

Model 2 clean_script

Jun Qian, Lucas Schroyer, Ryan Mitchell, Oliver Chang

4/8/2021

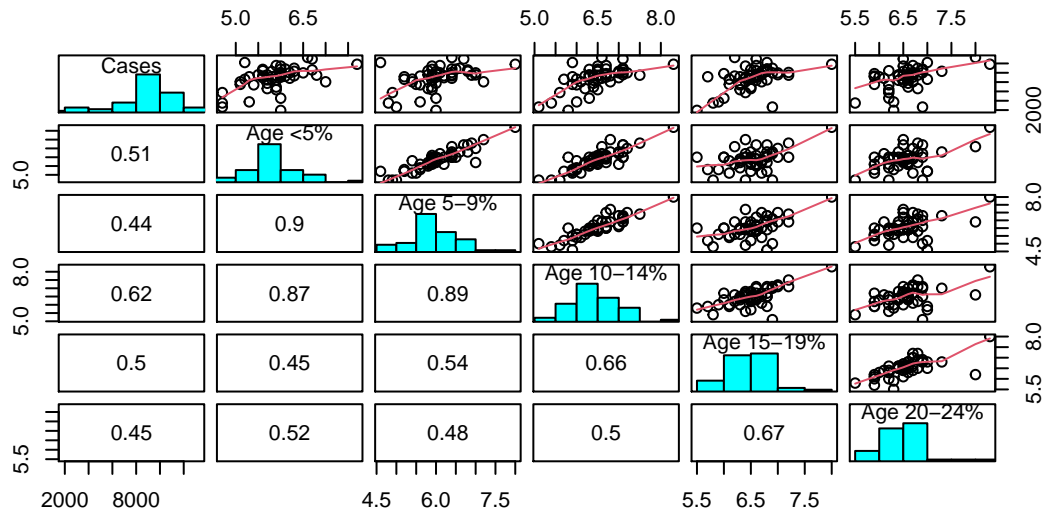
```
load("pwd/model_1_workspace.RData")
```

Second Model.

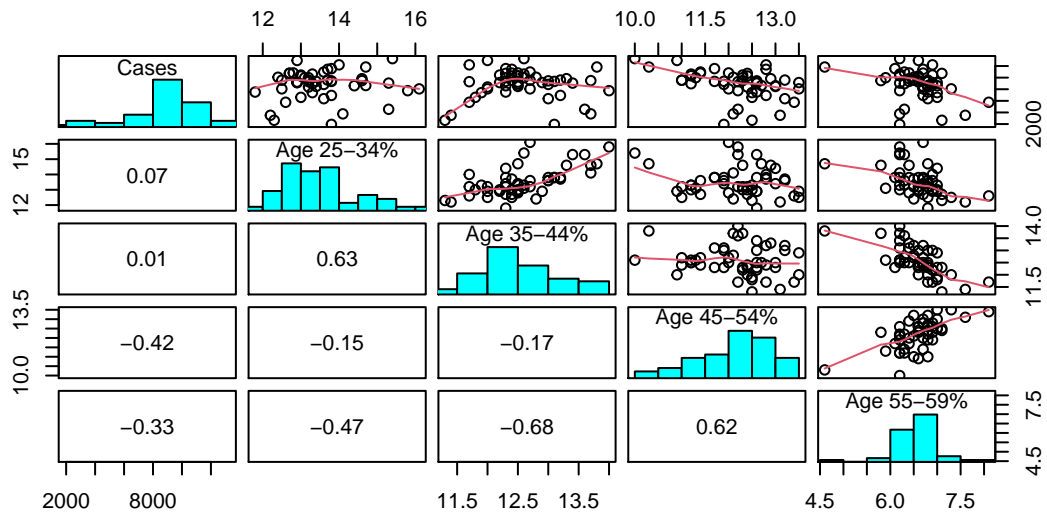
As we thought about an approach to building our Model 2 specifications, the general approach that our group decided to take was to incrementally test the addition of new features to our Model 1 specification in descending order of expected importance, according to our general understanding of how viruses spread. At each iteration, we began by examining the relationship (using scatterplots) between a new demographic or policy-related feature and our outcome of interest: COVID cases per 100K people 365 days after each state declared a state of emergency. If we observed a relationship, we ran a combination of t-Tests and/or ANOVA F-Tests to determine whether or not to add the feature to our Model 1 specification. We proceeded in this 'greedy' algorithmic fashion by adding variables to our Model 1 specification until we could no longer justify further additions based on results from ANOVA F-Tests. The first incremental feature we tested was derived from observational data from the census bureau on population age distributions by state.

One of the contributing factors to the spread of COVID-19 is asymptomatic spread. Younger people have been shown to have milder symptoms and therefore it stands to reason that they may be less likely to get tested, and ultimately end up spreading the disease at a greater rate than older age groups. To make matters worse, younger people tend to interact with more people (source?) as a result of being in school (switching between classrooms, touching desks and other surfaces where other students have been, congregating in large cafeterias, etc.), which increases the probability of viral transmission. Hence, for our second model, we were interested in testing our general hypothesis that age demographics may play a role in the spread of COVID-19. We began by doing some basic exploratory data analysis (EDA) with respect to age distributions and cumulative COVID-19 case counts per 100K, which is our target variable for Lab 2.

