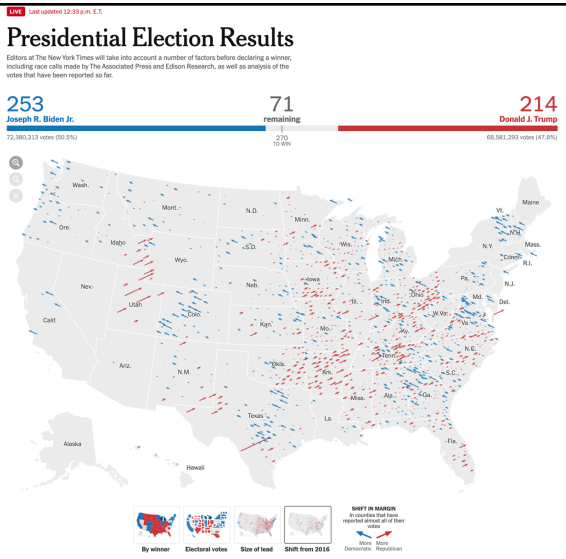
DECONSTRUCTING VISUALIZATIONS

CS 448B | Fall 2023

MANEESH AGRAWALA

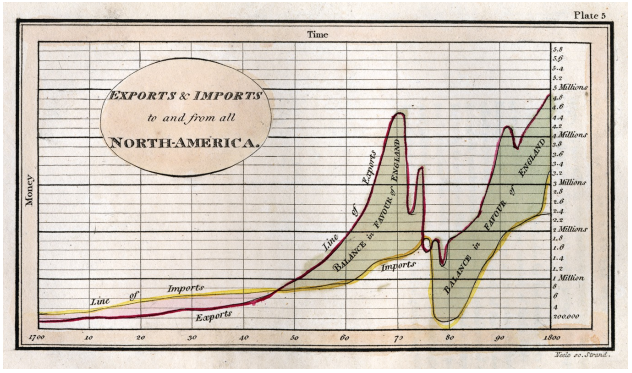
1



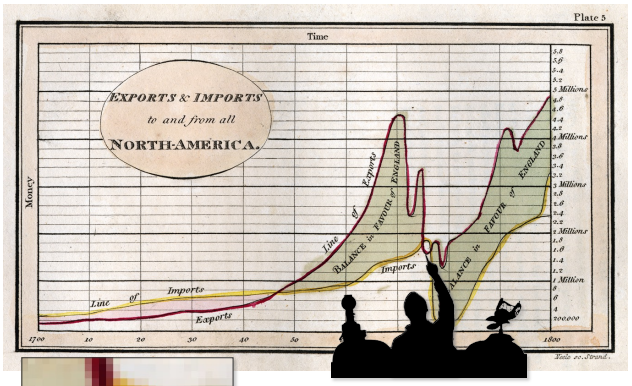
2

DECONSTRUCTING VISUALIZATIONS

3



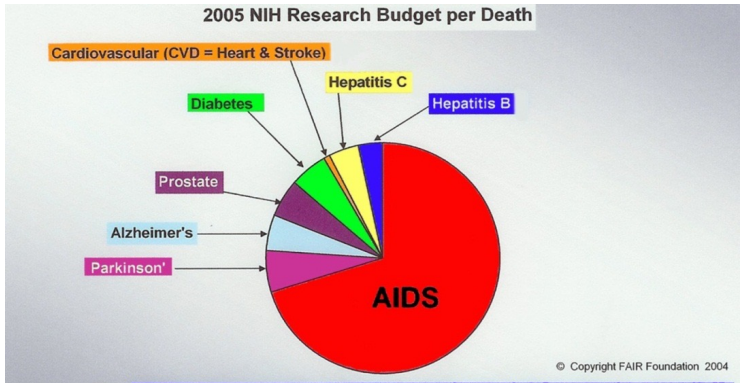
4



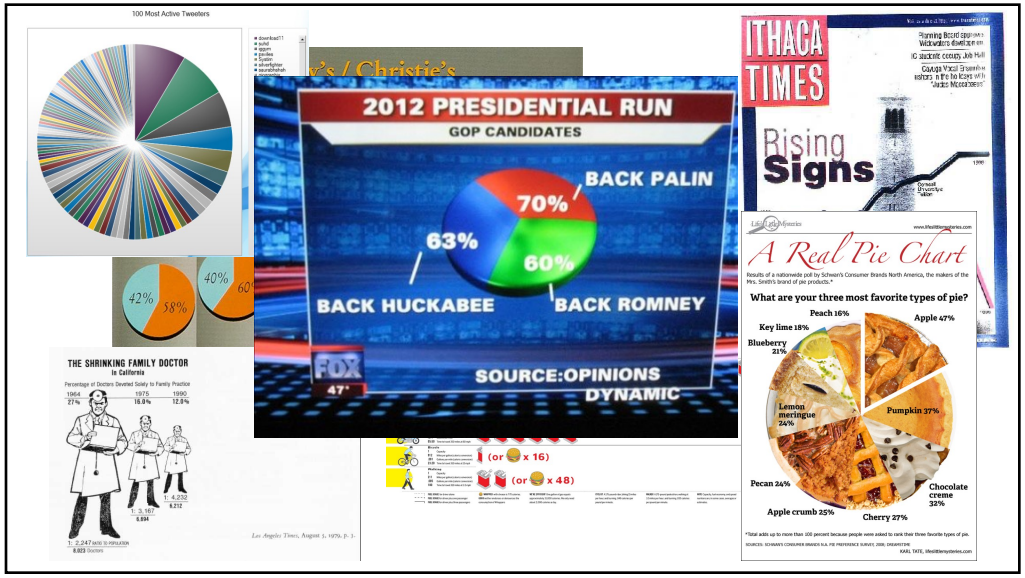
Pixels are poor representation

Hard for machines to retrieve data

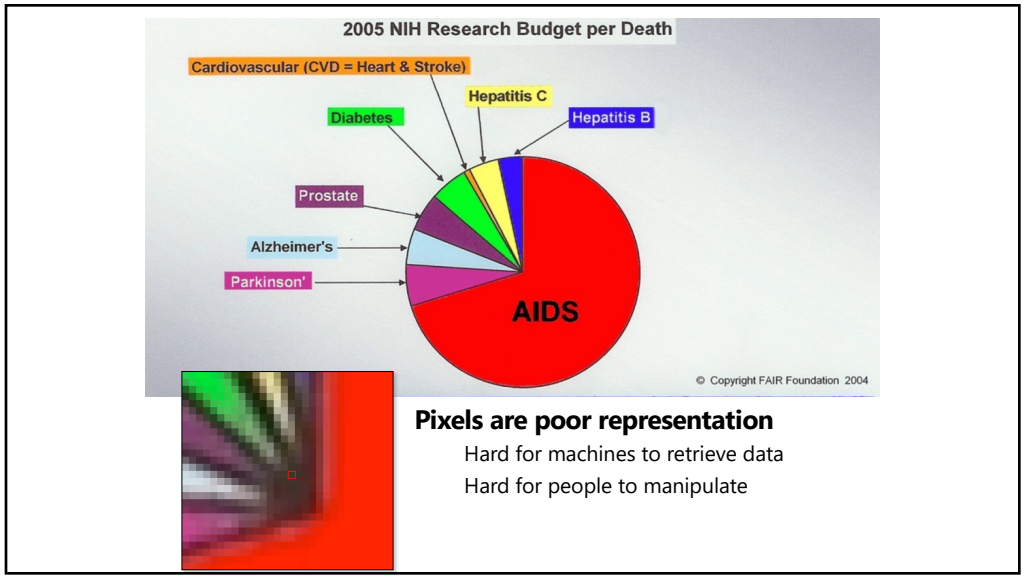
5



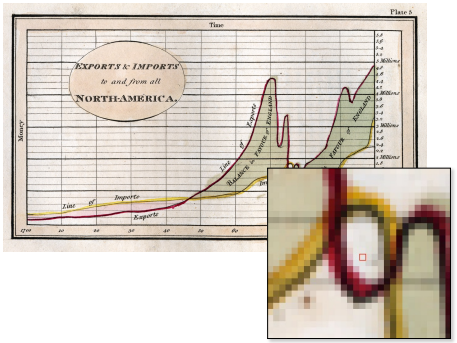
6

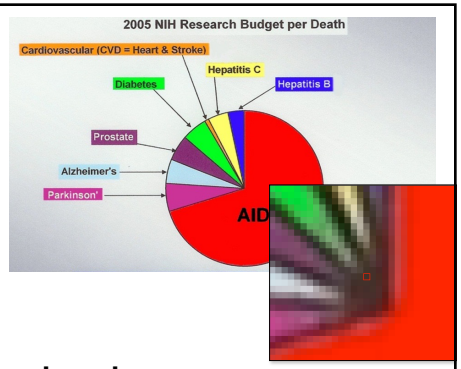

100 Most Active Tweeters
2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
FOX
47°
SOURCE:OPINIONS DYNAMIC
ITHACA TIMES
Rising Signs
A Real Pie Chart
What are your three most favorite types of pie?
Peach 16%
Apple 47%
Key lime 18%
Blueberry 21%
Lemon meringue 24%
Pumpkin 37%
Pecan 24%
Chocolate creme 32%
Apple crumb 25%
Cherry 27%
THE SHRINKING FAMILY DOCTOR
in California
42%
58%
40%
(or x 16)
(or x 48)

7


