

Overview of InfoVis



CS 7450 - Information Visualization
Aug. 25, 2011
John Stasko

Exercise



- Get out pencil and paper



Fall 2011

CS 7450

3

Data Explosion

- Society is more complex
 - There simply is more “stuff”
- Computers, internet and web give people access to an incredible amount of data
 - news, sports, financial, purchases, etc...



Fall 2011

CS 7450

4

How Much Data? (1)



- Estimated info added to digital universe each year will soon approach 1 ZB (zettabyte)*
 - 10000000000000000000000 (10²¹) bytes
 - From: http://www.emc.com/digital_universe viewed December 8, 2008

*But only half that goes to my email inbox

Fall 2011

CS 7450

5

How Much Data? (2)



- 6 million FedEx transactions per day
<http://www.fedex.com/us/about/today/companies/corporation/facts.html>
- Average of 98 million Visa credit-card transactions per day in 2005
<http://www.corporate.visa.com/md/nr/press278.jsp>
- Average of 5.4 petabytes of data crosses AT&T's network per day
<http://att.sbc.com/gen/investor-relations?pid=5711>
- Average of 610 to 1110 billion e-mails worldwide per year (based on estimates in 2000)
<http://www2.sims.berkeley.edu/research/projects/how-much-info/internet.html>

Fall 2011

CS 7450

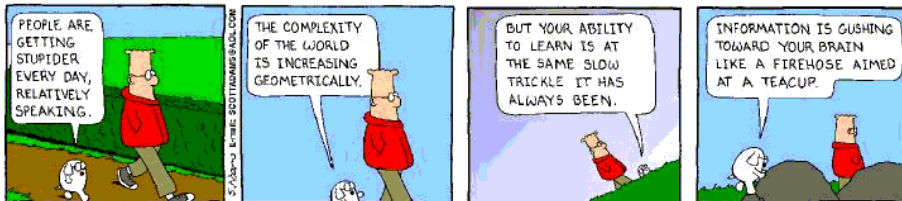
Slide courtesy Jim Thomas

6

Data Overload



- Confound: How to make use of the data
 - How do we make sense of the data?
 - How do we harness this data in decision-making processes?
 - How do we avoid being overwhelmed?



Fall 2011

CS 7450

7

The Challenge



- Transform the *data* into *information* (understanding, insight) thus making it useful to people



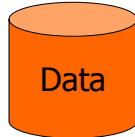
Fall 2011

CS 7450

8

The Problem

Web,
Books,
Papers,
Game scores,
Scientific data,
Biotech,
Shopping
People
Stock/finance
News



Data Transfer

How?



Vision: 100 MB/s
Ears: <100 b/s
Telepathy
Haptic/tactile
Smell
Taste

Two slides courtesy
of Chris North

Fall 2011

CS 7450

9

Human Vision

- Highest bandwidth sense
- Fast, parallel
- Pattern recognition
- Pre-attentive
- Extends memory and cognitive capacity
(Multiplication test)
- People think visually

Impressive. Lets use it!

Fall 2011

CS 7450

10

Some Examples



- Why visualization helps...

Fall 2011

CS 7450

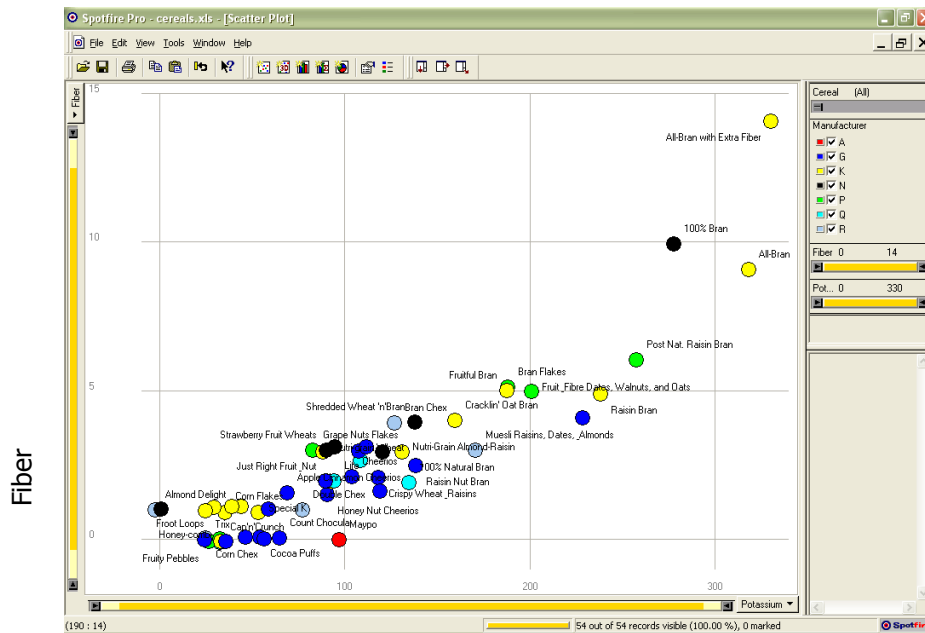
11

Questions:

Which cereal has the most/least potassium?
Is there a relationship between potassium and fiber?
If so, are there any outliers?
Which manufacturer makes the healthiest cereals?



	A	B	C	D					
1	Cereal	Manufacturer	Fiber	Potassium	28	Honey-comb	P	0	35
2	100% Bran	N	10	280	29	Just Right Fruit & Nut	K	2	95
3	100% Natural Bran	Q	2	135	30	Life	Q	2	95
4	All-Bran	K	9	320	31	Lucky Charms	G	0	55
5	All-Bran with Extra Fiber	K	14	330	32	Maypo	A	0	95
6	Almond Delight	R	1	0	33	Muesli Raisins, Dates, &	R	3	170
7	Apple Cinnamon Cheerios	G	1.5	70	34	Multi-Grain Cheerios	G	2	90
8	Bran Chex	R	4	125	35	Nutri-Grain Almond-Rais	K	3	130
9	Bran Flakes	P	5	190	36	Nutri-grain Wheat	K	3	90
10	Cap'n Crunch	Q	0	35	37	Oatmeal Raisin Crisp	G	1.5	120
11	Cheerios	G	2	105	38	Post Nat. Raisin Bran	P	6	260
12	Cocoa Puffs	G	0	55	39	Product 19	K	1	45
13	Corn Chex	R	0	25	40	Quaker Oatmeal	Q	2.7	110
14	Corn Flakes	K	1	35	41	Raisin Bran	K	5	240
15	Count Chocula	G	0	65	42	Raisin Nut Bran	G	2.5	140
16	Cracklin' Oat Bran	K	4	160	43	Rice Krispies	K	0	35
17	Cream of Wheat (Quick)	N	1	0	44	Shredded Wheat	N	3	95
18	Crispy Wheat & Raisins	G	2	120	45	Shredded Wheat 'n Bran	N	4	140
19	Double Chex	R	1	80	46	Shredded Wheat spoon	N	3	120
20	Froot Loops	K	1	30	47	Smacks	K	1	40
21	Frosted Flakes	K	1	25	48	Special K	K	1	55
22	Fruit & Fibre Dates, Wal	P	5	200	49	Strawberry Fruit Wheats	N	3	90
23	Fruitful Bran	K	5	190	50	Total Corn Flakes	G	0	35
24	Fruity Pebbles	P	0	25	51	Total Raisin Bran	G	4	230
25	Golden Grahams	G	0	45	52	Total Whole Grain	G	3	110
26	Grape Nuts Flakes	P	3	85	53	Trix	G	0	25
27	Honey Nut Cheerios	G	1.5	90	54	Wheaties	G	3	110
					55	Wheaties Honey Gold	G	1	60



Fall 2011

Potassium

CS 7450

13

Even Tougher?

- What if you could only see one cereal's data at a time? (e.g. some websites)
- What if I read the data to you?



Fall 2011

CS 7450

14



Another Illustrative Example

Fall 2011

CS 7450

15

Four Data Sets



- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: $y = 3 + 0.5x$
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression line) = 13.75
- Correlation coefficient = 0.82
- Coefficient of determination = 0.67

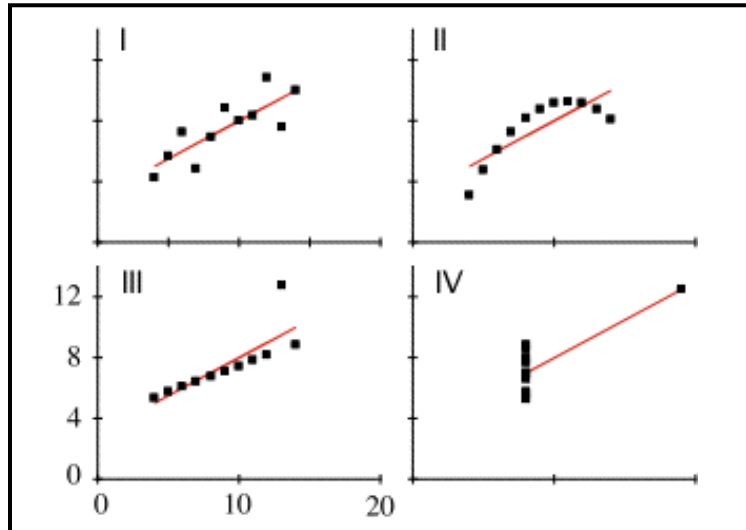
<http://astro.swarthmore.edu/astro121/anscombe.html>

Fall 2011

CS 7450

16

The Data Sets



Fall 2011

CS 7450

17

The Values

1	2	3	4
10.0, 8.04	10.0, 9.14	10.0, 7.46	8.0, 6.58
8.0, 6.95	8.0, 8.14	8.0, 6.77	8.0, 5.76
13.0, 7.58	13.0, 8.74	13.0, 12.74	8.0, 7.71
9.0, 8.81	9.0, 8.77	9.0, 7.11	8.0, 8.84
11.0, 8.33	11.0, 9.26	11.0, 7.81	8.0, 8.47
14.0, 9.96	14.0, 8.10	14.0, 8.84	8.0, 7.04
6.0, 7.24	6.0, 6.13	6.0, 6.08	8.0, 5.25
4.0, 4.26	4.0, 3.10	4.0, 5.39	19.0, 12.50
12.0, 10.84	12.0, 9.13	12.0, 8.15	8.0, 5.56
7.0, 4.82	7.0, 7.26	7.0, 6.42	8.0, 7.91
5.0, 5.68	5.0, 4.74	5.0, 5.73	8.0, 6.89

Fall 2011

CS 7450

18

Exercise Redux



- Let's check what you did...
- People work differently

Fall 2011

CS 7450

19

Visualization



- Definition
 - “The use of computer-supported, interactive visual representations of data to amplify cognition.”
From [Card, Mackinlay Shneiderman '98]

Fall 2011

CS 7450

20

Visualization



- Often thought of as process of making a graphic or an image
- Really is a cognitive process
 - Form a mental image of something
 - Internalize an understanding
- “The purpose of visualization is insight, not pictures”
 - Insight: discovery, decision making, explanation

Fall 2011

CS 7450

21

Main Idea



- Visuals help us think
 - Provide a frame of reference, a temporary storage area
- Cognition → Perception
- Pattern matching
- External cognition aid
 - Role of external world in thinking and reason

Larkin & Simon '87
Card, Mackinlay, Shneiderman '98

Fall 2011

CS 7450

22

Expressed Well



“Contained within the data of any investigation is information that can yield conclusions to questions not even originally asked. That is, there can be surprises in the data...To regularly miss surprises by failing to probe thoroughly with visualization tools is terribly inefficient because the cost of intensive data analysis is typically very small compared with the cost of data collection.”

