Uncovering Mask Wearing in the United States

Rena Cohen and Chloe Shawah

December 10, 2020

Introduction

As the number of COVID-19 cases soars to unprecedented heights around the United States, public health experts and many political figures continue to emphasize mask wearing as one of the most effective ways to slow the spread of the pandemic. But, as a New York Times survey from July 2020 shows, mask wearing adherence varies widely in counties around the nation. What predictors might explain this variation in mask wearing, and how might public health officials use this information to develop more effective mask-wearing interventions? To what extent can mask wearing predict the spread of the virus on a county level? In this paper, we will address these questions by building several models to best predict mask-wearing behavior at the county level, assess the potential effects of state and county-wide mask mandates, and explore whether mask-wearing data from July has any relationship to current rates of COVID-19 (as of December 2020). Throughout this process, we will pay particular attention to the relationship political party and mask wearing, which a recent study by researchers at the University of Chicago claims to be "the single most important predictor of local mask use," a statement we would like to investigate further.

Research and Hypothesis

Before delving into our model building and exploratory processes, we made several hypothesis about what variables might be useful in predicting mask usage based on research. As aforementioned, we hypothesized that political party would be an important predictor of mask-wearing, as President Trump has consistently questioned public health guidance about wearing masks, sometimes rejecting it outright. Indeed, numerous surveys have already shown that Republicans wear masks at lower rates than Democrats (generally about a 20%). We hypothesized that this relationship would hold on a countywide level, with Republican counties being less likely to wear masks on average: perhaps it would be even more amplified than the relationship between partisanship and mask-wearing on an individual level, for one would assume that a Republican in a Democratic county would be more likely to wear a mask than a Republican in a Republican county, thus compounding these disparities in rate. While our data does not include responses on an individual level, this would be an interesting question for further research.

As well as political party affiliation within the county, we hypothesized that several other predictors could have some bearing on mask-wearing. Researchers at the National Institute of Health have suggested that age and location (i.e. rural vs. urban setting) affect mask wearing behavior, so we included the percentage of seniors in a county (hypothesized to have a positive association with mask-wearing since COVID-19 most severely affects the elderly) and various measures of population density in our mask-wearing model (hypothesized to be positive, as urban residents are more likely to come into close contact with many people in day-to-day life). Race could also be an important variable, as data from the Pew Research Center has shown differences in mask wearing frequency by race among individuals, and COVID-19 has is proportionately impacted communities of color according to the CDC; accordingly, we hypothesize that counties with higher proportions of minority residents would be more likely to wear masks. Finally, given that this same Pew Research Center survey showed differences in mask-wearing behavior by education level, we hypothesized that counties with a greater rates of college education would have higher rates of mask-wearing. Other demographic variables in our model included poverty (predicted to have a negative association with mask wearing) and the percentage of adults who are female (predicted to have a positive association with mask waring).

Along with these demographic characteristics, we hypothesized that legal and public health conditions would impact mask-wearing frequency: in particular, that counties with more cases and deaths would have greater rates of mask wearing, as residents would be more likely to fear getting the virus. However, the relationship between cause and effect is muddy here: does less mask wearing predict higher rates of the virus or do higher rates of the virus predict higher mask wearing rates? We will not be making causal claims in this study, but it will be interesting to examine this relationship nevertheless. Finally, we hypothesize that state and countywide mask mandates will increase the frequency of mask wearing, a claim we will devote ample time to further investigating.

Data Collection and Wrangling

Mask-wearing frequency, which is our y-variable for much of this study, is from the aforementioned *New York Times* survey, which was conducted by the survey firm Dynata on behalf of the Times from July 2 to July 14. Aggregated at the county level, it sorts 250,000 individual responses into 3,000 U.S. counties. The survey asked respondents how often they wore a mask (choices were always, frequently, sometimes, rarely, or never) and presents the percentage of people who gave each answer for every county, which we combined into a single weighted average representing the probability that a randomly selected person is wearing a mask in the county; for more information about the rationale behind this, see the appendix. It is important to remember that this data likely carries a good amount of uncertainty stemming both from the possible variation in the interpretation of the survey response choices by individuals and because only about 80 individuals per county were surveyed (not to mention possible undercoverage/non-response bias; we do not know how the sampling procedure, but the NYT does write htat " overall the large number of responses provide rough comparisons across many areas").

