STEM Fellowship Big Data Challenge 2020

The Impacts of International, Domestic, and Local Mobility on COVID-19 Deaths

Cathy Yan[1], Simran Pal Waraich[2], and Liza Babaoglu[3]

[1]University of British Columbia
[2]Simon Fraser University
[3]University of Toronto

July 4, 2020

Abstract

As coronavirus disease 2019 (COVID-19) spread across the world, countries moved to restrict travel and mobility at the international, domestic, and local levels. However, these new policies have devastated the global tourism industry and been met with resistance from people. Due to the unprecedented nature of this global crisis, investigating the efficacy of these restrictions is important for future decision making by policymakers and individuals. Using datasets from the United Nations World Tourism Organization (UNWTO), Google, Apple, and John Hopkins, the relationships between mobility and the spread of COVID-19 was evaluated using linear regression and K-means clustering. We found that having fewer people enter the country correlated with fewer COVID-19 deaths, while visitations to local indoor public places were associated with a higher number of COVID-19 deaths. However, no conclusions were drawn between death rates and trips taken within a country, the mode of transportation used for local travel, or outings to parks. These findings, for the most part, support restrictions on international travel and stay-at-home orders, but will hopefully provide insights and opportunities for more nuanced guidelines that support economic and social needs without compromising public health.

Keywords
Coronavirus, Travel, Tourism, Big Data, Machine Learning

1 Introduction

Figure 1: Timeline of major travel restrictions with select landmark events from January 2020 to March 2020 as reported by the New York Times [1]

The global tourism industry is valued at 7.6 trillion USD and employs 1 in 10 workers [2]. However, due to coronavirus disease 2019 (COVID-19), it is expected to decline by as much as 80 percent over 2020 [3]. This is because travel to and from most countries have been limited to essential trips due to the highly transmissible nature of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [4]. Despite the efforts of governments to close international borders (Figure 1), COVID-19 was declared a global pandemic in March [5]. As of May 28, 2020, there have been more than 5.8 million coronavirus cases and more than 350,000 deaths worldwide [6, 7].

Tourism-centric countries such as Mexico, Spain, and the USA, among others, have been severely impacted economically as tourism contributes to more than 10 percent of their GDP [8, 9]. Additionally, restrictions on international travel have also been accompanied by regional orders to limit outings to essential trips only. These directives have faced resistance from citizens [10], as well as criticisms from government officials [11]. Therefore, in order to balance economic prerogatives and public health, we sought to quantify the strength of the relationship between travel in its various forms, and the spread of COVID-19.

In this study, datasets from the United Nations World Tourism Organization (UNWTO) were used to investigate trends in international and domestic tourism before travel bans took place [12, 13, 14]. Google's COVID-19 Community Mobility Report and Apple's Mobility Trends Report were utilized to analyze local mobility after lock-downs and restrictions [15, 16]. Data regarding cases of COVID-19 were accessed from Johns Hopkins' COVID-19 repository [17]. Understanding the relationships between COVID-19 and various forms of travel can help governments and communities make informed decisions about mobility-oriented guidelines.

2 Materials & Methods

2.1 Datasets Used

The Coronavirus disease 2019 (COVID-19) dataset [17] is sourced from a repository managed by the Johns Hopkins University Center for Systems Science and Engineering. It reports the total number of confirmed cases, recoveries, and deaths for each country on a day-to-day basis between January 22nd, 2020 to the present day. To account for differences in the quality and availability of tests, the number of deaths was deemed the defining metric for the prevalence of COVID-19.

Datasets for unemployment [18], poverty [19], sanitation [20], population density [21], total population [17], median age [22, 23], life expectancy [24], and income per person [25] were acquired from Gapminder. The most recent reported values were used for the analyses.

Tourism-related datasets were provided by the United Nations World Tourism Organization. Three were used for this paper: the number of inbound tourists [12], the number of outbound tourists [14], and the number of domestic trips [13]. Inbound tourists are non-resident visitors arriving at national borders, outbound tourists are people departing for other countries, and domestic trips are trips taken by visitors within a country. Data were available for the years 2014 to 2018 for each country. The most recent reported values were used and countries with no information were omitted from analyses.

Two datasets were used for assessing the impacts of local mobility. The first was the COVID-19 Community Mobility Report [15] compiled by Google. It communicates changes in visits to locations labelled as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential areas on a day-to-day basis from February 15th, 2020 to May 16th, 2020. It uses zero as the baseline, so a decrease in visitations would be reported as a negative value, while an increase would be positive.

The second was the Mobility Trends Report provided by Apple [16]. It reports changes in requests for directions from January 13th, 2020 to the present day. Requests are categorized as walking, transit, or driving. One hundred is used as the baseline, so a decrease in requests would be expressed as a value lesser than one hundred while an increase would be greater. For consistency with the previously described datasets, country-level data were used for both local mobility datasets. Values between March 1st, 2020 and April 30th, 2020 were averaged for each country.

2.2 Libraries Used

The libraries used in Python (Version 3.6.5) were Matplotlib 3.2.1, NumPy 1.14.3, Pandas 0.22.2, Scikit-learn 0.22.2.post1, SciPy 1.4.1, Seaborn 0.10.1, and Statsmodels 0.11.1.

The libraries used in R (Version 4.0.0) and RStudio (Version 1.2.5042) were ggplot2, caret and randomForest.

