and its visualizations, purpose, and data are presented. Proper onboarding can enhance both use and adoption of dashboards by giving an overview of the data, visuals, and interactions, and filling the users' knowledge gaps [8, 94, 159]. However, not all onboarding experiences are created equal—short of a personalized meeting between end user and the dashboard author, the most effective onboarding method is a guided tour of the dashboard components [197]. However, crafting such guided tours is labour-intensive and time-consuming because of a lack of tooling and standardization. Furthermore, once created, the tour remains static and cannot easily be adapted to a new end user with different skills and expertise, let alone to a new dashboard.

Here, we propose *interactive dashboard tours* (D-Tours) as an effective approach to designing dashboard *onboarding experiences* that preserve user agency while employing presentation techniques drawn from data-driven storytelling and open-world video games. A D-Tour is based on a sequence of dashboard components that are extracted semi-automatically from a dashboard. Similar to open-world games such as *Elden Ring* (2022) and *Hogwarts Academy* (2023), the presentation sequence is not linear but maintains the user's freedom in navigating the tour while keeping vital knowledge dependencies between components. In practice, this means that the user maintains agency in choosing when to view components, within author-defined constraints on the content. For example, certain Key Performance Indicators (KPIs) must be visited first or multiple components must all be visited.[1] Another benefit of semi-automatically generated D-Tours is that the narrative can be dynamically adapted to the user's domain and visualization expertise.

To demonstrate the applicability of our concept, we implemented D-TOUR PROTOTYPE, an in-situ toolkit for authoring, disseminating, and viewing D-Tours within the Microsoft Power BI [198] software suite. Using D-Tour Prototype, a dashboard author can generate a D-Tour in three steps: (i) extract the content (semi)-automatically from the dashboard; (ii) arrange the content; and (iii) share the D-Tour with the end users. End users navigate through the D-Tour while interacting with the dashboard; tailoring the content as they go, with the option of switching to a free exploration at any time.

We make the following contributions: (*i*) a concept for crafting and using (semi-)automated interactive dashboard tours (D-Tours) inspired by ideas from open-world

[1] Compare this to an open-world game where the player can pick and choose from various encounters across a map, which makes side quests and other mandatory main plotline encounters.

games; (*ii*) a prototype implementation of the concept in a web application that augments an embedded Microsoft Power BI dashboard; and (*iii*) results of evaluating the D-Tour Prototype in usage scenarios. Two qualitative user studies were done (one with five dashboard authors and one with six end users).

5.1 Related Work

We discuss existing research on onboarding for both single visualizations and dashboards. We also explore how data storytelling techniques can be used to create clear and engaging onboarding experiences. We finally examine authoring tools in data storytelling, as their design significantly informed our approach to author D-Tours.

5.1.1 Visualization and Dashboard Onboarding

Onboarding solutions, although common in user applications and video games, are less prevalent for visualization dashboards. Stoiber et al. [96] characterized the onboarding space for *single* visualizations, listing online guides [100] and cheat sheets [101] as well as more recent approaches, such as step-by-step guides and scrollytelling [102].

The need for dedicated *dashboard* onboarding was emphasized by Walchshofer et al. [8], Brehmer et al. [9], Tory et al. [20] and Sarikaya et al. [18]. While the need for onboarding might in theory be reduced by adhering to professional design guidelines for dashboards [30, 28], our observations from both the literature and our industrial collaborations highlight a significant demand for effective dashboard onboarding strategies in real-world scenarios. Tory et al. [20] mention the use of simplicity and training to onboard dashboard users with low data literacy. Previous work on dashboard onboarding includes annotating dashboards [160] and assisting users in learning public-access interactive tools [167]. Recent work by Chundury et al. [197] added data-driven, contextual, in-situ help features for visual data interfaces.

Help systems [199] and guidance [94] serve different purposes than onboarding; in this paper, we focus on the latter. While approaches to interactive onboarding for a single chart exist, most approaches in the literature are static, such as tooltips and annotations [200]. This is surprising given the interactive nature of dashboards.

In real-life user-onboarding scenarios, presentations [8] and static documentation of interactive dashboards are utilized for onboarding users. Our D-Tour concept allows an author to create and present onboarding experiences that are integrated into the dashboard to enhance user engagement and understanding.

5.1.2 Data-driven Storytelling

Visualization dashboards, much like their constituent visualizations, often convey data facts through storytelling. We propose extending storytelling concepts to dashboard onboarding to enhance user engagement [201, 202, 203, 204]. Segel and Heer's narrative visualization framework [171] identified seven narrative genres in storytelling that we adapted to create various onboarding styles. Zhao and Elmqvist [205] extended this narrative framework to include additional media types.

Our work drew inspiration not only from data-driven storytelling but also from other narrative-rich domains, such as movies and open-world video games [206]. Recent research into visualizing non-linear storytelling in movies has identified various categories of relationships between story order and narrative order [207, 208]. While dashboards lack the inherent temporal order of movies, non-linear narratives from open-world video games offer valuable insights [209]. In fact, such non-linear narratives have been found to increase user engagement [75] and have also been studied in the context of multi-user analyses [90]. We adopt various narrative styles, such as branching, parallel, and free-form narration. Our goal is to provide authors with control over the narrative structure of the interactive dashboard tour and users with a flexible level of agency.

Finally, we also adopted assistive methods for narrative and data-driven story creation. Chen [210] surveyed authoring tools in data-driven storytelling, focusing on automation in narrative visualizations. While semi-automation is an aspect of our work, we prioritized keeping the author involved in the creation of the onboarding experience. For clarity, we also made use of visualization sequencing from the work of Hullman et al. [211] and Kim et al. [212], to ensure an effective delivery of the intended message during the onboarding process.

5.1.3 Authoring Tools for Storytelling

Our review summarizes authoring tools that support narrative creation across various domains. Green et al. [213] introduced a design pipeline that focuses on user experience for interactive narrative authoring tools. Meixner et al.[214] presented an authoring tool for non-linear videos utilizing a tree structure, with annotations attached to scenes. Novella [215] facilitates interactive story creation in games. Metamorphers [216] provide storytelling templates which can be used to generate animated transitions for multiple data sets in molecular visualizations.

Several authoring tools have been proposed specifically for data-driven storytelling. ScrollyVis [204] is an interactive authoring tool for guided dynamic narrations that uses storytelling and integrates diverse resources, such as images, text, videos, and maps. Molecumentary [201] enables the creation of narrated documentaries about molecules and supports various media types, such as text, audio, and video. Roslingifier [202] offers a semi-automated approach to constructing data presentations using animated scatterplots, and ChartStory [217] provides a unique comic-style data narrative crafting method. Finally, InsideInsights [218] allow authors to organize facts into a hierarchy that can be dynamically navigated by the viewer. We implemented the interactive dashboard tours within the D-Tour Prototype by building upon insights gained from these authoring tools. AutoClips [175] offers a fully automated video generation approach to storytelling from data facts. Notable solutions also include Narvis [172] for narrative visualization, Temporal Summary Figures [179] for annotated temporal visualizations, and Erato [203] for data fact sheets. Unlike these works, our approach focuses on creating non-linear narratives derived from dashboard content to onboard new users effectively.

5.1.4 Tours in HCI and Visualization

Interface tours have long been a popular approach to onboarding users to an interface or tool in HCI practice, especially for the web [219, 220]. This practice has also been applied to both commercial and academic visualization systems [221]. Tours provide guided walkthroughs that allow users to explore an interface systematically, thereby reducing cognitive load. Commercial tools such as Tango [222] and Scribe [223] allow for the creation of annotations and guided tours for websites, which can also be used for dashboards. Their approach is mainly static and based on screenshots. Chundury et al. [197] present guided tours as one of several help mechanisms. Elmqvist et al. [224] study guided 3D tours for introducing users to information-rich 3D visualization environments. In general, allowing users to retain agency over navigation facilitates memory and recall. Similarly, the D-Tour concept proposed in this paper seeks to strike a balance between a fixed presentation sequence and user control to improve the onboarding experience.

5.2 Design Rationale

Recent studies on the use of visualizations [9, 18] and dashboards [8] in larger organizations have shown that onboarding is currently done primarily through oral presentations at the time of a visualization's or dashboard's launch. This process requires the user to go through a prepared, narrated script. There is little or no agency involved on the user's behalf. Additionally, there is a considerable time and effort involved in creating elaborate onboarding material, such as a video or a guided tour [9]. This means that most organizations opt for a single onboarding experience that must work for all users, regardless of their expertise. In most cases, documentation in the form of slides or a standard document may be provided as a supplementary aid to the dashboard to avoid repetitive onboarding scenarios. The apparent lack of interactivity and user agency in such fixed, "one-size-fits-all" onboarding can be particularly problematic for two different user groups:

- **Novices:** Low visualization literacy [225] and lack of experience with a dashboard tool and workplace practices means that novice users easily become *confused* or even *overwhelmed* by onboarding material that does not match their level of knowledge.
- **Experts:** High visualization literacy and extensive experience with similar dashboards means that an expert may become *disengaged* or even *bored* by material that does not recognize their level of expertise.

Design Sources. We derive the following *sources* for the design:

S1 *Dashboard Onboarding Space*—In prior work [6], we characterized the onboarding space and addressed *what* dashboard components must be explained while onboarding a user (*who*). We also reflected on other questions, such as when, where, why, and how to onboard.

S2 *Industrial Collaborators*—Our work is inspired by a long-term, ongoing collaboration with a large industrial manufacturing company. We found that dashboard onboarding can help users transition to a new visual analytics tool. Customized onboarding experiences can tackle specific challenges, such as end user fear of interacting with the dashboards and making sense of myriad charts and complex data [8].

S3 *Existing Storytelling Tools*—We also draw on prior findings from authoring tools based on data-driven storytelling (Section 5.1.3).

Design Goals. Because author and user needs and interactions with a dashboard and corresponding onboarding are inherently different, we list their design goals separately (following from **S1**, **S2**, and **S3**).

G1 Access Dashboard (Authors + Users)

To create a dashboard onboarding experience, the author must have access to the dashboard's visualizations and their relationships. This requires understanding how interactions with one visualization affect others, a key factor in determining *what* needs to be part of the onboarding [6] (**S1**).

It is equally important, although not at the same level as for the authors, that the onboarding *users* have access to the dashboard. With this, they can easily relate onboarding material to the visualizations in the dashboard.

G2 Retain Agency (Authors + Users)

The onboarding material is the subset of a dashboard's content functionality that must be explained to the user. The author initially selects the material and then decides how much agency the user should have. The onboarding tool should support the author by equipping them with ways of selecting this subset and providing or modifying the information shown during the onboarding.

For users, the tool should offer flexibility in selecting the level of detail at which they wish to engage with the material (**S2**). They should be able to (i) choose which aspects of the dashboard to explore in more detail, (ii) customize the amount of information they receive, and (iii) potentially alter the path they take through the onboarding based on their interactions.

G3 Craft Interactive Tours (Authors)

Tours are common onboarding mechanisms [197], and interactive ones preserve agency and engage the user. Employing data-driven storytelling (**S3**) in our context suggests using multiple tour structures [208] to enhance onboarding comprehension and potentially cover more than one way of using the dashboard. The responsibility of crafting these tours lies with the author, who must understand *why* the onboarding is needed and *how* it can best be explained (**S1**). The tool should provide ways of creating these tour structures (**S2**).

G4 Choose Presentation Mechanism (Authors + Users)

Onboarding can be delivered in several ways, ranging from interactive guides to annotated walkthroughs, video tutorials, and more dynamic exploratory modes. The tool should allow the author to choose a delivery mechanism (**S1**, **S2**).

The users should also be able to choose the presentation mechanism configured and provided by the authors.

G5 Explore the Tour (Authors + Users)

The author can use the tour to evaluate its effectiveness in conveying the intended message and make necessary modifications.

For the users, navigating through this tour is essential to understand the dashboard. Based on the tour structure produced, the tool should offer the user varying levels of autonomy (S3).

5.3 Interactive Dashboard Tours (D-Tours)

We propose an approach to semi-automatically generate *interactive dashboard tours* (*D-Tours*). This helps onboarding authors to create engaging onboarding experiences that can be navigated dynamically by the user (Figure 3.4). The design of D-Tours is inspired by concepts from storytelling, interactive narratives, and open-world games (Figure 5.1). Here, we explain how the concept can be applied to support both authors and users. We also describe how it addresses our design rationale.

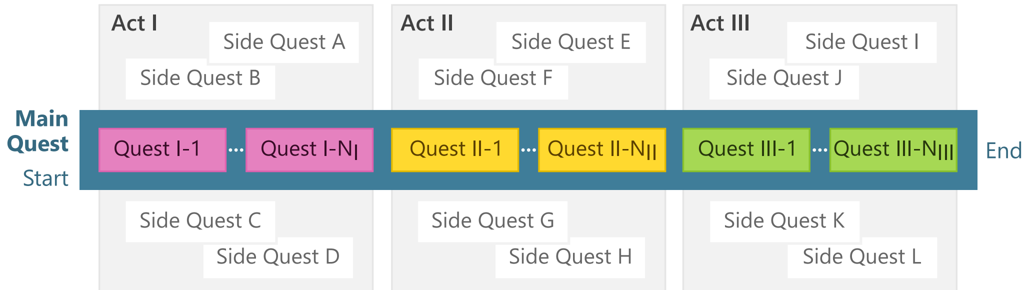


Fig. 5.1: Open-world video-game narrative graph. Video games based on an open world rather than a linear narrative grant the user maximum freedom in picking their actions. The main quest steps help players to progress towards completing the game. In contrast, side quests support but do not advance the story.

5.3.1 Design: Interactive Dashboard Tours

We define an *interface tour* as an annotated linear traversal of the components in a user interface intended to onboard a user [197, 219, 220]. In the context of a visualization dashboard, we refer more specifically to a *dashboard tour*. Most interface tours are fixed sequences—sometimes called *wizards*—where the user can move only forward and backwards. An *interactive tour*, in contrast, is a dashboard tour that preserves user agency by allowing them to navigate within components of the tour to a lesser or greater extent. The linear sequence thus becomes a narrative graph. Accordingly, interactive tours draw inspiration from dynamic storytelling and open-world non-linear narratives (Figure 5.1).

An interactive guided tour is a directed graph of story elements, where each element is either a visualization or a group element:

- **Optional:** The user can choose whether or not to visit the child elements in the container before proceeding. *Example:* a set of detailed charts not essential to understand a dashboard.

- **Visit at least one:** The user must visit at least one of the child elements before proceeding. *Example:* several line-series charts in a dashboard; understanding one is enough.

- **Visit one:** The user must visit exactly one of the child elements before proceeding. *Example:* visiting one of several related KPIs.

- **Visit all:** The user must visit all child elements before proceeding. The visiting order can be flexible or fixed. *Example:* key charts in a visualization dashboard that all must be explained.

A tour is a **Visit all** group consisting of one or more story elements. The approach can be used to emulate a fixed linear sequence.

5.3.2 D-Tour for Authors

In many organizations, creating a new dashboard typically means that the author must also create a dashboard onboarding experience for their end users. Whether manual or automated, this process can be structured into three steps: *extraction of the content* (Figure 3.4(a)) required for onboarding, *arranging the content* (Figure 3.4(b)) into an understandable sequence, and previewing the content to ensure that it aligns with the intended mental model before *sharing* (Figure 3.4(c)) it with the end user. We detail each of these steps below.

Content Extraction (G1, G4): The first step in authoring an onboarding experience is to list the content to be explained. For a dashboard, this includes identifying purpose, data, visualizations, and their interactions. We structure this content into a *component graph*, as outlined in our previous work [6]. Nodes represent the visual components (including data and insights) and edges represent their interactions. An edge from a filter node to a bar chart node indicates that the filter interacts with the bar chart through cross-filtering. To make the graph useful for onboarding, each component must be enriched with explanations, such as its description (type, mark, and encoding of visual components) and the insights it provides [102]. The edges representing the interactions must also be explained. We indicate all these types of explanations in Figure 3.4(a). Automating the content extraction and graph creation requires access to the components' low-level characteristics and the dashboard's data and interactions. In the next section, we describe how we accomplish this in the D-Tour Prototype. Creating an explicit component graph helps to automate and manage components and their relationships, especially when dashboard interactions occur and an update is required. In simple cases, the author can also use the component graph as a mental model.

Content Arrangement (G2, G3): The next step for the author is to arrange the content in a sequence, creating a path through the onboarding tour with directed edges (Figure 3.4(b)). If the content is structured as a component graph, the author can either employ traversal algorithms or manually create paths to arrange the components. In the latter case, the underlying component graph still remains valid and assists in managing the relationships between the components, while new paths determine an explicit order of the tour. Drawing on concepts from storytelling and open-world games, the content can be arranged in various ways, ranging from linear sequences to branched narratives and open-ended explorations. Such flexibility enables the author to tailor multiple tours to various types of users. This increases user agency by allowing them to choose the path that works best for them. The adaptability is crucial when introducing a single dashboard to a diverse audience with varying levels of domain and visualization expertise, such as managers, engineers, and sales personnel.

Dissemination (G4, G5): As with any created content, previewing the final version before delivery to the end user is crucial for testing (Figure 3.4(c)). An automated approach can show the tour to the author at any stage of its development. This can support the author in testing their onboarding tour throughout the creation process, allowing iterative improvements and refinements. The manual approach, however, might lack these iterative improvements and might involve only a cognitive walkthrough of the prepared onboarding tour.

5.3.3 D-Tour for End-Users

Our previous work [6] described different scenarios in which a user can be onboarded to a dashboard, including in-person meetings, textual documentation or video tutorials, guided tours, and an AI-based chat assistant. In each scenario, we highlighted user agency in relation to the onboarding content and how adaptable the onboarding could be. To maximize user agency and adaptivity, an onboarding experience should allow users to *choose an onboarding style* (predefined tour or self-guided exploration), *tailor the content* based on the style, and enable *interacting with the dashboard* and adapt accordingly.

Choose an onboarding style (G2, G4, G5): A user should have agency in choosing the style of their onboarding experience. For instance, novice users new to the domain of the dashboard and its visualizations might find it useful to go through an onboarding experience prepared by the author to help them understand the concepts. Experts, in contrast, might find it easier to explore the onboarding themselves, essentially creating their own onboarding experience (self-guided exploration). Additionally, if the author has configured multiple presentation mechanisms beyond text, users should be able to choose their preferred method of presentation. We explain how we incorporate these guided and self-guided modes in our implementation (Section 5.4).

Tailoring the content (G2): In addition to choosing the onboarding style and thus a path through the tour, the user should also be able to tailor the content to match their needs. For instance, a user might request more information on a specific visualization in a dashboard and less on others. Such dynamically adjusting content for each visualization level can be beneficial to all types of users.

Interacting with the dashboard (G1): Providing in-situ onboarding directly on the dashboard can be helpful, especially to users who prefer learning by doing or are afraid of "breaking the data" [8]. An onboarding tour should use the inherent interactive nature of dashboards to create an interactive onboarding tour for the end users.

5.4 The D-Tour Prototype Application

We propose the D-Tour Prototype as an implementation of our interactive dashboard tours. Similar to the description of D-Tours for authors and users, the D-Tour Prototype also has two modes: the *authoring* mode for onboarding authors and the *onboarding* mode for its users. We describe both modes in detail using a guiding example (Figure 5.2).

5.4.1 Authoring Mode

The authoring mode lets authors craft and preview interactive dashboard tours using extracted dashboard components. Based on the concept in Section 5.3, we divided the mode into *Content Extraction View*, *Content Arrangement View*, and *Dissemination View* (Figure 5.2). We describe the views in the following based on an exemplary use case which is described in more detail in Section 5.6. In the use case, an onboarding tour is created for a dashboard (Figure 5.2) that shows a company's market share and consists of two KPIs, a filter, a line chart, two-column charts, one combo chart (combination of line and column chart), and a table. This dashboard is available in Microsoft Power BI [198]. The D-Tour Prototype embeds the dashboard in a custom web application using Power BI REST API [226].

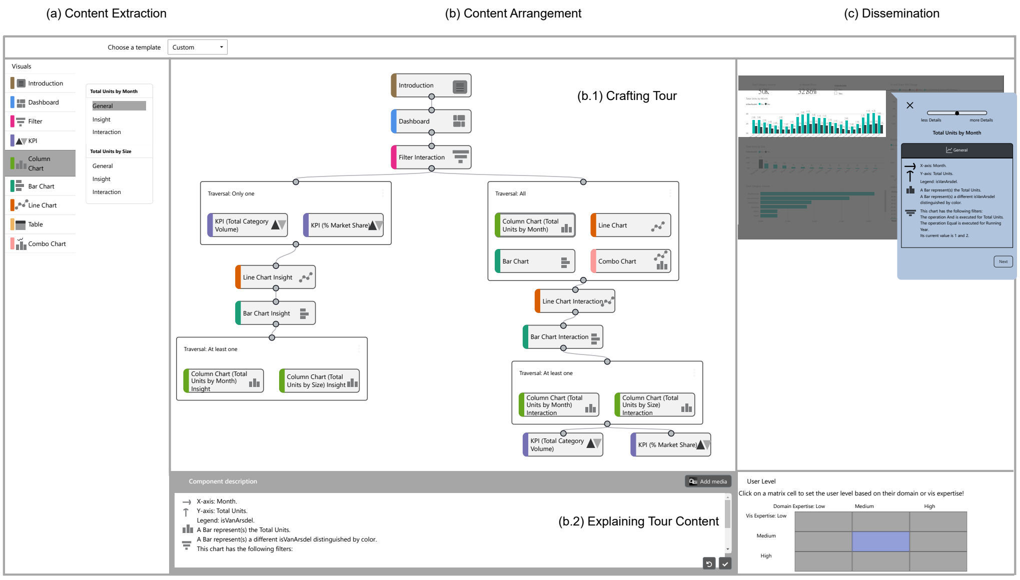


Fig. 5.2: D-Tour Prototype's Authoring Mode. Authors pick (a) automatically extracted visualization categories, *General*, *Insight*, or *Interaction*, from the Content Extraction View and drag them to the Content Arrangement View, where they (b) arrange them, (b.1) thus crafting a tour and (b.2) adding explanations to the tour content. In the Dissemination View they (c) test changes before disseminating them. A selection of the Column Chart *General* in the Content Extraction View is shown which is highlighted in the Content Arrangement View and in the Dissemination View. Its associated content can be seen in (b.2)

Content Extraction View

The Content Extraction View is an author's entry point into creating an onboarding experience. The dashboard content is presented in a simplified manner to the author, with the aim of providing an overview of the dashboard's purpose and a brief introduction to the visualizations and data. This is achieved through the use of two categories: *Introduction* and *Dashboard*. The Content Extraction View mirrors a real-world onboarding scenario that starts with the dashboard's goal and data.

The rest of the dashboard content, which is extracted and structured as a component graph in the background, is categorized first by type and then by the subcategory of the onboarding stage—reading, interacting with, and using a visualization—as proposed by Stoiber et al. [96]. We call these subcategories *General*, *Interaction*, and *Insight*, respectively.

Figure 5.2(a) shows the extracted visualizations and their subcategories. For example, the highlighted category of Column Chart displays two column charts present in the dashboard that are identified by their titles and listed with corresponding subcategories. These subcategories are simplified representations of the low-level characteristics of a visual component in the component graph (Section 5.3).

