FINDING A RELATIONSHIP BETWEEN RISK PERCEPTION OF CORONAVIRUS AND THE VARIOUS RESPONSES TO COMBAT THE PANDEMIC

Kelvin Nii Lartey-Abrahams

December 2020

ABSTRACT

The purpose of my project is to investigate the relationship between people's perception of risk posed by COVID-19 and governments' responses across countries. I used Mobility changes across Transit Stations, Workplaces and Retail & Recreational Centres as proxies for risk perception and collected daily data on them over a period of one and a half months. Corresponding daily data for 18 government response measures were collected. I used Multivariate Multiple Regression and Group Lasso Regression (which was capable of variable selection) in my work and compared the two models. In general, the Group Lasso method gave more accurate results than the Multivariate Multiple Regression. The results from the Group Lasso showed that the Containment measures were most effective in modelling mobility factors on the continental level and on the national level. Generally, the effects of economic response and public health sector responses on risk perception were mixed, but contact tracing was the most effective health sector response across continents and nations. One possible area for further research is to increase the datasets, since the pandemic is still active, to get even better results. Further analysis could be conducted into other models that could improve upon what has already been done in the study.

Contents

INTRODUCTION

We are in extraordinary times. As of 25 May 2020, more than 5 million cases of Coronavirus (COVID-19) have been recorded across the globe, with 346,283 reported deaths due to the virus. There have however been at least 2 million recoveries world-wide, which provides a glimmer of hope in these dark times. Nonetheless, it has not come as a surprise that the virus has affected all facets of life – sports, recreation, and the global economy. The global economy has been hit hardest, disrupting supply chains, and restricting movement and commerce, and this has invariably sped the notion (by most economists) that a recession by the end of 2020 is inevitable. Despite this phenomenon, the risk perception of Coronavirus varies across race, age groups, level of education, job, and race. According to a research conducted by the Pew Research Centre on the impact of the virus in the United States, concerns about contracting the virus are higher among Blacks and Hispanics than whites. Also, even though statistically, black Americans have been hit hard by the pandemic, dying at three times the rate of white people, they are more sceptical about medical scientists and experimental treatments for the Coronavirus than Hispanics and Whites. Interestingly, many Americans feel that the U.S. economy is under more threat compared to their health and finances and majority of people earning less than $50,000 say they would not be paid if the virus caused them to miss work for more than 2 weeks. Despite these statistics, majority of Americans think that the news media have exaggerated COVID-19 risks at least slightly, and Republicans are more likely to say so than democrats. Other measures of risk perception such as mobility trends across the globe are very important in predicting what kind of actions people are willing to take in relation to their risk perception. [Restructure] Nonetheless, governments across the world are taking certain measures to curb the impact of the virus. These measures are summarized into indicators such as Closures and Containment measures, Economic Response measures and Public Health measures.

Social media is very important in this modern age, and its activity has increased across the world, as health experts have embraced it as a tool to fight Coronavirus. Global Health experts are increasingly using platforms like Tik Tok and Twitter to communicate directly with the public about the rapid spread of the disease and some precautions to take to prevent the further spread of the virus.

Objectives

In this research, I will use changes in mobility trends across the globe as a proxy for measures of risk perception.

I seek to find out the following:

1. Find out if there is any correlation about how people perceive risk posed by COVID 19 and the various government responses
2. Find out which government response respective continents and select countries are sensitive to and how accurate the models used are at predicting risk perception.

Data Description and Preparation

In my project I have 3 mobility factors (Retail and Recreation, Transit Stations and Workplace Mobility) which will act as my predicted variables. The dataset used in this analysis also consists of government responses across 116 countries in the world which I will use as my predictor variables. The government response dataset was taken from the University of Oxford COVID-19 Government Website (https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker). Data collection is still on-going. The tracker system collects information on several different common policy responses governments have taken in response to the pandemic using 17 different indicators. The indicators are listed below:

- Closures and Containment
  C1 –school closing
  C2 –workplace closing
  C3 –cancel public events
  C4 –restrictions on gathering size
  C5 –close public transport
  C6 – "shelter-in-place" and home confinement orders
  C7 –restrictions on internal movement
  C8 –restrictions on international travel

- Economic response
  E1 –income support
  E2 –debt/contract relief for households
  E3 –fiscal measures
  E4 –giving international support

- Public health / health system
  H1 –public information campaign
  H2 –testing policy
  H3 –contact tracing
  H4 –emergency investment in healthcare
  H5 –investment in Covid-19 vaccines

Mobility data obtained from https://www.google.com/covid19/mobility/ shows how visits and length of stay at different places in the world change compared to a baseline. The changes are calculated using anonymous data used to show popular times for places in Google maps. Changes for each day are compared to a baseline value for that day of the week. The baseline is the median value for the corresponding day of the week, during the 5-week period Jan 3, 2020 to Feb 6, 2020. Mobility factors – and their descriptions - used in this study include:

- **Retail and Recreation** - Mobility trends for places like restaurants, cafes, shopping centres, theme parks, museums, libraries, and movie theatres.
- **Transit Stations** - Mobility trends for places like public transport hubs such as subway, bus, and train stations.
- **Workplaces** - Mobility trends for places of work
- **Grocery and Pharmacy** - Mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies.
- **Parks** - Mobility trends for places like national parks, public beaches, marinas, dog parks, plazas, and public gardens.

- **Residence** - Mobility trends for places of residence.

Only the most sensitive mobility factors were used in the study, and they are as follows:

- Retail and Recreation
- Transit Stations
- Workplaces

Modifications to Data

In preparing the data,

1. Blank cells were removed from the dataset and this reduced the number of rows to 4895. There are 23 columns with 3 columns representing my response variables (Retail and Recreation, Transit Stations and Workplaces) and 17 columns representing my explanatory variables (Government Responses C1-C8, E1-E5, and H1- H5). Three other columns consist of Country names, Period and Continents.
2. In the “Continents” column, Oceania and the Caribbean are not continents, but Puerto Rico geographically does not belong to any continent, so I classified it as part of the Caribbean. Eurasia is a special case because Turkey is both in Europe and Asia.
3. I took out 52 countries not included in mobility data and data for 13 countries not included in the Government response data.
4. Government Response Data is from 1 Jan 2020 -20 May 2020 and Mobility Data is from 16 Feb 2020 – 29 Mar 2020. I trimmed Government Response Data down to period 16 Feb 2020 – 29 Mar 2020 to match the Mobility data.
5. Insensitive factors **Grocery & Pharmacy**, **Parks** and **Residence** have been removed.
6. I split each sub-dataset into 705 train for the purposes of building the models and 30% for the purpose of prediction evaluation.

METHODOLOGY AND MODEL DESCRIPTION

Mobility changes are continuous variables. I will use Multivariate multiple regression and assume the linear regression model to predict mobility changes using the set of government responses I have. Alternative to Multivariate multiple regression, I will introduce Lasso regression, which can perform variable selection.

Multivariate Multiple Regression

Consider the problem of modelling the relationship between n responses $Y_1, Y_2, Y_3, \dots, Y_n$ and a single set of predictor variables $X_1, X_2, X_3, \dots, X_p$. Each response variable is assumed to follow its own regression model such that

$$\mathrm{Y}_1 = \beta_{01} + \beta_{11}x_1 + \cdots + \beta_{p1}x_p + \epsilon_1$$

$$\mathrm{Y}_2 = \beta_{02} + \beta_{12}x_1 + \cdots + \beta_{p2}x_p + \epsilon_2$$

$$\vdots$$

$$\mathrm{Y}_k = \beta_{0k} + \beta_{1k}x_1 + \cdots + \beta_{pk}x_p + \epsilon_k$$

The error term $\boldsymbol{\epsilon}' = [\epsilon_1 . \epsilon_1, \dots, \epsilon_1]$ has $E(\boldsymbol{\epsilon}) = 0$ and $Var(\boldsymbol{\epsilon}) = \Sigma$ where Σ represents the Variance-Covariance Matrix.

Conforming to notation in Classical Linear Regression, let $\boldsymbol{X'} = [x_{i0}, x_{i1}, \dots, x_{ip}]$ represent the values of the predictor variables for the ith trial, $\boldsymbol{Y^i} = [Y_{i1}, Y_{i2}, \dots, Y_{ik}]$be the response vector and $\boldsymbol{\epsilon}_i' = [\epsilon_{i1}, \epsilon_{i2}, \dots, \epsilon_{ik}]$ be the error vector.

In matrix notation,

$$\underbrace{X}_{(nx(p+1))} = \begin{bmatrix} x_{10} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n0} & \cdots & x_{np} \end{bmatrix}$$

$$\underbrace{Y}_{(nxk)} = \begin{bmatrix} y_{11} & \cdots & y_{1k} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{nk} \end{bmatrix} = [Y_{(1)} \vdots Y_{(2)} \vdots \cdots \vdots Y_{(k)}]$$

$$\underbrace{\beta}_{((p+1)x\,k)} = \begin{bmatrix} x_{01} & \cdots & x_{ok} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pk} \end{bmatrix} = [\beta_{(1)} \vdots \beta_{(2)} \vdots \cdots \vdots \beta_{(k)}]$$

$$\underbrace{\epsilon}_{(nxk)} = \begin{bmatrix} x_{01} & \cdots & x_{ok} \\ \vdots & \ddots & \vdots \\ x_{p1} & \cdots & x_{pk} \end{bmatrix} = [\epsilon_{(1)} \vdots \epsilon_{(2)} \vdots \cdots \vdots \epsilon_{(k)}]$$

The multivariate linear regression model is given as

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\epsilon}$$

Where $E[\epsilon_{(i)}] = 0$ and $Cov(\epsilon_{(i)}, \epsilon_{(m)}) = \sigma_{im} I \quad i,m = 1,2,\ldots k$

Simply stated, the ith response $Y_{(i)}$ follows the linear regression model

$\boldsymbol{Y_{(i)}} = \boldsymbol{X\beta_{(i)}} + \boldsymbol{\epsilon_{(i)}}$ i =1, 2…. k with $Cov(\epsilon_{(i)}) = \sigma_{ii} I$.

