Final Project 632 Rough Draft

Nic James, Sri Chandu, Thomas Li

05/21/2020

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 - Nic

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 base 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

key_google_mobility - Identifier used in Google Mobility Reports.

key_apple_mobility - Identifier used in Apple Mobility Reports.

wb - World Bank Data

Questions of Interest

Using the base data set

1. What model using the policy measures is the best predictor of deaths / confirmed?

- Response: deaths / confirmed
- Predictors: Base data set

2.

Using the World Bank Data

1. How does the economic profile of the country affect the mortality rate from COVID over the year 2020?

- Response: Number of deaths/Confirmed Cases
- Predictors: GDP per capita, GDP per capita growth, Poverty rate, base data set

2. What is the effect of air pollution (or exposure to air pollution) to the number of cases and the mortality rate from COVID?

- Response: Number of cases and Number of deaths/Confirmed Cases
- Predictors: Pollution in mcg, base data set

Using the Google Mobility Data

1. Are policy measures that are non-restrictive with movement significant in preventing spread of Covid-19?

- Response: deaths
- Predictors: any of the policy measures that don't specifically prevent people from freely moving/travelling (i.e. testing policy, info campaigns, contract tracing)

2. 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

Exploratory Analysis

Base Data We started 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.

```
## 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
```

```
## # A tibble: 6 x 38
## # Groups:   id [1]
##   id      date       vaccines tests confirmed recovered deaths  hosp  vent   icu
##   <chr>   <date>        <dbl> <dbl>     <dbl>     <int>  <int> <dbl> <int> <int>
## 1 1c20e6~ 2020-03-15       NA  8773       478        NA      6    NA    NA    NA
## 2 1c20e6~ 2020-03-16       NA 10874       588        NA     11    NA    NA    NA
## 3 1c20e6~ 2020-03-17       NA 13162       732        NA     14    NA    NA    NA
## 4 1c20e6~ 2020-03-18       NA 16063       893        NA     17    NA    NA    NA
## 5 1c20e6~ 2020-03-19       NA 19497      1067        NA     19    NA    NA    NA
## 6 1c20e6~ 2020-03-20       NA 25598      1283        NA     24    NA    NA    NA
## # ... with 28 more variables: population <dbl>, 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_3 <chr>, iso_alpha_2 <chr>, iso_numeric <int>, currency <chr>,
## #   administrative_area_level <int>, administrative_area_level_1 <chr>,
## #   administrative_area_level_2 <chr>, administrative_area_level_3 <lgl>,
## #   latitude <dbl>, longitude <dbl>, key <chr>, key_google_mobility <chr>,
## #   key_apple_mobility <chr>, key_numeric <int>, key_alpha_2 <chr>
```

```
## # A tibble: 6 x 38
## # Groups:   id [1]
```

```
##   id     date       vaccines  tests confirmed recovered deaths  hosp  vent   icu
##   <chr>  <date>        <dbl>  <dbl>     <dbl>     <int>  <int> <dbl> <int> <int>
## 1 1c20e~ 2021-03-09 10925581 4.33e7   3608022        NA  54620  4007    NA  1045
## 2 1c20e~ 2021-03-10 11062505 4.35e7   3611490        NA  54877  3900    NA  1009
## 3 1c20e~ 2021-03-11 11220508 4.36e7   3615049        NA  55132  3766    NA  1013
## 4 1c20e~ 2021-03-12 11428034 4.38e7   3617356        NA  55336  3649    NA   976
## 5 1c20e~ 2021-03-13 11881857 4.39e7   3621094        NA  56372  3842    NA   943
## 6 1c20e~ 2021-03-14 11883375 4.40e7   3623342        NA  56522  3513    NA   919
## # ... with 28 more variables: population <dbl>, 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_3 <chr>, iso_alpha_2 <chr>, iso_numeric <int>, currency <chr>,
## #   administrative_area_level <int>, administrative_area_level_1 <chr>,
## #   administrative_area_level_2 <chr>, administrative_area_level_3 <lgl>,
## #   latitude <dbl>, longitude <dbl>, key <chr>, key_google_mobility <chr>,
## #   key_apple_mobility <chr>, key_numeric <int>, key_alpha_2 <chr>
```

We then looked at all of the variables to look at what variables we should use. We narrowed it down to date, confirmed, deaths, tests, population, latitude, longitude, school closing, workplace closing, canceled events, transport closing, stay home restrictions, internal movement restrictions, international movement restrictions, information campaigns, testing policy, contact tracing, and stringency index for the base data set.

We realized that the policy measures needed to be turned into factors before we could run a regression model. We discovered that factors need to have 2 or more levels in order to work so we removed cancel_events, international_movement, and transport_closing.

```
## The following object is masked from package:tidyr:
##
##     population
```

We set the null and full models using the data with factors.