2.3 Feature Selection

COVID-19 deaths were expressed as a rate of units per 100,000 people. The following seven socioeconomic factors were merged with COVID-19 deaths into a dataframe for 140 countries: unemployment, poverty, sanitation, population density, median age, life expectancy, and income. Only the 51 countries with no missing values were analyzed. A random forest model was built utilizing the caret and randomForest packages, where the dependent variable was the COVID-19 deaths per 100,000 people and the independent variables were the socioeconomic factors. The importance function was used in order to extract the variable importance measures produced by the random forest model. Then, the Increase in Node Purity, also known as the Mean Decrease Gini, indices were analyzed. To visualize the importance of the features, the importance measures were plotted using the ggplot2 package, and the most correlated features were selected.

2.4 Trends in International Tourism

COVID-19 deaths, the number of inbound tourists, and the number of outbound tourists were expressed as a rate of units per 100,000 people. The natural logarithm of these values was taken and will be referred to as log-deaths, log-inbound, and log-outbound. Simple linear regression was performed to quantify relationships between log-inbound and log-deaths, and log-outbound and log-deaths using LinearRegression from Scikit-learn. The data were split into training (70%) and testing (30%) sets using train_test_split from Scikit-learn. R-values and p-values were calculated using linregress from SciPy.

To improve the model, socioeconomic factors determined through feature selection were incorporated as factors. The natural logarithm of these values was taken and multiple linear regression was performed with log-inbound using LinearRegression from Scikit-learn for log-deaths. The data were split into training (70%) and testing (30%) sets using train_test_split from Scikit-learn. P-values were calculated using OLS from Statsmodels.

The relationship between inbound and outbound tourists in the context of COVID-19 deaths was then examined using K-means clustering. The rates of inbound and outbound tourists were used as factors for the KMeans algorithm (random_state = 2) in Scikit-learn. The mean COVID-19 death rate for each cluster was graphed using Seaborn. A one-way ANOVA was performed on the means using f_oneway from SciPy to generate the p-value, and Tukey's Range Test was done using pairwise_tukey from Statsmodels. Cluster classification and the logarithmic values of the socioeconomic factors were used as features for multiple linear regression for log-deaths.

2.5 Trends in Domestic Tourism

A similar analytical progression as above was used to evaluate the impacts of domestic tourism on COVID-19 deaths. The number of domestic trips taken was converted to a rate of units per 100,000 people, and the natural logarithm was taken (log-domestic). Simple linear regression was performed to quantify the relationship between log-domestic and log-deaths. Next, the natural logarithms of the socioeconomic factors were included as additional factors for multiple linear regression.

2.6 Comparing the Impacts of International and Domestic Tourism

To analyze the relationship between international and domestic tourism in the context of COVID-19 deaths, values for inbound tourism and domestic tourism were used for K-Means clustering. Analyses were performed in the same way as described in section 2.4.

2.7 Trends in Local Travel

Using the data in Google's Mobility Reports, how local, routine trips contribute to COVID-19 deaths was evaluated. Each column represents changes in visitations to one of the following places: retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential areas. Exploratory data analysis was done using Tableau, where the colour of the country represents the deviation from the baseline for each location category, and the size of the overlaid circles represents COVID-19 deaths per capita. Data for each location category were then used as factors for multiple linear regression, performed as described in section 2.4, to assess the relationship between mobility and COVID-19 deaths.

Next, Apple's Mobility Trends Report was used to determine how the mode of transportation used - walking, transit, or driving - contributed to COVID-19 deaths. All three were first used as factors for multiple linear regression, performed as described in section 2.4, in order to confirm if mobility in general was correlated with COVID-19 deaths. Next the ratios of driving to walking, driving to transit, and transit to walking were calculated. The natural logarithm of each was taken and used to perform a linear regression to appraise the relationship between the mode of transportation and COVID-19 deaths.

3 Results

3.1 Feature Selection

3.1.1 Life expectancy and income per person are most highly correlated with COVID-19 death rates out of the examined socioeconomic factors

The relationship between COVID-19 deaths per 100,000 people and the following seven attributes were analyzed: unemployment, poverty, sanitation, population density, median age, life expectancy, and income. As seen in Table 1, Increase in Node Purity (IncNodePurity), also known as the Mean Decrease Gini, is a measure of variable importance calculated by the splits in random forest trees. The higher IncNodePurity signifies a greater feature importance. The random forest model revealed that the most significant features impacting the COVID-19 death rates were life expectancy, income and population density. These features were used alongside tourism and mobility datasets to impose a realistic relationship.

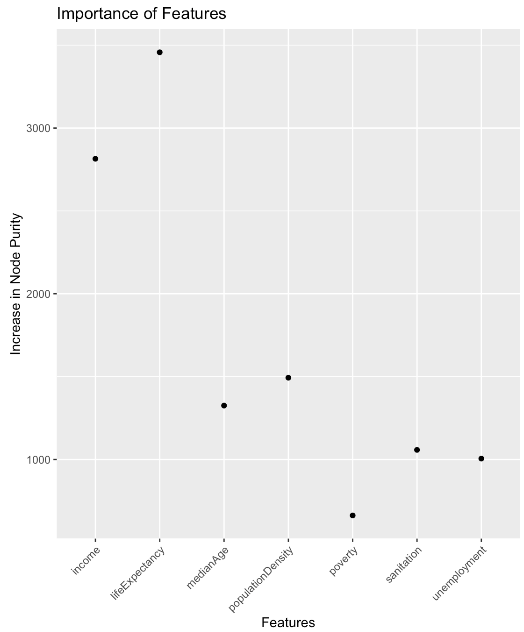

Figure 2: Increase in Node Purity for unemployment, poverty, sanitation, population density, median age, life expectancy, and income

