The World Pandemic Through The Stereotypes

Tatev Kyosababyan

7/20/2020

This document provides a simple analysis of the data set by the InfoGears data platform. The data has been collected through an online survey conducted by the NetGenix Inc. and Instigate Mobile partnership–survey website: https://infogears.org/. The targeted areas include Glendale, Burbank, Pasadena (California), and neighboring districts, although it includes responses from areas other than mentioned. Individual ZIP codes categorize the data, and the IP addresses of each submission are held in the record as well. The data-set contains information about the individuals' symptoms, activity, and demographic information. The survey takers have been encouraged to retake with intervals of a few days in between, to keep track of the changes over time. The goal of the data is to help local communities visualize the prevalence of coronavirus symptoms and other aspects of life impacted by the crisis. The analysis of such data is significant for the enhancement in the area of public health, so it can be used to acknowledge the further actions and visualize the reality to come up with the right plan for fighting against the virus.

Aside from the hard work conducted by professional medical workers, people, just like in any other situation, always find a way to come up with stereotypes and spread them more significantly. Although sometimes they math with the reality, these non-professional fictional facts often interfere with the procedure of fighting against the virus, as people find less urge to follow the directions given by the professionals, than those fictional beliefs.
I will be providing graphs that help drive into conclusions about the overall situation and people's behavior during the pandemic. With the help of the sketches based on the provided data-set, we will be able to make more precise determinations regarding some stereotypes people tend to believe in strongly.

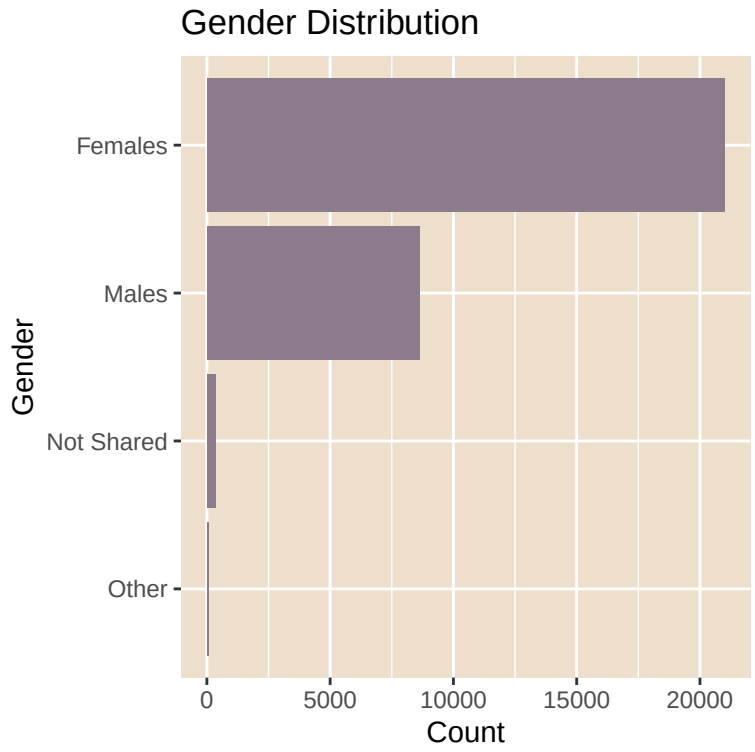
```
    The stereotypes we will be testing in this data analysis are:
    -> Mental health difficulties are common throughout the pandemic.
    -> Females are more likely to get infected by COVID-19 than males.
    -> Younger people are less likely to get infected.
```

```
## 'data.frame':    30058 obs. of  33 variables:
##  $ age_groups         : chr  "interval_36_45" "interval_26_35" "interval_36_45" "interval_36_45" ...
##  $ antibodyTest       : chr  "notTested" "notTested" "notTested" "notTested" ...
##  $ createdAt          : chr  "4/16/2020 2:11" "4/16/2020 2:14" "4/16/2020 2:18" "4/16/2020 2:36" ...
##  $ exposureLevel      : chr  "doNotKnow" "doNotKnow" "doNotKnow" "doNotKnow" ...
##  $ faceCovering       : chr  "" "" "" "" ...
##  $ gender             : chr  "male" "male" "male" "male" ...
##  $ guid               : chr  "b91c76ab-449a-4b2b-985a-60dd57d17145" "24ad589e-adc0-47ae-9515-63e7797
##  $ healthIssues       : chr  "noIssues" "noIssues" "noIssues" "noIssues" ...
##  $ householdHeadcount : num  4 2 4 4 4 8 4 1 3 4 ...
##  $ ip                 : chr  "174.193.218.227" "107.185.158.57" "137.25.127.165" "68.190.201.25" ...
##  $ leftHomeTimes      : chr  "didNotLeave" "didNotLeave" "didNotLeave" "oneTime" ...
##  $ mentalHealthImpact : chr  "someImpact" "someImpact" "significantImpact" "noImpact" ...
##  $ updatedAt          : chr  "4/16/2020 2:11" "4/16/2020 2:14" "4/16/2020 2:18" "4/16/2020 2:36" ...
##  $ userAgent          : chr  "Mozilla/5.0 (iPhone; CPU iPhone OS 13_3_1 like Mac OS X) AppleWebKit/60
##  $ virusTest          : chr  "notTested" "notTested" "notTested" "notTested" ...
```

```
##  $ zipCode            : chr  "91042" "91436" "91504" "91201" ...
##  $ 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 ...
##  $ id                 : chr  "26,784" "26,785" "26,786" "26,787" ...
##  $ irritatedEyes      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ leftForExercise    : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ leftForOther       : int  0 0 0 0 1 0 0 0 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 ...
##  $ lossOfSmell        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ noSymptoms         : int  1 1 1 1 1 1 1 1 0 1 ...
##  $ 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 ...
```