We ran a basic ANOVA to make sure that there was at least one sgnificant variable in our model. We accordingly can now reject the null hypothesis and assume that there is at least one significant predictor in this model.

```
## Analysis of Variance Table
##
## Model 1: deaths ~ 1
## Model 2: deaths ~ date + confirmed + tests + latitude + longitude + population +
##     fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + finternal_movement_restrictions +
##     finformation_campaigns + ftesting_policy + fcontact_tracing +
##     stringency_index
##   Res.Df        RSS Df  Sum of Sq      F    Pr(>F)
## 1    364 8.1640e+10
## 2    350 3.9805e+08 14 8.1242e+10 5102.5 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We ran the pairs() to start looking to see if the assumptions have been met however, there are too many variables for it to be useful. We then

```
##
## Call:
## lm(formula = deaths ~ date + confirmed + tests + latitude + longitude +
##     population + fschool_closing + fworkplace_closing + fgatherings_restrictions +
##     fstay_home_restrictions + finternal_movement_restrictions +
##     finformation_campaigns + ftesting_policy + fcontact_tracing +
##     stringency_index, data = fbase_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3436.3  -524.6     0.0   597.9  3357.6
##
## Coefficients: (5 not defined because of singularities)
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       4.305e+05  1.858e+05   2.316 0.021122 *
## date                             -2.467e+01  1.013e+01  -2.435 0.015407 *
## confirmed                        -5.084e-04  1.171e-03  -0.434 0.664484
## tests                             1.464e-03  1.596e-04   9.170  < 2e-16 ***
## latitude                                 NA         NA      NA       NA
## longitude                                NA         NA      NA       NA
## population                               NA         NA      NA       NA
## fschool_closing3                 -4.059e+03  5.083e+02  -7.985 2.05e-14 ***
## fworkplace_closing2               2.587e+03  1.345e+03   1.924 0.055162 .
## fworkplace_closing3               3.578e+03  1.670e+03   2.142 0.032864 *
## fgatherings_restrictions3        -1.007e+04  2.039e+03  -4.937 1.23e-06 ***
## fgatherings_restrictions4        -1.168e+04  2.208e+03  -5.288 2.18e-07 ***
## fstay_home_restrictions2         -6.785e+03  2.352e+02 -28.845  < 2e-16 ***
## finternal_movement_restrictions1  3.309e+03  4.645e+02   7.123 6.07e-12 ***
## finternal_movement_restrictions2         NA         NA      NA       NA
## finformation_campaigns2                  NA         NA      NA       NA
## ftesting_policy2                  1.447e+03  3.814e+02   3.795 0.000174 ***
## ftesting_policy3                  7.376e+02  4.857e+02   1.519 0.129761
## fcontact_tracing2                -1.376e+03  3.063e+02  -4.492 9.61e-06 ***
## stringency_index                  5.036e+02  7.795e+01   6.460 3.51e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1066 on 350 degrees of freedom
## Multiple R-squared:  0.9951, Adjusted R-squared:  0.9949
## F-statistic:  5102 on 14 and 350 DF,  p-value: < 2.2e-16
```

We ran through two more rounds of summary and removed information_campaign and internal_movement_restrictions due to singularities leaving with our final full model called mod.full3.

