Final Project Data, EDA, and Baseline

Chloe Shawah

11/20/2020

Data and Sources

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?

To address these questions, we plan to create two models: one to predict mask-wearing adherence by county based on a variety of county and state-wide predictors, and one to predict the spread of coronavirus in a county based on mask-wearing. Our data about mask-wearing (which is our outcome in the first model and a predictor in the second model) 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 (suggesting that a mixed effects model will likely be a useful approach). 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.

Our predictor variables were compiled from a variety of sources and joined with the mask wearing data by county FIPS code. We included gender, political party, education, and age statistics at the county level as all of these demographics have shown to differ in mask wearing frequency in prior surveys, with political party being especially significant. Other data, such as this poll from the Pew Research Center have suggested that mask wearing varies by race: this, combined with the fact that the pandemic has disproportionately impacted communities of color according to the CDC motivated us to include variables about the racial composition of counties in our baseline model. Researchers at the National Institute of Health have suggested that age and location (i.e. rural vs. urban setting) likewise affect mask wearing behavior, so we included the percentage of seniors in a county (since COVID-19 most severely affects the elderly) and various measures of population density in our mask-wearing model. Finally, we wanted to look beyond county demographics and determine whether coronavirus-related measures, including number of cases/deaths, growth rate of the virus at the time of the survey, and local/statewide mask mandates explained any of the variation in mask wearing by county.

For a complete list of the variables in our clean and compiled dataset and their sources, see below.

- `countyfp`
  - County level FIPS (Federal Information Processing System) code. Unique for each American county.
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `county_name`

- Name of the county

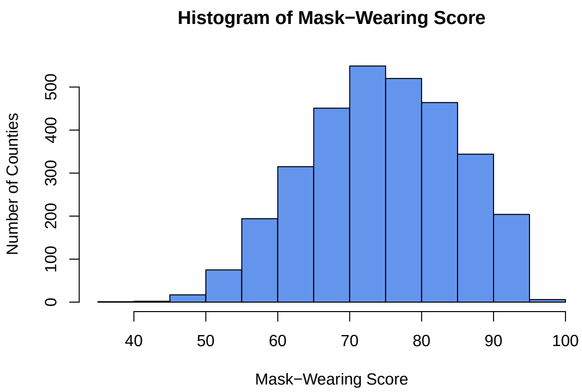
- `state`
  - state the county is located in
- `pct_mask`
  - An aggregate variable representing the probability that a randomly selected person in a county will wear a mask. Calculated by 1(always) + 0.75(frequently) + 0.5(sometimes) + 0.25(rarely) + 0(never)
- `always`
  - Percentage of people who answered they "always" wear a mask
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `frequently`
  - Percentage of peoplewho answered they frequently wear a mask
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `sometimes`
  - Percentage of people who answered they sometimes wear a mask
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `rarely`
  - Percentage of peoplewho answered they rarely wear a mask
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `never`
  - Percentage of people who answered they never wear a mask
  - New York Times
  - https://github.com/nytimes/covid-19-data/blob/master/mask-use/mask-use-by-county.csv
- `cases_02`
  - Number of COVID-19 cases on 07/02/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data
- `deaths_02`
  - Number of COVID-19 deaths on 07/02/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data
- `cases_14`
  - Number of COVID-19 cases on 07/14/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data
- `deaths_14`
  - Number of COVID-19 deaths on 07/14/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data

- `cases_27`
  - Number of COVID-19 cases on 07/27/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data
- `deaths_27`
  - Number of COVID-19 deaths on 07/27/2020
  - New York Times
  - https://github.com/nytimes/covid-19-data
- `case_growth_1`
  - cases_14/cases_02
- `case_growth_2`
  - cases_27/cases_14
- `pop_2019`
  - Population estimate in 2019
  - United States Census Bureau
  - https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html
- `ru_continuum`
  - 1 to 10 rating on the Rural-Urban Continuum
  - United States Census Bureau
  - https://www.census.gov/newsroom/press-kits/2019/national-state-estimates.html
- `density`
  - Population density of the county-
  - county_level_election.csv from class
- `pct_less_than_hs`
  - Percent of adults with less than a high school diploma,
  - 2014-18 & 2014-18 American Community Survey
  - https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- `pct_hs`
  - Percent of adults with a high school diploma only
  - 2014-18 & 2014-18 American Community Survey
  - https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- `pct_some_college`
  - Percent of adults completing some college or associate's degree
  - 2014-18 & 2014-18 American Community Survey
  - https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- `pct_college`
  - Percent of adults with a bachelor's degree or higher
  - 2014-18 & 2014-18 American Community Survey
  - https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- `pct_poverty`
  - Percentage of people estimated to be living in poverty in 2018
  - U.S. Census Bureau, Small Area Income and Poverty Estimates (SAIPE) Program
  - https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/

