Table of Contents

Abstract

The report is to interpreter the results of the analysis using Covid19 related indicators to find the relationship between various indicators as well as the difference and similarity between countries. The selection of variables is mainly based on Covid19 related research and data of Our World in Data and WHO (World Health Organization) Coronavirus Dashboard. The entity (country) selection criteria are to include countries with population greater than 5 million and in which at least 21 days had passed since the 100th confirmed case. (Exemplars in Global Health (EGH),2020). There are mainly two different multivariate techniques has been used in this report: Supervised methods and unsupervised methods. The former one is used to find the relationship between indicators and association between countries. The latter one is utilized to recognize counties which government underreacted Covid19 according to their situation.

1.0 Introduction

Since the outbreak of the 2019 novel coronavirus disease (COVID-19) in Wuhan, China, has spread rapidly worldwide (Lei S, 2020), every country in the world has been affected by the worldwide epidemic. Given different countries has various situation of economy, political, culture and so on, government and public of different countries have different attitudes and responses to Covid19. Therefore, the difference of situation of dealing with the virus between different countries is huge. To show countries' difference of Covid19 situation there are so many research using various indicators. The purpose of this project is to gain a better understanding of the meaning of those common indicators and the relationship between them by using multivariate statistical techniques.

The data used in the report is made by 19 variables dividing to 5 aspects and 114 countries. The data sources section gives their detail information.

The unsupervised method includes principal component analysis, factor analysis and cluster analysis. The former two are both to reduce dimensionality of variables and find the similarity and difference between countries. The cluster analysis finds the similar situation between countries according to those indicators.

The supervised method is comprised of discriminant analysis. All countries' government are classified into loose, moderate and strict according to the stringency index from Our World in Data. Discriminant analysis evaluates which countries' government overreacted and which countries' government underreacted in terms of some other indicators.

The last part of report is to summarize some valuable findings and make a conclusion as well as provide the limitation of the analysis.

2.0 Data Source

The data used for this analysis was from various sources. The main source was from Our World in Data (https://ourworldindata.org/grapher/covid-stringency-index) and WHO (https://www.who.int/data/gho/data/themes/health-workforce). The data has 4 parts: Covid19 Cases, Health Workforce, Government Response and General information. Health Workforce mainly was mainly gathered from WHO and rest of them were mainly collected from Our World in Data.

The following Table 1 shows the detail of variables used in the analysis. The original dataset has been manipulated for the analysis. The date of the variables collection was pointed out if the variables changed over time.

Table 1: Summary of Variables

Group	Variables	Description	Source and Year
Covid19 Cases	Total_cases	Confirmed cases in the country	Our World in Data 2021.6.8 (if not available, latest data will be used)
Covid19 Cases	Total_deaths	Confirmed Covid19 deaths in the country	Our World in Data 2021.6.8 (if not available, latest data will be used)
Covid19 Cases	New_cases_per_day	Confirmed new cases in one day in the country compared with last available day	Our World in Data 2021.6.8 (if not available, latest data will be used)
Covid19 Cases	Total_tests	The number of Covid19 tests in the country	Our World in Data 2021.6.8 (if not available, latest data will be used)
Health Workforce	Doctors	The number of medical doctors in the country	WHO latest available year in database
Health Workforce	Nurses	The number of nurses in the country	WHO latest available year in database
Health Workforce	Hosp_bed_per10000	The number of available hospital beds per 10000 people in the country	WHO latest available year in database
Government Response	Workplace_closures	The situation of workplace closures in the country	Our World in Data 2021.5.30 (if not available, latest

			data will be used)
Government Response	Facial_coverings	The situation of Facial_coverings in the country	Our World in Data 2021.5.30 (if not available, latest data will be used)
Government Response	School_closures	The situation of School_closures in the country	Our World in Data 2021.5.30 (if not available, latest data will be used)
Government Response	Public_transport_closures	The situation of Public_transport_closures in the country	Our World in Data 2021.5.30 (if not available, latest data will be used)
Government Response	International_travel_controls	The situation of International_travel_controls in the country	Our World in Data 2021.5.30 (if not available, latest data will be used)
Government Response	Government_response	Category grouping government strict actions for Covid19	Our World in Data 2021.6.8 (if not available, latest data will be used)
General information	Population	The number of the people in the country	Our World in Data
General information	Population_density	The number of the people per square mile in the country	Our World in Data
General information	GDP_per_capita	The number of the population in the country	Our World in Data
General information	Human_development_index	Average achievement score in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard living	Our World in Data
General information	Aged_65_older	Share of the population that is 65 years and older	Our World in Data
ID	Country	The name of the country	

3.0 Methodology

The variables' selection takes account of the availability of the variables and the frequency of using them in research. Because variables from Covid19 Cases group start with "total" are cumulative cases, 2021.6.8 data can be used directly. If it is unavailable, the latest available data can be used. However, average value was used in the variables from Government Response group and new_cases_per_day because average value is more representative than the value of a latest day for them.

The entities' selection leant from the article in Global Health (EGH) platform. There were two criteria applied in the analysis. The first one is to include only countries with population greater than 5 million. The second one is to include only countries in which at least 21 days had passed since the 100^{th} confirmed case.

Although the two criteria selected countries with good quality data, 4 countries that were quite unique about Covid19 were removed by the criteria. Therefore, they were added to the dataset in order to see their effect on analysis. They are Uruguay, Ireland, Iceland and New Zealand.

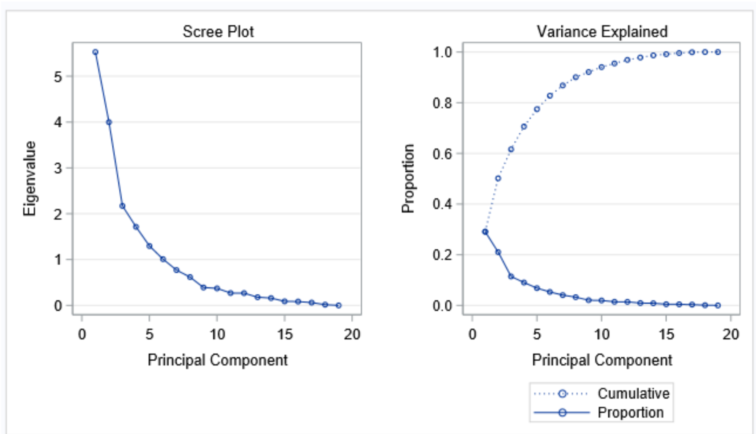
There were some log transformations for variables with skewness and outliers when it is necessary for certain analysis.

4.0 Unsupervised Methods Analysis

4.1 Principal Components Analysis

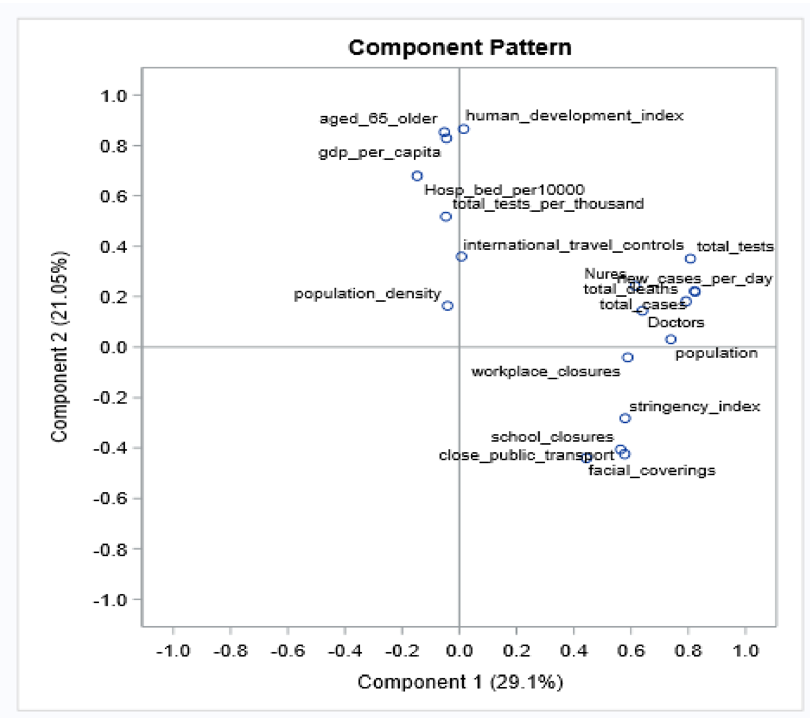
As the starter of the analysis, PCA (Principal Components Analysis) has two main purposes in the report. Given there are so many variables in the dataset, the first one is to reduce the dimensionality of the dataset and do 2D visualization. The second one is to get a basic understanding of the indicators and countries.

Figure 1: Scree Plot of PCA



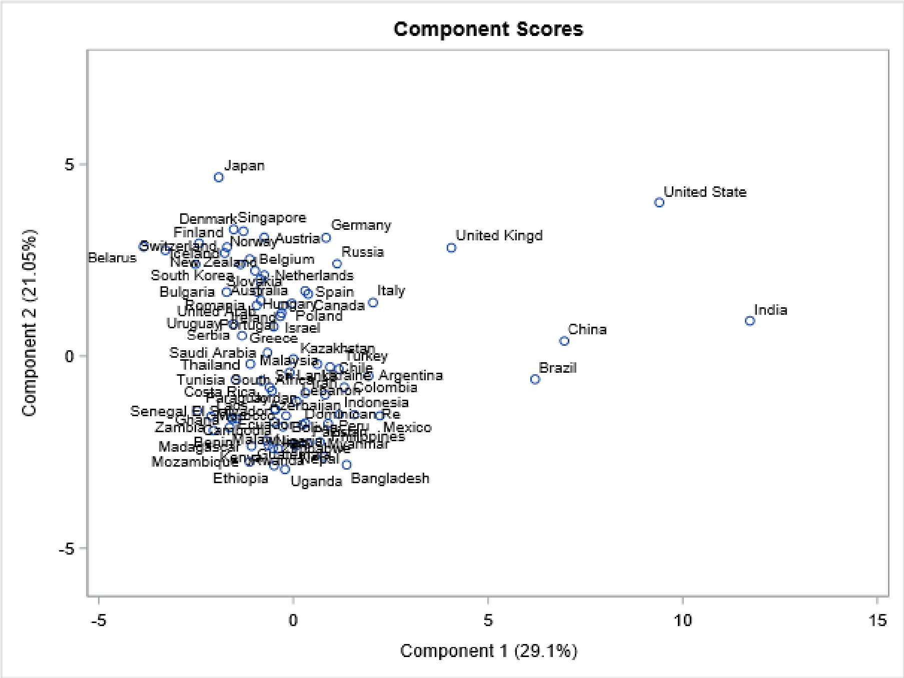
The following Figure 1 is the screen plot of PCA. There is an elbow in the third PC and the first three PCs explain over 60% variation from the original data. It is necessary to only use the first three PCs to represent original variables.

