

# Deconstructing Visualizations

*Maneesh Agrawala*

**CS 448B: Visualization**  
**Winter 2020**

1

## Announcements

2

# Final project

---

## New visualization research or data analysis project

- **Research:** Pose problem, Implement creative solution
- **Data analysis:** Analyze dataset in depth & make a visual explainer

## Deliverables

- **Research:** Implementation of solution
- **Data analysis/explainer:** Article with multiple interactive visualizations
- 6-8 page paper

## Schedule

- Project proposal: **Wed 2/19**
- Design review and feedback: **3/9 and 3/11**
- Final presentation: **3/16 (7-9pm) Location: TBD**
- Final code and writeup: **3/18 11:59pm**

## Grading

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

3

# Design Feedback (Week 10)

---

## Signup for a 10 min slot

<https://docs.google.com/spreadsheets/d/1BiXmbQHrC3-chPT6kKS51Q-2p9XhbiM3Qct0N847yPM/edit?usp=sharing>

- M 3/9 4-6pm
- T 3/10 7-8pm (SCPD only)
- W 3/11 4-6pm

**Plan to give a 5 min presentation (mostly demo) of work so far. We will give oral feedback.**

4

# Final Presentation

---

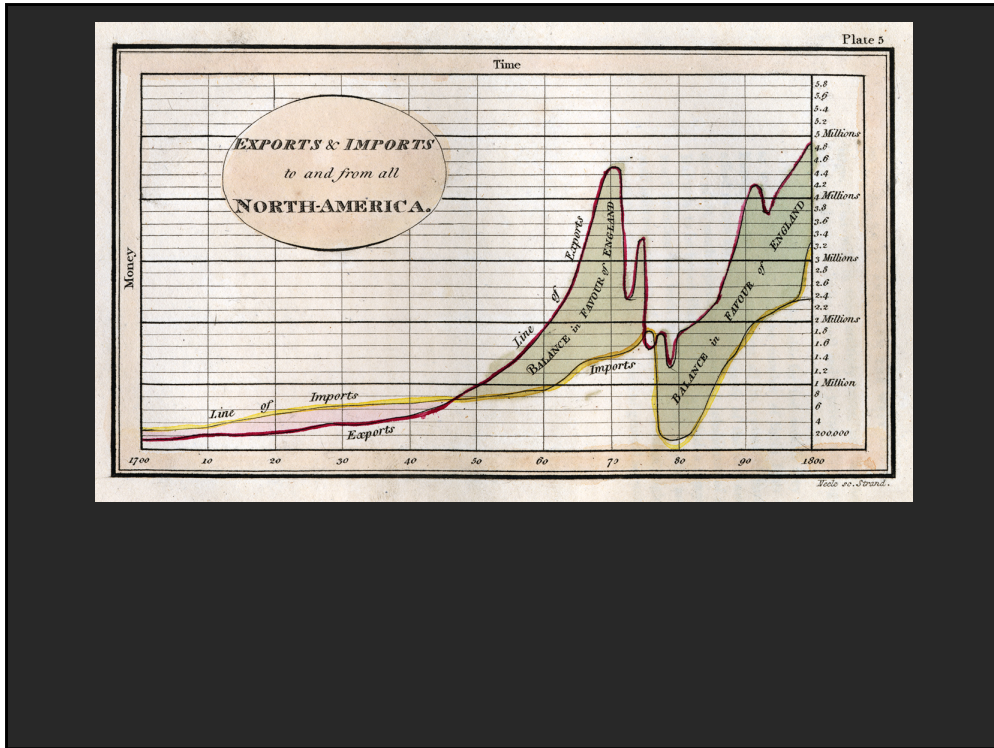
**M Mar 16 7-10pm, Location TBD**

- Short presentation (5 min, mostly demo)
- Make sure there is time for questions

5

## Deconstructing Visualizations

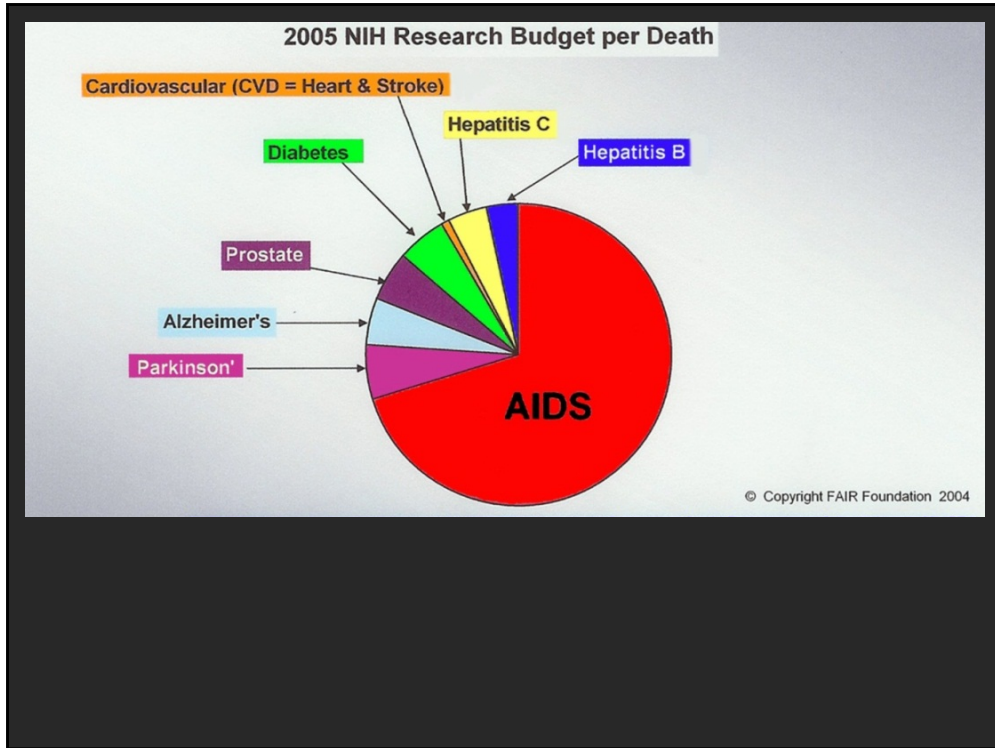
6



7



8



9

#### 100 Most Active Tweeters

- download11
- ashle
- oggin
- givine
- syrie
- slemphor
- savellefish
- pignacho
- Download
- cormenhouout
- shaffard
- LacOUD

#### Christie's Sales

#### 2012 PRESIDENTIAL RUN

##### GOP CANDIDATES

SOURCE: OPINIONS DYNAMIC

#### ITINERARIES

#### THE SHRINKING FAMILY in California

Percentage of Doctors Devoted Solely to Family Practice

Year	Percentage
1964	27%
1975	16.6%
1986	11.3%
1997	6.8%
2008	4.2%

Los Angeles Times, August 5, 1999, p. 3.

#### Pie Chart

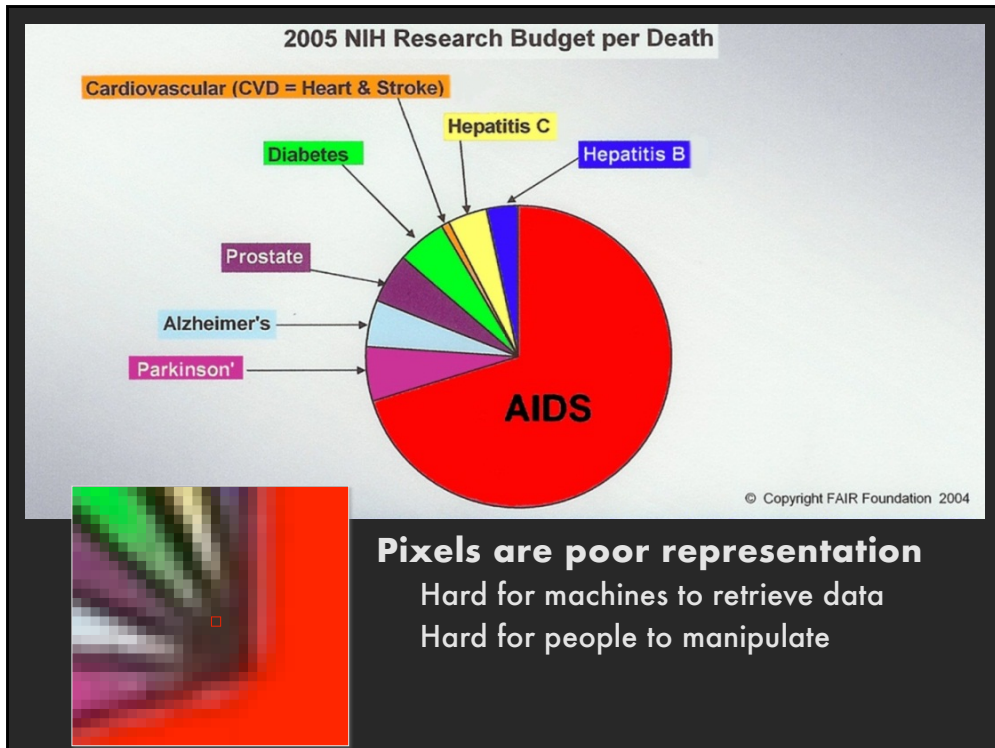
Consumer Brands North America, the makers of the most favorite types of pie?

