Show It or Tell It?
Text, Visualization, and Their Combination

Marti A. Hearst
hearst@berkeley.edu
University of California, Berkeley
Berkeley, CA, USA

Abstract

This article presents recent results and pressing questions about how language and visualization should be combined. These questions include: how much and what kind of text should be shown on visualizations? Under what circumstances do people prefer text over visualizations, and why? What do recent advances in machine learning – specifically large transformer models that combine language, visuals, and code – mean for the future of information visualization?

Keywords: information visualization, Natural Language Processing

ACM Reference Format:
Marti A. Hearst. 2023. Show It or Tell It? Text, Visualization, and Their Combination. In *Proceedings of Communications of the ACM (CACM)*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

The field of information visualization studies how visual representations can express relationships among abstract data. Visualization is often compared with alternative forms of presentation such as tables of numbers. Although research assumes that visualizations are embedded in context — within newspapers, textbooks, social media posts, and presentation slides – the composition and placement of the *language* used in charts is usually an afterthought. For example, Apple recently released a set of user interface guidelines which include patterns for designing charts [2] (see Figure 1[1]). This anatomy of a chart [29] contains a carefully designed layout, but entirely omits guidelines for the placement of the title and textual annotations. This omission reflects the assumptions within the field of information visualization, including many of its textbooks.

These assumptions persist despite the fact that the seminal work of Borkin et al. [6] showed that the language components of visualization are of key importance. That study compared a very large number of infographic designs, finding that the written text was the more memorable part of the visualization.

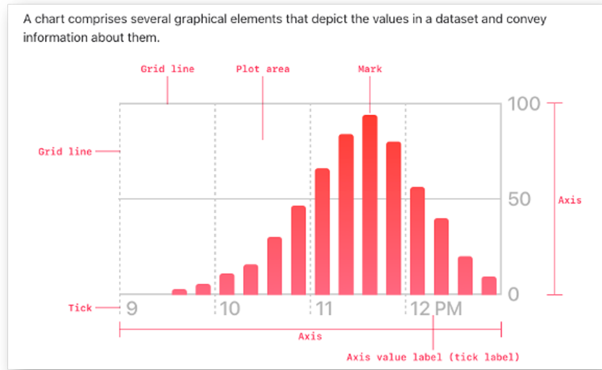

Figure 1. Apple Anatomy of a Chart, with no indication of title or text annotations.

This article argues that language should be considered as co-equal with visualization when communicating information. This is a rather radical statement for the visualization community; that said, there has been a recent surge of interest in this topic, including a new workshop on natural language and visualization [52]. This paper is an attempt to bring some of the questions and the results to a wider audience.[2]

The remainder of this essay discusses these main themes:

- **Combining Text + Viz**: How much text should be placed on a visualization and where should it go? What should that text consist of? How can results from linguistics be integrated into this research?
- **Text Alone**: Empirical work in visualization should, as a standard practice, compare charts against a baseline of no visualizations at all – a baseline of expressing the same information that is on the chart in language. Empirical evidence suggests there is a significant minority of people who tend to prefer no visualizations.
- **Literacy**: The visualization field has theories of visual literacy, but it should incorporate theories of reading literacy as well.

[1]From https://developer.apple.com/design/human-interface-guidelines/components/content/charts/, retrieved 4/23/23

[2]This paper uses the words "language" and "text" interchangeably, the same goes for "charts" and "visualizations".

CACM, 2023, NY, NY
2023. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

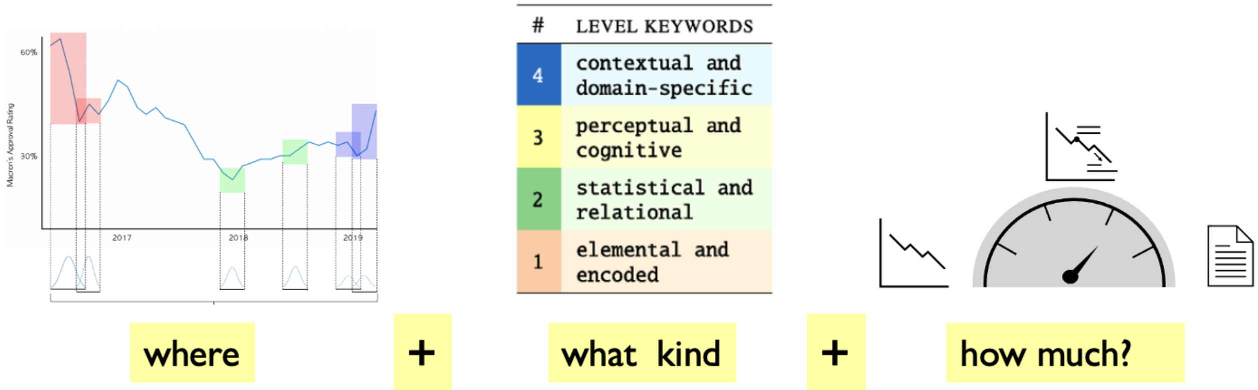

Figure 2. (Left) A study of where on a chart the most visual salient components are, from [26]. (Middle) Four levels of semantic description, from low (1) to high (4), from [37]. (Right) A schematic showing how much text to place on a chart, based on [55].

