Analysis of Government Policy’s Effect on COVID-19

Elias Karikas
Malick Tobe
Roberto Moreno
Siddharth Saraf
Spencer Ong

Northeastern University, Boston, MA, USA

Introduction

The COVID-19 pandemic has affected and changed all of our lives - from how we work to how we interact with one another. With the virus constantly transforming our society, government responses seem to play a key role in managing the pandemic. Despite the scientific proof, there is still a lot of public discourse around different covid policies (i.e. if facial coverings or border closures are effective in controlling the amount of COVID-19 cases). With this project, we are interested in examining the effectiveness of different government policies in each country and how much of a key role each policy played in preventing the spread of COVID-19.

Through data analysis, we hope to find policy effectiveness, and if certain indicators like wealth have an effect on COVID outcome. We believed that our findings would ultimately support tighter restrictions, showing a negative correlation between more strict policy and cases. If successful, our research could help support general discourse surrounding the pandemic, and potentially encourage individuals or countries to support certain policies.

Data Sources and Methods

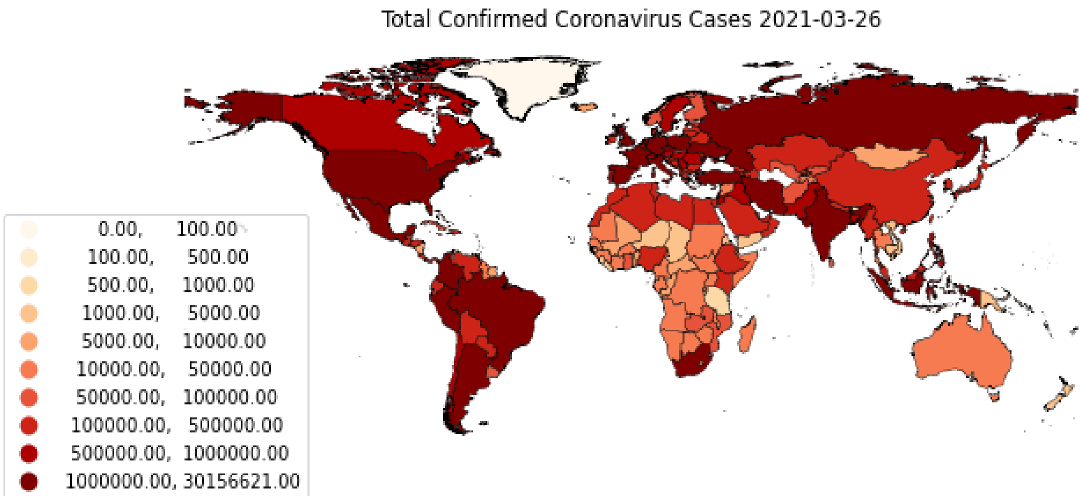
As we were searching for presentation topics, we got increasingly interested in conducting analysis on COVID-19 data. We found files from Our World in Data that had different governmental policies that immediately interested us. The files containing policy, ranked each country's policy on a day to day basis using a 0-4 index (0 being most lax, 4 being most strict) and starter code that contained source COVID-19 data. There are five CSV files we used in our analysis. The first file we used has the overall micro and macro data, displaying all countries by date with their total/new cases, total/new deaths, population densities, HDI, and various other metrics [1]. The other files we used contained dates, country data, and an index measuring various metrics such as: work/school closures for each country [2], face covering policies [3], covid testing or contract tracing data [4], and international travel restrictions [5]. For reference, the index range for a metric such as international travel restrictions is 0 - no measures, 1 - screenings, 2 - quarantine from high risk countries, 3 - ban on high risk regions, and 4 - total border closure.

Since our data was sourced from a site that was already using the CSV's for analysis, data cleaning was fairly simple. First we needed to convert our data to Pandas Dataframes, and merge various index files with our COVID-19 source code [1], removing or replacing N/A values depending on their column. For example, if a column contained an N/A on new cases, it was clear that the country either did not register a new case or it was 0, so it was replaced with a 0. The majority of our data cleaning was conducted by simplifying our source COVID-19 data set [1] per need. This could be only keeping columns on certain identifiers such as new cases, or deaths, or isolating columns.

Moreover, we required an additional dataset which contained the countries' shapes and boundaries in order to visualize the world with geopandas. This was quite a tough job since there were not many files containing such data and the ones that did were not really that compatible. Nonetheless, we came across some files in ArcGis Hub which contained the necessary information [9]. There were certain discrepancies regarding the country names between the datasets, but we were able to make the necessary adjustments.

Analysis and Results

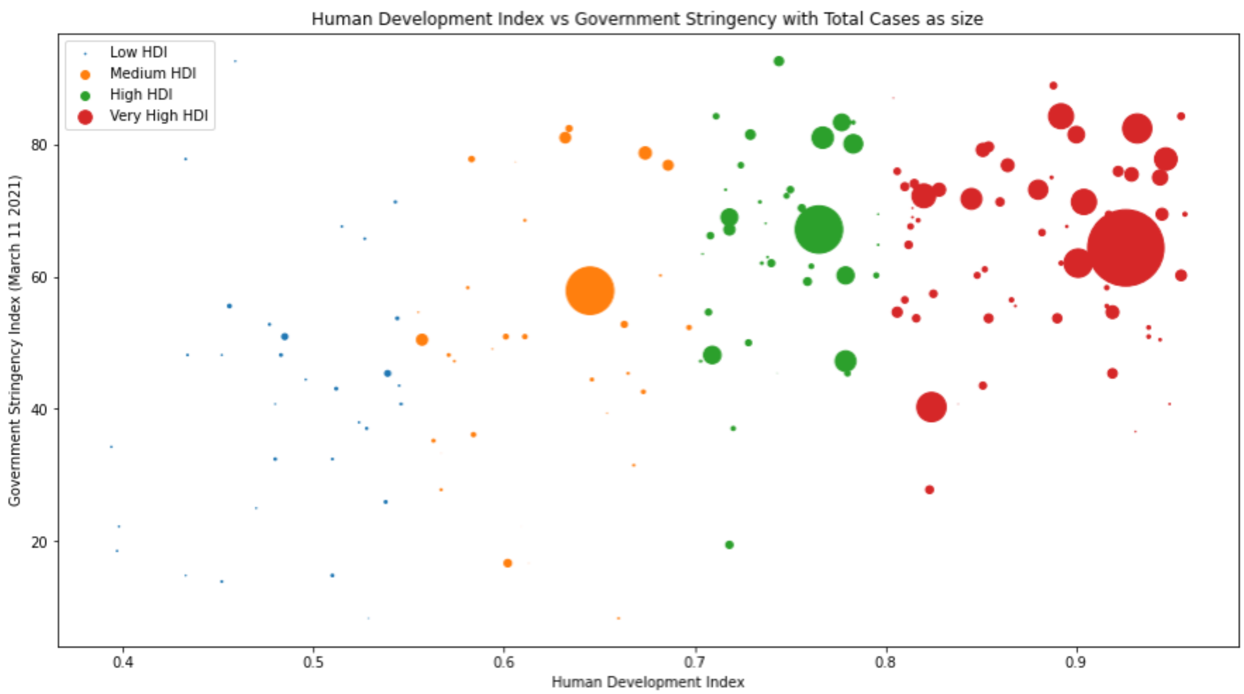
COVID-19 Around the World



The coronavirus has affected over 215 countries and territories all around the globe. Overall, it has impacted some countries more than others. Continents such as the Americas and Europe, as well as some parts of Asia, show higher rates of coronavirus confirmed infections than Africa and Oceania. There seems to be a geographical trend for the spread of the virus and the perceived unevenness among continents appears to suggest that core economic locations are hit the hardest. In this sense, we could take a further look into factors such as development and government stringency which may have the potential to point out how the pandemic plays out in different areas.

Human Development Index Analysis

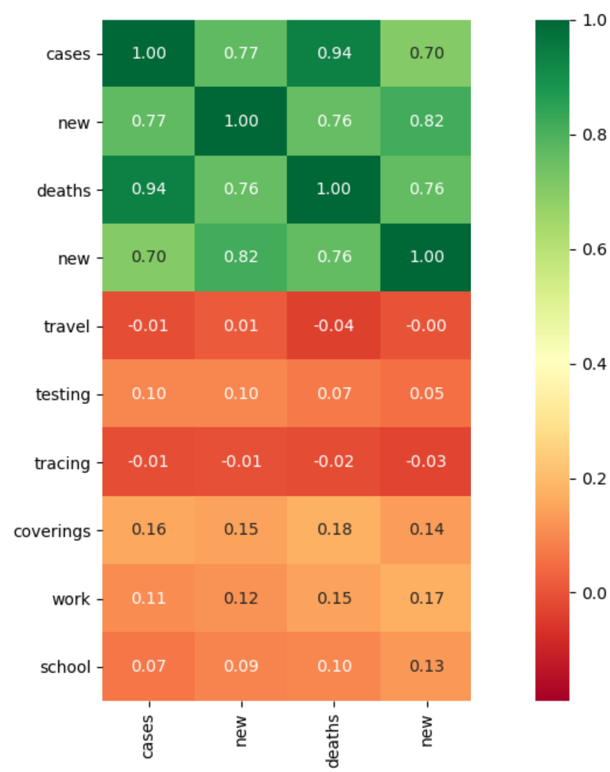
Given this information, we wanted to figure out if the wealth disparity between nations in the world actually played a significant part in the difference between government responses to the pandemic. To conduct our analysis, we used two different indexes: the government stringency index, and the human development index (HDI). The HDI divides the world in 4 different categories: low (0.000 to 0.549), medium (0.550 to 0.699), high (0.700 to 0.799), and very high (0.800 to 1.000) HDI. The higher the HDI, the more developed a country is said to be. We used filtering methods to group countries by HDI categories in different data frames, which makes them easier to distinguish on a visualization. To add more depth, we scaled the size of the scattered points to the number of total cases for each country.



