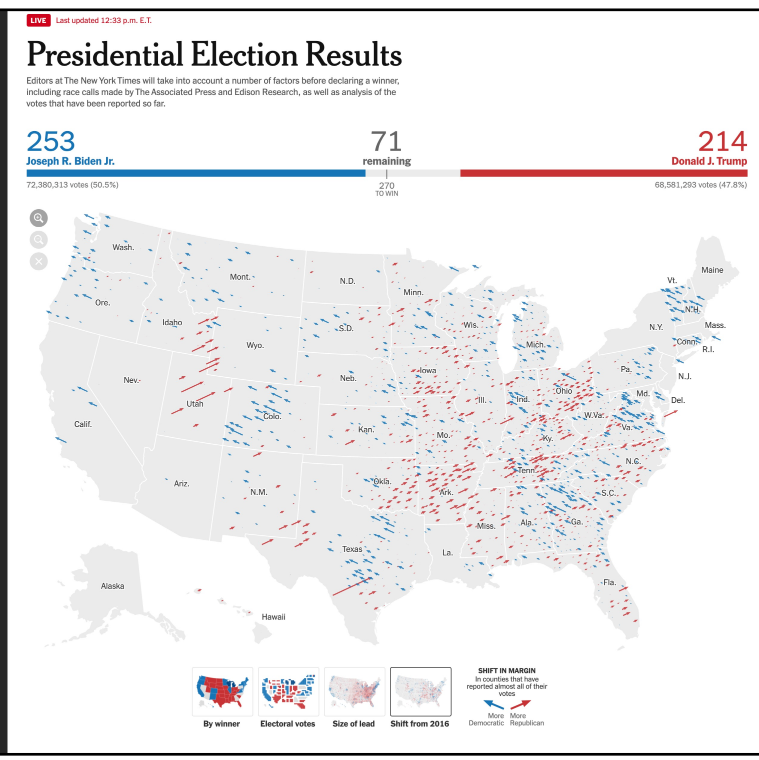
Deconstructing Visualizations

Maneesh Agrawala

CS 448B: Visualization
Fall 2021

1



2

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?

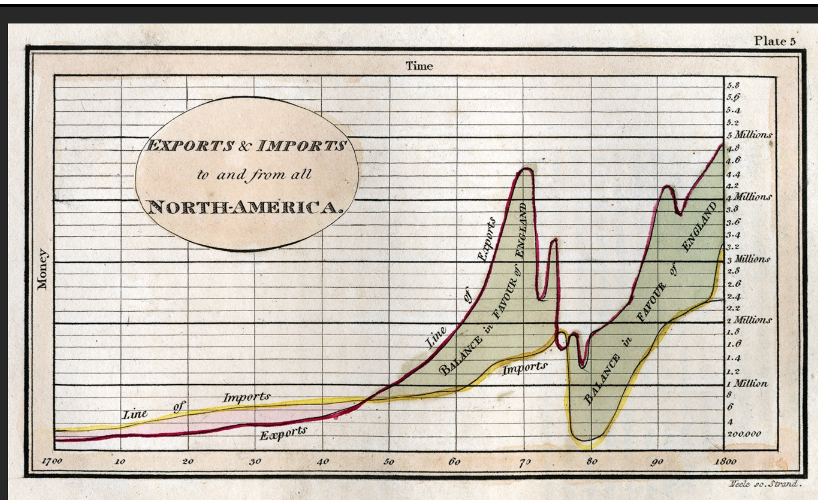
When using social network analysis, how do you validate your findings and/or determine if your findings are statistically significant? Is there an analogous "p value" standard for graph analysis? [Do you use qualitative or quantitative measures of validity?]

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?

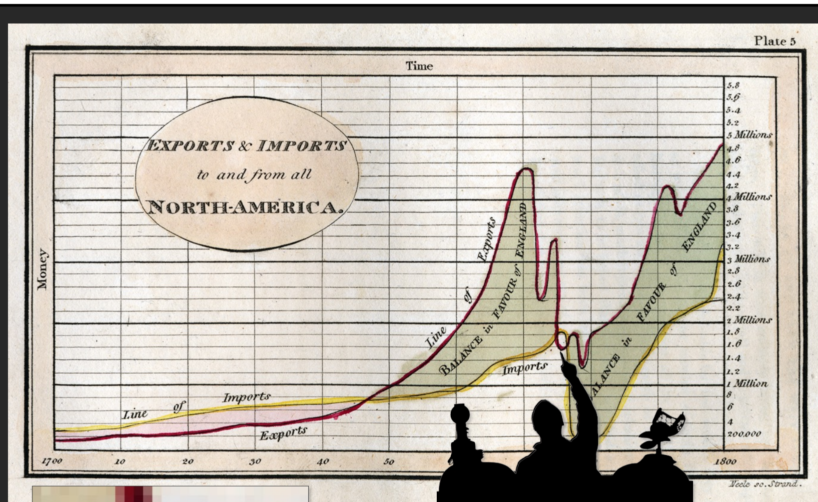
3

Deconstructing Visualizations

4



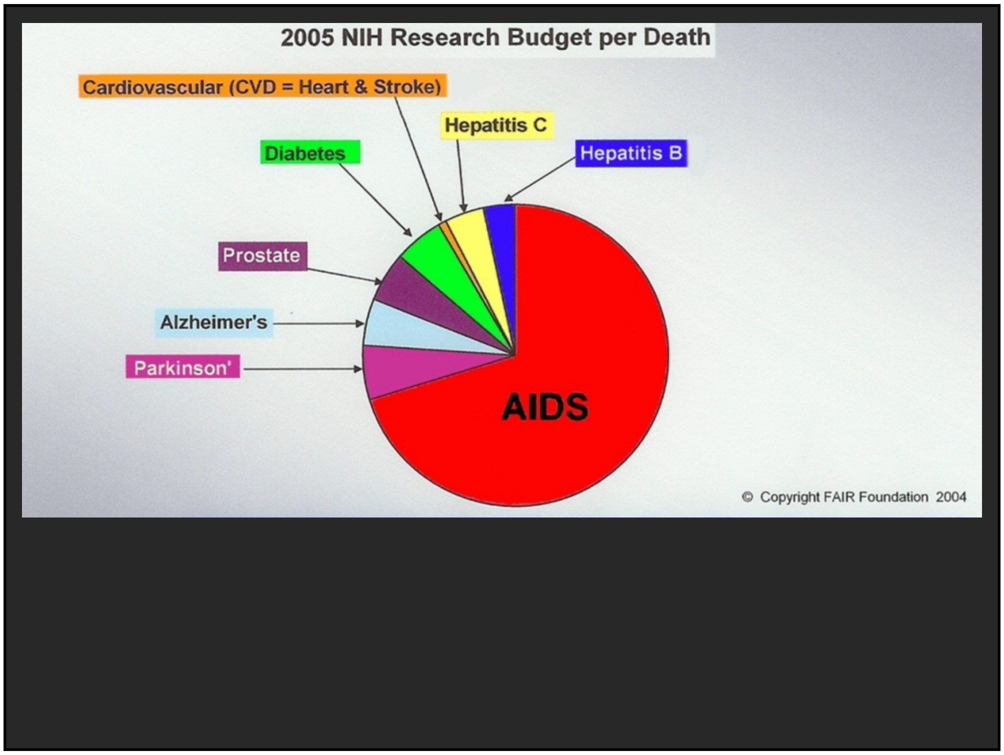
5



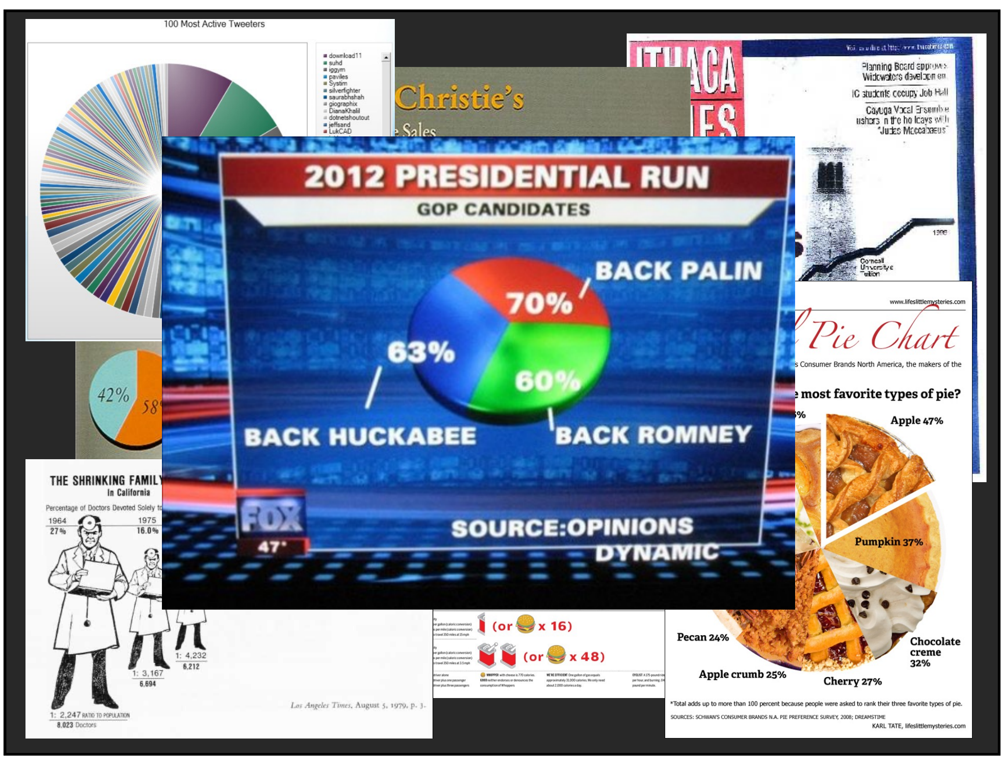
Pixels are poor representation

Hard for machines to retrieve data

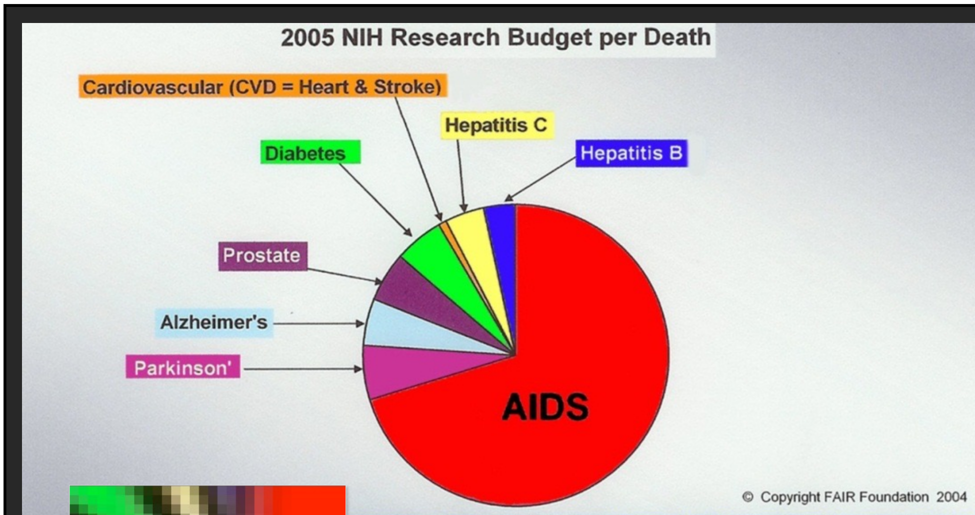
6

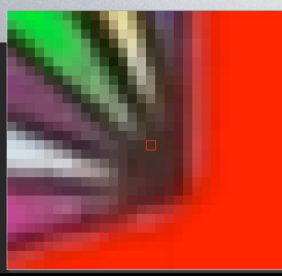

2005 NIH Research Budget per Death
Cardiovascular (CVD = Heart & Stroke)
Hepatitis C
Diabetes
Hepatitis B
Prostate
Alzheimer's
Parkinson'
AIDS
© Copyright FAIR Foundation 2004

