Authoring Interactive Data-Driven Reports for Business Data Analysts

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 1778

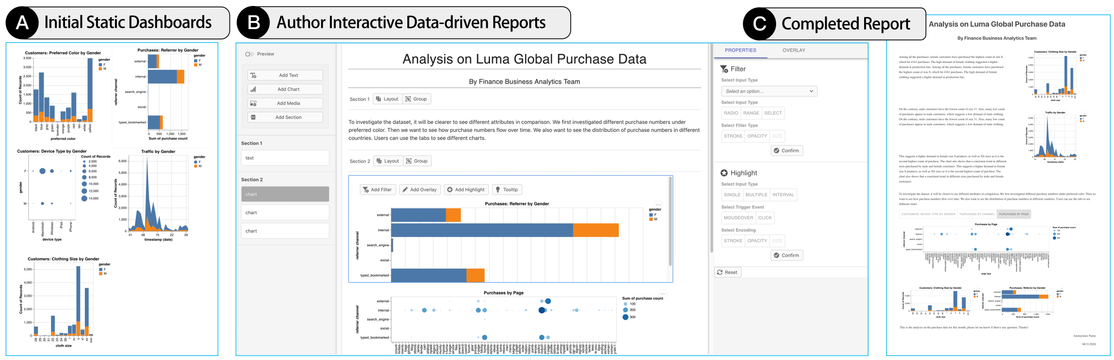

Fig. 1. The workflow overview: (A) Start with initial static dashboards from business analysts, (B) Author interactive data-driven reports using Codas, (C) An example final report in shareable link

Business analysts (BAs) create rich dashboards to find insights from data and communicate these findings with data-driven reports that combine screenshots of their visualizations with descriptive text. In contrast, data stories popularized on the web preserve the interactivity of visualizations while providing a textual narrative. While data stories have been studied for journalism, their applicability to business analytics provides new opportunities and challenges. In this work, we interviewed 15 business analysts and surveyed literature about interactive data stories to understand BAs' unique needs and identify gaps between analysts' reporting needs, technical skills, and existing tools. Towards bridging this gap, we implemented a prototype (Codas) that allows analysts to transform their dashboard into interactive, web-based reports through a UI with no coding. Our case study evaluation with two expert analysts demonstrates that Codas enables BAs to create interactive, data-driven reports that can replace the status quo.

CCS Concepts: • **Human-centered computing → Empirical studies in visualization**.

Additional Key Words and Phrases: data-driven reports, interactive charts, interaction authoring

ACM Reference Format:

Anonymous Author(s). 2021. Authoring Interactive Data-Driven Reports for Business Data Analysts. In *CHI '21: ACM CHI Conference on Human Factors in Computing Systems, May 08–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Manuscript submitted to ACM

1 INTRODUCTION

Business analysts' (BAs) responsibilities include findings insights from data and communicating these insights to stakeholders. During the exploratory data analysis (EDA) process, analysts create rich, interactive dashboards to help understand data, find trends, and explain them. Though these dashboards are sometimes the final artifact of the analysts, analysts often create secondary *data-driven reports* to present results to their stakeholders and audiences. The usual forms of data-driven reports include illustrated documents and emails, blog posts, infographics, or presentation slides.

Especially following the popularization of interactive data stories on the web in journalism, the process and tasks of authoring data-driven stories for general public consumption have been well studied [32, 41, 46]. However, these data stories are quite different from data-driven reports in business settings: often incorporating narrative text, engaging interactions, and highly customized visualizations with annotations. Although there has been some debate about whether interaction in data stories engages readers [1, 6, 9, 10], these are largely focused on the general public. We hypothesize that BAs' audiences would engage with and benefit from interactivity since they are more specialized than the general population and have a vested interest in the data. Few studies have focused on how business analysts create and share data-driven reports in corporate settings, and whether the more dynamic reports in the style of online data stories would be applicable. This paper seeks to answer the following questions:

- How do business analysts create data-driven reports?
- How do the workflows and needs of business analysts differ from the other data-driven report authors?
- Can techniques from journalistic data storytelling be leveraged in business data-driven reports?

To answer these questions, we conducted two formative studies to gather design requirements, built an early prototype, and evaluated the prototype with two expert case studies. We first conducted semi-structured interviews with 15 analysts of different job responsibilities to understand their current workflow, reporting needs, and pain-points. From here, we did a comparative analysis comparing the analysts' technical and content needs to existing data stories on the web and related literature. We identified several key findings that differentiate business analysts and their audiences:

- **Technical Expertise**: Many of the interactive data stories found on the web require specialized programming skills. However, most analysts do not have the programming skilled required to code these stories, nor do they have access to developers on their teams. Business analysts are usually more focused and specialized in non-technical domains like designing and curating visualization dashboards using graphic user interfaces.
- **Tools of Use**: Though tools exist for creating these stories, none are sufficient to meet the needs of the analysts we interviewed. Tools like Vega [43, 44] and Idyll [16] offer solutions for users to create visualization charts. Still they require programming skills and stop at chart creation. Other more accessible tools for business analysts like Tableau Stories [2, 58] and PowerBI [52] are also popular methods for data analysis, but are also limited in integrating interactivity authoring and layouts designing.
- **Specialized Needs**: We reveal specialized needs to business analysts as compared to online data stories (e.g., journalists). Our interview studies show that business analysts prefer to author shorter narratives, with customized explanations and captions per visualization. Additionally, due to the recurrent nature of the reports, business analysts also tend to preserve and templatize reports as the data stream refreshes or per customer.
- **Authoring Interactivity**: Business analysts expressed a desire to not only preserve the interactivity from their existing dashboards, but to author additional interactivity between text, visualizations, and page layout.

We distill these findings and others into seven design requirements, which we used to inform the design and development of a prototype system, Codas, a web-based prototype system for authoring data-driven reports for business analysts. We focus on analysts' process of creating a story from their dashboards by integrating, bridging and implementing the storywriting process with data. We evaluate the system with two expert case studies and demonstrate the utility of our system. The contribution of our work are the following: (i) two formative studies that contribute design requirements for business analysts' reporting needs, ii) A proof-of-concept prototype system Codas, and (iii) two expert case studies with real-word business analysts to evaluate the system.

2 RELATED WORK

In this section, we review prior work on authoring data-driven reports and stories, and interactive charts creation.

2.1 Data-driven Visual Storytelling

Data-driven visual storytelling has become a popular communication medium in a variety of fields like journalism, instructional videos, and blogs [46]. The growing popularity has triggered the visualization research community to investigate various aspects of data-driven visual storytelling. First, in order to distinguish it from general "storytelling," Lee et al. [32] defined visual data stories as a set of story pieces (e.g. charts with annotations and descriptions) in a meaningful order to highlight and emphasize the author's high-level communication goals. They also outlined the authoring process of data-driven storytelling with three stages: (i) exploring data to find facts (i.e. story pieces), (ii) making a story by connecting story pieces in a meaningful way, and (iii) telling a story to audience. According to Kosara et al. [30], the first stage is well supported by commercial tools (e.g. Tableau, Power BI) and research systems like Data Voyager 2 [62] and Lyra [42] for exploratory analysis. But such tools fall short in supporting people in the second and the third steps (i.e., to collect, organize, and present story pieces). Second, many existing works also explored data storytelling in its characteristics [15, 19, 21], methods [5], and applications [31]. Additionally, Weber et al. [61] and Viegas et al. [60] provided examples of presentation and live format in visual storytelling.

A few pioneering works did case studies of data stories published on the internet or news outlets. Back in 2010, Segel and Heer [46] articulated design spaces narrative visualizations for the first time. Hullman and Diakopoulos [25] extended Segel's design space with rhetorical techniques that affect audience's interpretation. Stolper et al. [59] further refined the design space by summarizing recurring storytelling techniques into four categories: communicating narrative and explaining data, linking separated story elements, enhancing structure and navigation, and providing controlled exploration. McKenna et al. [35] added seven factors of visual narrative flows (navigation input, level of control, navigation progress, story layout, role of visualization, story progression, and navigation feedback). Given that the tools and techniques used by practitioners are constantly shifting, in this paper we tried to hear from practitioners about how they currently create data-driven stories, and what kind of unmet needs they have.

Growing bodies of work focus on specific design factors. For instance, interactivity can make data storytelling more versatile and responsive to the information needs of presenter and/or audience [32]. Additional work focuses on how to effectively integrate text and visualization in storytelling. Ottley et al. [36] used eye tracking to examine how people extract information from text and visualization. Zhi et al. [63] explored strategies to combine text and visualization.

