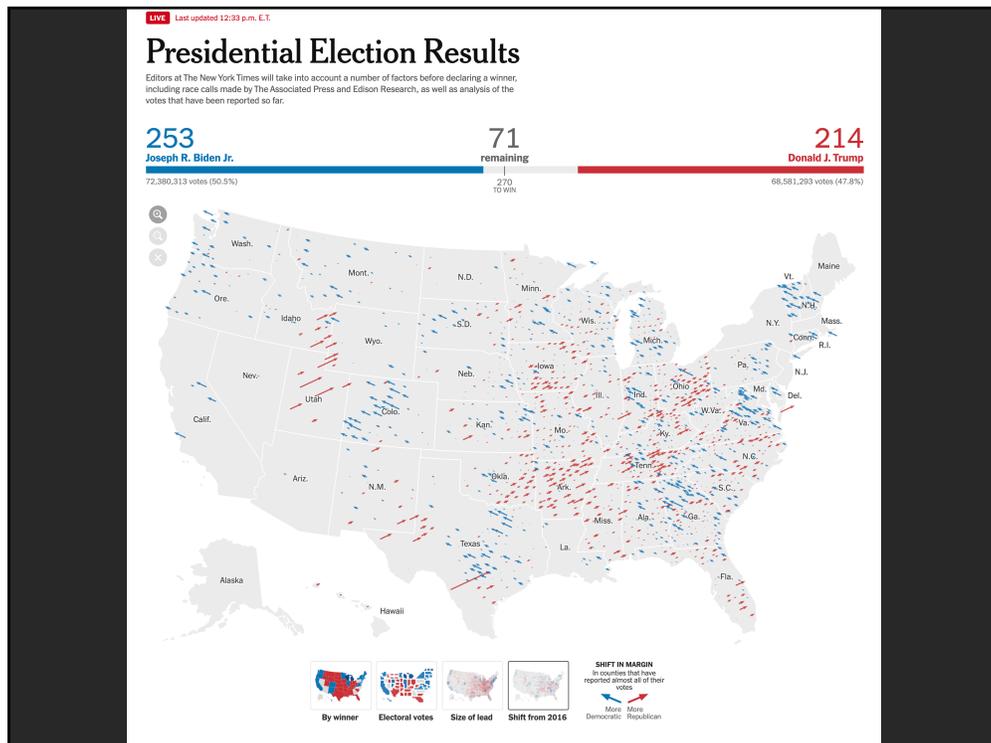


Deconstructing Visualizations

Maneesh Agrawala

CS 448B: Visualization
Fall 2020

1



2

Announcements

3

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 interactive visualizations
- **Research:** Implementation of solution and web-based demo if possible
- **Short video (2 min)** demoing and explaining the project

Schedule

- Project proposal: **Thu 10/29**
- Design Review and Feedback: **Tue 11/17 & Thu 11/19**
- Final code and video: **Sat 11/21 11:59pm**

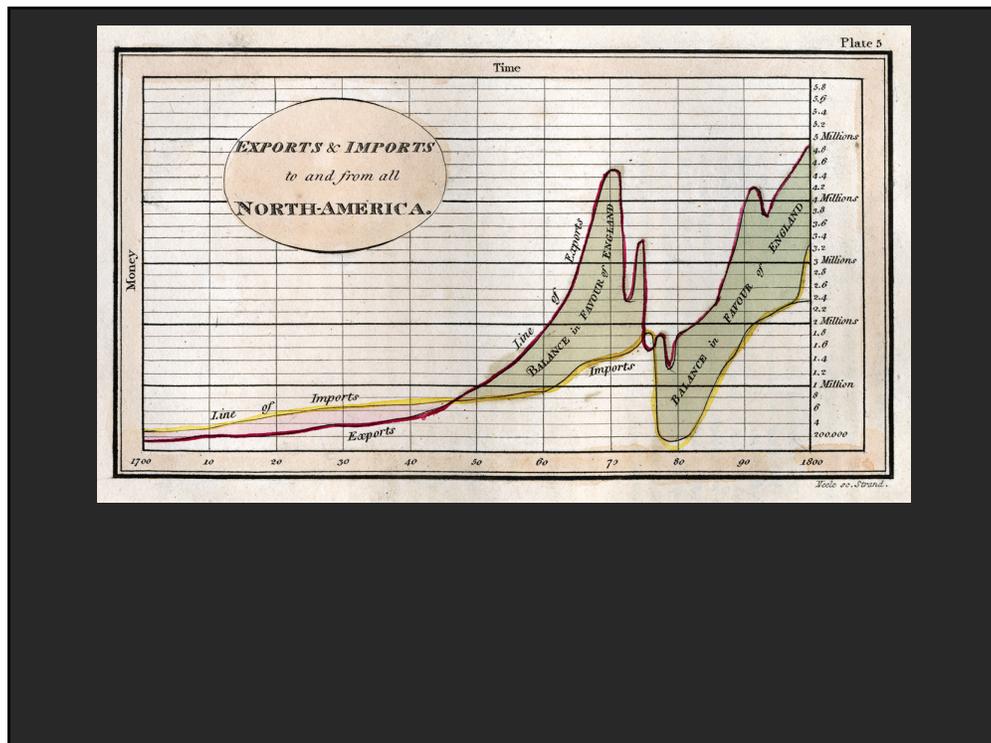
Grading

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

4

Deconstructing Visualizations

5



6

100 Most Active Tweeters

- download11
- andri
- lupin
- gillies
- skewfisher
- saubothah
- prognosis
- demarhah
- gabrielmiller
- amirad
- lucacab

Christie's Sales

2012 PRESIDENTIAL RUN GOP CANDIDATES

Candidate	Percentage
BACK PALIN	70%
BACK ROMNEY	60%
BACK HUCKABEE	63%
FOX	47%

Pie Chart

Consumer Brands North America, the makers of the most favorite types of pie?

Pie Type	Percentage
Apple	47%
Pumpkin	37%
Chocolate creme	32%
Cherry	27%
Apple crumb	25%
Pecan	24%

THE SHRINKING FAMILY in California

Percentage of Doctors Devised Safety

Year	Percentage
1964	27%
1975	18.6%

1: 2,247 (vs) 1964/1975
8,823 (vs) 1975

1: 3,167
8,823

1: 4,232
8,823

Los Angeles Times, August 3, 1975, p. 3.

(or x 16)

(or x 48)

*Total adds up to more than 100 percent because people were asked to rank their three favorite types of pie.
SOURCES: SHOWN'S CONSUMER BRANDS N.A. PREFERENCE SURVEY 2006; DEPARTMENT OF HEALTH & HUMAN SERVICES; KARL TATE, HealthMatters.com

9

2005 NIH Research Budget per Death

Disease	Relative Budget
AIDS	Highest
Cardiovascular (CVD = Heart & Stroke)	Second Highest
Diabetes	Third Highest
Hepatitis C	Fourth Highest
Hepatitis B	Fifth Highest
Prostate	Sixth Highest
Alzheimer's	Seventh Highest
Parkinson's	Eighth Highest

© Copyright FAIR Foundation 2004

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

10

Exports & Imports to and from all NORTH-AMERICA

2005 NIH Research Budget per Death

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

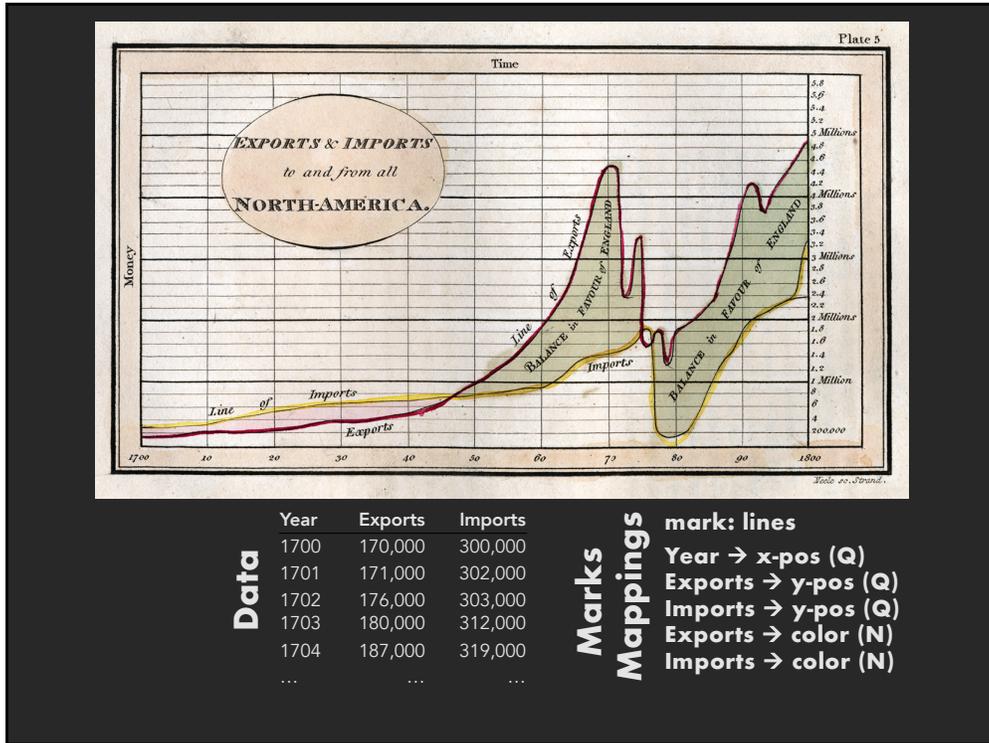
Pixels are a poor representation of charts and graphs
 Cannot index, search, manipulate or interact with the data