```
##
## Call:
## lm(formula = deaths ~ date + 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
## -3724.5  -599.6    54.7   595.4  3923.0
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 7.521e+05  1.926e+05   3.905 0.000113 ***
## date                       -4.146e+01  1.053e+01  -3.937 9.96e-05 ***
## confirmed                  -3.413e-03  1.173e-03  -2.909 0.003853 **
## tests                       1.731e-03  1.658e-04  10.444  < 2e-16 ***
## fschool_closing3           -4.973e+03  5.255e+02  -9.463  < 2e-16 ***
## fworkplace_closing2         6.514e+03  1.310e+03   4.971 1.04e-06 ***
## fworkplace_closing3         8.563e+03  1.620e+03   5.285 2.21e-07 ***
## fgatherings_restrictions3 -2.782e+03  1.885e+03  -1.476 0.140709
## fgatherings_restrictions4 -3.031e+03  1.971e+03  -1.538 0.124907
## fstay_home_restrictions2  -6.775e+03  2.513e+02 -26.959  < 2e-16 ***
## ftesting_policy2            1.843e+03  4.032e+02   4.572 6.72e-06 ***
## ftesting_policy3            9.742e+02  5.177e+02   1.882 0.060717 .
## fcontact_tracing2          -1.231e+03  3.265e+02  -3.771 0.000191 ***
## stringency_index            1.844e+02  6.816e+01   2.706 0.007141 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1139 on 351 degrees of freedom
## Multiple R-squared:  0.9944, Adjusted R-squared:  0.9942
## F-statistic:  4810 on 13 and 351 DF,  p-value: < 2.2e-16
```

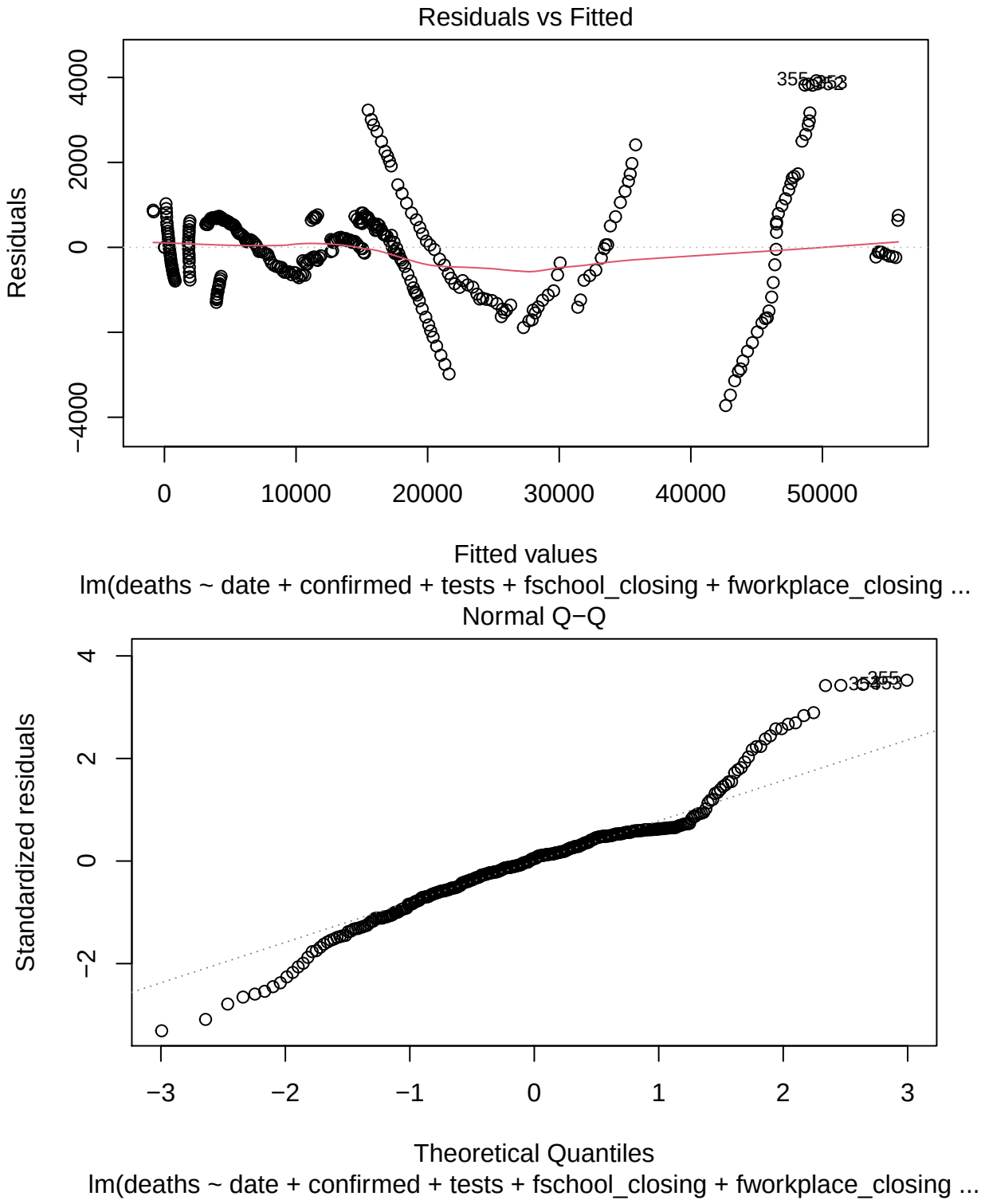
Google Mobility

Diagnostic Checks

```
plot(mod.full3, 1:2)
```

Base Data:

```
## Warning: not plotting observations with leverage one:
##   1, 2
```



This is heavy tailed and the residuals have a distinct pattern. So we need to look at transforming the data.

