TELLING STORIES WITH DATA VIS:

DISTRIBUTION, CHANGE OVER TIME, AND MAGNITUDE

Dr Alexandra Anderson

The University Of Sheffield.

Q-Step

A step-change in quantitative social science skills

Funded by the Nuffield Foundation, ESRC and HEFCE

Nuffield Foundation

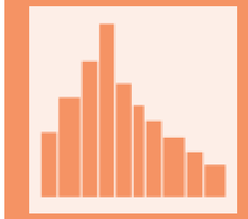
WE'LL BE DRAWING ON THE FINANCIAL TIME'S "VISUAL VOCABULARY"

- Distribution
- Change over time
- Magnitude

DISTRIBUTION

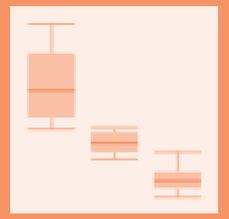
Chart types

histogram



The standard way to show a statistical distribution - keep the gaps between columns small to highlight the 'shape' of the data.

boxplot



Summarise multiple distributions by showing the median (centre) and range of the data

violin

Similar to a box plot but more effective with complex distributions (data that cannot be summarised with simple average).

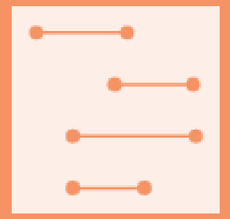
population-pyramis

A standard way for showing the age and sex breakdown of a population distribution; effectively, back to back histograms

dot-plot-strip

Good for showing individual values in a distribution, can be a problem when too many dots have the same value

dot-plot

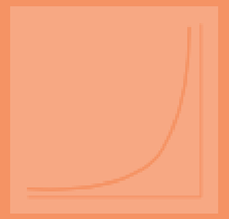


A simple way of showing the range (min/max) of data across multiple categories.

barcode

Like dot strip plots, good for displaying all the data in a table,they work best when highlighting individual values.