- **Cognitive Models**: The field does not make use of verified cognitive theories of how combinations of language and visualizations are read, perceived, and understood.
- **Language as the UI for Viz**: The rather spectacular advances that are happening in natural language processing may have major impacts on how visualizations are created in future.

2 Striking a Balance Between Text and Viz

Within the visualization community, Scott McCloud's brilliant book, *Understanding Comics* [39], is an inspirational classic. To illustrate the tradeoffs between what is depicted in images versus words, in Chapter 5, McCloud introduces a running example. In the first view, the message is expressed only with pictures, no words. The image shows the scene, the mood, and the action, allowing the words to be freed up to express something else, such as the inner thoughts of the character in the scene. In the reverse case, McCloud shows a comic consisting only of words above empty spaces. In this comic, the words carry the weight of describing the scene, action, and character's internal state. When images are added, they can zoom in to show just a piece of the action or to convey the mood.

McCloud shows that if the image takes on one part of the description, the text is freed up to show some other content, and vice versa. This framework can be applied to research questions about how text and visualization should be combined. We can break this further into: **What** is the nature of the text that should appear on the visualization? **Where** should it be placed? **How much** is too much? And how do the visual and the language components interact? The following subsections describe research addressing each of these questions.

Where? Kim et al. [26] investigated the question: How do captions influence what people take away from charts? To better observe the influence of the text, they developed a method to determine which parts of a univariate line chart are most visually salient. (See Figure 2 (Left)). After finding the salient regions (the "where"), Kim et al. [26] created captions that corresponded to each of these salient parts of the line chart (for instance, a sharp peak). The experimenters wanted to know if the content of the captions influence what people take away from the charts, and if that text can override the most visually salient parts.

The experimenters found that if the captions referred to the most salient parts of the chart, the most salient parts were recalled. But if the caption referred to the parts of the chart that were *not* the *most* visually salient, people recalled the parts called out by the text rather than the most salient regions; in other words, the text overruled the visuals. But in the final case, if the caption referred to something not visually salient *at all*, then what participants recalled was more influenced by the chart. (This work was recently verified and extended [11].)

These findings suggest that there is a complex relationship between the effects of the visuals versus the effects of the textual. There seemed to be a tipping point between when visual had more sway than textual.

What Kind? Lundgarden & Satyanarayan [37] investigated the question: What kind of language is preferred for describing charts by blind and low vision (BLV) people vs. sighted people?

In this study, the experimenters first asked participants to write descriptions of charts; they then analyzed these texts and identified four levels of semantics, L1 through L4 (see Figure 2 (Middle)). The lowest level, L1, describes the components of the chart, while the highest level, L4, describes external contextualizing information not visible on the chart. These semantic levels categorize "what kind" of text is used to describe visualizations.

Interestingly, the BLV participants preferred different kinds of textual information than sighted people. In particular, the majority of BLV readers opposed high level L4 expression, which by contrast was favored by the majority of sighted readers; the converse was true for low-level L1 language. These results are important in themselves for informing how to write alternative text for accessibility purposes.

How Much? Armed with "where" and "what kind", I and several collaborators conducted a study that asked: "how much" As in, how much text is too much for annotation as an overlay on a chart [57] (see Figure 2 (Right)).

Working with univariate line charts, we systematically varied chart design from all chart and no text, all the way to all text and no chart, shown in Figure 3. We created the charts by first finding the visually salient components as in [26], and then labeling those components with the different semantic levels as in [37], varying them for a controlled experiment with crowdworkers. We assessed these designs in two ways: with preference questions and according to how people took away information from the charts.

In terms of preference, more text context (type C in Figure 3) was preferred by a majority of participants. In a subsequent analysis [55], we found that although text can at first glance make the chart appear more cluttered, in actuality, this extra context was helpful and preferred, so long as the text was relevant and not redundant. This study also found the surprising result that 14% of participants ranked choice D, the all text paragraph in Figure 3, as their top choice.

In terms of what information the participants took away from the charts, our findings were: how much: use relevant text, do not worry extensively about the clutter issue, where: the best position depends on the type of semantic content (level) being shown, and what kind: the best semantic level depends on the message being conveyed. In summary, this study found that more text was better. That said, more research is needed to look at more complex and diverse chart types.