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

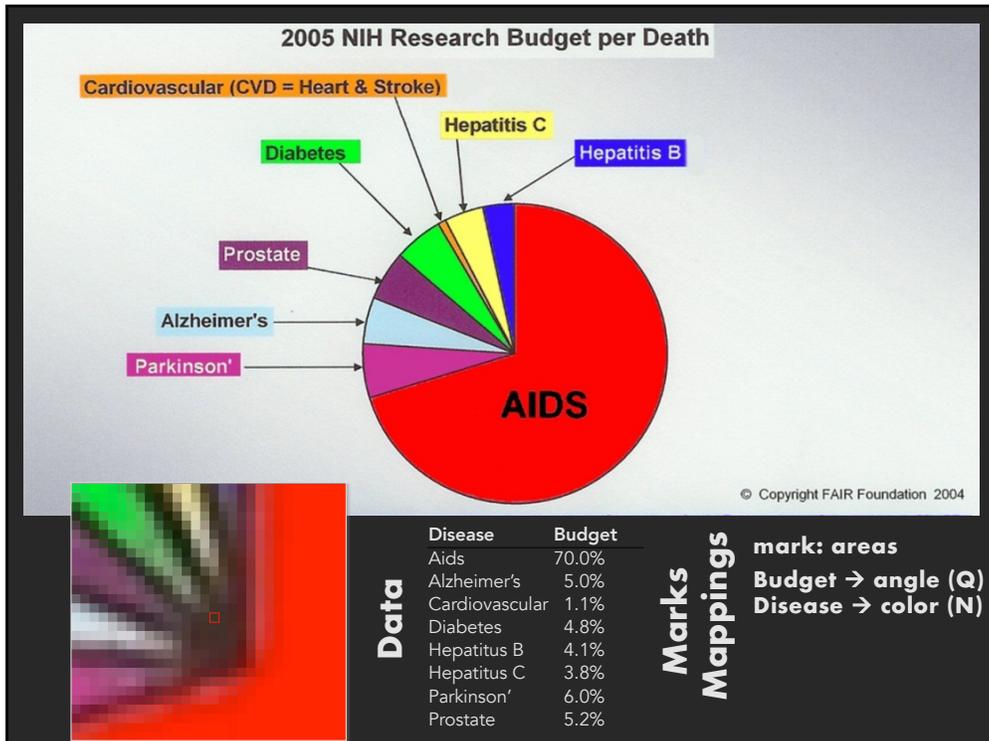
11

What is a good representation?

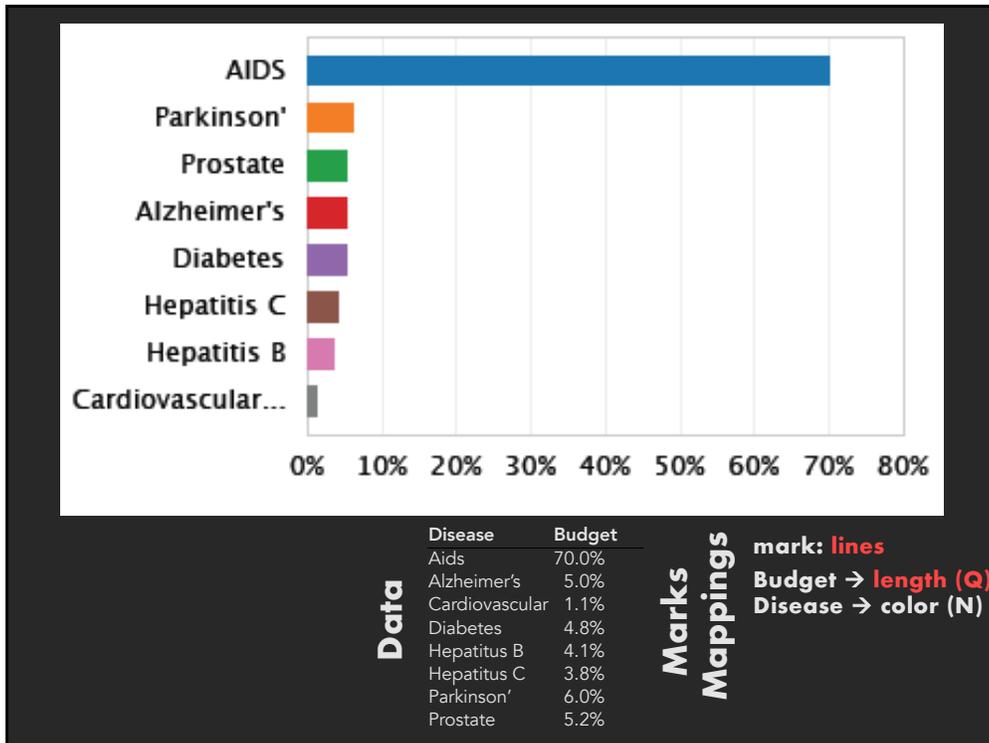
12



13



14



15

Approach

Classification: Determine chart type

Mark extraction: Retrieve graphical marks

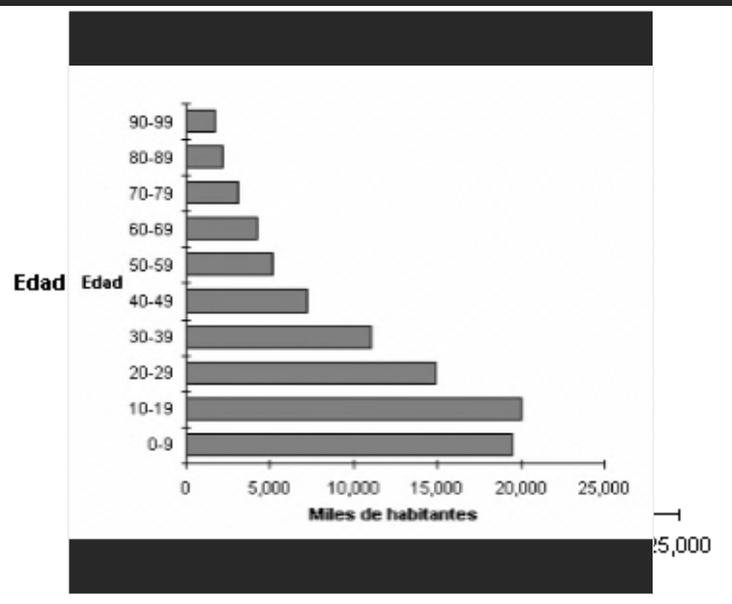
Data extraction: Retrieve underlying data table

