Data Visualization (CSCI 627/490)

Review

Dr. David Koop
Streamlines & Variants

• Steady vs. **Unsteady** flows
  - In unsteady flows, the vector field **changes** over time
• Variants: **Pathlines** and **Streaklines**
Streamlines & Variants

• Steady vs. **Unsteady** flows
  - In unsteady flows, the vector field **changes** over time

• Variants: **Pathlines** and **Streaklines**

All are identical in steady flows!
Streamline Variants

- **Stream Tubes** [Weiskopf/Machiraju/Möller]
  - Trace two close-by particles
  - Keep distance constant

- **Stream Ribbons** [Weiskopf/Machiraju/Möller]

- **Streak Surfaces** [Krishnan et al., 2009]

- **Streaklines** [NASA]

Fig. 7. A streak surface in the Ellipsoid dataset as depicted in our interactive visualization tool. The surfaces is seeded upstream of the ellipsoid in the initial timestep and shows a prominent bubble that precedes the vortex formation.

Top: Overview; a timeline texture provides temporal orientation.

Bottom left: Surface textured with streak ribbons.

Bottom right: Without texturing, spatial and temporal orientation on the surface is lost.

Fig. 8. Evolution of a time surface in the Ellipsoid dataset. The surface is seeded on rectangle located immediately downstream from the ellipsoid near the temporal beginning of the dataset and illustrates parts of the flow that remain close to the ellipsoid and twist to envelop the nascent vortex system as it forms. A two-dimensional color map helps identify distinct parts of the surface despite heavy overlap.

Fig. 9. Left images: Evolution of a time surface in the delta wing dataset, seeded parallel to the wing tip. The texture provides radial distance stripes to the wing tip for spatial orientation.

Right image: Despite numerical difficulties, the surface mesh remains well-conditioned.
Line Integral Convolution

- Goal: provide a global view of a steady vector field while avoiding issues with clutter, seeds, etc.
- Remember convolution?
- Start with random noise texture
- Smear according to the vector field
- Need structured data

![Input noise](https://example.com/input_noise.png) \( \ast \) ![Kernel](https://example.com/kernel.png) = ![Final image](https://example.com/final_image.png)

[Weiskopf/Machiraju/Möller]
Topology: Find Critical Points
Topology: Find Critical Points
Key development in topological data analysis (TDA)

1. Abstraction of the data: topological structures and their combinatorial representations

2. Separate features from noise: persistent homology

---

Scalar Field Topology

2D Scalar function

Reeb Graph/Contour Tree/Merge Tree

Morse-Smale Complex

[via Levine]
Vector Field Topology

• Instead of “guessing” correct seed points for streamlines to understand the field, try to identify structure (topology) of the field.
Text Visualization: Tag Cloud

- Derived data: number of occurrences of words
- Channel: Font size
- Potential problem: Amount of ink may not be proportional to occurrences…

[Scray, CC-BY-SA-3.0]
Word Tree

Many Eyes word tree provides a choice among three options. The states and make comments. In doing so, they may wish to point to and easily retreat from navigational dead ends. This helps users quickly switch between desired states for comparison.

The word tree animates smoothly to help make clear what is changing. Clicking, when the user switches between two of these options that are not directly connected, creates a smooth transition. If the created transitions are frequent, the user Control-clicks on the desired phrase to create a smooth transition. This feature allows the user to click on browser-like "back" and "forward" buttons to review her previous steps in the visualization.

As with all visualizations on Many Eyes, users can set parameters for the word tree. As the user interacts with the tree—she may click on a branch—these parameter choices are saved for subsequent visits to this word tree.

