Overview of InfoVis

CS 7450 - Information Visualization
Aug. 24, 2016
John Stasko

Learning Objectives

- Articulate definition and purpose of visualization
- Describe two main uses or applications of visualization
- List two primary components of visualizations
- Describe the different areas of academic visualization research
- Explain the infovis "pipeline" (process)

Exercise

- Get out pencil and paper

Electronics

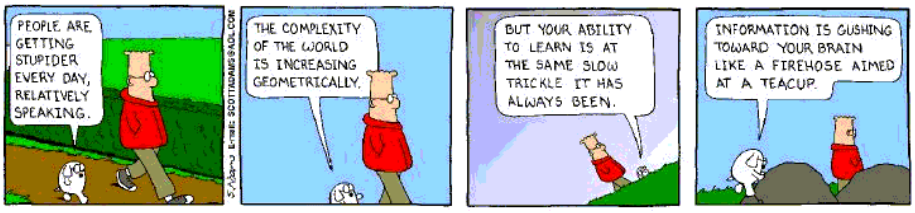
- This is now a NO laptops/cellphones class
- Exceptions will be noted (exercises, etc)
- Note-takers, see me

Reminder

Data

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?

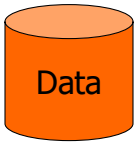


The Challenge

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

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
Haptic/tactile
Smell
Taste
Telepathy?

Two slides courtesy of Chris North

Human Vision

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

Impressive. Lets use it!

