



MASTER OF SCIENCE IN BUSINESS ANALYTICS

Data visualization: Examples of bad graphics and recommendations for doing better

Outline

The data viz hall of shame

How to do better

The data viz hall of shame

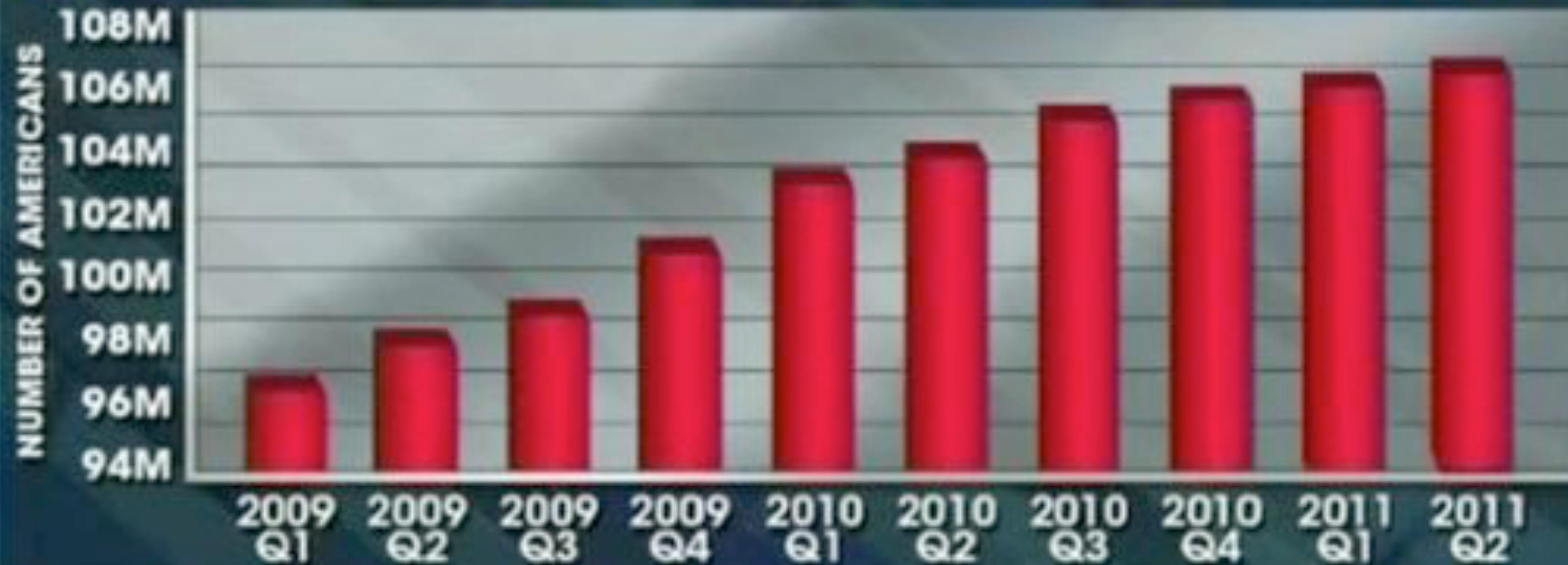
Plotting pitfalls

- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Plotting pitfalls

- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

FEDERAL WELFARE RECEIVED IN U.S.



FOX NEWS FOX
NEWS NETS
.COM

SOURCE: U.S. CENSUS SURVEY

:LIST TYLER HAMILTON OF HIS GOLD MEDAL FRO!

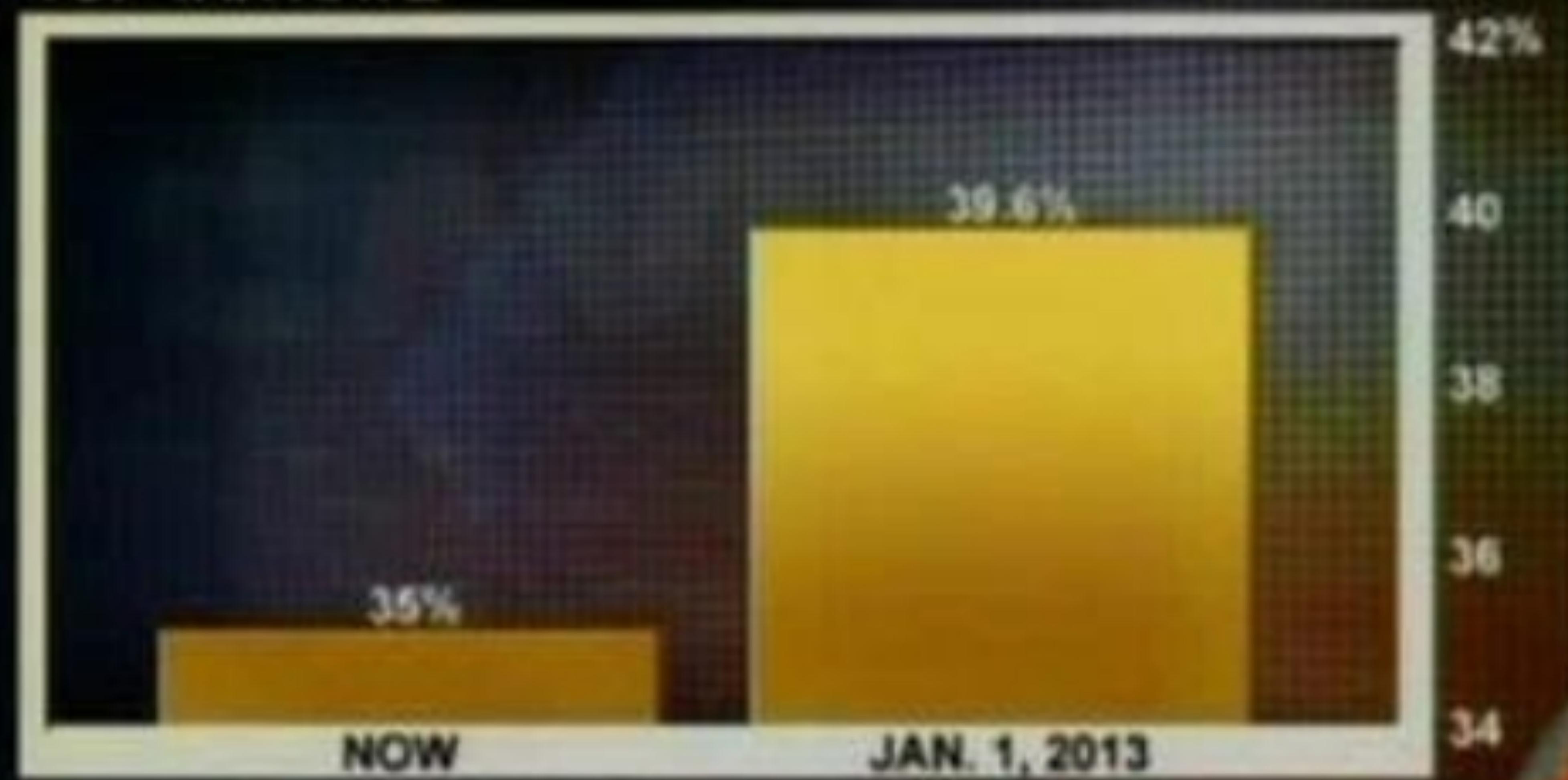
NAS



6.61

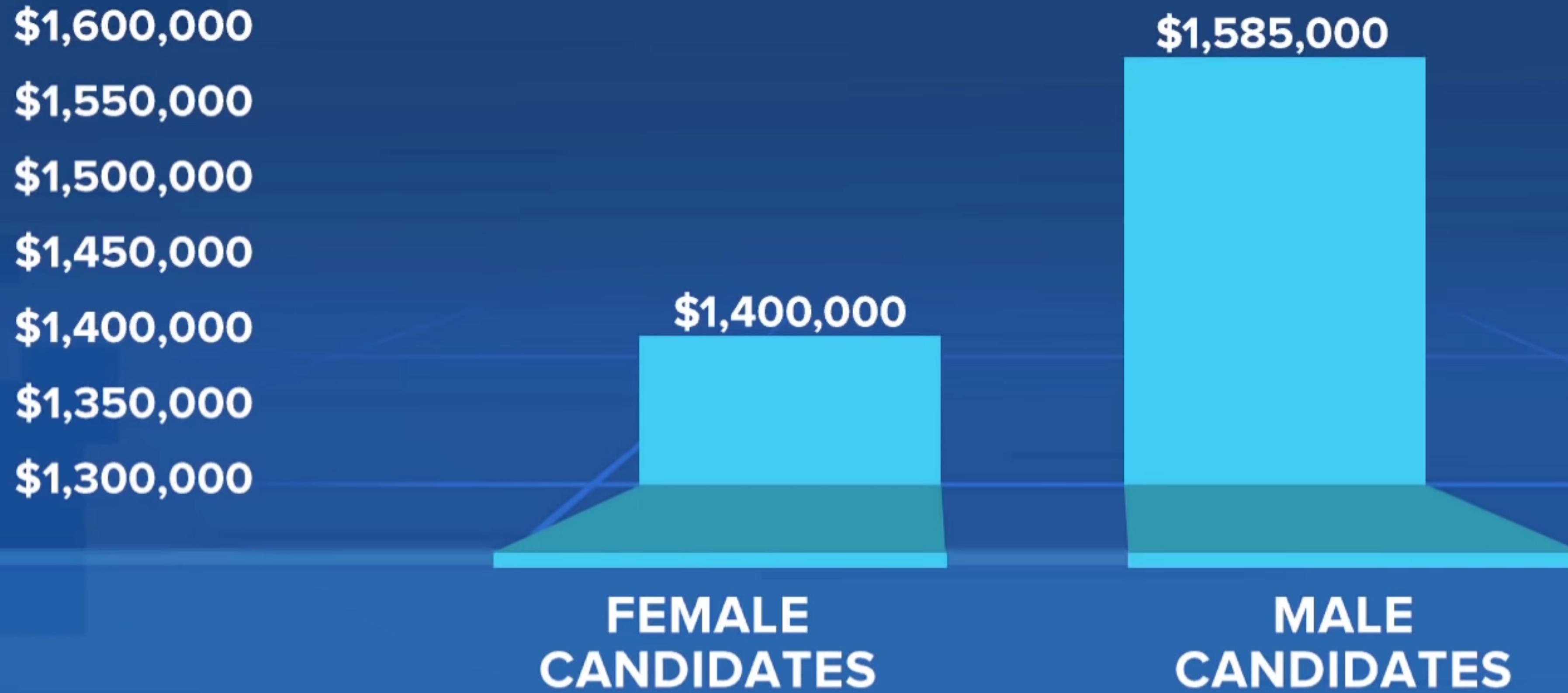
IF BUSH TAX CUTS EXPIRE

TOP TAX RATE



DEMOCRATIC WOMEN RAISE LESS MONEY THAN DEMOCRATIC MEN

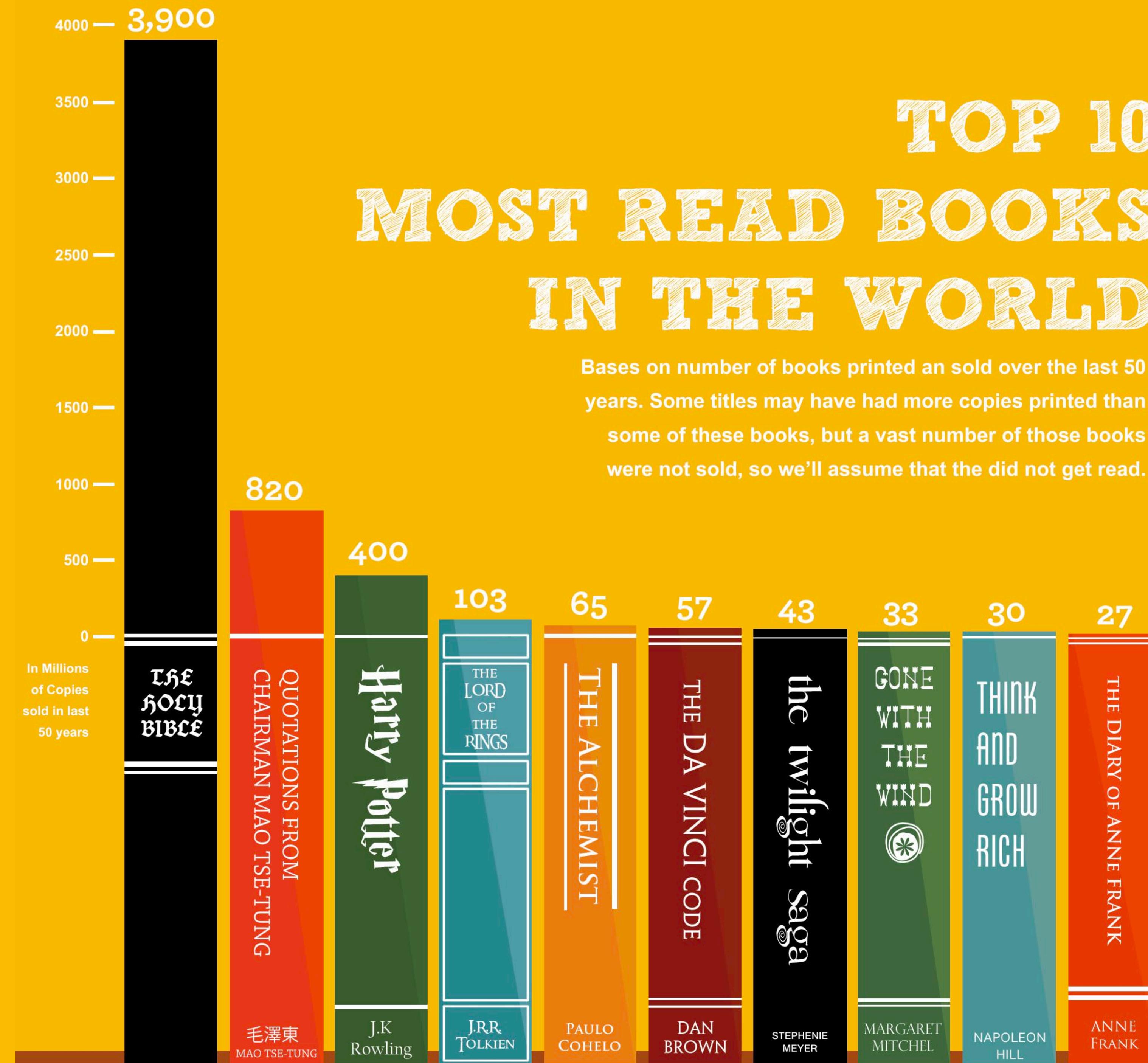
AVERAGE RAISED BY HOUSE PRIMARY WINNERS



SOURCE: CENTER FOR RESPONSIVE POLITICS

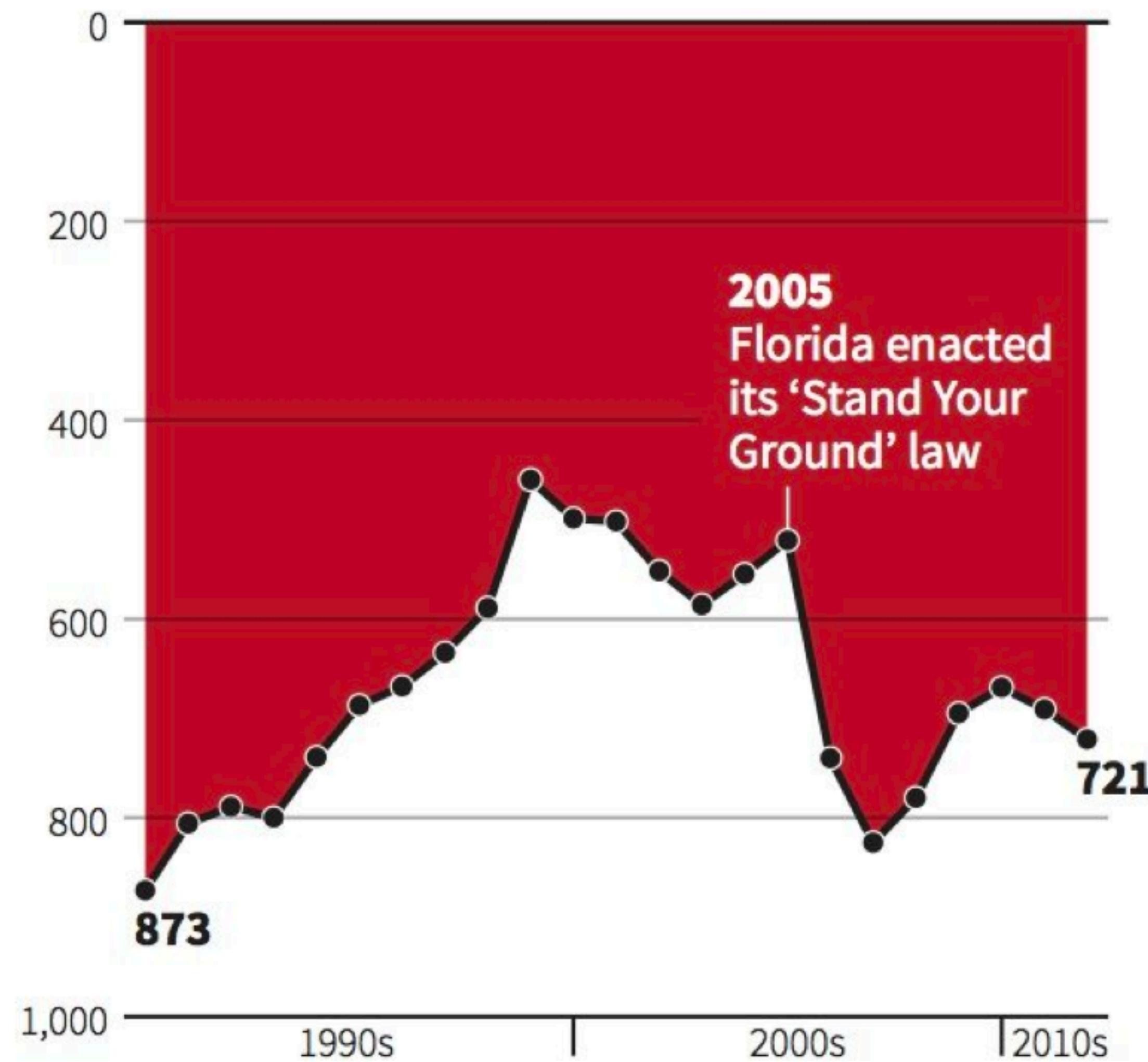
TOP 10 MOST READ BOOKS IN THE WORLD

Bases on number of books printed an sold over the last 50 years. Some titles may have had more copies printed than some of these books, but a vast number of those books were not sold, so we'll assume that they did not get read.



Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

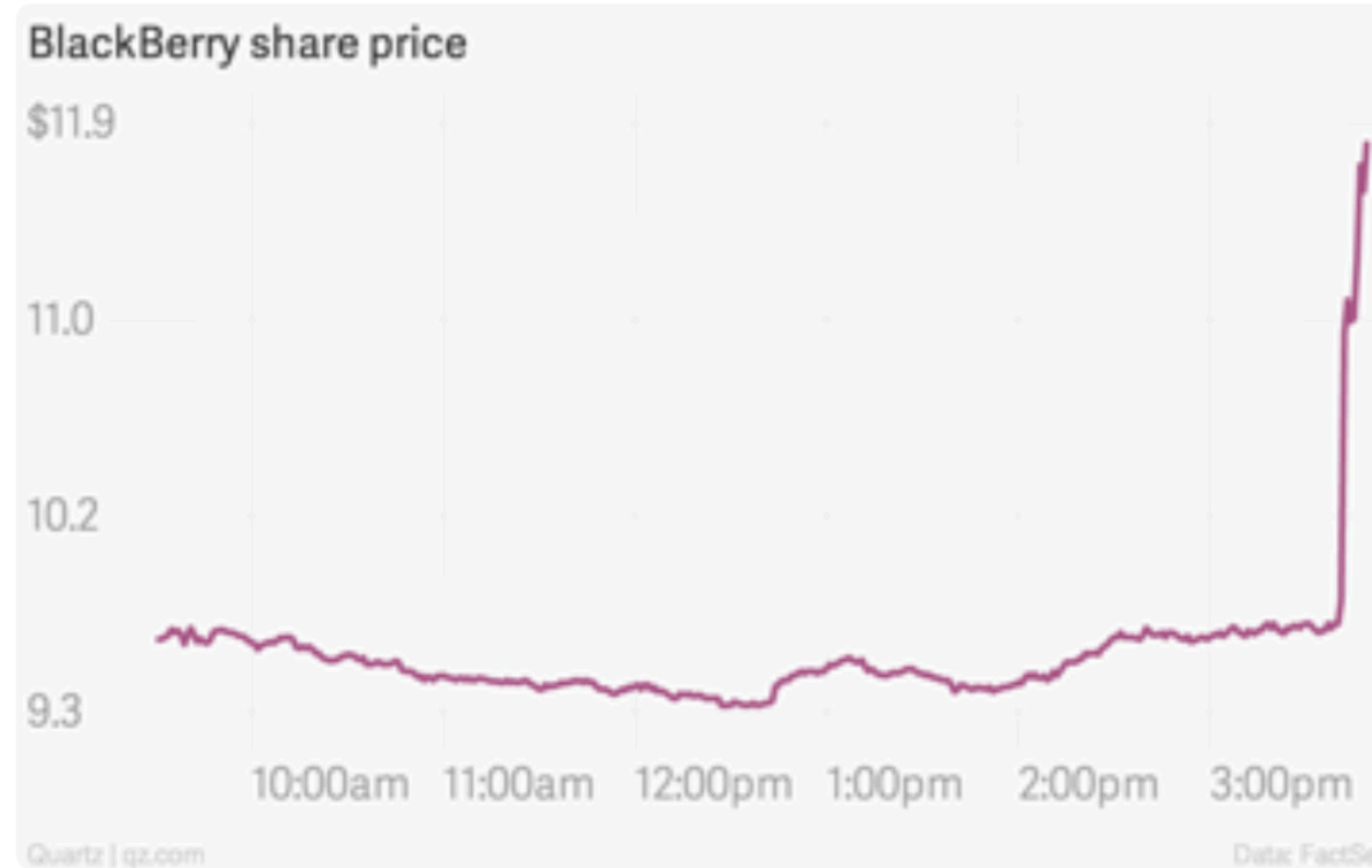
AND YET...



