

# Plot what you mean!

Plots/charts a great way how to deliver a message. Yet you should reveal, not conceal – the appropriate type of chart is needed. A simple design usually wins. BTW if you struggle with simple variants of charts (scatter plot), it makes little sense to try to compose something complex – most likely it will not work and your reader will be confused.

It is a sign of intellectual laziness to just copy and paste charts and tables from books and papers into your master theses. You should do them! Common guys... its a master level

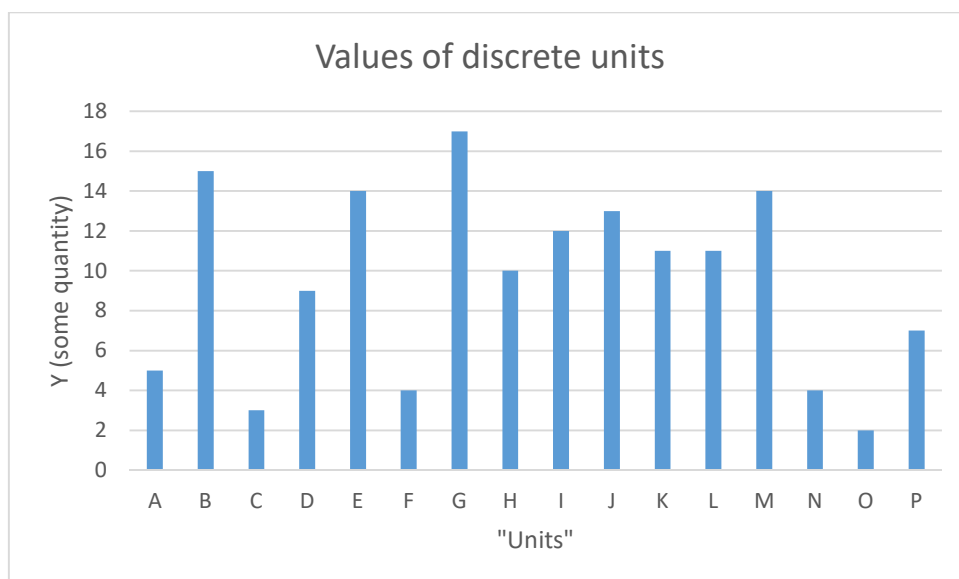
When you plot a chart, think twice if your reader can understand (ideally in a stand-alone format).

Don't forget to provide name, axis labels/descriptors, or legend (if needed). In the majority of situations, a source is a must! I mean the source of your data.

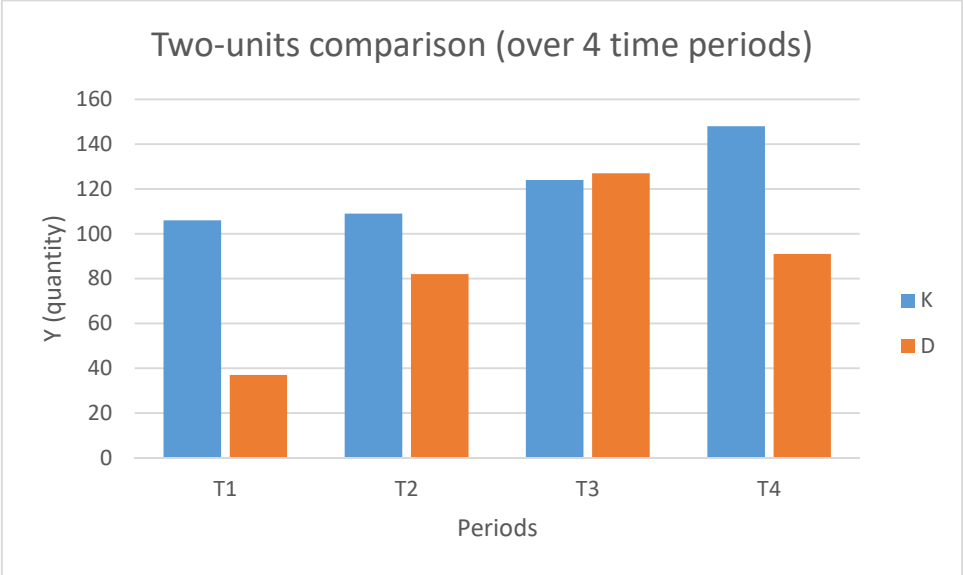
*Below we focus on a few basic types of charts – barplots/columncharts, line-charts, scatterplots, and plots of distribution. Make sure you can do these elementary types of charts. Try to emulate best practices (scientific papers, the Economist, 538, etc.).*

## BARS and COLUMNS

You have n discrete units (states, cities, regions, political parties) and you want to show how they fare concerning one variable (quantity). A simple **bar chart** will do the work! Sometimes, it makes sense to order your data first to obtain a decreasing or increasing order of units (it's better for comparison). However, if you give up on alphabetical order, it might be useful to highlight units of interest (if there are any) – like your city, state, etc. Do not plot trendline over this!