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

8





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*

9

TODAY

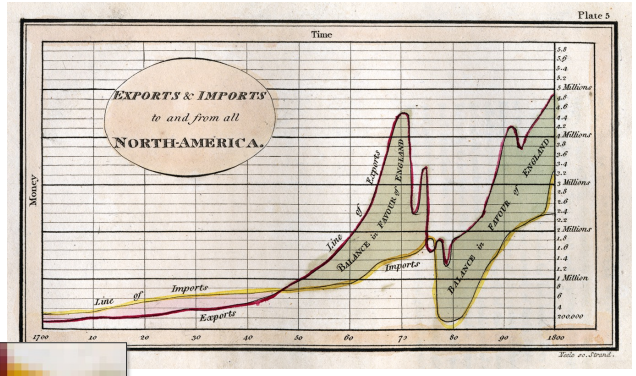
Learning Objectives

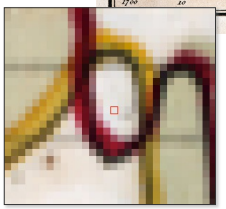
1. How to deconstruct charts and graphs into an editable representation
2. How to use this representation to support interactive reading of visualization

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
…	…	…

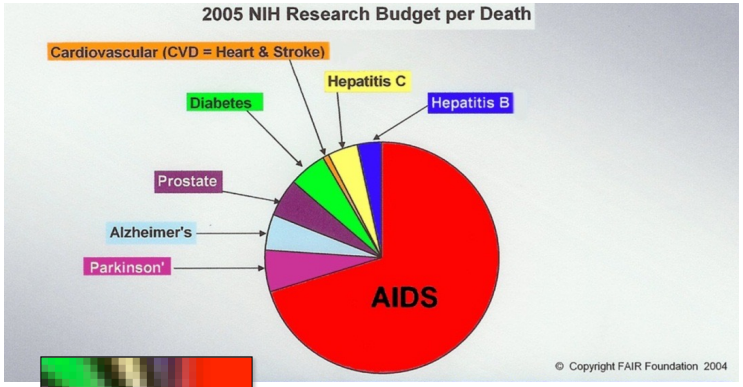
Marks

mark: lines

Mappings

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

12



Data

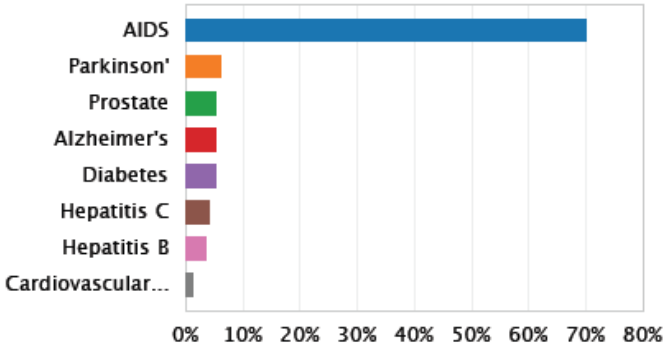
Disease	Budget
Aids	70.0%
Alzheimer's	5.0%
Cardiovascular	1.1%
Diabetes	4.8%
Hepatitus B	4.1%
Hepatitus C	3.8%
Parkinson'	6.0%
Prostate	5.2%

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

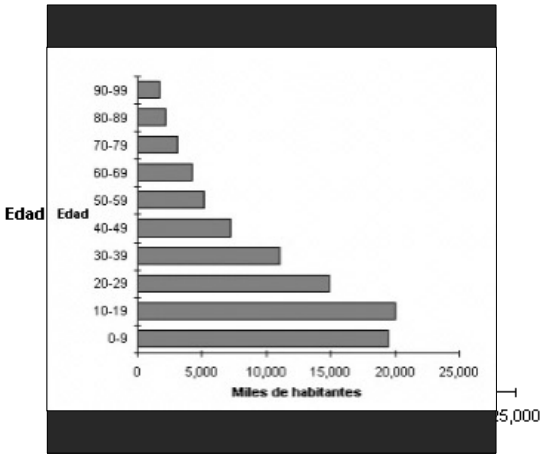
Classification: Determine chart type
Mark extraction: Retrieve graphical marks
Data extraction: Retrieve underlying data values

