

Open

gn in

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

M



# Why you sometimes need to break the rules in data viz

Best practices help us to avoid common pitfalls in data visualisation — but we shouldn't follow them blindly



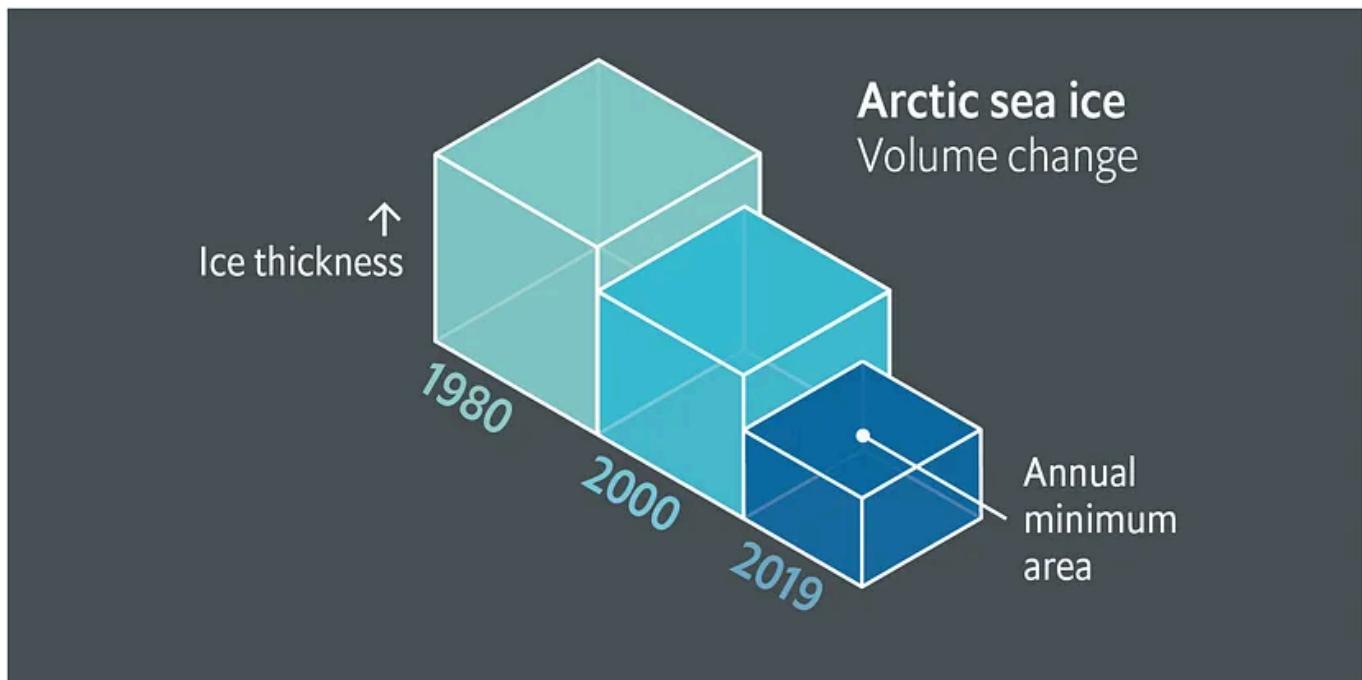
Rosamund Pearce · Follow



Published in The Economist · 10 min read · Feb 28, 2020

2.1K

Q 8



It's easy to visualise data badly. Choose the wrong chart type, omit labels or overload the graphic and you can quickly end up with something that is

## confusing, misleading or that fails to tell a story

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

To

field of data visualisation has matured. From steering clear of pie charts to getting rid of “chartjunk”, these are the conventions anyone dabbling in data viz should learn first.

But sometimes it’s helpful to revisit the things we learn early on with the benefit of experience. In the six years that I have been making data visualisations, I’ve found that sticking doggedly to the “rules” can sometimes impede effective visual communication. It’s easy to get bogged down in following technicalities, and lose sight of what actually *works*.

The trouble is that best practices tend to have little nuance, and there are times when they may be irrelevant, or worth the cost of breaking. Since joining *The Economist* I have broken many of them myself. But if you are going to do the same, it’s helpful to understand why they became rules of thumb in the first place. As the old saying goes, you have to know the rules before you can break them.

With this in mind, and with the help of *The Economist’s* archive, I’ve re-examined five things I was taught not to do when I first started out in data visualisation.

### 1: Don’t break the y-axis scale

*The Economist* occasionally gets complaints from readers about charts where the numerical axis doesn’t start at zero. It’s no wonder they are treated with

suspicion. A truncated axis can cause small differences to appear larger or

m To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

s

reputation. But sometimes small variations between data are important, and you need to limit the range of the chart in order to make the story visible:

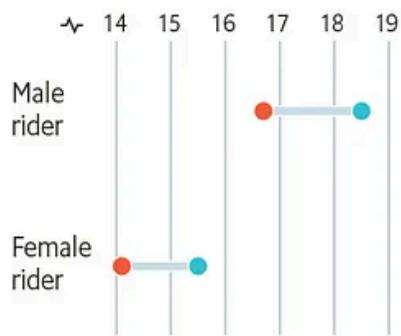
## Tipping points

United States, tipping habits of Uber passengers

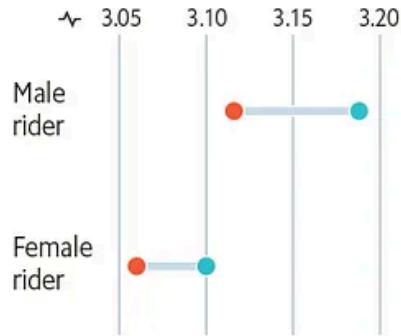
● Female driver ● Male driver

Aug-Sep 2017

Share of rides tipped, %

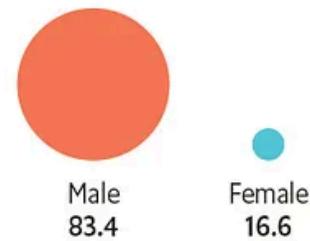


Average amount tipped, \$



Rides by driver's gender

% of total in study

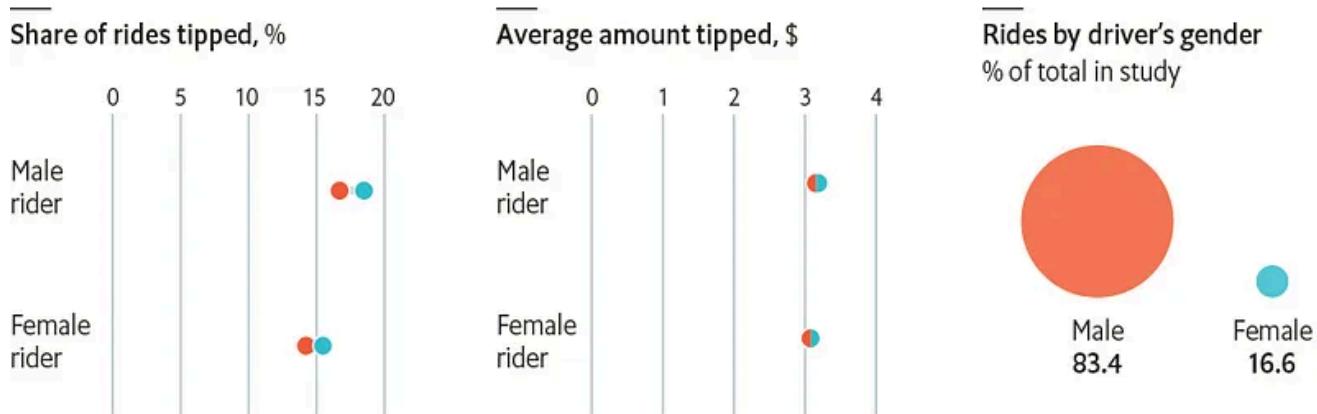


Source: "Evidence from a nationwide tipping field experiment" by B. Chandar et al., *National Bureau of Economic Research*, October 2019

Original chart

**TIP****UN****AUG**

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.



