Final Project 632 Rough Draft

Nic James, Sri Chandu, Thomas Li

05/21/2020

Contents

Abstract (100 words) - Nic

The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like.

Problem and Motivation (200 words) - Sri

The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like.

The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like.

Data Description

This data set is a collection of governmental sources at national, regional, and city levels from 190 countries for COVID19. It includes time series of vaccines, test, cases, deaths, recovered, intensive therapy, and policy measures by Oxford COVID-19 Government Response Tracker. We will used the World Bank Google Mobility Reports as well.

There are 16 variables in the base data set that we will be using for our regression. We will be limiting the location data strictly to California and using data from 3/15/2020 - 3/15/2021.

Our initial objective was to find out if running a linear regression of the Google Mobility data with the Covid-19 data had any significance in predicting the rate of deaths due to Covid-19. The Google mobility data recorded travel trends to categorized locations during the Covid-19 pandemic. This data is compared against a baseline reading; that is, the median value of each day of the week during a 5-week period (Jan 3 – Feb 6, 2020).

Variables in the original COVID-19 Data Hub data set:
date, confirmed, tests, population, latitude, longitude, school_closing, workplace_closing, cancel_events, transport_closing, stay_home_restrictions, internal_movement_restrictions, international_movement_restrictions, information_campaigns, testing_policy, contact_tracing, stringency_index

Variables used on top of base data set:
World Bank data set: GDP per capita, GDP per capita growth, Poverty rate, Pollution in mcg
Google Mobility data set: retail_and_recreation_percent_change_from_baseline, grocery_and_pharmacy_percent_change_from_baseline, parks_percent_change_from_baseline, transit_stations_percent_change_from_baseline, workplaces_percent_change_from_baseline, residential_percent_change_from_baseline

Questions of Interest

1. What model using the contact tracing is the best predictor of deaths? We plan to use *deaths* as the response and *confirmed*, *tests*, *contact tracing*, and *stringency index* from in the original COVID-19 data set as predictors to answer this question.

2. How does the economic profile of the country affect the mortality rate from COVID over the year 2020? We plan to use *deaths* as the response; *confirmed*, *tests*, *contact tracing*, and *stringency index* from in the original COVID-19 data set; and *GDP per capita*, *GDP per capita growth*, and *Poverty rate* from the World Bank data set to answer this question.

3. What is the effect of air pollution (or exposure to air pollution) to the number of cases and the mortality rate from COVID? We plan to use *deaths* as the response; *confirmed*, *tests*, *contact tracing*, and *stringency index* from in the original COVID-19 data set; and *GDP per capita*, *GDP per capita growth*, and *Pollution in mcg* from the World Bank data set to answer this question.

4. Using the Google Mobility Data, are policy measures that are non-restrictive with movement significant in preventing spread of Covid-19? We plan to use *deaths* as the response; *confirmed*, *tests*, *contact tracing*, and *stringency index* from in the original COVID-19 data set; and *retail and recreation percent change from baseline*, *grocery and pharmacy percent change from baseline*, *parks percent change from baseline*, *transit stations percent change from baseline*, *workplaces percent change from baseline*, *residential percent change from baseline* from the Google Mobility data set to answer this questions.

5. Using the Google Mobility Data, are policy measures that are restrictive with movement more significant than non-restrictive measures in preventing the spread of COVID-19 - Response: deaths - Predictors: looking at both movement restrictive and non-restrictive and comparing their significance

Regression Analysis, Results and Interpretation

Important Details

We did a hypothesis test to determine whether or not we had any significant variables in our full model to start. We used $H_0 : \beta_1 = \beta_2 = ... = \beta_9 = 0$ and H_1 : At least one $\beta_i \neq 0$ for $i = 1, 2, ..., 9$.

```
## Analysis of Variance Table
##
## Model 1: deaths ~ 1
## Model 2: deaths ~ confirmed + tests + fschool_closing + fworkplace_closing +
##     fgatherings_restrictions + fstay_home_restrictions + ftesting_policy +
##     fcontact_tracing + stringency_index
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    364 8.1640e+10
## 2    352 4.7587e+08 12 8.1164e+10 5003.1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# full fitted model w/ removed predictor variables
google.full <- lm(deaths ~ retail_and_recreation_percent_change_from_baseline + grocery_and_pharmacy_pe

google.null <- lm(deaths ~ 1, data = cacovid_mobility)
```

```
anova(google.null, google.full)
```

```
## Analysis of Variance Table
##
## Model 1: deaths ~ 1
## Model 2: deaths ~ retail_and_recreation_percent_change_from_baseline +
##     grocery_and_pharmacy_percent_change_from_baseline + parks_percent_change_from_baseline +
##     transit_stations_percent_change_from_baseline + workplaces_percent_change_from_baseline +
##     residential_percent_change_from_baseline + date + confirmed +
##     tests + fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + ftesting_policy + fcontact_tracing +
##     stringency_index
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  12972 3.0456e+12
## 2  12953 1.6726e+10 19 3.0288e+12 123455 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Exploratory Analysis I:

We started by creating a data frame by filtering the data first by United States of America, secondly by California, and finally by date. We ended with 365 rows of data for California. There is not any data for vaccines, recovered, hosp, vent, and icu so we removed these variables. We decided to use the following variables to create a new data frame date, *tests*, *confirmed*, *recovered*, *deaths*, *hosp*, *vent*, *icu*, *latitude*, *longitude*, *population*, *vaccines*, *school_closing*, *workplace_closing*, *cancel_events*, *gatherings_restrictions*, *transport_closing*, *stay_home_restrictions*, *internal_movement_restrictions*, *international_movement_restrictions*, *information_campaigns*, *testing_policy*, *contact_tracing*, and *stringency_index*. Since we don't have data for *vaccines*, *recovered*, *hosp*, *vent*, and *icu* we removed these variables. All variables now have data in every row. We then turned the policy measures (categorical variables) into factors before we fitted a regression model. However factors need to have 2 or more levels in order to work so we also removed *cancel_events*, *international_movement*, and *transport_closing*. We also removed *internal_movement_restrictions*, *information_campaigns*, *population*, *longitude*, and *latitude* as they have the same data for every row causing a singularity in the data. This left us with a base data set of *confirmed*, *tests*, *fschool_closing*, *fworkplace_closing*, *fgatherings_restrictions*, *fstay_home_restrictions*, *ftesting_policy*, *fcontact_tracing*, and *stringency_index* as the predictors to start looking for a linear regression model with.

We started by running a hypothesis test to see if we would prefer the null model against the full model. From Table 1 in Appendix 1, we can see that the p-value is $<$ 2.2e-16 so we reject the null model as at least one predictor in the full model is significant.

Next we looked at the scatter plots for all of the variables. This was less useful since there were so many plots that it was difficult to see in detail (Plot 1, Appendix 1) so we looked at the scatter plots of the numerical data (Plot 2, Appendix 1) to see if there was anything that we could derive from the data. From this we can see that confirmed and tests both have a positive linear relationship with deaths. This would lead us to assume that confirmed and tests would be a positive influence on the number of deaths. The stringency index has a clear patterning to it that does not show any linear trends making it difficult to make any assumptions about it. We also should note that none of the variables are spread out. The data creates a line with the data points we we will definitely need to transform this data to see if we can find a linear relationship. We then did an analysis of the categorical data using box plots (Plot 3, Appendix 1). From these we concluded that none of these variables have a constant variance and thus we will probably need to transform some if not all of the variables.

We also looked at the added variable plots (Plot 4, Appendix 1) and summary to see if we should remove any variables. From the added variable plots we assumed that we will probably remove testing policy, confirmed, and stringency index. While gathering restrictions 2 and 3 look like they should be removed, gathering restrictions 4 looks to have some influence and therefore we chose to keep the gathering restrictions. However, the summary table shows us that confirmed and gathering restrictions will probably be removed. We may keep testing policy since only one of the dummy variables is not significant.

Next we did a variable selection using, AIC and BIC stepwise selection (see Code 1-2, Appendix 1). We can see that the only variable that is not significant is a dummy variable and we cannot remove it without removing a significant variable so we leave it in and we are left with a model of

$$deaths = \beta_0 + \beta_1 fstay.home.restrictions + \beta_2 tests + \beta_3 fworkplace.closing + \beta_4 fschool.closing + \beta_5 fcontact.tracing + \beta_6 ftesting.policy + \beta_7 stringency.index$$

We looked the residuals vs fitted and Q-Q plot (Plot 5, Appendix 1) to see if the linear assumptions were violated and to check to see if there were any outliers and/or leverage points. From the plots we can see that there is definite patterning in the residuals vs fitted plot and the Q-Q plot is heavy tailed showing violations of normality. We then run a powerTransform (Code 3, Appendix 1) on the numerical predictors and see that *tests* needs a cube root transformation and *stringency index* needs a logarithmic transformation. After we transform these predictors we checked to see if we needed to transform *deaths* and can see that the boxCox (Plot 6, Appendix 1) suggests a cube root transformation.

After transformation we looked at the residual vs fitted and Q-Q plot again to check for linearity and check to see if there are still outliers. When we look we see that the residuals vs fitted plot is still very patterned and the Q-Q plot is still heavy-tailed but less so (Plot 7, Appendix 1). However we can still see that there is probably an outlier from the Cook's Distance plot (Plot 8, Appendix 1) so next we looked for outliers and leverage points (Code 4, Appendix 1). We removed the outliers and leverage points one at a time and stopped after removing rows 1, 2, 33, 34, and 35 (Code 5, Appendix 1). When we checked the diagnostics the Q-Q plot (Plot 9, Appendix 1) improved a little and the Cook's Distance plot again (Plot 10, Appendix 1) we were much happier with this result.

We checked the summary (Table 4, Appendix 1) again and saw that *fworkplace_gathering_restrictions* was no longer significant and so we removed it and checked the linear assumptions again (Plot 11, Appendix 1). There did not appear to be a difference between them. When we did a hypothesis test to see if we preferred the H_0 : mod.full6 or H_1 : mod.full5 (Code 6, Appendix 1), the p-value is 0.9087 showing us that we prefer the smaller model.

Diagnostic Checks I:

Interpretation I:

Exploratory Analysis II:

Full Model (Base Covid + Google Mobility) After working solely with the base Covid data, we decided to add in the Google mobility data. First, we read in and subset the Google mobility data. The data only included reports in CA and ranged from Mar 13, 2020 to Mar 14, 2021. We also took out 4 columns of data that were identifiers and not relevant for our data analysis. Lastly, we removed all rows with at least one NA and converted all data from percentages to decimals. After changing the Google mobility data, we merged the modified base Covid data and Google mobility data into one dataframe. The Google mobility variables added as predictors are *retail_and_recreation_percent_change_from_baseline*, *grocery_and_pharmacy_percent_change_from_baseline*, *parks_percent_change_from_baseline*, *transit_stations_percent_change_from_baseline*, *workplaces_percent_change_from_baseline*, and *residential_percent_change_from_baseline*. The modified base Covid data set included all variables with values in every row.

We started off by running a linear model summary of the the full model (Table 1, Appendix 2). We saw that the there were variables that were singularities (i.e. *longitude*, *latitude*, *population*,

finternal_movement_restrictions2, and *finformation_campaigns2*) so we removed these from the model as well. Next, we ran an ANOVA comparing the modified full model with the null model (Table 2, Appendix 2). The resulting p-value was <2.2e-16 so we reject the null model and conclude that there is at least one predictor variable in the full model that is significant. The next step we took was to check the QQ plot and residuals vs. fitted plot (Plot 1, Appendix 2). Visually, we saw that it did not meet the assumptions of linearity. The QQ plot did not follow a linear trend and the residuals vs. fitted plot showed obvious patterning.

Because the current model did not meet our assumptions of linearity, we decided to run a variable selection to help us narrow down significant predictors (Table 3, Appendix 2). We compared all eight models by looking at adjusted R-squared, CP values, and Bic values (Table 4, Appendix 2). The model we chose had seven predictor variables, but still did not show linearity (Plot 2, Appendix 2). Thus, we decided to check if there were any necessary transformations for the predictors.

We ran transformations of all non-factor predictors. This resulted in a square root transformation for *confirmed* (Table 5, Appendix 2). We also ran transformations for the response variable which resulted in a square root transformation for *deaths* (Plot 3, Appendix 2).

After transforming the model, we decided to check for outliers and high leverage points. We set $|r_i| > 2$ to identify outliers. Plotting the data showed that there were no bad leverage points (Plot 4, Appendix 2). Also, we decided to check for high Cook's distance values of which none were greater than 0.5 (Plot 5, Appendix 2). As a result, we decided not to remove any data points.

Our final model after variable selection and transformations does not include any Google mobility predictors. The response is $\sqrt{deaths}$ and the predictors are $\sqrt{tests}$, fschool_closing, fworkplace_closing, fstay_home_restrictions, ftesting_policy, and fcontact_tracing. Checking the final diagnostics plots, we see that the assumptions of linearity are not met (Plot 6, Appendix 2).

For our next model, we decided to remove all base Covid numerical data from the original data set. The model included categorical base Covid data and Google mobility data. After exploring this model, we came to the same conclusion as the previous model, that assumptions of linearity were not met. Because of these results, we decided to remove all base Covid data and work solely with Google mobility data.

Diagnostic Checks II:

And this plot shows us that we have a linear model and all of the assumptions are met as well as our data can.

We looked at adding a categorical variable to our model however the linear assumptions kept getting worse so we chose to leave the categorical variables out of our model and changed our questions to match this.

Interpretation II:

$Y_{ijk} = \beta_1 var1 + \beta_2 var2$

Exploratory Analysis III:

Due to the base data not being linear we decided to add but...

Interpretation

Conclusions (200 words) - Thomas

The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young

man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like.

The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like. The young man wanted a role model. He looked long and hard in his youth, but that role model never materialized. His only choice was to embrace all the people in his life he didn't want to be like.

Appendicies

Appendix 1: R Code for Original COVID-19 Data Set

Code 1: The AIC model kept fstay_home_restrictions, tests, fworkplace_closing, fschool_closing, fcontact_tracing, ftesting_policy, and stringency_index as the predictors in the ideal model. data

```
(step_aic <- step(mod.0, scope = list(lower = mod.0, upper = mod.full), trace = 0))
```

```
##
## Call:
## lm(formula = deaths ~ fstay_home_restrictions + tests + fworkplace_closing +
##     fschool_closing + fcontact_tracing + ftesting_policy + stringency_index,
##     data = fbase_data)
##
## Coefficients:
##              (Intercept)  fstay_home_restrictions2                    tests
##                -6.026e+03                -6.378e+03                1.133e-03
##      fworkplace_closing2       fworkplace_closing3          fschool_closing3
##                 5.127e+03                 8.405e+03                -5.347e+03
##        fcontact_tracing2          ftesting_policy2          ftesting_policy3
##                -1.676e+03                 9.524e+02                -2.803e+01
##         stringency_index
##                 1.162e+02
```

Code 2: The BIC is the same as the AIC, we chose to use BIC.

```
(step_bic <- step(mod.0, scope = list(lower = mod.0, upper = mod.full), trace = 0))
```

```
##
## Call:
## lm(formula = deaths ~ fstay_home_restrictions + tests + fworkplace_closing +
##     fschool_closing + fcontact_tracing + ftesting_policy + stringency_index,
##     data = fbase_data)
##
## Coefficients:
##              (Intercept)  fstay_home_restrictions2                    tests
##                -6.026e+03                -6.378e+03                1.133e-03
##      fworkplace_closing2       fworkplace_closing3          fschool_closing3
##                 5.127e+03                 8.405e+03                -5.347e+03
```

```
##        fcontact_tracing2          ftesting_policy2          ftesting_policy3
##                -1.676e+03                 9.524e+02                -2.803e+01
##         stringency_index
##                 1.162e+02
```

Code 3: The powerTransformation suggests we use a cube root transformation on *tests* and a logarithmic transformation on stringency_index.

```
pt <- powerTransform(cbind(tests, stringency_index) ~ 1, data = fbase_data)
summary(pt)
```

```
## bcPower Transformations to Multinormality
##                   Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## tests               0.2715        0.27       0.2178       0.3252
## stringency_index   -0.4417        0.00      -1.0026       0.1193
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                               LRT df       pval
## LR test, lambda = (0 0) 135.8532  2 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                               LRT df       pval
## LR test, lambda = (1 1) 420.7194  2 < 2.22e-16
```

Code 4: There are two leverage points and 3 outliers.

```
mod.full4 <- lm((deaths^{1/3}) ~ fstay_home_restrictions + I(tests^{1/3}) + fworkplace_closing +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data)
# leverage point calculations
p <- 10
n <- nrow(fbase_data)
mod.full4_hat <- hatvalues(mod.full4)

leverage <- which(mod.full4_hat > 4*(p+1)/n)

# Find outliers
mod.full4_out <- rstandard(mod.full4)
outliers<- which(abs(mod.full4_out) > 3)

# Cook's points and hat values
mod.full4.cooks <- cooks.distance(mod.full4)
cooks <- which(mod.full4.cooks > 4/(n-p-1))

leverage
```

```
## 1 2
## 1 2
```

```
outliers
```

```
## 33 34 35
## 33 34 35
```

```
cooks
```

```
##   1   2   3   4   5   6   7   8   9  10  11  29  30  31  32  33  34  35 352 353
##   1   2   3   4   5   6   7   8   9  10  11  29  30  31  32  33  34  35 352 353
## 354 355
## 354 355
```

Code 5: Remove all leverage and outliers from fbase_data.

