

Calling Bullsh*t on Data Visualisations!

CM4125 – Topic 3

What this lecture is about?

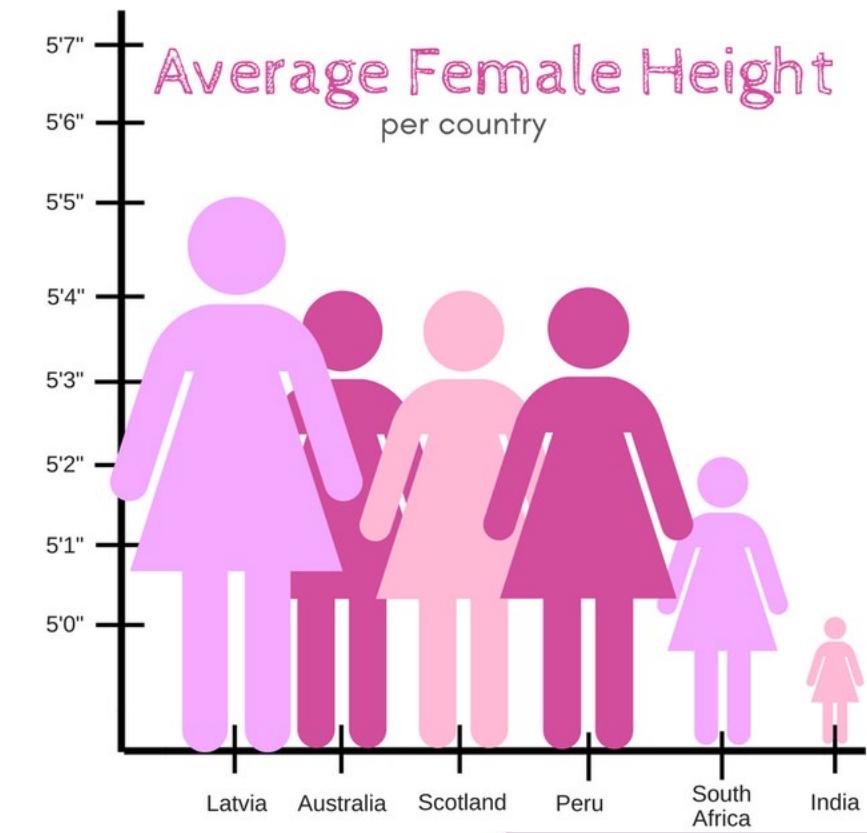
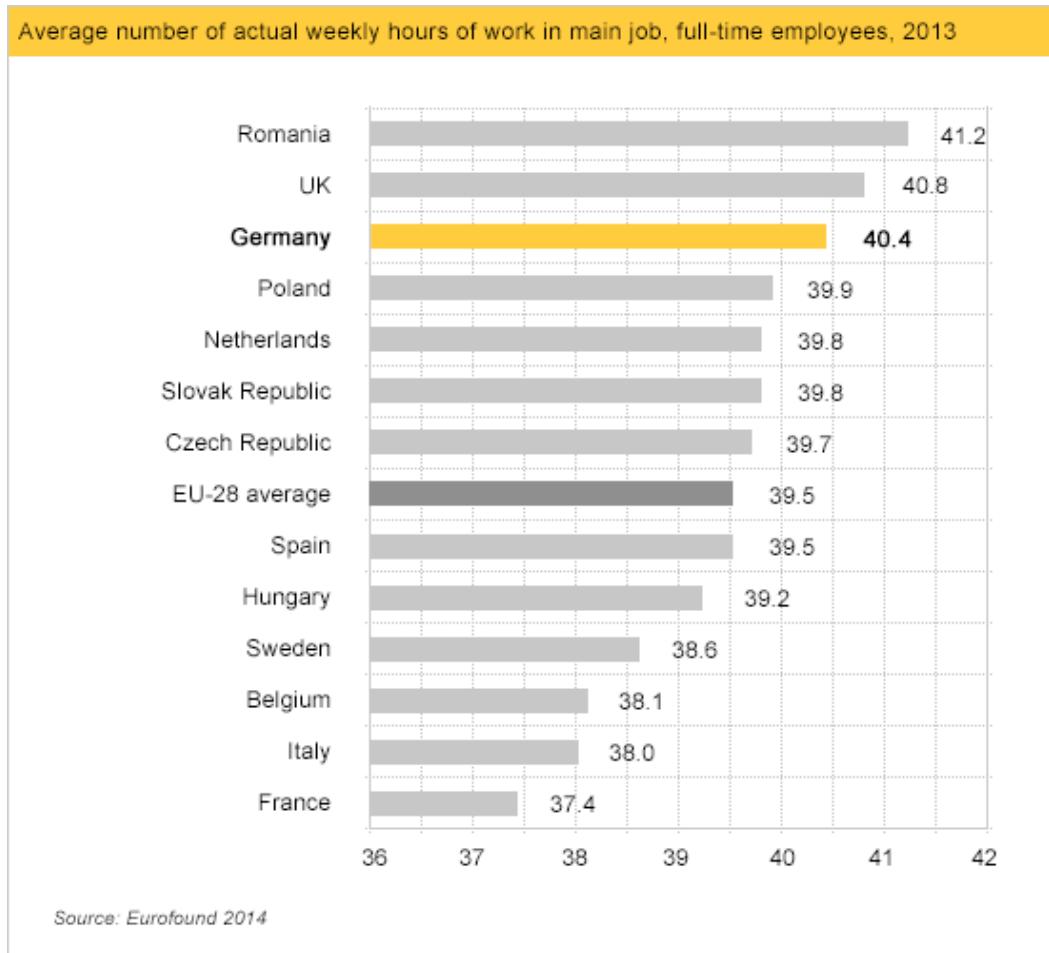
- Based on Week 6 of the Calling Bullsh*t course of the University of Washington by Carl T. Bergstrom and Jevin West
 - <https://www.callingbullshit.org/syllabus.html#Visual>
- Supplementary reading:
 - Alberto Cairo (2019) *How Charts Lie: Getting Smarter about Visual Information*. W.W. Norton and Company.
 - Edward Tufte (1983) *The Visual Display of Quantitative Information*. Chapters 2 (Graphical integrity) and 5 (Chartjunk: vibrations, grids, and ducks).

Misleading axis

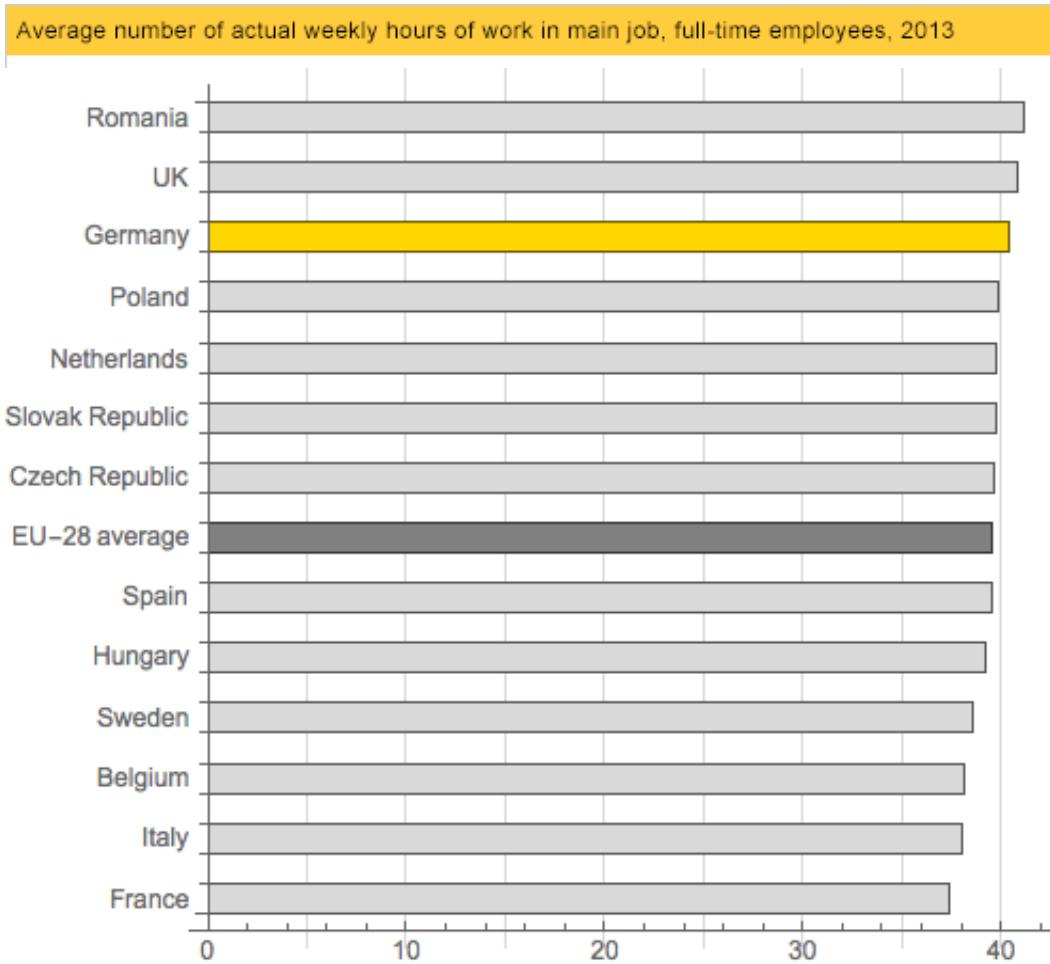
Charts can conceal or mislead if axes are set incorrectly

How to choose the range and scale of axes in a graph?

