

"Interior of a library filled with books, with a stock line chart on an easel in the center, oil painting"

**SHOW IT OR  
TELL IT?**  
IEEE Vis'22 KEYNOTE

**MARTI HEARST  
UC BERKELEY**

Show It or Tell It? Viz 2022 Keynote  
Marti Hearst  
UC Berkeley

I want to thank the conference organizers Danielle Szafir, David Ebert, and Hendrik Strobel for inviting me to speak. I'd also like to thank my PhD student Chase Stokes, as well as Cindy Xiong, and her students Aiman and Will for feedback on this talk.

The title of my talk is: Show It or Tell It: Text, Visualization, and their Combination.

In viz research, we often study which visualization best expresses something.

We also know that visualizations are embedded in context – within newspapers, textbooks, social media posts and powerpoint slides.

But the role of language used is often an afterthought.

Human Interface Guidelines for Charts are here! We've been working on these pages for a bit—hope you find them useful 📊👤

Patterns page: [developer.apple.com/design/human-i-...](https://developer.apple.com/design/human-interface-guidelines/patterns/)  
Components page: [developer.apple.com/design/human-i-...](https://developer.apple.com/design/human-interface-guidelines/components/)

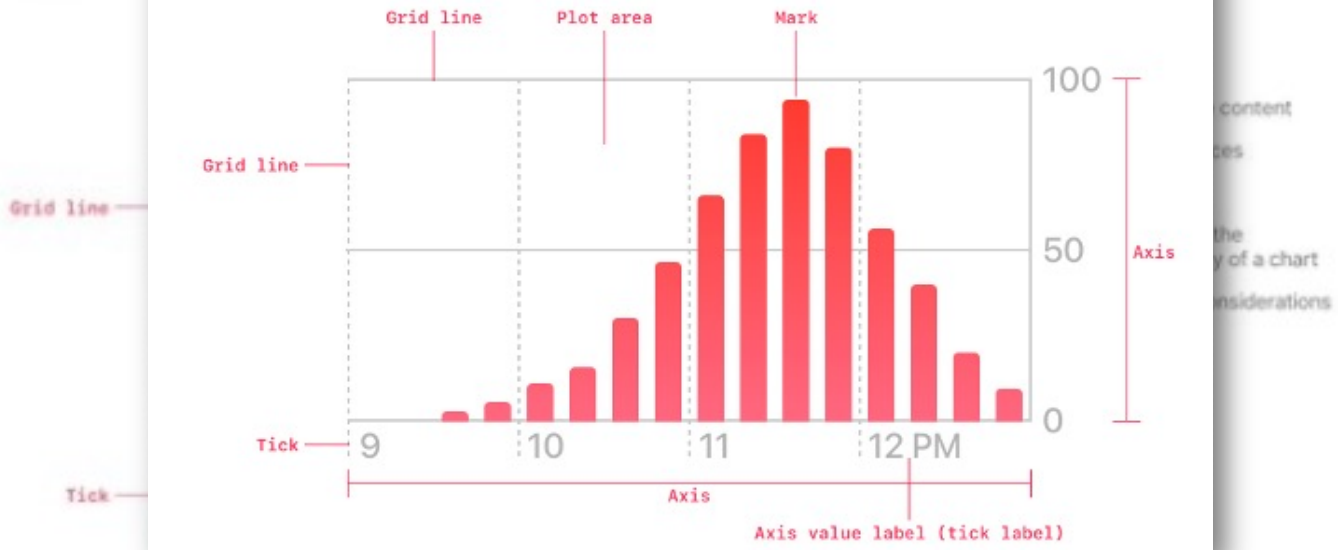
The image is a screenshot of the Apple Human Interface Guidelines (HIG) page for Charts. It features a navigation sidebar on the left with sections like 'Patterns', 'Components', and 'Charts'. The main content area is titled 'Presenting data in a chart can help you communicate information with clarity and appeal.' and includes a bar chart with three bars of increasing height. Below this, there's a section titled 'Charts' with the sub-heading 'A chart helps you communicate data in a graphical, approachable way.' and another bar chart with multiple bars. To the right, there are several key principles: 'Use a chart when you want to highlight important information about a dataset', 'Keep a chart simple', and 'Make every chart in your app accessible'. Each principle is accompanied by a brief explanation and a small icon. The page also includes a 'Support platforms' section and a 'Techniques' section at the bottom.

Here is an example from a recently released set of guidelines by Apple. These guidelines include patterns for designing visualizations.

through graphical representations of data and visual descriptions. In addition to the visual descriptions you display, it's crucial to provide both accessibility labels that describe chart values and components, and accessibility elements that help people interact with the chart. For guidance, see [Enhancing the accessibility of a chart](#).

A chart conveys information about a dataset and its components.

A chart comprises several graphical elements that depict the values in a dataset and convey information about them.



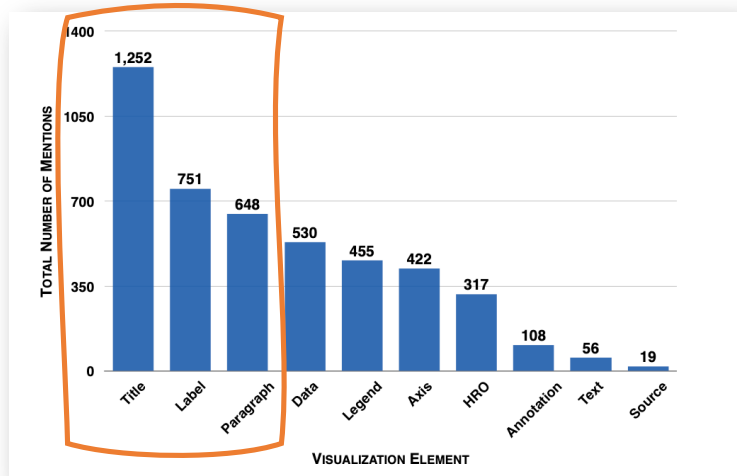
Let's zoom in to see the anatomy of a chart. This is nicely designed. But where are the guidelines for the use of language? Where is the title? The role of context is mentioned elsewhere, but not on the specifications for the chart.

This is a broader trend within our field, including many of our textbooks. Of course, a book about viz is not about language, but what I want to argue here is that text should get higher billing in our field.

## Beyond Memorability: Visualization Recognition and Recall

TVCG'16

Michelle A. Borkin\*, *Member, IEEE*, Zoya Bylinskii\*, Nam Wook Kim, Constance May Bainbridge,  
Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, *Senior Member, IEEE*, and Aude Oliva



Titles got the longest fixations during encoding and were the most likely to be mentioned during recall.

Title, Label, Paragraph received most mentions

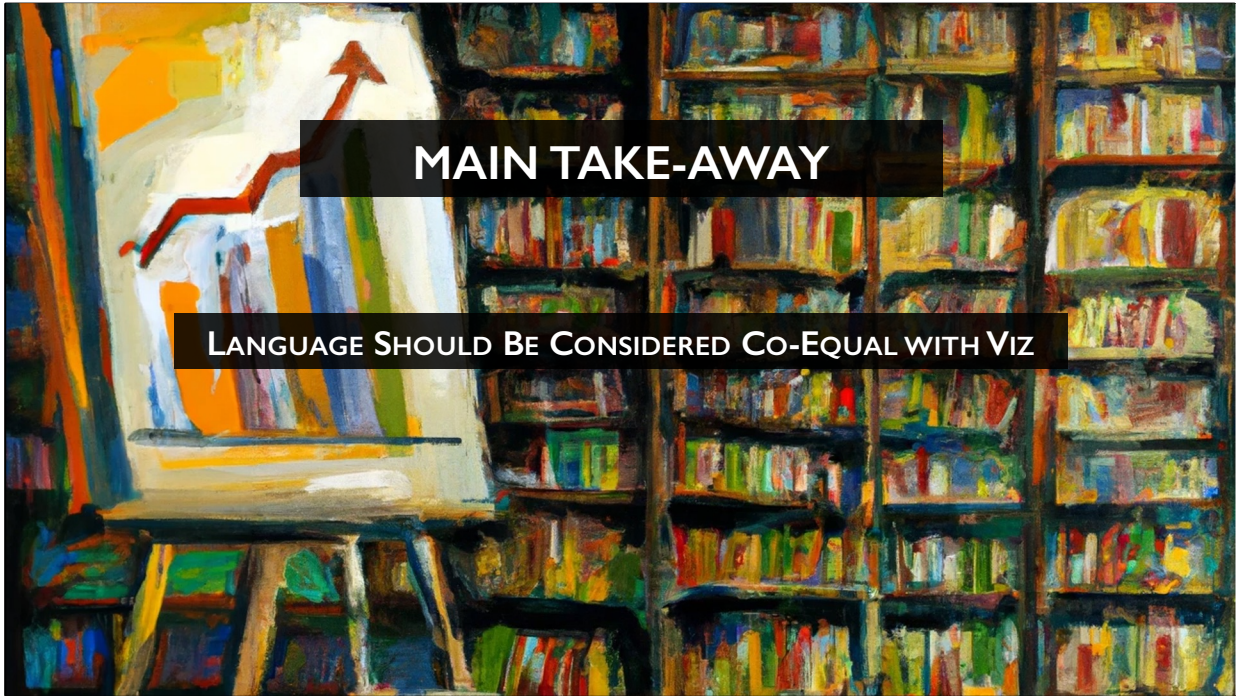
We know from the seminal work of Borkin et al. that the language component of visualization is key. They found that when comparing the memorability of a huge number of designs, the written text is the most memorable part of the visualization.



## MAIN TAKE-AWAY

The good news is that recently there has become a rise of interest in the role of language in information visualization. In fact, this is why I have been asked to give this keynote!

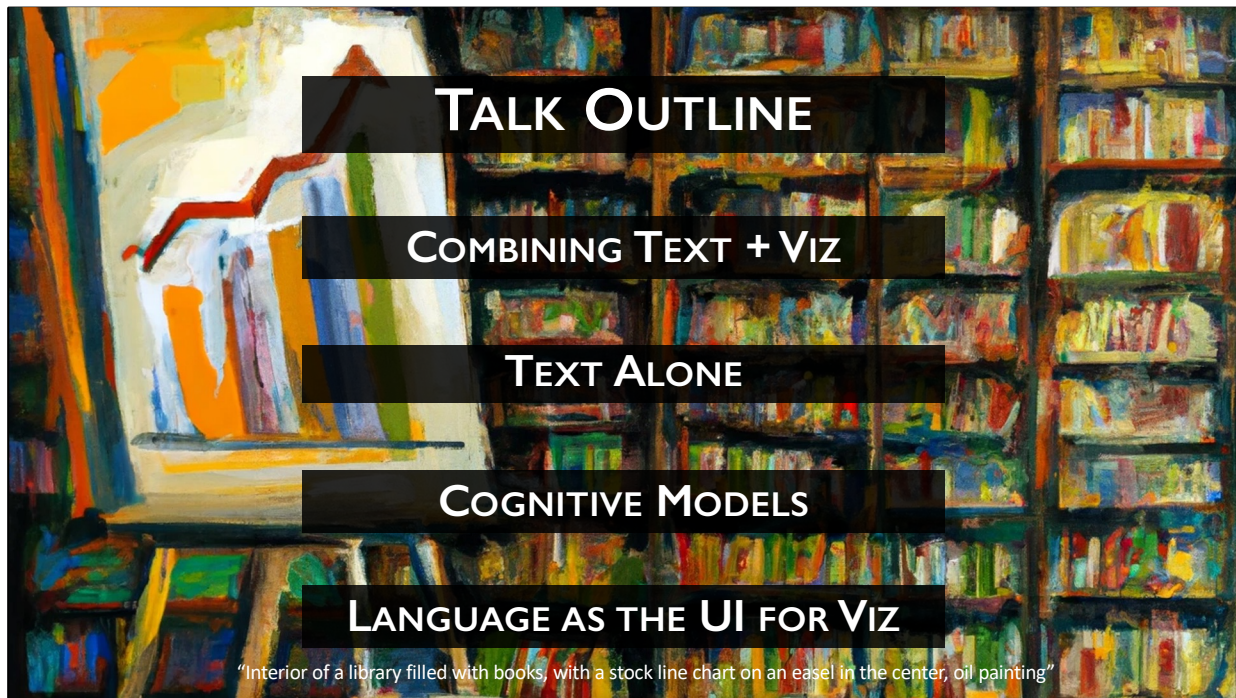
There is now a workshop on natural language and visualization here at Vis, being held for the second time. I see my role in this talk as bringing some of the questions and the results to a wider community.



## MAIN TAKE-AWAY

LANGUAGE SHOULD BE CONSIDERED CO-EQUAL WITH VIZ

**If you don't remember anything else from this talk, here is my intended Main Takeaway:** Language should be considered as co-equal with visualization.



With that preamble, here is the outline of the rest of my talk:

**Combining Text + Viz:**

As I mentioned, the viz community has under-studied what the actual text should be when combining text with viz. I'll discuss some research that has been done on this question, including interaction with cognitive linguistics.

**Text stand-alone:** I'd like to suggest that we consider comparing against a baseline of no visualizations at all – a baseline of expressing the same information in language – written or spoken. This should be a standard practice, as there is a significant minority of people who tend to prefer no visualizations in many cases.

**Cognitive models:** Our toolkit for cognitive and perceptual models include the psychology of reading language – especially when we embed visualizations within written text. I argue that we do not have sufficient cognitive theories of how combinations of language

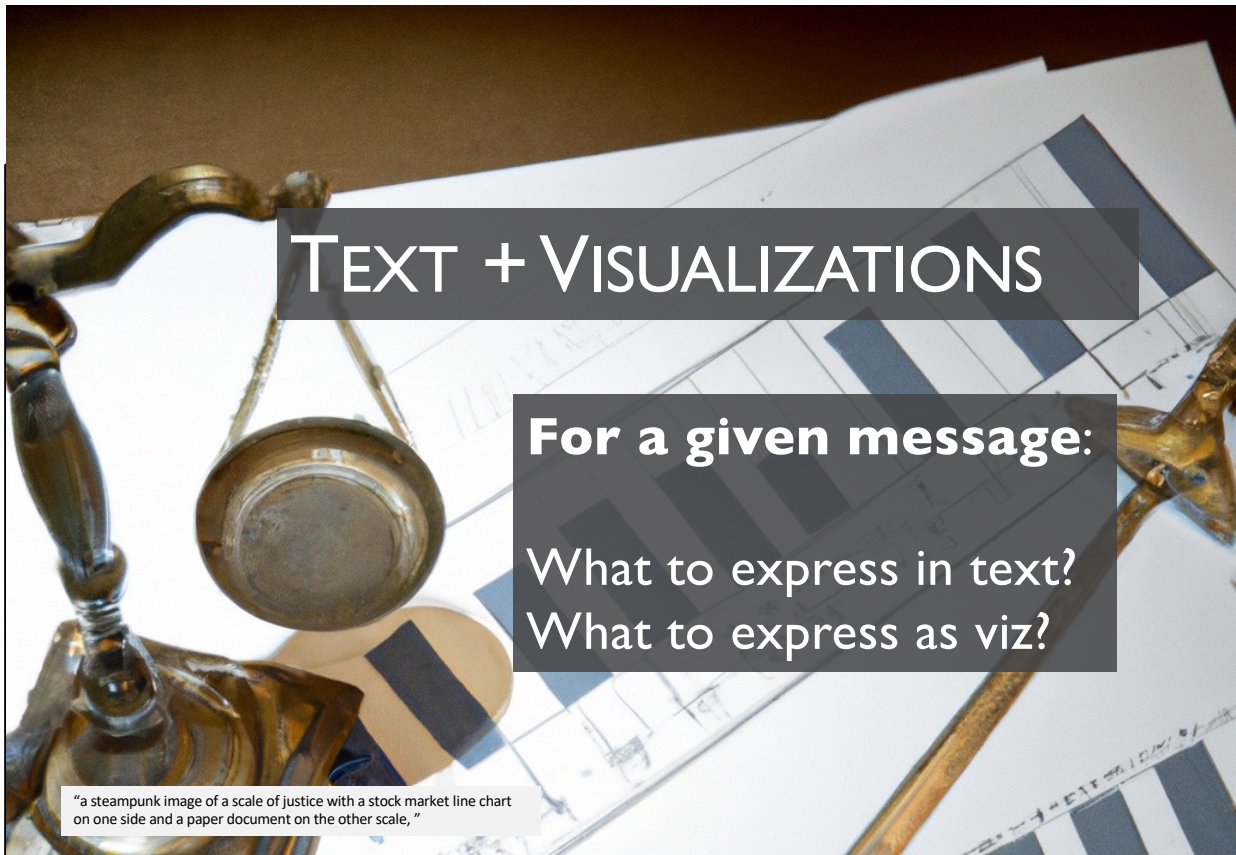
and visualizations are read, perceived, understood.

**Language as the UI for Viz:** I'd like to reflect on the rather spectacular advances that are happening in natural language processing, and speculate as to what this means for infoviz.

I should note that I am using language and text interchangeably here; much of it also can apply to spoken language.

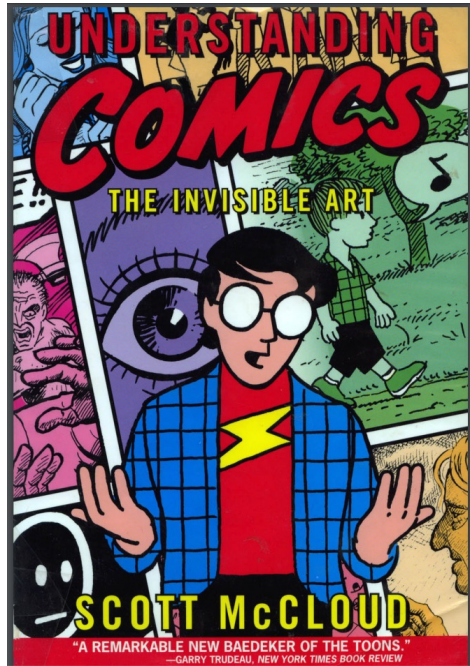
I also note that you will see several strange images in this talk; they are generated by one of these advances, the DALL-E2 system. You can see the text prompt at the bottom of each image.





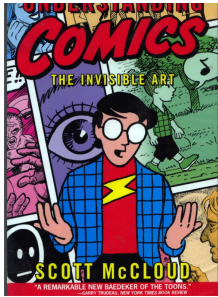
In our subject of combining text and visualizations together, I want to first address this question:

For a given message that you want to convey, what parts of it should be expressed as text, and what parts should be addressed visually? How do the two play off of one another?



