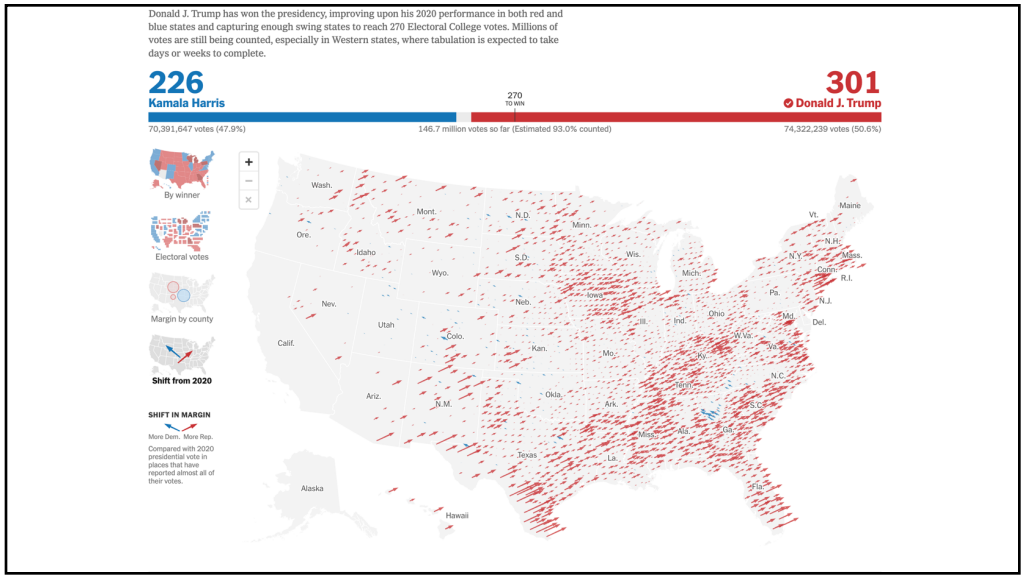
# DECONSTRUCTING VISUALIZATIONS

CS 448B | Fall 2024

MANEESH AGRAWALA

1




2

# READING RESPONSE: QUESTIONS/THOUGHTS

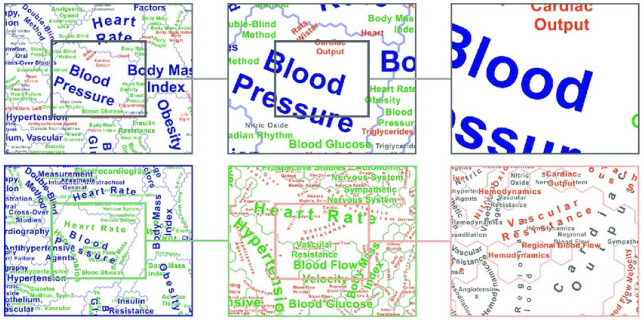
Are there scenarios where enhanced 2D techniques—such as interactive zoom, layered views, or even selective filtering—could achieve similar results without the visual and cognitive complexities introduced by 3D? When does 3D improve usability, rather than detract? ***In what ways can 3D UI design address the cognitive load that users experience when navigating dense information?***

However, while the article is comprehensive, it could benefit from a deeper examination of the contexts in which different centrality measures are most applicable. ***Centrality metrics might produce conflicting interpretations, particularly in complex networks.*** Additionally, the metrics assume that all connections are of equal significance, a simplification that can misrepresent social structures where connection strength varies.

3

# READING RESPONSE: QUESTIONS/THOUGHTS

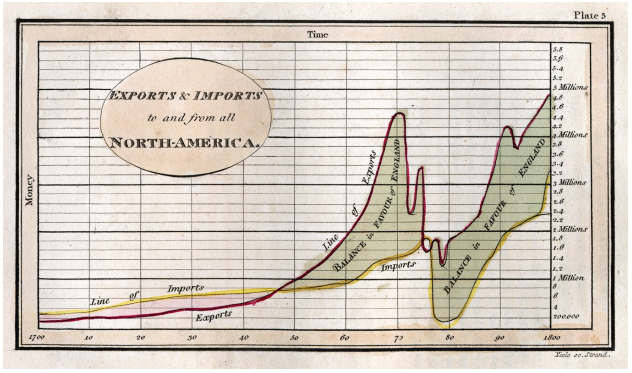
I also found it interesting that the Herman et al. paper distinguishes between two different types of zooms: ***geometric and semantic zooming***. I initially did not realize there was a difference, as I use zoom as a synonym for enlarge, but using real-world examples helped solidify my understanding. It seems that geometric zoom simply means enlarging, like zooming in on a photo taken with your phone, while semantic zooming is similar to zooming in on google maps, with more information being shown in the sense of continent -> country -> state -> city etc.



