

Graph Formats

Frederic Malter, PhD

University of Arizona
November 11, 2009

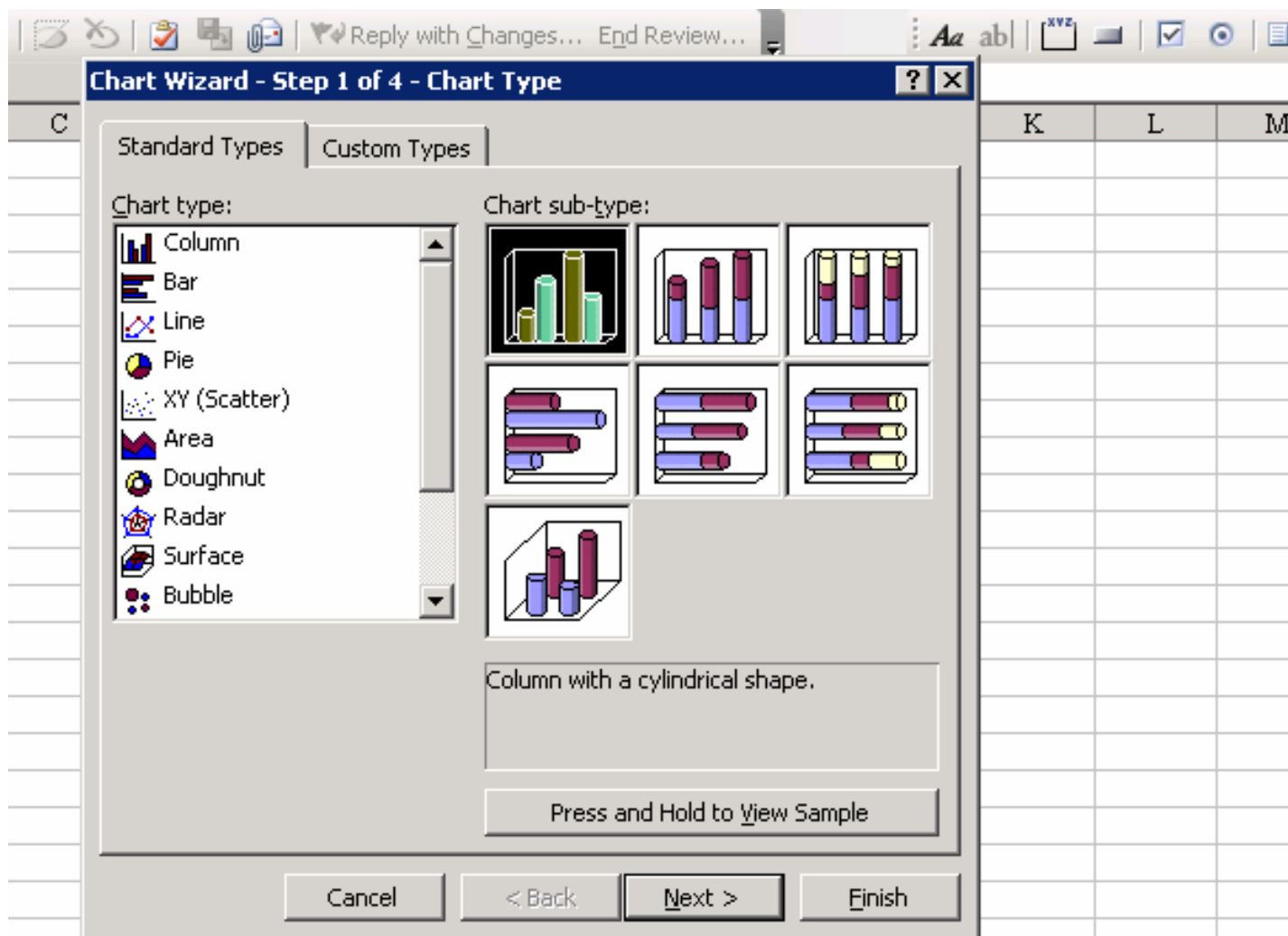
EVALUATION, RESEARCH AND DEVELOPMENT UNIT
THE UNIVERSITY OF ARIZONA

Funded by the Arizona Department of Health Services

Overview

- Pie chart
- Bar chart
- Line chart & „bumps“ chart
- Error bars
- Area chart
- Dot plot
- Bubble chart
- Scatterplot
- Small multiples
- Natural frequency graphs
- Consumer report-style graph tables
- Maps/spatial data

MS Excel graph menue

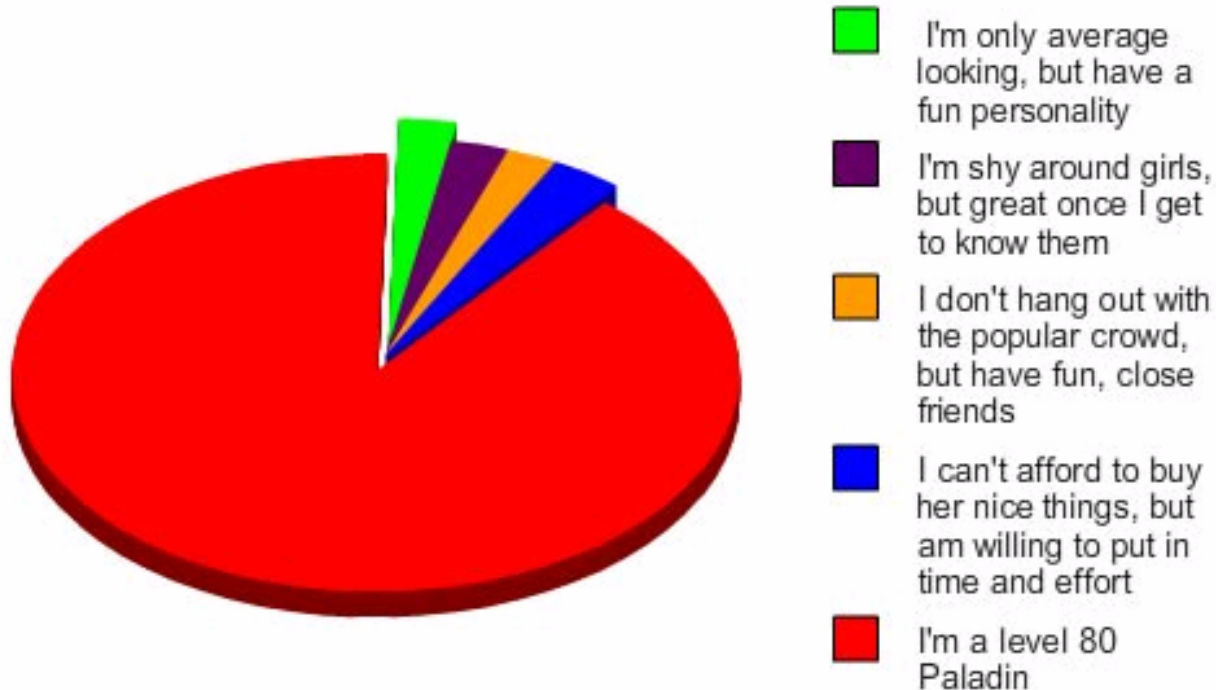


Pie Chart - Report Card

- Name: Pie Chart
- Purpose: part-of-whole (e.g. percentages adding up to 100)
- Application: percentages, comparison of distributions
- Grades
 - Data density: **F**
 - Popularity: **A**

Pie chart

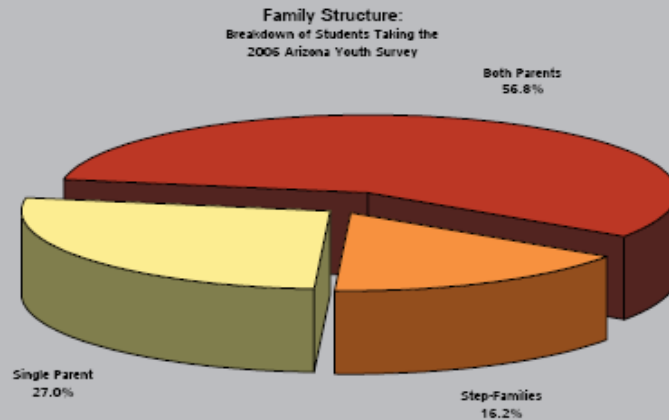
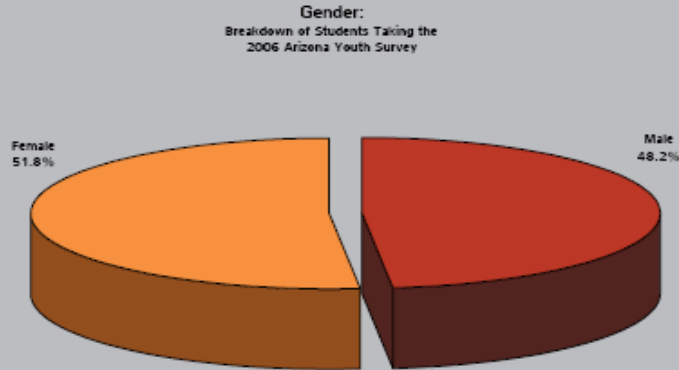
Why I Don't Have a Girlfriend



GraphJam.com

Pie Chart: Data-ink minimization

Figures 5 and 6



Let's count!

- 1) Pie charts
- 2) 3-D
- 3) Colors don't mean the same: not even a comparison
- 4) Numbers themselves, font size = 2pts.
- 5) Grey background
- 6) Empty spaces

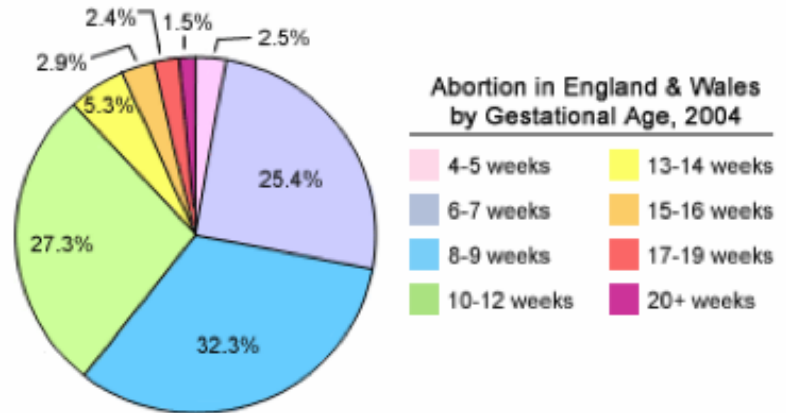
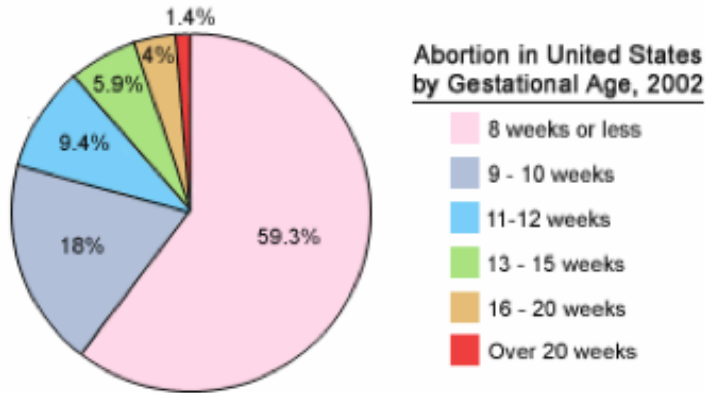
The best alternative: No graph at all!

No
decimals!

Gender	male	52%
	female	48%
Family structure	Both parents	57%
	Single parent	27%
	Step parents	16%

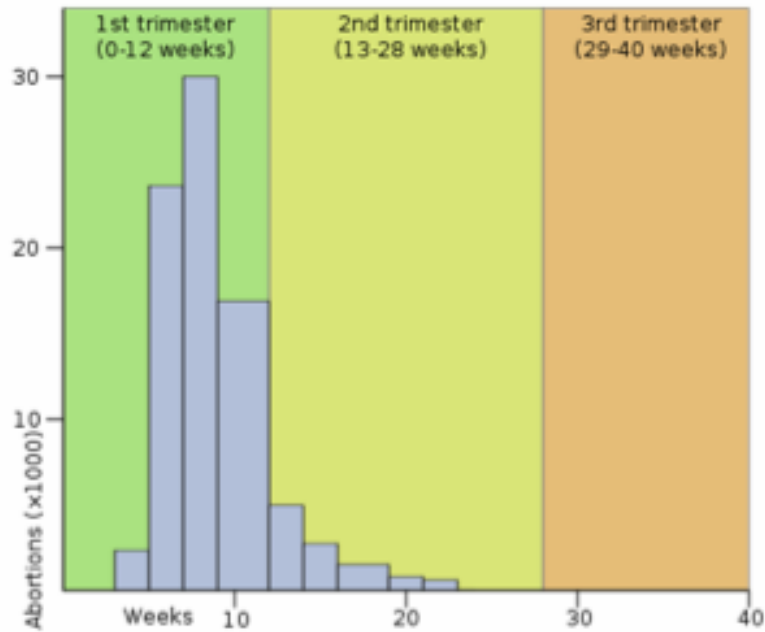
Sort according to size!

