The year 2020 has become synonymous with the word coronavirus, a brand-new illness, also known as COVID-19. First identified in Wuhan, China in late 2019, the virus spread quickly across the globe in 2020. It can be easily transmitted from human-to-human, but some experience little to no symptoms, similar to a common cold or allergies. On the other hand, due to the novelty of this virus, humans are lacking immunity, and it is leading to a high number of deaths. This has led to an inability to create an easy, standard mitigation tactic. The lack of symptoms in many cases, coupled with the lack of knowledge about this unique virus, has led to new developments and changes every day. One aspect that has been constant from the start is the collection of data. This data is now being heavily relied on to help overcome the spread of this deadly virus.

A handful of datasets with recorded COVID-19 observations broken down by day, by country, and by region, will be analyzed to determine what is already known about this virus and its spread, what can be predicted, and what prescriptions for improvements can be made. The first six datasets include data collected from January 22, 2020 through June 23, 2020 and the seventh is from December 31, 2019 through July 13, 2020.

This seventh dataset is from a non-profit organization, Our World in Data, with the goal of solving the world's problems through data. It includes daily numbers by country for total cases, new cases, total deaths, and new deaths similar to the other datasets. However, this one also includes factors unique to each country that could affect the spread of the virus or the number of deaths caused by the virus. Variables such as total tests, new tests, population density, percentage of population that is 65 or 70 years old and above, median age, diabetes prevalence, percentage who smoke, and number of handwashing facilities.

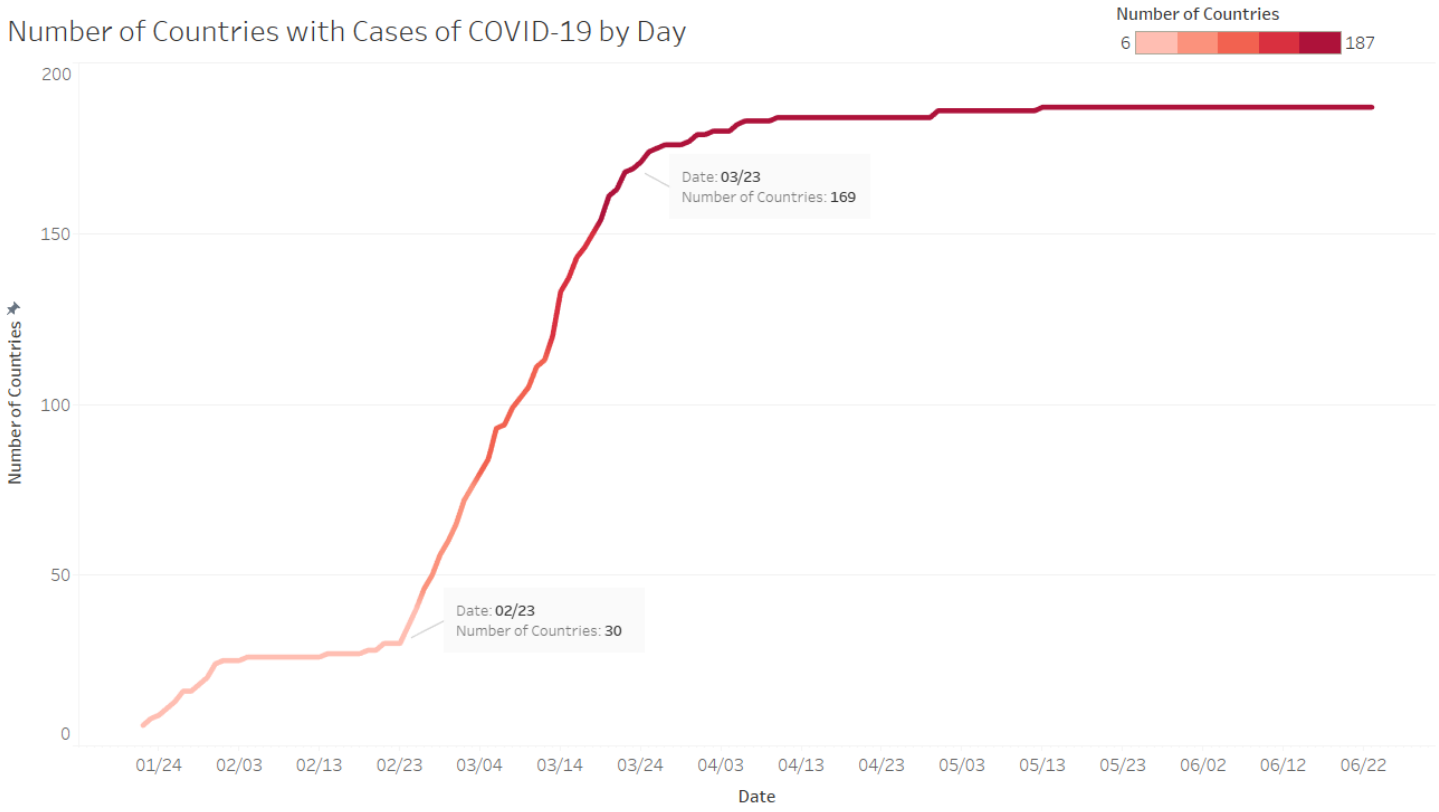
Analyzing COVID-19 data has a unique challenge. This pandemic is affecting the world; yet, no country is managing it the same, and every region is collecting and reporting different metrics. There is no standard method for data collection across the globe. Due to this, there is no simple solution to deal with the missingness in these datasets. Removing rows with missing observations would eliminate countries from the analysis but removing columns with a significant percentage of missing observations could remove potentially important predictor variables.

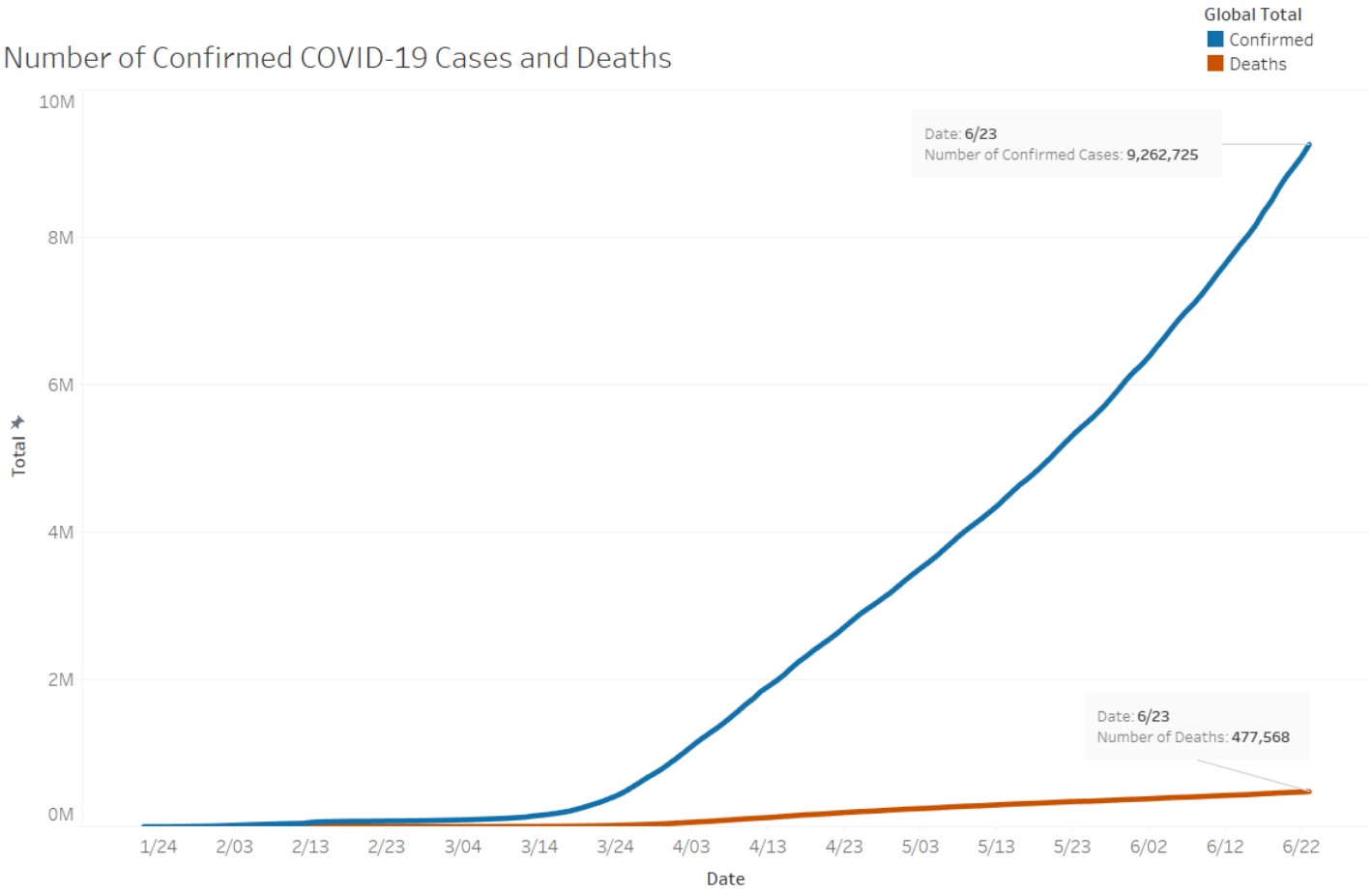
Imputation is also not an option in this specific case to manage the missingness in these datasets because it would skew the data. Imputation would completely ignore each country's immense differences in this situation. Replacing blank cells with zeroes for numeric variables, such as number of new tests, is also not a possibility. A blank cell does not necessarily equal no new tests,

it could mean the specific country does not collect or does not report that information. Therefore, with all these factors in mind, each blank cell will be filled with "NA" using Excel's IF formula.

The summary, structure, and vim plot R programming language commands help to easily determine the best way to clean the rest of the data. With the use of these, it becomes visible that there are 15 observations missing their designated World Health Organization region and the correct information will be filled in. This work in R Studio makes it clear that there are a couple of variables with zero observations recorded, such as "new tests smoothed," and therefore, these columns will be removed from the dataset.

