Individual Project

Knarik Manukyan

The problem/data description

- Why is this important/interesting

The coronovirus caused pandemic that the world didn't experience before. It's fascinating how a virus can have such a huge impact on the humanity, who think of themselves as the most powerful species. It had a huge impact on everyone living on this planet, on the economy of every country, on the social life of everyone and on a lot more areas. A proven vaccine or a medication is not found yet. Doctors and scientists are still doing experiments about this new virus. No one knows the exact impact it will have on our future, on the health of people who have this disease. No one knows if this will be a seasonal desease as the imunity after this virus is under a question. It's interesting and important to analyze data about this pandemic, as it's happening right now, and new discoveries are made every single day.

- The problem statement

What impact had the outbreak on the population of USA. What's the behavour of people.

- Where does the data come from?

The data comes from https://infogears.org/. The data is updated till 07.13.2020. The dataset is based on a survey responses.

Main hypotheses

The virus had a different impact on different age groups, gender and states.

Analysis

Introducing the data

In the filtered dataset from Infogears there are 30088 obervations and 25 features, from which 8 have character values and 17 numeric values. There are 8 age groups, 3 genders, there zip code and state. It mentions symptoms for every individual, their mental health impact, exposure level, how many times they left their house.

```
## 'data.frame':    30088 obs. of  26 variables:
##  $ age                : Factor w/ 8 levels "13-17","18-25",..: 4 3 4 4 5 4 4 4 5 4 ...
##  $ gender             : Factor w/ 3 levels "female","male",..: 2 2 2 2 2 2 1 1 1 1 ...
##  $ zipCode            : chr  "91042" "91436" "91504" "91201" ...
##  $ antibodyTest       : chr  "notTested" "notTested" "notTested" "notTested" ...
##  $ virusTest          : chr  "notTested" "notTested" "notTested" "notTested" ...
```

```
##  $ exposureLevel      : chr  "doNotKnow" "doNotKnow" "doNotKnow" "doNotKnow" ...
##  $ mentalHealthImpact : chr  "someImpact" "someImpact" "significantImpact" "noImpact" ...
##  $ householdHeadcount : num  4 2 4 4 4 8 4 1 3 4 ...
##  $ leftHomeTimes      : chr  "didNotLeave" "didNotLeave" "didNotLeave" "oneTime" ...
##  $ leftForExercise    : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ leftForShopping    : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ leftForWork        : int  0 0 0 1 1 0 1 0 1 0 ...
##  $ leftForOther       : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ healthIssues       : chr  "noIssues" "noIssues" "noIssues" "noIssues" ...
##  $ noSymptoms         : int  1 1 1 1 1 1 1 1 0 1 ...
##  $ bodyAche           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ diarrhea           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ difficultyBreathing: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ disorientation     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ fatigue            : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ headAche           : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ irritatedEyes      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ persistentCough    : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ soreThroat         : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ temperature        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ state              : chr  "CA" "CA" "CA" "CA" ...
```

In this dataset, there are 204 people who tested positive for coronovirus

```
##     n
## 1 204
```