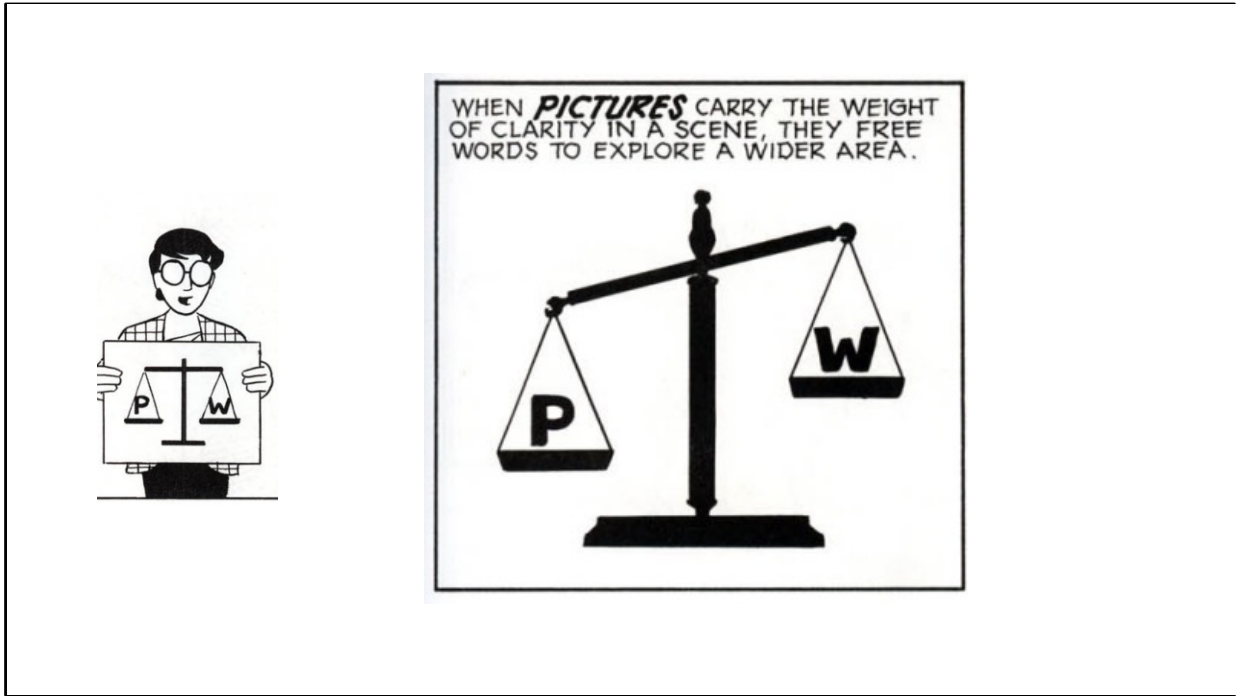
No viz keynote is complete without a reference to Scott McCloud's brilliant book, *Understanding Comics*.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.



In chapter 5, he considers the relationships between the W: words, and the P: pictures, in his nomenclature. He has a series of images showing these two as occupying two sides of a scale.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.



He then introduces a running example of a comic.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.

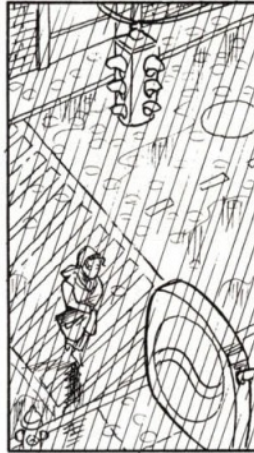


In the first view, it has only pictures, no words. What can this convey? What can it not convey?

His point is since the image is showing the action, the scene, and the mood, the words are free to do something else.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.

WHEN *PICTURES* CARRY THE WEIGHT OF CLARITY IN A SCENE, THEY FREE WORDS TO EXPLORE A WIDER AREA.

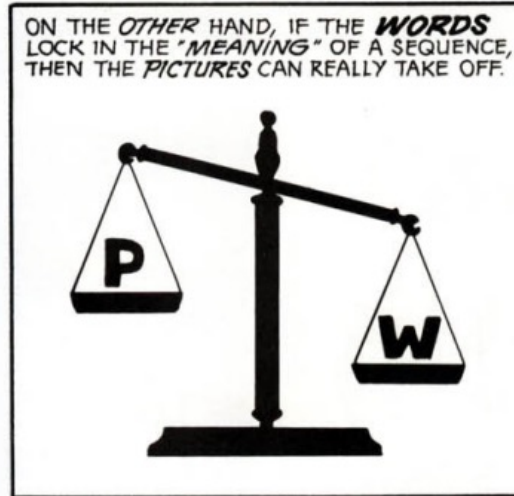
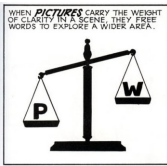


I MAY BE ALONE LIKE THIS FOR A VERY LONG TIME.



In this case, the words convey something about the internal state of the character, which the image cannot do.

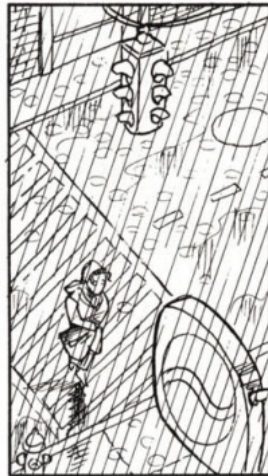
McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.



Now let's consider the reverse case where the words carry the weight of describing the scene and the action.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.

ON THE *OTHER* HAND, IF THE **WORDS**  
LOCK IN THE "*MEANING*" OF A SEQUENCE,  
THEN THE *PICTURES* CAN REALLY TAKE OFF.



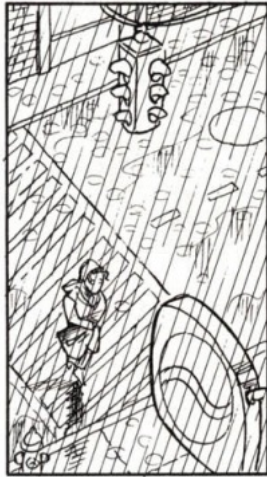
I CROSSED THE  
STREET TO THE  
CONVENIENCE STORE.  
THE RAIN SOAKED  
INTO MY BOOTS.

In this version, instead of the picture showing the scene and action, the words state them. You can see that the words are saying what the image conveyed.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.



ON THE OTHER HAND, IF THE **PICTURE**  
DOESN'T DO THE "WORK" OF A SCENE,  
THEN THE PICTURE CAN REALLY TAKE OFF!



I CROSSED THE  
STREET TO THE  
CONVENIENCE STORE.  
THE RAIN SOAKED  
INTO MY BOOTS.



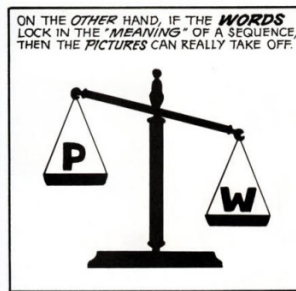
I CROSSED THE  
STREET TO THE  
CONVENIENCE STORE.  
THE RAIN SOAKED  
INTO MY BOOTS.



Since the words describe the scene, the image can zoom in to show just a piece of the action.

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.

1. How much text should a chart have?
2. What should it say?
3. Where should it go?



We can apply these observations from Scott McCloud to research questions about text + viz:

How much text should appear on a visualization?

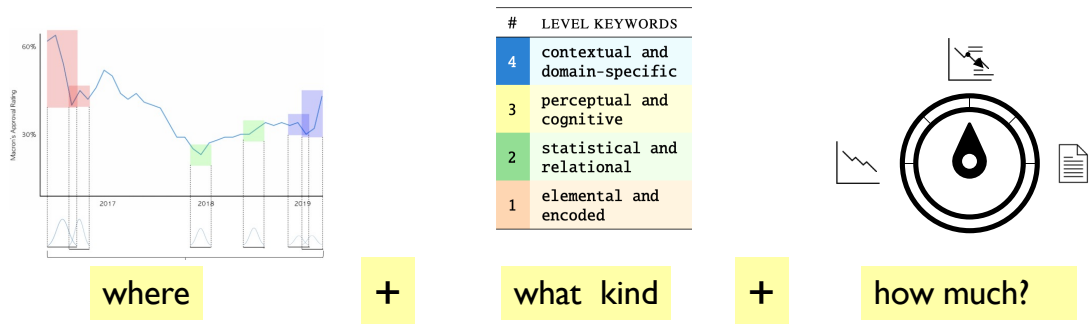
What should it say?

Where should it be placed?

And how do the visual and the language components interact?

McCloud, *Understanding Comics: The Invisible Art*, William Morrow Press, 1993.

1. How much text should a chart have?
2. What should it say?
3. Where should it go?



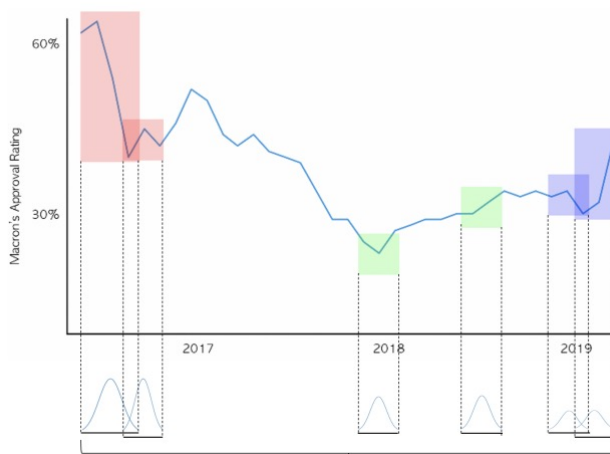
To answer these questions, we first need ways of measuring the where, the what, and the how much. I'm going to describe three pieces of prior research that each address one of these questions; the last puts all three together. These examples also tell us important information about how language is used in visualizations.

### Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts

Dae Hyun Kim

Vidya Setlur

Maneesh Agrawala



Research question:

To what extent do captions influence what people take away from charts?

What happens when the text highlights parts that are not most visually salient?

This piece of work asked: To what extent do captions influence what people take away from charts?

What happens when the text highlights parts that are not most visually salient?

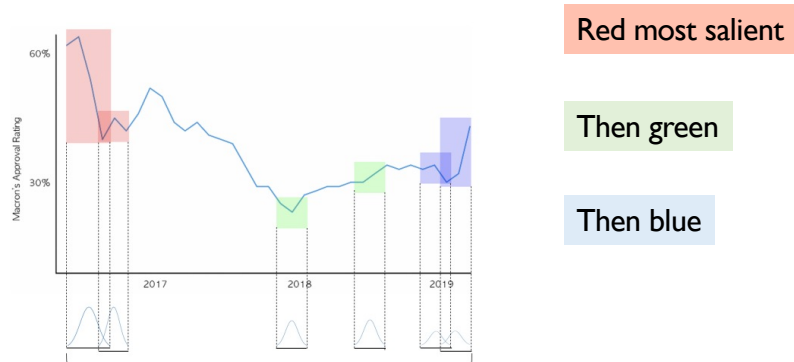
### Towards Understanding How Readers Integrate Charts and Captions: A Case Study with Line Charts

Dae Hyun Kim

Vidya Setlur

Maneesh Agrawala

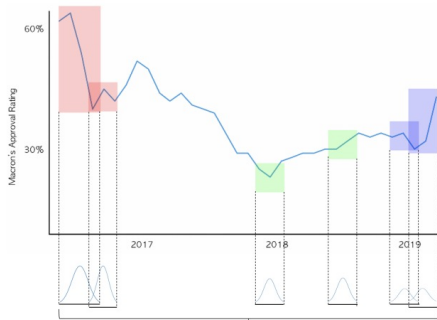
First, identify the most visually salient regions of a chart



This work developed a method to determine which parts of a univariate line chart are most visually salient: here shown in red, followed by green and blue.

The salience computation gives us information about the “where” to experiment with placing text, used in a different study I will discuss.

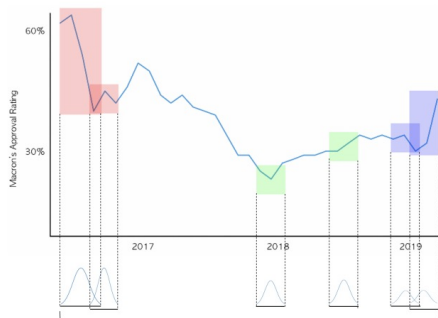
Second, write captions, one for each salient region.



“Macron’s approval rating steeply dropped between June and Aug 2017.”

They next created captions that corresponded to each of these salient parts of the line chart.

## Findings: the effect of caption depends on visual saliency



When the caption mentions a salient feature, reader takeaways more consistently mention the feature.

When the caption mentions a less salient feature, reader takeaways are more likely to mention the most salient **feature** than the what was described in the caption.

“Macron’s approval rating between June and July 2017”

### How Do Captions Affect Visualization Reading?

Hanxiu ‘Hazel’ Zhu \* †

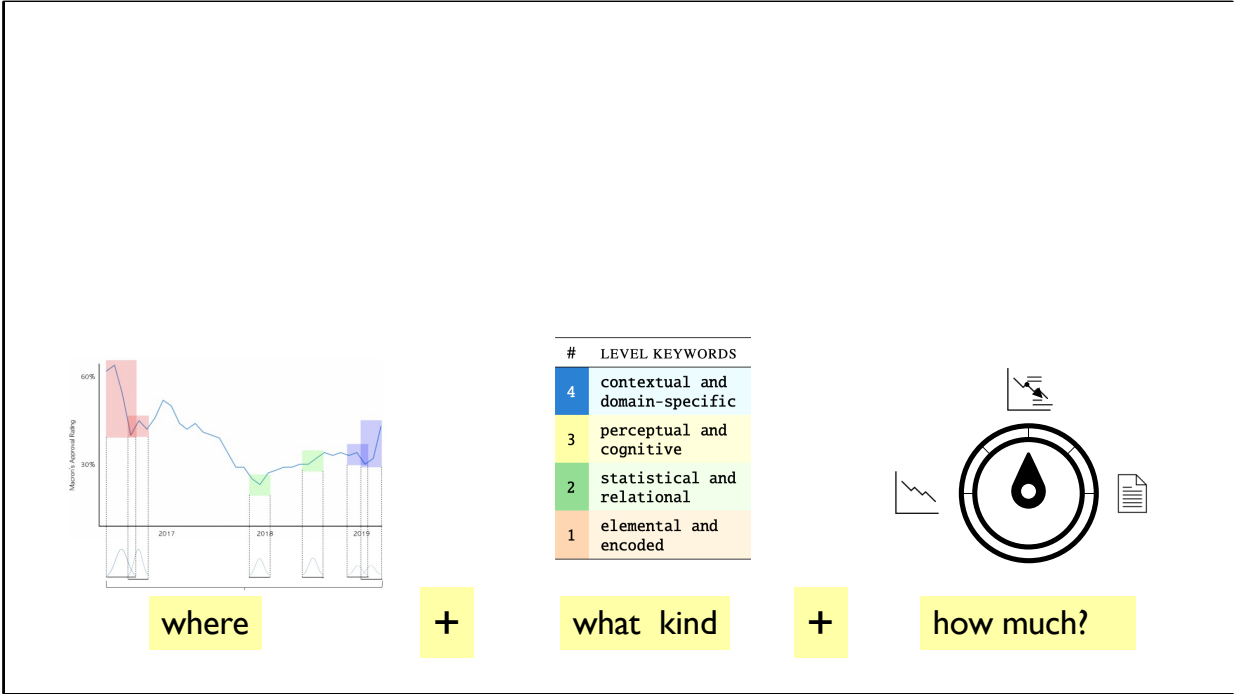
Shelly Shiyong Cheng \* †

Eugene Wu ‡

They found that when the captions referred to parts of the chart that were not the most visually salient, people recalled the parts associated with the less visually prominent parts of the chart. But if the caption referred to something not visually important at all, then what they recalled was more influenced by the chart.

These findings suggest that there is a complex relationship between the effects of the visuals versus the effects of the textual. There are a lot of other really interesting findings that I don’t have time to discuss here.

See also the paper by Zhu et al. that appeared at a workshop on Monday that extends this work.



The next ingredient is what kind of semantics, or meaning, the text should have.



# Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content

Alan Lundgard and Arvind Satyanarayan

TVCG'22

#	LEVEL KEYWORDS
4	contextual and domain-specific
3	perceptual and cognitive
2	statistical and relational
1	elemental and encoded

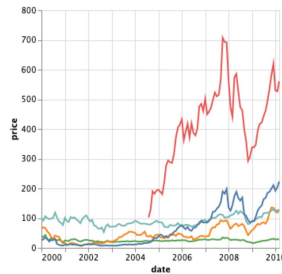
In this work, the experimenters were looking at another topic of great relevance to the text + viz question: what kind of language is preferred by blind and low vision people versus sighted people?

## Accessible Visualization via Natural Language Descriptions: A Four-Level Model of Semantic Content

Alan Lundgard and Arvind Satyanarayan

TVCG'22

#	LEVEL KEYWORDS
4	contextual and domain-specific
3	perceptual and cognitive
2	statistical and relational
1	elemental and encoded



[2] GOOG has the greatest price over time. MSFT has the lowest price over time. [3] Prices of particular Big Tech corporations seem to fluctuate but nevertheless increase over time. Years 2008-2009 are exceptions as we can see an extreme drop in prices of all given corporations. [4] The big drop in prices was caused by financial crisis of 2007-2008. The crisis culminated with the bankruptcy of Lehman Brothers on September 15, 2008 and an international banking crisis. [5] At the beginning of 2008, every of this stock price went down, likely due to the financial crisis. [6] Then they have risen again and dropped again, more so than previously.

The experimenters had participants write descriptions of visualizations. They then open coded the text, and found four levels of semantics. The lowest level names the components of the chart, and the highest level describes external contextualizing information.

These semantic levels for visualization text are the “what” for our question.

Accessible Visualization via Natural Language Descriptions:  
A Four-Level Model of Semantic Content

Alan Lundgard and Arvind Satyanarayan

TVCG'22

Compared views of blind/low vision (BLV) and sighted readers.

High-level language (L3 & L4)

Majority of **BLV**  
readers **opposed**

Majority of **sighted**  
readers **favored**

Low-level language (L1 & L2)

Majority of **BLV**  
readers **favored**

Majority of **sighted**  
readers **opposed**

Something very interesting that they found is that BLV people prefer different kinds of textual information than those who are sighted.

In particular, for high level expressions of language, the majority of BLV readers opposed this expression, while it was favored by the majority of sighted readers, and the converse was true for low-level language.

These results are important for the design of alternative text for visualizations.

They also give us the “what” for what kind of text should appear with visualizations.

Accessible Visualization via Natural Language Descriptions:  
A Four-Level Model of Semantic Content

Alan Lundgard and Arvind Satyanarayan

TVCG'22

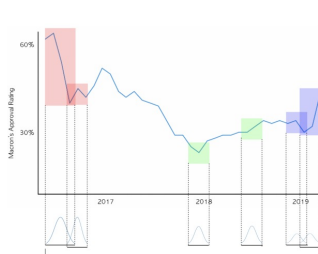
“Natural language should be seen as co-equal w/viz”

I also note that they had a sentence in their paper stating: “Natural language should be seen as co-equal with visualization.” This may strike some as a radical statement, but it aligns well with the point of this talk.

## Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts

Chase Stokes, Vidya Setlur (*Member*), Bridget Cogley, Arvind Satyanarayan, and Marti A. Hearst

1. How much text should a chart have?
2. What should it say?
3. Where should it go?



where

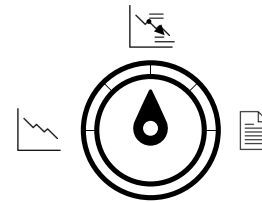
+

what kind

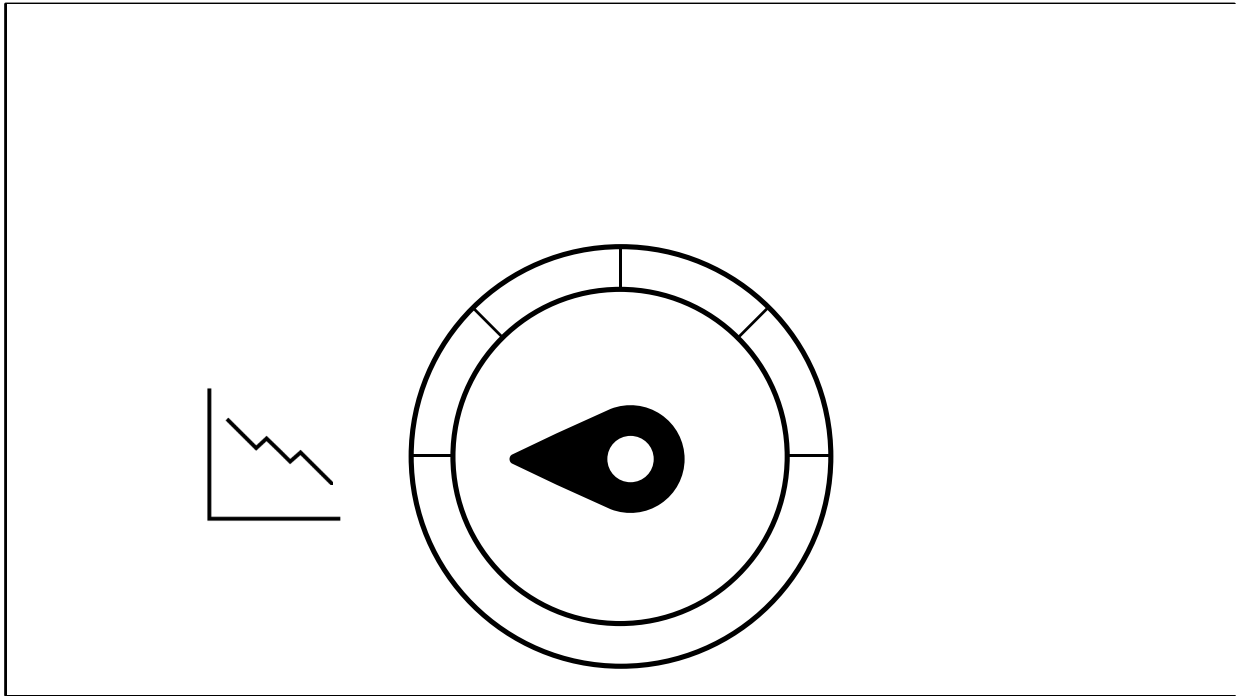
+

how much?

#	LEVEL KEYWORDS
4	contextual and domain-specific
3	perceptual and cognitive
2	statistical and relational
1	elemental and encoded

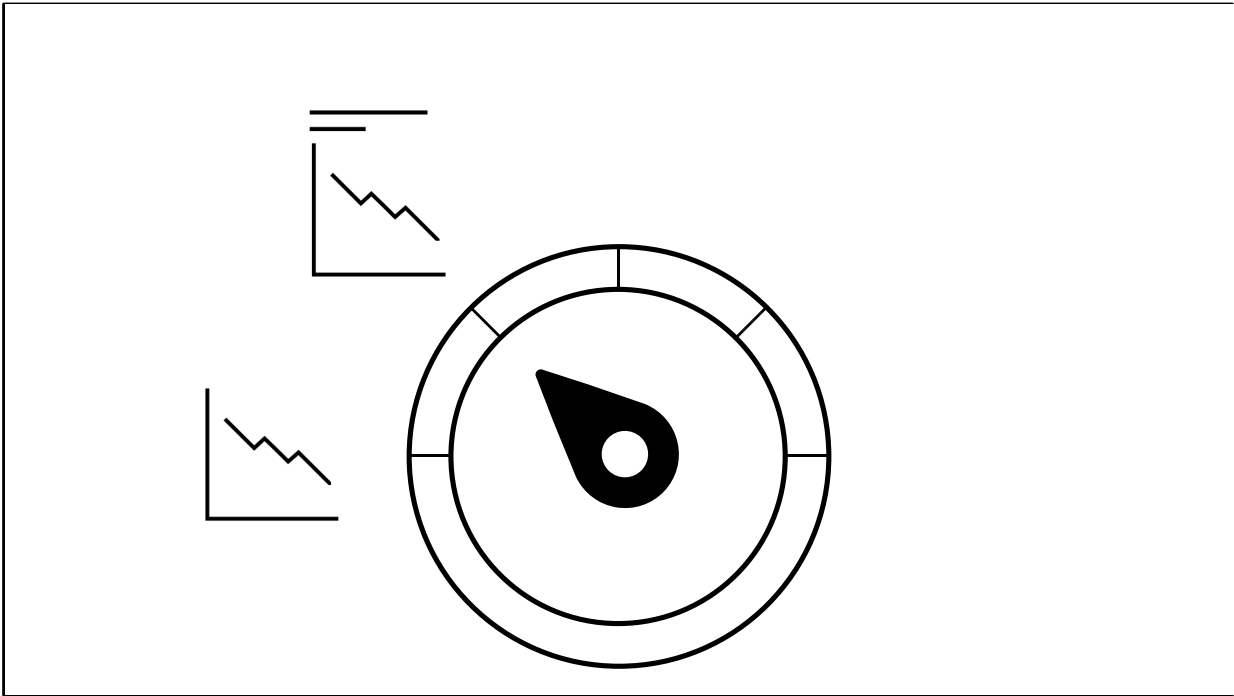


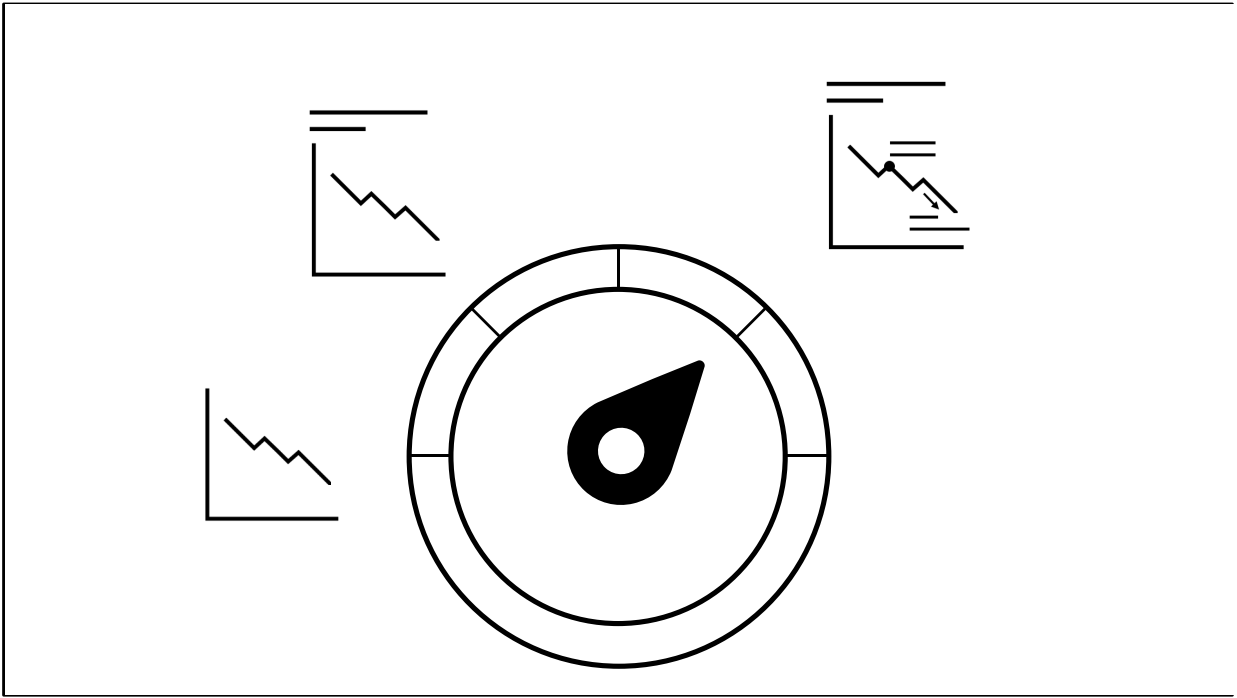
Armed with this “where” and “what”, we can now proceed to ask: “how much”? As in, how much text is too much for annotation as an overlay on a chart. This work is being presented in detail in this conference on Friday by my PhD student Chase Stokes.



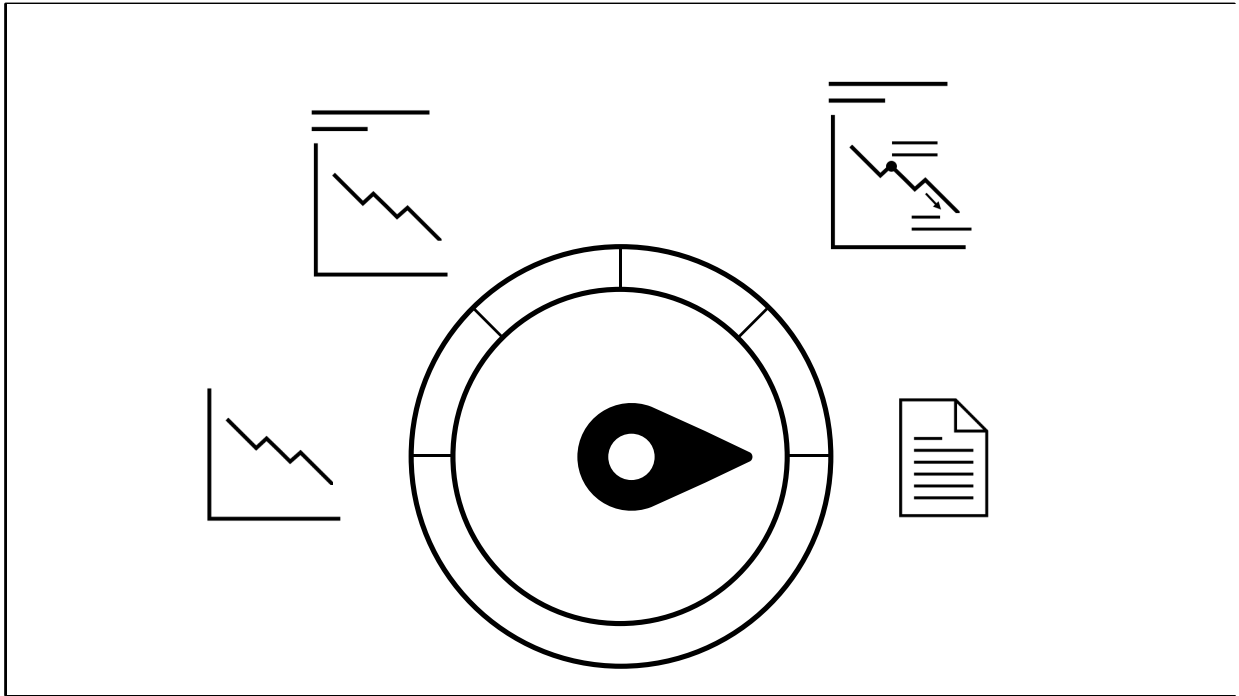
We varied the design from all chart and no text, to a bit more text, and still more, all the way to no chart at all.

.









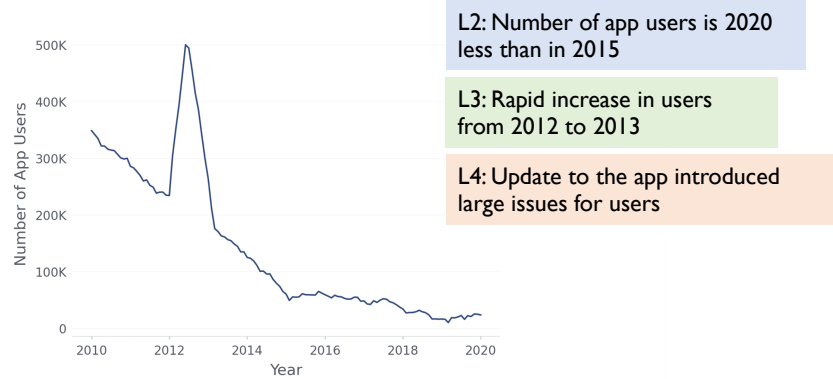
We varied the design from all chart and no text, to a bit more text, and still more, all the way to no chart at all.

First, identify the most visually salient regions of a chart



To do the study, we created a wide range of stimuli. We created the charts by first finding the visually salient components, ...

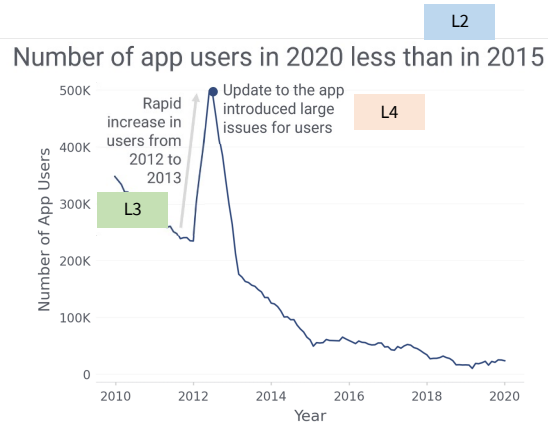
Next, write text at each semantic level of description



Next, we wrote text at each semantic level for each visually salient region.

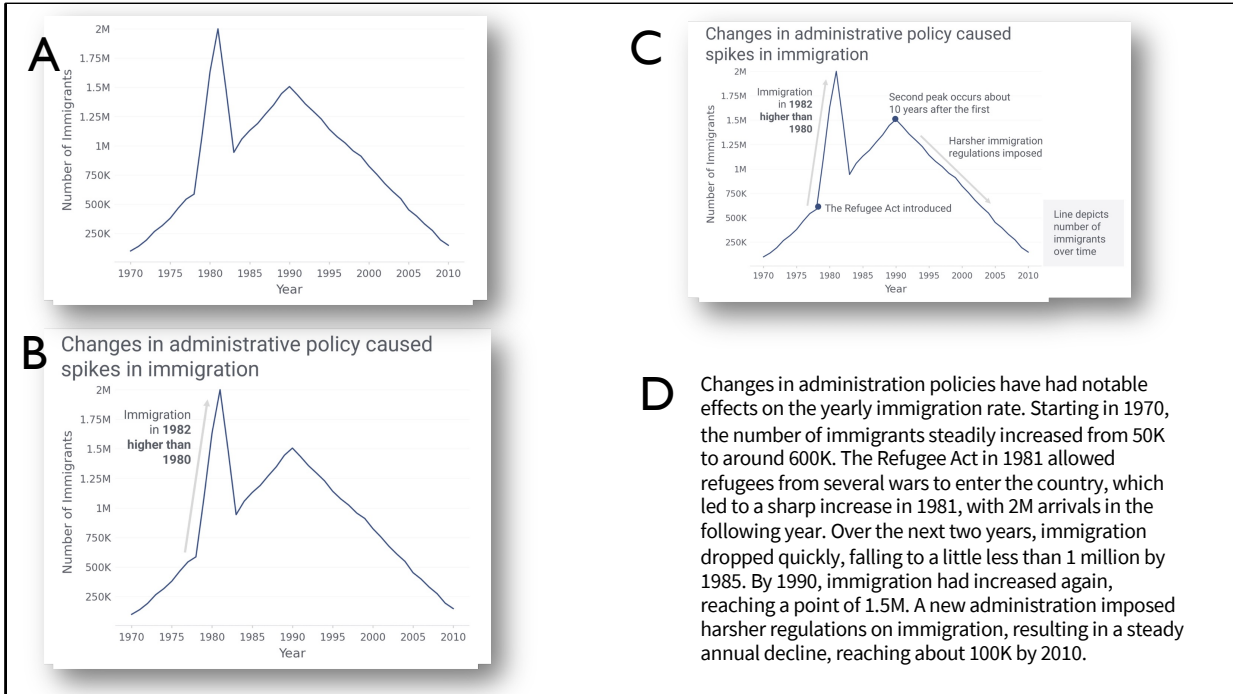
## Then, place the text into different locations

(title, annotation by salient region, etc.)



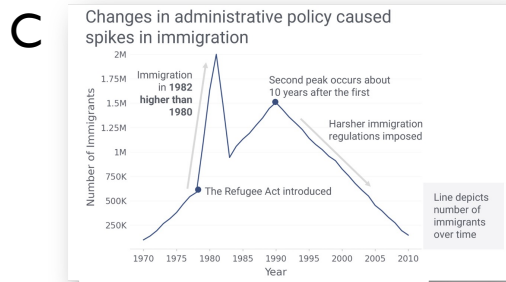
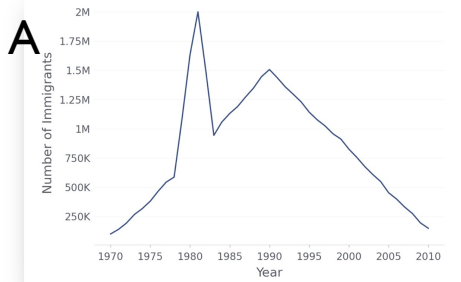
And then we placed the different types of text at different positions on the charts, varying them appropriately for a controlled experiment.

For example, in this case, the title is labeled with Level 2, the peak with L3 language, and the high level contextual information is positioned as a point, accompanied by a blue dot.

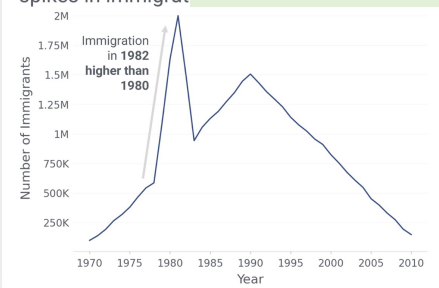


We assessed these designs both with preference questions, and with how well people took information away from the charts.

I want to do a quick quiz here and show you the four chart types that we compared in terms of user preferences, ranging from almost no text to the maximum amount for our study. We had crowdworker participants look at the designs one at a time, and then all at once, and rank order them by preference. Which chart type was most preferred? A B C D.



