

Graphical Unexcellence + Integrity

Data Viz Hall of Shame

(Professor Dave)²

The University of Austin

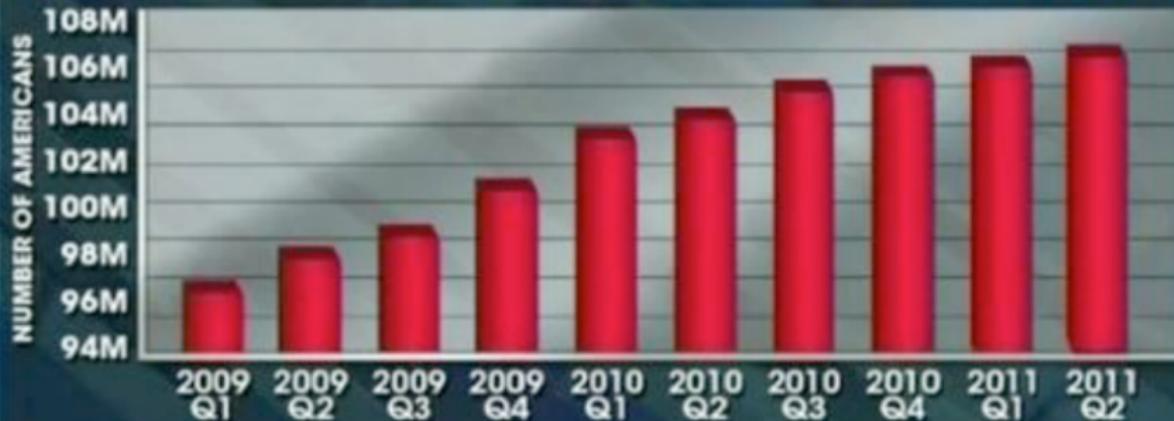
Outline

- Plotting pitfalls: Hall of shame
- Plot critiques
- How to do better
- The grammar of graphics (time permitting)

Plotting pitfalls

- Axis trickery (a.k.a. little y lies)
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

FEDERAL WELFARE RECEIVED IN U.S.



SOURCE: U.S. CENSUS SURVEY

C LIST TYLER HAMILTON OF HIS GOLD MEDAL FRO

NAS



6.61

DEMOCRATIC WOMEN RAISE LESS MONEY THAN DEMOCRATIC MEN

AVERAGE RAISED BY HOUSE PRIMARY WINNERS



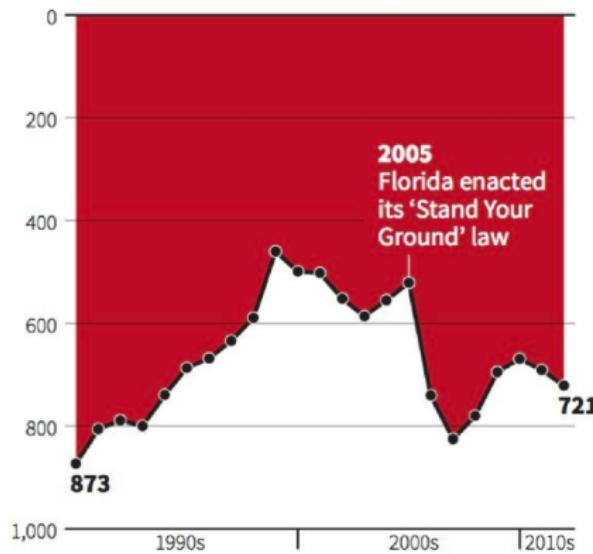
SOURCE: CENTER FOR RESPONSIVE POLITICS

SHOULD BRITAIN LEAVE EU?



Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

The Lie Factor

$$\text{Lie Factor} = \frac{\text{size of effect shown in graphic}}{\text{size of effect in data}}$$

If inflating actual effect, Lie Factor will be **greater than 1**. If deflating actual effect, Lie Factor will be **less than 1**.

And yet . . .

US GDP

\$16 trillion

12

8

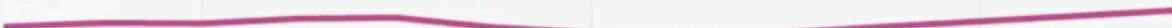
4

0

'08

'09

'10



US GDP

\$15.00 trillion

14.75

14.50

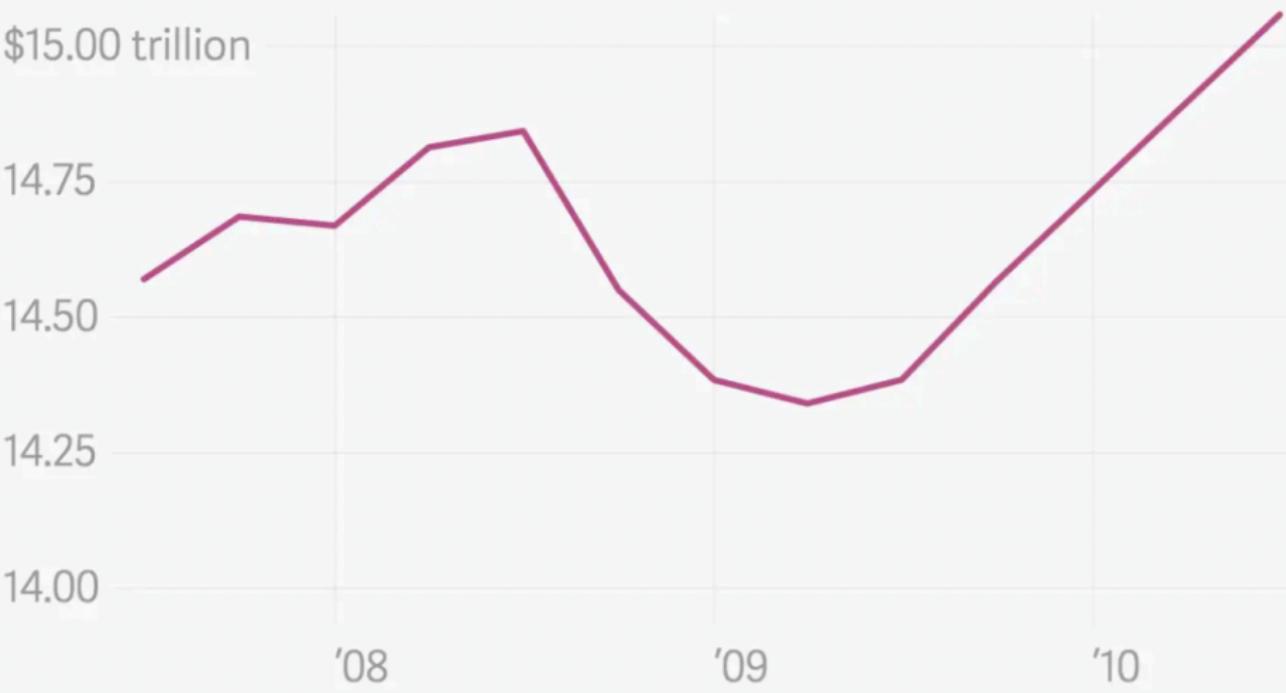
14.25

14.00

'08

'09

'10





Quartz

@qz

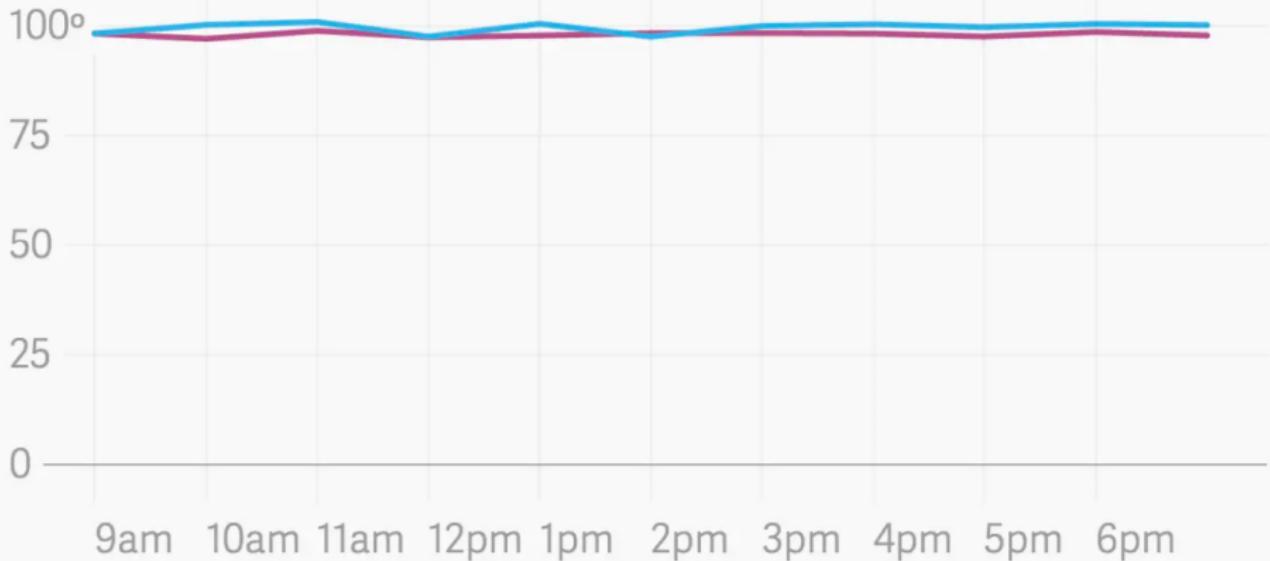
Acquisition rumors light a fire under BlackBerry's stock.
qz.com/note/326942/