Our predictor variables, which included a combination of countywide demographic measurements and coronavirus-specific values, came from a variety of sources detailed in the table below. Whenever possible, we used the most recent data we could; notably, this meant that we chose to use the percentage of people who voted for Trump in 2020 as rather than in 2016 as a measure of of partisanship, as we figured this would more closely reflect the political party affiliations of county residents in July 2020 when the NYT mask-wearing survey was administered. Whenever applicable, we transformed raw counts into rates so that predictors would be comparable across counties with different total population sizes.

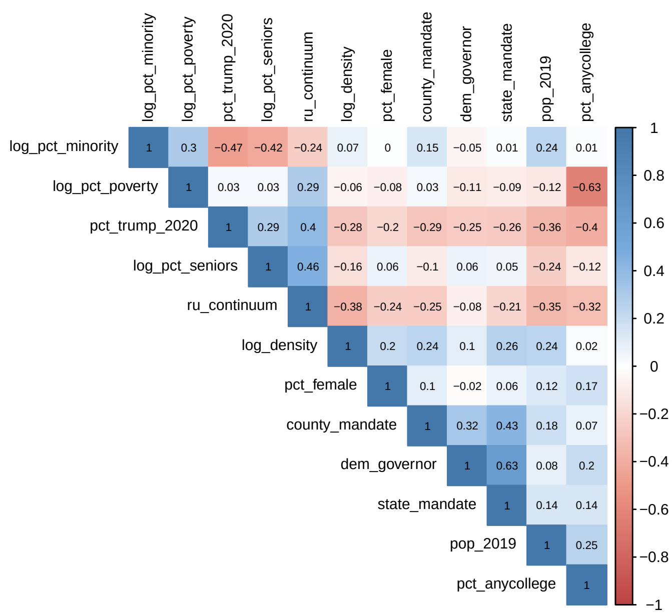
Compiling and cleaning this data was a (mostly) smooth process, thanks to the FIPS county code, which combines a two digit state code with a 3 digit county code to create a unique, standardized 5 digit code for every US county. When the merging proccess was over, we had 127 rows containing missing data, representing about 4% of the total number of counties in our dataset. The most concerning omission was the lack of political data for Alaska; after doing more research, we learned that Alaska treats election distributing in a different way than it treats FIPS county codes, a process that has led to unfortunate complications for many a data scientist. Because our class datasets for election results likewise did not include this data for Alaska at the county level, we simply carried forward our analysis without it, with the understanding that any results we came to should not be generalized to Alaska. Other than the large group of unused Alaskan counties, we did not think that removing the rest of the counties containing missing data would greatly impact our results. For more information about how we came to this conclusion, see the data cleaning section of our appendix.

Exploratory Data Analysis

After cleaning our data, our first step was to visualize our main response variable, mask-wearing score, to ensure that assumptions of our later models would be met. We found that the distribution of mask-wearing score to be approximately normal (see Appendix), which means that we will leave it untransformed for our later models. We then repeated this process for our predictors, log-transforming where needed. A final list of our predictors and their transformations is listed below:

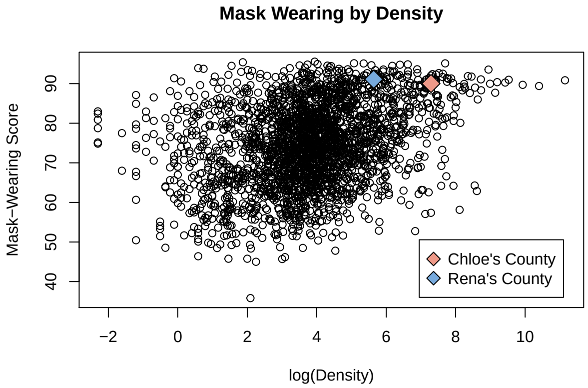
pop_2019:	population in 2019
log_density:	population density, log transformed
ru_continuum:	discrete score from 1 to 10 on the rural-Urban continuum with 1 being the most urban
log_pct_seniors:	percent of adults 65+ in 2019, log transformed
log_pct_minority:	percent of people from minority backgrounds in 2019, log-transformed
log_pct_poverty:	percent of people estimated to be living in poverty in 2018, log-transformed
pct_anycollege:	percent of adults who attended at least some college in 2018
pct_female:	percent of females in 2019
pct_trump2020:	percent of votes for Trump in 2020 election
dem_governor:	indicator of whether the county is in a state with a Democrat governor
state_mandate:	indicator of whether the county is in a state with a state-wide mask mandate
county_mandate:	indicator of whether the county has a county-wide mask mandate