After utilizing R and Excel to clean the data, Tableau will help with exploring the data to find insights from what is already known about this virus. COVID-19 began in the Hubei province of China and quickly spread. By February 23, 2020 thirty countries had cases of COVID-19. Within a month, by March 23, 2020, it had spread to 169 countries. In total, COVID-19 spread to 187 countries with over nine million confirmed cases and almost 500,000 deaths by the end of June 2020.





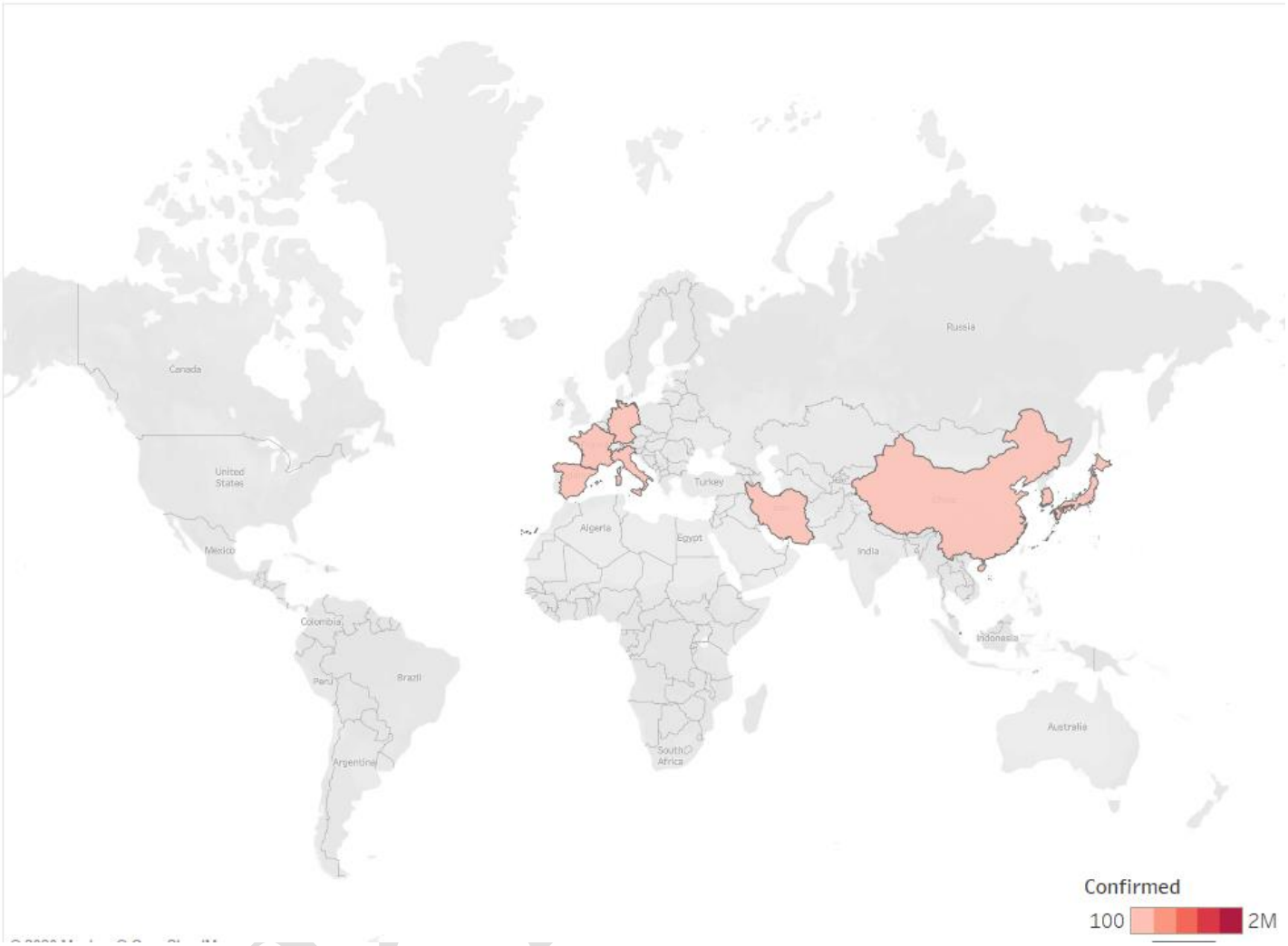
COVID-19 spread first to nearby South Korea and Japan. After, it jumped to Italy and then Iran. Eight days after Italy reached over 100 cases, it spread to France, Germany, and Spain. Two days later, COVID-19 reached the United States, over 7,000 miles away from its starting point. COVID-19 would spread across the majority of the globe and the United States would reach almost 1.5 million cases before China's Northern neighbor of Mongolia reports 100 cases. By the end of June, nearly the entire world had been affected by coronavirus, with the United States reaching over two million cases, but China's Southern neighbor, Bhutan, has less than 100 cases.

