

**USING DYNAMIC MODELS AND EMPIRICAL COVID-19 DATA TO SHOWCASE
EFFECTIVE PANDEMIC PREVENTION MEASURES**

An Interactive Qualifying Project Report:
submitted to the Faculty
of the

WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science by

Kade Woolverton
Matthew Withington
Ivan Nikulin

In collaboration with:

Astra Nikitina
Olga Skiba
Karina Nurgalieva
Anna Kozhieva

and

Professor Svetlana Nikitina, Advisor
Professor Losev Anton Alexeyevich, Advisor
Professor Dmitri Victorovich Korovin

This report represents the work of one or more WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on its web site without editorial or peer review.

I. Abstract

Pandemics, such as COVID-19, transcend borders and require adequate intervention measures from countries if they are to be contained. In collaboration with students and professors from the Financial University in Moscow, our team assessed six major countries' strategies for mitigating virus spread and developed pandemic models using AnyLogic, a dynamic modeling application, and Microsoft Excel. The international research and interactive models informed recommendations to governments on the following effective response protocols: contact tracing, prompt widespread testing, and uniform quarantining.

II. Table of Contents

I. Abstract	2
II. Table of Contents	3
III. Introduction	5
IV. Background/Literature Review	6
COVID-19	6
Factors Contributing to Spread	6
World Health Organization Recommended Response Protocol	7
Response Protocol and Factors Contributing to Spread by Country	8
United States	9
Russia	11
Republic of Korea	12
Italy	14
Germany	15
China	16
Modeling Novel Virus Spread	17
AnyLogic Predictive Modeling	20
V. Methodology	21
Teamwork and Project Development Process	21
Team Communications and Joint Work Organization	21
Developing a Thorough Understanding of COVID-19	21
Comparison of Response Protocols and Virus Impact across Target Countries	22
Tracking Historical COVID-19 Data (Case and Death Numbers) for the Six Countries	22
Demographic Data and Response Protocol Analysis for the Six Countries	22
Introduction to Different Excel Models	22
Exploring the AnyLogic Modeling Application	23
Models and Documentation Process	23
Use of Models and Other Data to Inform Recommendations	24
Compiling of Recommendations and Relevant Models onto a Website	24
VI. Results and Analyses	25
Correlation Between Response Protocols and Success in Preventing COVID-19 Spread	25
Correlation Between Age Demographics and Deaths per Capita	27
Correlation Between Population Density and Infections per Capita	27

Models	27
System Dynamic Models	28
Agent-Based Models	29
Econometric and Sigmoid Excel Models	30
Website	31
VII. Conclusion	32
Recommendations	33
Future Projects	33
Summary	34
VIII. Bibliography	35
IX. Appendices	39
Appendix A: Documentation of Models	39
The Significance of Using Dynamic Models and AnyLogic	39
Stock SIR Model	42
SEQJIR Model	43
HIRD Model	44
SEIR Model with Sliders	46
Town Model: Agent-Based	48
Econometric Models	53
Gamma Distribution	53
Exponential and Logistic Curves	54
Moving Averages	55
Forecasting Model	57
Chow Test for Model Stability	58
Quadratic Model	60
Sigmoid Models	61
Appendix B: Website	68
Accessibility	68
Homepage	69
Contributors Page	70
Models Page	71
Recommendations Page	72
Country Comparison Page	73

III. Introduction

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, is impacting nearly all major countries around the world in 2020. In light of the widespread disease, countries are being forced to coordinate efforts and learn from each other to limit their infection and death totals. The issue here lies in the fact that there is no one fundamental protocol to institute in response to such an infectious and mobile virus. The World Health Organization (WHO) and Centers for Disease Control (CDC) have released basic guidelines to follow, but other than that countries are forging their own paths in creating what they believe to be the best response protocols.

In the midst of this ever-evolving pandemic, certain countries are seeing successes in their efforts to combat the virus while others are struggling to get control of the situation. With a wealth of information available on countries' response protocols, as well as infection and death totals, correlations can be observed connecting the former to the latter. Beyond factors within the countries' control are unalterable properties like age demographics and population density that also contribute to the spread of COVID-19. This project looks into both these properties as well as into response protocols for six major countries in order to discern which actions seem to be mitigating virus spread effectively. Those countries are the United States, Russia, Italy, China, Germany, and the Republic of Korea.

In addition to researching and compiling information on the developments of the pandemic, this project focuses on modeling projections of disease spread to demonstrate the significance of certain response protocols like quarantine guidelines which mitigate virus spread. The modeling application, AnyLogic, is used as a tool to create dynamic models that display meaningful protocol impact correlations which can provide guidance for dealing with future pandemics.

The ultimate goal of this project is to learn from the models created and provide recommendations on the best practices a country can institute to prevent the spread of viruses such as this coronavirus. Through collaboration with four undergraduate students and two professors from the Financial University in Moscow, data on the development of COVID-19 and a variety of models are compiled to inform these recommendations.

IV. Background/Literature Review

COVID-19

COVID-19 is a respiratory disease that originated in Wuhan, China in late December of 2019. The disease is caused by a novel coronavirus that spreads primarily through respiratory droplets that travel through the air, although it can also live on surfaces for up to three days. Patients who contract this coronavirus may experience shortness of breath, fever, and dry cough, with critical cases resulting in pneumonia, organ failure, and, ultimately, death. Persons with heart-disease, diabetes and pre-existing respiratory diseases like asthma appear to be more at risk of dying from the virus. There is currently no treatment or vaccine for COVID-19 and practically every country has confirmed cases (CDC, 2020).

The virus itself that causes COVID-19 is known as SARS-CoV-2, which in long form spells out severe acute respiratory syndrome coronavirus 2. Many are confused why the virus and disease have different names, but it is essentially the same difference between HIV (the virus) and AIDs (the disease caused by the HIV virus). Despite the common misconception, COVID-19 can spread in all areas, including hot and humid climates. As of March 30th, 2020, 203 countries are reporting confirmed cases (World Health Organization, 2020c).

While COVID-19 contraction is relatively uniform between age demographics, although elderly populations exhibit a significantly higher mortality rate once the virus is contracted. A case study conducted on a population of infected persons from China showed that those in the elderly category had a higher chance of developing stage IV and V pneumonia and also a higher chance of dying than those in the middle-aged and young categories (Lia, Chen, Ruzheng, & Kunyuan, 2020). The international COVID-19 mortality rate actually climbs drastically in the upper age cohorts. Those in the 60-69 range showed a 3.6 percent mortality rate compared to an 8.0 in the 70-79 bracket and a 14.2 in the 80+ bracket (Worldometer, 2020). In addition to age demographics, the mortality rate varies wildly from country to country. As of March 30th, 2020 Italy's mortality rate is 11.39% while the United State's is 1.80% (World Health Organization, 2020).

Factors Contributing to Spread

Despite scientists' best estimates, there is still ambiguity about the exact pathways the SARS-CoV-2 virus spreads through. According to Krista Wigginton of Stanford, "A major reason for this is that the behaviors and traits of viruses are highly variable—some spread more easily through water, others through air; some are wrapped in layers of fatty molecules that help them avoid their host's immune system, while others are naked." Wigginton attributes the lack of information about SARS-CoV-2 to the deficit of research in the past decade on coronaviruses.

There has apparently been more funding for the study of non-enveloped viruses like norovirus and rotavirus. It is probable that SARS-CoV-2 uses proteins on the surface of its fatty layer to go undetected by its host's immune system, but with multiple strains of the virus emerging it is difficult to confirm this (Stanford University, 2020).

The virus does seem to be spreading both through airborne pathways and on surfaces touched by infected individuals who had touched their face with their hand. Airborne viruses like SARS-CoV-2 are emitted in droplets of saliva or mucus when an infected person coughs or sneezes. These droplets can travel multiple feet and, in some situations, can remain in the air for up to ten minutes (Aljazeera, 2020). It is still unknown how long the virus can survive on surfaces outside of a host, but the Centers for Disease Control and World Health Organization concur that it may be multiple days in some cases (CDC, 2020; World Health Organization, 2020). One pathway that is often a concern for virus spread is water. Fortunately, this transmission method can be ruled out since SARS-CoV-2 cannot be excreted in urine and feces, and henceforth cannot make its way into waterways (Wigginton & Boehm, 2020).

The main reason this coronavirus is running so rampant is because its incubation period ranges from one to fourteen days with the average falling around five days. Individuals infected with the virus may be contagious before they show symptoms, allowing them to spread the virus without knowing they have it. Another contributing factor is that some infected individuals will show no symptoms throughout the entirety of the time they are infected by the virus. A statistical analysis done on data from the Diamond Princess cruise ship showed that 17.9% of individuals who tested positive for the virus showed no symptoms at all (with a 95% confidence interval ranging from 15.5% to 20.2%). The testing occurred after a mandated sixteen-day quarantine so it can be assumed that all incubation periods were over. Both the long incubation period and ability for asymptomatic individuals to spread the virus, make the spread SARS-CoV-2 difficult to slow (Lee, 2020).

Understanding the factors that turned the COVID-19 disease into a global pandemic, is the first step towards developing useful response protocols for countries to follow. The succeeding sections look into the different ways that countries have been advised to combat the pandemic, and additionally, the actual measures they have undertaken as nations. Learning the best practices will inform the team's recommendations as well as the models created to emulate effective response protocols.

World Health Organization Recommended Response Protocol

A global crisis like a pandemic has to be addressed on multiple levels. To limit the spread of a disease, people have to organize as a community as a whole and as individuals. The World Health Organization (WHO) recommends that individuals consistently wash their hands with soap or clean them with alcohol-based rubs (WHO, 2020). This measure will reduce the amount of contamination spread by the hands. The WHO also recommends that people keep a distance of

at least one meter, this measure is expected to reduce the spread of the infection through the air (WHO, 2020). According to the personal recommendations, people should refrain from touching their face and covering up the nose and mouth when sneezing and coughing. Measures for social distancing have to be taken as well, such as avoiding unnecessary travel and public places; if an individual is not feeling well, they must stay home. Smoking and anything that weakens the lungs is highly unrecommended (WHO, 2020).

The WHO also made recommendations to countries on preventing the spread of the pandemic and keeping a strong medical care system. The testing and treatment of a pandemic has to be carried out, but many other health conditions should not be ignored during such situations. The organization gives recommendations on achieving the most useful balance between standard medical treatment and that of the pandemic (Maintaining essential health services and systems.2020).

Early investigation protocols are also provided by the World Health Organization. The WHO web page published five protocols that should help a country's specialists most efficiently limit the spread of the pandemic. The protocols address hospital infections as well as household ones. "The First Few X (FFX) Cases and contact investigation protocol for COVID-19 infection" is an example of a protocol that needs to be followed by professionals for tracking and stopping the spread of COVID-19 (Early investigations protocols.2020).

The WHO also provides a guide that is made to be used by the governments to prepare and handle the pandemic over a three month period. It states the priorities that need to be addressed in the protocol. Due to the swiftly changing situation, the guide and protocol might be updated accordingly (Country-level coordination, planning, and monitoring.). Comparing the WHO's recommendations to the actual actions of countries around the world, is a baseline gauge for how well governments are combating virus spread. The recommendations that show promise will be featured in this report and inform the models that demonstrate why they are effective.

Response Protocol and Factors Contributing to Spread by Country

In response to COVID-19, countries around the world are being forced to come up with protocols for combating the virus and keeping their citizens safe. Some countries are able to be more authoritative with their enforcement efforts due to their style of government. Technological and health care capabilities also play a significant role in a country's response to COVID-19. Factors that are out of countries' control, yet contribute to spread and mortality rates, are population density and age demographics. The six countries selected to be featured in this report are each significant in the COVID-19 pandemic for their own reasons. Germany and South Korea have managed to keep their death counts low to date, while Italy possesses the highest COVID-19 mortality rate in the world. China, dissimilarly, is of interest to this project because it completed a full cycling of the virus and is seeing few new cases. Lastly, Russia and the United

States are the two countries in which team members reside, so information and historical infection data is more easily accessible.

United States

The United States government issued a quarantine protocol for the entire country, but only after individual states began instituting their own precautionary actions. This federal protocol recommended social distancing which means no social gatherings of ten or more and limiting discretionary travel. There has not yet been a stay-home mandate and as of March 31, 2020, there is no enforcement of the quarantine protocol (usa.gov, 2020).

Many states have reacted to the pandemic by closing all non-essential businesses and public schools. Hospitals, grocery stores, banks, and fuel stations are all considered essential businesses and thus remain open with steps taken to protect the businesses that were shut down (Department of Homeland Security, 2020). In the United States, each state is responsible for the timeline and extent of their protocols and state legislators are adjusting those accordingly as the situation develops (Aljazeera, 2020).

In light of the exponential increase in infected individuals, many hospitals are experiencing shortages in the supplies they need to keep patients and staff safe. These supplies include masks, beds, and ventilators. The federal government has not yet released a distribution plan for these resources and states are being left to place their own orders for their hospitals and hope they get the supplies they need. Within individual hospitals, doctors are having to allocate resources for those most likely to survive because there are not enough to go around. One possible way to combat these resource shortages is with the Defense Production Act. This act, designed for wartime efforts, mandates that companies switch to production of goods that the country needs. While individual organizations have already begun producing masks and beds on their own agenda, the president has yet to put this act into effect (usa.gov, 2020).

The United States currently has 94 public health laboratories that are certified by the CDC and are actively testing samples for SARS-CoV-2. All fifty states have at least one laboratory that is conducting testing and these test data are informing infection counts (CDC, 2020). The first confirmed case in the United States occurred on January 20th, 2020 and as of March 30th, 2020 there are 163,539 cases in the US as seen in Figure 1 (CDC, 2020).

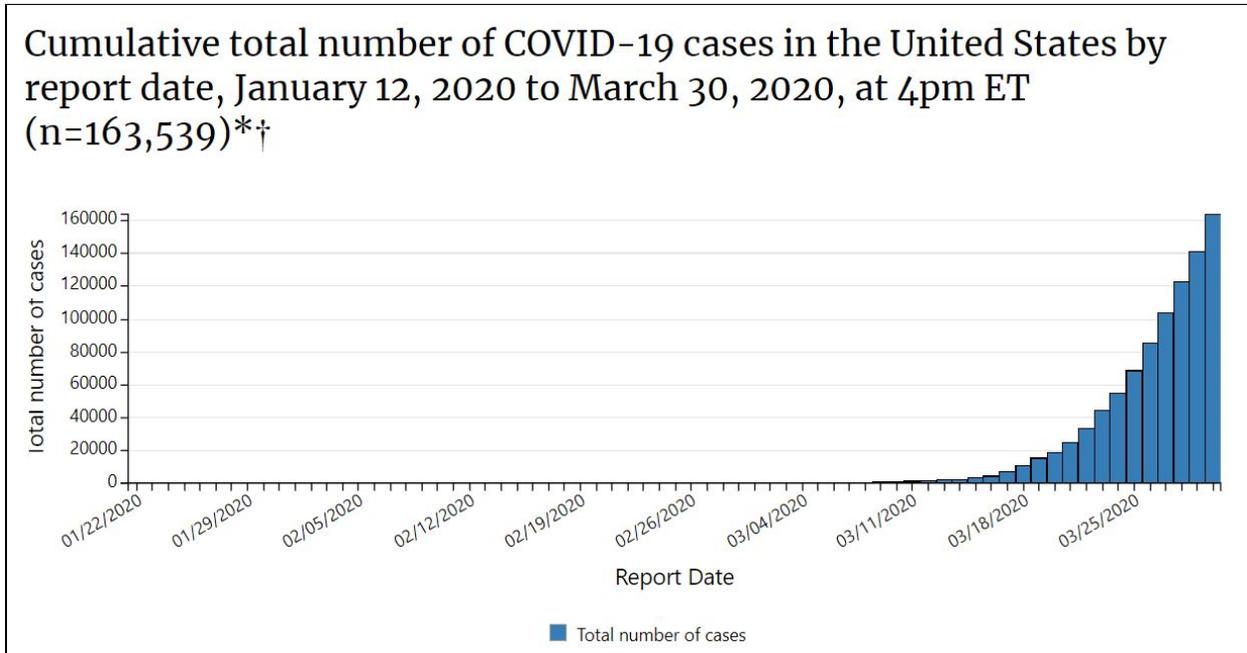


Figure 1: COVID-19 cases in the US over time (CDC, 2020).

Throughout the entirety of March, 2020, there has not been a day where the number of new confirmed cases did not exceed the prior day in the United States (CDC, 2020). It is not possible to accurately calculate a mortality rate in the United States at this time because there are too many new active cases. In terms of strict number of deaths, the United States is managing to stay below countries like Italy and Spain, despite having the highest number of confirmed cases (CDC, 2020). According to Peter Slavin, the president of Mass. General Hospital, “[lack of] testing is the biggest problem [the US] is facing.” Compared to South Korea which has tested 4000 out of every million citizens, the United States has only tested 5 out of every million as of March 30th (Thomke, 2020). Population density in cities is another factor contributing to the severity of the epidemic in the United States. Almost half of the total cases and total deaths occurred in New York state, with the vast majority in New York City (Worldometer, 2020). New York City is the most densely populated city in the United States with a density of 26,403 people per square mile (Population - new york city population.2015).

Given that the mortality rate for elderly citizens is significantly higher, the percentage of elderly persons in a population is significant. In the United States, individuals who are sixty years or older make up approximately 21% of the total population (Rogers, 2016).

The factors in the United States’ that may have impacted the spread of the COVID-19 pandemic were a slow initial response, a lack of prompt widespread testing, and a discrepancy in the ways that different states responded. These factors can be compared to other target countries and analyzed for their significance.

Russia

In January 2020, an operational headquarters was created in Russia to combat COVID-19 under the leadership of Deputy Prime Minister Tatyana Golikova. The committee started limiting the crossing of Russia's borders, by the end of the month the border with China was closed, and on February 20 citizens of China were banned from entering the country. The first two cases in Russia were recorded on January 31. The cases were identified, patients were isolated, and were recorded as recovered on February 11. The virus was recorded again on March 2 in Moscow. After the third case, the Russia's government started introducing quarantine and isolation measures to the public. Multiple measures were taken to promote the inflow of resources into the country, such as food and medical equipment. A period with the suspension of non-essential jobs was introduced on April 3 until the end of the month (Times, 2020).

The Russian Federation has one government certified testing laboratory and a number of labs in privately-owned medical centers. As of April 11, the amount of tests conducted is over a million, making it an average of 8.21 per one thousand people as of April 11th (Russia: The number of tests performed.2020). Russia as of April 11 has not yet reached the peak of COVID-19, as seen in Figure 2.

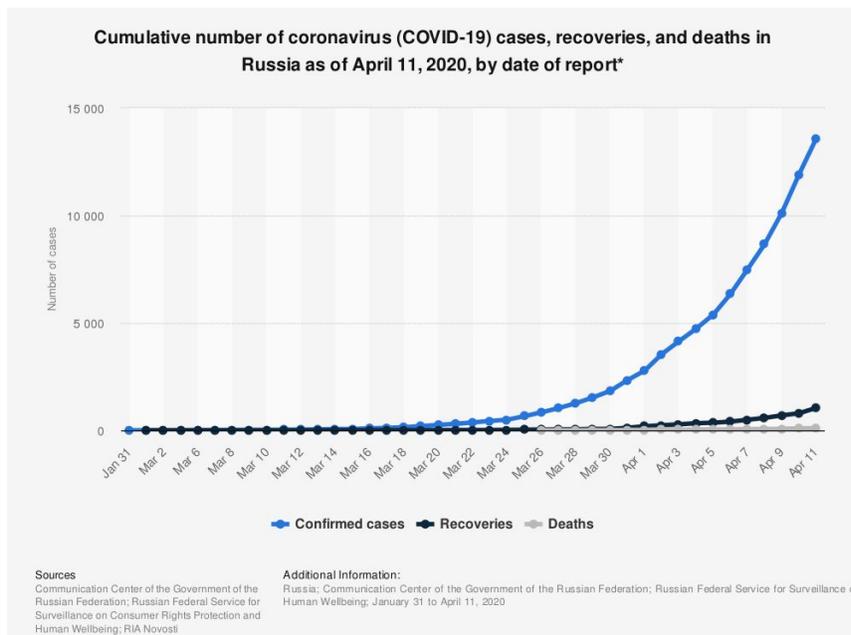


Figure 2: COVID-19 cases in Russia as of April 11, 2020 (Statista, 2020)

Population density in Russia is not uniform, with the highest densities being in the western part of the country in the largest of the country's cities with few exceptions on the east side of the Ural Mountains (Figure 3). Considering the given statistic is crucial, when determining the best tactics and strategies when trying to limit the spread of the epidemic. The epicenters of COVID-19 are located in the largest cities, which is an expected occurrence.

Moscow, being the most densely populated megapolis, is leading in the amount of virus cases (First Channel, 2020). Moscow is most likely the best prepared for the outbreak, having the most developed and modern healthcare system in the country, although the ratio of hospital beds per person is 5/1250 (Количество стационарных койко-мест в больницах.2019) Количество стационарных койко-мест в больницах.2019).



Figure 3: Distribution of Cases in Russia as of April 8, 2020 (Russian BBC service, 2020)

The lack of adequate financial government support of the population, scarcity of tests, low overall wealth of the population, overpopulated areas and low self-isolation indices of the population were factors that contributed to the quicker and wider spread of the pandemic. Factors for preventing the spread of COVID-19 were the encouragement of self-isolation in overpopulated areas and mandatory shutdown of non-essential businesses. The factors influencing the spread of the virus were considered by the project team and included in the analysis.