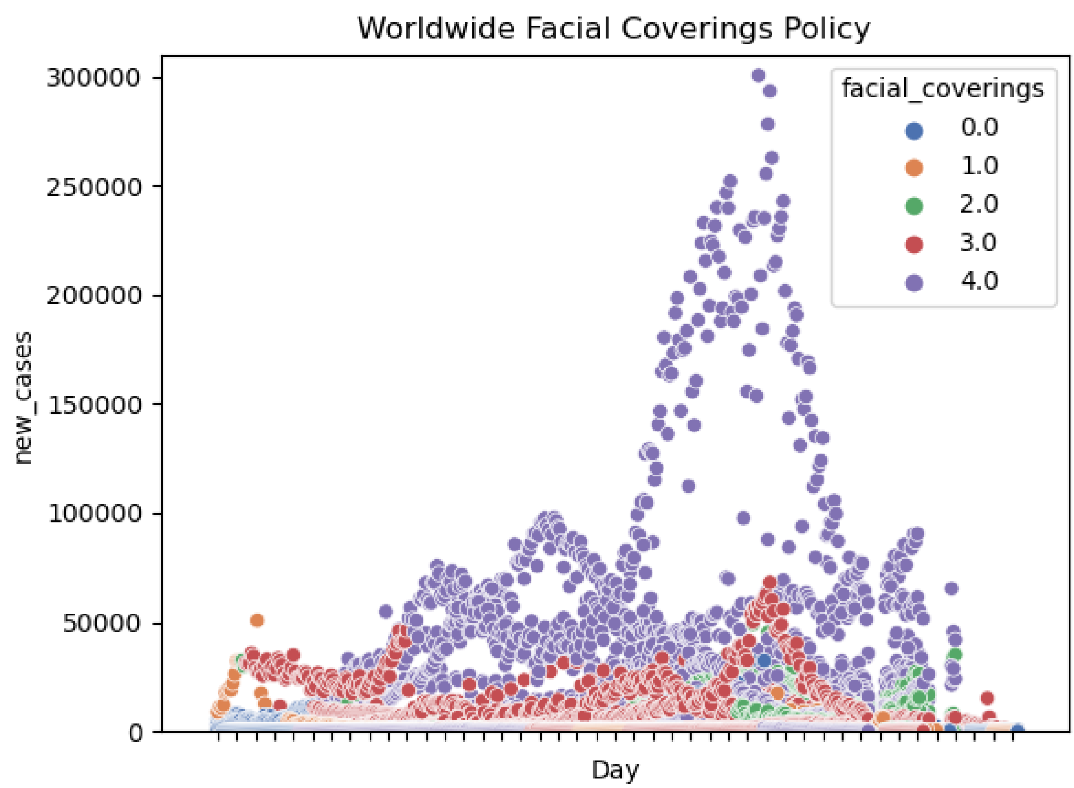
The number of total cases and the government stringency index are measure on March 11th, 2021, which is exactly one year after the World Health Organization announced the COVID-19 outbreak to be a global pandemic. With our plot result, we were able to draw important conclusions concerning the relationship between a nation's development and government response to the crisis: countries with High & Very High HDI tend to enforce stricter measures than countries with low & medium HDI, and countries with low HDI tend to have lower number of total cases than countries with high & very high HDI. Limits to our visualization include the small number of cases for low HDI countries which would make us tend to think that they are handling the situation better than high HDI countries, but that's not quite the case. Since a lot of low HDI countries economies rely on exports of goods, and border closures were enforced by high HDI countries, their exports were slower which led to their economies taking a significant hit [8]. Furthermore, the precariousness of public health services coupled with the low amount of government resources explain the small amount of tests conducted, which ultimately results in an underestimated amount of cases.

COVID-19 General Policy Analysis

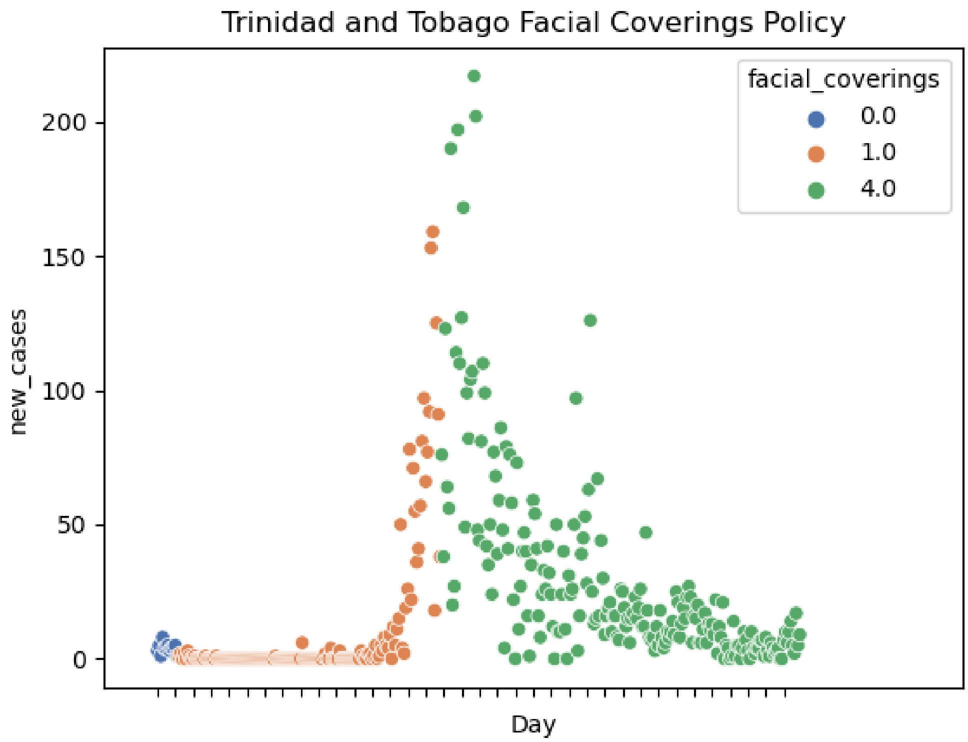
We also wanted to see the effect of government policy on COVID outcomes. For this goal it was important to find relative correlations for each variable of government policy (travel restrictions [5], severity and level of testing/contact tracing [4], public policy requiring face masks [3], and work/school closures [2]) and case data [1] (in this case we simplified this data to new/total cases and death to focus on basic COVID data and public policy). This was important to see if there was any relationship between certain variables and case data.



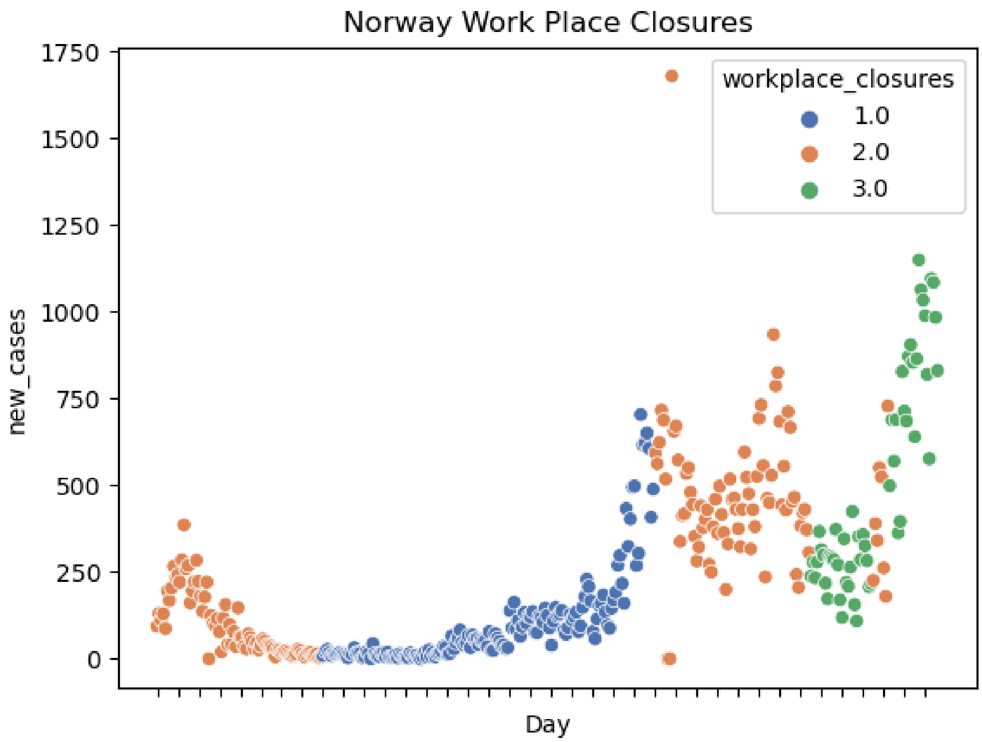
The difficulty here was that no policy was statistically significant when compared to case data. However, since we were using more than a year of data for over 210 countries, it was unrealistic to assume that this would be the case. Many countries had different outbreaks at different periods, irrespective of policy. Because of this, we then focused on the policies that had the highest correlation's relative to policy. Since facial coverings had correlations from .14-.18, this is where we focused first.



As we can see by the graph, more severe facial coverings policy saw more new cases for those countries. However, there is a fairly good explanation for this. By using a function that found the countries with the highest correlations between facial coverings and cases, I was able to find Trinidad and Tobago, which had a correlation of .94 when comparing facial coverings to new_cases data.