16

Classification

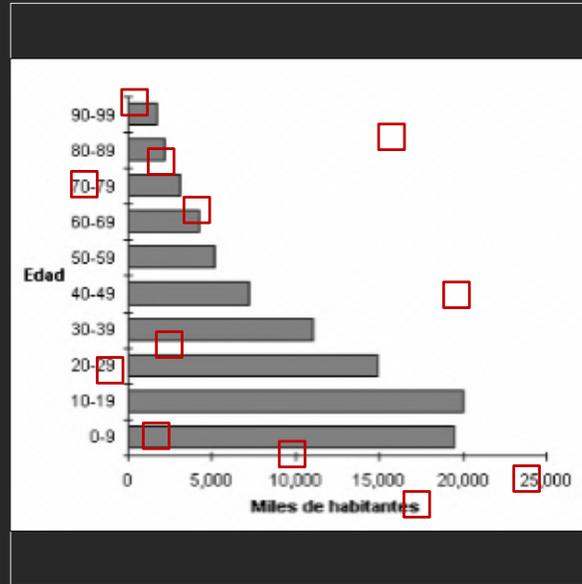
17

Training the Classifier



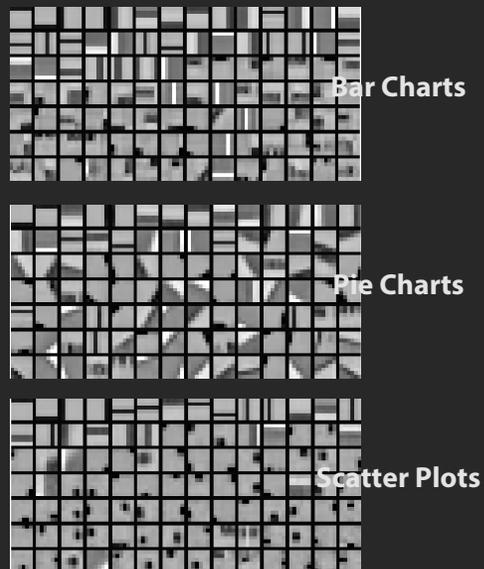
18

Training the Classifier



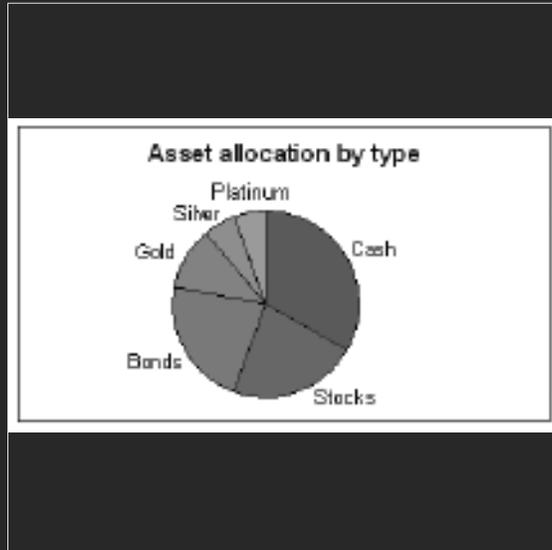
19

Training the Classifier



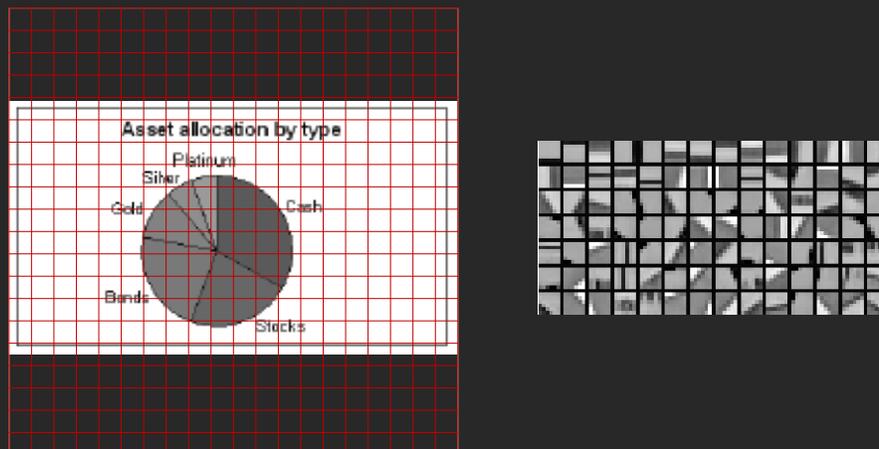
20

Classifying an Input Image



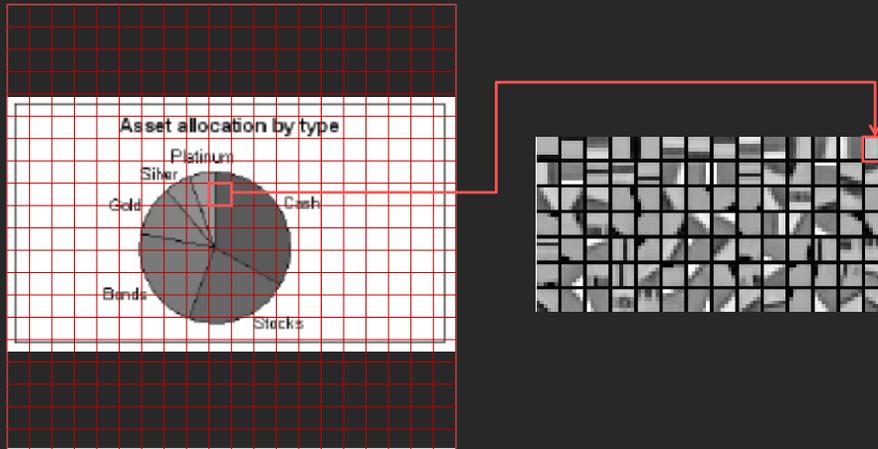
21

Classifying an Input Image



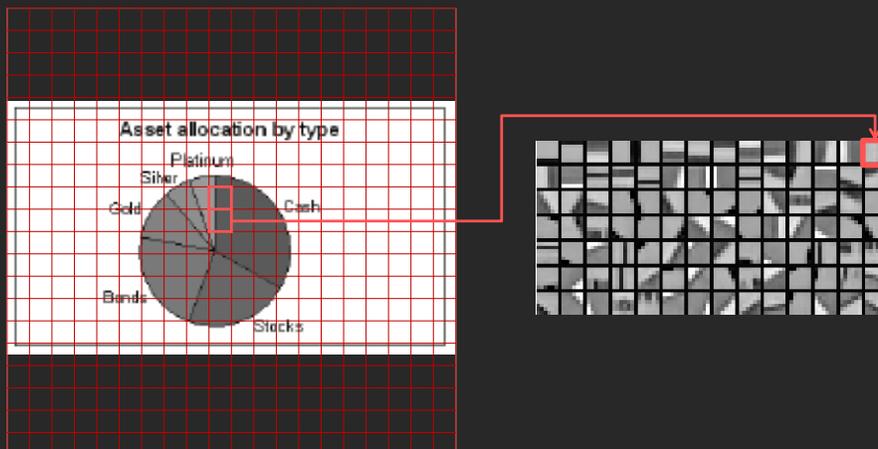
22

Classifying an Input Image



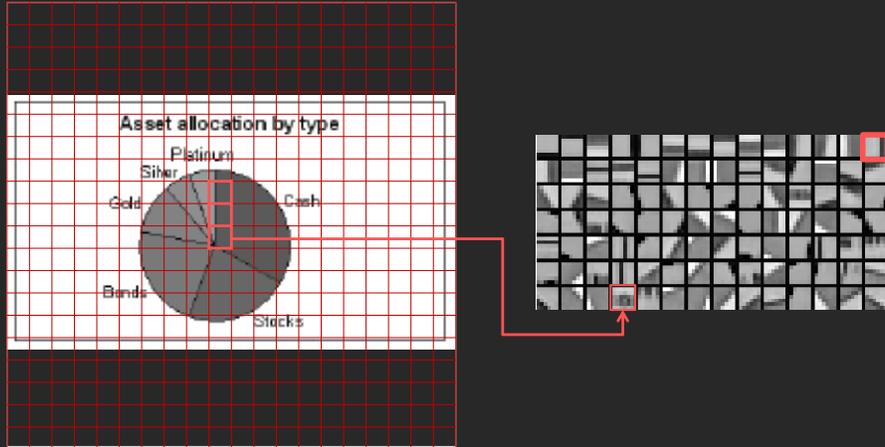
23

Classifying an Input Image



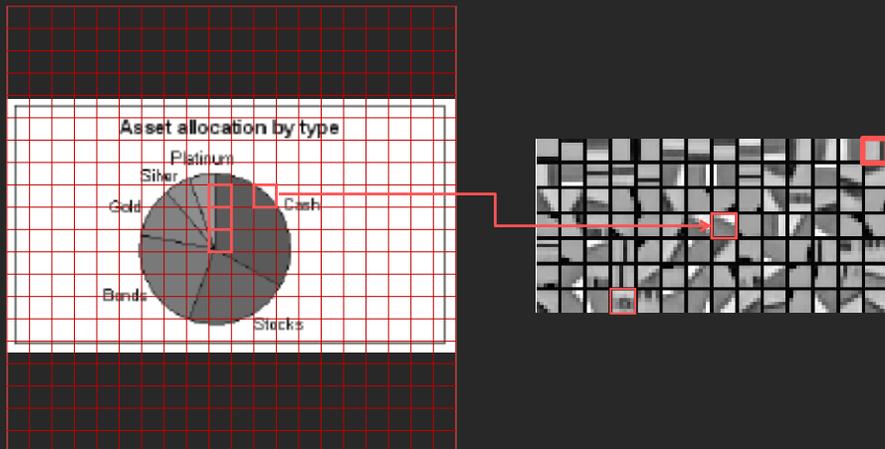
24

Classifying an Input Image



25

Classifying an Input Image



26

Classifying an Input Image



SVM Classifier



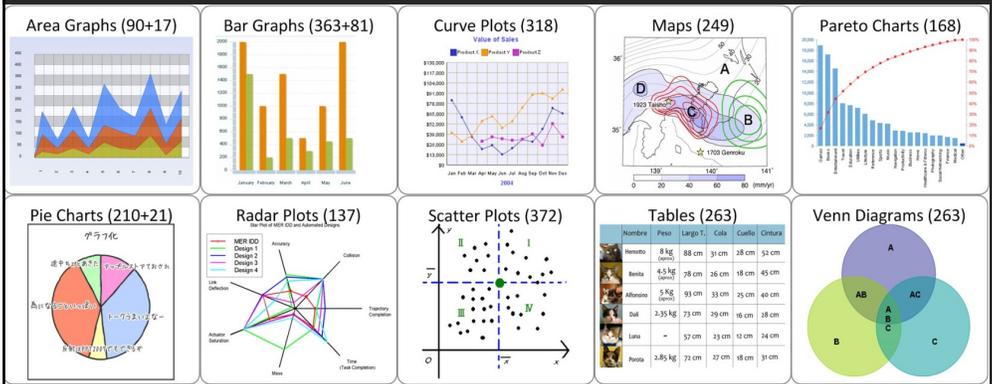
Pie Chart

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%

27

Our Corpus

Over 2500 labeled images and 10 chart types



ReVision binary SVMs give 96% classification accuracy

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

28

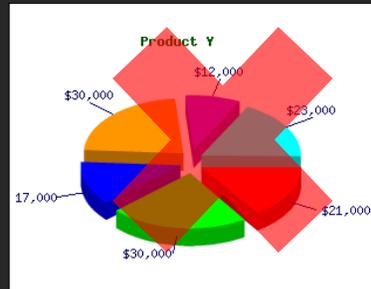
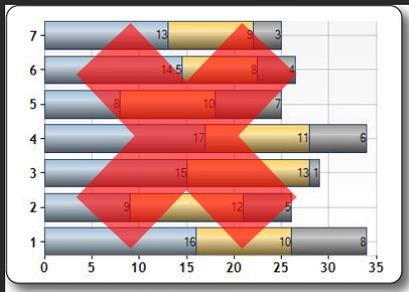
Mark and Data Extraction