```
pt <- powerTransform(cbind(confirmed, tests, fschool_closing,
            fworkplace_closing, fgatherings_restrictions, fstay_home_restrictions  ,
            ftesting_policy, fcontact_tracing, stringency_index) ~ 1, data = fbase_data)
```

```
## Warning in estimateTransform.default(X, Y, weights, family, ...): Convergence
## failure: return code = 1
```

```
summary(pt)
```

```
## bcPower Transformations to Multinormality
##                           Est Power Rounded Pwr Wald Lwr Bnd Wald Upr Bnd
## confirmed                    0.2307        0.23       0.2000       0.2615
## tests                        0.2637        0.26       0.2279       0.2996
## fschool_closing             42.3912       42.39      38.0423      46.7402
## fworkplace_closing           1.0507        1.00       0.7623       1.3391
## fgatherings_restrictions    16.6996       16.70      14.9401      18.4591
## fstay_home_restrictions     -3.2787       -3.28      -3.8568      -2.7006
## ftesting_policy              2.0644        2.00       1.7407       2.3882
## fcontact_tracing             2.5139        2.00       1.9626       3.0653
## stringency_index             1.0643        1.00       0.6628       1.4658
##
## Likelihood ratio test that transformation parameters are equal to 0
##  (all log transformations)
##                                             LRT df       pval
## LR test, lambda = (0 0 0 0 0 0 0 0 0) 3164.876  9 < 2.22e-16
##
## Likelihood ratio test that no transformations are needed
##                                            LRT df       pval
## LR test, lambda = (1 1 1 1 1 1 1 1 1) 3453.56  9 < 2.22e-16
```

Google Mobility

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.

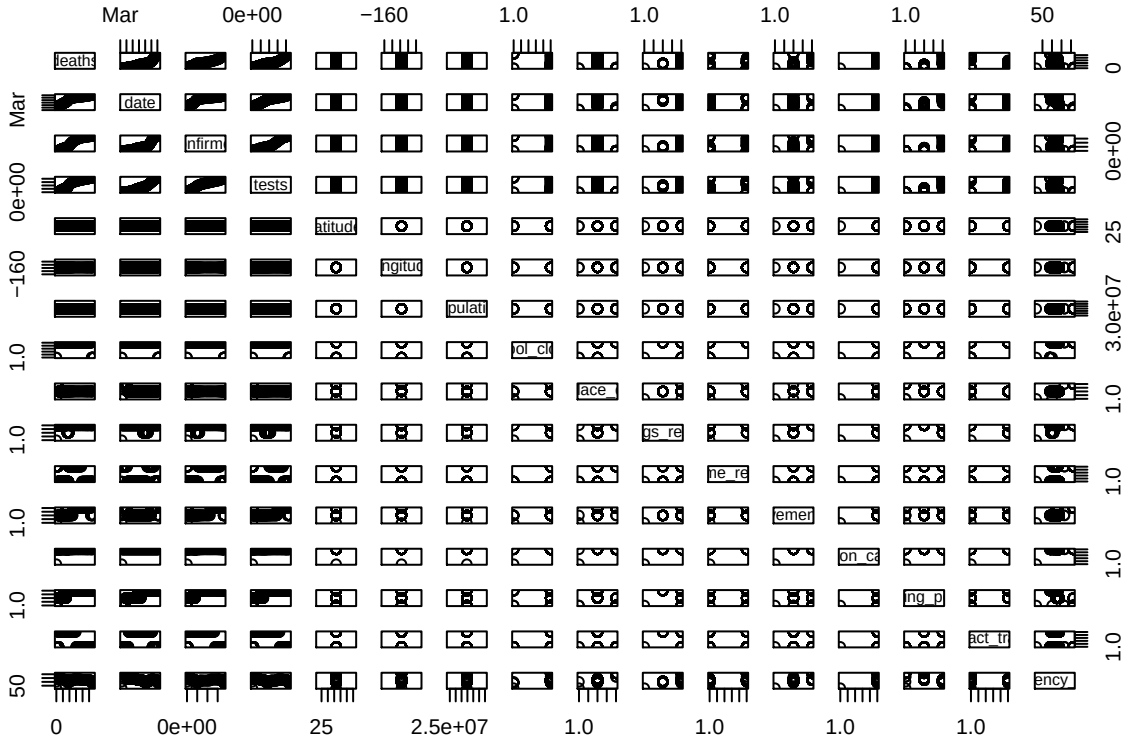
Appendicies

Appendix 1: R Code

Code Chunk 1:

Appendix 2 (optional): Exploratory analysis not used in final paper

```
pairs(deaths ~ date + confirmed + tests + latitude  + longitude + population +  fschool_closing +
            fworkplace_closing + fgatherings_restrictions + fstay_home_restrictions + finternal_movemen
            finformation_campaigns + ftesting_policy + fcontact_tracing + stringency_index, data = fbas
```



Google Mobility Data

Appendix 3: World Bank Data Story

Appendix 4: 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. ## 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).
```