Figure 2: Component Pattern of PCA



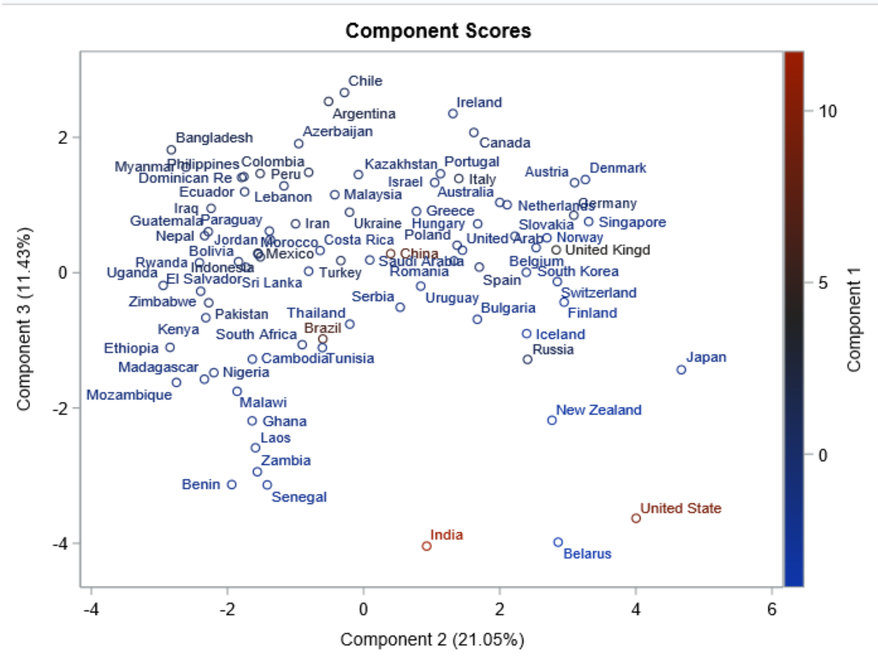
The Component Pattern (Figure 2) can illustrate the basic association between variables. As expectation, population is associated with total_cases, deaths, Nurses and Doctors which means countries with high population tend to get more Covid19 cases, deaths as the same time they also have more doctors and nurses. On the component 2, human _development_index, aged_65_older, gdp_per_capita and Hosp_bed_per10000 are linked with each other and all get high positive weightings. School_closures, close_public_transport and facial_coverings are highly associated with each other and they are negatively associated with population_density.

Figure 3: Component Scores of PCA (2 PCs)



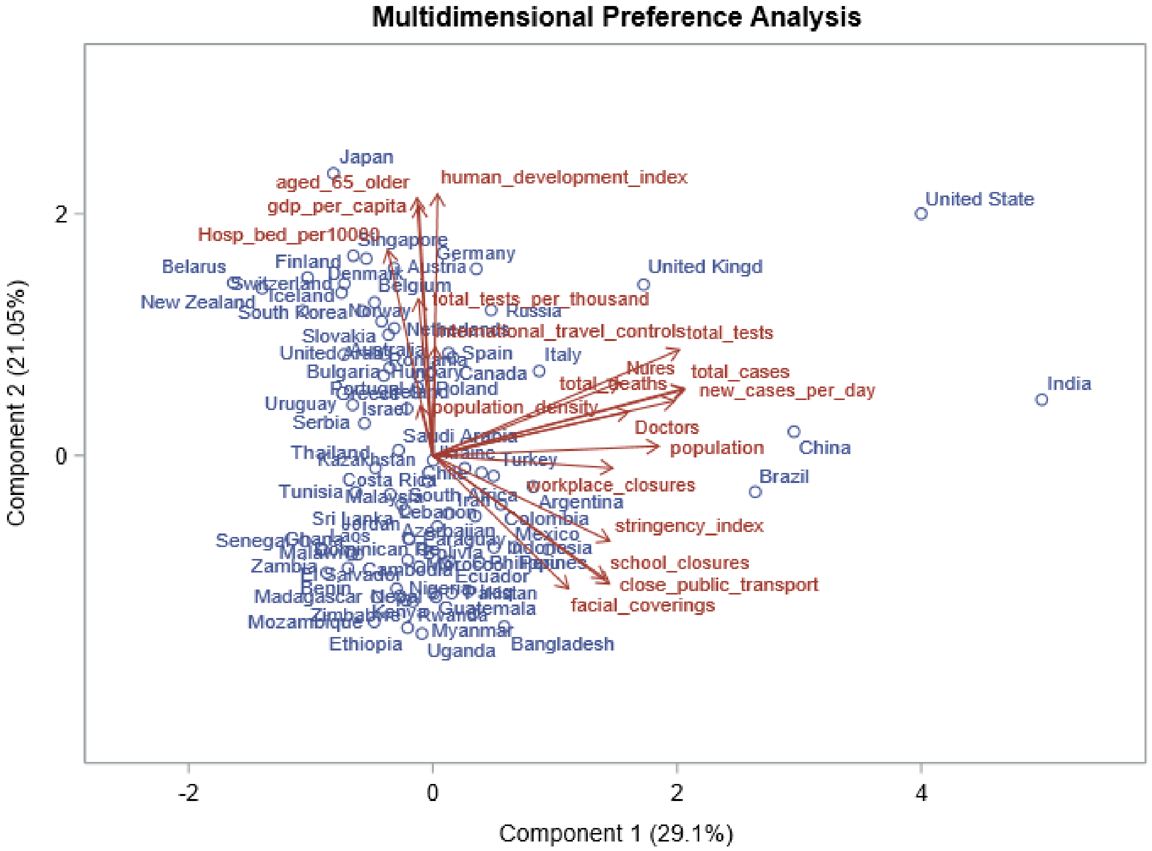
From the Component Scores (Figure 3) with PC 1 and 2, it is clear to see that China, India, Brazil, United State, United Kind and Japan are separated with a big cluster. In the big cluster, it can be divided into two parts: component 2 weighting is over 0 and below 0. Almost developed countries are in the top of PC 2 with positive weighting of PC 2 and developing countries are in the bottle of PC 2 with negative weighting of PC 2.

Figure 4: Component Scores of PCA (3 PCs)



Three PCs (Figure 4) has been visualized in the figure, Japan, New Zealand, United State, Belarus and India are in the right bottle and United State and India have red colour. Most of European countries are in the right top. Asian and American countries are mixed in the left top. Left bottle points mostly are African countries.

Figure 5: Biplot of PCA



The biplot (Figure 5) show an informative plot including variables and countries. Some developing countries with negative PC 2 and positive PC 1 weightings has high government strict actions for Covid19 like closing school, workplace, public transport and wearing facial coverings. Most developed countries with positive PC 2 weighting have high human_development_index, aged_65_older, gdp_per_capital and Hosp_bed_per10000. However, some developed countries with negative PC 1 weighting have very little government strict actions for Covid19 like New Zealand, Finland, South Korea, Belarus, Iceland and so on. The several outliers are certainly have high population, population_density, total_cases, total_deaths, new_cases_per_day,total_cases, Nures and Doctors.

4.2 Factor Analysis

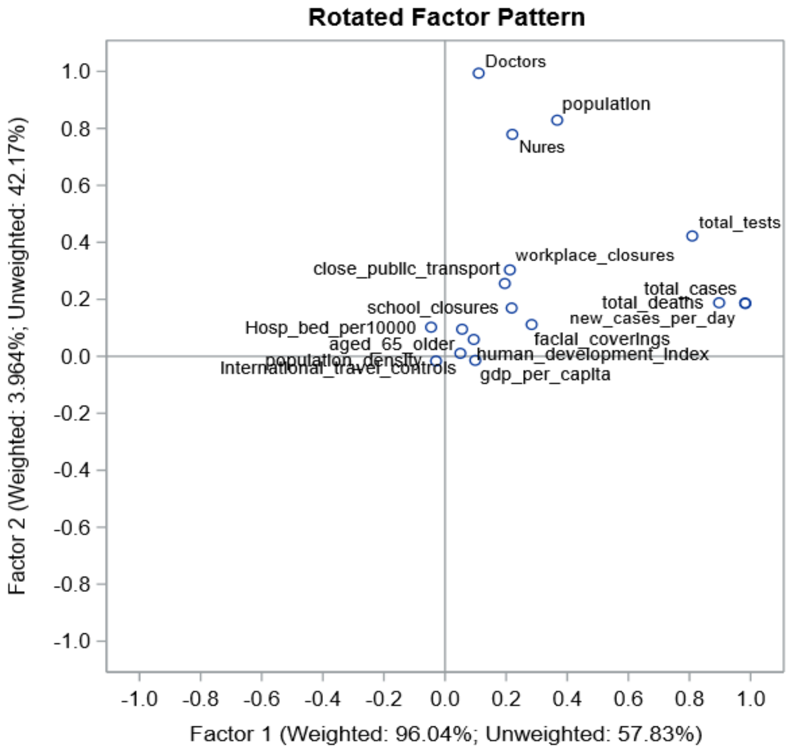
Factor analysis is a statistical method used to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors. (Wikipedia,2021)

Table 2: Rotated Factor Pattern

Rotated Factor Pattern				
	Factor1		Factor2	
total_cases	98	*	.	
new_cases_per_day	98	*	.	
total_deaths	90	*	.	
total_tests	81	*	42	*
Doctors	.		99	*
Nures	22		78	*
Hosp_bed_per10000	.		.	
workplace_closures	21		30	*
facial_coverings	28		.	
school_closures	22		.	
close_public_transport	.		26	
international_travel_controls	.		.	
population	37	*	83	*
population_density	.		.	
gdp_per_capita	.		.	
human_development_index	.		.	
aged_65_older	.		.	
Printed values are multiplied by 100 and rounded to the nearest integer. Absolute values greater than 0.3 are flagged with an '*'. Absolute values less than 0.2 are not printed.				

The left table 2 shows the loading of each variable in two factors. Only variables with loading over 0.3 are displayed with loading value.

