

# EMPHASISCHECKER: A Tool for Guiding Chart and Caption Emphasis

Dae Hyun Kim , Seulgi Choi , Juho Kim , Vidya Setlur , and Maneesh Agrawala 

**Abstract**—Recent work has shown that when both the chart and caption emphasize the same aspects of the data, readers tend to remember the doubly-emphasized features as takeaways; when there is a mismatch, readers rely on the chart to form takeaways and can miss information in the caption text. Through a survey of 280 chart-caption pairs in real-world sources (e.g., news media, poll reports, government reports, academic articles, and Tableau Public), we find that captions often do not emphasize the same information in practice, which could limit how effectively readers take away the authors' intended messages. Motivated by the survey findings, we present EMPHASISCHECKER, an interactive tool that highlights visually prominent chart features as well as the features emphasized by the caption text along with any mismatches in the emphasis. The tool implements a time-series prominent feature detector based on the Ramer-Douglas-Peucker algorithm and a text reference extractor that identifies time references and data descriptions in the caption and matches them with chart data. This information enables authors to compare features emphasized by these two modalities, quickly see mismatches, and make necessary revisions. A user study confirms that our tool is both useful and easy to use when authoring charts and captions.

**Index Terms**—Chart and text takeaways, visual prominence, authoring, captions

## 1 INTRODUCTION

Authors often pair charts and caption text together to convey information about data (e.g., in news articles, academic papers, and reports). For example, in Figure 1, the peak around 1981 is visually prominent. The caption text also emphasizes portions of the data by referring to specific points or ranges of data values. The caption text ‘peak in 1981’ emphasizes the visually prominent peak. When the chart and caption text emphasize the same aspects of the data, as in this example, people tend to remember those aspects of the data as takeaways [37, 79, 80]; when they emphasize different aspects of the data (e.g., the caption text emphasizes the ‘dip between 2008 and 2012’, but the fall is not visually prominent in the chart), people remember the visually prominent data features in the chart, not the data emphasized in the text [37, 80] and question the credibility of the information in the chart and text [39].

But how often do authors really create charts and caption text that emphasize the same aspects of data? Analyzing time-series charts and their captions in real-world publications (Section 3), we find that professional authors match chart and caption emphasis about 65% of the time and mismatches in the remaining 35%. In a survey of chart-caption pairs on Tableau Public [67], a community for the general public to share visualizations created using Tableau Software [68], we find that mismatches are even more common among the general public.

To help authors convey their messages effectively, we present EMPHASISCHECKER, a caption-and-chart checker tool that takes time-series charts as input. The tool highlights the visually emphasized chart features and the data features emphasized in the caption. The user can then decide how to update the caption or the chart to better align their emphasis. Our tool follows the model of spell- and grammar-checkers [25, 30, 42], facilitating the process of locating potential mismatches between chart and caption text, while leaving the final decision of how to resolve the issue to the author. In this work,

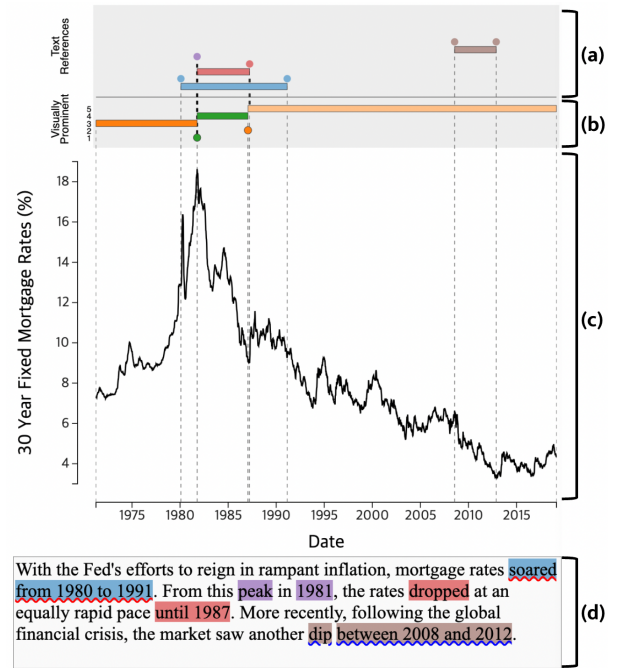








Fig. 1: As the author writes a caption about the chart (c) in the textbox (d), EMPHASISCHECKER shows the chart's visually prominent features in (b) and the text references to the chart features (a). The interface shows (b) visually prominent chart features (unmatched features in orange  and matched features in green , marks above the chart). It uses circles to depict point features (e.g., local extrema; the peak around 1981) and bars to depict trend features (e.g., the rising trend up to 1981). In addition, it shows (a) references between the chart and the text (i.e., blue , red , purple , and brown  marks at the top of the page and on the text). In the input text box (d), the tool adds a red squiggly underline (~~~~) on the phrase ‘soared from 1980 to 1991’, a type of ‘soared from 1980 to 1981’ because the phrase does not match the data in the chart. The tool also adds a blue squiggly underline (~~~~) on the phrase ‘dip between 2008 and 2012’ because the phrase does not match any of the prominent chart features.

- Dae Hyun Kim is with KAIST. This work began when Dae Hyun was at Stanford University. E-mail: [dhkim16@cs.stanford.edu](mailto:dhkim16@cs.stanford.edu)
- Seulgi Choi is with KAIST. E-mail: [igules8925@kaist.ac.kr](mailto:igules8925@kaist.ac.kr)
- Juho Kim is with KAIST. E-mail: [juhokim@kaist.ac.kr](mailto:juhokim@kaist.ac.kr)
- Vidya Setlur is with Tableau Research. E-mail: [vsetlur@tableau.com](mailto:vsetlur@tableau.com)
- Maneesh Agrawala is with Stanford University and Roblox. E-mail: [maneesh@cs.stanford.edu](mailto:maneesh@cs.stanford.edu)

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: [reprints@ieee.org](mailto:reprints@ieee.org). Digital Object Identifier: xx.xxx/TVCG.201x.xxxxxxx

we focus on time-series line charts as a first step towards more general tools as they are among the most common type of charts on the Web [2].

Our tool includes a *time-series prominent feature detector* that identifies the visually prominent features of the chart. It also includes a *text reference extractor* that identifies data emphasized in the text. The EMPHASISCHECKER interface is designed so that users can then compare the chart-emphasized data features with the text-emphasized data features (Figure 1). For the prominent feature detector, we introduce a new  $\epsilon$ -persistence technique based on the Ramer-Douglas-Peucker line simplification algorithm for approximating visual prominence. The chart-text reference extractor uses heuristics based on example analysis, modules within the Stanford CoreNLP toolkit [9, 12, 22, 48], and BERT embeddings [16] to find time references and data descriptions in the caption text and matches them with chart data. EMPHASISCHECKER visualizes the results of these components to the user as a part of a chart-caption authoring interface shown in Figure 1.

We evaluated the two components using the charts from Kim et al. [37]. We find that our time-series prominent feature detector outperforms a state-of-the-art method for prominent feature detection [31] while performing at the level of crowdsourced prominent features [37]. For human-written captions on the real-world charts in the corpus, we find that our text reference extractor correctly identifies text references to charts for 63.41% of the sentences. Finally, through a user study, we find that EMPHASISCHECKER is both useful and easy to use when authoring charts and captions.

In summary, the contributions of this work include:

- a survey of what charts and their captions in the real-world emphasize;
- EMPHASISCHECKER, a caption-and-chart checker tool that guides authors to create charts and captions whose emphasis match chart emphasis; and
- algorithms for detecting visually prominent features in time-series line charts and text references to chart features.

## 2 RELATED WORK

Our work is related to two main areas of prior work: (1) how people read charts and text and (2) supporting document authoring.

### 2.1 How People Read Charts and Text

Prior research has shown that visualizations and text complement each other when communicating information about the data. Specifically, Elzer et al. [19] and Carberry et al. [8] found that visualizations often contribute information not available in the text alone. On the other hand, researchers have found that text guides readers’ attention while viewing visualizations [29, 71, 79] and improves memorability of the key messages in visualizations [5, 37, 80].

The two representations facilitate different tasks for readers; Ottley et al. [54] found that readers easily *identify* key information with visualizations and easily *extract* key information from text. Despite the complementary nature of visualizations and text, prior studies did not find particularly synergistic benefits in comprehending information present in both chart and caption text [35, 50, 55, 56].

Researchers have noticed that one of the major problems limiting the synergistic benefits of using visualizations and text together is *split attention* as readers have to look back and forth between the spatially separated representations [4, 66, 71]. Whitacre and Saul [75] and Ottley et al. [54] concluded that readers tend not to integrate the information between the visualizations and text. The research community has put forth a range of solutions for the split attention problem. Tufte [71] introduced *sparklines*, which are word-sized line charts embedded into the text. Goffin et al. [24] and Beck and Weiskopf [3] extended the idea by adding interactivity to the sparklines. In addition, multiple systems (e.g., table-text reference display systems [1, 36], visualization-text reference display systems [40, 41, 43, 58]) display the references between visualizations and text to reduce the effort needed to locate relevant information.

The research community has recently begun diving deeper into the relationship between visualizations and text by studying how comprehension depends on the content of the visualizations and text. Kong

et al. [38, 39] studied how slants, framing in chart titles, as well as misalignments between charts and text, affect what people read from charts and their titles. They found that while people identify bias, they still consider charts to be impartial. Whitacre and Saul [75] studied high school students and found that they had difficulty identifying the inconsistencies between graphs and their captions. Kim et al. [37] found that when charts and captions emphasize the same feature in the data, people tend to take away the message related to the doubly emphasized feature. When the caption describes a feature that is not visually prominent in a chart, readers are more likely to consider the visually prominent chart features as carrying the key messages. Subsequent work found that semantic levels of text [47, 65, 80], as well as its placement [65], can influence how readers integrate charts with text. Although our work on the EMPHASISCHECKER tool is targeted toward authors, the principles behind how the tool helps authors write effective chart-text pairs, rely on these theories about how readers read charts and text together.

