Individual Project DS Summer

Smbat Gardilyan

The problem/data description

- Why is this important/interesting

  Nowadays, since the world is fighting a battle against the global pandemic, it is very crucial to be acknowledged and have a full understanding of COVID19 and in which patterns it impacts our lives. Therefore, having a reasonable dataset that provides proper statistics of various attributes of the virus, will hand us the opportunity to do some alluring discoveries that can help us overcome the situation in an easier way.

- The problem statement

Does the application of the masks prevent the spread of coronavirus?

-Where does the data come from?

The data is being taken from Infogears.org, which is a free Crowdsourced COVID19 local data platform that helps local communities track important data, such as Mask usage, Stay home, Symptoms and Mental Health. It is also a leading indicator for COVID19 hotspots.

-What was done on this data so far

There are more than 18,000 unique data points collected. In addition, some analysis has shown the predicted case increase in May 2020 alongside some specific findings such as the fact that marriages are in trouble, etc.

Main hypotheses

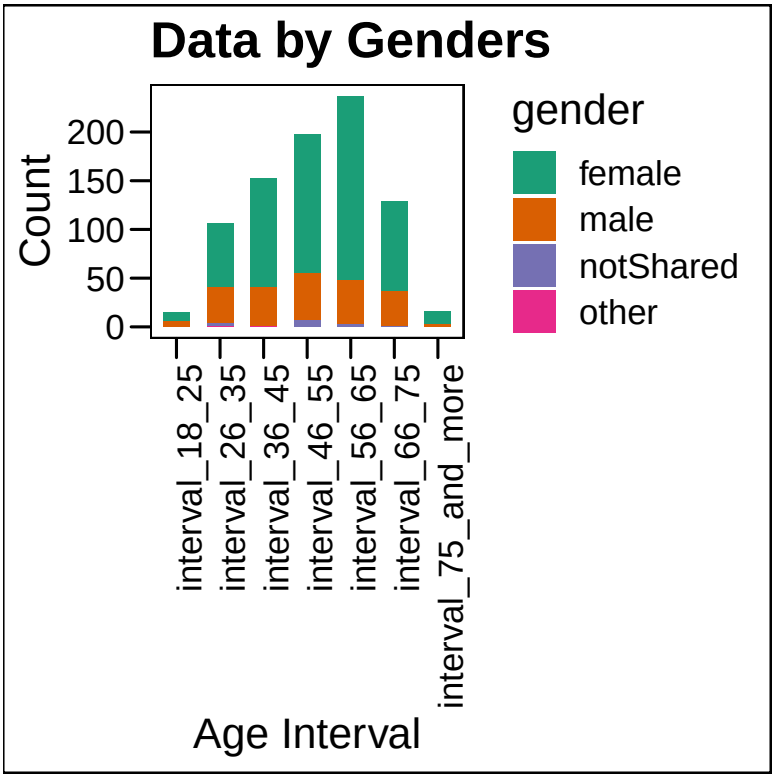
Here you write what are you trying to find in the data, what are some hypotheses that you are trying to test

The findings will present answers to the following questions: -How effective is the usage of the masks? -Are masks really preventing the spread of the virus in case of direct contact or is it just a myth? Since the collected data contains people from both tested and not test categories, separate data observations will be held on splitted tables.

Analysis

We have data of 16768 observations, however only 853 observations have results of virus test. I will make my analysis in that data.

First of all lets have a look at our data.



From this diagram we get basic information such as the fact most people that submitted data to the dataset belong to the age group 46-65 and the dominating majority are females.

Negative Results

```
##     n
## 1 803
```

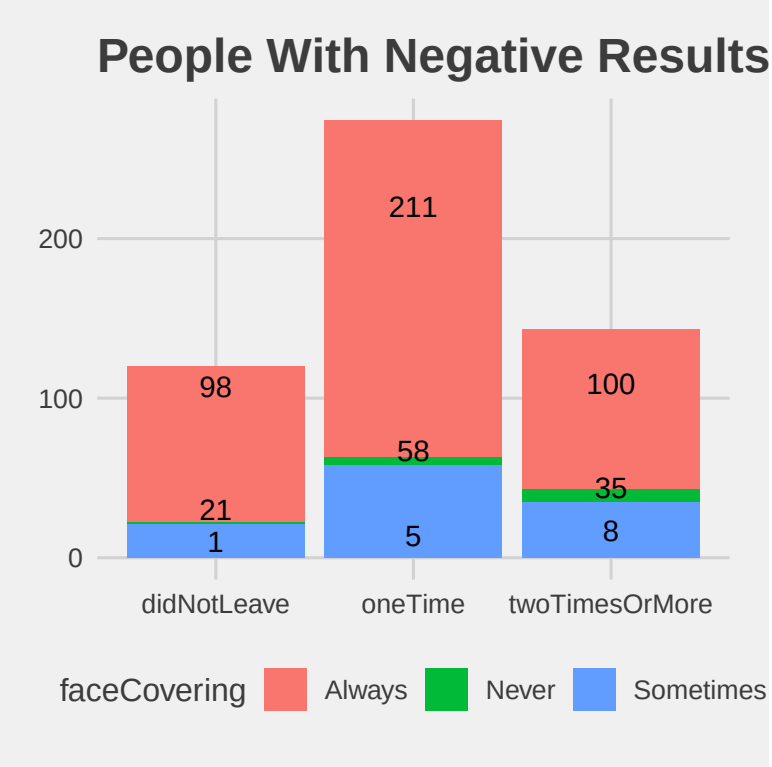
Positive Results

```
##    n
## 1 50
```

