User Tasks & Analysis

CS 4460 – Intro. to Information Visualization
August 28, 2014
John Stasko

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

- During intro:
  - 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...

Example from Earlier

Questions:

Which cereal has the most/least potassium?
Is there a relationship between potassium and fiber? If so, are there any outliers?
Which manufacturer makes the healthiest cereals?

<table>
<thead>
<tr>
<th>Brand</th>
<th>Manufacturer</th>
<th>Fiber</th>
<th>Potassium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honeycomb</td>
<td>P</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Just Right</td>
<td>K</td>
<td>2</td>
<td>95</td>
</tr>
<tr>
<td>Life</td>
<td>G</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Lids</td>
<td>G</td>
<td>0</td>
<td>55</td>
</tr>
<tr>
<td>Munch Chews</td>
<td>B</td>
<td>3</td>
<td>170</td>
</tr>
<tr>
<td>Multi Grain Chex</td>
<td>G</td>
<td>2</td>
<td>90</td>
</tr>
<tr>
<td>Non-Graham Flakes</td>
<td>K</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>Nature's Path</td>
<td>K</td>
<td>3</td>
<td>90</td>
</tr>
<tr>
<td>Original Size</td>
<td>G</td>
<td>15</td>
<td>120</td>
</tr>
<tr>
<td>Puff N' Stuff</td>
<td>B</td>
<td>6</td>
<td>260</td>
</tr>
<tr>
<td>Protein Plus</td>
<td>K</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>Quaker Oatmeal</td>
<td>G</td>
<td>2.7</td>
<td>170</td>
</tr>
<tr>
<td>Raisin Bran</td>
<td>K</td>
<td>6</td>
<td>240</td>
</tr>
<tr>
<td>Raisin Bran</td>
<td>G</td>
<td>3</td>
<td>140</td>
</tr>
<tr>
<td>Rice Krispies</td>
<td>K</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td>Rough Cut</td>
<td>G</td>
<td>0</td>
<td>95</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td>N</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td>N</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td>N</td>
<td>3</td>
<td>130</td>
</tr>
<tr>
<td>Shredded Wheat</td>
<td>N</td>
<td>4</td>
<td>140</td>
</tr>
<tr>
<td>Special K</td>
<td>K</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>Shredded Fruit</td>
<td>N</td>
<td>3</td>
<td>95</td>
</tr>
<tr>
<td>Total Corn Flakes</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Total Grain</td>
<td>G</td>
<td>4</td>
<td>230</td>
</tr>
<tr>
<td>Total Grain</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>Total Grain</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Total Wheat</td>
<td>G</td>
<td>2</td>
<td>110</td>
</tr>
<tr>
<td>Total Wheat</td>
<td>G</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Total Wheat</td>
<td>G</td>
<td>3</td>
<td>110</td>
</tr>
<tr>
<td>Total Wheat</td>
<td>G</td>
<td>1</td>
<td>60</td>
</tr>
</tbody>
</table>
Exercise

• What are the (types of) tasks being done here?
• Can you think of others?
  – Let’s develop a list

Task Taxonomies

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

- Wehrend & Lewis created a low-level, domain independent taxonomy of user tasks in visualization environments
- Eleven basic actions
  - identify, locate, distinguish, categorize, cluster, distribution, rank, compare within relations, compare between relations, associate, correlate

Another 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
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 (eg, 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 horsepowers?
- 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
Surveys

HW back next time (examples)

Project example

Interaction

• User goals and tasks carried out through interaction with visualization
  – The interactive dialog helps people explore
Interaction Framework

- Organized along user intent
- 7 categories
  - Select
  - Explore
  - Reconfigure
  - Encode
  - Abstract/elaborate
  - Filter
  - Connect

More to come later on interaction day

Interactive Dynamics

- “taxonomy of interactive dynamics that contribute to successful analytic dialogues”
  - part interaction, part task

<table>
<thead>
<tr>
<th>Data and View Specification</th>
<th>Visualize data by choosing visual encodings.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Filter out data to focus on relevant items.</td>
</tr>
<tr>
<td></td>
<td>Sort items to expose patterns.</td>
</tr>
<tr>
<td></td>
<td>Derive values or models from source data.</td>
</tr>
<tr>
<td>View Manipulation</td>
<td>Select items to highlight, filter, or manipulate them.</td>
</tr>
<tr>
<td></td>
<td>Navigate to examine high-level patterns and low-level detail.</td>
</tr>
<tr>
<td></td>
<td>Coordinate views for linked, multidimensional exploration.</td>
</tr>
<tr>
<td></td>
<td>Organize multiple windows and workspaces.</td>
</tr>
<tr>
<td>Process and Provenance</td>
<td>Record analysis histories for review, retrieval, and sharing.</td>
</tr>
<tr>
<td></td>
<td>Annotate patterns to document findings.</td>
</tr>
<tr>
<td></td>
<td>Share views and annotations to enable collaboration.</td>
</tr>
<tr>
<td></td>
<td>Middle users through analysis tasks or stories.</td>
</tr>
</tbody>
</table>

