i214 Reporting quantitative data

November 4, 2008

(with thanks to Gary Kass and Edward Tufte)

Why graphs?

Ex: The Harris Poll #100, Oct 15, 2007

American adults have their choice of many different cuisines ... From pasta to burritos, the choices are almost endless. But what do Americans choose when eating out? American food, of course!

Over one-quarter (28%) of U.S. adults say American food is what they are most likely to choose if they had the choice to go out to a restaurant and eat one type of food. Just under one-quarter (22%) say that they would most likely choose Italian while 17 percent would choose Mexican while 16 percent would choose Chinese if they had the choice to go out to a restaurant. Japanese is the next choice as seven percent say they would choose this type of cuisine. Much further down the list are Indian (2% say they would choose) and then French and Middle Eastern cuisine (1% would choose each). Finally, four percent say they would choose another type of food.

or...

TABLE 1

TYPE OF FOOD FOR EATING OUT

"Thinking of food now, if you had the choice to go out to a restaurant and eat one type of food, which of these are you most likely to choose?"

Base: All Adults

	Total	Region				Generation			
		East	Midwest	South	West	Echo Boomers (18-30)	Gen Xers (31-42)	Baby Boomers (43-61)	Matures (62+)
	%	%	%	%	%	%	%	%	%
American	28	29	27	33	21	21	25	28	41
Italian	22	31	21	18	20	23	18	24	22
Mexican	17	9	19	16	27	17	24	16	14
Chinese	16	14	21	15	17	18	12	19	13
Japanese	7	9	4	10	6	12	9	5	3
Indian	2	2	2	1	3	3	3	1	1
French	1	2	1	2	1	1	1	1	3
Middle-Eastern	1	1	1	1	1	1	*	2	1
Other	4	3	4	4	5	3	7	4	2
None of these	1	1	1	1	*	2	1	*	*

Note: Percentages may not add up to 100% due to rounding

Note: * = less than 0.5%

UC Berkeley School of Information

Needs and Usability Assessment 11.04.08

Reporting results the easy way

exact questions and answers

SE9. How confident do you feel about your own searching abilities
when using a search engine to find information online?

**Based on Internet users who have a used a search engine
[N=1,165]**
%
48   Very confident
44   Somewhat confident
6   Not too confident
2   Not confident at all
*   Don't know/Refused

http://www.pewinternet.org/pdfs/PIP_Search_Questions.pdf

Tabular displays

Comparing dependent to independent variables

	Undergrads % n=120	Grads % n=200	Total % n=320
Satisfied	60 *N=71*	13 *N=25*	31 *N=96*
Dissatisfied	40 *N=47*	87 *N=165*	69 *N=127*
Total	100 *n = 118*	100 *n = 190*	100 *n = 308*
No ans.	*n = 2*	*n = 10*	*n= 12*

UC Berkeley School of Information

Needs and Usability Assessment 11.04.08

Good information graphics

Efficient

minimizing non-data ink

Meaningful

displaying information relevant to the argument

Unambiguous

Self-explanatory, or at least difficult to misread

By 2030, Nearly Two-Thirds of the World's Population Will Live in Urban Areas.

Percent urban
■ 1950 ■ 2007 ■ 2030

World: 29, 49, 60
Africa: 15, 37, 51
Asia: 17, 41, 54
Europe: 51, 72, 78
Latin Am./Caribbean: 42, 76, 84
North America: 64, 79, 87

Source: United Nations, *World Urbanization Prospects: The 2005 Revision* (2006) and Carl Haub, *2007 World Population Data Sheet.*

