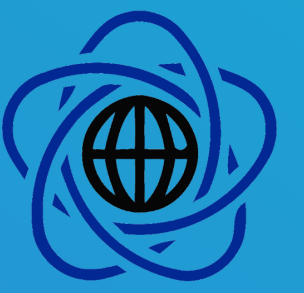


Choosing the chart's axes

good and bad practices

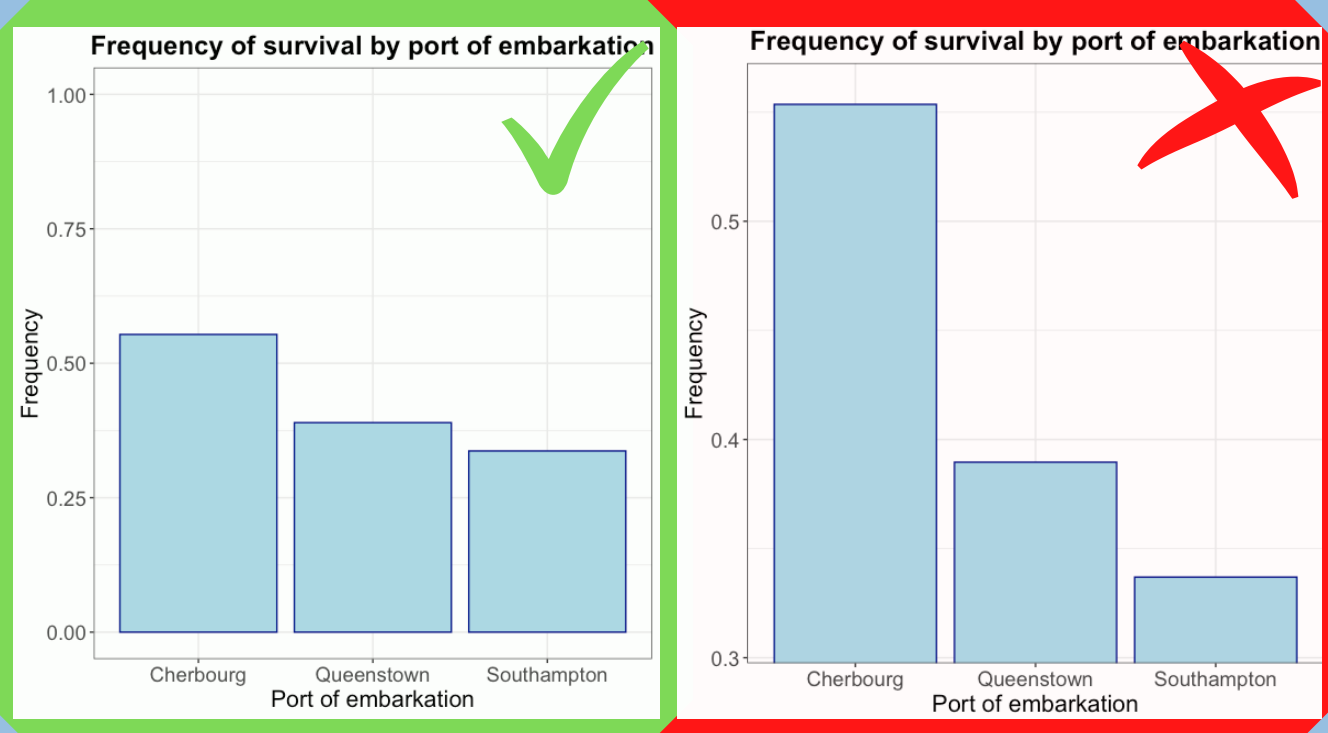


Patryk Tomaszewski
Urszula Białończyk
Mateusz Ziemiała

About the used dataset

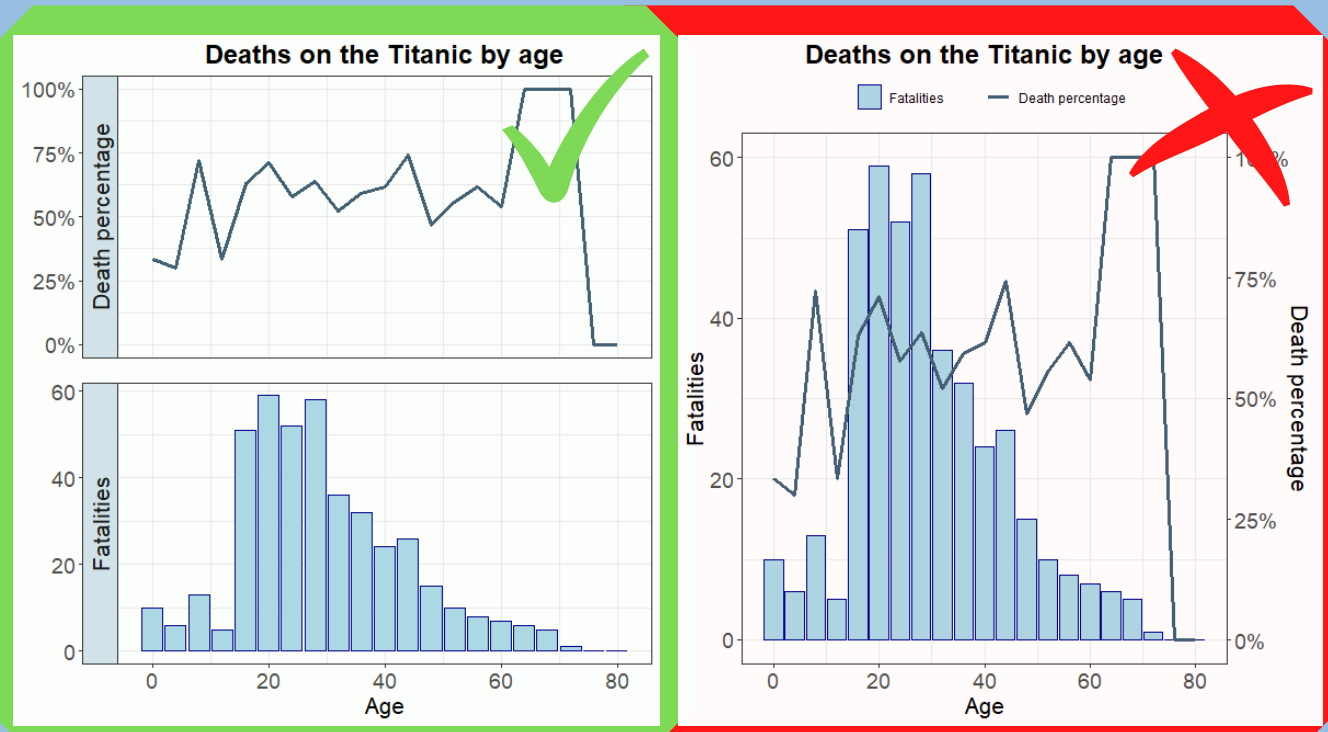
The Titanic dataset is a very popular dataset used mostly for classification problems. It consists of a various information on the passengers of Titanic, such