Quartz @qz

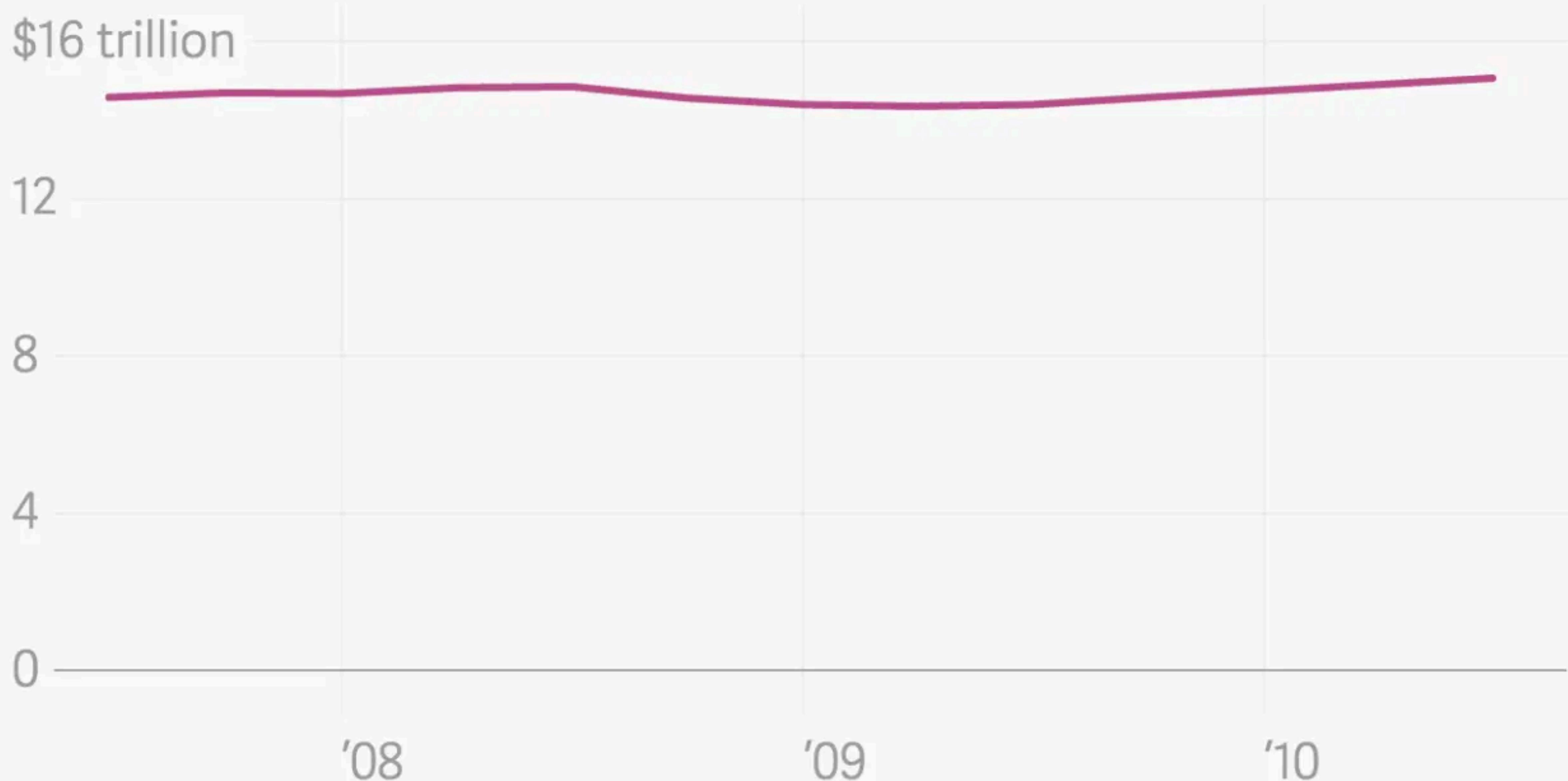
Acquisition rumors light a fire under BlackBerry's stock.

qz.com/note/326942/



6 4:44 PM - Jan 14, 2015

US GDP



US GDP

\$15.00 trillion

14.75

14.50

14.25

14.00

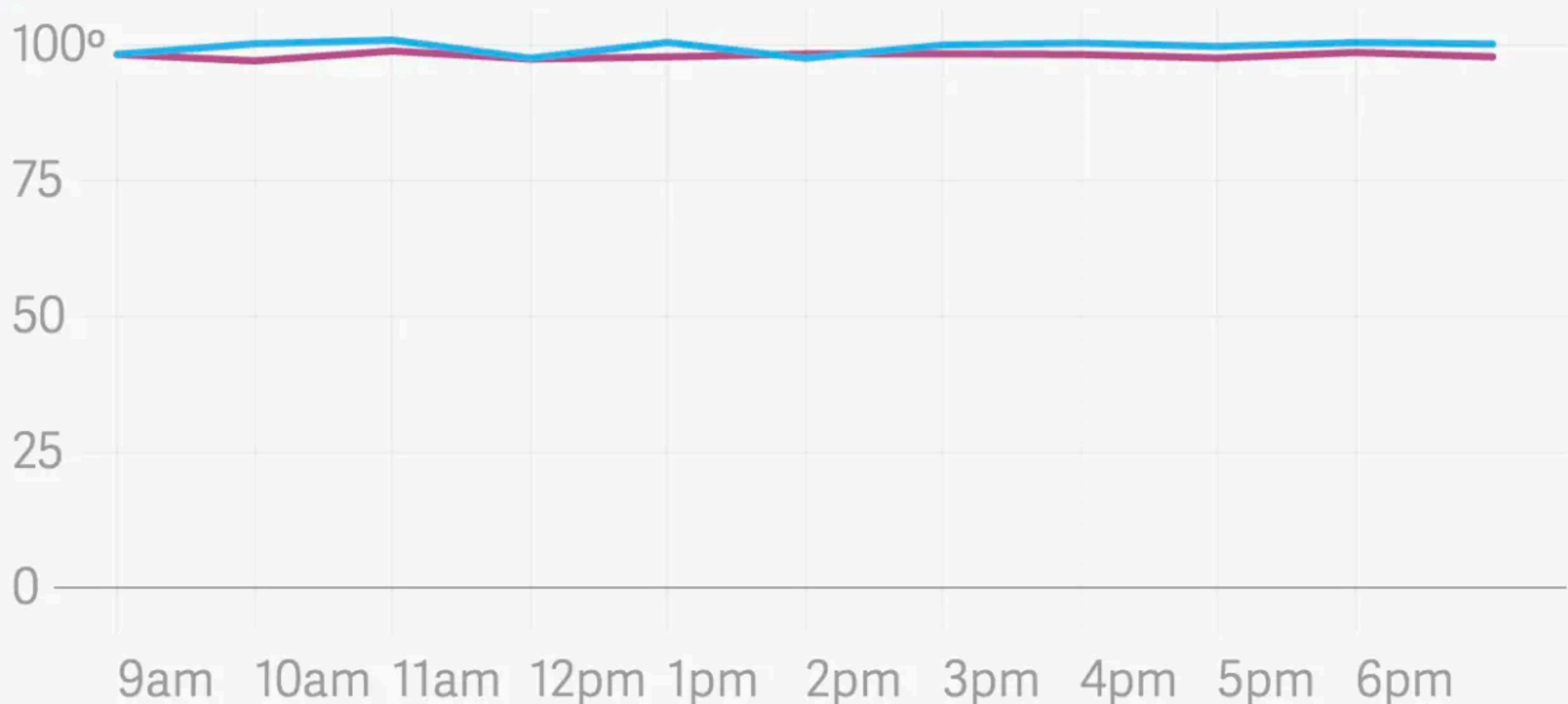
'08

'09

'10

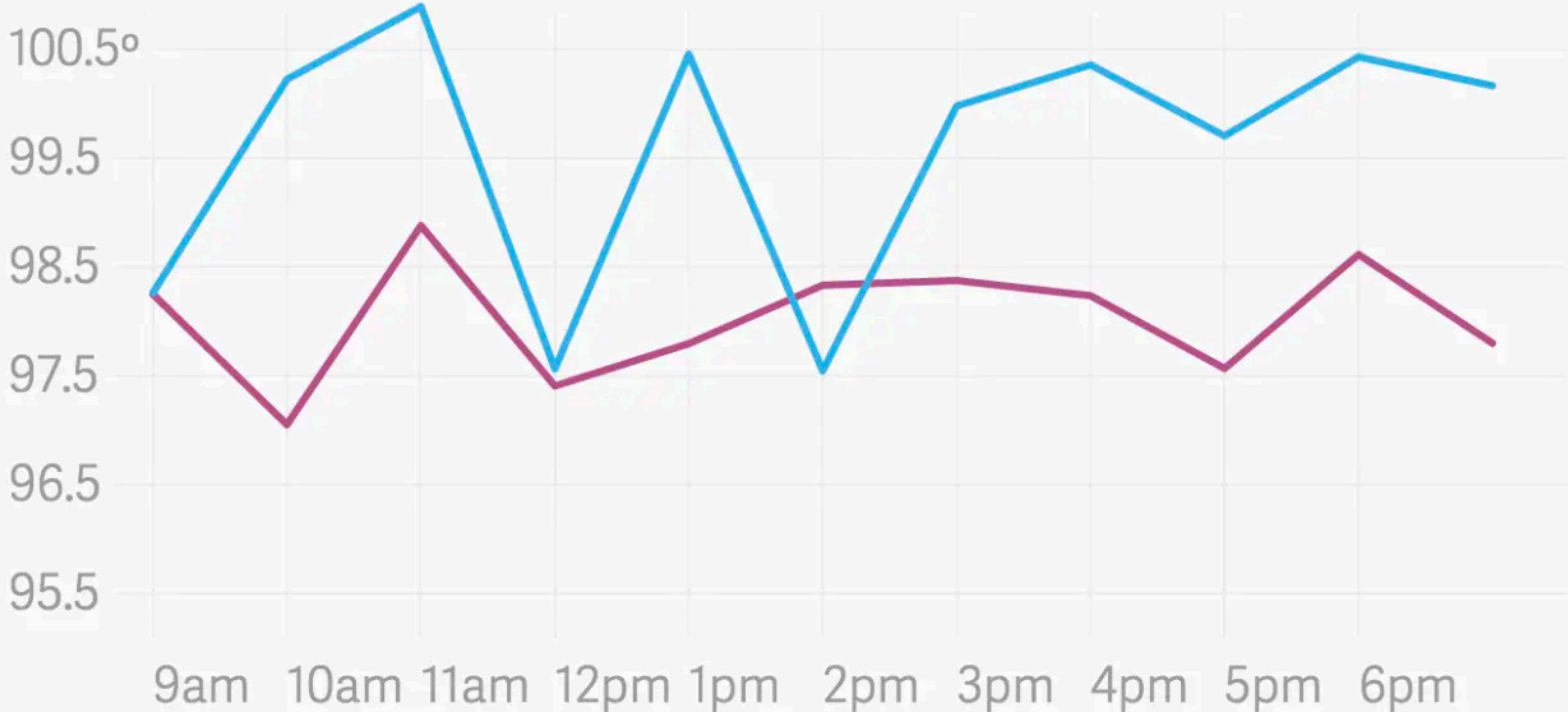
Oral temperature

Sara Bob



Oral temperature

Sara Bob



Truncating the y-axis is sometimes ok

- When you're trying to emphasize change, rather than relative magnitude.
- When you're plotting data over time.
- When zero is not a sensible baseline for comparison.

Bottom line: use your judgment; don't mislead people;
watch out for “little y lies.”

Plotting pitfalls

- Axis trickery
- **Violations of basic math**
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

RASMUSSEN REPORTS POLL

Did scientists falsify research to support their own theories on Global Warming?

59%

SOMEWHAT LIKELY

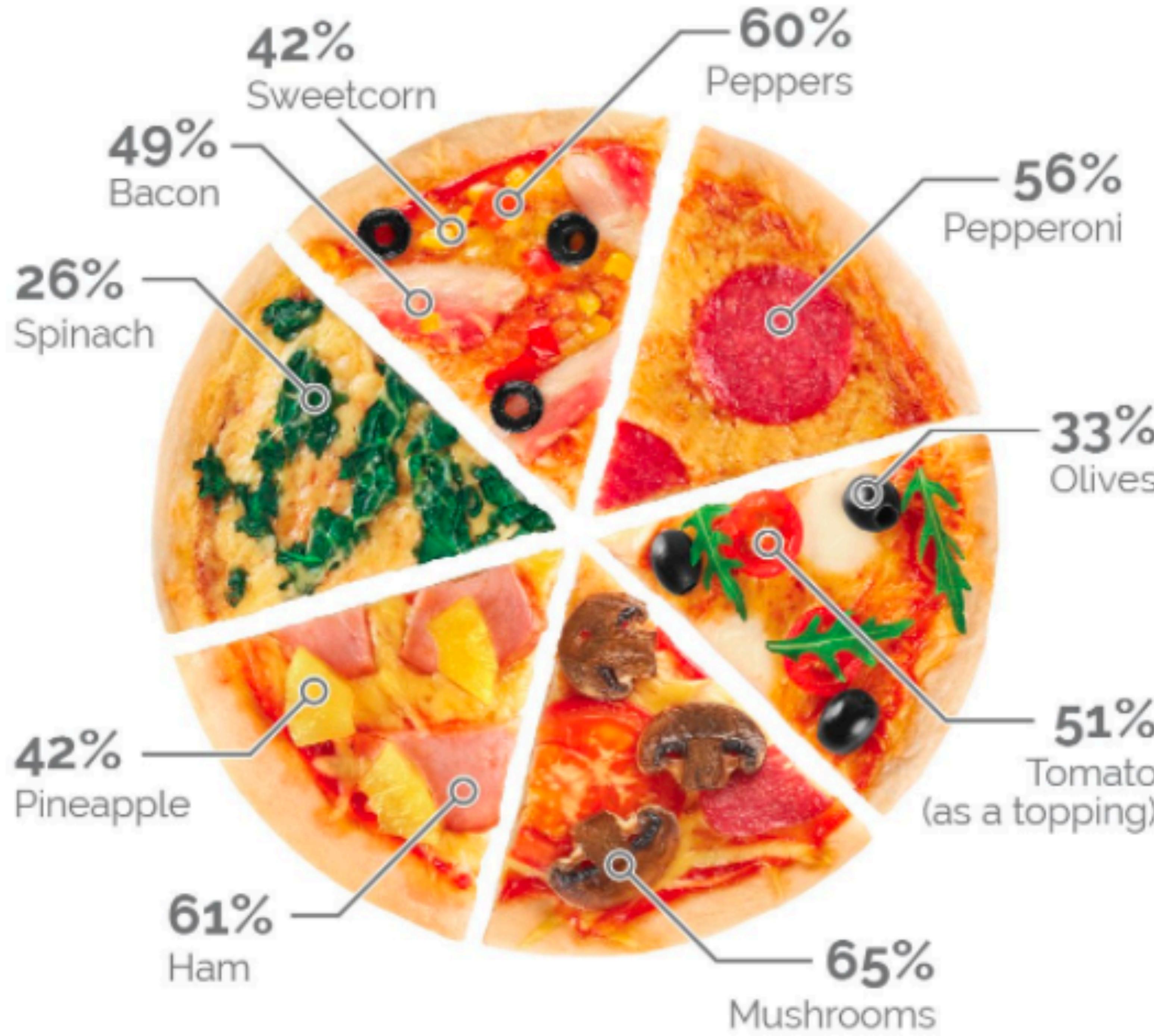
35%

VERY LIKELY

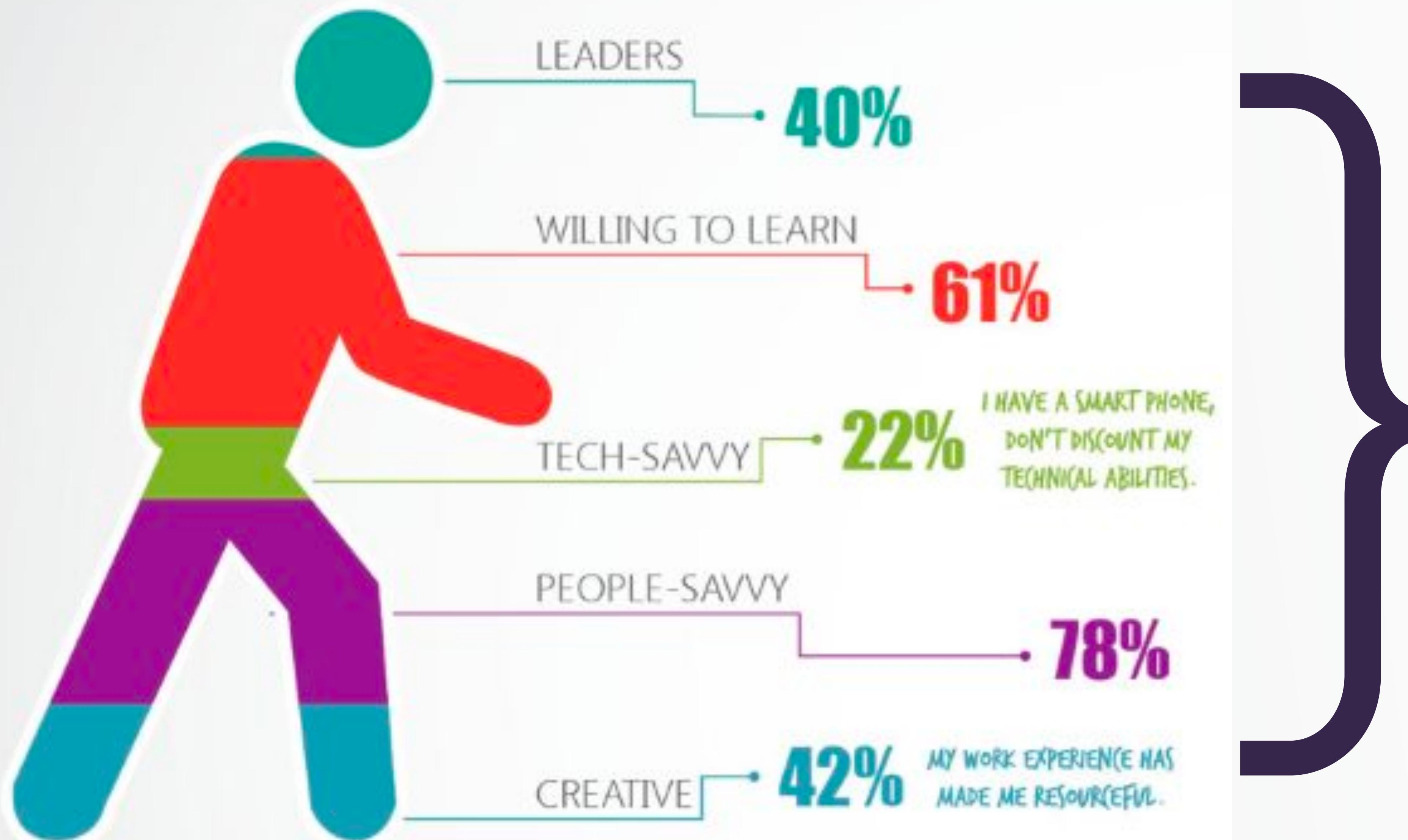
26%

NOT VERY LIKELY





HOW BABY BOOMERS DESCRIBE THEMSELVES



UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA



FOX
NEWS
radio

2011
SOURCE: BUREAU OF LABOR STATISTICS

8.6 > 8.8?

WHAT IS YOUR FAVORITE SEASON? @FCNMIKE

53%

FALL

17%

SUMMER

17%

WINTER

13%

SPRING



75°

6:51

FIRST COAST
NEWS

#KATIESCOMMUTERS

DEADLY CRASH

DEADLY HOME INVASION

DAILY FORECAST

LOW: 64° / HIGH: 75°

LOW: 64° / HIGH: 75°

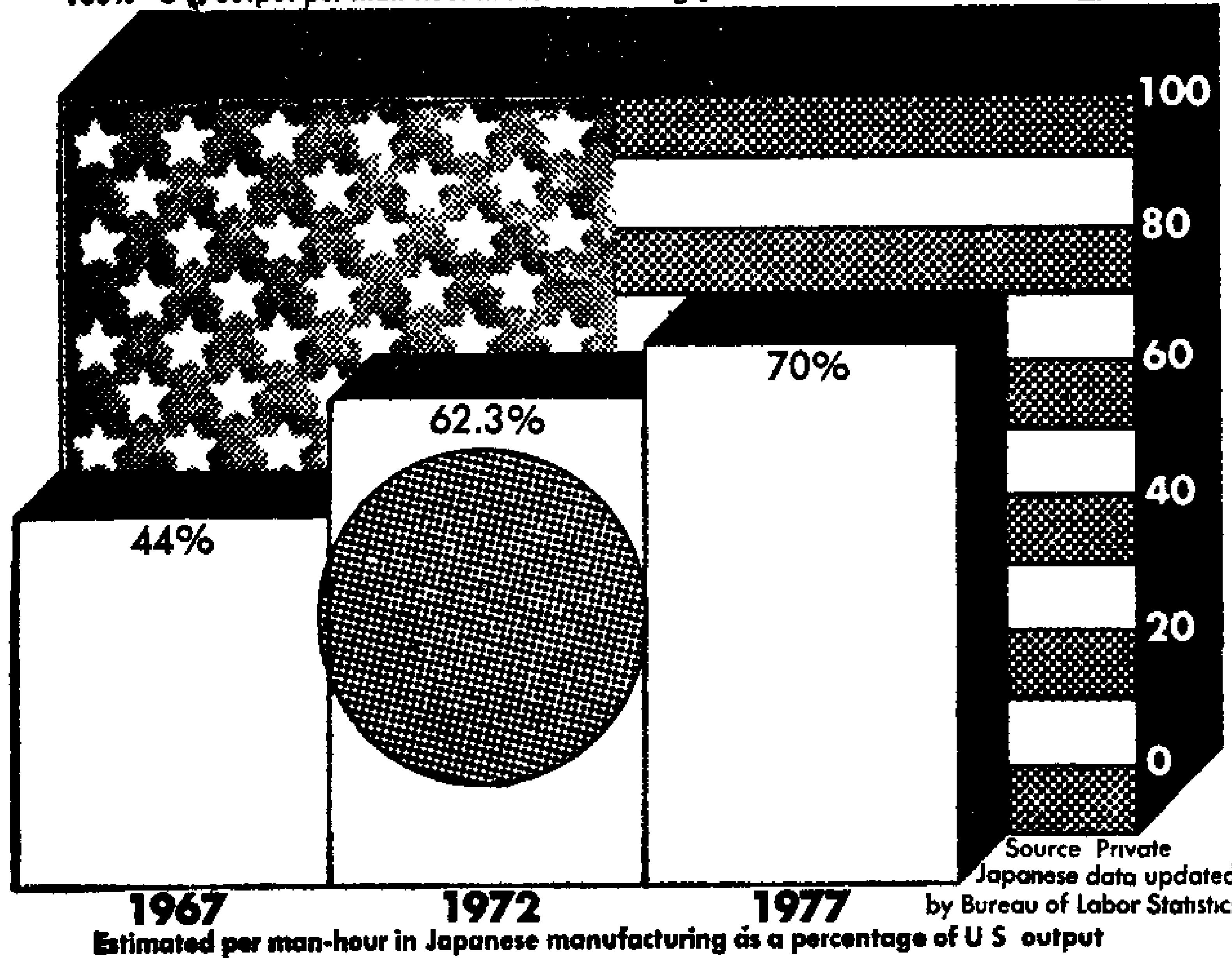
LOW: 64° / HIGH: 75°

Plotting pitfalls

- Axis trickery
- Violations of basic math
- **Nearly content-free figures**
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Labor Productivity: U.S. vs Japan