### 2.2 Supporting Document Authoring

Document authoring tools often support authors so that they can easily write high-quality text with a clear exposition. Spell-checkers [18] have been incorporated into many writing environments, including word processors and e-mail [25, 26, 45, 51]. Grammar checkers are also available in such environments through software extensions or plug-ins, such as Grammarly [30] and LanguageTool [42]. More recently, writing environments have incorporated AI-based autocompletion [27] that tries to predict what the writer intends to write. Similar to these tools, EMPHASISCHECKER is designed to help people write documents, but instead of purely analyzing text, EMPHASISCHECKER considers chart and text pairs and analyzes the visuals as well as the words together.

The research community has developed methods for automatically generating caption text for visualizations. Many of these techniques are designed to generate basic captions that simply explain how the visualization encodes data (e.g., [15, 68, 77, 81]). A few techniques go beyond basic captions and generate descriptions and summaries of features (e.g., ranking, extrema, trends) in the visualizations, such as techniques based on Bayesian models [8, 20] and neural networks [10, 11, 53, 60]. Rather than generating captions, Contextifier [31] adds annotations based on financial news headlines to noteworthy features of stock charts. Unlike the fully-automated systems, EMPHASISCHECKER follows the interface paradigm of a spell- or grammar-checker by guiding the author while they have complete control over their writing.

More advanced tools provide assistance beyond adding text to visualizations. TimeLineCurator [23] is a web-based timeline authoring tool that automatically extracts event data from temporal references in unstructured text documents using natural language processing, along with controls for curating and editing the events. VizByWiki [46] retrieves visualizations relevant to given news articles from Wikimedia Commons [76] to enrich articles. PostGraphe [21] creates text and graphics based on tabular input data and users’ intents. Kori [44] takes a mixed-initiative approach in helping authors construct interactive references between visualizations and text into their documents with suggestions based on natural language processing techniques. Cross-Data [13] links text and the underlying data to reduce efforts in writing data documents and data exploration. Although EMPHASISCHECKER leverages references between visualizations and text similarly to these prior works, our tool also analyzes the *chart data* while identifying references in order to establish comparisons between chart and text emphasis. Furthermore, the reference extraction is only a component of EMPHASISCHECKER whose end goal is to provide an interface for identifying mismatches between visualizations and the text.

## 3 A SURVEY OF LINE CHARTS IN THE WILD

To understand whether charts and captions emphasize the same information in practice, we conducted a survey of chart-text pairs in various real-world sources. We specifically looked at charts and captions written by professionals and circulated through various publishing venues,

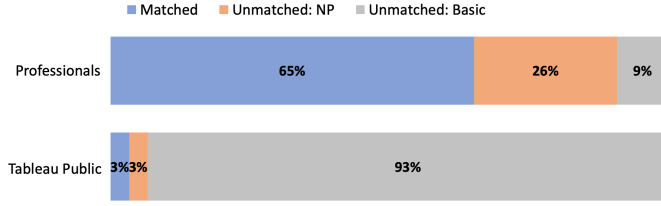


Fig. 2: Distribution of chart and caption emphasis in chart-article pairs. Professionals often match chart and text emphases (blue ■ segment in the top bar) but occasionally do not (orange ■ (Unmatched: NP (non-prominent); captions only describing non-prominent chart features) + gray ■ (Unmatched: basic; basic captions that do not point to specific chart features) segments in the top bar). On the other hand, captions on Tableau Public are predominantly basic captions (gray ■ segment in the bottom bar). The values have been rounded to the nearest whole numbers and may not sum to 100%.

as well as charts and captions published on Tableau Public [67], a community of professional and non-professional authors for sharing charts and captions created with Tableau Software [68], a tool that suggests basic captions about data fields and chart encoding to its users.

### 3.1 Dataset

For chart-caption pairs written by authors, we sampled a total of 250 chart-article pairs (189 unique articles) from various publishing venues to obtain a representative set. The venues include news media (New York Times [70], The BBC [69], Vox [74]), poll reports (Pew Research [57]), governmental and intergovernmental organization reports (US Treasury [73], International Monetary Fund [32], International Labour Organization [72], etc.), and academic articles (Nature [64]). In addition, we sampled 30 chart-article pairs (21 unique charts) from Tableau Public. We programmatically scraped the Web for articles published up to two years from the collection date and filtered for line charts using the chart classification pipeline from Poco and Heer [59]. Because Pew Research did not allow scraping without permission, we performed the collection process manually for that site. We include the links to these articles in the supplemental material.

### 3.2 Analysis Method

We performed analysis on the chart-article pairs in two steps; we identified (1) chart emphasis and then (2) text emphasis.

**Step 1: Identify chart emphasis.** For each chart image in the dataset, two of the authors of this paper independently labeled the visually prominent features. Afterward, the authors merged their labels and whenever the labels did not match, the two authors shared their reasoning for their annotations with one another to arrive at a consensus. If no consensus was reached, the authors used the opinion of a third individual who works in visualization research but is not a co-author to determine the prominent features of the chart. At this stage, the authors did not read any of the text of the article to avoid biasing their perception of visually prominent features [79].

**Step 2: Identify text emphasis.** The two authors read through the article text to identify all paragraphs that mention any information about the chart by looking for mentions of the phrases in the chart title, axis names, or chart numbering (e.g., ‘Figure X,’ ‘Chart X’). They each independently perused the sentences in each of the paragraphs, chart titles, and textual annotations within the charts to identify any references to prominent features and non-prominent features. The authors again discussed the annotations to arrive at a consensus, using the third individual for arbitration when necessary to resolve conflicts.

### 3.3 Results

Figure 2 shows the results of our analysis. We observed a visible dividing line between chart captions written by professional authors circulated through publishing organizations and chart-caption pairs on Tableau Public.

**Professionals often match emphasis but occasionally do not.** We found that 65% chart-caption pairs made by professional authors match emphases (blue ■ segment in the top bar of Figure 2); the chart emphasizes the author’s message in the text, and the text explains the visually prominent features in the chart. Yet, emphasis mismatches occur regularly; in 35% of the chart-caption pairs, the chart emphasizes a feature in the data different from the text, and the text either describes features that are not visually prominent (26%; first orange ■ bar in Figure 2) or only describes how the chart encodes information (9%; first gray ■ bar in Figure 2). Such mismatches suggest that there is some room for improvement, even in professionally authored documents.

**Tableau Public authors often write basic captions.** In Tableau Public, the overwhelming majority (93%) of the text descriptions for charts are basic captions that do not describe the specific features in the charts (gray ■ segment in the bottom bar of Figure 2). We hypothesize that this result is due in part because the Tableau Public authoring software defaults to providing basic captions. Authors do not have to put in the extra effort required to discuss the features visible in the charts. Unfortunately, prior work has found that readers find such basic captions of little use [47] and that the basic captions play no role in helping readers understand what they should take away [37]. Hence, the high ratio of basic captions indicates that these authors could benefit from further support and guidelines for authoring charts and text.

## 4 USAGE SCENARIO

We first describe a usage scenario of EMPHASISCHECKER to motivate the design of the tool and illustrate how it guides the user when writing a text caption for a chart. Tess, a policymaker, wants to add a chart showing the real home price index data to a presentation she plans to give to her fellow policymakers. As a part of her presentation, she would like to make a case for building more homes based on past data. To ensure that she effectively gets her message across, she decides to use the EMPHASISCHECKER tool.

**Viewing visually prominent features in the chart.** (Figure 3a above textbox) Tess starts by loading the data into EMPHASISCHECKER. The time-series chart shows the real home price index over time. The tool displays the visually prominent features in the chart as orange circles (●) and bars (■) just above the chart. EMPHASISCHECKER shows the recent low point in 1997 and the rising segment afterward as the top two most prominent features (darker shade of orange ■), followed by the global minimum point around 1920 and two other less prominent features (lighter shades of orange ■).

**Typing a basic caption.** (Figure 3a) Seeing the time range of the chart, she types “The chart shows the real home price index between 1890 and 2006.” in the text box below the chart and hits the [SHIFT-ENTER] command to run text analysis and assess text emphasis with respect to the chart. After the code runs to completion and the textbox is re-enabled, she sees that there is no change in the region above the chart. Based on this information, she understands that the text she typed in has no reference to any specific features in the chart.

**Typing caption text that matches the most prominent visual feature.** (Figure 3b) Tess then returns to thinking about the most prominent feature. She hovers over the orange circle (●) closest to the bottom and the orange bar (■) right above it to see where they lie on the chart. She observes the spike after 1997 in the context of the chart and convinces herself that the spike is not only visually prominent but also that she can point to this spike as proof of the dire situation of the housing market. She decides to cover the feature in her caption and types “The housing prices have skyrocketed starting around 1997 and we need to act.” After hitting [SHIFT-ENTER] to run analysis on the text, the interface highlights the phrase ‘skyrocketed starting around 1997’ in blue ■ and displays a blue circle (●) on the year 1997, the endpoint explicitly mentioned in the text, and a bar (■) starting in the year 1997 to show the reference between the chart and caption. The interface also highlights the top two prominent chart features in green ■ to indicate that Tess has matched the emphasis in the caption she has written so far.



