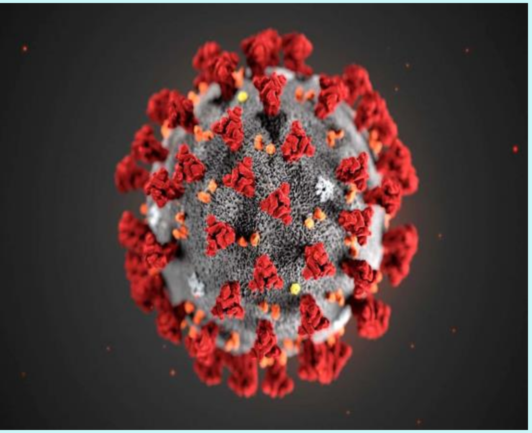
Nations’ Response to Covid-19: A Critical Look

By Keshav Narasimhan
Upper St. Clair High School

Why I chose this Project?



Covid-19 is a pandemic that has claimed over 400,000 lives in the US.

Despite having some of the most advanced medical infrastructure in the world, the United States has failed to contain this pandemic.

I wanted to understand if there are a few key factors which could have helped the US and other nations minimize the impact of this pandemic. If so, I wanted to find out what those factors are.

Hypotheses

Hypothesis # 1

Countries with an overall higher population density will have more cases per 1 million due to proximity, making the virus easier to spread.

Explanation: Higher population density will make social distancing lot more challenging, and the virus will spread easier.

Hypothesis # 2

Economic prosperity will not be as important of a variable when determining if a country successfully controlled the virus.

Explanation: By following the news, I knew that countries like the US, Germany and Belgium had a hard time controlling this virus.

Data Analyses Approach

Collect worldwide COVID-19 data from credible databases

Initially, keep the analysis as broad as possible. Collect data on all the countries.

Visually look for trends by plotting scatter plots.

Narrow down the countries to analyze based on the trust factor of the data; maintain adequate diversity in the countries chosen for analysis.

Statistically analyze the explanatory variables for inherent correlation.

Construct a regression model to predict the number of cases/deaths

Refine the model using R program to get the best fit. (high adjusted R-square value.)

Arrive at key conclusions from the model.

Data Collection

	ï..countries	Length.of.Lockdown	Population.Density	GDP.Per.Capita	HDI	Tests.1.million	Democracy.Index	Cases.Per..1M	Deaths.M	Continent	Population
1	Afghanistan	0	56.94	469.9191	0.496	5995	28.5	1384	60.0	Asia	39415253
2	Albania	80	104.61	5303.1978	0.791	112438	58.9	24567	453.0	Europe	2876025
3	Algeria	0	17.73	3975.5104	0.759	NA	40.1	2373	64.0	Africa	44292186
4	Andorra	0	163.84	40887.4216	0.857	2328674	0.0	121756	1203.0	Europe	77335
5	Angola	0	24.71	2670.8507	0.574	5208	37.2	576	14.0	Africa	33430593
6	Antigua and Barbuda	0	235.48	17112.8211	0.776	83347	0.0	1982	61.0	North America	98384
7	Argentina	52	16.16	10041.4633	0.830	127820	70.2	40808	1025.0	South America	45428197
8	Armenia	41	99.44	4622.7332	0.760	213273	55.4	55863	1021.0	Asia	2966361
9	Australia	52	3.00	54763.2024	0.938	492581	90.9	1120	35.0	Oceania	25665434
10	Austria	98	106.14	49700.7618	0.914	459984	82.9	44480	811.0	Europe	9035260
11	Azerbaijan	152	116.25	4781.9462	0.754	230408	27.5	22426	301.0	Asia	10190399
12	Bahamas	0	27.75	34863.7421	0.805	145401	0.0	20491	443.0	North America	395349
13	Bahrain	0	1982.91	23503.9771	0.838	1503508	25.5	57210	211.0	Asia	1734135
14	Bangladesh	51	1181.00	1846.4164	0.614	21317	58.8	3206	48.0	Asia	165608438
15	Barbados	36	667.50	18148.9585	0.813	323640	0.0	4020	31.0	North America	287573
16	Belarus	0	45.59	6673.5313	0.817	458373	24.8	24780	172.0	Europe	9447571
17	Belgium	89	375.73	46198.3103	0.919	669070	76.4	59118	1775.0	Europe	11617976
18	Belize	0	17.79	4815.1719	0.720	166621	0.0	29130	722.0	North America	401649
19	Benin	0	104.18	1220.4928	0.520	33180	50.9	296	4.0	Africa	12299096
20	Bhutan	163	21.75	3360.6172	0.617	530464	53.0	1096	1.0	Asia	776364
21	Bolivia	131	10.29	3552.0688	0.703	42892	48.4	16697	835.0	South America	11761996
22	Bosnia and Herzegovina	0	68.57	6108.5580	0.769	176185	48.6	36528	1393.0	Europe	3269277
23	Botswana	28	3.96	7961.3653	0.728	258172					
24	Brazil	0	25.00	8755.2712	0.761	134016					
25	Brunei	0	73.08	31086.3292	0.845	202155					