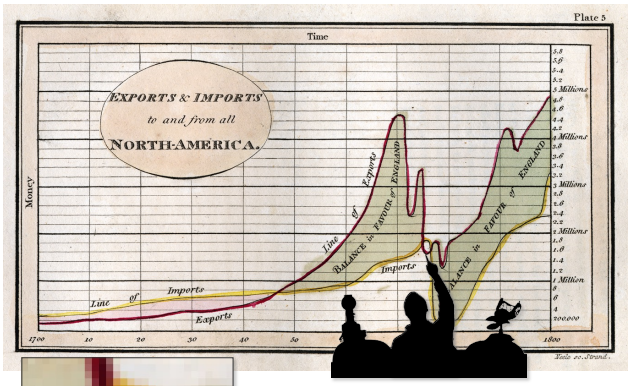

4

# DECONSTRUCTING VISUALIZATIONS

5



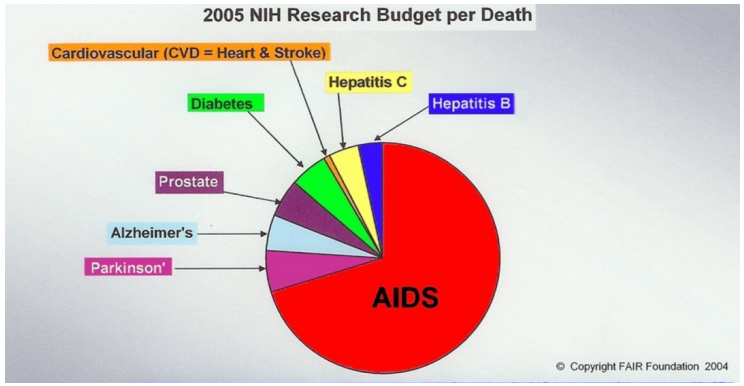

6



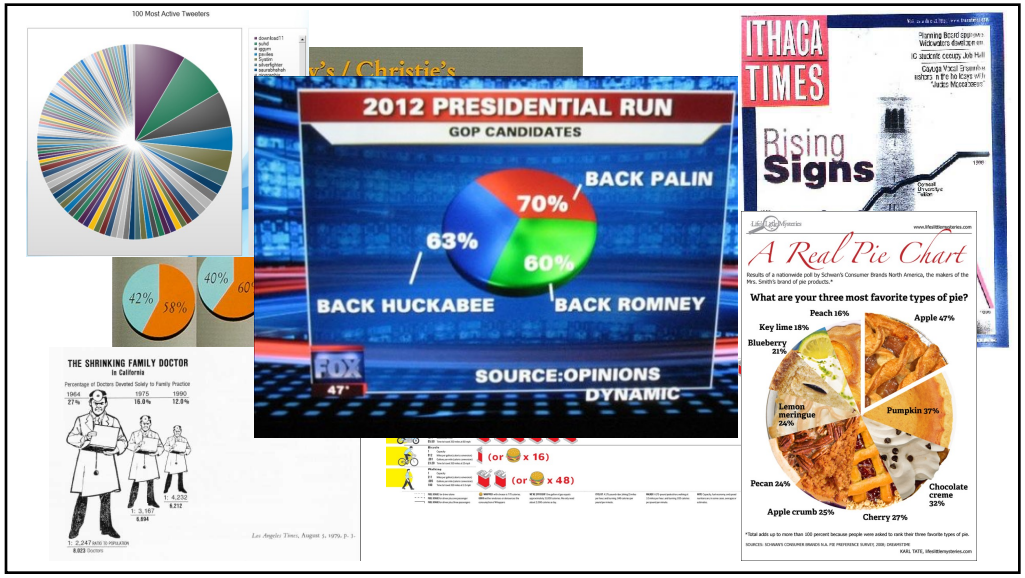

**Pixels are poor representation**

Hard for machines to retrieve data

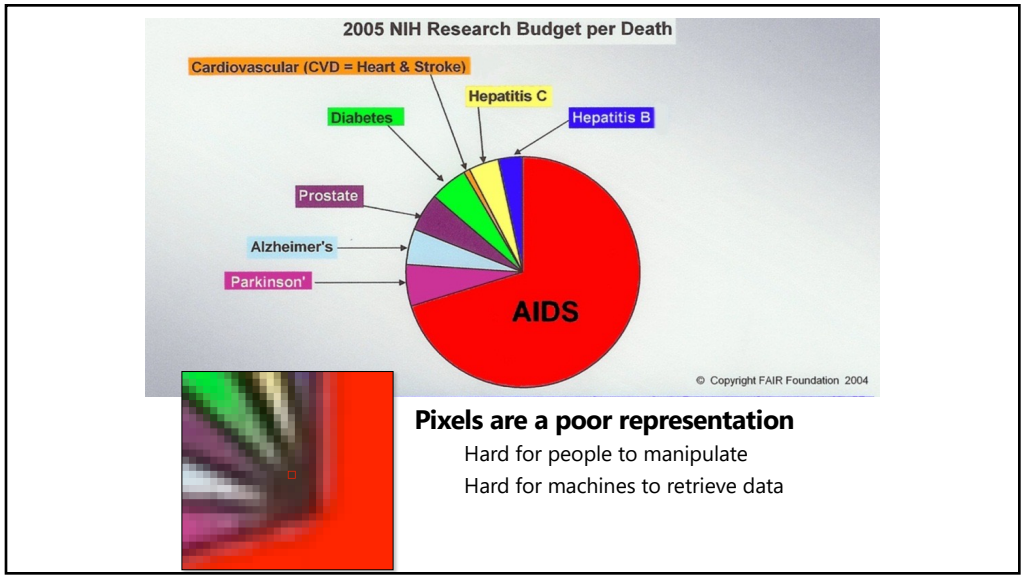
7



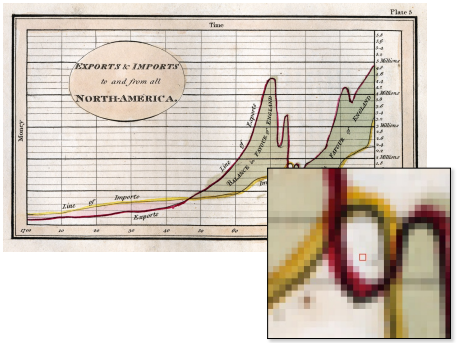

8

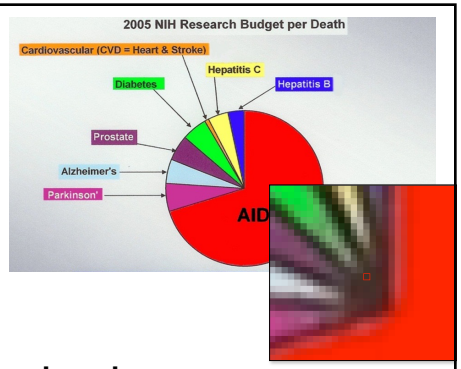

100 Most Active Tweeters
2012 PRESIDENTIAL RUN
GOP CANDIDATES
BACK PALIN
70%
63%
60%
BACK HUCKABEE
BACK ROMNEY
SOURCE:OPINIONS DYNAMIC
FOX
47°
ITHACA TIMES
Rising Signs
A Real Pie Chart
What are your three most favorite types of pie?
Peach 16%
Apple 47%
Key lime 18%
Blueberry 21%
Lemon meringue 24%
Pumpkin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

9


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 a poor representation
Hard for people to manipulate
Hard for machines to retrieve data

10




**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

# TODAY

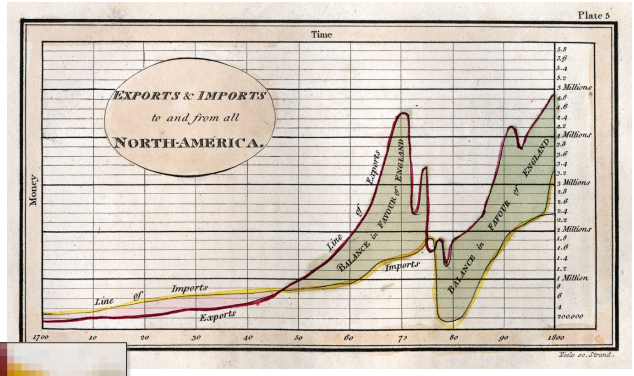
## Learning Objectives

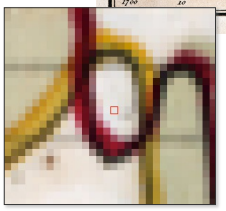
1. How to deconstruct charts and graphs into an editable representation
2. How to use this representation to support interactive reading of visualization