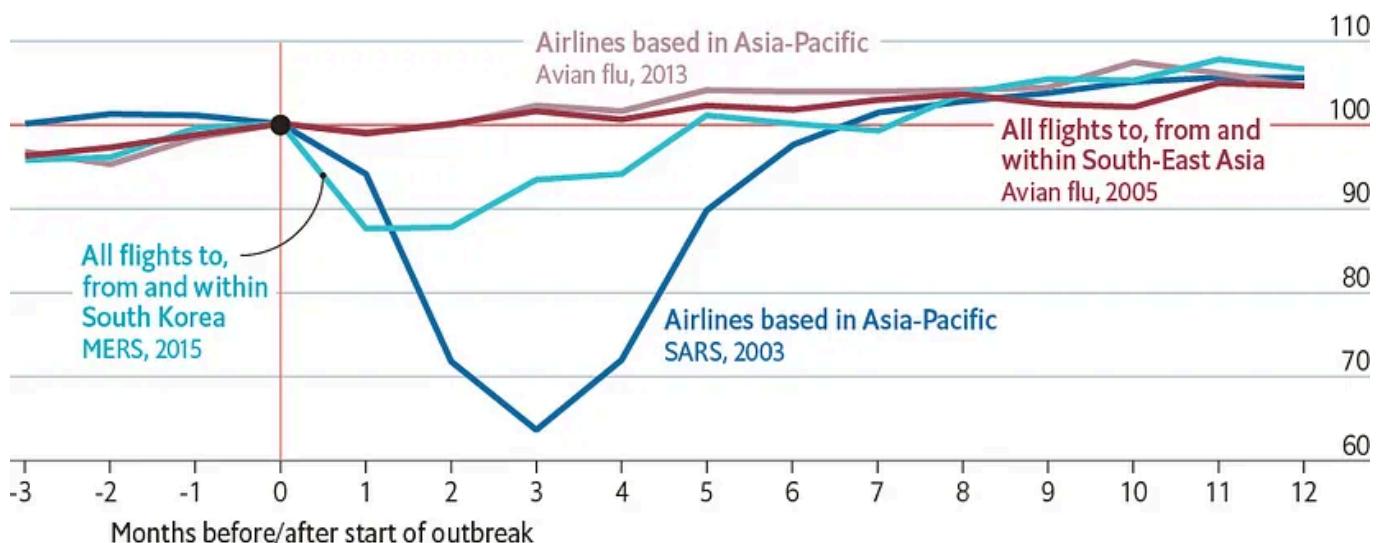
Source: "Evidence from a nationwide tipping field experiment" by B. Chandar et al., *National Bureau of Economic Research*, October 2019

The same chart with unbroken axes

In this chart I made about the tipping habits of Uber riders, starting the scale at zero hides the story because the differences — though significant — are small and not close to zero. It's necessary to break the axes to show these differences. At *The Economist* we always use a symbol to signpost a broken scale, except for index charts:

**Ta****Imp****Sta**

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.



Source: IATA

\*One revenue-passenger km equates to flying one paying passenger one kilometre

100 is treated as the baseline in index charts so a broken scale icon isn't necessary

The one time you shouldn't break an axis is when making a bar or column chart because it breaks the relationship between the rectangle's dimensions and the data.

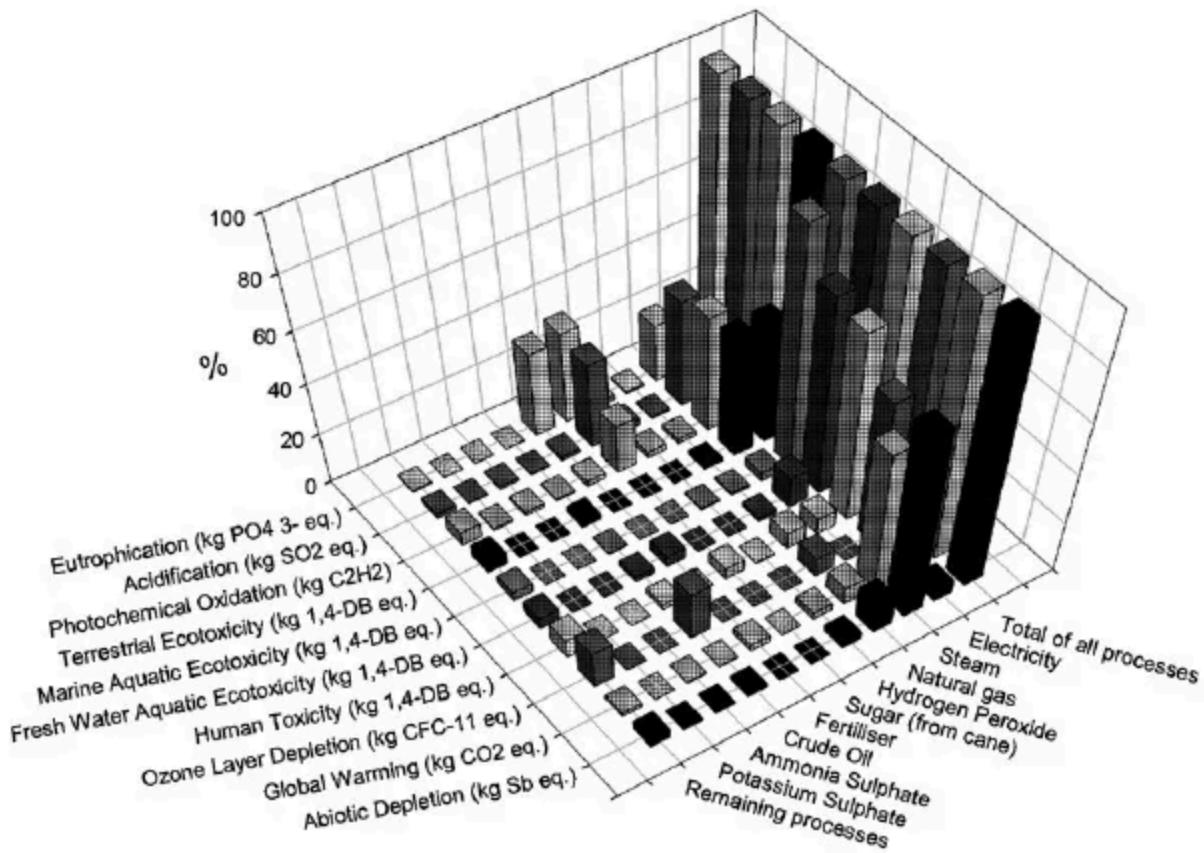


The Economist

In our modern style (right, from 2019) we use a thermometer chart instead of a bar chart (left, from 2006)