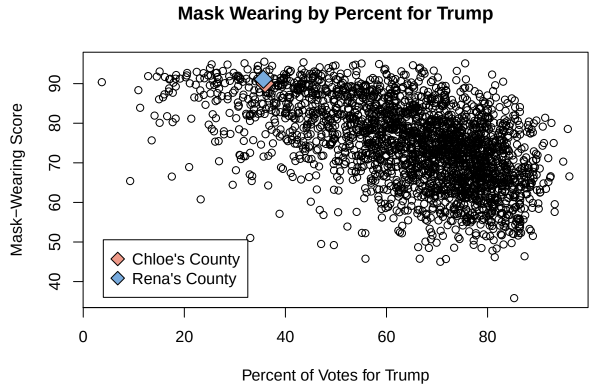
To ensure that none of these predictors were too colinear, we computed a correlation table to assess their correlations:



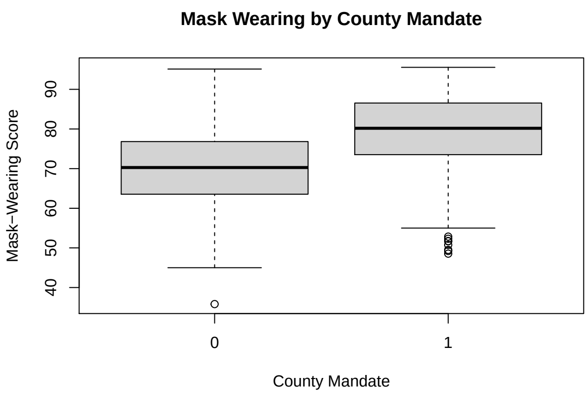
The strongest correlations we see between predictors is between percent poverty and percent college at -0.63 and between Democrat governor and state mandate at 0.63. Both of these associations do not surprise us, and are not so high that we are concerned about their impacts for our modeling. Another takeaway from this correlation table, however, is the extent to which political predictors like votes for Trump are correlated with seemingly apoltical predictors such as percent minority (-0.47). This will be important for us to keep in mind later in our analysis when we try to determine the effect of politics on mask-wearing behavior.

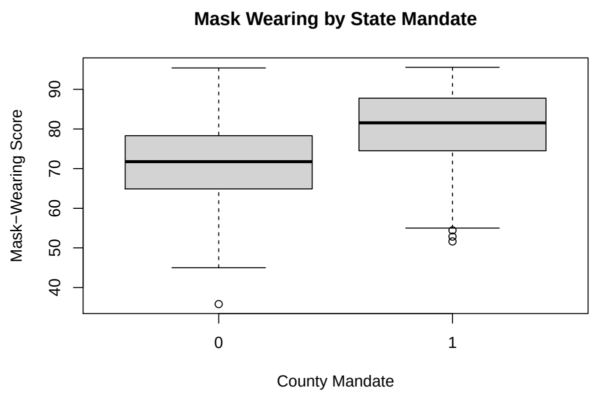
We then visualized the relationship between several of our predictors and mask-wearing score with scatterplots and boxplots. Two note-worthy plots were those for density and percent Trump votes.





It appears as through there might be a slight positive correlation between density and mask-wearing score, and a fairly strong negative correlation between percent Trump votes and mask-wearing. This shows us that both demographic characteristics and political characteristics of a county have associations with the county's mask-wearing behavior. (But it is also important to remember that density and percent Trump had a -0.28 correlation.) We then decided to visualize the association of mandates with mask-wearing compliance.





Based on the side-by-side boxplots above, we see that there appears to be a difference in mask-wearing behavior for counties with state-wide and county-wide mandates: a mandate looks like it is associated with an increase in mask-wearing. To determine if these differences are statistically significant, we ran two t-tests for a difference in means with $H_0 : \mu_1 = \mu_2$.