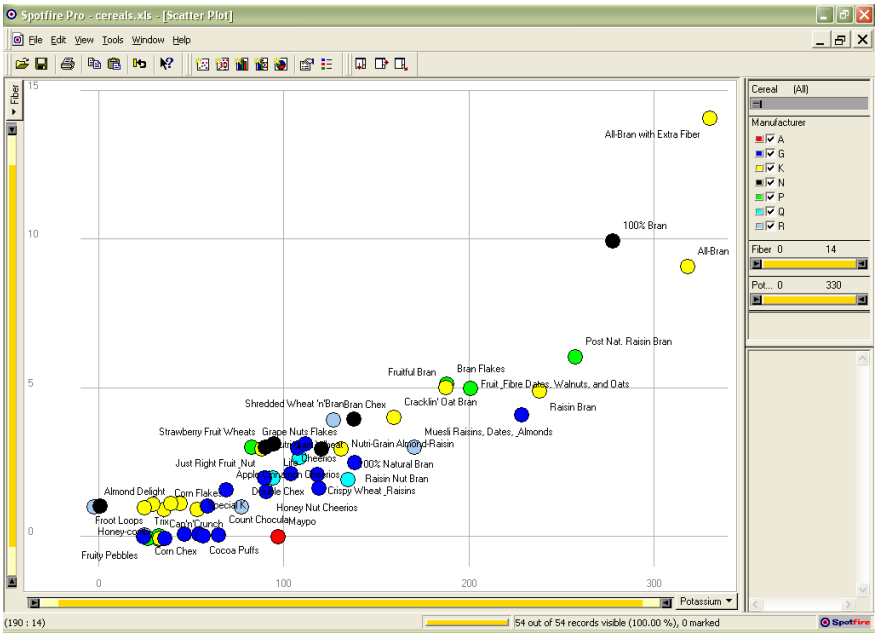
An Example

- Why visualization helps…

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



Fall 2016 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 2016 CS 7450 14

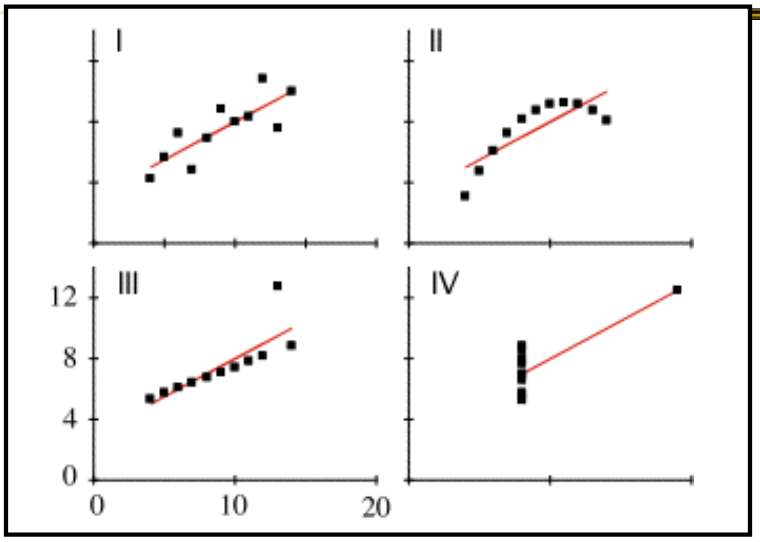
Another Illustrative Example

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

The Data Sets



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

More on this Topic

- "Value of visualization" lecture later in term

Exercise Redux

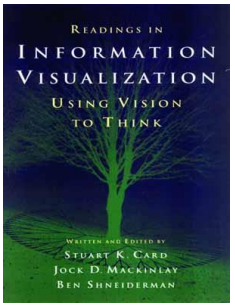
- Let's check what you did…

- People work differently

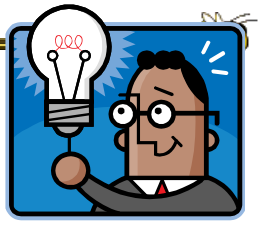
Visualization

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

From



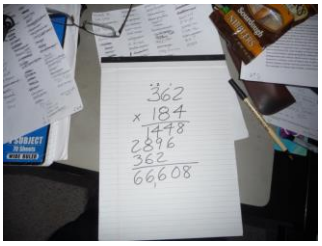
Visualization



- 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

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

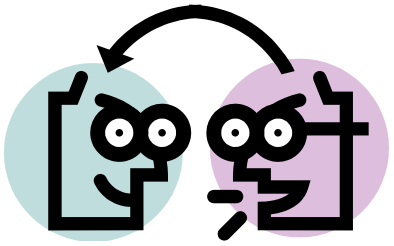
Expressed Well

"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

Part of our Culture

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



Admin Intermission

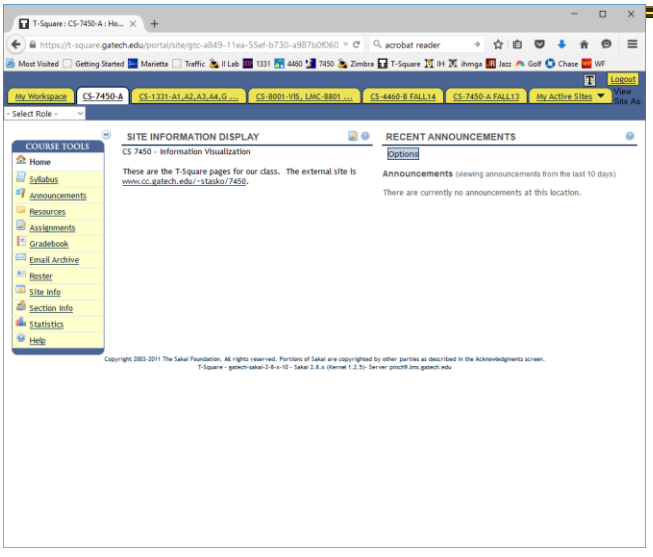
- Overloads
- Surveys
- More…

Administratia

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

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

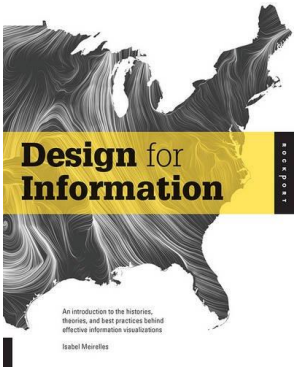
T-Square Site



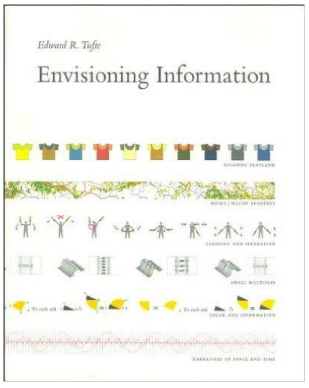
Get lecture slides from last time there

Books

Required

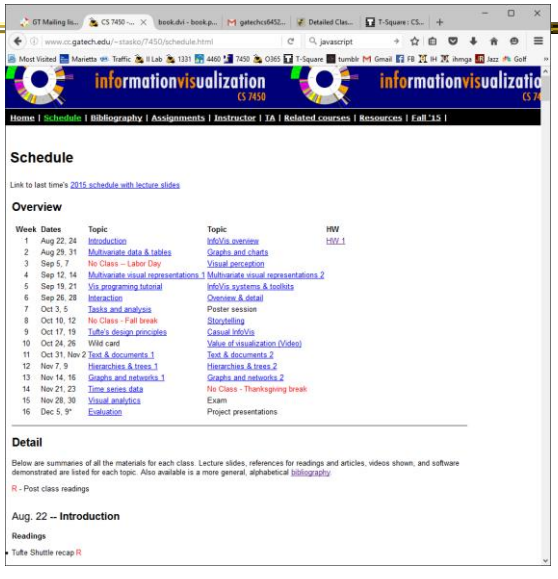


Recommended



Schedule

Info central



Flipped Readings

- No reading assignment previewing next class' material
- Instead, reading of papers and viewing of videos/websites from current class for next time
  - Potential pop quiz at start of next class

Look for R on Schedule webpage

Grading

- Participation
- Pop Quizzes
- Assignments
  - HWs (about 7)
- Project
- Final Exam

- (Details still being finalized)

Survey

- Who wasn't here on Monday and didn't fill out a survey?
- If you want to change your
  "I'm on WL and not sure I want in"
  to
  "I'm on WL and I definitely want in"
  see me after class

CAVEAT

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

Waitlist/Overload Update

- Status
  - Room capacity
  - Let in 15 yesterday
  - 29 of original 60 said "might drop"
  - Will let in more tomorrow
  - Please drop by Thursday noon
  - On Friday it becomes the "Wild West"

Electronics

- This is now a NO laptops/cellphones class
- Exceptions will be noted (exercises, etc)
- Note-takers, see me

Reminder

Back to content

Purpose

- Two main uses of infovis
  - Analysis – Understand your data better and act upon that understanding
  - Communication – 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

EDA example?

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?

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?

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

2. Communication

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

When to Apply?

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

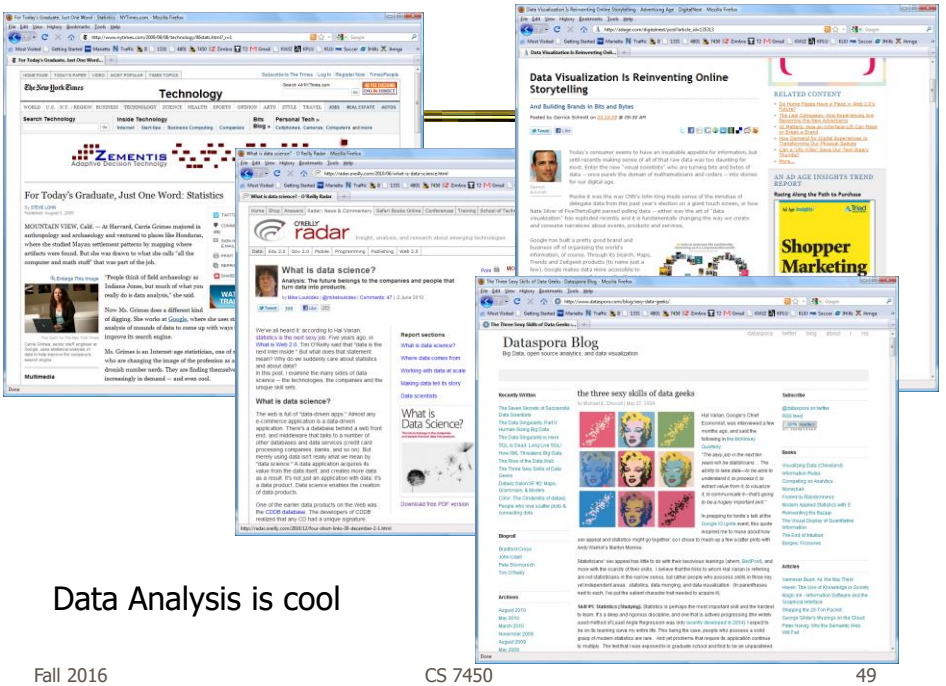
Key Benefits of Visualization

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

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

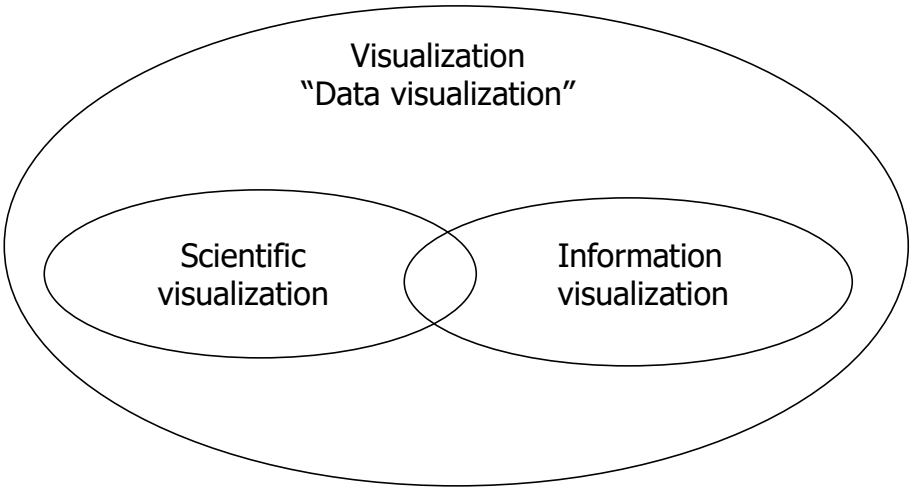


Data Analysis is cool

Academic Areas

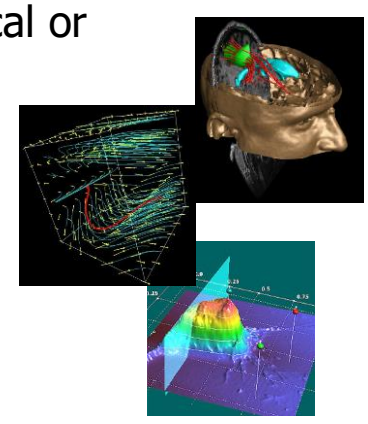
- Where does InfoVis fit in the academic world?

Overview



Scientific Visualization

- Primarily relates to and represents something spatial, 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"?
  - Non-spatial data: 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...

