

Infogears COVID Data Analysis

Sergey Stepanyan

The problem/data description

This report is being performed during the COVID pandemic which started in the first quarter of 2020. As it continues, it is important to understand the patterns which emerge as a result of the outbreak.

The data is provided by Infogears.org. It is a project made possible by the partnership between NetGenix Inc. and Instigate Mobile. Data collection is organized by a questionnaire on the Infogears website, where users fill in the relevant information (eg. what symptoms do they have, have they been tested or not, etc.) and send it for processing.

The dataset includes information about a person's symptoms, what their exposure level is, their gender, number of residents in their household, the mental health impact of the pandemic, whether or not they conducted an antibody and virus test, how often they wear a mask and their age group.

The data was filtered to exclude some misinforming results, duplicate entries and to generally place the data into a more presentable format (converting date strings into Date variables, etc.)

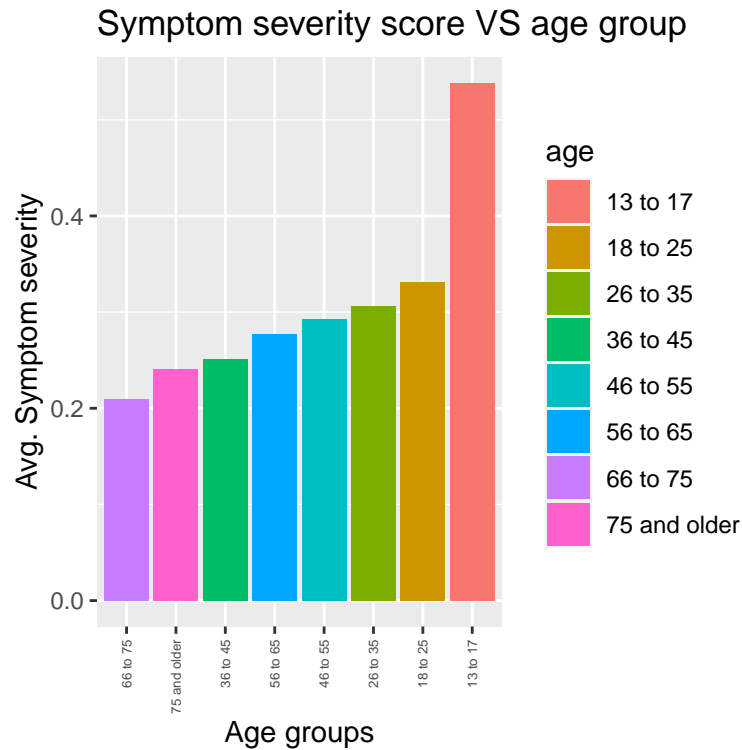
Main hypotheses

1. Generally the younger demographics have more symptoms.
2. From the people with the virus, older age groups are more likely to have symptoms, while younger ones are more likely to be asymptomatic.
3. More than 30 percent of virus "positive" cases have no symptoms.
4. The more severe the symptoms the more likely it is the person went to be tested. More people got tested over time.
5. Households with larger sizes are more likely to get infected. Inconclusive, since not enough data is available.
6. Mental health impact by age group. Youngest age demographic has had the least effect on mental health.
7. Mental health impact by gender. Females were impacted by the pandemic more than males.
8. People who never wear masks have had less mental impact than those who wear them.

Analysis

1. Generally the younger demographics have more symptoms.

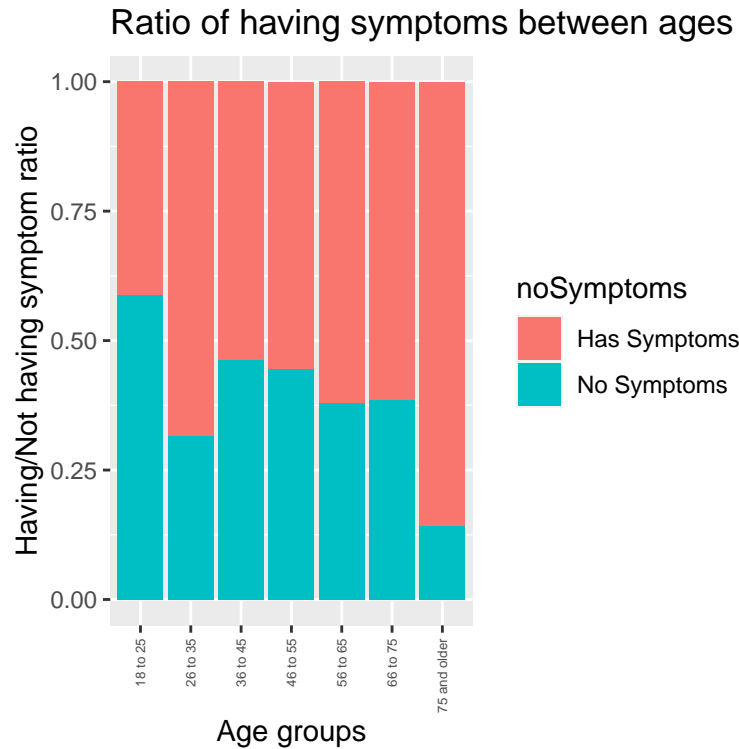
For this hypothesis, we are going to look at 8 different ages groups. In order to show the data, for each age group the average amount of symptoms was taken.



