

# Graphical Unexcellence

Data Viz Hall of Shame

---

(Professor Dave)<sup>2</sup>

# Outline

---

- test

**Figure 1**

# DATA VISUALIZATION

**Plotting pitfalls: the data vis hall of shame**

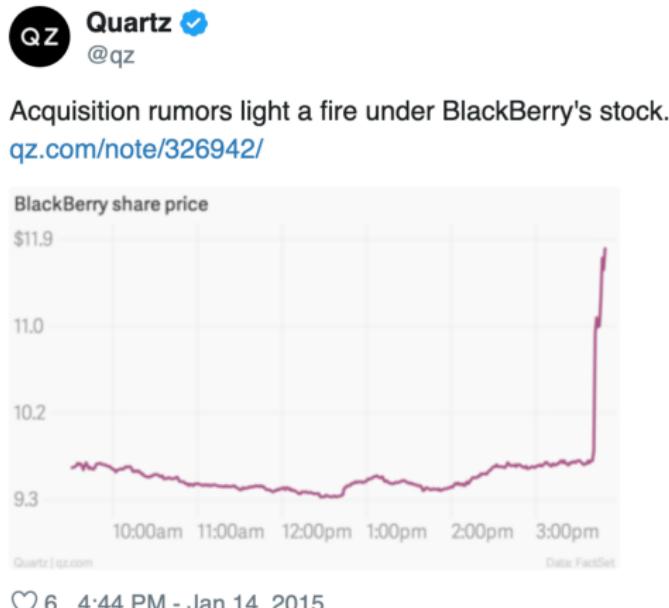
**Plot critique**

**The grammar of graphics**

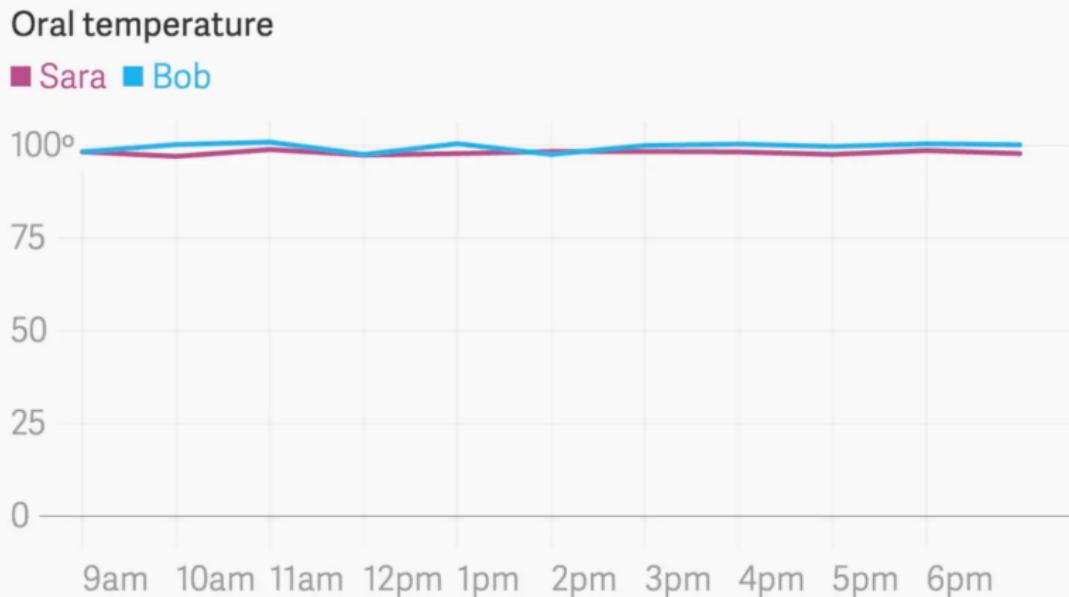
**The five most important plots**

**Enriching plots (color, faceting, labels, etc)**

# Figure 10



# Figure 11



## Figure 12

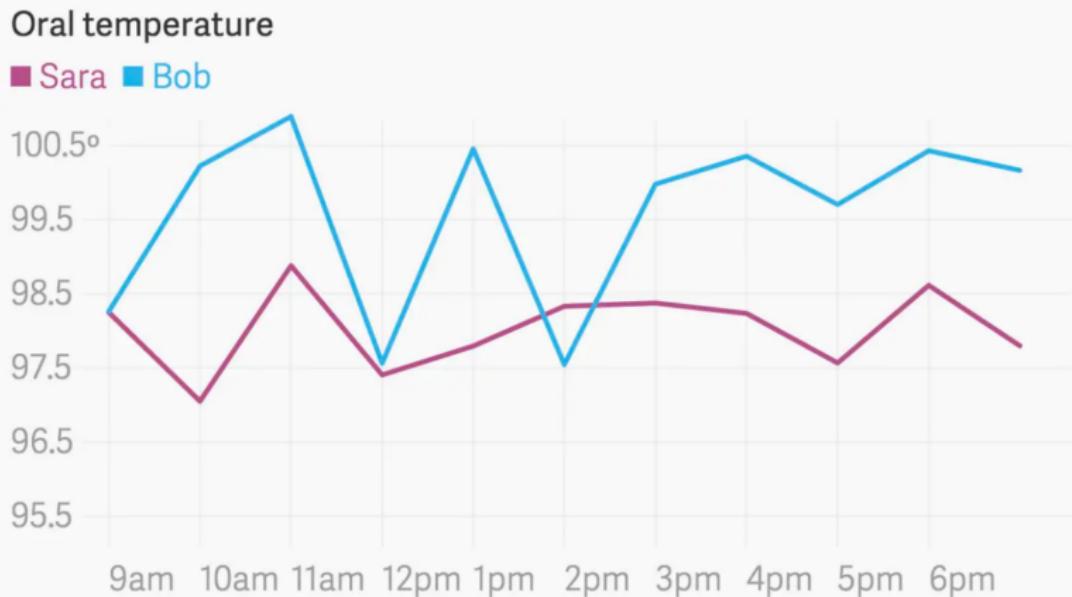
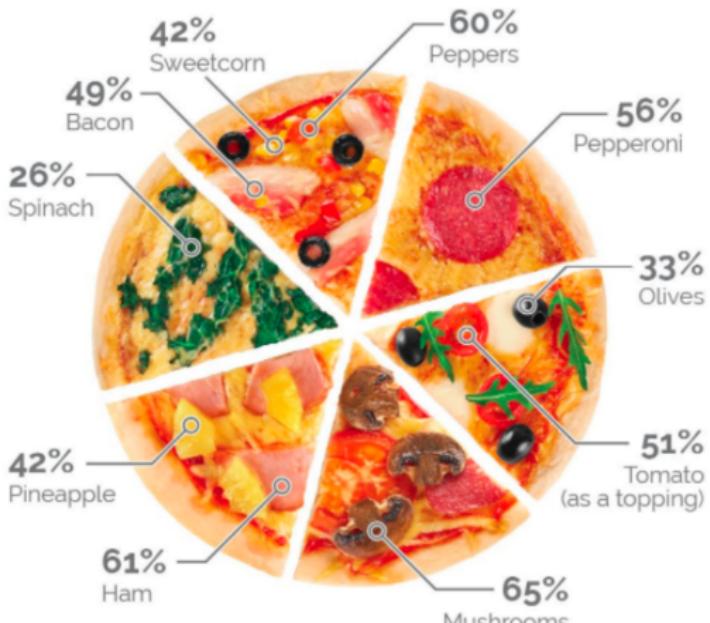