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

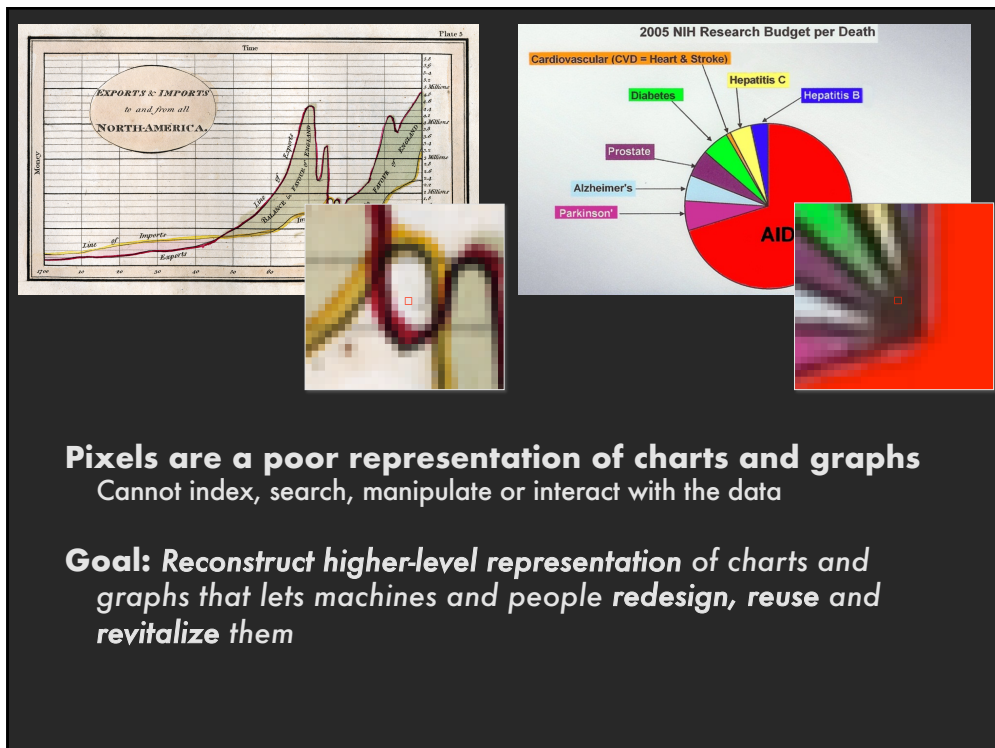
\*Total adds up to more than 100 percent because people were asked to rank their three favorite types of pie.

SOURCE: SHERWIN CONSUMER BRANDS N.A. THE PRESIDENTIAL SURVEY 2008. CREATED BY KARL TATE, [MediaSystem.com](http://MediaSystem.com)

10



11

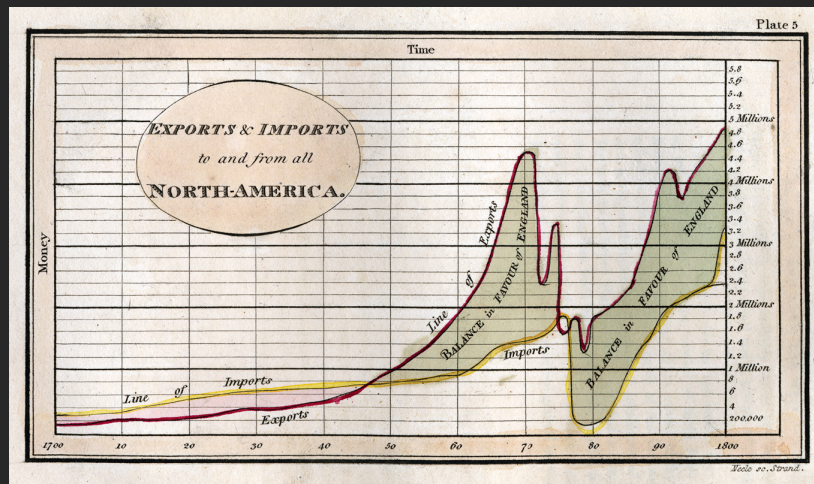


12



# What is a good representation?

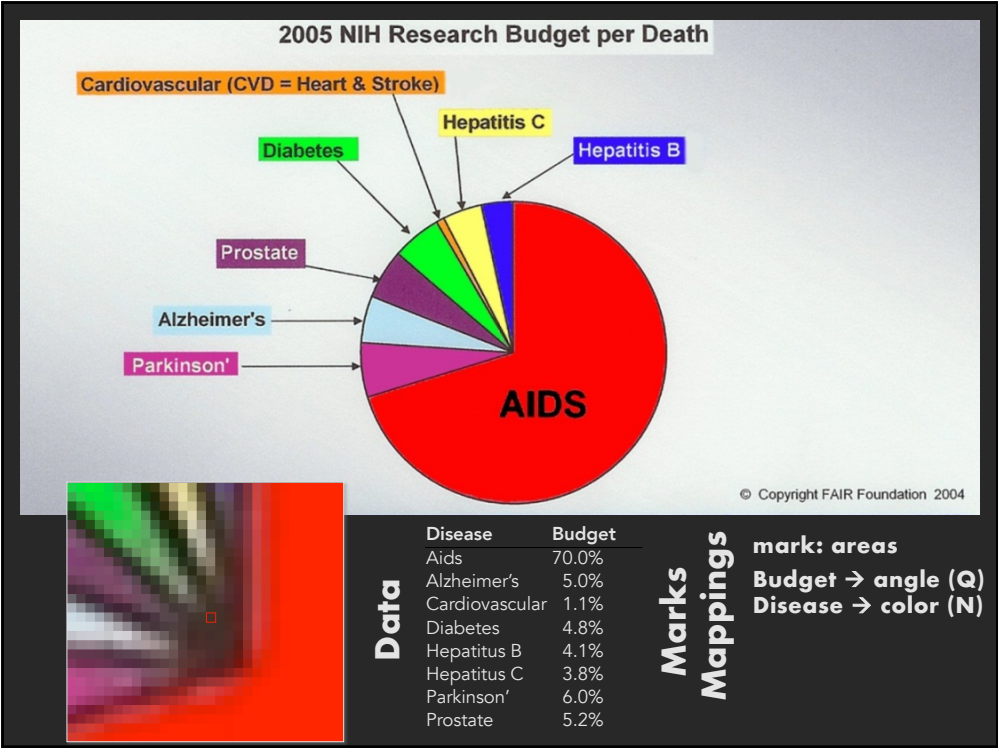
13



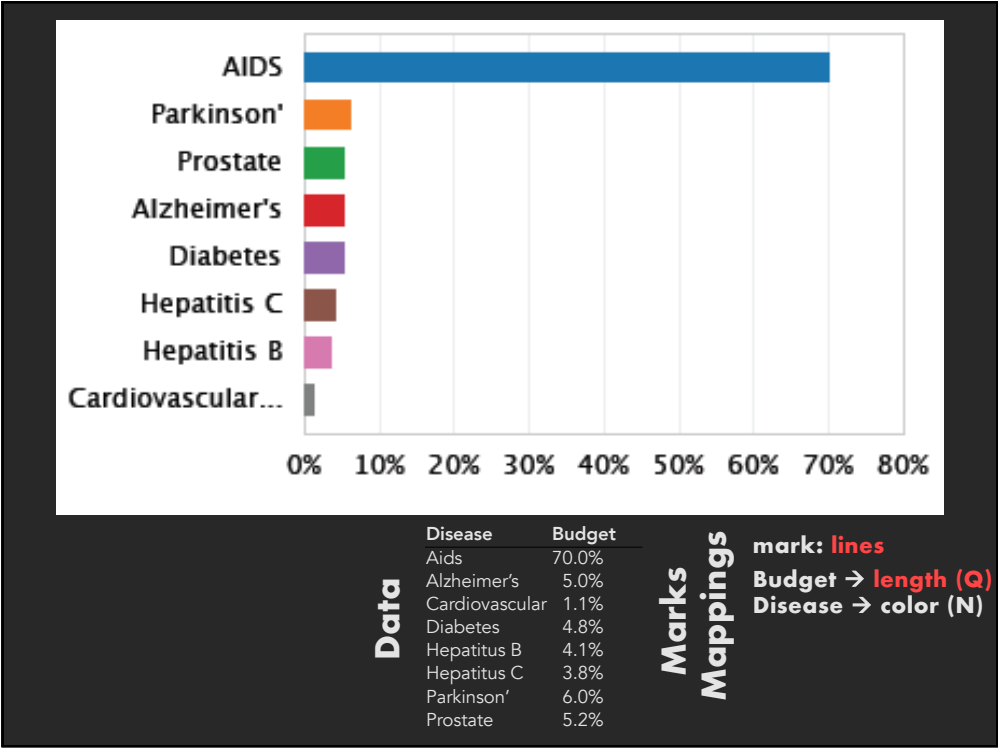
Year	Exports	Imports
1700	170,000	300,000
1701	171,000	302,000
1702	176,000	303,000
1703	180,000	312,000
1704	187,000	319,000
...	...	...

**Marks** mark: lines  
**Mappings** Year → x-pos (Q)  
 Exports → y-pos (Q)  
 Imports → y-pos (Q)  
 Exports → color (N)  
 Imports → color (N)