3 Case Study: Comparisons

A case study of the differences and dependencies between language and visualization can be understood through the case study of comparisons. Comparisons have been studied by both linguistics and visualization research. Both fields recognize the challenge of comparisons and both fields see this construct as being expressed in a diverse manner. On the linguistics side, for instance, Friedman [16] states that "The comparative is a difficult structure to process for both

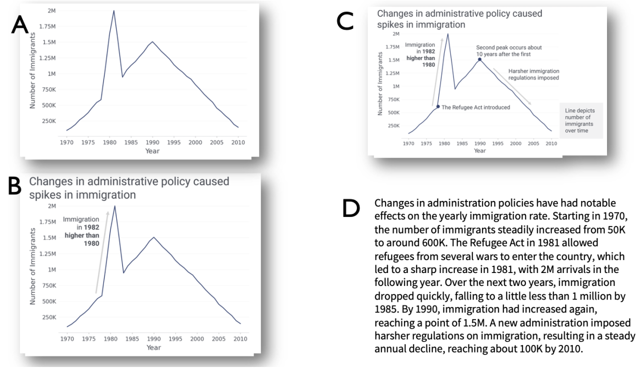

Figure 3. Four charts, from no text annotations (A) to all text (D), used to compare participant preferences, from the study of [55].

syntactic and semantic reasons. Syntactically, the comparative is extraordinarily diverse." On the visualization side, Gleicher [19] writes: "Supporting comparison is a common and diverse challenge in visualization."

Despite this commonality, their methods for addressing comparisons are quite divergent. In linguistics the difficulty is the variation in expression, the challenge of determining what entities are being compared, and what those relationships are. Below are two ways of expressing comparisons, taken from a camera reviews collection [25]. The syntactic structure and lexical choices between just these two examples are very different:

> "I felt more comfortable with XTi and some of my friends felt more comfortable with D80."
>
> "On the other hand I actually prefer the D80 handling with smaller lenses , which is what 's on my camera 80% of the time."

By contrast, the visualization literature assumes the entities and relationships being compared are known, and instead asks how to show those relationships, and how to make them scale. A single sentence of language can only compare a few things at a time, but visualization compares dozens, hundreds, or thousands of items at once.

Comparisons with Vague Modifiers. In the field of cognitive linguistics, Schmidt et al. [50] examine the question: how do people decide what the meaning is of "tall"? What is "tall" vs "not tall"? The answer, determined empirically and with modeling, is: it depends on the distribution of the data points. For instance, a step function yields more agreement than an exponential drop off (see Figure 4).

My colleagues and I used this result to determine how to show visualizations in response to natural language comparisons that contained vague modifiers like "tall" and "cheap" [23]. We used the results of Schmidt et al. [50] to determine

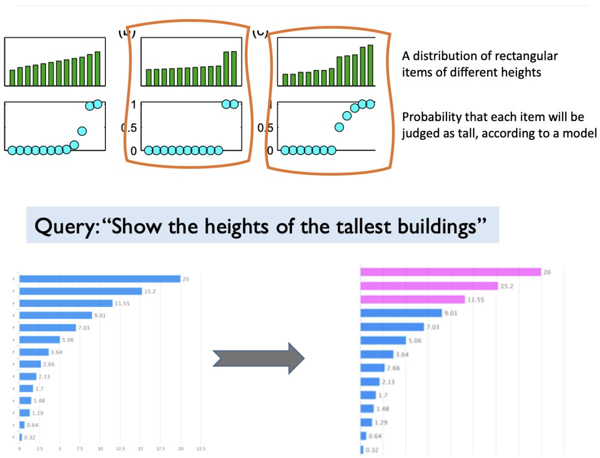

Figure 4. (Top) From Schmidt et al. [50], the degree of agreement between human judges for what items are considered "tall" varies based on the distribution of those items. (Bottom) Using this result to label charts in a conversational interface, as described in [23].

which bars to highlight for a response to a superlative comparison question such as "Show the heights of the tallest buildings". For instance, for the exponential drop off, the cognitive linguistics model shows us which bars we should highlight, depending on the shape of the curve (see Figure 4).

Another recent study [17] looked looked at several forms of linguistic expressions of comparisons, and for each type, determined which kind of visualization best expressed these comparisons according to crowdworkers.

Comparison Questions in Chat Interfaces. My colleagues and I did another study to assess a related question: how to show visualizations in a conversational interface for an intelligent assistant like Siri or Alexa [22]. The goal was to determine what kind of visual context people prefer after they ask a comparison question with a simple answer, such as "Which Olympic sport has the tallest players: rowing or swimming?"

We found that many people preferred bar charts so long as the chart did not get too long. However, we also found that *41% of participants did not want to see any chart at all* in this context. They preferred text alone. Statements by participants showed that those who preferred bar charts, preferred seeing the data points in the context of other bars. People who preferred text said that it is precise, and not overly complicated. A few participants switched from preferring text to preferring charts, when the situation merited it.

In summary, comparisons are a good case study for delving deeply into the questions about the differences between visualizing a concept versus expressing it in language; findings from cognitive and computational linguistics can help shed light.

4 Text Alone

The previous section recounted two cases in which text without a chart was preferred by a sizable minority of participants. This result is surprising from the perspective of information visualization research. Although not common, there are other examples to be found in the literature in which the text alone condition was tested.