Features	IncNodePurity
Unemployment	1005.3234
Poverty	662.2159
Life expectancy	3457.1474
Population density	1493.5164
Median age	1325.2598
Income	2814.6201
Sanitation	1057.8809

Table 1: Increase in Node Purity for unemployment, poverty, sanitation, population density, median age, life expectancy, and income

3.2 Trends in International Tourism

3.2.1 Inbound tourism is significantly correlated with COVID-19 deaths, but outbound tourism is not.

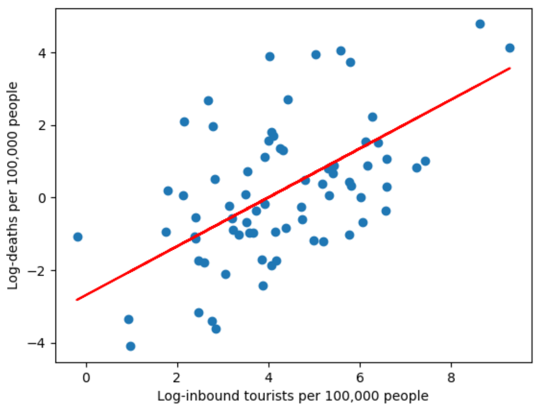

Figure 3: Scatter plot and regression line for log-COVID-19 deaths per 100,000 people vs log-inbound tourists per 100,000 people (coefficient: 0.57, R = 0.53, p = 1.28 e-4)

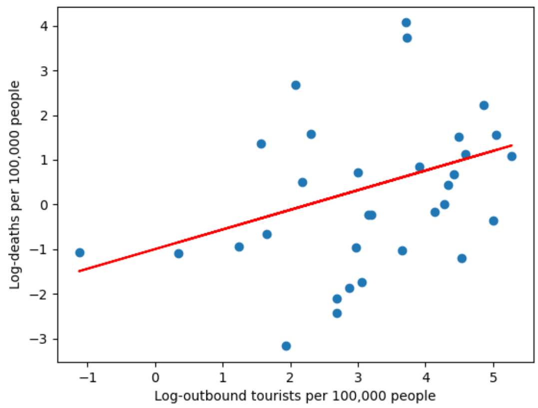

Figure 4: Scatter plot and regression line for log-COVID-19 deaths per 100,000 people vs log-outbound tourists per 100,000 people (coefficient = 0.39, R = 0.34, p = 0.058)

Simple linear regression revealed a significant, positive correlation between inbound tourism rates and COVID-19 death rates (Figure 3), but no significant correlation between outbound tourism rates and COVID-19 death rates (Figure 4). Moving forward with inbound tourism rates; life expectancy, and income per person were incorporated as additional factors to account for demographic and economic disparities between countries.

Using multiple linear regression, it was found that all three features have a significant effect on COVID-19 death rates (Table 2).

Feature	p-value	Coefficient
Inbound tourism	1.28×10^{-3}	0.57
Income per person	5.77×10^{-8}	0.40
Life expectancy	8.99×10^{-7}	1.03
Summary	4.30×10^{-6}	

Table 2: Feature-specific and summary p-values from multiple linear regression model (R-squared = 0.332, adjusted R-squared = 0.303)

3.2.2 There is no significant difference in COVID-19 death rates between countries clustered by levels of inbound and outbound tourism after accounting for socioeconomic factors

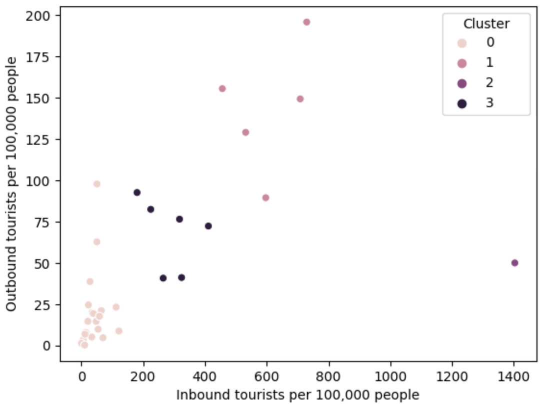

Figure 5: Scatter plot of inbound tourists per 100,000 people vs outbound tourists per 100,000 people coloured by cluster

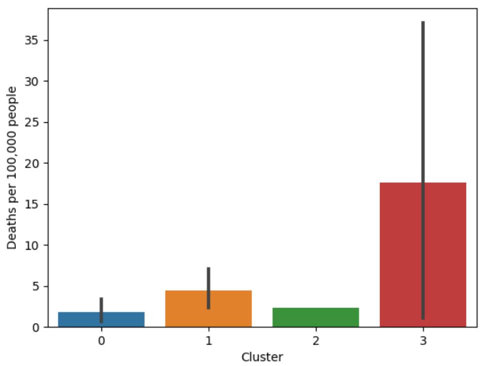

Figure 6: Bar chart of mean COVID-19 death rates for each cluster (error bars represent 95% CI, ANOVA p-value = 0.047

For K-means clustering, a k-value of 4 was deemed optimal by the inertia graph for k-values between 1 and 9. The clusters generated were low inbound-low outbound, moderate inbound-moderate outbound, high inbound-high outbound, and high inbound-low outbound

Group 1	Group 2	p-adj	Reject
0	0	0.90	False
0	2	0.9	False
0	3	0.027	True
1	2	0.90	False
1	3	0.25	False
2	3	0.59	False

Table 3: Results from Tukey's Range Test for comparing the mean COVID-19 death rates between clusters

(Figure 5). A one-way ANOVA revealed that at least two clusters have significantly different COVID-19 death rates from each other (Figure 6). Tukey's Range Test established that countries with moderate levels of inbound and outbound tourism have significantly higher COVID-19 death rates than countries with low rates of inbound and outbound tourism (Table 3).