In this dataset, there are 113 people who tested positive for antibodies

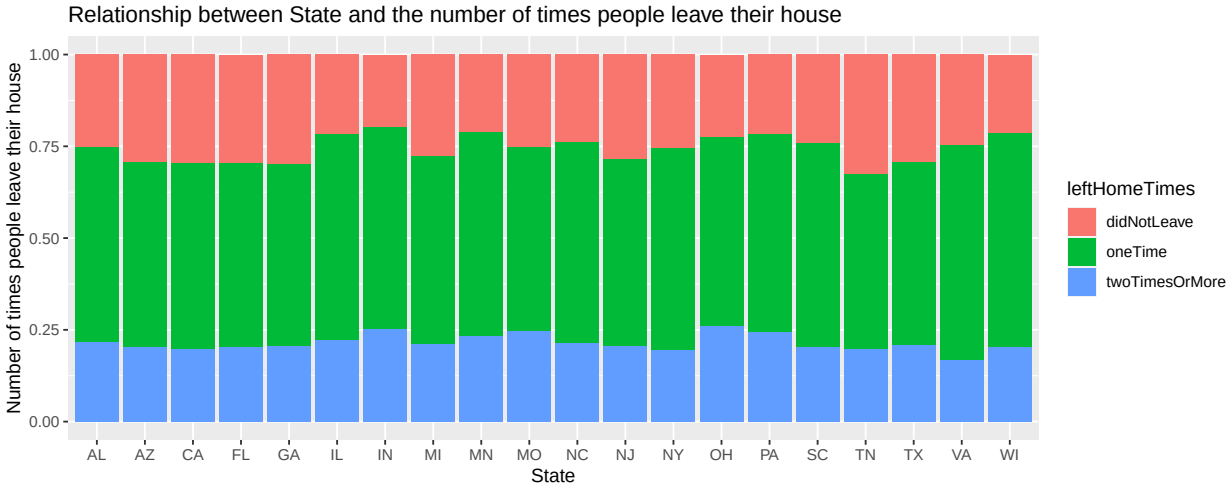
```
##     n
## 1 113
```

The dataset covers almost all states. Here are the top 10 most infected states according to the survey participants. (the number of currently infected or already recovered people is counted)

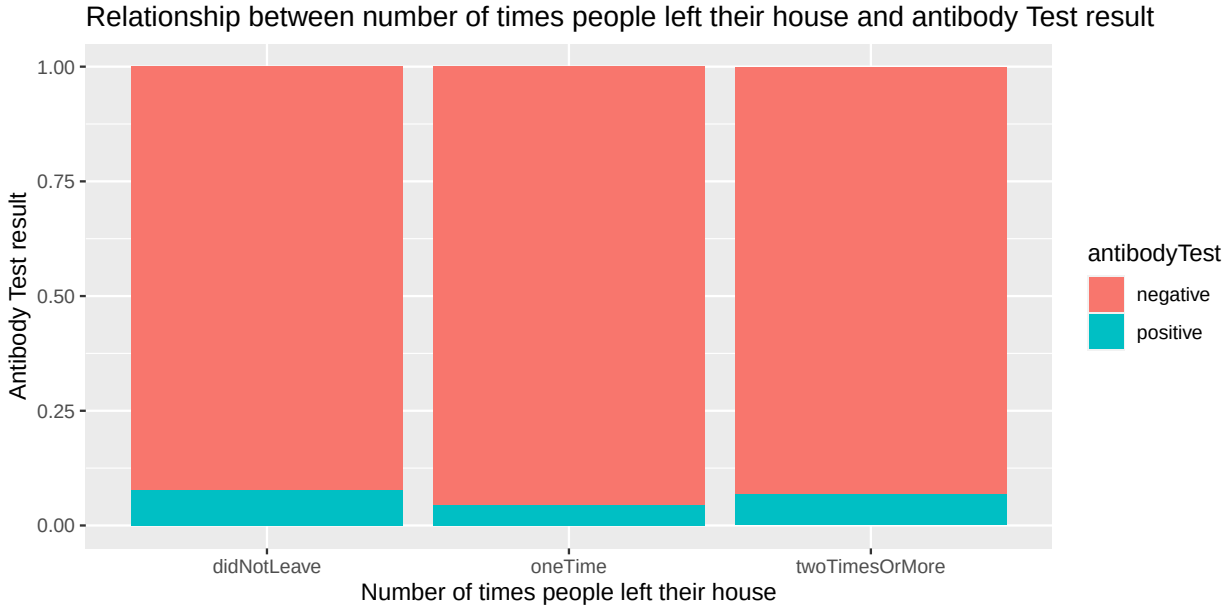
```
## # A tibble: 10 x 2
## # Groups:   state [10]
##    state     n
##    <chr> <int>
##  1 CA      153
##  2 TX       32
##  3 NY       23
##  4 FL       16
##  5 PA       12
##  6 NJ        6
##  7 IN        4
##  8 MI        4
##  9 CT        3
## 10 IA        3
```