29

Assumptions

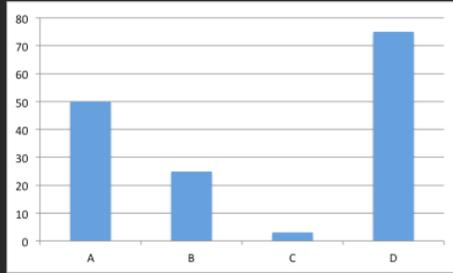
Bar charts and pie charts only

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

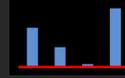


30

Bar Charts



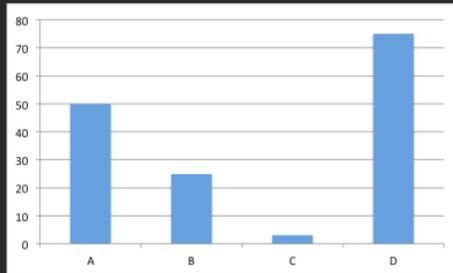
marks: lines



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

31

Bar Charts



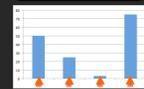
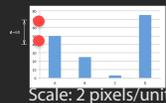
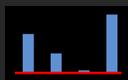
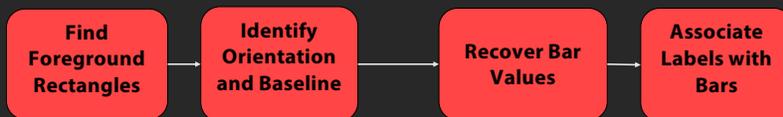
marks: lines



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

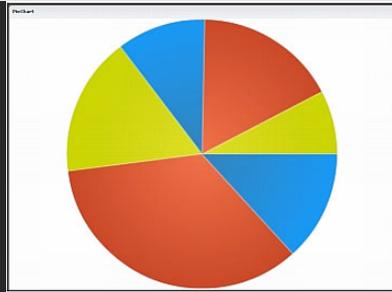
Extract Marks

Extract Data



42

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

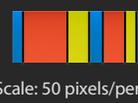
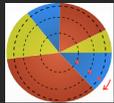
Extract Data