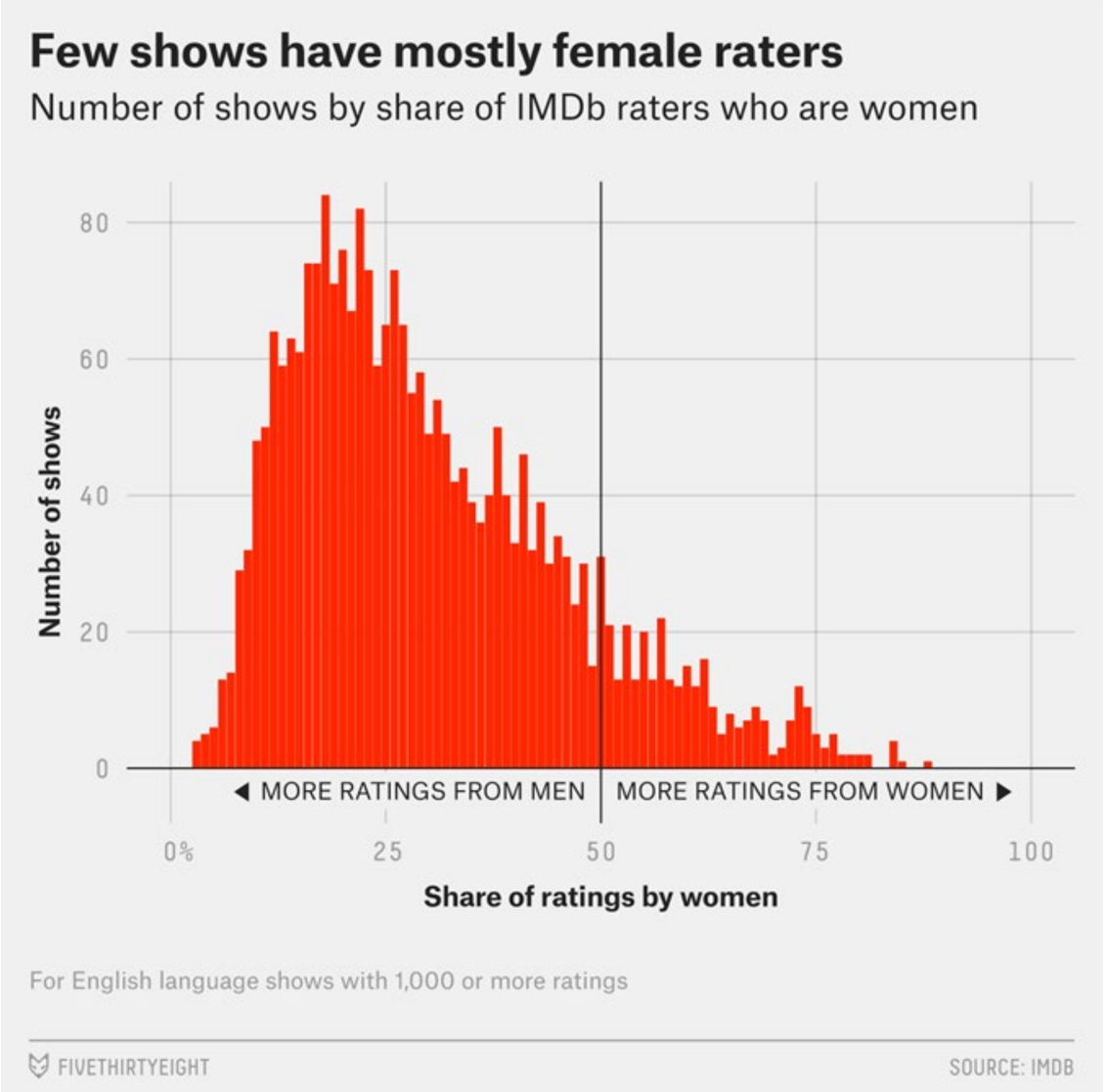
cumulative-curve



A good way of showing how unequal a distribution is: y axis is always cumulative frequency, x axis is always a measure.