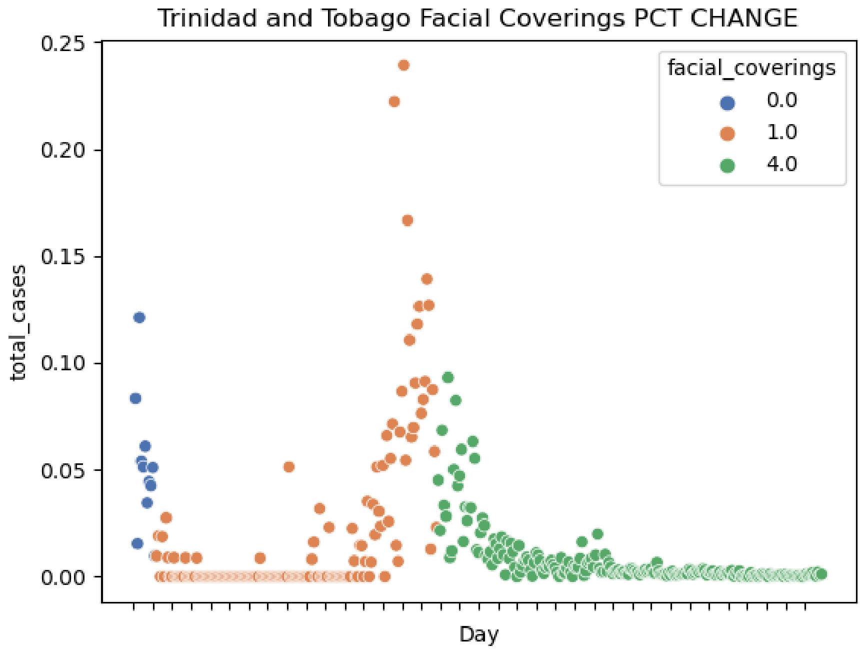


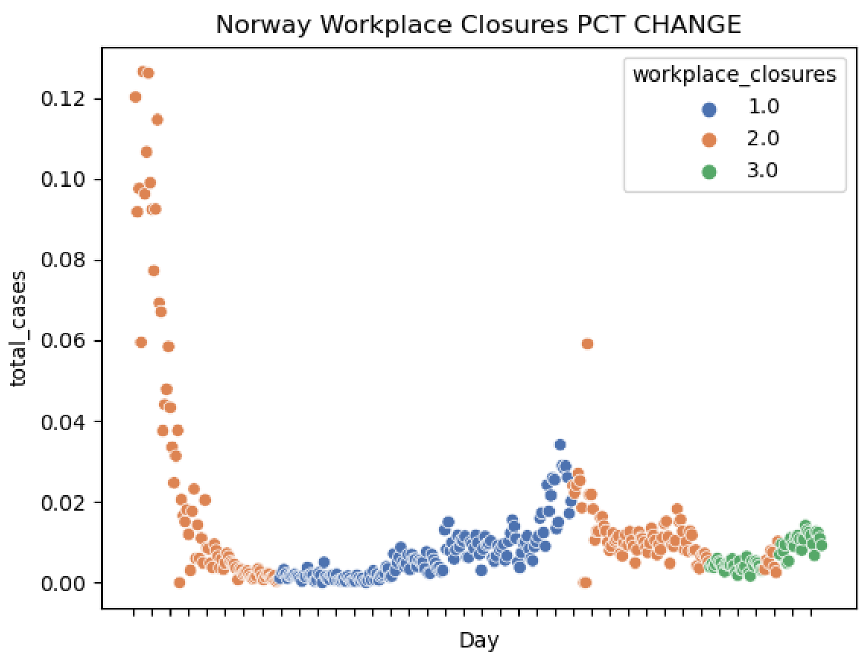
In the case of Trinidad and Tobago, a severe mask policy was adopted on the week of September 1st, as the country threatened fines from $1,000-$3,000 to those who didn't adhere to it [6]. Trinidad and Tobago, and many other countries like it have implemented severe mask policies only when cases have increased significantly, thus showing a positive correlation between facial covering policy and new cases. This was also the case for work closures. Norway has a correlation of .74 when comparing workplace closures to new cases.



Likewise, the severity of closures typically increase when spikes in cases occur, thus driving overall correlation between the two variables upwards. Many countries, but not all, followed the same ideology, being reactive in policy rather than being proactive.

A possible consideration would be to find the change in percentage of total cases, so as to not have misstated positive correlations by changes in policy occurring after spikes in COVID cases. The purpose of this is to see how effective a policy was following a change in severity, rather than seeing the correlation of when a policy was enacted.

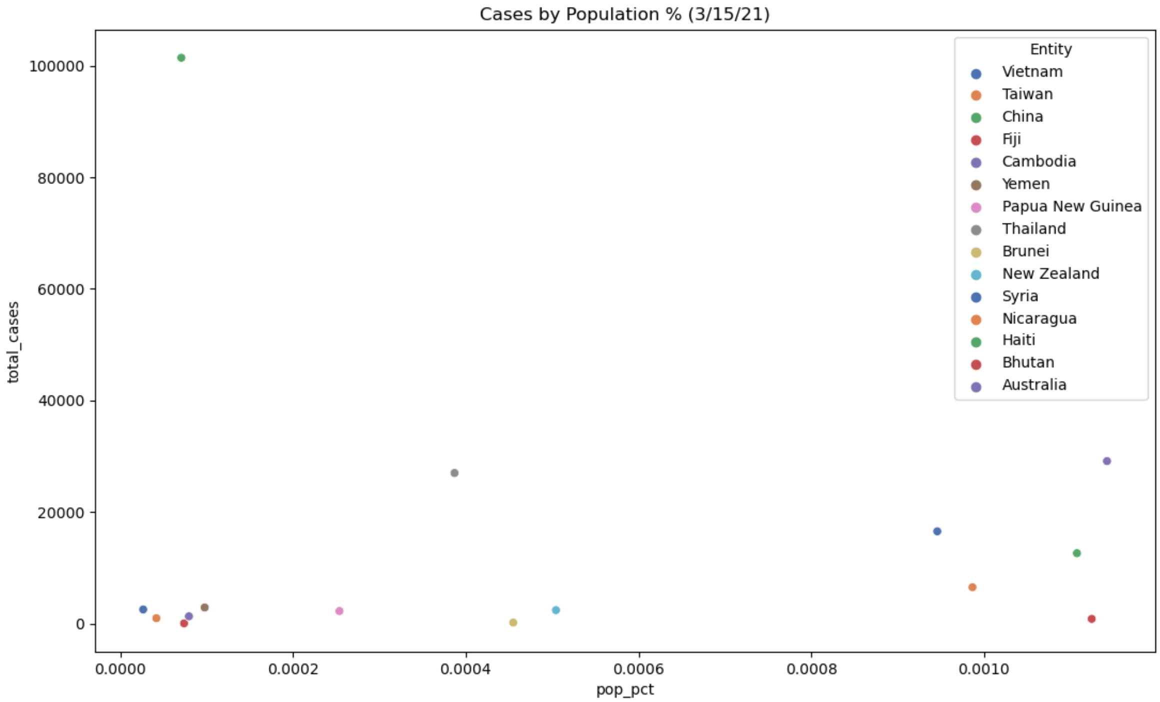


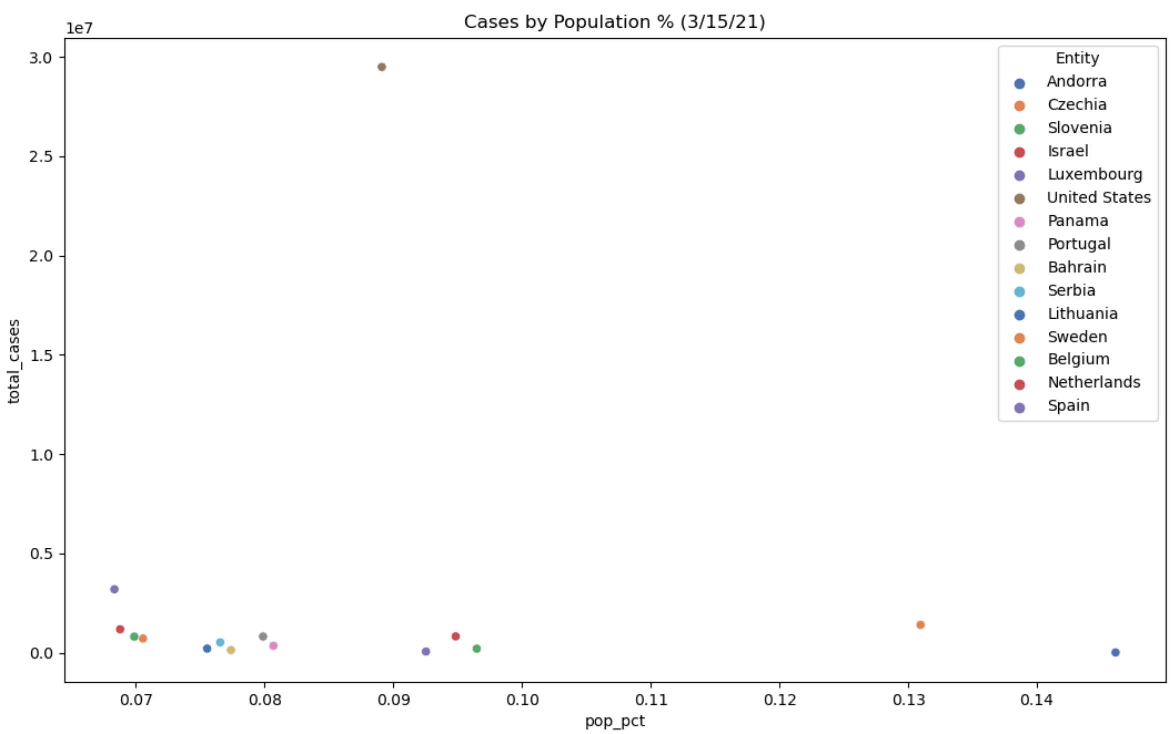


Here we can see that the relationship between variables shift. Trinidad and Tobago has a correlation of -.25 when compared to the percentage change in total cases, whereas

Norway has a correlation of -.09 when compared to the percentage change in total cases. Using this methodology, total correlation for face coverings and new cases dropped from .16 to -.11 (See Appendix A). Although this could be an important indicator in policy effectiveness, it is also possible that other expounding variables such as policy being enacted during spike's peaks, which could cause numbers to be overstated, or other increased stringent policies enacted during these periods had a greater effect on drops in cases. According to John Hopkins Medicine, public policy can take up to 30 days to be seen in statistics, whether that be lax or extreme policy, so immediate change in policy may not have the large effect that the data may show [7]. When trying to find dependence between variables by using Kendall rank correlation, Norway yielded -.02 correlation when comparing total cases to work closures, and Trinidad and Tobago .12 correlation when comparing total cases to facial coverings, showing a lack of dependence between the two variables (See Appendix B,C).

