Overview of InfoVis



CS 4460 – Intro. to Information Visualization Aug. 23, 2017 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)

(Will carry over into next class)

Exercise



Get out pencil and paper





Fall 2017

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



Fall 2017 CS 4460 7

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?

of Chris North

Fall 2017

Two slides courtesy

CS 4460

8

Human Vision



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



Impressive. Lets use it!

Fall 2017 CS 4460

An Example



Why visualization helps...

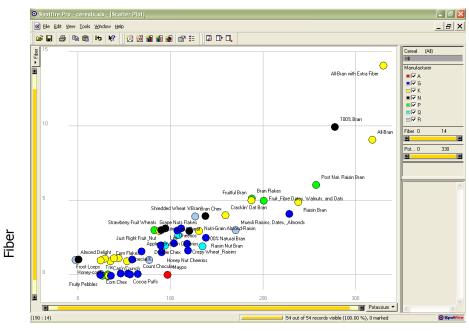
Which cereal has the most/least potassium?

Is there a relationship between potassium and fiber? If so, are there any outliers?

Questions:

Which manufacturer makes the healthiest cereals?

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



Fall 2017 Potassium

CS 4460

12

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 2017 CS 4460 13



Another Illustrative Example

Four Data Sets



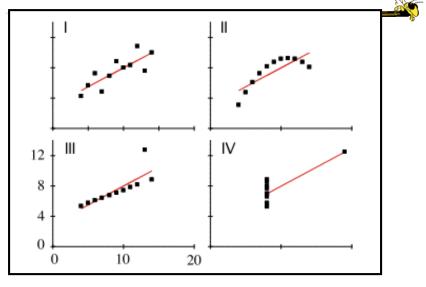
- Mean of the x values = 9.0
- Mean of the y values = 7.5
- Equation of the least-squared regression line is: y = 3 + 0.5x
- Sums of squared errors (about the mean) = 110.0
- Regression sums of squared errors (variance accounted for by x) = 27.5
- Residual sums of squared errors (about the regression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

Anscombe's quartet

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

Fall 2017 CS 4460 1

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

Fall 2017 CS 4460 17

Exercise Redux



• Let's check what you did...

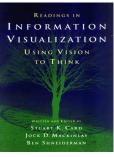
People work differently

Visualization



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

From



Fall 2017 CS 4460 19

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

Fall 2017 CS 4460 21

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"



Fall 2017 CS 4460 2

Admin Intermission



- Course policies
- Grading
- Overloads
- Surveys
- More...



Administratia



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

http://www.cc.gatech.edu/~stasko/4460



Fall 2017 CS 4460 25

T-Square Site

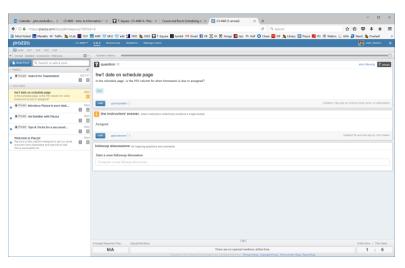


Should've seen an announcement

https://piazza.com/gatech/fall2017/cs4460

Piazza





Fall 2017 CS 4460 27

Responsibility



 You are responsible for examining and staying up-to-date with information on the class website and t-square website

Course Policies



- Most on class homepage or Assignments page
 - Format
 - Book
 - Attendance
 - Electronics
 - Academic Integrity
 - Grading

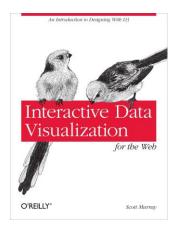
– ...

Fall 2017 CS 4460 29

Books



None required



Will use

Ebook free at http://chimera.labs.oreilly.com/books/1230000000345/

Colored Pencils





Please get some and bring to class when requested

Fall 2017 CS 4460 31

Attendance



- Expected
 - Part of your grade
- Will start promptly at 12:20 and end by 1:10
- Eating (quietly) is OK

Academic Integrity



- Do your own work, unless told otherwise
- Absolutely OK to consult me or Tas
- More to come on programming HWs...