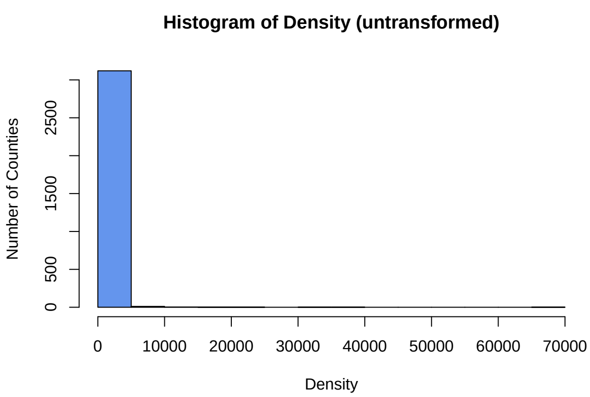
- `pct_female`
  - Percentage of females in county, 2019
  - U.S. Census Bureau
  - https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html
- `pct_black`
  - Percentage of Black/African-American residents in county, 2019
  - U.S. Census Bureau
  - https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html
- `pct_native`
  - Percentage of American Indian or Alaskan Native people in county, 2019
  - U.S. Census Bureau
  - https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html
- `pct_hispanic`
  - Percentage of Hispanic people in county, 2019
  - U.S. Census Bureau
  - https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html
- `pct_seniors`
  - Percentage of adults 65 or over in county, 2019
  - U.S. Census Bureau
  - https://www.census.gov/newsroom/press-kits/2020/population-estimates-detailed.html
- `pct_trump_2016`
  - Percentage of county who voted for Donald Trump in 2016
  - county_level_election.csv from class
- `pct_trump_2020`
  - Percentage of county who voted for Donald Trump in 2020
  - Scraped by GitHub user tonmcg from Fox News, Politico, and New York Times
  - https://github.com/tonmcg/US_County_Level_Election_Results_08-20
- `dem_governor`
  - Dummy variable coded 1 if the state has a Democratic governor
  - National Governor's Association
  - https://www.nga.org/wp-content/uploads/2019/07/Governors-Roster.pdf
- `state_mandate`
  - Dummy variable coded 1 if a statewide mask mandate was enacted before 07/14/2020
  - Axios
  - https://www.axios.com/states-face-coverings-mandatory-a0e2fe35-5b7b-458e-9d28-3f6cdb1032fb.html
- `county_mandate`
  - Dummy variable coded 1 if there was a county-wide mask mandate enacted before -7/14/2020
  - Harris Institute of Public Policy
  - https://www.austinlwright.com/covid-research

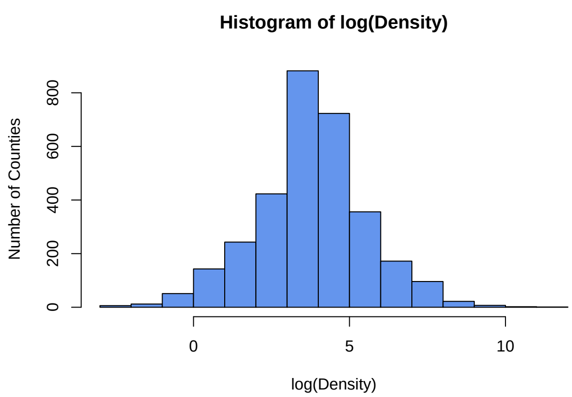
Exploratory Data Analysis

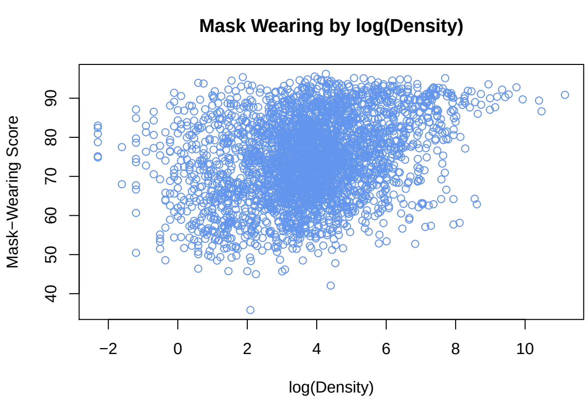
First, we wanted to make sure that our response variable `pct_mask` is distributed approximately normally. Based on this following histogram, the distribution of our response is roughly normal so we should not have to transform it.



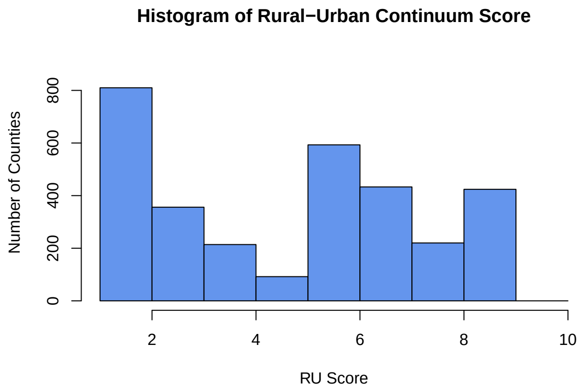
Next, we wanted to make sure that our predictors were roughly normally distributed. We see that for county population density, we have a heavy right skew, which is fixed with a log transformation. Based on the scatterplot, there does appear to be a positive correlation between population density and mask wearing percentage.