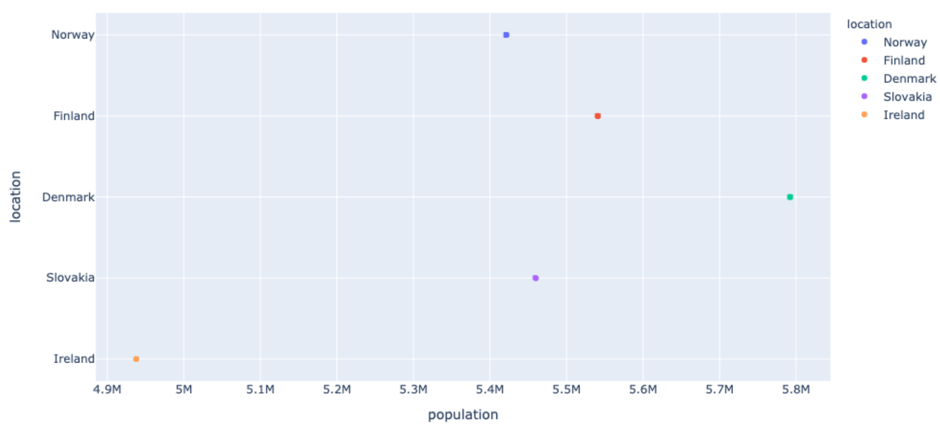
One thing that seemed fairly consistent was that nations that are fairly isolated (New Zealand, Taiwan, Vietnam, Fiji) fared well against the virus, however, countries that are landlocked or somewhat landlocked and larger in population (India, Belgium, United States) fared worse, although this wasn't always the case. The graphs below confirmed this. By finding the percentage between the total cases a country and its population it created a relative comparable indicator for case rates. It should be noted that only countries with HDI > .8 were included as those lower had less access to testing [8].

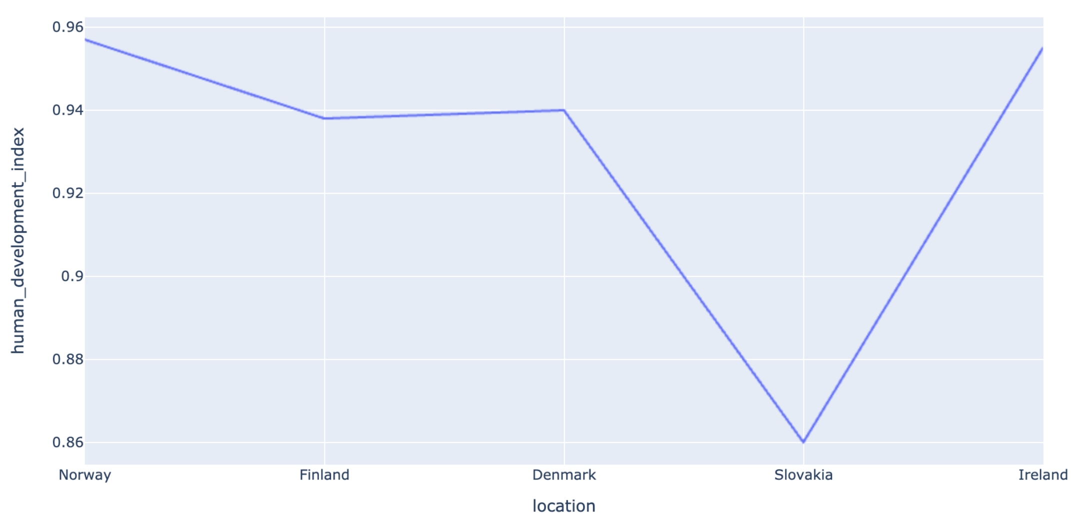

Cases by Population % (3/15/21)
Entity
Vietnam
Taiwan
China
Fiji
Cambodia
Yemen
Papua New Guinea
Thailand
Brunei
New Zealand
Syria
Nicaragua
Haiti
Bhutan
Australia
total_cases
pop_pct


Cases by Population % (3/15/21)
1e7
Entity
Andorra
Czechia
Slovenia
Israel
Luxembourg
United States
Panama
Portugal
Bahrain
Serbia
Lithuania
Sweden
Belgium
Netherlands
Spain
total_cases
pop_pct

EU Government Policy Analysis

Due to the size of our dataset, we decided to narrow it down to 5 test cases in a specified continent to get a more detailed look at the effectiveness of the 5 COVID government policies: contact tracing, testing, workplace closures, school closures, and facial coverings. To have an accurate representation, we chose the following countries from the EU based on their population density and HDI: Ireland, Finland, Denmark, Slovakia, and Norway. These countries have a population density of between 5 million to 6 million and are all relatively similar in terms of their human development index.

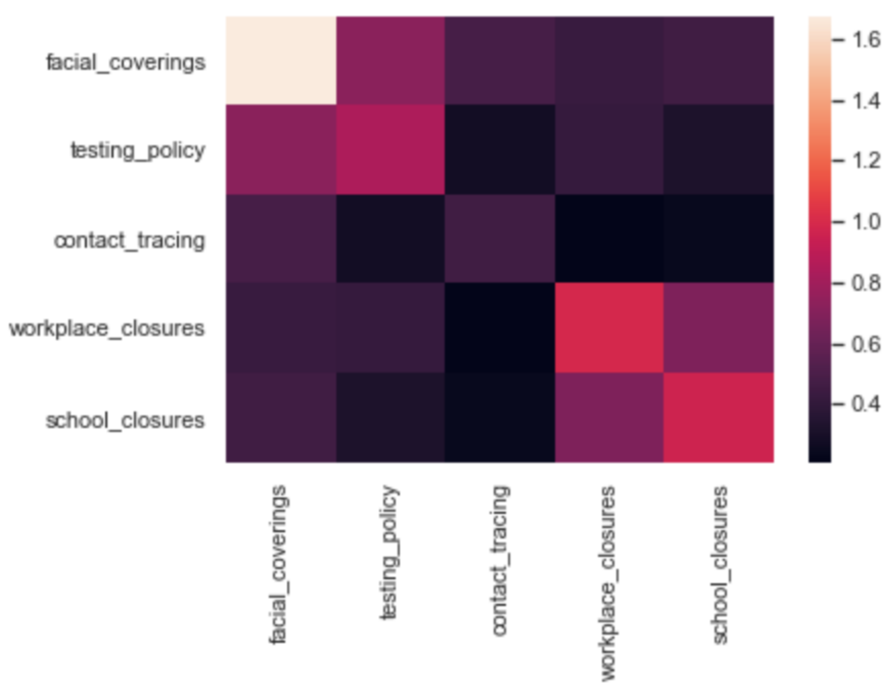




We constructed a covariance matrix heat map to determine whether there is a linear relationship between the 5 policies amongst each other. The diagonal elements contain the covariances of each pair of variables. The diagonal elements of the covariance matrix contain the variances of each variable. The variance measures how much the data are scattered about the mean. The strongest correlation within the set of variables for the 5 countries is the testing policies and facial coverings, with a positive linear relationship of 0.73.

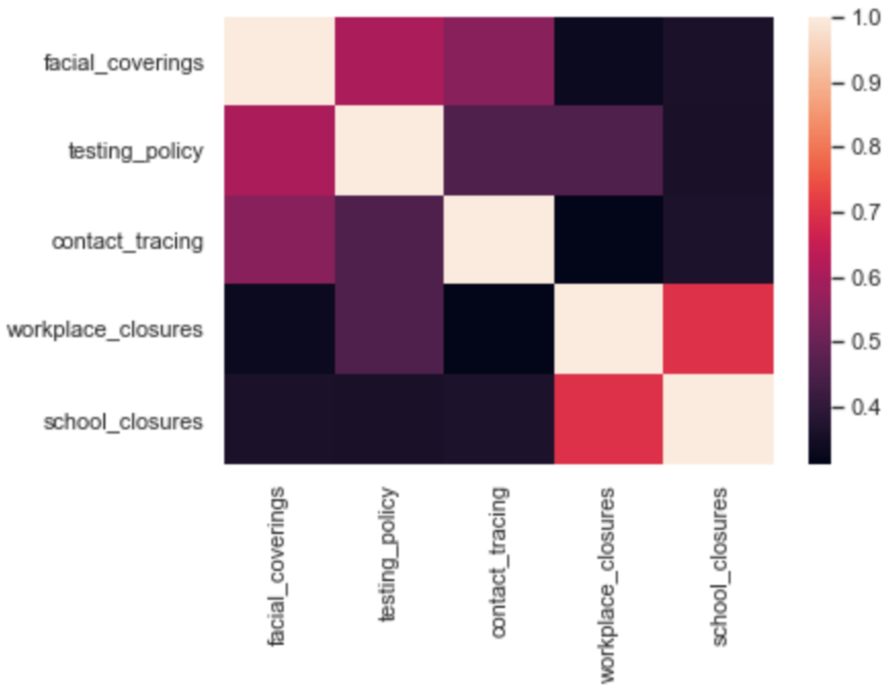
:

	facial_coverings	testing_policy	contact_tracing	workplace_closures	school_closures
facial_coverings	1.673049	0.720855	0.481022	0.427771	0.458559
testing_policy	0.720855	0.842452	0.283603	0.418694	0.322506
contact_tracing	0.481022	0.283603	0.455387	0.209455	0.241223
workplace_closures	0.427771	0.418694	0.209455	0.989827	0.682043
school_closures	0.458559	0.322506	0.241223	0.682043	0.962901

