Data Visualization (CSCI 627/490)

Tabular Data

Dr. David Koop
Visual Encoding

- **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
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?
## 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

[Source: Munzner (ill. Maguire), 2014]

D. Koop, CSCI 627/490, Fall 2023
How was this determined?
Perception Studies Summary

[Cleveland & McGill's Results]

[Crowdsourced Results]

[Munzner (ill. Maguire) based on Heer & Bostock, 2014]
Assignment 3

- Chicago Traffic Crashes
- Create the same stacked bar chart using
  - Tableau Public
  - Observable Plot
  - D3
- D3 Stacked Bar Chart:
  - Required for CSCI 627 students
  - CSCI 490 students can just do counts
Emergency Alert Test at 1:20pm
Project

- Start thinking about project dataset and questions
- Working on posting some example datasets
- Goal: Less explored datasets (more opportunity for design/questions)
- If you are doing research and can tie this project in, please talk with me
## Tables

<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</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</td>
<td>00224603</td>
<td>0011051</td>
<td>00000827</td>
<td>00003026</td>
<td>00000660</td>
<td>00089367</td>
<td>00199841</td>
</tr>
<tr>
<td>3</td>
<td>R046</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</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</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</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</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</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</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</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</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</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</td>
<td>00100742</td>
<td>00004273</td>
<td>00000185</td>
<td>00000704</td>
<td>00001241</td>
<td>00022808</td>
<td>00068216</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: categorical/ordinal/quantitative
• Levels: unique *values* of categorical or ordered attributes

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

- Express Values
- Separate, Order, Align Regions
- Separate
- Order
- Align
- Layout Density
  - Dense
  - Space-Filling
- Axis Orientation
  - Rectilinear
  - Parallel
  - Radial

1 Key
- List
2 Keys
- Matrix
3 Keys
- Volume
Many Keys
- Recursive Subdivision

[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 2.

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

geom_smooth(method = "rlm", se = F)
last_plot() +
ggplot(data = devi, aes(x = n, y = dist) + geom_point()
```

---

[Wickham, 2014]
To find these unusual points, we fit a robust linear model and plot the residuals, Figure 1. In (a) linear scales, samples have large variability. (b) Log-log plot makes it easy to see the pattern of variation as well as the line of best fit from a linear model.

```r
R> hod_unusual <- match_df(hod2, unusual)
R> unusual <- subset(devi, resid > 1.5)
R> geom_smooth(method = "rlm", se = F)
R>
```

---

**[Wickham, 2014]**
Bubble Plot

[Gapminder, Wealth & Health of Nations]

D. Koop, CSCI 627/490, Fall 2023
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

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

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

Grouped 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

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

Height shows weekly box office revenue.

Width shows longevity.

Transformers
The Simpsons Movie
Ocean's Thirteen
I Now Pronounce You Chuck and Larry
The Bourne Ultimatum
Live Free or Die Hard
Hairspray (2007)
Fantastic Four: Rise of the Silver Surfer
Superbad
Harry Potter and the Order of the Phoenix
Ratatouille
Evan Almighty
Rush Hour 3
National Treasure: Book of Secrets
Enchanted
I Am Legend
American Gangster
Juno
Alvin and the Chipmunks
Bea Movie

Streamgraphs

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

Aspect Ratio = 3.96

Aspect Ratio = 22.35
Multiscale Banking

Aspect Ratio = 4.23

![Graph showing a financial trend over time with an aspect ratio of 4.23.]

Aspect Ratio = 14.55

![Graph showing a financial trend with an aspect ratio of 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

Figure 2 parameterized by the mid-angle between the segments. Our approach to answering this question is twofold: (1) Based on insights gained from a series of pilot studies (Sections 6 and 7), we develop a model of human perception of slope ratios. The model fits observed data well and we provide a perceptual interpretation of its components (Sections 6 and 7).

For example, in early iterations we used a study design with 100 instruction users would naturally approach the slope estimation task. To explore a variety of model possibilities, we settled on the design shown in Figure 2 with 49 (7 × 7) slope ratios, but found that this resolution was unnecessary since the shape of the response function was relatively smooth. In later iterations, we relaxed our mid-angle sampling part way through our analysis, our mid-angle sampling was at the low end of their study (blue marks on right).

In our studies, we sampled with 100 points. We use a lower sampling rate makes better use of limited subject time. As far as we can tell, there is no other work in perception on slope ratio estimation. However, there is substantial work on related perception of line orientation. Other work on 2D angle perception have shown that the median segment slope was 1, but this is much smaller range. To keep the sample count manageable, we sample both lines is the same. However, we wonder if this is how most visualization users would naturally approach the slope estimation task.
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

![Slugging Percentage Heatmap](fastpitchanalytics.com)
Bertin Matrices

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

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

![Bertin Matrix Image]

[C.Perrin et al., 2014]
4.3 Human-Assisted Reordering

4.3.1 Manual Reordering Interactions

As in previous implementations, which is useful in many cases, including for automatically reordering visual groups (i.e., separators), we avoided the ambiguous term “group.” The principle of glueing several rows or columns together (Figure 49) is fairly simple and relies on the command is selected on mouse press, after which the user is allowed on post-processed values, these controls provide a way of fine-tuning

4.3.3 Visual Encodings

By supporting crossing-based interactions, the tool only shows a partial preview. The same is true for gestures have never been applied to manipulate multiple sliders at once, it has been conditioned and normalized (e.g., between

Automatic reordering is however disabled on text to reinforce the visual steering, a slow motor task. The remaining encodings are more common and all roughly follow Bertin’s Encodings

<table>
<thead>
<tr>
<th>QUANTITY OF INK ENCODINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grayscale</td>
</tr>
<tr>
<td>Circle</td>
</tr>
<tr>
<td>Dual bar chart</td>
</tr>
<tr>
<td>Bar chart</td>
</tr>
<tr>
<td>Line</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>POSITIONAL ENCODING</th>
</tr>
</thead>
<tbody>
<tr>
<td>Black and white bar chart</td>
</tr>
<tr>
<td>Average bar chart</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MEAN-BASED ENCODINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<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>
</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>
</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>
</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>
</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>
</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>
</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>
</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>
</tr>
</tbody>
</table>

Matrix Reordering
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?
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

[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
  - Radial axes

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

Diagram of a polar coordinate system with radial axes marked at various degrees: 0°, 30°, 60°, 90°, 120°, 150°, 180°, 210°, 240°, 270°, 300°, 330°.
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, M. Bostock, 2017]
Pie Chart

[Pie Chart, Bostock, 2017]