Data Visualization (CIS/DSC 468)

Marks & Channels

Dr. David Koop
D3 Pattern

• Select visual elements \(\text{d3.select, d3.selectAll} \)

• Join them with data items \(\text{.data(myData, key_function)} \)

• Using enter, update, and exit selections, append, update, or remove visual elements
  - \(\text{.enter().append("circle")} \)
  - \(\text{.exit().remove()} \)

• Modify the appearance of elements using attr, style setters
  - \(\text{.attr("width", 300)} \)
  - \(\text{.style("fill", function(d) { return d["Year"] > 2000 ? "red" : "blue"})} \)

• Optionally, use a transition to move from the current appearance to the modified appearance
  - \(\text{.transition().duration(3000)} \)
    - \(\text{.attr("x", function(d,i) { return 10*i; })} \)
D3 Data Joins

- Two groups: data and visual elements
- Three parts of the join between them: enter, update, and exit
- enter: `s.enter()`, update: `s`, exit: `s.exit()`
D3 v4 vs. v3

• v4 breaks a lot of v3 code…
• v4 is more modular, can build libraries that include only the parts you care about
  - Why worry about this?
• Result is that there is a flat namespace now
  - d3.scale.linear => d3.scaleLinear
• More important change: selections are **immutable** now
  - Used to be that enter() modified the selection to include any appended items
  - Use `merge` to explicitly merge the enter and update selections
    - `s.enter().append("rect")
      .merge(s)
    ...

D. Koop, CIS 468, Spring 2017
Assignment 2

- Use D3
  1. Repeat Part 3b of A1 using D3
  2. Extend Part 1 to create a **stacked** bar chart
  3. Create a line chart that shows a region's numbers that is linked to a dropdown menu allowing you to select the region. Use transitions!
Marks & Channels

- How should we generate a visualization of this data?

<table>
<thead>
<tr>
<th>Name</th>
<th>Region</th>
<th>Population</th>
<th>Life Expectancy</th>
<th>Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td>East Asia &amp; Pacific</td>
<td>1335029250</td>
<td>73.28</td>
<td>7226.07</td>
</tr>
<tr>
<td>India</td>
<td>South Asia</td>
<td>1140340245</td>
<td>64.01</td>
<td>2731</td>
</tr>
<tr>
<td>United States</td>
<td>America</td>
<td>306509345</td>
<td>79.43</td>
<td>41256.08</td>
</tr>
<tr>
<td>Indonesia</td>
<td>East Asia &amp; Pacific</td>
<td>228721000</td>
<td>71.17</td>
<td>3818.08</td>
</tr>
<tr>
<td>Brazil</td>
<td>America</td>
<td>193806549</td>
<td>72.68</td>
<td>9569.78</td>
</tr>
<tr>
<td>Pakistan</td>
<td>South Asia</td>
<td>176191165</td>
<td>66.84</td>
<td>2603</td>
</tr>
<tr>
<td>Bangladesh</td>
<td>South Asia</td>
<td>156645463</td>
<td>66.56</td>
<td>1492</td>
</tr>
<tr>
<td>Nigeria</td>
<td>Sub-Saharan Africa</td>
<td>141535316</td>
<td>48.17</td>
<td>2158.98</td>
</tr>
<tr>
<td>Japan</td>
<td>East Asia &amp; Pacific</td>
<td>127383472</td>
<td>82.98</td>
<td>29680.68</td>
</tr>
<tr>
<td>Mexico</td>
<td>America</td>
<td>111209909</td>
<td>76.47</td>
<td>11250.37</td>
</tr>
<tr>
<td>Philippines</td>
<td>East Asia &amp; Pacific</td>
<td>94285619</td>
<td>72.1</td>
<td>3203.97</td>
</tr>
<tr>
<td>Vietnam</td>
<td>East Asia &amp; Pacific</td>
<td>86970762</td>
<td>74.7</td>
<td>2679.34</td>
</tr>
<tr>
<td>Germany</td>
<td>Europe &amp; Central Asia</td>
<td>82338100</td>
<td>80.08</td>
<td>31191.15</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>Sub-Saharan Africa</td>
<td>79996293</td>
<td>55.69</td>
<td>812.16</td>
</tr>
<tr>
<td>Turkey</td>
<td>Europe &amp; Central Asia</td>
<td>72626967</td>
<td>72.06</td>
<td>8040.78</td>
</tr>
</tbody>
</table>
Potential Solution

[Gapminder, Wealth & Health of Nations]
Another Solution

[Gapminder, Wealth & Health of Nations]
Visual Encoding

• How do we encode data visually?
  - **Marks** are the basic graphical elements in a visualization
  - **Channels** are ways to control the appearance of the marks

• Marks classified by dimensionality:

  - **Points**
  - **Lines**
  - **Areas**

• Also can have surfaces, volumes

• Think of marks as a mathematical definition, or if familiar with tools like Adobe Illustrator or Inkscape, the path & point definitions
Visual Channels

- **Position**
  - Horizontal
  - Vertical
  - Both

- **Color**
  - Black
  - Red
  - Green

- **Shape**
  - ▲
  - ★
  - ↘
  - L

- **Tilt**
  - |
  - __|

- **Size**
  - Length
  - Area
  - Volume

[Munzner (ill. Maguire), 2014]
Channels

• Usually map an attribute to a single channel
  - Could use multiple channels but…
  - **Limited** number of channels

• Restrictions on size and shape
  - Points are nothing but location so size and shape are ok
  - Lines have a length, cannot easily encode attribute as length
  - Maps with boundaries have area, changing size can be problematic
Cartograms