Lauren Ferguson

Number of Cases: February 26, 2020

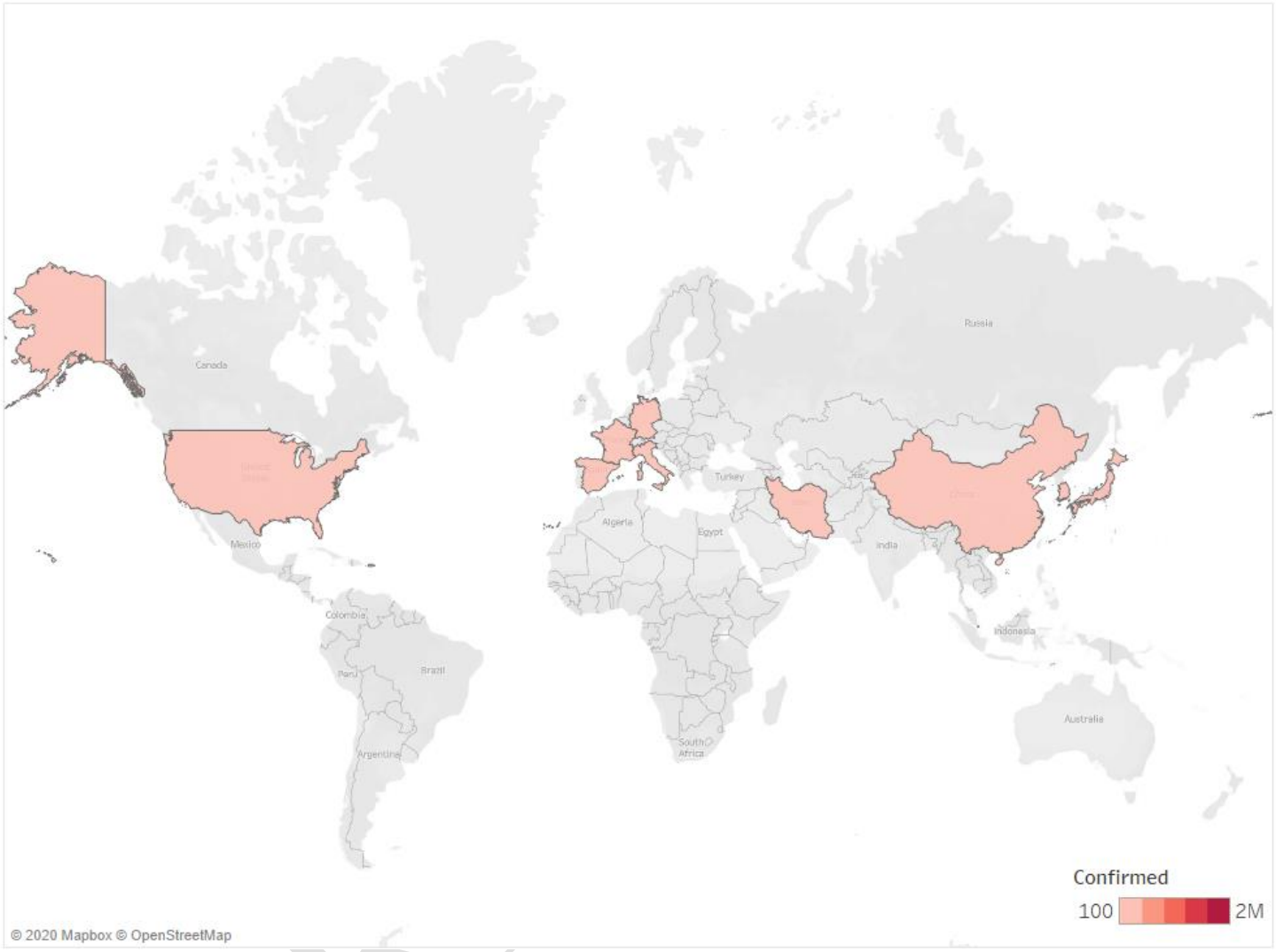
Lauren Ferguson

Number of Cases: March 2, 2020



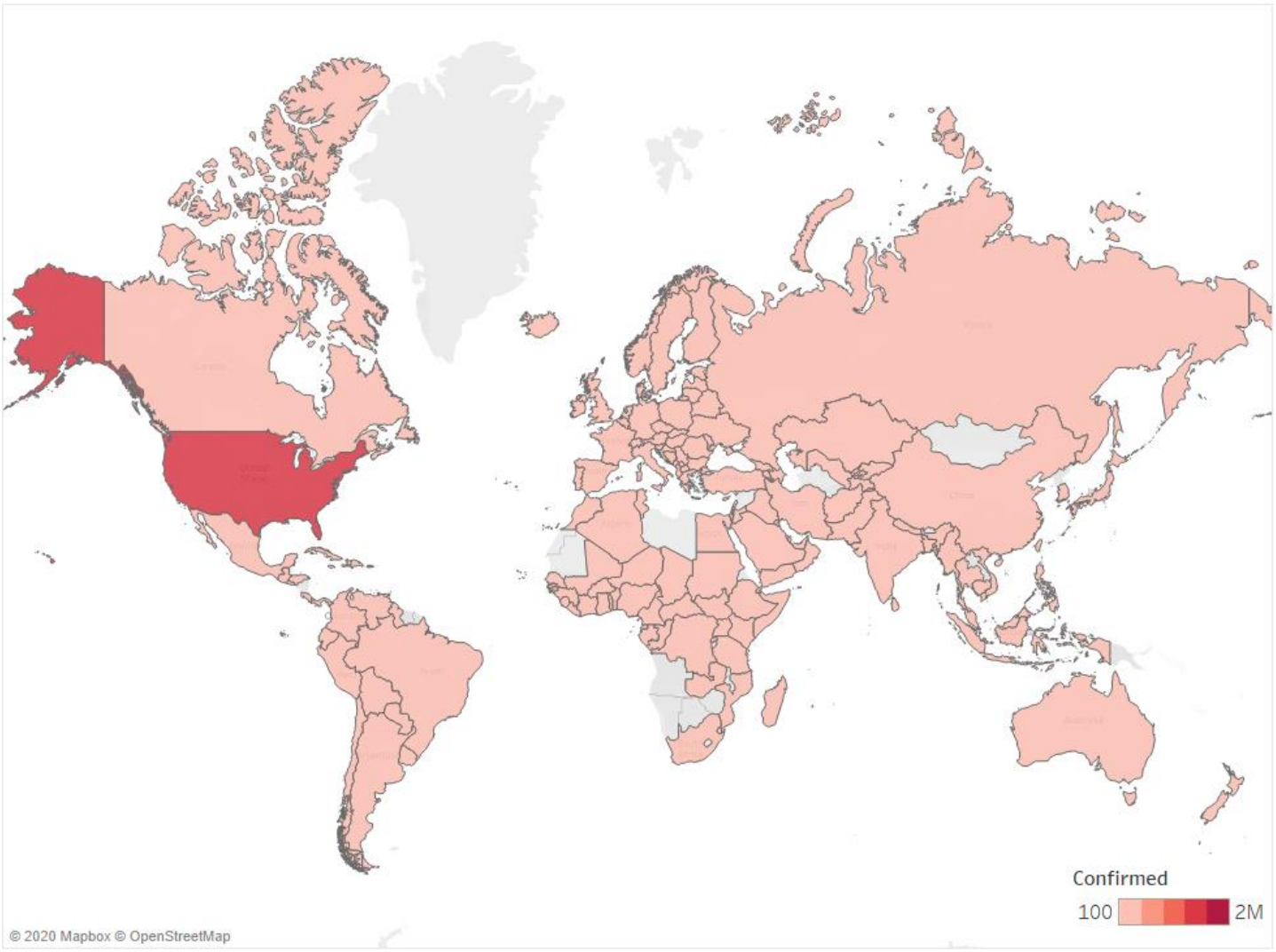
Lauren Ferguson

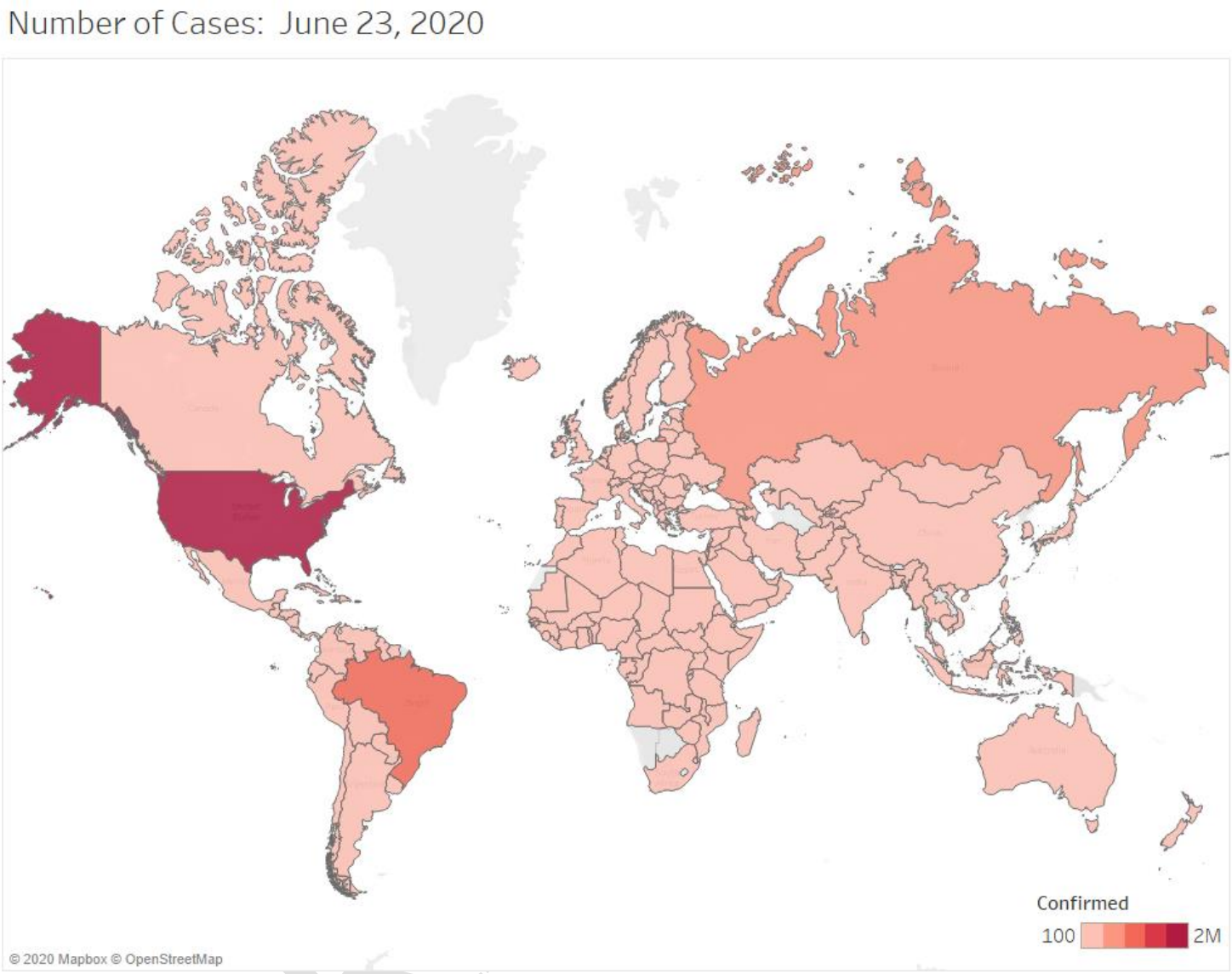
Number of Cases: March 4, 2020



Lauren Ferguson

Number of Cases: May 15, 2020

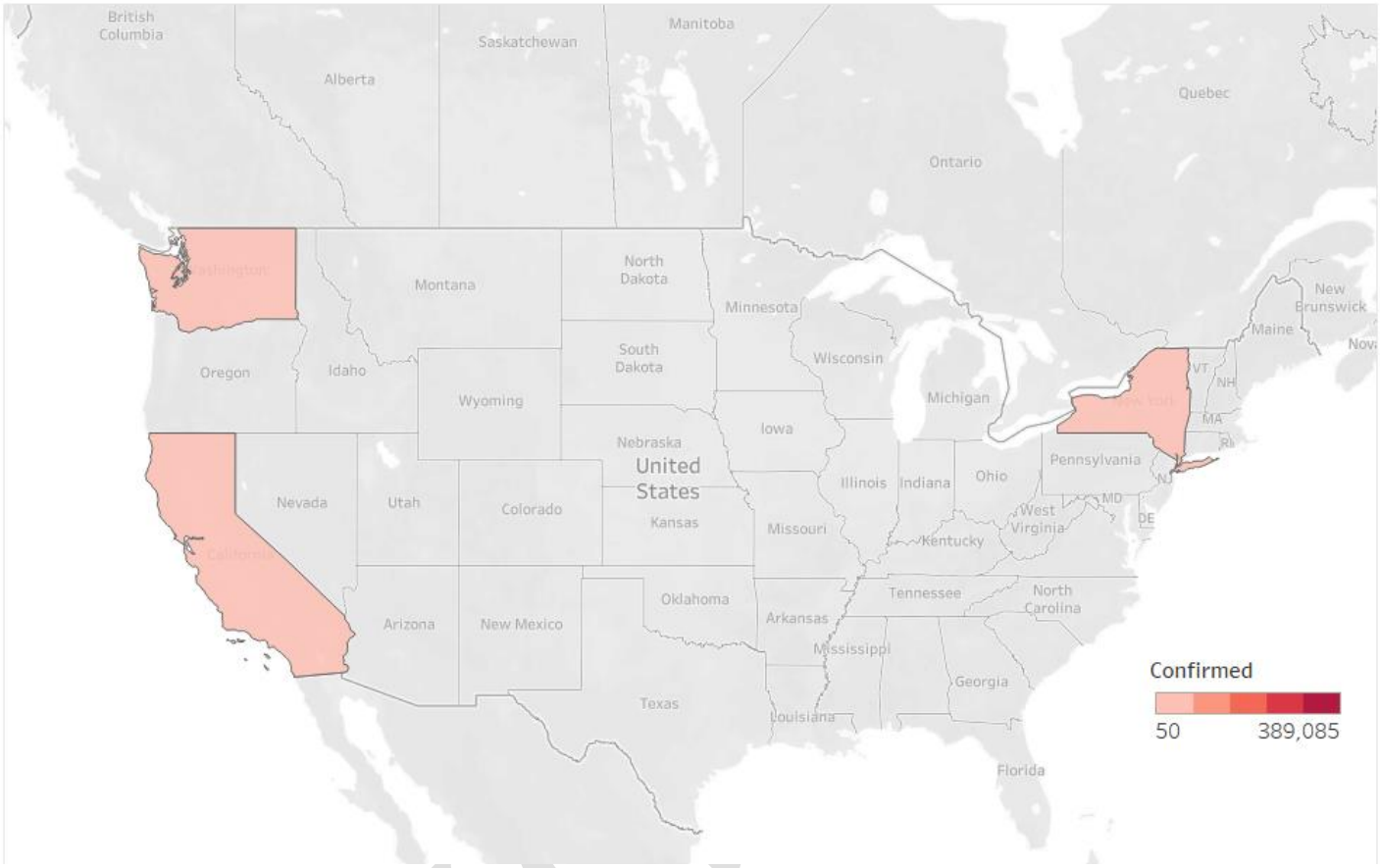


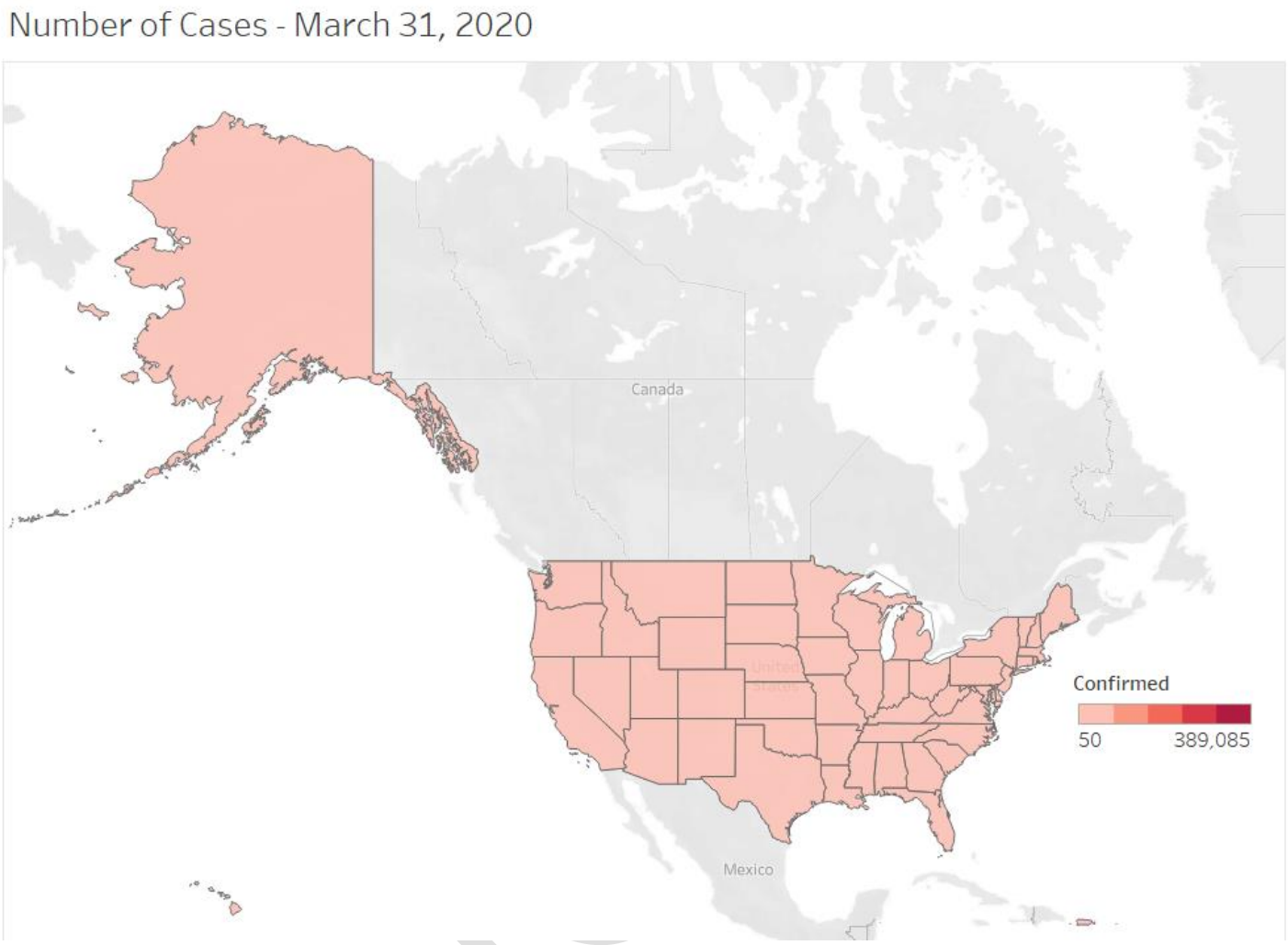


It took the United States no time for COVID-19 to jump from its starting point on the West Coast to the East Coast. On March 26th, the number of cases in the United States surpassed COVID-19's location of origin. By the end