From this picture, we can see that in in general older age groups have a lower average for the amount of symptoms, and the highest average is by “13 to 17” age group. This can be explained, by the lower age groups “less experienced” immune system. For instance, the older people may have had more illnesses in the past and are more prepared for the pandemic. It is also possible that given date which includes older folks, the picture would reverse, as the immune system of the more elderly would be less capable of defense.

2. From the people with the virus, older age groups are more likely to have symptoms, while younger ones are more likely to be asymptomatic.

Taking the same age groups, we can also try to understand the correlation between asymptomatic and symptomatic people, having been tested positive for the virus.

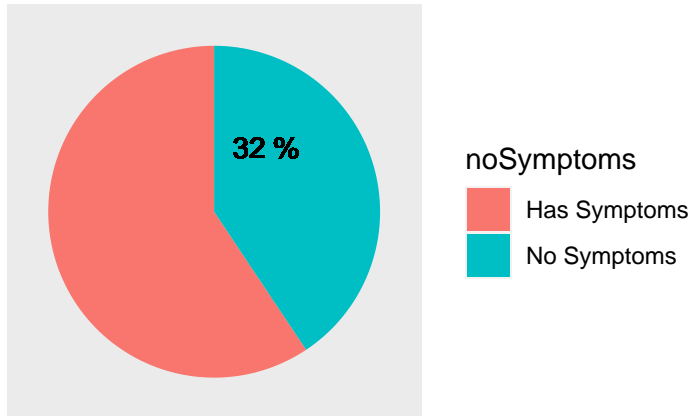


By the graph, we can tell that the “18 to 25” age group has the lowest proportion of symptomatic cases, and on the contrary the “75 and older” category has the largest proportion.

3. More than 30 percent of virus “positive” cases have no symptoms.

We are going to look through the data points where the virus test and antibody test results have been positive, using this information let’s construct the ratio between the symptomatic and asymptomatic cases.

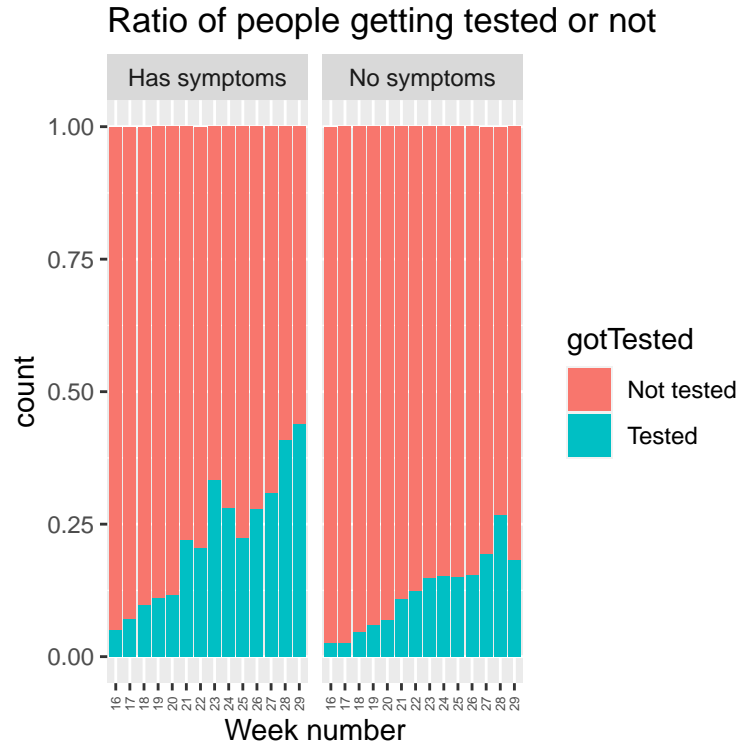
Ratio of asymptomatic cases



As we see from the graph above a large enough proportion (about 30 percent) of positive cases is asymptomatic. This result is including all age groups this time.