6 4:44 PM - Jan 14, 2015

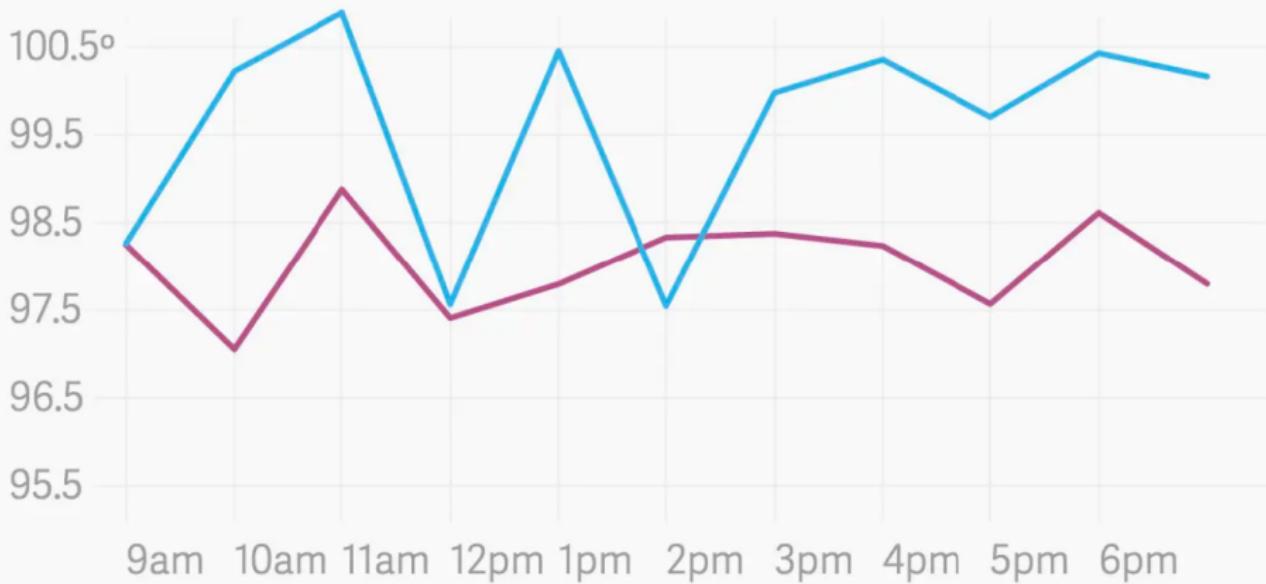
Oral temperature

■ Sara ■ Bob



Oral temperature

■ Sara ■ Bob



Truncating the vertical 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 (a.k.a. “little y lies”)
- **Violations of basic math**
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

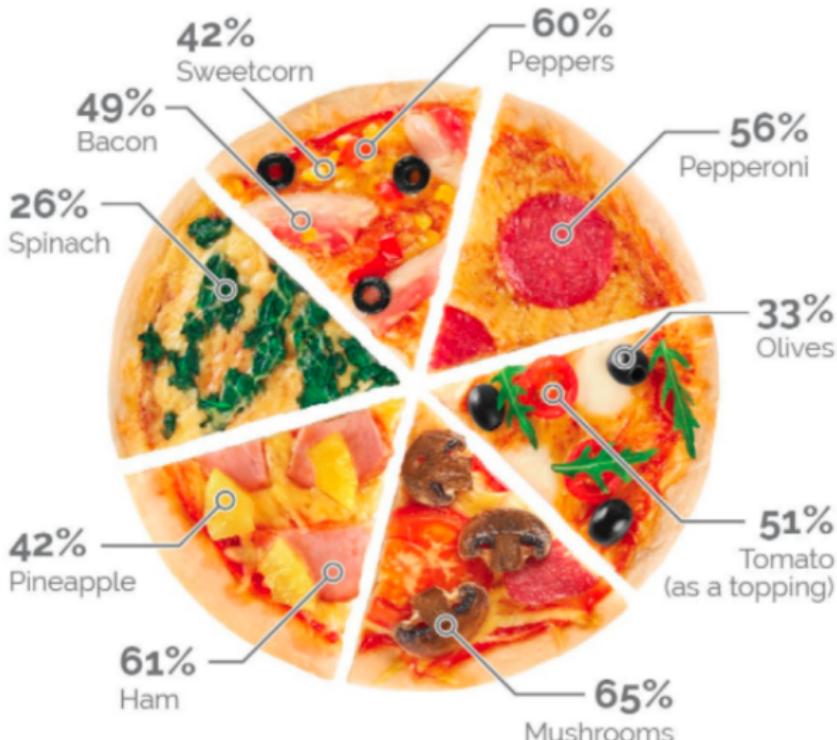
RASMUSSEN REPORTS POLL

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

- | | |
|------------|------------------------|
| 59% | SOMEWHAT LIKELY |
| 35% | VERY LIKELY |
| 26% | NOT VERY LIKELY |



CLIMATE CHANGE RESEARCH / FOX NEWS 11 GOP 5 NHL TOR 6 COB 3



HOW BABY BOOMERS DESCRIBE THEMSELVES



UNEMPLOYMENT RATE UNDER PRESIDENT OBAMA



2011

SOURCE: BUREAU OF LABOR STATISTICS

8.6 > 8.8?

WHAT IS YOUR FAVORITE SEASON? @FCNMIKE

53%



FALL

17%

SUMMER

17%

WINTER

13%

SPRING



#KATIESCOMMUTERS

DEADLY CRASH

DEADLY HOME INVASION

75°

6:51

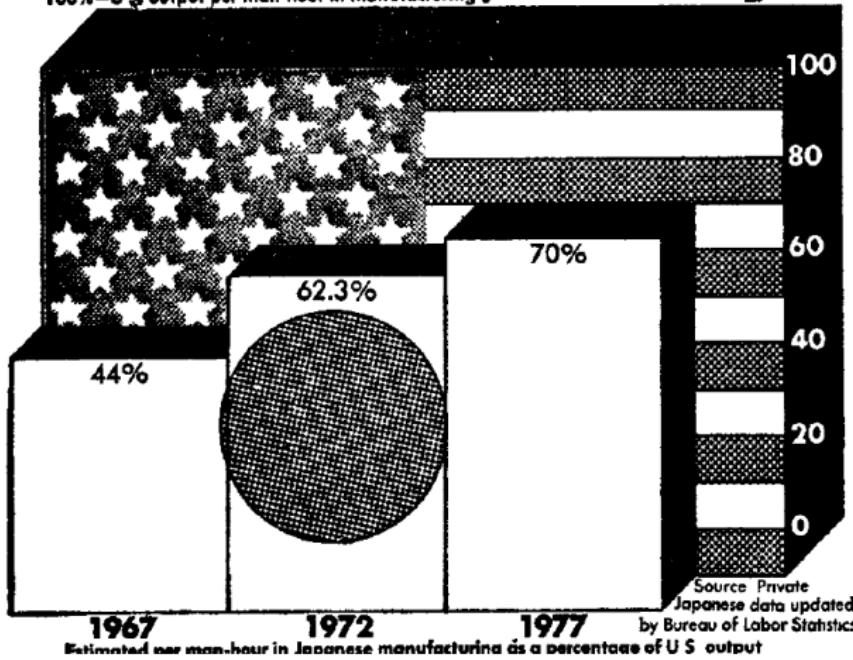
FIRST COAST
NEWS

Plotting pitfalls

- Axis trickery (a.k.a. “little y lies”)
- Violations of basic math
- **Nearly content-free figures**
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

Labor Productivity: U.S. vs Japan

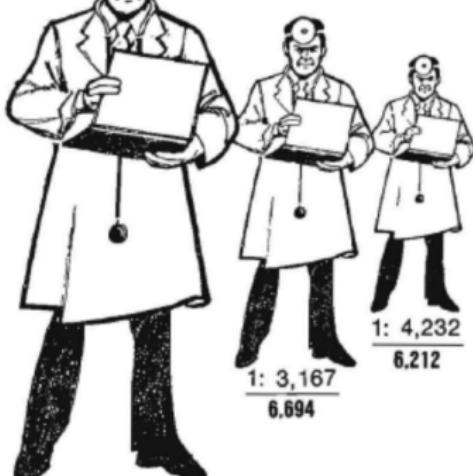
100% = U S output per man-hour in manufacturing