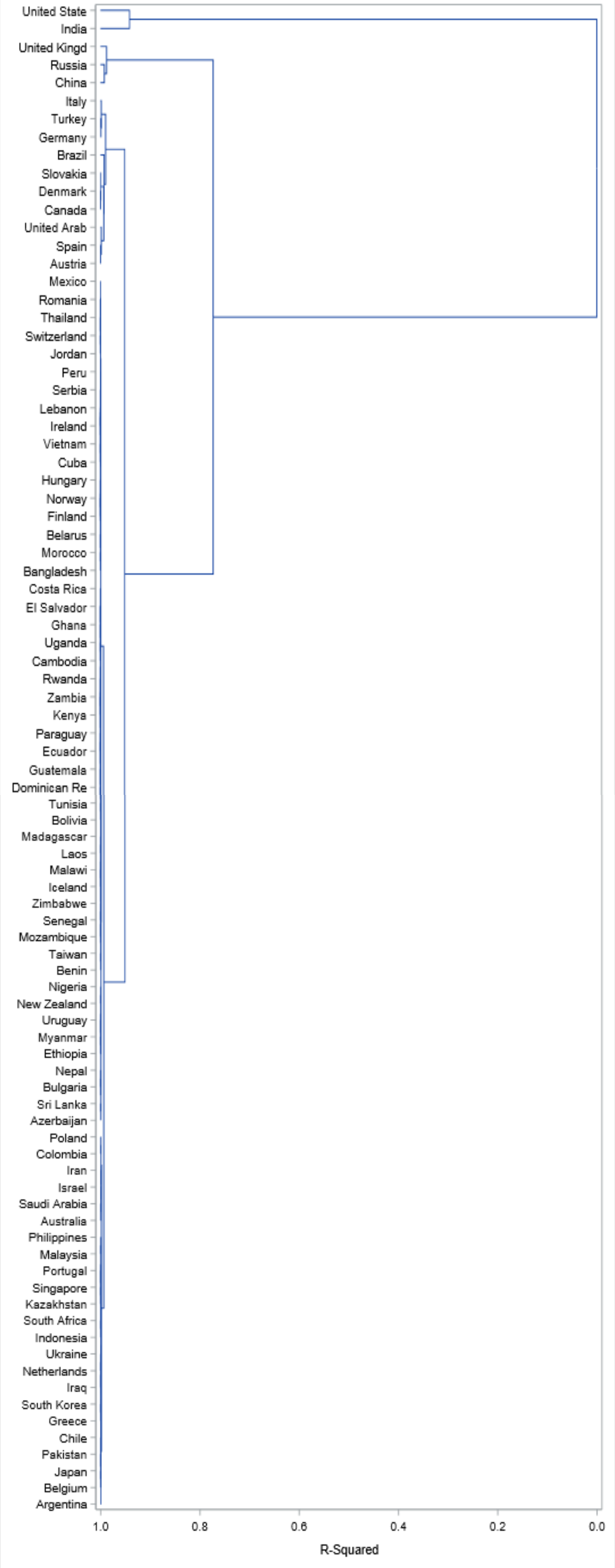
Figure 6: Rotated Factor Pattern



As we can see from the rotated factor pattern in figure 6, total_cases, total_deaths, new_cases_per_day are plotted together again and they all have high positive factor 1 loading. Doctors, Nures and population are plotted closely and with high positive high factor 2 loading. However, the factor analysis did not distinguish those variables well as the PCA did. For example, aged_ 65_older, human_development_index, gdp_per_capita, population_density, international_travel_controls and Hosp_bed_per10000 are plotted in the centre of the graph and are not explained very well by the two factors.

4.3 Cluster Analysis

Figure 7: Dendrogram of Cluster Analysis



A cluster analysis using variables from Covid19 Cases group and Government response group was used to find which countries are the most similar according to their current situation and response of dealing with Covid19.

The method of the cluster analysis is hierarchical agglomerative algorithm with average linkage. The dendrogram is on the left.

Unfortunately, it is hard to find the most similar countries in the figure 7 becaese of the overlaps. However, the plot still provide some valueable information. The United State and India were in a cluster and were distinguished from other countries because their both have very high confirmed Covid19 cases and deaths as well as new cases per day.

Russia and China are in a cluster and then they are in another cluster with United Kingdom.

As we expected New Zealand, Iceland and Taiwan are located in a big cluster with other countries.

4.4 Multi-Dimensional Scaling

Given there were many overlaps in the plot of cluster analysis, MDS (Multi-Dimensional Scaling) has been used to visualize the clusters.

The figure 8 is based on the Euclidean distance between countries by using the variables from Covid19 Cases group. It clearly shows that there is a cluster around the red dot and a horizontal “U” shape.

Figure 8: MDS for Covid19 Cases group variables

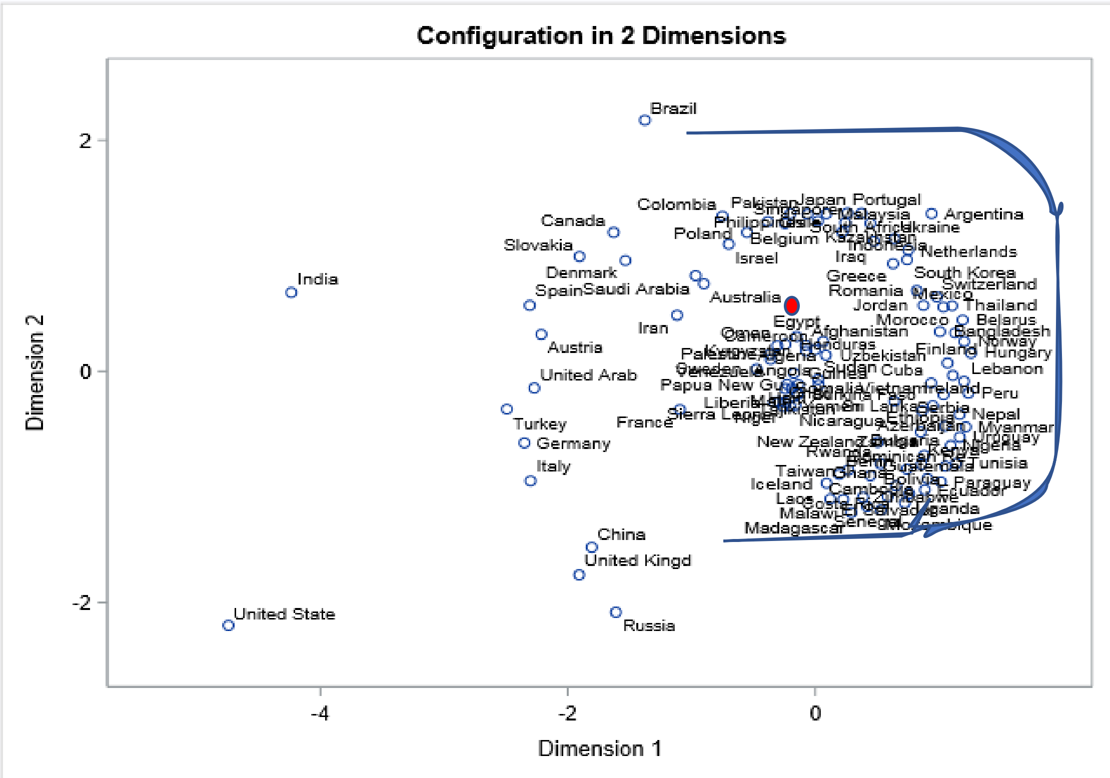
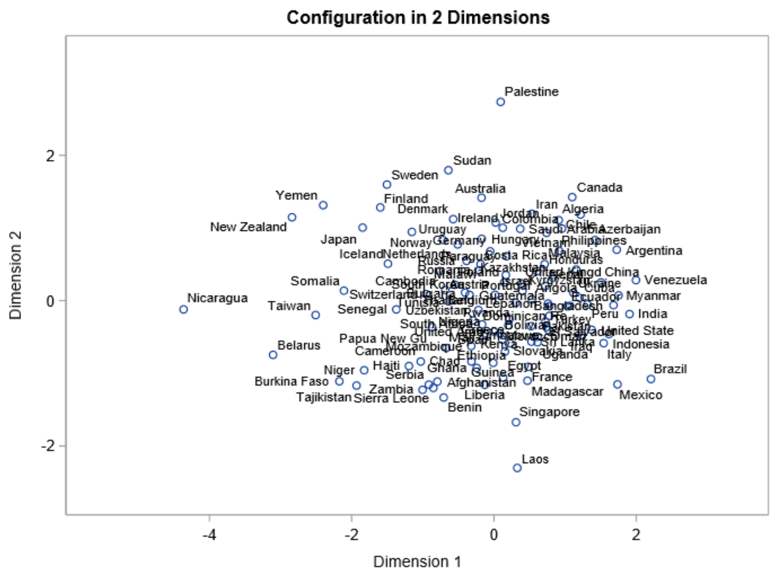


Figure 9: MDS for Government Response variables



The figure 9 on the left is based on the Euclidean distance between countries by using the variables from Government Response group.

There is no clear pattern and most countries randomly gather in a big and loose cluster. A few countries are outlier from the cluster like New Zealand, Palestine, Yemen, Nicaragua, Laos, Singapore, Japan and so on.

5.0 Supervised Method Analysis

5.1 Canonical Correlation Analysis

Canonical Correlation Analysis was used to find the relationship between a group of variables and another group of variables.