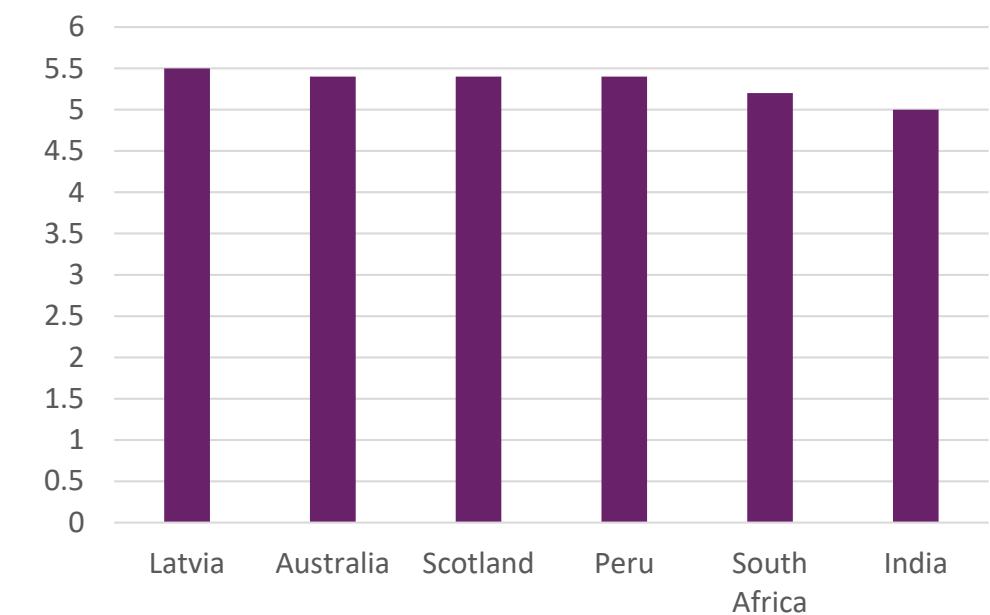
1. Bar chart axes need to include zero



1. Bar chart axes need to include zero



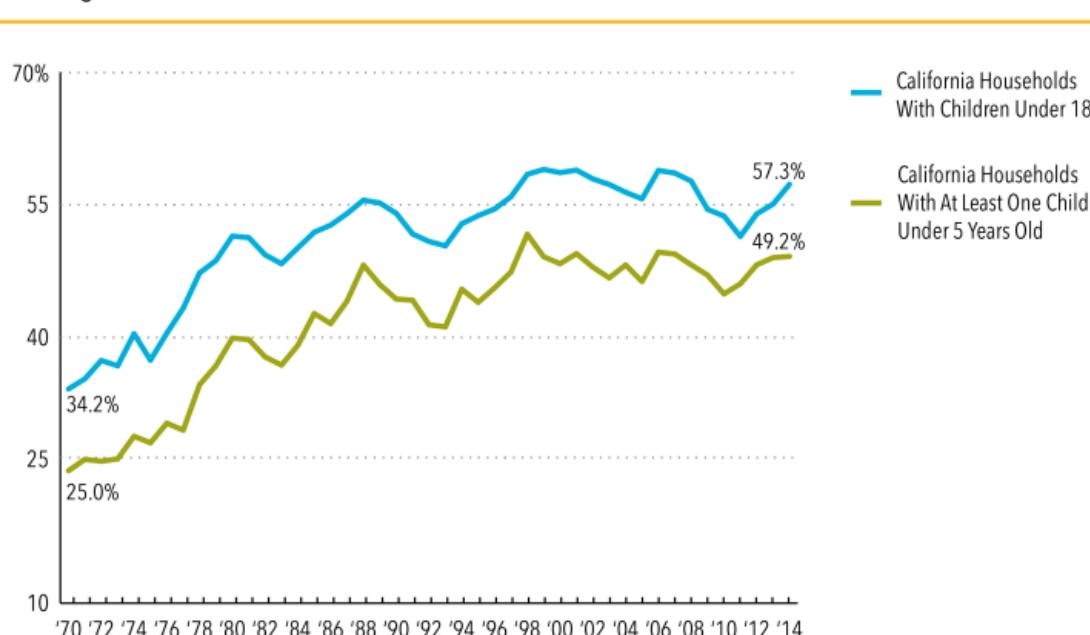
Average Female Height
Per country



2. Line graph axes don't need to include zero

More California Households Have All Parents Working, Making Access to Child Care an Important Priority

Percentage of California Households Where All Parents Work, 1970 to 2014

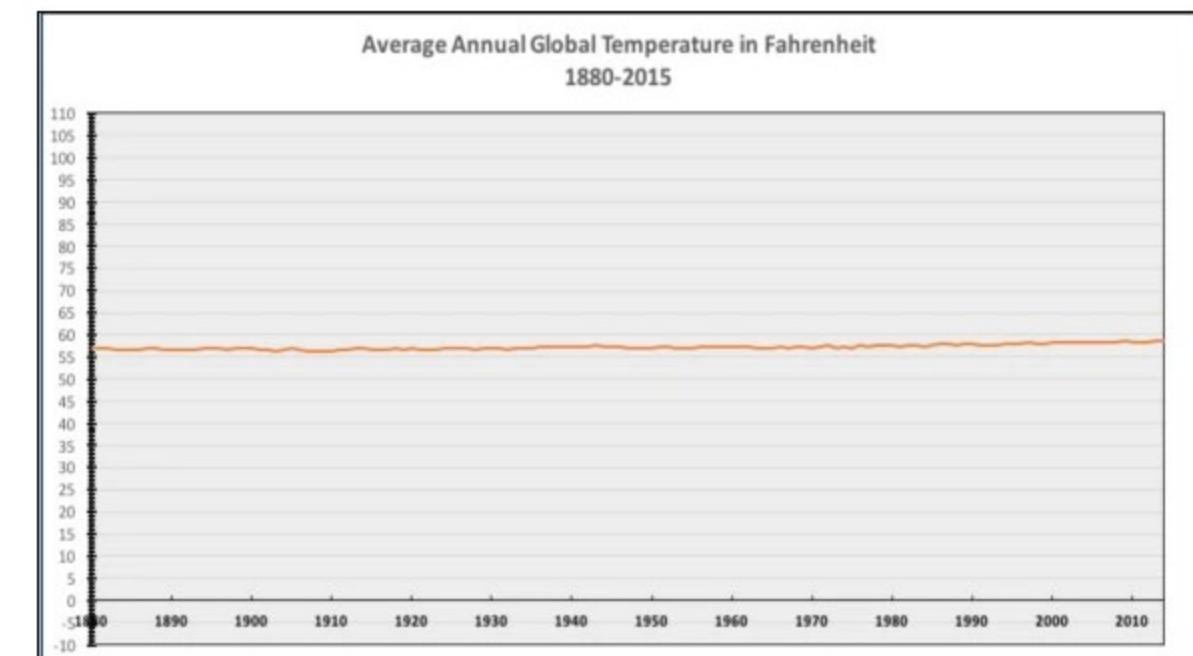


Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.

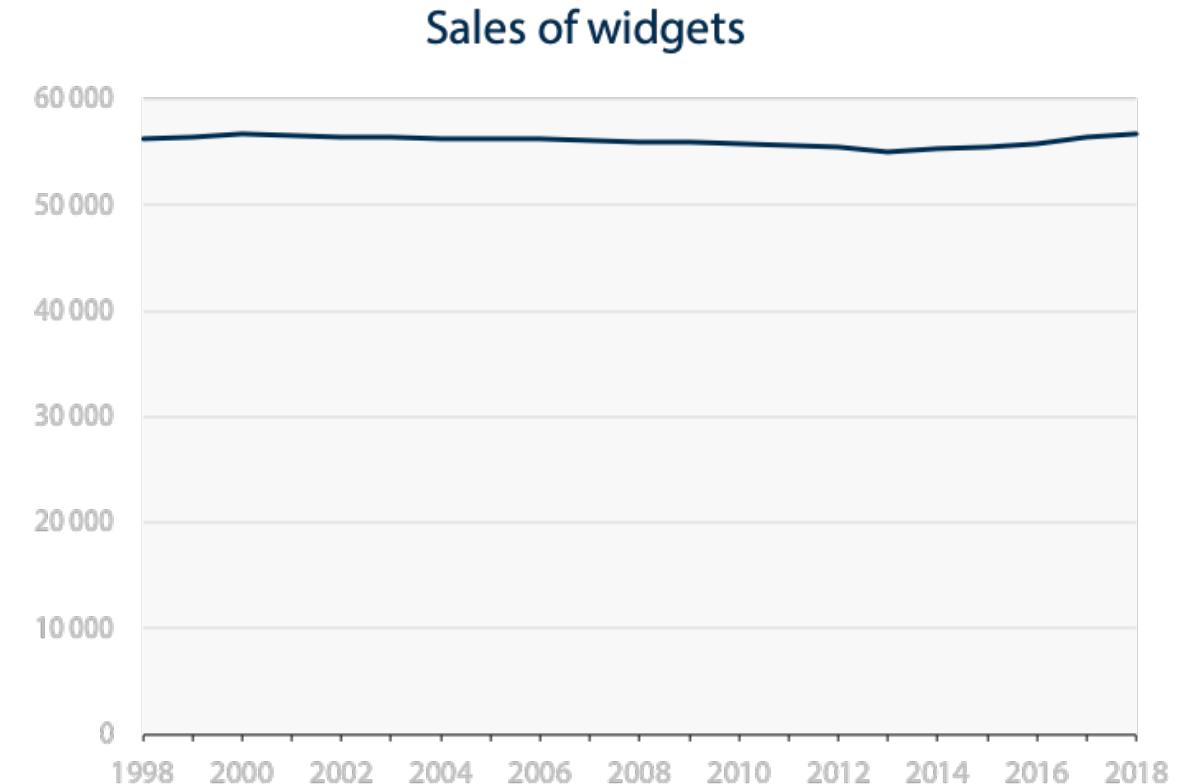
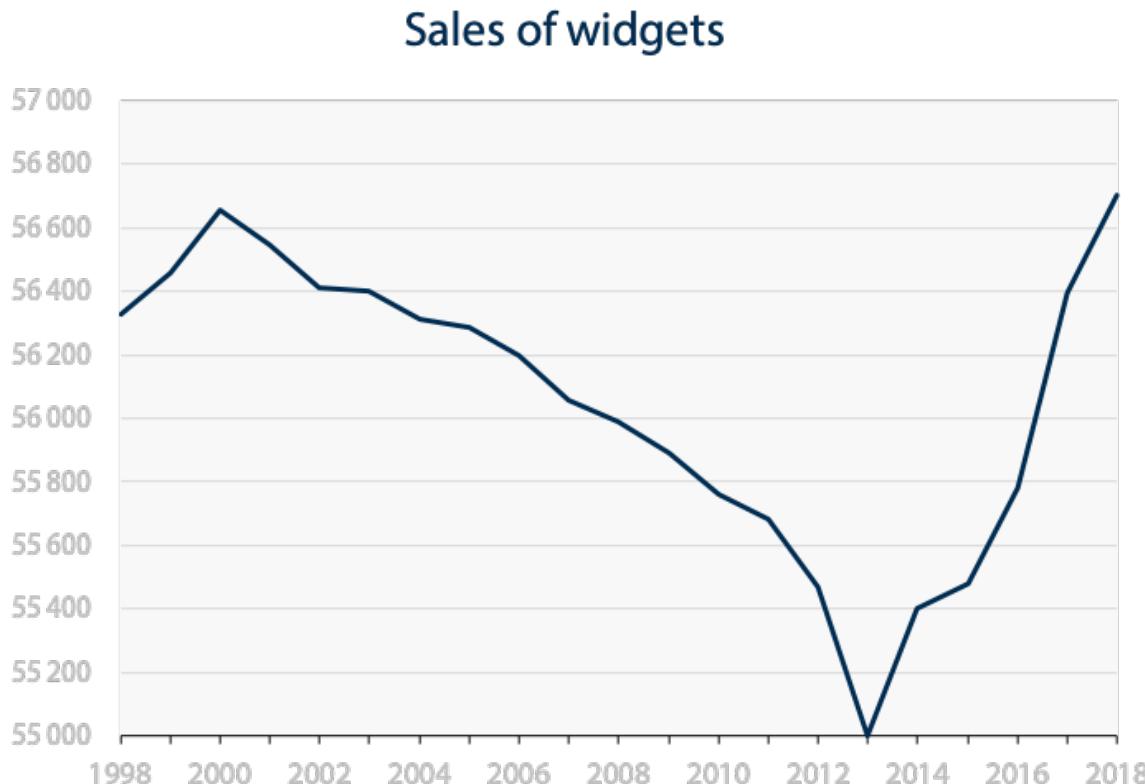
Source: Budget Center analysis of US Census Bureau data



