Deconstructing Visualizations

Maneesh Agrawala

CS 448B: Visualization
Fall 2018

Last Time: Visual Explainers

Narrative Storytelling

narrative (n): An account of a series of events, facts, etc., given in order and with the establishing of connections between them

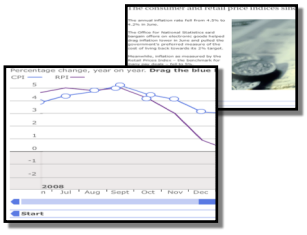
"… require[s] skills like those familiar to movie directors, beyond a technical expert's knowledge of computer engineering and science."

- Gershon & Page '01

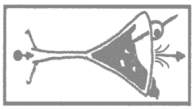
Genres + Interactivity + Messaging =

DESIGN SPACE

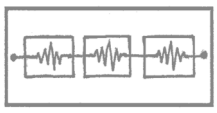
STORYTELLING	Author Driven	Reader Driven	ASK QUESTIONS
CLARITY	strong ordering	weak ordering	EXPLORE
SPEED	heavy messaging	light messaging	FIND
	limited interactivity	free interactivity	



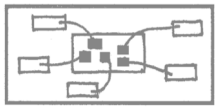
martini glass



interactive slideshow



drill-down story



Announcements

Final project

New visualization research or data analysis

- Pose problem, Implement creative solution
- Design studies/evaluations

Deliverables

- Implementation of solution
- 6-8 page paper in format of conference paper submission
- Project progress presentations

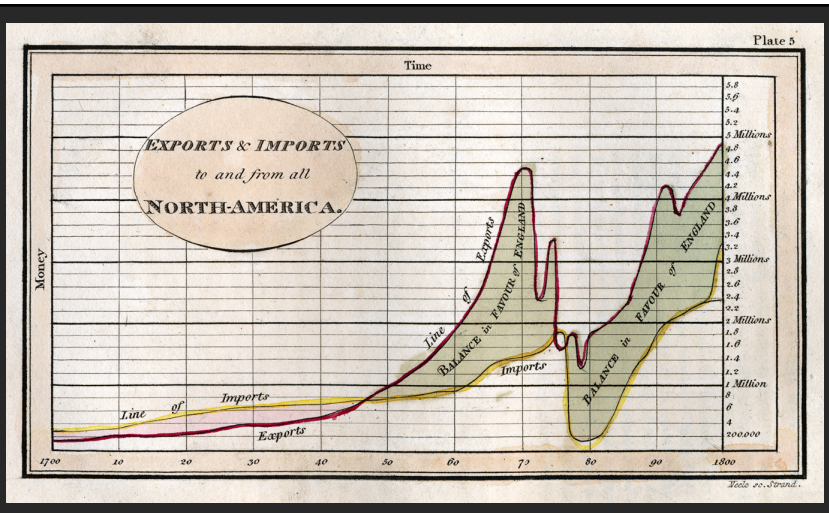
Schedule

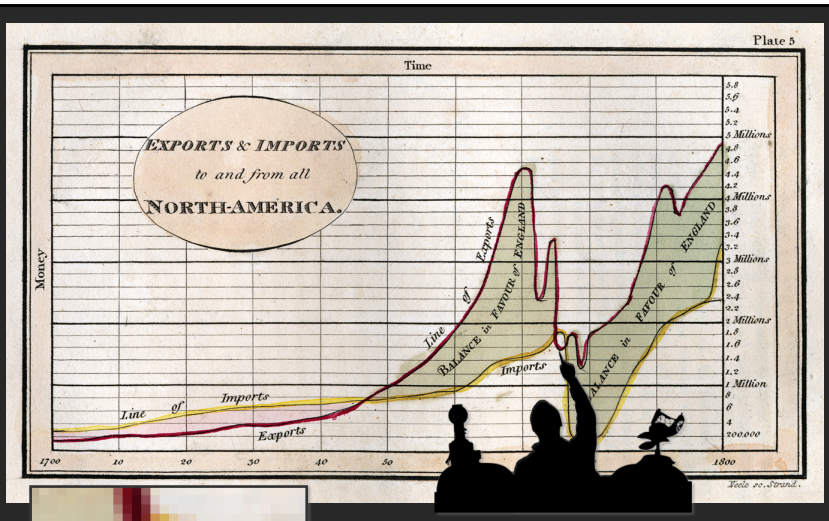
- Project proposal: Mon 11/5
- Project progress presentation: 11/12 and 11/14 in class (3-4 min)
- Final poster presentation: 12/5 Location: Lathrop 282
- Final paper: 12/9 11:59pm

Grading

- Groups of up to 3 people, graded individually
- Clearly report responsibilities of each member