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

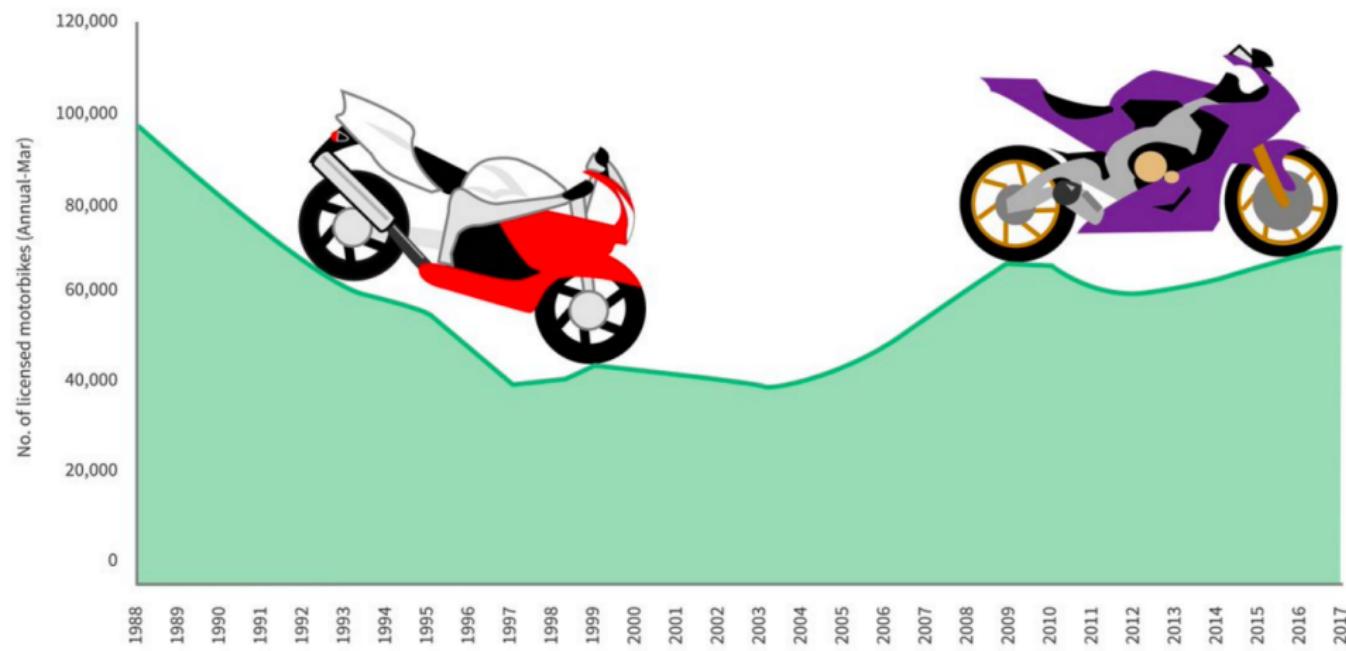


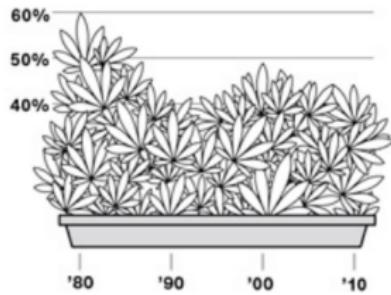
LIVE

MSNBC

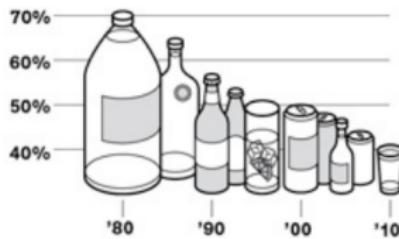
Plotting pitfalls

- Axis trickery (a.k.a. “little y lies”)
- Violations of basic math
- Nearly content-free figures
- **Gratuitous chartjunk**
- Poorly chosen 3D graphics
- Bad design choices

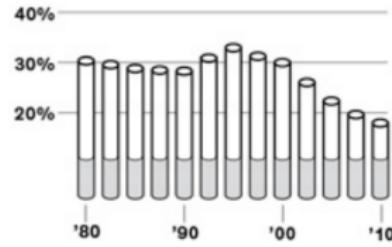




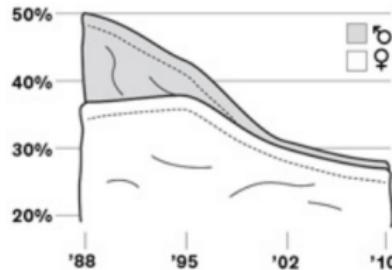
▲ Percentage of high-school seniors who have ever tried pot.



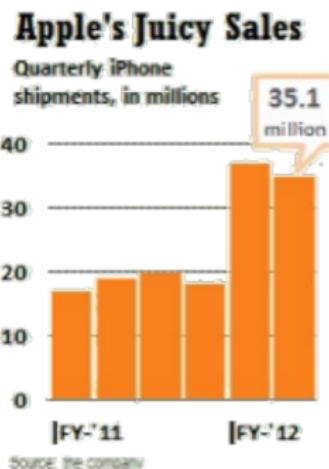
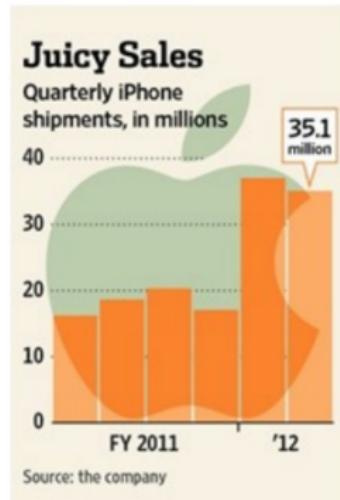
○ Alcohol use by high-school seniors over time.



■ Smoking habits of high-school seniors over time.

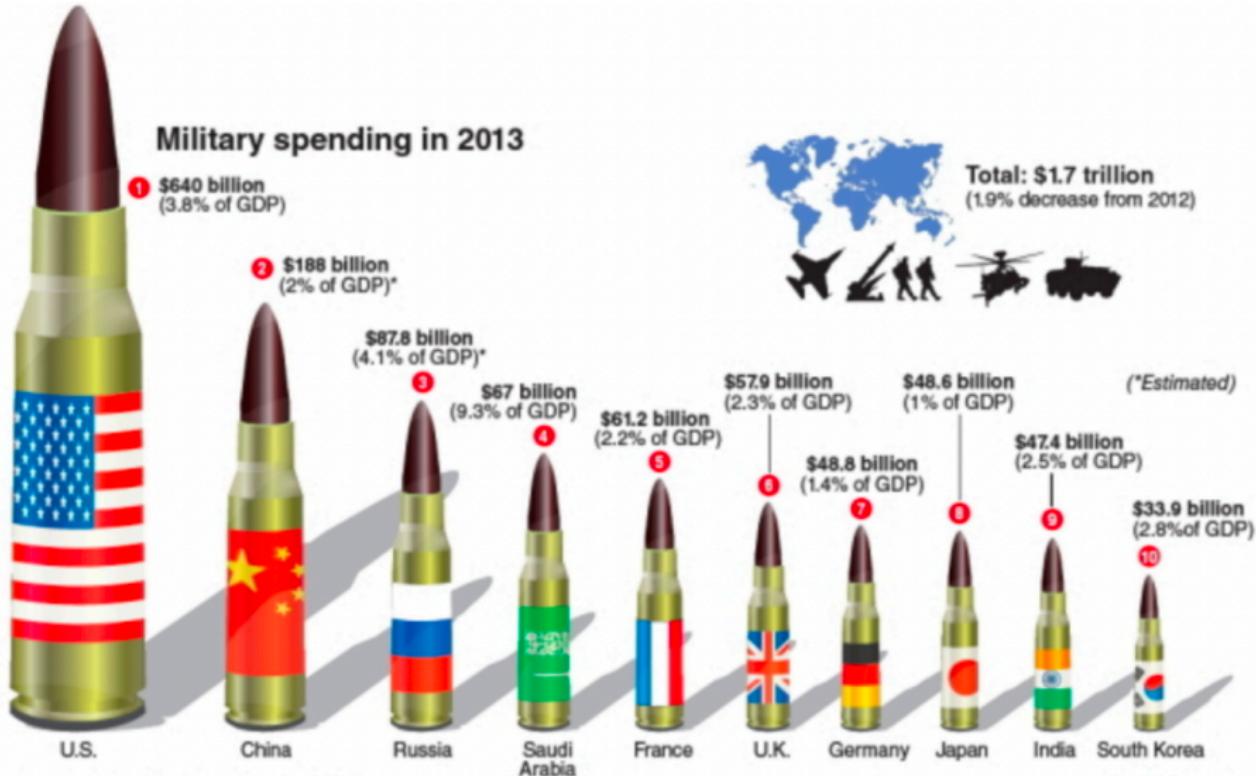


△ Percentage of 15-to-17-year-olds who have had sex.



"Be clear first and clever second. If you have to throw one of those out, throw out clever."

— Jason Fried



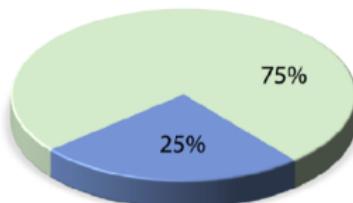
Plotting pitfalls

- Axis trickery (a.k.a. “little y lies”)
- Violations of basic math
- Nearly content-free figures
- Gratuitous chartjunk
- Poorly chosen 3D graphics
- Bad design choices