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

- Characteristics:
  - Taking things without a direct physical correspondence (non-spatial) 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/

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*

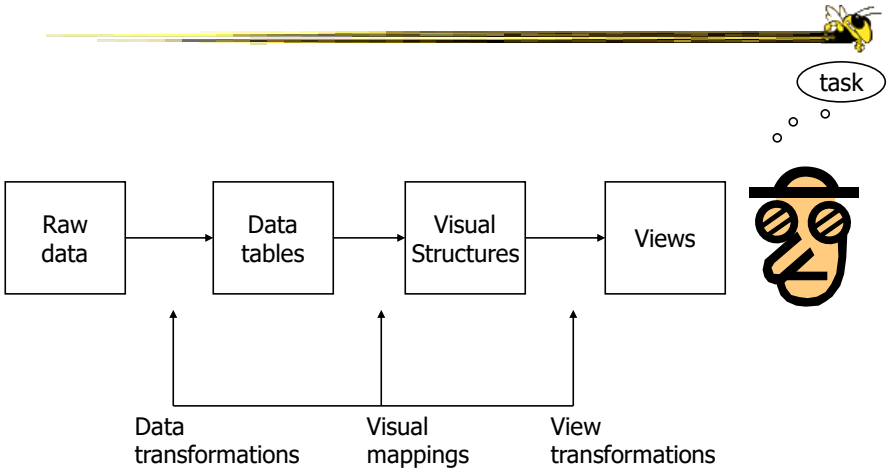
Two Key Challenges

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

Example Domains for Info Vis

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

InfoVis Process Model

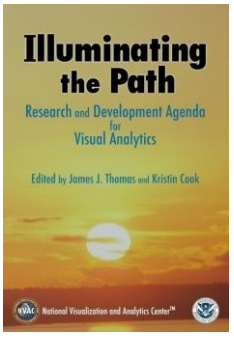


From: Card, Mackinlay, Shneiderman '99

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



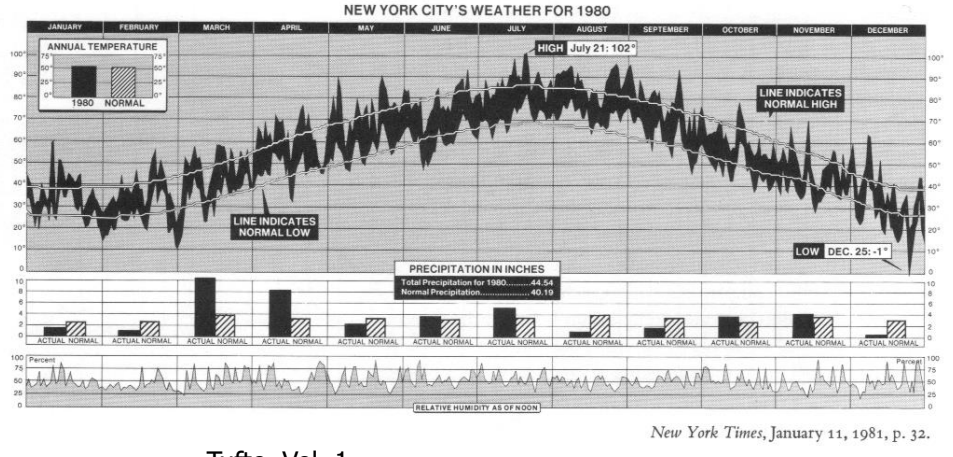
More to come later in term

Back to InfoVis (Examples)

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

NYC Weather

2220 numbers

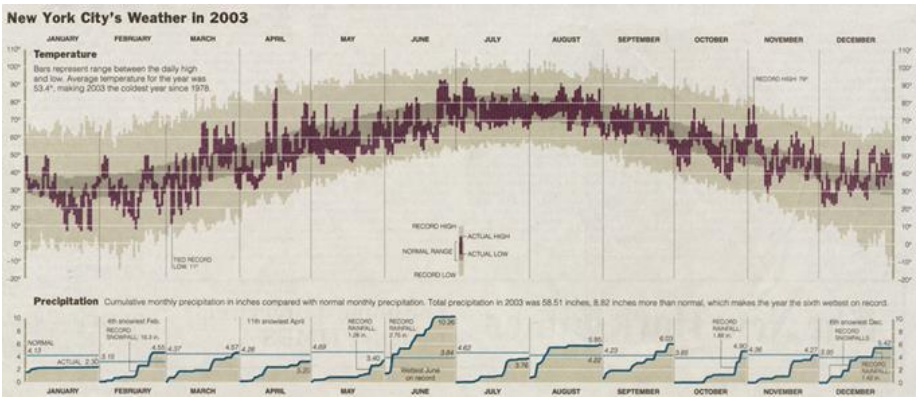


Tufte, Vol. 1

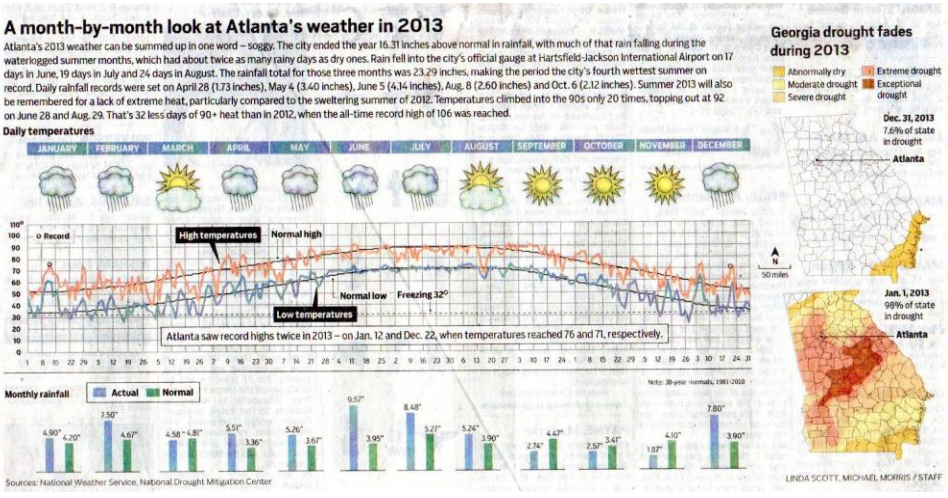
Data Values

365	High temp for each day
365	Low temp for each day
365	Avg high temp for each day
365	Avg low temp for each day
365	Precipitation for each day
365	Humidity for each day
12	Precipitation for each month
12	Avg precipitation for each month
1	Precipitation for the year
1	Avg precipitation per year
1	Highest temp (& day) for the year
1	Lowest temp (&day) for the year
1	Avg daily temp for the year
1	Avg daily temp per year