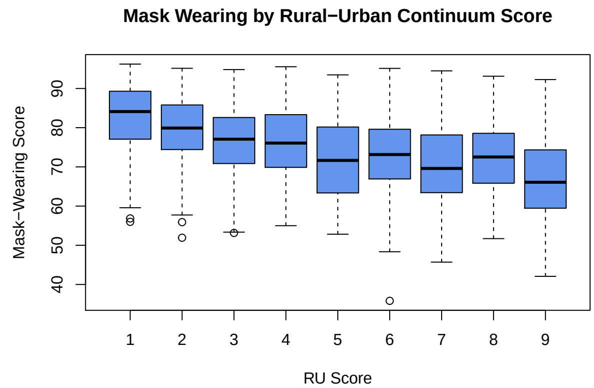




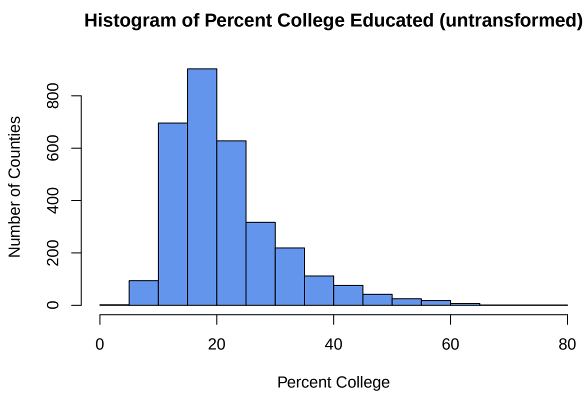


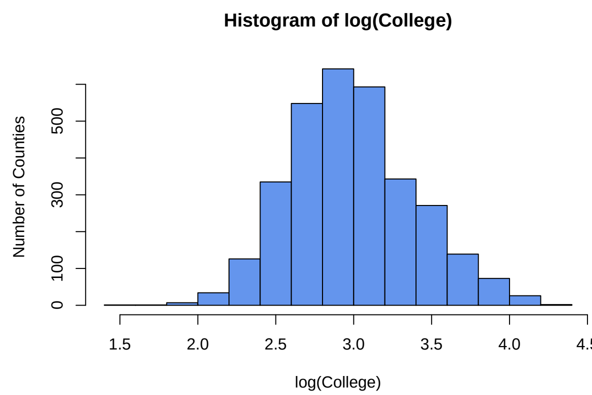
Next, we investigated the rural-urban continuum score. Its distribution is a little funky, but based on the side-by-side boxplots, it appears to have a negative correlation with mask-wearing. More rural areas appear to have lower mask-wearing percentages.

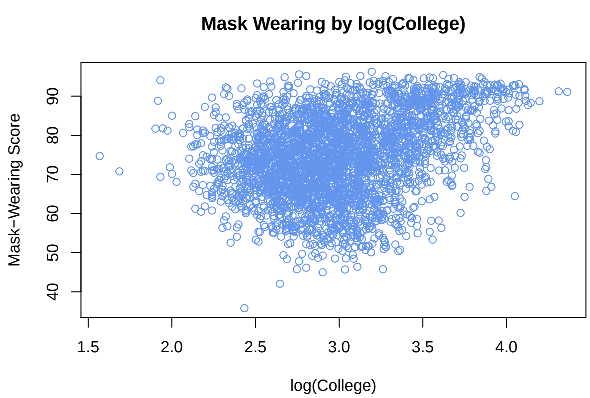




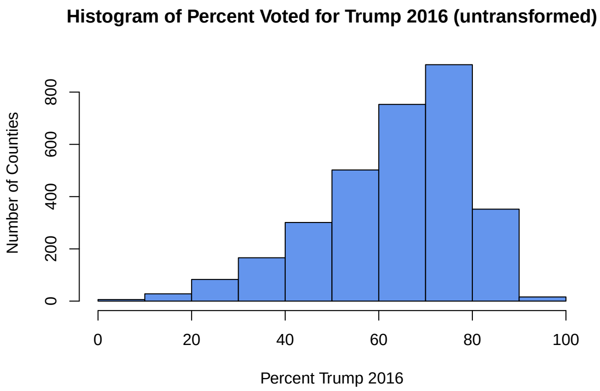
For the percent of the county adults who are college educated, we see a right skew that is easily fixed with a log transformation. Based on the scatterplot, there does appear to be a sliight positive correlation between percent college educated and mask-wearing.

