Data Visualization (DSC 530/CIS 568)

Tasks & D3

Dr. David Koop
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]
Attribute Types

- **Categorical**
- **Ordered**
  - **Ordinal**
  - **Quantitative**

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

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
<td></td>
<td>B</td>
<td>C</td>
<td>S</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----------</td>
<td>----------------------</td>
<td>---</td>
<td>---</td>
</tr>
<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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></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>
<td></td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td></td>
<td>0.6</td>
<td>6/6/05</td>
<td></td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td></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></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></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></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></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>Small Box</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>Medium Box</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>Small Box</td>
<td>0.6</td>
<td>5/11/08</td>
<td></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>
<td></td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Jumbo Box</td>
<td>0.69</td>
<td>6/14/06</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>
<td></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>
<td></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>
<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>
<td></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>
<td></td>
</tr>
</tbody>
</table>

- **Quantitative**
- **Ordinal**
- **Categorical**
3.1. Mathematical description and types of spirals

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 $r = \frac{a}{\phi}$. It is the inverse of Archimedes’ spiral with respect to the origin.

- More generally, spirals of the form $r = f(\phi)$ are called Archimedean spirals.

- **The logarithmic spiral** has the form $r = e^{a\phi}$. 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.

3.2. Mapping data to the spiral

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 marks with a corresponding distance to the spiral. However, since the $x$ and $y$ coordinates are needed to achieve the general form of the spiral their use is limited for the display of data values. One might consider to map data values to small absolute changes in the radius, i.e. $\delta r = f(\phi)$.

Yet, we have found this way of visualizing to be ineffective. We conclude that the general shape of the spiral should be untouched and other attributes should be used, such as:

- **colour**,
- **texture**, including line styles and patterns,
- **size**,

Figure 1: Two visualizations of sunshine intensity using about the same screen real estate and 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.
Tasks

<table>
<thead>
<tr>
<th>Actions</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analyse</strong></td>
<td><strong>All Data</strong></td>
</tr>
<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><strong>Produce</strong></td>
<td><strong>Attributes</strong></td>
</tr>
<tr>
<td>➔ Produce</td>
<td>➔ Attributes</td>
</tr>
<tr>
<td>➔ Annotate</td>
<td>➔ One</td>
</tr>
<tr>
<td>➔ Record</td>
<td>➔ Distribution</td>
</tr>
<tr>
<td>➔ Derive</td>
<td>➔ Extreme</td>
</tr>
<tr>
<td></td>
<td>➔ Many</td>
</tr>
<tr>
<td></td>
<td>➔ Dependency</td>
</tr>
<tr>
<td></td>
<td>➔ Correlation</td>
</tr>
<tr>
<td></td>
<td>➔ Similarity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Targets</th>
<th>Why?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search</strong></td>
<td>➔ Network Data</td>
</tr>
<tr>
<td>➔ Search</td>
<td>➔ Network Data</td>
</tr>
<tr>
<td></td>
<td>➔ Topology</td>
</tr>
<tr>
<td></td>
<td>➔ Paths</td>
</tr>
<tr>
<td></td>
<td>➔ Spatial Data</td>
</tr>
<tr>
<td></td>
<td>➔ Shape</td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td>➔ Query</td>
</tr>
<tr>
<td>➔ Identify</td>
<td>➔ Query</td>
</tr>
<tr>
<td>➔ Compare</td>
<td>➔ Compare</td>
</tr>
<tr>
<td>➔ Summarize</td>
<td>➔ Summarize</td>
</tr>
<tr>
<td><strong>What?</strong></td>
<td>➔ What?</td>
</tr>
<tr>
<td><strong>Why?</strong></td>
<td>➔ Why?</td>
</tr>
<tr>
<td><strong>How?</strong></td>
<td>➔ How?</td>
</tr>
</tbody>
</table>
Actions: Analyze

→ Consume
  → Discover

→ Present
  → Enjoy

→ Produce
  → Annotate
  → Record
  → Derive

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

Discover

NameVoyager: Explore baby names and name trends letter by letter
Looking for the perfect baby name? Sign up for free to receive access to our expert tools!