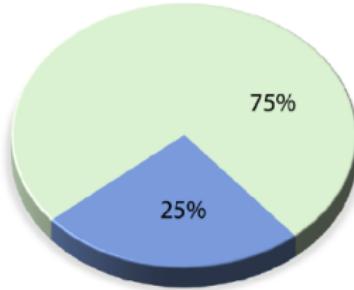
a



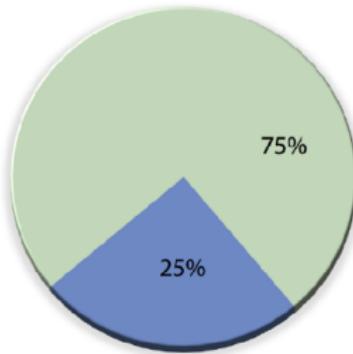
b



c



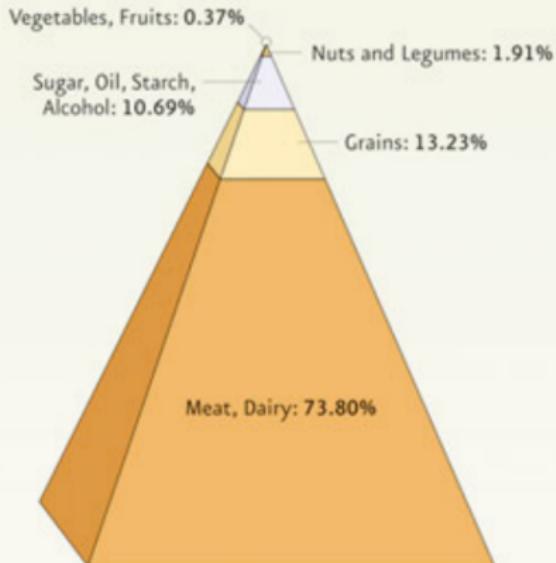
d



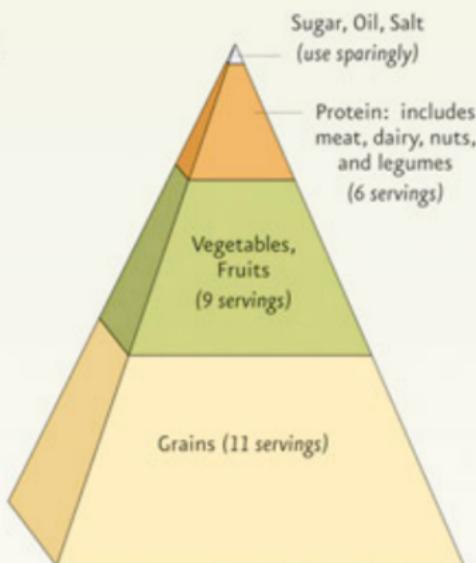
Same chart, four angles

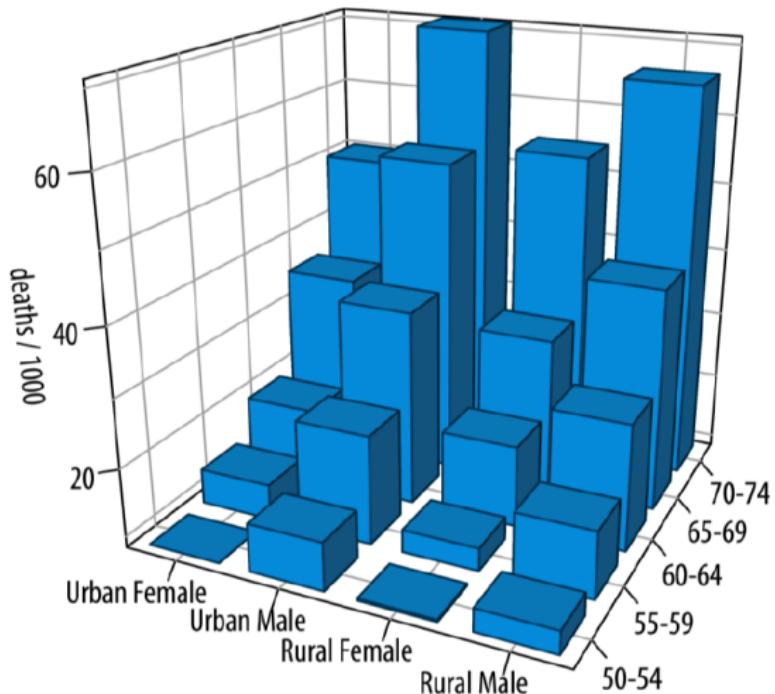
Why Does a Salad Cost More Than a Big Mac?

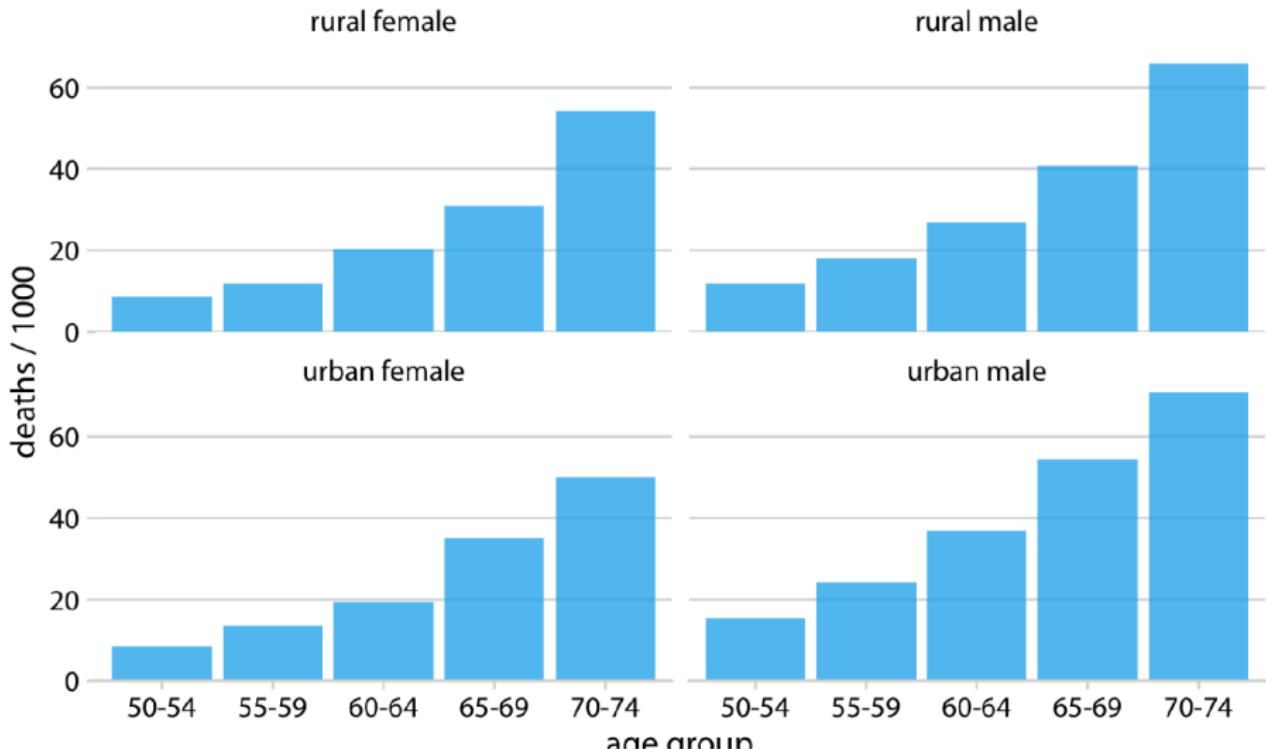
Federal Subsidies for Food Production, 1995-2005[®]



Federal Nutrition Recommendations



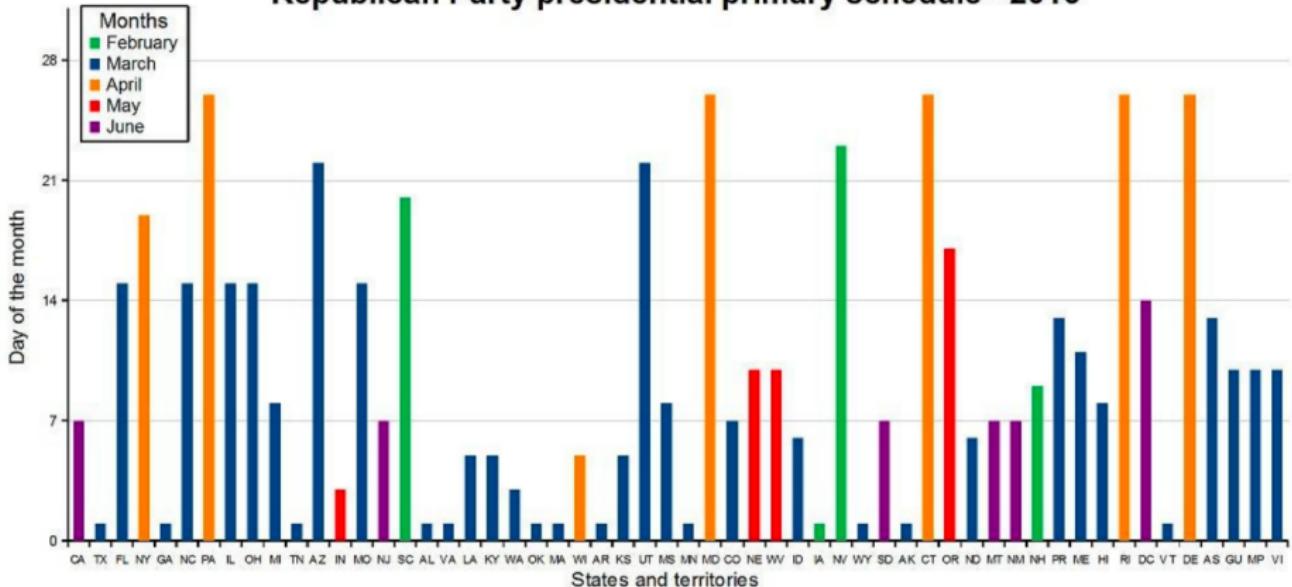




Plotting pitfalls

- Axis trickery (a.k.a. “little y lies”)
- 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 calendar?