Pie Chart: Inadequate for comparing

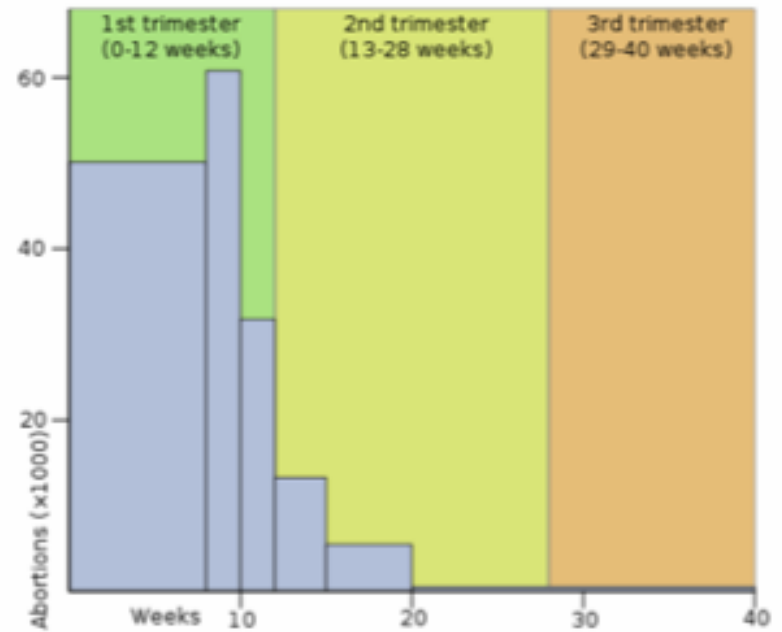


But there's help!

England & Wales, 2004



United States, 2002



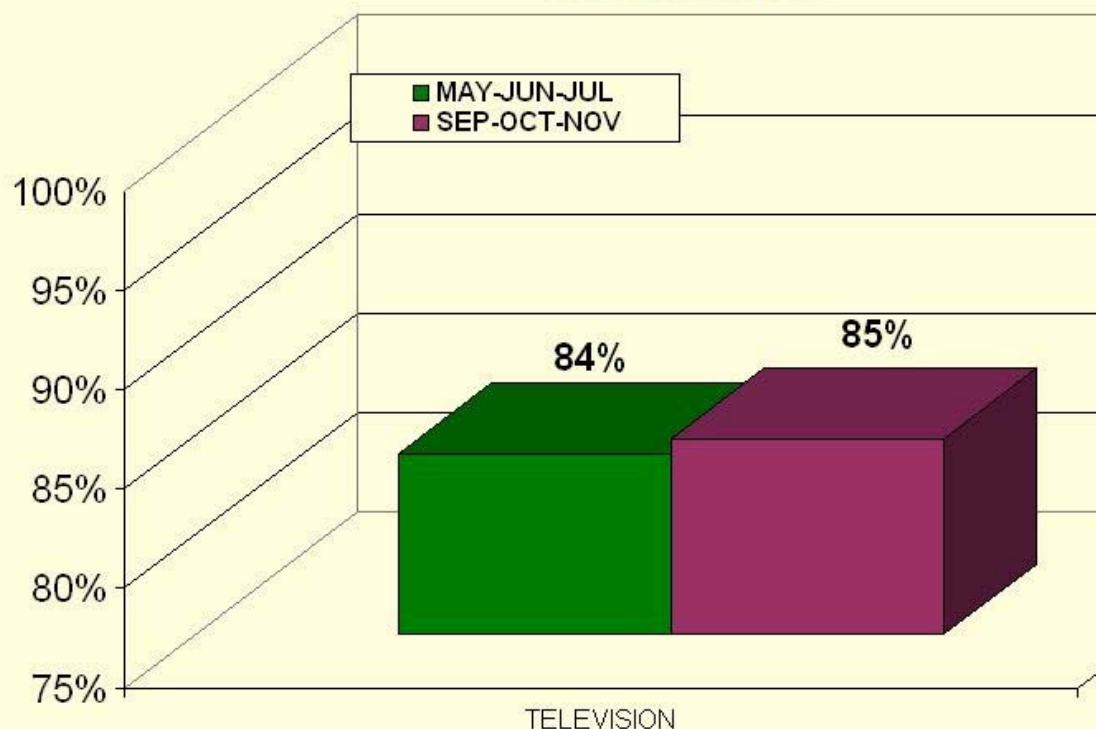
Bar Chart - Report Card

- Name: Bar Chart
- Purpose: discrete distribution (e.g. percentages), continuous histograms, comparing distributions
- Application: discrete distributions (percentages), comparison of distributions, covariation, time series, pretty much anything....
- Grades
 - Data density: **B-F (depends...)**
 - Popularity: **A**

Bar chart: example of nonsense

Where did you see or hear that advertising for the
Smokers Helpline?

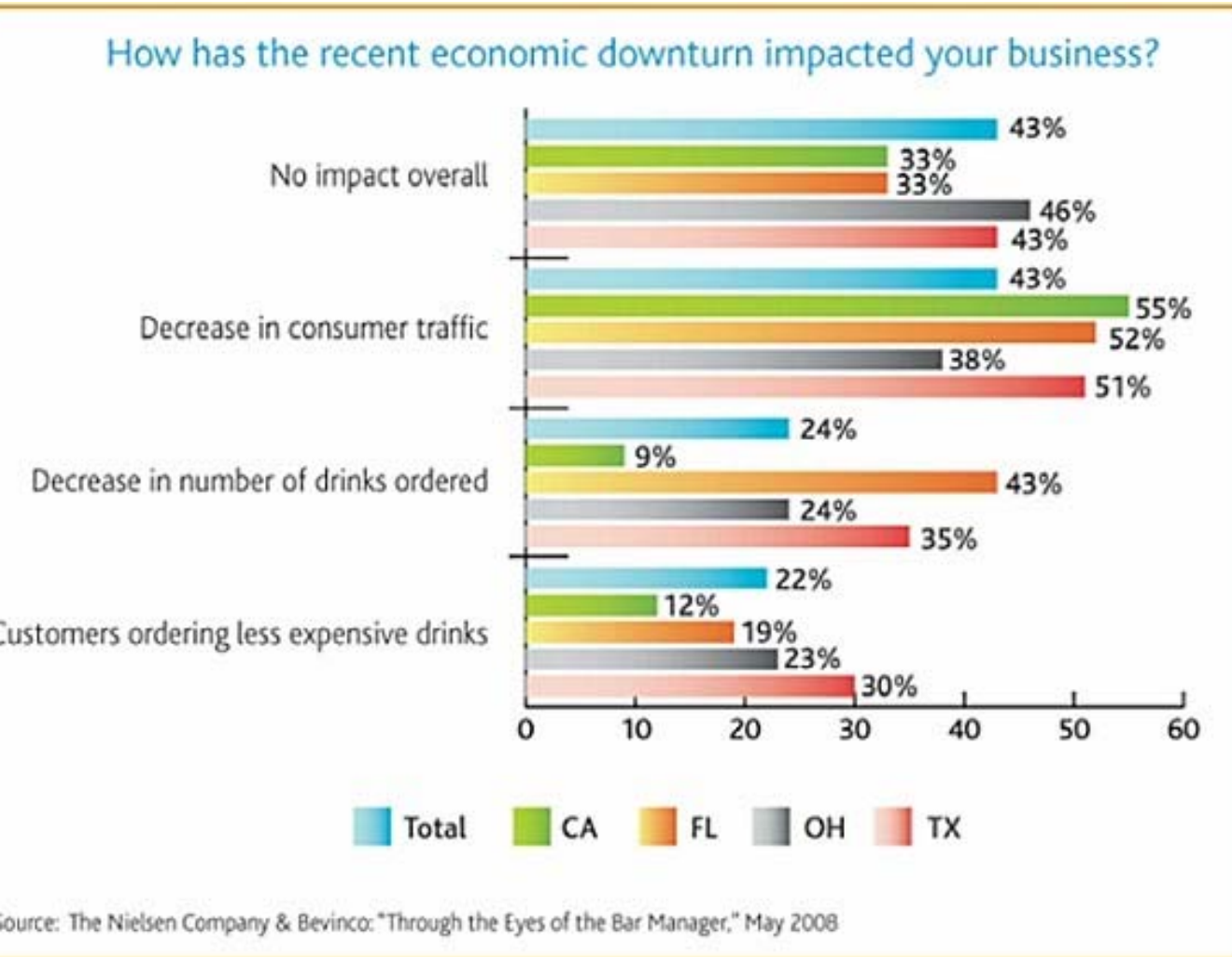
TELEVISION



Let's count!

- 1) Density: 2 numbers, no difference
- 2) Y-axis starts at 75%
- 3) 3-D
- 4) Gridlines
- 5) Aggregation?
- 6) Heading

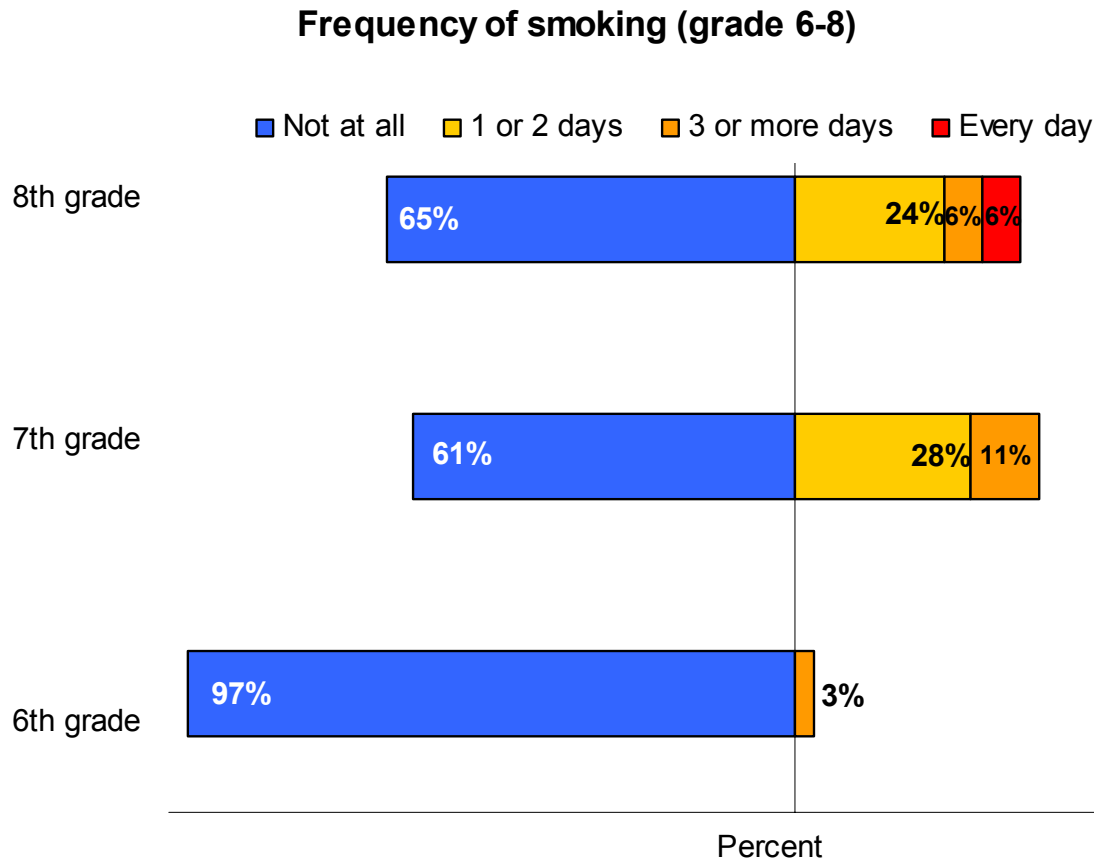
Bar chart: Loss aversion



Let's count!

