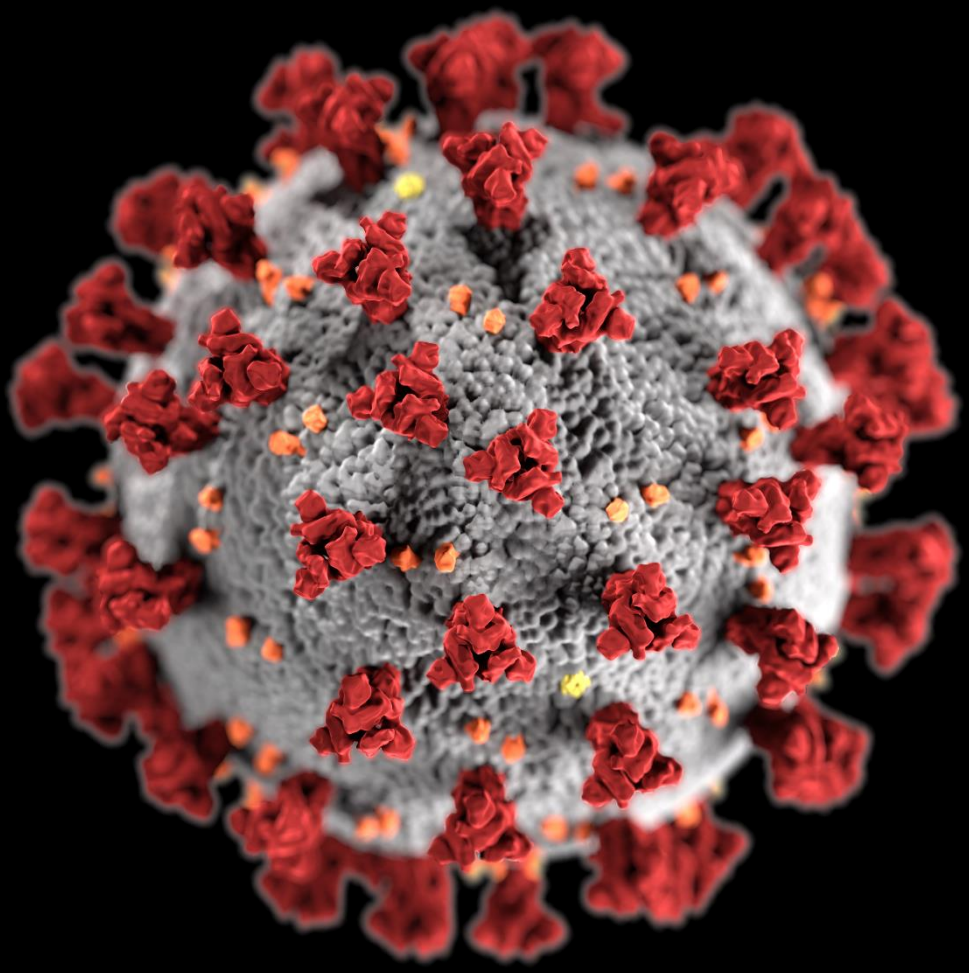


COVID-19 Pandemic

A Review of the Global Situations and Responses

Jean Liu

Submission for IBM Data Science Professional Certificate on COURSERA

26 May 2020

Table of Contents

Introduction 1
Data 2
Methodology 3
Reproduction Number (R nought / R0) 3
International COVID-19 Situation 4
Five Hardest Hit Major Cities 7
Clustering Model Development 12
Results 15
Country Clusters 15
Cluster Profiles 18
Government Responses 20
Discussion 22
Economic Catastrophe 22
Onward and Upward 23
Conclusion 24
References 26
Appendices 30
Appendix A. Data Dictionary 30
Appendix B. COVID-19 Individual Cluster Profiles 35
Appendix C. COVID-19 Attribute Correlation Heatmap 36
Appendix D. COVID-19 Attribute Correlation Scatter Plot 37

Introduction

The novel coronavirus outbreak caused by 'severe acute respiratory syndrome coronavirus 2' (SARS-CoV-2) originated from Wuhan City of Hubei Province in Mainland China back in early December 2019. It initially hit the neighbouring regions in Asia and then rapidly spread across other continent to the US, Europe, and then the rest of the world. On 11 March 2020 the World Health Organisation declared it a global pandemic and officially named the disease as 'COVID-19' [1]. So far this pneumonia-like disease has not only brought to the entire world public health crisis which has taken 341,722 lives (as of 26 May 2020), overwhelmed healthcare systems of numerous countries and has put more than half of world population into nation-wide lockdown restrictions as an emergency measure. The consequences are unprecedented in modern history and have become tremendously destructive to the global economy and people's livelihood.

Notably, national responses to the threat differ remarkably in the efficiency and consistency of decision making, stringency and creativity of containment strategies. Even countries across Europe which are experiencing similar stages of epidemic, no cohesion can be seen and lack of solidarity has been constantly on the table. Profound damage in public well-being, economic downturn, and social unrest within the community have even imposed extra crisis on everyone under those strict confinements. Arguably such disparity in government actions and their effectiveness could be attributed to availability of past experience in tackling with similar public health crisis, political agenda and diplomatic strategies of the nation, traditional and cultural values in society, healthcare capacity and infrastructure development, economic situation, population constitution and so on.

The situation has put the entire world under the greatest test. Every individual, household, business, government is so desperate to get their life back to normality, and have started easing the lockdown as soon as they have successfully 'flattened the curve'. For that, epidemiologists monitor whether the virus transmission measured by the 'basic reproduction number' (aka. R0) has been kept below 1 [2]. However, until an effective and safe vaccination has become globally available, the virus will never stop spreading and no one can be fully off their guard. In fact, second waves of infection have already emerged in a number of regions which at the onset of the pandemic succeeded in keeping the disease at bay, such as Singapore, South Korea, Iran, and even China [3] [4] [5].

Therefore, we will look into the current situations of selected countries and major cities which have been under recent spotlight in the media and examine why, what, when and how their country authorities have responded to this crisis. This study hopefully will help the general public and decision makers to get some questions answered and learn from what we have already been through for a better grip on the battle with the invisible enemy.

Data

The data to lay the groundwork of our analysis are of direct and indirect links to COVID-19. The data quality varies from different sources. Even large international institutions might not meet the standard of a cleaned, consistent, accurate, and timely fashion. WHO, the leading organisation to provide centralised global guidance during the pandemic, was the first place that came to our mind for retrieving daily number of cases and deaths. However, the data was found some underlying issues that miss our requirements. As the same information shared by European Centre for Disease Prevention and Control has no similar issues, it has been deployed for our analysis.

The disparity of government responses to the pandemic is a crucial part to be examined in this work, with particular focus on the effectiveness of various intervention measures and associated factors to consider. The Blavatnik School of Government at Oxford University has published 'COVID-19 Government Response Tracker' [6] and has shared the underlying time series data to the public, which plays a fundamental role in our study to complete the jigsaw puzzle.

Of the interest of this project, data for demographics, healthcare capacity and development, the level of public health, population, economic output and migration dynamics have also been collected, but met with no fewer challenges. Large part of these aggregated data in fact are not updated timely. Additionally, they are neither widely available or easily accessible even in large global bodies. Therefore, the figures presented here is the latest available online and should be interpreted with precaution. Pertinent sources of these national statistics include Worldmeters, The United Nations, Our World In Data, and Wikipedia.

Two separate datasets have been created from the sources of both Google COVID-19 Community Mobility Reports [7] and Foursquare API [8]. These two pieces of information look at five of the hardest hit cities in the west by coronavirus, namely, New York, London, Paris, Madrid, and Berlin. The data presents how daily footprints of Google users have been affected by restrictions on people's movements and the extent of impact on local businesses and public venues.

Upon completion of data acquisition, a series of data cleaning processes carried out includes setting data types, standardising country names, treatment of missing values, correcting misinformation, data reshaping, transformation, aggregation and normalisation for further analysis, updating column headers, and imputation of lost information for Taiwan and Singapore, two of the selected countries on the focus of this study. The final prepared data constitutes the following four datasets with corresponding data dictionaries available in Appendix A of this report.

- daily_covid_govresponse_df
- country_master_df
- goog_mob_df
- five_city_venues_df

Methodology

Reproduction Number (R nought / R0)

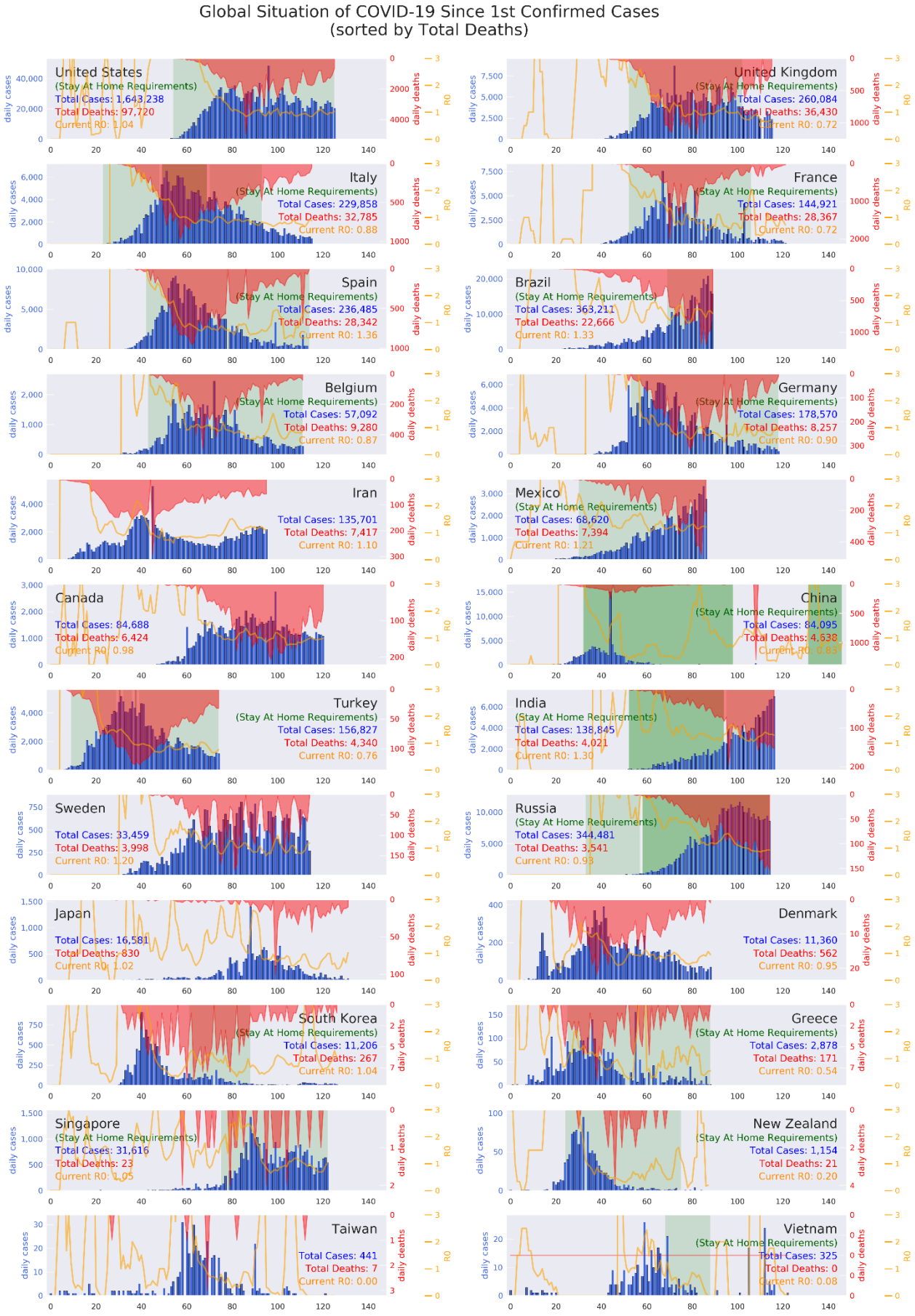
Over the last few days, a scientific term in epidemiology has been constantly under the spotlight and anxiously examined by numerous countries; i.e. 'Reproduction Number' [9]. This is an indicator showing the current speed of viral spreading and how the infected population will be scaling in the coming time period. Given no permanent solution is available for now, the spread of epidemic has to be kept within a limited scale to prevent draining the country's diagnostic and medical capacity, to allow the central government to be able to trace and track new infections, and most importantly to let a nation to be able to ease their lockdown and get the country economy back and running again.

However, estimating R0 with high accuracy is mathematically complicated and involves many factors and assumptions to be made. In this project, R0 is only to be used to present a general picture of current situations and not to predict future development of the pandemic. Our approach, a much simpler version, takes the average values to work out how many secondary infections have been resulted from one primary infection; i.e. the number of people infected by the same virus carrier. Here is the equation showing how R0 value has been estimated:

$$R0_{t0} = \frac{MA5_{t0}}{MA5_{t-5}}$$

When we look at the number of cases on a particular day, say, today (t0), and set them as secondary infection, we need to trace back few days ago to identify the group of people who potentially have passed on the virus. The evidences found so far by scientists indicate that the incubation period (from exposure to the virus to showing symptoms) could range between 1 and 14 days with 5 days most commonly observed [10]. Therefore, the number of infections registered 5 days ago (t-5) is the figure we would like to compare against. As the daily number generally fluctuates along the trend due to some randomness (aka. 'noise' in statistics) as well as other specific reasons, such as lag between reporting, causing its variation, we will average five days values to smooth out the noise to reflect current level of virus spreading. With the moving average window set on 5 days (aka. MA5), although the R0 estimate is for t0, it is actually closer to the truth 2 days ago. Theoretically the shorter the smoothing period, the more current R0 becomes. However, if there was any substantial and unusual spike or dip occurring within the last 5 days, it deserves to be accounted for and hence should add weight to the average value. This parameter has also been adopted by American Medical Association [11]. Therefore, 5 days smooth period keeps the balance between relevance and caution in our estimates. This result has been added to 'daily_covid_govresponse_df' which is to be visualized as part of the following time series chart.