7


100 Most Active Tweeters
Christie's
2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
SOURCE:OPINIONS
DYNAMIC
42%
THE SHRINKING FAMILY
in California
Los Angeles Times, August 5, 1979, p. 3.
Pie Chart
most favorite types of pie?
Apple 47%
Pumpkin 37%
Pecan 24%
Chocolate creme 32%
Apple crumb 25%
Cherry 27%
(or x 16)
(or x 48)

8



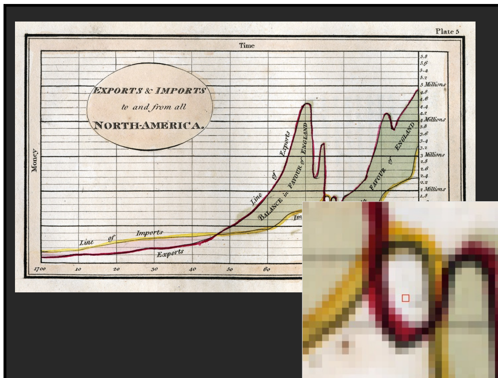


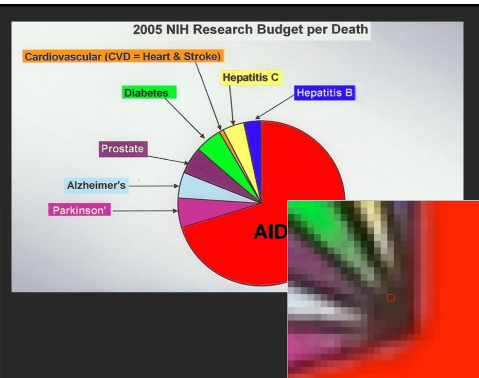
Pixels are poor representation

Hard for machines to retrieve data

Hard for people to manipulate

9





Pixels are a poor representation of charts and graphs

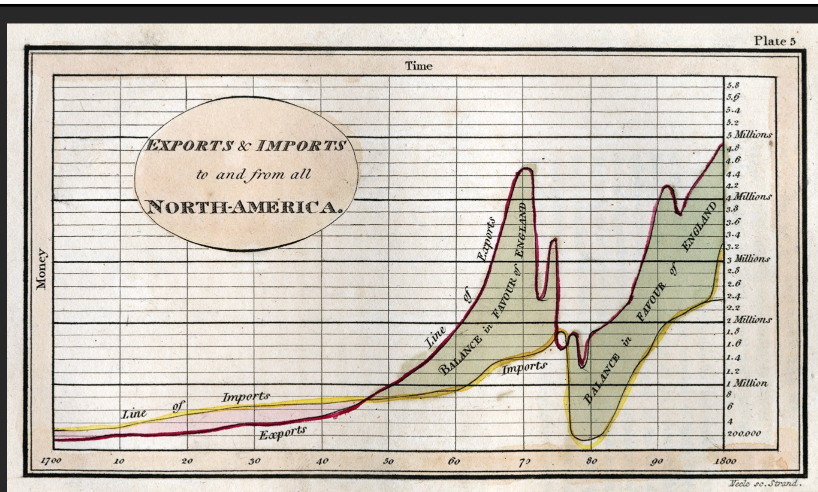
Cannot index, search, manipulate or interact with the data

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

10

What is a good representation?

11



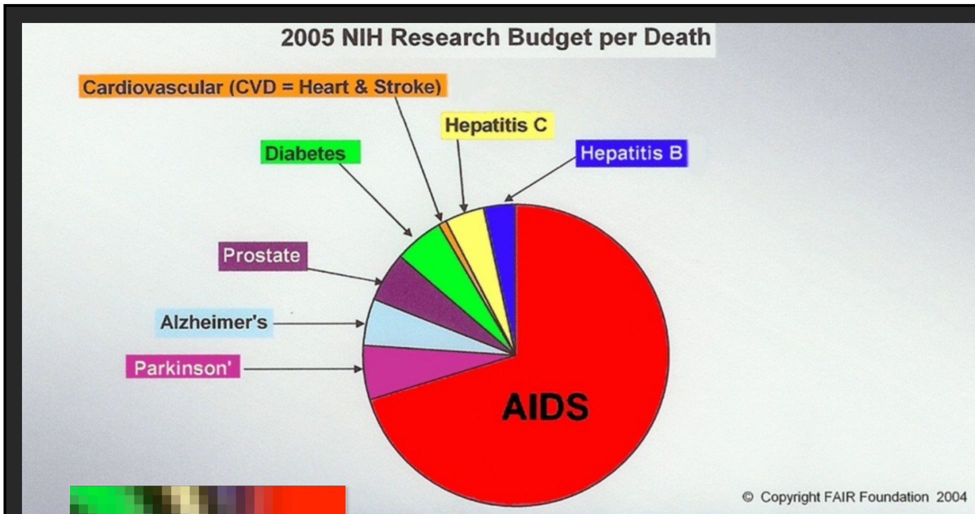
Data

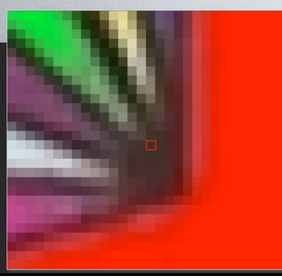
Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
1703	180,000	312,000
1704	187,000	319,000
…	…	…

Marks
mark: lines

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

12





Data

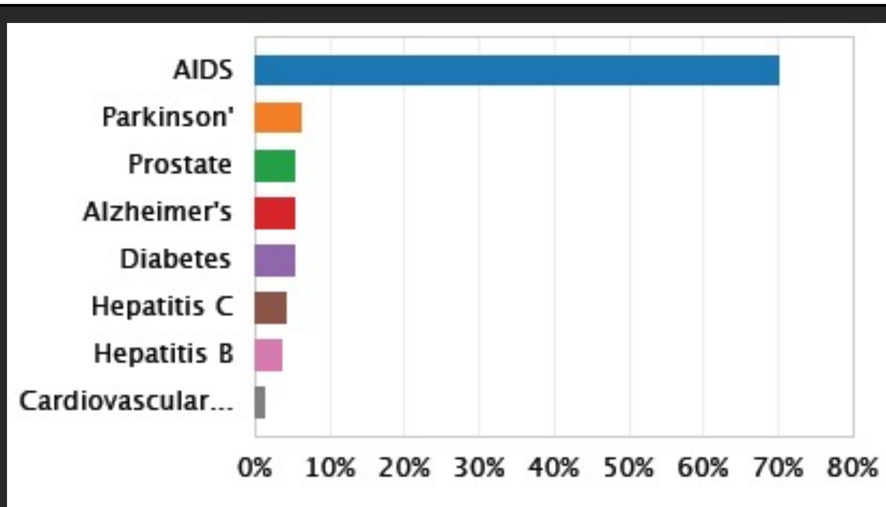
Disease	Budget
Aids	70.0%
Alzheimer's	5.0%
Cardiovascular	1.1%
Diabetes	4.8%
Hepatitus B	4.1%
Hepatitus C	3.8%
Parkinson'	6.0%
Prostate	5.2%

Marks
Mappings

mark: areas

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

13



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)