15

CLASSIFICATION OF CHART TYPE

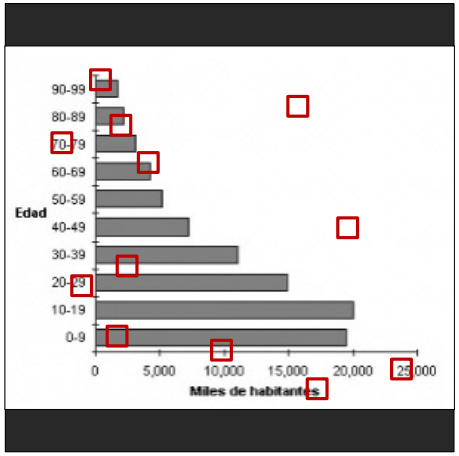
16

TRAINING



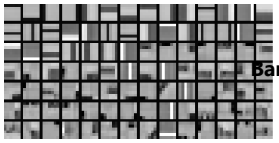
17

TRAINING

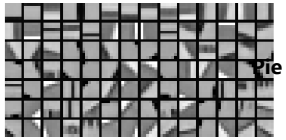


18

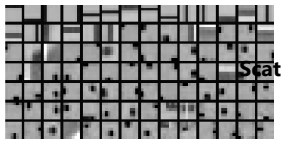
TRAINING



Bar Charts



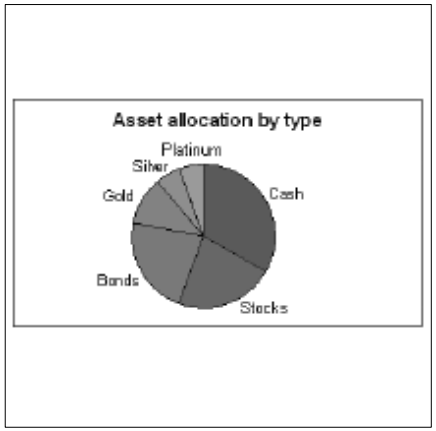
Pie Charts



Scatter Plots

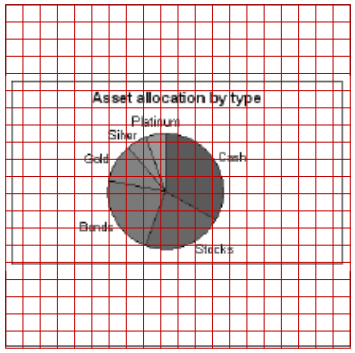
19

CLASSIFICATION



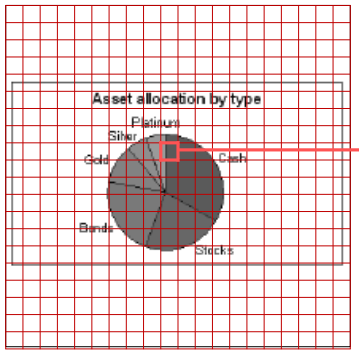
20

CLASSIFICATION



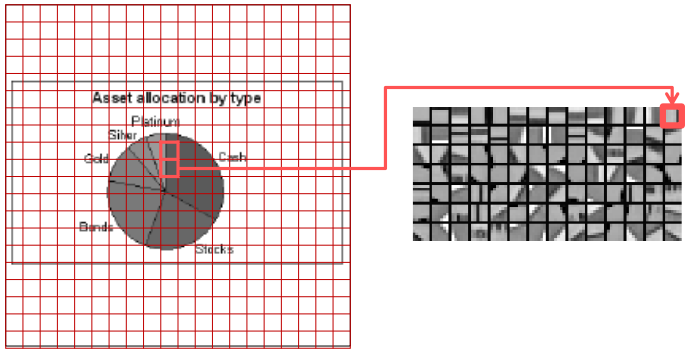
21

CLASSIFICATION



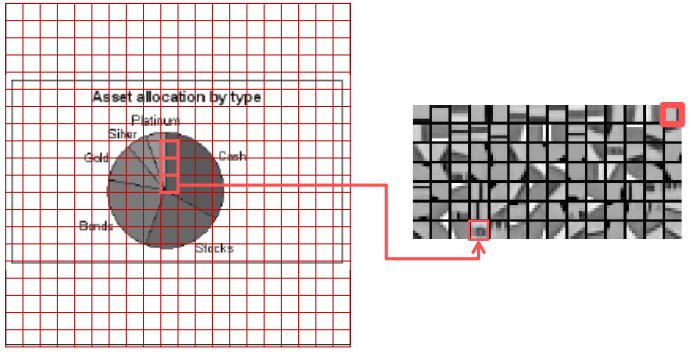
22

CLASSIFICATION



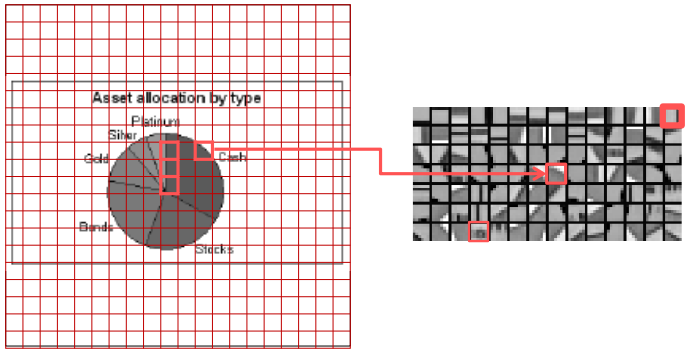
23

CLASSIFICATION



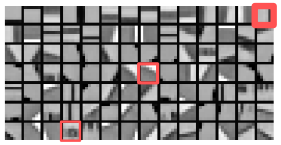
24

CLASSIFICATION



25

CLASSIFICATION



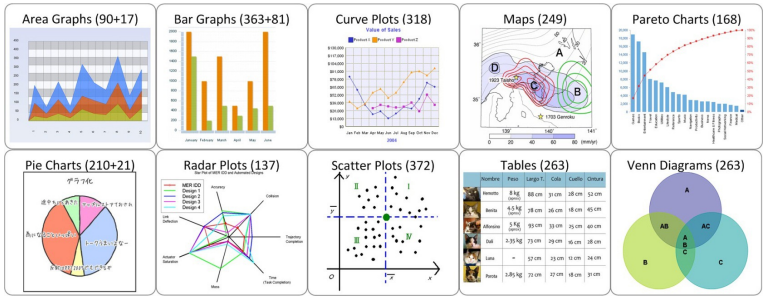
SVM Classifier → Pie Chart

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

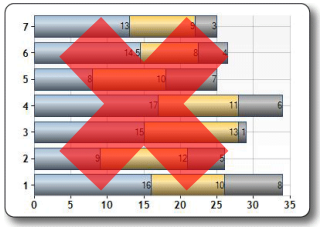
MARK AND DATA EXTRACTION

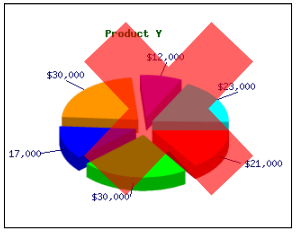
28

ASSUMPTIONS

Bar charts and pie charts only

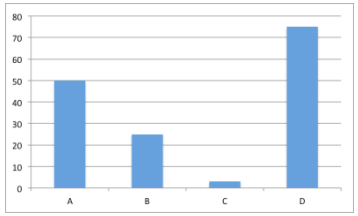
No shading or texture, 3D, stacked bars, or exploded pies