http://www.prb.org/Articles/2007/623Urbanization.aspx

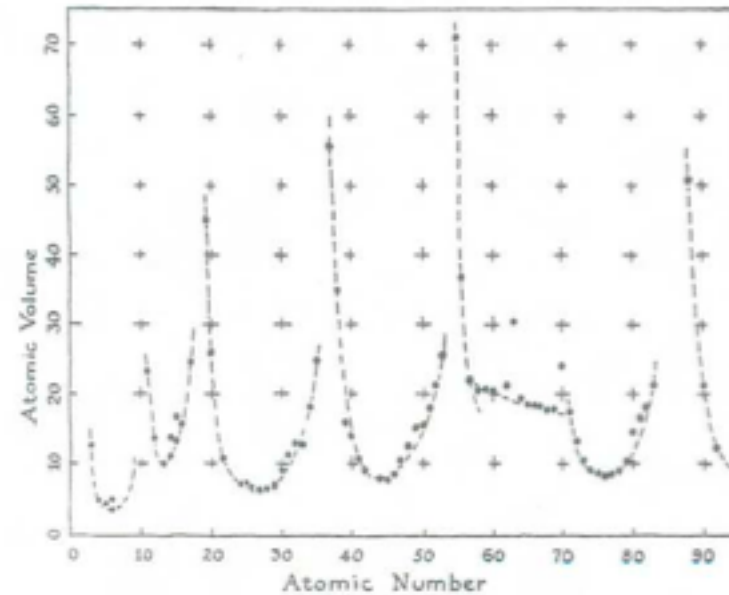# 5 principles from E. Tufte

Above all else show the data

Maximize the data-ink ratio

Erase non-data-ink, within reason

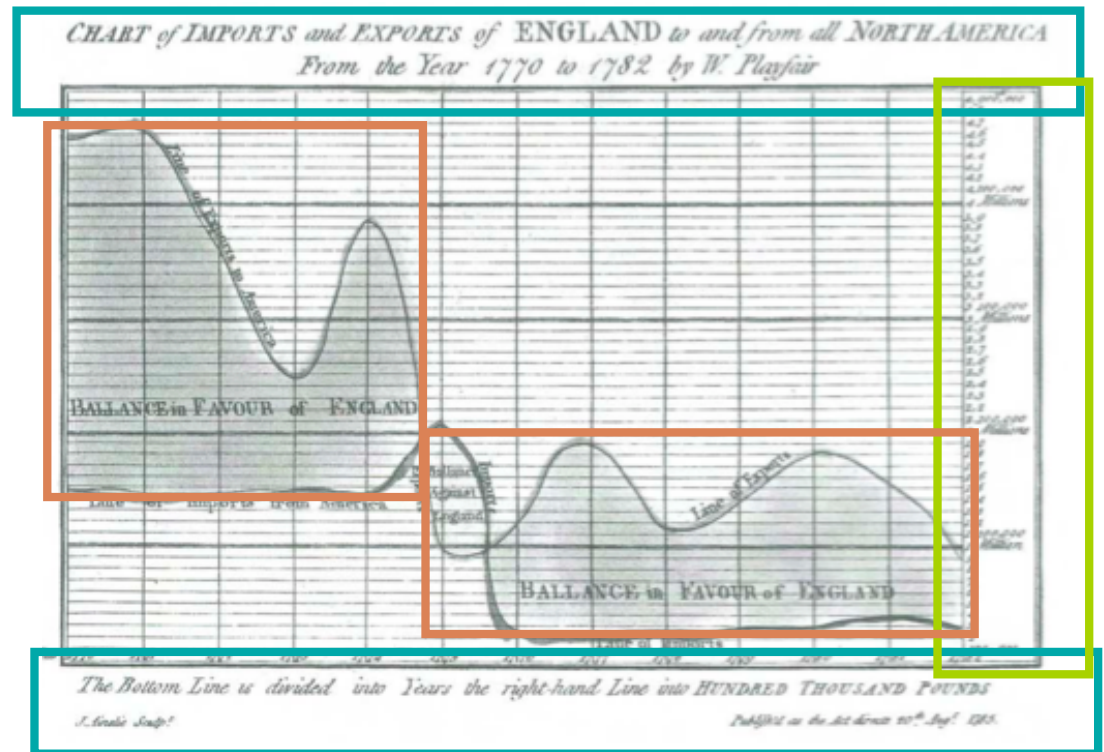Erase redundant data-ink

Revise and edit
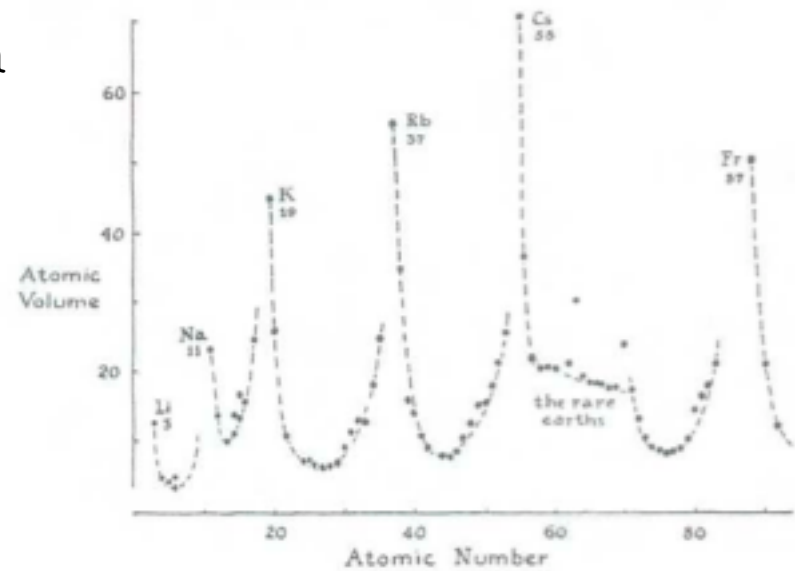
Components of a chart

labeling
Title
axis titles and labels
legends
notes

scales of X, Y axis

graphic
representations
of data

CHART of IMPORTS and EXPORTS of ENGLAND to and from all NORTH AMERICA
From the Year 1770 to 1782 by W. Playfair

BALLANCE in FAVOUR of ENGLAND

BALLANCE in FAVOUR of ENGLAND

The Bottom Line is divided into Years the right-hand Line into HUNDRED THOUSAND POUNDS

Area chart

How the dependent
relationship of parts of
a whole varies over an
independent variable