as age, sex, class, or an indicator whether a given passenger survived the disaster or not.

Starting the Y-axis at zero



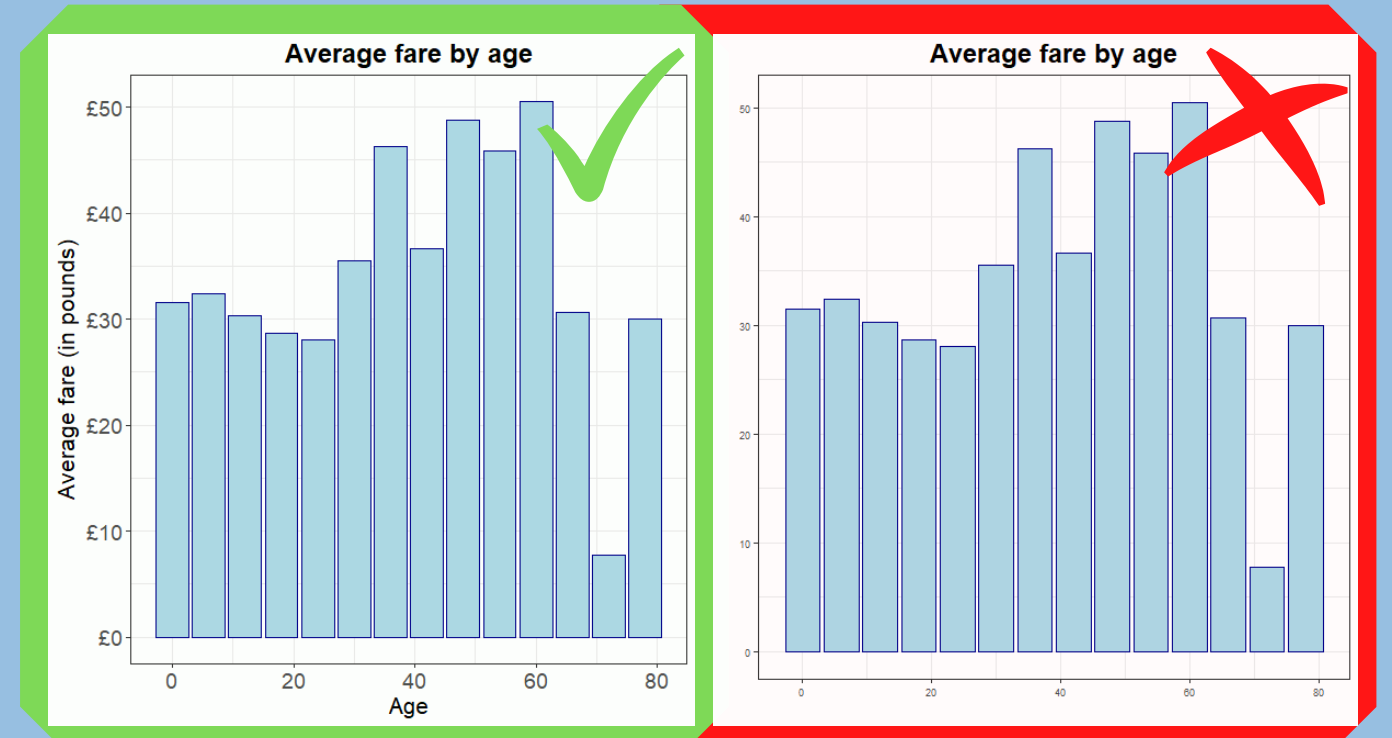
Starting the Y-axis at a point other than zero usually causes a distortion of proportions. This may lead to mistakes in the plot's interpretability: there seems to be a greater difference between the values that are actually quite similar to each other.