29

BAR CHARTS



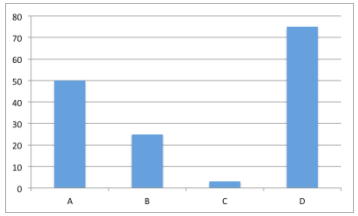
→

marks: lines

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

30

BAR CHARTS



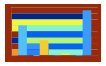
marks: lines

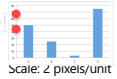
y-value	x-value
50	A
35	B
4	C
75	D

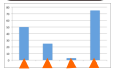
Extract Marks

Extract Data

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

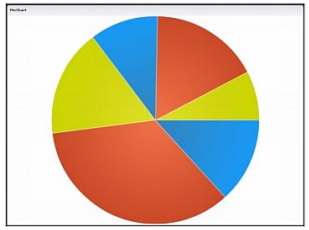






41

PIE CHARTS



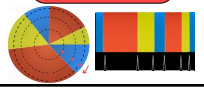
marks: areas

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

Extract Marks

Extract Data

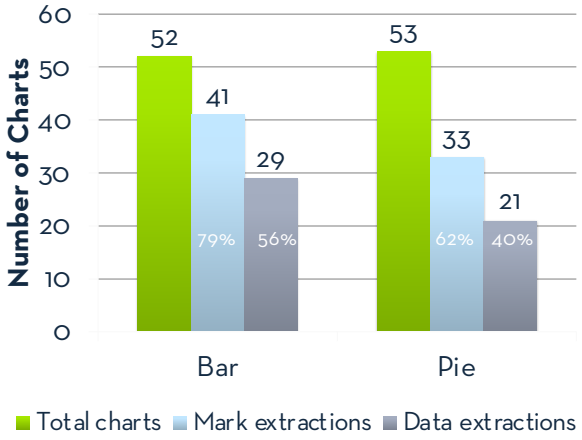
Fit Ellipse Using RNASAC → Unroll Pie and Find Transitions → Compute Area Percentages → Associate Labels with Areas



Scale: 50 pixels/percent

42

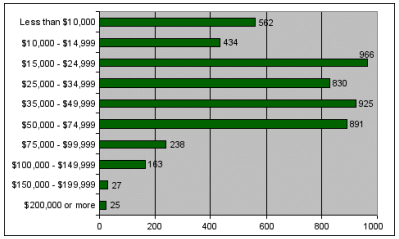
EXTRACTION RESULTS



43

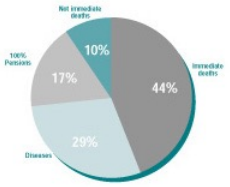
DATA EXTRACTION ERROR

Bar Charts



7.7%

Pie Charts



4.6%

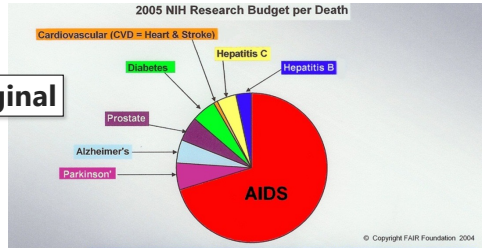
Average chart size: 342 x 452 pixels [Prasad 2007]