W. Cleveland
The Elements of Graphing Data

Fall 2011

CS 7450

23

Part of our Culture



- “I see what you’re saying”
- “Seeing is believing”
- “A picture is worth a thousand words”

Fall 2011

CS 7450

24

Purpose



- Two main uses of infovis
 - Analysis – Understand your data better and act upon that understanding
 - Presentation – Communicate and inform others more effectively

1. Analysis



- Given all the data, then
 - understand, compare, decide, judge, evaluate, assess, determine, ...
- Ultimately, about solving problems

When to Apply?



- Many other techniques for data analysis
 - Statistics, DB, data mining, machine learning
- Visualization most useful in **exploratory data analysis**
 - Don't know what you're looking for
 - Don't have a priori questions
 - Want to know what questions to ask

"A graphic display has many purposes but it achieves its highest value when it forces us to see what we were not expecting."

H. Wainer

Fall 2011

CS 7450

27

EDA Example 1



- Business
 - Why has Hyundai made such great strides in the US market?
 - How influential was their "Lose your job, we'll buy the car back" campaign?
 - Have their cars improved in quality? If so, in what major ways?
 - Is the Genesis as good of a car as the Lexus ES?

Fall 2011

CS 7450

28

EDA Example 2



- Airlines
 - What are the key factors causing flight delays in the US?
 - Are delays worse in the summer or winter?
 - Is the seasonal effect influenced by geographic location?
 - How does competition at an airport affect flight delays?

Fall 2011

CS 7450

29

More on EDA



“Information visualization is ideal for exploratory data analysis. Our eyes are naturally drawn to trends, patterns, and exceptions that would be difficult or impossible to find using more traditional approaches, such as tables or text, including pivot tables. When exploring data, even the best statisticians often set their calculations aside for a while and let their eyes take the lead.”

S. Few
Now you see it

Fall 2011

CS 7450

30

Tasks for Info Vis?



- Search (OK)
 - Finding a specific piece of information
 - How many games did the Braves win in 1995?
 - What novels did Ian Fleming author?
- Browsing (Better)
 - Look over or inspect something in a more casual manner, seek interesting information
 - Learn about crystallography
 - What has Jane been up to lately?

Fall 2011

CS 7450

31

Tasks in Info Vis



- Analysis
 - Comparison-Difference
 - Outliers, Extremes
 - Patterns
- Assimilation
- Monitoring
- Awareness

More to come in a future class...

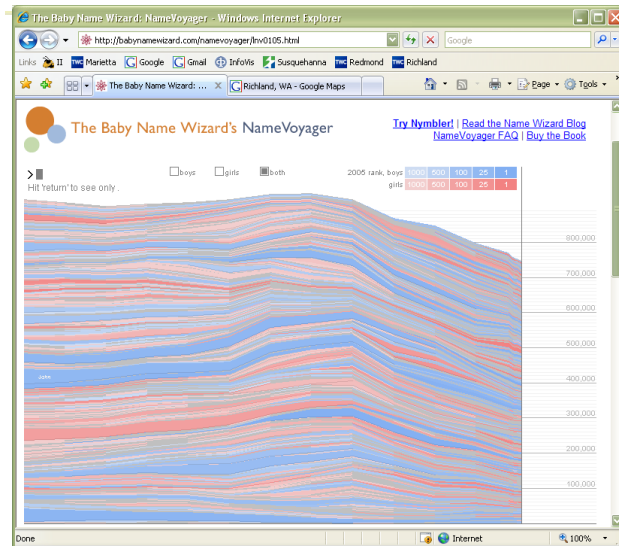
Fall 2011

CS 7450

32

<http://babynamewizard.com/namevoyager/>

Baby Name Wizard



Fall 2011

CS 7450

33

2. Presentation

- Use visualization to communicate ideas, influence, explain, persuade
- Visuals can serve as evidence or support

Fall 2011

CS 7450

When to Apply?



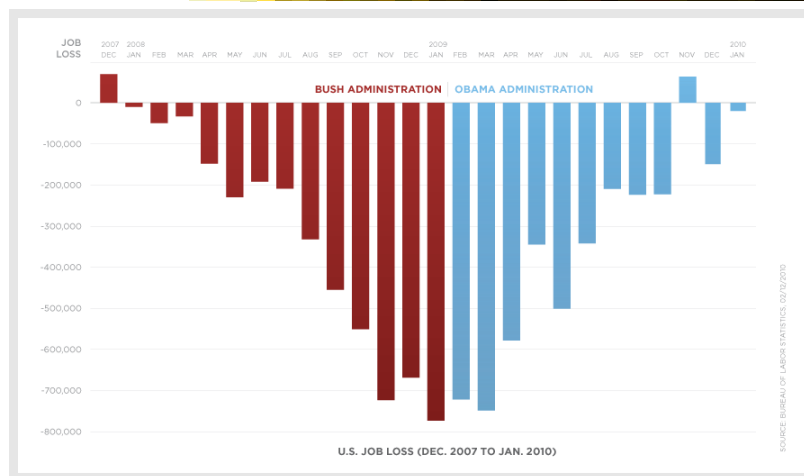
- Visuals can frequently take the place of many words
- Visuals can summarize, aggregate, unite, explain, ...
- Sometimes words are needed, however

Fall 2011

CS 7450

35

Job Losses



Controversial, see <http://soquelbythecreek.blogspot.com/2010/02/what-does-obama-job-chart-really-mean.html>

Fall 2011

CS 7450

36

Key Benefits of Visualization



- Facilitating awareness and understanding
- Helping to raise new questions and supply answers
- Generating insights
- Telling a story and making a point

Fall 2011

CS 7450

37

Key Challenge

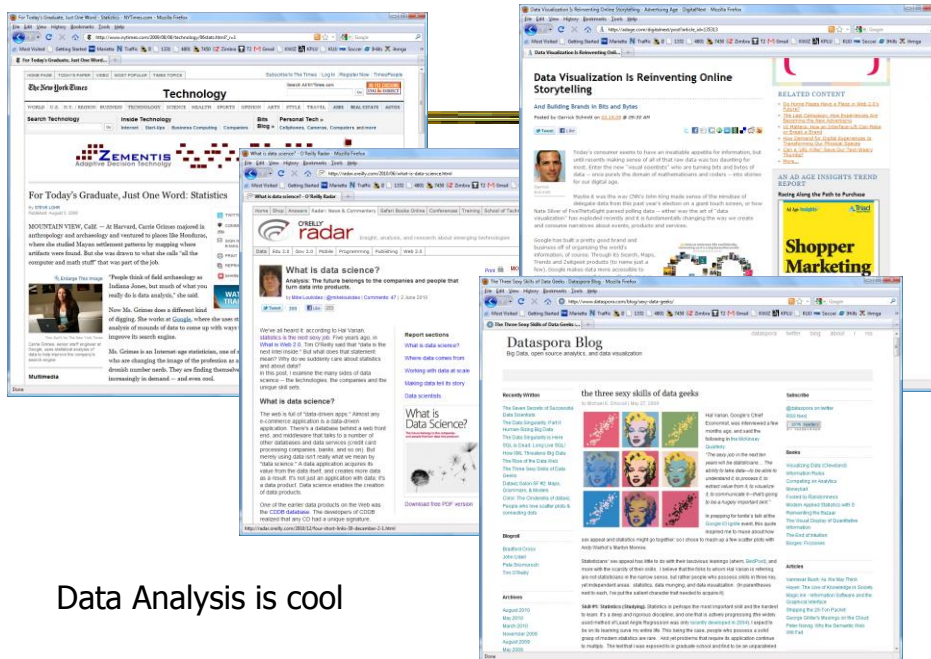


- How to measure and prove?
 - All those benefits are not easily quantifiable and measured
- Evaluation is perhaps primary open research challenge for visualization

More to come later in term

Fall 2011

CS 7450



Data Analysis is cool

Fall 2011

CS 7450

39

Academic Areas

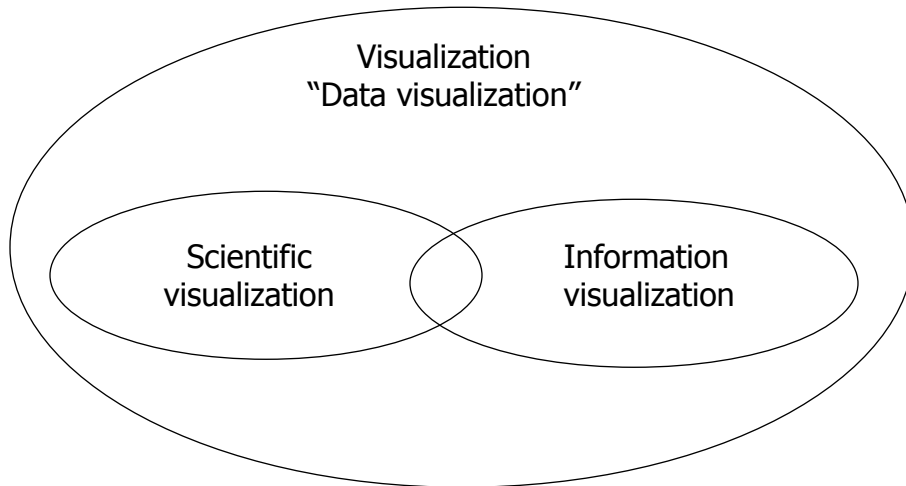
- Where does InfoVis fit in the academic world?