	No Mandate	Mandate	Difference	t-statistic	df	p-value
County-level Policy	70.29	79.38	-9.09	-24.52	2502.70	0.00
State-level Policy	71.34	80.26	-8.92	-23.08	1966.38	0.00

These results are unbelievably significant, so we reject the null hypothesis that there is no difference in the mask-wearing behavior in counties with and without a mandate. Yet county and state-level mandates are often a result of local politics and there has been writing, mentioned in our introduction, about how different Republicans are less likely to wear masks than Democrats. But we are curious about whether differences in mask-wearing behavior can be explained entirely by the county's political leaning, and if we can explain this difference using our demographic predictors such as density, percent minority, etc. This question motivates our next section, where will look more deeply into mask compliance by county through the lens of trying to decide if mask wearing is inherently political.

Modeling and Analysis

Is Mask Wearing Inherently Political?

To decide if mask-wearing behavior is inherently political, we will divide our predictors, removing state and county mandates, into two categories: political and apolitical. Even though these predictors can be correlated across categories as we saw in our correlation table, this does not undermine our goal of determining whether only demographic factors can explaining mask wearing more or less than only political factors.

political:	pct_trump_2020, dem_governor
apolitical:	pop_2019, log_pct_seniors, ru_continuum, log_pct_minority, pct_anycollege, log_density, pct_female, log_pct_poverty

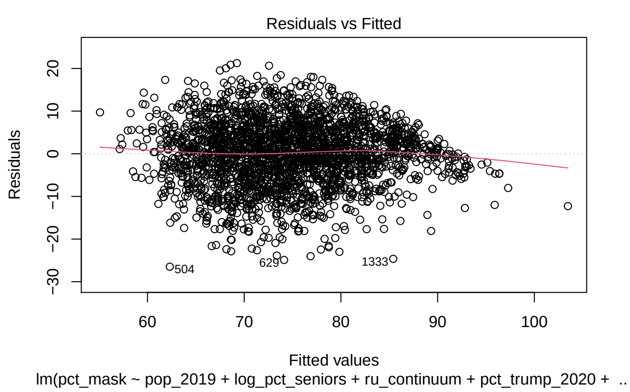
First, we will run 3 models with these two categories of predictors *combined* to predict mask-wearing score so we have an idea of our base ability explain mask-wearing behavior.

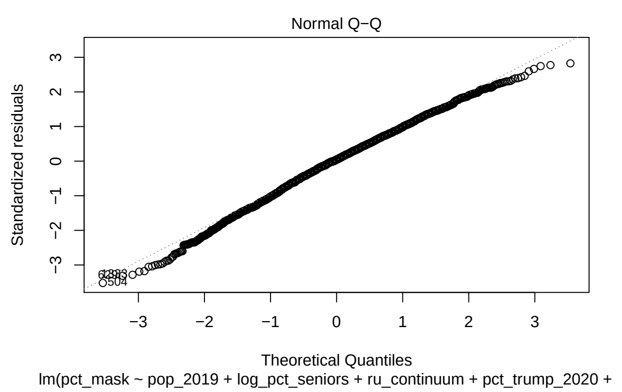
1) `linear full model`: an OLS model with single effects only

2) `linear interaction model`: an OLS model with single effects and all interaction terms

3) `linear selected model`: an OLS model using step-wise selection in both directions with an intercept-only model as an lower bound and starting point and the interaction model as an upper bound.

Included here are the assumptions checks for the first model `linear full model`. The normality and linearity assumption appear to be okay based on the residual and Q-Q plot, however we do see that the variance of residuals appears lower for very high $\hat{y}$, but does not seem severe enough to warrant a more robust method than OLS. The plots for the second two models are very similar to these plots for `linear full model`.





Now that we feel as though our assumptions are met, we will report our R^2 values for these 3 models. At the end of this section, we will include a full table with all models and their train and test RMSE.

