Data Visualization (CSCI 627/490)

Tasks & Design

Dr. David Koop
Data

• What is this data?

<table>
<thead>
<tr>
<th></th>
<th>42ND STREET &amp; 8TH AVENUE</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R011</td>
<td></td>
<td>00228985</td>
<td>00008471</td>
<td>00000441</td>
<td>000001455</td>
<td>00000134</td>
<td>00033341</td>
<td>00071255</td>
<td></td>
</tr>
<tr>
<td>R170</td>
<td>14TH STREET-UNION SQUARE</td>
<td>002264603</td>
<td>00011051</td>
<td>00000827</td>
<td>00003026</td>
<td>00000660</td>
<td>00089367</td>
<td>00199841</td>
<td></td>
</tr>
<tr>
<td>R046</td>
<td>42ND STREET &amp; GRAND CENTRAL</td>
<td>00207758</td>
<td>00007908</td>
<td>00000323</td>
<td>00001183</td>
<td>00003001</td>
<td>00040759</td>
<td>00096613</td>
<td></td>
</tr>
</tbody>
</table>

• **Semantics**: real-world meaning of the data

• **Type**: structural or mathematical interpretation

• Both often require **metadata**
  - Sometimes we can infer some of this information
  - Line between data and metadata isn’t always clear
Data Terminology

- Item (also Nodes): an entity
- Link: relationship between two items
- Attribute: property of an item
- Position: location in space
- Grid: how data is sampled
### Dataset Types

**Tables**
- **Attributes (columns)**
- **Items (rows)**
- **Cell containing value**

**Networks**
- **Link**
- **Node (item)**

**Fields (Continuous)**
- **Grid of positions**
- **Cell**
- **Attributes (columns)**
- **Value in cell**

**Geometry (Spatial)**
- **Position**

**Multidimensional Table**
- **Key 1**
- **Key 2**
- **Attributes**
- **Value in cell**

**Trees**

[Munzner (ill. Maguire), 2014]
Joe Carmanica recently wrote about this trend for the New York Times, arguing that it was led by Drake, who popularized the rapping-and-singing formula over the past decade.

A better benchmark for Lil Uzi Vert's word count (2,556) might be those of pop artists, such as Beyonce (2,433 words), or even one of his major influences: Marilyn Manson (2,466 words).

There are also genre-bending artists. If Childish Gambino's Awaken, My Love! is less hip-hop in the traditional '90s boom-bap sense, is it fair to compare it to vocabulary-dense Wu-Tang albums? Genre matters in vocabulary calculations—check out the chart below, which takes 500 random samples of 35,000 words from rock, country, and hip hop.

# of Unique Words Used in 500 Random Samples of 35,000 Lyrics from Country, Rock, Hip Hop

<table>
<thead>
<tr>
<th>Genre</th>
<th>Unique Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td>2,800</td>
</tr>
<tr>
<td>Rock</td>
<td>3,800</td>
</tr>
<tr>
<td>Hip Hop</td>
<td>4,800</td>
</tr>
</tbody>
</table>

Raw Lyrics Data via John W. Miller

In short, if artists depart from hip-hop song structure, we'd expect their vocabulary to go down in the number of unique words. That said, the results are still directionally interesting. Of the 150 artists in the dataset, let's take a look at who is on top.

[Sets & Lists]

[M. Daniels, 2019]
Assignment 2

• Process Data
• Create Bar Charts using SVGs and JavaScript
• Add Interaction
Attribute Types

