Visualization
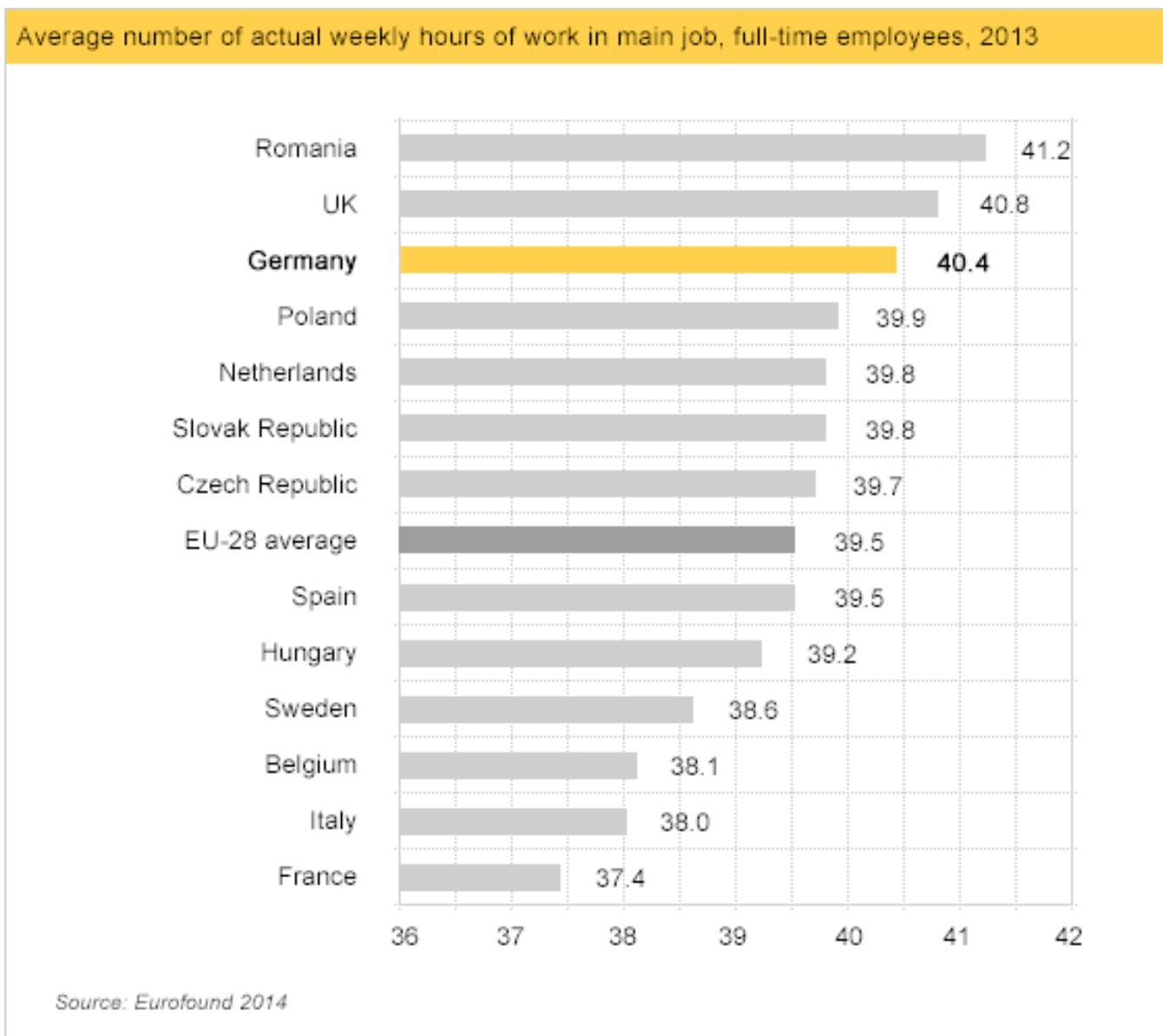
Misleading axes on graphs

[This website was] developed by Carl Bergstrom and Jevin West to meet what [they] see as a major need in higher education nationwide.

