

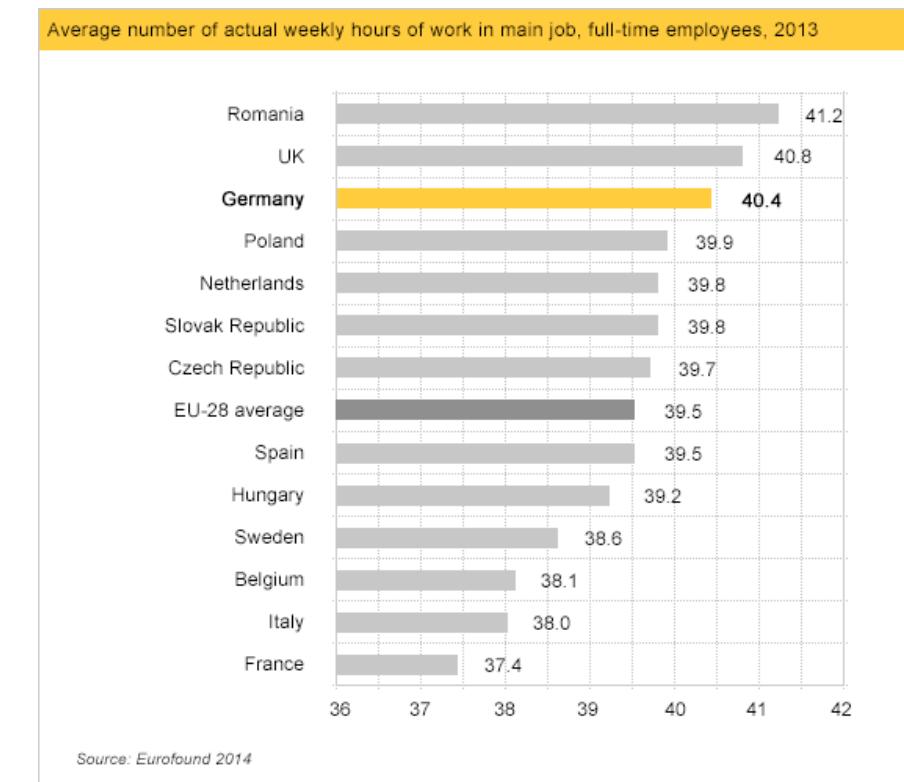
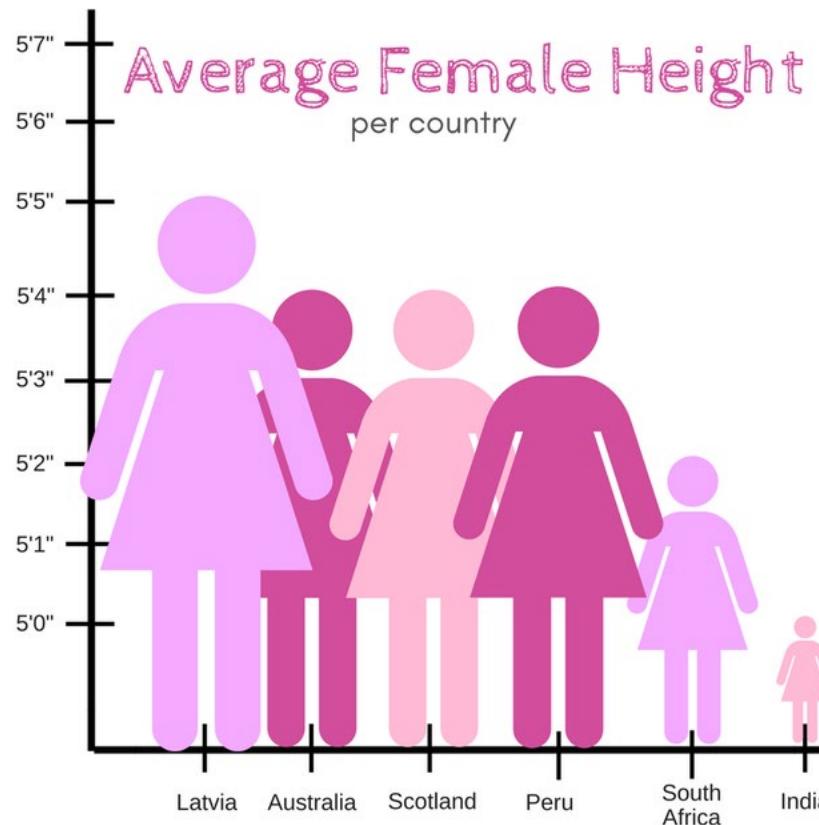
Calling Bullsh*t on Data Visualisations!

CM4125 – Topic 3

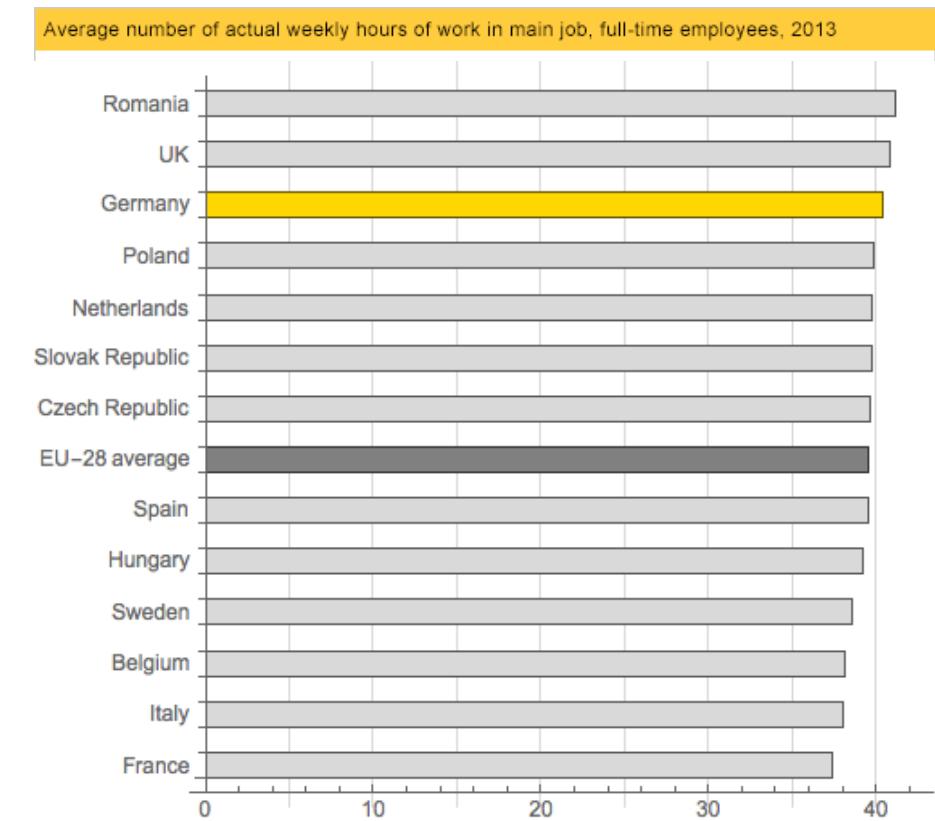
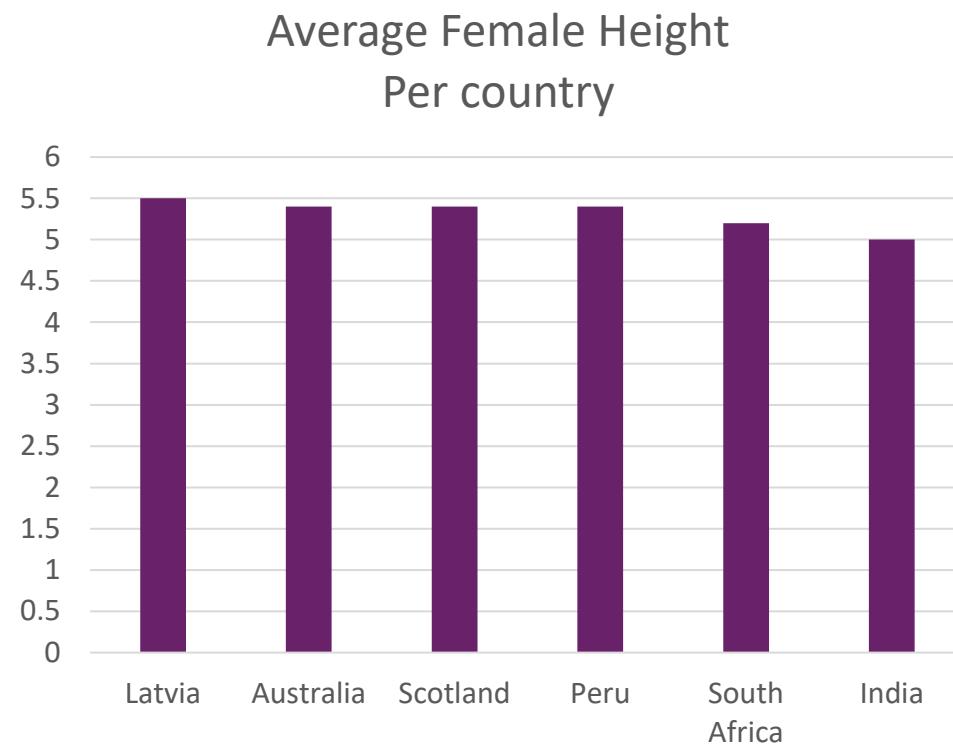
What this lecture is about?

- 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>
- Published as a book!
 - <https://www.amazon.co.uk/Calling-Bullshit-Skepticism-Data-Driven-World/dp/0525509186>
- 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).

1. Bar chart y-axis (dependent variable) need to include zero



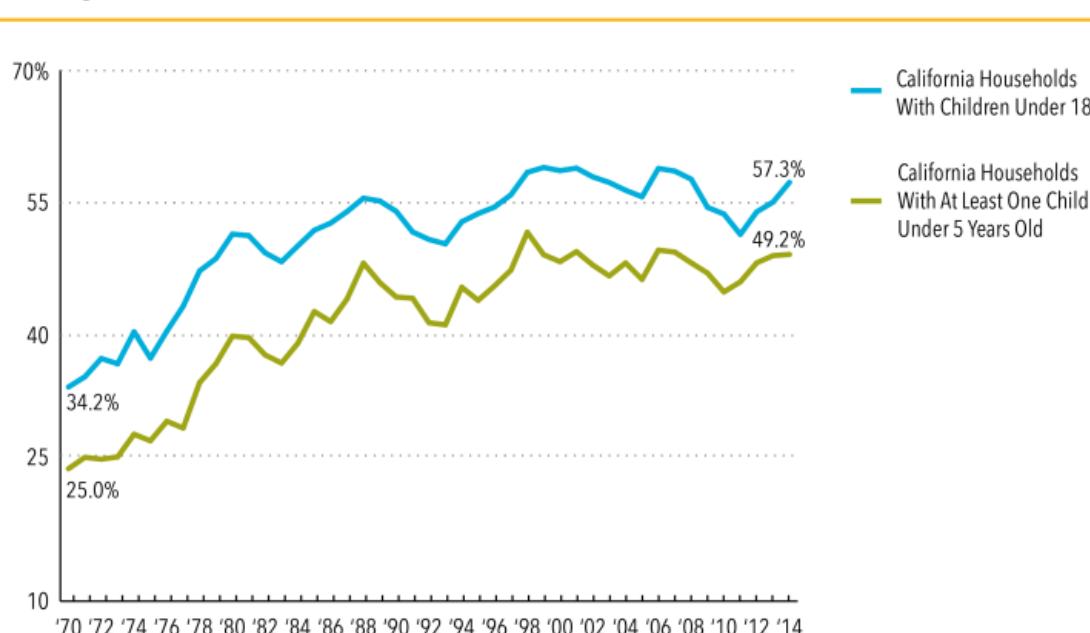
1. Bar chart y-axis (dependent variable) need to include zero



2. Line graph y-axis 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

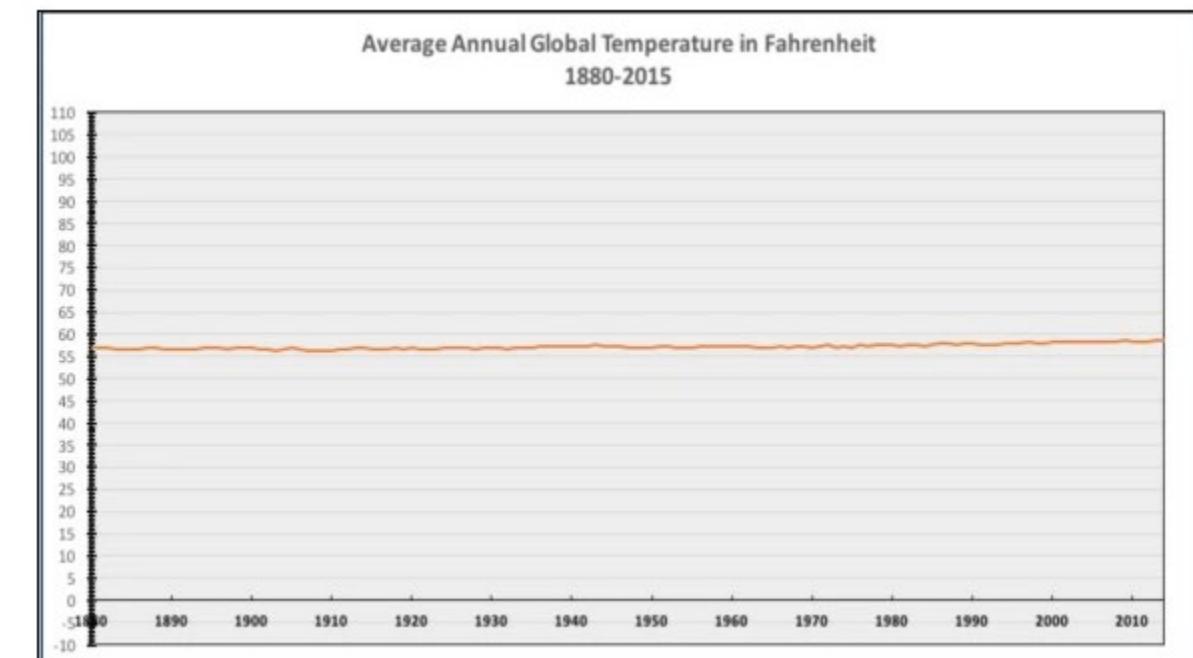


Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.

