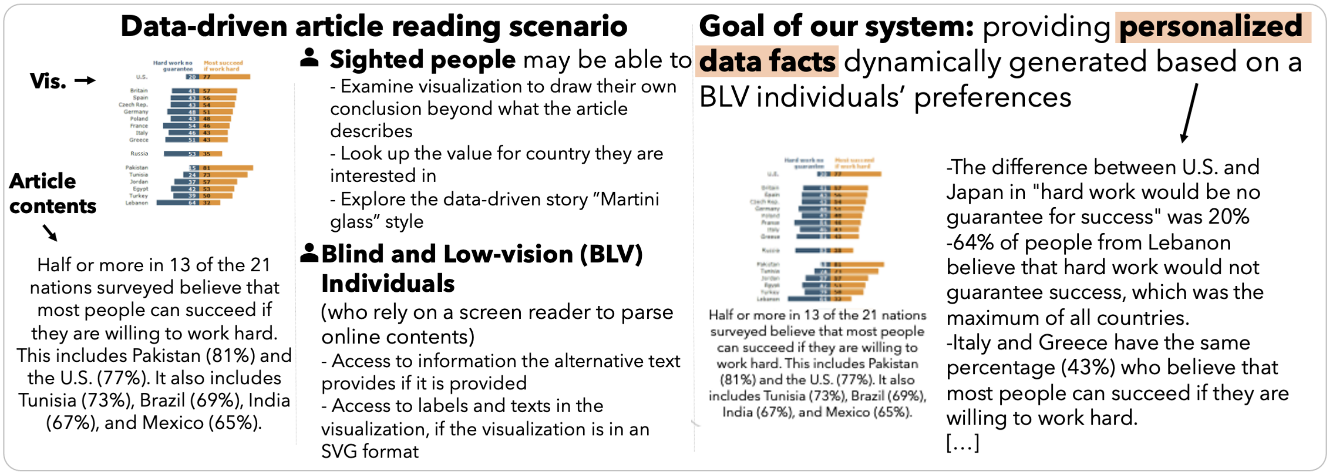
Making Data-Driven Articles more Accessible: An Active Preference Learning Approach to Data Fact Personalization

Yanan Wang
yanan@cs.wisc.edu
Department of Computer Sciences, University of Wisconsin-Madison
Madison, Wisconsin, USA

Yea-Seul Kim
yeaseul.kim@cs.wisc.edu
Department of Computer Sciences, University of Wisconsin-Madison
Madison, Wisconsin, USA


Data-driven article reading scenario
Vis.
Hard work no guarantee
Most succeed if work hard
U.S.
Britain
Spain
Czech Rep.
Germany
Poland
France
Italy
Greece
Russia
Pakistan
Tunisia
Jordan
Egypt
Turkey
Lebanon
Article contents
Sighted people may be able to
Blind and Low-vision (BLV) Individuals
Goal of our system: providing personalized data facts dynamically generated based on a BLV individuals' preferences

Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard. This includes Pakistan (81%) and the U.S. (77%). It also includes Tunisia (73%), Brazil (69%), India (67%), and Mexico (65%).

- Examine visualization to draw their own conclusion beyond what the article describes
- Look up the value for country they are interested in
- Explore the data-driven story "Martini glass" style

(who rely on a screen reader to parse online contents)

- Access to information the alternative text provides if it is provided
- Access to labels and texts in the visualization, if the visualization is in an SVG format

Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard. This includes Pakistan (81%) and the U.S. (77%). It also includes Tunisia (73%), Brazil (69%), India (67%), and Mexico (65%).

-The difference between U.S. and Japan in "hard work would be no guarantee for success" was 20%
-64% of people from Lebanon believe that hard work would not guarantee success, which was the maximum of all countries.
-Italy and Greece have the same percentage (43%) who believe that most people can succeed if they are willing to work hard.
[...]

Figure 1: Our system aims to support data-driven article readings for people with visual impairments. We use an active learning approach to learn individual users' preferences to generate personalized data facts presented alongside a data-driven article.

ABSTRACT

Data-driven news articles are widely used to communicate societal phenomena with concrete evidence. These articles are often accompanied by a visualization, helping readers to contextualize content. However, blind and low vision (BLV) individuals have limited access to visualizations, hindering a deep understanding of data. We explore the possibility of dynamically generating data facts (texts describing data patterns in a chart) for BLV individuals based on their preferences to aid the reading of such articles. We conduct a formative study to understand how they perceive system-generated data facts and the factors influencing their preferences. The results indicate the preferences are highly varied among individuals, and a simple preference elicitation alone induces noise. Based on the findings, we developed a method to personalize the data facts generation using an active learning approach. The evaluation studies demonstrate that our model converges effectively and provides more preferable sets of data facts than the baseline.

CCS CONCEPTS

• **Human-centered computing** → **Visualization**; **Visualization systems and tools**;

KEYWORDS

Accessibility, Data-driven article, Personalized data facts.

ACM Reference Format:
Yanan Wang and Yea-Seul Kim. 2023. Making Data-Driven Articles more Accessible: An Active Preference Learning Approach to Data Fact Personalization. In *Designing Interactive Systems Conference (DIS '23), July 10–14, 2023, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3563657.3595986

1 INTRODUCTION

As the world becomes data-driven, data journalism has emerged as a new standard for communicating societal phenomena. In fact, major media outlets publicly share the practice of using data in news articles (e.g., [1]); technical reports and books introduce principles around authoring data-driven articles (e.g., [10, 11]). One prominent strategy is the use of visualizations, as highlighted by Gray et al. in their popular handbook [10], where visualizations are described as "the workhorse of data journalism."

Sighted readers can benefit from visualizations in data-driven articles in many ways. They may be able to draw their own conclusions or examine other aspects of the data that the article does not

DIS '23, 10-14th July 2023, Pittsburgh, Pennsylvania

ACM ISBN 978-1-4503-9893-0/23/07…$15.00
https://doi.org/10.1145/3563657.3595986

cover. Also, the reader might be able to navigate the news article in a "Martini Glass fashion" where first a reader follows the provided narrative via the article text and other devices and explores the aspects of the visualization they consider most interesting later [66].

However, these tasks are not accessible to blind and low vision (BLV) individuals who rely on screen readers to parse information online. They can only *access* the data by reading the article text, the visualization's alternative text if provided, or text labels if the visualization is in an SVG format. Even with carefully crafted alternative text, it is often challenging to comprehend the various patterns of the underlying data since most alternative text describes the most salient patterns or provides a simple overview of the visualization. When inquired, BLV individuals expressed the desire to conduct similar tasks as what sighted people would do when they encounter visualizations, such as probing aspects of the data not described in the article or inspecting the claims of the author and drawing their own conclusions [38].

To support these needs further, we explore the possibility of dynamically generating *data facts* to aid the reading of data-driven articles for BLV individuals based on their preferences. Data facts refer to a textual description of the result of one or more statistical functions applied to the data used to create a visualization [73]. Since BLV individuals have limited access to various aspects of data, we wish to present extra data facts beyond what the article offers to enhance data accessibility. However, as illustrated in Figure 2 (a), the space of possible data facts can be vast, as it results from the combination of all data cells with the various types of data facts. Given the size of the data fact space, we need a method to *rank* the generated facts so that a handful of facts users might get the most benefit out of should be presented. Prior works propose to rank the facts based on several *factors* such as the statistical significance (e.g., prioritize facts that describe a sharp increase over facts, not a monotonic trend) and the impact of the data fact [72, 77]. Factors allow parameterizing the fact space by surfacing high-level characteristics of data facts. However, the factors used in the prior work may not be sufficient to support BLV individuals' needs and preferences. Therefore, we conducted a formative study with 11 BLV participants to understand how they perceive the value of system-generated data facts and their preferences on data facts and factors. Our result shows that participants were excited about the idea, stating that such a system could help them further engage with data-driven articles. We found that participants' preferences toward which data facts varied dramatically from user to user, thus motivating a *personalized approach* when ranking data facts for a user. Also, in addition to the four factors that we devised informed by prior work and hypothesis, several participants proposed an additional factor related to personal contexts. As we observed an elicitation noise in this study, we also decided to improve the robustness of our system by learning their preferences through active learning.

We build a system that ranks data facts based on a user's preferences given a data-driven news article. Our system uses high-level factors and features to model the data fact space effectively derived from prior work and our formative study. Our system uses a batch active learning model to learn individual preferences. It updates a user's preferences based on several example-based pair-wise queries answered by the user. The queries are formulated using maximum volume removal in combination with successive elimination [8] to maximize the learning from each query and minimize the number of the user's inputs. Our system further allows users to inform the model while reading a news article after the initial training to curate their preferences continuously. Our performance evaluation demonstrates that our model could learn users' preferences with less than ten batches of queries. User study results indicate that our model provides a preferable set of data facts compared to a set generated by user-indicated weights alone and a set that is randomly generated.

Our contribution includes a system that generates data facts based on BLV individuals' preferences. Our system offers dynamic support tailored to individual users while reading data-driven articles. Our system improves the accessibility of data by providing personalized data facts. Through our formative study, we contribute an empirical understanding of individual variances in data fact preferences and a set of requirements for preference-based data fact generation. Through our evaluative studies, we demonstrate the feasibility and the performance of the personalized model.

2 BACKGROUND & RELATED WORK

2.1 Supporting data-driven document reading

Narrative visualizations [66] can support the reading of data-driven documents by conveying stories around data with visuals. The visualizations surface patterns that the story describes, often accompanying annotations on the top of the visualizations, aiding readers to contextualize data with the story and vice versa. A great number of prior works explore how to design an effective narrative visualization (e.g., [12, 36, 49, 54, 63, 81]) and how to automatically generate the annotation overlay on top of the visualizations (e.g., [28, 36, 45]). In the context of a data-driven document, identifying links between the data and the corresponding text facilitates the development of useful features [42, 45, 46]. Several papers try different interfaces to display the identified link to support both the author [74] and the reader [4, 47, 55]. They find that providing the link between the text and the chart can help participants understand the document better. Some other works also explore the use of familiar analogies to help readers contextualize measurements and complex statistical information while reading data-driven documents [37, 43, 44].