Fall 2011

CS 7450

40

Overview



Fall 2011

CS 7450

41

Scientific Visualization



- Primarily relates to and represents something physical or geometric
 - Often 3-D
 - Examples
 - Air flow over a wing
 - Stresses on a girder
 - Torrents inside a tornado
 - Organs in the human body
 - Molecular bonding

Not the focus of this class

Fall 2011

CS 7450

42

Information Visualization



- 1. What is “information”?
 - Items, entities, things which do not have a direct physical correspondence
 - Notion of abstractness of the entities is important too
 - Examples: baseball statistics, stock trends, connections between criminals, car attributes...

Fall 2011

CS 7450

43

Information Visualization



- 2. What is “visualization”?
 - The use of computer-supported, interactive visual representations of data to amplify cognition.
From [Card, Mackinlay Shneiderman '98]

Fall 2011

CS 7450

44

Information Visualization



- Characteristics:
 - Taking items without a direct physical correspondence and mapping them to a 2-D or 3-D physical space.
 - Giving information a visual representation that is useful for analysis and presentation
 - “A key challenge in information visualization is designing a cognitively useful spatial mapping of a dataset that is not inherently spatial and accompanying the mapping by interaction techniques that allow people to intuitively explore the dataset. Information visualization draws on the intellectual history of several traditions, including computer graphics, human-computer interaction, cognitive psychology, semiotics, graphic design, statistical graphics, cartography, and art.”
<http://conferences.computer.org/infovis/>

Fall 2011

CS 7450

45

Constituents



- Two key aspects of infovis
 - Representation
 - Interaction (too often overlooked)

“The effectiveness of information visualization hinges on two things: its ability to clearly and accurately represent information and our ability to interact with it to figure out what the information means.”

S. Few, Now you see it

Fall 2011

CS 7450

46

Two Key Challenges



- Scale
 - Challenge often arises when data sets become large
- Diversity
 - Data of data types, forms, sizes

Fall 2011

CS 7450

47

Example Domains for Info Vis



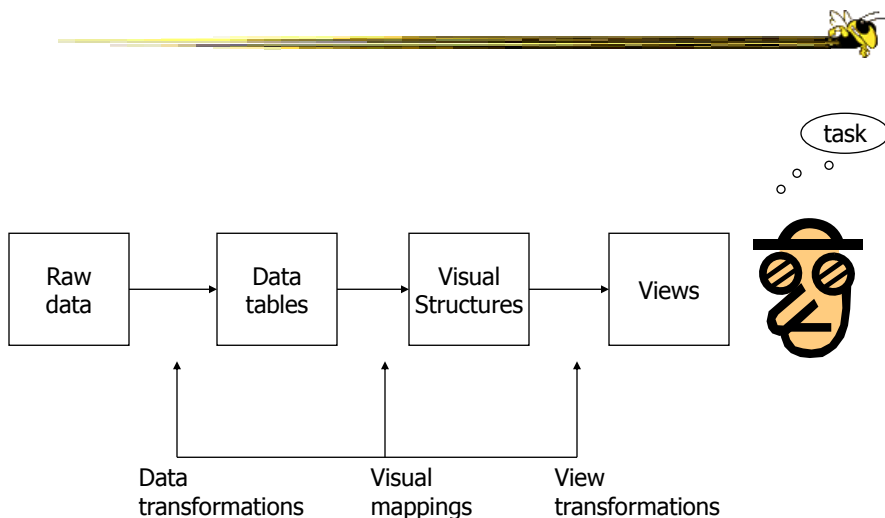
- Text
- Statistics
- Financial/business data
- Internet information
- Software
- ...

Fall 2011

CS 7450

48

InfoVis Process Model



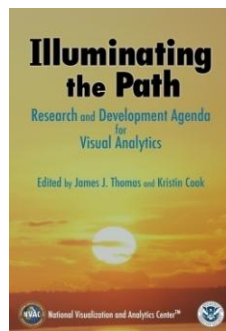
Fall 2011

CS 7450

49

New Area Emerging: Visual Analytics

Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces



Available at <http://nvac.pnl.gov/>
in PDF form

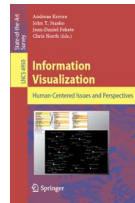
Fall 2011

CS 7450

Alternate Definition



Visual analytics combines automated analysis techniques with interactive visualizations for an effective understanding, reasoning and decision making on the basis of very large and complex data sets

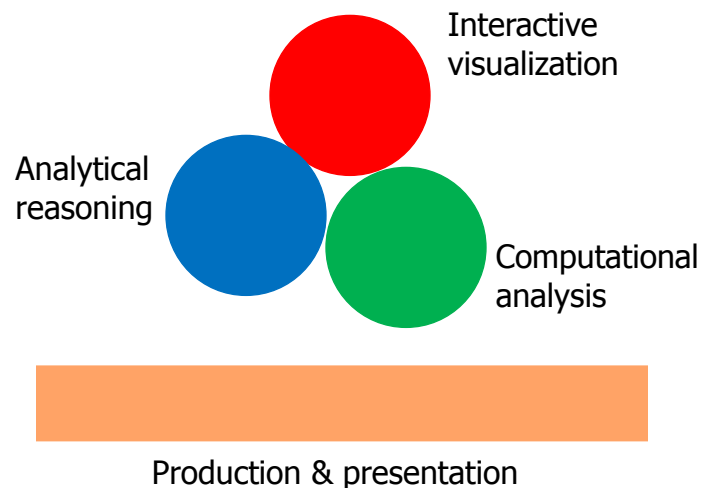


Keim et al
chapter in *Information Visualization: Human-Centered Issues and Perspectives*, 2008

Fall 2011

CS 7450

Main Components



Fall 2011

CS 7450

Back to InfoVis (Examples)



- Start with static pictures (InfoGraphics)
 - Very popular on the web
 - But are they information visualizations?

Fall 2011

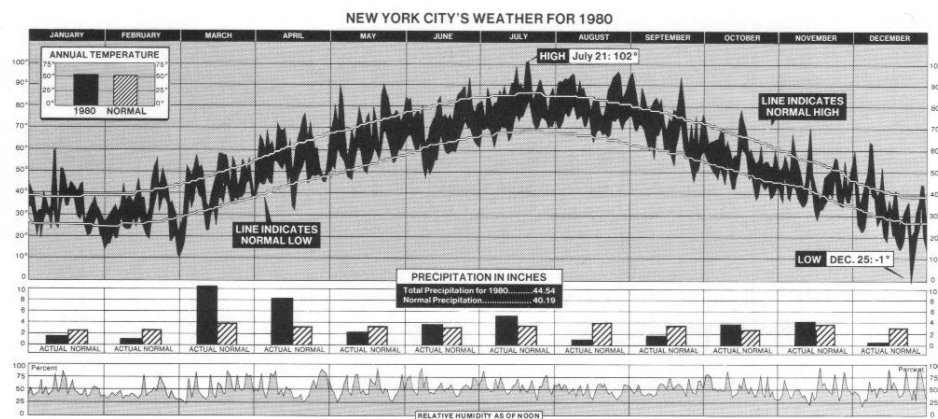
CS 7450

53

NYC Weather



2220 numbers



New York Times, January 11, 1981, p. 32.

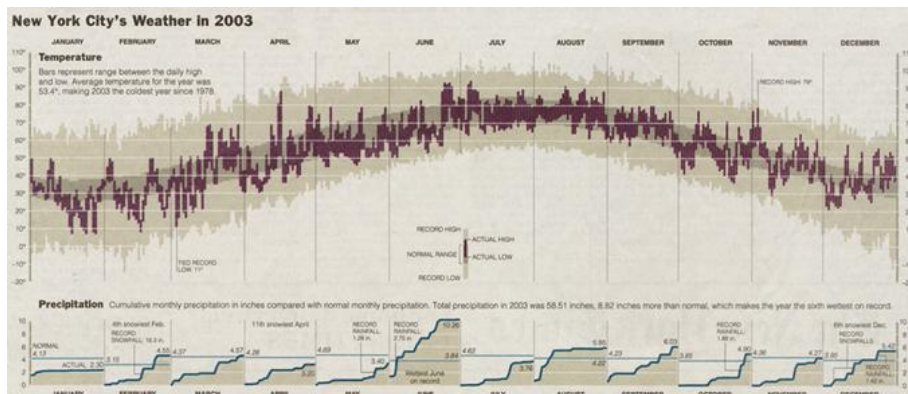
Tufte, Vol. 1

Fall 2011

CS 7450

54

Updated Version



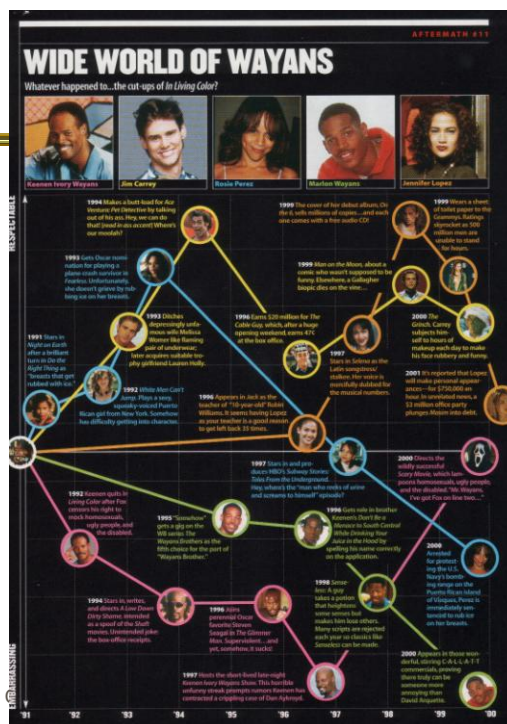
http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00014g