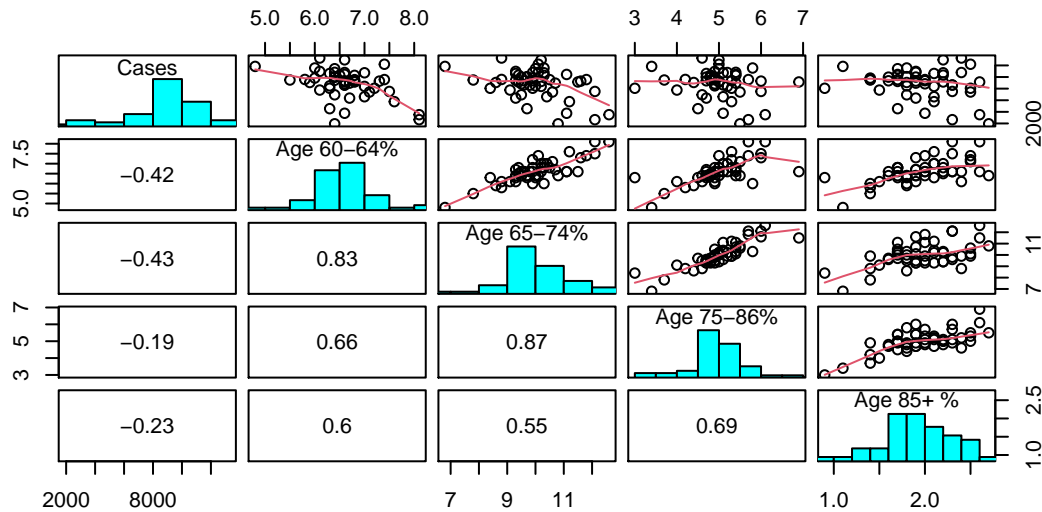
Case vs Age Bins Distribution and Scatter Plots



Case vs Age (Contin.) Bins Distribution and Scatter Plots



Case vs Age (Contin.) Bins Distribution and Scatter Plots



#Significant

```
anova(model_a, model_1_final, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)
```

```
## 1 47 184378849
```

```
## 2 48 245004785 -1 -60625936 15.454 0.0002765 ***
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
' ' 1
```

#Insignificant

```
anova(model_b, model_a, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)
```

```
## 1 46 170183297
```

```
## 2 47 184378849 -1 -14195553 3.837 0.05621 .
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
' ' 1
```

```
#Significant F-test
```

```
anova(model_c, model_a, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
```

```
Percent.Sex.And.Age.Total.Population.10.To.14.Years
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)
```

```
## 1 45 132785225
```

```
## 2 47 184378849 -2 -51593624 8.7424 0.00062 ***
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
```

```
' ' 1
```

```
#Insignificant relative to model_c
```

```
anova(model_d, model_c, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
```

```
Percent.Sex.And.Age.Total.Population.10.To.14.Years +
```

```
## Percent.Sex.And.Age.Total.Population.15.To.19.Years
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
```

```
Percent.Sex.And.Age.Total.Population.10.To.14.Years
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)
```

```
## 1 44 132369352
```

```
## 2 45 132785225 -1 -415873 0.1382 0.7118
```

```
anova(model_e, model_d, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
```

```
Percent.Sex.And.Age.Total.Population.10.To.14.Years +
```

```
## Percent.Sex.And.Age.Total.Population.15.To.19.Years +
```

```
Percent.Sex.And.Age.Total.Population.20.To.24.Years
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change
```

```
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
```

```
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
```

```
Percent.Sex.And.Age.Total.Population.10.To.14.Years +
```

```
## Percent.Sex.And.Age.Total.Population.15.To.19.Years
```

```
## Res.Df RSS Df Sum of Sq F Pr(>F)
```

```
## 1 43 124070515
## 2 44 132369352 -1 -8298836 2.8762 0.09713 .
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
```

```
anova(model_e, model_c, test = "F")
```

```
## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
Percent.Sex.And.Age.Total.Population.10.To.14.Years +
## Percent.Sex.And.Age.Total.Population.15.To.19.Years +
Percent.Sex.And.Age.Total.Population.20.To.24.Years
## Model 2: cases_per_100k_at_365d ~ median_transit_change
+ Percent.Sex.And.Age.Total.Population.Under.5.Years +
## Percent.Sex.And.Age.Total.Population.5.To.9.Years +
Percent.Sex.And.Age.Total.Population.10.To.14.Years
## Res.Df RSS Df Sum of Sq F Pr(>F)
## 1 43 124070515
## 2 45 132785225 -2 -8714710 1.5102 0.2324
```

When performing EDA on the categorical age group distributions and their relationship to our target variable, we noticed a strong positive correlation to the target among age groups 0-24. Age groups in the 24+ range did not appear to follow a consistent pattern with respect to correlation with our target variable and were therefore not a key focus area for our analysis.

