Data Visualization (CSCI 627/490)

Data & Tasks

Dr. David Koop
SVG Manipulation Example

- Draw a horizontal bar chart
  - `var a = [6, 2, 6, 10, 7, 18, 0, 17, 20, 6];`

- Steps:
  - Programmatically create SVG
  - Create individual rectangle for each item

- Link:
  - [https://codepen.io/dakoop/pen/mdbxQKe](https://codepen.io/dakoop/pen/mdbxQKe)
Data

• **What is this data?**

<table>
<thead>
<tr>
<th>ID</th>
<th>Location</th>
<th>Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>R011</td>
<td>42ND STREET &amp; 8TH AVENUE</td>
<td>00228985, 00008471, 00000441, 00001455, 00000134, 00033341, 00071255</td>
</tr>
<tr>
<td>R170</td>
<td>14TH STREET-UNION SQUARE</td>
<td>00224603, 00011051, 00000827, 00003026, 00000660, 00089367, 00199841</td>
</tr>
<tr>
<td>R046</td>
<td>42ND STREET &amp; GRAND CENTRAL</td>
<td>00207758, 00007908, 00000323, 00001183, 00003001, 00040759, 00096613</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
Assignment 2

• Link
• Three parts: table, horizontal bar chart, vertical bar chart
  - data processing
  - highlighting (CSCI 627)
• Vertical chart can be tricky
• Start early!
• Questions?
**Dataset Types**

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

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

- **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]
Table Visualizations

[Table diagrams showing relationships between various car specifications such as economy (mpg), cylinders, displacement (cc), power (hp), weight (lb), 0-60 mph (s), and year.]

[M. Bostock, 2011]
Networks

• Why networks instead of graphs?
• Tables can represent networks
  - Many-many relationships
  - Also can be stored as specific graph databases or files
Figure 7: US airlines graph (235 nodes, 2101 edges) (a) not bundled and bundled using (b) FDEB with inverse-linear model, (c) GBEB, and (d) FDEB with inverse-quadratic model.

Figure 8: US migration graph (1715 nodes, 9780 edges) (a) not bundled and bundled using (b) FDEB with inverse-linear model, (c) GBEB, and (d) FDEB with inverse-quadratic model. The same migration flow is highlighted in each graph.

Figure 9: A low amount of straightening provides an indication of the number of edges comprising a bundle by widening the bundle. (a) \( s = 0 \), (b) \( s = 10 \), and (c) \( s = 40 \). If \( s \) is 0, color more clearly indicates the number of edges comprising a bundle.

we generated use the rendering technique described in Section 4.1. To facilitate the comparison of migration flow in Figure 8, we use a similar rendering technique as the one that Cui et al. [CZQ 08] used to generate Figure 8c.

The airlines graph is comprised of 235 nodes and 2101 edges. It took 19 seconds to calculate the bundled airlines graphs (Figures 7b and 7d) using the calculation scheme presented in Section 3.3. The migration graph is comprised of 1715 nodes and 9780 edges. It took 80 seconds to calculate the bundled migration graphs (Figures 8b and 8d) using the same calculation scheme. All measurements were performed on an Intel Core 2 Duo 2.66GHz PC running Windows XP with 2GB of RAM and a GeForce 8800GT graphics card.

Our prototype was implemented in Borland Delphi 7.

© 2009 The Author(s)

Journal compilation © 2009 The Eurographics Association and Blackwell Publishing Ltd.
Figure 7: US airlines graph (235 nodes, 2101 edges) (a) not bundled and bundled using (b) FDEB with inverse-linear model, (c) GBEB, and (d) FDEB with inverse-quadratic model.

Figure 8: US migration graph (1715 nodes, 9780 edges) (a) not bundled and bundled using (b) FDEB with inverse-linear model, (c) GBEB, and (d) FDEB with inverse-quadratic model. The same migration flow is highlighted in each graph.

Figure 9: A low amount of straightening provides an indication of the number of edges comprising a bundle by widening the bundle. (a) \(s = 0\), (b) \(s = 10\), and (c) \(s = 40\). If \(s\) is 0, color more clearly indicates the number of edges comprising a bundle.

We generated these graphs using the rendering technique described in Section 4.1. To facilitate the comparison of migration flow in Figure 8, we use a similar rendering technique as the one that Cui et al. \[CZQ\]\(^\leftrightarrow\)08 used to generate Figure 8c.