For instance, McKenna et al. [40] conducted a study which compared different ways of presenting a scrollytelly design (a design in which visualizations appear dynamically within the text as the user scrolls down a long web document), including one design with no visualizations at all. Surprisingly, a notable minority of 10% of participants said they preferred this no-visual condition.

In an investigation by Ottley et al. [42], the researchers experimented with methods to help explain Bayesian reasoning. They compared text alone, visualization alone, and the two juxtaposed. They found that visualization was not more accurate than text for this purpose. They also found that when text and visualizations were presented together, participants did not seem to take advantage of the distinct affordances of each.

The last example is the well-known "Explaining the Gap" paper by Kim et al. [27] which was motivated by You Draw It interactive feature at the *New York Times* [24]. Its goal was to compare how well people recalled data depending on whether they had to first predict a trend or not. Their experiment design included a comparison between text-only and visualization-only conditions. The authors had three major findings with respect to text. First, presenting data as text helps people recall those values better than with a visualization. Second, the visualizations were better than text at helping people recall trends. Third, the aid to prediction was found for the visualization but not the text.

In summary, studies provide strong evidence that a text-only variant should be tested when assessing the efficacy of a design [56].

5 Preferences or Literacy?

One explanation for why people prefer text alone over visualization is personal preferences; it could be that some people prefer reading while others prefer visual information. But this begs the question as to *why* people prefer one over the other.

Within education circles, there is much discussion of math literacy (or numeracy), and computational literacy [59]. Within the visualization community, the notion of *visualization literacy* has recently interested researchers [7, 8, 34]. Solen [54] defines it as: "the ability to critically interpret and construct visualizations." An explanation for the differences in preferences be could that some people have not learned how to interpret charts, or have not had enough practice interpreting them, to be "fluent" at reading them, which results in their dispreferring them.

However, visualization research rarely considers the flip side – the role of reading literacy – the original meaning of the word. The reading expert Maryanne Wolf explains what is understood about cognition and reading, and relates it to the importance of fluency in reading [60]. Wolf notes that true literacy can be achieved only when readers become expert – that is fluent – readers. Wolf opens her book with: "We were never born to read", meaning that although most humans innately learn spoken language, reading is not innate. She points out that in order to learn to read, special pathways need to be formed across many different brain regions that were not evolved for reading.

Wolf points out that researchers have gathered extensive evidence that the processing of words occurs in the parafovea, before the word is directly fixated on. An expert reader uses their peripheral vision to pick up on visual characteristics such as word shape. This peripheral vision does not usually indicate the word's meaning, but it can approximate the general shape of what is to come. Wolf notes that this preview of what lies ahead on the line contributes to fluent reading. Wolf also talks about why fluent reading is so important – it gives enough time to the executive system to direct attention where it is most needed – to infer, to understand, to predict. In other words, to think while you are reading. Thus, literacy with fluent reading opens the door to developing new understanding while reading.

A growing trend in news reporting and scholarly publishing is to insert visualizations within the body of text paragraphs; Figure 5 shows an example from the *New York Times* [4]. This practice does not take into account how integrating visualizations within text can disrupt fluent reading.

In looking at the figure, consider how easy or difficult it is to read the text. Does you read the paragraph straight through, or does your eye dart from the visualizations to the text, and back again in an erratic manner?

There is evidence that, when the paragraph contains unexpected images, it can disrupt fluent reading. Consider hyperlinks. Fitzsimmons et al [15] found that readers focus on hyperlinks when skimming, and they tend to use these links as markers for important parts of the text. Similarly, studies show that emoji icons embedded within text can slow down reading [3, 12]. Thus, given the prior results, it is likely that embedded visualizations as shown in Figure 5 will have deleterious effects on fluent reading [20]

In summary, in order to shine light on the reasons for people's preferences for text versus visualization, future work should consider both reading and visualization literacy and fluency when combining language and charts.

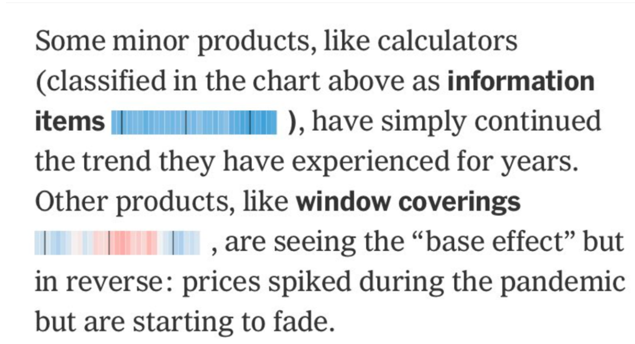
Some minor products, like calculators (classified in the chart above as **information items**), have simply continued the trend they have experienced for years. Other products, like **window coverings** , are seeing the "base effect" but in reverse: prices spiked during the pandemic but are starting to fade.

Figure 5. Visualizations inserted within a paragraph, from the *New York Times* [4].