```
fbase_data1<-fbase_data[-c(1,2,33,34,35), ]
```

Code 6:

```
## Analysis of Variance Table
##
## Model 1: (deaths^{
##     1/3
## }) ~ fstay_home_restrictions + I(tests^{
##     1/3
## }) + fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index)
## Model 2: (deaths^{
##     1/3
## }) ~ fstay_home_restrictions + I(tests^{
##     1/3
## }) + fworkplace_closing + fschool_closing + fcontact_tracing +
##     ftesting_policy + log(stringency_index)
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    352 103.02
## 2    351 103.02  1 0.0038684 0.0132 0.9087
```

Table 1: $H_0 : \beta_1 = \beta_2 = ... = \beta_9 = 0$
H_1 : At least one $\beta_i \neq 0$ for $i = 1, 2, ..., 9$

```
## Analysis of Variance Table
##
## Model 1: deaths ~ 1
## Model 2: deaths ~ confirmed + tests + fschool_closing + fworkplace_closing +
##     fgatherings_restrictions + fstay_home_restrictions + ftesting_policy +
##     fcontact_tracing + stringency_index
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    364 8.1640e+10
## 2    352 4.7587e+08 12 8.1164e+10 5003.1 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 2: This summary table shows us that confirmed and gathering restrictions will probably be removed. We may keep testing policy since only one of the dummy variables is not significant.

```
summary(mod.full)
```

```
##
## Call:
## lm(formula = deaths ~ confirmed + tests + fschool_closing + fworkplace_closing +
##     fgatherings_restrictions + fstay_home_restrictions + ftesting_policy +
##     fcontact_tracing + stringency_index, data = fbase_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3917.7  -536.0  -114.1   506.1  4360.5
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                -6.059e+03  3.202e+03  -1.892  0.05927 .
## confirmed                  -3.993e-04  9.071e-04  -0.440  0.66009
## tests                       1.169e-03  8.586e-05  13.614  < 2e-16 ***
## fschool_closing3           -5.411e+03  5.241e+02 -10.325  < 2e-16 ***
## fworkplace_closing2         6.002e+03  1.330e+03   4.511 8.79e-06 ***
## fworkplace_closing3         9.029e+03  1.649e+03   5.476 8.29e-08 ***
## fgatherings_restrictions3 -2.429e+03  1.921e+03  -1.265  0.20685
## fgatherings_restrictions4 -2.317e+03  2.002e+03  -1.157  0.24806
## fstay_home_restrictions2  -6.471e+03  2.440e+02 -26.518  < 2e-16 ***
## ftesting_policy2           9.374e+02  3.378e+02   2.775  0.00582 **
## ftesting_policy3           1.705e+01  4.664e+02   0.037  0.97085
## fcontact_tracing2          -1.732e+03  3.069e+02  -5.643 3.44e-08 ***
## stringency_index           1.391e+02  6.855e+01   2.030  0.04314 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1163 on 352 degrees of freedom
## Multiple R-squared:  0.9942, Adjusted R-squared:  0.994
## F-statistic:  5003 on 12 and 352 DF,  p-value: < 2.2e-16
```

Table 3: This is a summary of the model found after BIC Stepwise Selection.

```
summary(step_bic)
```

```
##
## Call:
## lm(formula = deaths ~ fstay_home_restrictions + tests + fworkplace_closing +
##     fschool_closing + fcontact_tracing + ftesting_policy + stringency_index,
##     data = fbase_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3925.3  -548.4  -118.5   559.9  4361.3
##
## Coefficients:
```

```
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -6.026e+03  1.972e+03  -3.056 0.002410 **
## fstay_home_restrictions2 -6.378e+03  2.051e+02 -31.101  < 2e-16 ***
## tests                     1.133e-03  1.147e-05  98.840  < 2e-16 ***
## fworkplace_closing2       5.127e+03  1.092e+03   4.695 3.82e-06 ***
## fworkplace_closing3       8.405e+03  1.217e+03   6.905 2.33e-11 ***
## fschool_closing3         -5.347e+03  4.619e+02 -11.576  < 2e-16 ***
## fcontact_tracing2        -1.676e+03  2.356e+02  -7.112 6.34e-12 ***
## ftesting_policy2          9.524e+02  3.350e+02   2.843 0.004723 **
## ftesting_policy3         -2.803e+01  4.298e+02  -0.065 0.948043
## stringency_index          1.162e+02  3.457e+01   3.360 0.000863 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1161 on 355 degrees of freedom
## Multiple R-squared:  0.9941, Adjusted R-squared:  0.994
## F-statistic:  6691 on 9 and 355 DF,  p-value: < 2.2e-16
```

```
summary(mod.full5)
```

Table 4:

```
##
## Call:
## lm(formula = (deaths^{
##     1/3
## }) ~ fstay_home_restrictions + I(tests^{
##     1/3
## }) + fworkplace_closing + fschool_closing + fcontact_tracing +
##     ftesting_policy + log(stringency_index), data = fbase_data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23640 -0.31719 -0.04829  0.21613  1.90884
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -21.143270   4.685643  -4.512 8.76e-06 ***
## fstay_home_restrictions2  -1.875099   0.095475 -19.640  < 2e-16 ***
## I(tests^{\n    1/3\n})     0.089391   0.000846 105.659  < 2e-16 ***
## fworkplace_closing3       -0.030021   0.261497  -0.115    0.909
## fschool_closing3          -2.000958   0.217052  -9.219  < 2e-16 ***
## fcontact_tracing2         -1.612712   0.092226 -17.487  < 2e-16 ***
## ftesting_policy2           3.274730   0.162020  20.212  < 2e-16 ***
## ftesting_policy3           3.402234   0.216109  15.743  < 2e-16 ***
## log(stringency_index)      6.000421   1.126961   5.324 1.81e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5418 on 351 degrees of freedom
## Multiple R-squared:  0.996,  Adjusted R-squared:  0.996
## F-statistic: 1.106e+04 on 8 and 351 DF,  p-value: < 2.2e-16
```

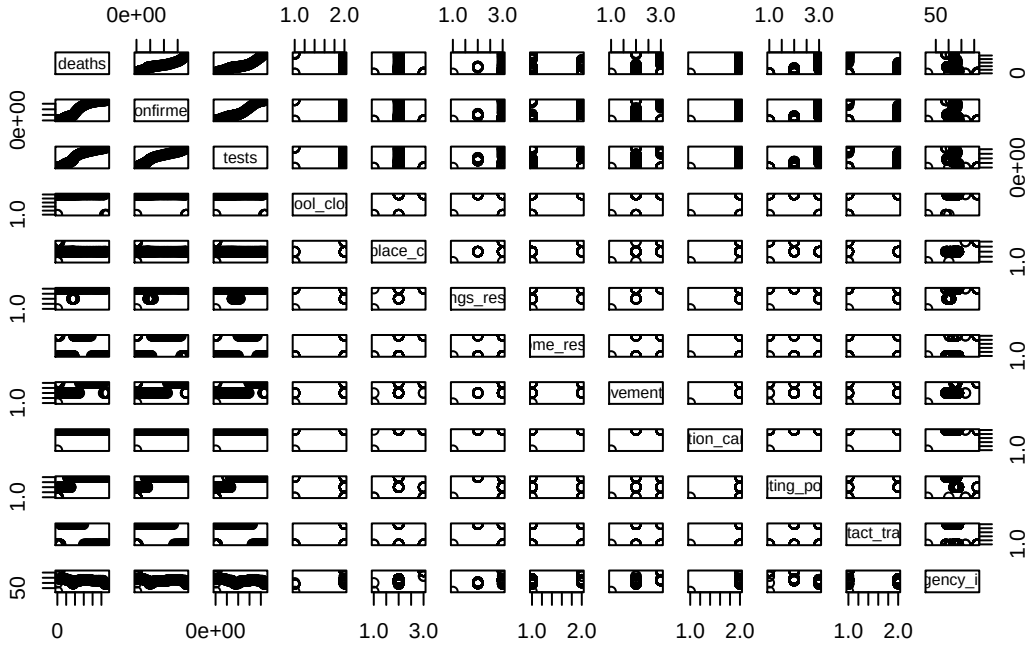
Table 5: This is our final model.

```
mod.full6 <- lm((deaths^{1/3}) ~ fstay_home_restrictions + I(tests^{1/3}) +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data1)
summary(mod.full6)
```

```
##
## Call:
## lm(formula = (deaths^{
##     1/3
## }) ~ fstay_home_restrictions + I(tests^{
##     1/3
## }) + fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
##     data = fbase_data1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.23530 -0.31336 -0.04861  0.21695  1.90872
##
## Coefficients:
##                             Estimate Std. Error t value Pr(>|t|)
## (Intercept)               -2.085e+01  3.928e+00  -5.308 1.97e-07 ***
## fstay_home_restrictions2 -1.877e+00  9.370e-02 -20.034  < 2e-16 ***
## I(tests^{\n    1/3\n})    8.943e-02  7.853e-04 113.871  < 2e-16 ***
## fschool_closing3          -1.992e+00  2.034e-01  -9.796  < 2e-16 ***
## fcontact_tracing2         -1.611e+00  9.136e-02 -17.638  < 2e-16 ***
## ftesting_policy2           3.281e+00  1.527e-01  21.481  < 2e-16 ***
## ftesting_policy3           3.402e+00  2.158e-01  15.765  < 2e-16 ***
## log(stringency_index)      5.925e+00  9.165e-01   6.465 3.38e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.541 on 352 degrees of freedom
## Multiple R-squared:  0.996,  Adjusted R-squared:  0.996
## F-statistic: 1.267e+04 on 7 and 352 DF,  p-value: < 2.2e-16
```

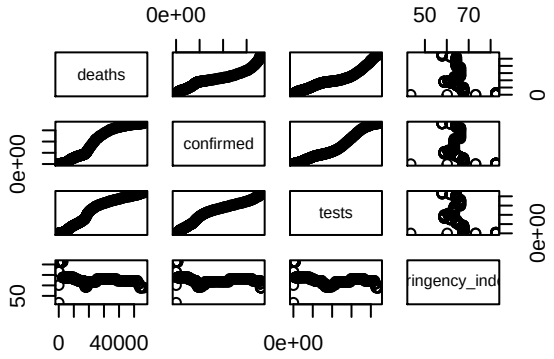
Plot 1: Each of these are really small and it is hard to derive anything useful from them.

```
pairs(deaths ~ confirmed + tests + fschool_closing +
                fworkplace_closing + fgatherings_restrictions + fstay_home_restrictions +
                finternal_movement_restrictions  +finformation_campaigns + ftesting_policy +
                fcontact_tracing + stringency_index, data = fbase_data)
```



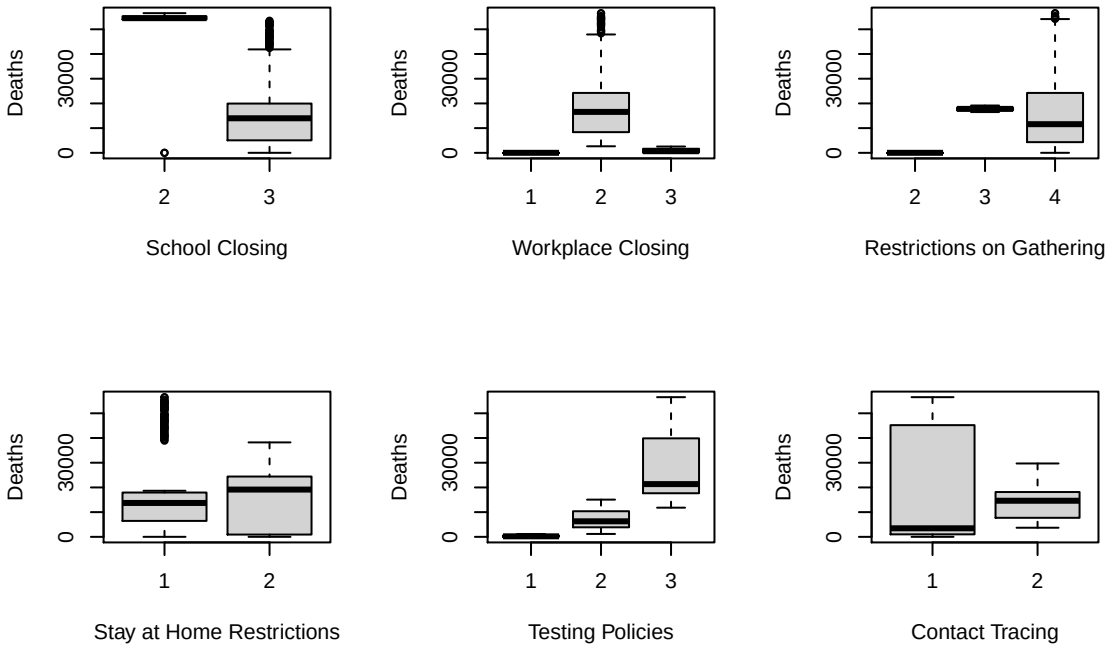
Plot 2: Scatterplots of the numerical variables in the original data set.

```
pairs(deaths ~ confirmed + tests +  stringency_index, data = fbase_data)
```



Plot 3: Box plots of the categorical variables in the original data set.

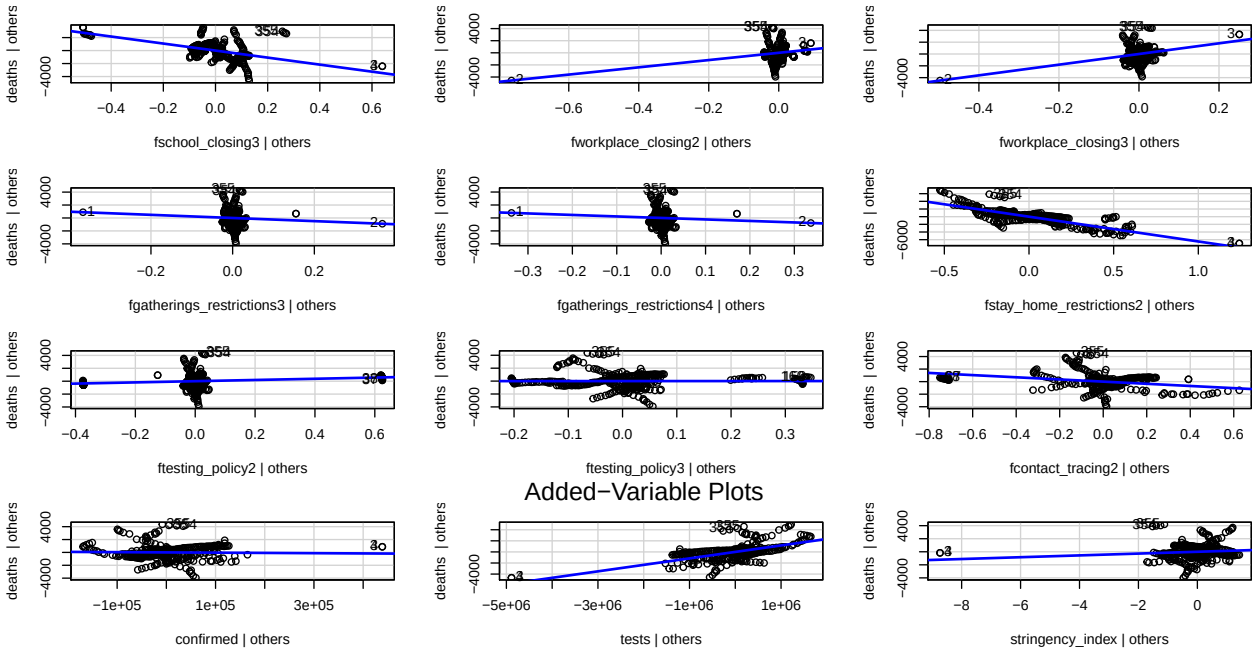
```
par(mfrow=c(2,3))
boxplot(deaths ~ fschool_closing, data = fbase_data, ylab = "Deaths", xlab = "School Closing")
boxplot(deaths ~ fworkplace_closing, data = fbase_data, ylab = "Deaths", xlab = "Workplace Closing")
boxplot(deaths ~  fgatherings_restrictions, data = fbase_data, ylab = "Deaths", xlab = "Restrictions on
boxplot(deaths ~  + fstay_home_restrictions, data = fbase_data, ylab = "Deaths", xlab = "Stay at Home R
boxplot(deaths ~  + ftesting_policy, data = fbase_data, ylab = "Deaths", xlab = "Testing Policies")
boxplot(deaths ~  + fcontact_tracing, data = fbase_data, ylab = "Deaths", xlab = "Contact Tracing")
```



Plot 4: Added variable plots for both the categorical variables.