12

# WHAT IS A GOOD REPRESENTATION?

13



**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 |
| … | … | … |

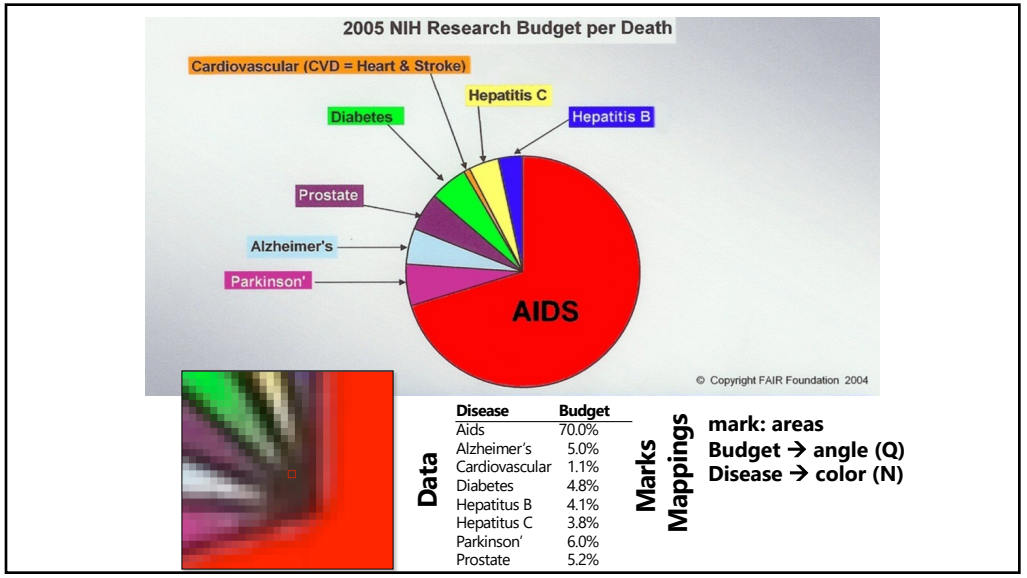
**Marks**

**mark: lines**

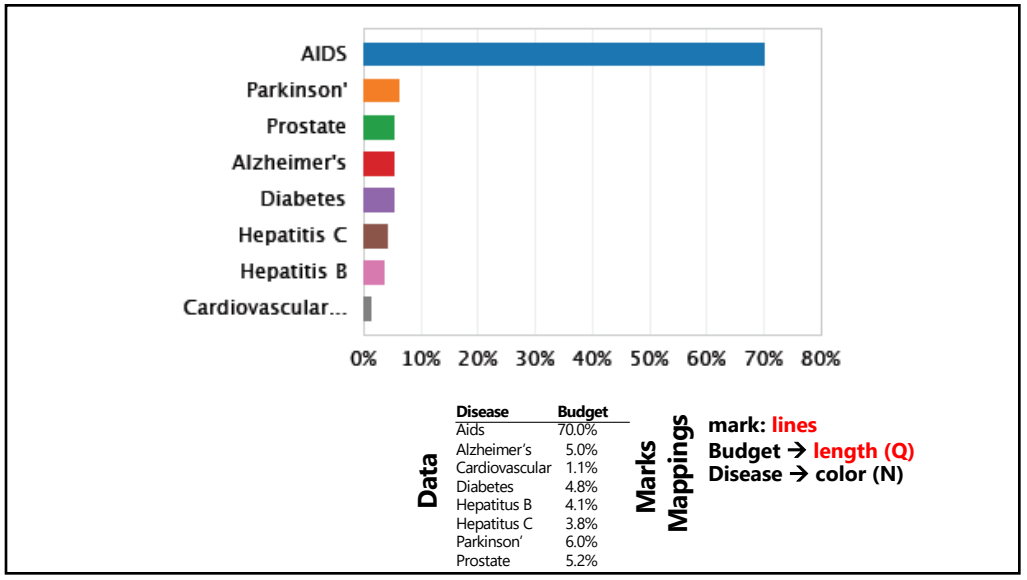
**Mappings**

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

14


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

15


AIDS
Parkinson'
Prostate
Alzheimer's
Diabetes
Hepatitis C
Hepatitis B
Cardiovascular...
0% 10% 20% 30% 40% 50% 60% 70% 80%
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)

16

## APPROACH

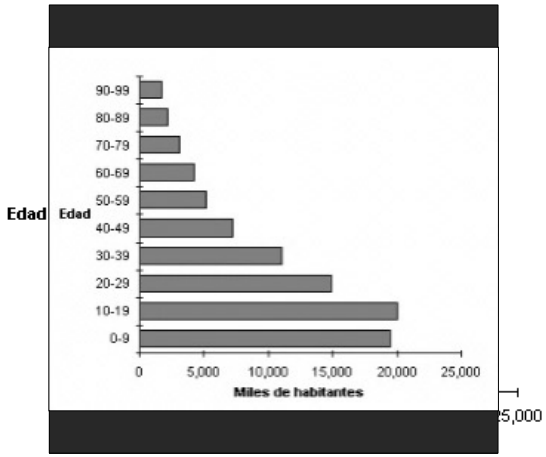
**Classification:** Determine chart type
**Mark extraction:** Retrieve graphical marks
**Data extraction:** Retrieve underlying data values