- Categorical
- Ordered
  - Ordinal
  - Quantitative

[Muñzner (ill. Maguire), 2014]
Categorial, Ordinal, and Quantitative

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>10/14/06</td>
<td>5-Low</td>
<td>Large Box</td>
<td>0.8</td>
<td>10/21/06</td>
</tr>
<tr>
<td>6</td>
<td>2/21/08</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.55</td>
<td>2/22/08</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Small Pack</td>
<td>0.79</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Jumbo Box</td>
<td>0.72</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.6</td>
<td>7/18/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.65</td>
<td>7/18/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Wrap Bag</td>
<td>0.52</td>
<td>10/24/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Small Box</td>
<td>0.58</td>
<td>10/25/07</td>
</tr>
<tr>
<td>36</td>
<td>11/3/07</td>
<td>1-Urgent</td>
<td>Small Box</td>
<td>0.55</td>
<td>11/3/07</td>
</tr>
<tr>
<td>65</td>
<td>3/18/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.49</td>
<td>3/19/07</td>
</tr>
<tr>
<td>66</td>
<td>1/20/05</td>
<td>5-Low</td>
<td>Wrap Bag</td>
<td>0.56</td>
<td>1/20/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>0.59</td>
<td>12/23/06</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>0.82</td>
<td>12/23/06</td>
<td></td>
</tr>
<tr>
<td>96</td>
<td>4/17/05</td>
<td>2-High</td>
<td>0.55</td>
<td>4/19/05</td>
<td></td>
</tr>
<tr>
<td>97</td>
<td>1/29/06</td>
<td>3-Medium</td>
<td>0.38</td>
<td>1/30/06</td>
<td></td>
</tr>
<tr>
<td>129</td>
<td>11/19/08</td>
<td>5-Low</td>
<td>0.37</td>
<td>11/28/08</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>0.37</td>
<td>5/9/08</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>0.38</td>
<td>5/10/08</td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>0.38</td>
<td>5/10/08</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>0.38</td>
<td>6/12/06</td>
<td></td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>0.38</td>
<td>6/12/06</td>
<td></td>
</tr>
<tr>
<td>134</td>
<td>5/1/08</td>
<td>4-Not Specified</td>
<td>Large Box</td>
<td>0.82</td>
<td>5/3/08</td>
</tr>
<tr>
<td>135</td>
<td>10/21/07</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.64</td>
<td>10/23/07</td>
</tr>
<tr>
<td>166</td>
<td>9/12/07</td>
<td>2-High</td>
<td>0.55</td>
<td>9/14/07</td>
<td></td>
</tr>
<tr>
<td>193</td>
<td>8/8/06</td>
<td>1-Urgent</td>
<td>Medium Box</td>
<td>0.57</td>
<td>8/10/06</td>
</tr>
<tr>
<td>194</td>
<td>4/5/08</td>
<td>3-Medium</td>
<td>Wrap Bag</td>
<td>0.42</td>
<td>4/7/08</td>
</tr>
</tbody>
</table>

quantitative

ordinal
categorical
Ordering Direction

- Sequential
  - Diverging
  - Cyclic

[Muñzner (ill. Maguire), 2014]
Sequential and Diverging Data

• Sequential: homogenous range from a minimum to a maximum
  - Examples: Land elevations, ocean depths

• Diverging: can be deconstructed into two sequences pointing in opposite directions
  - Has a zero point (not necessary 0)
  - Example: Map of both land elevation and ocean depth

[Rogowitz & Treinish, 1998]
3.1. Mathematical description and types of spirals

For the visualization of time-dependent data Archimedes spirals are appropriate. For the comparison of periodic data sets, the general shape of the spiral should be untouched and other attributes should be used, such as texture, including line styles and patterns, since the distance to other periods are equally important. This marks with a corresponding distance to the spiral. However, quantitative, qualitative changes in the radius, i.e., values. One might consider to map data values to small abscissas is always the same.

In general, markers, bars, and line elements can be used to visualize time-series data similar to standard point, bar, and line graphs on Spiral Graphs. For instance, quantitative, discrete data can be presented as bars on the spiral or by texture, including line styles and patterns, and for the display of data from different periods are equally important. This concludes that the general shape of the spiral should be used for the display of data. Yet, we have found this way of visualizing to be inefficient.

Archimedes spirals have the form $r = a \theta$, where $a$ is a real number and $\theta$ is the angle from the origin. The characteristic property that all arcs cut a ray emanating from the origin under the same angle.

