Data Visualization (DSC 530/CIS 602-01)

Tables

Dr. David Koop
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

[Visualization of Tables diagram]

[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!
To find these unusual points, we fit a robust linear model and plot the residuals, Figure 3. Overall, there is a clear pattern for the number of deaths, these points represent the diseases which depart the most from the overall pattern. We are interested in points that have high variability decreases with sample size. But on the log-log scale, Figure 2, there is a clear trend.
Bubble Plot

D. Koop, DSC 530, Spring 2018

[Gapminder, Wealth & Health of Nations]
Project Proposal

• Information

• Due Next Monday, March 5

• Choices:
  - [CreateVis] Create a complex, interactive visualization of a dataset
  - [Research] Research project that involves visualization (new technique, evaluation, etc.)

• Check with me if you're interested in the research project option

• Start:
  - [CreateVis] Looking for a dataset: Awesome Public Datasets, Kaggle, etc.
  - [Research] Surveying literature, identifying goals
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]
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

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

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
Aspect Ratios

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
Aspect Ratios

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
Aspect Ratios

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

[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.

[Fig. 2. Space of line comparisons parameterized by mid-angle $\theta_m$.]
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!

<table>
<thead>
<tr>
<th>Stripes</th>
<th>0 balls</th>
<th>1 balls</th>
<th>2 balls</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.503</td>
<td>0.492</td>
<td>0.431</td>
</tr>
<tr>
<td>1</td>
<td>0.538</td>
<td>0.518</td>
<td>0.450</td>
</tr>
<tr>
<td>2</td>
<td>0.560</td>
<td>0.552</td>
<td>0.508</td>
</tr>
<tr>
<td>3</td>
<td>0.543</td>
<td>0.690</td>
<td>0.512</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>QUANTITY OF INK ENCODINGS</th>
<th>POSITIONAL ENCODING</th>
<th>MEAN-BASED ENCODINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grayscale</td>
<td>Circle</td>
<td>Dual bar chart</td>
</tr>
<tr>
<td></td>
<td>Bar chart</td>
<td>Line</td>
</tr>
<tr>
<td></td>
<td>Black and white bar chart</td>
<td>Average bar chart</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grayscale</td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circle</td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dual bar chart</td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bar chart</td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Line</td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black and white bar chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average bar chart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>•</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

[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...
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

Math | Physics | Dance | Drama
---|---|---|---
100 | 90 | 80 | 70
90 | 80 | 70 | 60
80 | 70 | 60 | 50
70 | 60 | 50 | 40
60 | 50 | 40 | 30
50 | 40 | 30 | 20
40 | 30 | 20 | 10
30 | 20 | 10 | 0

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

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

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

[Munzner (ill. Maguire), 2014]
Radial Axes
Radial Axes

- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
Radial Axes

- Polar Coordinates (angle + position along the line at that angle)
- What types of encodings are possible for tabular data in polar coordinates?
  - Radial bar charts
  - Pie charts
  - Donut charts
Part-of-whole: Relative % comparison?

[Stacked Bar Chart, Bostock, 2017]
Normalized Stacked Bar Chart

[Normalized Stacked Bar Chart, Bostock, 2017]
Pie Chart

[Pie Chart, Bostock, 2017]
Pie Charts

• vs. bar charts [Munzner's Textbook, 2014]
  - Angle channel is lower precision than position in bar charts
• What about donut charts?
• Are we judging angle, or are we judging area, … or arc length?
  - "Arcs, Angles, or Areas: Individual Data Encodings in Pie and Donut Charts", D. Skau and R. Kosara, 2016
  - Summary: "An Illustrated Study of the Pie Chart Study Results"
Study Setup

- Three studies
- 80-100 participants each
- Each answered ~60 questions
- Computed results using 95% Confidence Intervals

[95% CI]

[Mean]

[R. Kosara and D. Skau, 2016]
Arcs, Angles, or Areas?

[R. Kosara and D. Skau, 2016]
Signed Error

[R. Kosara and D. Skau, 2016]
Absolute Error

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

[R. Kosara and D. Skau, 2016]
Donut Charts Width

[R. Kosara and D. Skau, 2016]

---

D. Koop, DSC 530, Spring 2018
Pie Chart Variations

[R. Kosara and D. Skau, 2016]
Pie Chart Variations

[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