

Pluto: Authoring Semantically Aligned Text and Charts for Data-Driven Communication

Arjun Srinivasan
Tableau Research
Seattle, Washington, USA
arjunsrinivasan@tableau.com

Vidya Setlur
Tableau Research
Palo Alto, California, USA
vsetlur@tableau.com

Arvind Satyanarayan
CSAIL
MIT
Cambridge, Massachusetts, USA
arvindsatya@mit.edu

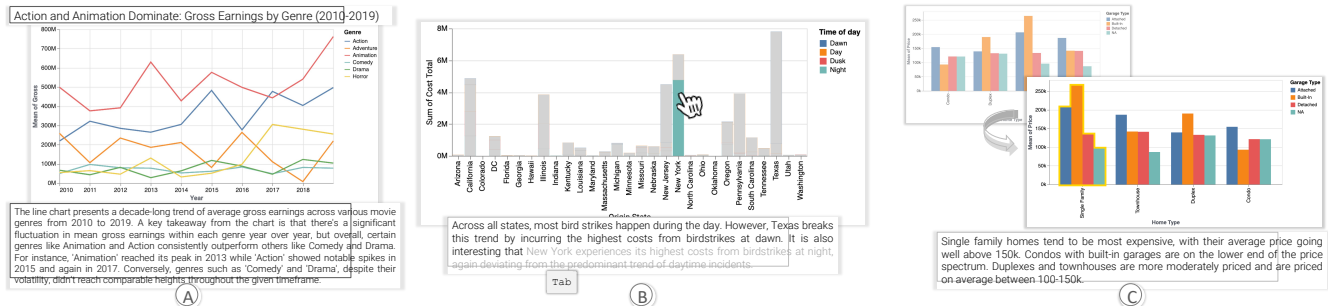


Figure 1: Examples of text and chart suggestions in PLUTO. (A) Coherent title and description that are auto-generated based on the chart. (B) Sentence completion is suggested based on multimodal input from the preceding text and a selection on the chart. (C) The chart is sorted and annotated to enhance coherence with the description.

Abstract

Textual content (including titles, annotations, and captions) plays a central role in helping readers understand a visualization by emphasizing, contextualizing, or summarizing the depicted data. Yet, existing visualization tools provide limited support for jointly authoring the two modalities of text and visuals such that both convey semantically-rich information and are cohesively integrated. In response, we introduce PLUTO, a mixed-initiative authoring system that uses features of a chart's construction (e.g., visual encodings) as well as any textual descriptions a user may have drafted to make suggestions about the content and presentation of the two modalities. For instance, a user can begin to type out a description and interactively brush a region of interest in the chart, and PLUTO will generate a relevant auto-completion of the sentence. Similarly, based on a written description, PLUTO may suggest lifting a sentence out as an annotation or the visualization's title, or may suggest applying a data transformation (e.g., sort) to better align the two modalities. A preliminary user study revealed that PLUTO's recommendations were particularly useful for bootstrapping the authoring process and helped identify different strategies participants adopt when jointly authoring text and charts. Based on study feedback, we discuss design implications for integrating interactive verification features between charts and text, offering control over

text verbosity and tone, and enhancing the bidirectional flow in unified text and chart authoring tools.

CCS Concepts

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

Keywords

Visualization, description, caption, mixed-initiative, recommendation.

ACM Reference Format:

Arjun Srinivasan, Vidya Setlur, and Arvind Satyanarayan. 2025. Pluto: Authoring Semantically Aligned Text and Charts for Data-Driven Communication. In *30th International Conference on Intelligent User Interfaces (IUI '25)*, March 24–27, 2025, Cagliari, Italy. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3708359.3712122>

1 Introduction

Research has shown that, perhaps paradoxically, text plays an important role in how readers interpret the data depicted by a visualization. These descriptions can explain how a chart is constructed, summarize statistical features (e.g., the minimum and maximum values), describe cognitive and perceptual phenomena (e.g., complex trends and patterns), or offer broader contextual and domain-specific explanations [34]. When well-authored, textual descriptions can helpfully reinforce visual features of the chart (e.g., high prominence characteristics) to reduce a reader's cognitive load while interpreting the chart [23] — indeed, in such cases, readers favor heavily annotated charts over simply charts or textual descriptions alone [56]. However, poorly written text can also slant



the takeaway message [26] in ways that impact readers' trust and recall [27].

Existing tools, however, provide almost no support for authoring text alongside the visualization. At best, a nascent body of work has begun to explore automated methods for generating titles [31] and descriptions [9, 20, 37] or, for pre-authored visualizations and textual descriptions, aligning emphasis [22], or constructing links and cross-references [29]. While valuable, these approaches leave unexplored a rich design space for *concurrently* authoring the two modalities together — that is, how might interactive mechanisms help users write a textual description while designing a chart; conversely, how might users update a chart's design to better reflect a textual description; and, how might a user iterate between the two modalities when authoring artifacts for data-driven communication?

To address this gap, we introduce PLUTO¹, a novel mixed-initiative tool designed to enable authors to craft semantically rich textual content for a variety of popular charts, including bar charts, histograms, line charts, and scatterplots. PLUTO facilitates an interactive composition process, allowing authors to leverage GPT for automatically generating chart descriptions (Figure 1A), use direct manipulation-based interactions with the chart to scope and guide text generation (Figure 1B), and even update the chart's design based on the composed text (Figure 1C). To operationalize this interactive authoring experience, we model a conceptual schema that captures key components of the chart (i.e., data, specification, selections, visual embellishments) and text (i.e., title, description, textual annotations along with the semantic information they convey [34]). We leverage this schema as part of a pipeline that blends heuristics and large language models (LLMs) to generate authoring recommendations in PLUTO.

Through a preliminary user evaluation, we explore the utility of PLUTO as a tool for helping authors create semantically relevant content when integrating text with the corresponding chart that it describes. We observe that the recommendations collectively afford a breadth of authoring workflows along a spectrum of using automatically generated text to leveraging recommendations only to augment manually authored content. Participant feedback also suggests that certain forms of assistance—such as generating text based on chart selections or suggesting chart design changes based on entered text—are deemed more useful than automated text generation and text editing suggestions. Our findings contribute to a growing body of literature on the integration of textual and visual elements for data-driven communication, offering insights into how future visualization authoring tools can better leverage the semantics and functional role of text with visualizations.

2 Related Work

Our work builds on several lines of research: exploring the role of text with visualizations, visualization and text systems, and image and text authoring interfaces.

¹Inspired by the celestial narrative where Pluto represents an entity once marginalized in astronomical classification, this work seeks to underscore the often-overlooked importance of text in data visualization. The name also aligns with the tradition of naming advances in visualization after celestial bodies, such as Polaris [57] and Vega [1].

2.1 The role of text with visualizations

The interplay between text and visual elements in data visualization has been a significant area of interest with increased advocacy for treating text as co-equal to visualization [34, 55]. Kim et al. [23] conducted a study to understand how readers integrate charts and captions in line charts. The study findings indicated that when both the chart and text emphasize the same prominent features, readers take away insights from both modalities. Their research underscores the importance of coherence between visual and textual elements and how external context provided by captions can enhance the reader's comprehension of the chart's message. Building on these insights, Lundgard and Satyanarayan [34] proposed a four-level model for content conveyed by natural language descriptions of visualizations. Their model delineates semantic content into four distinct levels: elemental and encoded properties (Level 1), statistical concepts (Level 2), perceptual and cognitive phenomena (Level 3), and contextual insights (Level 4).

Focusing on the role of textual annotations in visualization, Stokes et al. [56] observed that readers favored heavily annotated charts over less annotated charts or text alone. This preference highlights the added value of textual annotations in aiding data interpretation, with specific emphasis on how different types of semantic content impact the takeaways drawn by readers. Further contributions by Quadri et al. [40] and Fan et al. [15] explored high-level visualization comprehension and the impact of text details and spatial autocorrelation on reader takeaways in thematic maps. These studies collectively underline the critical role of textual elements in shaping viewer perceptions, understanding, and accessibility of visual data. Ottley et al. [38] and Stokes et al. [54] have also contributed to this body of research, focusing on how annotations influence perceptions of bias and predictions, reinforcing the multifaceted impact of text on visual data interpretation.

Our work further explores how text and charts can be better aligned with one another by offering a mixed-initiative authoring interface. Specifically, PLUTO allows leveraging both direct manipulation interactions and user-drafted text to generate recommendations for communicative text and chart design. Furthermore, PLUTO's text recommendations explicitly incorporate Lundgard and Satyanarayan's model [34] for semantic information conveyed by visualization descriptions. In doing so, the system ensures that the generated text has good semantic coverage and structure (e.g., generated descriptions start by conveying the chart's encodings and then list high-level trends) and is appropriate for the intended communicative use (e.g., the semantic information conveyed by titles is different from descriptions accompanying a chart or annotations on the chart).

2.2 Visualization and text systems

The integration of visualization and text has led to the development of various systems designed to facilitate the creation, interpretation, and enhancement of data visualizations with textual elements. He et al. [17] surveyed the leveraging of large models for crafting narrative visualizations, highlighting the potential of AI in supporting the narrative aspect of data visualization. This is complemented by AutoTitle, an interactive title generator for visualizations [31], and Vistext, a benchmark for semantically rich chart captioning [60].

VizFlow demonstrates the effectiveness of facilitating author-reader interaction by dynamically connecting text segments to corresponding chart elements to help enrich the storytelling experience [58]. This body of research highlights the need for tools to support more nuanced integration of text and visualization.

Supporting the co-authoring of text and charts, Latif et al. introduced Kori [29], an interactive system for synthesizing text and charts in data documents, emphasizing the seamless integration of visual and textual data for enhanced communication. CrossData [8] presents an interactive coupling between text and data in documents, enabling actions based on the document text and adjusting data values in the text through direct manipulation on the chart. Such systems illustrate the potential for the bidirectional linking between text and charts to assist rich authoring of data-driven narratives. Furthermore, systems like EmphasisChecker [22], Intentable [9], Chart-to-text [37], DataDive [25], InkSight [30], and FigurA11y [51] focus on guiding chart and caption creation, supporting readers' contextualization of statistical statements, and assisting in writing scientific alt text. Recent work like SciCapenter supports the composition of scientific figure captions using AI-generated content and quality ratings [20]. DataTales [59] is another example of a recent system using a large language model for authoring data-driven articles, indicating the growing interest in AI-assisted data storytelling. These systems collectively illustrate the expanding scope of text integration into visualization, from enhancing data document creation to improving accessibility and data-driven communication. Reviewing the aforementioned tools and the use of generative AI for visualization more broadly, Basole and Major [5] discuss how generative AI methods and tools offer creativity assistance and automation within the visualization workflow, specifically highlighting a shift towards "human-led AI-assisted" paradigms, where generative AI not only augments the creative process but also becomes a co-creator.