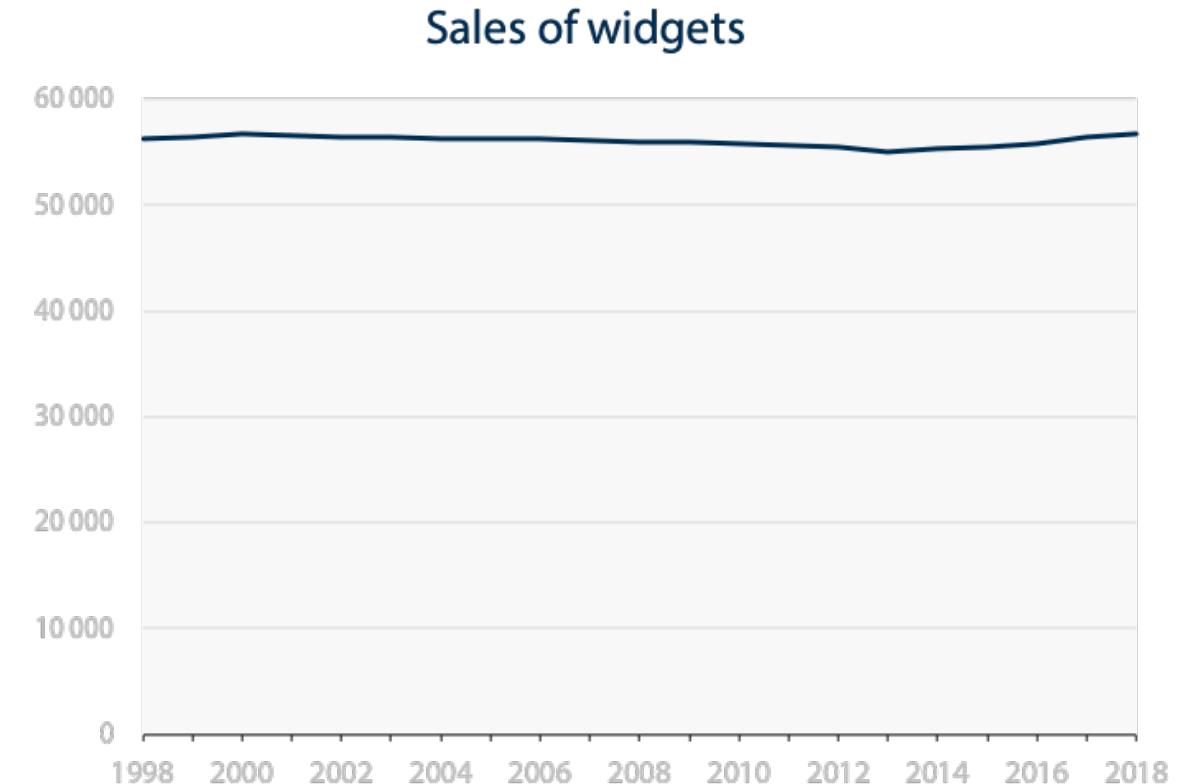
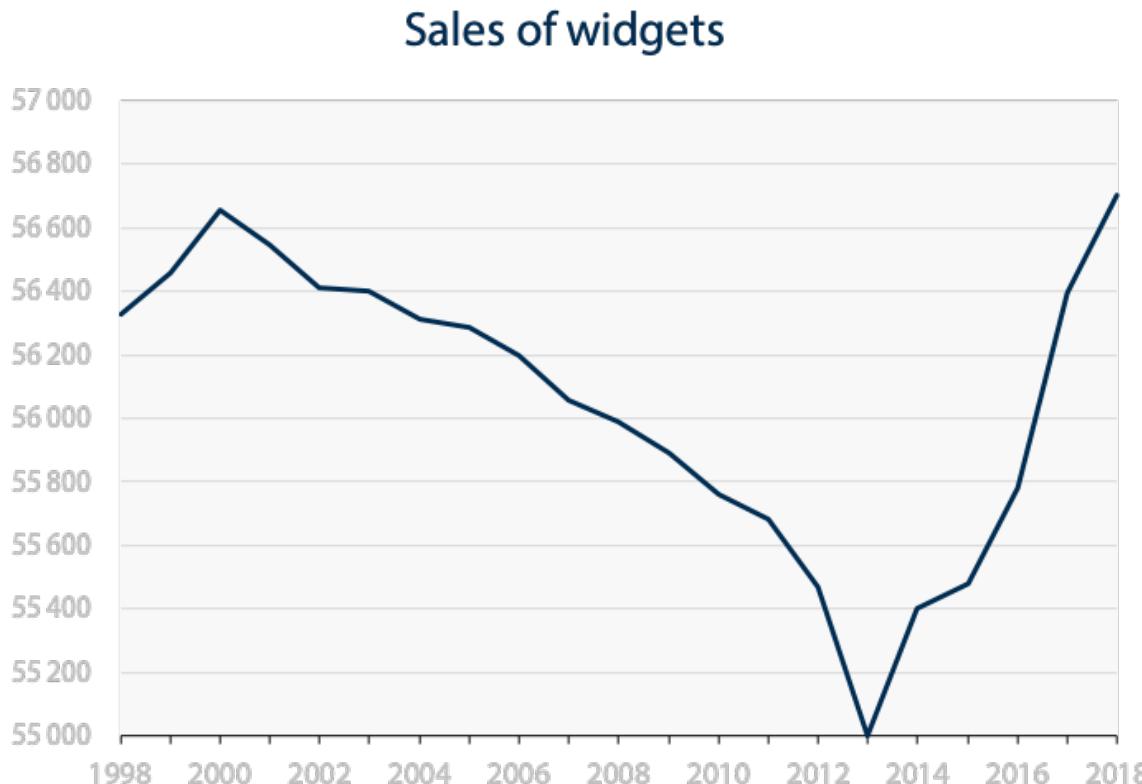
Source: Budget Center analysis of US Census Bureau data



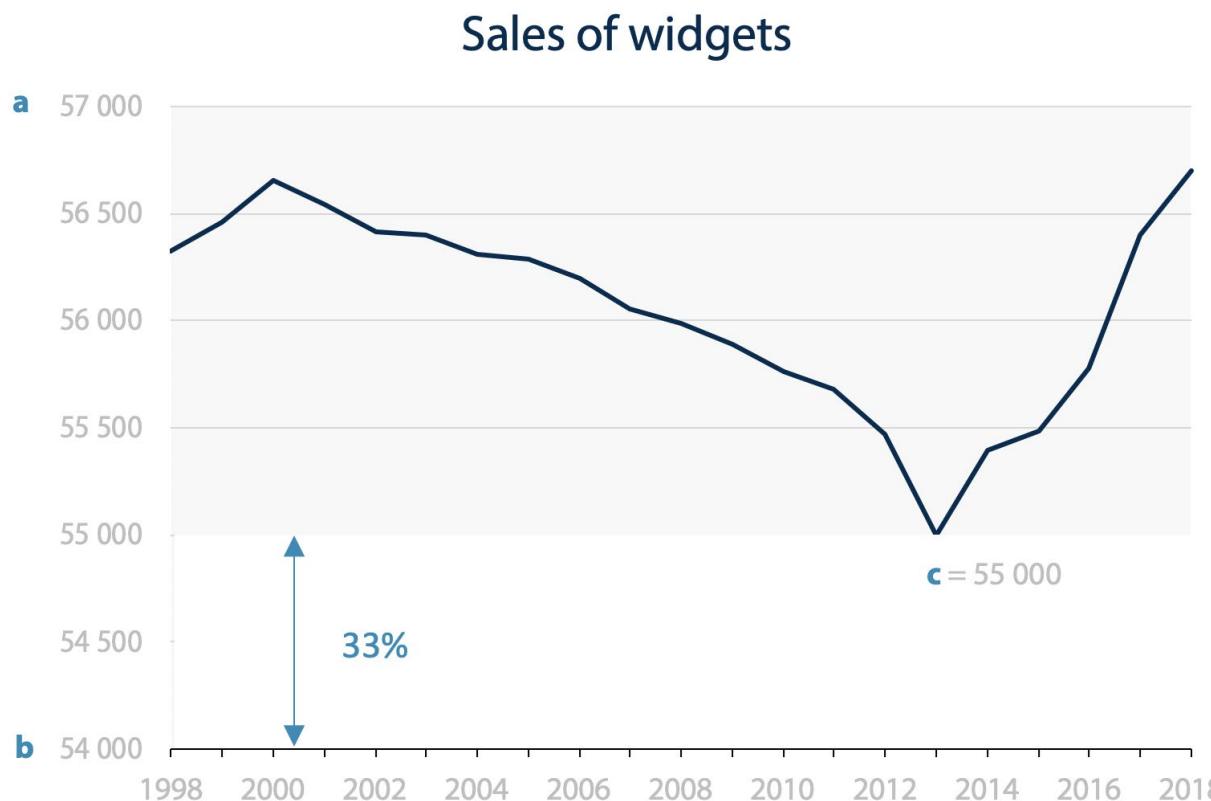
California Budget
& Policy Center
Independent Analysis. Shared Prosperity.



Well...



Solution



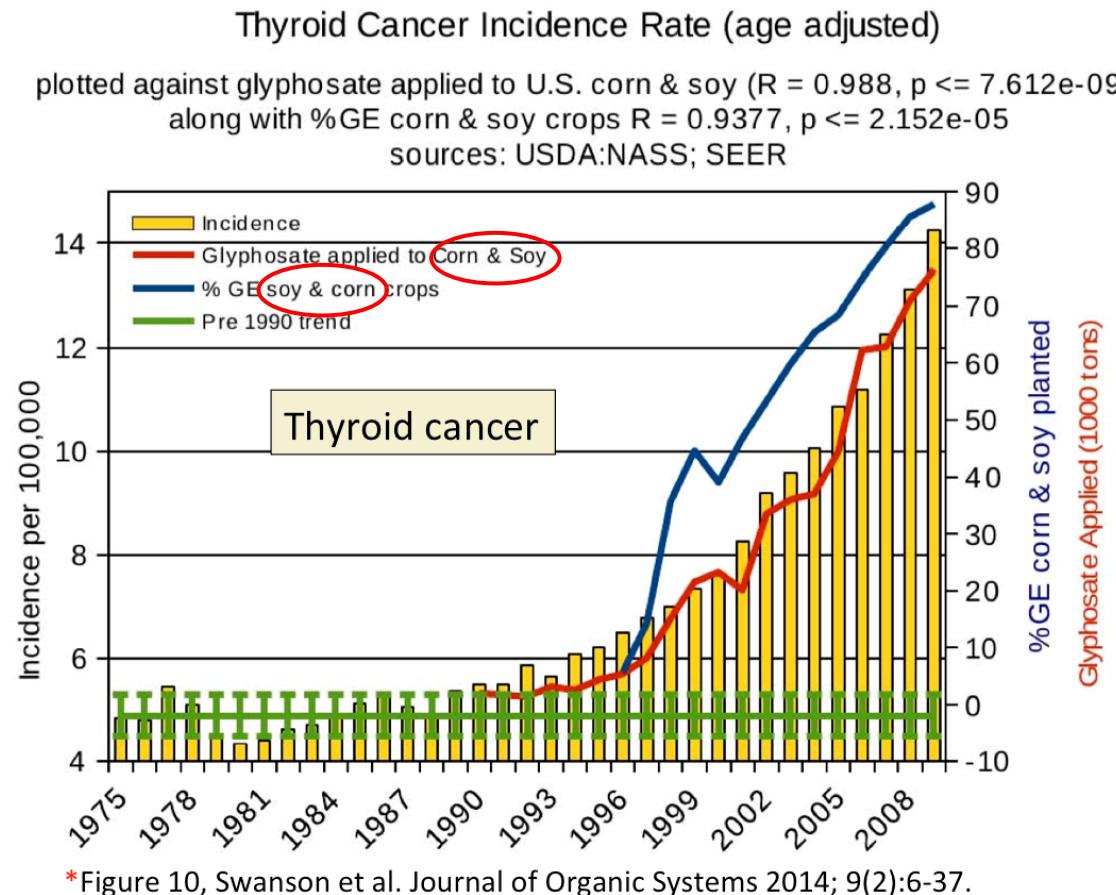
Empty space =
$$\frac{\text{Minimum data value} - \text{Minimum scale value}}{\text{Maximum scale value} - \text{Minimum scale value}}$$

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

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

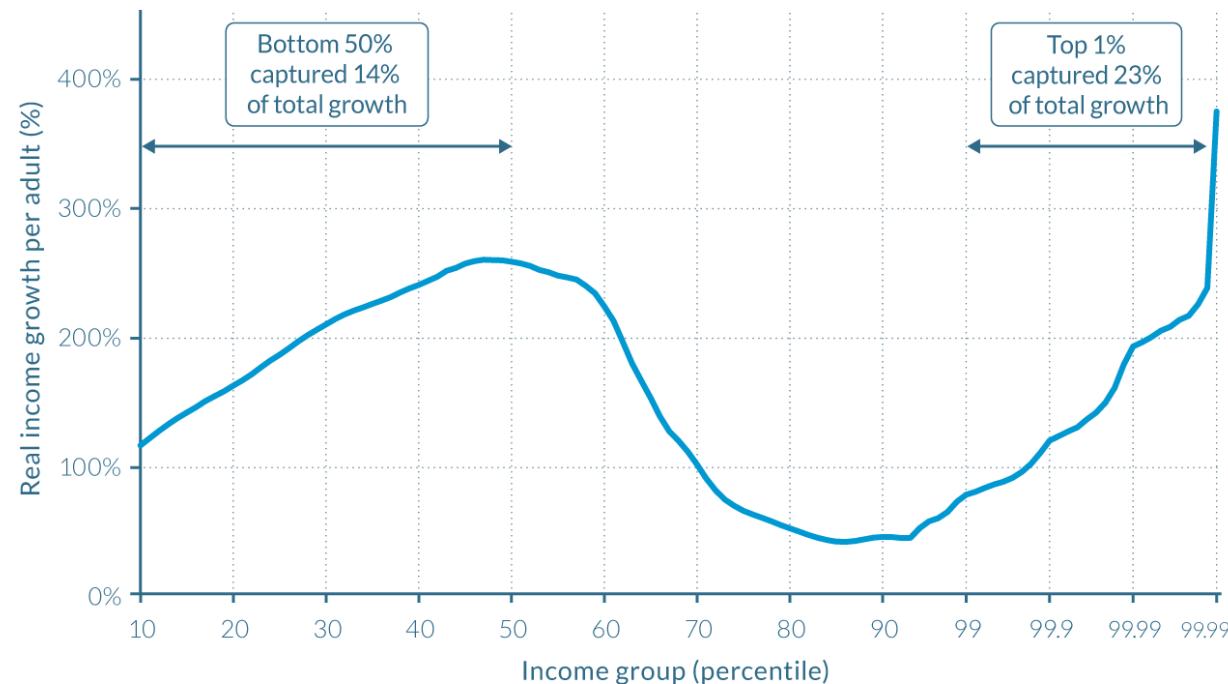
$$b = \frac{3 * 55,000 - 57000}{2} = 54\,000$$

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.