Figure 15



## Figure 16



## Figure 17

### HOW BABY BOOMERS DESCRIBE THEMSELVES

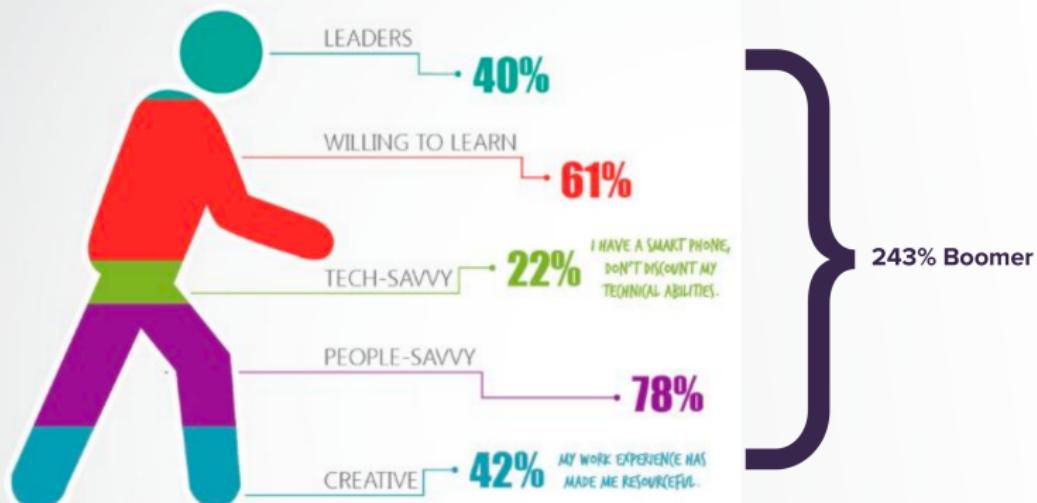


Figure 18

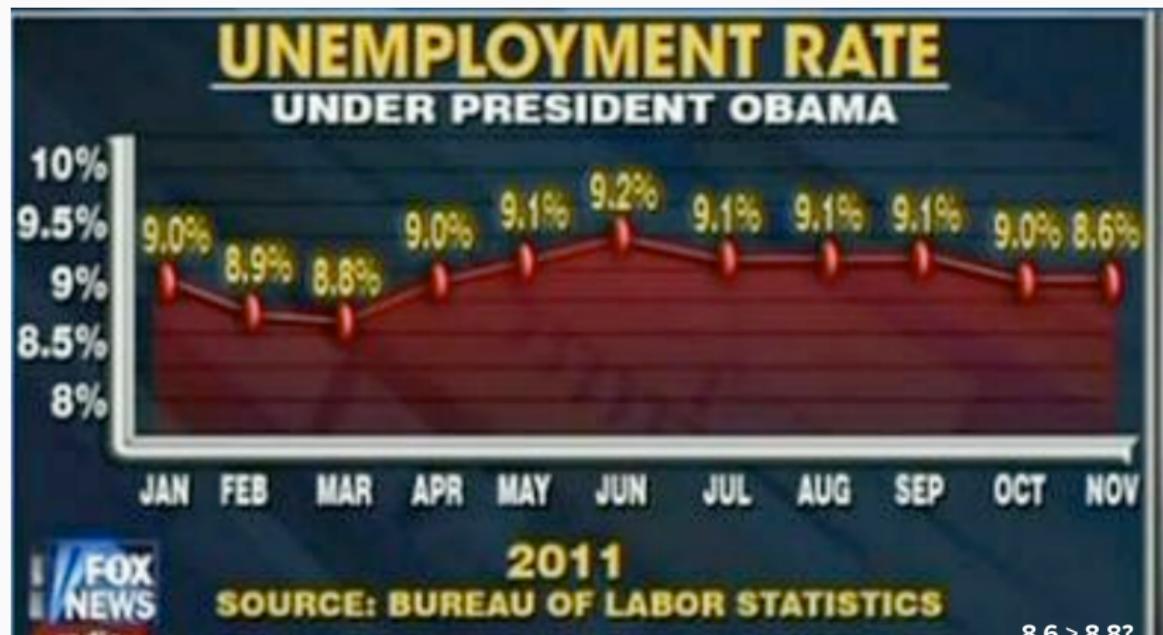


Figure 19

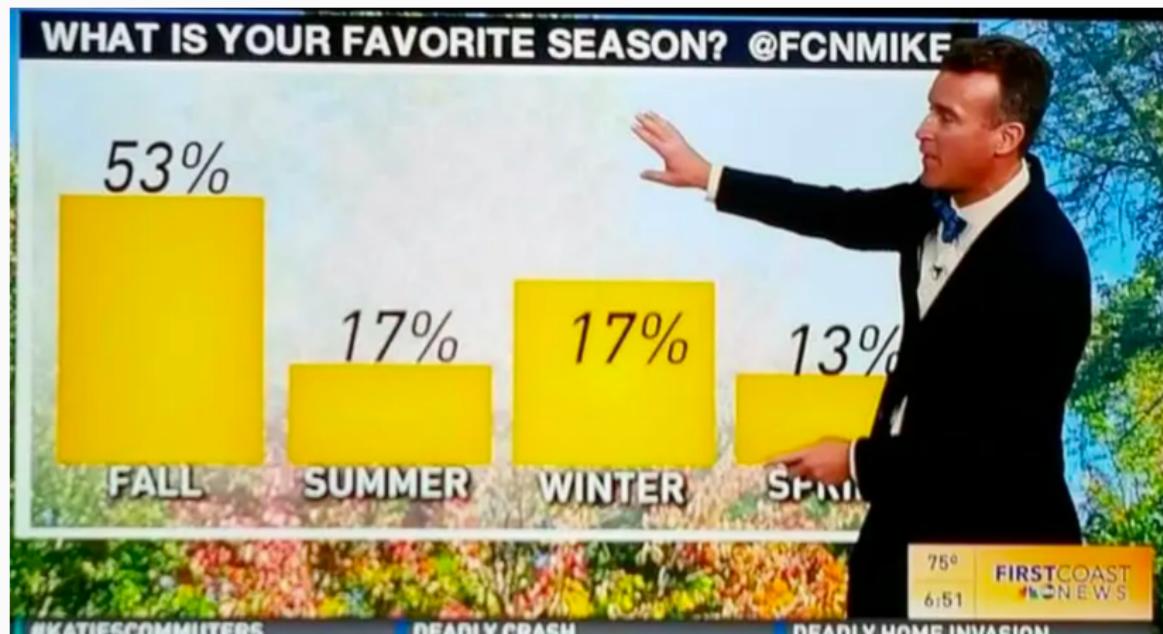
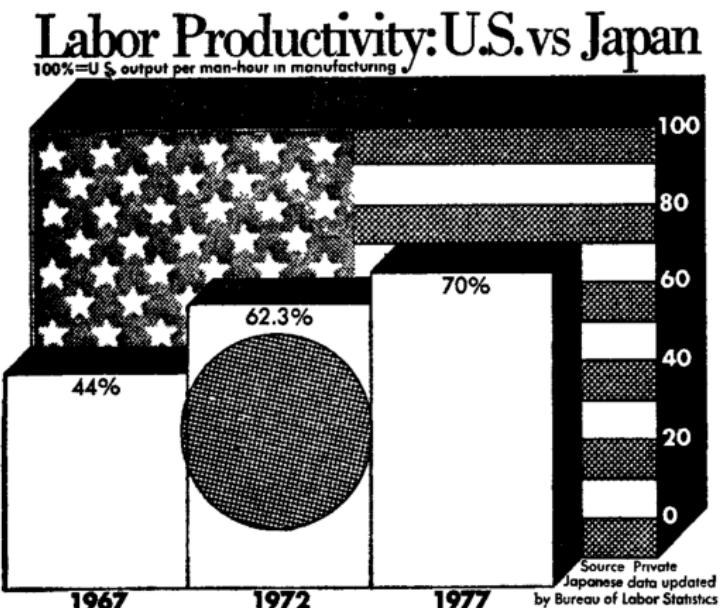