100% = U.S. output per man-hour in manufacturing

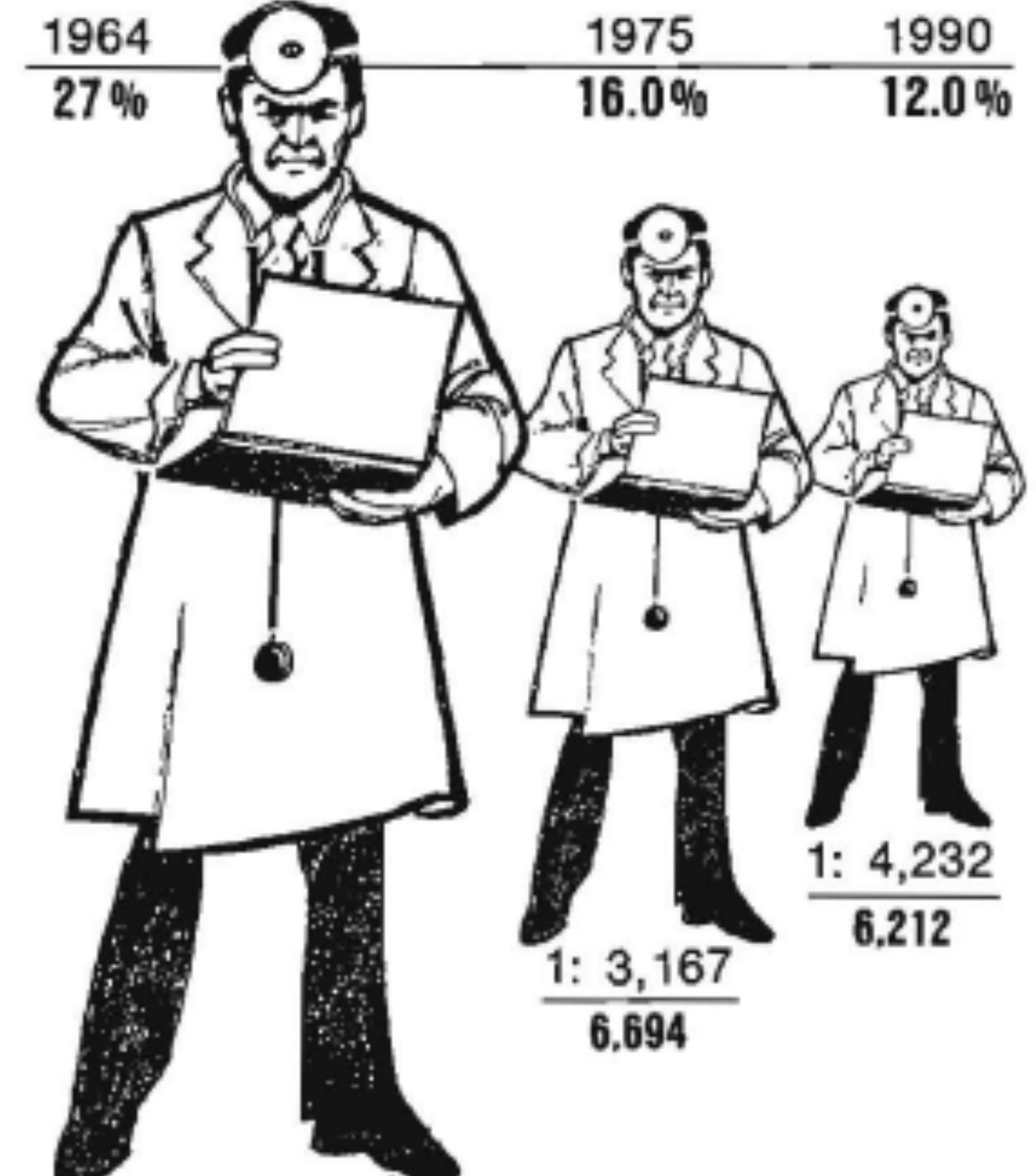


THE SHRINKING FAMILY DOCTOR

In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27%	16.0%	12.0%



TRUMP TWEETS MENTIONING MUELLER BY NAME



LIVE

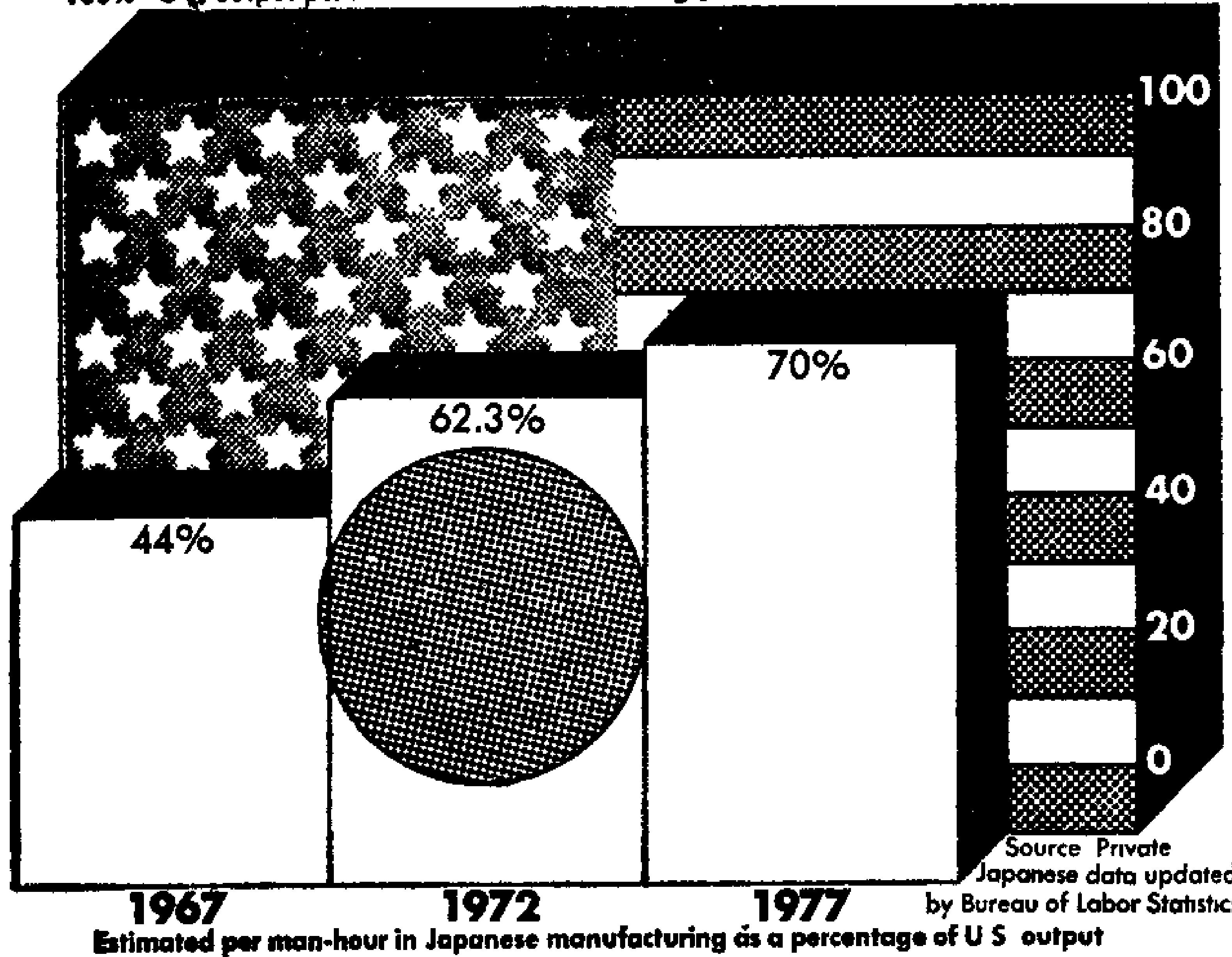
 MSNBC

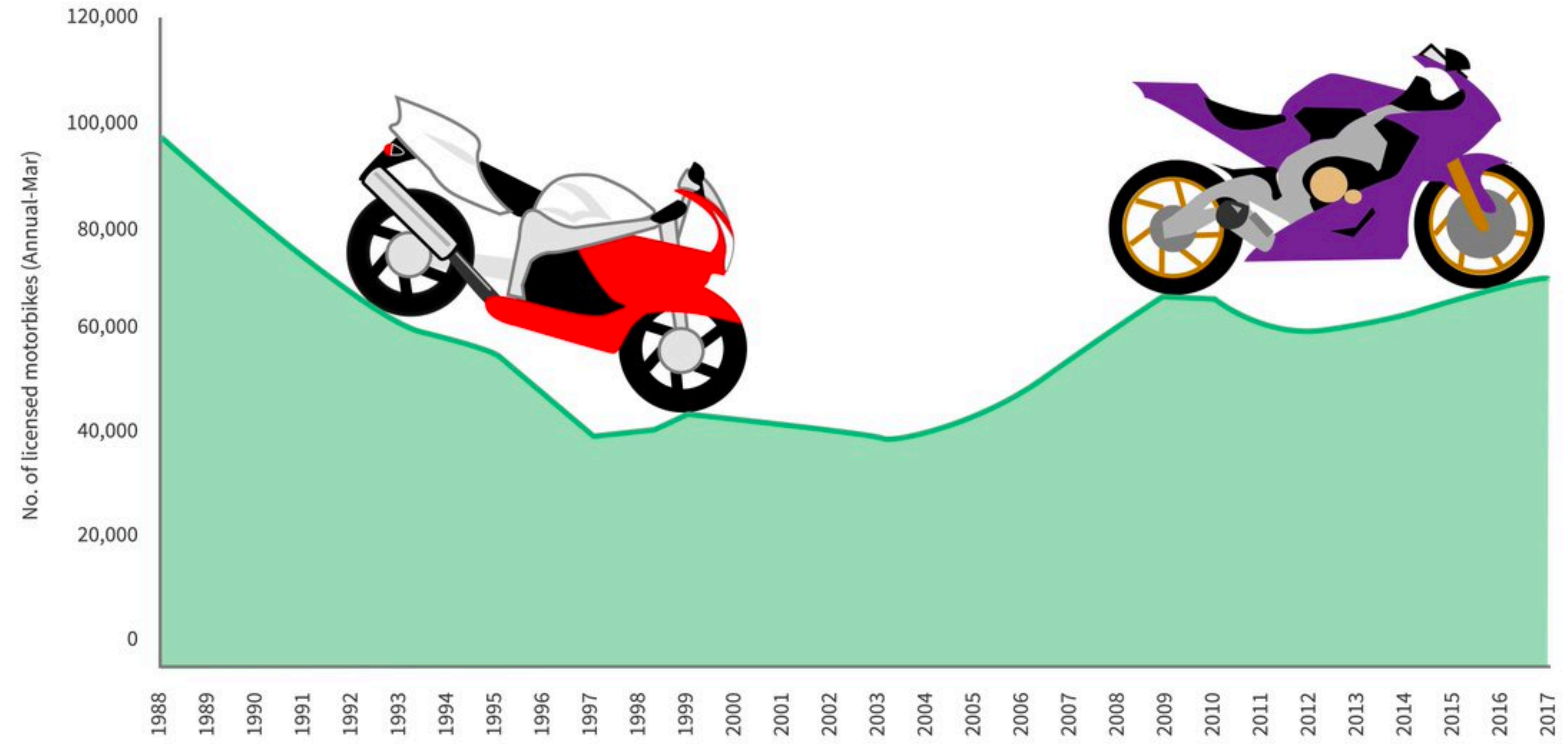
Plotting pitfalls

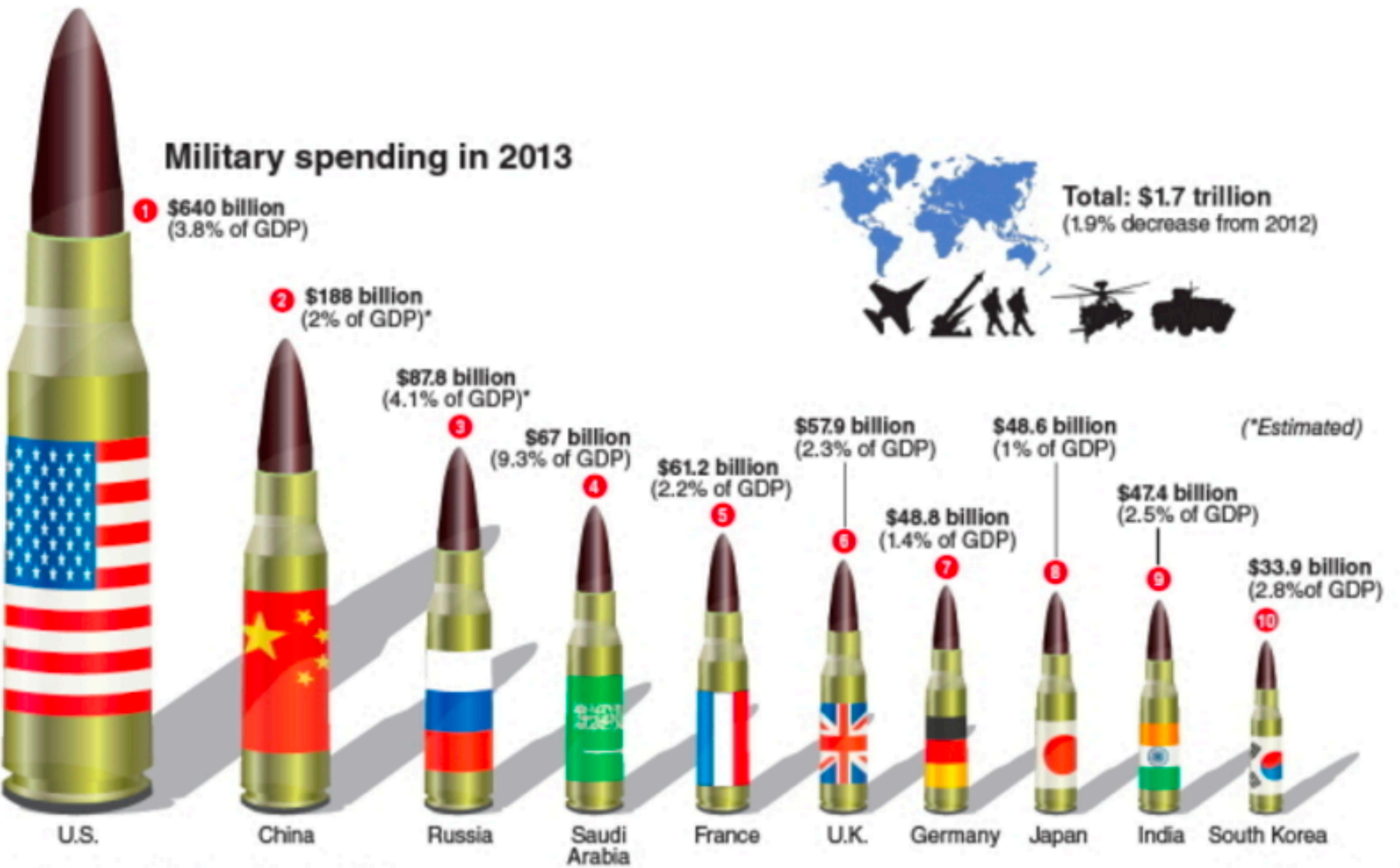
- Axis trickery
- Violations of basic math
- Nearly content-free figures
- **Gratuitous chartjunk**
- Poorly chosen 3D graphics
- Bad design choices

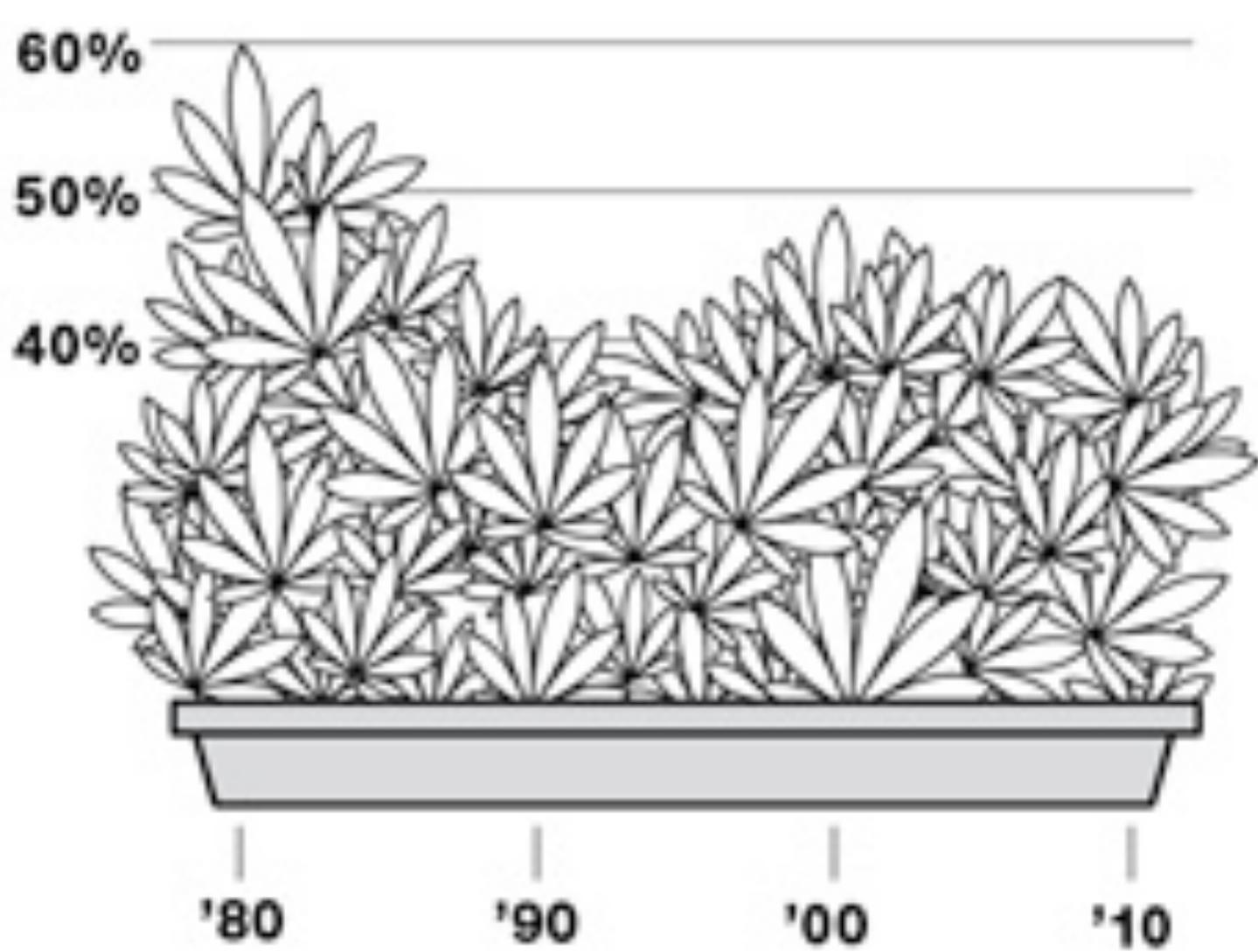
Labor Productivity: U.S. vs Japan

100% = U.S. output per man-hour in manufacturing

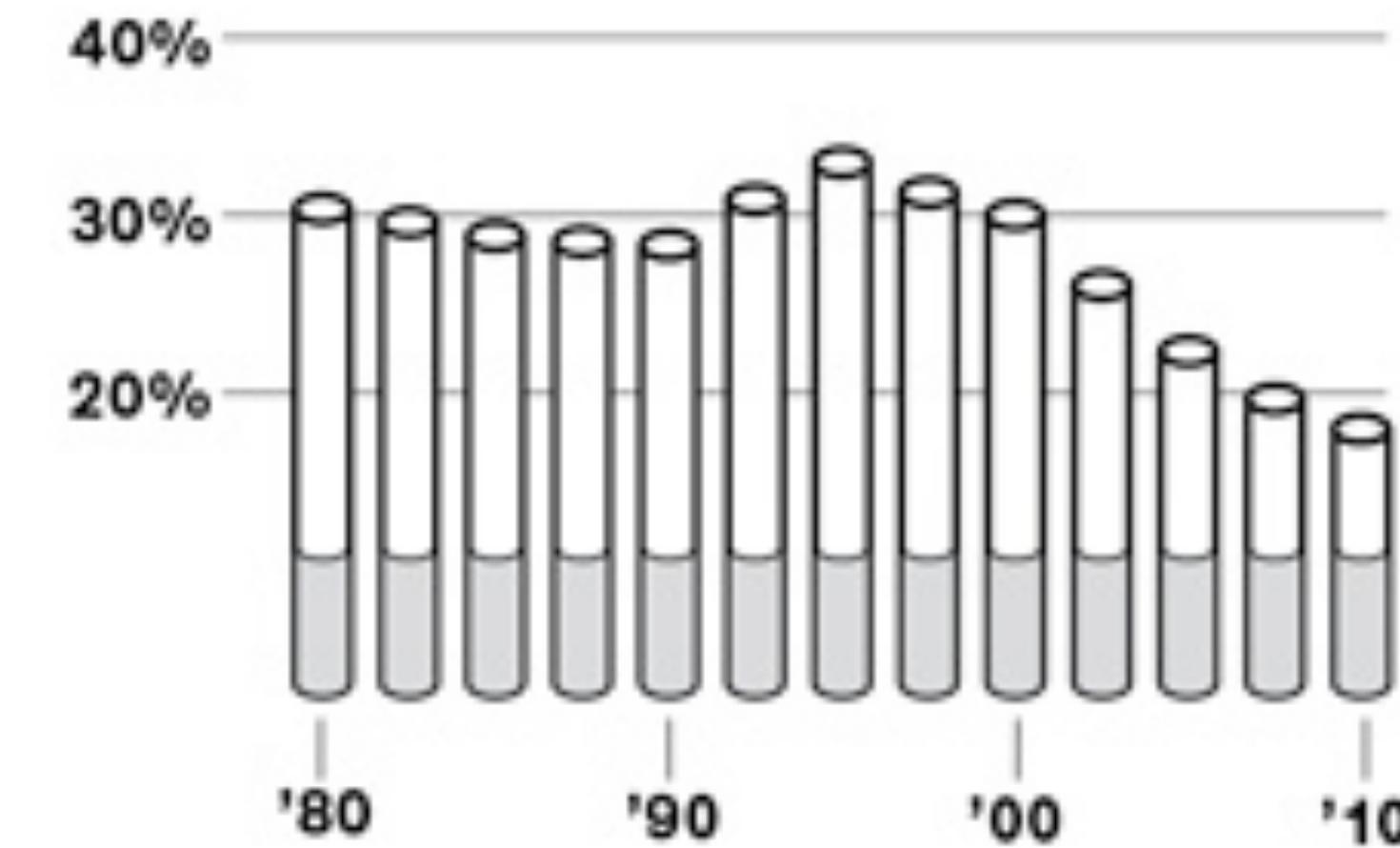




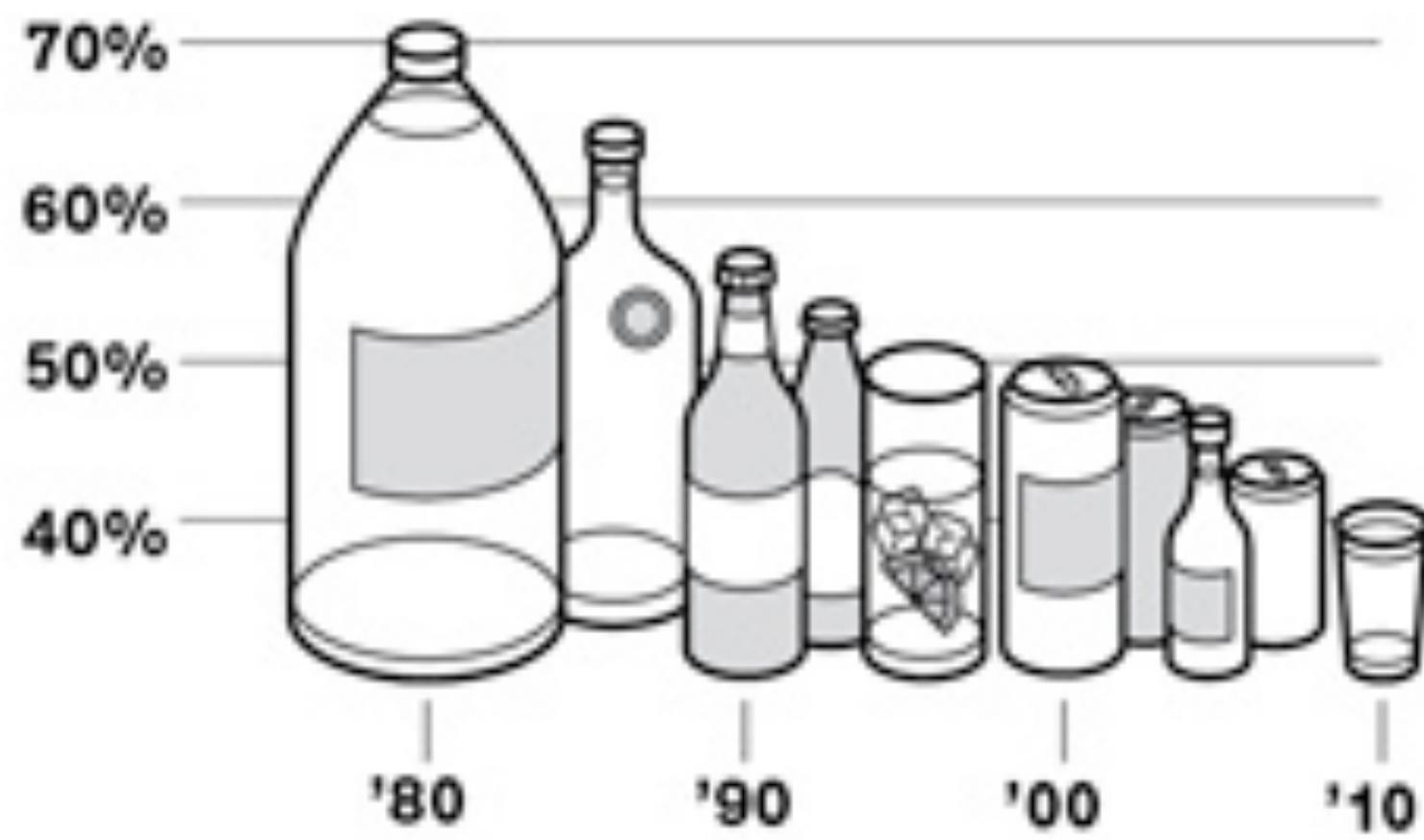




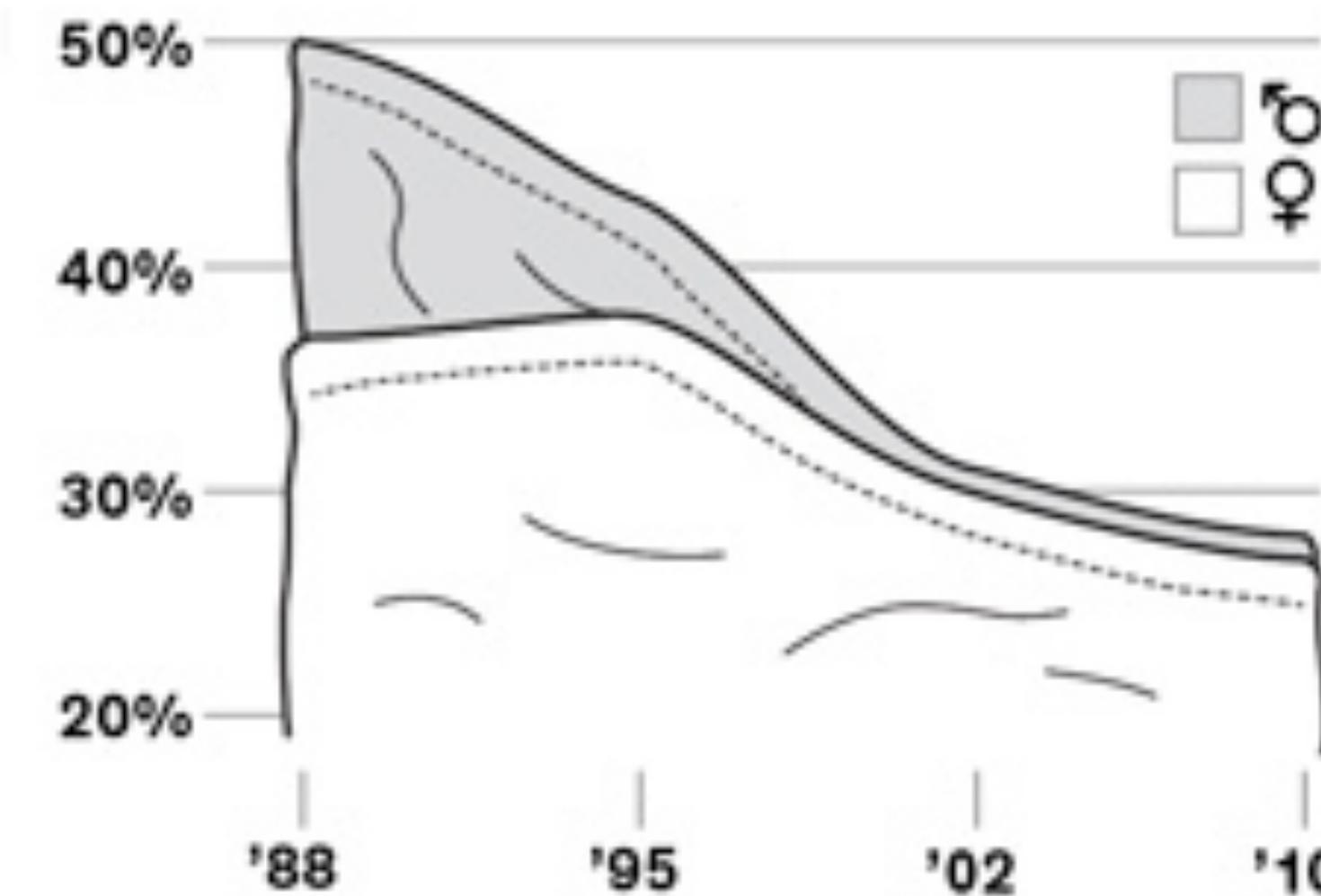
▲ Percentage of high-school seniors
who have ever tried pot.