17

## CLASSIFICATION OF CHART TYPE

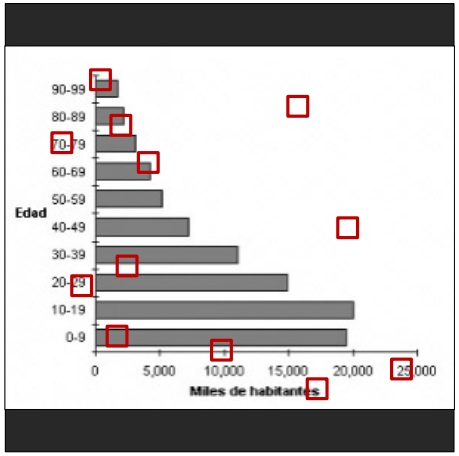
18

# TRAINING



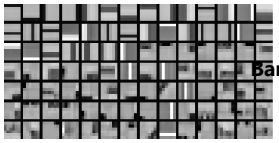

19

# TRAINING

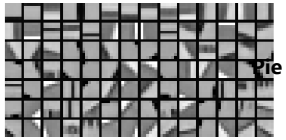



20

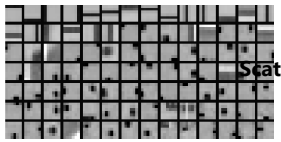
# TRAINING



Bar Charts



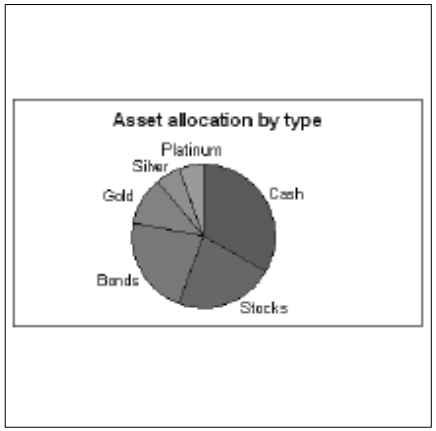
Pie Charts



Scatter Plots

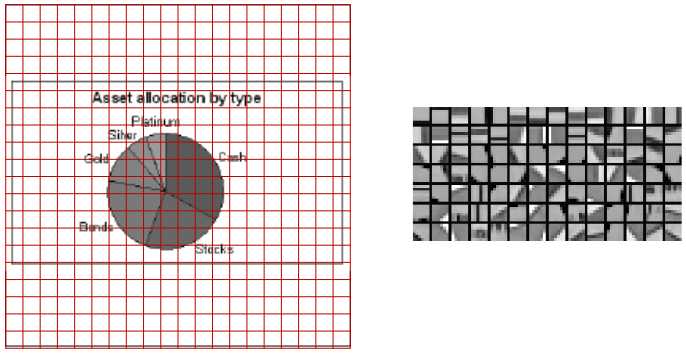
21

# CLASSIFICATION



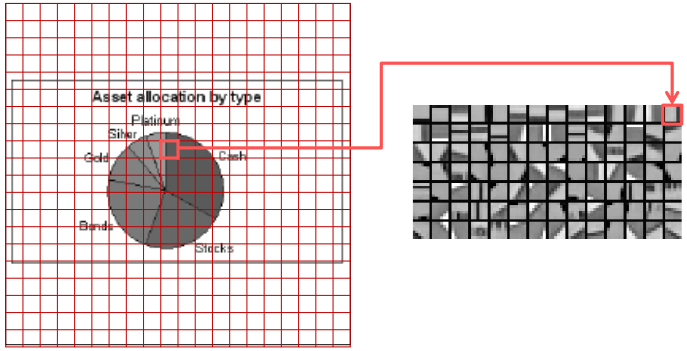

22

# CLASSIFICATION



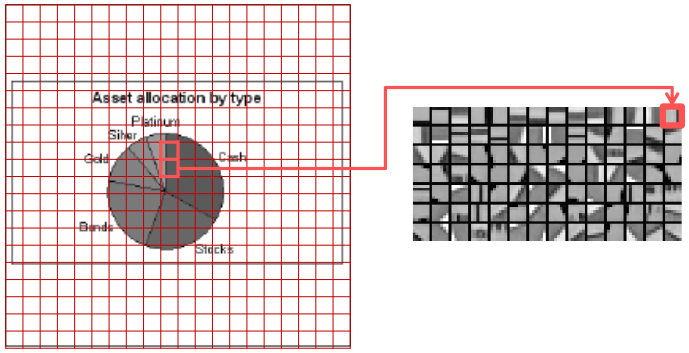

23

# CLASSIFICATION



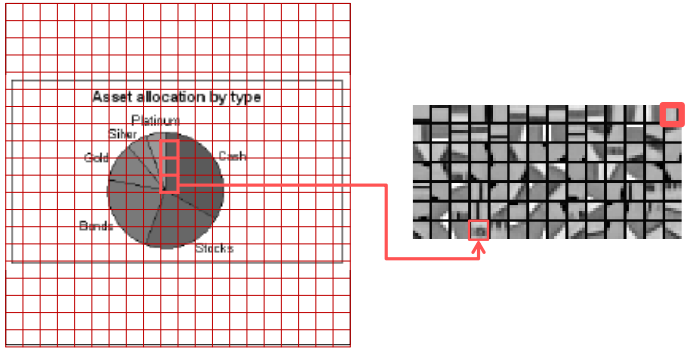

24

# CLASSIFICATION



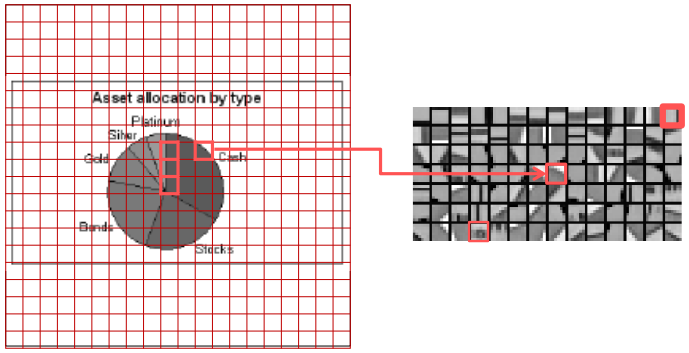

25

# CLASSIFICATION



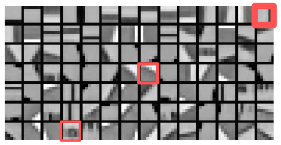

26

# CLASSIFICATION




27

# CLASSIFICATION



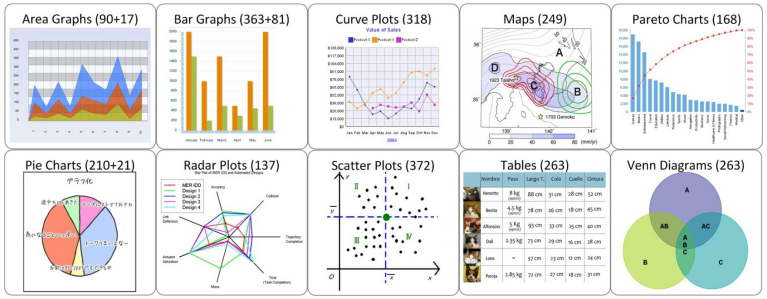
SVM Classifier → Pie Chart

| 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% |

28

# OUR CORPUS

Over 2500 labeled images and 10 chart types




**ReVision binary SVMs give 96% classification accuracy**

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

29

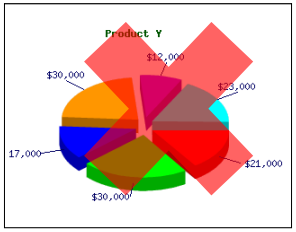
# MARK AND DATA EXTRACTION

30

# ASSUMPTIONS

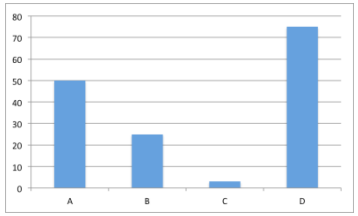
Bar charts and pie charts only

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




31

# BAR CHARTS



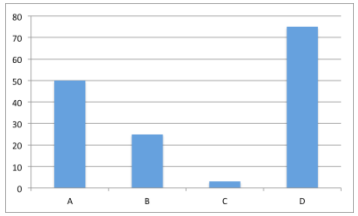
→

**marks: lines**

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

32

# BAR CHARTS

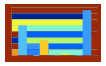


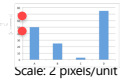

**marks: lines**

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

**Extract Marks** → **Find Foreground Rectangles** → **Identify Orientation and Baseline**

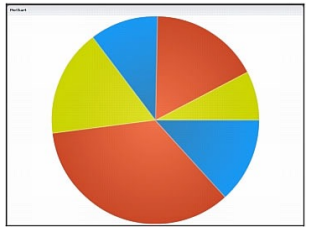
**Extract Data** ← **Recover Bar Values** ← **Associate Labels with Bars**





43

# 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** → **Fit Ellipse Using RNASAC** → **Unroll Pie and Find Transitions**

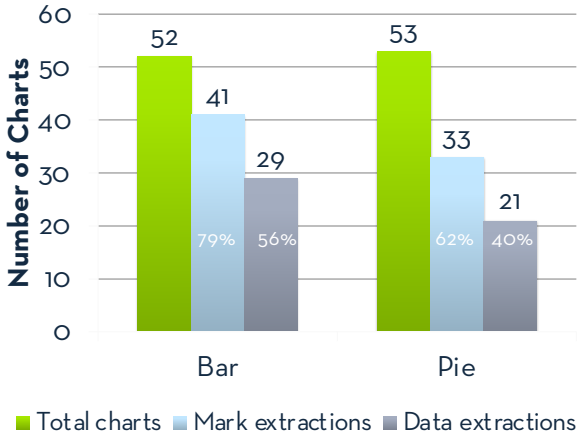
**Extract Data** ← **Compute Area Percentages** ← **Associate Labels with Areas**





44

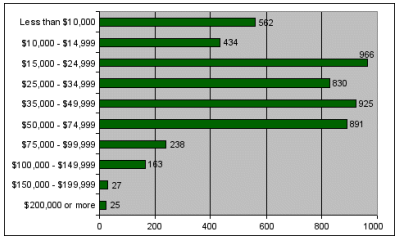
# EXTRACTION RESULTS




45

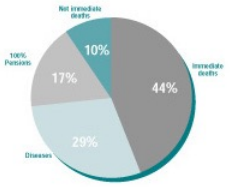
# DATA EXTRACTION ERROR

**Bar Charts**




**7.7%**

**Pie Charts**




**4.6%**

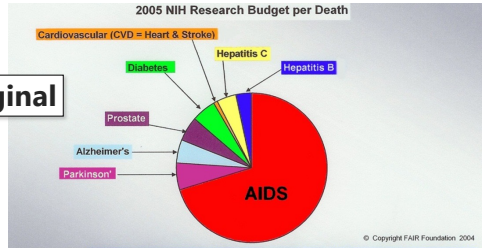
**Average chart size: 342 x 452 pixels** [Prasad 2007]