6 Cognitive Models for Combining Text + Vis

Cognitive models are used to understand why and how mental processes work and to aid in formulating predictions, such as how a person viewing a visualization will interpret it [43]. For instance, Padilla et al. [44] used a cognitive model to identify the specific process underlying why people misinterpreted hurricane forecast visualizations. However, within the visualization field, there is currently no commonly used cognitive model to shed light on the question of how text and visualizations are mentally processed when combined. In the Bayesian reasoning experiment described above, the authors noted that the field of visualization does not have sophisticated guidelines for understanding how to combine the two modalities [42].

Mayer [38] has done extensive research on combining language and visuals for the purposes of education. This work considers the placement and modality of text (written or audio) within images in the limited context of physical process explanations.

The cognitive theory that Mayer employs is the *dual-channel model* [38] which assumes separate cognitive systems or channels for processing pictoral versus verbal information. It assumes that each channel has limited capacity, and that meaningful learning involves actively building connections between the two.

From the field of journalism, Sundar [58] presents three main cognitive model theories. The first is the dual-channel theory just mentioned that states that there are two cognitive subsystems for language versus image, and they operate independently when coding information into memory. The next is *cue-summation theory* that posits that when text and visuals are presented together, text provides additional learning cues, particularly at memory retrieval time. The third is the *limited capacity information processing theory*, which states that combining multiple modalities overwhelms the system. Together, these theories cover all of the cases: text plus visuals are either independent, additive, or interfering.

There does not seem to be any consensus or even strong evidence for which is correct.

In summary, there is no widely accepted cognitive model for how text and images are perceived together, which may cause empirical results to be less predictive than if such a model existed. One remedy is for the visualization community to engage with cognitive scientists on this important and underexplored question.

7 Language as a User Interface for Vis

There has been extensive prior work on incorporating natural language processing (NLP) into information visualization systems [53]. This includes using language as a query against data to create a visualization [41, 51] and using natural language to build and refine designs of visualizations [13, 18]. However, these systems use technology that predates the recent advances in NLP. They often consist of a software pipeline of diverse algorithms – often including tokenizers, part-of-speech taggers, syntax parsers, entity recognizers, and a variety of semantic analyzers; each stage requires its own hand-labeled training data and format, and information is lost from one stage in the pipeline to the next. These systems are either limited in scope or are unable to get robust coverage of the possible ways to express relevant concepts.

Large generative language models (LLMs) such as GPT-3 and T5 are transforming the fields of both NLP and computer vision [5]. Although the models are very large in terms of input data and parameters trained, their architectures are in some sense simple compared to the NLP pipelines of the past. In the new transformer-based or diffusion-based approaches, one model is used for a wide variety of applications. For instance, T5 is trained on many tasks simultaneously, with the input represented as a textual description of the task [46]. For some tasks, training is self-supervised, meaning that the training phase does not require hand-labeled examples. These systems require enormous amounts of data and compute power to train on, but when that training is complete, the resulting models can be used as is, or fine-tuned on a specific problem, often with few labeled training examples [49].

Today, LLMs allow for language to be used as the interface to generate general images. Some of the new models train simultaneously on text and image input, creating a model that represents the two modalities jointly. Applications like DALLE-2, Imagen, and Midjourney can produce photo-realistic images in response to textual prompts and have become a popular way to for non-technical users to generate sophisticated, humorous, and surreal images. For instance, Figure 6 shows the output of DALLE-2 [48] in response to the text input “Interior of a library filled with books, with a stock line chart on an easel in the center, oil painting”.

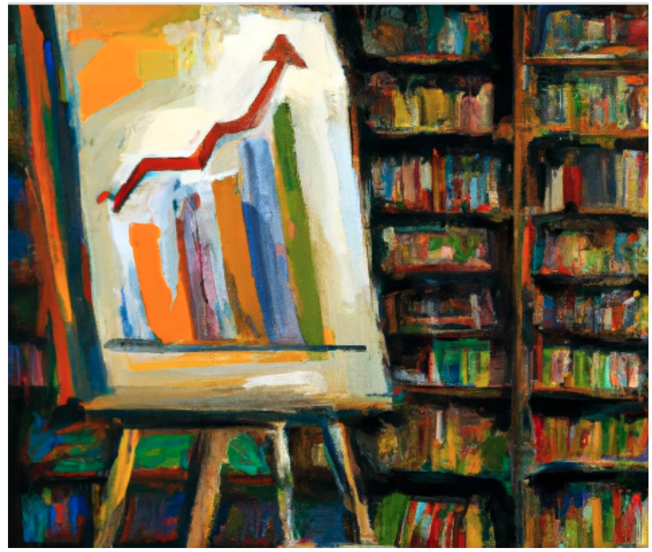

Figure 6. Image generated with the help of DALLE-2 [48] in response to the text prompt: “Interior of a library filled with books, with a stock line chart on an easel in the center, oil painting”.