**B** Changes in administrative policy caused spikes in immigration

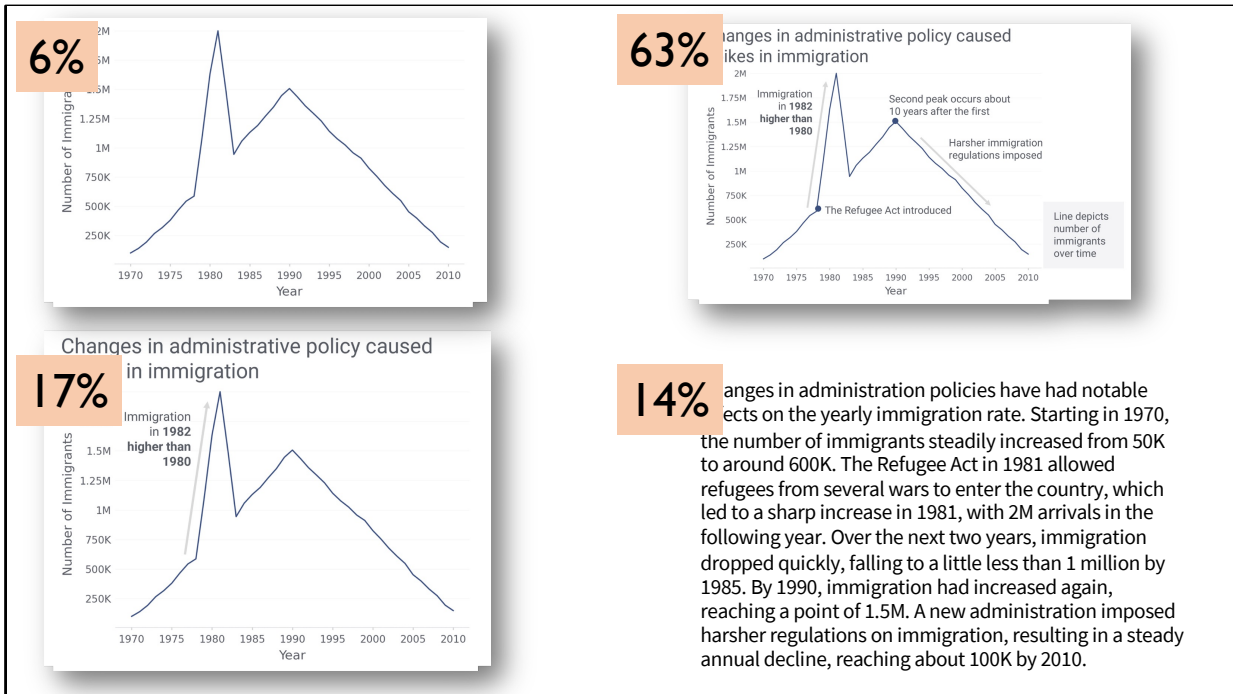


Which was ranked #1?

**D**

Changes in administrative policies have had notable effects on the yearly immigration rate. Starting in 1970, the number of immigrants steadily increased from 50K to around 600K. The Refugee Act in 1981 allowed refugees from several wars to enter the country, which led to a sharp increase in 1981, with 2M arrivals in the following year. Over the next two years, immigration dropped quickly, falling to a little less than 1 million by 1985. By 1990, immigration had increased again, reaching a point of 1.5M. A new administration imposed harsher regulations on immigration, resulting in a steady annual decline, reaching about 100K by 2010.

Which was ranked #1 most often?



It turns out that more text was not only most preferred, but also yielded the best outcomes for the takeaways.

In a paper in the NLViz workshop that took place here on Sunday, called “Why more text is often better”, we looked more deeply at the reasons for these findings, as expressed by participants in their comments. We found that although text can at first glance make the chart appear more cluttered, in actuality, this extra context was helpful. Overall, more context for the chart was seen as helpful by a majority of participants. This is not to say that any text works well; very likely if we had included irrelevant text, the reception for those charts would have been less warm. As it was, some participants expressed that they did not like the text when what it stated was redundant with the chart, such as labeling a point as a maximum.

## Striking a Balance: Reader Takeaways and Preferences when Integrating Text and Charts

Chase Stokes, Vidya Setlur (*Member*), Bridget Cogley, Arvind Satyanarayan, and Marti A. Hearst

### FINDINGS

1. How much text should be added?
2. Where should it go?
3. What should it say?

1. Rather than minimizing ink, annotate charts with relevant text.

2. The best position for annotation depends on the type of semantic level.

3. The best semantic level depends on the intended takeaway.

For the take-aways portion of the study, our findings were:  
For how much: use relevant text, don't worry so much about clutter  
For where: the position depends on the type of semantic level  
For what: the best semantic level depends on the message

We have only scratched the surface with this question; so much more needs to be done.





## COMPARISONS

"a line of buildings of different heights with a ruler next to each, in the style of Pissarro"

How do we compare things? This question has been studied both in linguistics and in visualization. The two areas conceptualize this problem differently. Now there is some interesting work looking at how to bring the two together, and this makes for a good case study.

# COMPARISONS

## LINGUISTICS

“The comparative is a difficult structure to process for both syntactic and semantic reasons. Syntactically the comparative is extraordinarily **diverse**.” – Friedman, 1989

## VISUALIZATION

“Supporting comparison is a common and **diverse** challenge in visualization.” – Gleicher 2017

Here are two different statements about comparisons; one is from linguistics and the other from visualization.

Friedman states that “The comparative is a difficulty structure to process for both syntactic and semantic reasons. Syntactically, the comparative is extraordinarily diverse.

On the visualization side, Gleicher writes: Supporting comparison is a common and diverse challenge in visualization.

I’d like to draw your attention to the emphasis on how both fields see this construct as being expressed in a diverse manner, which makes it more challenging.

# COMPARISONS

## LINGUISTICS

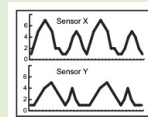
### Examples of Ranking Comparison:

I felt more comfortable with XTi and some of my friends felt more comfortable with D80 .

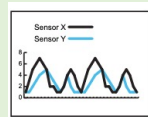
On the other hand I actually prefer the D80 handling with smaller lenses , which is what 's on my camera 80 % of the time.

Wiltrud & Kuhn. "A Corpus of Comparisons in Product Reviews." *LREC* (2014).

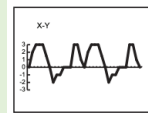
## VISUALIZATION



Juxtaposition



Superposition



Explicit Encoding

In linguistics the focus is on the variation in expression for similar concepts. Here I show different ways of expressing comparisons about cameras. Note the variation just with these examples. What is being compared to what? How much and to what degree is one thing being said to be different than the other. If you try to image how a program might determine the answers to these questions, you can see how indirect the information can be in language.

On the other hand, in the visualization literature, we focus a great deal on how to visually present comparisons when we know what the entities are that are being compared, and what the relationships are. In viz we are often focusing on how to show those relationships, and often on how to make them scale. One sentence of language can only compare a few things at a time.

# COMPARISONS: COMBINING LINGUISTICS WITH VIZ

Example: How do people judge tall vs not tall?



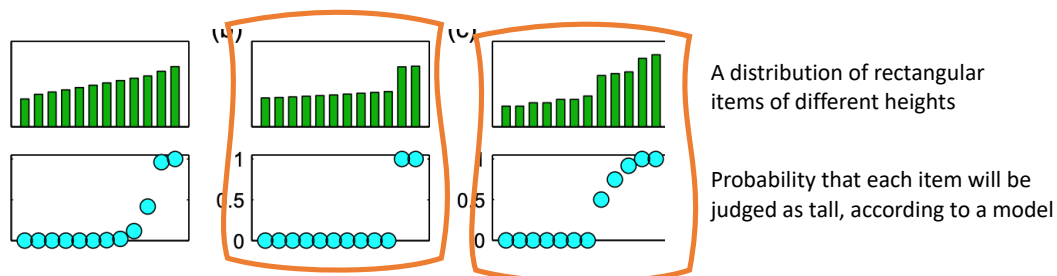
Let's take an example of using results from linguistics to inform us about visualization. The example is: how do people decide what the meaning is of tall? What is tall vs not tall?

## How Tall Is *Tall*? Compositionality, Statistics, and Gradable Adjectives

Lauren A. Schmidt<sup>1</sup> (lschmidt@mit.edu), Noah D. Goodman<sup>1</sup> (ndg@mit.edu), COGSCI 2009  
David Barner<sup>2</sup> (barner@ucsd.edu), and Joshua B. Tenenbaum<sup>1</sup> (jbt@mit.edu)

Example: How do people judge tall vs not tall?

A: It depends on the shape of the distribution of heights.



The answer according to cognitive linguistics is: it depends on the distribution of the data points.

Here we see two rows of images. The top row shows hypothetical distributions of heights among a set of things being compared.

The bottom row shows the probability that a person will judge that item is tall, according to a model based on empirical data about human judgements.

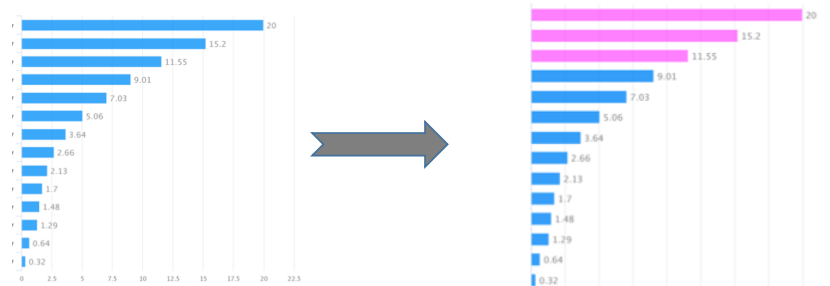
Note there is more agreement on tall vs not-tall for the step function than for the exponential curve.

## Toward Interface Defaults for Vague Modifiers in Natural Language Interfaces for Visual Analysis

Marti Hearst\* Melanie Tory† Vidya Setlur‡ Vis 2019

Use this finding from cognitive linguistics to  
determine which bars to highlight for a viz query.

Query: “Show the heights of the tallest buildings”



When I visited Tableau Research a few years ago, my colleagues and I were interested in how to handle vague modifiers like “tall” in a visualization system like Tableau’s Ask Data that tries to show visualizations in response to user’s written questions.

We used this finding to determine which bars to highlight in response to a query, such as the query “Show the heights of the tallest buildings”

We used this to determine what the defaults should be for an interface that showed answers to such questions. For instance, for the exponential drop off, the cognitive linguistics finding shows us what bars we should highlight, depending on the shape of the curve.



"a digital assistant chat bot lounging by the pool, late afternoon sun, photo realistic"

## Another Comparison Example

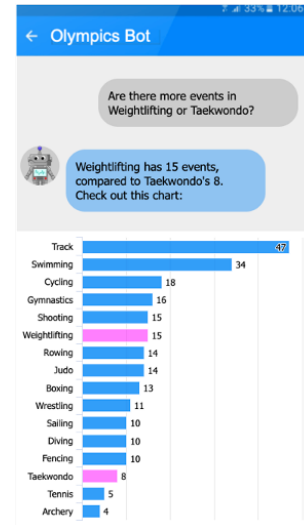
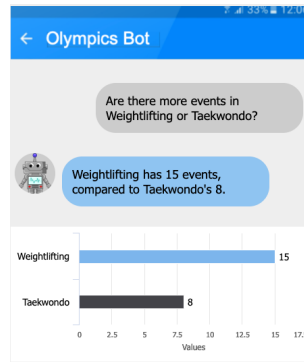
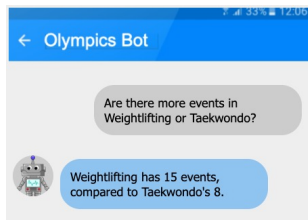
Should mobile digital assistants / chat bots show visualizations in response to comparison questions?

My Tableau colleagues and I did another study to assess a related question: how to show visualizations in a conversational interface for a mobile UI for an intelligent assistant like Siri or Alexa.

## Would You Like A Chart With That? Incorporating Visualizations into Conversational Interfaces

Marti Hearst\*

Melanie Tory†



Which is preferred?

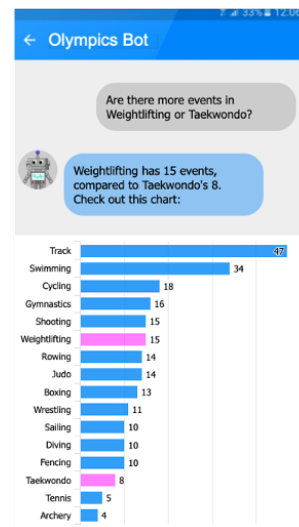
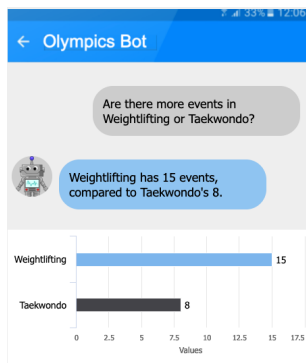
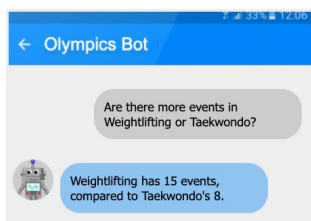
The investigation's goal was to determine what kind of visual context people prefer after they ask a comparison question of the assistant with a simple answer, such as "Which Olympic sport has the tallest players: rowing or swimming?" We wanted to know if people wanted to see a bar chart showing the results only for rowing and swimming, or if they would also like to see the values for other sports for comparison.



## Would You Like A Chart With That? Incorporating Visualizations into Conversational Interfaces

Marti Hearst\*

Melanie Tory†



More context preferred over showing only the data points visually.

We found that people did prefer more context, in the form of more bar charts, as long as the chart did not get too long.

...and, 41% of participants did not want to see viz in chat



However, we also found that 41% of participants did not want to see a chart at all in this context.



## Reasons for Preferences

### For Charts

Preferring more bars over a few:

“At first I thought I wanted a simple answer ... But after seeing [more bars], I realized I really liked getting the answer in context.”

### For Text

Liked the simplicity;  
felt the charts were extraneous:

“It’s precise and gives me enough details without too many details or too few.”  
“It was the easiest to understand and the answer was not overly complicated.”

Some reasons for these differences can be found in quotations from the study.

People who preferred bar charts, preferred having more bars as this gave the answer in context.

People who preferred text said that it is precise, and not overly complicated.

We should take note that we were only talking about a very few data points.

## Comparison Conundrum and the Chamber of Visualizations: An Exploration of How Language Influences Visual Design

Aimen Gaba, Vidya Setlur (*Member*), Arjun Srinivasan, Jane Hoffswell, and Cindy Xiong VIS 2022

“Compare a high-rated book with similar books”

What is the best way to visualize this kind of comparison?

Here is another piece of work looking at comparisons. This is also being presented in the Friday session I mentioned before.

They looked at several forms of linguistic expressions of comparisons, and for each type, they determined which kind of visualization best expressed these comparisons according to crowd workers.

## Comparison Conundrum and the Chamber of Visualizations: An Exploration of How Language Influences Visual Design

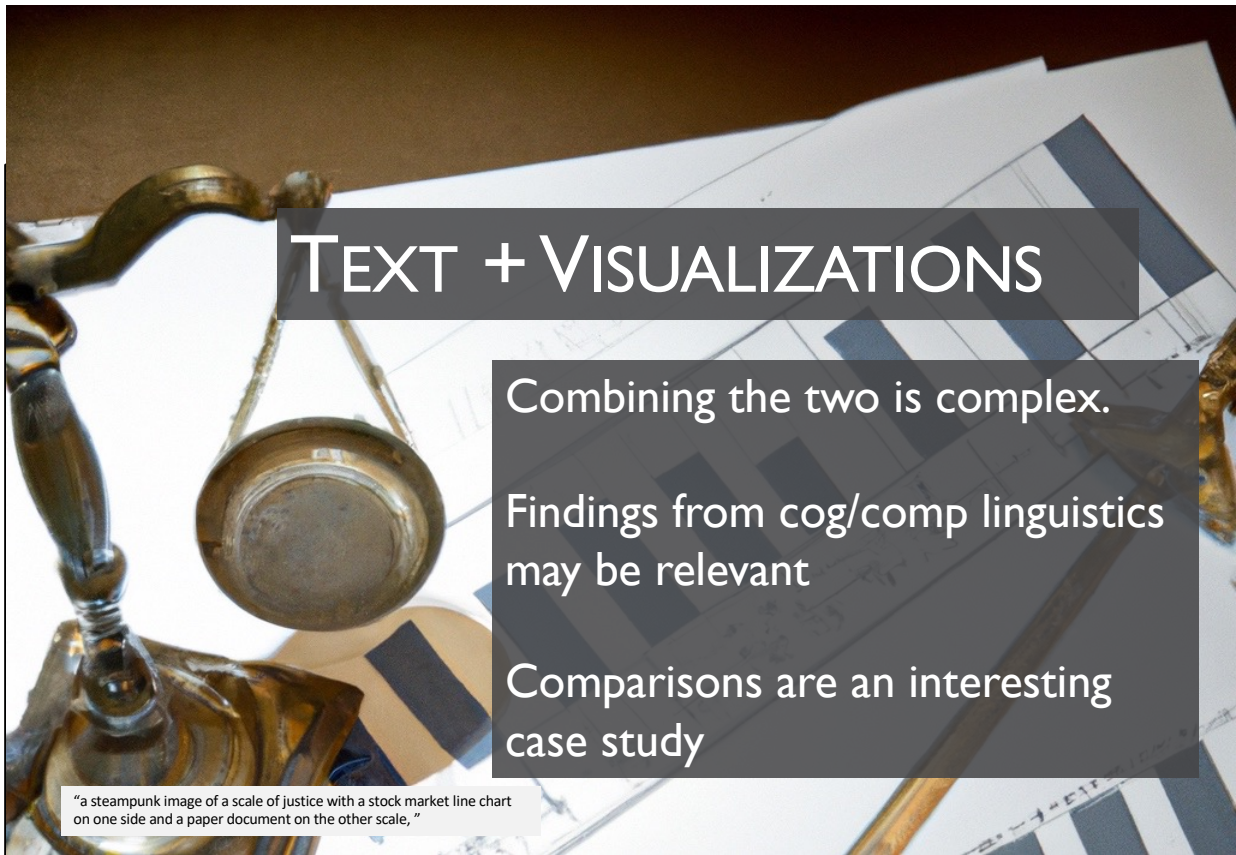
Aimen Gaba, Vidya Setlur (*Member*), Arjun Srinivasan, Jane Hoffswell, and Cindy Xiong VIS 2022

“Compare a high-rated book with similar books”



Here is another piece of work looking at comparisons by Gaba et al. This is also being presented in this conference.

The researchers looked at several forms of linguistic expressions of comparisons, and for each type, they determined which kind of visualization best expressed these comparisons according to crowd workers.



To summarize this section on text plus visualizations:  
Combining the two is complex; some research has been done but more is needed.  
Finding from cognitive and computational linguistics may be relevant for this work, and comparisons are a good case study.

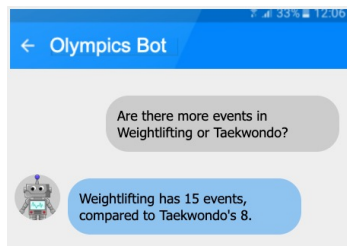


Now I'll talk about the importance of considering text as an option without a chart – text alone.



We've already seen 2 examples of text alone being favored by sizable minorities of participants.