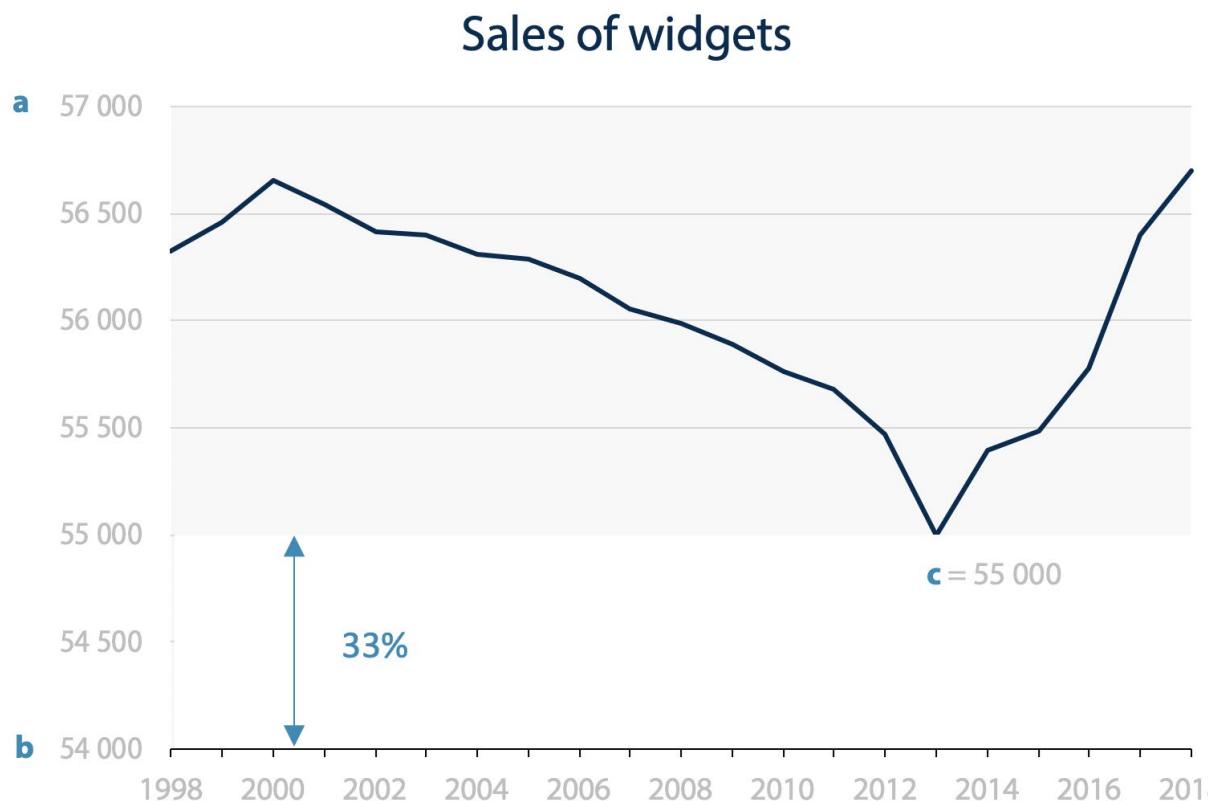
California Budget
& Policy Center
Independent Analysis. Shared Prosperity.



Well...



Solution



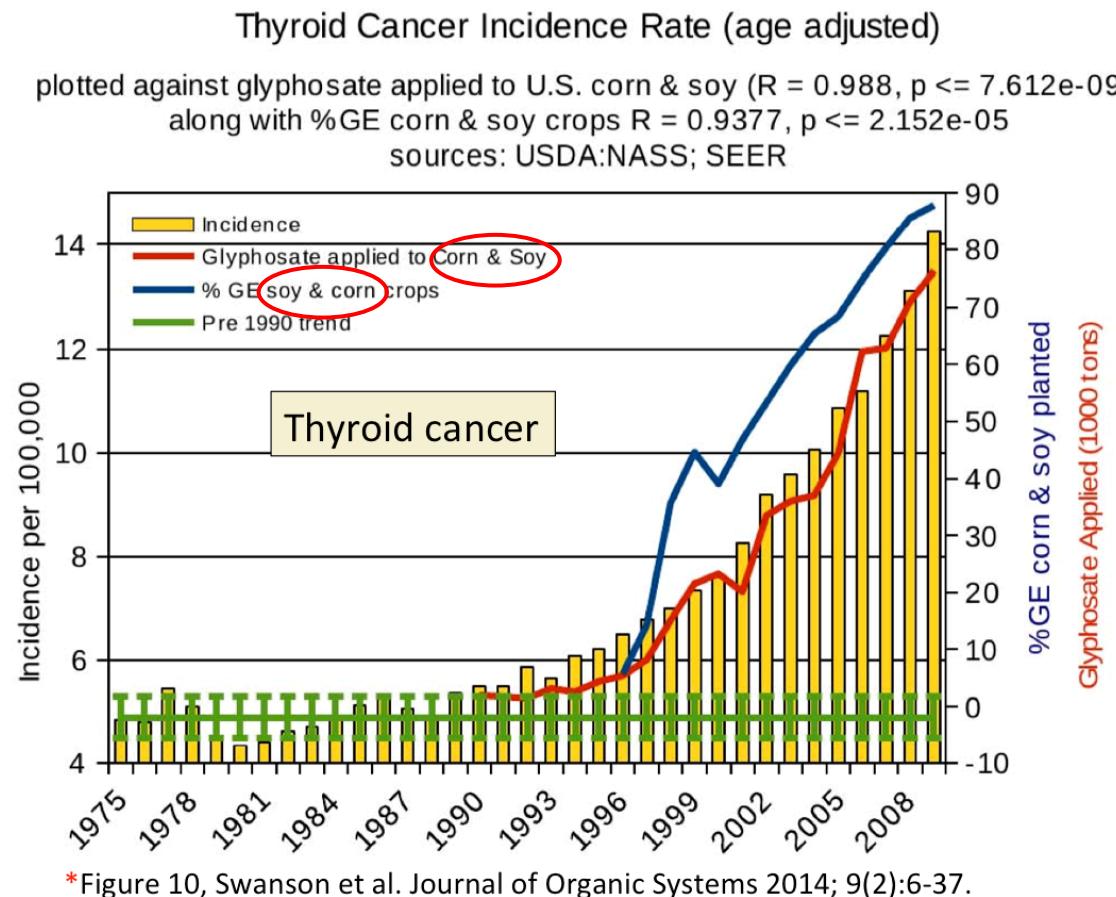
Empty space =
$$\frac{\text{Minimum data value} - \text{Minimum scale value}}{\text{Maximum scale value} - \text{Minimum scale value}}$$

$$x = \frac{(c - b)}{(a - b)}$$

$$b = \frac{3c - a}{2}$$

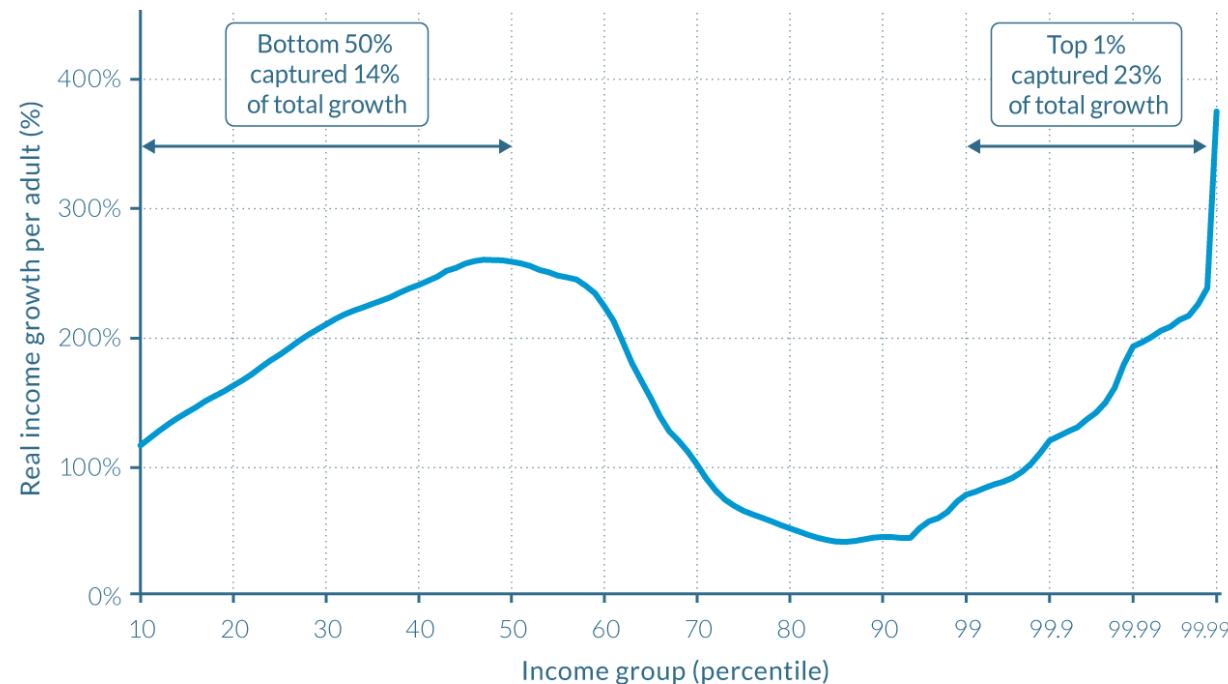
$$b = \frac{3 * 55,000 - 57000}{2} = 54\,000$$

3. No multiple axes on a single graph



How to solve this?