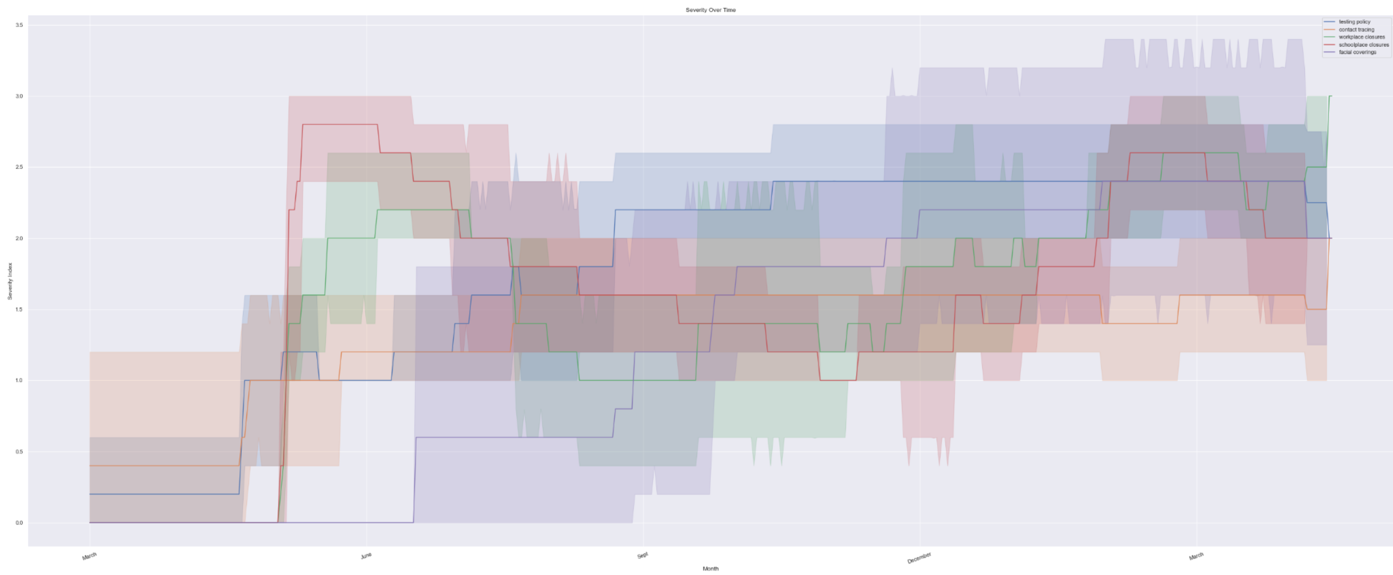


We also conducted a correlation test to extract the correlation coefficients between each policy. The difference between the covariance matrix and correlation matrix is that correlation measures both the strength and direction of 2 variables, while covariance only measures the direction of linear relationship. From the heat map and table below, we can see that school closures and workplace closures in the 5 countries have the strongest correlation with a score of 0.69. However, workplace closures and school closures are very similar so this can be ignored from the main objective of our analysis. The second highest correlation coefficient is between the testing policy and facial coverings, reiterating their strong relationship in usage against COVID-19. What is reassuring about these results is that all correlations are positive which was what we expected as all of these measures go hand in hand and together have an increased effect in combating COVID-19.

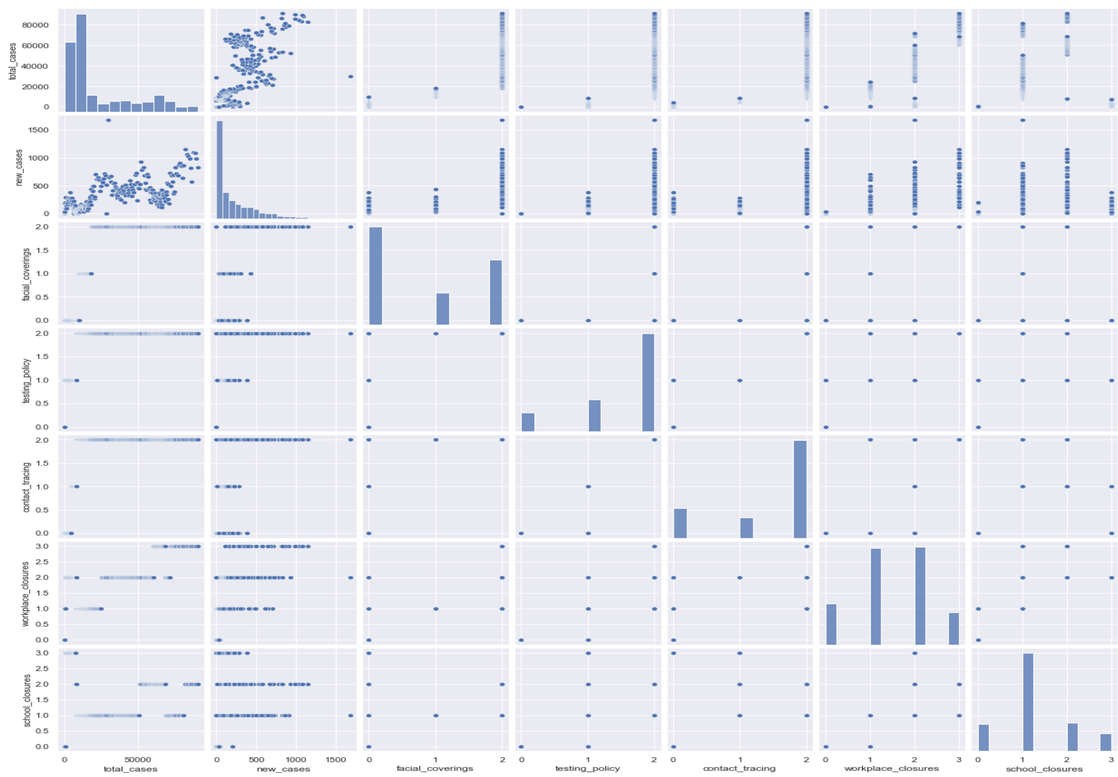
	facial_coverings	testing_policy	contact_tracing	workplace_closures	school_closures
facial_coverings	1.000000	0.607185	0.551087	0.332412	0.361285
testing_policy	0.607185	1.000000	0.457875	0.458505	0.358075
contact_tracing	0.551087	0.457875	1.000000	0.311975	0.364282
workplace_closures	0.332412	0.458505	0.311975	1.000000	0.698621
school_closures	0.361285	0.358075	0.364282	0.698621	1.000000



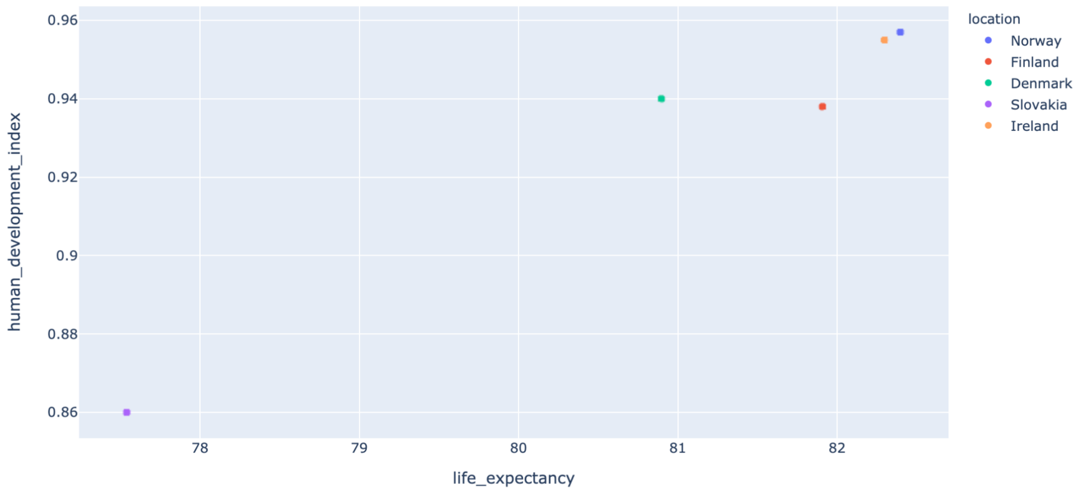
We plotted how much of a change each policy's severity index went during the COVID-19 Pandemic thus far. With this data, we can see how similar each country reacted to their respective cases and how their policies have changed over time. Below are line plots that show the average stringency index for each of the 5 COVID-19 policies enforced. Facial coverings had the highest variability in the stringency index, breaking over a level of 3 which shows severity and seriousness these countries give towards this policy. We can also see how countries utilized contact tracing from the very start but interestingly, facial coverings were not made mandatory or were not given a priority until much later. As one would expect, school closures had the highest slope which implies that the severity increased overnight and for the majority of the populations. Workplace closures saw a constant change overtime which means that the government was adapting and adjusting according to conditions and was the quickest with making changes to the severity of the policies. In the chart below, the color blue is testing policy, orange is contact tracing green is workplace closures, red is school closures, and purple is facial coverings (See Appendix D & E).

