CP SC 8810
Data Visualization

Joshua Levine
levinej@clemson.edu
Agenda

• Final Project Proposals
Last Time

Text
Text Data

- Documents
- Books
- Papers
- Webpages
- Emails
- Twitter posts

- Corpus: collection of documents
Text Data

- Documents
- Books
- Papers
- Webpages
- Emails
- Twitter posts

- Corpus: collection of documents
Tag Clouds / Word Clouds

http://www.tagcrowd.com

http://www.wordle.com
Text Arc

Wattenberg, Viegas 2008
DocuBurst
Arc Diagrams

Analysis of the Characters from Les Misérables: http://mbostock.github.io/protovis/ex/arc.html
Rule-Based: Poetry

Region 1. Phonetic Relations
Region 2. Phonetic Features
Region 3. Phonetic Units And Attributes
Region 4. Word Units And Attributes
Region 5. Semantic Relations

Abdul-Rahman et al. 2008
Text Visualization For Document Collections
Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora

Christopher Collins
Document Cards
(small multiples)
Showing Temporal Relationships: ThemeRiver (Stream Graph)
Jigsaw: Many Linked Views

Visual Analytics Support for Intelligence Analysis
Case Study: The 9/11 Report

Carsten Görg
Youn-ah Kang
Zhicheng Liu
John Stasko

Information Interfaces Group
Georgia Institute of Technology
Jigsaw: Many Linked Views

Visual Analytics Support for Intelligence Analysis
Case Study: The 9/11 Report

Carsten Görg
Youn-ah Kang
Zhicheng Liu
John Stasko

Information Interfaces Group
Georgia Institute of Technology
Sets
Mental Exercise: Consider The Following Dataset

- Item: Lego
- Attributes:
  - Color
  - Height
  - Width
  - Length
  - Shape
Dataset Could Have Other Attributes
The Dataset in Practice
Where Do We Start?
Organization

• Sort by color
Where Do We Start?
Organization

- Sort by size, shape
• Organization is a task where we group data into sets

• Drawbacks?
Set Theory
(Cantor, Dedekind 1870s)

• A **set** is a collection of objects
  • Use capital letters: A, B, C for sets

• An **object** is anything
  • Use lower case letters: a, b, c for objects
  • Objects are contained in sets: \( a \in A \) and \( z \in A \), but \( b \notin A \)

• Discrete sets often written with \{\}
  • Example: \( B = \{c,d,e\} \)
Venn Diagrams: Showing Set Relationships

- Show intersections, unions, complements. All combinations.

\[ A \cup B \]

\[ A \cap B \]
Venn Diagrams
Venn Diagrams: Can Get Messy And Non-Sensical
Euler Diagrams Show Only Valid Relationships
Converting From Venn to Euler
Untangling Euler Diagrams

Tim Dwyer

Published on Sep 10, 2013

Video to accompany our Infovis 2010 paper of the same name.

http://vis.pku.edu.cn/paper/vis2010/i...
Most Frequently Used Words in 10 of Shakespeare’s Plays

Riche, Dwyer 2009
Parallel Sets

Titanic Survivors

http://www.jasondavies.com/parallel-sets/

Bendix, Kosara, Hauser 2005
Bubble Sets:
Add Connection to Color

Collins, Penn, Carpendale 2009
Line Sets: Reduce Clutter

Restaurants

Social Communities

Alper, Riche, Ramos, Czerwinski 2011
Line Sets: Comparison

Alper, Riche, Ramos, Czerwinski 2011
Kelp Diagrams

Cities + Routes

Metabolic Network
Kelp Fusion: Combining Lines + Areas

Meulemans et al. 2013
Lec16
Required Reading
Chapter 8

Arrange Spatial Data

8.1 - 8.3 ONLY!!

8.1 The Big Picture

For datasets with spatial semantics, the usual choice for *arrange* is to *use* the given spatial information to guide the layout. In this case, the choices of *express, separate, order,* and *align* do not apply because the position channel is not available for directly encoding attributes. The two main spatial data types are geometry, where shape information is directly conveyed by spatial elements that do not necessarily have associated attributes, and spatial fields, where attributes are associated with each cell in the field. (See Figure 8.1.) For scalar fields with one attribute at each field cell, the two main visual encoding idiom families are isocontours and direct volume rendering. For both vector and tensor fields, with multiple attributes at each cell, there are four families of encoding idioms: flow glyphs that show local information, geometric approaches that compute derived geometry from a sparse set of seed points, texture approaches that use dense sets of seeds, and feature approaches where data is derived with global computations using information from the entire spatial field.