User Tasks & Analysis

CS 7450 - Information Visualization
October 3, 2016
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
• Understand the components of a successful design study
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**

---

<table>
<thead>
<tr>
<th>Name</th>
<th>Rating</th>
<th>Country</th>
<th>Category</th>
<th>Price</th>
<th>ABV</th>
<th>Age</th>
<th>Brand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyrconnell 10 Year Old Single Malt</td>
<td>100</td>
<td>Ireland</td>
<td>Single Malt</td>
<td>72</td>
<td>46</td>
<td>10</td>
<td>Tyrconnell</td>
</tr>
<tr>
<td>Dalmore 18 Year Old Single Highland</td>
<td>100</td>
<td>Scotland</td>
<td>Highlands</td>
<td>165</td>
<td>43</td>
<td>18</td>
<td>Dalmore</td>
</tr>
<tr>
<td>Powers 12 Year Old Irish Whiskey</td>
<td>99</td>
<td>Ireland</td>
<td>Blended</td>
<td>35</td>
<td>40</td>
<td>12</td>
<td>Powers</td>
</tr>
<tr>
<td>Suntory Yamazaki 18 Year Old Sherry</td>
<td>99</td>
<td>Japan</td>
<td>Single Malt</td>
<td>120</td>
<td>43</td>
<td>18</td>
<td>Suntory</td>
</tr>
<tr>
<td>Glenmorangie 10 Year Old Single Malts</td>
<td>99</td>
<td>Scotland</td>
<td>Highlands</td>
<td>42</td>
<td>40</td>
<td>10</td>
<td>Glenmorangie</td>
</tr>
<tr>
<td>Glenmorangie Single Malt Scotch</td>
<td>99</td>
<td>Scotland</td>
<td>Highlands</td>
<td>80</td>
<td>46</td>
<td>10</td>
<td>Glenmorangie</td>
</tr>
<tr>
<td>Bunnahabhain 18 Year Old Single Malt</td>
<td>99</td>
<td>Islay</td>
<td>Highlands</td>
<td>92</td>
<td>46.8</td>
<td>18</td>
<td>Bunnahabhain</td>
</tr>
<tr>
<td>Laphroaig 18 Year Old Single Malt Scotch</td>
<td>99</td>
<td>Islay</td>
<td>Highlands</td>
<td>107</td>
<td>48</td>
<td>18</td>
<td>Laphroaig</td>
</tr>
<tr>
<td>Cardhu 12 Year Old Single Malt Scotch</td>
<td>99</td>
<td>Highlands</td>
<td>Highlands</td>
<td>45</td>
<td>40</td>
<td>12</td>
<td>Cardhu</td>
</tr>
<tr>
<td>Aberlour 18 Year Old Single Malt Scotch</td>
<td>99</td>
<td>Speyside</td>
<td>Speyside</td>
<td>100</td>
<td>43</td>
<td>18</td>
<td>Aberlour</td>
</tr>
<tr>
<td>Balvenie 14 Year Old Single Malt Scotch</td>
<td>99</td>
<td>Speyside</td>
<td>Speyside</td>
<td>60</td>
<td>43</td>
<td>14</td>
<td>Balvenie</td>
</tr>
<tr>
<td>Caol Ila Single Malt Scotch Distillers</td>
<td>99</td>
<td>Islay</td>
<td>Islay</td>
<td>70</td>
<td>43</td>
<td>12</td>
<td>Caol Ila</td>
</tr>
<tr>
<td>Kingdom 17 Year Old Scotch</td>
<td>99</td>
<td>Scotland</td>
<td>Blended</td>
<td>50</td>
<td>40</td>
<td>17</td>
<td>Kingdom</td>
</tr>
<tr>
<td>Balvenie 12 Year Old Doublewood</td>
<td>99</td>
<td>Scotland</td>
<td>Speyside</td>
<td>45</td>
<td>40</td>
<td>12</td>
<td>Balvenie</td>
</tr>
<tr>
<td>Glen Garnoch Founders Reserve Scotch</td>
<td>99</td>
<td>Highlands</td>
<td>Highlands</td>
<td>45</td>
<td>48</td>
<td>12</td>
<td>Glen Garnoch</td>
</tr>
<tr>
<td>Bowmore 15 Year Old Single Malt Scotch</td>
<td>99</td>
<td>Islay</td>
<td>Highlands</td>
<td>70</td>
<td>43</td>
<td>12</td>
<td>Bowmore</td>
</tr>
<tr>
<td>Rebel Yell Kentucky Straight Bourbon</td>
<td>99</td>
<td>USA</td>
<td>Bourbon</td>
<td>11</td>
<td>50</td>
<td>14</td>
<td>Rebel Yell</td>
</tr>
<tr>
<td>Pappy Van Winkle 15 Year Old Family</td>