of March, all 50 states had at least 50 confirmed coronavirus cases. The United States has the highest case count with over two million, as of the end of June, and the highest death count, with over 100,000.

Lauren Ferguson

Number of Cases - March 7, 2020



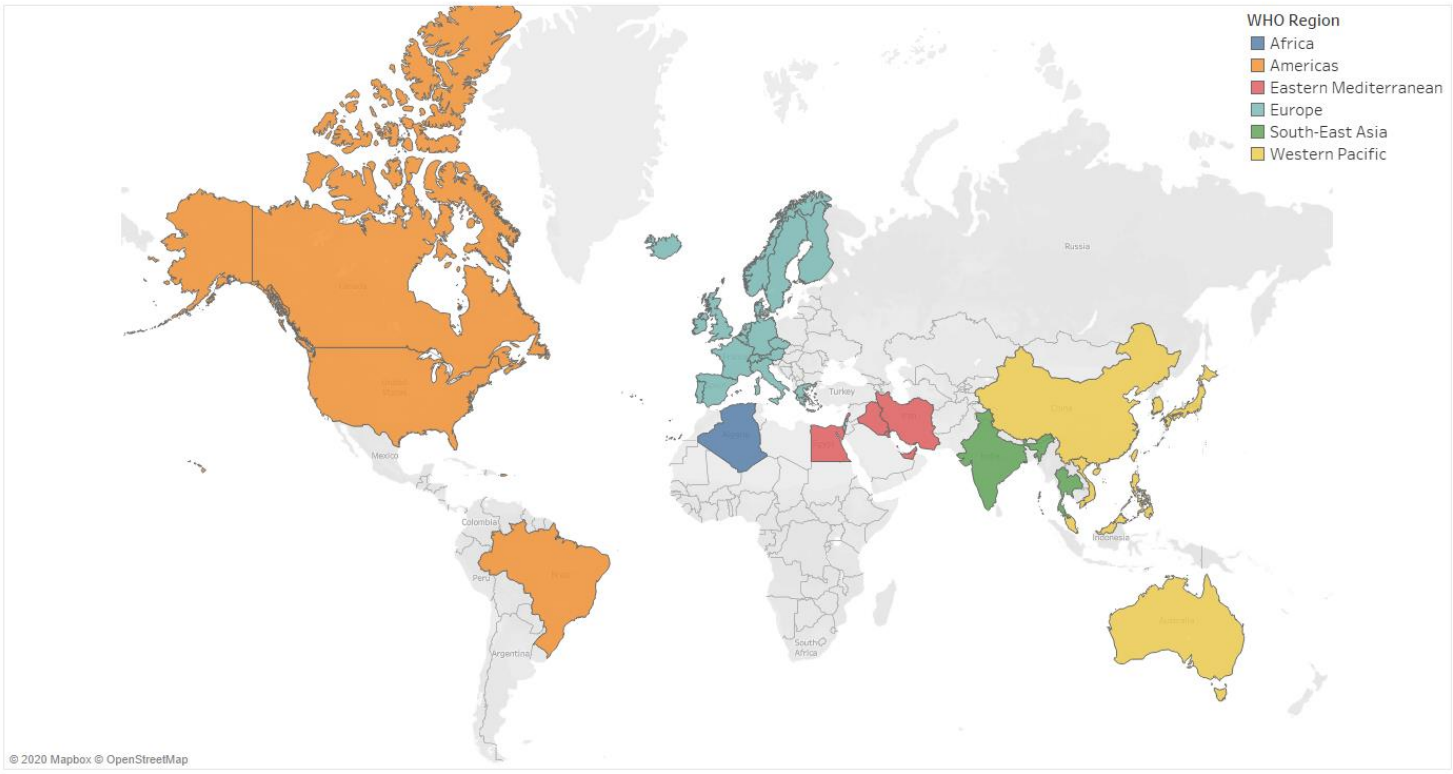


By March 9th, all five WHO regions had been affected by COVID-19. Africa was the last region to be impacted by coronavirus. Three days later, all five regions would experience death due to COVID-19. Globally, June 19th was the largest day of new cases with 181,495 and April 17th had the most reported number of deaths with 8,864.

