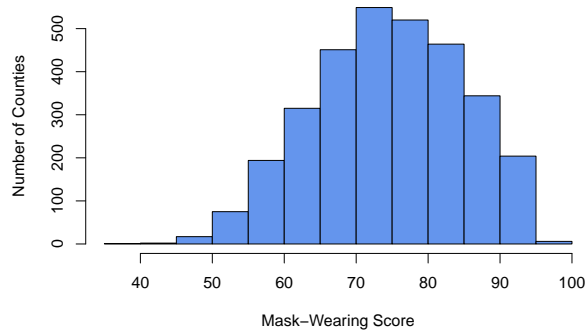


# EDA and Baseline Model

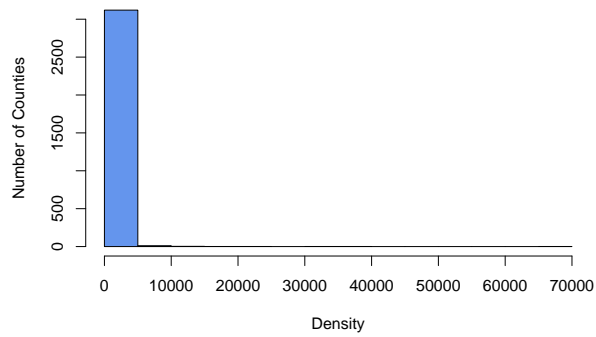
Chloe Shawah and Rena Cohen

11/15/2020

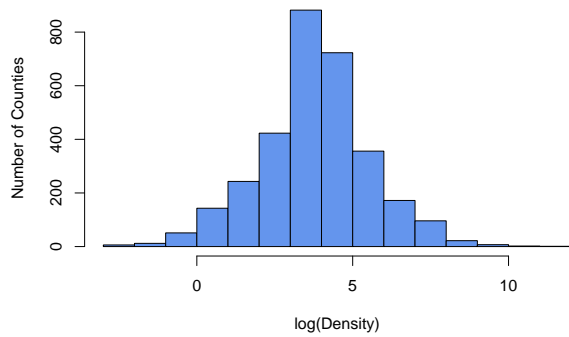
**Histogram of Mask-Wearing Score**



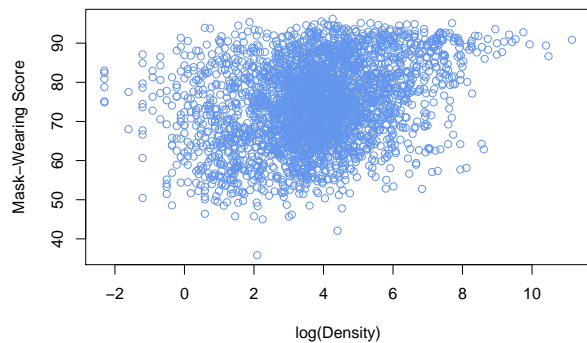
**Histogram of Density (untransformed)**