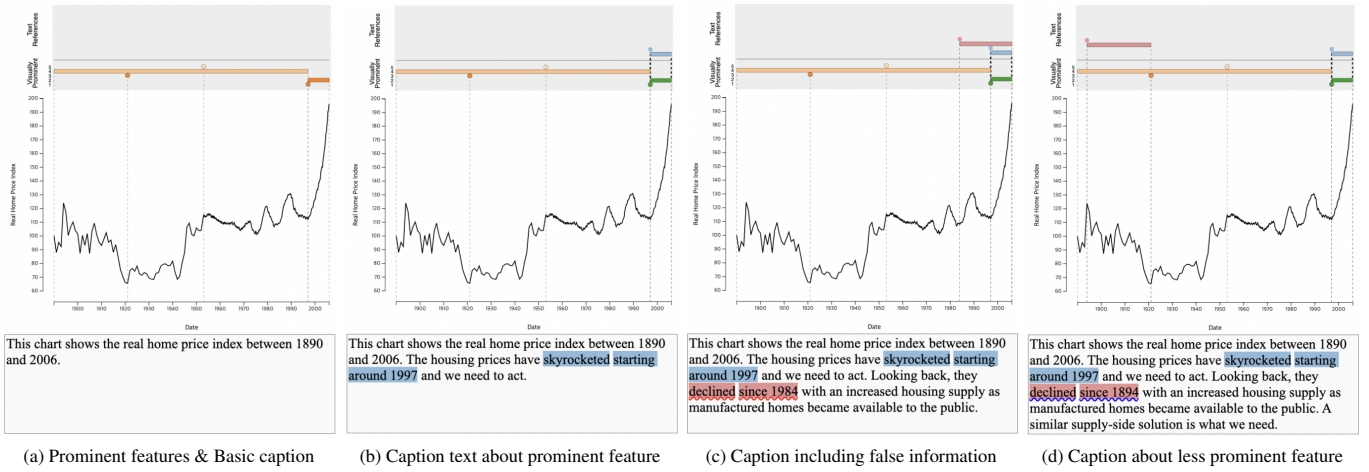


Fig. 3: Views of the EMPHASISCHECKER interface from the usage scenario. The chart shows the real home price index between 1890 and 2006. (a) Prominent features are shown on top with a basic caption not describing any specific feature. (b) Caption text matches the most prominent visual feature (sharp rise on the right; blue highlight in the UI). (c) Typo in the caption text indicated by a red squiggly underline (↗) on ‘declined since 1984’. (d) Caption text matching a less prominent feature, indicated by a blue squiggly underline (↘) on ‘declined since 1894.’

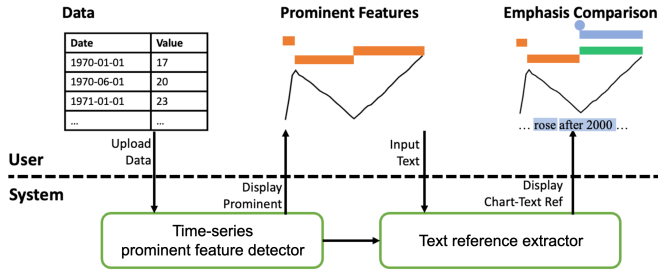


Fig. 4: EMPHASISCHECKER tool overview. A user first uploads time-series data (including axis ranges and aspect ratio). Based on the input, the *time-series prominent feature detector* detects the visually prominent features and displays them to the user (Section 5.1). The user can type text based on the prominent features they see in the chart. When triggered, the *text reference extractor* identifies the references between the chart and text and displays them to the user, with comparisons with the prominent features (Section 5.2).

**Typing caption text with an error.** (Figure 3c) Tess now looks for falling trends in the chart and discovers one between 1894 and 1921. She looks up the reason for the fall and thinks that it would support her message well; she types, “Looking back, they declined since 1984 with an increased housing supply as manufactured homes became available to the public.” Tess hits [SHIFT-ENTER] and this time, she is surprised to see a red squiggly underline (↗) under the phrase ‘declined since 1984’ in her caption. When she hovers over the text ‘declined since 1984’, she sees that the time segment that she is referring to (red bar (■) above the chart in Figure 3c) is not the one she intended and soon realizes that she mistyped ‘1984’ instead of ‘1894’.

**Pushing through with caption text about a less prominent feature.** (Figure 3d) Tess fixes the typo and completes her caption by adding the sentence “A similar supply-side solution is what we need.” She then confirms the change by pressing [SHIFT-ENTER]. She finds that the detected time range has been revised to the one she initially intended, but she realizes that there is a blue squiggly underline (↘) under the same phrase ‘declined since 1894’. She sees that the red bars (■) matched to the phrase do not match with any of the top five prominent features but decides that she wants to push forward with her caption. Before finalizing her caption, she looks over the unmatched prominent features, still shown in orange (●), and thinks about whether there are additional features she should describe in her caption. She sees that the third most prominent feature corresponding to the global minimum in 1921 is the other end point of the downward trend she just wrote

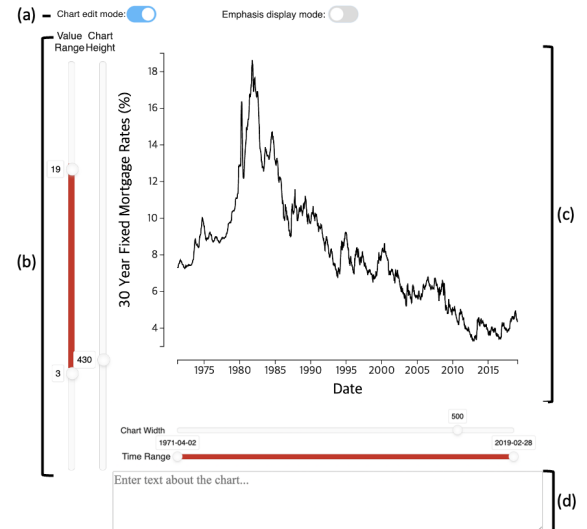


Fig. 5: EMPHASISCHECKER’s interface for authoring charts and captions. (a) The interface includes two switches, the *chart edit mode* for toggling the sliders (Component (b)) and the *emphasis display mode* that toggles the display of references (Figure 1a) and the visually prominent features (Figure 1b). The figure shows the state with chart edit mode on and emphasis display mode off. (b) When the chart edit mode is on, the user can manipulate the vertical and horizontal sliders to edit the dimensions of the chart and the axes ranges. (c) The user can hover over the chart to view tooltips showing the underlying data values. (d) The user can type the caption in the textbox.

about and decides that she does not need to describe the point explicitly. Finally, Tess looks over the fourth and the fifth most prominent features but decides that they are both irrelevant to her needs and not very prominent based on the lighter shade of orange (■). She concludes the chart-caption authoring process.

## 5 COMPONENTS OF EMPHASISCHECKER

EMPHASISCHECKER allows the user to edit both the chart and caption (Figure 5). The user can not only write captions (Figure 5d) but also edit the dimensions and the x- and y-axis ranges (Figure 5b). The tool provides two main features in addition to this basic chart and caption editing interface (Figure 4): (1) The *time-series prominent feature*

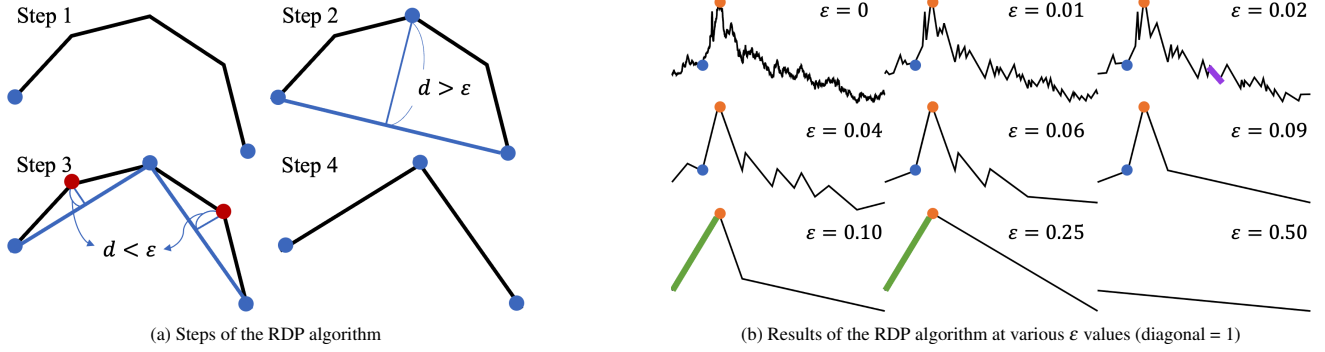


Fig. 6: (a) [Step 1] The RDP algorithm [17, 61] takes an input polyline and a distance threshold parameter  $\epsilon$ . [Step 2] It first checks whether the distance between the line segment connecting its endpoints and the farthest point on the polyline ( $d$ ) is greater than  $\epsilon$ . If so, the farthest point captures an important deviation in the polyline. [Step 3] The algorithm breaks the polyline at this point and is recursively applied to both halves. If  $d < \epsilon$ , the deviation at the farthest point is considered not important. [Step 4] The algorithm simplifies the polyline to the line segment between the endpoints, resulting in a simplified polyline. (b) Applying the RDP algorithm on the time series in Figure 3 with various  $\epsilon$  values yields simplified curves with various levels of detail; lower values of  $\epsilon$  preserves local details whereas higher values of  $\epsilon$  draws out the global patterns. For very high values of  $\epsilon$ , the polyline degenerates into a single line segment ( $\epsilon = 0.50$ ). The most prominent orange peak (●) in this chart persists through all  $\epsilon$  values between 0 and 0.25 and hence has an  $\epsilon$ -persistence of 0.25. In comparison, the blue point (●) to the left of the peak persists through  $\epsilon$  values between 0 and 0.09 and hence has an  $\epsilon$ -persistence of 0.09. The green rise up to the peak (■), the second most prominent feature in this chart, is first seen at  $\epsilon = 0.10$  and persists until  $\epsilon = 0.25$  and hence has an  $\epsilon$ -persistence of 0.15. On the other hand, the less prominent trend along the decrease (purple; ■) persists only briefly around the  $\epsilon$  value of 0.02 and hence has negligible  $\epsilon$ -persistence.

*detector* computes the visual prominence of the chart features in the time-series line chart (Figure 1b) and (2) the *text reference extractor* identifies references between the caption text and the time-series line chart (Figure 1a).

### 5.1 Time-Series Prominent Feature Detector

The *time-series prominent-feature detector* identifies visually prominent features of a given time-series line chart by looking for features that persist through multiple levels of detail.

Prior work [33, 34, 62, 63] has sought ways to simplify line charts to enhance the perceivability of patterns in line charts that are often obscured by the variations in the chart. However, these methods leave the interpretation of the features to the users and do not provide ways to compare the features. Based on the strength of the Ramer-Douglas-Peucker (RDP) line simplification algorithm (Figure 6a) in enhancing important extrema in time series [63], we modify the vanilla RDP algorithm to compute  $\epsilon$ -persistence for measuring visual prominence of both point and trend features.