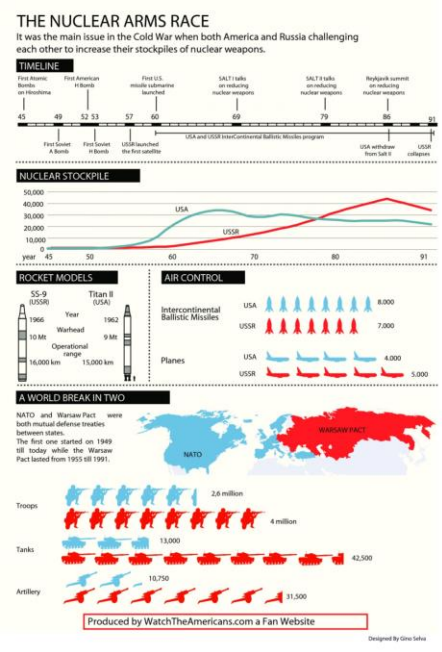
Updated Version



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

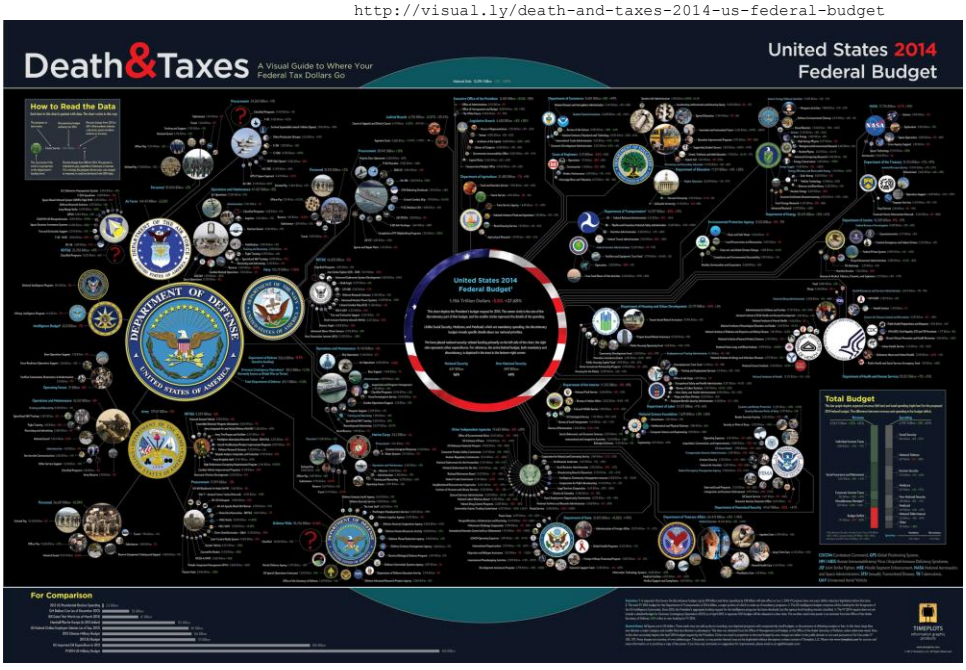


Atlanta Journal Constitution
Jan. 3, 2014



http://visual.ly/nuclear-arms-race

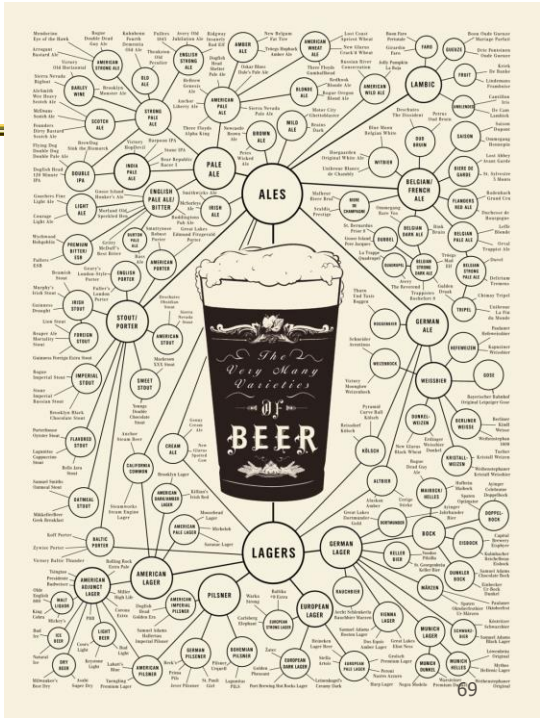
http://visual.ly/death-and-taxes-2014-us-federal-budget



http://www.mikewirthart.com/?cat=3

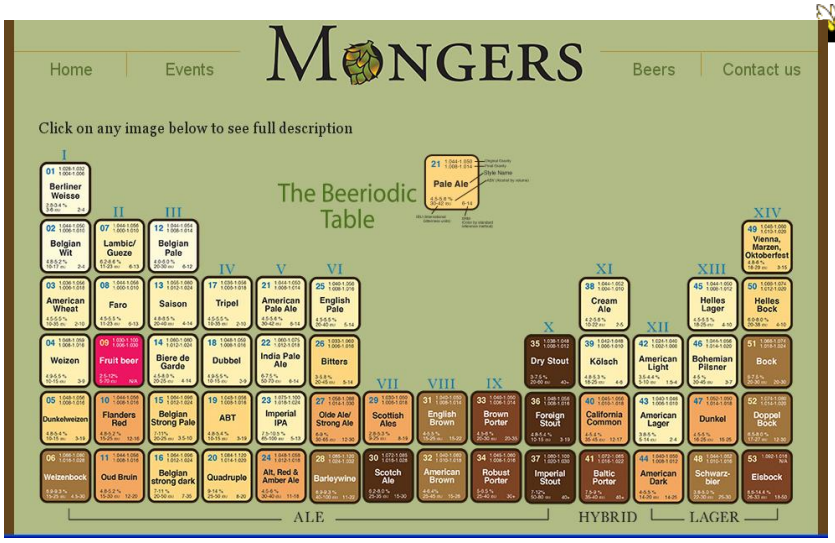
Beer

Beer!



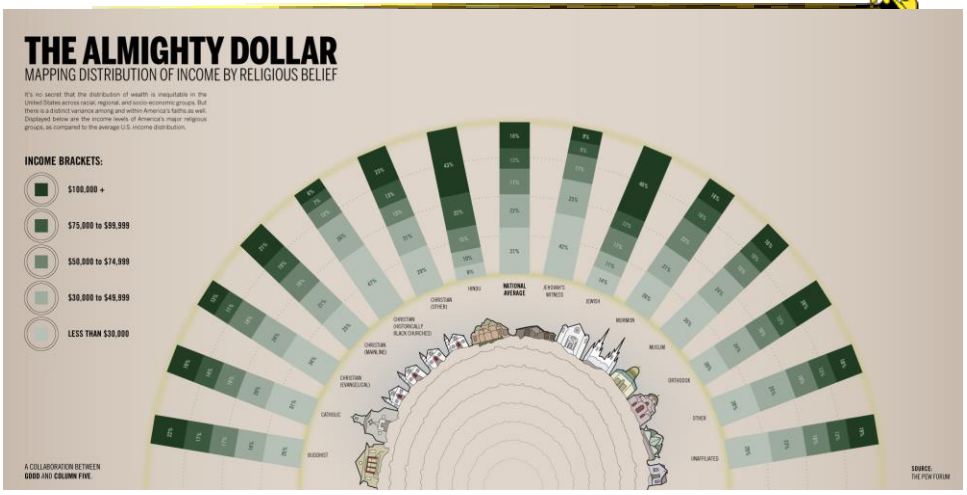
http://images.fastcompany.com/upload/poster_beer_1300.jpg

More Beer!