These prior work assumes that the user can *see* the visualization and leverage their visual perceptions to support readings and understanding the article. Our work supports *BLV individuals' data-driven article readings* by providing additional contexts around visualizations and the underlying data.

2.2 Generating data facts & description

A data fact refers to a statement that illustrates "patterns, relationships, or anomalies extracted from data under analysis" [17]. Many automated systems are proposed to generate various types of data facts based on their context via Natural Language Generation (NLG) using template-based approaches [17, 21, 72, 73, 77] or deep neural networks [16, 35, 59]. Several taxonomies have been proposed to classify the type of data facts. Chen et al. [17] propose a taxonomy consisting of 12 data fact types based on a literature survey on visualization task taxonomies, user studies, and domain experts'

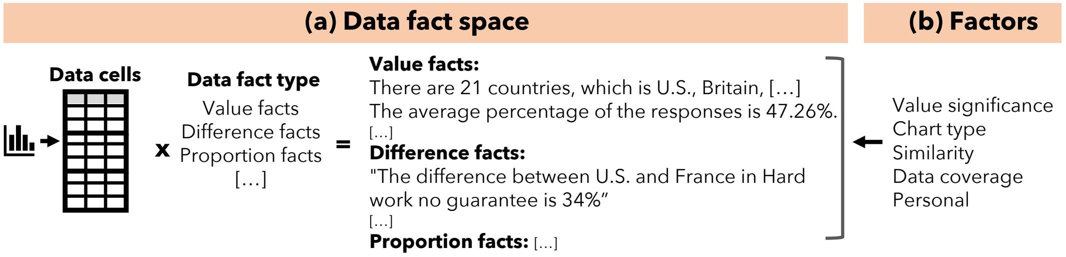


Figure 2: (a) The data fact space is created by conjugating all possible data cells combined with different fact types. (b) The factors to model the fact space, informed by prior work and formative study.

feedback. Recent work done by Wang et al. [77] refines this taxonomy by contextualizing the existing taxonomy with real-world examples, resulting in 11 categories. In addition to presenting the taxonomy, they propose a pipeline, DataShot, to generate a fact sheet (i.e., a set of data facts) automatically from tabular data [77]. Most recently, Calliope, a system, further streamlines the type of data facts based on the redundancy of fact definitions to generate data stories automatically [72] and ChartStory [80] focuses specifically on how to design the interface to present a set of charts based on clustering them with chart specifications. In our work, we followed Calliope's approach to generating data facts. We considered their proposed approaches, in addition to factors that we identified from other prior work and the formative study, to rank the data facts based on individual users' preferences. Also, none of the prior work offers a solution tailored to BLV individuals where some unique considerations should be applied, such as taking into account whether the data is covered by the article or not.

2.3 Making visualization accessible via text description

Recently, Marriott et al. [53] analyzed the current state of visualization for people with disabilities and raised the need to investigate multiple modalities for data presentation and develop automatic tools to improve the accessibility of data visualizations so that people with disabilities could consume them. Among others, describing visualizations with *text* is one of the economical and effective ways to convey visual information to BLV individuals. Prior work has investigated how to support BLV individuals when interacting with visualizations using text modality [14, 26, 27, 30, 58, 69]. Sharif et al. [68] conduct an empirical study exploring the experiences of screen-reader users when encountering online data visualization. The study reveals the lack of support for screen-reader users on online data visualization and raises several design recommendations implied from their study. They suggest that screen-reader users should be provided with both a holistic view of the data and drilled-down exploration given an online data visualization. In addition, because of the missing or unstable quality of the alternative text of visualization, they strongly suggest that alternative text could be auto-generated with the underlying data and be personalized to individual preferences. General guidelines [38], conceptual model [51], and interactive tool (e.g., VoxLens [70], SeeChart [2], ChartVi [57] are proposed to provide better alternative text to support the accessibility of visualization for BLV individuals. While the usefulness of some aspects of semantics and elements in the text description has been explored, none of the work aims to support BLV individuals when reading data-driven articles that include accompanying text descriptions. These descriptions might provide guidance on how to interpret visualizations, highlight crucial information, and determine what details may be unnecessary to elaborate upon. Also, none of them take the personalized approach to tailor to the individual's needs.

2.4 Elicitating preference & learning preference

Preference elicitation and learning *weights* (importance) of factors based on the users' responses have been investigated in various academic communities, including information retrieval, machine learning, and behavioral economics (e.g., [24, 25]). As opposed to directly eliciting the weights from users, *query-based learning approaches* leverage active learning, a machine learning approach, where the model iteratively queries users to label and learn from the labels. Prior works deployed the technique in the preference learning context with various models, including random, best N, max diversity, and max novelty [29] and balancing exploitation and exploration [13]. Biyik and Sadigh [9] investigated users' preferences using *batch* active learning of a reward function. They employ pair-wise queries to ask users to provide their preferences. Within the visualization community, some applications proposed to learn users' preferences on *data attributes* to support data exploration [15, 31, 60, 71, 75, 76]. For example, Podium ranks the data attributes based on the user's interaction and infers the true weights using the Ranking SVM.

In our work, we built upon the query selection method suggested by Biyik and Sadigh [9]. We further devised adjusted pair-wise queries (beyond pair-wise ones often used) by hypothesizing the noise reduction, and we explored how it performs compared with the pair-wise method. Contrary to most prior works related to active learning, which were evaluated with simulation, we evaluated our approach with users in a practical scenario.

3 FORMATIVE STUDY

We conducted a formative study to learn about BLV individuals' needs and preferences on a system that can generate data facts for them while reading online articles. Specifically, the goal of the study was to understand 1) how BLV individuals perceive the value of a system that generates extra data facts beyond the article text, 2) which data facts they prefer to see alongside articles, 3) which factors they value more when the system needs to rank data facts for them, and 4) any other factors that BLV individuals foresee their usefulness. Our study includes open-ended responses and ranking elicitation regarding participants' preferences.

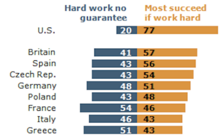
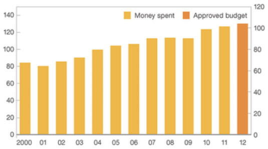
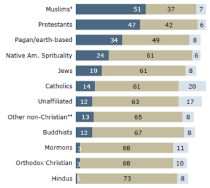
Article ID	Visualization	Alternative Text	Article Contents
A1		A bar chart depicting people's views of hard work and success in various countries, including the U.S., Britain, Spain, etc. Bars extending to the left side colored in dark blue represent hard work no guarantee, bars extending to the right side colored in orange represent most succeed if work hard. Bars are organized by country.	Half or more in 13 of the 21 nations surveyed believe that most people can succeed if they are willing to work hard. This includes Pakistan (81%) and the U.S. (77%). It also includes Tunisia (73%), Brazil (69%), India (67%), and Mexico (65%).
A2		A bar chart depicting European Union budgets over time since 2000. The x-axis ranges from 2000 in 1 year increment to 2012. The y axis on the left ranges from 0 to 140 billion, Euros in increments of 20. The y axis on the right ranges from 0 to 120 billion, Pounds in increments of 20. Yellow bars represent money spent and orange bars represent approved budget.	The 2013 EU budget is 132.8bn euros (£108bn; \$176bn). It is a 2.4% increase on the 2012 budget. The figure was reached in mid-December after much wrangling between the European Parliament, member states' governments and EU Commission. The Commission says it is 5bn euros below the draft budget, so there is still a risk of a funding shortfall. Over the past decade, the EU budget has risen considerably, as the graph below shows. Twelve countries have joined the EU since 2000. […]
A3		A horizontal stacked bar chart depicting the percentage of people's response to which religions are growing, staying the same, or shrinking regarding their size. Growing is represented in blue, staying the same is represented in beige, and shrinking is represented in light blue. The stacked bars are organized by religions, including Muslims, Protestants, Pagan/earth-based, Jews, Catholics, Buddhists, etc.	For nine of the 12 religious groups considered, however, a solid majority (61% or more) of chaplains answering the question report that the size of each group is stable. And for several religious groups, the chaplains are as likely, or even more likely, to report shrinkage as to report growth. For example, one-in-five chaplains answering this question (20%) say that the number of practicing Catholics behind bars is shrinking due to switching, while 14% say the ranks of Catholics are growing. […]

Figure 3: The visualization, the corresponding alternative text, and the article text used in the study. All participants examined all three stimuli.

3.1 Participants

We solicited the study using the emailing list of organizations serving BLV individuals. Our recruitment criteria were 1) at least 18 years old, 2) legally blind, and 3) screen-reader users. Eleven participants reached out to us. Once participants agreed to participate, we sent the Qualtrics[1] link that contains the study material. Participants completed the study by themselves. Five of them self-identified as female and six as male. The ages ranged from 25 to 53 (M=36.9, SD=9.86). The study took 38.8 minutes on average, with a standard deviation of approximately 21.5 minutes. Participants were compensated with a $20 gift card for their participation.

3.2 Data facts and factors

In our study, we used the fact types by Shi et al. [72], which includes nine types after excluding the distribution fact type since they are visually described (e.g., the distribution of the data is [an image of histogram]).

- **Value facts**: Facts related to individual data values or the aggregated value. "The average percentage of people who believe hard work doesn't guarantee success is 47.26%."
- **Difference facts**: Facts related to the difference between categories. "The difference of percentage of people who believe hard work doesn't guarantee success between U.S. and Britain is 21.0%."
- **Proportion facts**: Facts related to the proportion within and across categories. "The number of respondents who believe hard work doesn't guarantee success in Spain is 43.0% of the total respondents."
- **Ranking facts**: Facts related to ranking. "The top three countries where the most people believe they will succeed if they work hard are Tunisia, U.S., and Pakistan."
- **Categorization facts**: Facts related to categories presented in the data. "There are two response categories which are *Hard work no guarantee* and *Most succeed if work hard*."
- **Trend facts**: Facts related to trends over time. "The percentage of people who believe they will succeed if they work hard has a decreasing trend over the years between 2018 and 2021."
- **Extreme facts**: Facts related to minimum and maximum. "The minimum percentage of people who believe hard work doesn't guarantee success is 15.0% over all countries."
- **Association facts**: Facts related to the correlation between two variables. "The Pearson correlation between the year and the percentage of people who believe they will succeed if they work hard is -0.68."
- **Outlier facts**: Facts related to the outliers in the data. "The percentage of people who believe hard work doesn't guarantee success in Pakistan is detected as an outlier."

