Overview of InfoVis



CS 7450 - Information Visualization Jan. 13, 2011 John Stasko

Instructors



Prof: John StaskoOffice: TSRB 342Office Hours: ???

- Email: stasko@cc.gatech.edu

• TA: Hina Shah

Spring 2011 CS 7450 2

1

Administratia



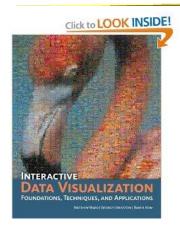
- Get it all from class website
 - Syllabus & Biblio
 - Assignments
 - Instructor & TA
 - Related Courses
 - InfoVis Resources

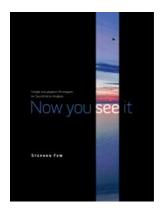


http://www.cc.gatech.edu/~stasko/7450

Spring 2011 CS 7450 3

Books



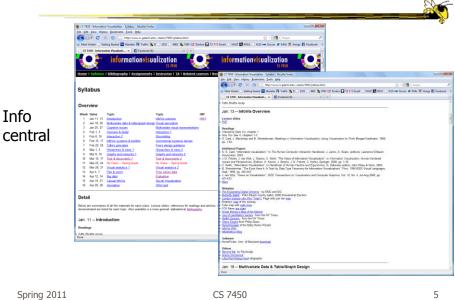


Supplemented by papers

Spring 2011 CS 7450

2

Syllabus



Grading



- Participation
- Assignments
 - HWs (about 7 or 8)
- Project
- Final Exam???
- (Details still being finalized)

Spring 2011 CS 7450

CAVEAT



- This course is a lot of work. If you're just looking for some easy grade, I would advise you to drop now.
 - Lots of reading too
- If you are sincerely interested in this topic, I hope you will enjoy the course and learn a lot

Spring 2011 CS 7450

Exercise



Get out pencil and paper



Spring 2011 CS 7450

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...

How Much Data? (1)



- Estimated info added to digital universe each year will soon approach 1 ZB (zettabyte)*
 - 1000000000000000000000 (10²¹⁾ bytes
 - From: http://www.emc.com/digital_universe viewed December 8, 2008

*But only half that goes to my email inbox

Spring 2011 CS 7450 11

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

Spring 2011 CS 7450 Slide courtesy Jim Thomas 12

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 decisionmaking processes?
 - How do we avoid being overwhelmed?



The Challenge



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



Spring 2011 CS 7450 14

7

The Problem



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

Two slides courtesy of Chris North



Data Transfer

How?

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

Taste

CS 7450

Spring 2011

Human Vision



15

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

Impressive. Lets use it!

CS 7450 Spring 2011 16

Some Examples



Why visualization helps...

Spring 2011 CS 7450 17

Which cereal has the most/least potassium?

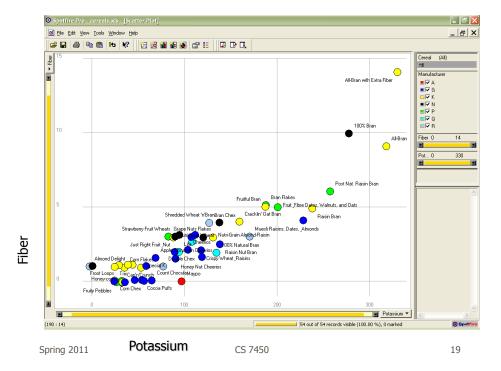
Is there a relationship between potassium and fiber?

If so, are there any outliers?

Questions: If so, are there any outliers?

Which manufacturer makes the healthiest cereals?





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?



Another Illustrative Example

Spring 2011 CS 7450 21

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

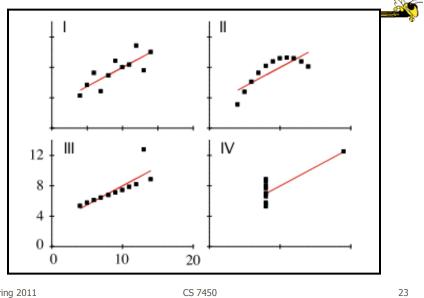
Spring 2011

Coefficient of determination = 0.67

http://astro.swarthmore.edu/astro121/anscombe.html
CS 7450 22

11

The Data Sets



CS 7450 Spring 2011

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

Spring 2011 CS 7450 24

Exercise Redux



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

Spring 2011 CS 7450

Visualization



- Definition
 - "The use of computer-supported, interactive visual representations of data to amplify cognition."

