Deconstructing Visualizations

Maneesh Agrawala

CS 448B: Visualization
Fall 2021
Reading Response Questions/Thoughts

For the data explainer project, do we have to find one dataset and create our three visualizations off of that one dataset, or is it alright if we find a high-level topic that we are interested in, and create three visualizations within that topic but using separate datasets?

When using social network analysis, how do you validate your findings and/or determine if your findings are statistically significant? Is there an analogous "p value" standard for graph analysis? [Do you use qualitative or quantitative measures of validity?]

Why do we go for complex graphs if we can break down a complex concept into multiple, easily digestible graphs [e.g. broken down into strongly connected components]? Wouldn’t this also help with making the structure more intuitive?

Deconstructing Visualizations
Pixels are poor representation
Hard for machines to retrieve data
Pixels are poor representation
Hard for machines to retrieve data
Hard for people to manipulate

Goal: Reconstruct higher-level representation of charts and graphs that lets machines and people redesign, reuse and revitalize them
What is a good representation?

<table>
<thead>
<tr>
<th>Year</th>
<th>Exports</th>
<th>Imports</th>
</tr>
</thead>
<tbody>
<tr>
<td>1700</td>
<td>170,000</td>
<td>300,000</td>
</tr>
<tr>
<td>1701</td>
<td>171,000</td>
<td>302,000</td>
</tr>
<tr>
<td>1702</td>
<td>176,000</td>
<td>303,000</td>
</tr>
<tr>
<td>1703</td>
<td>180,000</td>
<td>312,000</td>
</tr>
<tr>
<td>1704</td>
<td>187,000</td>
<td>319,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
2005 NIH Research Budget per Death

<table>
<thead>
<tr>
<th>Disease</th>
<th>Budget</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>70.0%</td>
</tr>
<tr>
<td>Alzheimer's</td>
<td>5.0%</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>1.1%</td>
</tr>
<tr>
<td>Diabetes</td>
<td>4.8%</td>
</tr>
<tr>
<td>Hepatitis B</td>
<td>4.1%</td>
</tr>
<tr>
<td>Hepatitis C</td>
<td>3.8%</td>
</tr>
<tr>
<td>Parkinson'</td>
<td>6.0%</td>
</tr>
<tr>
<td>Prostate</td>
<td>5.2%</td>
</tr>
</tbody>
</table>

Data marks:
- Budget → angle (Q)
- Disease → color (N)

AIDS
- Parkinson'
- Prostate
- Alzheimer's
- Diabetes
- Hepatitis C
- Hepatitis B
- Cardiovascular...
Approach

**Classification:** Determine chart type
**Mark extraction:** Retrieve graphical marks
**Data extraction:** Retrieve underlying data table
Training the Classifier
Training the Classifier

Bar Charts

Pie Charts

Scatter Plots

Classifying an Input Image

Asset allocation by type

Platinum
Silver
Gold
Bonds
Cash
Stocks
Classifying an Input Image
Classifying an Input Image

Asset allocation by type

[Graph showing asset allocation]

Classifying an Input Image
Classifying an Input Image

Corpus: 667 charts, 5 chart types [Prasad 2007]  
<table>
<thead>
<tr>
<th></th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Prasad 2007] Multi-class SVM</td>
<td>84%</td>
</tr>
<tr>
<td>ReVision: Multi-class SVM</td>
<td>88%</td>
</tr>
<tr>
<td>ReVision: Binary SVM (yes/no for each chart type)</td>
<td>96%</td>
</tr>
</tbody>
</table>
Our Corpus

Over 2500 labeled images and 10 chart types

http://vis.berkeley.edu/papers/revision

ReVision binary SVMs give 96% classification accuracy

Mark and Data Extraction
Assumptions
Bar charts and pie charts only
No shading or texture, 3D, stacked bars, or exploded pies

Bar Charts
marks: lines

<table>
<thead>
<tr>
<th>y-value</th>
<th>x-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>A</td>
</tr>
<tr>
<td>25</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>75</td>
<td>D</td>
</tr>
</tbody>
</table>
Bar Charts

- Find Foreground Rectangles
- Identify Orientation and Baseline
- Recover Bar Values
- Associate Labels with Bars

Extract Marks

Extract Data

marks: lines

<table>
<thead>
<tr>
<th>y-value</th>
<th>x-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>A</td>
</tr>
<tr>
<td>35</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
</tr>
<tr>
<td>75</td>
<td>D</td>
</tr>
</tbody>
</table>