5.1.1 Covid19 Cases and Government Response

Figure 10: Boxplot of total_cases

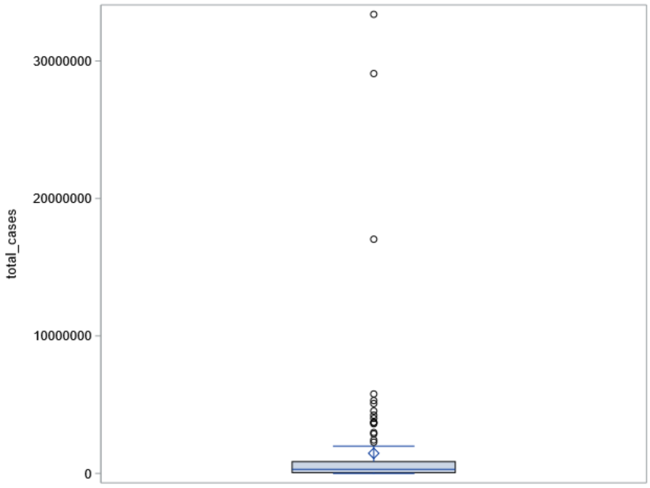


Figure 10: Boxplot of total_deaths

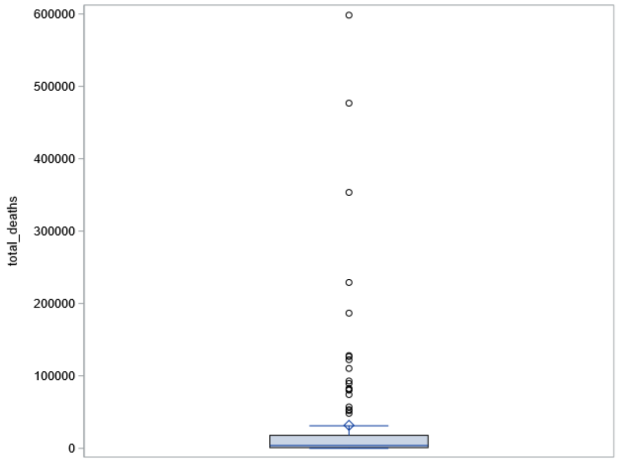


Figure 11: Box of total_tests

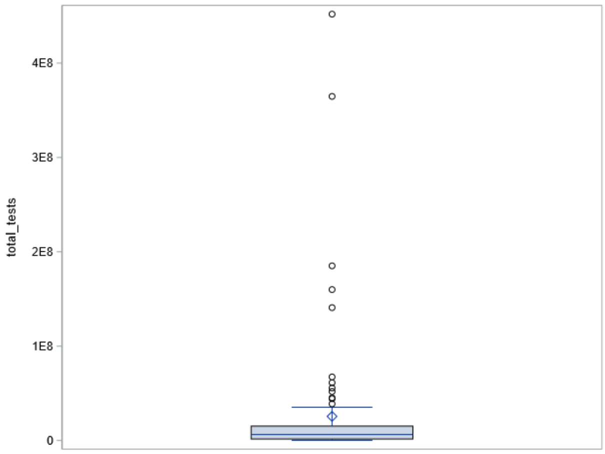
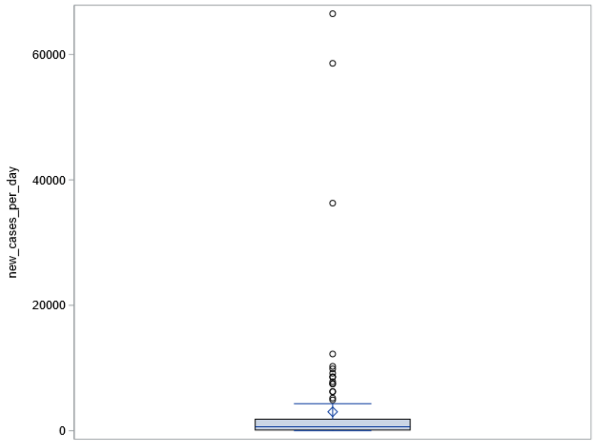


Figure 12: Box of new_cases_per_day



As we can see from the distribution of four variables from Covid19 Cases group (Figure 9 -12), they all have high skewness and some outliers, which may lead to insignificant statistical estimates. Therefore, they were taken log transformation to improve that.

Table 3: Canonical Correlation Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation
1	0.565345	0.483464	0.073368	0.319615
2	0.493473	0.480126	0.081574	0.243515
3	0.303650	0.266952	0.097890	0.092204
4	0.158267	0.157961	0.105132	0.025048

There are four correlations because the minmum number of variables is 4 between the two sets of varables.

Test of H0: The canonical correlations in the current row and all that follow are zero				
Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
0.45554025	3.47	20	259.65	<.0001
0.66953258	2.86	12	209.31	0.0012
0.88505750	1.68	6	160	0.1294
0.97495150	1.04	2	81	0.3579

According to the likelihood ratio, only the first two correlations are statistically different 0. Therefore, we mainly focus on the first two correlation.

Table 4: Canonical Structure

Correlations Between the VAR Variables and Their Canonical Variables		
	Cases1	Cases2
log_total_cases	0.9582	0.2211
log_total_deaths	0.9884	0.0492
log_total_tests	0.5864	0.7987
log_new_cases_per_day	0.9625	0.2091

Correlations Between the WITH Variables and Their Canonical Variables		
	Gov1	Gov2
workplace_closures	0.8390	0.4813
facial_coverings	0.2653	0.0052
school_closures	0.9518	-0.1687
close_public_transport	0.6663	-0.0372
international_travel_controls	-0.1906	0.5910

The canonical variate Cases1 seems to be highly correlated with log_total_cases, log_total_deaths and log_new_cases_per_day. It could refer to the risk of infecting Covid19 in the country.
The canonical variate Cases2 seems to be highly correlated with log_total_tests. Therefore, it could be the 1country's capability of testing people whether they have infected Covid19
The canonical variate Gov1 seems to be highly correlated with school_closures and workplace_closures.
The canonical variate Gov2 seems to be highly correlated with internation_travel_controls and workplace_closures.

Table 5: Canonical Structure

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables		
	Gov1	Gov2
log_total_cases	0.5417	0.1091
log_total_deaths	0.5588	0.0243
log_total_tests	0.3315	0.3941
log_new_cases_per_day	0.5441	0.1032

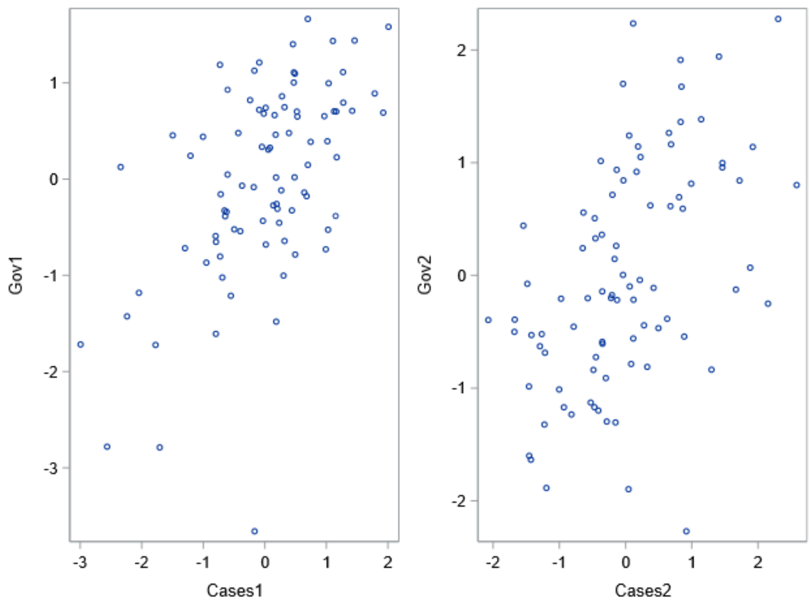
Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables		
	Cases1	Cases2
workplace_closures	0.4743	0.2375
facial_coverings	0.1500	0.0026
school_closures	0.5381	-0.0833
close_public_transport	0.3767	-0.0184
international_travel_controls	-0.1078	0.2916

The three variables that were strongly correlated with Cases1 are also highly correlated with Gov1. The two variables that were strongly correlated with Gov1 are also highly correlated with Cases1. Therefore, Comfirmed cases, deaths and new cases per day has a positive correlation (0.565345) with closing workplace and school.

However, there is an unexpected signal in the variable international_travel_controls. It shows that international_travel_control has a weak negative relationship with canonical variate Cases1, which means that even the situation of covid19 has been worse, the international travel seems not be affected by that.

Log_total_tests that was highly correlated with Cases2 now is also weakly correlated with Gov2. International_travel_controls and workplace_closures now are also weakly correlated with Cases2. Therefore, the countries with high capability of testing people have a weak intention to require closing workplace and controlling international travel.

Figure 13: Scatter plot of canonical variates



The figure 13 shows the two correlations between the two canonical variates.

5.1.2 Health workforce and Government Response

Given two variables Doctors and Nurses both have high skewness, log transformation has been taken for them.

Table 6: Canonical Correlation Analysis

Correlations Between the VAR Variables and Their Canonical Variables		
	Health1	Health2
log_Doctors	0.9303	0.3667
log_Nures	0.7897	0.4739
Hosp_bed_per10000	0.1033	0.9694

Correlations Between the WITH Variables and Their Canonical Variables		
	Gov1	Gov2
workplace_closures	0.8767	-0.0502
facial_coverings	0.2812	-0.5825
school_closures	0.8059	-0.4058
close_public_transport	0.7067	-0.4872
international_travel_controls	0.4653	0.6107

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables		
	Gov1	Gov2
log_Doctors	0.5352	0.1404
log_Nures	0.4543	0.1814
Hosp_bed_per10000	0.0594	0.3711

Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables		
	Health1	Health2
workplace_closures	0.5044	-0.0192
facial_coverings	0.1618	-0.2230
school_closures	0.4636	-0.1553
close_public_transport	0.4066	-0.1865
international_travel_controls	0.2677	0.2338

Canonical analysis output in figure indicates that both log_Doctors and log_Nures have highly positive correlation with the canonical variables of Gov1 indicators. Workplace_closures is the most positive correlated with the canonical variate Health1. Therefore, log_Doctors and log_Nures has a positive correlation with workplace_closures.

Hosp_bed_per10000 is positively correlated with the canonical variate Gov2. Facial_coverings has a weakly negative correlation with canonical variate Health2 and international_travel_controls has a weakly positive correlation with canonical variate Health2.

The figure 14 shows the two correlations between the two canonical variates.

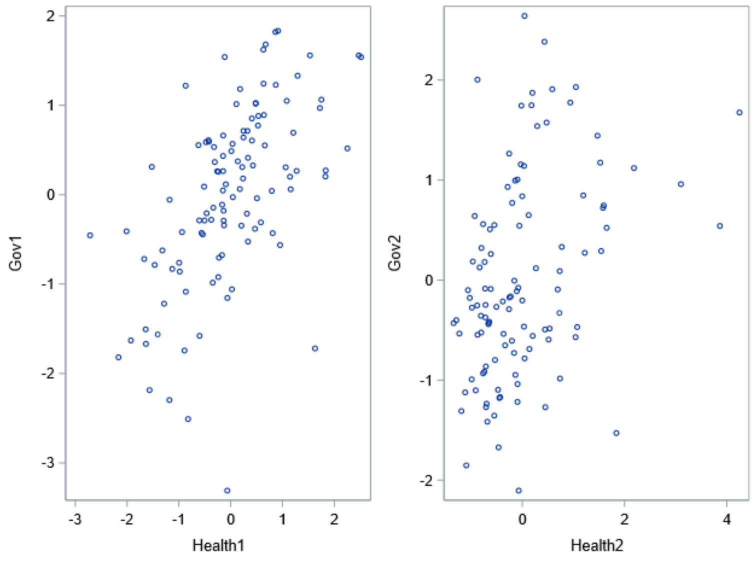


Figure 14 Scatter plot of canonical variates

5.1.3 General Information and Government Response

Given population has high skewness, log transformation has been taken for it.

