Homework3\_KBreest

Kazmiera Breest

8/12/2020

# COVID DATA

## Describing the Data

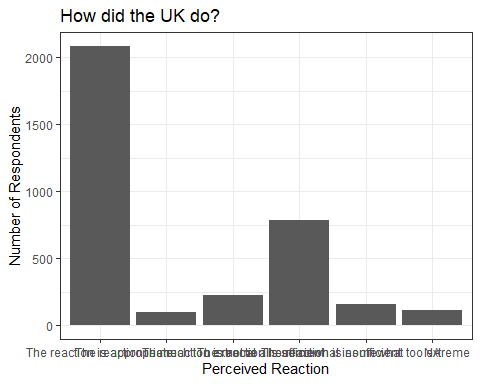
This dataset contains survey responses from a questionaire aimed at gathering insight on how folks in the UK experience the crisis caused by COVID-19. More specifically, Italian citizens took this optional survey which explores their self reported behaviors during the pandemic. The survey collected responses to 62 different questions, which resulted in a data set that has 80 variables, or columns, and 3,460 individual observations, or rows, with each row indicating a unique survey response. Outside of the 62 questions, the remaining 18 rows are made up of identifying information such as the start and end date of the survey response, contacts information, language, and names.

## Data Visualization

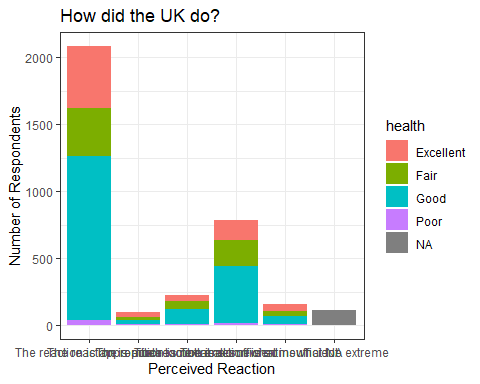
In order to create a useful data visualization, I am first going to check through the dataset to find a variable that I would like to highlight.

I realized that the headings of the columns in my data aren’t all easily interpreted. I looked to the codebook to understand what columns correspond to which questions in the survey, and luckily the headings in the data are displayed with the text in the survey. I decided to take a look at how the survey respondents percieved their government’s reaction, since the survey was taken by Italians who’s government had a very different course of action than the US did.

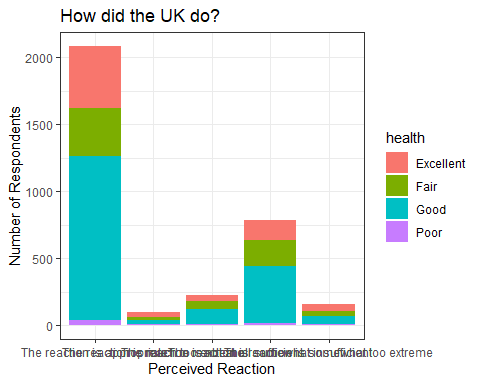
This is question 23 on the survey, presented as: “Do you think the reaction of the UK government to the current coronavirus outbreak is appropriate, too extreme, or not sufficient?” with the options for response being:  
(1) The reaction is much too extreme  
(2) The reaction is somewhat too extreme  
(3) The reaction is appropriate  
(4) The reaction is somewhat sufficient  
(5) The reaction is not at all sufficient



I think this graph is interesting, but it would be even more interesting to see how those of different health statuses fall into these response categories.



I also decided to remove the N/A categories to create a clearer picture



1. This visualization demonstrates the number of respondents who perceive the UK government’s reaction to the COVID-19 crisis in ranked categories by their reported health status. By cutting the percieved reaction data by health status, one might be able to identify correlation between an individuals response to how the government ‘should’ have reacted given their health status. What the bar graph shows is that the majority of respondents believe that the UK government’s reacton was appropriate, and also that the marjority of those who percieved the reaction as such indicated being in good health.
2. I chose the bar graph since there are only 5 categories of responses for the percieved reaction question. Since I also cut the data by health, which only has 4 categories, displaying these categories as colors was also easy to interpret. Had there been more categories of either variable, I would’ve chosen a scatterplot.
3. I would like to learn how to make the labels of the variables easier to read, as well as add data labels to the different bars to create a clearer depiction of the differences in responses from category to category. I think that being able to add the number of total valid responses as a note on this graph would also be useful.