Hipothesis 1: The behaviour of people in different states are significantly different

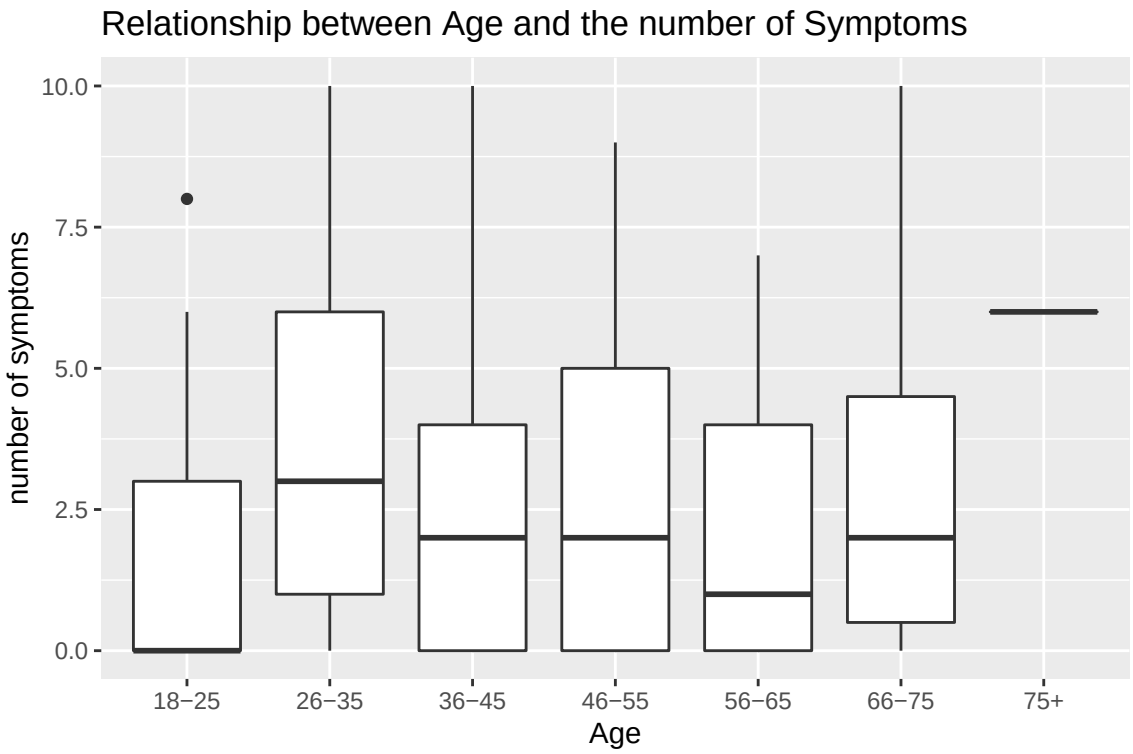
For this plot top 20 states by number of responders is considered. In this graph we can see, that Tennessee is the most responsible state, as they have the most amount of people who stay at home. Overall, the difference is not huge. Most people leave their house only once. So the hypothesis is wrong.



Hipothesis 2: People who do not leave their house, have less chance to get infected It's expected to see fully isolated people to have the lowest percentage of positive tests. From this plot we can see that people who leave their house have once left positive antibody tests. So, staying home doesn't guarantee full safety. The hipothesis is wrong.