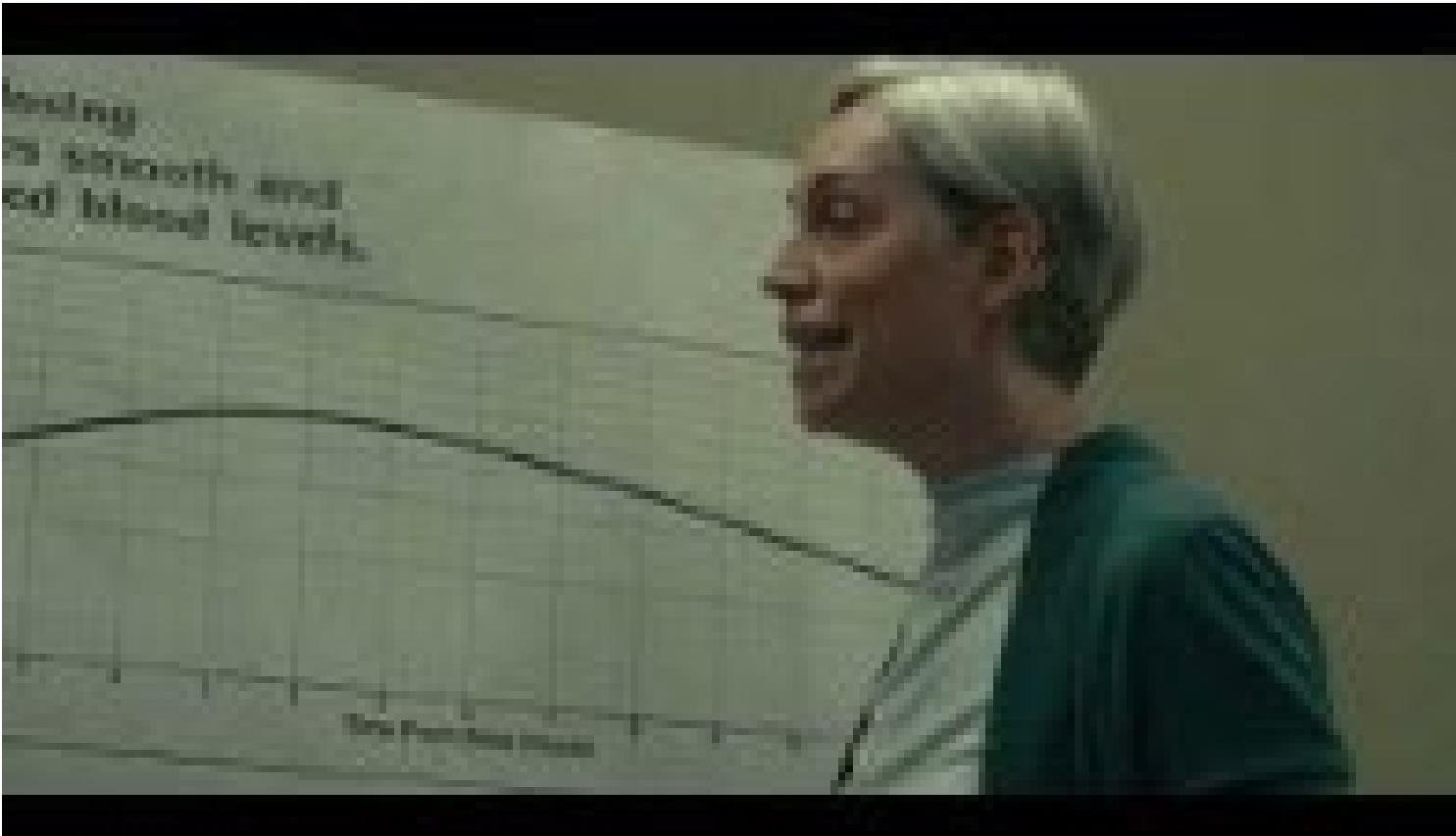
Use log scales!

Is this plot using
the “golden
ratio”?

Example (of what NOT to do)

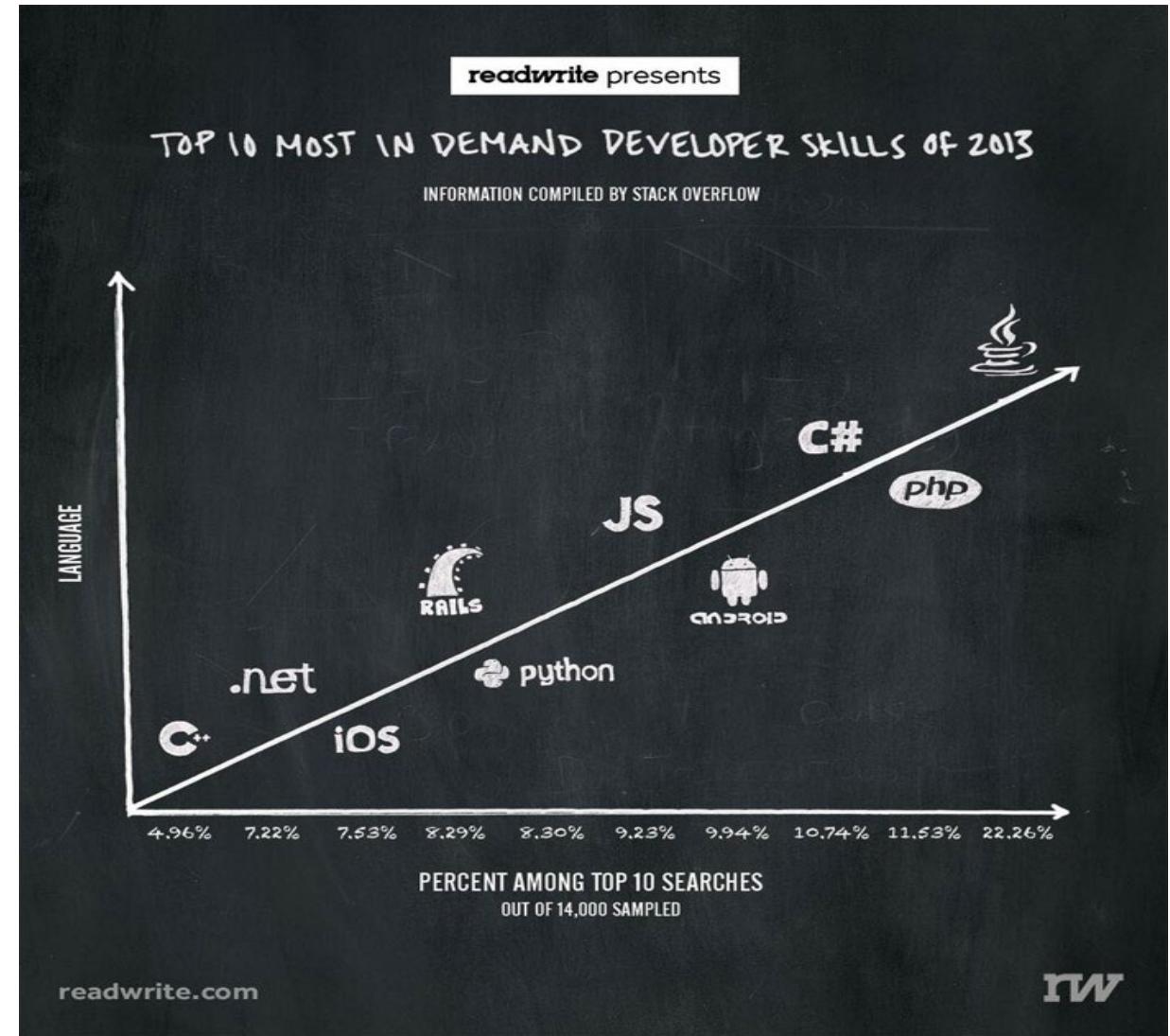


Example (of what NOT to do)



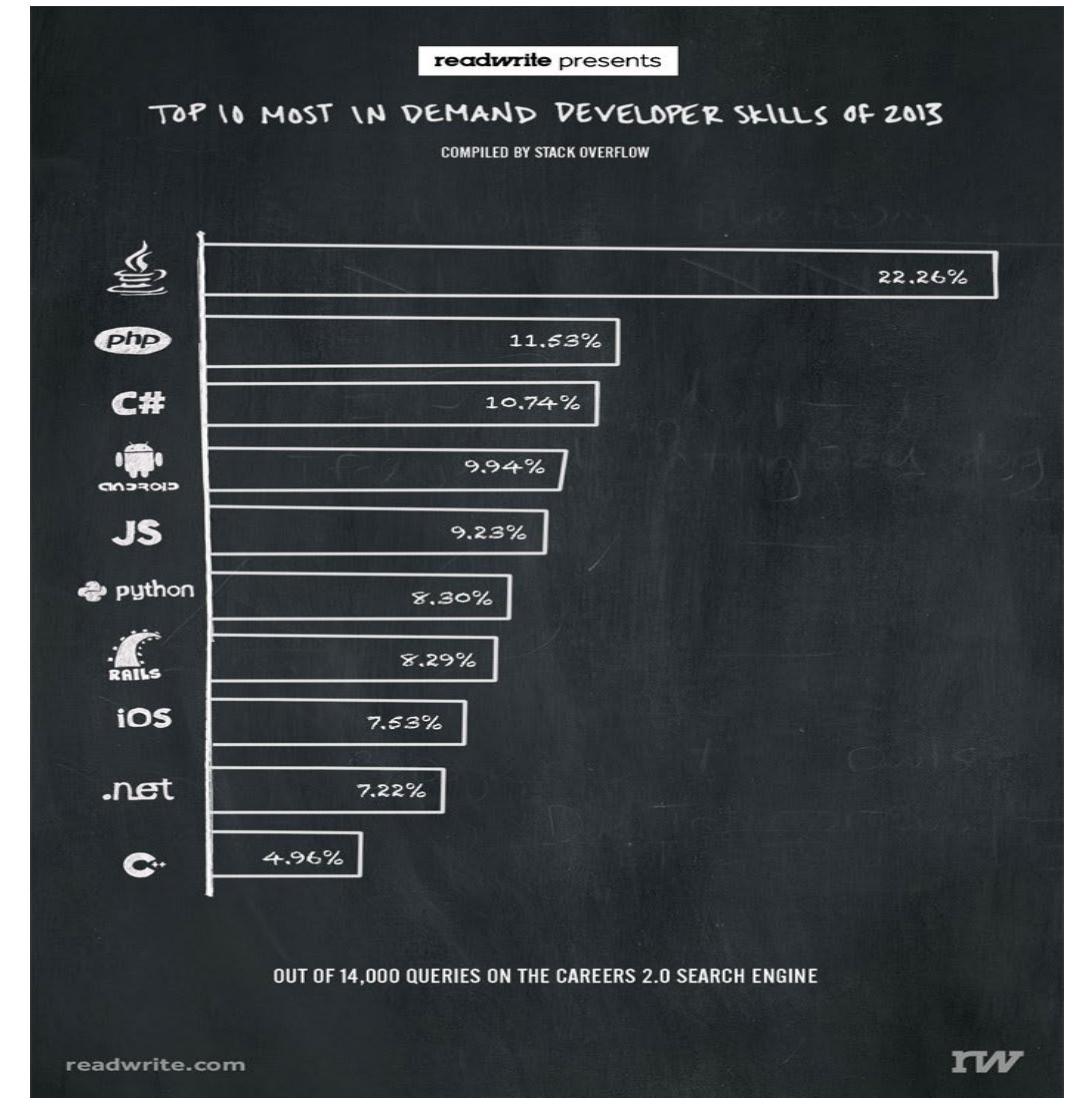


5. An axis should have *something* on it



Like this

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

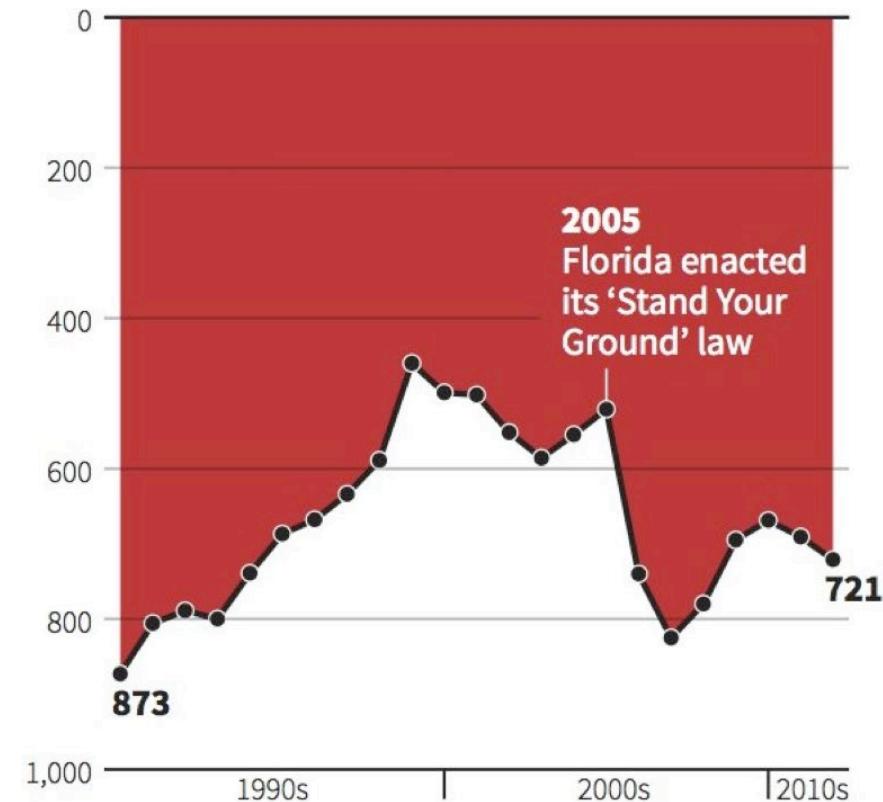


6. Don't invert the axis(?)

The author defended it!
What's your stance?