Republic of Korea

As of March 31, 2020, the Republic of Korea has 9,786 total confirmed cases (Figure 4) of COVID-19 and 162 total deaths resulting from COVID-19 (World Health Organization, 2020b). Additionally, as of March 31, 2020, the Republic of Korea has the 13th most reported cases of COVID-19 out of the 203 countries reporting confirmed cases (World Health Organization, 2020a). As mentioned in the US section of this report, the mortality rate for citizens above the age of 65, with COVID-19, is significantly higher than the younger age demographics (Ministry of Health and Wellness of Korea, 2020). 15.92% of the Republic of Korea's population falls in the 65 or older category (CIA, 2020b). For the citizens of South Korea that died due to COVID-19, the age demographic of 60 and older contributed to 92.40% of these deaths (Ministry of Health and Wellness of Korea, 2020).

Although the number of the Republic of Korea’s confirmed cases seems high, the nation has been extremely effective in mitigating the spread of the SARS-CoV-2 virus to the susceptible population. Through rapid and aggressive epidemiological testing and investigations, South Korea was able to severely retard the diffusion of the SARS-CoV-2 virus throughout the country (Ministry of Health and Wellness of Korea, 2020). Once a case is reported, the local and state governments work in tandem to track—through credit card details, cellular geolocation data, CCTV footage—the movement of an infected individual can be tracked (Ministry of Health and Wellness of Korea, 2020). Then, this information is transparently disclosed to prevent further transmission to the general public and for them to check whether or not they have come in contact with it themselves (Ministry of Health and Wellness of Korea, 2020). The Republic of Korea has quarantined tens of thousands of suspected cases through these methods and although most of these people are healthy, the people who do fall ill can be rapidly isolated and have little risk of further transmission to the susceptible population (Beaubien, 2020). Through the nation’s preventive measures, South Korea has been able to keep most factories, shopping malls and restaurants open (Beaubien, 2020).

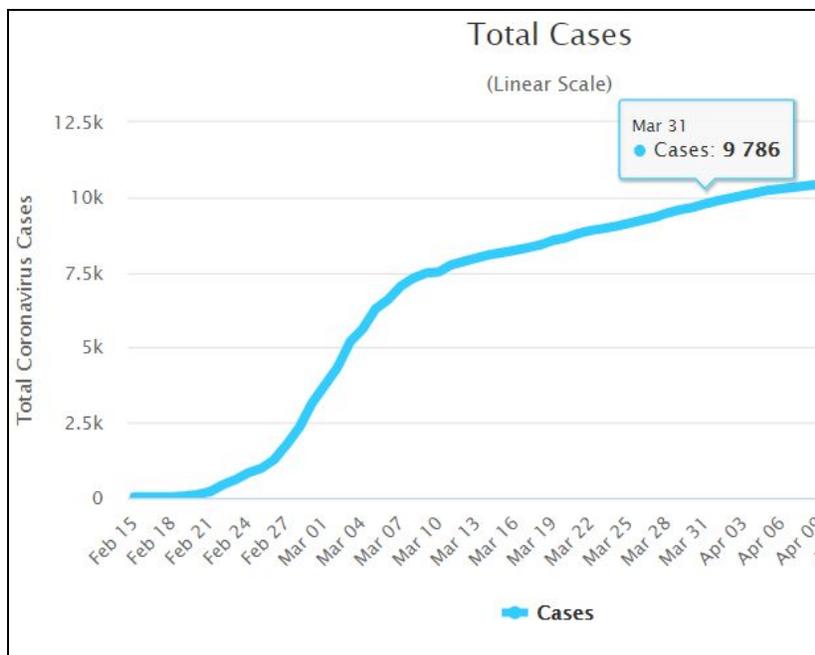


Figure 4: Total COVID-19 infections confirmed in Republic of Korea (World Health Organization, 2020b).

The government of South Korea was able to implement extensive measures in combating COVID-19 spread. The factors that contributed to the effective halting of the virus spread were the high testing rates, encouragement of quarantine, a high overall population wealth and higher quality medical care. The presence of overpopulated areas was a factor that promoted the holdback of effectiveness of the preventive measures. The influential factors to the spread of the

pandemic in South Korea were analyzed and used in the recommendations that were considered effective by the project team.

Italy

As of March 30, 2020, Italy has 101,739 total confirmed cases (Figure 5) of COVID-19 and 11,591 total deaths resulting from COVID-19 (World Health Organization, 2020b). Additionally, as of March 33, 2020, Italy has the 2nd most reported cases of COVID-19 out of the 203 countries reporting confirmed cases (World Health Organization, 2020a). As mentioned in the US section of this report, the mortality rate for citizens above the age of 65, with COVID-19, is significantly higher than the younger age demographics (Italian National Institute of Health, 2020). 22.08% of Italy's population falls in the 65 or older category (CIA, 2020a). For the citizens of Italy that died due to COVID-19, the age demographic of 60 and older contributed to 95.1% of these deaths (Italian National Institute of Health, 2020).

The state of Italy in view of the pandemic was rather grim, but the situation was improving due to the drastic measures the Italian government enforced to slow the infection rate of the susceptible population. Italy lagged in properly containing the spread of the virus. Nevertheless, a nationwide lockdown is being enforced by the military; factories and stores are also closed (nearly entirely shutting down the production of the country) (Italy, pandemic's new epicenter, has lessons for the world.2020).

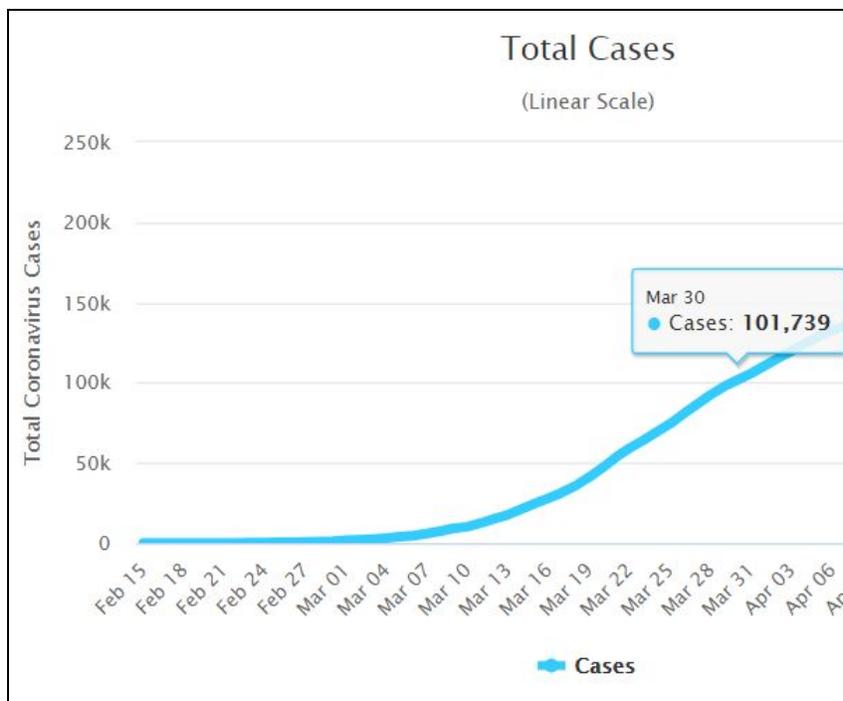


Figure 5: Total COVID-19 cases in Italy confirmed (World Health Organization, 2020b).

One factor that contributed to the high COVID-19 mortality rate in Italy was the high rate of people flowing through the country. The high quality of medical care, quarantine protocols and the closing of non-essential businesses were the factors that caused the plateau of new cases to occur sooner. The positive and negative influences on the situation with COVID-19 in Italy were analyzed by the project team and applied in making recommendations.

Germany

The first confirmed case of COVID-19 in Germany was recorded on January 28. The disease was spreading slowly until the beginning of March. On March 23 the first measure to prevent the spread of the virus was made by the German government, a contact ban was set in place on the national level. The ban forbids the gathering of two or more people in one place, with the exception of people that live together and public transportation. The contact ban is enforced by local police departments. Before the ban was declared by the government an advisory to all public institutions such as schools, museums, gyms and restaurants with the exception of grocery stores and pharmacies have been asked to close until further notice (Koptyug, 2020).

The number of labs reporting test results in Germany seems to vary from week to week, but more than 150 labs have been reporting the results. As for April 8th German labs have tested 1,317,887 samples. That number can also be represented as 15.97 tests per one thousand people. The number of tests per week has been growing and is expected to grow. Germany is opening drive-thru testing stations in order to maximize the coverage of COVID-19 (Germany COVID-19 cases and deaths statistics.2020).

The factors that may have helped Germany combat the pandemic were the high ratios of hospital beds per person (Loh & Kresge, 2020), and breadth of tests conducted early on. Based on the first two months of COVID-19 presence in Germany, the government's efforts at reducing spread seem to be effective. As is evident in Figure 6, the number of daily confirmed cases has begun to decrease at the end of March.

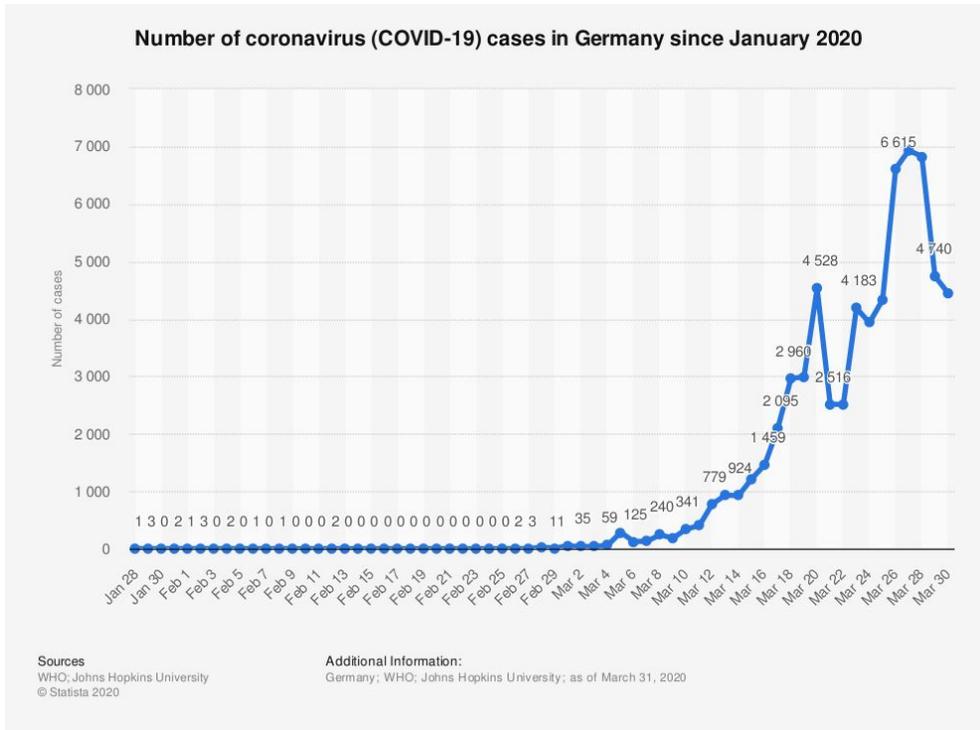


Figure 6: COVID-19 cases in Germany since January 2020 (Statista, 2020).

China

Wuhan, China is considered to be the origin of the global COVID-19 pandemic. It supposedly started in a seafood market in the beginning of December of 2019. The situation escalated quickly and multiple patients were hospitalized with cases of pneumonia. The Chinese government cancelled the official celebration of Chinese New Year and locked down 34 cities. Quarantine was very strictly enforced, with police and the military employed. To prevent overcrowding a hospital center was opened on February 3. With all the preventive measures taken the situation stabilized and a decrease in cases is happening (Thomala, 2020).

China's government has not published any official information on the amount or methods of COVID-19 tests conducted. Some information is kept out of the public reach, but China's COVID-19 protocol is published along with statistics on infection rates, recoveries and deaths. From what has been reported on the news, it is known that China's government employed the army to strictly enforce quarantine and isolation measures. Roads were blocked and people were denied leaving their homes other than to get groceries once or twice in two weeks. Such strict measures seem to have proven to be effective, but costly.

China was able to raise the number of tests they were conducting and saw a plateau in total numbers of cases as is evident in Figure 7. With the increased amount of testing, the death rate dropped, giving the authorities a more accurate view of the developing situation.

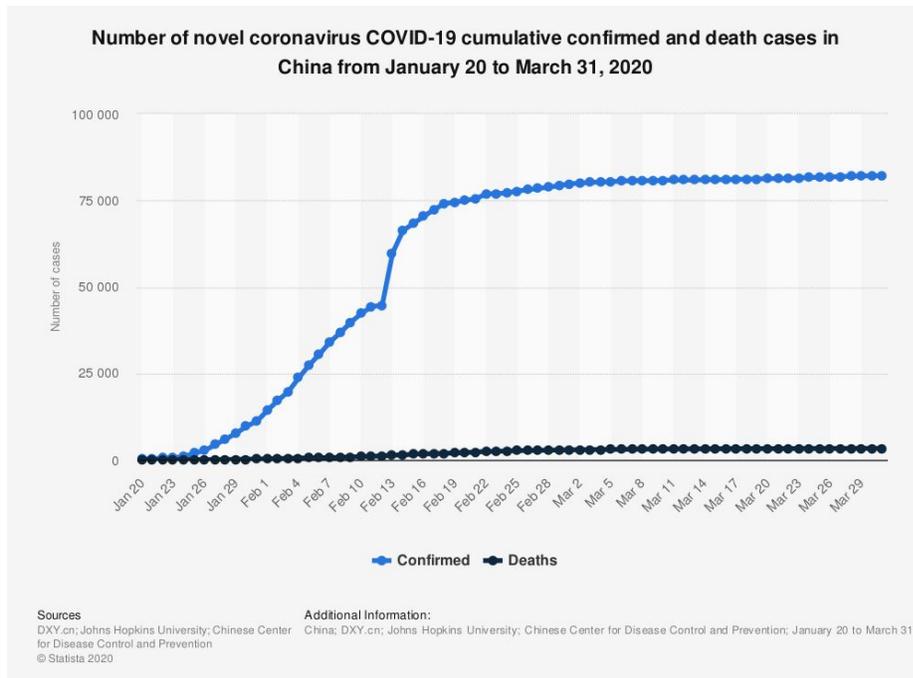


Figure 7: COVID-19 cases and deaths in China Jan 30-Mar 31, 2020 (Statista, 2020).

Factors that contributed to the stopping of growth of COVID-19 cases in China were strict quarantine measures, the shutdown of non-essential businesses, and a high availability of testing. The factors that caused the pandemic to spread faster was a high population density and a lower overall wealth of population. The factors that influenced the pandemic situation in China were used by the project team in making recommendations and assembly of dynamic models.

Modeling Novel Virus Spread

While it is useful to present the public with predictive COVID-19 case numbers and death totals in order to educate people on the situation the world faces, it is generally left unexplained what these predictive models are assuming when looking at future confirmed SARS-CoV-2 case totals. These models could be assuming no quarantine actions were taken in a specific area, which is frightening but unlikely as many countries have shown that proper intervention vastly mitigates virus spread. For instance, the Republic of Korea has demonstrated its ability to stop the SARS-CoV-2 virus from overwhelming their health infrastructure, seemingly by utilizing contact tracing to effectively isolate and quarantine people that have; been exposed to the virus; infected with the virus; or have a chance of coming in contact with areas/people that are in the two aforementioned states. For the rest of the world to learn how to replicate the success of virus control in South Korea, a model of the epidemic management in the nation must be developed. By developing various predictive models that were compared to the SARS-CoV-2 case data from South Korea and assuming quarantine actions were taken, the

nations' successful preventative measures could be extrapolated and applied to other nations around the world.

The most established way to predict the evolution of a population in an epidemic is to use compartmental models of epidemiology. Compartmental models can be used to understand and display the spread of the SARS-CoV-2 virus in a simple, effective manner. Furthermore, compartmental models can be used to predict the spreading of the SARS-CoV-2 and the mitigation of the spread via quarantine and isolation procedures; such as self-isolation (Brauer, 2008).

There are two underlying assumptions of compartmental models. The first is the “homogeneity of individuals”, meaning the members of a population have no individual qualities and act the same way in each compartment (Gallagher, 2017). The second assumption is that, “the rate of change of individuals in the compartment at the next step is proportional to the number of individuals in the compartment at the current step” (Gallagher, 2017). These two assumptions help describe how individuals in a population evolve over time. Therefore, compartmental models have a long and rich history of being used for modelling epidemics (like SARS and influenza) but there are some drawbacks to these models.

A rather sweeping assumption that a compartmental model makes is the presumption of homogeneity of individuals. If this principle is applied in a small commune, it is accurate and precise, yet when applied in a city like New York, it is hardly justified to assume. Various people have differing health histories and interactions with other members in the population, so assuming that all members of a society will act the same is to simplify the model.

Yet, predictive models that only use compartmental methods are now considered outdated (Brauer, 2008). Compartmental models have existed for nearly one hundred years, and with the advent of computer simulation; as well as the need to predict the use of hospital resources/infrastructure, hybrid models have become the cutting edge in epidemiology (Gallagher, 2017). Compartmental models can be altered and combined with a newer type of simulation, agent-based models, to reduce the impact of this assumption.

Agent-based models are more concerned with simulating the spread of a disease rather than the evolution of a population over time, in response to an epidemic (Gallagher, 2017). The driving assumption behind agent-based models is the “heterogeneity of agents”, where agents can be basically anything, including people, buildings, planes, airports and cities (Gallagher, 2017). Additionally, these agents have specific characteristics that distinguish them from other agents; this could mean one person has a weaker immune system than another, or one city has more immigration into it than another. Nevertheless, agent-based models are more complex to construct and are sometimes computationally demanding, which is why this project is primarily focused on constructing compartmental models, with some agent-based features, to display the effectiveness of isolation/quarantine measures in an epidemic situation.

Since we were concerned with modelling the mitigation of the SARS-CoV-2 virus via quarantine/isolation measures, a compartmental model, known as an “Epidemic Management

(Quarantine-Isolation) Model” (Brauer, 2008). In the case of the SARS-CoV-2 virus, there are many advantages to using the epidemic management model to understand why the Republic of Korea is so successful in retarding the spread of the virus.

The epidemic management model that we are using has the following base structure and assumptions, identified by Dr. Fred Brauer, an expert in the field of mathematical modelling of epidemiological problems and population systems (Brauer, 2008) :

- There are six compartmental sections of the model: $SEQIJR$

S: Susceptible - this is the population of healthy people that are at risk of contracting the SARS-CoV-2.

E: Exposed - this compartment is the section of people that have been exposed to the virus and are then sent to the quarantine section.

Q: Quarantined - this compartment is the section of people that have been exposed to the virus and are removed from the population to reduce contact rate.

I: Infected - this compartment is the section of people that have contracted the virus and are infectious (capable of spreading the SARS-CoV-2).

J: Isolated - this compartment is the section of people that have contracted the virus and have been removed from interacting with the rest of the population to reduce contact rate.

R: Recovered - this compartment is the section of people that no longer have the virus and come out of isolation or they didn’t contract the virus after being in quarantine.

Figure 8 represents the flow chart of the $SEQIJR$ model. It is an excellent illustration on how each compartment interacts with its respective neighbors

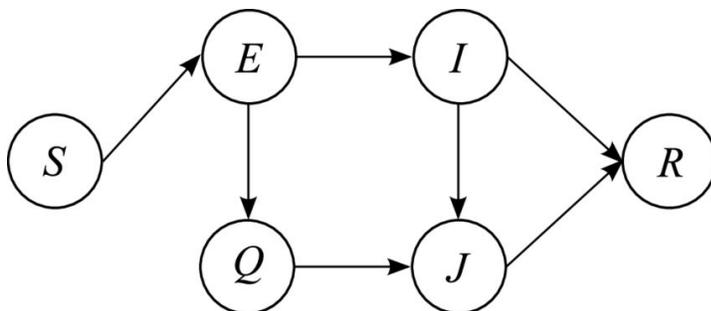


Figure 8: A flow chart of the $SEQIJR$ compartmental epidemiological model. In this form, it is far easier to realize how each compartment interacts with other compartments.

The advantages of this model lay in its ability to display perhaps the most important compartments of mitigating the spread of a virus; the quarantine and isolation compartments. In the quarantine and isolation compartments, members of a population are taken out of contact with healthy people, reducing contacts per day of those exposed people and waiting until they are recovered to return them to the rest of the population. If a normal SEIR model were used to

predict the evolution of a population in this epidemic, the model would immediately be at a disadvantage and would assume that no preventative measures were taken to halt the spread of the SARS-CoV-2 (Brauer, 2008).

Moreover, the SEQIJR model can be calibrated using the AnyLogic software and historical data collected from the six countries this report is focusing on (the Republic of Korea, Russia, the United States, Germany, Italy and China). The use of AnyLogic software would not have been possible without the recommendation of the software from Professor Anton Alexeyevich Losev and Professor Dmitri Victorovich Korovin. Without the recommendation and assistance that these professors offered with the AnyLogic platform, the rapid development of subsequent epidemiological models that were produced by our group would not have been possible.

AnyLogic Predictive Modeling

AnyLogic is a professional simulation and modelling program that enabled this group to create and study dynamic models of the SARS-CoV-2 virus in each of the following countries: the Republic of Korea, Russia, the United States, Germany, Italy and China. While compartmental models can be built in computational programs such as Microsoft Excel, agent-based models are extremely difficult to create in this program. AnyLogic grants the ability to create, modify and calibrate hybrid compartmental, agent-based models with discrete event simulation systems and share interactive versions of these models with the general public to share important information about quarantine measures. Programs like Microsoft Excel cannot compete with the modelling capabilities that AnyLogic has. Specifically in the discrete-event simulation sector.

Discrete-event simulations (DES) models a system as a sequence of occurrences that happen at distinct times (Gallagher, 2017). When an event occurs, the system changes, and between events the system is unchanged (Gallagher, 2017). Systems like hospital resource analysis are perfect for using DES, and AnyLogic is the best suited program to build these models in. Although no DES models were built for this project, the fact that these models can be constructed in AnyLogic made them the undisputable choice for the models developed in this project.