41% did not want to see viz in chat



14% preferred text alone over text + viz

Changes in administration policies have had notable effects on the yearly immigration rate. Starting in 1970, the number of immigrants steadily increased from 50K to around 600K. The Refugee Act in 1981 allowed refugees from several wars to enter the country, which led to a sharp increase in 1981, with 2M arrivals in the following year. Over the next two years, immigration dropped quickly, falling to a little less than 1 million by 1985. By 1990, immigration had increased again, reaching a point of 1.5M. A new administration imposed harsher regulations on immigration, resulting in a steady annual decline, reaching about 100K by 2010.

Are there more examples in the literature?

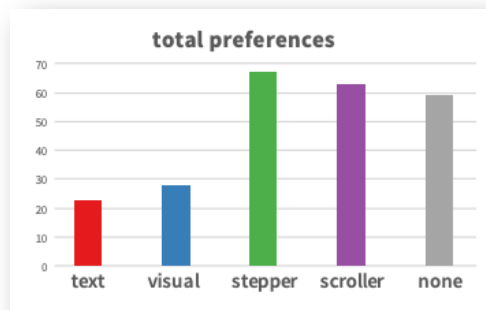
We've already seen two cases in which text without a chart was preferred by a sizable minority of participants. Can we find other examples in the literature?

## Visual Narrative Flow: Exploring Factors Shaping Data Visualization Story Reading Experiences

Eurographics'17

S. McKenna<sup>1,2</sup>, N. Henry Riche<sup>1</sup>, B. Lee<sup>1</sup>, J. Boy<sup>3</sup>, & M. Meyer<sup>2</sup>

### Compared scrollytelly presentation styles



10% prefer text alone

This study compared different ways of presenting a scrollytelly design, including with no visualizations at all. A notable minority of participants said they preferred the condition with no viz's at all.

## **The Curious Case of Combining Text and Visualization**

Eurovis'19

Alvitta Ottley<sup>1</sup>, Aleksandra Kaszowska<sup>2</sup>, R. Jordan Crouser<sup>3</sup>, and Evan M. Peck<sup>4</sup>

### Compared outcomes for Bayesian reasoning

In this work, Ottley et al. wanted to find the best way to help people do Bayesian reasoning.

## The Curious Case of Combining Text and Visualization

Eurovis'19

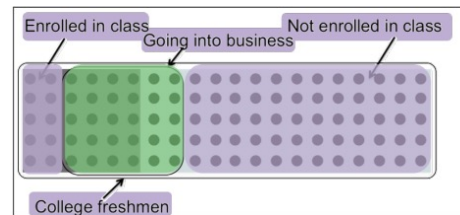
Alvitta Ottley<sup>1</sup>, Aleksandra Kaszowska<sup>2</sup>, R. Jordan Crouser<sup>3</sup>, and Evan M. Peck<sup>4</sup>

### Compared outcomes for Bayesian reasoning

Text alone

There is a total of 100 college freshmen in the population. Out of these 100 freshmen, 30 are enrolled in an introductory entrepreneurship course. Out of these 30 freshmen, 20 plan on going into business after graduation and 10 do not. On the other hand, 70 freshmen are not enrolled in an introductory entrepreneurship course. Out of these 70 freshmen, 10 plan on going into business after graduation and 60 do not.

Viz



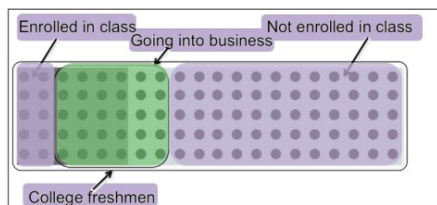
They compared text alone as well as viz alone, as well as a combined condition.

## The Curious Case of Combining Text and Visualization

Eurovis'19

Alvitta Ottley<sup>1</sup>, Aleksandra Kaszowska<sup>2</sup>, R. Jordan Crouser<sup>3</sup>, and Evan M. Peck<sup>4</sup>

There is a total of 100 college freshmen in the population. Out of these 100 freshmen, 30 are enrolled in an introductory entrepreneurship course. Out of these 30 freshmen, 20 plan on going into business after graduation and 10 do not. On the other hand, 70 freshmen are not enrolled in an introductory entrepreneurship course. Out of these 70 freshmen, 10 plan on going into business after graduation and 60 do not.



No significant effect of view on accuracy  
**(text as good as viz)**

When text and viz presented together, participants did not seem to take advantage of the distinct affordances of each.

“We currently do not have ... guidelines that maximize the impact of these two...”

For the point I'm trying to make here, they found that viz was not better than text for this; no significant difference in terms of accuracy.

They found other interesting effects as well, to my earlier points about the combination of viz and text.

They found that when text and viz were presented together, participants did not seem to take advantage of the distinct affordances of each.

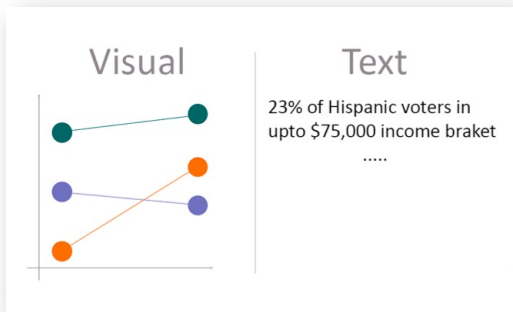
They also note that we do not have sophisticated guidelines for understanding how to combine the two modalities.

## Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data

Yea-Seul Kim

Katharina Reinecke

Jessica Hullman



### Study goal:

Compare how well people recalled data depending on whether they had to predict the data first or not.

Both a viz and a text condition.

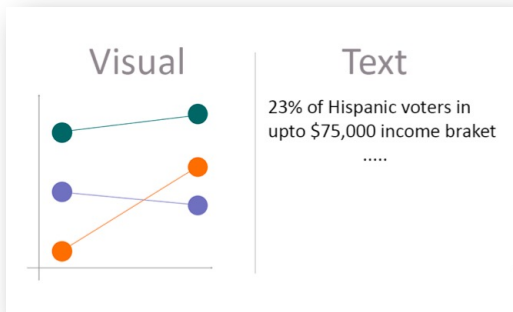
Our last example is the famous “Explaining the Gap” paper. This was motivated by the NYTimes You Draw It tool, and its goal was to compare how well people recalled data depending on whether they had to first predict it or not. They also compared text-only to viz-only conditions.

## Explaining the Gap: Visualizing One's Predictions Improves Recall and Comprehension of Data

Yea-Seul Kim

Katharina Reinecke

Jessica Hullman



Presenting stats as text helps people better recall the specific values later

Viz's help people better recall trends than text displays.

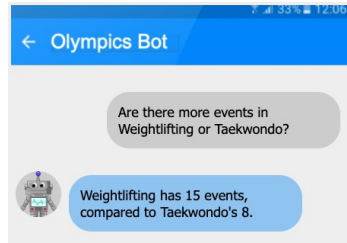
Seeing one's prediction against observed data is helpful only if you are using viz's.

There were three major findings with respect to text. The first was that presenting data as text helps people recall those values better than with a viz.

The other two findings favored vis over text.

The second was that the viz's were better at helping people recall trends. The third was that this prediction effect was only seen for viz, not text.

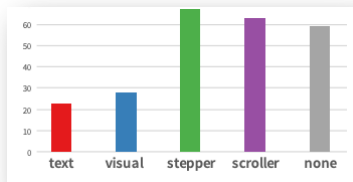
41% did not want to see viz in chat



14% preferred text alone over text + viz

Changes in administration policies have had notable effects on the yearly immigration rate. Starting in 1970, the number of immigrants steadily increased from 50K to around 600K. The Refugee Act in 1981 allowed refugees from several wars to enter the country, which led to a sharp increase in 1981, with 2M arrivals in the following year. Over the next two years, immigration dropped quickly, falling to a little less than 1 million by 1985. By 1990, immigration had increased again, reaching a point of 1.5M. A new administration imposed harsher regulations on immigration, resulting in a steady annual decline, reaching about 100K by 2010.

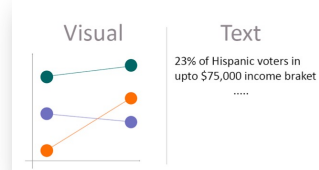
10% prefer text alone



Text alone as accurate as viz; more effective

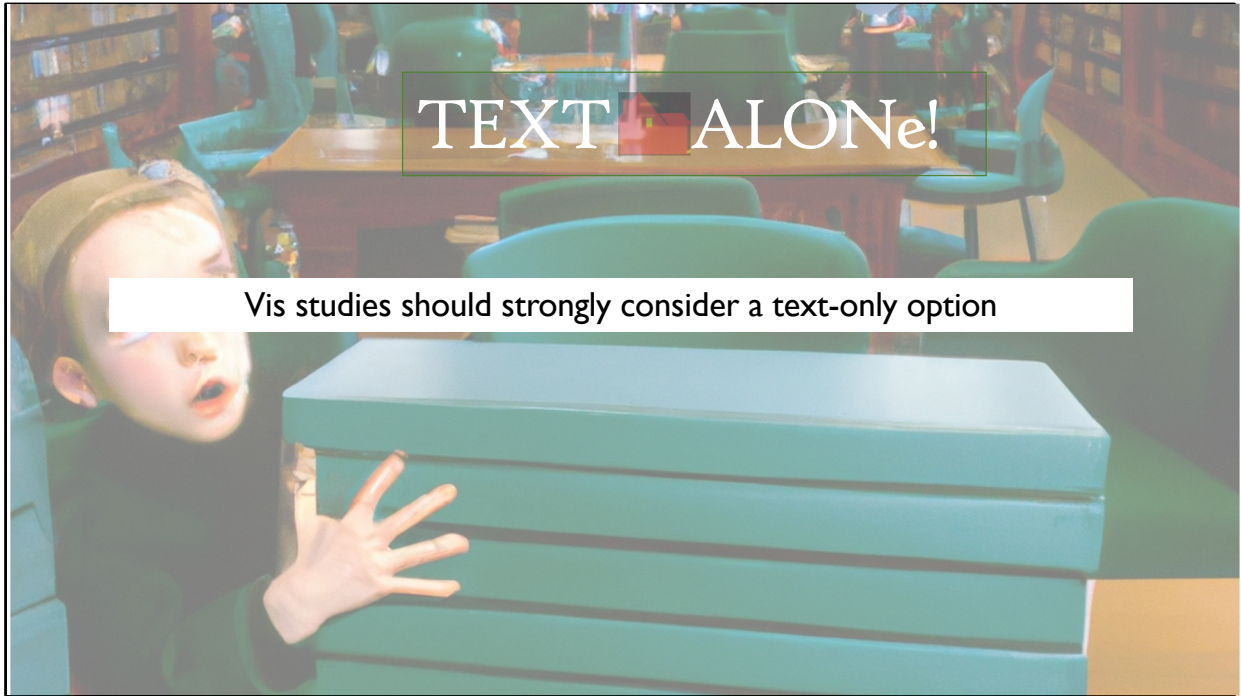
There is a total of 100 college freshmen in the population. 30 are enrolled in an introductory entrepreneurship course. Of these 30 freshmen, 20 plan on going into business after graduation and 10 do not. On the other hand, 70 freshmen are not enrolled in an entrepreneurship course. Out of these 70 freshmen, 10 plan on going into business after graduation and 60 do not.

Text alone more accurate for exact recall of stats



In summary, these examples show empirical studies with different findings. In each case, they have a text alone condition, and they determine different outcomes based on this in comparison to a design with a viz in it.





They show that vis studies should consider including a text-only variant.

Why do some people prefer text alone?

Now we might want to ask: why do some people prefer text alone?  
We have some answers for some situations, but I don't think we fully know why.

## LITERACY

### LANGUAGE

“Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts.”

UNESCO, 2004; 2017

### VISUALIZATION

Visualization literacy is “the ability to critically interpret and construct visualizations.”

Solen, VisComm, 2022

Can preferences be related to fluency and literacy? I realize this is a complicated and controversial topic, so I want to tread lightly here.

One point I can make is that we often talk about visualization literacy. But we might want to think about reading literacy as well.

We might want to consider questions such as: what is the role of literacy in these differences? Is there a difference between the two and does that cause differences in their performance? Or is it something else?

Here I have two definitions for literacy. For reading, I used the UNESCO definition. For visualization, I've put here a new definition for it, from a paper by Solen in the VisComm workshop from Monday.

I don't claim to know the answer to this question, but I do think when we talk about visualization literacy, we should think about how, if at all, it interacts with reading literacy.



## COGNITIVE MODELS



For Supporting Fluent Reading

For Combining Text + Visuals

"a brain composed of hundreds of tv sets, digital art"

This leads to my next topic on cognitive models. I want to start here with a caveat – I am not an expert by any stretch in this topic, and I'm sure I am missing a lot of key information. But I want to make two main points in this section. The first is that I think that vis researchers need to think carefully about how people read – that is, cognitively process written language when they design visualizations. And the second is that as far as I can tell, there are no heavily used cognitive models for how people combine language and visualizations.

Let's start with an example of visualization being inserted within a paragraph of text.

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.

<https://www.nytimes.com/interactive/2021/08/18/opinion/inflation-economy-transitory.html>

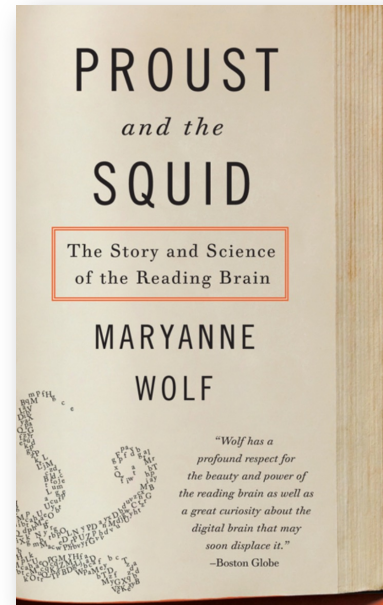
Take a moment to take a look. My question to you is: does your eye simply read this paragraph, or does it dart around from the visualizations, to the text, back to the visualizations in an erratic manner? I know that mine does.

## The Reading Brain

“We were never born to read.”

Reading researchers have gathered extensive evidence suggesting that the processing of words occurs in the parafovea before the word is directly fixated on.

Wolf notes that the preview of what lies ahead makes what follows easier to recognize, contributing to automaticity, which aids fluent reading.



P 3

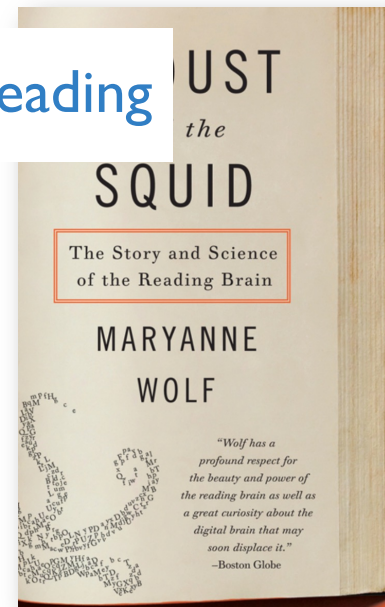
I really love this book by Maryanne Wolf. It explains at an intermediate level what is understood about cognition and reading, and relates these to the importance of fluency in reading.

She opens with this statement “we were never born to read”. What she means by this is that although humans innately learn spoken language in most cases, reading is something that requires new pathways across many different brain regions to be learned and formed.

Among other things, she points out that researchers have gathered extensive evidence that the processing of words occurs in the parafovea, before the word is directly fixated on. She notes that this preview of what lies ahead on the line contributes to fluent reading.

## The Importance of Fluent Reading

“Fluency does not ensure better comprehension; rather, fluency gives enough extra time to the executive system to direct attention where it is most needed – to infer, to understand, to predict...”



(pp 148-149)

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.

What I am focused on here is longer blocks of text, not short text like a phrase or sentence

So going back to visualization and reading, let's look at an example.



Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Habitasse platea dictumst vestibulum rhoncus est. In hendrerit gravida rutrum quisque non tellus orci ac auctor. Enim ut tellus elementum sagittis vitae et leo. Porta nibh venenatis cras sed felis. Laoreet sit amet cursus sit amet dictum sit. Praesent elementum facilisis leo vel fringilla est ullamcorper eget. Sed felis eget velit aliquet sagittis id consectetur purus. Risus ultricies tristique nulla aliquet enim tortor at auctor urna. Dignissim enim sit amet venenatis urna cursus eget nunc. Tincidunt dui ut ornare lectus sit amet est. Metus vulputate eu scelerisque felis imperdiet proin fermentum. Adipiscing bibendum est ultricies integer quis. Et malesuada fames ac turpis egestas integer eget aliquet. In tellus integer feugiat scelerisque varius morbi enim. Vitae elementum curabitur vitae nunc sed velit.

Here, we have a block of generic text.

An expert reader uses their peripheral vision to pick up on visual characteristics such as word shape. This peripheral vision doesn't usually give us semantic meaning, but it can approximate the general shape of what is to come.

>Lorem ipsum dolot sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

This is an exaggerated approximation of how your brain goes about reading. So here, you can see we are just starting the passage here, with attention focused on the first word. The rest of the passage, which we're seeing through our peripheral vision, seems as if it is behind frosted glass. This means that, when we run into unexpected or salient disruptors in the text, it can throw our visual system off.

We'll look at a few examples of what kinds of embedded graphics can have this effect, and I'll show both the clear version and this 'frosted glass' effect.

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut [labore](#) et dolore magna aliqua. Habitasse platea dictumst vestibulum rhoncus est. In hendrerit gravida rutrum quisque non tellus orci ac auctor. Enim ut tellus elementum sagittis vitae et leo. Porta nibh venenatis cras sed felis. Laoreet sit amet cursus sit amet dictum sit. Praesent elementum facilisis leo vel fringilla est ullamcorper eget. Sed felis eget velit aliquet sagittis id consectetur purus. Risus [ultrices](#) tristique nulla aliquet enim tortor at auctor urna. Dignissim enim sit amet venenatis urna cursus eget nunc. Tincidunt dui ut ornare lectus sit amet est. Metus vulputate eu scelerisque felis imperdiet proin fermentum. [Adipiscing](#) bibendum est ultricies integer quis. Et malesuada fames ac turpis egestas integer eget aliquet. In tellus integer feugiat scelerisque varius morbi [enim](#). Vitae elementum curabitur vitae nunc sed velit.

First, consider hyperlinks. Fitzsimmons et al found that readers focus on hyperlinks when skimming, and they tend to use these links as markers for important parts of the text. Seeing them here, they're visually salient, and they can draw attention away from the rest of the text.





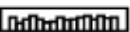
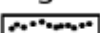
Lorem ipsum dolot sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut ☹️ et dolore magna aliqua. Habitasse platea dictumst vestibulum rhoncus est. In hendrerit gravida rutrum quisque non tellus orci ac auctor. Enim ut tellus elementum sagittis vitae et leo. Porta nibh venenatis cras sed felis. Laoreet sit amet cursus sit amet dictum sit. Praesent elementum facilisis leo vel fringilla est ullamcorper eget. Sed felis eget velit aliquet sagittis id consectetur purus. Risus ❤️ tristisque nulla aliquet enim tortor at auctor urna. Dignissim enim sit amet venenatis urna cursus eget nunc. Tincidunt dui ut ornare lectus sit amet est. Metus vulputate eu scelerisque felis imperdiet proin fermentum. 🍷 bibendum est ultricies integer quis. Et malesuada fames ac turpis egestas integer eget aliquet. In tellus integer feugiat scelerisque varius morbi 🍷. Vitae elementum curabitur vitae nunc sed velit.