Aligned with this paradigm shift, PLUTO adopts a mixed-initiative approach that leverages the capabilities of generative AI to help create semantic alignment between the chart and its corresponding text for effective data-driven communication. However, PLUTO differs from existing chart-and-text authoring tools in three significant ways. First, going beyond existing systems that primarily leverage unimodal information from the chart to generate text, PLUTO supports multimodal authoring combining information from both the chart (including any direct interactions with marks) and user-drafted text. Furthermore, unlike prior tools that focus exclusively on generating complete descriptions/captions or titles, PLUTO's recommendations can be leveraged in flexible ways to author not only titles and descriptions but also more fine-grained annotations and sentence completions. Second, while existing tools primarily recommend text for a given chart, PLUTO's recommendations are bidirectional. Specifically, the system suggests chart design changes like sorting or adding embellishments based on the authored text, resulting in artifacts that more clearly communicate takeaways via a combination of text and charts. Lastly, unlike existing tools that primarily rely on pre-trained knowledge in generative AI models, PLUTO's recommendations are grounded in a theoretical research-based model of semantic information conveyed in visualization text [34], ensuring the generated text covers the appropriate

level of detail and is effective for communication *alongside* the chart.

3 Design

The central idea of our work is exploring a unified visualization system for authoring well-integrated charts and text. Designing such a system, however, requires considering several open questions about the type of assistance the system should provide, and when and how system suggestions should be surfaced.

In exploring the chart-and-text authoring experience, several questions arise that warrant exploration. We must first discern which chart elements, ranging from axis labels and ticks to titles, descriptions, and annotations, necessitate the most authoring support. Additionally, we would need to determine the appropriate level of system assistance—whether to generate entire descriptions, fill in partially written text, or refine user-authored drafts—and how this assistance might vary with different types of text. Given the non-mutual exclusivity of text types, such as descriptions influencing titles, we must also consider if and how the system should sequence its suggestions. The timing of these suggestions is another critical factor: should they be offered immediately following the creation of a chart or once the user has initiated the authoring process? Deciding whether these suggestions should be proactive or solicited on-demand, along with the specific user actions that should trigger them, is also a direction worth considering. Lastly, we must explore the potential for a synergistic relationship between the chart and text, i.e., how interactions with each can be leveraged to enhance the other and what mechanisms would facilitate this interplay.

3.1 Design Goals

With the aforementioned considerations in mind, we iteratively compiled a list of design goals to guide our system's development. These goals were informed by prior research and systems focusing on authoring text for visualizations (e.g., [17, 22, 29, 31, 51, 60]), general principles of mixed-initiative user interfaces [18], as well as formative interviews with two experts on authoring text and charts for data-driven communication.

The two experts were a practitioner and a researcher who both author and critique text for data visualizations. Furthermore, they also regularly interact with end-users in the creation of text and charts. Both experts voluntarily participated in the interviews and were not financially compensated. We interviewed each expert twice over a span of three weeks. Each session lasted for 30-60 minutes. During the first set of interviews, we asked the experts about the key challenges users generally encounter during the authoring of text with charts. Additionally, to guide our design and identify critical features, we also presented an early version of our prototype with a basic set of functionality including generating titles and descriptions for a chart and supporting interactive highlighting of chart elements based on the text. Based on the initial feedback, we incorporated additional types of recommendations and refined the system design before the second meeting where the experts provided feedback on the overall utility and perceived usability of the different features. We subsequently developed the final version of



Figure 2: PLUTO’s user interface. The key components include a data panel (A), chart editor (B), chart title (C), main chart canvas (D), and a chart description (E). Here, the user has manually entered a description and clicked the **Suggest** button to get ideas on improving the chart and text for communication purposes. This results in the system suggesting a title and adding a highlight annotation for **Single Family** homes, while also generating a chart design recommendation (F) and a set of description editing recommendations (G).

PLUTO by iterating on this feedback and leveraging findings from related work on text+chart authoring systems (e.g., [9, 22, 29, 30, 59]).

DG1. Leverage the textual narrative to guide chart design. In line with prior work [23, 38, 56], the experts also stressed that the text should not only convey the right levels of information but also be well-aligned with the chart for a smooth reading experience. As text has an inherent narrative flow, we noted that the system should incorporate techniques from prior work on updating chart specification based on narrative text [8, 48, 62] to inspect the flow of information in the text and leverage it to augment the chart. This augmentation could involve making data transformation changes (e.g., sorting) or adding annotations to highlight portions of the chart that are emphasized in the text.

DG2. Support direct manipulation interactions with the chart for text generation. Visualizations make it easy to perceive trends in the data and identify points of interest. Phrasing something visually interesting as text can be challenging, however. For instance, one of the experts noted, “sometimes I notice something potentially interesting on the chart and want some quick text to verify what I’m seeing and get ideas for how to talk about it.” Given this multimodal nature of charts and text, in line with prior chart-and-text authoring systems (e.g., [8, 30]), we noted that the system should allow leveraging direct interactions with the chart (e.g., brushing a region or mark selection) to generate corresponding text.

DG3. Provide varying levels of assistance for text authoring. Both experts noted that users need different levels of assistance when writing text depending on their goals and experience level. For instance, novice and intermediate users may need auto-generated text to jump-start their authoring process, whereas domain experts may benefit from fine-tuning suggestions to improve manually written text. Combining this comment with prior work on text-chart authoring [22, 56], we noted that the system should not only recommend text for chart authors to add but also recommend editing actions (e.g., reordering sentence) or flag potential factual errors in the text (e.g., incorrectly stated trends).

DG4. Incorporate context-sensitive recommendations near their relevant targets to facilitate easier interpretation. System recommendations in the context of a unified text and chart authoring process could apply to different targets (e.g., chart, title, or description) and focus on either adding new content or editing existing content. Interpreting this broad set of recommendations can be challenging, however. For instance, in our early prototypes, we explored listing all recommendations in a side panel, but both experts noted that this was overwhelming and distracted them from the main content. Iterating on the designs, we noted that for improved usability, the system recommendations should be placed close to the targets they apply to and should also be presented differently (e.g., in-place overlays vs. suggested actions) based on the type of recommendation.

DG5. Recommendations should be unobtrusive during targeted authoring. While the recommendations are designed to help craft cohesive text and charts, there may be instances where the chart authors have clear authoring goals in mind. In such targeted authoring scenarios, the recommendations should not interfere with the users’ flow but still be available on demand if users want ideas for text content or chart design. Authors should have full control over the final content, however, and should be able to edit/update any suggestions made by the system.

Note that these goals are not exhaustive or mutually exclusive, nor are they meant to be prescriptive. For instance, we primarily focus on content suggestions and do not deeply consider operations like formatting as part of the recommendation space. Rather, **DG1-DG5** are only meant to be an initial set of goals to help ground our design and enable us to develop and test a viable prototype.

4 Pluto

Incorporating these design goals, we implemented PLUTO as a prototype system for authoring semantically-aligned text and charts.

4.1 Example Usage Scenarios

Figure 2 shows PLUTO’s interface. Users can drag and drop data fields onto visual encoding channels to create charts. To underscore a unified experience for authoring charts and text, the system also presents an explicit title and description region just above and below the chart. Users can also annotate the chart by creating text callouts via a context menu invoked on the chart, or by adding visual embellishments using the chart editor (e.g., borders to highlight marks). System recommendations are either directly applied to the title, chart, or description or displayed to the right of the chart and description (Figure 2F, G) for authors to review (**DG4, DG5**).

To illustrate how PLUTO’s interface and features collectively enable unified authoring of text and charts for data-driven communication, we now describe three vignettes². These examples are also illustrated in the supplementary video.

Augmenting generated text with built-in safeguards and chart annotations. Consider Dinah, an analyst at a movie production company. Dinah is tasked with summarizing a chart showing movie earnings across genres (Figure 1A) to share as part of a report her company plans to publish.

Dinah is unsure about how to start her description, so she uses the **Generate** feature to bootstrap her authoring process (**DG3**). In response, PLUTO inspects the chart and returns a description for Dinah to review (Figure 1A-bottom). Dinah peruses the generated text and manually edits it for conciseness. She then clicks **Suggest** to get ideas for using a combination of the description and the chart for better communication.

Analyzing the description, PLUTO makes three changes. The system flags description statements with ambiguous takeaways for review using a dashed border, suggesting that Dinah manually verifies the text with the chart before sharing it with others (Figure 3A) (**DG3**). Using PLUTO’s interactive highlighting feature, Dinah hovers over the statement in the description to see portions of the

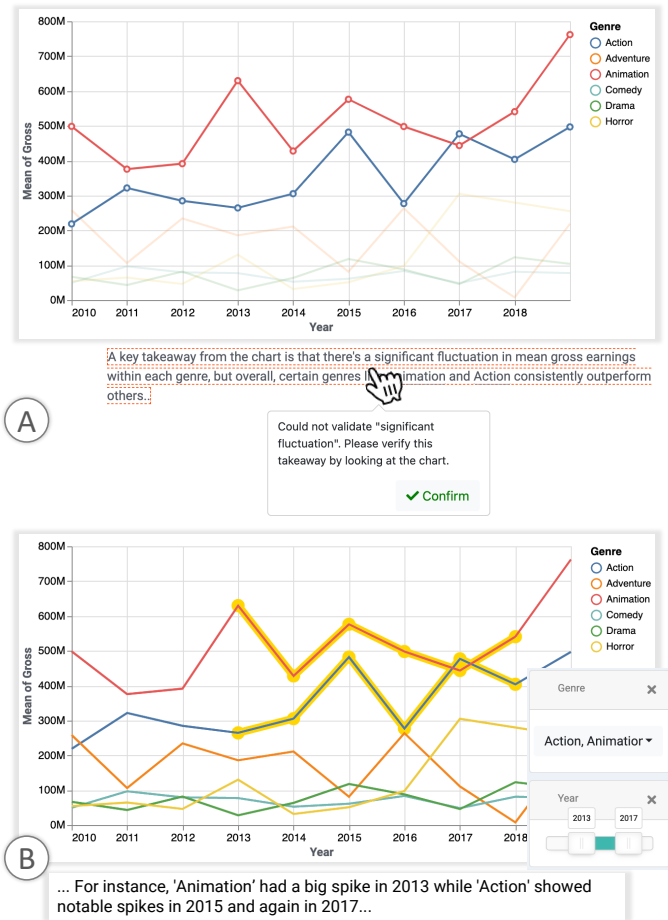


Figure 3: Upon processing a description, PLUTO flags statements that require manual verification (A) and automatically annotates the chart based on data references in the description (B).

chart it refers to. Reflecting on the flagged text, Dinah updates it to remove the modifier “significant” and make her description more objective for the readers’ interpretation.