The purpose of a graph is to tell a story. However, subtle choices about how to represent a dataset graphically can have a substantial influence on the story that a visualization tells. Good visualization can bring out important aspects of data, but visualization can also be used to conceal or mislead. In this discussion, we'll look at some of the subtleties surrounding the seemingly straightforward issue of how to choose the range and scale for the axes of a graph.

Bar chart axes should include zero

We begin with a well-known issue: drawing bar charts with a measurement axis that does not go to zero. The bar chart (below) was created by the German economic development agency GTAI, and comes from a webpage about the German labor market. In the accompanying text, the agency boasts that German workers are more motivated and work more hours than do workers in other EU nations.

Average number of actual weekly hours of work in main job, full-time employees, 2013

Country	Hours
Romania	41.2
UK	40.8
Germany	**40.4**
Poland	39.9
Netherlands	39.8
Slovak Republic	39.8
Czech Republic	39.7
EU-28 average	39.5
Spain	39.5
Hungary	39.2
Sweden	38.6
Belgium	38.1
Italy	38.0
France	37.4

Source: Eurofound 2014

NOTE: This document has been modified.

It looks like Germany has a big edge over other nations such as Sweden, let alone France, right? No. The size of this gap is an illusion. The graph is misleading because the horizontal axis representing working hours does not go to zero, but rather cuts off at 36. Below, we've redrawn the graph with an axis going all the way to zero. Now the differences between countries seem negligible.
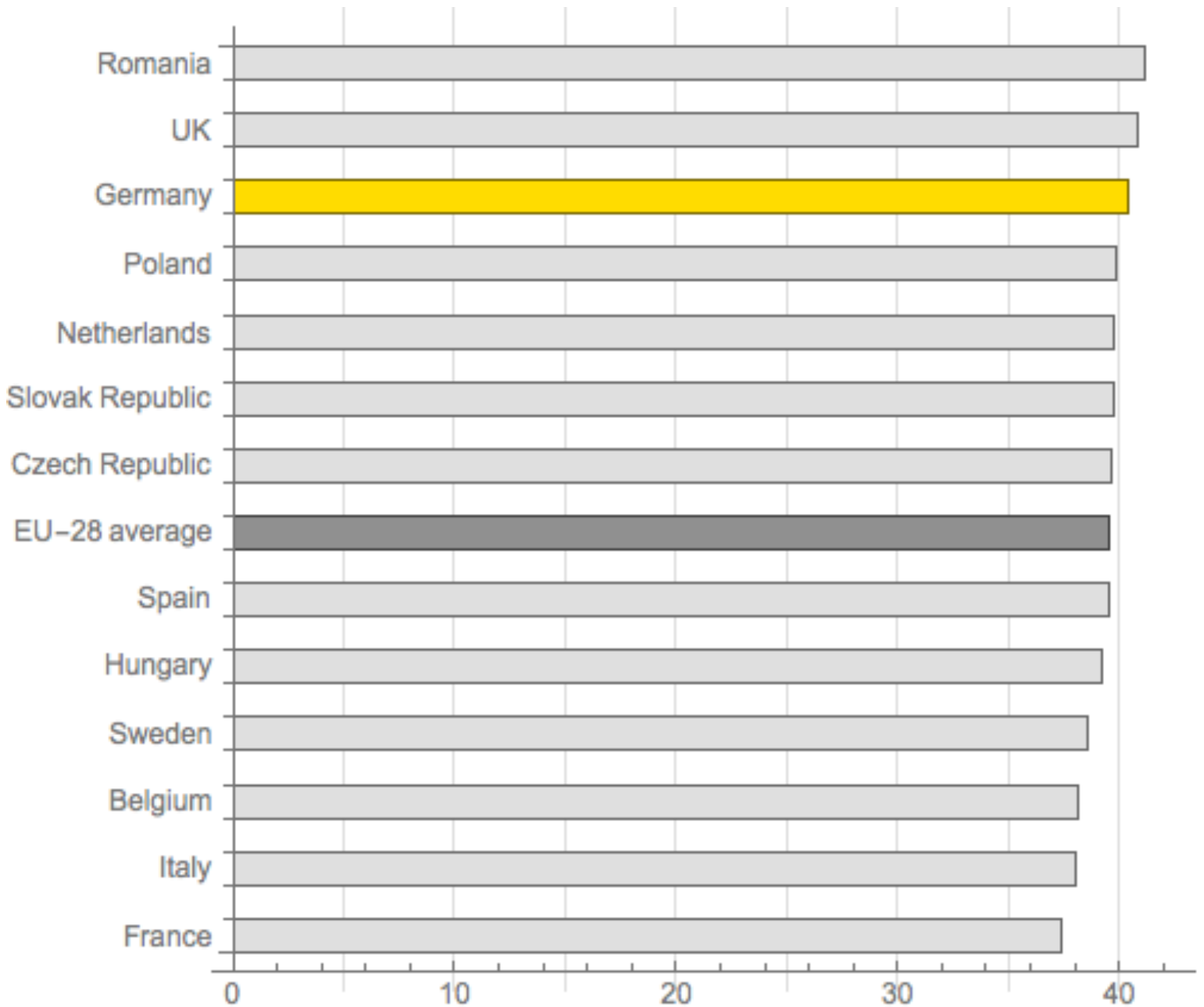

