CS EDA and Exploring

Chloe Shawah

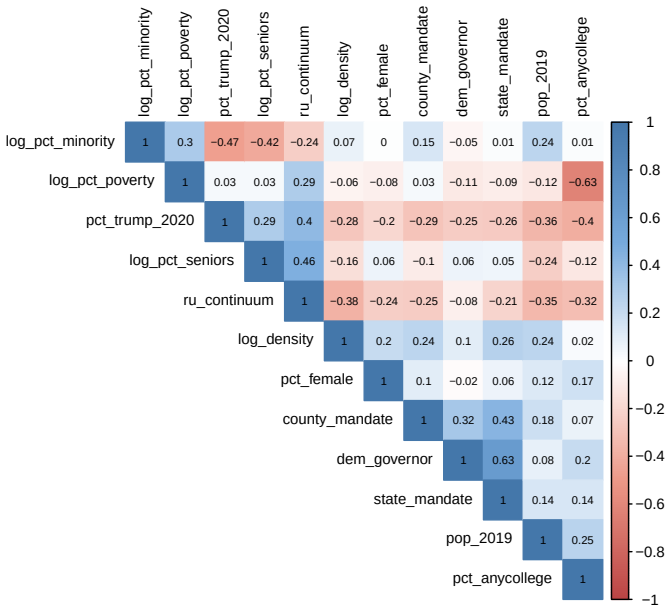
12/7/2020

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:

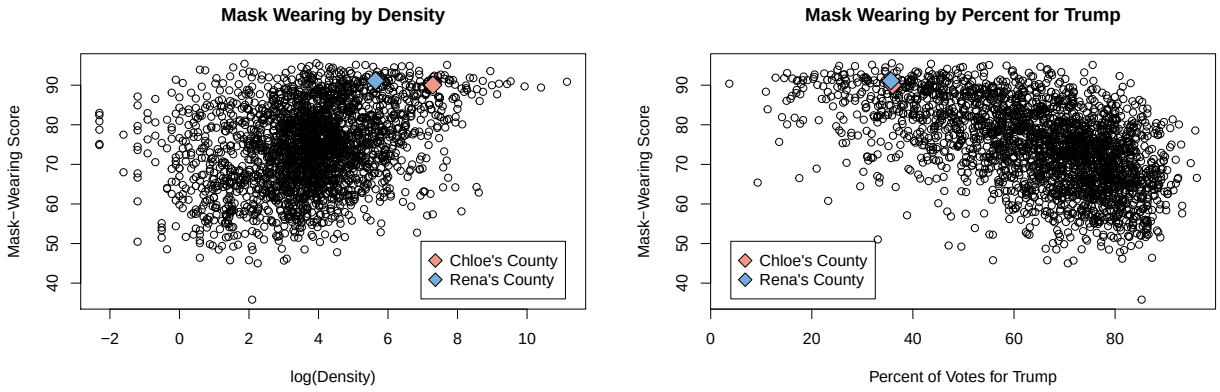
pop_2019:	population in 2019
log_density:	population density, log transformed
ru_continuum:	discrete score from 1 to 10 on the rural-Urban continuum with 1 being the most urban
log_pct_seniors:	percent of adults 65+ in 2019, log transformed
log_pct_minority:	percent of people from minority backgrounds 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:

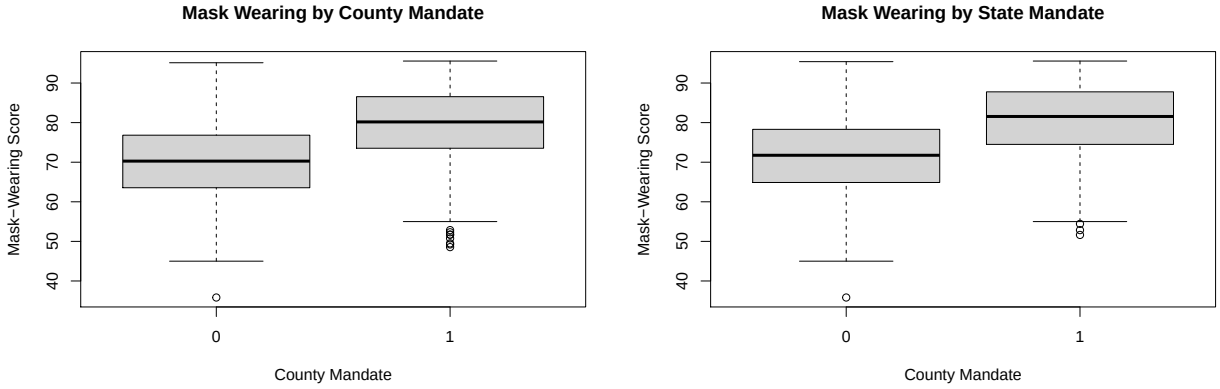


The strongest correlations we see between predictors is between percent poverty and percent college at -0.63 and between Democrat governor and state mandate at 0.63. Both of these associations do not surprise us, and are not so high that we are concerned about their impacts for our modeling. Another 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 $H_0 : \mu_1 = \mu_2$.

```
## \begin{table}[ht]
## \centering
## \begin{tabular}{lrrrrrr}
##   \hline
##   & No Mandate & Mandate & Difference & t-statistic & df & p-value \\
##   \hline
```

```
## 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 \\
##    \hline
## \end{tabular}
## \end{table}
```

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.

NEXT SECTION