Mask Efficacy

# **Adam Curry Summer Term - DSC680-T302 Applied Data Science (2217-1)**

In February 2020, if you told me I’d be wearing a mask to go grocery shopping, I would’ve laughed. However, as soon as the pandemic hit, I quickly realized this may be a necessary move to help save lives. I saw South Korea lead the way, as they implemented masking right away and saw few cases. I was on board with masking, as this appeared to be an indicator that masks worked. Once the panic of the virus started to settle and we had data to review, I wondered if masks truly worked. South Korea is still masking indoors, and noncompliance can lead to fines [1][2]. Yet they have seen case spikes higher than any point in this pandemic within recent months [Appendix A]. If masks are how we slow the spread, what is happening in these other countries that are seeing sudden spikes? This analysis will focus on the Unites States and the various mask policies implemented at the state level.

The purpose isn’t to uncover if masks work or not, rather to show the effectiveness of mandating them at the state level. Some claim that certain states followed the science based on mask mandates, and with the recent surge in COVID cases related to the delta variant, there have been new calls for mask mandates [3][4]. Are these calls warranted and are they being made as a result of empirical evidence? Do the data agree with statewide mask mandates? Some of the questions this analysis seeks to answer include:

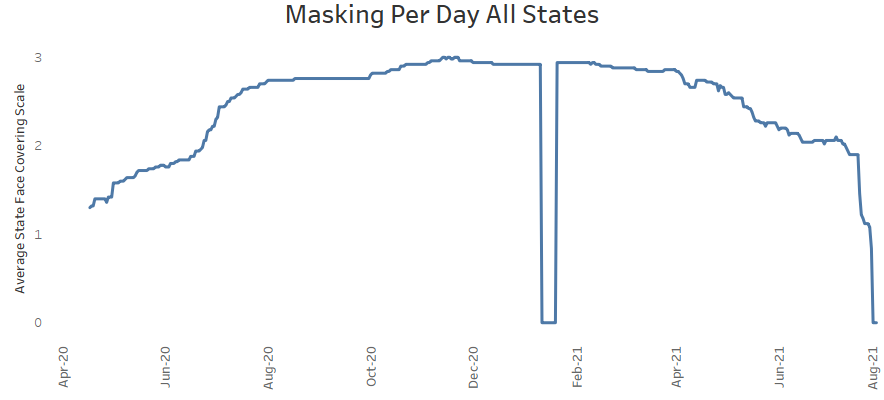
* What states have the most stringent masking policies?
* What states have the least stringent masking policies?
* Did statewide policies make a difference in suppressing covid case counts?
* Can covid case predictions be made based on masking policies?

Some questions I anticipate from more technically minded individuals include:

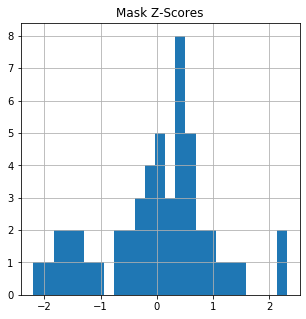
* Could this same method be applied at a more granular geographical level?
* You stated that the decision tree and random forest models had a decent prediction level, could this be an indicator that masks work?
* How would you alleviate the masking policy discrepancy in Texas compared to California?
* What were some of the correlation metrics that you noticed?
* Is a statewide mandate the best way to test if masks work or not?
* Could the droplets that masks stop have an impact on covid case numbers?
* Do you have plans to refresh the dataset with updated case counts and mask numbers?
* Why did you focus on 6 states in your neighboring state comparison?
* Is this an oversimplification of masking policies?
* Did you review hospitalizations and or deaths?

The dataset I found is one I used from a prior analysis, where I analyzed state government’s COVID response on the economy. The data is collected by Oxford university and contains a robust data dictionary with a historical record of cumulative cases and “facial covering” policies [5]. There are a wide variety of other metrics, even a stringency index that shows how governments responded to the virus, but the one two I was interested in were: H6\_Facial Coverings and ConfirmedCases. This dataset contains regions from across the world, so I had to reduce it down to just US state governments (plus DC). The facial coverings metric was defined five different ways [6]:

* 0 – meant there were no mask policies
* 1 – meant masks were merely recommended
* 2 – meant that masks were required in some specified shared/public spaces outside the home with other people present, or some situations when social distancing not possible
* 3 – meant masks were required in all shared/public spaces outside the home with other people present and/or all situations when social distancing was not possible
* 4 – meant that masks were required outside the home at all times regardless of location or presence of other people