Least Square Estimation of $\boldsymbol{\beta}$

Given the outcomes $\boldsymbol{Y}$ and the values of the predictor variables $\boldsymbol{X}$, the least square estimates $\boldsymbol{\hat{\beta}_{(i)}}$ are determined specifically from observing $Y_{(i)}$ on the ith response.

Taking $\boldsymbol{\hat{\beta}_{(i)}} = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y_{(i)}}$, we obtain the vector containing the univariate least square estimates:

$$\boldsymbol{\hat{\beta}} = [\boldsymbol{\hat{\beta}_{(1)}} \vdots \boldsymbol{\hat{\beta}_{(2)}} \vdots \cdots \vdots \boldsymbol{\hat{\beta}_{(m)}}] = (\boldsymbol{X'X})^{-1}\boldsymbol{X'}[\boldsymbol{Y_{(1)}} \vdots \boldsymbol{Y_{(2)}} \vdots \cdots \vdots \boldsymbol{Y_{(m)}}] = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}.$$

Using the least square estimates $\boldsymbol{\hat{\beta}}$, the matrix of predicted values and residuals are formed respectively:

Predicted values $\boldsymbol{\hat{Y}} = \boldsymbol{X\hat{\beta}} = \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$

Residuals $\boldsymbol{\hat{\epsilon}} = \boldsymbol{Y} - \boldsymbol{\hat{Y}} = [\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X'X})^{-1}\boldsymbol{X'}]\boldsymbol{X}$

Lasso Regression

The least squares fitting procedure estimates $\beta_0, \beta_1, \ldots, \beta_p$ using the values that minimize SSE $=\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2$. For Lasso regression, we modify this equation by trying to find coefficient estimates β_j that minimize the following expression:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p}|\beta_j|$$

where $\lambda \geq 0$ is a tuning parameter, which is determined separately by cross-validation. The tuning parameter, also known as the shrinkage penalty, serves as a control to the relative impact of these two terms in our expression on the regression coefficient estimates. When $\lambda = 0$, the penalty term has no effect, and lasso regression will produce the same results as the least squares estimates. However, as $\lambda \to \infty$, the impact of the shrinkage penalty grows, and the lasso regression coefficient estimates approach zero. Lasso regression produces a different set of coefficient estimates for each value of λ. This is unlike the least squares estimate but selecting a λ value is critical to getting the best results. To select the best λ, we obtain a grid of λ values, and compute the cross-validation error for each value of λ. Then we then select the tuning parameter value for which the cross-validation error is smallest. Finally, the model is re-fit using all the available observations and the selected λ value.

In the case of the lasso, the penalty forces some of the coefficient estimates to be exactly equal to zero when the tuning parameter λ is sufficiently large. As a result of this, we could say the lasso model performs variable selection and is much easier to interpret since unimportant variables are deleted.

