Data Visualization (CIS/DSC 468)

Focus+Context

Dr. David Koop
Overview

Reducing Items and Attributes

Filter

- Items
- Attributes

Aggregate

- Items
- Attributes

Reduce

Filter

- Items
- Attributes

Aggregate

- Items
- Attributes

Embed

[Munzner (ill. Maguire), 2014]
Two-Dimensional Binning
Spatial Aggregation

In cartography, changing the boundaries of the regions used to analyze data can yield dramatically different results.

[Penn State, GEOG 486]
Spatial Aggregation

In cartography, changing the boundaries of the regions used to analyze data can yield dramatically different results.
Spatial Aggregation

[Penn State, GEOG 486]
Gerrymandering

- Massachusetts
- Gov. Elbridge Gerry
- South Essex County
- Looks like a salamander

[E. Tisdale, Boston Gazette, 1812]
Modifiable Areal Unit Problem

• How you draw boundaries impacts the type of aggregation you get
• Similar to bins in histograms
• Gerrymandering

[Wonkblog, Washington Post, Adapted from S. Nass]
Boxplots

- Show **distribution**
- Single value (e.g. mean, max, min, quartiles) doesn't convey everything
- Created by John Tukey who grew up in New Bedford!
- Show **spread** and **skew** of data
- Best for **unimodal** data
- Variations like vase plot for multimodal data
- Aggregation here involves many different marks
Four Distributions, Same Boxplot...

[C. Choonpradub and D. McNeil, 2005]
D3 Multiple Views and Interaction

- [http://codepen.io/dakoop/pen/EWGRrE](http://codepen.io/dakoop/pen/EWGRrE)
- Table of games
- Map of schools
- Score (Bar Chart)
- Mouseover in table triggers updates:
  - Map highlights two schools playing in games
  - Score shows the score (home, away)
Exam 2

- **Wednesday, April 5, 12pm-1pm**
- Similar Format to Exam 1: Multiple Choice + Free Response
- Cumulative but emphasis on the topics covered since Exam 1
  - Trees
  - Geospatial Data
  - Color & Colormaps
  - Interaction
  - Multiple Views
  - Filtering & Aggregation
- Reading: Chs. 8.1-8.3, 10, 6, 11, 12, 13
- Be prepared for a similar D3 question to the last exam
Overview

Reducing Items and Attributes

✔️ Filter
- Items
- Attributes

✔️ Aggregate
- Items
- Attributes

Reduce

✔️ Filter

✔️ Aggregate

✔️ Embed

[Munzner (ill. Maguire), 2014]
Attribute Aggregation

• Remember reducing attributes—use statistics: either one variable matches another or doesn't change!

• We can also use similar criteria for aggregating attributes

• **Cluster** similar attributes together
  
  - How?
K-Means
K-Means Issues

Shape

Number of Clusters

[D. Robinson, 2015]
Dimensionality Reduction

• Attribute Aggregation: Use fewer attributes (dimensions) to represent items

• Combine attributes in a way that is more instructive than examining each individual attribute

• Example: Understanding the language in a collection of books
  - Count the occurrence of each non-common word in each book
  - Huge set of features (attributes), want to represent each with an aggregate feature (e.g. high use of "cowboy", lower use of "city") that allows clustering (e.g. "western")
  - Don't want to have to manually determine such rules

• Techniques: Principle Component Analysis, Multidimensional Scaling family of techniques
Principle Component Analysis (PCA)

original data space

Gene 1

Gene 3

Gene 2

PC 1

PC 2

component space

PC 1

PC 2

[ M. Scholz, CC-BY-SA 2.0 ]
Non-linear Dimensionality Reduction

\[ \Phi_{gen} : \mathcal{Z} \rightarrow \mathcal{X} \]

\[ \Phi_{extr} : \mathcal{X} \rightarrow \mathcal{Z} \]

original data space $\mathcal{X}$

component space $\mathcal{Z}$

[M. Scholz, CC-BY-SA 2.0]
Dimensionality Reduction in Visualization

[Glimmer, Ingram et al., 2009]
Tasks in Understanding High-Dim. Data

Task 1
- **What?**
  - In High-dimensional data
  - Out 2D data
- **Why?**
  - Produce
  - Derive

Task 2
- **What?**
  - In 2D data
  - Out Scatterplot Clusters & points
- **Why?**
  - Discover
  - Explore
  - Identify

Task 3
- **What?**
  - In Scatterplot Clusters & points
  - Out Labels for clusters
- **Why?**
  - Produce
  - Annotate

[Munzner (ill. Maguire), 2014]
Focus+Context

• Show everything at once but compress regions that are not the current focus
  - User shouldn't lose sight of the overall picture
  - May involve some aggregation in non-focused regions
  - "Nonliteral navigation" like semantic zooming

• Elision

• Superimposition: more directly tied than with layers

• Distortion
Focus+Content Overview

- **Embed**
  - Elide Data
  - Superimpose Layer
  - Distort Geometry

- **Reduce**
  - Filter
  - Aggregate
  - Embed

[Munzner (ill. Maguire), 2014]
Elision

• There are a number of examples of elision including in text, DOI Trees, …

• Includes both filtering and aggregation but goal is to give overall view of the data

• In visualization, usually correlated with focus regions
Elision: DOI Trees

[Heer and Card, 2004]
Superimposition with Interactive Lenses

(a) Alteration

(b) Suppression

[ChronoLenses and Sampling Lens in Tominski et al., 2014]
Superimposition with Interactive

(c) Enrichment

[Extended Lens in Tominski et al., 2014]
It can be difficult to observe micro and macro features simultaneously with complex graphs. If you zoom in for detail, the graph is too big to view in its entirety. If you zoom out to see the overall structure, small details are lost.