PLUTO also suggests a title, “Action and Animation Dominate: Gross Earnings by Genre (2010-2019)” based on both the narrative in the description and the underlying trends in the chart (Figure 1A-top).

Finally, besides suggestions for the text, PLUTO also adds an annotation to the chart highlighting the key regions the text describes (**DG1**). In this case, detecting the emphasis on the Action and Animation genres and their trends between 2013 and 2017, the system adds a gold stroke around the corresponding lines in the chart (Figure 3B).

Satisfied with her changes based on the system suggestions, Dinah shares the title, chart, and description with her colleagues for review.

²These usage scenarios are modeled on examples of how participants used PLUTO during the study described in Section 5.

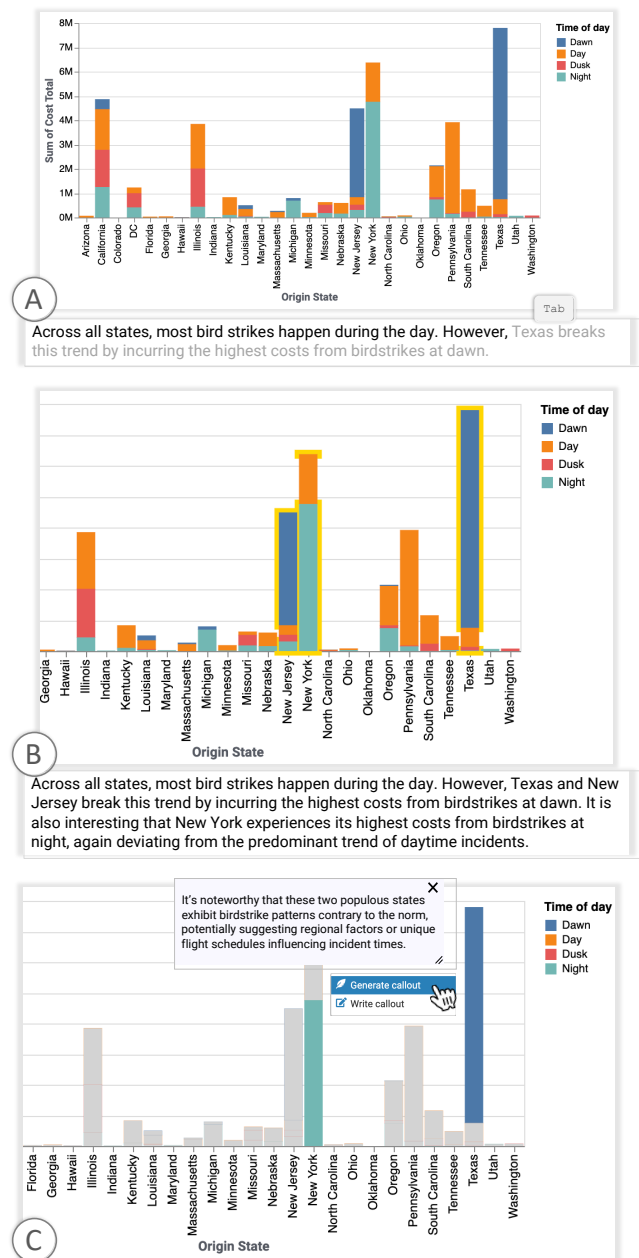


Figure 4: Examples of PLUTO’s recommendations including an in-place sentence completion (A), annotations based on a chart’s description (B), and a text callout generated based on marks selected on a chart (C).

Steering text competition through data-driven narratives and multimodal input. Imagine Ronnie, a financial analyst, writing a report on the monetary impact of bird strikes across the US based on the chart shown in Figure 4A.

Ronnie notices that most states, with the exception of Texas and New Jersey, have tall orange bars corresponding to costs incurred by

bird strikes during the day. Noting this observation, Ronnie types, “Across all states, most bird strikes happen during the day.” Wanting to emphasize the exception of Texas and New Jersey, Ronnie types “However,” and presses the `Tab` key to ask the system to finish the sentence. Parsing the preceding sentence and the data trends from the chart, PLUTO generates the completion “Texas breaks this trend by incurring the highest costs from birdstrikes at dawn.” (Figure 4A) Ronnie accepts this completion but edits it to include New Jersey.

Next, to emphasize the high cost incurred by incidents in New York at night, Ronnie types “It is also interesting that,” clicks on the teal bar showing the total cost for New York at Night, and again invokes a sentence completion. Using the multimodal input from the chart selection and the existing description text (DG2), PLUTO suggests the text “New York experiences its highest costs from birdstrikes at night, again deviating from the predominant trend of daytime incidents.” (Figure 1B)

Content with his description, Ronnie uses **Suggest** to see how he can further improve his text and the chart. PLUTO processes the description and adds a `stroke` to visually highlight the states Texas, New York, and New Jersey and the times Dawn and Night based on their high data values and the emphasis in the description (Figure 4B) (DG1). Seeing this annotation gives Ronnie an idea to explicitly call out the striking differences in values between Texas and New York. He selects the two tall bars within Texas and New York and uses **Generate Callout** to create a textual annotation directly overlaid onto the chart (Figure 4C) (DG2). Manually refining the generated callout and visual embellishments (DG5), Ronnie saves the annotated chart and description for his report.

Guiding manual authoring via system recommendations.

Imagine Fifi, a realtor who is using the grouped bar chart shown in Figure 2D to author an email blast on house pricing trends for her clients. Analyzing the chart, Fifi manually writes a description with three sentences, shown in Figure 2E. With this initial text, she invokes the **Suggest** feature to see how she can improve her text and chart for communication.

Parsing the description, PLUTO detects that it lacks a summary of the chart’s encodings and also detects that there is no higher-level statement encompassing a trend across multiple home types. Translating these into recommendations, PLUTO suggests adding a brief statement about the chart’s layout and a statement talking about the general impact of garage types across house types, respectively (Figure 2G) (DG3). Acknowledging these might be useful as overview statements for her readers, Fifi previews what her description would read like with the suggested text by hovering on the recommendations and subsequently accepting them. As with the other examples, PLUTO also suggests a title (*Home Type and Garage Influence on Property Prices*), but Fifi finds this too formal and manually adjusts it to make it more catchy: “Can I afford both a car and a home?: The Influence of Garage Type on Property Prices.”

Besides the text suggestions, PLUTO also detects that the description emphasizes the home types with highest and lowest values, whereas the home types in the chart are sorted alphabetically. To resolve this disparity, the system provides a chart design recommendation to sort the home types by price ranging from the highest to lowest (Figure 2F) (DG1). Fifi accepts this recommendation as it

```

Chart := (Data, Specification, ActiveSelection?,
  Annotation[]?)
Title := (Text, DataItem[]?)
Description := Statement[]

Statement := (Text, StatementType, DataItem[]?)
StatementType := encoding | perceptual-trend |
  data-fact | domain-specific | other

Annotation := TextAnnotation | MarkAnnotation |
  RegressionLine | ReferenceLine
TextAnnotation := (Text, DataItem[]?)
MarkAnnotation := (DataItem[], Style)
RegressionLine := Style
ReferenceLine := (Axis, Aggregation, Style)

Data := DataItem[]
ActiveSelection := DataItem[]
DataItem := (DataField, DataValue)

Axis := x | y
Style := (stroke | fill | opacity | ...)
Aggregation := mean | sum | ...
DataField := string
DataValue := string | number | datetime
Text := string

```

Figure 5: Conceptual schema representing the key text and chart elements in PLUTO’s interface.

can give her clients a glanceable summary of some key takeaways in her text (Figure 1C).

4.2 Conceptual Model

To enable the aforementioned workflows and recommendations, we model the various components across the text and the chart in PLUTO as a *conceptual schema* summarized in Figure 5. We use this schema in the subsequent sections to detail how the system tracks user input and generates recommendations.

Specifically, a **Chart** is represented by mapping **Data** onto specific visual encodings (in this case, a Vega-Lite **Specification** [44]). The **ActiveSelection** enumerates the data items that have been selected through direct manipulation interaction (e.g., in Figures 1B, 4C). Additionally, the chart can also have one or more **Annotations**. These can be textual comments on the chart, visual embellishments applied to individual marks (e.g., Figures 2D, 3B), or overlays like a regression line on a scatterplot or a line marking the average value across all bars in a bar chart.

The chart’s **Title** is **Text** that may include references to **DataItems** (e.g., *Genre: [Action, Animation]* in Figure 1A).

The **Description** is represented as a collection of **Statements**. Each statement maps to one of the four semantic statement types proposed by Lundgard and Satyanarayan [34]—namely, **encoding**, **perceptual-trend**, **data-fact**, **domain-specific**, or **other** (e.g.,

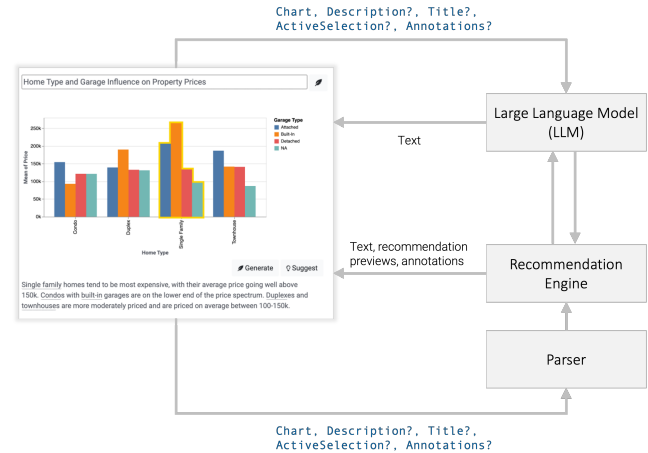


Figure 6: PLUTO’s system architecture overview

a statement about the data source for a chart). Additionally, similar to the title, statements in the description may also contain references to specific **DataItems**.

Note that this schema is not exhaustive (e.g., there may be additional types of annotations, statement types, or chart selections) and was primarily designed to operationalize the recommendations in PLUTO.

However, we hope that the idea of formalizing not only the chart but also its associated text can inspire future work on grammars and systems for data-driven communication through a *combination* of text and charts.