Figure 21



**Figure 22**

## **THE SHRINKING FAMILY DOCTOR** In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27 %	16.0 %	12.0 %

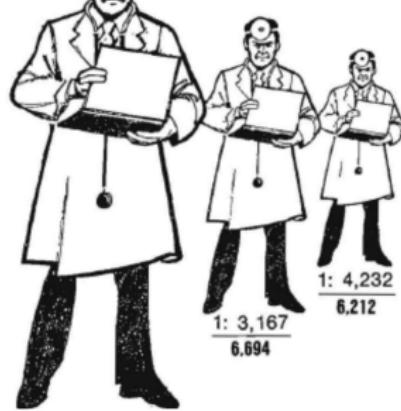
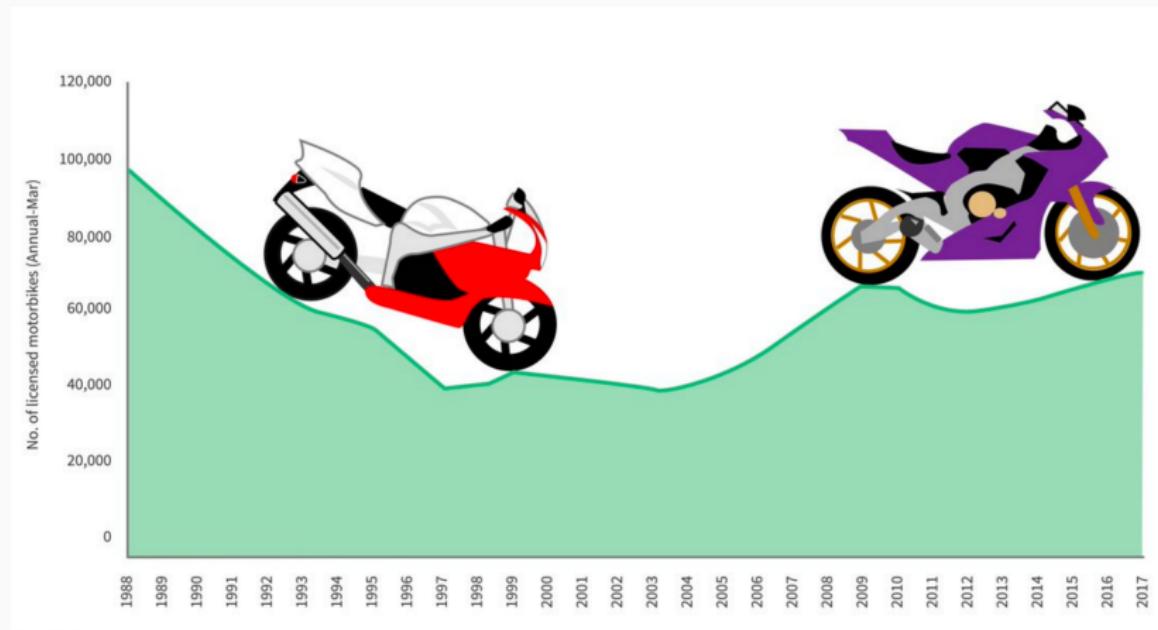


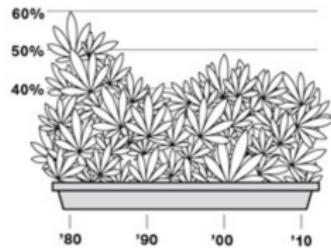
Figure 23



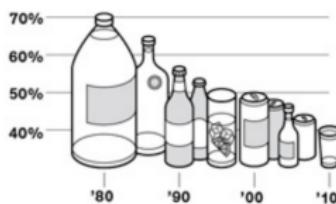
## Figure 25



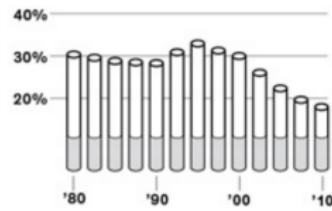
# Figure 26



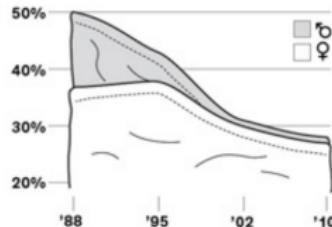
▲ Percentage of high-school seniors  
who have ever tried pot.



○ Alcohol use by high-school

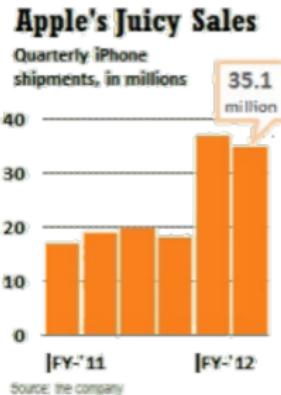
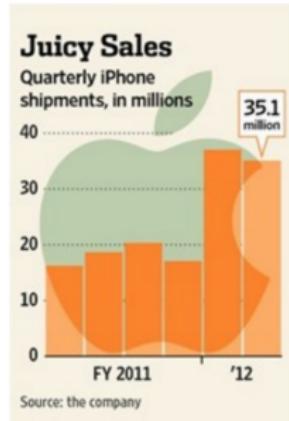


□ Smoking habits of high-school  
seniors over time.