The Hyperbolic spiral has the form $r \theta = a$. It is the inverse of Archimedes spirals. The logarithmic spiral has the form $r = ae^{\theta}$. It has the special property that a ray emanating from the origin crosses two consecutive arcs of the spiral in a constant distance.

Several simple functions are always the same. The determining function is the same color coding scheme. In the spiral visualization it is much easier to compare days, to spot cloudy time periods, or to see events like sunrise and sunset.

[Sunlight intensity, Weber et al., 2001]
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”

— T. Munzner
Tasks

• Why? Understand data, but what do I want to do with it?
• Levels: High (Produce/Consume), Mid (Search), Low (Queries)
• Another key concern: Who?
  - Designer <-> User (A spectrum)
  - Complex <-> Easy to Use
  - General <-> Context-Specific
  - Flexible <-> Constrained
  - Varied Data <-> Specific Data
Tasks

What?

Why?

How?

Actions

<table>
<thead>
<tr>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Data</td>
</tr>
<tr>
<td>Attributes</td>
</tr>
<tr>
<td>Network Data</td>
</tr>
<tr>
<td>Spatial Data</td>
</tr>
</tbody>
</table>

Targets

<table>
<thead>
<tr>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Data</td>
</tr>
<tr>
<td>Attributes</td>
</tr>
<tr>
<td>Network Data</td>
</tr>
<tr>
<td>Spatial Data</td>
</tr>
</tbody>
</table>

Analyze

- Consume
  - Discover
  - Present
  - Enjoy

- Produce
  - Annotate
  - Record
  - Derive

Search

<table>
<thead>
<tr>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>Location unknown</td>
</tr>
<tr>
<td><em>·</em> Lookup</td>
<td><em>·</em> Locate</td>
</tr>
<tr>
<td><em>·</em> Browse</td>
<td><em>·</em> Explore</td>
</tr>
</tbody>
</table>

Query

- Identify
- Compare
- Summarize

[Muñzner (ill. Maguire), 2014]
Actions: Analyze

➡️ Consume
  ➡️ Discover
  ➡️ Present
  ➡️ Enjoy

➡️ Produce
  ➡️ Annotate
  ➡️ Record
  ➡️ Derive

[Munzner (ill. Maguire), 2014]
Visualization for Consumption

• Discover new knowledge
  - Generate new hypothesis or verify existing one
  - Designer doesn’t know what users need to see
  - "why doesn't dictate how"

• Present known information
  - Presenter already knows what the data says
  - Wants to communicate this to an audience
  - May be static but not limited to that

• Enjoy
  - Similar to discover, but without concrete goals
  - May be enjoyed differently than the original purpose
Asking good **questions** is very important
Answers often lead to more questions
Explore MTA Fare Data
Present Known Information

Each solid circle represents a bee species active in Carlinville, Ill., in both the late 1800s and 2010.

Hatching represents a bee species active in the 1800s but now locally extinct.

The spot where each block rests on the circle indicates one of 26 plant species frequented by these bees.

In the 1880s scientists observed the following about the bee-plant encounters:
- Present
- Frequent
- Abundant

Studies in 2009 and 2010 showed many bee-plant interactions had changed:
- Lost
- Persisted
- New

[M. Stefaner, 2013]
Enjoy Visualizations of Names

Names starting with 'AN' per million babies

[www.babynamewizard.com] [Wattenberg, 2005]
“[W]e scientists now understand how important emotion is to everyday life, how valuable. Sure, utility and usability are important, but without fun and pleasure, joy and excitement, and yes, anxiety and anger, fear and rage, our lives would be incomplete.”

—D. Norman (Emotional Design)
Measuring User Experience in Visualization

• Memorability: Capability of maintaining and retrieving information [J. Brown et al., 1977]

• Engagement: Emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource. [S. Attfield et al., 2011]

• Enjoyment: Feeling that causes a person to experience pleasure. Pleasure is recognized with occurrent happiness and excitement, which can be explained in terms of belief, desire, and thought. [W. A. Davis, 1982]

[B. Saket et al., BELIV 2016]
Memorability

Figure 4: Policy shifts and interventions to mobile outlier positions in quantitative analysis of common socio-environmental factors.