International COVID-19 Situation



Data Source: European Centre for Disease Prevention and Control
(https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide)
Accessed on: 2020-05-25 09:41:40

The above time series reflect current situations as of the data accessed date indicated underneath the plots, presenting indicators including daily registered cases, daily deaths, daily R0, total cases, total deaths, the latest R0, and the duration of country lockdown. A few points help reading the charts with ease:

1) X axis represents the number of days after individual country's first reported case.
2) Y axis of daily death toll should be read top-down for the resemblance of distressful image for the loss of lives.
3) R0 values above 3 are truncated so that values below 1 which are where the containment efforts should be focused on can be read more clearly.
4) Green shaded areas in the background represent national lockdown enforcement, with two levels of darkness specifying the severity of legal enforcement.

Additionally, there are caveats when comparing real situations amongst countries, due to issues with both under and overreporting. Factors like discrepant measurement units, differential testing policies, non-standardised criteria and only certain places of deaths counted, lag between diagnosis and statistic reporting, political influence, population demographics, healthcare infrastructure and capacity, and etc. have all played comparable roles to deviate those figures [12]. Therefore, any COVID-19 data including the following results should be interpreted with these limitations in mind.

These selected countries are either previous or current epicentres across the globe. Several features have been observed and have been summarized as follows.

1. **Contagion Stage:**
   Countries in the east have generally reached the rightmost of the time span which is about where Mainland China current is, due to their geographical proximity. In the west, whilst the majority have past the peaks, a few of them appear to have a much slower and prolonged downward trail. Notably, Brazil, India, Mexico are on an increasingly upward trend and are continually seeing new highs at the moment [13].
2. **Transmission Rate:**
   As the daily number of cases stay low, the speed of infectious spread, indicated by R0 value, is highly volatile and could easily proliferate at a rate above 1, representing exponential growth of the disease. This is due to the amount of increment to small numbers would cause much larger growth than to large numbers. Impressively the virus transmission in Taiwan so far has been kept at zero as of the latest R0 on 26 May 2020.
3. **Seasonality of Reported Cases And Deaths:**
   Patterns of regularity can be observed in lots of regions like the United States, the United Kingdom, Brazil, Germany, Sweden, and Mexico, and usually are attributed to various administrative levels involved in data collection within a nation [14]. In the face of a crisis with rapid changes, whether such lag could be a pitfall for taking timely actions should be assessed.
4. **Timing of Introducing Movement Restrictions** [15]**:**
   The swiftest lockdown enforcement has been triggered by Turkey government with just about 9 days into the first reported case, followed by Italy, Greece, and New Zealand with less than 30 days into the first infections. Brazil has the longest waiting time to act, followed by the United States, the United Kingdom, France, and Germany. (Singapore is actually in its second wave of infections.) However, Greece and New Zealand have relatively much

less severity of contagion than other countries. Clearly this reflects that behind the same restrictive measure lie dissimilar perspectives and proactiveness of policymakers [16] [17].

5. **Variation of Cases and Deceases post Initiating Lockdown Restrictions:**
   Several countries under lockdown for more than 30 days have apparently slowed the viral growth. However, such drastic approach is not required for the contagion running down the curve in places like Taiwan, Japan, Denmark, Sweden and Canada. Although the escalation of restrictions does see an anticipated decrease of infected cases in Italy, in India and Mexico the effect appears to be the opposite.
6. **Lag between Peaks of Registered Cases And Deaths:**
   For countries which are currently downhill have shown gaps between infections and deceases of various size, ranging from 2 to 20 days. For instance, Italy has its infection peak on the 51st day and death at its top on the 57th day, while New Zealand on the 32nd day for confirmed cases and death on the 46th day. Relatively wider gaps are in Germany, South Korea, and New Zealand as opposed to Denmark, France and Italy. Assuming a constant mortality rate on any given day within the same country, it might imply the ability and capacity of how the medical services are treating their patients.
7. **Looming Threat and Existing Condition of Second infections:**
   A number of countries in Asia are having second waves of infection. In China where the disease explosion originated [5], localized lockdown has been resumed in Shulan of Jilin Province on the 10th of May when a cluster of infections was reported. Countries like Vietnam, South Korea and Singapore where national lockdown was not in consideration at their first infection surge have also imposed stay at home requirements to contain second outbreaks [4] [3]. In Iran the shape of a second wave is in a clear shape approaching similar level of severity as before.

The stringency of government responses widely differs; not every country sees it necessary to tame the contagion, even with widespread of virus within the territories, such as Japan, Sweden, and Denmark. Taiwan [18] and Vietnam [19] are the two places which have been highly praised for keeping the epidemic under control. In particular, while the entire human race has been suffering from global lockdown, Taiwan authorities have successfully managed the crisis without imposing movement restrictions and have kept its economic disruption to the minimum.

Five Hardest Hit Major Cities

From the top nations with the highest death tolls, five metropolitan cities in the west are selected to be explored on how government stringency on people's movements affect actual behaviours as well as local businesses. This is a extended study based on 'Google COVID-19 Community Mobility Reports' on top of the location data from Foursquare Places API. To begin with, we will compare city profiles with dimensions of local business venues and see how government responses have hit number of visits from Google account users.

1. Geo Data for City Metro Stations

To gather city venues, we use metro stations as central points, from which areas within a certain radius are places we want to extract through Foursquare API calls. Therefore, the latitude and longitude coordinates of these stations is what needs to be collected first. As our first step, we retrieve the lists of all metro stations from the websites stated in the table below, and then use Nominatim geocoder from GeoPy package to obtain station coordinates for Paris, Madrid, and Berlin as these websites do not provide location data.

City	Source for Lists of Metro Stations
New York City	https://geo.nyu.edu/download/file/nyu-2451-34503-geojson.json
London	https://wiki.openstreetmap.org/wiki/List_of_London_Underground_stations
Paris	https://en.wikipedia.org/wiki/List_of_Paris_M%C3%A9tro_stations
Madrid	https://www.metromadrid.es/en
Berlin	https://en.wikipedia.org/wiki/Berlin_U-Bahn https://en.wikipedia.org/wiki/Berlin_S-Bahn

Five output datasets are created for each district; i.e. 1) ldn_tube_df, 2) nyc_subway_df, 3) paris_metro_df, 4) madrid_metro_df, and 5) berlin_metro_df. The following screenshot for London underground stations exemplifies the typical output structure where a row represents a station, and there are 291 stations in London.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 291 entries, 0 to 290
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   station    291 non-null    object
 1   latitude   291 non-null    float64
 2   longitude  291 non-null    float64
dtypes: float64(2), object(1)
memory usage: 6.9+ KB
```

	station	latitude	longitude
0	Acton Town	51.5025	-0.27812

2. Foursquare Venue Data

With the location data of all metro stations in these cities, an API call then is made for each station to pull out 100 venues within a radius of 999 meters. Higher values for these two parameters did not work in our experiments, despite what has been stated on Foursquare Developers site [20]. Five datasets have then been produced as 1)

ldn_venue_df, 2) nyc_venue_df, 3) paris_venue_df, 4) madrid_venue_df, and 5) berlin_venue_df.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 42316 entries, 0 to 42315
Data columns (total 8 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   station              42316 non-null  object
 1   stn_latitude         42316 non-null  float64
 2   stn_longitude        42316 non-null  float64
 3   venue                42316 non-null  object
 4   venue_lat            42316 non-null  float64
 5   venue_lon            42316 non-null  float64
 6   venue_categories     42316 non-null  object
 7   venue_categories_id  42316 non-null  object
dtypes: float64(4), object(4)
memory usage: 2.6+ MB
```

The structure presented on the left is what these venue datasets are comprised of, with each row representing one venue. There are coordinates for both corresponding station and venue, as well as venue category. Foursquare uses venue category to label the type of business or purpose for a place, which is essential for us to relate what type of places are not allowed to open under lockdown restrictions. We will also translate these categories later on into those defined in Google COVID-19 Community Mobility Reports to get an idea of the extent of economic impact hitting these five regions. The snapshot below shows the first line of the dataset.

	station	stn_latitude	stn_longitude	venue	venue_lat	venue_lon	venue_categories	venue_categories_id
0	101_Van_Cortlandt_Park_242_St	40.889248	-73.898583	Van Cortlandt Park Running Track	40.888341	-73.897648	Track	4bf58dd8d48988d106941735

The summary displayed below tells us that New York City not only has the highest number of subway stations with most public venues, but also contains the highest venue density in the district. In contrast, Berlin is the least dense city. Let's take a look next at what these venues are within each district.

```
ldn_tube_df: 291 / ldn_venue_df: 18864 / Venues per station: 64.82
nyc_subway_df: 491 / nyc_venue_df: 42316 / Venues per station: 86.18
paris_metro_df: 304 / paris_venue_df: 22476 / Venues per station: 73.93
madrid_metro_df: 201 / madrid_venue_df: 12745 / Venues per station: 63.41
berlin_metro_df: 327 / berlin_venue_df: 12870 / Venues per station: 39.36
```

3. Top 10 City Venues

Each city has its unique characteristics, shaped by local culture, people, activities and places. We are interested in finding out any distinctions amongst them characterised by their commercial places. As the venue categories indicate types of activity taking place in these venues, we count the total number of venues for each category within a city and calculate the portion of category share and rank the weight of various public activities in these cities. The following table shows what the majority of venues are in these places.

city	Top 1st Venues	Top 2nd Venues	Top 3rd Venues	Top 4th Venues	Top 5th Venues	Top 6th Venues	Top 7th Venues	Top 8th Venues	Top 9th Venues	Top 10th Venues
Berlin	Hotel	Café	French Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Restaurant	Bakery	Bar	Supermarket
London	Hotel	French Restaurant	Coffee Shop	Café	Italian Restaurant	Pizza Place	Bakery	Bar	Restaurant	Park
Marid	Hotel	Café	French Restaurant	Coffee Shop	Italian Restaurant	Pizza Place	Restaurant	Bar	Bakery	Supermarket
New York City	Hotel	French Restaurant	Coffee Shop	Café	Italian Restaurant	Pizza Place	Bakery	Bar	Park	Restaurant
Paris	French Restaurant	Hotel	Coffee Shop	Café	Italian Restaurant	Pizza Place	Bakery	Bar	Restaurant	Park

These five cities have barely any noticeable differences and they all serve travelers with plenty of hotels, coffee shops, and places for food. However, Berlin and Madrid, unlike other cities, have fewer green places but with relatively more supermarkets.

4. Foursquare Venue Category Hierarchy

At this stage of venue analysis, we are going to identify which businesses are not allowed to open during lockdown period and then label them with Google place categories [21]. Due to the hierarchical structure of Foursquare venue categories [22], it is necessary to know the first level of category for each venue in that labelling them become plausible, as lockdown rules do not name particular type of restaurants for example. Hence an API call is made to Foursquare endpoint to request details of the hierarchy, followed by a series of manipulation to create a reference table showing different levels of categories with associated id as exemplified below:

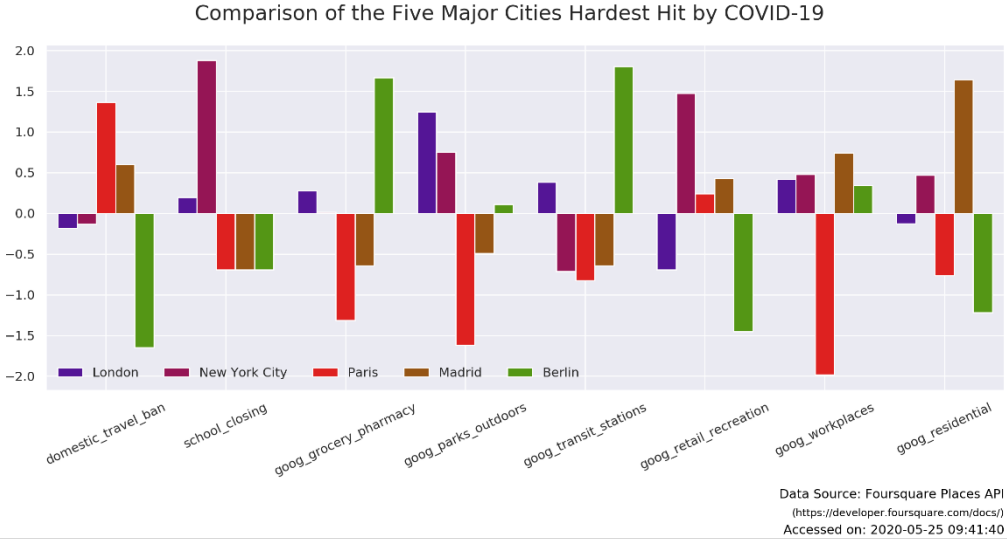
	id_1	categories_1	id_2	categories_2	id_3	categories_3	id_4	categories_4	id_5	categories_5
345	4d4b7105d754a06374d81259	Food	4bf58dd8d48988d1be941735	Latin American Restaurant	4bf58dd8d48988d1cd941735	South American Restaurant	4bf58dd8d48988d16b941735	Brazilian Restaurant	52939ab93cf9994f4e043a38	Northern Brazilian Restaurant