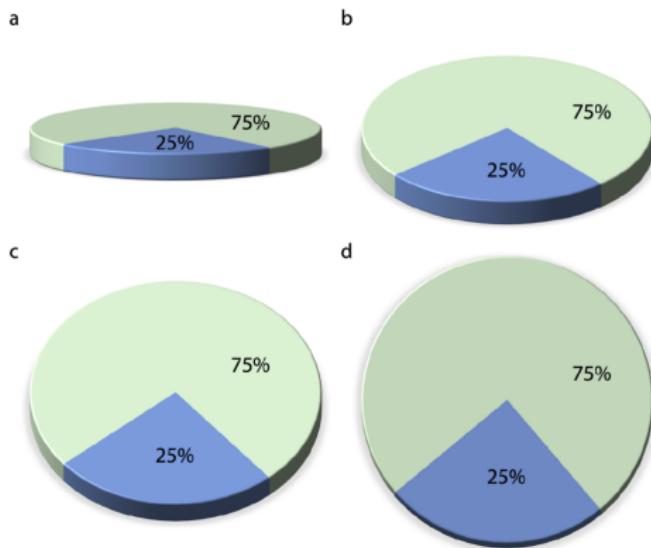
◆ Percentage of 15-to-17-year-olds

## Figure 27



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

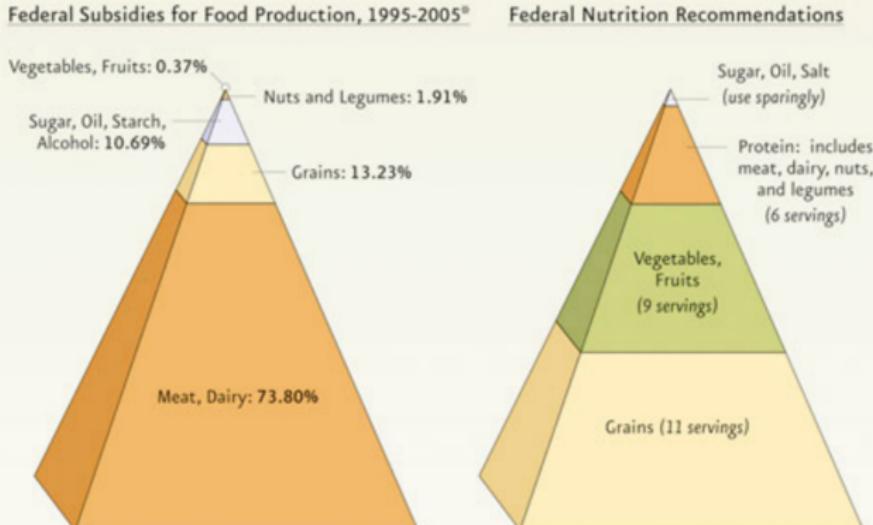
# Figure 29



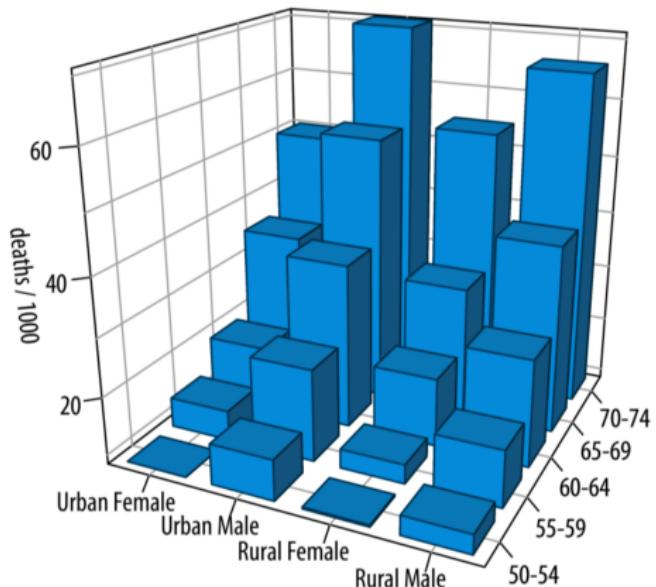
Same chart, four angles

# Figure 30

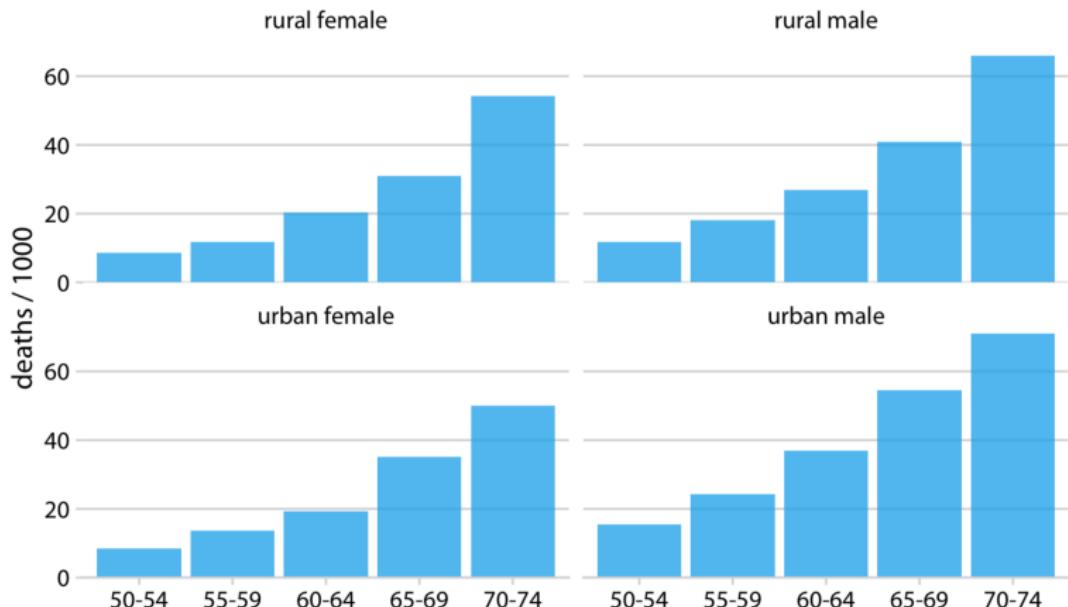
## Why Does a Salad Cost More Than a Big Mac?



## Figure 31

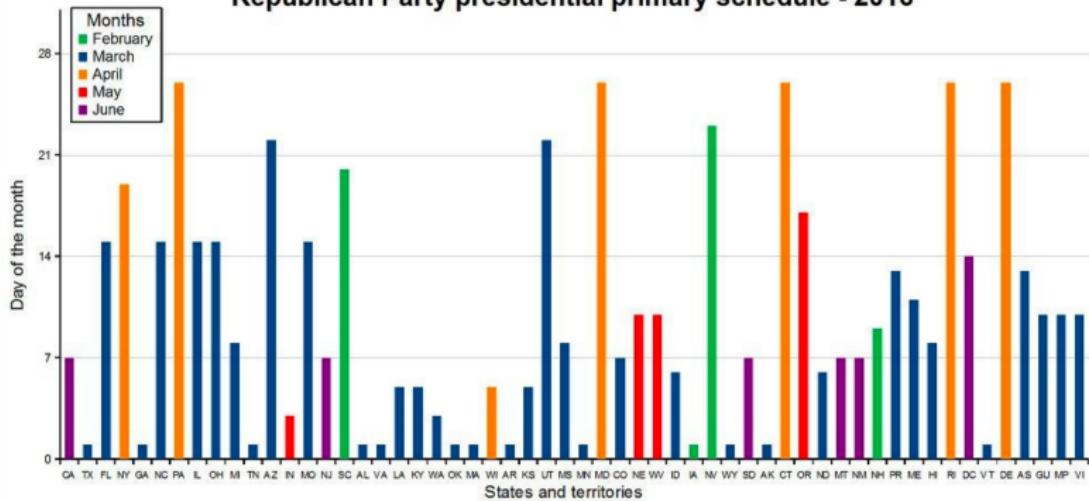


## Figure 32



# Figure 34

## Republican Party presidential primary schedule - 2016



# Figure 35

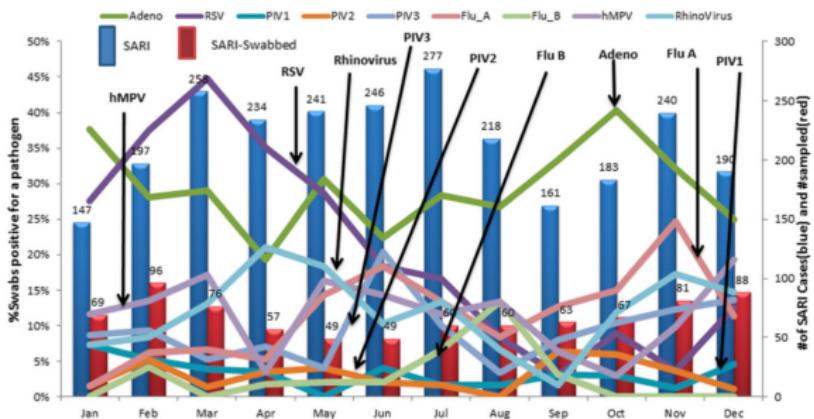
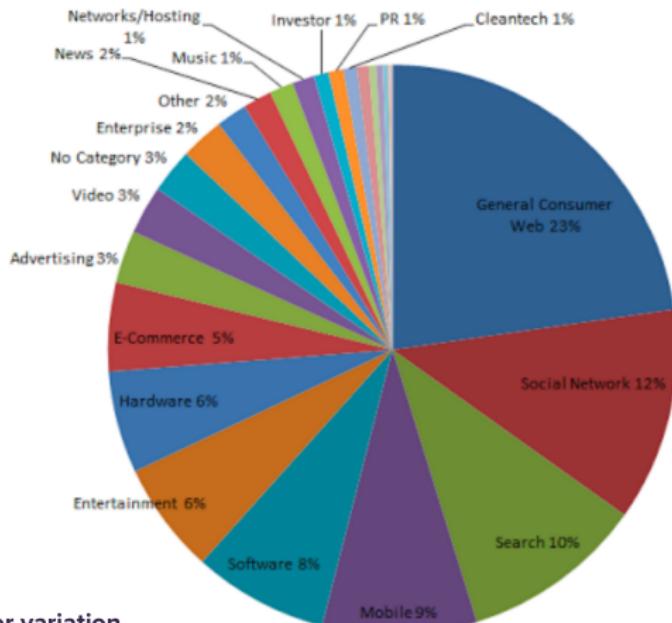
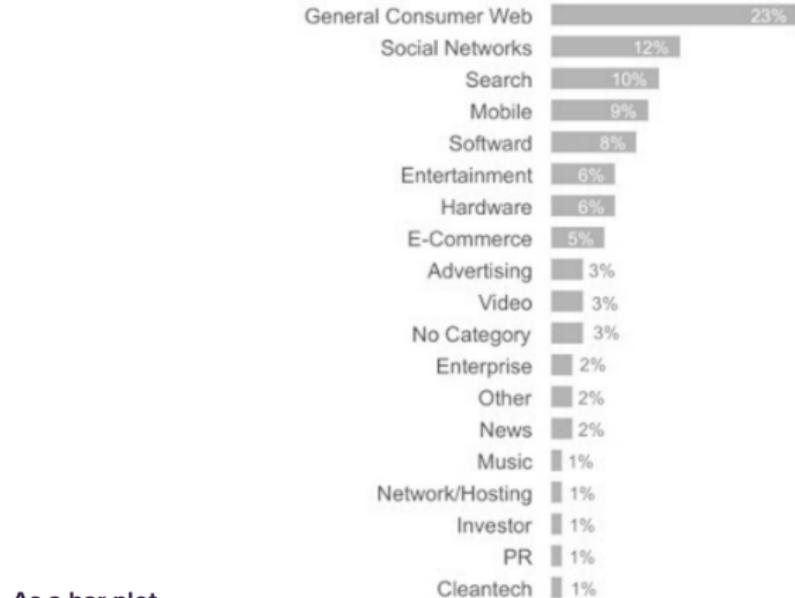


Figure 1 Monthly (aggregate) distribution of viral pathogens, March 1, 2007-Feb 28, 2011.