Finally, the word tree does not provide any sort of "overview" of the text nor does it present an initial search term, there is no obvious entry point—several alternatives are possible. For instance in the current system, the user has clicked on "blind," which appears in one of the branches under "if." This causes the visualization to recenter to the longer phrase "if love be blind, it best agrees with night." The word tree for "love cannot hit the mark" is not shown in the visualization to a "highlighter mode," where clicking on words like, "Note the position of God in this context," and highlighting "God" with translucent brown circles. Thus a user can leave a comment to the visualization to a "highlighter mode," where clicking on words

D. Koop, CSCI 627/490, Fall 2020

[Wattenberg & Viegas, 2007]
Fed Drapes
Clark Coolidge

FELL CAR BUT THE BARN (came) up & smacked me
Who're you, bleeding? Pled.
Blat in back of a Vistrola Car
is so red is such that sun
fell in the rushes & pen bear appear

the white wrong numeral on the wall
can't take if off with the clock
down with the clock it...
way
on the board - couch with brass, kindergarten clench;
joints
backed violet rip into the gas valve
it hemmed & snowed

the wrong way
remnant face
rubber
the pucker

<table>
<thead>
<tr>
<th>Rhymes</th>
<th>Phonetic Rhymes</th>
<th>Character Clusters</th>
<th>Levenshtein Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical Rhyme/Rhyme Riche</td>
<td>AAR AES AEKT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfect Rhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semirhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllabic Rhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consonant Slant Rhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vowel Slant Rhyme</td>
<td></td>
<td></td>
<td>OW1</td>
</tr>
<tr>
<td>Pararhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye Rhyme</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alliteration</td>
<td></td>
<td></td>
<td>KL1</td>
</tr>
<tr>
<td>Assonance</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

[N. McCurdy et al., 2015]
Figure 1: A PTC revealing the differences in drug prevalence amongst the circuits. Ally similar to the connected lists view of Jigsaw [28], however PTCs use size-weighting of words in the display. Shneiderman and Aris [26] have previously explored the contents of a faceted legal document databases using matrix-based visualizations to reveal the number and type of data items matching each facet value. Our work differs in that we seek to aggregate and visualize the contents of the data items, not only their presence or absence. A matrix visualization approach would not be appropriate as our word-selection method, described later, seeks to maximize the differences between corpus subsets. Rather than the single vertical column of words that a words ⇥ facets matrix would contain, our approach allows the entire space to be filled with a wide variety of words. VisGets, or visualization widgets, have been used to explore faceted collections of web-based streaming data [5]. Facets are filtered using scented visual widgets [34] appropriate for the data type, providing both an overview of the available data items and a method to drill down along several facets simultaneously. A tag cloud VisGet consists of a traditional tag cloud summarizing all available documents — text differentiation along a facet is only achieved through interactive brushing. The goal of VisGets is to provide coordinated overview and navigation tools in a faceted information space, where our work is customized to providing meaningful differentiating overviews across facets within large amounts of textual data.

2.2 U.S. Circuit Court Decisions

"Jargon serves lawyers as a bond of union: it serves them, at every word, to remind them of that common interest, by which they are made friends to one another, enemies to the rest of mankind."

Jeremy Bentham [2, 292]

Figure 2: US Court Circuits are multi-state regions. The words of the iconoclast Bentham were not the last written on the topic of legal language. Law and language meet in many academic ways: forensic linguists help solve crimes, judges make semantic rulings on unclear contract wording, and social scholars take a high-level view, studying the language of lawyers and judges [29]. By analyzing the written decisions of the US Circuit Courts of Appeal, we hope to shed light on thematic and potentially linguistic differences between subsets of the data. Differences in word usage between courts has been previously studied using legal databases as a source for historical lexicography [8]. However, in that work, text-based searches provided information on particular words of interest. Through text mining and visualization, we select words of interest and provide a broad overview as an entry point to deeper analysis.