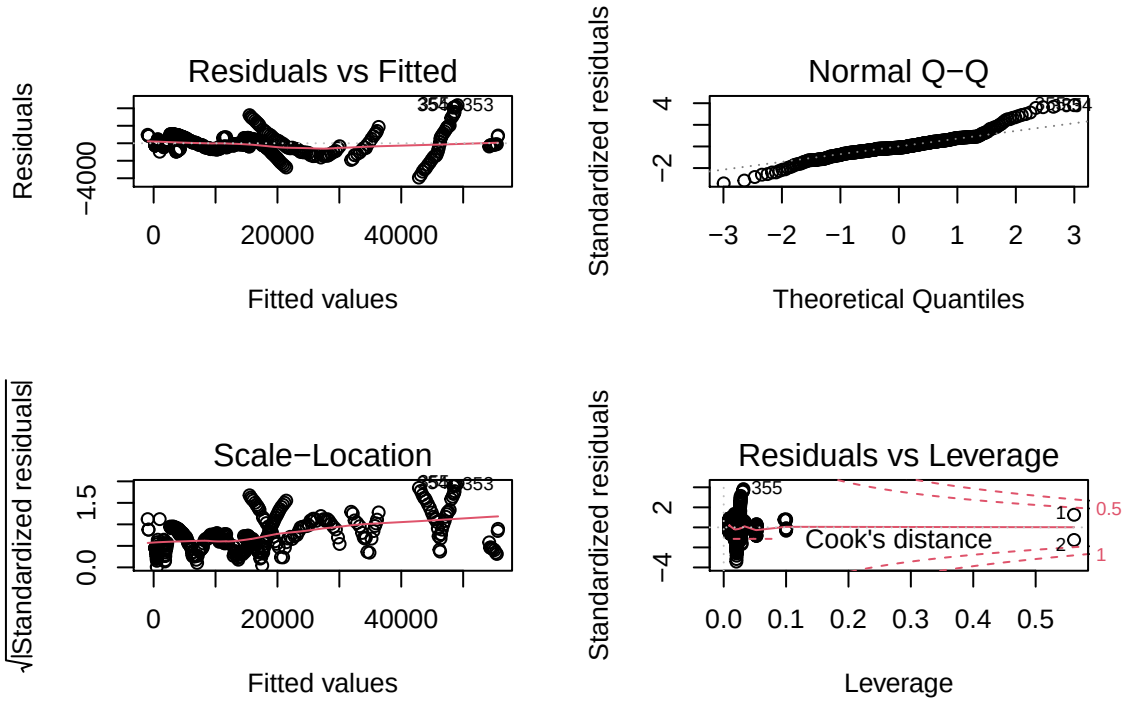
```
mod.full1 <- lm(deaths ~ fschool_closing +
                fworkplace_closing + fgatherings_restrictions + fstay_home_restrictions +
                ftesting_policy + fcontact_tracing + confirmed + tests + stringency_index, data = fbas

avPlots(mod.full1)
```



Plot 5: Checking normality prior to transformation

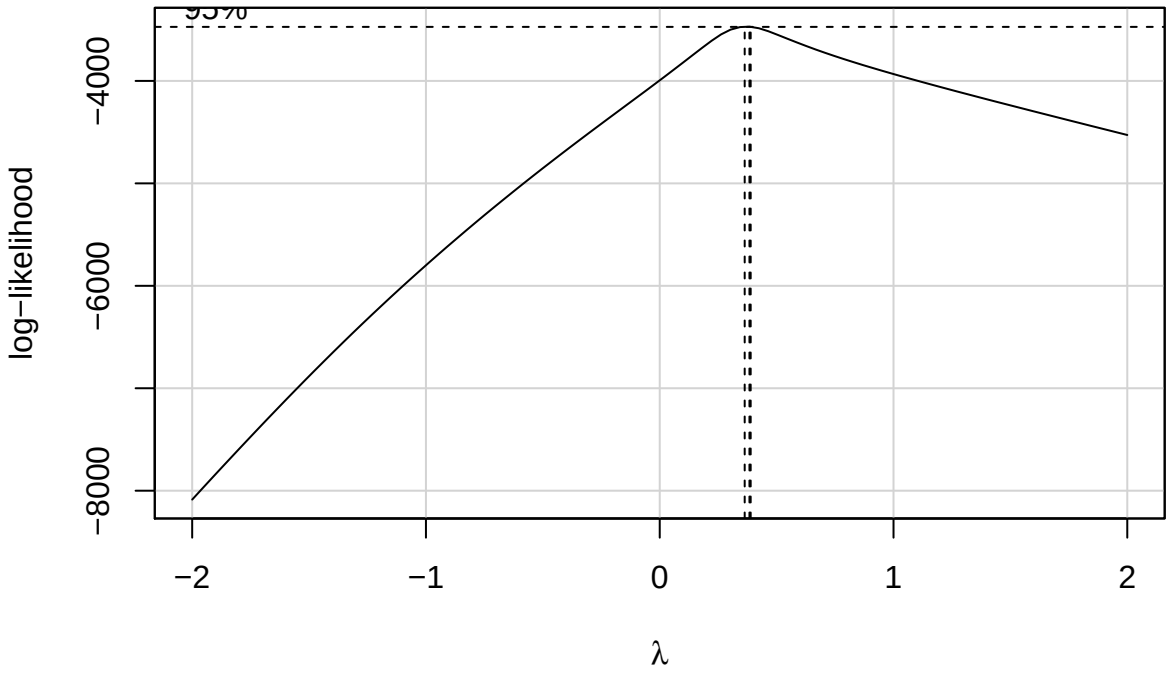
```
mod.full2 <- lm(deaths ~ fstay_home_restrictions + tests + fworkplace_closing +
    fschool_closing + fcontact_tracing + ftesting_policy + stringency_index,
    data = fbase_data)
par(mfrow=c(2,2))
plot(mod.full2)
```



Plot 6: According to boxCox we should do a cube root transformation on the response *deaths*.

```
mod.full3 <- lm(deaths ~ fstay_home_restrictions + I(tests^{1/3}) + fworkplace_closing +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data)

bcTrans <- boxCox(mod.full3)
```

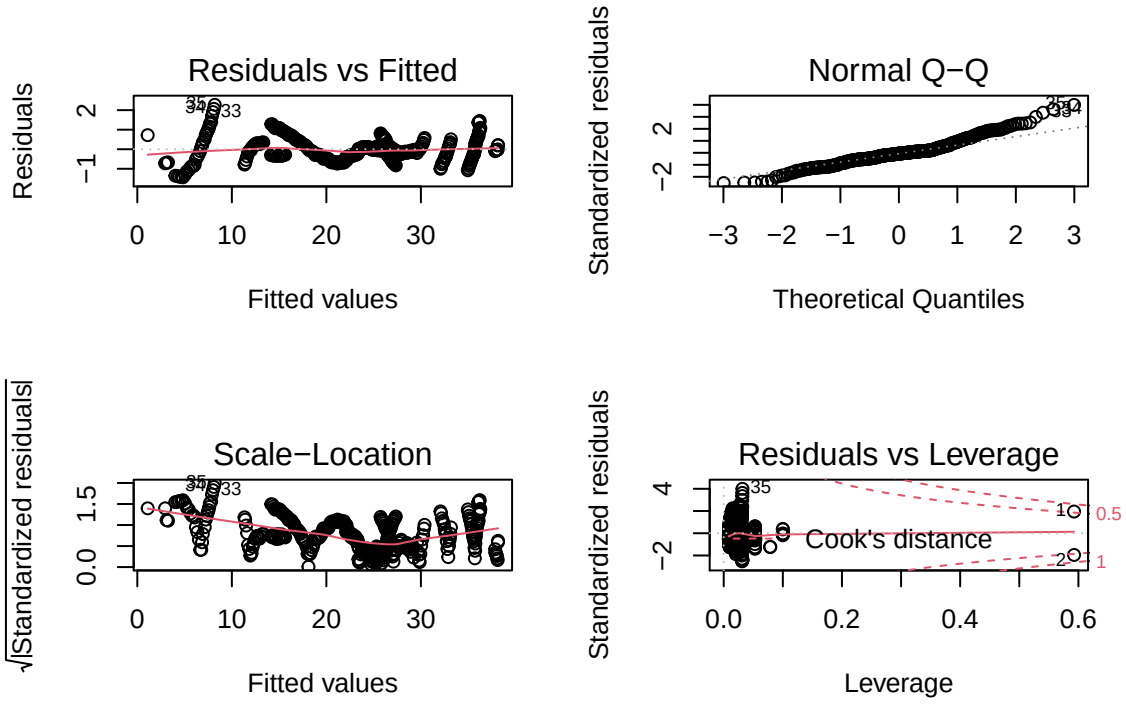



```
opt.lambda <- bcTrans$x[which.max(bcTrans$y)]
opt.lambda
```

```
## [1] 0.3838384
```

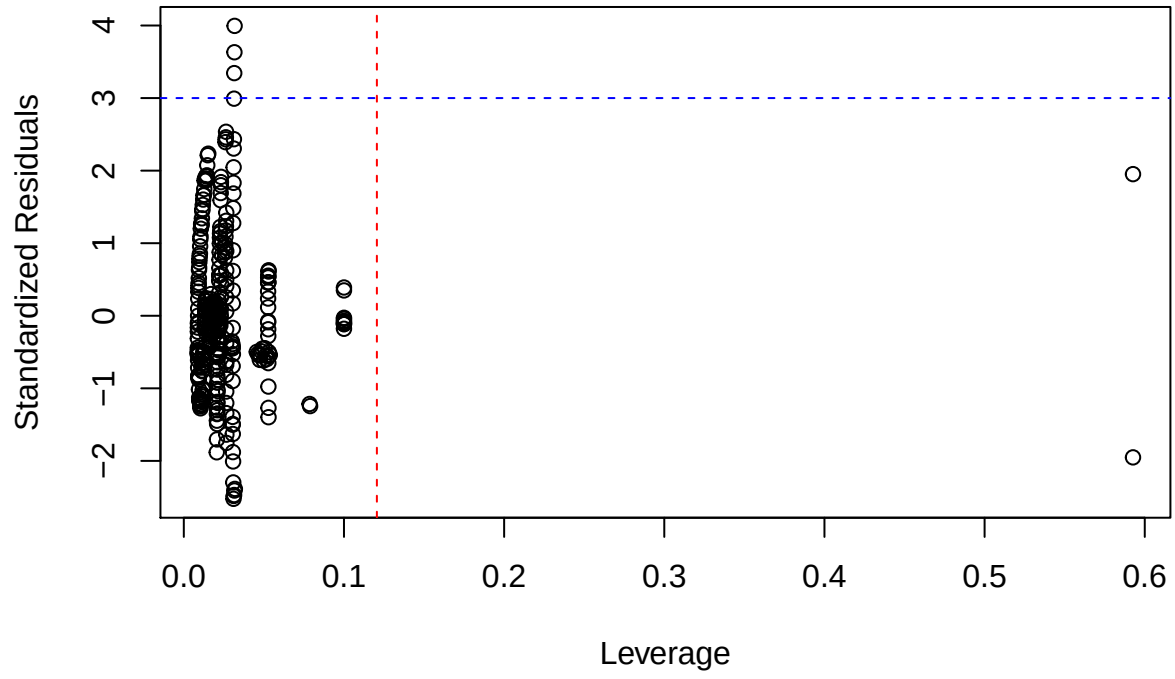
Plot 7: Post transformation linearity check. The residuals vs fitted plot is still very patterned and the Q-Q plot is still heavy-tailed but less so.

```
mod.full4 <- lm((deaths^{1/3}) ~ fstay_home_restrictions + I(tests^{1/3}) + fworkplace_closing +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data)
par(mfrow=c(2,2))
plot(mod.full4)
```

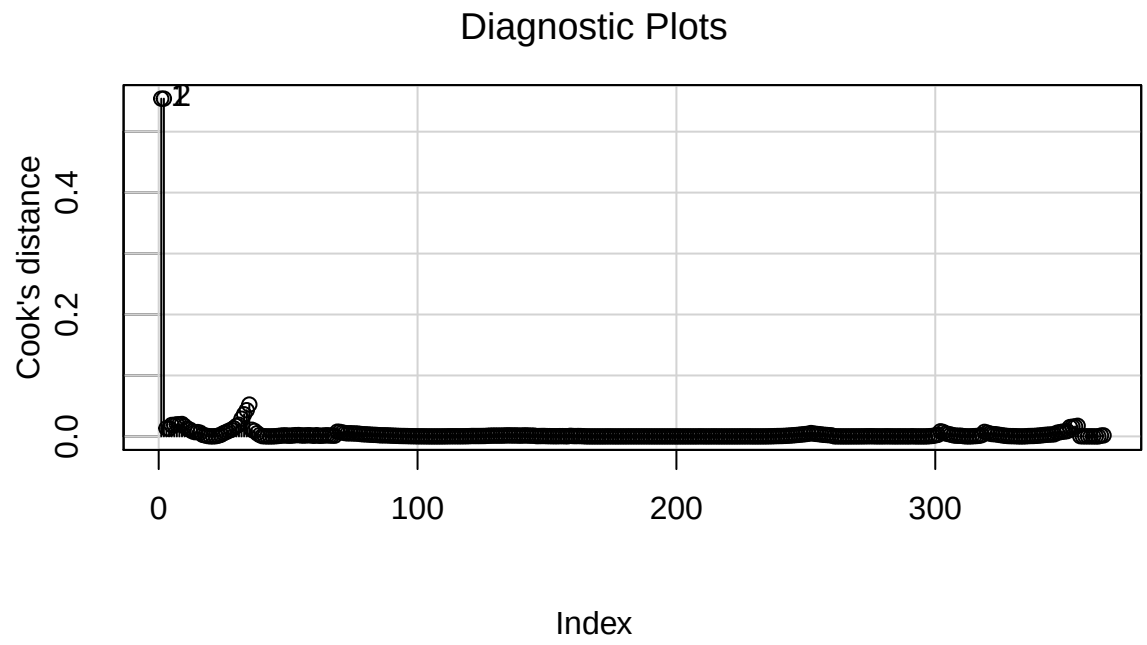


Plot 8: From the Cook's InfuleceIndexPlot and hat-values influenceIndexPlot we can see that we should definitely look at points 1 and 2.

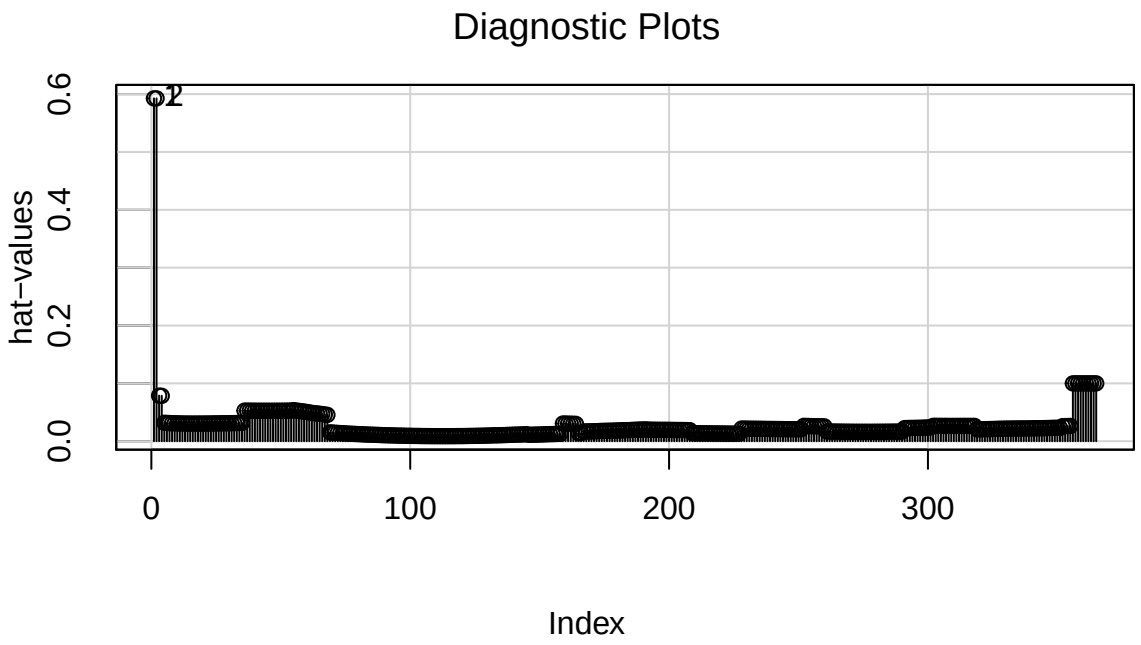
```
# plot of high leverage points and outliers
plot(hatvalues(mod.full4), rstandard(mod.full4), xlab = "Leverage",
     ylab = "Standardized Residuals")
abline(v = 4*(p+1)/n, col = "red", lty = 2)
abline(h = c(-3,3), col = "blue", lty =2)
```



```r
# Cook's points and hat values
mod.full4.cooks <- cooks.distance(mod.full4)
influenceIndexPlot(mod.full4, vars = "Cook")
```



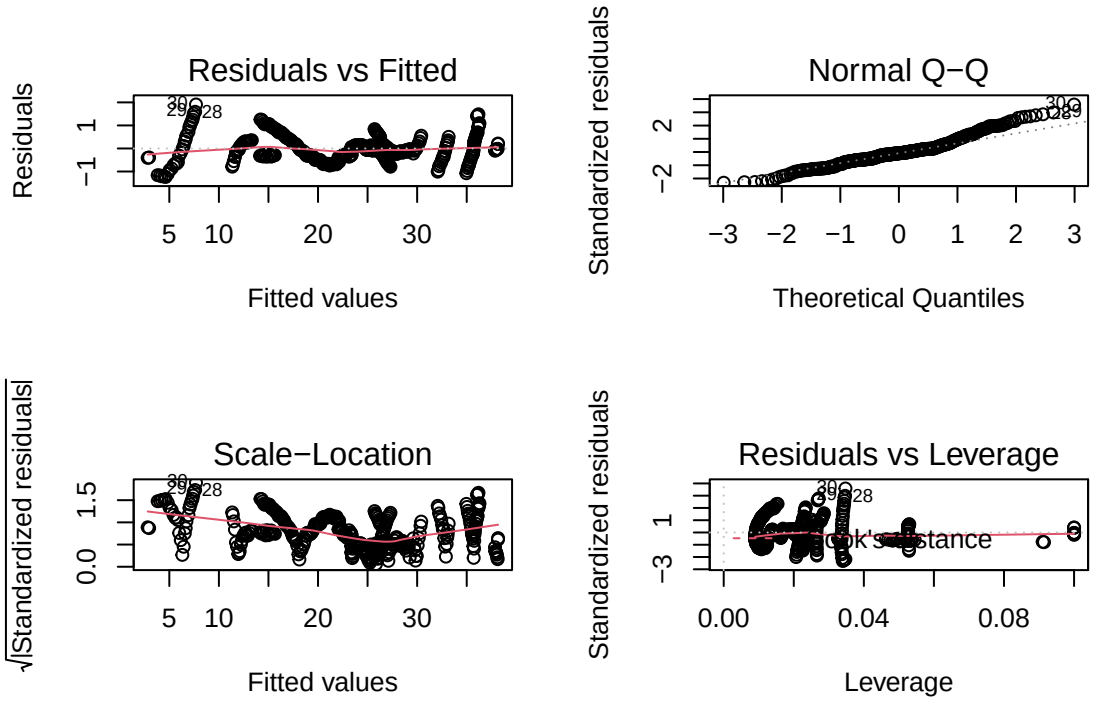

```r
influenceIndexPlot(mod.full4, vars = "hat")
```



Plot 9: Checking the linearity after the removal of rows 1, 2, 33, 34, and 35. Looks a little better.