V. Methodology

Teamwork and Project Development Process

This IQP project was completed in collaboration with the Financial University of Moscow (the Department of International Finance and the Department of Data Analysis, Decision-Making and Financial Technologies) with participation from two of its professors and four of its students. The original plan was to complete the project in Moscow, but due to quarantine protocol, it had to be completed domestically in an interactive and collaborative online mode. Both WPI and the Financial University instituted remote learning protocols shortly before the project started, so each team member was working on it from their home, dormitory, or apartment. The project benefited from the expertise of The Financial University's Professor Korovin and Professor Losev on pandemic modeling, as well as from the insight of the Financial University students on developments of the COVID-19 pandemic in Russia.

Team Communications and Joint Work Organization

Due to working on the project in a remote environment, it was vital to organize steady communication and establish a timeline for deliverables. The internal WPI team decided on meeting daily via Zoom video conference and communicating short correspondences via text message. Working with Professor Nikitina, it was determined that all Moscow project teams would meet weekly via Zoom video conference and update on project status and trouble-shoot any concerns. Similarly, for meetings with Moscow partners and professors, it was established that everyone would Zoom video conference weekly and develop plans for the upcoming week. The fourth type of meetings were with the internal WPI team and the Financial University students. Since the three partner students were enrolled in classes full time during this project, the group decided to meet on a will call basis via Zoom and communicate consistently through Telegram. A chat with the internal WPI team and all Russian partners and professors was also set up in Telegram (a messaging app). Harnessing consistent communication helped keep all project participants focused on short term tasks, while keeping the ultimate project goal in view.

Developing a Thorough Understanding of COVID-19

In order to begin to model the spread of COVID-19, and create recommendations for preventing spread, our team needed to first understand the nature of the SARS-CoV-2 virus and learn why it caused such a widespread pandemic so quickly. Background research on spread factors and general information on the virus was conducted, followed by a summary of some of its impacts on the planet. This information was then added to the background section and used as

a basis for developing recommendations on response protocols.

Comparison of Response Protocols and Virus Impact across Target Countries

After researching the protocol for preventing COVID-19 spread, as recommended by the World Health Organization, the protocols of the United States, Germany, Italy, Russia, Democratic Republic of Korea, and China were determined and compared. Graphs of infection numbers to date were also gathered for some of the countries and included in the background section. It was important to understand how each country's reactive measures to COVID-19 were influencing the trends for the number of individuals infected and number of individuals dead. The most current versions of the data in the background section were used to inform the recommendations and models presented as final deliverables.

Tracking Historical COVID-19 Data (Case and Death Numbers) for the Six Countries

After doing preliminary research on the virus's impact on China, the US, Russia, South Korea, Germany, and Italy, spreadsheets were created for each country to track the historical trends of the virus spread. Measurements such as total infected, total dead, total recovered, and total active cases were all tallied for each day since the initial case in each country. With the exception of Olga Skiba, each student collected historical data for one of the six countries. The goal in gathering such data was to compare the countries' successes at preventing virus spread and then relate the success to response protocols that would inform recommendations. The infection and death totals in the spreadsheets were used both to calibrate the compartmental models and to assess how well different countries' contained the pandemic.

Demographic Data and Response Protocol Analysis for the Six Countries

In addition to tracking COVID-19 numbers, demographic information was also gathered for the six countries and compared in a table. Factors attributed to virus spread and mortality rate such as percentage of population over the age of sixty and population density were considered. Additionally, the response protocols like travel bans, quarantine guidelines, and testing were added to this table and contrasted between the six countries. Compiling all these factors in one location allowed for correlations to be drawn between infection/death totals and the potential contributors. These correlations would eventually inform recommendations as well the models created to emulate the recommendations.

Introduction to Different Excel Models

During weekly international meetings, Professor Korovin from the Financial University demonstrated multiple Excel models for predicting the impact of a virus on a population. The

models were calibrated by determining specific coefficients that influenced the infection rate and virus spread. One common model that was demonstrated was the HIRD model that breaks a population into healthy, infected, recovered, and dead from infection. While this model was not ultimately used in the predictive dynamic model, it was instrumental in developing an understanding of basic virus spread in a population. Additionally, the Excel HIRD model demonstrated the necessity to calibrate coefficients that influence the rate at which individuals move from one category to another in the model.

Exploring the AnyLogic Modeling Application

AnyLogic for students (Personal Learning Edition) was downloaded on all team members' computers and multiple modeling methods were explored. The main two models of interest were system dynamic models and agent-based models. Fortunately, the application came installed with example models which contained actual data from different healthcare related models. Influenza, for instance, was an existing agent-based model that helped the team understand how introducing different agents of spread influenced the ultimate outcome. The SIR model for virus spread, which contained both agent-based and system dynamics components, was also integral in planning the final model.

After exploring the application, the *AnyLogic in Three Days* textbook was studied for in-depth information on setting up complex dynamic models. It was important to understand how to calibrate a dynamic model using existing data to create projections for that same data set. The textbook also helped in demonstrating additional cohorts to add to the SIR model like quarantined and isolated populations. These groups of a population were of the utmost relevance to this project as the ultimate goal was to demonstrate the importance of stringent quarantine protocols in efforts to contain a contagious virus such as SARS-CoV-2.

Models and Documentation Process

Once a basic understanding of AnyLogic was developed, it was most efficient for all project participants to brainstorm different types of models that could be beneficial to the general population. Having been exposed to a HIRD (healthy, infected, recovered, dead), SIR (susceptible, infected, recovered), SEIR (susceptible, exposed, infected, recovered), and an SEQJIR (mentioned in background) scaffolding, the next step was to create these models or hybrids of them. As new models were created, each iteration was recorded with documentation of the aspects that needed improvement and the areas that seemed to supply meaningful information. The empirical data, concerning numbers of infections and deaths for the six countries, was used to calibrate multiple system dynamic models in AnyLogic in an effort to make these models predictive. Additionally, the team used the empirical data to explore new models in excel like econometric and sigmoid models to determine how different parameters influenced the spread of COVID-19.

Beyond creating individual models for the six countries of interest, some general agent-based models were developed with interactive features. Adjustable values like infectivity and percentage of population in quarantine were represented with sliders where a user could drag to change the value and then run the model to see the impact on numbers of infected and dead individuals. All models made within AnyLogic generated a graph to portray the number of infected, recovered, and dead individuals throughout the allotted time period. These time periods were selected to include an entire cycle of the virus infection or until the new infection cases approached zero per day.

Documenting multiple iterations of created models was instrumental in understanding where improvements had to be made, but also to provide a baseline to future groups for developing more complex models. The keeping of consistent records served to provide a better flow of connections between models for readers of the report to follow.

Use of Models and Other Data to Inform Recommendations

After compiling documentation of models, tables of response protocols, and general information for the six countries, deductions were made for the overall significance of the data. Countries that had shown success in preventing virus spread were analyzed for their unique response to the pandemic in hopes that their techniques could be replicated by other countries. Response protocols instrumental in those countries' fight against the virus were condensed into recommendations for governments for future pandemic management. The completed models also helped confirm the recommendations by emulating the importance of certain response protocols during a pandemic.

Compiling of Recommendations and Relevant Models onto a Website

Once recommendations were deduced from empirical COVID-19 data and the methods countries' instituted to mitigate virus spread, they were organized and added to a website. This website was created using Wix and summarized response protocols for governments to follow in the future if such a pandemic was to happen again. In addition to the recommendations, models developed during this project were included on their own page of the website to demonstrate the significance of following certain response protocols like quarantining. Beyond the technical aspects of the project, the international value of the project was demonstrated on the contributions page of the website. Headshots and short biographies on all team members, including professors, were included on the contributions page. This website was assembled by both the WPI and Financial University students and was linked via url to the WPI Global Lab's website already established with its own domain.

VI. Results and Analyses

Correlation Between Response Protocols and Success in Preventing COVID-19 Spread

Once all data was gathered and tabulated, the six countries' successes in preventing virus spread could be compared. While no definitive conclusions could be generated considering the breadth of variables impacting virus spread, certain correlations between response protocols and preventing infection spread were observed. The goal in comparing successful measures in preventing virus spread was to offer recommendations that could help in managing future pandemics. With these recommendations summarized on the team's website, persons visiting the site could understand the significance and effectiveness of adhering to certain protocols.

First, the Republic of Korea, China, and Russia had significantly lower infections per populus 80 days after the first infection than the United States, Germany, and Italy (Figure 9). One primary difference between these subsets of countries is the extent in which the government enforced quarantine measures and the uniformity in enforcing such measures. In Russia, China, and South Korea, going out without an essential purpose was punishable by law, and this order was instituted nationally. By contrast, quarantine protocols were left up to the discretion of the states in the U.S. and many reacted later than they should have. There was also ambiguity in the extent that these measures were to be followed and fines for not adhering to them were instituted only in a few states. Italy and Germany also began enforcing quarantine measures too late, with Italy doing so nationally 37 days after the first domestic infection and Germany reacting similarly to the U.S. and only instituting protocol on a region by region basis at staggered times.

Another trend that became apparent when the countries were compared was the connection between prompt, widespread testing and low death per capita.. Germany, Russia, and the Republic of Korea all instituted mass testing early on and were able to keep their death totals low compared to many other large nations. While the number of tests conducted per capita slowly leveled between the six countries as the pandemic evolved (Figure 9), the difference in tests administered per capita were significant earlier on. For instance, the United States and the Republic of Korea (who saw the first domestic COVID-19 cases one day apart) had drastically different numbers of tests conducted as of March 30th—5/million versus 4000/million respectively.

The third factor that seemed to contribute significantly to death and infection totals was a country's ability to institute early contact tracing. The Republic of Korea, which saw great success in preventing virus spread, began practicing contact tracing through a number of methods following its first domestic COVID-19 cases. While their infection totals seemed comparatively high early on, this was primarily due to their ability to track individuals who had

tested positive for COVID-19 and quickly identify the people they had been in contact with. With the pandemic having evolved significantly since then, it is now clear that the Republic of Korea was able to keep their infections per capita and their deaths per capita comparatively low. China also apparently instituted similar contact tracing protocols, but there is less information available on the methods they used.

To summarize, the three biggest contributors to preventing virus spread, according to our research, seemed to be strict quarantine guidelines instituted uniformly on a national scale, prompt widespread testing, and contact tracing. Pandemics, by nature, involve viruses that spread at exponential rates if not contained, so instituting a national response one week after the first infection compared to three weeks can have a significant impact on total cases and deaths. For instance, the United States saw about 300 cases of COVID-19 in the first six weeks and then in the following two weeks saw more than 30,000 new cases. If the virus had continued to spread at this rate, the United States would have seen more than 1,000,000 cases by day 80 compared to the 505,000 recorded by that day. The perfect formula of response protocols can only be guessed at, but it is certain that all six countries would have seen far higher death and infection totals had they not responded at all.

Country	U.S.A.	Russia	R. Korea	Germany	China	Italy
Population in Millions	331	146	52	84	1,439	60
Percentage of Population over 60 Years	20.8	21.7	19.4	28.0	17.3	22.8
Average Population Density (persons/km ²)	35.8	8.4	503.0	227.0	148.0	206.0
Date of First Infection (m/d/yr)	1/21/20	1/31/20	1/20/20	1/28/20	11/17/20	1/31/20
Tests per Million Individuals (as of 4/17)	10,333	11,773	10,659	20,629	Data not available	19,490
Deaths per Million Population 80 Days After First Infection	66	3	4	51	<1	402
Infections per Million Population 80 Days After First Infection	1,528	323	200	1,683	19	3020

Figure 9: Table comparing various data pertaining to COVID-19 for the six countries of interest

Correlation Between Age Demographics and Deaths per Capita

Comparing the six countries' age demographics in Figure 9 provided minimal evidence pointing towards a direct correlation to deaths per capita. Italy, which had a significantly higher death count per capita than all the other countries, had the second highest percentage of individuals over 60 years of age, but this was not enough information to be definitive about the assertion. Likewise, China and the Republic of Korea, which had two of the three lowest deaths per capita, had the two lowest percentages of population over the age of 60. This seems to point at a connection between these two data points, but it is more likely that response protocols played a more significant role in death and infection totals. If no response protocols were instituted by all countries, then a more definitive trend between deaths per capita and percentage of population of 60 years could most likely have been traced. There are so many variables that contribute to these infection and death totals, making it impossible to isolate a sole cause or even a contributing cause. One theory about Italy's high rate of deaths per capita is the amount of elderly people living with younger family members (Italy, pandemic's new epicenter, has lessons for the world.2020), but the data collected for this project did not look into this factor.

Correlation Between Population Density and Infections per Capita

Another factor that was looked into was the correlation between population density and infections per populus in all six countries. While the data in Figure 8 shows no pattern pertaining to population density and infections, that is most likely because those density numbers are averages across the entire nation. For a country like Russia, which has vast areas that are unpopulated, it is not reasonable to use it as a gauge for how quickly viruses will spread. As is seen in almost all major countries, densely populated cities are the areas where the highest number of cases are seen per populus. To conclude, a country's population density on the whole would not be a good gauge of how fast viruses would spread in that country, but the number of densely populated cities in that country would provide insight to how severely pandemics might impact that country.

Models

During this project, many different models were created in an effort to simulate COVID-19 spread and pandemics. At first, the goal was to forecast the infections and deaths for the six countries of interest, but this objective slowly shifted towards modeling more nuanced situations that emphasized the importance of certain response protocols. AnyLogic modeling software was used to create the majority of models for this project. A more extensive portfolio of model iterations is included in Appendix A, while this section serves as a summary of results.

After learning about epidemic model fundamentals from Professor Korovin in Excel, the team compiled historical COVID-19 data for the six countries of interest. This data included total deaths, active cases, total infections, and total recovered from the first domestic infection to the current day. This data was put aside momentarily while models were assembled in AnyLogic with hopes that they would be able to be calibrated with the previously mentioned data. Calibration essentially involves imputing the raw data to date for a country, and informing the model so that it can predict beyond that date.

System Dynamic Models

The first set of models created were primarily system dynamic models which take an imputed population and separate cohorts of the population into compartmental sections that change in number as the virus runs its course. The models are dynamic because the flow rates at which people travel from one compartment to another change depending on the instantaneous number of individuals in a compartment. The goal for the system dynamic models was to calibrate them with the historical data for all six countries and enable them to predict future infection, death, and recovered totals specific to each country.

An SEQJIR model (Figure 14) was the first model scaffolding created in AnyLogic (a model scaffold is a skeleton version of a model that has not yet been calibrated with data). While this model was descriptive, it was ultimately too complex with six compartments—not including one for dead individuals—and too many coefficients that needed to be calculated. The following dynamic models created were a SIR model, a SEIR model, and a HIRD model (Figures 13-18). The former two did not contain a compartment for dead individuals, so the majority of the effort went into optimizing a HIRD (Healthy-Infected-Recovered-Dead) model.

The HIRD model had originally been created in Excel, but with generic data that did not represent a specific population of people. Additionally, the Excel model lacked the ability to adapt dynamically over time, so AnyLogic was necessary to accomplish the task. After the model scaffolding was set up, the raw historical data for one of the six countries (take the United States for instance) was inputted into AnyLogic in an effort to calibrate the model and make it predictive. At the time of calibration, about 85 days had passed since the first domestic infection. Sixty rows of data were inputted into AnyLogic so that upon forecasting past 60 days, the team could determine if the model was calibrated accurately. Unfortunately, this proved to be a dead end for the system dynamic models, as the calibration consistently forecasted the number of infected individuals increasing until virtually the entire population had become infected. Despite introducing quarantine coefficients into the model, the number of active cases never plateaued at a reasonable value. The inability to accurately model the lifecycle of the pandemic likely stemmed from the formulas that are innate to compartmental models. Since the number of individuals moving from the healthy compartment to the infected compartment is a product of the number of healthy individuals, the formula will always yield an increasing number of infected individuals until the healthy population is almost depleted.

After sharing the HIRD model results with Professor Korovin, he agreed that the model lacked the ability to be accurately calibrated, given its fundamental formulas. With more modeling expertise, the team could have potentially delved into the system dynamic models and reworked them, but this was deemed outside of the realistic scope for the project. Having exhausted the system dynamic models, the team moved onto agent-based models in efforts to model scaled down pandemic situations.

Agent-Based Models

Agent-based models in AnyLogic, unlike system dynamic models, follow individual agents as opposed to an entire population. In regards to a pandemic model, system dynamic models can only display the percentage of a population that is infected, while an agent-based model can display the specific individuals that are infected and determine how they became infected. Agent-based models, thus, are better suited for smaller populations as is evident in the town model that the team created.

The town model (seen in Figures 20-23) follows five-hundred agents as they navigate from their households to multiple grocery and convenience stores and graphs the instantaneous number of individuals who are infected, susceptible, exposed, and recovered. When the model is being run, a viewer can observe the agents interacting with one another in households and grocery stores as well as adjust the contact rate for the agents. Additionally, a quarantine event where the contact rate immediately drops to 5 contacts per day for the agents can be instituted at various times to demonstrate the influence of quarantining at different stages in a pandemic.

Given the assumption that agents go to grocery stores and convenience stores once every five days, have 60 contacts with other agents on a normal day, and develop immunity after an average illness duration of 15 days, multiple quarantine timelines were tested on the town model. When town-wide quarantine was instituted 21 days after the first infections, the peak number of active infections was 150 and about 400 agents were infected during the course of the entire pandemic (Figure 23). Comparatively, when town-wide quarantine was instituted 7 days after the first infections, the peak number of active cases was 26 and only about 80 individuals experienced infection (Figure 22). While some of the assumptions made for this simulation are not generated from real data, it is safe to discern that the earlier a population quarantines, the lower its total infections will be.

One exciting aspect of the town model is its ability to be adjusted as more data becomes available. If a community was able to gauge the average number of contacts its inhabitants had per day, better pandemic simulations could be conducted. While the town model has its limitations, like the absence of a category for dead individuals, there is much potential for improvement. One area that was looked into briefly is the introduction of hospitals with capacities that could be set for specific towns. With the introduction of this variable, local governments could prepare for the strain that pandemics might put on their hospitals.

Econometric and Sigmoid Excel Models

Alongside the AnyLogic models, were a number of excel models created by the Financial University students. These models primarily focused on analyzing different parameters that may have influenced COVID-19 infections in the six target countries and some even predicted trends for individual countries.

Before delving into an analysis of the econometric and sigmoid models, distribution tests were done to assess how empirical COVID-19 infection data compared to logistic and exponential trends. As is evident in Figures 27 and 28, the United States and Russia followed an exponential trend in the number of infections at first, but eventually the exponential trend grew to infection numbers far larger than would actually occur in a pandemic. While the logistic trend was not realistic for United States data, Figure 26 showed a possibility that Russian infection totals could begin to follow it in the future.

The first predictive model completed in Excel was a moving average prediction for the six target countries and demonstrated that countries following an upwards trend in new cases, like Russia and the United States, would continue to follow this trend in the near future, while countries still seeing new cases, but at a declining rate, would continue to follow their downward trend.

One especially interesting model was an econometric model that showed the target countries' infection trends before and after quarantine. While all six countries did institute a quarantine protocol of some kind, varying levels of successes were seen amongst the group. As can be seen in Figure 31, The Republic of Korea and China were most successful in leveling off total infections after quarantine protocols were implemented. China's strict lock-down policy and The Republic of Korea's targeted quarantine procedure both proved to be effective at mitigating spread. The United States and Italy, which both lacked uniformity and timeliness in their quarantine response, showed little success at preventing the spread of infections.

Another intriguing econometric model tested the influence that eight parameters had on the number of infections per country (Figures 33 and 34). The two factors that proved to have the highest correlation with the number of infections were the number of tests conducted per capita and population density. While population density is completely out of short term control, conducting widespread testing is an effective way to mitigate spread, according to this model. Although a country's ability to test large numbers of people depends on its wealth and medical infrastructure, it is still important to note this finding in an effort to inform recommendations.

In addition to the development of econometric models, was the exploration of sigmoid models to predict the number of infections by country. The sigmoid models used complex mathematical equations to find trends in countries' infection totals and used the trends to inform projections. While this type of model was only explored briefly, it did produce some convincing

forecasts for Germany, as is evident in Figure 37. The sigmoid models would certainly be an interesting area of study for examination in the future.

Website

Having produced a small portfolio of working models as well as a number of recommendations for response protocols during a pandemic, the team developed a website to share the accomplishments of the project. The website was created using Wix Website Developer and constructed to emphasize focal points of the project report. The website contains four pages excluding the home screen: a portfolio of models, a recommendations page, a contributions page, and a country comparison page.

The goal of the website is to share information with the public in a manner that is graspable, concise, and convincing. A lengthy project report could potentially deter people from accessing the important recommendations that were developed in this project and leave the work virtually archived. A visually aesthetic website on the other hand provides an easy platform for people to explore the IQP research and learn about effective ways to combat pandemics like COVID-19. The graphs displayed with a handful of the models emphasize the importance of quarantining and responding to a pandemic in a prompt manner. Additionally, the agent-based town model page on the website contains a link to the AnyLogic cloud where the model can be interacted with by any website visitors.

The website URL was linked to a Wordpress page already in existence, that was created by the WPI Global Labs Department. Their web page contains short synopses of the other remote IQP projects that involved various COVID-19 research and is an easy way to access the website. Visitors who navigate to the Global Labs web page can see the project abstract and click on the icon to access the breadth of information on the website.

VII. Conclusion

Pandemics are crises that span borders and affect populations all over the planet. Due to their evolving and fast moving nature, they are extremely difficult to manage and contain in an effective manner. During the COVID-19 pandemic, countries around the globe instituted a variety of protocols—some shared between nations, and others unique to specific countries—in an effort to combat the coronavirus and limit infection spread. Wildly varying levels of success were seen from country to country and this project was primarily interested in analyzing which response protocols worked best so that recommendations could be offered to anyone dealing with future pandemics.