Percentage: sum of the two
always represents 100%
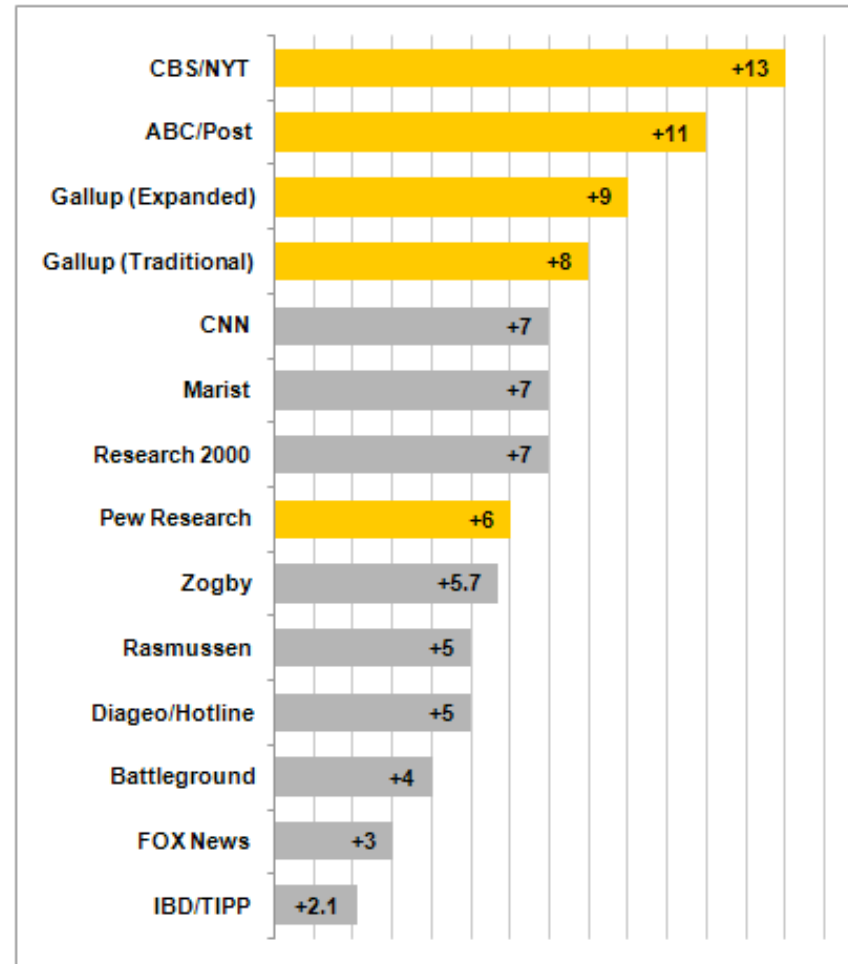
Cumulative: sum varies
according to elements
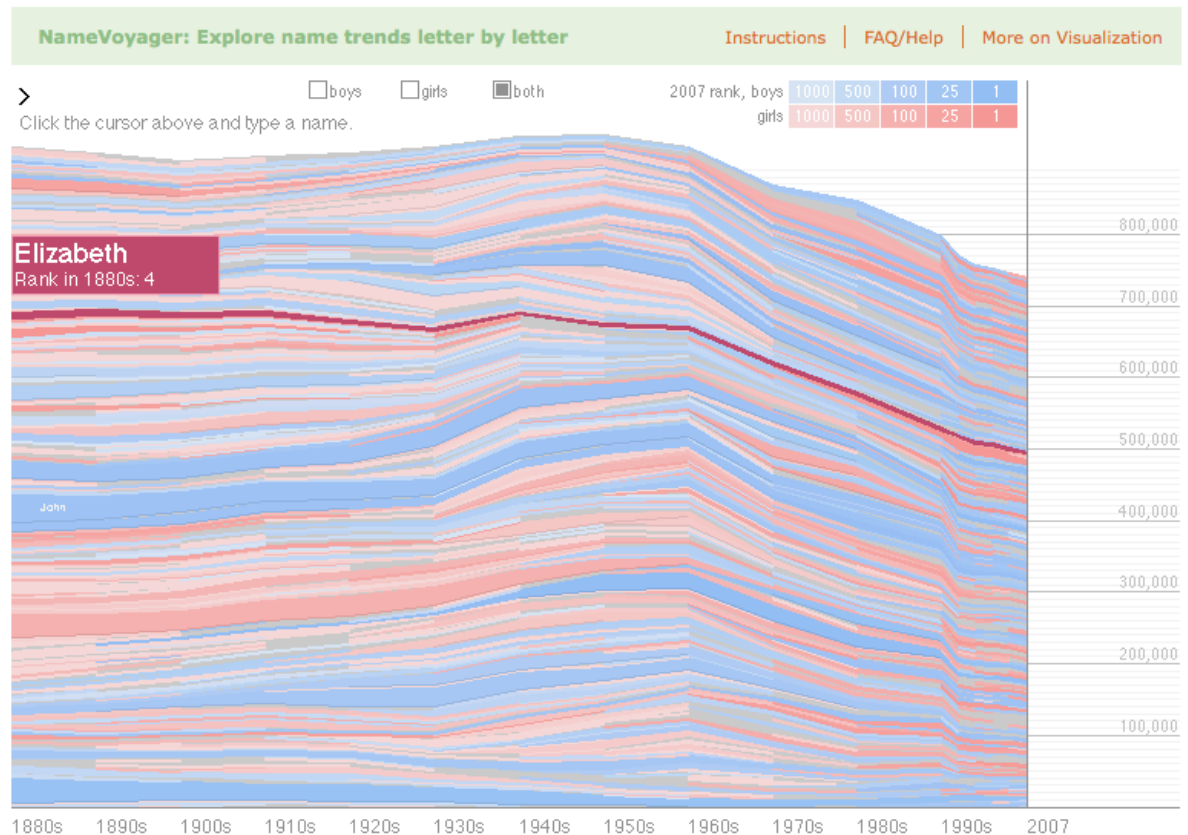
Column or bar