The following are factors that characterize the data fact space. We included a factor (value significance factor) from prior work directly related to data fact generation [72, 77], two factors (chart type factor, similarity factor) identified visualization perception experiments and one factor (data coverage factor) based on our assumption of their usefulness in supporting BLV individuals.

- **Value significance factor** is related to the statistical significance and the impact of the data cells that a data fact covers [72, 77]. For example, "The Pearson correlation between the budget and the year is 0.99." has a higher value significance than "On average 42% of the respondents think that hard work will guarantee success. " as the former example contains a higher significant number (0.99) and covers more cells (budget and year).
- **Chart type factor** measures the intuitiveness of data facts given the visualization type (e.g., bar chart, line chart). Prior work indicates people prefer different chart types based on the tasks [62]. For

[1] The formative study data for this paper was generated using Qualtrics software. Copyright © 2022 Qualtrics. Qualtrics and all other Qualtrics product or service names are registered trademarks or trademarks of Qualtrics, Provo, UT, USA. https://www.qualtrics.com

example, people might be more interested in learning proportional facts than other types given a pie chart.

- **Similarity factor** is related to how the data fact is similar to the data fact type manifested in the article texts, approximated by a similarity measure. Similar to the exploration and exploitation problem that an algorithm can encounter [20], users can also face the problem of balancing the breadth and depth of new information. A user might prefer to see more data facts constructed similar to what is already presented to explore the breadth of the dataset, or they might prefer to examine data from a different angle (i.e., different data fact type).
- **Data coverage factor** is related to whether the data point(s) that the data fact describes is mentioned in the article or not. Prior work indicates that BLV individuals wish to examine the data that are not mentioned in the article [38]. We envision providing information not mentioned in the article will add value to the BLV individuals' understanding of data.

3.3 Study design

Part 1: Demographic information We first asked for demographic information, including gender, education, and occupation. We also asked whether they have the functional vision or light sensitivity, their diagnosis, on-set age, and stability of the condition.

Part 2: Open-ended examination of visualization and text pairs In this part, we sought to understand which aspect of the data participants wanted to learn when reading data-driven articles. To situate participants in a real-world visualization reading scenario, we prepared three visualization and article pairs from Kong et al. dataset [46]. We prepared an alternative text for each visualization based on prior research [38]. However, to reduce the confounding effect of their perception of data facts, we removed the description of data trends from the alternative text. The visualizations, the alternative text, and the article text used in the study are shown in Figure 3. Participants were prompted to examine each visualization on a separate page. Each page contains a visualization and the corresponding article paragraph. To make the procedure smooth for participants with their screen readers, we added text before and after introducing the visualization and the article paragraph to help participants understand the structure of the page and locate themselves (Details can be found in the supplementary material). Then, we asked a question related to what other data-related information beyond the given article text they would like to know more about. We later analyzed these open-ended questions to see whether the existing data fact types were enough to fulfill what participants wanted. We did not introduce any notions of data fact or the factors in this stage to avoid priming the participants.

Part 3: Understanding preference & the perceived value Part 3 is designed to learn BLV individuals' preferences on data facts and factors as well as to understand BLV individuals' overall perceived usefulness of a system that offers additional data facts beyond those given in the article. We first explained to BLV individuals the different types of data facts with examples (Details can be found in the supplementary material). To reduce the cognitive load of understanding the concepts, we provided examples for each fact type using the datasets they examined in Part 2. Then, participants were asked to elaborate on what types of facts they preferred to see when they read any data-driven articles. We allowed participants to select multiple data fact types and elaborate on the reasons.

Next, we presented information related to the four factors, describing them using the examples in Part 2 when possible. Then, we prompted them to think about each factor. We asked questions related to the *directional preference* of two factors: similarity factor and data coverage factor. While it is apparent that valuable data facts should be statistically significant and intuitive given the visualization type, other factors like similarity and data coverage may depend on a user's preference. For example, some users might want to know more about the data presented in the article (negative weight for data coverage factor), while others may want to see some facts related to unmentioned data (positive weight for data coverage factor). Also, some users might prefer data facts structured similarly to the presented article (positive similarity factor), whereas other users might prefer otherwise (negative similarity factor).

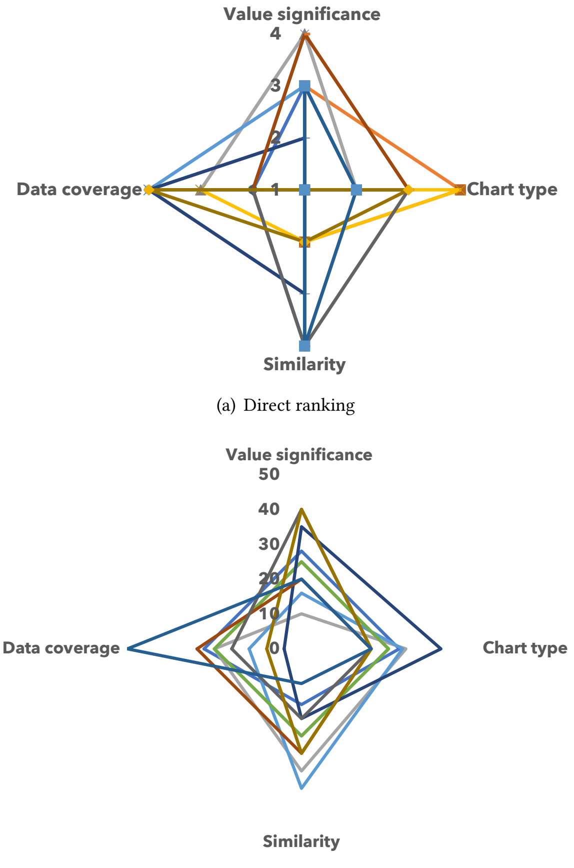
After prompting the directional preferences, we asked participants to rank the four factors in order of importance for data fact generation. We utilized two popular methods [24, 41] to compensate for the noise of the elicitation. First, we asked the participants to rank the factors using integers between 1 to 4 (1 being the most important and four being the least important). We also asked them to allocate 100 balls to these four factors based on their importance. We explained that they could enter 25 each if they equally valued each factor. We also mentioned the response doesn't need to sum up to exactly 100. While it might be difficult for participants who do not have expertise in data analytics and visualization, we asked them to describe *other factors* that may influence their preferences for data facts in case they have any.

After these questions, we asked participants to indicate the perceived usefulness of the system that generates extra data facts based on their preferences and willingness to use the system using the 7 Likert Scale (Extremely useful-extremely useless, extremely likely-extremely unlikely) and describe the rationales for their rating. More details are described in the supplementary material.

3.4 Results

3.4.1 Is the data fact taxonomy enough? We analyzed the result of the open-ended question from Part 2. Participants were not exposed to any data fact and factor taxonomy when they answered this question. Two coders reviewed the responses and classified them with the taxonomy suggested by Shi et al. [72]. If none of the types can be applied, the coder did not classify the response. Later, the two coders discussed their results until 100% agreement. The final result shows that 90% (30 out of 33 responses) of the responses fell into one of the data fact types suggested by Shi et al. [72] (Fig. 5), indicating that the existing taxonomy covers most of the data facts participants desired to know. The responses unable to be classified were about the data or visualization themselves, instead of data fact related. P10 stated they wished to know more about "specific measurements for the x and y-axis" and P11 was curious about "types of data."

3.4.2 Preference on data fact types. After the open-ended questions, participants were asked to read data fact taxonomy and choose the type of data facts they might be interested in knowing. Figure 5 shows the result. 7 out of 11 participants prefer to see extreme facts,



(a) Direct ranking

(b) Assigning balls

Figure 4: The result of direct ranking and the assigning balls approach of the four factors by each participant. Each line represents a participant. It shows a large variance among individuals.

followed by ranking facts (6/11). Value, Differences, and Trend facts were equally preferred by participants (5/11). No participants indicated they wished to see outlier facts. Their preferences varied in terms of the types of facts and the number of different types of facts they were interested in.

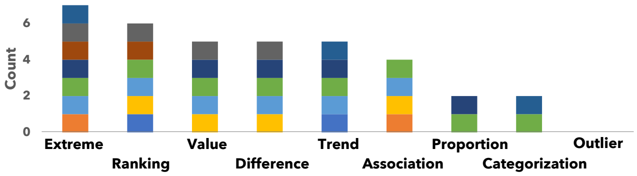


Figure 5: The result of data fact preferences. Each color represents a participant.

3.4.3 Preference of the factors. First, we analyzed the directional preference for the two factors (similarity and data coverage factor).

Directional preferences for similarity factor: 8 out of 11 participants indicated that they prefer to see more data facts of a similar type to those mentioned in the article. 3 participants indicated otherwise; they prefer to see more data facts that are different fact types than the article contained.

The directional preferences for data coverage factor: 6 out of 11 participants indicated that they prefer to see more data facts illustrated in the data that are not mentioned in the article, whereas 5 participants indicated otherwise.

Elicited ranking of the four factors: We elicited the preferences in two ways: direct ranking using numerical values ranging from 1 to 4 and assigning 100 balls to each factor. When we analyzed direct ranking answers, each factor's average ranking was similar. For example, the mean ranking value for the value significance factor and chart type factor was 2.4, the similarity factor was 2.3, and the data coverage factor was 2.5. These results indicate that all factors are similarly important on average. However, we observed a large individual variance. Figure 4(a) shows how individuals' ranking on each factor varies in their preferences. This large variance suggested that applying fixed weights for the factors (e.g., the average value) to generate data facts may not satisfy many users. A similar analysis conducted on the responses to the ball assignment elicitation technique corroborated these findings (Figure 4(b)). As expected, the results of the two elicitation methods were not always aligned. For example, P4 indicated that they preferred the similarity factor over the data coverage factor when we elicited directly but assigned more balls for the data coverage factor with the latter method. The existence of this misalignment underscores the presence of potential noise when participants attempt to express their preferences, thereby impeding their ability to articulate their true intentions and posing challenges for us to accurately capture them.