4. An axis should not change scales midstream

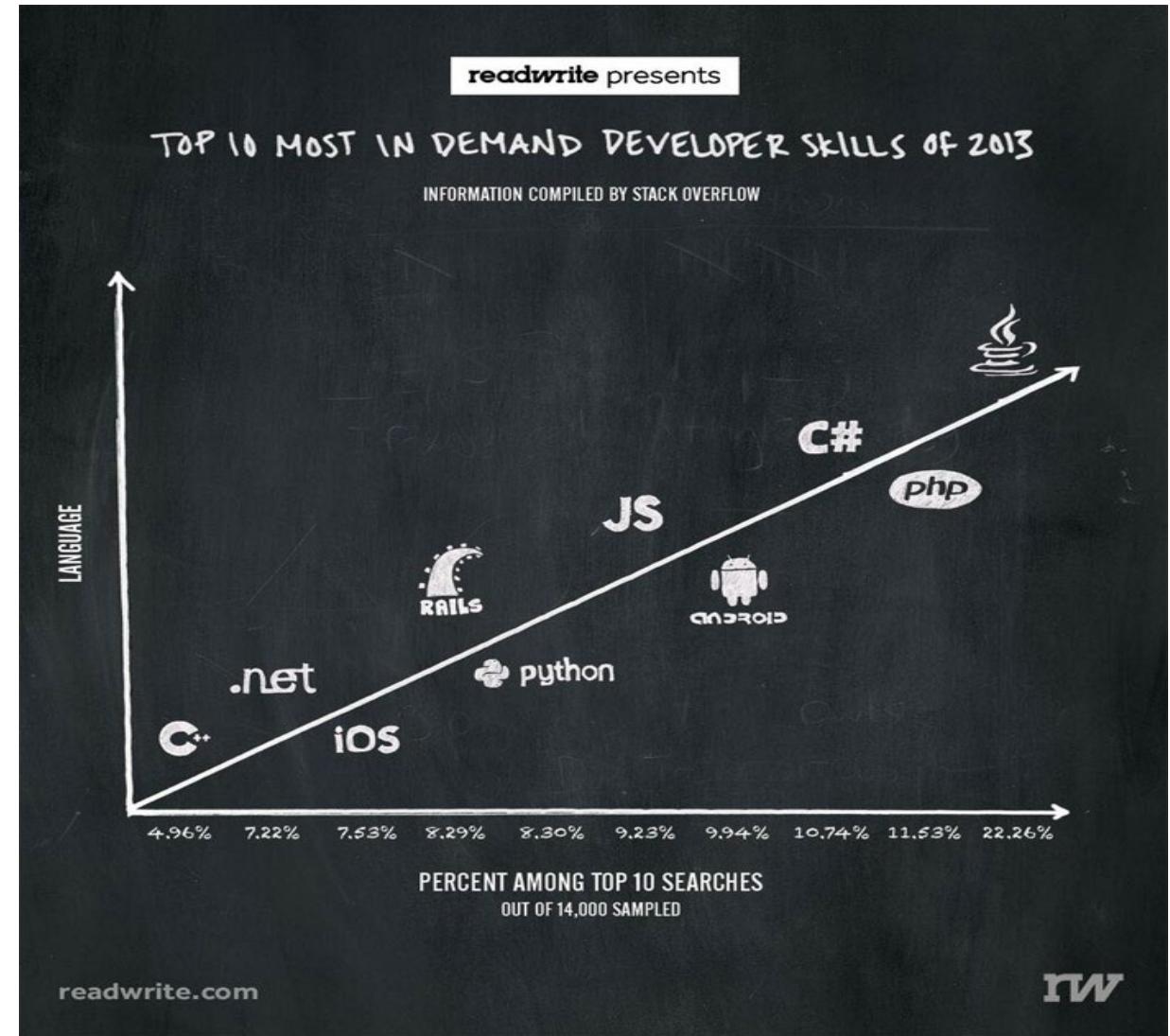


Source: WID.world (2017). See wid.world/methodology.html for data series and notes.

On the horizontal axis, the world population is divided into a hundred groups of equal population size and sorted in ascending order from left to right, according to each group's income level. The Top 1% group is divided into ten groups, the richest of these groups is also divided into ten groups, and the very top group is again divided into ten groups of equal population size. The vertical axis shows the total income growth of an average individual in each group between 1980 and 2016. For percentile group p99p99.1 (the poorest 10% among the world's richest 1%), growth was 77% between 1980 and 2016. The Top 1% captured 23% of total growth over this period. Income estimates account for differences in the cost of living between countries. Values are net of inflation.

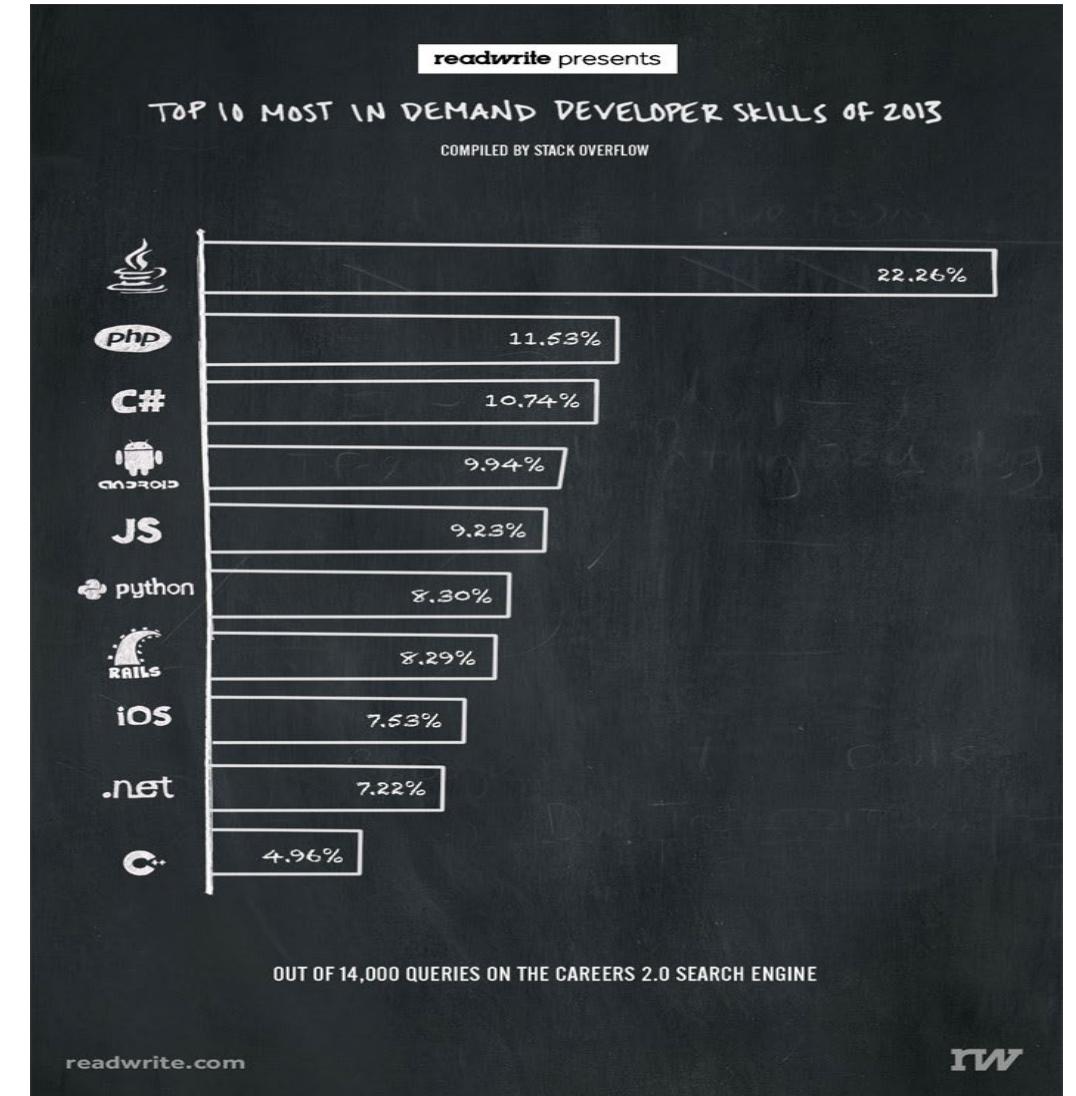
Use log scales!

5. An axis should have *something* on it



Like this

Tools such as
[WebPlotDigitizer](#) can
help you get the
missing info

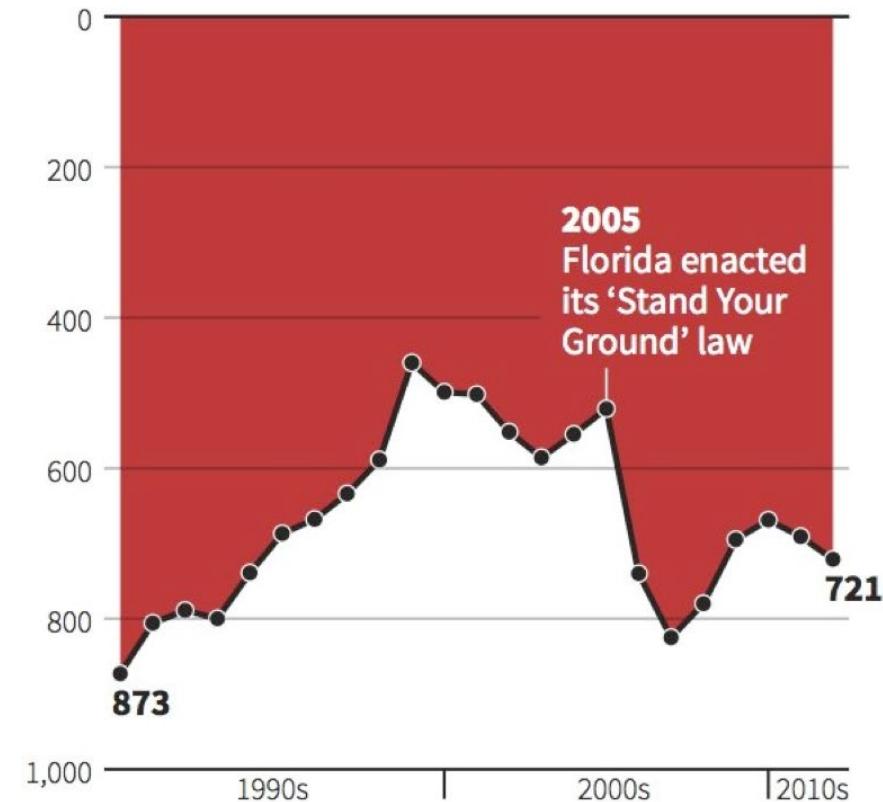


6. Don't invert the axis(?)

The author defended it!
What's your stance?