46

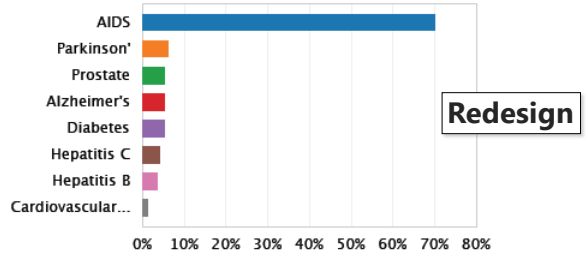
# REDESIGN

47

Original

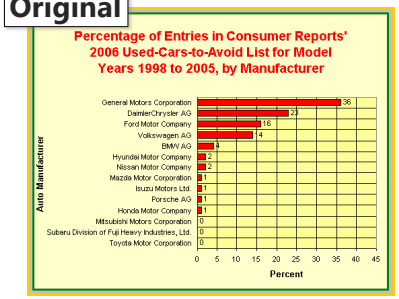



Redesign

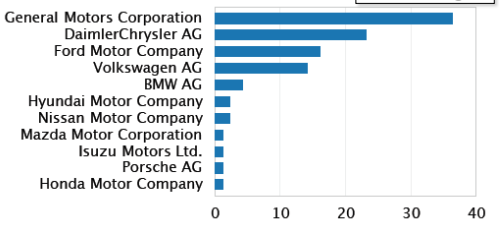



48

**Original**

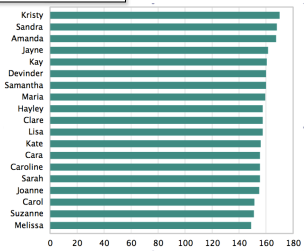



**Redesign**

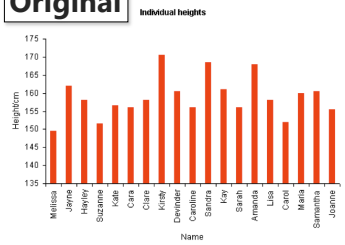



49

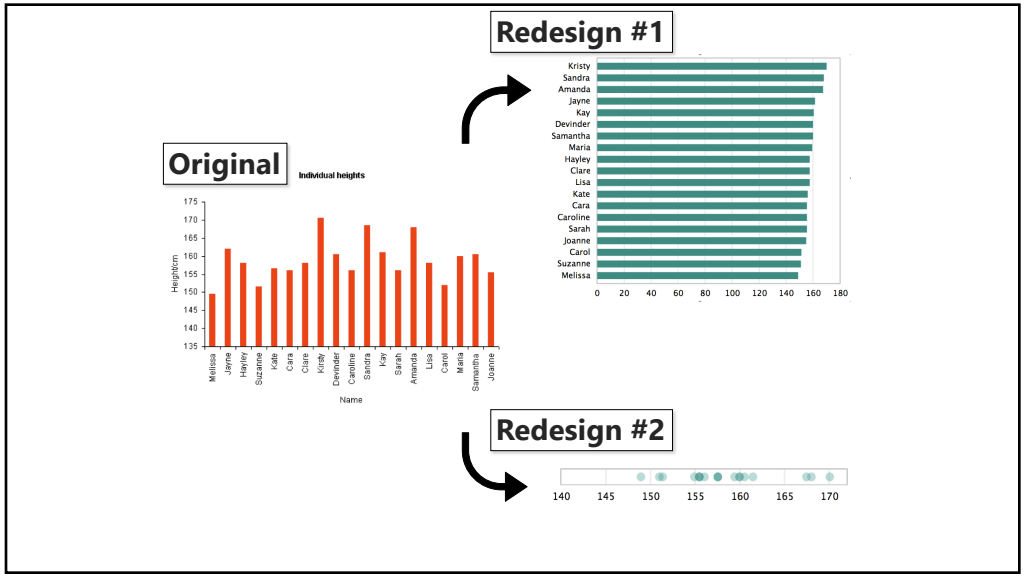
**Redesign #1**




**Original**



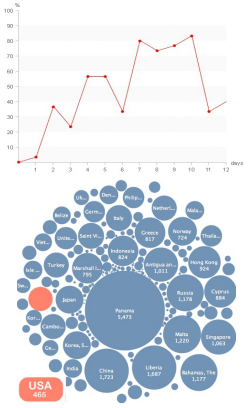

50



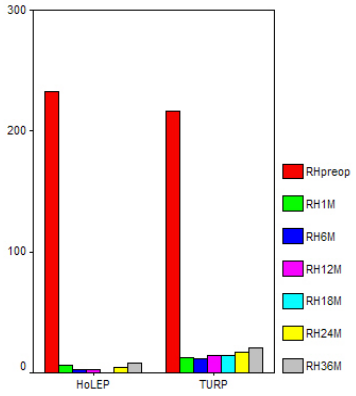

51

# LIMITATIONS

## Additional Chart Types



## Handling Legends




53

# ANNOUNCEMENTS

54

# FINAL PROJECT

## Design Reviews Dec 2 and Dec 4 (signups next week)

**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**

Design Review and Feedback: 10th week of quarter, 12/2 and 12/4
Final code and video: Sun 12/8 8pm

**Grading**

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

55

# THIS WEEK

**No lecture this Wed 11/13 (guest speaker can't make it)**

- Please do the readings
- Use time to focus on final project

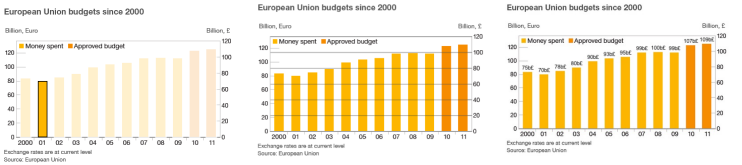
**A2 grades will be released later today**

- Late days not reflected in posted grades (will account for them in final course grades)