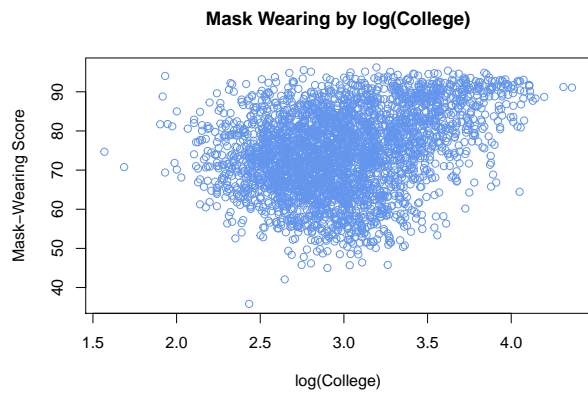
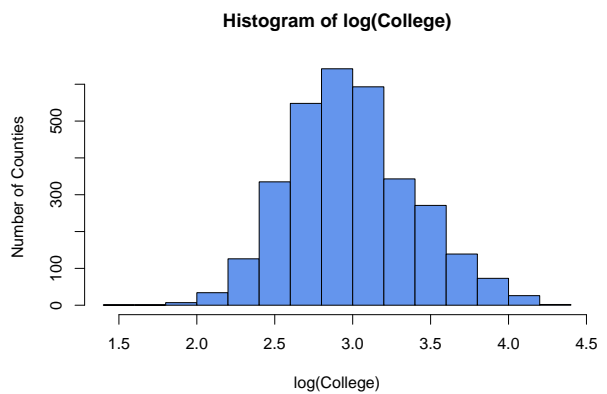
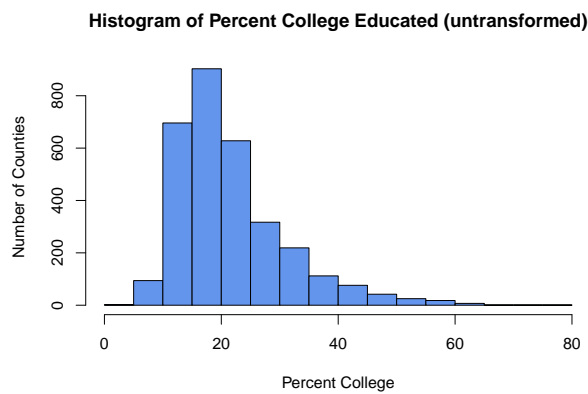
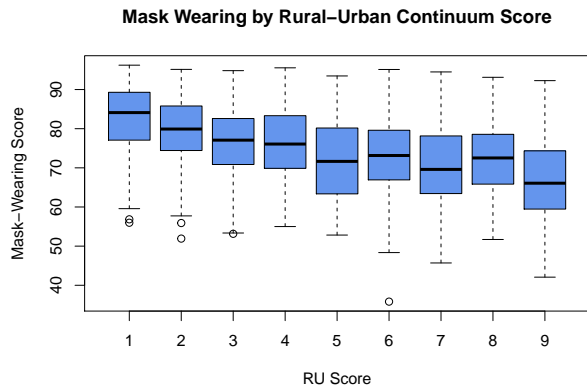
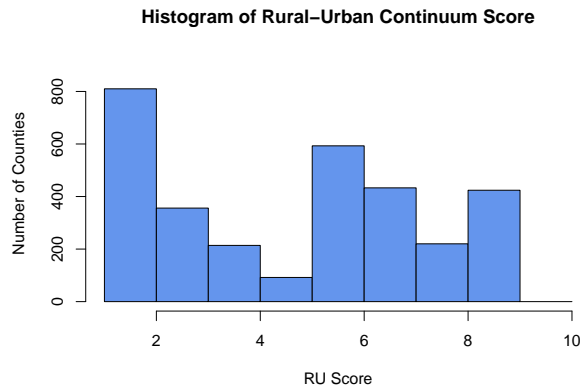


**Histogram of log(Density)**

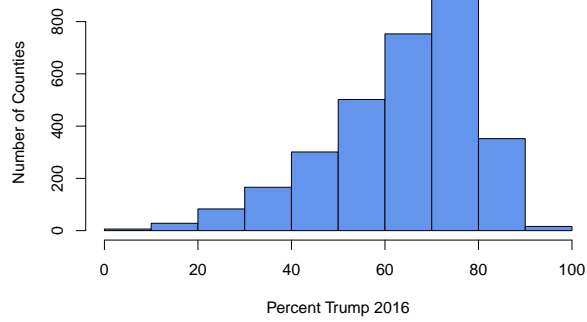


**Mask Wearing by log(Density)**

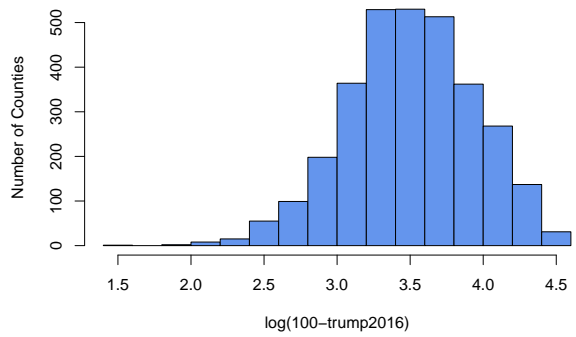




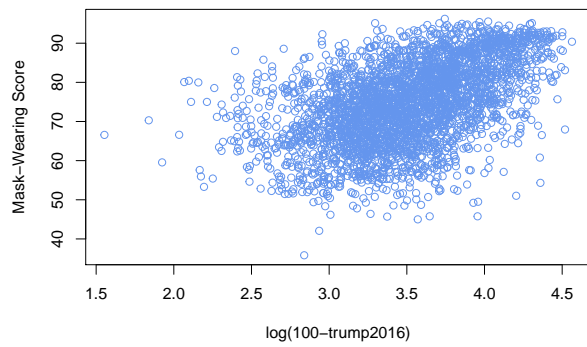
**Histogram of Percent Voted for Trump 2016 (untransformed)**



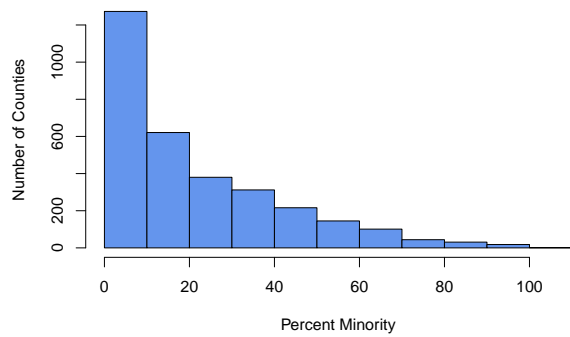
**Histogram of  $\log(100 - \text{Trump2016})$**