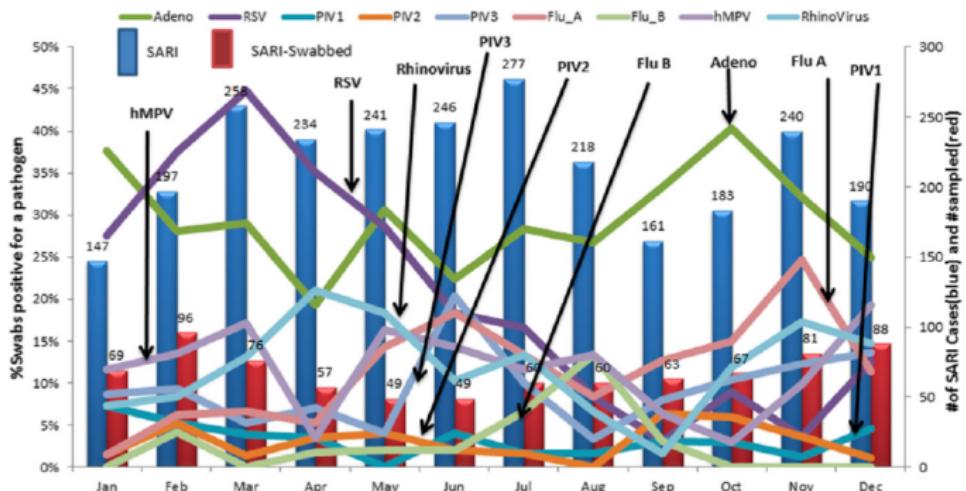
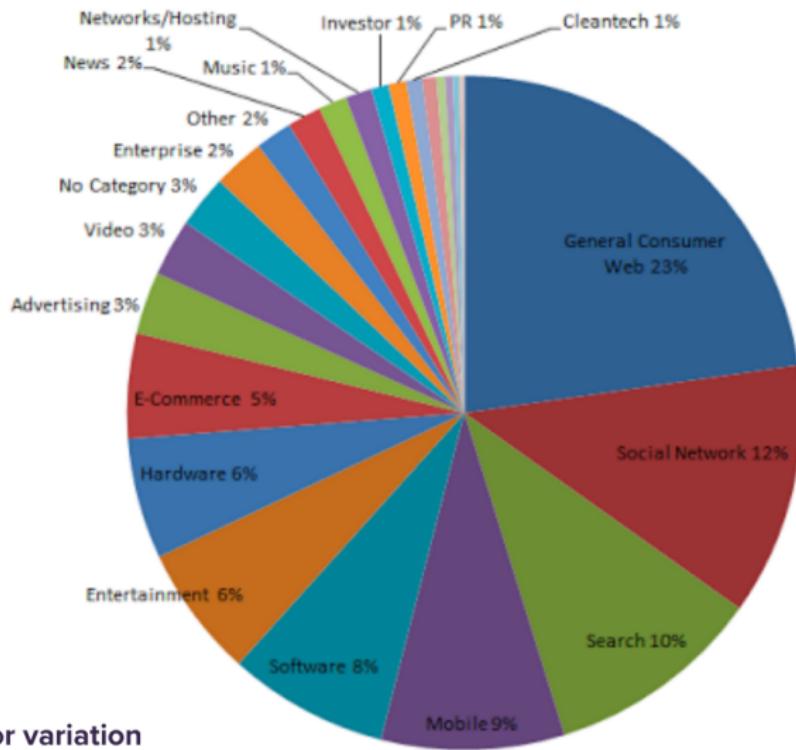
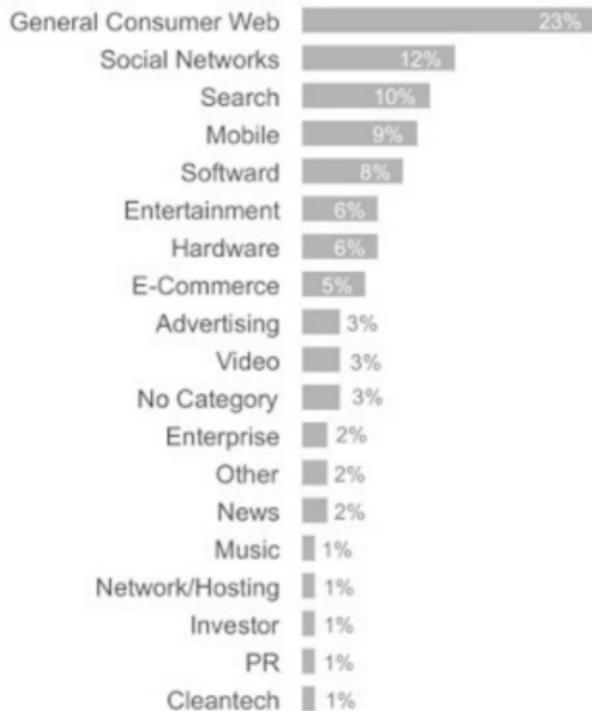


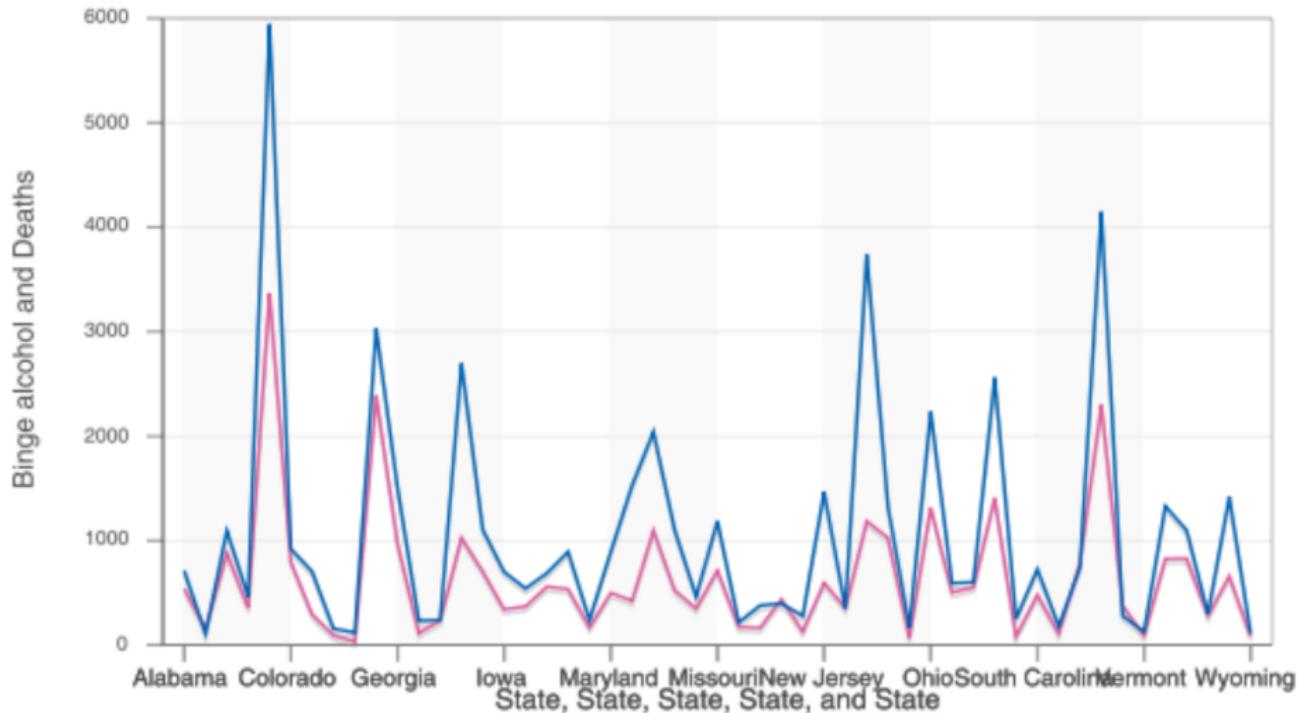
Figure 1 Monthly (aggregate) distribution of viral pathogens, March 1, 2007–Feb 28, 2011.