Fall 2017 CS 4460 3

Electronics



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





Grading



 Pop quizzes/Attendance 	10%
--------------------------------------------	-----

Start of class, 5 minutes

HW Assignments (5)20%

Programming Assignments (5) 28%

Last one bigger

Midterm Exam20%

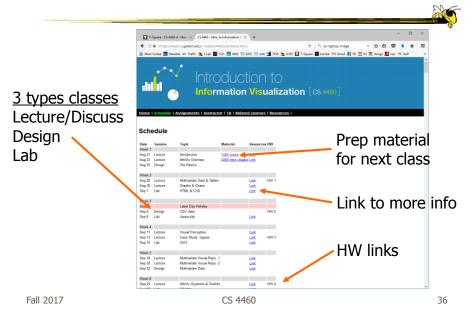
• Final Exam 22%

At worse 60-70-80-90 for final grades

Fall 2017 CS 4460 35

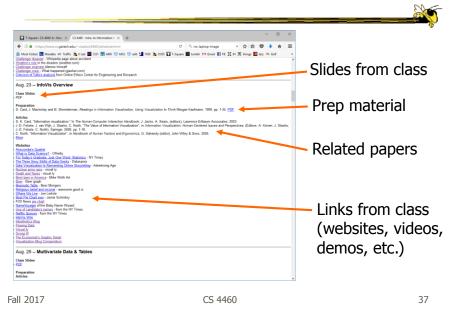
Schedule

Info central



Schedule

Lower on page



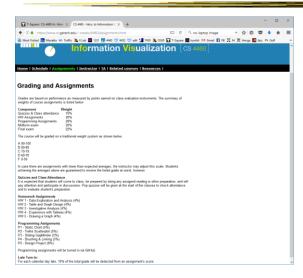
Class Preparation



- Paper, video, website to review for the next class
 - Multiple links from Schedule page
 - Potential pop quiz at start of next class

Assignments





Recaps weights

Will have links to HWs

Fall 2017 CS 4460

Instructor





Office number out of sequence with others (in corner of building)

About Me



CS PhD, 1989, Brown University

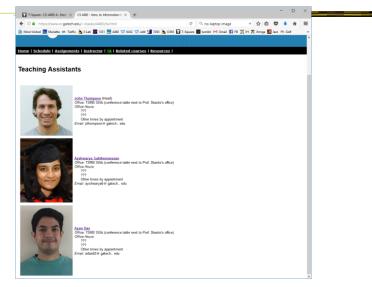
Researcher in information visualization, visual analytics, and HCI

Active in those academic research communities



Fall 2017 CS 4460 41

Teaching Assistants



CAVEAT



- This course is quite a bit of work. If you're just looking for some easy grade, I would advise you to drop now
- Graduating seniors: It's on you now to do the work so no problems later
- If you are sincerely interested in this topic, I hope you will enjoy the course and learn a lot

Fall 2017 CS 4460 43

Registration



- Will be able to add some students
 - Room has a little more capacity
 - More if there are drops
 - Cannot go too large
- Not a fan of "seat squatting"
- Please drop the class by Thursday noon

Survey



- Who wasn't here on Monday and didn't fill out a survey?
 - Please fill out even if on waitlist

Fall 2017 CS 4460 4



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

Fall 2017 CS 4460

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 2017 CS 4460 49



EDA example?

EDA Example



- 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 2017 CS 4460 51

2. Communication



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



When to Apply?



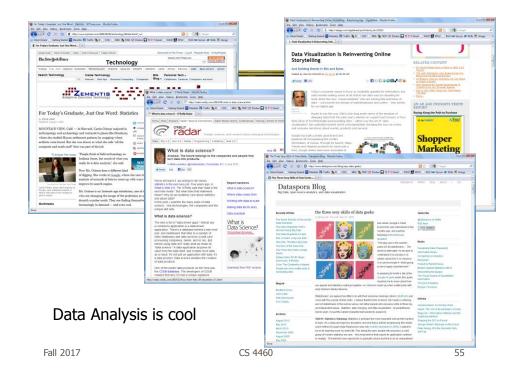
- Visuals can frequently take the place of many words
- Visuals can summarize, aggregate, unite, explain, ...
- Sometimes words are needed, however

Fall 2017 CS 4460 53

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



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]

Fall 2017 CS 4460 57

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

Fall 2017 CS 4460 59

Two Key Challenges



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

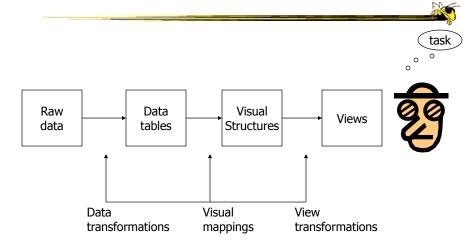
Example Domains for Info Vis



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

Fall 2017 CS 4460 61

InfoVis Process Model



From: Card, Mackinlay, Shneiderman '99

Back to InfoVis (Examples)



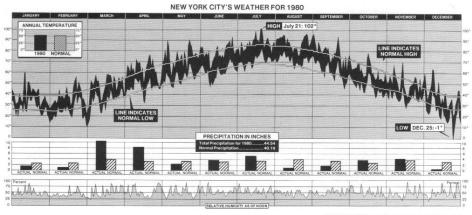
- Start with static pictures (InfoGraphics)
 - Very popular on the web
 - But are they information visualizations?