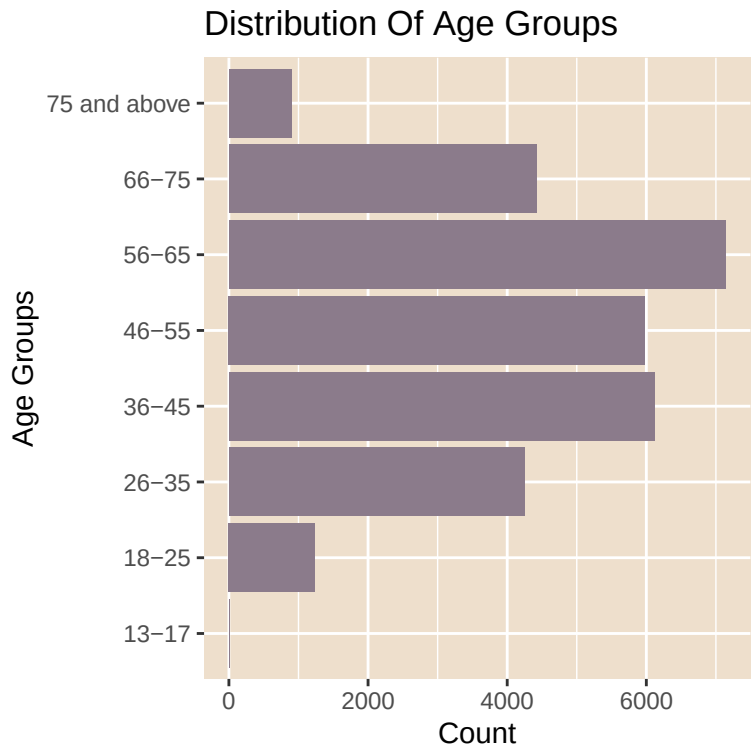
The data-set has overall 30058 observations, and 33 variables which include information regarding the individuals' activity, demographics and symptoms.

Let's first try and understand who were the survey-takers according to their age groups and genders.



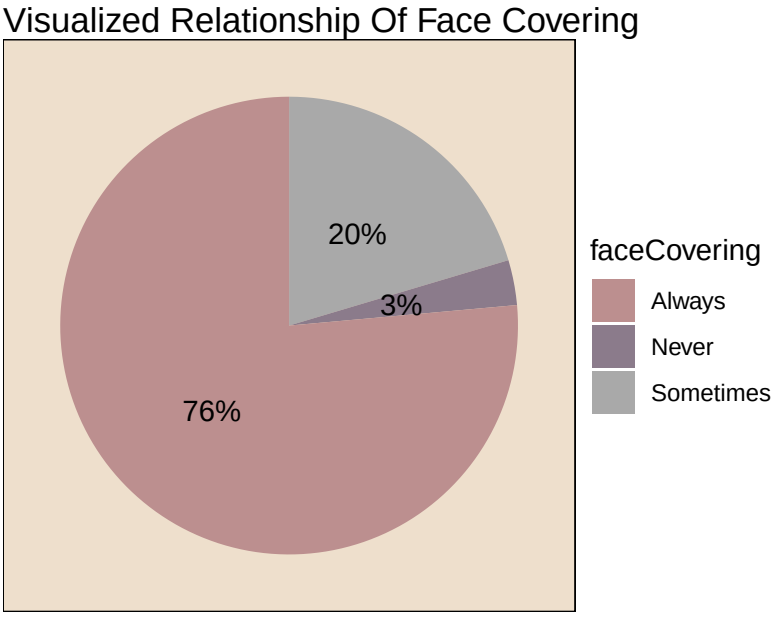
In the graphs we see that the number of female survey-takers is almost twice as much as the one of males. Maybe this is a sign analyzing the stereotype of females being infected more is a good idea? :)

A lot of people were arguing the virus is more likely to infect to those who are older, because younger people tend to have stronger immune systems. Will discuss that further later, but for now let us understand the age groups provided in our data-set.