After monitoring the infection and death numbers for six countries—the United States, Russia, Republic of Korea, Germany, China, and Italy—the team looked into the various response protocols they implemented as well as demographic differences between the nations. Due to the number of variables contributing to a country’s success in preventing virus spread, it was difficult to isolate specific factors that had the largest impact. Rather than definitively stating which methods worked and which did not, the team observed correlations and developed recommendations from collected data. Some factors contributing to virus spread and mortality, like population density and percentage of a population that is elderly, are virtually out of countries’ control. Conversely, response protocols like quarantine measures and social distancing guidelines are achievable in all countries. Somewhere in the middle, are procedures like contact tracing and widespread testing that may or may not be possible depending on a country’s government and medical capabilities.

Using the data gathered from the six target countries, the team developed models in AnyLogic and Excel to forecast COVID-19 infection totals and to demonstrate how certain response protocols helped mitigate infection spread. The original goal was to use system dynamic models to forecast the total infections and deaths each country would see once the pandemic had run its course, but this slowly fell outside of the realistic scope for this project. The team was not able to calibrate the compartmental models in AnyLogic as they continuously projected nearly the entire population of a country becoming infected. While the econometric and sigmoid models were able to produce some infection forecasts, they were based on too many assumptions to warrant confidence in the predictions. At this point, efforts switched to modeling smaller populations using agent-based models. These models required no calibration and were able to emulate pandemic protocols such as quarantine measures in a 500 person town. The advantage of the AnyLogic agent-based model was the interactive nature and easy visualization of virus spread in a small community. Harnessing the demonstrative power of user-friendly models allowed the team to best support the recommendations that were developed for pandemic response.

Recommendations

After identifying successful response protocols in countries, as well as in the AnyLogic and Excel models created, a number of pandemic containment recommendations were developed by the team. The first set of recommendations align with and reinforce what the World Health Organization suggested from the start. To begin, quarantining at home is an effective method to mitigate the spread of viruses like SARS-CoV-2. In the town model, (Figures 20-23) it was shown that as fewer people visit high risk areas like grocery stores, virus spread slows in a community. The same goes for the frequency with which persons visit high risk areas. Limiting trips to congested locations is thus an effective way to combat pandemic spread.

The next few recommendations, which are more difficult to implement nationally, are widespread testing and contact tracing. The Republic of Korea saw great success in combating the COVID-19 pandemic with these two methods and they could both ostensibly be carried out in all major first-world countries. While there is an argument for larger countries' (like the United States) inability to contact trace and test on a national scale, this concern would be partially addressed if the response protocols were instituted promptly after the first domestic infection—which leads to final and most significant recommendation of prompt and uniform action.

Responding to pandemics in a timely and uniform manner, greatly limits the scale of infections and deaths in a country. Having seen what exponential spread looks like, the team recommends that countries prepare a plan for future pandemics—using the previously mentioned guidelines—and implement that plan promptly and uniformly throughout the nation. When different states or territories follow different response protocols, confusion arises (as was seen in the United States) and success combating the pandemic decreases. Although the system dynamic models created were not able to be calibrated accurately for the six countries, they were run using influenza data and showed continuous exponential growth when no response protocols were applied. Only when a control like quarantining was applied to the population, was the curve of infections able to plateau without showing nearly the entire population becoming infected. The earlier the quarantine protocol was instituted in the model, the lower the number of total infections became. This points to the suggestion that whatever it is a country chooses to do to combat a pandemic, instituting the plan sooner rather than later will always yield better results.

Future Projects

As is true with all scientific progress, teams learn and build from work that was done before them—creating better iterations of the same thing, or developing a new thing using existing knowledge. This project, in its successes and remaining frontiers, provides a platform for future groups to make new models and extend their understanding of pandemics and how to combat them. One place to start could be a revisiting of the system dynamic models like the

HIRD compartmental model. With much more COVID-19 data to work with, future research groups could use the completed infection curves to calibrate the models and then begin adding different controls such as testing and quarantining. Another area of expansion could stem from the introduction of the agent-based town model. Using the basic scaffolding of the model, future groups could input new data to emulate other towns and cities and learn how effective prompt quarantining would be in that particular region. Additionally, the beginning of the hospital system in the town model is an area where much future modeling work could be done. If legitimate hospital data on staff and bed capacity was available, future teams could research how different infection numbers in a community influenced the strain on hospitals and their resources.

Summary

The goal of this project was to observe how six major countries' dealt with the COVID-19 pandemic and determine which response protocols mitigated virus spread most effectively. Based on those findings and the developed models which emphasized the significance of certain pandemic protocols, recommendations were compiled for countries to follow during pandemics. While predictive modeling success was modest, the team was able to develop smaller scale models that demonstrated the significance of prompt quarantining. After assessing the infection and death history for six countries, the team confirmed that contact tracing and widespread testing are key to combating a pandemic. Additionally, the promptness with which governments begin instituting uniform response protocols, impacts the entire course of a pandemic in a country. The team worked collaboratively to present its models and recommendations on a website accessible from the WPI Global Lab's webpage.

VIII. Bibliography

- Aljazeera. (2020). How does coronavirus spread and how can you protect yourself? Retrieved from <https://www.aljazeera.com/news/2020/01/coronavirus-spread-protect-200130115539072.html>
- Beaubien, J. (2020). How south korea reined in the outbreak without shutting everything down. Retrieved from <https://www.npr.org/sections/goatsandsoda/2020/03/26/821688981/how-south-korea-reigned-in-the-outbreak-without-shutting-everything-down>
- Brauer, F. (2008). Compartmental models in epidemiology. (pp. 19-79)
doi:10.1007/978-3-540-78911-6_2
- CDC. (2020). What you need to know about the coronavirus disease 2019 (COVID 19). Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/downloads/2019-ncov-factsheet.pdf>
- CIA. (2020a). The world factbook: Europe:Italy. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/geos/it.html>
- CIA. (2020b). The world factbook: South korea. Retrieved from <https://www.cia.gov/library/publications/the-world-factbook/geos/ks.html>
- Country-level coordination, planning, and monitoring. Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/country-readiness>
- Department of Homeland Security. (2020). List of essential industries. Retrieved from <https://www.cbia.com/resources/coronavirus/coronavirus-state-federal-updates/department-homeland-security-essential-industries/>
- Early investigations protocols. (2020). Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/early-investigations>

- First Channel. (2020). *В москве ситуация с коронавирусом остается самой сложной в стране.* (). WWW: Первый Канал (First Channel). Retrieved from https://www.1tv.ru/news/2020-04-12/383752-v_moskve_situatsiya_s_koronavirusom_ostaetsya_samoy_slozhnoy_v_strane
- Gallagher, S. (2017). Comparing compartment and agent-based models. Retrieved from http://www.stat.cmu.edu/~sgallagh/papers/gallagher_8-17.pdf
- Germany COVID-19 cases and deaths statistics. (2020). Retrieved from <http://virusncov.com/COVID-statistics/germany>
- Italian National Institute of Health. (2020). *Epedemia COVID-19.* Retrieved from https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_26-marzo%202020.pdf
- Italy, pandemic's new epicenter, has lessons for the world. (2020). Retrieved from <https://www.nytimes.com/2020/03/21/world/europe/italy-coronavirus-center-lessons.html>
- Koptyug, E. (2020). Coronavirus (COVID-19) in germany. Retrieved from <https://www.statista.com/topics/6176/coronavirus-COVID-19-in-germany/>
- Lee, B. (2020). Study: 17.9% of people with COVID-19 coronavirus had no symptoms. Retrieved from <https://www.forbes.com/sites/brucelee/2020/03/18/what-percentage-have-COVID-19-coronavirus-but-do-not-know-it/#339deb8c7e90>
- Lia, K., Chen, Y., Ruzheng, L., & Kunyuan, H. (2020). Clinical features of COVID-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection*, , 1-5.
- Loh, T., & Kresge, N. (2020, Apr 2,). Private labs helped germany test 1 million for COVID-19 virus. *Bloomberg.Com* Retrieved from <https://www.bloomberg.com/news/articles/2020-04-02/private-labs-helped-germany-test-1-million-for-COVID-19-virus>
- Maintaining essential health services and systems. (2020). Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/maintaining-essential-health-services-and-systems>

- Ministry of Health and Wellness of Korea. (2020). Coronavirus domestic occurrence status. Retrieved from http://ncov.mohw.go.kr/bdBoardList_Real.do?brdId=1&brdGubun=11&ncvContSeq=&contSeq=&board_id=&gubun=
- Population - new york city population. (2015). Retrieved from <https://www1.nyc.gov/site/planning/planning-level/nyc-population/population-facts.page>
- Rogers, L. (2016). America's age profile told through population pyramids.
- Russia: The number of tests performed. (2020). Retrieved from <https://ourworldindata.org/COVID-testing>
- Stanford University. (2020). How to identify factors affecting COVID-19 transmission. *Science Daily*, Retrieved from www.sciencedaily.com/releases/2020/03/200326160759.htm
- Thomala, L. L. (2020). Novel coronavirus COVID-19 in china. Retrieved from <https://www.statista.com/topics/5898/novel-coronavirus-COVID-19-in-china/>
- Thomke, S. (2020, March 30,). Why is the U.S. behind on coronavirus testing? Retrieved from <https://hbr.org/2020/03/why-is-the-u-s-behind-on-coronavirus-testing>
- Times, T. M. (2020, April 14,). Coronavirus in russia: The latest news. *The Moscow Times* Retrieved from <https://www.themoscowtimes.com/2020/04/14/coronavirus-in-russia-the-latest-news-april-14-a69117>
- usa.gov. (2020). Government response to coronavirus, COVID-19. Retrieved from <https://www.usa.gov/coronavirus>
- WHO. (2020). Coronavirus. Retrieved from <https://www.who.int/westernpacific/health-topics/coronavirus>
- Wigginton, K., & Boehm, A. (2020). Environmental engineers and scientists have important roles to play in stemming outbreaks and pandemics caused by enveloped viruses. *ACS Publications*,

World Health Organization. (2020a). Coronavirus disease (COVID-19) situation dashboard. Retrieved from <https://experience.arcgis.com/experience/685d0ace521648f8a5beeee1b9125cd>

World Health Organization. (2020b). Coronavirus disease 2019 (COVID-19) situation report - 71. Retrieved from <https://www.cdc.gov/coronavirus/2019-ncov/downloads/2019-ncov-factsheet.pdf>

World Health Organization. (2020c). Global research on coronavirus disease. Retrieved from <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov>

Worldometer. (2020). Age, sex, existing conditions of COVID-19 cases and deaths. Retrieved from <https://www.worldometers.info/coronavirus/coronavirus-age-sex-demographics/>

Количество стационарных койко-мест в больницах. (2019). Retrieved from <https://data.mos.ru/opendata/62045>

IX. Appendices

Appendix A: Documentation of Models

Many models were created in the course of project development. While the majority of them produced inconclusive results, or were impossible to calibrate altogether, each provided an opportunity to learn and improve future models. Each model, including all iterations of individual models, were documented and recorded in the Appendices. Since AnyLogic was a new application for all team members it was difficult to produce cutting edge models that others had not created before. With that being said, this Appendix could serve as a starting place for future groups interested in creating epidemiological models.

The Significance of Using Dynamic Models and AnyLogic

The main difference between dynamic and static models is captured in their names. Static models retain unchanging formulas which yield projections informed from initial conditions. Dynamic models, on the other hand, possess the ability to adjust the conditions and coefficients that are informing the projection, to develop a more accurate picture of the real world. Since epidemiological models need to first be calibrated with real world data, it is vital that the curve they project beyond the historical data can be influenced by new factors. If the projection is simply allowed to continue on the same trend as the historical data, the results will never emulate the real world. As is seen in Figure 10, where unchanging mathematical equations (polynomial, exponential etc.) are graphed against the United States' historical infection data (blue dots), the correlations are weak—their actual correlation errors can be observed in Figure 11. If even the most accurate of the functions was to be continued for a year, it would project more infections than there are people on the planet. Due to this discrepancy, it is necessary to have the ability to introduce new variables like percentages of the population that self-isolate. Only once dynamic factors are introduced, will infection and death curves level off before nearly the entire population becomes infected.

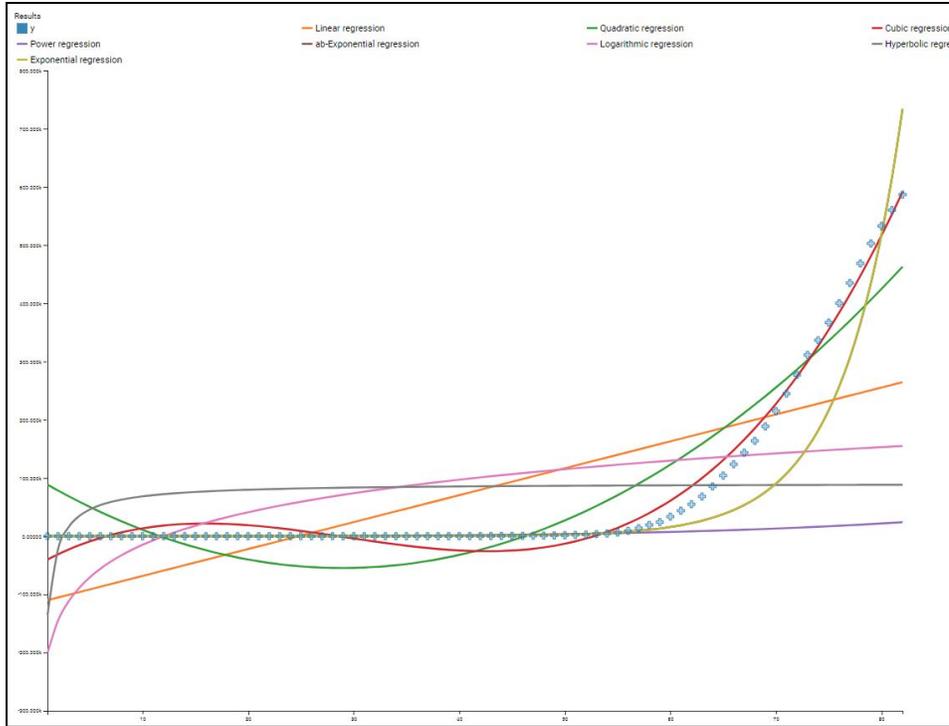


Figure 10: Common static equations mapped to correlate with the United States infection curve.

Linear regression $y = 4625.2014x + -114607.0777$		
Linear correlation coefficient 0.7135	Coefficient of determination 0.5091	Average relative error, % 461452.9488 %
Quadratic regression $y = 183.7477x^2 + -10625.8612x + 98907.7984$		
Correlation coefficient 0.9323	Coefficient of determination 0.8691	Average relative error, % 329121.0919 %
Cubic regression $y = 5.0223x^3 + -441.5281x^2 + 10259.3570x + -49907.8562$		
Correlation coefficient 0.9926	Coefficient of determination 0.9852	Average relative error, % 128883.1785 %
Power regression $y = 0.0013x^{3.7976}$		
Correlation coefficient 	Coefficient of determination 	Average relative error, % 562.0218 %
ab-Exponential regression $y = 0.4399 \cdot 1.1909^x$		
Correlation coefficient 0.9317	Coefficient of determination 0.8681	Average relative error, % 106.2367 %
Logarithmic regression $y = -200727.7396 + 80720.4593 \ln x$		
Correlation coefficient 0.4802	Coefficient of determination 0.2306	Average relative error, % 572196.0425 %
Hyperbolic regression $y = 91110.2507 + \frac{-226303.8108}{x}$		
Correlation coefficient 0.1878	Coefficient of determination 0.0353	Average relative error, % 446974.7691 %
Exponential regression $y = e^{-0.8211 + 0.1747x}$		
Correlation coefficient 0.9317	Coefficient of determination 0.8681	Average relative error, % 106.2367 %

Figure 11: Common static functions' correlation error to historical data points for United States infections.

While the team was not able to produce confident projections for death or infection curves, Figure 12 shows a death projection for the United States that was generated from dynamic modeling. This projection was created by the Institute for Health Metrics and Evaluation (IHME) and demonstrates how introducing social distancing as a contributing variable can influence when a death curve plateaus. This type of projection was initially the goal for the six countries researched, but was deemed too complex for the limited knowledge on modeling that the team possessed.

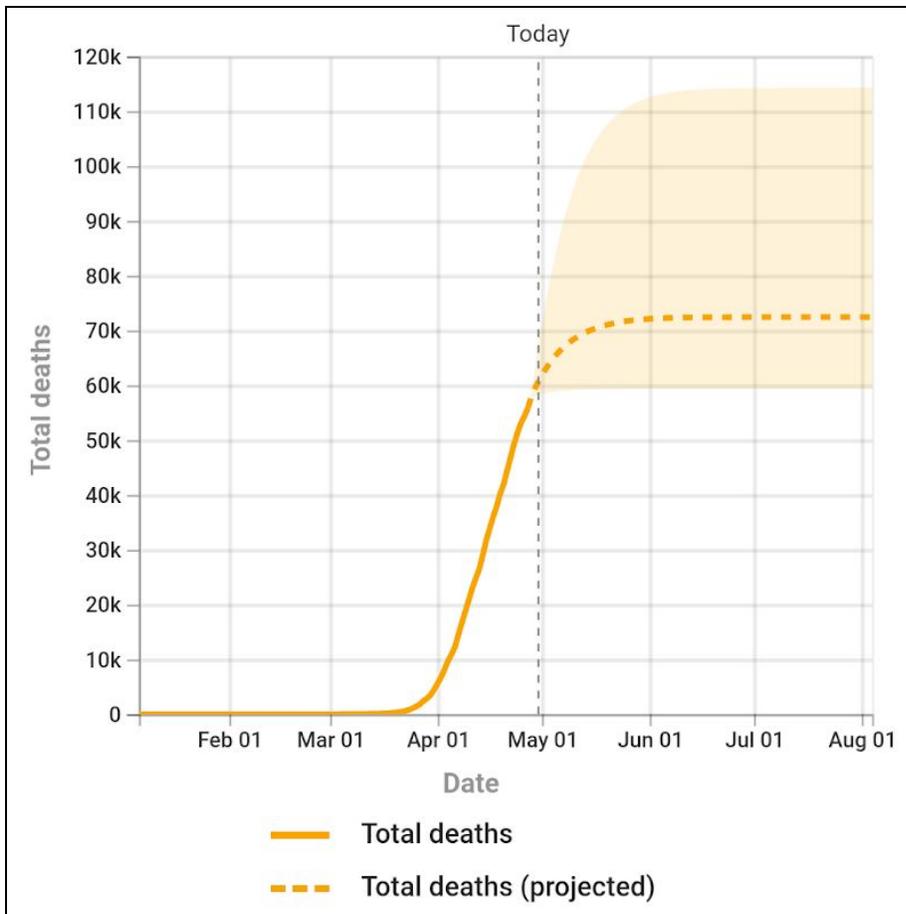


Figure 12: IHME dynamic modeling projection for total deaths in the United States.

Since it was decided that dynamic models were going to be instrumental in processing and displaying the evolving pandemic, the team needed to use an application that was specifically designed to host such models. AnyLogic is an effective application for developing visually informative models with little coding experience—perfect for a team with no computer science majors. Excel is great at graphing, but primarily with historical data or static extensions of that data. AnyLogic provides the ability to add dynamic variables and produce visually

engaging representations of a population. This application was recommended by Professor Losev and Professor Korovin and served its purpose well on this project.

Stock SIR Model

The first AnyLogic model worked with was a SIR model, which is a dynamic system model. This model (Figure 13) contains three different compartments which each represents a category an individual can be placed in during a pandemic—susceptible, infected, recovered. At any given time while this model is being run, the sum of the individuals in each compartment always adds up to the initial population input into the model. The smaller arrows pointing to the larger flow arrows all represent different variables that contribute to the rate at which individuals move from one compartment to another (Figure 13). In this particular model, population size, infectivity, and contact rate all impact infection rate while only the illness duration influences recovery rate. The latter is due to the fact that, in this model, all individuals recover from the virus and develop immunity, which moves them to the recovered compartment indefinitely. The graph displayed above the model in Figure 13 shows the percentage of individuals that fall into each compartment over an 85 day period—although that data is just randomly generated for the sake of showing what an infection graph could potentially look like.

The striking shortcoming of this SIR model is that there is no category for dead individuals which is a significant factor to consider when modeling COVID-19. Additionally, being a stock model, there are no variables to account for response protocols like quarantining. While fundamentally limited, the stock SIR model provided the team with a basic understanding of compartmental models and ways to improve future models.

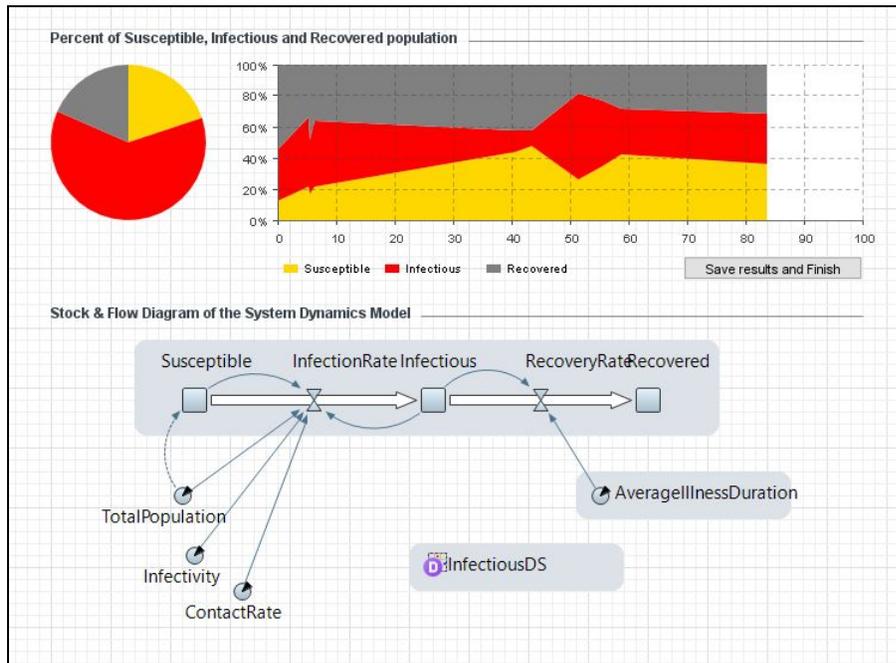


Figure 13: Stock SIR model scaffolding and accompanying graph.