Gun deaths in Florida

Number of murders committed using firearms



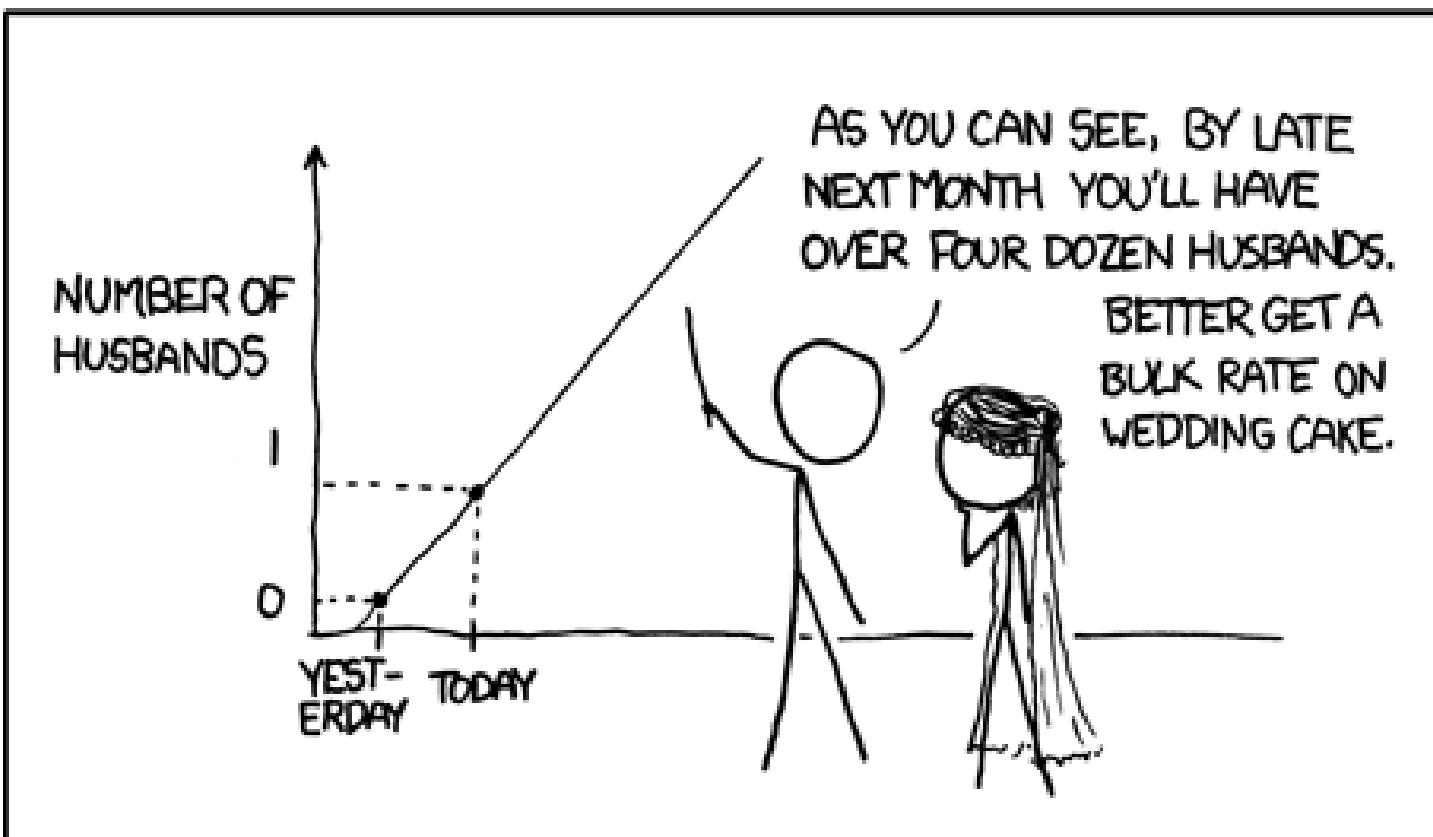
Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

REUTERS

7. Avoid extrapolation!

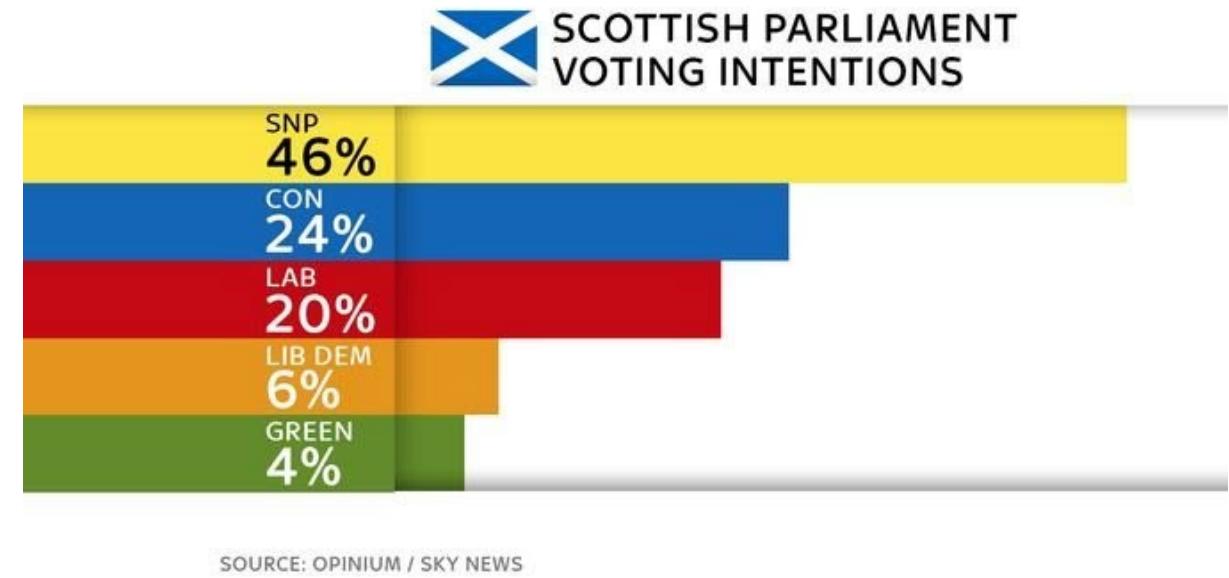
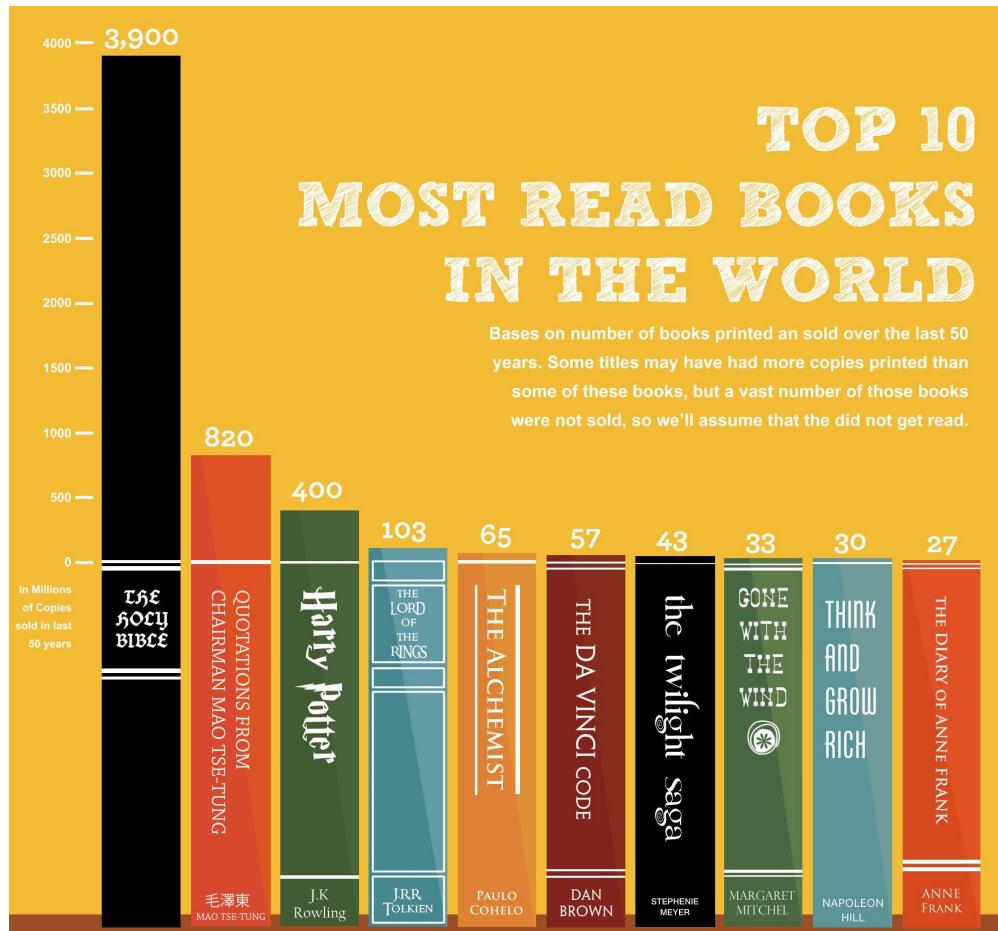
MY HOBBY: EXTRAPOLATING