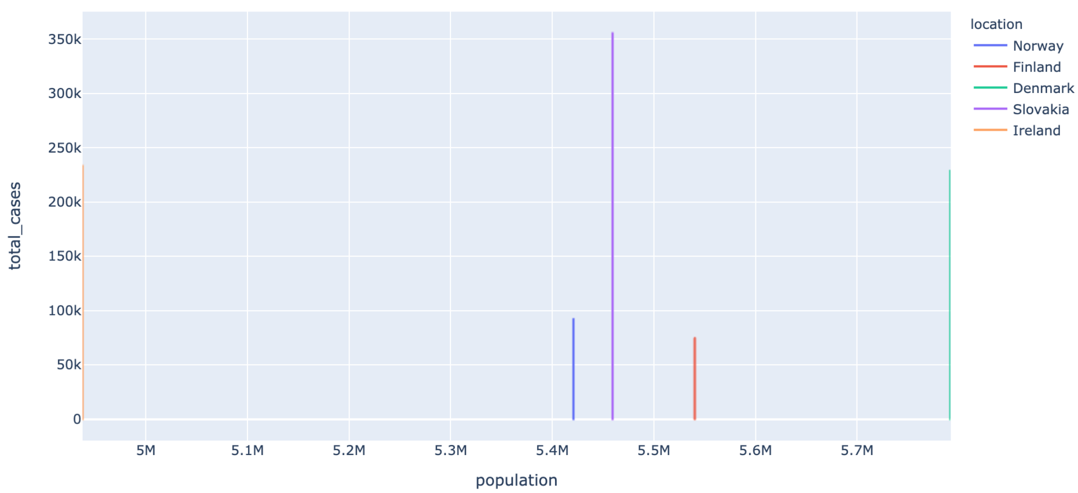


Next, we decided to create a pairplot for each country to get a clearer view of the relationships between each policy and the COVID-19 cases. Due to the nature of the data, all pairplots looked very similar and determined it was pointless to show 5 different pairplots. We chose to use Norway as our sample data for constructing the pairplot since it has the highest HDI out of the 5 countries. The data stretches to almost a year and as we know government policies and measures take time to show any real effect as it is a domino effect which is intended on having longer term implications and attain long run results. Due to this and the data being discrete, the pairplot cannot be used to infer much. But we can still see that when the school and workplace closure severity index was higher, the number of new cases per day was lower which is in line with our initial theory that these government interventions do have a controlling effect on covid cases. The analysis would have been more accurate if the data was more spread and of a longer period of time. In addition, finding how effective government interventions are when we look at a bunch of countries as a whole is harder as nations operate differently and work on their own pace (See Appendix F).



Finally, we wanted to see whether a country's HDI had any effect on a person's life expectancy. Below is a line chart that displays this relationship for the 5 countries. Slovakia has a HDI of 0.86, Finland a HDI of 0.94, Norway a HDI of 0.96, Ireland a HDI of 0.94 and Denmark a HDI of 0.94. The data shows that the more developed a country is, the higher life expectancy is for each person battling COVID-19. It is also interesting to note that although Finland and Norway have higher population densities than Slovakia, their number of total cases is much lower. This further reinforces that a country's HDI influences how well they handle the pandemic.





Conclusions

Overall it was difficult to have one direct answer to our hypothesis. After conducting extensive research and analysis, we found that there were too many expounding variables to find one consistent answer among test cases. Some countries fared well during COVID-19 because of geographic location (isolated countries fared well in controlling spread and enforcing policy), some countries more effectively implement the same policy than others, COVID spikes and policy effectiveness has a lag time [7], and other countries with higher wealth have more resources to effectively combat the virus [8].

Wealthier countries were more affected by the pandemic, and thus their governments had a tendency to enforce stricter measures on their populations compared to under-developed countries, which were less affected, leading to looser government actions. Nevertheless, poorer countries remain more vulnerable to the virus than their counterparts due to the precariousness of their public infrastructures and fragile economies that rely on the informal sector [8].

Based on our analysis of the 5 EU countries, we were able to see how all 5 government policies had a positive correlation with one another, showing their combined effectiveness against the pandemic. We determined how countries reacted through the severity index line chart, displaying the different behaviors of each policy in response to index changes over time. Finally, we found that a country's HDI is indicative of how well its government response is to the pandemic.

Author Contributions

Elias - Completed the entire COVID-19 General Policy Analysis, including all visualizations, analysis, understandings and conclusions from data. Also created a general data cleaning file for use by the team.

Roberto - Collected the datasets for the project, including those with the countries' shapes. Also, in charge of generating a geopandas visualization of the whole world in order to display the spread of COVID-19 through time.

Malick - Mainly focused on analyzing the data to find correlation between countries' wealth/development and stringency of government response. This work included data cleaning as well as visualizations making.

Siddharth - Focused on analyzing the 5 test case countries from the EU. Interpreted the covariance and correlation matrices, and visualisations about severity index of policies. Included making visualizations and interpreting results.

Spencer - Focused on creating and consolidating the data into a dataframe of 5 test case countries from the EU. Created and analyzed visualizations for the covariance and correlation matrices, along with the pairplot of Norway data. Created visualizations to display the relationship between a country's HDI and how well they are handling the COVID-19 pandemic. Assisted in cleaning overall dataset for more accurate representations.

References:

[1] - https://ourworldindata.org/policy-responses-covid

[2] - https://ourworldindata.org/covid-school-workplace-closures

[3] - https://ourworldindata.org/covid-face-coverings

[4] - https://ourworldindata.org/covid-testing-contact-tracing

[5] - https://ourworldindata.org/covid-international-domestic-travel

[6] - https://www.voanews.com/covid-19-pandemic/violators-trinidad-and-tobagos-new-no-mask-law-face-hefty-fines

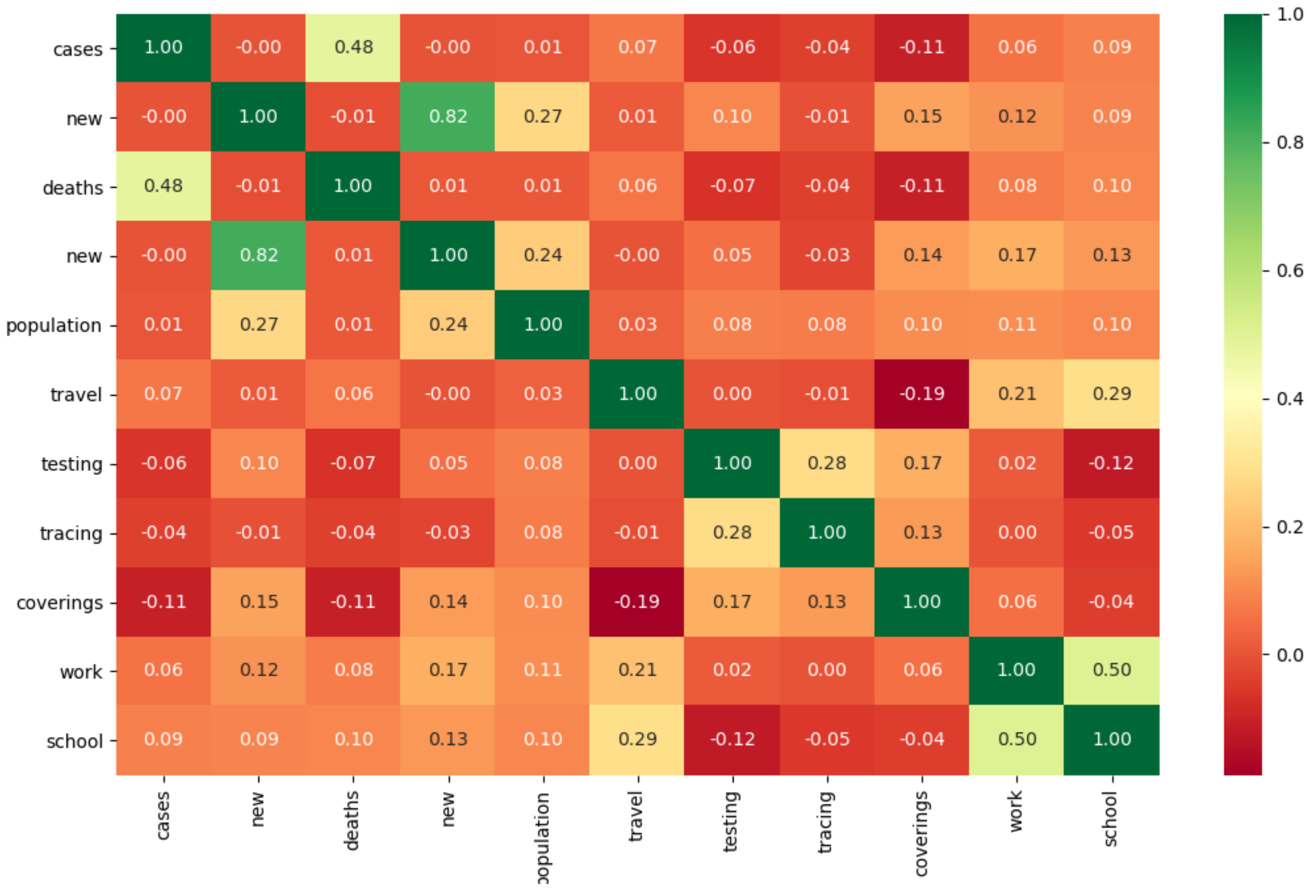
[7] - https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus/first-and-second-waves-of-coronavirus

[8] - https://www.un.org/ohrlls/news/world’s-most-vulnerable-countries-lack-capacity-respond-global-pandemic-credit-mfdelyas-alwazir

[9] - https://hub.arcgis.com/datasets/2b93b06dc0dc4e809d3c8db5cb96ba69_0?geometry=67.500%2C-89.382%2C-67.500%2C86.054

Appendix:

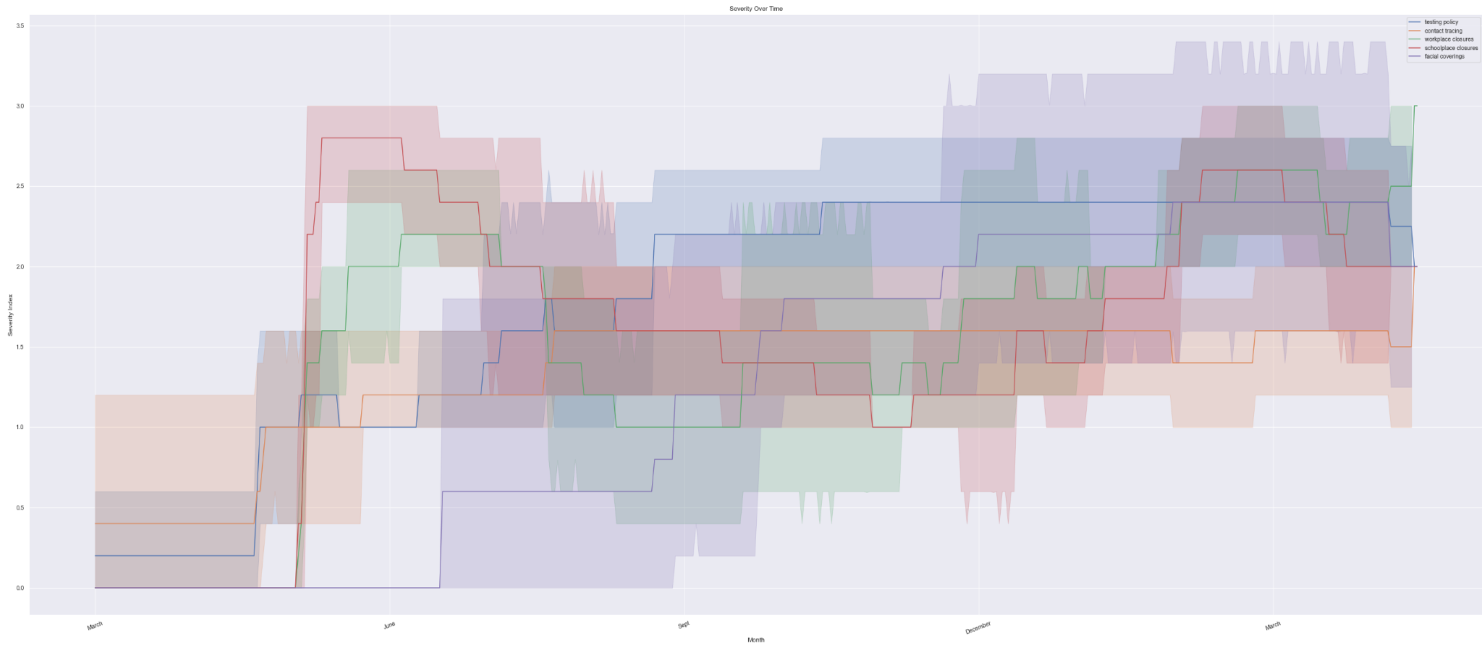
[A] - Correlation using pct_change.



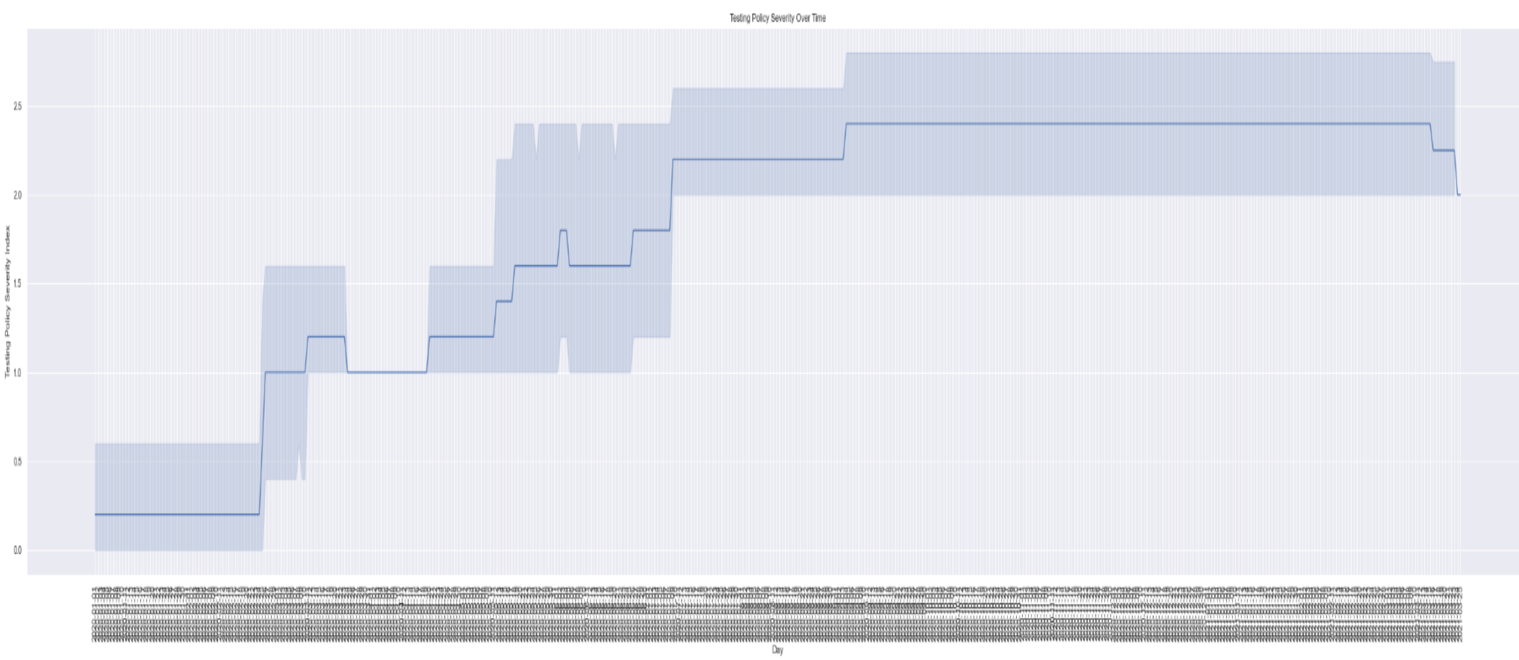
[B] - Norway Correlation using kendall method.

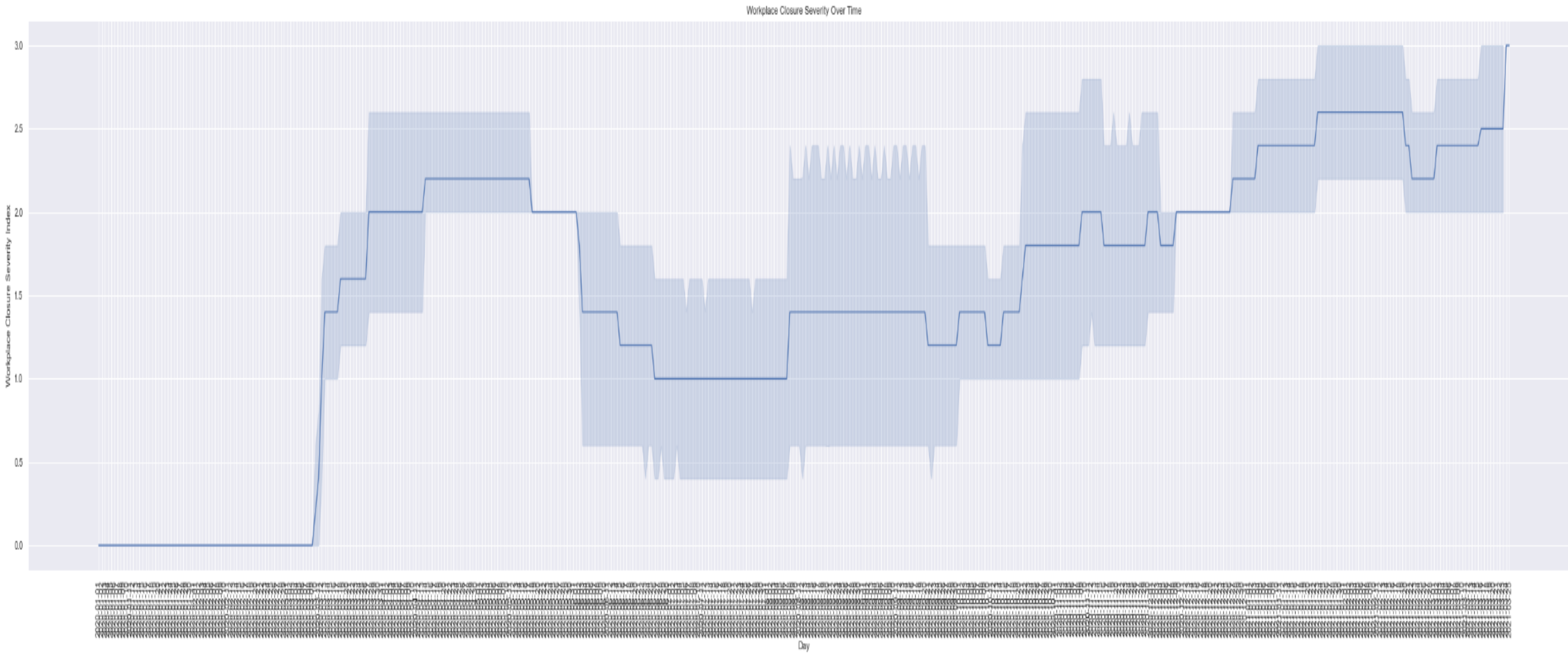
[C] - Trinidad and Tobago Correlation using kendall method.