The airlines graph is comprised of 235 nodes and 2101 edges. It took 19 seconds to calculate the bundled airlines graphs (Figures 7b and 7d) using the calculation scheme presented in Section 3.3. The migration graph is comprised of 1715 nodes and 9780 edges. It took 80 seconds to calculate the bundled migration graphs (Figures 8b and 8d) using the same calculation scheme. All measurements were performed on an Intel Core 2 Duo 2.66GHz PC running Windows XP with 2GB of RAM and a GeForce 8800GT graphics card.

Our prototype was implemented in Borland Delphi 7. ©2009 The Author(s)

Networks

[Holten & van Wijk, 2009]
Each point in space has an associated...

Scalar Fields
(Order-0 Tensor Fields)

Vector Fields
(Order-1 Tensor Fields)

Tensor Fields
(Order-2+)

\[ s_0 \]
Scalar

\[
\begin{bmatrix}
  v_0 \\
  v_1 \\
  v_2
\end{bmatrix}
\]
Vector

\[
\begin{bmatrix}
  \sigma_{00} & \sigma_{01} & \sigma_{02} \\
  \sigma_{10} & \sigma_{11} & \sigma_{12} \\
  \sigma_{20} & \sigma_{21} & \sigma_{22}
\end{bmatrix}
\]
Tensor
Fields

• Difference between *continuous* and *discrete* values
• Examples: temperature, pressure, density
• **Grids** necessary to sample continuous data:

  ![Uniform grid](image1)
  ![Rectilinear grid](image2)
  ![Structured grid](image3)
  ![Unstructured grid](image4)

  - uniform
  - rectilinear
  - structured
  - unstructured

• **Interpolation**: “how to show values between the sampled points in ways that do not mislead”
Spatial Data Example: MRI
Scivis and Infovis

• Two subfields of visualization
• **Scivis** deals with data where the spatial position is given with data
  - Usually continuous data
  - Often displaying physical phenomena
  - Techniques like isosurfacing, volume rendering, vector field vis
• In **Infovis**, the data has no set spatial representation, designer chooses how to visually represent data
SciVis

[Google Image Search for "scientific visualization", 2017]
InfoVis

[Google Image Search for "information visualization", 2017]
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 might be those of pop artists, such as Beyonce or even one of his major influences: Marilyn Manson. 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.

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

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

- **Country**: 2,800 unique words
- **Rock**: 3,800 unique words
- **Hip Hop**: 4,800 unique words

Raw Lyrics Data via John W. Miller

[M. Daniels, 2019]
35,000 words covers 3 to 5 studio albums and EPs. I included mixtapes if the artist was short of the 35,000 words. Quite a few rappers don't have enough of official material to be included (for example, Biggie, Chance the Rapper, Queen Latifah, and El-P).

Since the original release, there's now a notable trend of fewer unique words among newer artists. This is easier to see in the following chart, where I highlighted each artist's primary decade, based on album release dates for their vocabulary calculation:

---

Notes/sources:
All lyrics are via Genius.

[M. Daniels, 2019]
Some of the newer artists wield a smaller vocabulary comparatively, but this is not because hip hop has "dumbed down." The genre has evolved; it has moved away from complex lyricism toward elements traditionally associated with pop music: repetitive

Note/sources:
(1) Since this analysis uses an artist's first 35,000 lyrics (prioritizing studio albums), an artist's era is determined by the years the albums were released. Some artists may be identified with a certain era (for example, Jay-Z with the 1990s, with Reasonable Doubt in 1996, In My Lifetime, Vol. 1 in 1997, etc.) yet continue to release music in the present day. All lyrics are via Genius.

# of Unique Words Used Within Artist’s First 35,000 lyrics

<table>
<thead>
<tr>
<th>Era</th>
<th>Artists</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980s</td>
<td>DMX, 21 Savage, A Boogie wit...</td>
</tr>
<tr>
<td></td>
<td>Bone Thugs-n-Harmony, 50 Cent, Juicy J, Drake, Future, Kid Cudi, Kid Ink, Kodak Black, Lil Wayne, Missy Elliott, Trick Daddy, Trina, Young Jeezy, Big Sean, BoB, Childish Gambino, G-Eazy, J Cole, Machine Gun Kelly, Meek Mill, Nicki Minaj, Russ</td>
</tr>
<tr>
<td>1990s</td>
<td>Bone Thugs-n-Harmony, DMX, 21 Savage, A Boogie wit...</td>
</tr>
<tr>
<td></td>
<td>Common, Common, Das EFX, E-40, Goodie Mob, Nas, Redman, Brother Ali, Action Bronson, KAAN</td>
</tr>
<tr>
<td>2000s</td>
<td>Bone Thugs-n-Harmony, DMX, 21 Savage, A Boogie wit...</td>
</tr>
<tr>
<td>2010s</td>
<td>Bone Thugs-n-Harmony, DMX, 21 Savage, A Boogie wit...</td>
</tr>
<tr>
<td></td>
<td>Beastie Boys, Big Daddy Kane, LL Cool J, Busta Rhymes, Cypress Hill, De La Soul, Fat Joe, Gang Starr, KRS-One, Method Man, A Tribe Called Quest, Atmosphere, Ludacris, Lupe Fiasco, Mos Def, Murs, Talib Kweli, Xzibit, Flatbush Zombies, Joey BadA$$, Rittz</td>
</tr>
</tbody>
</table>
Attribute Types