From [Card, Mackinlay Shneiderman '98]

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

Spring 2011 CS 7450 27

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

Part of our Culture



- "I see what you're saying"
- "Seeing is believing"
- "A picture is worth a thousand words"

Spring 2011 CS 7450

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

Spring 2011

CS 7450

When to Apply?



- Many other techniques for data analysis
 - Statistics, DB queries, 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

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?

Spring 2011 CS 7450 33

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?

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?

Spring 2011 CS 7450 35

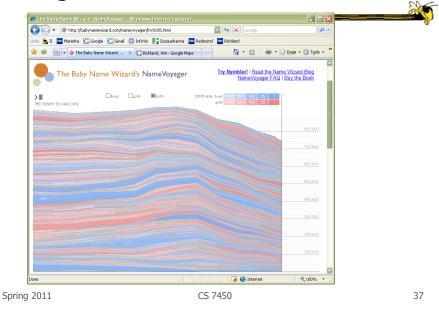
Tasks in Info Vis



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

More to come in a future class...

Baby Name Wizard



2. Presentation



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

\$8ring 2011 CS 7450

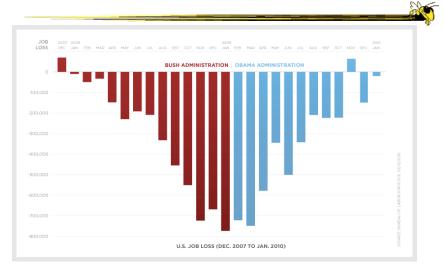
When to Apply?



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

Spring 2011 CS 7450 3

Job Losses



Controversial, see http://soquelbythecreek.blogspot.com/2010/02/what-does-obama-job-chart-really-mean.html
Spring 2011 CS 7450 40

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

Spring 2011 CS 7450 41

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

\$pring 2011 CS 7450

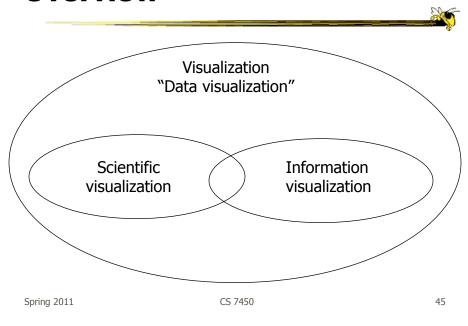


Academic Areas



 Where does InfoVis fit in the academic world?

Overview



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

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...

Spring 2011 CS 7450

Information Visualization



- 2. What is "visualization"?
 - The use of computer-supported, interactive visual representations of data to amplify cognition.

From [Card, Mackinlay Shneiderman '98]

Information Visualization



Components:

- 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/

Spring 2011 CS 7450 49

Two Key Attributes



- Scale
 - Challenge often arises when data sets become large
- Interactivity
 - Want to show multiple different perspectives on the data

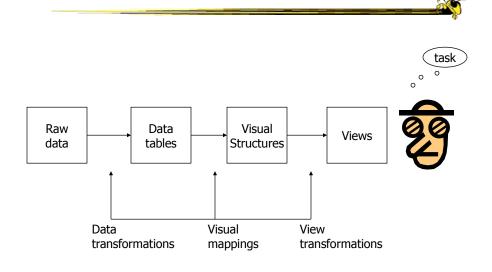
Example Domains for Info Vis



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

Spring 2011 CS 7450 51

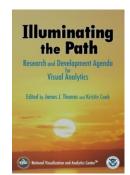
InfoVis Process Model



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

5₿ring 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

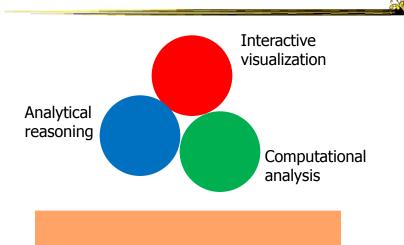


Keim et al chapter in *Information Visualization: Human-Centered Issues and Perspectives*, 2008

5pring 2011

CS 7450

Main Components



Production & presentation

Spring 2011 CS 74

Examples

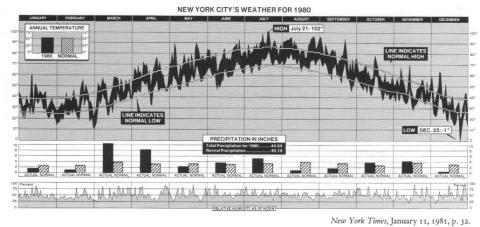


- Start with static pictures (InfoGraphics)
 - Very popular on the web
 - But are they information visualizations?