Fall 2017 CS 4460 63

NYC Weather



2220 numbers



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

Tufte, Vol. 1

Data Values

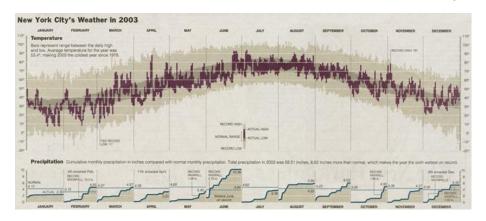


- 365 High temp for each day
- 365 Low temp for each day
- 365 Avg high temp for each day
- 365 Avg low temp for each day
- 365 Precipitation for each day
- 365 Humidity for each day
- 12 Precipitation for each month
- 12 Avg precipitation for each month
- 1 Precipitation for the year
- 1 Avg precipitation 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

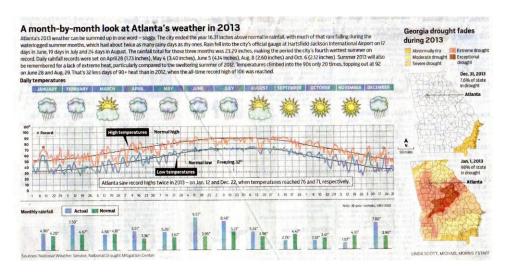
Fall 2017 CS 4460

Updated Version

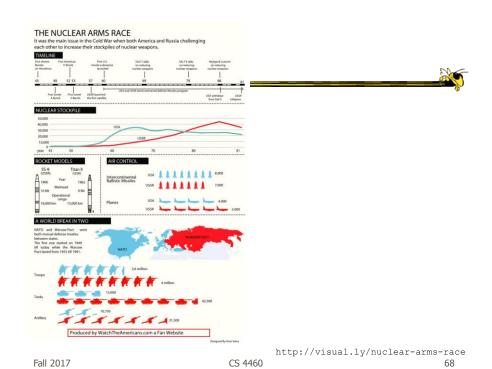




http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=00014g



Atlanta Journal Constitution Jan. 3, 2014



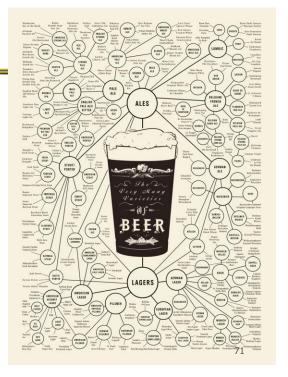


http://www.mikewirthart.com/?cat=3

Beer



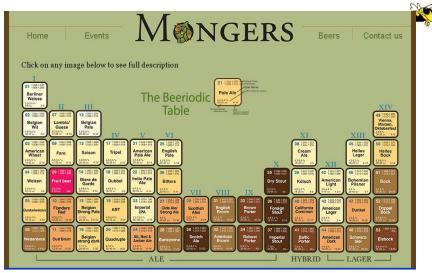
Beer!



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

Fall 2017

More Beer!