4.3 System Overview

PLUTO is implemented as a web-based application and is developed using Python, HTML/CSS, and JavaScript. Visualizations in the tool are created using Vega-Lite [44]. The system currently supports three encoding channels (x, y, color) and three mark types (bar, line, point). Collectively, this combination of encoding and mark types enables specifying several visualizations, including single- and multi-series bar charts, line charts, histograms, and scatterplots, covering a breadth of visualizations explored in prior systems [3, 9, 19, 22, 25, 29, 32, 37, 59].

Figure 6 depicts a high-level overview of the system architecture. Specifically, PLUTO uses a combination of an LLM (GPT-4 [2]) and a heuristics-based approach for generating suggestions. Specifically, requests like generating an entire description or a title from a chart are directly fulfilled using the LLM. In other cases, a custom parser extracts information from the text and chart and also classifies statements in the description based on their semantic levels [34]. This extracted information is leveraged by a heuristics-based recommendation engine to generate recommendations, including adding/editing text, adding mark annotations, and suggesting chart design changes such as sorting, among others. In cases where the recommendations involve generated text suggestions, the recommendation engine either uses its built-in templates or interacts with the LLM to pass it the required context for the text generation. In the subsequent

sections, we detail these components and PLUTO’s recommendation generation process.

4.4 Text and Chart Parsing

The parser extracts a number of features from the text and the chart that are used to determine system recommendations.

Text. The parser analyzes text in the description, title, and annotations to identify **DataItems**. The system uses a combination of a lexicon- and grammar-based approach adapted from prior natural language interfaces for visualization (e.g., [16, 36, 47]) to detect data item references. Specifically, given an input text, the parser extracts a list of N-grams and compares the N-grams to available data fields and values, looking for both syntactic (e.g., misspellings) and semantic similarities (e.g., synonyms) employing Levenshtein distance [66] and the Wu-Palmer similarity score [64], respectively. The extracted items are subsequently used to support features like adding mark annotations (Figures 3B, 4B) and highlighting relevant portions of the chart while hovering over statements in the description (Figure 3A).

In addition to detecting data item references, the parser also classifies description statements into one of the five statement types. We use a random forest classifier with BERT [42] to match a statement to one of the four semantic levels of text—**encoding**, **perceptual-trend**, **data-fact**, or **domain-specific** [34]. If the classification probability for all four types is below 60% (an empirically set threshold), a statement is labeled as **other**. The classifier is trained on a dataset of 2147 chart description statements curated by Lundgard and Satyanarayan [34]. Our choice for the classifier was based on comparing the results of 10-fold cross-validation between different techniques, including support vector machines [11], random forests [7], logistic regression [61], and naïve Bayes [14].

Chart. The parser also detects salient **DataItems** in the chart. For instance, for bar charts, the parser shortlists up to three categories with the highest and lowest values. For line charts, the system records time periods or specific timestamps with the most significant peaks and drops based on computing the smoothed z-scores, and so on. These chart-specific heuristics to determine salient targets are derived from prior “auto-insight” generating visualization systems (e.g., [12, 13, 52, 63]) and research on mappings between analytic tasks and visualizations (e.g., [4, 41, 45]). The salient items detected from the chart are subsequently used to suggest potential annotations and to generate text suggestions for verifying the description statements (e.g., Figure 3A).

4.5 Recommendation Generation

PLUTO uses a combination of heuristics, text templates, and an LLM to suggest changes to the text and the chart. The vignettes in §4.1 illustrate the breadth of PLUTO’s recommendations, which can broadly be categorized into three groups: 1) *full-text recommendations* to populate descriptions, titles, or text annotations, 2) *description statement recommendations* to fine-tune or update an existing description, and 3) *chart design recommendations* to ensure the chart is structurally aligned to its corresponding text.

Full-text Recommendations

These recommendations are invoked using the **Generate** button and suggest text for the title, description, or text annotations (e.g., Figure 1A and Figure 4C). All recommendations in this category are generated using the LLM, and Figure 7 presents an overview of the input/output for the recommendations. The LLM prompts are provided as part of the supplementary material.

Description. We use **StatementTypes** to systematically generate descriptions in PLUTO. Specifically, we provide the LLM with examples of the four statement types from Lundgard and Satyanarayan’s dataset [34]. Following the findings from Tang et al.’s qualitative analysis of the VisText chart caption dataset [60], we prompt the LLM to generate a description with a constraint that the text should start with an **encoding** statement and is followed by at least one **perceptual-trend**. This pattern follows the classic “*Overview first*” mantra for visualization design [49] and ensures the description talks about the chart and high-level takeaways before listing details of individual items and values. An example of this constraint in play can be noticed in the generated description in Figure 7 where the first statement, “*This line chart...*” describes the chart’s encodings and the second highlights how “*...certain genres like Animation and Action consistently outperform others...*” before talking about other lower-level observations from the chart.

By default, the LLM only uses the chart type and **Data** to generate a description. However, if the chart contains an **ActiveSelection**, has **TextAnnotations**, or the user has entered a **Title**, these are also used as context for generating the description.

Title. By default, the system uses a chart’s **Specification** and **Data** to generate a title. However, similar to generating descriptions, if there is additional context in the form of an **ActiveSelection**, **TextAnnotations**, or a **Description**, PLUTO leverages that information to generate a title that highlights the key message across the chart and previously added text. For instance, the suggested title in Figure 7 contains *Action* and *Animation* since these are called out as focal entities in the **Description**.

Text annotations. PLUTO also allows users to directly select items of interest on the chart and generate annotations based on the **ActiveSelection** (DG2). An example of this type of text generation is shown in Figure 7-bottom where the **TextAnnotation** is created based on the two selected bars for the states of *New York* and *Texas*, respectively. The example also illustrates the effect of including previously entered text as context for the generation. Specifically, notice that because the **Description** talks about *Texas* and *New York* deviating from the general trend, the generated annotation text also adopts that framing and phrasing (e.g., “*...contrary to the norm...*”) for consistency.

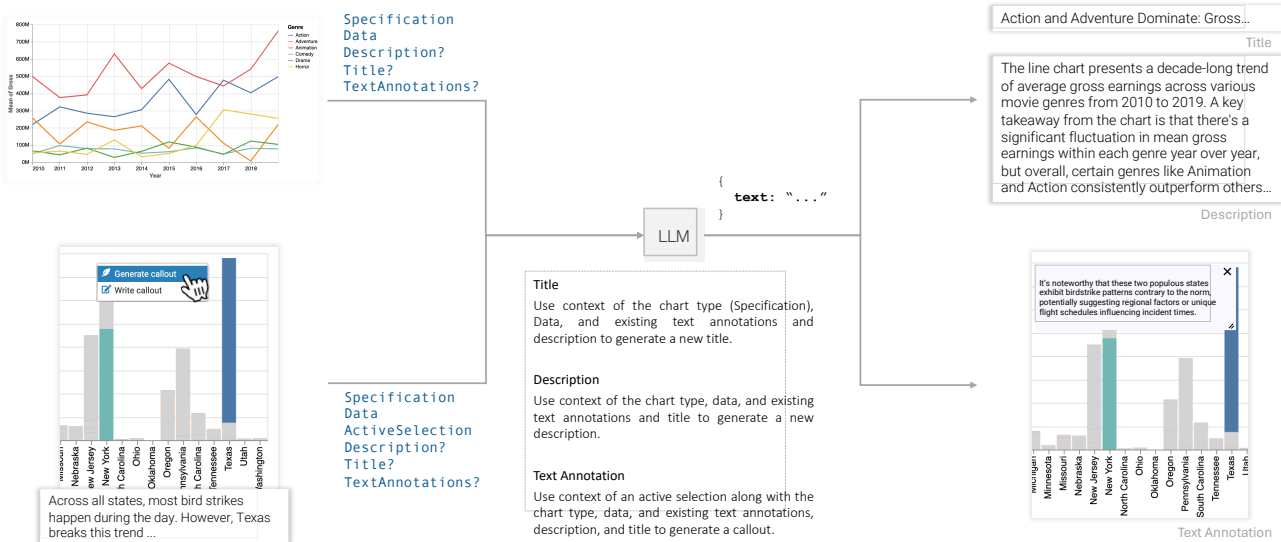


Figure 7: Overview of full-text recommendation generation. Given the context of the chart, data, and any existing text, PLUTO generates new text for the description, title, or annotations. Input parameters with a ? are optional only used if available.

Since all the above recommendations leverage the chart type (inferred via the **Specification**) and **Data**, from an implementation standpoint, we pass the chart type and the data to the LLM only once in an initial context setting prompt when a chart is created. Our choice to include the chart type as part of the context was motivated by our initial testing, during which we found that including the chart type improved the LLM’s performance in terms of detecting the most relevant data patterns (e.g., trends for line charts, extremes for bar charts, correlation for scatterplots).

Description Statement Recommendations

Besides suggesting text from scratch, PLUTO also recommends adding or editing **Statements** within an existing **Description** (DG3). Statement recommendations take different forms, including in-place suggestions to verify statement correctness (Figure 3A), statement addition/reordering recommendations presented to the side of an existing description (Figure 2G), and in-place text completions (Figure 1B and Figure 4A).

Statement addition and reordering. These recommendations are designed to ensure the description is semantically rich and has a good narrative structure. Figure 8 summarizes the logic used to generate statement recommendations. Note that because the rules used to generate these recommendations are already baked into the description generation prompt described above, the statement addition/reordering recommendations typically appear only for manually entered descriptions.

To generate the recommendations, we follow the same guidelines applied to generate a description from scratch.

PLUTO first checks for the presence of an **encoding** statement and at least one **perceptual-trend**. If these are absent or placed after other statements, the system recommends adding or reordering these statements. For instance, consider the example in Figure 8A. Detecting that the input description lacks both **encoding** and **perceptual-trend** statements, PLUTO suggests adding one statement of each type. We initially also explored suggesting adding **data-facts**. However, the formative studies and our testing revealed that these recommendations quickly became mundane and merely listed data values from the chart, leading to us subsequently disabling them. Drawing on prior work [60], we use a template-based approach to suggest **encoding** statements and invoke the LLM to suggest **perceptual-trends**.

To avoid overwriting the existing description, the recommendations are presented next to the description area instead of being directly applied to the text (Figure 2G) (DG5). Authors can preview the updated description with the suggested changes by hovering on the recommendations. Furthermore, because these recommendations are heuristically generated, PLUTO also provides an explanation for why a recommendation was shown. The output in Figure 8A shows examples of these explanations accompanying an **encoding** statement suggestion and a **perceptual-trend** statement suggestion generated when the input description only contains **data-facts**.