44

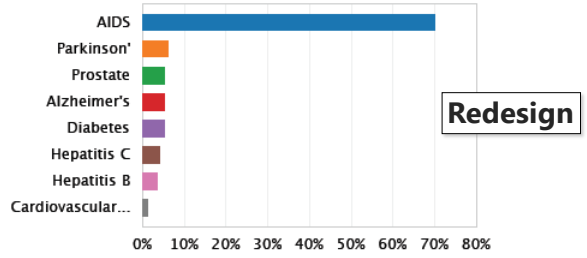
REDESIGN

45

Original

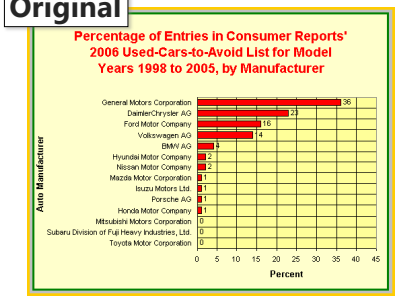


Redesign

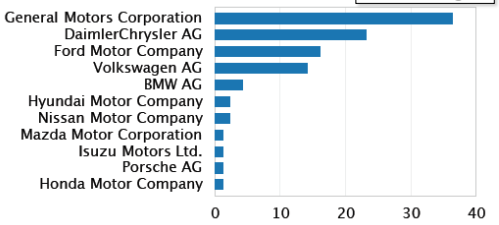


46

Original



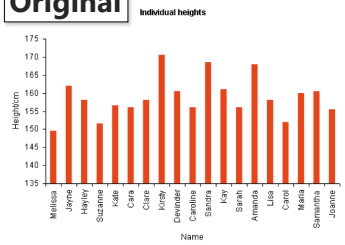
Redesign



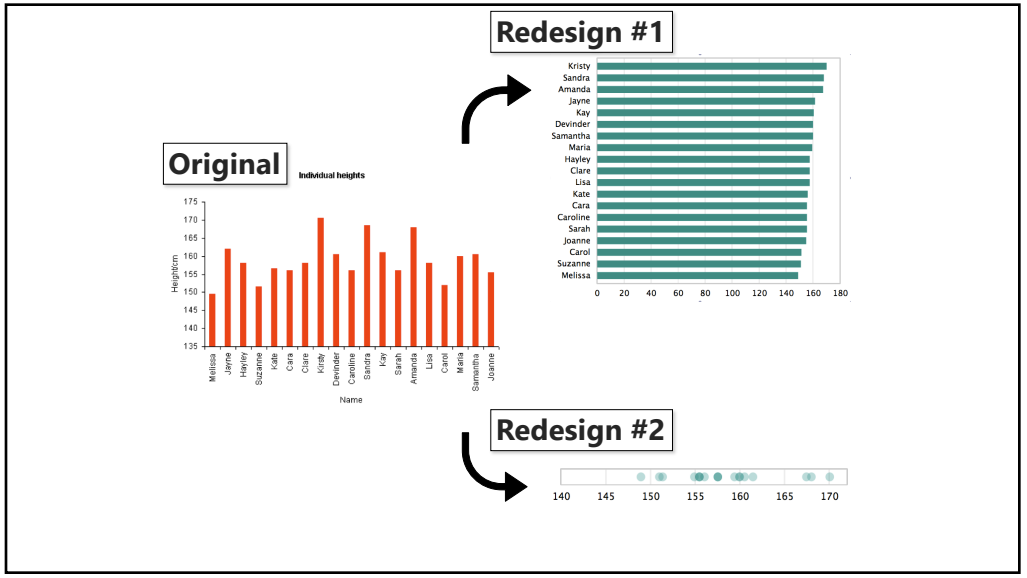
47

Redesign #1

Original



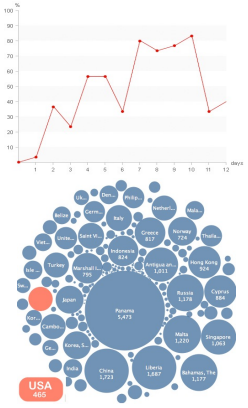
48



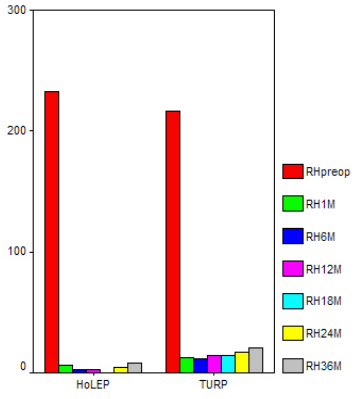
49

LIMITATIONS

Additional Chart Types



Handling Legends



51

ANNOUNCEMENTS

52

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

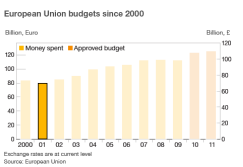
Grading

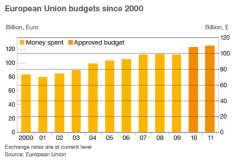
Groups of up to 3 people, graded individually
Clearly report responsibilities of each member

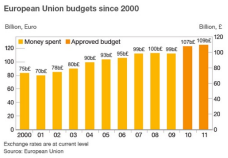
53

INTERACTIVE READING

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.

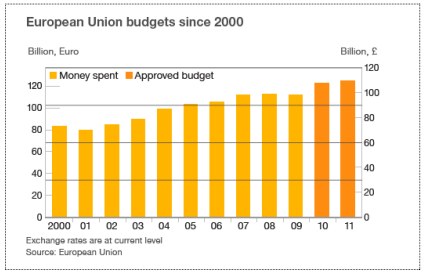


Chart type: Bar
Chart: 00281
Overlay type: Reference structures
Regular gridlines
Lines emanating from marks

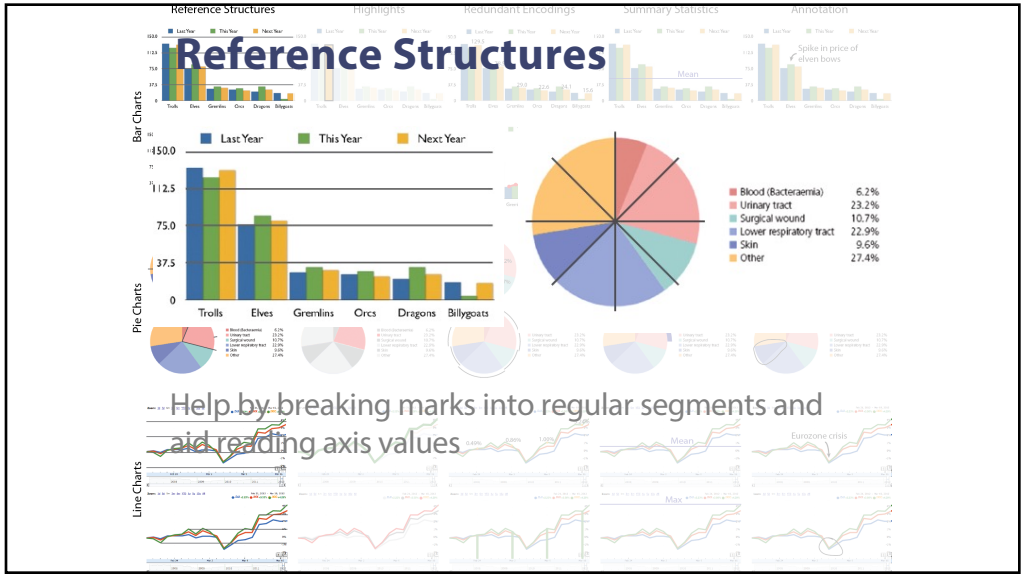
Parameters

Overlay Underlay
Static Interactive
Divisions: 4
Line thickness: 1

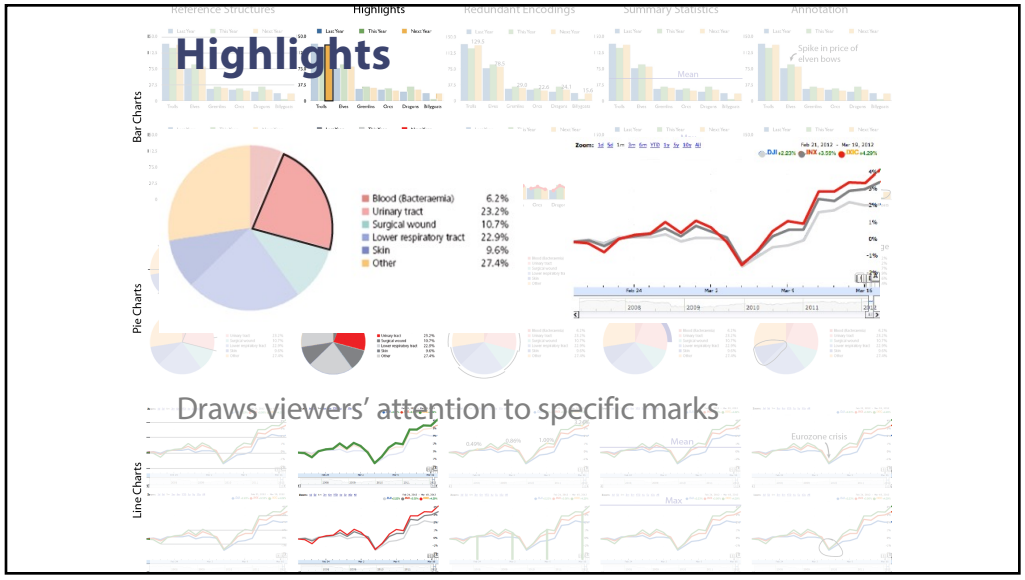
Places regular gridlines at user defined intervals.