56

# INTERACTIVE READING

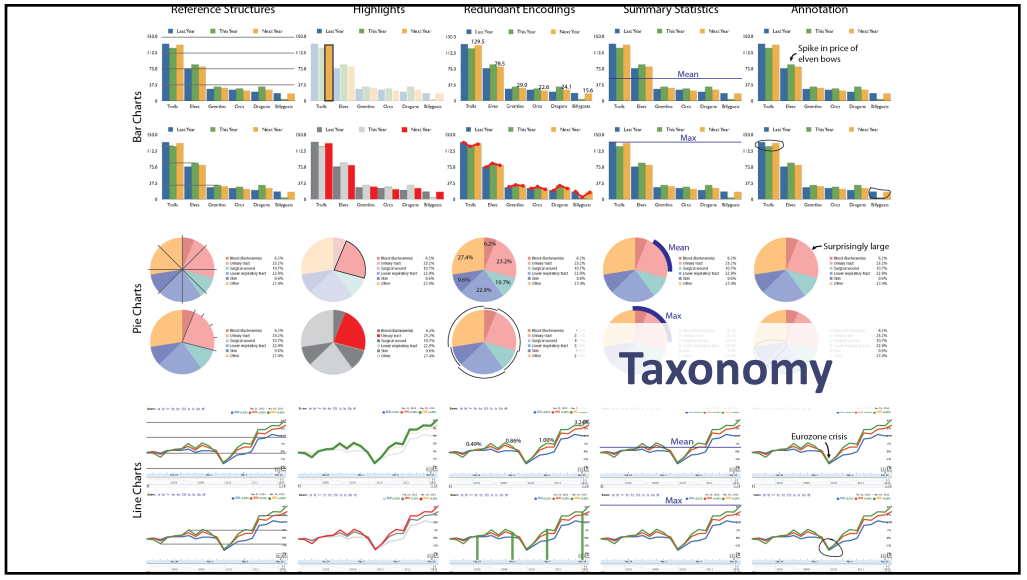
57




# Graphical Overlays

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

58




59

**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.

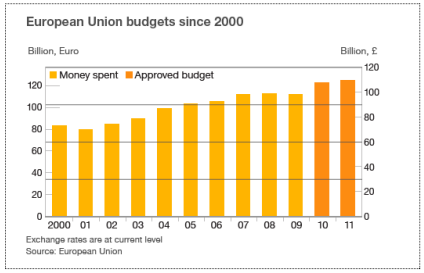



Chart type: Bar
Chart: 00281
Overlay type: Reference structures
Regular gridlines
Lines emanating from marks

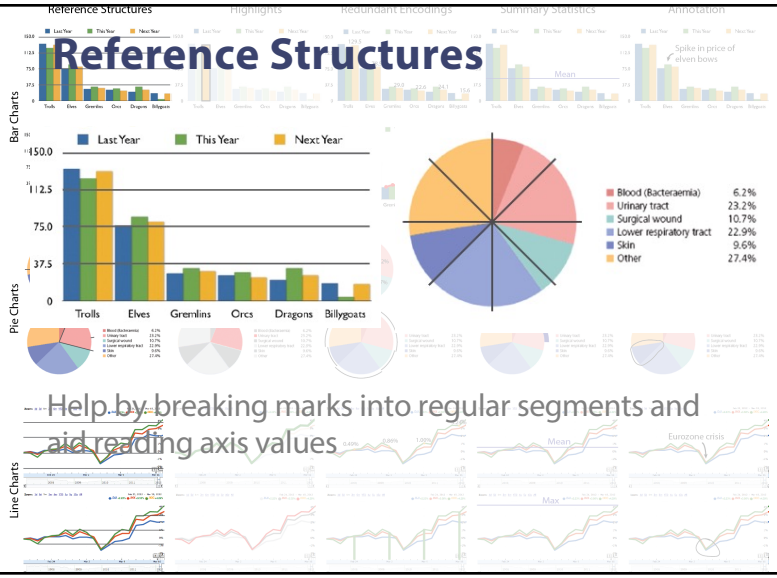
**Parameters**

Overlay Underlay
Static Interactive
Divisions: 4
Line thickness: 1

Places regular gridlines at user defined intervals.