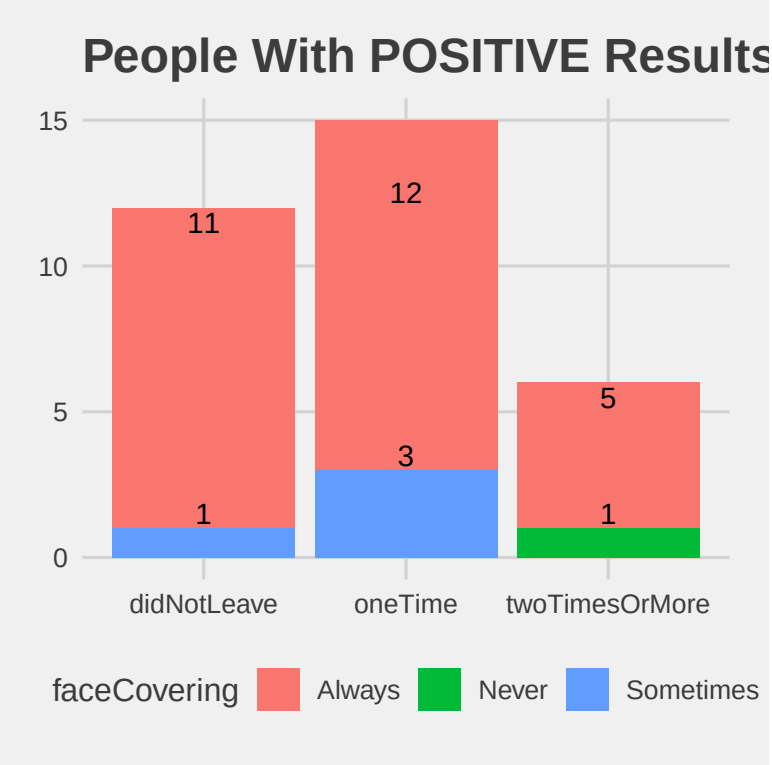
Now we can see that we have 853 observations from which 803 were tested negative while only 50 people had a positive result.

It is essential to acknowledge the patterns of leaving home; how often did people go outside and whether or not the respective actions affect the results of the tests.