Fit Ellipse
Using
RNASAC

Unroll Pie
and Find
Transitions

Compute
Area
Percentages

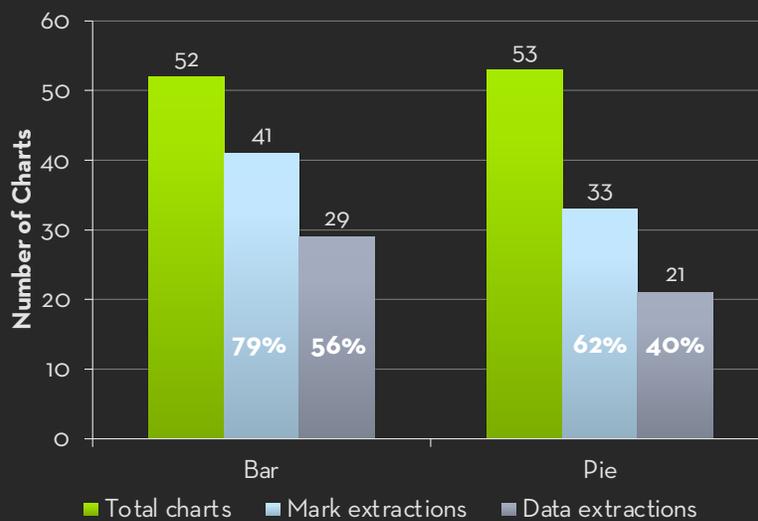
Associate
Labels with
Areas



Scale: 50 pixels/percent

43

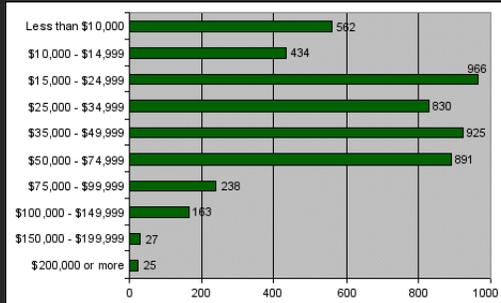
Extraction Results



44

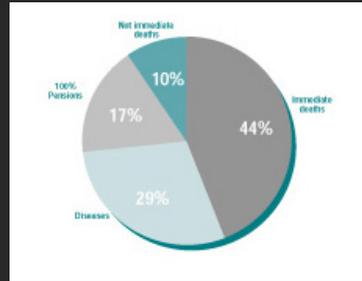
Data Extraction Error

Bar Charts



7.7%

Pie Charts



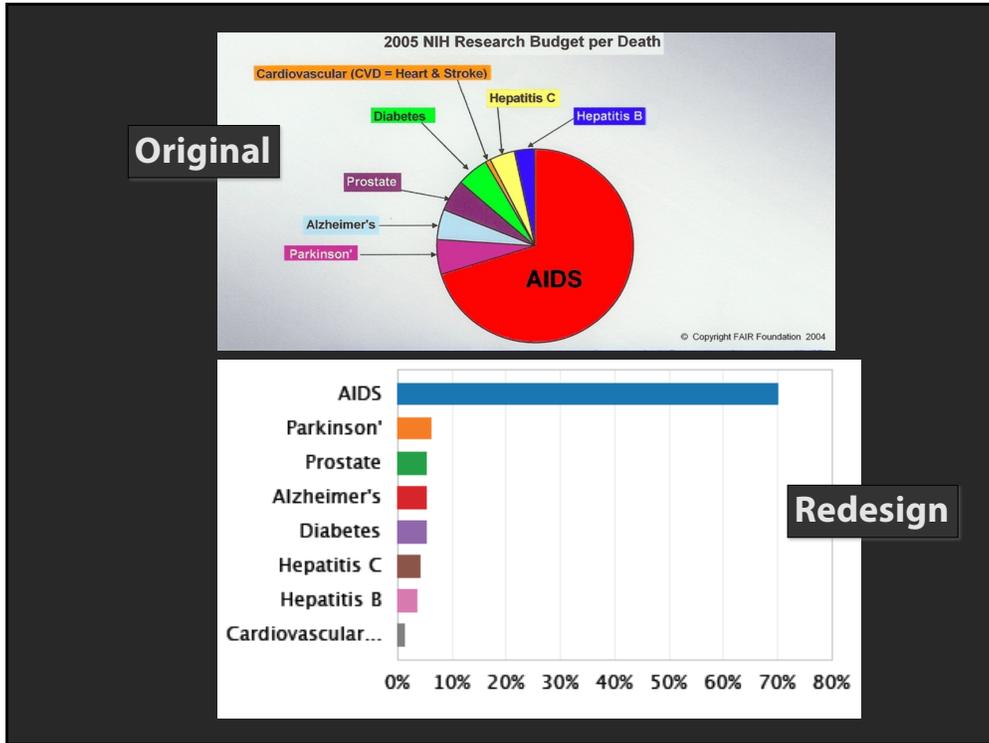
4.6%

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

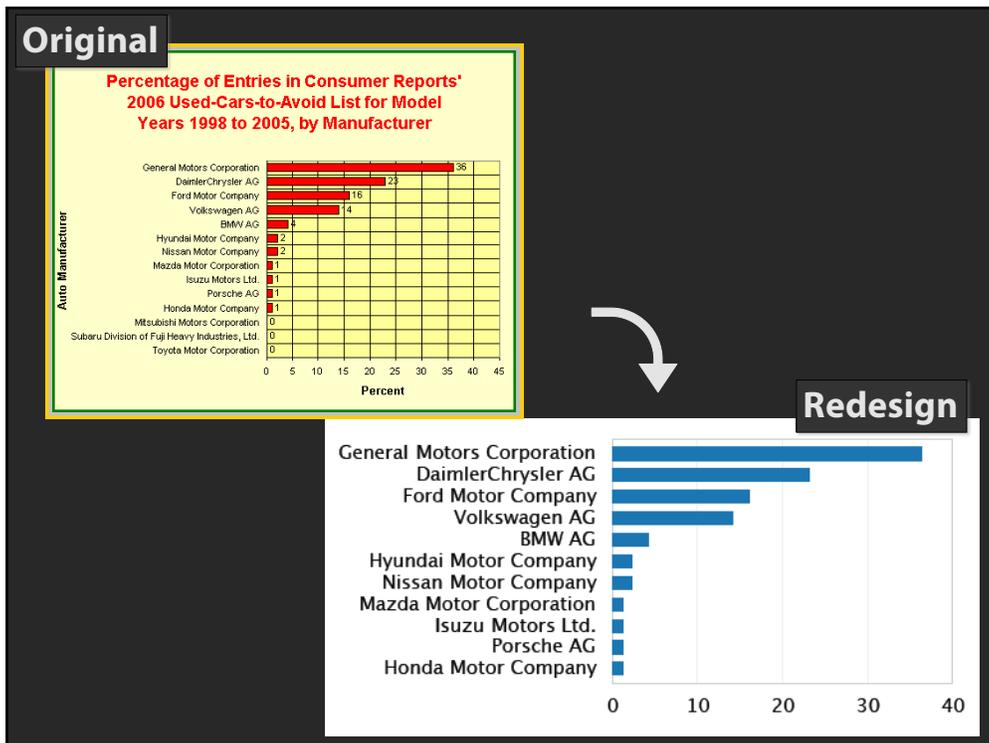
45

Redesign

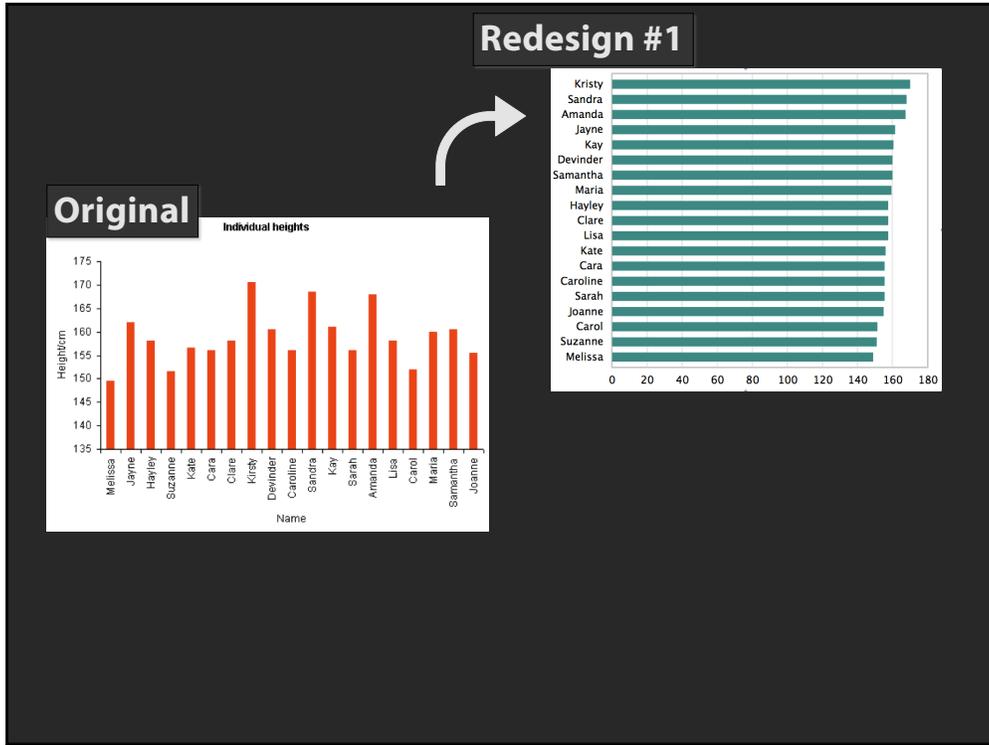
46



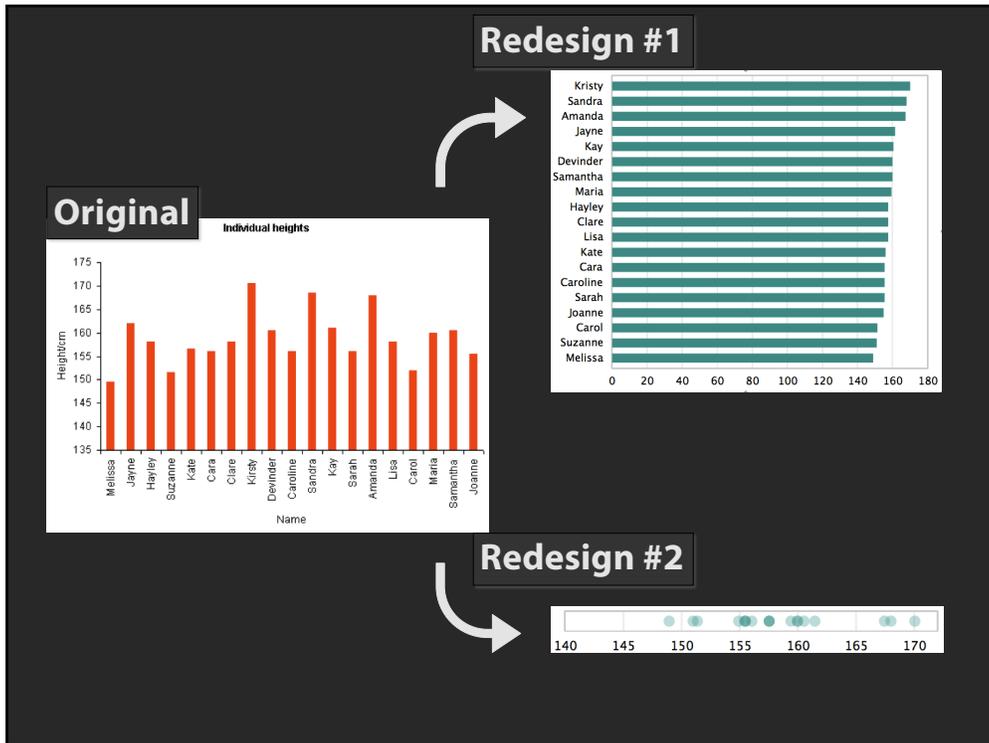
47



48



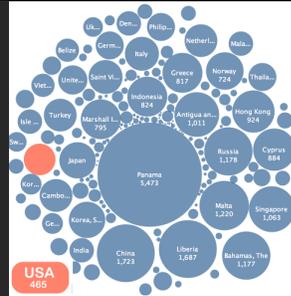
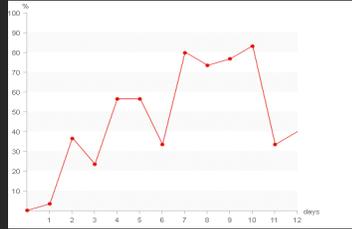
49



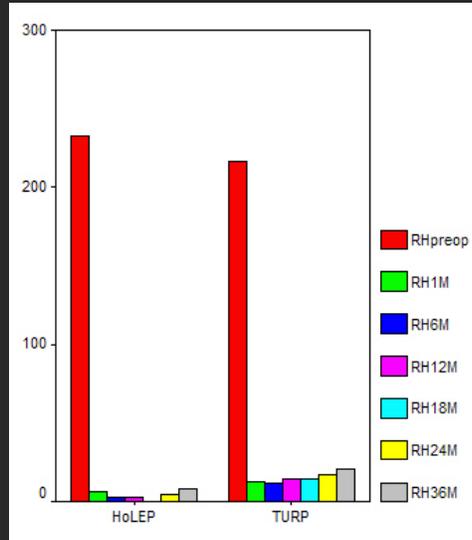
50

Limitations

Additional Chart Types

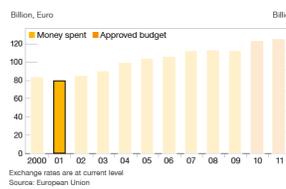


Handling Legends

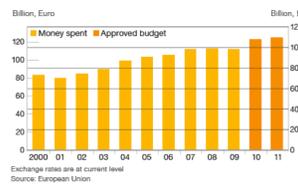


52

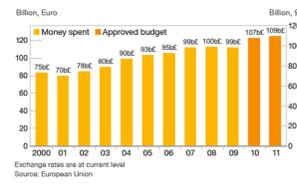
European Union budgets since 2000



European Union budgets since 2000



European Union budgets since 2000



Graphical Overlays

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

53



54

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.

European Union budgets since 2000

Chart type:

Chart:

Overlay type:

Regular gridlines

Lines emanating from marks

Parameters

Overlay Underlay

Static Interactive

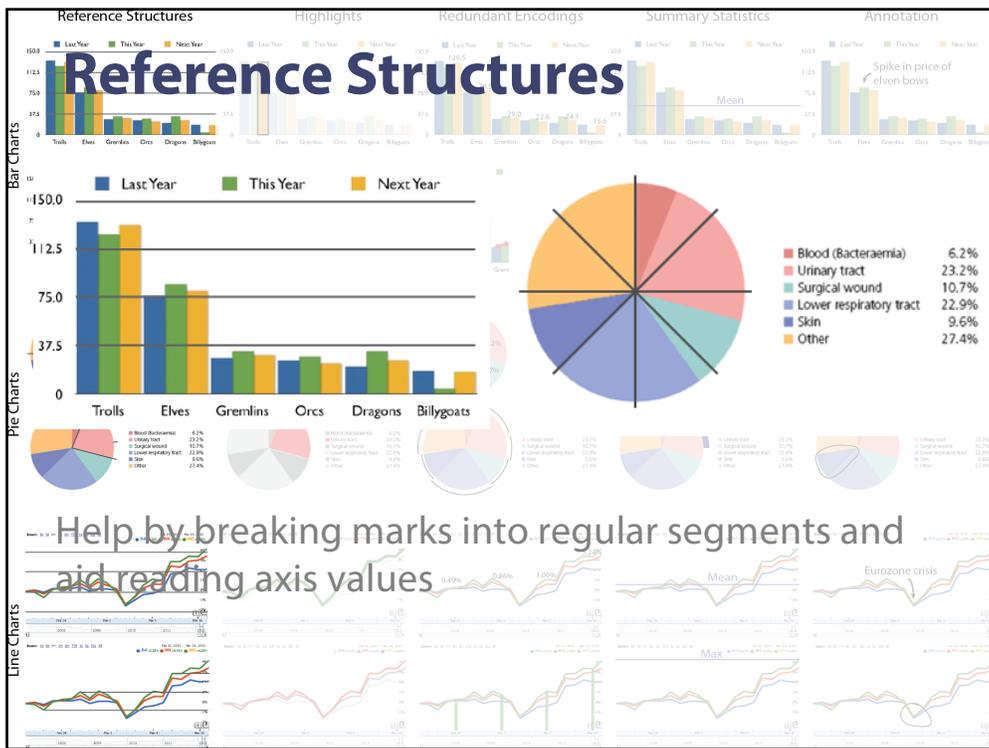
Divisions:

Line thickness:

Places regular gridlines at user defined intervals.