4. The more severe the symptoms the more likely it is the person went to be tested. More people got tested over time.

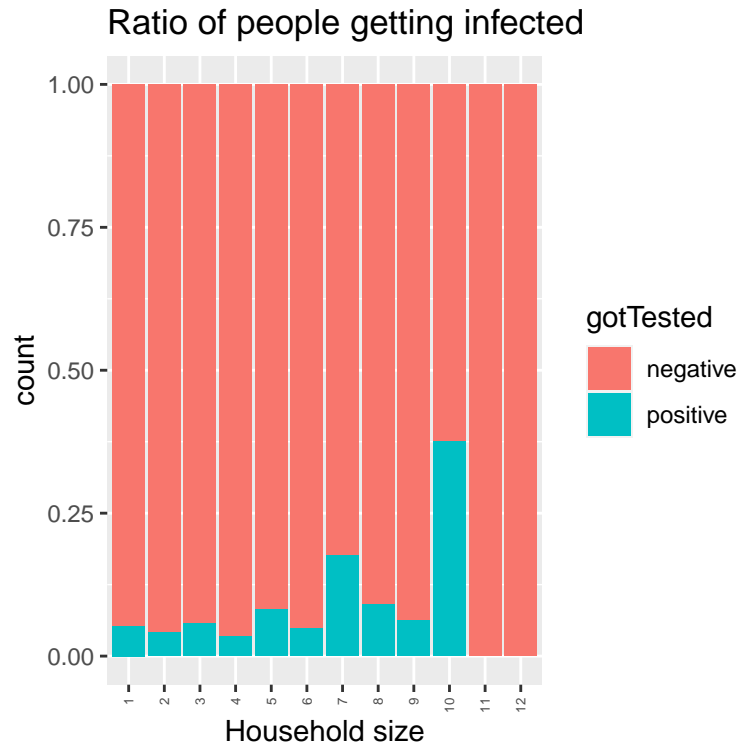
Next, we are going to look through the cases when the person went and had any kind of test taken (be it antibody or virus test).



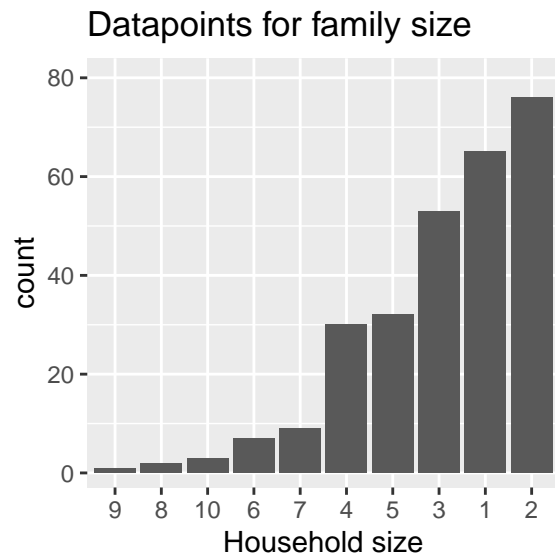
Judging by the result, it is safe to say that the vast majority of people without symptoms aren't tested and even the symptomatic people mostly also aren't tested. From the graph, we notice that more people get tested, if they actually have symptoms and in fact, a larger portion got tested over time.

5. Households with larger sizes are more likely to get infected. Inconclusive, since not enough data is available.

Let's look through data samples, where the household sizes are larger. Assuming that the chances of a person from a larger family going outside and getting infected, let's test if this is the case.



