What *Is* the Difference Between a Mountain and a Molehill? Quantifying Semantic Labeling of Visual Features in Line Charts

Dennis Bromley*
Tableau Research, Seattle, WA, USA

Vidya Setlur†
Tableau Research, Palo Alto, CA, USA

2023 IEEE Visualization and Visual Analytics (VIS) | 979-8-3503-2557-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/VIS54172.2023.00041

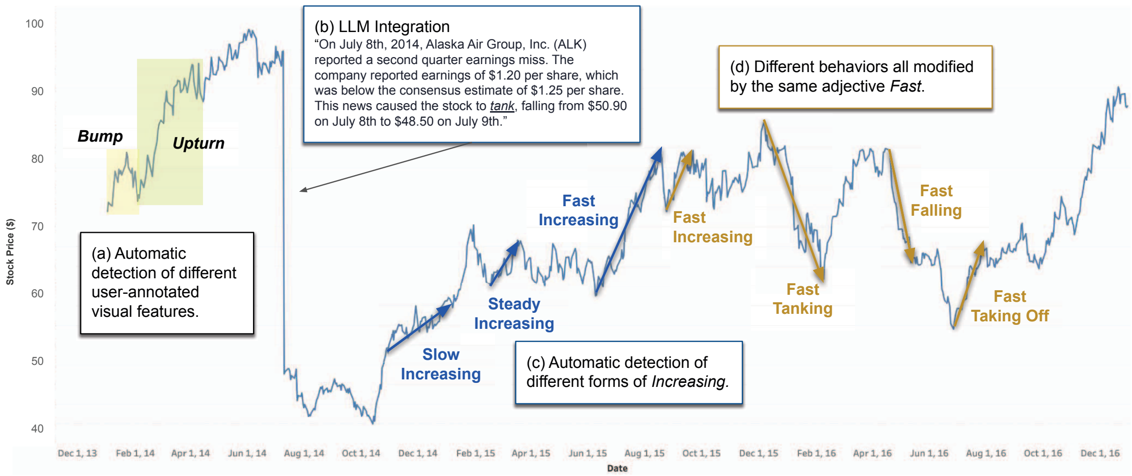

Figure 1: Automatic labeling of visual features in a line chart. (a) Automatic detection of user-annotated visual features. (b) Integration of discovered visual features with a large language model (LLM). (c, d) Automatic labeling of increases and decreases in trends using quantified semantics; the adjective/verb pairings encode specific empirically-derived line-slope information for generating corresponding annotations.

Abstract

Relevant language describing visual features in charts can be useful for authoring captions and summaries about the charts to help with readers' takeaways. To better understand the interplay between concepts that describe visual features and the semantic relationships among those concepts (e.g., *'sharp increase'* vs. *'gradual rise'*), we conducted a crowdsourced study to collect labels and visual feature pairs for univariate line charts. Using this crowdsourced dataset of labeled visual signatures, this paper proposes a novel method for labeling visual chart features based on combining feature-word distributions with the visual features and the data domain of the charts. These feature-word-topic models identify word associations with similar yet subtle differences in semantics, such as *'flat,' 'plateau,'* and *'stagnant,'* and descriptors of the visual features, such as *'sharp increase,' 'slow climb,'* and *'peak.'* Our feature-word-topic model is computed using both a quantified semantics approach and a signal processing-inspired least-errors shape-similarity approach. We finally demonstrate the application of this dataset for annotating charts and generating textual data summaries.

Keywords: Semantics, trends, annotation, text generation.

Index Terms: Human-centered computing—Visualization—Visualization systems and tools;

*e-mail: dbromley@tableau.com
†e-mail: vsetlur@tableau.com

1 Introduction

Data visualizations are often accompanied by captions and summaries describing their key takeaways to the reader [39, 47]. Studies have also found that captions help readers with takeaways by emphasizing visually prominent features in charts [25]. There are well-established visualization techniques to draw a reader's attention to the visually prominent features of a chart using techniques such as highlighting and annotating a portion of the chart or changing the scale or data extents to show prominent visual patterns [5].