14

Approach

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

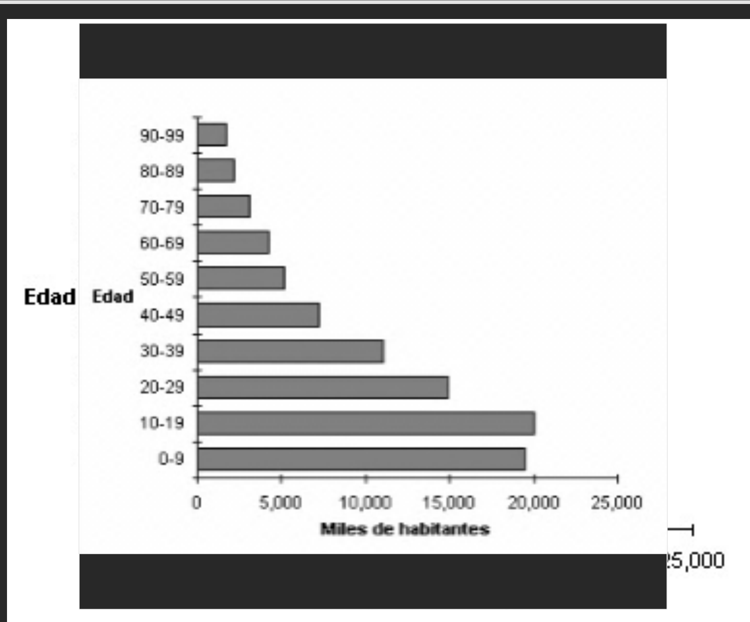
15

15

Classification

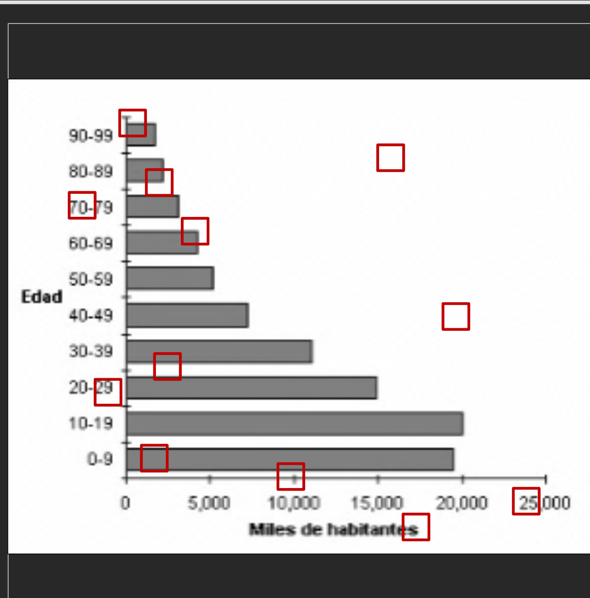
16

Training the Classifier



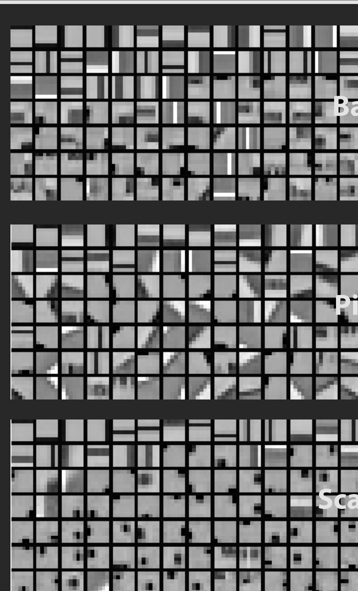
17

Training the Classifier



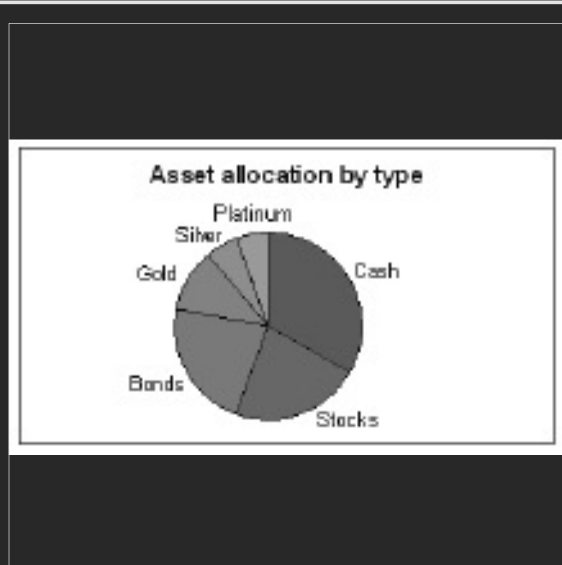
18

Training the Classifier



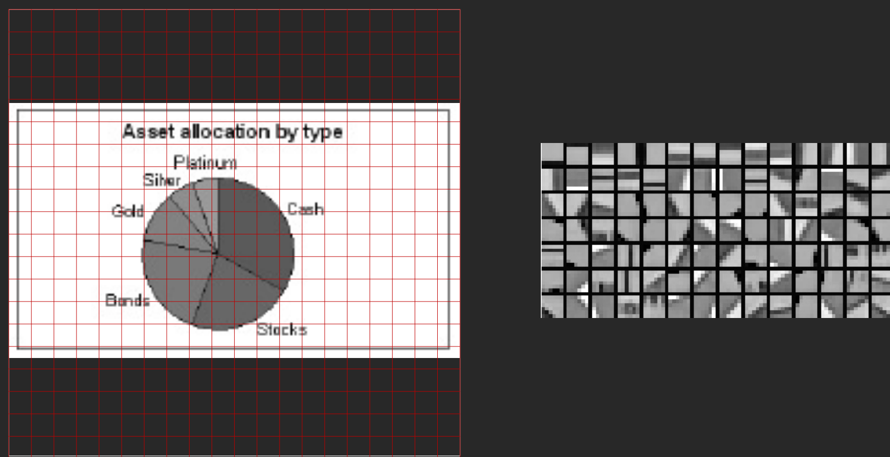
19

Classifying an Input Image



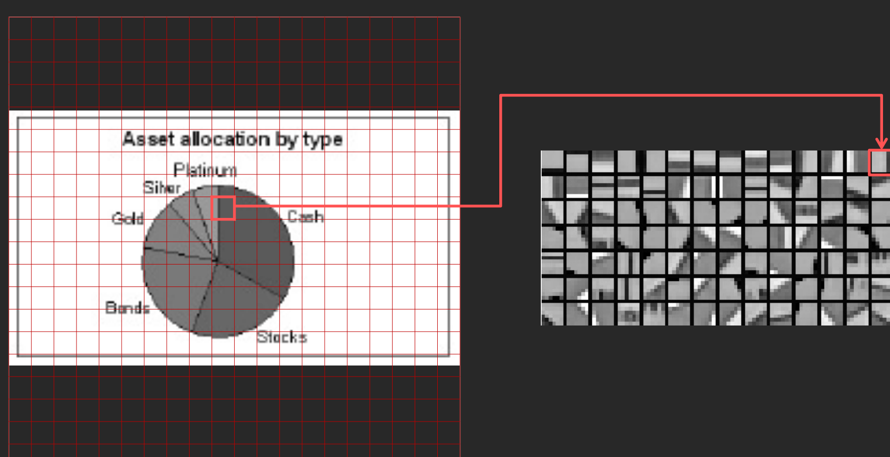
20

Classifying an Input Image



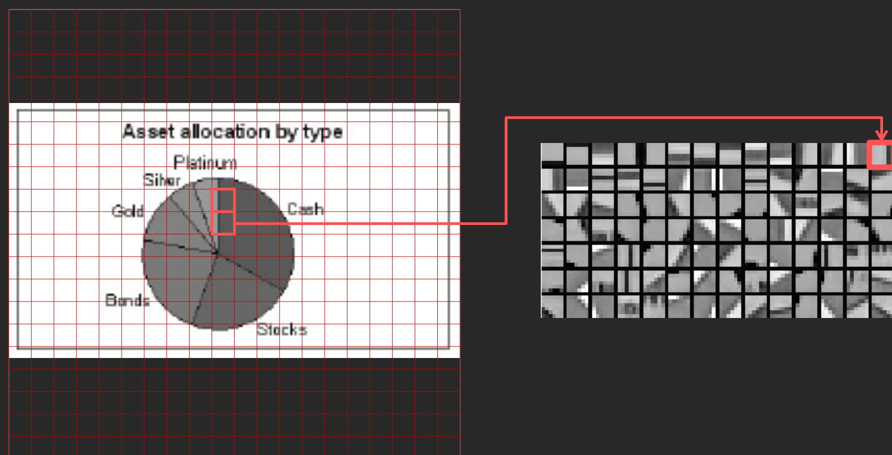
21

Classifying an Input Image



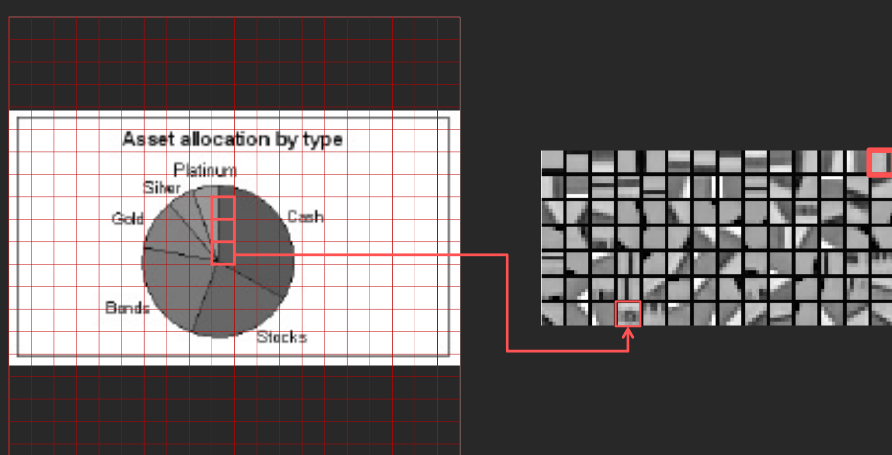
22

Classifying an Input Image



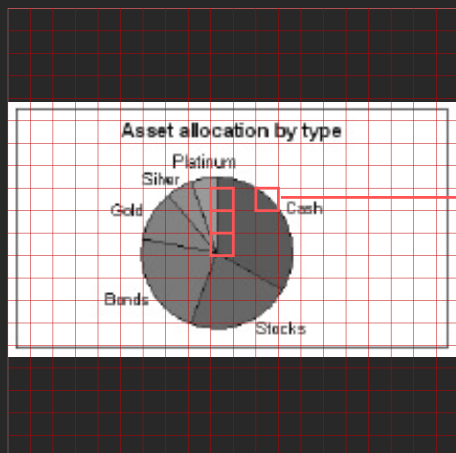
23

Classifying an Input Image



24

Classifying an Input Image



25

Classifying an Input Image

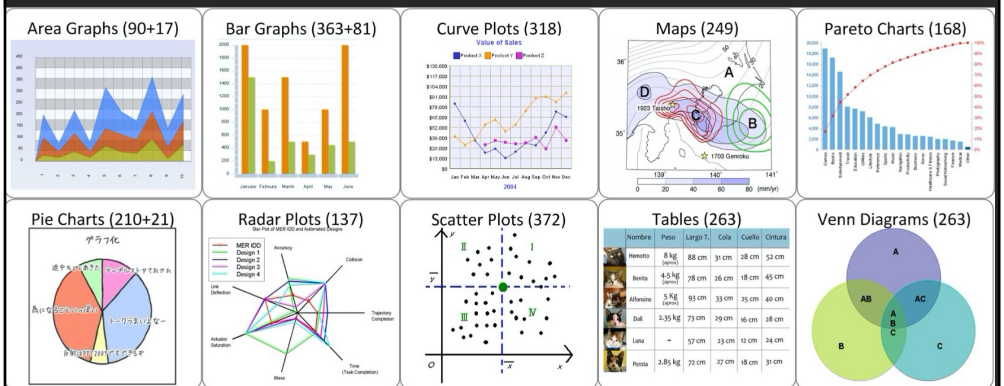
→ **SVM Classifier** → **Pie Chart**

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%

26

Our Corpus

Over 2500 labeled images and 10 chart types



ReVision binary SVMs give 96% classification accuracy