Source: https://ft-interactive.github.io/visual-vocabulary/

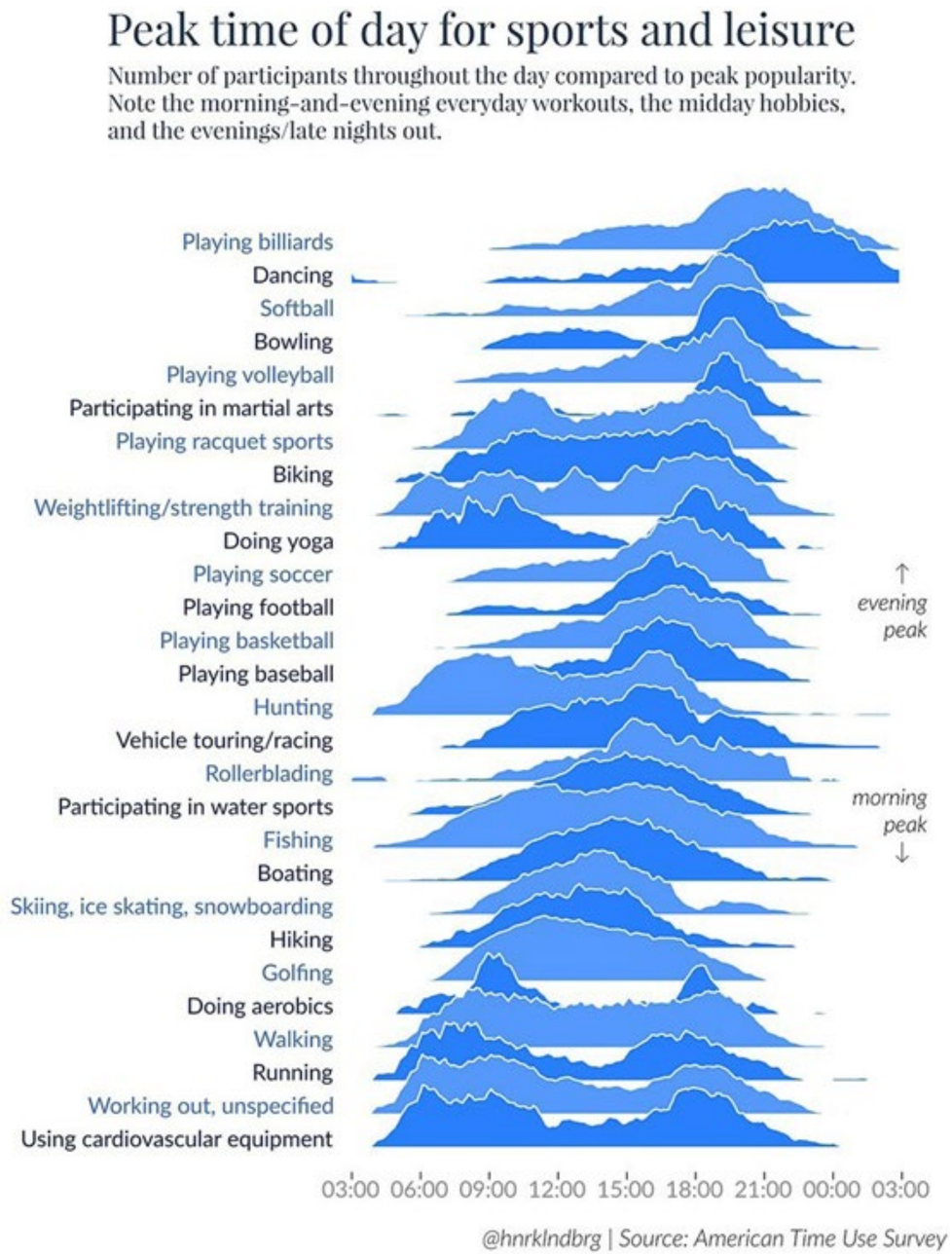
DISTRIBUTION



Source: FiveThirtyEight (2016). Men Are Sabotaging The Online Reviews Of TV Shows Aimed At Women

DISTRIBUTION

DISTRIBUTION

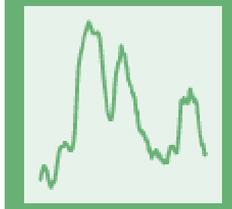


Source: FlowingData (2017). Peak Time for Leisure and Sports

CHANGE OVER TIME

Chart types

line



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

column-timeline

Columns work well for showing change over time - but usually best with only one series of data at a time

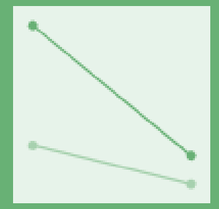
column-line-timeline

A good way of showing the relationship over time between an amount (columns) and a rate (line)

stock-price

Usually focused on day-to-day activity, these charts show opening/closing and hi/low points of each day

slope

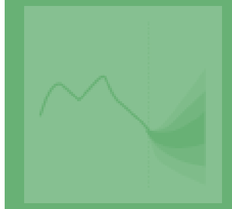


Good for showing changing data as long as the data can be simplified into 2 or 3 points without missing a key part of story

area

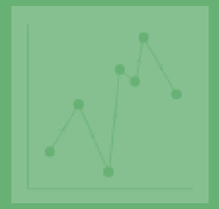
Use with care. These are good at showing changes to total, but seeing change in components can be very difficult.

fan



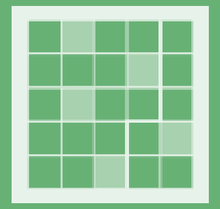
Use to show the uncertainty in future projections - usually this grows the further forward to projection

scatterplot-line-timeline



A good way of showing changing data for two variables whenever there is a relatively clear pattern of progression. Connected scatterplot

calendar-heatmap

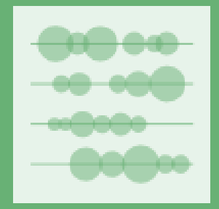


A great way of showing temporal patterns (daily, weekly, monthly), at the expense of showing precision in quantity

priestley timeline

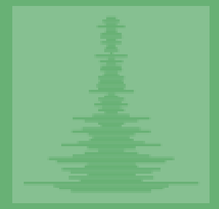
Great when date and duration are key elements of the story in the data

circles-timeline



Good for showing discrete values of varying size across multiple categories (eg earthquakes by contintent)