The highest level in the screenshot above is Food with Northern Brazilian Restaurant to be the lowest level in the hierarchy. With this knowledge, a large variety of restaurants or dinning places can then be attributed at once to Food venues, which belong to 'retail & recreation' category of Google and should not open to business per government rules. The list below shows all the top-level Foursquare categories. However some locations restricted by public orders might sit below the first level, and therefore has been taken into account of our mapping process.

```
0           Arts & Entertainment
1           College & University
2                          Event
3                           Food
4                   Nightlife Spot
5           Outdoors & Recreation
6    Professional & Other Places
7                      Residence
8                   Shop & Service
9             Travel & Transport
Name: name_1, dtype: object
```

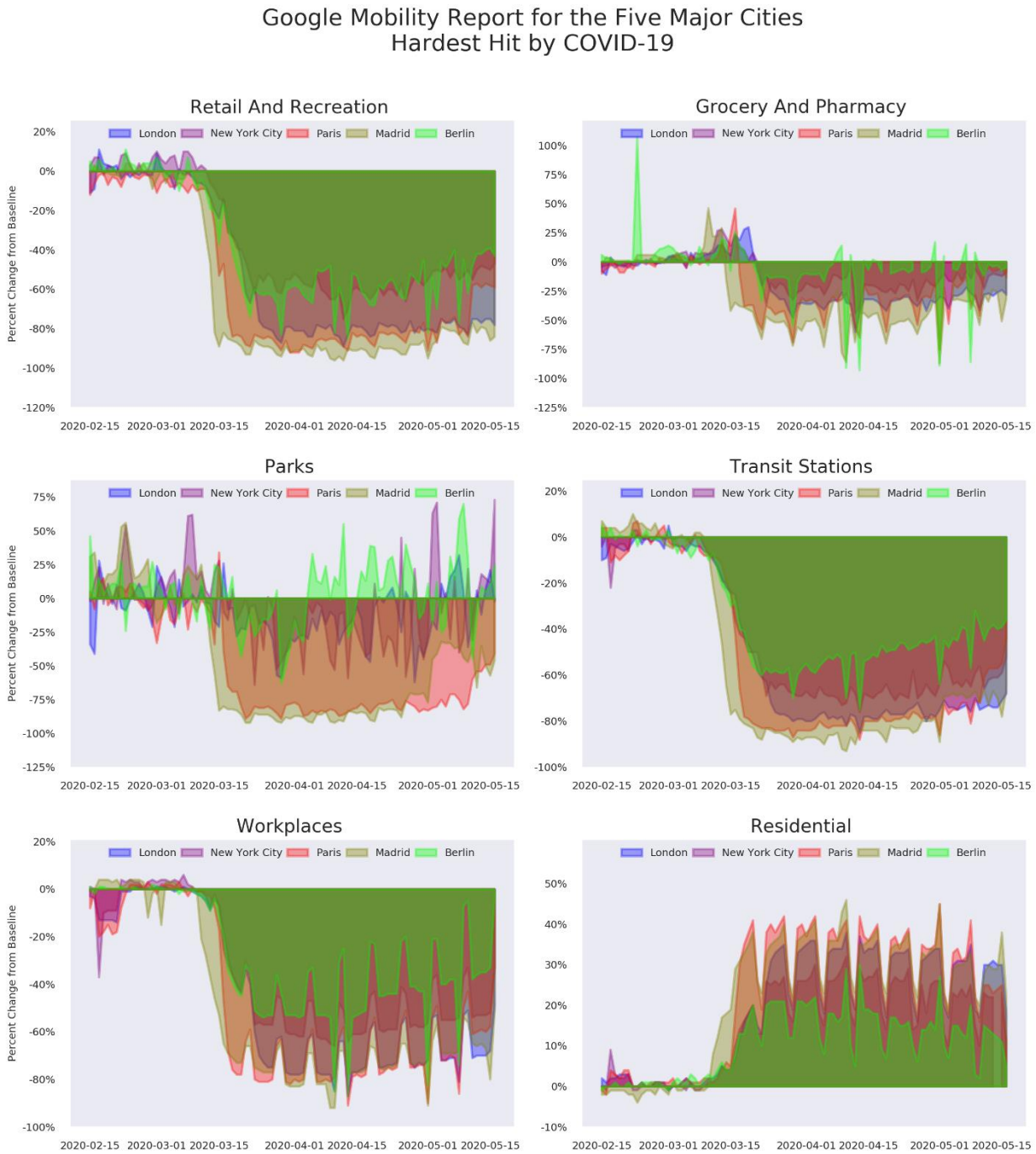
5. City Profiles for Movement Restriction Impacts

After categories have been mapped, the same method for top 10 city venues is applied here as well. Venues are counted per Google place categories along with places affected by local authority's restrictive measures. Next, the percentage shares for each city are compared against the average within each new category, as the visual below.



According to the plot, local venue constitution in Berlin has received the least impact from movement restrictions, while Parisian businesses are facing the hardest hit. School shut down affects New York the most, relatively to other four regions. Amongst the city profiles, London provides most outdoor places but with less recreational and retail venues. New York also has a high proportion of places for outdoors and parks and is the most recreational urban region. Paris does not stand out in any of these categories, but notably is the lowest in grocery and pharmacy, parks and outdoors, and workplaces categories. In Madrid, there are many workplaces and consists of the highest percentage of residential places. In Berlin, pharmacies and grocery shops can be found easily with the largest proportion of transportation links with relatively fewer retail & recreational venues and residential places.

6. Google Community Mobility Reports



Data Source: Google LLC "Google COVID-19 Community Mobility Reports"
(https://www.google.com/covid19/mobility/)
Accessed on: 2020-05-25 09:41:40

This part of graphical representation is an analogy of ‘Google COVID-19 Community Mobility Reports’ [7] but focuses on the five major cities seriously affected by the infectious illness. Around 90% of reduction in visits to non-essential places like parks, shops and restaurants, and public transports in Paris and Madrid indicates highest stringency is in place, followed by the two less strict regions, London and New York. In Berlin, German are not as confined by the government as much as other places. Several spikes at groceries and pharmacies surge in March signifies panic-buying occurring in Germany [23]. Interestingly, people on another side of the world were told ‘buy as much as possible’ – find out more about how Taiwanese authorities have acted differently from the news coverage cited in References section [24].

At the end of each category chart, we start seeing evidence of a few countries gradually resuming their normal life yet not in full swing. Berlin is the first city having its lockdown measure cautiouly eased since the 20 April. Spain and France are two of the strictest countries enforcing national confinements on their citizens, who were not even allowed to go to the parks. However, Paris has recently reopened small businesses, such as retail stores and shops, as well as the use of parks and public transports. More people are going back to their workplaces and staying less at home. Spain has not yet reopen non-essential shops, but people can stay in parks for longer than before since the beginning of May. Visits to transit stations have also slightly increased. In general, European countries have been implementing phased de-escalation of movement restrictions yet with no less concerns and debates on the way [25].

Clustering Model Development

Governments have been desperate to find out how to best manage the crisis from countries which are the most successful. However, are we comparing like for like or, in fact, apples with pears?

Now, there are 213 nations in our dataset, but only 139 with complete information remain in this part of analysis after rows of country with missing values being removed. These 139 nations are going to provide answer to one key question asked by the entire humanity: **'Who can global leaders really learn from in this historic crisis?'**

1. Modelling Feature Selection

Each nation has its distinct factors to consider when choosing the most effective but least costly responses. Wealthy country's approach might not be plausible for poor nation. Older country needs to focus on the elderly and has less reliance on the working age generation, so that it is less hit economically by national lockdown as opposed to younger nation. Therefore, we will segment countries based on similarities of demographics, healthcare development, population structure, and most importantly the stage of epidemic they are undergoing, and identify the most successful country within each cluster. The following list of variables are selected features for modelling. The meanings can be found in Appendix A. Data Dictionary.

1) population_2020
2) density_p_km2
3) migrants_net
4) med_age
5) urban_pop_%
6) world_share_%
7) life_exp_both_sex
8) life_exp_female
9) life_exp_male
10) old_dep_ratio_2020
11) hospital_beds_per_100k
12) physicians_per_k
13) gdp_capita_usd
14) gross_publicdebt_pct_of_gdp
15) days_1st_case
16) days_1st_death

2. Machine Learning Algorithm

Hierarchical Clustering and K-Means are the two most popular machine learning algorithm for segmentation. Each has its own pros and cons.

K-Means groups data points with the shortest distance to the centroids, an efficient algorithm suitable for large datasets. However, the centres need to be initialised at random with predefined number of clusters and then get constantly updated by the subsequent computations until convergence has been achieved. As a result, the clustering outcomes are not always identical each time when the algorithm is run given the same dataset used.

Hierarchical Clustering initiates with the number of data points to be the number of clusters; i.e. 139 for our dataset. The algorithm then merges different clusters based on the shortest distance between clusters. The computation process repeats until the assigned number of clusters has been reached or until no more clusters can be combined; i.e. all data points have been grouped into only one cluster. Unlike K-Means, this process takes longer runtime and therefore is not ideal for large datasets, but the result does not change as long as the same model has been run on the same dataset. As a result, this is chosen to be our clustering algorithm.

3. Data Normalisation

In statistical modelling, the huge differences in the value range of different features could cause bias to variables with large values. A remedy of this issue is to normalize modelling variables; that means transforming variables to the same value range. There are several feature scaling methods available such as Simple Feature Scaling, Standard Scaling (aka. Z-score Scaling) and Min-Max Feature Scaling. For the ease with setting measurement values in our data visualisation and proper modelling results, Min-Max Feature Scaling has been applied in our study. The following formula shows how it works:

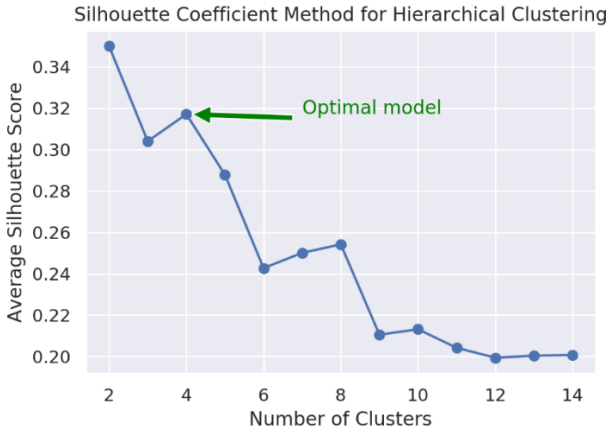
$$x_{std} = \frac{x_{old} - x_{min}}{x_{max} - x_{min}} \ , \ x_{std} \times (max - min) + min$$

Within each feature, the difference of each value x_{old} from the minimum x_{min} proportionate to the range between minimum x_{min} and maximum values x_{max}; this give us standardised values x_{std} between 0 and 1. Then this ratio scales up to the new range, defined by $max \ \& \ min$ here, within which all variables are converted to between 0 and 30 in our analysis.

4. Setting Number of Clusters

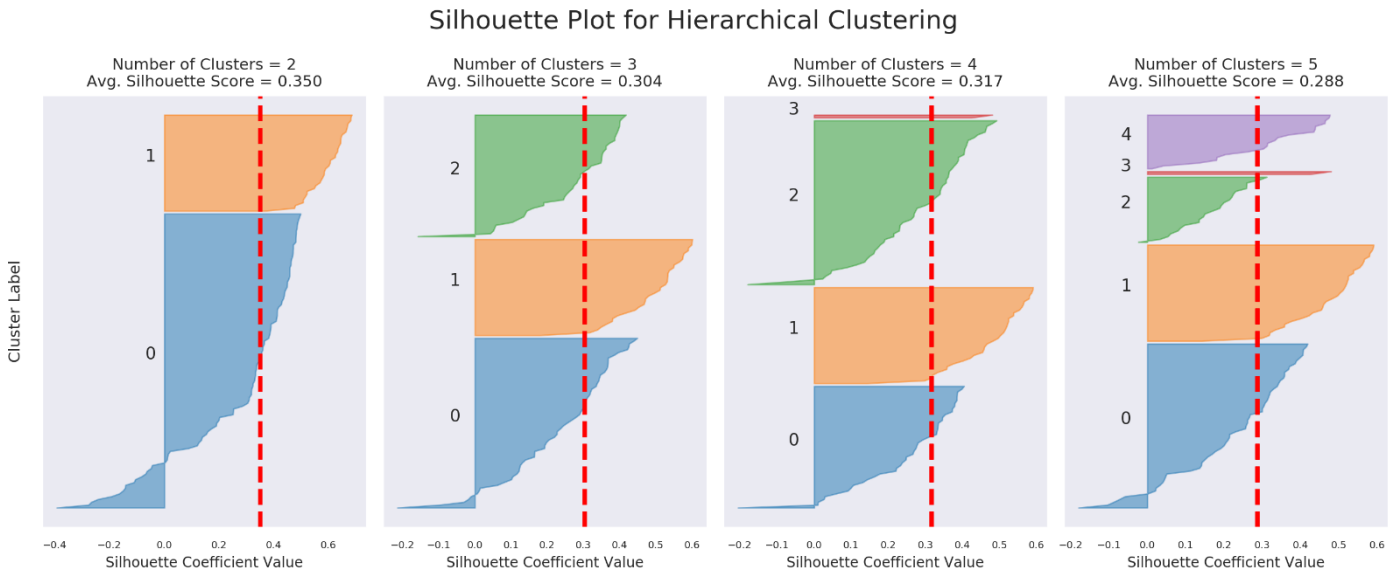
As mentioned before, we use hierarchical clustering algorithm to segment countries. This technique in theory does not require predefined number of clusters. However, the objective of this work is to provide customised resolutions for different circumstances in different nations, but how many clusters do we need?