Present

Enjoy

[M. Stefaner, M. Wattenberg]
Assignment 2

• Create a stacked bar chart using three different tools: Tableau, Vega-Lite, and D3

• Due Monday, Feb. 25
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

[Memorably FORGETTABLE]

Memorable

11

HIGH QUALITY DESCRIPTION

LOW QUALITY DESCRIPTION

FORGETTABLE

[1. Types of Preventative Interventions]

[60% of Americans feel Romney performed better than Obama in debates]

[D. Koop, DSC 530, Spring 2019]

[M. Borkin et al., InfoVis 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.

[B. Saket et al., BELIV 2016]
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]
Measuring Engagement

Grid is blurred, click for detail

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]
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, \( \eta_p^2=0.02 \)). 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?

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.

[S. Haroz et al., 2015]
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?
Visualization for Production

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

initial data
- corrected data
  - November FF
  - November 2 Data
  - August 16 Tab
  - Sum of FFS
  - 30-D Weekly
  - 161st-River

station locations
- Station Map
- Added Fares
- Difference

- Broadway Line
- August 16
- Broadway Diff Map

filtered
- with labels
- Concourse Line
- Heatmap

D. Koop, DSC 530, Spring 2019
Derived Data

Original Data

Derived Data

trade balance = \textit{exports} – \textit{imports}
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

#### Search

<table>
<thead>
<tr>
<th></th>
<th>Target known</th>
<th>Target unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location known</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

[Munzner (ill. Maguire), 2014]
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

[Munzner (ill. Maguire), 2014]
Analysis Example: Different “Idioms”

[SpaceTree, Grosjean et al.]  [TreeJuxtaposer, Munzner et al.]
“Idiom” Comparison

**SpaceTree**

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

**TreeJuxtaposer**

- **What?**
  - Tree
- **Why?**
  - Actions: Encode, Navigate, Select, Filter, Aggregate
  - Targets: Path between two nodes

**Comparison**


[Munzner (ill. Maguire), 2014]
“Idiom” Comparison

**SpaceTree**

- **Actions**
  - Present
  - Locate
  - Identify
- **Targets**
  - Path between two nodes

**TreeJuxtaposer**

- **Actions**
  - Encode
  - Navigate
  - Select
  - Filter
  - Aggregate
- **Targets**
  - Path between two nodes

**Why?**

- SpaceTree
- TreeJuxtaposer

**How?**

- Encode
- Navigate
- Select
- Filter
- Aggregate

**What?**


[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
- Out Filtered Tree

Why?
- Summarize
- Topology

How?
- Reduce
- Filter

[Munzner (ill. Maguire), 2014]
**D3.js** is a JavaScript library for manipulating documents based on data. **D3** helps you bring data to life using HTML, SVG, and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization.
Data-Driven Documents (D3)

- JavaScript Library
- http://d3js.org/
- Original Authors: Mike Bostock, Vadim Ogievestky, and Jeff Heer
- Open Source
- Focus on Web standards, customization, and usability
- Grew from work on Protovis: more standard, more interactive
- By nature, a **low-level** library; you have control over all elements and styles if you wish
- A top project on GitHub (over 80,000 stars as of 2/8/2017)
- Lots of impressive examples
  - Bostock was a New York Times Graphics Editor
  - http://bost.ocks.org/mike/
D3 Key Features

- Supports data as a core piece of Web elements
  - Loading data
  - Dealing with changing data (joins, enter/update/exit)
    - Correspondence between data and DOM elements
- Selections (similar to CSS) that allow greater manipulation
- Method Chaining
- Integrated layout algorithms, axes calculations, etc.
- Focus on interaction support
  - Straightforward support for transitions
  - Event handling support for user-initiated changes
D3 Introduction

- Ogievetsky has put together a nice set of interactive examples that show off the major features of D3
  - [http://dakoop.github.io/IntroD3/](http://dakoop.github.io/IntroD3/) (Updated from original for D3 v5)
- Other references:
  - Murray’s book on Interactive Data Visualization for the Web
  - The D3 website: [d3js.org](https://d3js.org)
  - Ros's Slides on v4: [https://iros.github.io/d3-v4-whats-new/](https://iros.github.io/d3-v4-whats-new/)