User Tasks & Analysis

CS 4460 – Intro. to Information Visualization
September 29, 2017
John Stasko

Learning Objectives

- Understand the importance of tasks, goals, and objectives for visualization
- Identify the common "low-level" tasks for visualizations
- Identify important "high-level" tasks for visualizations
What for?

- In order to build better visualizations, we need to understand what people might use them for
  - What tasks do they want to accomplish?

An Example

- search vs. browsing

- Value of Vis day (coming up):
  - Exploratory data analysis
  - Identifying better questions
  - Understanding, awareness, context, trust
Browsing vs. Search

• Important difference in activities
• Appears that information visualization may have more to offer to browsing

• But...browsing is a softer, fuzzier activity
• So, how do we articulate utility?
  – Maybe describe when it’s useful
  – When is browsing useful?

Browsing

• Useful when
  – Good underlying structure so that items close to one another can be inferred to be similar
  – Users are unfamiliar with collection contents
  – Users have limited understanding of how system is organized and prefer less cognitively loaded method of exploration
  – Users have difficulty verbalizing underlying information need
  – Information is easier to recognize than describe

Lin '97
Thought

- Maybe infovis isn’t about answering questions or solving problems... hmmm
- Maybe it’s about asking better questions

Tasks

- OK, but browsing and search are very high level
- Let’s be more specific...
Challenge

Whiskeys

Come up with analytic queries, tasks, goals...

Follow-on

- What are the (types of) tasks being done here?
  - Abstract away the domain
- Can you think of others?
Task Taxonomies

- Number of different ones exist, important to understand what process they focus on
  - Creating an artifact
  - Human tasks
  - Tasks using visualization system
  - ...

One (Famous) Perspective

- Shneiderman proposed task × data type taxonomy to understand what people do with visualization
- Mantra: “Overview first, zoom and filter, then details on demand”
  - Design paradigm for infovis systems

Shneiderman
VL ’96
Taxonomy

• Data Types
  1. 1D
  2. 2D
  3. 3D
  4. Temporal
  5. ND
  6. Tree
  7. Network

• Tasks
  1. Overview
  2. Zoom
  3. Filter
  4. Details-on-demand
  5. Relate
  6. History
  7. Extract

Another Task Taxonomy

• Amar, Eagan, & Stasko – InfoVis ‘05
Background

- Use "commercial tools" class assignment from this class
- Students generate questions to be answered using commercial infovis systems
- Data sets:

<table>
<thead>
<tr>
<th>Domain</th>
<th>Data cases</th>
<th>Attributes</th>
<th>Questions Generated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cereals</td>
<td>78</td>
<td>15</td>
<td>107</td>
</tr>
<tr>
<td>Mutual funds</td>
<td>987</td>
<td>14</td>
<td>41</td>
</tr>
<tr>
<td>Cars</td>
<td>407</td>
<td>10</td>
<td>153</td>
</tr>
<tr>
<td>Films</td>
<td>1742</td>
<td>10</td>
<td>169</td>
</tr>
<tr>
<td>Grocery surveys</td>
<td>5164</td>
<td>8</td>
<td>126</td>
</tr>
</tbody>
</table>

- Generated 596 total analysis tasks
Terminology

- *Data case* – An entity in the data set
- *Attribute* – A value measured for all data cases
- *Aggregation function* – A function that creates a numeric representation for a set of data cases (e.g., average, count, sum)

1. Retrieve Value

**General Description:**
Given a set of specific cases, find attributes of those cases.

**Examples:**
- What is the mileage per gallon of the Audi TT?
- How long is the movie Gone with the Wind?
2. Filter

**General Description:**
Given some concrete conditions on attribute values, find data cases satisfying those conditions.

**Examples:**
- What Kellogg's cereals have high fiber?
- What comedies have won awards?
- Which funds underperformed the SP-500?

3. Compute Derived Value

**General Description:**
Given a set of data cases, compute an aggregate numeric representation of those data cases.

**Examples:**
- What is the gross income of all stores combined?
- How many manufacturers of cars are there?
- What is the average calorie content of Post cereals?
4. Find Extremum

**General Description:**
Find data cases possessing an extreme value of an attribute over its range within the data set.

**Examples:**
- What is the car with the highest MPG?
- What director/film has won the most awards?
- What Robin Williams film has the most recent release date?

5. Sort

**General Description:**
Given a set of data cases, rank them according to some ordinal metric.

**Examples:**
- Order the cars by weight.
- Rank the cereals by calories.
6. Determine Range

**General Description:**
Given a set of data cases and an attribute of interest, find the span of values within the set.

**Examples:**
- What is the range of film lengths?
- What is the range of car horsepower?
- What actresses are in the data set?