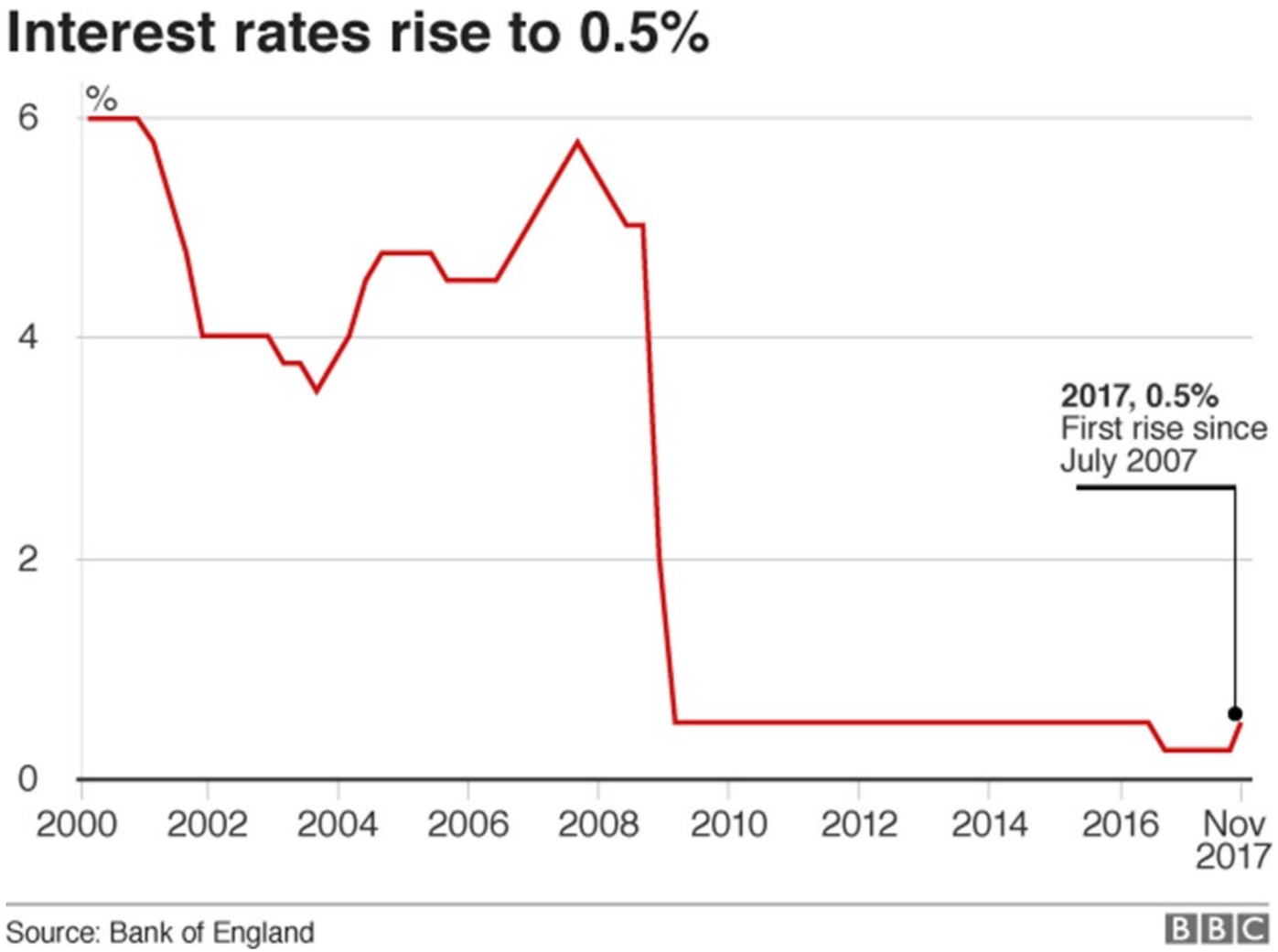
seismogram



Another alternative to the circle timeline for showing series where there are big variations in the data