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

27

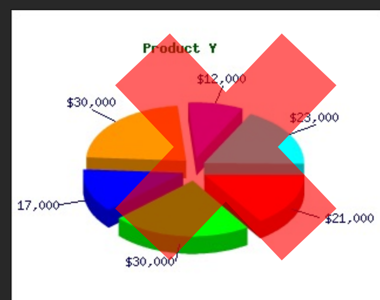
Mark and Data Extraction

28

Assumptions

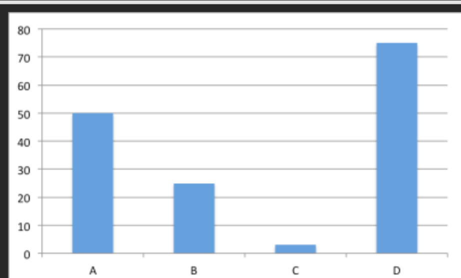
Bar charts and pie charts only

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



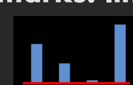
29

Bar Charts



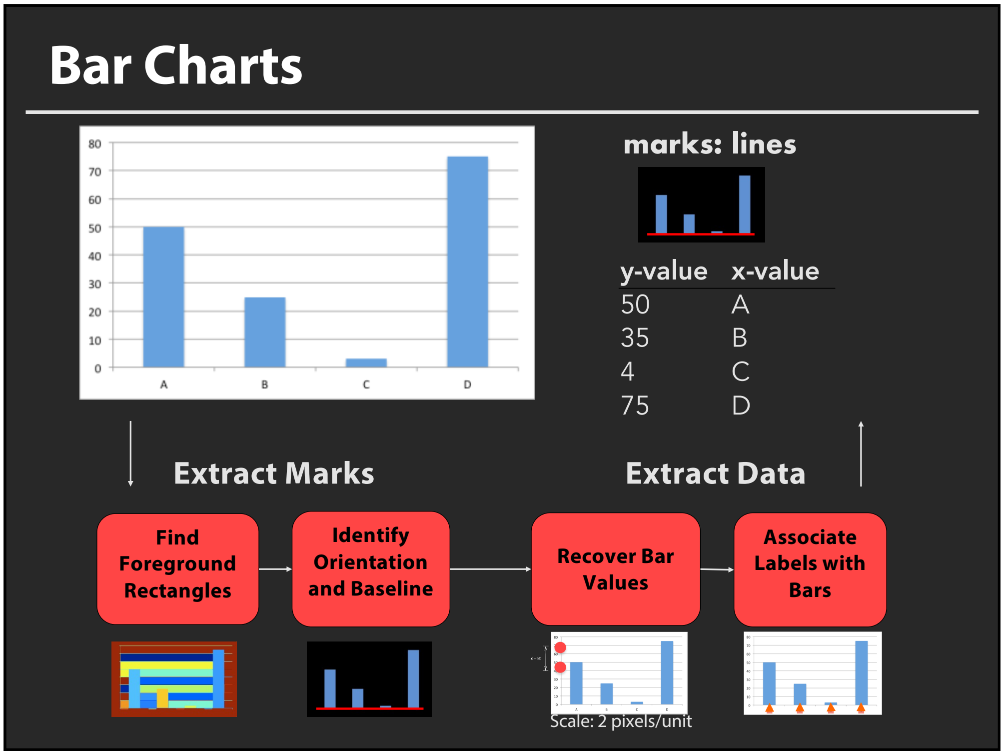
→

marks: lines

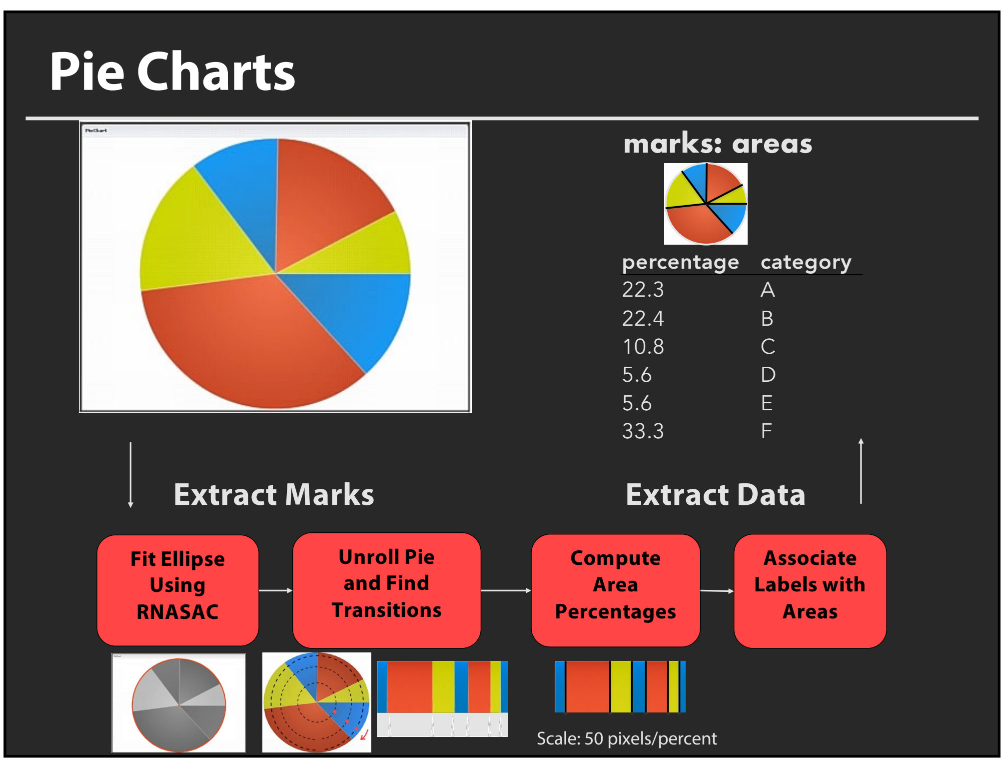


y-value	x-value
50	A
25	B
4	C
75	D

30

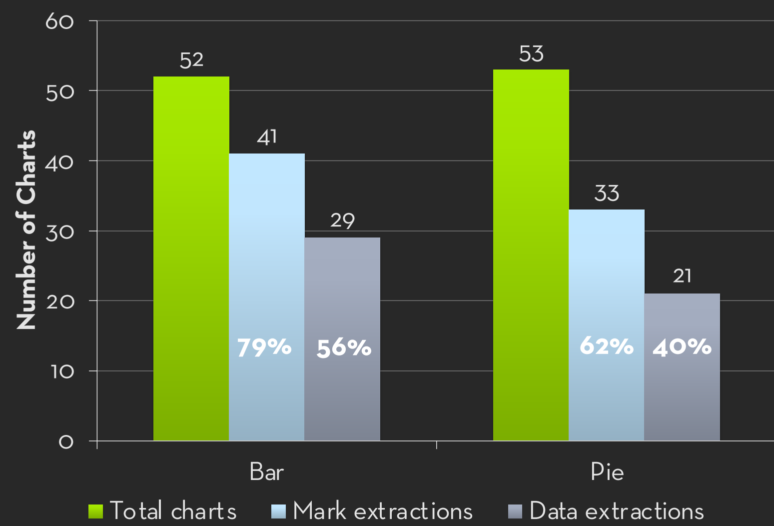

Bar Charts
marks: lines
y-value x-value
50 A
35 B
4 C
75 D
Extract Marks
Extract Data
Find Foreground Rectangles
Identify Orientation and Baseline
Recover Bar Values
Associate Labels with Bars
Scale: 2 pixels/unit

41


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

42

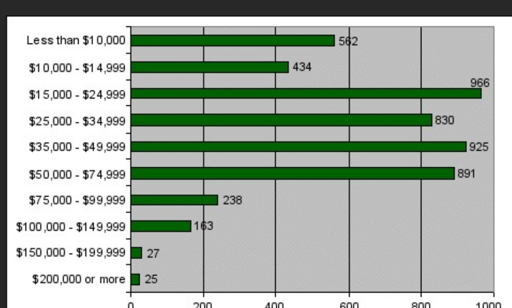
Extraction Results



43

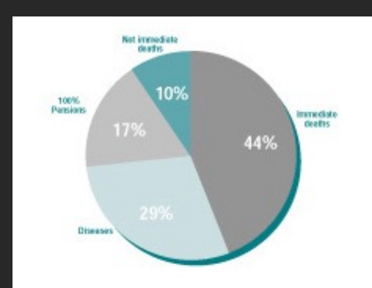
Data Extraction Error

Bar Charts



7.7%

Pie Charts



4.6%

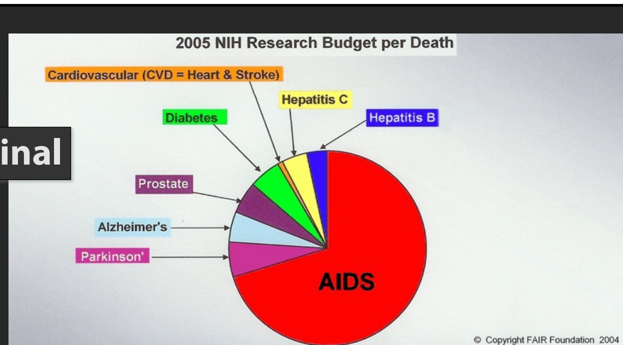
Average chart size: 342 x 452 pixels [Prasad 2007]