## Figure 36

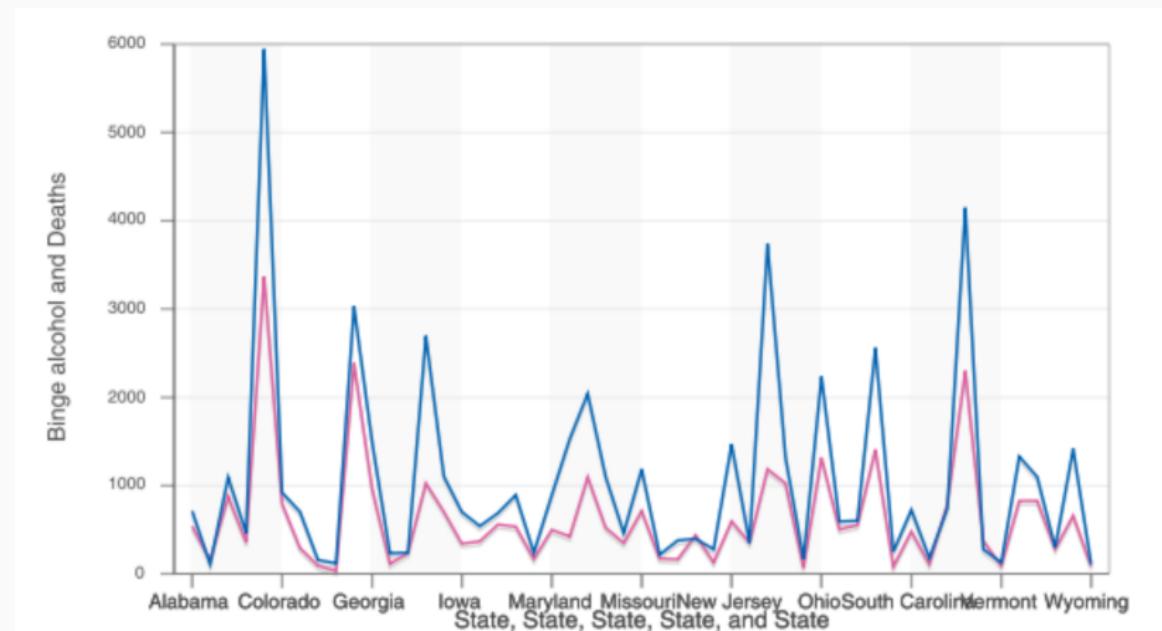


## Figure 37



As a bar plot

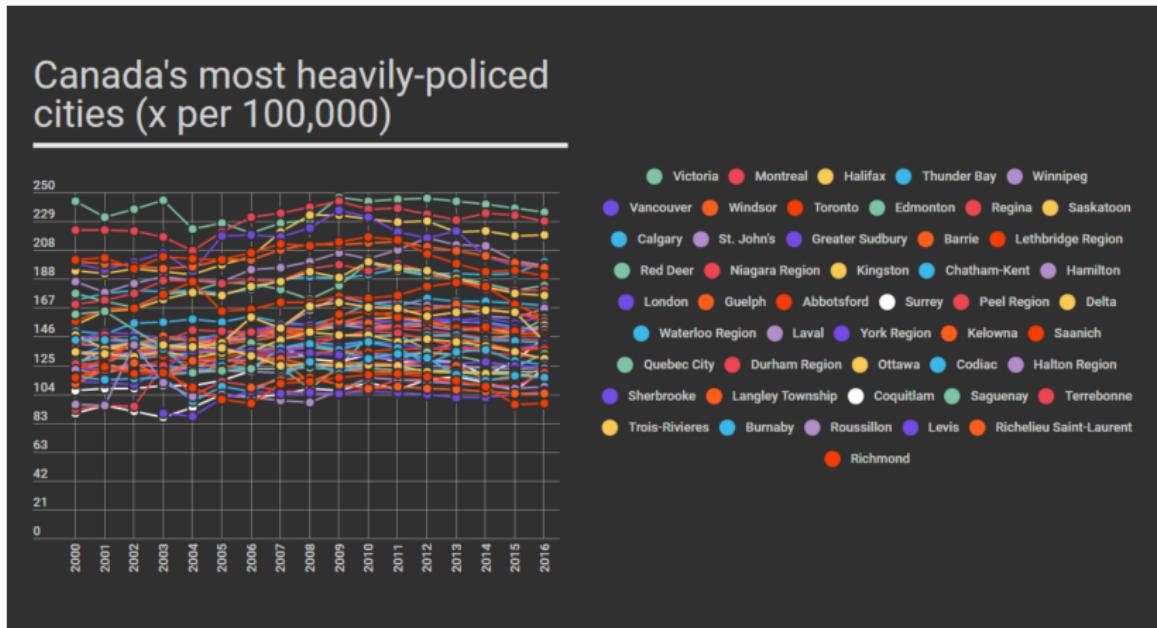
## Figure 38



# Figure 39



# Figure 43



## Figure 46

# 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

# Figure 47

## EXAMPLE: BIKE SHARE DATA

### Variables:

- Hour
- Median rentals
- Working day

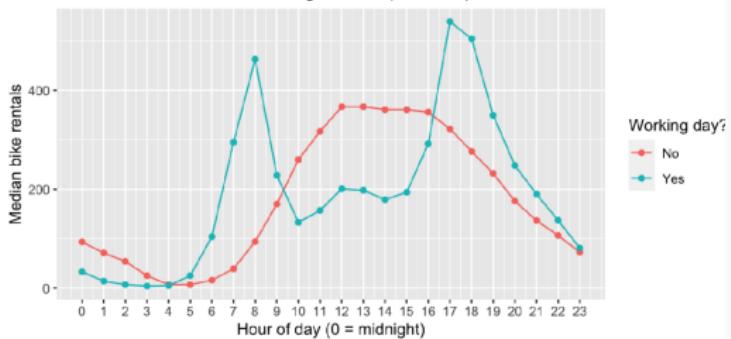
### Objects:

- Points/lines

### Mappings:

- Hour → x
- Median rentals → y
- Working day → color

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



Working day?

- No
- Yes

# Figure 48

## EXAMPLE: BIKE SHARE DATA

- Variables:

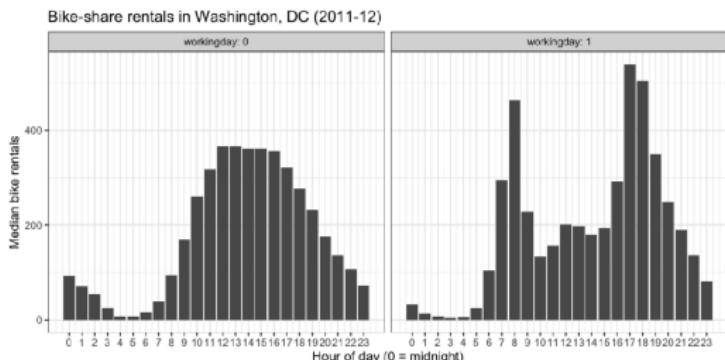
- Hour
- Median rentals
- Working day

- Objects:

- The two panels
- The bars in each panel

- Mappings:

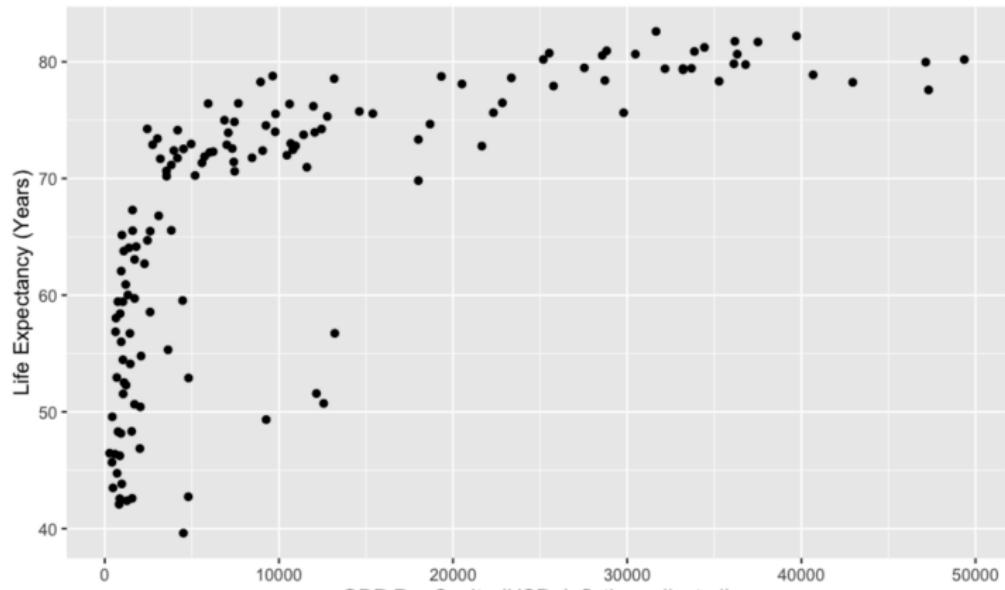
- Hour → location of bar
- Median rentals → height of bar
- Working day → left vs. right panel