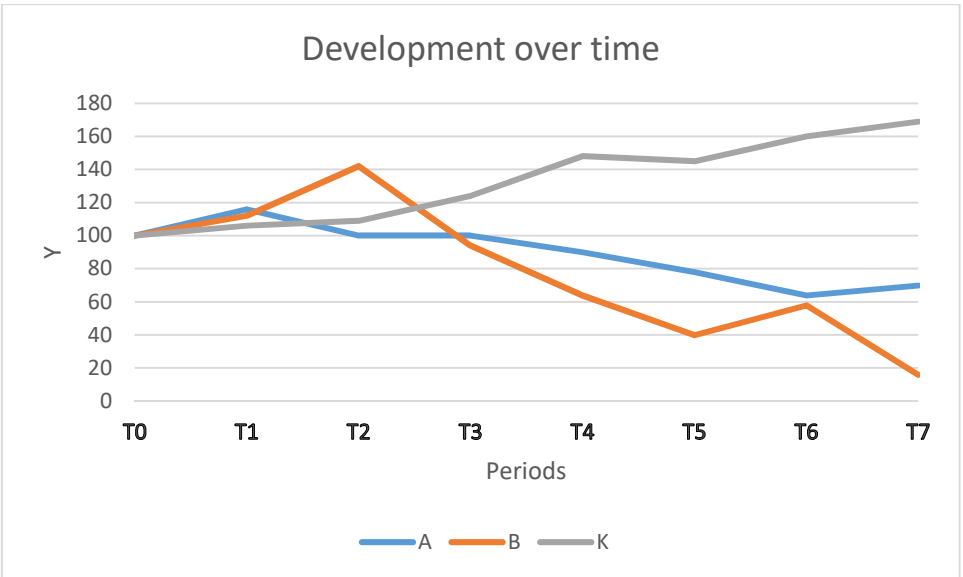


This chart can serve quite well even for limited-time series (if the number of periods is bigger than 12, consider the line chart). Might serve well even for comparison of a few units over time (**a very small number of units and periods is a key here**). Bar chart for longer time-series makes also sense when you want to show the total number of e.g. events/casualties per weak/month over a given time-span. It makes sense, sometimes, to add numerical values to the columns... btw trendline could be used here (yet think twice what will be its interpretation).



**LINES**

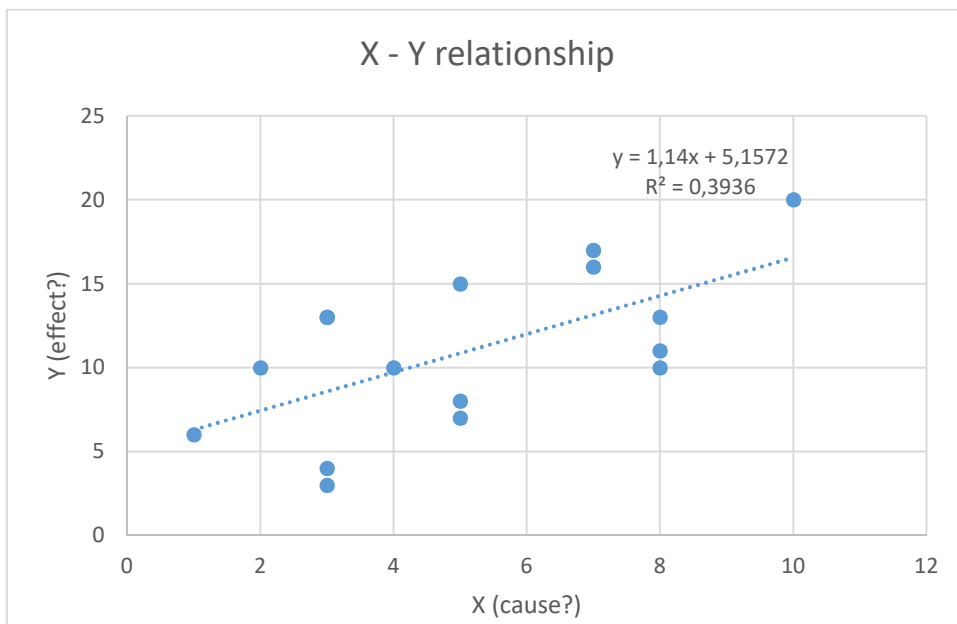
When the development of some variable over multiple periods is what we want to show, then chose line-chart! The bigger the number of the periods the better it is to use the line. Think twice about how many lines can be meaningfully plotted into one chart (usually six is the maximum). Think strategically about colors of lines (ideally they should evoke the units, but this might be hard to achieve).



Under specific circumstances, it might be useful to combine a bar and line chart. Lines and charts will capture two different quantities evolving over time. In this case, you will likely need the second Y (vertical) axis.

## DOTS

When we want to investigate the relationship between two (especially ratio/interval/ordinal) variables the best choice is old good **scatterplot**! Why is the scatter plot so great? It enables you to see immediately the X and Y values of any point (e.g. try to find point X=4, Y=10 in the chart below). At first sight, you can make quite a good guess if there is some relationship/correlation. It also helps you to identify outliers or deviant cases. It makes sense to plot over the scatter „trend“ (here de-facto a regression line), to check if your visual guess of a correlation was correct or not.



## BOXPLOT

Often, we want to compare the distribution of values (e.g. before and after some experimental manipulation). Boxplots offer a great tool for this comparison (and also they can tell us a lot about the distribution of a single variable – maybe a transformation of a variable is needed).