The python script has experimented models with various number of clusters from 2 to 14 and the fitness of each model has been measured by Silhouette coefficients. Silhouette coefficient values range between -1 and 1 to indicate how close each datapoint is within its cluster as opposed to its neighbouring clusters [26]. Value 1 means that a point perfectly approximates to its assigned cluster and is perfectly distant from its neighbouring clusters, and -1 otherwise. Value nought indicates a point sitting just at the boundary between two adjacent clusters.



The figure on the left shows that the average Silhouette coefficient for the model with 4 clusters is the optimal, which strikes a balance between sufficient number of groups and decent fitness of clusters. However, this only gives us an overall view of different models without the visibility of proximity within each cluster of individual model, which is what is to be presented next.

The next graph represents Silhouette coefficient value for each data point within individual model in comparison with the model's average Silhouette coefficient. Model performance with more than 5 clusters are not included here as their average Silhouette coefficients only get worse.



We can see from the plot that the last model with 5 clusters not only has the lowest score in average but also has its data points in Cluster 2 mostly standing below the model average. One thing to note is that Cluster 3 which exists in the last two models is significantly slim, given y axis representing the number of datapoints. We will explore more on this group. Please note that the use of colour coding for identifying clusters in this chart will be consistent throughout this report.

All in all, this result gives us confidence to build our model with 4 clusters and ends our analysis for Methodology section. Next, the modelling outcome will be elaborated in Result section.

Results

Country Clusters

After computation has been completed by the hierarchical clustering model, the result prints out the counts as follows for these 4 clusters and tells us that amongst the 139 states the most fatal is in Cluster 0.

```
The country with the highest "deaths_per_m" is in Cluster 0.

Cluster counts:
2    59
0    44
1    35
3     2
dtype: int64
```

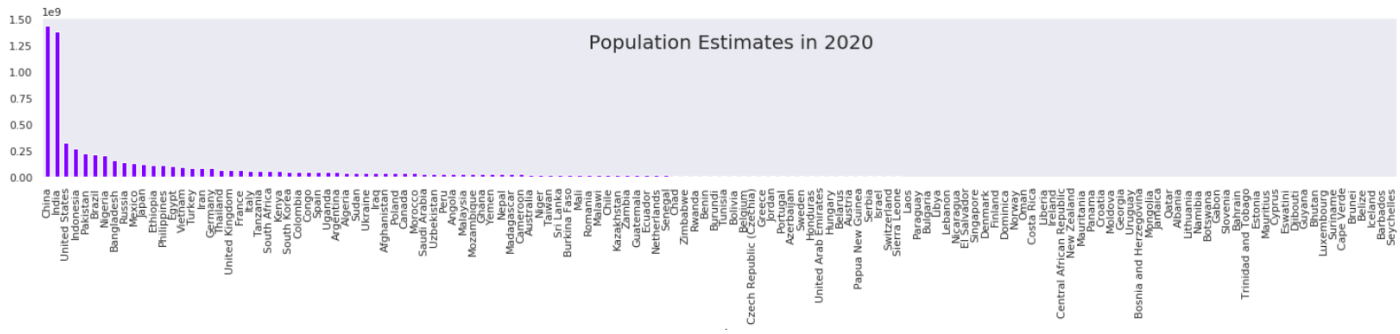
You must be able to recall the slim Cluster 3 we saw in the last two models from the previous section. Here we know that there are only 2 countries in this group.

The table below lists the most protected countries from each cluster in keeping the contagion at an impressively low level. (Please note that testing number has not been made available by Chinese authorities.) So now we know that China is part of Cluster 3, but who is the other member?

country	cluster	longitude	latitude	total_cases	total_deaths	total_recovered	total_tests	active_cases	serious_critical	cases_per_m	deaths_per_m	tests_per_m	tests_per_case	recovery_rate	critical_rate
Taiwan	0	120.992886	23.972586	441	7	415	70880	19	0	19.0	0.3	2977.0	160.725624	94.104308	0.000000
Laos	1	103.763406	18.502451	19	0	14	5795	5	0	3.0	0.0	798.0	305.000000	73.684211	0.000000
Bhutan	2	90.429433	27.415461	27	0	6	15607	21	0	35.0	0.0	20250.0	578.037037	22.222222	0.000000
India	3	79.618818	22.881321	144069	4117	58727	3033591	81225	8944	105.0	3.0	2200.0	21.056515	40.763107	11.011388

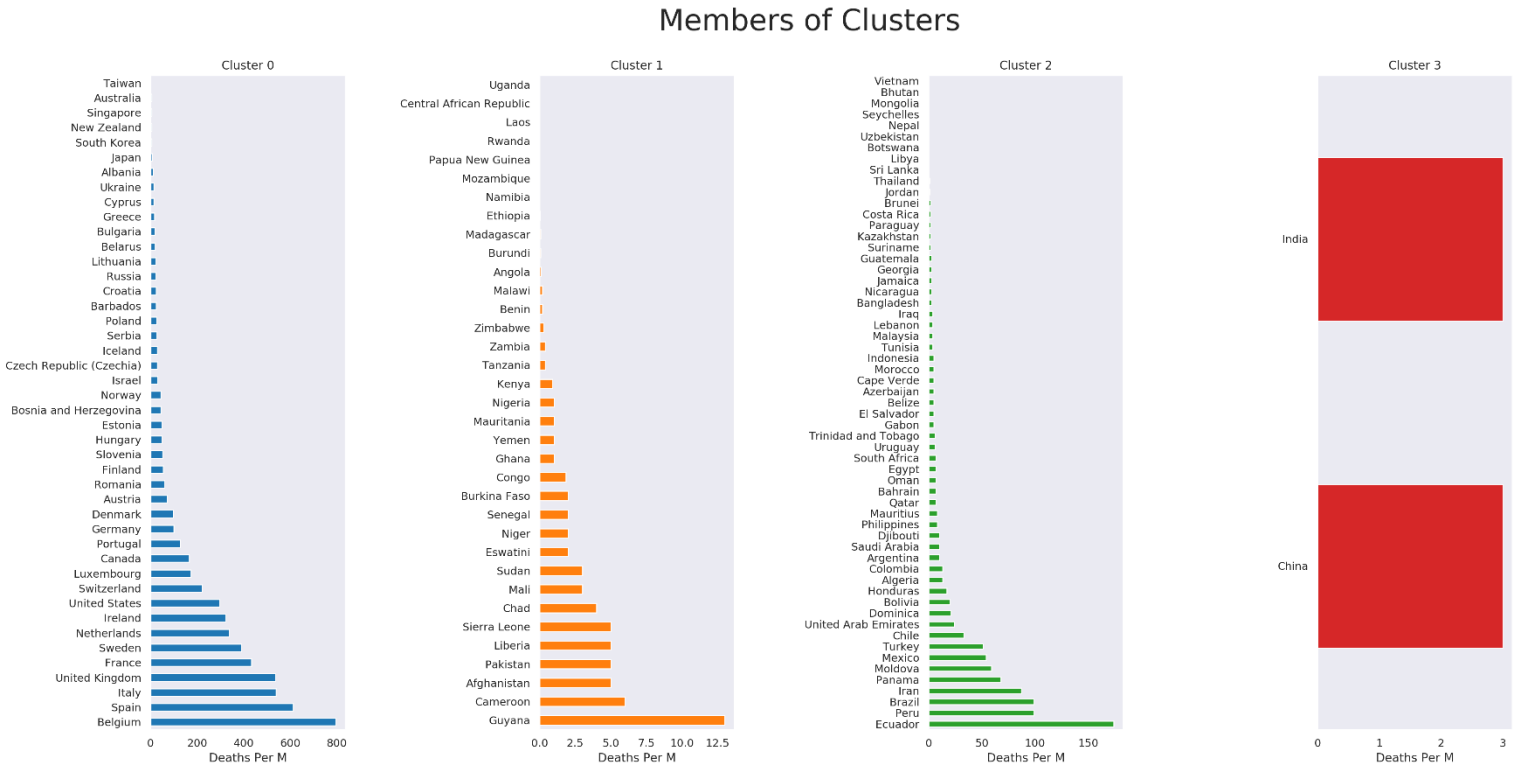
It is India, the second most populated country in the world, followed by the United States as ranked below but with just a quarter of India's population. Apparently, Indian government needs to take a completely different approach to deal with an emergency of 1.38 billion people as opposed to the rest of the world. Now let's shift our focus to the other clusters and see what are the associated member states within them.

country	cluster	longitude	latitude	total_cases	total_deaths	total_recovered	total_tests	active_cases	serious_critical	cases_per_m	deaths_per_m	tests_per_m	tests_per_case	recovery_rate	critical_rate
China	3	103.893473	36.518701	82971	4634	78258	0	79	9	58.0	3.0	0.0	0.000000	94.319702	11.392405
India	3	79.618818	22.881321	125149	3728	51824	2834798	69597	0	91.0	3.0	2056.0	22.651384	41.409839	0.000000

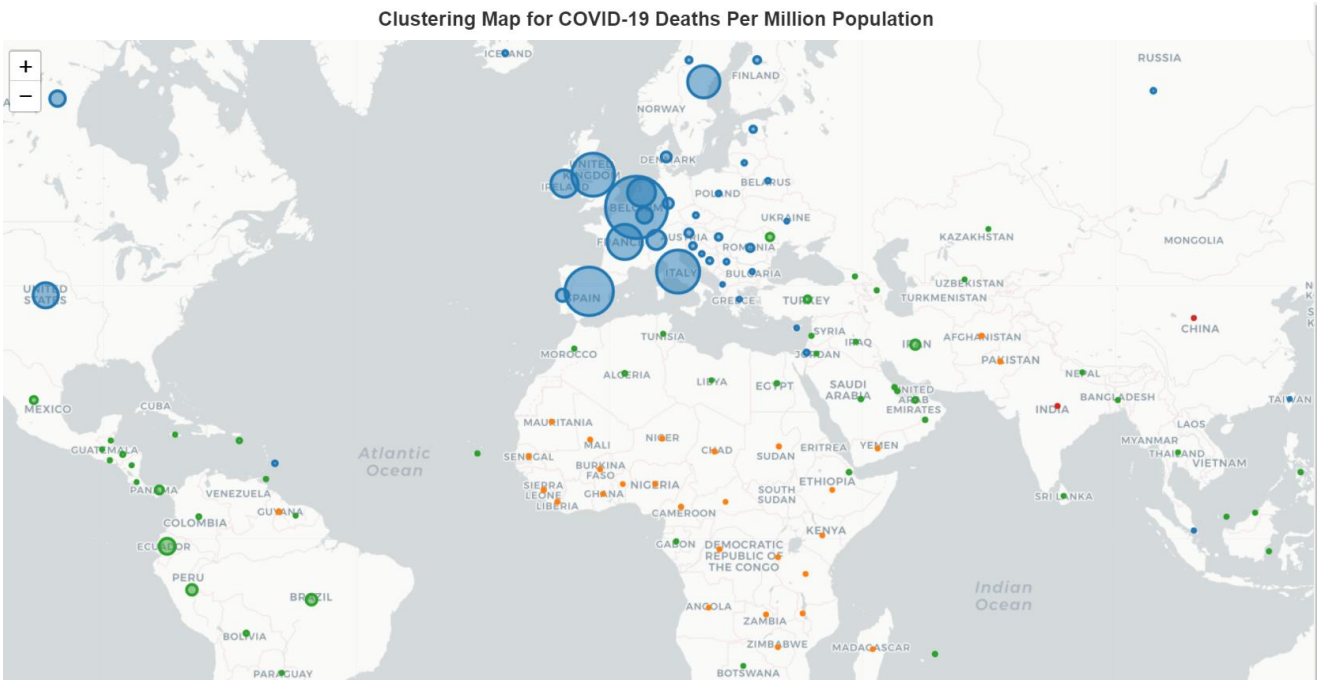


The following bar charts demonstrate countries in each cluster sorted by population deaths. As the model outcome we have seen earlier, the highest number of deaths is in Cluster 0, which is Belgium. This group primarily consists of European states and North America where the highest infection and death tolls so far have been reported, such as the

United States, the United Kingdom, Italy, France, Spain and so forth. The majority of African countries belong to Cluster 1. Cluster 2 mostly consists of countries in South America, Middle East, South and North Africa and numerous countries in Southern and Central Asia.



