Data Visualization (CSCI 627/490)

Tasks & Design

Dr. David Koop
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**
  - Items (rows)
  - Attributes (columns)
  - Cell containing value

- **Networks**
  - Nodes (item)
  - Links

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

- **Geometry (Spatial)**
  - Position

- **Multidimensional Table**

- **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.

Sets & Lists

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

Raw Lyrics Data via John W. Miller

[M. Daniels, 2019]
### 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>Order ID</td>
<td>Order Date</td>
<td>Order Priority</td>
<td>Product Container</td>
<td>Product Base Margin</td>
<td>Ship Date</td>
</tr>
<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>0.6</td>
<td>6/6/05</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Large Box</td>
<td>0.59</td>
<td>12/23/06</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.82</td>
<td>12/23/06</td>
</tr>
<tr>
<td>96</td>
<td>4/17/05</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.55</td>
<td>4/19/05</td>
</tr>
<tr>
<td>97</td>
<td>1/29/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.38</td>
<td>1/30/06</td>
</tr>
<tr>
<td>129</td>
<td>11/19/08</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.37</td>
<td>11/28/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.37</td>
<td>5/9/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.38</td>
<td>5/10/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.6</td>
<td>5/11/08</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>6/12/06</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Jumbo Box</td>
<td>0.59</td>
<td>6/14/06</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>Small Box</td>
<td>0.55</td>
<td>9/14/07</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**
A spiral is easy to describe and understand in polar coordinates, i.e. in the form $r = f(\phi)$. The distinctive feature of a spiral is that $f$ is a monotone function. In this work we assume a spiral is described by $r = f(\phi)$.

Several simple functions $f$ lead to well-known types of spirals:

- **Archimedes' spiral** has the form $r = a\phi$. It has the special property that a ray emanating from the origin crosses two consecutive arcs of the spiral in a constant distance.

- The **Hyperbolic spiral** has the form $r = a\phi^b$. It is the inverse of Archimedes' spiral with respect to the origin.

- More generally, spirals of the form $r = a\phi^k$ are called **Archimedean spirals**.

- The **logarithmic spiral** has the form $r = ke^{\phi/b}$. It has the special property that all arcs cut a ray emanating from the origin under the same angle.

For the visualization of time-dependent data Archimedes' spiral seems to be the most appropriate. In most applications data from different periods are equally important. This should be reflected visually in that the distance to other periods is always the same.

### Ordering Direction

- **Sequential**

- **Diverging**

- **Cyclic**

[Munzner (ill. Maguire), 2014; Rogowitz & Treinish, 1998; Weber et al., 2001]
Assignment 2

- Due Sept. 19
- Process Data
- Create Bar Charts using SVGs and JavaScript
- Interaction: Select by Decade
Tasks

What?

Why?

How?

<table>
<thead>
<tr>
<th>Actions</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyze</td>
<td>All Data</td>
</tr>
<tr>
<td>→ Consume</td>
<td>→ Trends</td>
</tr>
<tr>
<td>→ Discover</td>
<td>→ Outliers</td>
</tr>
<tr>
<td>→ Present</td>
<td>→ Features</td>
</tr>
<tr>
<td>→ Enjoy</td>
<td></td>
</tr>
<tr>
<td>Produce</td>
<td>Attributes</td>
</tr>
<tr>
<td>→ Annotate</td>
<td>→ One</td>
</tr>
<tr>
<td>→ Record</td>
<td>→ Many</td>
</tr>
<tr>
<td>→ Derive</td>
<td>→ Distribution</td>
</tr>
<tr>
<td>→ Extremes</td>
<td>→ Dependency</td>
</tr>
<tr>
<td>Search</td>
<td>→ Correlation</td>
</tr>
<tr>
<td>Target known</td>
<td>→ Similarity</td>
</tr>
<tr>
<td>Target unknown</td>
<td></td>
</tr>
<tr>
<td>Location known</td>
<td></td>
</tr>
<tr>
<td>→ Lookup</td>
<td></td>
</tr>
<tr>
<td>→ Browse</td>
<td></td>
</tr>
<tr>
<td>Location unknown</td>
<td></td>
</tr>
<tr>
<td>→ Locate</td>
<td></td>
</tr>
<tr>
<td>→ Explore</td>
<td></td>
</tr>
<tr>
<td>Query</td>
<td>Network Data</td>
</tr>
<tr>
<td>→ Identify</td>
<td>→ Topology</td>
</tr>
<tr>
<td>→ Compare</td>
<td>→ Paths</td>
</tr>
<tr>
<td>→ Summarize</td>
<td>→ Spatial Data</td>
</tr>
<tr>
<td></td>
<td>→ Shape</td>
</tr>
</tbody>
</table>

[Munzner (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
Each solid circle represents a bee species active in Carlinville, III., 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

“[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]
Memorability

Figure 6: Policy skills and interventions to solve the haze problem.

Memorable

Low Quality Description

Forgettable

[Memorability: M. Borkin et al., InfoVis 2015]
Memorability: Maps instead of Networks

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 [BMG†10, BARM†12, MPWG12, VMTW†12, IXTO11, BVB†13]. These seminal studies focused on which types of visual information are memorable [BVB†13]. 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

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]. 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.

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]. 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.

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]. 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.

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]. 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.

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.
Memorability: Maps instead of Networks

[Figure: Shows node-link and map-based visualizations of a network.]

[Caption: B. Saket et al., EuroVis 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

We ran 50 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)

![Image](A.png)

(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.

EXP 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?

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 inter-values 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 be seen in the middle right of the selection grid in (A) Fig. 14. ISOTYPE charts are best at initially engaging subjects to inspect information more closely.

Measuring Engagement

[S. Haroz et al., 2015]
Enjoyment: Name Voyager

Names starting with 'AN' per million babies

[Graph showing popularity of names starting with 'AN' over time]

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., BELIV 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
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?
Visualization for Production

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

initial data
- corrected data
  - November 17 data
  - November 2 data
  - August 16 Tab
- full fares map
  - Broadway line
  - August 16
- added fares
- station locations
- difference

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

with labels
- filtered
- heatmap

broadway diff map

D. Koop, CSCI 627/490, Fall 2022
Derived Data

Original Data

Derived Data

trade balance = \textit{exports} - \textit{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>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td><em>Lookup</em></td>
<td><em>Browse</em></td>
</tr>
<tr>
<td>Location unknown</td>
<td><em>Locate</em></td>
<td><em>Explore</em></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

[Note: The diagram includes various visual representations of data trends, outliers, features, distribution, dependency, correlation, similarity, topology, paths, and shape.]

[Munzner (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**

**TreeJuxtaposer**

**What?**
- Tree
- Actions
  - Present
  - Locate
  - Identify
- Targets
  - Path between two nodes

**Why?**
- SpaceTree
- Encode
- Navigate
- Select
- Filter
- Aggregate

- TreeJuxtaposer
- Encode
- Navigate
- Select
- Arrange

**How?**

[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

- **What?**
  - In Tree
  - Out Quantitative attribute on nodes

- **Why?**
  - Derive

Task 2

- **What?**
  - In Tree + In Quantitative attribute on nodes

- **Why?**
  - Summarize

- **How?**
  - Reduce

- **What?**
  - Out Filtered Tree

- **Why?**
  - Topology

- **How?**
  - Filter

[Munzner (ill. Maguire), 2014]