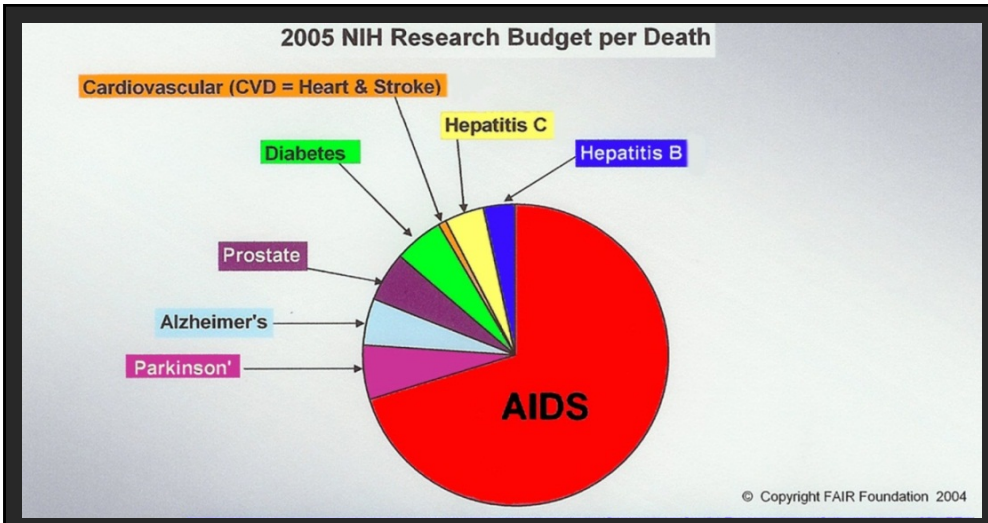
Deconstructing Visualizations

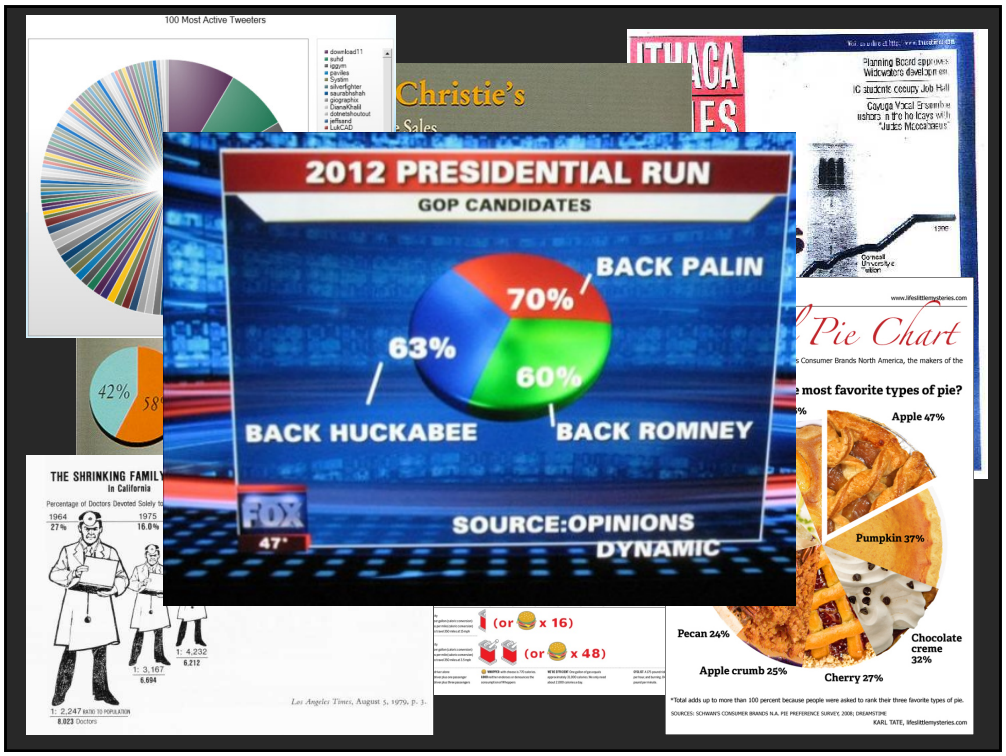


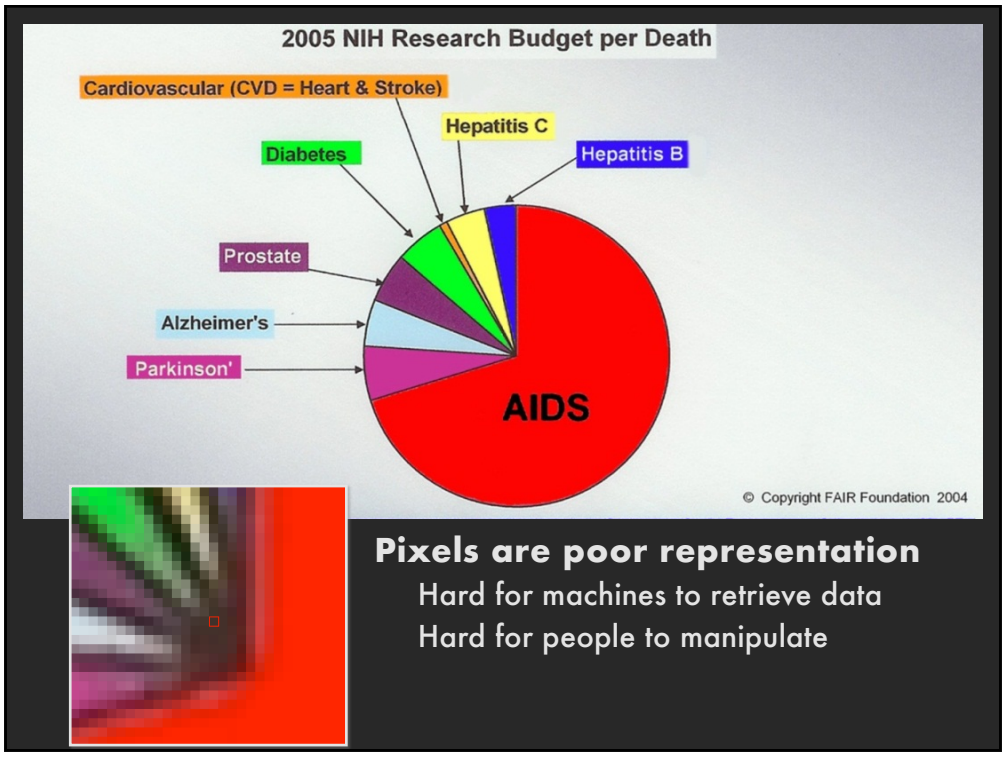


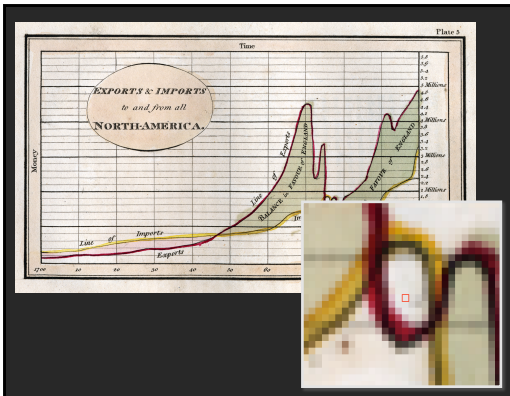
Pixels are poor representation

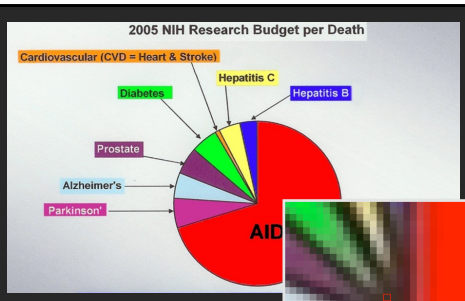
Hard for machines to retrieve data

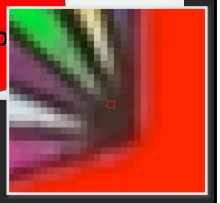



100 Most Active Tweeters
Christie's
2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47*
SOURCE:OPINIONS DYNAMIC
Pie Chart
most favorite types of pie?
Apple 47%
Pumpkin 37%
Pecan 24%
Chocolate creme 32%
Apple crumb 25%
Cherry 27%
THE SHRINKING FAMILY
in California
Las Angeles Times, August 5, 1979, p. 3.


2005 NIH Research Budget per Death
Cardiovascular (CVD = Heart & Stroke)
Hepatitis C
Diabetes
Hepatitis B
Prostate
Alzheimer's
Parkinson'
AIDS
© Copyright FAIR Foundation 2004
Pixels are poor representation
Hard for machines to retrieve data
Hard for people to manipulate



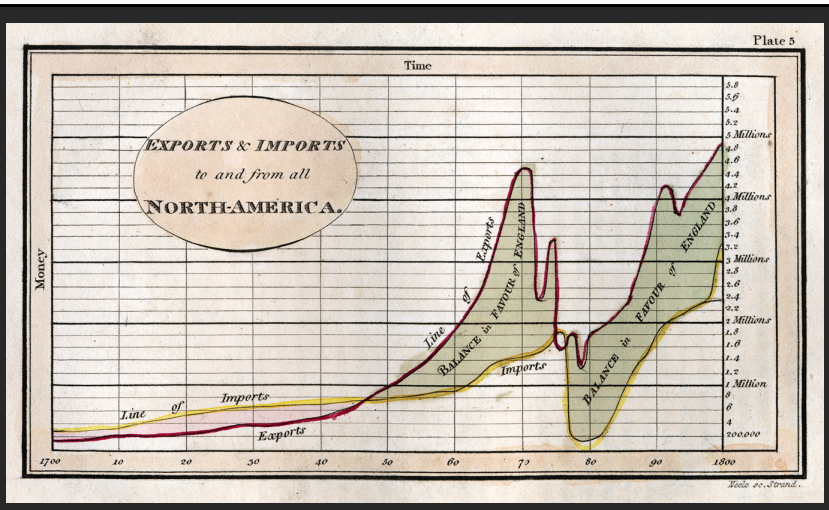




Pixels are a poor representation of charts and graphs
Cannot index, search, manipulate or interact with the data

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

What is a good representation?



Data

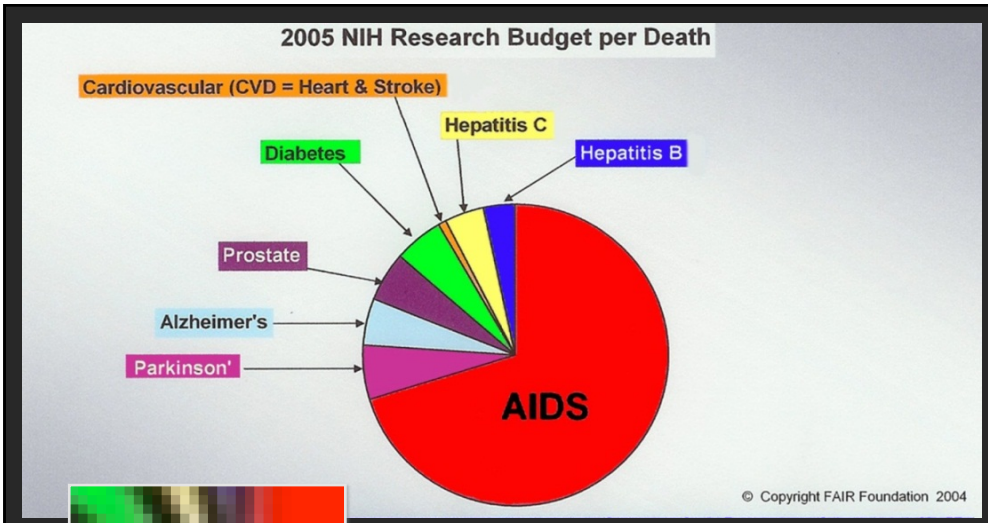
Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
1703	180,000	312,000
1704	187,000	319,000
…	…	…

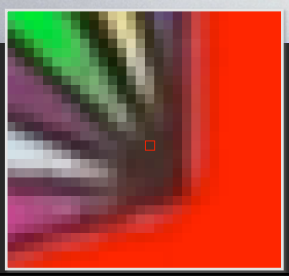
Marks

mark: lines

Mappings

Year → x-pos (Q)
Exports → y-pos (Q)
Imports → y-pos (Q)
Exports → color (N)
Imports → color (N)





Data

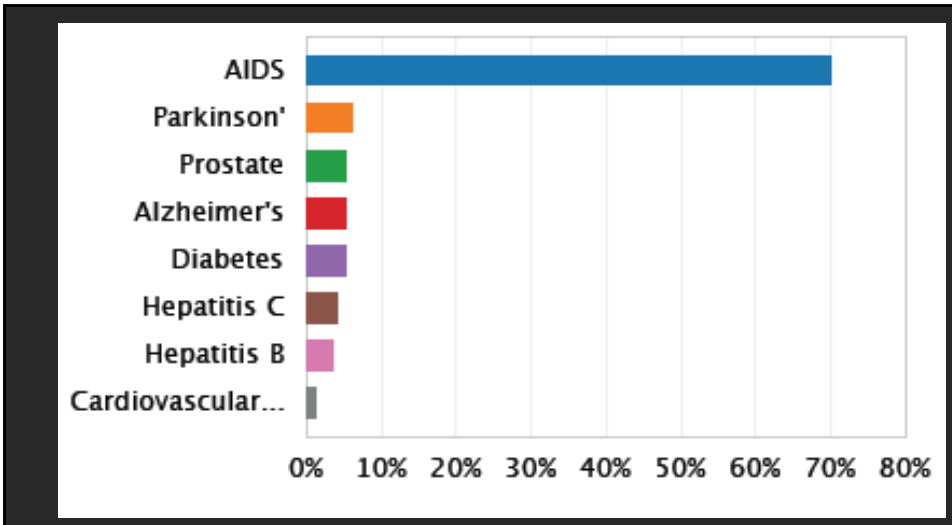
Disease	Budget
Aids	70.0%
Alzheimer's	5.0%
Cardiovascular	1.1%
Diabetes	4.8%
Hepatitus B	4.1%
Hepatitus C	3.8%
Parkinson'	6.0%
Prostate	5.2%