2.2 Tools for Creating Data-driven Reports

While most visualization tools lack supports for creating narrative structures, practitioners use a wide range of tools for their needs. For instance, researchers, data analysts, and even journalists are rapidly adopting computational notebooks (e.g., Jupyter) to create and share exploratory data analyses. However, according to two interview studies [29, 39], current notebooks are limited at supporting both exploration (i.e., stage 1 of the data-driven storytelling process) and explanation (i.e., stage 2) in a single document. Ellipsis [41] tried to address the issue by decoupling narrative structure from individual visualizations. Among commercial tools for visual analytic, Tableau Stories allows people to capture charts from a dashboard, add captions, and reorder to create narrative structure, though only allows for stepper layouts.

While the aforementioned tools are mostly WYSIWYG (What You See Is What You Get) editors, domain-specific languages can achieve a higher-level of expressiveness. For instance, Idyll [16] is a markup language that enables authors to create web-based interactive narratives. Compared to general purpose markup languages (e.g., Markdown [22], ArchieML [49]), Idyll allows users to implement interactive visualization components that respond to user input (e.g., click, page scroll) while maintaining readability and simplicity as a high-level markup language. Sketchstory [7] provided an engaging freeform sketching experience to author stories. Fulda et al. [18] proposed a visual timeline authoring system in TimeLineCurator to gather insights and author timelines from texts. To utilize novel interactions in order to explore data and externalize insights, Romat et al. proposed ActiveInk [38] for a natural experience of analysis. Another work which is highly related to our research is InsideInsights [34], which provided a system with interaction techniques to structure annotations and analytic components and link them for a presentation view. In this work, we focus on both a presentable and portable result.

2.3 Business Analyst Workflows and Tools

Much work has studied analysts' processes, discussed challenges within the process, and described collaborations among analysts [4, 14, 17, 27, 28, 37]. Fisher et al. [17] interviewed 16 data analysts within Microsoft about the special considerations for analytics on cloud architectures. Kandel et al. [27] studied enterprise analysts finding that there is much overlap in the high-level analytic process of intelligence and enterprise analysts, while low-level tasks, goals, and data may differ. They defined five high-level tasks: discovery, wrangling, profiling, modeling, and reporting.

Russell et al. [40] characterize high-level sensemaking activities necessary for analysis. Sedlmair et al. [45] discuss difficulties evaluating visualization tools in large corporations, including acquiring and integrating data. Kwon and Fisher [12] discuss challenges novices encounter when using visual analytic tools. Alspaugh et al. [4] found that their participants prefer tools that are open-source, well-supported by the participant's organization, and having staying power in the marketplace (i.e., being not likely to be made obsolete soon). Almost half of their participants reported using visual analytics tools as their primary analysis tool, including Tableau, SAS [55], Splunk [51], Stata [57], Alteryx [56], and Periscope [48]. Several researchers [13, 20, 23] have articulated the importance of capturing provenance to manage analytic workflows, and [26] has advocated the use of visualization across the analysis lifecycle.

Analysts often work individually, rather than as collaborative teams [14] and tend to collaborate asynchronously [27, 28]. They found that analysts collaborate on team to discuss each other's analysis and key findings, but once work is divided, then each part is done individually. When analysts finish their analysis, they need to convert the results into a concise format so that decision makers can understand their findings. This can be tedious and time-consuming part of the intelligence process due to verifying the insights and explaining them clearly.

3 FORMATIVE STUDIES

To confirm analysts' reporting needs and to further understand the differences between data-driven reports versus data stories in journalism contexts, we conducted two formative studies: (1) semi-structured interviews with fifteen professional analysts and (2) a systematic comparison of these analysts reports with literature and recent examples of interactive data stories on the web.

Table 1. Related information about the participants, including years of experience, title (job responsibilities), experience in using related tools

ID	Experience	Title	Domain
P01	15 years	Business Intelligence Analyst	Operations
P02	10 years	Principal Business Analyst	Finance
P03	21 years	Designer	Product Insights and Experiences
P04	3.5 years	Senior Data and Platform Engineer	Sales
P05	2 years	Sales Operations Process Analyst	Cloud Services
P06	6 years	Demand Operations Analyst	Finance
P07	7 years	Business Analyst	Digital Experience
P08	2 years	Intelligence Operations Analyst	Marketing Insights and Operations
P09	3 years	Marketing Specialist	Product
P10	8 years	Senior Business Intelligence Specialist	Product
P11	1 years	Customer Experience Analyst	Support Operations
P12	5 years	Infographics Designer	Customer Experience
P13	1 years	Qualitative Analyst	Sales
P14	2 years	Business Operations Analyst	Strategy
P15	3.5 years	Business Analyst	Strategy

3.1 Study 1: Semi-Structured Interviews with Business Analysts

The primary goal of the interview study [1] was to understand the needs of business analysts, especially in the context of reporting. While there has been much work on analysts workflow [17, 20, 23, 40], the majority focus on data wrangling and analysis process. We wanted to focus specifically on aspects of reporting, such as where it fits into their pipeline, what format are insights communicated, and the tools and methods for creating these artifacts. We also solicited feedback about specific features that could be included in an authoring tool to understand the importance of each.

3.1.1 Method. We conducted semi-structured interviews with 15 professional analysts at a large software company across 14 sessions (one session with two analysts). Participants were recruited through word-of-mouth and internal mailing lists for visualization tool users. Sessions were conducted remotely using video conference and screenshare and lasted between 47 to 67 minutes ($mean = 55m$). We asked each participant about their current workflow and tools, specifically around reporting, aspects of collaboration, and use of mobile devices such as mobile phones or tablets for work. We also asked participants to share an example of a recent artifact (dashboard or report) that they had created with the intent of communicating insights.

Participants. Fifteen participants consisted of twelve business analysts across various domains, two visualization designers, and one software engineer. All participants used visualization to find and communicate insights in their day-to-day job and their experience ranged from 1 to 21 years ($mean$ = 6 years). Ten participants were male, five female.

In terms of technical experience, most participants (80%) were familiar with GUI-based (Graphical User Interface) visual analytics tools such as Tableau, PowerBI, and Excel [53]. Some (27%) were familiar with programming languages, such as Python, R, or visualization libraries such as d3 [8]. All were familiar with some GUI document editor (e.g., Word, PowerPoint [53]) and a minority (27%) had experience with graphic design tools, such as Adobe Creative Suite [47].

Analysis. Interviews were recorded, transcribed, and coded. Participants' answers were thematically analyzed by two researchers. The artifacts shared during the screensharing portion were systematically reviewed for their visualization, interactivity, and text components.

3.1.2 Results. We present the results in three parts: (i) current workflow and tools, (ii) system needs, and (iii) content needs.

[1] Study materials are available in the Supplementary Materials

Current Workflow and Tools. Our findings about workflow were largely consistent with previous work [4]. Analysts followed the general steps of: gathering requirements, data wranglings, preparing the data model and exploring data (iteratively), then reporting the results. Analysts reported a wide range of time spent on the projects, ranging from several days to six months. Seven of the participants spent one to three weeks to create the artifacts, five spent 1-3 months, and the remaining spent over three months. The majority of time was spent coordinating with stakeholders to gather requirements and collaborating with engineers to access and prepare the data. In this work, we focus on the final step of crafting the story into a data-driven report.

In terms of distributing findings and reports, participants often used more than one method to share their reports. All but one participant shared findings through email, either as a link or attachment. Many would also upload documents to central locations accessible by colleagues (e..g, Sharepoint, wiki page). Often times, participants would also give live presentations of their findings (80%).

We asked participants about the role of collaboration in building reports, as this is an important aspect of interactive data stories on the web [24, 33]. Team sizes varied greatly, from three to thirty people, with ten to fifteen of these primarily being in stakeholder roles. Unlike in journalism teams, most common technical collaborators were data engineers who assisted in pulling and cleaning data, and reporting was done by one to three people.

As this study was conducted during the COVID-19 pandemic, all participants were working from home and we asked about any changes in workflow or collaboration as a result. Analysts reported mixed effects of working in fully remote teams. In general, analysts appreciated fewer interruptions and the ability to focus. One participant noted an unintended side effect of requiring better communication between team members resulting in increased documentation. The biggest pain points were a reduced ability to have ad-hoc conversations with colleagues and difficulty in using shared whiteboards when brainstorming as a team. However, several (33%) participants noted little to no change in the ways their teams works, since many had distributed members to begin with.

In terms of tools for collaboration, most were using video conferencing tools and screenshare to communicate. To replace ad-hoc conversation, participants relied more on workplace messaging applications (e.g., Slack).

System needs. To understand system needs, we asked participants about benefits and pain points of current tools, their use of different device types, and preferences about specific features.

In discussing benefits and pain points of current tools, participants noted that they preferred WYSIWYG document editors for communicating insights because it was easier to annotate charts, keep track of notes and insights, and incorporate text than in visualization tools. *"I actually really like PowerPoint because I can put a lot of notes of what I'm seeing in the notes area. I wish I could do that in Power BI or something. Just have a quick ' I'm writing all these notes of what I see,' and then have my own annotation. And, I would say I do that pretty often"* (P13).