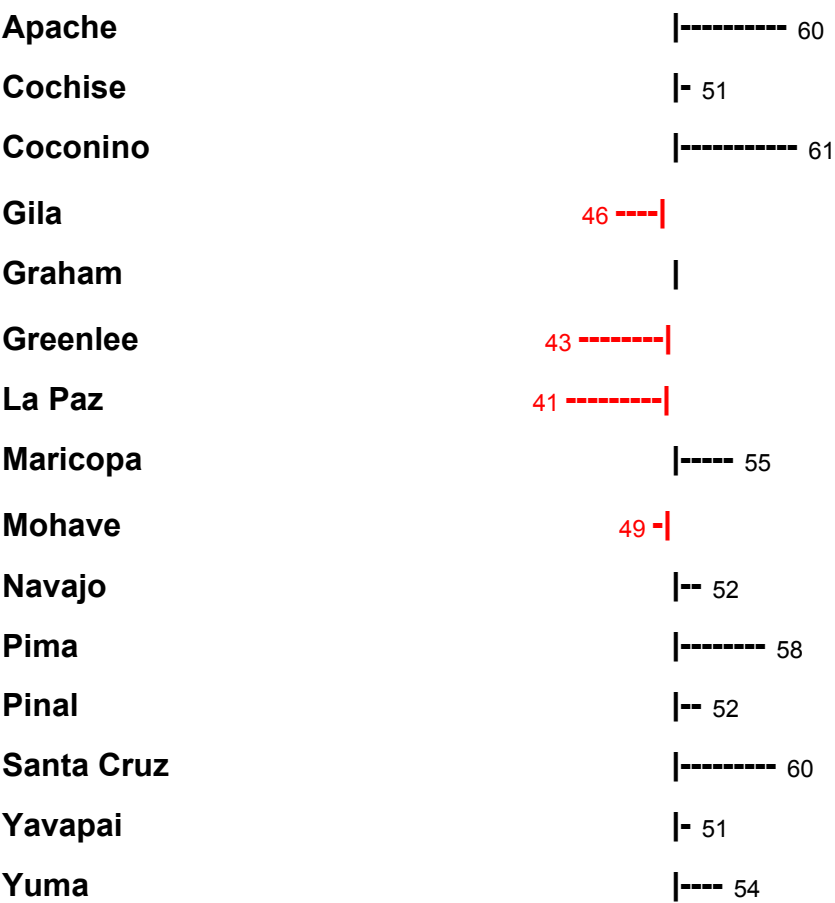
- 1) Message?
- 2) Sensible comparisons?
- 3) Color schemes
- 4) Shading
- 5) X-axis label

Alternatives to conventional bar charts: double-sided

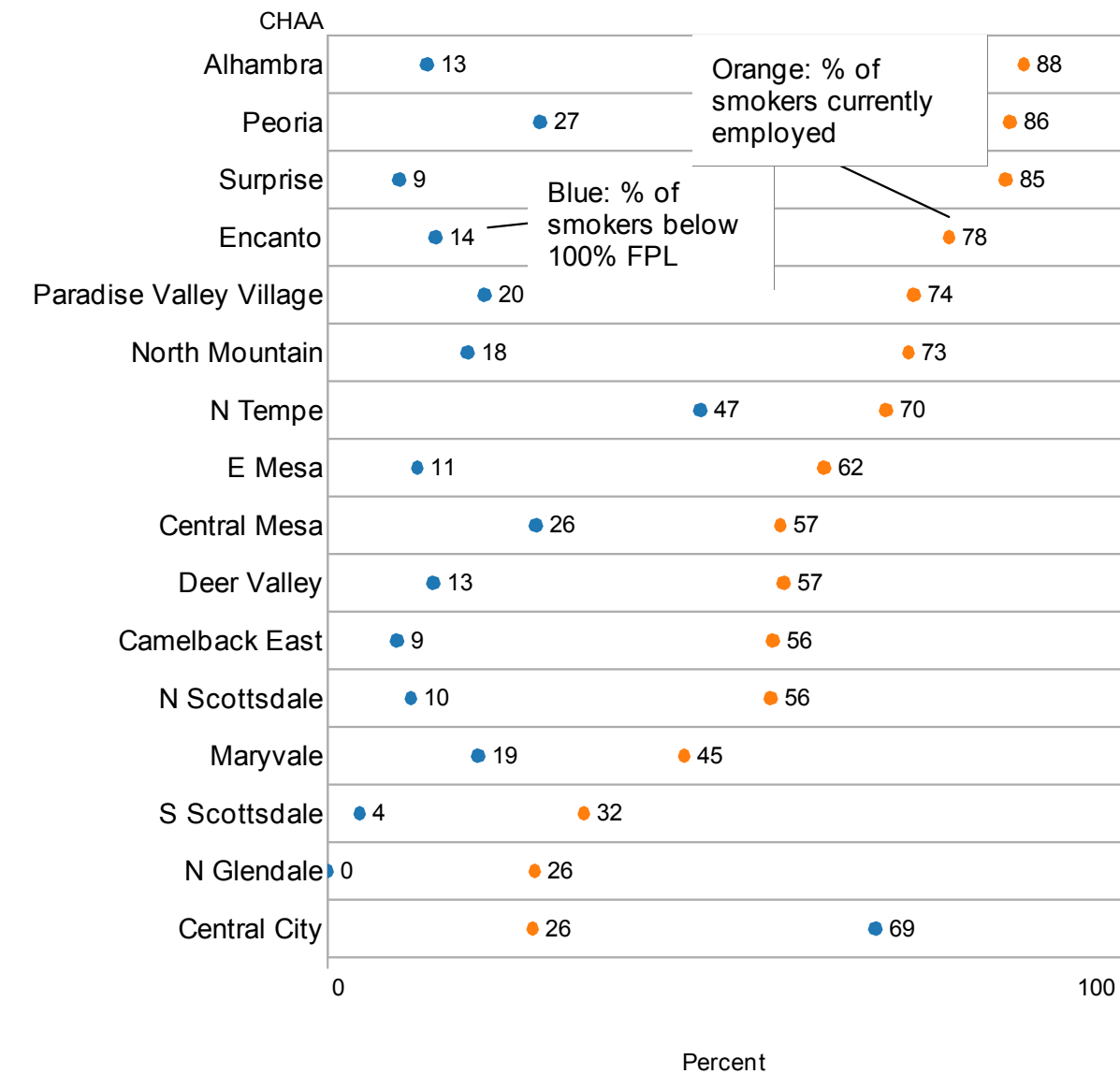


Deviation bars

Smoking Ban (Prop. 201): Deviation from 50%, by County



Dot plot: ranked by size of orange dots



Dot plot with intelligent labeling

Groups

\hat{F} : Female participants with children by mid-30s

\hat{M} : Male participants with children by mid-30s

F: Female participants without children by mid-30s

M: Male participants without children by mid-30s

Working no more than 60 hours per week

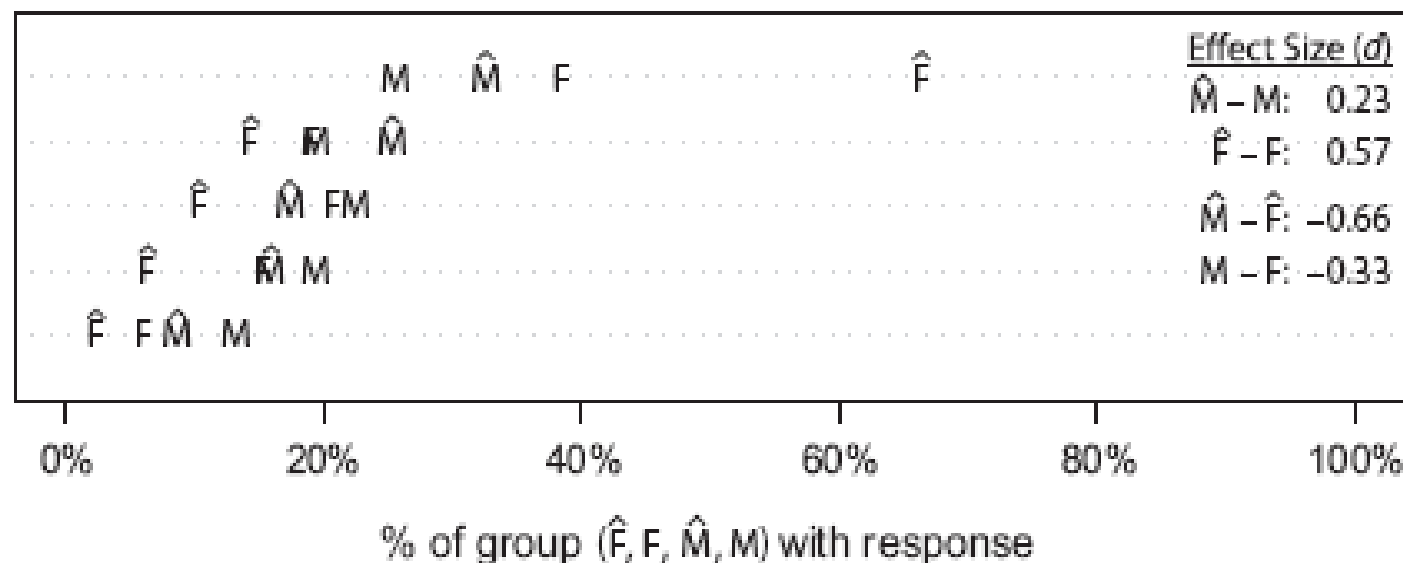


Tableau | Excel

X |

Trellis plot: dot plot's big brother

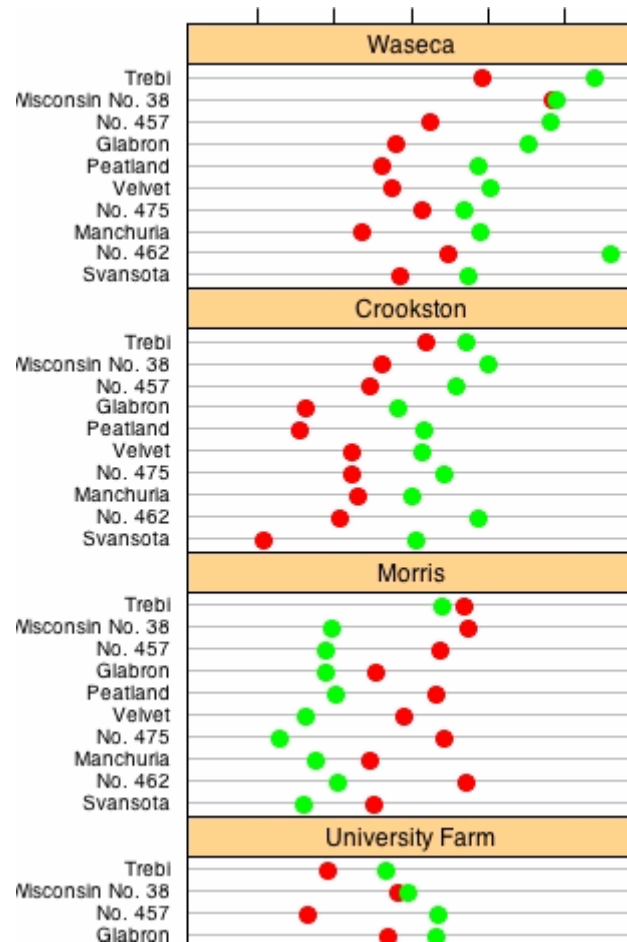
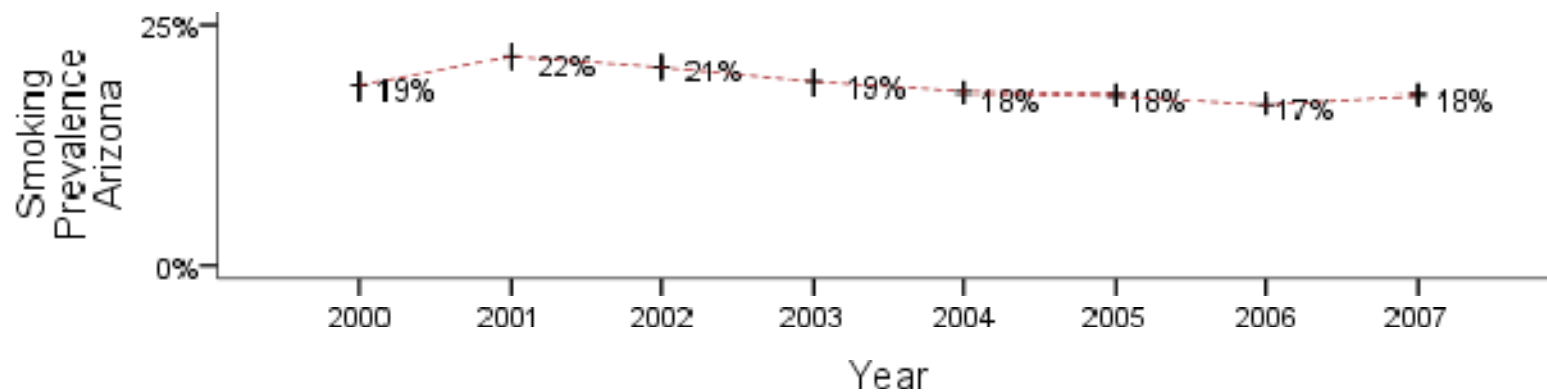
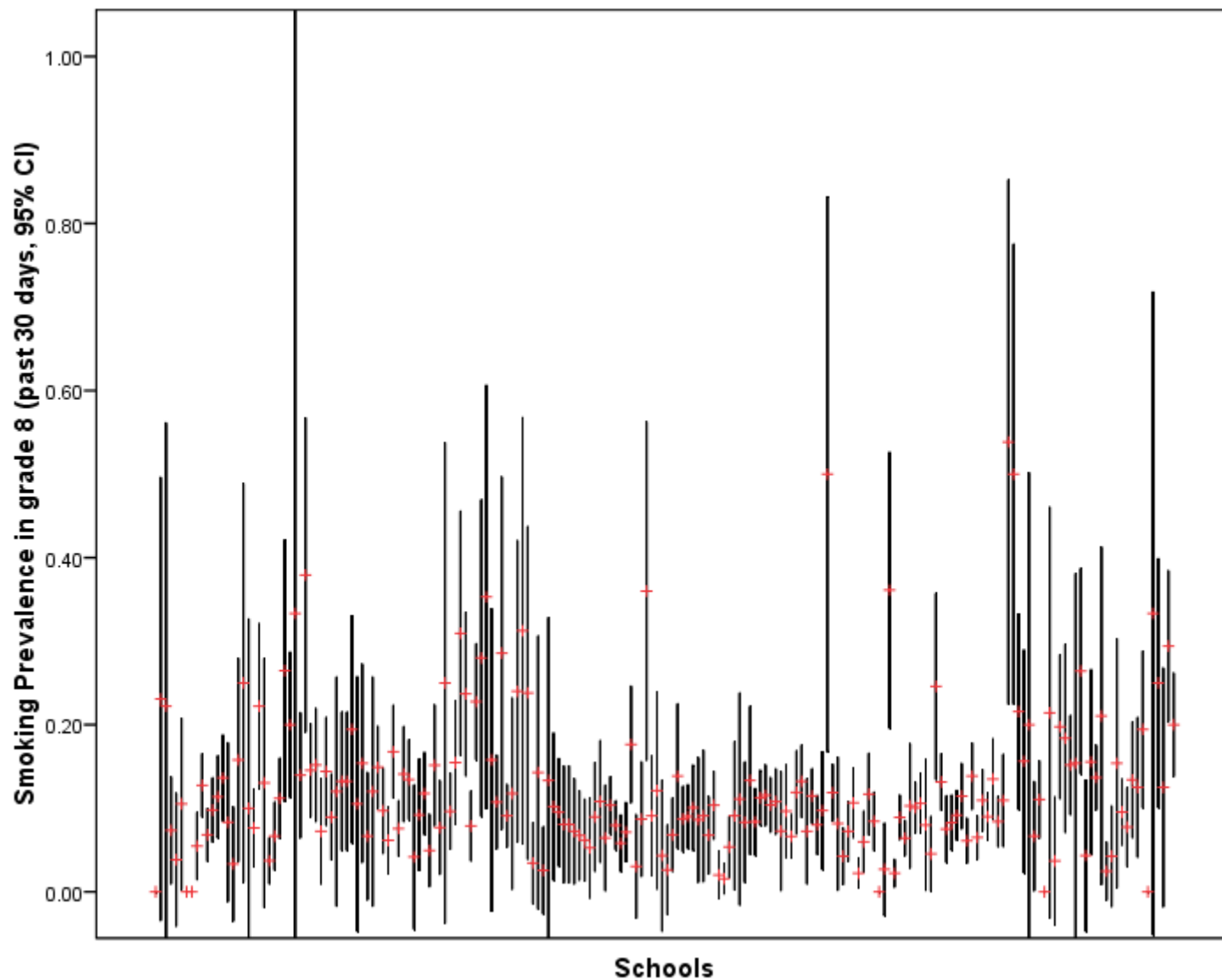