Marks

mark: areas

Mappings

Budget → angle (Q)
Disease → color (N)



Data

Disease	Budget
Aids	70.0%
Alzheimer's	5.0%
Cardiovascular	1.1%
Diabetes	4.8%
Hepatitus B	4.1%
Hepatitus C	3.8%
Parkinson'	6.0%
Prostate	5.2%

Marks Mappings

mark: lines
Budget → length (Q)
Disease → color (N)

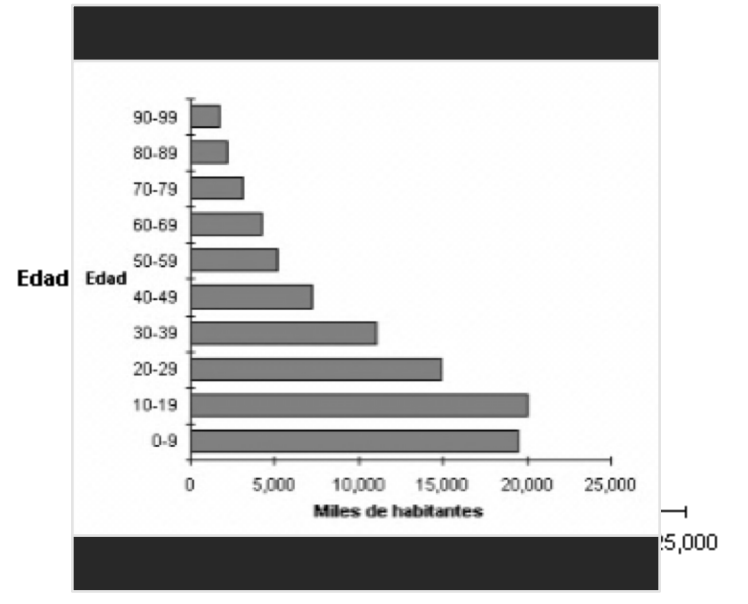
Approach

Classification: Determine chart type
Mark extraction: Retrieve graphical marks
Data extraction: Retrieve underlying data table