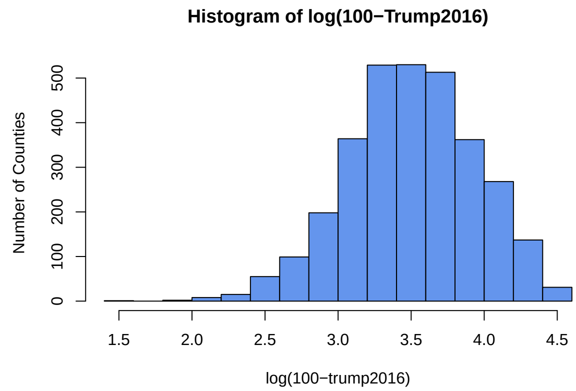


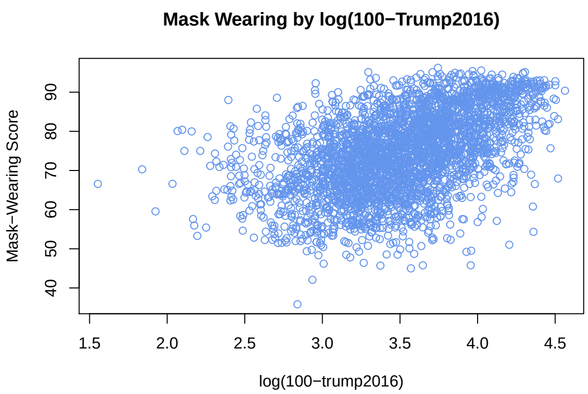




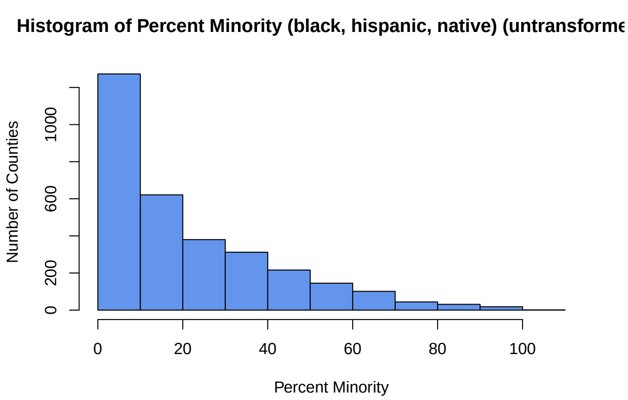
For the percent of voters who voted for Trump in 2016, we see a left skew. We fixed this by taking the log of 100-`trump2016`, which then becomes a measure of how many people did not vote for Trump. There appears to be a fairly strong correlation between not voting for Trump and mask usage.

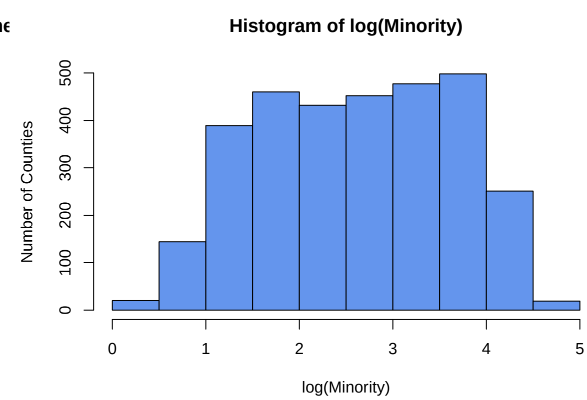


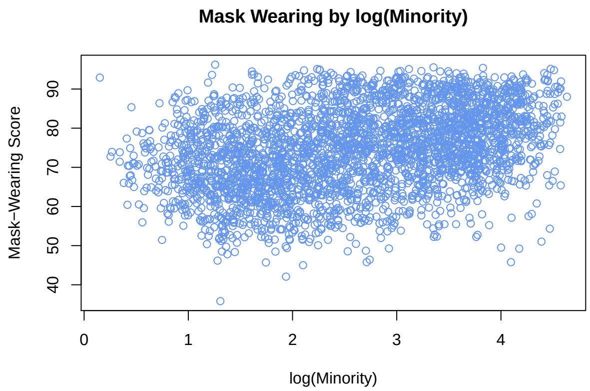




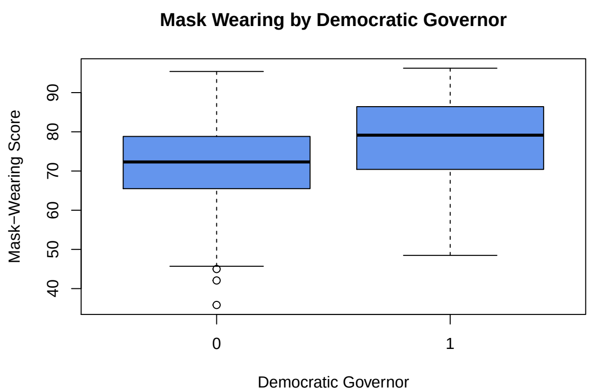
To investigate the effect of percent minority on mask wearing in a county, we added `pct_black`, `pct_native`, and `pct_hispanic` together, which appeared to have a right skew that was more or less fixed with a log transformation. Based on the scatterplot there does not seem to be a clear correlation between percent minority and mask usage, though in our model we will treat racial/ethnic groups separately.