Fall 2011

CS 7450

55

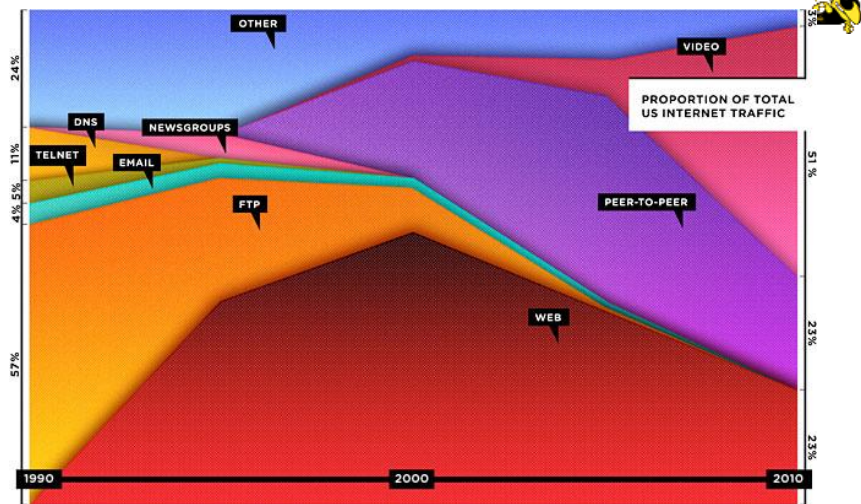
In Living Color



Maxim Magazine, July '01

Fall 2011

Internet Traffic



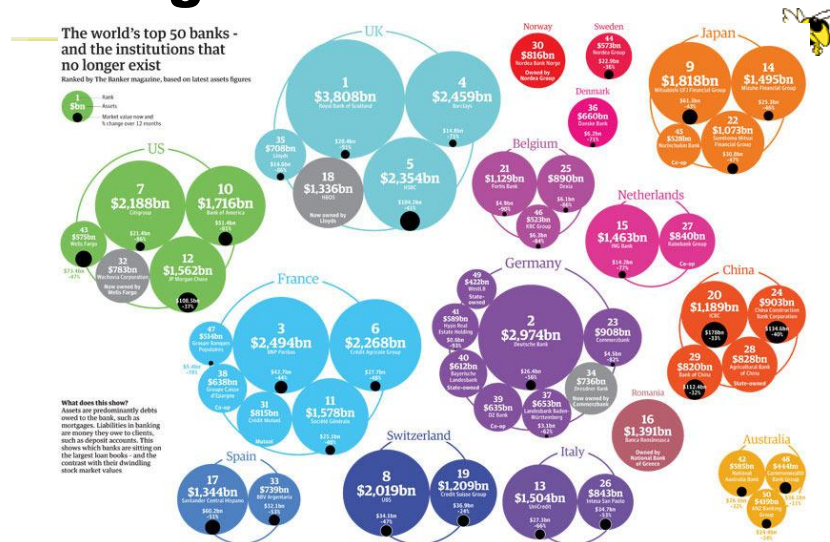
http://www.wired.com/magazine/2010/08/ff_webrip/all/1

Fall 2011

CS 7450

57

Banking



<http://www.guardian.co.uk/news/datablog/2009/mar/25/banking-g20#>

Fall 2011

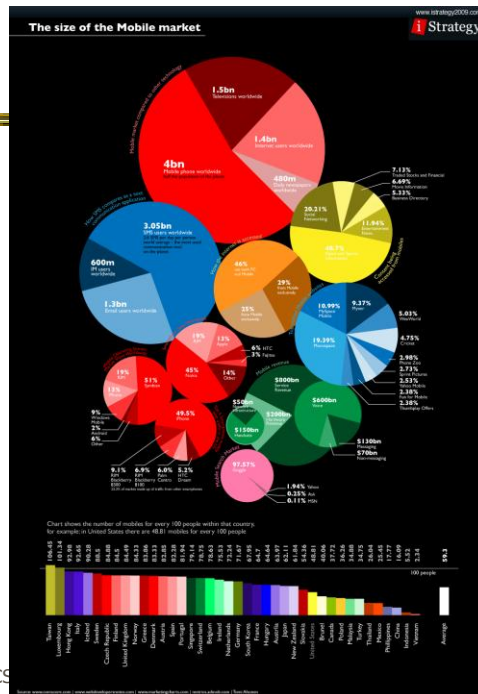
CS 7450

58

Mobile Market

<http://tecverse.com/design/cellphone-penetration-worldwide.html>

Fall 2011



<http://www.mikewirhart.com/?cat=3>

Beer

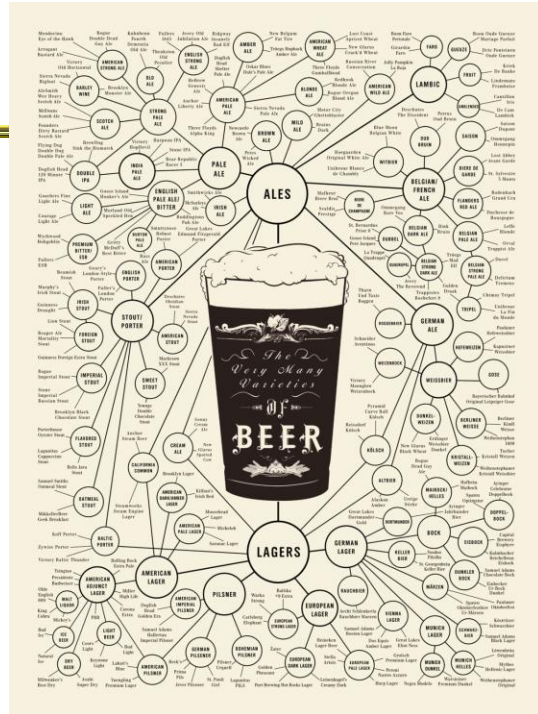


Fall 2011

CS 7450

60

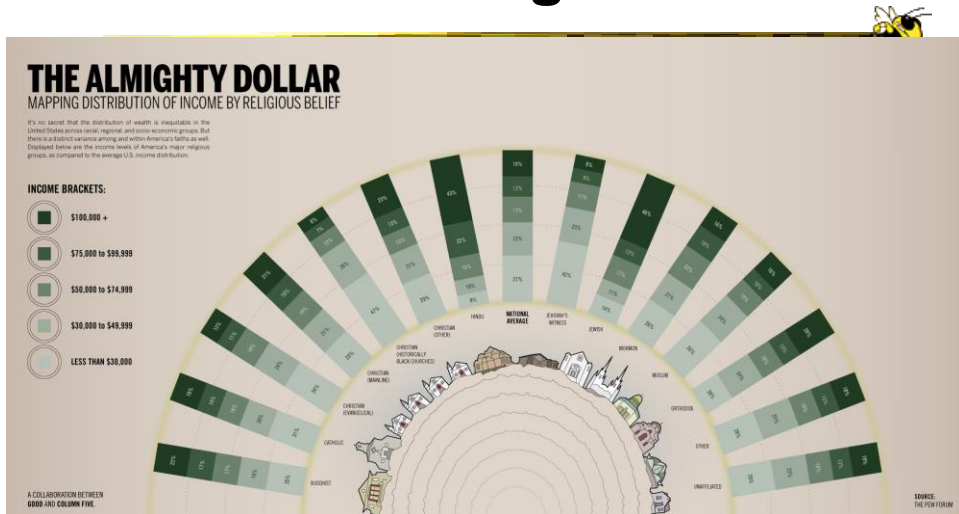
Beer!



http://images.fastcompany.com/upload/poster_beer_1300.jpg

Fall 2011

Income and Religion



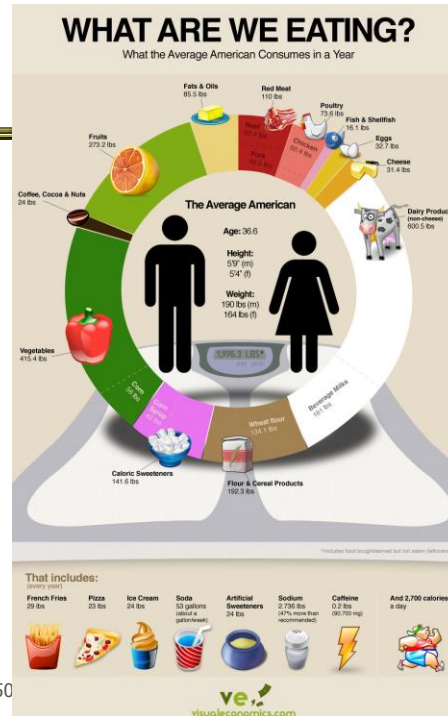
<http://awesome.good.is/transparency/web/1002/almighty-dollar/transparency.jpg>