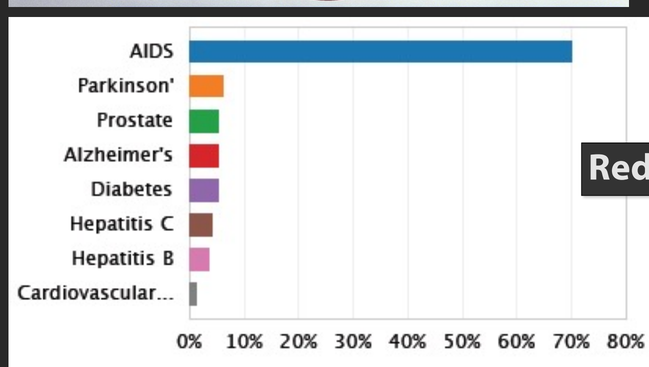
44

Redesign

45

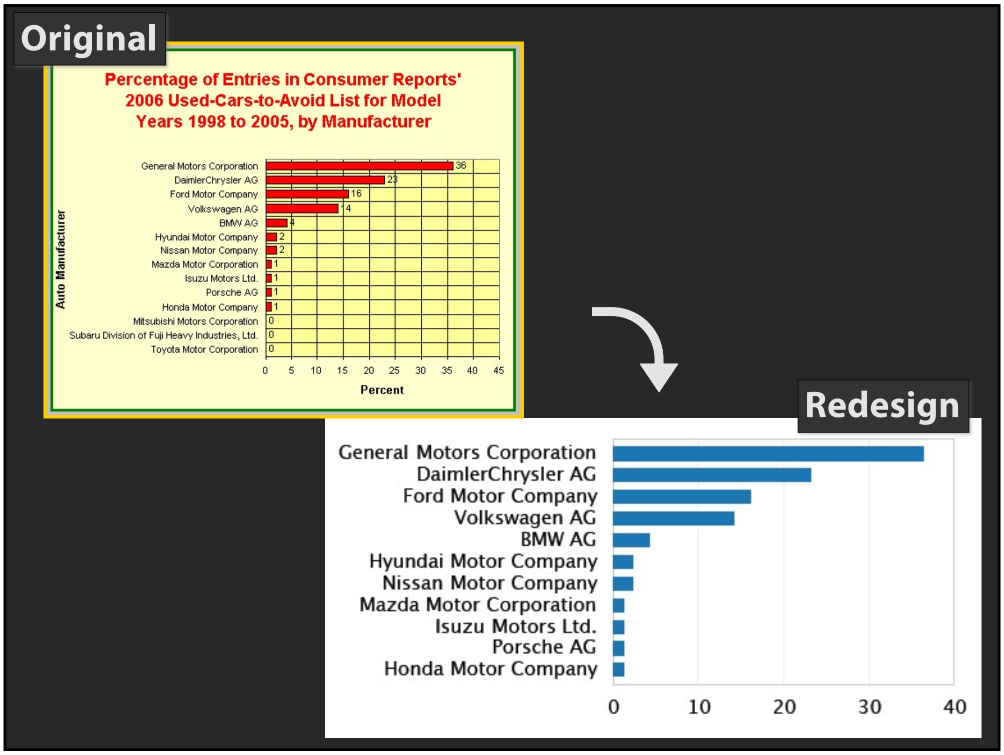
Original



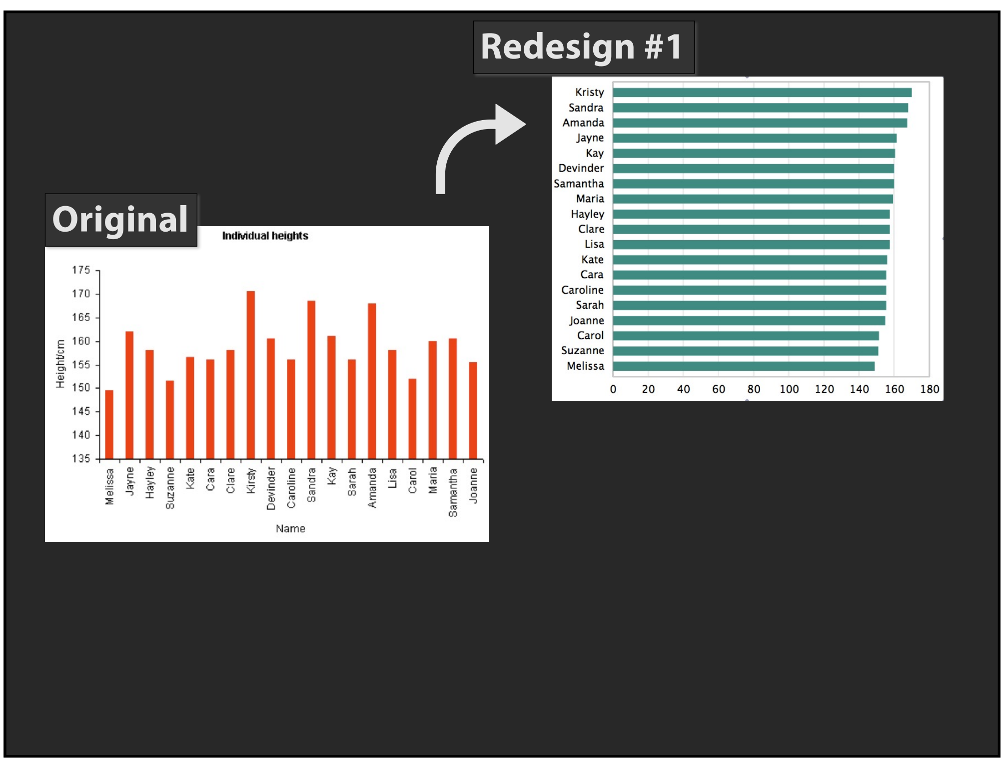


Redesign

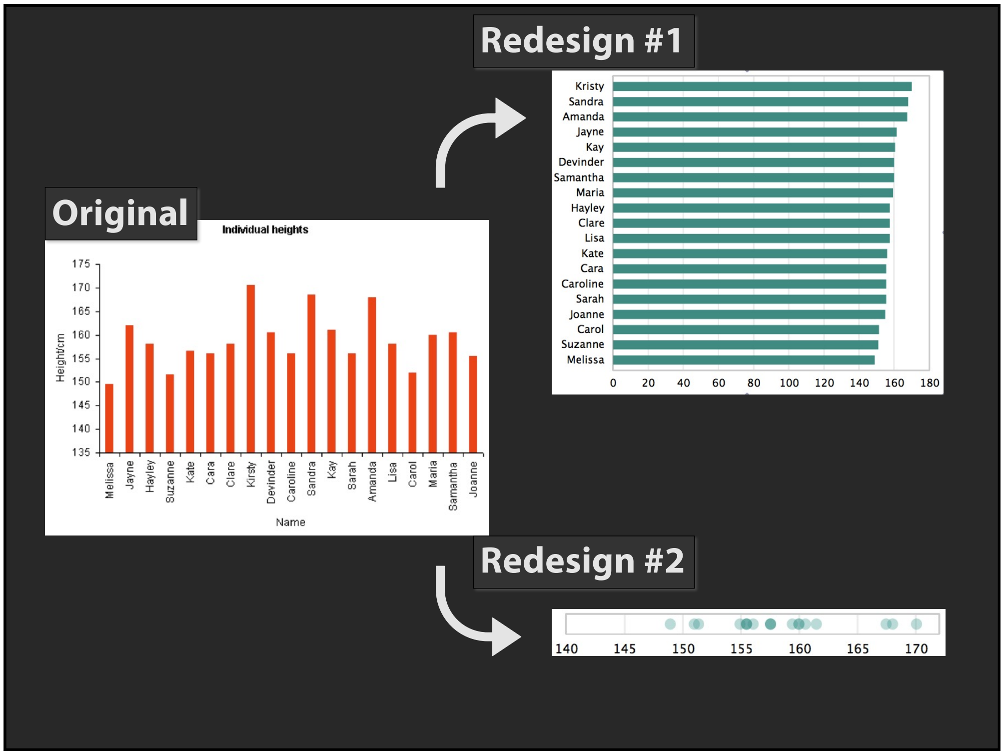
46


Original
Percentage of Entries in Consumer Reports' 2006 Used-Cars-to-Avoid List for Model Years 1998 to 2005, by Manufacturer
Auto Manufacturer
General Motors Corporation
DaimlerChrysler AG
Ford Motor Company
Volkswagen AG
BMW AG
Hyundai Motor Company
Nissan Motor Company
Mazda Motor Corporation
Isuzu Motors Ltd.
Porsche AG
Honda Motor Company
Mitsubishi Motors Corporation
Subaru Division of Fuji Heavy Industries, Ltd.
Toyota Motor Corporation
Percent
Redesign

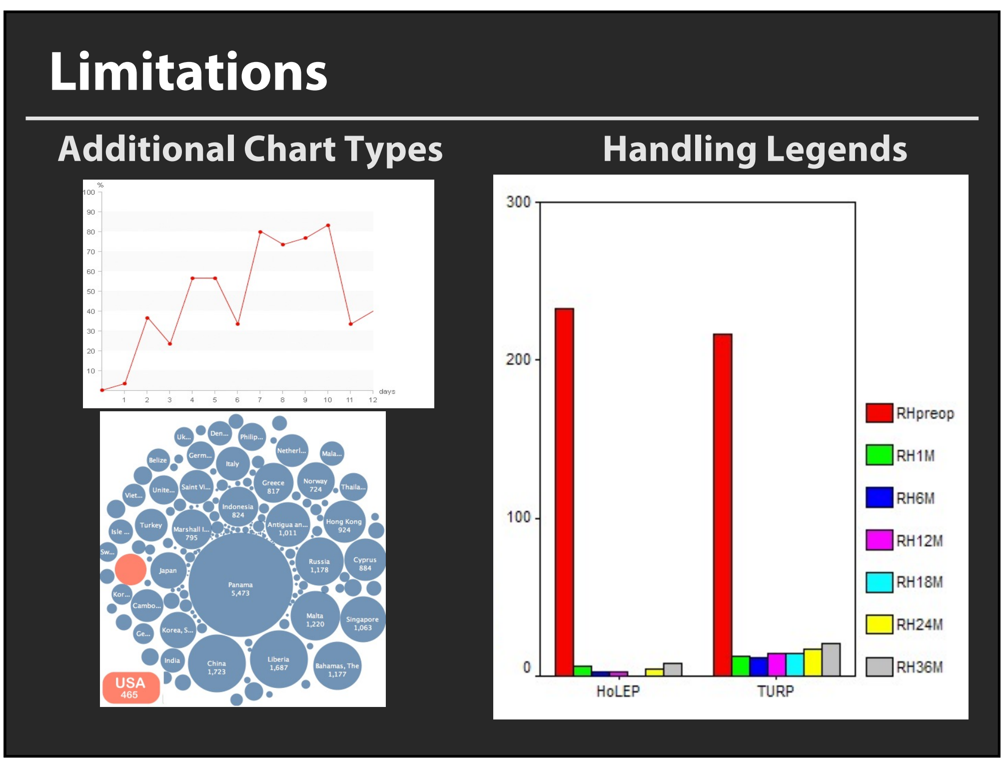
47


Redesign #1
Original
Individual heights
Height/cm
Name

48


Original
Individual heights
Height/cm
Name
Redesign #1
Redesign #2

49


Limitations
Additional Chart Types
Handling Legends
RHpreop
RH1M
RH6M
RH12M
RH18M
RH24M
RH36M
HoLEP
TURP

51

Announcements

52

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

53

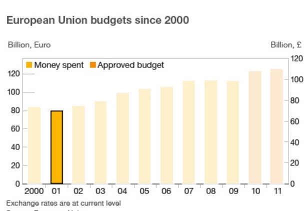
Guest Lecture on Wed

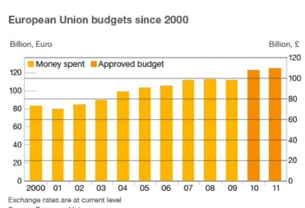
Visualization and NLP

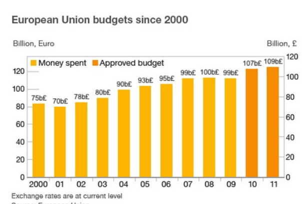
Dr. Dae Hyun Kim (Stanford)

Vidya Setlur (Tableau Research)