SEQJIR Model

The second AnyLogic model worked with was a more complex compartmental model that, at its core, contained the same scaffolding as a SIR model. The SEQJIR model differentiates itself from the SIR model with the addition of three new compartments and a plethora of new variables that influence flow rates. As can be seen in Figure 14, an individual first has to be exposed to another individual with the virus to move on to further compartments. Additionally, there are two different periods of time where an individual can quarantine: when they are exposed but not infected or when they are infected.

This model was initially created for the introduction of the quarantine compartments as well the addition that not all susceptible people can become infected as a number may already be isolated from society. The idea was to eventually add a new flow path from infectious to dead, but that never came to fruition because the model contained too many coefficients that required calibration. Even with the gathered historical data for the six countries, like death, infection, and recovered totals over time, much more data would be required to calibrate the coefficients in this model. The team looked to simplify for the coming models with hopes that one of them would occupy the ability to be calibrated with the gathered information.

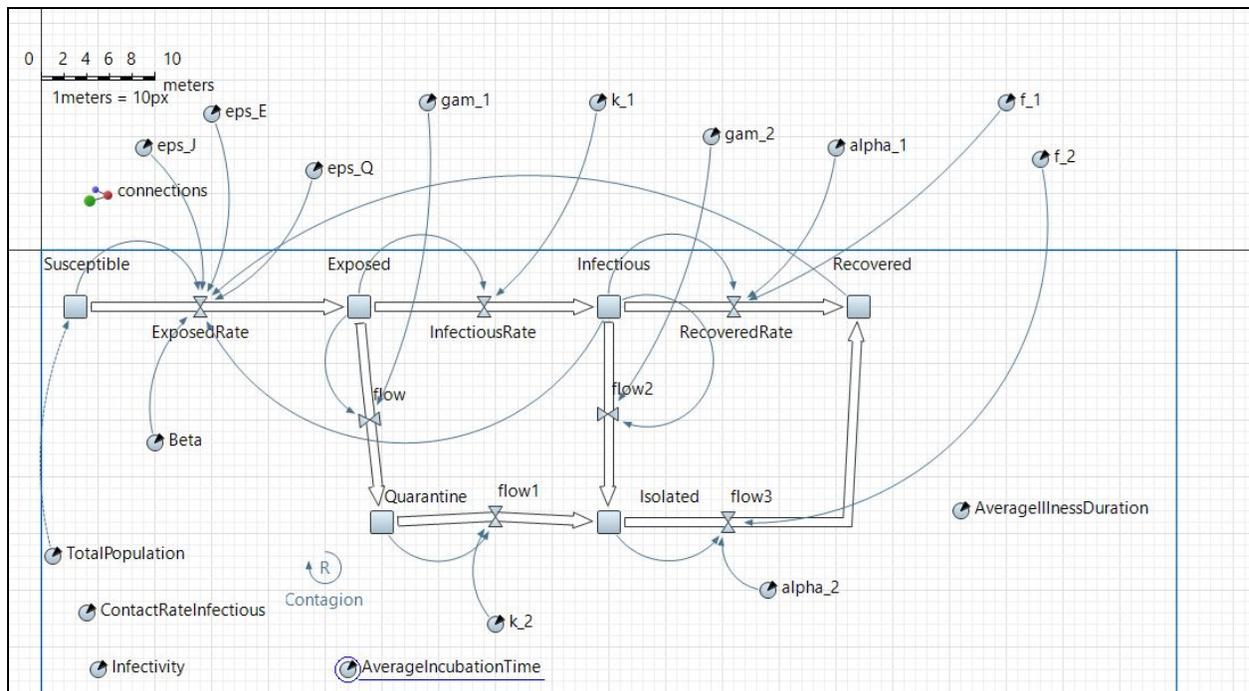


Figure 14: SEQJIR model scaffolding showing many uncalculated coefficients and six compartments.

In an effort to calibrate the infectivity coefficient, 60 days of United States' infections were first inputted into AnyLogic. Despite over 85 days of data since this first domestic infection available, the idea was to project beyond the 60th day and see if the prediction lined up with the empirical data. When this calibration produced inaccurate projections that continued to increase exponentially, 81 days of data were imputed to dial in the calibration and hopefully get the infection curve to peak earlier. The self isolation coefficient was also set to 0.5 to emulate half of the population quarantining strictly. As can be seen in Figure 16, the best feasible projection, shown in blue, does not follow the empirical data, shown in gray. While the calibration yielded an infectivity value, the team determined this value to be not useful. Nonetheless, the calibrated model was tested to see what kind of results it would produce.

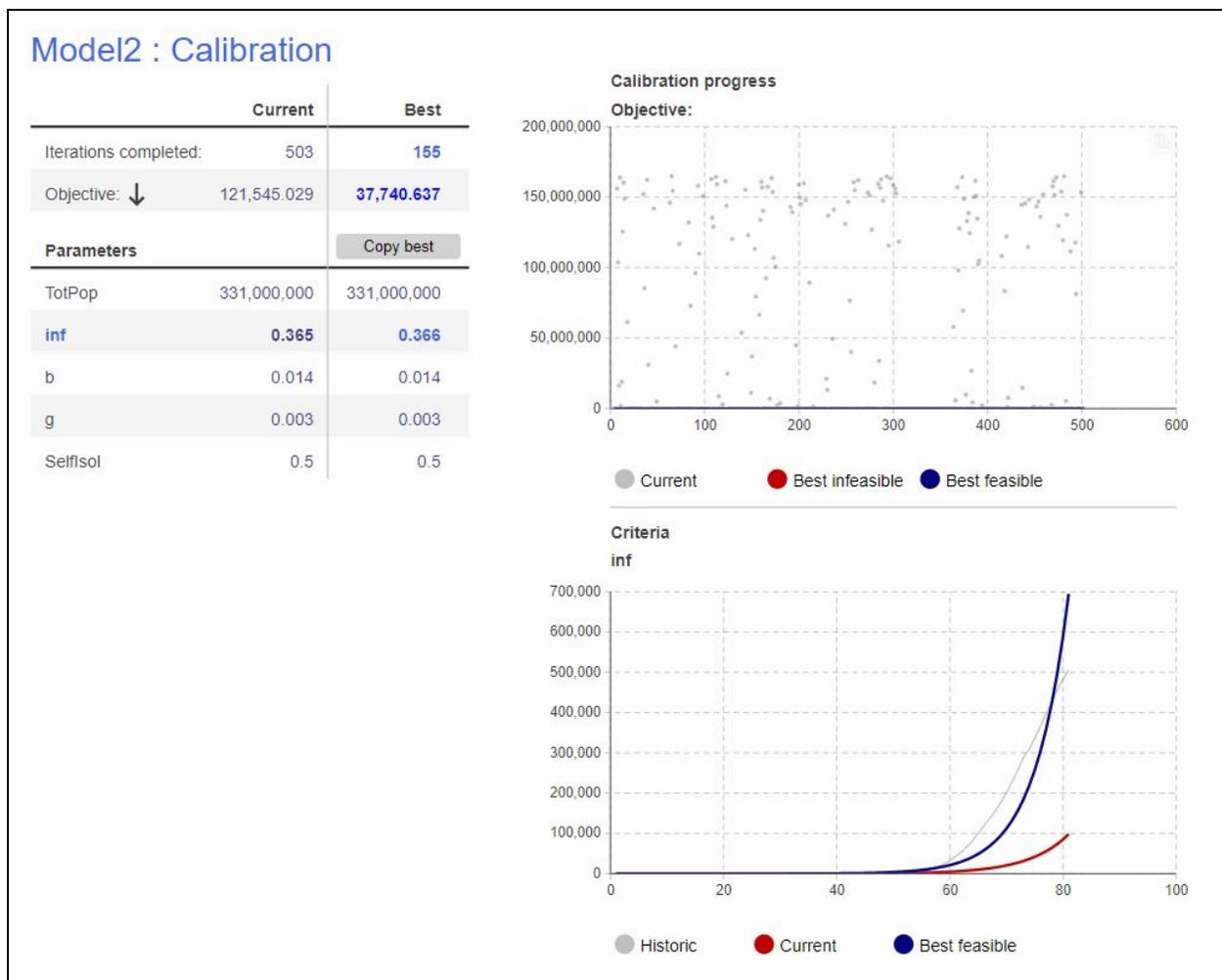


Figure 16: Calibration of HIRD model with US data yielding inaccurate infectivity.

When the calibrated HIRD model was run with the United States' population, it showed nearly 80 percent of the entire population becoming infected as can be seen in the upper right corner of Figure 17. As the team expected, the unsuccessful calibration led to inaccurate results. When Professor Korovin was consulted, he commented that the calibration lacked data and complexity. Ultimately, the historical death, infection, and recovered totals gathered for all six countries was not enough to accurately calibrate a system dynamic model in AnyLogic. When the Republic of Korea's data—a country that showed a complete plateau in total cases—was used to calibrate the same model, the application yielded similar results to the United States calibration. At this juncture, the team decided to cease creating predictive models and instead develop models that exemplify the importance of quarantining and social distancing.

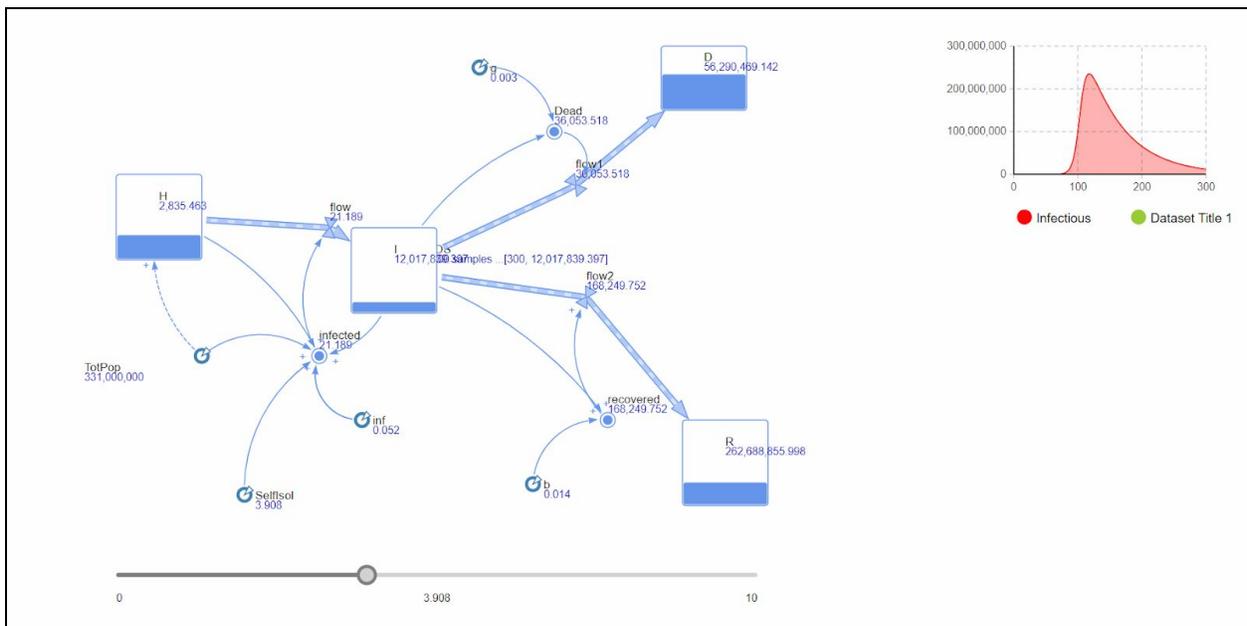


Figure 17: HIRD model results for the United States showing total infected, recovered, and dead after 300 days.

SEIR Model with Sliders

The last compartmental model created was a SEIR model that included adjustable coefficients. The two coefficients that can be adjusted via sliders are infectivity and contact rate, which both contribute to the rate individuals move from the susceptible compartment to the exposed compartment and eventually the infected compartment (Figure 18). While this model does not include a death category, its main purpose was to demonstrate how lowering the contact rate (i.e. the percentage of a population quarantining and social distancing) lowered the number of individuals in a population that got infected.

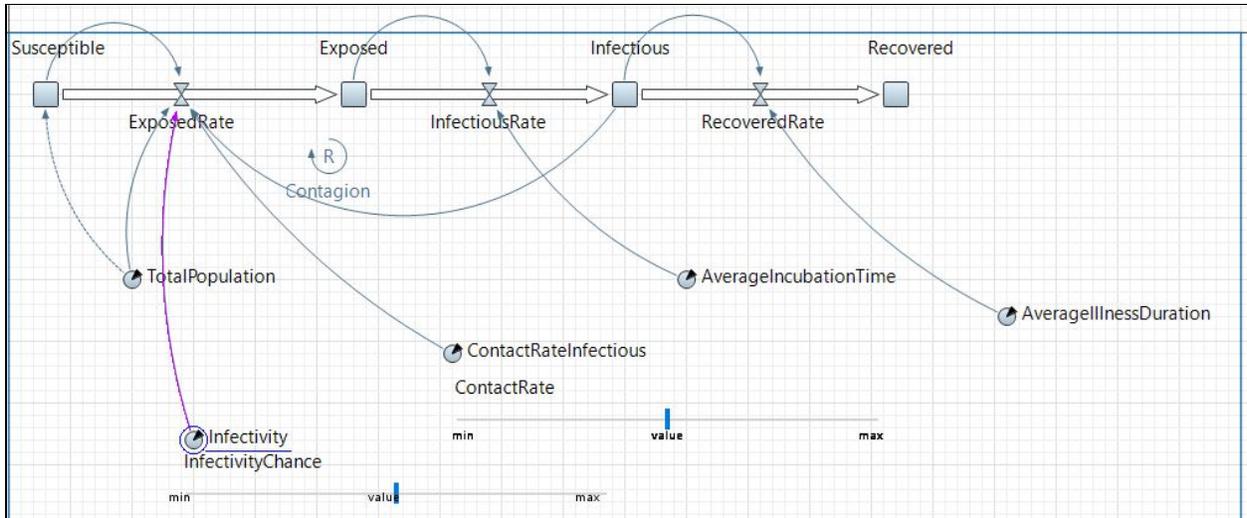


Figure 18: SEIR model scaffolding showing sliders for infectivity and contact rate.

A simulation was run with the SEIR model with an initial population of 10,000 individuals in which the infectivity remained constant, but the contact rate was varied at two places during the simulation. Initially, the contact rate was set to 10 contacts per day and the infectivity was set to 0.422. Essentially, if a susceptible individual came into contact with an infected individual, they had a 42.2 percent chance of contracting the virus. The initial number of infected individuals was set to 1 because each population always needs to have a first case. As can be seen in the graph shown in Figure 19, the number of infected individuals increased slowly at first, but increased quickly once more individuals moved into the exposed and infected compartments. The rate at which individuals moved from infected was simply determined by the average illness duration which was set to 15 days to emulate COVID-19.

At around 30 days into the simulation, the contacts rate was promptly changed to 0 contacts per day. As can be seen in Figure 19, the susceptible population flattened off immediately, since logically, no new individuals were susceptible to the virus. Meanwhile, the number of infected individuals increased for just a little while longer before dropping off to nearly zero as everyone recovered. Later, at around 160 days, the contact rate was moved to 3.15 contacts per day and the remainder of the population that had not yet developed immunity to the virus, became infected and then eventually recovered.

The simulation showed a virus passing through an entire population and demonstrated how a lowered contact rate can decrease the number of new infections. This model points to the significance of quarenting during a pandemic. When there are less healthy individuals exposed to infected individuals, the pandemic ends sooner with lower infection totals.

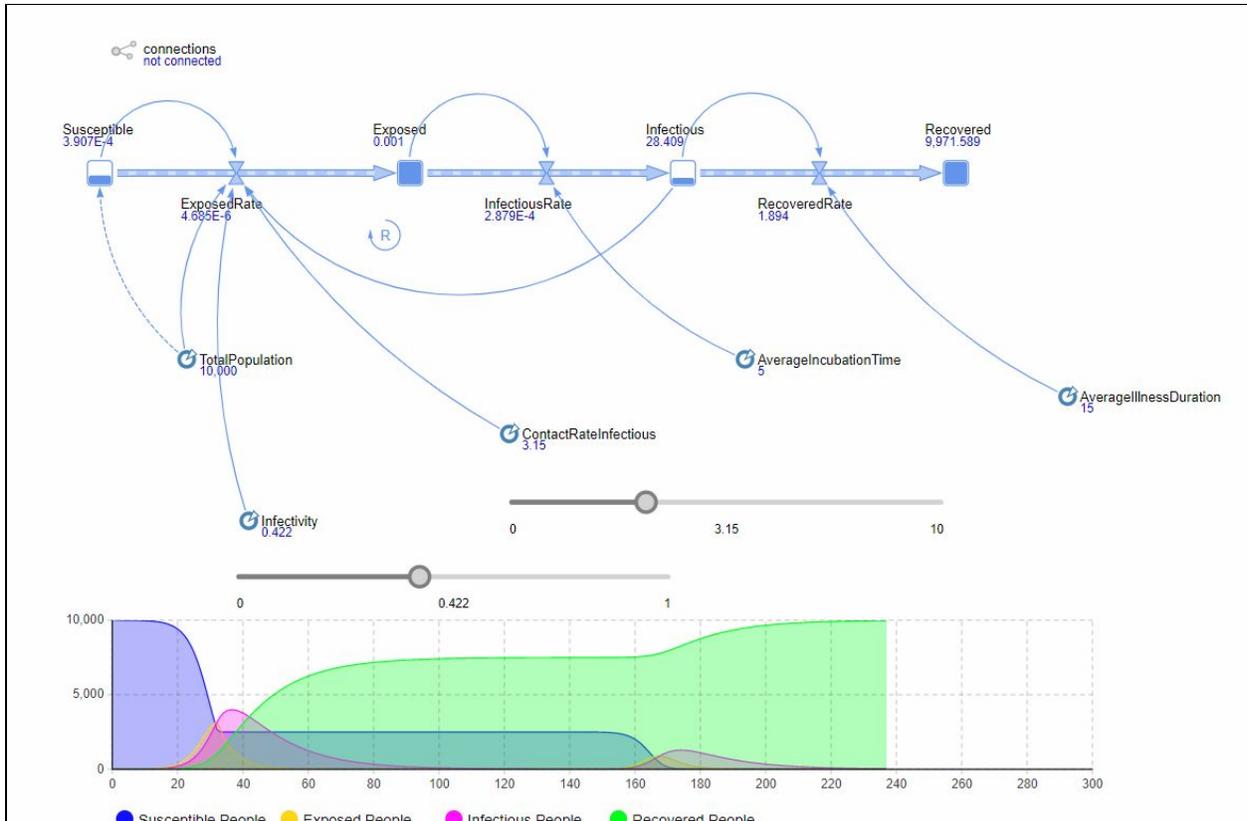


Figure 19: SEIR simulation showing how decreasing contact rate lowers new infections.

Town Model: Agent-Based

The agent-based town model was a huge leap for the modelling aspect of this project. The town model was, by far, the most visually understandable model developed in AnyLogic. In this town model, shown in Figure 20, a population of 500 people, who lived in 100 separate houses (approximately 5 people per house) traveled to and from grocery and convenience stores to get food for themselves (at a rate of approximately 1 visit per 5 days). A section of the GIS, Geographic Information System, the space in which the population resides. Yellow buildings are placed in the upper left corner of the image and denote the location of supermarkets in the town. The small green buildings represent convenience stores and the small blue buildings are homes where the residents of the town spend most of their time. Finally, the gray, windowed buildings near Sandwich Street represent clinics where the residents go to see if they have been infected.

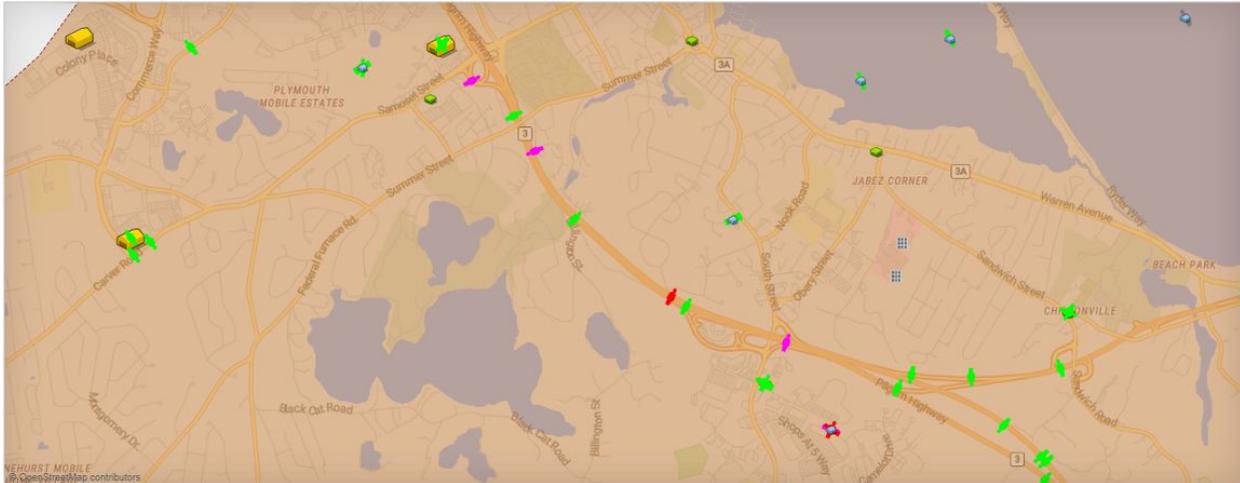


Figure 20: The town featured, Plymouth, MA, is where the agents reside. People are travelling along road networks to get from their houses to a grocery store.