Scale: 2 pixels/unit

marks: lines

Pie Charts

- Fit Ellipse Using RNASAC
- Unroll Pie and Find Transitions
- Compute Area Percentages
- Associate Labels with Areas

Extract Marks

Extract Data

marks: areas

<table>
<thead>
<tr>
<th>percentage</th>
<th>category</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.3</td>
<td>A</td>
</tr>
<tr>
<td>22.4</td>
<td>B</td>
</tr>
<tr>
<td>10.8</td>
<td>C</td>
</tr>
<tr>
<td>5.6</td>
<td>D</td>
</tr>
<tr>
<td>5.6</td>
<td>E</td>
</tr>
<tr>
<td>33.3</td>
<td>F</td>
</tr>
</tbody>
</table>

Scale: 50 pixels/percent
Extraction Results

<table>
<thead>
<tr>
<th>Number of Charts</th>
<th>Bar</th>
<th>Pie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total charts</td>
<td>52</td>
<td>53</td>
</tr>
<tr>
<td>Mark extractions</td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>Data extractions</td>
<td>29</td>
<td>21</td>
</tr>
</tbody>
</table>

- Bar Charts: 79% 56%
- Pie Charts: 62% 40%

Data Extraction Error

- Bar Charts: 7.7%
- Pie Charts: 4.6%

Average chart size: 342 x 452 pixels [Prasad 2007]
Redesign

Original

Redesign

AIDS
Parkinson
Prostate
Alzheimer's
Diabetes
Hepatitis C
Hepatitis B
Cardiovascular...

2005 NIH Research Budget per Death

- Cardiovascular (CVD, Heart & Stroke)
- Diabetes
- Hepatitis C
- Hepatitis B
- Prostate
- Alzheimer's
- Parkinson's

© Copyright PAR Foundation 2006
Limitations

Additional Chart Types

Handling Legends
Announcements

Final project

Data analysis/explainer or conduct research
- **Data analysis**: Analyze dataset in depth & make a visual explainer
- **Research**: Pose problem, Implement creative solution

Deliverables
- **Data analysis/explainer**: Article with multiple different interactive visualizations
- **Research**: Implementation of solution and web-based demo if possible
- **Short video (2 min)** demoing and explaining the project

Schedule
- Project proposal: **Wed 11/3**
- Design Review and Feedback: **10th week of quarter**
- Final code and video: **Fri 12/10 11:59pm**

Grading
- Groups of **up to 3 people**, graded individually
- Clearly report responsibilities of each member
Graphical Overlays

*Visual elements that are layered onto a chart to facilitate the perceptual and cognitive processes involved in chart reading*
Graphical overlay gallery

This gallery contains examples of graphical overlays, described in our paper. We have extracted marks and data from the charts using RevVis (for bars and pie charts) and Datamint (for line charts), but all of the overlays are generated in-browser. Try out some of the parameters, or click on an image thumbnail below the gallery to view some example overlays.

European Union budgets since 2000

Chart type: 
Chart: 00281
Overlay type: Reference structures

Parameters

Overlay
Underlay
Static
Interactive
Divisions
Line thickness:

Places regular gridlines at user defined intervals.

Demo
Reference Structures

- Help by breaking marks into regular segments and aid reading axis values

Highlights

- Draws viewers’ attention to specific marks
Redundant Encodings

Emphasize data values or trends

Summary Statistics

Enables comparison with statistics based on the data
Most overlays only require access to marks

Reference structures (marks)
Highlights (marks)
Redundant encodings (marks and data)
Summary statistics (marks)
Annotations (marks)
Interactive Documents

How can we facilitate reading text and charts together?

Goal: Extract references between text and chart
Skepticism for capitalism is lowest in Brazil (22%), China (19%), Germany (29%) (although East Germans are less supportive than West Germans) and the U.S. (24%). Skepticism for free markets is highest in Mexico (60%) and Japan (60%).
Skepticism for capitalism is lowest in Brazil (22%), China (19%), Germany (29%) (although East Germans are less supportive than West Germans) and the U.S. (24%). Skepticism for free markets is highest in Mexico (60%) and Japan (60%).

Example 1: Pew Research

Are People Better Off in a Free Market Economy?