14



15



16



# Approach

---

**Classification:** Determine chart type

**Mark extraction:** Retrieve graphical marks

**Data extraction:** Retrieve underlying data table

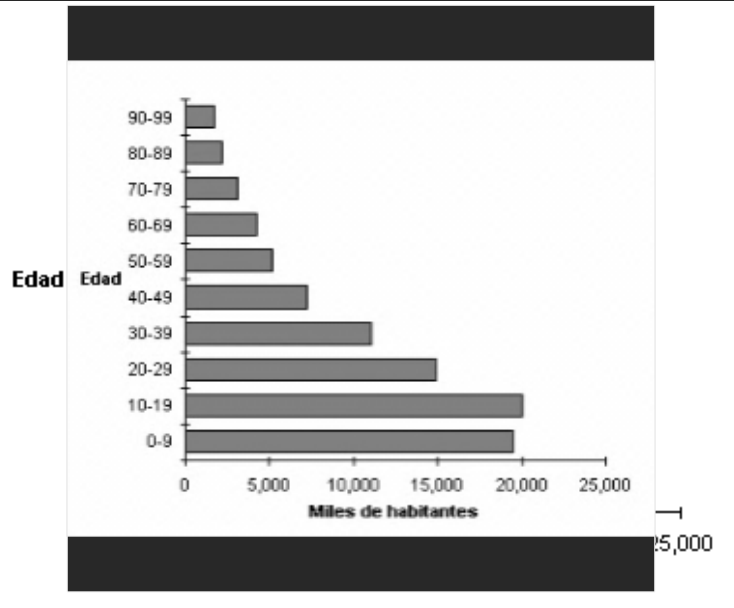
17

17

# Classification

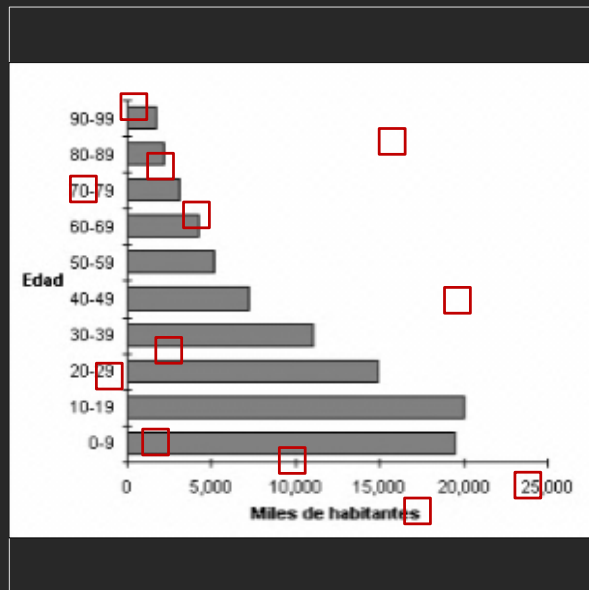
18

# Training the Classifier



19

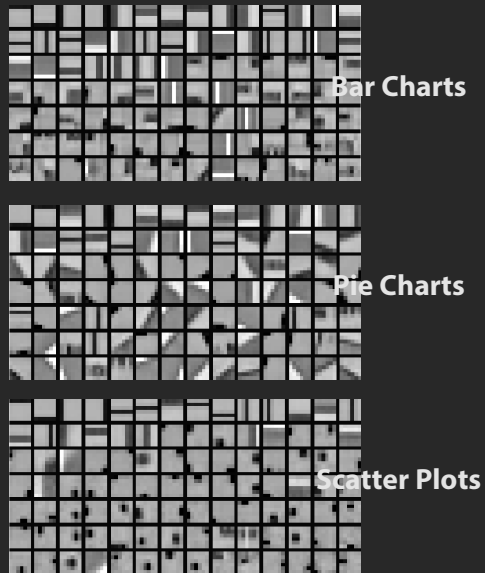
# Training the Classifier



20

# Training the Classifier

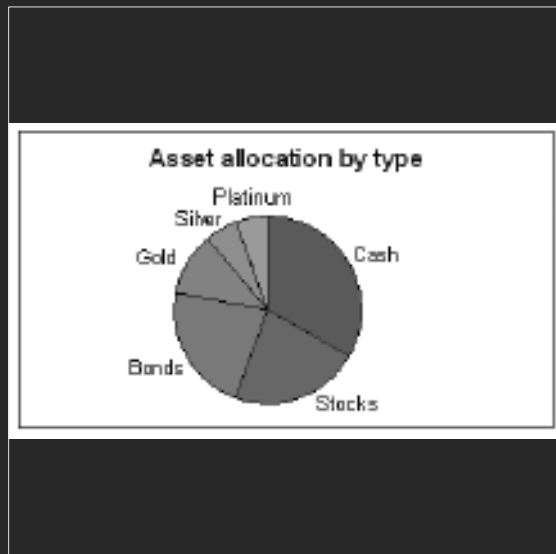
---



21

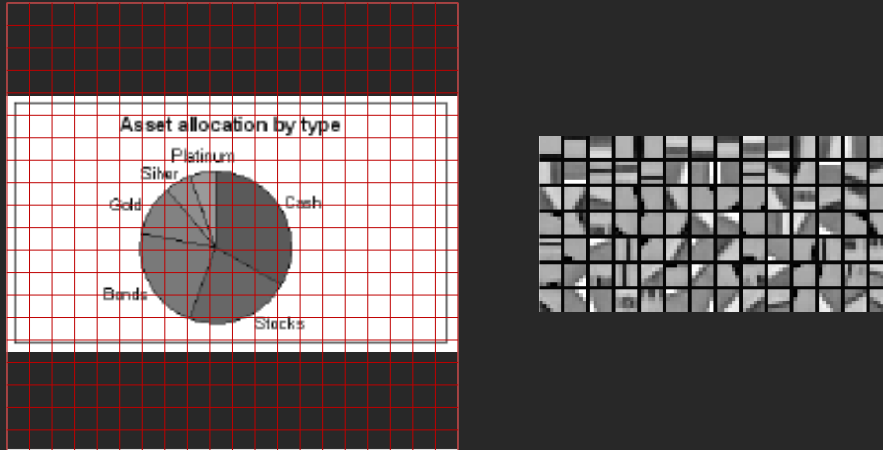
# Classifying an Input Image

---



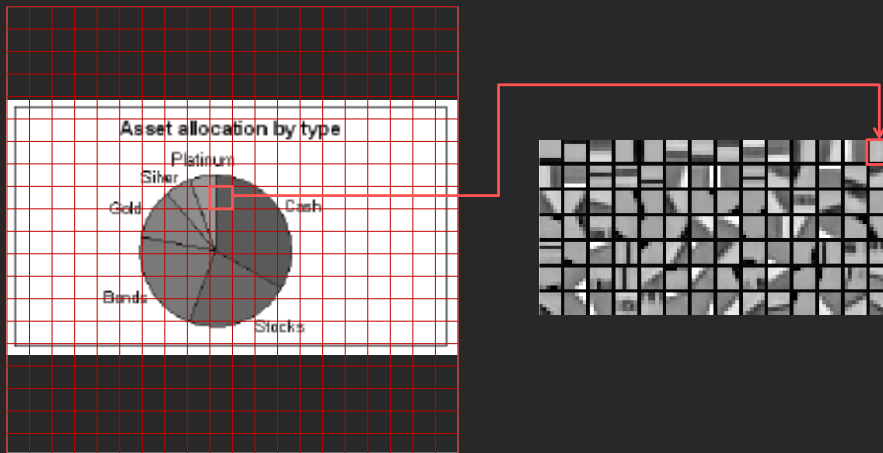
22

# Classifying an Input Image



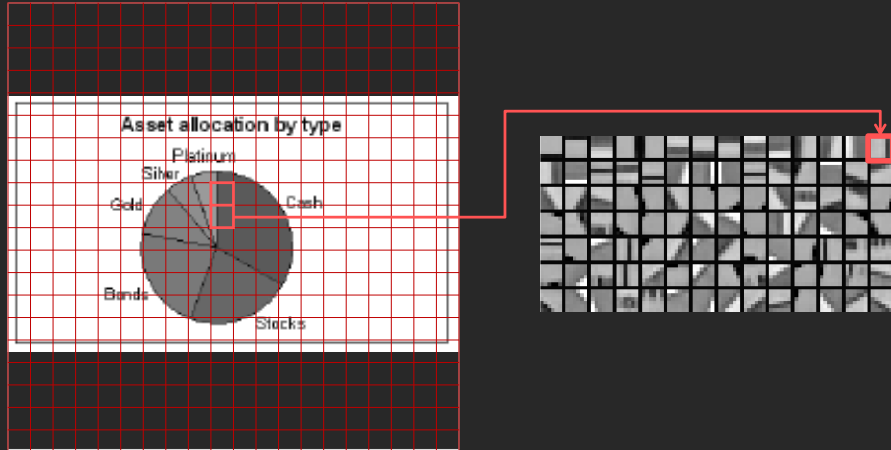
23

# Classifying an Input Image



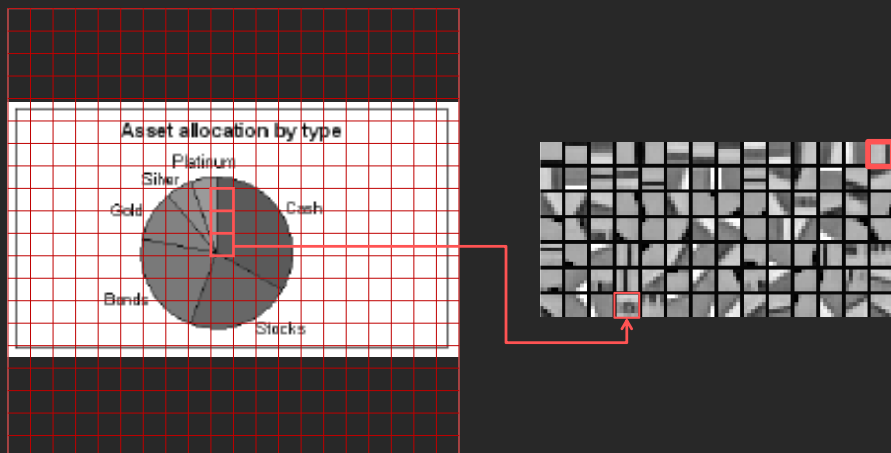
24

# Classifying an Input Image



25

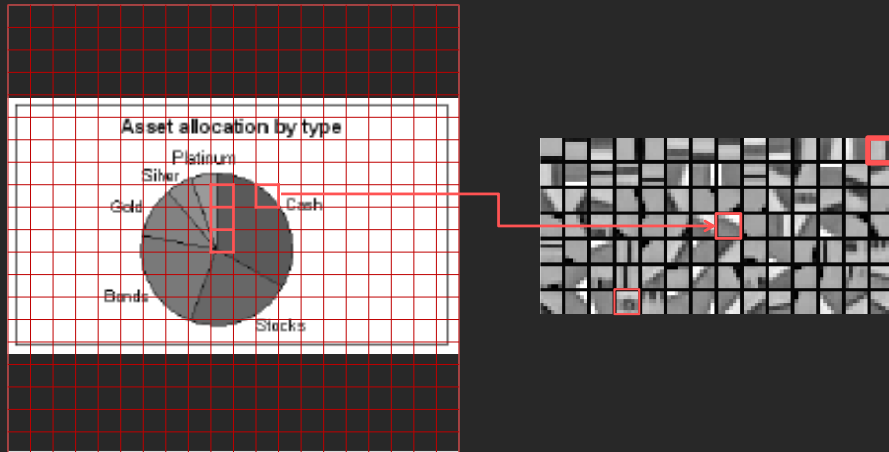
# Classifying an Input Image



26

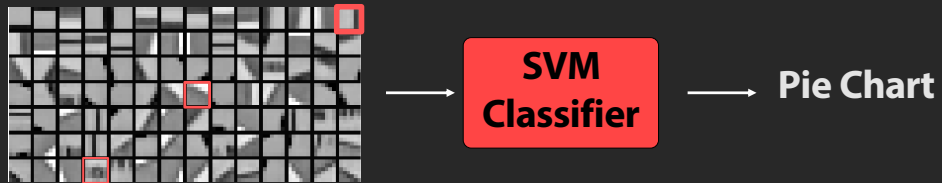


# Classifying an Input Image



27

# Classifying an Input Image

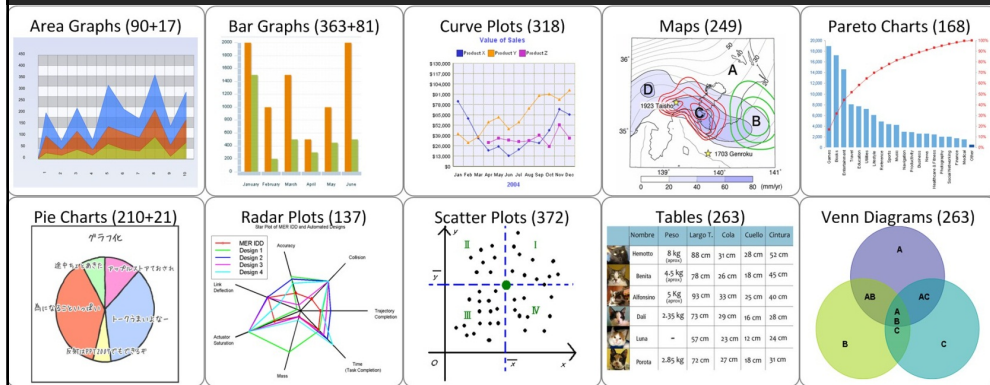


Corpus: 667 charts, 5 chart types [Prasad 2007]	Average Accuracy
[Prasad 2007] Multi-class SVM	84%
ReVision: Multi-class SVM	88%