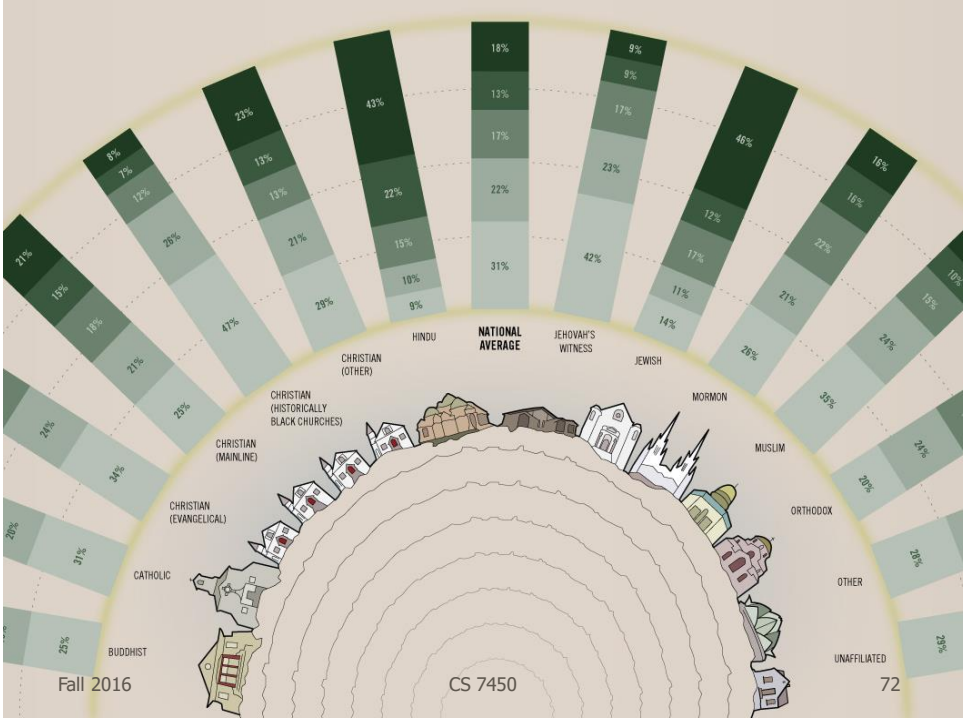


http://thebeermongers.com/beers/

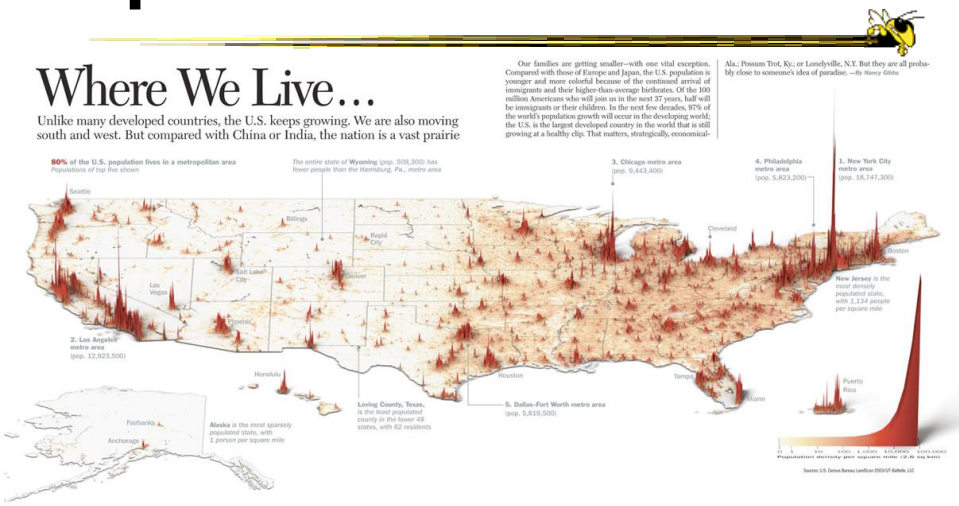
Income and Religion



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

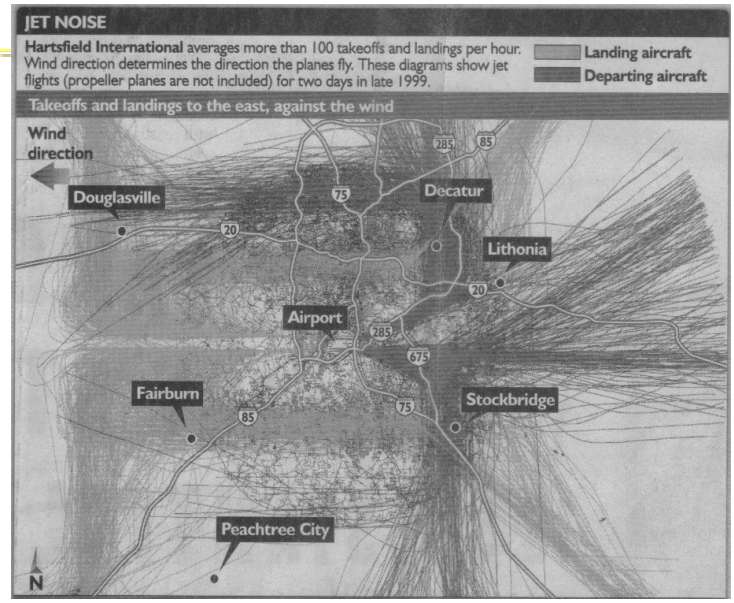


Population



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

Atlanta Flight Traffic



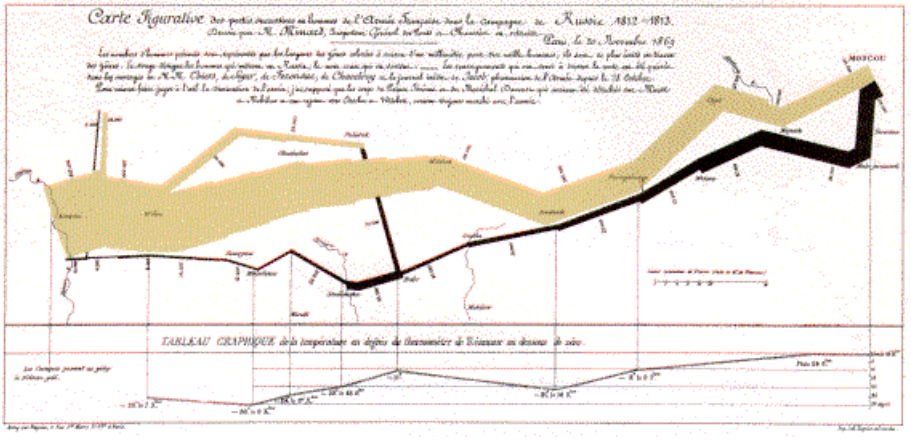
Atlanta Journal
April 30, 2000

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.