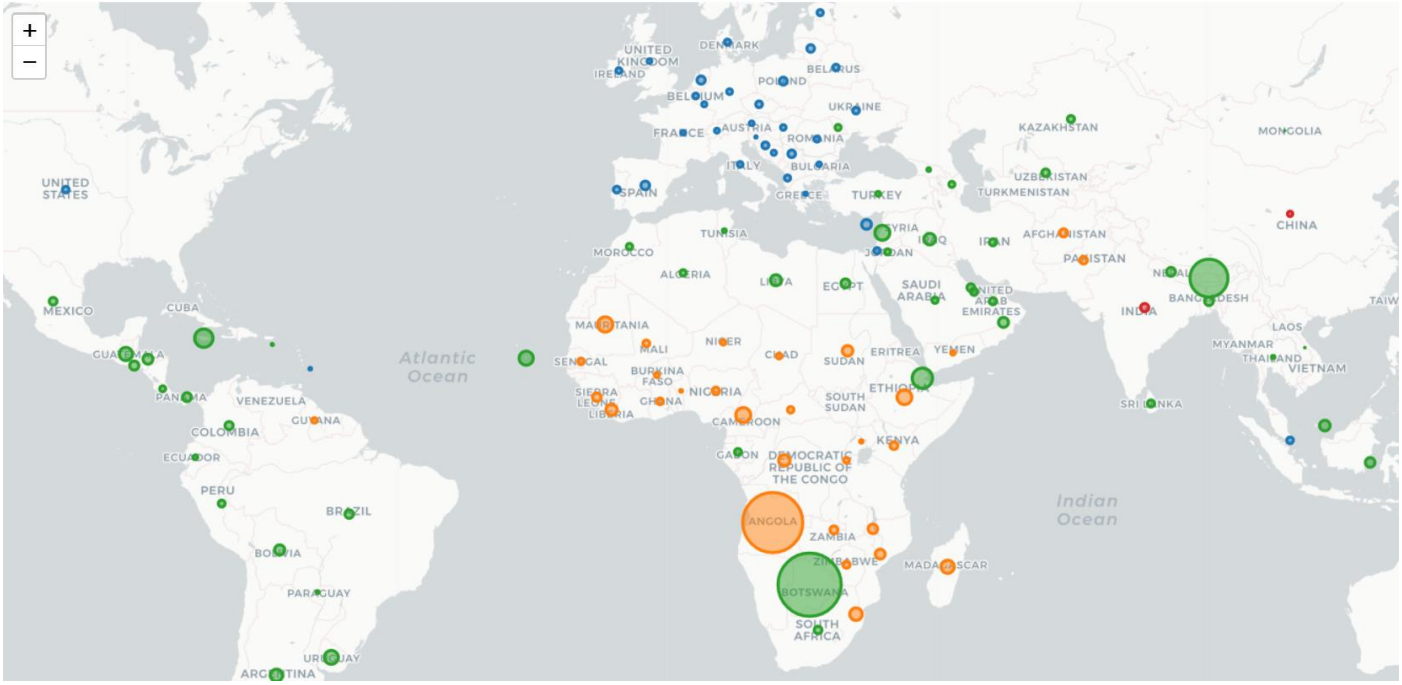
The map below created by folium package geographically presents the cluster distribution, and is showing the relative death counts, denoted by the size of circle marks. Countries in Cluster 0 are the deadliest. The highest number is from Belgium as of data retrieved on 26 May 2020, which stands at 801 deceased per million people.



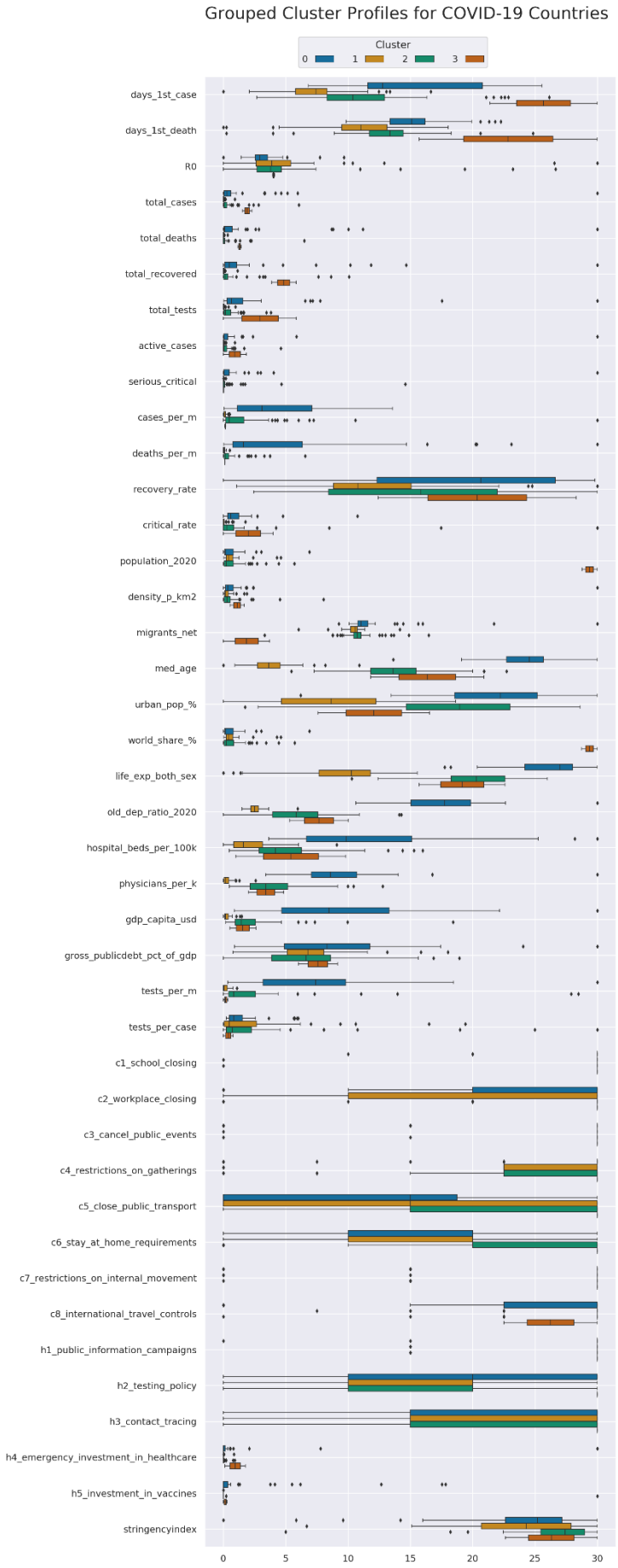
Another picture here displays several countries outside Europe are seeing the rise of transmission within the territories. Particularly for those less affluent countries in need of healthcare resources or constantly battered by civil wars in Africa and South America, are they ready to face the pandemic?

Potentially they can be guided from their peers in the same group to help them make the right decisions from the onset of the viral spread. In the next part of analysis result, we will dive into cluster profiles and find out what makes them distinct from one another.

Clustering Map for COVID-19 Reproduction Number



Cluster Profiles



This part of discussion breaks down each cluster into several attributes to describe countries in depth.

The grouped box plot we see on the left articulates current COVID-19 situations, the essential features of these nations, as well as government responses to address this crisis, in a top down sequence as plotted. However, details of government responses will be extended in a separate part and skipped here.

Variables used to develop the clustering model are those between 'population_2020' and 'gross_publicdebt_pct_of_gdp' on y axis in addition to the first two on the top, 'days_1st_case' and 'days_1st_death', which imply the stages of contagion which a country is currently undergoing and which entails different ways to combat the disease.

Several points to note on this plot. The values have been transformed into the range between 0 and 30 via prior data normalisation. As Cluster 3 contains only 2 countries and should not be generalized to other states, our discussion will primarily focus on the other 137 nations. To understand a box plot, a simple guide online has been provided in the corresponding reference [27] of this report. An ungrouped box plot is also available in Appendix B which allows clusters to be explored individually.

The epidemic initially began in China since December 2019. Then countries in Cluster 0 have also experienced a significantly long period of epidemic, followed by Cluster 2 and Cluster 1. However, both clusters have just started to see the spread of the virus over the last 2 months as more and more countries might be facing a series of fateful decisions that await. This is evident with the high reproduction numbers, with several worrying outliers with skyrocketing transmission within the

community. The good news is, death tolls in these areas are not as many as in Cluster 0 where this invisible enemy has horrendously overwhelmed these countries, careful steps taken at this stage [28] will be the game changer to fight against the disease.

As previously mentioned, most countries are more or less of similar population size and so are for the population density, yet urban districts in Cluster 0 have the highest percentage of people concentrated whereas Cluster 1 and Cluster 3 are less populous in cities. Urban areas are where citizen movements in public places are the most active and intense, that assists domestic spread of virus.

Due to the prevalence of globalisation, exchanging of commercial, professional, cultural, and recreational activities cross national borders are getting more and more frequent and easier. Foreigners flowing into Cluster 0 is the highest according to net migration metric, but is significantly lower in both China and India, which belong to Cluster 3. Arguably to control the growth of transmission, countries with large immigration volume need to reinforce virus detection at the borders and prevent the transmission circulating around the society.

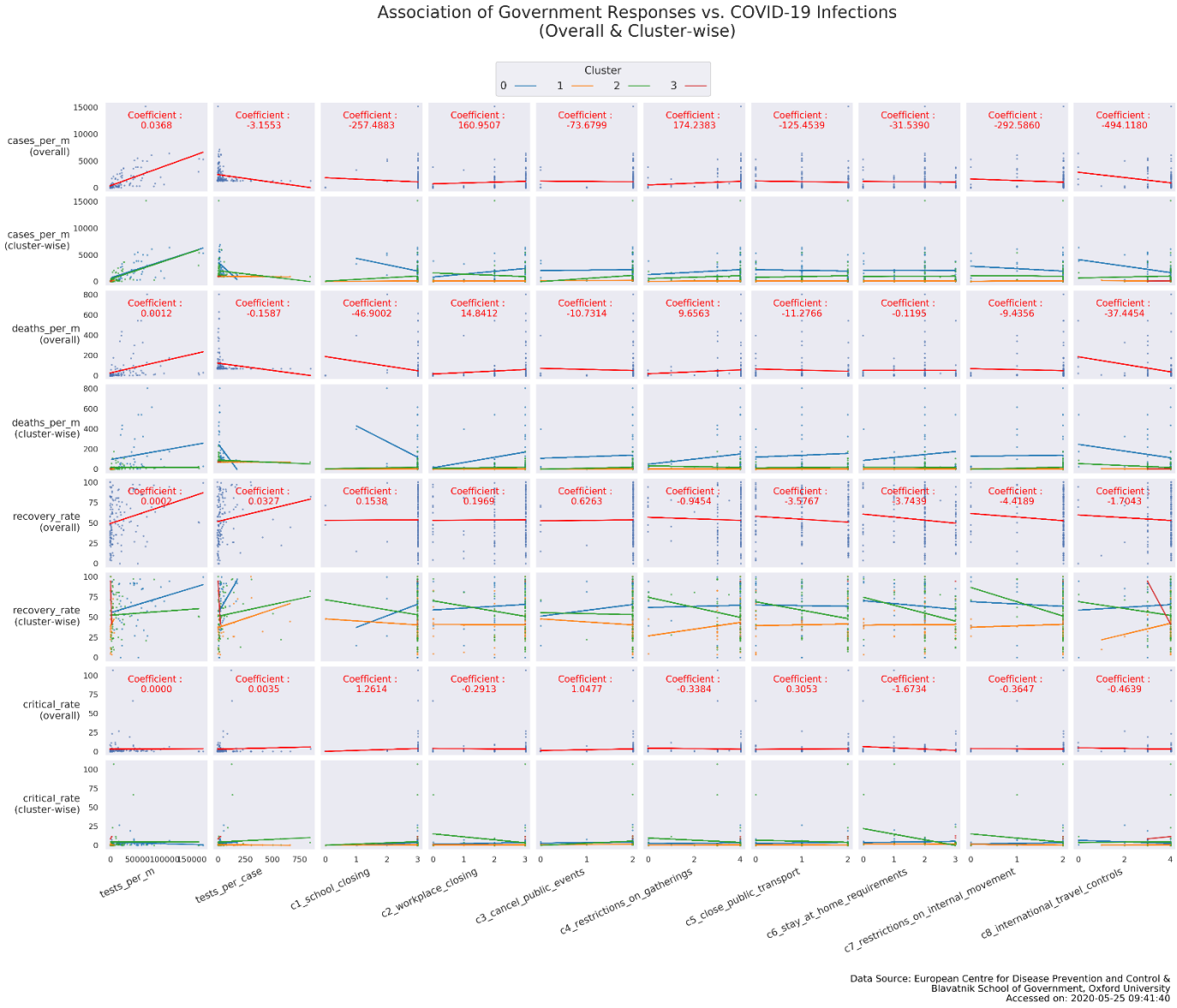
The demographics shows that Cluster 0 is significantly aged and the most elderly dependent. People there live much longer than the rest of the world. This cluster has been equipped with the most capacity in hospitals and the highest availability of medics for treating patients, but the range extends widely. As we knew from the media, health systems in several European countries have been exhausted during this pandemic. The condition is fine when life goes as normal but not in the times of extremely contagious epidemic like COVID-19, as critically ill patients require intensive care, closely monitoring the severity of symptoms, and respiratory aid equipments like ventilators. As infections scale up, people who require critical care services will soar proportionately.

From an economic perspective, Cluster 0 are predominantly affluent countries with highest GDP per capita but in the meantime have the highest share of national GDP in public debts. This signifies substantial emphasis of public funds and supports allocated to boost the momentum of economic and industrial developments before the crisis. Therefore, when the pandemic put a sharp and hard brake on all businesses, an economic tsunami very rapidly catches all private and public institutions unprepared and hits extremely fatally.

Now let's take a look at Cluster 1 where places are still in the early stage of infections. It is a much younger group with a lot fewer elderly relying on the younger generation. Citizens tend to more widely reside across the regions than in cities. Economy is least developed with all nations living with low GDP. Lifespan is the shortest in the world, with the worst level of medical services and healthcare personnel to treat diseases. Therefore, this is a relatively deprived corner where an outbreak could devastate relentlessly with very limited counter-actions in hand.

Yet, we learned that a large number of deaths linked to COVID-19 concentrate in the oldest cluster where the highest urban density resides. This is statistically evident in countries which have gone pass the (first) peak of the disease. Nonetheless, this should not be an excuse to leave these nations unguarded, because many other diseases have been poorly or even not treated and could become underlying factors causing complications and leading to deaths from COVID-19.

