Data Visualization (CIS 468)

Tables

Dr. David Koop
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

  ![Points](image1)  ![Lines](image2)  ![Areas](image3)

• 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
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]
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?
Cleveland & McGill Accuracy Experiments

Figure 4. Graphs from position–length experiment.

Figure 3. Graphs from position–angle experiment.

[Cleveland & McGill, 1984]
Results Summary

Cleveland & McGill’s Results

Crowdsourced Results

Positions

Rectangular areas
(aligned or in a treemap)

Circular areas

Angles

T1
T2
T3
T4
T5
T6
T7
T8
T9

Log Error

1.0 1.5 2.0 2.5 3.0

1.0 1.5 2.0 2.5 3.0

[Muñzner (ill. Maguire) based on Heer & Bostock, 2014]
Ranking Channels by Effectiveness

**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

[Image credit: Munzner (ill. Maguire), 2014]
Assignment 2

• Link
• Use Tableau and D3 to create stacked bar charts of Citibike data
• Due next Friday, Start now!
Tableau Example
D3 Examples

• Bar Chart:
  - Start: http://codepen.io/dakoop/pen/dNxjYL
  - Simple Solution: http://codepen.io/dakoop/pen/aJoLBp

• With Axes and Scales: http://codepen.io/dakoop/pen/WpeZOV

• With Objects and Margin Convention: http://codepen.io/dakoop/pen/MJNGwZ

• More on Margin Convention:
  - https://bl.ocks.org/mbostock/3019563 (Note this is D3 v3!)
Be Careful: More Channel Considerations
Discriminability

What is problematic here?

[Koop et al., 2013]
Discriminability

• Can a human tell the difference?
• How many values (bins) can be used so that a person can tell the difference?
• Example: Line width
  - Matching a particular width with a legend
  - Comparing two widths
Separability

- Cannot treat all channels as independent!
- **Separable** means each individual channel can be distinguished
- **Integral** means the channels are perceived together

[Munzner (ill. Maguire) based on Ware, 2014]
Separable or Integral?
Separable or Integral?

The map at right is a product of overlaying the three sets of data. The variation in hue and value has been produced from the data shown above. In general, darker counties represent a more educated, better paid population while lighter areas represent communities with fewer graduates and lower incomes.
Visual Popout
Visual Popout: Parallel Lines Require Search...

[Munzner (ill. Maguire), 2014]
Relative vs. Absolute Judgments

- Weber’s Law:
  - We judge based on relative not absolute differences
  - The amount of perceived difference depends is relative to the object’s magnitude!