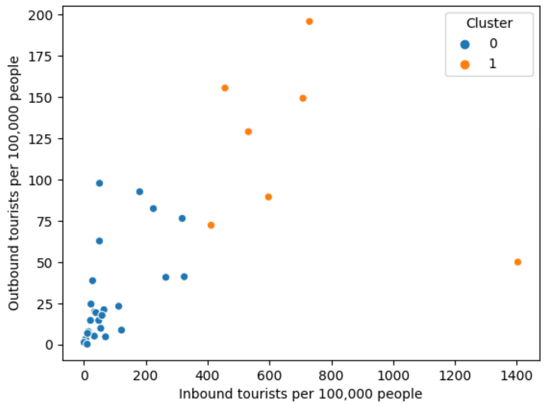

Figure 7: Scatterplot of inbound tourists per 100,000 people vs outbound tourists per 100,000 people coloured by cluster

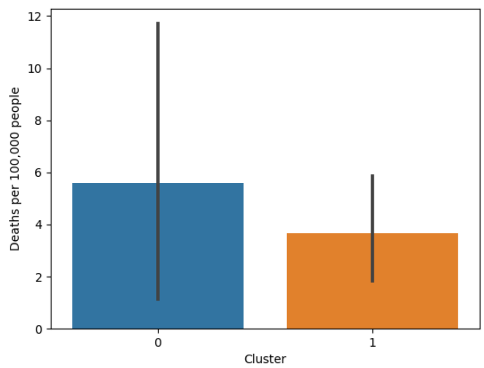

Figure 8: Bar chart of mean COVID-19 death rates for each cluster (error bars represent 95% CI, ANOVA p-value = 0.71)

K-means clustering was performed again with k=2 after visual inspection of Figure 4. The clusters generated were countries with lower levels of both inbound and outbound tourism, and countries with higher levels of either one or both (Figure 7). There was no significant difference in mean COVID-19 death rates between the two clusters (Figure 8).

Feature	p-value	Coefficient
Cluster assignment	0.966	-0.03
Income per person	0.002	1.33
Life expectancy	0.002	-2.96
Summary	0.068	

Table 4: Feature-specific and summary p-values from multiple linear regression model (R-squared = 0.330, adjusted R-squared = 0.260)

Using multiple linear regression, it was found that, even when socioeconomic factors are accounted for, there was no significant correlation between cluster assignment and COVID-19 death rates (Table 4).

3.3 Trends in Domestic Tourism

3.3.1 Domestic tourism is significantly correlated with COVID-19 deaths, but not when socioeconomic factors are taken into account

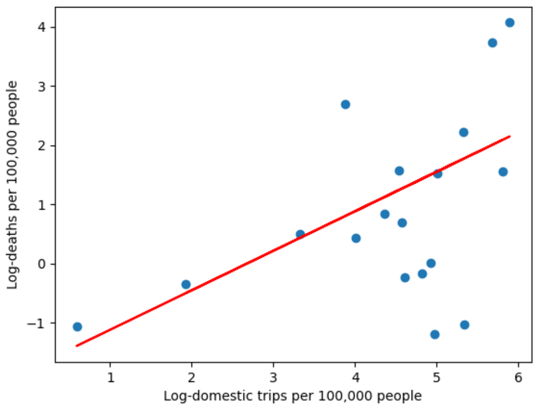

Figure 9: Scatterplot and regression line for log-COVID-19 deaths per 100,000 people vs log-domestic trips per 100,000 people (coefficient: 0.55, R = 0.47, p = 0.047)

Simple linear regression revealed a significant, positive correlation between the rate of domestic trips taken and COVID-19 death rates (Figure 9). Life expectancy, and income per person were once again incorporated as additional factors.

Using multiple linear regression, it was found that none of the four factors within the model significantly correlated with COVID-19 death rates (Table 5).

Feature	p-value	Coefficient
Domestic trips rate	0.184	0.45
Income per person	0.700	0.28
Life expectancy	0.553	-0.92
Summary	0.081	

Table 5: Feature-specific and summary p-values from multiple linear regression model (R-squared = 0.415, adjusted R-squared = 0.298)

3.4 Comparing the Impacts of International and Domestic Tourism

3.4.1 There is no significant difference in COVID-19 death rates between countries categorized by rates of inbound tourism and domestic trips

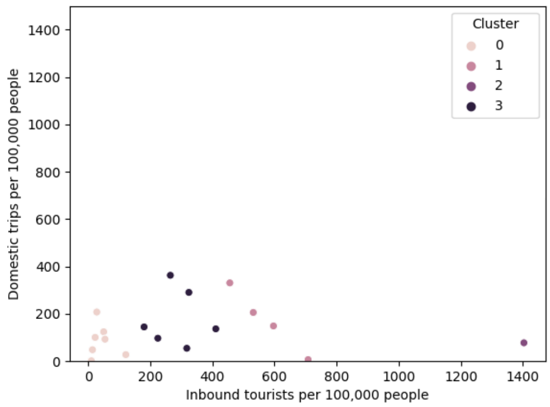

Figure 10: Scatterplot of inbound tourists per 100,000 people vs domestic trips per 100,000 people coloured by cluster