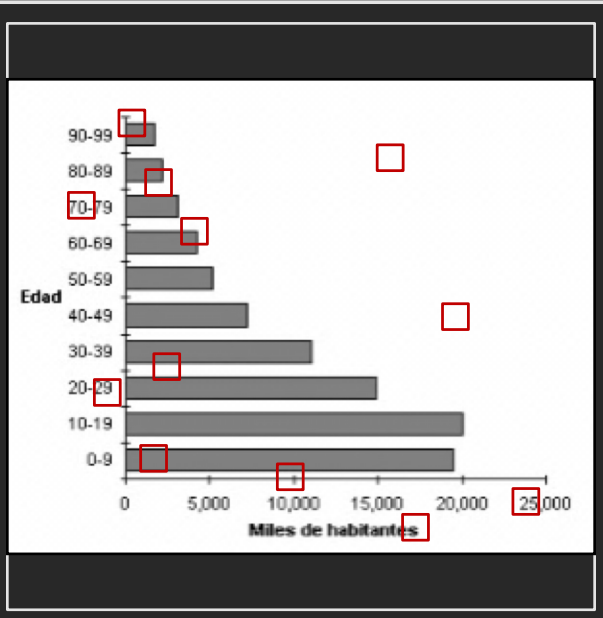
18

Classification

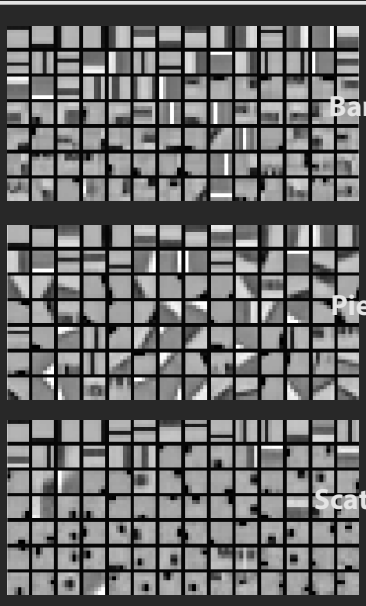
Training the Classifier



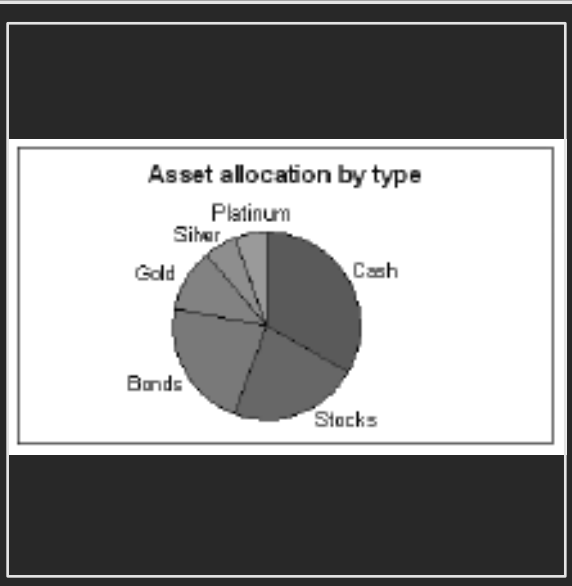
Training the Classifier



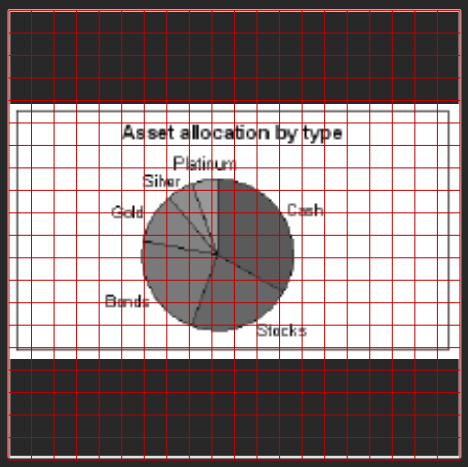
Training the Classifier



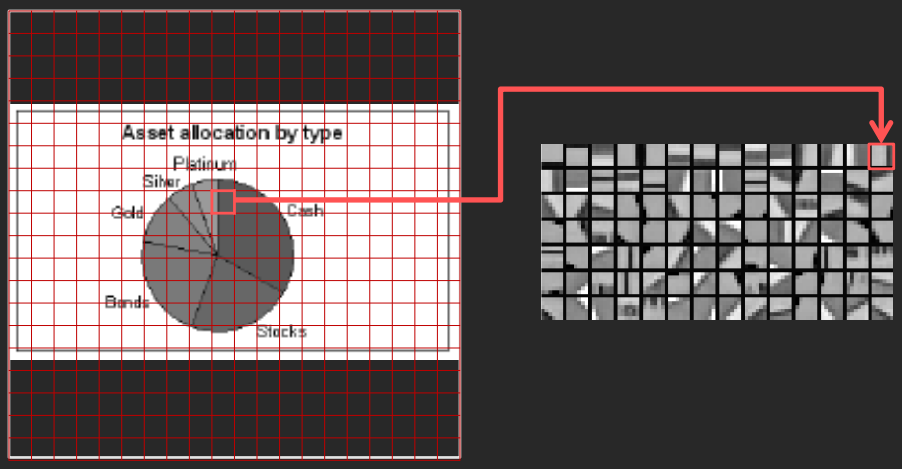
Classifying an Input Image



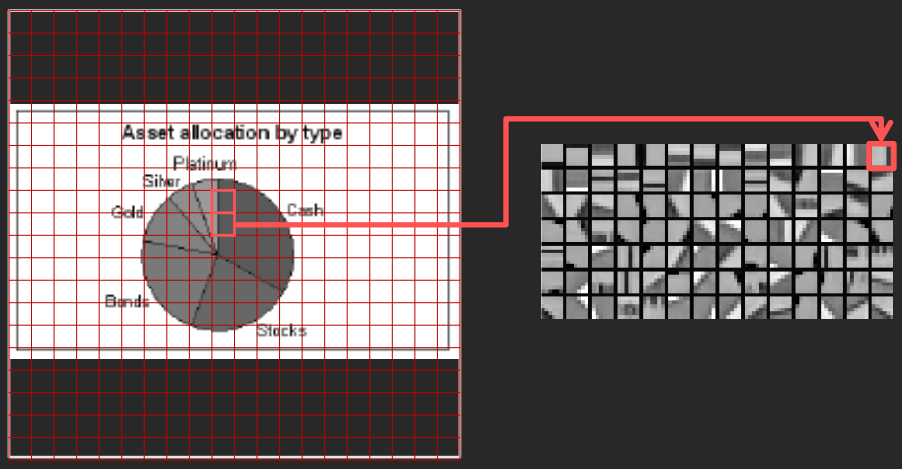
Classifying an Input Image



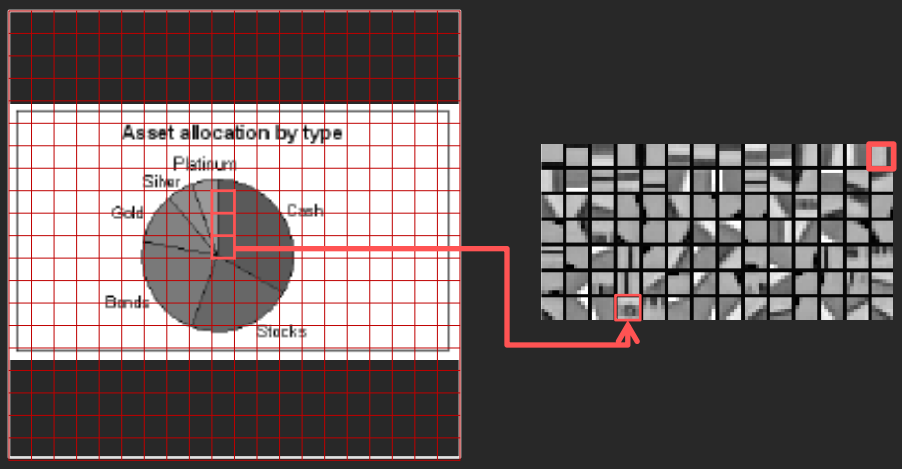
Classifying an Input Image



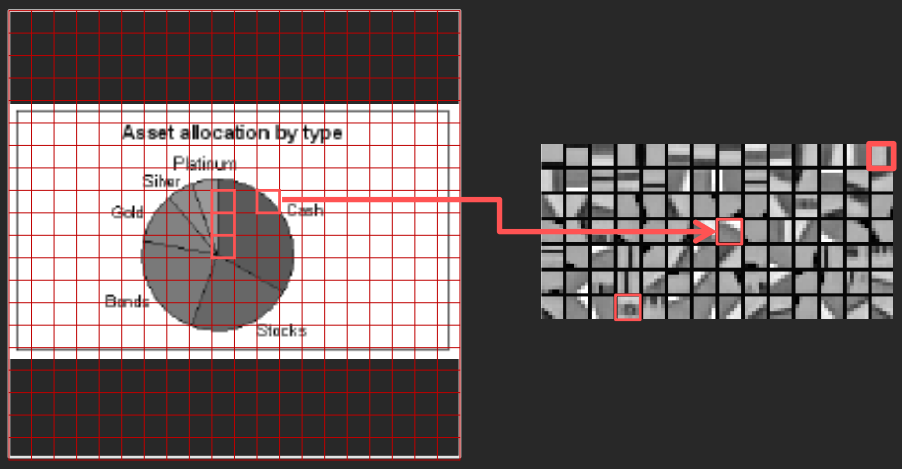
Classifying an Input Image



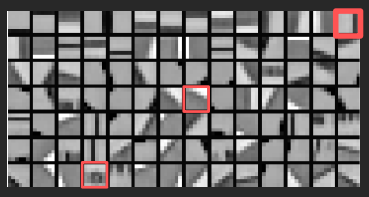
Classifying an Input Image



Classifying an Input Image



Classifying an Input Image

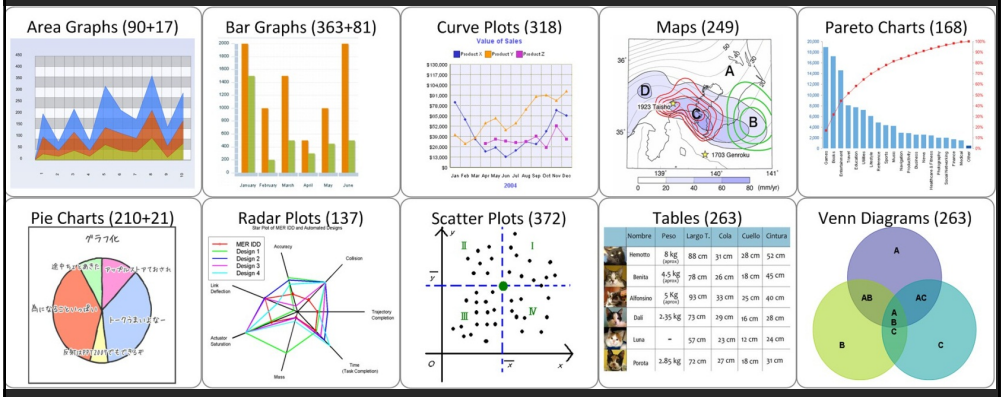


→ **SVM Classifier** → **Pie Chart**

Corpus: 667 charts, 5 chart types [Prasad 2007]	**Average Accuracy**
[Prasad 2007] Multi-class SVM	84%
ReVision: Multi-class SVM	88%
ReVision: Binary SVM (yes/no for each chart type)	96%