The most common limitation for current tools was in being able to incorporate more advanced charts (e.g., waterfall chars, spider charts), though the availability of plug-ins are helpful. The desire for more advanced charts was contrasted with awareness of the audience's visual literacy. *"You can get fancy with spider charts and flow charts..., but if your audience won't follow or understand it [and] it's not easy to interpret..., [they're] going to kind of pass over.* (P13) Participants noted a way to overcome this is through textual explanations of the charts.

We also considered the use of tablets or mobile devices in the design of our tool. No participant used tablets for work, and the majority (67%) had a strong preference to use only laptop or desktop devices. One participant (P6) had used a mobile device, but only to join remote meetings. In addition to rarely using tablets or mobile devices, participants rarely designed dashboards and reports for screens smaller than a laptop or desktop. Participants did not feel that the ability to sketch visualization and ideas (for brainstorming or for authoring) or the inclusion of touch input (e.g., for tablets or mobile) would be helpful for authoring. In fact, 73% of participants preferred either a laptop-only or a laptop plus tablet setup. *"If you can make [the system] work for a tablet and desktop, that's great, but tablet only–I'd be concerned"* (P14).

When asked about specific potential features, participants rated authoring interactions, linking text and annotations to charts, organizing feedback, and organizing charts as the most helpful and necessary for crafting interactive data-driven reports.

Content Needs. Finally, we asked participants about their needs as it pertains to the reports themselves, including use of interactive charts, annotations on charts, and text. The questions were informed by common techniques we found in the surveys of interactive data. We also asked them to share a dashboard or artifact that they had recently created, to capture any components that may not be present in our survey. Twelve participants shared an example.

With the exception of P5 who makes 100% static infographics, interactivity is incorporated dashboard. The degree of interactivity varies from simple things like tooltips to elaborate techniques like sliders, search, filters, zooming, clicking. P11 noted that interactivity was particularly useful for in meetings, when there are additional questions.

Though most participants created interactive dashboards, less than half (40%) delivered these artifacts to stakeholders as a result. Instead, almost all participants (80%) shared their findings with a secondary format by screenshotting their dashboards and using a document editor, such as Microsoft Word or PowerPoint. For participants who did present findings using their dashboards, they said it was particularly useful for answering questions on-the-fly: *"[Dashboards] are useful for in meetings, when there are additional questions. You don't have to create a static view for each question"* (P1).

Participants are easily able to add interactivity in dashboarding tools, like Tableau or Power BI, but felt these tools are limiting for adding text and annotations for presenting or standalone reports. One participant (P12) creates interactive PDFs called Playbooks - which has keys and buttons to show different content.

Many participants (50%) used multiple visualizations in their dashboards to help convey a single insight. These visualizations were often connected through common techniques like overview + detail-on-demand, filtering, and brushing and linking. Only two participants noted that they did not use coordinated visualizations because they were too complex and may lose focus on the story. *"If I were to do it, it'd be a visual breakout of maybe a pie chart into the next stage of it. But, no, I try to keep the, the visualizations pretty simple. A lot of times I've found if you make it too complicated, then you get more questions about it when you want them to focus in on the story"*

Only one participant (P15) indicated that an interactive data-driven report would not be sufficient for their audience, noting that, "the users will want to see the information in one place without scrolling." Remaining participants were enthusiastic about this story format, noting that it would offer a new, engaging way for them to share findings. Some noted that it would be useful depending on the situation: *"I wouldn't want it available during the presentation, but let's say I wasn't there to give the presentation, I could point them there and then they could walk through it."* (P13). There was a consistent theme that the reports could be improved by preserving the interactivity of their dashboards.

In terms of the role of text, usage varied quite a bit. A few participants noted the proportion of visualization to text was about 20% to 30%. For others, it was much more lightweight, using only in tooltips to explain fields (P15), explaining the data model (P8), or descriptions and definitions (P9). Only one participant (P14) noted heavy text usage, where they wrote paragraphs of text to explain the entire dashboard, coupled with captions for context and summarizing each visualization. Participants noted challenges with including text because dashboards needed to cater to many people whose data might be different and would need to be customized per user. Additionally, participants noted a challenge when creating recurring reports: captions largely followed a similar template, but needed to be manually updated every quarter with new insights. Participants also annotated charts directly (47%) using color highlights, guidelines, circles and arrows, notes, and icon markers to help mark insights.

3.2 Study 2: Survey Study and Comparison to Literature

3.2.1 Method. To better understand the state-of-the-art techniques used in recent data-driven reports and compare this with how business analysts would want to deliver the final artifacts, we conducted a survey of related literature and 80 recent data story examples and compared this with our findings from Study 1. This survey is also motivated by RQ3: can techniques from journalism data storytelling be leveraged for business analysts?

Dataset. Existing surveys on data-driven stories and reports [35, 59] have delivered categories and findings on techniques and flows used in stories and reports. Although they covered a comprehensive range of artifacts authored by journalists, BAs' unique needs and preferences have not been fully investigated. Our findings from the interview study motivated us to uncover more techniques used and preferred by BAs. Therefore, we gathered a corpus of recent stories and reports (from 2019 and 2020), analyzed the techniques used from BAs' perspective, then validated and compared the results with existing survey papers.

We first gathered a corpus of 80 recent data-driven reports (full list in Supplementary Materials). We referred to the collecting method used by Stolper et al. [59]: Using popular websites like The New York Times, The Guardian, The Washington Posts, and Bloomberg, which usually have an interactive graphics or stories section. We either used self-defined collections like "2019: The Year in Visual Stories and Graphics," [3] or searched for specific topics and screen for articles with static or interactive glyph-based data visualization charts to construct our corpus. This screening is to make sure that we were getting more "data-driven" reports. We also exclude reports with mostly data tables since they are presenting data in a straightforward format.

Analysis. We analyzed the collected reports by tagging and coding interaction techniques, chart types, layouts, and other elements. We generated an inductive codebook followed by a second (deductive) coding of the entire corpus. Finally, we compared our results to existing survey papers to validate our findings, including one about data-driven storytelling techniques (Survey 1 [59]) and another about visual data story flows (Survey 2 [35]).

3.2.2 Findings. We discuss our findings in three categories: annotations and external representations, interactivity in visualizations, and linking elements.

Annotations and External Representations. Authors apply various kinds of annotations and external representations in nearly every example. Most of these annotations appear in static charts or images. Although we learned from interviews that BAs prefer authoring interactivity in charts, they still use many static charts in reports. One possible reason is that embedding interactive charts requires more technical skill.

The most common and most straightforward technique was using user-generated **overlay annotations** to highlight illustrations. Survey 1 also concluded similarly (i.e., text annotations and labeling), but we extended the category to the broader idea of overlay annotations, including forms like texts, hand-written drawings, shapes, and arrows. BAs use them to highlight insights like trends, peaks, relations, similarities, and disparities.

Visual annotations are also commonly used in data stories and by analysts, including visual encodings like color, size, opacity, etc. Similarly, **tooltips and highlights** are also regularly used in data visualization charts with multiple layers. Tooltips can also add additional details by transferring contents into mouse hovering events. Survey 1 also identified them as essential components.

Survey 1 also brought up other techniques like textual and audio narratives and flowchart arrows, however these were not used by analysts in our interviews. Survey 2 focused more on the visual story flow and neglected how various annotations might affect the storytelling flow.

Interactivity in Visualization Charts. Interactivity is also vital in data-driven reporting. BAs tend to maintain interactivity in the visualization charts. The most common interaction used is **filters**. They usually provide query widgets to filter data. Filtering is an efficient way for the audience to focus on specific fragments of data. Besides filters, chart **groupings** are also commonly used, like **toggled tabs** to switch views and **coordinated views and dashboards** to assist data exploration for the audience, similar to how BAs create dashboards.

Survey 1 did not explicitly discuss interactions in the charts. Instead, it talked about specific interactive visualization types like timeline, geographic maps, and dynamic queries. Survey 2 also did not directly discuss chart interactions, which further supports our claim of addressing BAs' unique preferences and needs.

Linking Elements. The term "linking" here has a broad meaning in linking different objects related to data-driven reports. From a static point of view, **linking between texts and visualization charts** is a popular representation in data-driven reports. We refer to this linking as **layouts**. Survey 1 did not elaborate on different layouts, while Survey 2 explicitly stated story layout models and the role of visualization. We found that BAs prefer various layout options.

Linking between texts and visualization charts is often represented as **navigation**. Authors use various kinds of navigation techniques to enhance storytelling and reporting experiences. We identified similar navigation techniques to Survey 1, like transition animations, jump buttons, and breadcrumbs. Survey 2 also elaborated on navigations. It talked about navigation input, progress, and feedback explicitly. We recognize the importance of navigations in storytelling. We also found that navigations are more common in presentation-style artifacts like slides or multi-page reports like tutorials.