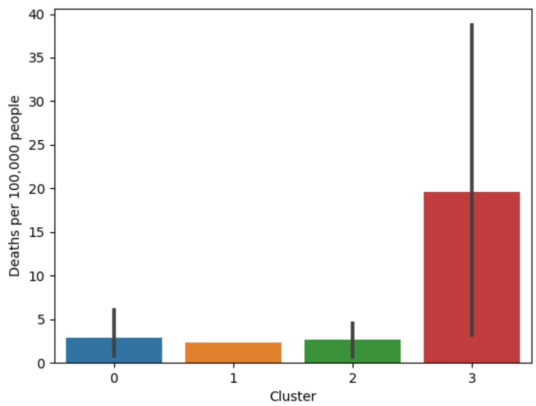

Figure 11: Bar chart of mean COVID-19 death rates for each cluster (error bars represent standard deviation, ANOVA p-value = 0.23)

For K-Means clustering, a k-value of 4 was deemed optimal by the inertia graph for k-values between 1 and 9. Each of the four clusters generated span a wide range of values for domestic trips, but a narrow range of values for inbound tourists (Figure 10). A one-way ANOVA revealed that there is no significant difference in COVID-19 death rates between the four clusters (Figure 11).

3.5 Trends in Local Travel

3.5.1 Changes in local travel are significantly correlated with COVID-19 death rates

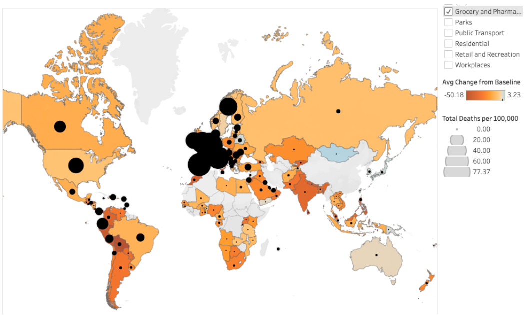

Figure 12: Changes in visitations to groceries and pharmacies, and COVID-19 deaths per capita by country

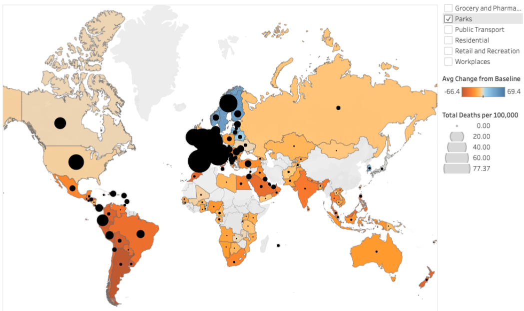

Figure 13: Changes in visitations to parks, and COVID-19 deaths per capita by country

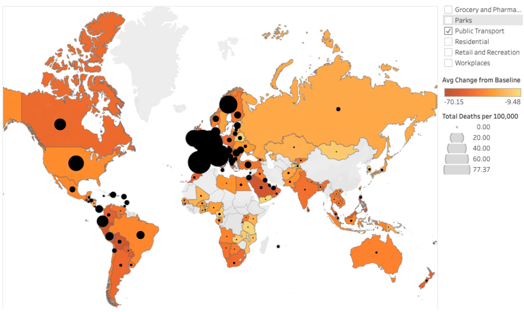

Figure 14: Changes in visitations to public transit stations, and COVID-19 deaths per capita by country

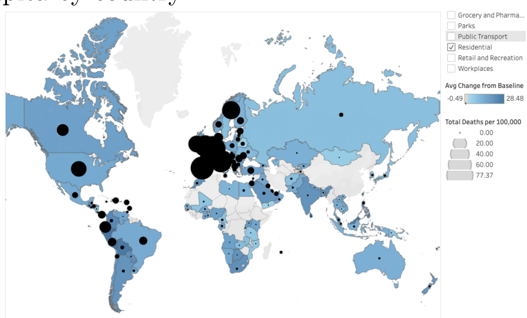

Figure 15: Changes in visitations to residential areas, and COVID-19 deaths per capita by country

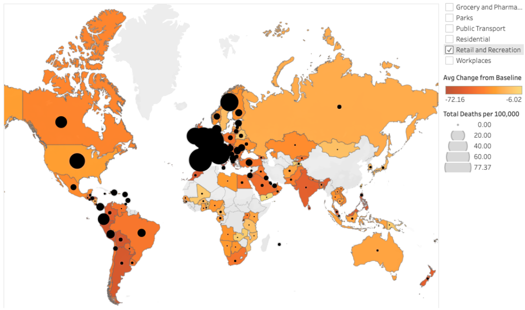

Figure 16: Changes in visitations to retail and recreation, and COVID-19 deaths per capita by country

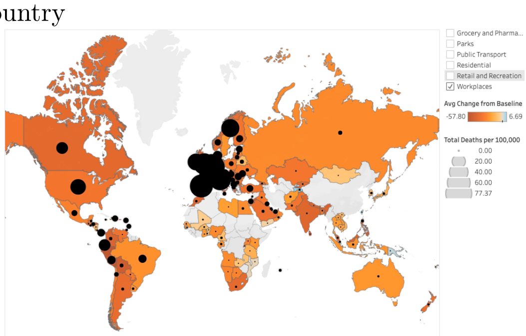

Figure 17: Changes in visitations to workplaces, and COVID-19 deaths per capita by country

Thus far, only tourism-related movement at the international and country levels have been considered. To explore the impact travel within communities has on COVID-19 death rates, changes in local mobility and COVID-19 death rates were mapped (Figures 12-17, Link to Interactive Map) [26]. Darker orange colours indicate decreases in visitations to that location type while darker blue colours represent increases. The larger the diameter of the overlaid circle, the more COVID-19 deaths per capita. It was found that, in general, visitations to all public and shared spaces have decreased, while the number of people in residential areas, presumably at home, have increased.