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

## 2: Don't go 3D



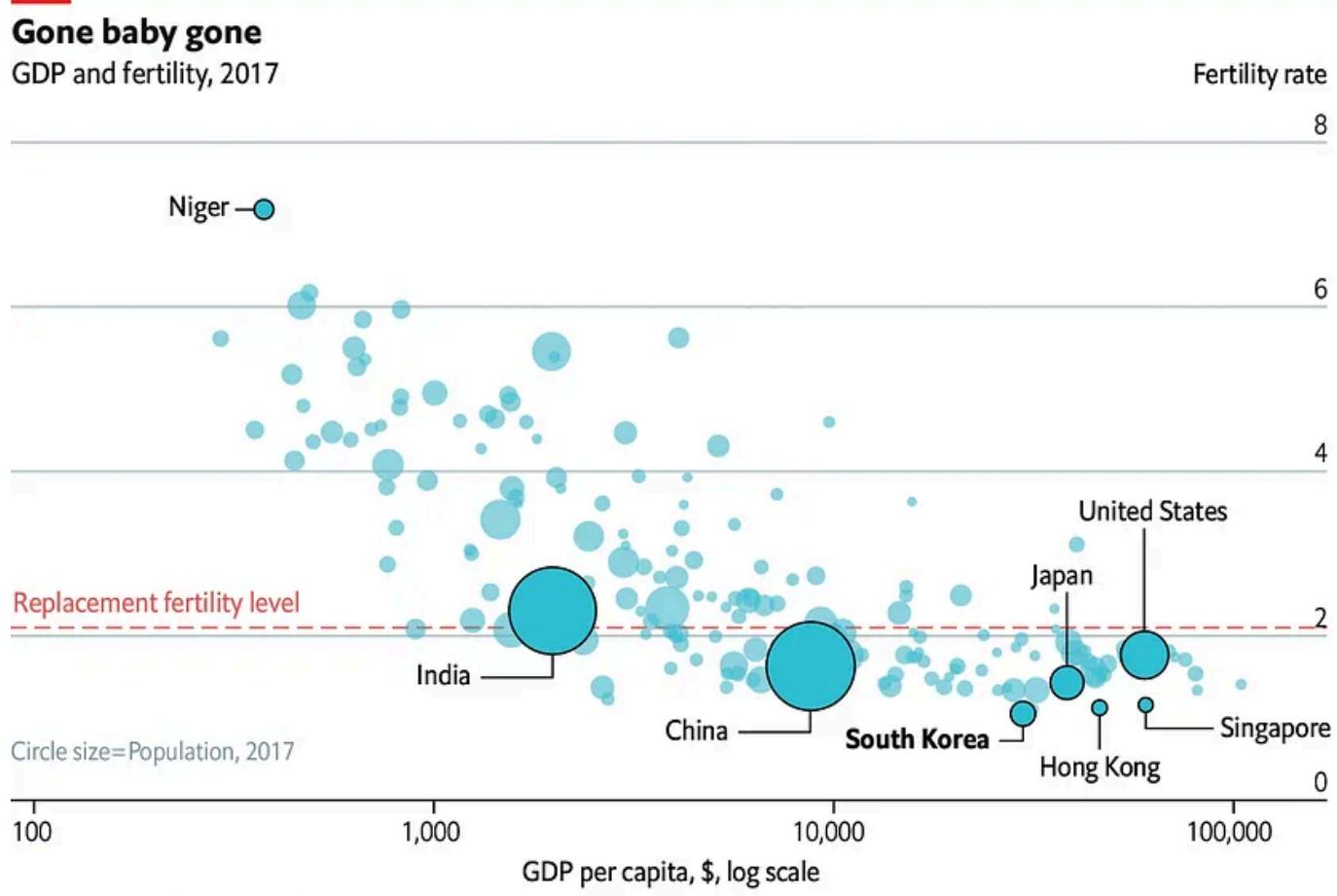
Can you accurately decipher this? I can't. Source: [WTF Viz](#)

Ever since Microsoft Excel launched a thousand 3D pie charts into business meetings, 3D charts have been among the most vilified of data visualisations. There are two types of 3D visualisations: ones where the third dimension is used to depict a third variable, and others where it is purely cosmetic. Decorative 3D is particularly frowned upon as it needlessly distorts the data in multiple ways: perspective causes nearer items to appear

larger, and using volume to represent a single variable magnifies the

ap To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

Using 3D to depict three variables, while slightly more justifiable, often results in a chart that is difficult to interpret. And for a static chart, it's impossible to pick an angle that provides one objective view of the data — whichever way you view it, some elements will be emphasised more than others. Mostly it's better to find another way to show a third variable — by bubble size, for instance:

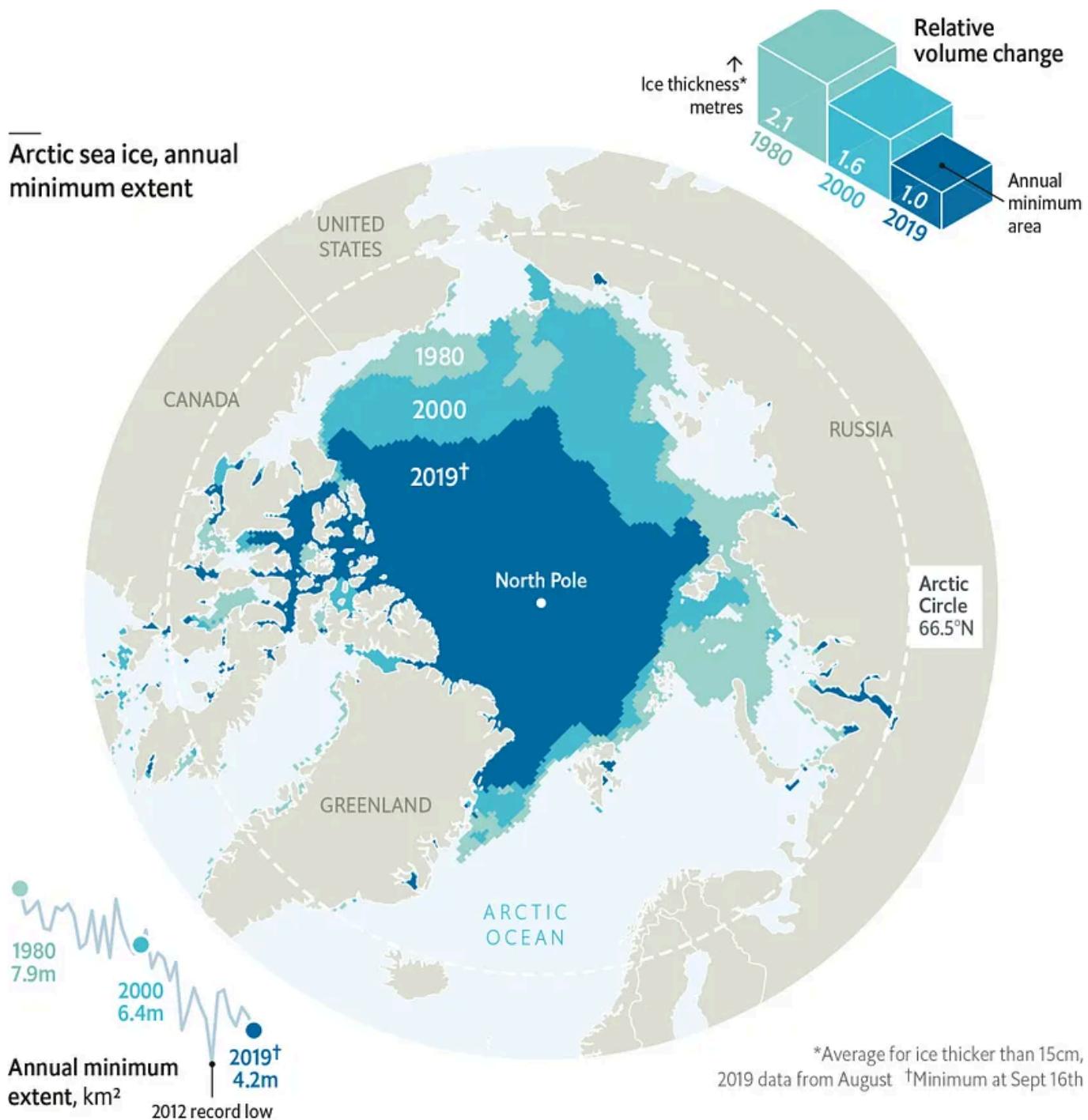


Three variables in 2D

But 3D can work well under certain circumstances. In the following visualisation for *The Economist's* Graphic Detail section, we used a series of

## cubes to show the relative volume change of Arctic Sea ice.

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.