Tableau | Excel
X |

Error bars



Smoking prevalence for 200 schools with 95% CI (this one not so great for communication)



S-PLUS

Line Chart - Report Card

- Name: Line Chart
- Purpose: time series, comparing distributions, comparing profiles (discrete distributions)
- Application: same as above
- Grades
 - Data density: **A**
 - Popularity: **A**

Line chart: Less is often better

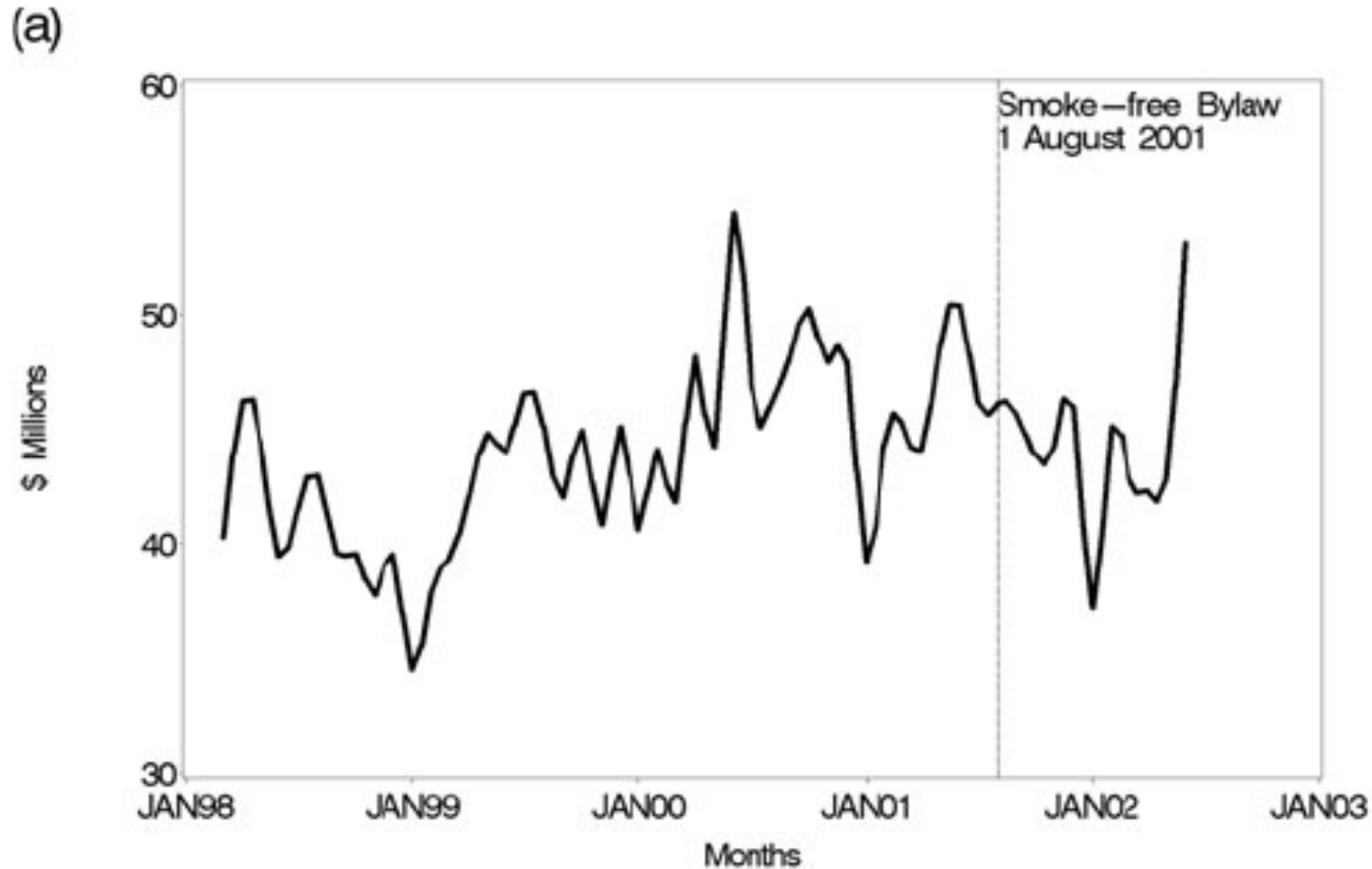
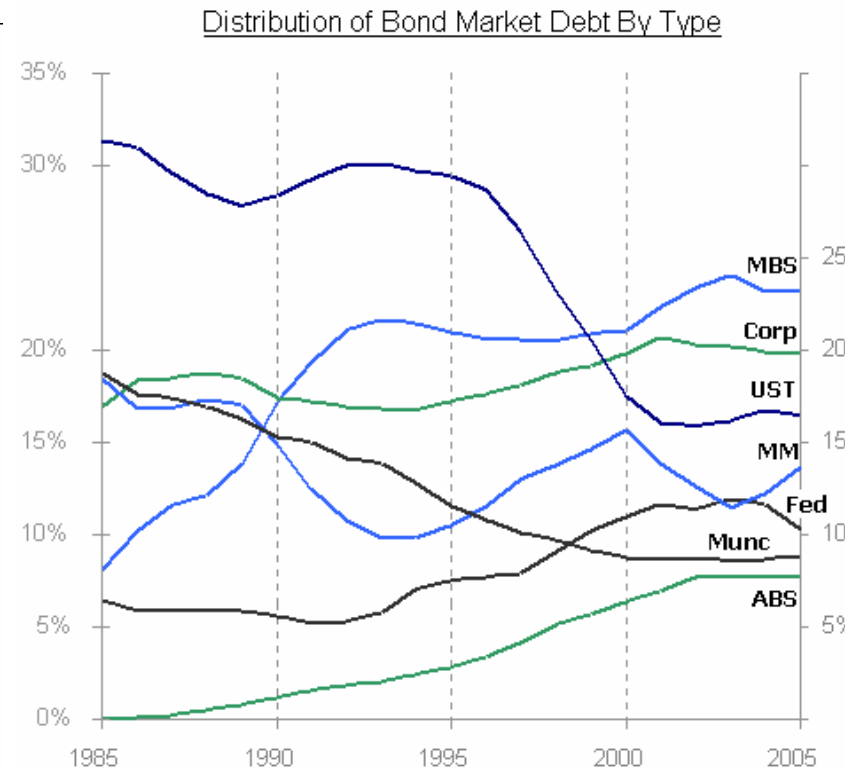
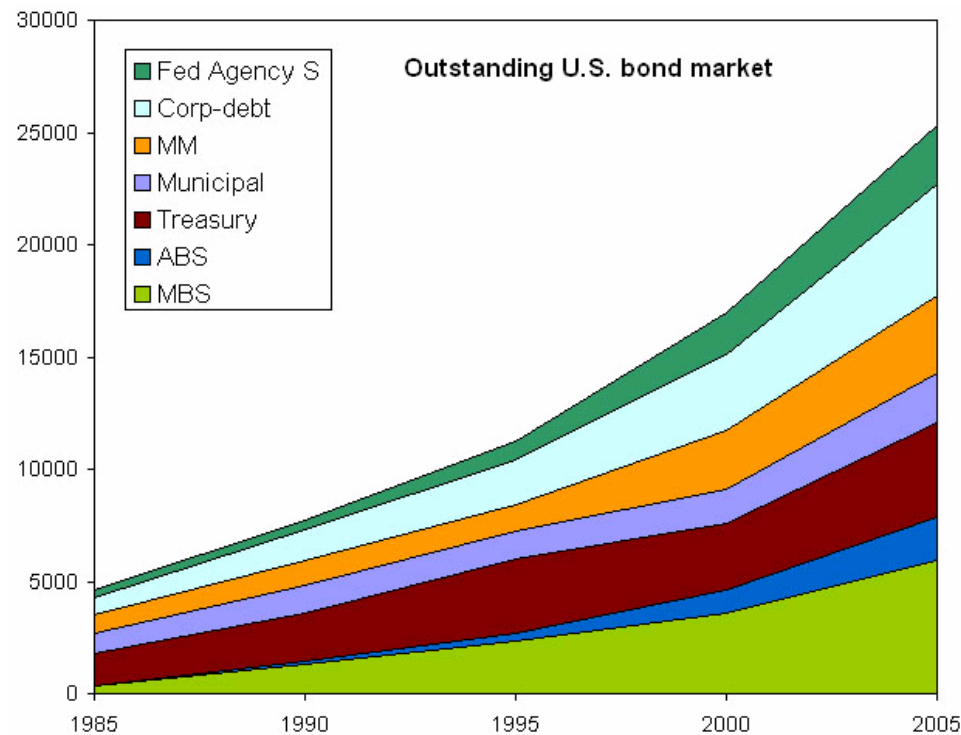


Tableau | Excel
X | X

Stacked area chart - a bad choice



“Sparkliness” (NY Times)