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

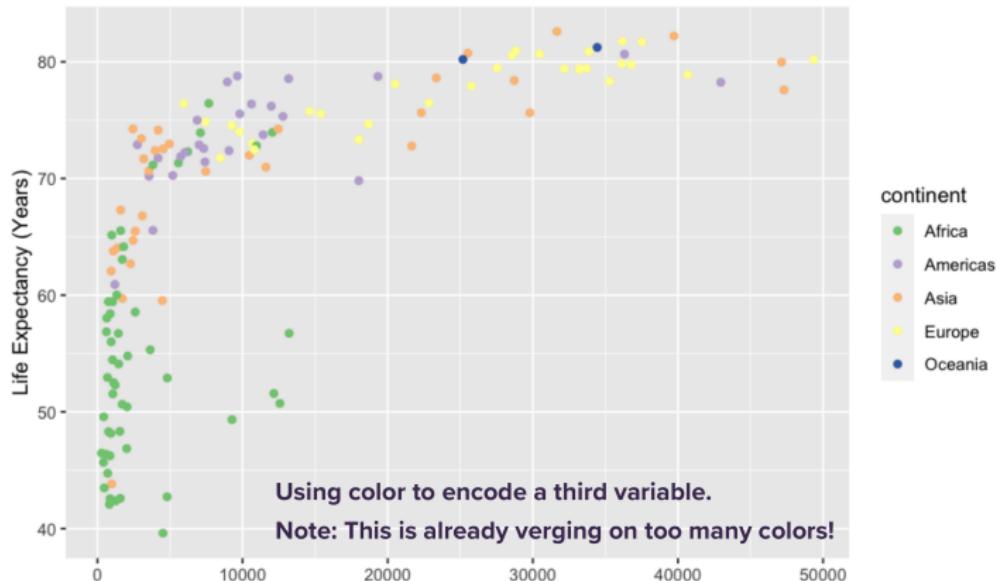
## Figure 53

Health versus Wealth for 142 Countries: 2007



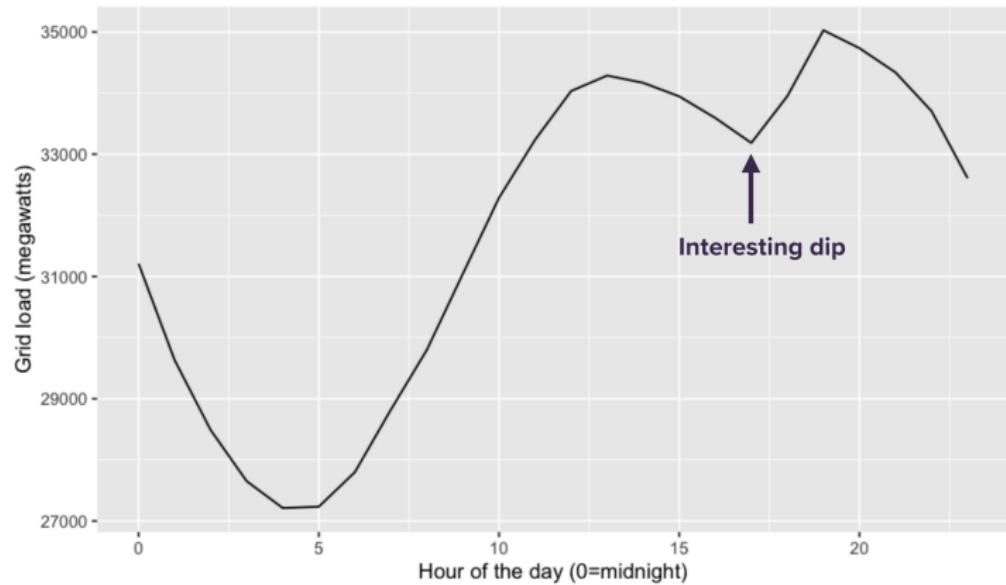
## Figure 54

### Health versus Wealth for 142 Countries: 2007

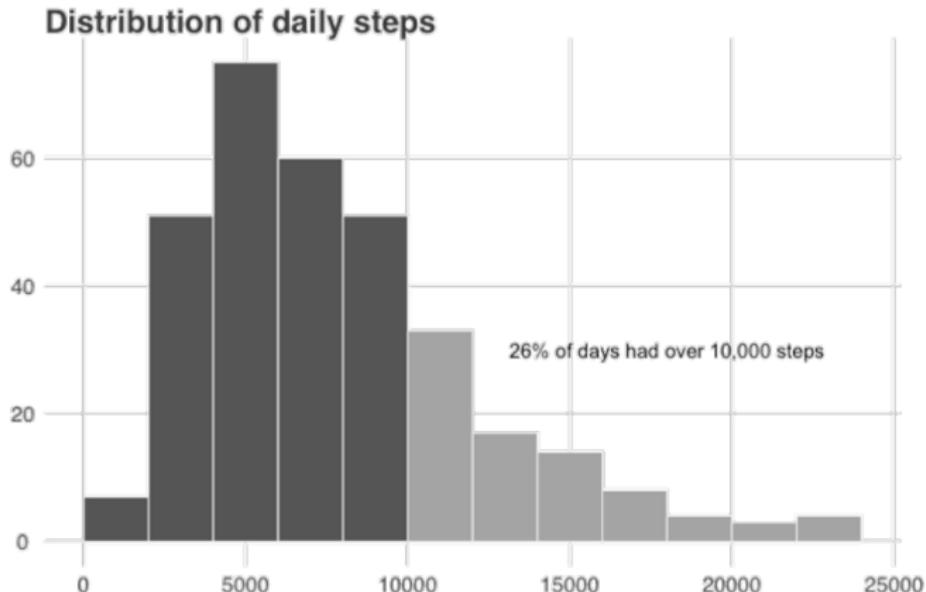


## Figure 58

Load on the Texas power grid, Christmas Day 2015



## Figure 61

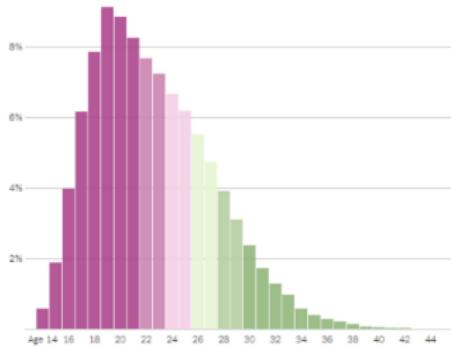


# Figure 62

## The Age That Women Have Babies: How a Gap Divides America

By QUOC-TRUNG BUI and CLAIRE CAIN MILLER AUG. 4, 2018

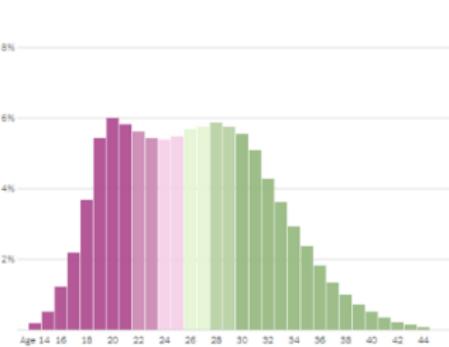
Ages of first-time mothers in 1980



Incorporate a categorical variable (here, two “snapshots” of 1980 and 2016) with side-by-side faceted histograms.

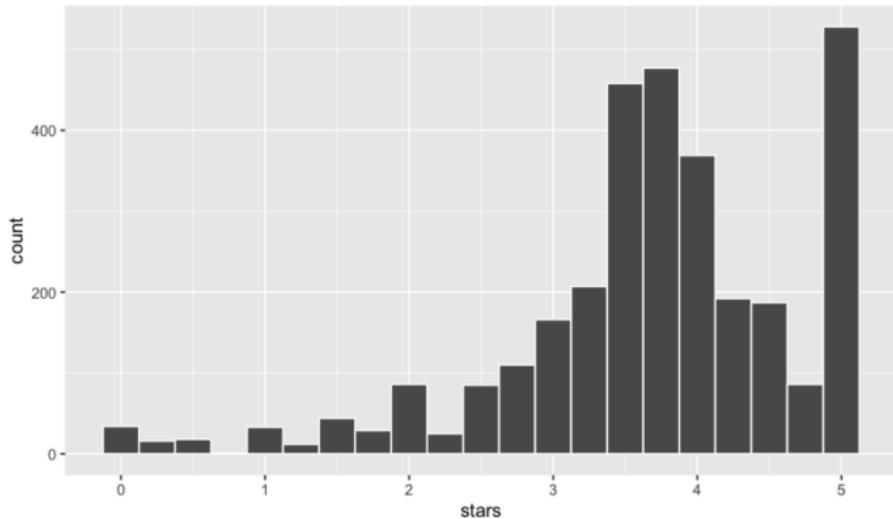
The colors allow you to compare specific age ranges across the two years. The proportion of 30+ year-old moms really stands out this way.