EXPLORATORY DATA ANALYSIS

Histograms of Response Variables

Figure 1. HISTOGRAM SHOWING MOBILITY CHANGES DUE TO RETAIL AND RECREATION

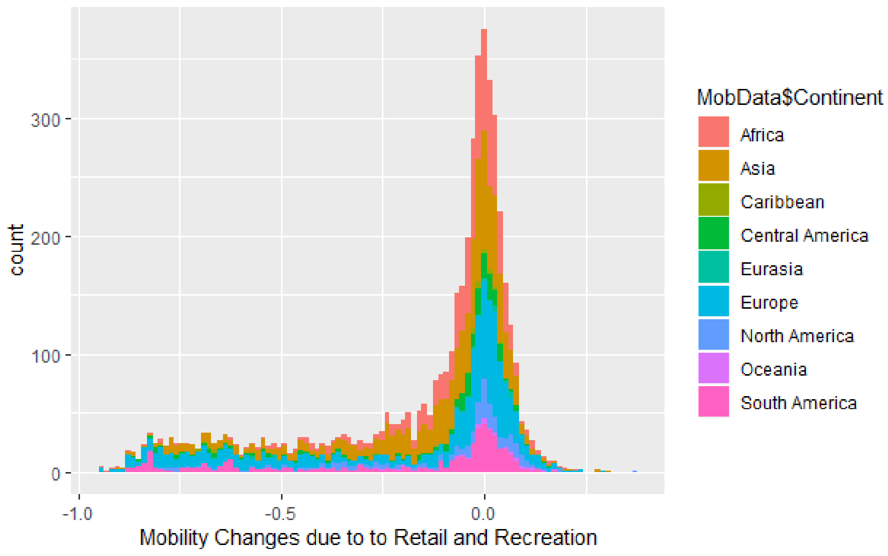


Figure 2 INDIVIDUAL FACETS SHOWING MOBILITY CHANGES DUE TO RETAIL AND RECREATION

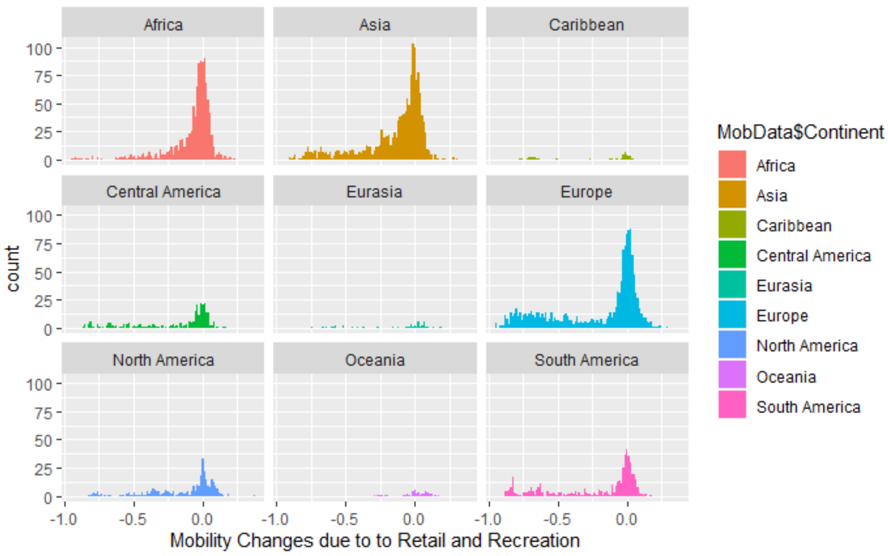


Figure 3 HISTOGRAM SHOWING TRANSIT STATION MOBILITY CHANGES

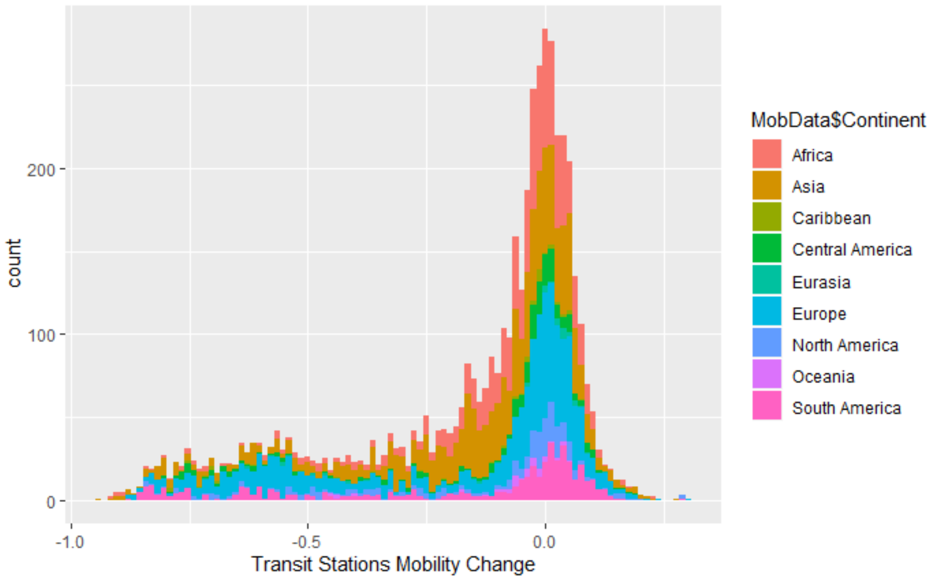


Figure 4 INDIVIDUAL FACETS SHOWING TRANSIT STATION MOBILITY CHANGE

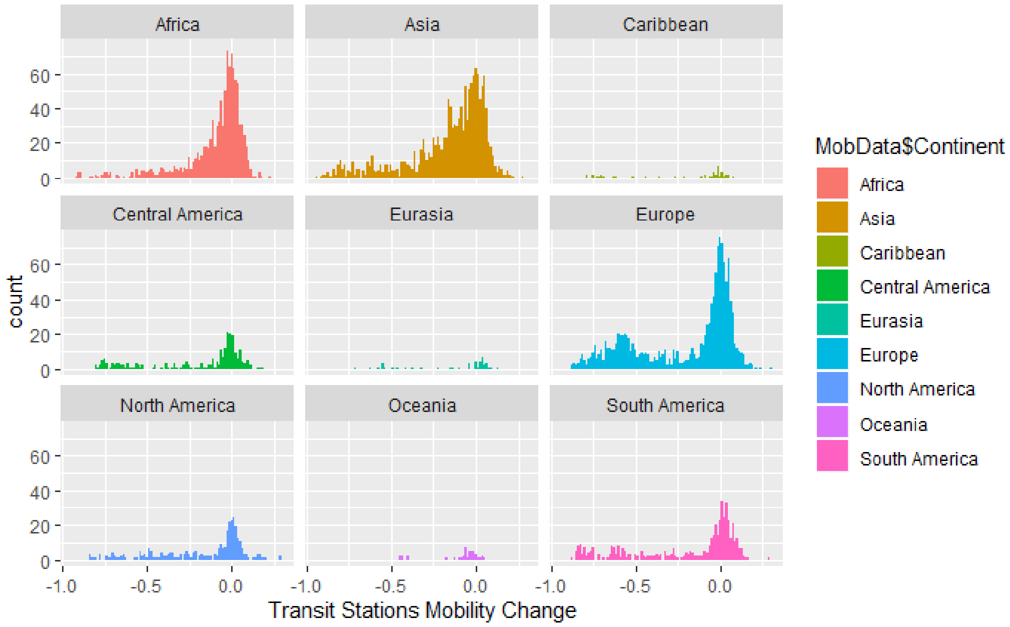


Figure 5 HISTOGRAM SHOWING MOBILITY CHANGES DUE TO RETAIL AND RECREATION

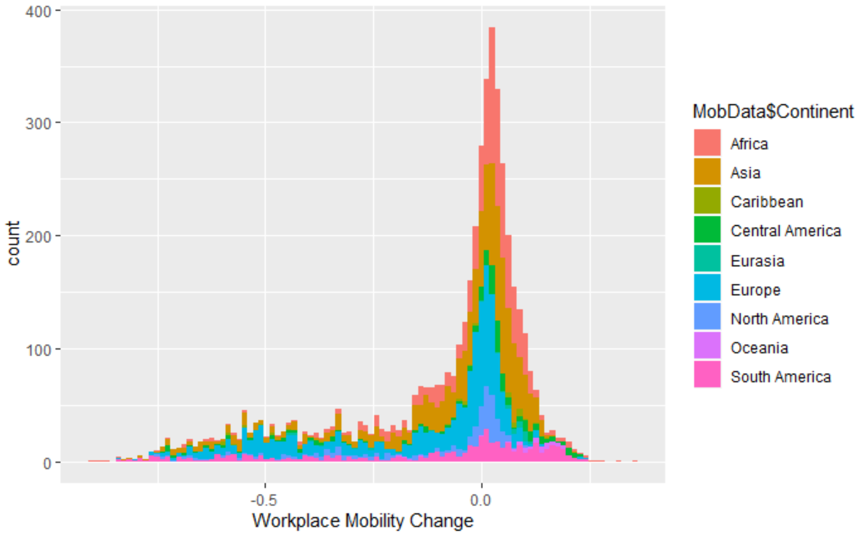
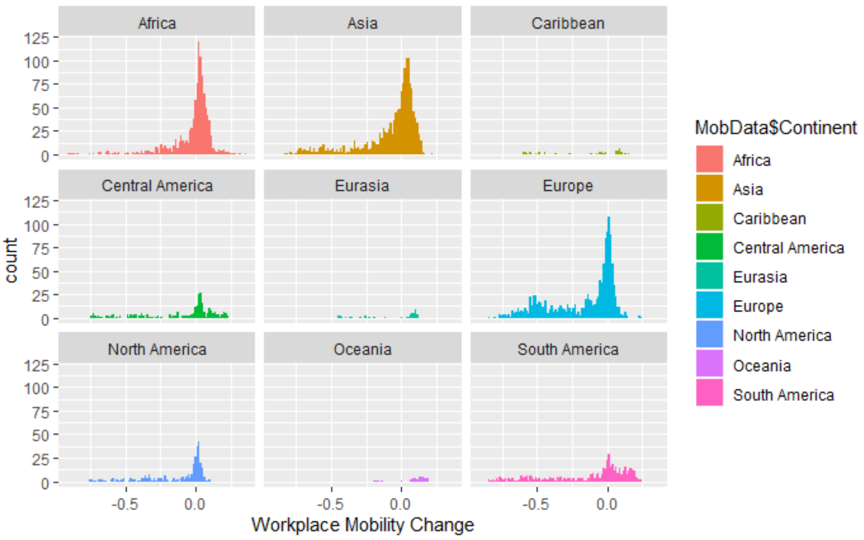


Figure 6 INDIVIDUAL FACETS SHOWING WORKPLACE MOBILITY CHANGE



Figures 1-6 show the frequency of changes in mobility factors across continents between the given period of 16 February 2020 and 29 March 2020. From the condensed graphs for all three mobility factors, it is apparent that the data is left skewed. The distributions therefore tell us that most people did not react much differently from how they would in terms of visiting parks, retailing, or using public transport for the workplace

The faceted histograms show us how mobility changed across Continents individually. From Figure 2. We realise Asia recorded the most inactivity in retail and recreation within the given period. This inactivity was followed by Africa and Europe, respectively. Africa and Asia however had a considerable amount of reduced retail and recreational activity compared to Europe who had more negative activity. Eurasia, the Caribbean, and Oceania recorded very negligible movements across the board, but this can be attributed to the fact that they are traditionally not continents, but just geographical regions. North and Central America are even in terms of inactivity, but South America but South America has a slightly higher frequency in both inactivity and negative activities.

From Figure 4., Africa, Europe, and Asia are relatively even in terms of little deviation from the normal activity at transit stations. The graph of Asia is however a little wider around 0 compared to Europe and Africa, which indicates a little more positive and negative movement. Europe however has a higher frequency in huge negative movements (-0.5 - -1.0), The graphs for Central, North and South America are very similar to the graphs for changes in Retail and Recreational activity.

From Figure 6., South America recorded more positive frequency in Workplace Mobility as compared to North and Central America. North America has very little positive change in Workplace mobility as compared South and Central America. Africa, Europe, and Asia are still even with respect to workplace mobility change, but Europe has more negative reaction in terms of the workplace.

Bar plots of Response Variables

Figure 7 MOBILITY CHANGE FOR 16 FEB 2020

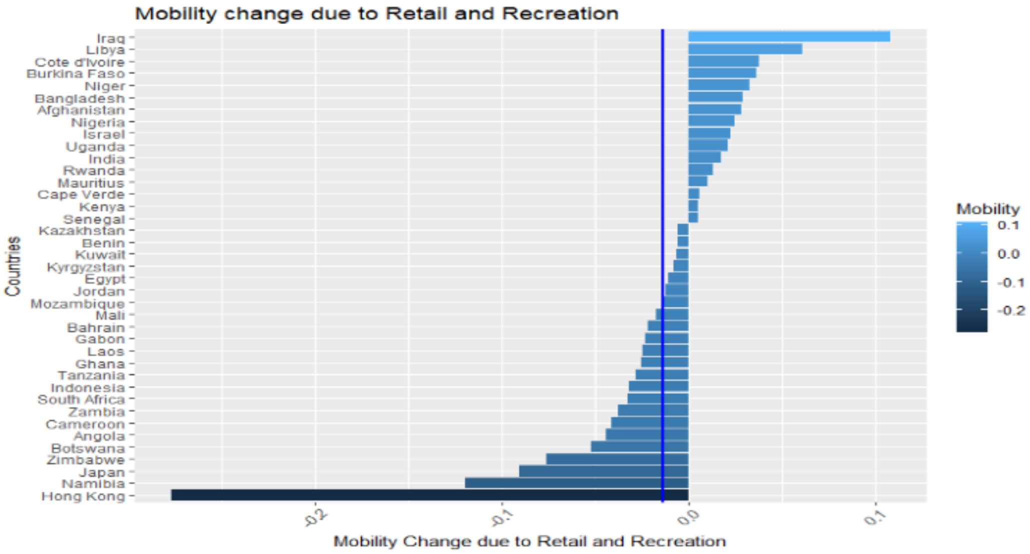
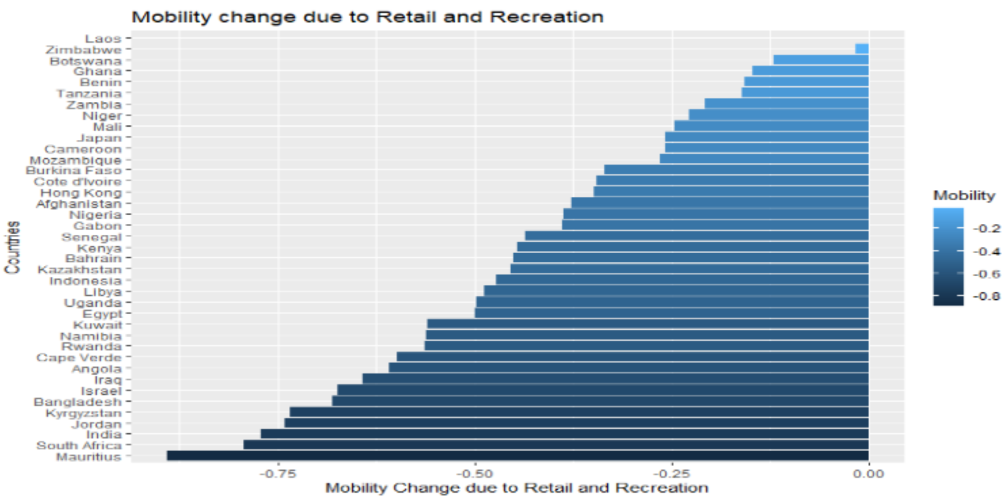


Figure 8 MOBILITY CHANGE FOR 29 MAR 2020



Figures 7 and 8 shows how mobility due to retail and recreation change across a sample of countries on 16 February 2020 and how it compares to March 29, 2020, respectively. On 16 February 2020, Iraq showed the most positive retail and recreation movements with Hong Kong reacting the most negatively. The average line is just a little to the left of zero, signifying a general negative movement across countries. On the 29th of March, however, all movements had become negative and this might have been due to the increasing spread of the pandemic, leading to more caution among countries. Laos recorded the least negative movement and Mauritius recorded the most negative movement.

Figure 9 MOBILITY CHANGE FOR 16 FEB 2020

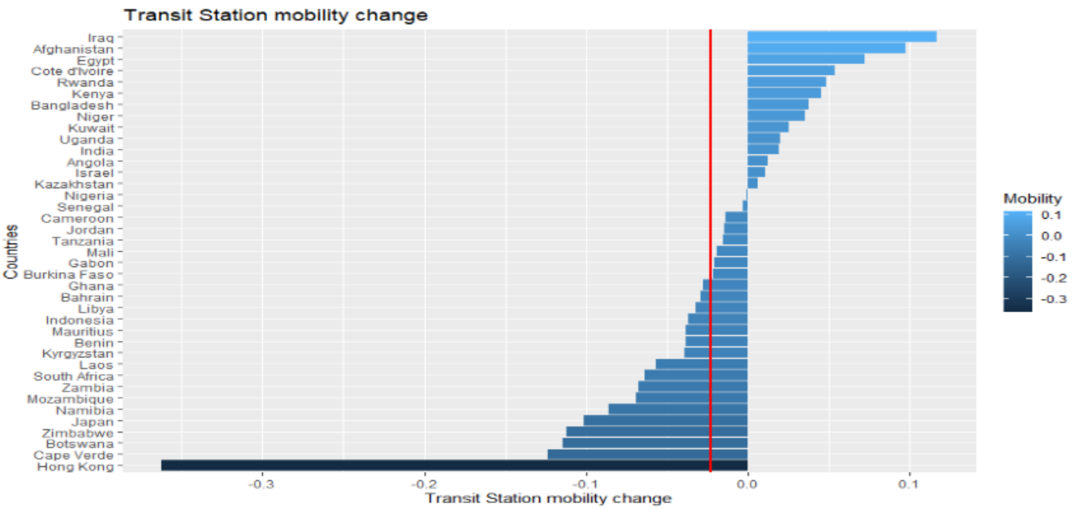
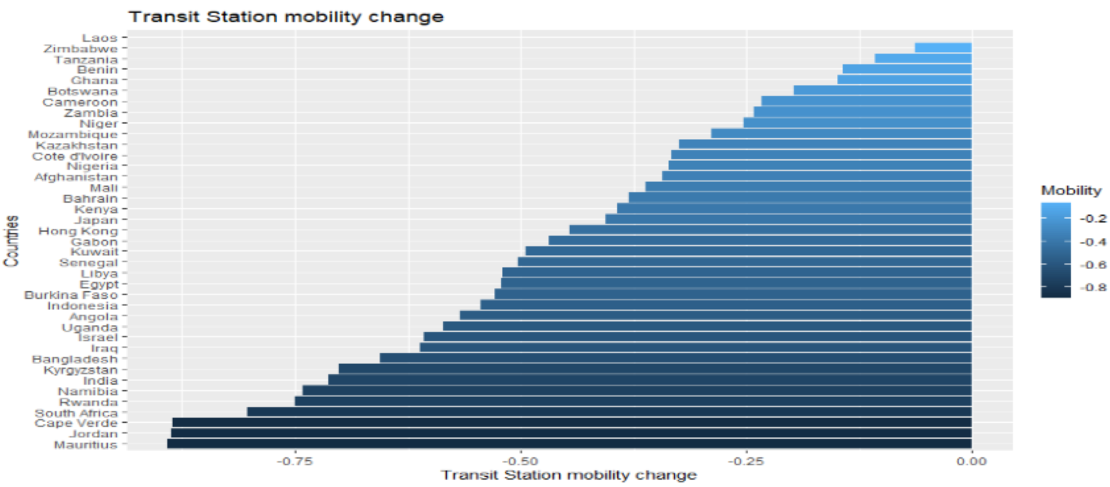


Figure 10 MOBILITY CHANGE FOR 29 MAR 2020



Figures 9 and 10 shows how transit stations mobility changes across a sample of countries on 16 February 2020 and how it compares to March 29, 2020. On 16 February 2020, Iraq showed the most positive transit station with Hong Kong still reacting the most negatively. Nigeria recorded a 0 which indicates that transit station activity was relatively neutral within the period. The average line for transit stations is also just a little to the left of zero, signifying a general negative movement across countries. On the 29th of March, all movements had become negative with Mauritius recording the most reduced transit station activity. Laos once again recorded the least negative movement.

Figure 11 MOBILITY CHANGE FOR 16 FEB 2020

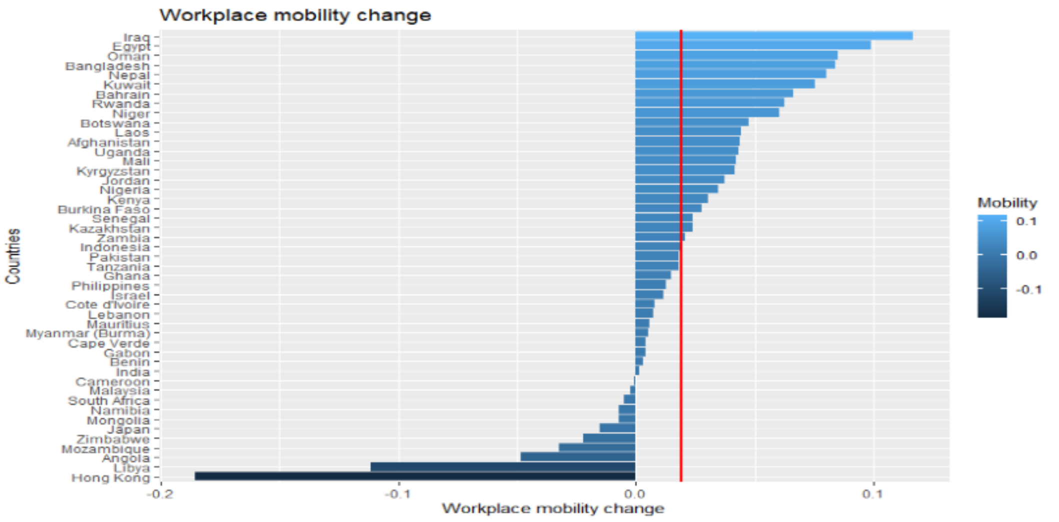
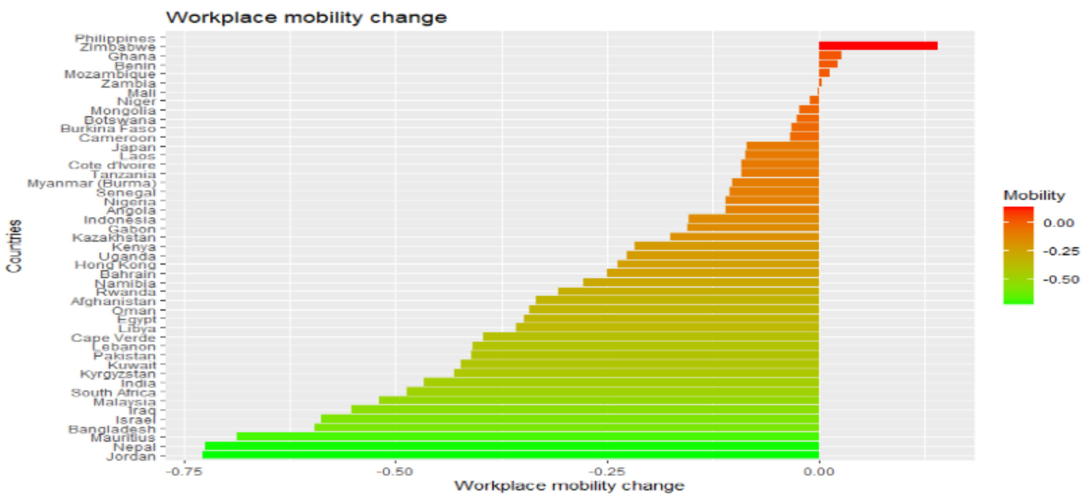


Figure 12 MOBILITY CHANGE FOR 29 MAR 2020



Figures 11 and 12 shows how transit stations mobility changes across a sample of countries on 16 February 2020 and how it compares to March 29, 2020. On 16 February 2020, Iraq showed the most positive transit station with Hong Kong still reacting the most negatively. Cameroon recorded a relatively neutral activity within the period. The average line for transit stations however is to the right of zero, signifying that workplace activity was still positive across countries. On the 29th of March, most countries had transitioned to working from home, as shown by the negative movement in the graph. A few countries like the Philippines, Zimbabwe and Ghana still had positive movements. Jordan recorded the most negative change in terms of workplace mobility.

MODEL FITTING: CONTINENTS

Multivariate Multiple Regression Model

Under the Multivariate Multiple regression model, I am primarily interested in finding out whether there is a significant relationship between perception of risk due to COVID-19 and the responses taken to combat the pandemic. I firstly run the regression model for the Continents of the World to get a sense of responses that are relevant in influencing the three mobility factors. The following table summarizes the adjusted R-Squares for each continent based on each response. The higher the adjusted R-square, the better the model is at explaining the variability in that particular response.

Table 1. Adjusted R-Squares for initial models (%)

Response	Africa	Asia	Central America	Europe	North America	South America
Retail and Recreation	69.38	86.10	89.25	86.63	93.89	92.62
Transit	72.10	84.23	90.24	87.95	94.77	91.28
Workplace	53.45	77.58	83.79	80.81	75.72	75.74

From Table 1, the lowest adjusted R-square is 69.38%, which means that the model built for Africa explains 69.38% of the variability in mobility changes due to retail and recreation. This is good and hence is a good initial fit to explain mobility changes due to retail and recreation in Africa.

Since I was interested in determining which explanatory variables are significant to the joint mobility factors, I run a multiple MANOVA test. to determine which response variables are jointly contributing to the 3 mobility factors. Table 2. summarizes all the Government response factors and a “Yes” signifies that that government response factor is significant to the joint mobility factors.

Table 2. Jointly important factors based on Multivariate Multiple Regression

	Africa	Asia	Central America	Europe	North America	South America
School Closing (C1)	Yes	Yes	Yes	Yes	Yes	Yes
Workplace Closing (C2)	Yes	Yes	Yes	Yes	Yes	Yes
Cancel Public Events (C3)	Yes	Yes		Yes	Yes	Yes
Restrictions on Gatherings (C4)	Yes	Yes	Yes	Yes	Yes	Yes
Close Public Transport (C5)	Yes	Yes	Yes	Yes	Yes	Yes
Stay at home req. (C6)	Yes	Yes		Yes	Yes	Yes
Restrictions on internal movement (C7)	Yes	Yes	Yes	Yes	Yes	Yes
International Travel Controls (C8)	Yes	Yes	Yes	Yes	Yes	Yes
Income Support (E1)	Yes			Yes		
Debt Contract Relief (E2)	Yes	Yes		Yes	Yes	Yes
Fiscal measures (E3)						
International Support (E4)						
Public Information Campaigns (H1)	Yes		Yes		Yes	Yes
Testing Policy (H2)	Yes	Yes	Yes		Yes	Yes
Contact Tracing (H3)	Yes	Yes	Yes		Yes	
Emergency investment in healthcare (H4)	Yes					
Investment in vaccines (H5)						

From Table 2, the following could be observed:

1. All the continents' containment measures (C1-C8) have a significant effect on the joint mobility factors.
2. For all continents, E3, E4 and H5 were not at all significant to mobility changes in retail and recreation, transit stations or the workplace.
3. Africa was the only continent where mobility was affected by Emergency Investment in Healthcare (E4).
4. The number of Government response factors in Africa contributed to the joint mobility factors than any other continent, with Asia and North America having 12 factors each affecting mobility.

Group Lasso Regression Model: Continents

To perform Group Lasso Regression and hence variable selection, I first needed to find the tuning parameter, λ, that minimizes the error in my model. Table 3 shows the appropriate λ value needed to minimize the MSE for models built for each continent. Figures 13, 14 and 15 in the appendix give a graphical perspective on the behaviour of each coefficient estimate as the tuning parameter gets very large for each continent. Although not easily seen, some coefficient estimates are shrunk to 0 before others as the tuning parameter grows. There is also one peculiarity about Group Lasso Regression: for categorical variables, Group Lasso Regression deletes levels and sometimes not just the whole variable. This implies that sometimes the Group Lasso Regression deems a variable important, but not all its levels. Table 4 provides a summary of all the Government response factors and a "Yes" signifies that that government response factor is significant to the joint mobility factors for the Group Lasso model. A "Yes" with an asterisk (*) signifies a variable for which some of the levels have been deleted by the Group Lasso.

Table 3. Best Tuning Parameter for Each Model

Model for Continent	Best Tuning Parameter (λ)
Africa	0.005420126
Asia	0.0004487489
Central America	0.00360207
Europe	0.002149771
North America	0.008916957
South America	0.001803525

Table 4. Jointly important factors based on Group Lasso Regression

	Africa	Asia	Central America	Europe	North America	South America
School Closing (C1)	Yes	Yes	Yes	Yes	Yes	Yes
Workplace Closing (C2)	Yes	Yes	Yes*	Yes	Yes	Yes
Cancel Public Events (C3)	Yes	Yes	Yes*	Yes	Yes	Yes
Restrictions on Gatherings (C4)		Yes	Yes	Yes	Yes*	Yes
Close Public Transport (C5)	Yes	Yes	Yes	Yes	Yes	Yes
Stay at home req. (C6)	Yes*	Yes	Yes*	Yes	Yes*	Yes
Restrictions on internal movement (C7)	Yes*	Yes		Yes	Yes	Yes
International Travel Controls (C8)	Yes	Yes	Yes*	Yes	Yes*	Yes
Income Support (E1)	Yes	Yes		Yes		Yes
Debt Contract Relief (E2)	Yes	Yes		Yes	Yes	Yes
Fiscal measures (E3)	Yes	Yes	Yes	Yes		Yes
International Support (E4)						
Public Information Campaigns (H1)	Yes*	Yes	Yes	Yes	Yes	Yes
Testing Policy (H2)	Yes*	Yes	Yes*	Yes	Yes*	Yes*
Contact Tracing (H3)	Yes	Yes	Yes	Yes	Yes	
Emergency investment in healthcare (H4)		Yes	Yes			
Investment in vaccines (H5)						

Before moving on to the specifics of which levels the Group Lasso removed from each categorical variable, I made a general comparison of both the Group Lasso model and the MMR. The following are my findings:

1. The Group Lasso shares striking similarities with the MMR, as all containment measures except C4 for Africa and C7 for Central America were relevant to the joint mobility factors.
2. Just like the MMR, the Group Lasso also shows E4 and H5 are irrelevant to predicting joint mobility factors.
3. For the Group Lasso, E3 is significant to predicting joint mobility in all continents except North America. The MMR did not deem E3 significant to mobility in any continent.

The following are the levels of the categorical variables that were deemed important to joint mobility by the Group lasso for each continent, with the ones not deemed significant added for completeness:

Africa

C6_Stay at home Requirements *Not leaving home with exceptions*	✔
C6_Stay at home Requirements *Recommend not leaving house*	✔
C6_Stay at home Requirements *Not leaving house with minimal exceptions*	✖
C7_Restrictions on Internal Movement *Require Closing*	✔
C7_Restrictions on Internal Movement *Recommend Closing*	✖
H1_Public Info campaigns *Public Officials Urging caution*	✔
H1_Public Info campaigns *No COVID-19 public campaign*	✖
H2_Testing Policy *Testing anyone showing symptoms*	✔
H2_Testing Policy *Those having symptoms and meeting specific criteria*	✔
H2_Testing Policy *Open public testing*	✖

South Africa

H2_Testing Policy *Testing anyone showing symptoms*	✔
H2_Testing Policy *Those having symptoms and meeting specific criteria*	✔

Asia

E1_Income Support *Government is replacing 50% or more of salary*	✔
E1_Income Support *Government is replacing less than 50% of salary*	✖

Central America

C2_Workplace Closing *No measures*	✔
C2_Workplace Closing *Closing some sectors*	✖
C6_Stay at home Requirements *Not leaving home with minimal exceptions*	✔
C6_Stay at home Requirements *Not leaving house with exceptions*	✖
C8_International Travel Controls *No measures*	✔
C8_International Travel Controls *Quarantine arrivals from high-risk regions*	✔
C8_International Travel Controls *Total Border Closure*	✔
C8_International Travel Controls *Screening*	✖
H2_Testing Policy *Those having symptoms and meeting specific criteria*	✔
H2_Testing Policy *Testing anyone showing symptoms*	✖

North America

H2_Testing Policy *Testing anyone showing symptoms*	✔
H2_Testing Policy *Open public testing*	✔
H2_Testing Policy *Those having symptoms and meeting specific criteria*	✖
C4_Restrictions on gatherings *Between ten and a hundred*	✔
C4_Restrictions on gatherings *Between zero and a thousand*	✔
C4_Restrictions on gatherings *No restrictions*	✔
C4_Restrictions on gatherings *Less than ten*	✖

C6_Stay at home Requirements *Recommend not leaving house* ✔

C6_Stay at home Requirements *Not leaving home with minimal exceptions* ✔

C6_Stay at home Requirements *Not leaving house with exceptions* ✖

Model Comparison and Selection

Table 5. compares both models fit to the joint mobility factors for each continent based on two metrics; the Adjusted R-square value (which shows how much the important factors in each model measure variability in joint mobility factors) and the Root Mean Square Error (which measures predictive accuracy of the model; the lower the RMSE, the better the model)

Table 5. Model Selection

	MMR		GROUP LASSO	
MODEL FOR CONTINENT	ADJ R-SQ (%)	RMSE	ADJ R-SQ. (%)	RMSE
Africa	71.58	0.096	61.81	0.103
Asia	75.49	0.102	82.55	0.102
Central America	67.73	0.136	84.48	0.131
Europe	80.66	0.116	85.26	0.114
North America	81.08	0.087	83.48	0.086
South America	79.00	0.116	85.29	0.107

From Table 5, we observe that apart from Africa, all other continental models have a comparable or better Adjusted R-square value and RMSE for the Group Lasso model as compared to the MMR. So overall the Group Lasso is a better model.

MODEL FITTING: COUNTRIES

Using the Simple Random Sampling Technique, I picked six of the top ten hit countries in the world according to John's Hopkins University COVID-19 database. The countries are Brazil, Italy, Germany, Spain, the U.S.A and France. I tried to determine which Government response factors were significant to the joint mobility factors in each of the 6 countries, just like was done for continents.

Table 6. Adjusted R-Squares for initial models (%)

Response	USA	Italy	Spain	France	Germany	Brazil
Retail and Recreation	97.33	61.29	80.67	97.39	93.42	79.65
Transit	95.14	72.73	81.84	97.54	95.49	71.68
Workplace	82.45	71.53	74.68	95.37	91.00	31.55

It is evident from Table 6 that the individual model for Mobility change in Workplaces for Brazil is very poor but on average, all the other models perform very well. In the case of Brazil, we could say that the Government responses do not predict Workplace mobility change well but may offer better predictions if the Government responses in Brazil predicted the responses jointly.

Table 7 summarizes all the Government response factors that are significant to the joint mobility factors for each country.

Table 7. Jointly important factors based on Multivariate Multiple Regression

	USA	Italy	Spain	France	Germany	Brazil
School Closing (C1)	Yes	Yes	Yes	Yes	Yes	Yes
Workplace Closing (C2)	Yes			Yes	Yes	Yes
Cancel Public Events (C3)	Yes			Yes	Yes	
Restrictions on Gatherings (C4)						
Close Public Transport (C5)	Yes	Yes		Yes		
Stay at home req. (C6)						
Restrictions on internal movement (C7)					Yes	
International Travel Controls (C8)						
Income Support (E1)	Yes	Yes				
Debt Contract Relief (E2)						
Fiscal measures (E3)		Yes				
International Support (E4)						
Public Information Campaigns (H1)						
Testing Policy (H2)		Yes				
Contact Tracing (H3)						
Emergency investment in healthcare (H4)			Yes			
Investment in vaccines (H5)						

From Table 7, the following could be observed:

1. From the table, it is apparent that across the six countries, most of the government responses have little effect on the joint mobility factors. School closing and workplace closing were significant for all countries except Italy and Spain.
2. Only mobility in Italy and Spain were affected by the Public health system. The public health system was deemed unimportant to mobility changes to the other countries.
3. For all continents, C4, C6, C7, C8, E2, E4, H1, H3 and H5 were not at all significant to mobility changes in retail and recreation, transit stations or the workplace.
4. Mobility in Brazil was affected by only two government responses (C1 and C2), and this represented the lowest out of the six countries.

Group Lasso Regression Model: Countries

Table 8. BEST TUNING PARAMETER FOR EACH MODEL

Model	Best Tuning Parameter (λ)
USA	0.02004496
Italy	0.0113946
Spain	0.02564251
France	0.005094611
Germany	0.001713891
Brazil	0.04583993

Table 8 shows the appropriate λ value needed to minimize the MSE for models built for each continent and Figures 16 ,17 and 18 gives a graphical perspective on the behaviour of each coefficient estimate as the tuning parameter gets very large. Table 9 provides a summary of all the Government response factors significant to the joint mobility factors for the Group Lasso model.

Table 9. Jointly important factors based on Group Lasso Regression

	Africa	Asia	Central America	Europe	North America	South America
School Closing (C1)	Yes	Yes	Yes		Yes	
Workplace Closing (C2)	Yes		Yes	Yes*	Yes	
Cancel Public Events (C3)	Yes	Yes	Yes	Yes	Yes	
Restrictions on Gatherings (C4)	Yes		Yes	Yes	Yes	Yes
Close Public Transport (C5)	Yes		Yes	Yes		
Stay at home req. (C6)	Yes		Yes	Yes		

Restrictions on internal movement (C7)	Yes		Yes	Yes	Yes	
International Travel Controls (C8)	Yes			Yes	Yes*	Yes
Income Support (E1)	Yes	Yes		Yes	Yes	
Debt Contract Relief (E2)	Yes	Yes		Yes		
Fiscal measures (E3)	Yes	Yes		Yes		
International Support (E4)						
Public Information Campaigns (H1)	Yes					
Testing Policy (H2)	Yes		Yes*			
Contact Tracing (H3)	Yes		Yes	Yes	Yes	
Emergency investment in healthcare (H4)	Yes					Yes
Investment in vaccines (H5)			Yes			

The following are my findings from a comparison of the MMR and the Lasso models from Table 9:

1. The Group Lasso is quite different to MMR for countries; only E4 was deemed to be insignificant to mobility in all countries for the Group Lasso.
2. Apart from E4, all economic support measures influenced joint mobility changes in the USA. Apart from H4, all public health responses were significant to explaining changes in mobility for the U.SA. All containment measures were significant in determining mobility changes in the USA.
3. Brazil once again had the lowest number of government responses affecting joint mobility (3 Government responses in all). These government responses were C4, C8 and H4 and this was contrary to the MMR choosing C1 and C2.

The following are the levels of the categorical variables that were deemed important to joint mobility by the Group lasso for each country with the ones not deemed significant added for completeness:

Italy

C6_Stay at home Requirements *Leaving house with minimal exceptions* ✔

C6_Stay at home Requirements *Not leaving house with exceptions* ✖

Spain

C2_Workplace Closing *No measures* ✔

C2_Workplace Closing *Closing some sectors* ✖

Germany

C8_International Travel Controls *Screening* ✔

C8_International Travel Controls *Total Border Closure* ✔

C8_International Travel Controls *No measures* ✖

France

H8_Testing Policy *Having Symptoms and meeting specific criteria* ✔

H8_Testing Policy *Testing anyone showing symptoms* ✖

Model Comparison and Selection

Table 10.

	MMR		GROUP LASSO	
MODEL FOR COUNTRY	ADJ R-SQ (%)	RMSE	ADJ R-SQ. (%)	RMSE
USA	95.60	0.044	78.04	0.036
Italy	72.87	0.165	96.24	0.082
France	97.10	0.060	98.29	0.063
Spain	51.13	0.270	96.56	0.055
Germany	94.77	0.058	98.07	0.075
Brazil	75.00	0.152	69.03	0.146

From Table 10, we observe that apart from USA and Brazil, all the other countries have a comparable or better Adjusted R-square value and RMSE for the Group Lasso model as compared to the MMR. So, generally, the Group Lasso is a better model because it gives us the lowest RMSE for all countries, compared to the MMR but the MMR does a better job at explaining the variability in joint mobility for USA and Brazil.

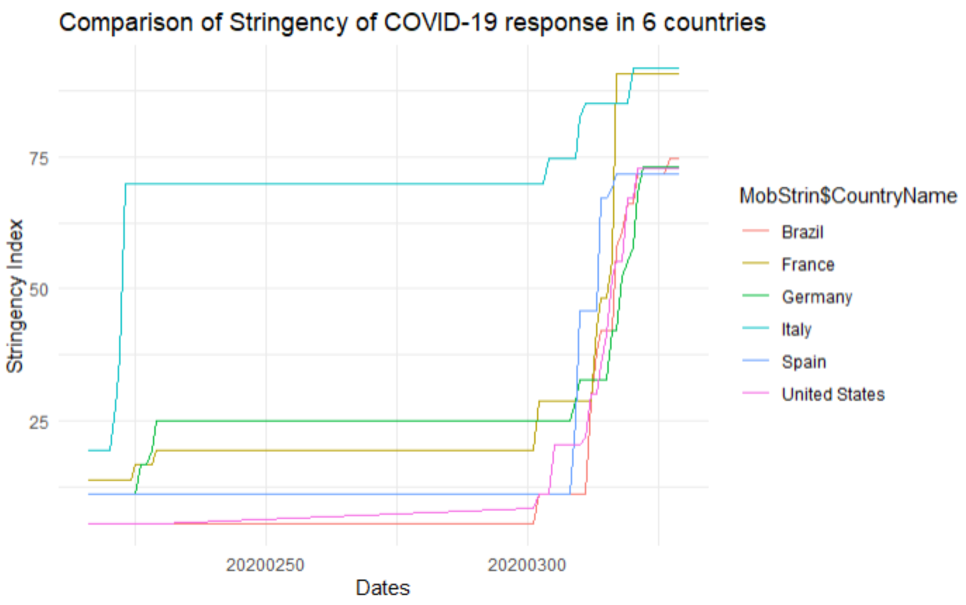
STRINGENCY COMPARISON

Since lockdown procedures vary from country to country, I also compared the stringency of lockdown measures across these countries. I attempted to compare the stringency of responses for the six countries within the periods 16 February 2020 and 29 March 2020 using 3 indices namely:

1. An overall Government Response Index – This index considers how responses of governments have changed over the indicators in the database.
2. A Containment Health Index -This index combines restrictions due to lockdown and closures with other measures like testing policy and investment in vaccines.
3. An Economic Support index – This index measures income support and debt relief levels.

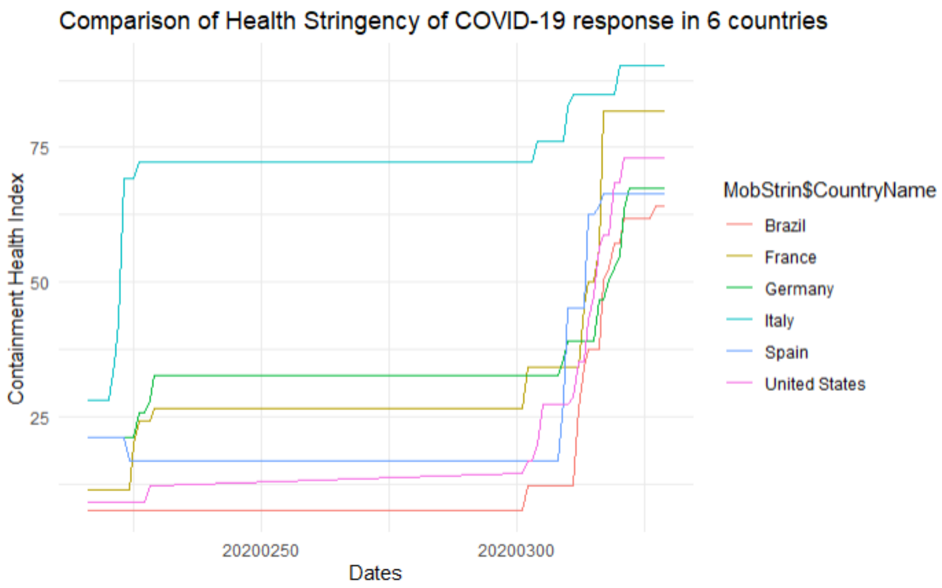
The four indices report a number from 0 – 100 but are however not used as a means of scoring performance levels of the 6 countries. The following graphs are a result of plotting these stringency indices for the 6 countries:

Figure 13



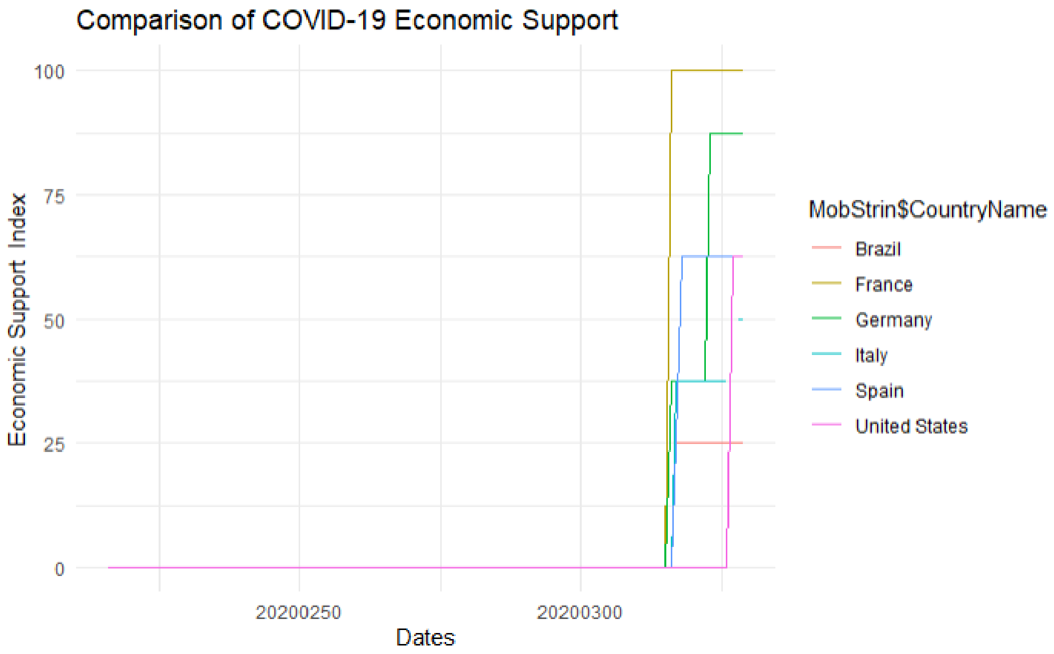
From Figure 13, it is apparent that overall stringency of COVID-19 response in the six countries grew over the period. Brazil had the weakest stringency response until March when the stringency of the responses became very steep. The U.S.A, just like Brazil were not stringent enough in their response to the pandemic, although they were slightly better at it than Brazil. Stringency however grew rapidly as the number of cases increased exponentially in the U.S.A. Italy, one of the countries to be deeply affected by the coronavirus in the early stages of the pandemic, showed a steep increase in stringency in the early days of February and this remained constant until early March where stringency grew in a stepwise manner.

Figure 14



From Figure 14, Health stringency of COVID-19 generally grew across the six countries, but Italy once again increased faster than any other country in the early stages, became constant and increased in a stepwise manner from early March onwards. Once again, the U.S.A and Brazil were slow out of the blocks in terms of health stringency, but this also increased as the number of cases grew world-wide. Spain's Health stringency index interestingly decreased slightly after being constant for a short period, became constant after that and steeply increased in March. Initially, Germany was more responsive than France but as time went on France overtook Germany in terms of stringency of health responses.

Figure 15



From the graph, it is apparent that economic support was virtually non-existent for all six countries until after mid-March, where some economic support became available in France, Italy, Germany, Spain, and Brazil. The U.S.A was the last country to offer any kind of economic support but eventually overtook Brazil in terms of the magnitude of economic support given.

CONCLUSION AND DISCUSSION

The project sought to find out whether there was any correlation between people's perception of risk posed by COVID-19 and the responses provided by governments across countries. The study also sought to compare the effectiveness of stringencies across countries. It was evident that across all continents, the risk perception of COVID-19 was not affected in any way by Fiscal Measures. However, as governments across all continents tightened their containment measures, mobility reduced drastically. Based on the best model (Group Lasso) picked, Asia had most government responses affecting mobility to recreational centres, transit stations and workplaces. From the results obtained after running similar tests for 6 of the top 10 worst hit countries by the pandemic, the United States of America had more than half of the government responses affecting mobility, with majority of the responses being containment measures. Mobility in the other countries used were similarly affected by the containment measures. The stringency plot also told an interesting story. The U.S.A had weak stringency at the beginning stages of the pandemic but picked up as the number of cases grew in America. Italy had the tightest stringency among all the six countries, and this was as expected as the death rate was very high.

There can be further study done into other different modelling techniques to increase the predictive power of the linear. Also, since this study was conducted over a short period (16 February to 29 March), the period of study could be extended to reflect the current situation and then observations can be made as to how the studies compare to each other.

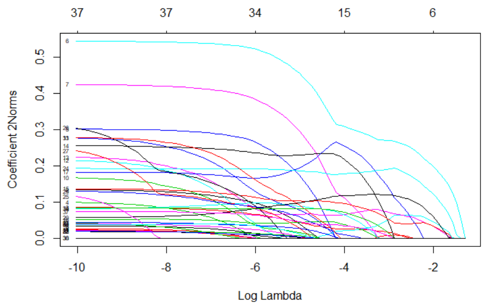
REFERENCES

1. Getting Started with Multivariate Multiple Regression (October 27, 2017). Retrieved from: **https://data.library.virginia.edu/getting-started-with-multivariate-multiple-regression/** on 29 May 2020
2. Understanding Diagnostic Plots for Linear Regression Analysis. (September 21, 2015). **Retrieved from: https://data.library.virginia.edu/diagnostic-plots/** on 27 May 2020
3. (n.d.) Retrieved from: **https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker** on 3 Jun 2020
4. Black Americans Dying at three times the rate of White People. (20 May, 2020). Retrieved from: **https://www.theguardian.com/world/2020/may/20/black-americans-death-rate-covid-19-coronavirus** on 5 Jun 2020
5. Johnson R. A. & Wichern D. W. (2007). Applied Multivariate Statistical Analysis. Prentice Hall.
6. Exploratory Data Analysis in R for Beginners (Oct 9, 2019) Retrieved from : https://towardsdatascience.com/exploratory-data-analysis-in-r-for-beginners-fe031add7072 on 14 June 2020
7. (n.d.) Retrieved from: https://rdrr.io/cran/tidyr/man/spread.html on 14 Jun 2020
8. R ggplot2 Histogram. (n.d.) Retrieved from: https://www.tutorialgateway.org/r-ggplot2-histogram/ on 19 Jun 2020
9. Top 2 coronavirus hit countries are nearly 12 lakh cases apart: COVID-19 global tally. (Jun 13 2020). Retrieved from: https://www.hindustantimes.com/world-news/us-and-brazil-top-2-coronavirus-hit-countries-are-nearly-12-lakh-cases-apart-covid-19-global-tally/story-CDiVdFBUq4U7sz1curvgxN.html on 7 July 2020.
10. (n.d.) Retrieved from: https://stackoverflow.com/questions/6572119/r-solvesystem-is-exactly-singular on 7 July 2020.
11. (n.d.) Retrieved from: https://r.789695.n4.nabble.com/Can-t-invert-matrix-td3052137.html on 7 July 2020.
12. Oxford Uni academics launch a tracker for COVID-19 policy interventions. Retrieved from: https://techcrunch.com/2020/03/25/oxford-uni-academics-launch-a-tracker-for-covid-19-policy-interventions/#:~:text=The%20tool%2C%20called%20the%20Oxford%20COVID-19%20Government%20Response,the%20stringency%20of%20policy%20responses%20around%20the%20world. On 8 July 2020.
13. (n.d.) Retrieved from: http://www.sthda.com/english/wiki/ggplot2-line-plot-quick-start-guide-r-software-and-data-visualization on 8 July 2020.
14. (n.d.) Retrieved from: https://datasharkie.com/how-to-test-for-normality-in-r/#:~:text=Shapiro-Wilk%20test%20in%20R%20Another%20widely%20used%20test,sample%20of%20observations%20came%20from%20a%20normal%20distribution. On 17 July 2020
15. Applied Statistics with R. (n.d.) Retrieved from: https://daviddalpiaz.github.io/appliedstats/transformations.html on 15 July 2020.
16. (n.d.) Retrieved from: https://stats.stackexchange.com/questions/247869/how-to-test-for-independence-of-residuals-in-linear-model on 31 July 2020
17. (n.d.) Retrieved from: https://boostedml.com/2019/03/linear-regression-plots-fitted-vs-residuals.html on 1 August 2020.
18. (n.d.) Retrieved from: https://www.statisticssolutions.com/autocorrelation/ on 13 October 2020

19. (n.d.) Retrieved from: https://www.statology.org/differences-between-anova-ancova-manova-mancova/ 15 October2020
20. (n.d.) Retrieved from: https://www.statology.org/polynomial-regression-r/#:~:text=How%20to%20Perform%20Polynomial%20Regression%20in%20R%20The,score%20%28response%20variable%29%20for%20a%20certain%20college%20course on 10 November 2020.
21. (n.d.) Retrieved from: https://www.sciencedirect.com/science/article/abs/pii/S0167947305001623 on 12 November 2020
22. (n.d.) Retrieved from: https://stats.stackexchange.com/questions/188664/how-incorrect-is-a-regression-model-when-assumptions-are-not-met on November 18 2020
23. (n.d.) Retrieved from: https://statmodeling.stat.columbia.edu/2019/08/21/you-should-usually-log-transform-your-positive-data/#:~:text=Non-normality%20is%20relatively%20unimportant%3B%20at%20worst%20you%20just,LOWEST%20priority.%20My%20experience%20supports%20this.%203%2F3%20pic.twitter.com%2FR0BfQCoxdK on December 1 2020
24. (n.d.) Retrieved from: https://stats.stackexchange.com/questions/2492/is-normality-testing-essentially-useless on December 2 2020
25. (n.d.) Retrieved from: https://stats.stackexchange.com/questions/163642/what-to-do-if-residuals-are-not-normally-distributed on December 4 2020

APPENDIX

Figure 16 AFRICA



ASIA

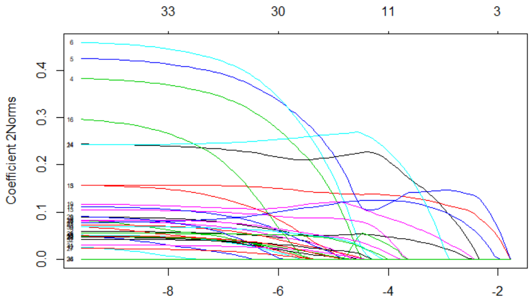
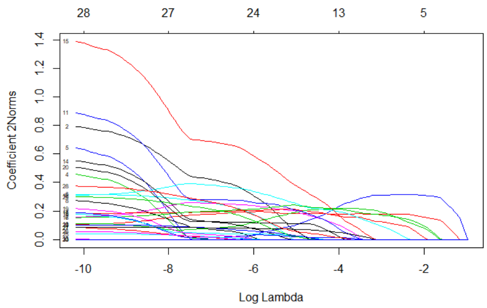


Figure 17 CENTRAL AMERICA



EUROPE

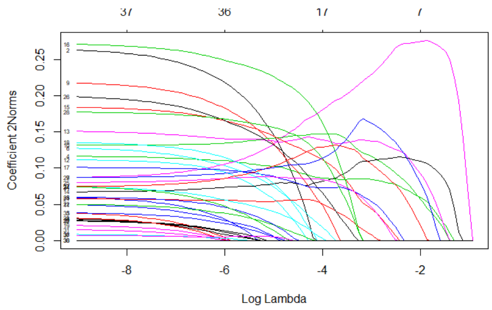
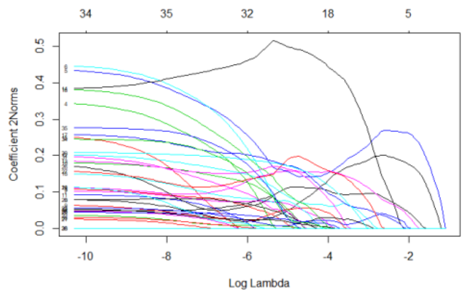


Figure 18 NORTH AMERICA



SOUTH AMERICA

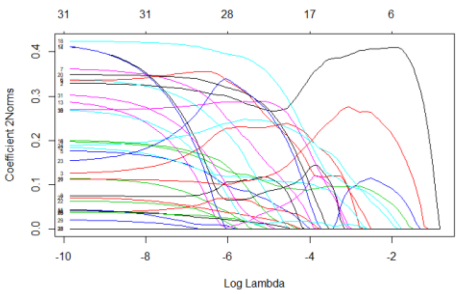
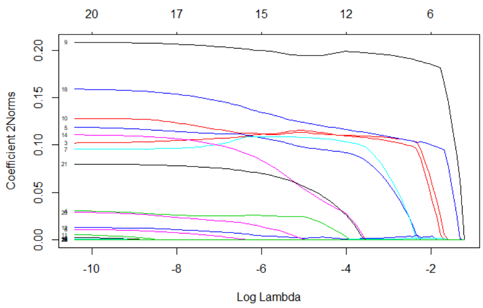


Figure 19 USA



ITALY

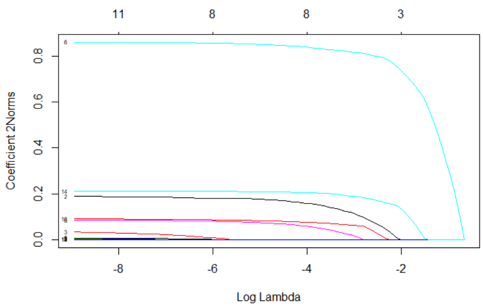
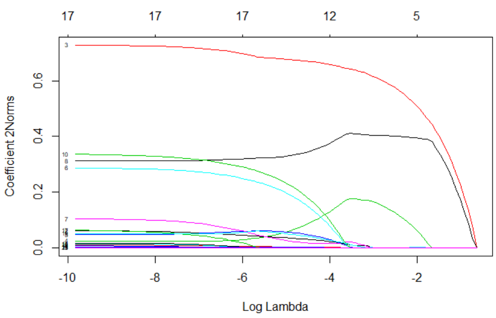


Figure 20 FRANCE



GERMANY

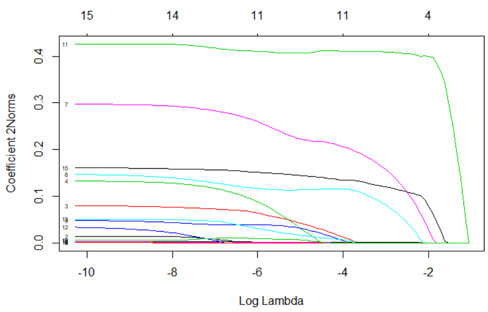
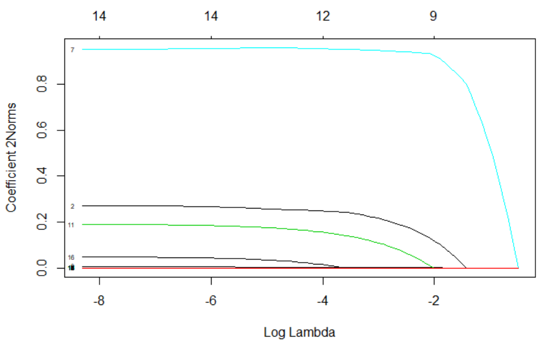


Figure 21 SPAIN



BRAZIL