Figure 7: US prevalence in women and men by age group, selected countries.

M. Borkin et al., InfoVis 2015
Memorability & Clutter

The bottom of Figure 15 shows the same bar graph embedded within a cartoonish monster. The addition of such pictorial elements that embellish data with anthropomorphic or metaphorical elements—intended to enhance engagement or memory—has been demonized as “chartjunk” (e.g., Tufte, 1983). Various studies have shown that adding these elements leads to no improvement in memory for the data (Helgeson & Moriarty, 1993; Kelly, 1989), mixed results depending on the details of the task and context (Gillan & Richman, 1994; Li & Moacdieh, 2014), or better memory for the data content or message (Bateman et al., 2010; Borkin et al., 2016; Haroz et al., 2015b). Like animation, these visual embellishments can increase ratings of engagement and aesthetic value (Li & Moacdieh, 2014). And despite mixed evidence as to whether their presence improves memory for the data, pictorial elements do improve memory for the fact that a visualization was previously seen, both in the short and the longer term (Borkin et al., 2013).

How to Design an Understandable Visualization

Use familiar designs to show data intuitively

Visualizations can be powerful, but a poorly designed visualization can easily confuse or even mislead (Burns et al., 2020; Cairo, 2019; Szafir, 2018). Because the interpretation of visualized data is in the eye and mind of the human beholder, we must consider the psychology of the observer as the translator of images into an understanding of the original data and the patterns that they hold. Below, we outline a set of common translation errors that can confuse and mislead.

Understanding a visualization can depend on a graph schema: a knowledge structure that includes default expectations, rules, and associations that a viewer uses to extract conceptual information from a data visualization. Figure 16 serves as an example of why a graph schema is often needed to interpret a data visualization. It depicts the GDP (on a log scale) and population of the 10 most populous countries. Take a moment to interpret the data. If you are having trouble extracting the data from this visualization, it is not your fault—you do not have the needed schema. First, if you have never seen this type of visualization, you cannot know which aspects of its variation are meaningful. The bubbles differ in Fig. 15.

A “cluttered” visualization (top), a minimalist “decluttered” version (middle), and a version that incorporates pictorial embellishment (bottom). The graph at the bottom was created by Nigel Holmes for TIME Magazine and was reprinted in his 1984 book, Designer’s Guide to Creating Charts & Diagrams. Used with permission.

Memorability & Clutter

MONSTROUS COSTS
Total House and Senate Campaign Expenditures

Cost, $ Millions

350
300
250
200
150
100
50
0

Memorability & Clutter

MONSTROUS COSTS
Total House and Senate Campaign Expenditures

MONSTROUS COSTS
Total House and Senate Campaign Expenditures, in Millions

Cost, $ Millions


[S. Franconeri et al., 2021]
Memorability & Clutter

The graph at the bottom was created by Nigel Holmes and was reprinted in his 1984 book. Used with permission.

Visualizations can be powerful, but a poorly designed visualization can easily confuse or even mislead (Burns et al., 2016; Haroz et al., 2015b). Like animation, these visual embellishments can increase ratings of engagement and retention. Anthropomorphic or metaphorical elements—intended to enhance engagement or memory—have shown that adding these elements leads to no improvement in memory for the data (Helgeson & Moriarty, 1993; Kelly, 1989), mixed results depending on the details of the task and context (Gillan & Richman, 1994; Borkin et al., 2010; Borkin et al., 2013). If you are having trouble extracting the data from such pictorial elements that embellish data with anthropomorphic or metaphorical elements—intended to enhance engagement or memory—has been demonstrated to enhance engagement or memory (Bateman et al., 1993; Cairo, 2019; Szafir, 2018). Because the interpretation of visualized data is in the eye and mind of the observer as the translator of images into an interpretation, you cannot know which aspects of the visualization were meaningful. The bubbles differ in type of visualization, you cannot know which aspects of the visualization were meaningful.