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

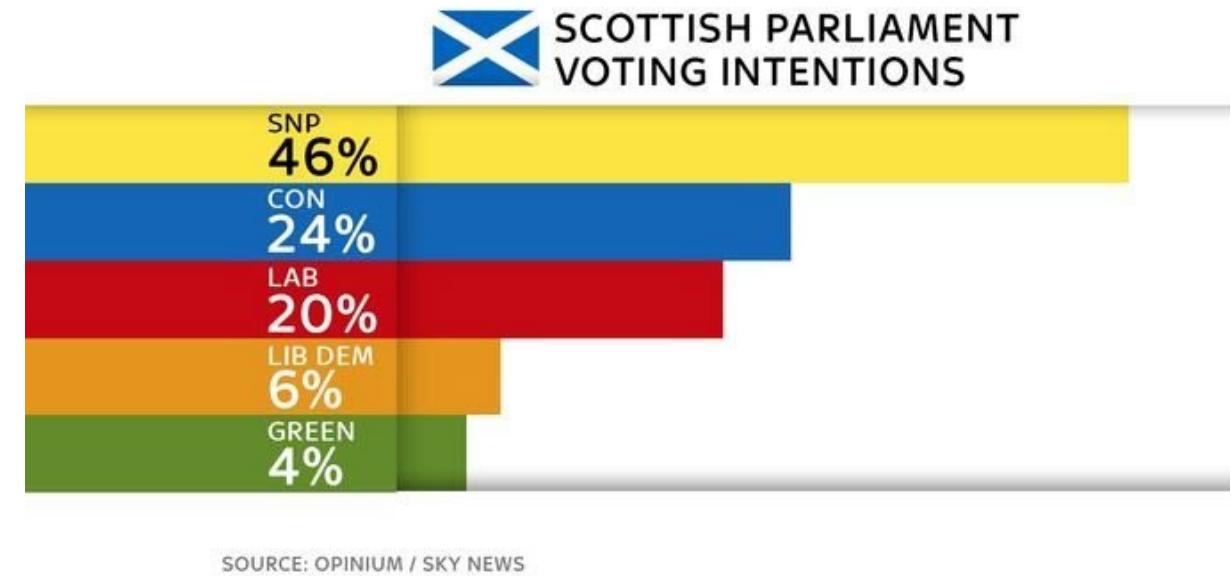
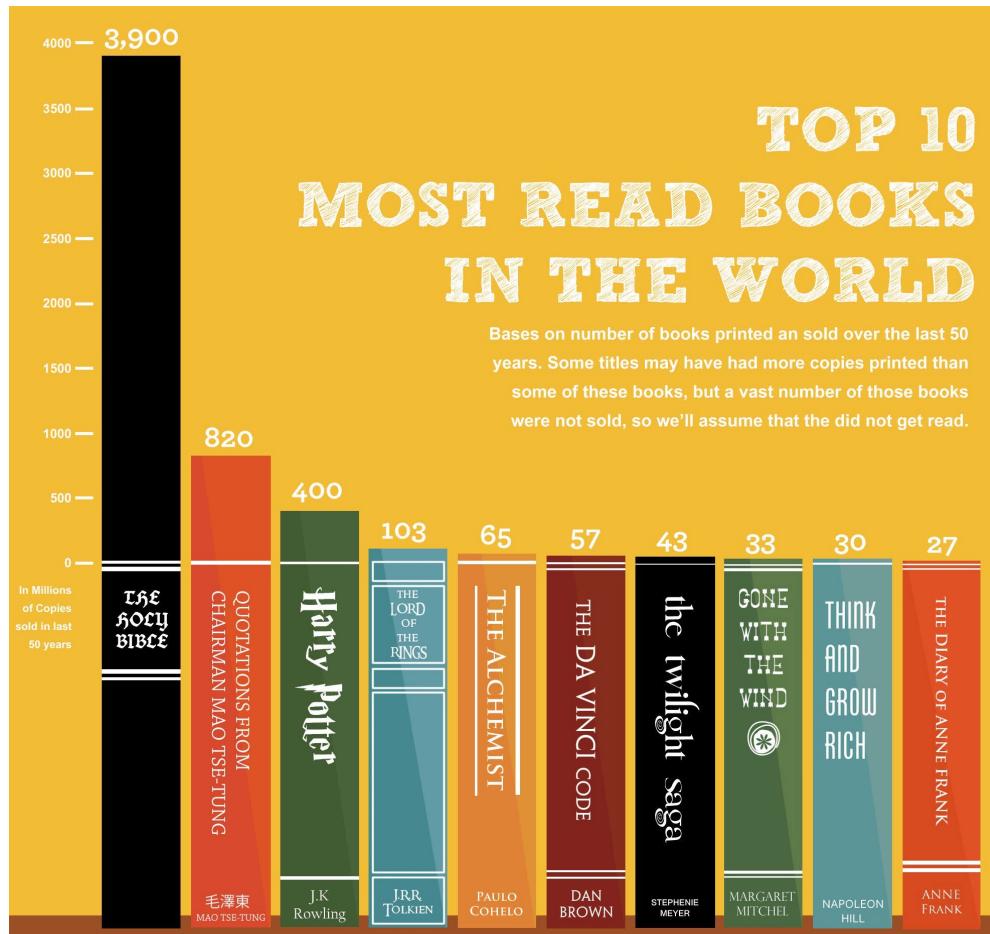
C. Chan 16/02/2014

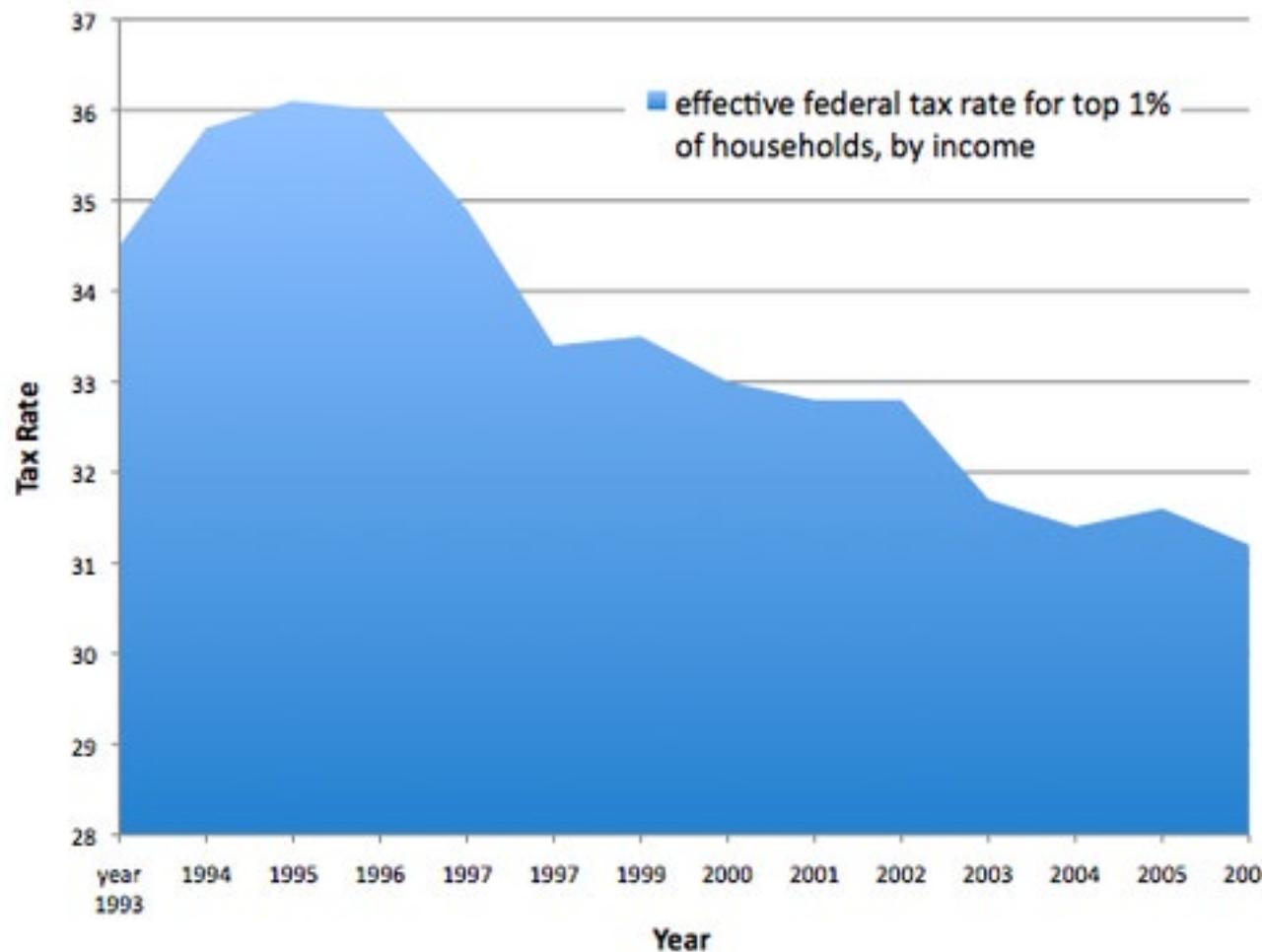
REUTERS

Proportional Ink

“When a shaded region is used to represent a numerical value, the area of that shaded region should be directly proportional to the corresponding value”

Extends the argument for misleading axes





Life expectancy, years 

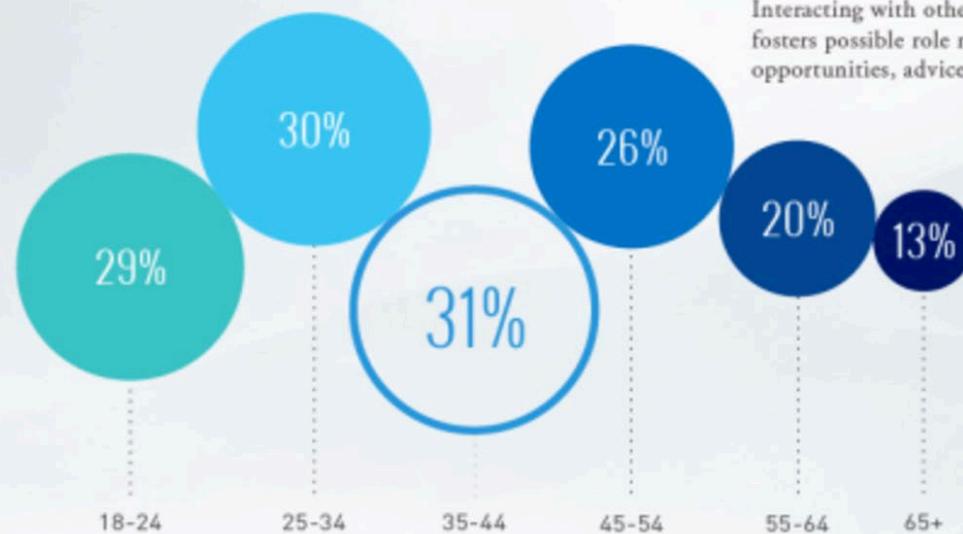


Should we violate
the principle?

