Data Visualization (CIS/DSC 468)

Tabular Data

Dr. David Koop
Channel Considerations

- Discriminability
- Separability
- Visual Popout
- Weber's Law
- Luminance Perception
Separability

- Cannot treat all channels as independent!

- **Separable** means each individual channel can be distinguished

- **Integral** means the channels are perceived together

---

[Position + Hue (Color)]

[Size + Hue (Color)]

[Width + Height]

[Red + Green]

- Fully separable
- Some interference
- Some/significant interference
- Major interference

[Munzner (ill. Maguire) based on Ware, 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]
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!
D3 Examples

• 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!)
Exam 1

• Wednesday, March 1 in class (12-12:50pm)

• Format:
  - Multiple Choice
  - Short Answer

• Sample questions on web site:
<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/RFM</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

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

- Express Values

- Separate, Order, Align Regions
  - Separate
  - Order
  - Align
  - 1 Key
  - 2 Keys
  - 3 Keys
  - Many Keys

- Axis Orientation
  - 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

[ Wickham, 2014 ]
Log-Log Plot

We are interested in points that have high unusually high values. The blue line is a robust line of best fit.

Figure 2: (a) Plot of n vs deviation. Variability of deviation is dominated by sample size: small variability decreases with sample size. But on the log-log scale, Figure [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

- Cool recent result from Harrison et al., "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

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

[Bostock, 2012]
Grouped Bar Chart

**Population**

- California (CA)
- Texas (TX)
- New York (NY)
- Florida (FL)
- Illinois (IL)
- Pennsylvania (PA)

- 65 Years and Over
- 45 to 64 Years
- 25 to 44 Years
- 18 to 24 Years
- 14 to 17 Years
- 5 to 13 Years
- Under 5 Years

[M. Bostock, http://bl.ocks.org/mbostock/3887051]
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]
NYTimes “Ebb and Flow of Movies"


Sources: Baseline StudioSystems; Box Office Mojo

Mathew Bloch, Lee Byron, Shan Carter and Amanda Cox
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

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

[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)
Sunspot Cycles

Aspect Ratio = 3.96

[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.

[Talbot et al., 2013]
Heatmaps

- Data: Two keys, one quantitative attribute
- Task: Find clusters, outliers, summarize
- How: area marks in grid, color encoding of quantitative attribute
- Scalability: number of pixels for area marks (millions), <12 colors
- Red-green color scales often used
  - Be aware of colorblindness!

Fast-Pitch Softball Slugging Percentage

[fastpitchanalytics.com]
Bertin Matrices

• Must we only use color?
  - What other marks might be appropriate?

[C.Perrin et al., 2014]
Bertin Matrices

- Must we only use color?
  - What other marks might be appropriate?

[C.Perrin et al., 2014]
Bertin’s Encodings

[Image: Bertin’s Encodings diagram showing encodings for different types of data visualization]

[C.Perrin et al., 2014]
Matrix Reordering

![Matrix Reordering Image]

[Bertin Exhibit (INRIA, Vis 2014), Photo by Robert Kosara]
Cluster Heatmap

[File System Similarity, R. Musăloiu-E., 2009]
Cluster Heatmap

- Data & Task: Same as Heatmap
- How: Area marks but matrix is ordered by cluster hierarchies
- Scalability: limited by the cluster dendrogram

- Dendrogram: a visual encoding of tree data with leaves aligned
Scatterplot Matrix (SPLOM)

- Data: Many quantitative attributes
- Derived Data: names of attributes
- Task: Find correlations, trends, outliers
- How: Scatterplots in matrix alignment
- Scale: attributes: ~12, items: hundreds?
- Visualizations in a visualization: at high level, marks are themselves visualizations...

[R command from Wikipedia]
Spatial Axis Orientation

• So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

• What other possibilities are there for axes?

[Munzner (ill. Maguire), 2014]
Spatial Axis Orientation

• So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

• What other possibilities are there for axes?
  - Parallel axes

[Parallel Coordinates]

[Munzner (ill. Maguire), 2014]

D. Koop, CIS 468, Spring 2017
Spatial Axis Orientation

• So far, we have seen the vertical and horizontal axes (a rectilinear layout) used to encode almost everything

• What other possibilities are there for axes?
  - Parallel axes
  - Radial axes

[Parallel Co ordinates: Math, Physics, Dance, Drama]

[Munzner (ill. Maguire), 2014]
Pie Chart

[Pie Chart, Bostock, 2017]
Pie Charts

• vs. bar charts [Munzner's Textbook, 2014]
  - Angle channel is lower precision then position in bar charts
• What about donut charts?
• Are we judging angle, or are we judging area, … or arc length?
  - See "An Illustrated Study of the Pie Chart Study Results" by R. Kosara
Arcs, Angles, or Areas?

[R. Kosara and D. Skau, 2016]
Absolute Error Relative to Pie Chart

[R. Kosara and D. Skau, 2016]
Conclusion: We do not read pie charts by angle

[R. Kosara and D. Skau, 2016]
Pies vs. Bars

- …but area is still harder to judge than position
- Screens are usually not round