

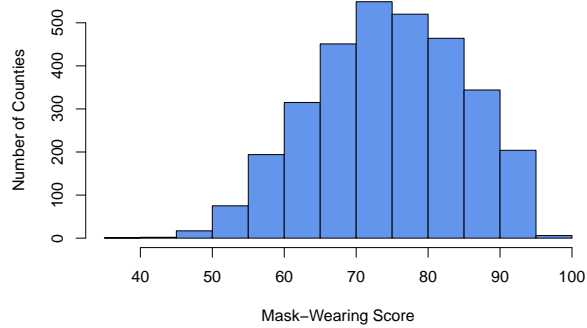
# EDA and Baseline Model

Chloe Shawah and Rena Cohen

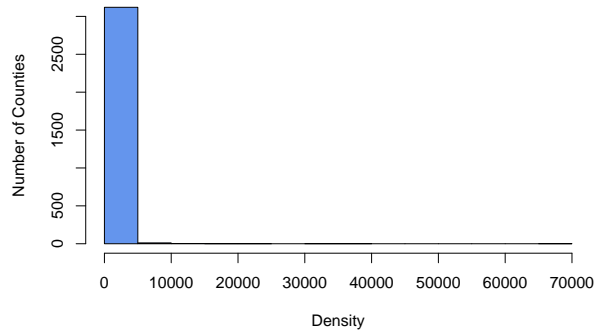
11/15/2020

## Exploratory Data Analysis

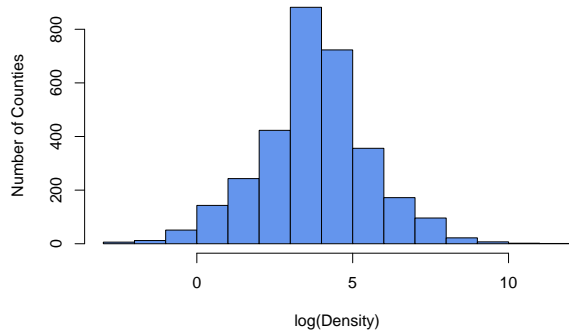