Source: https://ft-interactive.github.io/visual-vocabulary/

CHANGE OVER TIME

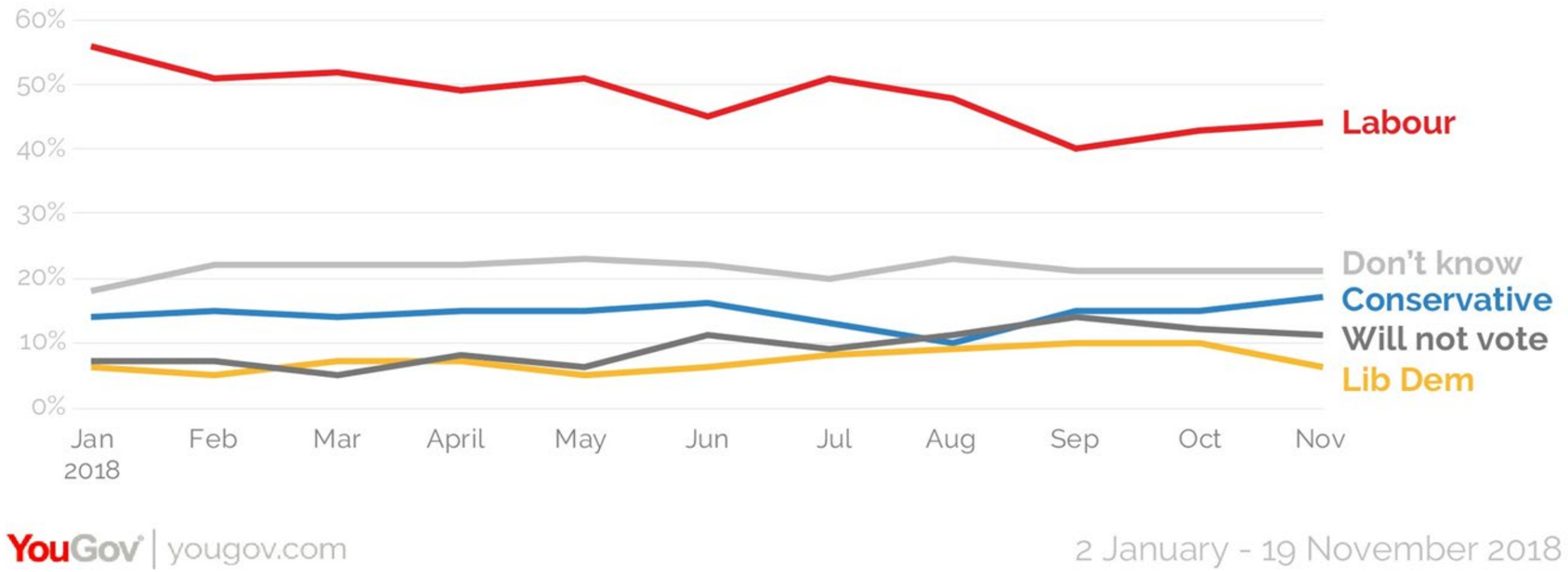


Source: BBC News

CHANGE OVER TIME

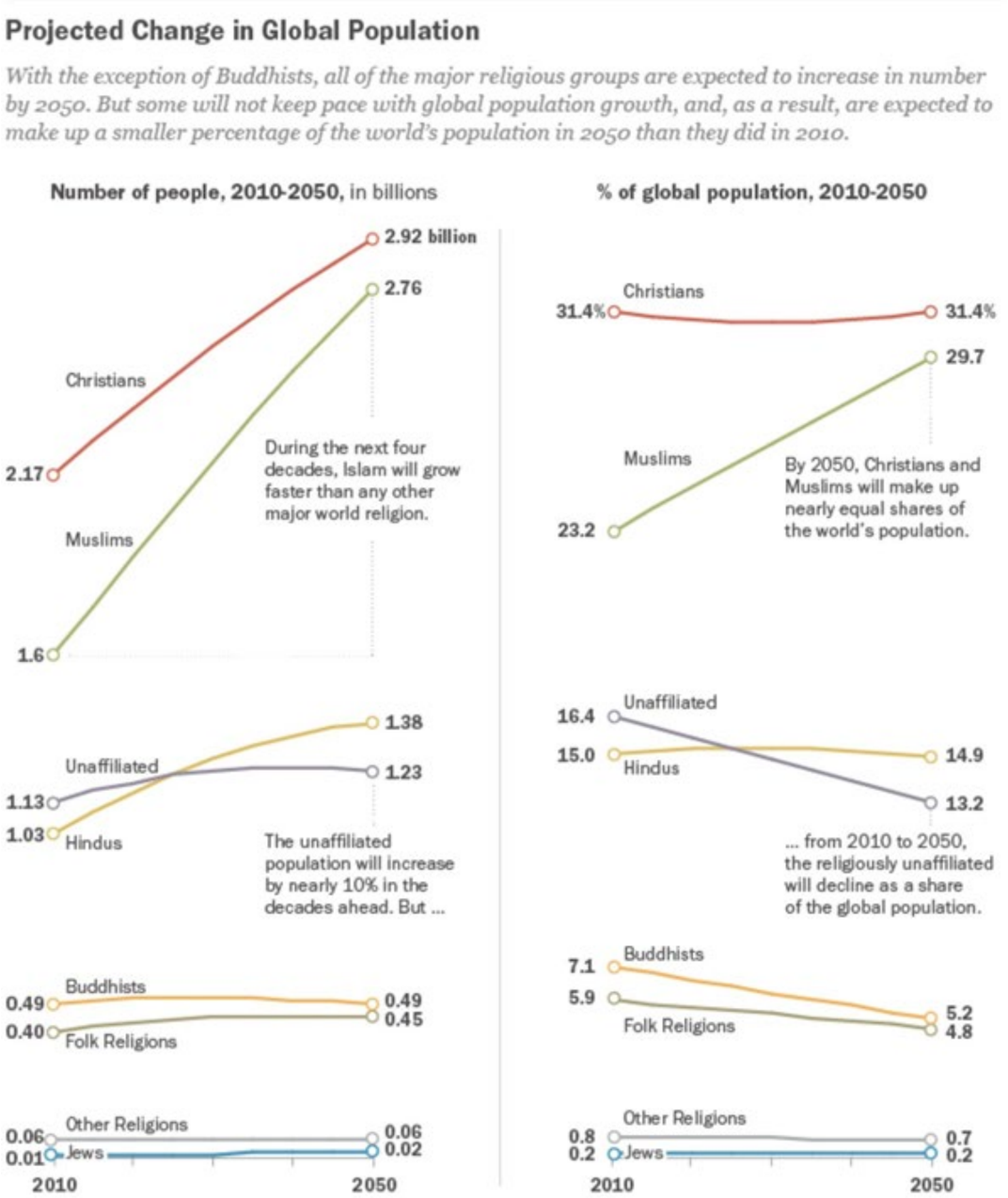
Young people are less likely to say they would vote Labour - but the Conservatives have not been the beneficiaries

Figures shown are monthly averages. Please note that these are not headline voting figures, but raw voting intention i.e. including those who said 'don't know' or 'will not vote'



Source: YouGov

CHANGE OVER TIME



Source: Pew Research Center

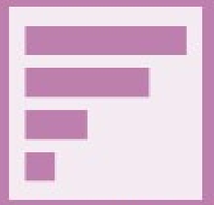
MAGNITUDE

Chart types

Column

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

Bar

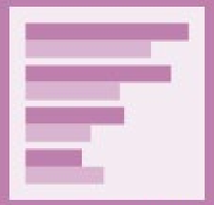


The standard way to compare the size of things. Must always start at 0 on the axis. Good when the data are not time series and labels have long category names

column-grouped

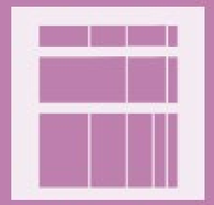
As per standard column but allows for multiple series. Can become tricky to read with more than 2 series

bar-grouped



As per standard bar but allows for multiple series. Can become tricky to read with more than 2 series

bar-stacked-proportional



A good way of showing the size and proportion of data at the same time - as long as the data are not too complicated

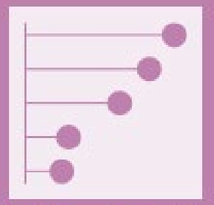
symbol-proportional

Use when there are big variations between values and/or seeing fne differences between data is not so important

isotope (pictogram)

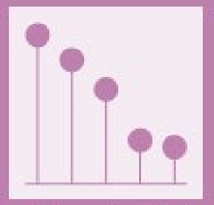
Excellent solution in some instances - use only with whole numbers (do not slice off an arm to represent a decimal).

lollipop-h



Lollipop charts draw more attention to the data value than standard bar/column and can also show rank effectively

lollipop-v



Lollipop charts draw more attention to the data value than standard bar/column and can also show rank effectively

Radar

A space-efficient way of showing value pf multiple variables - but make sure they are organised in a way that makes sense to reader.