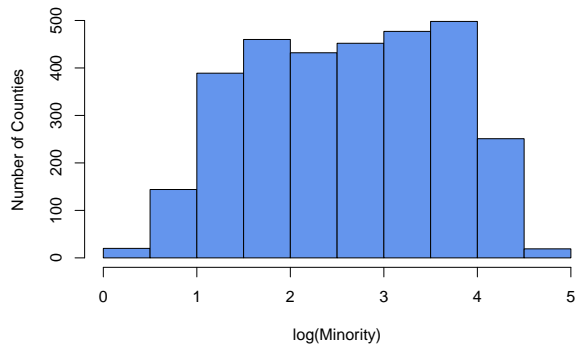
**Mask Wearing by  $\log(100 - \text{Trump2016})$**

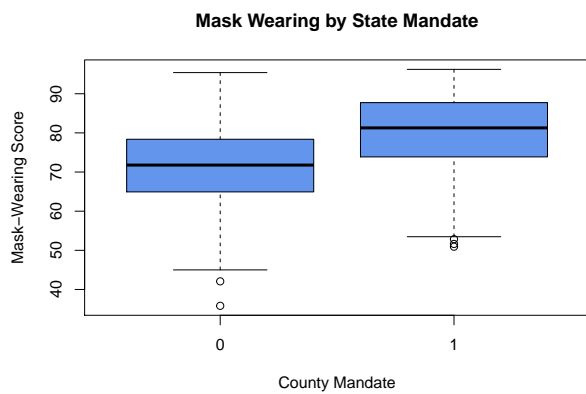
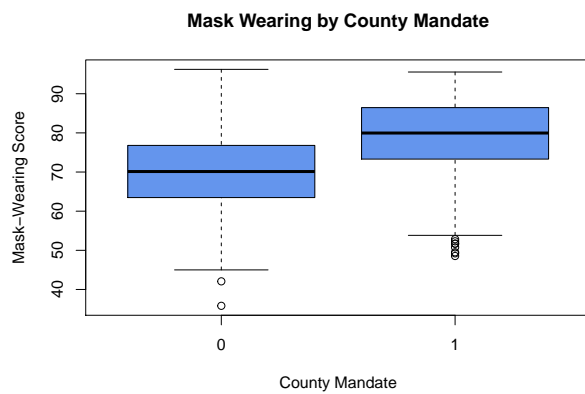
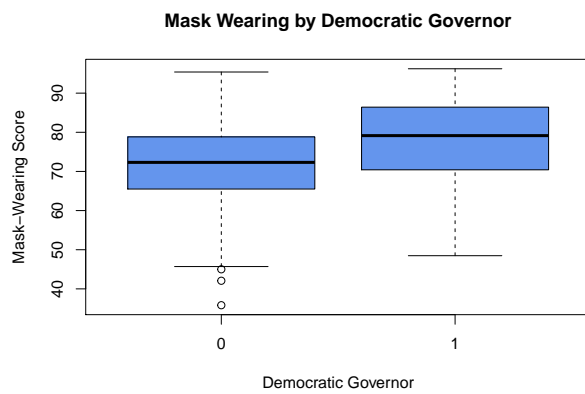
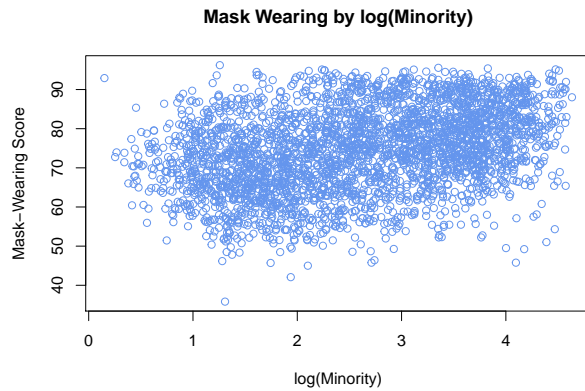


**Histogram of Percent Minority (black, hispanic, native) (untransformed)**



**Histogram of  $\log(\text{Minority})$**





Other transformations:  $\log(\text{pct\_seniors})$ ,  $\log(\text{pct\_poverty})$  each ethnicity separately  $\text{pct\_hs}$  leave alone  
 $\text{\%female}$  is icky but let's leave it

had to do  $1 + \log$  left out 2 education categories bc of collinearity removed  $\text{trump2020}$  bc of collinearity

Thoughts: maybe  $\text{\%minority}$  and  $\text{\%college}$

```
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_hispani
               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)
```

Things to think about: imputing data

```
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