Homework3\_KBreest

Kazmiera Breest

8/12/2020

# COVID DATA

## Describing the Data

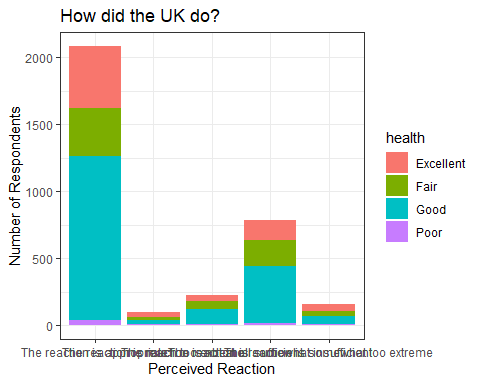
This dataset contains survey responses from a questionaire aimed at gathering insight on how folks in the UK experience the crisis caused by COVID-19. More specifically, Italian citizens took this optional survey which explores their self reported behaviors during the pandemic. The survey collected responses 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



## RELATIONSHIPS BETWEEN VARIABLES

*1. Descriptions of the variables, including the coding of the variables, summary statistics, and, if appropriate, visualizations.*

After looking into the variable percieved reaction, i thought it might be interesting to see how a person’s reaction is correlated with how effective they think social distancing measures are. Percieved effectiveness is measured in the survey by asking:

“What do you think: How effective are social distancing measures (e.g., through a general curfew) to slow down the spread of the coronavirus?” with the following options:

(1)Not at all effective  
(2)Not effective  
(3)Neither effective nor ineffective  
(4)Effective  
(5)Very effective

To get summary statistics on these categorical variables, I started by making a proportional table (converted then into a percentage) of the responses for each category. I stored this table as a tibble, and then summariezed each one.

## effective n   
## Length:5 Min. : 2.388   
## Class :character 1st Qu.: 6.567   
## Mode :character Median :13.164   
## Mean :20.000   
## 3rd Qu.:33.254   
## Max. :44.627

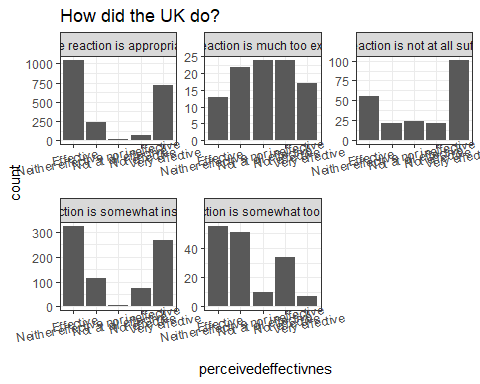
## reaction n   
## Length:5 Min. : 2.985   
## Class :character 1st Qu.: 4.687   
## Mode :character Median : 6.687   
## Mean :20.000   
## 3rd Qu.:23.403   
## Max. :62.239

*2. Description of the relationship between the variables, including a hypothesis (or hypotheses) about the relationship.*

Both of these variables measure the respondent’s perception of aspects of COVID responses by their government. Percieved reaction measures the level of appropriateness each respondent believes their government reacted with, while percieved effectiveness measures how effective respeondents feel that the measures that their government put into place are. A person who thinks their government responded appropriately will likely believe that the imposed social distancing guidelines are effective to some extent. It is important to note that the government’s reaction is not restricted to social distancing, and likely includes other actions such as business shut-downs, financial relief, or even press coverage. That being said, I hypothesize that if a respondent believes their government reacted appropriately, they will also believe that the social distancing measures are effective.

*3. Initial demonstration of the relationship, which could include correlation, visualization, or model.*

Initally, I used a variation of my graph from HW3 to check the relationship between the two variables. The Facet\_Wrap function was useful to make a clear chart which shows the breakdown between categories. From this visual, it is clear that out of those who felt the reaction was appropriate, they also felt social distancing was effective over any other effectiveness measure. However, almost every other category had a high percentage of those who felt the measures were effective as well.



Next, I attempted to fit a model and find the correlation between the two variables, and recoded them numerically in attemps to solve the errors i was getting, but I am not sure what I am doing wrong here. I think my issue is that these are categorical variables and are not continuous, but I am not sure how I should manipulate these data in order to fit a model. My work is below for reference!! (I’ll make an appointment next week for office hours)

#recode(effective, `Not at all effective` = 1, `Not effective` = 2, `Neither effective nor ineffective`= 3, `Effective` = 4, `Very effective` = 5)  
  
#recode(reaction, `The reaction is much too extreme` = 1, `The reaction is somewhat too extreme` = 2, `The reaction is appropriate`= 3, `The reaction is somewhat sufficient` = 4, `The reaction is not at all sufficient` = 5)  
  
#\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
  
#cor(effective, reaction)  
  
#fit <- lm(effective, reaction, data = covidData)  
  
#cor(reaction, effective)  
  
#fit <- lm(reaction,effective, data = covidData)