**Histogram of Mask-Wearing Score**

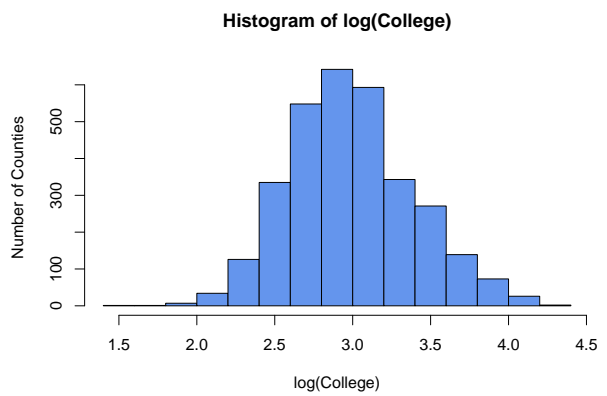
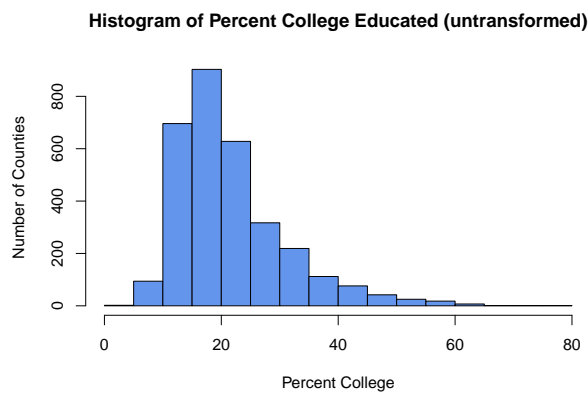
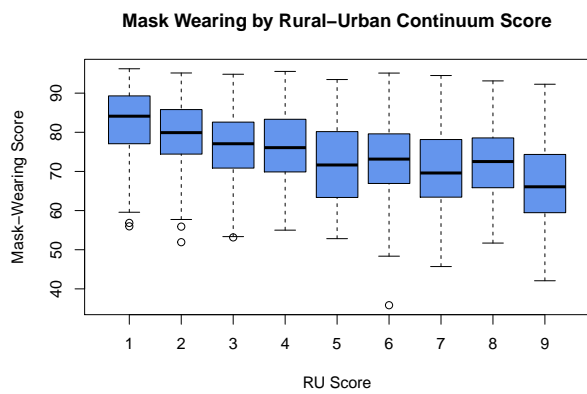
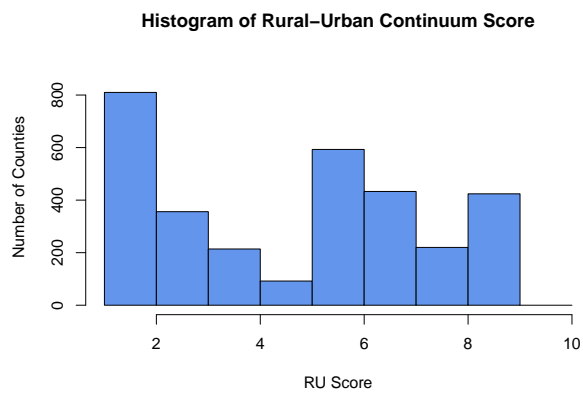
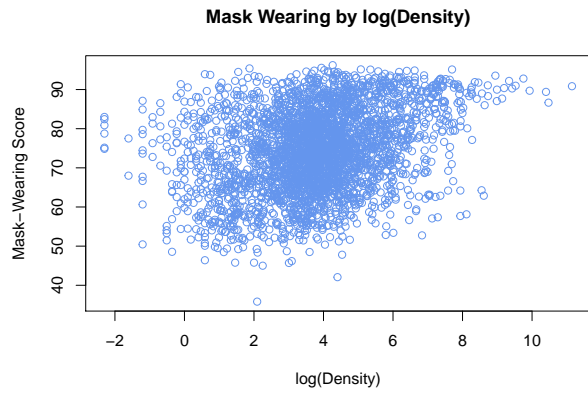