ReVision: Binary SVM (yes/no for each chart type)	96%

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

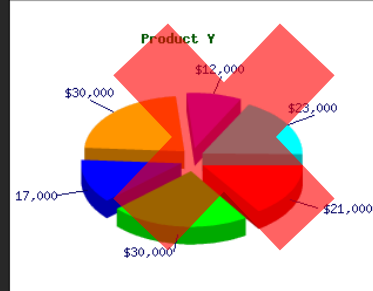
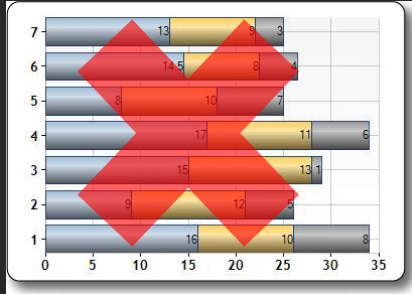
# Mark and Data Extraction

30

# Assumptions

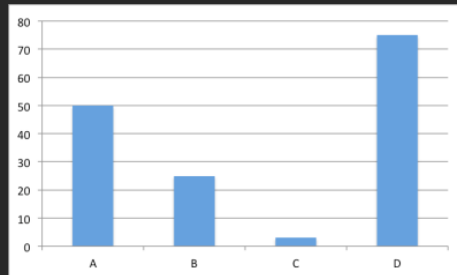
Bar charts and pie charts only

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



31

# Bar Charts



marks: lines

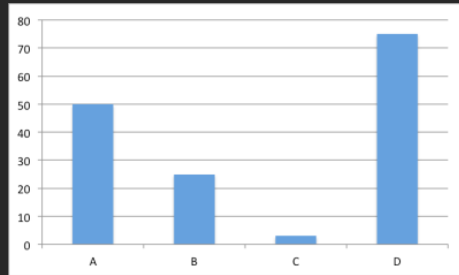


y-value x-value

50	A
25	B
4	C
75	D

32

# Bar Charts



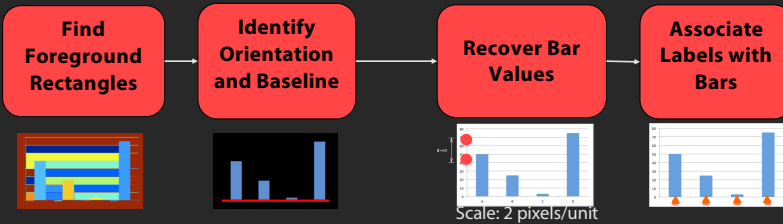
marks: lines



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

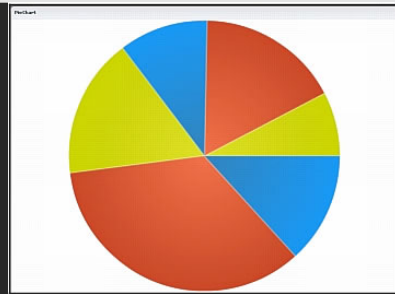
Extract Marks

Extract Data



43

# Pie Charts



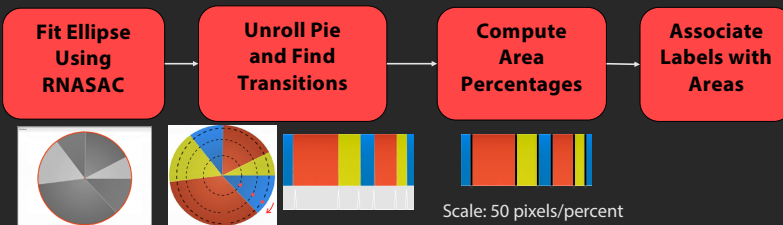
marks: areas



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

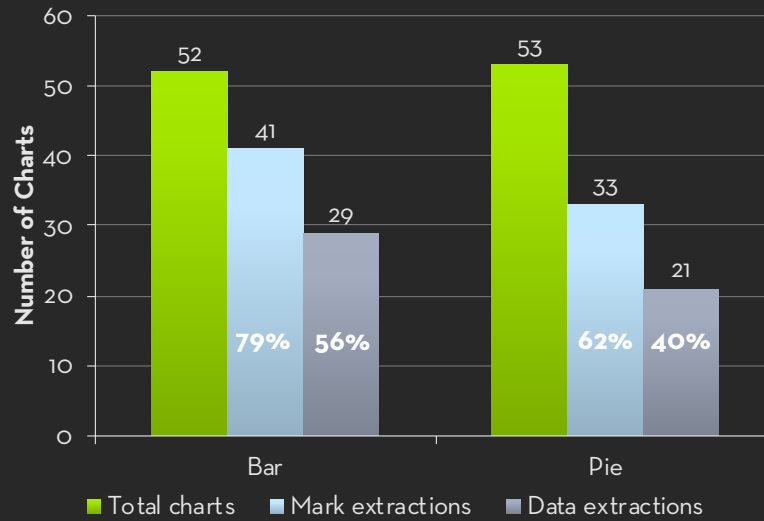
Extract Marks

Extract Data



44

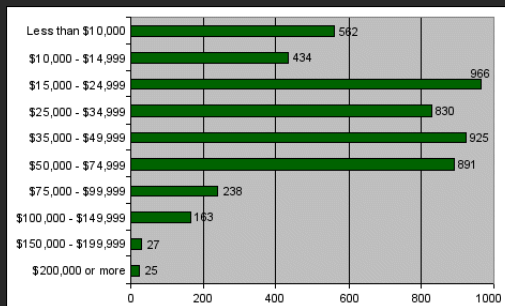
## Extraction Results



45

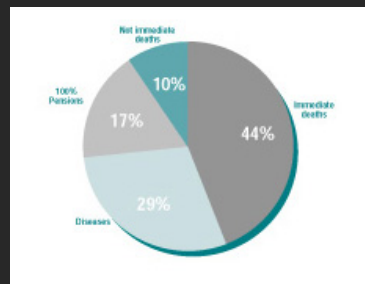
## Data Extraction Error

### Bar Charts



7.7%

### Pie Charts



4.6%

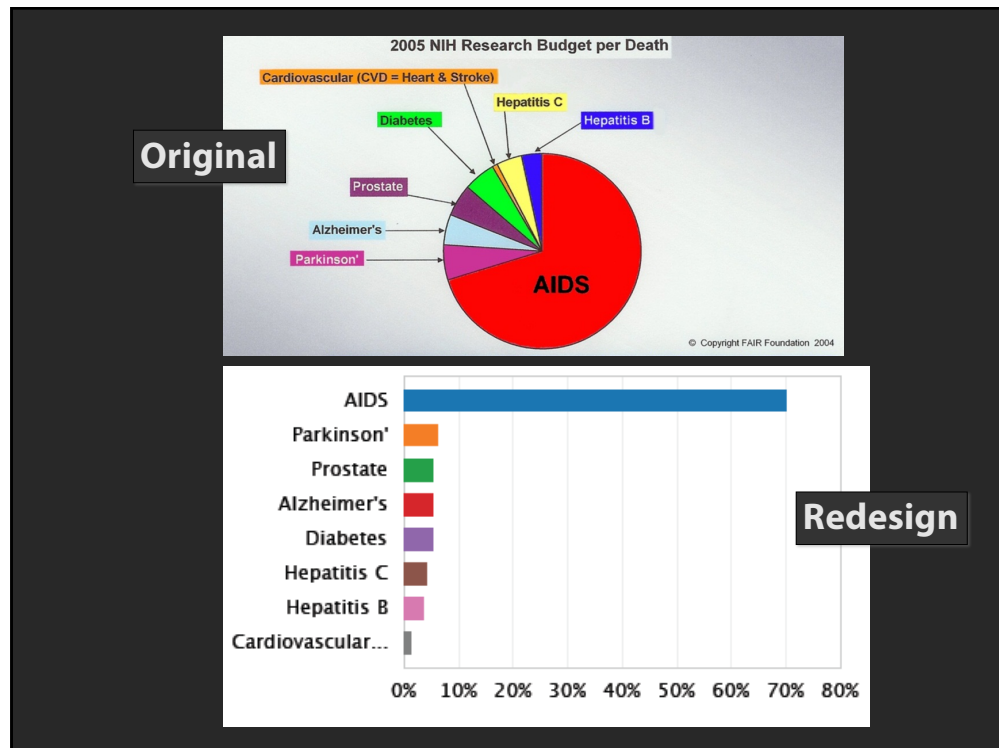
Average chart size: 342 x 452 pixels [Prasad 2007]