Radius ≠ Area

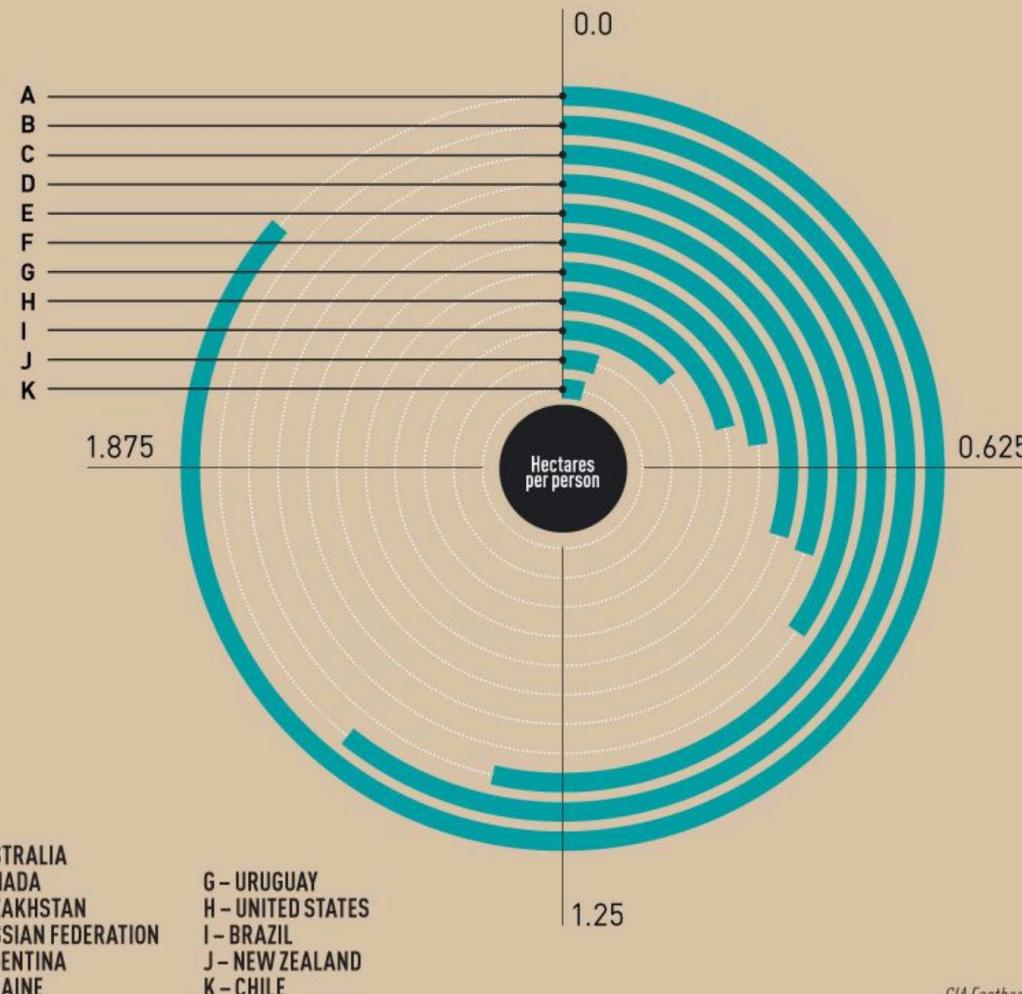
It's Who You Know

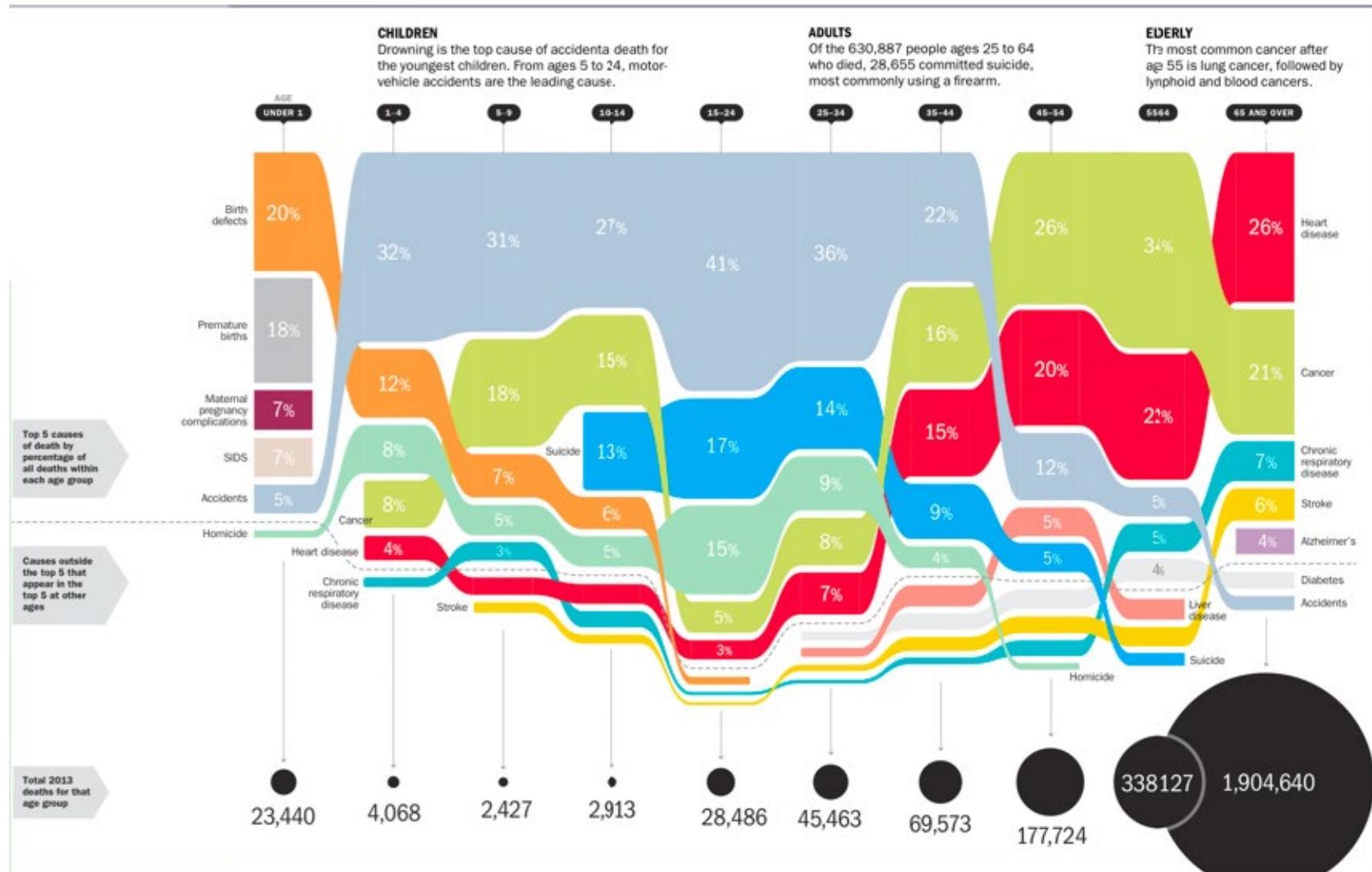
By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these personal social networks begin to decline over time.



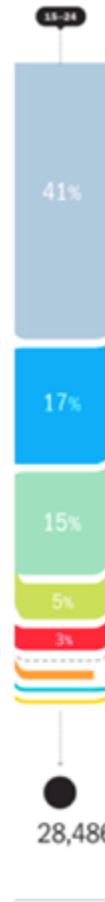
Interacting with other entrepreneurs fosters possible role models, networking opportunities, advice, and encouragement.

ARABLE LAND PER CAPITA (HECTARES PER PERSON)

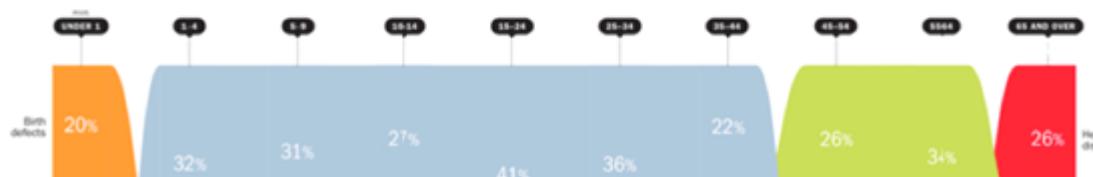




Along vertical slices, ink is proportional to value because shaded areas represent the fraction of a fixed number of deaths (here 28,486) from each cause.

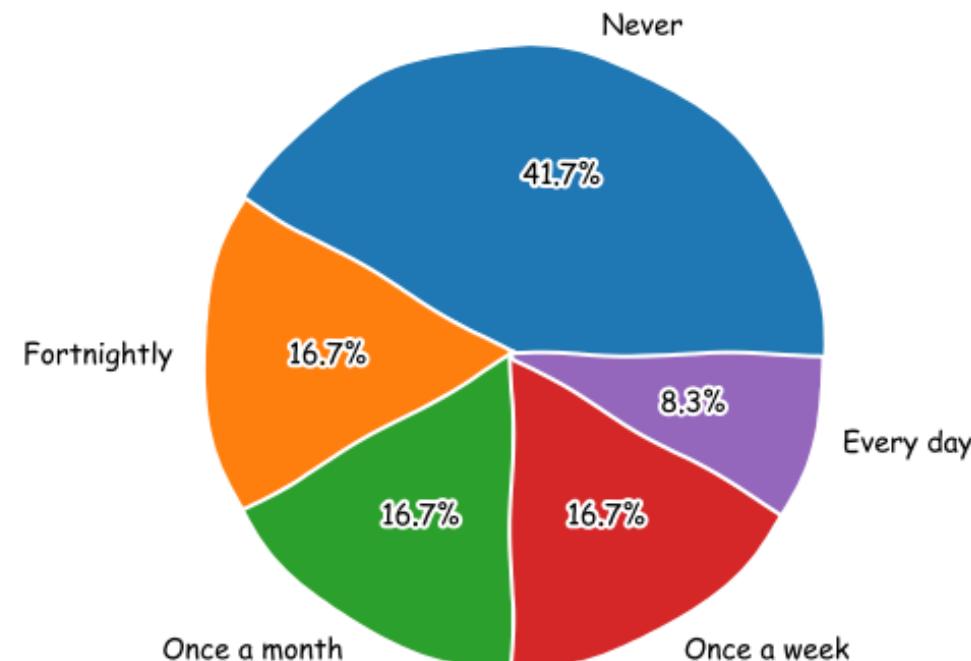


Along horizontal slices, ink is not proportional because total deaths differ widely by age group. Far more people 65 and older died of heart disease (red) than children age 1-4 die of accidents (blue-gray), but the latter takes more ink because it represents a larger percentage of the (relatively few) total deaths at that age.

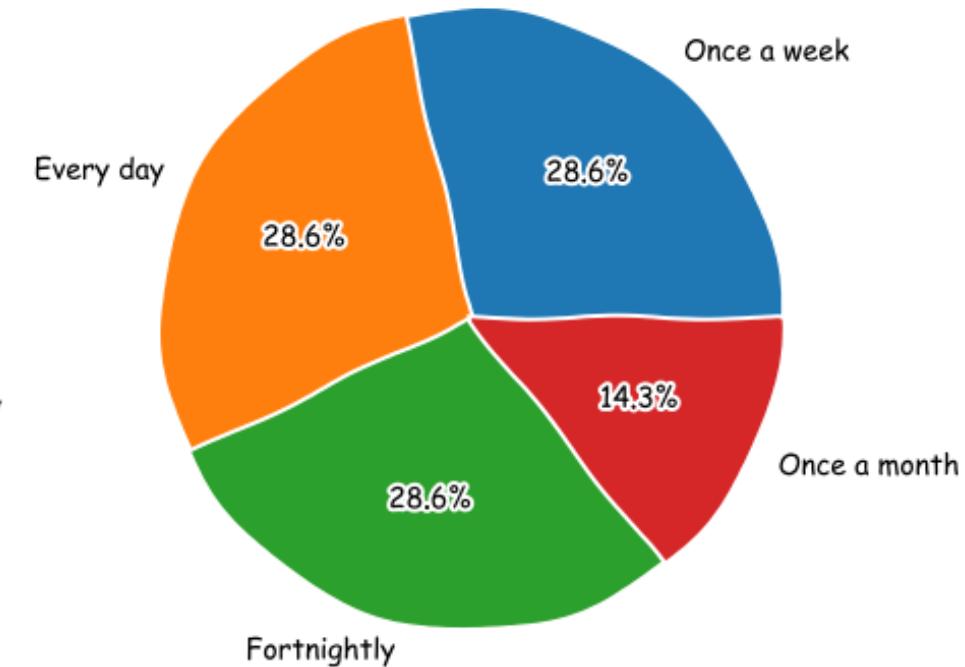


How often do you use a dating app during a month?

Male



Female



What is wrong in
this one?

