

Title: "Analysis of Four Types of Masks Tweeted During the Covid-19 Pandemic"

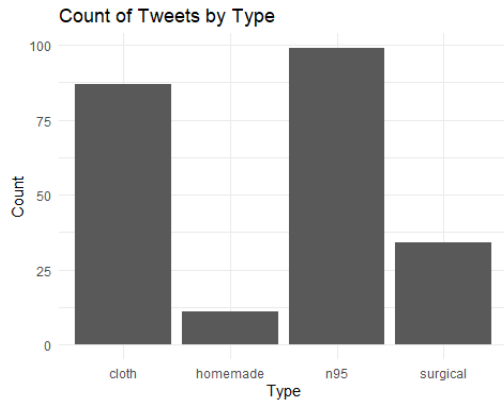
Author: "Aretha Miller" "DS710" "May 6, 2020"

Executive Summary

According to the Centers for Disease Control and Prevention ([CDC], 2020), many people may have contracted COVID19, but not show symptoms. As a result, people could unknowingly transmit the virus to others via sneezing, coughing, or speaking. Considering this knowledge, the CDC recommends that civilians wear cloth masks for facial coverage while out in public. Alternately, frontline healthcare workers (FHWs) should wear respirators called N95s, when treating patients with COVID19. Nevertheless, there is a global shortage of masks and a strategy to optimize the use of N95s for FHWs is to encourage the public to wear other versions of face coverings, for instance homemade masks. The purpose of this work is to collect data from Twitter and analyze what people are saying about face masks. We wish to discover if the distribution of people who commented about cloth, homemade, surgical masks, or N95s occurred in equal proportions, or whether there were more comments about one or more types of masks. The results may be informative to public health leaders and FHWs to reveal whether civilians were discussing issues of facemasks in the current pandemic.

Keywords, Data Collection and Analysis

Keywords used to collect data from Twitter were "mask" and "N95." The word "mask" was chosen to get maximum amounts of tweets, while the word "respirator" was omitted, since the words "N95 & respirator" together produced limited tweets. These keywords include the recommended face coverings of the CDC and therefore appropriate to address the research question. The dataset contained 1670 tweets and Python software was used to extract tweets from Twitter's RepresentationalStateTransfer (REST) architecture and streaming Application Programming Interfaces (APIs). The REST API collected data over seven days, while data from the streaming API were collected May 3, 2020. Data were cleaned and duplicates and retweets excluded. Tweets were then formatted and saved to a file for statistical computing and graphical analysis with another programming language called R. A chi-squared goodness of fit test was used to determine whether the frequencies of the observed sample differed significantly from the expected frequencies. The null hypothesis assumed no difference between the observed and expected values of tweets that appeared with the keywords. In essence, the prediction was that each of the four types of masks were equally distributed or equally likely to be tweeted about.



Counts table

| type | count | percentage |
|----------|-------|------------|
| homemade | 11 | 4.76 |
| cloth | 87 | 37.66 |
| surgical | 34 | 14.72 |
| n95 | 99 | 42.86 |

Hypothesis Testing

Chi-squared Goodness of fit test was used under the assumption that the distribution of tweets among the four types of masks were similar. The null hypothesis was no difference between the observed and expected values ($p_1 = p_2 = p_3 = p_4$). Each of the four types of masks were equally distributed or were equally likely to be tweeted about. While the alternative hypothesis expected a difference between the observed and expected values. In other words, at least one of the four types of masks were unequally distributed or unequally likely to be tweeted about ($1/4$). A chi-squared goodness of fit test was used to determine whether the frequencies of the observed sample differed significantly from the expected frequencies specified in the null hypothesis.

Findings and Conclusion

Based on a p-value of $< 2.2e-16$ (significance level of 0.05), there were enough evidence to reject the null hypothesis and support the claim that there was a significant difference between the observed and expected values. In other words, at least one of the four types of masks were unequally distributed, or unequally likely to be tweeted about ($1/4$). Likewise, real time sentiment analysis revealed that people's attitudes and written opinions about masks were neutral. Those who wear face coverings commonly state how uncomfortable some masks can be. Yet, tweets mentioned N95 respirators the most, while homemade masks were referred to the least (Figures 1 & 2). The social implications of these findings suggest that people from the "Twitter verse" may not be very concerned about face coverings as personal protection and may continue to spread the virus.