7. Characterize Distribution

**General Description:**
Given a set of data cases and a quantitative attribute of interest, characterize the distribution of that attribute’s values over the set.

**Examples:**
- What is the distribution of carbohydrates in cereals?
- What is the age distribution of shoppers?
8. Find Anomalies

**General Description:**
Identify any anomalies within a given set of data cases with respect to a given relationship or expectation, e.g., statistical outliers.

**Examples:**
- Are there any outliers in protein?
- Are there exceptions to the relationship between horsepower and acceleration?

9. Cluster

**General Description:**
Given a set of data cases, find clusters of similar attribute values.

**Examples:**
- Are there groups of cereals w/ similar fat/calories/sugar?
- Is there a cluster of typical film lengths?
10. Correlate

**General Description:**
Given a set of data cases and two attributes, determine useful relationships between the values of those attributes.

**Examples:**
- Is there a correlation between carbohydrates and fat?
- Is there a correlation between country of origin and MPG?
- Do different genders have a preferred payment method?
- Is there a trend of increasing film length over the years?

**Discussion/Reflection**

- **Compound tasks**
  - “Sort the cereal manufacturers by average fat content”
    - Compute derived value; Sort
  - “Which actors have co-starred with Julia Roberts?”
    - Filter; Retrieve value
**Discussion/Reflection**

- What questions were left out?
  - Basic math
    “Which cereal has more sugar, Cheerios or Special K?”
    “Compare the average MPG of American and Japanese cars.”
  - Uncertain criteria
    “Does cereal (X, Y, Z…) sound tasty?”
    “What are the characteristics of the most valued customers?”
  - Higher-level tasks
    “How do mutual funds get rated?”
    “Are there car aspects that Toyota has concentrated on?”
  - More qualitative comparison
    “How does the Toyota RAV4 compare to the Honda CRV?”
    “What other cereals are most similar to Trix?”

**Concerns/Limitations**

- InfoVis tools may have influenced students’ questions
- Graduate students as group being studied
  - How about professional analysts?
- Subjective – Not an exact science
- Data was really quantitative so may get a different set of tasks for relational/graph data
  - See Lee et al, BELIV ’06
Contributions

• Set of *grounded* low-level analysis tasks
• Potential use of tasks as a language/vocabulary for comparing and evaluating infovis systems

Another Perspective

• Taxonomy proposed
• “…used specifically for multidimensional visualizations, taking into account the generic objectives that a user has when using such techniques to perform exploratory analyses as a previous step of statistical analysis.”
Task Taxonomy

- 7 tasks in 2 categories
  - User goals
    - Identify – Find, discover new information
    - Determine – Calculate, define a precise value
    - Compare – Compare data & values
    - Infer – Infer knowledge, generate hypotheses
    - Locate – Search and identify information
  - Intermediate level tasks to support analysis
    - Visualize – Represent the data a certain way
    - Configure – Normalize, filter, reorder, etc.

More Details

- Each task has “parameters”
  - Identify
    - clusters
    - correlations
    - categories
    - properties
    - patterns
    - characteristics
    - thresholds
    - similarities
    - differences
    - dependencies
    - uncertainties
    - variations
Abstract Tasks

Framework/Typology of abstract visualization tasks

Brehmer & Munzner
TVCG (InfoVis) '13

Why?

What are the top-level categories (answers) to the “Why?” question?
Discover

High to low level

How?

Fall 2016  CS 7450  39

Fall 2016  CS 7450  40
Targets

What are the types of targets?

Visual Analytic Activity

Sedig, Parsons, Babanski

JMPT’12
Learning Objectives

- Understand the importance of tasks, goals, and objectives for visualization
- Identify the common "low-level" tasks for visualizations
- Identify important "high-level" tasks for visualizations

HW 4

- Questions?
  - Get Tableau
  - Pick data set

- Due next Friday, 6th at 1pm (T-Square)
Midterm Exam

- About 15 short answer questions
- Closed book/notes
- Examples
  - Concepts (majority)
  - Analyze a vis
  - Design a vis
  - Analyze some code

Upcoming

- Midterm Exam

- No (Scheller 300) class next Wed & Fri
  - John T. and I are away all next week

- Value of Vis
  - Prep: Norman & Chang articles
  - Watch My EuroVis '14 Capstone talk
References

- Spence & CMS texts
- All referred to papers