Data Visualization (DSC 530/CIS 602-02)

Networks

Dr. David Koop
Arrange Tables

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

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!
Line and Bar Charts, and Proper Use

[Image: Bar charts and line graphs showing height differences between genders and age groups.]

[Zacks and Tversky, 1999, Munzner (ill. Maguire), 2014]
Sunspot Cycles

Aspect Ratio = 3.96

Aspect Ratio = 22.35

[Heer and Agrawala, 2006]
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!
- **Also Bertin matrices**

---

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>.503</td>
<td>.492</td>
<td>.431</td>
</tr>
<tr>
<td>1</td>
<td>.538</td>
<td>.518</td>
<td>.450</td>
</tr>
<tr>
<td>2</td>
<td>.560</td>
<td>.552</td>
<td>.508</td>
</tr>
<tr>
<td>3</td>
<td>.543</td>
<td>.690</td>
<td>.512</td>
</tr>
</tbody>
</table>

Fast-Pitch Softball
Slugging Percentage

[fastpitchanalytics.com]
Part-of-whole: Relative % comparison?
Normalized Stacked Bar Chart

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

[Pie Chart, Bostock, 2017]
Arcs, Angles, or Areas?

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

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

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

[R. Kosara and D. Skau, 2016]
Assignment 2

- Due next Monday
- 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 (3:30-4:45pm)
- Format:
  - Multiple Choice
  - Short Answer
- Sample questions on web site:
- Questions may involve papers discussed in class
- Questions may involve interpreting code or writing pseudocode
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

[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]
Parallel Coordinates

• Data: many quantitative attributes
• Task: Find trends, extremes, correlation
• How: vertical spatial position for each attribute, connection marks for identity, axes horizontally spaced
• Scalability: <40 attributes, hundreds of values

• Connection marks help visualize trends between particular values
• **Ordering** the horizontal axes is important
• Not as well-known, often requires learning
Comparing SPLOMs and Parallel Coordinates

Scatterplot Matrix

Parallel Coordinates

[Math, Physics, Dance, Drama]

[Math, Physics, Dance, Drama]

[Math, Physics, Dance, Drama]

[Math, Physics, Dance, Drama]

[Munzner (ill. Maguire), 2014]
Correlation in Parallel Coordinates

[Wegman, 1990]
Overdraw in Parallel Coordinates

[Fua et al., 1999]
Hierarchical Parallel Coordinates

[Figure 4: This image sequence shows a Fatal Accident data set of 230,000 data elements at different levels of detail. The first image shows a cut across the root node. The last image shows the cut chaining all the leaf nodes. The rest of the images show intermediate cuts at varying levels of detail. (See Color Plates).]

[Figure 6: Left image shows Iris data set without proximity-based coloring. Right image shows Iris data set with proximity-based coloring revealing three distinct clusters depicted by concentrations of blue, green and pink lines. (See Color Plates).]

[Fua et al., 1999]
Networks

• Why not graphs?
  - Bar graph
  - Graphing functions in mathematics

• Network: nodes and edges connecting the nodes

• Formally, \( G = (V,E) \) is a set of nodes \( V \) and a set of edges \( E \) where each edge connects two nodes.

• Nodes == items, edges connect items

• **Both** nodes and edges may have **attributes**
Arrange Networks and Trees

Node–Link Diagrams
Connection Marks

Adjacency Matrix
Derived Table

Enclosure
Containment Marks

[Munzner (ill. Maguire), 2014]
Molecule Graph
Molecule Graph

- Nodes may have attributes (e.g. element)
Molecule Graph

- Nodes may have attributes (e.g. element)
- Edges may have attributes (e.g. number of bonds)
Web Sites as Graphs (amazon.com)

[M. Salathe, 2006]
Social Networks

[P. Butler, 2010]
Graphs as Data

Nodes

<table>
<thead>
<tr>
<th>ID</th>
<th>Atom</th>
<th>Electrons</th>
<th>Protons</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>N</td>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

Edges

<table>
<thead>
<tr>
<th>ID1</th>
<th>ID2</th>
<th>Bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Node-Link Diagrams

• Data: nodes and edges
• Task: understand connectivity, paths, structure (topology)
• Encoding: nodes as point marks, connections as line marks
• Scalability: hundreds

• ...but high **density** of links can be problematic!

• Problem with the above encoding?
Arc Diagram

[D. Eppstein, 2013]
Network Layout

• Need to use spatial position when designing network visualizations
• Otherwise, nodes can **occlude** each other, links hard to distinguish
• How?
  - With bar charts, we could order using an attribute…
  - With networks, we want to be able to see connectivity and topology (not in the data usually)
• Possible metrics:
  - Edge crossings
  - Node overlaps
  - Total area
Force-Directed Layout

• Nodes push away from each other but edges are springs that pull them together

• Weakness: nondeterminism, algorithm may produce difference results each time it runs

[M. Bostock, 2012]
sfdp

[Hu, 2005]
“Hairball”

[Hu, 2014]
Hierarchical Edge Bundling

Fig. 13. A software system and its associated call graph (caller = green, callee = red). (a) and (b) show the system with bundling strength $\beta = 0.85$ using a balloon layout (node labels disabled) and a radial layout, respectively. Bundling reduces visual clutter, making it easier to perceive the actual connections than when compared to the non-bundled versions (figures 2a and 11a). Bundled visualizations also show relations between sparsely connected systems more clearly (encircled regions); these are almost completely obscured in the non-bundled versions. The encircled regions highlight identical parts of the system for (a), (b), and figure 15.