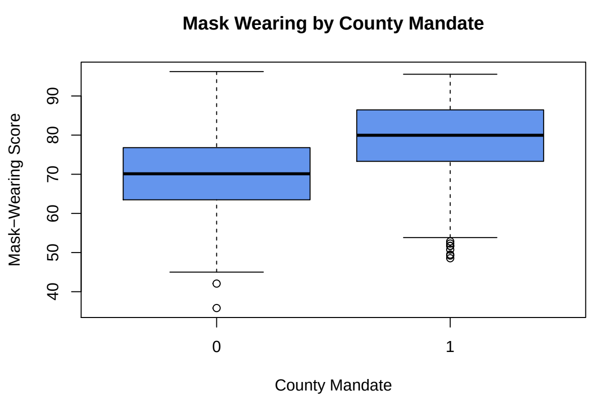


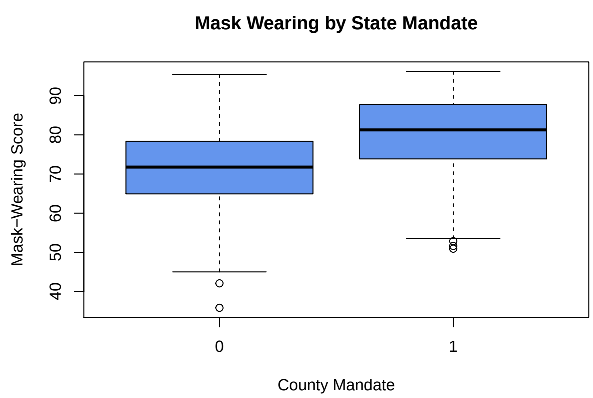


This boxplot of mask usages by whether or not a county is in a state with a Democratic governor suggests that there is higher usage in states with Democratic governors.



The following boxplots by state and county also suggest that the presence of a mandate has a positive effect on mask usage. We can follow this up with t-tests later to see if this is statistically significant.





Other variables that needed to be log transformed were `pct_seniors`, `pct_poverty`, and all individual race/ethnicity categories. `pct_hs` did not need to be log transformed, and `pct_female` looked skewed both with and without a transformation, so we left it untransformed.

We also have some missing variables in our dataset which we will have to figure out how to impute. The variables misisng the most values are case and death counts which we do not want to ignore because their missingness may not be random. We are also missing the percent of Trump votes for a decent number of counties, per the output below.

```
sapply(clean_data_2, function(x) sum(is.na(x)))
```

```
##         countyfp      county_name            state         pct_mask
##                0               30                0                0
##           always       frequently        sometimes           rarely
##                0                0                0                0
##            never         cases_02        deaths_02         cases_14
##                0               97               97               59
##        deaths_14         cases_27        deaths_27     case_growth_1
##               59               42               42               97
##    case_growth_2         pop_2019     ru_continuum          density
##               59                0                0                3
## pct_less_than_hs           pct_hs pct_some_college      pct_college
##                0                0                0                0
##      pct_poverty       pct_female        pct_black       pct_native
##                1                0                0                0
##     pct_hispanic        pct_asian      pct_seniors   pct_trump_2016
##                0                0                0               30
##   pct_trump_2020     dem_governor    state_mandate   county_mandate
##               32                0                0               10
```

Baseline Models

In order to run a linear model, we had to change some of the transformations to `log(1+X)` so avoid taking the log of 0. Specifically, we had to do this for all 4 race/ethnicity variables and `pct_college`. We left out two of the education categories to `pct_less_than_hs` and `pct_some_college` to avoid multicolinearity, but moving forward, it might be best to create a variables that sums `pct_college` and `pct_some_college`. Finally, we removed `pct_trump2020` from the model because the multicolinearity between that and the 2016 percent was inflating the standard errors.

We ran two baseline models:

1.) A basic linear regression model with all predictors besides those mentioned in the paragraph above
2.) An sequential variable selection model using AIC starting with the first model and with a lower bound of an intercept-only model and an upper bound of a model with predictors and their two-way interactions.