The RDP algorithm (Figure 6a) takes a polyline and a distance threshold parameter  $\epsilon$ , and modifying  $\epsilon$  changes the level of detail in the resulting simplified polyline (Figure 6b). Our key observation is that visually prominent features persist through multiple  $\epsilon$  values (e.g., orange peak (●) in Figure 6b).

Based on the observation, our persistence algorithm initiates by running the RDP algorithm on the input time series data multiple times. On each run, the algorithm varies the  $\epsilon$  threshold, stepping through the range of values  $[0.0, 0.25]$  with a step size of 0.01, relative to a chart whose diagonal length is normalized to equal 1. This normalization allows us to reduce the dependencies of the algorithm on the scale of the rendered charts (e.g., the algorithm would treat a 300px-by-400px chart the same as a 600px-by-800px chart). We chose the step size to have a sufficient amount of resolution into the range of  $\epsilon$  values while being able to be run in real-time. We do not proceed to higher values of  $\epsilon$  as most charts degenerate into just two points and no longer provide useful signal (Figure 6b,  $\epsilon = 0.50$ ).

Finally, we define the  $\epsilon$ -persistence of each point in the input time series as the greatest  $\epsilon$  value at which the point is deemed important and included in the simplified polyline by the RDP algorithm, capped to 0.25 by our choice of the range of  $\epsilon$  values. In Figure 6b, for example, the prominent orange peak (●) persists all the way up to  $\epsilon = 0.25$  and therefore has an  $\epsilon$ -persistence of 0.25.

We extend this  $\epsilon$ -persistence measure for points to a measure of  $\epsilon$ -persistence of trends between any two points  $p_i$  and  $p_j$  on the original polyline. Observe the green trend (■) to the left of the orange peak (●) in Figure 6b. Below  $\epsilon = 0.10$ , the trend is broken into smaller trends by the blue point (●) left of the orange peak (●). Somewhere above  $\epsilon = 0.25$ , the orange peak (●) is removed and the trend is absorbed into a larger trend. Thus, the green trend (■) persists from  $\epsilon = 0.10$  to  $\epsilon = 0.25$  and hence has  $\epsilon$ -persistence of 0.15. Generalizing this observation, we can compute  $\epsilon$ -persistence of trends between any two points  $p_i$  and  $p_j$  by subtracting the maximum  $\epsilon$ -persistence of points lying between  $p_i$  and  $p_j$  from the minimum  $\epsilon$ -persistence for  $p_i$  and  $p_j$  and adding 0.01, the  $\epsilon$  step size we use.

EMPHASISCHECKER displays the top five most prominent features above the chart in order (Figure 1b), the most prominent closest to the chart, using circles to depict point features and bars to depict trend features. To help users get a sense of the degree of prominence of these features, the tool displays more prominent features using darker shades (■/■; e.g., the two most prominent features in Figure 3a) and less prominent features using lighter shades (■/■; e.g., the fifth most prominent feature in Figure 3a).

### 5.2 Text References Extractor

Our *text reference extractor* uses a five-step pipeline to determine the matches between chart features and caption text (Figure 7).

**Step 1. Extract time references in caption text.** To detect the time ranges described in the caption text, our tool first uses the Named Entity Module in the Stanford CoreNLP toolkit [9, 22, 48]. It first identifies phrases describing points in time (e.g., ‘1970’, ‘March 2020’) or durations (e.g., ‘the last six months’). Since the time-series data in the chart is often represented at a finer granularity than the dates and durations mentioned in the caption text, we convert the detected time points and durations into a set of time points at the finest granularity of the time-series data. In practice, the time-series data we have worked with is at a granularity of days, weeks, months or years.

The time points mentioned in the caption text occasionally signify the start or the end point of a time range. For example, ‘between 1970 and 1980’ or ‘from 1970 to 1980’ indicates the start point 1970 and 1980. To detect these start and end points, we utilize the context template patterns near the time words (Table 1). In the example phrases, our tool would detect the time range 1970-1980. However, caption text sometimes gives only one endpoint of a time range (e.g., ‘since Nov

**Caption Text:** “The 30-year mortgage rates peaked in 1981 and then declined sharply until 1987.”

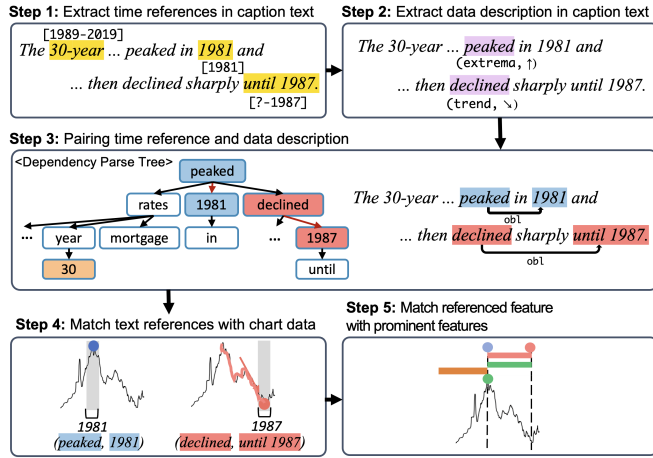


Fig. 7: The five steps of the *text references extractor*.

1997', 'after March 2020'). In these cases, we detect one endpoint and leave the other endpoint undetermined (e.g., 1997/11-?, 2020/03-?).

**Step 2. Extract data descriptions in caption text.** We next extract trend descriptions (i.e., upward, downward) and local extrema (i.e., maximum and minimum) in the caption text by looking for a predefined set of keywords (Table 2) in the caption. We compiled these keywords by examining captions collected through our survey of line charts in the wild and expanding them using a thesaurus [49].

We first compared lemmas (i.e., base forms of words; e.g., ‘*rising*’ - ‘*rise*’, ‘*soared*’ - ‘*soar*’) of each word in the input sentence with the lemmas of words in our compiled list of keywords to find any exact matches. To capture any synonyms that we may have missed in our list, we use the cosine similarity of the BERT contextual embedding vectors [16] between the words in the input sentence and the words in the keywords list. We obtained the BERT contextual embedding vectors of the keywords by running a pre-trained BERT model [16] on the sentences from which the words originated and extracting the results of the final layer. For each of the words in the input sentence, we similarly obtain its BERT contextual embedding vector. We consider the word in the sentence a description of the data if the cosine similarity of its BERT contextual embedding vector and any of the BERT contextual embedding vectors of the keywords is greater than the empirically determined threshold of 0.7.

**Step 3. Pairing time references with data descriptions.** Complex caption sentences may sometimes refer to more than one point or duration in time along with more than one time feature. For example, the sentence “*The 30-year fixed mortgage rates peaked in 1981 and then declined sharply until 1987*” includes the time duration ‘30-year’, and points in time ‘1981’ and ‘until 1987,’ with features ‘peaked’ and ‘declined.’

To connect the times with their features, we first use the Stanford CoreNLP dependency parser [12, 48] to obtain a dependency tree for the caption (Figure 7 Step 3 left). We match each of the time references to the closest data description within the dependency tree that does not have a closer time reference. When two time references complement each other, with one only mentioning the start point and the other only mentioning the end point, we combine them into a single time range. For example, in Figure 8b, Sentence 2, “*From 1950, North Korea’s GDP increased quite rapidly until 1985,*” the data description ‘*increased*’ is closest to the two time references, ‘*from 1950*’ (1950-?) and ‘*until 1985*’ (?-1985) and the tool combines the two end dates into a single range, 1950-1985. We discard time references or data descriptions not matched through this process. In our example, we follow the relation from ‘1981’ and ‘1987,’ to pair the corresponding time references with the data descriptions ‘*peaked*’ and ‘*declined*’ at a distance of 1, respectively. On the other hand, the time phrase, ‘*30-year*’ remains unmatched and is therefore discarded.

Table 1: List of context template patterns near the mention of time  $T$  that we use to determine whether  $T$  is a start point or an end point of a time range.

|                    | Context Template Patterns  |
|--------------------|--|
| <b>Start point</b> | since $T$ , after $T$ , from $T$ (to ...), following $T$ , between $T$ and ..., start / commence in $T$  |
| <b>End point</b>   | until $T$ , till $T$ , up to $T$ , preceding $T$ , prior to $T$ , previous to $T$ , (from ...) to $T$ , between ... and $T$ , end / stop / finish / terminate in $T$ |

Table 2: The word list EMPHASISCHECKER uses to identify data descriptions in caption text.

| Feature  | Type | Data Description Words  |
|----------|------|---|
| Trend    | ↗    | boost, climb, gain, grow, heighten, higher, increase, increment, jump, raise, rise, skyrocket, soar, spike, surge, up, uptick, upward                 |
|          | ↘    | contract, decline, decrease, diminish, dip, drop, down, downward, dwindle, lessen, lower, plummet, fall, falloff, reduce, shrink, sink, subside, wane |
| Extremum | ↑    | apex, biggest, cap, ceiling, greatest, highest, largest, max, maximum, most, peak, pinnacle, top, topmost, zenith                                     |
|          | ↓    | base, bottom, fewest, floor, least, lowest, min, minimum, smallest, tiniest, trough   |

**Step 4. Match text references with chart data.** The data descriptions and time references in the caption text obtained from the previous three steps may require further disambiguation to accurately pinpoint the referenced features in the chart. For example, the time reference ‘1981’ could refer to any of the points in the chart whose year is 1981. The time reference ‘until 1987’ could end at any of the points in the chart whose year is 1987, and the start point is yet to be determined.

To find a suitable selection, we join the information specified in the data descriptions and the time references with the chart data. When the data description refers to a local maximum, our tool infers that the chart point being referenced is the point within the time reference with the greatest value; if it refers to a local minimum, our tool infers that the chart point being referenced is the point within the time reference with the smallest value. Hence, in our example, we select the global maximum point in the chart for the pair (*'peaked'*, *'1981'*) (blue point (●) in Figure 7). For rising trends, we select the point with the minimum value among the start point candidates and the point with the maximum value among the endpoint candidates, and vice versa for falling trends. Thus, for the pair (*'decreased'*, *'until 1987'*), the tool matches the decreasing trend between the maximum point in 1981 and the minimum point in 1987 (green line (■) in Figure 7).