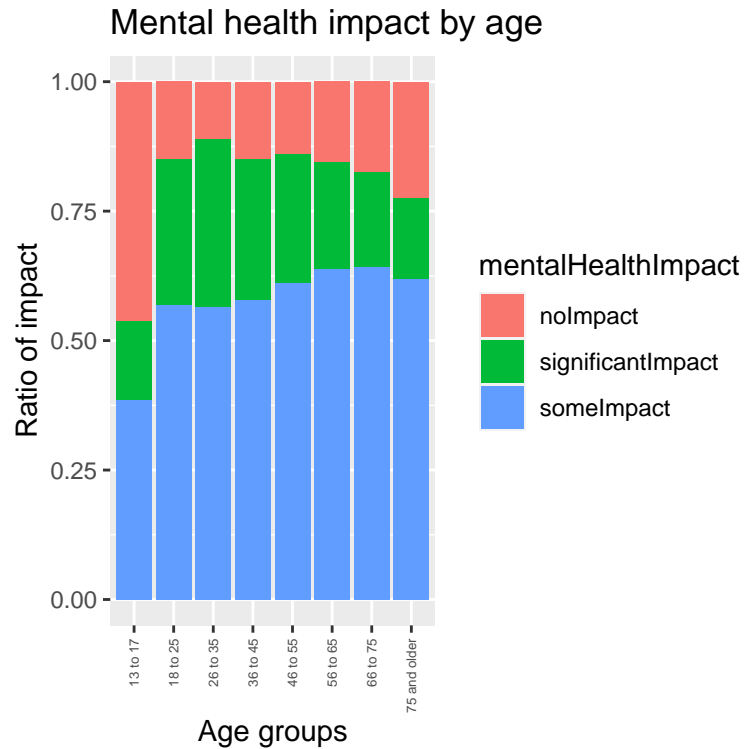
From the first graph, we see that there seems to be an emerging pattern, but the as the household size gets larger it becomes more and more difficult to justify this hypothesis.



Therefore, from the second graph which shows the amount of observations for the households, we can say that this hypothesis is rather inconclusive as not enough datapoints are available.

6. Mental health impact by age group. Youngest age demographic has had the least effect on mental health.

Once again, let's turn our attention the different age groups and see how the pandemic affected the mental health of the population.

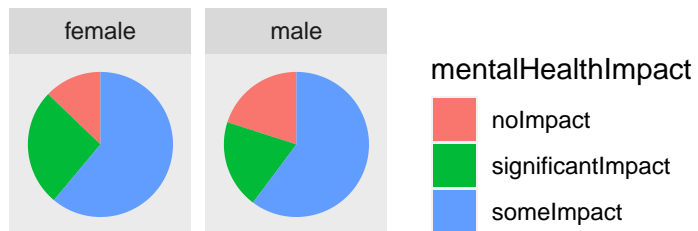


From the data we can tell, that the least effect was on the “13 to 17” age group, while the most significant impact was on the “26 to 35” age group.

7. Mental health impact by gender. Females were impacted by the pandemic more than males.

Looking through the mental impacts, moving on to the gender.

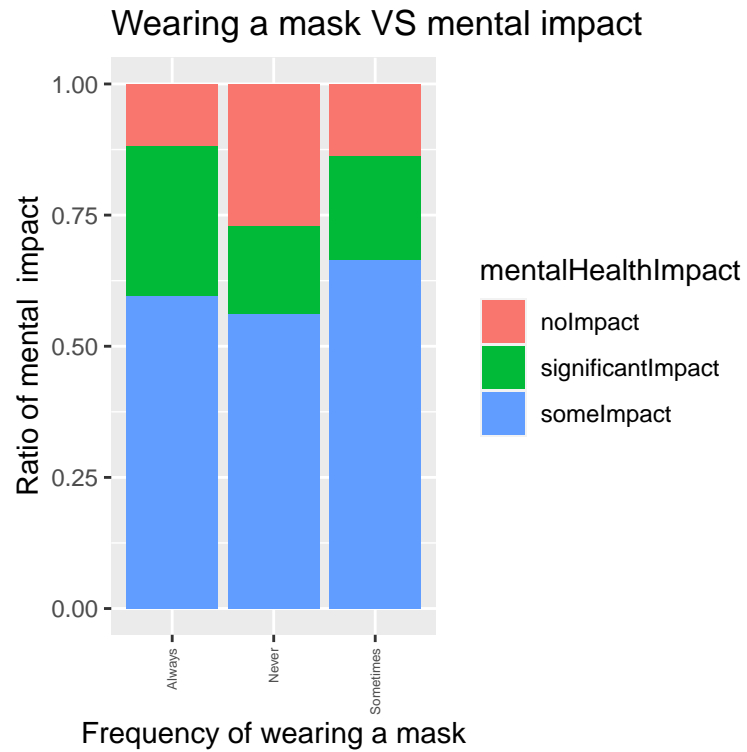
Mental health impact by gender



We can see that males, by proportion, are less likely to be mentally impacted by the pandemic.

8. People who never wear masks have had less mental impact than those who wear them.

In the category of protection, we can see from the data, that there is a connection between the people not impacted by the pandemic and those not wearing a mask.



Summary of findings and recommendatins if any

1. Most people regardless whether or not they have symptoms, didn't get tested.
2. People are getting tested more, but the data currently shows that the progress isn't fast.
3. Those who aren't phased mentally by the pandemic are wearing masks the least.