NYC Weather



2220 numbers

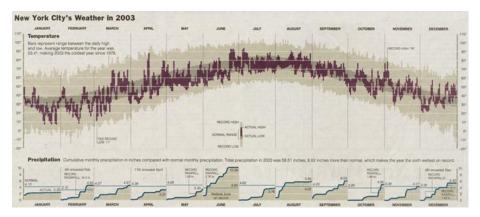


Tufte, Vol. 1

Spring 2011 CS 7450

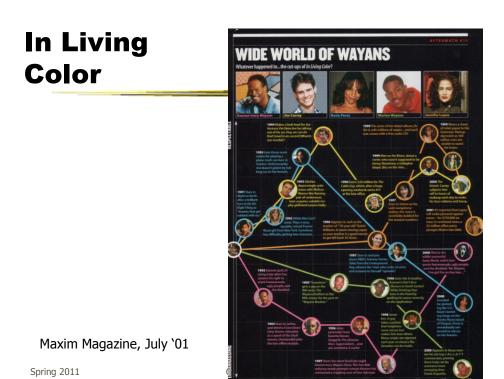
Updated Version





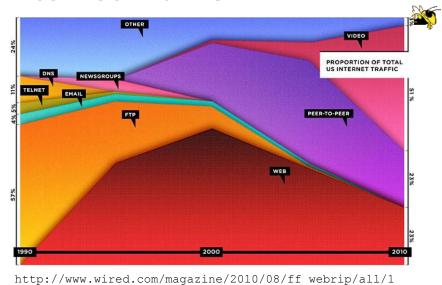
http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00014g

Spring 2011 CS 7450 58



Internet Traffic

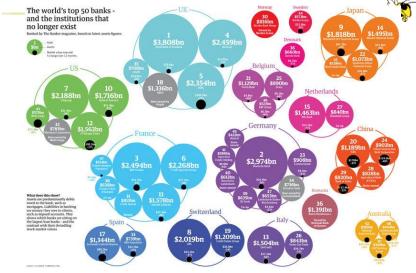
Spring 2011



CS 7450

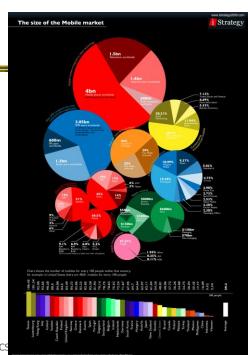
60

Banking



 $\label{local-could} $$ $$ $$ http://www.guardian.co.uk/news/datablog/2009/mar/25/banking-g20\# Spring 2011 $$ CS 7450 $$ 61$

Mobile Market



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

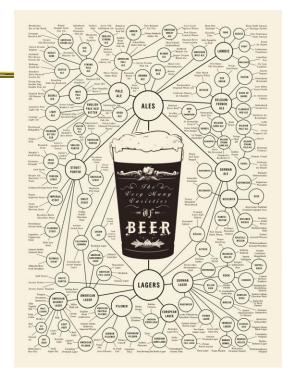
Spring 2011

Beer



Spring 2011 CS 7450 63

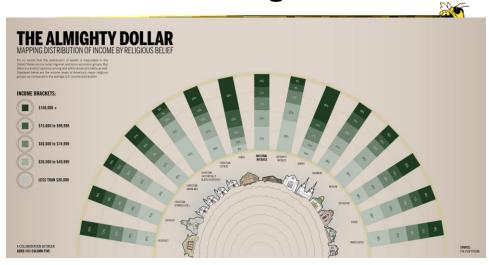
Beer!



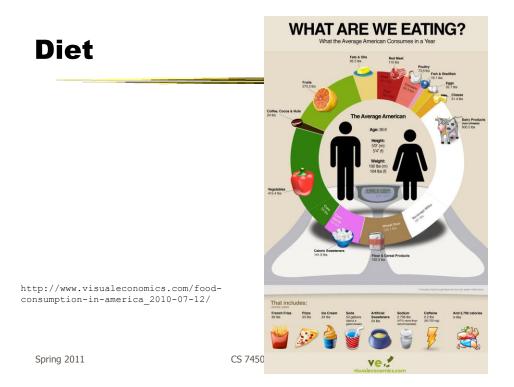
http://images.fastcompany.com/upload/poster_beer_1300.jpg