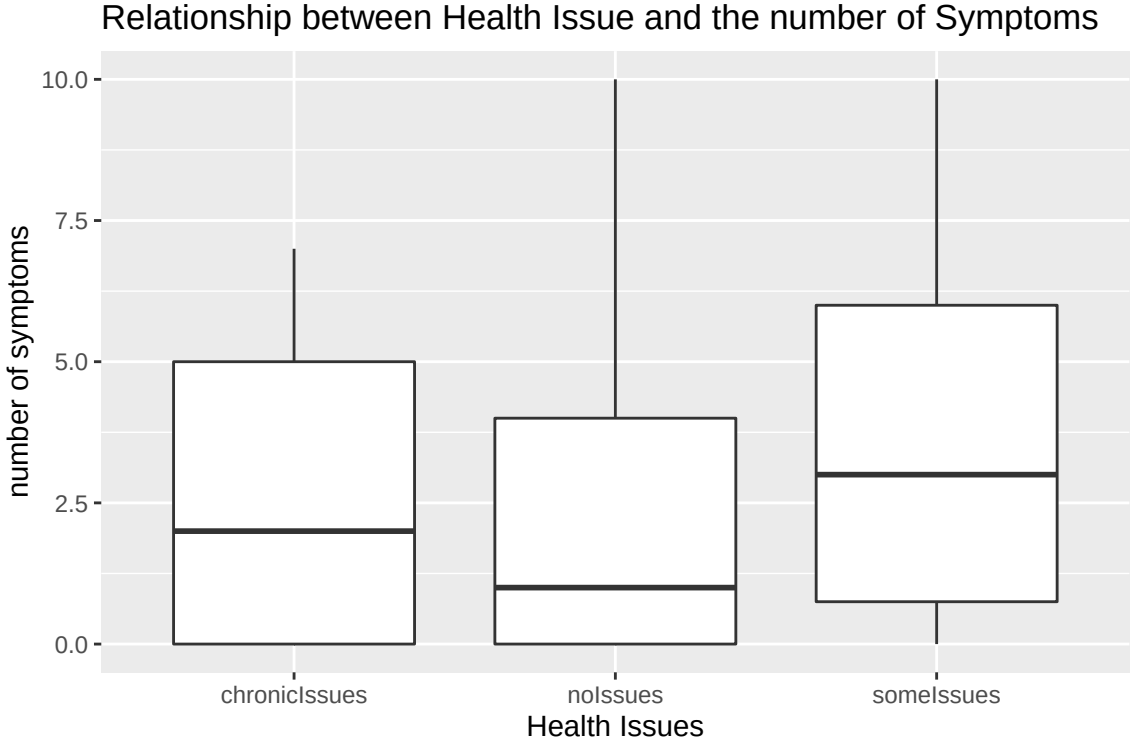


Hipothesis 3: For older people the number of symptoms is more In this graph we can notice that people from age interval of 18 to 25 usually have no symptom (as it's the mean). The only outlier has more than 7.5 symptoms. What's interesting is that people from age interval 26 to 35 have the highest mean from interval 26 to 75. As we were expecting to see people from older generation have the highest mean of the number of symptoms, which is approximately 6. So as we expected the number of symptoms is higher by the age. The hypothesis is right.



###

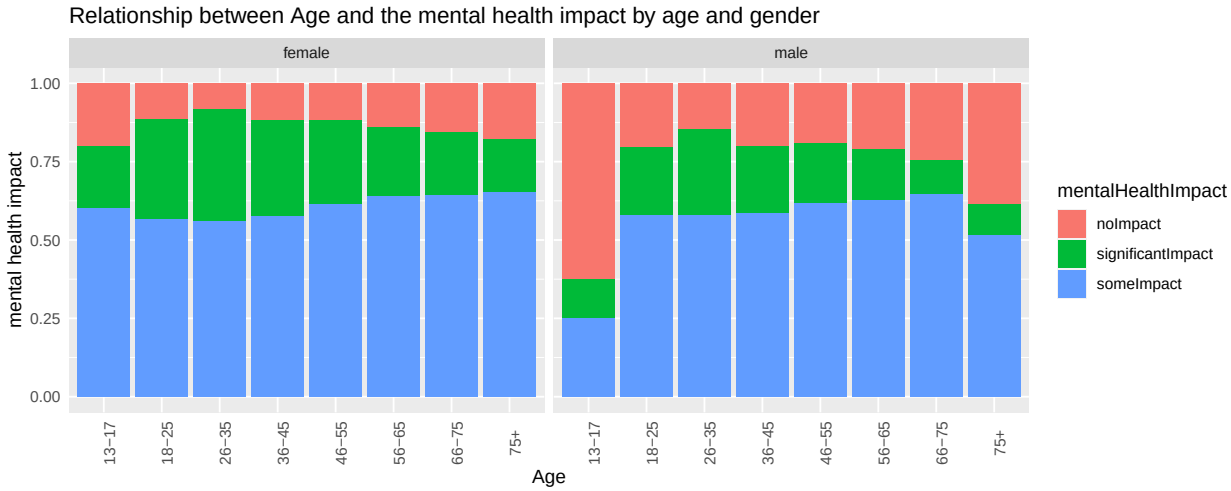
Hipothesis 4: People with cronic and some diseases have more symptoms. As was expected people who have no health issue have less symptoms then people who have either some or chronic issues, have higher chance of getting more symptoms. Which seems weird is that people with some health issues report more symptoms than people with chronic issues. The hypothesis is right.