Memorability of data depends on an understanding of the original data and the patterns that the needed schema. First, if you have never seen this visualization, it is not your fault—you do not have an understanding of the original data and the patterns that the needed schema. Second, even if you do have an understanding of the original data and the patterns that the needed schema, you may still have difficulty understanding the visualization. It may be that the visualization was not designed with an understanding of the original data and the patterns that the needed schema. Third, even if you have an understanding of the original data and the patterns that the needed schema, you may still find it difficult to understand the visualization. It may be that the visualization was not designed with an understanding of the original data and the patterns that the needed schema. Fourth, even if you have an understanding of the original data and the patterns that the needed schema, you may still find it difficult to understand the visualization. It may be that the visualization was not designed with an understanding of the original data and the patterns that the needed schema. Fifth, even if you have an understanding of the original data and the patterns that the needed schema, you may still find it difficult to understand the visualization. It may be that the visualization was not designed with an understanding of the original data and the patterns that the needed schema.

Use familiar designs to show data. The bottom of Figure 15 shows the same bar graph of total House and Senate campaign expenditures, in millions. It depicts the GDP (on a log scale) and population for the 10 most populous countries. Take a moment to understand the data. It is important to understand the data. The graph at the bottom was created by Nigel Holmes and was reprinted in his 1984 book. Used with permission.

[Helgeson & Moriarty, 1993] and [N. Holmes, 2014] and [S. Franconeri et al., 2021]
Memorability: Maps instead of Networks

Map-based Visualizations Increase Recall Accuracy of Data

Bahador Saket, Carlos Scheidegger, Stephen G. Kobourov, and Katy Börner

1 Department of Computer Science, University of Arizona, Tucson, AZ, USA
2 Department of Information and Library Science, Indiana University, Bloomington, IN, USA

Figure 1: We investigate the memorability of relational data represented with node-link (left-side) and map-based (right-side) visualizations; shown are a node-link and a map-based visualization with 200 nodes and 500 links from the LastFM dataset.

Abstract

We investigate the memorability of data represented in two different visualization designs. In contrast to recent studies that examine which types of visual information make visualizations memorable, we examine the effect of different visualizations on time and accuracy of recall of the displayed data, minutes and days after interaction with the visualizations. In particular, we describe the results of an evaluation comparing the memorability of two different visualizations of the same relational data: node-link diagrams and map-based visualization. We find significant differences in the accuracy of the tasks performed, and these differences persist days after the original exposure to the visualizations. Specifically, participants in the study recalled the data better when exposed to map-based visualizations as opposed to node-link diagrams. We discuss the scope of the study and its limitations, possible implications, and future directions.

1. Introduction

Researchers have long recognized that the visual display of information can be more effective than tables and numeric summaries [Ans73]. We also know that different visual designs offer significantly different reading precision [CM84]. In contrast, we do not understand nearly as well the memorability of the data that underlies the visualization. Is the design of a visualization responsible for how well users will remember its content?

In this paper, we present evidence that different visual designs can impact the recall accuracy of the data being visualized. Several recent studies have tested the memorability of different types of visualizations [BMG10, BARM12, MPWG12, VMTW12, IXTO11, BVB13]. These seminal studies focused on which types of visual information are memorable [BVB13]. To the best of our knowledge, no study has yet been performed to assess long-term memorability of the underlying data represented in these visualizations.

In this paper, we focus on two alternative visualizations for relational data. Specifically, we compare node-link visualizations to map-based visualizations. Node-link visualizations date back to 1735 and are a standard way of depicting relational datasets. In node-link diagrams, entities are depicted as points (typically dots or circles) in low-dimensional space, and two related entities are connected with a curve (typically a straight-line segment). Cluster membership is typically indicated by filling each circle with a color that is unique for each cluster.
Memorability: Maps instead of Networks

D. Koop, CSCI 627/490, Fall 2023
Memorability: Maps instead of Networks

In this paper, we focus on two alternative visualizations for relational data. Specifically, we compare different visualizations of the same relational data: node-link diagrams and map-based visualization. We find significant differences in the accuracy of the tasks performed, and these differences persist days after the original exposure to the visualizations. Specifically, participants in the study recalled the data better when exposed to different visualizations of the same relational data: node-link diagrams and map-based visualization. We find different visualizations on time and accuracy of recall of the data.