When these people entered their house or a store, they established a contact network with all other people in the same building. Once this contact network was established, each person had a significant contact rate of 100 contacts per day. A significant contact implied that if a person were to be infectious with SARS-CoV-2, they would have 100 instances throughout the day where they could potentially spread the virus to other people. To assume that every person in a grocery store was equally likely to have significant contact with every other person in that store was unrealistic, so the contact rate for when a person entered a store was changed from 100 to 10 significant contacts per day. Meanwhile, when a person is in their home, they have a much higher chance to infect their family members.

A SEIR state chart, seen in Figure 21, which is a process for agents to go through the specific steps of an epidemic, was used within the agent based model to simulate the spread of the infection. When a person was depicted as green, they were healthy but at risk of contracting the virus. The exposed state was depicted as magenta, when a person was this color it meant that they had come in contact with the virus and would transition to the infectious state in a few hours. A person in the infectious state was colored red, these individuals were responsible for the spread of the virus. Finally, the gray state meant that people had recovered from having the virus and were immune from catching the virus again. All of the states can be seen in action in Figure 20 as well.

Another state chart was built to simulate the mobility of an agent going to and from the store, (Figure 21). The shopping state chart allowed for the model to be dynamic and resulted in each agent having distinct movement patterns.

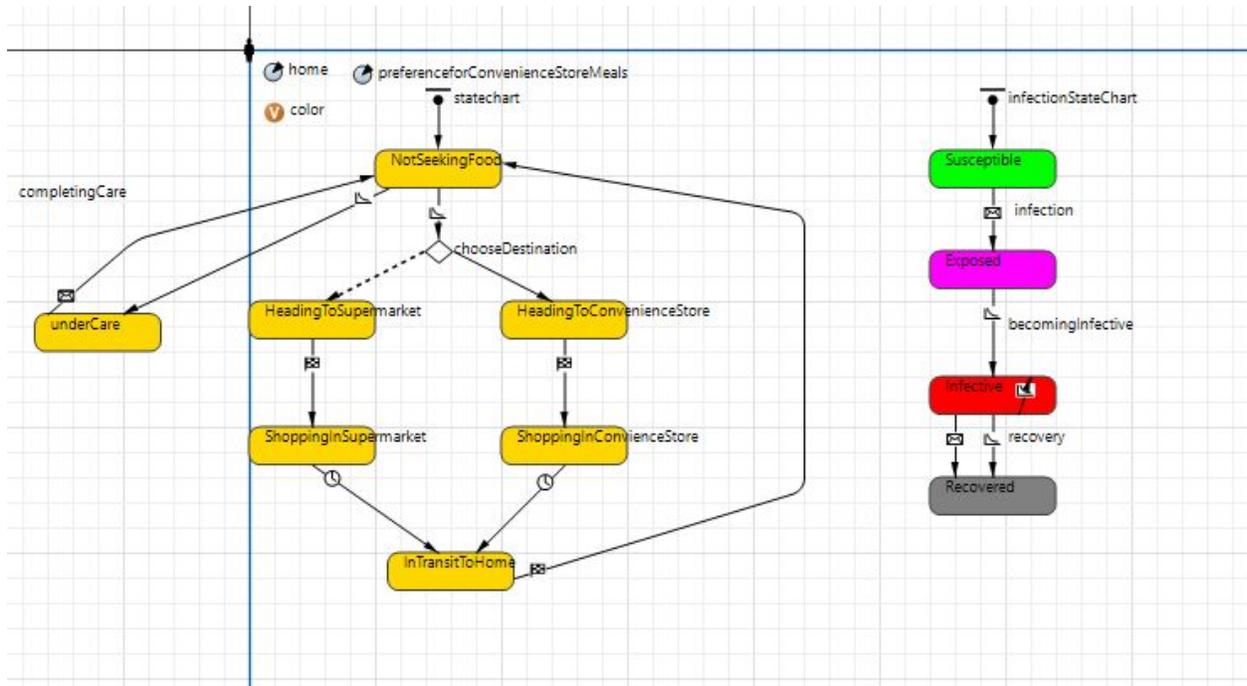


Figure 21: The two state charts that determine the location, the yellow state chart, and the infection progression, the green-magenta-red-gray chart, for each person in the town.

Finally, two quarantine methods were implemented on separate model runs, each instituted a certain number of days after the initial infection event occurred. One quarantine measure was applied at seven days from the start of the simulation (Figure 22), while the other was applied at 21 days from the start of the simulation (Figure 23). These measures set the home contact rate and store contact rate to 5 per day, from 100 contacts per day in the home setting and 10 contacts per day in the store setting. Eventually, the quarantine resulted in the retardation of the spread of SARS-CoV-2 in the town’s population.

Yet, the seven day quarantine measure was far superior in mitigating the spread of the virus than the 21-day quarantine measure. The peak number of infected individuals for the seven day quarantine scenario was 26 at one time. Meanwhile, the peak number of infected individuals for the 21-day quarantine measure was over 150 people at one time.

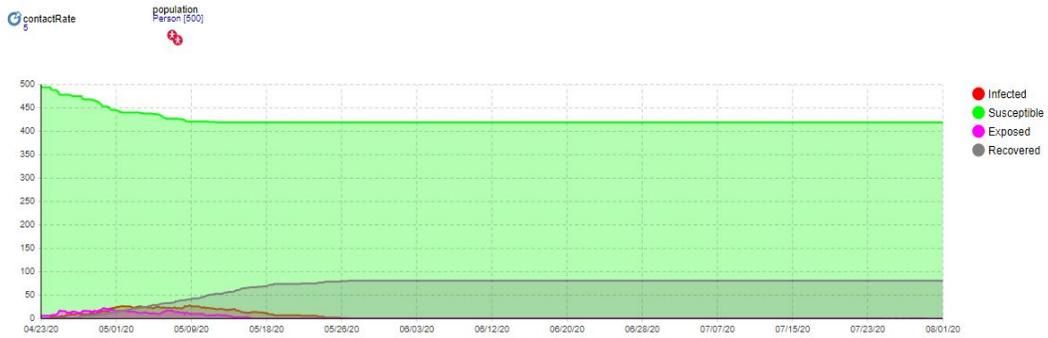
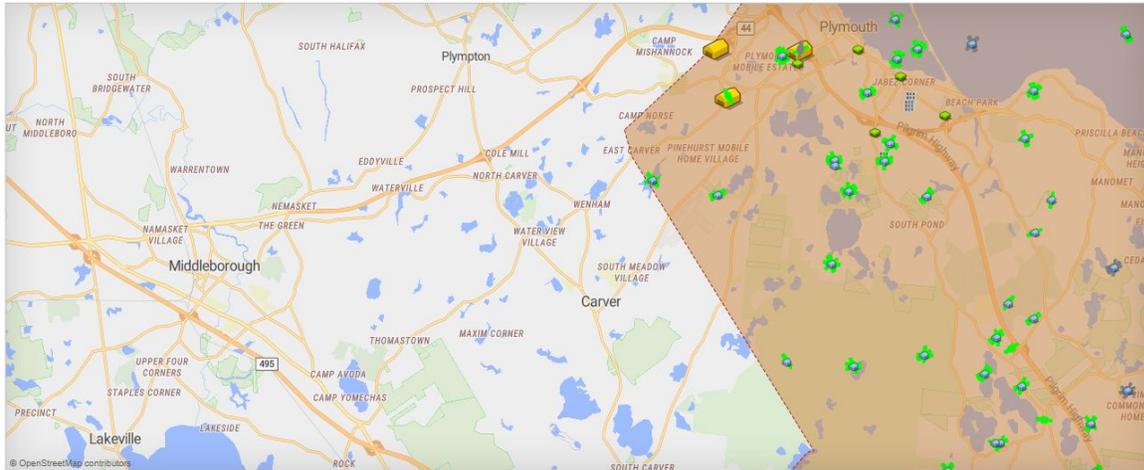


Figure 22: Representation of a 500 person population if a town-wide quarantine is put into place, 7 days after the first infection occurs, which reduces individual contacts per day from 100 to 5.

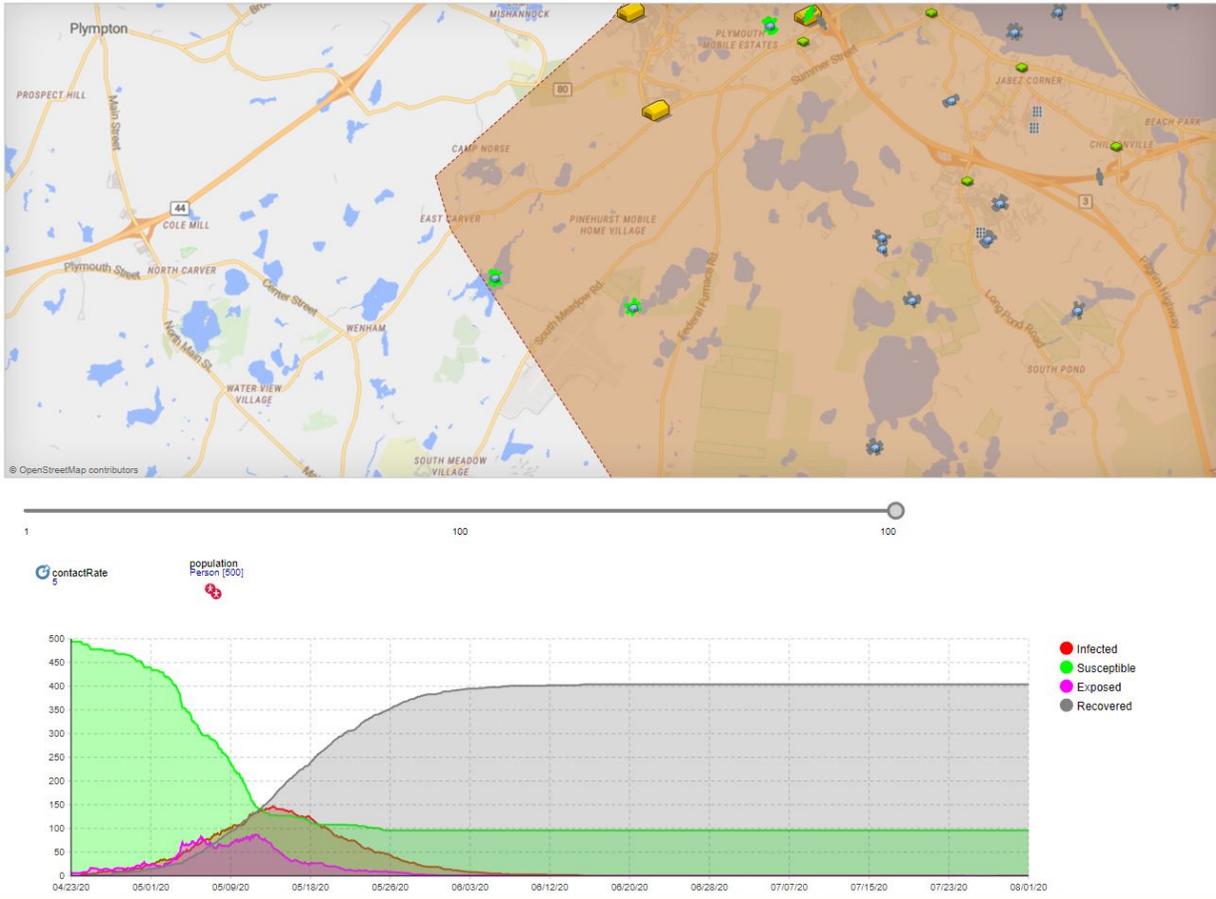


Figure 23: Representation of a 500 person population if a town-wide quarantine is put into place, 21 days after the first infection occurs, which reduces individual contacts per day from 100 to 5.

Not only does the later quarantine reaction result in a higher number of people being infected over the entire duration of the pandemic, but it also has a much greater strain on the healthcare infrastructure of a town or city. Although the strain on healthcare infrastructure was not considered in any of the models this team produced, it is the main concern of many epidemiologists and health policy modellers. If more members of a population are infected at one time, the higher the potential strain could be on medical resources and the medical workforce when treating all of these people. The goal of this model was to illustrate how a 14-day difference in reacting with quarantine measures can entirely alter the course of a pandemic through a population.

The town model is available to be interacted with in the AnyLogic cloud and can be accessed [here](#). Users can adjust the contact rate for all agents and observe how the virus spreads, both on the GIS map and in the SEIR graph.

Econometric Models

Gamma Distribution

Gamma distribution is a type of distribution that occurs naturally in processes for which the waiting times between events are relevant. It also can be used in making predictions.

Gamma distribution equation with 2 parameters:

$$f(t, a) = \frac{1}{b^a \cdot \Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

Gamma distribution function with one parameter:

$$f(t, a) = \frac{t^{a-1} \cdot e^{-t}}{\Gamma(a)}$$

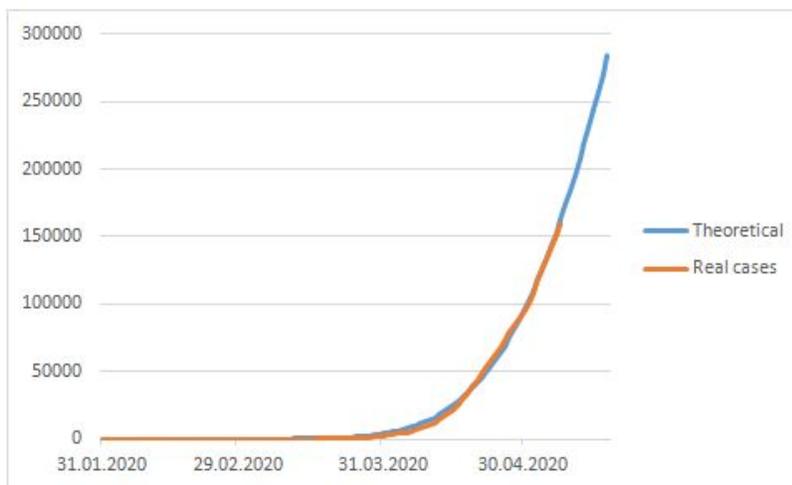


Figure 24: Gamma distribution - Russia

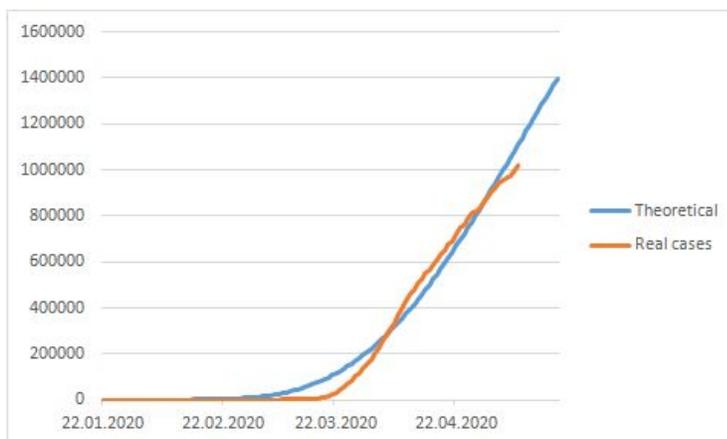


Figure 25: Gamma distribution - United States

The team constructed graphs for all the countries where the number of infected people depends on the number of days since the beginning of the epidemic (Figure 24) and (Figure 25). From research we found that Gamma distribution works well for the prediction in the countries which have increasing growth rates of cases, but as soon as the country reaches the plateau, Gamma distribution works worse for predicting a trend.

Exponential and Logistic Curves

The reason to use Exponential Growth for modeling Coronavirus outbreaks is that epidemiologists have studied those types of outbreaks. It is well known that the first period of an epidemic involves Exponential Growth. The exponential curve is characterized by the following formula:

$$f(x) = a_1 \cdot e^{a_2 \cdot x} + a_3 \cdot x^2 + a_4 \cdot x + a_5$$

Logistic regression is used to predict the probability of occurrence of an event by fitting data to a logistic curve. A logistic regression model is built according to the following expression:

$$f(x) = \frac{b_1 + b_2 \cdot e^{b_5 \cdot x}}{b_3 + b_4 \cdot e^{b_5 \cdot x}}$$

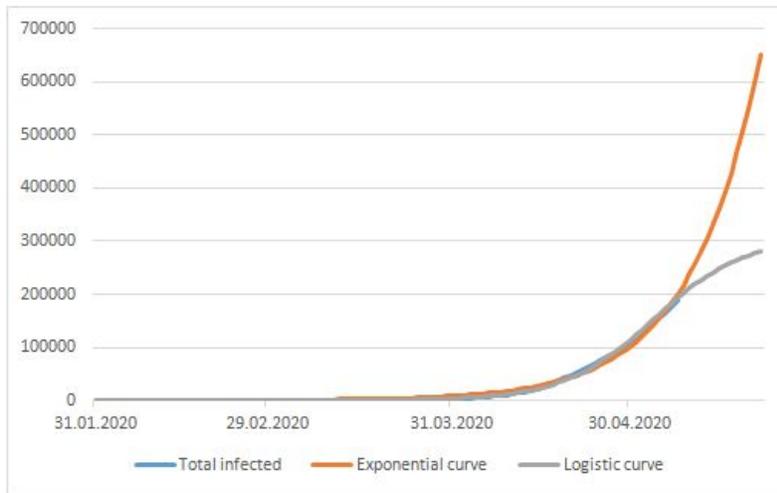


Figure 26: Exponential and Logistic curves for total number of infected - Russia

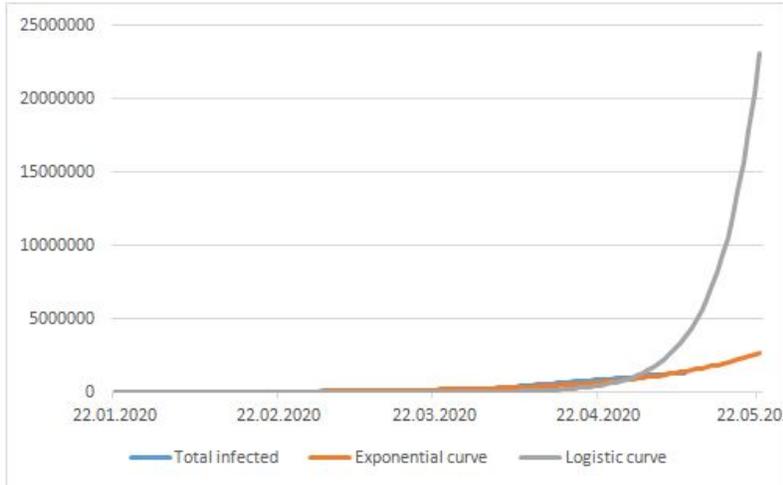


Figure 27: Exponential and Logistic curves for total number of infected - United States

From the graphs we constructed for all 6 countries we could see that the exponential curve prediction was often more accurate at the beginning of the function, but later the exponential curve would grow at a higher rate. The logistic curve could be used as a much better approximation. A good example of the approximation can be seen in Figure 26, which shows a prediction for Russia's case of the virus. Figure 27 shows an exponential curve and that it is possible to find coefficients that work better than those for logistic graphs. Nevertheless, the prediction curve peels off.

Moving Averages

Moving averages are a simple and common type of smoothing used in time series analysis and time series forecasting.

Simple moving average for 7 periods:

$$M_7 = \frac{y_t + y_{t-1} + \dots + y_{t-8}}{7}$$

Double moving average for 7 periods:

$$M'_7 = \frac{M_t + M_{t-1} + \dots + M_{t-8}}{7}$$

Level at time t:

$$a_t = 2M_t - M'_t$$

Slope at time t:

$$b_t = \frac{2}{n-1}(M_t - M'_t)$$

Forecast at time t:

$$F_t = a_t + b_t t$$

Moving averages were used to make predictions for all of the countries' trends. It was one of the most accurate predictions delivered by the team. Figure 28 below shows such predictions for active cases in Russia, US, Italy, Germany, China and South Korea. A quotient of the number of people in each country was used to make the comparison more clear.

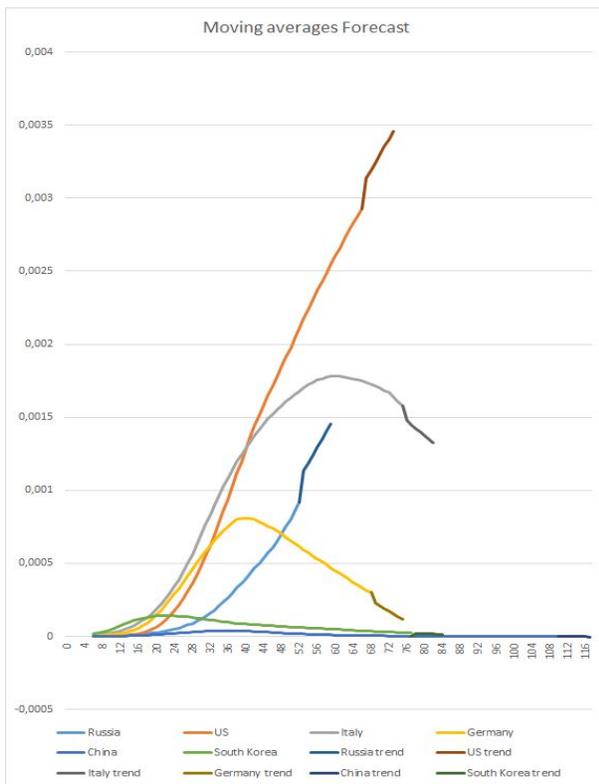


Figure 28. Moving average Forecast for Russia, US, Italy, Germany, China and South Korea.

Moving Averages technique was also applied to make a comparison of new cases per day out of total cases (Figure 29). The similar trends were noted, and were used to apply them to the prediction of the country's future new cases of infected people.