Demo

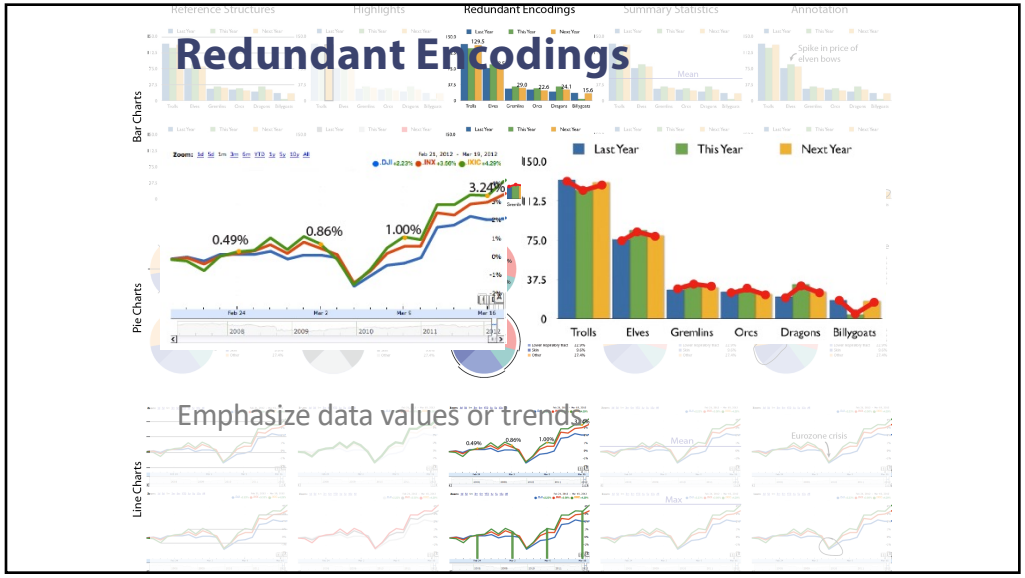
57


Reference Structures
Highlights
Redundant Encodings
Summary Statistics
Annotation
Reference Structures
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
Blood (Bacteraemia) 6.2%
Urinary tract 23.2%
Surgical wound 10.7%
Lower respiratory tract 22.9%
Skin 9.6%
Other 27.4%
Help by breaking marks into regular segments and aid reading axis values
Mean
Max
Eurozone crisis

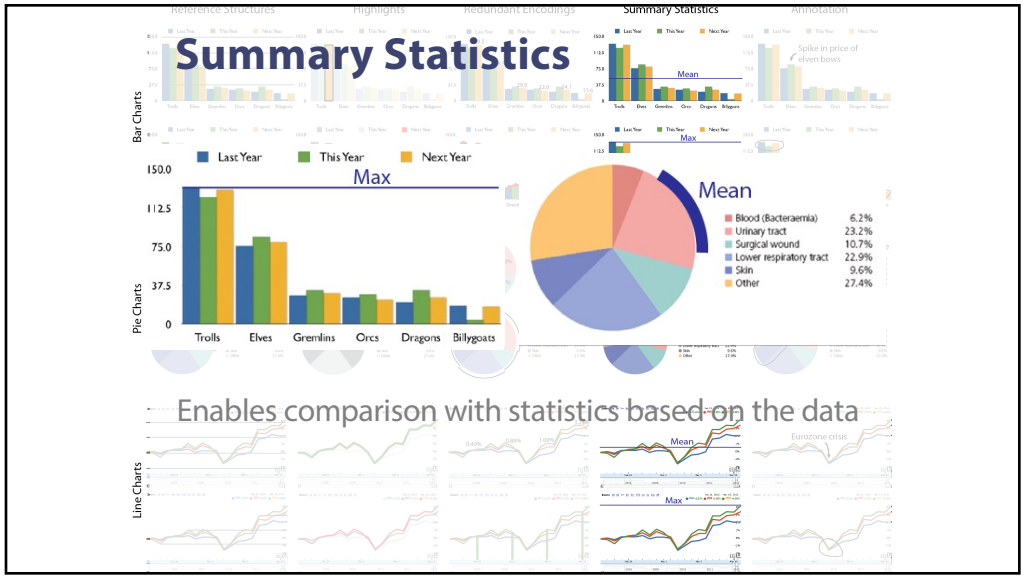
58


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%
Draws viewers' attention to specific marks
Mean
Max
Eurozone crisis