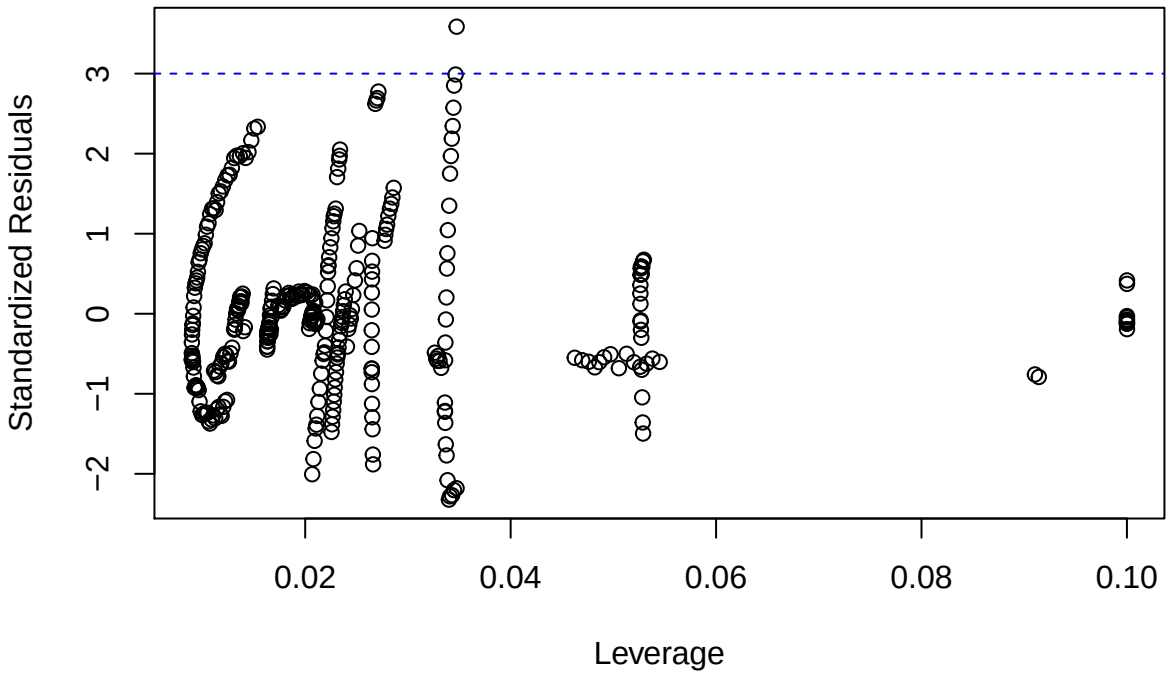
```r
mod.full5 <- lm((deaths^{1/3}) ~ fstay_home_restrictions + I(tests^{1/3}) + fworkplace_closing +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data1)
```

```
par(mfrow=c(2,2))
plot(mod.full5)
```

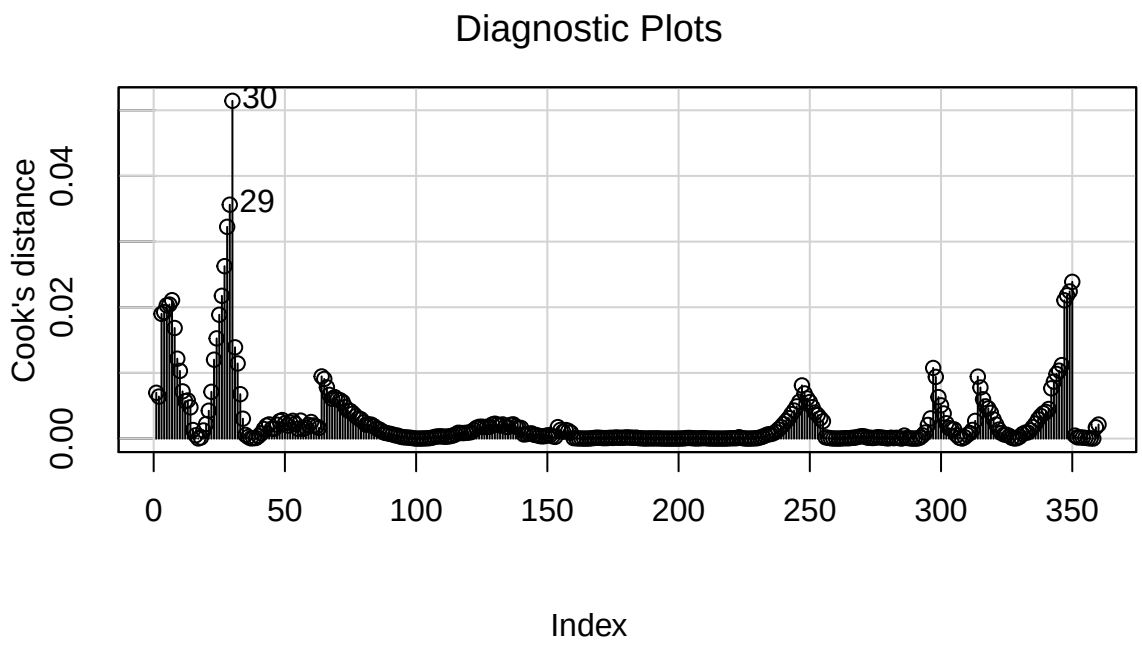


Plot 10: Checking Cook's plot again after the removal of rows 1, 2, 33, 34, and 35. Looks better.

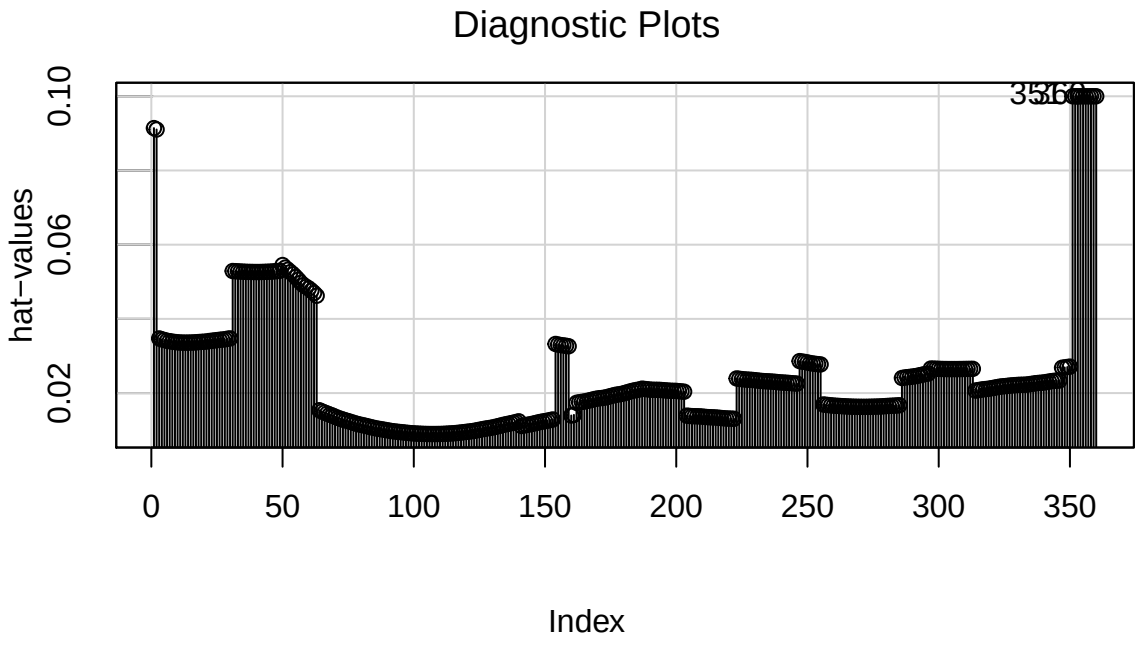
```
# plot of high leverage points and outliers
plot(hatvalues(mod.full5), rstandard(mod.full5), xlab = "Leverage",
     ylab = "Standardized Residuals")
abline(v = 4*(p+1)/n, col = "red", lty = 2)
abline(h = c(-3,3), col = "blue", lty =2)
```




```
# Cook's points and hat values
mod.full4.cooks <- cooks.distance(mod.full5)
influenceIndexPlot(mod.full5, vars = "Cook")
```

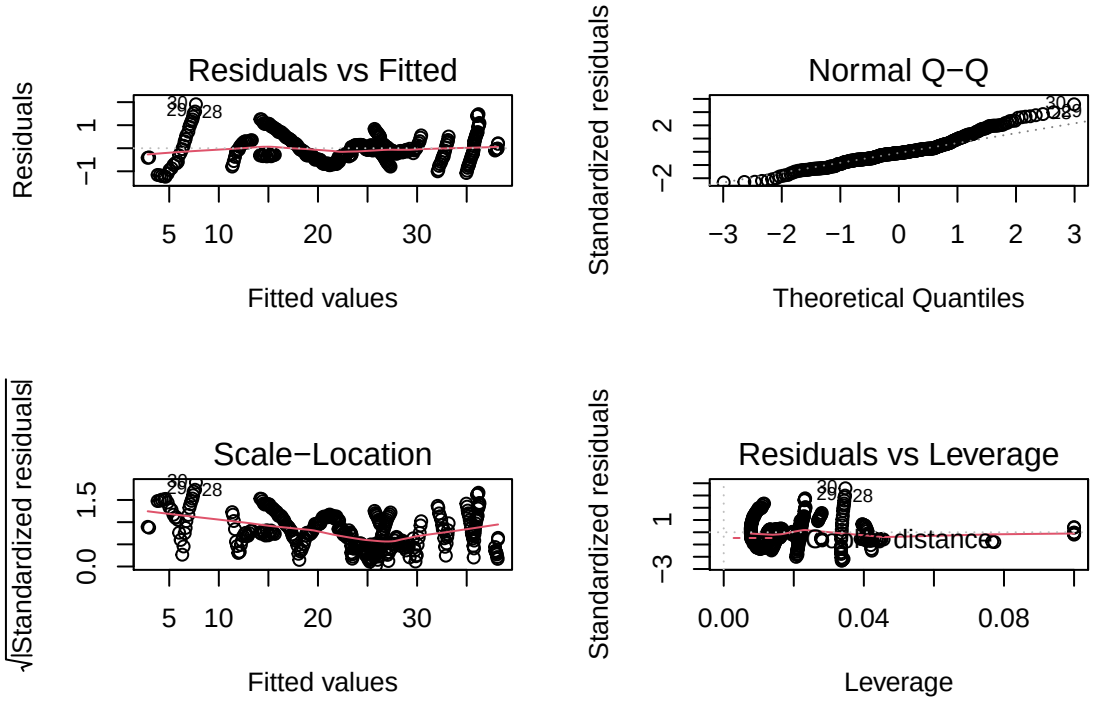



```
influenceIndexPlot(mod.full5, vars = "hat")
```



Plot 11: After removing the outliers work_place_restrictions is no longer significant so we removed it and the plots look roughly the same so we remove it from our model.

```
par(mfrow=c(2,2))
mod.full6 <- lm((deaths^{1/3}) ~ fstay_home_restrictions + I(tests^{1/3}) +
    fschool_closing + fcontact_tracing + ftesting_policy + log(stringency_index),
    data = fbase_data1)
plot(mod.full6)
```



Appendix 2: R Code for Google Mobility

```
regsubset.trans1 <- lm(sqrt(deaths) ~ date + sqrt(tests) + fschool_closing +
                         fworkplace_closing + fstay_home_restrictions +
                         ftesting_policy +fcontact_tracing, data = cacovid_mobility)
# leverage point calculations
p <- 7
n <- nrow(cacovid_mobility)
cacovid_hat <- hatvalues(regsubset.trans1)
which(cacovid_hat > 4*(p+1)/n)
```