Change within small data sets that vary over a nominal (ie, category) or interval scale

Column: vertical rectangles

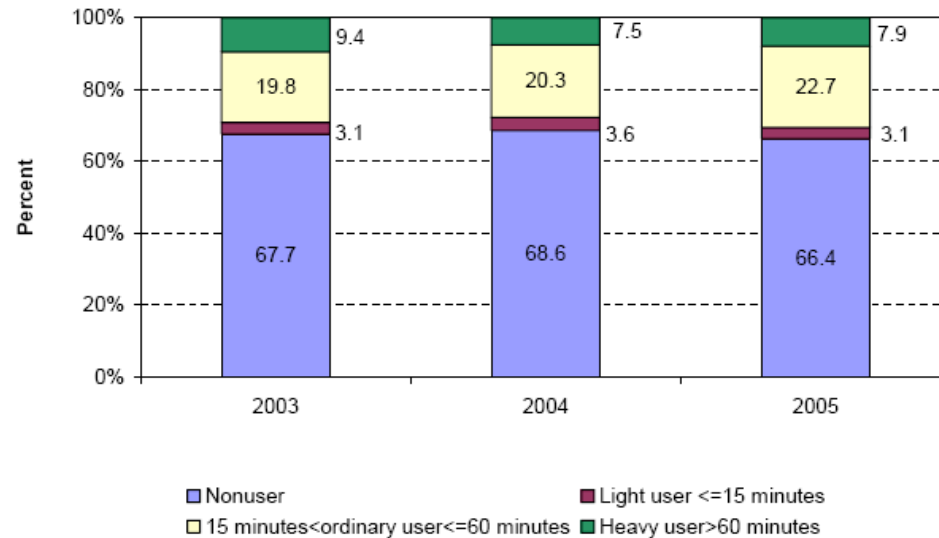Bar: horizontal

Segmented column

A column chart that represents the part-whole relationships of elements

II. SURVEY RESULTS

1. Internet Use: What Percentage, How Much, When, Where, What Activities and with Whom?

1a. Percentage: Among 4,384 respondents 18 to 64 years old, 33.6% used the Internet on the day before being surveyed. This result is an increase of 2.2 percentage points compared to last year. Over the past three years, the percentage of Internet users has been fairly stable (around 32%). (See Graph 1)
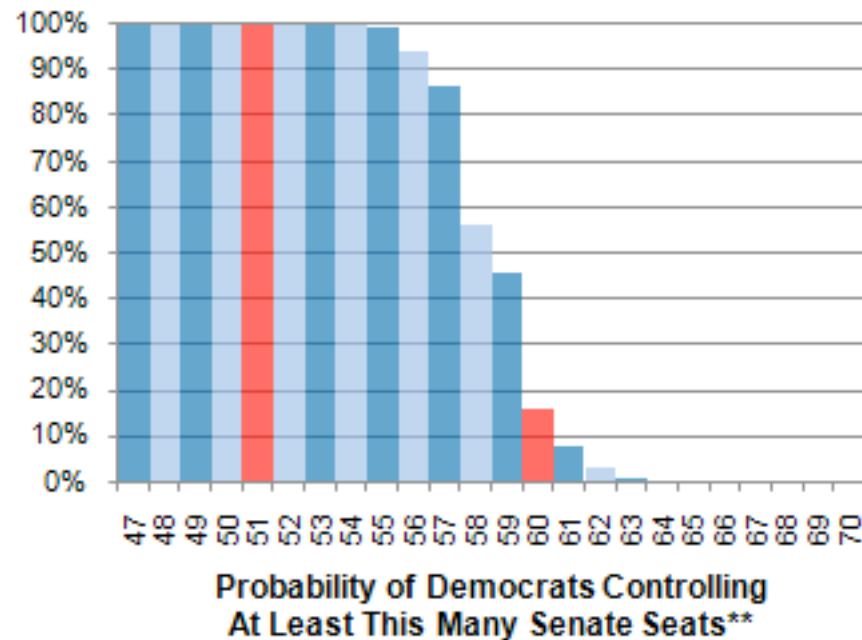
Graph 1: Level of Internet Use and Percentages
(Based on 6 Hour Results)

Year	Nonuser	Light user	15<ordinary≤60	Heavy user
2003	67.7	3.1	19.8	9.4
2004	68.6	3.6	20.3	7.5
2005	66.4	3.1	22.7	7.9

- Nonuser
- Light user <=15 minutes
- 15 minutes<ordinary user<=60 minutes
- Heavy user>60 minutes