Source: Worldometer Website

www.worldometer.com

Response Variable:

1. No. of Covid-19 cases
2. No. of Covid- 19 deaths

Explanatory Variable:

1. Population density
2. GDP Per Capita
3. Human Development Index (HDI)
4. Tests per million
5. Democracy Index
6. Length of Lockdown.
7. Mask Mandate Implementation

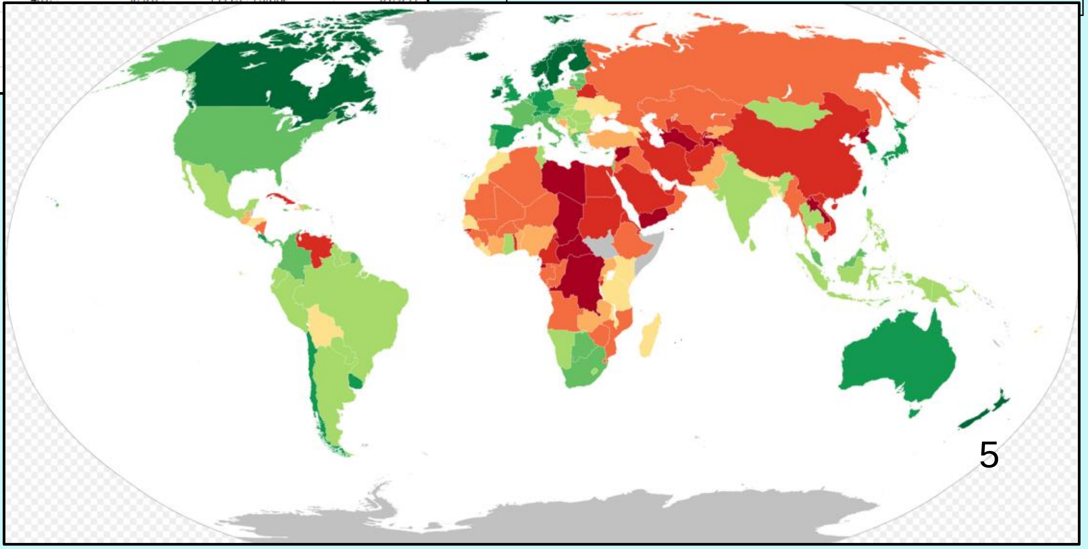
Graphical Representation of Democracy Index.