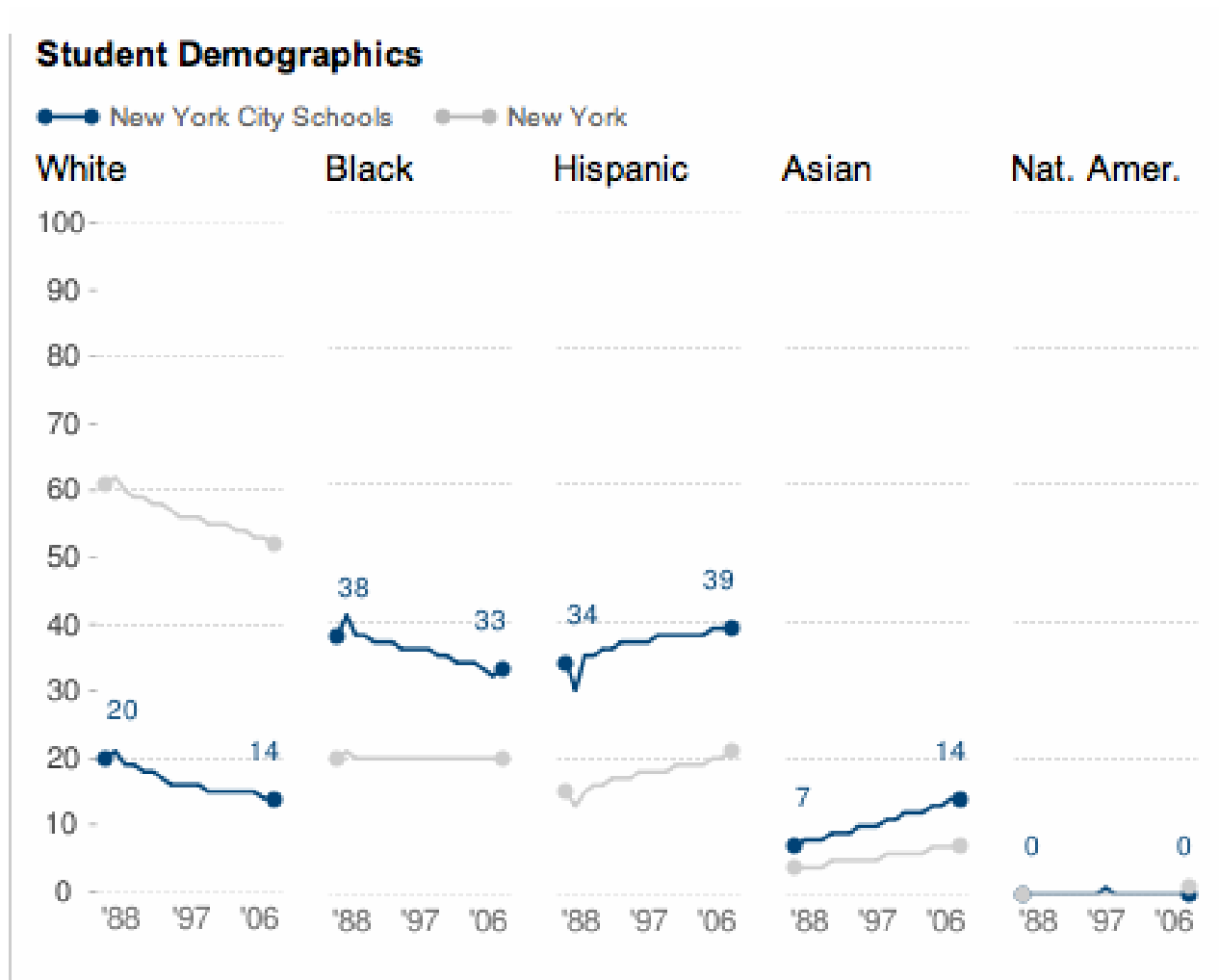


Tableau | Excel
X |

“Bumps” Chart

- Special case of line chart
- Purpose: change over time in ranking
- Grades
 - Data density: **A**
 - Popularity: **F**

The New York Times

April 6, 2009



Scatter plot - Report Card

- Name: Scatterplot
- Purpose: Joint distribution of two continuous variables, permutation of joint distribution by categorical third variable
- Application: same
- Grades
 - Data density: **A**
 - Popularity: **A**

Scatter plot with intelligent labels

Red: Republican

Blue: Democrats

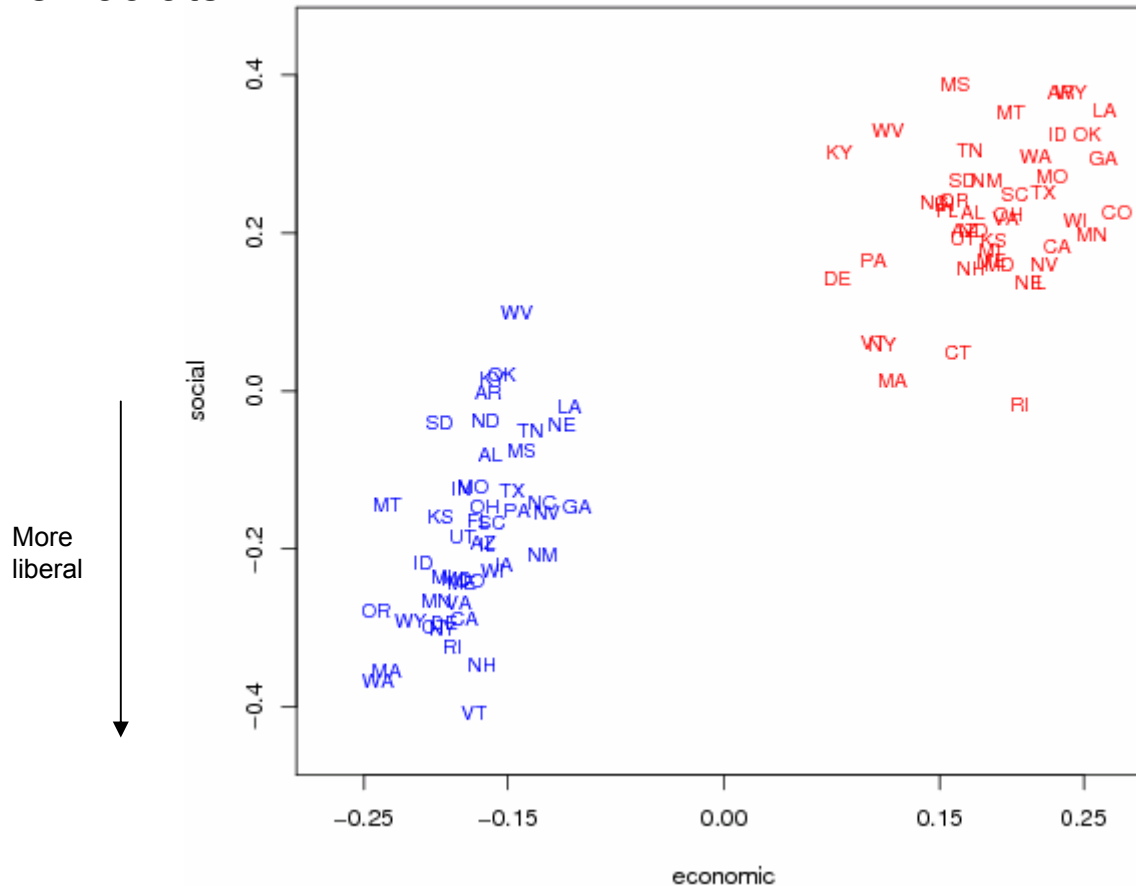
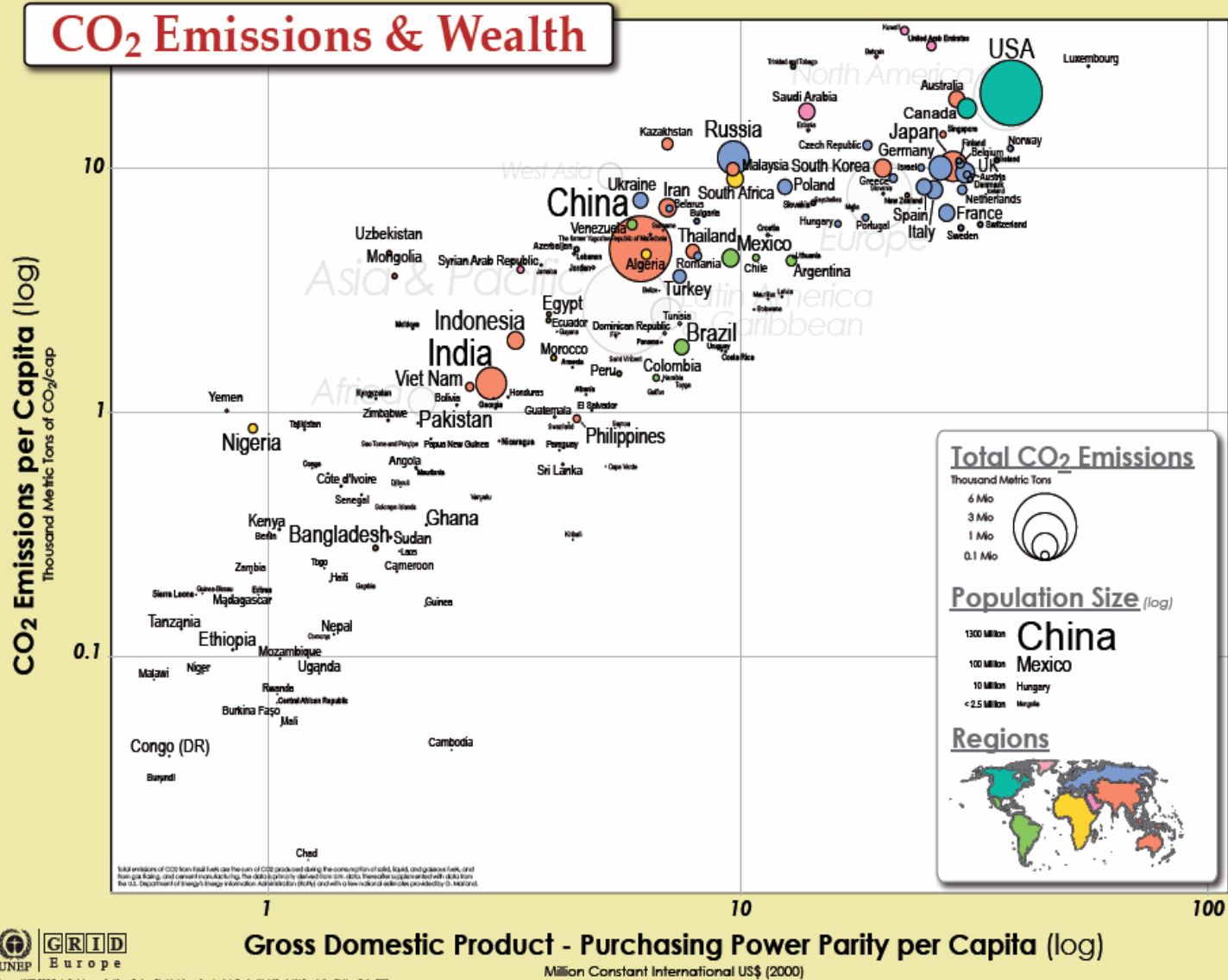


Tableau | Excel

X |

Scatter plot with 5 (!) dimensions

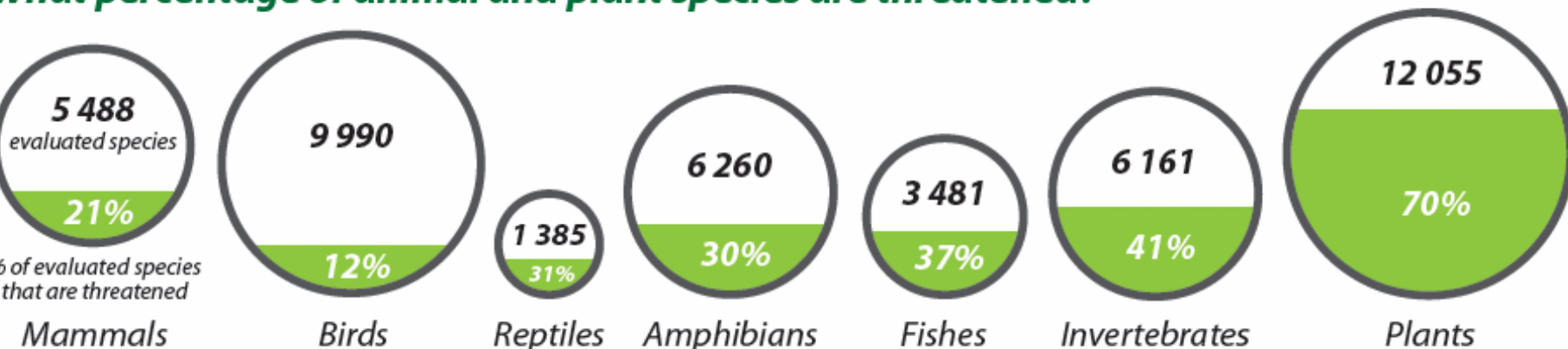


Bubble Charts - Report Card

- Name: Bubble charts
- Purpose: magnitudes for categories, comparing categories on indicator magnitude, percentages
- Application: same
- Grades
 - Data density: **C-F**
 - Popularity: **B**

Bubble chart

What percentage of animal and plant species are threatened?