<td>99</td>
<td>USA</td>
<td>Bourbon</td>
<td>58</td>
<td>53.5</td>
<td>15</td>
<td>Pappy Van Winkle</td>
</tr>
<tr>
<td>Thomas H. Handy Kentucky Straight</td>
<td>99</td>
<td>USA</td>
<td>Rye</td>
<td>67</td>
<td>66.4</td>
<td>6</td>
<td>Thomas H. Handy</td>
</tr>
<tr>
<td>Ardbeg Uigeadail</td>
<td>99</td>
<td>Scotland</td>
<td>Islay</td>
<td>80</td>
<td>54.2</td>
<td>12</td>
<td>Ardbeg</td>
</tr>
<tr>
<td>Noah's Mill Bourbon</td>
<td>99</td>
<td>USA</td>
<td>Bourbon</td>
<td>60</td>
<td>57.15</td>
<td>12</td>
<td>Noah's Mill</td>
</tr>
<tr>
<td>Parker's Heritage Bourbon</td>
<td>99</td>
<td>USA</td>
<td>Bourbon</td>
<td>80</td>
<td>62</td>
<td>15</td>
<td>Parker's Heritage</td>
</tr>
<tr>
<td>Glenlivet 21 Year Old Single Malt Scotch</td>
<td>97</td>
<td>Scotland</td>
<td>Speyside</td>
<td>123</td>
<td>40</td>
<td>21</td>
<td>Glenlivet</td>
</tr>
<tr>
<td>Macallan 21 Year Old Fine Oak Scotch</td>
<td>96</td>
<td>Scotland</td>
<td>Speyside</td>
<td>220</td>
<td>43</td>
<td>21</td>
<td>Macallan</td>
</tr>
<tr>
<td>George T. Stagg Kentucky Straight</td>
<td>96</td>
<td>USA</td>
<td>Bourbon</td>
<td>70</td>
<td>45</td>
<td>14</td>
<td>George T. Stagg</td>
</tr>
<tr>
<td>Parker's Heritage Collection 10 Year Old</td>
<td>96</td>
<td>USA</td>
<td>Bourbon</td>
<td>80</td>
<td>63</td>
<td>10</td>
<td>Parker's Heritage</td>
</tr>
<tr>
<td>Rowan's Creek Bourbon</td>
<td>96</td>
<td>USA</td>
<td>Bourbon</td>
<td>50</td>
<td>50.05</td>
<td>12</td>
<td>Rowan's Creek</td>
</tr>
<tr>
<td>Woodford Reserve Master's Collection</td>
<td>96</td>
<td>USA</td>
<td>Bourbon</td>
<td>80</td>
<td>48.2</td>
<td>12</td>
<td>Woodford Reserve</td>
</tr>
<tr>
<td>Lagavulin 21 Scotch</td>
<td>96</td>
<td>Scotland</td>
<td>Islay</td>
<td>300</td>
<td>56.5</td>
<td>21</td>
<td>Lagavulin</td>
</tr>
<tr>
<td>Highland Park 30 Scotch</td>
<td>96</td>
<td>Scotland</td>
<td>Islands</td>
<td>355</td>
<td>48.1</td>
<td>30</td>
<td>Highland Park</td>
</tr>
<tr>
<td>King Car Single Malt Whisky</td>
<td>96</td>
<td>Taiwan</td>
<td>Single Malt</td>
<td>84</td>
<td>46</td>
<td>0</td>
<td>Kavalan</td>
</tr>
<tr>
<td>Rye Dog Whiskey</td>
<td>96</td>
<td>USA</td>
<td>Rye</td>
<td>65</td>
<td>50</td>
<td>0</td>
<td>Delaware Phoenix</td>
</tr>
<tr>
<td>Thirteen Colony Southern Corn Whisky</td>
<td>96</td>
<td>USA</td>
<td>Corn</td>
<td>30</td>
<td>47.5</td>
<td>0</td>
<td>Thirteen Colony</td>
</tr>
<tr>
<td>Glenfiddich 12 Year Old Single Malt Scotch</td>
<td>96</td>
<td>Scotland</td>
<td>Speyside</td>
<td>43</td>
<td>40</td>
<td>12</td>
<td>Glenfiddich</td>
</tr>
<tr>
<td>Oban 15 Year Old Single Malt Scotch</td>
<td>96</td>
<td>Highlands</td>
<td>Highlands</td>
<td>89</td>
<td>43</td>
<td>15</td>
<td>Oban</td>
</tr>
<tr>
<td>Old Pulteney 30 Year Old Single Malt Scotch</td>
<td>96</td>
<td>Scotland</td>
<td>Highlands</td>
<td>400</td>
<td>44</td>
<td>30</td>
<td>Old Pulteney</td>
</tr>
</tbody>
</table>

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
  - ...
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

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
Organizational operators

VS. (filter, cluster, category)

Transformation ops (avg, count, ...)