▣ Smoking habits of high-school
seniors over time.



○ Alcohol use by high-school
seniors over time.

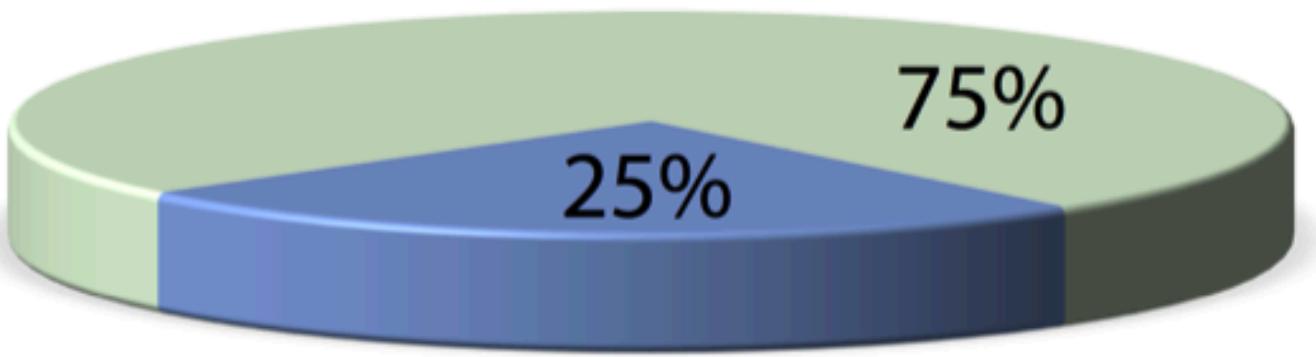


↳ Percentage of 15-to-17-year-olds
who have had sex.

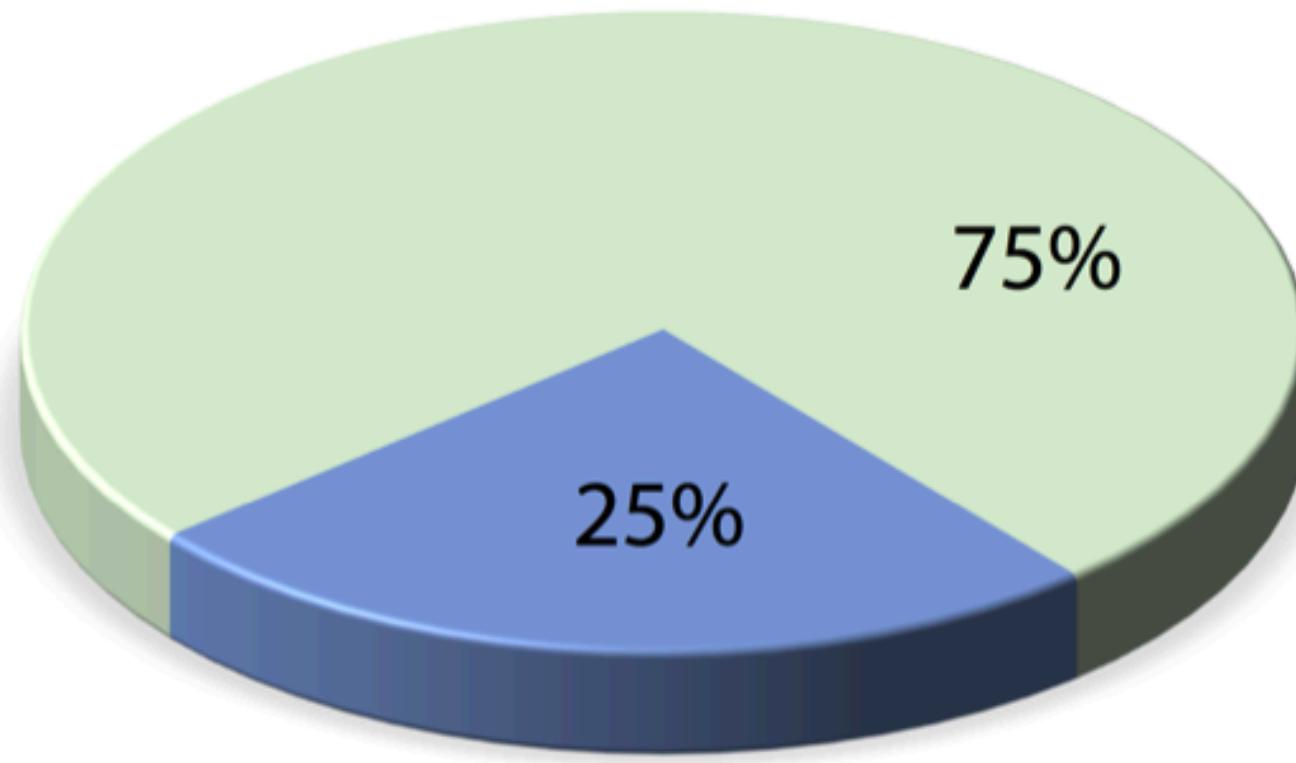
Plotting pitfalls

- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- **Poorly chosen 3D graphics**
- Bad design choices

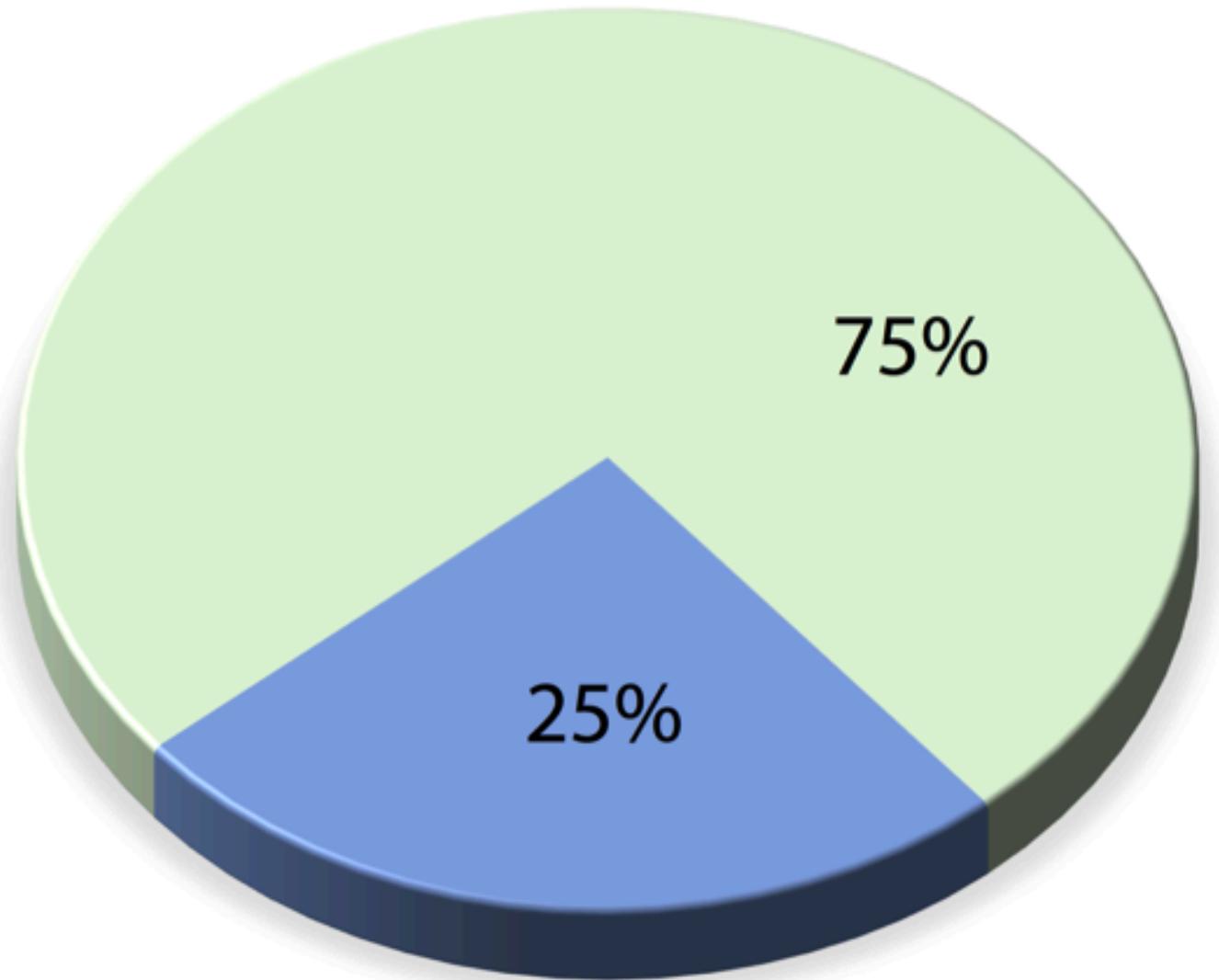
a



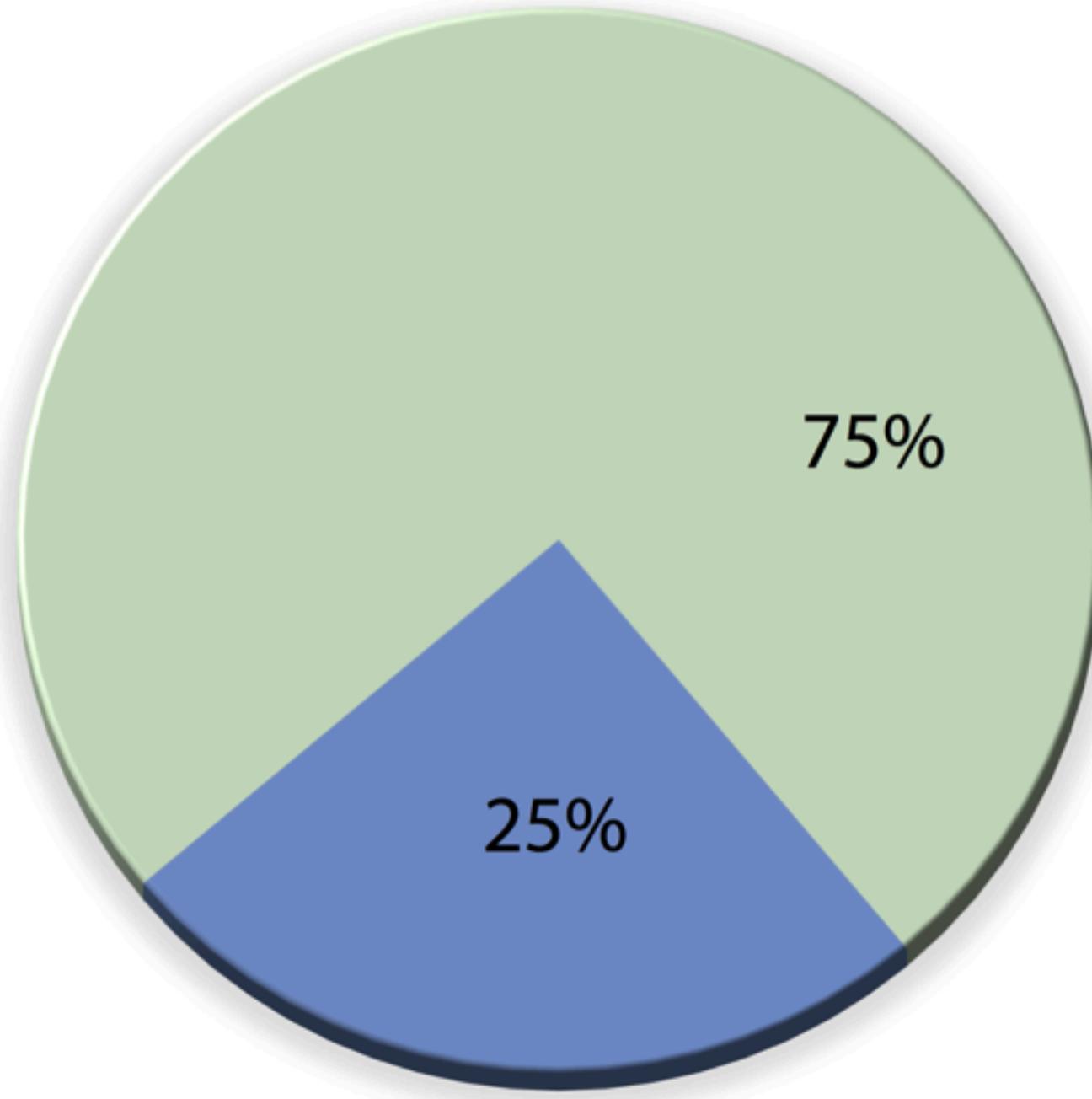
b



c



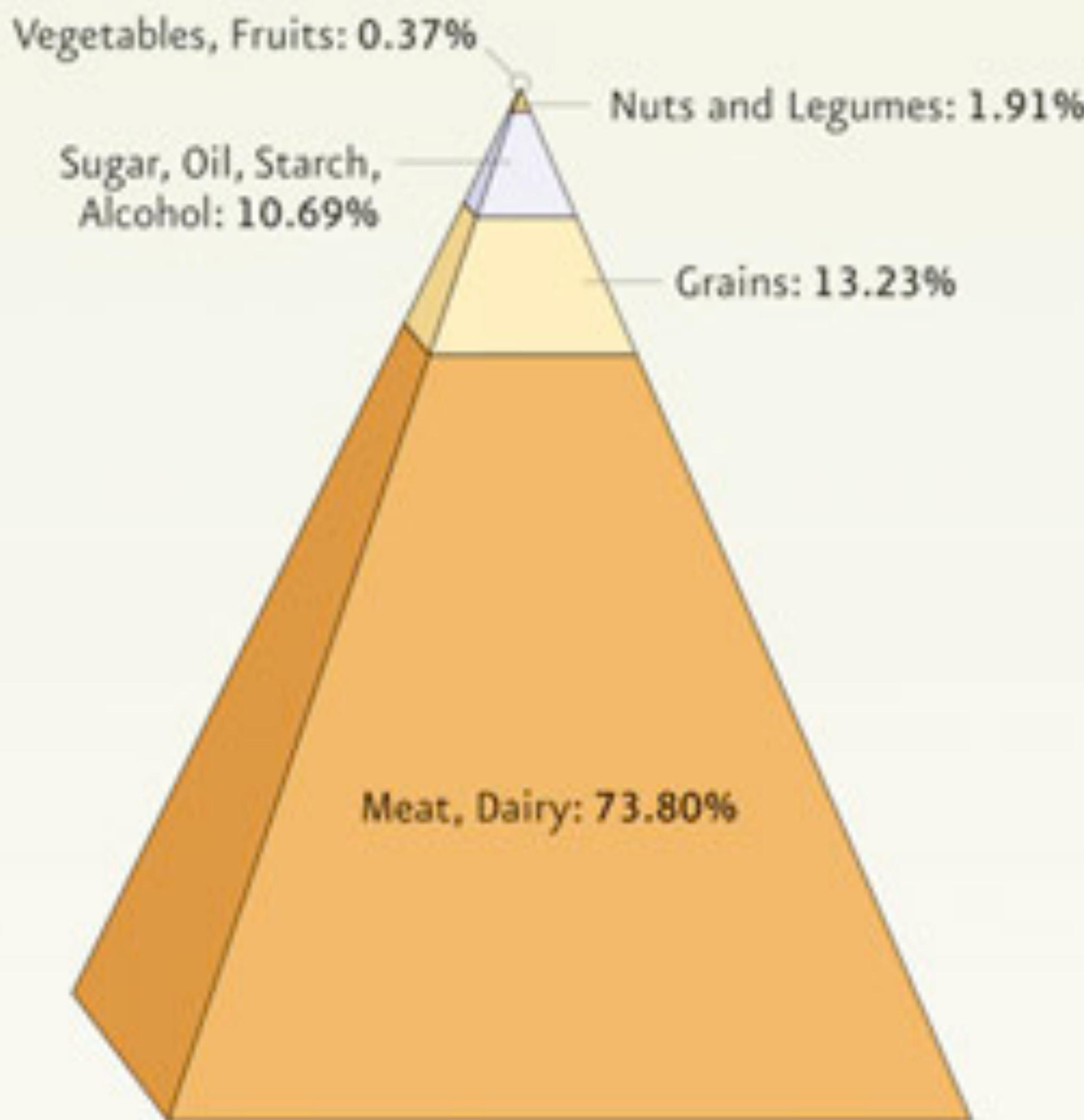
d



Same chart, four angles

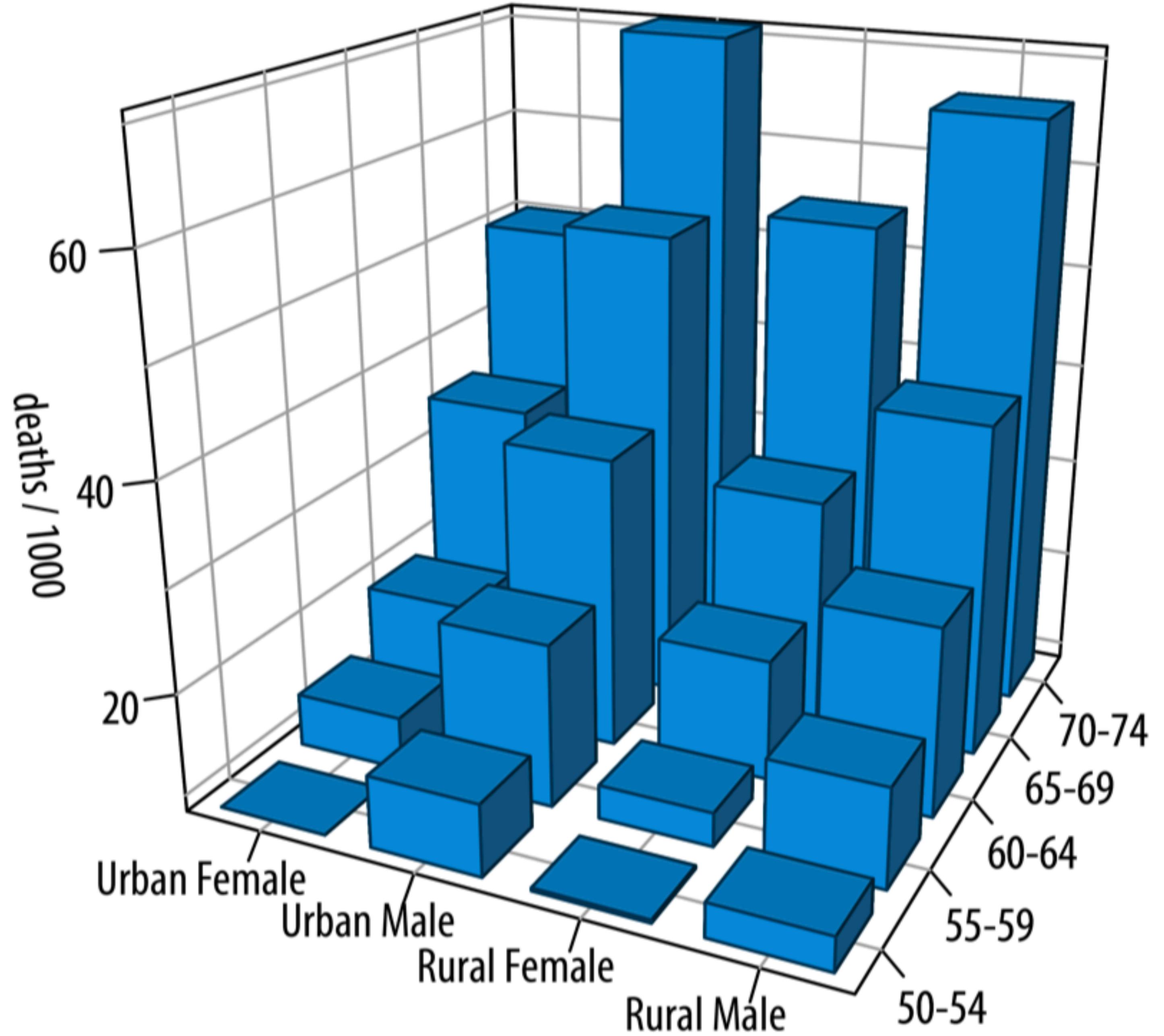
Why Does a Salad Cost More Than a Big Mac?