Napolean's March

From E. Tufte
The Visual Display of Quantitative Information



Minard graphic

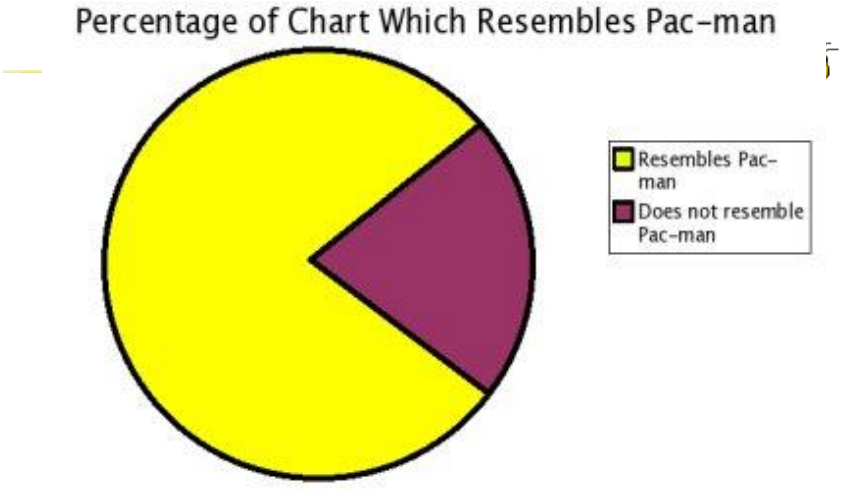
size of army
direction

latitude
longitude

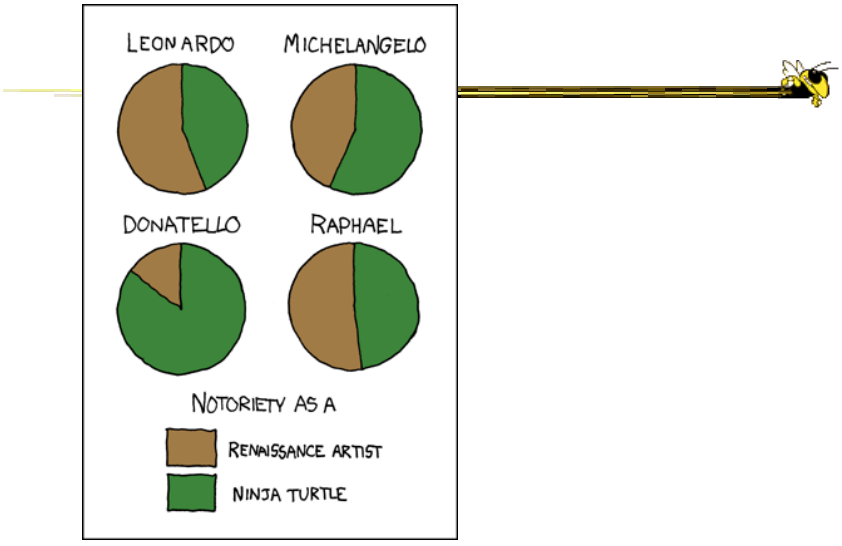
temperature
date

Or, for fun…

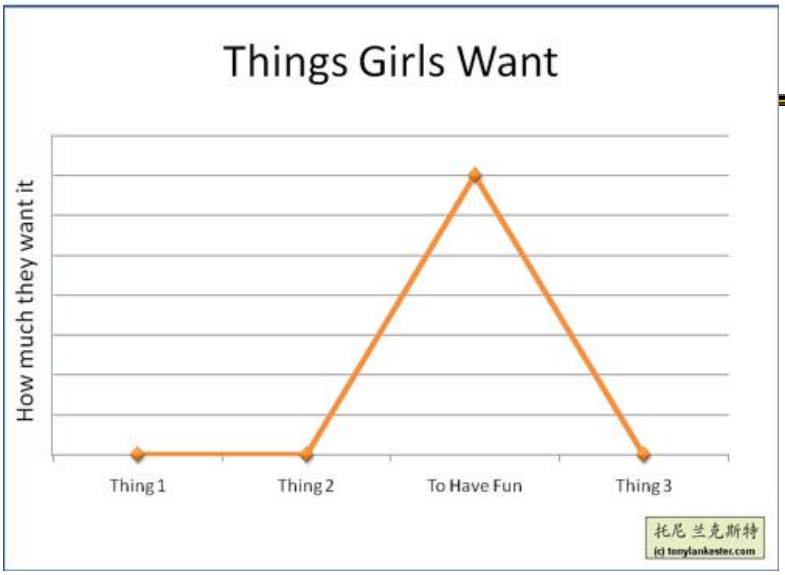
http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html



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



http://xkcd.com/197/

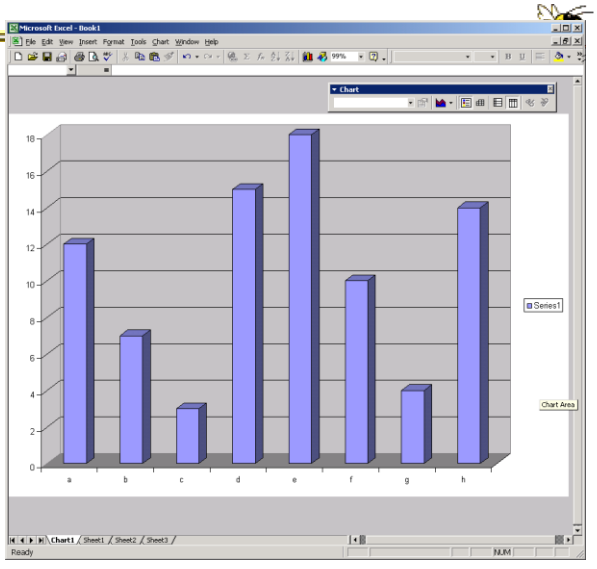


http://www.flickr.com/photos/91884218@N00/3108768440/in/pool-songchart

But Don't Do This

Excel

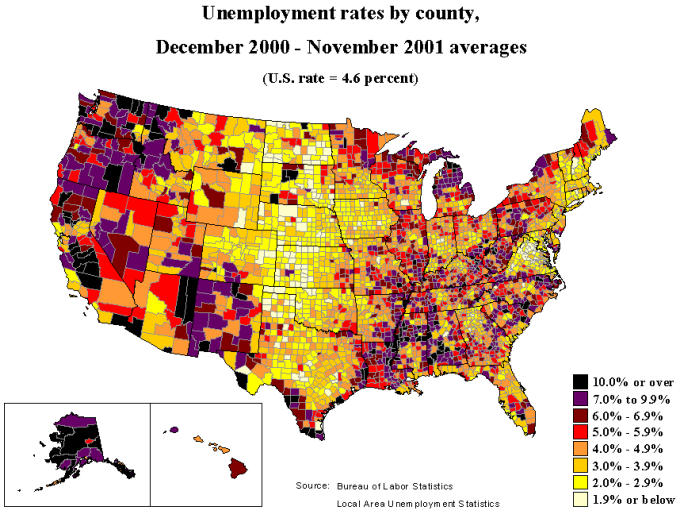
Get rid of those darn 3D bars!



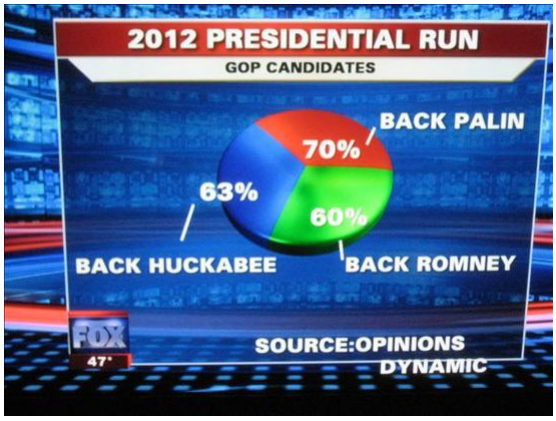
USA Today Graphics

Or worse yet…

Unemployment Rates



FOX “News”



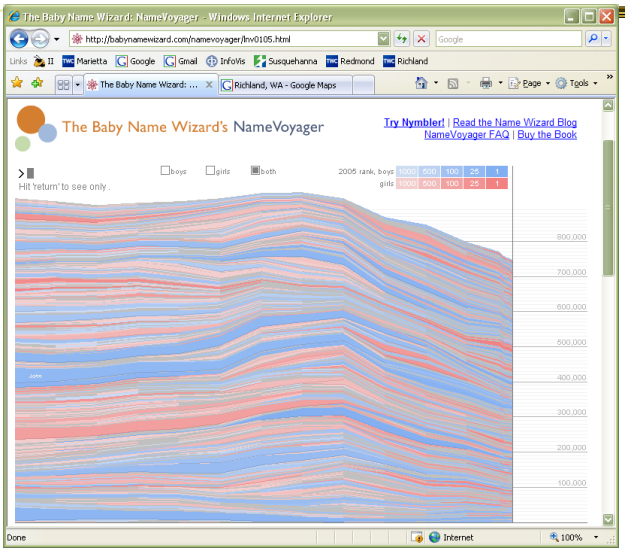
http://wonkette.com/412361/all-193-of-republicans-support-palin-romney-and-huckabee