54



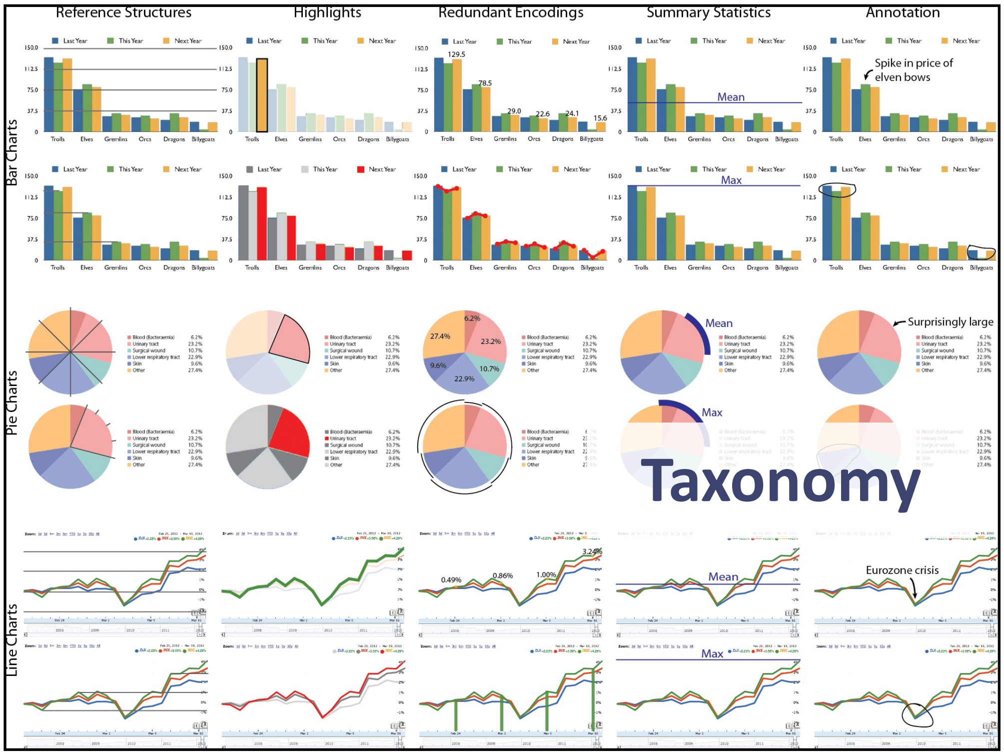




Graphical Overlays

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

55



56

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.

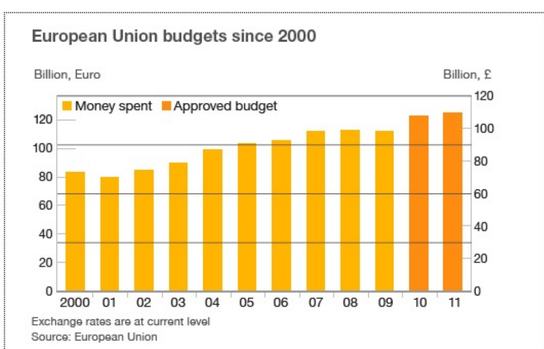


Chart type: Bar
Chart: 00281
Overlay type: Reference structures
- Regular gridlines
- Lines emanating from marks

Parameters

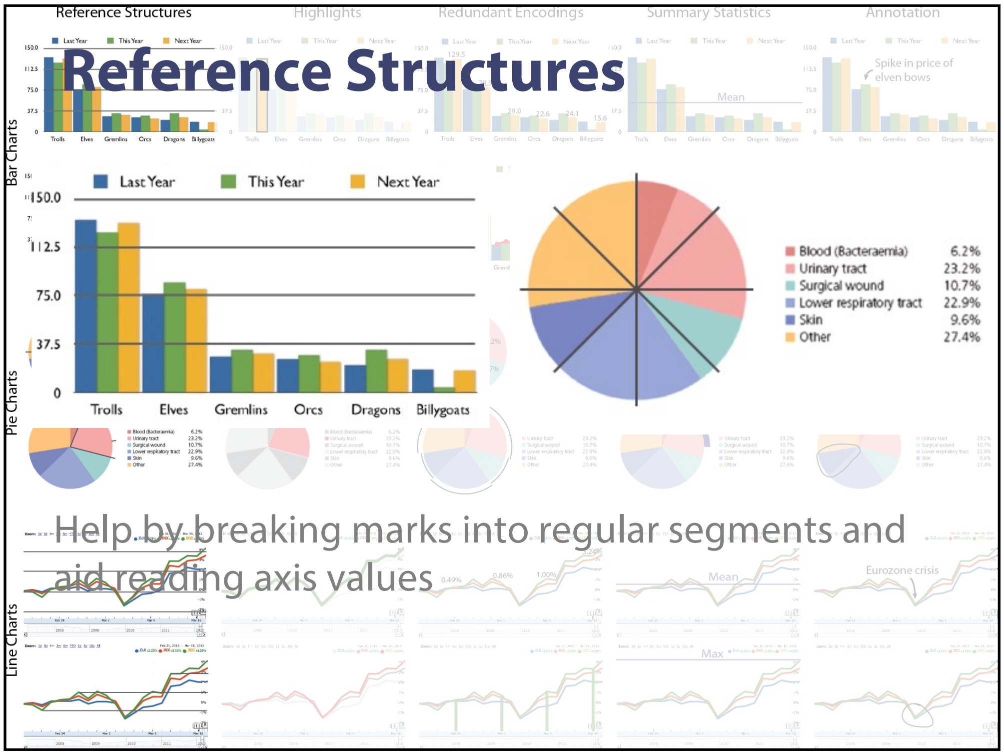
- Overlay
- Underlay
- Static
- Interactive

Divisions: 4
Line thickness: 1

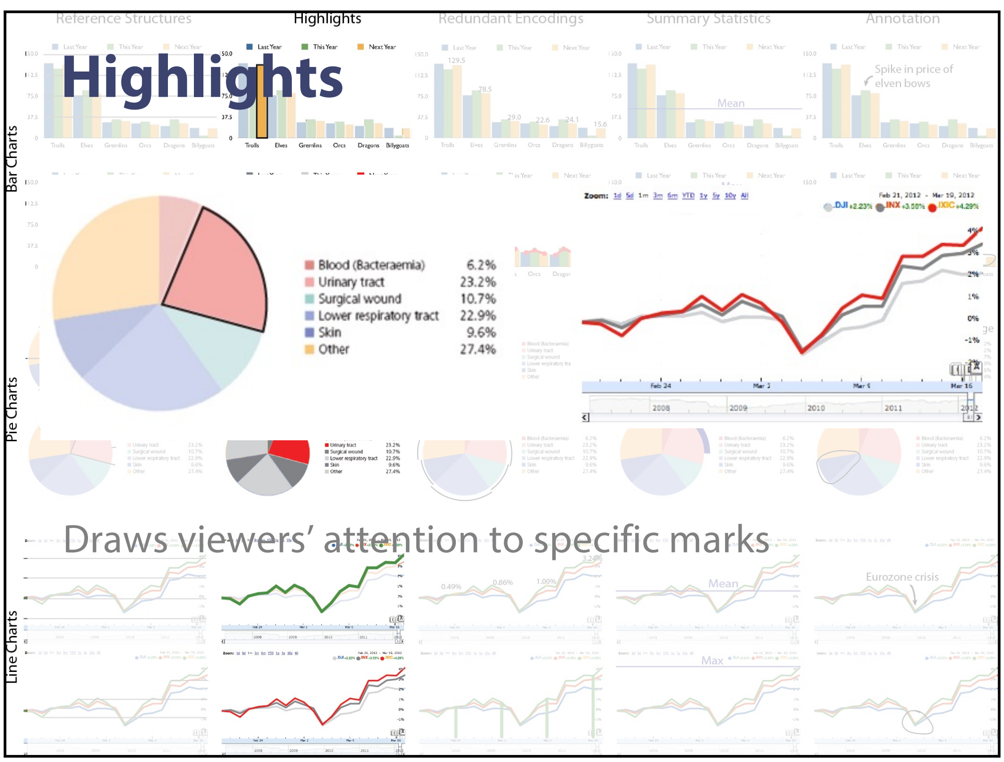
Places regular gridlines at user defined intervals.

Demo

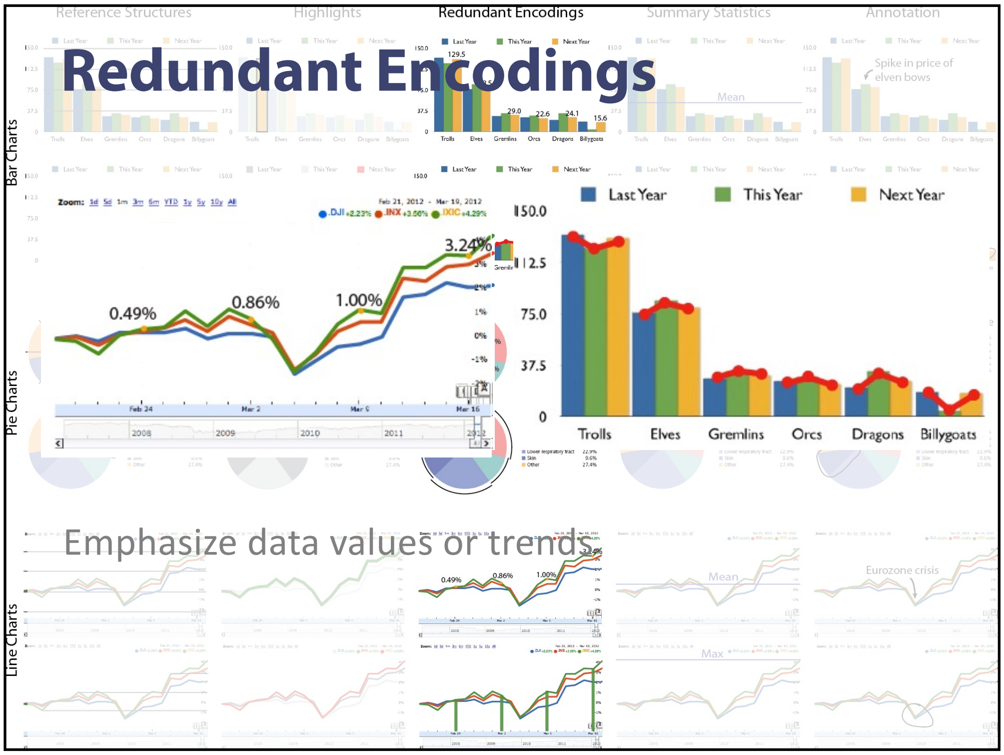
57


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Bar Charts
Pie Charts
Line Charts
Reference Structures
Last Year
This Year
Next Year
150.0
112.5
75.0
37.5
0
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
Mean
Max
Eurozone crisis
Help by breaking marks into regular segments and aid reading axis values