The component graph is derived using (i) the meta-model by Ingelmo et al. [176], (ii) the Vega-Lite visualization grammar [177], and (iii) Krist's visualization component grammar [196]. It is then populated with information extracted from the Power BI REST APIs [226] and the data visualization catalogue [100], combined with text templates to ensure meaningful sentences. The graph is created automatically and updated whenever the dashboard changes, which in turn updates the *General*, *Interaction*, and *Insight* subcategories.

For crafting an interactive dashboard tour, the subcategories can be *picked* by dragging and dropping them onto the Content Arrangement View. Additionally, selecting a subcategory triggers two simultaneous actions: (*i*) highlighting the corresponding visualization in the Dissemination View and (*ii*) displaying the default description for the subcategory in the Content Arrangement View. This dual-display functionality helps authors to easily correlate all types of information about a visualization with its location in the dashboard. Figure 5.2 shows the subcategory of a column chart highlighted in the Dissemination View, which can be dragged and dropped into the Content Arrangement View.

Note that not every visualization will include all three subcategories. For instance, Key Performance Indicators (KPIs) typically lack interactivity and have overlapping general and insight information, so they are represented only by the subcategory *General*. Visualizations that are only filters, such as drop-down lists, do not provide insights and thus do not have the corresponding subcategory *Insight*.

Content Arrangement View

The interface of the Content Arrangement View enables the author to craft interactive dashboard tours with the subcategories picked from the Content Extraction View. Authors can either utilize predefined tours or create entirely new ones from scratch. Figure 5.2 shows an example tour created from scratch. In the following we explain the process of crafting a tour and then the means used for explaining the tour.

Crafting Interactive Tours. Our approach to interactive dashboard tours is inspired by non-linear storytelling and open-world game narratives (Section 5.3.1). This supports authors in crafting tours with linear, branching, or completely free narrative structures. We use a directed graph to represent D-Tours. The author can use the subcategories from the Content Extraction View (Subsection 5.4.1) as *story elements* for the *narrative* *structure* of the tour. To provide a sequence in this structure, the author can draw explicit edges between the elements. The branching structure can be specified by having multiple incoming and outgoing connections. These explicit edges indicate only the next element in the narrative of the tour. The implicit edges from the component graph obtained in Subsection 5.4.1 still remain valid and maintain the interaction relationships between the visualizations but are not shown to the author.

Figure 5.2 gives an example of an onboarding journey. It was crafted by dragging and dropping subcategories, such as a *General* component, from the column chart into the Content Arrangement View. The edges are created explicitly by connecting the elements, providing also a direction in the tour. Since this visual and interactive tour crafting is easy to use and requires no coding expertise, it is accessible to a broad audience.

Authors can group elements to improve the readability of the crafted journey. A *group* is a collection of elements bundled together by logical operations. They govern how users can navigate the tour (**G2**): ⊕ visit at least one, ! visit one, and ★ visit all elements of a group. Authors can thus specify the conditions for a user to proceed to the next step.

The choice of grouping depends on the author's objectives. For instance, in scenarios with multiple visualizations of the same type, an author might prefer the ? optional or ⊕ visit one choice to streamline the process. In this manner, the author can ensure that users grasp the key concepts without having to explore all similar visualizations.

In addition to giving authors the option to craft the sequence themselves (Figure 5.3 (Custom)), we also provide a few predefined tours based on traversing the component graph (**G3**). These predefined tours can be easily customized by adding more nodes from the Content Extraction View or by deleting existing ones. Currently, we support the following narrative templates (Figure 5.3):

- **Deep-Dive:** A detailed, sequential walkthrough of all visualizations in a dashboard, starting from the top. It covers all the subcategories of a visualization—*General*, *Insight*, and *Interaction*—before moving to the next visualization.

- **Similarity-based:** A traversal of visualizations from top to bottom grouped by visualization type. The categorization criteria can extend beyond just the type of visualization. They may include aspects such as data characteristics and insights. This approach is based on the work of Hullman et al. [84] and on GraphScape [212].

- **Martini Glass:** A traversal based on the Martini glass storytelling metaphor [171]. It starts with a broad overview of the dashboard and gradually focuses on specific elements, mimicking the narrow shape of a Martini glass.

The supplementary material details various design considerations for the nodes and the narrative structures.[2]

Explaining Tour Content. Individual elements in an interactive tour can also be explained to the end user in various ways using, for instance, text, audio, or video depending on user needs (**G4**). With the D-Tour Prototype the author can also *adjust the displayed content detail* by specifying the perceived or assumed expertise of the end user through a *user-level matrix*. Our prototype supports text, audio, video, and screen recording and

[2] https://osf.io/uf8ew

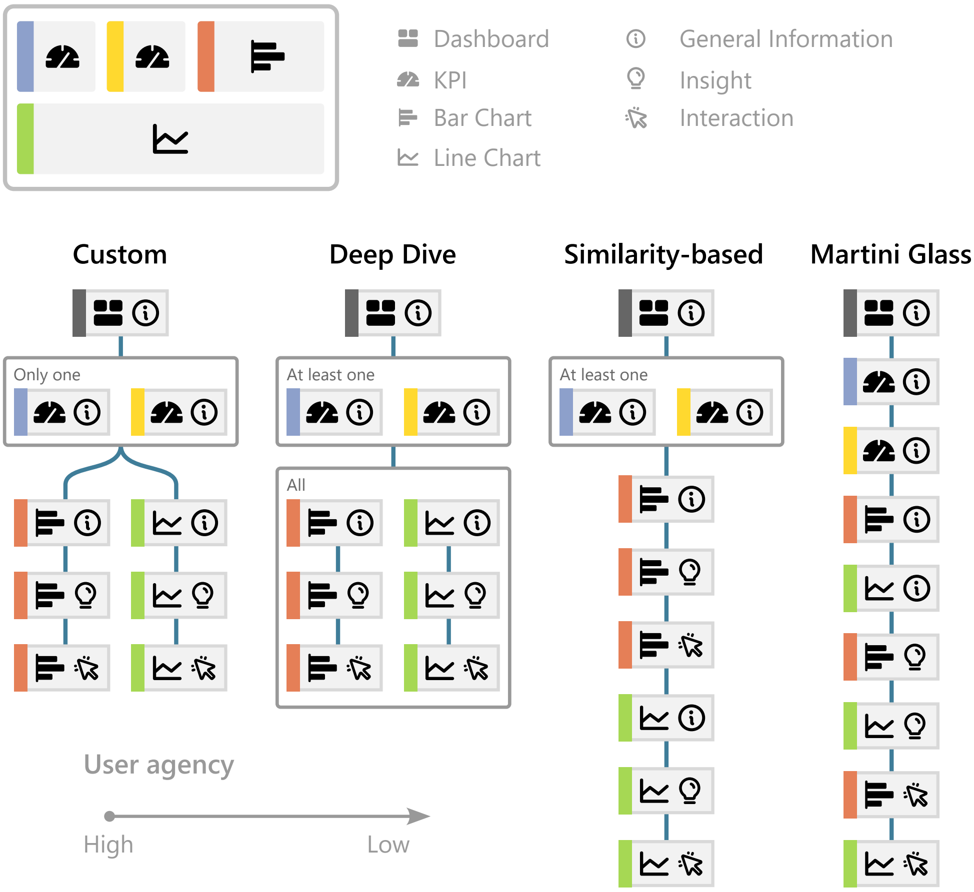


Fig. 5.3: Predefined Narrative Templates in the D-Tour Prototype. The templates allow authors to quickly create tours based on established narrative structures. Here, the simple example dashboard consists of two KPIs, a bar chart, and a line chart. At all times, authors can move to a completely customized narrative.

provides text as a default. The explanations are displayed at the bottom of the Content Arrangement View (Figure 5.2). They appear when the author selects a subcategory, such as *General* in this example from the Content Extraction View or an element (essentially a subcategory) in the Content Arrangement View.

Textual explanations are automatically generated from the component graph that is created and populated in the Content Extraction View.

In the Content Arrangement View, the explanations are attached to individual nodes. Herewith, authors can edit node-specific explanations even if they belong to the same subcategory of the same visualization in the Content Extraction View. This is helpful in cases where a node must be explained multiple times in different tour segments. In addition to text the author may also attach other means, such as a video, audio, or screen recording. The intention is to explain specific nodes, such as a filter interaction [8], that may be difficult for a new user to understand.

Beyond the node-specific explanations, the user-level matrix at the bottom right of the Content Arrangement View (Figure 5.2) can be used to set the expertise of the target user. This matrix displays two dimensions—domain and visualization expertise—on a scale from low to high. If the author sets the level of visualization expertise to high, only little *General* information will be shown, as the user is expected to be familiar with the visualization. Likewise, if the author sets the level of domain expertise to high, the default *Insight* description becomes concise. This supports **G2** by enabling authors to adjust the depicted details to the user group.

The Content Arrangement View is directly connected to the Dissemination View. All changes in the narrative structure of the tour or the content explanation immediately influence the Dissemination View.

Dissemination View

The Dissemination View is a miniature, real-time representation of the dashboard and incorporates interactions from the Content Extraction View and Content Arrangement View. As explained in Section 5.4.1, selections made in the Content Extraction View are concurrently highlighted in the Dissemination View (**G1**). Similarly, the nodes in the Content Arrangement View are also reflected in real time in the Dissemination View, which allows the authors to test both individual nodes and the entire narrative structure of the D-Tour as it is developed. This facilitates seamless exploration (**G5**), testing, and refinement of the tour, and eliminates the need to switch between modes.

If a subcategory is selected in the Content Extraction View or an element is selected in the Content Arrangement View, the corresponding visualization is highlighted in the Dissemination View with a pop-up showing the associated resources. Buttons facilitate navigation.

Once authors are satisfied with the onboarding narrative, they may optionally publish it, making the onboarding accessible to users. The authors can also edit and republish the onboarding story if necessary.

5.4.2 Onboarding mode

In the Onboarding mode, both authors and users can explore the onboarding story, which allows authors to experience the tour from a user's perspective (**G5**). Based on the concepts in Subsection 5.3.3, we explain how D-Tour incorporates user agency in this mode.

- **Choosing an onboarding style:** Users can navigate through the onboarding story in a guided manner. They either follow the predefined onboarding tour crafted by the author or choose a self-guided exploration mode. In the guided mode, the tour appears as a sequence of pop-up boxes with *next* and *previous* buttons. If a branch is encountered in the tour, users are prompted to select one of the paths indicated by green frames around specific visualizations. Their choice determines the course of the onboarding journey. Users can also backtrack and alter choices, to follow a different onboarding path.

  In the self-guided exploration mode, the onboarding journey is driven solely by user choice. The onboarding content for each visualization is accessible upon clicking it. There is no *next* and *previous* navigation, as the user is in charge of the narrative. Additionally, the users can access multiple presentation resources (such as video, audio or text) provided by the author (**G4**).

- **Adjusting the level of detail:** Users can adjust on the fly the detail level of the content they see during the onboarding.

- **Interacting with the dashboard:** As the onboarding is integrated into the dashboard, users can access the dashboard content at all times (**G1**). They can interact with the visualizations, and the onboarding content will update to reflect the changes. For interactive elements in the D-Tour, a *Try it out* option appears, encouraging active engagement with the dashboard.

The maximum level of control over the content and narrative gives users agency in their onboarding (**G2**). The onboarding finishes with a *close* button, but it can also be restarted.

5.5 Implementation

We implemented the D-Tour Prototype using the Microsoft Power BI Embedded report [227], which provides a practical dashboard solution and eliminates the need to build one from scratch. Using Power BI's REST API [226], we access comprehensive information about all the visualizations in any Power BI dashboard, making the D-Tour Prototype generalizable for dashboards with default Power BI visualizations. The source code of the D-Tour Prototype is available online [228].

We used the React framework [229] for the D-Tour Prototype front-end [229]. We employed Bootstrap [230] to structure the Content Extraction View, utilizing User Interface (UI) elements such as navigation bars, tabs, and accordions. Finally, we ensure that the component graph we construct and update via the Power BI REST APIs is in sync with the dashboard's current state.

The Content Arrangement View uses React-Flow [231], which offers support for building node-based editors. The Dissemination View integrates the Power BI dashboard and uses CSS and TypeScript for dynamic elements. We use FastAPI [232] and Python for the backend. These tools support multimedia content, such as audio and video.

5.6 Use Case: Sales and Marketing Dashboard

Here, we present the onboarding example of Section 5.4 and authored with the D-Tour Prototype in more detail. Marie is a mid-level executive working for a retail and manufacturing company. She has been asked to prepare an onboarding experience around the dashboard in Figure 5.2 for managers and employees to help them to monitor and analyze the company's industry standing. Managers are interested to see an overview of their sales and the most important insights while the employees need a more thorough onboarding on all the visualizations.

Authoring Mode. Marie starts the authoring process for the onboarding using the D-Tour Prototype. An overview of all the visualization types in the Content Extraction View helps her pick the elements required for the D-Tour. Instead of creating two different onboarding tours, Marie decides to create a single adaptable tour that accommodates both user groups. To create this tour, she starts from scratch and adds common categories from the Content Extraction View that are relevant to both user groups, such as *Introduction*, *Dashboard*, and filter *Interaction*. She also adds a supporting video to enhance the textual description of the filter.

The narrative then branches as depicted in Figure 5.2. The left branch is targeted towards the managers. It features grouped KPIs with an *Only one* option (), which allows managers to focus on KPIs of their particular interest. To explain the most important insights from each visualization type, the author creates a linear sequence of *Insight* elements for each unique visualization type. Since there are two column charts, Marie simply groups them with an *At least one* option ().

The right branch, specifically designed for employees, provides a detailed explanation of the visualizations. Therefore, Marie uses *All* option (★) to group all the visualizations' *General* elements. She wants to ensure that employees are familiarized with all visualizations before moving on to interactions. To explain the interactions, Marie creates a linear sequence of *Interaction* elements for each unique visualization type. This mirrors the left branch, but it is designed for the *Interaction* elements. The story concludes with a branch on each KPI, giving users a choice in their onboarding journey.

Onboarding Mode. Both user groups are given access to the onboarding and follow the prepared onboarding paths. The managers gravitate towards the KPI-focused branch, while employees' choices are influenced by their expertise. They can adjust the level of detail and interact with the dashboard while going through the prepared tour. After the onboarding has been completed, they can choose to close it or go back and follow the same or a different path. That both paths are accessible to both user groups makes onboarding flexible and adaptive.

Conclusion. Integrating different onboarding experiences into one dashboard decreases the effort required from the authors and allows them to create a single tour to onboard a variety of end users. A video demonstrating the scenario is available in the supplementary material.[3]

5.7 Evaluation

We validated the D-Tour Prototype in a user study with five onboarding authors and six users from our industrial collaborators.

5.7.1 User Study: Authoring Mode

The goal of the user study was to understand the challenges of current onboarding practices and how the D-Tour Prototype can address them and support authors in creating onboarding experiences for end users.

Study Method

We interviewed five authors from various departments and divisions of our industrial collaborators, who are active in the production of steel and steel-based technologies worldwide. The authors are experts in their domains, knowledgeable about visualization, and regularly onboard users. For each interview, we used a dashboard they had recently created and had planned to onboard end users. Each interview was conducted in person

[3] https://osf.io/q492g

and individually. We chose the authors for three main reasons, i.e., their: (i) experience with real-world onboarding scenarios, (ii) role as actual onboarding authors, and (iii) usage of the same technology stack as our D-Tour Prototype.

The interviews began with preliminary questions,[4] including a self-assessment of the authors' data and domain knowledge, and their experiences and challenges with onboarding. We then provided an introduction to the D-Tour Prototype and explained its features. This took 15-20 minutes, depending on the questions posed by the participants. They were subsequently requested to perform basic actions, such as clicking on the categories in the Content Extraction View and dragging and dropping them to the Content Arrangement View. This was done in order to facilitate familiarity with the system.

The main task of the study was for the authors to create an onboarding experience for one of their own dashboards. They were asked to incorporate as many details as necessary to onboard their end users. All authors had different onboarding goals in mind that were based on their dashboards and domains. For instance, an author from process engineering wanted to onboard a new colleague on the workings of a manufacturing plant by explaining indicators, thresholds, and system failures, if any. In contrast, an author who creates dashboards for the finance team wanted to convey the finance KPIs and the estimated cost-to-profit ratio of certain products to their users. They each were given 1.5 hours to create an onboarding experience. After completing the onboarding authoring, we invited the end users to onboard themselves with the resulting material. The authors of the onboarding material observed the onboarding sessions.

We ended sessions by asking authors to fill out a post-experiment questionnaire on ease of learning, ease of use, efficiency, and satisfaction with and intuitiveness of the D-Tour Prototype. Questions used a 7-point Likert scale (1: weakest agreement and 7: strongest agreement). We also asked open-ended questions about their experience with and opinions of the prototype. We recorded all sessions and interviews.

Session and Results

The average age of the authors was 38.8 years ($\sigma = 8.38$) and their experience in the company ranged from 5 to 25 years.

Challenges with current practices. All authors reported that they typically onboarded end users via documentation, videos, or online presentations, explaining each use case and the associated visualizations in the dashboard. Author A1 said that exposure to a new visualization type typically triggered questions that prompts them to *"explain every detail about it."* Author A2 mentioned being *"asked to create a video"* to provide reusable onboarding material. However, they felt that they are *"not a media person"*, so they instead *"ended up doing personal meetings"* for the onboarding. Author A3 observed that users had a *"fear of ruining the dashboard"* and *"do not know the [interactions] they can use within the dashboard."* If not properly onboarded, the users would *"just take the easiest way or go with the first guess"* without learning interactive ways to explore the material, which would *"hinder the analysis of the data."* Author A5 reported that, despite the dashboards having a consistent overall design due to *"corporate identity"*,

[4] https://osf.io/cedzu

they varied significantly, and without proper onboarding *"users could spend hours"* attempting to understand their functionality.

Using the D-Tour Prototype to create new onboarding experiences. Most authors opted to build their D-Tours from scratch, except A3, who switched from a predefined template to a custom tour due to an unanticipated technical issue. A1 started with sketching the intended onboarding experience on paper, where they picked the most important points that should be made clear. For example, they picked one parameter for which a value of zero indicates a production failure. After noting more of these points based on their domain knowledge, they switched to the D-Tour Prototype to directly craft their narrative. As the main language used in the company was German, almost all the authors, except A2, translated the default content, either partially or completely, to German. A1 found that *"you can guess 90 % of the questions in advance because of the automation and the flexibility of the message."* A3 attached a small video to explain the bar chart's drill-down functionality. Almost all authors combined elements with the *group* feature. After creating the narrative, all the authors thoroughly tested their D-Tours in the Dissemination View and Onboarding mode.

Experience with the D-Tour Prototype. A1 mentioned that using the D-Tour Prototype might *"reduce the training time of the colleagues"* and save them time as they *"did not need to be there all the time."* For authors A2, A3, and A5 the support of video content was valuable, especially for new users who are afraid of breaking the dashboard. A3 and A4 liked the self-guided option of onboarding but suggested that *"a small help button above the visualizations might be helpful."* All the authors found the D-Tour Prototype useful, time-saving, and easily adaptable to their dashboards. Nearly all authors highlighted the need for additional language support and adjustable font sizes, emphasizing the importance of accessibility in the onboarding experiences.

Almost all authors rated the experience metrics post-experiment as 6 or 7 on a 1–7 Likert scale. The exception was A1, who gave a score of 5 for intuitiveness, as they felt that the implementation could be improved, especially for creating groups. To create a group, an author must first select the components by clicking on them while pressing the Shift key. This was not obvious to A1, even though it was briefly explained before the study. To address this issue, we plan to add a group icon that can be dragged on existing components to form a group.

5.7.2 User Study: Onboarding Mode

We conducted another study to assess how users perceive the onboarding experience crafted with the D-Tour Prototype.

Study Method

The authors released the finished onboarding material to their end users. With the exception of author A4, who had two end users U4 and U5, all authors had one end user each. Nearly all the interviews were conducted in person. Only end user U6 was online, but they were given control of the application using an online remote conferencing tool.

The interviews began with asking end users preliminary questions[5], for instance, about self-assessing the data, domain knowledge, and visualization knowledge, how they are currently onboarded, and what difficulties they experience.

Sessions and Results

The average user was 34 years old (σ= 5.09) and spent an average of about 10 minutes on the onboarding sessions.

Challenges with current practices. End users U1 and U5 had joined the company three months before the interview. They reported being onboarded routinely via in-person meetings or *WebEx* calls. According to U1, this *"was an overwhelming experience as it was too much information in a short time."* The other experienced users preferred to be onboarded on the data and the domain, as they were typically familiar with the visualizations. End user U3 reported that they sometimes are onboarded via the bookmarks feature of Microsoft Power BI [233], which can be used to share insights into the data. However, they found that it affected the dashboard's performance and the rearrangement of the bookmarked elements consumed considerable time.

Using the D-Tour Prototype to explore onboarding experiences. All users intuitively used the D-Tour Prototype to view the created onboarding experience by the author. U6 also tried the free exploration to learn more about the visualizations that were not part of the prepared D-Tour. Almost all the users found the translated text by the authors helpful in understanding the dashboard. U3 faced technical difficulties with video resizing. U2 mentioned that *"more description on how to* try out *the dashboard interactions would have been helpful"*.

Experience with the D-Tour Prototype. All users responded positively to the onboarding experience. U1 mentioned that *"the interaction with the dashboard"* alongside the onboarding was very convenient. Similar thoughts were expressed by U2 and U3. U1 pointed out *"missing contextual knowledge"*, which author A1 had forgotten to add. This also provided hints to A1 for their next onboarding session. Similar to U1, U2 also mentioned that they would have wished for *"more descriptive text"* concerning some visualizations. Nearly all the users mentioned that with the D-Tour Prototype, they could onboard at their own pace, potentially reducing calls or meetings with the authors. U2 liked the choice of less or more details in the content. The other users found the translated content helpful. U3 suggested that an animation might be more effective than a video, although they commended the overall presentation style. U2 and U5 liked the simple design. Almost all users found the onboarding easy to use and self-explanatory.

5.8 Discussion

We have presented interactive dashboard tours (D-Tours) as an effective approach to creating a dashboard *onboarding experiences* that preserves user agency. We implemented the concept in a D-Tour Prototype and evaluated it to assess usability in real-life

[5] https://osf.io/cedzu

onboarding scenarios. The D-Tour Prototype significantly decreases the authoring effort necessary to create reusable onboarding experiences. The effectiveness of the onboarding ultimately depends on the decisions made by the authors. We discuss the limitations and future research directions based on the evaluation of the D-Tour Prototype, existing research, and an interview with an expert who holds a PhD in visualization. With this expert, we also discussed the challenges and the opportunities in onboarding from both an author's and a user's perspective[6]. The topics are discussed in Subsections 5.8.1 Design, 5.8.2 Development, and 5.8.3 Evaluation to highlight the challenges and opportunities in each area. While all mentioned points are relevant, we use ✧ to indicate those particularly interesting from a research perspective.

5.8.1 Design

Difference from open-world game design: While our design of interactive dashboard tours is inspired by open-world games, there the next level unlocks only after completing the previous one successfully. In our component onboarding, we have no such verification in place to check successful completion. Currently, Microsoft Power BI Rest APIs [226] fetch visualization data to help the authoring process. Future APIs could also help in verifying the onboarding success, for example by checking if a filter was correctly applied before moving on to the next component in the onboarding.