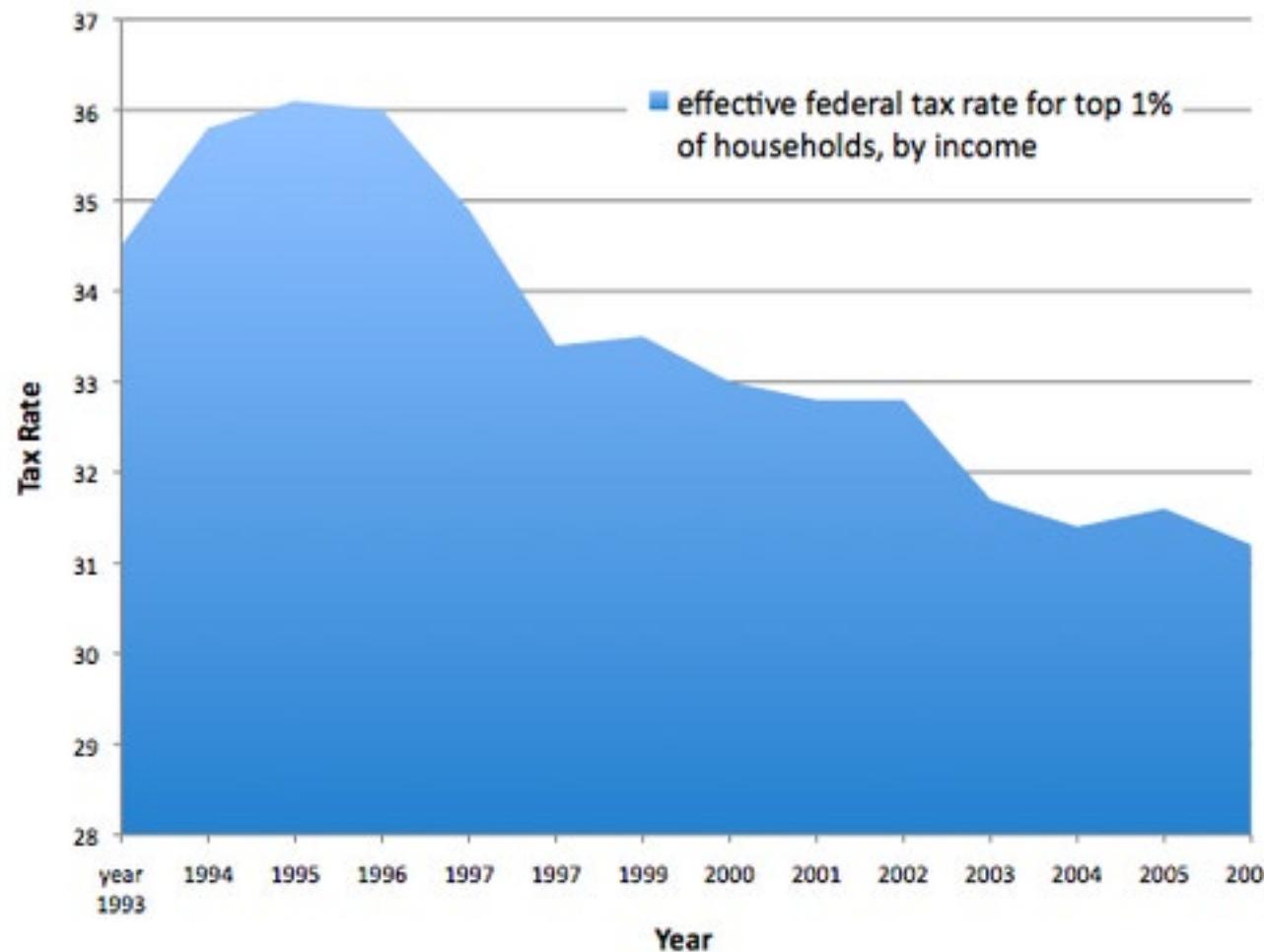


8. Proportional Ink

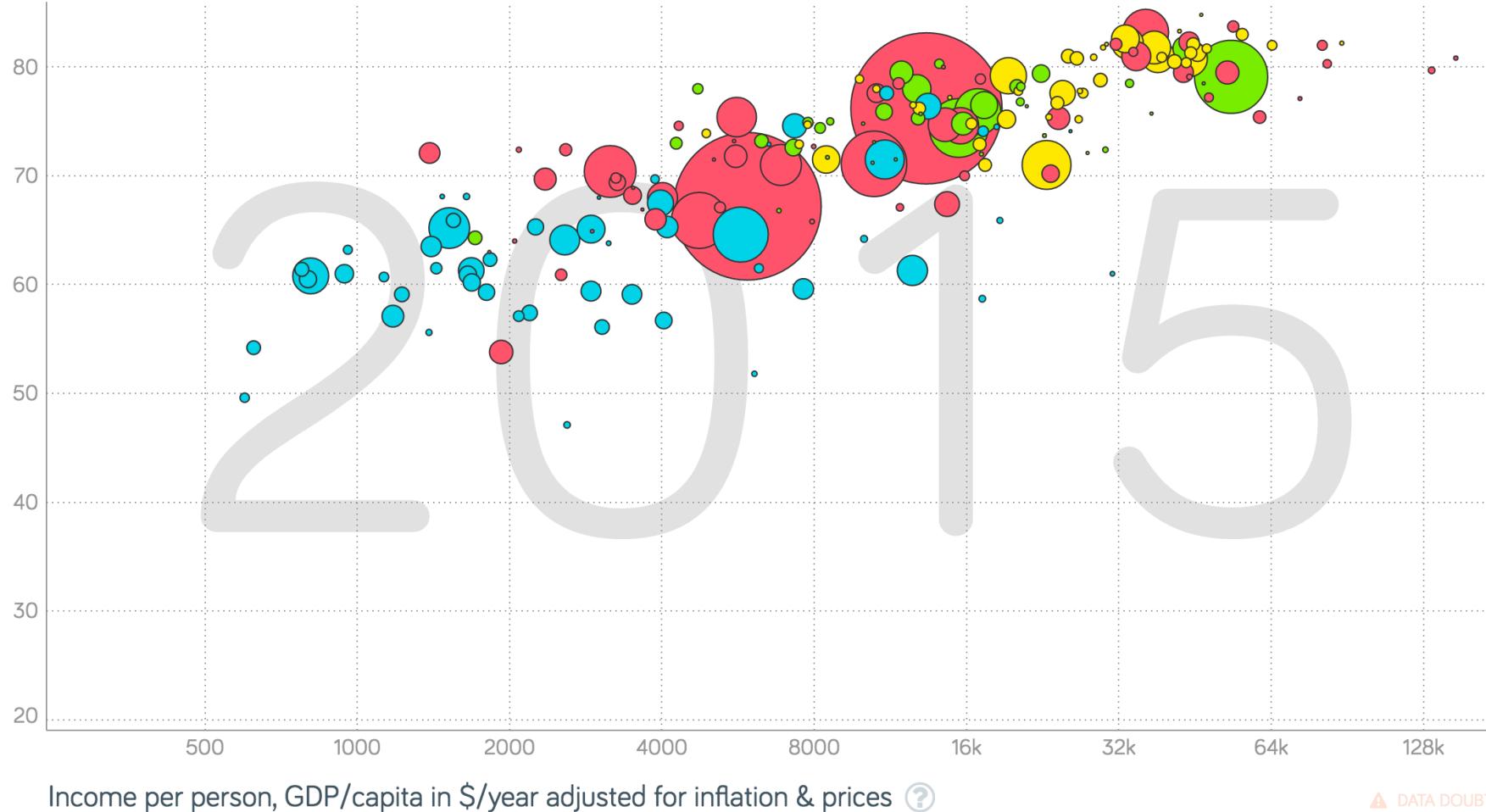
“When a shaded region is used to represent a numerical value, the area of that shaded region should be directly proportional to the corresponding value”

Extends the argument for misleading axes





Life expectancy, years 

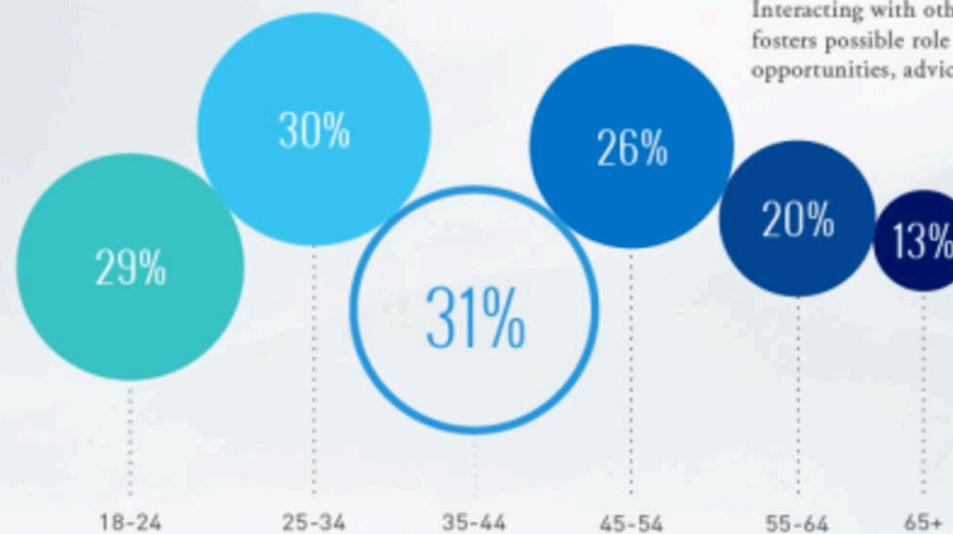


Should we violate
the principle?

Radius ≠ Area

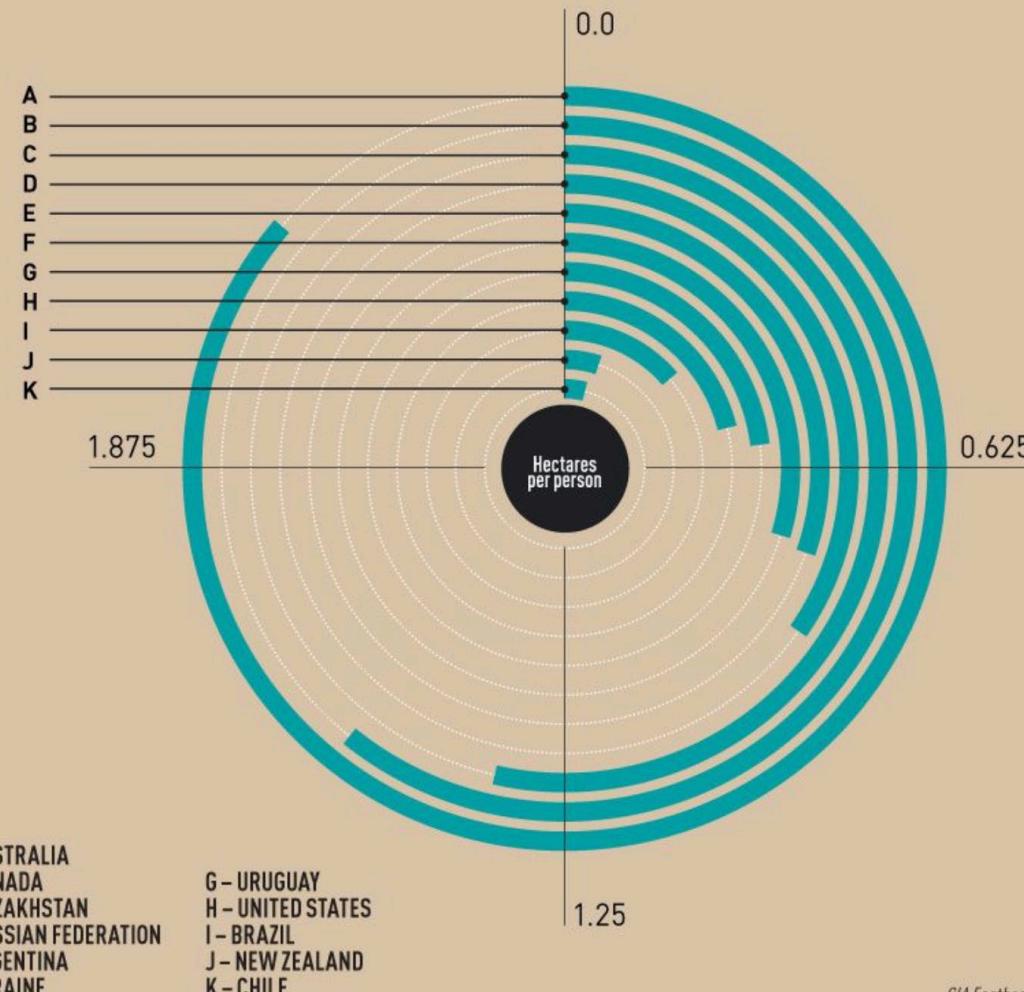
It's Who You Know

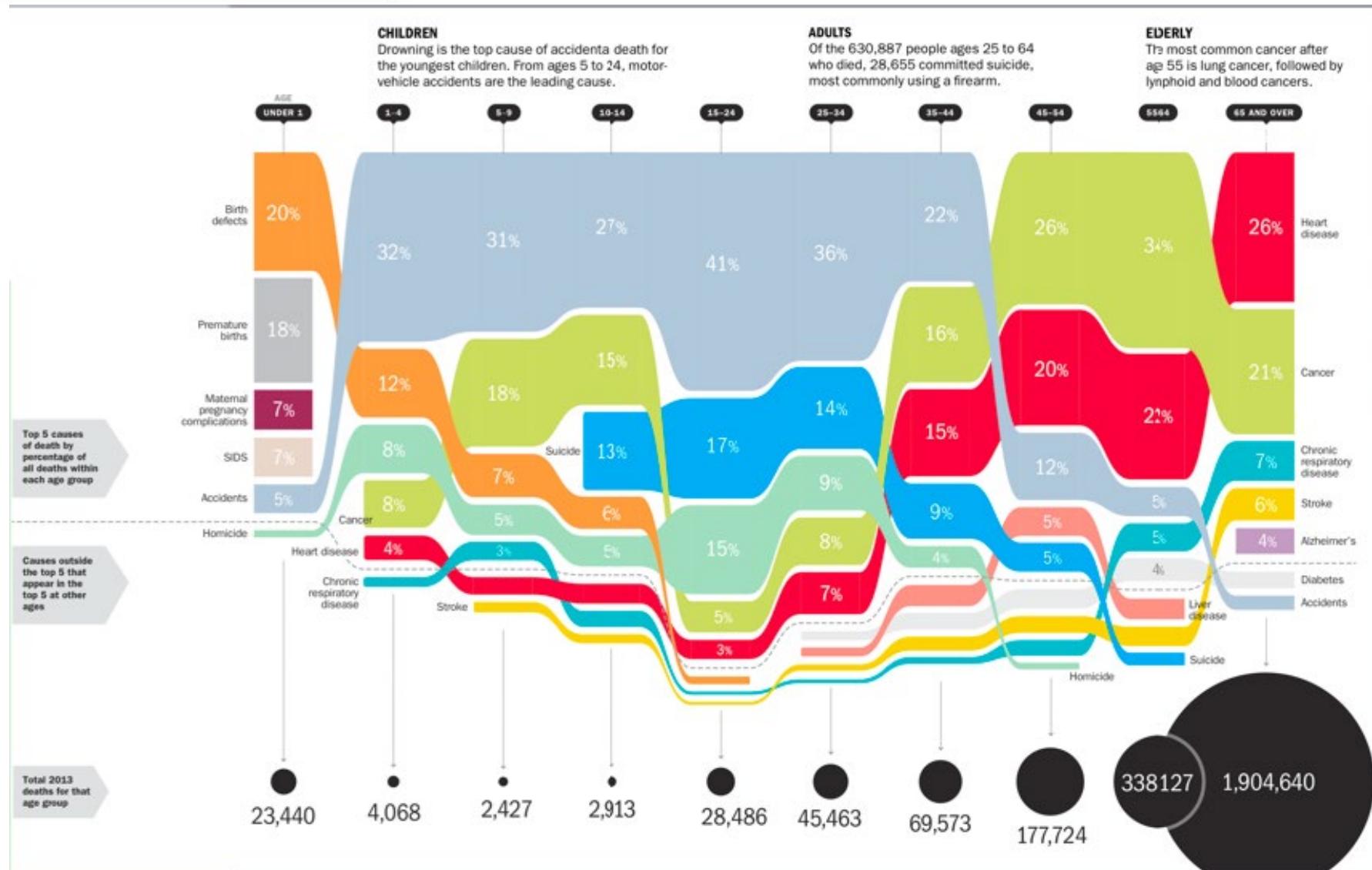
By ages 35–44, people are likely to have the strongest network of fellow entrepreneurs, but these **personal social networks** begin to decline over time.