Statement verification. Prior research has shown that both manually-written and LLM-generated descriptions can contain erroneous mentions of data trends or values [22, 60]. For instance, a category stated to have the highest value in

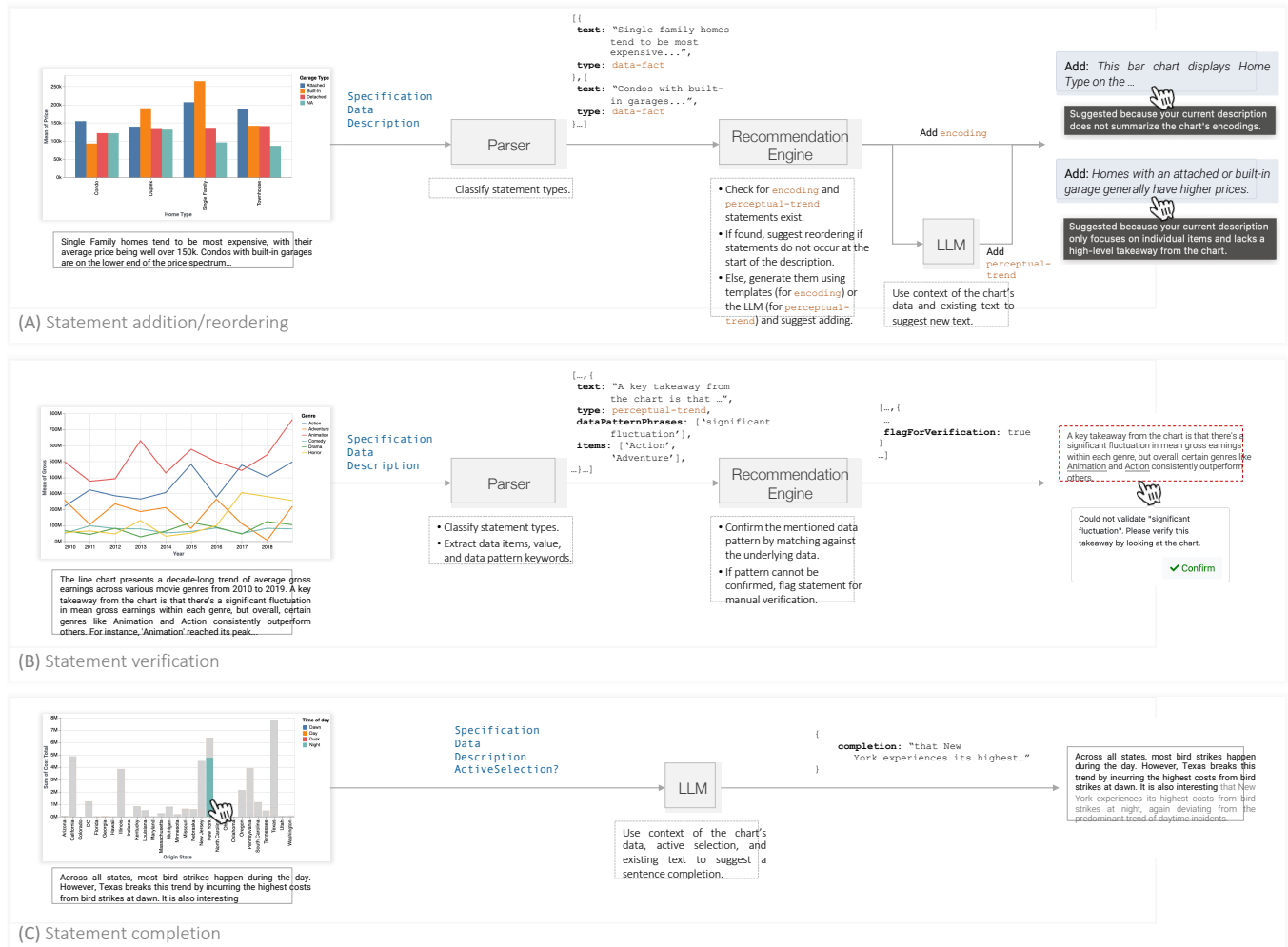


Figure 8: Overview of the description statement recommendations in PLUTO. The system uses a combination of the chart’s specification, data, the active description, and selections on the chart to recommend changes to the description.

the text may not actually be the category with the highest value in the chart. Motivated by this prior work and the experts’ feedback on earlier prototypes, we check for potentially incorrect data references in the text and flag them for authors to manually verify (DG3).

Figure 8B gives an overview of how PLUTO flags statements for verification. Specifically, the system first checks if the statement contains one or more **DataItems** (typically found in **data-fact** and **perceptual-trend** statements) and the type of takeaway the statement calls out (e.g., min/max, trend, correlation). PLUTO then uses this extracted information to validate the mentioned items and values against the underlying **Data**, flagging the statement for review if it fails to detect a match. An example of this is shown in Figure 8B, where the sentence is flagged because the system is unable to confirm

if the fluctuation in values is “significant”. Authors can click on a flagged statement to **Confirm** its correctness.

Statement text completion. In addition to retrospective recommendations on the description text, PLUTO also allows users to request text completion suggestions while writing their descriptions by pressing the **Tab** key. To generate these completions, the system sends the current **Description** along with any **ActiveSelections** on the chart to the LLM and prompts it to complete or suggest the last sentence. Examples of these suggestions can be seen in Figure 4A, where the system generates a text completion based on the **Description** alone and Figures 1B and 8C, where the completion is generated based on multimodal input, including the **description** text and the **ActiveSelection** of **New York** on the chart. If the text completion recommendation is invoked with an empty description, the system follows the same rules

as it does when generating descriptions from scratch and starts by suggesting an **encoding** statement followed by a **perceptual-trend** before other statements.

Chart Design Recommendations

Following **DG1**, PLUTO’s recommendations are geared not only to improve the text but also to align the chart with the text, resulting in a better-combined reading experience. Specifically, once a description is entered, the system generates two types of recommendations for updating the chart.

Annotations. PLUTO recommends visual embellishments based on the description to help emphasize the key takeaways from the text in the chart following the approach summarized in Figure 9. The recommended annotations are applied by default since they do not impact the chart’s structure/layout, but authors are provided with controls to refine or remove the applied annotations (e.g., Figures 2B and 3B).

The system first extracts a list of potential **DataItems** from both the text and the chart and assigns a saliency score to these items based on saliency or “interestingness” metrics [13, 34, 52, 63] (e.g., categories with extreme values in bar charts have higher saliency scores, time ranges in line charts with more variability have higher saliency than those with lower variability, targets mentioned in **perceptual-trend** statements are considered more salient than those referenced in **data-facts**).

Next, PLUTO checks for overlaps in **DataItems** extracted from the chart and the text to shortlist candidates for annotation based on a combined saliency score. Checking for the combined saliency scores across the text and chart ensures that the emphasized items are important in both the underlying data and the author’s interpretation of the chart conveyed via the text. In cases where there are no overlaps between **DataItems** in the chart and the text, PLUTO defaults to adjusting the chart to match the author’s description and adds an annotation for the most salient **DataItems** in the text. An example of this is shown in Figure 9 where the system highlights *Home Type=Single Family* since it has the highest average value (i.e., it is a salient data item in the chart) and is also explicitly called out in the description.

Besides annotating specific marks or regions on the chart, the system also adds overlay annotations based on references to aggregate values (e.g., adding a line to highlight the average value across categories in a bar chart or a regression line to emphasize the correlation between fields on a scatterplot).

Sorting. During our initial testing and formative interviews, we noted that there was often a disparity in the order in which data items or marks appear on a chart and the order in which they are discussed in the text. For example, consider the bar chart in Figure 2C. Although the bars are sorted alphabetically by home type, the description highlights takeaways starting with the home type having the

highest value (i.e., *Single Family* homes). To enable a more aligned chart and text reading experience, PLUTO checks for such disparities and recommends sorting order changes for alignment.

Specifically, as summarized in Figure 9, the system extracts an ordered list of **DataItems** in the description and compares it to the items in the chart. If the description starts by focusing on the highest or lowest category and the chart is not ordered to match that narrative, PLUTO suggests sorting the chart in a descending or ascending order, respectively. An example of the sorting recommendation in action can be seen in Figure 9 where applying the sort aligns the chart and text with both emphasizing *Single Family* homes having the highest values and *Condos* having the lowest values.

Unlike the annotation recommendations, however, sorting recommendations are presented next to a chart (Figure 2F) and not applied by default to prevent an abrupt visual layout change by the system. As shown in Figure 9, authors can preview the suggested order by hovering on the recommendation and clicking to apply that recommendation.

5 Preliminary User Study

Using PLUTO as a design probe, we conducted a preliminary user study to assess the utility of the proposed interactive experience for authoring semantically aligned text and charts for data-driven communication. Specifically, we focused on two goals: 1) understand if and how the system suggestions aid the joint authoring of text and charts, and 2) gather feedback on PLUTO’s current recommendations and features.

5.1 Participants and Setup

We recruited ten participants (*P1-P10*) through mailing lists at a data analytics software company. The recruitment call sought for individuals who use a combination of charts and text for data-driven communication. Participation in the study was voluntary and participants were recruited on a first-come, first-serve basis. Seven participants rated themselves as visualization experts, and three participants had a moderate level of expertise in visualization. Regarding participants’ professional backgrounds, six participants were solution architects, two participants worked as visualization consultants, and two were data analysts. Of the ten participants, six participants reported that they used a combination of text and charts to communicate data once in a few weeks, two frequently used text and charts as part of their jobs, and two participants had only recently started using charts and text for communication as part of their new roles. This mix of participant backgrounds w.r.t. chart and text authoring ensured that the study captured holistic feedback on PLUTO from varying users, including novices, moderate-level users, and experts.