**Histogram of Density (untransformed)**

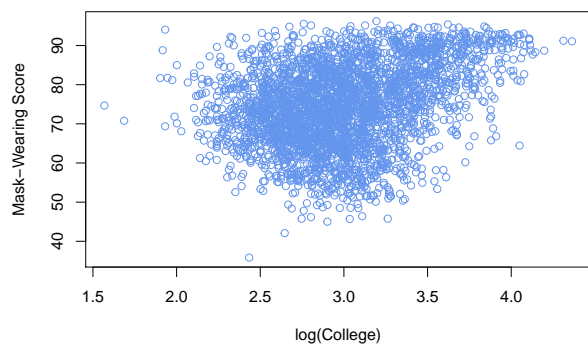


**Histogram of log(Density)**

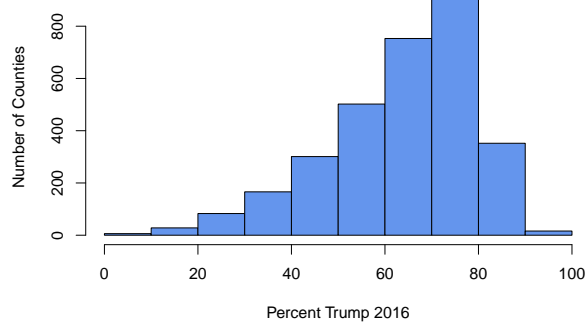




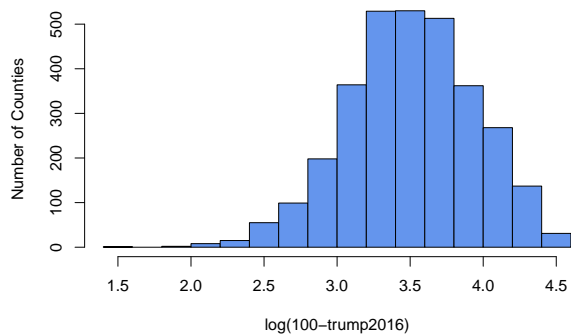
**Mask Wearing by log(College)**



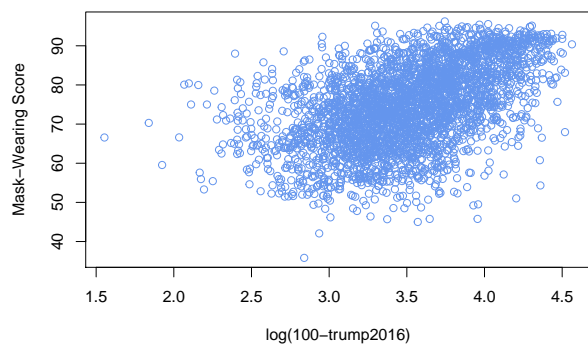
**Histogram of Percent Voted for Trump 2016 (untransformed)**



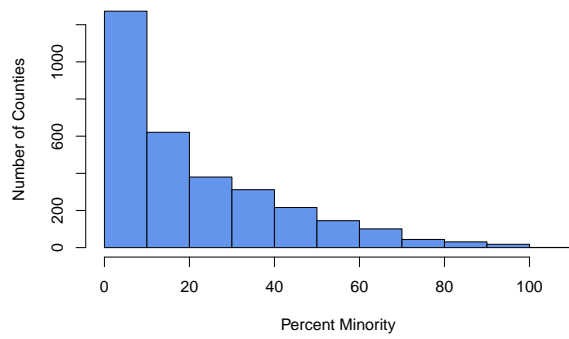
**Histogram of log(100–Trump2016)**



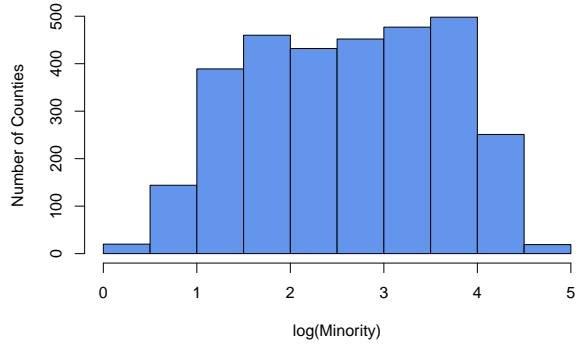
**Mask Wearing by log(100–Trump2016)**



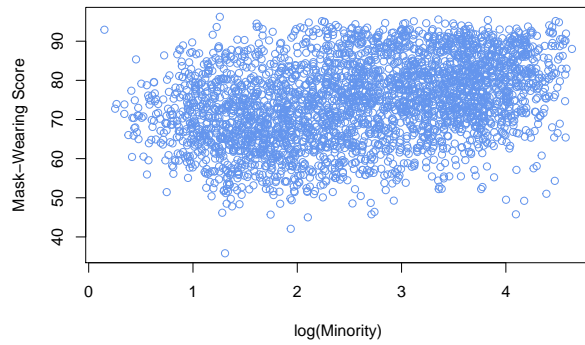
**Histogram of Percent Minority (black, hispanic, native) (untransforme**