Spring 2011

Income and Religion



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



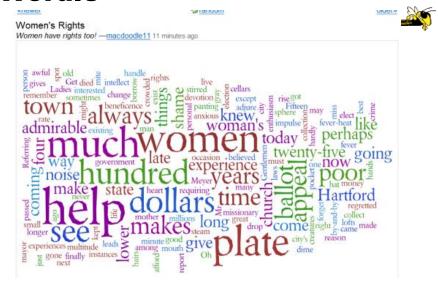
HOW MUCH DOES IT COST TO RAISE A KID? -**Kids** TOTAL EXPENSES PER CHILD, BY AGE, WITH CATERGORIZED SPENDING PERCENTAGES (2009)

http://www.visualeconomics.com/wp-content/uploads/2010/11/kidcosts-FINAL.jpg CS 7450

Wordle

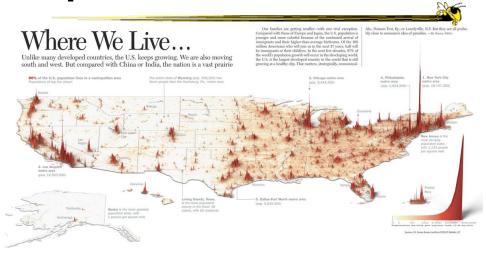
Spring 2011

http://www.wordle.net



6βring 2011 CS 7450

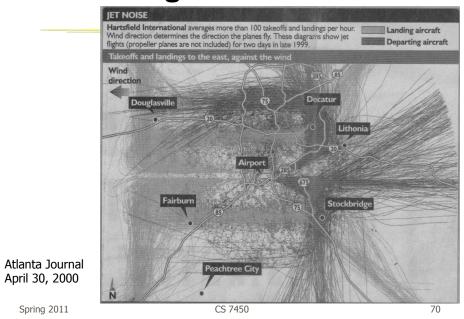
Population



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

Spring 2011 CS 7450 69

Atlanta Flight Traffic



35

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.