Government Responses



In this part of the analysis results, government responses are broken down into clusters in comparison with overall effects to all countries. We use scatter plots for our exploration. The coefficient is the slope of regression line (fitted by Least Squares approach) on the chart and represents the unit of influence to infections for every unit of government response escalation. A simple guide to understand linear regression in statistics can be found in the corresponding reference provided in this report [29]. Be wary of comparing coefficient of test volume with other government responses, due to perception bias which potentially occurs due to the huge difference in measurement units. The underlying data is untransformed at country level. Metrics on y axis reflect the latest statistics as of the data accessed date stated under the graph, whereas those on x axis except for tests carried out are the highest level of severity ever imposed within a country and therefore might not be current status. As regression coefficients explain association in data rather than causation, basic understanding about this communicable disease [10] should also be applied on our interpretation.

Generally speaking, we can see that countries with large number of infection testing implemented are associated with higher number cases per capita, because a generalized detection approach allows more infections to be discovered and acted upon quickly. Conversely, higher number of tests performed per case is linked to countries with lower population infections when transmission is trackable. Both strategies are seeing positive moves on recovery rate much to do with early actions on the transmission allowing health services to be able to treat patients without being overwhelmed. Patterns are consistent across clusters as well.

Role models for these two different strategies are South Korea and Taiwan. Both have conducted rigorous detection from the start of the outbreak to allow every case to be traceable and isolated. However, in South Korea the 31st case unfolded a large scale of clustering infections within the community due to the patient's attendance to several public gatherings densely seated thousands of people. It was a series of decisive actions taken, including rapidly ramping up testing and medical capacity, to successfully turn the corner after a surge of demands in tracing the virus as well as patients in hospitals [30]. No safety net is perfect. Despite a few community transmissions were inevitable, the fact that Taiwanese authorities have enforced requirements of wearing surgical masks in public venues since the start of spread in the nation has played a crucial role to break the chain of infections [18].

When it comes to government restrictions to public activities, to our surprise national blanket lockdown, represented by 'c6_stay_at_home_requirements', is not showing impressive effects on taming the disease despite the level of escalation. In fact, this is an approach with least association with alleviating the epidemic. As more and more statistics are out for unemployment claims, job furloughs, business rescue applications in recent days, an increasingly hot debates have been on the table questioning the moves and justifications of policymakers. Indeed, this is a life and death decision – in public health, in people's livelihood, in national development, in the future of our next generation, and perhaps, in politician's making of their political history and future [31]. Nevertheless, this emergency has swept through the world too vastly and too quickly and has cruelly caught our country leaders unprepared. Hardly anyone could have made instant calculations with great precision, and now we have seen it.

Our analysis results finish here. Further observations and recommendations come next.

Discussion

Economic Catastrophe

According to a medical journal published on 30 March 2020 for a comprehensive review of the severity of coronavirus outbreak in China, there is in average 1.38 loss of lives for every 100 infections [32]. So, the next big question should be asked: **What is our loss of economy and livelihood?**

Around the world, records of government's fiscal spending [33], borrowing, jobless figures, business bailouts and bankruptcy, have still been continuously climbing at an 'R0' rate much higher than the virus spread – to a record high never seen in our lifetime!

Amid a row about China's plan to introduce new security law for Hong Kong, the economic powerhouse in the east has made its latest announcement on 22 May to drop its GDP forecast for this year, pledging further and deeper economic stimulus to restore the aftermath of COVID-19 [34]. Before that, the world's second largest economy saw its GDP contracted by 6.8% during the first quarter of the year [35]. The largest world economy, the United States, has also published its figure for the same period, which has declined to a 4.8% annual rate but the surge of unemployment and large scale of economic fallout did not emerge until late March. Economists and experts expect that the worst is still to come [36].

In the EU where free movement of people, goods, services and capital is the centre of the union's principle, GDP is 2.2 % down in Germany [37], whereas France, Italy, and Spain have also gone into recession showing GDP figures well below 3.8% fall in the Eurozone. Even in the least strictly confined country, Sweden has its first quarter GDP lowered by 0.3%, but the country's central bank has estimated its GDP this year will drop by 7-10 per cent, with unemployment peaking at between 9 and 10.4 per cent [38]. So how bad the situation will be? Some experts believe that the world economy is set to plummet by 5.5% this year [39], while some see the loss of global output in between 6.4% and 9.7% [40]. In other words, that is more than triple of the damage in 2008 financial crisis.

Although there have been projections on various shapes of recovery for post pandemic economy [41], the trajectory will depend not only on government's stringency to address the disease but also heavily on consumer confidence and purchasing power, despite the lifting of lockdown orders. Data has revealed drops of optimism in consumer sentiments from the vast majority of world economies [42], signifying weakening strength and outlook on the demand side. Moreover, in a deeply and broadly globalised era, no country can stay intact from this economic fallout if actions of world leaders have not been taken in an integrated and unanimous fashion. In Asia, reopened sectors which are mostly reliant on international movements are still struggling with empty flight seats, restaurant tables, hotel rooms due to shortage of customer flow [43].

Onward and Upward

All we have discussed so far have now been in history. We need to look ahead and know which tool is the best to pick from our lessons learned. As of now, some countries have already been through their worst situations, some are going back to the old status again and perhaps even worse, still some are just struggling with their mounting threats and challenges. No matter in what stage they are currently, the same mistakes cannot repeat.

Bill Gates has a fantastic and comprehensive post about several scientific advances required for this pandemic [44], which he has already given warning to the world in his TED talk in 2015. He stressed on the vital focus required from the all countries on the five innovations to defeat this invisible enemy: treatments, vaccines, testing, contact tracing, and policies for opening up. The first two categories, as we know, are still in an unprecedentedly 'fast-tracking' progress. For the last three, they are paramount to reopening our economy in a safe and sustainable manner.

All in all, these are what governments need to prioritise on their agenda: rationing and expanding the capacity and medical resource to key persons, deploying technology and data analytics to get a grip on the transmission trajectory, devising policies which are both sensible and sensitive to exert effective influence on people's behaviours to respect public efforts to fight against the epidemic.

Another spectrum that has attracted fierce international debates is Sweden's strategy [45]. Swedish authorities count on Swedes respecting social distancing rules without ordering people staying at home and shutting businesses around the country, with only restrictions on public gatherings and international travel ban in force. Criticisms of its exceptionalism express concerns of wiping out the lockdown efforts from neighbouring nations to contain the contagion as the European continent is anxious and cautious about returning to normality [46]. Suspicion of the controversial concept of 'herd immunity' behind Swedish state epidemiologist Dr. Anders Tegnell's calculated move has been denied [45].

Nonetheless, given the unmeasurable damage to the world economy and the questionable efficacy of the Great Lockdown [47], herd immunity might have actually been happening in some regions with population infection rate higher than Sweden. Although expert's view on the economic loss of Sweden this year is not much different from the global average, we may want to find out more for Swedish government's post-pandemic debt level, an indicator of how bright the future of a country's next generation is like.

Before moving to our final section, here ends our discussion of this analysis and we will draw some conclusions next.

Conclusion

No evidence has been found that shutting down the world could do the magic of effectively halting the disease spreading, if not causing more harmful problems for decades to come. Similar picture has also been outlined by Bloomberg's data journalist with the latest published results [48]. The WHO's role and performance of leading the world to fight against this pandemic will be left with the upcoming independent investigation called for by its 192 member states [49].

Crucially, collective efforts and globally integrated collaboration are key to supporting countries around the world to quickly restore health in all aspects from this crisis. Virus detection at the early stage of infection is essential to curb it spreading rather than paralysing general movements of people, goods and resources, which has also limited support access from abroad.

However, no single research is able to fully cover all the details that are needed for all sorts of circumstances. Particularly this is absolutely an ambitious endeavour, trying to obtain a comprehensive understanding of an unprecedented event never happening in the lifetime of most of us. Data keep generating every day, but we still need more information to help us examine more about the following areas:

1. **Mortality rate in existing health conditions, age, ethnicity, care home, public health insurance cost per capita:** Data has not been widely available and shared internationally, although such information has been published in a few nations. A global view will allow us to apply the knowledge and approach to different situation.
2. **Historic death rate in all causes, smoking, suicide, pneumonia, flu, national death, before and during last financial crisis and in comparison with COVID-19:** Direct impacts from the pandemic have been broadly assessed by the press and experts, but indirect consequences cannot be overlooked and require policymakers to address the risks.
3. **Infection testing:** It was decided not to include this piece of information in our analysis, as the approach of test counting largely varies in different countries. Same person could be tested for more than 6 times. In some regions the number of tested persons is counted whilst in some the number of tests carried out has been recorded. This will require extra efforts to classify them and computation to standardise testing units.
4. **Situation in countries with existing humanitarian crisis (such as severely battered by civil wars and living under poverty line):** Many countries on the planet have already been suffering severe humanitarian crisis due to violent conflicts or poverty, but virus does not discriminate borders and ethnics. These are areas most in need and short of even basic supplies that are desperate for global assistance, given that life in other family is equally important as in our own.
5. **Economic impact to different industrial structures, government rescue packages and public debt level:** Movement restrictions not just take place domestically but also internationally. Nations with heavy reliance of global trade and weak domestic demand or output might get the hardest hit when internationally social distancing has been put in place. On top of that, how public spending and burden has soared following the crisis

definitely will cause no less impact than GDP losses to our future and deserves further study.

6. **Implication to climate change:**

   COVID-19 has changed the way we live, work, and learn, perhaps permanently. How we address the threat of further infections as part of global recovery should be put in the context of climate crisis [50], which unfortunately might not have any earthly solution which any world leaders can provide. By then, the loss of human lives will be so many times of what COVID-19 has caused.

In a nutshell, we cannot stress more that it is of paramount importance for world leaders to pull together in solidarity, sharing their success stories with their neighbours of similar experience and nature as well as stretching out their helping hands to where they are most and desperately needed. In a time of humanity crisis, protectionism and nationalism could only make individual countries less protected and nationally weaker with immediate effect. Key takeaway from this pandemic for leaders of all institutions and countries is to lead with vision of future and humanity, which is what this study aims to contribute and inspire.

Here ends the entire project, which is a submission for IBM Data Science Professional Certificate on COURSERA, dedicated for everyone on this planet, for the first global pandemic in modern history – **COVID-19**.

References

[1] “COVID-19 pandemic,” The Wikimedia Foundation, Inc., 14 05 2020. [Online]. Available: https://en.wikipedia.org/wiki/COVID-19_pandemic. [Accessed 20 05 2020].

[2] P. Beech, "COVID-19: what is the R number?," World Economic Forum, 08 05 2020. [Online]. Available: https://www.weforum.org/agenda/2020/05/covid-19-what-is-the-r-number/. [Accessed 20 05 2020].

[3] A. Jones, "Coronavirus: Should the world worry about Singapore's virus surge?," BBC News, Singapore, 10 04 2020. [Online]. Available: https://www.bbc.com/news/world-asia-52232147. [Accessed 20 05 2020].

[4] K. R. Lai, "Why Coronavirus Cases Have Spiked in Hong Kong, Singapore and Taiwan," The New York Times, 09 04 2020. [Online]. Available: https://www.nytimes.com/interactive/2020/04/09/world/asia/coronavirus-hong-kong-singapore-taiwan.html. [Accessed 20 05 2020].

[5] L. Z. L. G. R. L. a. R. W. Se Young Lee, "China's Wuhan reports first coronavirus cluster since lockdown lifted," Reuters, 11 05 2020. [Online]. Available: https://uk.reuters.com/article/uk-health-coronavirus-china-toll/chinas-wuhan-reports-first-coronavirus-cluster-since-lockdown-lifted-idUKKBN22N02H. [Accessed 20 05 2020].

[6] T. S. W. A. P. T. P. a. B. K. (. Hale, "Oxford COVID-19 Government Response Tracker," Blavatnik School of Government, Oxford University, 2020. [Online]. Available: https://covidtracker.bsg.ox.ac.uk/. [Accessed 20 05 2020].

[7] "Google COVID-19 Community Mobility Reports," Google LLC, 12 05 2020. [Online]. Available: https://www.google.com/covid19/mobility/. [Accessed 20 05 2020].

[8] "Foursquare Places API," Foursquare, [Online]. Available: https://developer.foursquare.com/docs/places-api/. [Accessed 20 05 2020].

[9] H. F. Sophia Wagner, "What's the reproduction number R?," Deutsche Welle, 24 04 2020. [Online]. Available: https://www.dw.com/en/whats-the-reproduction-number-r/av-53234319. [Accessed 20 05 2020].

[10] "Q&A on coronaviruses (COVID-19)," The World Health Organisation, 04 05 2020. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-coronaviruses. [Accessed 20 05 2020].

[11] M. J. H. B. S. o. P. H. 6. E. P. S. B. M. 2. (. Thomas V. Inglesby, "Public Health Measures and the Reproduction Number of SARS-CoV-2," American Medical Association, 01 05 2020. [Online]. Available: https://jamanetwork.com/journals/jama/fullarticle/2765665. [Accessed 20 05 2020].

[12] C. M. &. A. Reuben, "Coronavirus: Why are international comparisons difficult?," BBC, 11 05 2020. [Online]. Available: https://www.bbc.com/news/52311014. [Accessed 20 05 2020].

[13] "COVID-19 pandemic by country and territory," Wikipedia, 2020. [Online]. Available: https://en.wikipedia.org/wiki/COVID-19_pandemic_by_country_and_territory. [Accessed 20 05 2020].

[14] H. R. E. O.-O. a. J. H. Max Roser, "Why are there delays in death reports?," Our World In Data, 15 05 2020. [Online]. Available: https://ourworldindata.org/covid-deaths#why-are-there-delays-in-death-reports.

[Accessed 20 05 2020].