Our Corpus

Over 2500 labeled images and 10 chart types



ReVision binary SVMs give 96% classification accuracy

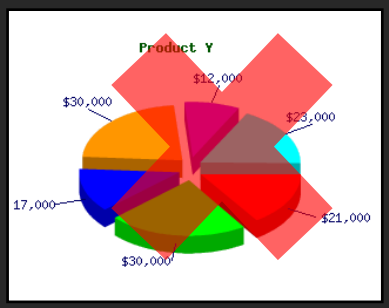
http://vis.berkeley.edu/papers/revision

Mark and Data Extraction

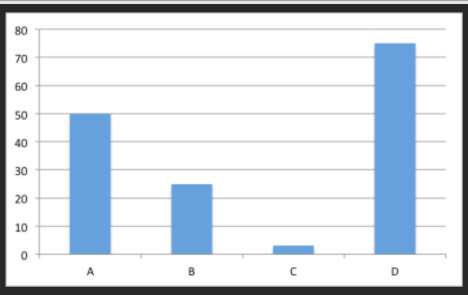
Assumptions

Bar charts and pie charts only

No shading or texture, 3D, stacked bars, or exploded pies



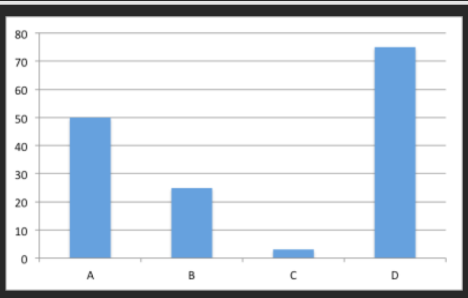
Bar Charts



marks: lines

y-value	x-value
50	A
25	B
4	C
75	D

Bar Charts



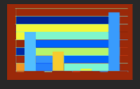
marks: lines

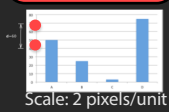
y-value	x-value
50	A
35	B
4	C
75	D

Extract Marks

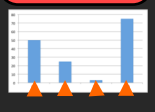
Extract Data

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

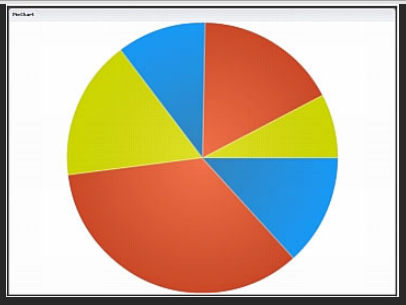




Scale: 2 pixels/unit



Pie Charts



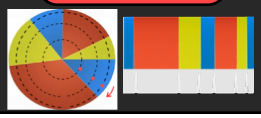
marks: areas

percentage	category
22.3	A
22.4	B
10.8	C
5.6	D
5.6	E
33.3	F

Extract Marks ↓

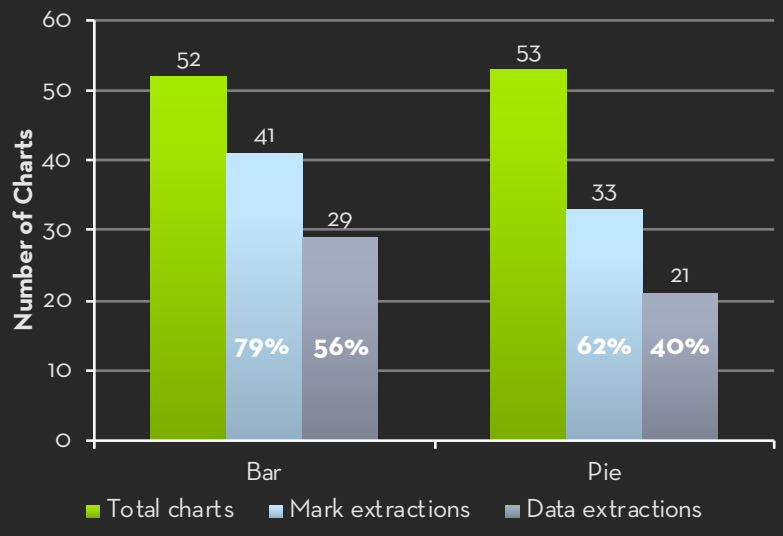
Extract Data ↑

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



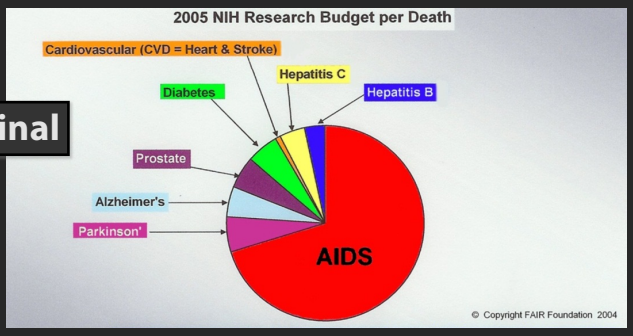
Scale: 50 pixels/percent

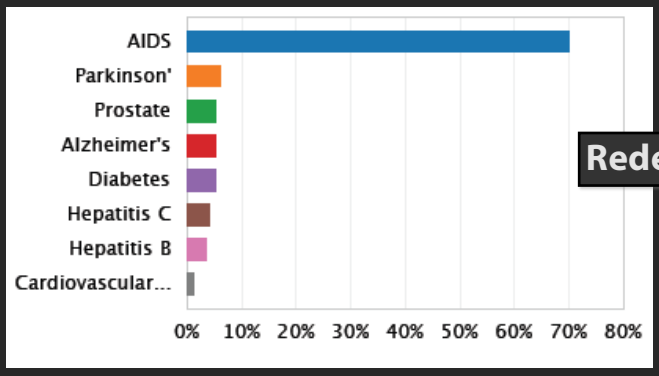
Extraction Results



Redesign

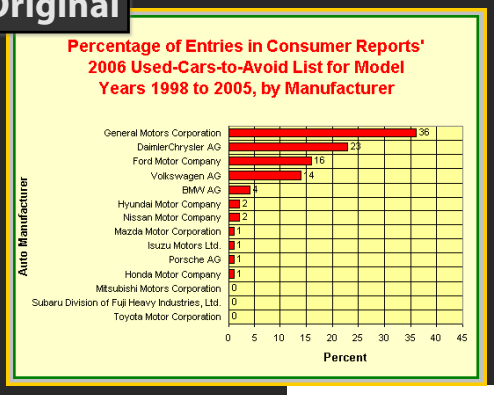
Original



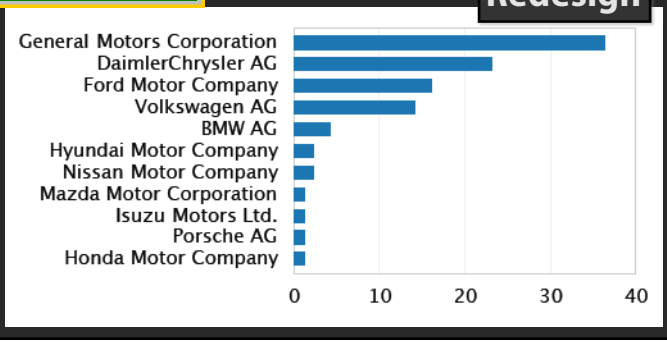


Redesign

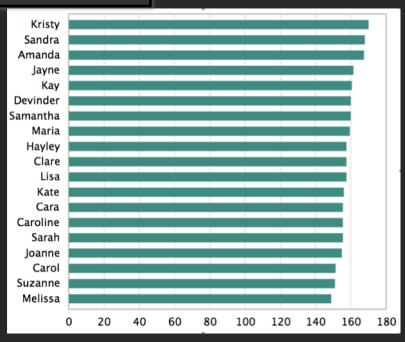
Original



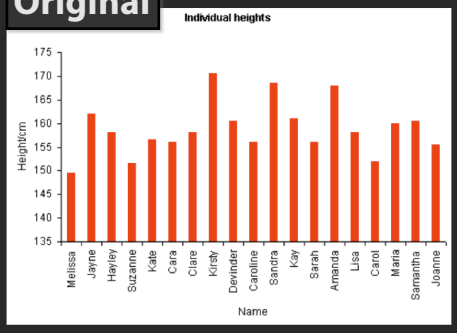
Redesign

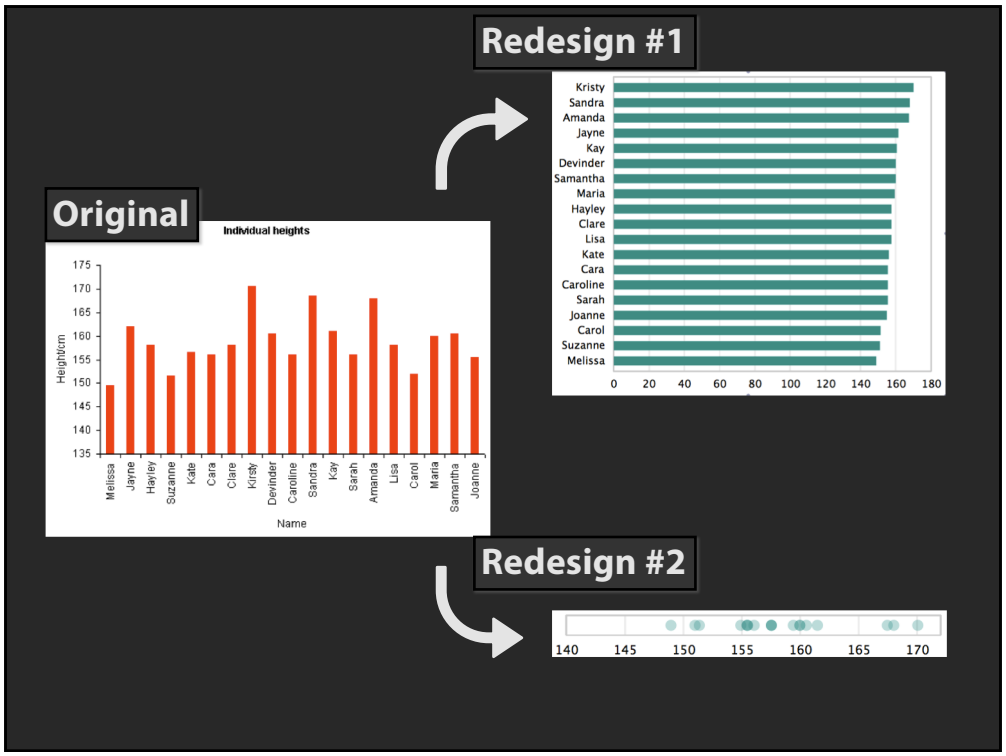


Redesign #1



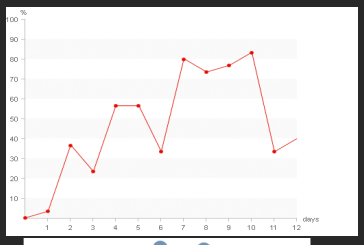
Original

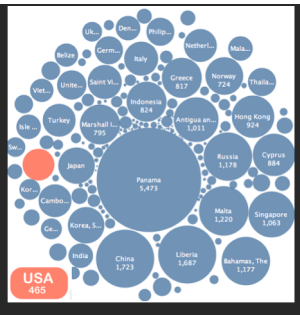




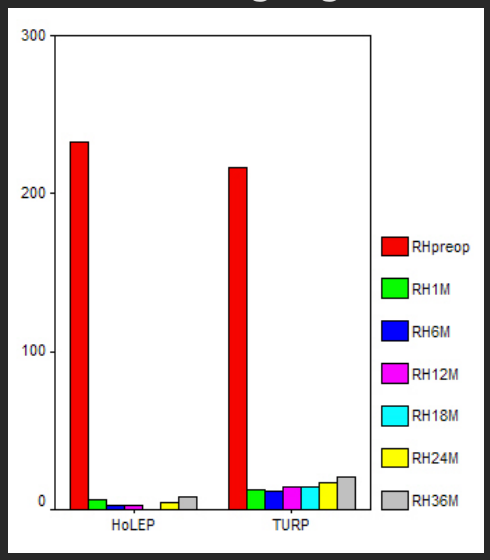
Limitations

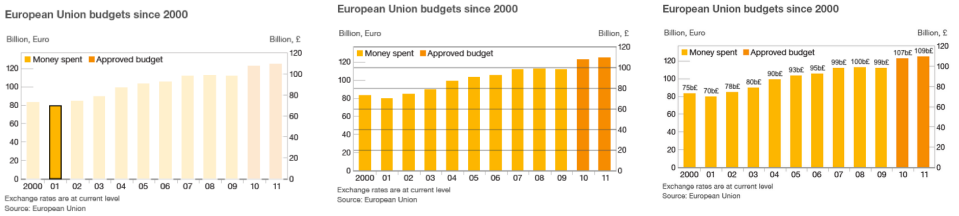
Additional Chart Types





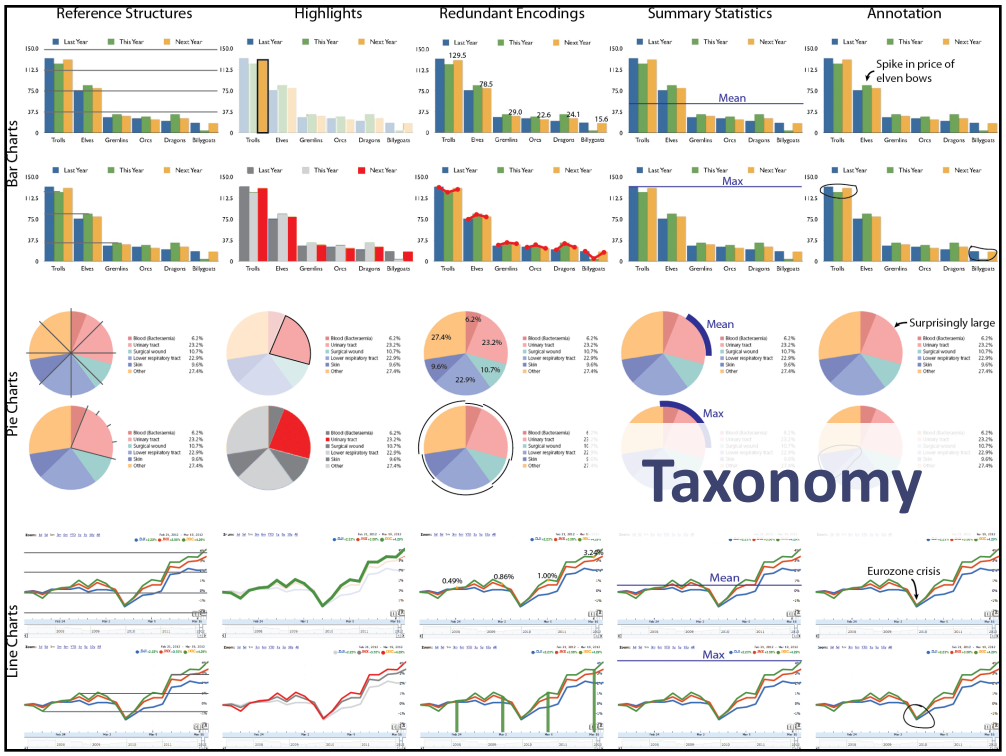
Handling Legends





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 ReVision (for bars and pie charts) and Datathief (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.