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

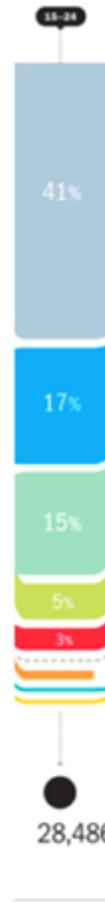
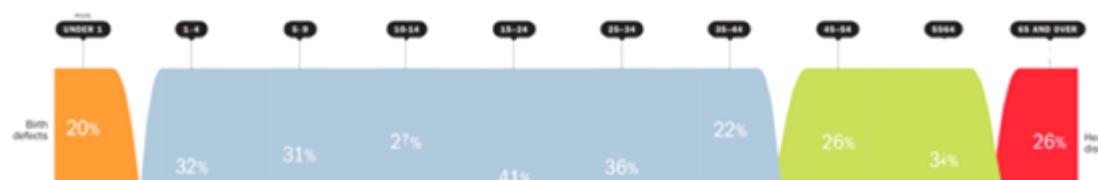
ARABLE LAND PER CAPITA (HECTARES PER PERSON)





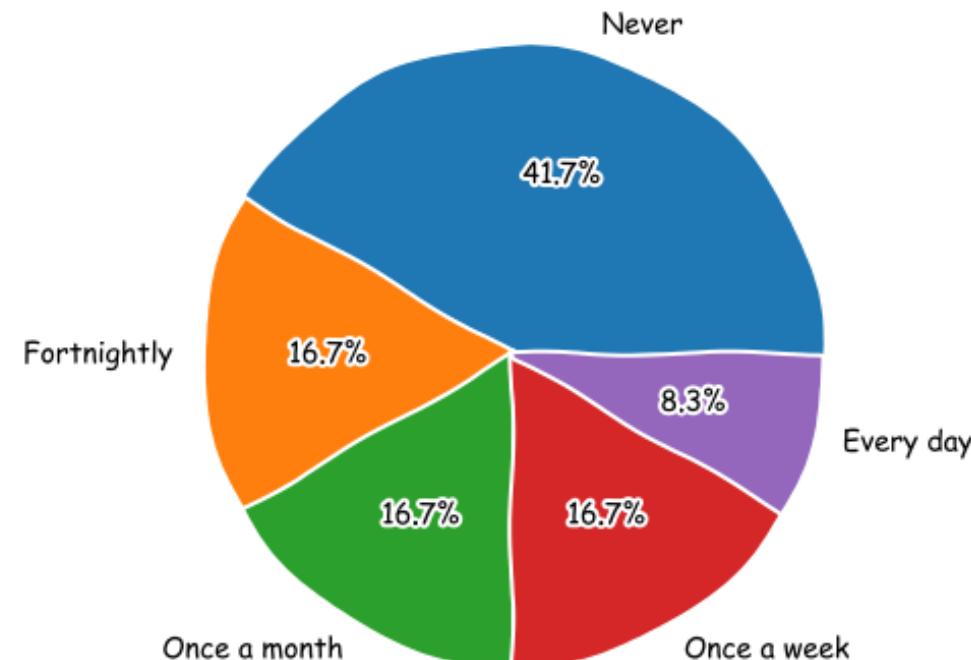
Along vertical slices, ink is proportional to value because shaded areas represent the fraction of a fixed number of deaths (here 28,486) from each cause.

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.

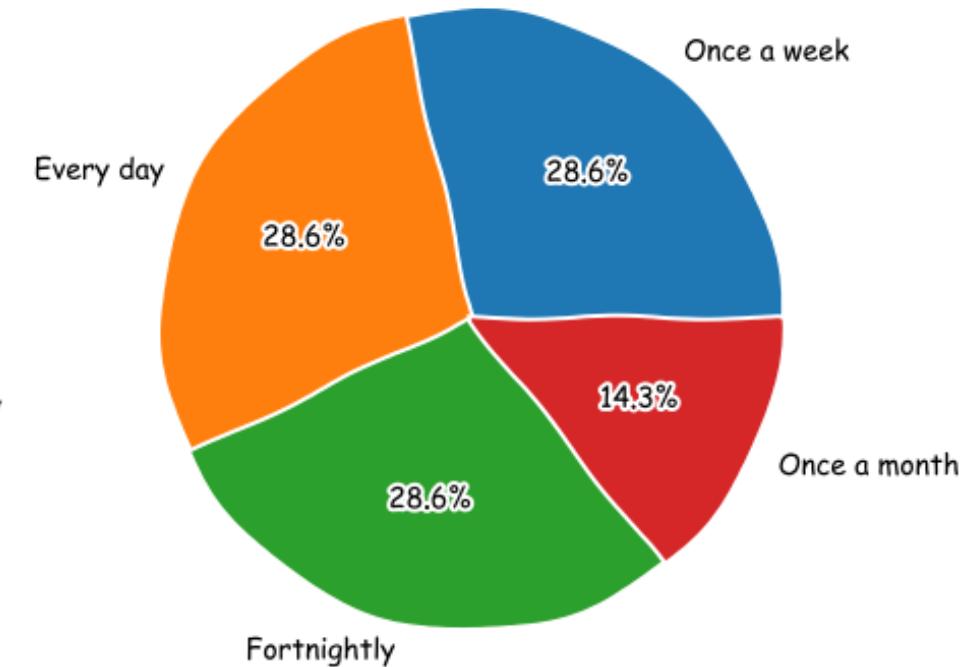


How often do you use a dating app during a month?

Male



Female



What is wrong in
this one?

9. Perspective

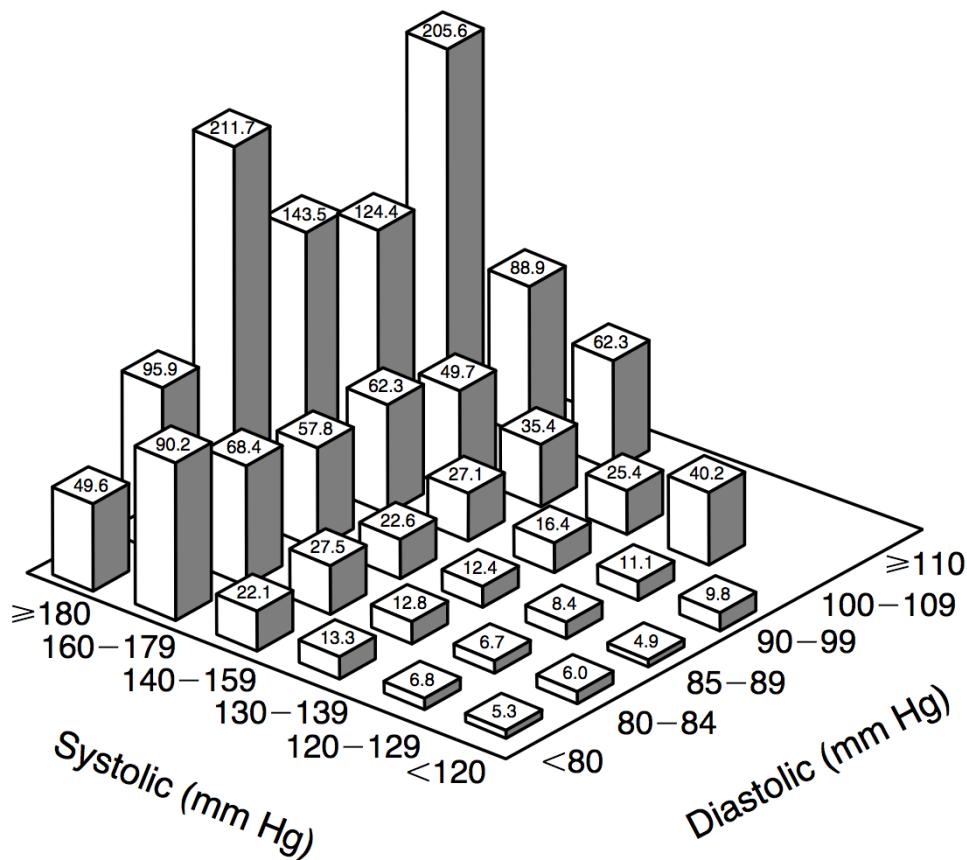
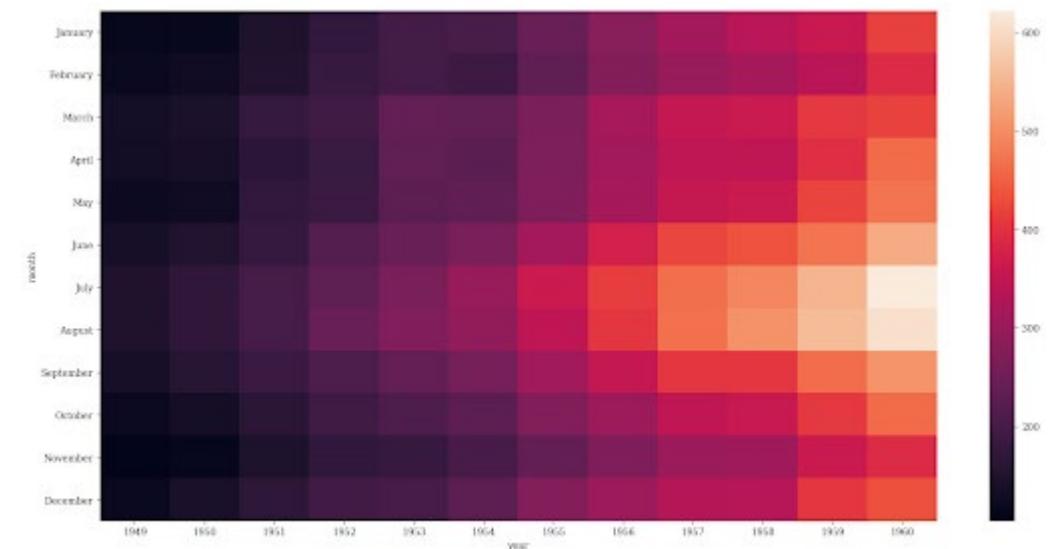
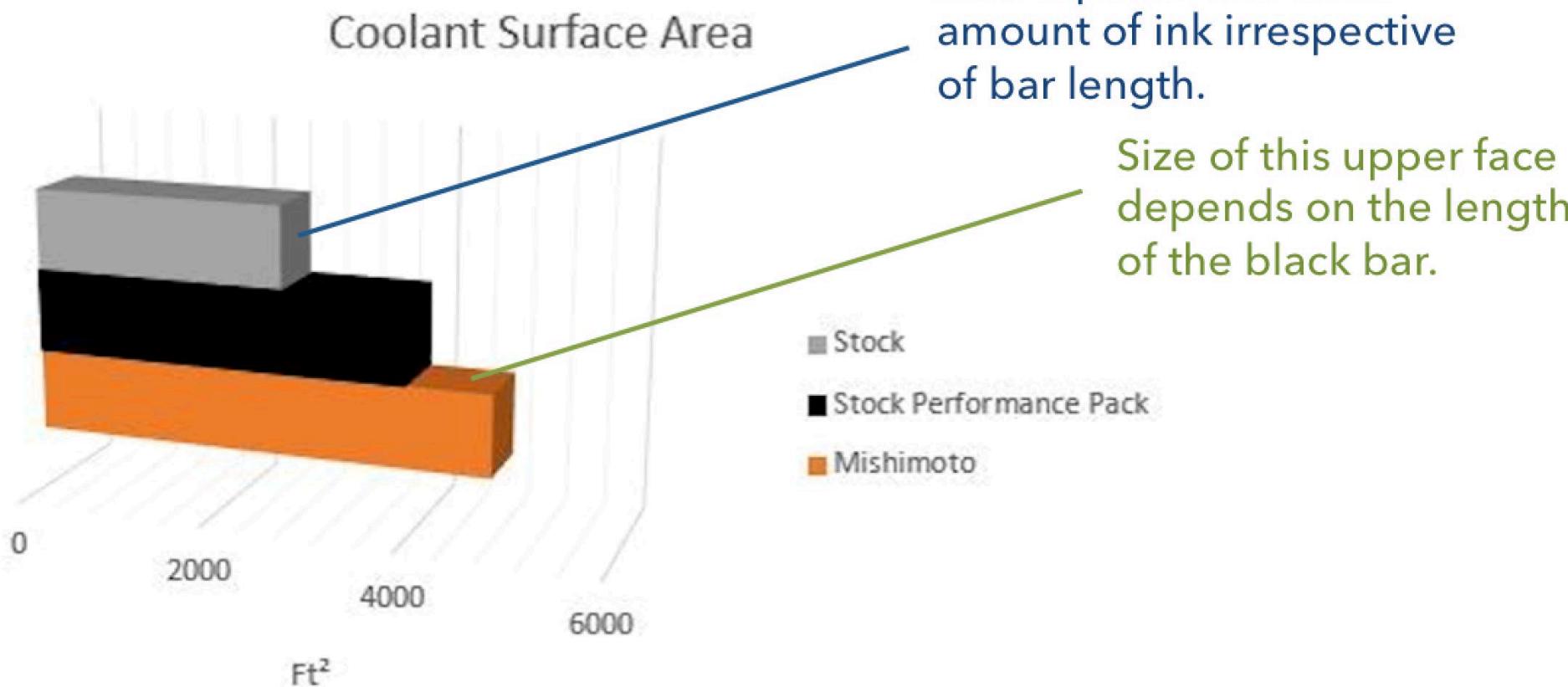
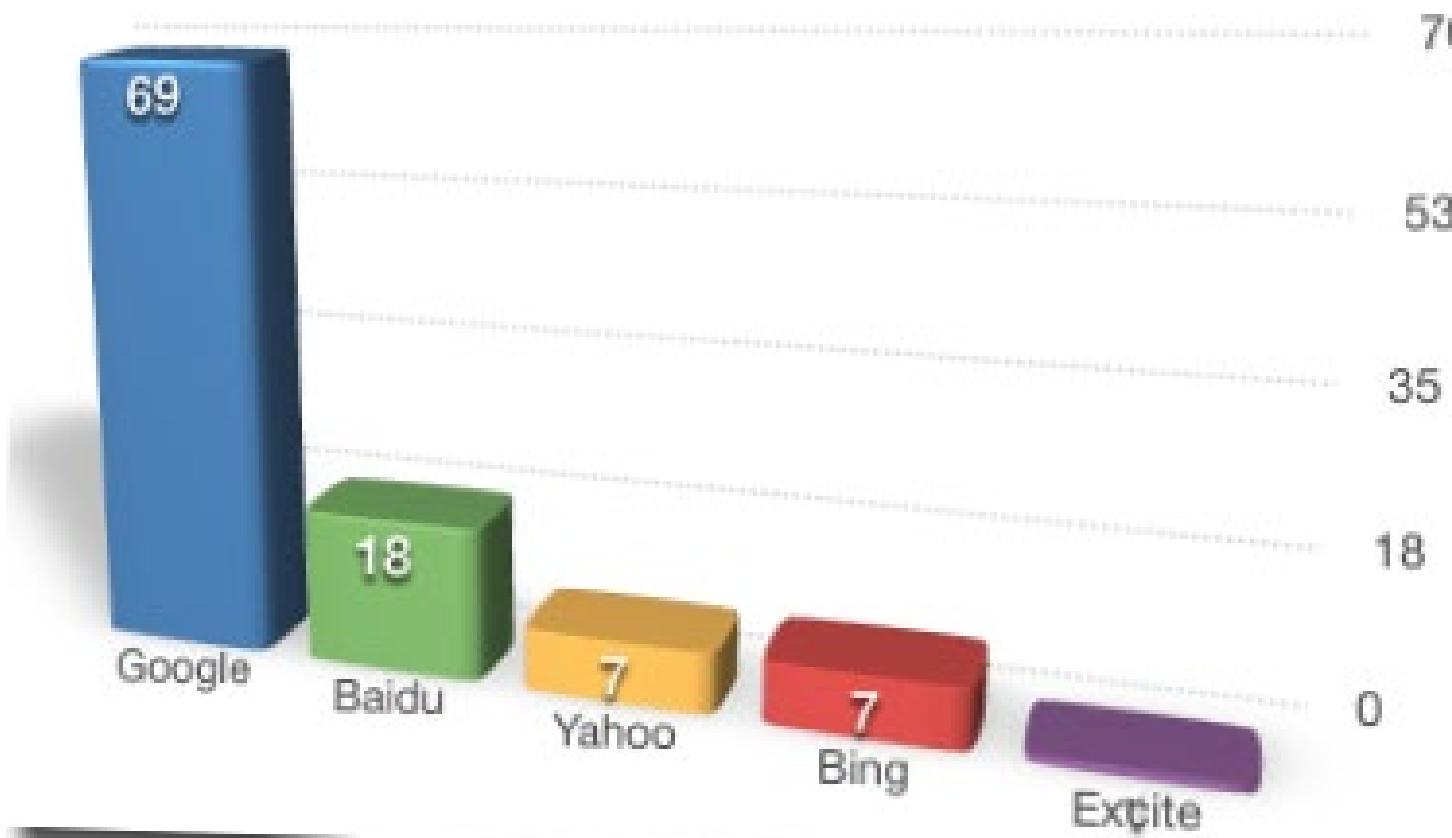