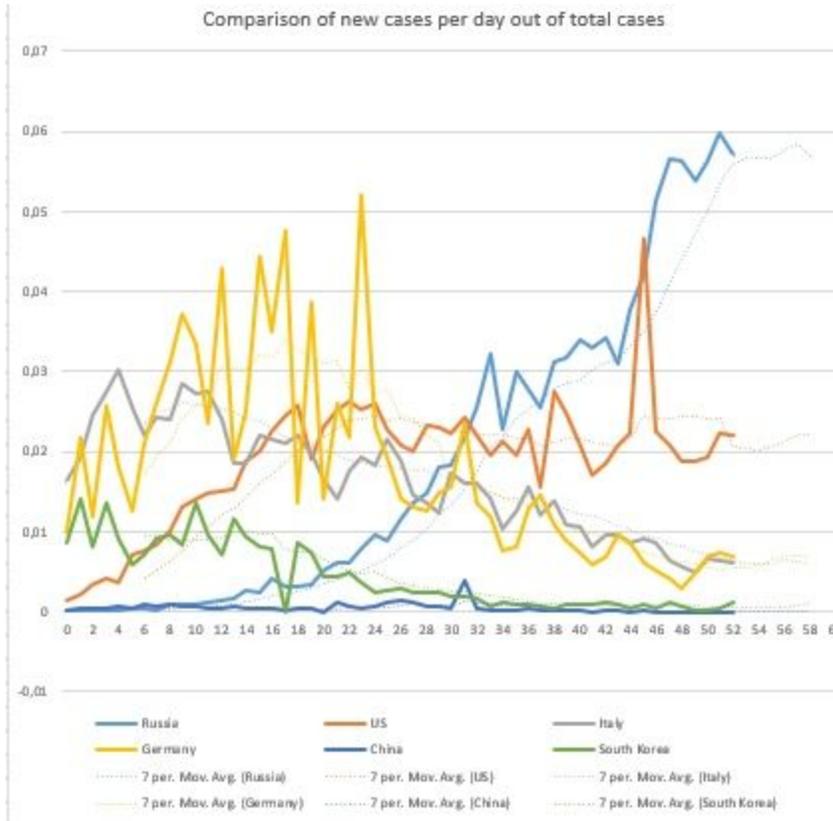


Figure 29. Comparison of new cases per day out of cumulative total cases for Russia, US, Italy, Germany, China and South Korea.

Forecasting Model

Some countries have trends similar to those of other countries. The similar trends were used to predict future values in the models. The next graph shows a prediction for the models based on the trends of other countries (Figure 20).

$$\begin{cases}
 y = a * t^2 + b * t + c; t - \text{number of days since start of epidemics;} \\
 a = a1 * p1 + a2 * p2 + a0; \\
 b = b1 * p1 + b2 * p2 + b0 \\
 c = c1 * p1 + c2 * p2 + c0; p1 - \text{density per km}^2, p2 - \text{total number of tests for COVID} - 19.
 \end{cases}$$

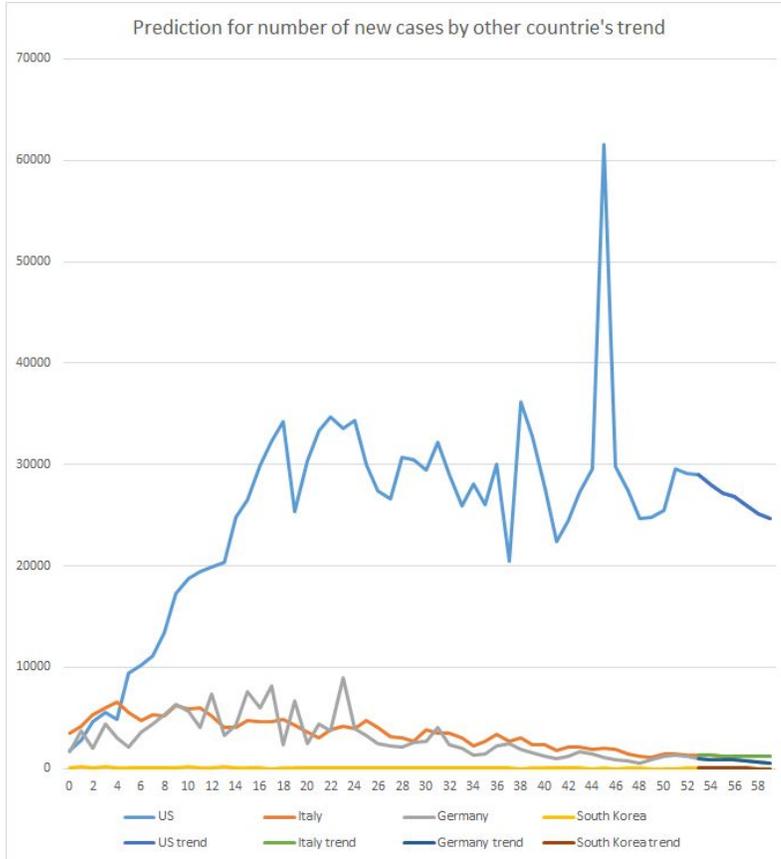


Figure 20. Prediction of other countries' trends for the US, Italy, Germany and South Korea.

The graph of Moving averages for Italy at the final dates is similar to the graph for Germany from day 45 to day 52, which made it eligible to be used for making predictions. Also the graph of South Korea looks similar to that of China at the period from day 33 to day 39. The team was not able to make predictions for Russia, due to the highest numbers for moving averages. For the US a section from day 17 to day 23 for Italy could be used. Predictions also could not be made for China due to the smallest moving averages.

Chow Test for Model Stability

The Chow test allows to determine the structural stability of variables in a model. A statistical test was used to assess significance, improving the regression model after dividing the original sample into parts.

The team noted the dates of quarantine enforcement in the six countries: quarantine started March 9 in Italy, March 13 in the USA, January 22 in China, March 16 in Russia, March 6 in South Korea, March 22 in Germany (Figure 31). The hypothesis about the significant impact of quarantine implementation in selected countries was checked with the Chow test. The samples of each country were divided into those of quarantine and the n-th day of quarantine implementation.

For Italy, a split of data did not show an actually high effectiveness of the implemented quarantine. The effectiveness largely depended on how people followed such restrictions and recommendations. There were 50 confirmed cases per day, when people were not being isolated.

In Russia $F_{crit} > F$ means that splitting data did not make sense, as the Chow test also showed, when the probability of accepting a hypothesis does not increase by taking as example 45, 69, 87 days. Data was not stable. Russia's way of treatment with pandemics was not efficient, it has still not reached plateau, numbers of infected were only increasing.

In South Korea the hypothesis was neglected as single models quality oversaw the quantity of overall model of regression. In South Korea quarantine was not universal, but aimed at some individual elements of human activity.

For China the data was structurally stable, the dependencies of indicators before quarantine and after do not differ, the probability of accepting the hypothesis that changed effect was more than 5%. Dependence was true for both samples.

For the USA probability for implementing hypotheses was relatively small ($< 5\%$), the dependencies of before and after implementing quarantine differed, data were not structurally stable.

In Germany, through 62,81,93 days, taken for instance, the probability of taking the hypothesis was not increasing. The data was not stable. An extensive testing program and relative success in protecting the most vulnerable parts of the population has led to a decrease in mortality in Germany compared to that of Italy, Spain and the UK.

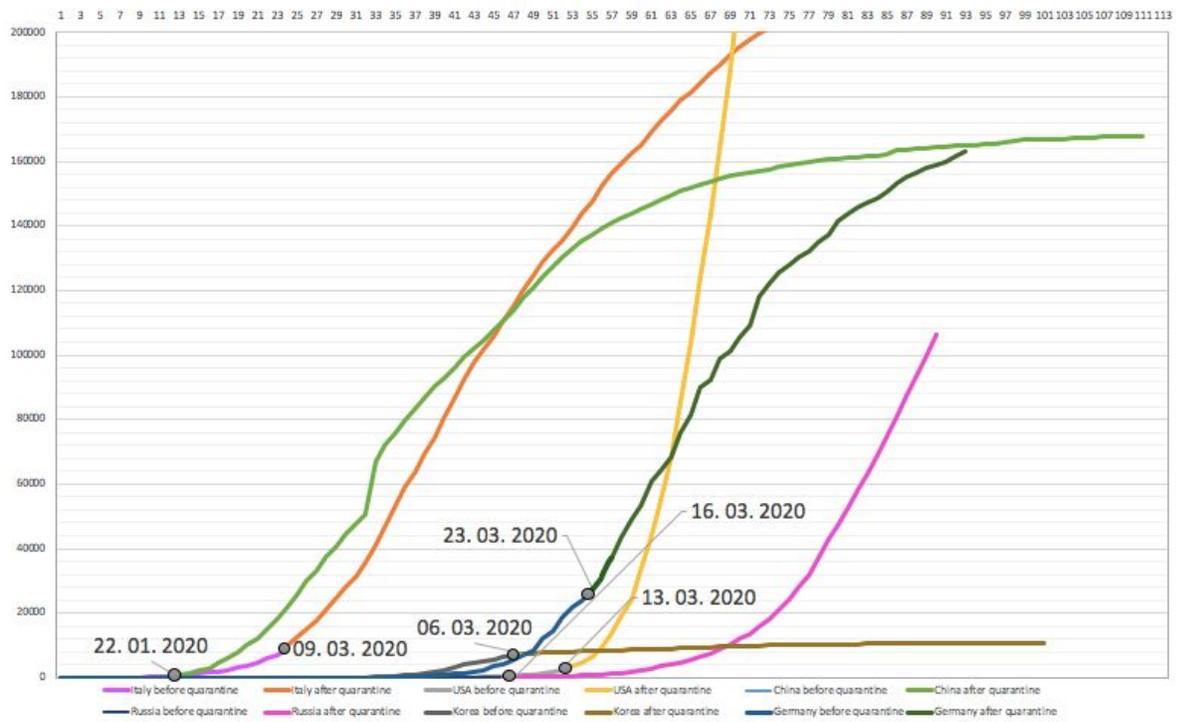


Figure 31. Amount of total infected in six target countries before and after quarantine.

Quadratic Model

The team decided to use 8 parameters for testing the econometric models for each country. Fertility rate was implemented, showing the quantity of children born by one mother; employment rate was taken as a possible indicator of showing dependence of wealth and ability of getting COVID-19; high population densities were considered if they could act as catalysts to evolving epidemics; net migration rate – difference between inward coming and outwards leaving; health care index - overall quality of the health care system, including health care infrastructure; health care professionals (doctors, nursing staff, and other health workers) competencies, cost (USD per capita), quality medicine availability, and government readiness; the amount of hospital beds was an important parameter; the population over 60 years old was considered to be mostly receptive of the coronavirus; the total number of COVID-19 tests made in each country.

The mathematical representation of an econometric quadratic model with 2 parameters looked like this:

$$\left\{ \begin{array}{l} y = a * t^2 + b * t + c; t - \text{number of days since start of epidemics}; \\ a = a_1 * p_1 + a_2 * p_2 + a_0; \\ b = b_1 * p_1 + b_2 * p_2 + b_0 \\ c = c_1 * p_1 + c_2 * p_2 + c_0; p_1 - \text{density per km}^2, p_2 - \text{total number of tests fpr COVID - 19.} \end{array} \right.$$

In order to find coefficients, the project team used the least squares method. As a result, two graphs were produced, one for the real data (Figure 32) and one with the Econometric model data (Figure 33). These factors had a significant influence on spread of COVID-19, as evidence the data presented in the final graph looks close to real data. Other tested factors have had less significant influence on the outcome.

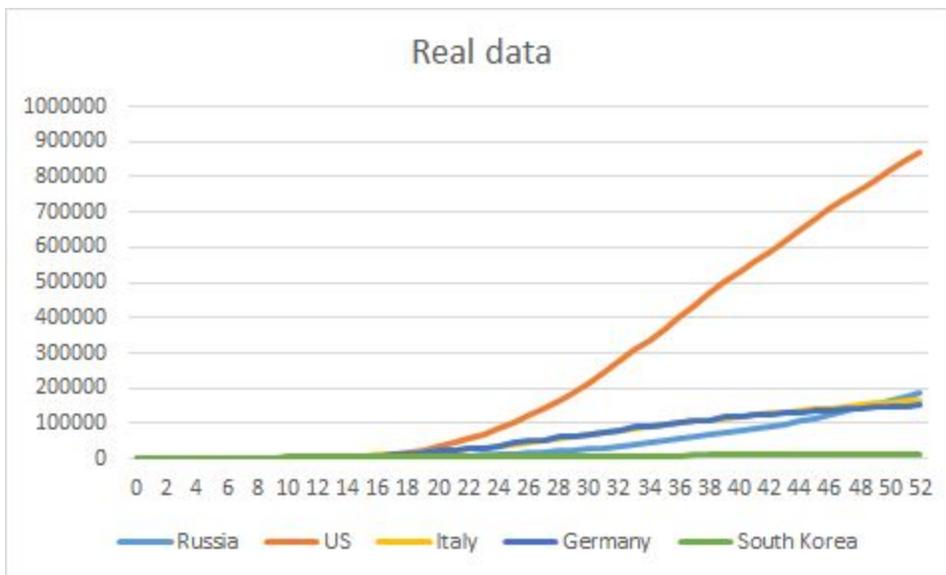


Figure 32: Graph of total number of infected people in 5 countries

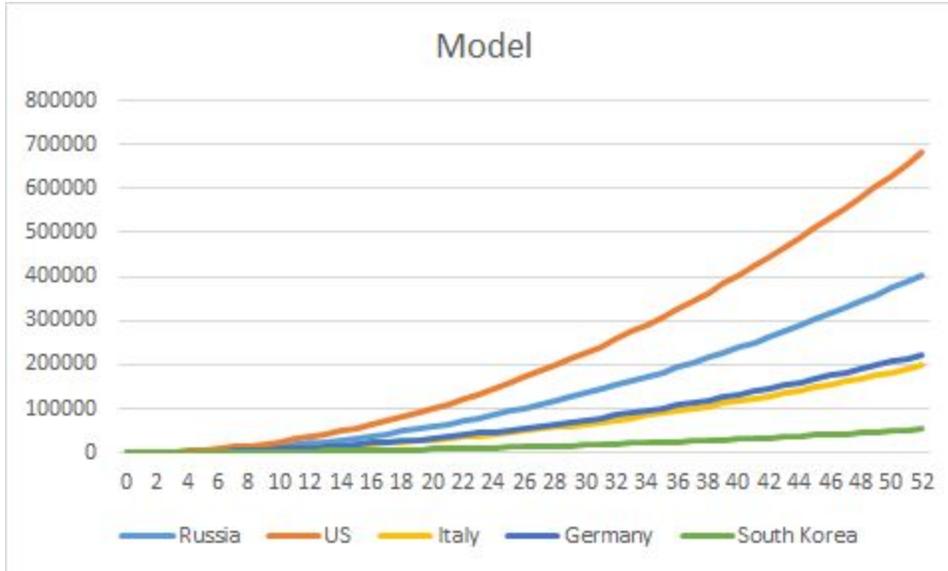


Figure 33: Econometrics model of total number of infected people with 2 parameters: density and number of tests

Sigmoid Models

The geometric progression worked well in the initial period of the spread of the virus, but then began to produce very high results. The inaccuracy forced the team to try another model, which would show realistic results. Logistic curve, which was discussed previously, also a type of sigmoid model, but it was a very simplified version, which caused it to not always work properly.

For the first model we had $P = \text{const}$ be a population in the country. In Germany, which model we constructed, $P = 83719792$ people. For the first model the team used the Malthus-Verhulst equation:

$$\frac{dI}{dt} = rI \left(1 - \frac{I}{P}\right), I(0) = 1$$

t – number of days from the beginning of epidemic.

$I(t)$ – total number of infected people.

r - how many healthy people can be infected on average by one infected per day.

For our case we have 1 person infected in the beginning of the epidemic.

The equation solution was a sigmoide (Sig), and its first derivative was the Hubb function:

$$I(t) = \text{Sig}(t) = \frac{P}{1 + e^{-r(t-t_0)}}$$

$$\frac{dI}{dt} = \text{Hubb}(t) = \frac{rP e^{-r(t-t_0)}}{[1 + e^{-r(t-t_0)}]^2}$$

t_0 - maximum point of Hubb function, so we can calculate this.

At day 0 the epidemic had only one person infected, so the team applied it to the Sig function:

$$I(0) = \frac{P}{1 + e^{-r(t-t_0)}} = \frac{P}{1 + e^{rt}} = 1$$

$$t_0 = \frac{1}{r} \log(P - 1) \approx \frac{\log(P)}{r}$$

If this infection had an infinite infecting period, $r = \text{const}$. For the first model the project team used $r = 0,267$.

In this case $t_0 = \frac{\log(P)}{r} = 68$

After that the team applied everything in the Sig formula (Figure 34).

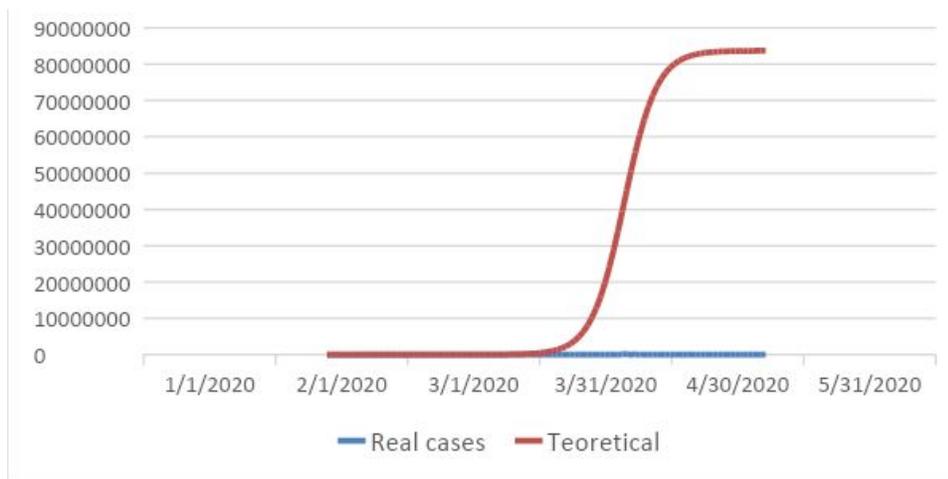


Figure 34. 1st Sigmoid model for Germany with infinite time of spread of COVID-19

The model representing theoretical data grew much faster than the real data model. The first idea for improvement of the model was a method to calculate r properly. The following formula was used to calculate it:

$$R_0 = r \int_0^t \rho(t) dt$$

R_0 - basic reproduction number.

Chinese scientists approximated this number for COVID-19 as 2-2,5; for construction of the model the value 2,4 was used.

$$0 \leq \rho(t) \leq 1$$

The area of the subgraph of the infectivity function given in the table was multiplied by the transmissibility r . The unit of measure was people. As many people as were already infected could infect the same amount on average over a period of contagiousness if no protective measures were taken and there were enough candidates around to be infected.

The example was constructed table function for $\rho(t)$ as a table with 20 days of possible infectivity distributed by days. Integral of this function was equal to 9, which was close to the situation with other diseases with similar levels of infectivity.

The second idea was to divide people into 3 groups:

$I(t)$ – number of infected people, which already was used at the beginning

$C(t)$ – number of contagious people

$S(t)$ – number of susceptible people.

Such division is often used in the models of diseases spread.

Formulas, which describe the model:

$$\frac{dI}{dt} = r \frac{S(t)C(t)}{P}, I(0) = 1$$

$$\frac{dS}{dt} = -r \frac{S(t)C(t)}{P}, S(0) = P - 1$$

$$C(t) = \int_0^t \frac{dI(\tau)}{d\tau} \rho(\tau) d\tau$$

Number of Contingent people was calculated by multiplying the number of Infected people for the previous 20 days by the coefficients from the table to find the possible number of contingent people. The resulting graph for the second model (Figure 35):

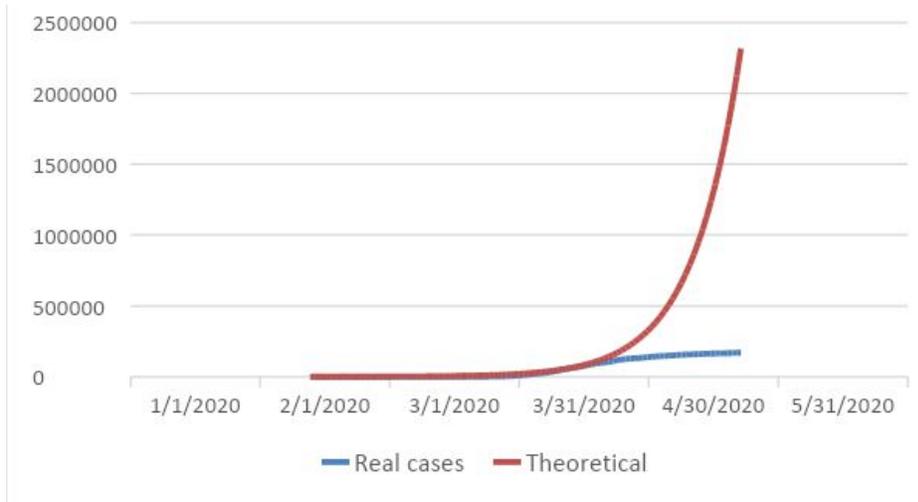


Figure 35. 2nd Sigmoid model for Germany with infectivity distribution

The model showed the level of explanation of the model significantly increased, but the theoretical model still has risen much faster than the real model. The differences existed because the team assumed previously that people did not know about the symptoms. But in case of COVID-19 many people had a clear symptom of the infection. The percentage of people who had the infection symptoms for different sources went from 5% to 75% (because there was no significant research for this topic). For this model the team used 45% who had the infection symptoms, and this number had the best fit for the model.