However, the language for describing and emphasizing visual features in the captions is a less-studied topic. Automatic chart captioning and summarization tools [3, 6, 9, 17] support authors with caption generation that describes visual features in charts; the language employed in the recommended captions and annotations tends to be simple, ranging from describing the domain, axes, and encodings to specific statistical information (e.g., extrema) describing specific marks in the chart [25]. Little research has explored the nuances in language to add emphasis to the characteristics of the data encoded in a chart. *Hedging* is a communicative strategy used in language for increasing or reducing the force of statements to emphasize and bring a reader's attention to specific portions of the text that is important to the overall takeaway of the intended message [21]. Prior work in linguistics research indicates the benefits of employing hedge words to provide additional texture and emphasis to textual discourse for the reader [32].

The goal of this work is to explore language that emphasizes visual features in charts and the semantic relationships that expose the nuances among those language concepts. We specifically focus on how people label and associate hedge words such as *'sharp'* or *'gradual'* when describing visual features in univariate line charts. To better understand the correspondence between language and

DOI 10.1109/VIS54172.2023.00041

visual features, we crowdsource a dataset of labeled visual chart features based on combining feature-word distributions with the visual features and the data domain of the charts. These feature-word-topic models identify word associations with similar yet subtle differences in semantics, such as *'flat,' 'plateau,'* and *'stagnant,'* and descriptors of the visual features, such as *'sharp increase,' 'slow climb,'* and *'peak,'* for example. Our feature-word-topic model is computed using both a quantified semantics approach and a signal processing-inspired multi-resolution approach wherein windowed versions of crowdsource-labeled chart segments are applied to unlabeled charts to find regions of low-mean absolute error (MAE) shape similarity. We demonstrate the utility of this dataset for automatically annotating line charts and for generating data summaries using large language models (LLMs), as shown in Figure 1. We also describe future research directions for this work, such as incorporating domain-specific descriptions, leveraging LLMs for semantic enrichment, and supporting search and natural language (NL) interaction.

2 Related Work

This work builds on prior work that explores chart annotation techniques and linguistic approaches for generating data narratives.

2.1 Text Annotation for Charts

There is a growing body of research that focuses on the role and importance of text in visual analysis [38, 44]. Kong et al. [27, 28] evaluated how titles can impact the perceived message of a chart and found that people were more likely to recall information conveyed by slanted framings (e.g., emphasizing only part of a chart's message) than the actual chart's visuals. Kim et al. found that when both the chart and text described a high-prominence feature, readers treated the doubly emphasized high-prominence feature as the takeaway [25]. When the text described a low-prominence chart feature, readers relied mostly on the chart alone and usually reported a higher-prominence feature as the takeaway. Hearst & Tory examined participant preferences for text with visualizations in the context of chatbot interaction [15]. Their study found that when participants preferred to see charts, they also preferred to view additional contextual data to be provided in the chart.

Prior research has also explored how authors add annotations and descriptions to charts guiding a reader's attention to visual features in the chart, explaining what the underlying data means, and providing additional context [18, 26, 42]. Kong and Agrawala developed techniques for analyzing charts to recover visually salient features of the data-encoding marks (e.g., min, max, mean values). Users can interactively add graphical and text annotations to facilitate chart reading [29, 30]. Kandogan [23] introduced just-in-time descriptive analytics by employing statistics to automatically generate annotations for clusters and outliers. Contextifier [19] uses news headlines to provide external contextual annotations for line charts. They consider linguistic relevance, the number of article views, and the visual saliency of chart peaks to identify the headlines and chart features to annotate. Henkin & Turkay [16] have done extensive work quantifying crowdsourced semantics for scatter plots.

Our work contributes to this body of work by exploring how hedge words can further express and describe visual features in line charts. We also quantify the semantics to identify language subtleties to automatically label and generate text summaries for describing the magnitude of the slopes and characteristics of the features.