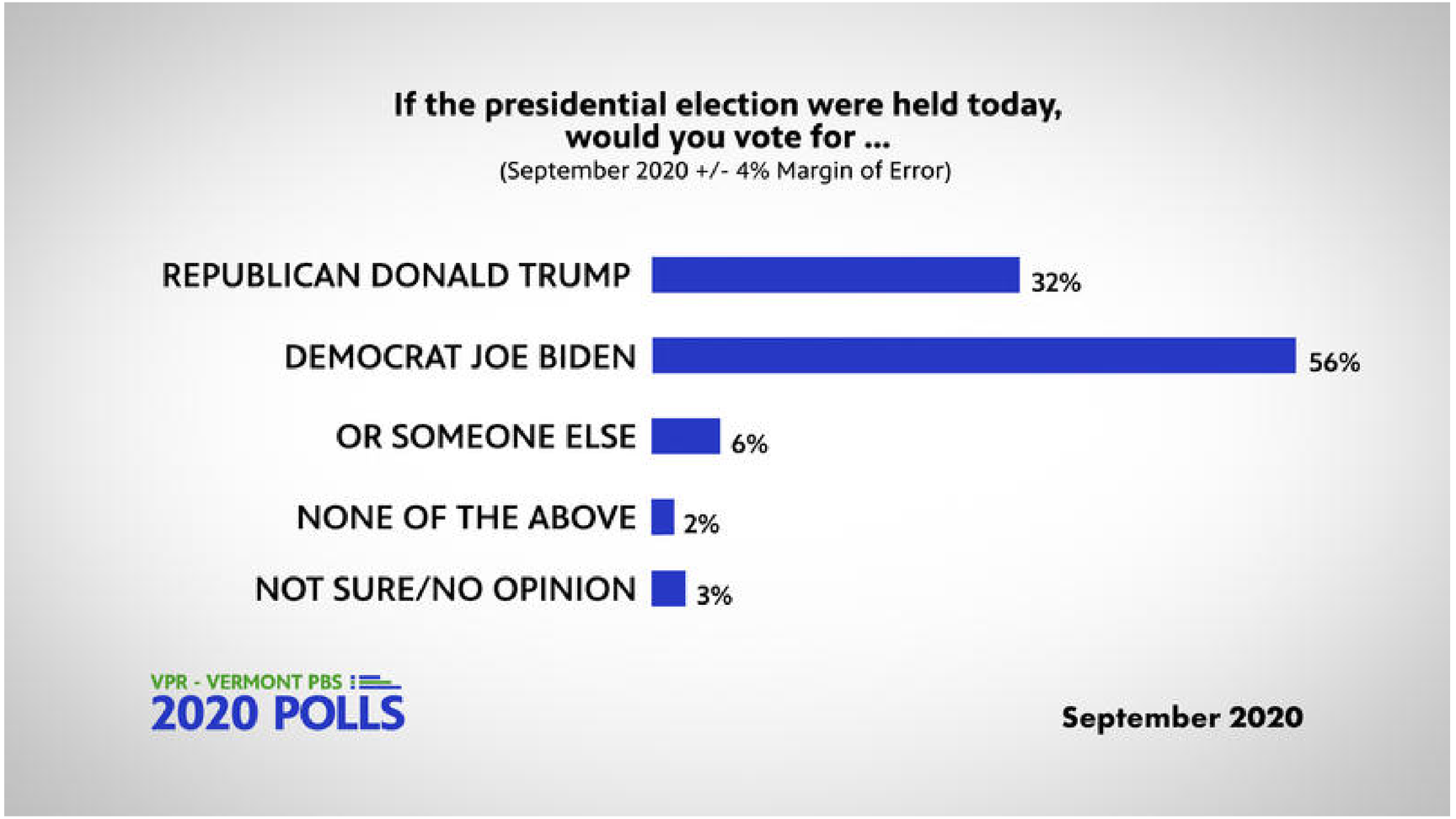
Bullet

Parallel coordinates

An alternative to radar charts - again, the arrngement of the variables is important. Usually benefits from highlighting values