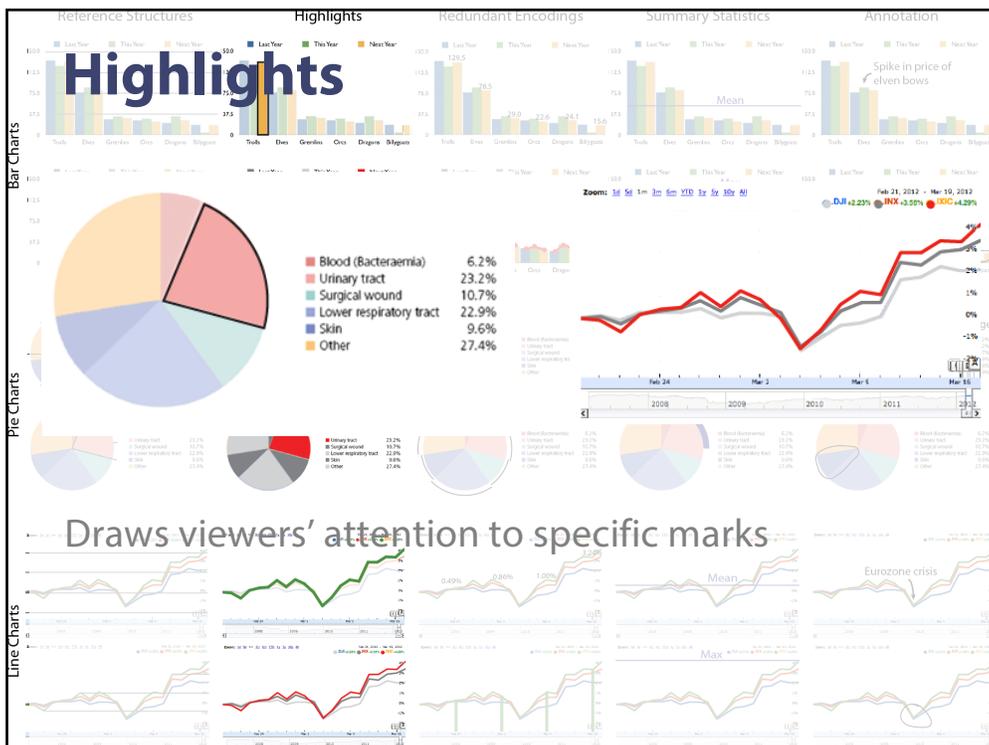
Demo

55



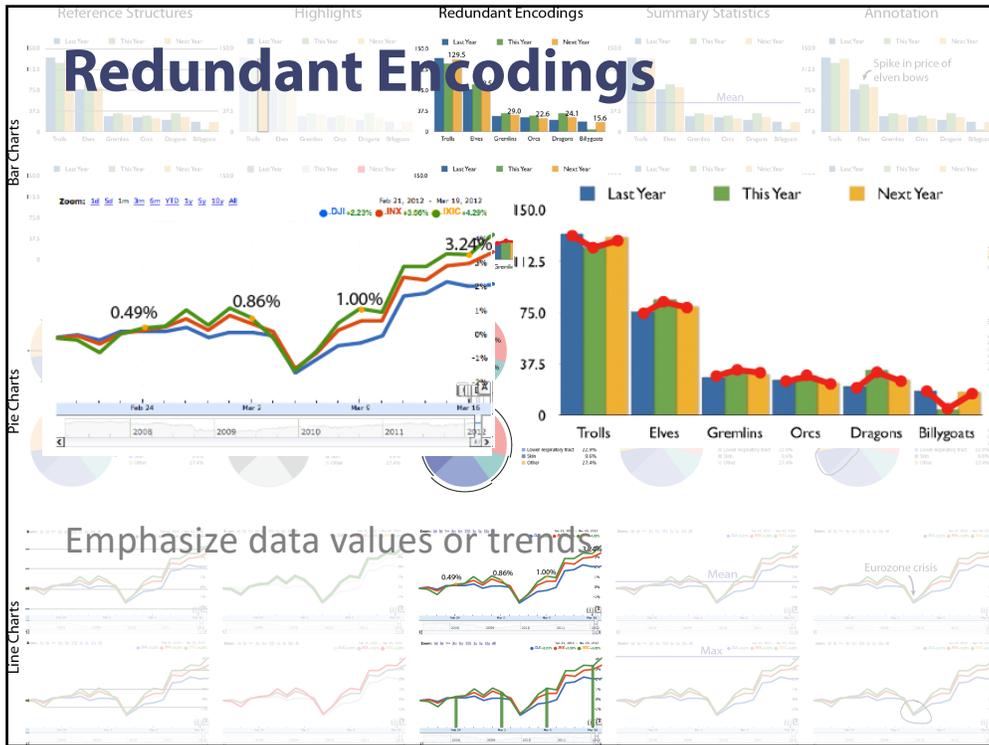
Help by breaking marks into regular segments and aid reading axis values

56



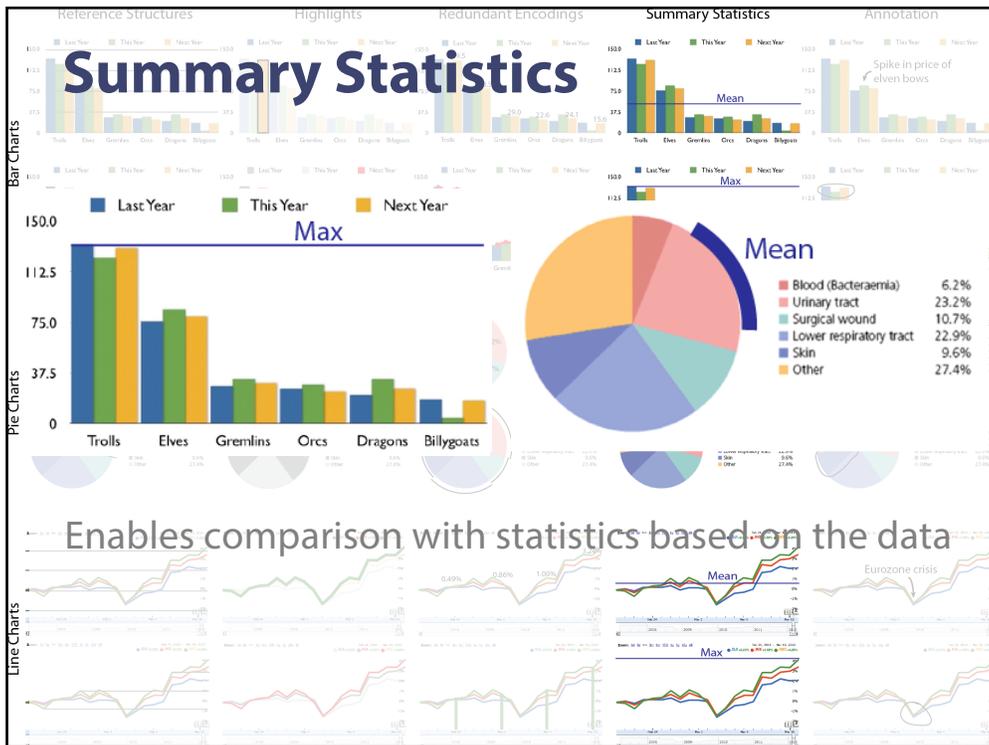
Draws viewers' attention to specific marks

57



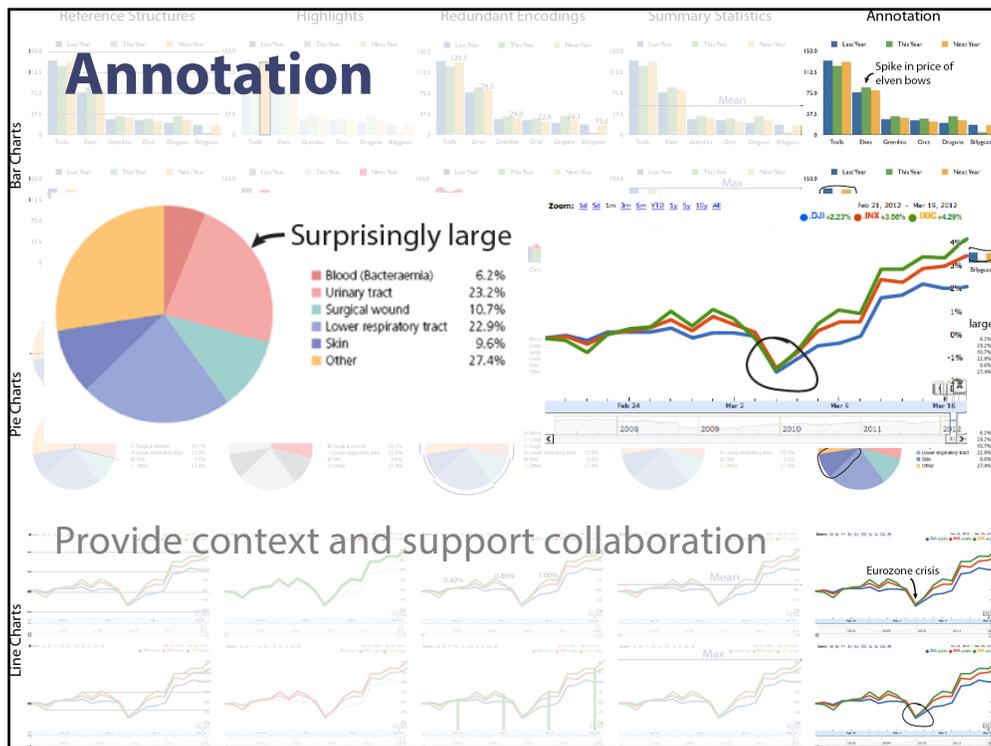
Emphasize data values or trends.