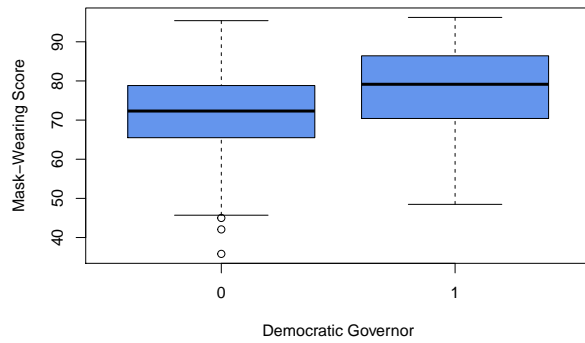
**Histogram of log(Minority)**

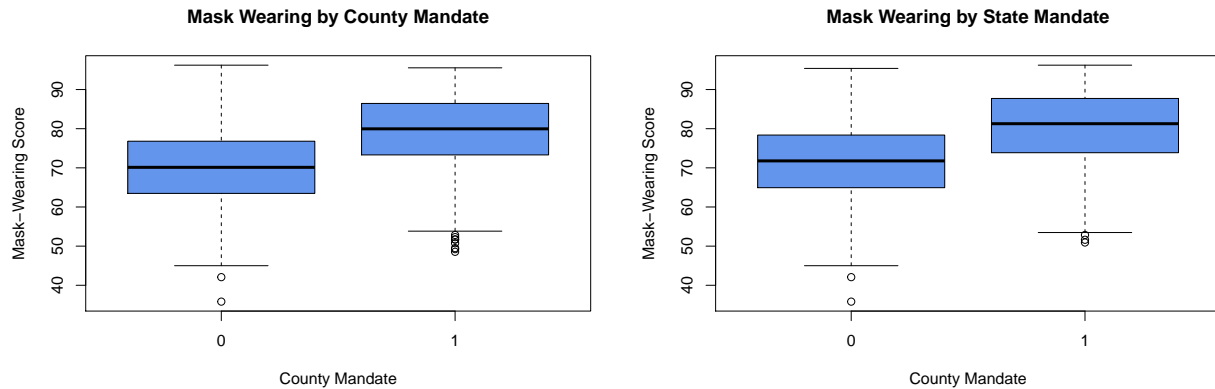


**Mask Wearing by log(Minority)**



**Mask Wearing by Democratic Governor**





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.

```
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 the 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 multicollinearity, but moving forward, it might be best to create two a variables that sums `pct_college` and `pct_some_college`. We might consider doing the same thing with minority groups. Finally, we removed `pct_trump2020` from the model because the multicollinearity between that and the 2016 percent was inflating the standard errors.

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

```
fullmodel = lm(pct_mask ~ pop_2019 + 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 ~ pop_2019 + 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
## -25.3324  -4.6905   0.3858   4.7527  22.9165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.501e+01  4.737e+00   7.391 1.87e-13 ***
## pop_2019          1.239e-07  4.542e-07   0.273 0.785092
## ru_continuum      -8.997e-01  6.660e-02 -13.510 < 2e-16 ***
## log(density)       5.013e-01  8.762e-02   5.721 1.16e-08 ***
## pct_hs            -6.896e-02  3.254e-02  -2.119 0.034141 *
## log(1 + pct_college)  9.460e-01  6.656e-01   1.421 0.155320
## log(pct_poverty)    3.961e-01  4.725e-01   0.838 0.401932
## pct_female        -2.325e-01  6.281e-02  -3.702 0.000218 ***
## log(1 + pct_black)   1.185e+00  1.537e-01   7.711 1.67e-14 ***
## log(1 + pct_native)  -1.101e+00  2.321e-01  -4.743 2.20e-06 ***
## log(1 + pct_hispanic) 3.502e+00  1.910e-01  18.332 < 2e-16 ***
## log(1 + pct_asian)   -4.157e-01  4.076e-01  -1.020 0.307857
## log(pct_seniors)     7.650e+00  6.760e-01  11.317 < 2e-16 ***
## log(100 - pct_trump_2016) 5.295e+00  3.867e-01  13.690 < 2e-16 ***
## dem_governor        7.017e-01  3.442e-01   2.039 0.041571 *
## state_mandate        3.459e+00  3.684e-01   9.389 < 2e-16 ***
## county_mandate       3.815e+00  2.940e-01  12.977 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.089 on 3092 degrees of freedom
## (33 observations deleted due to missingness)
## Multiple R-squared:  0.5335, Adjusted R-squared:  0.5311
## F-statistic: 221 on 16 and 3092 DF, p-value: < 2.2e-16

interactionmodel = lm(pct_mask ~ (pop_2019 + 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(interact.

```