I did some basic data exploration and found that level four was rarely enforced. Delaware, Illinois, Massachusetts, Maryland, Montana, North Carolina, Pennsylvania, South Carolina, and Utah were the only states that enforced outdoor mask mandates. I also looked for gaps in the data. As you can see, the image on the right shows a massive gap in the facial covering metric. This doesn’t make logical sense, as January 2020 was the peak pandemic within the states. To mitigate this, I had to group the data by state and grab the mode for the month of January. I then imputed these values into the dataset, and this smoothed out the January gap [Appendix A]. I also noticed that the covid case column was cumulative case counts, which isn’t helpful when looking for correlation and daily case distribution over a timeline. To mitigate this, I partitioned the data by state, ordered by date, and subtracted the record from the prior record. This fixed the daily case distribution. Lastly, I had to normalize for population, as Texas would easily see more cases than Nebraska. Once this was complete, I had a complete dataset.

As a note, I did notice some potential data discrepancies. For example, Texas appeared to have ended their mask mandate after California (June 13th), according to this data. However, a quick Google search shows Texas ended their mandate in March [7]. Either way, the data appeared to be mostly accurate, and I need to make the assumption that this is a metric that has specific parameters defined by Oxford. Therefore, I proceeded with my analysis.

Now that the dataset was defined and cleaned, I was able to begin my analysis. The first hypothesis question was: What states have the most stringent masking policies? To find this, I utilized the face coverings metric, which is an ordinal scaled metric. I added up the total number of days with a mask mandate, and the associated 0-4 mask scale. I found that North Carolina and Massachusetts had the most stringent mask policies, as they were greater than two standard deviations below the mean. The opposite showed that Georgia was the least stringent at greater than two standard deviations above the mean followed closely by Minnesota and Nebraska who were about two standard deviations above the mean.

**Hypothesis Question # 1 - What states have the most stringent masking policies?**

North Carolina and Massachusetts had the highest 0-4 scale per month than any other state. However, New York had the most months that required a mask. Remember, some states required masks outdoors, even when nobody was around (scale 4). These mask policies pushed North Carolina and Massachusetts above the rest.

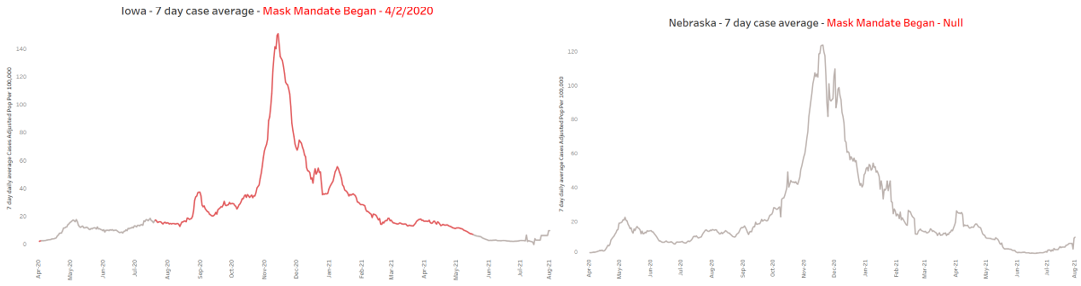
**Hypothesis Question # 2 - What states have the least stringent masking policies?**

Nebraska, Minnesota, and Georgia had the lowest 0-4 scale per month than any other state. However, South Dakota and Wyoming also never had a mask mandate. Remember, some states “recommended” masks (scale 1), which added to the totals.

**Hypothesis Question # 3 - Did statewide policies make a difference in suppressing covid case counts?**

For this I created two separate datasets, one for the high masked states and one for the low masked states. Correlation heatmaps in the high mask population show that there is a .45 positive correlation with masking outside and a -.25 negative correlation when masking at scale 2 (required in some specified shared/public spaces outside the home with other people present). For the low mask rate there is a .22 positive correlation with masking scale 2 and a .14 negative correlation when masking is recommended. The scatter plots below represent a month for each of the populations. The charts below show that the more months that required masks, the higher the cases were (for both populations).