EMPHASISCHECKER highlights the time references and data descriptions in the caption text and adds bars (for trend features) and circles (for point features) in the region above the prominent features (Figure 1a) using the same colors. Users can hover over either the text highlighting or the bars and circles to view the set of referenced points on the chart. During this process, EMPHASISCHECKER performs a basic check for factual errors between references to trend features in the text and the actual change of data between the two endpoints of the trend feature. If the data reference is an upward trend but the data decreases, or vice versa, our system detects a factual error. For example, we detect pair (*‘soared’, ‘from 1980 to 1991’*) in the first sentence of Figure 1. Although the text indicates that the value in 1991 would be higher than the value in 1980, this actually is not the case. The tool alerts authors of such factual errors by drawing a red squiggly underline (~~~~) on the time references and data descriptions in the text, similar to spell-checkers.

**Step 5. Match referenced feature with prominent features.** We finally compare the references with the prominent chart features returned by our prominent feature detector. We consider points features as a match if they are exactly the same. On the other hand, we detect a match between text references to time segments and a visually prominent trend if the intersection of the points in the two sets covers at least



95% of the union of the two sets. This way, we are able to detect that ('peaked', '1981') matches the most prominent feature and that trend feature referred to by ('decreased', 'until 1987') matches the fourth most prominent feature, whereas its endpoint of in 1987 does not match any feature as it is off from the third most prominent feature by couple weeks.

If a prominent feature is matched to a reference, EMPHASIS-CHECKER highlights the feature in green ■ (e.g., Figure 1 most prominent and fourth most prominent features). On the contrary, if the feature the caption text refers to is not matched to any text (e.g., Figure 1 Sentence 3), the tool alerts the authors of the emphasis mismatch by drawing a blue squiggly underline (~~~~) on the time references and data descriptions in the caption task, similarly to grammatical errors in grammar-checkers.

## 6 RESULTS AND EVALUATION

Figure 8 shows results of running the EMPHASISCHECKER tool on various charts and captions. The tool's components function accurately and provide correct guidance in Figure 8a-c, but err in cases shown in Figure 8d. To understand how well each component of EMPHASIS-CHECKER performs, we ran an evaluation on the time-series prominent feature detector and the text reference extractor.

### 6.1 Evaluation: Time-Series Prominent Feature Detector

To evaluate our time-series prominent feature detector, we used Kim et al.'s [37] corpus of 43 synthetic and real-world examples. We used their crowdsourced prominent features as a gold feature set to compare against. As another baseline for comparison, we also generated chart features using a state-of-the-art prominent feature detector, Contextifier [31], which uses the value and the first derivative at each point to compute the visual saliency of each point.

The features we generated with EMPHASISCHECKER matched the gold feature set far better than features generated by Contextifier. On average, 1.47 (49%) of EMPHASISCHECKER's top three prominent features matched the gold features, whereas only 0.81 (27%) of Contextifier's top three prominent features matched the gold features. We further analyzed the variance among crowd workers who labeled the gold feature set. That is, we compared each crowd worker's labeling of the top-three features against the average crowd workers' labels of these features and found a match of 1.72 (57%) of the time. This suggests that there is high variation among what people think of as the top three visually prominent features.

We include further analysis of our prominent feature detection algorithm in the supplemental material.

### 6.2 Evaluation: Text Reference Extractor

To understand how reliably our text reference extractor identifies text references to chart features, we ran our text reference extractor on 24 chart-caption pairs with 82 sentences collected through a user study (Section 7). Two of the authors of this paper independently reviewed each of the captions at sentence level along with the participant's stated intended message to determine the references included in each sentence. The two authors discussed any mismatches and asked a third-party visualization expert for interventions whenever the conflicts could not be resolved between the two authors.

Based on the labels, we ran our text reference extractor on the captions and classified errors in each sentence as one of the following:

- *False negative* (FN): the tool failed to extract an existing reference (e.g., Figure 8c Sentence 3).
- *False positive* (FP): the tool extracted a reference that is non-existent.
- *Intention mismatch* (IM): the tool correctly extracts the reference based on what is said in the sentence but detects a feature that is different from the author's intention (e.g., Figure 8d Sentence 1; the endpoint of the range is underspecified in the text, but the participant's intended endpoint is clearly 2008, not the end of the chart).

A sentence can include several of these errors and is considered correct if it includes no errors. For three sentences that included apparent typos (one written with the baseline tool (Figure 8a), two written with

EMPHASISCHECKER tool), we considered them correct if the text reference extractor correctly identified the mentioned time range.

Of the 81 sentences, the text reference extractor correctly identified the references in 57 sentences (70%), included FN errors in 22 sentences (27%), FP errors in 4 sentences (5%), and IM errors in 2 sentences (2%). Four of the sentences included multiple errors and were double-counted into multiple categories. We include further analysis of the error cases in the supplemental material.

## 7 USER STUDY

To evaluate how the EMPHASISCHECKER tool helps authors write charts and captions, we conducted a within-subjects study comparing the complete EMPHASISCHECKER tool (EMPHASISCHECKER condition) against a baseline system that only includes the chart and text authoring interface in Figure 5 (baseline condition). We consider the following two hypotheses:

[H1] Users will find the features in EMPHASISCHECKER useful in authoring charts and captions.

[H2] The features in EMPHASISCHECKER are easy to use.

We note that the accuracy of the text reference extractor was lower at 56% at the time of the user study due to an issue with our implementation of the time reference extraction that was later revised.

### 7.1 Participants

We recruited 12 participants through online communities within KAIST and personal referrals. We required participants to be capable of reading and writing in English and we administered a test to ascertain skills such as understanding the chart encoding, reading off values, and recognizing trends and extrema [37]. In order to model the target population of the tool, we further required that the participants regularly author charts and captions. All of the participants reported having regular exposure to chart-caption authoring for academic purposes (12 participants, P1-P12), presentations (11 participants), coursework (5 participants), industrial purposes (4 participants), articles (2 participants), social media (2 participants), and personal messages (2 participants).

### 7.2 Procedure

We conducted the study online through Google Meet [28] with screen sharing. We began the study with a pre-survey asking about the participant's background in authoring charts and text as well as their domain of expertise. We then showed the participant a chart and three types of captions: a basic caption, a caption about prominent chart features, and a caption about non-prominent chart features. After reading through the chart-caption pairs, they answered which caption is the most effective and their rationale. We then provided instructions about the baseline and EMPHASISCHECKER tools for authoring chart-caption pairs.

During the study, the participant chose two distinct time-series line charts either from the real-world corpus from Kim et al. [37] or one of their own. The participant authored one chart-caption pair with the baseline tool and another pair with the EMPHASISCHECKER tool in a counterbalanced order. After the use of each tool, the participant completed a post-task reflection survey that comprised three parts: (1) the participant's intended messages; (2) a usefulness assessment of the guidance provided by the tool and the expected capability of captions generated with the tool in communicating authors' main messages on 5-point Likert scales and free form comments about the benefits and disadvantages of using the tool; and (3) a usability assessment using the System Usability Scale (SUS) [6] and free-form comments about the usability of the tool.

After completing the two chart-caption pair authoring tasks, we asked the participant to *compare* their experience with the two tools as a post-survey. We specifically asked the participants to compare the usefulness and ease of use of the two tools in chart-caption authoring on 5-point scales (i.e., Baseline+2, Baseline+1, Neutral, EMPHASIS-CHECKER+1, EMPHASISCHECKER+2) and asked for rationales for their scores. We finally asked for any free-form comments on the two tools.

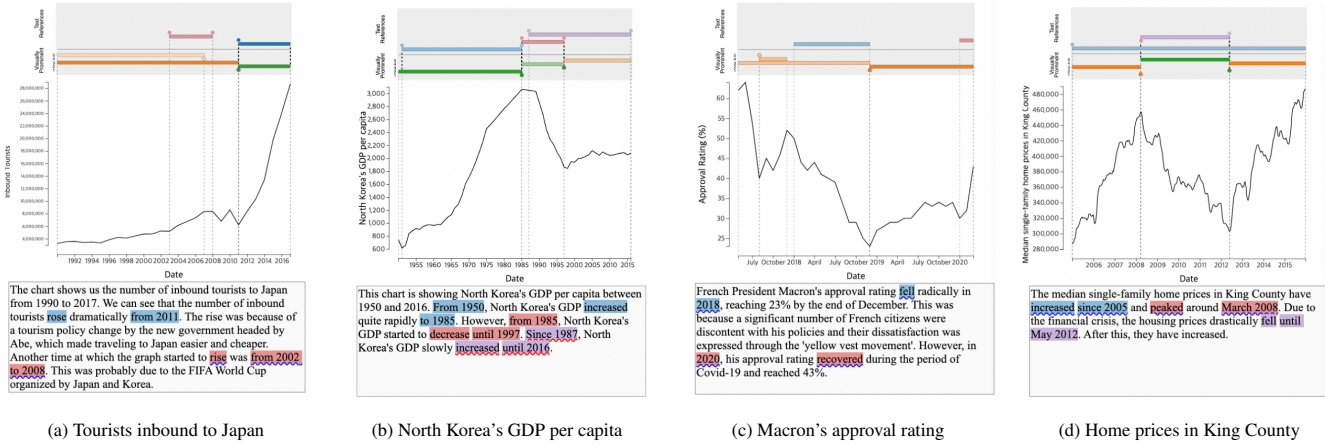


Fig. 8: Results of running EMPHASISCHECKER on charts and captions written by participants in the user study (Section 7). Participants wrote the chart-caption pairs in (a) and (b) with the EMPHASISCHECKER tool and the chart-caption pairs in (c) and (d) with the baseline tool. (a), (b), and (c) show chart-caption pairs for which EMPHASISCHECKER would provide suitable guidance. (a) Sentence 4 and (c) Sentence 3 mention non-prominent features and are underlined in blue (~~~~). (b) Sentence 4 includes a typo ('1987' instead of '1997'), and the tool underlines the typo in red (~~~). However, for (d) Sentence 1, EMPHASISCHECKER fails to correctly capture the intent in the phrase 'increased since 2015' and captures a wrong time range and also misses the peak around March 2008 due to an artifact of the RDP algorithm and underlines both blue (~~~~) despite being mentions of the most and the third most prominent features. The tool also completely misses references in (d) Sentence 3. Minor typos and grammatical errors (verb tense, plural-singular, etc.) have been fixed for illustration.