Next, we performed a series of ANOVA F-tests on a nested set of linear regression models with parameter estimates for this set of age groups (Under 5, 5-9, 10-14, 15-19, and 20-24) as well as our main variable of interest (median transit mobility change). Our motivation here was to understand whether the regression residuals were measurably different from one another between the different (nested) model specifications. For context, the null hypothesis for an F-test is that fitting additional coefficients for a longer model does not measurably reduce the residuals relative to a nested model with fewer parameters.

Our first ANOVA F-test compared our Model 1 with a new model that had an additional parameter estimate for the percentage of the population under 5 years old. With a p-value of 0.0003, we rejected the null hypothesis and used this new model (model_a) as the baseline model for subsequent comparisons with additional parameter estimates for different age categories.

We failed to reject the null hypothesis when adding an estimator for ages 5-9 (model_b), and succeeded in rejecting the null hypothesis when adding an estimator for ages 10-14 (model_c), with a p-value of 0.001. Adding additional estimators for 15-19 and 20-24 failed to reject the null hypothesis that these models (model_d and model_e) were measurably better at reducing residuals than model_c. These results suggested creating two new features for the percentage of the population ages 0-9 and 10-24 and adding them to our base Model 1 to create the first specification for our Model 2.

```
final_df <- final_df %>%
  mutate(pop_pct_age_0_9 = Percent.Sex.And.Age.Total.Population.Under.5.Years +
    Percent.Sex.And.Age.Total.Population.5.To.9.Years,
    pop_pct_age_10_24 =
    Percent.Sex.And.Age.Total.Population.10.To.14.Years +
    Percent.Sex.And.Age.Total.Population.15.To.19.Years +
```

```

Percent.Sex.And.Age.Total.Population.20.To.24.Years)

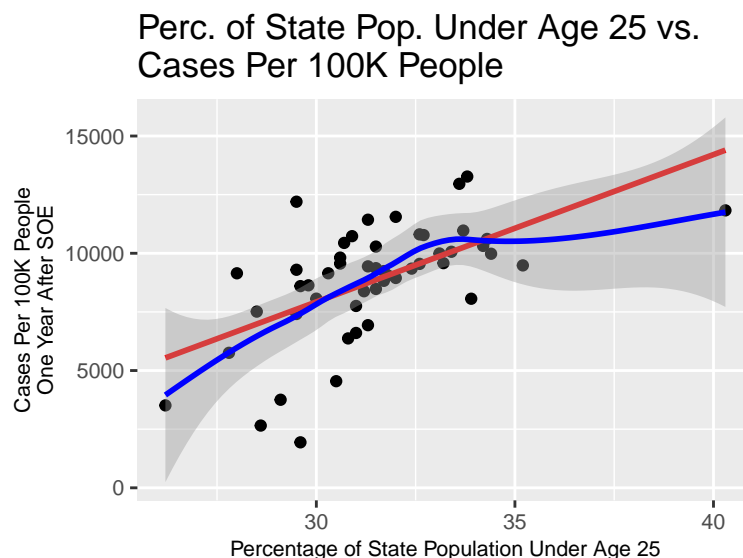
cor(final_df$pop_pct_age_0_9, final_df$pop_pct_age_10_24)

## [1] 0.7701834

final_df <- final_df %>%
  mutate(pop_pct_age_0_24 = Percent.Sex.And.Age.Total.Population.Under.5.Years +
    Percent.Sex.And.Age.Total.Population.5.To.9.Years +
    Percent.Sex.And.Age.Total.Population.10.To.14.Years +
    Percent.Sex.And.Age.Total.Population.15.To.19.Years +
    Percent.Sex.And.Age.Total.Population.20.To.24.Years)

```

However, after measuring a 0.77 correlation between these two features - and with the goal of increased model parsimony - we decided to group them together to prevent the standard errors for their respective coefficients from increasing substantially. Let's take a look and assess whether this new variable for percentage of the population < 24 years old visually satisfies the conditional linearity expectation with respect to our target.



The conditional linear expectation between our population percentage aged 0-24 and our target appears to be met. In addition to evaluating the linear relationships between the demographic age variables and our target, we also evaluated the log, square root, and square transformations to the age distribution data but did not find them to aid in reducing the frequency or magnitude of outlier data points.

Indeed, including a feature for the percentage of the population aged < 24 years old appears to have improved our model's performance. An ANOVA F-test returned a p-value of 0.0001, suggesting that we reject the null hypothesis that the (interim) Model 2's residuals were not measurably different from the residuals of Model 1.