A similar effect can be seen here with icons embedded in text. They're not the shape of letters or words, which our visual system knows how to recognize. They're eye-catching, and a couple of studies show that they can slow down reading.

N. Cohn, T. Roijackers, R. Schaap, and J. Engelen. Are emoji a poor substitute for words? sentence processing with emoji substitutions. In CogSci, 2018.

E. Barach, L. B. Feldman, and H. Sheridan. Are emojis processed like words?: Eye movements reveal the time course of semantic processing for emoji-fied text. Psychonomic Bulletin & Review, pages 1–14, 2021.



Lorem ipsum dolot sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut  et dolore magna aliqua. Habitasse platea dictumst vestibulum rhoncus est. In hendrerit gravida rutrum quisque non tellus orci ac auctor. Enim ut tellus elementum sagittis vitae et leo. Porta nibh venenatis cras sed felis. Laoreet sit amet cursus sit amet dictum sit. Praesent elementum facilisis leo vel fringilla est ullamcorper eget. Sed felis eget velit aliquet sagittis id consectetur purus. Risus  tristique nulla aliquet enim tortor at auctor urna. Dignissim enim sit amet venenatis urna cursus eget nunc. Tincidunt dui ut ornare lectus sit amet est. Metus vulputate eu scelerisque felis imperdiet proin fermentum.  bibendum est ultricies integer quis. Et malesuada fames ac turpis egestas integer eget aliquet. In tellus integer feugiat scelerisque  varius morbi. Vitae elementum curabitur vitae nunc sed velit.

Finally, we have these embedded visualizations. On one hand, they seem as though they should be useful, putting the visual information right next to the textual information, allowing the reader to get exactly what they need in both formats. However, as you might pick up on here, they can feel distracting from the surrounding text.

Lorem ipsum dolot sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit anim id est laborum.

And again, even in the peripheral, these graphics can be salient and overall should be assessed in the way they might affect readability of the text passage.

However, as far as I know, the relevant studies have not been done.





To support fluent reading, legibility of long text spans should receive high priority.

Insertion of non-alphanumeric visuals into paragraphs may impede fluent reading

So to summarize this point ... oh, let me fix this

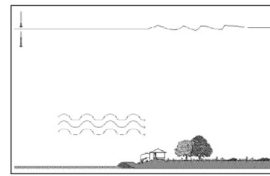
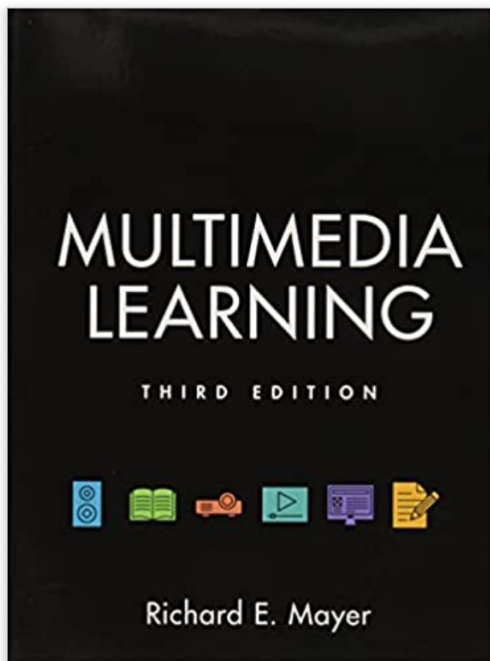
To support fluent reading,  legibility of long text spans should receive high **!!** priority

**Insertion** of non-alphanumeric  visuals into paragraphs may  impede fluent reading

To summarize this point, to support fluent reading, legibility of long text spans should receive high priority. Insertion of non-alphanumeric visuals into paragraphs may impede fluent reading.

What are the cognitive models of how text and images are perceived together?

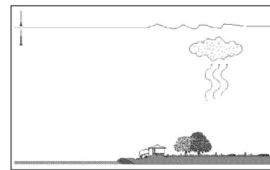
What are the cognitive models of how text and images are perceived together?



"Cool moist air moves over a warmer surface and becomes heated."



"Warmed moist air near the earth's surface rises rapidly."



"As the air in this updraft cools, water vapor condenses into water droplets and forms a cloud."

Extensive research on combining language and images for explanations of processes.

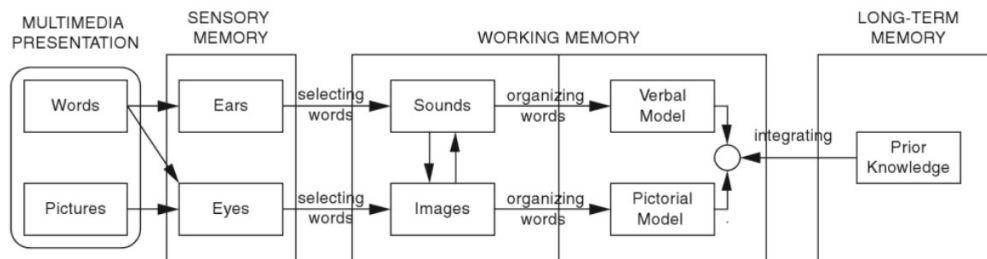
Dual-channel cognitive model

Richard Mayer has done extensive research on combining language and visuals for the purposes of education. He looks specifically at describing physical processes that can be shown visually. He and his collaborators have done a lot of work comparing written language to audio in this situation, and also looking at placement of text with respect to the visuals. However, his focus has been limited to this use case of physical process explanations, and short pieces of text only.

The cognitive model that he says explains this interactions is the dual-channel model.

## Dual-Channel Model assumes:

- **Separate systems** for processing pictorial and verbal
- Each channel has **limited capacity**
- Meaningful learning involves **actively building connections** between the two



Mayer, Richard, Multimedia Learning, third edition

The dual-channel model assumes separate cognitive systems or channels for processing pictorial and verbal information. It assumes that each channel has limited capacity, and that meaningful learning involved actively building connections between the two.

As far as I know, there has not yet been work done that verifies this model in neurophysiological terms.

***MULTIMEDIA EFFECTS ON PROCESSING  
AND PERCEPTION OF ONLINE NEWS:  
A STUDY OF PICTURE, AUDIO, AND  
VIDEO DOWNLOADS***

---

*By S. Shyam Sundar*

*J&MC Quarterly  
Vol. 77, No. 3  
Autumn 2000  
480-499  
©2000 AEJMC*

I looked for other cognitive models in the visualization literature, but did not find a lot. However, when I looked at papers from journalism, I did find a few examples.

This work often referenced this paper as a starting place.

## Dual-coding theory

There are two cognitive sub-systems that operate independently as far as encoding into memory is concerned.

## Cue-summation theory

When textual information is presented along with images it provides additional learning cues, particularly at the time of retrieval from memory

## Limited capacity information processing theory

Media messages, delivered simultaneously in several modalities are cognitively complex and serve to overload the processing system.

Sundar, Multimedia effects on the Processing of Online News, JMCQ 2000

The three main theories are:

The dual-coding theory, which I already talked about, that posits that there are two cognitive subsystems for language vs image, and they operate independently when coding information into memory.

The next is cue-summation theory that posits that when the two are presented together, text provides additional learning cues, particularly at memory retrieval time.

The third is the limited capacity information processing theory. This states that combining multiple modalities overwhelms the system.

As far as I can tell, these each cover all the cases: independent, additive, and interfering or subtractive. As far as I can tell, it is not known which of these is actually the case.

So my main message here is that a lot of work remains to be done

on cognitive modeling for text + viz, and I hope that those who do this kind of work will follow up and publish at viz!





My final topic is the use of language to query for and to generate visualizations. The organizers who invited me to speak asked me to discuss the role of AI on my topic.

## Large Language Models are Transforming the Field



Language as UI to Language

Input a text prompt, get out new text

Large language models are changing the fields of both NLP and vision, and even more relevant for this talk, of the two used together in tandem.

In this section, I would like to talk about what these new models mean for the future of Infoviz.

I'll start with the models that create text from text, such as GPT-3 and T5.

## Using a Large Language Model to Produce Text

Input Prompt: `Recite the first law of robotics`



Output: `A robot may not injure a human being or, through inaction, allow a human being to come to harm.`

<https://jalammar.github.io/how-gpt3-works-visualizations-animations/>

To use a trained model, you give it an input text and it generates some output

I'm not going into detail on how they work here, but the link shows a great post about how they work.

<https://jalammar.github.io/how-gpt3-works-visualizations-animations/>



when I gave GPT-3 the prompt:

***In 1958, Frank Rosenblatt introduced the Perceptron, an early version of a neural net that could only detect the location of a square on a printed card. Fifty years later, neural nets are capable of...***

The software dutifully completed the thought, and then continued on, picking up on the historical framing of the initial text:

***... detecting and classifying faces, finding patterns in data, and emulating the human brain. While the neural net matured in academia, it also found its way into the tech industry. In the late 1980s and early 90s, neural nets were used in speech recognition and character recognition applications. The growth in computing power and the ability to collect larger data sets allowed other neural net applications to emerge.***

Most of you have probably heard about the major advancements happening in natural language processing around large language models such as GPT-3 and T5. You might have seen text that they can generate.

Here is an example of its use in this NYTimes article.

<https://www.nytimes.com/2022/04/15/magazine/ai-language.html>

## NLP before ~2020:

First I'll give some background and caveats about these models. Then I'll speculate about what this means for the future of viz.

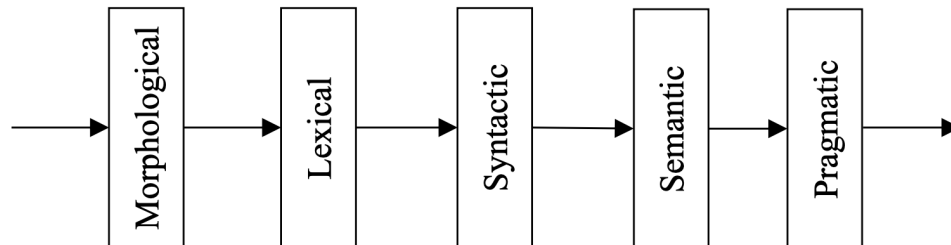
I assume that most people in the audience are not tracking this topic very carefully, so I apologize in advance to those who know a lot about it already.

I have been working in NLP since 1987, and I have to say that I have never experienced the speed of advancement that we see today. I will also admit that I was initially skeptical of the new developments, mainly because of the hype and overclaiming. I will be talking about their drawbacks in a few minutes.

Perhaps most compelling of all, from an NLP perspective, is that the models are very large, but in some sense very simple,

compared to how we have built NLP pipelines in the past.

## NLP before ~2020:



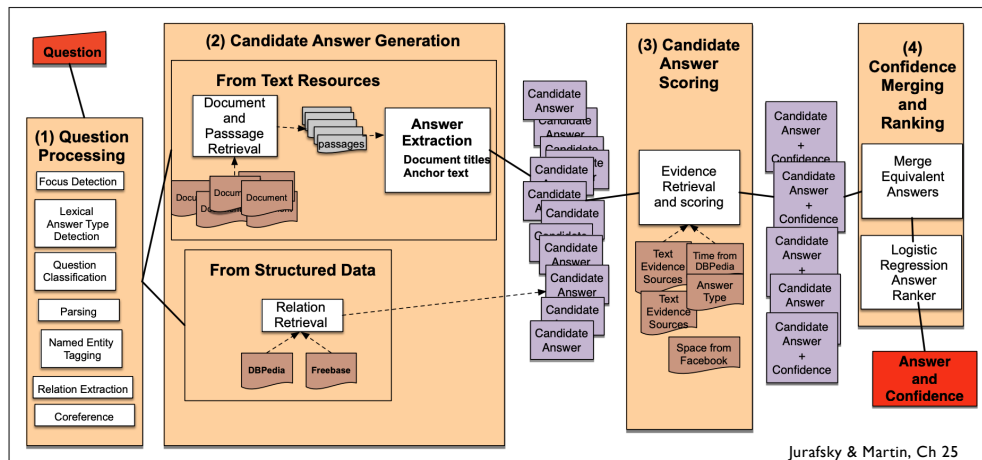
Each component needs an algorithm & hand-labeled training data  
Going from one stage to the next is complex and loses information

Shi, Pustejovsky, & Verhagen. "A conceptual framework of online natural language processing pipeline application." *Proceedings of the Workshop on OIAF, for HLT*. 2014.

Here is an example of typical NLP pipeline from as recently as 2014. The key thing to notice is that each part of the pipeline is a special kind of processing, usually with hand-coded rules and specialized hand-labeled training data.



## The “Classic” Paradigm: IBM Watson’s DeepQA



Built by a huge team over about 7 years; it worked really well!

Here is an example of a famous NLP pipeline from 2004-2011. It's goal is to perform automated question answering for the game of Jeopardy. The key thing to notice is that each part of the pipeline is a special kind of processing, usually with hand-coded rules and specialized hand-labeled training data.

## Viz Research / Products Building NL Interfaces

### **DataTone: Managing Ambiguity in Natural Language Interfaces for Data Visualization** UIST 2015

Tong Gao<sup>1</sup>, Mira Dontcheva<sup>2</sup>, Eytan Adar<sup>1</sup>, Zhicheng Liu<sup>2</sup>, Karrie Karahalios<sup>3</sup>

### **Eviza: A Natural Language Interface for Visual Analysis**

Authors:  Vidya Setlur,  Sarah E. Battersby,  Melanie Tory,  Rich Gossweiler,  Angel X. Chang [Authors li](#)

### **NL4DV: A Toolkit for Generating Analytic Specifications for Data Visualization from Natural Language Queries**

TVCG 2021

Arpit Narechania\*, Arjun Srinivasan\*, and John Stasko

### **Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements**

120

Weiwei Cui, Xiaoyu Zhang, Yun Wang, He Huang, Bei Chen, Lei Fang, Haidong Zhang, Jian-Guan Lou, and Dongmei Zhang



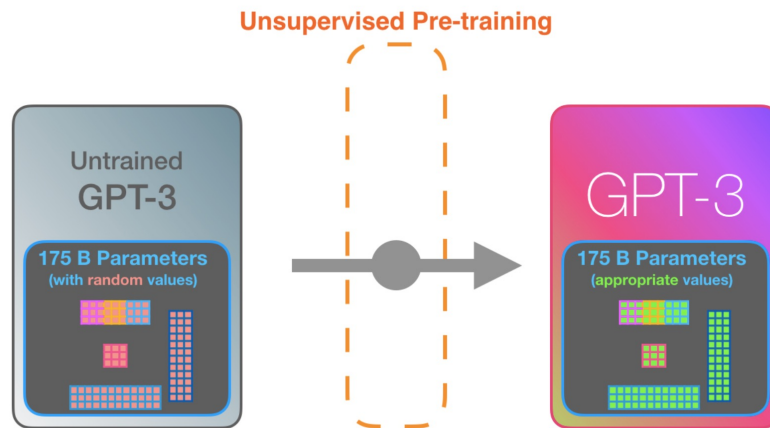
ThoughtSpot.



ableau narrative science

There has been a lot of great research in the viz literature working on how to convert data and queries to visualizations. These use the prior methods of building NLP models, and so are specialized to their tasks. It is challenging to get good coverage of the possible ways to express ideas in language in a robust manner.

## Large Language Models



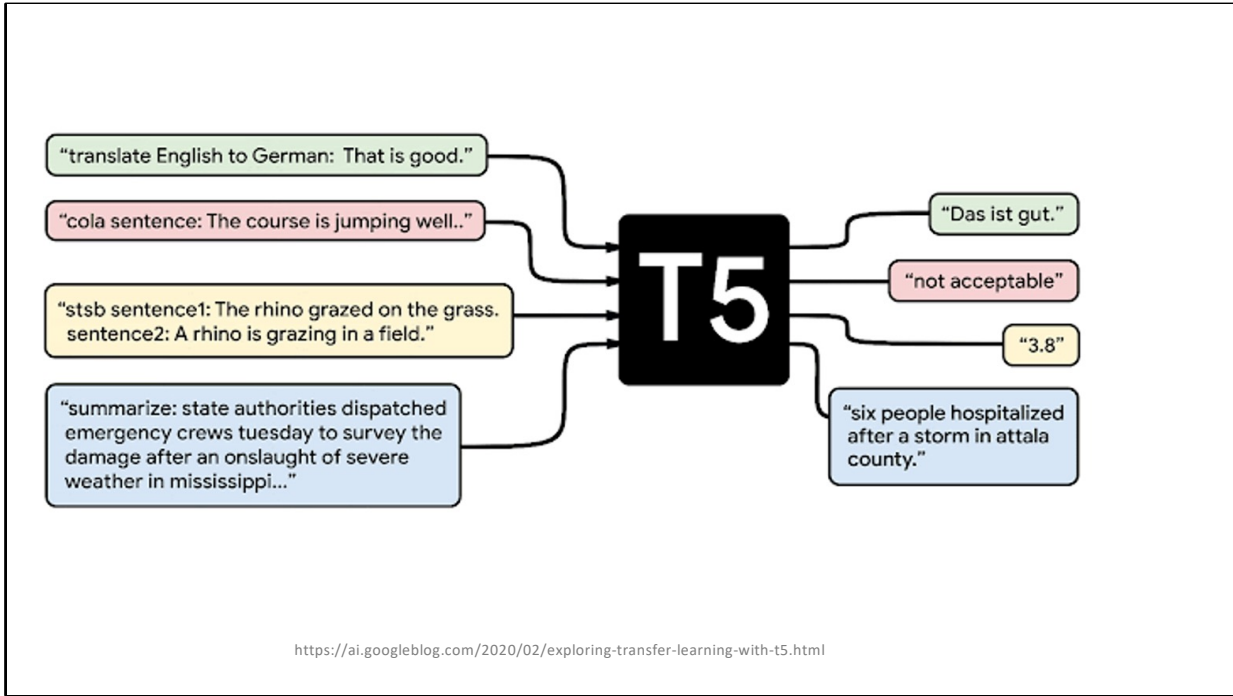
<https://jalammar.github.io/how-gpt3-works-visualizations-animations/>

Language models are context-sensitive deep learning models that learn the probabilities of a sequence of words, be it spoken or written, in a common language such as English

With these probabilities it then predicts the next word in that sequence, known as “next word prediction.”

GPT-3 is trained on 40GB of text and contains 175 billion parameters. The training is unsupervised; it can be done directly from text so long as there is a lot of it to train on.

<https://medium.com/sogetiblogsnl/language-models-battle-of-the-parameters-natural-language-processing-on-steroids-rocket-101b51fdf8>



Large language models are trained on different tasks by stating the task as part of the text input.