Table 7: Canonical Correlation Analysis

Correlations Between the VAR Variables and Their Canonical Variables		
	GI1	GI2
log_population	0.5744	0.3088
population_density	-0.1768	0.1504
aged_65_older	-0.6744	0.6232
gdp_per_capita	-0.6324	0.6539
human_development_index	-0.4132	0.8823

Correlations Between the WITH Variables and Their Canonical Variables		
	Gov1	Gov2
workplace_closures	0.2160	0.8192
facial_coverings	0.5455	0.0105
school_closures	0.6747	0.5920
close_public_transport	0.7559	0.3745
international_travel_controls	-0.1057	0.6642

Correlations Between the VAR Variables and the Canonical Variables of the WITH Variables		
	Gov1	Gov2
log_population	0.3834	0.1839
population_density	-0.1180	0.0896
aged_65_older	-0.4501	0.3711
gdp_per_capita	-0.4221	0.3894
human_development_index	-0.2758	0.5254

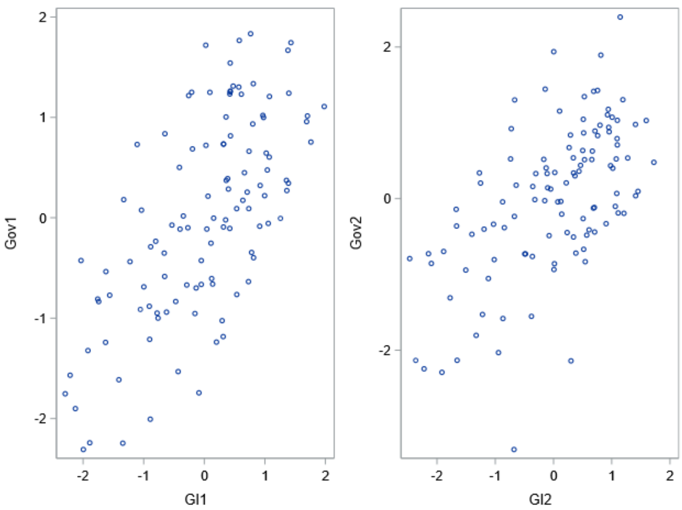
Correlations Between the WITH Variables and the Canonical Variables of the VAR Variables		
	GI1	GI2
workplace_closures	0.1442	0.4878
facial_coverings	0.3641	0.0062
school_closures	0.4503	0.3525
close_public_transport	0.5045	0.2230
international_travel_controls	-0.0705	0.3955

Canonical analysis output in figure indicates that both age_65older and gdp_per_capita have highly negative correlation with the canonical variables of Gov1 indicators. Close_public_transport is the most positively correlated with the canonical variate GI1. It indicates that countries with more older people and high GDP per capita tend not to close public transport during Covid19 period.

Huamn_development_index is positively correlated with the canonical variate Gov2. Workplace_closures has a positive correlation with canonical variate GI2. It indicates that countries with high human development index tend to close workplace during Covid19 period.

The figure 15 shows the two correlations between the two canonical variates.

Figure 15 Scatter plot of canonical variates



5.2 Discriminant Analysis

The discriminant analysis is mainly used non-Government Response group variables to classify countries' government into loose and strict government response. The assumption of the analysis is that different governments made different strategies to deal with the Covid19 based on their economy, culture, policy and so on. Therefore, the result of the discriminant analysis would indicate which countries' government overreacted to Covid19 and which countries' government underreacted to Covid19 in terms of those available indicators in the analysis.

It's important to note that this categorical variable simply records the strictness of government policies. It does not measure or imply the appropriateness or effectiveness of a country's response. Strict group members do not necessarily have 'better' government response than others from loose group. (Thomas Hale, 2021)

Table 8: Output of CANDISC

Total Sample Size	84	DF Total	83
Variables	12	DF Within Classes	82
Classes	2	DF Between Classes	1

Number of Observations Read	114
Number of Observations Used	84

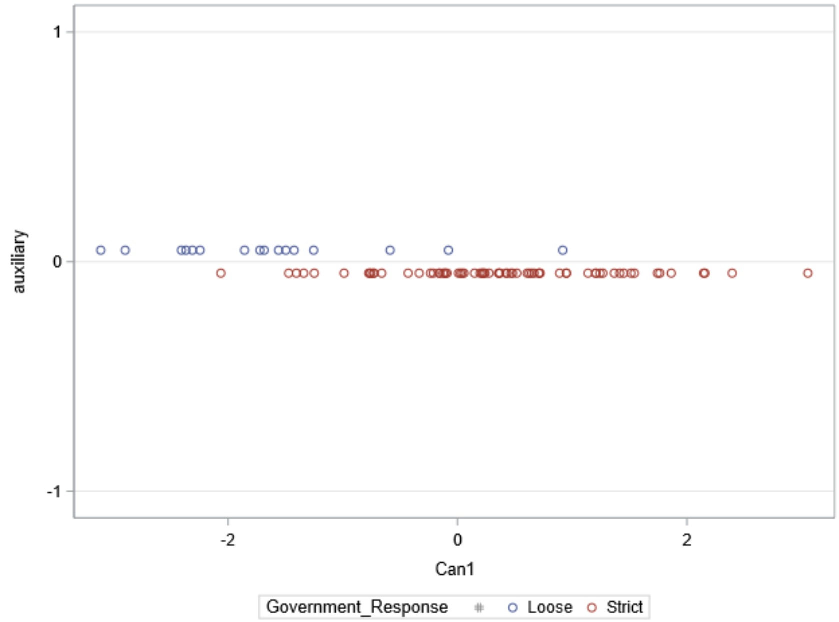
Class Level Information				
Government_Response	Variable Name	Frequency	Weight	Proportion
Loose	Loose	16	16.0000	0.190476
Strict	Strict	68	68.0000	0.809524

The partial output from CANDISC procedure is shown left table 8. There are 84 countries and 12 variables to classify into 2 groups in the discriminant analysis.

19.0476% countries' government are from loose group and 80.9524% are from strict group

Because there are only two groups, there is one discriminant function. To avoid the overlaps, the auxiliary axis has been created to separate loose and strict groups. However, the auxiliary axis has no mathematical meaning in the discriminant analysis.

Figure 16 Scatter plot of discriminant analysis



Given there are 12 variables used in above discriminant analysis, a stepwise discriminant analysis (Table 9) has been used to reduce the number of variables.

Only two variables meet the criteria that p-value is smaller than 0.05 and they were selected to continue to next analysis.

Table 9: Stepwise Selection

Stepwise Selection Summary										
Step	Number In	Entered	Removed	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	1	log_new_cases_per_day		0.2033	20.92	<.0001	0.79674760	<.0001	0.20325240	<.0001
2	2	Hosp_bed_per10000		0.0707	6.16	0.0151	0.74042234	<.0001	0.25957766	<.0001

The table 10 shows the result of the test of homogeneity of within covariance matrices. Given the Chi-Square value is significant at 0.05 level, the null hypothesis that covariance matrices are homogeneous is rejected. Therefore, a quadratic discriminant analysis was utilized.

Table 10: Homogeneity Test

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
9.721080	3	0.0211

The overall misclassification rate is 18.18% (20 out of 110 countries). 14 errors occurred in the loose group. However, the proportion of the loose group is only 27.27%. The misclassification rate of strict is only 7.5%.

It is worthy pointing out that the estimates are quite optimistic because the same countries that were classified were also used to determine the discriminant function.

Number of Observations and Percent Classified into Government_Response

From Government_Response	Loose	Strict	Total
Loose	16 53.33	14 46.67	30 100.00
Strict	6 7.50	74 92.50	80 100.00
Total	22 20.00	88 80.00	110 100.00
Priors	0.27273	0.72727	

Table 11: Error Count Estimates

Error Count Estimates for Government_Response

	Loose	Strict	Total
Rate	0.4667	0.0750	0.1818
Priors	0.2727	0.7273	

Table 12: Misclassified countries list

Posterior Probability of Membership in Government_Response

country	From Government_Response	Classified into Government_Response		Loose	Strict
Afghanistan	Loose	Strict	*	0.2366	0.7634
Australia	Strict	Loose	*	0.5871	0.4129
Bulgaria	Loose	Strict	*	0.4830	0.5170
Cambodia	Loose	Strict	*	0.4953	0.5047
Cameroon	Loose	Strict	*	0.2689	0.7311
Chad	Strict	Loose	*	0.8605	0.1395
Finland	Loose	Strict	*	0.3648	0.6352
Ghana	Loose	Strict	*	0.2099	0.7901
Guinea	Strict	Loose	*	0.5756	0.4244
Liberia	Strict	Loose	*	0.9085	0.0915
Malawi	Loose	Strict	*	0.4675	0.5325
Russia	Loose	Strict	*	0.0892	0.9108
Rwanda	Strict	Loose	*	0.5424	0.4576
Senegal	Loose	Strict	*	0.4122	0.5878
South Korea	Strict	Loose	*	0.9486	0.0514
Sudan	Loose	Strict	*	0.4557	0.5443
Switzerland	Loose	Strict	*	0.1044	0.8956
Thailand	Loose	Strict	*	0.1578	0.8422
Uzbekistan	Loose	Strict	*	0.3518	0.6482
Zambia	Loose	Strict	*	0.2389	0.7611

The table 12 left is the list of countries misclassified. It shows that Australia, Chad, Guinea, Liberia, Rwanda and South Korea overreacted the Covid19 according the discriminant model. Finland, Russia, Switzerland, Thailand and some countries underreacted the Covid19 according the model.

5.3 Partial Least Square Analysis

Given discriminant analysis estimate which countries' government overreacted or underreacted Covid19, it is worth continuing to see the detail of government response. Therefore, partial least square analysis was used to find how variables from Government Response group are predicted by rest of other variables.

The regression structure is that the variables workplace_closures, school_closures, close_public_transport and international_travel_controls are response variables. Rest of other continuous variables are predictor variables. The variables total_cases, total_deaths, total_tests, new_cases_per_day, population, doctors and nurses have taken a log transformation.

The program performs a partial least squares analysis using the NIPALS algorithm and leave one out cross validation. The following table 13 shows that details of the fitted model. There are 5 response variables and 12 predictor parameters. 84 out of 114 observations were used in the analysis.

Table 13: Basic information of PLS

Data Set	WORK.COVID19
Factor Extraction Method	Partial Least Squares
PLS Algorithm	NIPALS
Number of Response Variables	5
Number of Predictor Parameters	12
Missing Value Handling	Exclude
Maximum Number of Factors	12
Validation Method	Leave-one-out Cross Validation
Validation Testing Criterion	Prob T**2 > 0.1
Number of Random Permutations	1000
Random Permutation Seed	608789001

Number of Observations Read	114
Number of Observations Used	84

The minimum PRESS occurs at two factors same as the result of CVTEST. This is a good reduction from the original 12 predictor variables.