The US Circuit Courts of Appeal are made up of 12 regionally-based court divisions (numbered First through Eleventh, plus the DC Circuit) and the Federal Circuit, which hears cases of national relevance, such as patent-related appeals (see Fig. 2). This data contains 628,000 court decisions, each labeled by circuit. The judgments are faceted, because they can be organized along several dimensions, such as the lead authoring judge, the decision length, the date of the decision, or whether the lower court was upheld or overturned. For [Collins et al., 2009]
Project

- Presentations on **Thursday**:
  - Turn in code for the visualization to **Blackboard** by **Dec. 2 at 11:59pm**
  - **5 minutes** per presenter/group
  - Showcase the visualization (not slides)
    - **Brief** introduction to your data and questions
    - Discuss **design** decisions
    - **Demonstrate** the interactive features of your project
  - For groups, one person should drive but both can help present
- Have until Dec. 6 to turn in final code and report
- Note two assignments on Blackboard (one for presentation, one for report)
Final Exam

- December 10, 2020, 10-11:50am
- Covers all topics but emphasizes second half of the course
- Similar format as Midterm (multiple choice, free response)
- 627 Students will have extra questions related to the research papers
- Questions?
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”

— T. Munzner
## Dataset Types

- **Tables**
  - Attributes (columns)
  - Items (rows)
  - Cell containing value

- **Networks**
  - Link
  - Node (item)

- **Fields (Continuous)**
  - Grid of positions
  - Cell
  - Attributes (columns)
  - Value in cell

- **Geometry (Spatial)**
  - Position

- **Multidimensional Table**
  - Key 1
  - Key 2
  - Attributes
  - Value in cell

- **Trees**
  - + Sets
  - + Text

[Munzner (ill. Maguire), 2014]
Tasks

Why?

Analyse
- Consume
  - Discover
  - Present
  - Enjoy
- Produce
  - Annotate
  - Record
  - Derive

Search
- Target known
  - Location known
    - Lookup
    - Browse
  - Location unknown
    - Locate
    - Explore
- Target unknown

Query
- Identify
- Compare
- Summarize

Why?

All Data
- Trends
- Outliers
- Features

Attributes
- One
  - Distribution
- Many
  - Dependency
  - Correlation
  - Similarity

Network Data
- Topology
  - Paths
- Copy Data
  - Shape

Spatial Data

[D. Koop, CSCI 627/490, Fall 2020]
## How do we do visualization?

### Encode

<table>
<thead>
<tr>
<th>Arrange</th>
<th>Express</th>
<th>Separate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order</td>
<td>Align</td>
<td></td>
</tr>
<tr>
<td>Use</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Map

- **from categorical and ordered attributes**
- **Color**
  - Hue
  - Saturation
  - Luminance
- **Size, Angle, Curvature, ...**
- **Shape**
  - +
  - ○
  - ■
  - △
- **Motion**
  - Direction, Rate, Frequency, ...

### Manipulate

- **Change**
- **Select**
- **Navigate**

### Facet

- **Juxtapose**
- **Partition**
- **Superimpose**

### Reduce

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

---

[Muñoz (ill. Maguire), 2014]
Visual Encoding

• How do we encode data visually?
  - **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
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]
Design

- **Domain situation**
  You misunderstood their needs

- **Data/task abstraction**
  You’re showing them the wrong thing

- **Visual encoding/interaction idiom**
  The way you show it doesn’t work

- **Algorithm**
  Your code is too slow

[Munzner (ill. Maguire), 2014]
Design Guidelines

- Tufte:
  - Show data variation, not design variation
  - Clear, detailed, and thorough labeling and appropriate scales
  - Size of the graphic effect should be directly proportional to the numerical quantities ("lie factor")
Design Analysis: What is Wrong Here?
D3

- http://d3js.org/
- Supports data as a core piece of Web elements
  - Correspondence between data and DOM elements
  - Dealing with changing data (joins, enter/update/exit)
  - Data drives the marks and channels
- Selections (similar to CSS) that allow greater manipulation
- Integrated layout algorithms, axes calculations, etc.
- Focus on interaction support
  - Straightforward support for transitions
  - Event handling support for user-initiated changes
Arrange Tables

Express Values

Separate, Order, Align Regions

Separate

Order

Align

1 Key
List

2 Keys
Matrix

3 Keys
Volume

Many Keys
Recursive Subdivision

Axis Orientation

Rectilinear

Parallel

Radial

[Munzner (ill. Maguire), 2014]
Categorical Map

[Map of the United States showing different food categories by state]
Map with Two Variables

This map removes mostly uninhabited areas, revealing Mr. Bush’s suburban and rural support in the East and South.