58

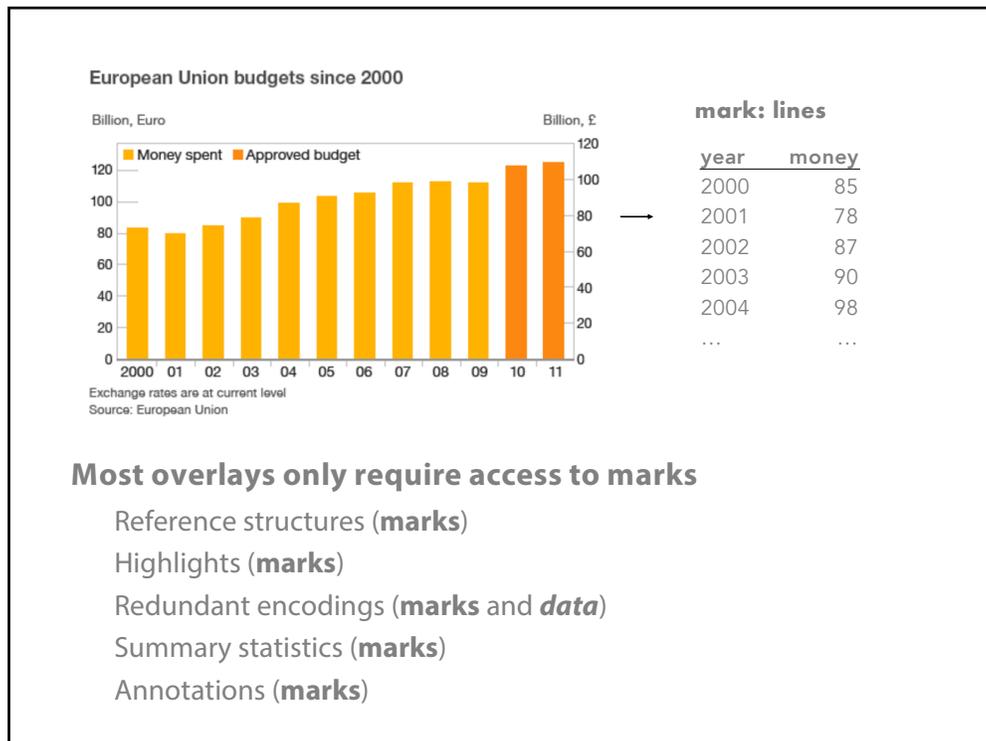


Enables comparison with statistics based on the data

59



60



61

Interactive Documents

How can we facilitate reading text and charts together?

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

The humanitarian fallout of the conflict in Syria reaches new proportions as the number of estimated refugees reaches one million

- Download the data
- More data journalism and data visualisations from the Guardian

Mona Chalabi & Simon Rogers
theguardian.com, Wednesday 6 March 2013 13:03 GMT
Jump to comments (0)

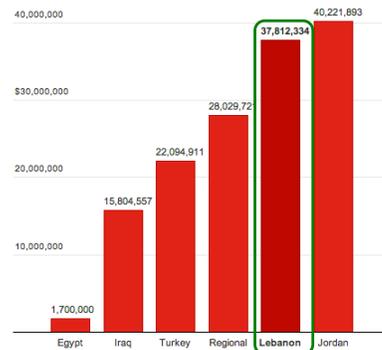
Share 2
Tweet 0
Print 1
Out 0
Share 0
Email

Article history

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 refugees themselves, most funds are funnelled towards Jordan (28%), followed by Lebanon (26%), Turkey (15%) and Iraq (11%).

Where the money goes

Where the international community has donated to help Syria's refugees



SOURCE: UNHCR
GET THE DATA EMBED FULLSCREEN

theguardian

62

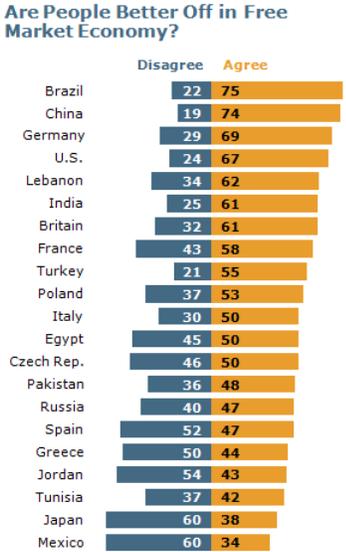
Goal: Extract references between text and chart

63

Problem: Diversity of writing styles

64

Example 1: Pew Research



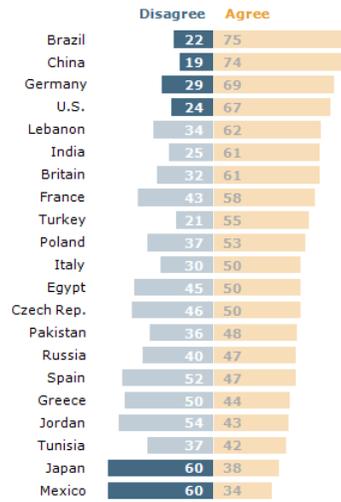
Skepticism for capitalism is lowest in Brazil (22%), China (19%), Germany (29%) (although East Germans are less supportive than West Germans) and the U.S. (24%). Skepticism for free markets is highest in Mexico (60%) and Japan (60%).

PEW RESEARCH CENTER Q26.

65

Example 1: Pew Research

Are People Better Off in Free Market Economy?



PEW RESEARCH CENTER Q26.

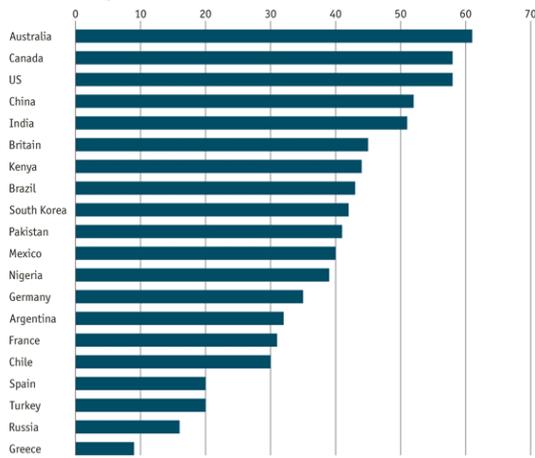
Skepticism for capitalism is lowest in **Brazil (22%), China (19%), Germany (29%)** (although East Germans are less supportive than West Germans) and the **U.S. (24%)**. Skepticism for free markets is highest in **Mexico (60%)** and **Japan (60%)**.

66

Example 2: Economist

Public opinion on the rich

Respondents who agree* that most rich people in their country deserve their wealth, %
December 2011–February 2012



Source: GlobeScan

**Strongly agree* and *Somewhat agree"

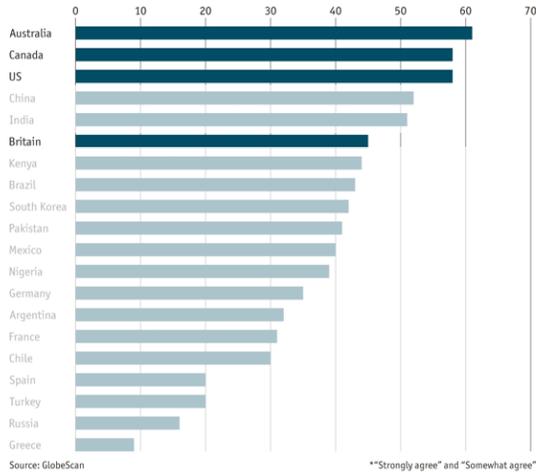
Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions. Europeans and Latin Americans tend to have similar attitudes to the rich; the Anglo-Saxon world is a bit more forgiving.

67

Example 2: Economist

Public opinion on the rich

Respondents who agree* that most rich people in their country deserve their wealth, %
December 2011–February 2012



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.