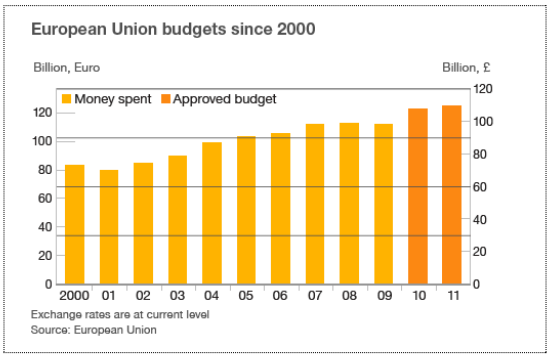


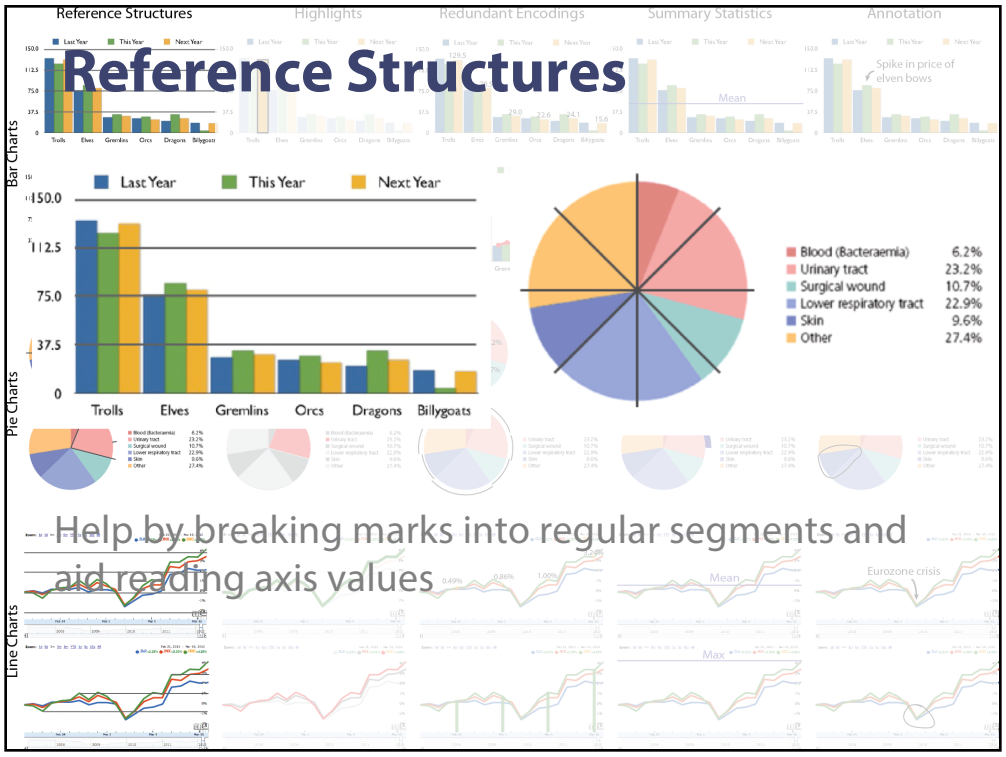
Chart type: Bar
Chart: 00281
Overlay type: Reference structures
Regular gridlines
Lines emanating from marks

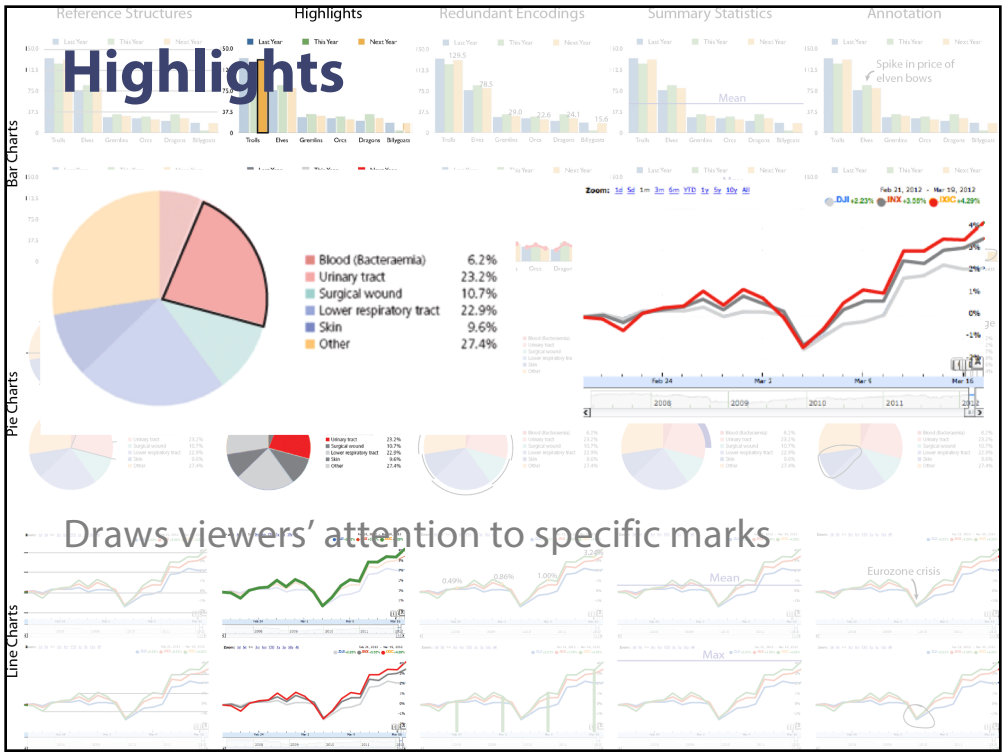
Parameters

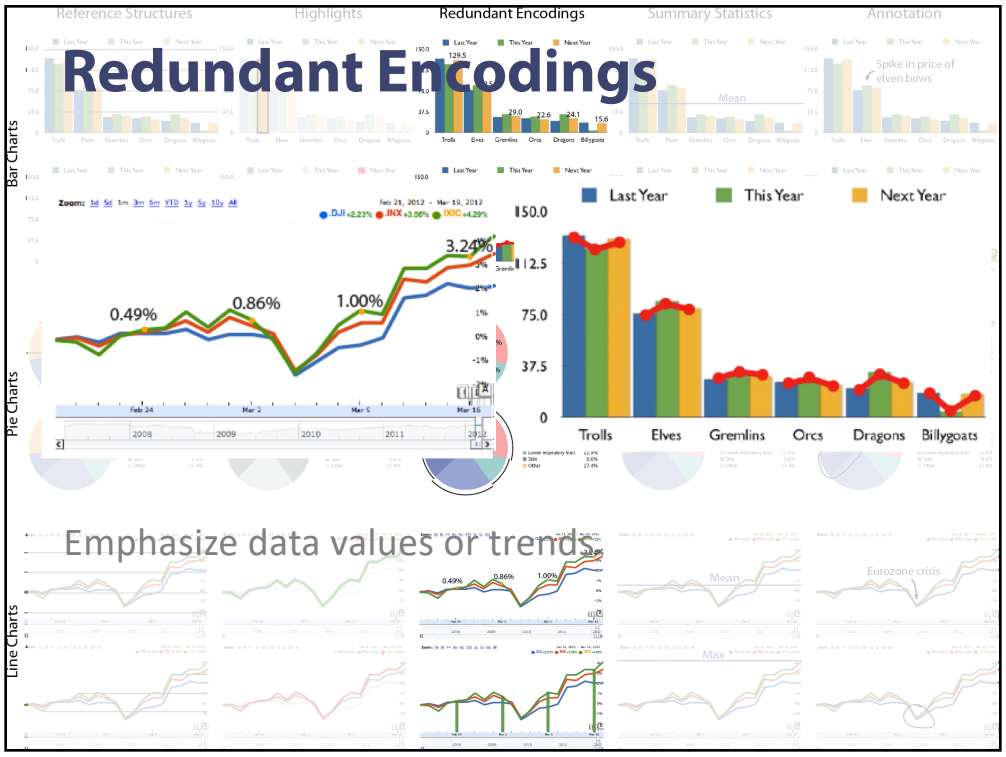
Overlay Underlay
Static Interactive
Divisions: 4
Line thickness: 1