Fig. 14. Using the bundling strength $\beta$ to provide a trade-off between low-level and high-level views of the adjacency relations. The value of $\beta$ increases from left-to-right; low values mainly provide low-level, node-to-node connectivity information, whereas high values provide high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of explicit adjacency edges between their respective child nodes.

More specifically, most of the participants particularly valued the fact that relations between items at low levels of the hierarchy were automatically lifted to implicit relations between items at higher levels by means of bundles. This quickly gave them an impression of the high-level connectivity information while still being able to inspect the low-level relations that were responsible for the bundles by interactively manipulating the bundling strength. This is illustrated in figure 14, which shows visualizations using different values for the bundling strength $\beta$. Low values result in visualizations that mainly provide low-level, node-to-node connectivity information. High values result in visualizations that provide high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of explicit adjacency edges between their respective child nodes.

Another aspect that was commented on was how the bundles gave an impression of the hierarchical organization of the data as well, thereby strengthening the visualization of the hierarchy. More specifically, a thick bundle shows the presence of two elements at a fairly high level of the hierarchy, whereas the fanning out of a bundle shows the subdivision of an element into subelements.

Most participants preferred the radial layout over the balloon layout and the squarified treemap layout. Another finding was the fact that the rooted layout and the slice-and-dice treemap layout were considered less pleasing according to several participants. This is probably due to the large number of collinear nodes within these layouts, which causes bundles to overlap along the collinearity axes. This is illustrated in figure 17.

Although our main focus while developing hierarchical edge bundling was on the visualization itself, interaction is an important aspect in determining the usability of our technique. Based on our own insight and feedback gathered from participants, we contend that bundle-based interaction as described below could provide a convenient way of interacting with the visualizations.

Figure 16 shows how the bundling strength $\beta$ could be used in conjunction with different interaction techniques. For example, increasing $\beta$ could be used to selectively hide or emphasize certain parts of the visualization, while decreasing $\beta$ could be used to explore the underlying structure and relationships between items at different levels of the hierarchy.

[Holten, 2006]
Hierarchical Edge Bundling

[Holten, 2006]
Hierarchical Edge Bundling

- Flexible and generic method
- Reduces visual clutter when dealing with large numbers of adjacency edges
- Provides an intuitive and continuous way to control the strength of bundling.
  - Low bundling strength mainly provides low-level, node-to-node connectivity information
  - High bundling strength provides high-level information as well by implicit visualization of adjacency edges between parent nodes that are the result of explicit adjacency edges between their respective child nodes

[Holten, 2006]
Bundling Strength

\[ \beta = 0 \quad \beta = 0.25 \quad \beta = 0.5 \quad \beta = 0.75 \quad \beta = 1 \]

[Holten, 2006]
Adjacency Matrix

• Change network to tabular data and use a matrix representation
• Derived data: nodes are keys, edges are boolean values
• Task: lookup connections, find well-connected clusters
• Scalability: millions of edges

• Can encode edge weight, too

Figure 7.5: Comparing matrix and node-link views of a five-node network.

a) Matrix view. b) Node-link view. From [Henry et al. 07], Figure 3b and 3a.

(Permission needed.)

Matrix views of networks can achieve very high information density, up to a limit of one thousand nodes and one million edges, just like cluster heatmaps and all other matrix views that uses small area marks.

Technique: network matrix view
Data Types: network
Derived Data: table: network nodes as keys, link status between two nodes as values
View Comp.: space: area marks in 2D matrix alignment
Scalability: nodes: 1K, edges: 1M

[Henry et al., 2007]
Cliques in Adjacency Matrices

[Gehlenborg and Wong]
Structures from Adjacency Matrices
Node-Link or Adjacency Matrix?

- Empirical study: For most tasks, node-link is better for small graphs and adjacency better for large graphs
- Multi-link paths are hard with adjacency matrices
- Immediate connectivity or neighbors are ok, estimating size (nodes & edges also ok)
- People tend to be more familiar with node-link diagrams
- Link density is a problem with node-link but not with adjacency matrices
Trees

- Trees are directed acyclic graphs
  - each edge has a direction: the origin is the parent, the destination is the child
  - cannot get back to a node after leaving it
- A tree has a root (every other node hangs off it)
- Can consider enclosure in trees using parent-child relationships
Tree Visualizations

![Diagram of tree visualizations](image)

[McGuffin and Robert, 2010]
Node-Link Diagram

- Trees are graphs
- …but we have more structure
- Horizontal or vertical
- Idea 1: partition space for each node via recursion
- Idea 2: “Tidy” Drawing
  - Wetherell & Shannon: Don’t waste space (overlapping parent nodes is ok)
  - Reingold and Tilford: Keep symmetry, subtrees look similar

[WS Alg., Reingold and Tilford, 1981]
Reingold-Tilford Algorithm

- Recurse on left and right subtrees
- Shift subtree over as long as it doesn’t overlap
- Place parent centered above the subtrees
- Originally, only binary trees, extended by Walker

[Reingold and Tilford, 1981]
Icicle Plot

• Line marks
• Vertical position shows depth
• Horizontal position shows links and sibling order
• Scalability: 1 pixel leaves, but harder to label

[Bostock, 2011]
Radial Node-Link

- Use polar coordinates instead of rectilinear
- Same layout algorithms work (e.g. Reingold-Tilford)
- Benefit: space usage, labels
Sunburst

- Icicle plot in a radial layout
- Reading labels?
- Intuitive navigation

[Heer et al., 2012]