Fall 2011

CS 7450

62

Diet

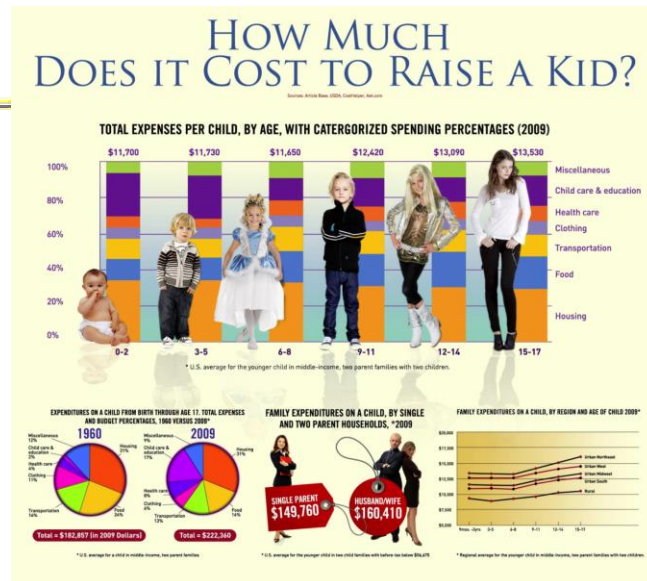


http://www.visualeconomics.com/food-consumption-in-america_2010-07-12/

Fall 2011

CS 7450

Kids



<http://www.visualeconomics.com/wp-content/uploads/2010/11/kidcosts-FINAL.jpg>

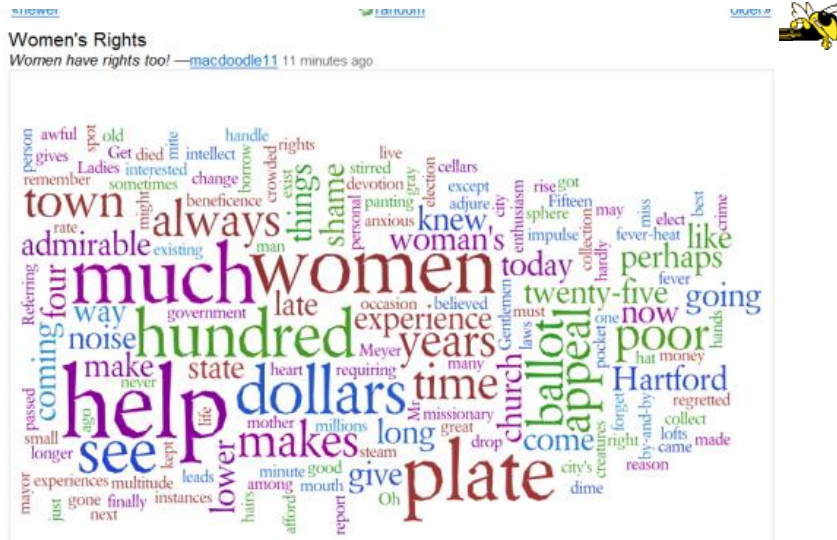
Fall 2011

CS 7450

64

Wordle

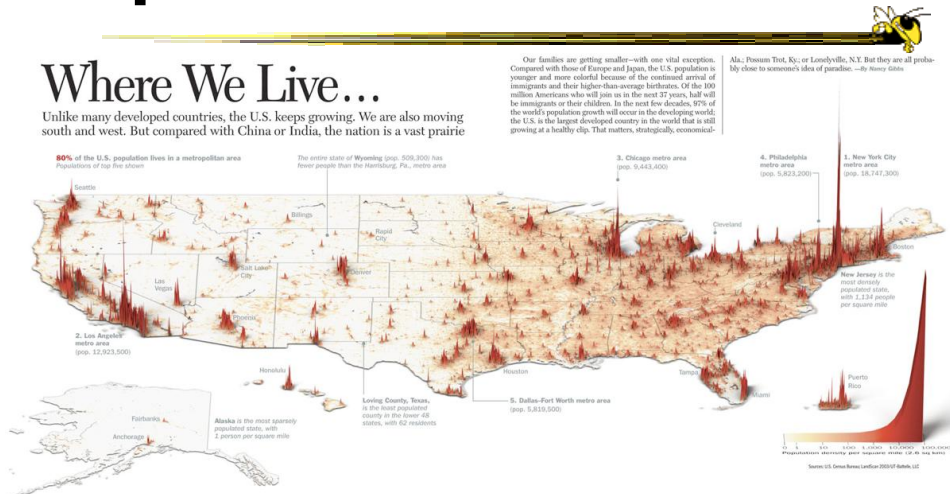
<http://www.wordle.net>



55 | 2011

CS 7450

Population



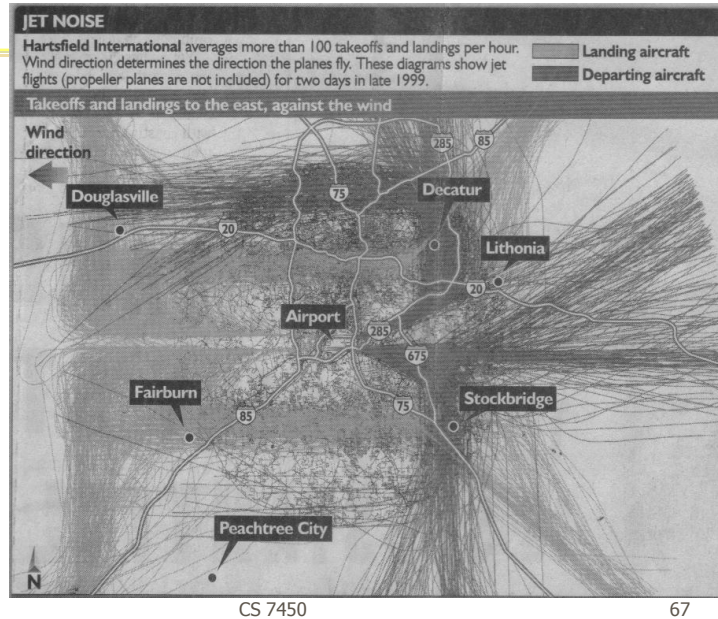
<http://infographicsnews.blogspot.com/2009/04/mantras-joe-lertolas-maps.html>

Fall 2011

CS 7450

66

Atlanta Flight Traffic



Atlanta Journal
April 30, 2000

Fall 2011

Country Music

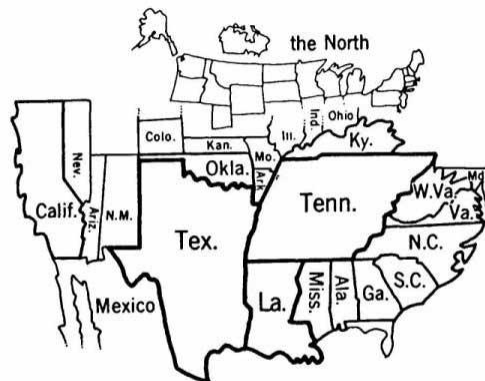


Figure 14. States Mentioned in Country-Music Lyrics
Source: Ben Marsh, "A Rose-Colored Map," *Harper's*, July 1977, 80. Used by permission.
Note: The size of each state is proportional to the number of times it is mentioned.

Fall 2011

CS 7450

68

London Subway

www.thetube.com



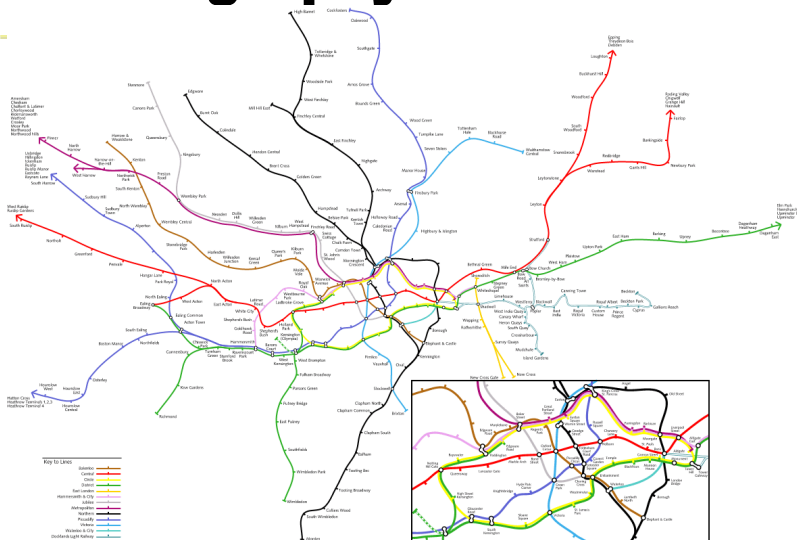
Fall 2011

CS 7450

69

www.kottke.org/plus/misc/images/tubegeo.gif

True Geography



Fall 2011

CS 7450

70

Easy Walking Lines Added



Fall 2011

CS 7450

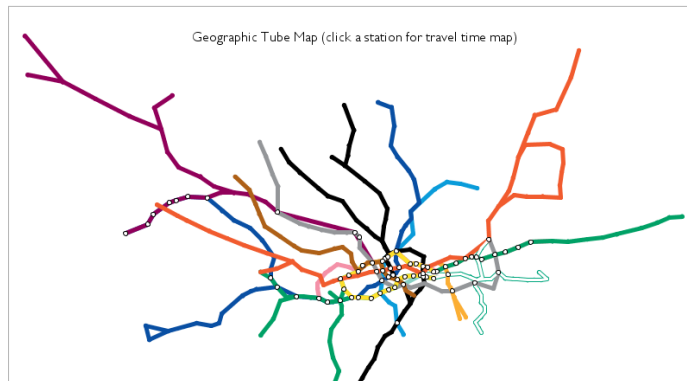
71

Travel Help

Travel Time Tube Map

Geographic Layout

Click on (or select, above) a station to see the London Underground map reorganise around the times of travel from that station. Shortest paths are used to place the other stations - radius is proportional to time to travel, and angle should be correct for as-the-crow-flies direction on a map. The concentric circles are at 10 minute intervals. Press 'g' to get back to the geographical tube map.



Fall 2011

CS 7450

72

Interaction

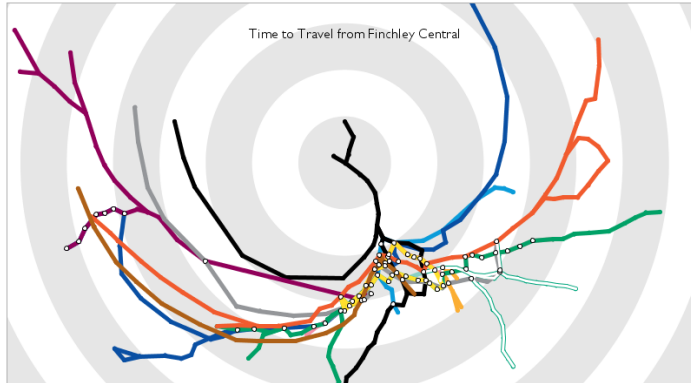
Shows travel times



Travel Time Tube Map

Gallions Reach

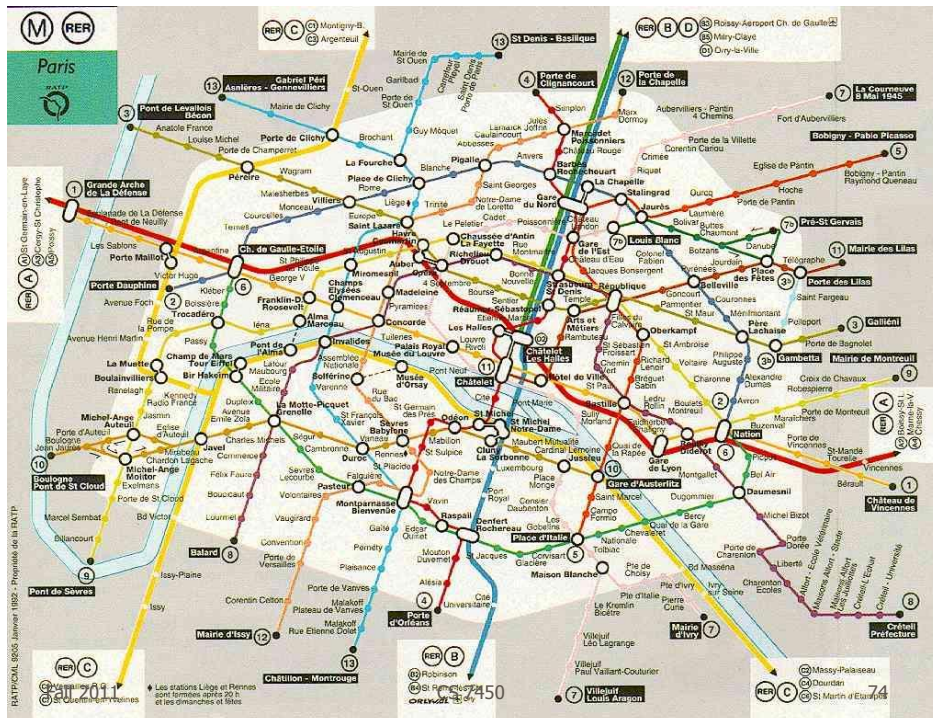
Click on (or select, above) a station to see the London Underground map reorganise around the times of travel from that station. Shortest paths are used to place the other stations - radius is proportional to time to travel, and angle should be correct for as-the-crow-flies direction on a map. The concentric circles are at 10 minute intervals. Press 'g' to get back to the geographical tube map.