Other suggested factors While most participants indicated that the four factors presented were sufficient to cover their preference, two participants stated another factor, which P1 called the "personal interest" factor. P1 shared that data facts would be more valuable if they had "geographic relevance to me". P3 also stated, "people will always care about something more if it personally relates to them."

3.4.4 Perceived usefulness and willingness to use. Participants were very positive about the potential of a system that would generate personalized data facts catering to their preferences. On average, the rating was 6.2 (out of 7, SD=0.8) when asked to rate the usefulness of a system that provides additional data facts based on their preference. P6 mentioned that the system would be useful because it would allow "*to have equal access to information with sighted people and to be able to engage with the information in the article better.*" P11 echoed this sentiment: *"Some articles do not contain all the facts I would like to know about the topic I'm reading about."* When we asked to rate how likely they would use the system if it's available, the majority of participants indicated that they were likely to use it (M=6.3, SD=1.0).

4 SYSTEM: MAKE DATA-DRIVEN ARTICLES MORE ACCESSIBLE

We designed a system that makes data-driven articles more accessible by generating personalized data facts based on user preference.

4.1 System requirements

We derive system requirements based on study findings.

- **Personalize weights of each factor for a user**
In the study, we observed participants' preferences toward factors vary individually, including their rankings (e.g., which factor they prefer the most) and directional preferences of two factors. This finding indicates that the system should accommodate individuals' different preferences when generating data facts.
-> Our system should provide data facts generated by personalized weights on each factor to serve an individual's unique preferences.
- **Personalization factor**
Two participants indicated they would like to see data facts generated by considering their own circumstances. For example, participants wanted to prioritize data facts related to their own location/region. Prior work [38] also indicates that BLV individuals would like to locate data values for their own city and state when they examine a map.
-> Our system should include a factor that can reflect people's personal condition.
- **Better elicitation method**
In the study, we asked about participants' preferences in two different ways that can reflect their perceived importance of the factors. Two participants' responses were irrational (e.g., fewer balls being allocated to the factor identified as the most important). These inconsistencies clearly show the responses obtained from elicitation methods can be noisy.
-> Our system should use a better way of understanding the user's true weights of each factor, such as by implementing a learning approach.

4.2 System workflow

The system is assumed to get two pieces of information investigated by prior work. First, the system requires to get the data extracted from the visualization. ReVision [64], ChartSense [39] and Scatteract [19] offer to extract the data from a visualization formatted as bit map images and other proposed systems (e.g., [33]) can provide the data from the visualization. Second, the system requires to have a link between data and the article paragraph pairs [46]. Kim et al. extensively investigated the methods to achieve this goal [42]. Given the data and the corresponding article paragraph parsed from a data-driven article, the system follows the following five steps to provide a personalized set of data facts.

4.2.1 Formulating data facts. Based on the approach suggested by Shi et al. [72], the system generates all possible facts from the given data table (Fig 2(a)). First, the system classifies the values of each column into three categories, numerical ('N'), categorical ('C'), and temporal ('T'). Then the system extracts unique values of each column and makes all possible combinations of those values among non-numerical columns (i.e., categorical and temporal columns). Then, the system uses them as *filters* to create subspaces. The different subspaces of the data are associated with unique values in non-numerical columns or combinations of unique values across all possible pairs of columns. To improve computational efficiency, subspaces containing only one value are eliminated (their record is reflected in value facts). From each subspace, the system computes values for all data facts (e.g., values, differences, proportions, trends, etc.) and generates data facts based on all five aggregation methods (i.e., count, sum, average, maximum, and minimum). We used the natural language generation template from prior work [72]. We further tweaked some of the templates to make the sentences more natural.

We took the following considerations:

Category expression: When referring to the columns, the original template use syntax like "There are (count) (column name)(s) which are (column values)." However, if the categorical column name is not a typical categorical column name (e.g., trade provision), readers may not realize they are reading the column name. We added an indication to solve this. For the former example, the syntax will be 'There are (count) (column name) *categories* which are (column values)." In addition, we modified the expression to "(a category in column 1) in (a category in temporal column 2)" (e.g., the stock market in September 2012) to show the category relationship when a temporal category is used together with another category.

Subspace expression: We removed expressions about subspace if the subspace is the entire dataset to make the fact more concise.

Numerical value expression: We used rounded percentages instead of decimals (e.g., 1% instead of 0.01) in the proportion fact.

Aggregation method expression: We used "total" instead of "sum" to make it sound less mathematical. Also, instead of using an expression like "max (column name)," we used "(column name) ... regarding their maximum value" to avoid misunderstanding.

4.2.2 Assigning weights to each data fact. Given a generated data fact, we quantified the value of the data fact regarding the five factors and calculated the score of each data fact by multiplying the data fact value with the weights of the factors. To parameterize data facts with the five factors (including the personalization factor derived from the formative study), we curated features for each factor.

- **Value significance factor:** To represent the factor, we used three quantifications suggested by Wang et al. [77]: statistical significance (i.e., the degree to which the fact entails a large value) , context impact (i.e., whether the subspace from which the data fact is derived from belongs to more or less frequently mentioned categories) and focus impact (i.e., the number of cells covered by the subspace corresponding to a data fact). All three quantities take values between 0 and 1. Following [77], the value significance factor is summarized as a weighted average of the three quantities, with weights of 0.6 for statistical significance, 0.2 for context impact, and 0.2 for focus impact.
- **Chart type factor:** To quantify the factor, we consulted with prior work that measures the effectiveness and preferences of analytical tasks based on chart type [62]. This prior work uses a task taxonomy suggested by Amar et al. [3] to investigate how different chart types can effectively support the tasks. We bridged the taxonomy and the data fact taxonomy that our system used [72] by comparing its similarity in functioning (e.g., retrieve value ↔ value fact, find anomalies ↔ outlier fact, find extreme ↔ extreme fact). The two taxonomies are aligned with one exception of determining range. We used finding extreme and determining range tasks to match with extreme facts. Then, we used the preference rankings derived

from [62] to calculate how a data fact type can be preferred in different chart types. We normalized the values of the rank of a chart type of each fact type with the min-max normalization to scale the range of the importance distribution to [0, 1]. Figure 6 shows the weight of different fact types in different types of charts.

- **Similarity factor:** To calculate the factor, the system constructs a Word2Vec representation of each sentence in the article and each data fact by averaging the Word2Vec vector of each word contained in the sentence and the data fact, respectively. Then calculate the cosine similarity between the aggregated Word2Vec representation of each sentence in the article and the data fact to quantify the factor. We used pre-trained Word2Vec embeddings that contain 300-dimensional vectors for 3 million words and phrases trained on the part of the Google News dataset, with approaches provided in [56].
- **Data coverage factor:** To quantify the factor, the system takes the percentage of data cells covered in the data fact covered by the original article. To establish links between cells and sentences in the original article covers, we used a method from Kim et al. [42]. The system subtracts this value from 1 to make the higher value means less coverage.
- **Personalization factor:** To calculate the factor, the system computes the cosine similarity between Word2Vec vectors of the data fact and the user's location string. We get the user's location by analyzing their IP addresses. As Word2vec vectors are represented with their semantics, we assumed that this similarity reflects how the data fact is related to the location.

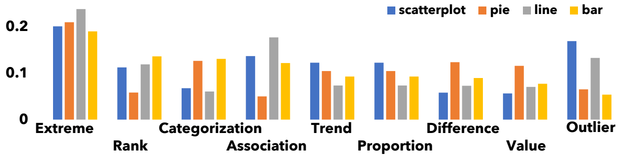


Figure 6: Quantification derived from Saket et al. [62]. The normalized values are calculated by data fact type and chart type.

Calculating the final score for each data fact Each data fact (x) is represented by a vector consisting of five values from the five factors, and the system predicts individual user's data fact preferences through a personalized linear model (i.e., by learning a 5-dimensional vector w for each user) to weight the importance of the five factors. Once trained, the system predicts how likely a user, represented by the learned weights w, prefers a data fact, represented by a feature vector x, through the dot product $\sum_{k=1}^{5} w_i x_i$. This score represents the user's preferences for each data fact. A data fact with a higher score is more likely to be preferred by the user.

4.2.3 *Learning true weights from users.*

To learn about users' individual preferences, we applied an active learning approach to reduce users' effort as much as possible while still learning a stable distribution of their preferences.

Step 1: eliciting initial weights We directly elicited each user's preferred weights for the five factors. The answers from direct elicitation contain some signals, but they can be noisy. Thus, the system uses these answers as the initial weights for the personalized model. For the two factors that users show a variance in the directionality, we also asked users to provide an initial direction (i.e., whether to prioritize or downplay a data fact based on the factor).

Step 2: solicit labels from users To train a personalized model, we need labeled data indicating a user's preferences. The goal is to minimize the number of user interactions to minimize the user burden while maximizing learning.

Our system uses a maximum volume removal approach [8] to achieve these goals. This approach uses a pairwise query to prompt users' preferences. More specifically, two data facts (we will refer to this as a query hereafter) can be given to the user, which is prompted to choose the preferred option. The model is then trained based on this response. Formulating the right set of queries is crucial to maximize learning and minimize the number of queries required to prompt the user. In other words, depending on which query the system asks, the learning outcome will be drastically different. Our system constructs query sets that maximize the expected information obtained from the queries.

Specifically, the system considers all possible pairs of data facts in the data fact space. For each specific pair, e.g., composed of data facts A and B - represented by x_a and x_b, we compute the difference between the five factors, denoted as d.

$$d_{AB} = x_A - x_B \tag{1}$$

Then, based on the current user weight w, the probability that option A will be chosen is modeled with the following equation.

$$p\left(I_{AB} \mid w\right) = \min\left(1, \exp\left(I_{AB} w^T d_{AB}\right)\right) \tag{2}$$

Conversely, the probability that option B will be chosen is modeled with the following equation.

$$p\left(-I_{AB} \mid w\right) = \min\left(1, \exp\left(-I_{AB} w^T d_{AB}\right)\right) \tag{3}$$

I_{AB} is the $sign(w^T d_{AB})$. I_{AB} indicates which option (A or B) should be selected ($I_{AB} = 1$ for option A and $I_{AB} = -1$ for option B) according to the current model.

