DECONSTRUCTING VISUALIZATIONS

CS 448B | Fall 2023

MANEESH AGRAWALA

Presidential Election Results

253
71
214

2020 Election Results

2020 US Presidential Election

Source: The New York Times
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
Pixels are a poor representation of charts and graphs
Cannot index, search, manipulate or interact with the data

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

Learning Objectives

1. How to deconstruct charts and graphs into an editable representation
2. How to use this representation to support interactive reading of visualization
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>
</tbody>
</table>

...
APPROACH

**Classification**: Determine chart type
**Mark extraction**: Retrieve graphical marks
**Data extraction**: Retrieve underlying data values
TRAINING

Bar Charts

Pie Charts

Scatter Plots

CLASSIFICATION

Asset allocation by type

- Platinum
- Silver
- Gold
- Bonds
- Sticks
- Cash
CLASSIFICATION

Class diagrams by type

[Prasad 2007]

Average Accuracy

<table>
<thead>
<tr>
<th>Corpus: 667 charts, 5 chart types [Prasad 2007]</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

ReVision binary SVMs give 96% classification accuracy

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

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

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>

Extract Marks
- Find Foreground Rectangles
- Identify Orientation and Baseline

Extract Data
- Recover Bar Values
- Associate Labels with Bars

PIE CHARTS

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>

Extract Marks
- Fit Ellipse Using RNASAC
- Unroll Pie and Find Transitions

Extract Data
- Compute Area Percentages
- Associate Labels with Areas
EXTRACTION RESULTS

DATA EXTRACTION ERROR

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

FINAL PROJECT

**Design Review Nov 27 and 29 (Signups released yesterday)**

**Data analysis/explainer**
- Analyze dataset in depth & make a visual explainer

**Deliverables**
- An article with multiple different interactive visualizations
- Short video (2 min) demoing and explaining the project

**Schedule**
- Project proposal: Mon 11/6
- Design Review and Feedback: 9th week of quarter, 11/27 and 11/29
- Final code and video: Sun 12/10 8pm

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

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 Rvizviz (for bars and pie charts) and dataplot (for line charts). Click on any image to download the gallery to view some example overlays.

- **Chart type:** line
- **Overlay type:** line
- **Regular gridlines:**
- **Lines emanating from marks**

**Parameters**
- **Overlay**
- **Underlay**
- **Static**
- **Interactive**

**Line thickness:**

Plots 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

Syrian refugees: how many are there and where are they?

Some contributions are made on a regional basis, but many donors prefer to contribute to efforts in a specific country. In line with the distribution of the refugee numbers, the funds are funnelled towards Jordan (28%), followed by Lebanon (26%), Turkey (19%) and Iraq (17%).
**Problem:** Diversity of writing styles

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%).

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.
EVALUATION

Avg. $F_1$ distance: expert specified references vs. crowd specified references
DECONSTRUCTING D3 CHARTS

Narrowly defined unemployment rates: top 20 countries (2013)

<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>
<tr>
<td>Estonia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lithuania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serbia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Georgia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yemen, Rep.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slovak Republic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dominican Republic</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tunisia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Albania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ireland</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jordan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greece</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

AUTOMATIC REDESIGN

Data Source

Style Target

Result
**DOCUMENT COLLECTIONS**

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*