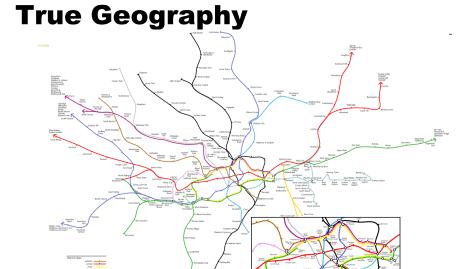
Spring 2011 CS 7450 71

London Subway

www.thetube.com



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



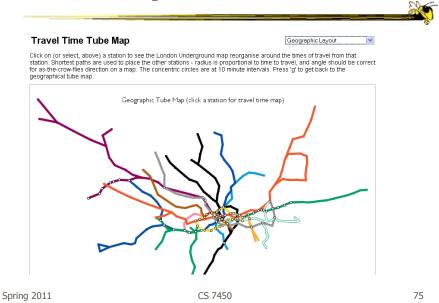
Spring 2011 CS 7450 73

rodcorp.typepad.com/photos/art_2003/tube_walklines_final_lmfaint.html

Easy Walking Lines Added



Travel Help

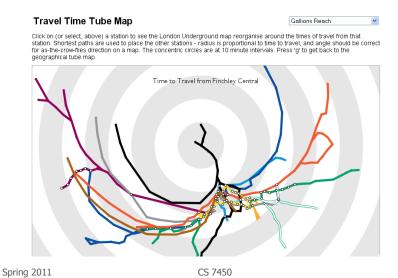


Interaction

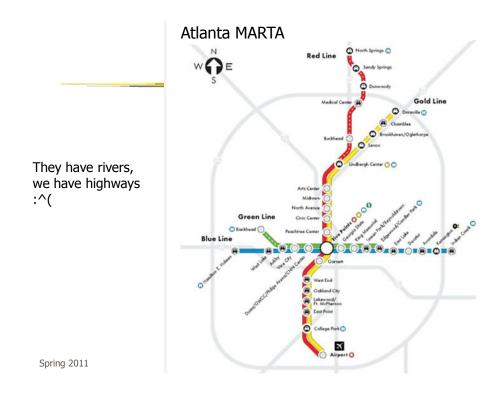
Shows travel times



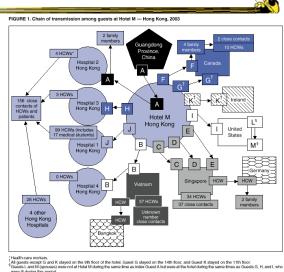
76







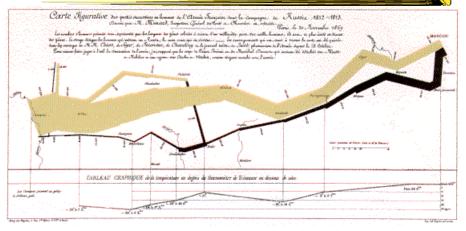
SARS Outbreak



Spring 2011 CS 7450

Napolean's March

From E. Tufte The Visual Display of Quantitative Information



Minard graphic

size of army direction

latitude longitude temperature date

Spring 2011

CS 7450

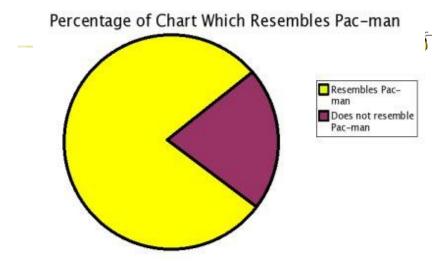


Or, for fun...

Spring 2011 CS 7450 81

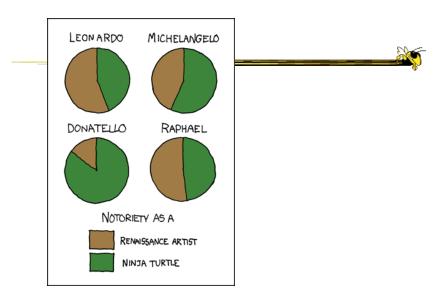


http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html

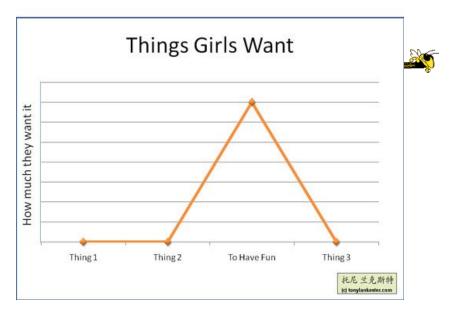


http://www.boingboing.net/2006/11/02/hilarious-piechartvi.html

Spring 2011 CS 7450 83



http://xkcd.com/197/

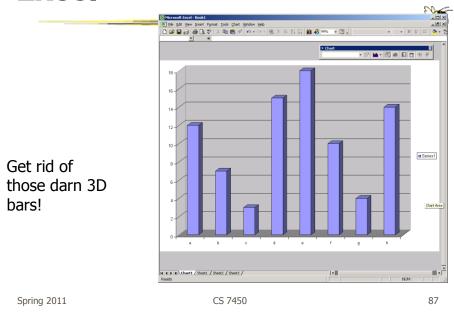


 $\label{linear_http://www.flickr.com/photos/91884218@N00/3108768440/in/pool-song chart} $$\operatorname{Spring} 2011 $$\operatorname{CS} 7450 $$85$



But Don't Do This

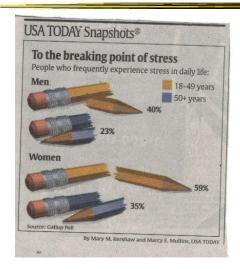
Excel



USA Today Graphics

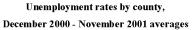


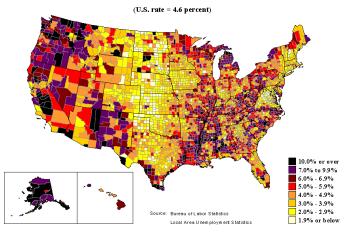
Or worse yet...



Unemployment Rates



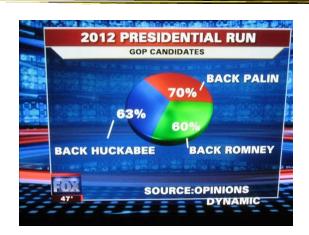




Spring 2011 CS /450 89

FOX "News"





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

Examples



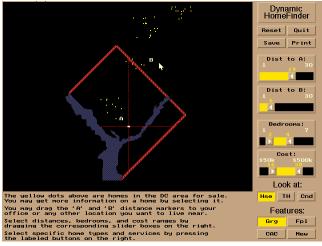
- Tools/Systems
 - Now interaction becomes important...

Spring 2011 CS 7450 9

HomeFinder



HCIL Univ. Maryland



Spring 2011 CS 7450 92

Map of the Market



Some Days It Looks Like This...