UC Berkeley School of Information

Needs and Usability Assessment 11.04.08

Frequency polygon; Histogram

Overall shape
represents
distribution of
elements

**Probability of Democrats Controlling
At Least This Many Senate Seats**

** Counting Sanders (I-VT), Lieberman (I-CT)

Line charts

Used to interpolate between data points and extrapolate future values
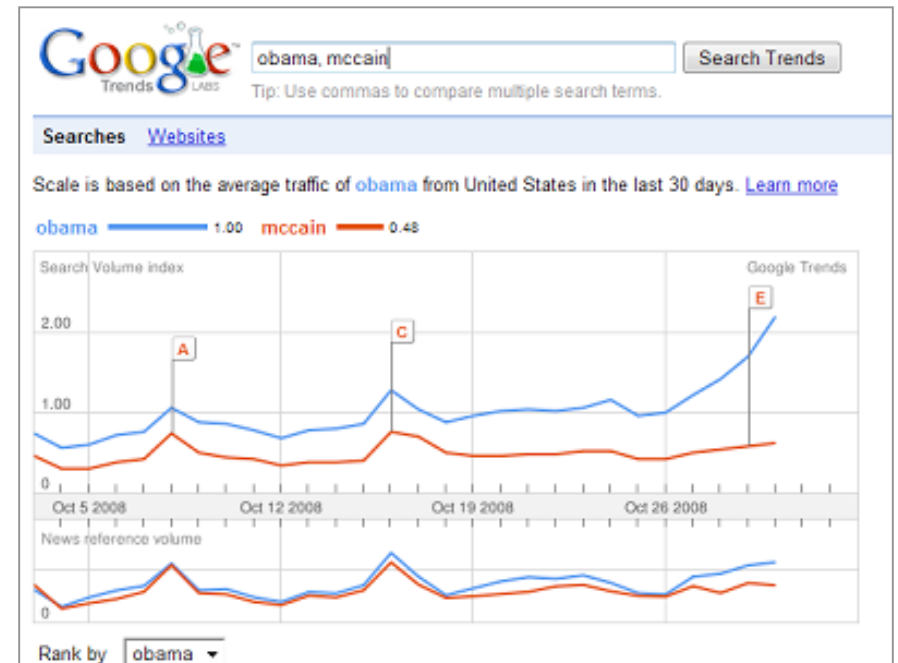
Trends or change over time

Looking for correlations between variables

Lines: Connect "real" data points

Curves: represent functional relations between data points or to interpolate data

X must have numerical values!

http://www.fivethirtyeight.com/2008/11/google-traffic-suggests-mccain-not.html
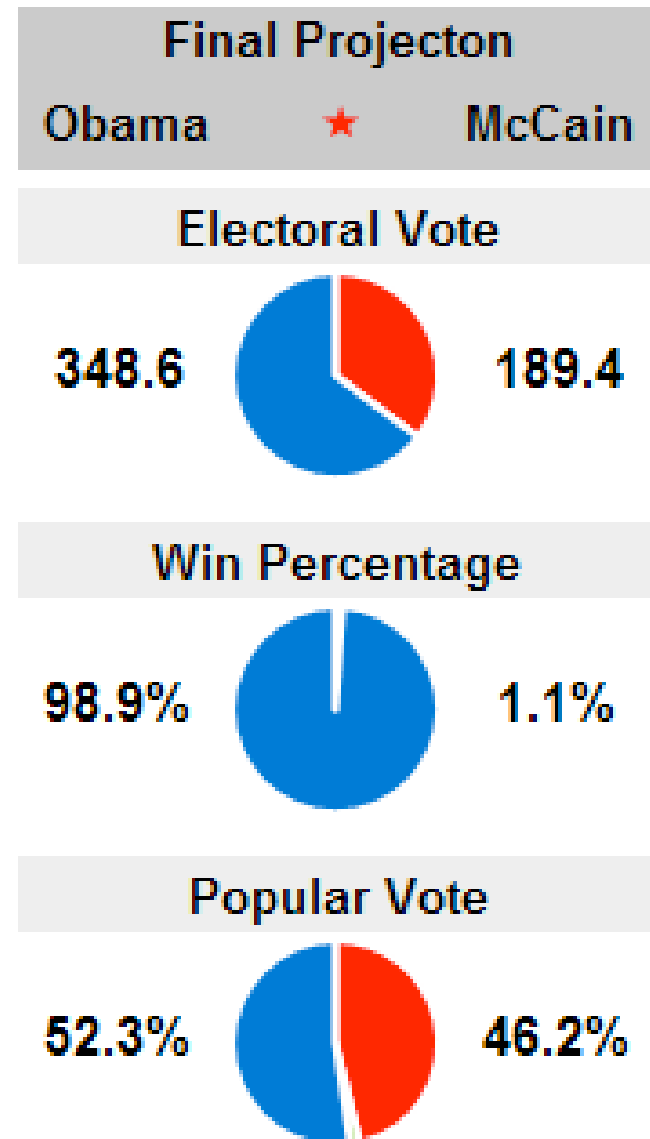
Pie chart

Good for displaying
approximate percentages
at a single point in time

Caution:

Makes exact comparisons
difficult

You cannot necessarily
compare pie charts to each other

Maybe use a bar chart instead?