LLMs are also having great success at automatic code suggestions based on natural language commands, as seen in Github CoPilot, based on Codex [10], and AlphaCode [35]. These LLMs can aid in the creation of programmatically-defined visualizations. For example when Copilot is given the commented command:

```
# plot this as a bar chart with bars colored blue
unless the car name starts with 'M'
```

CoPilot responds with:

```
plt.bar(cars, values, color= ['blue' if not
car.startswith('M') else 'red' for car in cars]
```

An entire matplotlib program can be quickly specified that both makes up data and plots it on a chart. CoPilot, as part of Github, is already used by hundreds of thousands of programmers, many of whom think it enhances their productivity [61] (although studies suggest that its use can result in programming errors or security flaws [45]).

As the models improve, it will become more feasible to express what is wanted using natural language than by writing commands, code, or even using a graphical user interface. There has been a long running debate about which is better for exploring, analyzing, and visualizing data: using a GUI or writing code. People’s preference depends on which tool they are most comfortable with, and many experts use a combination of the two [1]. The very rapid uptake and popularity of tools like CoPilot suggests that the answer in the long run is going to be: not GUI, not code, but language. We will simply speak or type how we want the data to be visualized, perhaps augmented with pointing or gesturing.

(It will be important to have the models trained on high quality designs, so that they do not reproduce designs with poor usability.)

This preference for using language as the user interface is not new. In the late 1990's, a search engine called Ask Jeeves that purported to let people enter natural language questions rather than type keywords was enormously popular, despite being brittle and having low coverage [30]. Ask Jeeves employed people to create an enormous databases of questions and answers; it took 20 years of development before major search engines could do this task reliably.

It is important to draw attention to the many known problems and concerns with LLMs. First, they are trained on huge collections of data, and so if care is not taken, they repeat the biases and injustices that are inherent in those collections. Another major problem is that the field does not really understand how they work, and furthermore, the results they produce cannot be predicted or explained in a way that makes sense to people. The third problem is that they are far from perfect, and they produce compelling output without having what we would consider an understandable internal representation of what it is they are producing. They are huge, not available to all researchers or users due to their size, and they are costly to train in terms of compute time, and to a lesser degree, energy consumption (although some of these drawbacks may subside due to research efforts). There are concerns about how the building of models from other people's intellectual property perhaps violates their ownership rights. Perhaps the biggest drawback of all is how these models can contribute to misinformation and make it very hard to determine what information is generated by humans versus by computer software.

In summary, although large language models today are still far from fully able to be used to generate visualizations, they are likely to significantly transform how we generate them in future. For instance, there is already some preliminary work in using them to generate captions for charts [36].

8 Conclusions

This essay has advocated for and described research about the complex interactions when combining text with visualizations, the importance of considering text alone when assessing visualization designs, the need for better cognitive models that combine reading and understanding visualizations, and a future projection of language as the user interface for visualizations.

There area many other topics in this space that are not covered here, including the role of bias and slant [28], misinformation and deception [14, 33], multi-lingual text [47], visualizing text itself [9, 21], linking interactive visualizations within documents [31**?** , 32], and spoken versus written text in combination with visualization. The field is ripe for innovation and increased understanding.

Acknowledgments

I thank David Ebert, Hendrick Strobelt, and Danielle Szafir for inviting me to give the IEEE Vis 2022 keynote talk, upon which this paper is based, and Chase Stokes, Cindy Xiong, Aiman Gaba, and Huichen (Will) Wang for feedback on the talk. Finally, I thank the anonymous reviewers for their feedback and suggestions for improvement of this manuscript.

References

[1] Sara Alspaugh, Nava Zokaei, Andrea Liu, Cindy Jin, and Marti A Hearst. 2018. Futzing and moseying: Interviews with professional data analysts on exploration practices. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2018), 22–31.

[2] Apple. Retrieved Oct 30, 2022. Human Interface Guidelines. https://developer.apple.com/design/human-interface-guidelines/guidelines/overview/.

[3] Eliza Barach, Laurie Beth Feldman, and Heather Sheridan. 2021. Are emojis processed like words?: Eye movements reveal the time course of semantic processing for emojified text. *Psychonomic Bulletin & Review* (2021), 1–14.

[4] Josh Bivens and Stuart A. Thompson. 2021. 179 Reasons You Probably Don't Need to Panic About Inflation. https://www.nytimes.com/interactive/2021/08/18/opinion/inflation-economy-transitory.html. *New York Times* (18 August 2021).

[5] Rishi Bommasani, Drew A Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, et al. 2021. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258* (2021).

[6] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. 2015. Beyond Memorability: Visualization Recognition and Recall. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2015), 519–528.

[7] Katy Börner, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1857–1864.

[8] Jeremy Boy, Ronald A Rensink, Enrico Bertini, and Jean-Daniel Fekete. 2014. A principled way of assessing visualization literacy. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 1963–1972.