46

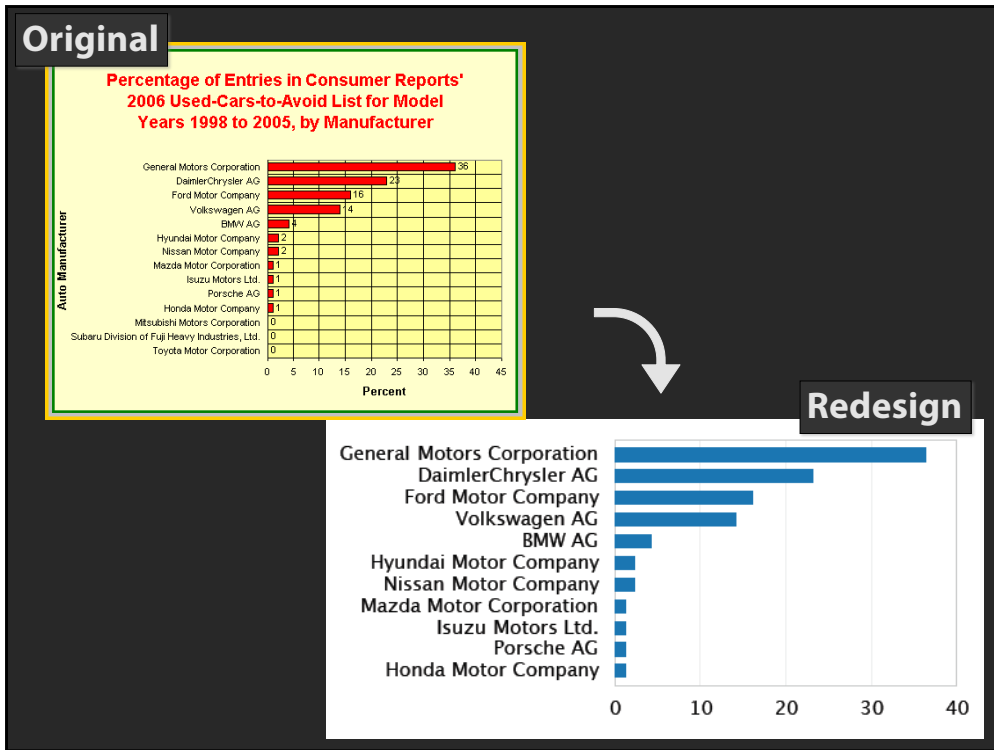


# Redesign

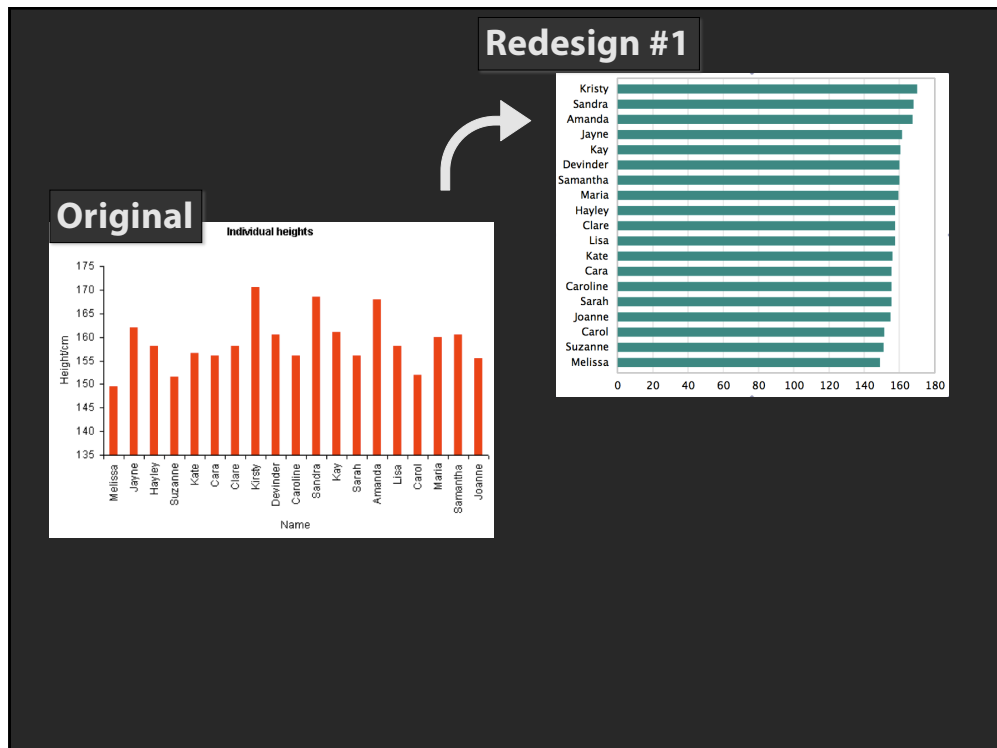
47



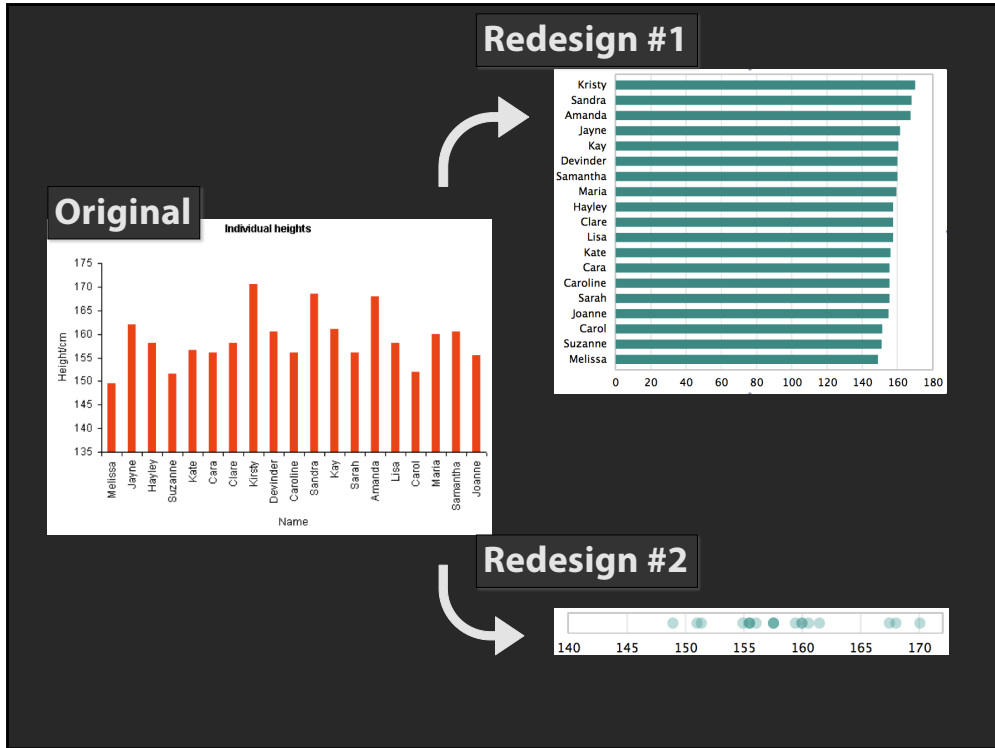
48



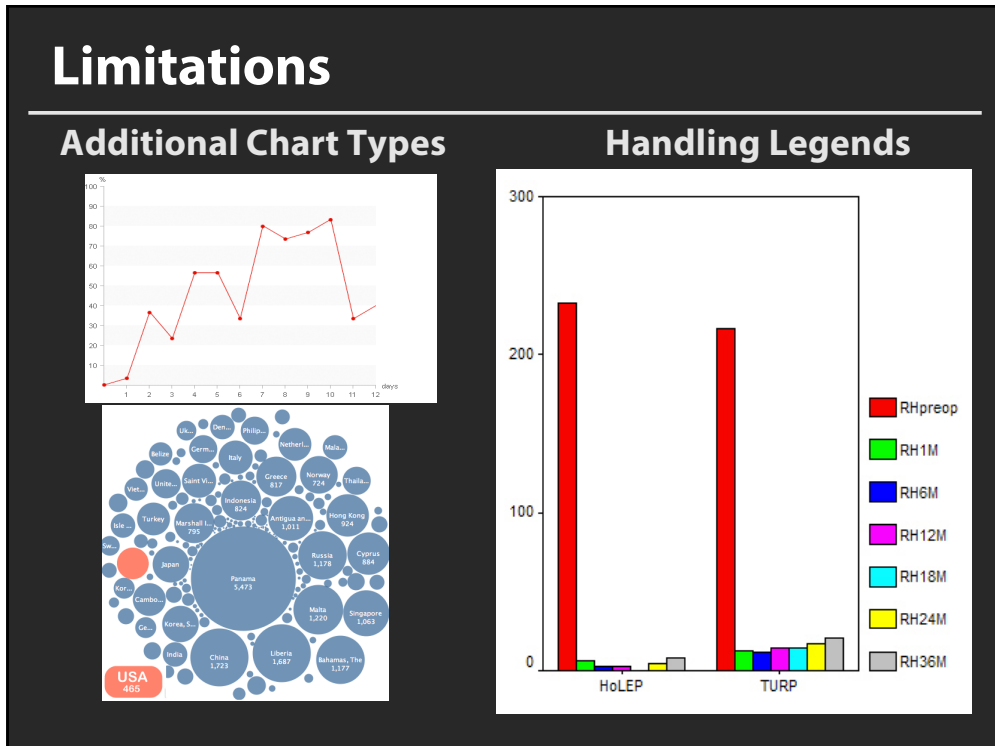
49



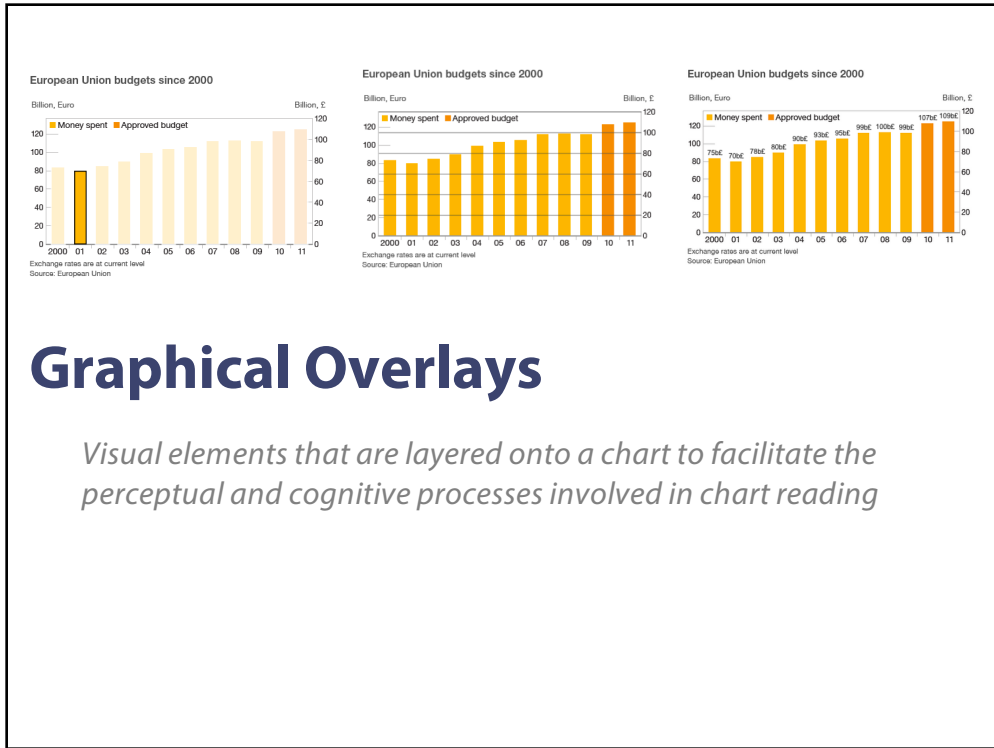
50



51



53



# Graphical Overlays

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

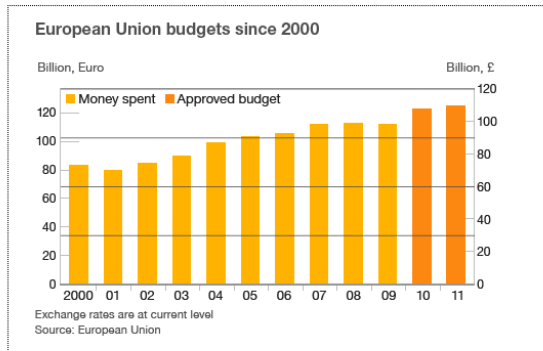