Final Projecton

Obama   ★   McCain

Electoral Vote

348.6   189.4

Win Percentage

98.9%   1.1%

Popular Vote

52.3%   46.2%

Scatterplot

One dimensional
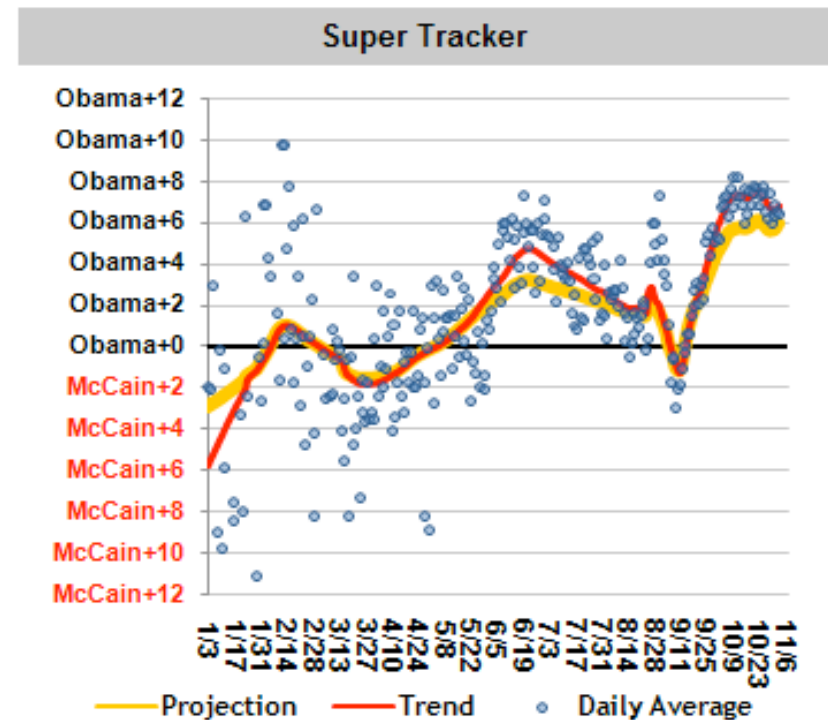
Show measurements over time

Two-dimensional

Convey an overall impression of
the relation between two variables

Some caution:

Not good for following trends

Should not be used with more
than one independent variable

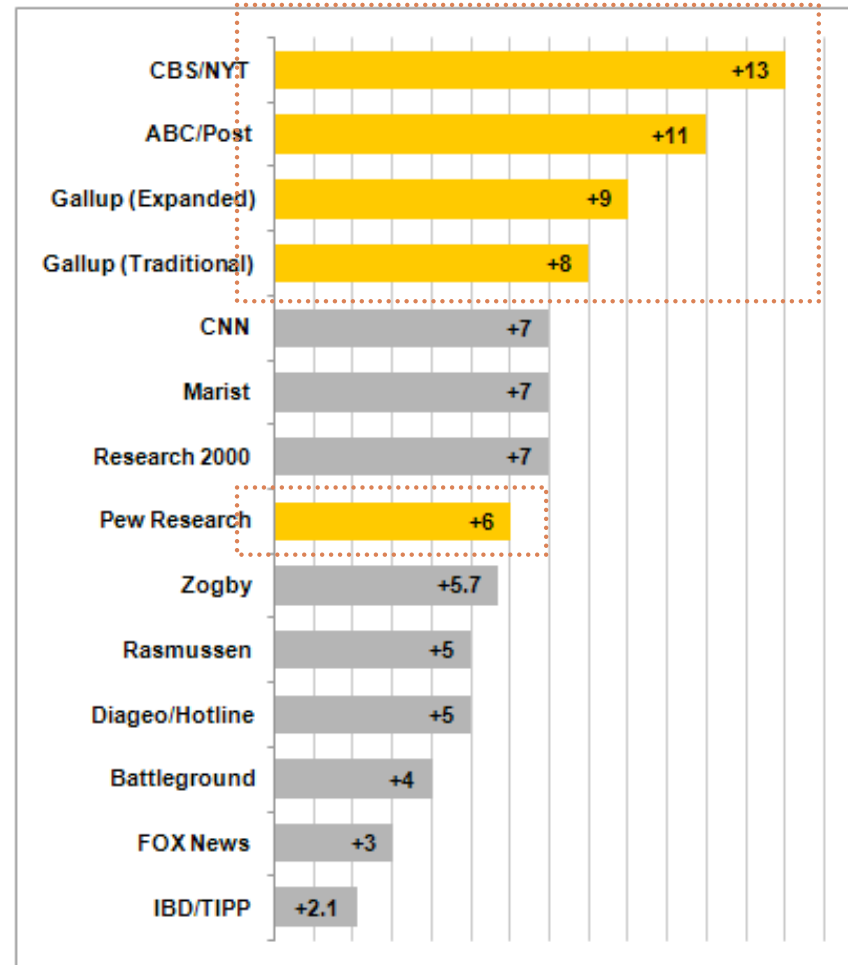Needs and Usability Assessment 11.04.08

Analysis

Looking for **significant** differences across **important** groups

Looking for **correlation** – inferring causality

Polls including mobile phones

Poll	Value
CBS/NYT	+13
ABC/Post	+11
Gallup (Expanded)	+9
Gallup (Traditional)	+8
CNN	+7
Marist	+7
Research 2000	+7
Pew Research	+6
Zogby	+5.7
Rasmussen	+5
Diageo/Hotline	+5
Battleground	+4
FOX News	+3
IBD/TIPP	+2.1