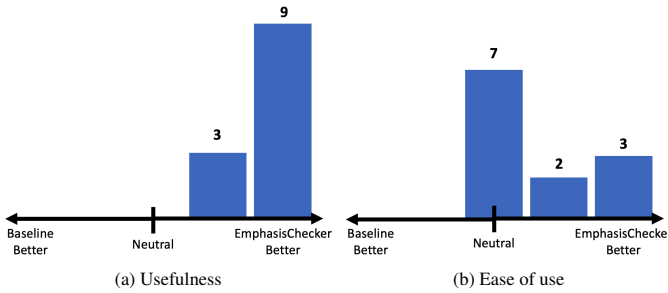


Fig. 9: Participants' comparison of baseline and EMPHASISCHECKER tools based on their usefulness and ease of use when writing charts and captions. The participants leaned towards EMPHASISCHECKER for usefulness and were neutral or slightly leaning towards EMPHASISCHECKER for ease of use.

The study lasted 60-75 minutes, and we compensated each participant with an equivalent of 15 USD as a direct deposit. We include the survey materials and responses in the supplemental material.

### 7.3 Results and Discussion

**Assessing H1.** All 12 participants stated that EMPHASISCHECKER is more useful than the baseline tool in authoring (Figure 9a), with 4 participants explicitly stating their excitement about the technology potentially being applied to their actual authoring routines in their free-form responses. Comparisons of the ratings of each of the two tools are also in line with their comparisons. The participants ranked the usefulness of the guides provided by EMPHASISCHECKER as 4.33 ( $\sigma = 0.85$ ), which is significantly higher than the baseline tool's 2.75 ( $\sigma = 1.09$ ) (two-sided Wilcoxon signed-rank test;  $W(11) = 5.5 < 10$ , the critical  $W$  value for  $\alpha = 0.05$ ). The participants also ranked the messaging capability of captions written with EMPHASISCHECKER as 4.17 ( $\sigma = 0.69$ ), also significantly higher than the baseline tool's 3.18 ( $\sigma = 0.55$ ) (two-sided Wilcoxon signed-rank test;  $W(8) = 0.0 < 3$ , the critical  $W$  value for  $\alpha = 0.05$ ).

When we asked the participants about the rationale for their assessment of the usefulness of EMPHASISCHECKER, 6 of the 12 participants explicitly mentioned that the check for the alignment between the chart and text helped with their authoring process. For example, P7

wrote, "(translated) [EMPHASISCHECKER's checking feature] gave me a chance to reflect on whether other people would agree on what I described as important in the text."

An additional six participants specifically pointed to the prominent feature detection as a component that contributes to the usefulness of EMPHASISCHECKER. P8 went further to suggest that the tool could be useful for exploratory data analysis while also precautioning that the features being shown up front in these cases could bias analysis.

All captions we collected through the study described prominent chart features. We hypothesize that this is because all the participants we had in the pool were not only experienced with authoring charts and captions but also were reminded of the properties of effective chart-caption pairs through a pre-study exercise.

**Assessing H2.** While five participants stated that authoring charts and text with EMPHASISCHECKER was easier, the majority (7 of 12) of the participants were neutral on the comparison of the ease of use of the two tools (Figure 9b). This result is similar to our findings on the SUS scores of each of the two tools; the average SUS score of our tool was 88.33 ( $\sigma = 8.06$ ) was on par with that of the baseline tool at 82.08 ( $\sigma = 11.22$ ) (two-sided paired t-test:  $t(11) = 1.96, p = 0.08 > 0.05$ ).

While it is counterintuitive that EMPHASISCHECKER with more features was deemed easier to use than the simpler baseline tool by some participants, diving deeper into the participants' comments provides an insight into why this was the case. Looking first at the comments on the ease of use from the seven participants who were neutral, six of them commented that while EMPHASISCHECKER includes additional features, the two tools were both very intuitive and easy to use. The participants who rated EMPHASISCHECKER as easier to use pointed out that the synergistic effect between the basic authoring interface and the added features reduces the mental burden on the users. For instance, P4 wrote, "While rescaling the graph and seeing prominent points, it is easier to focus on the main characteristics of the graph," pointing to the synergistic effect between the chart-editing interface and the prominent feature display. P10, who was neutral about the comparison, stated, "(translated) Physically, Tool A [baseline] was easier because it was simpler but Tool B [EMPHASISCHECKER] required less effort in authoring charts and was psychologically easier to use."

**Participants' reactions to errors in text references.** While writing captions with EMPHASISCHECKER, 10 of the 12 participants experienced errors. All of the participants immediately noticed errors. We



note that this is different from how *readers* often fail to see errors in reference extraction methods [36]. We hypothesize that this is because the authors of our tool have a clear intention and expectation of the correct results in mind when they run it, whereas readers do not know the expected results beforehand. After noticing the error, three of the participants attempted a revision of the text so that their text would be detected correctly, whereas the others read over the text and continued. In summary, the results suggest that while authors are unlikely to be misled by the system's errors, the errors can lead authors to spend unnecessary efforts in trying to satisfy the system.

However, the prevalence of the errors in the tool occasionally resulted in the failure of participants to realize their error even when EMPHASISCHECKER behaved correctly. For example, P12 typed in the wrong year (2018 instead of 2008), which was out of the time range shown in the chart (up to 2015), and EMPHASISCHECKER correctly identified no match between the text and chart. However, the participant thought that it was an error of the system for not detecting the time range in the chart and moved on. P6, on the other hand, made a typo by typing in the year 1987 instead of 1997, causing EMPHASISCHECKER to detect a time range different from what the participant intended (Figure 8b Sentence 4). As with the other participant, the participant deemed it an error of our tool and continued on. We believe that these errors are partially due to the authors trusting themselves more than EMPHASISCHECKER and that such issues will diminish as EMPHASISCHECKER becomes more accurate.

## 8 LIMITATIONS AND FUTURE WORK

**Providing additional guidance in authoring charts and captions.** Our tool currently provides rudimentary guidance in authoring charts and captions with matched emphasis but can be improved to provide further guidance to support authors. As P3 and P8 suggested, the tool can be extended to provide more information about the detected features (e.g., if the feature is rising/falling, the slope). Future work could also apply natural language generation techniques and suggest automatically generated captions for selected prominent features to further help with the authoring process. Alternatively, future work could provide further explanations and revision suggestions on detected emphasis mismatches or factual errors to help authors make more informed decisions. Furthermore, during the pre-study activity of choosing the most effective caption and citing rationale, four of the participants mentioned that useful integration of external information not available in the chart is a property of effective captions, which is in line with the findings of Kim et al. [37]. Based on these observations, we believe that suggestion of suitable external information in the process could also provide useful additional guidance. Another potential future direction would be to extend the tool to assist in writing chart descriptions for accessibility. As blind and sighted individuals perceive the usefulness of chart descriptions differently [47], properties of effective text descriptions for blind individuals should be incorporated into the extended tool design.

**Avoiding introducing bias to chart-caption pairs.** While all participants agreed on the usefulness of the tool, some participants suggested that guidance needs to be provided to the authors with careful consideration. P8's concern about the potential biasing of the chart and caption was shared by two other participants. They mentioned that being shown the prominent features too early while exploring the data could skew what messages they decide to include in the caption. From these comments, we believe that the guidance provided by the tool could be personalized and adjusted based on the state of the writing process, avoiding too much guidance during the exploratory stages. Moreover, focusing only on the match between chart and caption emphasis can lead to biased messages or violations of principles of good visualization design, such as the banking to 45 degrees principle [14]. Future developments of the tool can include detectors for biased representation of data and violations of good visualization design to help authors stay faithful to the data within the boundaries of good visualization design while matching chart and caption emphasis.

**Generalizing tool to other chart features & chart types.** Our tool is designed to work with time-series line charts, with a focus on univer-

sally present features: local extrema and trends. Yet, depending on the domain and context of the data, time-series charts can include other types of features, such as seasonal or cyclic patterns in monthly sales of AC units, L- or V-shapes in curves showing economic recovery, or double bottoms in stock charts ('W'-shaped features signaling potential future increases in stock price). We believe that future work can expand EMPHASISCHECKER by implementing such features for authors in specific domains and contexts. Furthermore, future work can go beyond univariate time-series line charts and cover other chart types, starting with the more common charts [2] such as multi-line charts, bar charts, and scatter plots. However, we expect that the generalization of EMPHASISCHECKER to other chart features and chart types will require a deeper understanding of the features not currently covered in this work. EMPHASISCHECKER is based on being able to compare the prominence of the chart features, and generalization of the tool will require similar levels of understanding of the chart features not covered in our work. Moreover, the design of the tool is grounded on the assumption that there is a one-to-one mapping between the x-axis and the data points. While the current design would hold for chart types that follow this assumption (e.g., horizontal bar charts), the design of the tool may need modifications for other chart types to help users recognize features while avoiding visual clutter.

**Incorporating information outside sentence-chart pairs.** Our text references extractor operates at the sentence level and utilizes information available strictly within the sentence and the chart. The text references extractor currently does not utilize the information available in prior sentences. For instance, our tool fails to detect any references in Figure 8 (d) Sentence 3 because the reference to time ('May 2012') appears in the previous sentence. Anaphora resolution methods [52] would be a starting point for future work. In addition, while our use of the Named Entity Module in the Stanford CoreNLP toolkit [9, 22, 48] allows our text references extractor to cover a variety of time expressions (e.g., year, date, time, seasons, quarters, duration, etc.), it is unable to comprehend time expressions requiring external knowledge. For example, were it not for the explicit mention of the year '2020' in Figure 8 (c) Sentence 3, the tool would not have been able to comprehend the time range described as 'the period of Covid-19.' Allowing the text references extractor to draw information from external knowledge bases (e.g., WolframAlpha [78]) or LLM-based tools (e.g., GPT [7]) could allow it to detect more descriptions of time reliably.