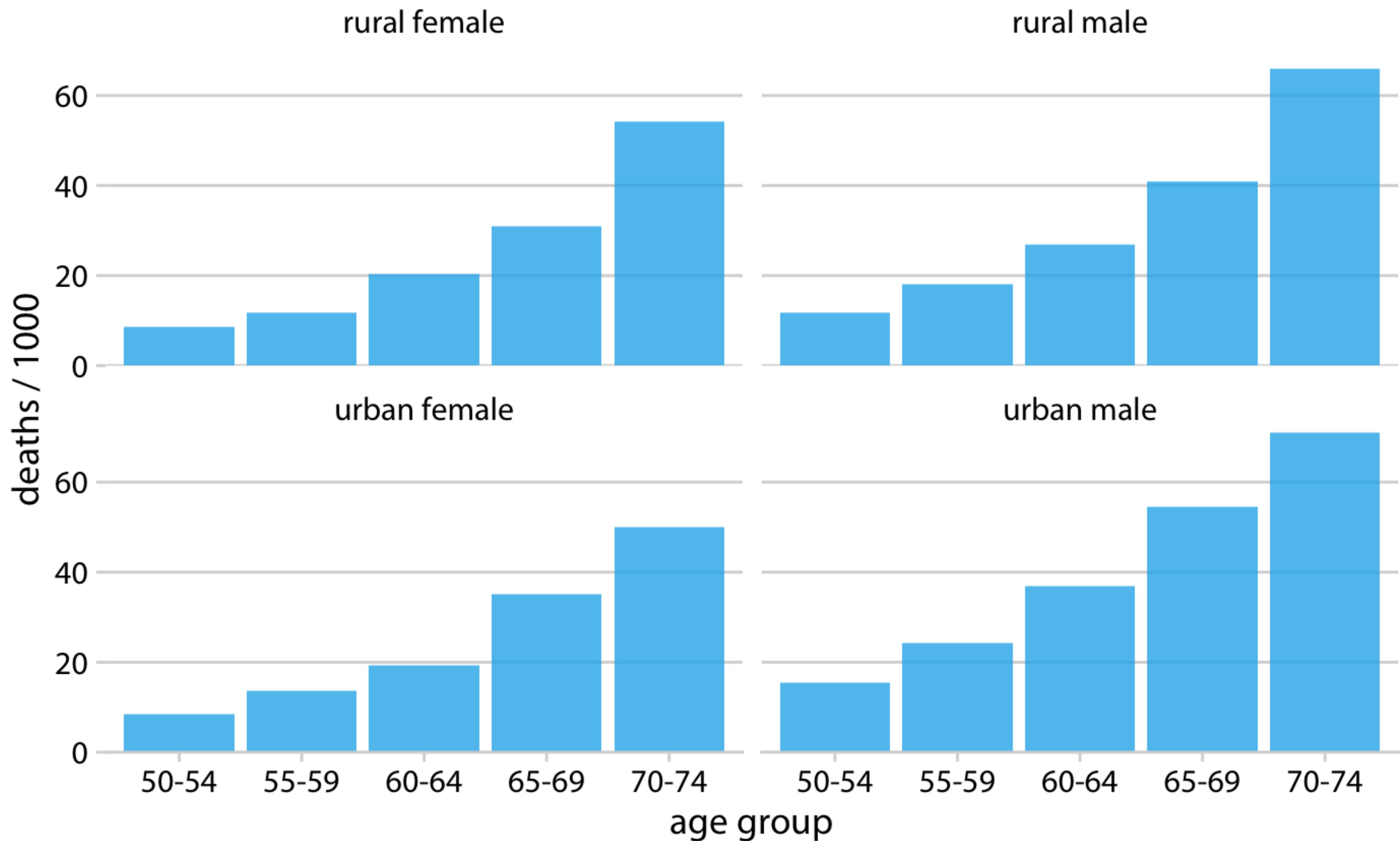
Federal Subsidies for Food Production, 1995-2005[®]



Federal Nutrition Recommendations



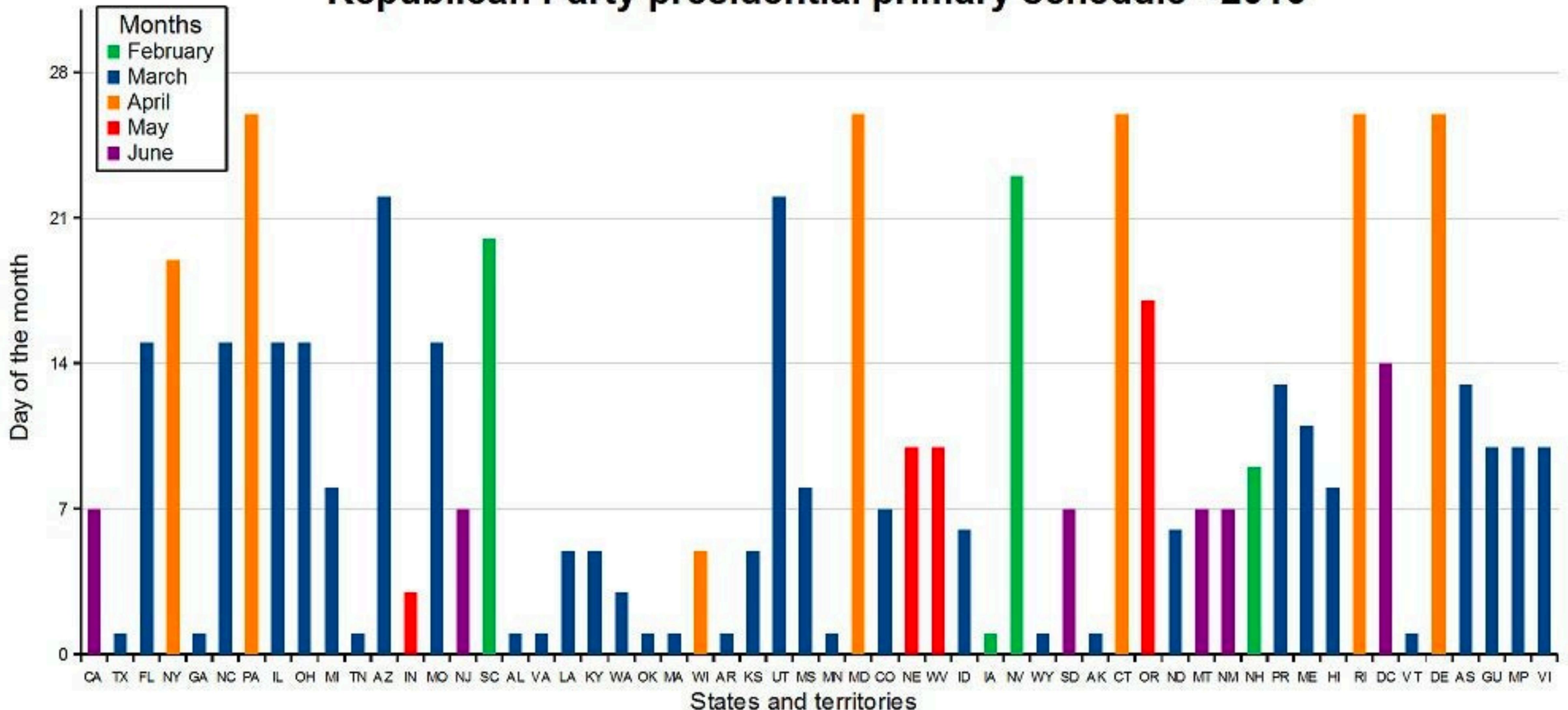




Plotting pitfalls

- Axis trickery
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- **Bad design choices**

Republican Party presidential primary schedule - 2016



May we suggest... a calendar³⁷

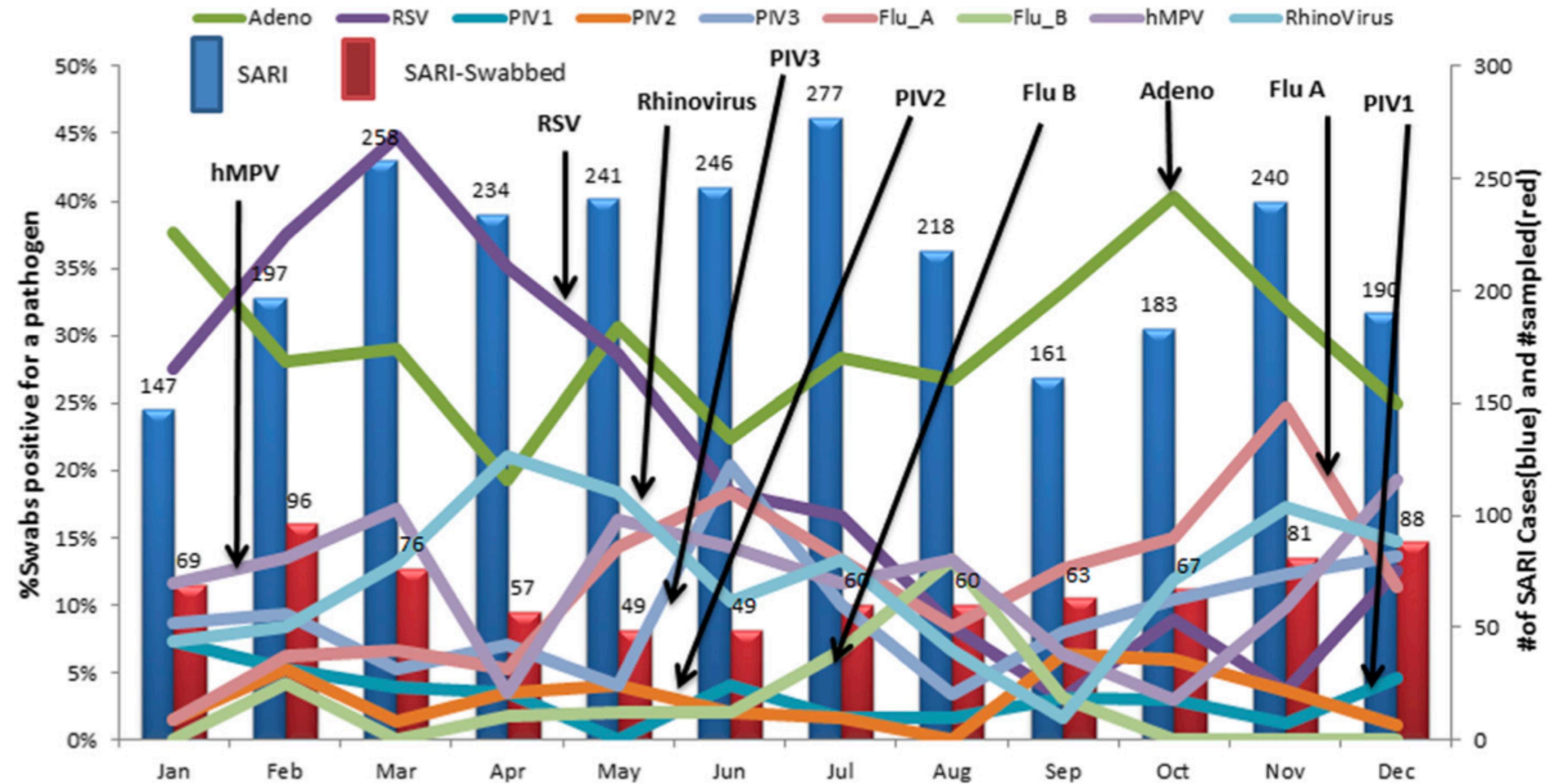
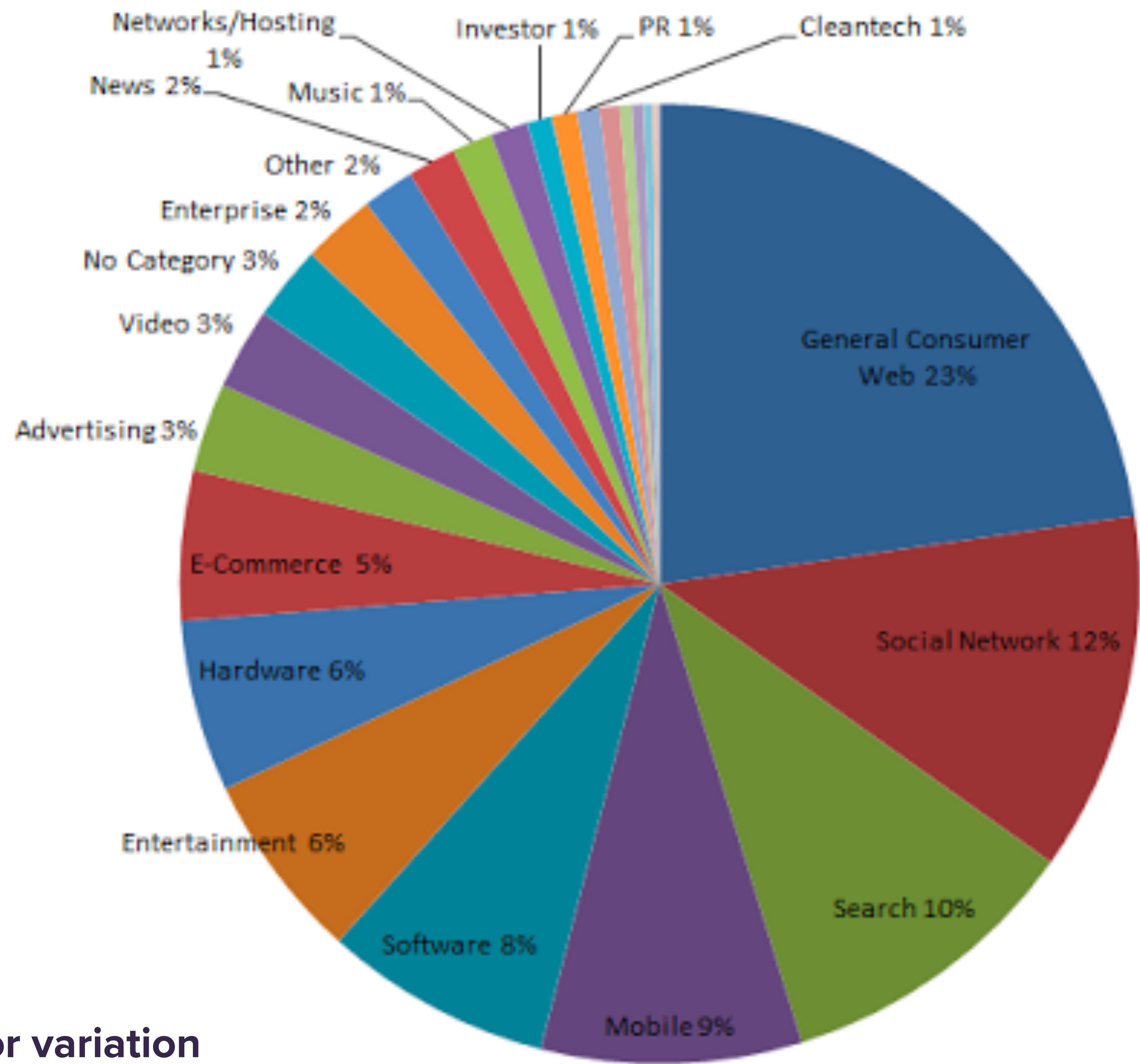
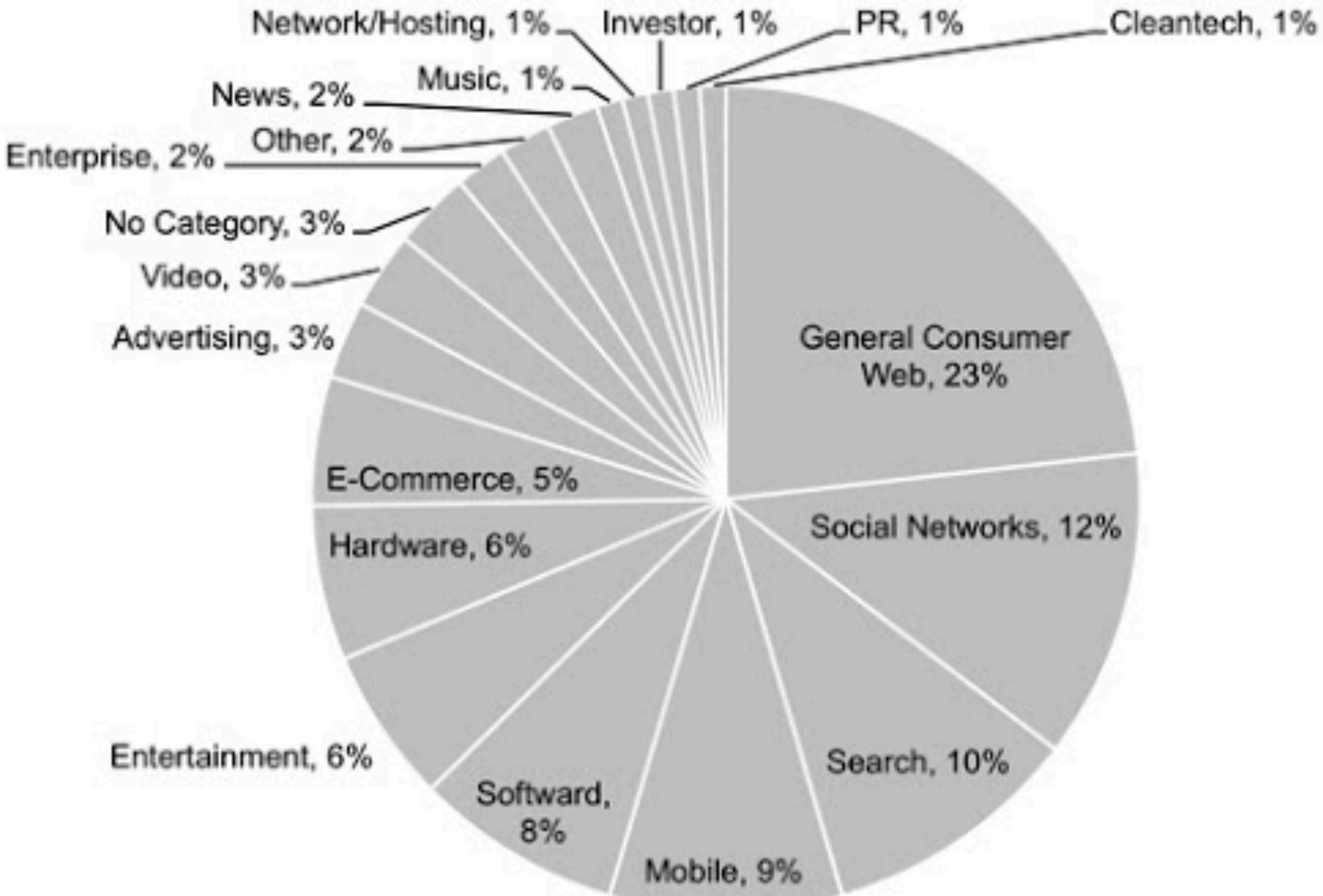
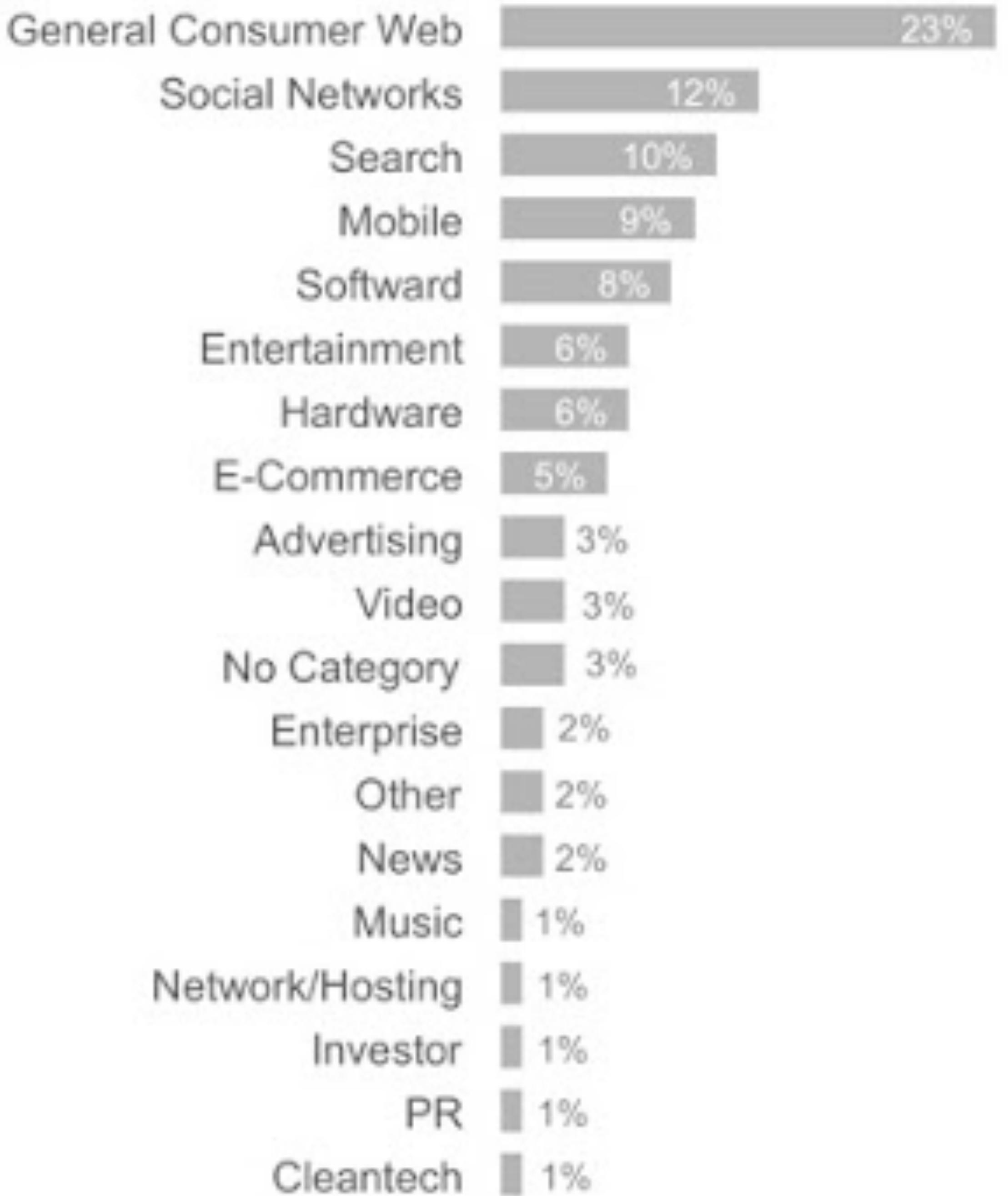


Figure 1 Monthly (aggregate) distribution of viral pathogens, March 1, 2007-Feb 28, 2011.