Linking between data and charts is a unique need we discovered from recent reports. Many of the existing data-driven reports are updated daily, like COVID-19 health analysis and political data reports. For BAs, it is also common to update data frequently. However, sometimes data values are also referred to and linked in texts, which caused unnecessary efforts. Therefore, we concluded this unique need for BAs, which is not mentioned in either survey.

4 DESIGN REQUIREMENTS

Based on the formative interviews and survey findings from Section 3, we propose 7 important *design* requirements (**R1-R7**) for developing effective authoring tools for business analysts creating interactive, data-driven reports.

- **R1: GUI-based creation of data stories.** Tools like d3, Vega [44], Vega-Lite [43], and Idyll are powerful programming languages for creating web-based visualizations. However, general business analysts are usually more focused and specialized in non-technical domains. Additionally, in contrast to journalism teams [11], many analysts do not work in teams with software engineers for reporting needs. Our participants expressed that other more accessible tools for business analysts like Tableau Stories and PowerBI were limited in integrating functions of combining narrative with interactive charts and designing narrative-based layouts.
- **R2: Consider the role of collaboration throughout the process.** Though working on teams for data wrangling and requirement collection, analysts are largely creating the dashboard or report alone. We aim to overlap with their existing tools and skills as much as possible.

R3: Preserve interactivity from dashboards. There was a consistent theme that the reports could be improved by preserving the interactivity of their dashboards. This was often lost from dashboards because in order to add contextual text and annotations, analysts were screenshotting the visualizations from their dashboard and importing them into tools like Word and PowerPoint. Recent work studying data journalism found that general audiences to not use interaction provided in online articles. In contrast, analyst's audiences regularly interact with the dashboards they create in order to filter and explore data tailored to them.

R4: Consider the role of shorter, data-driven narratives. Most audiences are key decision makers, so additional context is important (self-service dashboard is not enough), but narrative text is not needed. Our interview studies show that business analysts prefer to author shorter narratives with customized explanations and captions. Due to the nature of recurring reports, business analysts also tend to preserve and templatize reports for reuse between time periods (e.g., quarterly reports).

R5: Resulting document must be both presentable and portable. Most analysts are presenting their reports in meetings *and* distributing through email and shared repositories.

R6: Include additional methods to link multiple visualizations. Similar to data stories, analysts often used coordinated visualizations to explain a single insight and communicate multiple dimensions of the data. However, data stories provide additional methods of grouping charts together, such as tabbed views.

R7: Design for laptop and desktop first. Analysts had a strong preference for using a laptop or desktop, and expected their audience to be viewing the data stories similarly. Though open to the use of tablet or mobile, they expressed interest in doing so in tandem with a computer.

5 THE CODAS SYSTEM

To address our design requirements, we developed a proof-of-concept prototype system called Codas.

5.1 System Overview

Codas is a web-based system using React.js and Vega (R7). The system's overall goal is to support the report-authoring experience from dashboards, including combining storytelling elements (like text, interactive charts, and other media), authoring multiple levels of interactions, organizing story elements, and generating the final artifact. In this section, we provide details of the system in four main parts: (i) input and preprocessing, (ii) arranging the storyline, (iii) developing the story, and (iv) previewing and rendering.

5.2 Input and Preprocessing

We split the input into two parts: dataset (as a CSV) and a list of Vega-Lite [43] chart specifications (as JSON). Vega-Lite provides declarative specifications for quickly creating exploratory data analysis visualizations and many GUI tools exist to create Vega or Vega-Lite specifications (e.g., Data Voyager 2, Lyra). The Codas system preprocesses the chart specifications by parsing the specification's parts of interactions, and adjusting them to fit with our standard format (Fig. 3C). This prepares the charts for annotations, tooltips, filters, and highlighting, as well as responsive layouts.

5.3 Arranging the Storyline

The Codas system contains a visual data storyline approach (R1), where users can view an overview of the report in a condensed format. From this panel, they can add, remove, edit, and rearrange *story elements*, including text, charts, media (images and video), and sections (Fig. 2A).

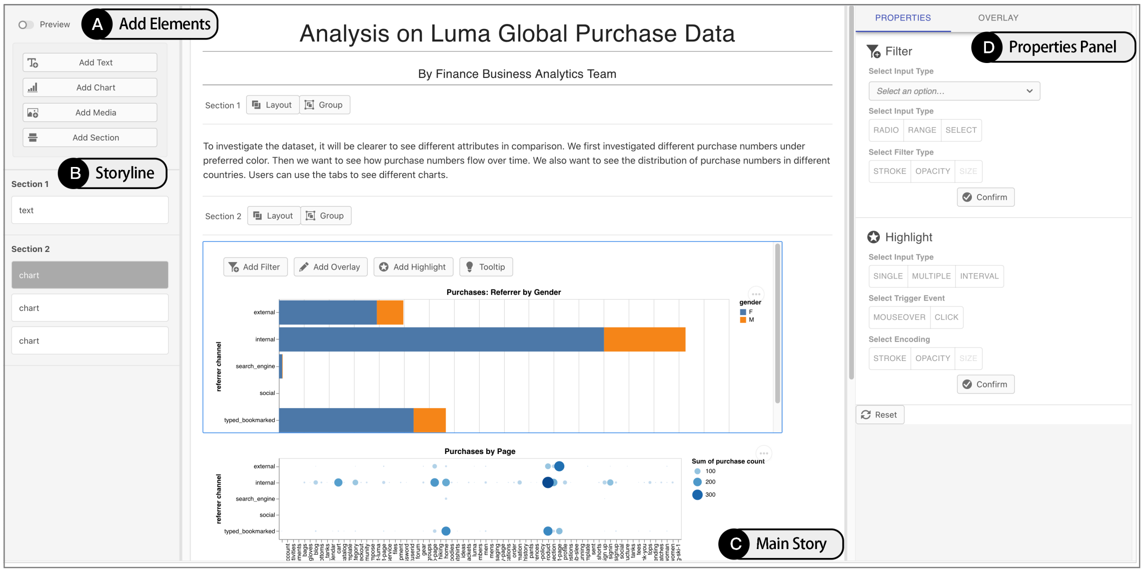

Fig. 2. The Codas system: (a, b) Storyline panel which contains an overview of the data storyline and allows users to add, remove, and rearrange story elements, (c) Main panel which contains the primary story rendering and options to edit individual story elements, and (d) Properties panel, area that contains chart properties and annotation overlay.

Sections. A section in Codas is a container for a series of story piece elements. It is the basic unit for performing interactions between elements. We created this ease-of-use workflow to minimize the analysts' efforts when authoring section-level interactions (R6). We designed section-level interactions include grouping interactions among charts and layout interactions between charts and texts (Fig. 4A).

From the storyline panel, users can easily drag-and-drop story elements to rearrange their position within and between sections. Changes are rendered immediately in the main story panel, which is described next.

5.4 Developing the Story

The next panel comprises the main view for authoring the story and allows users to edit individual story elements.

5.4.1 Text and media. We integrate an open-source WYSIWYG system as the text input method into our tool, which provides a rich set of text editing tools including formatting, linking, and attachments.

5.4.2 Charts. We use the *react-vega* [54] component to render visualizations in the main editing panel. We also designed a toolbar to support users in authoring interactivity (R3). All the authored interactions appear as list of properties in the property panel (Fig. 2D).

Authoring Interactions in Charts. To maintain interactivity when transferring dashboards of charts into data-driven reports, we designed highly simplified GUI for authoring interactions in charts. We developed a simple Vega specification parser (Fig. 3C) and an easy-to-use model to author interactions (Fig. 4) with a WYSIWYG-style toolbar which contains all the options to add interactions at the chart level.