For negative results we have

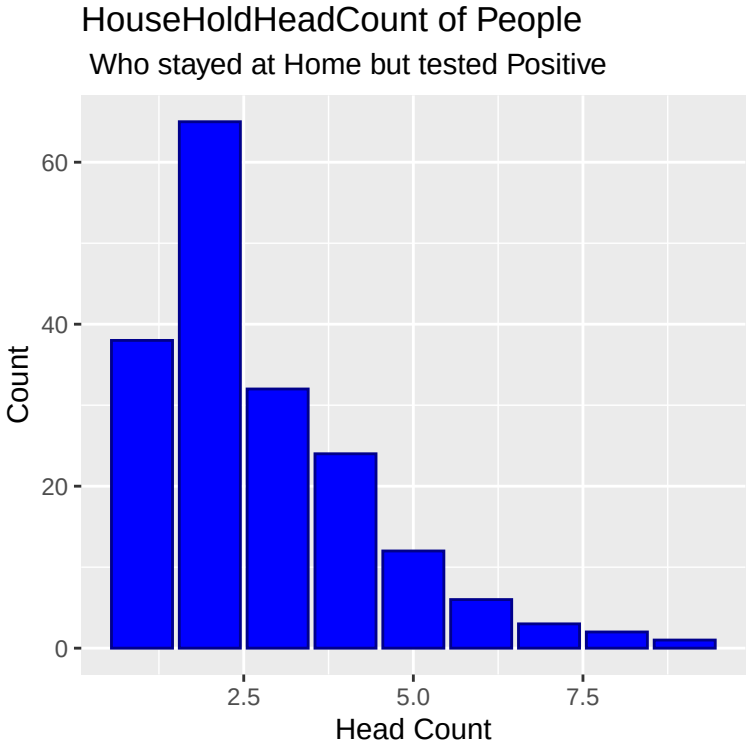


For positive Results



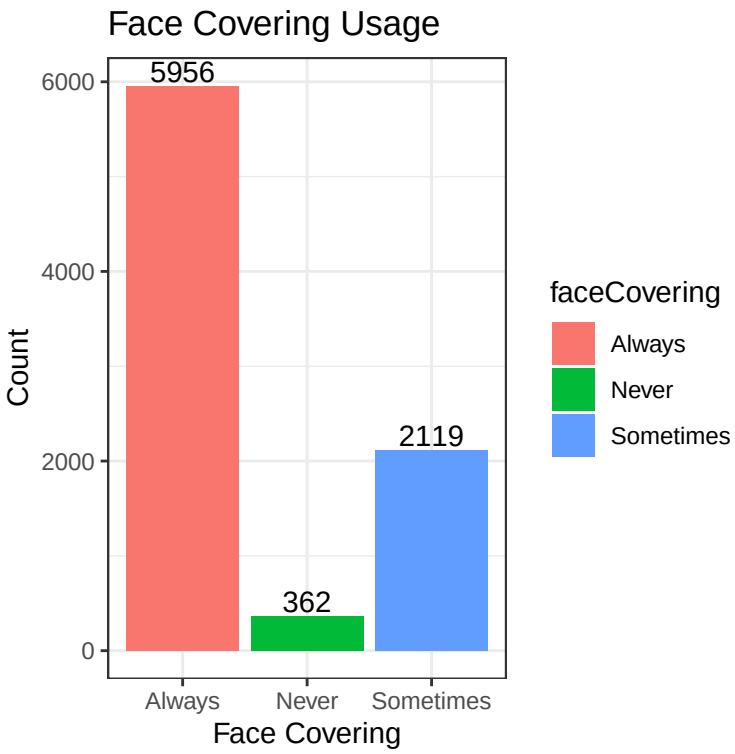
The majority of the people having negative results left home only once, the rest either stayed at home or went outside twice or more. Out of these people the vast majority protected their face by wearing masks. The surprising fact here is that most people who tested positive also used face covering, which supports our hypotheses, meaning that mask usage did not really affect the spread of the virus.

From here we can conclude that if one did not leave the house and still tested positive must have had a direct contact with an affected person who is from their household, which brings us to our next test; how many people live under the same roof.



As we can see from the diagram, in average families, the households consists of more than 2.5 people which supports the idea that people carrying the virus also affected their family members who did not leave home.

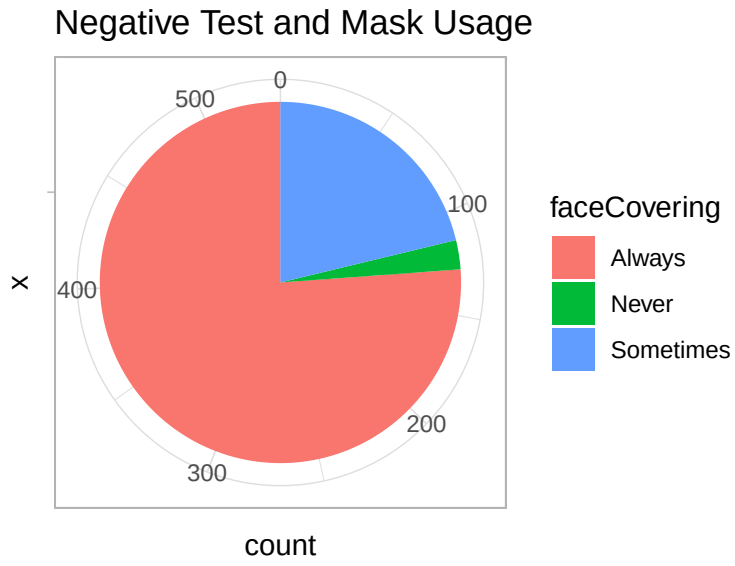
Before testing our main hypotheses, we also need to consider the whole dataset in terms of face coverage usage. Let's see how many people use the face coverage.



We see that 5956 people who submitted their answers to the dataset always wear masks. Following, we have 2119 people wearing it sometimes and only 362 people, who never wear masks.