To identify which queries maximize learning, we compute how much the model's uncertainty could be improved based on the query's response. The amount of potential uncertainty reduction is captured by the information entropy of each data fact pair:

$$U_i = \min\left\{E\left[1 - p\left(I_i \mid w\right)\right], E\left[1 - p\left(-I_i \mid w\right)\right]\right\} \tag{4}$$

In other words, U_i quantifies how much information the system could gain by asking a specific pair. One criterion to choose the *batch of queries* that maximizes learning is to select queries the model is most uncertain, i.e., queries for which U_{AB} is the largest.

However, greedy selection based on only model uncertainty alone could be suboptimal because a batch may contain redundant queries. To further improve the efficiency, the system uses the successive elimination approach suggested in prior work [8]. The goal of the method is to maximize conditional entropy.

$$\max H\left(d_1, d_2, d_3, \ldots, d_n \mid w\right) \tag{5}$$

Successive elimination starts from a large batch of queries (i.e., a batch of data fact pairs) to maximize conditional entropy. Then, it selects the two queries that are closest to each other at each iteration, removes the one that has the lowest information entropy, and iterates this procedure until only the expected number of queries (a small batch that will be provided to the user) remains.

To summarize, the system first shuffles the list of all potential queries and randomly samples 5000 of them. Then, it selects the top 200 queries based on the information entropy U_{AB}. Finally, it filters this set down to 10 queries by applying successive elimination. The number of final queries presented to the user (ten) was determined based on our simulation (refer to Sec. 5.1).

Then the user is presented with 10 batches of queries with a data-related article. The system asks the user to choose their preferred data fact between two options for each query. The model uses this response to label the difference between the two options' vector d.

Step 3: updating weights based on users' responses Once the system receives the responses, the model performs a Bayesian update of the weights based on the following equation.

$$p\left(w \mid I_{AB}\right) \propto p\left(I_{AB} \mid w\right) p(w) \tag{6}$$

Since the distribution of $p(w)$ is unknown, the Metropolis algorithm [32] is used to estimate the distribution of $p(w)$.

4.2.4 Presenting data facts based on personalized weights. Using the learned weights and a given new data-driven article, the system can rank the data facts based on the user's preferences. The system presents the top 10 results, but the user can retrieve more data facts by clicking the "read more" button.

4.2.5 Soliciting labels during daily use. Our envisioned system accommodates the user's new labels. For example, in reading the presented data facts, the user can simply press the *u* key to inform the model the given data fact is useful or press the *n* key to indicate it is not. After receiving these additional inputs, the model will update their weights (Eq. 6) by creating similar comparisons that the system makes for the initial update. Specifically, once the user marks a data fact as useful, the system pairs this marked data fact with all other data facts in the list to calculate d. Similarly, once the user indicates a data fact is not useful, the model considers all the pairs in the article with the marked data fact to update the model. This capability allows the personalized model to keep the model up-to-date with the users' preferences.

5 EVALUATION

We evaluate our system using two approaches. First, we evaluate how fast the model can converge by simulation and how accurate the learned weights are. We simulate the model using the (hypothetical) user responses with varying noise levels. Second, we conduct a user study to evaluate whether our learning approach offers a preferable set of data facts compared to a set generated based on the direct initial weights and a set generated with random weights.

5.1 Simulating model's convergence

We conducted a simulation with two goals: (1) to *evaluate* how fast our personalized model can converge and learn the true weights with a reasonable amount of human noise and (2) to *inform the system deployed in our user study* regarding the number of labels we need to solicit from users to train a personalized model and the elicitation method that reduces the noisy signal.

	Adjust Pair-wise # of batches to converge (alignment value)	**Pair-wise # of batches to converge (alignment value)**
True weights +/- 0	5.68 (0.79)	5.72 (0.97)
True weights +/- 0.05	5.98 (0.84)	6.50 (0.97)
True weights +/- 0.1	7.02 (0.87)	7.22 (0.97)

Table 1: The simulation result. Adjust pairwise method reached the convergence a bit faster but performed a much lower alignment.

5.1.1 Parameterizations. We first designed the noises to replicate a realistic setting and the two strategies for the elicitation.

Human noise We assume that users have internal ground-truth preferences (i.e., true weight distribution). The training process aims to learn these true weights from the user through learning and elicitation. However, users may not always follow their true weights when making choices. In other words, when the user expresses their preferences by choosing the data fact that they prefer, their responses can be noisy. To account for elicitation noise (i.e., the extent to which their choice of data facts differs from their true preferences), we parameterized the noise in our simulation (Table 1).

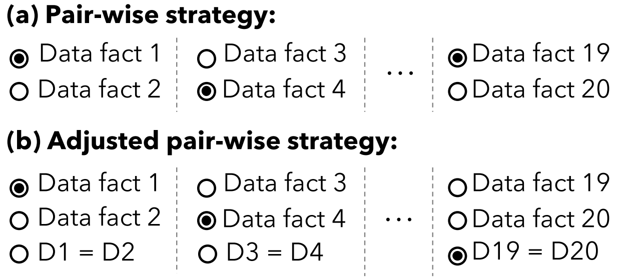


Figure 7: Two different elicitation strategies we simulated.

Elicitation method Active learning methods often solicit user input using a shortlist of options to get labels from agents. In our case, the system asks users to choose one data fact from two presented options to inform the model of their preferences (Fig. 7(a)). In addition to this strategy, we designed an alternative approach, named adjusted pairwise elicitation, where the user is allowed to indicate the two options are equally preferable (Fig. 7(b)). The adjusted pairwise elicitation was designed after we observed that some generated data facts share similar scores. We conjectured that the space of possible data facts is relatively small compared to other domains, leading to many data facts sharing similar scores. We hypothesize that providing the option to indicate the two presented data facts are equally preferable will reduce the noise for learning. In this simulation, we set out to compare the performance of these two strategies and choose the one that induces less noise and helps the model converge faster to use in our user study.

5.1.2 Set-up & measures. We randomly chose ten visualization/article pairs from Kong et al. [46] and simulated the experiment 50 times. We used the equal weights for each experiment for the five factors as their initial weights (i.e., 0.2) and set the true weights as randomly generated weights. Then we generated a batch of queries (a batch consisting of 10 queries) based on our query strategy (Step2: Solicit Labels from Users under Sec. 4.2.3) and simulated the user responses according to the study setup (true weight +/- noise). To measure convergence, we used the metric presented in a prior work [61]. This method proposes the alignment metric of the two weight distributions (e.g., the alignment between the weight distributions at time t-1 and weight distributions at time t, Eq. 7). We determined the model convergence when the similarity between the weights distributions at time t-1 and t (m) is more than 0.99. Once the model is converged, we calculated two evaluative metrics: (1) the number of the responses that the model needs to converge (i.e., t) and (2) the alignment between the learned weight distributions (i.e., weight distribution at t) and the true weight distributions. Metric (1) demonstrates how fast the model can be stabilized (i.e., how few queries does the model need to converge?), and metric (2) demonstrates how the model can accurately learn about the true weights.

$$m = \frac{\boldsymbol{w}_{\text{true}}\, E(\hat{\boldsymbol{w}})}{\|\boldsymbol{w}_{\text{true}}\|_2\, \|\hat{\boldsymbol{w}}\|_2} \tag{7}$$

5.1.3 Results. The results (Tab. 1) demonstrated that, on average, both elicitation methods could converge with a reasonable number of inputs (6-8 batches of queries), even though the adjusted pairwise method required slightly fewer queries. While there is no considerable difference in the number of queries to converge between the two elicitation methods, the alignment metric of the pairwise method was higher than the adjusted pairwise method, indicating the pairwise method learned the true weights more accurately.

5.2 User study: can learning approach outperform?

We conducted a user study to evaluate how our system performs with actual users' input. Since the overall motivation for a personalized system has been validated through our formative study, we focused on the quantitative performance of the system in the evaluative study.

5.2.1 Stimuli & procedure. We used the same ten visualization/article pairs from Kong et al. [46] used in the simulation. We formulated an alternative text based on prior work [38]. Again, we removed any data trends in the alternative text to avoid a potential confounding effect with the presented data facts. We used ARIA radio buttons and ARIA-Labeled buttons to ensure accessibility. Informed by our simulation, we used the ten batches (to be conservative) with the pairwise elicitation method in soliciting labels from users. The overall procedure is illustrated in Figure 8. They were first asked to provide their initial weights of the five factors and the directionality preferences for the similarity factor and the data coverage factor (Fig 8 (a)). After that, we solicited labels to learn their preferences (Fig 8(b). In this stage, participants were asked to read a randomly selected visualization (represented by its alternative text) and the paired article. The interface prompted them to complete 10 batches of queries using the pairwise elicitation methods. These questions were generated based on their initial weights, followed by our query formulation strategies (Step2: Solicit Labels from Users under Sec. 4.2.3). The responses were used to train a personalized model for the participant. Figure 9 shows the structure of the main study page given to a participant (The actual page can be found in the supplementary material). Next, we tested the performance of the personalized model (Fig 8(c)). Participants were prompted to read *two other* randomly selected visualization/article pairs at a time. In each article, participants were asked to choose the one set of data facts they found most useful among the three, which are generated based on the personalized model, initial weights and random weights, respectively. The presentation order was randomized. This set-up of evaluating performance would not allow us to conduct statistical tests due to the aggregated nature of the responses, compared to asking participants to evaluate each data fact using a Likert scale, for example. However, we designed it with two rationales: it is more trustful to our system's mechanism as the system updates the weights after each batch (not after every data fact generation), and the responses would be more accurate since participants need to form an impression about a set of data facts as a whole instead of investigating a single data fact at a time and assign a numerical score to it.