Source: Economist Intelligence Unit

Dark Green – More Democratic

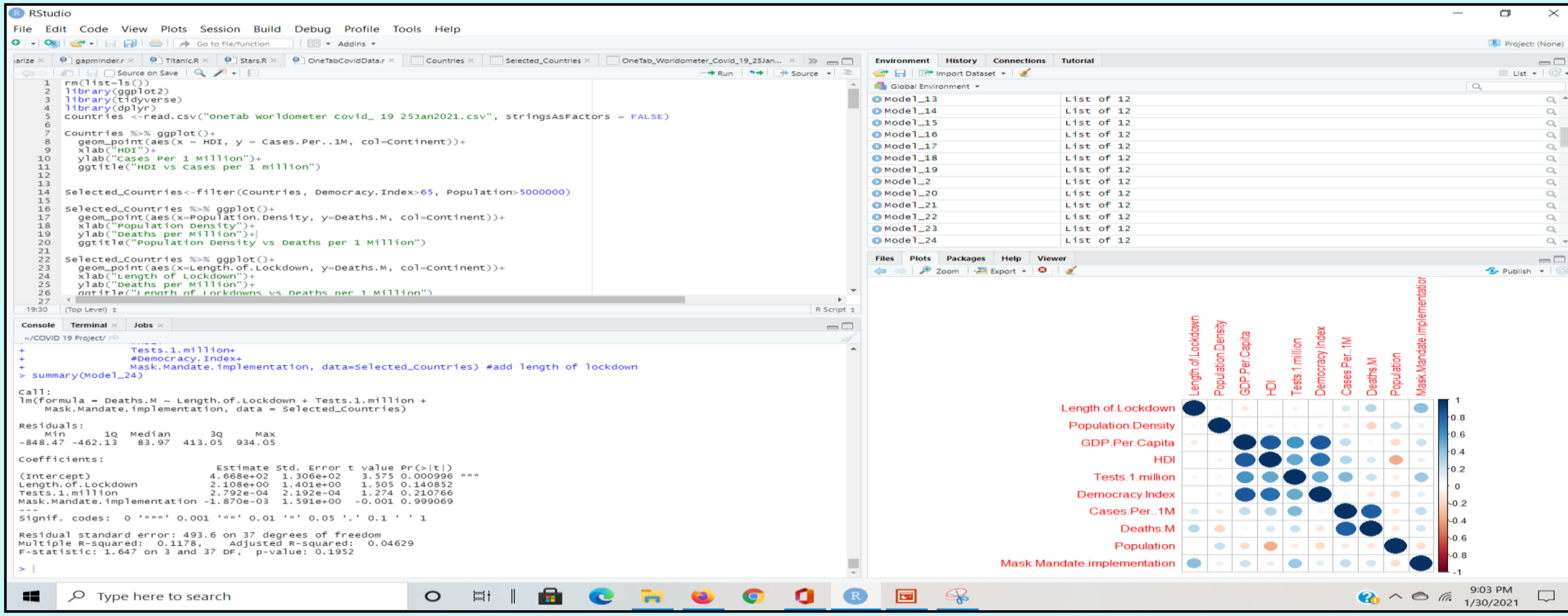
Dark Red – Less Democratic

*Note China, a highly populated country is very red. Therefore, I will be excluding this country.