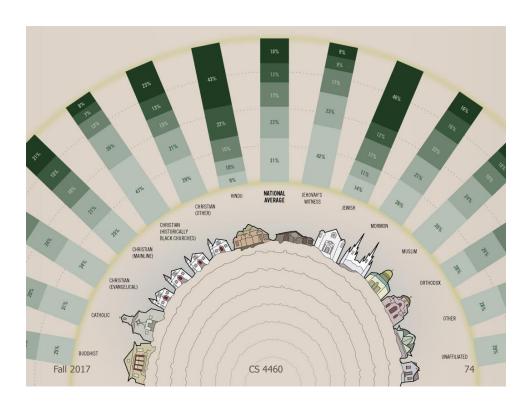


http://thebeermongers.com/beers/

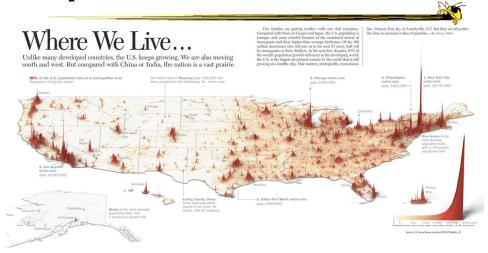
Income and Religion



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



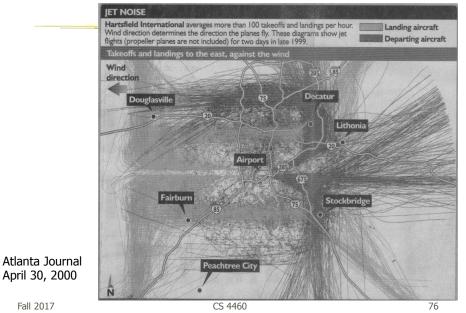
Population



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

Fall 2017 CS 4460

Atlanta Flight Traffic



Country Music



77

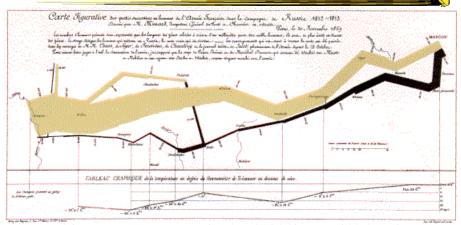


Figure 14. States Mentioned in Country-Music Lyrics Source: Ben Marsh, "A Rose-Colored Map," Hanper's, July 1977, 80. Used by permission. Note: The size of each state is proportional to the number of times it is mentioned.

Fall 2017 CS 4460

Napolean's March

From E. Tufte
The Visual Display of
Quantitative Information



Minard graphic

size of army direction

latitude longitude temperature date

Fall 2017

CS 4460

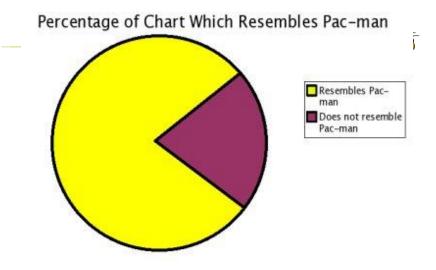


Or, for fun...

Fall 2017 CS 4460 79

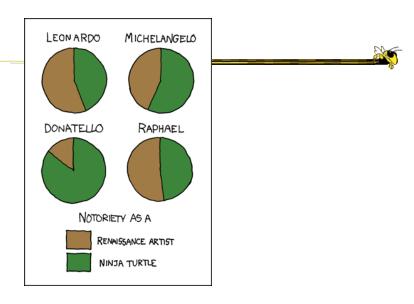


http://infosthetics.com/archives/2008/09/funniest_pie_chart_ever.html

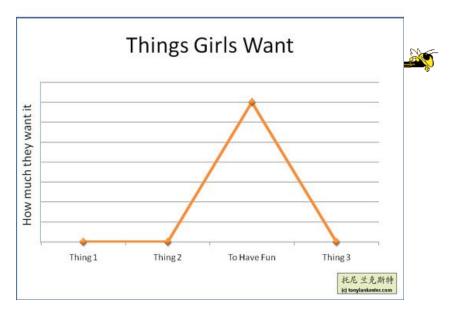


http://www.boingboing.net/2006/11/02/hilarious-piechartvi.html

Fall 2017 CS 4460 81



http://xkcd.com/197/



 $\label{linear_com_photos_91884218@N00/3108768440/in/pool-songchart} $$\text{CS 4460}$ 83$

Fall 2017

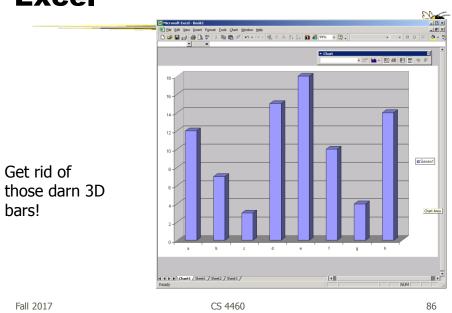




But Don't Do This

Fall 2017 CS 4460 8

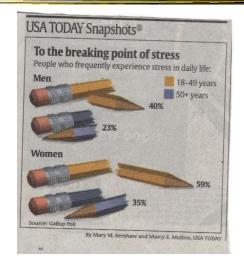
Excel



USA Today Graphics



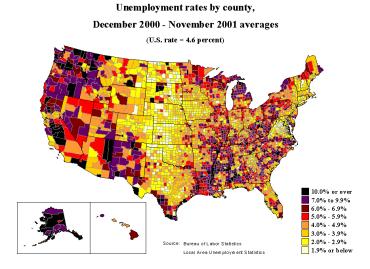
Or worse yet...