- Categorical
- Ordered
- Ordinal
- Quantitative

[Munzner (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>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>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Wrap Bag</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>Small 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>Large 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>Large 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>6/1/08</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>6/12/06</td>
</tr>
<tr>
<td>132</td>
<td>6/1/08</td>
<td>3-Medium</td>
<td>Jumbo Box</td>
<td>0.69</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>
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>10/23/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.49</td>
<td>3/19/07</td>
</tr>
<tr>
<td>32</td>
<td>10/13/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.55</td>
<td>10/23/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>5-Low</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>Small Pack</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>35</td>
<td>10/23/07</td>
<td>4-Not Specified</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>Medium Box</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>Medium 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 Box</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>90</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>Medium 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>Medium Box</td>
<td>0.69</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
Data Model vs. Conceptual Model

• Data Model: raw data that has a specific data type (e.g. floats):
  - Temperature Example: [32.5, 54.0, -17.3] (floats)

• Conceptual Model: how we think about the data
  - Includes semantics, reasoning
  - Temperature Example:
    • Quantitative: [32.50, 54.00, -17.30]
Data Model vs. Conceptual Model

• Data Model: raw data that has a specific data type (e.g. floats):
  - Temperature Example: [32.5, 54.0, -17.3] (floats)

• Conceptual Model: how we think about the data
  - Includes semantics, reasoning
  - Temperature Example:
    • Quantitative: [32.50, 54.00, -17.30]
    • Ordered: [warm, hot, cold]

[via A. Lex, 2015]
Data Model vs. Conceptual Model

- **Data Model**: raw data that has a specific data type (e.g., floats):
  - Temperature Example: [32.5, 54.0, -17.3] (floats)

- **Conceptual Model**: how we think about the data
  - Includes semantics, reasoning
  - Temperature Example:
    - Quantitative: [32.50, 54.00, -17.30]
    - Ordered: [warm, hot, cold]
    - Categorical: [not burned, burned, not burned]

[via A. Lex, 2015]
Ordering Direction

- Sequential
- Diverging
- Cyclic

[Munzner (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]
Cyclic Data

For the visualization of time-dependent data Archimedes spirals can be used. The logarithmic spiral has the form \( r = a e^{\phi/a} \), which has the special property that all arcs cut a ray emanating from the origin under the same angle. The hyperbolic spiral has the form \( r = \frac{a}{\phi + b} \). It is the inverse of Archimedes spirals.

In general, markers, bars, and line elements can be used for the visualization of time-dependent data. The choice of the best type of visualization should be reflected visually in that the distance to other periods are equally important. This is mainly the case when the data is always the same. Yet, we have found this way of visualizing to be inefficient in certain cases, especially when this is not the case. In such a scenario, other attributes should be used, such as colour, texture, including line styles and patterns.

Discrete data can be presented as bars on the spiral or by marks with a corresponding distance to the spiral. However, the use is limited for the display of data from different periods are equally important. The logarithmic spiral seems to be the most appropriate. In most applications the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate.

The continuity of the data is expressed by using a spiral in the form of the spiral their use is limited for the display of data. The logarithmic spiral can be described by \( r = e^{\phi} \). The hyperbolic spiral has the form \( r = \frac{a}{\phi + b} \). It is the inverse of Archimedes spirals. Archimedes spirals have the form \( r = a \phi \).

A spiral is easy to describe and understand in polar coordinates. Several simple functions are always the same.

The inverse of Archimedes spirals is the form of the spiral. The logarithmic spiral has the form \( r = a e^{\phi/a} \). The hyperbolic spiral has the form \( r = \frac{a}{\phi + b} \). It is the inverse of Archimedes spirals. Archimedes spirals have the form \( r = a \phi \).

Spirals can be used for the visualization of time-series data similar to standard point, bar, and line graphs on Spiral Graphs. For instance, quantitative, qualitative, and categorical data can be shown in this manner.