Adding Filters & Highlights. Filtering and highlighting are popular and essential features preferred by BAs. Users can add multiple filters or highlights to one chart, and the added filters will appear in the Properties panel (Fig. 5D) as a list

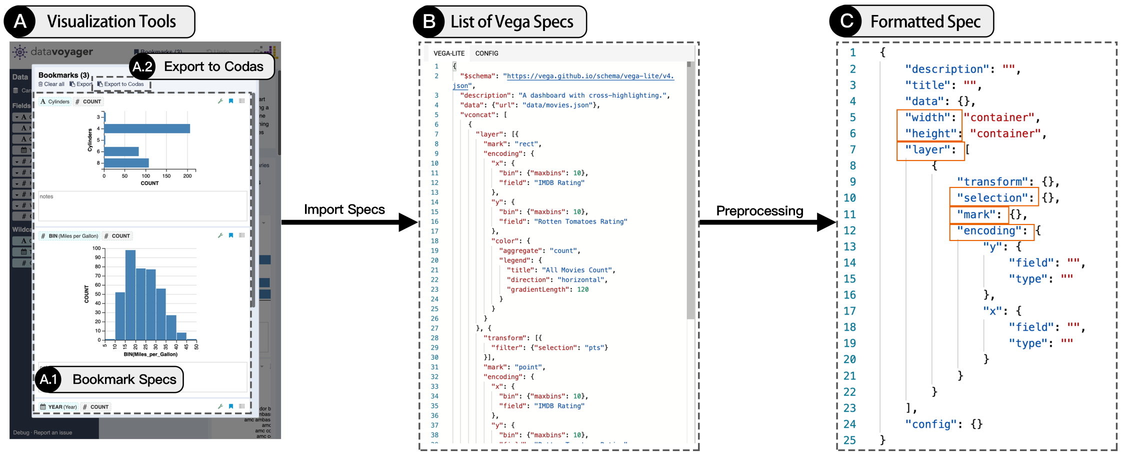

Fig. 3. The input process of the CODAS system: (A): Use Data Voyager 2's bookmark system to export Vega specifications, (B): Import them into CODAS system, (C): Parse and preprocess to make sure the charts have the standard format. The specifications should include "selection", "mark" and "encoding" in every object of the "layer" list

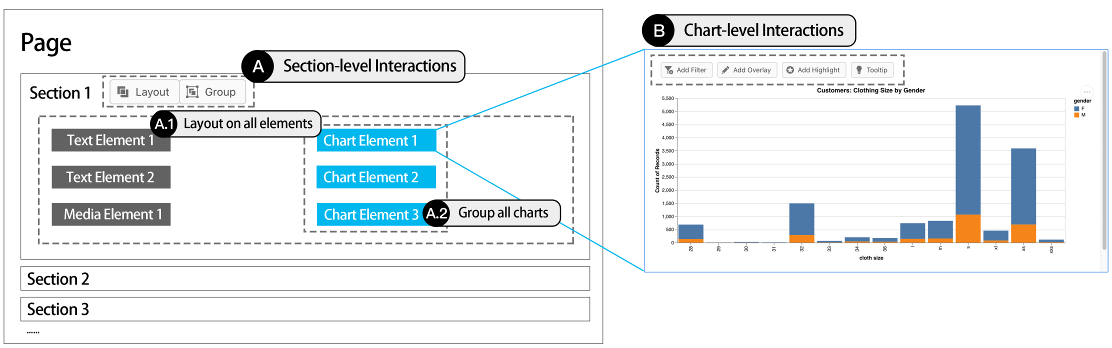

Fig. 4. The interactions in the CODAS system: (A): Section-level interactions, including grouping aggregations on all charts in one section, and layout styling on all elements in the section. (B): Chart-level interactions on single charts, including filter, overlay, highlight, and tooltip

of properties. For each filter, CODAS provides three simple inputs (Fig. 5C): (1) An attribute in the dataset to filter, (2) input type from checkbox (for boolean values), range (for numeric values), selections or radio (for nominal values), and (3) encoding type from size, stroke width, or opacity. Similarly, CODAS also provides three inputs for highlighting (Fig. 5C): (1) input type from single, multiple, or interval selection, (2) trigger event from mouseover or click, and (3) encoding type which is the same as filter. The supported filter and highlight aggregations are a subset of what Vega supports as query widgets.

Creating Overlays. Static charts with user-generated overlay are also popular in data stories. BAs tend to add arrows, links, self-defined annotations to highlight insights manually. The CODAS system provides a drafting panel for users. The draft panel enables users to add, edit, and delete overlay annotations freely by providing PhotoShop-style components

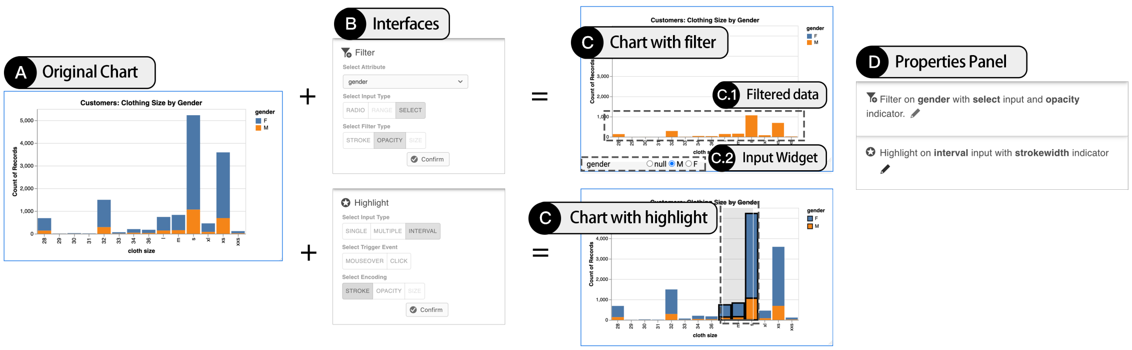

Fig. 5. An example of adding filter: (1) Use the interface to select attribute to filter, input type, and encoding type. (2) The new chart will have the input query widget for the selected attribute, and a new property will be added to the properties panel of this chart

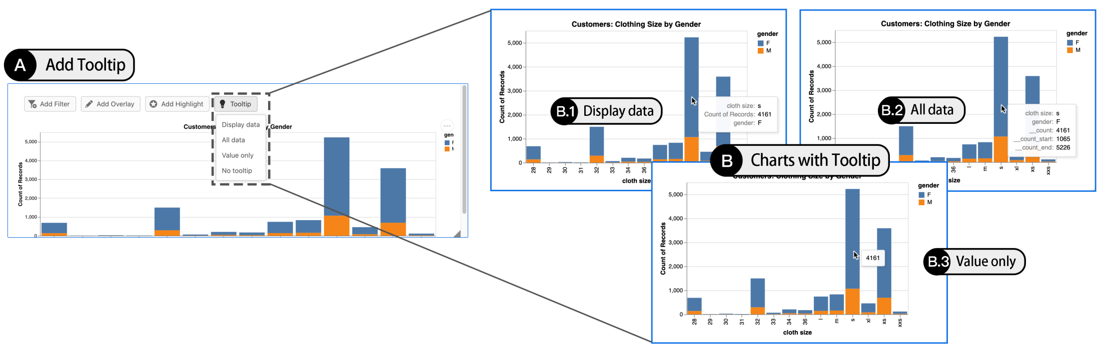

Fig. 6. Authors can choose one of the three tooltip types: (1) Display data, where the audience will see the relevant data shown in the chart, (2) All data, where the audience will see all the data associated with this datum piece in the dataset, (3) Value only, where the audience will only see the current data value

like text, pencil, line, rectangle, circle, selection tools, and other useful tools for drafting. After users have finished overlaying annotations, the Codas system will adapt the changes and reflect the final artifact to the main editing panel.

Adding Tooltips. Similar to highlights, users can also add tooltips to help the audience access the additional details in the data. The Codas system provides tooltip types supported by Vega. They are generated based on different factors, including the encoding, the underlying data point, and the `tooltip` channel directly. Users will only need to select from one of the three tooltip types (Fig. 6).

5.4.3 *Layout Styles.* Another kind of interaction between charts and other elements is the layout (R6). Different layout styles could indicate the relationships between charts and story elements, link different components, or serve as navigation to other elements. Layout style takes effect as a section-level attribute, which means that all the elements in one section will be rendered with layouts accordingly. The Codas system supports four types of layouts (Fig. 7 A): (1) inline, the default style that all elements will be in a linear mode; (2) aside, all charts will show aside with the paragraphs, which indicates a parallel relationship between texts and charts; (3) aside-fixed, a similar layout to "aside," with the difference of fixing the charts on the screen like a reference as text flows; (4) overlay-text, put the charts as

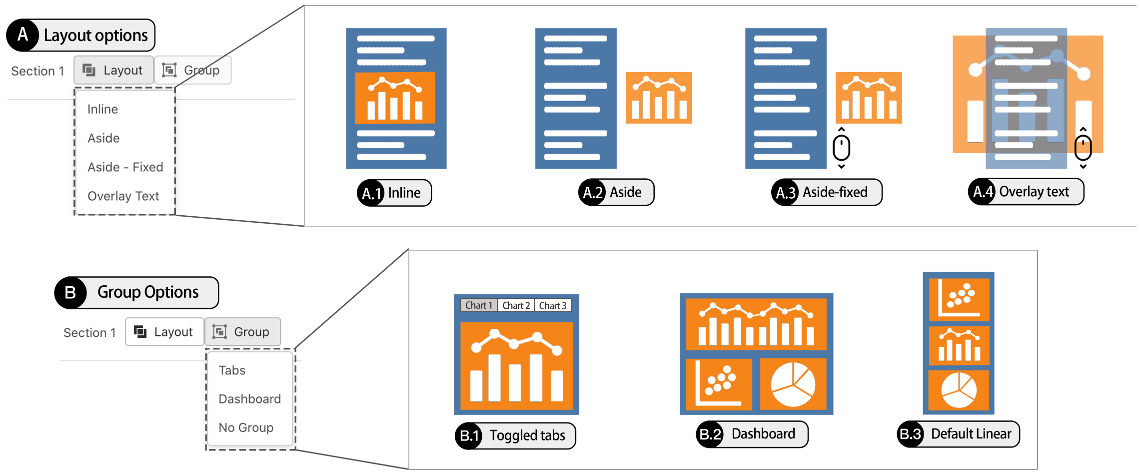

Fig. 7. (A): Grouping aggregations supported in the Codas system: (1) Toggled tabs for switching views, (2) Resizable dashboards where authors can adjust and reposition the grouped charts for the audience; (B): Layout styles supported in the Codas system: (1) Inline (2) Aside (3) Aside-fixed (4) Overlay text

fixed backgrounds as the text paragraphs scroll down, which is typical when the charts are too complicated to be put aside, like a world map visualization.