54



55

## 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.

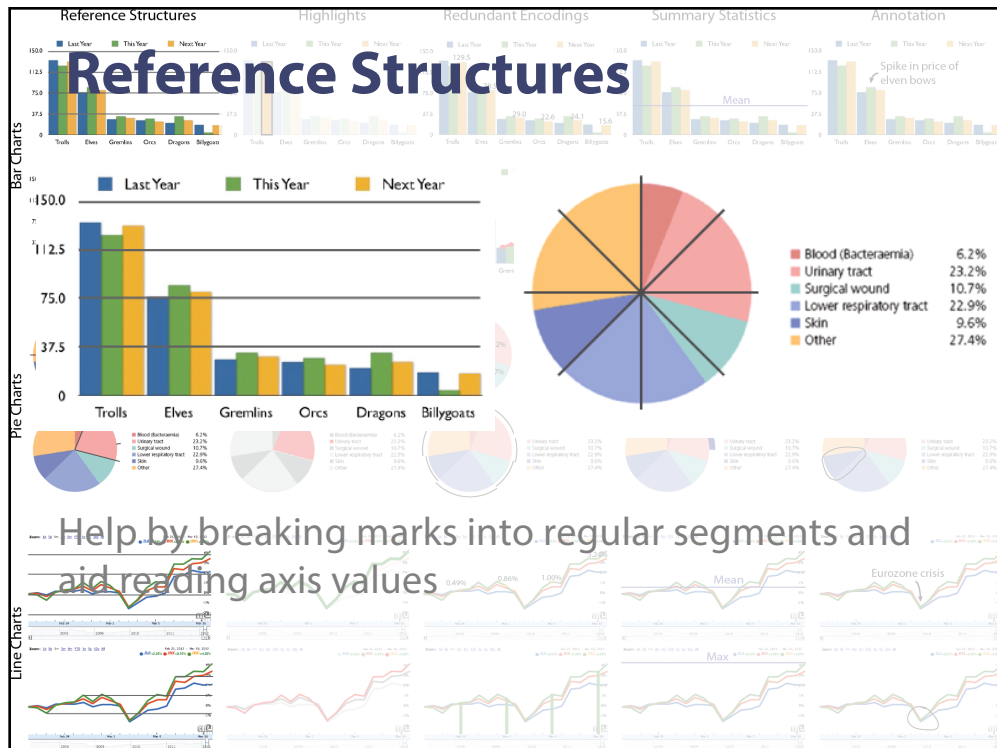


- Chart type: Bar
- Chart: 00281
- Overlay type: Reference structures
- Regular gridlines
  - Lines emanating from marks
- Parameters
- Overlay  Underlay
  - Static  Interactive
  - Divisions: 4
  - Line thickness: 1

Places regular gridlines at user defined intervals.