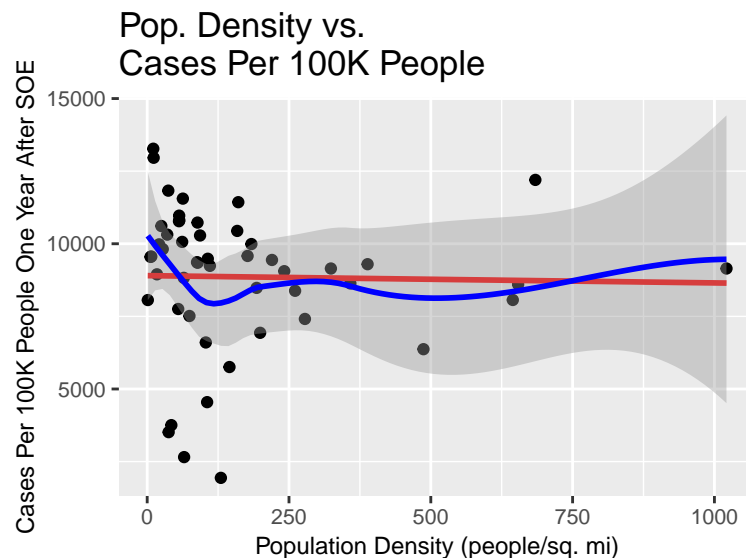
```

## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24
## Model 2: cases_per_100k_at_365d ~ median_transit_change
##   Res.Df      RSS Df Sum of Sq    F    Pr(>F)

```

```
## 1      47 176654365
## 2      48 245004785 -1 -68350419 18.185 9.595e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

After accounting for changes in mobility and demographic age differences between states, the next variable we wanted to explore as part of our descriptive model for COVID case counts was population density. According to the World Health Organization, “COVID-19 virus is primarily transmitted between people through respiratory droplets and contact routes.” (<https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations>). In other words, COVID spreads primarily through physical interactions between infected and uninfected hosts, regardless of whether the actual mechanism of transmission is airborne or surface based contact. Hence, it stands to reason that more densely populated areas would see greater rates of infections, because the frequency of these physical interactions will increase with population density. This was the motivation for our group exploring whether a relationship existed between state population density and our outcome variable of interest.



To our surprise, however, there was no clear relationship between population density and our outcome variable. A t-Test on the parameter estimate for population density’s relationship with our target variable failed to reject the null hypothesis that the coefficient value was not measurably different from zero.

```
##
## Call:
## lm(formula = cases_per_100k_at_365d ~ population_density, data = final_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6930.5  -799.2   487.9  1416.9  4372.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    8902.5248   454.6106  19.583  <2e-16 ***
## population_density  -0.2508     1.6991  -0.148    0.883
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2477 on 48 degrees of freedom
## Multiple R-squared:  0.0004535, Adjusted R-squared:  -0.02037
## F-statistic: 0.02178 on 1 and 48 DF,  p-value: 0.8833
```

In spite of this test, our group decided to move forward and try including the population density feature in Model 2, as we believed it to be a conceptually meaningful variable in describing the population prevalence of COVID in each state one year after each state declared a state of emergency. Therefore, we proceeded with an ANOVA F-Test to test whether the incremental population density feature measurably improved model performance (via reduction of residuals) relative to our current model with features for median transit mobility change and percentage of the population < 24 years old. This ANOVA F-test returned a p-value of 0.002, enough to reject the null hypothesis that the model residuals were not measurably different from one another.

```
##
## -----Model results-----

##
## Call:
## lm(formula = cases_per_100k_at_365d ~ median_transit_change +
##     pop_pct_age_0_24 + population_density, data = final_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3666.7 -1367.9  -257.5  1191.5  3684.0
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -10160.586    3917.035   -2.594  0.01268 *
## median_transit_change     82.223     23.514    3.497  0.00106 **
## pop_pct_age_0_24      628.592    119.313    5.268 3.55e-06 ***
## population_density      4.665      1.446    3.225  0.00232 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1770 on 46 degrees of freedom
## Multiple R-squared:  0.5109, Adjusted R-squared:  0.479
## F-statistic: 16.02 on 3 and 46 DF,  p-value: 2.875e-07
```

```
##
## -----ANOVA F-Test Significance Test Relative To Model 1-----
```

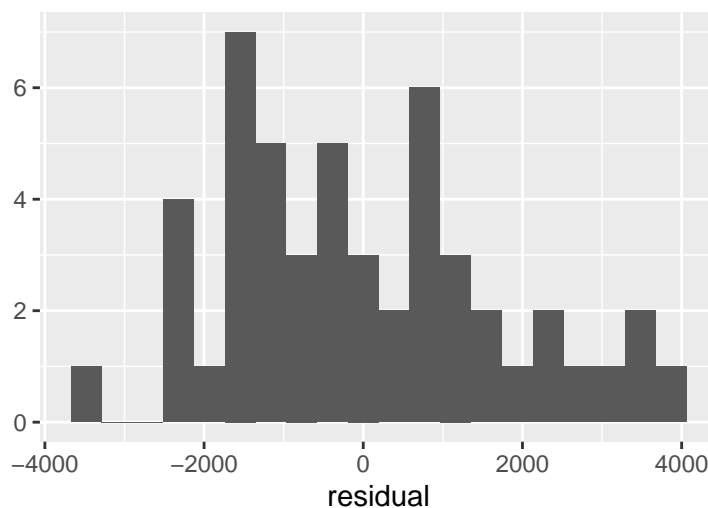
```
## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##     population_density
## Model 2: cases_per_100k_at_365d ~ median_transit_change
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      46 144077125
## 2      48 245004785 -2 -100927659 16.112 4.974e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