[Election Results by Population, M. Newman, 2012]
Channel Types

- Identity => what or where, Magnitude => how much

**Magnitude Channels: Ordered Attributes**
- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

**Identity Channels: Categorical Attributes**
- Spatial region
- Color hue
- Motion
- Shape

[Munzner (ill. Maguire), 2014]
Mark Types

- Can have marks for items and **links**
  - Connection => pairwise relationship
  - Containment => hierarchical relationship

**Marks as Items/Nodes**
- Points
- Lines
- Areas

**Marks as Links**
- Containment
- Connection

[Munzner (ill. Maguire), 2014]
Expressiveness and Effectiveness

• Expressiveness Principle: all data from the dataset and nothing more should be shown
  - Do encode ordered data in an ordered fashion
  - Don’t encode categorical data in a way that implies an ordering

• Effectiveness Principle: the most important attributes should be the most salient
  - Saliency: how noticeable something is
  - How do the channels we have discussed measure up?
  - How was this determined?
How do we test effectiveness?
To analyze responses, we replicated Cleveland & McGill's experiment adhered to the original within-subjects format. Signed each chart as an individual task. Since the vast majority of responses—but ensured that subjects understood the instructions—thus did not filter inaccurate subjects—which would bias the results. Our experiments were initially launched with a limited number of assignments (typically 3) to serve as a pilot. Upon completion of the trial assignments and verification of success, the number of assignments was increased.

As such, it is a natural extension of automated presentation techniques [7, 21], and been successfully extended by others (e.g., [14]). As such, it is a natural extension of automated presentation techniques [7, 21], and been successfully extended by others (e.g., [14]).

RIGHT: T reemap (T9).

Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Top right: Spatially sorted rectangles (T6), Bottom right: Circular area encodings. Cleveland & McGill constructed 7 types of charts, for a total of 70 trials (HITs). We mimicked the number, area (as a bubble chart, see Figure 2), involving transparency (luminance) adjustment of chart grid lines. Our second goal was to conduct additional experiments that demonstrate the use of Mechanical Turk for data exploration, using their log absolute error measure of prediction error and the confidence intervals for each judgment type using bootstrap.

Next we computed the log absolute error means and 95% confidence intervals for each judgment type using bootstrap.

Types 1 and 2 are length encodings (such as Cleveland & McGill's original position-length encoding). Cleveland & McGill corresponded to a visual encoding using position, length or angle (as a pie chart) and type 7 uses circular area encoding. The first question served broadly to assess the use of Mechanical Turk as an experimental platform. Our third goal was to analyze data from across our experiments to characterize more general nature are visited in our performance and cost analysis; for example, we delay discussion of response time and focus on details specific to visualization. Results of a set of perceptual tasks, we replicated Cleveland & McGill to enable comparison, and then investigated optimal encoding still significantly outperformed length encoding. Cleveland & McGill's classic study (Exp. 1A) of proportionality estimation set of perceptual tasks, we replicated Cleveland & McGill's seminal study [7, 21], and been successfully extended by others (e.g., [14]).

TEST % difference in length between elements

[Heer & Bostock, 2010]
Test % difference in **length** between elements

![Chart showing % difference in length between elements A and B.](image)

[Heer & Bostock, 2010]
Test % difference in **length** between elements

![Bar chart showing % difference in length between elements A and B](image)

[Heer & Bostock, 2010]
Test % difference in length between elements

Answer: Left is ~5.6x longer than Right

[Heer & Bostock, 2010]
Test % difference in **length** between elements

[Modified from Heer & Bostock, 2010]
Test % difference in **length** between elements

Answer: Right is 4x larger than Left

[Modified from Heer & Bostock, 2010]
Test % difference in area between elements

[Heer & Bostock, 2010]
Test % difference in area between elements

Answer: A is ~2.25x larger (in area) than B

[Heer & Bostock, 2010]
Test % difference in area between elements

[Heer & Bostock, 2010]
Test % difference in **area** between elements

Answer: A is ~6.1x larger (in area) than B

[Heer & Bostock, 2010]
Test % difference in **area** between elements

[Heer & Bostock, 2010]
Test % difference in area between elements

Answer: B is ~2.5 larger (in area) than A

[Heer & Bostock, 2010]
Cleveland & McGill Experiments

Figure 4. Graphs from position–length experiment.

Figure 3. Graphs from position–angle experiment.

[Cleveland & McGill, 1984]
Heer & Bostock Experiments

- Rerun Cleveland & McGill’s experiment using Mechanical Turk
- … with more tests

Figure 2: Area judgment stimuli. Top left: Bubble chart (T7), Bottom left: Center-aligned rectangles (T8), Right: Treemap (T9).

[Heer & Bostock, 2010]
Psychophysics

- How do we perceive changes in stimuli
- The Psychophysical Power Law [Stevens, 1975]: All sensory channels follow a power function based on stimulus intensity ($S = I^n$)
- Length is fairly accurate
- Magnified vs. compressed sensations

Steven’s Psychophysical Power Law: $S = I^N$

[Stevens, 1975]

[D. Koop, CIS 468, Spring 2017]

[Munzner (ill. Maguire), 2014]
Ranking Channels by Effectiveness

**Channels:** Expressiveness Types and Effectiveness Ranks

- **Magnitude Channels:** Ordered Attributes
  - Position on common scale
  - Position on unaligned scale
  - Length (1D size)
  - Tilt/angle
  - Area (2D size)
  - Depth (3D position)
  - Color luminance
  - Color saturation
  - Curvature
  - Volume (3D size)

- **Identity Channels:** Categorical Attributes
  - Spatial region
  - Color hue
  - Motion
  - Shape

---

[Munzner (ill. Maguire), 2014]
Results Summary

Cleveland & McGill's Results

Crowdsourced Results

Positions

Rectangular areas (aligned or in a treemap)

Angles

Circular areas

Results Summary

[Munzner (ill. Maguire) based on Heer & Bostock, 2014]