[Munzner (ill. Maguire), 2014]
Luminance Perception

Edward H. Adelson

[E. H. Adelson, 1995]
Luminance Perception

[Edward H. Adelson, 1995]
<table>
<thead>
<tr>
<th>REMOTE</th>
<th>STATION</th>
<th>FF</th>
<th>SEN/DIS</th>
<th>7-D AFAS UNL</th>
<th>D AFAS/RMF</th>
<th>JOINT RR TKT</th>
<th>7-D UNL</th>
<th>30-D UNL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>R011 42ND STREET &amp; 8TH AVENUE</td>
<td>00228985</td>
<td>00008471</td>
<td>00000441</td>
<td>00001455</td>
<td>00000134</td>
<td>00033341</td>
<td>00071255</td>
</tr>
<tr>
<td>2</td>
<td>R170 14TH STREET-UNION SQUARE</td>
<td>00224603</td>
<td>00011051</td>
<td>00000827</td>
<td>00003026</td>
<td>00000660</td>
<td>00089367</td>
<td>00199841</td>
</tr>
<tr>
<td>3</td>
<td>R046 42ND STREET &amp; GRAND CENTRAL</td>
<td>00207758</td>
<td>00007908</td>
<td>00000323</td>
<td>00001183</td>
<td>00003001</td>
<td>00040759</td>
<td>00096613</td>
</tr>
<tr>
<td>4</td>
<td>R012 34TH STREET &amp; 8TH AVENUE</td>
<td>00188311</td>
<td>00006490</td>
<td>00000498</td>
<td>00001279</td>
<td>00003622</td>
<td>00035527</td>
<td>00067483</td>
</tr>
<tr>
<td>5</td>
<td>R293 34TH STREET – PENN STATION</td>
<td>00168768</td>
<td>00006155</td>
<td>00000523</td>
<td>00001065</td>
<td>00005031</td>
<td>00030645</td>
<td>00054376</td>
</tr>
<tr>
<td>6</td>
<td>R033 42ND STREET/TIMES SQUARE</td>
<td>00159382</td>
<td>00005945</td>
<td>00000378</td>
<td>00001205</td>
<td>00000690</td>
<td>00058931</td>
<td>00078644</td>
</tr>
<tr>
<td>7</td>
<td>R022 34TH STREET &amp; 6TH AVENUE</td>
<td>00156008</td>
<td>00006276</td>
<td>00000487</td>
<td>00001543</td>
<td>00000712</td>
<td>00058910</td>
<td>00110466</td>
</tr>
<tr>
<td>8</td>
<td>R084 59TH STREET/COLUMBUS CIRCLE</td>
<td>00155262</td>
<td>00009484</td>
<td>00000589</td>
<td>00002071</td>
<td>00000542</td>
<td>00053397</td>
<td>00113966</td>
</tr>
<tr>
<td>9</td>
<td>R020 47-50 STREETS/ROCKEFELLER</td>
<td>00143500</td>
<td>00006402</td>
<td>00000384</td>
<td>00001159</td>
<td>00000723</td>
<td>00037978</td>
<td>00090745</td>
</tr>
<tr>
<td>10</td>
<td>R179 86TH STREET-LEXINGTON AVE</td>
<td>00142169</td>
<td>00010367</td>
<td>00000470</td>
<td>00001839</td>
<td>00000271</td>
<td>00050328</td>
<td>00125250</td>
</tr>
<tr>
<td>11</td>
<td>R023 34TH STREET &amp; 6TH AVENUE</td>
<td>00134052</td>
<td>00005005</td>
<td>00000348</td>
<td>00001112</td>
<td>00000649</td>
<td>00031531</td>
<td>00075040</td>
</tr>
<tr>
<td>12</td>
<td>R029 PARK PLACE</td>
<td>00121614</td>
<td>00004311</td>
<td>00000287</td>
<td>00000931</td>
<td>00000792</td>
<td>00025404</td>
<td>00065362</td>
</tr>
<tr>
<td>13</td>
<td>R047 42ND STREET &amp; GRAND CENTRAL</td>
<td>00100742</td>
<td>00004273</td>
<td>00000185</td>
<td>00000704</td>
<td>00001241</td>
<td>00022808</td>
<td>00068216</td>
</tr>
<tr>
<td>14</td>
<td>R031 34TH STREET &amp; 7TH AVENUE</td>
<td>00095076</td>
<td>00003990</td>
<td>00000232</td>
<td>00000727</td>
<td>00001459</td>
<td>00024284</td>
<td>00038671</td>
</tr>
<tr>
<td>15</td>
<td>R017 LEXINGTON AVENUE</td>
<td>00094655</td>
<td>00004688</td>
<td>00000190</td>
<td>00000833</td>
<td>00000754</td>
<td>00020018</td>
<td>00055066</td>
</tr>
<tr>
<td>16</td>
<td>R175 8TH AVENUE-14TH STREET</td>
<td>00094313</td>
<td>00003907</td>
<td>00000286</td>
<td>00001144</td>
<td>00000256</td>
<td>00038272</td>
<td>00074661</td>
</tr>
<tr>
<td>17</td>
<td>R057 BARCLAYS CENTER</td>
<td>00093804</td>
<td>00004204</td>
<td>00000454</td>
<td>00001386</td>
<td>00001491</td>
<td>00039113</td>
<td>00068119</td>
</tr>
<tr>
<td>18</td>
<td>R138 WEST 4TH ST-WASHINGTON SQ</td>
<td>00093562</td>
<td>00004677</td>
<td>00000251</td>
<td>00000965</td>
<td>00000127</td>
<td>00031628</td>
<td>00074458</td>
</tr>
</tbody>
</table>
Visualization of Tables

- Items and attributes
- For now, attributes are not known to be positions
- Keys and values
  - **key** is an independent attribute that is unique and identifies item
  - **value** tells some aspect of an item
- Keys: categorical/ordinal
- Values: +quantitative
- Levels: unique values of categorical or ordered attributes

[Munzner (ill. Maguire), 2014]
Arrange Tables

- Express Values
  - Separate, Order, Align Regions
    - Separate
    - Order
    - Align
    - 1 Key
    - 2 Keys
    - 3 Keys
    - Many Keys
    - List
    - Recursive Subdivision
    - Volume
    - Matrix
  - Rectilinear
  - Parallel
  - Radial
- Layout Density
  - Dense
  - Space-Filling

[Munzner (ill. Maguire), 2014]
Express Values: Scatterplots

- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions

- Correlation: dependence between two attributes
  - Positive and negative correlation
  - Indicated by lines
- Coordinate system (axes) and labels are important!
Coordinate Systems

[Figure 2: (a) Plot of n vs deviation. Variability of deviation is dominated by sample size: small sample sizes have large variability, and the variability decreases with sample size.]

```
R> hod_unusual <- match_df(hod2, unusual)
R> unusual <- subset(devi, resid > 1.5)
R> geom_smooth(method = "rlm", se = F)
R> scale_x_log10() +
R> ggplot(data = devi, aes(x = n, y = dist) + geom_point()
```

[Source: Wickham, 2014]
Log-Log Plot

---

R> hod_unusual <- match_df(hod2, unusual)
R> unusual <- subset(devi, resid > 1.5)

We are interested in points that have high unusually high values. The blue line is a robust line of best fit. Small samples have large variability. (b) Log-log plot makes it easy to see the pattern of variation as well as variability decreases with sample size. But on the log-log scale, Figure 2 (a) Plot of n vs deviation. Variability of deviation is dominated by sample size: small