1-1 \( n \times n \)

Find extremum

\[ \text{extremum} \]

Simple comparison

- What is the car with the highest MPG?
- What is the MP3 of Adele?
- How long is Gone with the Wind?
- Given cases(s), what are some attributes?

Distribution

- Characterization of distribution of an attribute over set of cases
- What is the distribution of carbs in cereals?
- What is the age distribution of shoppers?

Clustering

- Find clusters of attribute values I'm a set of cases
- Are there groups of cereals w/ similar fat, sugar?
- Is there a cluster of typical film lengths?

Save

Please do not erase

Filter

- What Kellogg's cereals have 3g fiber?
- What comedies have won awards?
- What films have the release date?
- What is Robin Williams film with release date?

 aggregate

Cluster

Categorize

- What is the highest award
- What directory has won the most awards?
- What film?
- What is the Robin Williams film with release date?
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 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.”

Valiati et al
BELIV ’06
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
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

*We saw this earlier*

Yi et al
*TVCG '07*
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. Filter out data to focus on relevant items. Sort items to expose patterns. Derive values or models from source data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>View Manipulation</td>
<td>Select items to highlight, filter, or manipulate them. Navigate to examine high-level patterns and low-level detail. Coordinate views for linked, multidimensional exploration. Organize multiple windows and workspaces.</td>
</tr>
<tr>
<td>Process and Provenance</td>
<td>Record analysis histories for revisitation, review, and sharing. Annotate patterns to document findings. Share views and annotations to enable collaboration. Guide users through analysis tasks or stories.</td>
</tr>
</tbody>
</table>
A Science of Interaction

Blend of interaction and tasks

Pike, Stasko, Chang, O'Connell
Information Visualization '08
Abstract Tasks

Framework/Typology of abstract visualization tasks

Brehmer & Munzner

TVCG (InfoVis) '13

Chapter 3
Why?

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

High to low level
How?

- encode
- select
- navigate
- arrange
- change
- filter
- aggregate
- annotate
- import
- derive
- record
Targets

What are the types of targets?
Task Cube

Perspective
- why: objectives
- how: actions

Abstraction
- concrete
- abstract

Composition
- high-level
- low-level

Rind et al
*Information Visualization '15*
Visual Analytic Activity

Sedig, Parsons, Babanski

JMPT'12
Another Question?

- Are the visualizations helping with exploratory analysis enough?
- Are they attempting to accomplish the right goals?
Status Quo Limitations

• Current Information Visualization systems inadequately support decision making:
  – Limited Affordances
  – Predetermined Representations
  – Decline of Determinism in Decision-Making

• “Representational primacy” versus “Analytic primacy”

Amar & Stasko
TVCG ’05
Goal: High-Level Tasks

- Complex decision-making, especially under uncertainty
- Learning a domain
- Identifying the nature of trends
- Predicting the future
- ...

Fall 2016  CS 7450
Analytic Gaps

- Analytic gaps – “obstacles faced by visualizations in facilitating higher-level analytic tasks, such as decision making and learning.”
  - Worldview Gap
  - Rationale Gap
Knowledge Precepts

• For narrowing these gaps
  – Worldview-Based Precepts
    (“Did we show the right thing to the user?”)
    Determine Domain Parameters
    Expose Multivariate Explanation
    Facilitate Hypothesis Testing
  – Rationale-Based Precepts
    (“Will the user believe what they see?”)
    Expose Uncertainty
    Concretize Relationships
    Expose Cause and Effect
Put Them Together

- Combine the ideas:
  - Use computational, statistical analysis more
  - Cater to the user’s analytic reasoning needs
- And put together with infovis

- Leads to...
Visual Analytics

• “The science of analytical reasoning facilitated by interactive visual interfaces”

• Combines
  – Data analysis
  – Infovis
  – Analytical reasoning

• Grew from view that infovis was neglecting these other aspects
  – True?

Thomas & Cook
_Illuminating the Path_
Visual Analytics

• Grew from stimulus in the homeland security area
  – Need for better data analysis methods
  – Really big data

• Topic for later in term...
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

![Diagram showing problem suitability based on task clarity and information location. The diagram is divided into sections for 'not enough data', 'design study methodology suitable', and 'algorithm automation possible'.]
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

- Examples
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
- Understand the components of a successful design study
Reading

- Brehmer & Munzner '13
HW 4

- Questions?

- Due next Weds, 12th
  - If you haven’t started yet...
Upcoming

- Poster Session
- Fall Break
- Storytelling (don't miss it)
References

• Spence & CMS texts
• All referred to papers