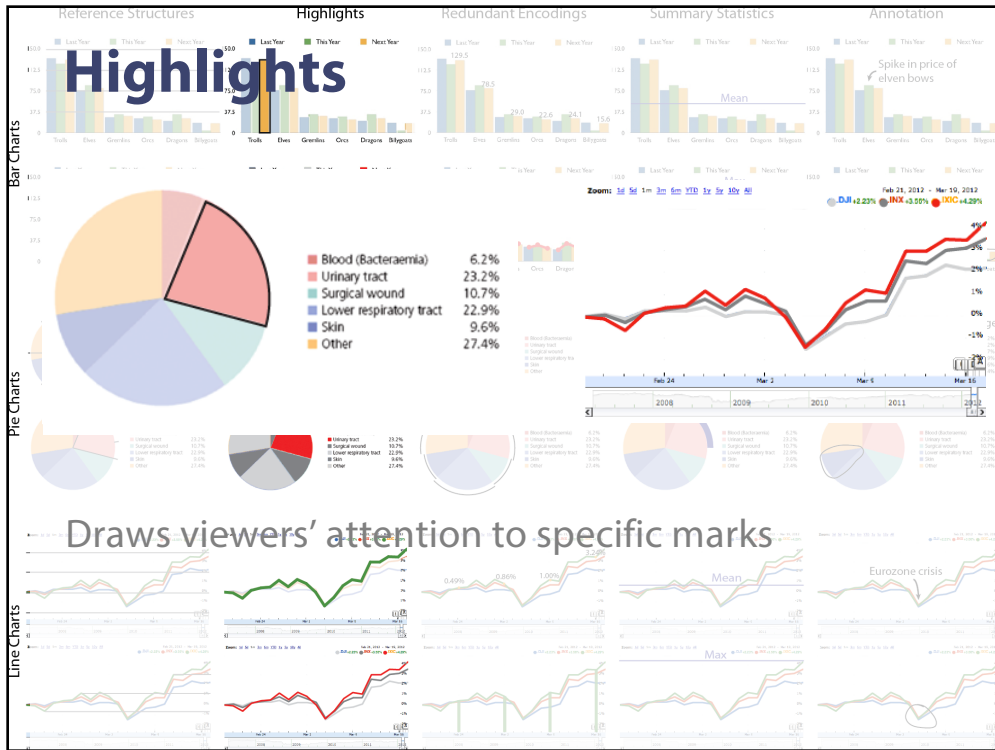
# Demo

56



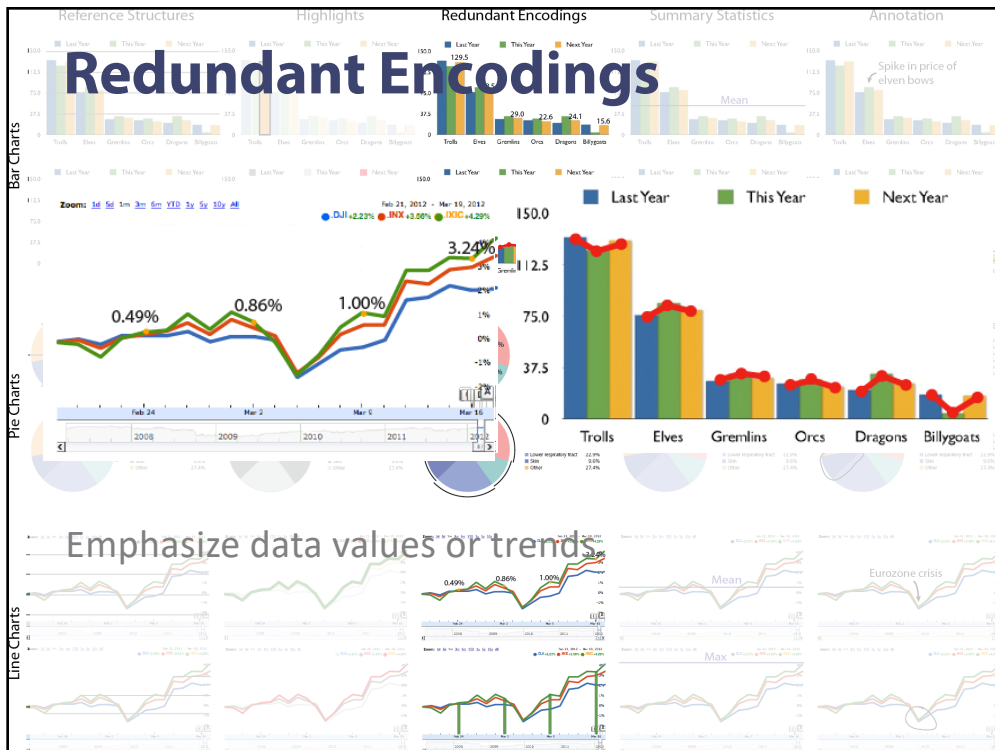
57





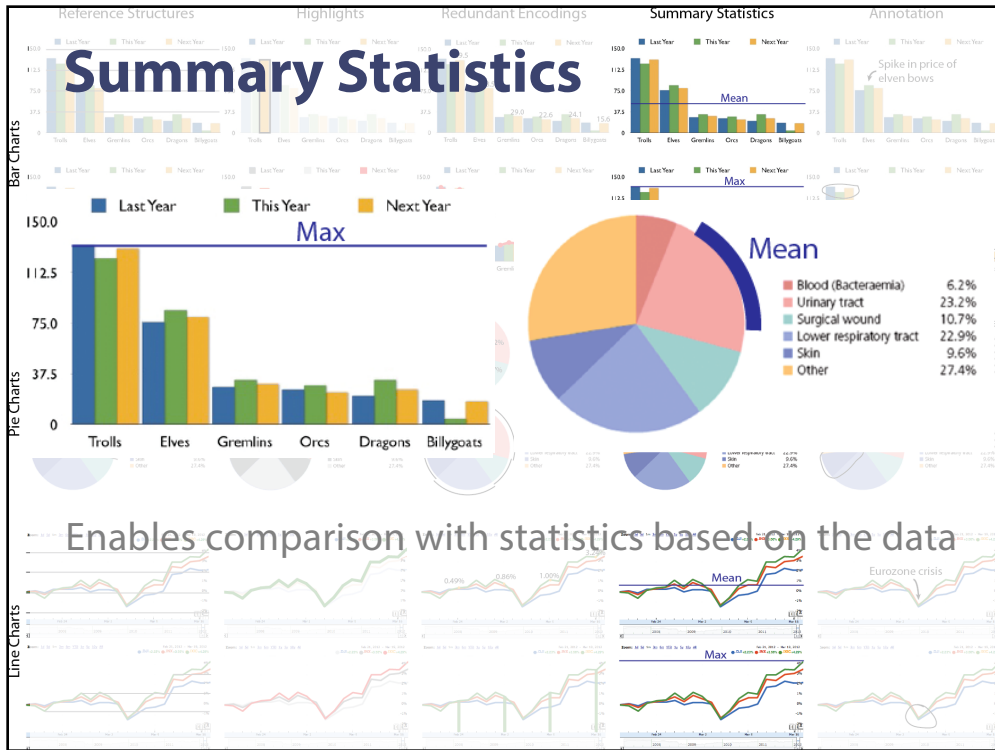
Draws viewers' attention to specific marks

58

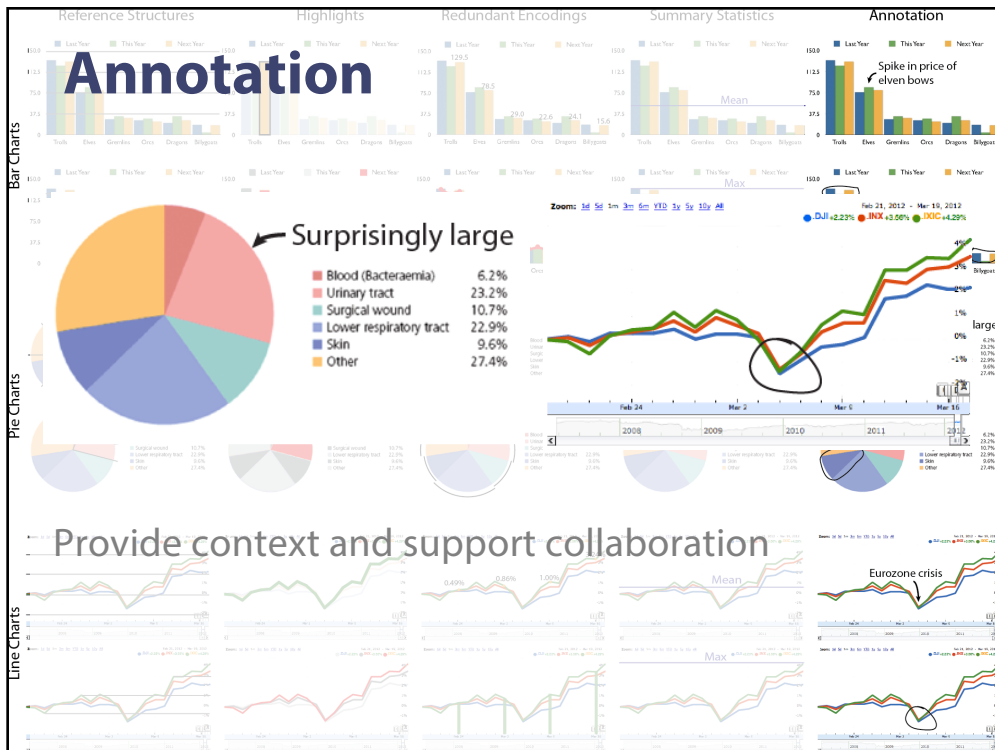


Emphasize data values or trends

59



60



61

### European Union budgets since 2000

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

Exchange rates are at current level  
Source: European Union

**Most overlays only require access to marks**

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

62

## 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

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

Region/Country	Amount
Egypt	1,700,000
Iraq	15,804,557
Turkey	22,094,911
Regional	28,029,722
Lebanon	37,812,334
Jordan	40,221,893

SOURCE: UNHCR  
GET THE DATA EMBED FULLSCREEN

theguardian

63

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

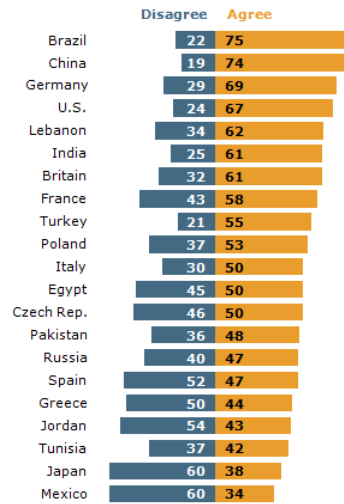
64

**Problem:** Diversity of writing styles

65

## Example 1: Pew Research

### Are People Better Off in Free Market Economy?



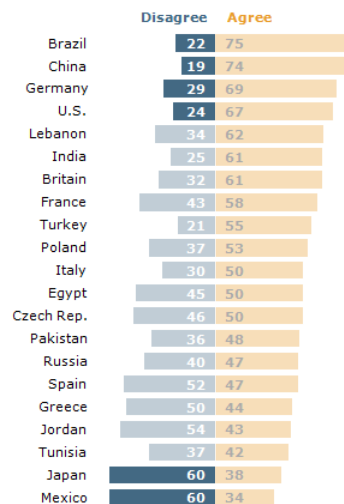
PEW RESEARCH CENTER Q26.

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 1: Pew Research

### Are People Better Off in Free Market Economy?



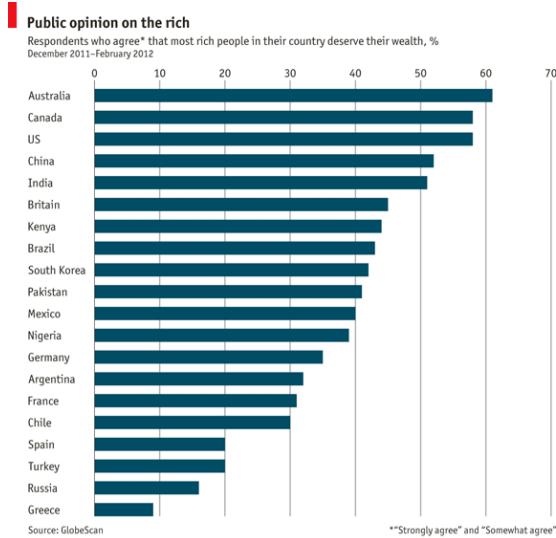
PEW RESEARCH CENTER Q26.

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%)**.

67



## Example 2: Economist

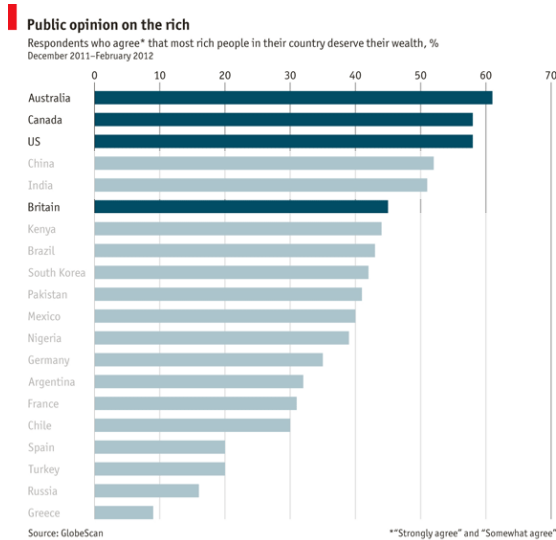


Top earners have attracted more opprobrium as their salaries and the performance of the economy have headed in opposite directions.

Europeans and Latin Americans tend to have similar attitudes to the rich; the Anglo-Saxon world is a bit more forgiving.