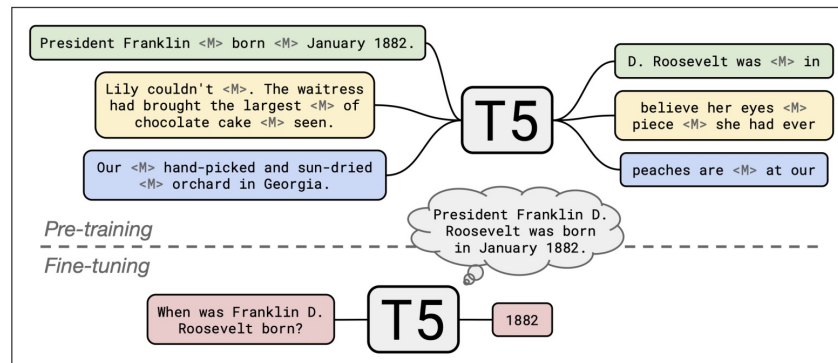
The four tasks here are: translation, judging if a sentence is grammatically and semantically meaningful, semantic similarity comparison, and summarization.

## Example: Question Answering

Pre-training is done in advance for a huge amount of input.

In this case, the model learns to fill in the blanks <M>

Then fine-tuned on QA datasets without any additional context or passages.



Roberts, Raffel, and Shazeer. "How much knowledge can you pack into the parameters of a language model?." EMNLP 2020

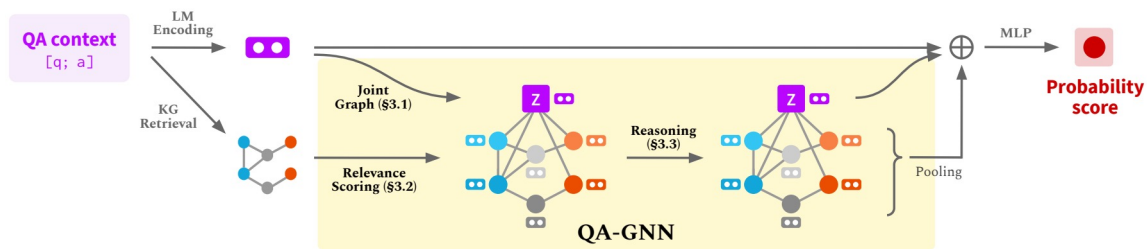
Here is an example of T5 applied to question answering. Pre-training is done in advance for a huge amount of input. In this case, the model learns to fill in the blanks <M> It is then fine-tuned on QA datasets, given the question, without adding any additional context or passages.

QA-GNN: Reasoning with Language Models and Knowledge Graphs  
for Question Answering

NAACL 2021

Michihiro Yasunaga Hongyu Ren Antoine Bosselut  
Percy Liang Jure Leskovec

Research into models incorporating knowledge representations / external data

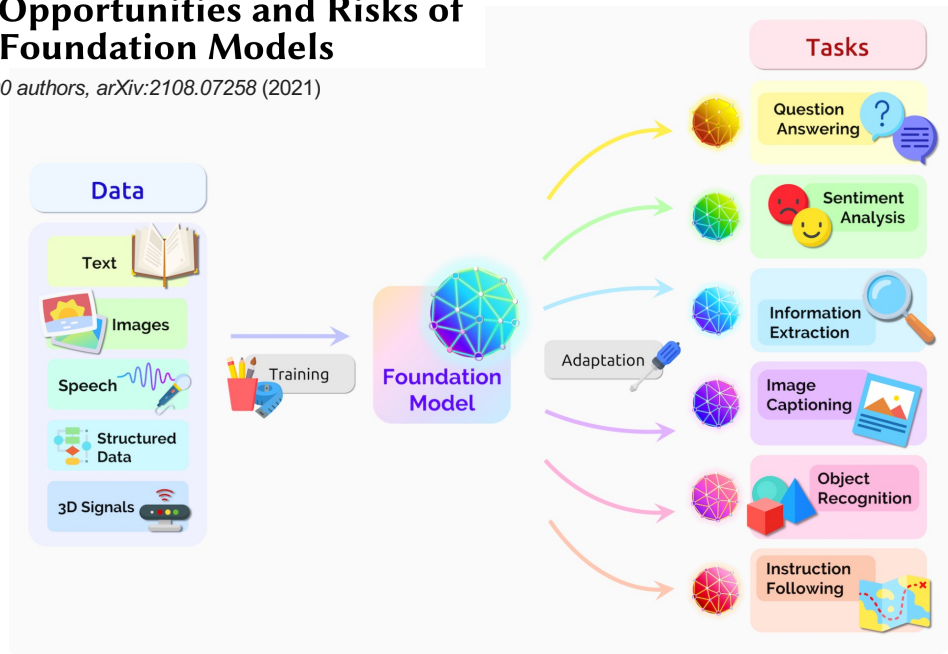


99

That was an example of how simple can you go. People are experimenting with incorporating knowledge representations into these models; this is an example architecture for question answering.

## On the Opportunities and Risks of Foundation Models

~200 authors, *arXiv:2108.07258* (2021)



Part of what is new about this approach is that it can combine input from many different modalities and produce applications of many different types.

## Large Language Models are Transforming the Field



Language as UI to Generate Images

Large language models are changing the fields of both NLP and vision, and even more relevant for this talk, of the two used together in tandem.

Here I show two major tasks that are being transformed by large language models: automated co-writing of code, and image generation.

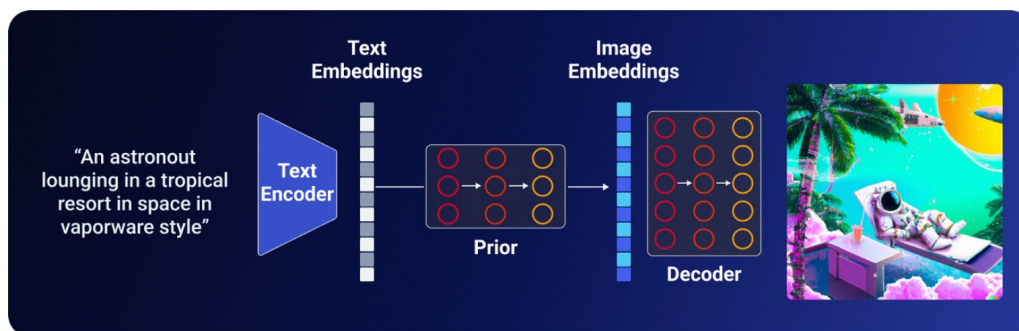
In this section, I would like to talk about what, if anything, these new models mean for the future of Infoviz.



## DALL-E2

### Hierarchical Text-Conditional Image Generation with CLIP Latents

Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. *arXiv:2204.06125* (2022).



Learns a **joint** representation space for text and images

<https://medium.com/augmented-startups/how-does-dall-e-2-work-e6d492a2667f>

From the perspective of this talk, one of the most compelling aspects of these new models is the fact that they train on representations of text and images simultaneously. There are many different model architectures out there, but what is really interesting is that when you start with a sentence of text as the input, and indicate that an image should be the output, the model learns a joint representation of the two. Given the thesis of my talk, this is very relevant!

## Images generated in response to a text prompt

DALL-E2



"a sea otter with a pearl earring" by Johannes Vermeer

IMAGEN



Vines in the shape of text 'Imagen' with flowers and butterflies bursting out of an old TV.

PARTI



A. A photo of a frog reading the newspaper named "Tooday" written on it. There is a frog printed on the newspaper too.

A lot of you have probably also seen images generated by systems like DALL-E2, Imagen, Parti, and MidJourney. I've been showing images generated by DALL-E2 in this talk, not because it is the best, but because you can use a web interface with it to generate images. In many ways these are really amazing.

## Where DALL-E2 currently has difficulty

Generating text

“a visualization of income inequality”

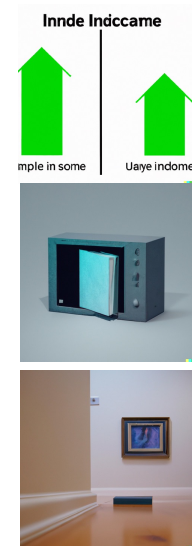
Generating attributes and relations between objects, comparisons

“a book turning into a tv”

Requesting objects be shown differently than is typical

“surveillance video of a book being chased down a corridor by a framed painting”

<https://medium.com/augmented-startups/how-does-dall-e-2-work-e6d492a2667f>

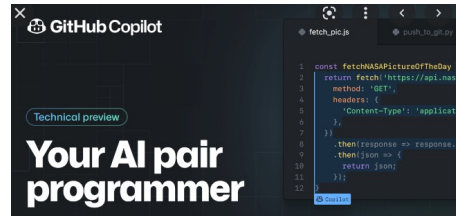


How to write the text to get the kind of output you want has become something of a dark art. These are areas where Dall-e2 currently has difficulty; some of the other systems do better on these challenges.

## Large Language Models are Transforming the Field



Language as UI to Generate Images



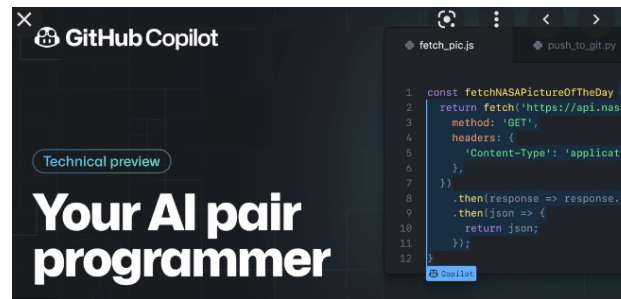
Language as UI to Generate Code

Large language models are changing the fields of both NLP and vision, and even more relevant for this talk, of the two used together in tandem.

Here I show two major tasks that are being transformed by large language models: automated co-writing of code, and image generation.

In this section, I would like to talk about what, if anything, these new models mean for the future of Infoviz.

## Natural Language as the UI for Code



This works for language plus code, as well as for language plus images. Codex is a large model, built by OpenAI, that trains on language – specifically comments – and software code simultaneously. It is used in an application called CoPilot that is part of Github, and as far as I can tell, is now very widely used.

## Example: Use Copilot to write viz code

**Goal:** Create an example for class

Generate some random data about car sales  
Generate a bar chart, with differential coloring

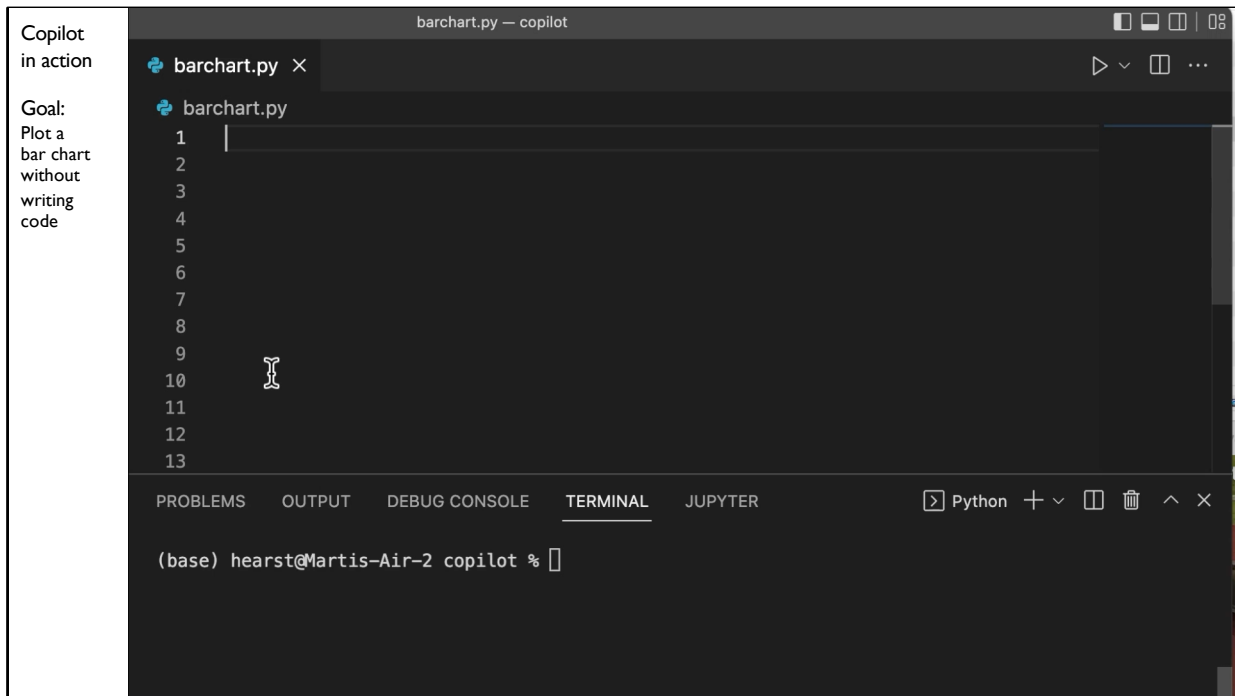
**Problem:** I can never remember the matplotlib syntax!

In the following example, I only write natural language

I'm about to show an example of using copilot to generate viz code.  
My goal is to create some example code for class.  
I want it to generate some data about cars with minimal work on my part  
I want this to create a bar chart with differential coloring.

The problem is that I can never remember the syntax for matplotlib lib

Notice in the following example that I don't write any code, only comments



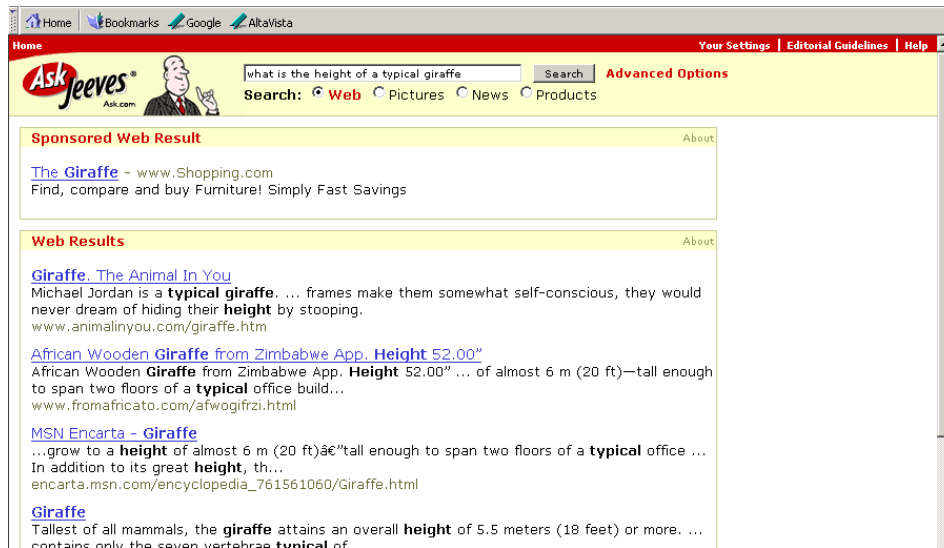
Here is an example of me using co-pilot to generate a bar chart from data.

CoPilot allows a programmer to type in a text comment, and have the system suggest some code that corresponds to that comment.

Now, there have been some empirical studies that show that the accuracy is relatively low, that it often has security issues, and that it can generate only simple code. All that said, many many people are using it in a human-computer interaction for coding.

Adapted from <https://www.youtube.com/watch?v=oFwG0la8gjY>

# Natural Language as the UI for Search



This is not an isolated case. I've also been in the field of search, or information retrieval for many years. Even back in the 1990's, people loved the Ask Jeeves search engine, even though it did not work well. Why did they love it? They loved the idea of asking it questions rather than typing in keyword queries.



# Natural Language as the UI for Search

The screenshot shows a Google search interface with the query "what is the height of a typical giraffe". The search results are filtered to "All" and show a "Giraffe / Height" section. This section contains two summary boxes: "16 – 19 ft. Male, Adult" and "14 – 17 ft. Female, Adult". Below these are three search results:

- [https://zoologist.com > how-tall-is-a-giraffe](https://zoologist.com/how-tall-is-a-giraffe)  
**How Tall is a Giraffe - Giraffe Height - Zoologist**  
Apr 23, 2019 — The average height of a Giraffe in feet is 16.4 to 19.7 feet. 2. How Tall is a Male Giraffe. The average height of a male Giraffe is 5 to 6 ...
- [https://www.pbs.org > Home > Blog](https://www.pbs.org/Home/Blog) 43  
**Giraffe Fact Sheet | Blog | Nature - PBS**  
Jul 15, 2020 — Size and Weight: ... Giraffes are the tallest land animals. Female giraffes are up to 14 feet tall and weigh up to 1,500 pounds. Meanwhile, males ...
- [https://animals.mom.com > tall-giraffes-1592](https://animals.mom.com/tall-giraffes-1592)  
**How Tall are Giraffes? - Pets on Mom.com**

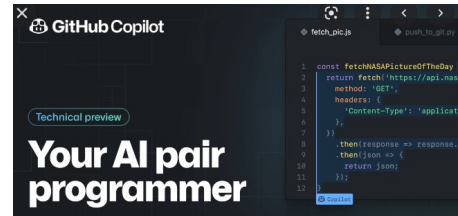
On the right side of the search results, there is a "More Images" section with a grid of giraffe photos. Below the images is a card for "Giraffe" (Animal) with a share icon and a brief description: "The giraffe is a large African hoofed mammal".

Fast forward 20 years, and you can often do just that successfully in a search engine.

## Large Language Models are Transforming the Field



Language as UI to Generate Images



Language as UI to Generate Code

Large language models are changing the fields of both NLP and vision, and even more relevant for this talk, of the two used together in tandem.

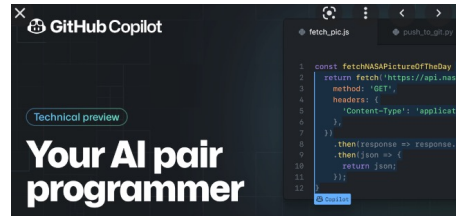
Here I show two major tasks that are being transformed by large language models: automated co-writing of code, and image generation.

In this section, I would like to talk about what, if anything, these new models mean for the future of Infoviz.

## Large Language Models are Transforming the Field



Language as UI to Generate Images



Language as UI to Generate Code



Language as UI to Generate Visualizations?

Large language models are changing the fields of both NLP and vision, and even more relevant for this talk, of the two used together in tandem.

Here I show two major tasks that are being transformed by large language models: automated co-writing of code, and image generation.

In this section, I would like to talk about what, if anything, these new models mean for the future of Infoviz.