[Map with Two Variables]

[Source: M. Ericson, New York Times]
Rectangular Cartogram

1860 Presidential Election

<table>
<thead>
<tr>
<th>CANDIDATE</th>
<th>PARTY</th>
<th>ELECTORAL VOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abraham Lincoln</td>
<td>Republican</td>
<td>180</td>
</tr>
<tr>
<td>John C. Breckinridge</td>
<td>Southern Democratic</td>
<td>72</td>
</tr>
<tr>
<td>John Bell</td>
<td>Constitutional Union</td>
<td>39</td>
</tr>
<tr>
<td>Stephen A. Douglas</td>
<td>Northern Democratic</td>
<td>12</td>
</tr>
</tbody>
</table>

Each state is sized by the number of electoral votes it had in 1860.
Arrange Networks and Trees

- **Node–Link Diagrams**
  Connection Marks
  
  ![Node–Link Diagrams]

- **Adjacency Matrix**
  Derived Table
  
  ![Adjacency Matrix]

- **Enclosure**
  Containment Marks
  
  ![Enclosure]

[Munzner (ill. Maguire), 2014]
Tree Visualizations

Fig. 1. Several basic kinds of tree representations, here each node-link [23, 7].

Abstract

A variety of graphical representations are available for depicting tree structures (Figure 1), from "classical" node-link diagrams [23, 7], to treemaps [14, 26, 6, 30], concentric circles [2, 27, 31], and many others.

Index Terms

—Tree visualization, graph drawing, efficiency metrics.

Quantifying the Space-Efficiency

of area, aspect ratio, label size, or other measures. However, there is more to space-efficiency than total area. Treemaps are often described as optimally space-efficient, not just because they have a total area of 1, but also because they allow for weighted partitioning of area, and incidentally have no need for margins between the borders of nodes as treemaps often do.

This article identifies several metrics related to space-efficiency, and in particular, introduces a metric of total area; (2) analyzing the area of smallest nodes (i.e. the leaf nodes) in the representation, in addition to the overall representation might be possible. Likewise, concentric circles and nested circles (Figures 1E and 1F) are equally efficient (and both optimal) according to this metric of space-efficiency used, and "optimal" space-efficiency is defined as a total area of 1 (possibly partitioned by weight), then we can distinguish the four possibilities in Figure 2, e.g. in terms of their respective mean area exponent as well as the area they allocate to non-labels, suggesting that a more space-efficient (in terms of label size) representation might be possible.

Clearly, it would be useful to have some way to quantitatively distinguish the four cases. If alternative metrics are defined as a total area of 1 (possibly partitioned by weight), then we can indeed desirable properties, however they are not unique to treemaps.

This article shows that there are finer ways of distinguishing efficiency, i.e. assuming the representation is bound within a 1-square, both have a total area of 1, and are equally efficient (and both optimal) according to this metric of space-efficiency used, and "optimal" space-efficiency is defined as a total area of 1 (possibly partitioned by weight), then we can

Fig. 2 shows that icicle diagrams also allow for a weighted partitioning of area, and incidentally have no need for margins between the borders of nodes as treemaps often do.