Table 14: Selection of Factors' number

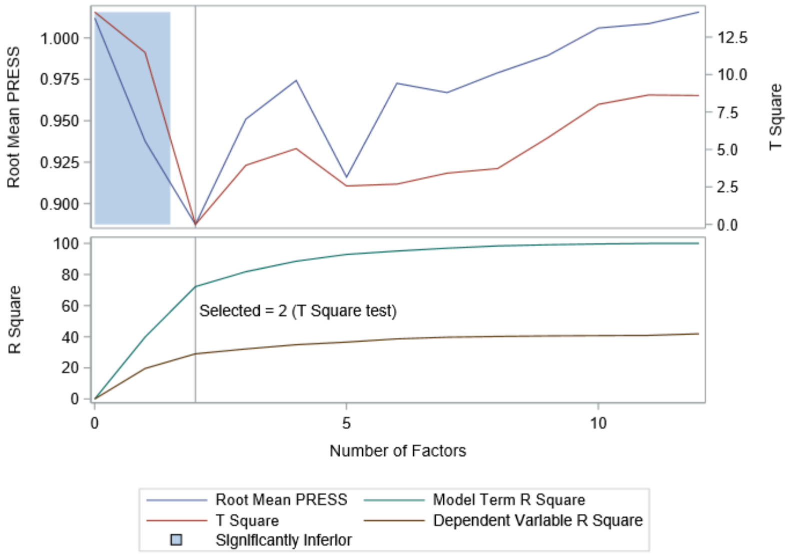
Minimum root mean PRESS	0.8874
Minimizing number of factors	2
Smallest number of factors with p > 0.1	2

The percent variation table 15 shows how well predictor and response variables are represented by each factor. The model effect shows that the two factors account for about 72.24% of variance in predictor variables. The two factors explain about 29.03% variability in response variables.

Table 15: Percent variation

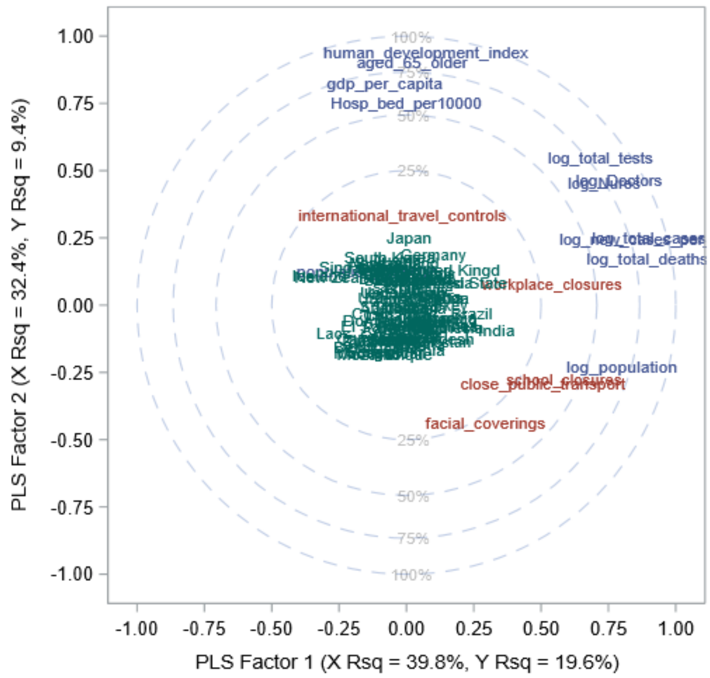
Percent Variation Accounted for by Partial Least Squares Factors				
	Model Effects		Dependent Variables	
Number of Extracted Factors	Current	Total	Current	Total
1	39.7968	39.7968	19.5775	19.5775
2	32.4431	72.2399	9.4500	29.0274

Figure 17: Cross-Validation Analysis



The left figure 17 is the cross-validation analysis plot showing the variance explained and root mean PRESS for 12 factors. As we can see from the figure, after 2 factors there is obvious evidence showing overfitting. Therefore, 2 factors were retained in the model.

Figure 18: Correlation Loading Plot



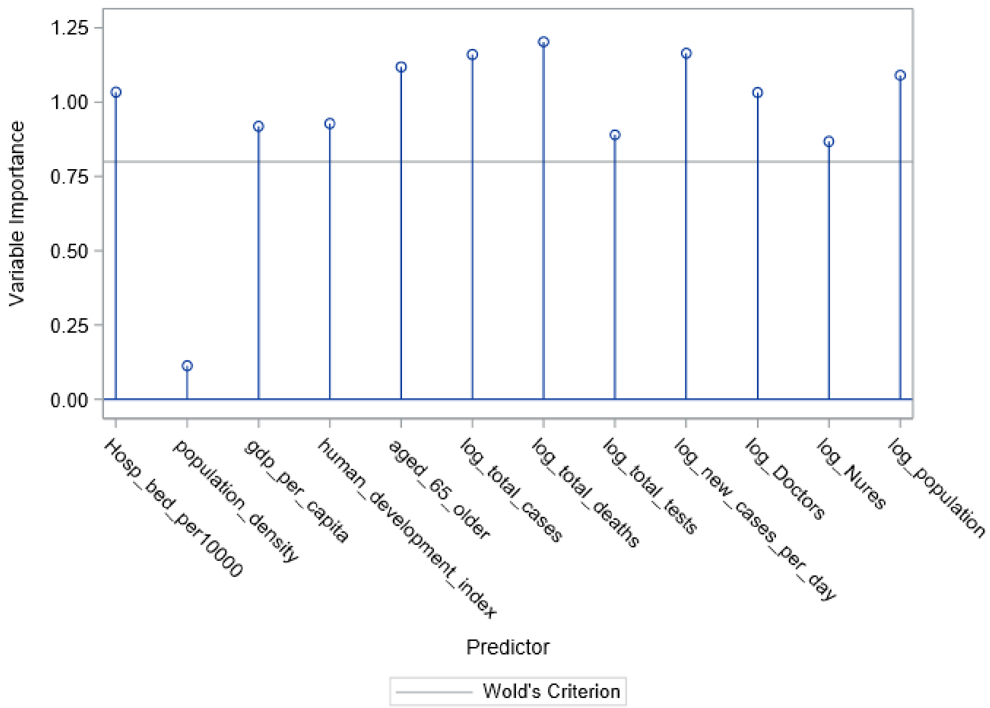
The correlation loading plot (Figure 18) shows the correlation with the two factors along the horizontal and vertical axes and the percent variation accounted for by each factor on the axis labels. The 12 predictor variables are plotted in blue, the 5 response variables are plotted in red and 84 countries selected by the model are plotted in green.

There is no obvious outlier and almost countries gathered together. However, Japan, United Kingdom, India, Brazil, Singapore and South Korea can be seen and are different from the cluster.

Huamn_development_index is most strongly correlated with Factor 2. Other variables from general information group like aged_65_older and gdp_per_capita are also positively correlated with Factor 2 as well as Hosp_bed_per10000. Log_total_deaths, Log_total_cases and Log_new_cases_per_day are strongly positively correlated with Factor 1. Log_population, Log_tests, Log_nurses, Log_doctors have a moderately positive correlation with Factor 1. Most of predictor variables are predicted well by the factors.

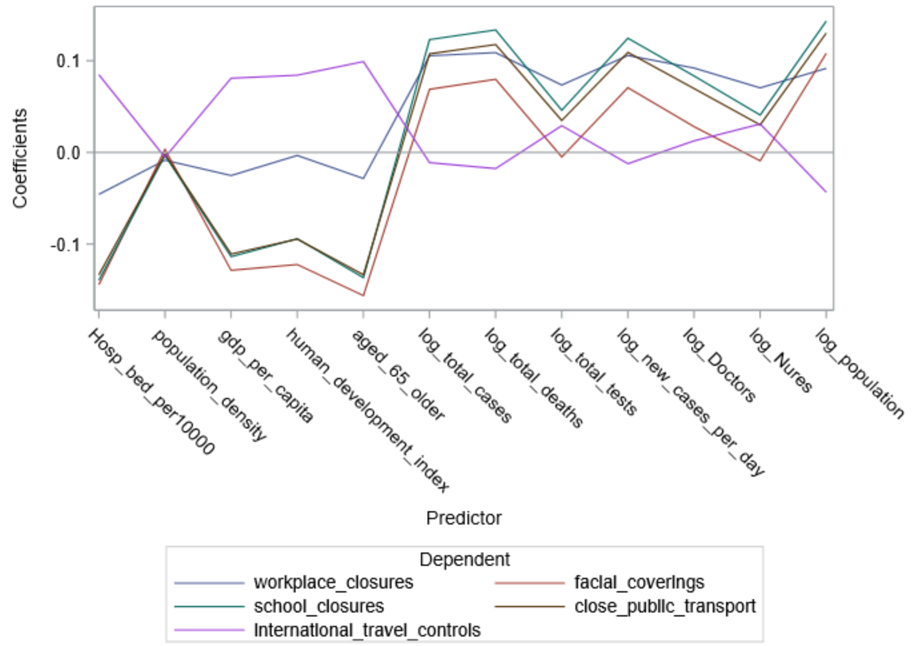
International_travel_controls is positively correlated with Factor 2 and workplace_closures is positively correlated with Factor 1. Facial_coverings, close_publica_transport and school_closures are weakly positively correlated with Factor 1 and weakly negatively correlated with Factor 2.

Figure 19: Variables Importance Plot



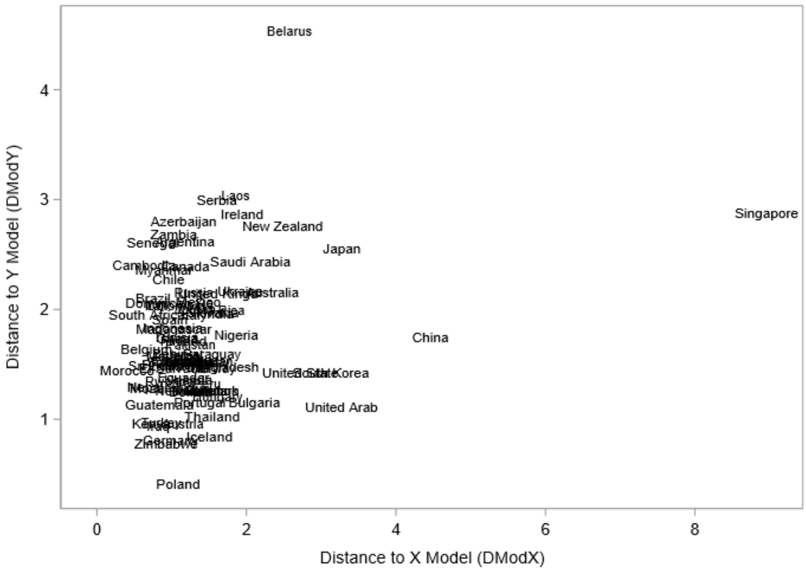
The figure 19 is variable importance plot showing the most useful predictors for predicting the response variables. The importance of variables is based on the contribution of each predictor in fitting the PLS model for both predictors and response. As we can see, except population_density, all predictor variables' importance is over 0.8 so they are important to the model. Log_total_deaths is the most important variable among them.