6 of evaluated species
that are threatened

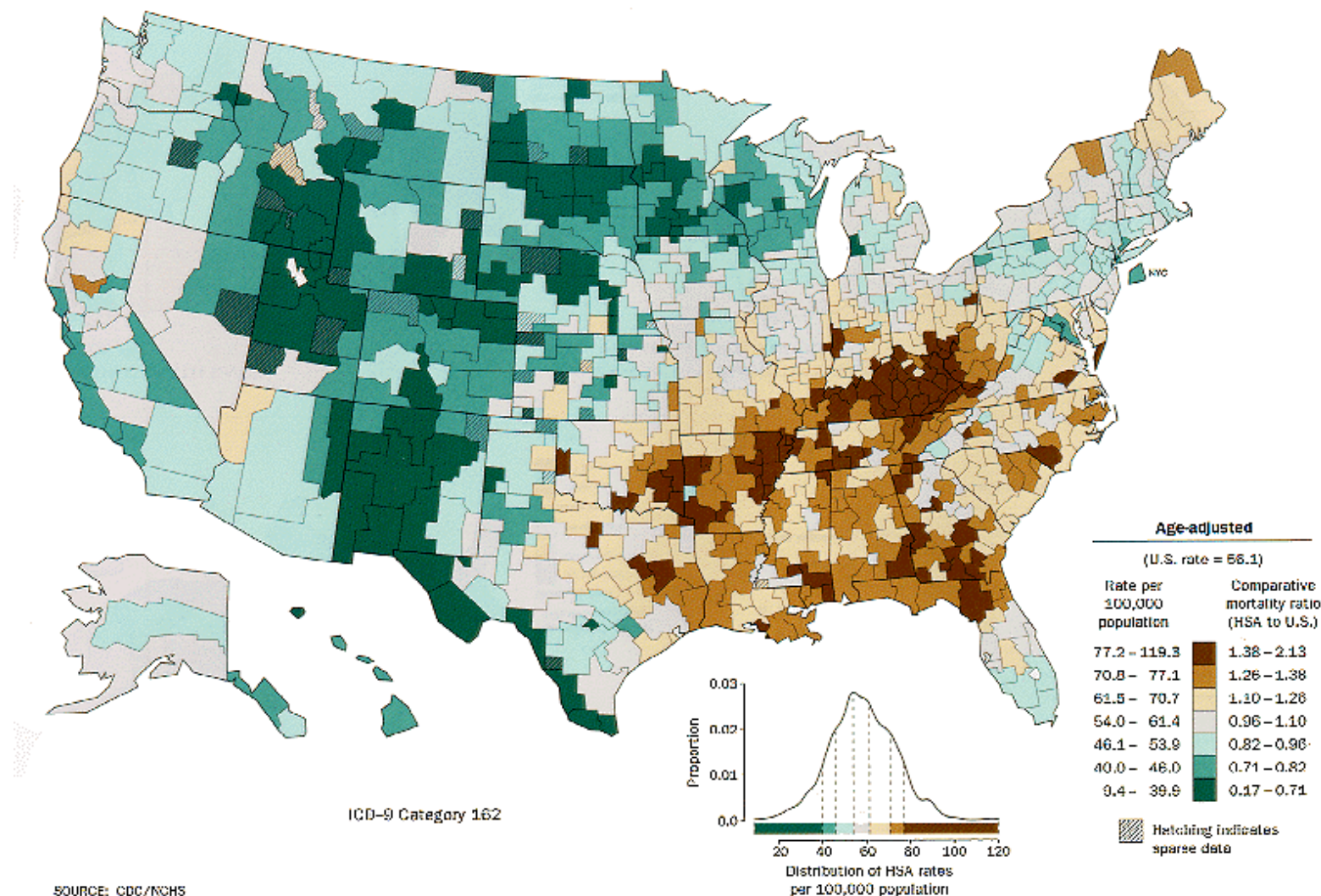
Source: UNEP GEO Data Portal, compiled from IUCN/SSC; IUCN 2008. The IUCN Red List of Threatened Species

Data maps - Report Card

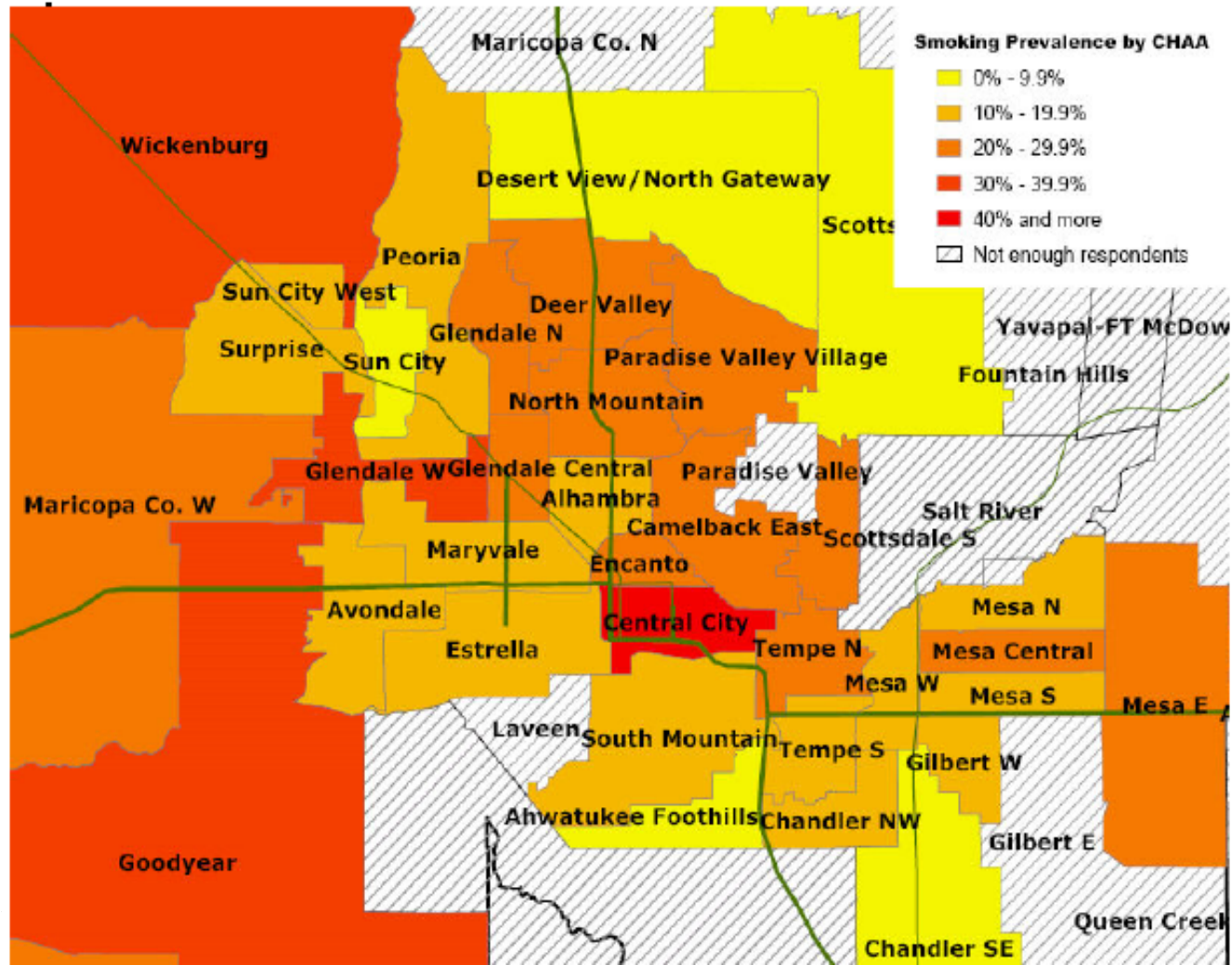
- Name: Data maps
- Purpose: continuous variables with spatial distribution
- Application: magnitude of indicator color-coded („cloropleth“) for a geographical point, line or area
 - Many hybrids: bubble maps, bar maps, pie-chart maps (→usually much worse than cloropleth map)
- Grades
 - Data density: **A-C**
 - Popularity: **A**

Graphical excellence I:

Data maps (Lung cancer rates by county, CDC 2004)



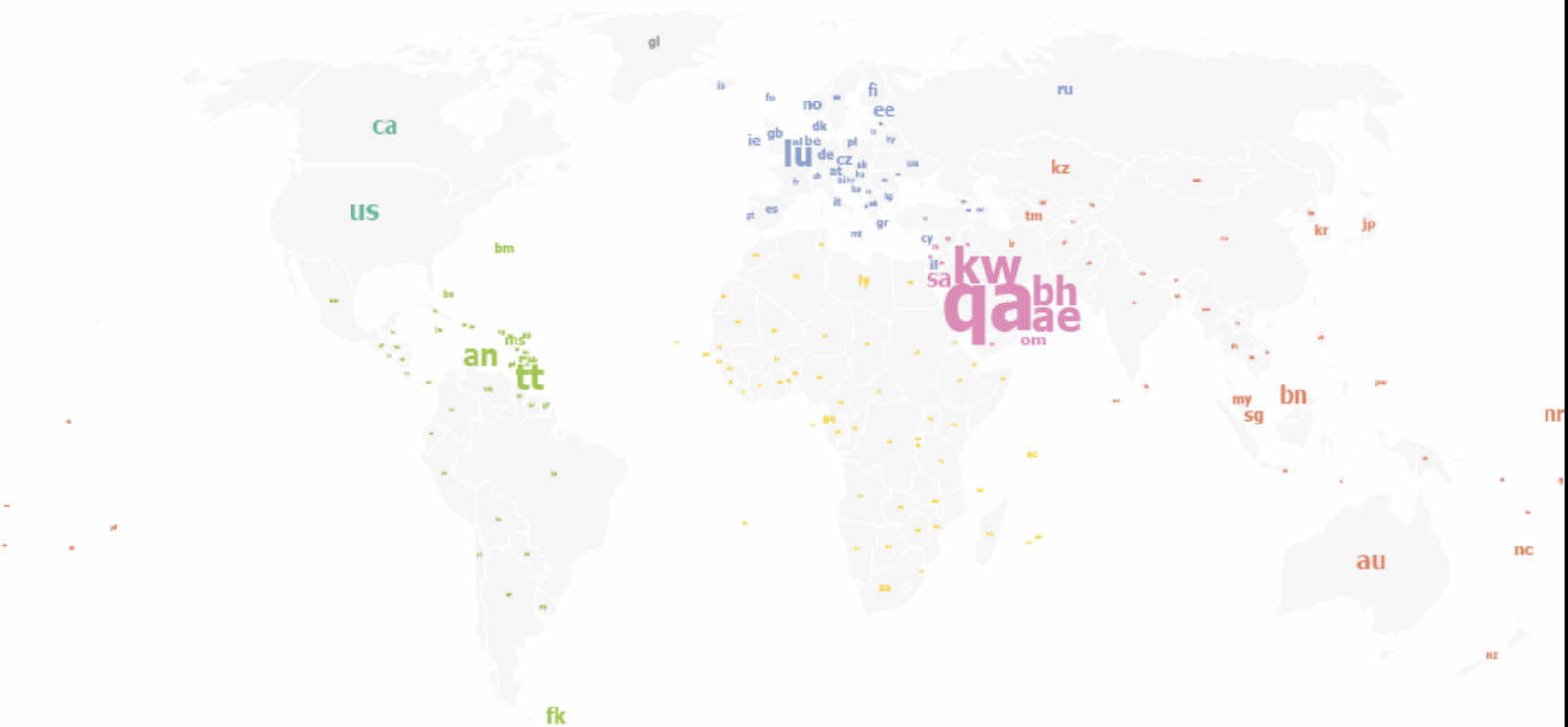
My first choropleth map!



“Map wordle”

Per Capita CO₂ Emissions

Total emissions of CO₂ from fossil fuels are the sum of CO₂ produced during the consumption of solid, liquid, and gaseous fuels, and from gas flaring, and cement manufacturing. The data is primarily derived from U.N. data. Thereafter supplemented with data from the U.S. Department of Energy's Energy Information Administration (EIA) and with a few national estimates provided by G. Marland.