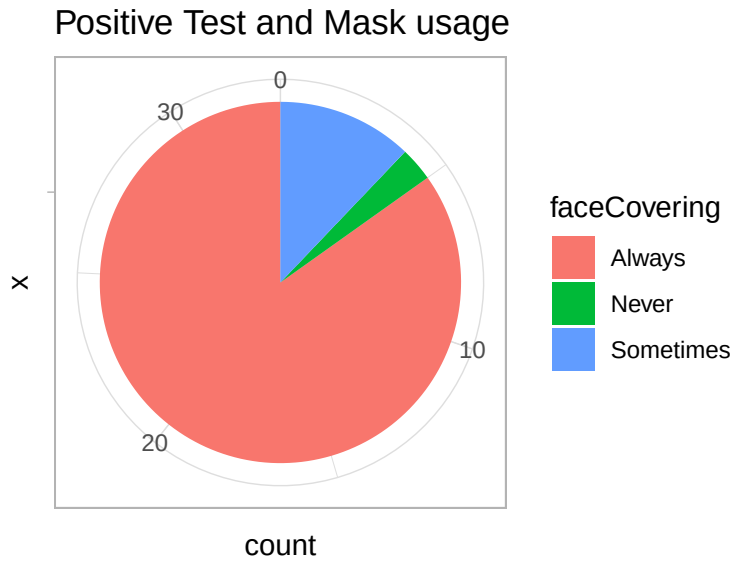
In order to answer to the question whether or not the application of the masks affect the spread of the virus, lets look at our next graph.

Negative ones



This demonstrates that out of 803 people who were tested negative, about ¾ of them were wearing masks, the rest wore them sometimes and only a few people never used face coverage.

Now lets have a look to positive Tests



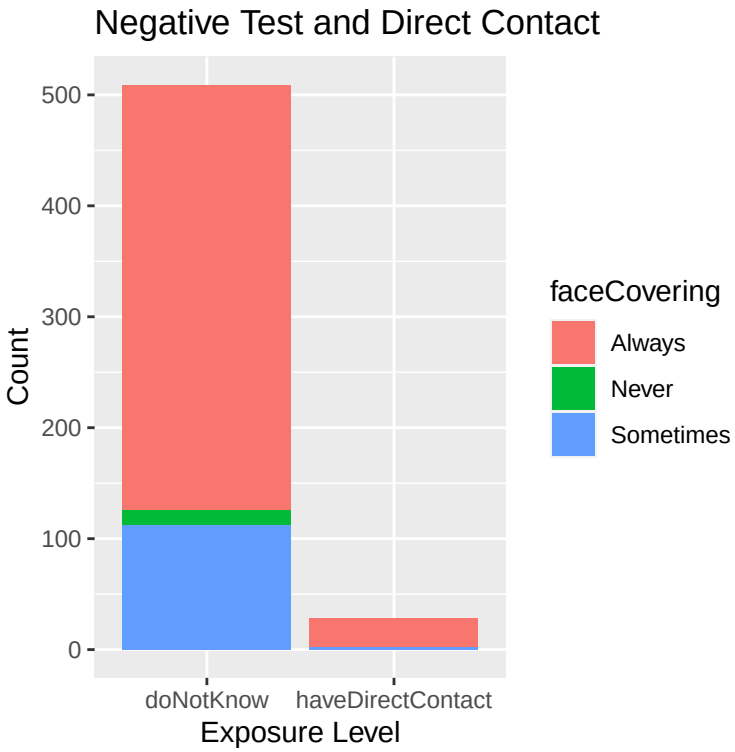
From this graph, we realize that the picture of positive tested people does not really differ from the one of negative tested people, meaning that again almost 75% of the people used masks the whole time.

Now, lets understand the relationship between the mask usage and direct contacts.

Positive


Positive Test and Direct Contact
25
20
15
10
5
0
Count
doNotKnow
haveDirectContact
Exposure Level
faceCovering
Always
Never
Sometimes

Negative



It turns out, that the majority of the people who were affected by the virus and had direct contact with an affected person were wearing masks. Same goes for negative tested people who had direct contact.

Summary of findings and recommendatins if any

To summarize, we can surely state that our hypotheses has been proved. Our findings showed that mask usage has little to no influence on whether or not the virus will be spread. We saw that even people who always wore masks, didn't leave home or had direct contact with an affected person were still tested positive which is an indicator that masks had little to no role in terms of virus prevention. Moreover, it is more crucial to realize that keeping the social distance, staying at home and avoiding direct contact with people who are not from our household is a way better solution rather than face coverage. Even though the provided data was not sufficient enough and there was no balance, we still managed to find the answers to our main hypotheses in these limited conditions.