Model Name	R^2
linear full model	0.4751
linear interaction model	0.5329
linear selected model	0.5309

This tells us that a few things: adding the interaction terms does improve the power of our model, and cutting our full interaction model with 55 predictors to 31 well-chosen predictors only decreases the R^2 of our model slightly. We compared the coefficients for the single effects in the `linear full model` and `linear selected model` and noticed that when we added the interaction terms, some of the coefficients flipped signs for the single effect: `log_pct_seniors`, `pct_trump_2020`, and `pct_female`. For `pct_trump_2020`, this coefficient was negative in `linear full model`, but flipped to positive in the `linear selected model` while almost all of the interaction terms involving `pct_trump_2020` were negative. This is still a little surprising: when we control for many interaction terms such as `pct_trump_2020:pct_anycollege`, an increase in Trump votes while holding all of these other interactions constant is associated with an increase in mask-wearing score. Very interesting! Perhaps this indicates that the relationship between politics and mask-wearing is a more complex than we had thought.

Next, we will fit two more models:

4) `linear political model`: an OLS model with just our political predictors and their interactions

5) `linear apolitical model`: an OLS model with just our apolitical predictors and their interactions

Model Name	R^2
linear political model	0.3222
linear apolitical model	0.4287

Based on these two R^2 scores, it seems like our apolitical model is better able to explain variance in mask-wearing behavior by county than using our two political predictors. Even though we only have two political predictors, this still helps us debunk the theory that mask wearing behavior is completely tied to support for President Trump, because we are able to explain just as much variability, if not more, using only our demographic variables.

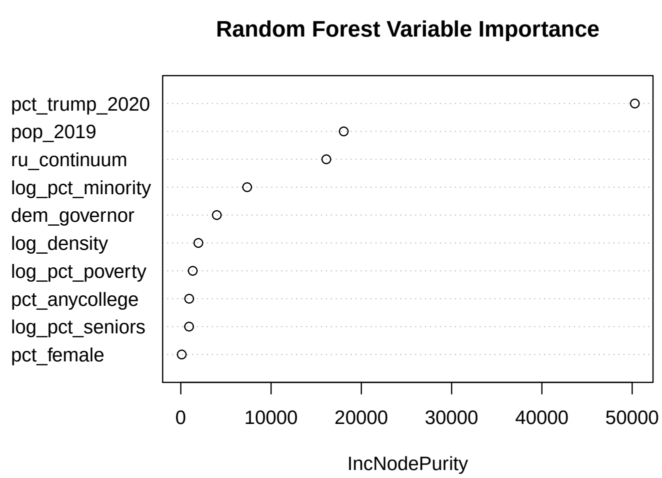
For our final models for this section, we will fit 2 random forest models and 1 decision tree that will have an added advantage of being able to catch nonlinear trends in our data. They will be:

6) `rf full`: a random forest with all predictors

7) `rf apolitical`: a random forest with only apolitical predictors

8) `dt political`: a decision tree with the two political predictors

Both random forests were fit with mtry = 6 and maxmodes = 10, while the decision tree was fit with maxdepth = 10. After fitting the `rf full` model, we checked the variable importance plot pictured below.



We found that `pct_trump_2020` was most often used to split the nodes. This does not contradict our early discussion of apolitical factors being able to explain more than political factors, but certainly adds some nuance: on its own, `pct_trump_2020` may be the most the most important predictor of mask-wearing score, but at least in the linear models, when we include lots of other, less individually important demographic predictors, we can get a model that is just as good if not better. We also noticed that pop_2019 is the second most important predictor according to this plot, even though it was deemed unimportant in some of the linear models above; this indicates to us that there may be a nonlinear relationship between pop_2019 and pct_mask.

Final RMSE Table:

Model Name	Train RMSE	Test RMSE
linear full model	7.437	7.953
linear interaction model	7.011	7.609
linear selected model	7.032	7.648
linear political model	8.451	8.634
linear apolitical model	7.745	8.120
rf full model	7.601	8.127
dt political model	8.497	8.683
rf apolitical model	7.959	8.465