Fall 2011

CS 7450

73



Atlanta MARTA

They have rivers,
we have highways
:^(

Fall 2011

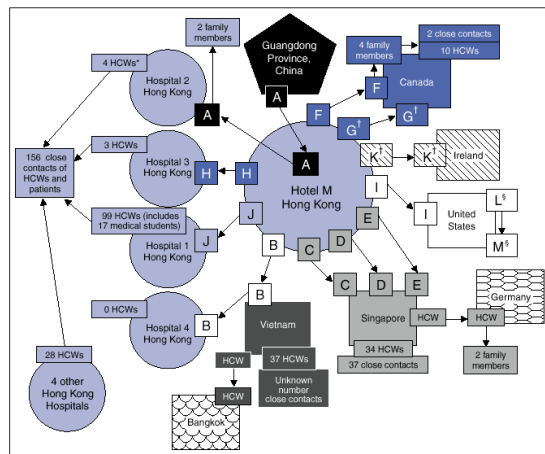


www.cdc.gov/mmwr/preview/mmwrhtml/mm5212a1.htm

SARS Outbreak



FIGURE 1. Chain of transmission among guests at Hotel M — Hong Kong, 2003



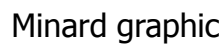
*Health-care workers.
†All guests except G and K stayed on the 9th floor of the hotel. Guest G stayed on the 14th floor, and Guest K stayed on the 11th floor.
‡Guests L and M (spouses) were not at Hotel M during the same time as index Guest A but were at the hotel during the same times as Guests G, H, and I, who were ill during this period.

Fall 2011

CS 7450

76

From E. Tufte
*The Visual Display of
Quantitative Information*



latitude
longitude

```
temperature
date
```

77

78



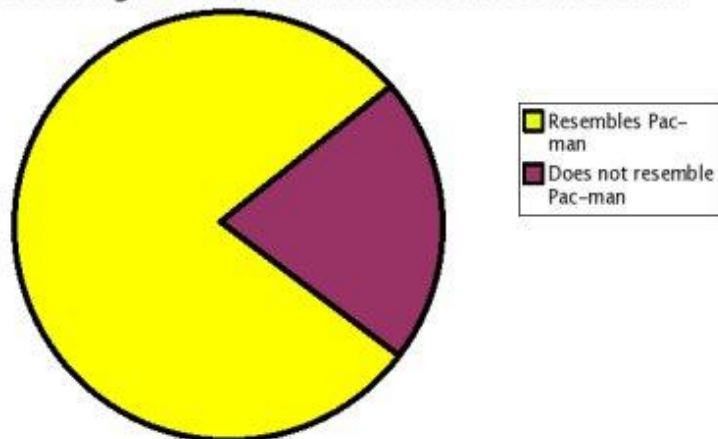
http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html

Fall 2011

CS 7450

79

Percentage of Chart Which Resembles Pac-man

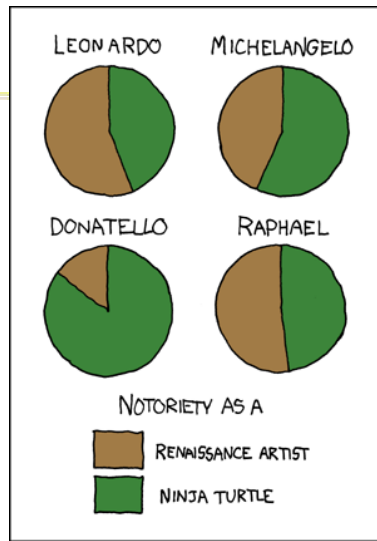


<http://www.boingboing.net/2006/11/02/hilarious-piechartvi.html>

Fall 2011

CS 7450

80

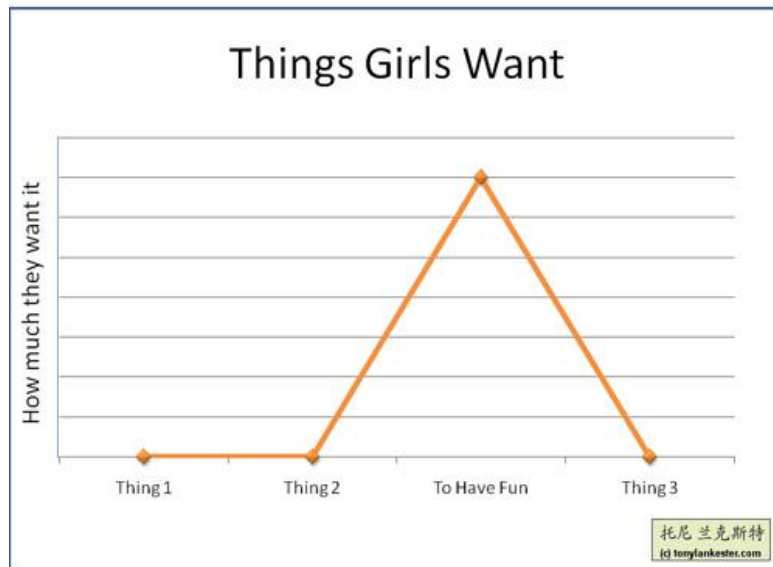


<http://xkcd.com/197/>

Fall 2011

CS 7450

81



<http://www.flickr.com/photos/91884218@N00/3108768440/in/pool-songchart>

Fall 2011

CS 7450

82



But Don't Do This

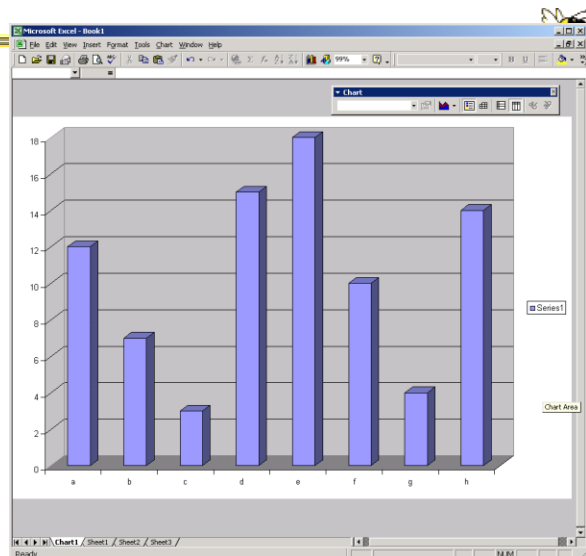
Fall 2011

CS 7450

83

Excel

Get rid of
those darn 3D
bars!



Fall 2011

CS 7450

84

USA Today Graphics



Or worse
yet...



Fall 2011

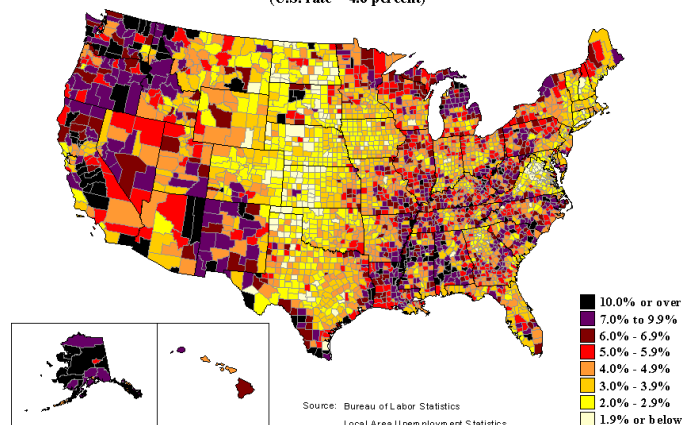
CS 7450

85

Unemployment Rates



Unemployment rates by county,
December 2000 - November 2001 averages
(U.S. rate = 4.6 percent)

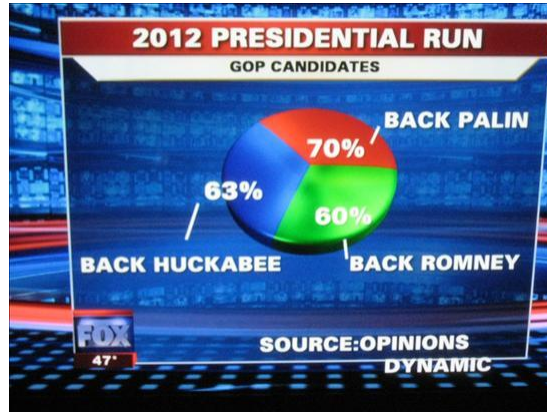


Fall 2011

CS 7450

86

FOX “News”



<http://wonkette.com/412361/all-193-of-republicans-support-palin-romney-and-huckabee>

Fall 2011

CS 7450

87

Examples



- Tools/Systems
 - Now interaction becomes important...

