DATA VISUALIZATION USING GGPLOT2: 5 features worth knowing

Xyz min ggplot2 tips

Whether you are working on your school project, term papers, master thesis or a business presentation, the quality of the data translates well to the audience if the data is presented in a meaningful and aesthetically pleasing way. Although making an impactful figure from the data is challenging and time consuming, but some small tips if implemented properly can make the workflow simpler and save a lot of time over a long run.

In this post, I will be discussing 5 tips or features that I find useful and I implement on a regular basis while working with ggplot2 package. May be some of you will already know them already and for some it will be something new.

To give a little background about ggplot2, it is developed based on the grammar of graphic in short *gg*. The plot consists of three fundamental parts, namely:

1. Data: consisting of data frame
2. Aesthetics: defining the x and y variable and detailing the color, size, and shape of the data points
3. Geometry: defines the type of plot e.g. bar plot, line plot, scatter, histogram etc.

So, lets dive in to see the 5 features/tips that I want to share with you all that have been very handy to me.

Tip #1: Using parentheses while assigning ggplot function to a variable

A cool thing in ggplot2 package is the use of parentheses. In general, while plotting using ggplot function, if we assign it to a variable then in order to plot, the variable is executed in the next line. So, this is a two-step process. By putting the ggplot function in the parentheses when assigning helps plot the figure directly without executing the variable again.

Tip #2: Minimizing repetitive ggplot function by switching data

This command simply is very useful during the EDA stage of data when I try plotting different sets of variables to understand the relationships. Instead of writing ggplot commands again and again, I prefer to use %+% from the tidyverse package. By using %+% we can override the current data with the new data.

Warning: For %+% to work, make sure the column names of the new data matches the column name of the dataset being replaced.

Tip #3: Freedom to use col/color/colour in aesthetics

The aesthetics of the ggplot2 package gives the freedom to choose between col/color/colour to assign the colours to the plot. So, it doesn’t matter if you use British English or American English, you are covered by ggplot.

Tip #4: Reordering data using fct\_reorder() function

To make aesthetically pleasing plots, generally, most of us use group\_by(), summarize(), and order() functions to arrange the data. This being a lengthy step which could be replaced by a single function from the forcats package called fct\_order().

Tip #5: ggsave() for saving plots

ggsave() function gives flexibility to users looking to save high resolution for publications, presentations or for printing. The function gives the flexibility to fix the image size, define the resolution, if the image be saved as transparent or with background, and flexibility to choose the file extension.

These are the 5 features that I use quite often when plotting with ggplot2 package. Did you guys find these interesting? Were any of them new to you? Do share your thoughts in the comments and any tips/tricks that you use for plotting not covered in here.

You can connect with me on LinkedIn.