58

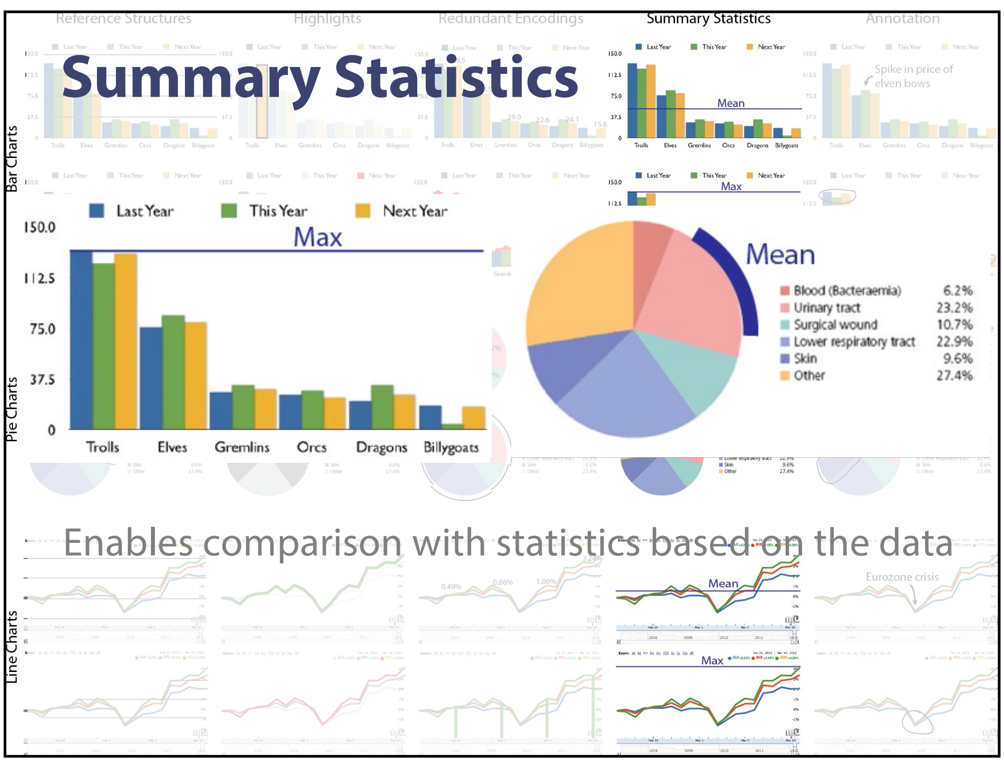

Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Bar Charts
Pie Charts
Line Charts
Highlights
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Spike in price of elven bows
Mean
Max
Eurozone crisis
Draws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

59



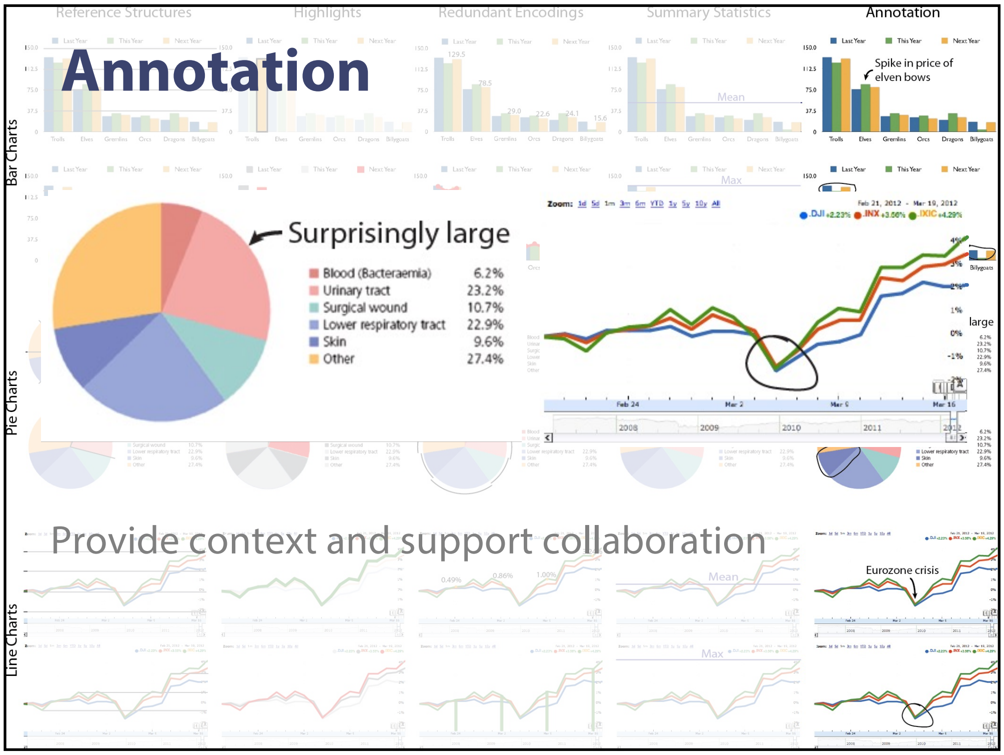
Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Redundant Encodings
Bar Charts
Pie Charts
Line Charts
Last Year
This Year
Next Year
150.0
112.5
75.0
37.5
0
Trolls
Elves
Gremlins
Orcs
Dragons
Billygoats
0.49%
0.86%
1.00%
3.24%
Emphasize data values or trends

60

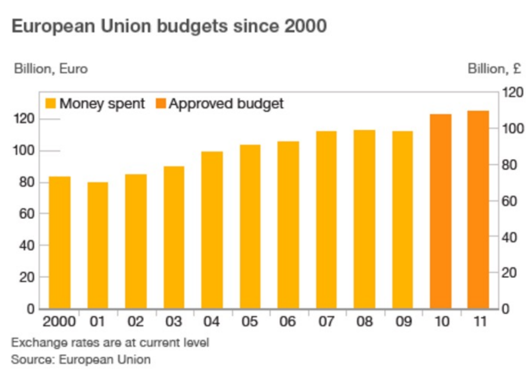


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Summary Statistics
Bar Charts
Pie Charts
Line Charts
Last Year
This Year
Next Year
150.0
112.5
75.0
37.5
0
Max
Trolls
Elves
Gremlins
Orcs
Dragons
Billygoats
Mean
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Enables comparison with statistics based on the data

61



62



→

mark: lines

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

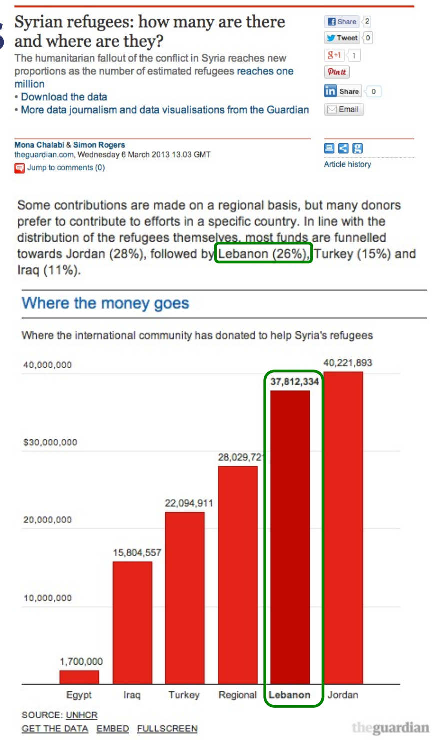
Most overlays only require access to marks

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

63

Interactive Documents

How can we facilitate reading text and charts together?



64

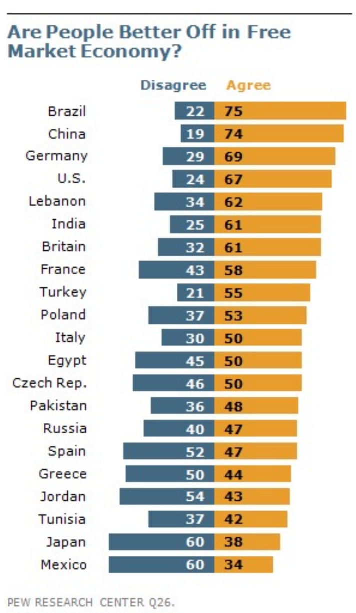
Goal: Extract references between text and chart

65

Problem: Diversity of writing styles

66

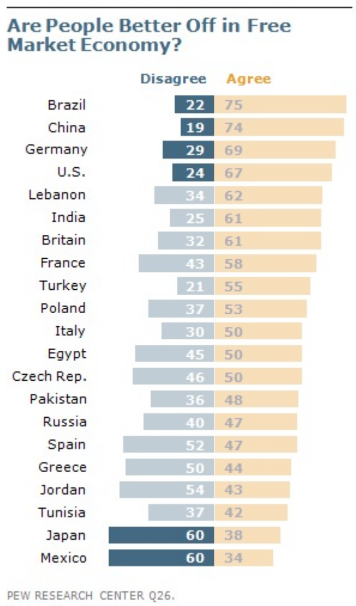
Example 1: Pew Research



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%).

67

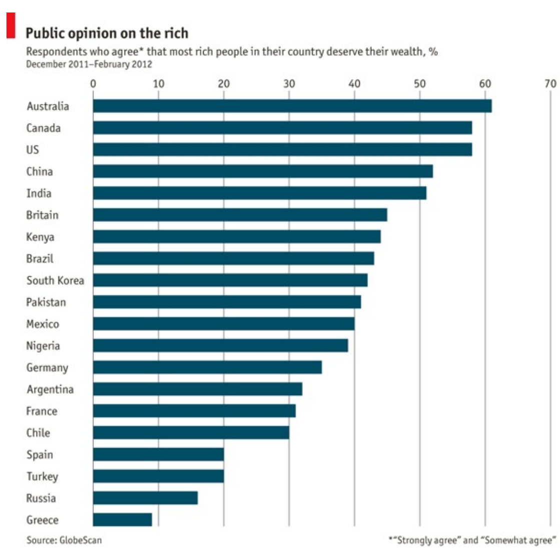
Example 1: Pew Research



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%).**

68

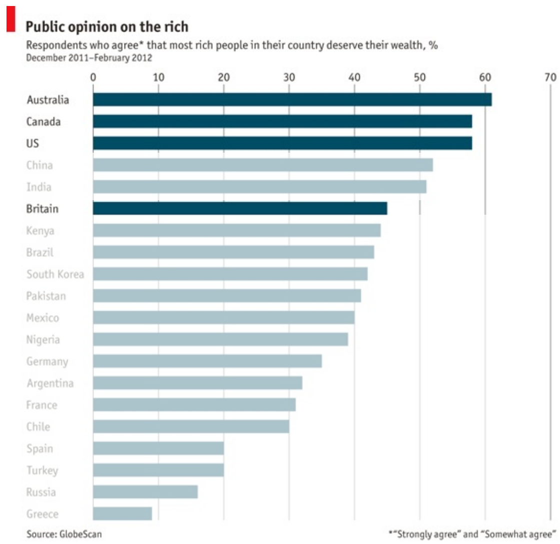
Example 2: Economist



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.