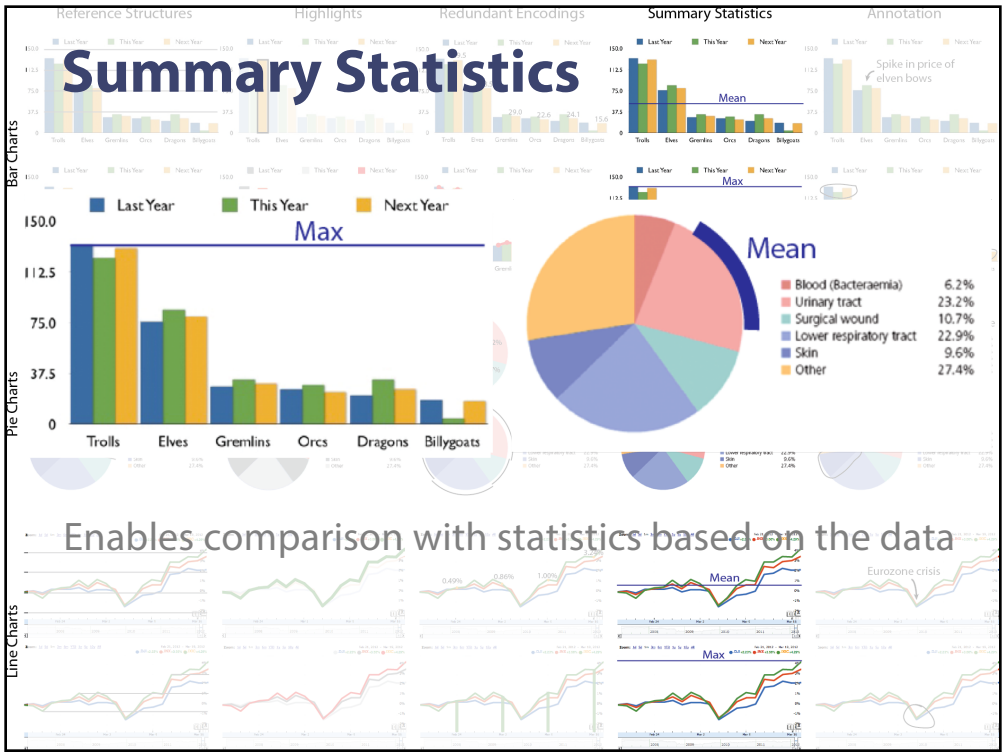
Places regular gridlines at user defined intervals.

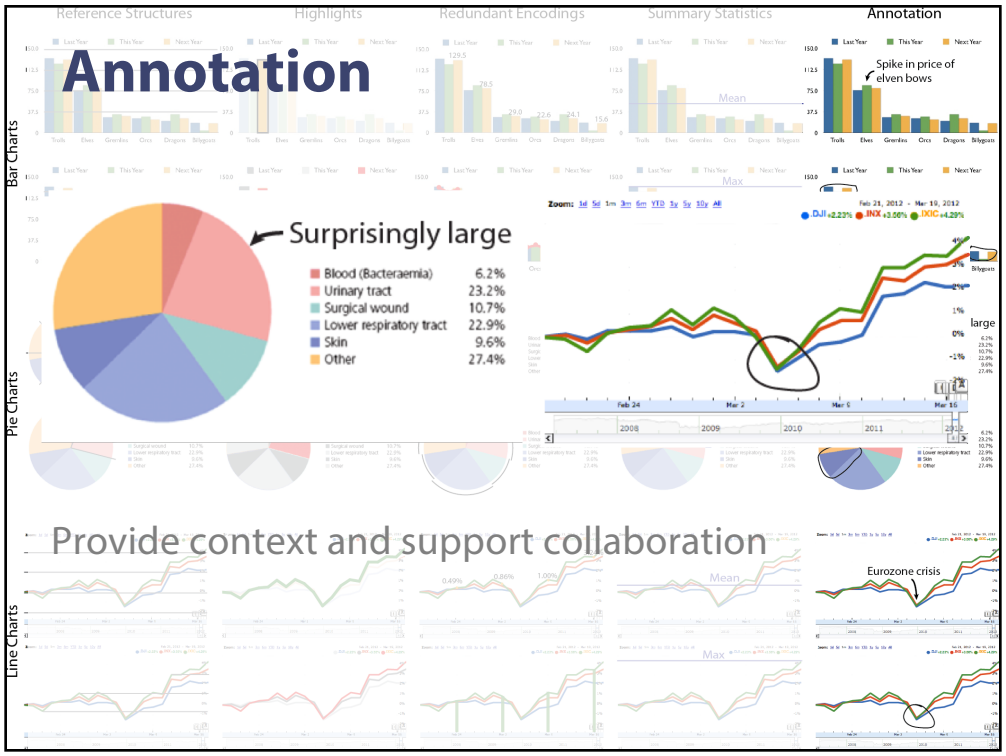
Demo

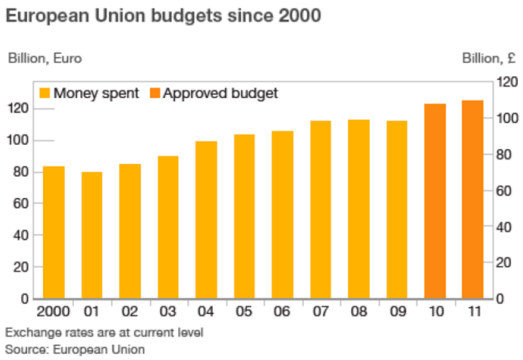



Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Highlights
Bar Charts
Pie Charts
Line Charts
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Mean
Spike in price of elven bows
Eurozone crisis
Max
Draws viewers' attention to specific marks


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Redundant Encodings
Bar Charts
Pie Charts
Line Charts
Last Year
This Year
Next Year
0.49%
0.86%
1.00%
3.24%
Trolls
Elves
Gremlins
Orcs
Dragons
Billygoats
Mean
Spike in price of elven bows
Eurozone crisis
Max
Emphasize data values or trends


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Summary Statistics
Bar Charts
Pie Charts
Line Charts
Last Year
This Year
Next Year
Max
Mean
Trolls
Elves
Gremlins
Orcs
Dragons
Billygoats
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Spike in price of elven bows
Eurozone crisis
Enables comparison with statistics based on the data


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Annotation
Bar Charts
Pie Charts
Line Charts
Surprisingly large
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Spike in price of elven bows
Mean
Max
Eurozone crisis
Provide context and support collaboration



mark: lines

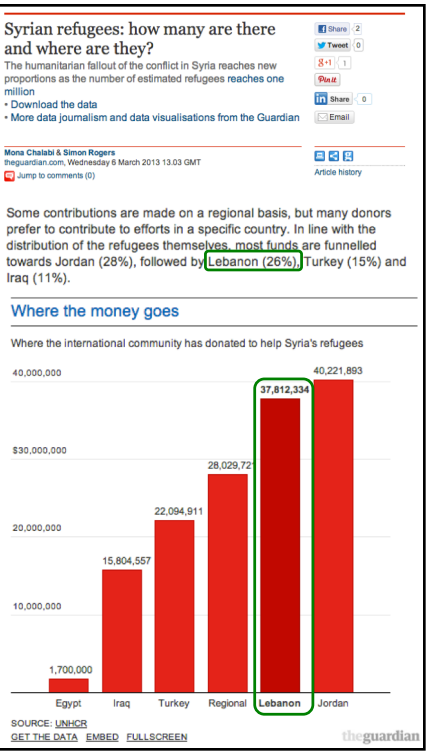
year	money
2000	85
2001	78
2002	87
2003	90
2004	98
…	…

Most overlays only require access to marks

- Reference structures (**marks**)
- Highlights (**marks**)
- Redundant encodings (**marks** and ***data***)
- Summary statistics (**marks**)
- Annotations (**marks**)

Interactive Documents

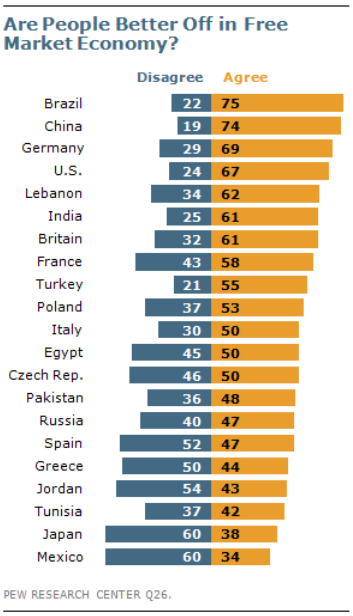
How can we facilitate reading text and charts together?