Examples

- Tools/Systems
  - Now interaction becomes important…

http://babynamewizard.com/namevoyager/

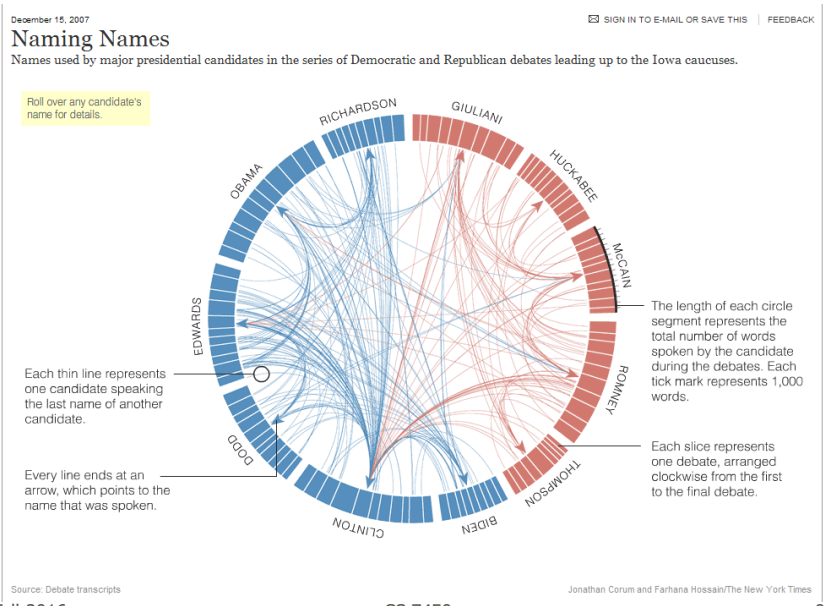
Baby Name Wizard



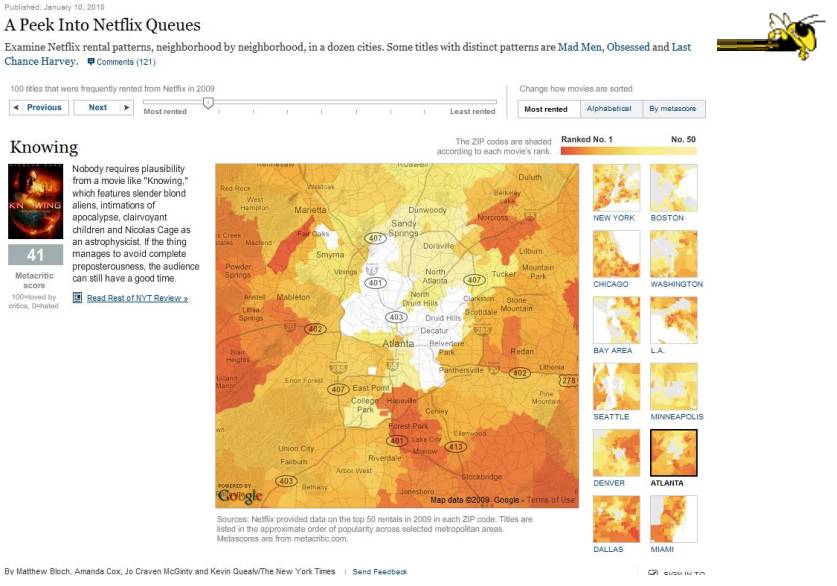
NY Times

- 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

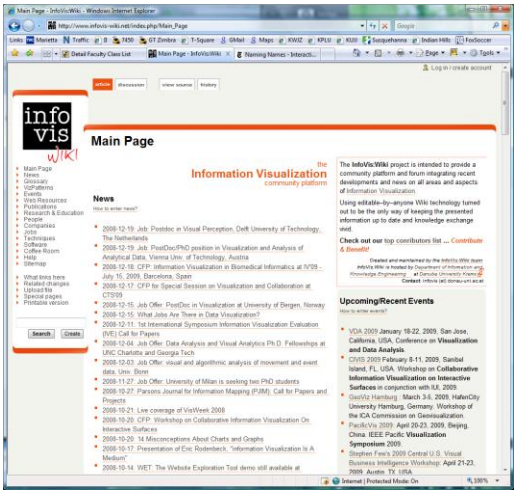


Good Resources

- Some places to look for more information

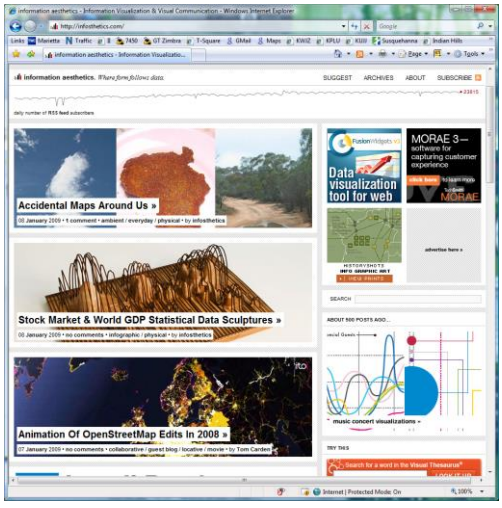
InfoVis Wiki

http://www.infovis-wiki.net



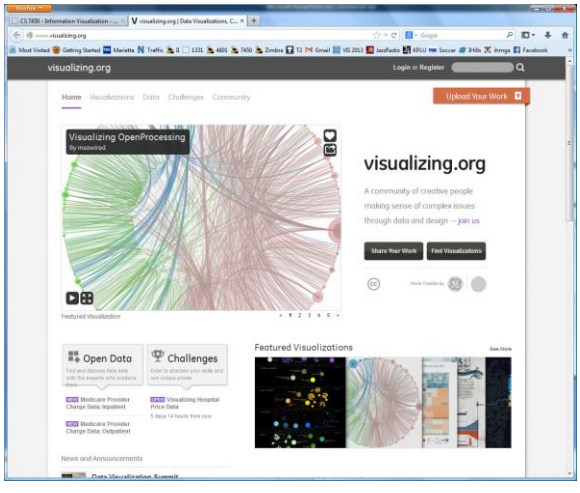
Infosthetics Blog

http://infosthetics.com/



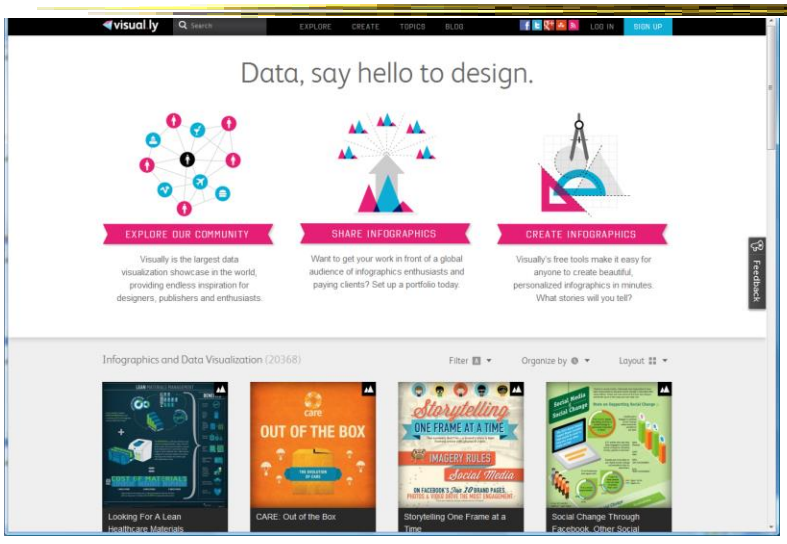
Visualizing.org

http://www.visualizing.org



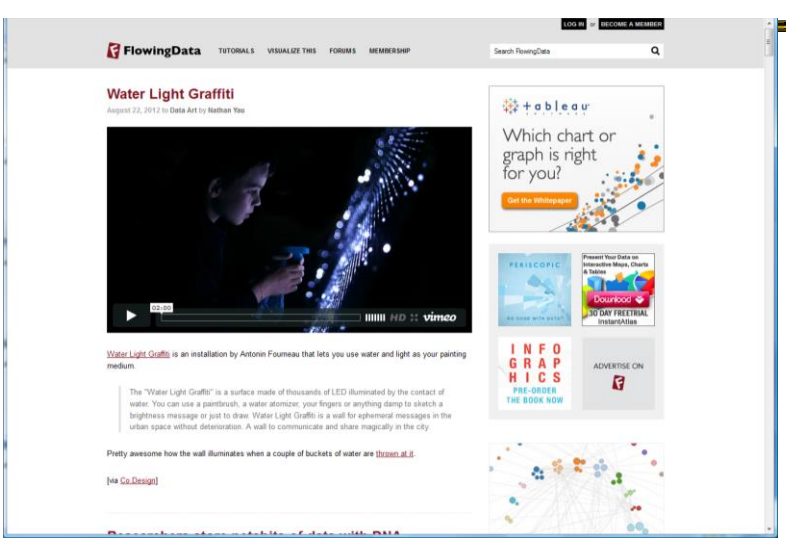
Visual.ly

http://visual.ly/



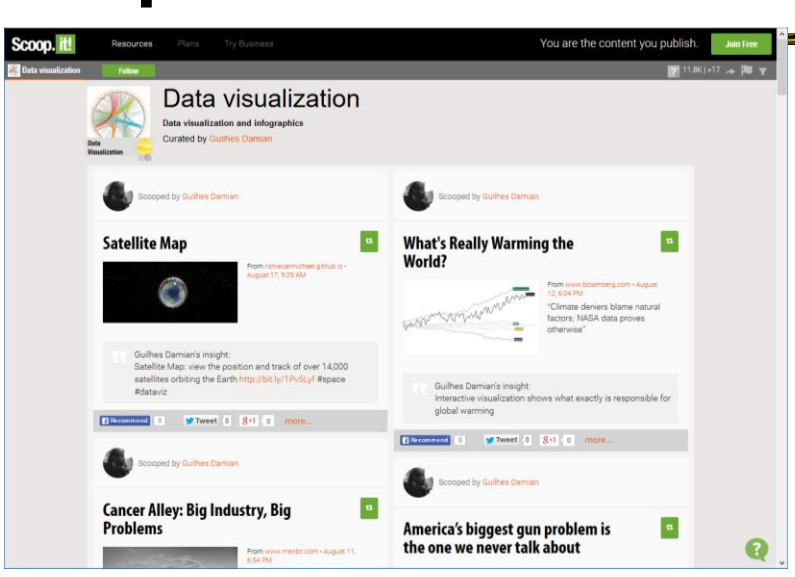
Flowing Data

http://flowingdata.com/



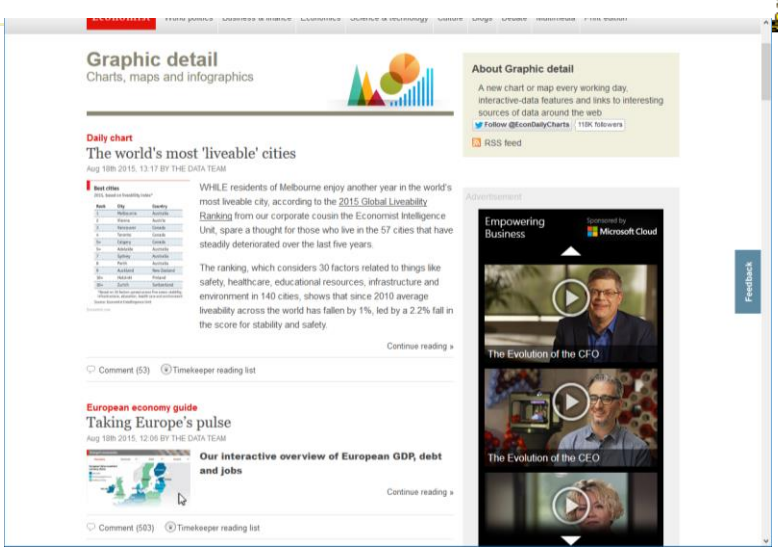