2.2 Linguistic Approaches for Generating Narratives

The computational linguistics community has implemented techniques for identifying hedging patterns in text and conversational transcripts to determine their effectiveness in debating or communicating a point of view to the reader [14, 22]. Other work has focused on creating datasets containing hedge cues, curated from open-access text, that are fed into a multitask learning model for text classification and generation [13]. However, none of these linguistic approaches have explored hedging and its associated semantics for specifically describing visual features in charts.

Visual analytics systems incorporate generated text with visualization responses to help communicate key insights to the user [7, 24, 37, 40, 43]. Other tools produce text summaries with statistical descriptions shown in the visualizations [2, 3]. Data storytelling incorporates textual narratives with visuals, communicating insights that are more memorable, persuasive, and engaging than statistics alone [31, 35, 42]. Systems like Kori [33, 34] and VizFlow [45] provide explicit linking strategies between text and charts to support design patterns for data storytelling, narrative sequencing, and rhetoric [18, 20]. In this paper, we further explore the interplay between text and charts by the automatic labeling of visual features in charts and text generation containing hedge word descriptors using a crowdsourced dataset of labeled visual signatures.

3 Crowdsourcing Labeled Visual Features

The motivation for crowdsourcing a labeled dataset of terms and visual features is two-fold: 1) capture semantic descriptions of different visual features in univariate line charts and 2) elucidate and quantify the relationships among those semantic descriptions.

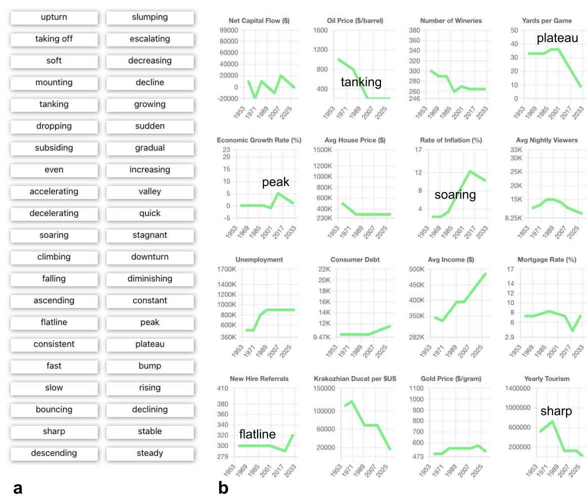

Figure 2: The annotation-collection tool. Participants drag words from the left (a) over to visual features of the charts on the right (b). The words are snapped to the nearest chart position. Words may be moved or deleted once they are attached to a chart. Individual words may be used on multiple charts and multiple times on a single chart.

We collected labeled annotations for visual features in univariate line charts by implementing a data collection tool, shown in Figure 2. The tool was implemented as a Typescript frontend and a Django backend attached to a PostgreSQL database.

The interface has two parts: the left side (Figure 2a) comprises 42 word labels consisting of: 1) words related to the basic shape descriptors, *'up,' 'down,'* and *'flat,'* 2) adjectives that describe such shapes (e.g., *'slow,' 'sudden,'*) and 3) words that describe the emergent shapes created by such regions (e.g., *'plateau'* or *'valley'*). To find these word descriptors, we leveraged the hierarchy of hypernyms and hyponyms from Wordnet [11], whose depth typically ranges up or down to two hierarchical levels (e.g., '*up*' → ['*increasing*,' '*ascending*']), as well as word2vec [36] to identify related concepts, such as *'sharp'* and *'increasing.'* In total, the list contained 8 nouns, 13 adjectives, and 21 verbs. Note that while this list is not exhaustive, we considered the set of words as a starting point for collecting nuanced language that describes common features found in line charts. The words were displayed in a randomized order in the interface to avoid positional bias.