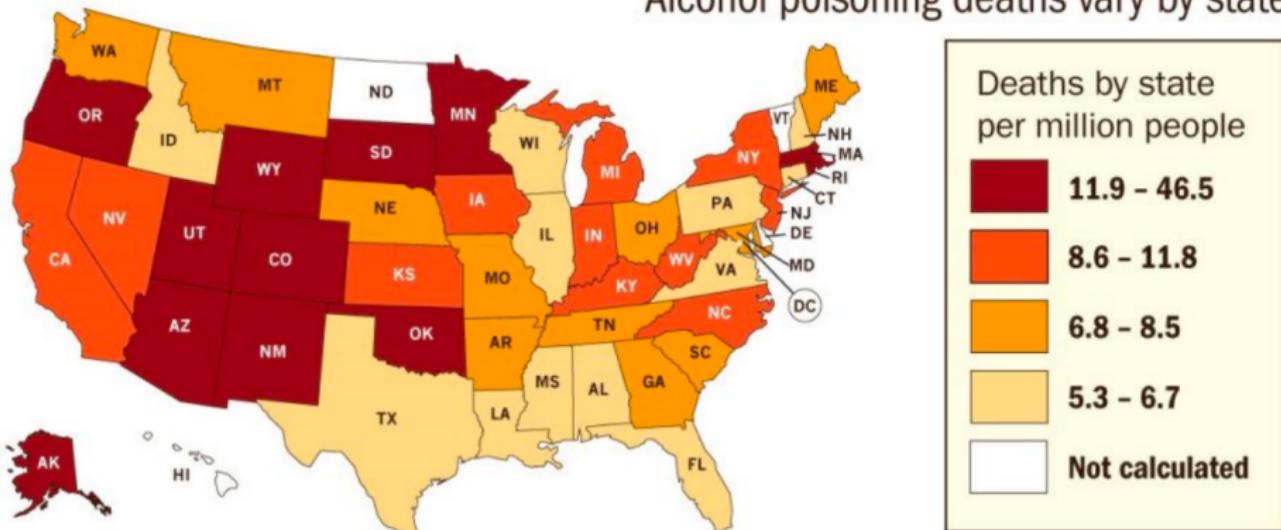
Meaningless color variation



As a bar plot



Alcohol poisoning deaths vary by state

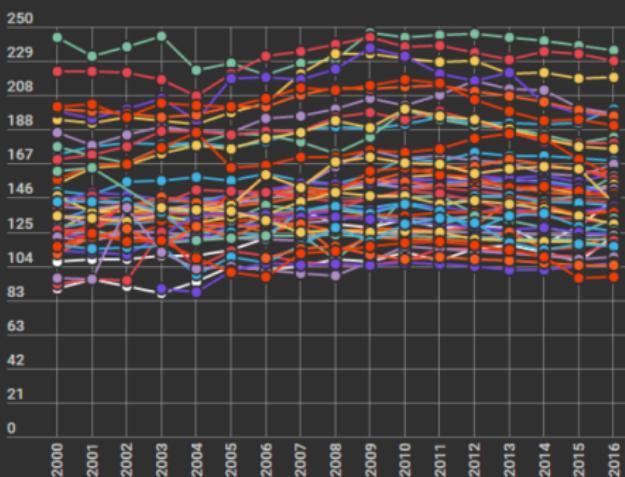


SOURCE: National Center for Health Statistics Mortality Multiple Cause Files, 2010-2012.

Your turn!

- Break in the groups of 2-3 and answer the following questions about the next figure:
- What is this figure trying to convey, and is it successful or not?
- If successful, what features make it so? If unsuccessful, what would you change to make it the best graphic ever?

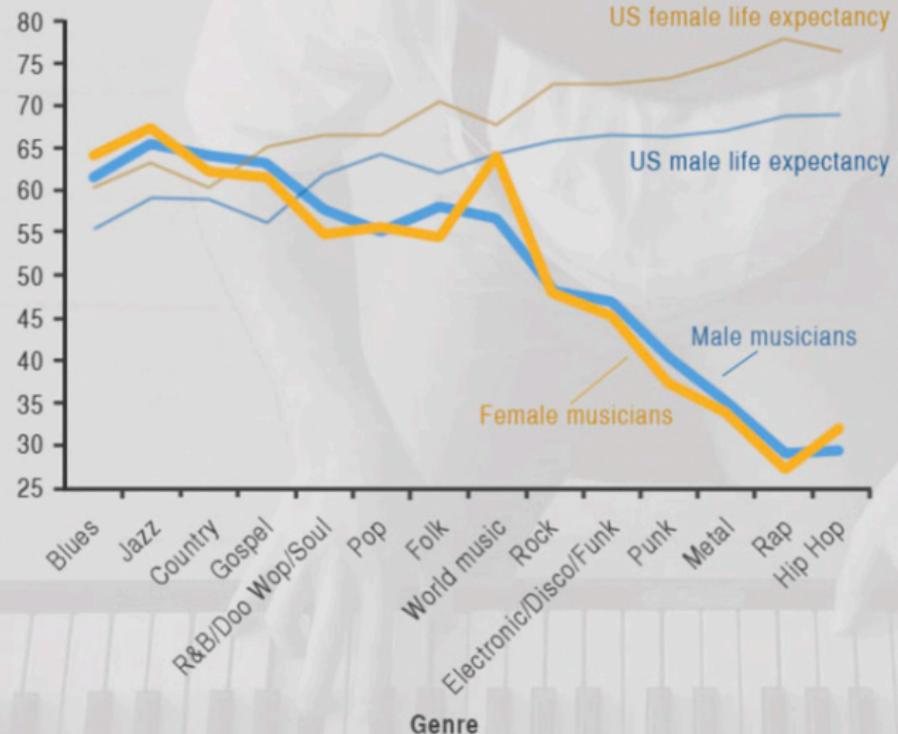
Canada's most heavily-policed cities (x per 100,000)



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>

How to do better . . .

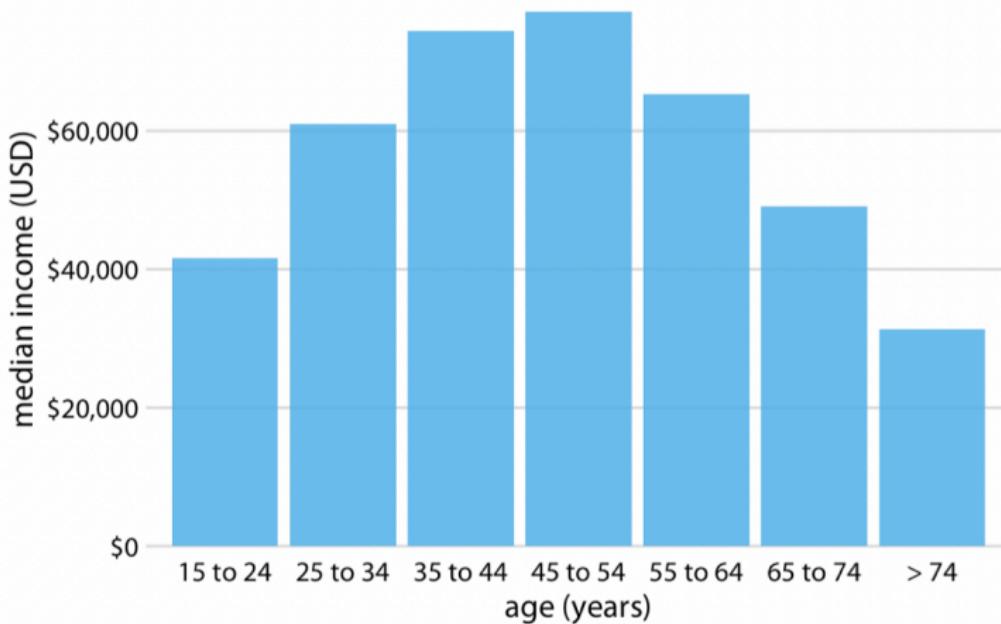
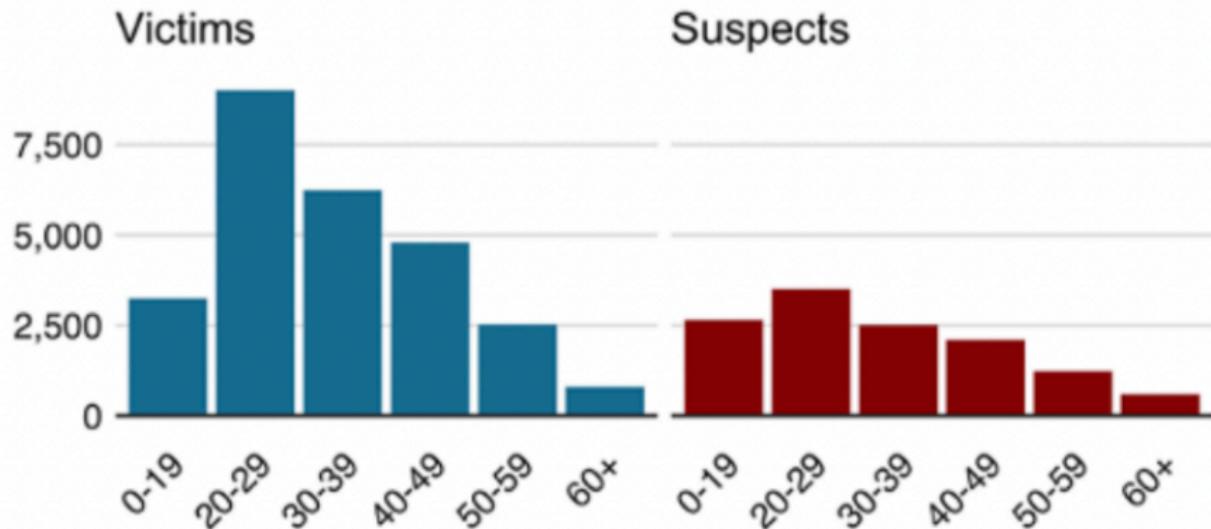


Figure 6.5: 2016 median U.S. annual household income versus age group. The 45–54 year age group has the highest median income. Data source: United States Census Bureau