Avoiding the use of dual axes



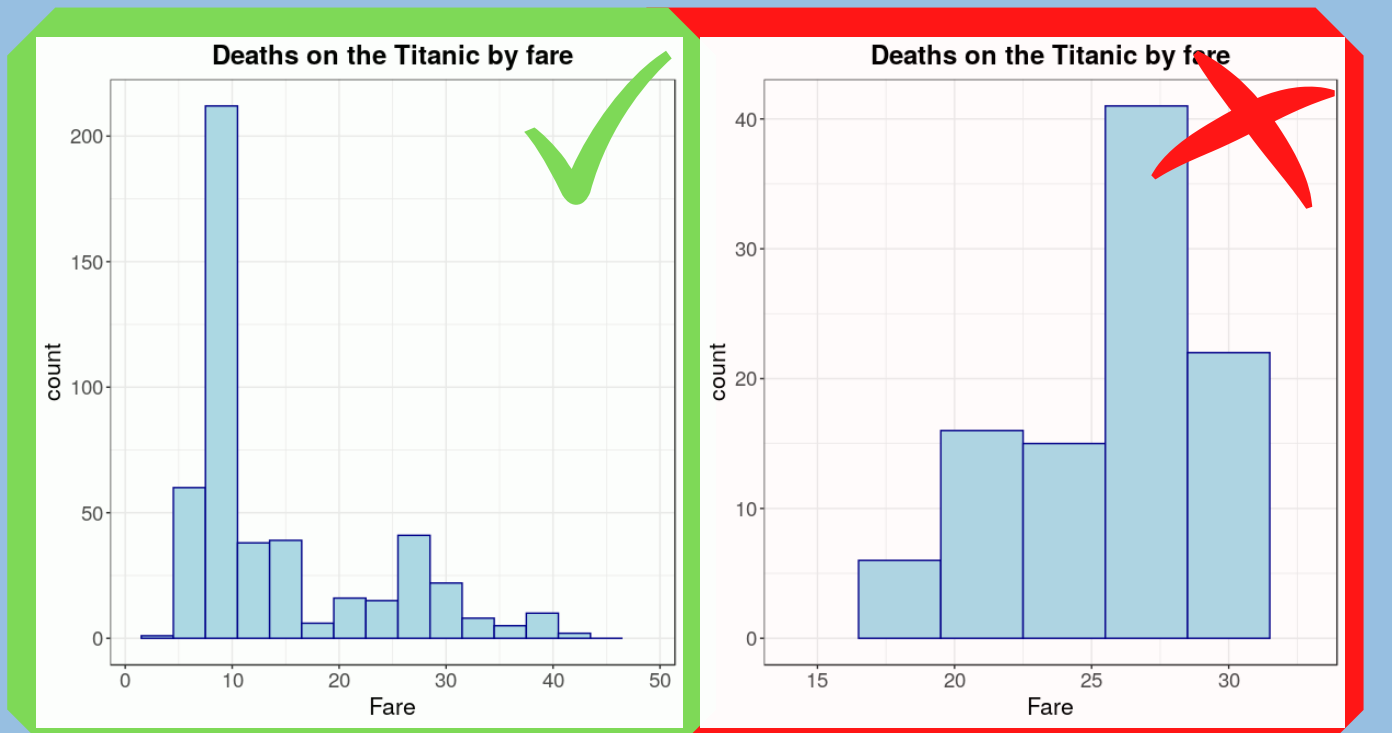
Overlapping multiple different axes on a single chart makes it more complex and forces the recipient to spend more time reading it in order to understand the displayed data. Therefore, in most cases it is better to display two different charts side by side instead.