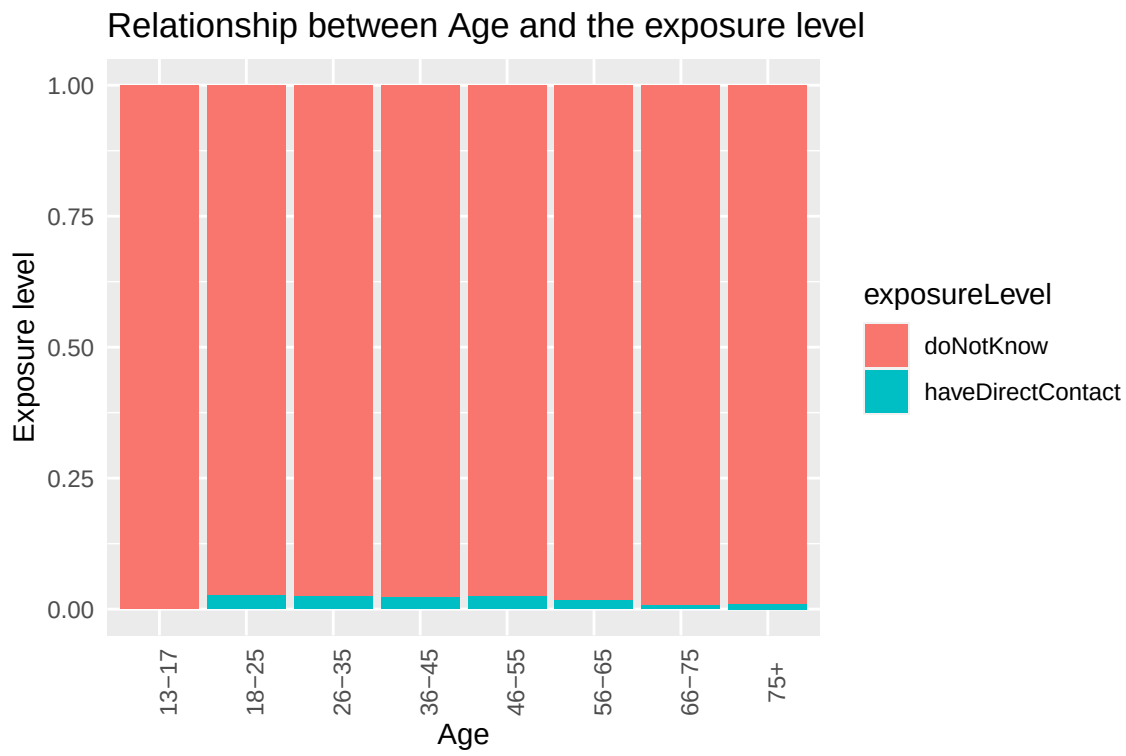
###

Hipothesis 5: Younger and older people are less valnurable for mental health impacts In this graph we can

see that therre is a difference between mental health impact on different ages and genders. We can see that male participants from age group 13 to 17, and 75 and more had the least impact from all genders and age groups. The sensitiveness of females is proven here, as there are less females with no impact in any age group. In females we have the same picture (the youngest and the oldest had the least impact, probably as they are the ones who aren't working). In females and males people in the age interval 26 to 35 had the most impact. They have the highest percentage of significant Impact. As the age becomes higher the impact becomes less. So there is negative relationship between age and mental health impact. The hipothesis is proven.