Goal: Extract references between text and chart

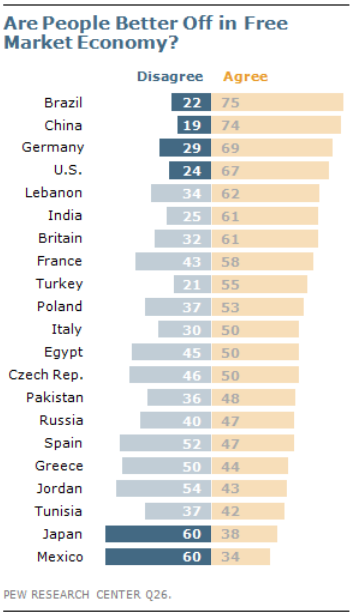
Problem: Diversity of writing styles

Example 1: Pew Research



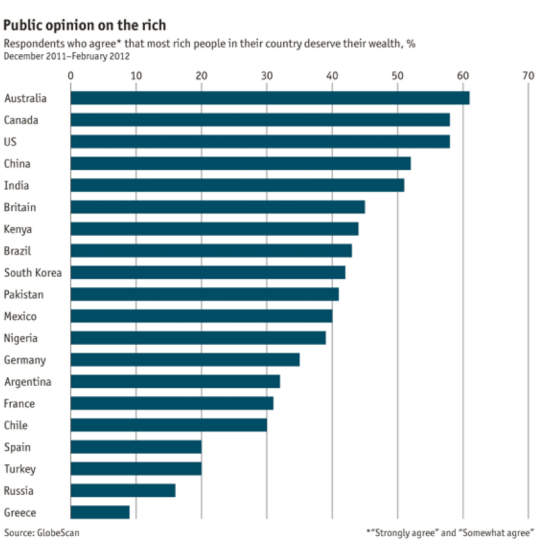
Skepticism for capitalism is lowest in Brazil (22%), China (19%), Germany (29%) (although East Germans are less supportive than West Germans) and the U.S. (24%). Skepticism for free markets is highest in Mexico (60%) and Japan (60%).

Example 1: Pew Research



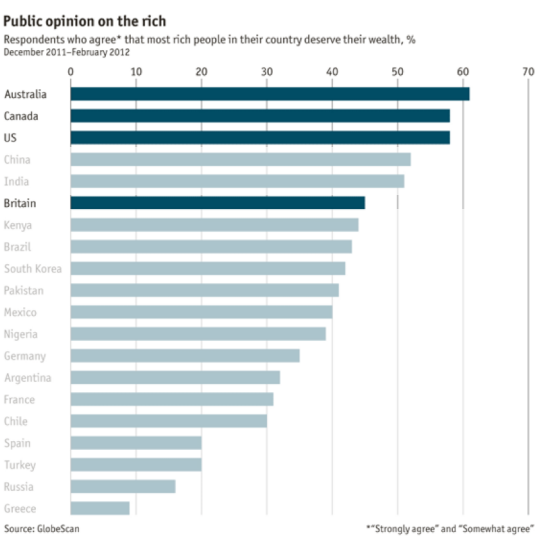
Skepticism for capitalism is lowest in **Brazil (22%), China (19%), Germany (29%)** (although East Germans are less supportive than West Germans) and the **U.S. (24%)**. Skepticism for free markets is highest in **Mexico (60%)** and **Japan (60%).**

Example 2: Economist

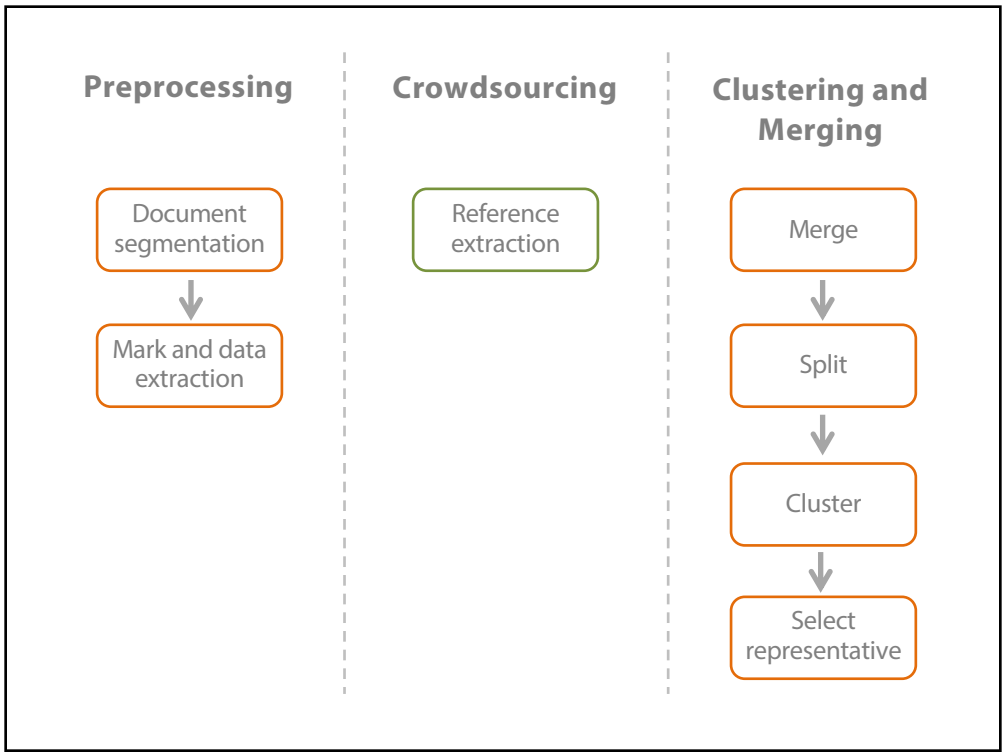


Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions. Europeans and Latin Americans tend to have similar attitudes to the rich; the Anglo-Saxon world is a bit more forgiving.

Example 2: Economist

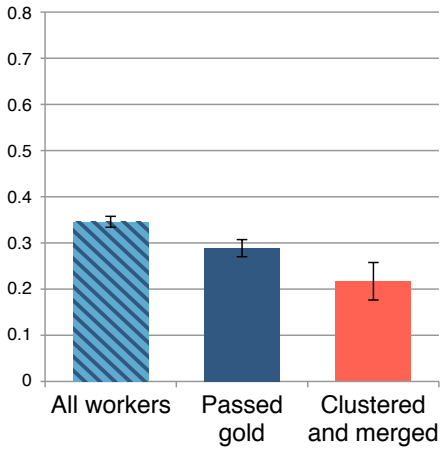


Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions. Europeans and Latin Americans tend to have similar attitudes to the rich; **the Anglo-Saxon world** is a bit more forgiving.


Preprocessing
Crowdsourcing
Clustering and Merging
Document segmentation
Mark and data extraction
Reference extraction
Merge
Split
Cluster
Select representative

Demo

Evaluation

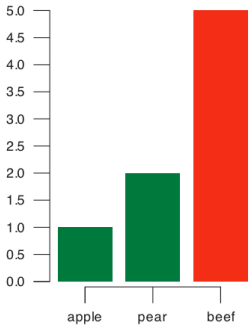


Avg. F_1 distance: expert specified references vs. crowd specified references

Ongoing and Future Work

Deconstructing D3 Charts

```
1 items = [{name: "apple",type: "fruit", cost: 1.00},
2          {name: "pear", type: "fruit", cost: 2.00},
3          {name: "beef", type: "meat",  cost: 5.00}]
4 var bars = svg.selectAll("rect")
5                .data(items)
6                .enter()
7                .append("rect");
8 bars.attr("x", function(d, i)
9         {return i * 25;})
10    .attr("y", function(d)
11      {return h - d.price * 10;})
12    .attr("height", function(d)
13      {return d.price * 10;})
14    .attr("fill", function(d, i)
15      {if(d.type == "fruit"){return "green";}
16       else if (d.type == "meat"){return "red";}})
17    .attr("width", "20px")
18    .attr("stroke-width", 0);
```



Data

deconID	name	type	cost
2	apple	fruit	1.00
3	pear	fruit	2.00
4	beef	meat	5.00

Marks

fill	xPosition	height
green	35 px	20 px
green	60 px	40 px
red	85 px	100 px

Mappings

type ↔ *fill*
cost ↔ *height*
cost ↔ *yPos*
cost ↔ *area*
deconID ↔ *xPos*

D3 Code **D3 Chart** **Our Deconstruction**

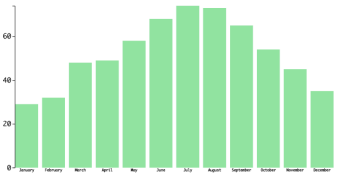
Automatically convert D3 code into mapping based representation to enable redesign and style reuse

Deconstructing and Restyling D3 Visualizations. Jonathan Harper and Maneesh Agrawala. User Interface Software Technology (UIST) 2014.

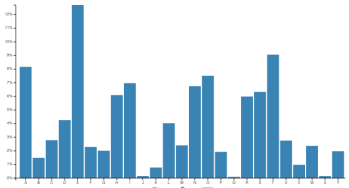
Automatic Redesign

Can we automatically redesign charts to improve

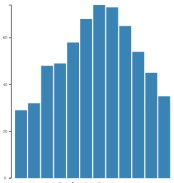
Perceptual effectiveness?
Visual aesthetics?
Accessibility for vision impaired users?



Data Source

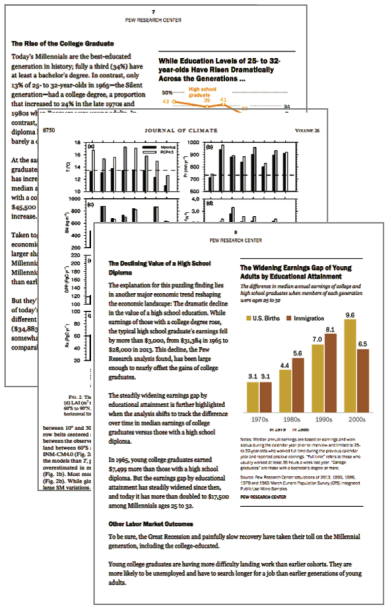


Style Target



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***