Relative volume change of Arctic sea ice, as three ice cubes (top right)

I think we got away with breaking the convention against using 3D because:

- There is a good reason for going 3D, given that we want to represent

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

- We used an isometric projection — which lacks perspective — so there is no distortion between elements in the foreground and background
- Rather than using a scale, which can be harder to read in 3D, we directly labelled the data
- Smaller elements go in front, so as little is obscured as possible

3D can also work better in visualisations that are animated or interactive, because you aren't restricted to one view.

### 3: Get rid of “chartjunk” to maximise the data-ink ratio

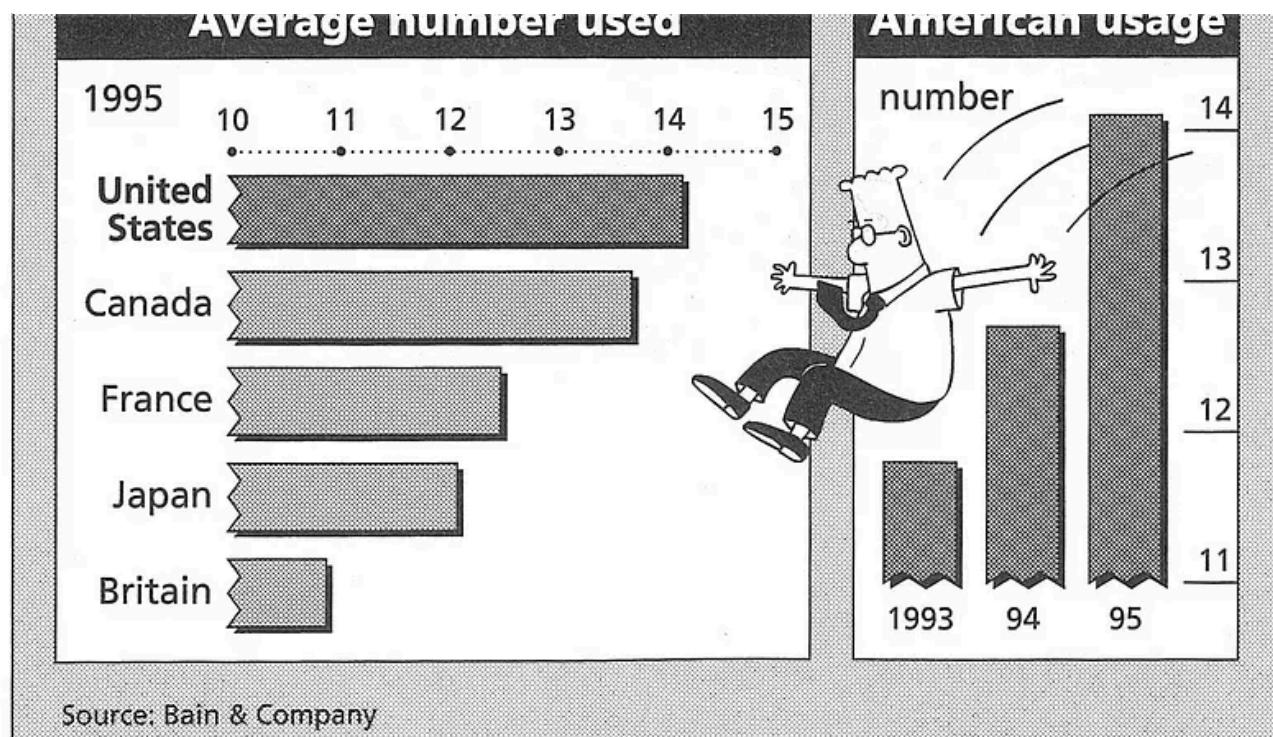
The term “chartjunk” was coined by the statistician Edward Tufte to describe anything on a chart that is not central to interpreting it. Tufte deemed things like illustrations and fancy fonts to be an unwelcome distraction:

“Cosmetic decoration, which frequently distorts the data, will never salvage an underlying lack of content. If the numbers are boring, then you've got the wrong numbers.”

“Every bit of ink on a graphic requires a reason. And nearly always that reason should be that the ink presents new information”

Edward Tufte, *Visual Display of Quantitative Information*, 1983

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.



Tufte would probably not approve of the cartoon in this chart from a 1997 edition of The Economist. Or the shading. Or the broken scales...

Tufte's ideas helped prompt a shift towards minimalism in data visualisation that remains influential. It was probably one of the first rules that I learned. But there has been some pushback against stripping charts back to their bare bones, including from data visualisation consultant Alberto Cairo.

*“...gridlines in a time-series chart, or unobtrusive and subtle icons that identify the topic the chart discusses — might not be junk at all. Far from obstacles to understanding, they may enhance understanding.”*

Alberto Cairo, *The Functional Art*

A study in 2010 also questioned a wholly minimalist approach to chart

de To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy,  
including cookie policy.