[15] "National responses to the COVID-19 pandemic," Wikipedia, 2020. [Online]. Available: https://en.wikipedia.org/wiki/National_responses_to_the_COVID-19_pandemic. [Accessed 20 05 2020].

[16] H. Smith, "How Greece is beating coronavirus despite a decade of debt," The Guardian, 14 04 2020. [Online]. Available: https://www.theguardian.com/world/2020/apr/14/how-greece-is-beating-coronavirus-despite-a-decade-of-debt?CMP=fb_gu&utm_medium=Social&utm_source=Facebook&fbclid=IwAR00aux6cXXOSi2LN_3iJfQ5gvAikX4PkFASXvdiBqPohIeqexIE6-6E9Xg#Echobox=1586844440. [Accessed 20 05 2020].

[17] A. M. a. G. Zampano, "COVID-19: What went wrong in Italy and Spain?," Anadolu Agency, 08 04 2020. [Online]. Available: https://www.aa.com.tr/en/europe/covid-19-what-went-wrong-in-italy-and-spain/1797461. [Accessed 20 05 2020].

[18] W. Yang, "How has Taiwan kept its coronavirus infection rate so low?," Deutsche Welle, 09 04 2020. [Online]. Available: https://www.dw.com/en/taiwan-coronavirus/a-52724523?fbclid=IwAR3C_FCuxRr7H64307E1PTK8QGbNcLkhPtAP05WrZxg2_7yL_FSB2-vyDF4. [Accessed 22 05 2020].

[19] A. Jones, "Coronavirus: How 'overreaction' made Vietnam a virus success," BBC News, 15 05 2020. [Online]. Available: https://www.bbc.com/news/world-asia-52628283. [Accessed 20 05 2020].

[20] "Foursquare Places API Reference - Venue Recommendations," Foursquare, 2020. [Online]. Available: https://developer.foursquare.com/docs/api-reference/venues/explore/. [Accessed 20 05 2020].

[21] "Mobility Report CSV Documentation," Google LLC, 2020. [Online]. Available: https://www.google.com/covid19/mobility/data_documentation.html?hl=en. [Accessed 20 05 2020].

[22] "Foursquare Venue Category Hierarchy," Foursquare, 2020. [Online]. Available: https://developer.foursquare.com/docs/build-with-foursquare/categories/. [Accessed 20 05 2020].

[23] E. K. Besson, "COVID-19 (coronavirus): Panic buying and its impact on global health supply chains," The World Bank Group, 28 04 2020. [Online]. Available: https://blogs.worldbank.org/health/covid-19-coronavirus-panic-buying-and-its-impact-global-health-supply-chains. [Accessed 20 05 220].

[24] S. C.-M. Ben Blanchard, "'Buy as much as possible' - Taiwan sees boon to panic buying," Reuters, 20 03 2020. [Online]. Available: https://www.reuters.com/article/us-healthcare-coronavirus-taiwan-premier/buy-as-much-as-possible-taiwan-sees-boon-to-panic-buying-idUSKBN2170CF. [Accessed 20 05 2020].

[25] "Coronavirus: How lockdown is being lifted across Europe," BBC News, 22 05 2020. [Online]. Available: https://www.bbc.com/news/explainers-52575313. [Accessed 23 05 2020].

[26] "Selecting the number of clusters with silhouette analysis on KMeans clustering," Scikit Learn, [Online]. Available: https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html. [Accessed 20 05 2020].

[27] "Understanding and interpreting box plots," wellbeing@school, [Online]. Available: https://www.wellbeingatschool.org.nz/information-sheet/understanding-and-interpreting-box-plots. [Accessed 21 05 2020].

[28] C. Sibthorpe, "Coronavirus: What are the four stages of the UK's response plan?," Sky News, 12 03 2020. [Online]. Available: https://news.sky.com/story/coronavirus-what-are-the-four-stages-of-the-uks-response-

plan-11950264. [Accessed 21 05 2020].

[29] C. Taylor, "The Slope of the Regression Line and the Correlation Coefficient," ThoughtCo (part of the Dotdash publishing family), 06 02 2020. [Online]. Available: https://www.thoughtco.com/slope-of-regression-line-3126232. [Accessed 22 05 2020].

[30] L. Bicker, "Coronavirus in South Korea: How 'trace, test and treat' may be saving lives," BBC News, 12 05 2020. [Online]. Available: https://www.bbc.com/news/world-asia-51836898. [Accessed 22 05 2020].

[31] M. H. &. L. Chadwick, "Coronavirus: Why did European leaders' approval ratings rise during lockdown?," euronews, 21 05 2020. [Online]. Available: https://www.euronews.com/2020/05/21/coronavirus-why-did-european-leaders-approval-ratings-rise-during-lockdown. [Accessed 22 05 2020].

[32] L. C. O. I. D. P. W. C. W. N. I. G. C.-D. e. a. Robert Verity, "Estimates of the severity of coronavirus disease 2019: a model-based analysis," *The Lancet,* p. 9, 2020.

[33] W. R. L. a. E. T. Bryn Battersby, "Global fiscal support during the COVID pandemic, in one chart," International Monetary Fund, 20 05 2020. [Online]. Available: https://www.weforum.org/agenda/2020/05/9-trillion-global-fiscal-support-covid-19/. [Accessed 24 05 2020].

[34] J. H. Kevin Yao, "China drops mention of GDP goal as parliament opens, virus slams economy," Reuters, 22 05 2020. [Online]. Available: https://www.reuters.com/article/us-china-parliament-opening/china-drops-mention-of-gdp-goal-as-parliament-opens-virus-slams-economy-idUSKBN22Y05K. [Accessed 22 05 2020].

[35] K. Vaswani, "China GDP: Bleak outlook for economic recovery post-virus," BBC News, 17 04 2020. [Online]. Available: https://www.bbc.com/news/business-52305259. [Accessed 22 05 2020].

[36] B. Casselman, "Worst Economy in a Decade. What's Next? 'Worst in Our Lifetime.'," The New York Times, 29 04 2020. [Online]. Available: https://www.nytimes.com/2020/04/29/business/economy/us-gdp.html. [Accessed 23 05 2020].

[37] "Coronavirus lockdown pushes Germany into recession," Deutsche Welle, 15 05 2020. [Online]. Available: https://www.dw.com/en/coronavirus-lockdown-pushes-germany-into-recession/a-53446866. [Accessed 23 05 2020].

[38] R. Milne, "Sweden unlikely to feel economic benefit of no-lockdown approach | Free to read," The Financial Times Ltd., 10 05 2020. [Online]. Available: https://www.ft.com/content/93105160-dcb4-4721-9e58-a7b262cd4b6e. [Accessed 22 05 2020].

[39] A. D. &. AP, "Coronavirus: how much is the world economy expected to shrink?," euronews, 20 05 2020. [Online]. Available: https://www.euronews.com/2020/05/20/coronavirus-how-much-is-the-world-economy-expected-to-shrink. [Accessed 23 05 2020].

[40] "Coronavirus 'could cost global economy $8.8tn' says ADB," BBC News, 15 05 2020. [Online]. Available: https://www.bbc.com/news/business-52671992. [Accessed 23 05 020].

[41] P. Hoskins, "Coronavirus: What shape will the recession be?," BBC News, 10 05 2020. [Online]. Available: https://www.bbc.com/news/business-52450854. [Accessed 24 05 2020].

[42] K. Jones, "These charts show how COVID-19 has changed consumer spending around the world," World Economic Forum, 02 05 2020. [Online]. Available: https://www.weforum.org/agenda/2020/05/coronavirus-covid19-consumers-shopping-goods-economics-industry/. [Accessed 24 05 2020].

[43] D. Moss, "This Recovery Is Doomed Before It Even Begins," Bloomberg, 22 05 2020. [Online]. Available:

https://www.bloomberg.com/opinion/articles/2020-05-21/coronavirus-cautious-reopenings-mean-weak-economic-recoveries. [Accessed 24 05 2020].

[44] B. Gates, "The scientific advances we need to stop COVID-19," Bill & Melinda Gates Foundation, 24 04 2020. [Online]. Available: https://www.linkedin.com/pulse/scientific-advances-we-need-stop-covid-19-bill-gates/. [Accessed 24 05 2020].

[45] M. Savage, "Coronavirus: Has Sweden got its science right?," BBC News, 25 04 2020. [Online]. Available: https://www.bbc.com/news/world-europe-52395866. [Accessed 24 05 2020].

[46] D. T. H. Kim, "Sweden's coronavirus exceptionalism will not be remembered favourably by Europe | View," euronews, 29 04 2020. [Online]. Available: https://www.euronews.com/2020/04/27/sweden-s-covid-19-exceptionalism-will-not-be-remembered-favourably-by-europe-view. [Accessed 24 05 2020].

[47] G. Gopinath, "The Great Lockdown: Worst Economic Downturn Since the Great Depression," International Monetary Fund, 14 04 2020. [Online]. Available: https://blogs.imf.org/2020/04/14/the-great-lockdown-worst-economic-downturn-since-the-great-depression/. [Accessed 24 05 2020].

[48] E. He, "The Results of Europe's Lockdown Experiment Are In," Bloomberg, 20 05 2020. [Online]. Available: https://www.bloomberg.com/graphics/2020-opinion-coronavirus-europe-lockdown-excess-deaths-recession/. [Accessed 24 05 2020].

[49] "WHO members agree to independent probe of coronavirus response," Deutsche Welle, 19 05 2020. [Online]. Available: https://www.dw.com/en/who-members-agree-to-independent-probe-of-coronavirus-response/a-53496467. [Accessed 23 05 2020].

[50] "Earth Day: UN chief warns world is facing a bigger threat than coronavirus," Sky News, 22 04 2020. [Online]. Available: https://news.sky.com/story/earth-day-un-chief-calls-for-green-recovery-from-coronavirus-pandemic-11976836. [Accessed 24 05 2020].

Appendices

Appendix A. Data Dictionary

1) daily_covid_govresponse_df

Variable Name	Variable Description	Data Source
country	Country names cleaned and standardised from multiple data sources	Derived
date_reported	Dates when cases were registered	European Centre for Disease Prevention and Control: https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide
day_reported	Day element of *date_reported*	Ditto
month_reported	Month element of *date_reported*	Ditto
year_reported	Year element of *date_reported*	Ditto
covid_cases	Number of registered cases on the corresponding date	Ditto
covid_deaths	Number of registered deaths related to COVID-19 on the corresponding date	Ditto
first_case_date	Date of first registered case	Derived
first_death_date	Date of first death	Derived
days_1st_case	Number of days after first registered case	Derived
days_1st_death	Number of days after first death	Derived
R0	Reproduction rate of infections, calculated from the ratio of moving average of registered cases of the last 3 days (aka. ma3) over the ma3 value for 14 days ago	Derived
c1_school_closing	Record closing of schools and universities (Defined in the Codebook section of BSG Working Paper 'BSG-WP-2020-032-v5.0_0.pdf' which can be downloaded from https://covidtracker.bsg.ox.ac.uk/)	Oxford University: https://oxcgrtportal.azurewebsites.net/api/CSVDownload
c1_flag	Flag for general nation-wide implementation	Ditto
c2_workplace_closing	Record closings of workplaces	Ditto
c2_flag	Flag for general nation-wide implementation	Ditto
c3_cancel_public_events	Record cancelling public events	Ditto
c3_flag	Flag for general nation-wide implementation	Ditto
c4_restrictions_on_gatherings	Record the cut-off size for bans on private gatherings	Ditto
c4_flag	Flag for general nation-wide implementation	Ditto
c5_close_public_transport	Record closing of public transport	Ditto
c5_flag	Flag for general nation-wide implementation	Ditto
c6_stay_at_home_requirements	Record orders to 'shelter-in-place' and otherwise confine to home	Ditto

Variable Name	Variable Description	Data Source
c6_flag	Flag for general nation-wide implementation	Ditto
c7_restrictions_on_internal_movement	Record restrictions on internal movement	Ditto
c7_flag	Flag for general nation-wide implementation	Ditto
c8_international_travel_controls	Record restrictions on international movement	Ditto
e1_income_support	Record if the government is covering the salaries or providing direct cash payments, universal basic income, or similar, of people who lose their jobs or cannot work. (Includes payments to firms if explicitly linked to payroll/ salaries)	Ditto
e1_flag	Flag for general nation-wide implementation	Ditto
e2_debt/contract_relief	Record if govt. is freezing financial obligations (eg stopping loan repayments, preventing services like water from stopping, or banning evictions)	Ditto
e3_fiscal_measures	What economic stimulus policies are adopted? (USD)	Ditto
e4_international_support	Announced offers of COVID-19 related aid spending to other countries (USD)	Ditto
h1_public_information_campaigns	Record presence of public info campaigns	Ditto
h1_flag	Flag for general nation-wide implementation	Ditto
h2_testing_policy	Who can get tested?	Ditto
h3_contact_tracing	Are governments doing contact tracing?	Ditto
h4_emergency_investment_in_healthcare	Short-term spending on, e.g, hospitals, masks, etc	Ditto
h5_investment_in_vaccines	Announced public spending on vaccine development	Ditto
m1_wildcard	Record policy announcements that do not fit anywhere else	Ditto
stringencyindex	Index of stringency for national governments' responses to contain infection population	Ditto

2) country_master_df

Variable Name	Variable Description	Data Source
country	Country names cleaned and standardised from multiple data sources	Derived
longitude	Longitude value of country location	The World Health Organisation: https://www.who.int/data/gho or https://data.humdata.org/dataset/coronavirus-covid-19-cases-data-for-china-and-the-rest-of-the-world
latitude	Latitude value of country location	Ditto
total_cases	Accumulated number of registered cases	Worldometers: https://www.worldometers.info/coronavirus/
total_deaths	Accumulated number of registered	Ditto