Meaningless color variation

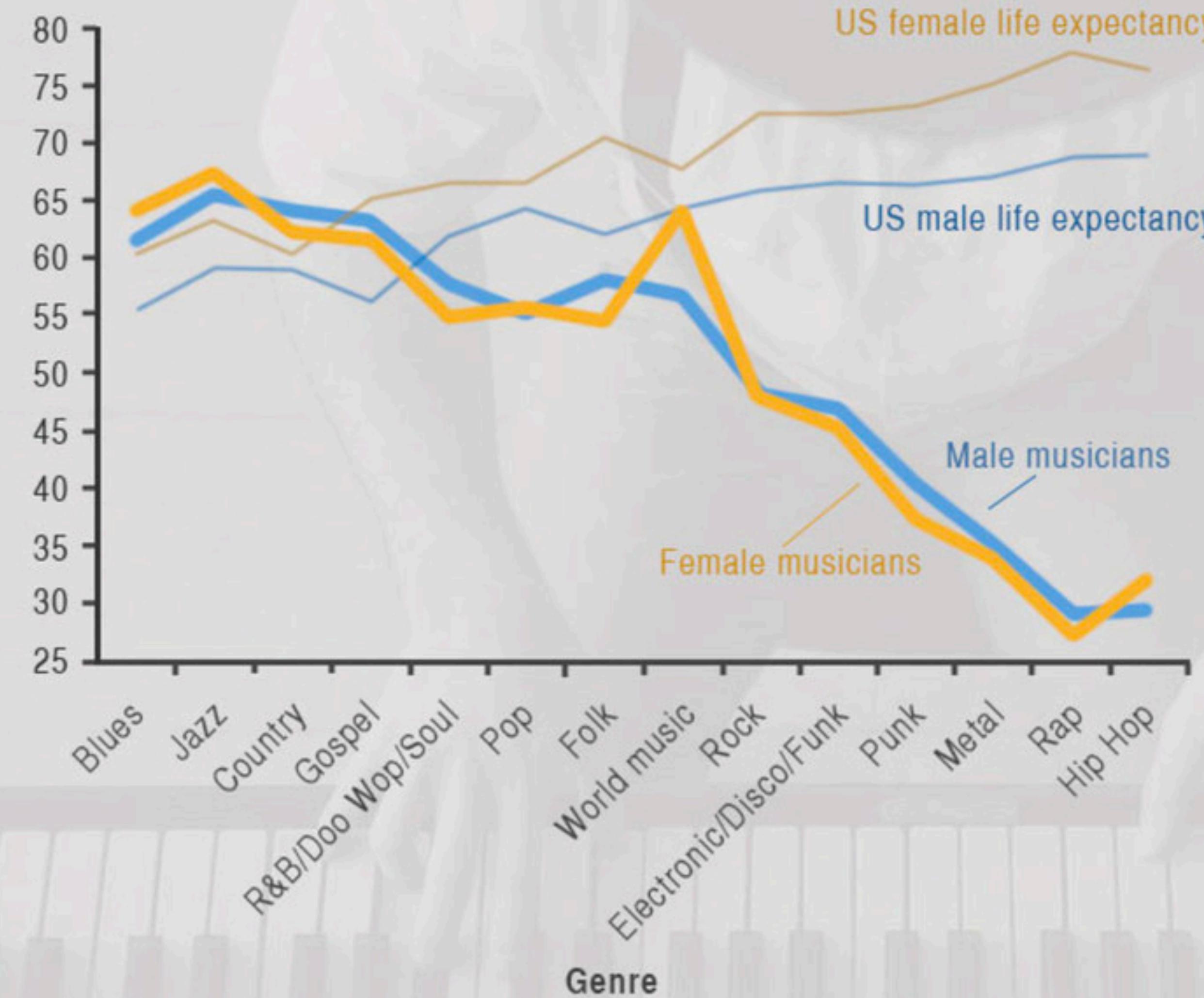




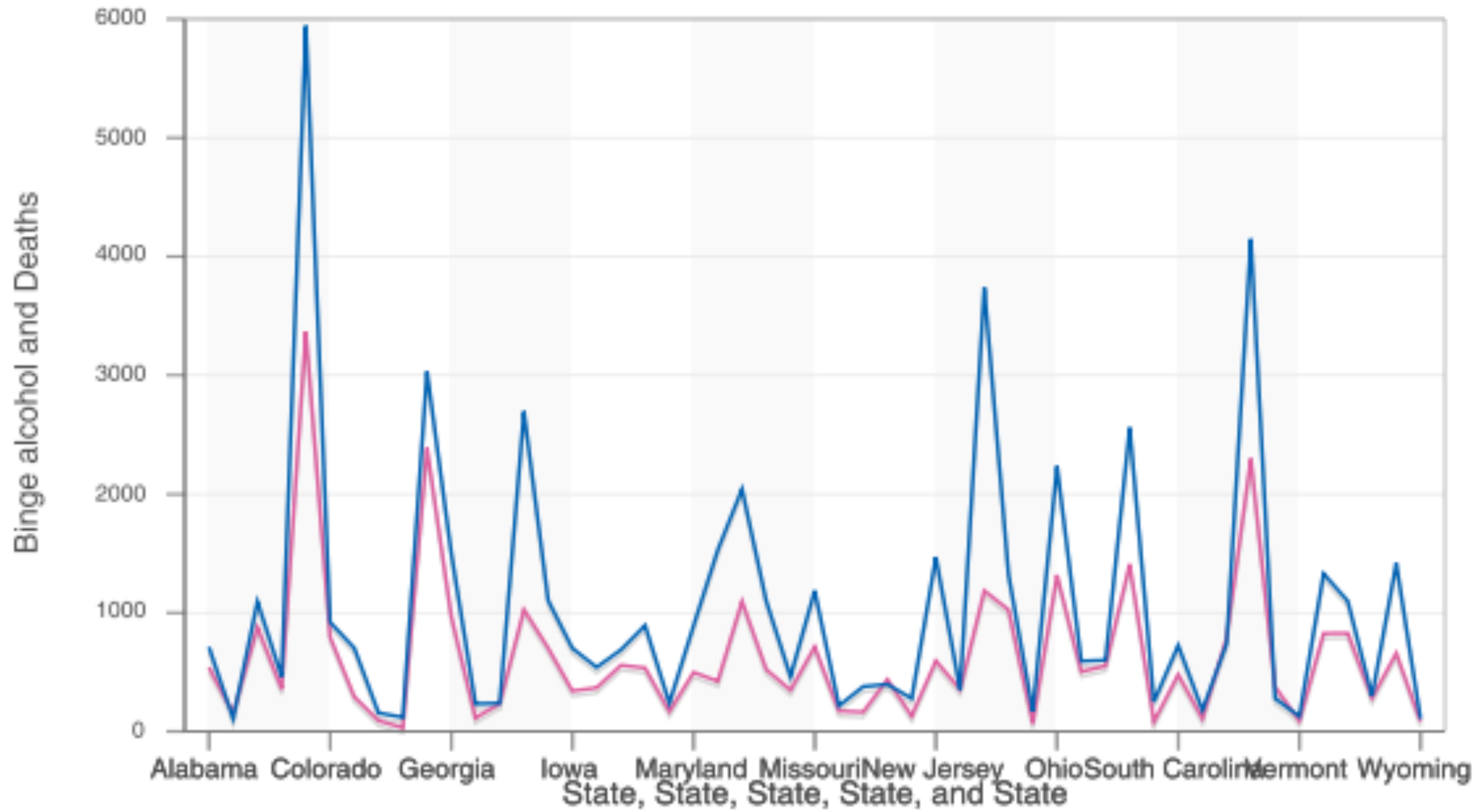
Age of death and musical genre

Average age of death for popular musicians by genre and sex

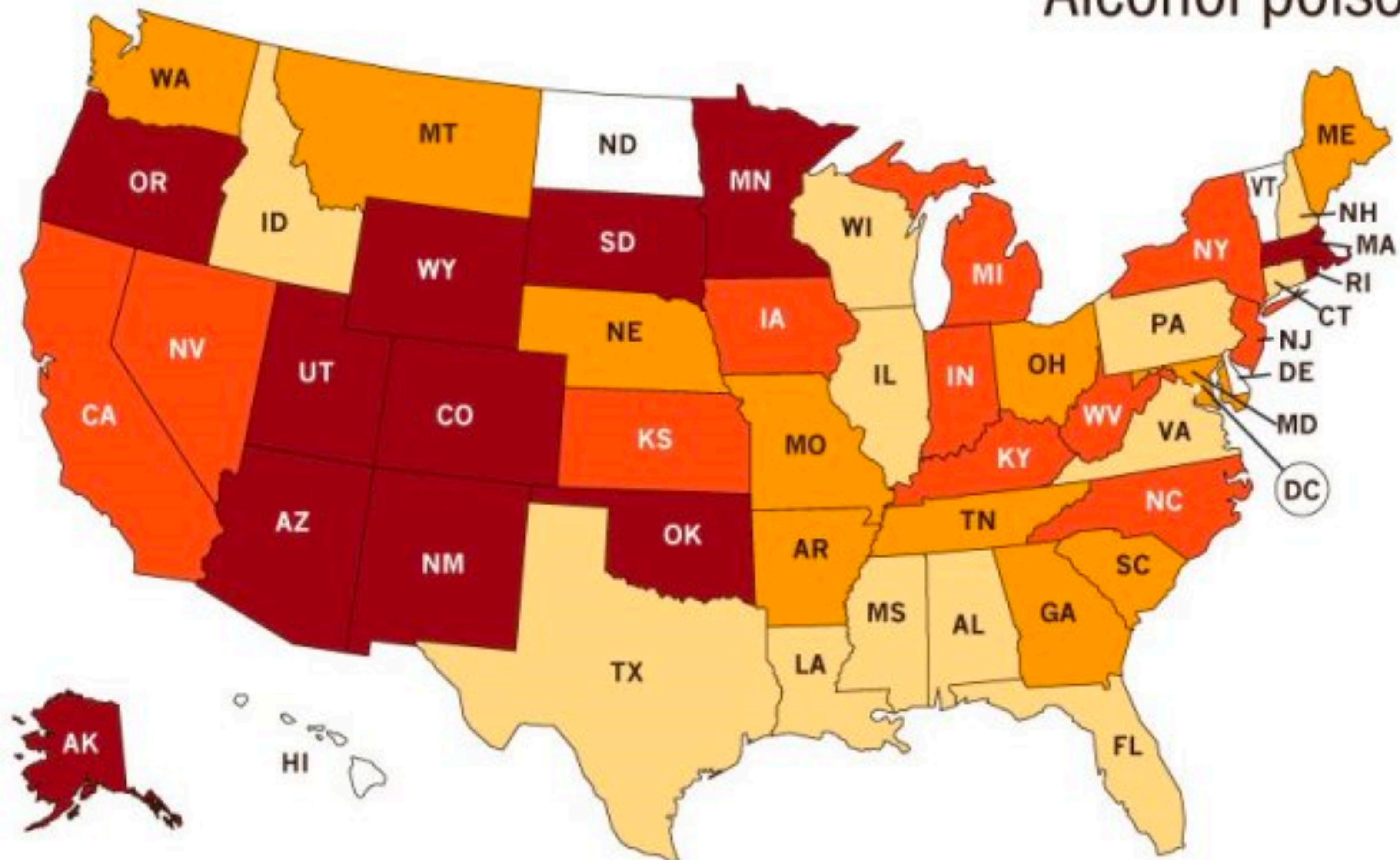
Average age at death



Life expectancy data from: <http://demog.berkeley.edu/~andrew/1918/figure2.html>



Alcohol poisoning deaths vary by state



Deaths by state
per million people

11.9 - 46.5

8.6 - 11.8

6.8 - 8.5

5.3 - 6.7

Not calculated

SOURCE: National Center for Health Statistics Mortality Multiple Cause Files, 2010-2012.

How to do better

Good plots ...

- Make relevant comparisons and avoid irrelevant ones. Are truthful about magnitude.
- Minimize distractions.
- Introduce relevant conditioning variables.
- Stay firmly 2D (unless visualizing an actual 3D object, like a mountain or a building).
- Avoid meaningless variation (e.g. color differences that don't encode information).
- Have clear axis labels and informative titles/annotations.
- Highlight both overall trends and individual variation.
- Sometimes aren't plots at all – they're just tables.
- Are sensitive to the knowledge and limitations of the viewer.

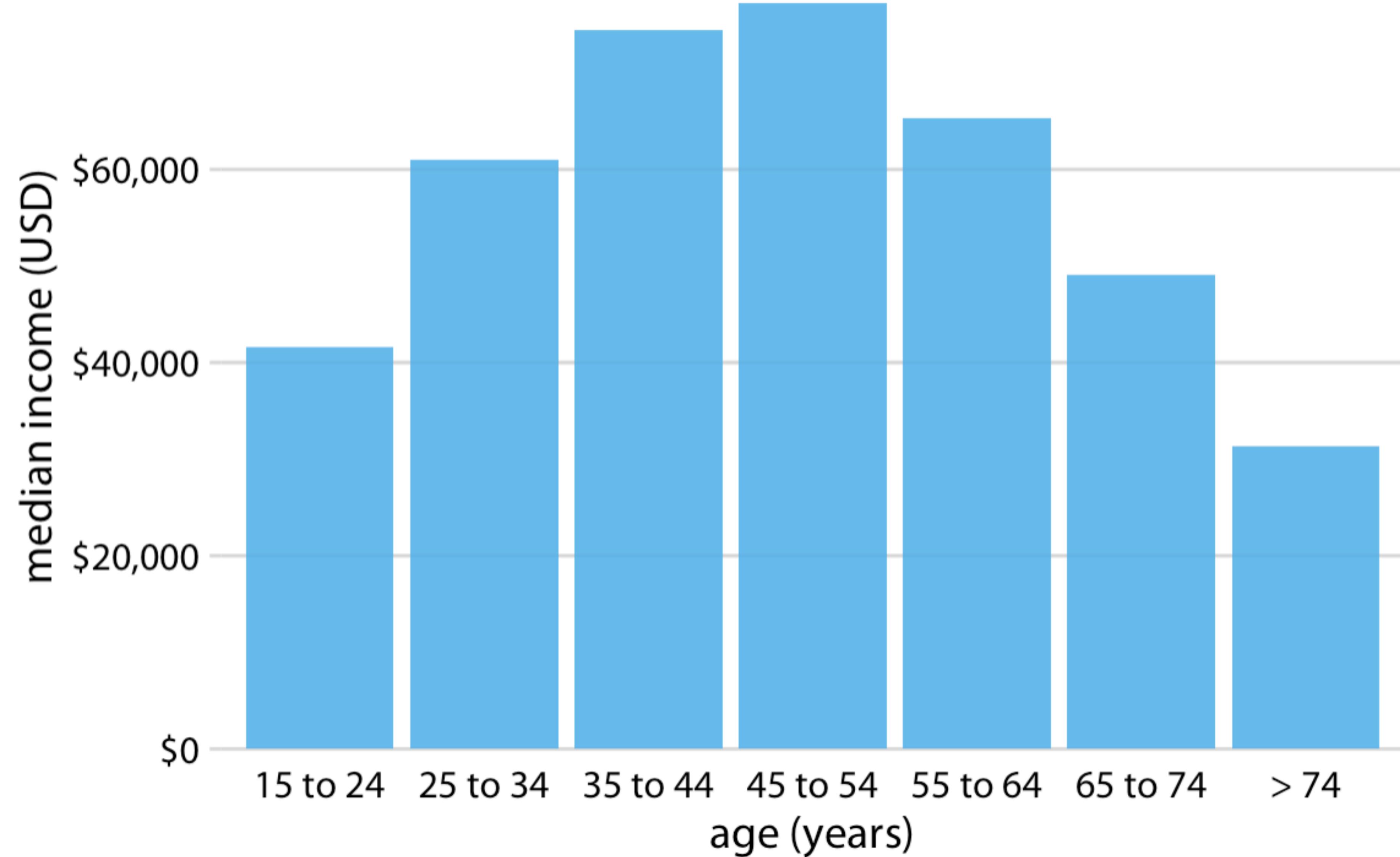
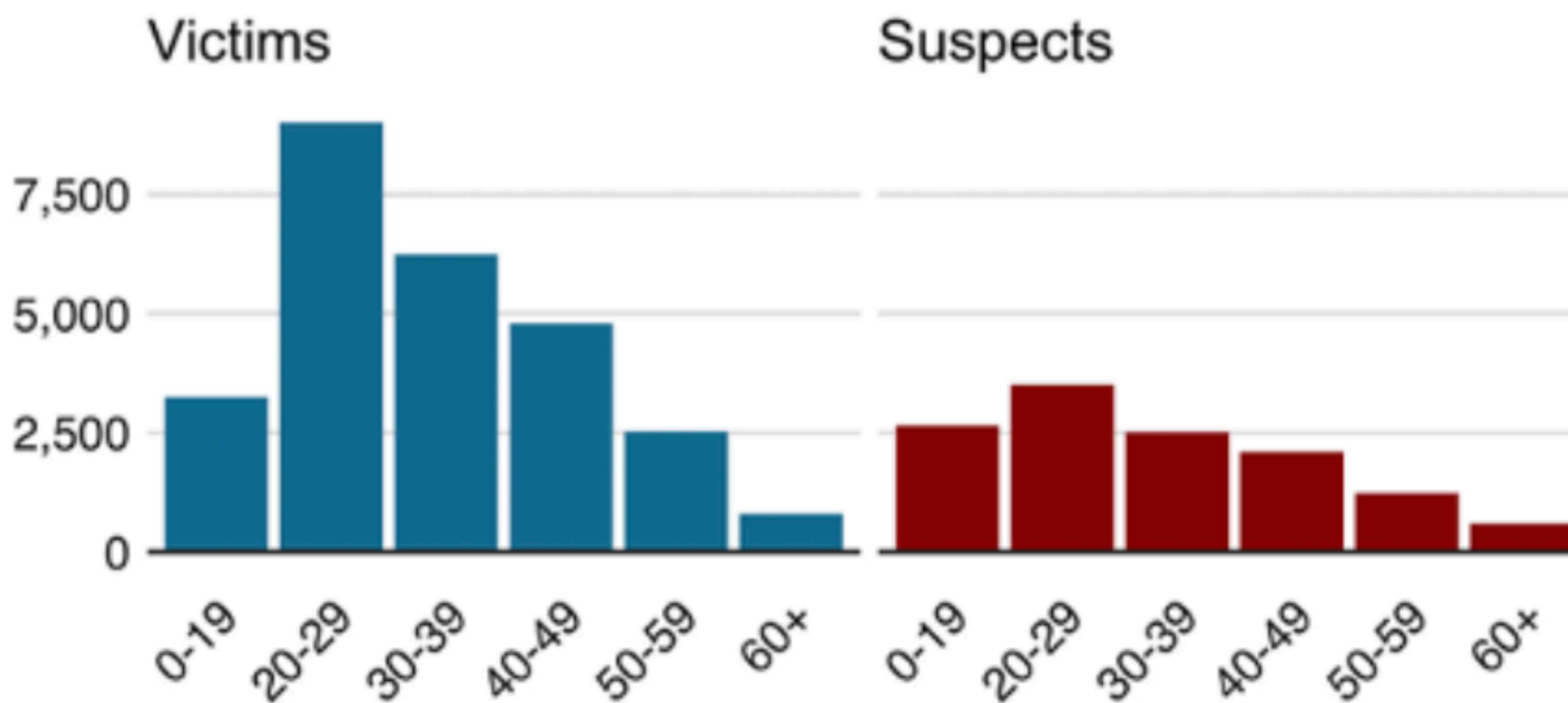


Figure 6.5: 2016 median U.S. annual household income versus age group. The 45–54 year age group has the highest median income. Data source: United States Census Bureau

Homophobic hate crimes are mainly committed by young people on young people

Number in each age group 2014 - 2017



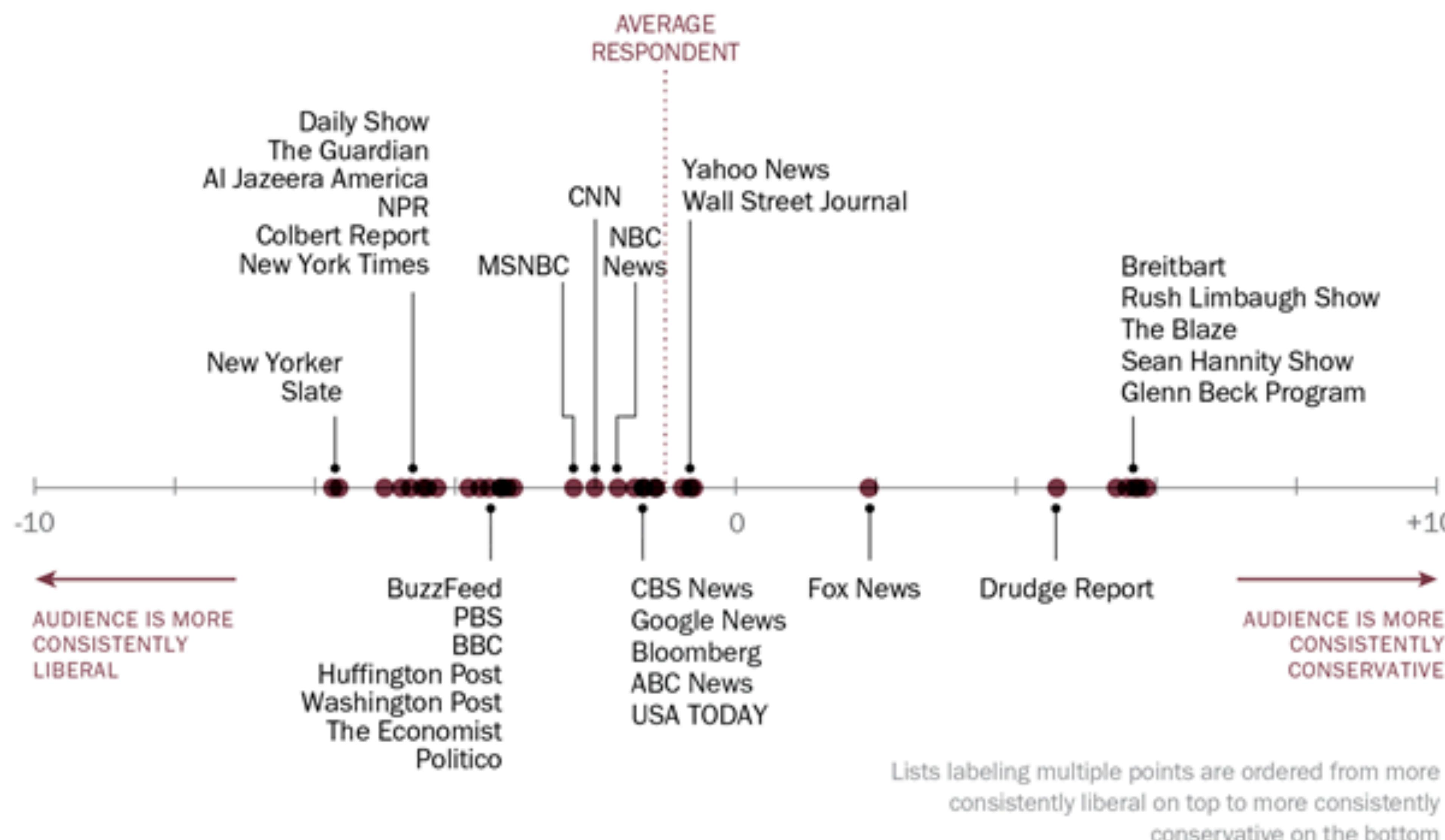
NB. Analysis excludes 3,474 victims and 2,003 suspects whose ages are unknown

Source: BBC Freedom of Information requests to UK police forces

BBC

Ideological Placement of Each Source's Audience

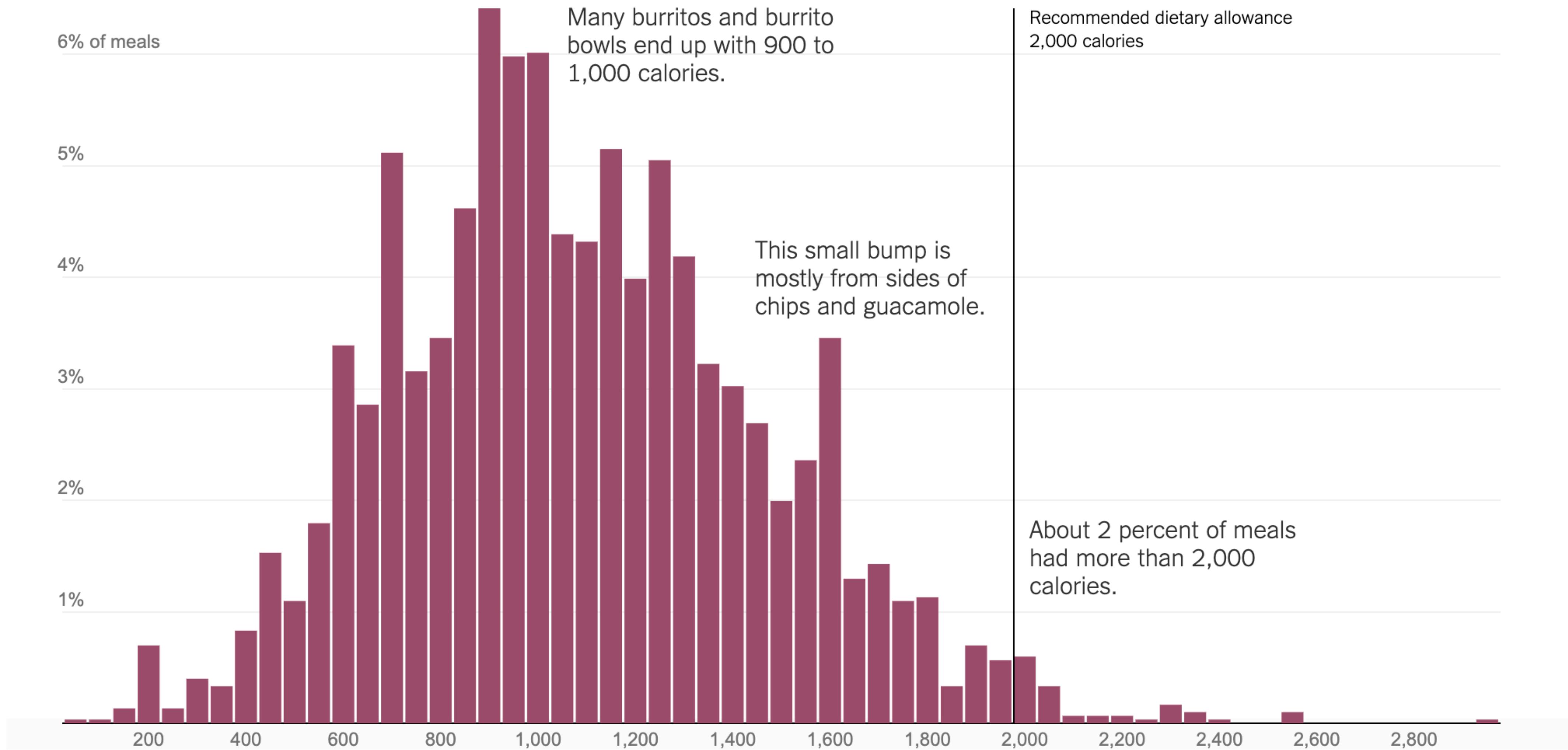
Average ideological placement on a 10-point scale of ideological consistency of those who got news from each source in the past week...