The right side of the tool interface (Figure 2b) displayed 16 line charts shown in random order to each participant to mitigate any positional bias. The same charts were shown to all participants. The charts were generated in Chart.JS [1], showing years on the x-axis, ranging from 1960 to 2030. The title and its corresponding y-axis range were randomly assigned from one of the following topics: Average Income ($), Unemployment, Yards per Game, New Hire Referrals, Yearly Tourism, Rate of Inflation (%), Average House Price ($), Krakozhian Ducats per $US, Average Nightly Viewers, Economic Growth Rate (%), Gold Price ($/gram), Oil Price ($/barrel), Consumer Debt, Number of Wineries, Mortgage Rate (%), and Net Capital Flow ($). Each chart is a line graph constructed by connecting seven sequential line segments end to end. Similar to the chart stimuli generated in [25], each segment is randomly assigned one of nine different slopes: *Up*, *Down*, *Flat* with slopes [1, -1, 0], *SteepUp*, *SteepDown*, *SteepFlat* with slopes [3, -3, 0] *GentleUp*, *GentleDown*, *GentleFlat* with slopes [0.5, -0.5, 0].

We recruited 67 participants through a mailing list at a data analytics software company. Participants were required to pass a chart literacy test before proceeding to the annotation labeling exercise. Participants annotated the charts by dragging words from the left (Figure 2a) onto the charts on the right (Figure 2b) in the interface. Multiple words could be dragged to the same feature in a chart. We recorded the chart identifier, the annotation, the position along the line graph where the annotation occurred, the date the annotation occurred, and a unique anonymous participant identifier. The study details and instructions are found in the supplementary material.

4 Analysis of the Dataset

4.1 Analysis Technique

We calculate term co-occurrence and perform annotation clustering to identify quantifiable relationships among the different annotation terms. Annotation co-occurrence helps us understand how often different annotation terms are used to label the same visual feature; for each annotation, the co-occurrence of every other word is calculated as the average of per-segment % representation. For example, consider two segments that contain the annotation *'quick.'* If the term *'fast'* represents 50% of the annotations on the first segment and 30% of the annotations on the second segment, then the overall co-occurrence of *'fast'* with respect to *'quick'* is $\frac{50\%+30\%}{2} = 40\%$. Note that co-occurrence is not symmetric as *'quick'* may co-occur with different annotations than *'fast.'*

Annotations are clustered using hierarchical clustering and Ward's linkage [46] calculated with Euclidean distance; these approaches tend to identify dense clusters while making a minimum number of assumptions about cluster size, shape, and count. Position matrix entries are assigned by segment co-occurrence. For example, if *'quick'* and *'fast'* co-occurred 10 times, then each would have the position 10 on the other's axis. The matrix is then scaled so all values are in $[0,1]$, and values of 1.0 are placed along the diagonal.

4.2 Findings

A total of 67 participants generated $1,892$ annotations, with an average of 28.2 annotations per user, and 118.3 annotations per chart. 7 segments and 16 charts provided a total of 112 different segments. On average, there were 17 annotations per segment, allowing us to empirically derive various inter-word relationships (Figure 3). Term co-occurrence analysis quantifies which words are typically present together. Agglomerative hierarchical clustering of term co-occurrence results in distinct groups, suggesting a high degree of semantic agreement among participants. The crowdsourced dataset and analysis are provided in the supplementary material.

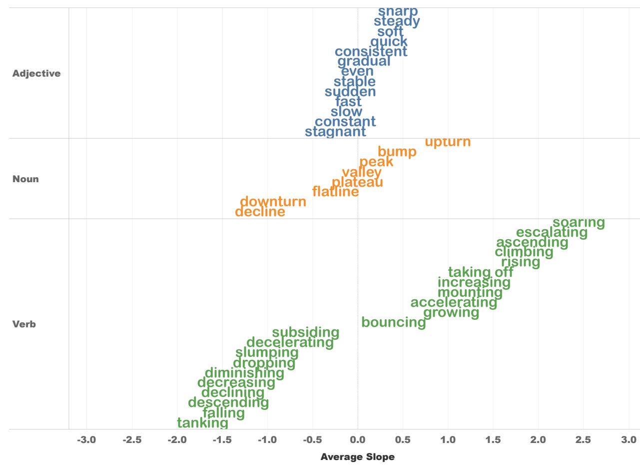

Figure 3: Average segment slope of the annotations. The maximum possible slope range available for the charts is -3 to +3.