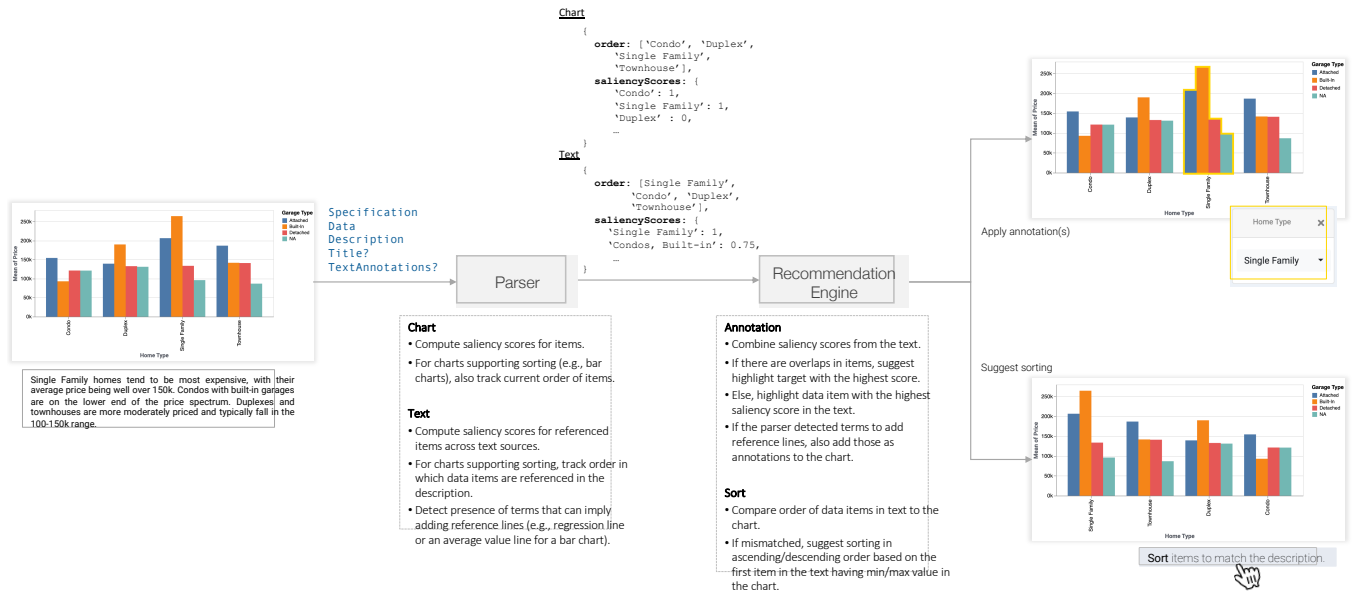


Figure 9: Summary of PLUTO’s process for recommending chart design changes based on the authored text. Given a chart and accompanying text, the system extracts data references from both the chart and the text, and compares the references to suggest potential design changes to make the chart more structurally aligned to the text.

Speaking about their existing authoring experience, all participants noted they manually inspected the chart and wrote a corresponding text blurb while sharing. Five participants said they sometimes annotated a chart’s screenshot with text during communication. The two participants who frequently communicated using text and charts also noted using the dashboard authoring features in tools like Tableau.

All sessions were conducted remotely via the Zoom video conferencing software. The prototype was hosted on a local server on the experimenter’s laptop. Participants were granted control over the experimenter’s screen during the session, and all studies followed a think-aloud protocol. The audio, video, and on-screen actions were recorded for all sessions with permission from the participants.

5.2 Procedure

We initially considered an evaluation of PLUTO against an existing chart-and-text authoring tool. However, we did not find a freely available baseline that provided equivalent features to PLUTO in terms of supporting multimodal authoring, bidirectional editing of text and charts, and using the semantic structure of descriptions [34] during text recommendation. We also considered an ablation study to compare the system’s output to that from the underlying GPT-4 model. However, we did not see value in this approach as PLUTO’s utility stems from the integration of multimodal interactions and recommendations, and not one standalone feature focused on text generation.

We ultimately decided on a qualitative study where all participants interact with PLUTO and perform the same set of tasks as this would allow us to observe usage patterns and assess the utility of the recommendations. Sessions lasted between 44-57 minutes (μ : 53 min., σ : 4 min.) and were organized as follows:

Introduction [~10min]: Participants were given an overview of the study and asked to share their background information. This briefing was followed by an introduction to PLUTO’s interface and key features. We used a simple bar chart about US college costs as a running example for the introduction.

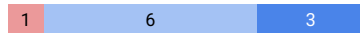
Tasks [~30min]: Participants were presented with three charts and were asked to write complementary text (including title, description, and annotations) for the chart so they could share their findings with others who may be interested in the data. To ensure the tasks were realistic and of appropriate difficulty, we consulted the two experts from the formative study (Section 3.1) and conducted three pilot studies with participants from academic and journalism backgrounds.

The three charts included a multi-series line chart about movie earnings by genre, a stacked bar chart about costs incurred by airplane bird strikes in the US, and a grouped bar chart about US housing prices shown in Figures 1A, 4A, and 2D, respectively. We selected these chart types based on their frequency of use in prior chart-and-text authoring systems (e.g., [8, 9, 22, 30, 59]) as well as the general prevalence of bar

Overall, how likely are you to use such a system in practice for authoring text and charts for data-driven communication?



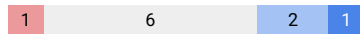
How helpful were the text generation suggestions?



How helpful was it to use selections on the chart to guide the text generation features?



How helpful were the system recommendations for editing the description?



How helpful were the system recommendations for updating the chart?



Very unhelpful/unlikely Unhelpful/unlikely Moderately helpful/likely Helpful/likely Very helpful/likely

Figure 10: Participant responses to post-session questions about PLUTO’s recommendations.

and line charts across visualization platforms (e.g., [6, 39]) that are commonly leveraged for data-driven communication. The order of charts was randomized across participants. We asked the participants to use the system as they saw fit (e.g., start with auto-generated text, manually write text and use the suggestions to edit, or manually compose the chart and text without using system recommendations).

Debrief [~10min]: Sessions concluded with a semi-structured interview discussing the overall experience, support for different authoring tasks, and areas for improvement. Participants also filled out a questionnaire rating the quality and utility PLUTO’s recommendations and features.

5.3 Results

Overall, participants were receptive to PLUTO’s features and recommendations, noting they would use such a system in practice for data-driven communication (Figure 10). We summarize general themes from participant feedback.

Full-text recommendations are helpful for bootstrapping. Participants generally found the text generation recommendations useful, with nine participants rating them as ‘helpful’ or ‘very helpful’ (Figure 10). Participants noted that these suggestions were particularly useful for bootstrapping the authoring process. For instance, commenting on the generated description, P1 said, “I would love to use these as a first draft. Just helps get my juices flowing.” However, participants’ feedback on the quality of generated descriptions was mixed, with some participants (P2, P4, P7) finding the suggested text too verbose. Both P4 and P7, for instance, noted that

they would prefer the system generate a bullet list of key “insights,” allowing them to use the insights to manually craft a narrative. P2 and P4 (both frequent users of text and charts for communication) also stated they would like more control over the generated text in terms of its verbosity and writing style (e.g., configuring the text to have a more “casual” versus “formal” tone depending on the communication context).

Feedback on the suggested titles was unanimously positive, however. For example, commending a suggested title for the house pricing chart, P8 said, “*Terms like ‘Influence of really makes chart feel less templated and more informative.’*” Upon seeing the title “Action and Adventure Dominate:...”, P5 noted, “*I am very impressed with the title. It’s almost like it took all my changes and gave me a summary.*” Participants were also pleasantly surprised by the quality of text completions for individual sentences (e.g., Figure 4A) stating “*the text completion followed my lead very well*” (P6).

Using selections to guide recommendations saves time and instills confidence. Overall, participants appreciated the ability to directly select items on the chart to drive text recommendations, with 8/10 participants noting this feature was ‘helpful’ or ‘very helpful’ (Figure 10). The positive feedback for the multimodal text generation feature often stemmed from its ability to assist in faster writing and to adjust the scope of the generated text.

P8, for instance, particularly appreciated the ability to use chart selections to drive auto-complete and annotations (e.g., Figures 1B and 4C) and said, “*It was nice to be able to point at things and have the system give the text. It saved me a lot of time.*” P2 and P4, who found the description generation feature too verbose, switched to using selection-based text generation, with P4 stating that “*the selection at least ensured the generated text was about something I want to talk about.*”

Description editing recommendations are primarily useful for validation. Participants’ reactions to recommendations for verifying and editing an entered description (e.g., Figures 3A and 2G) were more neutral (Figure 10, Q4).

Seven participants explicitly noted that they appreciated that the system flagged text that needed verification. P6, for instance, said, “*Humans are lazy. Having that extra step is going to save a lot of people a lot of embarrassment.*” Noting that the verification suggestions made him more critical, P3 said, “*It’s important that we think about what charts are saying...and it’s referring back and making sure. It’s definitely a step in the right direction.*” The remaining participants either felt the suggestions should not appear for manually written text (P4) or that statements could be flagged more conservatively, reducing the work on the authors’ part (P7, P8).

Participants used recommendations to add or reorder text in the description only three times across all sessions. Participants commented that these recommendations were too

“obvious” (particularly referring to the **encoding** statement suggestions). P9, for instance, “*This [encoding statement] suggestion is too generic and not as interesting as the others that tell me key points about the data.*” However, all participants noted that the recommendations were appropriately placed on the side and did not impede their workflow (DG5).

Chart annotation and design recommendations foster an integrated reading experience. Participants were generally very impressed by PLUTO’s suggestions to annotate or sort the chart based on an entered description, with 8/10 participants rating this feature as ‘helpful’ or ‘very helpful.’

For example, even P7, who was generally critical about the other features, exclaimed, “*Loved that just loved that! It’s easy to forget the chart when writing because I know what I should be focusing on, but someone else looking at the chart may not.*” Appreciating the ability to adjust the system-suggested annotations further (DG5), P1 commented, “*It’s great that it highlighted some elements in the chart based on my text. It made me see things from a reader’s perspective and go back and make additional changes.*”

Usage patterns. As typical with mixed-initiative interfaces, there was a constant back and forth between the participants’ authoring actions and the system recommendations. We observed four high-level usage strategies around how participants started the authoring process by writing descriptions. We summarize these strategies below, as they can help inform the user experience of future systems³.

- **Generate then edit.** The most common strategy across participants (7/10) was to start with auto-generated descriptions (✍ **Generate**). Once the description text was generated, participants would either first peruse through it and make edits or directly request suggestions for improvements to the text and the chart (🔍 **Suggest**).
- **Guide text completion.** Four participants started writing their descriptions leveraging the in-place text completion suggestions (via the `Tab` key). Participants would typically mention a data entity (e.g., New York) or a narrative hook (e.g., “however,” “but,” “sadly”) in their text and have the system suggest the remaining statement that they would use as-is or edit further. The 🔍 **Suggest** feature was primarily used to verify the statements and get chart annotation recommendations to complement the text.
- **Clipboard text generation.** On two occasions, participants used the system text suggestions to create a clipboard of ideas. Specifically, participants started by selecting visually salient entities on the chart and asking PLUTO to generate a series of text annotations. Subsequently, the participants went through these annotations and either

edited and kept them on the chart, moved them to the description, or deleted them. The 🔍 **Suggest** feature was used after this initial drafting of the description and text annotations to get further editing suggestions for the chart.

- **Manual writing with chart design recommendations.** Two participants (both experts at communicating data using text and charts) typically started their process by manually drafting the description and then asking the system to 🔍 **Suggest** improvements. These participants also noted that they utilized the suggest feature to obtain chart annotations and design suggestions based on their input text (DG1).