Focus + context techniques allow interactive exploration of an area. Mouseover to distort the nodes.

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Distortion Choices

• How many focus regions?
  - One
  - Multiple

• Shape of the focus?
  - Radial
  - Rectangular
  - Other

• Extent of the focus
  - Constrained similar to magic lenses
  - Entire view changes

• Type of interaction:
  - Geometric, moveable lenses, rubber sheet
Overplotting

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Cartesian Distortion

[M. Bostock, http://bostocks.org/mike/fisheye/]
Cartesian Distortion

[M. Bostock, http://bost.ocks.org/mike/fisheye/]
Figure 3. LiveRAC shows a full day of system management time-series data using a reorderable matrix of area-aware charts. Over 4000 devices are shown in rows, with 11 columns representing groups of monitored parameters. (a): The user has sorted by the maximum value in the CPU column. The first several dozen rows have been stretched to show sparklines for the devices, with the top 13 enlarged enough to display text labels. The time period of business hours has been selected, showing the increase in the In pkts parameter for many devices. (b): The top three rows have been further enlarged to show fully detailed charts in the CPU column and partially detailed ones in Swap and two other columns. The time marker (vertical black line on each chart) indicates the start of anomalous activity in several of spire’s parameters. Below the labeled rows, we see many blocks at the lowest semantic zoom level, and further below we see a compressed region of highly saturated blocks that aggregate information from many charts.

Principle: multiple views are most effective when coordinated through explicit linking.

The principle of linked views [15] is that explicit coordination between views enhances their value. In LiveRAC, as the user moves the cursor within a chart, the same point in time is marked in all charts with a vertical line. Similarly, selecting a time segment in one chart shows a mark in all of them. This technique allows direct comparison between parameter values at the same time on different charts. In addition, people can easily correlate times between large charts with detailed axis labels, and smaller, more concise charts.

Assertion: showing several levels of detail simultaneously provides useful high information density in context.

Several technique choices are based on this assertion. First, LiveRAC uses stretch and squish navigation, where expanding one or many regions compresses the rest of the view [11, 17]. The accompanying video shows the look and feel of this navigation technique. The stretching and squishing operates on rectangular regions, so expanding a single chart also magnifies the entire row for the device it represents, and the entire column for the parameters that it shows. The edges of the display are fixed so that all cells remain within the visible area, as opposed to conventional zooming where some regions are pushed off-screen. There are rapid navigation shortcuts to zoom a single cell, a column, an aggregated group of devices, the results of a search, or to zoom out to an overview. Users can also directly drag grid lines or resize freely drawn on-screen rectangles. Navigation shortcuts can also be created for any arbitrary grouping, whose cells do not need to be contiguous. This interaction mechanism affords multiple focus regions, supporting multiple levels of detail.

Second, charts in LiveRAC dynamically adapt to show visual representations adapted in each cell to the available screen space. This technique, called semantic zooming [13], allows a hierarchy of representations for a group of device-parameter time-series. In Figure 3, the largest charts have multiple overlaid curves and detailed axis and legend labels. Smaller charts show fewer curves and less labeling, and at smaller sizes only one curve is shown as a sparkline [24]. On each curve, the maximum value over the displayed time period is indicated with a red dot, the minimum with a blue dot, and the current value with a green one. All representation levels color code the background rectangle according to dynamically changeable thresholds of the minimum, maximum, or average values of the parameters within the current time window. The smallest view is a simple block, where this color coding is the only information shown.

Third, aggregation techniques achieve visual scalability by ensuring dense regions show meaningful visual representations. Given our target scale of dozens of parameters and thousands of devices, the size of the matrix could easily surpass 100,000 cells. Stretch and squish navigation allows users to quickly create a mosaic with cells of many different sizes. [McLachlan et al., 2008]
Distortion Concerns

• Distance and length judgments are **harder**
  - Example: Mac OS X Dock with Magnification
  - Spatial position of items changes as the focus changes

• Node-link diagrams not an issue… why?

• Users have to be made aware of distortion
  - Back to scatterplot with distortion example
  - Lenses or shading give clues to users

• **Object constancy**: understanding when two viewsrames show the same object
  - What happens under distortion?
  - 3D Perspective is distortion… but we are well-trained for that

• Think about **what** is being shown (filtering) and method (fisheye)