Demo

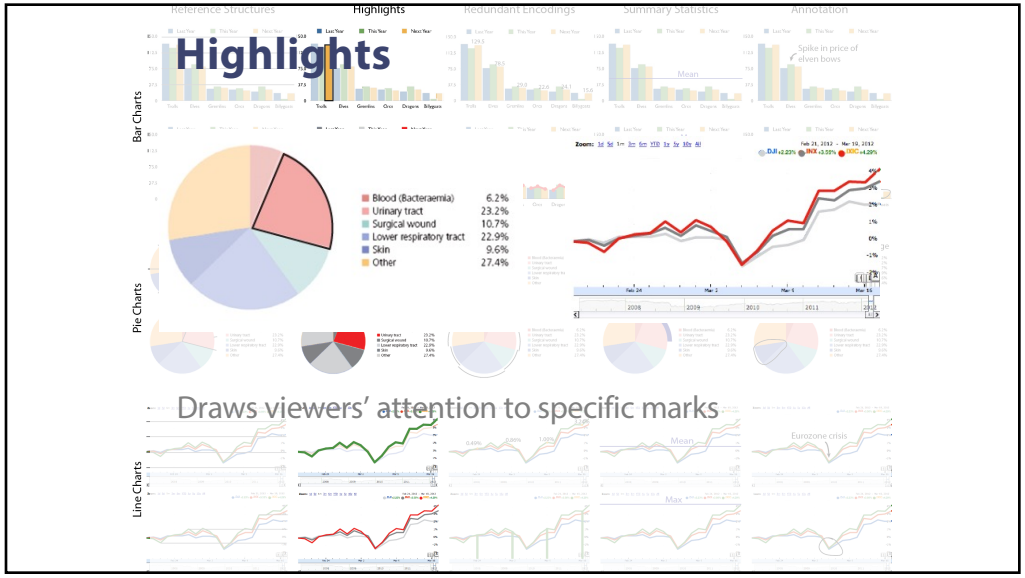
60



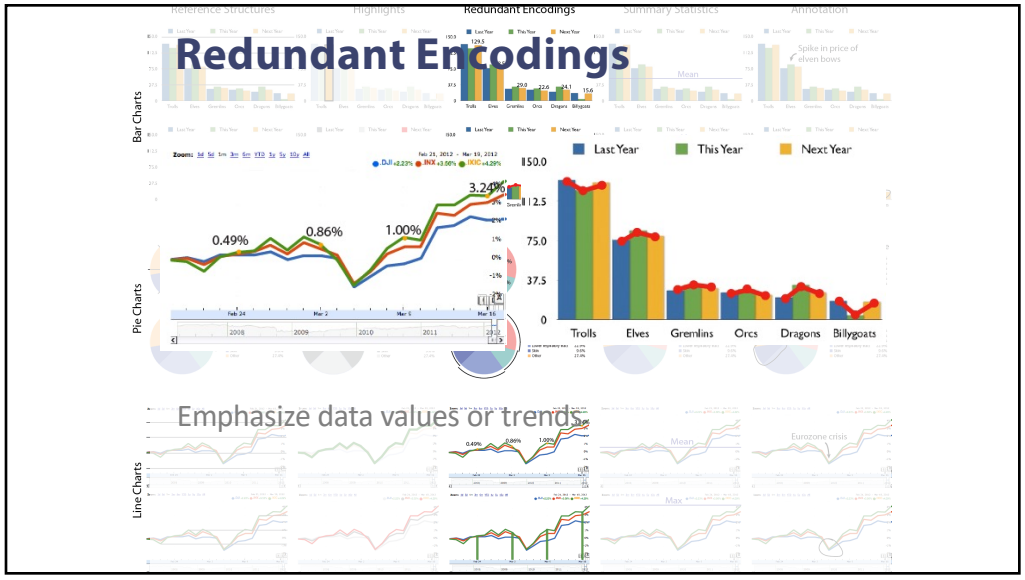

Reference Structures

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

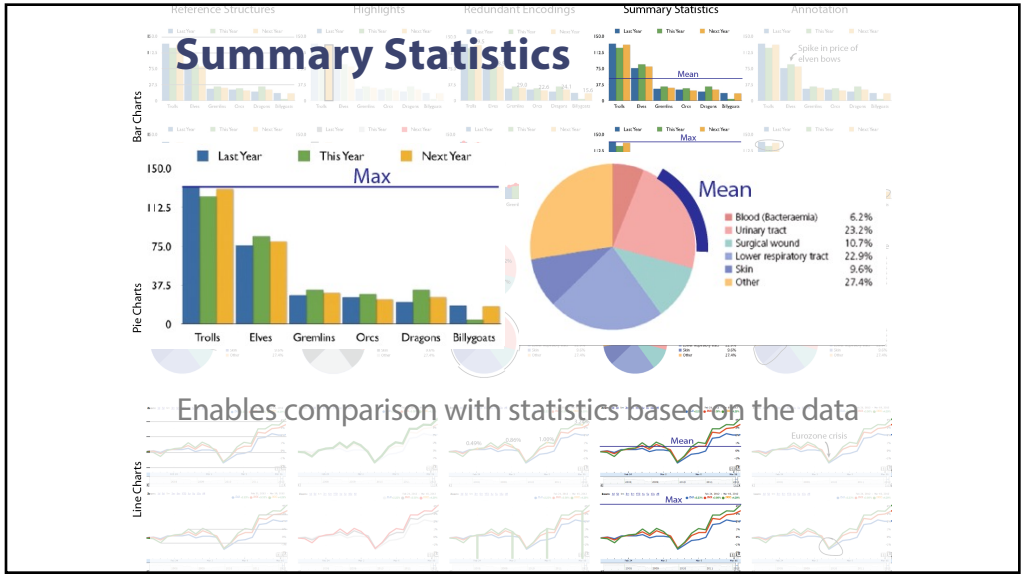
61


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

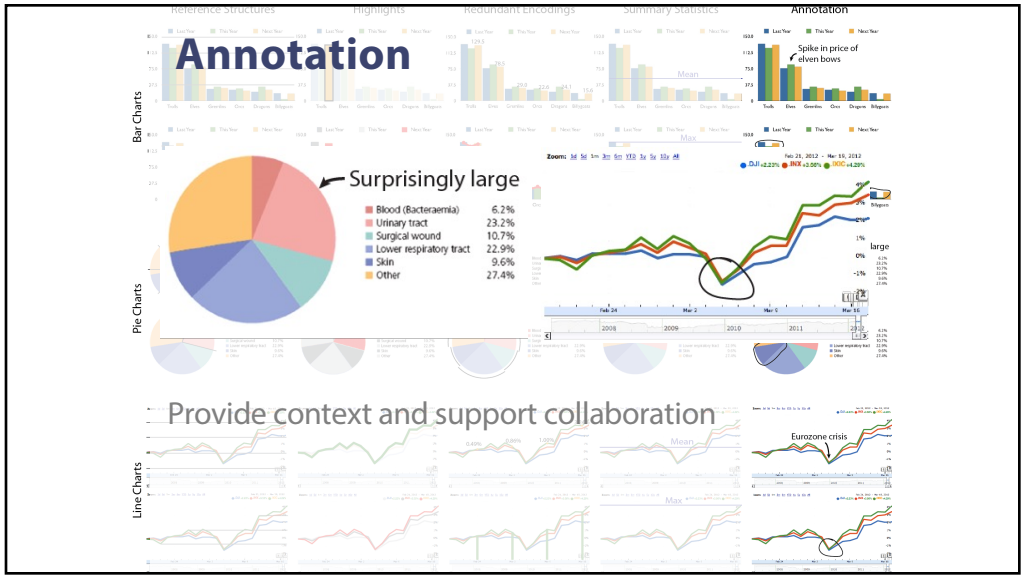
62


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
Emphasize data values or trends

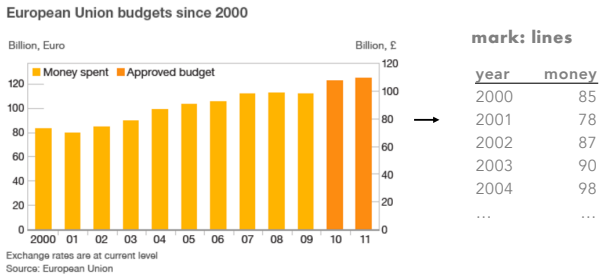
63


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

64


Annotation
Bar Charts
Pie Charts
Line Charts
Spike in price of elven bows
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%
Eurozone crisis
Provide context and support collaboration

65




**Most overlays only require access to marks**

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

66

# INTERACTIVE DOCUMENTS

How can we facilitate reading text and charts together?


67

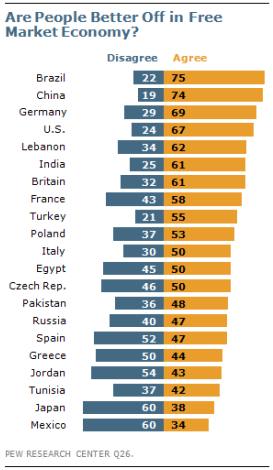
**Goal:** Extract references between text and chart

68

**Problem:** Diversity of writing styles

69

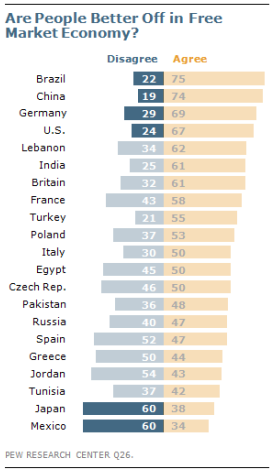
# EXAMPLE 1: PEW RESEARCH




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%).

70

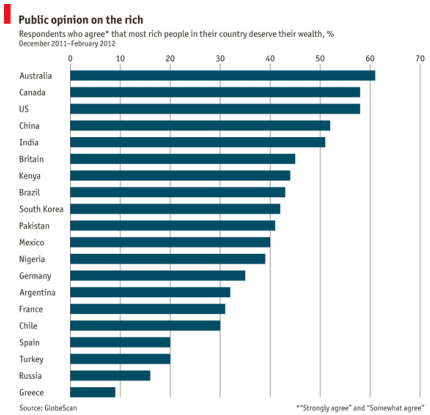
# EXAMPLE 1: PEW RESEARCH




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%).**

71

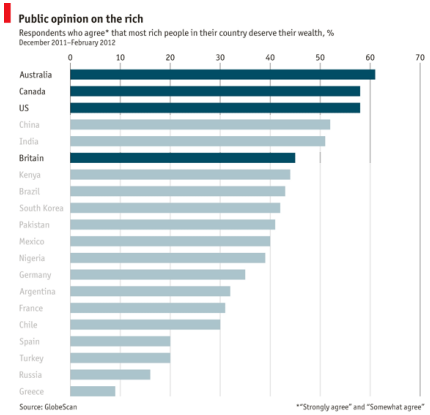
# EXAMPLE 2: ECONOMIST




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.