68

Preprocessing

Document segmentation



Mark and data extraction

Crowdsourcing

Reference extraction

Clustering and Merging

Merge



Split



Cluster



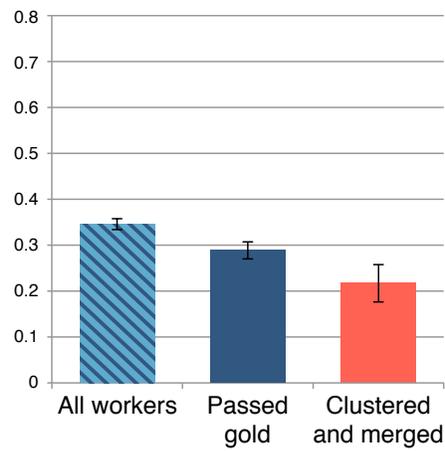
Select representative

69

Demo

70

Evaluation



Avg. F₁ distance: expert specified references vs. crowd specified references

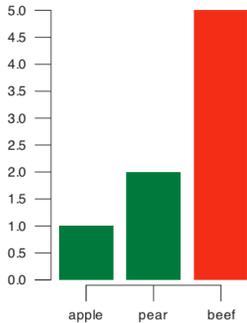
71

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

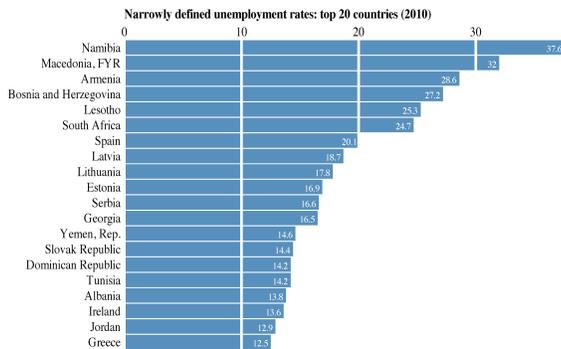
Our Deconstruction

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.

73

Deconstructing D3 Charts



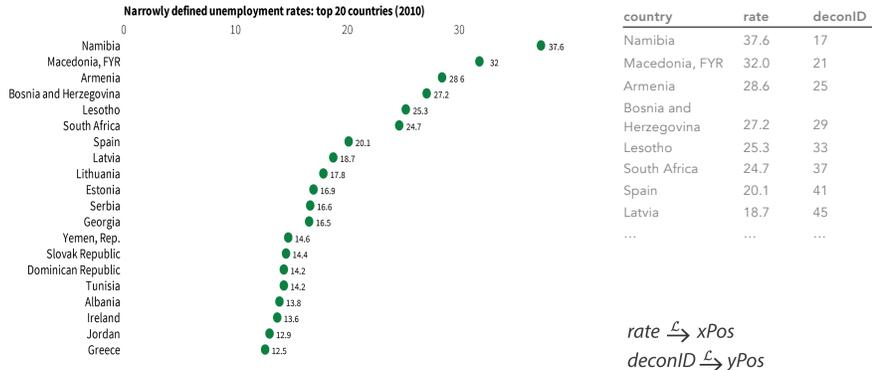
country	rate	deconID
Namibia	37.6	17
Macedonia, FYR	32.0	21
Armenia	28.6	25
Bosnia and Herzegovina	27.2	29
Lesotho	25.3	33
South Africa	24.7	37
Spain	20.1	41
Latvia	18.7	45
...

rate ↪ *width*
rate ↪ *area*
rate ↪ *xPos*
deconID ↪ *yPos*

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

74

Deconstructing D3 Charts

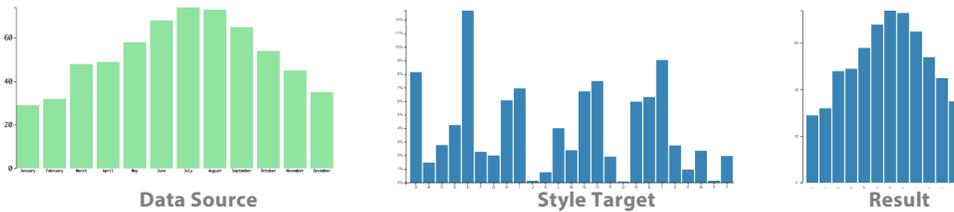


Deconstructing and Restyling D3 Visualizations. Jonathan Harper and Maneesh Agrawala. User Interface Software Technology (UIST) 2014.

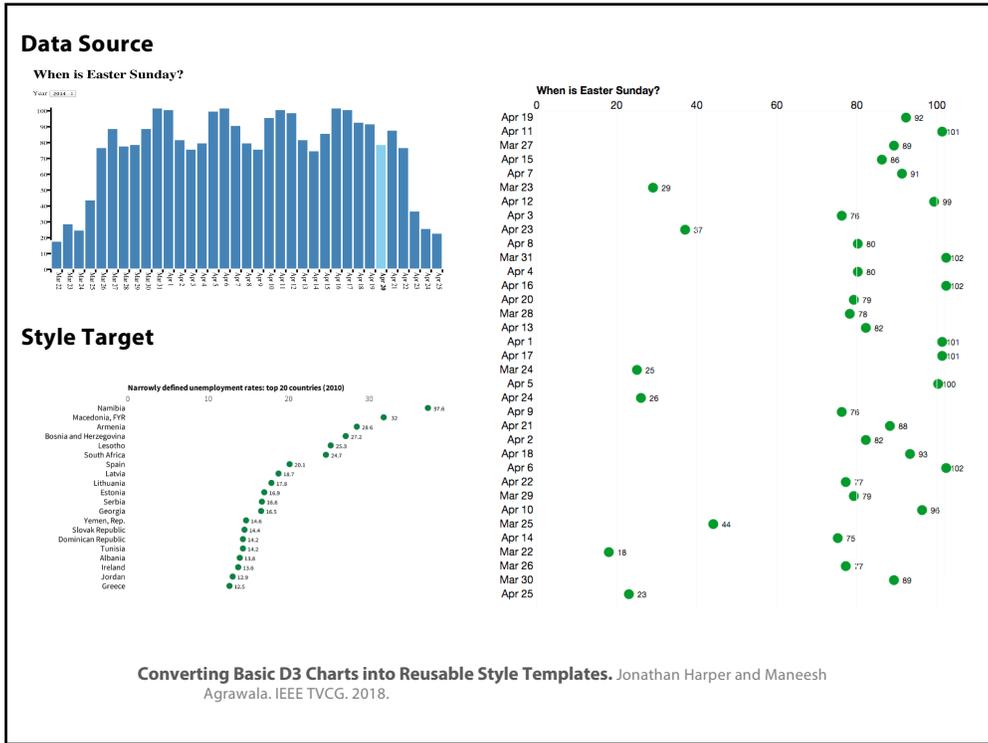
75

Automatic Redesign

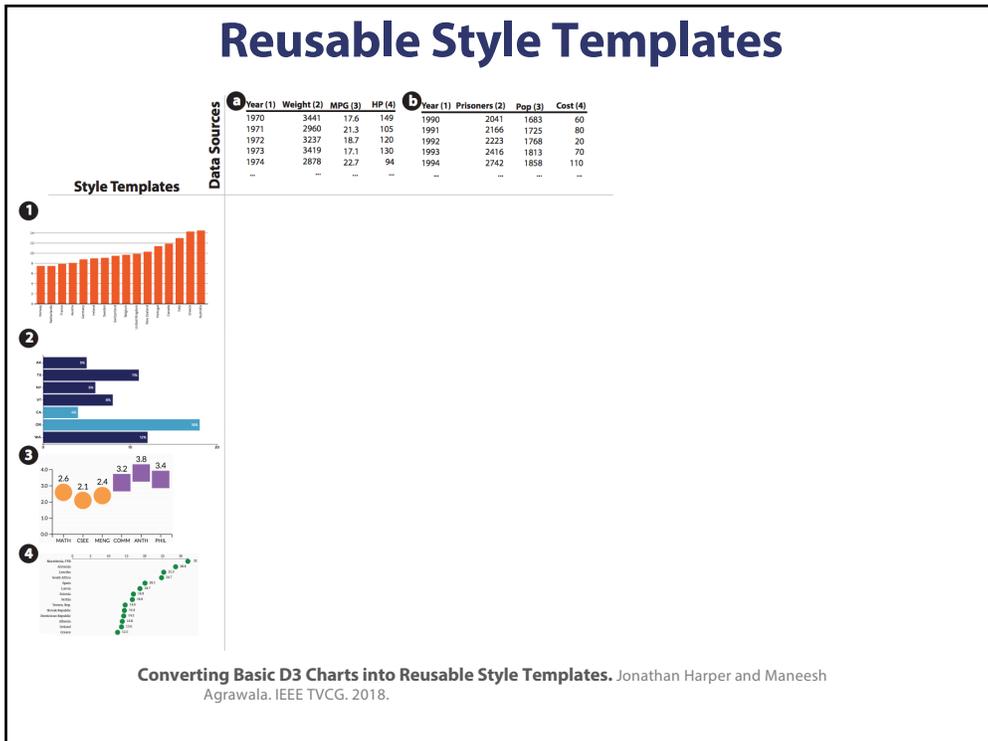
Can we automatically redesign charts to improve
 Perceptual effectiveness?
 Visual aesthetics?
 Accessibility for vision impaired users?



76

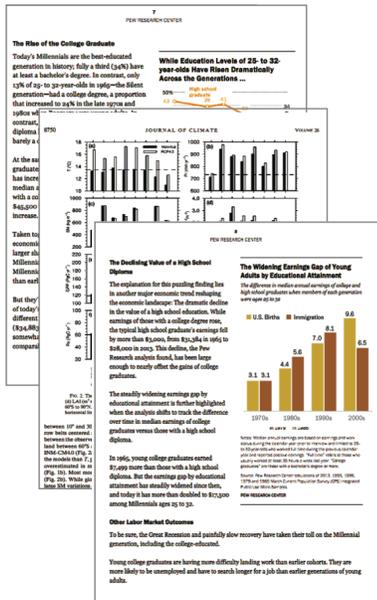


77



78

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

79

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

80