6 Discussion

6.1 Grammar-based Approaches for Chart & Text Interfaces

While we currently implement only a set of charts and recommendations in PLUTO, the underlying schema (Figure 5) is not limited to this set. Since the framework is based on the underpinning data and conceptual elements of the chart and text, the presented schema can be generalized to other cases. Figure 11 highlights examples of this generalization by illustrating PLUTO’s text and embellishment suggestions for a geographic map and a graph dataset represented as a network matrix.

Furthermore, the idea of breaking individual **Statements** into **Text**, **StatementType**, and **DataItems** enables interactive text-chart linking (Figure 3A) as well as targeted recommendations for description statements (Figure 8). Beyond these recommendations, however, this breakdown can also be leveraged more generally to design better linting systems for writing text for charts. For example, by inspecting the **Text** and **StatementTypes** in a given description and applying findings from studies on accessible visualization descriptions [21, 24, 34, 65], visualization systems can flag inaccessible descriptions and suggest improvements to make the descriptions more accessible.

While these are just examples, leveraging logical concepts such as those in the presented schema (Figure 5) affords a compelling opportunity to create a unifying grammar to capture the multimodal nature of chart + text authoring interfaces. Besides streamlining the interface and interaction design of future tools, such a grammar that is centered around abstract concepts underpinning text and charts can also facilitate effective communication with LLMs by allowing systems to only share required task-specific context with the models in a structured representation.

³Note that some participants adopted different strategies across tasks, resulting in the participant count across strategies adding up to more than 10

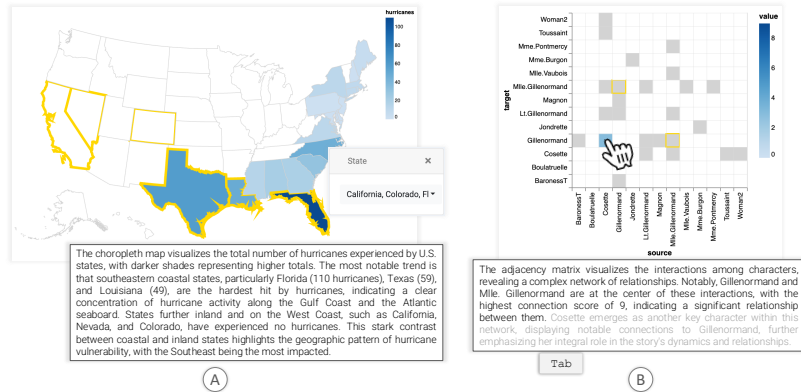


Figure 11: Examples illustrating the extensibility afforded by the proposed conceptual schema. Here, PLUTO’s text generation and recommendation modules are used *as-is* to author text for a geographic map and an adjacency matrix. In (A), the system generates a description based on the map and subsequently highlights six states in the chart that are emphasized in the text. In (B), PLUTO suggests a sentence completion based on multimodal input of the previously drafted description and a user selection on the chart.

6.2 Design Considerations

Based on the study observations and feedback, we derive four design considerations for future systems exploring AI-based suggestions to assist unified text and chart authoring.

Format and phrasing of text are as important as its content. Participants recognized the value in text generation and noted that the generated text often picked up on the salient data points and trends in the data. However, the two participants who frequently used text and charts for communication, in particular, critiqued the verbosity and tone of the generated text, indicating that the text was “*too formal or complex*” for their consumers. Future systems should explore providing users with control over the configuration of the properties of the generated text (e.g., choosing between paragraphs and bullets, adjusting the level of verbosity, and setting the tone or writing style) [33].

Leverage multimodal context during text generation.

Participants particularly appreciated text suggestions that were based on chart selections or built upon previously authored text. *P5*, for instance, referred to the interaction experience of having some text and then using the chart in tandem to guide the system (Figure 1B) as being a “*smooth and controlled authoring flow*.” To support similar fluid and coherent authoring experiences, future systems should consider multimodal context from both the chart and previously written text when suggesting new text.

Include techniques to help verify the text with the chart. Participants appreciated the interactive highlighting and statement verification features in PLUTO (Figure 3A), noting that the features encouraged them to be more critical of the text (regardless of whether it was written by them or was

system-generated). With the growing prevalence of generated text, future systems should continue incorporating such interactive verification features to mitigate false statements and enhance synergy between the text and the chart.

Adjust chart design to align with the text description.

Unlike prior systems that focus on the unidirectional task of generating text from charts, PLUTO introduces a bidirectional flow by also recommending changes to the chart design based on the text (Figure 9). We observed that participants extensively used and appreciated these suggestions, commenting that the chart design recommendations helped them gain a better reader’s perspective. To this end, future systems should continue to explore techniques to leverage the bidirectional flow of information between text and charts and generate chart design suggestions for an integrated reading experience.

7 Limitations and Future Work

Incorporate data context into generated text. PLUTO currently only considers the active chart and data fields while suggesting text. However, since the recommendations are provided within a general-purpose visualization specification tool, the system can access other data fields not displayed in the active chart (Figure 2A). For instance, during the study, viewing the generated text for the bird strikes chart (Figure 4A), *P8* said, “*I wish it could generate some text explaining why the costs were high based on the number of incidents.*” Although *P8* was able to work around this limitation by creating the second chart, inspecting the new visualization, and returning to the original chart, automatically considering other fields as part of the generated text is an open area for future work. Such dataset-level text (as opposed to

chart-level text) can also help make the authored descriptions semantically rich by including more **domain-specific** statements.

Text formatting recommendations. The current implementation leans heavily on content generation; however, future iterations should also focus on providing sophisticated formatting suggestions that can add expressivity to the text based on the semantic levels. For example, titles can be rendered more prominently with annotations and footnotes shown to support the chart. Complementing chart design suggestions and providing structural suggestions for the description, such as including line breaks or using bullet points instead of paragraphs, can also help improve readability.

Supporting multiple views and articles. PLUTO currently supports authoring a single chart and assumes the resulting chart and text are static. While such an approach covers a popular scenario for data-driven communication as evidenced by prior research focusing on this setup (e.g., [9, 19, 32, 37, 56, 60]), people also use dashboards and interactive storytelling articles when communicating with charts [43, 46]. Investigating support for multiple views and operationalizing PLUTO's recommendations in more expressive interactive data-driven article authoring tools such as Idyll Studio [10] is an open topic for future work.

Manage data sharing and combine heuristics with LLMs. As discussed in §4.5, we currently provide the context of the chart's specification and the data to the LLM. However, depending on the size of the data and or the usage context, sharing data directly with the LLM may not always be feasible. Exploring alternatives to generate text without sharing the underlying data is an important direction for future work. For instance, one possible approach could be to use heuristics-based approaches and prior knowledge of analytic tasks (e.g., [22, 35, 53]) to identify key trends from a chart and then leverage the LLM for appropriately consolidating the extracted information into a coherent text narrative.

Conduct longitudinal evaluation across data domains. The combination of two expert interviews, three pilots, and an evaluation study with ten participants helped us validate PLUTO as a proof-of-concept for the general idea of a mixed-initiative interface for authoring semantically aligned charts and text. However, this evaluation is only preliminary, and assessing the practical value of a tool like PLUTO deems more longitudinal studies spanning different data domains and individuals with varying levels of expertise [28, 50].

Incorporating safeguards for machine failures. With the current implementation of PLUTO, we utilize a pretrained LLM to generate text and a heuristic parser to analyze the generated text for additional suggestions. To make users

aware of potential errors in either of these steps, we provide interactive visual previews that can be viewed before accepting system suggestions (e.g., Figure 3A and Figure 9). While PLUTO serves as a proof-of-concept, developing a more generalized system requires a deeper investigation of the types of errors that might be generated both at the LLM and the parser level. Categorizing and understanding the distribution of such errors can ultimately help develop better recommendation modules, but also design interface and interaction strategies to mitigate system errors.

8 Conclusion

The integration of AI-driven recommendations with user-driven customization presents a promising direction toward authoring more semantically aligned charts and text. This work introduces PLUTO, a tool that incorporates a semantic and interactive framework that is informed by a nuanced understanding of the interplay between text and visuals. The system's recommendations, ranging from text generation to annotation and title suggestions, demonstrate the viability of a mixed-initiative approach to support richer chart-and-text authoring. A preliminary evaluation of the system revealed several key takeaways: participants highly valued the ability to use multimodal input for generating text, found text generation recommendations particularly useful for initiating the authoring process, and appreciated the system's ability to suggest annotations and chart design modifications based on the entered description. These findings underscore the potential of systems like PLUTO to facilitate the creation of more semantically coherent and integrated text and chart content. We envision a continued exploration of this research space, extending PLUTO's capabilities to accommodate a broader array of visualization types and narrative styles, wherein text and visuals are not mere complements but coequals in data-driven communication.

Acknowledgments

This work was supported, in part, by the National Science Foundation under III-1900991.

References

- [1] 2014. Vega: A Visualization Grammar. <https://vega.github.io/vega/>. Accessed: January 2025.
- [2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. GPT-4 Technical Report. *arXiv preprint arXiv:2303.08774* (2023).
- [3] 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.
- [4] 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.
- [5] Rahul C Basole and Timothy Major. 2024. Generative AI for Visualization: Opportunities and Challenges. *IEEE Computer Graphics and Applications* 44, 2 (2024), 55–64.