One of the goals of this work is to understand the hierarchical semantics of the visual feature/annotation pairs. Using line slope as a fundamental component of signal shape, we analyzed the average slope associated with each annotation. As shown in Figure 3, slope analysis distributed the various annotation words across a broad continuum from steeply descending to steeply ascending. Not only does this suggest an empirically derived semantic hierarchy e.g., *'soaring'* is a steeper increase than *'taking off'* while *'tanking'* is a steeper dropoff than *'slumping,'* the quantification of that hierarchy allows us to make concrete NL recommendations when generating labels for previously unlabeled signals (Figures 1c and 1d).

5 Applications

We demonstrate two applications for this crowdsourced dataset of labeled visual signatures.

5.1 Automatic Labeling of Visual Features in Line Charts

Using the dataset, we develop two techniques for automatic labeling of visual features: *shape identification*, which is useful for discovering concrete shapes such as *'peak'* or *'valley'*, and *slope identification*, which is useful for describing how univariate data changes along the y-axis. In this discussion, we borrow from the signal processing lexicon and refer to a univariate data set as a *signal* and the small annotated source signal whose shape were are looking for in a larger unlabeled signal as the *kernel*.

Shape identification tries to find an annotated visual feature in a larger unlabeled signal. Figure 1a shows the detection of a *'bump'* and an *'upturn.'* This shape identification approach is particularly applicable to finding visual features that are constructed from multiple segments, e.g., a *'peak'* consists of a rising segment followed by a falling segment. The following algorithm describes the process of identifying a kernel signal's shape within a larger unlabeled signal.

5.1.1 Shape Discovery Algorithm

1. Begin with an unlabeled signal in which we would like to find a visual feature.
2. Collect all 112 (16 line plots * 7 segments/plot) annotatable segments and the annotations associated with them.
3. For each segment, build a five-segment *kernel signal* consisting of the annotated segment and two segments on each side of that annotated segment. Note that kernels near the edge may consist of fewer than five segments. For each such kernel signal, create

shallow and deep variants of it where the normalized variant heights range from $[0.1, 1.0]$ in units of 0.1.

4. For each variant, perform normalization, smoothing, and take the first derivative. We employ standard Savitzky-Golay smoothing [41] with a smoothing factor proportional to the kernel size for efficient smoothing and derivation.
5. Similarly, normalize the unlabeled signal and apply Savitzky-Golay smoothing and derivation.
6. Calculate a windowed-mean-absolute-error (MAE) by sliding the kernel past the unlabeled signal, much like convolution. The window size is parameterizable to allow the algorithm to search for visual features of different sizes.
7. Accrue these errors for every variant of every kernel.
8. For every kernel, calculate MAE z-scores.
9. Filter MAE scores using two criteria: max acceptable MAE score and z-score. Keep points below either threshold.
10. Mark points that meet the criteria threshold. The presence of points indicates that a visual feature is found in the chart.
11. Merge neighboring (<= 2 points) qualifying points into larger annotated regions.

In addition to the least-errors shape identification approach taken above, the quantified slope semantics shown in Figure 3 provides us with an additional tool for visual feature identification. Specifically, the quantified slope semantics helps identify specific relationships among line slope, hedge words, and the hedge word's semantic modifiers. For example, Figure 3 shows us a rough hierarchy of single-word slope descriptions from which we might decide to label a line as *'soaring'* rather than *'growing.'* However, if we look at verb annotations *and their adjective modifiers* as a single unit that encodes line-slope information, we become much more precise; for example, *'taking off'* in the context of *'gradual'* has an average slope of only 0.7, but *'taking off'* in the context of *'quick'* has a much steeper average slope of 2.7. Using this information along with word co-occurrence data for the specific <*adjective*><*verb*> pairs, we are able to annotate the different regions of the analyzed signals (Figures 1c and 1d) (refer to supplementary materials for expanded <*adjective*><*verb*> data). Selecting an <*adjective*><*verb*> annotation for a given chart region uses the following protocol:

1. Determine the slope of a given region using Ramer-Douglas-Peuker piecewise-linear decomposition [10].
2. Find all <*adjective*><*verb*> pairs whose average slope falls within a window (default = 0.5) of the desired slope.
3. From that set, select the <*adjective*><*verb*> pair with the highest annotation co-occurrence. The window in step 2 allows us to use annotation co-occurrence to select more common expressions like *'fast tanking'* instead of *'stagnant accelerating.'*

One of the goals of this work is to quantify relationships between visual features and annotations. Figures 1c and 1d and Figure 3 show that these terms do, in fact, work together to encode specific slope information that can be used to automatically annotate a univariate signal. Among terms, annotation clustering (refer to supplementary data) shows that terms tend to cluster in semantically intuitive ways. Collectively, these findings support the hypothesis that quantitative analysis of semantic labels may be capable of generating visual feature labels that are not only human-accessible but also quantitatively accurate. Providing language descriptors for accurately describing data insights can provide useful 'guard rails' and guidance as data summary generation becomes prevalent with the use of LLMs.

5.2 Visual Feature Integration with LLMs

To leverage the generative language of LLMs, we combine the semantic labels with additional information from the data set to form input prompts; for example, we employ the stock symbol and the dates of the discovered visual feature to ask the GPT 3.5 LLM [4] the templated question, "What happened between <*July 8, 2014*> and <*July 9, 2014*> that caused the stock symbol <*ALK*> to <***tank***>?". (Refer to supplementary materials for the prompt template.)

The specific LLM response is shown in Figure 1b. Notice that the model's responses implicitly integrate additional data into the user's investigation. For example, no data involving share price, earnings reports, or even the company name is explicitly linked to our chart. While these results are preliminary, additional research needs to explore the effectiveness of LLMs as ad hoc data sources.

6 Future Directions

The crowdsourced data has several layers of contribution. At a high level, simple shape identification enables the labeling of charts with appropriate language, whether for colloquial phrasing or domain-specific terminology. At a deeper level, we have begun to quantify the relationships between visual features and various hedge words, suggesting the possibility of numerically-accurate NL data descriptions. Finally, the data shows promise for analyzing the relationships among the different hedge words. While we believe that our work is an initial step in exploring the interplay of language and visual features, we identify the following future research directions:

Incorporate additional charts and domain-specific descriptions. Our dataset currently applies to univariate line charts. Future work should investigate language descriptors for other chart types, as well as labels for describing concepts for specific domains, such as associating *'flat'* with sales trends and *'constant'* with temperatures [32].

Use of LLMs to further semantic enrichment. We explored the use of GPT for summary generation using the labels describing visual features in a chart. The models do have limitations around higher-order numeracy reasoning and context [12]. For example, in Figure 1b, while the LLM provided several reasons for the stock price decline, the model missed the fact that there was a stock split. Custom-trained GPT models could potentially bridge this gap in higher-order analytical reasoning by incorporating additional knowledge. Other utilities for these custom LLMs could explore the automatic enrichment of additional descriptors for the dataset.

Supporting the search of shape descriptors. Annotations and summaries describing visual features in charts could be used as metadata in search interfaces to find pre-authored charts based on search queries such as, "find me the sales chart that has a spike in 2009, followed by a gradual decline," or in a voice assistant to ask for real-time notifications about data - "Hey Siri, tell me if this stock *tanks*." The work could also provide language prompts to LLMs to support sketching interfaces used for generating data stories [8].