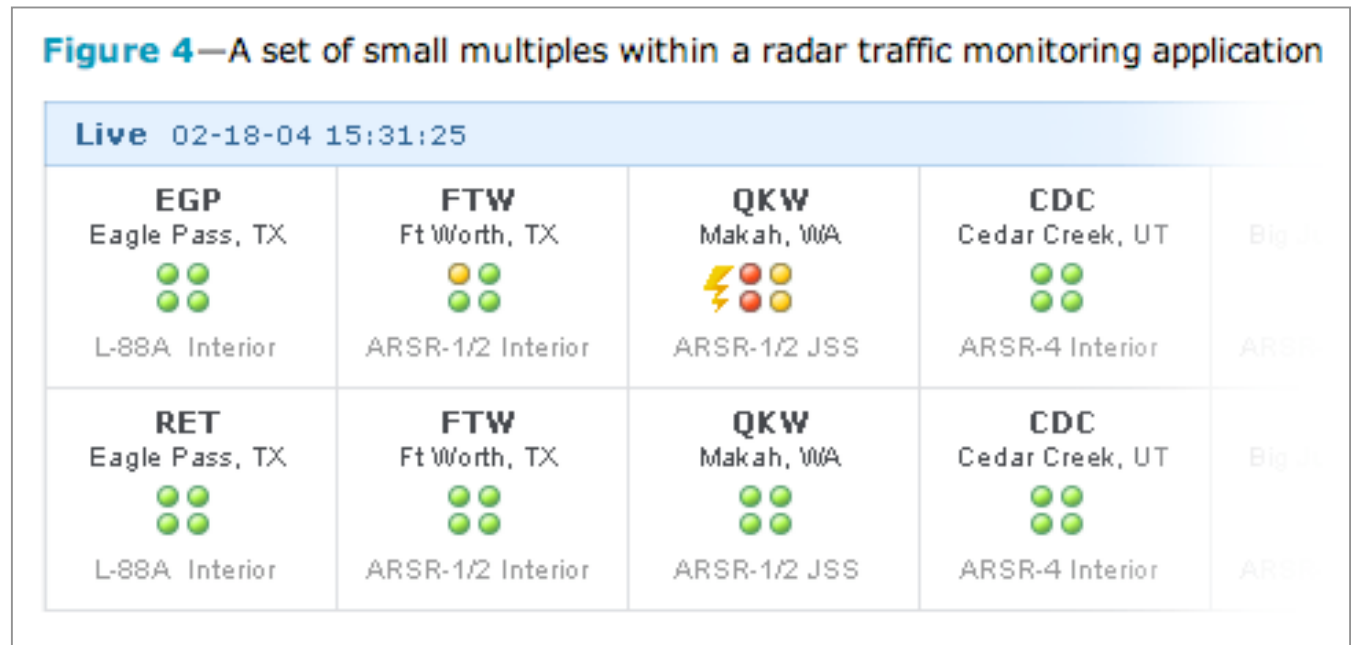# Maps

Placing data on physical space
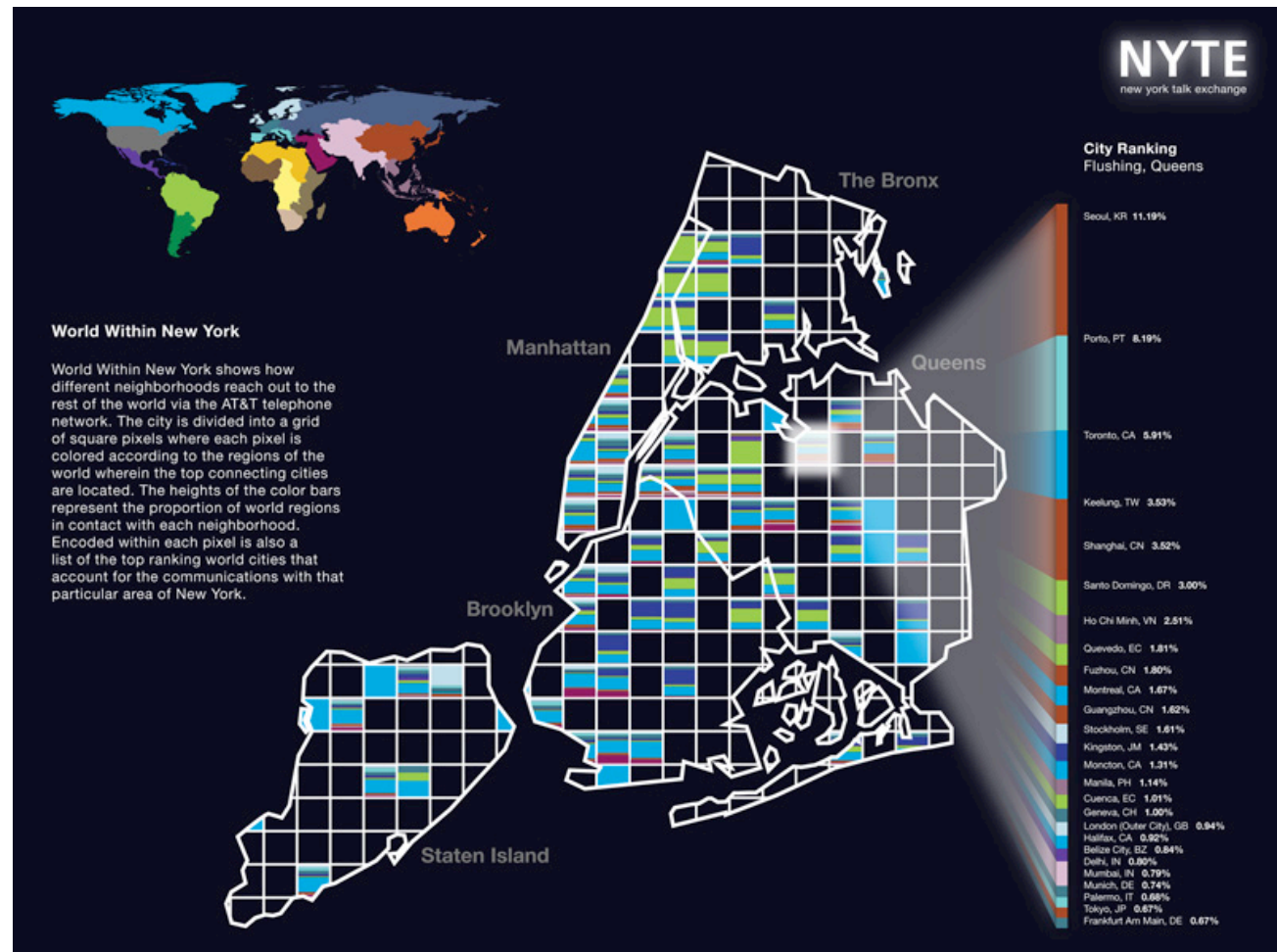
Representing change with small multiples