Researchers have long recognized that the visual display of information can be more effective than tables and numeric summaries. These seminal studies focused on which types of visual information are memorable. To the best of our knowledge, no study has yet been performed to assess long-term memorability of the data that underlies the visualization.

In contrast, we do not understand nearly as well the possible implications, and future directions.

In this paper, we present evidence that different visual designs can impact the recall accuracy of the data being visualized. Several recent studies have tested the memorability of the visualizations; shown are a node-link and a map-based visualization with 200 nodes and 500 links from the LastFM dataset.

Figure 1: Memorability: Maps instead of Networks

[B. Saket et al., 2015]
ISOTYPE Visualizations

• Study [Haroz et al., 2015]
  - Want quick understanding and ease of remembering
  - Does ISOTYPE help?

• Results:
  - Stacked icons allow both length and quantity encoding
  - Icons are more memorable
  - Images that aren’t used to show data are distracting

[Image by O. and M. Neurath, Study by S. Haroz et al., 2015]
Memorability

- Capability of maintaining and retrieving information
  [J. Brown et al., 1977]
- How to measure?
  - test users
- How long?
  - short-term, intermediate, or long-term?
- What types of visualizations?
  - bar/line/pie, networks, graphs, etc.
Engagement

• "Emotional, cognitive and behavioral connection that exists, at any point in time and possibly over time, between a user and a resource." [S. Attfield et al., 2011]

• How to measure? total time spent looking at a chart
Measuring Engagement

Grid is blurred, click for detail

(A) Tool Use For Typical Repairs  Cup Use in Restaurants  Dessert Preferences

Large Animals  Hat Purchases in June  Mammals

Knight’s Weapons  Organisms on One Island  Preference for Fun Gliding

(B) Mammals
Mammals are distinguished from reptiles and birds by the possession of hair, three middle ear bones, mammary glands in females, and a neocortex (a region of the brain).

[S. Haroz et al., 2015]
We ran 500 subjects on Amazon Mechanical Turk in 200 trials (5 chart types × 2 questions × 20 repetitions) blocked by chart type. Each subject was paid 8 US Dollars for the 30-minute study, and all participants were from the USA.

Exp 4 Results
All subjects showed over 92% accuracy, allowing incorrect responses to be dropped from analysis without substantially affecting statistical power. We also collapsed across the 'More' vs 'Fewer' condition to yield approximately 40 trials per chart type per subject. As with the previous experiments, we analyzed the results within-subject to determine the performance relative to that of the simple bar charts. We found a main effect of graph type on response time ($F[4, 49]=20, p<0.05, \eta^2_p=0.02$). A Tukey HSD-corrected comparison of all the graph types found that only the superfluous condition was significantly different from the standard bar graph ($p<0.05$) as can be seen in Fig. 13.

This result combined with the results of experiment 1 show that superfluous images hurt both memorability and speed of usability of charts.

EXPERIMENT 5: INITIAL ENGAGEMENT
Although speed can be an important benchmark, the aim of some visualizations is to make people pause and look—as is often the case in news articles. Designers often rely on pictographs because they are thought to draw the attention of a reader. When perusing through a collection of articles, an enticing visualization may increase the likelihood that an article will be inspected more closely. Will an ISOTYPE visualization be better at capturing attention than a simple bar chart?

We ran an experiment that simulated how visualizations are commonly encountered in a peripheral glimpse, as thumbnails among a collection of text and other visualizations competing for interest.