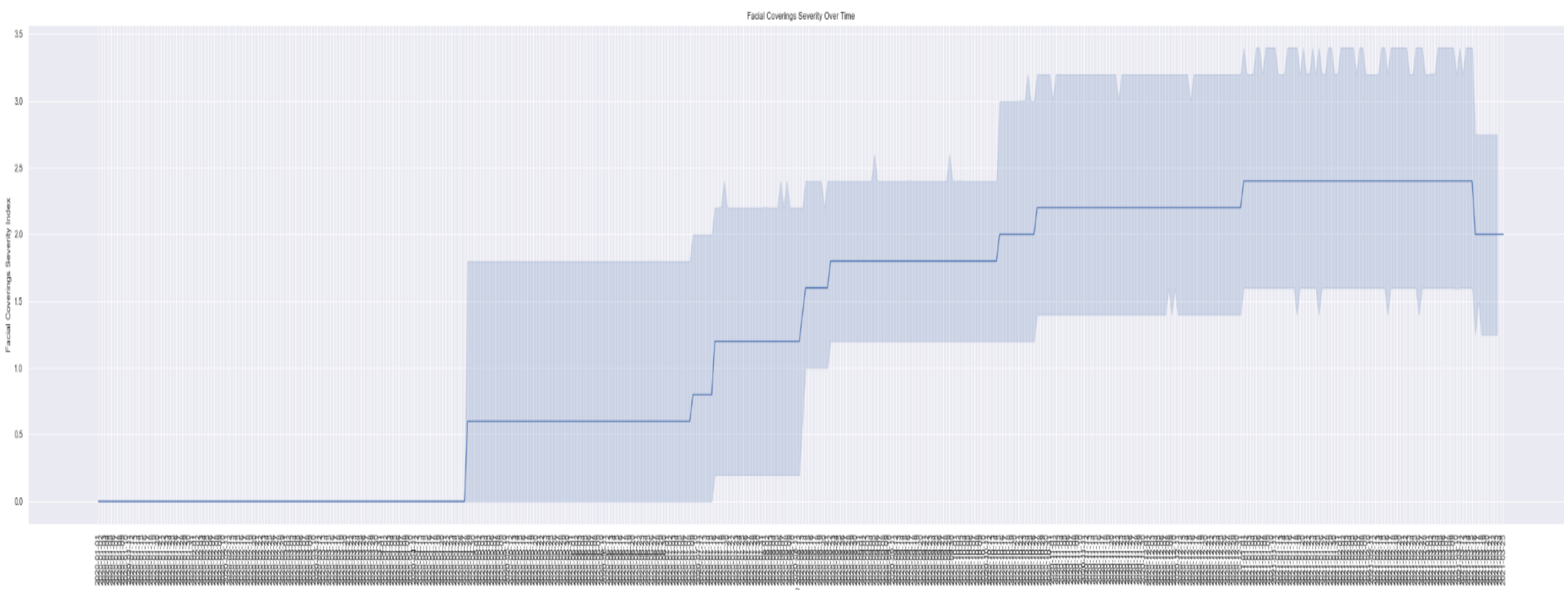
[D] - Severity Index Over Time

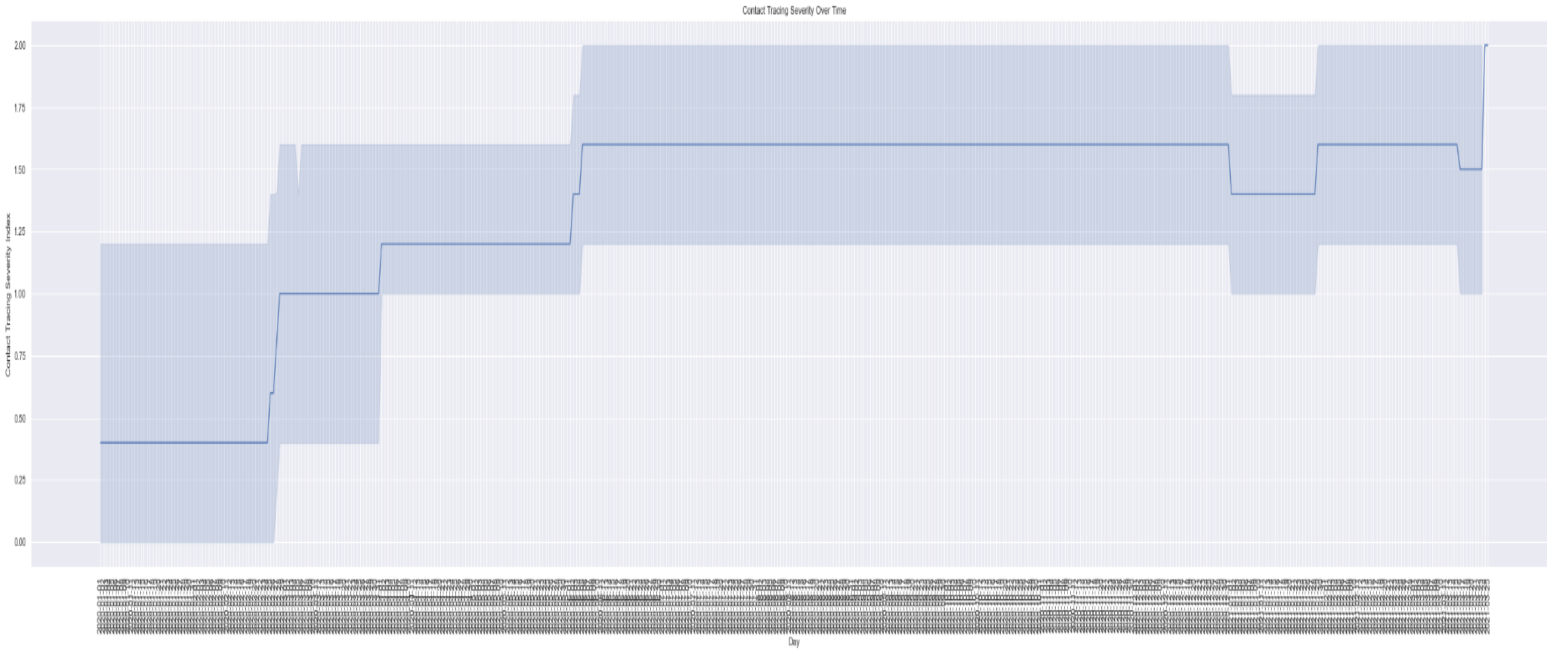


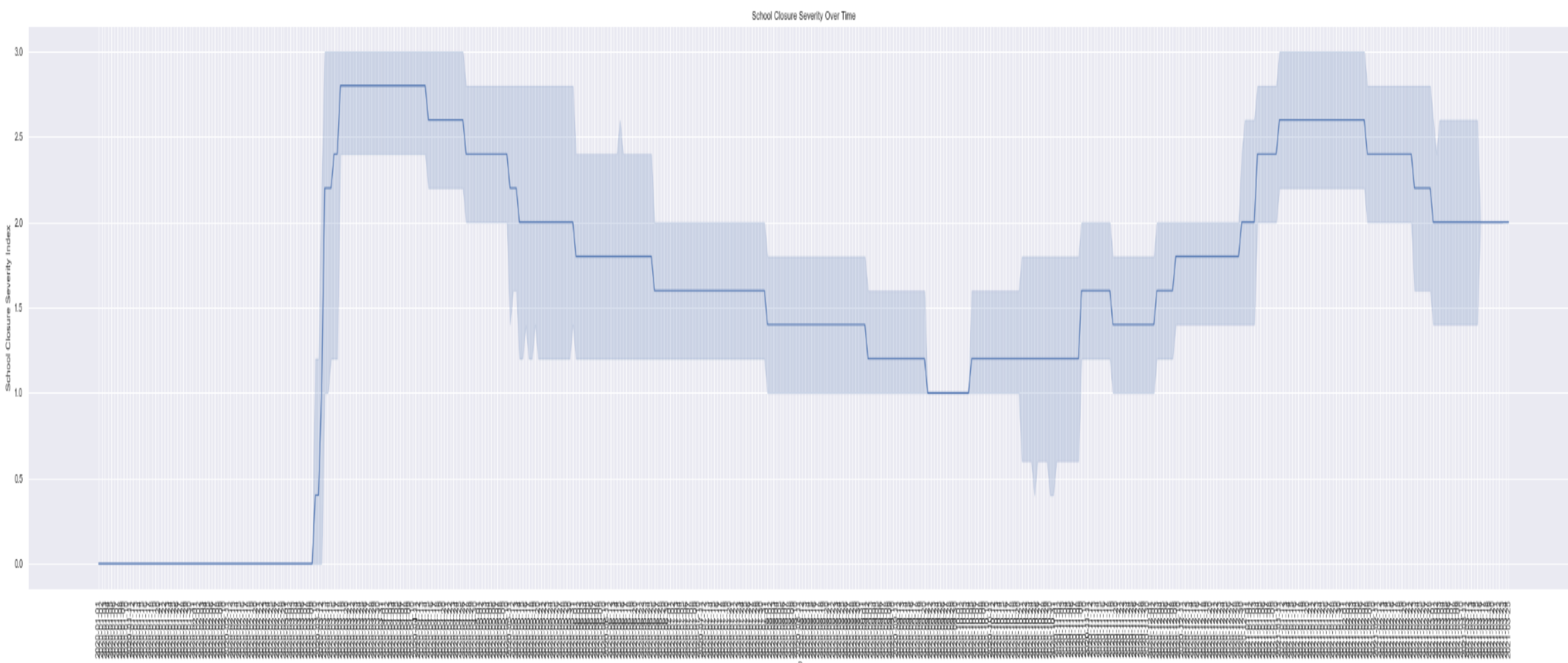
[E] - Severity Index Over Time For Each Policy




Workplace Closure Severity Over Time
Workplace Closure Severity Index
Day


Facial Coverings Severity Over Time
Facial Coverings Severity Index
Day


Contact Tracing Severity Over Time
Contact Tracing Severity Index
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Day


School Closure Severity Over Time
School Closure Severity Index
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Day

[F] - Norway Data Pairplot