One basic metric of space-efficiency is the total area of a representation, and it is unclear what approach would be generally accepted standard set of metrics for evaluating the space-efficiency of such representations is how efficiently they use area, aspect ratio, label size, or other measures. However, there is more to space-efficiency than total area.
Treemaps

- Containment marks instead of connection marks
- Encodes some attribute of the items as the size of the rectangles
- Not as easy to see the intermediate rectangles
- Scalability: millions of leaf nodes and links possible
- Need a layout algorithm!
  - Slice-and-Dice vs. Squarify
  - Viewing Hierarchy: Cushion Treemap
Set Visualizations

- How to show the intersection of sets?
Human Color Perception

Metamerism: same three responses == same color

[via M. Meyer]
Avoid Rainbow Colormaps!

[Borland & Taylor, 2007]
Colormaps

D. Koop, CSCI 627/490, Fall 2020

[Munzner (ill. Maguire), 2014]
Interaction Overview

→ Change over Time

→ Navigate

→ Item Reduction

→ Zoom Geometric or Semantic

→ Pan/Translate

→ Constrained

→ Attribute Reduction

→ Slice

→ Cut

→ Project

[Munzner (ill. Maguire), 2014]
Staged Animated Transitions

[M. Bostock]
Staged Animated Transitions

Grouped

Stacked
Multiple Views

- Juxtapose and Coordinate Multiple Side-by-Side Views
  - Share Encoding: Same/Different
    - Linked Highlighting
  - Share Data: All/Subset/None
  - Share Navigation

<table>
<thead>
<tr>
<th>Encoding</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same</td>
<td>Redundant</td>
</tr>
<tr>
<td></td>
<td>Overview/Detail</td>
</tr>
<tr>
<td></td>
<td>Small Multiples</td>
</tr>
<tr>
<td>Different</td>
<td>Multiform</td>
</tr>
<tr>
<td></td>
<td>Multiform, Overview/Detail</td>
</tr>
<tr>
<td></td>
<td>No Linkage</td>
</tr>
</tbody>
</table>

- Partition into Side-by-Side Views
- Superimpose Layers

[Munzner (ill. Maguire), 2014]
Filtering and Aggregation

Reducing Items and Attributes

- **Filter**
  - Items
  - Attributes

- **Aggregate**
  - Items
  - Attributes

Reduce

- Filter
- Aggregate
- Embed

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

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

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

[Munzner (ill. Maguire), 2014]
Fields in Visualization

Scalar Fields

Vector Fields

Tensor Fields

Each point in space has an associated...
Fields in Visualization

Scalar Fields
(Order-0 Tensor Fields)

Vector Fields
(Order-1 Tensor Fields)

Tensor Fields
(Order-2+)

Each point in space has an associated...

Scalar:

\[ s_0 \]

Vector:

\[
\begin{bmatrix}
  v_0 \\
  v_1 \\
  v_2
\end{bmatrix}
\]

Tensor:

\[
\begin{bmatrix}
  \sigma_{00} & \sigma_{01} & \sigma_{02} \\
  \sigma_{10} & \sigma_{11} & \sigma_{12} \\
  \sigma_{20} & \sigma_{21} & \sigma_{22}
\end{bmatrix}
\]
Isosurfacing

(a) An isosurfaced tooth.

(b) Multiple isosurfaces. [J. Kniss, 2002]
Volume Rendering

[Image: Black and white rabbit scan next to a 3D color rendering of a structure.

[J. Kniss, 2002]
Vector Fields

Streaklines [NASA]

Stream Ribbons [Weiskopf/Machiraju/Möller]

Stream Tubes [Weiskopf/Machiraju/Möller]

Mapping Methods Based on Particle Tracing
- Stream ribbons
  - Trace two close-by particles
  - Keep distance constant

- Stream tubes
  - Specify contour, e.g. triangle or circle, and trace it through the flow

Streaklines in real life
© Weiskopf/Machiraju/Möller
The purpose of computing is about insight, not numbers

– R. W. Hamming
The purpose of visualization is about **insight**, not pictures

– Card, Mackinlay, Schneiderman
Projects Thursday