Exp 5 Methods
Subjects were presented with a 3x3 grid of items (Fig. 14). Each item included a short title above a small, slightly blurred thumbnail. The thumbnail was either a set of sentences about the topic from Wikipedia or a chart related to the topic. The subjects were given two minutes to look through the thumbnails. They could click whichever item interested them to view the information in full screen without pixilation or blur. Clicking again returned them to the grid, where they could repeat the process. No limit was placed on the number or duration of views for each item. However, after the trial's time had finished, everything was removed from the screen. They were then presented with a button to begin the next trial.

We selected 36 topics from the previous experiments' categories and constructed text, a bar chart, and a stacked pictograph chart for each. Throughout the experiment, each subject encountered each topic exactly once (9 items × 4 trials). A trial included 3 bar charts, 3 stacked pictograph charts, and 3 pieces of text. We tracked the start time and duration of each view.

10 subjects (4 women) participated in this experiment. Because it was implemented as a Windows desktop application, it was run in the lab. All subjects were undergraduates and were paid 5 US dollars for the 15 minute duration.

Exp 5 Results
We binned the first minute of viewing into one-second intervals and found the portion of subjects viewing each type of item. Fig. 15 shows a linear fit of these results collapsed across trial. For the first few seconds, most are at the selection grid. However, the ISOTYPE visualization takes a quick (A) Fig. 14. (A) An example of the selection grid for experiment 5. The title is readable, but the details of the content are unrecognizable beyond the type of information. (B) An example text display that can also been seen in the middle right of the selection grid in (A). [S. Haroz et al., 2015]
Enjoyment: Name Voyager

Measuring Enjoyment

- Difference from engagement (e.g. may be for a job)
- Self-reporting (e.g. comparison between different charts)
- Measure why someone enjoys a visualization:
  - Challenge
  - Focus
  - Clarity
  - Feedback
  - Control
  - Immersion

[B. Saket et al., 2016]
“Visualizations don’t need to be designed for memorability – they need to be designed for comprehension. For most visualizations, the comprehension that they provide need only last until the decision that it informs is made. Usually, that is only a matter of seconds.”

— S. Few
Reaction

• B. Jones (paraphrased): People make decisions using visualizations but this isn't instantaneous like robots or algorithms; they often chew on a decision for a while

• R. Kosara: there are cases where people benefit from remembering a visualization (e.g. health-related visualization)

• Are there tradeoffs between the characteristics?
Finally, the format of a visualization can also guide the types of conclusions that viewers draw from the underlying data. Imagine data showing that students who eat breakfast more often tend to have higher GPAs. A viewer might see this correlation and assume a causal relationship whereby a good breakfast causes better grades. Although plausible, this conclusion cannot be drawn from these data. When shown visualizations like these, viewers made unwarranted claims about similar correlational data, and they did so more often when the visualizations aggregated the data into fewer groups (e.g., a two-bar graph), compared with more groups (e.g., a scatterplot showing all of the individual data values; Xiong, Shapiro, et al., 2020), perhaps because seeing the data in fewer groups is implicitly associated with those data being gathered by an experimental manipulation.

Avoid taxing limited working memory

Given that comparisons are already highly capacity limited, any extraneous demands on working memory due to the design of visualizations should be avoided. Interpreting the graphs in the middle and right sides of Figure 13 requires individuals to map the symbols and colors in the graphs to their referents in the legends below. This task is highly demanding of limited working memory resources. If information is lost in interpreting a graph, viewers might make interpretation errors or require extra time to reinspect the legend. Indeed, one...

Fig. 12. An example of emphasizing different perspectives in a single data set (inspired by Bostock et al., 2012). One data set can be seen with dramatically different perspectives, depending on which patterns an observer does and does not extract.

Fig. 13. A demonstration of the advantage of direct labels over legends. Take a moment to state the names of the four groups shown in the line graph at left in top-to-bottom order. (Answer: b, d, a, c.) Now do the same for the graphs at center and right, which require coordination with color and shape legends. You should notice a substantial slowdown because of the need to frequently look back and forth between the graph and the legend. If you attempt to memorize the legend first, you will experience the capacity limit of your working memory.