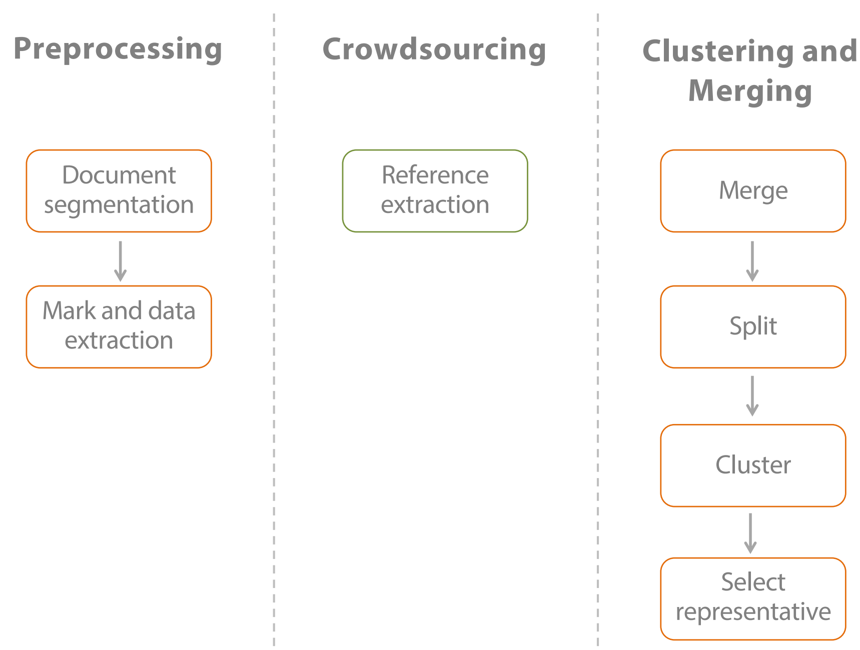
69

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.

70

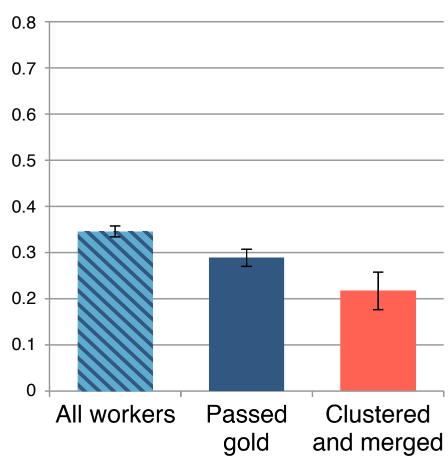


71

Demo

72

Evaluation

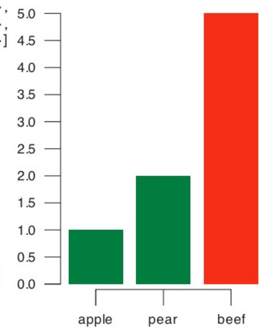


Avg. F_1 distance: expert specified references vs. crowd specified references

73

Deconstructing D3 Charts

```
items = [{name: "apple",type: "fruit", cost: 1.00},
         {name: "pear", type: "fruit", cost: 2.00},
         {name: "beef", type: "meat",  cost: 5.00}]
var bars = svg.selectAll("rect")
              .data(items)
              .enter()
              .append("rect");
bars.attr("x", function(d, i)
        {return i * 25;})
    .attr("y", function(d)
      {return h - d.price * 10;})
    .attr("height", function(d)
      {return d.price * 10;})
    .attr("fill", function(d, i)
      {if(d.type == "fruit"){return "green";}
       else if (d.type == "meat"){return "red";}})
    .attr("width", "20px")
    .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 $\xrightarrow{f}$ *fill*
cost $\xrightarrow{f}$ *height*
cost $\xrightarrow{f}$ *yPos*
cost $\xrightarrow{f}$ *area*
deconID $\xrightarrow{f}$ *xPos*

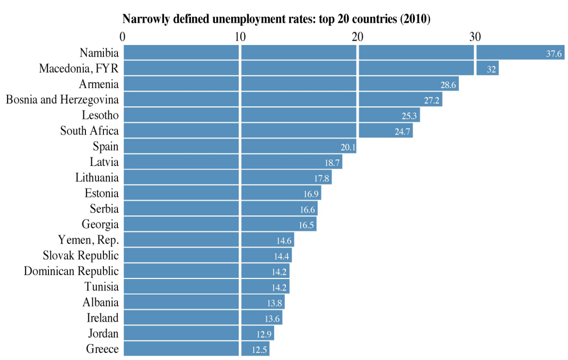
D3 Code | **D3 Chart** | **Our Deconstruction**

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.

75

Deconstructing D3 Charts



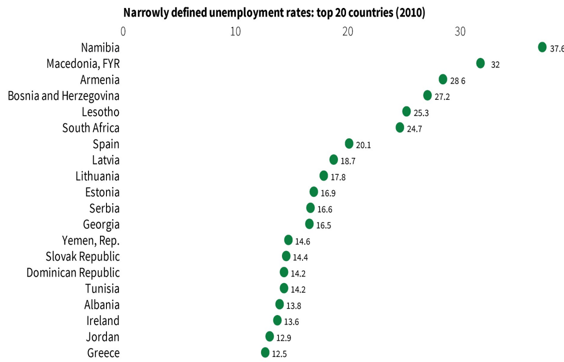
country	rate	deconID
Namibia	37.6	17
Macedonia, FYR	32.0	21
Armenia	28.6	25
Bosnia and Herzegovina	27.2	29
Lesotho	25.3	33
South Africa	24.7	37
Spain	20.1	41
Latvia	18.7	45
…	…	…

rate $\xrightarrow{f}$ *width*
rate $\xrightarrow{f}$ *area*
rate $\xrightarrow{f}$ *xPos*
deconID $\xrightarrow{f}$ *yPos*

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

76

Deconstructing D3 Charts



country	rate	deconID
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
...	...	...

$rate \xrightarrow{f} xPos$

$deconID \xrightarrow{f} yPos$

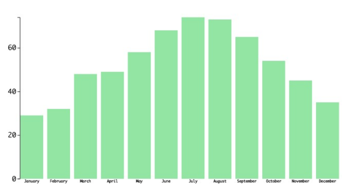
Deconstructing and Restyling D3 Visualizations. Jonathan Harper and Maneesh Agrawala. User Interface Software Technology (UIST) 2014.

77

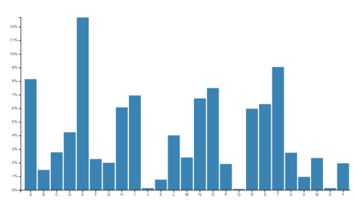
Automatic Redesign

Can we automatically redesign charts to improve

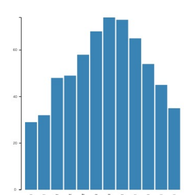
- Perceptual effectiveness?
- Visual aesthetics?
- Accessibility for vision impaired users?



Data Source



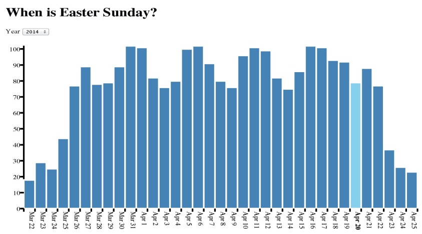
Style Target



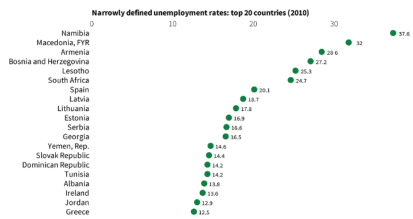
Result

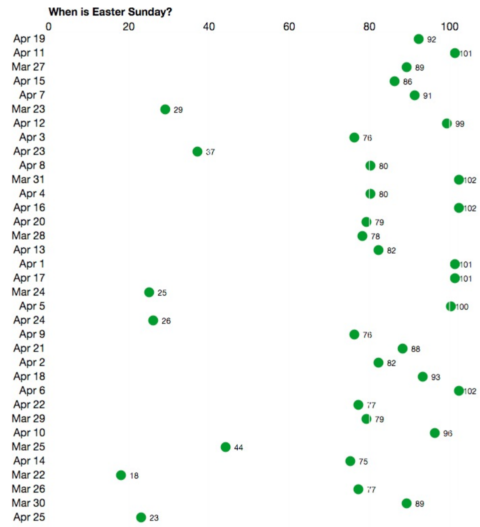
78

Data Source



Style Target

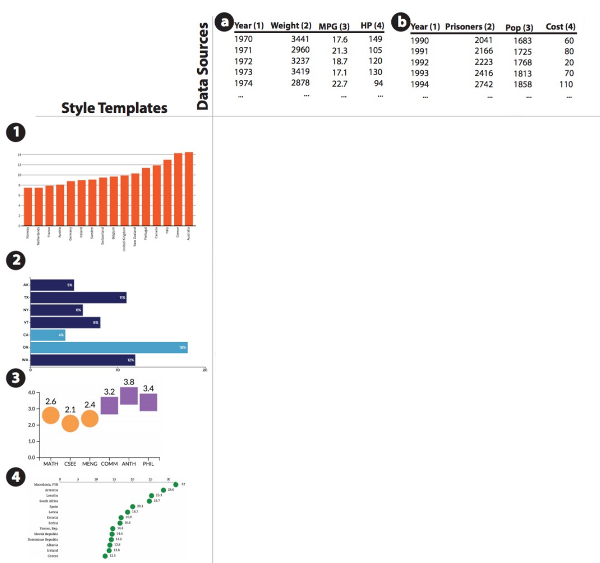




Converting Basic D3 Charts into Reusable Style Templates. Jonathan Harper and Maneesh Agrawala. IEEE TVCG. 2018.

79

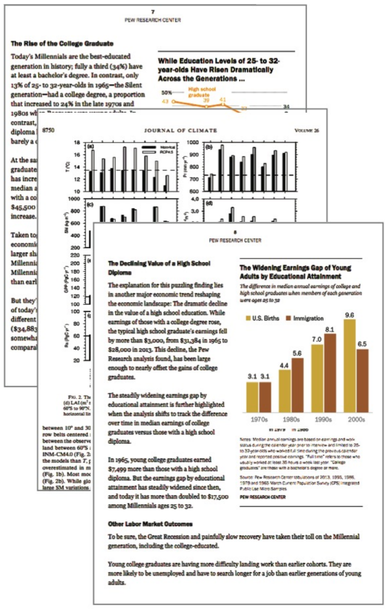
Reusable Style Templates



Converting Basic D3 Charts into Reusable Style Templates. Jonathan Harper and Maneesh Agrawala. IEEE TVCG. 2018.

80

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

81

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***

82