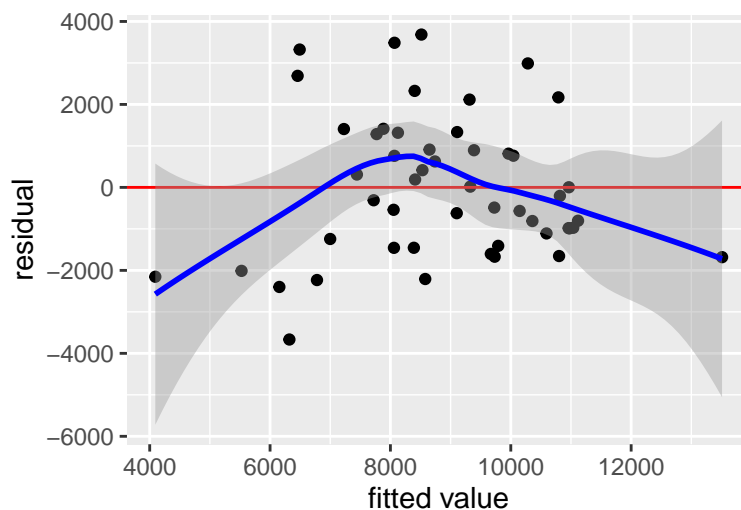
```
##
## -----ANOVA F-Test Significance Test Relative To Interim Model 2-----

## Analysis of Variance Table
##
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
##   population_density
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24
##   Res.Df      RSS Df Sum of Sq    F Pr(>F)
## 1      46 144077125
## 2      47 176654365 -1 -32577240 10.401 0.00232 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Histogram of Model 2 Residuals



Plot of Residuals vs Fitted Values



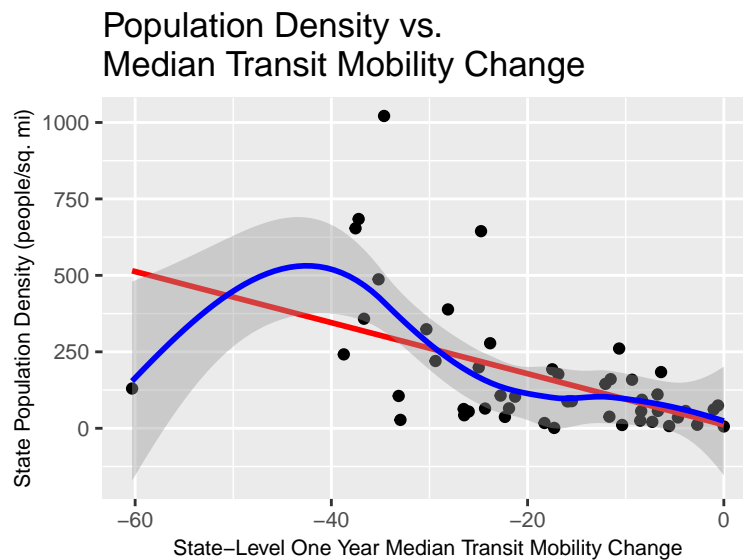
```
##
## -----Homoskedasticity Test-----
```

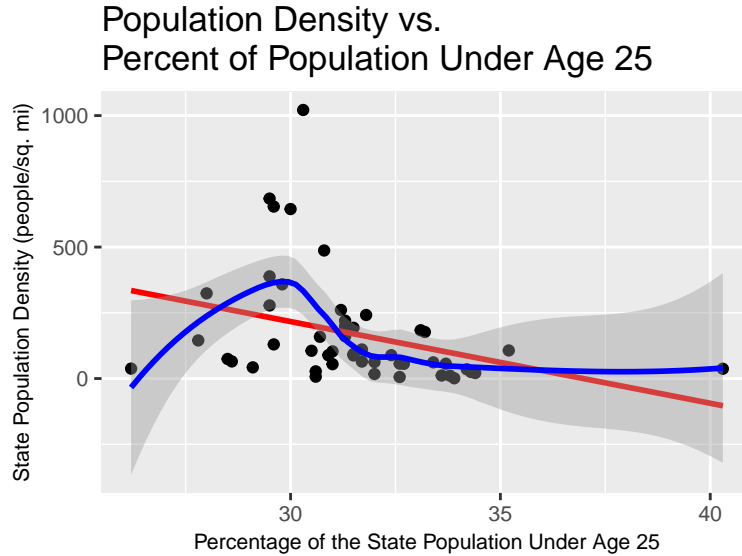
```
##
## studentized Breusch-Pagan test
##
## data: model_2_final
## BP = 8.3386, df = 3, p-value = 0.03951
```