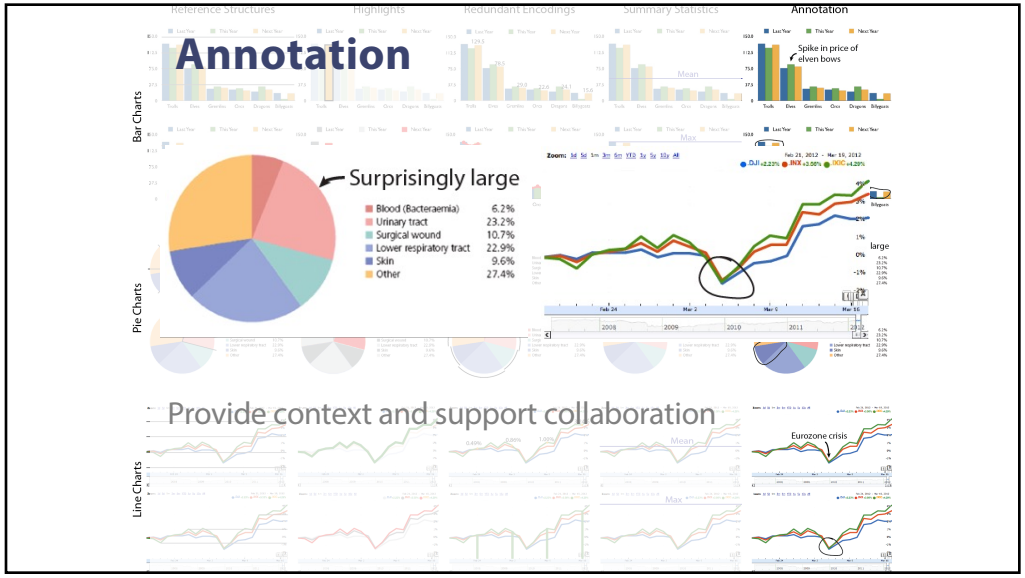
59


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

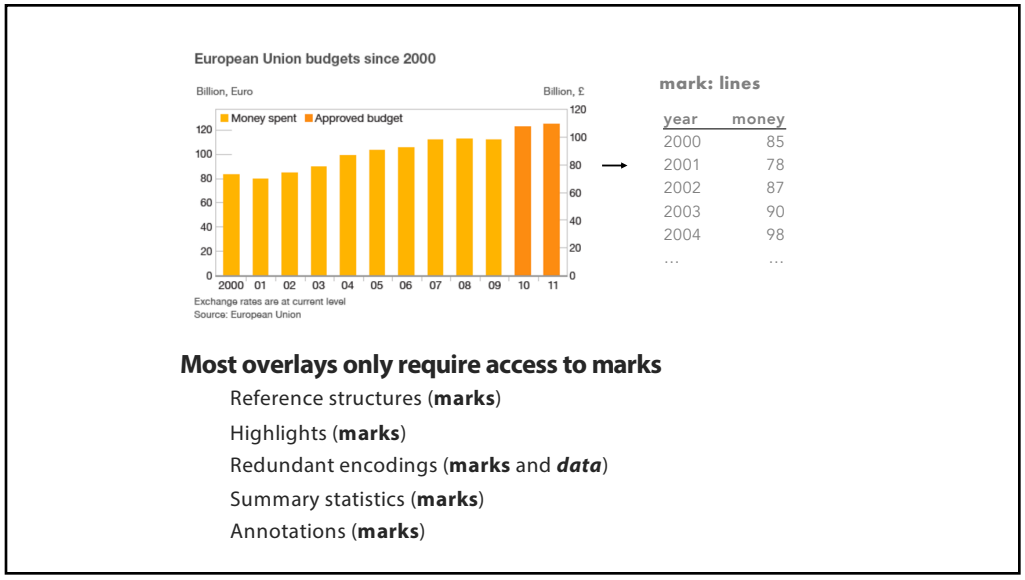
60


Summary Statistics
Bar Charts
Pie Charts
Line Charts
Last Year
This Year
Next Year
Max
Mean
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%
Enables comparison with statistics based on the data

61



62



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

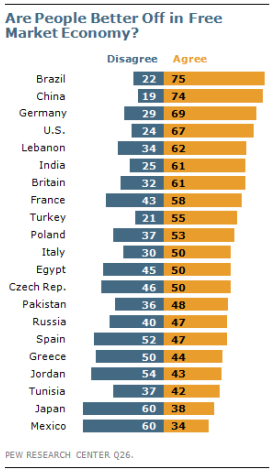
Goal: Extract references between text and chart

65

Problem: Diversity of writing styles

66

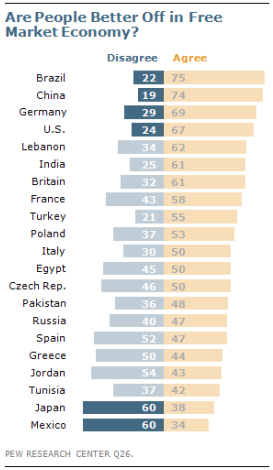
EXAMPLE 1: PEW RESEARCH



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

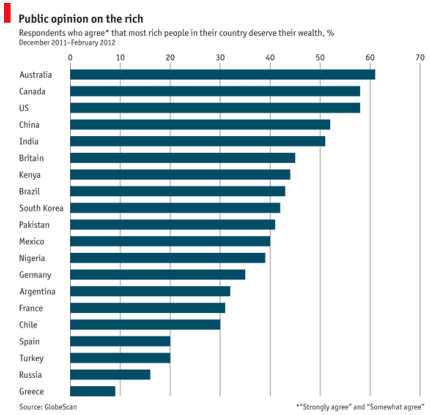
EXAMPLE 1: PEW RESEARCH



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

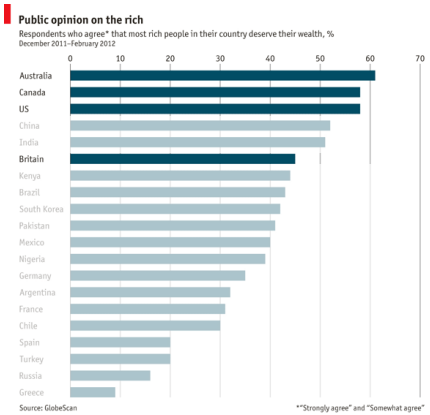
EXAMPLE 2: ECONOMIST



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.