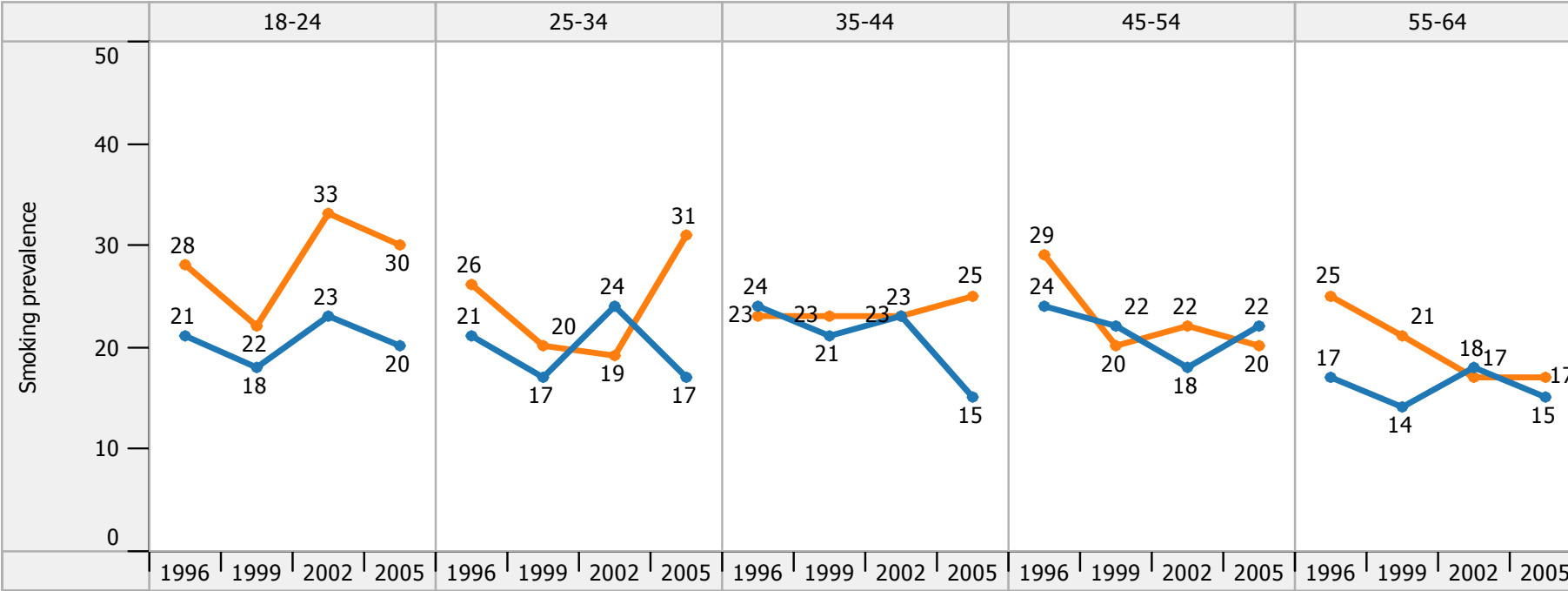
ISO Country	Population (Million)	Per Capita CO ₂ Emissions (kg)	ISO Country	Population (Million)	Per Capita CO ₂ Emissions (kg)	ISO Country	Population (Million)	Per Capita CO ₂ Emissions (kg)	ISO Country	Population (Million)	Per Capita CO ₂ Emissions (kg)
AFRICA			US	310	18.5	RU	143	12.5	EU	730	11.5
Algeria	34.0	0.8	Canada	34.0	18.5	Russia	143	12.5	France	65.0	11.5
Angola	16.0	0.8	Chad	16.0	0.8	Ukraine	46.0	12.5	Germany	82.0	11.5
Benin	18.0	0.8	Comoros	0.8	0.8	Belarus	9.5	12.5	Italy	60.0	11.5
Burkina Faso	18.0	0.8	Cote d'Ivoire	18.0	0.8	Poland	38.0	12.5	Spain	45.0	11.5
Burundi	10.0	0.8	Egypt	80.0	0.8	Czechia	11.0	12.5	Sweden	9.5	11.5
Cameroon	22.0	0.8	Ethiopia	100.0	0.8	Slovakia	5.5	12.5	Switzerland	3.5	11.5
Cape Verde	0.5	0.8	Ghana	22.0	0.8	Slovenia	2.1	12.5	Norway	4.5	11.5
Central African Rep.	4.5	0.8	Guinea	12.0	0.8	Croatia	4.5	12.5	Netherlands	16.0	11.5
Chad	16.0	0.8	Guinea-Bissau	1.5	0.8	Cyprus	0.8	12.5	Belgium	10.0	11.5
Cote d'Ivoire	18.0	0.8	Kenya	38.0	0.8	Denmark	5.5	12.5	Austria	8.5	11.5
DRC	75.0	0.8	Lesotho	2.5	0.8	Finland	5.5	12.5	Luxembourg	0.5	11.5
Egypt	80.0	0.8	Liberia	4.5	0.8	France	65.0	11.5	Ireland	0.4	11.5
Ethiopia	100.0	0.8	Madagascar	22.0	0.8	Germany	82.0	11.5	Malta	0.4	11.5
Ghana	22.0	0.8	Mali	18.0	0.8	Italy	60.0	11.5	Monaco	0.03	11.5
Guinea	12.0	0.8	Morocco	34.0	0.8	Japan	125.0	11.5	San Marino	0.03	11.5
Guinea-Bissau	1.5	0.8	Mozambique	22.0	0.8	South Korea	48.0	11.5	South Korea	48.0	11.5
Kenya	38.0	0.8	Niger	18.0	0.8	Taiwan	23.0	11.5	Taiwan	23.0	11.5
Lesotho	2.5	0.8	Rwanda	12.0	0.8	Thailand	65.0	11.5	Thailand	65.0	11.5
Liberia	4.5	0.8	Senegal	14.0	0.8	Vietnam	80.0	11.5	Vietnam	80.0	11.5
Madagascar	22.0	0.8	Sierra Leone	6.0	0.8	China	1400.0	11.5	China	1400.0	11.5
Mali	18.0	0.8	South Africa	55.0	0.8	India	1300.0	11.5	India	1300.0	11.5
Morocco	34.0	0.8	South Sudan	11.0	0.8	USA	310.0	11.5	USA	310.0	11.5
Mozambique	22.0	0.8	Tanzania	55.0	0.8	Canada	34.0	11.5	Canada	34.0	11.5
Niger	18.0	0.8	Togo	7.5	0.8	Australia	23.0	11.5	Australia	23.0	11.5
Nigeria	180.0	0.8	Tunisia	11.0	0.8	Brazil	200.0	11.5	Brazil	200.0	11.5
Rwanda	12.0	0.8	Uganda	28.0	0.8	Argentina	40.0	11.5	Argentina	40.0	11.5
Senegal	14.0	0.8	Zambia	11.0	0.8	Chile	18.0	11.5	Chile	18.0	11.5
Sierra Leone	6.0	0.8	Zimbabwe	12.0	0.8	Colombia	45.0	11.5	Colombia	45.0	11.5
South Africa	55.0	0.8				Ecuador	16.0	11.5	Ecuador	16.0	11.5
South Sudan	11.0	0.8				Peru	31.0	11.5	Peru	31.0	11.5
Tanzania	55.0	0.8				Venezuela	28.0	11.5	Venezuela	28.0	11.5
Togo	7.5	0.8									
Tunisia	11.0	0.8									
Uganda	28.0	0.8									
Zambia	11.0	0.8									
Zimbabwe	12.0	0.8									

Small multiples - Report Card

- Name: Small multiples
- Purpose: permutation of distributions or relationships, multivariate relationships
- Application: same
- Grades
 - Data density: **A**
 - Popularity: **D**

Small multiples: 4-dimensional line chart

(Prevalence by sex, year age group)

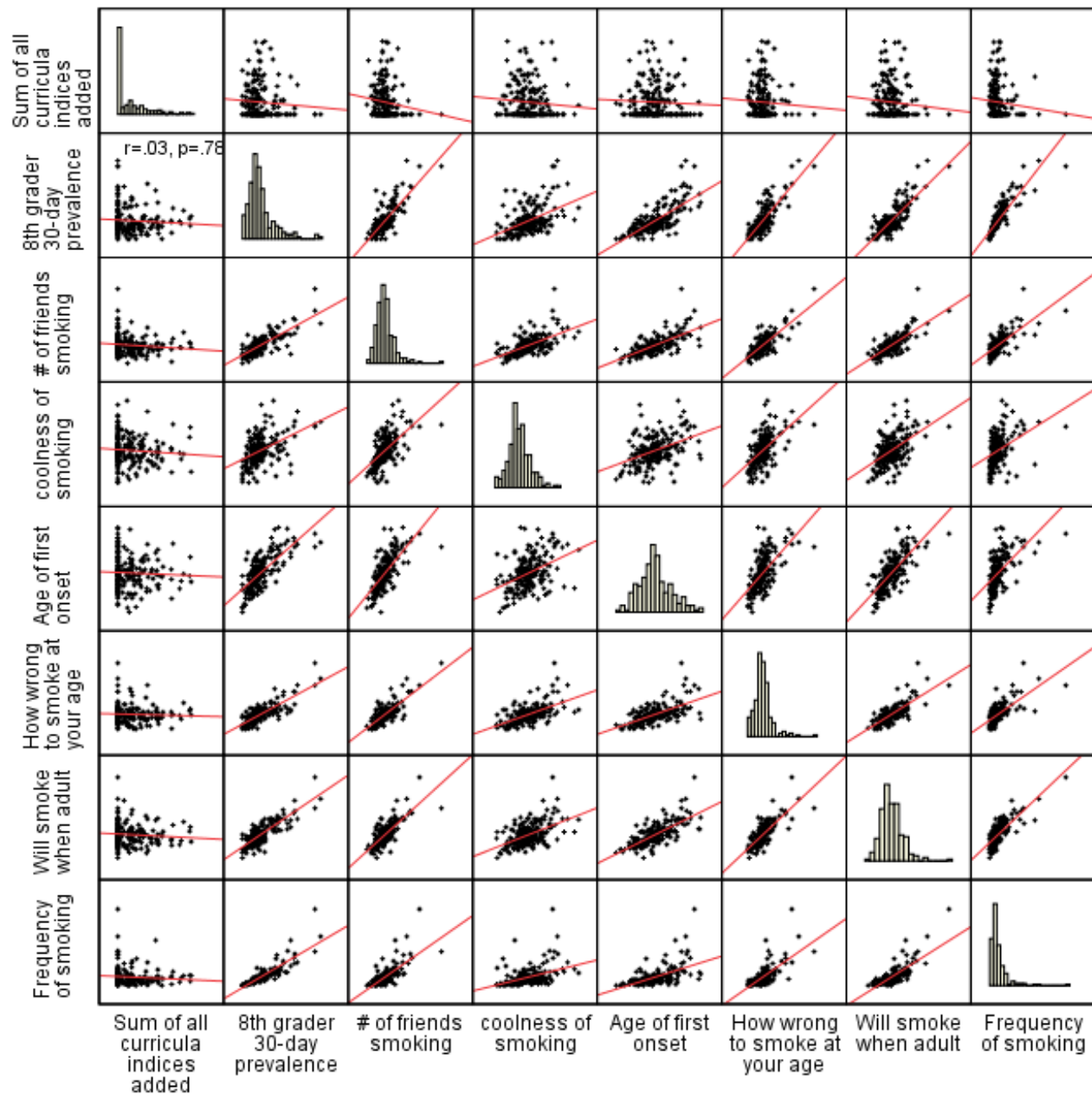


Smoking prevalence for each year broken down by Age group. Color shows details about Sex. The view is filtered on Age group, which keeps 18-24, 25-34, 35-44, 45-54 and 55-64. The marks are labeled by sum of Smoking prevalence.

Sex

Female

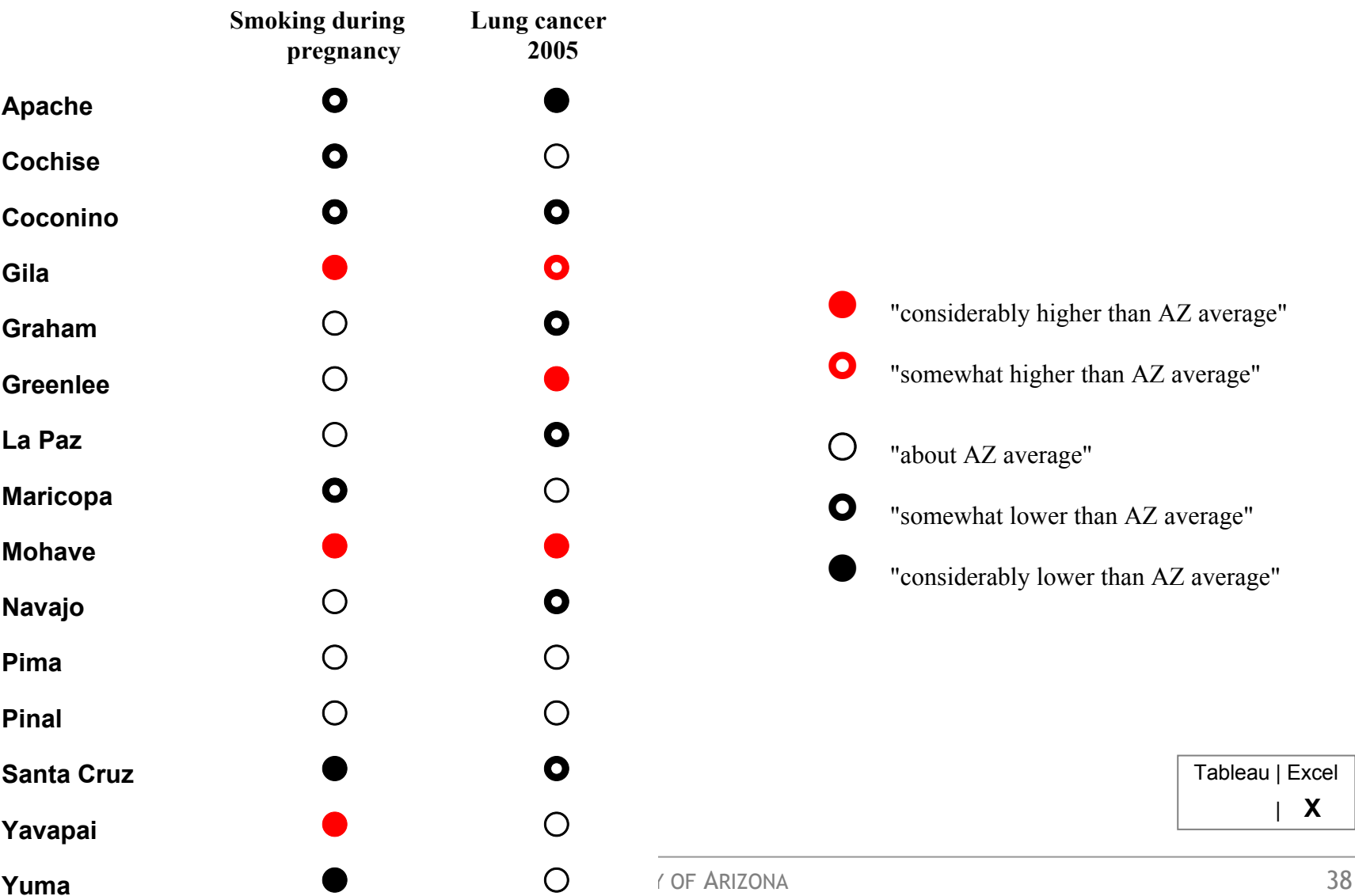
Male



Consumer report graph tables - Report Card

- Name: Consumer report graph tables
- Purpose: relational comparisons (standardized comparisons)
- Application: Comparing many objects on many dimensions
- Grades
 - Data density: **A**
 - Popularity: **F**