Yi et al. TVCG ‘07

Heer & Shneiderman CACM ‘12
Abstract Tasks

Framework/Typology of abstract visualization tasks

Why?

What are the top-level categories (answers) to the "Why?" question?
Actions

What are the three types of actions?

```
<table>
<thead>
<tr>
<th>use</th>
<th>produce</th>
<th>consume</th>
</tr>
</thead>
<tbody>
<tr>
<td>present</td>
<td>enjoy</td>
<td>annotate</td>
</tr>
<tr>
<td>discover</td>
<td>record</td>
<td>derive</td>
</tr>
<tr>
<td>generate</td>
<td>verify</td>
<td></td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>query</th>
<th>identify</th>
<th>compare</th>
<th>summarize</th>
</tr>
</thead>
<tbody>
<tr>
<td>one</td>
<td>some</td>
<td>all</td>
<td></td>
</tr>
</tbody>
</table>
```

Targets

What are the types of targets?

```
<table>
<thead>
<tr>
<th>targets</th>
<th>all data</th>
<th>trends</th>
<th>outliers</th>
<th>features</th>
</tr>
</thead>
<tbody>
<tr>
<td>attributes</td>
<td>one</td>
<td>distribution</td>
<td>extreme</td>
<td>value</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dependency</td>
<td>correlation</td>
<td>similarity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>network data</td>
<td>topology</td>
<td>paths</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>spatial data</td>
<td>shape</td>
</tr>
</tbody>
</table>
```
Can InfoVis Be More?

- Is InfoVis helping people enough?
- What do we need to do to provide even more value?

Providing Better Analysis

- Combine computational analysis approaches such as data mining with infovis
  - Too often viewed as competitors in past
- Each has something to contribute
Issues

• Issues influencing the design of discovery tools:
  – Statistical Algorithms vs. Visual data presentation
  – Hypothesis testing vs. exploratory data analysis

• Pro’s and Con’s?

Differing Views

• Hypothesis testing
  – Advocates:
    By stating hypotheses up front, limit variables and sharpens thinking, more precise measurement
  – Critics:
    Too far from reality, initial hypotheses bias toward finding evidence to support it

• Exploratory Data Analysis
  – Advocates:
    Find the interesting things this way, we now have computational capabilities to do them
  – Skeptics:
    Not generalizable, everything is a special case, detecting statistical relationships does not infer cause and effect
Recommendations

- Integrate data mining and information visualization
- Allow users to specify what they are seeking
- Recognize that users are situated in a social context
- Respect human responsibility

Related Detour

- Your projects are “design studies”
  - Problem-driven visualization research
  - Assist clients with data who want to understand it better
  - Design and build visualization system

- How do you do it well?
Reflects on 21 design studies from 3 authors & reviewing others

Definition

“...A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines.”
Problem Suitability

Framework

Fig. 2. Nine-stage design study methodology framework classified into three top-level categories. While outlined as a linear process, the overlapping stages and gray arrows imply the iterative dynamics of this process.
Considerations

- **Practical**
  - Data: Does data exist, is it enough, can you get it?
  - Engagement: How much time do they and you have for the project? How much time can you spend in their environment?

- **Intellectual**
  - Problem: Is there a vis research question lurking?
  - Need: Is there a real need or are existing approaches good enough?
  - Task: Are you addressing a real task? How long will need persist? How many people care?

- **Interpersonal**
  - What is your rapport with clients?

Pitfalls

- 32 pitfalls to design study projects listed, organized by framework phase
  - Examples
    - No real data available
    - No need for vis, problem can be automated
    - Nonrapid prototyping
    - Premature and insufficient deployment
Design Project

- Read description on website
- Form your team
  - Help with pairing now

- Examples

Upcoming

- Storytelling
  - Reading:

- Multivariate Data & Table/Graph Design
  - Reading:
    - Munzner chapter 2

- (start reading Few book chapters 5-12)
References

- Spence & CMS texts
- All referred to papers