```
##
## -----Normality of Residuals Test-----
```

```
##
## Shapiro-Wilk normality test
##
## data: model_2_final$residuals
## W = 0.97275, p-value = 0.2984
```

Because the population density feature only becomes statistically significant at the $p = 0.05$ level in our model when we include features for median change in transit mobility and the percentage of the population under age 25, we say that the population density has a conditional relationship with our outcome of interest. Both of these two co-variables (median transit mobility change and percentage of the population under age 25) are negatively correlated with population density and positively correlated with our outcome variable.





However, we are only interested in the unique variation of population density with respect to our outcome variable. When the OLS regression algorithm calculates the parameter estimate for population density, it starts by regressing population density on the other model co-variate (input) features. The residuals from that regression represent the portion of population density that is *not* colinear with median transit mobility change and percentage of the population under age 25. Then OLS regresses our target values on those residuals to derive an estimate for the population density parameter. The model summary tells us that if we hold the percentage of the population under 25 and the median transit mobility change for a state constant (we do not allow them to co-vary), that there exists a positive correlation between population density and our outcome variable at a statistically significant level ($p = 0.002$) using classical standard errors.

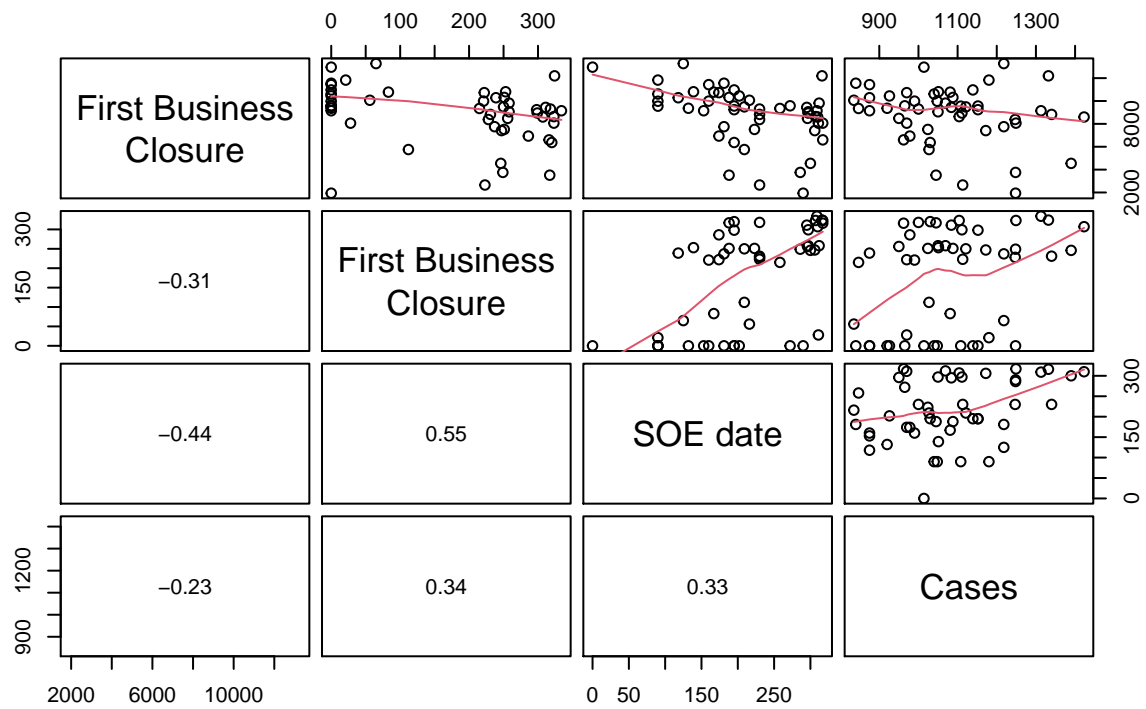
At this stage, we wanted to explore whether any state policy changes aimed at reducing the spread of COVID-19 added incremental descriptive power beyond our current Model 2 specification. In particular, we wanted to examine whether the timing of mask mandates, length and amount of increased government assistance via enhanced unemployment benefits, business closures, stay at home mandates, and travel quarantine restrictions had measurable effects on our outcome of interest after accounting for the features already in our Model 2 specification (which included state-level features for median change in transit mobility, percentage of the population under 25 years of age, and population density). All the policy-related features were recorded as dates, except for the increased unemployment insurance amount, which was recorded as an integer.