**Computing visual prominence in rendered charts.** Our tool, like that of Hullman et al. [31], utilizes only the properties of the underlying time-series data to approximate visual prominence; the tool does not consider the rendering of the chart. But, chart authors often use specific encodings (e.g., color, size) and annotations that guide the readers' attention to specific chart features. Such encodings and annotations can affect the visual prominence of features. Future work could analyze the visual properties of the rendered chart as well as the statistical properties of the underlying data to develop additional measures of visual prominence.

## 9 CONCLUSION

Writing caption text whose emphasis matches chart emphasis is important for communicating the intended message to the reader. Through a survey of chart-caption pairs in the real world, we find that there is often a mismatch between chart emphasis and caption emphasis. To address this issue, we introduce EMPHASISCHECKER, a tool to help authors align the message in the caption text with the prominent features in the chart. The tool comprises the *time-series prominent feature detector* that identifies chart emphasis and performs at the level of prior crowdsourcing algorithm and the *text reference extractor* that extracts text emphasis and matches them with chart emphasis. Based on feedback we collected from users, EMPHASISCHECKER is both useful and easy to use for writing effective captions. We identify future directions in this space for further supporting chart authors with better caption-writing tools. Our code is available at: <https://github.com/dhkim16/EmphasisChecker-release>.

## ACKNOWLEDGMENTS

This work is supported by the Brown Institute for Media Innovation.