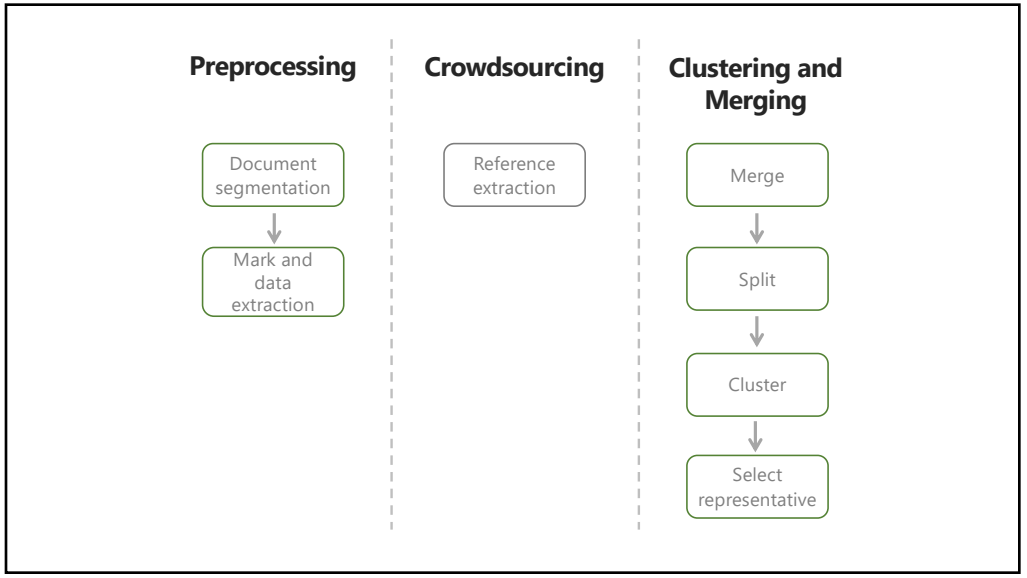
72

# 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.

73

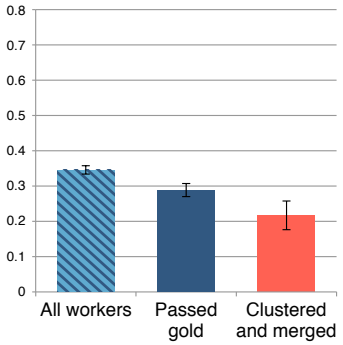

Preprocessing
Crowdsourcing
Clustering and Merging
Document segmentation
Mark and data extraction
Reference extraction
Merge
Split
Cluster
Select representative

74


DEMO

75

# EVALUATION




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

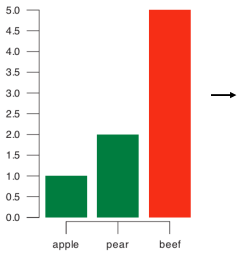
76

# 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*

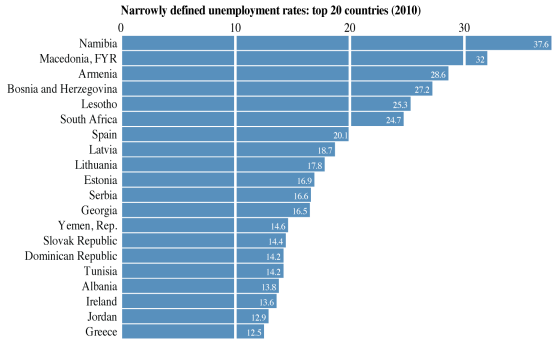
**D3 Code** **D3 Chart** **Our Deconstruction**

**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.

78

# DECONSTRUCTING D3 CHARTS



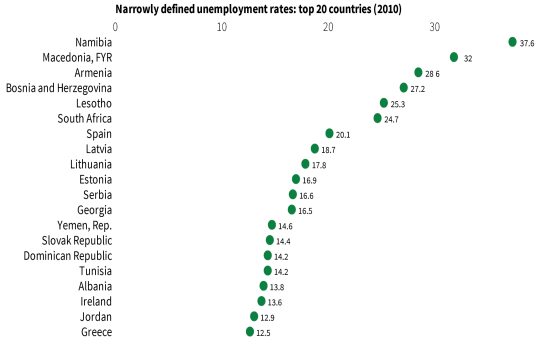

| country | rate | deconID |
|---|---|---|
| Namibia | 37.6 | 17 |
| Macedonia, FYR | 32.0 | 21 |
| Armenia | 28.6 | 25 |
| Bosnia and Herzegovina | 27.2 | 29 |
| Lesotho | 25.3 | 33 |
| South Africa | 24.7 | 37 |
| Spain | 20.1 | 41 |
| Latvia | 18.7 | 45 |
| ... | ... | ... |

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

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

79

# 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*

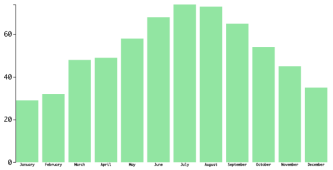
**Deconstructing and Restyling D3 Visualizations.** Jonathan Harper and Maneesh Agrawala. User Interface Software Technology (UIST) 2014.

80

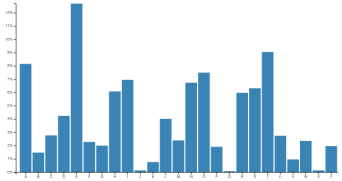
# AUTOMATIC REDESIGN

**Can we automatically redesign charts to improve**

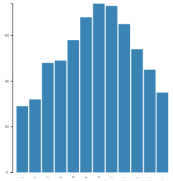
Perceptual effectiveness?
Visual aesthetics?
Accessibility for vision impaired users?



**Data Source**



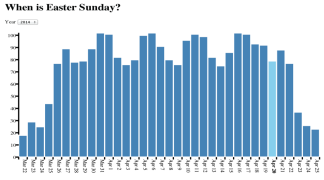
**Style Target**

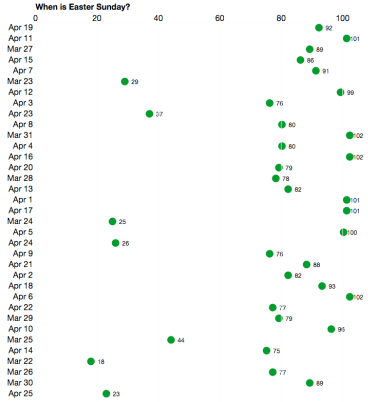


**Result**

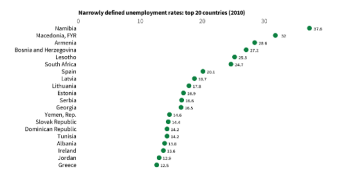
81

## Data Source



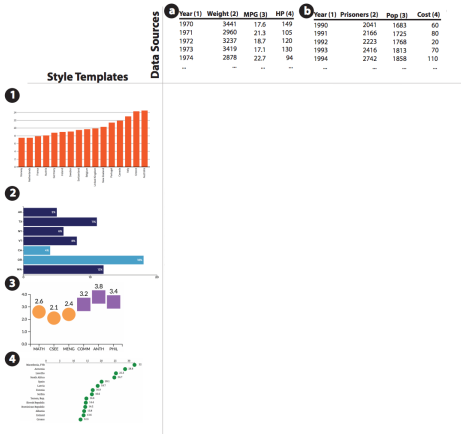

## Style Target




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

82

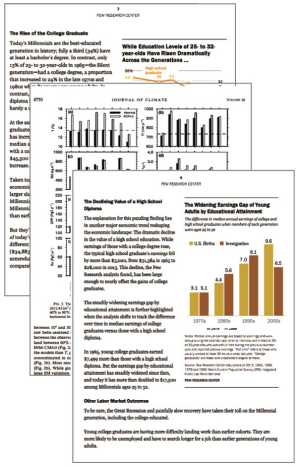
# REUSABLE STYLE TEMPLATES




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

83

# 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

84

# 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***

85