The final formulas explaining the model:

$$\frac{dI}{dt} = \frac{S(t)}{P} [r_N C_N(t) + r_s C_s(t)], I(0) = 1$$

$$\frac{dS}{dt} = \frac{S(t)}{P} [r_N C_N(t) + r_s C_s(t)], S(0) = P - 1$$

$$M(t) = \int_0^t \frac{dI(\tau)}{dt} \varphi(\tau) d\tau$$

$$r_s C_s(t) = r_s \int_0^t \frac{dI(\tau)}{dt} \rho_s d\tau, \rho_s = \max(p - \varphi, 0)$$

$$r_N C_N(t) = r_n \int_0^t \frac{dI(\tau)}{dt} \rho_n(\tau) d\tau, \rho_n = p - \rho_s$$

$$R_0 = r \int_0^t \rho(t) dt$$

Here R_N - basic reproduction number for the no symptoms people, it equals to R_0 , because people did not distance themselves from other people, because they did not know that they were infected.

R_s - basic reproduction number for the people with symptoms, it equals to $R_0/2$, because people distanced themselves from other people, because they did not want to infect other people (and vice versa).

$C_N(t)$ – number of contagious people with no symptoms.

$C_s(t)$ – number of contagious people with symptoms.

$\varphi(\tau)$ - likelihood of symptoms functions as a table for 20 days probabilities.

$M(t)$ - number of overt patients (the ones that are used in statistics).

Constructed third model (Figure 36):

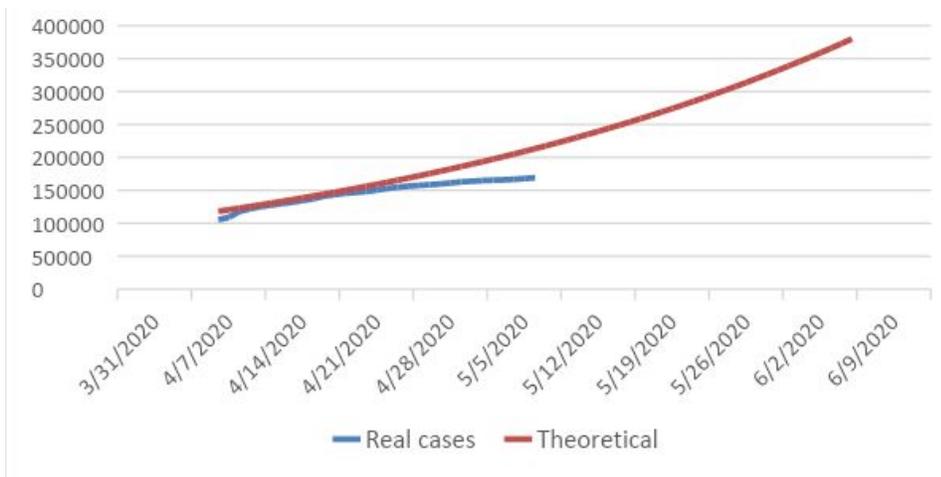


Figure 36. 3rd Sigmoid model for Germany with self-isolation and social distance

Here it is clearly seen that the level of explanation became closer to the real data, but the model was still slightly higher than the real data. The situation exists because the team did not make any quarantine assumptions. But in Germany the government used quarantine measures to reduce the number of infected people. Quarantine changed only basic reproduction numbers for the no symptoms people, because people with symptoms were already highly avoided by the public. So, the reproduction number for people with quarantine was approximated as $R_0/1,3$; because the number of contacts was decreasing, but not as significantly, as for the case of a person with symptoms. All the contagious people also would be divided into a group of

not-working people 70% with reproduction number for people with quarantine with R_0/a ; and a people who had to work $b = 30\%$ of contagious with no symptoms with R_0 .

We construct the final model:

$$\frac{dI}{dt} = \frac{S(t)}{P} \left[\left(\frac{a}{b} + 1 - a \right) r_N C_N(t) + r_s C_s(t) \right], I(0) = 1$$

$$\frac{dS}{dt} = \frac{S(t)}{P} \left[\left(\frac{a}{b} + 1 - a \right) r_N C_N(t) + r_s C_s(t) \right], S(0) = P - 1$$

$$M(t) = \int_0^t \frac{dI(\tau)}{dt} \varphi(\tau) d\tau$$

$$r_s C_s(t) = r_s \int_0^t \frac{dI(\tau)}{dt} \rho_s d\tau, \rho_s = \max(p - \varphi, 0)$$

$$r_N C_N(t) = r_n \int_0^t \frac{dI(\tau)}{dt} \rho_n(\tau) d\tau, \rho_n = \rho - \rho_s$$

$$R_0 = r \int_0^t \rho(t) dt$$

The result of the construction (Figure 37):

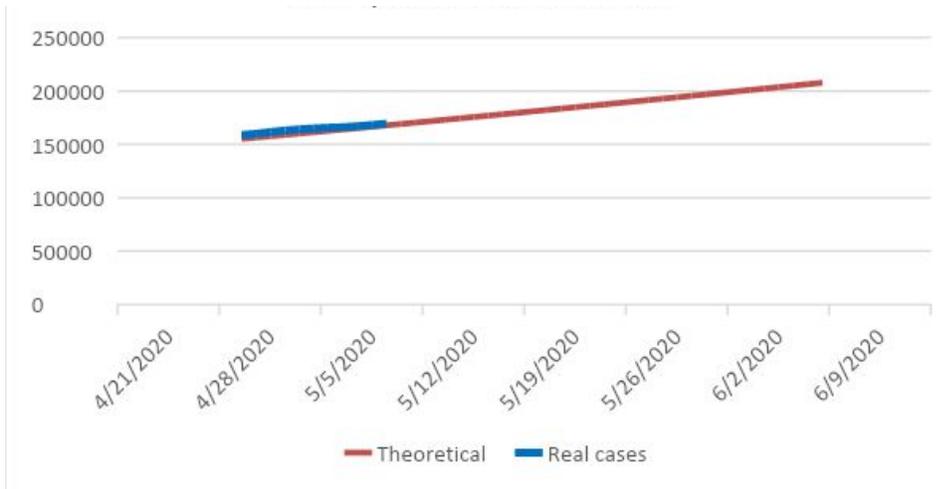


Figure 37. 4th Sigmoid model for Germany with quarantine measures

The final model was very close to real cases, which means that this model was reasonable and could predict the number of infected people in the future periods. The team compared the current graph to the previous one, and saw that the application of quarantine measures significantly decreased the number of infected. In the previous model it was almost 400,000, but

here it was 200000 people. From which it can be concluded that quarantine had a significant effect on the number of infected.

Appendix B: Website

The deliverable for this project was a website that was created using Wix Site Developer and linked to the WPI Global Labs' webpage for COVID-19 related IQP projects. The website was created in collaboration with the Financial University students and serves as a summary of the work done during this project. Highlighted on the site are three models, a comparison of the six countries researched, the team's recommendations for preventing virus spread, and a contributions page to share a little bit about each team member.

Accessibility

The project website can be accessed from the WPI Global Labs' Wordpress page (Figure 38), created specifically to display all of the D-term IQP projects that researched some aspects of the COVID-19 pandemic. The abstract can be found on the Wordpress page in addition to the project name and photo. The Wixsite's URL—<https://covid19iqpsite.wixsite.com/2020>—is linked to these items so that when a site visitor clicks on them, they are transported to the project website.



Figure 38: Global Lab's Wordpress web page providing a link to the team's website.

Homepage

The website homepage shares this project's title as well the nature of collaboration between WPI and the Financial University (Figure 39). Also included on this page, but not shown in Figure 21 is the inclusion of a brief description of what an IQP entails, as well as the project's abstract.

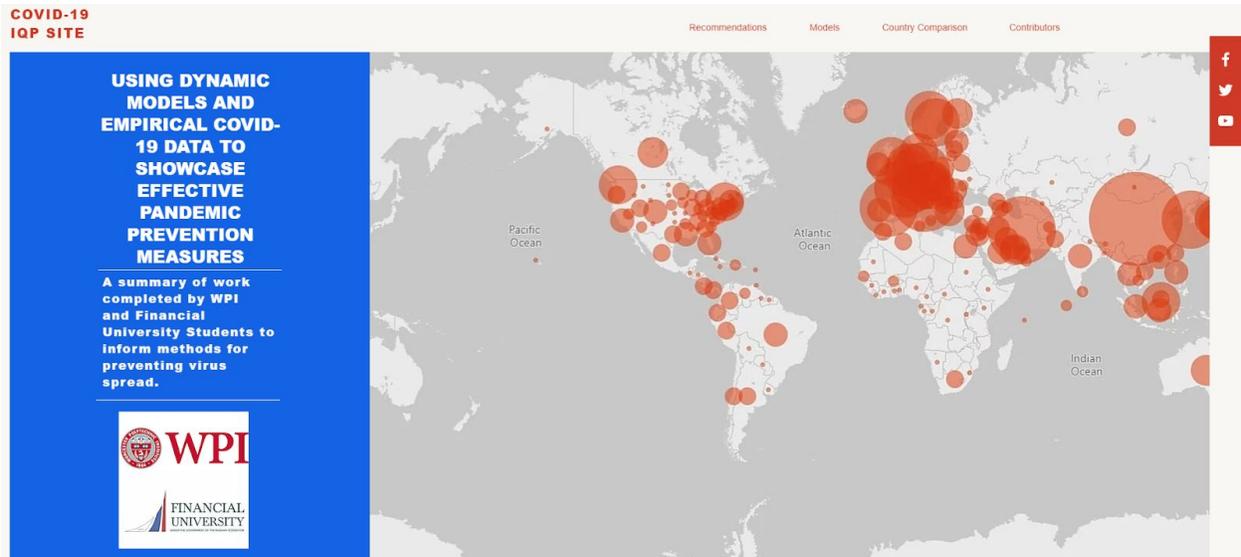


Figure 39: Website homepage with project title.

Contributors Page

This page (Figure 40) is designed to showcase all contributors of the project and share a little bit about each member as well as the advisers who oversaw the project. Per the ideas of the Financial University students and Professor Nikitina, each team member has a headshot featured on this page as well as a brief biography. This page emphasizes that the project was completed as a joint effort between the United States and Russia despite being done remotely.

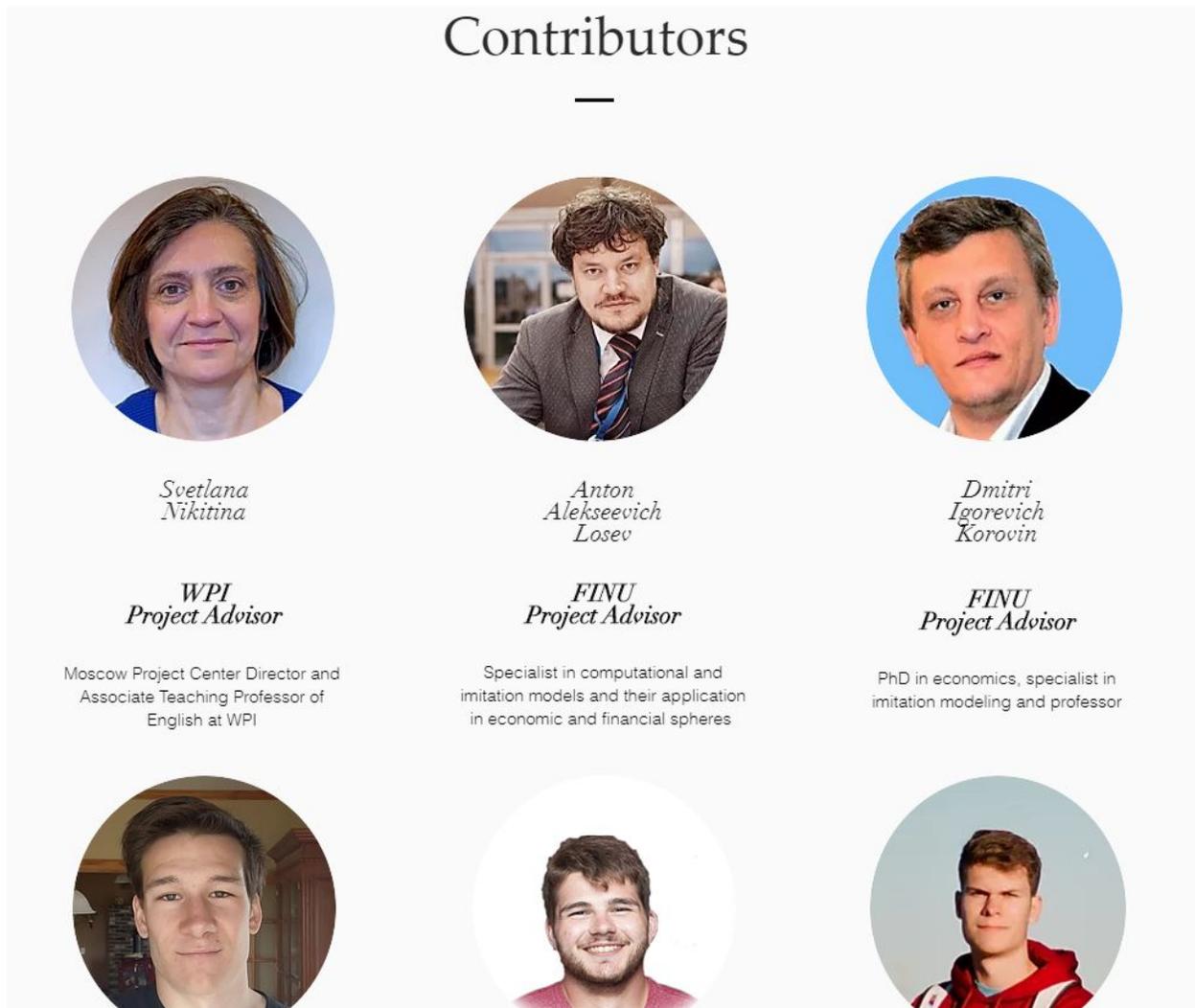


Figure 40: Contributors Page

Models Page

The models page (Figure 41) of the website contains a documentation of the models created during this project. The models are each displayed on their own individual pages all accessible from the models page. While the system dynamic models are only displayed as screen shots, the robust agent-based town model is linked through the AnyLogic Cloud. This link allows users to interact with the parameters of the model and run their own simulations.

Models

The Models page of the website contains descriptions of all the models that the project team worked with. The earlier models are less advanced and not as complex as ones that were used by the team later in the project work, but all are able to be useful for one to understand dynamic modeling better.

Model #6: First Agent-Based

For the team's sixth model a new approach was used. At first the models were built around a population as a whole, but agent...

1 view Write a comment

System Dynamics Diagram:

The diagram shows a complex system dynamics model. It features several stocks (H, I, D, R) and flows (flow, flow1, flow2, recovered). Key variables include TotPop, Contact_Rate, SelfIsol, and b. The diagram illustrates the interactions between these variables, showing how they influence the system's behavior over time.

Figure 41: Models Page of Website

Recommendations Page

The recommendations page (Figure 42) showcases the short list of response protocols that the team deemed effective in preventing virus spread during a pandemic. Sharing these recommendations concisely on the website makes them more readily available to the public and increases the likelihood that they are shared broadly in society. The goal in developing these recommendations was to help governments and individual citizens combat a pandemic together using a common set of institutionalized guidelines.

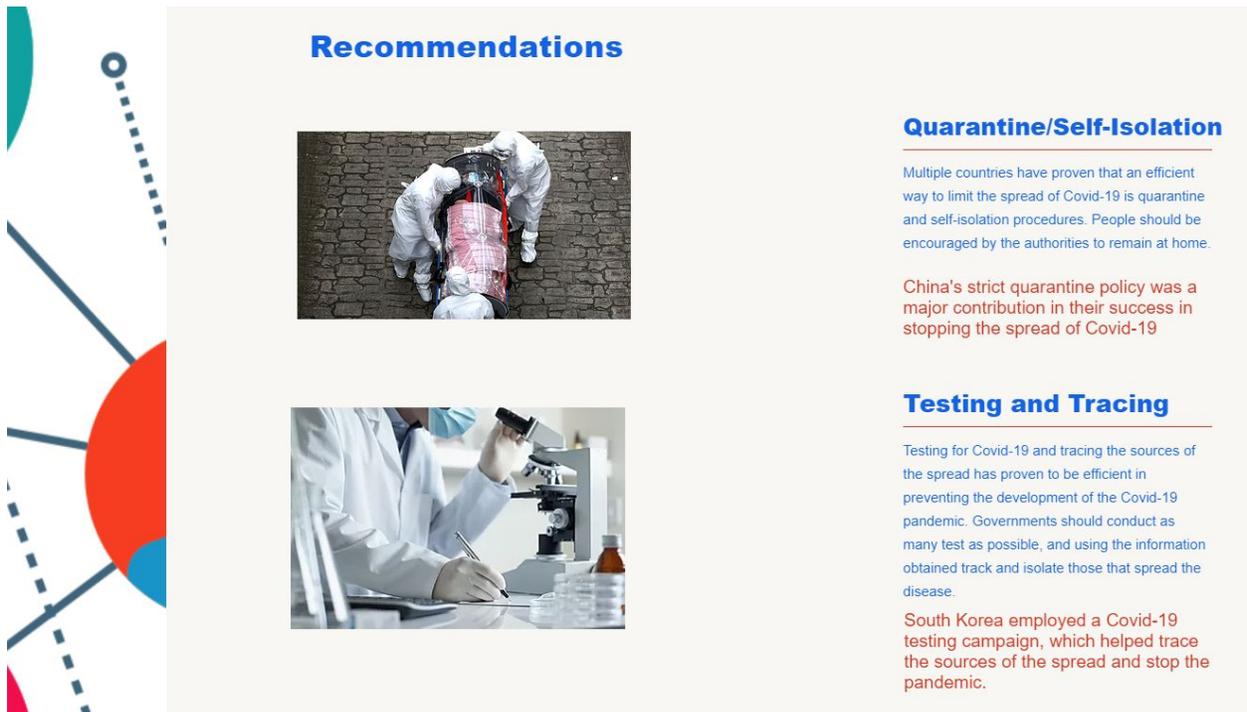


Figure 42: Recommendations Page

Country Comparison Page

A significant portion of this project was dedicated towards comparing the response protocols of six major countries as well as the number of COVID-19 infections and deaths seen in those countries. In light of this fact, a comparison page (Figure 43) is included on the website to highlight the main differences in how each country combatted the virus and how their efforts paid off in terms of spread mitigation.

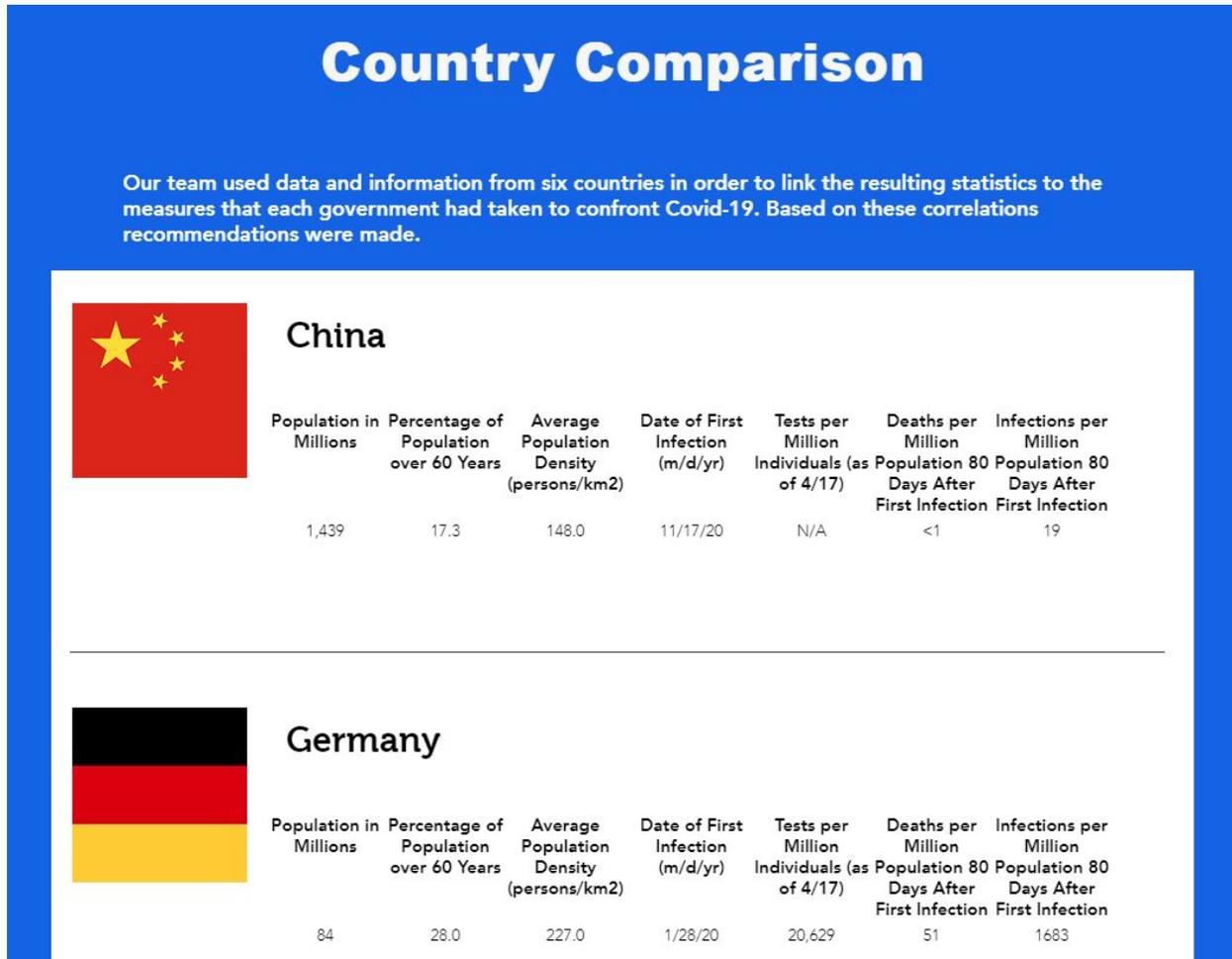


Figure 43: Country Comparison Page