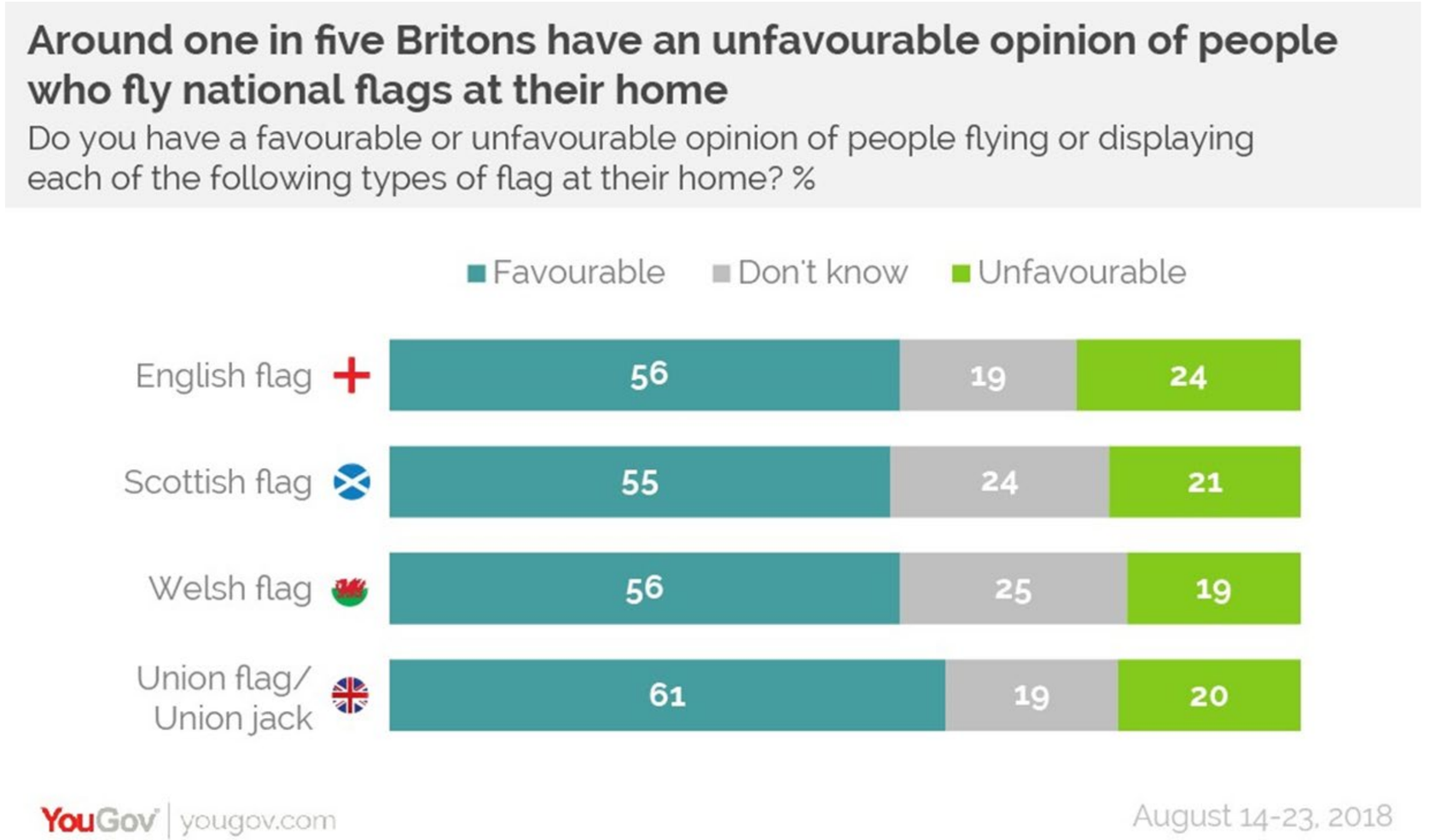
Source: https://ft-interactive.github.io/visual-vocabulary/

MAGNITUDE



Source: VPR.org

MAGNITUDE



Source: YouGov