Ages of first-time mothers in 2016



## Figure 63

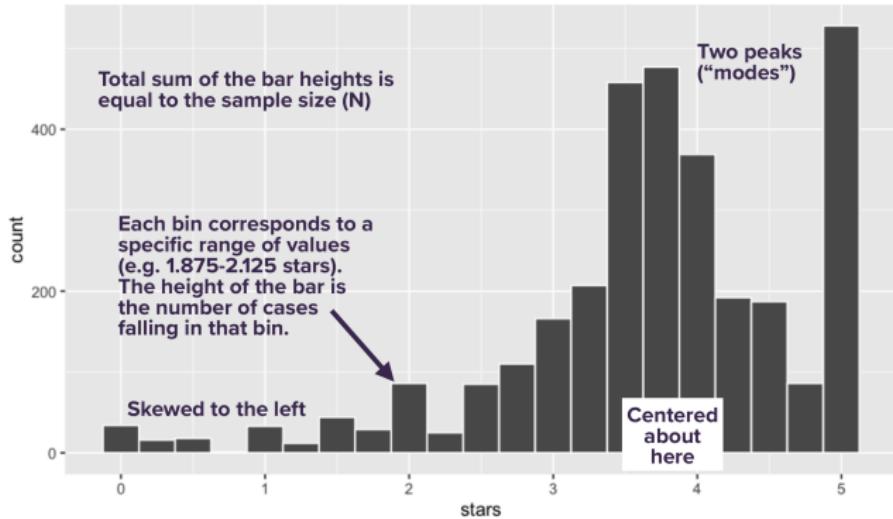
Distribution of ratings (0-5 scale) across 2764 ramen reviews



Data from The Ramen Rater, <https://www.theramenrater.com/resources-2/the-list/>

# Figure 64

Distribution of ratings (0-5 scale) across 2764 ramen reviews

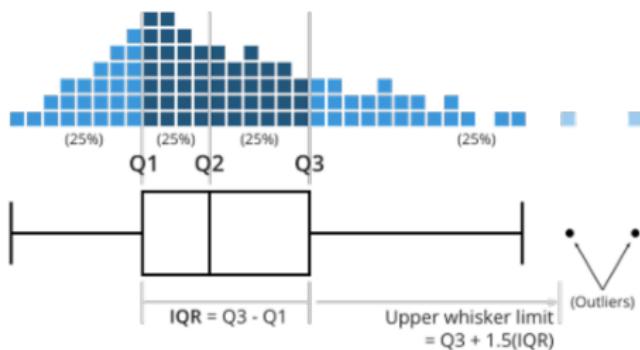


Data from The Ramen Rater, <https://www.theramenrater.com/resources-2/the-list/>

# Figure 70

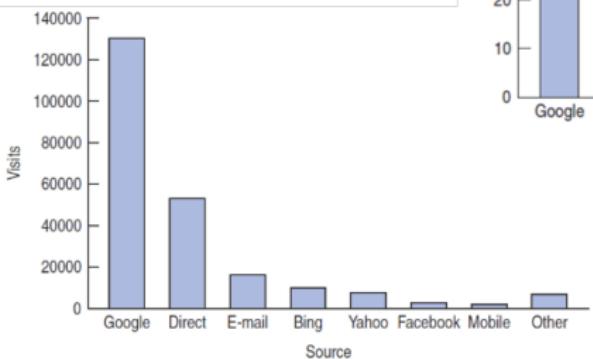
## BOXPLOTS

- The central box shows the middle half of the data: 25th to 75th percentiles, the interquartile range (IQR).
- If the median is roughly centered in the box, the middle half of the data is roughly symmetric. Here we see some mild skewness.
- The whiskers also show skewness based on comparing their lengths.
- A common (though not universal) convention is to cut the whiskers off at 1.5 times the IQR, rather than min/max.

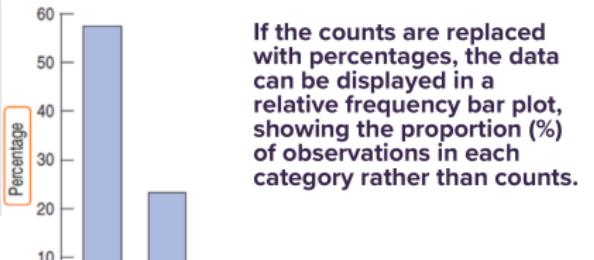


## Figure 75

A bar plot displays the distribution of a categorical variable, showing the counts for each category next to each other for easy comparison.



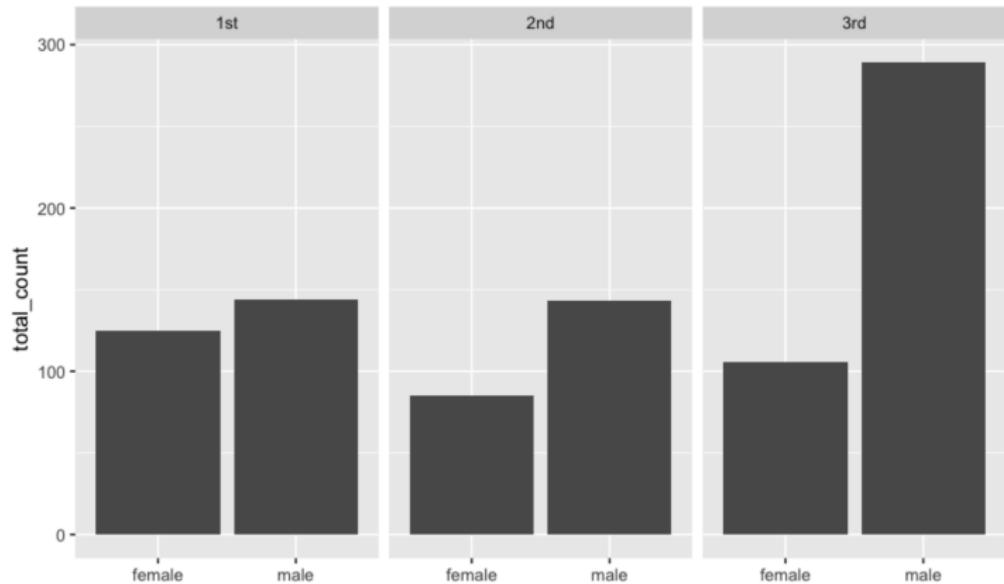
If the counts are replaced with percentages, the data can be displayed in a relative frequency bar plot, showing the proportion (%) of observations in each category rather than counts.



Source	Visits	Visits by %
Google	130,758	57.36
Direct	52,969	23.34
E-mail	16,084	7.09
Bing	9,581	4.22
Yahoo	7,439	3.28
Facebook	2,253	0.99
Mobile	1,701	0.75
Other	6,740	2.97
Total	226,925	100.00