To align with our target variable of COVID-19 Cases per 100K one year after state of emergency declaration, we encoded the date related policy variables as the total number of days each policy was in place for after the state of emergency was announced for each state (up to 365 days). For transparency:

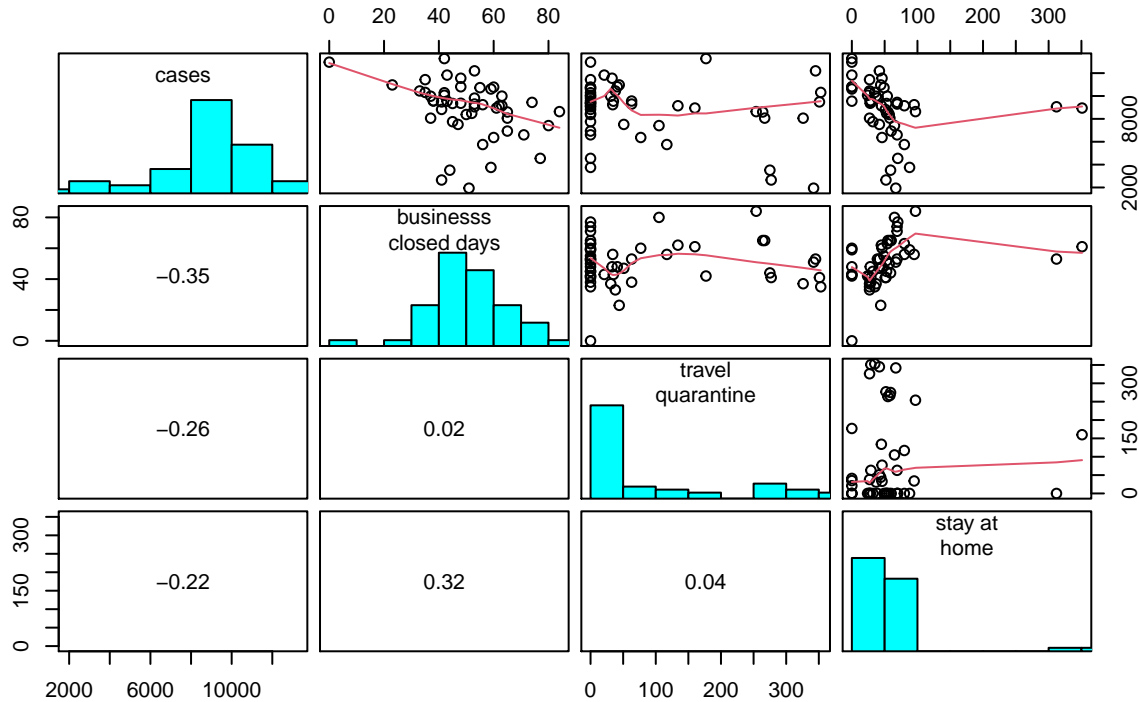
1. If a policy had no beginning and end dates, the total days were assigned as zero.
2. If a policy had beginning but not end dates, the total days were calculated by the date of state emergency declared + 364 days - the date of the beginning of the policy.
3. If a policy had both beginning and end dates, the total days were calculated by the difference in days of the two dates.

Once the policy-related features of our interest were transformed into days in force, we conducted EDA on the policy-features and COVID cases per 100k population using scatter plot and correlation matrices.

Policy Days vs Cases Per Capita Distribution and Scatter Plots



Policy Days (Contin.) vs Cases Per Cap. Distrib. and Scatter Plots



From the scatterplots, we see varying degrees of linearity between the policy features and our outcome of interest. We tested each of these features iteratively in the same manner as before, using significance from ANOVA F-tests as the benchmark to decide whether or not to include incremental policy features as part of our final Model 2. Ultimately, none of these policy variables (length of mask mandates, length and amount of increased government assistance via enhanced unemployment benefits, length business closures, length of stay at home mandates, and length of travel quarantine restrictions) returned a p-value that would allow us to reject the null hypothesis that the model residuals had not measurably improved.

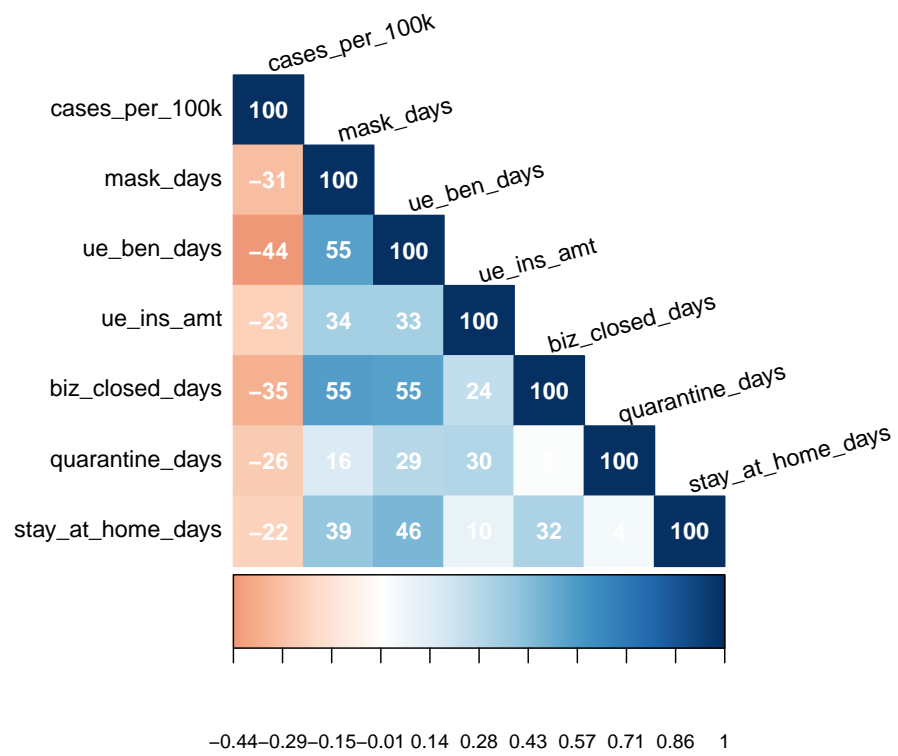
Still, it is worth noting that all of the aforementioned policy features, which were declared to mitigate COVID spread, demonstrated a negative correlation (from approximately -0.5 to -0.2) with our outcome variable of COVID cases per 100k people. The fact that these policy features failed to reject the null hypothesis in the ANOVA F-Test relative to the features already in our Model 2 was not entirely unexpected. Conceptually, several of the features share a significant amount information with median transit mobility change. One could make the argument that business closures, quarantine mandates, and stay at home mandates are all captured, to some extent, in the transit mobility change.

Ryan fleshing out writing here

Unemployment insurance extensions and benefit increases

We were surprised, however, that mask mandates

Correlation Matrix: State Policies vs. Cases



Jun's policy tests (all failed) for Model 2 here (to be reviewed...):

#Testing mask mandate days: Failed F-Test