<table>
<thead>
<tr>
<th>Country</th>
<th>Disagree</th>
<th>Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazil</td>
<td>72</td>
<td>28</td>
</tr>
<tr>
<td>China</td>
<td>74</td>
<td>26</td>
</tr>
<tr>
<td>Germany</td>
<td>29</td>
<td>71</td>
</tr>
<tr>
<td>U.S.</td>
<td>24</td>
<td>76</td>
</tr>
<tr>
<td>Lebanon</td>
<td>34</td>
<td>66</td>
</tr>
<tr>
<td>India</td>
<td>35</td>
<td>65</td>
</tr>
<tr>
<td>Britain</td>
<td>37</td>
<td>63</td>
</tr>
<tr>
<td>France</td>
<td>43</td>
<td>57</td>
</tr>
<tr>
<td>Turkey</td>
<td>21</td>
<td>79</td>
</tr>
<tr>
<td>Poland</td>
<td>37</td>
<td>63</td>
</tr>
<tr>
<td>Italy</td>
<td>30</td>
<td>70</td>
</tr>
<tr>
<td>Egypt</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td>Czech Rep.</td>
<td>46</td>
<td>54</td>
</tr>
<tr>
<td>Pakistan</td>
<td>36</td>
<td>64</td>
</tr>
<tr>
<td>Russia</td>
<td>40</td>
<td>60</td>
</tr>
<tr>
<td>Spain</td>
<td>52</td>
<td>48</td>
</tr>
<tr>
<td>Greece</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Jordan</td>
<td>54</td>
<td>46</td>
</tr>
<tr>
<td>Tunisia</td>
<td>37</td>
<td>63</td>
</tr>
<tr>
<td>Japan</td>
<td>60</td>
<td>40</td>
</tr>
<tr>
<td>Mexico</td>
<td>60</td>
<td>40</td>
</tr>
</tbody>
</table>

Example 2: Economist

Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions. Europeans and Latin Americans tend to have similar attitudes to the rich; the Anglo-Saxon world is a bit more forgiving.
Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions. Europeans and Latin Americans tend to have similar attitudes to the rich; the Anglo-Saxon world is a bit more forgiving.
Demo

Evaluation

Avg. $F_1$ distance: expert specified references vs. crowd specified references
Deconstructing D3 Charts

D3 Code

```javascript
var bars = svg.selectAll('rect')
  .data(items)
  .enter()
  .append('rect');

bars.attr('x', function(d, i) {
  return i * 35;
}).
.attr('y', function(d) {
  return h - d.price * 38;
}).
.attr('height', function(d) {
  return d.price * 38;
}).
.attr('fill', function(d) {
  return d.type == 'fruit' ? 'green' : 'red';
}).
.attr('width', function(d) {
  return 20;
}).
.attr('stroke-width', 0);
```

D3 Chart

Our Deconstruction

Automatically convert D3 code into mapping based representation to enable redesign and style reuse

User Interface Software Technology (UIST) 2014.

Deconstructing D3 Charts

Narrowly defined unemployment rates: top 20 countries (2010)

<table>
<thead>
<tr>
<th>country</th>
<th>rate</th>
<th>deconID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Namibia</td>
<td>37.6</td>
<td>17</td>
</tr>
<tr>
<td>Macedonia, FYR</td>
<td>32.0</td>
<td>21</td>
</tr>
<tr>
<td>Armenia</td>
<td>28.6</td>
<td>25</td>
</tr>
<tr>
<td>Bosnia and Herzegovina</td>
<td>27.2</td>
<td>29</td>
</tr>
<tr>
<td>Lesotho</td>
<td>25.3</td>
<td>33</td>
</tr>
<tr>
<td>South Africa</td>
<td>24.7</td>
<td>37</td>
</tr>
<tr>
<td>Spain</td>
<td>20.1</td>
<td>41</td>
</tr>
<tr>
<td>Latvia</td>
<td>18.7</td>
<td>45</td>
</tr>
</tbody>
</table>

rate ➔ width
rate ➔ area
rate ➔ xPos
deconID ➔ yPos

User Interface Software Technology (UIST) 2014.
Deconstructing D3 Charts

Namibia 37.6 17
Macedonia, FYR 32.0 21
Armenia 28.6 25
Bosnia and Herzegovina 27.2 29
Lesotho 25.3 33
South Africa 24.7 37
Spain 20.1 41
Latvia 18.7 45
...

rate xPos
deconID yPos

Automatic Redesign

Can we automatically redesign charts to improve
Perceptual effectiveness?
Visual aesthetics?
Accessibility for vision impaired users?

Many specialized collections
Scientific: PLOS, JSTOR, ACM DL, ...
Web visualizations: D3, Processing, ...
News: New York Times, Pew research, ...

How can deconstruction aid search?
Search by chart type, data type, marks, data, ...
Similarity search with inexact matching
Query expansion

Takeaways

A chart is a collection of mappings between data and marks

We can reconstruct this representation from chart bitmaps

Such reconstruction enables redesign, reuse and revitalization