Feature	p-value	Coefficient
Recreation & retail	0.010	5.31
Grocery & pharmacy	0.037	0.0063
Parks	0.41	-0.47
Transit stations	0.011	-7.47
Workplaces	0.010	0.678
Residential areas	0.056	1.86
Summary	0.021	

Table 6: Feature-specific and summary p-values from multiple linear regression model (R-squared = 0.691, adjusted R-squared = 0.523)

Multiple linear regression was then performed. It was found that changes in visitations to all locations except for parks and residential areas significantly affects COVID-19 death rates (Table 6). The overall model significantly correlates the 6 features with COVID-19 death rates.

3.5.2 The mode of transportation used for local travel is not correlated with COVID-19 death rates

Feature	p-value	Coefficient
Driving	0.437	-0.65
Transit	0.909	-0.10
Walking	0.209	1.04
Summary	0.015	

Table 7: Feature-specific and summary p-values from multiple linear regression model (R-squared = 0.415, adjusted R-squared = 0.298)

To explore the role the mode of transportation plays in COVID-19 deaths, multiple linear regression was performed. No correlation was found between any individual method and COVID-19 deaths per capita, but the three factors combined were significantly correlated with COVID-19 death rates (Table 7).

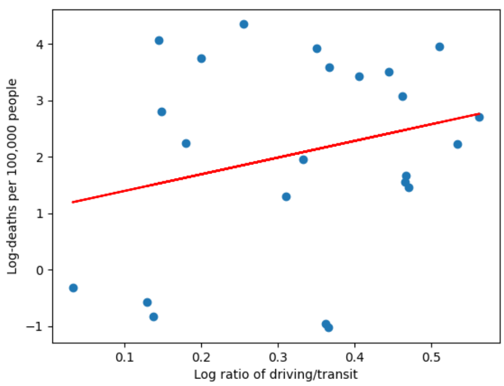

Figure 18: Scatterplot and regression line for log-COVID-19 deaths per 100,000 people vs log-ratio of driving/transit (coefficient: 2.95, R = 0.25, p = 0.24)

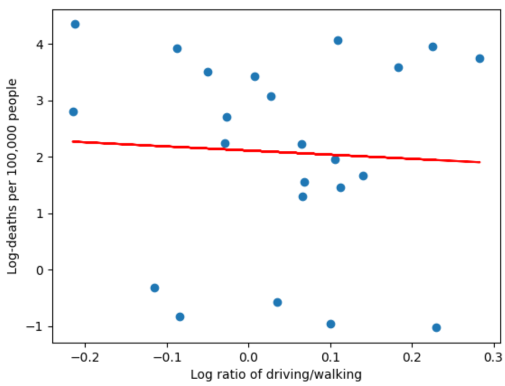

Figure 19: Scatterplot and regression line for log-COVID-19 deaths per 100,000 people vs log-ratio of driving/walking (coefficient: -0.73, R = -0.05, p = 0.80)

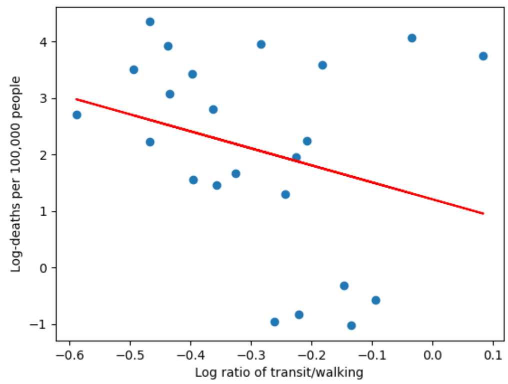

Figure 20: Scatterplot and regression line for log-COVID-19 deaths per 100,000 people vs log-ratio of transit/walking (coefficient: -3.00, R = -0.28, p = 0.20)

Next, the ratios between the types of mobility (driving to transit, driving to walking, and transit to walking) were calculated, and a linear regression was performed with COVID-19 death rates for each individually. No correlation was found between preferentially choosing one mode over another and COVID-19 death rates (Figures 18-20).

4 Discussion

As tourism becomes more convenient and affordable, more people are travelling internationally and domestically more than ever before. However, with greater mobility comes the greater likelihood of pandemics, where pathogens are transported as quickly and globally as people. As COVID-19 spread, governments and regions enacted lock-downs on various scales, preventing the entrance and/or exit of people (Figure 1). However, these measures have faced criticisms from economists [27], politicians [28], and health officers [29] for their inefficacy. To evaluate the validity of these measures, we studied the relationship between COVID-19 death rates and mobility on three different scales: international, domestic, and local.

To account for socioeconomic disparities between countries, a feature selection process was implemented to determine which socioeconomic factors were most highly correlated with COVID-19 death rates. Originally, we planned on incorporating three factors. However, after seeing the calculated IncNodePurity scores (Figure 2 and Table 1), only the top two features were selected. This is because the score for the second-best feature, income per person, was twice the score of the third-best feature (population density). The top two features, life expectancy and income per person were logically expected to associate closely with COVID-19 death rates. The risk of serious illness is known to increase with age [30]. Additionally, it is speculated that people in lower-income brackets are more susceptible to COVID-19 [31]. Altogether, these two factors are suitable for inclusion in future models.