[9] Richard Brath. 2020. *Visualizing with Text* (1 ed.). A K Peters/CRC Press.

[10] Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde de Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, et al. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374* (2021).

[11] Shelly Cheng, Hazel Zhu, and Eugene Wu. 2022. How Do Captions Affect Visualization Reading?. In *IEEE Vis Workshop on Visualization for Communication (VisComm).*

[12] Neil Cohn, Tim Roijackers, Robin Schaap, and Jan Engelen. 2018. Are emoji a poor substitute for words? Sentence processing with emoji substitutions.. In *Annual Cognitive Science Society Meeting: Changing/Minds.*

[13] Weiwei Cui, Xiaoyu Zhang, Yun Wang, He Huang, Bei Chen, Lei Fang, Haidong Zhang, Jian-Guan Lou, and Dongmei Zhang. 2019. Text-to-viz: Automatic generation of infographics from proportion-related natural language statements. *IEEE Transactions on Visualization and Computer Graphics* 26, 1 (2019), 906–916.

[14] Arlen Fan, Yuxin Ma, Michelle Mancenido, and Ross Maciejewski. 2022. Annotating Line Charts for Addressing Deception. In *CHI Conference on Human Factors in Computing Systems*. 1–12.

[15] Gemma Fitzsimmons, Lewis T Jayes, Mark J Weal, and Denis Drieghe. 2020. The impact of skim reading and navigation when reading hyperlinks on the web. *PloS One* 15, 9 (2020), e0239134.

[16] Carol Friedman. 1989. A general computational treatment of the comparative. In *27th Annual Meeting of the Association for Computational Linguistics*. 161–168.

[17] Aimen Gaba, Vidya Setlur, Arjun Srinivasan, Jane Hoffswell, and Cindy Xiong. 2022. Comparison conundrum and the chamber of visualizations: An exploration of how language influences visual design. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[18] Tong Gao, Mira Dontcheva, Eytan Adar, Zhicheng Liu, and Karrie G Karahalios. 2015. Datatone: Managing ambiguity in natural language interfaces for data visualization. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*. 489–500.

[19] Michael Gleicher. 2017. Considerations for visualizing comparison. *IEEE Transactions on Visualization and Computer Graphics* 24, 1 (2017), 413–423.

[20] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. 2014. Exploring the placement and design of word-scale visualizations. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2291–2300.

[21] Marti Hearst. 2009. *Search user interfaces*. Cambridge University Press.

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

[23] Marti Hearst, Melanie Tory, and Vidya Setlur. 2019. Toward interface defaults for vague modifiers in natural language interfaces for visual analysis. In *2019 IEEE Visualization Conference (VIS)*. IEEE, 21–25.

[24] Josh Katz. 2017. You Draw It: Just How Bad is the Drug Overdose Epidemic? https://www.nytimes.com/interactive/2017/04/14/upshot/drug-overdose-epidemic-you-draw-it.html. *New York Times* (26 October 2017).

[25] Wiltrud Kessler and Jonas Kuhn. 2014. A corpus of comparisons in product reviews. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*. 2242–2248.

[26] Dae Hyun Kim, Vidya Setlur, and Maneesh Agrawala. 2021. Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[27] Yea-Seul Kim, Katharina Reinecke, and Jessica Hullman. 2017. Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1375–1386. https://doi.org/10.1145/3025453.3025592

[28] Ha-Kyung Kong, Zhicheng Liu, and Karrie Karahalios. 2018. Frames and slants in titles of visualizations on controversial topics. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

[29] Stephen M. Kosslyn. 2006. *Graph design for the eye and mind*. OUP USA.

[30] Cody C.T. Kwok, Oren Etzioni, and Daniel S Weld. 2001. Scaling question answering to the Web. In *Proceedings of the 10th International Conference on the World Wide Web*. 150–161.

[31] Sébastien Lallé, Dereck Toker, and Cristina Conati. 2021. Gaze-Driven Adaptive Interventions for Magazine-Style Narrative Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 27, 6 (2021), 2941–2952. https://doi.org/10.1109/TVCG.2019.2958540

[32] Shahid Latif, Zheng Zhou, Yoon Kim, Fabian Beck, and Nam Wook Kim. 2022. Kori: Interactive Synthesis of Text and Charts in Data Documents. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2022), 184–194. https://doi.org/10.1109/TVCG.2021.3114802

[33] Crystal Lee, Tanya Yang, Gabrielle D Inchoco, Graham M Jones, and Arvind Satyanarayan. 2021. Viral visualizations: How coronavirus skeptics use orthodox data practices to promote unorthodox science online. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–18.

[34] Sukwon Lee, Sung-Hee Kim, and Bum Chul Kwon. 2016. Vlat: Development of a visualization literacy assessment test. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2016), 551–560.

[35] Yujia Li, David Choi, Junyoung Chung, Nate Kushman, Julian Schrittwieser, Rémi Leblond, Tom Eccles, James Keeling, Felix Gimeno, Agustin Dal Lago, et al. 2022. Competition-level code generation with alphacode. *arXiv preprint arXiv:2203.07814* (2022).