Figure 20: Profiles of regression coefficient estimates



The figure 20 is profiles of regression coefficient estimates showing the predictors' importance in the prediction of each response. Similar with last plot, predictors with big VIP (variable importace) tend to have big coefficients (in absolute value) like log_total_deaths, log_total_cases, log_new_cases_per_day, log_population aged_65_older and Hosp_bed_per10000. However, the predictor population_density that had small VIP also has a relatively small coefficient here.

Figure 21: Distance to Response and Predictor Models



The figure 21 above shows the distance of each country from the model in x and y space. It should be that none of country is dramatically far away from other countries in the model. However, Singapore, Belarus and China are very father from than the rest of countries with different distance and direction.

6.0 Key Findings

This chapter will summarize some key finding and some interesting insights from the analysis.

The principal component analysis indicated that the number of doctors, nurses and people are highly associated with PC 1 as well as total Covid19 confirmed cases, deaths and new cases per day. It also showed that there is a strong correlation between GDP per capita, share of population that is 65 years and older, human development index and PC 2. Closures of school, public transport and facial coverings are highly associated with each other and they are negatively associated with population density.

The biplot from PCA showed that some developing countries with negative PC 2 and positive PC 1 weightings has high government strict actions for Covid19 like closing school, workplace, public transport and wearing facial coverings. Most developed countries with positive PC 2 weighting have high human development index, share of population that is 65 years and older, GDP per capita and the number of hospital beds per 10000 people. However, some developed countries with negative PC 1 weighting have very little government strict actions for Covid19 like New Zealand, Finland, South Korea, Belarus and so on.

The factor analysis showed similar finding with principal component analysis. For example, the number of doctors, nurses and people are associated with each other, total Covid19 confirmed cases, deaths and new cases per day are associated with each other. However, the factor analysis can not explain other variables well.

Although the cluster did not find the most similar counties by their Covid19 situation and response because of the overlaps, it still has some interesting insights. The United State and India were in a cluster and were distinguished from other countries. Russia and China are in a cluster and then they are in another cluster with United Kingdom. As we expected New Zealand, Iceland and Taiwan are located in a big cluster with other countries.

The MDS showed that India, Brazil and United State are the most far away from other countries according to Covid19 cases situation. New Zealand, Palestine, Yemen, Nicaragua, Laos, Singapore, Japan and some countries are different from most countries according to government response.

The canonical correlation analysis showed that countries with high confirmed cases, deaths and new cases per day tend to require close workplace and school. Countries with high capability of testing people have a weak intention to require closing workplace and controlling international travel. International travel seems not be affected by Covid19 cases. Countries with a lot of doctors and nurses tend to require workplace closures. Countries with high human development index tend to close workplace during Covid19 period. Countries with more older people and high GDP per capita tend not to close public transport during Covid19 period.

The discriminant analysis indicated that there are 20 countries' government "overreacted" or "underreacted" the Covid19 according to the new cases per day and the number of hospital beds per 10000 people. There are 6 countries' government "overreacted" the Covid19 like South Korea, Liberia, Australia, Chad, Guinea and Rwanda. There are 14 countries' government "underreacted the Covid19 like Finland, Russia, Sudan, Thailand and so on.

The partial analysis indicated which variables are more important and more correlated with a set of government response variables. It showed that log of total deaths is the most important variables for predicting government response. Rest of variable except population density are all relatively important for predicting government response.

7.0 Conclusion and Limitation

The whole analysis captures the overall relationship between common variables using in the research about Covid19 and difference and similarity between countries. Although those common variables are created to analysis in Covid19 research, each variable stands for its own meaning and also associated with each other at the same time.
Given analysing the Covid19 indicator is a rapidly changed and broad topic, a lot of indicators which was not used in this report and combinations of countries could be analysed in Covid19 research. For example, number of Covid19 patients and number of Covid19 patients in intensive care (ICU) can be added into Covid19 Cases group. They show how dangerous Covid19 can be and how they affect Government Response variables.
There is another limitation of the analysis that it is a cross-section data and small sample size were the main limitations to the current study (Ali A Ghweil, 2020, July 17). Different date and time data can be collected and create a panel data to see the relationship of the trend of variables among countries.

8.0 Reference

1. David Kennedy (2020, June 30) How experts use data to identify emerging COVD-19 success stories. Retrieved from https://ourworldindata.org/identify-covid-exemplars

2. Lei S, Jiang F, Su W, et al. (2020, March 14) Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection.

3. Wikipedia. Factor analysis. Retrieved June 2 2021. Retrieved from *https://en.wikipedia.org/wiki/Factor_analysis#:~:text=Factor%20analysis%20is%20a%20statistical,of%20unobserved%20variables%20called%20factors*.

4. Thomas Hale. (2021, Jan 1) A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker. Retrieved from https://doi.org/10.1038/s41562-021-01079-8.

5. Ali A Ghweil. (2020, July 17). Characteristics, Outcomes and Indicators of Severity for COVID-19 Among Sample of ESNA Quarantine Hospital's Patients, Egypt: A Retrospective Study.

9.0 Appendices

9.1 Python Code (Data Cleaning)

```
!pip install ipython-sql
import pandas as pd
import sqlite3
import datetime as dt
import numpy as np
import requests
from bs4 import BeautifulSoup
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\demograohy.csv')

df['country'].unique()

case_test = pd.read_csv('D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\owid-covid-data.csv')

pd.to_datetime(case_test['date'],format = '%Y-%m-%d')

case_test_dropcountries=
case_test.loc[:,['continent','location','date','total_cases','new_cases','total_deaths','total_tests']
]

df_case_test_dropcountries =
pd.DataFrame({'continent':[],'location':[],'date':[],'total_cases':[],'new_cases':[],'total_deaths':[],
'total_tests':[]})
for i in list(df['country']):
    df_case_test_dropcountries =
df_case_test_dropcountries.append(case_test_dropcountries[case_test_dropcountries['locatio
n']==i])

# total_cases(last day of 2020)

case_test

# Delecting unneccesary columns
total_cases = case_test.loc[:,['location','date','total_cases']]
# Counting the last day in 2020
```

```
total_cases = total_cases[total_cases['date'] == '2021-06-08']
#subseting countries which exists in df's criteria
df_total_cases = pd.DataFrame({'location':[],'date':[],'total_cases':[]})
for i in list(df['country']):
	df_total_cases = df_total_cases.append(total_cases[total_cases['location']==i])
# resetting index
df_total_cases['index'] = list(range(0,114))
df_total_cases.set_index('index',inplace = True)
#Adding the column to df
df['total_cases']=df_total_cases['total_cases']
df

#	New cases per day(last day of 2020)

New_cases_per_day = case_test.loc[:,['location','date','new_cases']]

pd.to_datetime(New_cases_per_day['date'],format = '%Y-%m-%d')

New_cases_per_day = New_cases_per_day[New_cases_per_day['date']<'2021-06-08']

New_cases_per_day = New_cases_per_day.groupby('location').mean()

New_cases_per_day.reset_index(level = [0],inplace=True)

df_New_cases_per_day = pd.DataFrame({'location':[],'date':[],'new_cases':[]})

for i in list(df['country']):
	df_New_cases_per_day =
df_New_cases_per_day.append(New_cases_per_day[New_cases_per_day['location']==i])

df_New_cases_per_day ['index'] = list(range(0,114))
df_New_cases_per_day .set_index('index',inplace = True)

df_New_cases_per_day

df['new_cases_per_day']=df_New_cases_per_day['new_cases']
df

# Total death(last day of 2020)

# Delecting unneccesary columns
total_deaths = case_test.loc[:,['location','date','total_deaths']]
# Counting the last day in 2020
total_deaths = total_deaths[total_deaths['date'] == '2021-06-08']
```

```
#subseting countries which exists in df's criteria
df_total_deaths = pd.DataFrame({'location':[],'date':[],'total_deaths':[]})
for i in list(df['country']):
	df_total_deaths = df_total_deaths.append(total_deaths[total_deaths['location']==i])
# resetting index
df_total_deaths['index'] = list(range(0,114))
df_total_deaths.set_index('index',inplace = True)
#Adding the column to df
df['total_deaths']=df_total_deaths['total_deaths']
df

# Total test

# Delecting unneccesary columns
total_tests = case_test.loc[:,['location','date','total_tests']]
# Counting the last day in 2020
total_tests = total_tests[total_tests['total_tests'].notnull()]
total_tests = total_tests.sort_values('date').groupby('location').tail(1)
#subseting countries which exists in df's criteria
df_total_tests = pd.DataFrame({'location':[],'date':[],'total_tests':[]})
for i in list(df['country']):
	df_total_tests= df_total_tests.append(total_tests[total_tests['location']==i])
#Merge df with total_tests
df = pd.merge(df,df_total_tests,how = 'left',left_on='country',right_on='location')
df.drop(columns = ['location','date'],inplace = True)
df

# Number of Doctors

Doctors = pd.read_csv('D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\Doctors.csv')

fix_list = ['Bolivia','Iran','Laos','Russia','South Korea','Taiwan','United Kingdom','United
States','Vietnam']

losted=['Bolivia (Plurinational State of)','Iran (Islamic Republic of)',"Lao People's Democratic
Republic",'Russian Federation','Republic of Korea','United Kingdom of Great Britain and
Northern Ireland','United States of America','Viet Nam']

fix_dict = dict(zip(losted,fix_list))

Doctors['Location'].replace(fix_dict,inplace=True)

Doctors = Doctors.loc[:,['Location','Period','FactValueNumeric']]
```

```
Doctors = Doctors.sort_values('Period').groupby('Location').tail(1)
Doctors.rename(columns={'FactValueNumeric':'Doctors'},inplace=True)

df = pd.merge(df,Doctors,how='left',left_on='country',right_on='Location')

df.drop(columns=['Period','Location'],inplace = True)
df

df[df['Doctors'].isnull()]

# Number of nurses

Nures = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\Nurses.csv')

fix_list = ['Bolivia','Iran','Laos','Russia','South Korea','Taiwan','United Kingdom','United
States','Vietnam']

losted=['Bolivia (Plurinational State of)','Iran (Islamic Republic of)',"Lao People's Democratic
Republic",'Russian Federation','Republic of Korea','United Kingdom of Great Britain and
Northern Ireland','United States of America','Viet Nam']

fix_dict = dict(zip(losted,fix_list))

Nures['Location'].replace(fix_dict,inplace=True)

Nures = Nures.loc[:,['Location','Period','FactValueNumeric']]
Nures = Nures.sort_values('Period').groupby('Location').tail(1)
Nures.rename(columns={'FactValueNumeric':'Nures'},inplace=True)

df = pd.merge(df,Nures,how='left',left_on='country',right_on='Location')
df.drop(columns=['Period','Location'],inplace = True)
df

# Hospital bed per 10000

Hosp_bed_per10000 = pd.read_csv('D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\Hosp_bed_per10000.csv')

fix_list = ['Bolivia','Iran','Laos','Russia','South Korea','Taiwan','United Kingdom','United
States','Vietnam']

losted=['Bolivia (Plurinational State of)','Iran (Islamic Republic of)',"Lao People's Democratic
Republic",'Russian Federation','Republic of Korea','United Kingdom of Great Britain and
```

```
Northern Ireland','United States of America','Viet Nam']

fix_dict = dict(zip(losted,fix_list))

Hosp_bed_per10000['Location'].replace(fix_dict,inplace=True)

Hosp_bed_per10000    = Hosp_bed_per10000 .loc[:,['Location','Period','FactValueNumeric']]
Hosp_bed_per10000    = Hosp_bed_per10000 .sort_values('Period').groupby('Location').tail(1)
Hosp_bed_per10000.rename(columns={'FactValueNumeric':'Hosp_bed_per10000'},inplace=True)

df = pd.merge(df,Hosp_bed_per10000,how='left',left_on='country',right_on='Location')
df.drop(columns=['Period','Location'],inplace = True)
df

# Workplace_Closures

Workplace_Closures = pd.read_csv('D:\Master of Analytics\Multivariate Analysis\Final prejeject\Datasets\workplace-closures-covid.csv')

Workplace_Closures = Workplace_Closures.groupby('Entity').mean()
Workplace_Closures.reset_index(inplace=True)

df = pd.merge(df,Workplace_Closures,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# Face_Covering

Face_Covering = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final prejeject\Datasets\face-covering-policies-covid.csv')

Face_Covering = Face_Covering.groupby('Entity').mean()
Face_Covering.reset_index(inplace=True)
df = pd.merge(df,Face_Covering,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# School_Closures

School_Closures    = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final prejeject\Datasets\school-closures-covid.csv')

School_Closures
```

```
School_Closures = School_Closures.groupby('Entity').mean()
School_Closures.reset_index(inplace=True)
df = pd.merge(df,School_Closures,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# Public_Transport

Public_Transport = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\public-transport-covid.csv')

Public_Transport
Public_Transport = Public_Transport.groupby('Entity').mean()
Public_Transport.reset_index(inplace=True)
df = pd.merge(df,Public_Transport,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# International_travel

International_travel = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\international-travel-covid.csv')

International_travel = International_travel.groupby('Entity').mean()
International_travel.reset_index(inplace=True)
df = pd.merge(df,International_travel,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# Stringency_index

Stringency_index = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\covid-stringency-index.csv')

Stringency_index = Stringency_index.groupby('Entity').mean()
Stringency_index.reset_index(inplace=True)
df = pd.merge(df,Stringency_index,how='left',left_on='country',right_on='Entity')
df.drop(columns=['Entity'],inplace = True)
df

# General info

general = pd.read_csv(r'D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\owid-covid-data.csv')
```

```
general =
general.loc[:,['location','population','population_density','gdp_per_capita','human_development_index','aged_65_older']]

general = general.drop_duplicates(subset=['location'])

df = pd.merge(df,general,how='left',left_on='country',right_on='location')

df.drop(columns=['location'],inplace=True)

df.isnull().sum()

df[df['total_tests_per_thousand'].isnull()]

df[df['stringency_index']<=50]

df.loc[df['stringency_index']<=50,'Government_Response']='Loose'
df.loc[df['stringency_index']>50,'Government_Response']='Strict'

# df.to_csv('D:\Master of Analytics\Multivariate Analysis\Final preject\Datasets\Covid19.csv',index=False)
```