Fall 2017 CS 4460 87

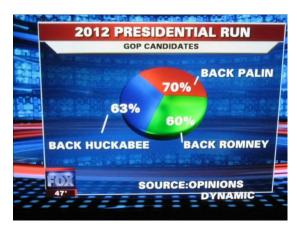
Unemployment Rates





FOX "News"





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

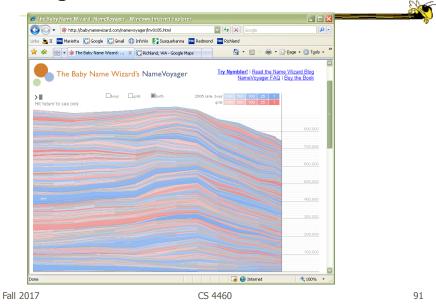
Fall 2017 CS 4460 89

Examples



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

Baby Name Wizard

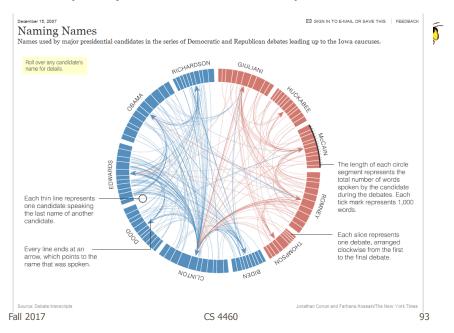


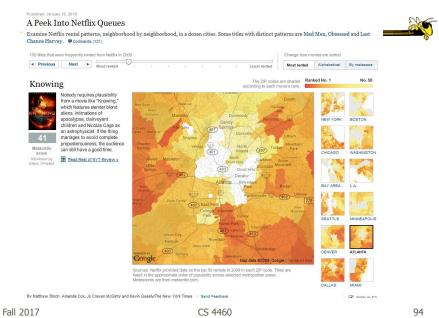
NY Times



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

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





Good Resources



Some places to look for more information

Fall 2017 CS 4460 99

InfoVis Wiki

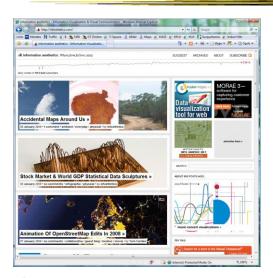
http://www.infovis-wiki.net





Infosthetics Blog



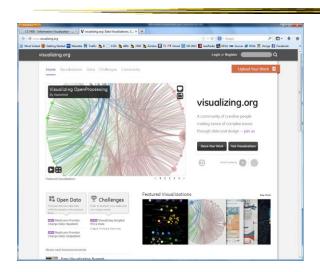


Fall 2017 CS 4460

Visualizing.org

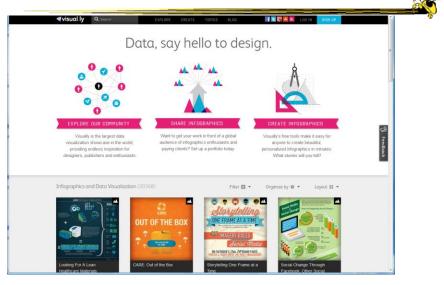
http://www.visualizing.org





Visual.ly

http://visual.ly/



Fall 2017 CS 4460 99

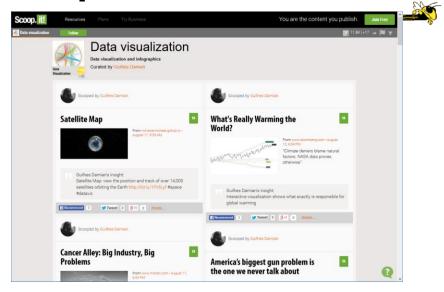
Flowing Data

http://flowingdata.com/



 $\verb|http://www.scoop.it/t/data-visualization-by-guilhes-damian|$

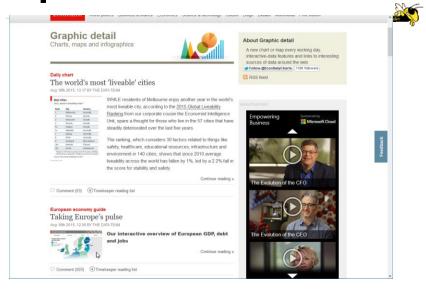
Scoop.It!



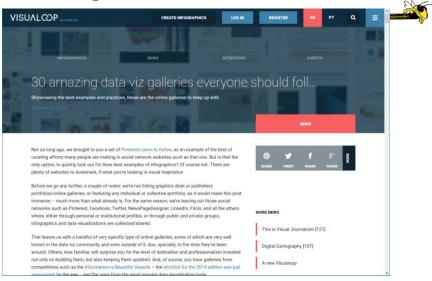
Fall 2017 CS 4460 101

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

Graphic Detail - Economist



A Compendium



Fall 2017 CS 4460 103

Upcoming



- Design Exercise
 - Bring colored pencils
- Data & Tables