For the visualization of time-dependent data Archimedes spirals can be used. The logarithmic spiral has the form \( r = a e^{\phi/a} \), which has the special property that all arcs cut a ray emanating from the origin under the same angle. The hyperbolic spiral has the form \( r = \frac{a}{\phi + b} \). It is the inverse of Archimedes spirals.

In general, markers, bars, and line elements can be used for the visualization of time-dependent data. The choice of the best type of visualization should be reflected visually in that the distance to other periods are equally important. This is mainly the case when the data is always the same. Yet, we have found this way of visualizing to be inefficient in certain cases, especially when this is not the case. In such a scenario, other attributes should be used, such as colour, texture, including line styles and patterns.

Discrete data can be presented as bars on the spiral or by marks with a corresponding distance to the spiral. However, the use is limited for the display of data from different periods are equally important. The logarithmic spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate. In most applications, the spiral seems to be the most appropriate.
“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?

Action:
- Analyze
  - Consume
    - Discover
    - Present
    - Enjoy
  - Produce
    - Annotate
    - Record
    - Derive

Search:

<table>
<thead>
<tr>
<th>Location known</th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td>&quot;.&quot; &quot;Lookup&quot;</td>
<td>&quot;.&quot; &quot;Browse&quot;</td>
</tr>
<tr>
<td>Location unknown</td>
<td>&quot;.&quot; &quot;Locate&quot;</td>
<td>&quot;.&quot; &quot;Explore&quot;</td>
</tr>
</tbody>
</table>

Query:
- Identify
- Compare
- Summarize

Why?

Target:
- All Data
  - Trends
  - Outliers
  - Features
- Attributes
  - One
    - Distribution
  - Many
    - Dependency
    - Correlation
    - Similarity
- Network Data
  - Topology
    - Paths
- Spatial Data
  - Shape

[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, 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

“[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 shifts and interventions to middle entailed prudent in accomodation instead of emergency action and frame health.

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

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, BARM, MPWG, VMTW, IXTO, BVB]. These seminal studies focused on which types of visual information are memorable [BVB]. 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.
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 [BMG10, BARM12, MPWG12, VMTW12, IXTO11, BV13]. These seminal studies focused on which types of visual information are memorable [BV13]. 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

Researchers have long recognized that the visual display of data can influence how well users will remember summaries or summaries of the data. In this paper, we present evidence that different visualizations of the same relational data: node-link diagrams and map-based visualizations. We find that map-based visualizations increase recall accuracy of data summaries. In particular, we describe the results of an evaluation comparing the memorability of two different visualizations on time and accuracy of recall of the underlying data represented in these visualizations. Specifically, we compare map-based visualizations as opposed to node-link diagrams. We discuss the scope of the study and its limitations, and future directions.

Abstract

We investigate the memorability of relational data represented with visualizations; shown are a node-link and a map-based visualization with 200 nodes and 500 links from the LastFM dataset.

Figure 1: Map-based Visualizations Increase Recall Accuracy of Data

In this paper, we focus on two alternative visualizations for relational data. Specifically, we compare map-based visualizations to node-link visualizations. To the best of our knowledge, no study has yet been performed to assess long-term memorability of data visualizations. These seminal studies focused on which types of visual information are possible implications, and future directions.

Figure 1: Map-based Visualizations Increase Recall Accuracy of Data

We also know that different visual designs can impact the recall accuracy of the data being visualized. In particular, we describe the results of an evaluation comparing the memorability of two different visualizations on time and accuracy of recall of the underlying data represented in these visualizations.

Acknowledgments

This work was supported by the National Science Foundation under Grant No. IIS-1528646. We would like to thank the anonymous reviewers for their valuable comments and suggestions.

References

[1] B. Saket et al., EuroVis 2015

Jonny Greenwood

Type O Negative

Behemoth

Massive Attack

Sahani

Archive

Apparat

Iron & Wine

Jonny Greenwood

BVB

CM84

VMTW

IXTO11

Ans73

D. Koop, CSCI 627/490, Fall 2020

Northern Illinois University
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
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

[B. Saket et al., BELIV 2016]
We ran 500 subjects on Amazon Mechanical Turk in 200 trials (5 chart types \(\times 2\) questions \(\times 20\) 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 \((F[4, 49] = 20, p < 0.05, \hat{\eta}^2 = 0.02)\). 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 \(\times 4\) 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)

![](https://via.placeholder.com/150)

**Grid is blurred, click for detail**

(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 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?

**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 be seen in the middle right of the selection grid in (A).

Fig. 15. ISOTYPE charts are best at initially engaging subjects to inspect information more closely.
Enjoyment: Name Voyager

Names starting with 'AN' per million babies

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
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?