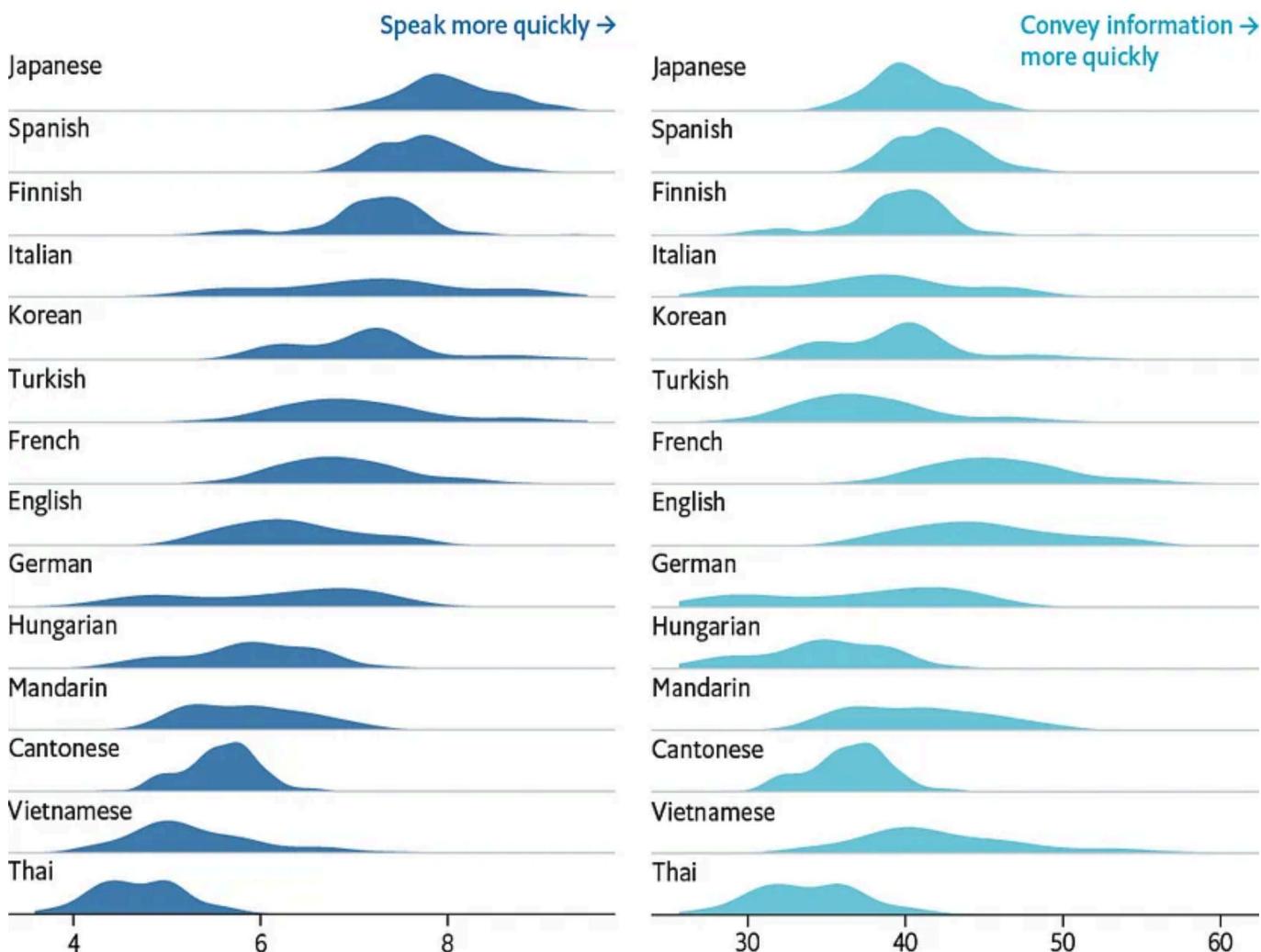
an “redundant” labels — that reiterate the main trend in a visualisation — were better understood, while charts with icons were better recalled. No wonder most information designers don’t follow Tufte’s advice to the letter.

While I favour decluttering charts, I also want to help the reader out. There is a distinction between superfluous decoration and extra visual elements that provide a cue to the topic or reinforce a key finding. There’s nothing wrong with a bit of hand-holding — particularly with a challenging subject.

In the density plot that I made below, the labels “Speak more quickly” and “Convey information more quickly” repeat information already given in the subtitles. Under Tufte’s definition they would be considered as “junk”, as they don’t present new information — but they help the viewer understand the axes much faster.

**Sa**

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

**Syl**

Source: "Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche" by Christophe Coupé, Yoon Mi Oh, Dan Dediu and François Pellegrino, *Science Advances* (2019)

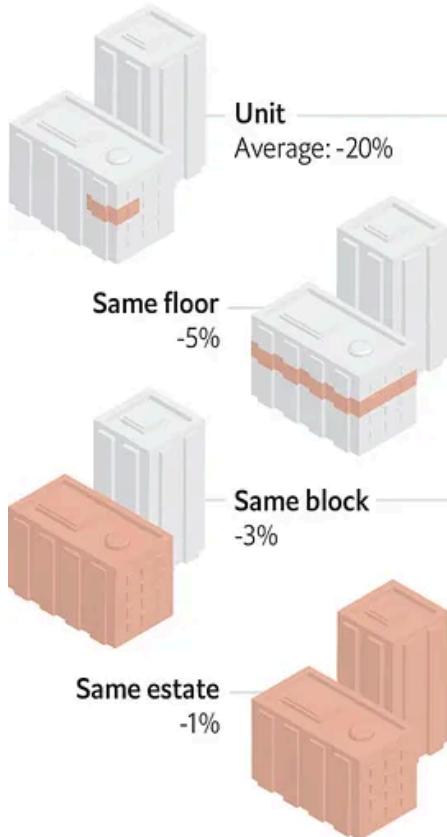
Why are some languages spoken faster than others?

Another chart I made about haunted houses in Hong Kong probably didn't *need* the illustrations of a Hong Kong tower block — but the infographic is more visually interesting with them, and they allow you to identify the topic at a glance.