5.4.4 Chart Grouping. The Codas system supports another primary type of interaction among the charts (R6), of which we use the term "grouping." Grouping also takes effect as a section-level interaction, which means all the charts in one section will be grouped accordingly. The Codas system supports two types of grouping (Fig. 7 B): (1) toggled tabs, an integrated group of visualization charts where users can use toggle buttons to switch similar or related views; (2) dashboards, a resizable and repositionable set of coordinated views, which may contain different subsets of the same data and facilitate comparison across subsets. Furthermore, interactions of grouping and layout could be applied together to one section. In this case, all the charts will be combined into a chart group, and it will become a new story element. The Codas system will render the new element in different layout styles accordingly. By default, the Codas system renders the assembled story elements in linear order.

5.5 Previewing and Rendering

Like most authoring and creation tools, we provide a preview function for analysts. Analyst users can preview the current final artifact at any point during the authoring process.

When the user is finished, they can export Codas into a shareable link, which can be easily shared and viewed in any native browser (R5). Analysts may also download the raw data story structure as a JSON, that can be shared directly with team members for collaborative editing. The file accommodates all the interactivity, grouping, and layout styles authored by users, and it could be re-rendered using the Codas system. This method could apply to situations where analysts iterate the reports with team members (R2).

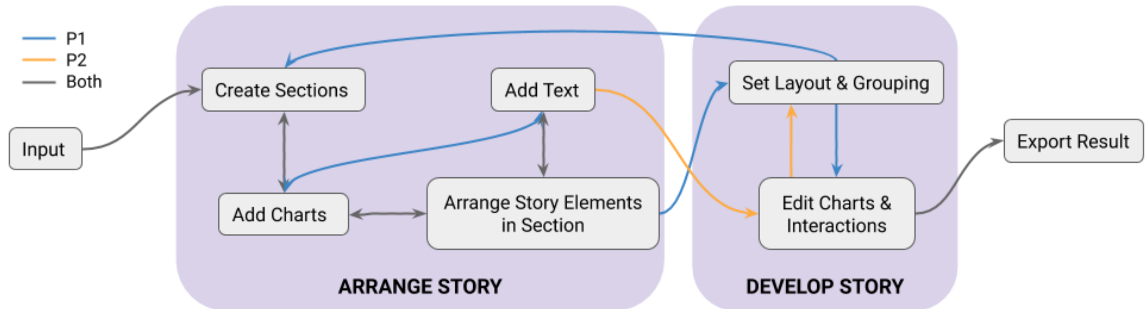

Fig. 8. The workflow for P1 and P2 in case studies. Arrows indicate the flow directions under certain participants, showing different possible work loops in the open-ended system

6 CASE STUDIES

To solicit feedback and insights on the usability, potential usage, and future improvements of Codas, we conducted two case studies [2] with experienced business analysts. The analysts were recruited from the pool of analysts who participated in our formative interviews. The analysts shared a dashboard that they wanted to create a report from ahead of time which was imported into Codas. Each participant was trained on the features of system then asked to create a report using their own data and visualizations using a think-aloud protocol. Two researchers observed and took notes. Participants were encouraged to try as many features as possible and create the most complete data story possible. After they created their reports, they provided their feedback in an open-ended interview. Each session lasted one hour.

6.1 Financial Operations

The first participant (P1) was an operations demand analyst responsible for managing and tracking spend across his organization and providing insights to its leadership team. He makes quarterly reports about target metrics, such as pacing and budget forecasting. His role also included consulting with other teams to optimize their analysis processes.

Dataset. The dataset consisted of spending and budget details for 791 teams in one organization. Each team had data from one quarter in 2020 about committed budget, actual spend, spend type, and metadata about the team, such as reporting executive and organization. The original dashboard consisted of five visualizations across three tabs, including a grouped bar chart, table summary display, and three bar charts.

Workflow. P1 began by reviewing the available charts. His overall approach was to linearly create the sections. In each section, he added the relevant charts, text, and other elements. Only after all elements were added to a section did he begin to group charts and focus on the section layout. He would then move onto the next section. P1 used grouping and layout options to (i) aggregate related charts into tabs and (ii) set the charts to the "aside" position.

As he created new sections, P1 would also go over to previous sections to view the overall flow of the report. After the overall layout and grouping designs were finished, P1 started to author interactions on single charts, including adding tooltips and filters to provide readers with the ability to tailor and dive deep into the data.

[2] Study procedure is available in the Supplementary Materials

Feedback and Takeaways. P1's overall feedback about the system was positive. What he liked most about the system was the ease of adding story pieces (both chart and text) and quickly arrange the sections to form a narrative. *"It's simple in a good way. The simplicity of adding things to the report and moving things around is really straightforward."* He noted this would be useful in many use cases, especially when building regularly recurring reports. *"his would be really valuable to do once a quarter as a post-mortem to describe what the VP spent, with more granular insights [or] weekly business reports. I currently build everything in Excel, and put it into PowerPoint manually, for each spend leader."* P1 mentioned and emphasized that data is not static and changes over time in their work scenarios, so analysts need to update the reports periodically. We have identified this need from the formative interviews and decoupled the data from the visualizations so it can be updated independently or automatically refreshed. He noted it is a manual process for him currently: *"I have to refresh the spreadsheet in the dashboard every week."* Though we offered the ability to add other media, such as images or video, P1 said he would not use them for his reports.

P1 also expressed directions for future work. In particular, he wished to be able to templatize the captions more, in order to provide personalized insights across spender leaders and quarters. For example, he might want to call out the top spend categories for each audience member in a caption. It would also be necessary to allow for calculated values, such as pacing based on the current week. Although we allowed authoring filters on the individual charts, P1 expressed the need for global (page-level) *and* section-level filters that would be applied to the charts. He also mentioned that he would like more design control over the page, such as selecting background color and fonts or applying a theme. Lastly, P1 noted he would want the reporting feature to be more tightly integrated with the analysis process, in order to refine existing charts, create new ones, or preview and analyze the data.

6.2 Customer Insights

6.2.1 Background. Our second participant (P2) was an analyst with 7 years of experience in marketing insights. He supports multiple business units in meeting their visual reporting needs, from dashboards to standalone reports. In this case, he had created a dashboard to share the results of a customer satisfaction survey for a software company.

6.2.2 Dataset. The dataset consisted of over 18,000 customer responses about satisfaction related to a set of software products. Dimensions included customer demographics such as gender, age, and job type (e.g., professional, student), as well as responses to ten 5-point Likert questions about the product buying and usage experience. The shared dashboard consisted of ten visualizations: three bar charts, four donut/pie charts, a treemap, and two gauge charts.

6.2.3 Workflow. P2 began by immediately adding all the charts to the story, then creating sections. He then arranged the charts into their respective sections and arranged the overall flow in the Storyline panel. After the flow of the sections was laid out, he turned to the Story panel to add text, starting with section headers, then description captions, and finally paragraphs explaining the insights. From here, he edited individual charts to add filters and highlighting interactions. Lastly, he arranged the charts into groups and selected the layout for each section.

6.2.4 Feedback and Takeaways. Overall, P2 was felt the tool was very useful and showed a lot of potential. P2 found the "dashboard" layout particularly helpful, noting that even in dashboarding tool, it is difficult to flexibly resize and arrange charts. In terms of the final data story, P1 said it would be useful in replacing live presentations, for which he current uses PowerPoint or similar. Interestingly, P2 noted that it is particularly important to be able to edit the charts outside of the dashboard because the final report might not require the same amount of detail due to the additional text. In contrast to P1, P2's workflow focused first on developing the overall story, then refining individual sections (Fig. 8).

In terms of improvements, P2 expressed a need for templatized annotations as well. For example, wanting to author an annotation around the "maximum" peak, that was update based on data values. As for the exporting options, P2 suggested that it could be more diverse, like exporting to existing formats like PowerPoint, documents, and interactive PDFs. P2 also noted that it would be useful to allow for additional page layouts, such as horizontal scrolling.