## REFERENCES

- [1] S. K. Badam, Z. Liu, and N. Elmqvist. Elastic Documents: Coupling Text and Tables through Contextual Visualizations for Enhanced Document Reading. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):661–671, 2019. doi: 10.1109/TVCG.2018.2865119 2
- [2] L. Battle, P. Duan, Z. Miranda, D. Mukusheva, R. Chang, and M. Stonebraker. Beagle: Automated Extraction and Interpretation of Visualizations from the Web. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–8. ACM, New York, USA, 2018. doi: 10.1145/3173574.3174168 2, 9
- [3] F. Beck and D. Weiskopf. Word-Sized Graphics for Scientific Texts. *IEEE Transactions on Visualization and Computer Graphics*, 23(6):1576–1587, 2017. doi: 10.1109/TVCG.2017.2674958 2
- [4] A. D. Biderman. The Graph as a Victim of Adverse Discrimination and Segregation. *Information Design Journal*, 1(4):232–241, 1979. doi: 10.1075/idj.1.4.03bid 2
- [5] M. A. Borkin, Z. Bylinskii, N. W. Kim, C. M. Bainbridge, C. S. Yeh, D. Borkin, H. Pfister, and A. Oliiva. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):519–528, 2016. doi: 10.1109/TVCG.2015.2467732 2
- [6] J. Brooke. SUS: A ‘Quick and Dirty’ Usability Scale. In *Usability Evaluation In Industry*, chap. 21, pp. 189–194. Taylor & Francis Group, London, UK, 1996. doi: 10.1201/9781498710411-35 7
- [7] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. Language Models Are Few-Shot Learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*. 159. Curran Associates Inc., Red Hook, USA, 2020. doi: 10.48550/arXiv.2005.14165 9
- [8] S. Carberry, S. Elzer, and S. Demir. Information Graphics: An Untapped Resource for Digital Libraries. In *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 581–588. ACM, New York, USA, 2006. doi: 10.1145/1148170.1148270 2
- [9] A. X. Chang and C. Manning. SUTime: A Library for Recognizing and Normalizing Time Expressions. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*, pp. 3735–3740. ELRA, Paris, France, 2012. 2, 5, 9
- [10] C. Chen, R. Zhang, S. Kim, S. Cohen, T. Yu, R. Rossi, and R. Bunescu. Neural Caption Generation over Figures. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, pp. 482–485. ACM, New York, USA, 2019. doi: 10.1145/3341162.3345601 2
- [11] C. Chen, R. Zhang, E. Koh, S. Kim, S. Cohen, T. Yu, R. Rossi, and R. Bunescu. Figure Captioning with Reasoning and Sequence-Level Training. *ArXiv*, abs/1906.02850, 2019. doi: 10.48550/arXiv.1906.02850 2
- [12] D. Chen and C. Manning. A Fast and Accurate Dependency Parser Using Neural Networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 740–750. ACL, Stroudsburg, USA, 2014. doi: 10.3115/v1/D14-1082 2, 6
- [13] Z. Chen and H. Xia. CrossData: Leveraging Text-Data Connections for Authoring Data Documents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 95. ACM, New York, USA, 2022. doi: 10.1145/3491102.3517485 2
- [14] W. S. Cleveland, M. E. McGill, and R. McGill. The Shape Parameter of a Two-Variable Graph. *Journal of the American Statistical Association*, 83(402):289–300, 1988. doi: 10.1080/01621459.1988.10478598 9
- [15] Z. Cui, S. K. Badam, M. A. Yalcin, and N. Elmqvist. DataSite: Proactive Visual Data Exploration with Computation of Insight-Based Recommendations. *Information Visualization*, 18(2):251–267, 2019. doi: 10.1177/1473871618806555 2
- [16] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. ACL, Stroudsburg, USA, 2019. doi: 10.18653/v1/N19-1423 2, 6
- [17] D. H. Douglas and T. K. Peucker. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, 1973. doi: 10.3138/FM57-6770-U75U-7727 5
- [18] L. Earnest. The First Three Spelling Checkers. Retrieved July 18, 2023 from <https://web.archive.org/web/20121022091418/http://www.stanford.edu/~learnest/spelling.pdf>, 2011. 2
- [19] S. Elzer, S. Carberry, D. Chester, S. Demir, N. Green, I. Zukerman, and K. Trnka. Exploring and Exploiting the Limited Utility of Captions in Recognizing Intention in Information Graphics. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 223–230. ACL, Stroudsburg, USA, 2005. doi: 10.3115/1219840.1219868 2
- [20] S. Elzer, S. Carberry, and I. Zukerman. The Automated Understanding of Simple Bar Charts. *Artificial Intelligence*, 175(2):526–555, 2011. doi: 10.1016/j.artint.2010.10.003 2
- [21] M. Fasciano and G. Lapalme. PostGraphe: A System for the Generation of Statistical Graphics and Text. In *Eighth International Natural Language Generation Workshop*. ACL, Stroudsburg, USA, 1996. 2
- [22] J. R. Finkel, T. Grenager, and C. Manning. Incorporating Non-Local Information into Information Extraction Systems by Gibbs Sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pp. 363–370. ACL, Stroudsburg, USA, 2005. doi: 10.3115/1219840.1219885 2, 5, 9
- [23] J. Fulda, M. Brehmer, and T. Munzner. TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):300–309, 2016. doi: 10.1109/TVCG.2015.2467531 2
- [24] P. Goffin, W. Willett, J.-D. Fekete, and P. Isenberg. Design Considerations for Enhancing Word-Scale Visualizations with Interaction. In *Posters of the Conference on Information Visualization (InfoVis)*. IEEE, New York, USA, 2015. 2
- [25] Google. Check Your Spelling & Grammar in Google Docs. Retrieved July 18, 2023 from <https://support.google.com/docs/answer/57859?hl=en>, 2023. 1, 2
- [26] Google. Fix Spelling & Grammar as You Type in Gmail. Retrieved July 18, 2023 from <https://support.google.com/mail/answer/7987?hl=en>, 2023. 2
- [27] Google. Use Smart Compose. Retrieved July 18, 2023 from <https://support.google.com/mail/answer/9116836?hl=en>, 2023. 2
- [28] Google Meet. Retrieved July 18, 2023 from <https://meet.google.com/>, 2023. 7
- [29] J. D. Gould. Looking at Pictures. In R. A. Monty and J. W. Senders, eds., *Eye Movements and Psychological Processes*, chap. 5.2, pp. 323–345. Lawrence Erlbaum Associates, Inc., Hillsdale, USA, 1976. 2
- [30] Grammarly. Retrieved July 18, 2023 from <https://www.grammarly.com>, 2023. 1, 2
- [31] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: Automatic Generation of Annotated Stock Visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2707–2716. ACM, New York, USA, 2013. doi: 10.1145/2470654.2481374 2, 7, 9
- [32] International Monetary Fund. Retrieved July 18, 2023 from <https://www.imf.org/en/Home>, 2023. 3
- [33] U. Jügel, Z. Jerzak, G. Hackenbroich, and V. Markl. M4: A Visualization-Oriented Time Series Data Aggregation. *Proceedings of the VLDB Endowment*, 7(10):797–808, 2014. doi: 10.14778/2732951.2732953 5
- [34] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases. *Knowledge and Information Systems*, 3(3):263–286, 2001. doi: 10.1007/PL00011669 5
- [35] A. Khan, S. Breslav, M. Glueck, and K. Hornbæk. Benefits of Visualization in the Mammography Problem. *International Journal of Human-Computer Studies*, 83:94–113, 2015. doi: 10.1016/j.ijhcs.2015.07.001 2
- [36] D. H. Kim, E. Hoque, J. Kim, and M. Agrawala. Facilitating Annual Reading by Linking Text and Tables. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 423–434. ACM, New York, USA, 2018. doi: 10.1145/3242587.3242617 2, 9
- [37] D. H. Kim, V. Setlur, and M. Agrawala. 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*. 610. ACM, New York, USA, 2021. doi: [10.1145/3411764.3445443](https://doi.org/10.1145/3411764.3445443) 1, 2, 3, 7, 9
- [38] H.-K. Kong, Z. Liu, and K. Karahalios. Frames and Slants in Titles of Visualizations on Controversial Topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12. ACM, New York, USA, 2018. doi: [10.1145/3173574.3174012](https://doi.org/10.1145/3173574.3174012) 2
- [39] H.-K. Kong, Z. Liu, and K. Karahalios. 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*, pp. 1–13. ACM, New York, USA, 2019. doi: [10.1145/3290605.3300576](https://doi.org/10.1145/3290605.3300576) 1, 2
- [40] N. Kong, M. A. Hearst, and M. Agrawala. Extracting References between Text and Charts via Crowdsourcing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 31–40. ACM, New York, USA, 2014. doi: [10.1145/2556288.2557241](https://doi.org/10.1145/2556288.2557241) 2
- [41] C. Lai, Z. Lin, R. Jiang, Y. Han, C. Liu, and X. Yuan. Automatic Annotation Synchronizing with Textual Description for Visualization. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, New York, USA, 2020. doi: [10.1145/3313831.3376443](https://doi.org/10.1145/3313831.3376443) 2
- [42] LanguageTool. Retrieved July 18, 2023 from <https://languagetool.org/>, 2023. 1, 2
- [43] S. Latif, D. Liu, and F. Beck. Exploring Interactive Linking between Text and Visualization. In *Proceedings of the 2018 Eurographics/IEEE VGTC Conference on Visualization: Short Papers*, pp. 91–94. The EG Association, Eindhoven, Netherlands, 2018. doi: [10.2312/eurovisshort.20181084](https://doi.org/10.2312/eurovisshort.20181084) 2
- [44] S. Latif, Z. Zhou, Y. Kim, F. Beck, and N. W. Kim. Kori: Interactive Synthesis of Text and Charts in Data Documents. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):184–194, 2022. doi: [10.1109/TVCG.2021.3114802](https://doi.org/10.1109/TVCG.2021.3114802) 2
- [45] LibreOffice. Checking Spelling and Grammar. Retrieved July 18, 2023 from [https://help.libreoffice.org/6.2/en-US/text/swriter/guide/spellcheck\\_dialog.html](https://help.libreoffice.org/6.2/en-US/text/swriter/guide/spellcheck_dialog.html), 2023. 2
- [46] A. Y. Lin, J. Ford, E. Adar, and B. Hecht. VizByWiki: Mining Data Visualizations from the Web to Enrich News Articles. In *Proceedings of the 2018 World Wide Web Conference*, pp. 873–882. IW3C2, Geneva, Switzerland, 2018. doi: [10.1145/3178876.3186135](https://doi.org/10.1145/3178876.3186135) 2
- [47] A. Lundgard and A. Satyanarayan. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1073–1083, 2022. doi: [10.1109/TVCG.2021.3114770](https://doi.org/10.1109/TVCG.2021.3114770) 2, 3, 9
- [48] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, and D. McClosky. The Stanford CoreNLP Natural Language Processing Toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55–60. ACL, Stroudsburg, USA, 2014. doi: [10.3115/v1/P14-5010](https://doi.org/10.3115/v1/P14-5010) 2, 5, 6, 9
- [49] Merriam-Webster. Retrieved July 18, 2023 from <https://www.merriam-webster.com>, 2023. 6
- [50] L. Micallef, P. Dragicevic, and J.-D. Fekete. Assessing the Effect of Visualizations on Bayesian Reasoning through Crowdsourcing. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2536–2545, 2012. doi: [10.1109/TVCG.2012.199](https://doi.org/10.1109/TVCG.2012.199) 2
- [51] Microsoft. Editor Settings in Outlook.com and Outlook on the Web. Retrieved July 18, 2023 from <https://support.microsoft.com/en-us/office/editor-settings-in-outlook-com-and-outlook-on-the-web-c6b1283d-81a2-47f1-bc85-9e8dfc0cbf15>, 2023. 2
- [52] R. Mitkov. *Anaphora Resolution*. Taylor & Francis Group, London, UK, 2014. doi: [10.4324/9781315840086](https://doi.org/10.4324/9781315840086) 9
- [53] J. Obeid and E. Hoque. Chart-to-Text: Generating Natural Language Descriptions for Charts by Adapting the Transformer Model. In *Proceedings of the 13th International Conference on Natural Language Generation*, pp. 138–147. ACL, Stroudsburg, USA, 2020. 2
- [54] A. Ottley, A. Kaszowska, R. J. Crouser, and E. M. Peck. The Curious Case of Combining Text and Visualization. In *Proceedings of the 2019 Eurographics/IEEE VGTC Conference on Visualization: Short Papers*. The EG Association, Eindhoven, Netherlands, 2019. doi: [10.2312/evs.20191181](https://doi.org/10.2312/evs.20191181) 2
- [55] A. Ottley, B. Metevier, P. Han, and R. Chang. Visually Communicating Bayesian Statistics to Laypersons. In *Technical Report 2012-02*. Tufts University, Medford, USA, 2012. 2
- [56] A. Ottley, E. M. Peck, L. T. Harrison, D. Afegan, C. Ziemkiewicz, H. A. Taylor, P. K. J. Han, and R. Chang. Improving Bayesian Reasoning: The Effects of Phrasing, Visualization, and Spatial Ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2016. doi: [10.1109/TVCG.2015.2467758](https://doi.org/10.1109/TVCG.2015.2467758) 2
- [57] Pew Research Center. Retrieved June 18, 2023 from <https://www.pewresearch.org/>, 2023. 3
- [58] J. Pinheiro and J. Poco. ChartText: Linking Text with Charts in Documents. *ArXiv*, abs/2201.05043, 2022. doi: [10.48550/arXiv.2201.05043](https://doi.org/10.48550/arXiv.2201.05043) 2
- [59] J. Poco and J. Heer. Reverse-Engineering Visualizations: Recovering Visual Encodings from Chart Images. *Computer Graphics Forum*, 36(3):353–363, 2017. doi: [10.1111/cgf.13193](https://doi.org/10.1111/cgf.13193) 3
- [60] X. Qian, E. Koh, F. Du, S. Kim, J. Chan, R. A. Rossi, S. Malik, and T. Y. Lee. Generating Accurate Caption Units for Figure Captioning. In *Proceedings of the Web Conference 2021*, pp. 2792–2804. ACM, New York, USA, 2021. doi: [10.1145/3442381.3449923](https://doi.org/10.1145/3442381.3449923) 2
- [61] U. Ramer. An Iterative Procedure for the Polygonal Approximation of Plane Curves. *Computer Graphics and Image Processing*, 1(3):244–256, 1972. doi: [10.1016/S0146-664X\(72\)80017-0](https://doi.org/10.1016/S0146-664X(72)80017-0) 5
- [62] K. Rong and P. Bailis. ASAP: Prioritizing Attention via Time Series Smoothing. *Proceedings of the VLDB Endowment*, 10(11):1358–1369, 2017. doi: [10.14778/3137628.3137645](https://doi.org/10.14778/3137628.3137645) 5
- [63] P. Rosen and G. J. Quadri. LineSmooth: An Analytical Framework for Evaluating the Effectiveness of Smoothing Techniques on Line Charts. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1536–1546, 2021. doi: [10.1109/TVCG.2020.3030421](https://doi.org/10.1109/TVCG.2020.3030421) 5
- [64] Springer Nature. Retrieved July 18, 2023 from <https://www.nature.com/>, 2023. 3
- [65] C. Stokes, V. Setlur, B. Cogley, A. Satyanarayan, and M. A. Hearst. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):1233–1243, 2023. doi: [10.1109/TVCG.2022.3209383](https://doi.org/10.1109/TVCG.2022.3209383) 2
- [66] J. Sweller, P. Ayres, and S. Kalyuga. The Split-Attention Effect. In *Cognitive Load Theory*, chap. 9, pp. 111–128. Springer, New York, USA, 2011. doi: [10.1007/978-1-4419-8126-4\\_9](https://doi.org/10.1007/978-1-4419-8126-4_9) 2
- [67] Tableau Public. Retrieved July 18, 2023 from <https://public.tableau.com/app/discover>, 2023. 1, 3
- [68] Tableau Software. Retrieved July 18, 2023 from <https://www.tableau.com>, 2023. 1, 2, 3
- [69] The British Broadcasting Corporation (BBC). Retrieved July 18, 2023 from <https://www.bbc.com/>, 2023. 3
- [70] The New York Times. Retrieved July 18, 2023 from <https://www.nytimes.com/>, 2023. 3
- [71] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, USA, 2001. 2
- [72] United Nations. International Labour Organization. Retrieved July 18, 2023 from <https://www.ilo.org/global/lang-en/>, 2023. 3
- [73] U.S. Department of the Treasury. Retrieved July 18, 2023 from <https://home.treasury.gov/>, 2023. 3
- [74] Vox Media. Retrieved July 18, 2023 from <https://www.vox.com/>, 2023. 3
- [75] M. P. Whitacre and E. W. Saul. High School Girls' Interpretations of Science Graphs: Exploring Complex Visual and Natural Language Hybrid Text. *International Journal of Science and Mathematics Education*, 14(8):1387–1406, 2016. doi: [10.1007/s10763-015-9677-7](https://doi.org/10.1007/s10763-015-9677-7) 2
- [76] Wikimedia Commons. Retrieved July 18, 2023 from [https://commons.wikimedia.org/wiki/Main\\_Page](https://commons.wikimedia.org/wiki/Main_Page), 2023. 2
- [77] G. Wills and L. Wilkinson. AutoVis: Automatic Visualization. *Information Visualization*, 9(1):47–69, 2010. doi: [10.1057/ivs.2008.27](https://doi.org/10.1057/ivs.2008.27) 2
- [78] WolframAlpha. Retrieved July 18, 2023 from <https://www.wolframalpha.com/>, 2023. 9
- [79] C. Xiong, L. Van Weelden, and S. Franconeri. The Curse of Knowledge in Visual Data Communication. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):3051–3062, 2020. doi: [10.1109/TVCG.2019.2917689](https://doi.org/10.1109/TVCG.2019.2917689) 1, 2, 3
- [80] H. H. Zhu, S. S. Cheng, and E. Wu. How Do Captions Affect Visualization Reading? *ArXiv*, abs/2205.01263, 2022. doi: [10.48550/arXiv.2205.01263](https://doi.org/10.48550/arXiv.2205.01263) 1, 2
- [81] Çagatay Demiralp, P. J. Haas, S. Parthasarathy, and T. Pedapati. Foresight: Rapid Data Exploration Through Guideposts. *ArXiv*, abs/1709.10513, 2017. doi: [10.48550/arXiv.1709.10513](https://doi.org/10.48550/arXiv.1709.10513) 2