# Figure 76

## Who sailed the Titanic? Passenger count by class and sex



## Figure 77

Who survived the Titanic? Survival percentages by class and sex

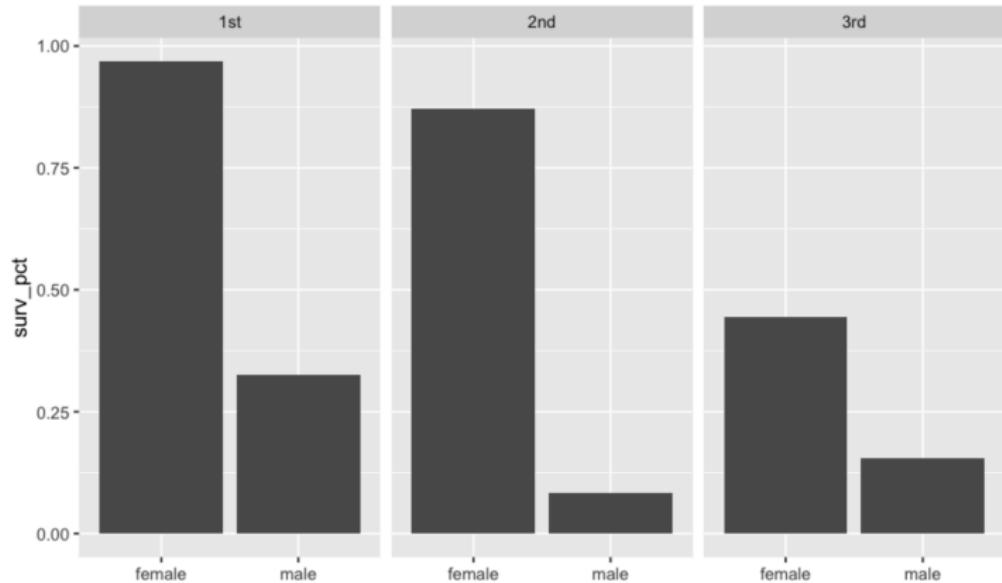


Figure 5

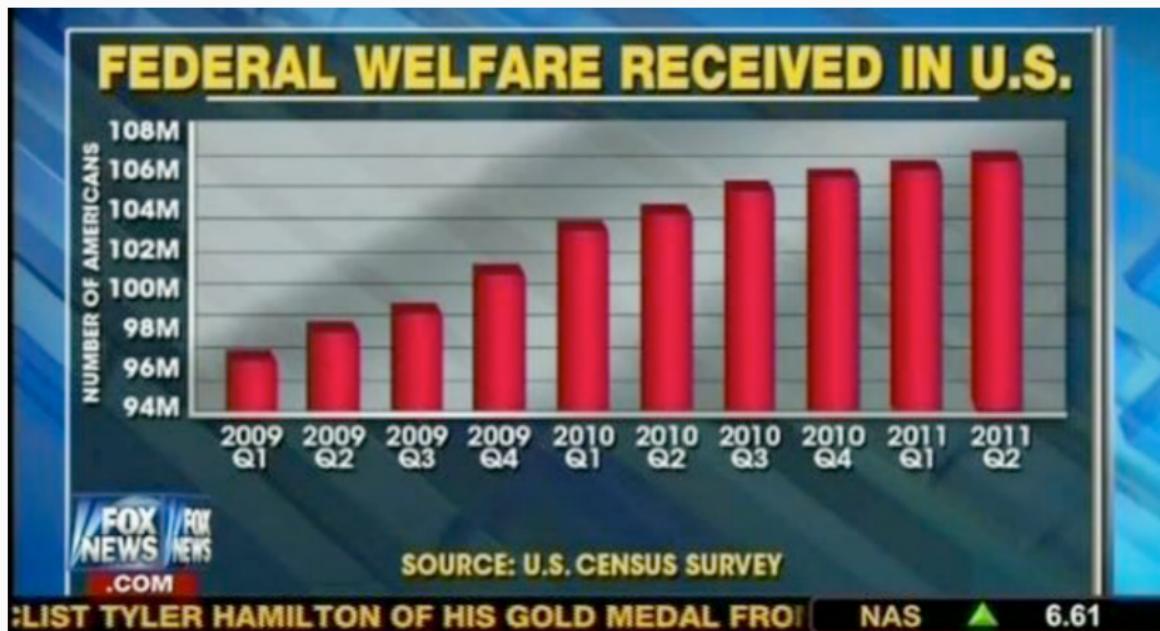


Figure 6

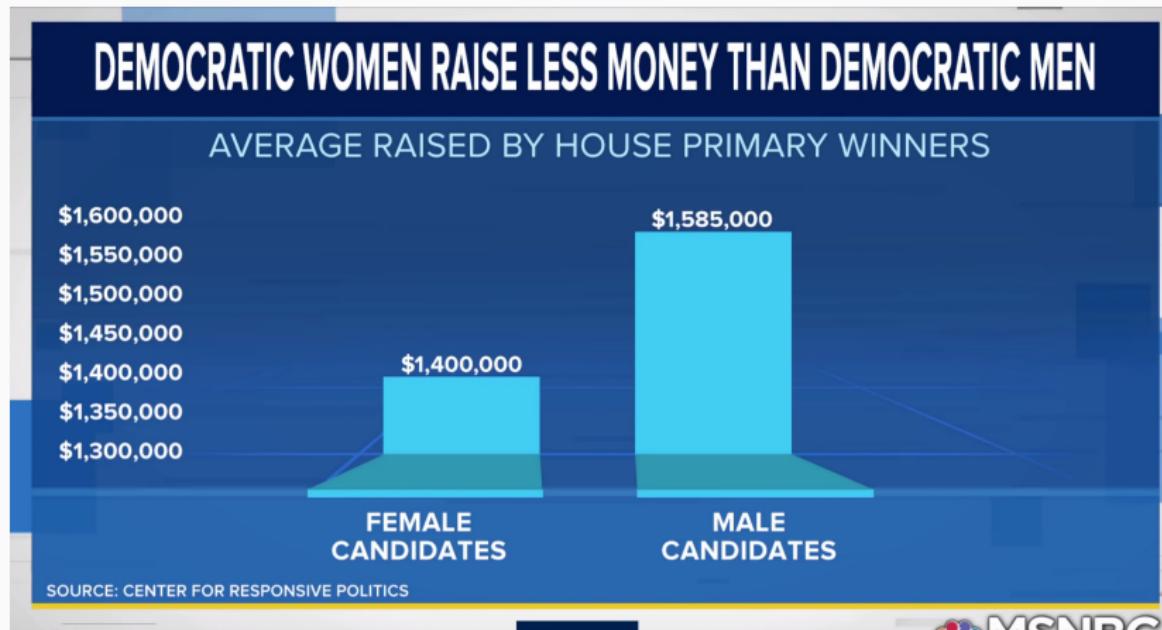
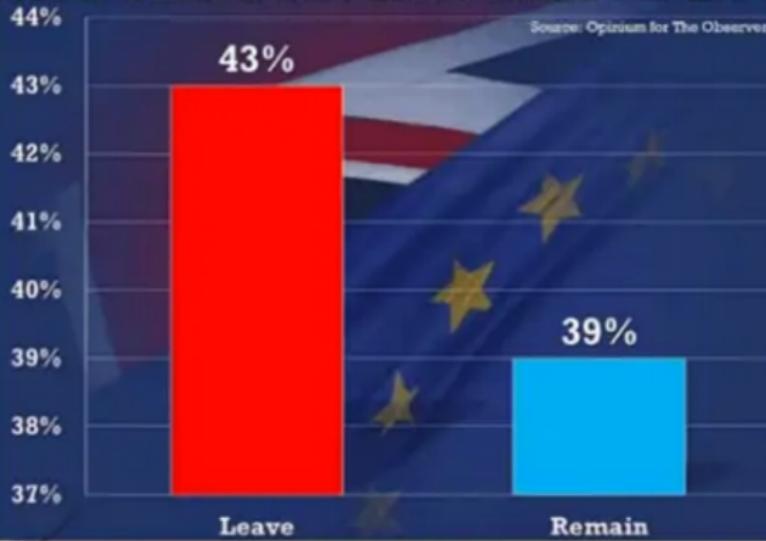
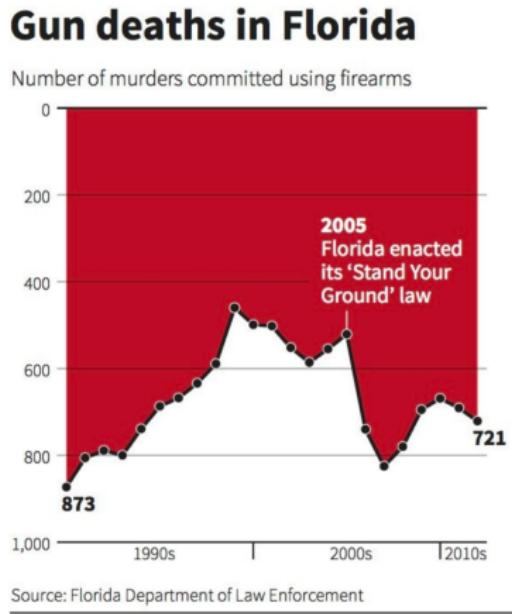


Figure 7

## SHOULD BRITAIN LEAVE EU?



**Figure 8**



111