- [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. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [7] Leo Breiman. 2001. Random Forests. *Machine Learning* 45, 1 (Oct 2001), 5–32. <https://doi.org/10.1023/A:1010933404324>
- [8] Zhutian Chen and Haijun Xia. 2022. Crossdata: Leveraging Text-Data Connections for Authoring Data Documents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [9] Jiwon Choi and Jaemin Jo. 2022. Intentable: A Mixed-initiative System for Intent-based Chart Captioning. In *2022 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 40–44.
- [10] Matthew Conlen, Megan Vo, Alan Tan, and Jeffrey Heer. 2021. Idyll Studio: A Structured Editor for Authoring Interactive & Data-Driven Articles. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1–12.
- [11] Corinna Cortes and Vladimir Naumovich Vapnik. 1995. Support-Vector Networks. *Machine Learning* 20 (1995), 273–297. <https://api.semanticscholar.org/CorpusID:52874011>
- [12] Zhe Cui, Sriram Karthik Badam, M Adil Yalçın, and Niklas Elmqvist. 2019. Dat-asite: Proactive Visual Data Exploration with Computation of Insight-based Recommendations. *Information Visualization* 18, 2 (2019), 251–267.
- [13] Çağatay Demiralp, Peter J. Haas, Srinivasan Parthasarathy, and Tejaswini Pedapati. 2017. Foresight: Recommending Visual Insights. *Proc. VLDB Endow.* 10, 12 (Aug. 2017), 1937–1940. <https://doi.org/10.14778/3137765.3137813>
- [14] Richard O. Duda and Peter E. Hart. 1974. Pattern Classification and Scene Analysis. In *A Wiley-Interscience Publication*. <https://api.semanticscholar.org/CorpusID:12946615>
- [15] Arlen Fan, Fan Lei, Michelle Mancenido, Alan M. Maceachren, and Ross Maciejewski. 2024. Understanding Reader Takeaways in Thematic Maps Under Varying Text, Detail, and Spatial Autocorrelation. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 213, 17 pages. <https://doi.org/10.1145/3613904.3642132>
- [16] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G Karahalios. 2015. Datatone: Managing Ambiguity in Natural Language Interfaces for Data Visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST)*. 489–500.
- [17] Yi He, Shixiong Cao, Yang Shi, Qing Chen, Ke Xu, and Nan Cao. 2024. Leveraging Large Models for Crafting Narrative Visualization: A Survey. *arXiv preprint arXiv:2401.14010* (2024).
- [18] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) (*CHI '99*). Association for Computing Machinery, New York, NY, USA, 159–166. <https://doi.org/10.1145/302979.303030>
- [19] Ting-Yao Hsu, C Lee Giles, and Ting-Hao Huang. 2021. SciCap: Generating Captions for Scientific Figures. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih (Eds.). Association for Computational Linguistics, Punta Cana, Dominican Republic, 3258–3264. <https://doi.org/10.18653/v1/2021.findings-emnlp.277>
- [20] Ting-Yao Hsu, Chieh-Yang Huang, Shih-Hong Huang, Ryan Rossi, Sungchul Kim, Tong Yu, C Lee Giles, and Ting-Hao Kenneth Huang. 2024. SciCapenter: Supporting Caption Composition for Scientific Figures with Machine-Generated Captions and Ratings. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 284, 9 pages. <https://doi.org/10.1145/3613905.3650738>
- [21] 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* 28, 1 (2021), 1095–1105.
- [22] Dae Hyun Kim, Seulgi Choi, Juho Kim, Vidya Setlur, and Maneesh Agrawala. 2023. EmphasisChecker A Tool for Guiding Chart and Caption Emphasis. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [23] Dae Hyun Kim, Vidya Setlur, and Maneesh Agrawala. 2021. Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [24] Gyeongri Kim, Jiho Kim, and Yea-Seul Kim. 2023. “Explain What a Treemap is”: Exploratory Investigation of Strategies for Explaining Unfamiliar Chart to Blind and Low Vision Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [25] Hyunwoo Kim, Khanh Duy Le, Gionnieve Lim, Dae Hyun Kim, Yoo Jin Hong, and Juho Kim. 2024. DataDive: Supporting Readers’ Contextualization of Statistical Statements with Data Exploration. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 623–639.
- [26] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [27] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2019. Trust and Recall of Information Across Varying Degrees of Title-visualization Misalignment. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [28] Heidi Lam, Enrico Bertini, Petra Isenberg, Catherine Plaisant, and Sheelagh Carpendale. 2012. Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics* 18, 9 (2012), 1520–1536. <https://doi.org/10.1109/TVCG.2011.279>
- [29] 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.
- [30] Yanna Lin, Haotian Li, Leni Yang, Aoyu Wu, and Huamin Qu. 2023. Inksight: Leveraging Sketch Interaction for Documenting Chart Findings in Computational Notebooks. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [31] Can Liu, Yuhang Guo, and Xiaoru Yuan. 2023. AutoTitle: An Interactive Title Generator for Visualizations. *IEEE Transactions on Visualization and Computer Graphics* (2023).
- [32] Can Liu, Liwenhan Xie, Yun Han, Datong Wei, and Xiaoru Yuan. 2020. AutoCaption: An Approach to Generate Natural Language Description from Visualization Automatically. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 191–195.
- [33] Annie Louis and Ani Nenkova. 2014. Verbose, Laconic or Just Right: A Simple Computational Model of Content Appropriateness under Length Constraints. *14th Conference of the European Chapter of the Association for Computational Linguistics 2014, EACL 2014*. <https://doi.org/10.3115/v1/E14-1067>
- [34] 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* 28, 1 (2021), 1073–1083.
- [35] Denis M. Manelski and Gilbert K. Krulec. 1965. A Heuristic Approach to Natural Language Processing. In *COLING 1965*. <https://aclanthology.org/C65-1018>
- [36] Arpit Narechania, Arjun Srinivasan, and John Stasko. 2020. NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 369–379.
- [37] Jason Obeid and Enamul Hoque. 2020. Chart-to-Text: Generating Natural Language Descriptions for Charts by Adapting the Transformer Model. In *International Conference on Natural Language Generation*. <https://api.semanticscholar.org/CorpusID:224704949>
- [38] Alviitta Otley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. 2019. The Curious Case of Combining Text and Visualization. *EuroVis 2019-Short Papers* (2019).
- [39] Joanna Purich, Arjun Srinivasan, Michael Correll, Leilani Battle, Vidya Setlur, and Anamaria Crisan. 2023. Toward a Scalable Census of Dashboard Designs in the Wild: A Case Study with Tableau Public. *arXiv preprint arXiv:2306.16513* (2023).
- [40] Ghulam Jilani Quadri, Arran Zeyu Wang, Zhehao Wang, Jennifer Adorno, Paul Rosen, and Danielle Albers Szafir. 2024. Do You See What I See? A Qualitative Study Eliciting High-Level Visualization Comprehension. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '24*). Association for Computing Machinery, New York, NY, USA, Article 204, 26 pages. <https://doi.org/10.1145/3613904.3642813>
- [41] 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.
- [42] Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, A Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter. *arXiv preprint arXiv:1910.01108* (2019).
- [43] Alper Sarikaya, Michael Correll, Lyn Bartram, Melanie Tory, and Danyel Fisher. 2018. What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 682–692.
- [44] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, and Jeffrey Heer. 2016. Vega-lite: A Grammar of Interactive Graphics. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 341–350.
- [45] Hans-Jörg Schulz, Thomas Nocke, Magnus Heitzler, and Heidrun Schumann. 2013. A Design Space of Visualization Tasks. *IEEE Transactions on Visualization and Computer Graphics* 19, 12 (2013), 2366–2375.
- [46] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 1139–1148.
- [47] Vidya Setlur, Sarah E Battersby, Melanie Tory, Rich Gossweiler, and Angel X Chang. 2016. Eviza: A Natural Language Interface for Visual Analysis. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST)*. 365–377.
- [48] Leixian Shen, Haotian Li, Yun Wang, Tianqi Luo, Yuyu Luo, and Huamin Qu. 2024. Data Playwright: Authoring Data Videos with Annotated Narration. *IEEE Transactions on Visualization and Computer Graphics* (2024).
- [49] Ben Shneiderman. 2003. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*. Elsevier,

- 364–371.
- [50] Ben Shneiderman and Catherine Plaisant. 2006. Strategies for Evaluating Information Visualization Tools: Multi-Dimensional In-Depth Long-Term Case Studies. In *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. 1–7.
 - [51] Nikhil Singh, Lucy Lu Wang, and Jonathan Bragg. 2024. FigurA11y: AI Assistance for Writing Scientific Alt Text. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 886–906.
 - [52] 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.
 - [53] Arjun Srinivasan and Vidya Setlur. 2021. Snowy: Recommending Utterances for Conversational Visual Analysis. In *The 34th Annual ACM Symposium on User Interface Software and Technology (UIST) (Virtual Event, USA) (UIST '21)*. Association for Computing Machinery, New York, NY, USA, 864–880. <https://doi.org/10.1145/3472749.3474792>
 - [54] Chase Stokes, Cindy Xiong Bearfield, and Marti A Hearst. 2023. The Role of Text in Visualizations: How Annotations Shape Perceptions of Bias and Influence Predictions. *IEEE Transactions on Visualization and Computer Graphics* (2023).
 - [55] Chase Stokes and Marti A Hearst. 2021. Give Text A Chance: Advocating for Equal Consideration for Language and Visualization. In *NL VIZ: Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis*. IEEE.
 - [56] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A Hearst. 2022. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1233–1243.
 - [57] Chris Stolte, Diane Tang, and Pat Hanrahan. 2008. Polaris: A System for Query, Analysis, and Visualization of Multidimensional Databases. *Commun. ACM* 51, 11 (Nov 2008), 75–84. <https://doi.org/10.1145/1400214.1400234>
 - [58] 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 (Yokohama, Japan) (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 16, 17 pages. <https://doi.org/10.1145/3411764.3445354>
 - [59] Nicole Sultanum and Arjun Srinivasan. 2023. Datatales: Investigating the Use of Large Language Models for Authoring Data-driven Articles. In *2023 IEEE Visualization and Visual Analytics (VIS)*. IEEE, 231–235.
 - [60] Benny J Tang, Angie Boggust, and Arvind Satyanarayan. 2023. Vistext: A Benchmark for Semantically Rich Chart Captioning. *arXiv preprint arXiv:2307.05356* (2023).
 - [61] Strother H. Walker and David B. Duncan. 1967. Estimation of the Probability of an Event as a Function of Several Independent Variables. *Biometrika* 54, 1/2 (1967), 167–179. <http://www.jstor.org/stable/2333860>
 - [62] Yun Wang, Zhitao Hou, Leixian Shen, Tongshuang Wu, Jiaqi Wang, He Huang, Haidong Zhang, and Dongmei Zhang. 2022. Towards Natural Language-based Visualization Authoring. *IEEE Transactions on Visualization and Computer Graphics* 29, 1 (2022), 1222–1232.
 - [63] Yun Wang, Zhida Sun, Haidong Zhang, Weiwei Cui, Ke Xu, Xiaojuan Ma, and Dongmei Zhang. 2019. Datashot: Automatic Generation of Fact Sheets from Tabular Data. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 895–905.
 - [64] Zhibiao Wu and Martha Palmer. 1994. Verbs Semantics and Lexical Selection. In *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics (Las Cruces, New Mexico) (ACL '94)*. Association for Computational Linguistics, USA, 133–138. <https://doi.org/10.3115/981732.981751>
 - [65] Victoria Yaneva, Irina Temnikova, and Ruslan Mitkov. 2015. Accessible Texts for Autism: An Eye-Tracking Study. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 49–57.
 - [66] Li Yujian and Liu Bo. 2007. A Normalized Levenshtein Distance Metric. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29, 6 (2007), 1091–1095.