Multi-component onboarding: Currently, multiple components can be onboarded by grouping them together. However, the author has to check and update the description of each chart separately. An efficient approach could be to automatically update similar charts if one description changes.

Progress of the onboarding story: We received feedback during one of our pilot interviews to add a progress bar for conveying where the user is in the onboarding story. However, progress bars are directly applicable only to completely linear stories and may be misleading if users navigate through branched or free-form narratives.

5.8.2 Development

Optimizing onboarding narratives: Even with the D-Tour Prototype, choosing the onboarding material and narrative style for a given use case remains challenging. Additionally, currently, there is no support for automatically updating the narrative if the dashboard changes, requiring manual adjustments by the authors. Another limitation is the lack of functionality to export and import onboarding templates to other dashboards. Similar dashboards could benefit from using onboarding templates developed for a previous one. This would save authors time and potentially help create a standardized onboarding process.

Supporting more interactions and visualization types: One of the technical limitations of the D-Tour Prototype—but not of the general D-Tour concept behind it—stems from the constraints of the Microsoft Power BI APIs. While the Microsoft Power BI REST API [226] provides extensive details about the visualizations, it lacks specific interaction information, such as whether a visualization is highlighted. We infer missing information by analyzing changes in the displayed data and opacity to determine if parts are highlighted or filtered. Another limitation arises with non-standard visualizations

[6] https://osf.io/vqz76

from Microsoft AppSource [234], as the REST APIs do not provide information about their components. This makes it challenging to support custom visualizations. Multi-modal generative models could analyze dashboard screenshots to enhance support for non-standard visualizations.

Generalizability beyond Microsoft Power BI: Although the D-Tour Prototype uses Microsoft technology, the concept could be applied to other dashboard systems as well. The essential requirement is the ability to extract information about the visualizations within a dashboard. Once extracted, this information can be provided to the component graph, which can support additional visualization types and is applicable to non-Power BI dashboards as well.

Accessibility: The automatically generated descriptions for the components in the D-Tour Prototype are currently in English. User feedback sessions revealed that authors frequently translate the descriptions to German to align them to their users' preferences. Adding support for multiple languages or translations would enhance accessibility. Additionally, improving font colors, background contrast, and font sizes could increase usability.

Conversational onboarding interface: Microsoft Power BI has recently launched an AI interface (co-pilot), which can help the authors create a dashboard from the given data and create a narrative about the dashboard. The main focus of this and similar AI interfaces for other BI platforms is on visualization creation and presentation. The approaches currently do not help authors create onboarding strategies or enhance onboarding material like text. A combination of onboarding created through the D-Tour Prototype and conversational interfaces to ask further questions could prove to be effective in this regard.

Collaborative onboarding process: Similar to a collaborative dashboard design process, onboarding could also be designed collaboratively by teaming up expert users before rolling out the onboarding material to a larger audience. This might enhance the *effectiveness of the created onboarding* and help with its wider adoption.

Adapting to different dashboard design patterns: Bach et al. [17] introduced design patterns for dashboards. The D-Tour Prototype already supports onboarding for dashboards that use many of these patterns. We plan to provide additional assistance for patterns like multi-page dashboards, parallel structures, and a wider range of interactions. A tabular summary of design patterns supported by the D-Tour Prototype can be found in our supplementary material [7].

5.8.3 Evaluation

Number of participants and dashboards: Due to the small number of participants and consequently dashboards, some of the advanced features of the D-Tour Prototype have not been utilized, such as multiple branchings. We may have missed edge cases that are not properly supported by the D-Tour Prototype. As the goal of the user study was to collect rich qualitative feedback to improve future versions of the D-Tour Prototype, we focused on the insights gained from interviews. In the absence of established guidelines for dashboard onboarding, it also becomes difficult to create a baseline for comparison.

Effectiveness of the created onboarding: A primary challenge in measuring the effectiveness of an onboarding experience is the subjectivity of learning, as it is a personal

[7] https://osf.io/f5jru

process. A longitudinal study might provide quantitative numbers by deploying the D-Tour Prototype at the collaborator's site over an extended period of time. Through tracking user interaction logs, we could assess whether the onboarding helped users get started with the dashboards, accomplish their tasks, discover new insights, and reduce communication time with the author.

Role-based usage analytics: We plan to deploy the D-Tour Prototype to study the *effectiveness of onboarding* and how much time authors might save as compared to their current onboarding practices. Based on the expert interview and participants' feedback from the user study in Subsection 5.7.1, it takes authors up to 30 minutes to onboard a single user in an in-person meeting. This depends on the size of the dashboard, the pages, the complexity of the visualizations, and other factors. While some authors prefer documenting the dashboard and insights, others spend more time on user questions. We plan to also save user interaction logs, which we already started by integrating the Trrack library [235][8]. This can help us identify the level of expertise so that the author does not have to manually specify it for each onboarding.

5.9 Conclusion

We propose *interactive dashboard tours* (*D-Tours*) as semi-automated onboarding experiences that support users with various levels of expertise. The interactive tours concept draws from open-world game design to give the user freedom in choosing their path in the onboarding. To demonstrate the applicability of our work, we implemented the concept in a tool called D-Tour Prototype, which allows authors to craft custom and interactive dashboard tours from scratch or use automatic templates. We validated the prototype with user studies for the authoring and onboarding modes. The authors' feedback was positive, as they successfully created reusable onboarding experiences with little effort. End-users also found the onboarding narrative engaging and expressed a desire to continue using such tours for their future onboarding needs.

[8] https://shorturl.at/W8qHN

6 Discussion & Outlook

The proliferation of commercial visual analytics tools, such as Microsoft Power BI, Tableau, and Spotfire, has accelerated the use of interactive visualizations across all organizational layers. This rapid adoption of visualization-based systems has brought about significant changes to the foundational models, such as the 1999 information visualization pipeline by Stuart K. Card et al. [1] and the 2008 visual analytics model by Daniel Keim et al. [3].

The advancement of technology has led to significant improvements in data preparation and analysis steps in the visual analytics pipeline. Additionally, the need for real-time data analysis, scalability, and collaboration has further introduced deployment and communication steps.

This thesis focuses on the challenges concerning the analysis and communication components of the visual analytics pipeline while also addressing data preparation and deployment to some extent. The first part of the thesis focuses on analyzing uncertain information in tabular data within spreadsheets using compact in-cell visualizations. We refer to it as the *Fuzzy Spreadsheet* approach that augments well-established spreadsheets with uncertain information and other contextually relevant details, such as impact and relationship between cells.

However, as spreadsheets offer limited visualization and interaction capabilities and often run into performance issues with large datasets, we observed that users shift toward more advanced visual analytics tools. While many resources support authors in designing and creating effective dashboards, we found little to no support for the dashboard users in the literature and real life.

Therefore, the second part of the thesis focuses on communicating to users how to use a dashboard. We focused on improving the dashboard experience for users by emphasizing the importance of onboarding. The process of dashboard onboarding involves introducing a user to the purpose, data, visualizations, interactions, and exemplary insights present in the dashboard. First, we created a *process model for dashboard onboarding* that introduces an onboarding loop alongside the dashboard usage loop. We expand on these loops and illustrate common onboarding scenarios using our model. In the second work, we focus on both the dashboard authors and users in the process of dashboard onboarding. Utilizing the semi-automatic dashboard tours, i.e., the *D-Tour* concept, the dashboard author can design onboarding experiences that preserve the agency of users with various levels of expertise to keep them interested and engaged. This concept is also implemented in a tool called D-Tour Prototype, which allows authors to craft custom interactive dashboard tours from scratch or to employ automatic templates.

6.1 Reflecting on the Research Questions

- **RQ1**: How can we support users in comprehending and exploring uncertainty during the analysis process?

We developed the Fuzzy Spreadsheet approach, which augments well-established spreadsheets with compact in-cell visualizations. These visualizations help users assess probability distributions and understand computational relationships directly within the spreadsheet cells. We collected the most prevalent user tasks that are supported by various tools for uncertainty analysis in tabular layout and assigned those tasks into the context of existing analysis tasks frameworks [134, 135]. Fuzzy Spreadsheet addresses the analysis tasks by adding compact visualizations embedded within the spreadsheet and by introducing a side panel with a user interface that shows additional information useful for the analysis. It allows users to understand the uncertain information present within a cell and how this uncertainty propagates to its related cells. The users can also compare alternative scenarios utilizing what-if analyses. We performed a small-scale user study to compare Fuzzy Spreadsheet with traditional spreadsheets regarding answer correctness, response time, mental effort, and usability. The results indicate that Fuzzy Spreadsheet outperforms traditional spreadsheets and empowers users to understand and explore uncertain information.

- **RQ2**: How can we support users in understanding complex, interlinked visualizations, such as in a dashboard?

  To learn how we can assist users in understanding multiple interlinked visualizations, we delved into the literature on visualization onboarding, guidance, and data-driven storytelling. We focused on the concept of onboarding as opposed to guidance, as our goal has been to provide explanations on the interpretation and interactions of the visualizations, as well as of the underlying data. The goal of guidance is to support users in performing specific tasks with visual analytics tools [94]. In our work [6], we analyzed various onboarding strategies and developed a conceptual model that unifies and formalizes different dashboard onboarding strategies. We demonstrated the generalizability of our model using four real-world examples (video, textual explanation, programmed guided tour, and interactive onboarding with a human presenter). The effectiveness of the onboarding is challenging to measure as it depends on various factors, such as the quality of the onboarding material, presentation and engagement skills of the onboarding author, and the skill set of the user, among others. To quantitatively assess these onboarding and learning experiences, we might need to deploy various onboarding strategies at a larger organization. This would involve monitoring data such as the user interaction logs and collecting qualitative feedback on whether the onboarding process helped the users accomplish their tasks, discover new insights, and decrease communication time with the author.

- **RQ3**: How can we support dashboard authors in developing onboarding approaches tailored to their end users while maintaining the users' autonomy?

  During our onboarding experience with collaborators and through existing literature, we realized the importance of user autonomy concerning the onboarding material. Without this autonomy, the users might find the onboarding tedious and boring. Additionally, based on the knowledge and skills of the users, they may want to go into more detail about the prepared onboarding or choose fewer details. Therefore, we focused on developing an onboarding approach that can be tailored to individual users' needs and preserve their agency. We drew inspiration from the field of data-driven storytelling and open-world video games. Similar to open-world games such as Elden Ring (2022) and Hogwarts Academy (2023), users can choose when to view the components within author-defined constraints on the content. We adapt this approach in our D-Tour Prototype, which allows authors to craft onboarding experiences using existing templates or from scratch. Users can go through these tours while interacting with the dashboard and tailor the content as they go, with the option of switching to free exploration at any time.

6.2 Generalizability of the Results

The prototypes presented in this thesis use the Microsoft technology stack because Microsoft APIs are easily accessible to developers. Fuzzy Spreadsheet approach uses Microsoft Excel to demonstrate its uncertainty authoring technique. However, this approach with in-cell encoding and other contextually relevant information, such as impact, relationship, and likelihood of the cell, may be applied to any tabular data format with uncertainty. The underlying calculation currently depends on an Excel-based formula, which can be replaced with the formula of any other spreadsheet-based program.

D-Tour Prototype uses Microsoft Power BI REST APIs [226] to implement the D-Tour approach. These APIs allow developers to embed Microsoft Power BI dashboards into custom web applications and grant access to important information about the visualizations and their interactions. Even though we use Microsoft technology, the underlying concept behind D-Tour is adaptable to other dashboard systems. Any platform that provides information about visualizations and their interactions can leverage our implementation to integrate D-Tour Prototype into their respective systems.

6.3 Future Research Directions

Drawing from my PhD research experiences working with industry partners and supervising several students, I plan to focus on three primary areas for my future research: storytelling, artificial intelligence, and accessibility.

6.3.1 Storytelling: Survey on Authoring Tools

While researching data-driven storytelling to craft engaging onboarding narratives, I realized that existing works on data-driven storytelling [236, 237, 238, 239] tend to focus on the *consumer experience* of data-driven stories: viewing and interacting with already created data stories. However, the experience of authors creating the stories seems to be less investigated. While arguably any form of data-driven story can be created from scratch given enough time and expertise, dedicated authoring software for data-driven storytelling (e.g., data-driven story editors) has several benefits:

- Lowering barriers: enable novices to author data-driven stories without writing code;

- Streamlining the process: facilitate rapid authoring of data-driven stories; and

- Guiding narration: promote efficient use of narrative elements specific to the storytelling media.

Accordingly, many data-driven authoring tools are now available, such as DataClips [240] for designing data animations, Roslingifier for constructing data presentations [91], and DataComicsJS for data comics [241]. However, these tools are still isolated examples with little standardization, shared practices and terminologies, and few common design patterns. In an attempt to understand existing storytelling mechanisms and categorize them from a narrative perspective, such that authors can understand the tools available for data-driven stories, we plan to survey the current state of the art in *authoring tools for data-driven storytelling*.

Data-driven storytelling has been covered in several recent surveys, but they tend to focus on aspects other than the authoring experience. The data-driven storytelling book by Henry Riche et al. [237] gives an overview of the area and its research challenges. Tong et al. [239] surveys the scope of storytelling in various domains such as scientific visualization, information visualization, and geospatial visualization. Ren et al. [242] look at authoring tools in data storytelling from a narrative perspective. Similar to their work, we also aim to survey authoring for data-driven storytelling. However, our approach will identify advanced narrative elements, storytelling mechanisms, and design patterns and suggest new areas of work. More recently, Chen et al. [236] also surveyed the field of authoring tools in data storytelling, focusing on automation in narrative visualization. Our survey will study all forms of story authoring, both automated and manual. Finally, Zhao and Elmqvist [243] present an updated look at new media for data-driven storytelling

With our planned survey, we aim to fill a gap in the current literature, offering visualization content creators a comprehensive understanding of the tools available for authoring data-driven stories. Many studies have focused on specific storytelling modalities, overlooking the broader spectrum essential for crafting effective data narratives. Our survey educates on existing tools and suggests directions for future tool development. We also propose a shift in perspective: from treating storytelling as an afterthought to making it an active component in data visualization creation. This can foster greater accessibility, inclusivity, and visualization literacy. Furthermore, we highlight new data storytelling mechanisms that have evolved since Segel's 2010 discussion on narrative visualization [35].

6.3.2 AI-Enhanced Dashboard Onboarding

With recent advancements in artificial intelligence (AI) and machine learning (ML) research, tools that automate the analysis and communication processes are becoming common in both research and industrial domains [43, 16]. Microsoft recently launched Copilot [43] that can generate an automated summary of the data model in a dashboard, generate automated reports, and answer various questions related to the dashboard content. This effectively supports users in both analysis and communication tasks. Meanwhile, in the literature, there are several examples of automated report creation [16] and automated data-driven narrative creation for communication purposes [245]. While the automation offered by these tools can be useful, they should focus on empowering and not replacing humans, a concern discussed in various literature sources [246, 247].

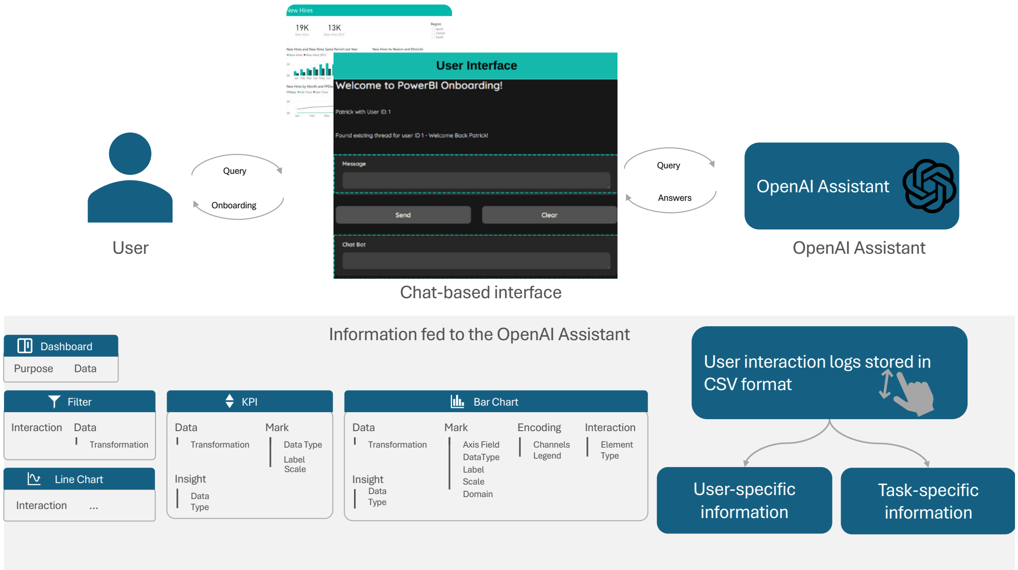


Fig. 6.1: Onboarding Assistant Interface for an exemplary Microsoft Power BI Dashboard based on the work by Patrick Mair [244].

Ben Shneiderman [246], in his book Human-Centered AI provides recommendations on how various user roles, such as programmers, leaders, educators, and policymakers, can implement human-centered AI (HCAI), which places human values and agency at the forefront and empowers users. This implementation can be done using AI-infused supertools, which "enable people to see, think, create, and act in extraordinary ways by combining potent user experiences with embedded AI support services." Hoque et al. [248] point out the lack of a clear definition of how to build such tools. They propose interactive visualizations as a key enabling technology for building such HCAI software tools. Visualization is often used for explaining AI models due to their effectiveness in representing complex and large-scale datasets [13]. An approach that combines visualizations with data-driven, semantic, and unified interactive feedback loops to integrate AI models into the loop with human users is referred to as visualization-enabled HCAI tools [248]. Hoque et al. [248] discuss the capabilities of and concerns with such tools. They use examples from their work to extract design guidelines on constructing visual representations in support of HCAI principles.

As my research is motivated by real-world scenarios, using AI to enhance my work, especially in the field of dashboard onboarding, can *amplify*, *augment*, *empower*, and *enhance* the capabilities [246] of the currently existing D-Tour Prototype while keeping the user in the loop. We have already started research in this direction that aims to enhance the understanding of a dashboard by applying Large Language Models (LLMs). This approach encompasses several strategic measures:

- Crafting onboarding content by engineering prompts based on specific keywords of each visualization.
- Leveraging the LLM's capabilities to determine the onboarding sequence by the type, location, and data inherent to each visual.
- Dynamically adjusting the narrative to cater to varied user expertise levels.

- Implementing a responsive interface where users can seek further clarification, for instance, by prompting questions like "What's the purpose of the legend?". The LLM-backed onboarding then adapts its responses, ensuring a personalized and enriched user experience inside data visualization platforms.

- Automated text generation to explain a visualization that is further enhanced by explaining certain words or phrases the user might not know.

A preliminary implementation of such an approach can be found at `https://github.com/PatrickJKU/Dashboard-Onboarding` [244]. In this work, shown in Figure 6.1, we use Retrieval Augmented Generation (RAG) combined with LLMs to enhance the onboarding experience. The domain-specific data is provided in the form of interaction logs and via component graphs explained in Chapter 4. We use the OpenAI API for creating *Assistants* [249], which supports users in understanding dashboard content. Assistants communicate through a chat-based interface embedded in the dashboard, as shown in Figure 6.1. The assistant can be tuned to the user level, i.e., beginner, intermediate, advanced, and prompted to provide an onboarding tour. Preliminary results using a sample dashboard with three different user assistants show that they provide thorough explanations and give step-by-step instructions on how to interact with a visualization, for instance. While hallucination is still an issue in this process, we are working on improving our approach. Additionally, we plan to compare our work with Microsoft Copilot [43], which aims to support users throughout the dashboard creation and usage process.

In another line of research, we are working on generating an automated sequence for onboarding using LLMs and applying it to multiple dashboards with multiple users.

6.3.3 Accessibility for Dashboards

Elmqvist [14] wrote an article on Visualization for the Blind, where he highlighted the need to make visualizations accessible for blind and low vision (BLV) users. As visualization researchers, we should aim to reduce the barriers to data visualization for everyone and make data accessible. Kim et al. [250] explore chart question-answering systems to support visualization interpretation and evaluation for BLV users. Efforts in the field of making visualization accessible include Chartability [251], Susurrus [252], and TactualPlot [253]. Chartability [251] helps evaluate visualizations for visual, motor, vestibular, neurological, and cognitive accessibility. Susurrus [252] uses sonification for accessible data representation. It maps visualization marks, such as a bar from a bar chart to natural sound drawn from an ambient theme. The data values are conveyed based on the loudness of sounds in decibels proportional to the bars. Based on their user study results, Susurrus helped improve user performance, especially for people who do not have musical training. TactualPlot [253] is an approach to *sensory substitution*, where touch interaction provides auditory feedback. They combine both touch and sound for a scalable data exploration using two-dimensional scatterplots. Srinivasan et al. [200] identified design goals for accessible dashboards through an interactive co-design study with two blind users and demonstrated them through their prototype, Azimuth. Azimuth takes a JSON-based dashboard specification and converts it into a web-based dashboard, which is optimized for screen readers, along with descriptions to help with dashboard understanding and interaction. Our work on dashboards focuses on explaining dashboard visualizations and their interactions. As a natural next step, we can extend our work to make dashboards more accessible. We can also incorporate different output modalities in our dashboards, such as audio and touch, that have so far been applied to single visualizations.

I believe that as visualization researchers it is our responsibility to not just develop novel visualizations, but also provide a platform for users to understand and use those visualizations in their daily workflow.

References

[1] Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman. *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1999. ISBN: 978-1-55860-533-6.

[2] James J. Thomas and Kristin A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society Press, 2005. ISBN: 978-0-7695-2323-1.

[3] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. "Visual Analytics: Definition, Process, and Challenges". In: *Information Visualization*. Ed. by Andreas Kerren, John T. Stasko, Jean-Daniel Fekete, and Chris North. Lecture Notes in Computer Science 4950. Springer Berlin Heidelberg, 2008, pp. 154–175. ISBN: 978-3-540-70955-8.

[4] Anamaria Crisan, Brittany Fiore-Gartland, and Melanie Tory. "Passing the Data Baton : A Retrospective Analysis on Data Science Work and Workers". In: *IEEE Transactions on Visualization and Computer Graphics* 27.2 (2021), pp. 1860–1870. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2020.3030340. URL: https://ieeexplore.ieee.org/document/9222030/ (visited on 11/26/2021).

[5] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Eduard Gröller, and Marc Streit. "Fuzzy Spreadsheet: Understanding and Exploring Uncertainties in Tabular Calculations". In: *IEEE Transactions on Visualization and Computer Graphics* 29.2 (2023), pp. 1463–1477. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2021.3119212.

[6] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Holger Stitz, Eduard Gröller, and Marc Streit. "A Process Model for Dashboard Onboarding". In: *Computer Graphics Forum* 41 (2022), pp. 501–513. DOI: 10.1111/cgf.14558.

[7] Vaishali Dhanoa, Andreas Hinterreiter, Vanessa Fediuk, Niklas Elmqvist, Eduard Gröller, and Marc Streit. "D-Tour: Semi-Automatic Generation of Interactive Guided Tours for Visualization Dashboard Onboarding". In: *IEEE Transactions on Visualization and Computer Graphics* (2024). DOI: 10.1109/TVCG.2024.3456347.