The summaries of these two models can be found below.

```
interceptmodel = lm(pct_mask ~ 1, data = clean_data_2)

fullmodel = lm(pct_mask ~ ru_continuum + log(density) + pct_hs + log(1+pct_college) +
               log(pct_poverty) + pct_female + log(1+pct_black) + log(1+pct_native) +
               log(1+pct_hispanic) + log(1+pct_asian) + log(pct_seniors) +
               log(100-pct_trump_2016) + dem_governor + state_mandate + county_mandate,
               data = clean_data_2)

summary(fullmodel)
```

```
##
## Call:
## lm(formula = pct_mask ~ ru_continuum + log(density) + pct_hs +
```

```
##     log(1 + pct_college) + log(pct_poverty) + pct_female + log(1 +
##     pct_black) + log(1 + pct_native) + log(1 + pct_hispanic) +
##     log(1 + pct_asian) + log(pct_seniors) + log(100 - pct_trump_2016) +
##     dem_governor + state_mandate + county_mandate, data = clean_data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.3273  -4.6977   0.3866   4.7483  22.9209
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)               34.90981    4.72162   7.394 1.83e-13 ***
## ru_continuum              -0.90107    0.06640 -13.571  < 2e-16 ***
## log(density)               0.50240    0.08751   5.741 1.03e-08 ***
## pct_hs                    -0.06883    0.03253  -2.116 0.034436 *
## log(1 + pct_college)       0.94914    0.66535   1.427 0.153819
## log(pct_poverty)           0.40250    0.47179   0.853 0.393651
## pct_female                -0.23172    0.06273  -3.694 0.000225 ***
## log(1 + pct_black)         1.18468    0.15365   7.710 1.68e-14 ***
## log(1 + pct_native)       -1.10148    0.23208  -4.746 2.17e-06 ***
## log(1 + pct_hispanic)      3.50514    0.19061  18.389  < 2e-16 ***
## log(1 + pct_asian)        -0.38162    0.38790  -0.984 0.325285
## log(pct_seniors)           7.65386    0.67580  11.326  < 2e-16 ***
## log(100 - pct_trump_2016)  5.29667    0.38661  13.700  < 2e-16 ***
## dem_governor               0.69969    0.34406   2.034 0.042073 *
## state_mandate              3.45866    0.36834   9.390  < 2e-16 ***
## county_mandate             3.81805    0.29378  12.996  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.088 on 3093 degrees of freedom
##   (33 observations deleted due to missingness)
## Multiple R-squared:  0.5335, Adjusted R-squared:  0.5312
## F-statistic: 235.8 on 15 and 3093 DF,  p-value: < 2.2e-16
```

```r
interactionmodel = lm(pct_mask ~ (ru_continuum + log(density) + pct_hs + log(1+pct_college) +
                      log(pct_poverty) + pct_female + log(1+pct_black) + log(1+pct_native) +
                      log(1+pct_hispanic) + log(1+pct_asian) + log(pct_seniors) +
                      log(100-pct_trump_2016) + dem_governor +
                      state_mandate + county_mandate)^2,
                      data = clean_data_2)

selected_model = step(fullmodel, scope = list(lower = formula(interceptmodel),
                                              upper = formula(interactionmodel)),
                                              direction = "both", trace = 0)

summary(selected_model)
```

```
##
## Call:
## lm(formula = pct_mask ~ ru_continuum + log(density) + pct_hs +
##     log(1 + pct_college) + log(pct_poverty) + pct_female + log(1 +
##     pct_black) + log(1 + pct_native) + log(1 + pct_hispanic) +
##     log(1 + pct_asian) + log(pct_seniors) + log(100 - pct_trump_2016) +
```

```
##     dem_governor + state_mandate + county_mandate + log(1 + pct_college):log(100 -
##     pct_trump_2016) + dem_governor:state_mandate + log(100 -
##     pct_trump_2016):state_mandate + log(density):log(1 + pct_black) +
##     log(pct_poverty):log(1 + pct_native) + ru_continuum:county_mandate +
##     log(1 + pct_hispanic):county_mandate + log(1 + pct_native):log(1 +
##     pct_hispanic) + log(1 + pct_hispanic):state_mandate + ru_continuum:log(1 +
##     pct_hispanic) + ru_continuum:log(1 + pct_native) + log(1 +
##     pct_native):state_mandate + log(1 + pct_native):county_mandate +
##     log(pct_poverty):state_mandate + log(100 - pct_trump_2016):dem_governor +
##     log(pct_poverty):log(100 - pct_trump_2016) + log(1 + pct_college):log(pct_poverty) +
##     log(pct_poverty):log(1 + pct_asian) + log(pct_poverty):log(1 +
##     pct_black) + log(1 + pct_black):dem_governor + log(1 + pct_black):log(1 +
##     pct_native) + log(1 + pct_black):log(1 + pct_hispanic) +
##     log(density):log(100 - pct_trump_2016) + log(100 - pct_trump_2016):county_mandate +
##     pct_hs:state_mandate + pct_hs:log(pct_poverty) + log(1 +
##     pct_college):log(1 + pct_hispanic) + pct_hs:log(1 + pct_hispanic) +
##     pct_hs:log(1 + pct_native) + pct_female:log(100 - pct_trump_2016) +
##     log(1 + pct_hispanic):log(100 - pct_trump_2016) + log(1 +
##     pct_hispanic):log(pct_seniors) + log(pct_seniors):state_mandate +
##     log(pct_seniors):dem_governor + log(1 + pct_asian):state_mandate +
##     log(1 + pct_asian):dem_governor + log(density):log(1 + pct_asian) +
##     log(1 + pct_college):log(1 + pct_native) + log(density):log(pct_seniors) +
##     log(1 + pct_native):dem_governor + pct_female:log(1 + pct_hispanic) +
##     log(1 + pct_asian):log(100 - pct_trump_2016) + log(1 + pct_college):log(1 +
##     pct_black) + log(1 + pct_native):log(100 - pct_trump_2016) +
##     pct_female:dem_governor + pct_female:county_mandate + ru_continuum:log(1 +
##     pct_college) + log(1 + pct_college):log(pct_seniors) + ru_continuum:log(pct_seniors) +
##     