However, when comparing the same variable (“required in some” – scale 2), we can see the opposite occurring in the “high masked” population. Appendix C contains four charts. The first set shows that the left chart (high masking) shows cases dropping with “required in some” while the right chart (low masking) shows cases increasing with “required in some”. This leads me to believe that you can’t tell one way or the other if state mandates worked based on correlation alone. What we are seeing on the left are months where cases were already low, therefore the state’s mandates were at the lower end of the scale, while the right shows cases were already high when mask requirements were made [Appendix C].

Showing neighboring states shows that states with a mask mandate have the exact same curve and case counts as those without mandates. Looking at Nebraska and Iowa’s seven day moving case average, adjusted for population, shows near identical movement and case counts. The same can be seen for other states Minnesota compared to Michigan, Georgia compared to North Carolina [Appendix D].

The final step to help answer this question, is a multi linear regression model using both populations. The analysis yields a relatively low R2 where the highest was at .28 and the at lowest .1. This seems to indicate that state mask mandates account for little variance in covid cases. All p-values are statistically insignificant ranging from 0.161 to 0.674. Examining the scatter plots seem to indicate that covid cases went up with the implementation of masking. Meaning we can't reject the null hypothesis - masked states vs unmasked states don't appear to stop the spread of covid cases at the state level. To illustrate this, I created a time laps of the American map (minus Alaska and Hawaii), showing weekly cases aggregations in red, and a red X for weeks a mask mandate was implemented.

**Hypothesis Question # 4 - Can covid case predictions be made based on masking policies?**

The final step was to predict masking policies based on covid cases. I did this with the entire population of states and yielded pretty good results. The Support Vector Machine yielded 71% accuracy, Decision Tree 66% accuracy, and Random Forest 66% accuracy. However, these results are very misleading. In fact, I’d like to rerun the analysis in a couple of weeks, as predictions are being made that case counts are spiking. The difference in the analysis would be that states aren’t mandating masks with the rise in cases as the vaccine has been rolled, whereas before they “played it safe” and mandated masks to err on the side of caution.

The “next steps” I could take for this analysis would require a deeper dive into the data set. As I pointed out earlier, there were some data discrepancies. In addition, I would like to review this same analysis at the county level. Perhaps group counties by features (i.e., economy, population density) along with disparate mask policies could yield interesting results. I’d also like to do a full analysis on the spread within schools. Many studies and analyses have shown that children don’t get nearly as sick and are not the vectors of the disease [8][9][10][11][12]. As a parent, I make calculated risks with my children. I can’t shelter them forever, that wouldn’t be fun for them or me. Instead, I can interpret empirical data rather than anecdotes and weigh the risk/reward as a parent. This will ultimately be the next analysis I do around masking policies.

**Appendix/References:**

1 <https://www.loc.gov/item/global-legal-monitor/2020-12-07/south-korea-mask-rule-violators-may-be-punished-by-fines/>

2<https://www.google.com/search?q=south+korea+covid+cases&rlz=1C1CHBF_enUS850US886&oq=south+korea+covid+cases&aqs=chrome..69i57j0i131i433i457i512j0i512l8.4502j0j4&sourceid=chrome&ie=UTF-8>

3 <https://bestlifeonline.com/first-states-mandate-masks/>

4 <https://www.nbcnewyork.com/news/coronavirus/calls-for-renewed-mask-mandate-grow-as-delta-fuels-covid-surge-in-nyc-u-s/3161139/>

5 Hale, T. Angrist, N. Cameron-Blake, E. Hallas, L. Kira, B. Majumdar, S. Petherick, A. Phillips, T. Tatlow, H. Webste, S. (28, June 2021). Variation in government responses to COVID-19. Retrieved from https://www.bsg.ox.ac.uk/sites/default/files/2020-09/BSG-WP-2020-032-v7.0.pdf

6 <https://github.com/OxCGRT/covid-policy-tracker/blob/master/documentation/codebook.md>

7 <https://www.texastribune.org/2021/03/25/conventions-texas-mask-mandate/>

8 <https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(21)00198-X/fulltext>