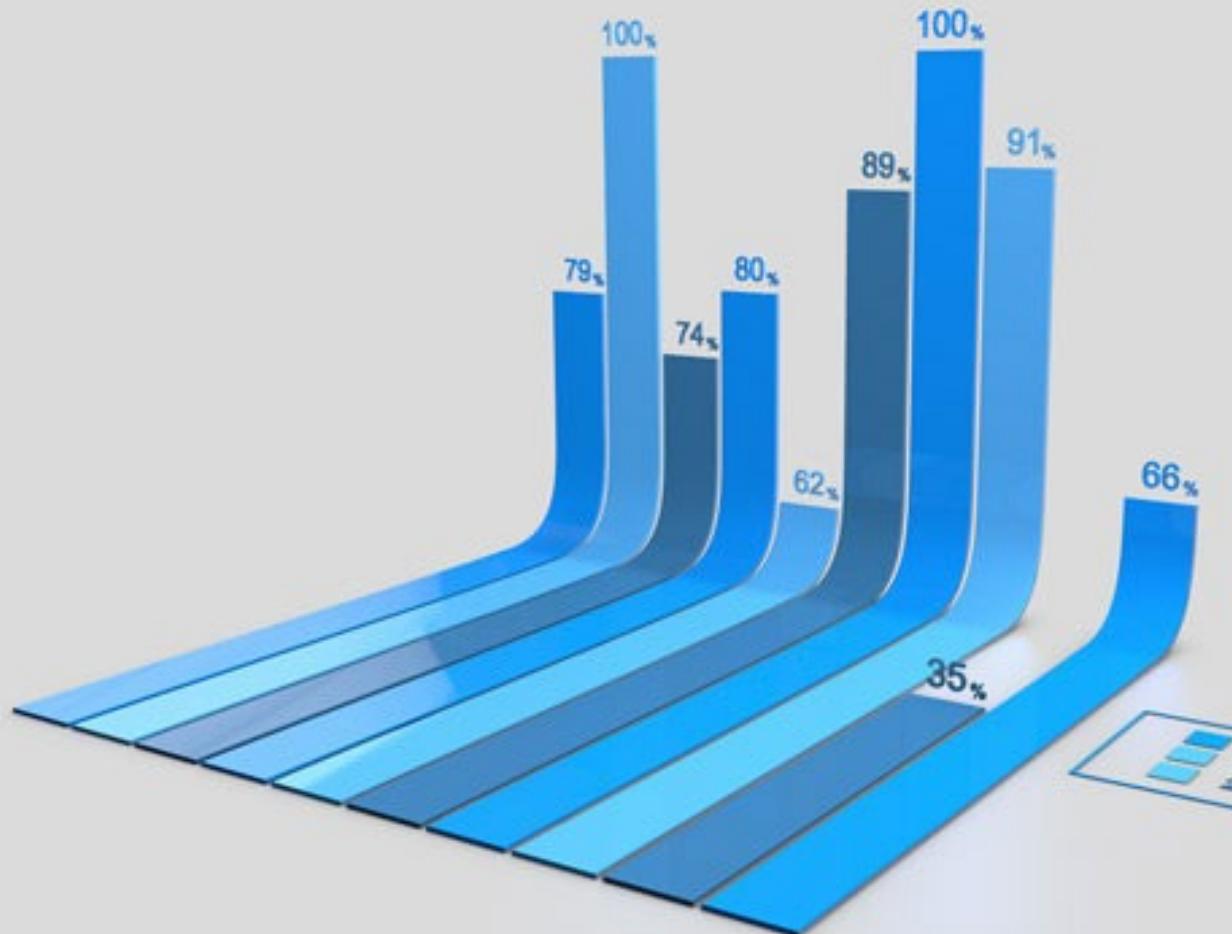
Figure 2. Age-Adjusted Rate of End-Stage Renal Disease Due to Any Cause per 100,000 Person-Years, According to Systolic and Diastolic Blood Pressure in 332,544 Men Screened for MRFIT.