```

direction = "both", trace = 0)

summary(selected_model)

##
## Call:
## lm(formula = pct_mask ~ pop_2019 + 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 + ru_continuum:log(pct_poverty) +
##     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) + 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) +
##     pct_female:log(1 + pct_hispanic) + log(1 + pct_native):dem_governor +
##     log(1 + pct_native):log(100 - pct_trump_2016) + log(1 + pct_college):log(1 +
##     pct_black) + log(1 + pct_asian):log(100 - pct_trump_2016) +
##     pop_2019:log(100 - pct_trump_2016) + pop_2019:log(pct_seniors) +
##     pct_female:dem_governor + pct_female:county_mandate + log(1 +
##     pct_college):log(pct_seniors) + ru_continuum:log(1 + pct_college) +
##     ru_continuum:pct_female, data = clean_data_2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -27.7905  -3.9539   0.1619   3.9442  21.6951
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                       4.640e+01  3.469e+01   1.338
## pop_2019                          3.949e-05  1.126e-05   3.508
## ru_continuum                      -2.164e+00  1.441e+00  -1.502
## log(density)                       6.934e-01  1.237e+00   0.561
## pct_hs                            8.454e-01  1.879e-01   4.498
## log(1 + pct_college)               -1.165e+01  6.122e+00  -1.903
## log(pct_poverty)                   1.506e+01  5.620e+00   2.680

```

## pct_female	9.482e-01	5.312e-01	1.785
## log(1 + pct_black)	1.264e+01	2.135e+00	5.921
## log(1 + pct_native)	-2.429e+01	5.628e+00	-4.316
## log(1 + pct_hispanic)	2.511e+01	4.093e+00	6.136
## log(1 + pct_asian)	-1.409e+01	3.817e+00	-3.690
## log(pct_seniors)	-4.641e+00	5.236e+00	-0.886
## log(100 - pct_trump_2016)	-2.646e+01	7.925e+00	-3.339
## dem_governor	3.012e+01	8.053e+00	3.741
## state_mandate	-4.928e+01	7.328e+00	-6.725
## county_mandate	-7.985e+00	6.481e+00	-1.232
## ru_continuum:log(pct_poverty)	3.232e-01	1.779e-01	1.817
## log(1 + pct_college):log(100 - pct_trump_2016)	9.578e+00	1.079e+00	8.880
## dem_governor:state_mandate	5.202e+00	8.424e-01	6.175
## log(100 - pct_trump_2016):state_mandate	4.892e+00	1.115e+00	4.387
## log(density):log(1 + pct_black)	-3.141e-01	8.591e-02	-3.657
## ru_continuum:county_mandate	5.260e-01	1.070e-01	4.915
## log(1 + pct_hispanic):county_mandate	1.784e+00	3.376e-01	5.286
## log(1 + pct_native):log(1 + pct_hispanic)	1.611e+00	3.288e-01	4.900
## log(1 + pct_hispanic):state_mandate	-1.830e+00	4.102e-01	-4.462
## ru_continuum:log(1 + pct_hispanic)	-4.686e-01	7.688e-02	-6.095
## ru_continuum:log(1 + pct_native)	4.227e-01	9.676e-02	4.368
## log(1 + pct_native):state_mandate	3.910e+00	7.005e-01	5.581
## log(1 + pct_native):county_mandate	-7.234e-01	4.989e-01	-1.450
## log(pct_poverty):state_mandate	-3.428e+00	8.234e-01	-4.162
## log(100 - pct_trump_2016):dem_governor	3.859e+00	1.030e+00	3.746
## log(pct_poverty):log(100 - pct_trump_2016)	5.526e+00	1.071e+00	5.162
## log(1 + pct_college):log(pct_poverty)	-8.317e+00	1.115e+00	-7.457
## log(pct_poverty):log(1 + pct_asian)	2.284e+00	8.271e-01	2.761
## log(pct_poverty):log(1 + pct_black)	-2.063e+00	4.193e-01	-4.921
## log(1 + pct_black):dem_governor	-1.687e+00	2.932e-01	-5.756
## log(1 + pct_black):log(1 + pct_native)	5.034e-01	2.723e-01	1.849
## log(1 + pct_black):log(1 + pct_hispanic)	-4.101e-01	1.859e-01	-2.206
## log(density):log(100 - pct_trump_2016)	6.808e-01	2.120e-01	3.212
## log(100 - pct_trump_2016):county_mandate	-2.349e+00	6.987e-01	-3.363
## pct_hs:state_mandate	3.094e-01	5.088e-02	6.081
## pct_hs:log(pct_poverty)	-2.568e-01	6.437e-02	-3.989
## log(1 + pct_college):log(1 + pct_hispanic)	-2.577e+00	5.862e-01	-4.395
## pct_hs:log(1 + pct_hispanic)	-1.796e-01	3.141e-02	-5.718
## pct_hs:log(1 + pct_native)	2.197e-01	5.320e-02	4.130
## pct_female:log(100 - pct_trump_2016)	-3.016e-01	1.331e-01	-2.267
## log(1 + pct_hispanic):log(100 - pct_trump_2016)	-1.503e+00	3.936e-01	-3.818
## log(1 + pct_hispanic):log(pct_seniors)	2.042e+00	7.030e-01	2.905
## log(pct_seniors):state_mandate	9.452e+00	1.782e+00	5.304
## log(pct_seniors):dem_governor	-7.499e+00	1.663e+00	-4.509
## log(1 + pct_asian):state_mandate	5.464e+00	9.898e-01	5.520
## log(1 + pct_asian):dem_governor	-4.141e+00	8.557e-01	-4.840
## log(density):log(1 + pct_asian)	-4.869e-01	1.594e-01	-3.054
## log(1 + pct_college):log(1 + pct_native)	1.481e+00	9.524e-01	1.555
## log(density):log(pct_seniors)	-6.568e-01	3.246e-01	-2.023
## pct_female:log(1 + pct_hispanic)	-1.484e-01	5.940e-02	-2.498
## log(1 + pct_native):dem_governor	-2.093e+00	6.157e-01	-3.399
## log(1 + pct_native):log(100 - pct_trump_2016)	1.490e+00	5.911e-01	2.521
## log(1 + pct_college):log(1 + pct_black)	-1.071e+00	4.421e-01	-2.423
## log(1 + pct_asian):log(100 - pct_trump_2016)	2.207e+00	8.282e-01	2.665



## pop_2019:log(100 - pct_trump_2016)	-5.701e-06	1.993e-06	-2.860
## pop_2019:log(pct_seniors)	-5.878e-06	2.560e-06	-2.296
## pct_female:dem_governor	-3.052e-01	1.182e-01	-2.583
## pct_female:county_mandate	2.708e-01	1.180e-01	2.294
## log(1 + pct_college):log(pct_seniors)	3.559e+00	1.580e+00	2.253
## ru_continuum:log(1 + pct_college)	-4.145e-01	1.971e-01	-2.103
## ru_continuum:pct_female	4.193e-02	2.462e-02	1.703
##	Pr(> t )		
## (Intercept)	0.181124		
## pop_2019	0.000457	***	
## ru_continuum	0.133297		
## log(density)	0.575025		
## pct_hs	7.11e-06	***	
## log(1 + pct_college)	0.057143	.	
## log(pct_poverty)	0.007407	**	
## pct_female	0.074351	.	
## log(1 + pct_black)	3.55e-09	***	
## log(1 + pct_native)	1.64e-05	***	
## log(1 + pct_hispanic)	9.53e-10	***	
## log(1 + pct_asian)	0.000228	***	
## log(pct_seniors)	0.375455		
## log(100 - pct_trump_2016)	0.000852	***	
## dem_governor	0.000187	***	
## state_mandate	2.09e-11	***	
## county_mandate	0.218047		
## ru_continuum:log(pct_poverty)	0.069271	.	
## log(1 + pct_college):log(100 - pct_trump_2016)	< 2e-16	***	
## dem_governor:state_mandate	7.49e-10	***	
## log(100 - pct_trump_2016):state_mandate	1.19e-05	***	
## log(density):log(1 + pct_black)	0.000260	***	
## ru_continuum:county_mandate	9.33e-07	***	
## log(1 + pct_hispanic):county_mandate	1.34e-07	***	
## log(1 + pct_native):log(1 + pct_hispanic)	1.01e-06	***	
## log(1 + pct_hispanic):state_mandate	8.40e-06	***	
## ru_continuum:log(1 + pct_hispanic)	1.23e-09	***	
## ru_continuum:log(1 + pct_native)	1.29e-05	***	
## log(1 + pct_native):state_mandate	2.59e-08	***	
## log(1 + pct_native):county_mandate	0.147200		
## log(pct_poverty):state_mandate	3.24e-05	***	
## log(100 - pct_trump_2016):dem_governor	0.000183	***	
## log(pct_poverty):log(100 - pct_trump_2016)	2.60e-07	***	
## log(1 + pct_college):log(pct_poverty)	1.15e-13	***	
## log(pct_poverty):log(1 + pct_asian)	0.005792	**	
## log(pct_poverty):log(1 + pct_black)	9.08e-07	***	
## log(1 + pct_black):dem_governor	9.49e-09	***	
## log(1 + pct_black):log(1 + pct_native)	0.064582	.	
## log(1 + pct_black):log(1 + pct_hispanic)	0.027483	*	
## log(density):log(100 - pct_trump_2016)	0.001331	**	
## log(100 - pct_trump_2016):county_mandate	0.000782	***	
## pct_hs:state_mandate	1.34e-09	***	
## pct_hs:log(pct_poverty)	6.79e-05	***	
## log(1 + pct_college):log(1 + pct_hispanic)	1.15e-05	***	
## pct_hs:log(1 + pct_hispanic)	1.18e-08	***	
## pct_hs:log(1 + pct_native)	3.73e-05	***	