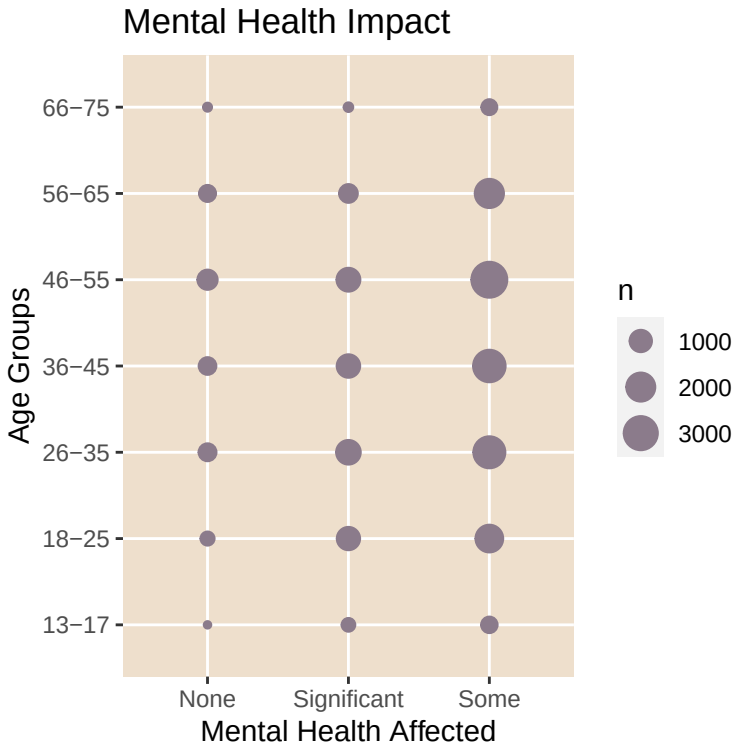
We can vividly see, that the significant amount of survey-takers belong to the age group 26-75. Age groups like 18-25 have 1237 responses, so we are mainly dealing with individuals starting from the adulthood until the old ages.

Initially, when the pandemic first started, wearing masks was optional, and hence, at that time the response for the question regarding the face covering was optional. These responses make the 27.6488966% of the entire data, and for the sake of getting a better sight of the situation I decided to disregard those cases in this chart below.

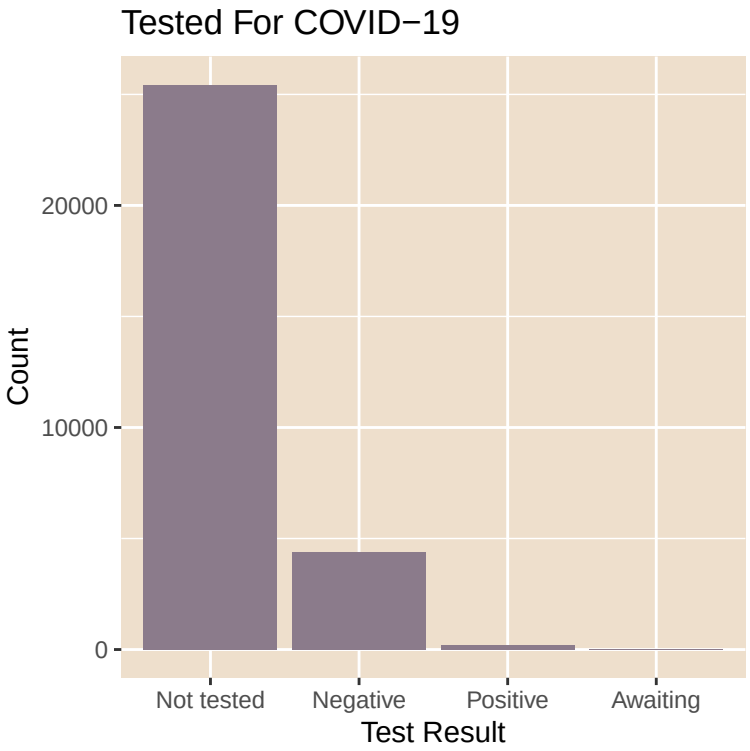


Here we see that most people who took the survey and gave valid responses have been wearing masks always, one-fifth uses masks sometimes and only a small portion does not ever wear masks which makes 692 people.