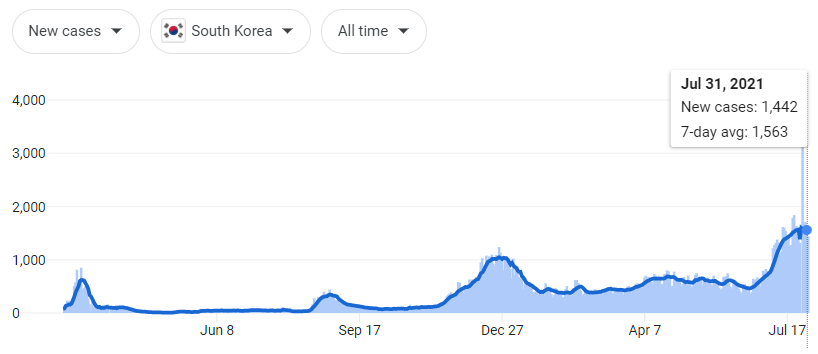
9 <https://statsiq.co1.qualtrics.com/public-dashboard/v0/dashboard/5f78e5d4de521a001036f78e#/dashboard/5f78e5d4de521a001036f78e?pageId=Page_f6071bf7-7db4-4a61-942f-ade4cce464de>

10 <https://www.cmaj.ca/content/193/17/E601>

11 <https://www.medrxiv.org/content/10.1101/2020.11.01.20222315v1>

12 <https://www.medrxiv.org/content/10.1101/2020.09.21.20196428v1>

Image A:



<https://www.google.com/search?q=south+korea+covid+cases&rlz=1C1CHBF_enUS850US886&oq=south+korea+covid+cases&aqs=chrome..69i57j0i131i433i457i512j0i512l8.4502j0j4&sourceid=chrome&ie=UTF-8>

Image B:

|  |  |
| --- | --- |
| **state** | **face\_covering\_scale** |
| Alabama | 3 |
| Alaska | 3 |
| Arizona | 3 |
| Arkansas | 3 |
| California | 3 |
| Colorado | 3 |
| Connecticut | 3 |
| Delaware | 4 |
| Florida | 3 |
| Georgia | 2 |
| Hawaii | 3 |
| Idaho | 3 |
| Illinois | 3 |
| Indiana | 2 |
| Iowa | 3 |
| Kansas | 3 |
| Kentucky | 3 |
| Louisiana | 3 |
| Maine | 3 |
| Maryland | 4 |
| Massachusetts | 4 |
| Michigan | 3 |
| Minnesota | 2 |
| Mississippi | 3 |
| Missouri | 3 |
| Montana | 3 |
| Nebraska | 2 |
| Nevada | 3 |
| New Hampshire | 3 |
| New Jersey | 3 |
| New Mexico | 3 |
| New York | 3 |
| North Carolina | 4 |
| North Dakota | 3 |
| Ohio | 3 |
| Oklahoma | 3 |
| Oregon | 3 |
| Pennsylvania | 3 |
| Rhode Island | 3 |
| South Carolina | 3 |
| South Dakota | 2 |
| Tennessee | 3 |
| Texas | 3 |
| Utah | 2 |
| Vermont | 3 |
| Virginia | 3 |
| Washington | 3 |
| Washington DC | 3 |
| West Virginia | 3 |
| Wisconsin | 3 |
| Wyoming | 2 |