Code 1:

```
##     1     2     3     4     5     6     7     8     9    10    11    12    13
##     1     2     3     4     5     6     7     8     9    10    11    12    13
##    14    15    16    17    18    19    20    21    22    23    24    25    26
##    14    15    16    17    18    19    20    21    22    23    24    25    26
##    27    28    29    30    31    32    33    34    35    36    37    38    39
##    27    28    29    30    31    32    33    34    35    36    37    38    39
##    40    41    42    43    44    45    46    47    48    49    50    51    52
##    40    41    42    43    44    45    46    47    48    49    50    51    52
##    53    54    55    56    57    58    59    60    61    62    63    64    65
##    53    54    55    56    57    58    59    60    61    62    63    64    65
##    66    67    68    69    70    71    72    73 12597 12598 12599 12600 12601
##    66    67    68    69    70    71    72    73 12597 12598 12599 12600 12601
## 12602 12603 12604 12605 12606 12607 12608 12609 12610 12611 12612 12613 12614
## 12602 12603 12604 12605 12606 12607 12608 12609 12610 12611 12612 12613 12614
## 12615 12616 12617 12618 12619 12620 12621 12622 12623 12624 12625 12626 12627
## 12615 12616 12617 12618 12619 12620 12621 12622 12623 12624 12625 12626 12627
## 12628 12629 12630 12631 12632 12633 12634 12635 12636 12637 12638 12639 12640
## 12628 12629 12630 12631 12632 12633 12634 12635 12636 12637 12638 12639 12640
## 12641 12642 12643 12644 12645 12646 12647 12648 12649 12650 12651 12652 12653
## 12641 12642 12643 12644 12645 12646 12647 12648 12649 12650 12651 12652 12653
## 12654 12655 12656 12657 12658 12659 12660 12661 12662 12663 12664 12665 12666
## 12654 12655 12656 12657 12658 12659 12660 12661 12662 12663 12664 12665 12666
## 12667 12668 12669 12670 12671 12672 12673 12674 12675 12676 12677 12678 12679
## 12667 12668 12669 12670 12671 12672 12673 12674 12675 12676 12677 12678 12679
## 12680 12681 12682 12683 12684 12685 12686 12687 12688 12689 12690 12691 12692
## 12680 12681 12682 12683 12684 12685 12686 12687 12688 12689 12690 12691 12692
## 12693 12694 12695 12696 12697 12698 12699 12700 12701 12702 12703 12704 12705
## 12693 12694 12695 12696 12697 12698 12699 12700 12701 12702 12703 12704 12705
## 12706 12707 12708 12709 12710 12711 12712 12713 12714 12715 12716 12717 12718
## 12706 12707 12708 12709 12710 12711 12712 12713 12714 12715 12716 12717 12718
## 12719 12720 12721 12722 12723 12724 12725 12726 12727 12728 12729 12730 12731
## 12719 12720 12721 12722 12723 12724 12725 12726 12727 12728 12729 12730 12731
## 12732 12733 12734 12735 12736 12737 12738 12739 12740 12741 12742 12743 12744
## 12732 12733 12734 12735 12736 12737 12738 12739 12740 12741 12742 12743 12744
## 12745 12746 12747 12748 12749 12750 12751 12752 12753 12754 12755 12756 12757
## 12745 12746 12747 12748 12749 12750 12751 12752 12753 12754 12755 12756 12757
## 12758 12759 12760 12761 12762 12763 12764 12765 12766 12767 12768 12769 12770
## 12758 12759 12760 12761 12762 12763 12764 12765 12766 12767 12768 12769 12770
## 12771 12772 12773 12774 12775 12776 12777 12778 12779 12780 12781 12782 12783
```

```
## 12771 12772 12773 12774 12775 12776 12777 12778 12779 12780 12781 12782 12783
## 12784 12785 12786 12787 12788 12789 12790 12791 12792 12793 12794 12795 12796
## 12784 12785 12786 12787 12788 12789 12790 12791 12792 12793 12794 12795 12796
## 12797 12798 12799 12800 12801 12802 12803 12804 12805 12806 12807 12808 12809
## 12797 12798 12799 12800 12801 12802 12803 12804 12805 12806 12807 12808 12809
## 12810 12811 12812 12813 12814 12815 12816 12817 12818 12819 12820 12821 12822
## 12810 12811 12812 12813 12814 12815 12816 12817 12818 12819 12820 12821 12822
## 12823 12824 12825 12826 12827 12828 12829 12830 12831 12832 12833 12834 12835
## 12823 12824 12825 12826 12827 12828 12829 12830 12831 12832 12833 12834 12835
## 12836 12837 12838 12839 12840 12841 12842 12843 12844 12845 12846 12847 12848
## 12836 12837 12838 12839 12840 12841 12842 12843 12844 12845 12846 12847 12848
## 12849 12850 12851 12852 12853 12854 12855 12856 12857 12858 12859 12860 12861
## 12849 12850 12851 12852 12853 12854 12855 12856 12857 12858 12859 12860 12861
## 12862 12863 12864 12865 12866 12867 12868 12869 12870 12871 12872 12873 12874
## 12862 12863 12864 12865 12866 12867 12868 12869 12870 12871 12872 12873 12874
## 12875 12876 12877 12878 12879 12880 12881 12882 12883 12884 12885 12886 12887
## 12875 12876 12877 12878 12879 12880 12881 12882 12883 12884 12885 12886 12887
## 12888 12889 12890 12891 12892 12893 12894 12895 12896 12897 12898 12899 12900
## 12888 12889 12890 12891 12892 12893 12894 12895 12896 12897 12898 12899 12900
## 12901 12902 12903 12904 12905 12906 12907 12908 12909 12910 12911 12912 12913
## 12901 12902 12903 12904 12905 12906 12907 12908 12909 12910 12911 12912 12913
## 12914 12915 12916 12917 12918 12919 12920 12921 12922 12923 12924 12925 12926
## 12914 12915 12916 12917 12918 12919 12920 12921 12922 12923 12924 12925 12926
## 12927 12928 12929 12930 12931 12932 12933 12934 12935 12936 12937 12938 12939
## 12927 12928 12929 12930 12931 12932 12933 12934 12935 12936 12937 12938 12939
## 12940 12941 12942 12943 12944 12945 12946 12947 12948 12949 12950 12951 12952
## 12940 12941 12942 12943 12944 12945 12946 12947 12948 12949 12950 12951 12952
## 12953 12954 12955 12956 12957 12958 12959 12960 12961 12962 12963 12964 12965
## 12953 12954 12955 12956 12957 12958 12959 12960 12961 12962 12963 12964 12965
## 12966 12967 12968 12969 12970 12971 12972 12973
## 12966 12967 12968 12969 12970 12971 12972 12973
```

```
# outlier calculations
cacovid_std <- rstandard(regsubset.trans1)
which(abs(cacovid_std) > 2)
```

```
##  1254  1255  1256  1257  1258  1259  1260  1261  1262  1263  1264  1265  1266
##  1254  1255  1256  1257  1258  1259  1260  1261  1262  1263  1264  1265  1266
##  1267  1268  1269  1270  1271  1272  1273  1274  1275  1276  1277  1278  1279
##  1267  1268  1269  1270  1271  1272  1273  1274  1275  1276  1277  1278  1279
##  1280  1281  1282  1283  1284  1285  1286  1287  1288  1289  1290  1291  1292
##  1280  1281  1282  1283  1284  1285  1286  1287  1288  1289  1290  1291  1292
##  1293  1294  1295  1296  1297  1298  1299  1300  1301  1302  1303  1304  1305
##  1293  1294  1295  1296  1297  1298  1299  1300  1301  1302  1303  1304  1305
##  1306  1307  1308  1309  1310  1311  1312  1313  1314  1315  1316  1317  1318
##  1306  1307  1308  1309  1310  1311  1312  1313  1314  1315  1316  1317  1318
##  1319  1320  1321  1322  1323  1324  1325  1326  1327  1328  2527  2528  2529
##  1319  1320  1321  1322  1323  1324  1325  1326  1327  1328  2527  2528  2529
##  2530  2531  2532  2533  2534  2535  2536  2537  2538  2539  2540  2541  2542
##  2530  2531  2532  2533  2534  2535  2536  2537  2538  2539  2540  2541  2542
##  2543  2544  2545  2546  2547  2548  2549  2550  2551  2552  2553  2554  2555
##  2543  2544  2545  2546  2547  2548  2549  2550  2551  2552  2553  2554  2555
##  2556  2557  2558  2559  2560  2561  2562  2563  2564  2565  2566  2567  2568
##  2556  2557  2558  2559  2560  2561  2562  2563  2564  2565  2566  2567  2568
```

```
##  2569  2570  2571  2572  2573  2574  2575  2576  2577  2578  2579  2580  2581
##  2569  2570  2571  2572  2573  2574  2575  2576  2577  2578  2579  2580  2581
##  2582  2583  2584  2585  2586  2587  2588  2589  2590  2591  2592  2593  2594
##  2582  2583  2584  2585  2586  2587  2588  2589  2590  2591  2592  2593  2594
##  2595  2596  2597  8432  8433  8434  8435  8436  8437  8438  8439  8440  8441
##  2595  2596  2597  8432  8433  8434  8435  8436  8437  8438  8439  8440  8441
##  8442  8443  8444  8445  8446  8447  8448  8449  8450  8451  8452  8453  8454
##  8442  8443  8444  8445  8446  8447  8448  8449  8450  8451  8452  8453  8454
##  8455  8456  8457  8458  8459  8460  8461  8462  8463  8464  8465  8466  8467
##  8455  8456  8457  8458  8459  8460  8461  8462  8463  8464  8465  8466  8467
##  8468  8469  8470  8471  8472  8473  8474  8475  8476  8477  8478  8479  8480
##  8468  8469  8470  8471  8472  8473  8474  8475  8476  8477  8478  8479  8480
##  8481  8482  8483  8484  8485  8486  8487  8488  8489  8490  8491  8492  8493
##  8481  8482  8483  8484  8485  8486  8487  8488  8489  8490  8491  8492  8493
##  8494  8495  8496  8497  8498  8499  8500  8501  8502  8503  8504  8505  8506
##  8494  8495  8496  8497  8498  8499  8500  8501  8502  8503  8504  8505  8506
##  8507  8508  8509  8510  8511  8512  8513  8514  8515  8516  8517  8518  8519
##  8507  8508  8509  8510  8511  8512  8513  8514  8515  8516  8517  8518  8519
##  8520  8521  8522  8523  8524  8525  8526  8527  8528  8529  8530  8531  8532
##  8520  8521  8522  8523  8524  8525  8526  8527  8528  8529  8530  8531  8532
##  8533  8534  8535  8536  8537  8538  8539  8540  8541  8542  8543  8544  8545
##  8533  8534  8535  8536  8537  8538  8539  8540  8541  8542  8543  8544  8545
##  8546  8547  8548  8549  8550  8551  8552  8553  8554  8555  8556  8557  8558
##  8546  8547  8548  8549  8550  8551  8552  8553  8554  8555  8556  8557  8558
##  8559  8560  8561  8562  8563  8564  8565  8566  8567  8568  8569  8570  8571
##  8559  8560  8561  8562  8563  8564  8565  8566  8567  8568  8569  8570  8571
##  8572  8573  8574  8575  8576  8577  8578  8579  8580  8581  8582  8583  8584
##  8572  8573  8574  8575  8576  8577  8578  8579  8580  8581  8582  8583  8584
##  8585  8586  8587  8588  8589  8590  8591  8592  8593  8594  8595  8596  8597
##  8585  8586  8587  8588  8589  8590  8591  8592  8593  8594  8595  8596  8597
##  8598  8599  8600  8601  8602  8603  8604  8605  8606  8607  8608  8609  8610
##  8598  8599  8600  8601  8602  8603  8604  8605  8606  8607  8608  8609  8610
##  8611  8612  8613  8614  8615  8616  8617  8618  8619  8620  8621  8622  8623
##  8611  8612  8613  8614  8615  8616  8617  8618  8619  8620  8621  8622  8623
##  8624  8625  8626  8627  8628  8629  8630  8631  8632  8633  8634  8635  8636
##  8624  8625  8626  8627  8628  8629  8630  8631  8632  8633  8634  8635  8636
##  8637  8638  8639  8640  8641  8642  8643  8644  8645  8646  8647  8648  8649
##  8637  8638  8639  8640  8641  8642  8643  8644  8645  8646  8647  8648  8649
##  8650  8651  8652  8653  8654  8655  8656  8657 11197 11198 11199 11200 11201
##  8650  8651  8652  8653  8654  8655  8656  8657 11197 11198 11199 11200 11201
## 11202 11203 11204 11205 11206 11207 11208 11209 11210 11211 11212 11213 11214
## 11202 11203 11204 11205 11206 11207 11208 11209 11210 11211 11212 11213 11214
## 11215 11216 11217 11218 11219 11220 11221 11222 11223 11224 11225 11226 11227
## 11215 11216 11217 11218 11219 11220 11221 11222 11223 11224 11225 11226 11227
## 11228 11229 11230 11231 11232 11233 11234 11235 12295 12296 12297 12298 12299
## 11228 11229 11230 11231 11232 11233 11234 11235 12295 12296 12297 12298 12299
## 12300 12301 12302 12303 12304 12305 12306 12307 12308 12309 12310 12311 12312
## 12300 12301 12302 12303 12304 12305 12306 12307 12308 12309 12310 12311 12312
## 12313 12314 12315 12316 12317 12318 12319 12320 12321 12322 12323 12324 12325
## 12313 12314 12315 12316 12317 12318 12319 12320 12321 12322 12323 12324 12325
## 12326 12327 12328 12329 12330 12331 12332 12333 12334 12335 12336 12337 12338
## 12326 12327 12328 12329 12330 12331 12332 12333 12334 12335 12336 12337 12338
## 12339 12340 12341 12342 12343 12344 12345 12346 12347 12348 12349 12350 12351
## 12339 12340 12341 12342 12343 12344 12345 12346 12347 12348 12349 12350 12351
```

```
## 12352 12353 12354 12355 12356 12357 12358 12359 12360 12361 12362 12363 12364
## 12352 12353 12354 12355 12356 12357 12358 12359 12360 12361 12362 12363 12364
## 12365 12366 12367 12368 12369 12370 12371 12372 12373 12374 12375 12376 12377
## 12365 12366 12367 12368 12369 12370 12371 12372 12373 12374 12375 12376 12377
## 12378 12379 12380 12381 12382 12383 12384 12385 12386 12387 12388 12389 12390
## 12378 12379 12380 12381 12382 12383 12384 12385 12386 12387 12388 12389 12390
## 12391 12392 12393 12394 12395 12396 12397 12398 12399 12400 12401 12402 12403
## 12391 12392 12393 12394 12395 12396 12397 12398 12399 12400 12401 12402 12403
## 12404 12405 12406 12407 12408 12409 12410 12411 12412 12413 12414 12415 12416
## 12404 12405 12406 12407 12408 12409 12410 12411 12412 12413 12414 12415 12416
## 12417 12418 12419 12420 12421 12422 12423 12424 12425 12426 12427 12428 12429
## 12417 12418 12419 12420 12421 12422 12423 12424 12425 12426 12427 12428 12429
## 12430 12431 12432 12433 12434 12435 12436 12437 12438 12439 12440 12441 12442
## 12430 12431 12432 12433 12434 12435 12436 12437 12438 12439 12440 12441 12442
## 12443 12444 12445 12446 12447 12448 12449 12450 12451 12452 12453 12454 12455
## 12443 12444 12445 12446 12447 12448 12449 12450 12451 12452 12453 12454 12455
## 12456 12457 12458 12459 12460 12461 12462 12463 12464 12465 12466 12467 12468
## 12456 12457 12458 12459 12460 12461 12462 12463 12464 12465 12466 12467 12468
## 12469 12470 12471 12472 12473 12474 12475 12476 12477 12478 12479 12480 12481
## 12469 12470 12471 12472 12473 12474 12475 12476 12477 12478 12479 12480 12481
## 12482 12483 12484 12485 12486 12487 12488 12489 12490 12491 12492 12493 12494
## 12482 12483 12484 12485 12486 12487 12488 12489 12490 12491 12492 12493 12494
## 12495 12496 12497 12498 12499 12500 12501 12502 12503 12504 12505 12506 12507
## 12495 12496 12497 12498 12499 12500 12501 12502 12503 12504 12505 12506 12507
## 12508 12509 12510 12511 12512 12513 12514 12515 12516 12517 12518 12519 12520
## 12508 12509 12510 12511 12512 12513 12514 12515 12516 12517 12518 12519 12520
## 12521 12522 12523 12524 12525 12526 12527 12528 12529 12530 12531 12532 12533
## 12521 12522 12523 12524 12525 12526 12527 12528 12529 12530 12531 12532 12533
## 12534 12535 12536 12537 12538 12539 12540 12541 12542 12543 12544 12545 12546
## 12534 12535 12536 12537 12538 12539 12540 12541 12542 12543 12544 12545 12546
## 12547 12548 12549 12550 12551 12552 12553 12554 12555 12556 12557 12558 12559
## 12547 12548 12549 12550 12551 12552 12553 12554 12555 12556 12557 12558 12559
## 12560 12561 12562 12563 12564 12565 12566 12567 12568 12569 12570 12571 12572
## 12560 12561 12562 12563 12564 12565 12566 12567 12568 12569 12570 12571 12572
## 12573 12574 12575 12576 12577 12578 12579 12580 12581 12582 12583 12584 12585
## 12573 12574 12575 12576 12577 12578 12579 12580 12581 12582 12583 12584 12585
## 12586 12587 12588 12589 12590 12591 12592 12593 12594 12595 12596
## 12586 12587 12588 12589 12590 12591 12592 12593 12594 12595 12596
```

```
# full fitted model
google.full1 <- lm(deaths ~ retail_and_recreation_percent_change_from_baseline + grocery_and_pharmacy_p

summary(google.full1)
```

Table 1:

```
##
## Call:
## lm(formula = deaths ~ retail_and_recreation_percent_change_from_baseline +
##     grocery_and_pharmacy_percent_change_from_baseline + parks_percent_change_from_baseline +
##     transit_stations_percent_change_from_baseline + workplaces_percent_change_from_baseline +
```

```
##     residential_percent_change_from_baseline + date + confirmed +
##     tests + fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + finternal_movement_restrictions +
##     finformation_campaigns + ftesting_policy + fcontact_tracing +
##     stringency_index, data = cacovid_mobility)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3438.2  -583.0   -14.4   655.7  3375.1
##
## Coefficients: (2 not defined because of singularities)
##                                                    Estimate Std. Error
## (Intercept)                                       4.939e+05  3.286e+04
## retail_and_recreation_percent_change_from_baseline  3.842e+02  1.516e+02
## grocery_and_pharmacy_percent_change_from_baseline  -1.958e+01  1.600e+02
## parks_percent_change_from_baseline                  1.096e+02  3.218e+01
## transit_stations_percent_change_from_baseline      -8.237e+01  5.132e+01
## workplaces_percent_change_from_baseline            -1.987e+01  1.644e+02
## residential_percent_change_from_baseline           -1.476e+03  4.710e+02
## date                                               -2.814e+01  1.785e+00
## confirmed                                          -8.406e-04  1.937e-04
## tests                                               1.518e-03  2.682e-05
## fschool_closing3                                   -4.083e+03  8.308e+01
## fworkplace_closing2                                 2.678e+03  2.189e+02
## fworkplace_closing3                                 3.266e+03  2.735e+02
## fgatherings_restrictions3                          -1.040e+04  3.426e+02
## fgatherings_restrictions4                          -1.216e+04  3.725e+02
## fstay_home_restrictions2                           -6.714e+03  3.924e+01
## finternal_movement_restrictions1                    3.355e+03  7.746e+01
## finternal_movement_restrictions2                           NA         NA
## finformation_campaigns2                                    NA         NA
## ftesting_policy2                                    1.500e+03  6.814e+01
## ftesting_policy3                                    7.963e+02  8.899e+01
## fcontact_tracing2                                  -1.466e+03  5.076e+01
## stringency_index                                    5.314e+02  1.327e+01
##                                                    t value Pr(>|t|)
## (Intercept)                                         15.029  < 2e-16 ***
## retail_and_recreation_percent_change_from_baseline   2.534 0.011281 *
## grocery_and_pharmacy_percent_change_from_baseline   -0.122 0.902596
## parks_percent_change_from_baseline                   3.408 0.000657 ***
## transit_stations_percent_change_from_baseline       -1.605 0.108545
## workplaces_percent_change_from_baseline             -0.121 0.903807
## residential_percent_change_from_baseline            -3.134 0.001729 **
## date                                               -15.758  < 2e-16 ***
## confirmed                                           -4.340 1.44e-05 ***
## tests                                               56.605  < 2e-16 ***
## fschool_closing3                                   -49.142  < 2e-16 ***
## fworkplace_closing2                                 12.238  < 2e-16 ***
## fworkplace_closing3                                 11.942  < 2e-16 ***
## fgatherings_restrictions3                          -30.370  < 2e-16 ***
## fgatherings_restrictions4                          -32.647  < 2e-16 ***
## fstay_home_restrictions2                          -171.094  < 2e-16 ***
## finternal_movement_restrictions1                    43.316  < 2e-16 ***
## finternal_movement_restrictions2                        NA       NA
```

```
## finformation_campaigns2                                NA       NA
## ftesting_policy2                                   22.008  < 2e-16 ***
## ftesting_policy3                                    8.948  < 2e-16 ***
## fcontact_tracing2                                 -28.878  < 2e-16 ***
## stringency_index                                   40.060  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1062 on 12952 degrees of freedom
## Multiple R-squared:  0.9952, Adjusted R-squared:  0.9952
## F-statistic: 1.344e+05 on 20 and 12952 DF,  p-value: < 2.2e-16
```

```
# full fitted model w/ removed predictor variables
google.full <- lm(deaths ~ retail_and_recreation_percent_change_from_baseline + grocery_and_pharmacy_pe

google.null <- lm(deaths ~ 1, data = cacovid_mobility)

anova(google.null, google.full)
```