[8] Conny Walchshofer, Vaishali Dhanoa, Marc Streit, and Miriah Meyer. "Transitioning to a Commercial Dashboarding System: Socio-technical Observations and Opportunities". In: *IEEE Transactions on Visualization and Computer Graphics* 30.1 (2023), pp. 381–391. DOI: 10.1109/TVCG.2023.3326525.

[9] Matthew Brehmer and Robert Kosara. "From Jam Session to Recital: Synchronous Communication and Collaboration Around Data in Organizations". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (2022). DOI: 10.1109/TVCG.2021.3114760.

[10] Bongshin Lee, N.H. Riche, P. Isenberg, and S. Carpendale. "More Than Telling a Story: Transforming Data into Visually Shared Stories". In: *IEEE Computer Graphics and Applications* 35.5 (2015), pp. 84–90. ISSN: 0272-1716. DOI: 10.1109/MCG.2015.99.

[11] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. "SketchStory: Telling More Engaging Stories with Data through Freeform Sketching". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (Dec. 2013), pp. 2416–2425. ISSN: 1077-2626. DOI: 10.1109/TVCG.2013.191. URL: http://ieeexplore.ieee.org/document/6634113/ (visited on 10/12/2023).

[12] Christina Stoiber, Conny Walchshofer, Florian Grassinger, Holger Stitz, Marc Streit, and Wolfgang Aigner. "Design and Comparative Evaluation of Visualization Onboarding Methods". en. In: *Symposium on Visual Information Communication and Interaction (VINCI'21)*. ACM, 2021, pp. 1–5. ISBN: 978-1-4503-8647-0. DOI: 10.1145/3481549.3481558. URL: https://dl.acm.org/doi/10.1145/3481549.3481558 (visited on 04/19/2022).

[13] Peter Kerpedjiev, Nezar Abdennur, Fritz Lekschas, Chuck McCallum, Kasper Dinkla, Hendrik Strobelt, Jacob M. Luber, Scott B. Ouellette, Alaleh Azhir, Nikhil Kumar, Jeewon Hwang, Soohyun Lee, Burak H. Alver, Hanspeter Pfister, Leonid A. Mirny, Peter J. Park, and Nils Gehlenborg. "HiGlass: web-based visual exploration and analysis of genome interaction maps". In: *Genome Biology* 19.1 (Aug. 2018), p. 125. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1486-1. URL: https://doi.org/10.1186/s13059-018-1486-1 (visited on 10/29/2020).

[14] Niklas Elmqvist. "Data Analytics Anywhere and Everywhere". en. In: *Communications of the ACM* 66.12 (Dec. 2023), pp. 52–63. ISSN: 0001-0782, 1557-7317. DOI: 10.1145/3584858. URL: https://dl.acm.org/doi/10.1145/3584858 (visited on 08/26/2024).

[15] Maggie Sparkman, Mona Berdugo, Paul Inbar, Tim Sherer, Justin Piesco, David Iseminger, Michele Hart, and Kent Sharkey. *Sales and Marketing sample for Power BI: Take a tour*. Last accessed on 2023/11/29. June 27, 2023. URL: https://learn.microsoft.com/en-us/power-bi/create-reports/sample-sales-and-marketing (visited on 11/29/2023).

[16] Arjun Srinivasan and Vidya Setlur. "BOLT: A Natural Language Interface for Dashboard Authoring". In: *EuroVis 2023 - Short Papers*. Ed. by Thomas Hoellt, Wolfgang Aigner, and Bei Wang. The Eurographics Association, 2023. ISBN: 978-3-03868-219-6. DOI: 10.2312/evs.20231035.

[17] Benjamin Bach, Euan Freeman, Alfie Abdul-Rahman, Cagatay Turkay, Saiful Khan, Yulei Fan, and Min Chen. "Dashboard Design Patterns". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2023), pp. 342–352. DOI: 10.1109/TVCG.2022.3209448.

[18] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. "What Do We Talk About When We Talk About Dashboards?" In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 682–692. DOI: 10.1109/TVCG.2018.2864903.

[19] Christina Stoiber, Davide Ceneda, Markus Wagner, Victor Schetinger, Theresia Gschwandtner, Marc Streit, Silvia Miksch, and Wolfgang Aigner. "Perspectives of visualization onboarding and guidance in VA". In: *Visual Informatics* 6.1 (2022), pp. 68–83. DOI: 10.1016/j.visinf.2022.02.005.

[20] Melanie Tory, Lyn Bartram, Brittany Fiore-Gartland, and Anamaria Crisan. "Finding Their Data Voice: Practices and Challenges of Dashboard Users". en. In: *IEEE Computer Graphics and Applications* (2022), pp. 1–1. ISSN: 0272-1716, 1558-1756. DOI: 10.1109/MCG.2021.3136545. URL: https://ieeexplore.ieee.org/document/9656613/ (visited on 11/09/2022).

[21] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. "What Do We Talk About When We Talk About Dashboards?" In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019), pp. 682–692. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2864903. URL: https://ieeexplore.ieee.org/document/8443395/ (visited on 10/11/2023).

[22] Datapine. *Management Overview*. Accessed: 2022-08-26. 2015. URL: https://www.datapine.com/blog/wp-content/uploads/2016/09/time-interval-widget.gif (visited on 2015).

[23] Olivia Padilla Jackson. *Economic Dashboard by City of Albuquerque*. Accessed: 2024-08-26. 2015. URL: https://public.tableau.com/profile/olivia.padilla.jackson#!/vizhome/EconomicDashboard_2/CABQEconomicDashboard (visited on 2015).

[24] Tableau. *50 Years of Crime in the US*. Accessed: 2022-08-26. 2022. URL: https://public.tableau.com/en-us/s/gallery/50-years-crime-us (visited on 2022).

[25] UNHCR. *Focus on Refugees*. Accessed: 2024-08-26. 2022. URL: https://focusonrefugees.org/wp-content/uploads/2016/05/UNHCR-migration-data-visualisation.png (visited on 2022).

[26] Empower. *Personal Capital Portfolio Management*. Accessed: 2024-08-26. 2022. URL: https://www.personalcapital.com/financial-software/mobile-app (visited on 2022).

[27] Hongsuda Tangmunarunkit, Cheng-Kang Hsieh, Brent Longstaff, S. Nolen, John Jenkins, Cameron Ketcham, Joshua Selsky, Faisal Alquaddoomi, Dony George, Jinha Kang, Z. Khalapyan, Jeroen Ooms, Nithya Ramanathan, and Deborah Estrin. "Ohmage". In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 6 (2015), pp. 1–21. URL: https://api.semanticscholar.org/CorpusID:14256939.

[28] Stephen Few. *Information Dashboard Design: The Effective Visual Communication of Data*. eng. OCLC: 946802363. O'Reilly Media, 2006. ISBN: 978-1-60033-019-3.

[29] Perceptual Edge Stephen Few. *Dashboard Confusion Revisited*. Accessed: 2024-08-26. 2007. URL: https://www.perceptualedge.com/articles/visual_business_intelligence/dboard_confusion_revisited.pdf (visited on 2024).

[30] Steve Wexler, Jeffrey Shaffer, and Andy Cotgreave. *The big book of dashboards: visualizing your data using real-world business scenarios*. Hoboken, New Jersey: Wiley, 2017. ISBN: 978-1-119-28271-6.

[31] Rob Kitchin, Tracey P. Lauriault, and Gavin McArdle. "Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards". en. In: *Regional Studies, Regional Science* 2.1 (Jan. 2015), pp. 6–28. ISSN: 2168-1376. DOI: 10.1080/21681376.2014.983149. URL: https://www.tandfonline.com/doi/full/10.1080/21681376.2014.983149 (visited on 08/28/2024).

[32] Benjamin Bach, Zezhong Wang, Matteo Farinella, Dave Murray-Rust, and Nathalie Henry Riche. "Design Patterns for Data Comics". en. In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC Canada: ACM, Apr. 2018, pp. 1–12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173612. URL: https://dl.acm.org/doi/10.1145/3173574.3173612 (visited on 10/12/2023).

[33] Nils Rasmussen, Claire Y. Chen, and Manish Bansal. *Business dashboards: a visual catalog for design and deployment*. eng. Chichester: John Wiley, 2009. ISBN: 978-0-470-41347-0.

[34] Bryan A. Wilbanks and Patsy A. Langford. "A Review of Dashboards for Data Analytics in Nursing". en. In: *CIN: Computers, Informatics, Nursing* 32.11 (Nov. 2014), pp. 545–549. ISSN: 1538-2931. DOI: 10.1097/CIN.0000000000000106. URL: https://journals.lww.com/00024665-201411000-00006 (visited on 08/28/2024).

[35] Edward Segel and Jeffrey Heer. "Narrative Visualization: Telling Stories with Data". In: *IEEE Transactions on Visualization and Computer Graphics* 16 (2010), pp. 1139–1148. ISSN: 1941-0506. DOI: 10.1109/TVCG.2010.179. URL: https://ieeexplore.ieee.org/document/5613452.

[36] Matej Vukovic, Vaishali Dhanoa, Markus Jäger, Conny Walchshofer, Josef Küng, Petra Krahwinkler, Belgin Mutlu, and Stefan Thalmann. "A Forecasting Model-Based Discovery of Causal Links of Key Influencing Performance Quality Indicators for Sinter Production Improvement". In: *AISTech2020 Proceedings of the Iron and Steel Technology Conference*. AIST, 2020, pp. 2028–2038. ISBN: 978-1-935117-88-9. DOI: 10.33313/380/218.

[37] Klaus Eckelt, Andreas Hinterreiter, Patrick Adelberger, Conny Walchshofer, Vaishali Dhanoa, Christina Humer, Moritz Heckmann, Christian Steinparz, and Marc Streit. "Visual Exploration of Relationships and Structure in Low-Dimensional Embeddings". In: *IEEE Transactions on Visualization and Computer Graphics* 29.7 (2023), pp. 3312–3326. DOI: 10.1109/TVCG.2022.3156760.

[38] Palisade. *@RISK*. https://www.palisade.com/risk/. Accessed: Sep. 1, 2021. 2020.

[39] Oracle Corporation. *Oracle Crystal Ball*. https://www.oracle.com/applications/crystalball/. Accessed: Sep. 1, 2021. 2020.

[40] Probability Management. *Probability Management*. https://www.probabilitymanagement.org/sipmath. Accessed: Sep. 1, 2021. 2021.

[41] Guesstimate Inc. *Guesstimate*. https://www.getguesstimate.com/. Accessed: Sep. 1, 2021. 2019.

[42] Andrea Batch and Niklas Elmqvist. “The Interactive Visualization Gap in Initial Exploratory Data Analysis”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (Jan. 2018), pp. 278–287. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2017.2743990. URL: https://ieeexplore.ieee.org/document/8017577/ (visited on 07/30/2024).

[43] Microsoft Power BI. *Copilot for Power BI*. Accessed: 2024-08-22. 2021. URL: https://learn.microsoft.com/en-us/power-bi/create-reports/copilot-introduction (visited on 2024).

[44] Karl W. Broman and Kara H. Woo. “Data Organization in Spreadsheets”. en. In: *The American Statistician* 72.1 (Jan. 2018), pp. 2–10. ISSN: 0003-1305, 1537-2731. DOI: 10.1080/00031305.2017.1375989. URL: https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989 (visited on 09/03/2024).

[45] J.D. Pemberton and A.J. Robson. “Spreadsheets in business”. en. In: *Industrial Management & Data Systems* 100.8 (Nov. 2000), pp. 379–388. ISSN: 0263-5577. DOI: 10.1108/02635570010353938. URL: https://www.emerald.com/insight/content/doi/10.1108/02635570010353938/full/html (visited on 09/03/2024).

[46] E.H. Chi, J. Riedl, P. Barry, and J.A. Konstan. “Principles for information visualization spreadsheets”. In: *IEEE Computer Graphics and Applications* 18.4 (Aug. 1998), pp. 30–38. ISSN: 02721716. DOI: 10.1109/38.689659. URL: http://ieeexplore.ieee.org/document/689659/ (visited on 09/03/2024).

[47] Mihai Budiu, Parikshit Gopalan, Lalith Suresh, Udi Wieder, Han Kruiger, and Marcos K. Aguilera. “Hillview: A trillion-cell spreadsheet for big data”. In: *Proceedings of the VLDB Endowment* 12.11 (July 2019). arXiv:1907.04827 [cs], pp. 1442–1457. ISSN: 2150-8097. DOI: 10.14778/3342263.3342279. URL: http://arxiv.org/abs/1907.04827 (visited on 09/04/2024).

[48] David Birch, David Lyford-Smith, and Yike Guo. *The Future of Spreadsheets in the Big Data Era*. arXiv:1801.10231 [cs]. Jan. 2018. URL: http://arxiv.org/abs/1801.10231 (visited on 09/04/2024).

[49] Mangesh Bendre, Vipul Venkataraman, Xinyan Zhou, Kevin Chang, and Aditya Parameswaran. “Towards a Holistic Integration of Spreadsheets with Databases: A Scalable Storage Engine for Presentational Data Management”. In: *2018 IEEE 34th International Conference on Data Engineering (ICDE)*. Paris: IEEE, Apr. 2018, pp. 113–124. ISBN: 978-1-5386-5520-7. DOI: 10.1109/ICDE.2018.00020. URL: https://ieeexplore.ieee.org/document/8509241/ (visited on 09/03/2024).

[50] Mangesh Bendre, Tana Wattanawaroon, Sajjadur Rahman, Kelly Mack, Yuyang Liu, Shichu Zhu, Yu Lu, Ping-Jing Yang, Xinyan Zhou, Kevin Chen-Chuan Chang, Karrie Karahalios, and Aditya Parameswaran. “Faster, Higher, Stronger: Redesigning Spreadsheets for Scale”. In: *2019 IEEE 35th International Conference on Data Engineering (ICDE)*. Macao, Macao: IEEE, Apr. 2019, pp. 1972–

1975. ISBN: 978-1-5386-7474-1. DOI: 10.1109/ICDE.2019.00217. URL: https://ieeexplore.ieee.org/document/8731567/ (visited on 09/03/2024).

[51] Microsoft. *Learn to combine multiple data sources (Power Query)*. Accessed: 2024-08-26. 2024. URL: https://support.microsoft.com/en-us/office/learn-to-combine-multiple-data-sources-power-query-70cfe661-5a2a-4d9d-a4fe-586cc7878c7d (visited on 2024).

[52] Felienne Hermans, Bas Jansen, Sohon Roy, Efthimia Aivaloglou, Alaaeddin Swidan, and David Hoepelman. "Spreadsheets are Code: An Overview of Software Engineering Approaches Applied to Spreadsheets". In: *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*. Suita, Osaka, Japan: IEEE, Mar. 2016, pp. 56–65. ISBN: 978-1-5090-1855-0. DOI: 10.1109/SANER.2016.86. URL: http://ieeexplore.ieee.org/document/7476773/ (visited on 09/03/2024).

[53] Lyn Bartram, Michael Correll, and Melanie Tory. "Untidy Data: The Unreasonable Effectiveness of Tables". In: *IEEE Transactions on Visualization and Computer Graphics* 28.1 (Jan. 2022), pp. 686–696. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2021.3114830. URL: https://ieeexplore.ieee.org/document/9555457/ (visited on 09/03/2024).

[54] A. Streit, Binh Pham, and R. Brown. "A Spreadsheet Approach to Facilitate Visualization of Uncertainty in Information". In: *IEEE Transactions on Visualization and Computer Graphics* 14.1 (2008), pp. 61–72. ISSN: 1077-2626. DOI: 10.1109/TVCG.2007.70426. URL: http://ieeexplore.ieee.org/document/4359495/ (visited on 03/12/2020).

[55] R. A. Choudhury, W. F. Mahaffee, N. McRoberts, and W. D. Gubler. "Modeling Uncertainty in Grapevine Powdery Mildew Epidemiology Using Fuzzy Logic". en. In: *BioRxiv* 264622 (2018). DOI: 10.1101/264622. URL: http://biorxiv.org/lookup/doi/10.1101/264622 (visited on 04/24/2020).

[56] Google Developers. *Google Charts*. Accessed: 2024-08-26. 2024. URL: https://developers.google.com/chart/interactive/docs/spreadsheets (visited on 2024).

[57] Microsoft. *Excel specifications and limits*. Accessed: 2024-08-26. 2024. URL: https://support.microsoft.com/en-us/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3 (visited on 2024).

[58] Sean Kandel, Andreas Paepcke, Joseph M. Hellerstein, and Jeffrey Heer. "Enterprise Data Analysis and Visualization: An Interview Study". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2917–2926. DOI: 10.1109/TVCG.2012.219.

[59] Danyel Fisher, Rob DeLine, Mary Czerwinski, and Steven Drucker. "Interactions with big data analytics". In: *Interactions* 19.3 (2012), pp. 50–59. ISSN: 1072-5520. DOI: 10.1145/2168931.2168943. URL: https://dl.acm.org/doi/10.1145/2168931.2168943 (visited on 03/28/2023).

[60] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. "The emerging role of data scientists on software development teams". en. In: *International Conference on Software Engineering*. ACM, 2016, pp. 96–107. ISBN: 978-1-4503-3900-1. DOI: 10.1145/2884781.2884783. URL: https://dl.acm.org/doi/10.1145/2884781.2884783 (visited on 03/25/2023).

[61] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A. Hearst. "Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 22–31. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2865040. (Visited on 09/14/2020).

[62] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. "LineUp: Visual Analysis of Multi-Attribute Rankings". English. In: *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)* 19.12 (2013), pp. 2277–2286. DOI: 10.1109/TVCG.2013.173.

[63] Marc Streit, Samuel Gratzl, Holger Stitz, Andreas Wernitznig, Thomas Zichner, and Christian Haslinger. "Ordino: a visual cancer analysis tool for ranking and exploring genes, cell lines and tissue samples". en. In: *Bioinformatics* 35.17 (2019), pp. 3140–3142. DOI: 10.1093/bioinformatics/btz009. (Visited on 02/13/2019).

[64] Peter McLachlan, Tamara Munzner, Eleftherios Koutsofios, and Stephen North. "LiveRAC: interactive visual exploration of system management time-series data". en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Florence Italy: ACM, Apr. 2008, pp. 1483–1492. ISBN: 978-1-60558-011-1. DOI: 10.1145/1357054.1357286. URL: https://dl.acm.org/doi/10.1145/1357054.1357286 (visited on 08/27/2024).

[65] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. "Temporal event sequence simplification". eng. In: *IEEE transactions on visualization and computer graphics* 19.12 (Dec. 2013), pp. 2227–2236. ISSN: 1941-0506. DOI: 10.1109/TVCG.2013.200.

[66] Michael Sedlmair, Miriah Meyer, and Tamara Munzner. "Design Study Methodology: Reflections from the Trenches and the Stacks". In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (Dec. 2012), pp. 2431–2440. ISSN: 1077-2626. DOI: 10.1109/TVCG.2012.213. URL: http://ieeexplore.ieee.org/document/6327248/ (visited on 09/03/2024).

[67] Chris Stolte, Diane Tang, and Pat Hanrahan. "Polaris: A System for Query, Analysis, and Visualization of Multidimensional Relational Databases". In: *IEEE Transactions on Visualization and Computer Graphics* 8.1 (2002), pp. 52–65. DOI: 10.1109/2945.981851.

[68] Christopher Ahlberg. "Spotfire: An Information Exploration Environment". In: *ACM SIGMOD Record* 25.4 (1996), pp. 25–29. ISSN: 0163-5808. DOI: 10.1145/245882.245893. (Visited on 12/10/2017).

[69] S.G. Eick. "Visual discovery and analysis". In: *IEEE Transactions on Visualization and Computer Graphics* 6.1 (2000). ISSN: 1941-0506. DOI: 10.1109/2945.841120.

[70] Leishi Zhang, Andreas Stoffel, Michael Behrisch, Sebastian Mittelstadt, Tobias Schreck, Rene Pompl, Stefan Weber, Holger Last, and Daniel Keim. "Visual analytics for the big data era — A comparative review of state-of-the-art commercial systems". In: *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*. Seattle, WA, USA: IEEE, Oct. 2012, pp. 173–182. ISBN: 978-1-4673-4753-2. DOI: 10.1109/VAST.2012.6400554. URL: http://ieeexplore.ieee.org/document/6400554/ (visited on 09/03/2024).

[71] Michael Behrisch, Dirk Streeb, Florian Stoffel, Daniel Seebacher, Brian Matejek, Stefan Hagen Weber, Sebastian Mittelstadt, Hanspeter Pfister, and Daniel Keim. "Commercial Visual Analytics Systems–Advances in the Big Data Analytics Field". In: *IEEE Transactions on Visualization and Computer Graphics* 25.10 (Oct. 2019), pp. 3011–3031. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2859973. URL: https://ieeexplore.ieee.org/document/8423105/ (visited on 09/03/2024).

[72] Robert Kosara and Jock Mackinlay. "Storytelling: The Next Step for Visualization". In: *Computer* 46 (2013). Conference Name: Computer, pp. 44–50. ISSN: 1558-0814. DOI: 10.1109/MC.2013.36. URL: https://ieeexplore.ieee.org/document/6412677.

[73] Arjun Srinivasan, Steven M. Drucker, Alex Endert, and John Stasko. "Augmenting Visualizations with Interactive Data Facts to Facilitate Interpretation and Communication". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 672–681. ISSN: 1941-0506. DOI: 10.1109/TVCG.2018.2865145. URL: https://ieeexplore.ieee.org/abstract/document/8440860.

[74] Eric Mörth, Stefan Bruckner, and Noeska N. Smit. "ScrollyVis: Interactive Visual Authoring of Guided Dynamic Narratives for Scientific Scrollytelling". In: *IEEE Transactions on Visualization and Computer Graphics* (2022). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 1–12. ISSN: 1941-0506. DOI: 10.1109/TVCG.2022.3205769. URL: https://ieeexplore.ieee.org/abstract/document/9887905.

[75] Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. "Storytelling in Information Visualizations: Does it Engage Users to Explore Data?" In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, Apr. 2015, pp. 1449–1458. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702452.

[76] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. "Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences". en. In: *Computer Graphics Forum* 36.3 (2017). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf pp. 377–387. ISSN: 1467-8659. DOI: 10.1111/cgf.13195. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13195.

[77] Bahador Saket, Alex Endert, and John Stasko. "Beyond Usability and Performance: A Review of User Experience-focused Evaluations in Visualization". In: *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*. BELIV '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 133–142. ISBN: 978-1-4503-

4818-8. DOI: 10.1145/2993901.2993903. URL: https://dl.acm.org/doi/10.1145/2993901.2993903.

[78] Yang Shi, Xingyu Lan, Jingwen Li, Zhaorui Li, and Nan Cao. "Communicating with Motion: A Design Space for Animated Visual Narratives in Data Videos". In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. CHI '21. New York, NY, USA: Association for Computing Machinery, 2021, pp. 1–13. ISBN: 978-1-4503-8096-6. DOI: 10.1145/3411764.3445337. URL: https://dl.acm.org/doi/10.1145/3411764.3445337.

[79] Paul J. Zak. "Why Your Brain Loves Good Storytelling". In: *Harvard Business Review* (Oct. 2014). Section: Business communication. ISSN: 0017-8012. URL: https://hbr.org/2014/10/why-your-brain-loves-good-storytelling (visited on 10/11/2023).