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

Number in each age group 2014 - 2017



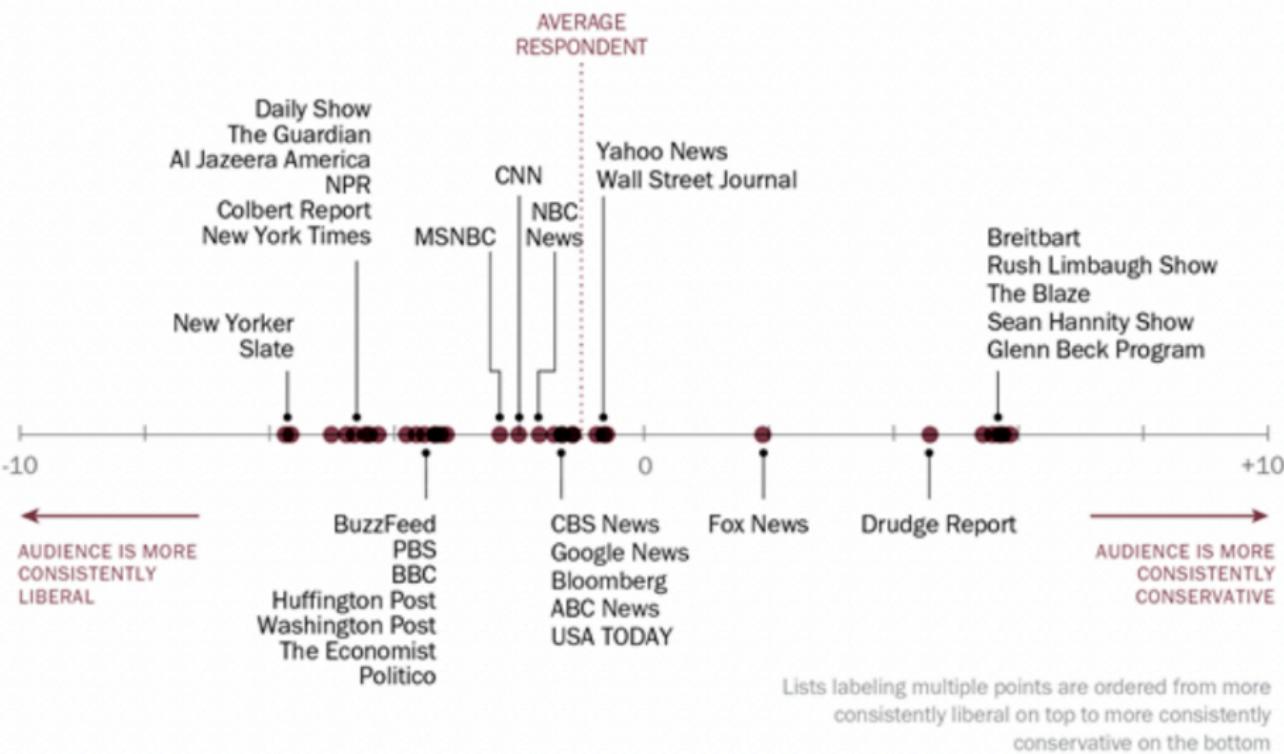
N.B. Analysis excludes 3,474 victims and 2,003 suspects whose ages are unknown

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...

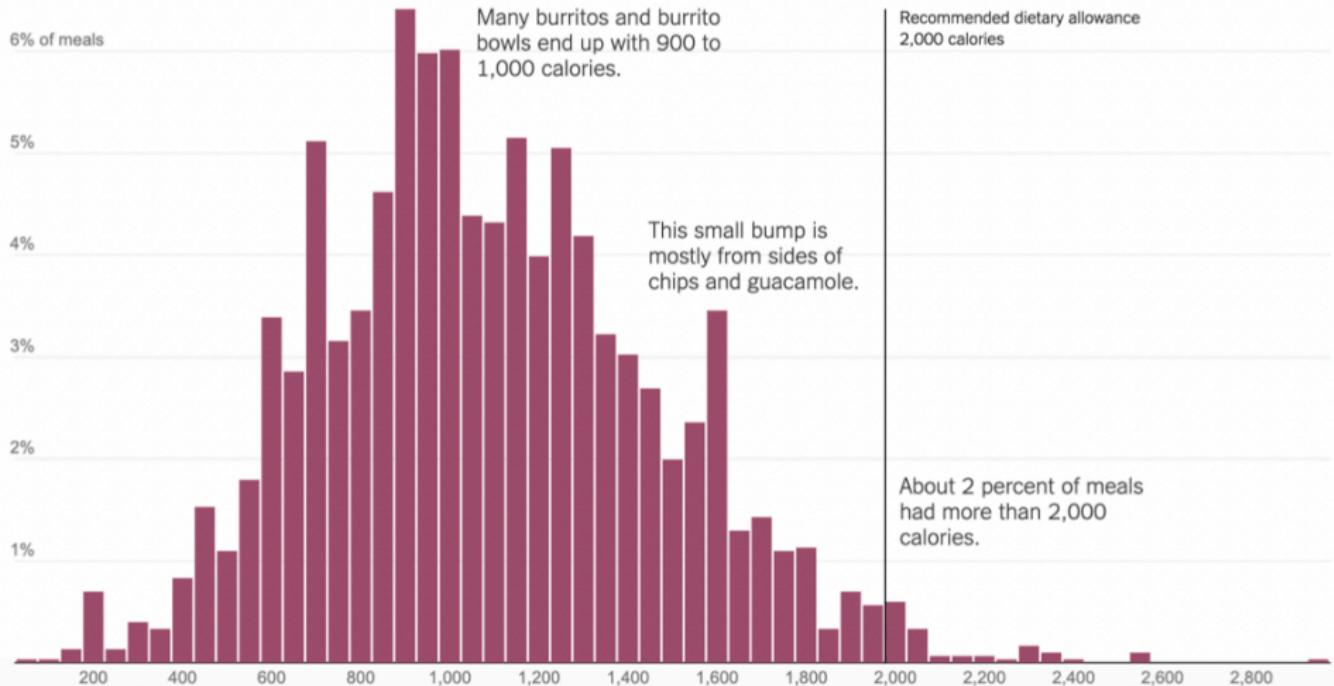


At Chipotle, How Many Calories Do People Really Eat?

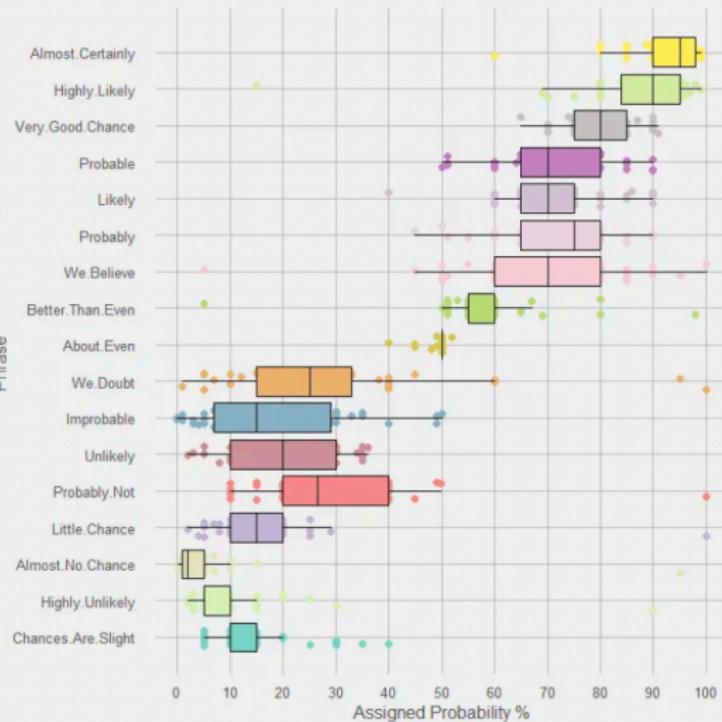
By KEVIN QUEALY, AMANDA COX and JOSH KATZ FEB. 17, 2015

Most meals have more than 1,000 calories and almost a full day's worth of sodium.

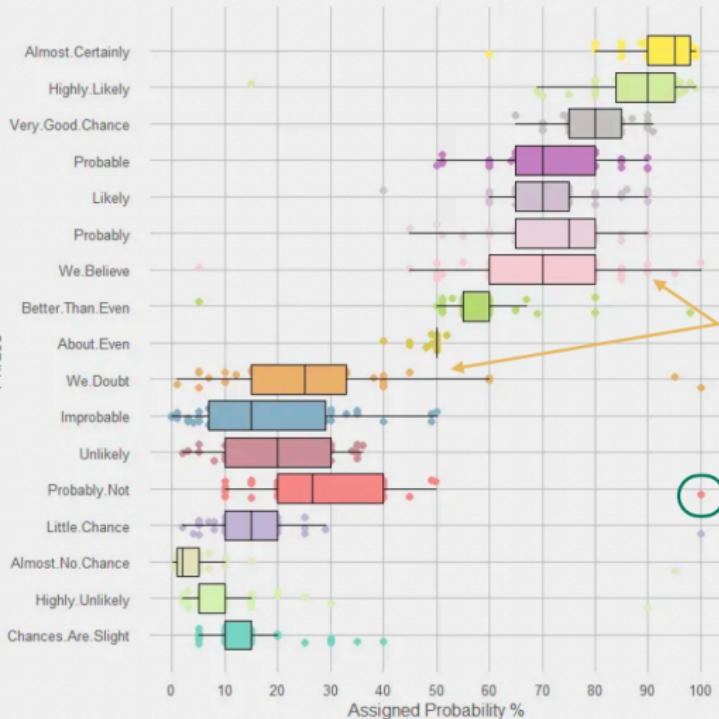
[RELATED ARTICLE](#)



Perceptions of Probability



Perceptions of Probability



The vertical lines show the median in each group.