```

## pct_female:log(100 - pct_trump_2016)          0.023462 *
## log(1 + pct_hispanic):log(100 - pct_trump_2016) 0.000137 ***
## log(1 + pct_hispanic):log(pct_seniors)         0.003704 **
## log(pct_seniors):state_mandate                 1.21e-07 ***
## log(pct_seniors):dem_governor                  6.76e-06 ***
## log(1 + pct_asian):state_mandate                3.68e-08 ***
## log(1 + pct_asian):dem_governor                1.37e-06 ***
## log(density):log(1 + pct_asian)                0.002275 **
## log(1 + pct_college):log(1 + pct_native)        0.120150
## log(density):log(pct_seniors)                  0.043122 *
## pct_female:log(1 + pct_hispanic)               0.012557 *
## log(1 + pct_native):dem_governor               0.000685 ***
## log(1 + pct_native):log(100 - pct_trump_2016)  0.011748 *
## log(1 + pct_college):log(1 + pct_black)         0.015433 *
## log(1 + pct_asian):log(100 - pct_trump_2016)   0.007735 **
## pop_2019:log(100 - pct_trump_2016)            0.004260 **
## pop_2019:log(pct_seniors)                     0.021726 *
## pct_female:dem_governor                        0.009854 **
## pct_female:county_mandate                      0.021840 *
## log(1 + pct_college):log(pct_seniors)          0.024356 *
## ru_continuum:log(1 + pct_college)              0.035529 *
## ru_continuum:pct_female                        0.088712 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.304 on 3041 degrees of freedom
## (33 observations deleted due to missingness)
## Multiple R-squared:  0.6372, Adjusted R-squared:  0.6292
## F-statistic: 79.72 on 67 and 3041 DF, p-value: < 2.2e-16

```