DACSS 601 / PS 601 - Assignment Three

Mustafa Erdem Kusdil

August 13, 2020

Visualization

The visualization demonstrates the comparison between two variables with a colored classification variable added. The variable on the x-axis represents the tendency to stay at home during the Covid-19 pandemic on 0 to 100 scale. The variable on the y-axis represents the amount of fines that should be enforced for the individuals who go out while demonstrating Covid-19 symptoms on a 0 to 500 British pounds scale. The colored classification is the health condition of the respondents ranging from "Poor" to "Excellent".

When we interpret the plot, we see that the two variables have a positive correlation between them. The density of the scatters in the upper right side tells us that most of the people with a tendency to stay at home during the Covid-19 pandemic are also supportive of the possibility to enforce a higher fine for going out with symptoms. The plot also shows us that the health condition of the individuals do not make a significant difference.

Before starting the assignment, I considered other visualization options, for instance, the bar charts with the options of 'geom_col' and 'geom_bar'. However, there were a lot of overlapping values that could make it very difficult to understand. That is why I decided to utilize the smooth option since it shows the trend and comparison much clearer. Also, it would be an option to include the 'facet_grid' option to split the plot into different facets that represent the 'health' variable. However, after reviewing the plot with the 'color' option, I concluded that it is relatively easier to understand the plot as it is. I believe the coloring of the scatters made the plot more compact and interesting.

For future work, I would like to include other variables such as the questions of anxiety. I was unable to include this set of questions because these variables required recoding as they were initially coded discretely. I would also like to learn more about the data and treat some groups of questions as single scale measures and add them to the comparison as one variable.

Note: The only problem I encountered while creating the plot was the orange-colored line under the health legend. I could not find a solution to delete or hide it.

Scatterplot
Stay at Home or Pay the Fine