[80] Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N Kostis. "Scientific Storytelling Using Visualization". In: *IEEE Computer Graphics and Applications* 32.1 (Jan. 2012), pp. 12–19. ISSN: 0272-1716. DOI: 10.1109/MCG.2012.24. URL: http://ieeexplore.ieee.org/document/6111347/ (visited on 10/11/2023).

[81] Humphrey O. Obie, Caslon Chua, Iman Avazpour, Mohamed Abdelrazek, John Grundy, and Tomasz Bednarz. "A study of the effects of narration on comprehension and memorability of visualisations". en. In: *Journal of Computer Languages* 52 (June 2019), pp. 113–124. ISSN: 25901184. DOI: 10.1016/j.cola.2019.04.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S1045926X19300059 (visited on 10/11/2023).

[82] Monique Meuschke, Laura A. Garrison, Noeska N. Smit, Benjamin Bach, Sarah Mittenentzwei, Veronika Weiß, Stefan Bruckner, Kai Lawonn, and Bernhard Preim. "Narrative medical visualization to communicate disease data". en. In: *Computers & Graphics* 107 (Oct. 2022), pp. 144–157. ISSN: 00978493. DOI: 10.1016/j.cag.2022.07.017. URL: https://linkinghub.elsevier.com/retrieve/pii/S009784932200139X (visited on 10/16/2023).

[83] Qianwen Wang, Zhen Li, Siwei Fu, Weiwei Cui, and Huamin Qu. "Narvis: Authoring Narrative Slideshows for Introducing Data Visualization Designs". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (Jan. 2019). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 779–788. ISSN: 1941-0506. DOI: 10.1109/TVCG.2018.2865232. URL: https://ieeexplore.ieee.org/abstract/document/8444072/.

[84] Jessica Hullman, Steven Drucker, Nathalie Henry Riche, Bongshin Lee, Danyel Fisher, and Eytan Adar. "A Deeper Understanding of Sequence in Narrative Visualization". en. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2406–2415. DOI: 10.1109/TVCG.2013.119.

[85] Faten El Outa, Patrick Marcel, Veronika Peralta, Raphaël da Silva, Marie Chagnoux, and Panos Vassiliadis. "Data Narrative Crafting via a Comprehensive and Well-Founded Process". In: *Advances in Databases and Information Systems*. Ed. by Silvia Chiusano, Tania Cerquitelli, and Robert Wrembel. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022, pp. 347–360. ISBN: 978-3-031-15740-0. DOI: 10.1007/978-3-031-15740-0_25.

[86] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. "Understanding Data Videos: Looking at Narrative Visualization through the Cinematography Lens". In: *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. CHI '15. New York, NY, USA: Association for Computing Machinery, Apr. 2015, pp. 1459–1468. ISBN: 978-1-4503-3145-6. DOI: 10.1145/2702123.2702431. URL: https://dl.acm.org/doi/10.1145/2702123.2702431.

[87] David Kouřil, Ondřej Strnad, Peter Mindek, Sarkis Halladjian, Tobias Isenberg, M. Eduard Gröller, and Ivan Viola. "Molecumentary: Adaptable Narrated Documentaries Using Molecular Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 29.3 (Mar. 2023). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 1733–1747. ISSN: 1941-0506. DOI: 10.1109/TVCG.2021.3130670. URL: https://ieeexplore.ieee.org/document/9627526.

[88] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. "Calliope: Automatic Visual Data Story Generation from a Spreadsheet". In: *IEEE Transactions on Visualization and Computer Graphics* 27 (2021). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 453–463. ISSN: 1941-0506. DOI: 10.1109/TVCG.2020.3030403. URL: https://ieeexplore.ieee.org/abstract/document/9222368.

[89] Britta Meixner, Katarzyna Matusik, Christoph Grill, and Harald Kosch. "Towards an easy to use authoring tool for interactive non-linear video". In: *Multimedia Tools and Applications* 70.2 (May 2014), pp. 1251–1276. DOI: 10.1007/s11042-012-1218-6.

[90] Shenyu Xu, Chris Bryan, Jianping Kelvin Li, Jian Zhao, and Kwan-Liu Ma. "Chart Constellations: Effective Chart Summarization for Collaborative and Multi-User Analyses". en. In: *Computer Graphics Forum* 37.3 (2018), pp. 75–86. DOI: 10.1111/cgf.13402.

[91] Minjeong Shin, Joohee Kim, Yunha Han, Lexing Xie, Mitchell Whitelaw, Bum Chul Kwon, Sungahn Ko, and Niklas Elmqvist. "Roslingifier: Semi-Automated Storytelling for Animated Scatterplots". In: *IEEE Transactions on Visualization and Computer Graphics* 29.6 (June 2023). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 2980–2995. ISSN: 1941-0506. DOI: 10.1109/TVCG.2022.3146329. URL: https://ieeexplore.ieee.org/document/9695173.

[92] Mengdi Sun, Ligan Cai, Weiwei Cui, Yanqiu Wu, Yang Shi, and Nan Cao. "Erato: Cooperative Data Story Editing via Fact Interpolation". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (Jan. 2023). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 983–993. ISSN: 1941-0506. DOI: 10.1109/TVCG.2022.3209428. URL: https://ieeexplore.ieee.org/abstract/document/9903555.

[93] Tan Tang, Junxiu Tang, Jiewen Lai, Lu Ying, Yingcai Wu, Lingyun Yu, and Peiran Ren. "SmartShots: An Optimization Approach for Generating Videos with Data Visualizations Embedded". In: *ACM Transactions on Interactive Intelligent*

Systems 12 (Mar. 2022), 4:1–4:21. ISSN: 2160-6455. DOI: 10.1145/3484506. URL: https://dl.acm.org/doi/10.1145/3484506.

[94] Davide Ceneda, Theresia Gschwandtner, Thorsten May, Silvia Miksch, Hans-Jorg Schulz, Marc Streit, and Christian Tominski. "Characterizing Guidance in Visual Analytics". In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (Jan. 2017), pp. 111–120. DOI: 10.1109/TVCG.2016.2598468.

[95] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. "Investigating Aspects of Data Visualization Literacy Using 20 Information Visualizations and 273 Science Museum Visitors". en. In: *Information Visualization* 15.3 (2016), pp. 198–213. ISSN: 1473-8716. DOI: 10.1177/1473871615594652. (Visited on 05/18/2017).

[96] Christina Stoiber, Florian Grassinger, Margit Pohl, Holger Stitz, Marc Streit, and Wolfgang Aigner. "Visualization Onboarding: Learning How to Read and Use Visualizations". In: *Proceedings of the Visualization for Communication Workshop at IEEE VIS*. 2019. DOI: 10.31219/osf.io/c38ab.

[97] Arthur Winzenried, Derek Law, Phillip Hughes, Doug Johnson, Sue Healey, David Warner, Katie Hannan, and Giuseppe Giovenco. "Towards an organisational theory for information professionals". en. In: *Visionary Leaders for Information*. Elsevier, 2010, pp. 23–61. ISBN: 978-1-876938-85-7. DOI: 10.1016/B978-1-876938-85-7.50002-6. URL: https://linkinghub.elsevier.com/retrieve/pii/B9781876938857500026 (visited on 09/03/2024).

[98] Howard J. Klein, Beth Polin, and Kyra Leigh Sutton. "Specific Onboarding Practices for the Socialization of New Employees". en. In: *International Journal of Selection and Assessment* 23.3 (2015), pp. 263–283. ISSN: 0965-075X, 1468-2389. DOI: 10.1111/ijsa.12113. URL: https://onlinelibrary.wiley.com/doi/10.1111/ijsa.12113 (visited on 09/03/2024).

[99] Hulick and Kunders. *User Onboarding*. Accessed: 2024-08-26. 2020. URL: https://www.useronboard.com/ (visited on 2020).

[100] Severino Ribecca. *The Data Visualisation Catalogue*. Last accessed on 2023/11/29. 2020. URL: https://datavizcatalogue.com/.

[101] Zezhong Wang, Lovisa Sundin, Dave Murray-Rust, and Benjamin Bach. "Cheat Sheets for Data Visualization Techniques". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2020. DOI: 10.1145/3313831.3376271.

[102] Christina Stoiber, Conny Walchshofer, Margit Pohl, Benjamin Potzmann, Florian Grassinger, Holger Stitz, Marc Streit, and Wolfgang Aigner. "Comparative evaluations of visualization onboarding methods". In: *Visual Informatics* 6.4 (Dec. 2022), pp. 34–50. DOI: 10.1016/j.visinf.2022.07.001.

[103] L. Bartram, M. Correll, and M. Tory. "Untidy Data: The Unreasonable Effectiveness of Tables". In: *IEEE Transactions on Visualization and Computer Graphics* 01 (2022), pp. 1–1. ISSN: 1941-0506. DOI: 10.1109/TVCG.2021.3114830.

[104] Andrea Kohlhase. “Human-Spreadsheet Interaction”. en. In: *Human-Computer Interaction – INTERACT 2013*. Ed. by Paula Kotzé, Gary Marsden, Gitte Lindgaard, Janet Wesson, and Marco Winckler. Lecture Notes in Computer Science. Springer, 2013, pp. 571–578. ISBN: 978-3-642-40498-6. DOI: 10.1007/978-3-642-40498-6_47.

[105] Carl V Phillips and Luwanna M LaPole. “Quantifying errors without random sampling”. en. In: *BMC Medical Research Methodology* 3.1 (2003), p. 9. ISSN: 1471-2288. DOI: 10.1186/1471-2288-3-9. URL: http://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-3-9 (visited on 04/27/2020).

[106] Microsoft Corporation. *Microsoft Excel*. https://www.microsoft.com/en-us/microsoft-365/excel/. Accessed: Sep. 1, 2021. 2020.

[107] Google LLC. *Google Sheets*. https://www.google.com/sheets/about/. Accessed: Sep. 1, 2021. 2020.

[108] The Apache Software Foundation. *Apache OpenOffice Calc*. https://www.openoffice.org/product/calc.html. Accessed: Sep. 1, 2021. 2021.

[109] Yingcai Wu, Guo-Xun Yuan, and Kwan-Liu Ma. “Visualizing Flow of Uncertainty through Analytical Processes”. In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2526–2535. ISSN: 1077-2626. DOI: 10.1109/TVCG.2012.285.

[110] Jessica Hullman. “Why Authors Don’t Visualize Uncertainty”. en. In: *IEEE Transactions on Visualization and Computer Graphics* 26.1 (2020), pp. 130–139. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2019.2934287. URL: https://ieeexplore.ieee.org/document/8805422/ (visited on 05/17/2021).

[111] Joke Bilcke, Philippe Beutels, Marc Brisson, and Mark Jit. “Accounting for Methodological, Structural, and Parameter Uncertainty in Decision-Analytic Models: A Practical Guide”. en. In: *Medical Decision Making* 31.4 (2011), pp. 675–692. ISSN: 0272-989X, 1552-681X. DOI: 10.1177/0272989X11409240. URL: http://journals.sagepub.com/doi/10.1177/0272989X11409240 (visited on 03/16/2020).

[112] Markus Knauff and Ann G. Wolf. “Complex cognition: the science of human reasoning, problem-solving, and decision-making”. en. In: *Cognitive Processing* 11.2 (2010), pp. 99–102. ISSN: 1612-4782, 1612-4790. DOI: 10.1007/s10339-010-0362-z. URL: http://link.springer.com/10.1007/s10339-010-0362-z (visited on 04/23/2020).

[113] Alex T. Pang, Craig M. Wittenbrink, and Suresh K. Lodh. “Approaches to Uncertainty Visualization”. In: *The Visual Computer* 13.8 (1997), pp. 370–390.

[114] Georges-Pierre Bonneau, Hans-Christian Hege, Chris R. Johnson, Manuel M. Oliveira, Kristin Potter, Penny Rheingans, and Thomas Schultz. “Overview and State-of-the-Art of Uncertainty Visualization”. In: *Scientific Visualization*. Ed. by Charles D. Hansen, Min Chen, Christopher R. Johnson, Arie E. Kaufman, and Hans Hagen. Series Title: Mathematics and Visualization. London: Springer,

2014, pp. 3–27. ISBN: 978-1-4471-6497-5. URL: http://link.springer.com/10.1007/978-1-4471-6497-5_1 (visited on 03/12/2020).

[115] Claes Lundström, Patric Ljung, Anders Persson, and Anders Ynnerman. "Uncertainty Visualization in Medical Volume Rendering Using Probabilistic Animation". In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1648–1655. ISSN: 1077-2626. DOI: 10.1109/TVCG.2007.70518. URL: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4376198 (visited on 07/07/2011).

[116] J Sanyal, Song Zhang, J Dyer, A Mercer, P Amburn, and R J Moorhead. "Noodles: A Tool for Visualization of Numerical Weather Model Ensemble Uncertainty". en. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1421–1430. ISSN: 1077-2626. DOI: 10.1109/TVCG.2010.181. URL: http://ieeexplore.ieee.org/document/5613483/ (visited on 04/27/2020).

[117] Zana Vosough, Dietrich Kammer, Mandy Keck, and Rainer Groh. "Visualizing uncertainty in flow diagrams: a case study in product costing". en. In: *Symposium on Visual Information Communication and Interaction (VINCI '17)*. ACM Press, 2017, pp. 1–8. ISBN: 978-1-4503-5292-5. DOI: 10.1145/3105971.3105972. URL: http://dl.acm.org/citation.cfm?doid=3105971.3105972 (visited on 04/24/2020).

[118] Tony Pitcher. "Beverton and Holt Equations: Spreadsheet functions and Uncertainty". In: *Naga, The ICLARM Quarterly* 22 (1 1999), pp. 37–41. URL: https://www.worldfishcenter.org/content/beverton-and-holt-equations-spreadsheet-functions-and-uncertainty-0.

[119] Avi Parush, Adi Hod, and Avy Shtub. "Impact of visualization type and contextual factors on performance with enterprise resource planning systems". en. In: *Computers & Industrial Engineering* 52.1 (2007), pp. 133–142. ISSN: 03608352. DOI: 10.1016/j.cie.2006.11.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0360835206001975 (visited on 04/23/2020).

[120] Kristin Potter, Paul Rosen, and Chris R. Johnson. "From Quantification to Visualization: A Taxonomy of Uncertainty Visualization Approaches". In: *Uncertainty Quantification in Scientific Computing*. London: Springer, 2012, pp. 226–249. ISBN: 978-3-642-32677-6. URL: http://link.springer.com/10.1007/978-3-642-32677-6_15.

[121] J. Sanyal, Song Zhang, G. Bhattacharya, P. Amburn, and R. Moorhead. "A User Study to Compare Four Uncertainty Visualization Methods for 1D and 2D Datasets". In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1209–1218. ISSN: 1077-2626. DOI: 10.1109/TVCG.2009.114. URL: http://ieeexplore.ieee.org/document/5290731/ (visited on 04/22/2020).

[122] Rita Borgo, Johannes Kehrer, David H. S. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications". en. In: *Eurographics 2013 - State of the Art Reports* (2012), pp. 39–63. ISSN: 1017-4656. DOI: 10.2312/CONF/EG2013/STARS/039-063. URL: http://diglib.eg.org/handle/10.2312/conf.EG2013.stars.039-063 (visited on 04/08/2020).

[123] D. Feng, L. Kwock, Yueh Lee, and R. M Taylor. “Matching Visual Saliency to Confidence in Plots of Uncertain Data”. English. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 980–989. ISSN: 1077-2626. DOI: 10.1109/TVCG.2010.176.

[124] Haidong Chen, Song Zhang, Wei Chen, Honghui Mei, Jiawei Zhang, Andrew Mercer, Ronghua Liang, and Huamin Qu. “Uncertainty-Aware Multidimensional Ensemble Data Visualization and Exploration”. en. In: *IEEE Transactions on Visualization and Computer Graphics* 21.9 (2015), pp. 1072–1086. ISSN: 1077-2626. DOI: 10.1109/TVCG.2015.2410278. URL: http://ieeexplore.ieee.org/document/7055260/ (visited on 04/24/2020).

[125] D. Spiegelhalter, M. Pearson, and I. Short. “Visualizing Uncertainty About the Future”. en. In: *Science* 333.6048 (2011), pp. 1393–1400. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1191181. URL: https://www.sciencemag.org/lookup/doi/10.1126/science.1191181 (visited on 04/23/2020).

[126] D. van der Laan, Edwin de Jonge, and Jessica Solcer. “Effect of Displaying Uncertainty in Line and Bar Charts – Presentation and Interpretation:” in: *Conference on Information Visualization Theory and Applications (IVAPP ’15)*. 2015, pp. 225–232. ISBN: 978-989-758-088-8. DOI: 10.5220/0005300702250232. URL: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0005300702250232 (visited on 04/23/2020).

[127] Michael Fernandes, Logan Walls, Sean Munson, Jessica Hullman, and Matthew Kay. “Uncertainty Displays Using Quantile Dotplots or CDFs Improve Transit Decision-Making”. en. In: *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI ’18)*. ACM Press, 2018, pp. 1–12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173718. URL: http://dl.acm.org/citation.cfm?doid=3173574.3173718 (visited on 04/24/2020).

[128] Jerry L. Hintze and Ray D. Nelson. “Violin Plots: A Box Plot-Density Trace Synergism”. en. In: *The American Statistician* 52.2 (1998), pp. 181–184. ISSN: 0003-1305, 1537-2731. DOI: 10.1080/00031305.1998.10480559. URL: http://www.tandfonline.com/doi/abs/10.1080/00031305.1998.10480559.

[129] Clemens Holzhüter, Alexander Lex, Dieter Schmalstieg, Hans-Jörg Schulz, Heidrun Schumann, and Marc Streit. “Visualizing Uncertainty in Biological Expression Data”. In: *Visualization and Data Analysis*. International Society for Optics and Photonics, 2012, 82940O. DOI: 10.1117/12.908516.

[130] D Fisher, S M Drucker, R Fernandez, and S Ruble. “Visualizations everywhere: A Multiplatform Infrastructure for Linked Visualizations”. en. In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1157–1163. ISSN: 1077-2626. DOI: 10.1109/TVCG.2010.222. URL: http://ieeexplore.ieee.org/document/5613454/ (visited on 05/17/2021).

[131] Charts API. *Google Charts*. https://developers.google.com/chart. Accessed: Sep. 1, 2021. 2019.

[132] Nick French and Laura Gabrielli. "Discounted cash flow: accounting for uncertainty". en. In: *Journal of Property Investment & Finance* 23.1 (2005). Ed. by Eamonn D'Arcy, pp. 75–89. ISSN: 1463-578X. DOI: 10.1108/14635780510575102. URL: https://www.emerald.com/insight/content/doi/10.1108/14635780510575102/full/html (visited on 03/16/2020).

[133] Daniel G Goldstein and David Rothschild. "Lay understanding of probability distributions". en. In: *Judgment and Decision Making* 9.1 (2014), p. 14.

[134] Matthew Brehmer and Tamara Munzner. "A Multi-Level Typology of Abstract Visualization Tasks". In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2376–2385. ISSN: 1077-2626. DOI: 10.1109/TVCG.2013.124. URL: http://ieeexplore.ieee.org/document/6634168/ (visited on 02/28/2020).

[135] R. Amar and J. Stasko. "A Knowledge Task-Based Framework for Design and Evaluation of Information Visualizations". In: *IEEE Symposium on Information Visualization*. 2004, pp. 143–150. DOI: 10.1109/INFVIS.2004.10.

[136] Carolina Nobre, Marc Streit, and Alexander Lex. "Juniper: A Tree+Table Approach to Multivariate Graph Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 544–554. DOI: 10.1109/TVCG.2018.2865149. (Visited on 04/11/2018).

[137] Katharina Furmanova, Samuel Gratzl, Holger Stitz, Thomas Zichner, Mirsolava Jaresova, Martin Ennemoser, Alexander Lex, and Marc Streit. "Taggle: Scalable Visualization of Tabular Data through Aggregation". In: *Information Visualization* (2019). DOI: 10.1177/1473871619878085. URL: https://arxiv.org/abs/1712.05944.

[138] Dietmar Jannach, Thomas Schmitz, Birgit Hofer, and Franz Wotawa. "Avoiding, finding and fixing spreadsheet errors – A survey of automated approaches for spreadsheet QA". In: *Journal of Systems and Software* 94 (2014), pp. 129–150.

[139] Ying Chen and Hock Chuan Chan. "Visual Checking of Spreadsheets". en. In: *Proceedings of the European Spreadsheet Risks Interest Group 1st Annual Conference (EuSpRIG 2000),London, United Kingdom* (2000).

[140] Felienne Hermans, Martin Pinzger, and Arie van Deursen. "Supporting professional spreadsheet users by generating leveled dataflow diagrams". en. In: *Conference on Software Engineering (ICSE '11)*. ACM Press, 2011, p. 451. ISBN: 978-1-4503-0445-0. DOI: 10.1145/1985793.1985855. URL: http://portal.acm.org/citation.cfm?doid=1985793.1985855 (visited on 04/27/2020).

[141] Alexander Lex, Christian Partl, Denis Kalkofen, Marc Streit, Samuel Gratzl, Anne Mai Wassermann, Dieter Schmalstieg, and Hanspeter Pfister. "Entourage: Visualizing Relationships between Biological Pathways using Contextual Subsets". English. In: *IEEE Transactions on Visualization and Computer Graphics (InfoVis '13)* 19.12 (2013), pp. 2536–2545. DOI: 10.1109/TVCG.2013.154.

[142] Anna Gogolou, Theophanis Tsandilas, Themis Palpanas, and Anastasia Bezerianos. "Comparing Similarity Perception in Time Series Visualizations". en. In: *IEEE Transactions on Visualization and Computer Graphics (InfoVis '18)* 25.1 (2019), pp. 523–533. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2018.2865077. URL: https://ieeexplore.ieee.org/document/8440826/ (visited on 06/06/2019).

[143] Microsoft Corporation. *Understanding the Office JavaScript API*. https://docs.microsoft.com/en-us/office/dev/add-ins/develop/understanding-the-javascript-api-for-office/. Accessed: Sep. 1, 2021. 2020.

[144] Jacob Menick. *Sampling – JavaScript for sampling from a handful of discrete probability distributions*. https://github.com/jacobmenick/sampling. Accessed: Sep. 1, 2021. 2016.

[145] Matthew Williams and Trevor Norris. *jStat – JavaScript Statistical Library*. https://github.com/jstat/jstat. Accessed: Sep. 1, 2021. 2020.

[146] Jos de Jong. *math.js – An extensive math library for JavaScript and Node.js*. https://mathjs.org/. Accessed: Sep. 1, 2021. 2019. (Visited on 2021).

[147] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. "D3: Data-Driven Documents". In: *IEEE Transactions on Visualization and Computer Graphics (InfoVis '11)* 17.12 (2011), pp. 2301–2309. DOI: 10.1109/TVCG.2011.185.

[148] Fred G. W. C. Paas, Jeroen J. G. van Merriënboer, and Jos J. Adam. "Measurement of Cognitive Load in Instructional Research". en. In: *Perceptual and Motor Skills* 79.1 (1994), pp. 419–430. ISSN: 0031-5125, 1558-688X. DOI: 10.2466/pms.1994.79.1.419. URL: http://journals.sagepub.com/doi/10.2466/pms.1994.79.1.419 (visited on 04/19/2020).