Fall 2011

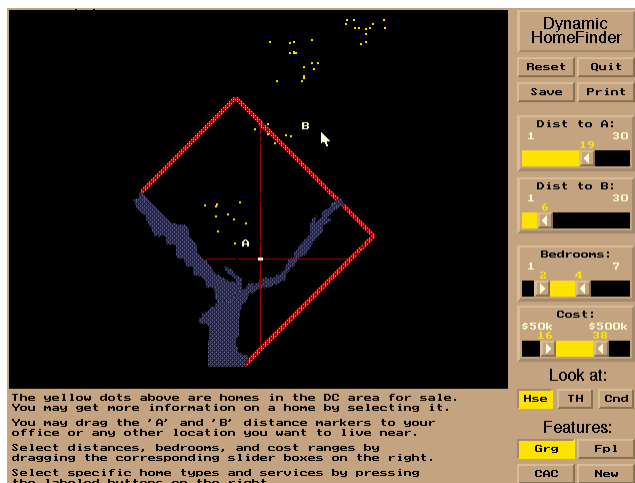
CS 7450

88

HomeFinder



HCIL
Univ. Maryland



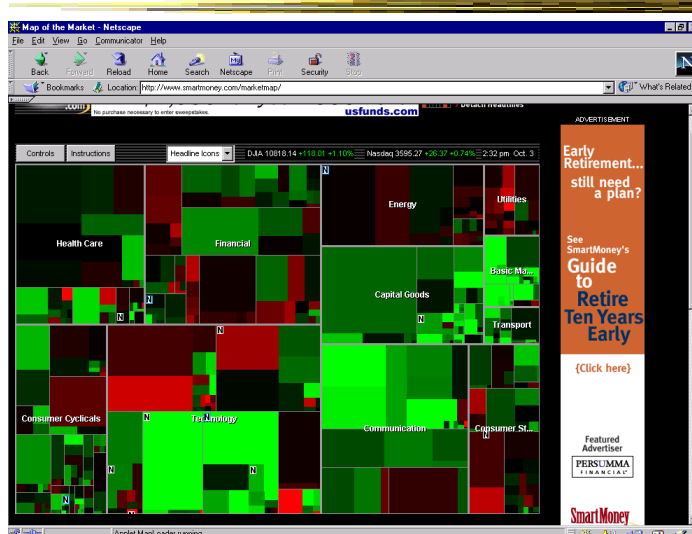
Fall 2011

CS 7450

89

www.smartmoney.com/marketmap

Map of the Market



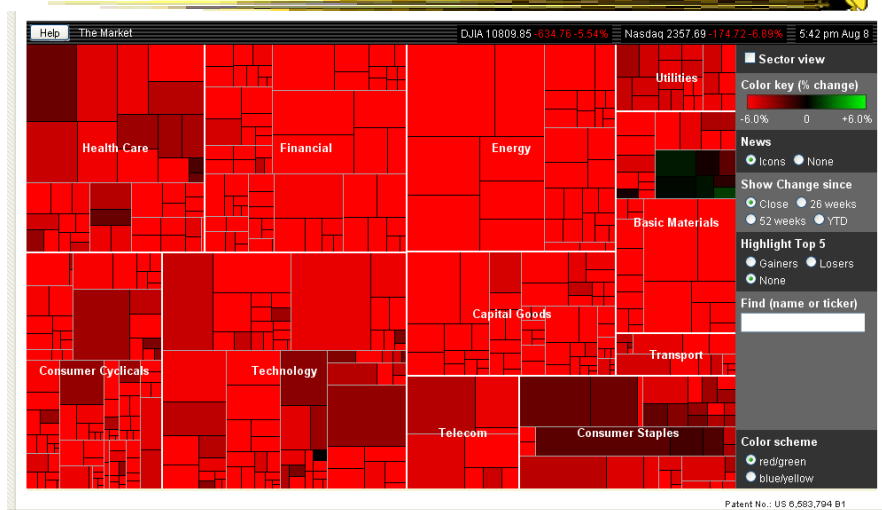
Demo

Fall 2011

CS 7450

90

Some Days It Looks Like This...



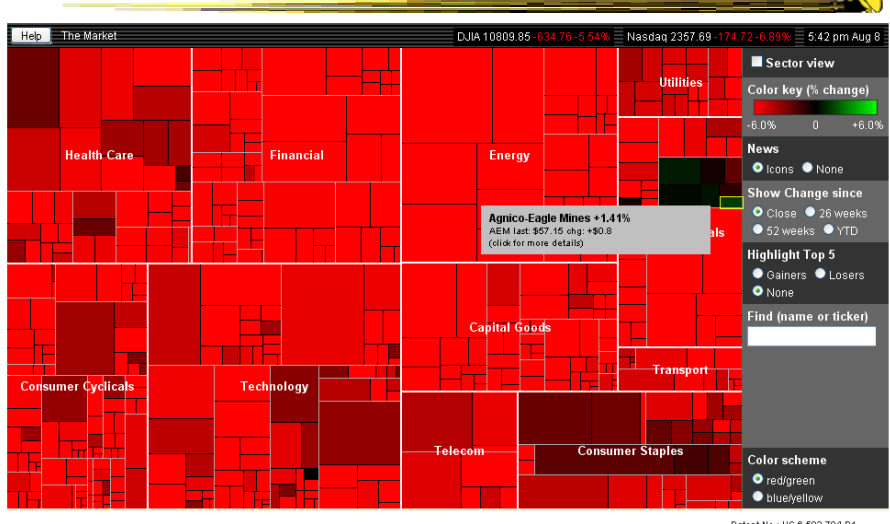
Fall 2011

CS 7450

Aug. 8, 2011

91

Some Days It Looks Like This...



Fall 2011

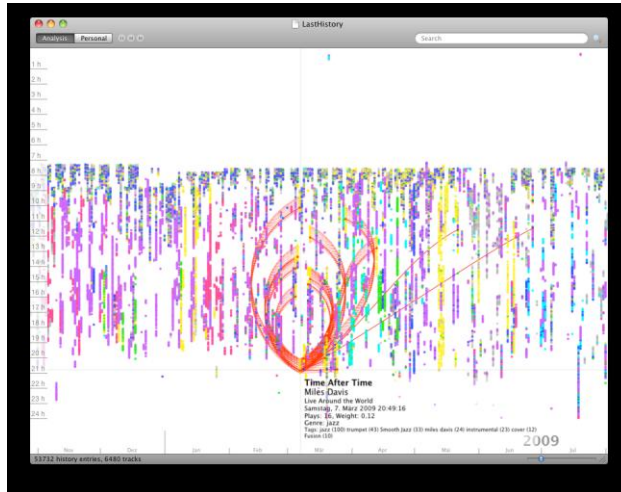
CS 7450

Aug. 8, 2011

92

<http://www.frederikseiffert.de/lasthistory/>

Your Music Listening History



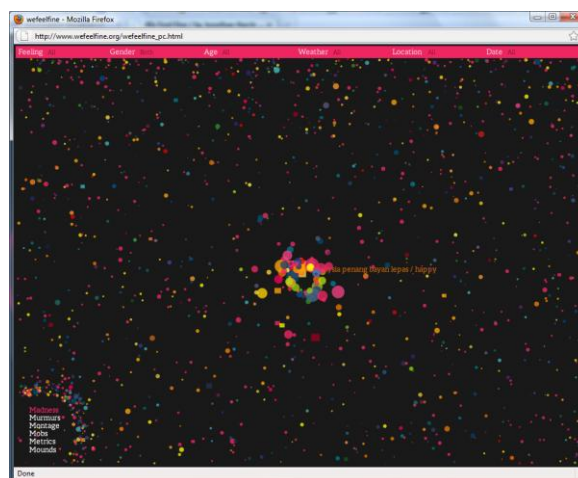
Video

Fall 2011

CS 7450

93

IM/Tweets



<http://www.wefeelfine.org>

Fall 2011

CS 7450

94

Philip Glass Music

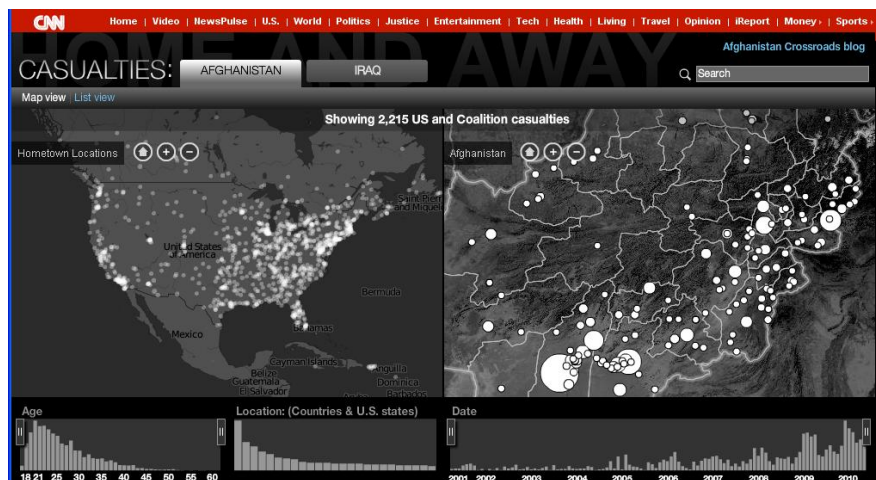


Fall 2011

CS 7450

95

Casualties of War



56 | 2011

CS 7450

NY Times



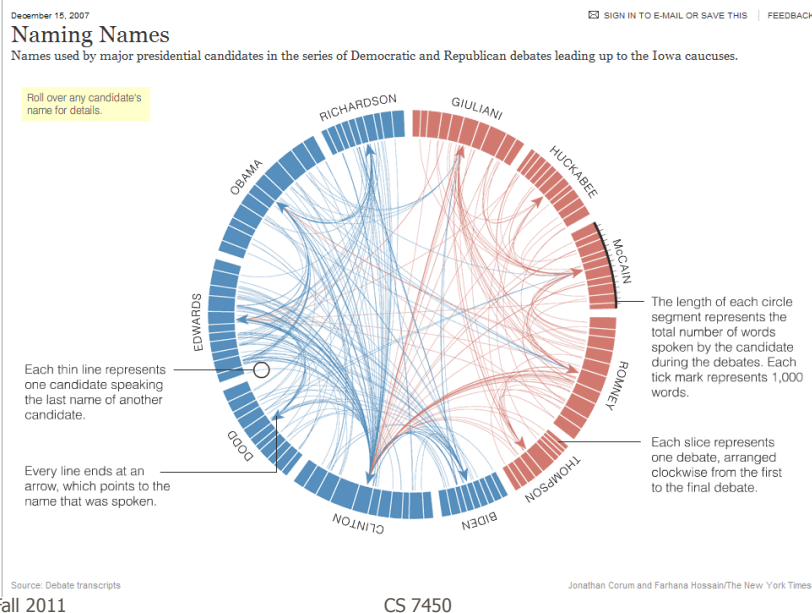
- Has been a wonderful source of interactive data visualizations
- Some examples...