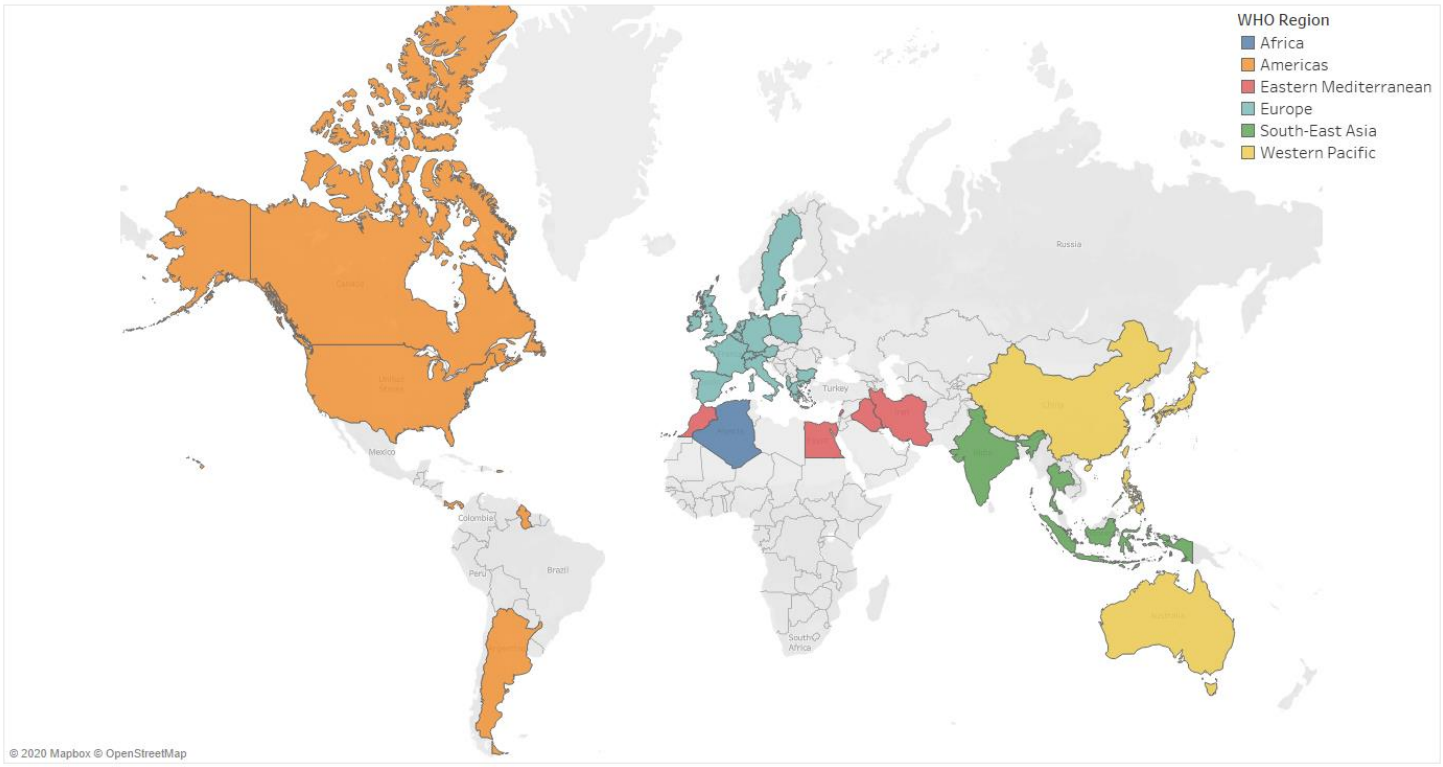
The Americas region has the highest number of cases with over four and a half million and the highest number of deaths with over 230,000. However, Europe has seen a larger amount of cases result in death. Seven percent of all of Europe's coronavirus cases ended in death, while the Americas had five percent of cases cause death.

Lauren Ferguson

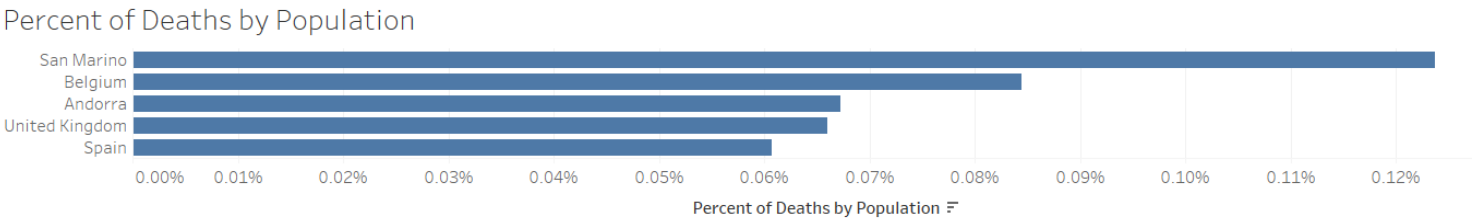
Number of Cases: March 9, 2020



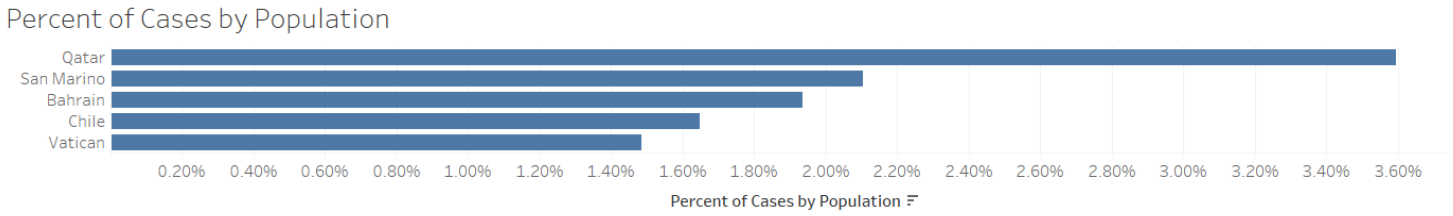
Number of Deaths: March 12, 2020



Lesotho was the last country to have a confirmed COVID-19 case. That day occurred on May 13th, although the country sits in the middle of South Africa, who saw their first case on March 5th. Similarly, San Marino is surrounded by Italy who had almost 250,000 cases by the end of June while San Marino reported 700. However, with San Marino's small population of just under 34,000 people, they have the highest percentage of deaths by population, with 0.12 percent of their citizens dying from COVID-19. San Marino also had 2.1 percent of its population with confirmed coronavirus cases, the second highest of any country.



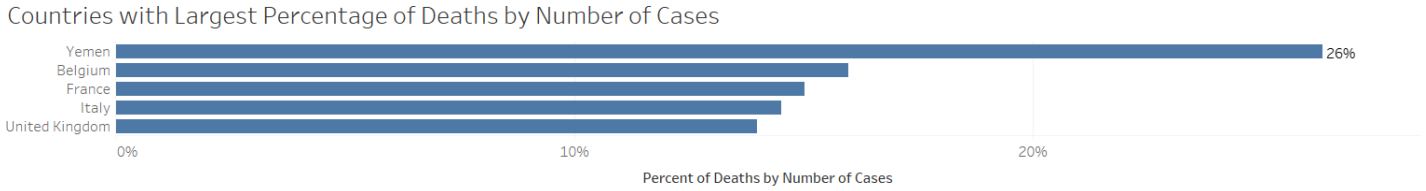
Qatar saw its first confirmed coronavirus case at the end of February, days before its much larger neighbor, Saudi Arabia saw a case, with which it shares a border. By the end of June, Saudi Arabia had almost 130,000 more cases than Qatar, but due to Saudi Arabia having almost 32 million more citizens, COVID-19 has affected a much larger percentage of Qatar's population. Qatar has had 3.6 percent of its population report positive coronavirus cases, the most of any country. Bahrain, 139 nautical miles across the Persian Gulf from Qatar, has had almost two percent of its population affected by COVID-19, the third largest percentage across the globe.



The first country to experience death from coronavirus, outside of COVID-19's country of origin, China, was the Philippines. Eleven countries reported deaths due to COVID-19 on March 1, 2020 and by the end of the month, the number rose to 125. By this time, the United States had also surpassed China's death count. As of the end of June, 167 countries would suffer deaths due to the COVID-19 pandemic.

The country of Yemen has been hit the hardest by COVID-19's death toll. While their total deaths, 261, is much lower than most countries, 26 percent of all Yemen's coronavirus cases have

ended in death. It is the only country out of the top five that is not in Europe and the country in second place is a distant second with 16 percent of its cases adding to the death count.



For the world as a whole, 0.17 percent of its population reported positive coronavirus cases as of July. There are 75 countries who have higher percentages of their residents testing positive. Thirty countries have larger than 0.01 percent, the world's percentage, of their populace who have died from COVID-19. As of July 13th, the world has almost 13 million reported COVID-19 cases, almost 600,000 of those ended in death, meaning 4.41 percent of the global coronavirus cases resulted in a death. Fifty-one countries have a larger percentage of their case count end with a fatality. There are 20 countries that are above the world's percentages for all three: percentage of cases by population, percentage of deaths by population, and percentage of deaths by number of cases.

Eighty-five percent of these 20 countries have an above average median age for the world. Seventy-seven percent have greater than average populations of 65 and 70-year old residents and above. All are way below the average population of extreme poverty for the world and the majority are well above the gross domestic product (GDP) per capita. A country's GDP reveals its economic standing and per capita is dividing its economic worth by the size of its population. This suggests that COVID-19 spread has more to do with the age of the country's population than economic standing, but what about the health of its population?

Of these 20 countries who are above the world's averages, 80 percent are below average for diabetes prevalence. Seventy-seven percent are well above average for percent of female smokers, but only fifteen percent are barely above average for male smokers. Half of the countries are below average in number of hospital beds per 1,000 people. However, all of them have above average life expectancies. This strengthens the case that a larger spread and increased death count from COVID-19 are due to larger populations of older generations, which goes hand-in-hand with a higher life expectancy, instead of at-risk health factors, like diabetes.

Further evidence of these factors mattering more than others appears when analyzing the opposite case, countries with zero deaths. The averages for these countries are exactly the opposite of the countries with the highest percentages of deaths by population and by cases. Their average population density is higher than the world's average and the percentage of their populations that are 65 or 70 years old and above is lower than the world's average. Fifty-two percent have below average life expectancies and the other 48 percent all have life expectancies hovering around the world's average. There is above average diabetes prevalence, below average female smoker population, and slightly above average male smokers.

Based on the observations above, if countries with higher populations of older generations and higher life expectancies are seeing higher numbers of deaths due to COVID-19 then what about the countries with the highest percentages of these populations and the highest life expectancy? Japan has the largest percentage of both the 65 and older population and 70 and older population in the world. Due to Japan's enormous population, of almost 130 million, they have managed to be in the lower half of cases and deaths by population. They rank 45th in number of cases by population with 0.017 percent of their population reporting positive cases and 86th in number of deaths by population with 0.0008 percent of its population dying from coronavirus. However, Japan falls into the last quarter of countries for highest percentages of COVID-19 cases that end in death. Their percentage is higher than the world's with 4.6 percent of their positive cases resulting in death.

The observations are even more drastic for the country with the highest life expectancy. Monaco's life expectancy is 86.75, the oldest in the world, and Monaco falls above average for percent of cases and deaths by population as well as the number of cases that end in death. Monaco ranks 156th with 0.28 percent of its population reporting positive COVID-19 cases, compared to the world's 0.17 percent. Monaco ranks 177th with 0.013 percent of its population dying from coronavirus and it ranks 162nd with 4.59 percent of its COVID-19 cases ending in death. All three of these fall into the last quarter of countries when ranked lowest to highest by percentage. Both Japan and Monaco having high percentages of their COVID-19 cases resulting in death adds to the evidence that higher death rates due to COVID-19 may be predicated on the percentage of older populace.