[149] Patrick W. Jordan, ed. *Usability evaluation in industry*. CRC Press, 1996. ISBN: 978-0-7484-0314-1.

[150] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Eduard Gröller, Marc Streit. *Supplementary Material for Fuzzy Spreadsheets*. https://osf.io/8ypux/. Accessed: Sep. 1, 2021. 2021. DOI: 10.17605/OSF.IO/8YPUX.

[151] Zoom Video Communications, Inc. *Zoom*. https://zoom.us/. Accessed: Sep. 1, 2021. 2020.

[152] TeamViewer Germany GmbH. *TeamViewer*. https://www.teamviewer.com/. Accessed: Sep. 1, 2021. 2020.

[153] LimeSurvey GmbH. *Professionelle Online-Umfragen mit LimeSurvey*. https://www.limesurvey.org/. Accessed: Sep. 1, 2021. 2020.

[154] Wolfram Alpha LLC. *WolframAlpha*. https://www.wolframalpha.com/. Accessed: Sep. 1, 2021. 2021.

[155] Stephen Few. *There's Nothing Mere About Semantics*. Accessed: 2021-12-2. 2021. URL: https://www.perceptualedge.com/blog/?p=2793 (visited on 12/13/2017).

[156] Lauren. *The 34 Best Interactive Data Visualizations from the New York Times*. Accessed: 2021-11-24. 2018. URL: https://getdolphins.com/blog/interactive-data-visualizations-new-york-times/ (visited on 02/28/2018).

[157] Mohan S Gounder, Vani Vasudevan Iyer, and Abdulaziz Al Mazyad. "A survey on business intelligence tools for university dashboard development". In: *MEC Conference on Big Data and Smart City (ICBDSC'16)*. IEEE, 2016, pp. 1–7. ISBN: 978-1-5090-1365-4. DOI: 10.1109/ICBDSC.2016.7460347. URL: https://ieeexplore.ieee.org/document/7460347/ (visited on 11/24/2021).

[158] Yuzuru Tanahashi, Nick Leaf, and Kwan-Liu Ma. "A Study On Designing Effective Introductory Materials for Information Visualization". In: *Computer Graphics Forum* 35.7 (Oct. 2016), pp. 117–126. DOI: 10.1111/cgf.13009.

[159] Christopher Collins, Natalia Andrienko, Tobias Schreck, Jing Yang, Jaegul Choo, Ulrich Engelke, Amit Jena, and Tim Dwyer. "Guidance in the human–machine analytics process". In: *Visual Informatics* 2.3 (2018), pp. 166–180. DOI: 10.1016/j.visinf.2018.09.003.

[160] Micheline Elias and Anastasia Bezerianos. "Annotating BI visualization dashboards: needs and challenges". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems* (2012), pp. 1641–1650. DOI: 10.1145/2207676.2208288.

[161] Garett Dworman and Stephanie Rosenbaum. "Helping users to use help: improving interaction with help systems". en. In: *Extended Abstracts of the Conference on Human Factors and Computing Systems (CHI '04)*. ACM, 2004, p. 1717. ISBN: 978-1-58113-703-3. DOI: 10.1145/985921.986198. URL: http://portal.acm.org/citation.cfm?doid=985921.986198 (visited on 07/09/2021).

[162] Erik Andersen, Eleanor O'Rourke, Yun-En Liu, Rich Snider, Jeff Lowdermilk, David Truong, Seth Cooper, and Zoran Popovic. "The impact of tutorials on games of varying complexity". en. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 2012, p. 59. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2207687. URL: http://dl.acm.org/citation.cfm?doid=2207676.2207687 (visited on 07/09/2021).

[163] Raphaël Morin, Pierre-Majorique Léger, Sylvain Senecal, Marie-Christine Bastarache-Roberge, Mario Lefèbrve, and Marc Fredette. "The Effect of Game Tutorial: A Comparison Between Casual and Hardcore Gamers". en. In: *Proceedings of the Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY '16)*. ACM, 2016, pp. 229–237. ISBN: 978-1-4503-4458-6. DOI: 10.1145/2968120.2987730. URL: https://dl.acm.org/doi/10.1145/2968120.2987730 (visited on 07/09/2021).

[164] Wei Li, Tovi Grossman, and George Fitzmaurice. "GamiCAD: a gamified tutorial system for first time autocad users". en. In: *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, 2012, p. 103. ISBN: 978-1-4503-1580-7. DOI: 10.1145/2380116.2380131. URL: http://dl.acm.org/citation.cfm?doid=2380116.2380131 (visited on 07/09/2021).

[165] Yan Holtz and Conor Healy. *From Data to Viz*. From data to Viz | Find the graphic you need. Library Catalog: www.data-to-viz.com. URL: data-to-viz.com (visited on 03/12/2020).

[166] Severino Ribecca. *The Data Visualisation Catalogue*. Accessed: 2021-11-19. 2020. URL: https://datavizcatalogue.com/ (visited on 11/19/2021).

[167] Hyunmo Kang, Catherine Plaisant, and Ben Shneiderman. "New Approaches to Help Users Get Started with Visual Interfaces: Multi-Layered Interfaces and Integrated Initial Guidance". In: *Proceedings of the National Conference on Digital Government Research*. Digital Government Society of North America, 2003, pp. 1–6. DOI: 10.5555/1123196.1123269.

[168] Micheline Elias, Marie-Aude Aufaure, and Anastasia Bezerianos. "Storytelling in Visual Analytics Tools for Business Intelligence". In: *Human-Computer Interaction (INTERACT'13)*. Vol. 8119. Springer Berlin Heidelberg, 2013, pp. 280–297. ISBN: 978-3-642-40476-4. DOI: 10.1007/978-3-642-40477-1_18. URL: http://link.springer.com/10.1007/978-3-642-40477-1_18 (visited on 11/19/2021).

[169] Elif E. Firat, Alena Denisova, and Robert S. Laramee. "Treemap Literacy: A Classroom-Based Investigation". In: *Education Papers (EG'20)* (2020). ISBN: 9783038681021 Publisher: The Eurographics Association Version Number: 029-038, 10 pages. ISSN: 1017-4656. DOI: 10.2312/EGED.20201032. URL: https://diglib.eg.org/handle/10.2312/eged20201032 (visited on 11/18/2021).

[170] Bum Chul Kwon and Bongshin Lee. "A Comparative Evaluation on Online Learning Approaches using Parallel Coordinate Visualization". en. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, 2016, pp. 993–997. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858101. (Visited on 06/07/2018).

[171] Edward Segel and Jeffrey Heer. "Narrative Visualization: Telling Stories with Data". In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1139–1148. DOI: 10.1109/TVCG.2010.179.

[172] Qianwen Wang, Zhen Li, Siwei Fu, Weiwei Cui, and Huamin Qu. "Narvis: Authoring Narrative Slideshows for Introducing Data Visualization Designs". In: *IEEE Transactions on Visualization and Computer Graphics* 25.1 (2019), pp. 779–788. DOI: 10.1109/TVCG.2018.2865232.

[173] Feng M, Deng C, Peck E. M, and Harrison L. "HindSight: Encouraging Exploration through Direct Encoding of Personal Interaction History." In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), pp. 351–360. DOI: 10.1109/TVCG.2016.2599058.

[174] Yu-Rong Cao, Jia-Yu Pan, and Wen-Chieh Lin. "User-oriented Generation of Contextual Visualization Sequences". en. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI'20)*. ACM, 2020, pp. 1–8. ISBN: 978-1-4503-6819-3. DOI: 10.1145/3334480.3383057. URL: https://dl.acm.org/doi/10.1145/3334480.3383057 (visited on 11/11/2021).

[175] Danqing Shi, Fuling Sun, Xinyue Xu, Xingyu Lan, David Gotz, and Nan Cao. "AutoClips: An Automatic Approach to Video Generation from Data Facts". In: *Computer Graphics Forum* 40.3 (2021), pp. 495–505. DOI: 10.1111/cgf.14324.

[176] Andrea Vázquez-Ingelmo, Francisco J. García-Peñalvo, and Roberto Therón. "Capturing high-level requirements of information dashboards' components through meta-modeling". In: *Proceedings of the Conference on Technological Ecosystems for Enhancing Mulitculturality*. New York, NY, USA: ACM, 2019, pp. 815–821. ISBN: 978-1-4503-7191-9. DOI: 10.1145/3362789.3362837.

[177] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. “Vega-Lite: A Grammar of Interactive Graphics”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), pp. 341–350. DOI: 10.1109/TVCG.2016.2599030.

[178] Paolo Federico, Markus Wagner, Alexander Rind, Albert Amor-Amorós, Silvia Miksch, and Wolfgang Aigner. “The Role of Explicit Knowledge: A Conceptual Model of Knowledge-Assisted Visual Analytics”. In: *IEEE Conference on Visual Analytics Science and Technology (VAST’17)*. IEEE, 2017, pp. 92–103. DOI: 10.1109/VAST.2017.8585498. URL: https://publik.tuwien.ac.at/files/publik_261674.pdf.

[179] Chris Bryan, Kwan-Liu Ma, and Jonathan Woodring. “Temporal Summary Images: An Approach to Narrative Visualization via Interactive Annotation Generation and Placement”. In: *IEEE Transactions on Visualization and Computer Graphics* 23.1 (2017), pp. 511–520. DOI: 10.1109/TVCG.2016.2598876.

[180] Micheline Elias and Anastasia Bezerianos. “Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices”. In: *Human-Computer Interaction (INTERACT’11)*. Vol. 6949. Springer Berlin Heidelberg, 2011, pp. 274–291. ISBN: 978-3-642-23767-6. DOI: 10.1007/978-3-642-23768-3_23. URL: http://link.springer.com/10.1007/978-3-642-23768-3_23 (visited on 11/19/2021).

[181] J Andreeva, I Dzhunov, E Karavakis, L Kokoszkiewicz, M Nowotka, P Saiz, and D Tuckett. “Designing and developing portable large-scale JavaScript web applications within the Experiment Dashboard framework”. en. In: *Journal of Physics: Conference Series* 396.5 (2012), p. 052069. ISSN: 1742-6588, 1742-6596. DOI: 10.1088/1742-6596/396/5/052069. URL: https://iopscience.iop.org/article/10.1088/1742-6596/396/5/052069 (visited on 11/19/2021).

[182] CA: Peachpit Press Berkeley. *Adobe Photoshop CS*. Accessed: 2021-11-22. 2004. (Visited on 11/22/2021).

[183] Vaishali Dhanoa, Conny Walchshofer, Andreas Hinterreiter, Holger Stitz, Eduard Groeller, and Marc Streit. *Supplementary Material for Dashboard Onboarding*. Accessed: 2021-12-2. 2021. URL: https://osf.io/zsb5a/ (visited on 12/02/2021).

[184] Maggie Sparkman, Beth Harvey, David Iseminger, and Kent Sharkey. *Human Resources*. Accessed: 2021-11-22. 2021. URL: https://docs.microsoft.com/en-us/power-bi/create-reports/sample-human-resources (visited on 11/22/2021).

[185] P. Pirolli and S. Card. “The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis”. In: *Proceedings of Conference on Intelligence Analysis*. 2005. URL: https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf.

[186] Qian Ma and Barbara Millet. “Design Guidelines for Immersive Dashboards”. en. In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 65.1 (2021), pp. 1524–1528. ISSN: 2169-5067, 1071-1813. DOI: 10.1177/1071181321651177. URL: http://journals.sagepub.com/doi/10.1177/1071181321651177 (visited on 11/19/2021).

[187] R. Amar, J. Eagan, and J. Stasko. "Low-level components of analytic activity in information visualization". In: *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '05)*. IEEE, 2005, pp. 111–117. DOI: 10.1109/INFVIS.2005.1532136.

[188] Data Bear Power BI Training. *Sales Analysis*. Accessed: 2021-11-22. 2021. URL: https://www.youtube.com/watch?v=XceUqw7kVmc&t=198s&ab_channel=DataBearPowerBITraining (visited on 08/09/2017).

[189] IPCC-WG1. *IPCC WGI Interactive Atlas: Regional information*. Accessed: 2021-11-22. 2021. URL: https://interactive-atlas.ipcc.ch/regional-information (visited on 11/22/2021).

[190] IPCC-WG1. *Atlas Data Sources*. Accessed: 2021-11-22. 2021. URL: https://github.com/IPCC-WG1/Atlas/tree/main/data-sources (visited on 11/22/2021).

[191] Zening Qu, Fan Du, Ryan A. Rossi, and Bill Howe. "Aunt Lily Can Say Her Visualizations". en. In: *Proceedings of the Conference of the North (2019)*. Association for Computational Linguistics, 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423. URL: http://aclweb.org/anthology/N19-1423 (visited on 11/25/2021).

[192] Gerwald Tschinkel, Robert Hafner, Peter Hasitschka, and Vedran Sabol. "Using Micro-Visualisations to Support Faceted Filtering of Recommender Results". In: *Conference of Information Visualisation (IV'16)*. IEEE, 2016, pp. 318–323. ISBN: 978-1-4673-8942-6. DOI: 10.1109/IV.2016.77. URL: http://ieeexplore.ieee.org/document/7557946/ (visited on 07/09/2021).

[193] Mona Berdugo, Kesem Sharabi, Alex Buck, Kent Sharkey, and Michelrj. *Power BI Embedded Analytics*. Accessed: 2021-11-22. 2021. URL: https://docs.microsoft.com/en-us/power-bi/developer/embedded/embedded-analytics-power-bi (visited on 10/19/2021).

[194] Microsoft Power BI. *Power BI embedded analytics Client APIs*. Accessed: 2021-11-22. 2021. URL: https://docs.microsoft.com/en-us/javascript/api/overview/powerbi/ (visited on 2021).

[195] Tableau. *Tableau*. Accessed: 2021-11-22. 2021. URL: https://www.tableau.com/ (visited on 11/29/2021).

[196] Krist Wongsuphasawat. "Encodable: Configurable Grammar for Visualization Components". In: *Proceedings of the IEEE Visualization Conference*. Los Alamitos, CA, USA: IEEE, 2020, pp. 131–135. DOI: 10.1109/VIS47514.2020.00033.

[197] Pramod Chundury, Mehmet Adil Yalçin, Jonathan Crabtree, Anup Mahurkar, Lisa M. Shulman, and Niklas Elmqvist. "Contextual in situ help for visual data interfaces". In: *Information Visualization* 22.1 (2023), pp. 69–84. DOI: 10.1177/14738716221120064.

[198] Microsoft Corporation. *Microsoft Power BI*. Last accessed on 2023/11/29. 2023. URL: https://powerbi.microsoft.com/ (visited on 11/29/2023).

[199] Sylvain Delisle and Bernard Moulin. "User Interfaces and Help Systems: From Helplessness to Intelligent Assistance". In: *Artificial Intelligence Review* 18.2 (2002), pp. 117–157. DOI: 10.1023/A:1015179704819.

[200] Arjun Srinivasan, Tim Harshbarger, Darrell Hilliker, and Jennifer Mankoff. "Azimuth: Designing Accessible Dashboards for Screen Reader Users". In: *Proceedings of the ACM Conference on Computers and Accessibility*. New York, NY, USA: ACM, Oct. 2023, pp. 1–16. ISBN: 9798400702204. DOI: 10.1145/3597638.3608405.

[201] David Kouřil, Ondřej Strnad, Peter Mindek, Sarkis Halladjian, Tobias Isenberg, M. Eduard Gröller, and Ivan Viola. "Molecumentary: Adaptable Narrated Documentaries Using Molecular Visualization". In: *IEEE Transactions on Visualization and Computer Graphics* 29.3 (2023), pp. 1733–1747. DOI: 10.1109/TVCG.2021.3130670.

[202] Minjeong Shin, Joohee Kim, Yunha Han, Lexing Xie, Mitchell Whitelaw, Bum Chul Kwon, Sungahn Ko, and Niklas Elmqvist. "Roslingifier: Semi-Automated Storytelling for Animated Scatterplots". In: *IEEE Transactions on Visualization and Computer Graphics* 29.6 (2023), pp. 2980–2995. DOI: 10.1109/TVCG.2022.3146329.

[203] Mengdi Sun, Ligan Cai, Weiwei Cui, Yanqiu Wu, Yang Shi, and Nan Cao. "Erato: Cooperative Data Story Editing via Fact Interpolation". In: *IEEE Transactions on Visualization and Computer Graphics* 29.1 (2023), pp. 983–993. DOI: 10.1109/TVCG.2022.3209428.

[204] Eric Mörth, Stefan Bruckner, and Noeska N. Smit. "ScrollyVis: Interactive Visual Authoring of Guided Dynamic Narratives for Scientific Scrollytelling". In: *IEEE Transactions on Visualization and Computer Graphics* 29.12 (2023), pp. 5165–5177. DOI: 10.1109/TVCG.2022.3205769.

[205] Zhenpeng Zhao and Niklas Elmqvist. "The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization". In: *IEEE Computer Graphics and Applications* 43.4 (2023), pp. 97–110. DOI: 10.1109/MCG.2023.3269850.

[206] Nam Wook Kim, Benjamin Bach, Hyejin Im, Sasha Schriber, Markus H. Gross, and Hanspeter Pfister. "Visualizing Nonlinear Narratives with Story Curves". In: *IEEE Transactions on Visualization and Computer Graphics* 24.1 (2018), pp. 595–604. DOI: 10.1109/TVCG.2017.2744118.

[207] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale, eds. *Data-Driven Storytelling*. eng. A K Peters visualization series. Boca Raton, FL, USA: CRC Press, 2018.

[208] Rob Munday. *A Guide to Interactive Documentary: Structure, Tools & Narrative*. Last accessed on 2023/11/29. Aug. 8, 2016. URL: https://directorsnotes.com/2016/08/08/interactive-documentary-guide/ (visited on 11/29/2023).

[209] Ben McIntosh, Randi Cohn, and Lindsay D. Grace. *Nonlinear Narrative in Games: Theory and Practice*. Game Careerguide. Last accessed via they Wayback machine at https://web.archive.org/web/20150406221523/http://www.gamecareerguide.com/features/882/features/882/nonlinear_narrative_in_games_.php on 2023/11/30. 2010. URL: http://www.gamecareerguide.com/features/882/features/882/nonlinear_narrative_in_games_.php.

[210] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. "How Does Automation Shape the Process of Narrative Visualization: A Survey on Tools". In: *IEEE Transactions on Visualization and Computer Graphics* (2023). DOI: 10.1109/TVCG.2023.3261320.

[211] Jessica Hullman, Robert Kosara, and Heidi Lam. "Finding a Clear Path: Structuring Strategies for Visualization Sequences". In: *Computer Graphics Forum* 36.3 (2017), pp. 365–375. DOI: 10.1111/cgf.13194.

[212] Younghoon Kim, Kanit Wongsuphasawat, Jessica Hullman, and Jeffrey Heer. "GraphScape: A Model for Automated Reasoning about Visualization Similarity and Sequencing". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2017, pp. 2628–2638. DOI: 10.1145/3025453.3025866.

[213] Daniel Green, Charlie Hargood, and Fred Charles. "A Novel Design Pipeline for Authoring Tools". In: *Proceedings of the International Conference on Interactive Digital Storytelling*. Lecture Notes in Computer Science 12497 (2020), pp. 102–110. DOI: 10.1007/978-3-030-62516-0_9.

[214] Britta Meixner, Katarzyna Matusik, Christoph Grill, and Harald Kosch. "Towards an easy to use authoring tool for interactive non-linear video". In: *Multimedia Tools and Applications* (2014), pp. 1251–1276. DOI: 10.1007/s11042-012-1218-6.

[215] Daniel Green. "Novella: An Authoring Tool for Interactive Storytelling in Games". In: *Interactive Storytelling*. Vol. 11318. Lecture Notes in Computer Science. Cham: Springer Publishing, 2018, pp. 556–559. DOI: 10.1007/978-3-030-04028-4_66.

[216] Johannes Sorger, Peter Mindek, Peter Rautek, Eduard Gröller, Graham Johnson, and Ivan Viola. "Metamorphers: storytelling templates for illustrative animated transitions in molecular visualization". In: *Proceedings of the 33rd Spring Conference on Computer Graphics*. SCCG '17. Mikulov, Czech Republic: ACM, 2017. ISBN: 9781450351072. DOI: 10.1145/3154353.3154364. URL: https://doi.org/10.1145/3154353.3154364.

[217] Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. "ChartStory: Automated Partitioning, Layout, and Captioning of Charts into Comic-Style Narratives". In: *IEEE Transactions on Visualization and Computer Graphics* 29.2 (2023), pp. 1384–1399. DOI: 10.1109/TVCG.2021.3114211.

[218] Andreas Mathisen, Tom Horak, Clemens Nylandsted Klokmose, Kaj Grønbæk, and Niklas Elmqvist. "InsideInsights: Integrating Data-Driven Reporting in Collaborative Visual Analytics". In: *Computer Graphics Forum* 38.3 (2019), pp. 649–661. DOI: 10.1111/CGF.13717.

[219] Elmer Sandvad, Kaj Grønbæk, Lennert Sloth, and Jørgen Lindskov Knudsen. "A metro map metaphor for guided tours on the Web: the Webvise guided tour system". In: *Proceedings of the ACM World Wide Web Conference*. New York, NY, USA: ACM, 2001, pp. 326–333. DOI: 10.1145/371920.372079.

[220] Randall H. Trigg. "Guided Tours and Tabletops: Tools for Communicating in a Hypertext Environment". In: *ACM Transactions on Information Systems* 6.4 (1988), pp. 398–414. DOI: 10.1145/58566.59299.

[221] Robert Kosara. *Story Points in Tableau Software*. Keynote at Tableau Customer Conference. 2013.

[222] Tango Technology, Inc. *Tango*. Last accessed on 2023/11/29. 2023. URL: https://www.tango.us/ (visited on 11/29/2023).

[223] Colony Labs, Inc. *Scribe*. Last accessed on 2023/11/29. 2023. URL: https://scribehow.com/ (visited on 11/29/2023).

[224] Niklas Elmqvist, Mihail Eduard Tudoreanu, and Philippas Tsigas. "Evaluating motion constraints for 3D wayfinding in immersive and desktop virtual environments". In: *Proceedings of the ACM Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2008, pp. 1769–1778. DOI: 10.1145/1357054.1357330.

[225] Jeremy Boy, Ronald A. Rensink, Enrico Bertini, and Jean-Daniel Fekete. "A Principled Way of Assessing Visualization Literacy". In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1963–1972. DOI: 10.1109/TVCG.2014.2346984.

[226] Kesem Sharabi, Mahir Diab, laurent-mic, Owen Duncan, Jian Dong, and Sudeep Kumar. *Using the Power BI REST APIs*. Last accessed on 2023/11/29. Aug. 31, 2023. URL: https://learn.microsoft.com/en-us/rest/api/power-bi/ (visited on 11/29/2023).

[227] Microsoft Corporation. *Power BI Embedded*. Last accessed on 2023/11/29. 2023. URL: https://azure.microsoft.com/en-us/products/power-bi-embedded/ (visited on 11/29/2023).

[228] Vaishali Dhanoa, Andreas Hinterreiter, Vanessa Fediuk, Niklas Elmqvist, Eduard Groeller, and Marc Streit. *D-Tour GitHub Repository*. 2023. URL: https://github.com/jku-vds-lab/dashboard-onboarding.