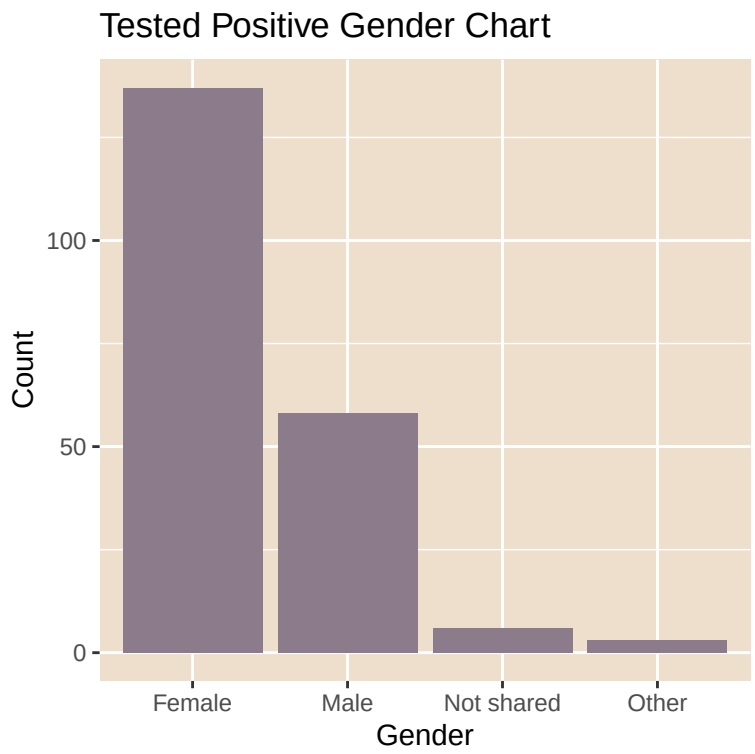
The pandemic has had its impact on people's mental health and studies conducted in the United States investigating COVID-19 patients found a high level of post-traumatic stress symptoms (PTSS) (96.2%).



From the density in the graph above we see that people mostly insisted the pandemic has influenced their mental state at least somehow. The graph as well illustrates the fact that those who noticed no impact on their mental health make 14.9228928% of everyone taking the survey. These results are not very far from the scientific research, and this time people are actually proved to be right, pandemic does have its consequences on the population's mental health.



There are 4626 people from our survey that took a COVID-19 test, and only 204 have tested positive. This is great news from the health standpoint.
As I mentioned before, there was a stereotype insisting that females are more likely to catch the virus than males. Let us find out the accuracy of that information using our data-set.



From the first sight, the stereotypes seem to be winning the argument because of the significantly large difference between the females testing positive and the males testing positive. However, the number of females here is 137, while the number of the males is 58.
Let's recall the the beginning of our analysis. We saw that the number of female survey-takers was significantly larger than that of the males. More specifically, the number of females in the full data-set equals 20998, and the males is 8636. To be more precise, the ratio of females and males testing positive is 2.362069, and that of females and males overall taking this survey is 2.4314497. The numbers are pretty close, thus, the chart is not the most accurate representation of the ratio between female infected patients and the male female patients, in spite of its accuracy of representation for our data-set. Therefore, the stereotype of females getting infected more is proved to be wrong.

Another famous stereotype regarding the spread of the virus, is that since younger people have stronger immune systems, they are able to handle the pandemic better and not be infected as much as the older people. Let's try to drive into conclusions through sketching graphs using our data.

Here again, we see that the stereotypes are not the most accurate facts to follow. Despite the fact that the number of survey takers in the age group 18-25 was significantly less than the number of the older age groups, it is the second with its largest number of positive testers.

As a suggestion, I thought it would be a better idea if the survey-takers could have their own profiles and make the updates there, instead of having same individuals submitting responses as new. This way we could track the changes considering individual cases. Another idea, is that having the settings of the data collection more precisely would reduce a few steps in the overall data analyzing process. As mentioned before, there were some invalid responses, which at some point could create more vague results and interfere with the accuracy of the data.

Overall, we found that despite stereotypes can coincide with the reality, sometimes, it is better always to educate ourselves from valid sources to avoid unnecessary complications. The world pandemic has caused astonishingly many hardships for people, making life even more miserable with its sudden changes. The latter has caused a significant impact on people's mental health, which itself brings new challenges to the standard path of sustaining life. Contrary to the stereotypes, we saw that catching the virus is equally likely to happen no matter the gender. Healthy immune systems are not the best guarantee for staying healthy and not becoming one of the millions of others infected. Conclusion: always stay careful, take care of oneself because, at this point, everything we do affect the lives of our loved ones as well.

Github Link: https://github.com/tatevkyosababyan/individual_project Sources Used: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7260522/