ru_continuum:log(1 + pct_black), data = clean_data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.7951  -3.9996   0.1689   3.9072  21.3930
##
## Coefficients:
##                                                 Estimate Std. Error t value
## (Intercept)                                     18.34886   31.99446   0.574
## ru_continuum                                    -0.81398    0.87554  -0.930
## log(density)                                     0.69911    1.24929   0.560
## pct_hs                                           0.81423    0.18838   4.322
## log(1 + pct_college)                           -10.10513    5.99143  -1.687
## log(pct_poverty)                                18.75093    5.04423   3.717
## pct_female                                       1.46358    0.47150   3.104
## log(1 + pct_black)                              11.42189    2.21334   5.160
## log(1 + pct_native)                            -28.67218    6.17911  -4.640
## log(1 + pct_hispanic)                           25.17130    4.10570   6.131
## log(1 + pct_asian)                             -11.44573    3.65069  -3.135
## log(pct_seniors)                                -8.67769    6.05802  -1.432
## log(100 - pct_trump_2016)                      -19.00745    7.79552  -2.438
## dem_governor                                    30.03318    8.04732   3.732
## state_mandate                                  -49.55233    7.33535  -6.755
## county_mandate                                  -6.01635    6.47998  -0.928
## log(1 + pct_college):log(100 - pct_trump_2016)   8.92740    1.05656   8.449
## dem_governor:state_mandate                       5.45334    0.84301   6.469
## log(100 - pct_trump_2016):state_mandate          4.99680    1.11595   4.478
```

```
## log(density):log(1 + pct_black)                    -0.30630    0.08668  -3.534
## log(pct_poverty):log(1 + pct_native)                1.06575    0.70481   1.512
## ru_continuum:county_mandate                         0.50297    0.10762   4.673
## log(1 + pct_hispanic):county_mandate                1.72356    0.33798   5.100
## log(1 + pct_native):log(1 + pct_hispanic)           1.62402    0.32916   4.934
## log(1 + pct_hispanic):state_mandate                -1.83650    0.41038  -4.475
## ru_continuum:log(1 + pct_hispanic)                 -0.48169    0.07649  -6.298
## ru_continuum:log(1 + pct_native)                    0.44964    0.09902   4.541
## log(1 + pct_native):state_mandate                   3.83623    0.70063   5.475
## log(1 + pct_native):county_mandate                 -0.78777    0.50184  -1.570
## log(pct_poverty):state_mandate                     -3.60118    0.81982  -4.393
## log(100 - pct_trump_2016):dem_governor              3.71951    1.03194   3.604
## log(pct_poverty):log(100 - pct_trump_2016)          4.99395    1.08216   4.615
## log(1 + pct_college):log(pct_poverty)              -8.58194    1.08570  -7.905
## log(pct_poverty):log(1 + pct_asian)                 2.07213    0.81440   2.544
## log(pct_poverty):log(1 + pct_black)                -2.13887    0.43096  -4.963
## log(1 + pct_black):dem_governor                    -1.74035    0.29333  -5.933
## log(1 + pct_black):log(1 + pct_native)              0.48413    0.27261   1.776
## log(1 + pct_black):log(1 + pct_hispanic)           -0.40446    0.18981  -2.131
## log(density):log(100 - pct_trump_2016)              0.67020    0.21222   3.158
## log(100 - pct_trump_2016):county_mandate           -2.37426    0.69825  -3.400
## pct_hs:state_mandate                                0.31853    0.05080   6.270
## pct_hs:log(pct_poverty)                            -0.24984    0.06434  -3.883
## log(1 + pct_college):log(1 + pct_hispanic)         -2.51285    0.58737  -4.278
## pct_hs:log(1 + pct_hispanic)                       -0.17971    0.03138  -5.727
## pct_hs:log(1 + pct_native)                          0.21907    0.05332   4.109
## pct_female:log(100 - pct_trump_2016)               -0.37610    0.12997  -2.894
## log(1 + pct_hispanic):log(100 - pct_trump_2016)  -1.55350    0.39341  -3.949
## log(1 + pct_hispanic):log(pct_seniors)              1.98960    0.69833   2.849
## log(pct_seniors):state_mandate                      9.42277    1.78199   5.288
## log(pct_seniors):dem_governor                      -7.72912    1.66843  -4.633
## log(1 + pct_asian):state_mandate                    5.37235    0.98891   5.433
## log(1 + pct_asian):dem_governor                    -4.21206    0.85556  -4.923
## log(density):log(1 + pct_asian)                    -0.52821    0.15838  -3.335
## log(1 + pct_college):log(1 + pct_native)            2.24891    1.09269   2.058
## log(density):log(pct_seniors)                      -0.64315    0.33113  -1.942
## log(1 + pct_native):dem_governor                   -1.91571    0.62729  -3.054
## pct_female:log(1 + pct_hispanic)                   -0.14566    0.05927  -2.457
## log(1 + pct_asian):log(100 - pct_trump_2016)        1.70251    0.79676   2.137
## log(1 + pct_college):log(1 + pct_black)            -0.82113    0.45883  -1.790
## log(1 + pct_native):log(100 - pct_trump_2016)       1.19192    0.65187   1.828
## pct_female:dem_governor                            -0.27975    0.11785  -2.374
## pct_female:county_mandate                           0.23998    0.11766   2.040
## ru_continuum:log(1 + pct_college)                  -0.49650    0.16216  -3.062
## log(1 + pct_college):log(pct_seniors)               3.81417    1.61750   2.358
## ru_continuum:log(pct_seniors)                       0.54872    0.23498   2.335
## ru_continuum:log(1 + pct_black)                     0.13299    0.05294   2.512
##                                                  Pr(>|t|)
## (Intercept)                                      0.566348
## ru_continuum                                     0.352608
## log(density)                                     0.575792
## pct_hs                                           1.59e-05 ***
## log(1 + pct_college)                             0.091783 .
## log(pct_poverty)                                 0.000205 ***
```