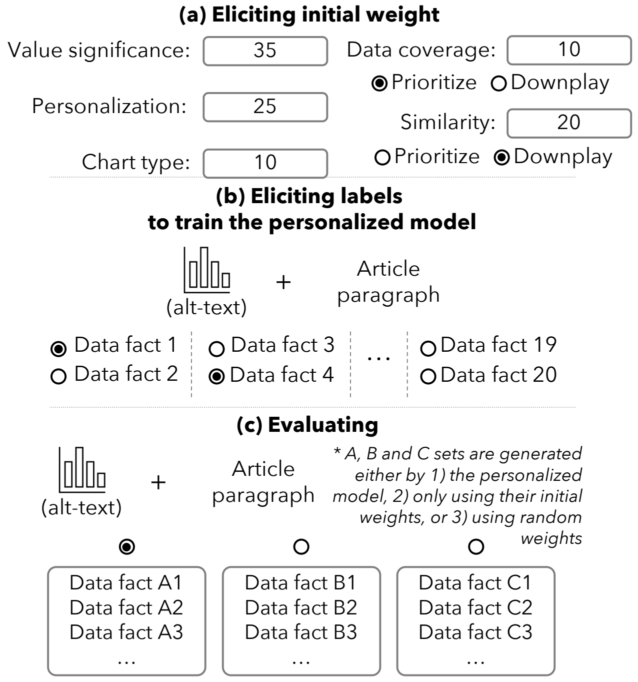
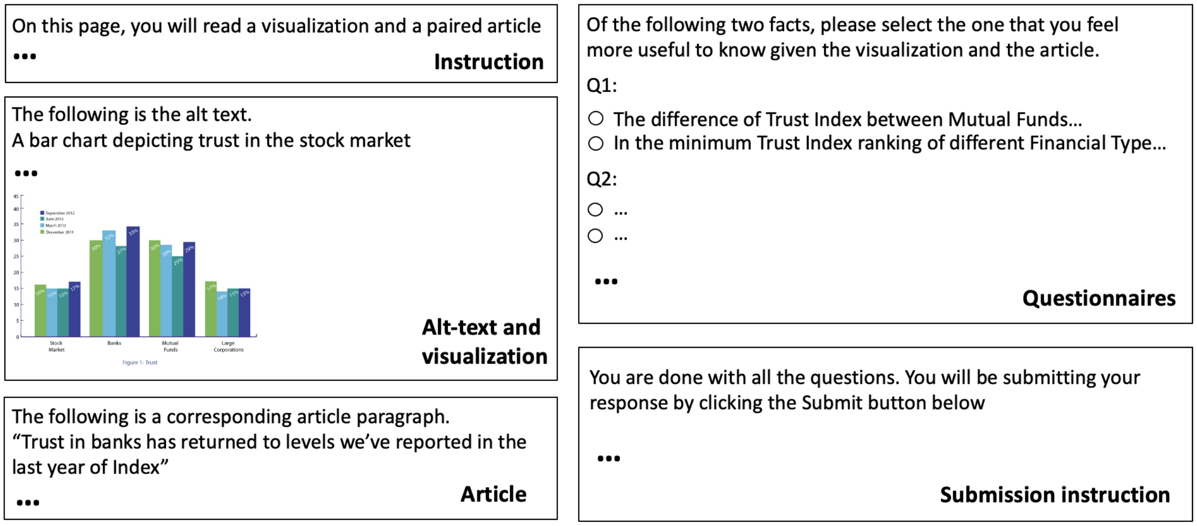


Figure 8: The procedure of the study. (a) asking for the initial weights of the five factors and the directional preferences for the two factors. (b) prompting 10 questions to learn their weights each time. (c) evaluating with three sets of data facts generated by either our personalized model, users' initial weights, or generated randomly.

5.2.2 Participants. We recruited 17 participants through the same mailing list with the same recruitment criteria used in our formative study. The average task time was 32 minutes (SD=18). Participants were compensated with a $25 gift card for their participation. Eleven



On this page, you will read a visualization and a paired article

•••

The following is the alt text.
A bar chart depicting trust in the stock market

•••

The following is a corresponding article paragraph.
"Trust in banks has returned to levels we've reported in the last year of Index"

•••

Of the following two facts, please select the one that you feel more useful to know given the visualization and the article.

Q1:

○ The difference of Trust Index between Mutual Funds...
○ In the minimum Trust Index ranking of different Financial Type...

Q2:

○ ...
○ ...

•••

You are done with all the questions. You will be submitting your response by clicking the Submit button below

•••

Figure 9: The structure of the main page of the study. Participants can access all the contents through screen readers.

of them self-identified as female, and six of them self-identified as male. The age ranged from 23 to 42 (M=31, SD=6).

5.2.3 *Results.* We first analyzed the results from the perspective of usability and performance. We further analyzed participants' weight change after the learning process to gain more insight.

Informal Observation on Usability Our research team was standing-by via email at the assigned time slot for participants, and participants were informed about our availability. We did not get any emails during the participants' sessions, which may indicate the participants did not need help completing the study. Also, participants completed the study in a reasonable time frame, as we piloted (around 1 hour). When we solicited, participants mentioned the interface was usable to complete the given tasks.

Performance We collected two evaluative responses from the participants (i.e., testing on two sets of visualization/article pairs), resulting in 34 data points. Among those, 41% times (14 out of 34), participants chose the set generated by the personalized model (generated using the learned weights). 32% (11 out of 34) of the times, participants chose the set generated by the model using their initial weights. 26% (9 out of 34) of the times, participants chose the sets generated by random weights. Notably, more than 33% of the times (which is equivalent to a random chance), participants chose the set generated by their personalized model.

Comparison between users' initial weights vs. learned weights Figure 10 shows the individual participants' plots visualizing initial weights and the learned weights, sorted by the amount of the gap between those two. We did not find any systematic patterns in the difference between the initial weights and the learned weights by factor. We observed many participants' initial weights differed from their final learned weights. The average difference across all five factors was 0.17 (SD=0.15). The minimum gap that participants exhibited was 0.06 (across all five factors), and the maximum gap was 0.32. The observed difference between the initial and the learned weight demonstrates the effectiveness of the learning

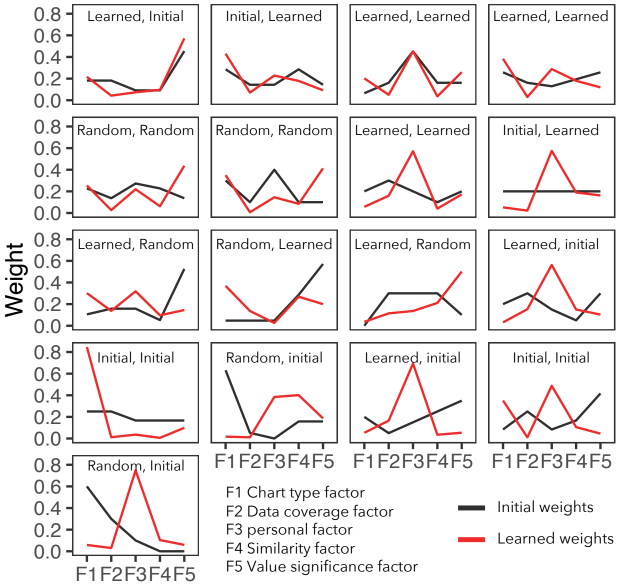


Figure 10: Individual participants' plots that visualize their initial elicited weights and the learned weight. The text annotation in each plot represents their choice of sets in the evaluation phase. "Learned" indicates that of the three provided data fact sets, the participant chose the one generated based on the learned weight as the most useful. Similarly, "Initial" indicates that they chose the one developed with the initial weight they provided. "Random" indicates that they chose the one generated with a random weight. Each participant will examine two visualizations and answer two questions in the evaluation phase, so each plot has two text annotations.

approach as well as demonstrates the inaccuracy of direct elicitation of the preferences. The annotated texts in each plot represent the two sets they chose in the evaluation phase (Fig 8(c)).

6 DISCUSSION

Our formative study findings show that BLV individuals wish to navigate aspects of data when reading data-driven articles and their preferences are highly individualized. Motivated by the study, we built a system that generates personalized data facts using a batch active learning approach. We show the system’s feasibility by demonstrating its accuracy and efficiency via an evaluative user study. We believe our work paves the way for personalizing data fact space using factors that characterize the space at a high level. We envision that many other visualization applications can take a similar approach to further support BLV individuals. For example, the portion of alternative texts that describes the data trend can be dynamically generated based on a user’s needs as their preferences can be varied [51]. Also, we can apply a similar approach to personalize the column order of data tables to improve its accessibility along with other works (e.g., [78]).

In applying batch active learning in the domain of data facts, we proposed a new elicitation method (namely an adjusted pairwise approach) that potentially can reduce the noise when many entities share the same ranking score. While this method did not exceed the conventional method in terms of accuracy, it shows some promising properties regarding minimizing the number of user inputs. We would like to keep working on improving the method to tailor its properties specific to data fact-related domains.

6.1 Envisioning a pipeline combined with the prior work

Our system focuses on designing a system to *generate personalized data facts* given a data-driven article. Combined with prior work, we envision a workflow that allows detecting visualizations in an article to provide data facts based on user preferences. Techniques to detect a visualization [6, 18], its chart type [5, 6, 18, 52, 65] and extract the data from either SVG formatted visualization [6, 23, 33, 34, 67] or bitmap images [5, 7, 18, 19, 22, 40, 50, 52] can feed data and chart type to our system. Techniques to detect the links between the data and the article paragraphs can also be improved using prior work [42, 48, 74, 79].

6.2 Extension for sighted individual

We carried out studies and developed a system to support BLV individuals. However, we envision this system can be easily transferable to support sighted people. One straightforward future work will be tailoring the system more for sighted people by conducting user studies and building the modified system. An interesting next step would be collecting sighted participants’ weights and comparing them with the weight distributions from our study to observe any differences in their preferences. Understanding the difference will be able to inform designers of data communication systems on how they should consider the design differently for BLV individuals.

6.3 Limitations & future work

In this paper, we focused on investigating how we could incorporate people’s preferences in generating data facts and demonstrating the algorithm’s effectiveness. Future work can continue to implement the envisioned system by combining with plenty of existing methods proposed for extracting expected data from a visualization ((e.g., ReVision [64], ChartSense [39], Scatteract [19]).