Lastly, P2 reiterated many of the takeaways as P1: closer integration with the analysis step, to iterative analyze and create the data story at once, usefulness of live data integration,

7 DISCUSSION

In this section, we discuss our findings in better supporting business analysts in creating interactive data stories, specifically in respect to adherence to our design requirements and new requirements from the case studies.

Shareable interactive reports. BAs use presentation-style artifacts to convey insights, but they also desire a way to share reports and maintain their interactivity (R5). A web link is one of the most popular methods for embedding interactive charts and distributing data-driven reports. Other trending formats like interactive PDF [50] could also be used as a portable document format.

Templatized reports. BAs' job responsibilities require them to update reports with massive new data periodically. To support analysts in creating that, designers and developers should further support templatized reports for data refreshing. This is necessary for both the narrative text and annotations on the charts themselves.

Integration with exploratory data analysis tools. Existing tools like Tableau and PowerBI are used by analysts for exploratory data analysis and to create dashboards. Though analysts use these dashboards as a basis for their data-driven reports, during the storytelling process, analysts may find the need for new charts or insights to be conveyed. It is critical to incorporate functions like exploring the dataset table into the new authoring tool. BAs also need to refer to datasets to make design decisions.

Multi-level interaction authoring. Though we identified the use of incorporating chart-level filters from our formative study (R3), we found that additional filters and interactions are required at the page- and section-levels.

More flexibility in page design. It is essential to provide BAs with the overall control of layout styles. They need to have more flexibility in repositioning elements to make a more personalized report, closer to what is available in word processing tools. Additionally, they appreciate more control over the look-and-feel of the page.

Support collaborative feedback. Although collaborative tools of data-driven reporting are trending nowadays [34], collaboration typically happened in high-level design requirement communications based on our findings. When analysts are engaged in authoring reports and delivering insights, they do not need synchronous collaborations to author design changes. Instead, they are in need of gathering feedback collaboratively afterwards.

We believe that Codas is a first step towards better support BAs' need to create interactive data reports. Next, we will discuss the limitations of our prototype system and some interesting future research directions.

7.1 Limitations

While we gathered plenty of insightful feedback from two expert case studies, our prototype system can benefit from a more controlled user study to prove its generality, usefulness, and user-friendliness among analysts. We could remedy this limitation by deploying this system to a broader range of use and conduct more quantitative analysis on the system usage, like a weekly diary study of recording analysts' activities when using this tool to curate interactive data-driven reports for distribution. Although we recruited many interview participants with very different job responsibilities to increase our findings' diversity, they were from the same company. The disparities in job responsibilities could efficiently neutralize the single-company effect, but it might still influence the system's generality.

Another limitation lies in the prototype implementation. Since some of the functionalities require integration of many existing state-of-the-art techniques to reach a highly smooth system, we did not include them in the prototype. Instead, we implemented critical functions to support an end-to-end authoring workflow to validate our overall framework. Besides, the prototype also needs a more systematic review of real analysts' reports and artifacts to cover a broader range of supported input materials rather than mostly dashboards. Incorporating the Vega ecosystem, our prototype is also limited to what Vega can achieve in chart creation. Finally, the early research prototype is still not comparable to real industry tools in many details and nuances.

7.2 Future Work

Our framework and prototype's key contribution is to bridge the gap between BAs' workflow and needs and the skills required to create a sophisticated interactive data-driven report. The long-term goal of our work is to deeply understand BAs' workflow and enable them to author interactive data-driven reports effortlessly. To achieve that, we will work closely with more analysts to iterate the prototype in many aspects. First, we will incorporate more chart creation specifications beyond Vega and more exporting options other than web links. Then we will also integrate state-of-the-art techniques of data exploration and other layout styling methods to make a smooth system for real use. Additionally, we only assessed the system from the BAs' perspective, who assumed their audience would engage more with the data reports mostly. Future work could include a more controlled study on whether or not this system increases engagement or suits the needs of BAs' audience.

For future researchers, designers, and developers, it is also worthy of exploring an intelligent system for authoring data-driven reports. One of our expert participants mentioned the need for a recommendation system to provide suggestions when selecting visualization charts, grouping, and layout options. The system could directly recommend visualization charts to use either from the dataset directly or from the user's dashboards based on the user's general goal. The system could also utilize attributes and distribution of the report elements for recommending elegant layouts and groupings. These improvements could better support analysts in creating a clear, engaging, and insightful data-driven report.

8 CONCLUSION

In this work, we aimed to answer, *"Can interactive data stories be used in business contexts, and if so, how can we support analysts in creating them?"*. Towards answering this, we conducted two formative studies to understand the workflow and find the needs of BAs as it related to creating data stories. We collected findings from 15 interviews of BAs and a comparative analysis of literature into seven design requirements. We implemented an early prototype called Codas based on these requirements and evaluated our system using two case studies. Through the formative studies, prototype development, and case study evaluation, we found that (i) analysts have limitations in their current reporting workflows, (ii) business analysts have unique needs that are not filled by existing data storytelling tools, and (iii) analysts were enthusiastic about creating interactive data stories for their audiences.

9 ACKNOWLEDGEMENTS

Removed for anonymity.

REFERENCES

[1] 2016. Gregor aisch. https://vimeo.com/182590214

[2] 2016. Tableau(Version. 9. 1). *Journal of the Medical Library Association : JMLA* 104, 2 (April 2016), 182–183. https://doi.org/10.3163/1536-5050.104.2.022

[3] 2019. 2019: the year in visual stories and graphics. *The New York Times* (Dec. 2019). https://www.nytimes.com/interactive/2019/12/30/us/2019-year-in-graphics.html

[4] S. Alspaugh, N. Zokaei, A. Liu, C. Jin, and M. A. Hearst. 2019. Futzing and Moseying: Interviews with Professional Data Analysts on Exploration Practices. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 22–31.

[5] Daniel Andrews and Chris Baber. 2014. Visualizing interactive narratives: employing a branching comic to tell a story and show its readings. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, Toronto, Ontario, Canada, 1895–1904. https://doi.org/10.1145/2556288.2557296

[6] Dominikus Baur. 2017. The death of interactive infographics? https://medium.com/@dominikus/the-end-of-interactive-visualizations-52c585dcafcb

[7] Bongshin Lee, Rubaiat Habib Kazi, and Greg Smith. 2013. Sketchstory: telling more engaging stories with data through freeform sketching. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (Dec. 2013), 2416–2425. https://doi.org/10.1109/TVCG.2013.191

[8] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2301–2309.

[9] Jeremy Boy, Francoise Detienne, and Jean-Daniel Fekete. 2015. Storytelling in information visualizations: does it engage users to explore data?. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 1449–1458. https://doi.org/10.1145/2702123.2702452

[10] Jeremy Boy, Louis Eveillard, Francoise Detienne, and Jean-Daniel Fekete. 2016. Suggested interactivity: seeking perceived affordances for information visualization. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 639–648. https://doi.org/10.1109/TVCG.2015.2467201

[11] Brian Boyer. 2014. How we work. http://blog.apps.npr.org/2014/06/04/how-we-work.html

[12] B. c. Kwon, B. Fisher, and J. S. Yi. 2011. Visual analytic roadblocks for novice investigators. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 3–11.

[13] Steven P. Callahan, Juliana Freire, Emanuele Santos, Carlos E. Scheidegger, Cláudio T. Silva, and Huy T. Vo. 2006. VisTrails: Visualization Meets Data Management. In *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data* (Chicago, IL, USA) *(SIGMOD '06)*. Association for Computing Machinery, New York, NY, USA, 745–747. https://doi.org/10.1145/1142473.1142574

[14] George Chin, Olga A. Kuchar, and Katherine E. Wolf. 2009. Exploring the Analytical Processes of Intelligence Analysts *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 11–20. https://doi.org/10.1145/1518701.1518704

[15] Eun Kyoung Choe, Bongshin Lee, and m.c. schraefel. 2015. Characterizing visualization insights from quantified selfers' personal data presentations. *IEEE Computer Graphics and Applications* 35, 4 (July 2015), 28–37. https://doi.org/10.1109/MCG.2015.51

[16] Matthew Conlen and Jeffrey Heer. 2018. Idyll: A markup language for authoring and publishing interactive articles on the web. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*. 977–989.

[17] Danyel Fisher, Rob DeLine, Mary Czerwinski, and Steven Drucker. 2012. Interactions with Big Data Analytics. *ACM Interactions* (May 2012). https://www.microsoft.com/en-us/research/publication/interactions-with-big-data-analytics/ This is a pre-publication draft; do not distribute or reproduce. The official version may be found at http://dl.acm.org/citation.cfm?doid=2168931.2168943.