To see if machine-learning algorithms agree with these factors being the most important, the data will be sorted by date and imported into JMP Pro 15. After, the data will be prepared

through cross-validation to guard against poor forecasts from predictive modeling. This will split the data into three parts. The first will be the training data composed of 60 percent of the observations. This portion will be used to estimate the models. The second section will be the validation data consisting of 20 percent of the data. This segment will be used to assess the model's estimations and be an indicator to stop modeling. The last 20 percent of the data will not be used to create the model, but instead, will supply new, never-before-seen observations to perform an unbiased analysis of the predictive ability of the model.

Once the data has been split, a number of statistical models will be run to predict the response variable, new deaths. To compare all the models, the r squared value in the test portion of the data will be examined. The r squared value defines what percent of variability in the dependent variable is explained by the model; therefore, the model chosen will have the highest r squared value in the last 20 percent, the test portion of the data. This is a mark of the best model because it demonstrates how the model performs when introduced to brand new data.

The first model will be a standard Ordinary Least Squares (OLS) model. With big data, OLS models tend to model random noise and produce poor forecasts, but it is a good benchmark model to use for the other, more complex models. The second model will be an elastic net penalized regression. Elastic net is a best of both worlds regression model because it combines methods from both ridge and LASSO models. This model will penalize variables that have little or no impact and only allow the most informative variables to enter the model. If two or more variables are highly correlated, elastic net will keep both in the model.

The next step will be more complex models that can handle highly advanced, nonlinear relationships. First, a decision tree, which will be the benchmark model for the second, a random forest model that is an ensemble decision tree. This model will construct multiple decision trees using only part of the data and a selection of the variables. Each tree will be uncorrelated with the other ones and in the end, the random forest will be an average of all the predictions.

Subsequently, a set of neural networks will be run on the data. These models transform input variables by applying an activation function. They also use penalty methods to prevent modeling random noise. After completing a neural network with two layers of three nodes each using the absolute penalty method, a boosted neural network will be run. Boosting starts with a relatively simple model and builds, learning from the mistakes it made in the previous models. Each time it runs, the model fit is improved by correcting the areas of bad fit in the previous model,

then in a similar fashion as a random forest model, all the models are averaged together to produce a prediction for each observation. The averaging done with these ensemble models helps ensure the predictions will not be unstable. The final model will be a boosted tree, which does the same construction as a boosted neural network and the same tree building as a random forest.

The only other details of these models to note are the random seed and informative missing options. Each model will be created using random seed 123 to make it possible to replicate the results. Also, each model will run with and without the informative missing option to see which one performs better. Informative missing is a feature of JMP that tells the model to handle observations with missing variables in a different class as ones with no missing information. In some instances, not reporting certain pieces of data could be an indicator that will change the model if observations of full data and observations with missing data are seen as two different groups.

Measures of Fit for new_deaths

Holdback	Predictor	Creator	.2 .4 .6 .8	RSquare	RASE	AAE	Freq
0	Pred Formula new_deaths OLS woIM	Fit Least Squares		0.5016	32.158	12.015	1203
0	new_deaths Prediction Formula ENet	Fit Generalized Standard Least Squares		0.5016	32.158	12.015	1203
0	new_deaths Predictor DT	Partition		0.6848	212.09	32.057	15257
0	new_deaths Predictor RF wIM	Bootstrap Forest		0.7212	199.47	21.995	15257
0	new_deaths Predictor 2 RF woIM	Bootstrap Forest		0.7039	205.59	25.503	15257
0	new_deaths Predictor BT wIM 2	Boosted Tree		0.6693	217.25	40.219	15257
0	Predicted new_deaths BoostedNN 3Nodes 40Models Squared 20Tours wIM	Neural		0.7256	197.89	26.137	15257
0	Predicted new_deaths NN 2Layers 3Nodes Absolute 20Tours wo IM	Neural		0.6673	26.274	10.538	1203
1	Pred Formula new_deaths OLS woIM	Fit Least Squares		0.1829	138.81	43.894	495
1	new_deaths Prediction Formula ENet	Fit Generalized Standard Least Squares		0.1829	138.81	43.894	495
1	new_deaths Predictor DT	Partition		0.8215	137.05	29.667	5150
1	new_deaths Predictor RF wIM	Bootstrap Forest		0.8757	114.36	22.959	5150
1	new_deaths Predictor 2 RF woIM	Bootstrap Forest		0.8448	127.82	24.926	5150
1	new_deaths Predictor BT wIM 2	Boosted Tree		0.7981	145.76	38.399	5150
1	Predicted new_deaths BoostedNN 3Nodes 40Models Squared 20Tours wIM	Neural		0.8784	113.11	29.588	5150
1	Predicted new_deaths NN 2Layers 3Nodes Absolute 20Tours wo IM	Neural		0.3086	127.69	45.814	495
2	Pred Formula new_deaths OLS woIM	Fit Least Squares		-2.255	304.79	163.93	395
2	new_deaths Prediction Formula ENet	Fit Generalized Standard Least Squares		0.2290	148.34	64.152	395
2	new_deaths Predictor DT	Partition		0.7834	165.81	44.449	5142
2	new_deaths Predictor RF wIM	Bootstrap Forest		0.8313	146.30	43.281	5142
2	new_deaths Predictor 2 RF woIM	Bootstrap Forest		0.8191	151.51	33.177	5142
2	new_deaths Predictor BT wIM 2	Boosted Tree		0.7498	178.18	53.010	5142
2	Predicted new_deaths BoostedNN 3Nodes 40Models Squared 20Tours wIM	Neural		-0.890	489.77	176.09	5142
2	Predicted new_deaths NN 2Layers 3Nodes Absolute 20Tours wo IM	Neural		-0.226	187.06	111.15	395

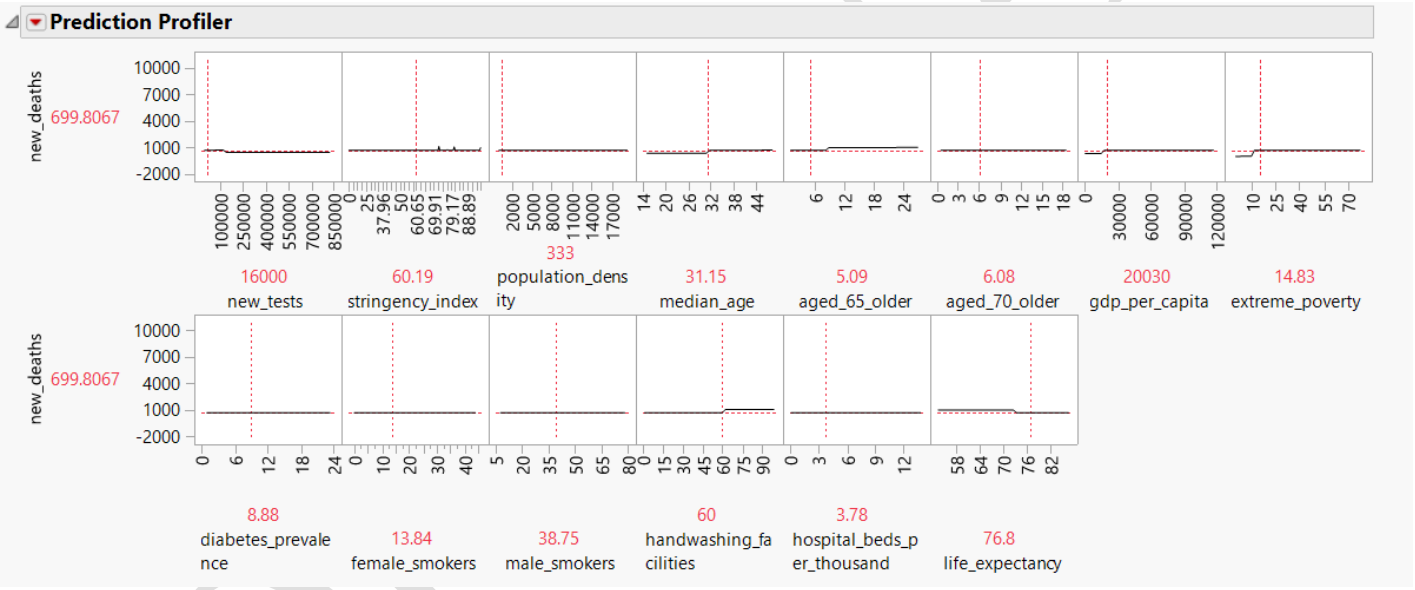
After comparing all the models against each other, the random forest model with the informative missing option selected has the highest r squared value at 0.83 and the lowest root average squared error at 146. The lowest average absolute error, which is a useful metric when the data has unusual, extreme observations is the random forest model without informative missing. Both root average squared error and average absolute error designate how much error is in the model; therefore, the lower the value, the better. The random forest with informative missing explains 83 percent of variability in the new deaths variable and is the best model.

Column Contributions

Term	Number of Splits	SS	Portion
aged_65_older	222	191878866	0.1966
extreme_poverty	196	135222566	0.1385
stringency_index	676	115255162	0.1181
aged_70_older	184	107656701	0.1103
median_age	255	105191833	0.1078
gdp_per_capita	216	96629303.8	0.0990
male_smokers	165	54896977.2	0.0562
handwashing_facilities	144	50480266.4	0.0517
new_tests	387	44296027.5	0.0454
diabetes_prevalence	229	34016336.8	0.0348
life_expectancy	203	22099652.9	0.0226
female_smokers	164	10343097	0.0106
hospital_beds_per_thousand	188	4379288.52	0.0045
population_density	237	3812485.91	0.0039

According to the random forest model, the percentage of a country's population who are 65 or older is the most important variable in predicting new deaths due to COVID-19. This percentage explains 20 percent of the fluctuations in the number of new deaths. The second most important variable is the country's extreme poverty level, it explains 14 percent. After that, the stringency index, or strict rules countries put in place to try and halt the rising number of cases, explains 12 percent. The percentage of population who are 70 or older and the median age of the population both explain 11 percent of vacillations in new death numbers. The gross domestic product by population explains ten percent.