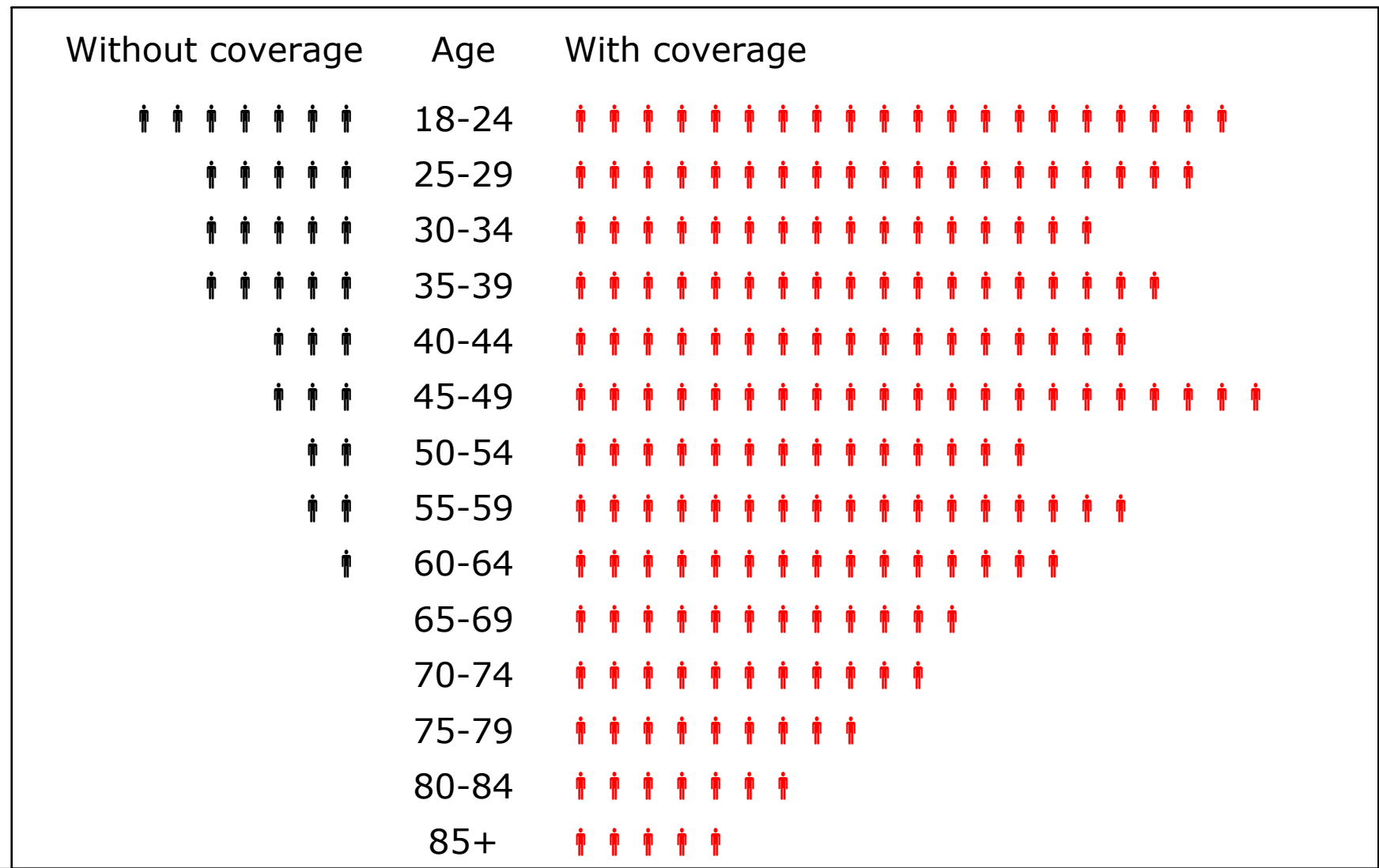
Consumer report graph tables



Natural frequency graph- Report Card

- Name: Natural frequency graph
- Purpose: number of events per category, distributions expressed in whole numbers
- Application: same
- Grades
 - Data density: **A**
 - Popularity: **F**

Natural frequency graph



Chicks Rule!

Gender balance on social networking sites

			% M*
equality	LinkedIn		50
	YouTube		50
	deviantART		50
	del.icio.us	♂♀	52
matriarchy	hi5	♀♀♀♀	54 2.4
	flickr	♀♀♀♀♀	55 6.0
	friendfeed	♀♀♀♀♀	55 0.2
	twitter	♀♀♀♀♀♀♀	57 7.7
	facebook	♀♀♀♀♀♀♀	57 46
	Ning	♀♀♀♀♀♀♀♀♀	59 2.5
	gaia ONLINE	♀♀♀♀♀♀♀♀♀♀♀	61 0.3
	classmates.com	♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀	64 2.2
	myspace	♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀	64 27
	BUZZnet	♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀	64 1.4
	TAGGED	♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀	64 3.6
	bebo	♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀♀	68 3.2
patriarchy	digg	♂♂♂♂♂♂♂♂♂♂♂♂♂♂♂♂	64 4.7

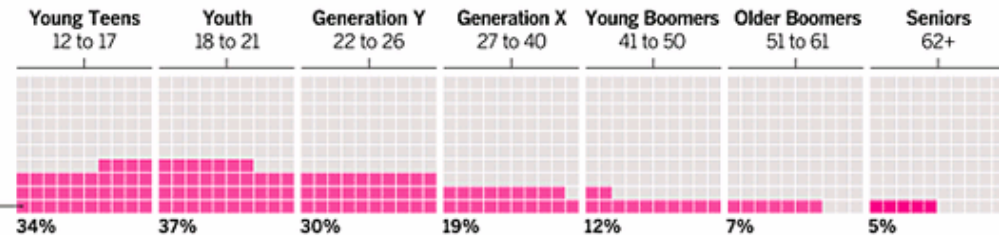
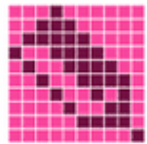
Tableau | Excel
| X

This could have been a bar chart...

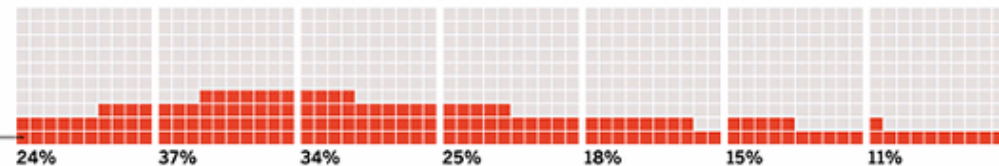
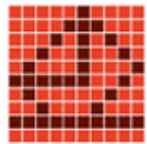
What people are doing

Who participates (U.S. online users)

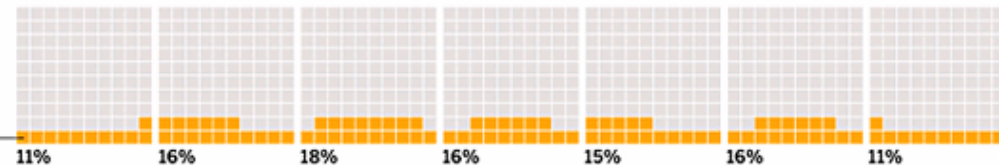
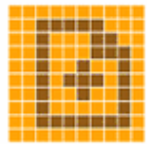
Creators publish Web pages, write blogs, upload videos to sites like YouTube.



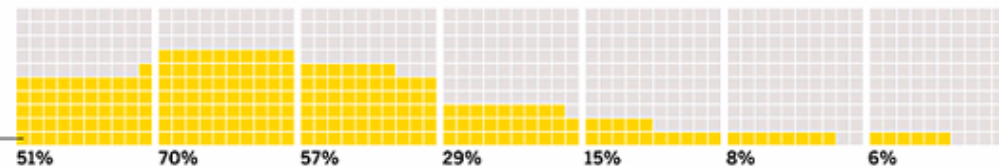
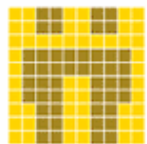
Critics comment on blogs and post ratings and reviews.



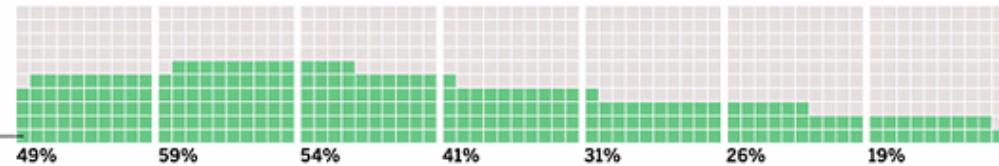
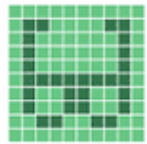
Collectors use Really Simple Syndication (RSS) and tag Web pages to gather information.



Joiners use social networking sites.



Spectators read blogs, watch peer-generated videos, and listen to podcasts.



Inactives are online but don't yet participate in any form of social media.

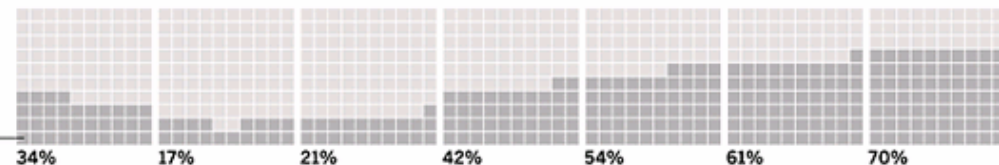
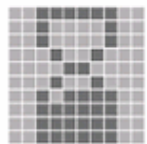


CHART BY ARNO GHELF

Gigerenzer et al. 2007

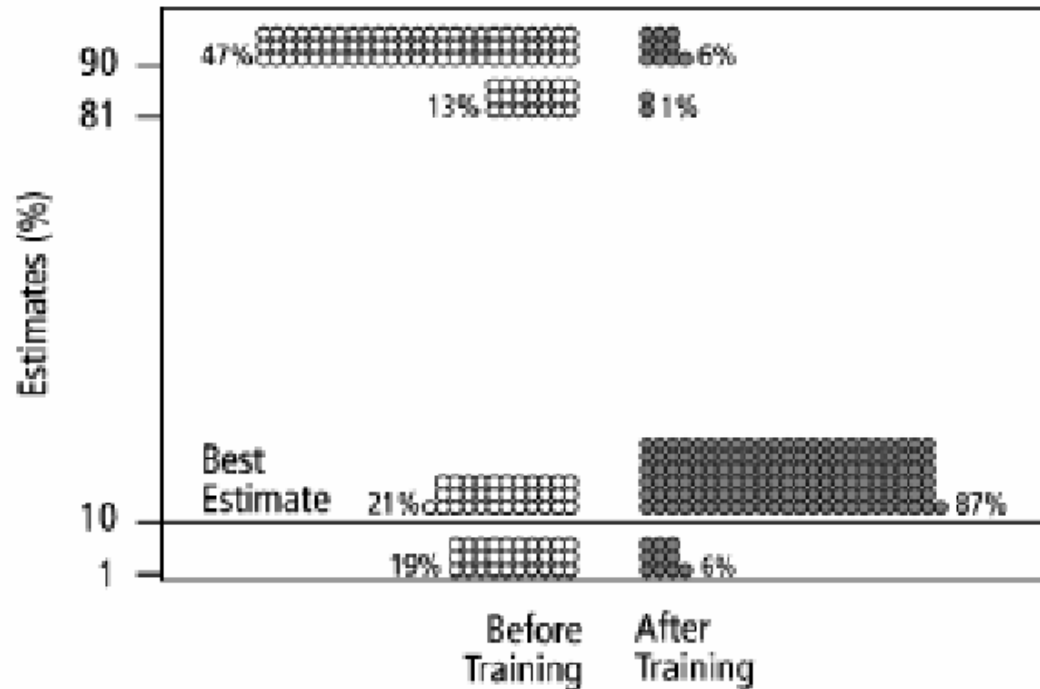


Fig. 2. Estimates by 160 gynecologists of the probability that a woman has breast cancer given a positive mammogram, before and after receiving training in how to translate conditional probabilities into natural frequencies.