Table 2:

```
## Analysis of Variance Table
##
## Model 1: deaths ~ 1
## Model 2: deaths ~ retail_and_recreation_percent_change_from_baseline +
##     grocery_and_pharmacy_percent_change_from_baseline + parks_percent_change_from_baseline +
##     transit_stations_percent_change_from_baseline + workplaces_percent_change_from_baseline +
##     residential_percent_change_from_baseline + date + confirmed +
##     tests + fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + ftesting_policy + fcontact_tracing +
##     stringency_index
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1  12972 3.0456e+12
## 2  12953 1.6726e+10 19 3.0288e+12 123455 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# best subset regression
subset.summary <- summary(regsubsets(deaths ~ retail_and_recreation_percent_change_from_baseline + groc
            fworkplace_closing + fgatherings_restrictions + fstay_home_restrictions + ftesting_policy +

subset.summary
```

Table 3:

```
## Subset selection object
## Call: regsubsets.formula(deaths ~ retail_and_recreation_percent_change_from_baseline +
```

```
##     grocery_and_pharmacy_percent_change_from_baseline + parks_percent_change_from_baseline +
##     transit_stations_percent_change_from_baseline + workplaces_percent_change_from_baseline +
##     residential_percent_change_from_baseline + date + confirmed +
##     tests + fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + ftesting_policy + fcontact_tracing +
##     stringency_index, data = cacovid_mobility)
## 19 Variables  (and intercept)
##                                                  Forced in Forced out
## retail_and_recreation_percent_change_from_baseline     FALSE      FALSE
## grocery_and_pharmacy_percent_change_from_baseline      FALSE      FALSE
## parks_percent_change_from_baseline                     FALSE      FALSE
## transit_stations_percent_change_from_baseline          FALSE      FALSE
## workplaces_percent_change_from_baseline                FALSE      FALSE
## residential_percent_change_from_baseline               FALSE      FALSE
## date                                                   FALSE      FALSE
## confirmed                                              FALSE      FALSE
## tests                                                  FALSE      FALSE
## fschool_closing3                                       FALSE      FALSE
## fworkplace_closing2                                    FALSE      FALSE
## fworkplace_closing3                                    FALSE      FALSE
## fgatherings_restrictions3                              FALSE      FALSE
## fgatherings_restrictions4                              FALSE      FALSE
## fstay_home_restrictions2                               FALSE      FALSE
## ftesting_policy2                                       FALSE      FALSE
## ftesting_policy3                                       FALSE      FALSE
## fcontact_tracing2                                      FALSE      FALSE
## stringency_index                                       FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          retail_and_recreation_percent_change_from_baseline
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
##          grocery_and_pharmacy_percent_change_from_baseline
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
##          parks_percent_change_from_baseline
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
```

```
## 7  ( 1 ) " "
## 8  ( 1 ) " "
##          transit_stations_percent_change_from_baseline
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
##          workplaces_percent_change_from_baseline
## 1  ( 1 ) " "
## 2  ( 1 ) " "
## 3  ( 1 ) " "
## 4  ( 1 ) " "
## 5  ( 1 ) " "
## 6  ( 1 ) " "
## 7  ( 1 ) " "
## 8  ( 1 ) " "
##          residential_percent_change_from_baseline date confirmed tests
## 1  ( 1 ) " "                                      " "  "*"       " "
## 2  ( 1 ) " "                                      " "  "*"       " "
## 3  ( 1 ) " "                                      " "  " "       "*"
## 4  ( 1 ) " "                                      " "  " "       "*"
## 5  ( 1 ) " "                                      " "  " "       "*"
## 6  ( 1 ) " "                                      " "  " "       "*"
## 7  ( 1 ) " "                                      " "  " "       "*"
## 8  ( 1 ) " "                                      "*"  " "       "*"
##          fschool_closing3 fworkplace_closing2 fworkplace_closing3
## 1  ( 1 ) " "              " "                 " "
## 2  ( 1 ) " "              " "                 " "
## 3  ( 1 ) " "              " "                 "*"
## 4  ( 1 ) "*"              " "                 " "
## 5  ( 1 ) "*"              " "                 " "
## 6  ( 1 ) "*"              " "                 "*"
## 7  ( 1 ) "*"              "*"                 "*"
## 8  ( 1 ) "*"              "*"                 "*"
##          fgatherings_restrictions3 fgatherings_restrictions4
## 1  ( 1 ) " "                       " "
## 2  ( 1 ) " "                       " "
## 3  ( 1 ) " "                       " "
## 4  ( 1 ) " "                       " "
## 5  ( 1 ) " "                       " "
## 6  ( 1 ) " "                       " "
## 7  ( 1 ) " "                       " "
## 8  ( 1 ) " "                       " "
##          fstay_home_restrictions2 ftesting_policy2 ftesting_policy3
## 1  ( 1 ) " "                      " "              " "
## 2  ( 1 ) "*"                      " "              " "
## 3  ( 1 ) "*"                      " "              " "
## 4  ( 1 ) "*"                      " "              " "
## 5  ( 1 ) "*"                      " "              " "
## 6  ( 1 ) "*"                      "*"              " "
```

```
## 7  ( 1 ) "*"                    "*"              " "
## 8  ( 1 ) "*"                    "*"              " "
##          fcontact_tracing2 stringency_index
## 1  ( 1 ) " "               " "
## 2  ( 1 ) " "               " "
## 3  ( 1 ) " "               " "
## 4  ( 1 ) " "               "*"
## 5  ( 1 ) "*"               "*"
## 6  ( 1 ) "*"               " "
## 7  ( 1 ) "*"               " "
## 8  ( 1 ) "*"               " "
```

```
data.frame(subset.summary$adjr2,
subset.summary$cp,
subset.summary$bic)
```

Table 4:

```
##   subset.summary.adjr2 subset.summary.cp subset.summary.bic
## 1            0.9605253        80128.9929          -41912.02
## 2            0.9780352        38831.2128          -49508.58
## 3            0.9881143        15062.2614          -57466.84
## 4            0.9922731         5256.1681          -63044.89
## 5            0.9930667         3385.6083          -64442.39
## 6            0.9933658         2681.1650          -65006.04
## 7            0.9939441         1318.7697          -66180.73
## 8            0.9941868          747.6231          -66702.84
```

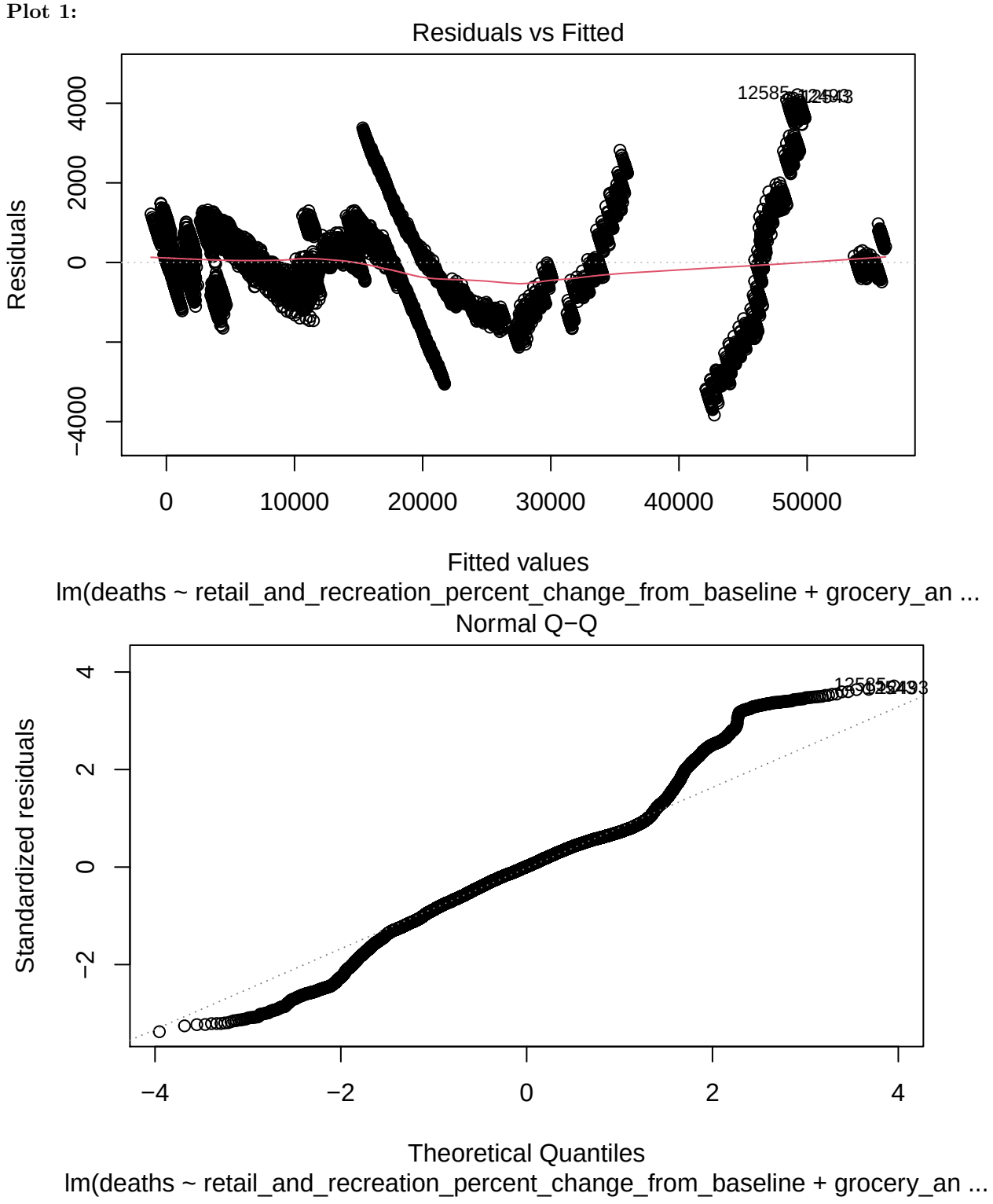
```
pt <- powerTransform(cbind(cacovid_mobility$tests) ~ 1)
summary(pt)
```

Table 5:

```
## bcPower Transformation to Normality
##    Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## Y1    0.3642        0.36       0.3529       0.3754
##
## Likelihood ratio test that transformation parameter is equal to 0
##  (log transformation)
##                            LRT df       pval
## LR test, lambda = (0) 5177.776  1 < 2.22e-16
##
## Likelihood ratio test that no transformation is needed
##                            LRT df       pval
## LR test, lambda = (1) 8574.034  1 < 2.22e-16
```

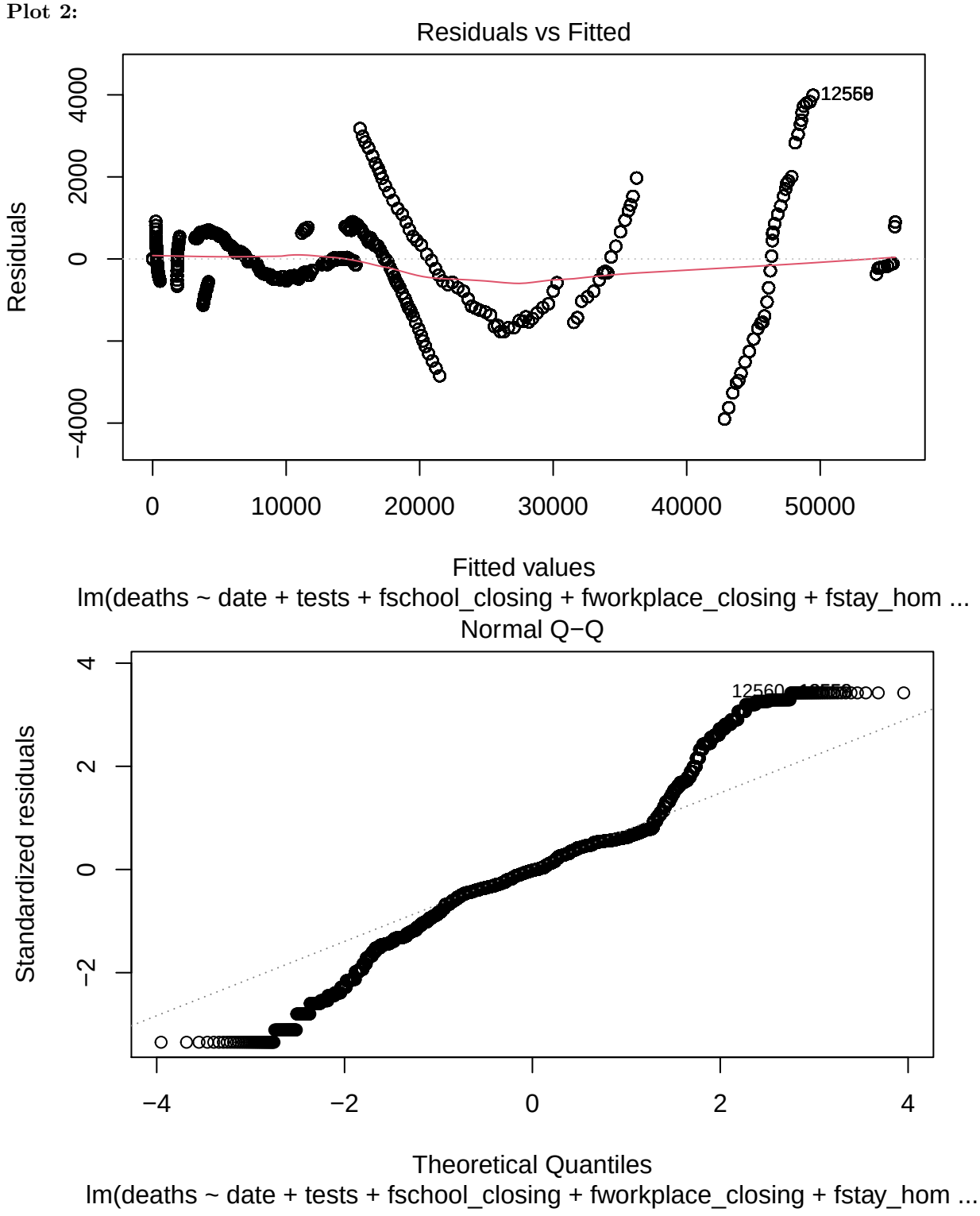
```
# diagnostic plot for full model w/ removed singularities
plot(google.full, which = c(1,2))
```

Plot 1:



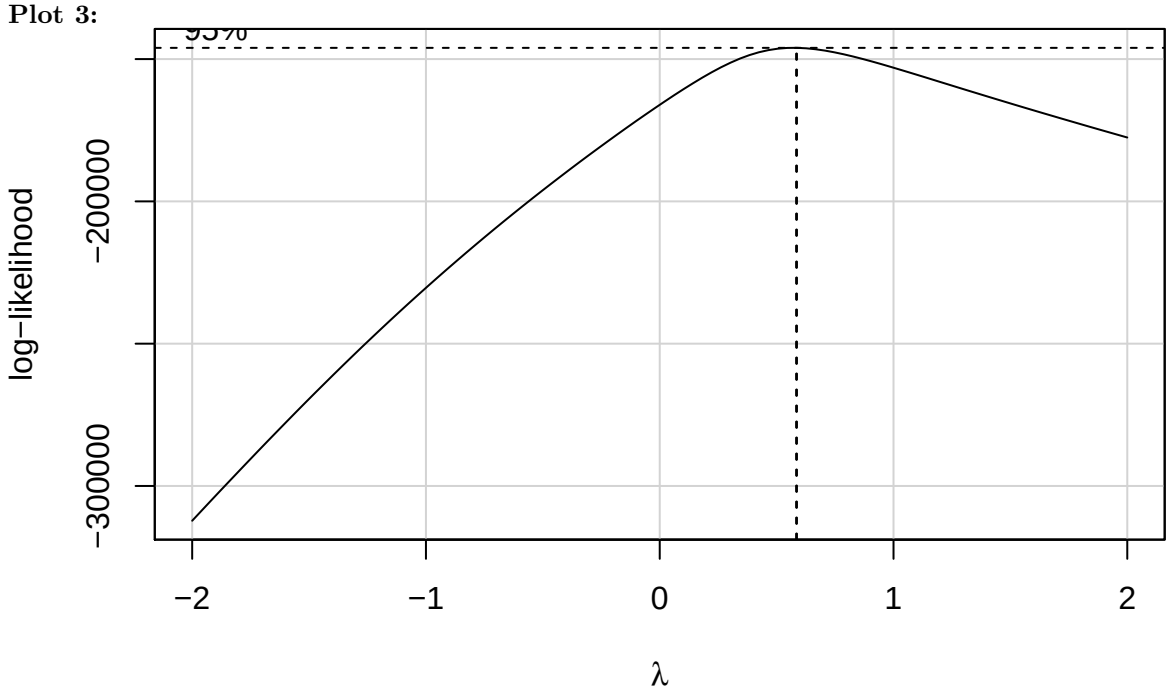
```r
# diagnostic plots after model selection (7 predictors)
regsubset.fit <- lm(deaths ~ date + tests + fschool_closing + fworkplace_closing + fstay_home_restricti
plot(regsubset.fit, which = c(1,2))
```