Variable Name	Variable Description	Data Source
	deaths related to COVID-19	
total_recovered	Accumulated number of registered cases discharged from hospitalisation or quarantine	Ditto
total_tests	Accumulated number of tests for the presence of COVID-19 virus carried out	Ditto
active_cases	Accumulated number of registered cases under hospitalisation or quarantine	Ditto
serious_critical	Accumulated number of hospitalised cases in need of intensive critical care	Ditto
cases_per_m	Accumulated number of registered cases per million of population	Ditto
deaths_per_m	Accumulated number of registered deaths related to COVID-19 per million of population	Ditto
tests_per_m	Accumulated number of tests for the presence of COVID-19 virus carried out per million of population	Ditto
tests_per_case	Ratio of *total_tests* over *total_cases*	Derived
recovery_rate	Ratio of *total_recovered* over *total_cases*	Derived
critical_rate	Ratio of *serious_critical* over *active_cases* (difference between *total_cases* and *total_deaths* substitutes if *active_cases* is not available)	Derived
population_2020	Country population estimated in 2020	Worldometers: https://www.worldometers.info/world-population/population-by-country/
density_p_km2	Number of population per square kilometre of country area	Ditto
migrants_net	The difference between the number of immigrants and the number of emigrants throughout the year	Ditto
med_age	Median value or midpoint of population age	Ditto
urban_pop_%	Percent of country residents living in urban areas	Ditto
world_share_%	Country population in proportion of world population	Ditto
life_exp_both_sex	The average length of time that people of the country are normally likely to live, regardless of gender	Worldometers: https://www.worldometers.info/demographics/life-expectancy/#countries-ranked-by-life-expectancy
life_exp_female	The average length of time that the female population of the country are normally likely to live	Ditto
life_exp_male	The average length of time that the male population of the country are normally likely to live	Ditto
old_dep_ratio_2020	Old-age dependency ratio (age 65+ / age	The United Nations:

Variable Name	Variable Description	Data Source
	15-64) de facto population	https://population.un.org/wpp/Download/Standard/Population/
hospital_beds_per_100k	Total number of hospital beds per 100 thousand of country population	Our World In Data: blob:https://ourworldindata.org/562c6cd1-8da3-4dfc-8d70-8a80ed4f9cad
physicians_per_k	Total number of physicians per thousand of country population	Our World In Data: blob:https://ourworldindata.org/bc19aa7c-bf6b-419f-9d06-d3101310dd5d
gdp_capita_usd	2019 estimates of nominal GDP per capita from IMF	Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(nominal)_per_capita
gross_publicdebt_pct_of_gdp	Latest available figures for public debt in % of GDP from CIA	Wikipedia: https://en.wikipedia.org/wiki/List_of_countries_by_public_debt
days_1st_case	Maximum value from *daily_covid_govresponse_df*	Derived
days_1st_death	Maximum value from *daily_covid_govresponse_df*	Ditto
R0	Last value from *daily_covid_govresponse_df*	Ditto
c1_school_closing	Maximum value from *daily_covid_govresponse_df*	Ditto
c1_flag	Ditto	Ditto
c2_workplace_closing	Ditto	Ditto
c2_flag	Ditto	Ditto
c3_cancel_public_events	Ditto	Ditto
c3_flag	Ditto	Ditto
c4_restrictions_on_gatherings	Ditto	Ditto
c4_flag	Ditto	Ditto
c5_close_public_transport	Ditto	Ditto
c5_flag	Ditto	Ditto
c6_stay_at_home_requirements	Ditto	Ditto
c6_flag	Ditto	Ditto
c7_restrictions_on_internal_movement	Ditto	Ditto
c7_flag	Ditto	Ditto
c8_international_travel_controls	Ditto	Ditto
e1_income_support	Ditto	Ditto
e1_flag	Ditto	Ditto
e2_debt/contract_relief	Ditto	Ditto
e3_fiscal_measures	Ditto	Ditto
e4_international_support	Ditto	Ditto
h1_public_information_campaigns	Ditto	Ditto
h1_flag	Ditto	Ditto
h2_testing_policy	Ditto	Ditto
h3_contact_tracing	Ditto	Ditto

Variable Name	Variable Description	Data Source
h4_emergency_investment_in_healthcare	Ditto	Ditto
h5_investment_in_vaccines	Ditto	Ditto
m1_wildcard	Ditto	Ditto
stringencyindex	Ditto	Ditto

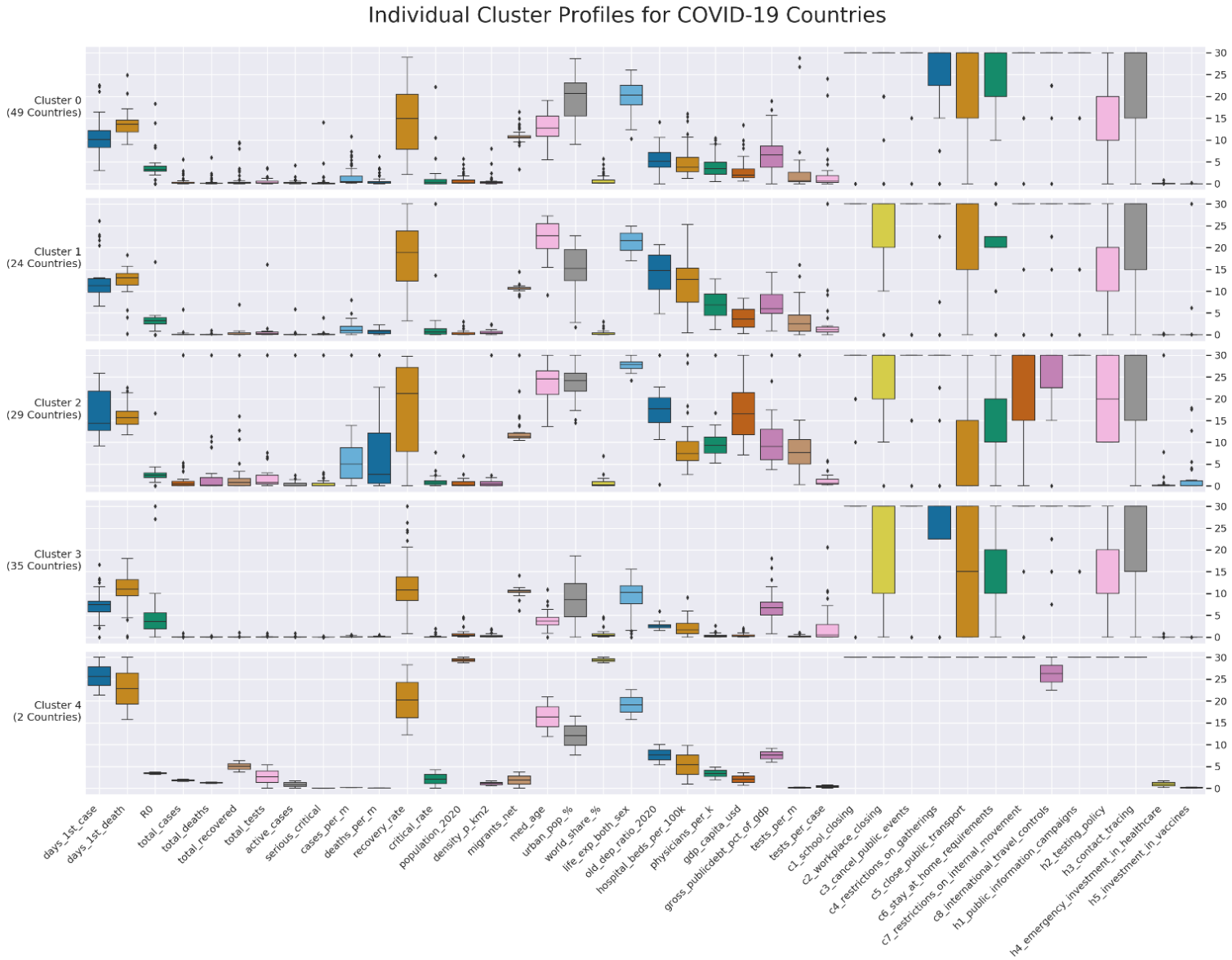
3) goog_mob_df

Variable Name	Variable Description	Data Source
sub_region_1	City	Google COVID-19 Community Mobility Reports: https://www.google.com/covid19/mobility/index.html?hl=en
date	Date of visits of users who have opted-in to Location History for their Google Account	Ditto
retail_and_recreation_percent_change_from_baseline	Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters	Ditto
grocery_and_pharmacy_percent_change_from_baseline	Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies	Ditto
parks_percent_change_from_baseline	Mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens	Ditto
transit_stations_percent_change_from_baseline	Mobility trends for places like public transport hubs such as subway, bus, and train stations	Ditto
workplaces_percent_change_from_baseline	Mobility trends for places of work	Ditto
residential_percent_change_from_baseline	Mobility trends for places of residence	Ditto

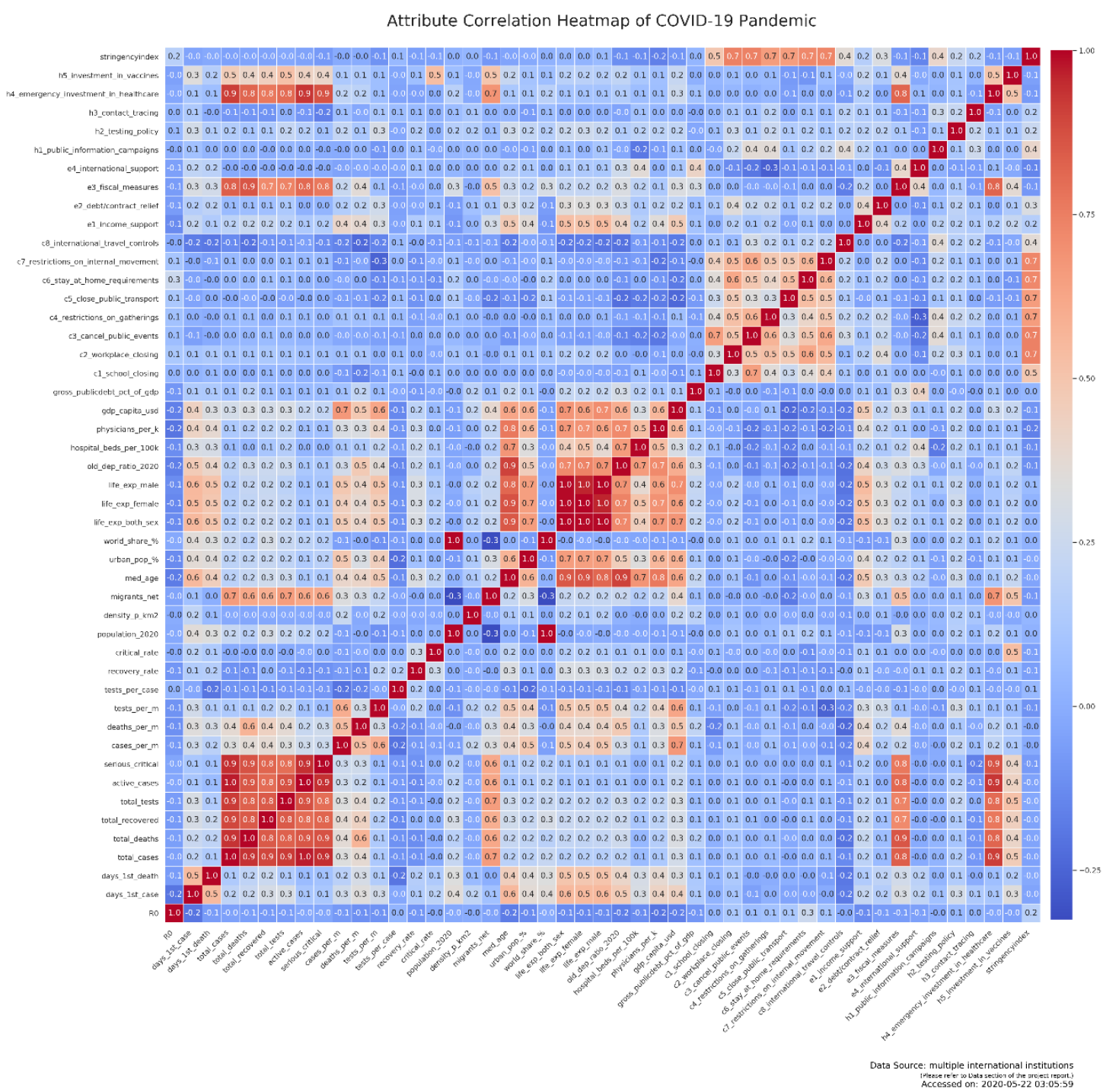
4) five_city_venues_df

Variable Name	Variable Description	Data Source
station	Metro station of the city	Foursquare Places API
stn_latitude	Latitude value of the metro station	Ditto
stn_longitude	Longitude value of the metro station	Ditto
venue	The best known name for this venue	Ditto
venue_lat	Latitude value of the venue	Ditto
venue_lon	Longitude value of the venue	Ditto
venue_categories	Place category of the venue	Ditto
venue_categories_id	ID of the Place category	Ditto
city	City	Derived

Appendix B. **COVID-19 Individual Cluster Profiles**



Appendix C. COVID-19 Attribute Correlation Heatmap



Data Source: multiple international institutions
(Please refer to Data section of the project report.)
Accessed on: 2020-05-22 03:05:59

Appendix D. COVID-19 Attribute Correlation Scatter Plot