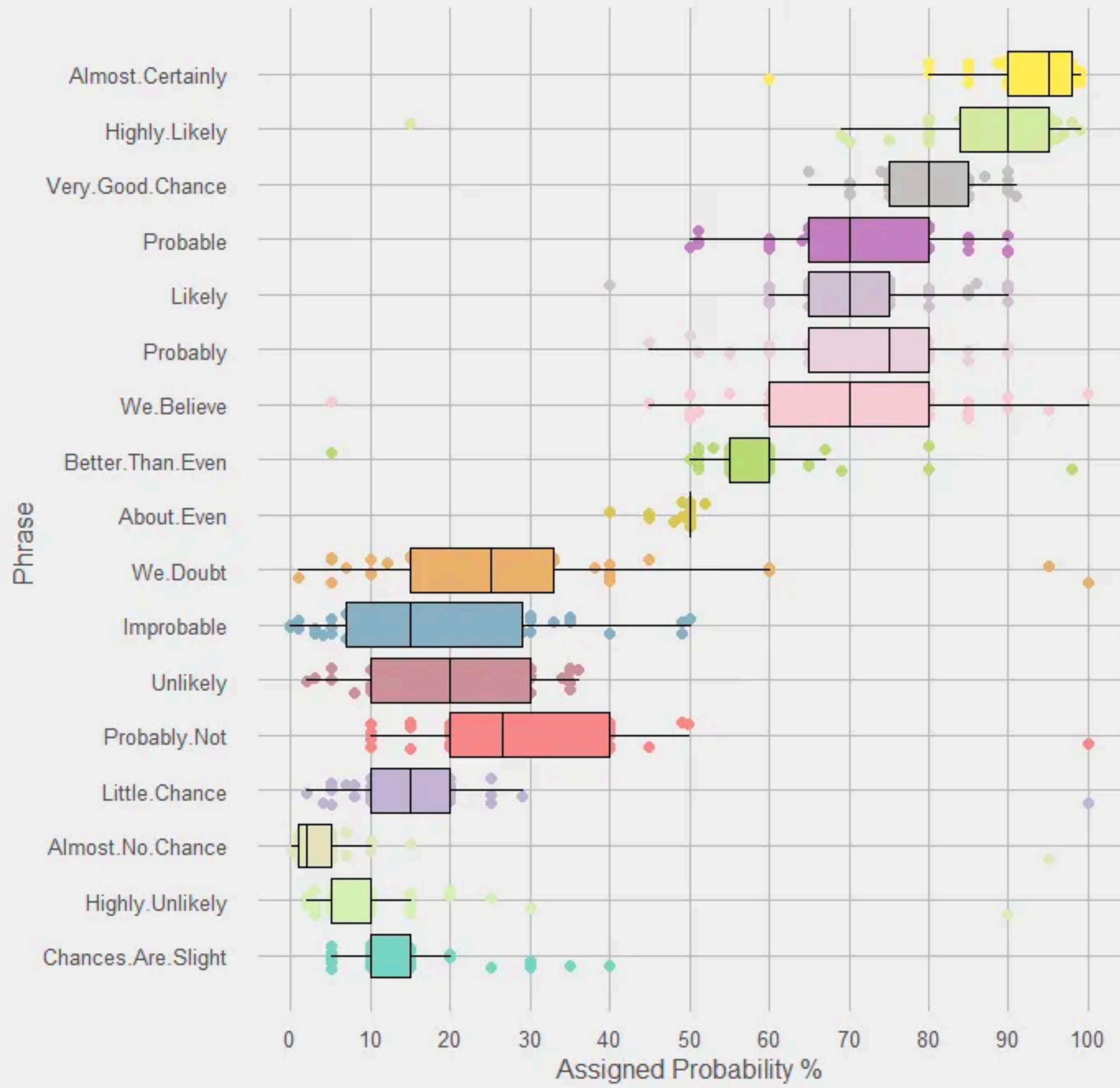
At Chipotle, How Many Calories Do People Really Eat?

By KEVIN QUEALY, AMANDA COX and JOSH KATZ FEB. 17, 2015

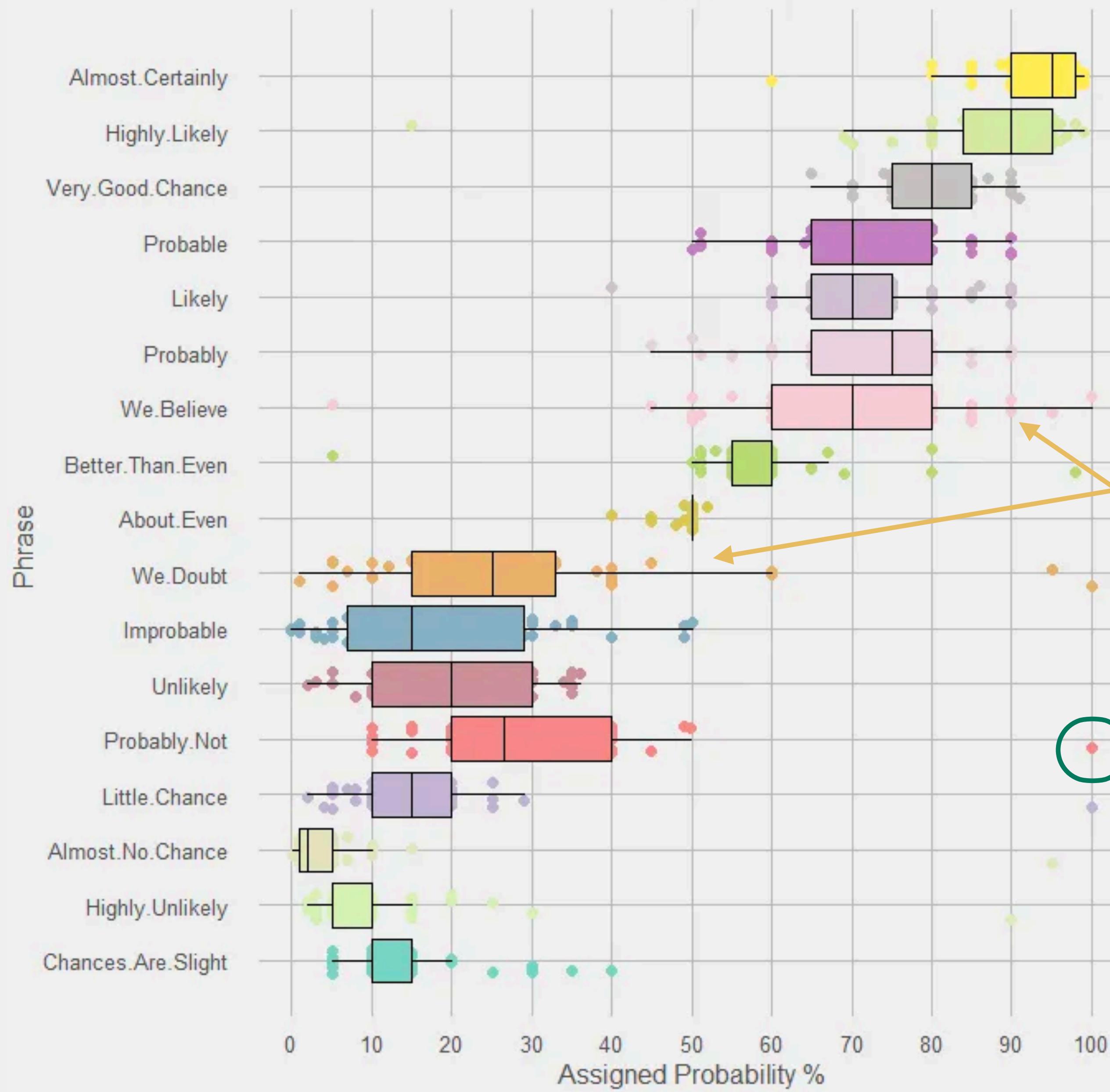
Most meals have more than 1,000 calories and almost a full day's worth of sodium. [RELATED ARTICLE](#)



Perceptions of Probability



Perceptions of Probability



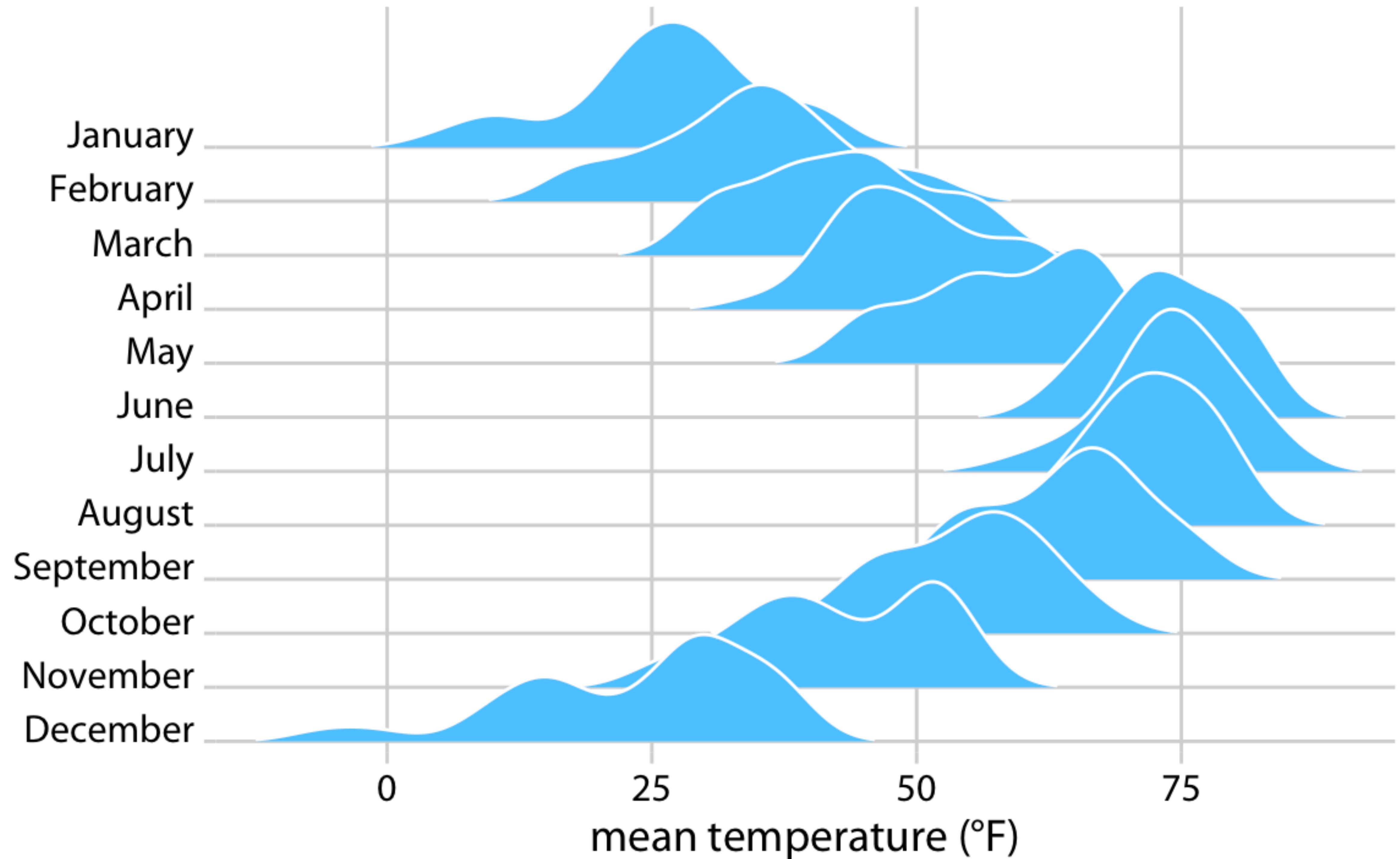
The vertical lines show the median in each group.

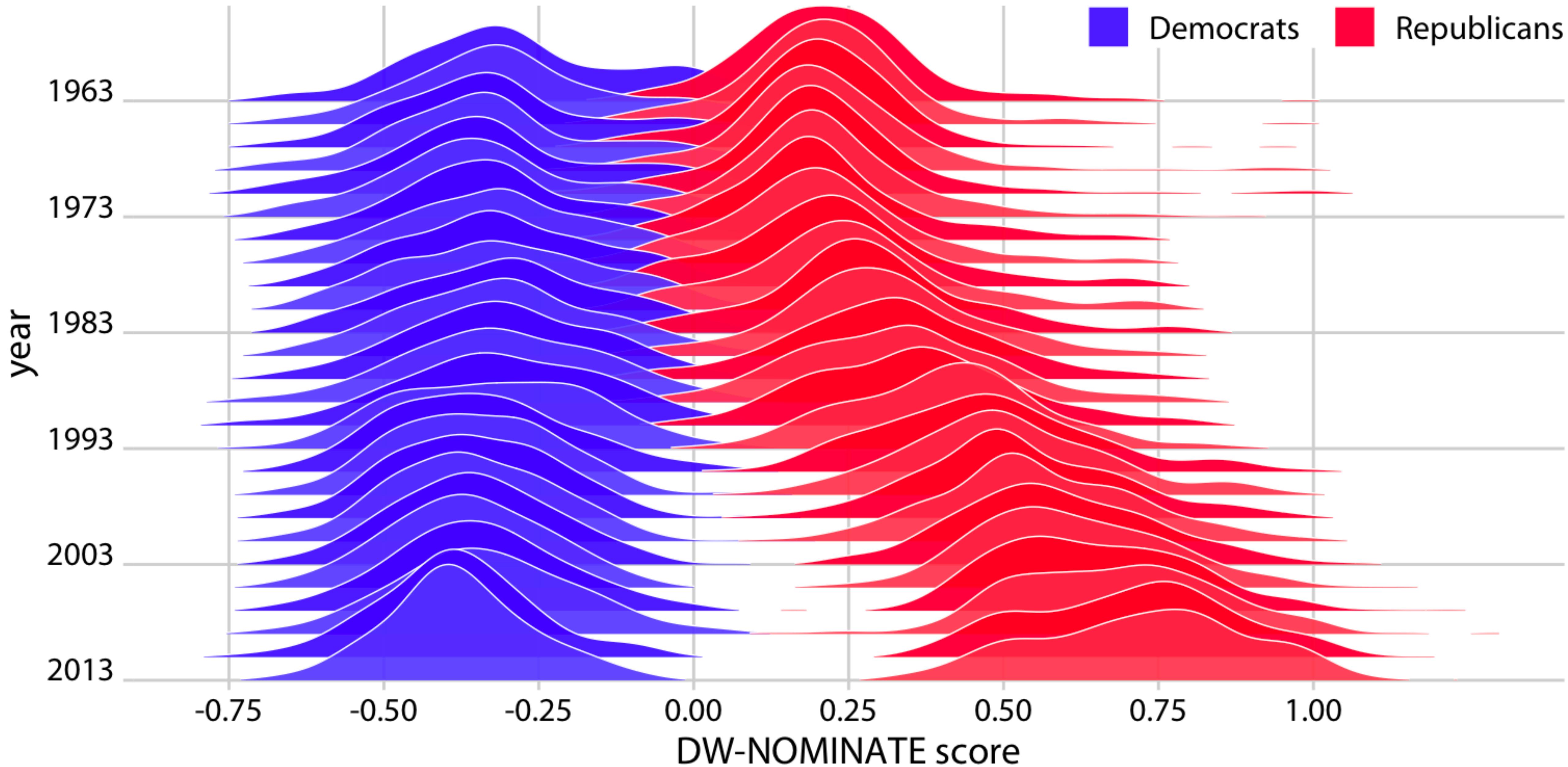
The boxes show the middle 50% of the data in each group.
This is the IQR: the range from the 25th to 75th percentile.

The whiskers typically go out 1.5 times the width of the box.

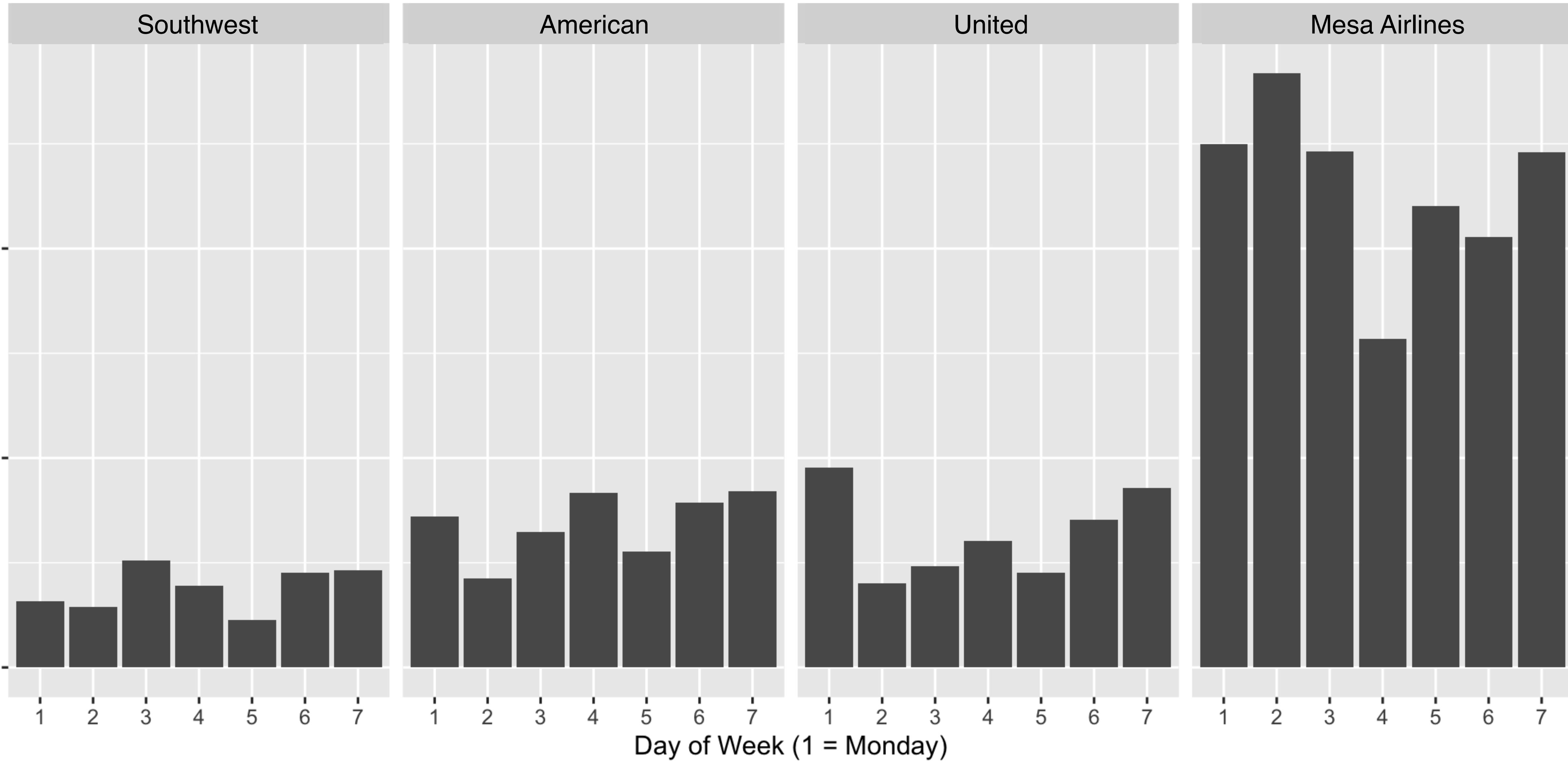
Data points falling outside the whiskers are shown individually, like this one here.

(This plot is a bit unusual in that all points falling outside the box, not just the whiskers, are shown individually.)

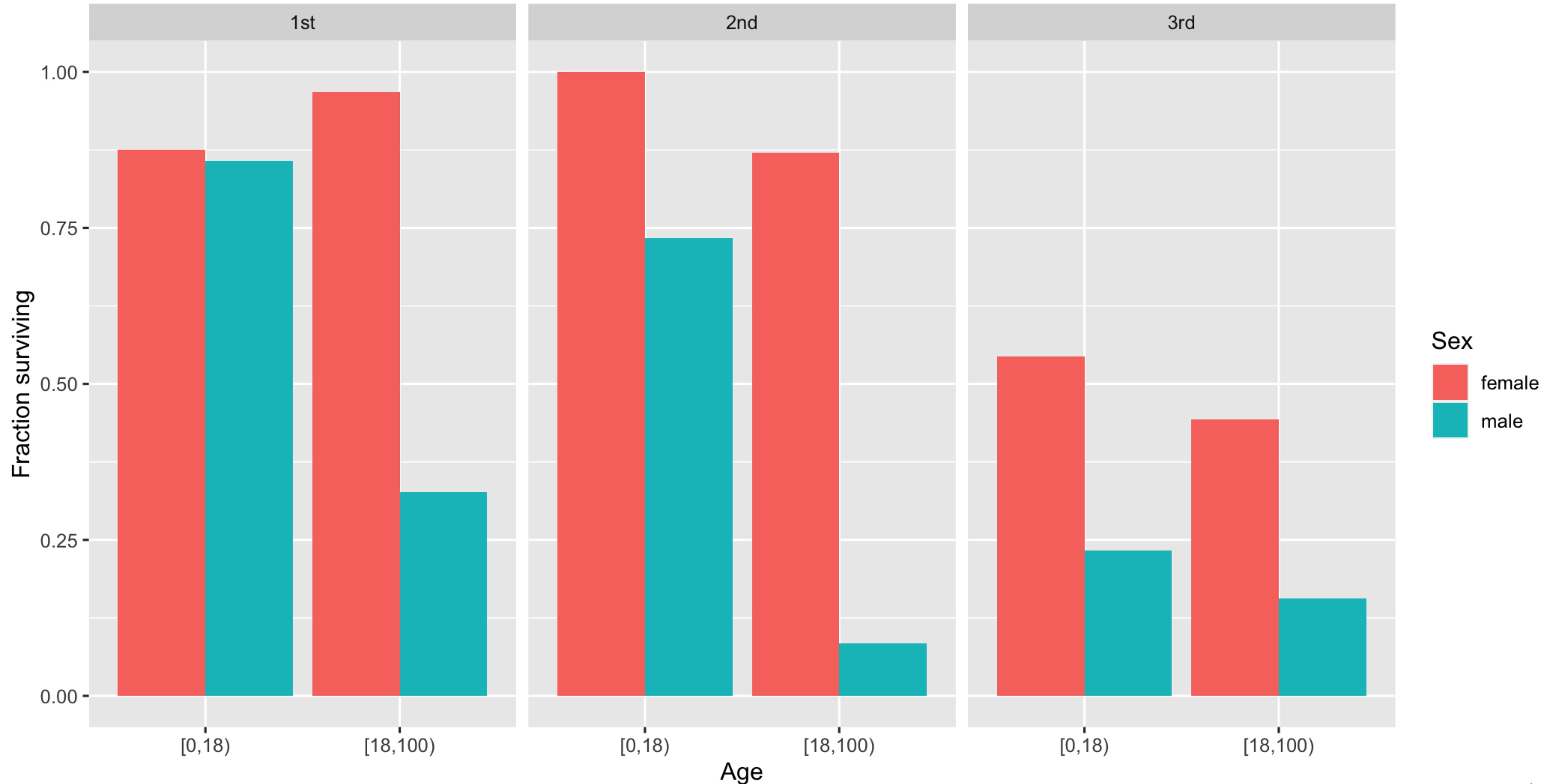




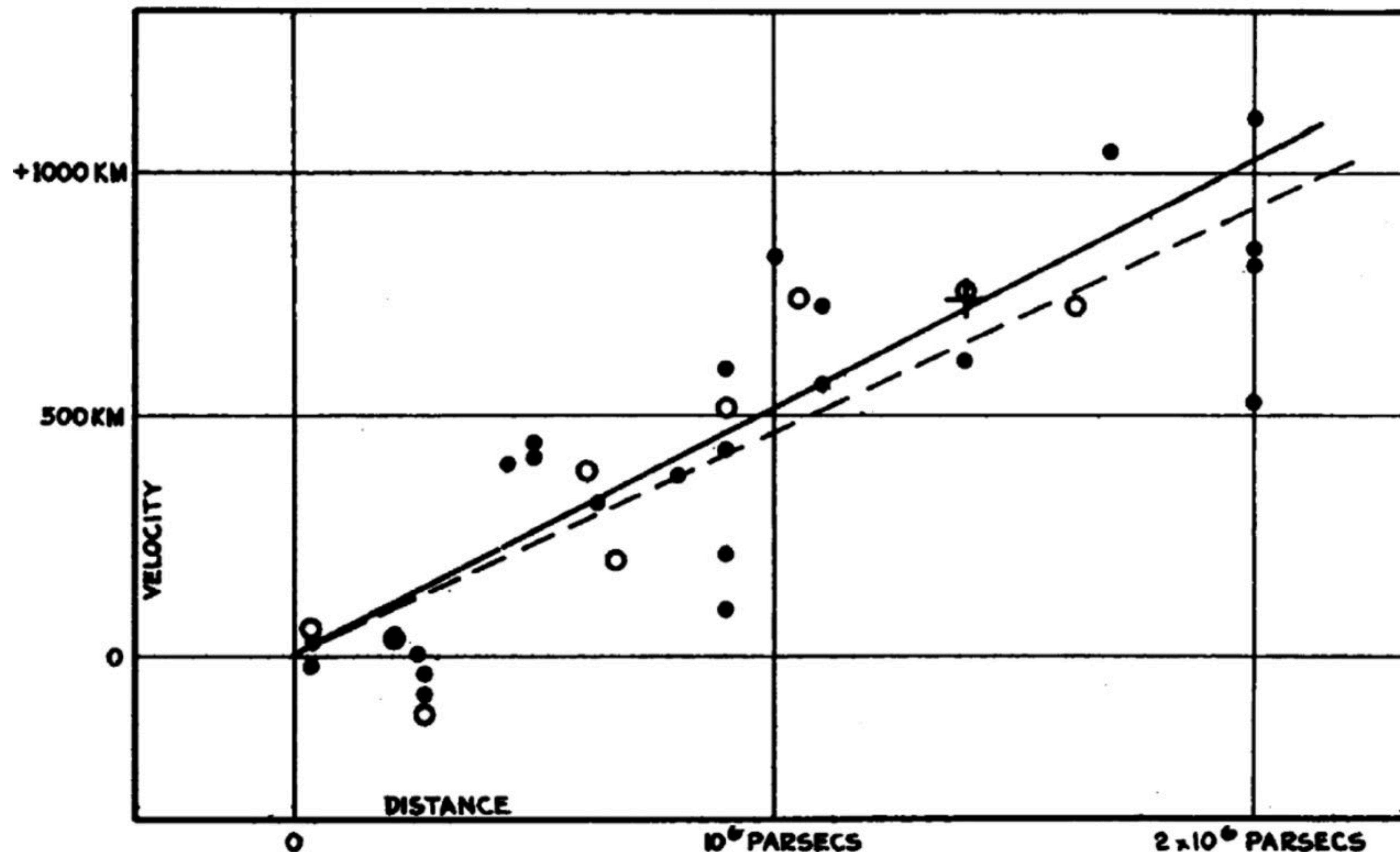
ABIA Delays by Airline and Day of Week (2011)