Fall 2011

CS 7450

97

<http://www.nytimes.com/interactive/2007/12/15/us/politics/DEBATE.html#>

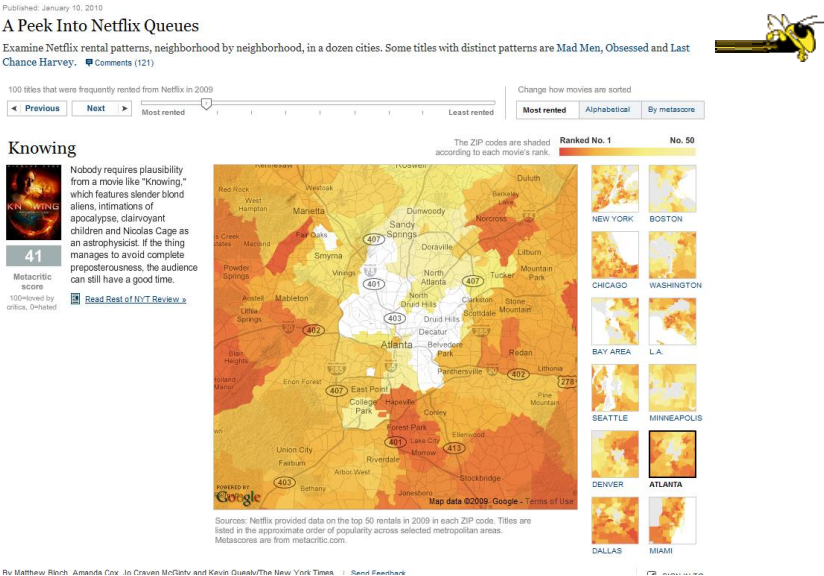


Fall 2011

CS 7450

98

http://www.nytimes.com/interactive/2010/01/10/nyregion/20100110-netflix-map.html?hp

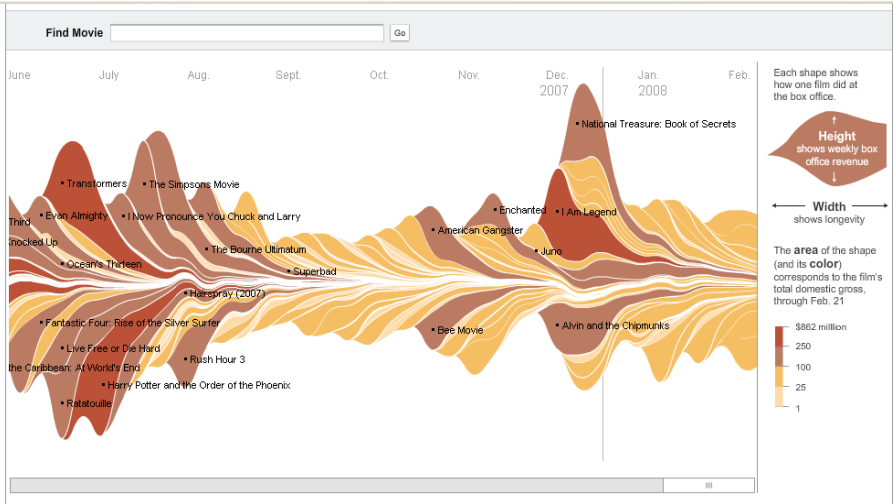


Fall 2011

CS 7450

99

http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html



Fall 2011

CS 7450

100

http://www.nytimes.com/ref/us/20061228_3000FACES_TAB2.html

Casualties of War

FACES ANALYSIS THEIR STORIES

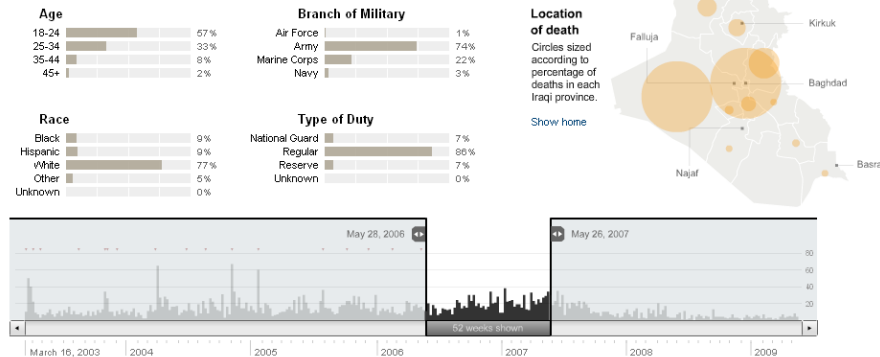
E-MAIL FEEDBACK

Use the slider below to investigate the demographics and military status of U.S. service members who died during the war in Iraq.

MAY 28, 2006 — MAY 26, 2007 (52 WEEKS)

Show all Initial invasion First invasion of Falluja Second invasion of Falluja Since troop buildup began

992 deaths



Fall 2011

CS 7450

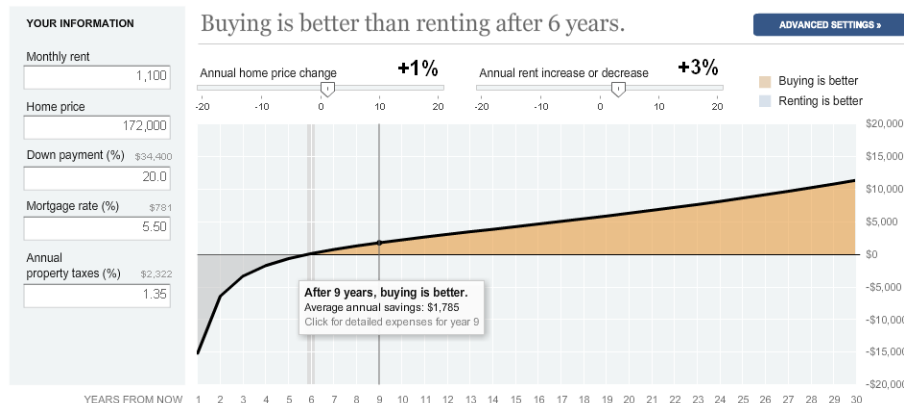
101

http://www.nytimes.com/interactive/business/buy-rent-calculator.html

UPDATED April 21, 2010

Is It Better to Buy or Rent?

Whether renting is better than buying depends on many factors, particularly how fast prices and rents rise and how long you stay in your home. Compare the costs of buying and renting a home in the calculator below. Click the **ADVANCED SETTINGS** button to change inputs such as your rate of return on investments, condo/common fees and your tax bracket.

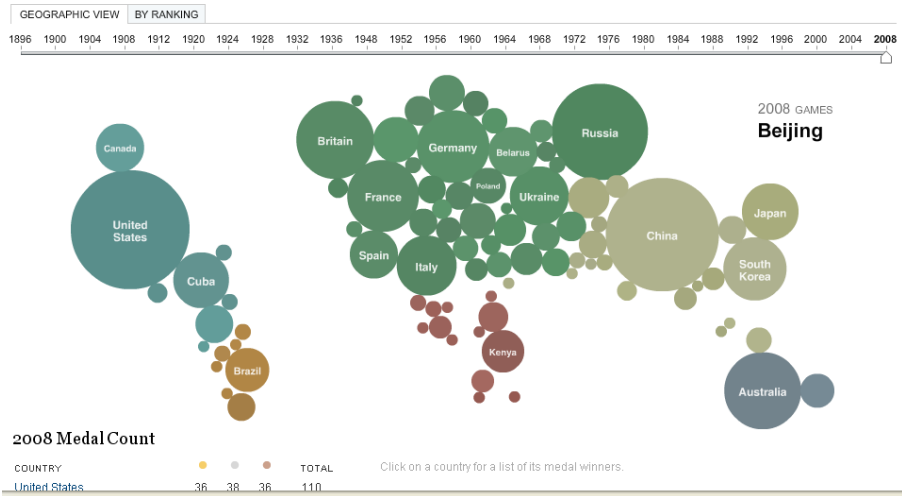


Fall 2011

CS 7450

102

http://www.nytimes.com/interactive/2008/08/04/sports/olympics/20080804_MEDALCOUNT_MAP.html



Fall 2011

CS 7450

103

Good Resources

- Some places to look for more information



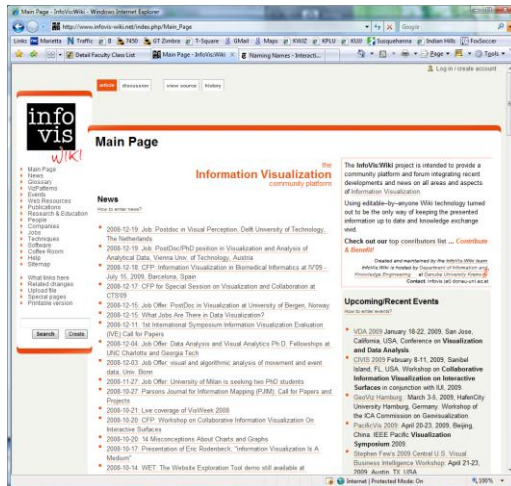
Fall 2011

CS 7450

104

<http://www.infovis-wiki.net>

InfoVis Wiki



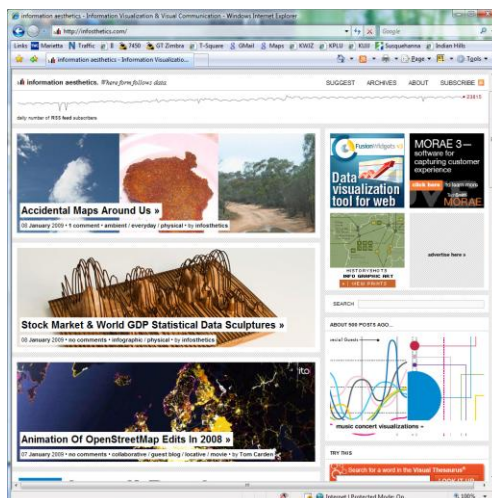
Fall 2011

CS 7450

105

<http://infosthetics.com/>

Infosthetics Blog



Fall 2011

CS 7450

106

S. Few Book



- Chapters 1-3
- Discuss

Fall 2011

CS 7450

107

Administrative Matters



- Review
- Questions?

Fall 2011

CS 7450

108

HW



- HW1 due next Tuesday
 - Data Exploration and Analysis

Fall 2011

CS 7450

109

Upcoming



- Data & Graph/Chart Design
 - Reading:
 - S. Few – web article
- Visual Perception
 - Reading:
 - M. Stone – web article

Fall 2011

CS 7450

110