---

[Wickham, 2014]
Bubble Plot

[Gapminder, Wealth & Health of Nations]
Scatterplot

- Data: two quantitative values
- Task: find trends, clusters, outliers
- How: marks at spatial position in horizontal and vertical directions
- **Scalability**: hundreds of items

- "Ranking Visualizations of Correlation Using Weber’s Law", 2014:
  - Correlation perception can be modeled via Weber’s Law
  - Scatterplots are one of the best visualizations for both positive and negative correlation
Separate, Order, and Align: Categorical Regions

- Categorical: =, !=
- Spatial position can be used for categorical attributes
- Use regions, distinct contiguous bounded areas, to encode categorical attributes
- Three operations on the regions:
  - Separate (use categorical attribute)
  - Align (use some other ordered attribute)
  - Order
- Alignment and order can use same or different attribute
List Alignment: Bar Charts

- Data: one quantitative attribute, one categorical attribute
- Task: lookup & compare values
- How: line marks, vertical position (quantitative), horizontal position (categorical)
- What about length?
- Ordering criteria: alphabetical or using quantitative attribute
- Scalability: distinguishability
  - bars at least one pixel wide
  - hundreds

[Muñzner (ill. Maguire), 2014]
Stacked Bar Charts

[Stacked Bar Chart, M. Bostock, 2017]
Stacked Bar Charts

- Data: multidimensional table: one quantitative, two categorical
- Task: lookup values, part-to-whole relationship, trends
- How: line marks: position (both horizontal & vertical), subcomponent line marks: length, color
- Scalability: main axis (hundreds like bar chart), bar classes (<12)

- Orientation: vertical or horizontal (swap how horizontal and vertical position are used.)
Streamgraphs

- Include a time attribute
- Data: multidimensional table, one quantitative attribute (count), one ordered key attribute (time), one categorical key attribute
- + derived attribute: layer ordering (quantitative)
- Task: analyze trends in time, find (maxmial) outliers
- How: derived position+geometry, length, color
- Scalability: more categories than stacked bar charts

[Byron and Wattenberg, 2012]
Streamgraphs

Each shape shows how one film did at the box office.

Height shows weekly box office revenue

Width shows longevity

The area of the shape (and its color) corresponds to the film's total domestic gross, through Feb. 21

1 25

3662 million

250

100

25

Dot and Line Charts

- Data: one quantitative attribute, one ordered attribute
- Task: lookup values, find outliers and trends
- How: point mark and positions
- Line Charts: add connection mark (line)
- Similar to scatterplots but allow ordered attribute

[Munzner (ill. Maguire), 2014]
Proper Use of Line and Bar Charts

[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]
Proper Use of Line and Bar Charts

- What does the line indicate?
- Does this make sense?

[Adapted from Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]
Aspect Ratio

- Trends in line charts are more apparent because we are using angle as a channel.
- Perception of angle (and the **relative difference** between angles) is important.
- Initial experiments found people best judge differences in **slope** when angles are around 45 degrees (Cleveland et al., 1988, 1993).
Multiscale Banking

Sunspot Cycles
Aspect Ratio = 3.96

Aspect Ratio = 22.35

Power Spectrum

Figure 5. Sunspot observations, 1700-1987. The first plot shows low-frequency oscillations in the maximum values of sunspot cycles. The second plot brings the individual cycles into greater relief.

Carbon Dioxide Measurements
Aspect Ratio = 1.17
Aspect Ratio = 7.87

Power Spectrum

Figure 6. Monthly atmospheric CO$_2$ measurements. The first plot shows a baseline trend of increasing values, with a slight inflection. The second plot more clearly communicates the yearly oscillations.

Figures 5-8 show the results of applying multi-scale banking to real-world data sets. Data sets are plotted at each computed aspect ratio, with banked trend lines shown in red. The power spectrum plot shows a frequency-domain representation of the data, annotated with potential scales of interest. The aspect ratio plot shows the banked aspect ratios for each possible lowpass filtering of the data, annotated with the final aspect ratios returned by the algorithm.

PRMTX Mutual Fund
Aspect Ratio = 4.23
Aspect Ratio = 14.55

Power Spectrum

Figure 7. PRMTX mutual fund performance, 1997-2006. The first plot shows the boom and bust of the “dot-com” bubble and subsequent recovery. The second plot affords closer consideration of short-term variations.

Downloads of the prefuse toolkit
Aspect Ratio = 1.44
Aspect Ratio = 2.89
Aspect Ratio = 8.81

Power Spectrum

Figure 8. Daily download counts of the prefuse visualization toolkit. The first plot shows a general increase in downloads. The second plot shows weekly variations, including reduced downloads on the weekends. The third plot enables closer inspection of day-to-day spikes and decays.

[Heer and Agrawala, 2006]
Multiscale Banking

PRMTX Mutual Fund

Aspect Ratio = 4.23

Aspect Ratio = 14.55

[Heer and Agrawala, 2006]
Expanding the Study

- Cleveland et al. did not study the entire space of slope comparisons and 45 degrees was at the low end of their study (blue marks on right)
- Talbot et al. compared more slopes and found that people do better with smaller slopes
- Baselines may aid with this