7 Conclusion

This work explores the interplay of language and hedge words that describe visual features and their semantic relationships in line charts. We conducted a crowdsourced study to collect a range of label and visual feature pairs for these charts. Using this dataset of labeled visual signatures, we demonstrated its application for labeling charts and generating text summaries. The quantitative semantics presented in this work suggest a path forward for converting the crowdsourced dataset of feature-word descriptions into a semantic library of concepts that can distinguish between a *'rise'* and a *'gradual increase,'* for example. By making this dataset available to the broader research community, we believe that the work has useful implications for labeling and summarizing concepts for other chart types and their features, as well as for specific data domains. Our work suggests that, for the most part, people have a shared sense of semantic meaning. While the common saying, "Don't make a mountain out of a molehill," reprimands the exaggeration of a minor issue, perhaps exploring the *actual* difference between a mountain and a molehill is an important step towards better language and data understanding.

References

[1] Chart.js. https://www.chartjs.org/, 2022. 3

[2] Microsoft Power BI. https://powerbi.microsoft.com/en-us/, 2022. 2

[3] Tableau Software. https://www.tableau.com, 2023. 1, 2

[4] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020. 4

[5] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision To Think*. Morgan Kaufmann, 01 1999. 1

[6] C. Chen, R. Zhang, S. Kim, S. Cohen, T. Yu, R. Rossi, and R. Bunescu. Neural caption generation over figures. In *Joint Conference on Pervasive and Ubiquitous Computing and International Symposium on Wearable Computers*, pp. 482–485, 2019. 1

[7] C. Chen, R. Zhang, E. Koh, S. Kim, S. Cohen, and R. Rossi. Figure captioning with relation maps for reasoning. In *IEEE/CVF Conference on Applications of Computer Vision*, pp. 1537–1545, 2020. 2

[8] J. J. Y. Chung, W. Kim, K. M. Yoo, H. Lee, E. Adar, and M. Chang. Talebrush: Sketching stories with generative pretrained language models. In *SIGCHI Conference*, pp. 1–19, 2022. 4

[9] Z. Cui, S. K. Badam, M. A. Yalçin, and N. Elmqvist. Datasite: Proactive visual data exploration with computation of insight-based recommendations. *Information Visualization*, 18(2):251–267, 2019. 1

[10] 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. 4

[11] C. Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998. 2

[12] S. Frieder, L. Pinchetti, R.-R. Griffiths, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner. Mathematical capabilities of ChatGPT, 2023. 4

[13] T. Ghosal, K. K. Varanasi, and V. Kordoni. Hedgepeer: A dataset for uncertainty detection in peer reviews. In *ACM/IEEE Joint Conference on Digital Libraries*. ACM, 2022. 2

[14] T. Goodluck Constance, N. Bajaj, M. Rajwadi, H. Maltby, J. Wall, M. Moniri, C. Woodruff, T. Laird, J. Laird, C. Glackin, and N. Cannings. Resolving ambiguity in hedge detection by automatic generation of linguistic rules. In *International Conference on Artificial Neural Networks*, p. 369–380. Springer-Verlag, 2021. 2

[15] M. Hearst and M. Tory. Would you like a chart with that? Incorporating visualizations into conversational interfaces. In *2019 IEEE Visualization Conference (VIS)*, pp. 1–5. IEEE, 2019. 2

[16] R. Henkin and C. Turkay. Words of estimative correlation: Studying verbalizations of scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 28(4):1967–1981, 2020. 2

[17] K. Hu, D. Orghian, and C. Hidalgo. Dive: A mixed-initiative system supporting integrated data exploration workflows. In *Workshop on human-in-the-loop data analytics*, pp. 1–7, 2018. 1

[18] J. Hullman and N. Diakopoulos. Visualization rhetoric: Framing effects in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2231–2240, 2011. 2

[19] J. Hullman, N. Diakopoulos, and E. Adar. Contextifier: Automatic generation of annotated stock visualizations. In *SIGCHI Conference*, p. 2707–2716. ACM, 2013. 2

[20] J. Hullman, S. Drucker, N. H. Riche, B. Lee, D. Fisher, and E. Adar. A deeper understanding of sequence in narrative visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2406–2415, 2013. 2

[21] K. Hyland. *Hedging in Scientific Research Articles*. Pragmatics & beyond. John Benjamins Publishing Company, 1998. 1