After these top six variables, the amount explained by the remaining variables drops to single digits. Male smokers is six percent and female smokers is one percent. The number of handwashing facilities and new tests are five percent. Diabetes prevalance is three percent, life expectancy two, and hospital beds per thousand is 0.5. Surprisingly, due to how easily this virus is spread from person-to-person, population density is the least important variable, explaining 0.4 percent.



If all these independent variables are slightly higher than the average for all countries, then that location could expect 1,000 new deaths due to COVID-19. Even with a stringency index as high as 60.19, if the location has a higher percentage of older residents, it is going to experience more people dying from coronavirus. If all variables stayed the exact same, but the 65 and older population was half the size, 5.09 percent, instead of 10.18, that location would experience 301 less deaths than the one with the higher percentage of the 65 and older population.

A detrimental side effect of this novel virus is the amount of lives it is cutting short. Large numbers of individuals who are contracting the virus are having harsh reactions causing extended

intensive care hospital stays resulting in death. Experts have not been able to fully determine what is causing the difference in one person's reaction to the virus compared to another individual.

According to the chosen random forest model, the most important variable in predicting new deaths from COVID-19 is the percentage of population that is 65 or older. This is understandable for a number of reasons. As humans get older, their bodies produce fewer immune cells and the little they have do not communicate as well with the others. This in turn creates a situation where the body of an older individual takes longer to react to harmful germs entering the body. Losing immune cells happens the same way any feature of aging occurs, such as developing gray hair, everyone is different, and it occurs at varying times. There is no specific age where an individual will start losing immune cells. This helps explain why individuals of the same age are seeing vastly different reactions to COVID-19.

This weakened immune system also leads to underlying illnesses that are more common in older adults. For example, scientists are not completely sure what causes patients who are 65 and older to be more likely to be diagnosed with cancer, but one possibility is their lowered immunity does not recognize and target mutated cells to eradicate them like the immunity of a younger individual. Another possibility is the increased length of time a patient has been exposed to negative factors that can cause cancer, such as UV rays, poor diet, alcohol abuse, or smoking.

Other underlying health risks can come from the physiological changes our bodies go through during the aging process. As someone grows older, blood vessels change, and the heart is unable to beat as fast. This combined with the buildup and hardening of plaque in arteries that occurs over time make older individuals more susceptible to heart attacks, strokes, heart disease, high blood pressure or heart failure. All these circumstances combined are reasons for why a country's vulnerable 65 and older population is the most significant factor in predicting new deaths from COVID-19.

To safeguard this population from contracting the virus, there are a number of protections that can be put in place. First, a substantial issue that needs to be addressed regarding this group of people is assisted living and retirement facilities. This population is already more susceptible to a very adverse reaction to this virus, and it is spread easily and quickly from person-to-person, making group homes a disastrous situation if the virus is able to find its way in. Due to this, strict guidelines must be put in place at any congregate elderly care residences.

All communal spaces and events in the community will have to be completely shut down until there is a way to cure or prevent the virus. No visitors will be allowed inside any of the buildings. Visiting is only allowed to occur in outside areas, with masks on, with a guaranteed six feet of distance or more. More rapid tests need to be produced and delivered in bulk quantities to these communities. All staffers, visitors, or delivery workers will need to be tested before allowed onto the property.

For the older populations not living in communal spaces, community structures need to be put in place to look out for these individuals. We need to ensure that this population has family, friends, or neighbors to rely on to help them get necessary items or willing volunteers to run errands for them. All essential stores, grocery stores, drug stores, and others, should be required to have special hours for this older population. During those hours, far fewer amounts of people may be allowed in the store at once, masks must be required, and staff should keep their distance from the patrons, even when being asked for help.

This vulnerable population is also more likely to need doctors' appointments or to have an unexpected hospital visit. Protocols need to be put in place to protect these individuals at these high-risk locations as well. Doctors' offices should have special times to see their older patients and hospitals should set up special areas away from any COVID-19 patients to intake seniors.

Similar to the older population, the second most important variable in predicting new deaths, the percentage in extreme poverty, also need to be protected. Comparably, this population is also more likely to be residing in a communal living space, whether it be a homeless shelter or government subsidized housing. Equal protections need to be established in homeless shelters as senior care facilities.

However, it is not as easy to enforce these protections at low-income housing complexes because there are no caretakers or on-campus staff. The best that can be done is to post signs with suggested rules, such as not sharing an elevator, always wearing a mask, and keeping hands washed. Hand sanitizer stations need to be set up around the complex as well as regular deliveries of soap to all residents.

Another reason individuals in extreme poverty are susceptible to dying from COVID-19 is the lack of healthcare and nutritious foods available. The government and non-profits need to instill programs to remedy this situation, especially during this dangerous time. Affordable, healthy foods need to be made accessible to low-income areas through weekly deliveries or pop-up markets. The

same can be done for healthcare as well. Monthly visits from an array of specialists funded by Medicaid to help protect this at-risk population.

While we can put protections in place to look after our most vulnerable populations, the only new death predictor variables we have any control to change are new tests and the stringency index. New tests, especially ones with rapid results, can quickly help to stop the spread of the virus. If staff at restaurants, community living environments, hospitals, doctors' offices, and more can guarantee they are negative before stepping into work, this will not only curtail a fast spread, but also give people more comfort and confidence to go back to a more normal lifestyle. This is particularly important with COVID-19 because there is a significant percentage of individuals who are positive with the virus but have zero symptoms. These asymptomatic carriers, combined with a lengthy incubation for patients who end up having symptoms, drive a quick and stealthy spread of the virus. Constant, recurring testing is a clear-cut way to minimize the spread.

The other factor we can control, stringency index, is the governmental response to COVID-19, including school, work, and business closures. A stringency index of zero means no rules are in place to combat the virus and 100 is the strictest response. This aspect of COVID-19 prevention is a balancing act. Citizens look up to authority figures in government and expect them to know more about what is going on. Due to this, these citizens rely on the leaders of their country to know what is safe and unsafe in this unprecedented time. By shutting down schools or restaurants, it is a clear message that going to these locations is risky. However, keeping them closed has negative effects. Students who rely on school for meals will be left without it. Parents who rely on school for childcare will not have it. Children who do not have the technology capabilities to learn remotely will fall behind. Workers from a closed business will be without any income. Closed businesses will be without customers or revenue to sustain themselves.

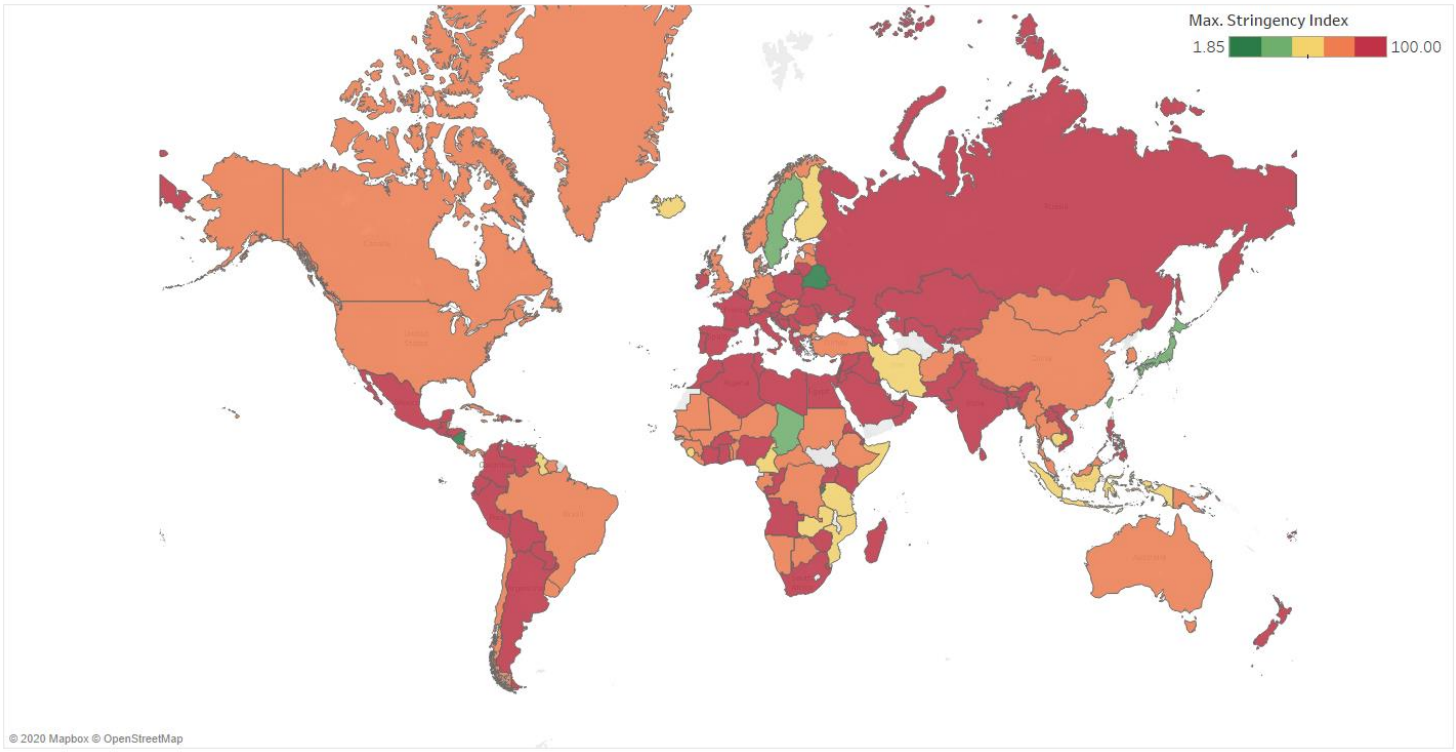
Taiwan and Indonesia were the first to put in place any rigid rules to try and combat COVID-19 at the beginning of January. They put them in place before ever seeing one case. China followed a few days later, but they already had 59 reported cases. Taiwan did not see a case appear in their country until January 21st. Indonesia had much better luck, keeping COVID-19 at bay until its first two cases materialized on March 2nd, even after their Northern neighbor, Malaysia, found its first three coronavirus cases on January 25th.