On the international level, it was found that countries with a higher rate of inbound tourists have a higher rate of COVID-19 deaths (Figure 3). The model retained significance even when income per person, and life expectancy were incorporated (Table 2). However, since the R-squared and adjusted R-squared values are roughly equal and low, the model is incomplete. Interestingly, the rate of outbound tourists did not significantly correlate with COVID-19 death rates despite being another metric for international mobility (Figure 4). The reason for this could be that not all outbound tourists return to their countries within the viral incubation period, and if infected, are accounted for in the country they are visiting instead. However, these results were confounded by a smaller dataset available compared to inbound tourism. Nevertheless, these initial findings support the closing of international borders to incoming visitors, but there is no evidence to suggest that preventing people from leaving the country has any effect on COVID-19 spread within that country.

Since inbound tourism and outbound tourism impact COVID-19 deaths with different levels of significance, it was inferred that their relationship with each other was not as linear as assumed. Following that line of reasoning, it was predicted that they could be independent of one another, and countries could partake in them to different extents (i.e. countries could have a high volume of inbound tourists, but a low volume of outbound tourists and vice versa). It was hypothesized that countries would fall into one of four clusters: low inbound-low outbound, high inbound-low outbound, low inbound-high outbound, and high inbound-high outbound. With K-means clustering, it was found that countries with moderate levels of inbound and outbound tourism have significantly higher COVID-19 death rates than countries with low rates of inbound and outbound tourism (Figure 6 and Table 3). However, based on the initial linear model, the cluster with the highest rates of inbound and outbound tourism should have the highest death rates.

Additionally, these findings are questionable as the clusters were not in each corner of the graph as anticipated (Figure 5). Instead, it seemed that having two clusters, one for low inbound-low outbound and one for everything else, would be more appropriate (Figure 7). For this model, there was no significant difference in mean COVID-19 death rates between the two clusters (Figure 8). Despite this, it was suspected that a trend might emerge once socioeconomic factors were included in the analysis since inbound and outbound tourism are, to a certain extent, reflective of economic status. This turned out to not be the case (Table 4), enforcing that there is no significant difference in the mean death rates between these two clusters. Additionally, these findings and the distribution of the clusters suggest that clustering countries based on inbound and outbound tourism rates may not be an appropriate method.

At the domestic level, it was found that countries with more domestic trips taken have a higher rate of COVID-19 deaths (Figure 9). However, when socioeconomic factors are added to the analysis, the model loses significance (Table 5). Comparing the results from this dataset to the larger one used for inbound tourism, it is likely that the lack of available data for domestic trips is a confounding factor. Known correlations between socioeconomic factors and COVID-19 death rates are not accurately represented by this data. Therefore, despite the promising nature of restricting domestic mobility, no conclusions can be drawn.

The domestic trips taken into account are only trips by visitors (ie. inbound tourists). As inbound tourism rates are significantly correlated with COVID-19 death rates and domestic trips rates are not, it prompted an inquiry into the mean death rates of countries clustered based on inbound tourism and domestic trips. The rationale for this analysis was the same as for clustering by inbound and outbound - despite the presumption of linear relation, these two factors impact COVID-19 death rates to different extents, suggesting a degree of independence.

It was hypothesized that countries would fall into one of four clusters: low inbound-low domestic, high inbound-low domestic, low inbound-high domestic, and high inbound-high domestic. However, there was no clear relationship between the location of the clusters (Figure 10), and there was no significant difference in the mean COVID-19 death rates (Figure 11). The clusters each spanned the same range of values for the rate of domestic trips, suggesting that the data is not conducive to being sorted in this manner. As a result, the fact that there was no significant difference in mean death rates was not surprising given that the clusters themselves were not representative of the hypothesized relationships for which this analysis was devised. As a result, no conclusions can be drawn about how domestic trips rates and inbound tourism rates contribute independently to COVID-19 death rates.

Additionally, to investigate the efficacy of stay-at-home orders, we evaluated how changes in visitations to different types of places have impacted COVID-19 death rates. By mapping the data, it was found that, across the world, visitations to public places have decreased and more people are located in residences (Figures 12-17). Quantitatively, every factor was significantly correlated with COVID-19 deaths except for parks and residential areas, and only visitations to transit stations were significantly negatively correlated with COVID-19 death rates (Table 6). Regarding parks, the lack of correlation is potentially due to the slower spread of the virus outdoors since there is more space to practice physical distancing [32]. Although it was expected that staying home would have a significant negative correlation with COVID-19 death rates, as the values for residential areas do not differentiate between one's own home and the home of others, the results are likely confounded. Overall, though, these findings demonstrate that stay-at-home orders are effective at curbing the spread of COVID-19, but that access to parks can be allowed.

Finally, Visitations to transit stations were unexpectedly significantly negatively correlated with COVID-19 death rates, which inspired an investigation into whether transit is safer somehow than other modes of transportation. This notion is contrary to the advice of infectious diseases experts and epidemiologists, who recommend driving a personal vehicle, or walking or biking, instead of taking public transport to decrease risks of infection [33]. However, multiple linear regression revealed only that general mobility, and not any specific mode of transportation, is correlated with COVID-19 death rates (Table 7). Although this fact is valuable as it supports our findings from Google's dataset, it does not provide insights into whether one mode is safer than another.

To better answer that question, we calculated the ratios between different forms of transportation, and performed linear regressions with COVID-19 death rates. It was expected that, if preference for one mode (ex. driving) over another (ex. transit) was associated with lower levels of viral infections, there would be a significant correlation between the respective ratio (driving/transit in this case) and COVID-19 death rates. However, there was no significant correlation between any of the ratios and COVID-19 death rates (Figures 18-20). These results, though, are limited by the fact that the dataset only reports changes in requests for directions. Consequently, trips to familiar locations are not taken into account, and there is no guarantee that the person actually travelled along that route. Therefore, it can only be said that there is no correlation between the intended mode of travel to a destination and COVID-19 deaths.

Conclusions

The purpose of this study was to evaluate the efficacy of lock-downs at the international, domestic, and local levels in reducing the spread of COVID-19. Inbound tourism was found to be significantly positively correlated with increased COVID-19 deaths, but outbound tourism was not. Any significant correlation between domestic trips and COVID-19 deaths disappears when socioeconomic factors are accounted for. Visitations to local indoor public places are significantly correlated with COVID-19 deaths, but the mode of transportation used to get there is not. These findings overall support the actions taken to restrict mobility in an effort to slow the pandemic.

Currently, countries across the world are at different stages in the implementation of mobility restrictions. However, although this paper has demonstrated that reducing mobility can inhibit the spread of COVID-19, cultural norms and misinformation have led to push-back from citizens. To combat this, based on our research, governments should encourage the use of outdoor spaces to alleviate frustration and push campaigns to emphasize the importance of travelling only for essential purposes. Furthermore, regions lifting restrictions, from our findings, need to carefully monitor for new outbreaks and implement other procedures to inhibit spread.

However, further research is needed to determine what types and level of completeness of restrictions are required. As international, domestic, and local mobility are essential to economies and people's access to essential goods and services, limitations need to be strategically devised and implemented. Evaluating how mobility restrictions work in tandem with other behaviour-oriented legislation such as social distancing and wearing face coverings in curbing the spread of COVID-19 will be the next step in making more effective and nuanced recommendations. Comprehensively understanding how the movement of people at different levels impacts the spread of the virus is essential to combating the pandemic.

Acknowledgements

We would like to thank our mentor Sean La for his support and guidance in this project. Additionally, we would also like to thank STEM Fellowship for hosting this event and providing us with the opportunity to engage in data science research.

References

[1] Derrick Bryson Taylor. How the coronavirus pandemic unfolded: a timeline, Feb 2020.

[2] Daala. 10 facts about global tourism industry, Aug 2018.

[3] International tourist numbers could fall 60-80 percent in 2020, unwto reports, Jun 2019.

[4] Anna Smith. What are the early symptoms of coronavirus (covid-19)?, May 2020.

[5] Ryan Tumility. It's official: 'deeply concerned' who declares the covid-19 outbreak a global pandemic, Mar 2020.

[6] Joshua Berlinger and Brett McKeehan. Coronavirus pandemic: Updates from around the world. may 28 coronavirus news, May 2020.

[7] Coronavirus updates, May 2020.

[8] Dorothy Neufeld. Visualizing the countries most reliant on tourism, May 2020.

[9] A Yu Aleksandrova. Typology of countries of the world according to the development level of international tourism. *Geography and Natural Resources*, 37(1):18–25, 2016.

[10] Nicole Mortillaro. Mixed messages, frustration with lockdowns fuel some skepticism about pandemic, May 2020.

[11] Joe Nocera. Lockdown critics may have some valid points, May 2020.

[12] World Tourism Organization. All countries: Inbound tourism: Arrivals 2014 - 2018, 2020.

[13] World Tourism Organization. All countries: Domestic tourism: Arrivals 2014 - 2018, 2020.

[14] World Tourism Organization. All countries: Outbound tourism, 2020.

[15] Google. Community mobility reports, 2020.

[16] Apple. Covid-19 - mobility trends reports.

[17] Datasets. datasets/covid-19, May 2020.

[18] Gapminder. Aged 15+ unemployment rate, Jan 2017.

[19] Gapminder. Extreme poverty, percentage of people below 3.20 dollars a day, Jan 2018.

[20] Gapminder. Basic sanitation overall access, Jan 2017.

[21] Gapminder. Population density per square km, Jan 2020.

[22] World population prospects - population division, May 2020.

[23] Gapminder. Median age, Jan 2020.

[24] Gapminder. Life expectancy at birth, Jan 2020.

[25] Gapminder. Gdp per capita, constant ppp dollars, Jan 2020.

[26] Cathy Yan. Covid-19 deaths and changes in mobility by country, Jun 2020.

[27] Abigail Klein Leichman. Lockdown only made corona crisis worse, claim experts, Apr 2020.

[28] Sam Jones. Spain's path out of covid lockdown complicated by polarised politics, May 2020.

[29] Fraser Nelson. Norway health chief: lockdown was not needed to tame covid, May 2020.

[30] Older adults, Jun 2020.

[31] Abby Vesoulis. Coronavirus may disproportionately hurt the poor-and that's bad for everyone, Mar 2020.

[32] Jane Seyd. Risk of transmitting covid-19 outdoors is 'negligible' says health officer, Apr 2020.

[33] Ania Bessonov and Andrew Culbert. Planes, trains and automobiles. what's the safest way to travel domestically? your covid-19 questions answered | cbc news, Jun 2020.