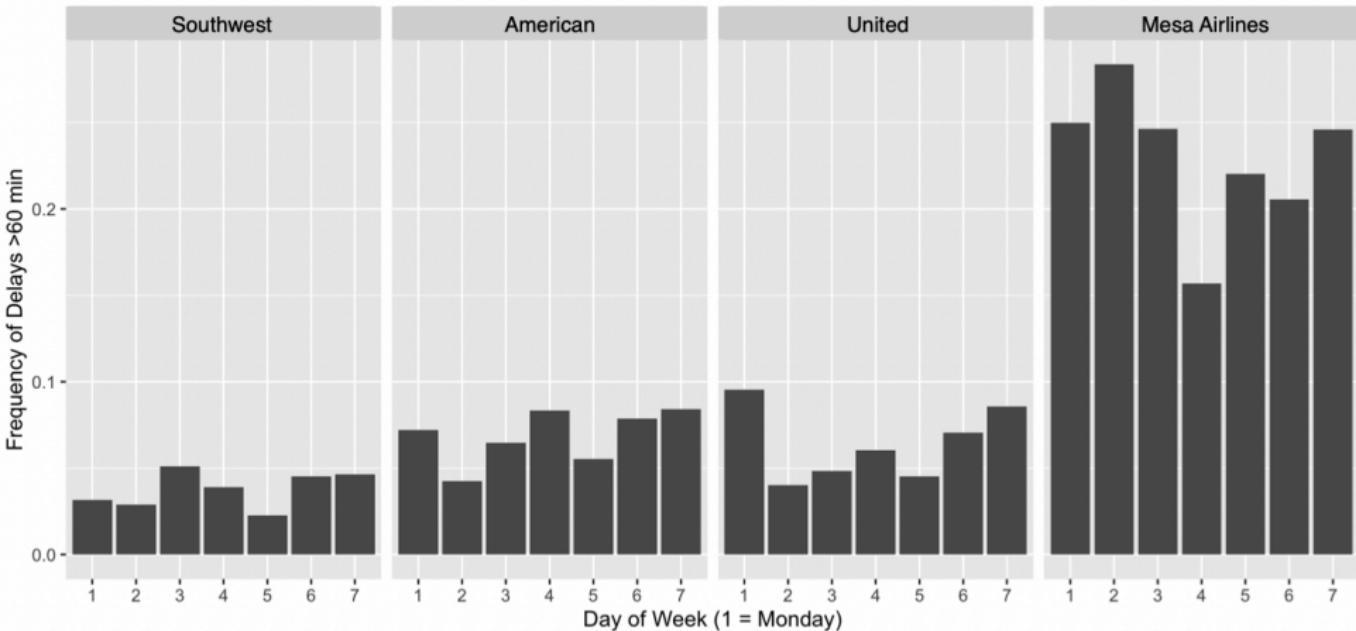
The boxes show the middle 50% of the data in each group. This is the IQR: the range from the 25th to 75th percentile.

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

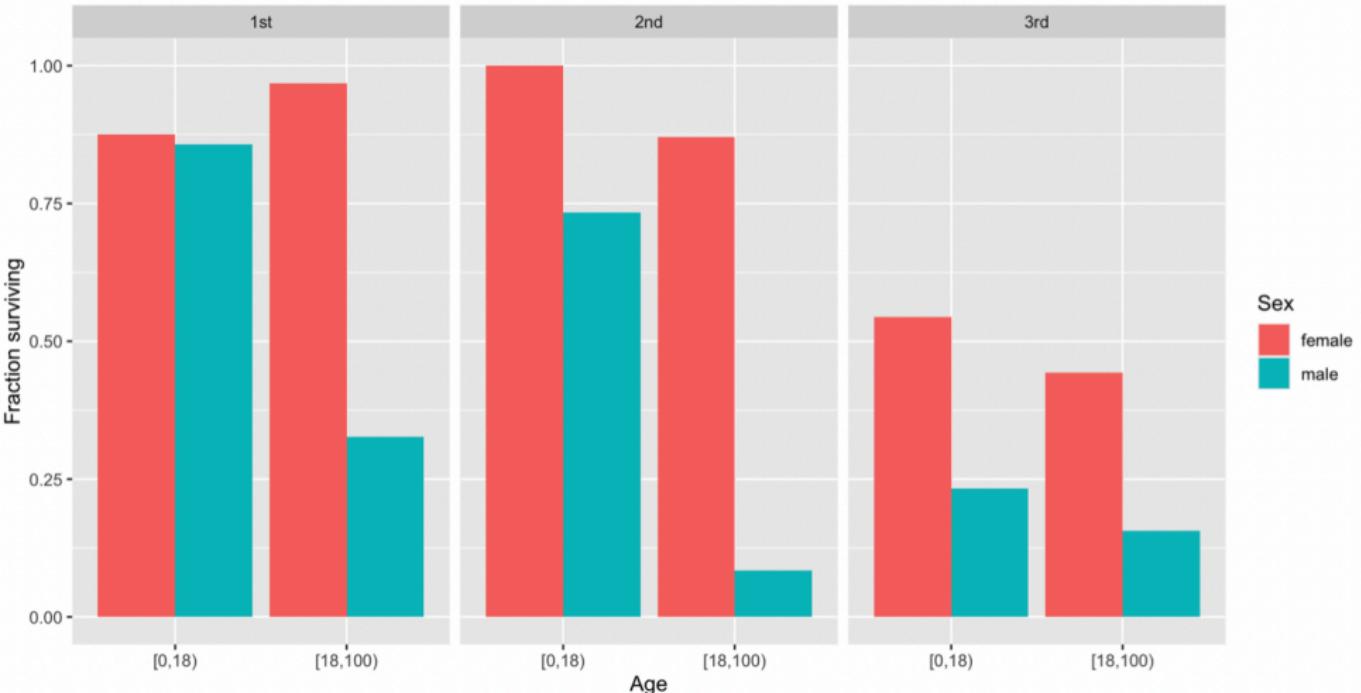
Data points falling outside the whiskers are shown individually, like this one here.

(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



THE GRAMMAR OF GRAPHICS

- Consider sentences that an English learner (e.g. a toddler) might say:
 - “Squirrel climbs tree.”
 - “Anna eats cake.”
 - “Jimmy needs potty.”
- These sentences obey a simple, consistent grammar with three core elements:
 - a *subject* that performs an action (squirrel, Anna).
 - an *object* that receives the action (tree, cake).
 - a *verb*, i.e. the action itself (climbs, eats).
- If you swap out different subjects/objects/verbs in the basic grammar, you can generate lots of different, fully comprehensible sentences.

THE GRAMMAR OF GRAPHICS

- Statistical plots are just the same.
- They obey a simple, consistent grammar with three basic elements

Variables in a data set (like the weight or engine size of a vehicle). These are like the subject of a sentence.

Objects: specifically, geometric objects (like dot or lines or bars). These are like the object of a sentence.

Mappings from the data variables to aesthetic properties of the geometric objects (like their size, location, or color). These are like the verb in a sentence.

- If you swap out different variables/objects/mappings in the basic grammar, you can generate lots of different plots.

EXAMPLE: BIKE SHARE DATA

Here you see the first several rows of a data set on median hourly demand for bike-share rentals in Washington, DC's Capital Bikeshare program, stratified by whether it's a working day (1) or not (0).

hr	workingday	median_rentals
0	0	94
0	1	33
1	0	71
1	1	14
2	0	54
2	1	7

EXAMPLE: BIKE SHARE DATA

— Variables:

Hour

Median rentals

Working day

— Objects:

Points/lines

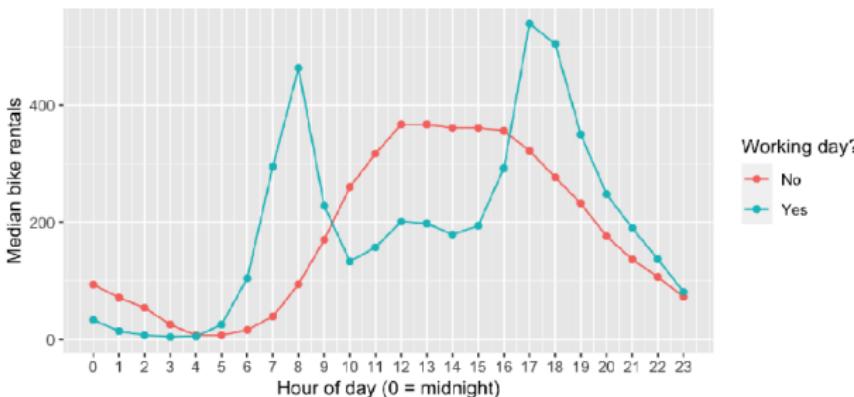
— Mappings:

Hour → x

Median rentals → y

Working day → color

Bike-share rentals in Washington, DC (2011-12)



EXAMPLE: BIKE SHARE DATA

- Variables:

- Hour

- Median rentals

- Working day

- Objects:

- The two panels

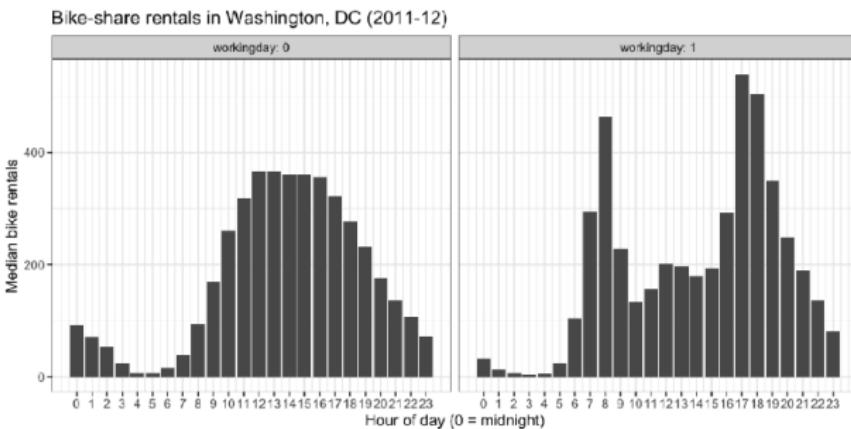
- The bars in each panel

- Mappings:

- Hour → location of bar

- Median rentals → height of bar

- Working day → left vs. right panel



Same variables, but different objects/
mappings, and so a different plot.

A plot is a mapping of data variables to the aesthetic properties of geometric objects.