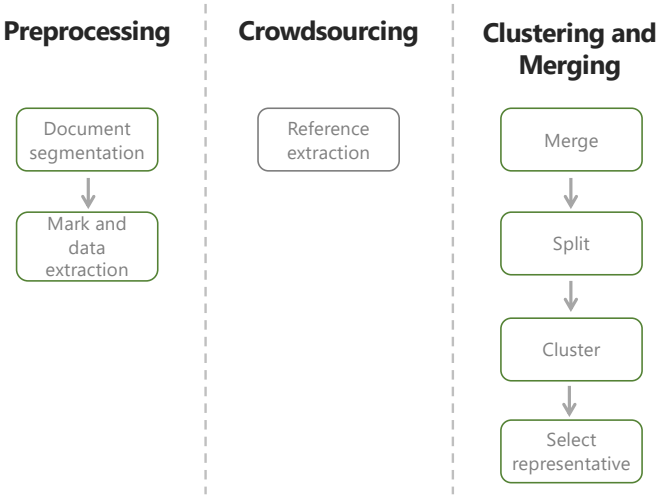
69

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

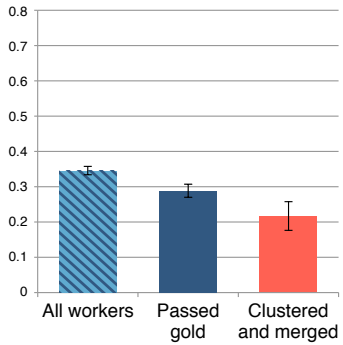


71

DEMO

72

EVALUATION



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

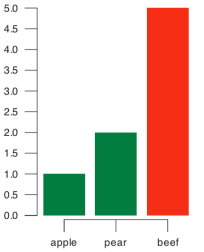
73

DECONSTRUCTING D3 CHARTS

```
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

deconI D	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*

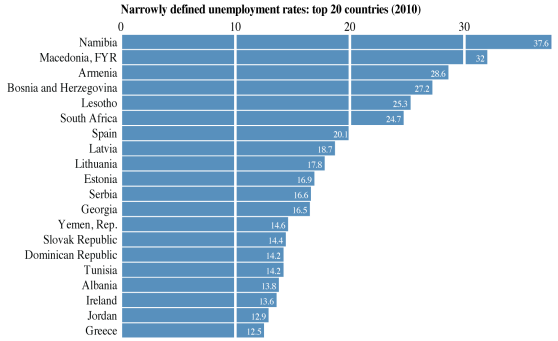
D3 Code **D3 Chart** **Our Deconstruction**

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.

75

DECONSTRUCTING D3 CHARTS



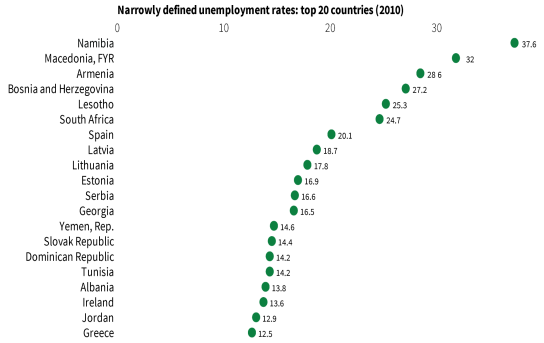
country	rate	deconID
Namibia	37.6	17
Macedonia, FYR	32.0	21
Armenia	28.6	25
Bosnia and Herzegovina	27.2	29
Lesotho	25.3	33
South Africa	24.7	37
Spain	20.1	41
Latvia	18.7	45
...	...	...

rate → *width*
rate → *area*
rate → *xPos*
deconID → *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 ↦ *xPos*
deconID ↦ *yPos*

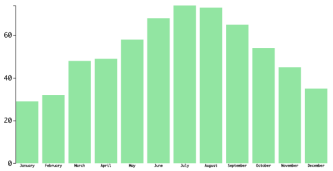
Deconstructing and Restyling D3 Visualizations. Jonathan Harper and Maneesh Agrawala. User Interface Software Technology (UIST) 2014.

77

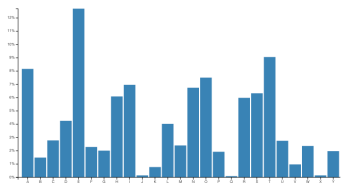
AUTOMATIC REDESIGN

Can we automatically redesign charts to improve

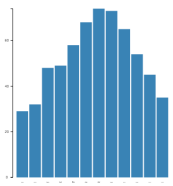
Perceptual effectiveness?
Visual aesthetics?
Accessibility for vision impaired users?



Data Source



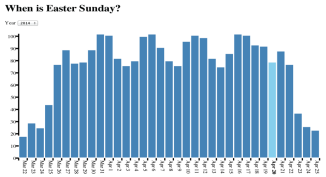
Style Target

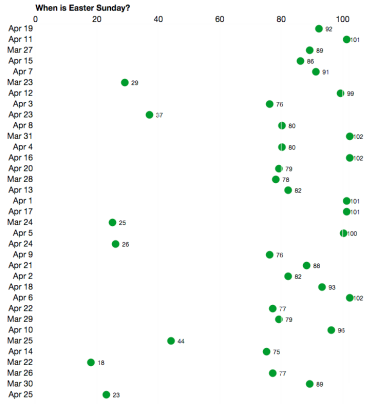


Result

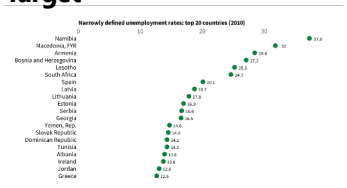
78

Data Source





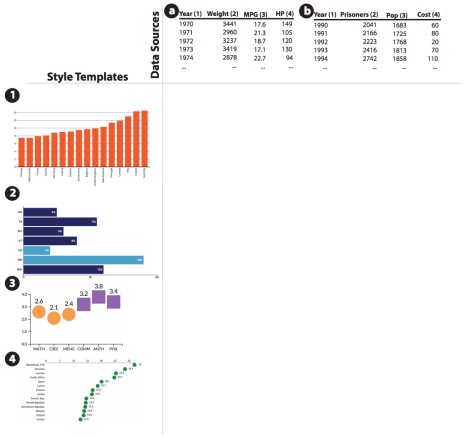
Style Target



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

79

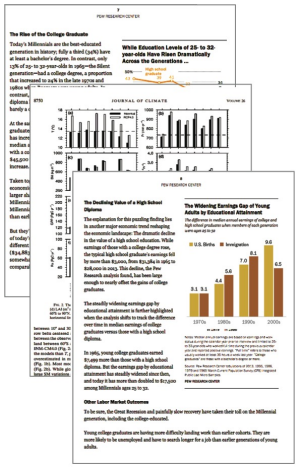
REUSABLE STYLE TEMPLATES



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

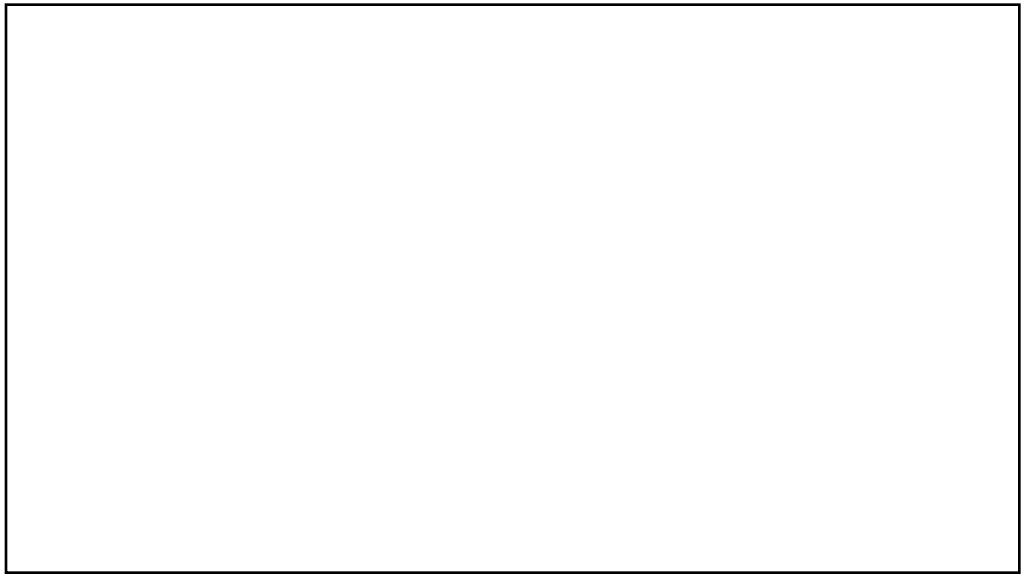
TAKEAWAYS

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



83