(You might notice that in the redrawn graph we've removed the horizontal gridlines separating the countries. These were not particularly misleading, but they add visual clutter without serving any purpose whatsoever.)

Line graph axes need not include zero

While the bars in a bar chart should (almost) always extend to zero, a line graph does not need to include zero on the dependent variable axis. For example, we consider the line graph below from the [California Budget and Policy Center](#) to be perfectly fine, despite the fact that the y-axis does not 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
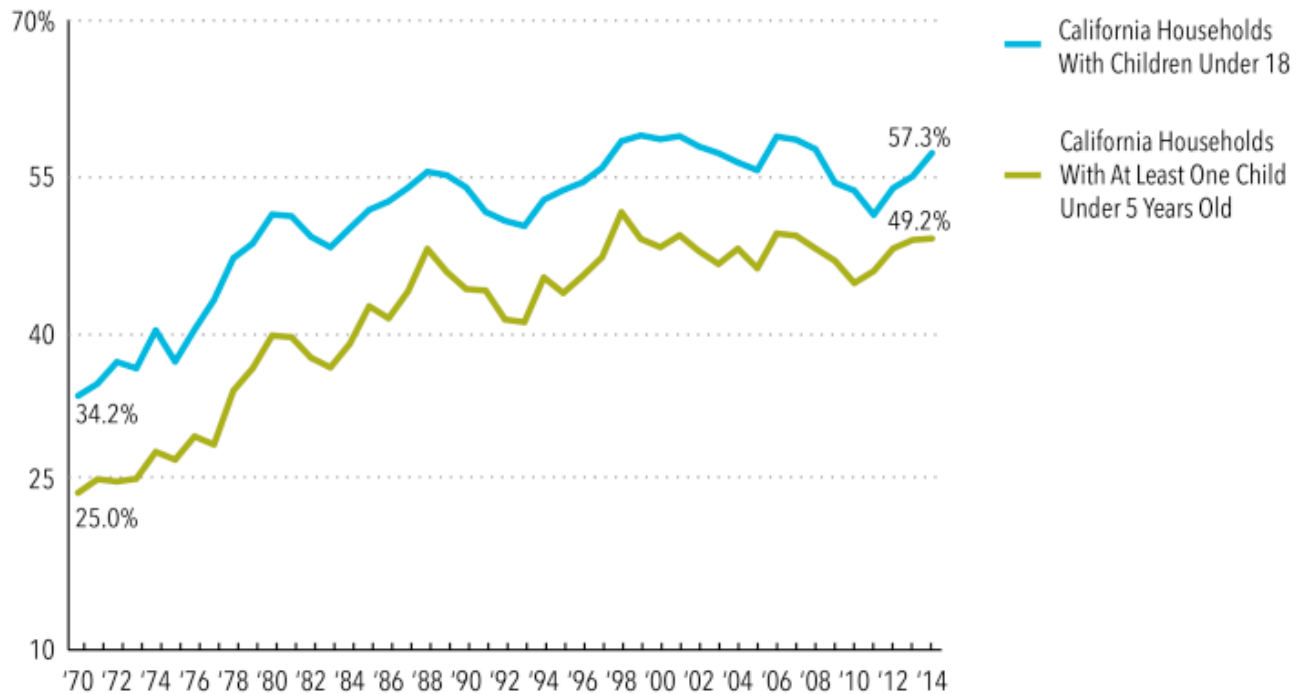
Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.
Source: Budget Center analysis of US Census Bureau data