[S. Franconeri et al., 2019]
Present to Persuade

**Fig. 13.** A demonstration of the advantage of direct labels over legends. Take a moment to state the names of the four groups shown in the line graph at left in top-to-bottom order. (Answer: b, d, a, c.)
Influencing Messages in Visualizations

• Perception is influenced by existing biases (e.g. unemployment numbers)
• Perception is influenced by visualization's title [Kong et al., 2019]
• Perception can be biased by social influence [Hullman et al., 2011]
• See A. Cairo's books
Visualization for Production

• Generate new material
• Annotate
• Record
• Derive (Transform)
Annotation: Circle Annotations
Record: Provenance of MTA Data Exploration

Initial data
- Station locations
  - Station map
  - Added fares
    - Full fares map
    - Difference
  - Broadway line
  - August 16
    - Broadway diff map

Corrected data
- November ff
  - November 2 data
  - August 16 Tab

Filtered
- Concourse line
- Heatmap

Sum of ffs
- 30-D weekly
- 161st-River

D. Koop, CSCI 627/490, Fall 2023
Derived Data

Original Data

Derived Data

trade balance = exports − imports

[Munzner (ill. Maguire), 2014]
Visualization for Production

- Generate new material
- Annotate:
  - Add more to a visualization
  - Usually associated with text, but can be graphical
- Record:
  - Persist visualizations for historical record
  - Provenance (graphical histories): how did I get here?
- Derive (Transform):
  - Create new data
  - Create derived attributes (e.g. mathematical operations, aggregation)
Actions: Search

• What does a user know?
  - Lookup: check bearings
  - Locate: find on a map
  - Browse: what’s nearby
  - Explore: where to go
  - Patterns

<table>
<thead>
<tr>
<th></th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td><img src="image" alt="Lookup" /></td>
<td><img src="image" alt="Browse" /></td>
</tr>
<tr>
<td>Location unknown</td>
<td><img src="image" alt="Locate" /></td>
<td><img src="image" alt="Explore" /></td>
</tr>
</tbody>
</table>
Query

- Number of targets: One, Some (Often 2), or All
- Identify: characteristics or references
- Compare: similarities and differences
- Summarize: overview of everything

[Munzner (ill. Maguire), 2014]
Targets

- **ALL DATA**
  - Trends
  - Outliers
  - Features

- **ATTRIBUTES**
  - One
    - Distribution
    - Extremes
  - Many
    - Dependency
    - Correlation
    - Similarity

- **NETWORK DATA**
  - Topology
    - Paths

- **SPATIAL DATA**
  - Shape

[McNzner (ill. Maguire), 2014]
Roadmap

What? → Data
- Types
- Semantics

Why? → Tasks
- Actions
- Targets

How → Vis Idioms/Techniques
- Data Representation
- Visual Encoding
- Interaction Encoding
Analysis Example: Different “Idioms”

[SpaceTree, Grosjean et al.]

[TreeJuxtaposer, Munzner et al.]
"Idiom" Comparison

**SpaceTree**

What?
- Tree

Why?
- Actions
  - Present
  - Locate
  - Identify

How?
- Targets
  - Path between two nodes

**TreeJuxtaposer**

What?
- Tree

Why?
- Actions
  - Encode
  - Navigate
  - Select
  - Filter
  - Aggregate

How?
- Targets
  - Encode
  - Navigate
  - Select
  - Arrange

[Munzner (ill. Maguire), 2014]


Analysis Example: Derivation

• Strahler number
  – centrality metric for trees/networks
  – derived quantitative attribute
  – draw top 5K of 500K for good skeleton


Task 1

- In Tree
- Out Quantitative attribute on nodes

What?
- In Tree
- Out Quantitative attribute on nodes

Why?
- Derive

Task 2

- In Tree
- Out Quantitative attribute on nodes

What?
- In Tree
- In Quantitative attribute on nodes

Why?
- Derive

How?
- Summarize
- Topology
- Filter

Out Filtered Tree

Removed unimportant parts

[Munzner (ill. Maguire), 2014]