In our system, we demonstrated *a method to parametrize the data fact space* to personalize the generation process. Future work can extend the list of factors and their featurization. For example, adding more detailed factors regarding the content (e.g., preference on the subject of the paper) may improve user experience. Evaluating different factors and features to measure the impact on users’ satisfaction can be an interesting research direction, which allows researchers to deepen the understanding of what information people want.

In the current system, one of the factors, the personalization factor, uses the user’s geographical semantic proximity. However, the personalization can be extended beyond their location. For example, if the model can access other demographic information (e.g., religion, political party), our approach can easily incorporate the new information.

While our study offers insights into the algorithmic performance as well as usability, future work can explore *how* people use the system by conducting in-depth qualitative studies through video conferencing tool or in-person with which more user feedback can be provided through think aloud technique and observation and thus more insights can be gained to improve the system further. Also, we studied our system with relatively small numbers of people to gain initial insights. The results show the difference in the performance between our approach and the baseline. We expect to see the difference will be larger as the number of observations increases.

Our evaluative user study used the datasets from Kong et al. [46] where all of the examples are bar charts (e.g., simple bar charts, grouped bar charts, stacked bar charts) due to the availability of gold standard datasets, which can limit the generalizability of the system and conclusions. However, considering during the process of learning the weight and selecting the data facts, the system does not take any chart-specific features into account, we believe that our findings from the evaluative study will be generalizable to other data types and other chart types. Future work may consider creating more gold-standard datasets regarding different chart types and generalizing the approach to different chart types.

Regarding the time efficiency of the proposed system, the initial learning process may be somewhat time-consuming, as each weight learning iteration with ten responses takes approximately 40 seconds. However, it is important to note that this learning process only needs to be executed once in order to obtain satisfactory post-training user preference weights. With the intention of facilitating long-lasting updates during daily use, it is proposed that new updates can be processed in the background and applied upon completion rather than instantaneously, thus rendering the algorithm’s time requirements tolerable. Furthermore, it should be noted that all computations were performed on a CPU. As potential avenues for future research, the possibility of migrating the process to a cloud-based infrastructure and employing GPU or distributed computing to minimize user wait times could be explored. Additionally, future studies may seek to evaluate the system's utility by examining the extent to which users value the generated data insights in relation to the time invested.

7 CONCLUSION

We proposed a system that generates personalized data facts for data-driven articles. We explored factors that may influence BLV individuals' perception of the usefulness of the data facts. The results from the study showed that people's preferences are varied, leading us to build a personalized data fact generation system. We demonstrated the feasibility and usefulness of our system implemented with batch active learning. We hope our work impacts many accessibility visualization applications by considering the preferences of individuals with blindness or low vision (BLV), as well as accommodating their varied conditions and circumstances.

REFERENCES

[1] 2022. Data Journalism. https://open.nytimes.com/tagged/data-journalism.

[2] Md Zubair Ibne Alam, Shehnaz Islam, and Enamul Hoque. 2023. SeeChart: Enabling Accessible Visualizations Through Interactive Natural Language Interface For People with Visual Impairments. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 46–64.

[3] Robert Amar, James Eagan, and John Stasko. 2005. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 111–117.

[4] Sriram Karthik Badam, Zhicheng Liu, and Niklas Elmqvist. 2018. Elastic documents: Coupling text and tables through contextual visualizations for enhanced document reading. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 661–671.

[5] Abhijit Balaji, Thuvaarakkesh Ramanathan, and Venkateshwarlu Sonathi. 2018. Chart-Text: A Fully Automated Chart Image Descriptor. *ArXiv* abs/1812.10636 (2018).

[6] Leilani Battle, Peitong Duan, Zachery Miranda, Dana Mukusheva, Remco Chang, and Michael Stonebraker. 2018. *Beagle: Automated Extraction and Interpretation of Visualizations from the Web.* Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3173574.3174168

[7] Aaron Baucom and Christopher Echanique. 2013. ScatterScanner: Data extraction and chart restyling of scatterplots. In *Proceedings of the 2013 CHI Conference on Human Factors in Computing Systems.*

[8] Erdem Biyik and Dorsa Sadigh. 2018. Batch active preference-based learning of reward functions. In *Conference on robot learning*. PMLR, 519–528.

[9] Erdem Biyik and Dorsa Sadigh. 2018. Batch active preference-based learning of reward functions. In *Conference on robot learning*. PMLR, 519–528.

[10] Liliana Bounegru and Jonathan Gray. 2021. *The Data Journalism Handbook: Towards a Critical Data Practice.* Amsterdam University Press.

[11] Paul Bradshaw. 2017. Data journalism. In *The Online Journalism Handbook*. Routledge, 250–280.

[12] Matthew Brehmer, Bongshin Lee, Benjamin Bach, Nathalie Henry Riche, and Tamara Munzner. 2016. Timelines revisited: A design space and considerations for expressive storytelling. *IEEE transactions on visualization and computer graphics* 23, 9 (2016), 2151–2164.

[13] Eric Brochu, Nando De Freitas, and Abhijeet Ghosh. 2007. Active Preference Learning with Discrete Choice Data.. In *NIPS*. 409–416.

[14] Sandra Carberry, Stephanie Elzer Schwartz, Kathleen Mccoy, Seniz Demir, Peng Wu, Charles Greenbacker, Daniel Chester, Edward Schwartz, David Oliver, and Priscilla Moraes. 2013. Access to multimodal articles for individuals with sight impairments. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 2, 4 (2013), 1–49.

[15] Giuseppe Carenini and John Loyd. 2004. Valuecharts: analyzing linear models expressing preferences and evaluations. In *Proceedings of the working conference on Advanced visual interfaces.* 150–157.

[16] Charles Chen, Ruiyi Zhang, Eunyee Koh, Sungchul Kim, Scott Cohen, Tong Yu, Ryan Rossi, and Razvan Bunescu. 2019. Figure Captioning with Reasoning and Sequence-Level Training. *arXiv:1906.02850 [cs]* (June 2019). arXiv:1906.02850 [cs]

[17] Yang Chen, Jing Yang, and William Ribarsky. 2009. Toward effective insight management in visual analytics systems. In *2009 IEEE Pacific Visualization Symposium.* IEEE, 49–56.

[18] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. 2019. Visualizing for the Non-Visual: Enabling the Visually Impaired to Use Visualization, In Computer Graphics Forum. *Computer Graphics Forum* 38, 3, 249–260. https://doi.org/10.1111/cgf.13686

[19] Mathieu Cliche, David Rosenberg, Dhruv Madeka, and Connie Yee. 2017. Scatteract: Automated Extraction of Data From Scatter Plots. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases.* Springer, 135–150.

[20] Matej Črepinšek, Shih-Hsi Liu, and Marjan Mernik. 2013. Exploration and exploitation in evolutionary algorithms: A survey. *ACM computing surveys (CSUR)* 45, 3 (2013), 1–33.

[21] Zhe Cui, Sriram Karthik Badam, M Adil Yalçin, and Niklas Elmqvist. 2019. Datasite: Proactive visual data exploration with computation of insight-based recommendations. *Information Visualization* 18, 2 (2019), 251–267.

[22] Paramita De. 2018. Automatic Data Extraction from 2D and 3D Pie Chart Images. In *2018 IEEE 8th International Advance Computing Conference (IACC).* 20–25. https://doi.org/10.1109/IADCC.2018.8692104

[23] Peitong Duan. 2017. *Beagle: automated extraction and interpretation of visualizations from the web.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[24] Ward Edwards. 1977. How to use multiattribute utility measurement for social decisionmaking. *IEEE transactions on systems, man, and cybernetics* 7, 5 (1977), 326–340.

[25] Franz Eisenführ, Martin Weber, and Thomas Langer. 2010. *Rational decision making.* Springer.

[26] Stephanie Elzer, Edward Schwartz, Sandra Carberry, Daniel Chester, Seniz Demir, and Peng Wu. 2007. A Browser Extension for Providing Visually Impaired Users Access to the Content of Bar Charts on the Web.. In *WEBIST (2).* Citeseer, 59–66.

[27] Leo Ferres, Petro Verkhogliad, Gitte Lindgaard, Louis Boucher, Antoine Chretien, and Martin Lachance. 2007. Improving accessibility to statistical graphs: The iGraph-lite system. *ASSETS'07: Proceedings of the Ninth International ACM SIGACCESS Conference on Computers and Accessibility* (2007), 67–74. https://doi.org/10.1145/1296843.1296857

[28] Tong Gao, Jessica R Hullman, Eytan Adar, Brent Hecht, and Nicholas Diakopoulos. 2014. NewsViews: an automated pipeline for creating custom geovisualizations for news. In *Proceedings of the SIGCHI conference on human factors in computing systems.* 3005–3014.

[29] Melinda T Gervasio, Michael D Moffitt, Martha E Pollack, Joseph M Taylor, and Tomas E Uribe. 2005. Active preference learning for personalized calendar scheduling assistance. In *Proceedings of the 10th international conference on Intelligent user interfaces.* 90–97.

[30] Heather Granz, Merve Tuccar, Shweta Purushe, and Georges Grinstein. 2013. Implementing Disability Accommodations in a Widely Distributed Web Based Visualization and Analysis Platform – Weave. In *Universal Access in Human-Computer Interaction. Design Methods, Tools, and Interaction Techniques for eInclusion*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, Constantine Stephanidis, and Margherita Antona (Eds.). Vol. 8009. Springer Berlin Heidelberg, Berlin, Heidelberg, 31–39. https://doi.org/10.1007/978-3-642-39188-0_4

[31] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. 2013. Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2277–2286.

[32] Heikki Haario, Eero Saksman, and Johanna Tamminen. 2001. An adaptive Metropolis algorithm. *Bernoulli* (2001), 223–242.

[33] Jonathan Harper and Maneesh Agrawala. 2014. Deconstructing and restyling D3 visualizations. In *Proceedings of the 27th annual ACM symposium on User interface software and technology.* 253–262.