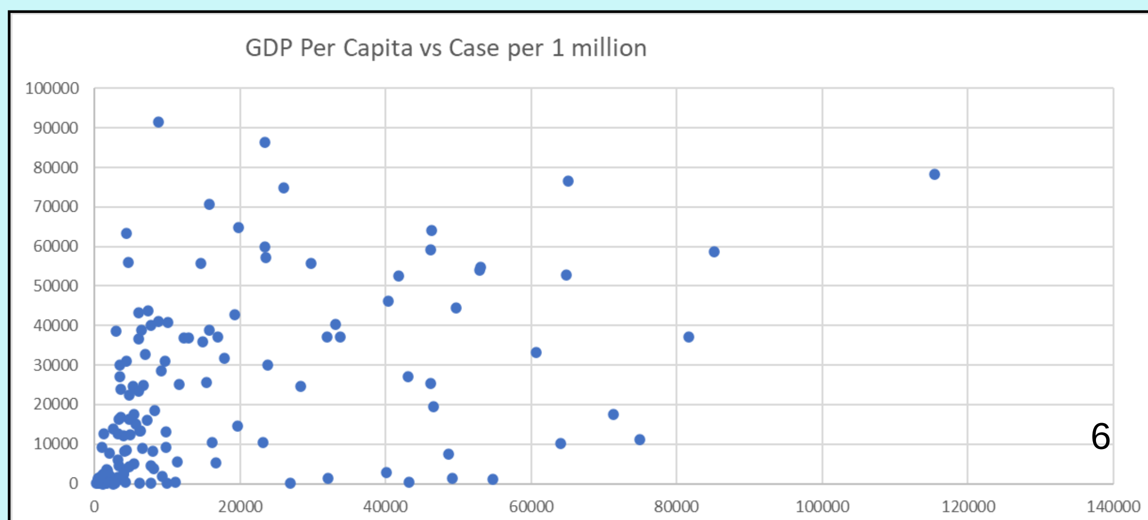
Statistical Program

R-Studio



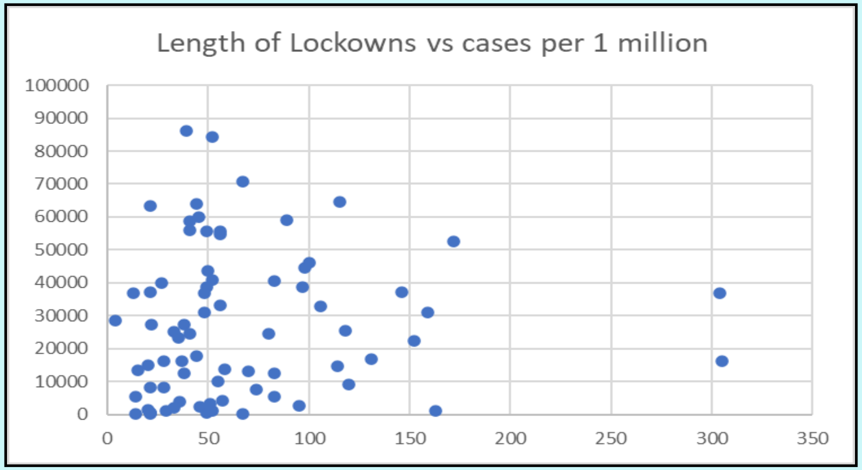
I used statistical program R studio in conjunction with Excel to look for trends.

Excel

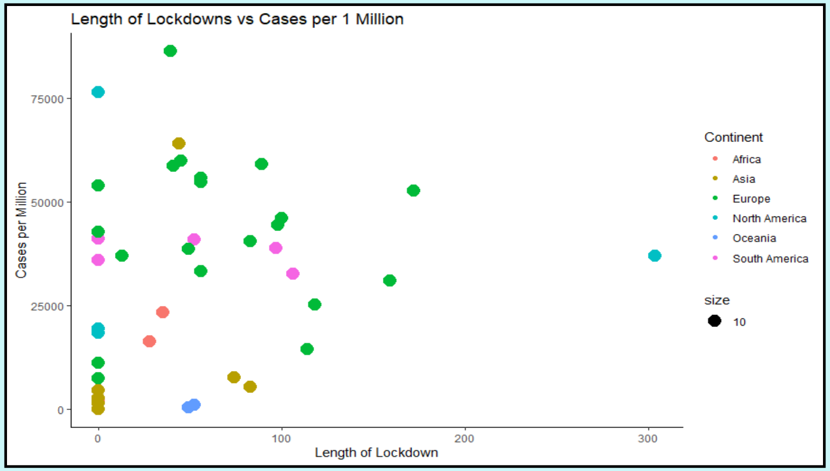


Scatter Plots to Search for Trends

All 190 countries



41 selected countries



Looking at all 190 countries in a single scatterplot made it difficult to see any obvious patterns. Therefore, I used two criteria to further filter the data:

1. Population > 5M (to ensure countries are chosen with a substantial number of cases and deaths)
2. Democracy Index > 65 (to ensure that a free media can monitor the accuracy of government data.)

41 countries selected based on the criteria listed above.

Even amongst these selected countries, data was still scattered and hard to spot a pattern.

– This is a hard problem to solve. Will require analysis of multiple variables to predict the number of cases.

Correlation Plot in R

Since the scatterplot did not show any strong pattern, I used the R program to look for correlation statistically.

The analysis showed that the following 4 factors were strongly correlated:

1. GDP Per Capita
2. HDI
3. Tests per 1 million, and
4. Democracy Index

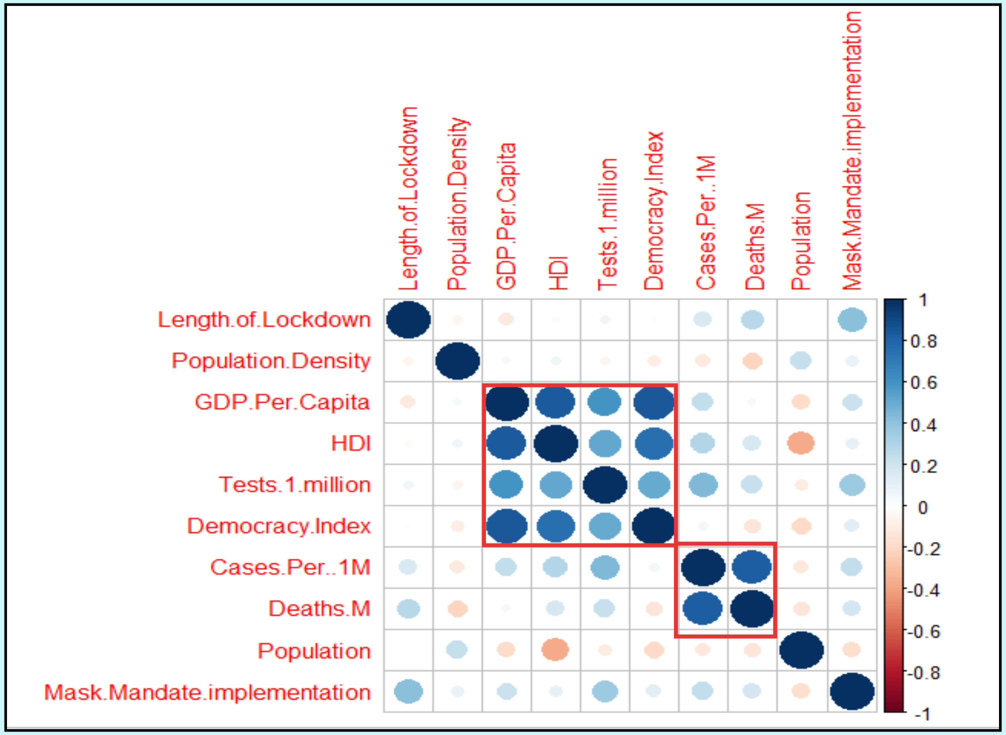
This meant I could potentially use fewer explanatory variables to construct the linear model. I used adjusted R-square value to confirm this.

Similarly,

1. Death per million, and
2. Cases per million

were also strongly correlated.

It might be possible to use only 1 of these two response variables.




```
selected_countries<-filter(selected_countries, Democracy.Index>65, Population>5000000)
drop.cols <- c('ï..countries', 'Continent')
NDF <- selected_countries %>% select(-one_of(drop.cols))
cor.table = cor(NDF)
corrplot(cor.table)
```

Multivariate Linear Regressions

```
Call:
lm(formula = Cases.Per..1M ~ Length.of.Lockdown + Population.Density +
    GDP.Per.Capita + HDI + Tests.1.million + Democracy.Index +
    Mask.Mandate.implementation, data = selected_countries)

Residuals:
   Min     1Q Median     3Q    Max
-35163  -9716  -2489  11128  47874

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  6.908e+04  6.316e+04   1.094   0.2820
Length.of.Lockdown           6.116e+01  6.050e+01   1.011   0.3194
Population.Density          -2.694e+01  2.062e+01  -1.307   0.2003
GDP.Per.Capita               3.697e-01  3.701e-01   0.999   0.3251
HDI                          9.594e+04  6.755e+04   1.420   0.1649
Tests.1.million              2.278e-02  1.076e-02   2.118   0.0418 *
Democracy.Index             -1.767e+03  6.694e+02  -2.640   0.0126 *
Mask.Mandate.implementation  1.145e+01  6.768e+01   0.169   0.8667
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19790 on 33 degrees of freedom
Multiple R-squared:  0.3774,    Adjusted R-squared:  0.2454
F-statistic: 2.858 on 7 and 33 DF,  p-value: 0.01907

>
```

In this model I used all 7 explanatory variables to construct the model that had an adjusted R-Squared value of 0.2454*. This is typically considered a low value.

*Note: A higher adjusted R-Squared value is better.