Perspective

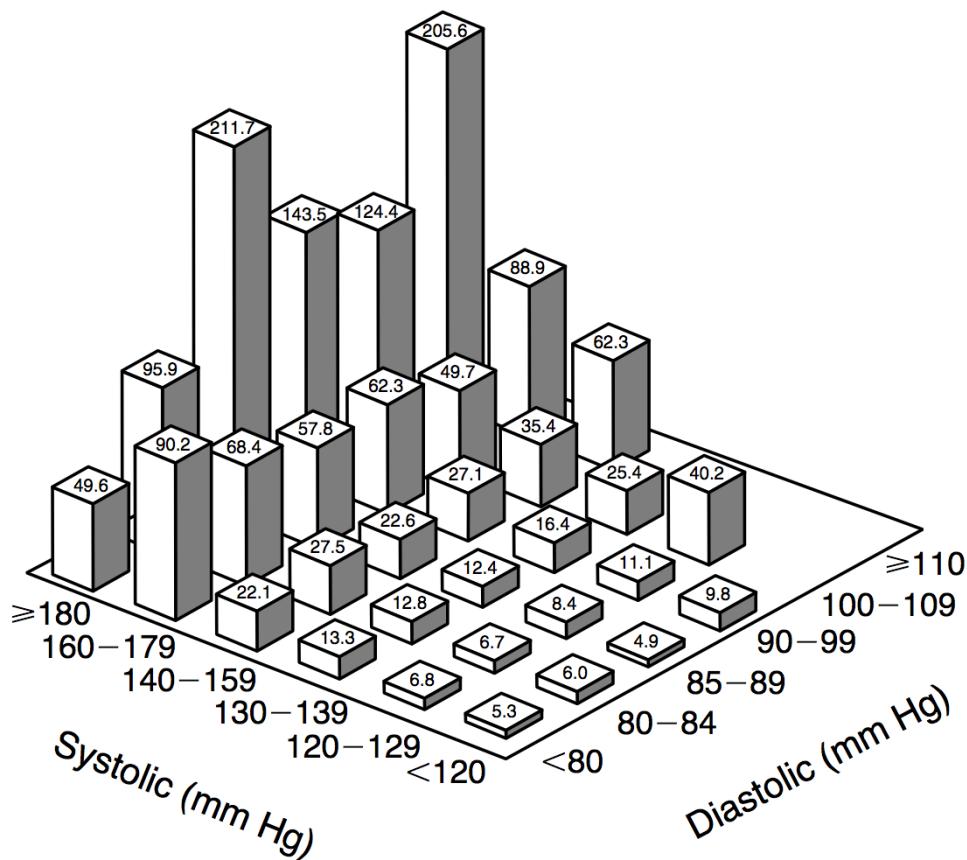
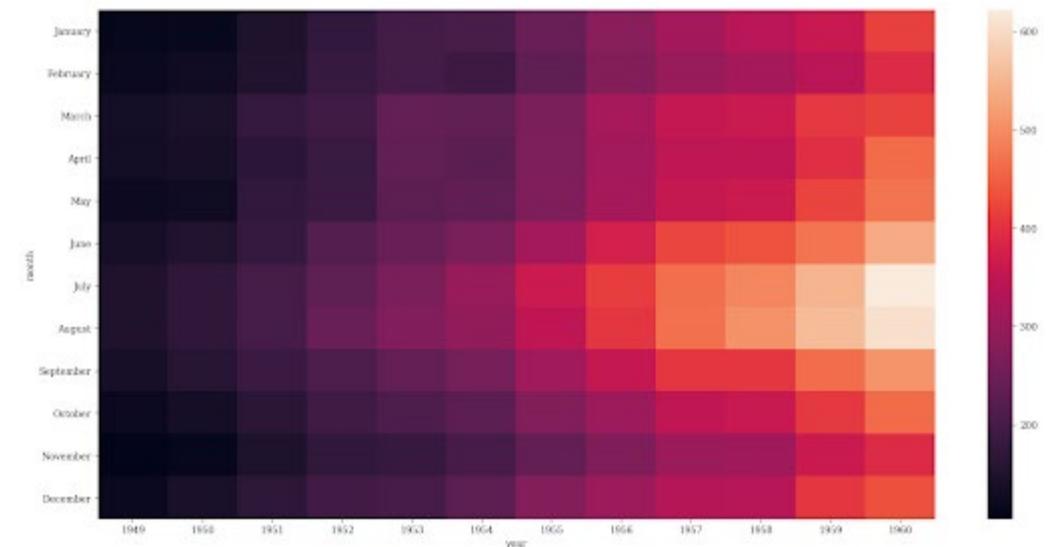
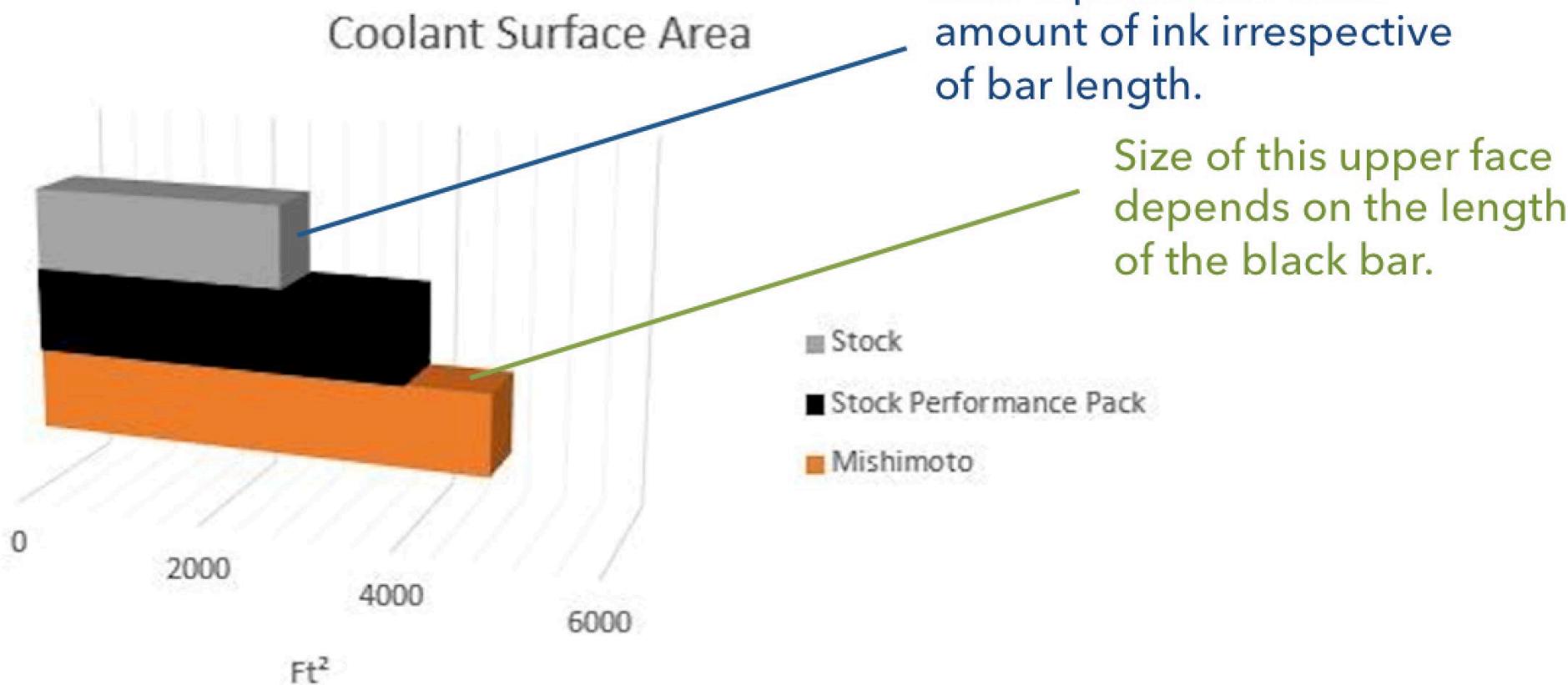
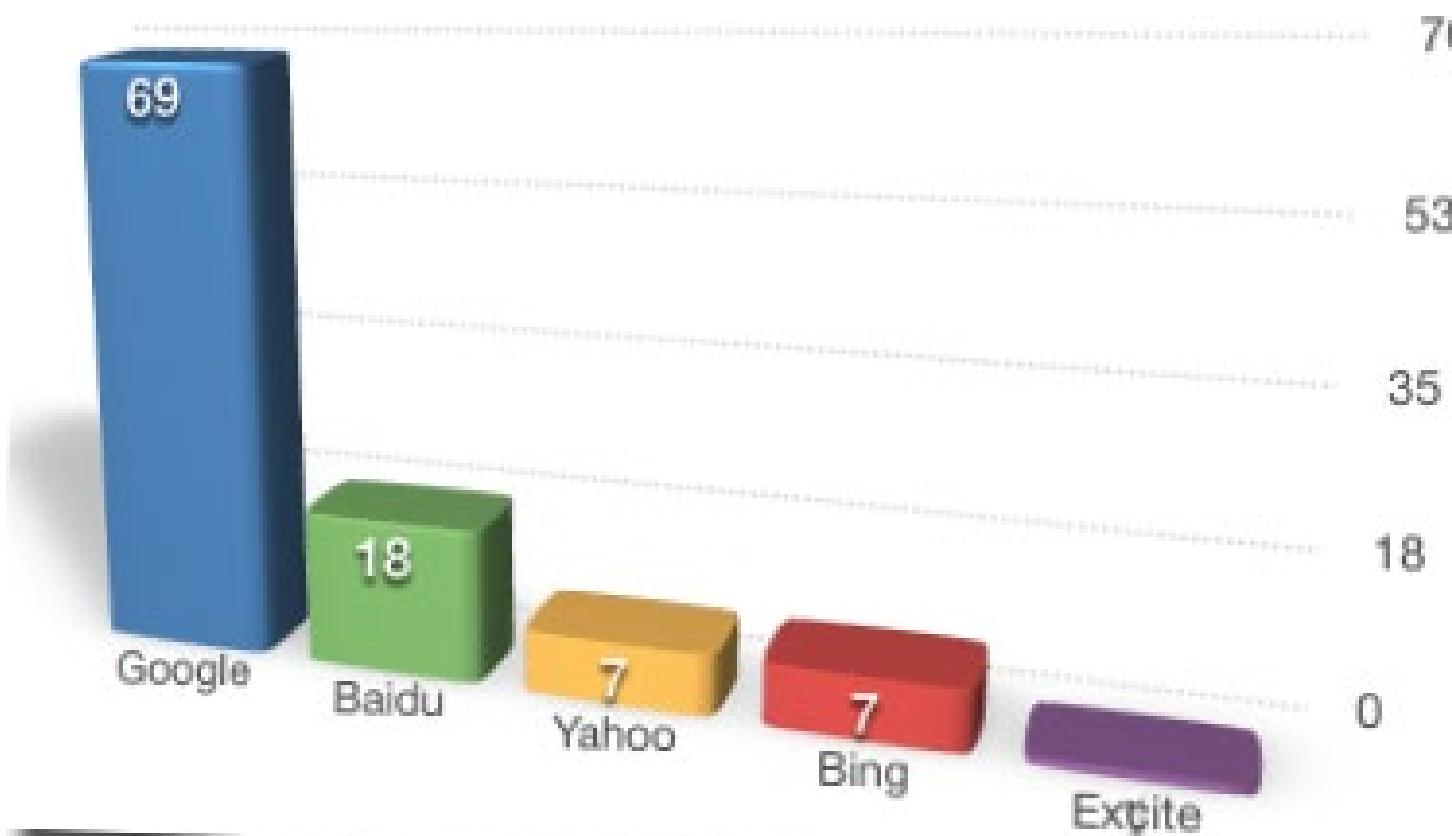