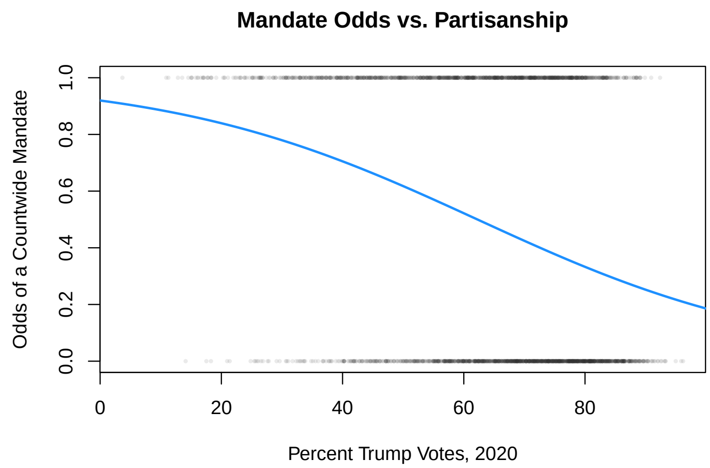
From this final RMSE table, we can make a few observations. The model that performed best on the test data was the linear model with all interaction terms. Some of our models appear to be a little bit overfit, but nothing drastic. Finally, and most importantly, in both the linear models and decision tree/random forest models, the apolitical models outperformed the political models on train and test data.

This tells us, tentatively, that we can explain more variability in mask-wearing behavior between counties with demographic predictors as we can with politcal predictors. Mask-wearing may not be inherently political.

Examining the Efficacy of Mask Mandates In Promoting Mask Wearing Behavior

Mask mandates at the county and state levels have been widely proposed as an effective and low-cost strategy for reducing the spread of COVID-19. This strategy assumes that wearing a mask truly does reduce the spread of COVID-19 (very reasonable given the preponderance of scientific evidence supporting this claim). It also makes a second assumption: that people will follow a mask mandate if it is in place. This assumption is more questionable, as there are reasons to believe that a mask mandate may have no effect on some individuals who are skeptical about masks (most of whom tend to be Republican, as discussed in this recent Pew Research Study) given that in many states, law enforcement is unwilling to enforce mask mandates. In this portion of the project, we wish to further investigate the relationship between mask mandates and mask wearing behavior. Is there evidence that counties with mandates have higher mask-wearing behavior after taking into account partisanship (which has become a proxy for attitudes towards mask wearing)? If so, might counties with different political leanings respond to county-wide mask mandates in different ways? While we cannot make causal claims with any of our data, we hope to identify patterns that may inform the often-contentious discussions about local mask mandates as a strategy for COVID-19 mitigation.

While counties with local mask mandates did have higher mask-wearing adherence on average (as shown in a two-sample t-test in our introduction), counties with a larger share of Republicans were also much less likely to have mask mandates in general. A logistic regression predicting the existence of a mask mandate in early July from the percent of the county who voted for Trump in the 2020 presidential election showed that partisanship is a very significant predictor of a county-wide mask mandate ($z = -13.87$, $p < 0.0001$). An increase of 1 percent in 2020 votes for Trump is associated with an odds ratio of 0.9615872, or a multiplicative increase of 2.615845. As shown in the plot below, this model meant that a county that was 80% Democratic had about an 82% chance of having a mask mandate, a county equally split between the counties had about a 57% chance of a mandate, and a county that was 80% Republican had about a 27% chance of a mandate.



Clearly, partisanship is a strong predictor of whether a county has a mask mandate in place in the first place (which makes sense given that leaders in local governments who make masking policies naturally represent the political beliefs of their constituents). Are the differences in what types of counties have mandates enough to explain away differences in mask wearing behavior? To address this question, we conducted an ESS F test to compare a linear model predicting mask wearing (lm_trump_quad) from just partisanship on the training set to a linear model with both partisanship and a mask mandate as a predictor (in both these cases, partisanship was fitted using a polynomial of degree 2, as this seemed to better account for regression assumptions and explain non-linearities in the data). We found that the addition of the mask mandate was statistically significant (F = 356.78, $p < 0.0001$): After controlling for partisanship, the predicted increase in the probability that a person in a county with a mandate was wearing a mask was 6.3918896.