CS 7450

Spring 2011

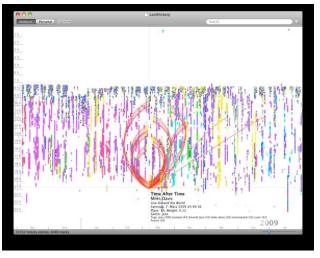
Sept. 29, 2008

US Open Tennis



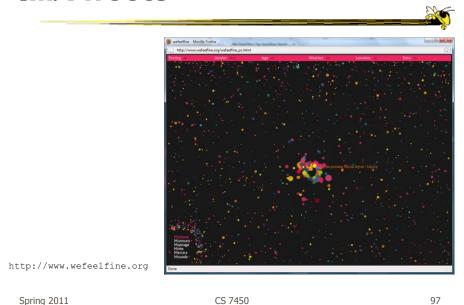
http://www.frederikseiffert.de/lasthistory/

Your Music Listening History



Video

IM/Tweets

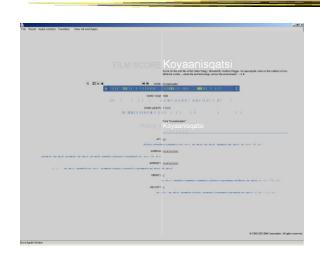


Spring 2011 CS 7450

www.philipglass.com/glassengine

Philip Glass Music





Spring 2011 CS 7450 98

Casualties of War



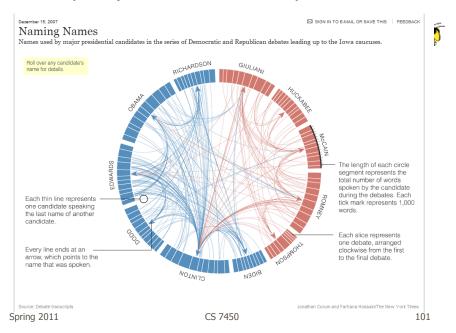
99ring 2011 CS 7450

NY Times



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

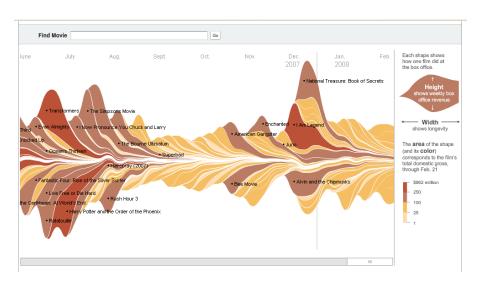
http://www.nytimes.com/interactive/2007/12/15/us/politics/DEBATE.html#



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

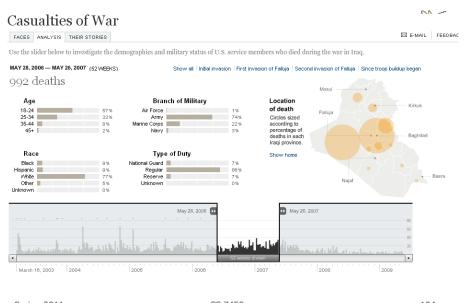


http://www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html



Spring 2011 CS 7450 103

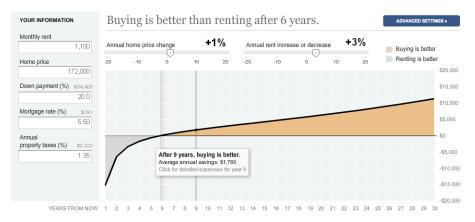
http://www.nytimes.com/ref/us/20061228_3000FACES_TAB2.html



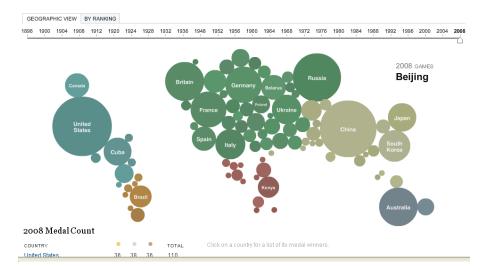
UPDATED April 21, 2010

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.



Spring 2011 CS 7450 105



Good Resources



Some places to look for more information

Spring 2011 CS 7450 107

InfoVis Wiki

http://www.infovis-wiki.net





http://infosthetics.com/

Infosthetics Blog





Spring 2011 CS 7450 109



InfoVis meet MTV



Spring 2011 CS 7450 111

HW



- HW1 due next Thursday
 - Data Exploration and Analysis

Upcoming



- Data & Graph/Chart Design
 - Reading:

Ward – chapters 2 and 4

S. Few – web article

- Visual Perception
 - Reading:

Ward – chapter 3

M. Stone – web article