Small multiples

"Graphical depictions of variable information that share context, but not content."
Edward Tufte

Figure 4—A set of small multiples within a radar traffic monitoring application

Tufte, E. *Envisioning Information*. Cheshire, CT: Graphics Press (1990).

graphic: www.uxmatters.com/MT/archives/000044.php

Reporting guidelines

The reader should be able to reconstruct what you did

Show the question

Include *all* data

ie, percentages *and* numbers of respondents

Table 1: Opinion on food

Are you satisfied or disssatisfied with the bagels at the Free Speech Movement Café?

	Undergrads % n=120	**Grads** % n=200	**Total** % n=320
Satisfied	60 *N=71*	13 *N=25*	31 *N=96*
Dissatisfied	40 *N=47*	87 *N=165*	69 *N=127*
Total	100 *n = 118*	100 *n = 190*	100 *n = 308*
No ans.	*n = 2*	*n = 10*	*n= 12*

UC Berkeley School of Information
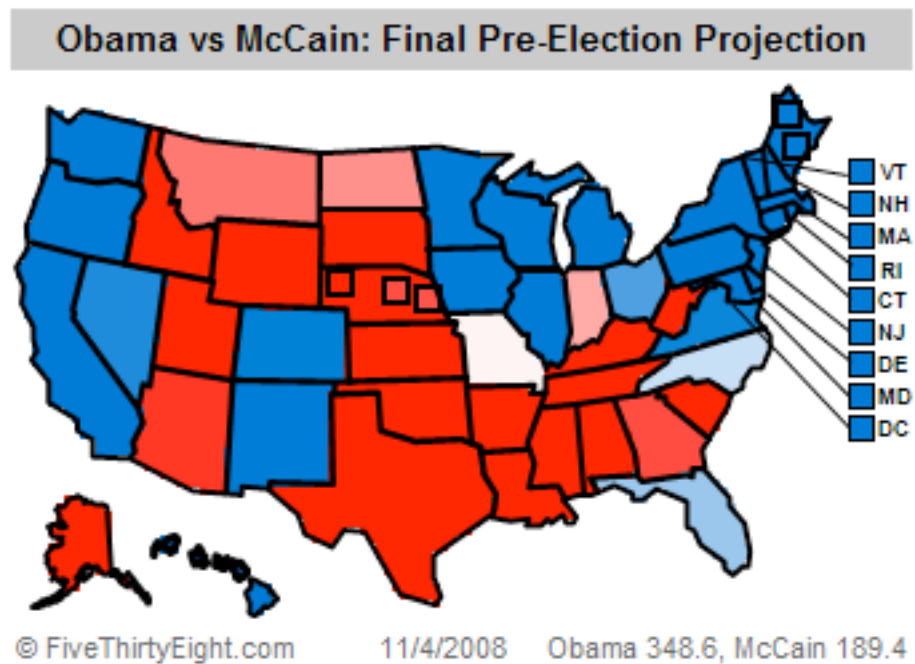
Needs and Usability Assessment 11.04.08

Reporting: last words

Graphics are visual metaphors.

They are rhetoric.

Do you know the story you want to tell?

...the story your audience is seeing?

Obama vs McCain: Final Pre-Election Projection

VT
NH
MA
RI
CT
NJ
DE
MD
DC

© FiveThirtyEight.com   11/4/2008   Obama 348.6, McCain 189.4

Appendix A:
Good places to look for example reports

Pew Internet

http://www.pewinternet.org/pdfs/PIP_Searchengine_users.pdf pp. i-iv

http://www.pewinternet.org/pdfs/PIP_Future_of_Internet_2006.pdf

http://www.pewinternet.org/pdfs/PIP_Search_Questions.pdf

"What Do Americans Do on the Internet?" Appendix B

Harris Interactive Poll

http://www.harrisinteractive.com/harris_poll/

Stanford Time Study

http://www.stanford.edu/group/siqss/SIQSS_Time_Study_04.pdf