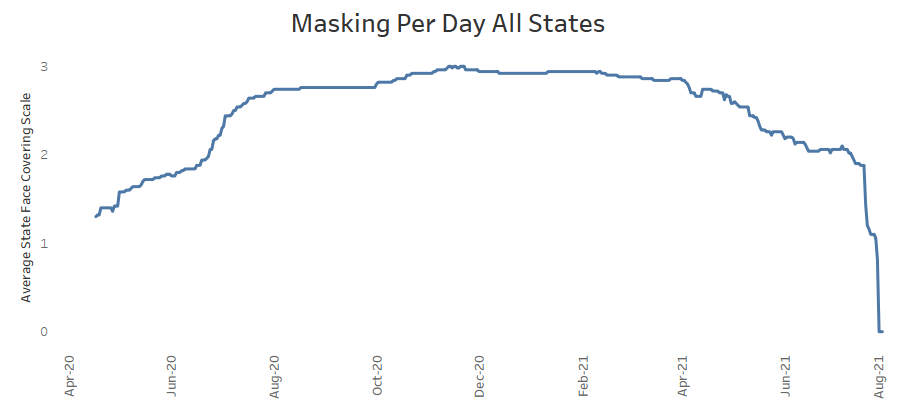
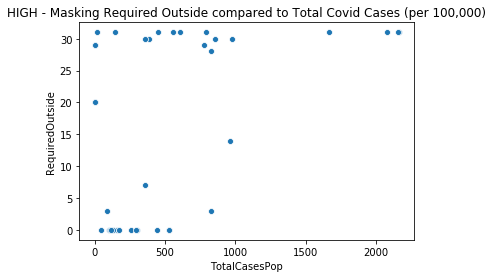
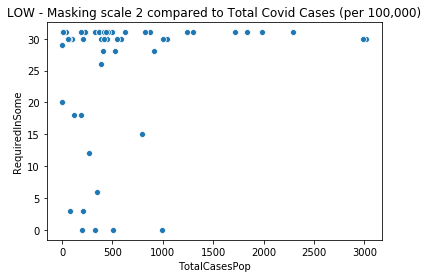
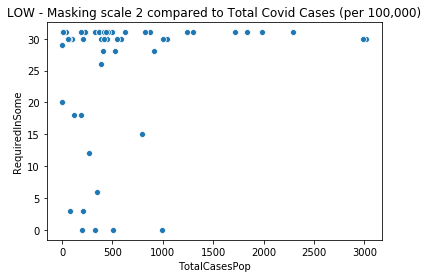
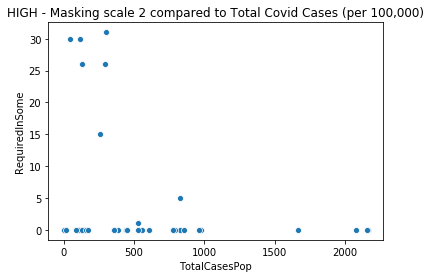


Image C:



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **state** | **YYYYMM** | **RequiredInSome** | **Sumface\_covering\_scale** | **TotalCasesPop** |
| Georgia | 202001 | 20 | 40 | 0 |
| Georgia | 202002 | 29 | 58 | 0 |
| Georgia | 202003 | 31 | 62 | 37.005213 |
| Georgia | 202004 | 3 | 33 | 210.36178 |
| Georgia | 202005 | 0 | 31 | 195.894993 |
| Georgia | 202006 | 0 | 30 | 322.375778 |
| Georgia | 202007 | 0 | 31 | 989.515064 |
| Georgia | 202008 | 15 | 46 | 792.273229 |
| Georgia | 202009 | 30 | 60 | 447.895878 |
| Georgia | 202010 | 31 | 62 | 402.771934 |
| Georgia | 202011 | 30 | 60 | 1043.31343 |
| Georgia | 202012 | 31 | 62 | 1835.5584 |
| Georgia | 202101 | 31 | 62 | 2288.62503 |
| Georgia | 202102 | 28 | 56 | 914.308491 |
| Georgia | 202103 | 31 | 62 | 499.433808 |
| Georgia | 202104 | 30 | 60 | 382.757662 |
| Georgia | 202105 | 31 | 62 | 222.784757 |
| Georgia | 202106 | 30 | 60 | 101.079141 |
| Georgia | 202107 | 26 | 52 | 388.71956 |
| Massachusetts | 202001 | 0 | 80 | 0.014509 |
| Massachusetts | 202002 | 0 | 116 | 0.014509 |
| Massachusetts | 202003 | 0 | 124 | 143.95351 |
| Massachusetts | 202004 | 5 | 72 | 826.158509 |
| Massachusetts | 202005 | 0 | 93 | 440.957371 |
| Massachusetts | 202006 | 0 | 90 | 104.171155 |
| Massachusetts | 202007 | 0 | 93 | 107.304995 |
| Massachusetts | 202008 | 0 | 93 | 133.217207 |
| Massachusetts | 202009 | 0 | 90 | 161.015527 |
| Massachusetts | 202010 | 0 | 123 | 383.895372 |
| Massachusetts | 202011 | 0 | 120 | 980.137404 |
| Massachusetts | 202012 | 0 | 124 | 2162.43649 |
| Massachusetts | 202101 | 0 | 124 | 2159.5493 |
| Massachusetts | 202102 | 0 | 112 | 828.770042 |
| Massachusetts | 202103 | 0 | 124 | 789.727621 |
| Massachusetts | 202104 | 0 | 119 | 774.653272 |
| Massachusetts | 202105 | 15 | 78 | 260.747075 |
| Massachusetts | 202106 | 30 | 60 | 43.830231 |
| Massachusetts | 202107 | 26 | 52 | 128.87916 |

Image D:

