Multivariate Statistical Analysis

A Study of COVID-19 Data - Do Intervention Policies Matter?

2017010914 Wenxin Zhang

Abstract

As COVID-19 continues to spread, the eyes of the world have been focused on it. Different demographic and cultural characteristics of the population play an important role in determining the outbreak trajectories and clinical outcomes at the population level [1]. However, effective interventions may also be vital when combating COVID-19. In this study, we divide the process of the spread into two phases: the initial period where the government and the public are not fully aware and the second period after community transmission that arouses people's alertness. We focus on the following questions.

- How are countries performing when faced with the virus?
- Are developed countries better at combatting the virus?
- Do intervention policies matter?

We find that most countries are doing better in the second phase and the spread of COVID-19 has slowed down. Among them, developed countries are not performing better. Some of them have a lot of cumulative cases and they even have higher case fatality rate. Furthermore, by using CCA, linear regression, and kernel CCA, we find that intervention policies do matter.

Background and Significance

Since the beginning of 2020, COVID-19 has developed into a pandemic, influencing millions of people around the world. With few new cases confirmed in China, we are happy to see that our battle towards the virus has entered the last phase. However, the number of confirmed cases is still rising rapidly in other countries like the USA. Globally, as of 6:45 pm CEST, 16 May 2020, there have been 4,434,653 confirmed cases of COVID-19, including 302,169 deaths, reported to WHO [2].

Four months ago, when the virus first outbroke in China, many foreigners were confident of their healthcare systems and national strength. Some argue that how a country performs when facing the virus is related to its comprehensive capability in economic and health expenditure. However, as the virus continues to spread rapidly across the globe, governments realized that this may not be the case. They then start to weigh decisions about the stringency of their policies against social and economic concerns.

Critics blame some countries for their over-stringent policies while others claim that stringency is all we need. In this study, we aim to identify the relationship between a country's comprehensive strength and the performance when combating the virus, and further examine the effectiveness of control measures to inform policymakers and leaders in formulating management guidelines, and to provide directions for future research.

Most of the previous researches on COVID-19 focused on epidemiology. See Park et.al [1] for an excellent survey of the epidemiological characteristics of COVID-19. Other papers focus on using mathematical models to predict the spread and influence of the virus. Joseph T et.al estimated the basic reproductive number and warned that the virus is no longer contained in Wuhan. [3] There are fewer papers concerning policy and interventions. Wang et.al [4] presents evaluated the impact of non-pharmaceutical interventions on the epidemic in Wuhan. Hale et.al [5] tracked the stringency of government responses to COVID-19 and explore whether rising stringency of response affects the rate of infection. In our study, however, we also consider a country's comprehensive strength. Moreover, we look at cross-sectional data instead of time-series data and focus on variance and correlation.

Methods

Data Source

We use (1) basic economic and healthcare-related data, data from Jan 22 to May 10 (110 days in total) on (2) the number of confirmed, deaths and recovered cases, and (3) policies of 85 countries around the world. For simplicity, we will call them countryData, covidData and policyData in the remainder of the report.

- https://www.kaggle.com/dumbgeek/countries-dataset-2020
- https://www.kaggle.com/fernandol/countries-of-the-world
- https://github.com/covid19datahub/COVID19

Variable Creation

Just as Hellewell J et.al [6] pointed out, the course of an epidemic is defined by a series of key factors, some of which are poorly understood at present for COVID-19. Therefore, in this study, we will try to focus on some of the variables that are often used as indicators of the development of the virus. For detailed variables description, see Appendix.

Analytical Methods

First, we use K-means to classify the countries into 3 categories based on countryData, which we assume indicates the country's capability and healthcare strength. We also conduct principal component analysis(PCA) to see if there is any chance for dimension reduction.

Second, with covidData, we create several variables as described above based on related epidemiology researches. We conduct normality tests on the data and perform factor analysis(FA) to find the latent factor variables. Again, we use K-means to classify the countries into 3 categories and compare the new classification with the one obtained using countryData. In this section, we will try to show how countries perform when faced with the virus and whether developed countries perform better.

Third, to see whether policies matter, we conduct CCA on the pair of covidData and countryData and the pair of covidData and policData to identify and quantify the associations between these sets of variables. Meanwhile, we try to conduct linear regression to quantify the influence of other variables on R1 and R2. We also use kernel CCA to identify the possible nonlinear relationship between covidData and policyData. At last, we perform regression analysis to quantify the possible relationship between R1, R2 and other variables.

Multiple plots are also used for better visualization and to facilitate illustration.

Results

Part I. countryData

Descriptive statistics

Use boxplots for better illustration. (Data used: standardized countryData)

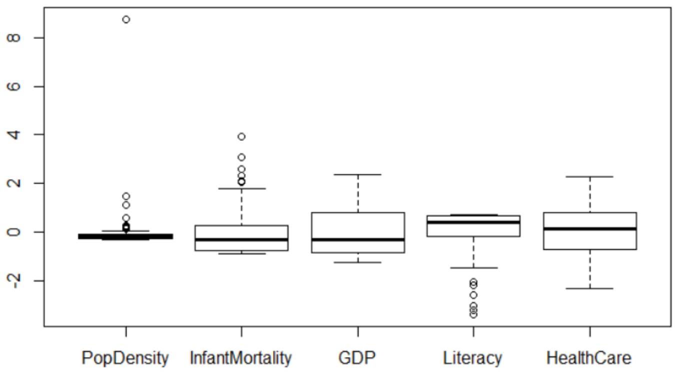


Figure 1: Boxplot for countryData

We can see from Figure 1 that the variances mainly lie in GDP and HealthCare. Moreover, there re many outliers when it comes to PopDensity and Literacy. Therefore, the level of development in various regions of the world is still uneven.

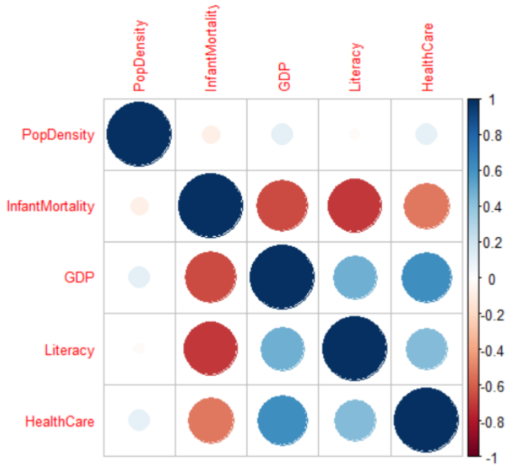


Figure 2: A Visualization of the Correlation Matrix of countryData

Based on Figure 2, we can see that GDP, Literacy, and HealthCare are positively correlated while InfantMortality is negatively correlated with them, which is in accordance with common knowledge. With high GDP, Literacy, and HealthCare but low InfantMortatlity, we may regard the country as a developed one.

Clustering Results

Now we will use K-means and to cluster the countries. See Appendix for how to choose the number of clusters.

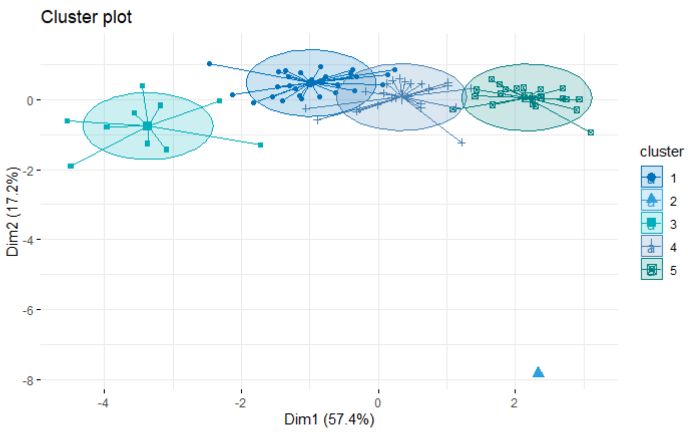


Figure 3: Cluster Plot for countryData

We get K-means clustering with 5 clusters of sizes 27, 1, 10, 24, 23.

$$\frac{\text{between\_ss}}{\text{total\_ss}} = 75.6\%$$

For cluster means, see Appendix. Here we present some examples.

Cluster ID	Examples
1	Russia, Iran, Greece
2	Singapore
3	Iraq, India, Pakistan
4	Argentina, Chile, South Africa
5	United States, Japan, Canada

So the countries in Cluster1 have low population density and average scores in other indexes. Cluster2 can be seen as an outlier. It only involves Singapore, which has a large population density, low InfantMortality and high GDP, Literacy, and HealthCare. Cluster3 contains countries with high population density, InfantMortality but low GDP, Literacy, and HealthCare, indicating that they are developing countries. Countries in Cluster4 are similar to the ones in Cluster1 but with high HealthCare. Cluster5 involves developed countries.

Part II: covidData

Descriptive statistics

Use boxplots for better illustration. (Data used: standardized covidData)

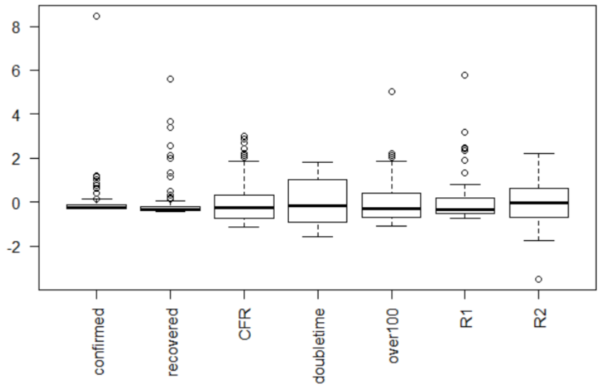


Figure 4: Boxplot for covidData

We can see from Figure 4 that the variances mainly lie in Doubletime and R2. Since population density varies, it is reasonable that Doubletime varies. Moreover, there re many outliers when it comes to the number of recovered people and R1. Therefore, the spread of COVID-19 in various regions of the world is quite different.

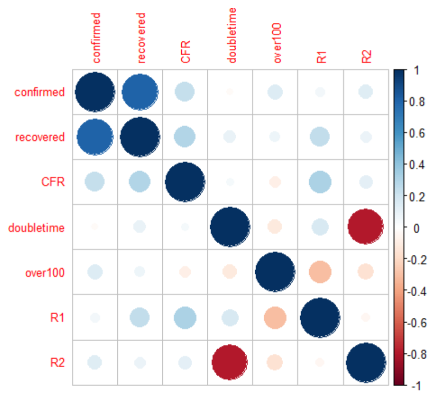


Figure 5: A Visualization of the Correlation Matrix of covidData

Based on Figure 5, we can see that confirmed and recovered are positively correlated while R2 and doubletime is negatively correlated, which is in accordance with common knowledge. Apart from them, most variables are weakly correlated. Therefore, we don't perform any type of dimension reduction.

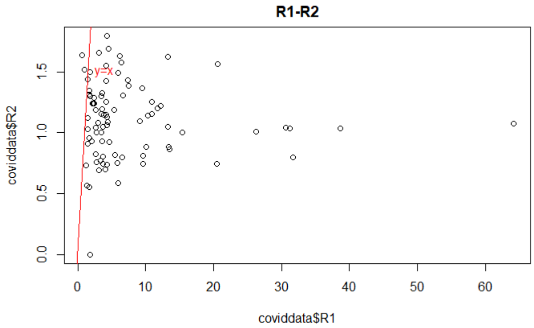


Figure 6: Scatterplot of R1-R2

Based on Figure 6, we can see that most of the countries lie below the line $y = x$, indicating a decrease in the effective reproductive number. The two exceptions are Singapore and Kuwait, which is in accordance with the reality. For example, as a global transportation hub, Singapore has witnessed a rise in confirmed cases after months of optimistic reactions and seemingly mild spread. Further, from Figure 5, we can see that the distributions of R1 and R2 are different. Let us conduct a Kolmogorov-Smirnov test.

```
        Two-sample Kolmogorov-Smirnov test

data:  coviddata$R1 and coviddata$R2
D = 0.88235, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Since $\text{p-value} < 2.2e-16$, we reject the null hypothesis, indicating that though community transmission starts in the second phase, most countries start to perform better than in the first phase. We may assume that public awareness, government interventions, and healthcare resources deployment play an important role in fighting COVID-19.

Factor Analysis

In this section, we want to perform factor analysis on covidData to describe variability among variables in terms of a potentially lower number of unobserved factors.

First, we need to test the data for normality.

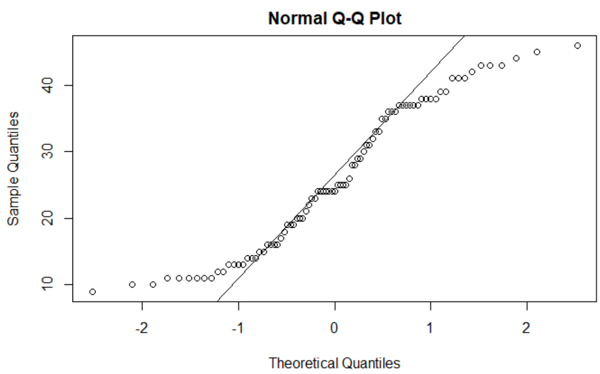


Figure 7: Normal Q-Q Plot for Doubletime

Based on Q-Q plots, we can find that some variables may not satisfy the normality assumption, like Doubletime as shown in Figure 7. Therefore, we use PC to perform factor analysis.

```
Loadings:
           RC1    RC2    RC3   
confirmed   0.921              
recovered   0.926         0.115
CFR         0.419         0.520
doubletime         0.935  0.164
over100     0.266        -0.755
R1          0.178  0.150  0.776
R2                -0.951  0.126

                 RC1   RC2   RC3
SS loadings    1.992 1.814 1.502
Proportion Var 0.285 0.259 0.215
Cumulative Var 0.285 0.544 0.758
```

The first factor can be interpreted as the number of confirmed cases since the number of covered cases is highly correlated with it. It can illustrate the current pressure on healthcare resources and reflect how big the influence is in the country. Doubletime and R2 contribute oppositely to the second factor. The third factor focuses on the speed of spread during the first phase.

In conclusion, we cannot obtain a feasible set of factors to represent covidData because variables like Doubletime and R2 focus more on recent performance which may depend on how the public reacts to the virus and how effective the actions taken by the country are. Therefore, we will stick to the original covidData in the remainder of the report.

Clustering Results

Now we will use K-means to cluster the countries. See Appendix for how to choose the number of clusters.

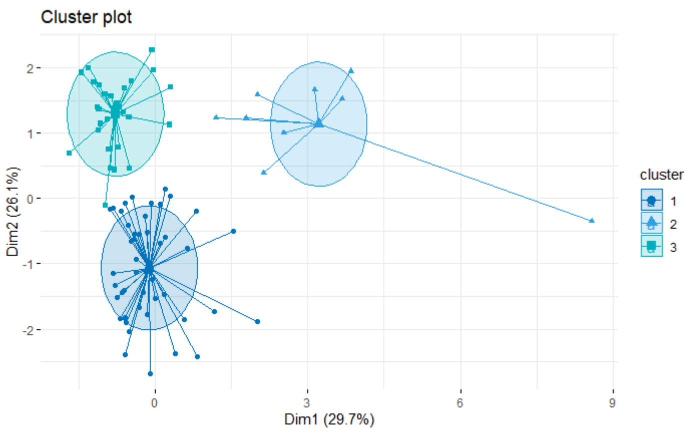


Figure 8: Cluster Plot for countryData

Based on Figure 8, the clustering result is We get K-means clustering with 3 clusters of sizes 46, 9, 30.

$$\frac{\text{between\_ss}}{\text{total\_ss}} = 39.8\%$$

For cluster means, see Appendix. Here we present some examples.

Cluster ID	Examples
1	Argentina,Brazil,Canada
2	Iran, United States, Italy
3	New Zealand, Nepal, Norway

Based on the above results, Cluster1 involves countries that have relatively low confirmed cases, Doubletime, and R1. However, they tend to have high R2, indicating invalid intervention actions.

The countries in Cluster2 show high confirmed cases, CFR and R1, indicating unready when first faced with the virus. But they have relatively low R2, which means they start to take effective actions to tackle the issue. Note that though the United States is in Cluster2, it is more like an outlier. Roaring confirmed cases may be the reason. Also, note that both R1 and R2 of the United States are not very high despite lots of confirmed cases since the former reflects the spread of the virus while the latter shows how big the influence is.

	Country <fctr>	confirmed <int>	recovered <int>	CFR <dbl>	doubletime <int>	over100 <int>	R1 <dbl>	R2 <dbl>
82	United States	1309550	212534	0.06542528	24	41	2.437123	1.237526

Cluster3 contains countries that are less affected by the virus. They have low confirmed cases and low R1, R2. Doubletime is higher since they have just got influenced by the virus lately and the cumulative quantity is still small.

Canonical Correlation Analysis

CCA is concerned with explaining covariance structure of two sets of variables through a few linear combinations of them. The main objective of CCA is to identify and quantify the associations between two sets of variables.

countryData & covidData - Are developed countries better at combatting the virus?

For simplicity, we won't list all clusters obtained above. If we fix the number of clusters to be 3 in countryData and perform K-means again, the two clustering results are mismatching. We can see clearly that developed countries are not always performing better when faced with the virus. In this section, we will examine the relationship between countryData and covidData.

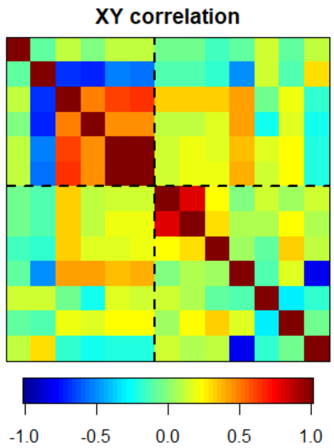


Figure 9: Correlation Matrix of countryData and covidData

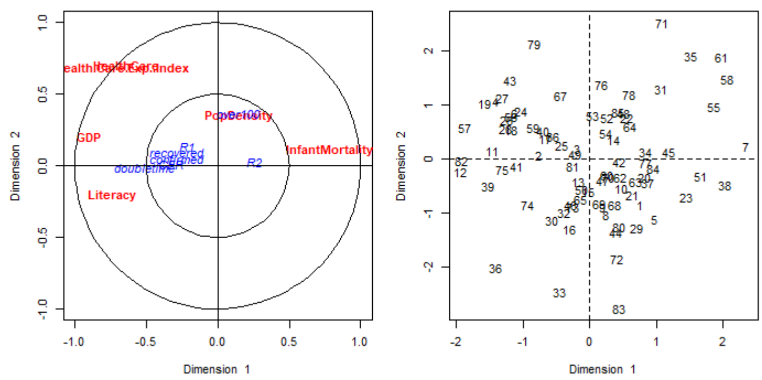


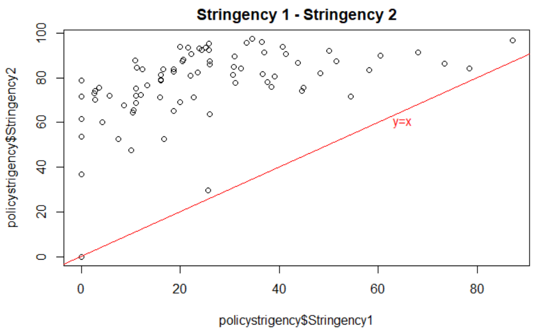
Figure 10: Graphical outputs for CCA of countryData and covidData

Based on Figure 10, we can see that Popdensity and over100 are close to each other, indicating that with high population density, the number of days to 100 confirmed cases is small, which is in accordance with common sense. Moreover, variables in covidData are close to each other.

However, there is no obvious relationship between countryData and covidData. Therefore, developed countries are not better at combating the virus.

covidData & policyData - Do intervention policies matter?

First, let us look at the policyData.




```
  Stringency1      Stringency2   
 Min.   : 0.00   Min.   : 0.00  
 1st Qu.:10.55   1st Qu.:70.14  
 Median :20.64   Median :80.49  
 Mean   :23.88   Mean   :73.14  
 3rd Qu.:34.56   3rd Qu.:87.55  
 Max.   :87.06   Max.   :97.54  
```

Figure 11: Scatter Plot of Stringency1 - Stringency2

Based on Figure 13, we can see clearly that all countries are more stringent than before, though to different degrees. As COVID-19 continues to spread, governments all over the world have come to realize the importance of intervention actions like shut-down and social distances. Let us conduct a Kolmogorov-Smirnov test to see if the distributions are the same.

```
        Two-sample Kolmogorov-Smirnov test

data:  policystrigency$Stringency1 and policystrigency$Stringency2
D = 0.81176, p-value < 2.2e-16
alternative hypothesis: two-sided
```

Since $\text{p-value} < 2.2e - 16,$ we reject the null hypothesis.

Next, let us take a look at the correlation between covidData and policyData.

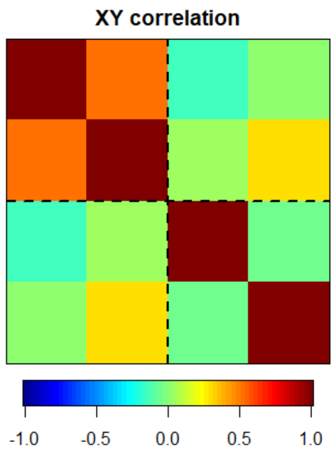


Figure 12: Correlation Matrix of policyData and covidData[R1,R2]

Based on Figure 12, we can see there exists weak correlation between the two sets of variables.

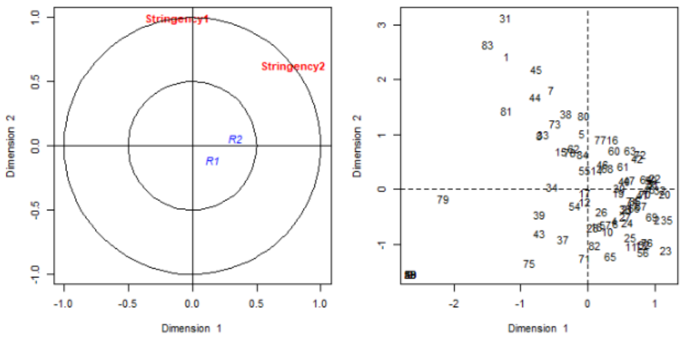


Figure 13: Graphical outputs for CCA of policyData and covidData[R1,R2]

Based on the above results, we cannot say that the stringency of policies matters a lot. Maybe we should divide the period into more time slots and consider the delay of the effectiveness of policies. For example, we can look at the effective reproductive numbers after one policy is launched. Moreover, even with stringent policies, the public may not conform to them. In other words, the stringency index only partly demonstrates the stringent of interventions taken place. Moreover, in countries with serious situations, the stringency index tends to be higher. Demographic and other cultural factors may also play an important role here.

Kernel CCA

Further, we use Kernel CCA to explore the possible nonlinear relationship between policyData and covidData. The resulting correlation coefficients in the feature space are -0.9456146 and 0.9456146, indicating strong associations between the two sets of variables.

Regression Analysis

In this part, we will use linear regression to explore the relationship between effective reproductive coefficients and other variables of countryData, policyData, and covidData. The detailed output can be found in the Appendix.

First, let us run a regression with R1 as the response variable.

We can see that R1 is closely related to over100, which is in accordance with common knowledge. R1 reflects the speed of spread before community transmission when people are not alert. With a high R1, the cumulative number of cases rise rapidly, leading to a small number of days to 100 cases, which is over100 by definition. R1 is also related to HealthCare, indicating that good health care quality can help with containing COVID-19 at the beginning.

Then we treat R2 as the response variable.

We can see that R2 is closely related to doubletime, over100, and Stringency2. For the first two variables, the relationship can be similarly explained as above. For Stringency2, we can conclude that the R2 is closely related to policy. However, the coefficient is positive, which is out of expectation. My interpretation of the result involves two possible explanations. First, just as in the canonical correlation analysis, we assume there exists a nonlinear relationship between them. Second, since it takes time for intervention policies to take effect, Stringency1 may play a more important role than Stringency2. Meanwhile, in countries with serious situations during the second phase, the governments are likely to be more anxious and thus the stringency index tends to be higher. Either way, we can find that intervention policies do matter.

Discussion

In the previous sections, we discuss the clustering of countries in terms of their overall strengths and performances when faced with COVID-19 respectively. We conclude that most countries are doing better than before and the spread of COVID-19 has greatly slowed down. Developed countries are not performing better than other countries, even with better health care facilities and strong economic strength. Moreover, we identify and quantify the effectiveness of intervention policies when combating COVID-19.

The limitations of the study are as follows.

- With detection capacity and other constraints, the data on the number of confirmed cases can only be treated as an approximate estimation of the real infected number. As richer countries have stronger testing ability, they may tend to have a larger number of reported cases.
- For lack of data, we only involve 85 countries in our study. Meanwhile, we cannot go further without data on the number of patients hospitalized or in ICUs.
- In our study, we treat a country as a whole while in reality, infected people tend to appear in clusters within certain regions. For example, Wuhan in China and New York in the USA. Meanwhile, as we treat a country as a whole, we cannot take temperature and humidity into consideration, which are proved to be important. One-degree Celsius increase in temperature and one percent increase in relative humidity lower R by 0.0225 and 0.0158, respectively[^9].
- Since the virus’s incubation period is so long, hospitalization and death figures only give an accurate picture of a few weeks ago[7].
- Though we claim developed countries do not perform better when faced with the virus, we must admit that in developed countries, population mobility tends to be higher and people are apt to contact foreigners, which all contribute to the spread of the virus. We have to admit there are many other underlying factors that play important roles in the spread or containment of the virus. Further researches are needed to identify and quantify them.
- As we have pointed out in the previous section, the stringency index only partly demonstrates the stringent of interventions taken place. The public may not conform to all of them. Moreover, in countries with serious situations, the stringency index tends to be higher.
- We may further look at different types of policies and interventions to see which is more important or effective.

Appendix

Variables Description

For simplicity, we only introduce some variables not familiar to us.

(1) Health Care Index

An estimation of the overall quality of the health care system, health care professionals, equipment, staff, doctors, cost, etc. For more information, see https://www.numbeo.com/health-care/rankings_by_country.jsp

(2) CFR (Case fatality rate)

Adjusted case fatality rate.

$$aCFR = \frac{D(t)}{C(t-T)}$$

where $D(t)$ is the number of cumulative deaths, $C(t-T)$ is the number of cumulative cases at time $t-T$, and T is the average time from case confirmation to death. Here we assume $T=5$. [8]

(3) Doubletime

Days for the confirmed cases to double recently. It implies the speed of the spread and ignores the impact of the base since the population varies from country to country.

(4) Over100

Days to go from the 1^{th} to 100^{th} cases. This period is also treated as the first phase. Since 100 is often treated as a sign of community transmission, it implies the progress has entered a new phase [9]. After the unconstrained spread, governments and the public get aware of the problem and start to take action to fight the virus. The remaining period is treated as the second phase.

(5) R1 (R2)

Effective reproductive number during the first(second) phase. R is the average number of secondary cases per infectious case in a population made up of both susceptible and non-susceptible hosts. If R>1, the number of cases will increase, such as at the start of an epidemic. Note that we use both R and cumulative cases because even if we have a large number of cases and an R of 1 or just below, that still equates to a large number of infections. Using both of the indicators helps us see the big picture. We use package R0 to calculate R1 and R2.

(6) Stringency1 (Stringency2)

Stringency index during the first(second) phase. A group of researchers from Oxford University collects publicly available information on 17 indicators of government responses, involving information on containment and closure policies, economic policies, and health system policies. See OxCGRT [5] for more information. It implies the number and strictness of government policies.

Variable Name	Description
countryData	
Country	Name for the country
PopDensity	Population density per square meter
InfantMortality	Infant Mortality per 1000 births
GDP	GDP per capita ($)
Literacy	Proportion of literate people (%)
HealthCare	Health care index
covidData	
Confirmed	Number of cumulative confirmed cases
Recovered	Number of cumulative recovered cases
CFR	Case fatality rate
Doubletime	Number of days used for confirmed cases to double
Over100	Number of days used for reaching 100 cumulative cases
R1	Effective reproductive number during the first phase
R2	Effective reproductive number during the second phase
policyData	
Stringency1	Mean policy stringency during the first phase
Stringency2	Mean policy stringency during the second phase

Descriptive analysis of data

countryData

```
   PopDensity       InfantMortality        GDP             Literacy          HealthCare
 Min.    :   2.9   Min.    : 2.29   Min.    :  900   Min.    : 40.40   Min.    :39.66
 1st Qu.:   36.2   1st Qu.: 5.53   1st Qu.: 4800   1st Qu.: 86.50   1st Qu.:56.15
 Median :   79.8   Median :14.35   Median :10200   Median : 95.40   Median :64.58
 Mean   : 211.1   Mean    :20.36   Mean    :13492   Mean    : 89.39   Mean    :63.39
 3rd Qu.: 135.7   3rd Qu.:25.95   3rd Qu.:21600   3rd Qu.: 99.00   3rd Qu.:71.58
 Max.    :6482.2   Max.    :98.80   Max.    :37800   Max.    :100.00   Max.    :86.71
```

covidData

```
   confirmed          recovered            CFR             doubletime         over100              R1                  R2
 Min.    :    110   Min.    :     31   Min.    :0.00000   Min.    : 9.00   Min.    :  7.00   Min.    : 0.6178   Min.    :0.0003564
 1st Qu.:   1733   1st Qu.:    660   1st Qu.:0.01692   1st Qu.:16.00   1st Qu.: 13.00   1st Qu.: 2.6205   1st Qu.:0.8817754
 Median :   8964   Median :   2499   Median :0.03544   Median :24.00   Median : 19.00   Median : 4.2041   Median :1.0870250
 Mean   :  45381   Mean    : 14776   Mean    :0.04626   Mean    :25.94   Mean    : 23.91   Mean    : 7.6884   Mean    :1.1014034
 3rd Qu.:  27219   3rd Qu.:   7809   3rd Qu.:0.05976   3rd Qu.:37.00   3rd Qu.: 30.00   3rd Qu.: 9.5541   3rd Qu.:1.3026120
 Max.    :1309550   Max.    :212534   Max.    :0.16626   Max.    :46.00   Max.    :103.00   Max.    :64.1209   Max.    :1.7972880
```

Histogram of log(covidData$confirmed)

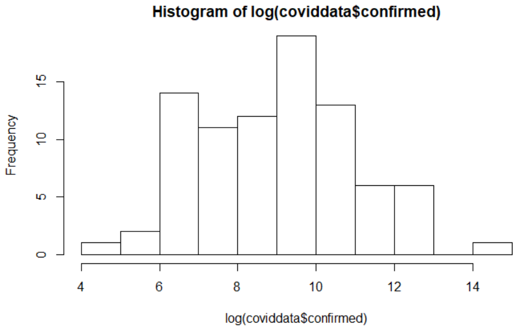


Figure 14: Histogram of log(covidData$confirmed)

We can see that log(Confirmed) tend to be normal. The median of Confirmed is 8,964 while the mean is 45,381, indicating that the variance is huge and half of the countries have more than 9,000 confirmed cases. Though we don't have a benchmark to evaluate the current development of COVID-19 around the world, the number itself is appalling.

Choose k for K-Means Clustering

countryData

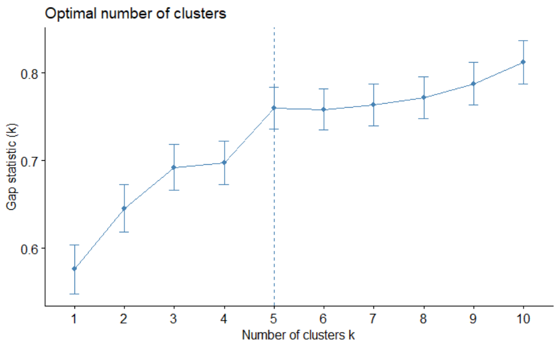


Figure 15: Gap statistic - Optimal number of clusters

Therefore, we will choose the number of clusters as 5.

covidData

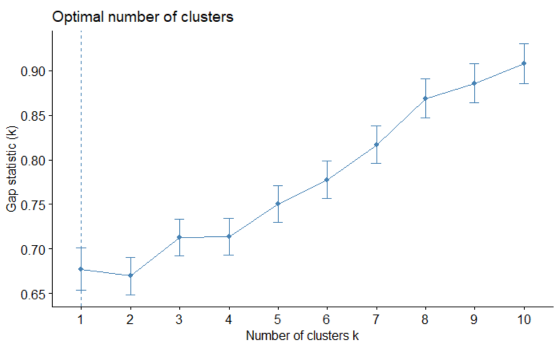


Figure 16: Gap statistic - Optimal number of clusters

Based on Figure 9, the optimal number of clusters should be 1. However, 3 may also be a plausible number. Thus, here we choose the number of clusters to be 3.

Cluster means

countryData

```
Cluster means:
    PopDensity InfantMortality        GDP   Literacy HealthCare
1 -0.19122242     0.001776197 -0.4283984 0.08250765 -0.8480914
2  8.77108926    -0.904381131  0.9966114 0.21462871  0.7218182
3  0.02132131     2.108620724 -1.0916553 -2.07339130 -1.1949479
4 -0.09266817    -0.120819432 -0.3718102 0.21692869  0.3869189
5 -0.06944613    -0.753483351  1.3221802 0.56892562  1.0800032
```

covidData