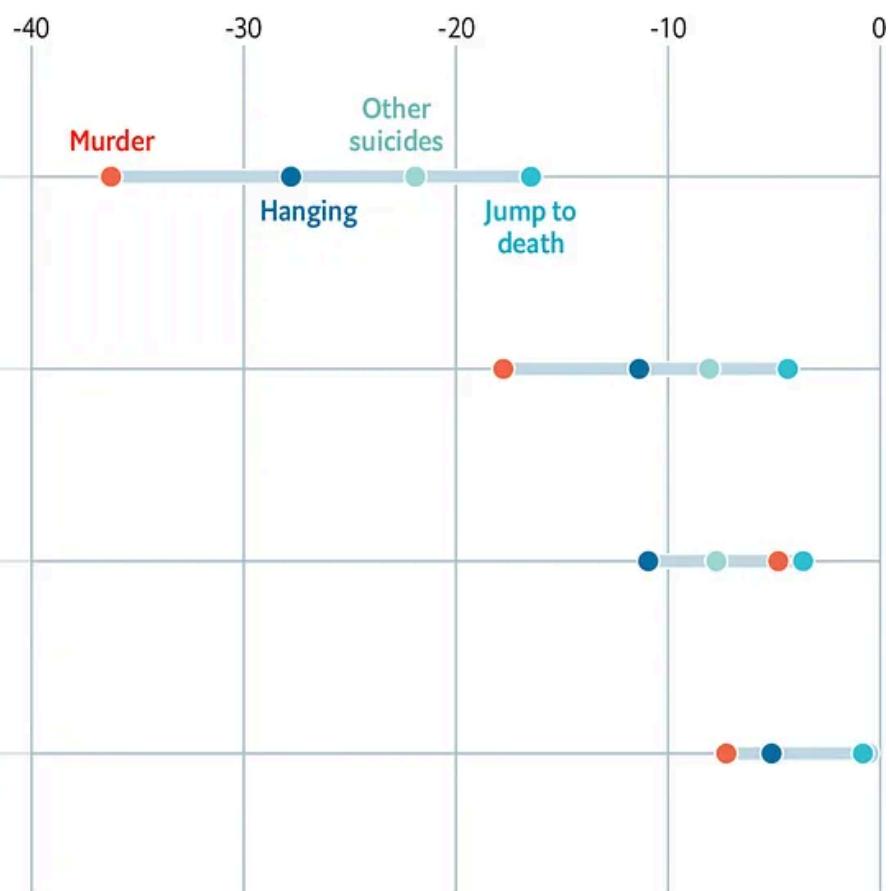
**Gr**

**Ho** To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

### By proximity to "haunted"† unit



### By cause of death



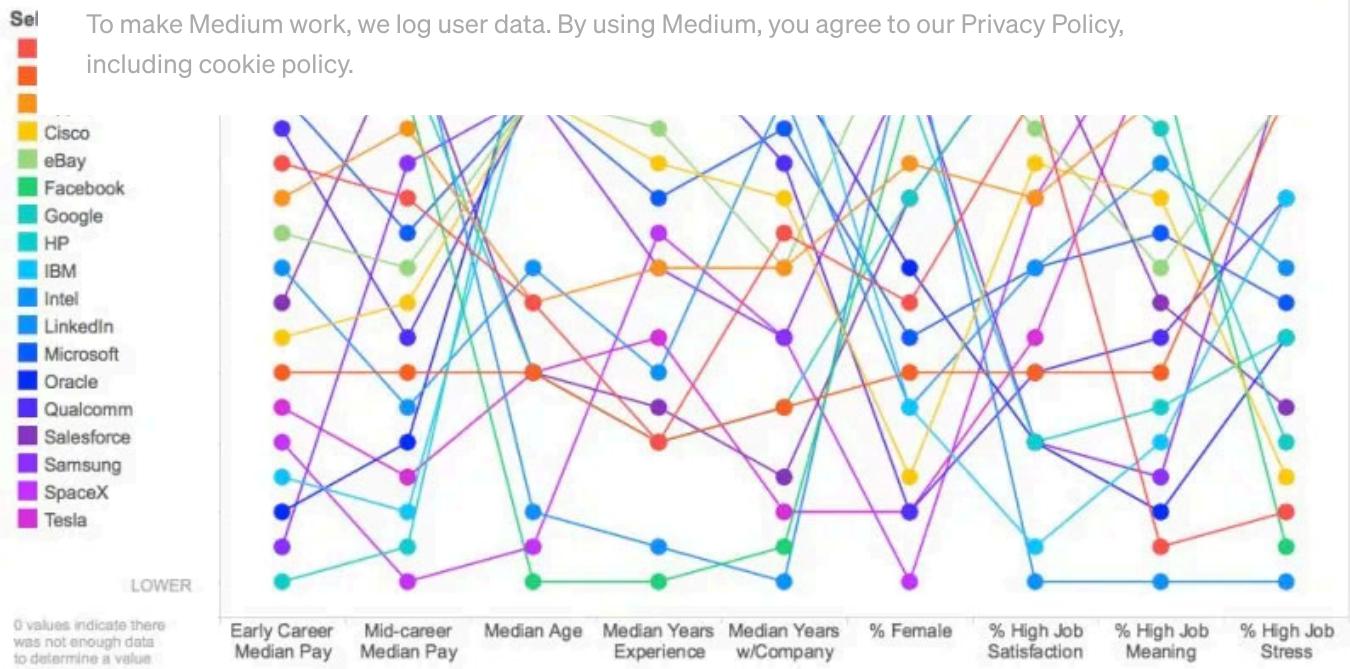
Source: "Spillovers in asset prices: the curious case of haunted houses" by Utpal Bhattacharya, Daisy Huang and Kasper Meisner Nielsen, Hong Kong University of Science and Technology working paper (2019)

\*2000-2015

†Due to an unnatural death

Haunted houses have a chilling effect on the property market

## 4: Use fewer than seven colours

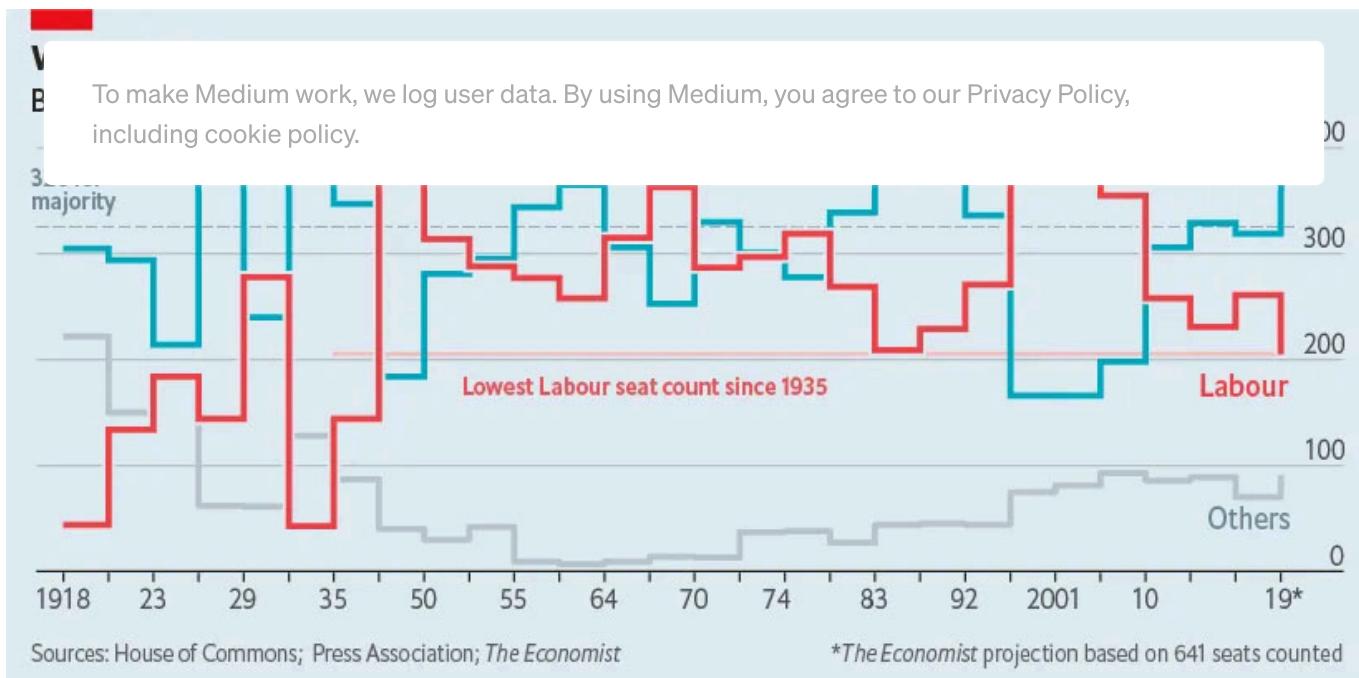


Loads of colours = loads of confusion. Source: [WTF Viz](#)

While the exact figure varies somewhat, there is a general consensus that using too many colours in a categorical colour scale is a bad idea. Our eyes aren't great at distinguishing subtle differences in hue, compared with other visual methods of depicting data. If you use a lot of colours it can be hard to work out what's going on at a glance, and similar-looking categories may be misread, particularly if the coloured area is small (eg, a thin line).

It often looks a bit messy too. Using a limited palette can look cleaner and more striking, especially when used consistently throughout an infographic or data dashboard.

So what do you do if your data has a lot of categories? You could use another visual method to show the grouping, such as a label or placing related elements adjacent to one another. Another solution is to limit the number of categories by creating an 'other' grouping:



Using an 'other' category makes the chart less noisy and places the focus on the two major parties. From [December 13th 2019](#)

Using different tints of the same shade can also make the design look cleaner, if you have a number of related categories.