California Budget & Policy Center
Independent Analysis. Shared Prosperity.

What is the difference? Why does a bar graph need to include 0 on the dependent axis whereas a line graph need not do so? Our view is that the two types of graphs are telling different stories. By its design bar graph emphasizes the absolute magnitude of values associated with each category, whereas a line graph emphasizes the change in the dependent variable (usually the y value) as the independent variable (usually the x value) changes.

For a bar graph to provide a representative impression of the values being plotted, the visual weight of each bar — the amount of ink on the page, if you will — must be proportional to the value of that bar. Setting the axis above zero interferes with this. For example, if we create a bar graph with values 15 and 20 but set the axis at 10, the bar corresponding to 20 has twice the visual weight of the bar corresponding to 15, despite the value 20 being only 4/3 of the value 15.

A line graph doesn't draw the attention to the absolute magnitudes of the values, because there is little visual density — i.e., ink — below the curve being plotted. (The exception is line graph in which the area under the curve is filled; we believe these line graphs need to have a zero axis in the vast majority of cases.) As a

result, the line graph is freed from the constraint of including 0 as the axis, and thus can zoom into the relevant range to better reveal changes in the dependent variable as the independent variable changes. Thus while people may gripe about line graphs that don't include zero on the dependent axis, we are unconcerned by this display decision.
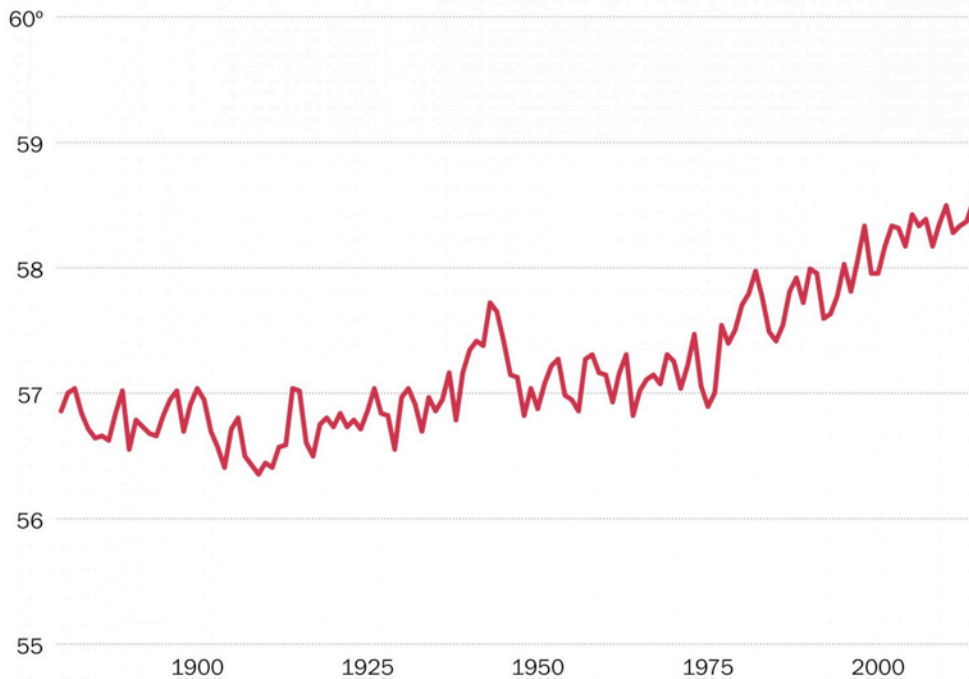
When line graphs ought not include zero

