Assignment 2: Exploratory Data Analysis

Use existing software to formulate & answer questions

First steps
- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate

Create visualizations
- Interact with data
- Question will evolve
- Tableau

Make wiki notebook
- Keep record of all steps you took to answer the questions

Due before class on Sep 30, 2013
Announcements

Auditors, please enroll in the class (1 unit P/NP)
- Requirements: Come to class and participate (online as well)
- Requirements: Assignment 1

Course takers (including 3 unit P/NP)
- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question by 3pm on day of lecture
- All 3 assignments and final project

All, add yourself to participants page on the wiki

Class wiki
http://vis.berkeley.edu/courses/cs294-10-fa13/wiki/

Announcements

M Oct 7: Introduction to D3 (Scott Murray)
- Look at his online tutorials (part of reading)
- Bring your laptop on Monday
- Get a webserver running on your laptop

W Oct 9: Wrangling Data (Sean Kandel)

Class wiki
http://vis.berkeley.edu/courses/cs294-10-fa13/wiki/
Assignment 3: Visualization Software

Create a small interactive visualization application – you choose data domain and visualization technique.

1. Describe data and storyboard interface
2. Implement interface and produce final writeup
3. Submit the application and a final writeup on the wiki

Can work alone or in pairs
Final write up due before class on Oct 16, 2013

Interaction
Gulf of Execution
The difference between the user’s intentions and the allowable actions.

Gulf of Evaluation
The amount of effort that the person must exert to interpret the state of the system and to determine how well the expectations and intentions have been met.
Conceptual model: x,y correlated?

Real world:

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.67</td>
<td>0.79</td>
</tr>
<tr>
<td>0.32</td>
<td>0.85</td>
</tr>
<tr>
<td>0.39</td>
<td>0.72</td>
</tr>
<tr>
<td>0.27</td>
<td>0.85</td>
</tr>
<tr>
<td>0.71</td>
<td>0.43</td>
</tr>
<tr>
<td>0.63</td>
<td>0.09</td>
</tr>
<tr>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>0.20</td>
<td>0.54</td>
</tr>
<tr>
<td>0.51</td>
<td>0.38</td>
</tr>
<tr>
<td>0.11</td>
<td>0.33</td>
</tr>
<tr>
<td>0.46</td>
<td>0.46</td>
</tr>
</tbody>
</table>
Gulf of evaluation

- Conceptual model: \( x, y \) correlated?
- Real world: \( \rho = -0.29 \)

Gulf of execution

- Conceptual model: Draw a scatterplot
- Real world:
  - Move 90 30
  - Rotate 35
  - Pen down
  - ...
Gulf of execution

Conceptual model: Draw a scatterplot

Real world

Execution

Topics

Early interactive systems
Pointing
Brushing and linking
Dynamic queries
Sorting
Generalized selections
Early Systems

[Graphics and Graphic Information Processing, Bertin 81]
[Graphics and Graphic Information Processing, Bertin 81]
Basic Pointing Methods

- Point Selection
- Mouse Hover / Click
- Touch / Tap
- Select Nearby Element (e.g., Bubble Cursor)
Basic Pointing Methods

Point Selection
Mouse Hover / Click
Touch / Tap
Select Nearby Element (e.g., Bubble Cursor)

Region Selection
Rubber-band or Lasso
Area Cursors ("Brushes")
Brushing and Linking

Highlighting
Focus user attention on a subset of the data within one graph [from Wills 95]
Brushing

- Interactively select subset of data
- See selected data in other views
- Two things (normally views) must be *linked* to allow for brushing

Brushing Scatterplots, Becker & Cleveland 1982
Baseball statistics [from Wills 95]

- How long in majors
- Avg assists vs avg putouts (fielding ability)
- Distribution of positions played
- Select high salaries
- Avg career HRs vs avg career hits (batting ability)

Linking assists to positions
GGobi: Brushing

http://www.ggobi.org/

Dynamic Queries
Query and results

SELECT house
FROM east bay
WHERE price < 1,000,000 AND bedrooms > 2
ORDER BY price

Issues

1. For programmers
2. Rigid syntax
3. Only shows exact matches
4. Too few or too many hits
5. No hint on how to reformulate the query
6. Slow question-answer loop
7. Results returned as table
HomeFinder

[Ahlberg and Schneiderman 92]

Direct manipulation

1. Visual representation of objects and actions
2. Rapid, incremental and reversible actions
3. Selection by pointing (not typing)
4. Immediate and continuous display of results
Zipdecode [from Fry 04]

http://acg.media.mit.edu/people/fry/zipdecode/

NameVoyager

http://www.babynamewizard.com/voyager
**TimeSearcher** [Hochheiser & Schneiderman 02]

Based on Wattenberg's [2001] idea for sketch-based queries of time-series data.

---

**3D dynamic queries** [Akers et al. 04]
3D dynamic queries [Akers et al. 04]

Pros and cons

Pros
- Controls useful for both novices and experts
- Quick way to explore data

Cons
- Simple queries
- Lots of controls
- Amount of data shown limited by screen space

Who would use these kinds of tools?
Sorting

Trellis
[Becker, Cleveland, and Shyu 96]
Trellis
[Becker, Cleveland, and Shyu 96]
Generalized Selection

Visual Queries

Model selections as declarative queries

(-118.371 ≤ lon AND lon ≤ -118.164) AND (33.915 ≤ lat AND lat ≤ 34.089)
Visual Queries

Model selections as declarative queries
Applicable to dynamic, time-varying data
Retarget selection across visual encodings
Perform operations on query structure

“Select items like this one.”
Generalized Selection

Point to an example and define an abstraction based on one or more properties [Clark, Brennan]

“Blue like this”
“The same shape as that”

Abstraction may occur over multiple levels

This is not a sentence.
Generalized Selection

Provide *generalization mechanisms* that enable users to *expand a selection query* along *chosen dimensions* of interest.

Expand selections via *query relaxation*.
Query Builder

Click: Select Items
(id = 'China')

Drag: Select Range
(2000 < gni AND gni < 10000) AND (.1 < internet AND internet < .2)

Legend: Select Attributes
(region = 'The Americas')

Interactor

Query Visualizer

Query Builder

(id = 'China')
Query Relaxation

Generalize an input query to create an expanded selection, according to:

1. A semantic structure describing the data
2. A traversal policy for that structure
Relaxation using Hierarchies

Relax using abstraction hierarchies of the data
Traverse in direction of increasing generality

Examples
A Priori: Calendar, Categories, Geography
Data-Driven: Nearest-Neighbor, Clustering
Relaxation using Attributes

If no explicit semantic structure is available, treat data itself as a “flat” hierarchy

Select all items with matching values along the attributes chosen for relaxation

Relaxation of Networks

![Diagram of network relaxation](image)
Lesson

Consider how the structure and/or semantics of the data might be leveraged to aid analysis

Extension: look beyond data features to incorporate perceptual features of the display

Perceptual Annotation [Kong & Agrawala 09]
Summary

Most visualizations are interactive
- Even passive media elicit interactions

Good visualizations are task dependant
- Choose the right space
- Pick the right interaction technique

Human factors are important
- Leverage human strengths
- Assist to get past human limitations