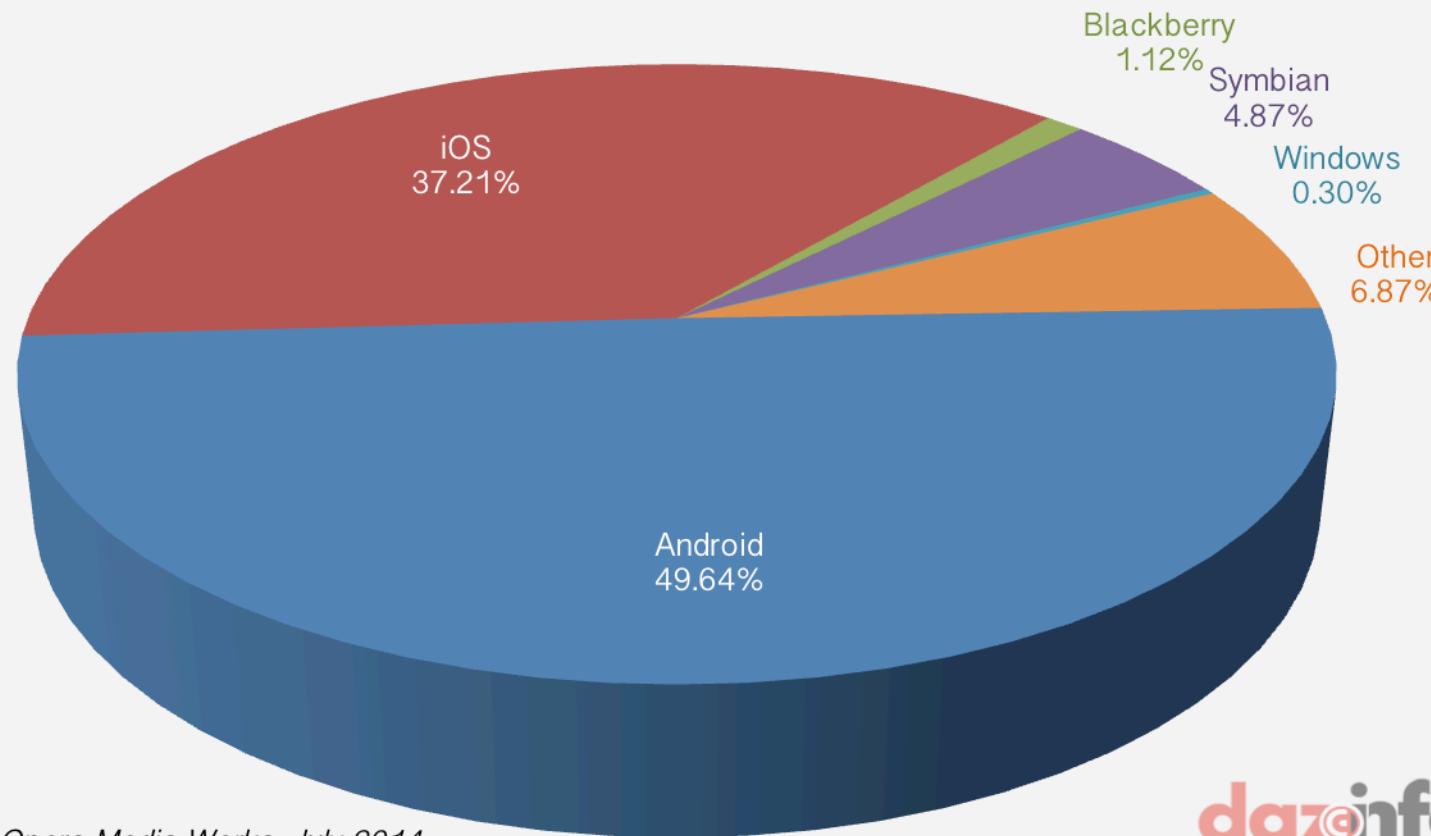


Search Engine Market Share



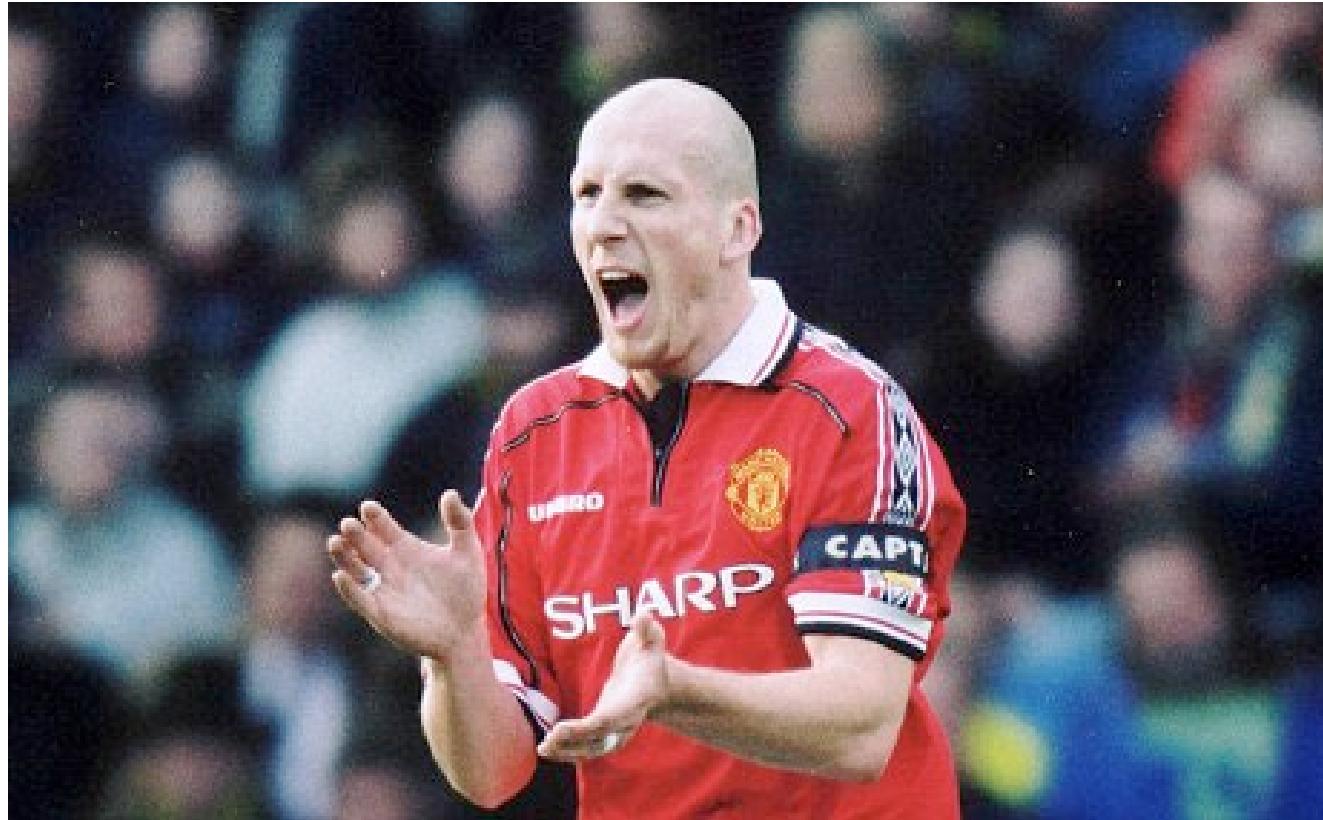


Mobile Phone OS Traffic Share Q2 2014

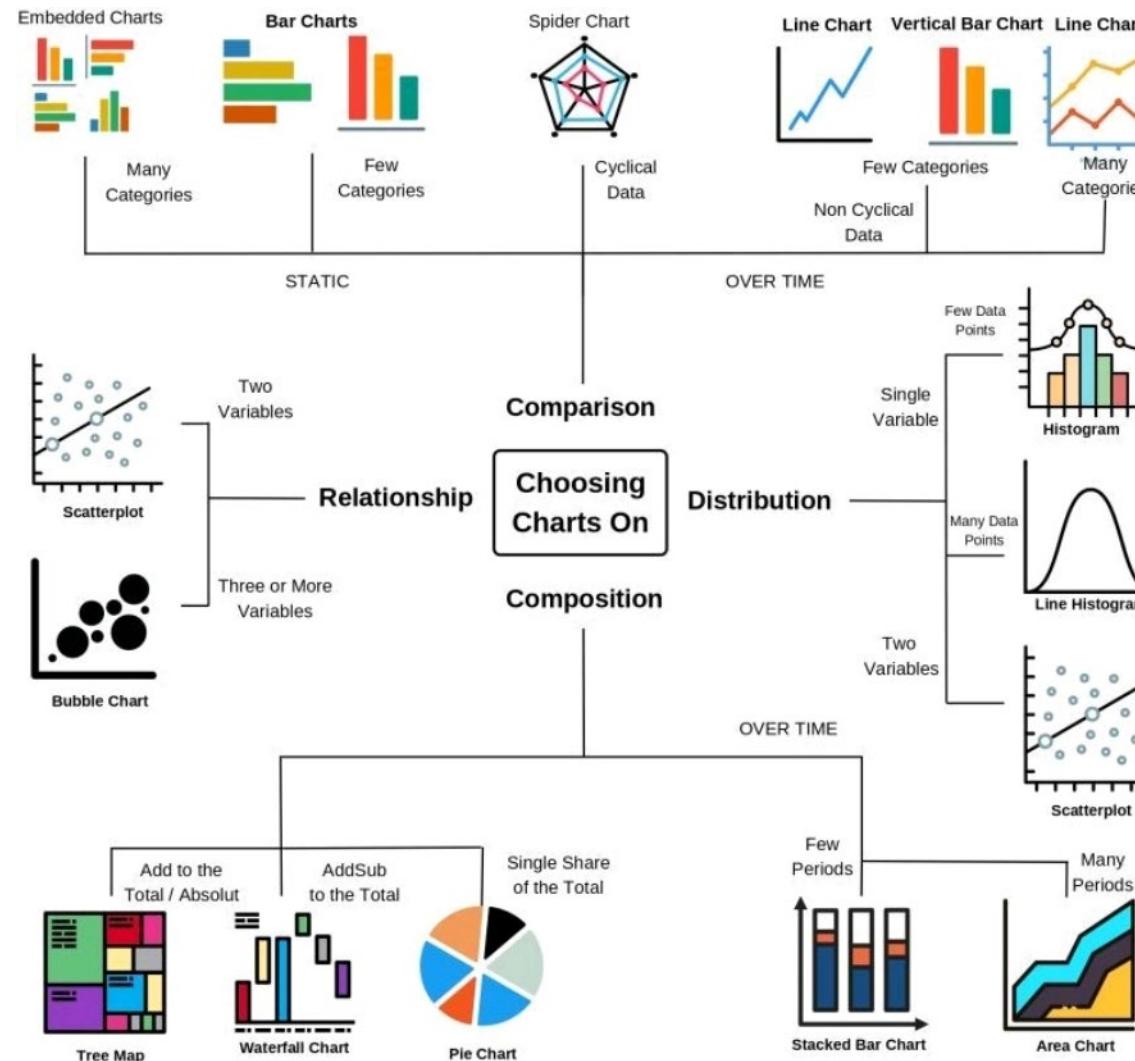


10. Don't jump into early conclusions

Anyone knows who this is?



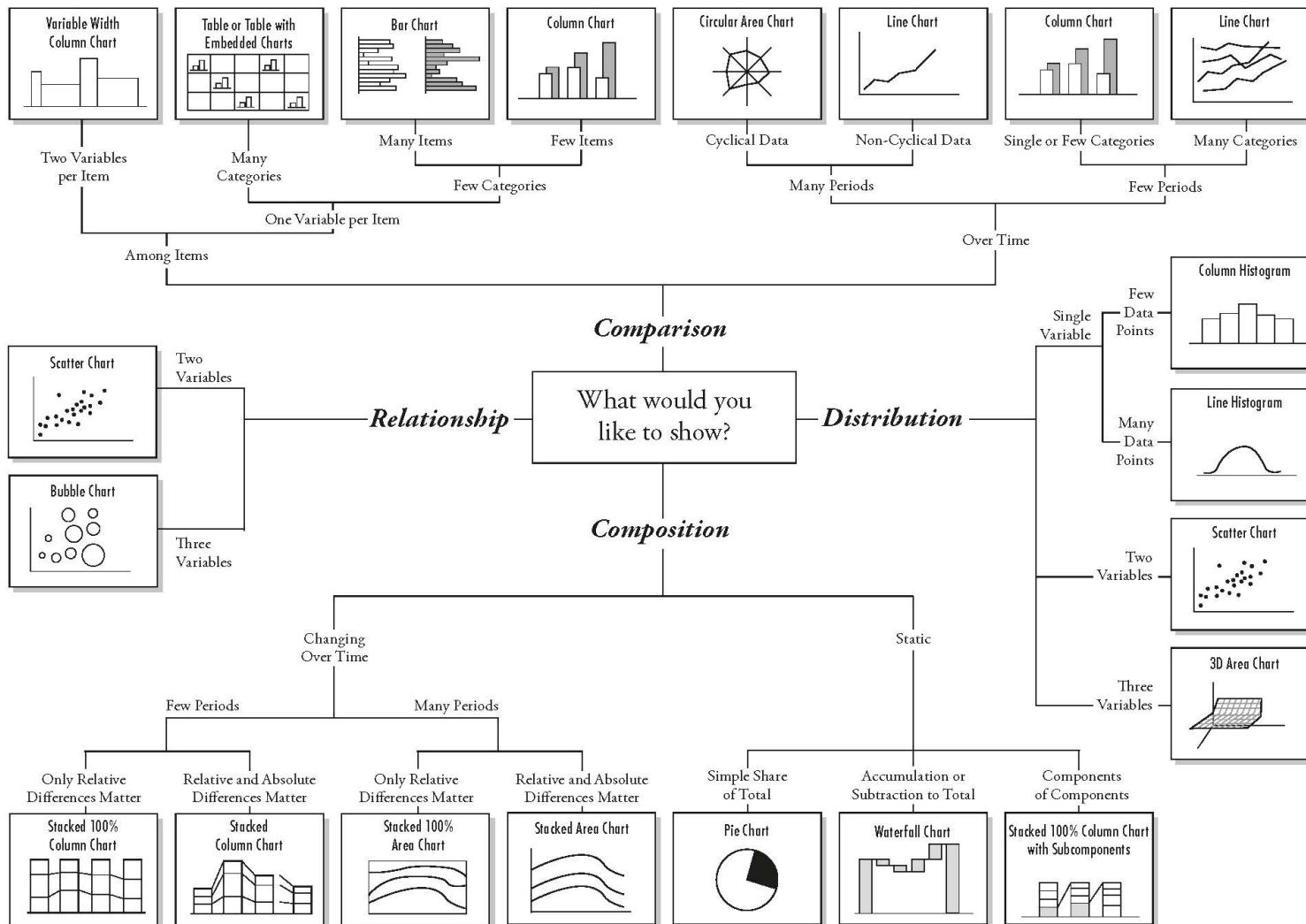
Is there a way to choose the “right” plot?



By Johnny Sholla

Add me on: [in](#) [git](#) [tw](#) [yt](#)

Chart Suggestions—A Thought-Starter



Deviation

Emphasise variation (CV) from a fixed reference point. Typically the reference point is zero but it can also be a target or a mean. Can also be used to show to show symmetry (positive/negative).

Example FT uses
Trade surplus/deficit, climate change

Diverging bar

A simple standard bar chart that can handle both negative and positive magnitude values.

Correlation

Show the relationship between two or more variables. Be mindful that unless you tell them otherwise, many readers will assume the relationships you show them to be causal (i.e. one causes the other).

Example FT uses
Inflation/unemployment, income and life expectancy

Scatterplot

The standard way to show the relationship between two continuous variables, each of which has its own axis.

Ranking

Use when an exact position in an ordered list is more important than its absolute or relative value. Don't be afraid to highlight the points of interest.

Example FT uses
Wealth, deprivation, league tables, constituency election results

Ordered bar

Standard bar charts display the ranks of values much more easily when sorted into order.

Distribution

Show values in a dataset and how often they occur. The shape or 'skew' of a distribution can be a memorable way of highlighting the lack of uniformity or inequality in the data.

Example FT uses
Income distribution, population (age/sex) distribution, revealing inequality

Histogram

The standard way to show a statistical distribution - keep the bars the same width to highlight the 'shape' of the data.

Change over Time

Give emphasis to changing trends. These can be short (interim) movements or extended series (long-term trends). Choosing the correct time period is important to provide suitable context for the reader.

Example FT uses
Share price movements, economic time series, sectoral changes in a market

Line

The standard way to show a changing time series. If data are irregular, consider markers to represent data points.

Column

The standard way to compare the size of things. Must always start at 0 on the axis.

Magnitude

Show the magnitude. These can be relative (just being able to see larger/larger) or absolute (need to see the numbers). Consider to show a 'count' number (for example, barrels, dollars or people) rather than a calculated per cent.

Example FT uses
Fiscal budgets, company structures, national election results

Stacked column/bar

A simple way of showing data on a map - should always be used with totals and use a sensible base geography.

Marimekko

A good way of showing the size and proportion of data at a single point - as long as the data are not too complicated.

Paired column

As per standard column but allows for multiple categories (columns) and a rate (line).

Paired bar

See above.

Donut

Similar to a pie chart - but the segments can be a good way of using space to include more information about the data (eg. roles).

Contour map

For showing areas of equal value on a map. Consider using colour schemes for showing +/- values.

Equalised cartogram

Converting each unit on a map to a regular and equal size - good for representing voting regions with equal numbers.

Chord

A complex but powerful diagram that can show 2-way (or net winner) in a matrix.

Network

Used for showing the strength and complexity of relationships of varying types.

Part-to-whole

Show how a single entity can be broken down into its component elements. If the reader's interest is solely in the size of the components, consider a magnitude-type chart instead.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Basic choropleth (categorical)

The standard approach for putting data on a map - should always be used with totals and use a sensible base geography.

Proportional symbol (categorical)

Designed to show the size of data rather than rates - be wary that small differences in data will be hard to see.

Flow map

For showing unambiguous movement across a map.

Contour map

For showing areas of equal value on a map. Consider using colour schemes for showing +/- values.

Equalised cartogram

Converting each unit on a map to a regular and equal size - good for representing voting regions with equal numbers.

Chord

A complex but powerful diagram that can show 2-way (or net winner) in a matrix.

Network

Used for showing the strength and complexity of relationships of varying types.

Spatial

Above from higher magnify only used when precise locations or geographical patterns in data are more important to the reader than anything else.

Example FT uses
Population density, natural resource locations, natural disaster risk/impact, catchment areas, variation in election results

Sankey

Show the reader volume or intensity of movement between different states or conditions. These might be logical sequences or geographical locations.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram that can show 2-way (or net winner) in a matrix.

Network

Used for showing the strength and complexity of relationships of varying types.

Flow

Show the reader volume or intensity of movement between different states or conditions. These might be logical sequences or geographical locations.

Example FT uses
Movement of funds, trade, migrants, lawsuits, information, relationship graphs.

Sankey

Show the reader volume or intensity of movement between different states or conditions. These might be logical sequences or geographical locations.

Waterfall

Designed to show the sequencing of data through a flow process, typically budgets. Can include +/- components.

Chord

A complex but powerful diagram that can show 2-way (or net winner) in a matrix.

Network

Used for showing the strength and complexity of relationships of varying types.

Visual vocabulary

Designing with data

There are so many ways to visualise data - how do we know which one to pick? Use the categories across the top to decide which data relationship is most important in your story, then look at the different types of chart within the category to form some initial ideas about what might work best. This list is not meant to be exhaustive, nor a wizard, but is a useful starting point for making informative and meaningful data visualisations.

FT graphic: Alan Smith; Chris Campbell; Ian Burt; Liz Fawcett; Graham Parish; Billy Ehrenberg; Shannon; Paul McCallum; Martin Shale



ft.com/vocabulary

FT

© Financial Times 2024-2025
This work is licensed under a Creative Commons
Attribution-ShareAlike 4.0 International License.