[22] J. Islam, L. Xiao, and R. E. Mercer. A lexicon-based approach for detecting hedges in informal text. In *Language Resources and Evaluation Conference*, pp. 3109–3113, 2020. 2

[23] E. Kandogan. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 73–82, 2012. 2

[24] S. Kantharaj, R. T. Leong, X. Lin, A. Masry, M. Thakkar, E. Hoque, and S. Joty. Chart-to-text: A large-scale benchmark for chart summarization. In *Association for Computational Linguistics*, pp. 4005–4023, 2022. 2

[25] D. H. Kim, V. Setlur, and M. Agrawala. Towards understanding how readers integrate charts and captions: A case study with line charts. In *SIGCHI Conference*, pp. 1–11, 2021. 1, 2, 3

[26] J. Kim, J. Jo, and J. Seo. Autotator: Semi-automatic approach for accelerating the chart image annotation process. In *ISS Conference*, p. 315–318. ACM, 2019. 2

[27] H.-K. Kong, Z. Liu, and K. Karahalios. Frames and slants in titles of visualizations on controversial topics. In *SIGCHI Conference*, pp. 1–12, 2018. 2

[28] H.-K. Kong, Z. Liu, and K. Karahalios. Trust and recall of information across varying degrees of title-visualization misalignment. In *SIGCHI Conference*, pp. 1–13, 2019. 2

[29] N. Kong and M. Agrawala. Perceptual interpretation of ink annotations on line charts. In *UIST Conference*, pp. 233–236. ACM, 2009. 2

[30] N. Kong and M. Agrawala. Graphical overlays: Using layered elements to aid chart reading. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2631–2638, 2012. 2

[31] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013. 2

[32] G. Lakoff. *Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts*. Journal of Philosophical Logic. University of Chicago Press, 1973. 1, 4

[33] S. Latif, K. Su, and F. Beck. Authoring combined textual and visual descriptions of graph data. In *EuroVis*, pp. 115–119, 2019. 2

[34] 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, 2021. 2

[35] B. Lee, N. H. Riche, P. Isenberg, and S. Carpendale. More than telling a story: Transforming data into visually shared stories. *IEEE Computer Graphics and Applications*, 35(5):84–90, 2015. 2

[36] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. In Y. Bengio and Y. LeCun, eds., *ICLR Workshop*, 2013. 2

[37] V. O. Mittal, J. D. Moore, G. Carenini, and S. Roth. Describing complex charts in natural language: A caption generation system. *Computational Linguistics*, 24(3):431–467, 1998. 2

[38] A. Ottley, A. Kaszowska, R. J. Crouser, and E. M. Peck. The curious case of combining text and visualization. In *EuroVis*, pp. 121–125, 2019. 2

[39] M. Pérez-Echeverría, Y. Postigo, and C. Marín. Understanding of graphs in social science undergraduate students: selection and interpretation of graphs. *Irish Educational Studies*, 37:1–23, 02 2018. 1

[40] 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 *Web Conference*, pp. 2792–2804, 2021. 2

[41] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964. 4

[42] E. Segel and J. Heer. Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1139–1148, 2010. 2

[43] A. Srinivasan, S. M. Drucker, A. Endert, and J. Stasko. Augmenting visualizations with interactive data facts to facilitate interpretation and communication. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):672–681, 2018. 2

[44] C. Stokes and M. A. Hearst. Give text a chance: Advocating for equal consideration for language and visualization. In *NL VIZ Workshop*. IEEE, 2021. 2

[45] N. Sultanum, F. Chevalier, Z. Bylinskii, and Z. Liu. Leveraging text-chart links to support authoring of data-driven articles with vizflow. In *SIGCHI Conference*, pp. 1–17, 2021. 2

[46] J. H. Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963. 3

[47] D. Whitaker and T. Jacobbe. Students' understanding of bar graphs and histograms: Results from the locus assessments. *Journal of Statistics Education*, 25(2):90–102, 2017. 1