http://www.scoop.it/t/data-visualization-by-guilhes-damian

Scoop.It!



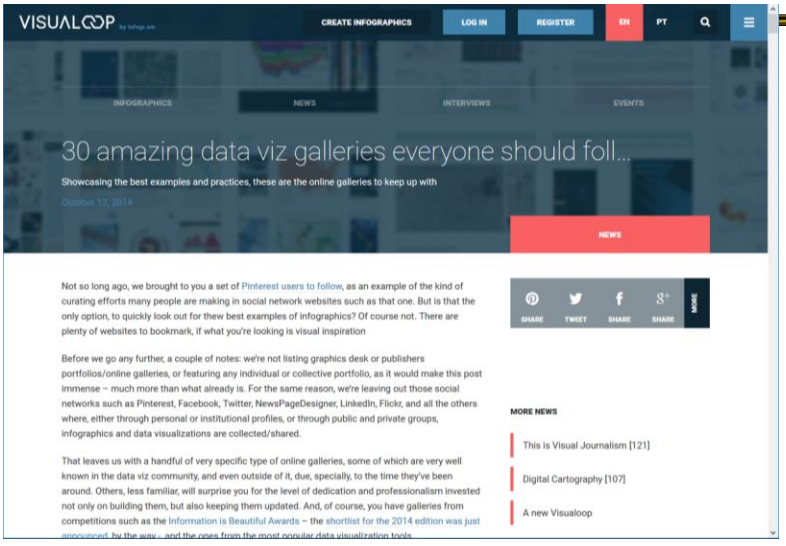
http://www.economist.com/blogs/graphicdetail

Graphic Detail - Economist



http://visualoop.com/blog/11044/30-amazing-data-viz-galleries-everyone-should-follow

A Compendium



Learning Objectives

- Articulate definition and purpose of visualization
- Describe two main uses or applications of visualization
- List two primary components of visualizations
- Describe the different areas of academic visualization research
- Explain the infovis "pipeline" (process)

HW

- HW1 due next Monday
  - Data Exploration and Analysis
  - Bring 2 hardcopies

Reading

- Card, Mackinlay, Shneiderman
  Chapter 1 of their book
- Check out some of the websites on the Schedule page

Upcoming

- Multivariate data & tables
- Graphs & Charts