Indeed, line graphs can obscure important patterns if their axes do go to zero. One example is below.
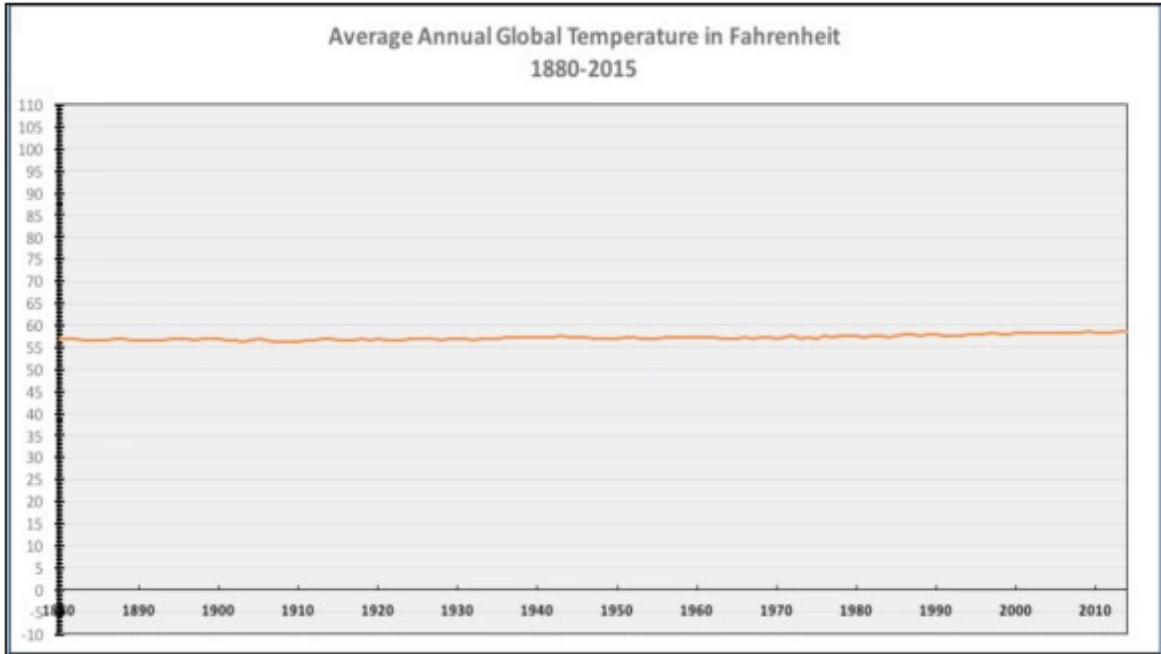
Philip Bump discusses this graph in a Washington Post article. He points out that for the purpose of considering climate change, the proper representation of these data would look something like the following:
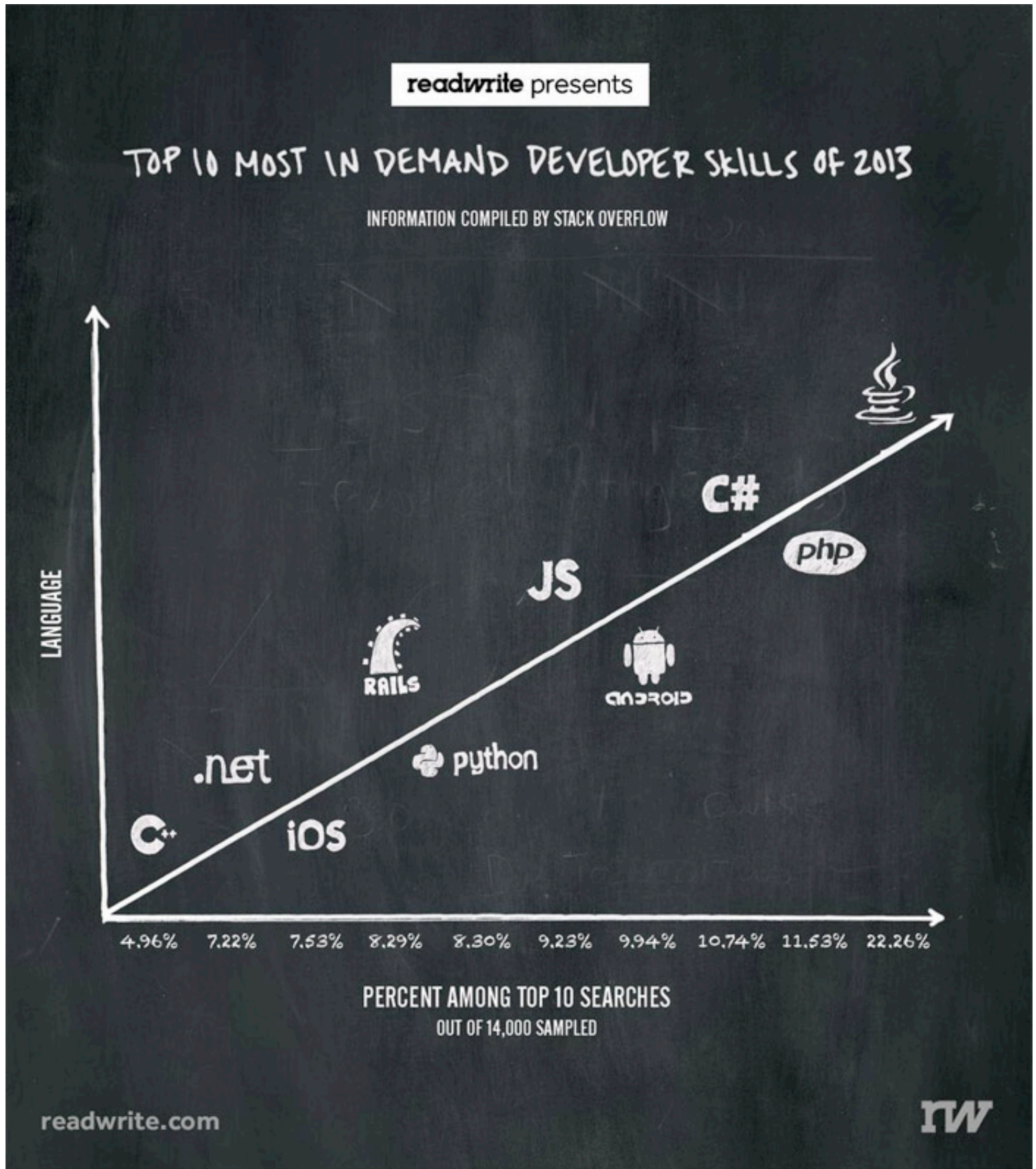
An axis should have something on it

We would have thought this obvious, but a line graph should have something numerical on each axis. The graph below does not. Its vertical axis is labeled "Language" but even in the most generous interpretation this is a categorical variable and thus not appropriate for display using a line graph.

This graph isn't so much misleading as it is just plain perplexing. Max Woolf does a nice job of [discussing its design flaws](#) and suggesting appropriate alternative ways to present the same information.

Conclusion

In summary, data visualizations tell stories. Relatively subtle choices, such as the range of the axes in a bar chart or line graph, can have a big impact on the story that a figure tells. When you look at data graphics, you want to ask yourself whether the graph has been designed to tell a story that accurately reflects the underlying data, or whether it has been designed to tell a story more closely aligned with what the designer would like you to believe.