```
model_2_final_incremental_test1 <- lm(cases_per_100k_at_365d ~
                                     median_transit_change +
                                     pop_pct_age_0_24 + population_density +
                                     mask_mandate_days, data = final_df)
anova(model_2_final_incremental_test1, model_2_final, test = "F")
```

Analysis of Variance Table

##

Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density + mask_mandate_days

Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
## 1	45	138865334				
## 2	46	144077125	-1	-5211792	1.6889	0.2004

#Testing extended unemployment benefit days: Failed F-Test

```
model_2_final_incremental_test2 <- lm(cases_per_100k_at_365d ~
                                     median_transit_change +
                                     pop_pct_age_0_24 + population_density + unemployment_benefits_d
                                     )
anova(model_2_final_incremental_test2, model_2_final, test = "F")
```

Analysis of Variance Table

##

Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density + unemployment_benefits_days

Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
## 1	45	137886393				
## 2	46	144077125	-1	-6190732	2.0204	0.1621

#Testing increased unemployment insurance amount: Failed F-Test

```
model_2_final_incremental_test3 <- lm(cases_per_100k_at_365d ~
                                     median_transit_change +
                                     pop_pct_age_0_24 + population_density + increased_weekly_unempl
                                     )
anova(model_2_final_incremental_test3, model_2_final, test = "F")
```

Analysis of Variance Table

##

Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density + increased_weekly_unemployment_insurance_amt_thru_jul31

Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +
population_density

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
## 1	45	135803753				
## 2	46	144077125	-1	-8273373	2.7415	0.1047

```
#Testing mask mandate days: Failed F-Test
```

```
model_2_final_incremental_test4 <- lm(cases_per_100k_at_365d ~  
                                     median_transit_change +  
                                     pop_pct_age_0_24 + population_density + business_closed_days_round1  
anova(model_2_final_incremental_test4, model_2_final, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density + business_closed_days_round1
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density
```

```
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      45 140581017
```

```
## 2      46 144077125 -1   -3496108 1.1191 0.2958
```

```
#Testing quarantine mandate days: Failed F-Test
```

```
model_2_final_incremental_test5 <- lm(cases_per_100k_at_365d ~  
                                     median_transit_change +  
                                     pop_pct_age_0_24 + population_density + travel_quarantine_mandate_days  
anova(model_2_final_incremental_test5, model_2_final, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density + travel_quarantine_mandate_days
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density
```

```
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      45 140733564
```

```
## 2      46 144077125 -1   -3343562 1.0691 0.3067
```

```
#Testing stay at home mandate days: Failed F-Test
```

```
model_2_final_incremental_test6 <- lm(cases_per_100k_at_365d ~  
                                     median_transit_change +  
                                     pop_pct_age_0_24 + population_density +  
                                     stay_at_home_days, data = final_df)  
anova(model_2_final_incremental_test6, model_2_final, test = "F")
```

```
## Analysis of Variance Table
```

```
##
```

```
## Model 1: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density + stay_at_home_days
```

```
## Model 2: cases_per_100k_at_365d ~ median_transit_change + pop_pct_age_0_24 +  
##      population_density
```

```
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
```

```
## 1      45 144017384
```

```
## 2      46 144077125 -1    -59741 0.0187 0.8919
```

```
#model_base=cases_per_100k_at_365d ~ median_transit_change +  
#pop_pct_age_0_24 + population_density
```



```
#Model_3=cases_per_100k_at_365d ~ median_transit_change +  
#pop_pct_age_0_24 + population_density + mask_mandate_days+  
#unemployment_benefits_days  
#+increased_weekly_unemployment_insurance_amt_thru_jul31+  
#business_close_open_days+travel_quarantine_mandate_days+stay_at_home_days  
  
#lm.II=lm(Model_3,data=final_df)  
#summary(lm.II)  
#resid(lm.II)  
#anova(lm.II, lm.I, test = "F")  
#shapiro.test(lm.II$residuals)  
#bptest(lm.II)
```

```
save.image(file = "pwd/model_2_workspace.RData")
```