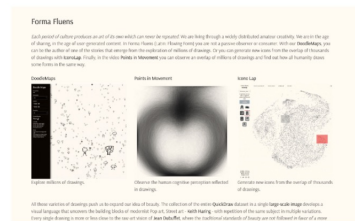
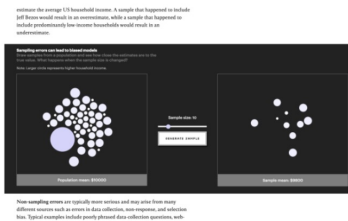
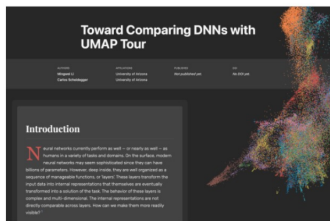
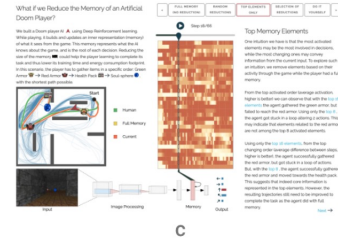
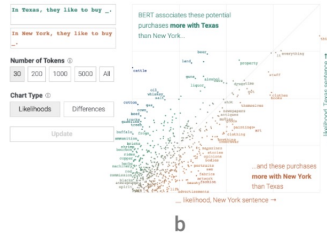
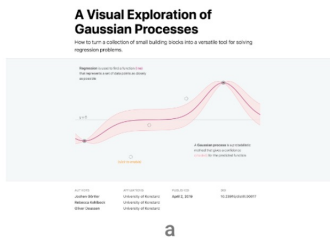
# 5<sup>th</sup> Workshop on Visualization for AI Explainability

October 17th, 2022 at IEEE VIS in Oklahoma City, Oklahoma



The role of visualization in artificial intelligence (AI) gained significant attention in recent years. With the growing complexity of AI models, the critical need for understanding their inner-workings has increased. Visualization is potentially a powerful technique to fill such a critical need.

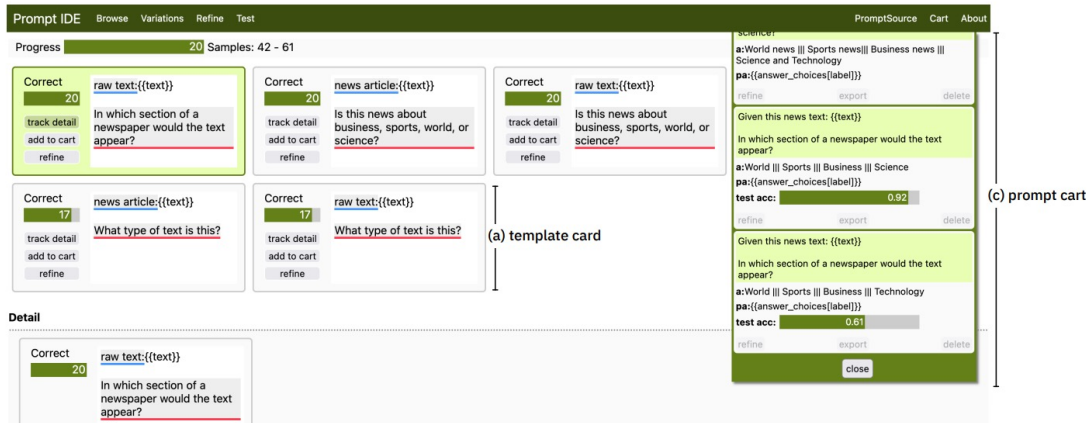
The goal of this workshop is to initiate a call for "explainables" / "explorables" that explain how AI techniques work using visualization. We believe the VIS community can leverage their expertise in creating visual narratives to bring new insight into the often obfuscated complexity of AI systems.



So what does all this mean for viz? There is a lot of fantastic work in the infviz community on visualizing the capabilities of ML in general and some on visualizing large language models directly.

# Interactive and Visual Prompt Engineering for Ad-hoc Task Adaptation with Large Language Models

Hendrik Strobelt, Albert Webson, Victor Sanh, Benjamin Hoover,  
Johanna Beyer, Hanspeter Pfister, and Alexander M. Rush

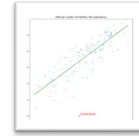


Here is an example in this conference on viz to help with prompt engineering.

## Using Large Language Models to Generate Engaging Captions for Data Visualizations

Ashley Liew\*

Klaus Mueller† NLViz Workshop 2022



Experimented w GPT-3 to generate captions for scatter plots

Text Prompt

+

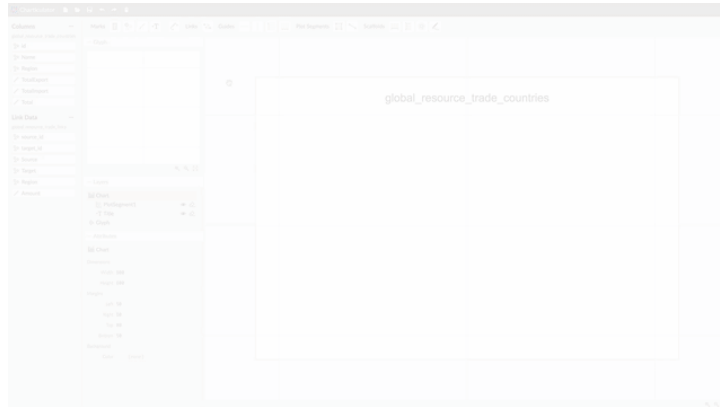
Both	
Title, Axes labels, Value ranges, Other columns in the data set	
Linear Regression	Clustering
Regression coefficients, Outliers by studentized residual distances, Pearson's correlation	Number of clusters, Cluster sizes, Description of clusters



*There is a strong positive correlation between GDP per capita and Healthy life expectancy. A country's GDP per capita is indicative of the average income of its citizens and the overall wealth of the country. A higher GDP per capita generally means that citizens have more disposable income, which can be used to purchase goods and services that improve their health and wellbeing. The outlier in this data is Swaziland, which has a lower healthy life expectancy than would be expected of its GDP per capita. ...*

Here is an example from the NLViz workshop. It feeds in the text prompt and stats about a dataset, and outputs creative captions.

## Cool UIs for Building Viz's



Charticulator: Interactive Construction of Bespoke Chart Layouts

Donghao Ren, Bongshin Lee, and Matthew Brehmer

TVCG 2018

We have really amazing interfaces in viz for building visualizations, like Charticulator.

## Natural Language as the UI for Viz?

What does all this mean for interfaces FOR viz? Well, there is a long running debate about which is better: using a GUI or using the command line, to create user interfaces and visualizations. People's preference depends on which tool they are most comfortable with.

Perhaps 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, augmented with a bit of pointing.

The copilot demo I showed you suggests the way forward. And new capabilities are being developed at an astonishing pace.



## PROBLEMS WITH LARGE LANGUAGE MODELS

- Repeat **biases** and **injustices** from the training data
- Still often **inaccurate** (but improving rapidly)
- Do not reflect “**understanding**”
- Currently require **huge resources** for training
- May aid **misinformation** / undermine foundations of real

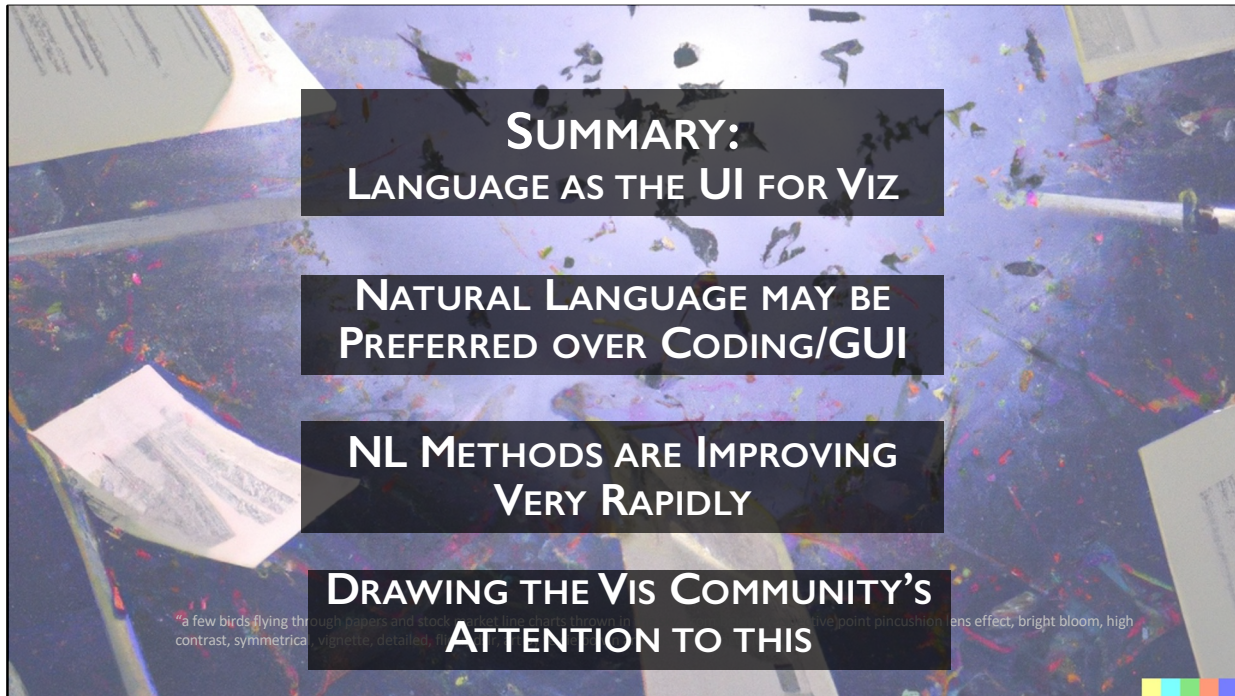
I want to pause here to say there are a lot of problems with these models. One of the most well-known problems is that they are trained on huge collections of “found data”, and so if care is not taken, they repeat the biases and injustices that are inherent in those datasets.

They are still far from perfect,

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.

They are huge, not available to all researchers or users due to their size, and they are costly to train in terms of compute, and to a lesser degree, energy consumption.

And I think the biggest drawback of all is how they could contribute to misinformation and make it very hard to determine what information is real and what is computer generated.



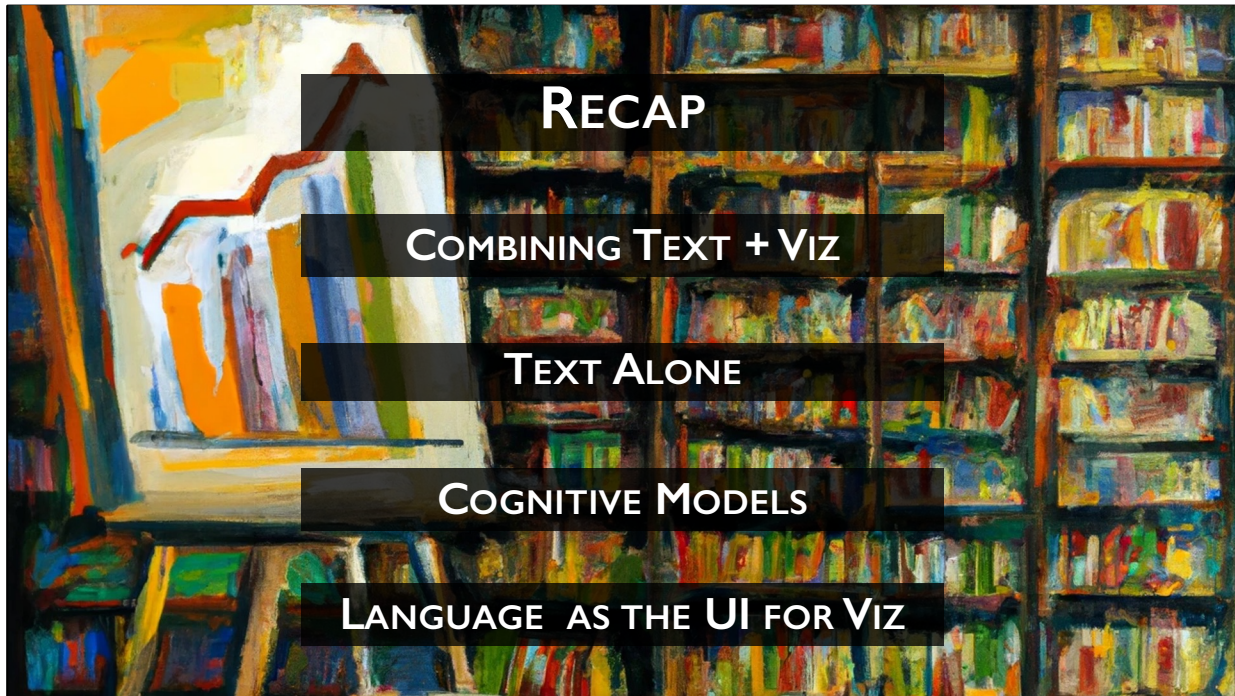
**SUMMARY:  
LANGUAGE AS THE UI FOR VIZ**

**NATURAL LANGUAGE MAY BE  
PREFERRED OVER CODING/GUI**

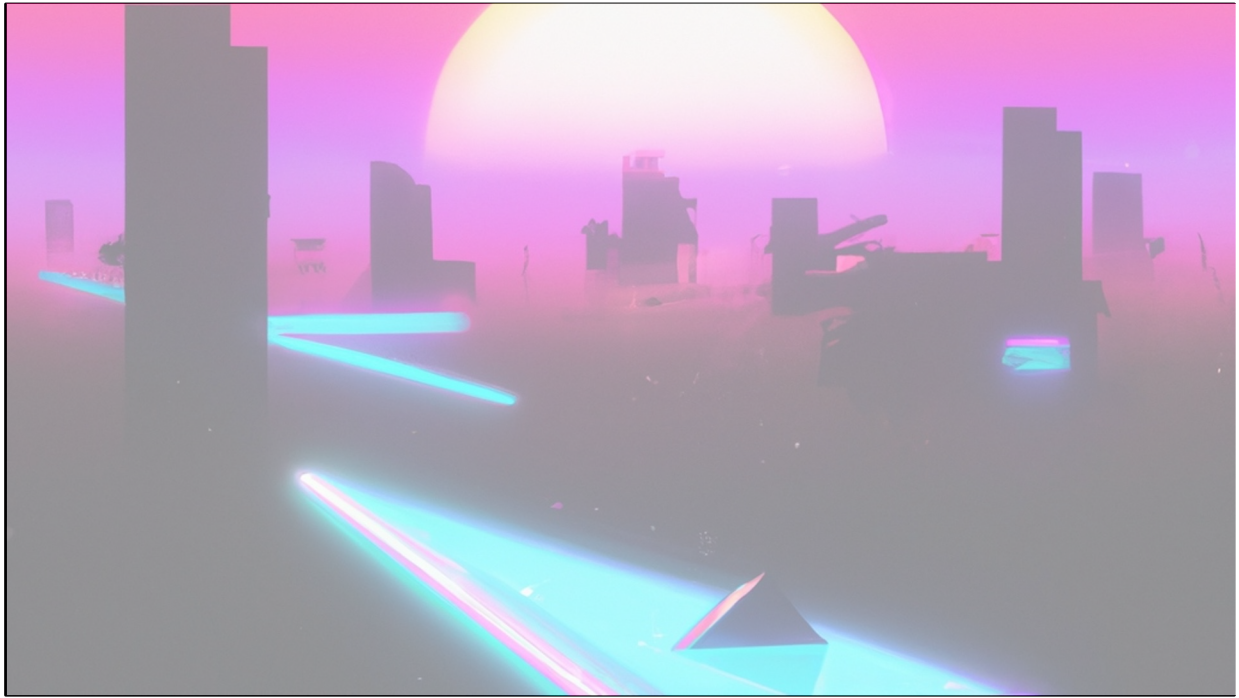
**NL METHODS ARE IMPROVING  
VERY RAPIDLY**

**DRAWING THE VIS COMMUNITY'S  
ATTENTION TO THIS**

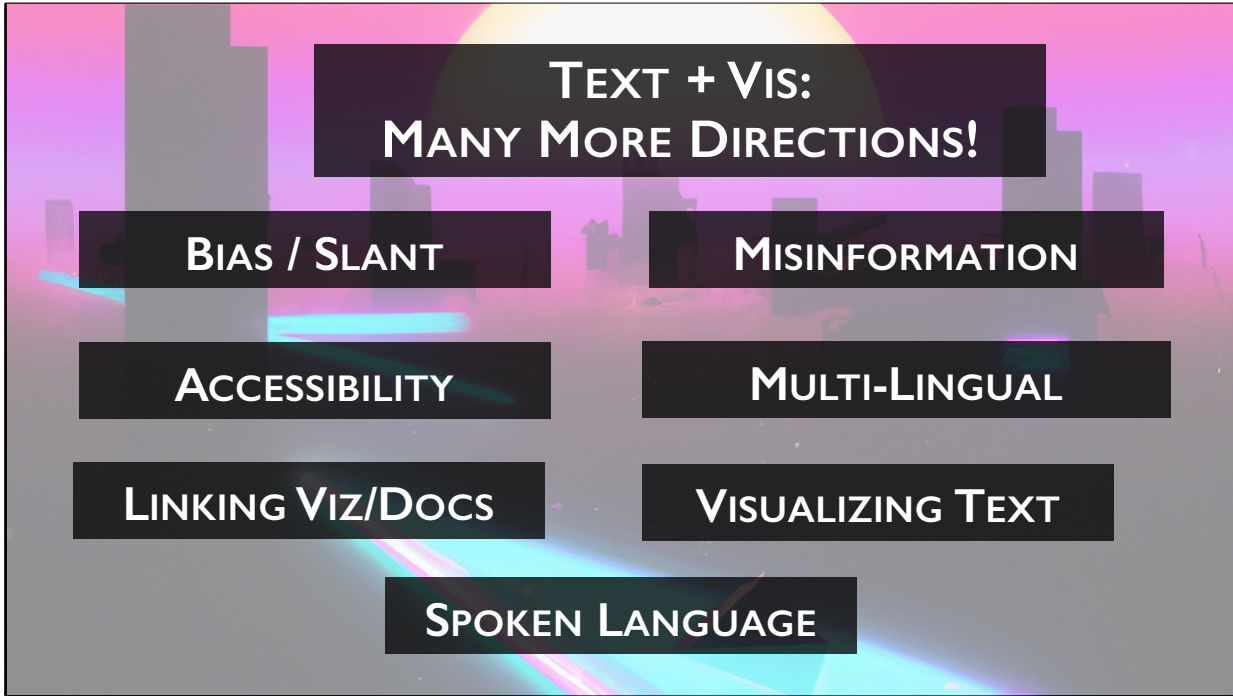
To conclude, I want to make it clear that these large models today are still far from fully able to be used in this manner, and as I said before, they have a lot of issues and drawbacks. I used to be rather skeptical about them, but I have to say that the more I see, the less skeptical I become about what they will eventually be able to do in this space. This was recently debated at the North American ACL: will language models do everything, or will they require a knowledge of linguistics and semantics. Most people said they thought that some kind of linguistic and semantic representation will be needed. But I do think they are going to have a large impact on how we create visualizations in the future.



To recap, I've talked about combining text with viz, and its complex interactions, about the importance of considering text alone, about the need for better cognitive models that combine reading and understanding visualizations, and the future of language as the UI for visualizations.



There are many additional future directions.



These include bias / slant, misinformation and the protection from it, accessibility, multi-lingual and multicultural text, dynamic documents, especially scientific ones, visualizing text, and the spoken medium. And there are more!

# Show It or Tell It?

Thank you!

Marti Hearst  
UC Berkeley

"A female minion scientist giving a keynote talk about visualization in front of an audience of 1000 minions, photorealistic"



In closing: what do you think? Show it or tell it? Thank you!