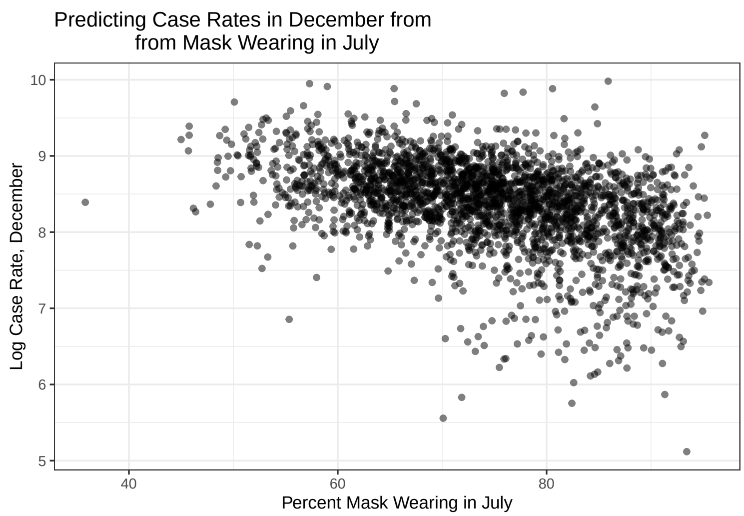
In order to consider whether the relationship between mandates and mask-wearing varied based on partisanship, we considered adding an interaction term between these two variables. The addition of this term was not significant in an ESS-F test ($F_2 = 1.5243$, p = 0.218), suggesting that the relationship between mask mandates and mask-wearing behavior does not vary based on partisanship.

One other question that arose was whether county-wide mandates still increased mask wearing when a statewide mandate was already in place. In order to account for this, we refit our previous model adding in a term for statewide mandate, and refit again with an interaction term between state and county mandates. In each of these models, the addition of the new terms added significant predictive power to the model as determined by an appropriate ESS F Test (F = 144.35, and F = 42.791 for each of the successive tests). Focusing more closely on the interaction term, the 95% bootstrap confidence interval for the predicted increase in mask-wearing for a county with a county-wide mandate when a state mandate is already in place is (0.5037052, 3.0058353) . This suggests that, while there is a slight benefit to having both a county and a state mask mandate in place, it is not nearly as large as the increase in mask wearing associated with going from having no mandates to a single one.

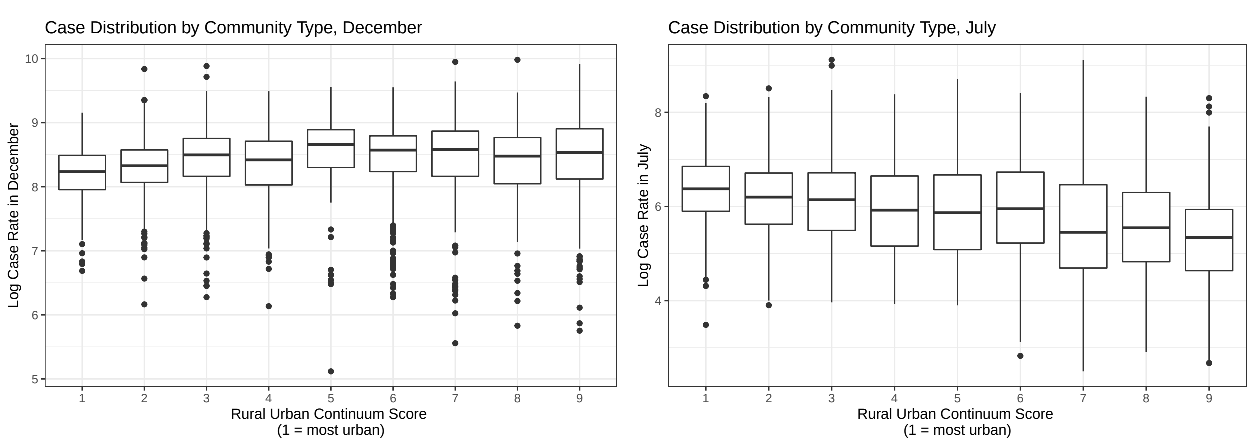
Future COVID-19 Spread and Mask Wearing Behavior

Thus far, we have concentrated our efforts on exploring what variables contribute to mask-wearing behavior as measured by the NYT survey, with underlying idea that mask-wearing behavior is a useful thing to know about because it directly impacts the spread of the pandemic. As a final exploration in our project, we wanted to put this claim to the test and investigate whether mask wearing behavior as measured by a survey in July could provide any predictive power in determining COVID-19 case rates in a county in December. While the link between masks and COVID-19 is causal (i.e. wearing a mask directly increases or reduces your chances of obtaining the virus), this is in no way a causal modelz, as behaviors may have changed greatly within the past 6 months and we are dealing with observational data to begin with. Rather, we intend to treat $mask_{pct}$ as a proxy for the "carefulness" of a county, as measured early in the pandemic.