Plot 2:

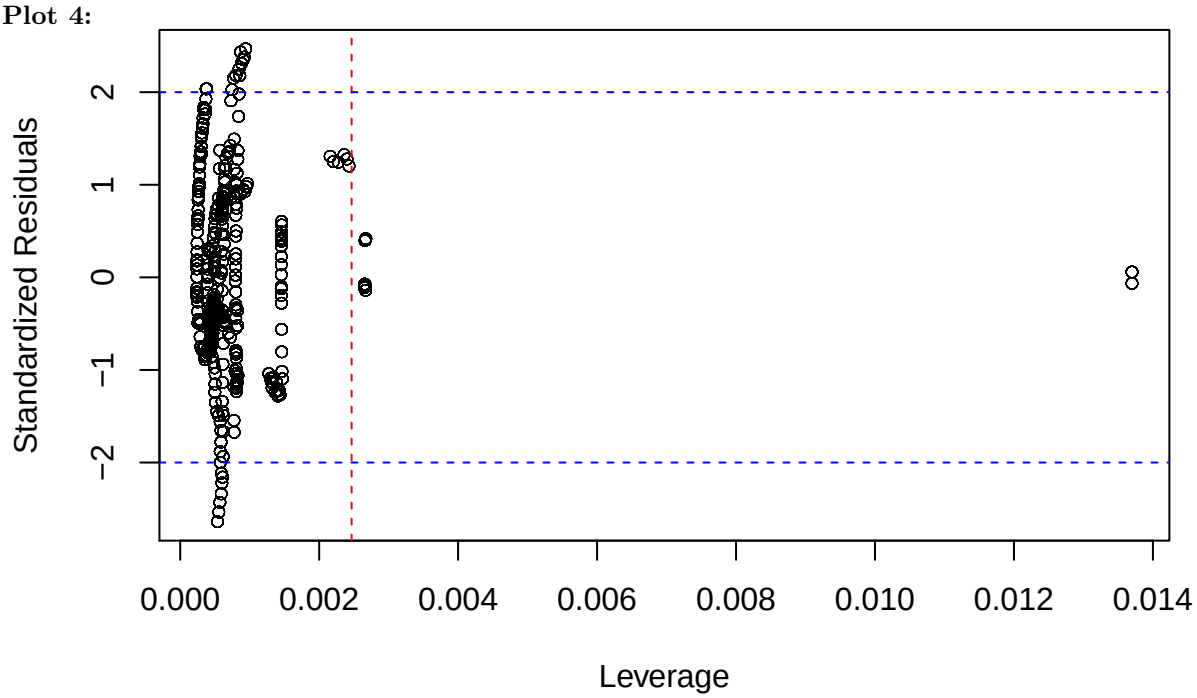


```r
# optimal lambda for response from full model
boxCox(regsubset.fit)
```

Plot 3:



Plot 4:



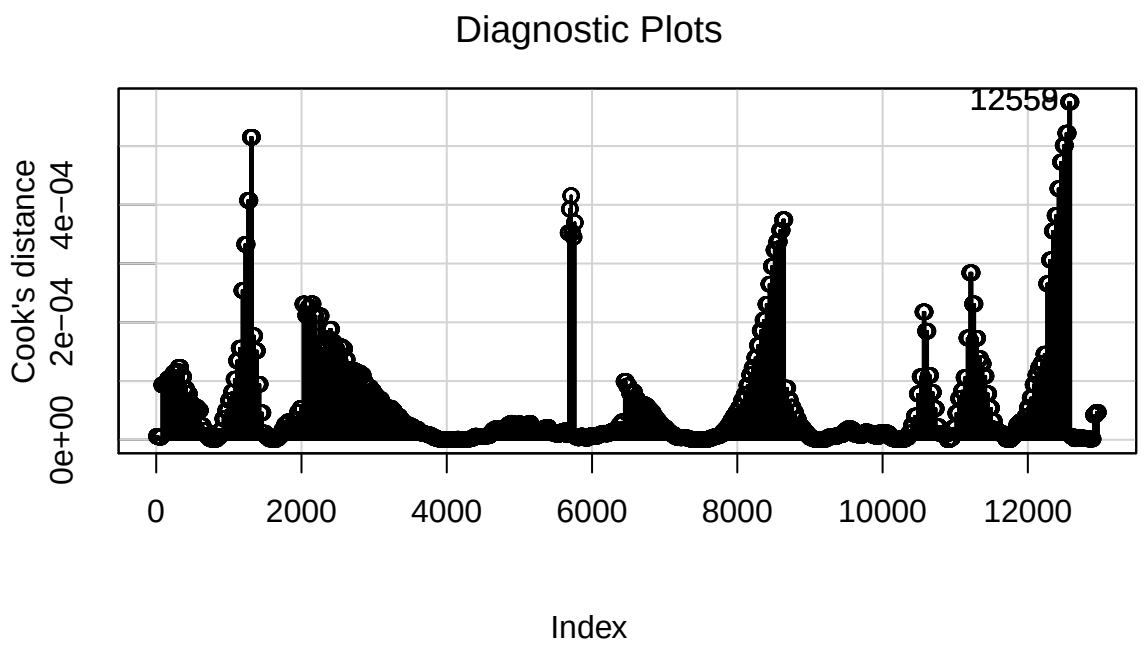
```r
# Cook's distance
cacovid.cooks <- cooks.distance(regsubset.trans1)
which(cacovid.cooks > 4/(n-p-1))
```

Plot 5:

```
##  1215  1216  1217  1218  1219  1220  1221  1222  1223  1224  1225  1226  1227
##  1215  1216  1217  1218  1219  1220  1221  1222  1223  1224  1225  1226  1227
##  1228  1229  1230  1231  1232  1233  1234  1235  1236  1237  1238  1239  1240
##  1228  1229  1230  1231  1232  1233  1234  1235  1236  1237  1238  1239  1240
##  1241  1242  1243  1244  1245  1246  1247  1248  1249  1250  1251  1252  1253
##  1241  1242  1243  1244  1245  1246  1247  1248  1249  1250  1251  1252  1253
##  1254  1255  1256  1257  1258  1259  1260  1261  1262  1263  1264  1265  1266
##  1254  1255  1256  1257  1258  1259  1260  1261  1262  1263  1264  1265  1266
##  1267  1268  1269  1270  1271  1272  1273  1274  1275  1276  1277  1278  1279
##  1267  1268  1269  1270  1271  1272  1273  1274  1275  1276  1277  1278  1279
##  1280  1281  1282  1283  1284  1285  1286  1287  1288  1289  1290  1291  1292
##  1280  1281  1282  1283  1284  1285  1286  1287  1288  1289  1290  1291  1292
##  1293  1294  1295  1296  1297  1298  1299  1300  1301  1302  1303  1304  1305
##  1293  1294  1295  1296  1297  1298  1299  1300  1301  1302  1303  1304  1305
##  1306  1307  1308  1309  1310  1311  1312  1313  1314  1315  1316  1317  1318
##  1306  1307  1308  1309  1310  1311  1312  1313  1314  1315  1316  1317  1318
##  1319  1320  1321  1322  1323  1324  1325  1326  1327  1328  5669  5670  5671
##  1319  1320  1321  1322  1323  1324  1325  1326  1327  1328  5669  5670  5671
##  5672  5673  5674  5675  5676  5677  5678  5679  5680  5681  5682  5683  5684
##  5672  5673  5674  5675  5676  5677  5678  5679  5680  5681  5682  5683  5684
##  5685  5686  5687  5688  5689  5690  5691  5692  5693  5694  5695  5696  5697
##  5685  5686  5687  5688  5689  5690  5691  5692  5693  5694  5695  5696  5697
##  5698  5699  5700  5701  5702  5703  5704  5705  5706  5707  5708  5709  5710
##  5698  5699  5700  5701  5702  5703  5704  5705  5706  5707  5708  5709  5710
##  5711  5712  5713  5714  5715  5716  5717  5718  5719  5720  5721  5722  5723
##  5711  5712  5713  5714  5715  5716  5717  5718  5719  5720  5721  5722  5723
##  5724  5725  5726  5727  5728  5729  5730  5731  5732  5733  5734  5735  5736
##  5724  5725  5726  5727  5728  5729  5730  5731  5732  5733  5734  5735  5736
##  5737  5738  5739  5740  5741  5742  5743  5744  5745  5746  5747  5748  5749
##  5737  5738  5739  5740  5741  5742  5743  5744  5745  5746  5747  5748  5749
##  5750  5751  5752  5753  5754  5755  5756  5757  5758  5759  5760  5761  5762
##  5750  5751  5752  5753  5754  5755  5756  5757  5758  5759  5760  5761  5762
##  5763  5764  5765  5766  5767  5768  5769  5770  5771  8503  8504  8505  8506
##  5763  5764  5765  5766  5767  5768  5769  5770  5771  8503  8504  8505  8506
##  8507  8508  8509  8510  8511  8512  8513  8514  8515  8516  8517  8518  8519
##  8507  8508  8509  8510  8511  8512  8513  8514  8515  8516  8517  8518  8519
##  8520  8521  8522  8523  8524  8525  8526  8527  8528  8529  8530  8531  8532
##  8520  8521  8522  8523  8524  8525  8526  8527  8528  8529  8530  8531  8532
##  8533  8534  8535  8536  8537  8538  8539  8540  8541  8542  8543  8544  8545
##  8533  8534  8535  8536  8537  8538  8539  8540  8541  8542  8543  8544  8545
##  8546  8547  8548  8549  8550  8551  8552  8553  8554  8555  8556  8557  8558
##  8546  8547  8548  8549  8550  8551  8552  8553  8554  8555  8556  8557  8558
##  8559  8560  8561  8562  8563  8564  8565  8566  8567  8568  8569  8570  8571
##  8559  8560  8561  8562  8563  8564  8565  8566  8567  8568  8569  8570  8571
##  8572  8573  8574  8575  8576  8577  8578  8579  8580  8581  8582  8583  8584
##  8572  8573  8574  8575  8576  8577  8578  8579  8580  8581  8582  8583  8584
```

```
##  8585  8586  8587  8588  8589  8590  8591  8592  8593  8594  8595  8596  8597
##  8585  8586  8587  8588  8589  8590  8591  8592  8593  8594  8595  8596  8597
##  8598  8599  8600  8601  8602  8603  8604  8605  8606  8607  8608  8609  8610
##  8598  8599  8600  8601  8602  8603  8604  8605  8606  8607  8608  8609  8610
##  8611  8612  8613  8614  8615  8616  8617  8618  8619  8620  8621  8622  8623
##  8611  8612  8613  8614  8615  8616  8617  8618  8619  8620  8621  8622  8623
##  8624  8625  8626  8627  8628  8629  8630  8631  8632  8633  8634  8635  8636
##  8624  8625  8626  8627  8628  8629  8630  8631  8632  8633  8634  8635  8636
##  8637  8638  8639  8640  8641  8642  8643  8644  8645  8646  8647  8648  8649
##  8637  8638  8639  8640  8641  8642  8643  8644  8645  8646  8647  8648  8649
##  8650  8651  8652  8653  8654  8655  8656  8657 12334 12335 12336 12337 12338
##  8650  8651  8652  8653  8654  8655  8656  8657 12334 12335 12336 12337 12338
## 12339 12340 12341 12342 12343 12344 12345 12346 12347 12348 12349 12350 12351
## 12339 12340 12341 12342 12343 12344 12345 12346 12347 12348 12349 12350 12351
## 12352 12353 12354 12355 12356 12357 12358 12359 12360 12361 12362 12363 12364
## 12352 12353 12354 12355 12356 12357 12358 12359 12360 12361 12362 12363 12364
## 12365 12366 12367 12368 12369 12370 12371 12372 12373 12374 12375 12376 12377
## 12365 12366 12367 12368 12369 12370 12371 12372 12373 12374 12375 12376 12377
## 12378 12379 12380 12381 12382 12383 12384 12385 12386 12387 12388 12389 12390
## 12378 12379 12380 12381 12382 12383 12384 12385 12386 12387 12388 12389 12390
## 12391 12392 12393 12394 12395 12396 12397 12398 12399 12400 12401 12402 12403
## 12391 12392 12393 12394 12395 12396 12397 12398 12399 12400 12401 12402 12403
## 12404 12405 12406 12407 12408 12409 12410 12411 12412 12413 12414 12415 12416
## 12404 12405 12406 12407 12408 12409 12410 12411 12412 12413 12414 12415 12416
## 12417 12418 12419 12420 12421 12422 12423 12424 12425 12426 12427 12428 12429
## 12417 12418 12419 12420 12421 12422 12423 12424 12425 12426 12427 12428 12429
## 12430 12431 12432 12433 12434 12435 12436 12437 12438 12439 12440 12441 12442
## 12430 12431 12432 12433 12434 12435 12436 12437 12438 12439 12440 12441 12442
## 12443 12444 12445 12446 12447 12448 12449 12450 12451 12452 12453 12454 12455
## 12443 12444 12445 12446 12447 12448 12449 12450 12451 12452 12453 12454 12455
## 12456 12457 12458 12459 12460 12461 12462 12463 12464 12465 12466 12467 12468
## 12456 12457 12458 12459 12460 12461 12462 12463 12464 12465 12466 12467 12468
## 12469 12470 12471 12472 12473 12474 12475 12476 12477 12478 12479 12480 12481
## 12469 12470 12471 12472 12473 12474 12475 12476 12477 12478 12479 12480 12481
## 12482 12483 12484 12485 12486 12487 12488 12489 12490 12491 12492 12493 12494
## 12482 12483 12484 12485 12486 12487 12488 12489 12490 12491 12492 12493 12494
## 12495 12496 12497 12498 12499 12500 12501 12502 12503 12504 12505 12506 12507
## 12495 12496 12497 12498 12499 12500 12501 12502 12503 12504 12505 12506 12507
## 12508 12509 12510 12511 12512 12513 12514 12515 12516 12517 12518 12519 12520
## 12508 12509 12510 12511 12512 12513 12514 12515 12516 12517 12518 12519 12520
## 12521 12522 12523 12524 12525 12526 12527 12528 12529 12530 12531 12532 12533
## 12521 12522 12523 12524 12525 12526 12527 12528 12529 12530 12531 12532 12533
## 12534 12535 12536 12537 12538 12539 12540 12541 12542 12543 12544 12545 12546
## 12534 12535 12536 12537 12538 12539 12540 12541 12542 12543 12544 12545 12546
## 12547 12548 12549 12550 12551 12552 12553 12554 12555 12556 12557 12558 12559
## 12547 12548 12549 12550 12551 12552 12553 12554 12555 12556 12557 12558 12559
## 12560 12561 12562 12563 12564 12565 12566 12567 12568 12569 12570 12571 12572
## 12560 12561 12562 12563 12564 12565 12566 12567 12568 12569 12570 12571 12572
## 12573 12574 12575 12576 12577 12578 12579 12580 12581 12582 12583 12584 12585
## 12573 12574 12575 12576 12577 12578 12579 12580 12581 12582 12583 12584 12585
## 12586 12587 12588 12589 12590 12591 12592 12593 12594 12595 12596
## 12586 12587 12588 12589 12590 12591 12592 12593 12594 12595 12596
```