68

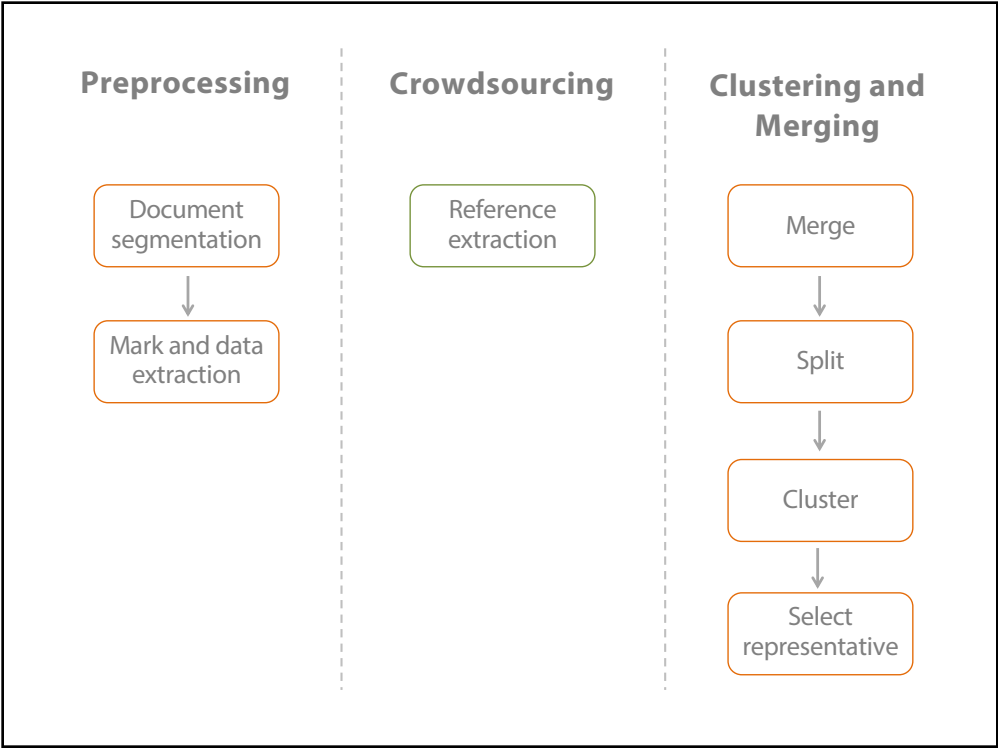
## Example 2: Economist



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

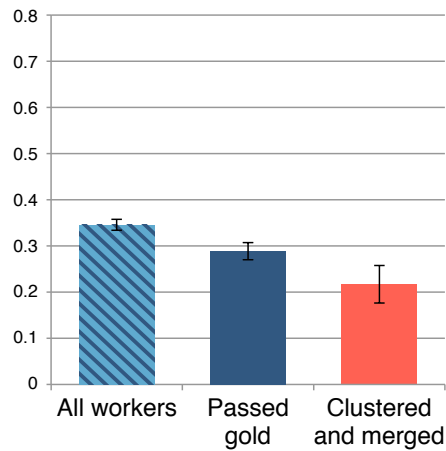


70

**Demo**

71

## Evaluation



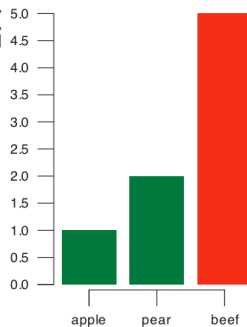
Avg. F<sub>1</sub> distance: expert specified references vs. crowd specified references

72

## 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  $\hookrightarrow$  fill
- cost  $\hookrightarrow$  height
- cost  $\hookrightarrow$  yPos
- cost  $\hookrightarrow$  area
- deconID  $\hookrightarrow$  xPos

D3 Code

D3 Chart

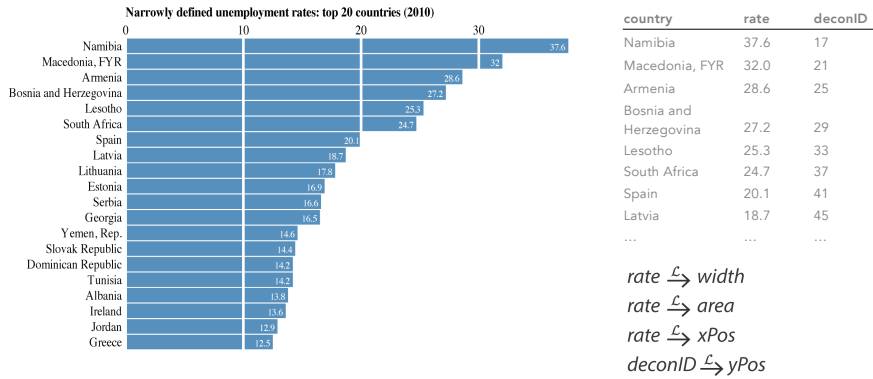
Our Deconstruction

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.

74

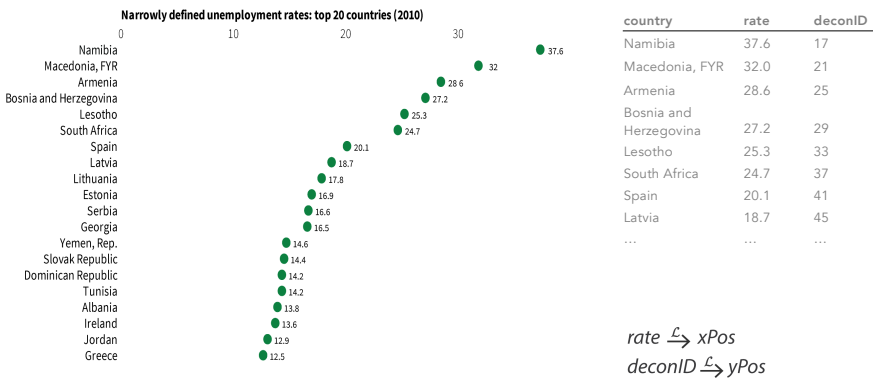
# Deconstructing D3 Charts



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

75

# Deconstructing D3 Charts



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

76

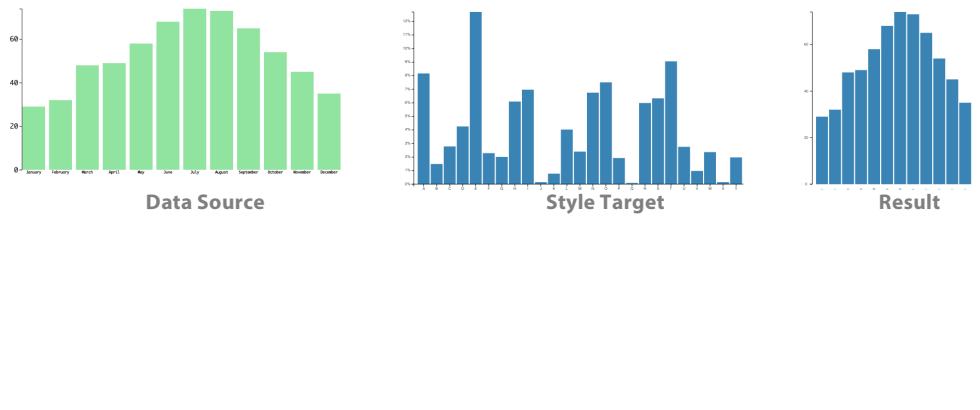
# Automatic Redesign

Can we automatically redesign charts to improve

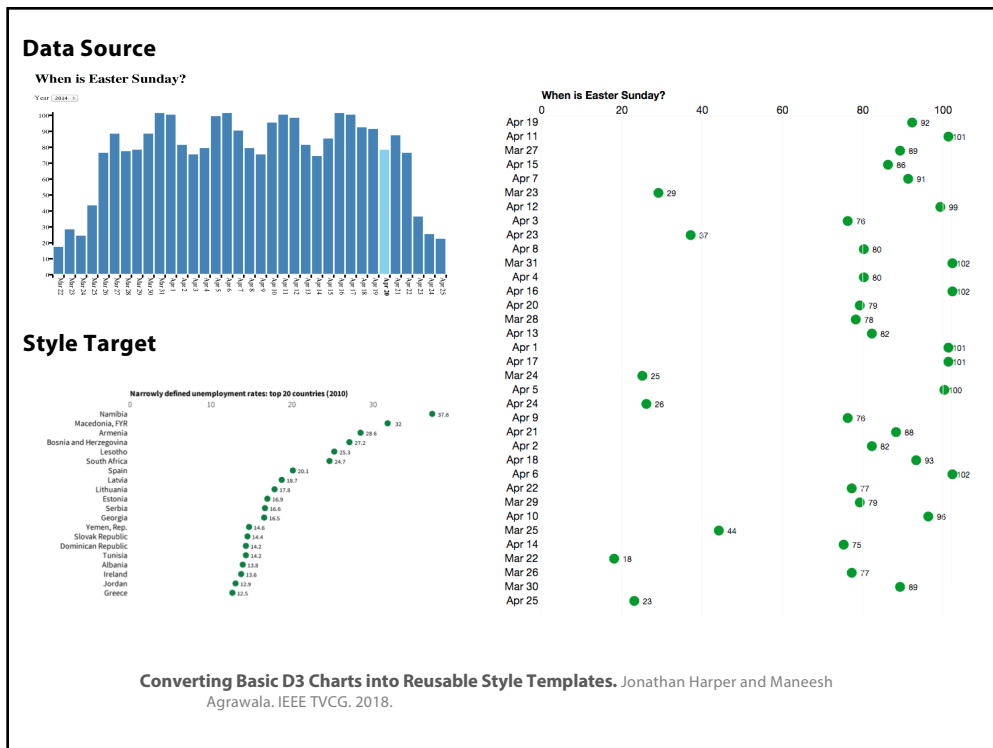
Perceptual effectiveness?

Visual aesthetics?

Accessibility for vision impaired users?



77

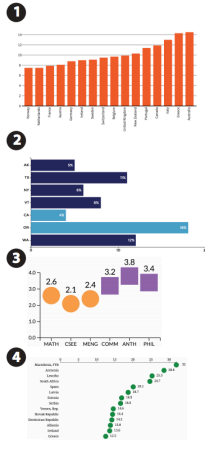


78

# Reusable Style Templates

Year (1)	Weight (2)	MPG (3)	HP (4)	Year (1)	Prisoners (2)	Pop (3)	Cost (4)
1970	3441	17.6	149	1990	204	1683	60
1971	2960	21.3	105	1991	2166	1725	80
1972	3237	18.7	120	1992	2223	1768	20
1973	3419	17.1	130	1993	2416	1813	70
1974	2878	22.7	94	1994	2742	1858	110
...	...	...	...	...	...	...	...

## Style Templates



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

79

# Document Collections

**The Rise of the College Graduate**  
Today's Millennials are the best-educated generation in history, fully a third (34%) have at least a bachelor's degree. In contrast, only 12% of age- to 29-year-olds in the same generation had a college degree, a proportion that increased to 24% in the late 1990s and 34% in the late 2000s.

**White Education Levels of 25- to 39-year-olds Have risen Dramatically Across the Generations ...**

**The Declining Value of a High School Diploma**  
The explanation for this troubling finding lies in another major economic trend reshaping the economic landscape: The dramatic decline in the value of high school education. While earnings of those with a college degree rose, the typical high school graduate's earnings fell by more than 50% since 1980. This decline, the Pew Research Center found, has been large enough to nearly offset the gains of college graduates.

**The Widening Earnings Gap of Young Adults by Educational Attainment**  
The difference in median annual earnings of college and high school graduates when members of each generation were age 25 to 29.

**U.S. Births by Investigation**  
1970: 8.1, 1980: 6.4, 1990: 7.0, 2000: 9.6, 2010: 6.8

**Other Labor Market Outcomes**  
To be sure, the Great Recession and possibly slow recovery have taken their toll on the Millennial generation, including the college-educated. Young college graduates are having more difficulty finding work than earlier cohorts. They are more likely to be unemployed and have to search longer for a job than earlier generations of young adults.

## 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

80

## 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***