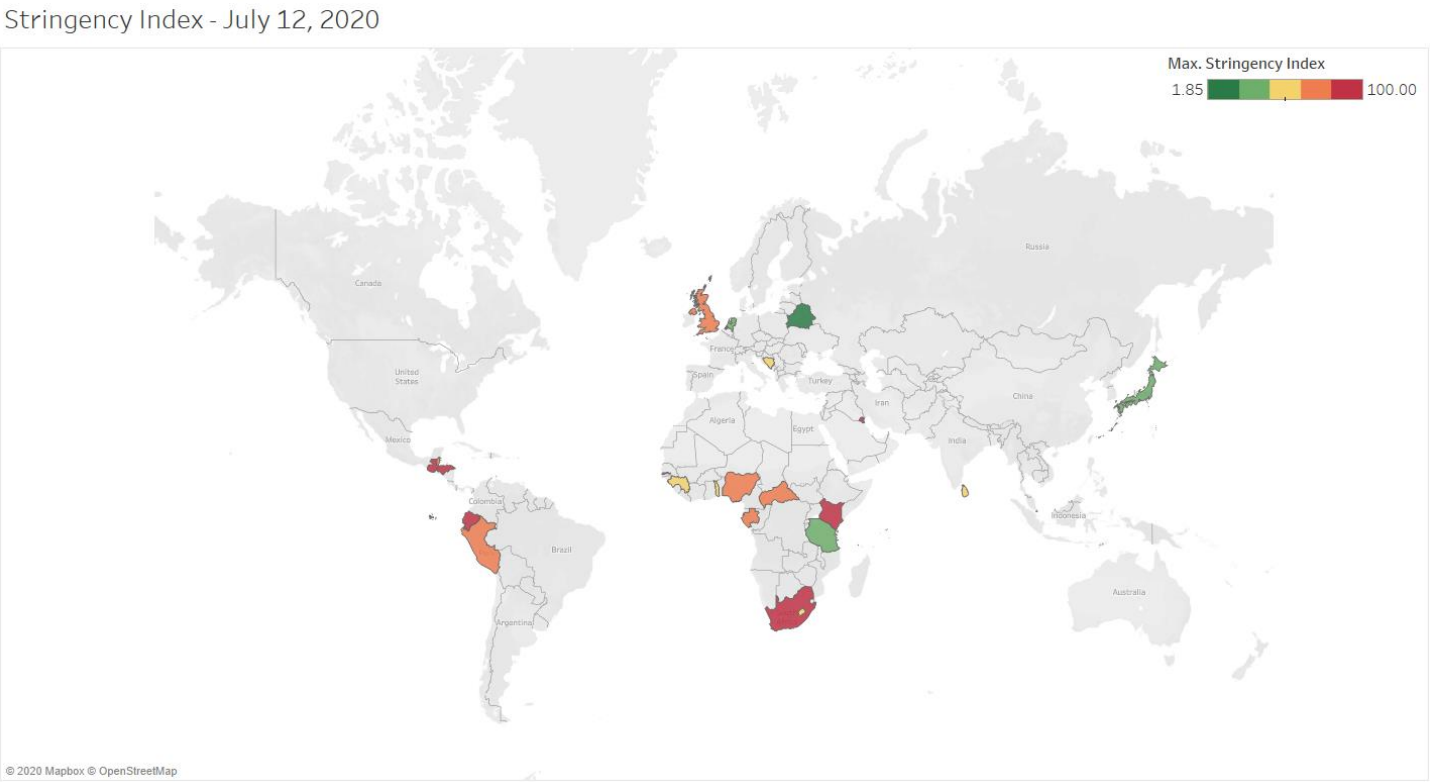
China never introduced a stringency index higher than 81.94. For comparison, 108 countries had stringency indexes higher than this at some point during the pandemic. The United

States was not one of those 108 countries, its highest was 72.69. Thirteen countries went all the way to 100.

By April 1st, most countries had some level of rules in place to try and combat the spreading virus. However, by July 12th, only 25 countries would have any restrictions left in place even though on that day the world saw almost 218,000 new cases and almost 5,000 new deaths.

Stringency Index - April 1, 2020





This movement towards lower stringency indexes could remain, without having the virus spread like wildfire, if enough testing is available. According to this random forest predictive model, the more tests we have, the less we need higher stringency indexes. This model predicts the world will have 3,660 new deaths at the end of August, if there are zero new tests and a stringency index of 81.02. If the stringency index is bumped up to almost 100, we could save almost 1,000 lives. However, the entire world would be almost completely shut down with barely any production and massive loss in jobs with the majority of the population unable to pay for shelter and food. This model predicts the world would see 2,671 new deaths by the end of August with zero new tests and a 98.15 stringency index.

On the other hand, if the world distributed eight million new tests, the stringency index could drop all the way down to 15.74 and save over 2,500 lives. With the percentage of the world's population who are 65 years and older, 8.7, and the percentage in extreme poverty, 10, not changing, the random forest model predicts the world would experience 1,132 new deaths at the end of August as long as we supply eight million new tests. This would protect over 1,500 more lives with a stringency index of 15.74, 82.41 lower than the one needed with zero tests.

Therefore, producing and distributing more tests would allow the world to open back up with very few restrictions. Individuals could get back to work and businesses could start to rebuild after so much loss. The only requirement to save lives without completely shutting down the world is ample testing and standard protocols of masks, gloves, and staying distant.

According to this COVID-19 data, there are a number of recommendations decision-makers can put in place to save the world from more tragedy. First, a higher stringency index must remain in place while production of rapid COVID-19 tests ramps up. Second, protections must be put in place around our most vulnerable populations at senior living facilities and homeless shelters. Along with procedures to keep them safe, they should also receive the initial batch of tests to ensure they remain cared for, but protected, and to safeguard against the virus ever finding its way inside these communities. Other communal spaces, such as hospitals and penitentiaries, should also receive testing for anyone that comes inside.

The next group of tests should be delivered to businesses deemed essential, such as grocery stores, drug stores, and convenience stores to ensure their workers are remaining healthy before they come in contact with the general population. Even with testing available, these places of business must still be required to maintain special hours for vulnerable populations as well as face masks for all workers and customers.

Essential healthcare facilities should be the third group to receive testing. This can be determined through health insurance. Specialists that insurance companies assess as elective will not get priority for COVID-19 tests with this third group. These tests will be reserved for the offices of primary care physicians, oncologists, cardiologists, and other necessary health needs to make sure citizens are not falling behind on important health checks.

Once there is enough testing for a certain location, then the stringency index can be lowered to allow that location to open back up. For example, once there are enough tests for students, teachers, bus drivers, and other school staff to be tested before attending school each day, schools can reopen. Once there is enough testing for restaurants and bars to test staff and patrons, their business may open. Once salons and spas have the capability to test all customers and workers, then their services will be available again.

Companies who are able to continue practicing business with their employees working from home must purchase enough testing for anyone who steps inside their offices to be tested before entering. If they are unable or do not want to comply, then their staff must continue working

from the safety of non-communal workspaces. In the same regard, public events and large gatherings can only take place if testing is provided for all participants.

Similar to businesses with work from home capabilities, religious services will also need to remain online. The only exception to this will be in instances where the service can occur outside with a limited number of people spaced six feet apart and wearing masks. No sharing or touching with other members of the congregation will be allowed.

While this distribution is occurring to make it possible to go back to a normal lifestyle, ample, free testing must be set up for all people. Each case will be assessed separately on whether or not a rapid test is needed. If the individual needing a test can quarantine at home without losing a job or infecting anyone, then the rapid tests can be saved for a more dire situation.

If all this testing becomes available, restrictions will start to be lifted, which will restore wages to citizens. With incomes back in place, stimulus money from the government will no longer be necessary and that money can be used towards the production and distribution of more tests. With testing, travel bans will no longer be necessary as long as visitors are able to be tested upon arrival.

Most businesses and healthcare facilities are relying on temperature-taking as an easy, prevention method to send away potentially sick individuals. However, taking one's temperature does not factor in those who are asymptomatic carriers, currently in the incubation period, or do not have fever as a symptom. There are vastly different reactions to COVID-19, which means checking for a fever, while easier to accomplish than testing, will not provide the same level of confidence as negative test results.

A big area of concern with these recommendations is the accuracy of testing, especially ones with rapid results. In order to ensure a staff member arriving at a senior care facility or a restaurant does not test positive for COVID-19 before entering the premises and starting work, the results must be returned quickly. However, the faster the results, the more likely the patient will receive a false positive or negative. If it is a false positive, the only consequence would be sending home a perfectly healthy worker when the help is needed. On the other hand, a false negative would negate the use of testing. This worry can be tampered with the use of masks. With constant use of masks, even with testing in place, the spread of the virus will still be hampered even if testing produces false negatives on occasion.

The second area of concern is the reliance on all community members to follow basic guidelines set forth by governmental officials. These guidelines consist of wearing a mask, keeping distance from others, and washing hands often. Some countries will be able to force their residents to obey these rules, but not all of them have the ability to do this and having individuals not comply will put all steps for improvement at risk. The same goes for ensuring compliance from establishments not allowed to be open yet or maintaining testing once they are allowed to open.

The third area of concern is the novelty of this virus. Insights found today from analyzing past COVID-19 data can be different tomorrow because it is a rapidly changing situation. Also, the varying metrics reported by each country can lead to missing data observations that can alter the results or skew the analysis. However, as new data becomes available or missing data is filled in, this analysis can be replicated and redone.

According to this analysis, there is a solution that will save lives without causing further economic distress. To successfully prevent deaths, a standard set of protocols must be followed, such as wearing masks, maintaining a safe distance, and washing hands. Additionally, the top priority needs to be producing more tests along with increased distribution starting with the most vulnerable populations. If this is accomplished, then this model predicts 2,528 lives will be saved globally, by the end of August, from an unnecessary COVID-19 death.