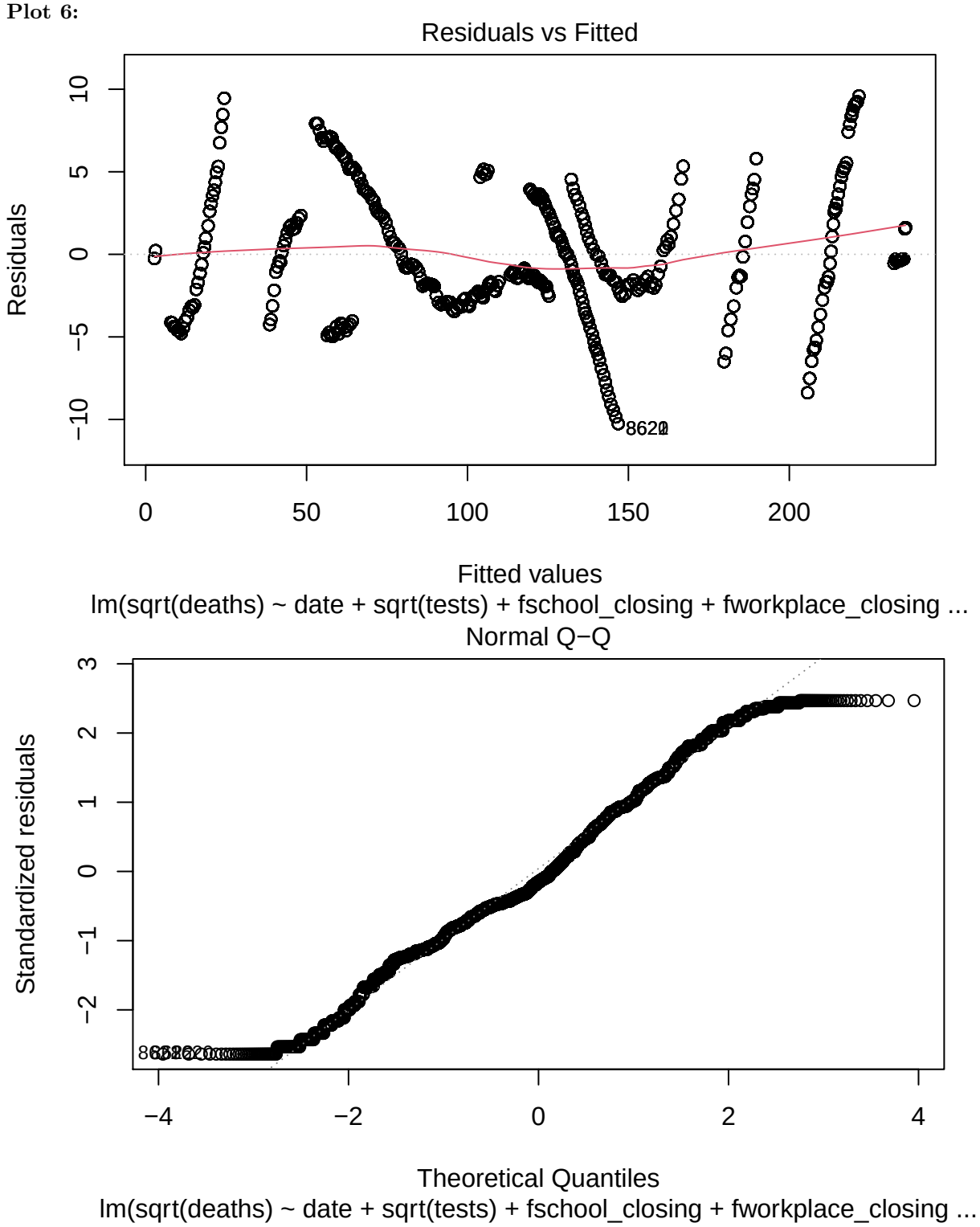
```
influenceIndexPlot(regsubset.trans1, vars = "Cook")
```




```
regsubset.trans1 <- lm(sqrt(deaths) ~ date + sqrt(tests) + fschool_closing +
                         fworkplace_closing + fstay_home_restrictions +
                         ftesting_policy +fcontact_tracing, data = cacovid_mobility)

plot(regsubset.trans1, which = c(1,2))
```

Plot 6:



Appendix 3: R Code for World Bank

Base Data with World Bank imported, GDP, GDP growth, Hospital beds/1,000 ppl, Poverty= Poverty head count ratio at 1.90 a day(% of pop), CO2em: CO2 emissions (metric tons/capita), Air pollution: %

Pop exposed to levels exceeding WHO guidelines,

```
x <- covid19()
```

```
## We have invested a lot of time and effort in creating COVID-19 Data Hub, please cite the following wh
##
##   Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open
##   Source Software 5(51):2376, doi: 10.21105/joss.02376.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {COVID-19 Data Hub},
##     year = {2020},
##     doi = {10.21105/joss.02376},
##     author = {Emanuele Guidotti and David Ardia},
##     journal = {Journal of Open Source Software},
##     volume = {5},
##     number = {51},
##     pages = {2376},
##   }
##
## To retrieve citation and metadata of the data sources see ?covid19cite. To hide this message use 'ver
```

```
covid19<- covid19(level = 1, start = "2020-03-15", end = "2021-03-15",)
wb <- c("gdp" = "NY.GDP.MKTP.CD", "hosp_beds" = "SH.MED.BEDS.ZS","gdp_grow" = "NY.GDP.MKTP.KD.ZG","pover
wbdcovid  <- covid19(wb = wb)
```

```
## We have invested a lot of time and effort in creating COVID-19 Data Hub, please cite the following wh
##
##   Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open
##   Source Software 5(51):2376, doi: 10.21105/joss.02376.
##
## A BibTeX entry for LaTeX users is
##
##   @Article{,
##     title = {COVID-19 Data Hub},
##     year = {2020},
##     doi = {10.21105/joss.02376},
##     author = {Emanuele Guidotti and David Ardia},
##     journal = {Journal of Open Source Software},
##     volume = {5},
##     number = {51},
##     pages = {2376},
##   }
##
## To retrieve citation and metadata of the data sources see ?covid19cite. To hide this message use 'ver
```

```
wbdcovid
```

```
## # A tibble: 94,726 x 42
## # Groups:   id [199]
```

```
##    iso_alpha_3 id    date       vaccines tests confirmed recovered deaths  hosp
##    <chr>       <chr> <date>        <dbl> <int>     <int>     <int>  <int> <dbl>
##  1 AFG         AFG   2020-01-22       NA    NA        NA        NA     NA    NA
##  2 AFG         AFG   2020-01-23       NA    NA        NA        NA     NA    NA
##  3 AFG         AFG   2020-01-24       NA    NA        NA        NA     NA    NA
##  4 AFG         AFG   2020-01-25       NA    NA        NA        NA     NA    NA
##  5 AFG         AFG   2020-01-26       NA    NA        NA        NA     NA    NA
##  6 AFG         AFG   2020-01-27       NA    NA        NA        NA     NA    NA
##  7 AFG         AFG   2020-01-28       NA    NA        NA        NA     NA    NA
##  8 AFG         AFG   2020-01-29       NA    NA        NA        NA     NA    NA
##  9 AFG         AFG   2020-01-30       NA    NA        NA        NA     NA    NA
## 10 AFG         AFG   2020-01-31       NA    NA        NA        NA     NA    NA
## # ... with 94,716 more rows, and 33 more variables: vent <int>, icu <int>,
## #   population <int>, school_closing <int>, workplace_closing <int>,
## #   cancel_events <int>, gatherings_restrictions <int>,
## #   transport_closing <int>, stay_home_restrictions <int>,
## #   internal_movement_restrictions <int>,
## #   international_movement_restrictions <int>, information_campaigns <int>,
## #   testing_policy <int>, contact_tracing <int>, stringency_index <dbl>,
## #   iso_alpha_2 <chr>, iso_numeric <int>, currency <chr>,
## #   administrative_area_level <int>, administrative_area_level_1 <chr>,
## #   administrative_area_level_2 <lgl>, administrative_area_level_3 <lgl>,
## #   latitude <dbl>, longitude <dbl>, key <lgl>, key_apple_mobility <chr>,
## #   key_google_mobility <chr>, gdp <dbl>, hosp_beds <dbl>, gdp_grow <dbl>,
## #   poverty <dbl>, co2em <dbl>, pollution <dbl>
```

Only variables from WB and Confirmed and Death

```
wbcovdata<- subset(wbdcovid, select = c("date", "confirmed", "deaths", "iso_alpha_3", "administrative_a
```

```
fgpd<- as.integer(wbcovdata$gdp)
```

```
## Warning: NAs introduced by coercion to integer range
```

```
fgdp_grow<- as.integer(wbcovdata$gdp_grow)
fhosp_beds<- as.integer(wbcovdata$hosp_beds)
fpoverty<- as.integer(wbcovdata$poverty)
fco2em<- as.integer(wbcovdata$co2em)
fpollution<- as.integer(wbcovdata$pollution)

#wbcovdata$gdp %>% replace_na(0)
#wbcovdata$gdp_grow %>% replace_na(0)
#wbcovdata$hosp_beds %>% replace_na(0)
#wbcovdata$poverty %>% replace_na(0)
#wbcovdata$co2em %>% replace_na(0)
#wbcovdata$pollution %>% replace_na(0)


cleandata <- na.omit(wbcovdata)
cleandata
```

```
## # A tibble: 51,351 x 11
```

```
##    date       confirmed deaths iso_alpha_3 administrative_are~      gdp gdp_grow
##    <date>         <int>  <int> <chr>       <chr>                   <dbl>    <dbl>
##  1 2020-03-11        12      1 ALB         Albania              1.53e10     2.24
##  2 2020-03-12        23      1 ALB         Albania              1.53e10     2.24
##  3 2020-03-13        33      1 ALB         Albania              1.53e10     2.24
##  4 2020-03-14        38      1 ALB         Albania              1.53e10     2.24
##  5 2020-03-15        42      1 ALB         Albania              1.53e10     2.24
##  6 2020-03-16        51      1 ALB         Albania              1.53e10     2.24
##  7 2020-03-17        55      1 ALB         Albania              1.53e10     2.24
##  8 2020-03-18        59      2 ALB         Albania              1.53e10     2.24
##  9 2020-03-19        64      2 ALB         Albania              1.53e10     2.24
## 10 2020-03-20        70      2 ALB         Albania              1.53e10     2.24
## # ... with 51,341 more rows, and 4 more variables: hosp_beds <dbl>,
## #   poverty <dbl>, co2em <dbl>, pollution <dbl>
```

Graphs to check for Normality and variance

```
economic<- lm(deaths ~ confirmed + gdp + gdp_grow + poverty, data= cleandata) # Economic
airqual<- lm(deaths ~ co2em + pollution, data= wbcovdata) #Air Quality

summary(economic)
```

```
##
## Call:
## lm(formula = deaths ~ confirmed + gdp + gdp_grow + poverty, data = cleandata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -173216   -3394   -1551     870  170438
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.446e+03  8.413e+01   52.85   <2e-16 ***
## confirmed    1.813e-02  3.412e-05  531.29   <2e-16 ***
## gdp          1.217e-09  2.644e-11   46.03   <2e-16 ***
## gdp_grow    -7.526e+02  2.024e+01  -37.18   <2e-16 ***
## poverty     -4.573e+01  3.287e+00  -13.91   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12650 on 51346 degrees of freedom
## Multiple R-squared:  0.8934, Adjusted R-squared:  0.8934
## F-statistic: 1.076e+05 on 4 and 51346 DF,  p-value: < 2.2e-16
```

```
summary(airqual)
```

```
##
## Call:
## lm(formula = deaths ~ co2em + pollution, data = wbcovdata)
##
## Residuals:
##    Min     1Q Median     3Q    Max
```

```
## -29024  -6005  -4523  -3352 555315
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 23220.796    467.545   49.66   <2e-16 ***
## co2em         317.907     22.419   14.18   <2e-16 ***
## pollution    -188.355      4.664  -40.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 32960 on 70074 degrees of freedom
##   (24649 observations deleted due to missingness)
## Multiple R-squared:  0.03154,    Adjusted R-squared:  0.03151
## F-statistic:  1141 on 2 and 70074 DF,  p-value: < 2.2e-16
```

```
wbdcovid19 <- covid19("US", level = 2, verbose=FALSE, wb = wb) %>%
  filter(administrative_area_level_2 == "California", date >= "2020-03-15", date <= "2021-03-14")


wb_small <- subset(wbdcovid19, select = c("date", "confirmed", "deaths", "gdp", "gdp_grow",
                                          "hosp_beds", "poverty", "co2em", "pollution"))
```

```
summary(wb_small)
```

```
##       date              confirmed           deaths           gdp
##  Min.   :2020-03-15   Min.   :    478   Min.   :    6   Min.   :2.143e+13
##  1st Qu.:2020-06-14   1st Qu.: 152953   1st Qu.: 5089   1st Qu.:2.143e+13
##  Median :2020-09-13   Median : 763389   Median :14386   Median :2.143e+13
##  Mean   :2020-09-13   Mean   :1131651   Mean   :16914   Mean   :2.143e+13
##  3rd Qu.:2020-12-13   3rd Qu.:1590670   3rd Qu.:21043   3rd Qu.:2.143e+13
##  Max.   :2021-03-14   Max.   :3623342   Max.   :56522   Max.   :2.143e+13
##     gdp_grow       hosp_beds       poverty      co2em        pollution
##  Min.   :2.161   Min.   :2.87   Min.   :1   Min.   :15.5   Min.   :3.343
##  1st Qu.:2.161   1st Qu.:2.87   1st Qu.:1   1st Qu.:15.5   1st Qu.:3.343
##  Median :2.161   Median :2.87   Median :1   Median :15.5   Median :3.343
##  Mean   :2.161   Mean   :2.87   Mean   :1   Mean   :15.5   Mean   :3.343
##  3rd Qu.:2.161   3rd Qu.:2.87   3rd Qu.:1   3rd Qu.:15.5   3rd Qu.:3.343
##  Max.   :2.161   Max.   :2.87   Max.   :1   Max.   :15.5   Max.   :3.343
```

```
n_distinct(wb_small$gdp)
```

```
## [1] 1
```

```
n_distinct(wb_small$gdp_grow)
```

```
## [1] 1
```

```
n_distinct(wb_small$hosp_beds)
```

```
## [1] 1
```

```
n_distinct(wb_small$poverty)
```

```
## [1] 1
```

```
n_distinct(wb_small$co2em)
```

```
## [1] 1
```

```
n_distinct(wb_small$pollution)
```

```
## [1] 1
```

Appendix 4: Exploratory analysis not used in final paper

Github Link:

Appendix 5: Data Variable Description

- **date** - Observation date
- **confirmed** - Cumulative number of confirmed cases
- **tests** - Cumulative number of tests
- **population** - Total population
- **latitude** - Latitude (Check to see if more than 1 since we are only using CA)
- **longitude** - Longitude (Check to see if more than 1 since we are only using CA)
- **school_closing** - 0: No measures - 1: Recommend closing - 2: Require closing (only some levels or categories, eg just high school, or just public schools - 3: Require closing all levels
- **workplace_closing** - 0: No measures - 1: Recommend closing (or work from home) - 2: require closing for some sectors or categories of workers - 3: require closing (or work from home) all-but-essential workplaces (eg grocery stores, doctors).
- **cancel_events** - 0: No measures - 1: Recommend canceling - 2: Require canceling gatherings_restrictions 0: No restrictions - 1: Restrictions on very large gatherings (the limit is above 1000 people) - 2: Restrictions on gatherings between 100-1000 people - 3: Restrictions on gatherings between 10-100 people - 4: Restrictions on gatherings of less than 10 people.
- **gatherings_restrictions** - 0: No restrictions - 1: Restrictions on very large gatherings (the limit is above 1000 people) - 2: Restrictions on gatherings between 100-1000 people - 3: Restrictions on gatherings between 10-100 people - 4: Restrictions on gatherings of less than 10 people.
- **transport_closing** - 0: No measures - 1: Recommend closing (or significantly reduce volume/route/means of transport available) - 2: Require closing (or prohibit most citizens from using it).
- **stay_home_restrictions** - 0: No measures - 1: recommend not leaving house - 2: require not leaving house with exceptions for daily exercise, grocery shopping, and "essential" trips - 3: Require not leaving house with minimal exceptions (e.g. allowed to leave only once every few days, or only one person can leave at a time, etc.).
- **internal_movement_restrictions** - 0: No measures - 1: Recommend closing (or significantly reduce volume/route/means of transport) - 2: Require closing (or prohibit most people from using it).
- **international_movement_restrictions** - 0: No measures - 1: Screening - 2: Quarantine arrivals from high-risk regions - 3: Ban on high-risk regions - 4: Total border closure.
- **information_campaigns** - 0: No COVID-19 public information campaign - 1: public officials urging caution about COVID-19 - 2: coordinated public information campaign (e.g. across traditional and social media).

- **testing_policy** - 0: No testing policy - 1: Only those who both (a) have symptoms AND (b) meet specific criteria (eg key workers, admitted to hospital, came into contact with a known case, returned from overseas) - 2: testing of anyone showing COVID-19 symptoms - 3: open public testing (eg "drive through" testing available to asymptomatic people).
- **contact_tracing** - 0: No contact tracing - 1: Limited contact tracing, not done for all cases - 2: Comprehensive contact tracing, done for all cases.
- **stringency_index** - Stringency of governmental responses.
- **retail_and_recreation_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as retail and recreation
- **grocery_and_pharmacy_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as grocery stores and pharmacies
- **parks_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as outdoor parks
- **transit_stations_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as transit stations
- **workplaces_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as work places
- **residential_percent_change_from_baseline** - comparison of pre-Covid-19 pandemic to Covid-19 pandemic travel trends to destinations classified as residential

wb - World Bank Data

Source

<URL: https://covid19datahub.io>

References

Guidotti, E., Ardia, D., (2020), "COVID-19 Data Hub", Journal of Open Source Software 5(51):2376, doi: 10.21105/joss.02376 (URL: https://doi.org/10.21105/joss.02376).