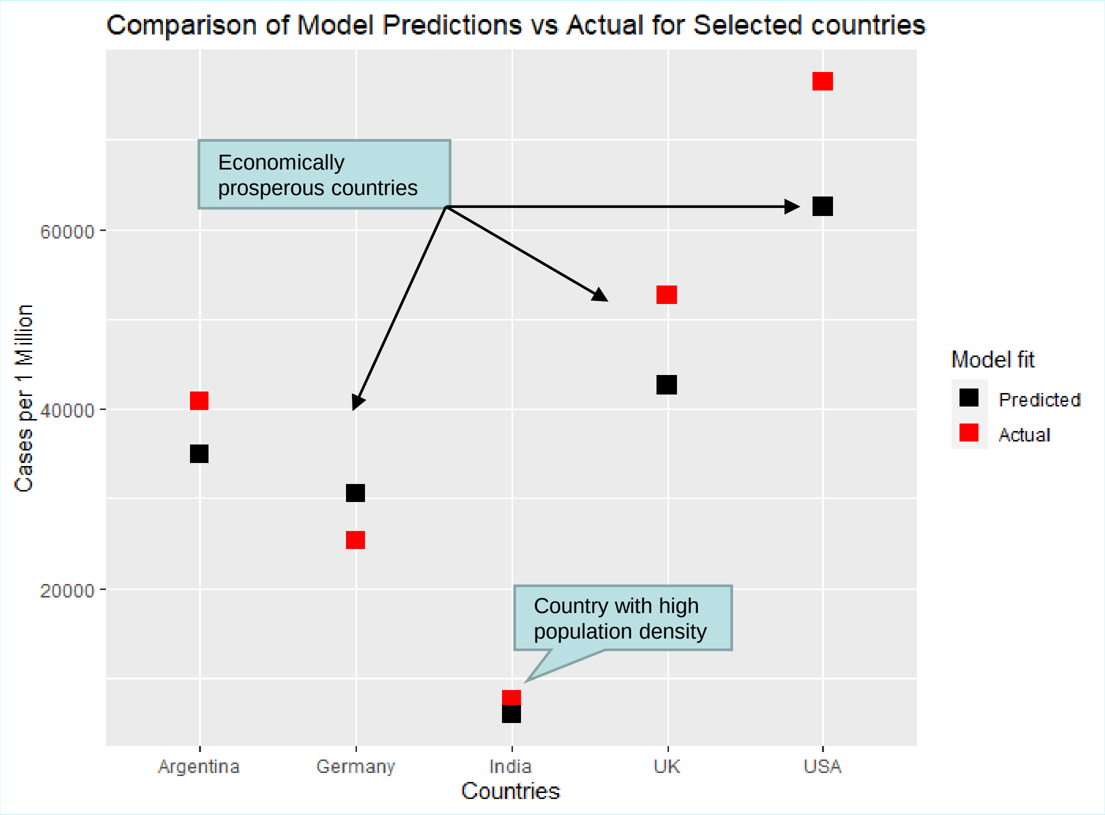
Attempts to Improve Fit…

```
Call:
lm(formula = Cases.Per..1M ~ Population.Density + GDP.Per.Capita +
    HDI + Tests.1.million + Democracy.Index, data = Selected_Countries)

Residuals:
   Min     1Q Median     3Q    Max
-37033 -12011  -2260  11420  45519

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         6.181e+04  5.966e+04   1.036   0.3072
Population.Density -2.654e+01  2.032e+01  -1.306   0.1999
GDP.Per.Capita      2.712e-01  3.409e-01   0.796   0.4317
HDI                 1.017e+05  6.512e+04   1.562   0.1273
Tests.1.million     2.543e-02  1.013e-02   2.511   0.0168 *
Democracy.Index    -1.664e+03  6.508e+02  -2.557   0.0150 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19680 on 35 degrees of freedom
Multiple R-squared:  0.3476,	Adjusted R-squared:  0.2544
F-statistic:  3.73 on 5 and 35 DF,  p-value: 0.008231
```

Even after trying different combinations of the explanatory variables, this is the best adjusted R-Square value that I could get.

It proved that it is a very challenging problem to solve.

Model Vs. Actual



Despite the model having a low adjusted R-Square value, the model is reasonably good in predicting the cases per 1 million.

```
ggplot(Selected_Countries[c(1,16,20,40,41),]) + geom_point(aes(x=ï..countries, y=Cases.Per..1M, color="Actual"), shape=15, size=4) +
  geom_point(aes(x=ï..countries, y=pred5, color="Prediction"),  shape=15, size=4) +
  xlab("Countries")+
  ylab("Cases per 1 Million")+
  ggtitle("Comparison of Model Predictions vs Actual for Selected countries")
```

Conclusions

I have three conclusions to make:

1. My Hypothesis #1 is rejected: Data showed that the countries that had a higher population density did not show higher numbers of cases.

2. My hypothesis #2 is correct: Economically prosperous countries, such as USA and Germany, did not show better capability in controlling the number of Covid-19 cases.

3. Predicting the number of cases in a country is a very difficult task. No one single factor was more dominant than the others in predicting the number of Covid-19 cases.