Ensuring the legibility of the axes



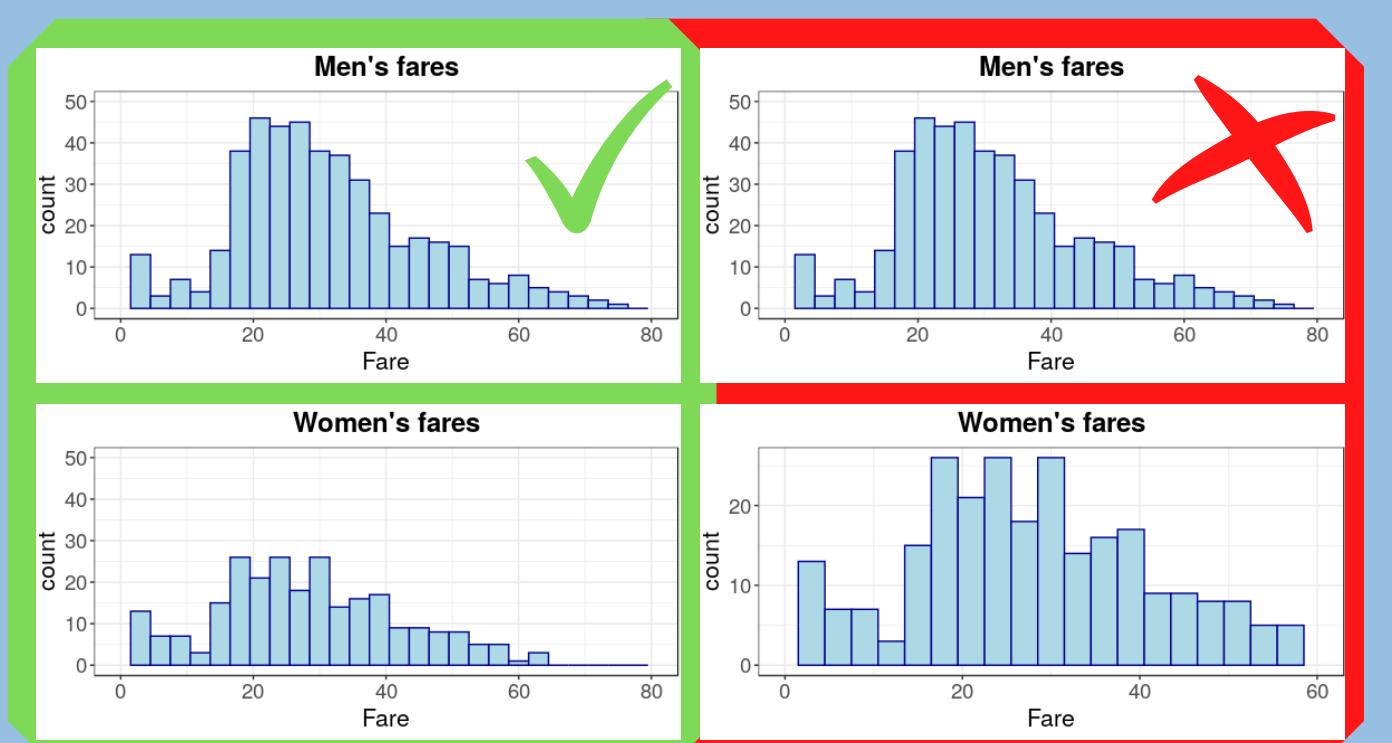
Legibility of the axes is crucial in charts. All axes should have a concise description, preferably with units. Labels need to have proper size and should be formatted to be easily readable, for example by the use of separators in large numbers. Any non-linear axis should be visibly marked as such.

Not zooming in on favourable data



Not showing the entire or at least the trimmed dataset can maliciously suggest trends antithetical to the ones actually present in the data. Any zoom that leaves data out should be clearly marked as such.

Using the same scale with comparisons



With direct comparisons, using a different scale on any axis makes it difficult to easily compare data, since for any comparison the reader would be forced to read the values off the axes instead of easily comparing the different shapes.