As expected the current case rate is quite right-skewed; most counties have between 0 and 10,000 cases per 100,000 residents, but some counties have as many as 20,000 cases per 100,000 residents. For this reason, we log transformed case_rate_december, our y variable for this investigation. An initial plot of the transformed case rate variable against $mask_pct$ shows a weak negative association, albeit with many influential points, as there seemed to be several counties with very high mask-wearing scores that had an especially low case rate in December.



However, rather than fitting an OLS model straightaway, we decided to begin with a mixed effects model to predict cases by state using random intercepts. This is because different states have enacted different policies (i.e. shutdowns, prevalence/availability of tests, etc.) that could help explain differing numbers in current cases. We also considered adding density to this initial model (as one would hypothesize that more people coming into contact with one another more frequently would predict the spread of the virus), but decided not to, as it did not add any predictive power with the ESS F Test. As illustrated by the boxplots below, while urban counties had far higher case rates than rural counties back in July, now in December, the case rates have more or less evened out and population density is no longer a useful predictor of what counties will have higher rates of spread.



Based on this model, the average case rate among states is estimated to be 3,603.785 per 100,000 residents. Next, we fit another random intercepts model which included pct_mask as a fixed effect; in this model, on average, the an additional 1 percent in mask wearing within a county in July was associated with an increase of about 1 person (exp(-0.003855)) in the case rate per 100,000 people. This result, while quite small in magnitude, added significant explanatory power to the model ($\chi^2_1 = 11.393$, $p = 0.0007 < 0.05$); it also held when we calculated the RMSE for our two models on test data. This suggests that how careful a county was about wearing masks in July is an important predictor for case rates in December even after controlling for statewide differences, suggesting that fostering cultures of county-wide mask wearing should be a priority.

To conclude this section, we wanted to take a closer look at the counties that this mixed effects model had mis-predicted the most (i.e. the counties with the largest residuals). Counties that had lower rates than their mask wearing score would suggest tended to be small, rural communities in the Southwestern United States; while their success in controlling case rates could simply be due to chance given their small size (there are plenty of small towns that have been unsuccessful, so it could be due to random chance alone), it could be worthwhile to do further research on these counties in order to determine if they instituted any policy strategies at a local level that could be applicable to other scenarios. Likewise, further investigating the counties that comparatively did the worst (which tended to be midsize cities, spread out around the nation) given their mask wearing score could also be an interesting case study about what policy practices were ineffective.

Here are the counties with the most negative residuals (indicating COVID-19 rates below what their mask-wearing status would predict).

	county_name	state	pct_mask	pop_2019	ru_continuum	residuals	log_cases_dec
2029		AK	82.42	1592.00	9.00	-5.28	5.75
1461	King County	TX	70.30	272.00	9.00	-4.43	6.60
2229	San Juan County	WA	91.33	17582.00	9.00	-4.40	5.87
1149	Borden County	TX	78.55	654.00	8.00	-4.24	6.64
135	Sierra County	CA	76.10	3005.00	8.00	-4.18	6.34
1229	Daggett County	UT	55.35	950.00	9.00	-4.02	6.85

And here were the counties with the most positive residuals (indicating COVID-19 rates above what their mask-wearing status would predict).

	county_name	state	pct_mask	pop_2019	ru_continuum	residuals	log_cases_dec
1348	Lincoln County	AR	80.58	13024.00	3.00	3.52	9.88
2080	Rockland County	NY	89.97	325789.00	1.00	3.57	8.84
822	Chattahoochee County	GA	77.75	10907.00	2.00	3.70	9.84
262	Luce County	MI	84.62	6229.00	7.00	3.76	9.64
1676	Malheur County	OR	61.75	30571.00	6.00	4.00	9.01
226	Crowley County	CO	85.90	6061.00	8.00	4.96	9.98

Conclusions

Bibliography