[229] Meta Platforms, Inc. *React: The library for web and native user interfaces*. Last accessed on 2023/11/29. 2023. URL: https://react.dev/ (visited on 11/29/2023).

[230] Bootstrap Team. *Build fast, responsive sites with Bootstrap*. Last accessed on 2023/11/29. 2023. URL: https://getbootstrap.com/ (visited on 11/29/2023).

[231] Christopher, Hayleigh, John, Moritz, and Peter. *React Flow*. Last accessed on 2023/11/29. Aug. 8, 2019. URL: https://reactflow.dev/ (visited on 11/29/2023).

[232] Sebastián Ramírez. *Fast API*. Last accessed on 2023/11/29. Aug. 8, 2018. URL: https://fastapi.tiangolo.com/ (visited on 11/29/2023).

[233] Michaela Hart, Mohit Patel, Maggie Sparkman, Rolevass, Jodi Martis, William Anton Rohm, David Iseminger, Kent Sharkey, and Hemant Arya. *Create report bookmarks in Power BI to share insights and build stories*. Last accessed on 2023/11/29. June 27, 2023. URL: https://learn.microsoft.com/en-us/power-bi/create-reports/desktop-bookmarks (visited on 11/29/2023).

[234] Microsoft Corporation. *Microsoft AppSource*. Last accessed on 2023/11/29. 2023. URL: https://appsource.microsoft.com/en-us/ (visited on 11/29/2023).

[235] Zachary T Cutler, Kiran Gadhave, and Alexander Lex. "Trrack: A Library for Provenance Tracking in Web-Based Visualizations". In: *IEEE Visualization Conference (VIS)*. 2020, pp. 116–120. DOI: 10.1109/VIS47514.2020.00030.

[236] Qing Chen, Shixiong Cao, Jiazhe Wang, and Nan Cao. "How Does Automation Shape the Process of Narrative Visualization: A Survey of Tools". In: (2022). DOI: 10.48550/ARXIV.2206.12118. URL: https://arxiv.org/abs/2206.12118.

[237] Nathalie Henry Riche, Christophe Hurter, Nicholas Diakopoulos, and Sheelagh Carpendale, eds. *Data-Driven Storytelling*. AK Peters Visualization Series. Boca Raton, FL, USA: CRC Press, 2018.

[238] J. Thompson, Z. Liu, W. Li, and J. Stasko. "Understanding the Design Space and Authoring Paradigms for Animated Data Graphics". en. In: *Computer Graphics Forum* 39.3 (2020), pp. 207–218. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.13974. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.13974.

[239] Chao Tong, Richard Roberts, Rita Borgo, Sean Walton, Robert Laramee, Kodzo Wegba, Aidong Lu, Yun Wang, Huamin Qu, Qiong Luo, and Xiaojuan Ma. "Storytelling and Visualization: An Extended Survey". en. In: *Information* 9.3 (Mar. 2018), p. 65. ISSN: 2078-2489. DOI: 10.3390/info9030065. URL: http://www.mdpi.com/2078-2489/9/3/65 (visited on 10/11/2023).

[240] Fereshteh Amini, Nathalie Henry Riche, Bongshin Lee, Andres Monroy-Hernandez, and Pourang Irani. "Authoring Data-Driven Videos with DataClips". In: *IEEE Transactions on Visualization and Computer Graphics* 23 (Jan. 2017). Conference Name: IEEE Transactions on Visualization and Computer Graphics, pp. 501–510. ISSN: 1941-0506. DOI: 10.1109/TVCG.2016.2598647. URL: https://ieeexplore.ieee.org/abstract/document/7539370.

[241] Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. *Data Comics: Sequential Art for Data-Driven Storytelling*. Tech. rep. 15. Human-Computer Interaction Laboratory, University of Maryland, College Park, 2015.

[242] Pengkun Ren, Yi Wang, and Fan Zhao. "Re-understanding of data storytelling tools from a narrative perspective". In: *Visual Intelligence* 1 (June 2023), p. 11. ISSN: 2731-9008. DOI: 10.1007/s44267-023-00011-0. URL: https://link.springer.com/10.1007/s44267-023-00011-0.

[243] Zhenpeng Zhao and Niklas Elmqvist. "The Stories We Tell About Data: Surveying Data-Driven Storytelling Using Visualization". In: *IEEE Computer Graphics & Applications* 43.4 (2023), pp. 97–110. DOI: 10.1109/MCG.2023.3269850.

[244] Patrick Mair. *LLM Enhanced Dashboard Onboarding*. Accessed: 2024-08-26. 2024. URL: https://github.com/PatrickJKU/Dashboard-Onboarding (visited on 2024).

[245] Mohammed Saidul Islam, Md Tahmid Rahman Laskar, Md Rizwan Parvez, Enamul Hoque, and Shafiq Joty. *DataNarrative: Automated Data-Driven Storytelling with Visualizations and Texts*. Version Number: 2. 2024. DOI: 10.48550/ARXIV.2408.05346. URL: https://arxiv.org/abs/2408.05346 (visited on 08/26/2024).

[246] Ben Shneiderman. *Human-centered AI*. eng. Oxford: Oxford University Press, 2022. ISBN: 978-0-19-284529-0.

[247] Fei-Fei Li. *How to Make A.I. That's Good for People*. Accessed: 2024-08-26. 2018. URL: https://www.nytimes.com/2018/03/07/opinion/artificial-intelligence-human.html (visited on 03/07/2018).

[248] Md. Naimul Hoque, Sungbok Shin, and Niklas Elmqvist. "Harder, Better, Faster, Stronger: Interactive Visualization for Human-Centered AI Tools". In: *ArXiv* abs/2404.02147 (2024). URL: https://api.semanticscholar.org/CorpusID:268857110.

[249] Open AI. *Assistants API Overview*. Accessed: 2024-08-26. 2024. URL: https://platform.openai.com/docs/assistants/overview (visited on 2024).

[250] Jiho Kim, Arjun Srinivasan, Nam Wook Kim, and Yea-Seul Kim. "Exploring Chart Question Answering for Blind and Low Vision Users". en. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Hamburg Germany: ACM, Apr. 2023, pp. 1–15. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581532. URL: https://dl.acm.org/doi/10.1145/3544548.3581532 (visited on 09/04/2024).

[251] Frank Elavsky, Cynthia Bennett, and Dominik Moritz. "How accessible is my visualization? Evaluating visualization accessibility with Chartability". en. In: *Computer Graphics Forum* 41.3 (June 2022), pp. 57–70. ISSN: 0167-7055, 1467-8659. DOI: 10.1111/cgf.14522. URL: https://onlinelibrary.wiley.com/doi/10.1111/cgf.14522 (visited on 09/04/2024).

[252] Md Naimul Hoque, Md Ehtesham-Ul-Haque, Niklas Elmqvist, and Syed Masum Billah. "Accessible Data Representation with Natural Sound". en. In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Hamburg Germany: ACM, Apr. 2023, pp. 1–19. ISBN: 978-1-4503-9421-5. DOI: 10.1145/3544548.3581087. URL: https://dl.acm.org/doi/10.1145/3544548.3581087 (visited on 09/04/2024).

[253] Pramod Chundury, Yasmin Reyazuddin, J. Bern Jordan, Jonathan Lazar, and Niklas Elmqvist. "TactualPlot: Spatializing Data as Sound using Sensory Substitution for Touchscreen Accessibility". In: *IEEE Transactions on Visualization and Computer Graphics* (2023), pp. 1–11. ISSN: 1077-2626, 1941-0506, 2160-9306. DOI: 10.1109/TVCG.2023.3326937. URL: https://ieeexplore.ieee.org/document/10308620/ (visited on 09/04/2024).

A A Process Model for Dashboard Onboarding Additional Information

Creation of the component graph

We hereby show the creation of our component graph with edges indicating the relationship between each component in a stepwise manner. We start by extracting all components with their low-level characteristics from the dashboard before showing the effect of filtering/highlighting by interacting with each component.

Extract all components

First, we extracted all components from our exemplary use case scenario using the "New Hires" dashboard provided by Microsoft Power BI. We have uploaded the interactive dashboard to our OSF project. The dashboard consists of six components: one global filter, two key performance indicators (KPI), two bar charts and one line chart, see Figure A.1.

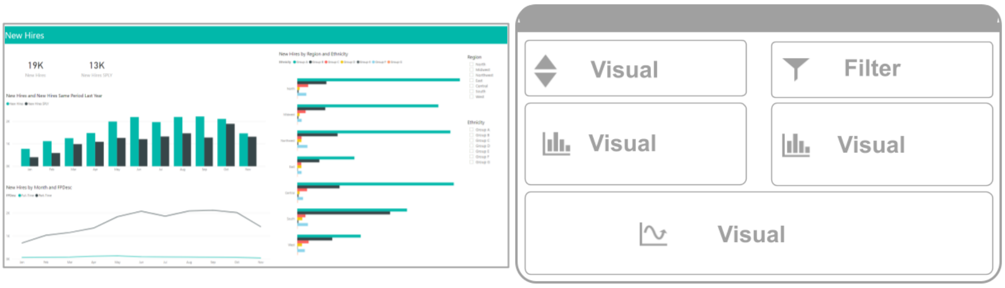


Fig. A.1: Example dashboard using the "new hires" data. (a) represents the actual interactive dashboard created and (b) the schematic representation highlighting the six components

As already outlined in Chapter 4, the six components with their low-level characteristics can be extracted first without any relationship, see Figure A.2.

Connect the components through edges at a high level We started by demonstrating the influence of each component on the others using edges connecting the components. We first connected the dashboard to all components as it includes the data. For the sake of completeness, we have added the influence of the dashboard on all the other components, however, we do not consider the dashboard as an actual influencing factor as the data is presumed to be provided in advance and cannot be changed by neither the author nor the user.

We continued by indicating the influence of filtering or highlighting effect in a generalized way without explicitly showing the effect on the low-level characteristics. As no interaction on the KPI visual was supported, no edges can be drawn from them to other visuals. However, the global filter, the bar chart, and the line chart affect almost all other visuals, see Figures A.4, A.5, A.6, and A.7

When combining all filter effects, we end up with the following component graph.

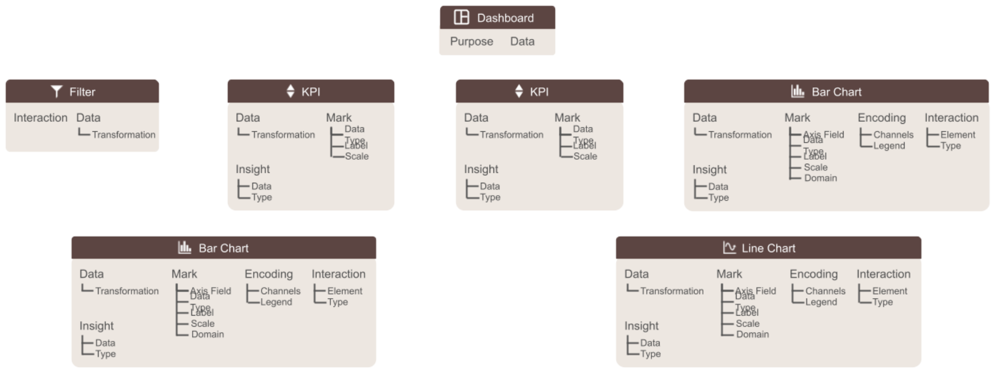


Fig. A.2: Component graph without relationships for the example dashboard

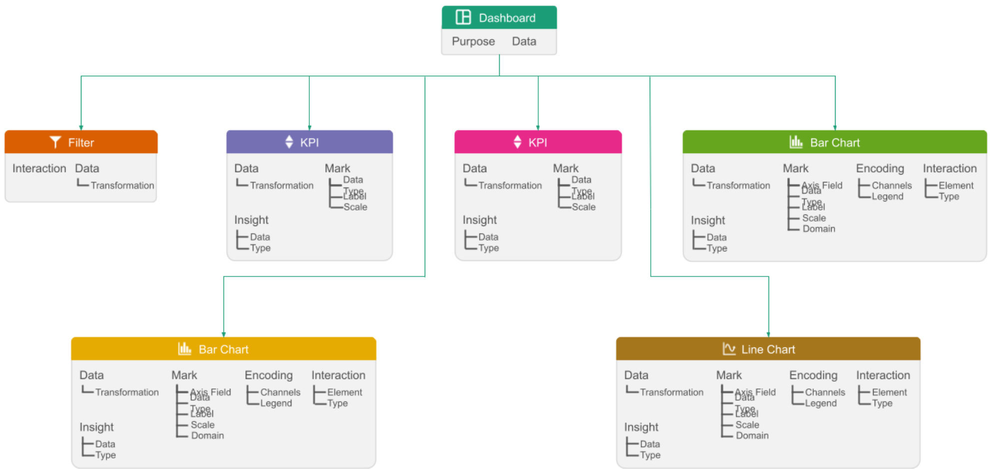


Fig. A.3: Influence of the dashboard (data) on the other visuals

Add influence on low-level characteristics This graph gives a good overview of how each component is related to each other. However, it does not yet reflect the effect of filter interaction on low-level components such as data, marks, or insight. Thus, we have revised the influences of each component and highlighted in the respective components' color the influence, see Figures A.9, A.10, A.11, and A.12.

Consequently, when combining all four models, the effect of each interaction becomes apparent; see Figure A.13. It has to be said that we have tried to keep the visual clutter at a minimum as this model is still a simplified version of the relational component graph. It highlights the effect of filtering by keeping the edges visible, and affected low-level characteristics are highlighted. Note: Non-affected low-level characteristics are displayed in smaller font and in light gray.

We extended our component graph with edges on the level of our low-level components. Thus, for instance, a selection of a check box in the filter visual always triggers a transformation of the data for the KPI visuals, which might change the type of insight that can be gained from the updated bar chart and might change the displayed scale and domain of the line chart. The effect of these interactions are reflected in the following

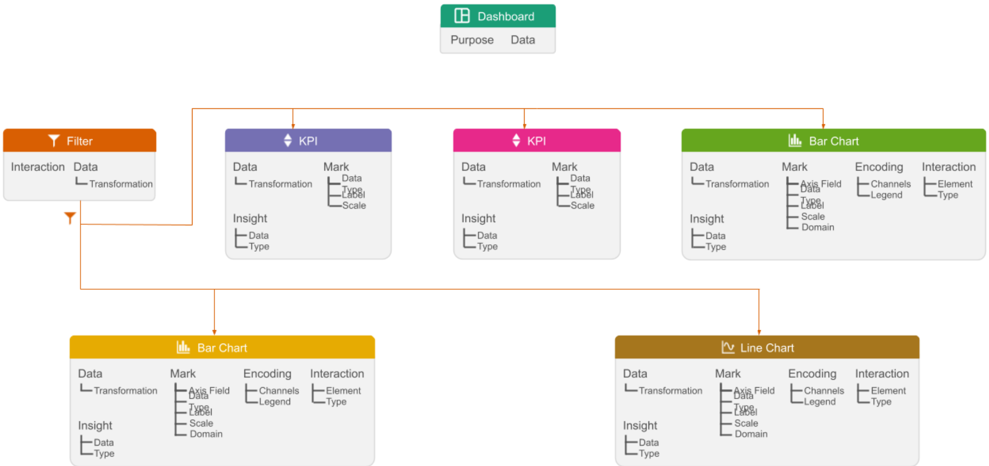


Fig. A.4: Influence of global filter

Figures A.14, A.15, A.16, and A.17 for the global filter, both bar charts and the line chart.

Combining all of these individual component graphs, results in the following Figure, where the effect of interaction for the filter, both bar charts and the line chart are highlighted

Thus, for our paper, we decided to reduce the visual clutter and therefore only show (1) the dashboard, (2) the component graph with its high-level connectors, and (3) an excerpt of a depth-first narrative in Figure A.19.