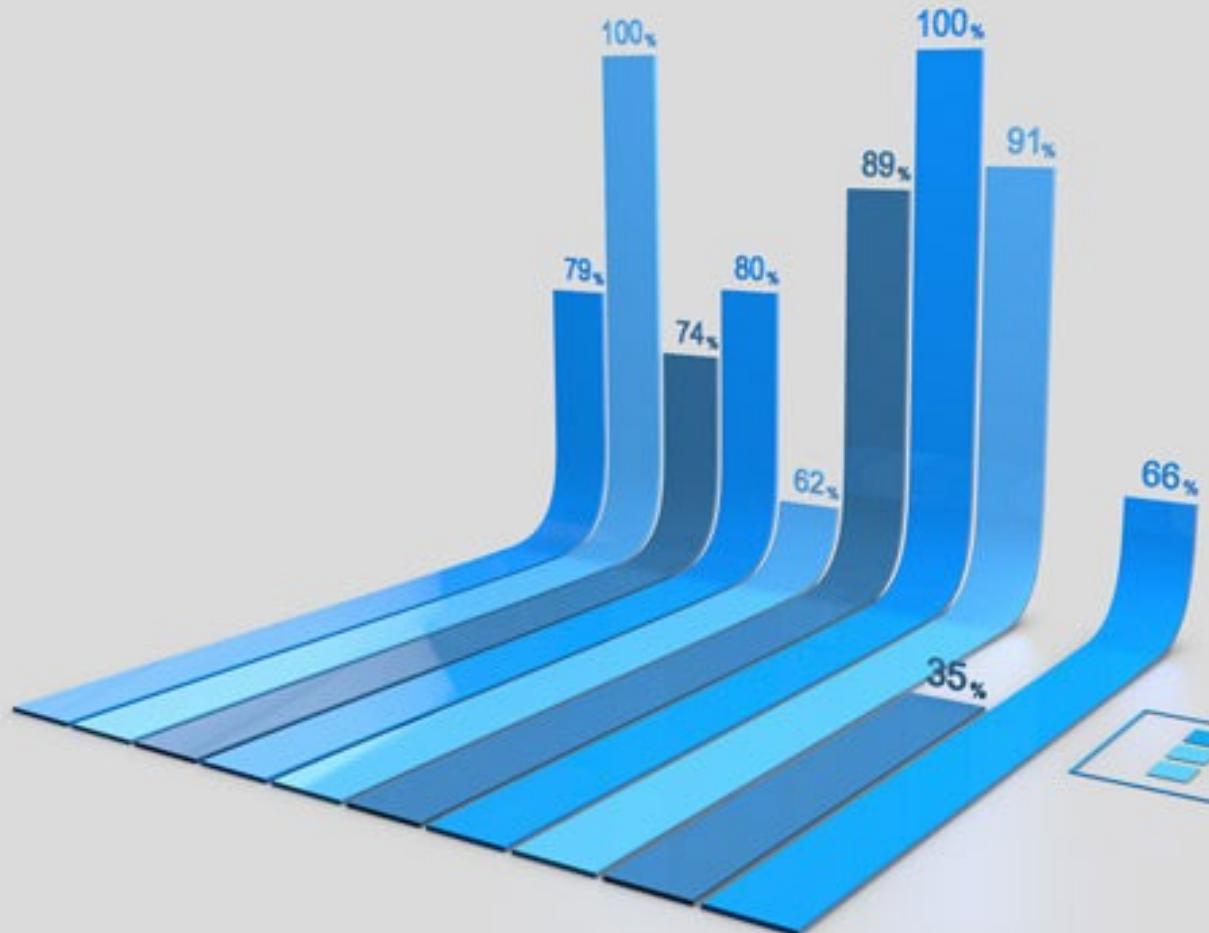
Figure 2. Age-Adjusted Rate of End-Stage Renal Disease Due to Any Cause per 100,000 Person-Years, According to Systolic and Diastolic Blood Pressure in 332,544 Men Screened for MRFIT.



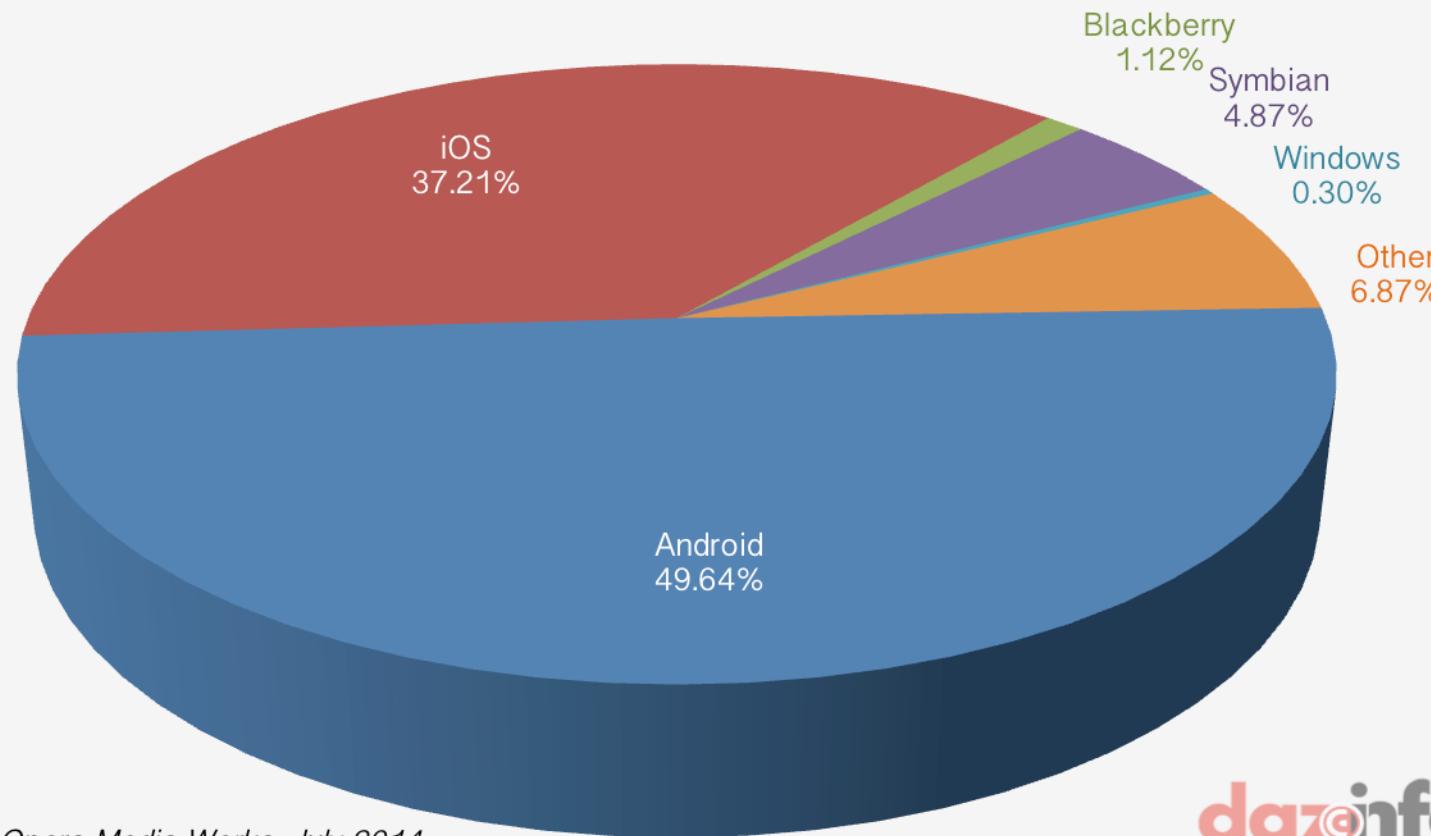


Search Engine Market Share





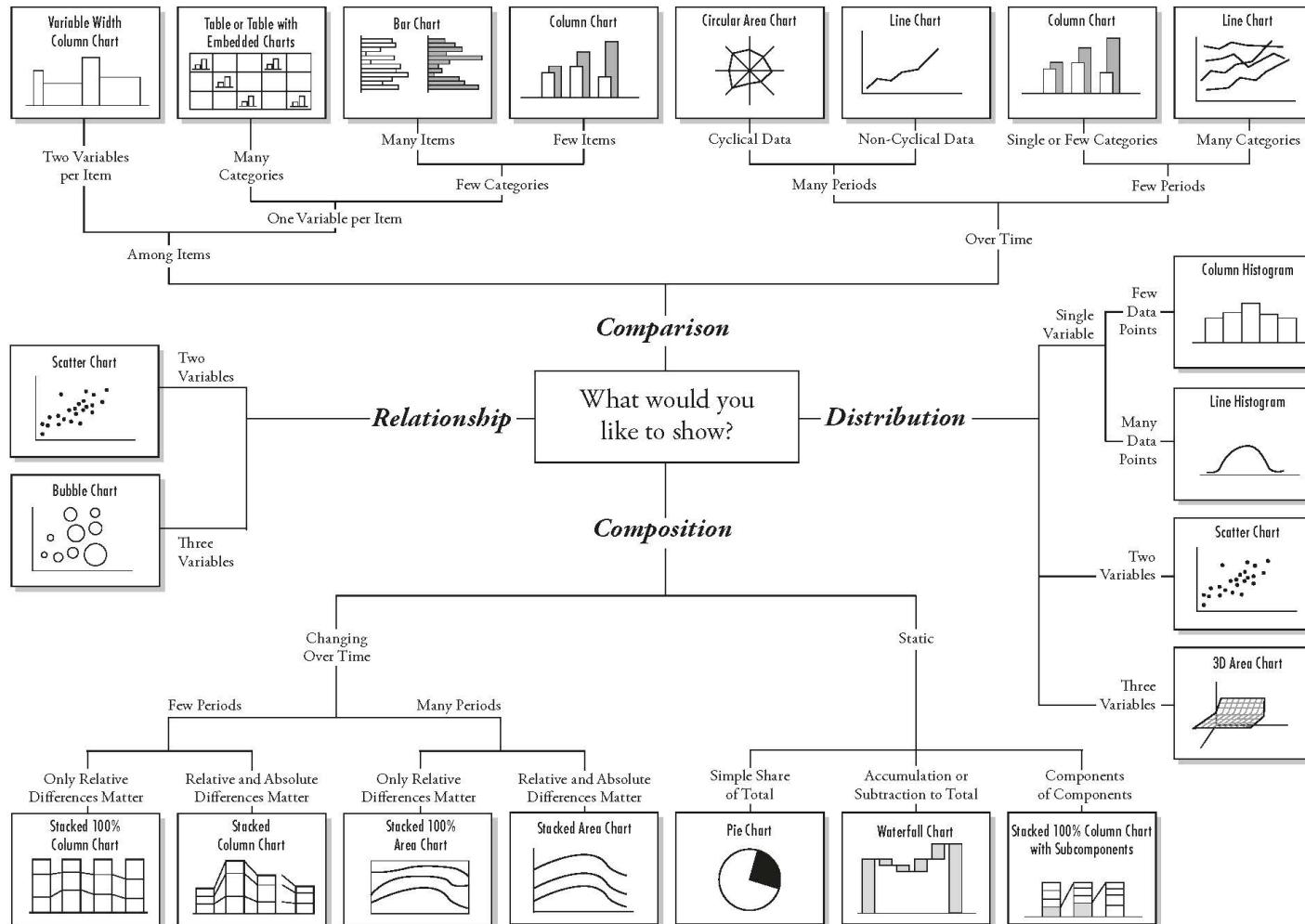
Mobile Phone OS Traffic Share Q2 2014



Source: Opera Media Works, July 2014

dazinfo
STAY AHEAD OF TECH CURVE

Chart Suggestions—A Thought-Starter



Is there a way
to choose the
“right” plot?

If there's time...

Check some examples from last lab