Similar tints like this are better labelled directly. From 14th November 2019

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

But

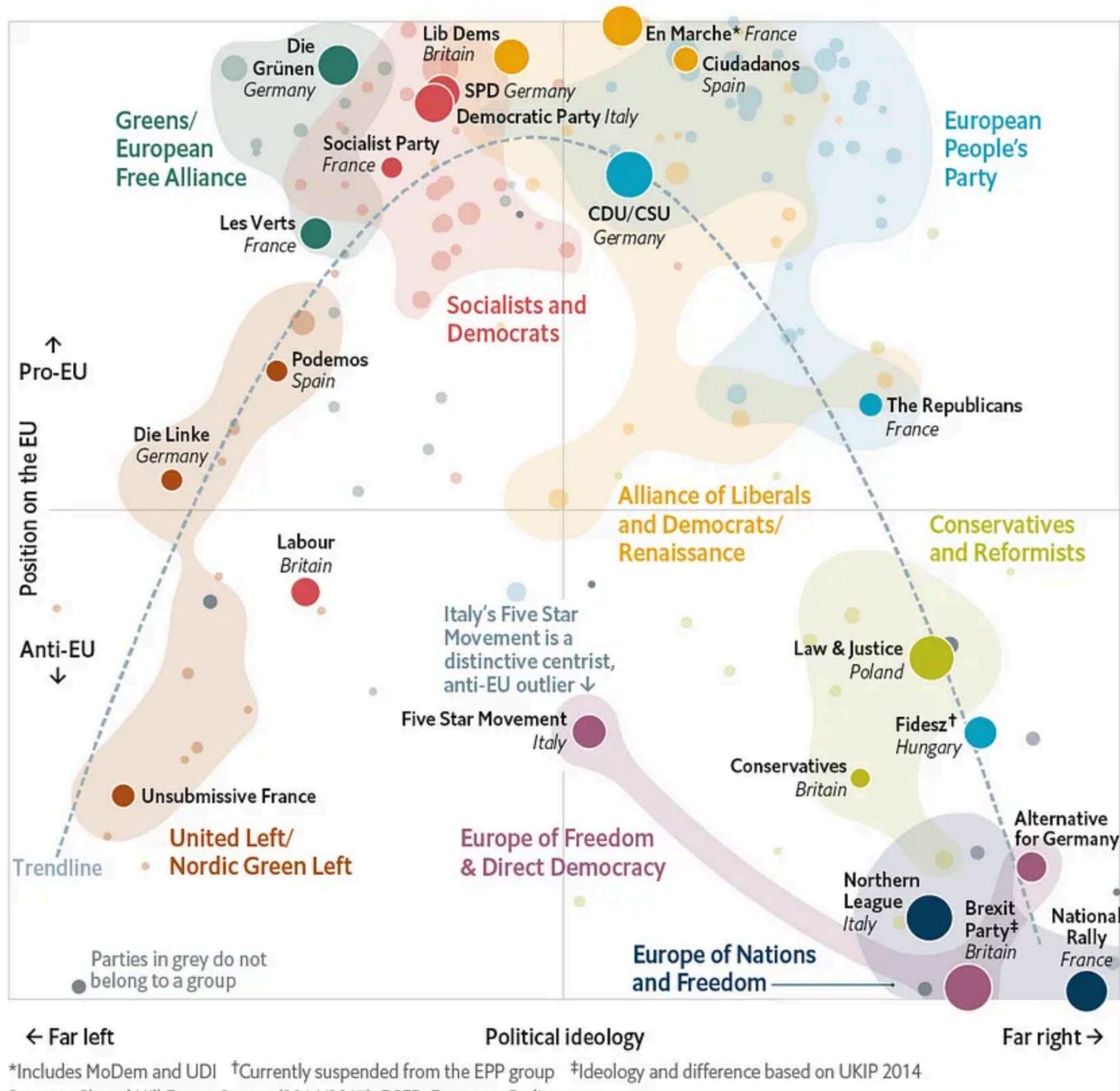
can use to make your visualisation easier to read. Consider the following from *The Economist's* Graphic Detail page:

## Art of Data Visualization: Political parties in the European Parliament

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

By

Party name —●— Parliamentary grouping 2019 election, provisional results, seats ○ 10 ○ 20



\*Includes MoDem and UDI

†Currently suspended from the EPP group

Sources: Chapel Hill Expert Survey (2014/2017); ECFR; European Parliament

Centrist liberals gained the most power in the EU parliament

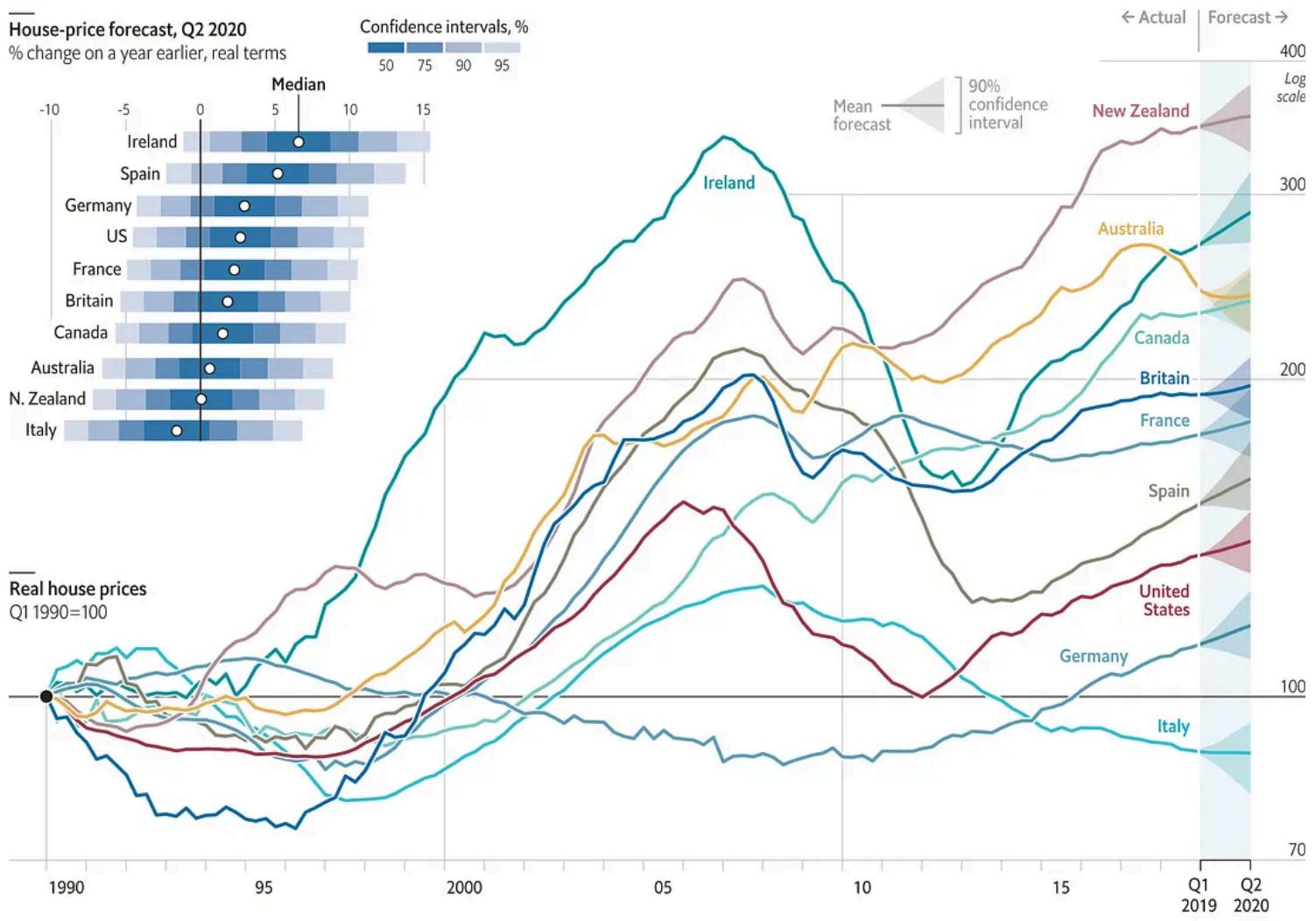
This chart uses nine colours. I think it gets away with it because:

- Categories are labelled adjacent to the group, so you don't need to match

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

- A shape helps reinforce the grouping
  - The coloured regions cover a relatively large area and use sufficiently distinct hues