[36] Ashley Liew and Klaus Mueller. 2022. Using Large Language Models to Generate Engaging Captions for Data Visualizations. In *IEEE Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis (NL Vis)*.

[37] Alan Lundgard and Arvind Satyanarayan. 2021. Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (2021), 1073–1083. https://doi.org/10.1109/TVCG.2021.3114770

[38] Richard E. Mayer. 2020. *Multimedia Learning* (3rd ed.). Cambridge University Press.

[39] Scott McCloud. 1993. *Understanding Comics: The Invisible Art*. Northampton, Mass.

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

[41] Arpit Narechania, Arjun Srinivasan, and John Stasko. 2020. NL4DV: A toolkit for generating analytic specifications for data visualization from natural language queries. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2020), 369–379.

[42] Alvitta Ottley, Aleksandra Kaszowska, R Jordan Crouser, and Evan M Peck. 2019. The Curious Case of Combining Text and Visualization.. In *EuroVis (Short Papers)*. 121–125.

[43] Lace M. Padilla. 2018. A Case for Cognitive Models in Visualization Research : Position paper. In *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*. 69–77. https://doi.org/10.1109/BELIV.2018.8634267

[44] Lace M Padilla, Ian T Ruginski, and Sarah H Creem-Regehr. 2017. Effects of ensemble and summary displays on interpretations of geospatial uncertainty data. *Cognitive research: principles and implications* 2 (2017), 1–16.

[45] Hammond Pearce, Baleegh Ahmad, Benjamin Tan, Brendan Dolan-Gavitt, and Ramesh Karri. 2022. Asleep at the keyboard? assessing the security of github copilot's code contributions. In *2022 IEEE Symposium on Security and Privacy (SP)*. IEEE, 754–768.

[46] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research* 21, 140 (2020), 1–67.

[47] Noëlle Rakotondravony, Yiren Ding, and Lane Harrison. 1912. Probablement, Wahrscheinlich, Likely? A Cross-Language Study of How People Verbalize Probabilities in Icon Array Visualizations. *IEEE Transactions on Visualization and Computer Graphics* (1912).

[48] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv:2204.06125* (2022).

[49] Adam Roberts, Colin Raffel, and Noam Shazeer. 2020. How Much Knowledge Can You Pack Into the Parameters of a Language Model?. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 5418–5426.

[50] Lauren A Schmidt, Noah D Goodman, David Barner, and Joshua B Tenenbaum. 2009. How tall is tall? Compositionality, statistics, and

gradable adjectives. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Citeseer, 2759–2764.

[51] Vidya Setlur, Sarah E Battersby, Melanie Tory, Rich Gossweiler, and Angel X Chang. 2016. Eviza: A natural language interface for visual analysis. In *Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology*. 365–377.

[52] Vedya Setlur and Arjun Srinivasan (Eds.). 2021. *IEEE Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis (NL Viz)*. IEEE.

[53] Leixian Shen, Enya Shen, Yuyu Luo, Xiaocong Yang, Xuming Hu, Xiongshuai Zhang, Zhiwei Tai, and Jianmin Wang. 2021. Towards Natural Language Interfaces for Data Visualization: A Survey. *IEEE transactions on visualization and computer graphics* PP (2021).

[54] Mara Solen. 2022. Scoping the Future of Visualization Literacy: A Review. In *IEEE Vis workshop on Visualization for Communication (VisComm)*.

[55] Chase Stokes and Marti Hearst. 2022. Why More Text is (Often) Better: Themes from Reader Preferences for Integration of Charts and Text. In *IEEE Vis Workshop: Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis (NL Vis)*.

[56] Chase Stokes and Marti A Hearst. 2021. Give Text A Chance: Advocating for Equal Consideration for Language and Visualization. In *IEEE Workshop on Exploring Opportunities and Challenges for Natural Language Techniques to Support Visual Analysis (NL Viz)*. IEEE.

[57] Chase Stokes, Vidya Setlur, Bridget Cogley, Arvind Satyanarayan, and Marti A Hearst. 2022. Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[58] S Shyam Sundar. 2000. Multimedia effects on processing and perception of online news: A study of picture, audio, and video downloads. *Journalism & Mass Communication Quarterly* 77, 3 (2000), 480–499.

[59] Uri Wilensky, Corey E Brady, and Michael S Horn. 2014. Fostering computational literacy in science classrooms. *Commun. ACM* 57, 8 (2014), 24–28.

[60] Maryanne Wolf. 2007. *Proust and the Squid: The Story and Science of the Reading Brain*. Harper Perennial New York.

[61] Albert Ziegler, Eirini Kalliamvakou, X Alice Li, Andrew Rice, Devon Rifkin, Shawn Simister, Ganesh Sittampalam, and Edward Aftandilian. 2022. Productivity assessment of neural code completion. In *Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming*. 21–29.