```
Cluster means:
   confirmed  recovered         CFR doubletime        over100          R1         R2
1 -0.1188627 -0.2166297  0.02511982 -0.7559149  0.007378684 -0.24663926  0.7007823
2  1.5108952  2.2275631  1.40183592  0.6165646  0.005994163  1.50586389 -0.1227727
3 -0.2710125 -0.3361034 -0.45906783  0.9741001 -0.013112232 -0.07357897 -1.0377011
```

Other results

Output for regression1

```
Call:
lm(formula = R1 ~ confirmed + doubletime + over100 + Stringency1 +
    PopDensity + InfantMortality + GDP + Literacy + HealthCare,
    data = regdata)

Residuals:
   Min     1Q Median     3Q    Max
-9.658 -4.547 -2.210  1.753 48.978

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)     -8.771e+00  1.242e+01  -0.706 0.482422
confirmed        3.941e-06  7.415e-06   0.531 0.596684
doubletime       8.738e-02  1.101e-01   0.794 0.429819
over100         -2.537e-01  7.219e-02  -3.514 0.000753 ***
Stringency1     -7.538e-02  6.111e-02  -1.234 0.221203
PopDensity      -6.815e-04  1.443e-03  -0.472 0.638002
InfantMortality  1.241e-01  8.438e-02   1.471 0.145544
GDP              3.470e-05  1.549e-04   0.224 0.823373
Literacy         1.317e-02  1.009e-01   0.131 0.896506
HealthCare       2.816e-01  1.362e-01   2.067 0.042154 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.112 on 75 degrees of freedom
Multiple R-squared:  0.2226,    Adjusted R-squared:  0.1293
F-statistic: 2.386 on 9 and 75 DF,  p-value: 0.01964
```

Output for regression2

```
Call:
lm(formula = R2 ~ confirmed + doubletime + over100 + Stringency1 +
    Stringency2 + PopDensity + InfantMortality + GDP + Literacy +
    HealthCare, data = regdata)

Residuals:
     Min       1Q   Median       3Q      Max
-0.81940 -0.08644  0.00193  0.07489  0.32786

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.373e+00  2.234e-01   6.145 3.69e-08 ***
confirmed        2.060e-07  1.281e-07   1.607 0.112245
doubletime      -2.425e-02  1.920e-03 -12.635  < 2e-16 ***
over100         -5.313e-03  1.277e-03  -4.159 8.50e-05 ***
Stringency1     -2.837e-03  1.210e-03  -2.345 0.021733 *
Stringency2      3.211e-03  8.824e-04   3.639 0.000504 ***
PopDensity       4.168e-05  2.502e-05   1.666 0.099893 .
InfantMortality  1.040e-03  1.456e-03   0.714 0.477256
GDP              1.701e-06  2.677e-06   0.636 0.527017
Literacy         1.379e-03  1.760e-03   0.784 0.435768
HealthCare       2.086e-03  2.379e-03   0.876 0.383604
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1572 on 74 degrees of freedom
Multiple R-squared:  0.7808,    Adjusted R-squared:  0.7511
F-statistic: 26.35 on 10 and 74 DF,  p-value: < 2.2e-16
```

References

1. Park, Minah and Cook, Alex R. and Lim, Jue Tao and Sun, Yinxiaohe and Dickens, Borame L. A Systematic Review of COVID-19 Epidemiology Based on Current Evidence. *J. Clin. Med.* **2020**, *9*(4), 967 ↩ ↩

2. WHO Coronavirus Disease (COVID-19) Dashboard ↩

3. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling studyWu, Joseph T et al. The Lancet, Volume 395, Issue 10225, 689 - 697 ↩

4. Wang, Chaolong & Liu, Li & Hao, Xingjie & Guo, Huan & Wang, Qi & Huang, Jiao & He, Na & Yu, Hongjie & Lin, Xihong & Pan, An & Wei, Sheng & Wu, Tangchun. (2020). Evolving Epidemiology and Impact of Non-pharmaceutical Interventions on the Outbreak of Coronavirus Disease 2019 in Wuhan, China. 10.1101/2020.03.03.20030593. ↩

5. Anna Petherick, Beatriz Kira, Noam Angrist, Thomas Hale, Toby Phillips. 29 April 2020. Variation in government. BLAVATNIK SCHOOL WORKING PAPER ↩ ↩

6. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. Hellewell, JoelSun, Fiona et al. The Lancet Global Health, Volume 8, Issue 4, e488 - e496 ↩

7. Linton, N.M.; Kobayashi, T.; Yang, Y.; Hayashi, K.; Akhmetzhanov, A.R.; Jung, S.-M.; Yuan, B.; Kinoshita, R.; Nishiura, H. Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J. Clin. Med.* **2020**, *9*, 538. ↩

8. Jing Yuan, Minghui Li, Gang Lv, Z. Kevin Lu, Monitoring transmissibility and mortality of COVID-19 in Europe, International Journal of Infectious Diseases, Volume 95, 2020, Pages 311-315, ISSN 1201-9712 ↩

9. The Lancet. "Singapore modelling study estimates impact of physical distancing on reducing spread of COVID-19." ScienceDaily. ScienceDaily, 24 March 2020. ↩