Survival on the Titanic by Age, Sex, and Passenger Class

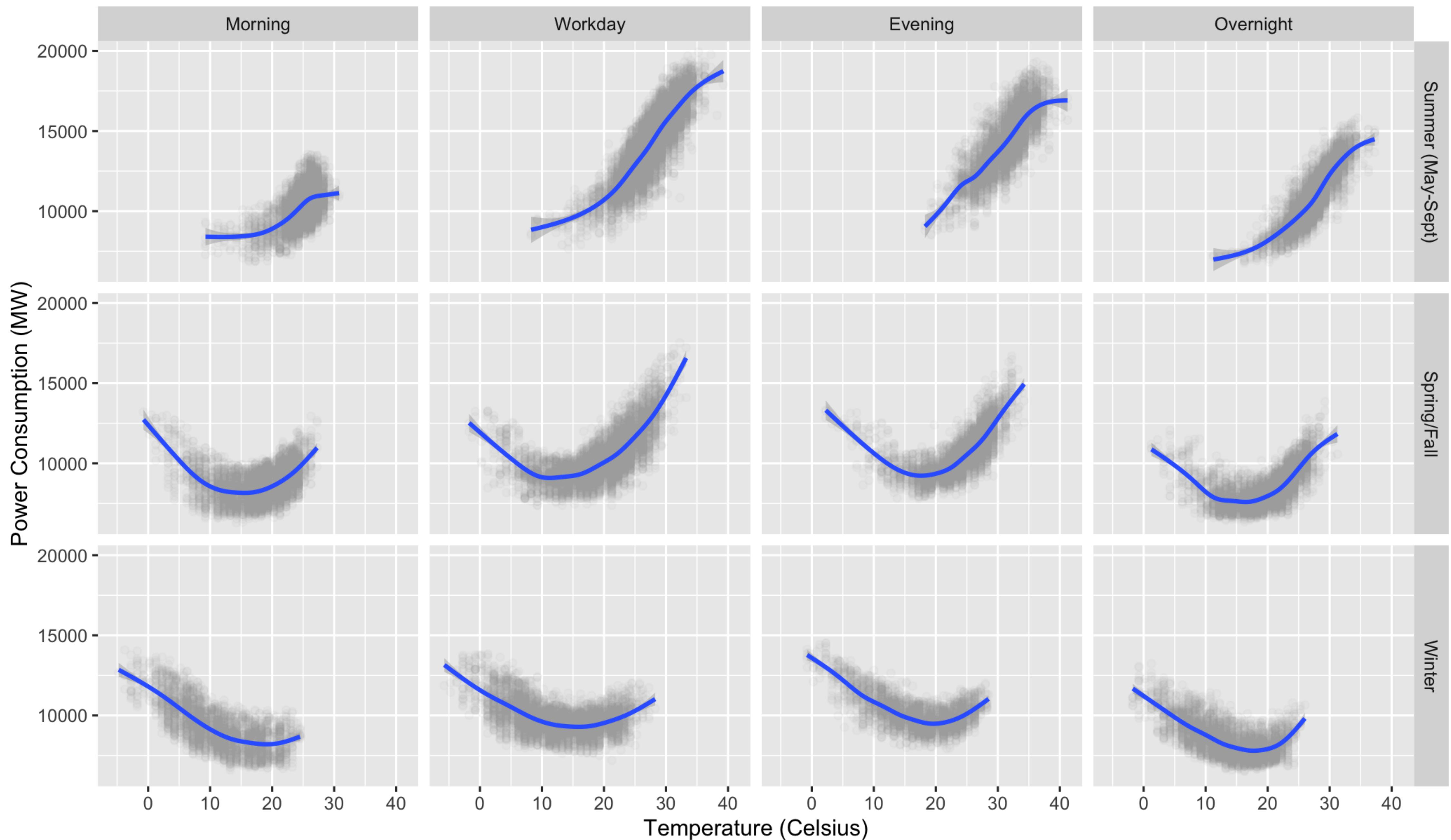


Velocity-Distance Relation among Extra-Galactic Nebulae.



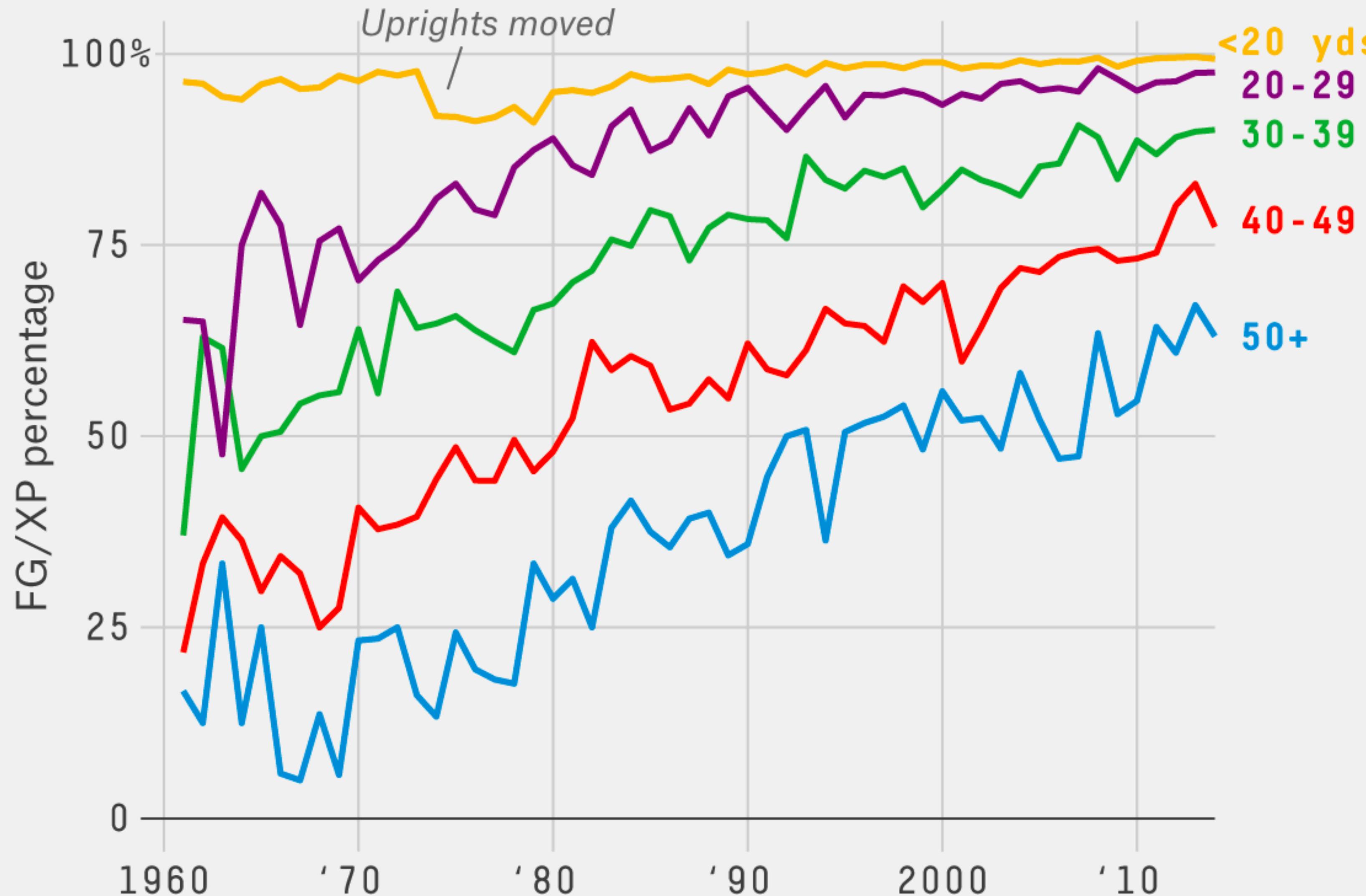
How Edwin Hubble proved that the universe was expanding

Power Consumption vs Temperature: Houston Metro Area

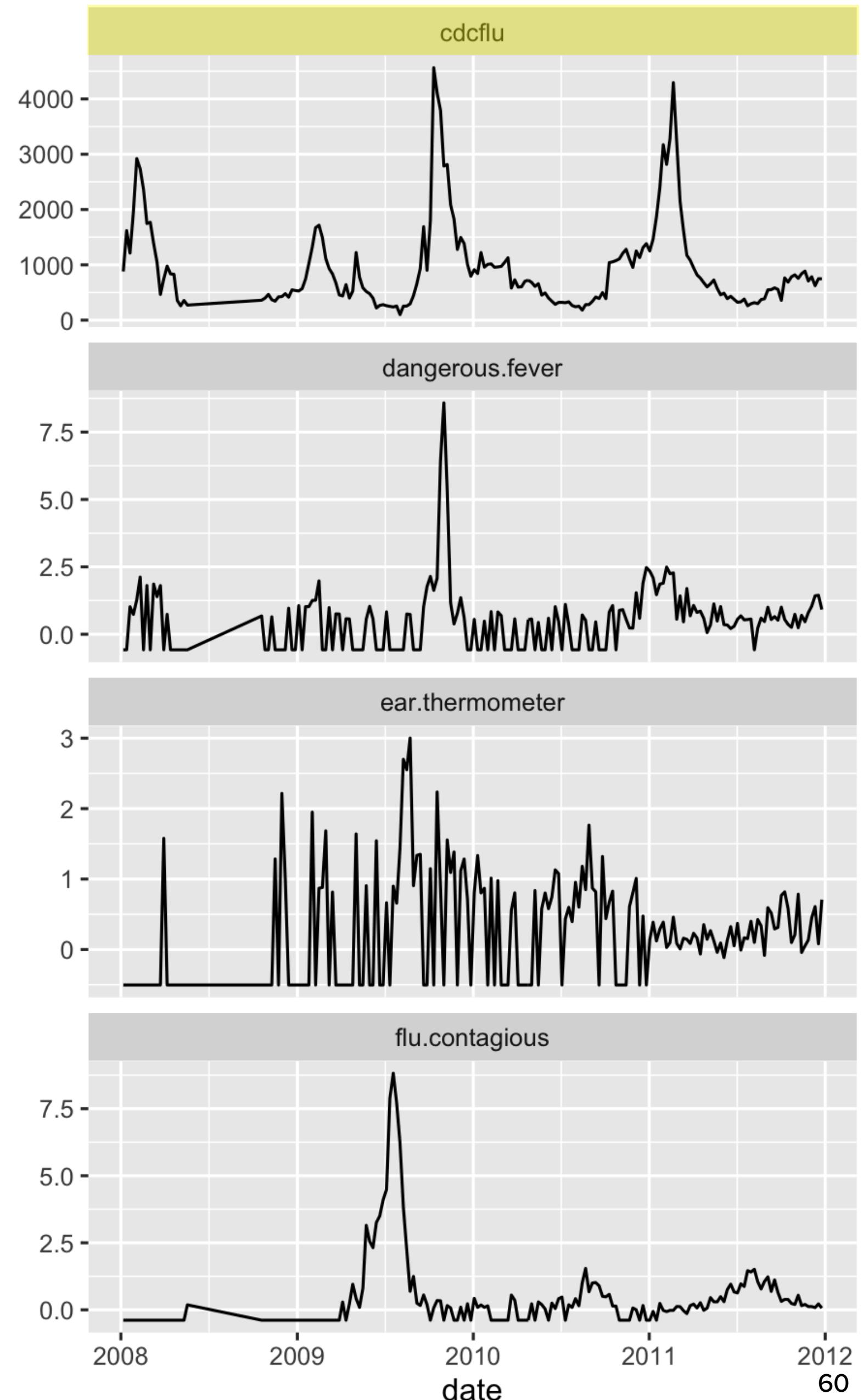
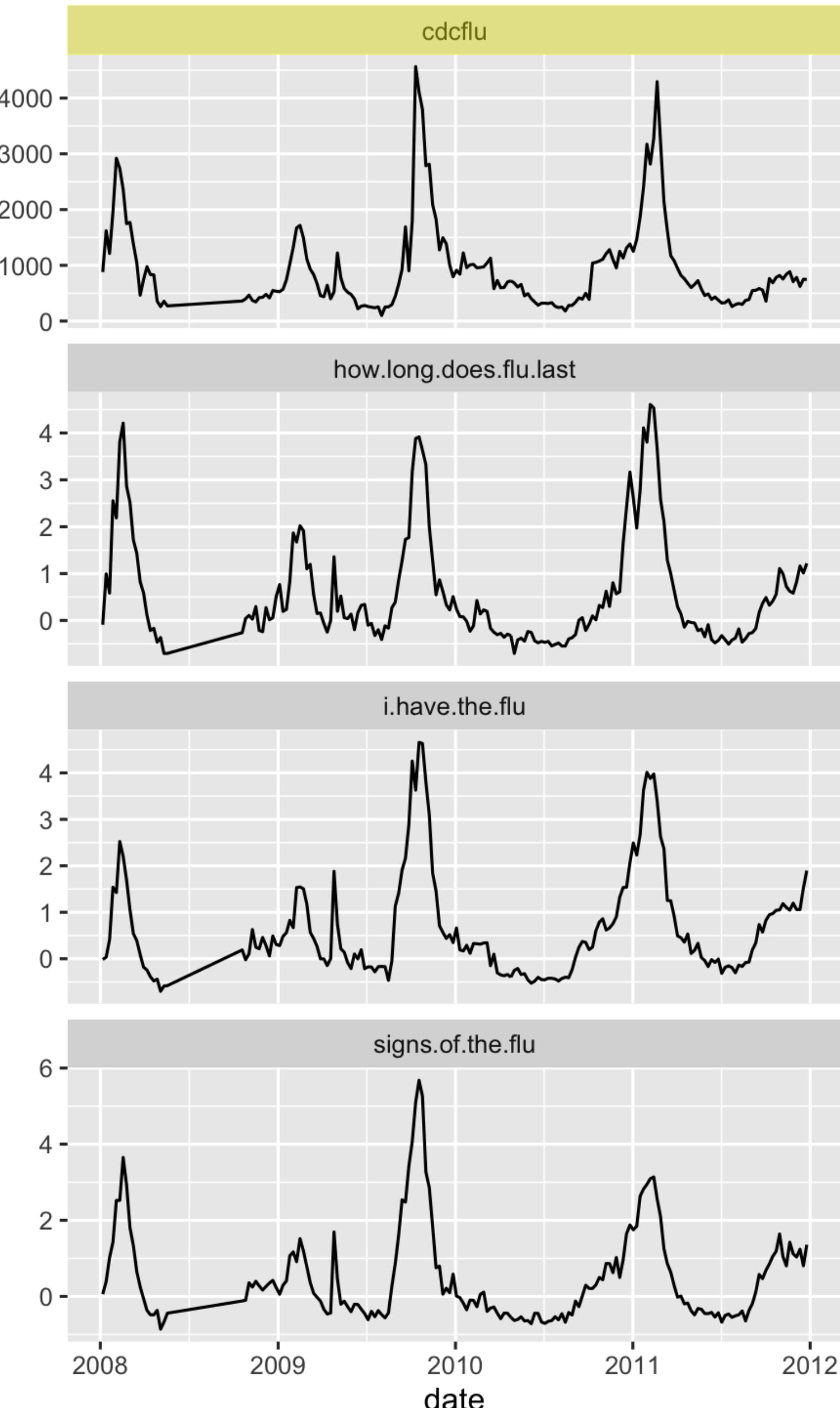
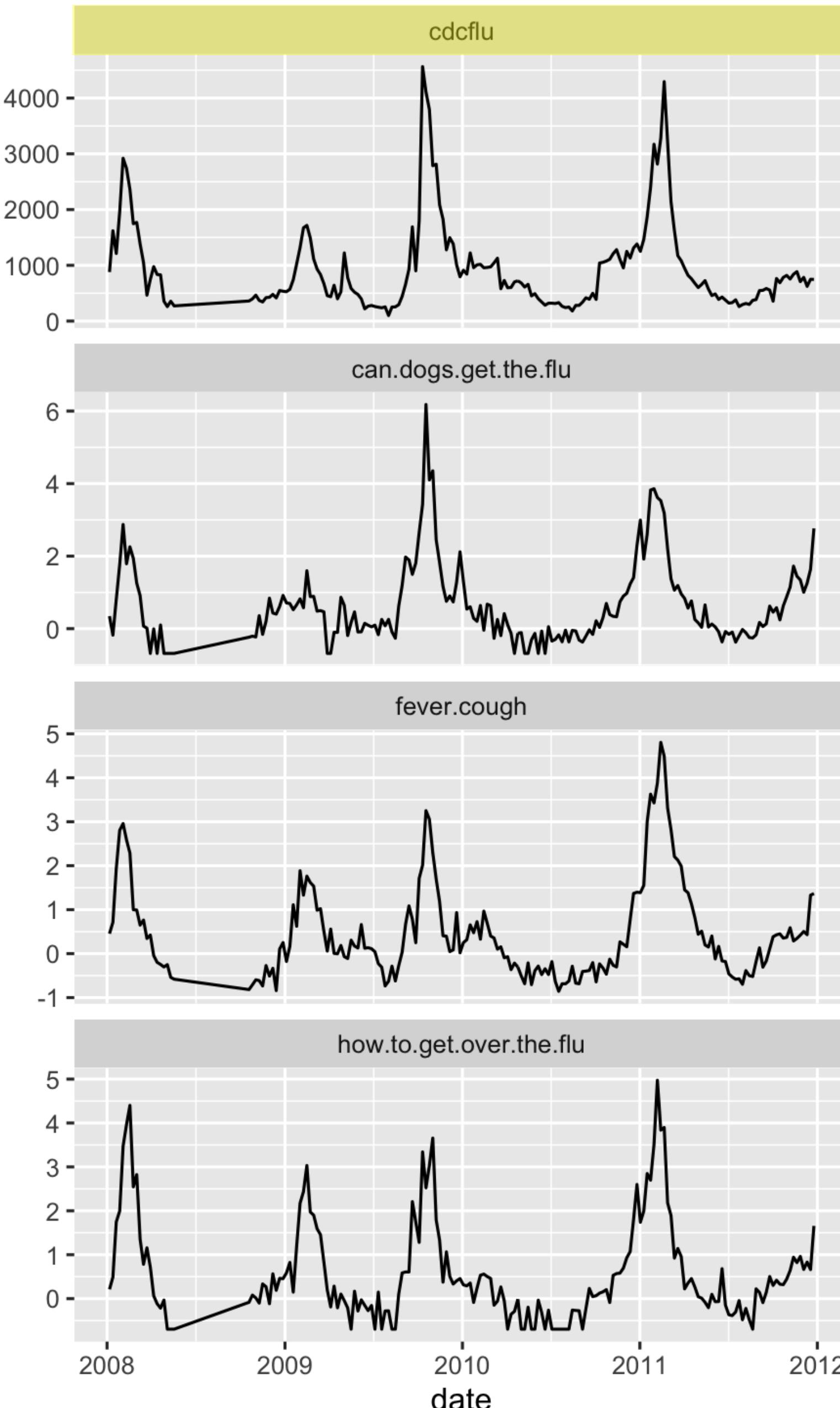


NFL Kickers Have Improved At Every Distance

Field-goal and extra-point accuracy, 1961-2014



Some “obvious” Google search terms track CDC flu data well...



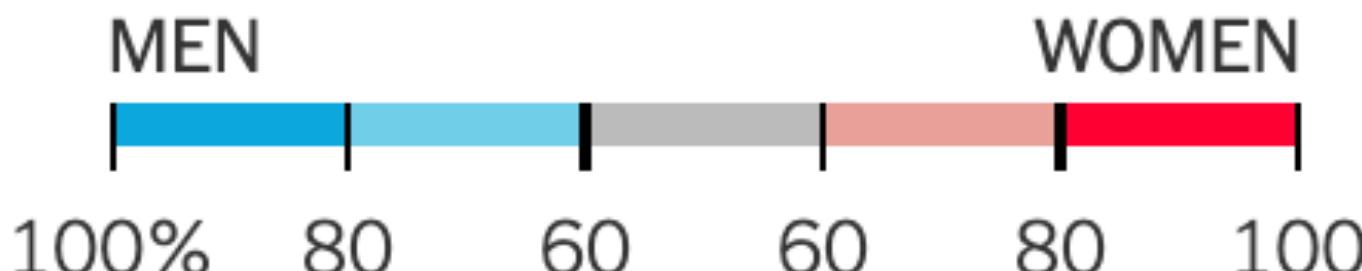
...and others don't.

Jobs per 1,000 middle-class jobs

Top 20 occupations that gained share

	1980	2012	Change
Registered nurses	19	39	+20
Mathematical and computer scientists	6	25	+19
Supervisors and proprietors of sales jobs	18	34	+16
Adjusters and investigators	6	22	+15
Managers and administrators	70	84	+14
Health technologists and technicians	8	21	+13
Other teachers and counselors	7	19	+12
Elementary school teachers	41	51	+9
Health therapists	3	11	+8
Accountants and auditors	15	23	+8
Health assistants	4	11	+7
Writers, artists, entertainers and athletes	11	18	+7
Legal assistants and aides	1	6	+6
Personal service occupations	3	8	+5
Police and detectives	14	18	+4
Management analysts	1	6	+4
Financial managers	6	10	+4
Food preparation and service occupations	6	10	+4
Financial specialists	5	9	+3
Personnel and human resources specialists	7	9	+3

Gender breakdown of each occupation



Jobs per 1,000 middle-class jobs

Top 20 occupations that lost share

	1980	2012	Change
Machine operators and tenders, except precision	51	18	-33
Production supervisors or foremen	33	11	-23
Construction trades, except electricians	38	24	-15
Freight, stock and material handlers	25	11	-14
Assemblers of electrical equipment	15	6	-9
Precision metal working occupations	15	6	-9
Vehicle and mobile equipment mechanics and repairers	31	22	-9
Truck, delivery and tractor drivers	39	31	-8
Transportation occupations	14	7	-8
Material recording, scheduling and distributing clerks	23	16	-7
Graders and sorters in manufacturing	13	6	-7
Welders and metal cutters	11	5	-6
Engineering and related technologists and technicians	17	10	-6
Sales representatives, commodities	43	38	-5
Engineers	20	15	-5
Mail and message distributing occupations	15	10	-5
High school teachers	17	12	-5
Printing machine operators	6	2	-4
Precision workers, except metal	10	7	-3
Janitors and other commercial building services	14	11	-3

Start with what's simple and effective

Simple and effective plots:

- Density plots (histograms, boxplots, violin plots)
- Barplots: to show summary statistics (means/counts/proportions) by group
- Scatter plots: to show how two numerical variables are related
- Line graphs: to show how one numerical variable changes continuously versus another (e.g. time)

Simple and effective strategies for enriching these plots:

- Variation in size, shape, and/or color (but be colorblind-friendly and don't go crazy)
- Faceting: showing the same basic plot across multiple conditions
- Labels