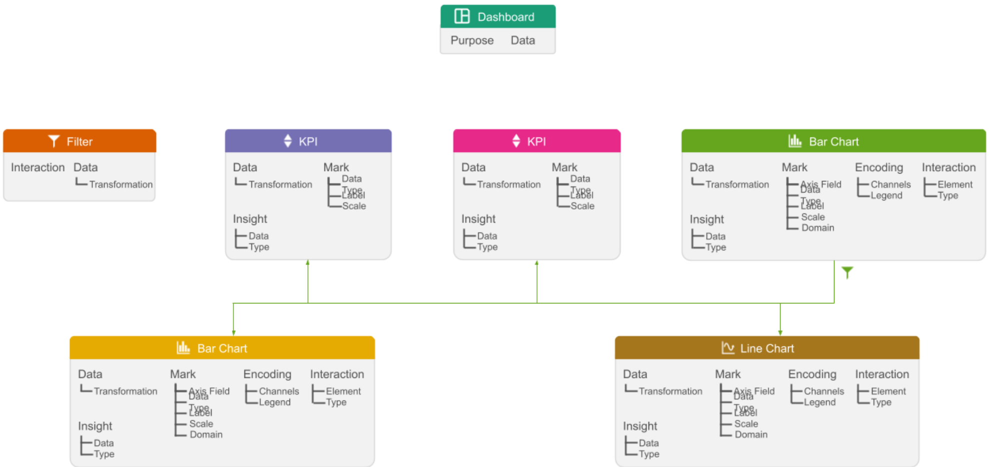


Fig. A.5: Influence of bar chart

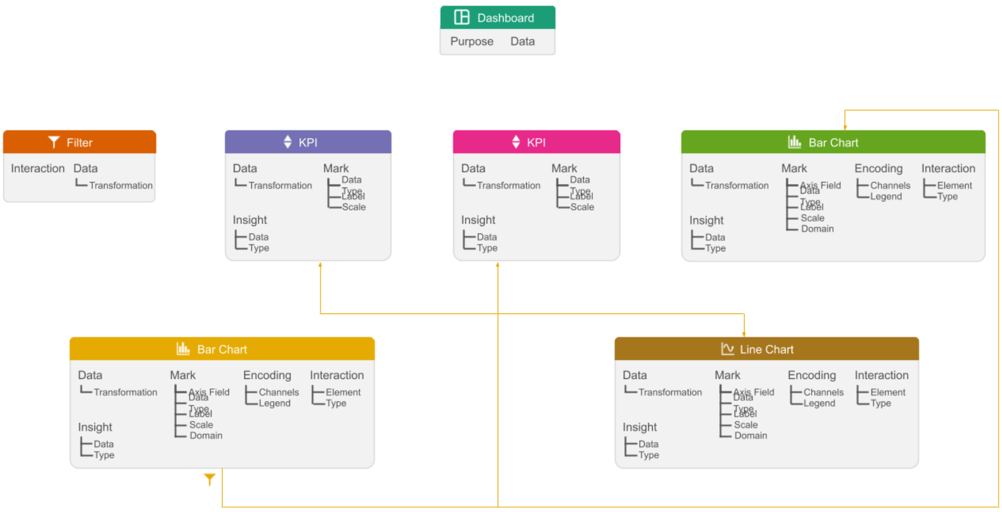


Fig. A.6: Influence of second bar chart

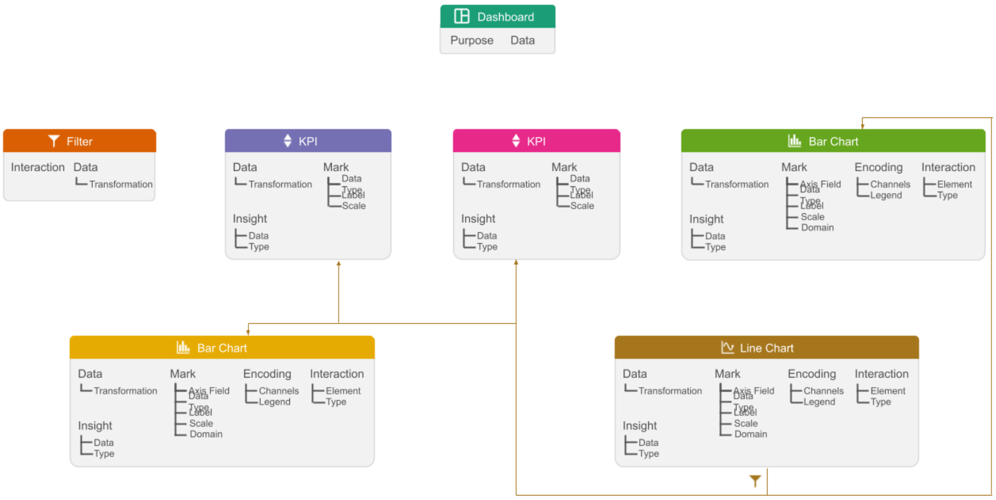


Fig. A.7: Influence of line chart

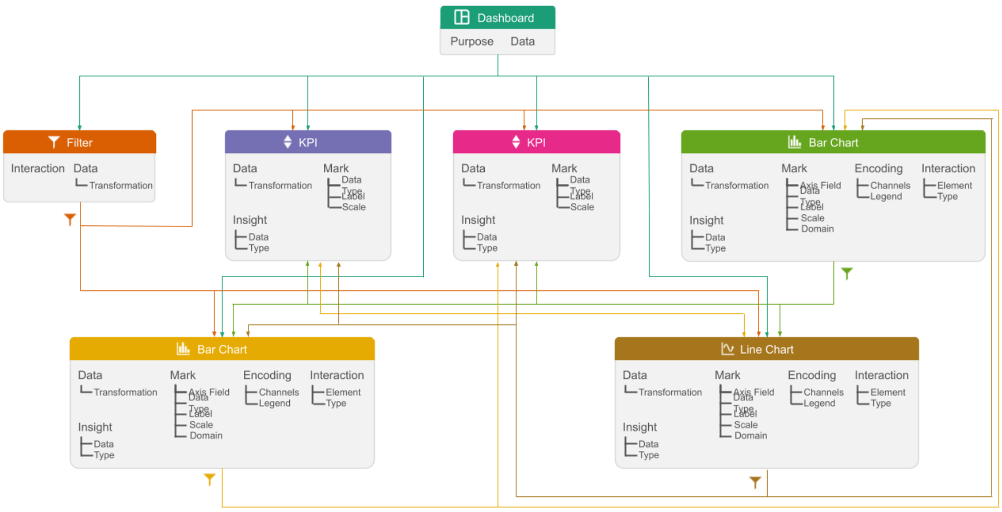


Fig. A.8: Relational component graph at a high level of detail

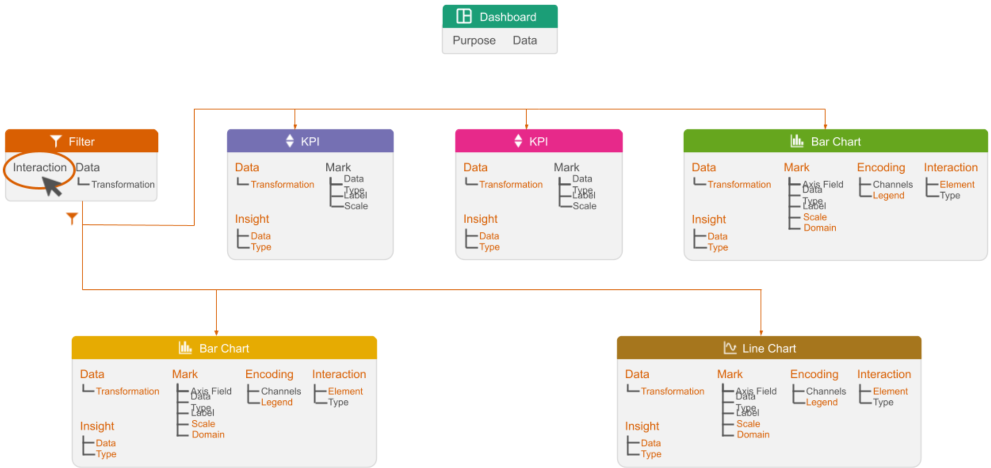


Fig. A.9: Influence of global filter

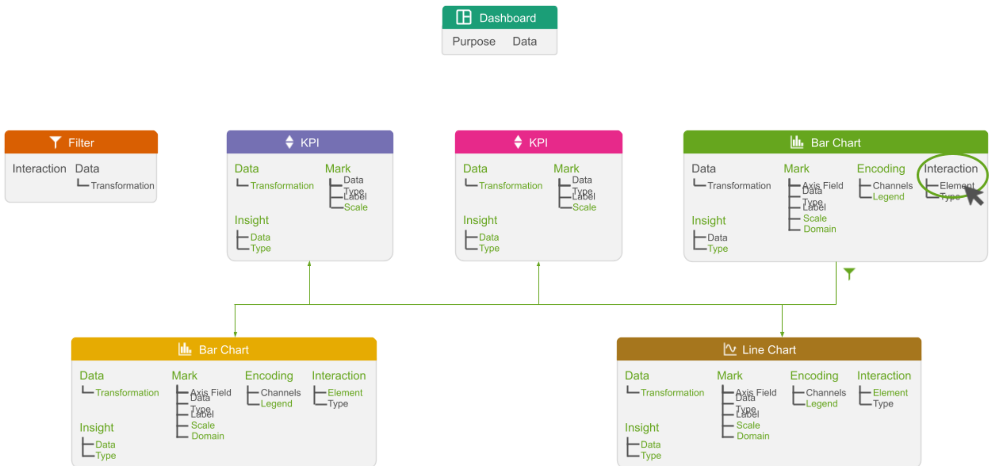


Fig. A.10: Influence of bar chart

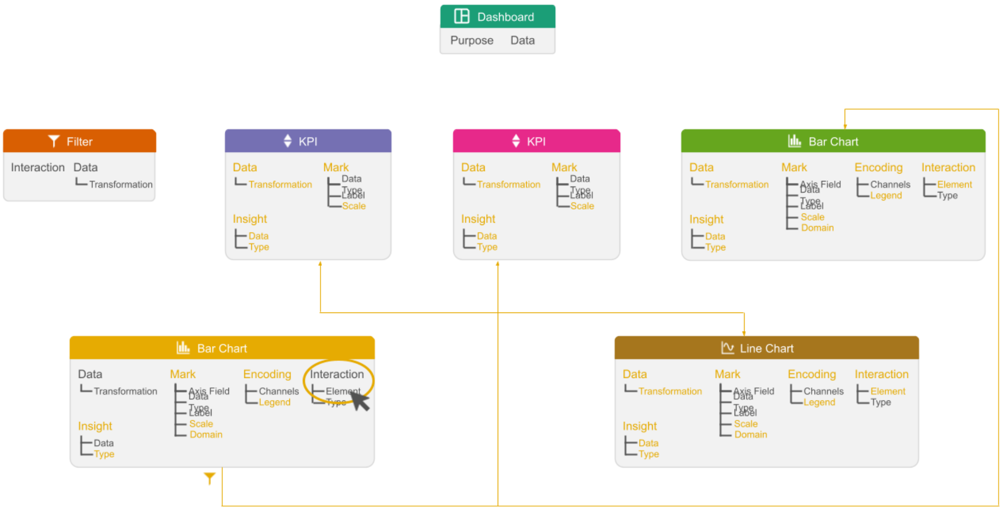


Fig. A.11: Influence of second bar chart

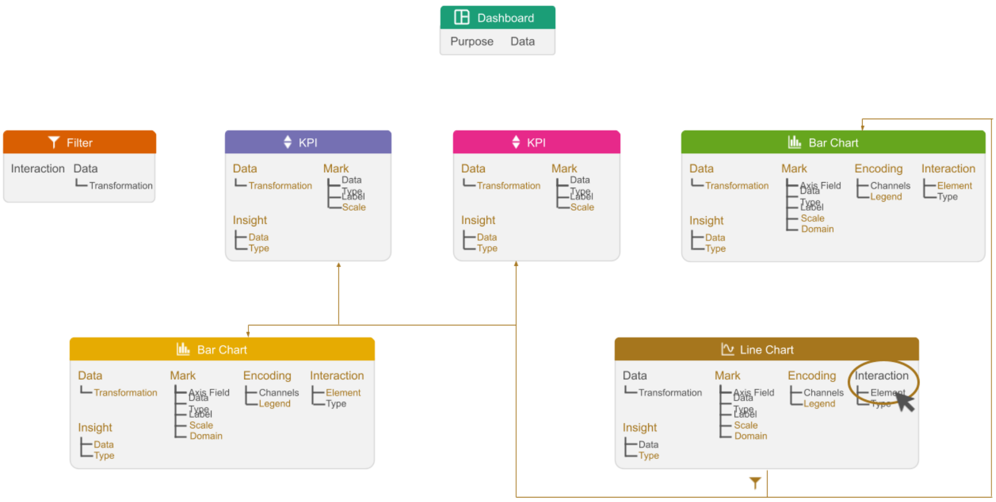


Fig. A.12: Influence of line chart

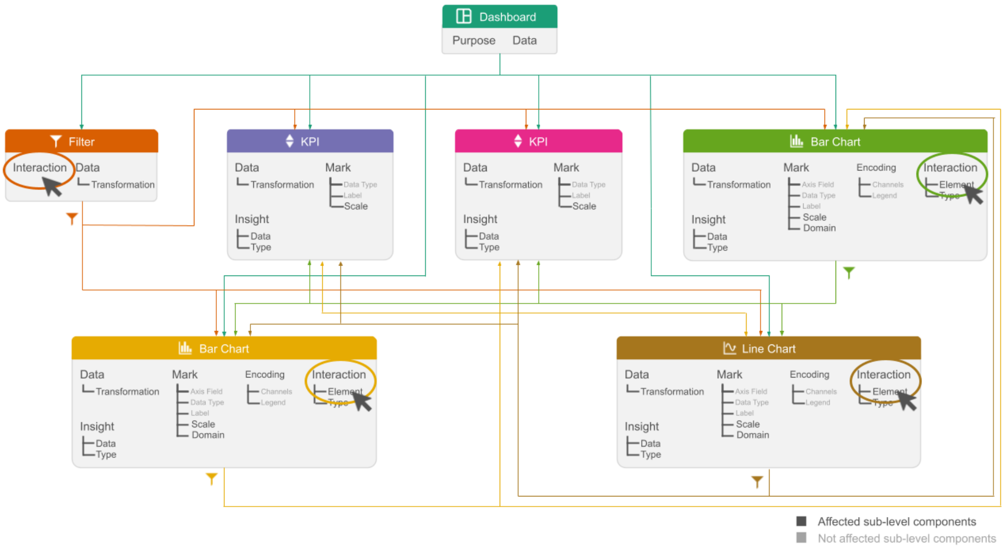


Fig. A.13: Component graph with edges going from one visual to the others. Additionally, the affected low-level characteristics are highlighted for each visual component.

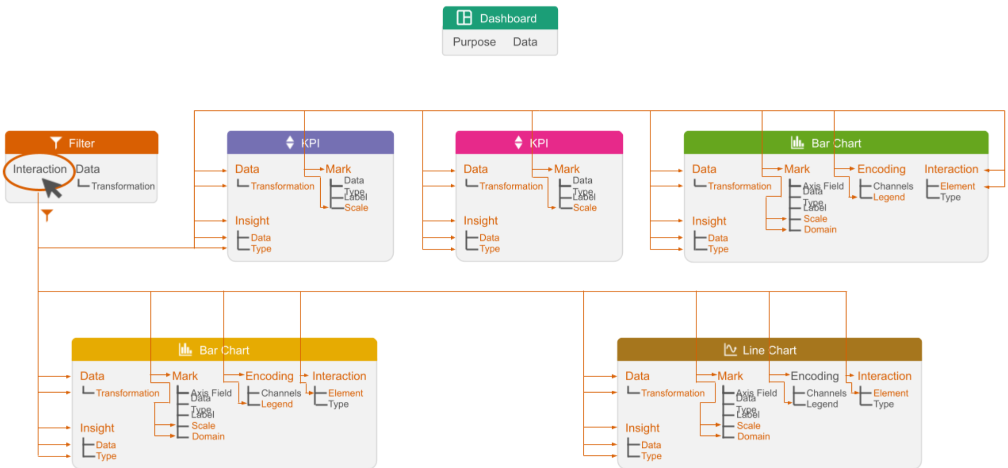


Fig. A.14: Influence of global filter

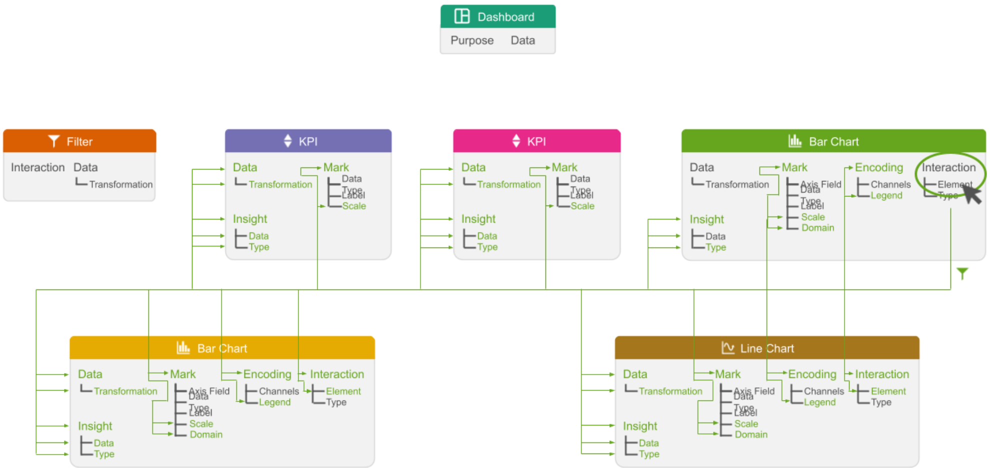


Fig. A.15: Influence of bar chart

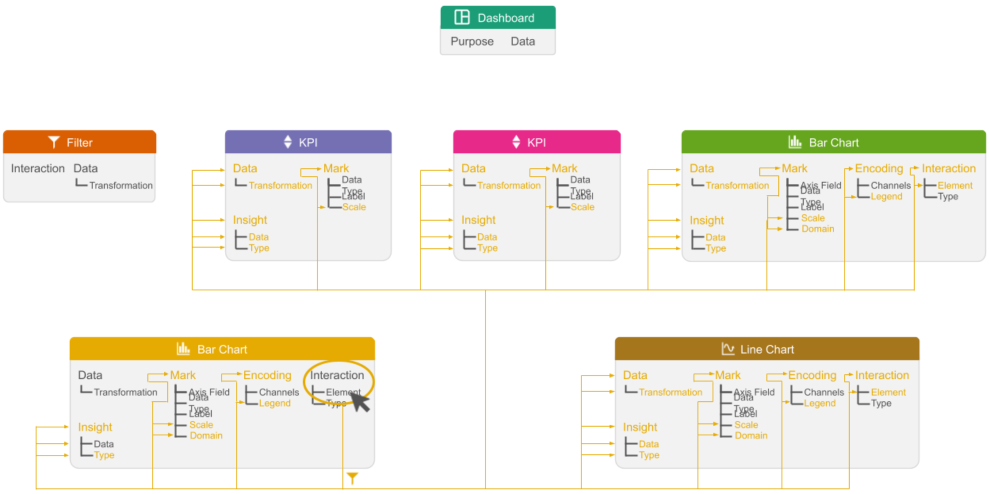


Fig. A.16: Influence of second bar chart

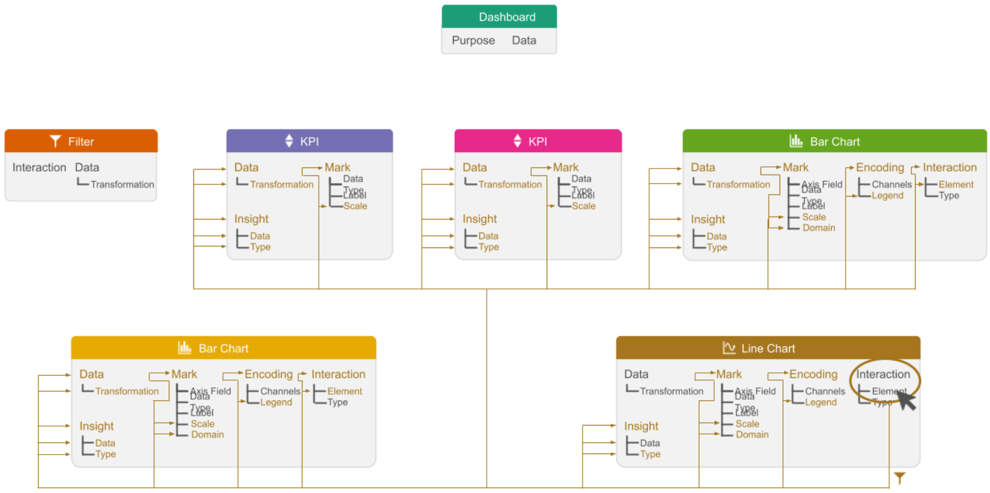


Fig. A.17: Influence of line chart

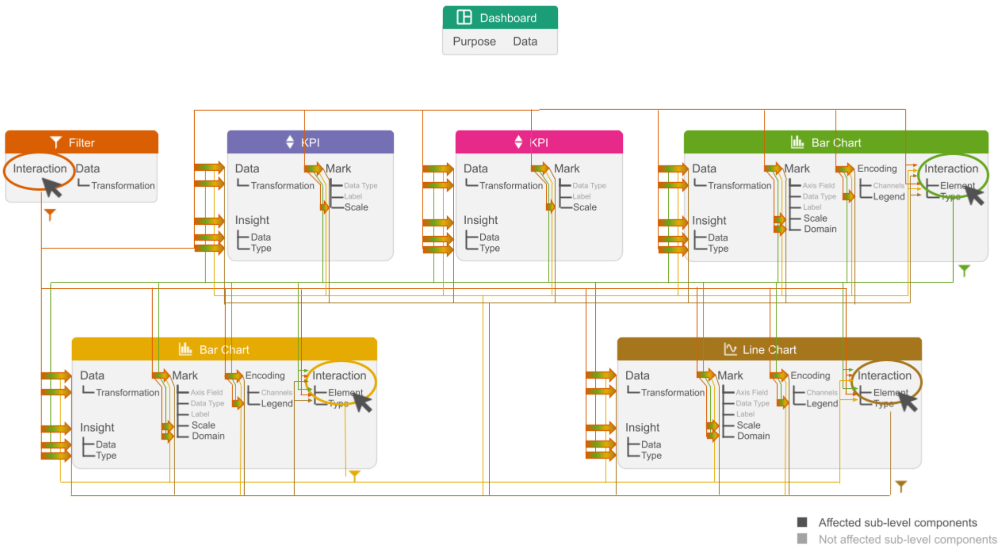


Fig. A.18: Combined Influences of all components on the level of low-level characteristics.

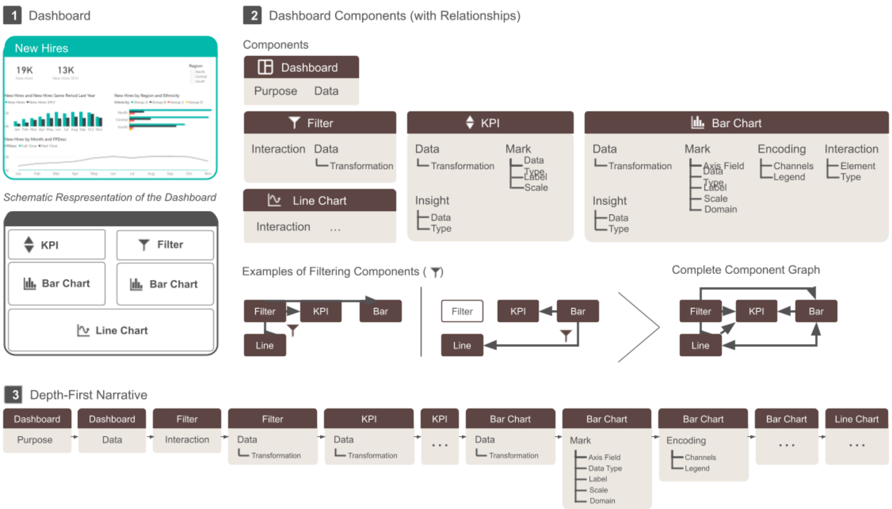


Fig. A.19: An excerpt of the complex component graph with relationship information.

B D-Tour Additional Information

B.1 Summary of the Expert Interview

Original interview: https://osf.io/v8n93.

We interviewed an expert with a Ph.D. in visualization to explore the challenges and the opportunities in onboarding from both author and user's perspectives. She has published several papers in the field of visualization onboarding, evaluated various onboarding methods, and worked extensively on onboarding v/s guidance.

The interview aimed to learn from her extensive experience in visualization onboarding, informed by her work and the workshops she has led. The objectives were to identify challenges, discuss commonalities between visualization and dashboard onboarding, recognize common pitfalls, and explore the potential of advancing the field of visualization and dashboard onboarding.

The interview lasted an hour and was conducted online using Zoom https://www.zoom.com/en/products/virtual-meetings/.

It followed a structured format with questions that were already prepared. The session was recorded and notes were taken, which are accessible at https://osf.io/v8n93.

B.1.1 Onboarding from author's perspective

Based on her workshop experience with data journalists, she observed that authors often face difficulties deciding whether to first describe the data behind a visualization or the visualization itself. In another workshop experience with biomedical researchers, she observed clear preferences for understanding the data and its preprocessing to enhance the analysis process. This experience highlights the importance of data analytics onboarding, a significant topic that helps expert users onboard themselves on data. She emphasized the importance of designing onboarding alongside the dashboard, so that users can onboard themselves while engaging with the dashboard, aligning with our previous work.. Similar to the collaborative dashboard design process, she suggested that onboarding should also be designed collaboratively with expert users before rolling it out to a larger audience. She also realized the need for multilingual onboarding material, similar to our findings from the user study. On the topic of automated onboarding, she recognized its effectiveness for specific tasks in customized domain-specific solutions but pointed its limitations (similar to Clippit: https://en.wikipedia.org/wiki/Office_Assistant). in context-sensitive assistance during the analysis phase. In such a situation, direct human interaction is often more effective, as it circumvents the need for "specific keywords" in addressing user queries.

B.1.2 Onboarding from user's perspective

The primary challenge users face when interacting with a dashboard is determining "where to begin" and identifying the "main visualization", and with "finding help when needed". In her experience, in-situ onboarding –whether through text, audio, or video– is highly effective. It should allow users the flexibility to access help as needed (similar to our design framework and observation). Although the effectiveness of an onboarding experience is tough to measure as "learning is personal", it should at least help the users in starting with the dashboard. Other effective measures could include helping the users accomplish their tasks and discover new insights, especially with complex visualizations.

B.2 Feedback workshop

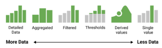
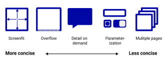
We gave a one-hour workshop presentation to 15 participants from our industrial collaborators who attended remotely, with the session recorded for later company-wide access. The workshop's purpose was to share the d-tour concept and the D-tour Prototype to gather feedback on our work and explore potential future improvements. The workshop participants regularly created dashboards using Microsoft Power BI [33] and onboarded users via personal or online meetings. Additionally, they also received onboarding on new dashboards. Consequently, they fulfilled the requisite criteria of being both the author and the user of dashboards with a distinct domain, data, and visualization expertise. Preparing and presenting at the workshop. We prepared a slide deck to explain d-tours and the D-Tour Prototype. We discussed challenges in dashboard onboarding, explored various onboarding methods, and presented the d-tour concept, explaining it from both the author's and the user's perspectives. We then played a video of an example dashboard, which was followed by a live demonstration of the D-Tour Prototype.

B.2.1 Conclusions from the workshop

We received positive feedback on this workshop, with participants understanding the need for the d-tour concept. They further discussed integrating the D-Tour Prototype into their dashboard environments. The online format of the workshop may have inhibited audience discussion as we only received a few questions during the workshop. However, after the interview, we received personal messages from participants and others as further people viewed the recorded presentation. The questions were both on a conceptual and technical level. One of the participants asked about the learning curve of the D-Tour Prototype and how much time they can save when creating multiple onboarding experiences for various dashboards. In our prior studies, it took about an hour, including explanations of features and usage, to create the first onboarding experience for a given dashboard. This time span is expected to decrease as authors continue to create more onboarding experiences, thereby becoming more familiar with the system. Another question was whether the d-tour could update automatically in response to changes in the dashboard content, such as a bar chart being replaced by a column chart. Although we do not currently support this level of automation, we recognize its potential and consider it a valuable direction for future development.

B.3 Learning from Dashboard Design Patterns

Bach et al. [17] introduced design patterns for dashboards. The D-Tour Prototype currently supports onboarding for dashboards that use many of these patterns. We plan to provide additional assistance for patterns like multi-page dashboards, parallel structures, and a wider range of interactions. A tabular summary of design patterns supported by the D-Tour Prototype can be found in Figure B.1.

Index	Figures	Dashboard Design	Support in D-Tour	Explanation
1		Data information	●	Most of the data information is covered in the subcategories General and Insight.
2		Meta information	● ⏳	Although some meta information is covered in insight, the rest has to be added by the author manually. In the future, we plan to include data source, disclaimer, annotations.
3		Visual representation	● ⏳	Currently, we support lists, tables, numbers and detailed visualizations. In the future, we plan to support miniature charts, progress bars, pictograms and trend arrows.
4		Interactions	● ⏳	The subcategory of Interact provides authors with automatic description of what filters show and how to use them. However, not all interaction types are supported, especially zoom and naivgate. We plan to cover these interactiopn types in the future, as they are commonly used in the dashboards. Also, we plan to support the validity of the applied filter.
5		Screenspace	● ⏳	We support screenfit, overflow and parameterized composition. The details on demand need to be manually adjusted for now. We plan to offer support for multiple pages, as it was one of the highly requested features during the user feedback session.
6		Page layout	● ⏳	The onboarding standard story templates offer basic support for any type of dashboard layout. However, specific templates can be designed for stratified, group and and schematic dashboards as they have a natural order of the onboarding story based on their ordering.For example, stratified layouts emphasize a top-down ordering of widgets and their information
7		Structure	● ⏳	Currently, we only designed and developed for a single page dashboard structure. In the future, we plan to support multiple pages as well and the structures such as parallel and hierarchical can help us define an order of the dashboards to be onboarded.
8		Color	● ⏳	Apart from identifying the colors of the visual marks in our prototype, in the future we plan to offer support for different colors, such as in Colorbrewer for accessibility. Harrower, M., & Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. The Cartographic Journal, 40(1), 27–37.

Support in D-Tour (available ● or future ⏳)

Fig. B.1: Summary of desing patterns suported by D-Tour Prototype.