```
## pct_female                                         0.001926 **
## log(1 + pct_black)                                 2.62e-07 ***
## log(1 + pct_native)                                3.63e-06 ***
## log(1 + pct_hispanic)                              9.87e-10 ***
## log(1 + pct_asian)                                 0.001734 **
## log(pct_seniors)                                   0.152123
## log(100 - pct_trump_2016)                          0.014815 *
## dem_governor                                       0.000193 ***
## state_mandate                                      1.70e-11 ***
## county_mandate                                     0.353247
## log(1 + pct_college):log(100 - pct_trump_2016)      < 2e-16 ***
## dem_governor:state_mandate                         1.15e-10 ***
## log(100 - pct_trump_2016):state_mandate            7.83e-06 ***
## log(density):log(1 + pct_black)                    0.000416 ***
## log(pct_poverty):log(1 + pct_native)               0.130607
## ru_continuum:county_mandate                        3.09e-06 ***
## log(1 + pct_hispanic):county_mandate               3.61e-07 ***
## log(1 + pct_native):log(1 + pct_hispanic)          8.50e-07 ***
## log(1 + pct_hispanic):state_mandate                7.92e-06 ***
## ru_continuum:log(1 + pct_hispanic)                 3.45e-10 ***
## ru_continuum:log(1 + pct_native)                   5.82e-06 ***
## log(1 + pct_native):state_mandate                  4.72e-08 ***
## log(1 + pct_native):county_mandate                 0.116570
## log(pct_poverty):state_mandate                     1.16e-05 ***
## log(100 - pct_trump_2016):dem_governor             0.000318 ***
## log(pct_poverty):log(100 - pct_trump_2016)         4.10e-06 ***
## log(1 + pct_college):log(pct_poverty)              3.73e-15 ***
## log(pct_poverty):log(1 + pct_asian)                0.010997 *
## log(pct_poverty):log(1 + pct_black)                7.32e-07 ***
## log(1 + pct_black):dem_governor                    3.31e-09 ***
## log(1 + pct_black):log(1 + pct_native)             0.075841 .
## log(1 + pct_black):log(1 + pct_hispanic)           0.033183 *
## log(density):log(100 - pct_trump_2016)             0.001603 **
## log(100 - pct_trump_2016):county_mandate           0.000682 ***
## pct_hs:state_mandate                               4.11e-10 ***
## pct_hs:log(pct_poverty)                            0.000105 ***
## log(1 + pct_college):log(1 + pct_hispanic)         1.94e-05 ***
## pct_hs:log(1 + pct_hispanic)                       1.12e-08 ***
## pct_hs:log(1 + pct_native)                         4.08e-05 ***
## pct_female:log(100 - pct_trump_2016)               0.003833 **
## log(1 + pct_hispanic):log(100 - pct_trump_2016) 8.03e-05 ***
## log(1 + pct_hispanic):log(pct_seniors)             0.004414 **
## log(pct_seniors):state_mandate                     1.33e-07 ***
## log(pct_seniors):dem_governor                      3.76e-06 ***
## log(1 + pct_asian):state_mandate                   5.99e-08 ***
## log(1 + pct_asian):dem_governor                    8.97e-07 ***
## log(density):log(1 + pct_asian)                    0.000863 ***
## log(1 + pct_college):log(1 + pct_native)           0.039661 *
## log(density):log(pct_seniors)                      0.052191 .
## log(1 + pct_native):dem_governor                   0.002278 **
## pct_female:log(1 + pct_hispanic)                   0.014050 *
## log(1 + pct_asian):log(100 - pct_trump_2016)       0.032695 *
## log(1 + pct_college):log(1 + pct_black)            0.073617 .
## log(1 + pct_native):log(100 - pct_trump_2016)      0.067576 .
```

```
## pct_female:dem_governor                          0.017665 *
## pct_female:county_mandate                        0.041479 *
## ru_continuum:log(1 + pct_college)                0.002219 **
## log(1 + pct_college):log(pct_seniors)            0.018433 *
## ru_continuum:log(pct_seniors)                    0.019598 *
## ru_continuum:log(1 + pct_black)                  0.012053 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.309 on 3043 degrees of freedom
##   (33 observations deleted due to missingness)
## Multiple R-squared:  0.6364, Adjusted R-squared:  0.6286
## F-statistic: 81.94 on 65 and 3043 DF,  p-value: < 2.2e-16
```

First Model: Our first model deemed all of our predictors significant except `log(1 + pct_college)`, `log(pct_poverty)`, and `log(1 + pct_asian)`. This model has an R^2 of 0.53.

Second Model: Our selection model kept several dozen of our predictors and their interaction terms, including several that were not deemed significant in our first model such as `log(1 + pct_college)`. This model has an R^2 of 0.63.

As part of our final paper, we will dive into the interpretations of the coefficients of these variables, and will look more critically at whether or not multicolinearity may be affecting our standard errors and p-values.