9.2 SAS Code (Data Modelling)

```
proc import datafile="D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\Covid19.csv" dbms=csv out=Covid19 ;
proc import datafile="D:\Master of Analytics\Multivariate Analysis\Final
preject\Datasets\Covid19_droped.csv" dbms=csv out=Covid19_droped ;


ods graphics on;
proc princomp data=covid19
               plots=(matrix score(ncomp=3) patternprofile
               pattern(ncomp=3));
var total_cases--aged_65_older;
id country;
run;


proc prinqual data=covid19  mdpref;
  transform identity(total_cases--aged_65_older);
```

```
id country;
run;


proc factor data=covid19 plots=all method=ml priors=smc HEYWOOD;
    var total_cases--aged_65_older;
run;

ods select orthrotfactpat patternplot;
proc factor data=covid19 plots=loadings out=factor
              method=ml priors=smc n=2 r=v flag=.3 fuzz=.2 HEYWOOD;
var total_cases--international_travel_controls population--aged_65_older;
run;


proc sgplot data=factor;
bubble x=factor1 y=factor2 size=international_travel_controls / transparency=0.4;
run;

15


ods graphics on;
proc cluster data=covid19 method=average ccc pseudo outtree=tree print=15
plots=den(height=rsq);
var total_cases--total_tests workplace_closures--international_travel_controls;
id country;
run;


proc distance data=covid19 method=euclid out=dis;
var INTERVAL(total_cases--total_tests);
id country;
run;
proc mds data=dis dim=2 converge=0.0001 level=ordinal;
id country;
run;


proc distance data=covid19 method=euclid out=dis;
var INTERVAL(workplace_closures--international_travel_controls);
id country;
run;
```

```
proc mds data=dis dim=2 converge=0.0001 level=ordinal;
id country;
run;



data covid19_per;
set covid19;
total_cases_per1000=(total_cases/population)*1000;
total_deaths_per1000=(total_deaths/population)*1000;
total_tests_per1000=(total_tests/population)*1000;
new_cases_per_day_per1000=(new_cases_per_day/population)*1000;
Doctors_per1000=(Doctors/population)*1000;
Nures_per1000=(Nures/population)*1000;
run;





data covid19_test;
set work.covid19_per;
if country="China" then delete ;
if country="United State" then delete;
if country="India" then delete;
if country="United Kingd" then delete;
if country="Russia" then delete;
run;

ods graphics on;
proc cluster data=covid19_test method=average ccc pseudo outtree=tree print=15
plots=den(height=rsq);
var total_cases--total_tests;
id country;
run;


proc means data=covid19;
var total_cases--total_tests;
run;

proc sgplot data=covid19;
vbox total_cases;
run;
```

```
proc sgplot data=covid19;
vbox total_deaths;
run;
proc sgplot data=covid19;
vbox total_tests;
run;
proc sgplot data=covid19;
vbox new_cases_per_day;
run;
data covid19;
set covid19;
log_total_cases=log(total_cases);
log_total_deaths=log(total_deaths);
log_total_tests=log(total_tests);
log_new_cases_per_day=log(new_cases_per_day);
run;

ods output cancorr=a;
proc cancorr data=covid19 vprefix=Cases wprefix=Gov ncan=2 out=cancorr ;
var log_total_cases--log_new_cases_per_day;
with workplace_closures--international_travel_controls;
run;
proc sgscatter data=cancorr;
   plot Gov1*Cases1 Gov2*Cases2;
run;



data covid19;
set covid19;
log_Doctors=log(Doctors);
log_Nures=log(Nures);
run;

ods output cancorr=a;
proc cancorr data=covid19 vprefix=Health wprefix=Gov ncan=2 out=cancorr ;
var log_Doctors log_Nures Hosp_bed_per10000  ;
with workplace_closures--international_travel_controls;
run;
proc sgscatter data=cancorr;
   plot Gov1*Health1 Gov2*Health2;
run;
```

```
data covid19;
set covid19;
log_population=log(population);
run;
proc cancorr data=covid19 vprefix=GI wprefix=Gov ncan=2 out=cancorr ;
var log_population population_density aged_65_older gdp_per_capita
human_development_index ;
with workplace_closures--international_travel_controls;
run;
proc sgscatter data=cancorr;
   plot Gov1*GI1 Gov2*GI2;
run;
```


```
proc cancorr data=covid19 vprefix=GI wprefix=Gov ncan=2 out=cancorr ;
var log_population population_density aged_65_older gdp_per_capita
human_development_index ;
with workplace_closures--international_travel_controls;
run;
proc sgscatter data=cancorr;
   plot Gov1*GI1 Gov2*GI2;
run;
```


```
ods output canonicalmeans=b(rename=(can1=can1c));
proc candisc data=covid19 out=candout;
class Government_Response;
var log_total_cases--log_population Hosp_bed_per10000 population_density--
aged_65_older  ;
run;
data plot;
set candout b;
if Government_Response="Loose" then auxiliary=0.05;
if Government_Response="Strict" then auxiliary=-0.05;
auxiliaryc=auxiliary;
run;

proc sort data=plot;
by Government_Response fromGovernment_Response;
run;
proc sgplot data=plot nocycleattr;
scatter x=can1 y=auxiliary / group=Government_Response;
scatter x=can1c y=auxiliaryc / group=fromGovernment_Response
```

```
markerattrs=(size=20);
yaxis grid values=(-1 to 1 by 1);
run;


proc stepdisc data=covid19 method=stepwise;
class Government_Response;
var log_total_cases--log_population Hosp_bed_per10000 population_density--
aged_65_older   ;
run;



proc discrim data=covid19 pool=test slpool=.05 listerr;
class Government_Response;
priors prop;
var &_STDVAR;
id country;
run;

proc discrim data=covid19 pool=test slpool=.05 listerr;
class Government_Response;
priors prop;
var log_total_cases--log_population Hosp_bed_per10000 population_density--
aged_65_older ;
id country;
run;


* Partial least square;
ods graphics on;
proc pls data = covid19 method = pls(algorithm=nipals)
       cv=one cvtest(seed=608789001) plot=(vip xyscores xscores dmodxy
parmprofiles dmod corrload);
model workplace_closures facial_coverings school_closures close_public_transport
international_travel_controls = Hosp_bed_per10000 population_density--
aged_65_older log_total_cases--log_population;
id country;
run;
```