[18] Johanna Fulda, Matthew Brehmer, and Tamara Munzner. 2016. Timelinecurator: interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (Jan. 2016), 300–309. https://doi.org/10.1109/TVCG.2015.2467531

[19] Nahum Gershon and Ward Page. 2001. What storytelling can do for information visualization. *Commun. ACM* 44, 8 (Aug. 2001), 31–37. https://doi.org/10.1145/381641.381653

[20] D. Gotz and M. X. Zhou. 2008. Characterizing users' visual analytic activity for insight provenance. In *2008 IEEE Symposium on Visual Analytics Science and Technology*. 123–130.

[21] S. Gratzl, A. Lex, N. Gehlenborg, N. Cosgrove, and M. Streit. 2016. From visual exploration to storytelling and back again. *Computer Graphics Forum* 35, 3 (June 2016), 491–500. https://doi.org/10.1111/cgf.12925

[22] John Gruber. [n.d.]. Daring Fireball: Introducing Markdown. https://daringfireball.net/2004/03/introducing_markdown

[23] J. Heer, J. Mackinlay, C. Stolte, and M. Agrawala. 2008. Graphical Histories for Visualization: Supporting Analysis, Communication, and Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1189–1196.

[24] Jane Hoffswell, Wilmot Li, and Zhicheng Liu. 2020. Techniques for flexible responsive visualization design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, Honolulu, HI, USA, 1–13. https://doi.org/10.1145/3313831.3376777

[25] Jessica Hullman and Nick Diakopoulos. 2011. Visualization rhetoric: Framing effects in narrative visualization. *IEEE transactions on visualization and computer graphics* 17, 12 (2011), 2231–2240.

[26] Sean Kandel, Jeffrey Heer, Catherine Plaisant, Jessie Kennedy, Frank van Ham, Nathalie Henry Riche, Chris Weaver, Bongshin Lee, Dominique Brodbeck, and Paolo Buono. 2011. Research Directions in Data Wrangling: Visualizations and Transformations for Usable and Credible Data. *Information Visualization* 10, 4 (Oct. 2011), 271–288. https://doi.org/10.1177/1473871611415994

[27] S. Kandel, A. Paepcke, J. M. Hellerstein, and J. Heer. 2012. Enterprise Data Analysis and Visualization: An Interview Study. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2917–2926.

[28] Y. Kang and J. Stasko. 2011. Characterizing the intelligence analysis process: Informing visual analytics design through a longitudinal field study. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)*. 21–30.

[29] Mary Beth Kery, Marissa Radensky, Mahima Arya, Bonnie E John, and Brad A Myers. 2018. The story in the notebook: Exploratory data science using a literate programming tool. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–11.

[30] Robert Kosara and Jock Mackinlay. 2013. Storytelling: The next step for visualization. *Computer* 46, 5 (2013), 44–50.

[31] Kwan-Liu Ma, I. Liao, J. Frazier, H. Hauser, and H.-N Kostis. 2012. Scientific storytelling using visualization. *IEEE Computer Graphics and Applications* 32, 1 (Jan. 2012), 12–19. https://doi.org/10.1109/MCG.2012.24

[32] Bongshin Lee, Nathalie Henry Riche, Petra Isenberg, and Sheelagh Carpendale. 2015. More than telling a story: Transforming data into visually shared stories. *IEEE computer graphics and applications* 35, 5 (2015), 84–90.

[33] Wesley Lindamood. 2018.

[34] A. Mathisen, T. Horak, C. N. Klokmose, K. Grønbæk, and N. Elmqvist. 2019. Insideinsights: integrating data-driven reporting in collaborative visual analytics. *Computer Graphics Forum* 38, 3 (2019), 649–661. https://doi.org/10.1111/cgf.13717

[35] S. McKenna, N. Henry Riche, B. Lee, J. Boy, and M. Meyer. 2017. Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences. *Computer Graphics Forum* 36, 3 (June 2017), 377–387. https://doi.org/10.1111/cgf.13195

[36] Alvitta Ottley, Aleksandra Kaszowska, R. Jordan Crouser, and Evan M. Peck. 2019. The Curious Case of Combining Text and Visualization. *EuroVis 2019 - Short Papers* (2019), 5 pages. https://doi.org/10.2312/EVS.20191181 Artwork Size: 5 pages ISBN: 9783038680901 Publisher: The Eurographics Association Version Number: 121-125.

[37] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis.

[38] Hugo Romat, Nathalie Henry Riche, Ken Hinckley, Bongshin Lee, Caroline Appert, Emmanuel Pietriga, and Christopher Collins. 2019. Activeink: (Th)inking with data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, Glasgow, Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300272

[39] Adam Rule, Aurélien Tabard, and James D. Hollan. 2018. Exploration and Explanation in Computational Notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173606

[40] Daniel Russell, Mark Stefik, Peter Pirolli, and Stuart Card. 1993. The cost structure of sense making.

[41] Arvind Satyanarayan and Jeffrey Heer. 2014. Authoring narrative visualizations with ellipsis. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 361–370.

[42] Arvind Satyanarayan and Jeffrey Heer. 2014. Lyra: an interactive visualization design environment. *Computer Graphics Forum* 33, 3 (2014), 351–360. https://doi.org/10.1111/cgf.12391

[43] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2016. Vega-lite: A grammar of interactive graphics. *IEEE transactions on visualization and computer graphics* 23, 1 (2016), 341–350.

[44] Arvind Satyanarayan, Kanit Wongsuphasawat, and Jeffrey Heer. 2014. Declarative Interaction Design for Data Visualization. In *ACM User Interface Software & Technology (UIST)*. http://idl.cs.washington.edu/papers/reactive-vega

[45] Michael Sedlmair, Petra Isenberg, Dominikus Baur, Andreas Butz, and Bmw Group. 2010. Evaluating Information Visualization in Large Companies: Challenges, Experiences and Recommendations. *Conference on Human Factors in Computing Systems - Proceedings* (04 2010). https://doi.org/10.1145/2110192.2110204

[46] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010), 1139–1148.

[47] Software. Retrieved 2020-09-17. Adobe creative cloud. https://www.adobe.com/creativecloud.html

[48] Software. Retrieved 2020-09-17. Advanced analytics & cloud data platform for data teams. https://www.sisense.com/product/data-teams/

[49] Software. Retrieved 2020-09-17. Archie Markup Language (ArchieML). http://archieml.org/

[50] Software. Retrieved 2020-09-17. Create an interactive PDF. https://helpx.adobe.com/indesign/how-to/indesign-create-interactive-pdf.html

[51] Software. Retrieved 2020-09-17. The data-to-everything platform built for the cloud. https://www.splunk.com/en_us/homepage.html

[52] Software. Retrieved 2020-09-17. Data visualization | microsoft power bi. https://powerbi.microsoft.com/en-us/

[53] Software. Retrieved 2020-09-17. Microsoft Office 365 is a part of Microsoft 365. https://www.microsoft.com/en-us/microsoft-365/office-365

[54] Software. Retrieved 2020-09-17. React-vega. https://www.npmjs.com/package/react-vega

[55] Software. Retrieved 2020-09-17. Sas visual analytics. https://www.sas.com/en_us/software/visual-analytics.html

[56] Software. Retrieved 2020-09-17. Self-service analytics, data science & process automation | alteryx. https://www.alteryx.com/

[57] Software. Retrieved 2020-09-17. Stata: software for statistics and data science. https://www.stata.com/

[58] Software. Retrieved 2020-09-17. Stories - Tableau. https://help.tableau.com/current/pro/desktop/en-us/stories.htm. (Accessed on 09/13/2020).

[59] Charles D. Stolper, Bongshin Lee, Nathalie Henry Riche, and John Stasko. 2016. *Emerging and Recurring Data-Driven Storytelling Techniques: Analysis of a Curated Collection of Recent Stories*. Technical Report MSR-TR-2016-14. https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent-stories/

[60] Fernanda B. Viegas and Martin Wattenberg. 2006. Communication-minded visualization: A call to action [Technical forum]. *IBM Systems Journal* 45, 4 (2006), 801–812. https://doi.org/10.1147/sj.454.0801

[61] Anke Weber, Holger M. Kienle, and Hausi A. Müller. 2002. Live documents with contextual, data-driven information components. In *Proceedings of the 20th annual international conference on Computer documentation - SIGDOC '02*. ACM Press, Toronto, Ontario, Canada, 236–247. https://doi.org/10.1145/584955.584990

[62] Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand, Jock Mackinlay, Bill Howe, and Jeffrey Heer. 2017. Voyager 2: Augmenting Visual Analysis with Partial View Specifications. In *ACM Human Factors in Computing Systems (CHI)*. http://idl.cs.washington.edu/papers/voyager2

[63] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and Layout: Exploring the Integration of Text and Visualization in Storytelling. *Computer Graphics Forum* 38, 3 (2019), 675–685. https://doi.org/10.1111/cgf.13719 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.13719.