[34] Jonathan Harper and Maneesh Agrawala. 2018. Converting Basic D3 Charts into Reusable Style Templates. *IEEE Transactions on Visualization and Computer Graphics* 24, 3 (2018), 1274–1286. https://doi.org/10.1109/TVCG.2017.2659744

[35] Ting-Yao Hsu, C Lee Giles, and Ting-Hao'Kenneth' Huang. 2021. SciCap: Generating Captions for Scientific Figures. *arXiv preprint arXiv:2110.11624* (2021).

[36] Jessica Hullman, Nicholas Diakopoulos, and Eytan Adar. 2013. Contextifier: automatic generation of annotated stock visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 2707–2716.

[37] Jessica Hullman, Yea-Seul Kim, Francis Nguyen, Lauren Speers, and Maneesh Agrawala. 2018. Improving comprehension of measurements using concrete re-expression strategies, In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. *IEEE transactions on visualization and computer graphics* 24, 1, 1–12.

[38] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2021. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[39] Daekyoung Jung, Wonjae Kim, Hyunjoo Song, Jeong-in Hwang, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2017. Chartsense: Interactive data extraction from chart images. In *Proceedings of the 2017 chi conference on human factors in*

computing systems. 6706–6717.

[40] Daekyoung Jung, Wonjae Kim, Hyunjoo Song, Jeong-in Hwang, Bongshin Lee, Bohyoung Kim, and Jinwook Seo. 2017. ChartSense: Interactive Data Extraction from Chart Images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* ACM, Denver Colorado USA, 6706–6717. https://doi.org/10.1145/3025453.3025957

[41] Ralph L Keeney, Howard Raiffa, and Richard F Meyer. 1993. *Decisions with multiple objectives: preferences and value trade-offs.* Cambridge university press.

[42] Dae Hyun Kim, Enamul Hoque, Juho Kim, and Maneesh Agrawala. 2018. Facilitating Document Reading by Linking Text and Tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology.* ACM, Berlin Germany, 423–434. https://doi.org/10.1145/3242587.3242617

[43] Yea-Seul Kim, Jake Hofman, and Daniel Goldstein. 2022. Putting scientific results in perspective: Improving the communication of standardized effect sizes. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* ACM.

[44] Yea-Seul Kim, Jessica Hullman, and Maneesh Agrawala. 2016. Generating personalized spatial analogies for distances and areas. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 38–48.

[45] Nicholas Kong and Maneesh Agrawala. 2012. Graphical Overlays: Using Layered Elements to Aid Chart Reading. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (Dec. 2012), 2631–2638. https://doi.org/10.1109/TVCG.2012.229

[46] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. 2014. Extracting References between Text and Charts via Crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, Toronto Ontario Canada, 31–40. https://doi.org/10.1145/2556288.2557241

[47] Shahid Latif, Zheng Zhou, Yoon Kim, Fabian Beck, and Nam Wook Kim. 2021. Kori: Interactive Synthesis of Text and Charts in Data Documents. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 184–194.

[48] Shahid Latif, Zheng Zhou, Yoon Kim, Fabian Beck, and Nam Wook Kim. 2021. Kori: Interactive Synthesis of Text and Charts in Data Documents. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. https://doi.org/10.1109/TVCG.2021.3114802

[49] 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.

[50] Xiaoyi Liu, Diego Klabjan, and Patrick N. Bless. 2019. Data Extraction from Charts via Single Deep Neural Network. *CoRR* abs/1906.11906 (2019). arXiv:1906.11906 http://arxiv.org/abs/1906.11906

[51] Alan Lundgard and Arvind Satyanarayan. 2021. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* (2021), 1–1. https://doi.org/10.1109/TVCG.2021.3114770

[52] Junyu Luo, Zekun Li, Jinpeng Wang, and Chin-Yew Lin. 2021. ChartOCR: Data Extraction from Charts Images via a Deep Hybrid Framework. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV).* 1916–1924. https://doi.org/10.1109/WACV48630.2021.00196

[53] Kim Marriott, Bongshin Lee, Matthew Butler, Ed Cutrell, Kirsten Ellis, Cagatay Goncu, Marti Hearst, Kathleen McCoy, and Danielle Albers Szafir. 2021. Inclusive data visualization for people with disabilities: a call to action. *Interactions* 28, 3 (2021), 47–51.

[54] Sean McKenna, Nathalie Henry Riche, Bongshin Lee, Jeremy Boy, and Miriah Meyer. 2017. Visual narrative flow: Exploring factors shaping data visualization story reading experiences. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 377–387.

[55] Ronald Metoyer, Qiyu Zhi, Bart Janczuk, and Walter Scheirer. 2018. Coupling story to visualization: Using textual analysis as a bridge between data and interpretation. In *23rd International Conference on Intelligent User Interfaces.* 503–507.

[56] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems* 26 (2013).

[57] Prerna Mishra, Santosh Kumar, Mithilesh Kumar Chaube, and Urmila Shrawankar. 2022. ChartVi: Charts summarizer for visually impaired. *Journal of Computer Languages* 69 (2022), 101107.

[58] Priscilla Moraes, Gabriel Sina, Kathleen McCoy, and Sandra Carberry. 2014. Evaluating the Accessibility of Line Graphs through Textual Summaries for Visually Impaired Users, In Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '14. *ASSETS14 - Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility*, 83–90. https://doi.org/10.1145/2661334.2661368

[59] Jason Obeid and Enamul Hoque. 2020. Chart-to-text: Generating natural language descriptions for charts by adapting the transformer model. *arXiv preprint arXiv:2010.09142* (Nov. 2020). http://arxiv.org/abs/2010.09142 arXiv: 2010.09142.

[60] Ramana Rao and Stuart K Card. 1994. The table lens: merging graphical and symbolic representations in an interactive focus+ context visualization for tabular information. In *Proceedings of the SIGCHI conference on Human factors in computing systems.* 318–322.

[61] Dorsa Sadigh, Anca D Dragan, Shankar Sastry, and Sanjit A Seshia. 2017. *Active preference-based learning of reward functions.*

[62] Bahador Saket, Alex Endert, and Çağatay Demiralp. 2018. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics* 25, 7 (2018), 2505–2512.

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

[64] Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. Revision: Automated classification, analysis and redesign of chart images. In *Proceedings of the 24th annual ACM symposium on User interface software and technology.* 393–402.

[65] Manolis Savva, Nicholas Kong, Arti Chhajta, Li Fei-Fei, Maneesh Agrawala, and Jeffrey Heer. 2011. ReVision: Automated Classification, Analysis and Redesign of Chart Images. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology - UIST '11.* ACM Press, Santa Barbara, California, USA, 393. https://doi.org/10.1145/2047196.2047247

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

[67] Mingyan Shao and Robert P. Futrelle. 2005. Recognition and Classification of Figures in PDF Documents. In *Proceedings of the 6th International Conference on Graphics Recognition: Ten Years Review and Future Perspectives* (Hong Kong, China) *(GREC'05).* Springer-Verlag, Berlin, Heidelberg, 231–242. https://doi.org/10.1007/11767978_21

[68] Ather Sharif, Sanjana Shivani Chintalapati, Jacob O Wobbrock, and Katharina Reinecke. 2021. Understanding screen-reader users' experiences with online data visualizations. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility.* 1–16.

[69] Ather Sharif and Babak Forouraghi. 2018. evoGraphs — A jQuery Plugin to Create Web Accessible Graphs. In *2018 15th IEEE Annual Consumer Communications Networking Conference (CCNC).* IEEE, 1–4. https://doi.org/10.1109/CCNC.2018.8319239

[70] Ather Sharif, Olivia H Wang, Alida T Muongchan, Katharina Reinecke, and Jacob O Wobbrock. 2022. VoxLens: Making Online Data Visualizations Accessible with an Interactive JavaScript Plug-In. In *CHI Conference on Human Factors in Computing Systems.* 1–19.

[71] Conglei Shi, Weiwei Cui, Shixia Liu, Panpan Xu, Wei Chen, and Huamin Qu. 2012. Rankexplorer: Visualization of ranking changes in large time series data. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2669–2678.

[72] Danqing Shi, Xinyue Xu, Fuling Sun, Yang Shi, and Nan Cao. 2021. Calliope: Automatic Visual Data Story Generation from a Spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (Feb. 2021), 453–463. https://doi.org/10.1109/TVCG.2020.3030403

[73] Arjun Srinivasan, Steven M Drucker, Alex Endert, and John Stasko. 2018. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE transactions on visualization and computer graphics* 25, 1 (2018), 672–681.

[74] Nicole Sultanum, Fanny Chevalier, Zoya Bylinskii, and Zhicheng Liu. 2021. Leveraging Text-Chart Links to Support Authoring of Data-Driven Articles with VizFlow. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* ACM, Yokohama Japan, 1–17. https://doi.org/10.1145/3411764.3445354

[75] Romain Vuillemot and Charles Perin. 2015. Investigating the direct manipulation of ranking tables for time navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 2703–2706.

[76] Emily Wall, Subhajit Das, Ravish Chawla, Bharath Kalidindi, Eli T Brown, and Alex Endert. 2017. Podium: Ranking data using mixed-initiative visual analytics. *IEEE transactions on visualization and computer graphics* 24, 1 (2017), 288–297.

[77] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. 2020. DataShot: Automatic Generation of Fact Sheets from Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (Jan. 2020), 895–905. https://doi.org/10.1109/TVCG.2019.2934398

[78] Yanan Wang, Ruobing Wang, Crescentia Jung, and Yea-Seul Kim. 2022. What makes web data tables accessible? Insights and a tool for rendering accessible tables for people with visual impairments. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems.* ACM. https://doi.org/10.1145/3491102.3517469

[79] Chia-Kai Yang and Chat Wacharamanotham. 2020. *Asymmetric Effect of Text-Chart Proximity on Reading Behavior.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3419249.3420184

[80] Jian Zhao, Shenyu Xu, Senthil Chandrasegaran, Chris Bryan, Fan Du, Aditi Mishra, Xin Qian, Yiran Li, and Kwan-Liu Ma. 2021. Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. *arXiv preprint arXiv:2103.03996* (2021).

[81] Qiyu Zhi, Alvitta Ottley, and Ronald Metoyer. 2019. Linking and layout: Exploring the integration of text and visualization in storytelling. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 675–685.