**A decade after the financial crisis, house prices are at new highs**



For now, residential-property prices are likely to keep rising

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

It's like this one, because:

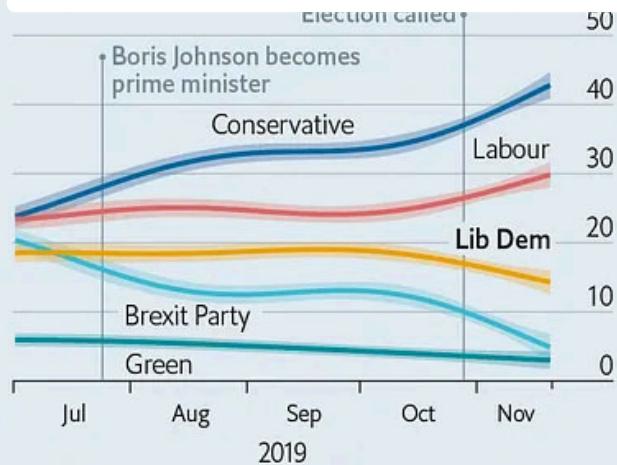
- The colours here have been applied to a smaller area, but this is less of a problem with a line chart as each colour is used for just one shape. You don't rely on colour to identify a group of many disparate shapes as you do in the bubble scatter above
- Care has been taken when choosing colours to ensure that they are easy to follow when they cross over each other. A subtle white outline also helps to distinguish each line
- Again, categories are labelled directly, so correctly matching colours isn't essential to reading the chart

## 5: Avoid colour scales with red and green

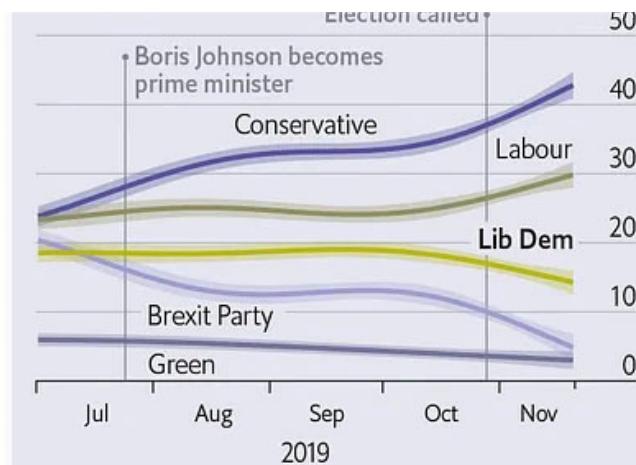
One in 12 men have colour-blindness, the most common form of which is red-green colour blindness (Deutanopia). Because of this, I was taught to avoid creating data visualisations that use red and green. And this is certainly a good starting point for making your visualisations more accessible.

But sometimes it's hard to avoid a particular selection of colours. For example, using anything other than the main colour associated with a particular political party may confuse the reader. In the UK that means using red for Labour and green for the Green Party. Luckily, there are accessible combinations of red and green:

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.



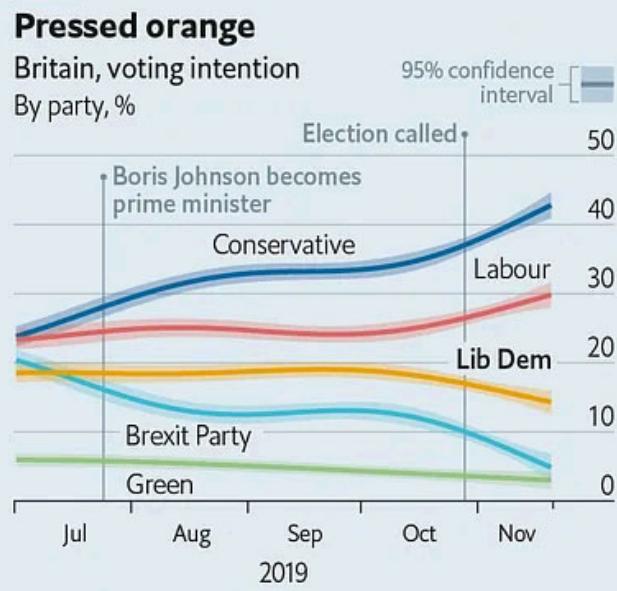
Source: Politico Poll of Polls



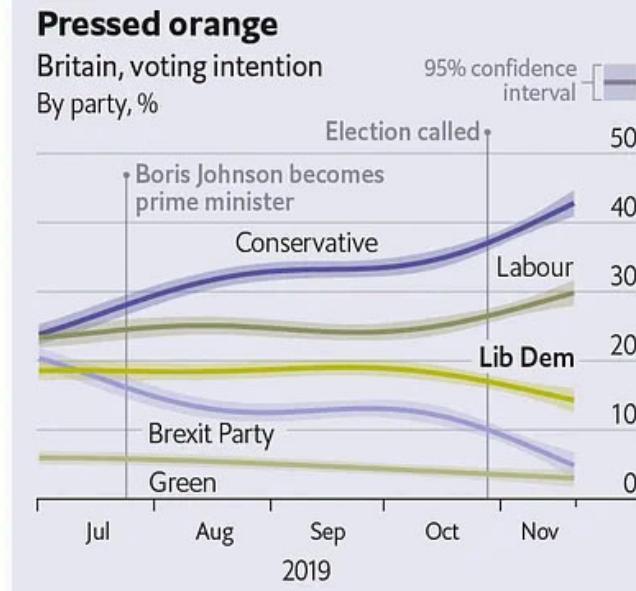
Source: Politico Poll of Polls

The red and teal in The Economist chart colour scheme are OK for people with Deutanopia (simulation on right)

'Red' and 'green' are nebulous and somewhat subjective categories. Here, the chart uses a shade of teal that is at the bluer end of what I would describe as green. The [colour-blindness simulator](#) shows that this is fine to use with this shade of red.



Source: Politico Poll of Polls



Source: Politico Poll of Polls

When I tweak the hue to a grass-green they look a lot more similar but are

sti To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

sa

difference between a light shade of green and a dark shade of red. Playing around with hue, saturation and lightness can help you find shades that are distinguishable.

Another way to get away with using red and green is to ensure that colour isn't essential to understanding the visualisation, by using other elements such as icons or labels, or by adding a subtle outline to an element (see above). Personally I think it's good practice to avoid relying too much on colour anyway, since red and green isn't the only colour combination that those with colour-blindness find challenging.

We need best practices to guide us through the numerous decisions that go into making a visualisation. But each data visualisation comes with its own unique design challenges, and these rules — which are often a broad simplification of a topic — may not work in every case. We shouldn't be afraid to break them if it makes our visualisation clearer or more engaging. Clever bending of the rules — as in other areas of design — can result in our most creative work.

That said, there are still some rules that I've yet to find a good reason to break:

- ~~Die charts are always better as a single stacked bar or another chart~~

To make Medium work, we log user data. By using Medium, you agree to our Privacy Policy, including cookie policy.

.e

- Rainbow sequential colour scales: as well as being inaccessible, the abrupt changes in luminance create “false boundaries”
- Blow-apart effects: makes it harder to judge areas and compare elements

If you know a good reason for breaking these rules, or can think of any other cases, I'd love to hear about them.

Rosamund Pearce is a visual data journalist at The Economist. You can follow The Economist Data Team's work on Twitter.

[Data Visualization](#)[Infographics](#)[Information Design](#)[Data](#)[Journalism](#)

E

## Written by Rosamund Pearce

481 Followers · Writer for The Economist

Visual Data Journalist at The Economist

[Follow](#)