Hipothesis 6: People in middle age have more direct COntact with an infected person From this plot we can see that people from 18 to 55 have more direct contacts than other groups. This can be explained by their working status. The hipothesis is correct.



###

Hipothesis 7: Fatigue, headache and sore throat are the most common symptoms of coronovirus from this data we can make top 3 symptoms. 1. fatigue - 1463 cases 2. headache - 1422 cases 3. sore throat - 1228 cases

```
##   sum(bodyAche) sum(diarrhea) sum(difficultyBreathing) sum(disorientation)
## 1           830           631                      419                 284
##   sum(fatigue) sum(headAche) sum(irritatedEyes) sum(persistentCough)
## 1         1463          1422                603                  816
##   sum(soreThroat) sum(temperature)
## 1            1228             224
```

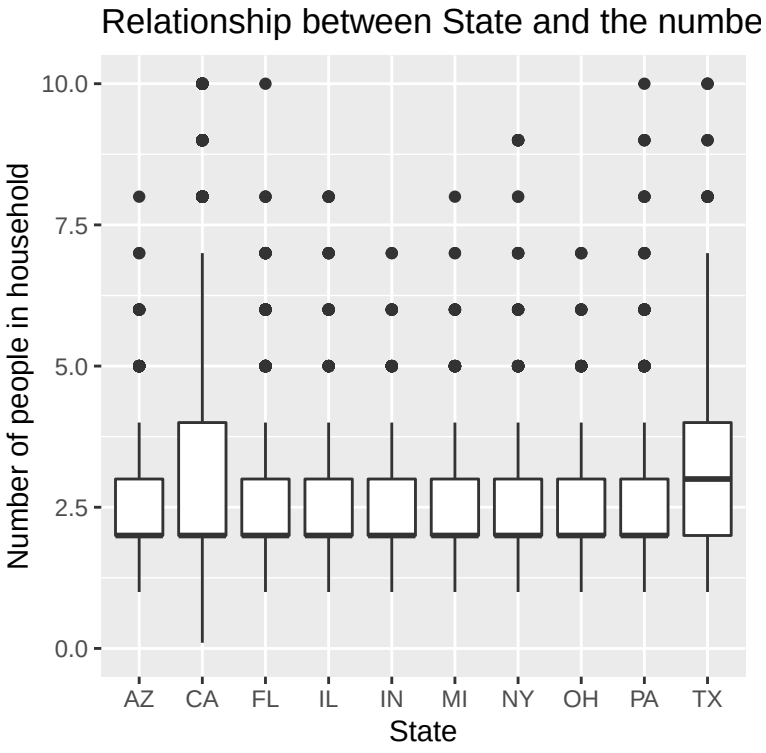
Hipothesis 8: Most people who were infected with coronovirus in the past had no symptom

From all people who tested positive for antibody tests mostly had no symptom with 60 from 113 cases

```
##   peopleWithNoSymptom totalNumber
## 1                  60         113
```

Hipothesis 9: Different states have different number of people in a house, so if one is infected they will infect more people

The mean is mostly same in all states, except Texas. The mean is approximately 3.



Summary of findings and recommendatins if any

The analysis showed that the coronavirus pandemic had a significant impact on people's lives. The youngest and the oldest people are the ones who suffered less mentally, however the older generation and people with some or chronic health issues suffer the most from the disease, as they have the highest number of symptoms. Mostly people who are infected recover without any symptom, and the most common symptoms are fatigue, headache and sore throat. The recommendation that can be made, is to stay home to protect our older generation as they suffer the most.