

Virus Spread Simulation Using the SIR Model

David Clark

MH04

Arkansas School for Mathematics, Sciences, and the Arts

International Science and Engineering Fair 2021

February 17, 2021

Introduction

“An estimated 300 million people died from smallpox in the 20th century alone” is a terrifying statistic and even more so when taken into account that smallpox is a single virus among countless others; many of them having the possibility to cause disease, including death, among humans.³ While their existence may be hard and costly to eradicate fully, understanding how they spread and what factors affect this most can save countless lives and be done with minimal resources. That’s the focus of the project described in this paper, simulating accurate virus spread using the SIR model and the factors that affect it.

The SIR model stands for simulated, infected, and recovered.¹ The idea of applying mathematic models to epidemiology was brought about by Bernoulli in 1760 and later Kermack and McKendrick in 1927 furthered the framework; however, the SIR model is credited to Sir Ronald Ross and a few others in the early 20th century.¹ The SIR model can be simulated multiple ways, and at points can be boiled down to a few equations to find extremes in virus outbreaks, but in general a SIR model starts with an index case, which is the first introduction of an infected individual, that can spread it to other susceptible individuals and they can go on to spread the virus more and cause a chain of infections.⁶

In contrast to speaking about past research, further research regarding the study of viruses, how they spread, and how simple simulations could save millions of lives one would have to look no further than the current sars-covid-19 pandemic. A research paper by Dmitry Ivanov shows that computer simulations of viruses are used quite frequently and even goes on to prove their effectivity by demonstrating a fairly accurate one.⁴ It’s quite clear how data like this can be incredibly useful; projections like this could give hospitals a rough estimate of what patient load to expect, the quantity of medicine needed, and the space required.

Even more relevant, projections like this can show the necessary requirement of the general population's reaction to a virus. For example, if 2 projections were estimated, one where people wore masks and social distanced and another where people didn't, it could show how necessary or unnecessary practices like this are. However, these estimations are based on past research which already achieves the goal of seeing how important a population's reaction is, it just puts it into perspective. Research goes on to say, that at least for the sars-covid-19 virus, "along with preventing someone from transmitting the coronavirus, a range of new research shows that the risk of infection to the wearer is decreased by 65%" and similar statistics can be expected when looking at air born or saliva dependent spread viruses.⁵

Other factors than just the amount of susceptible, infected, and recovered individuals effect a population. For instance, certain occupations and areas will be subject to a higher risk of virus spread rather than just an entire population being at risk equally. This can be seen in a study from 2015 that stated "02% of Guinea's population had died due to Ebola, compared with 45% of the country's doctors, nurses, and midwives. In Liberia and Sierra Leone, the differences are more dramatic, with 11% and 06% of the general population killed by Ebola versus 07% of the health-care workers in Liberia, and 85% in Sierra Leone ... health-care workers are at greater risk of contracting Ebola".² The SIR model fails to take factors like occupation, age, or even general health of the infected population which offers fair reasoning to any inaccuracies from its predictions.

The goal of the described experiment is to yield realistic data that can mirror actual viruses and to see if enough change in any virus condition can lower a virus's spread. It was hypothesized that the tested independent variables, such as the spread radius, the infection chance, the required recovery period, the use of masks, and the effectiveness of masks and how

all of these affect virus spread; all have the ability to keep at least 50% of the population from becoming infected; however, some tested variables will meet this threshold before others.

This list of goals and hypothesis wouldn't be achievable without actually proceeding with the project and everything that comes with it. This would include how the simulation works, how the experiments were ran, and how the data was collected.

Procedures

Before going into how the procedure and goal of the project within this paper, it would be beneficial to understand how the simulation that this project utilizes works. In a computer simulation, 100 individuals represented by colored triangles wander around in an 100 by 100 pixel square. These individuals can be blue, signifying they are susceptible, red, signifying they are infected, or grey, signifying they are recovered. At the start of each experiment an index case is set, meaning a random individual becomes infected. The infected individual continues to wander among the susceptible individuals and has a chance to spread the virus if they are within the infection radius. If in a frame an infected individual and a susceptible individual are within the infect radius of each other then a calculation based off the simulations values is ran that determines if the virus spreads to that susceptible individual. When an individual becomes infected another calculation can be done to determine how many frames the individual will be infected; after that many frames pass the individual becomes recovered and can no longer spread the virus or become infected again. The calculation for how long a population member is infected is taking a gaussian distribution of the average infect period with at most 40 frames of variation. The simulation runs until there isn't any infected individuals left.

Looking at figures 1,2, and 3 below, these images were taken from a single control data run. Near the beginning, figure 1 was taken as the index case has barely spread to others and few have recovered. Not too far after figure 1, figure 2 was taken as the index case has spread enough to others to cause a large chain reaction and result in many infected individuals. After that, figure 3 was taken as the infected population member's infection period runs out and they become recovered. For more explanation of how a simulation as a whole runs, look to figure 4 below for a simplified flow chart of a simulation run.

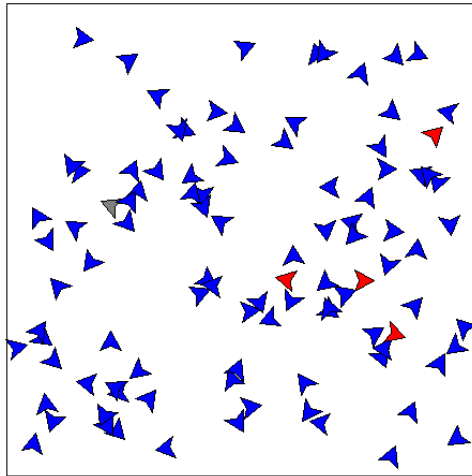


Figure 1. Susceptible Snapshot. Frame with many susceptible individuals taken from the beginning of a control experiment.

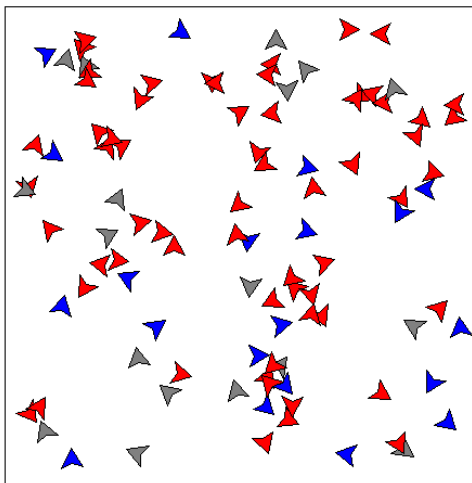


Figure 2. Infected Snapshot. Frame with many infected individuals taken from the about halfway between the beginning and end of a control experiment.

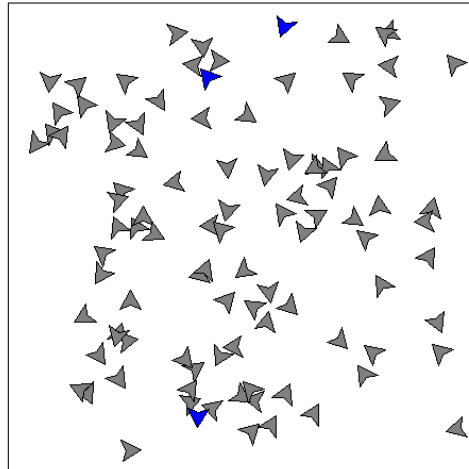


Figure 3. Recovered Snapshot. Frame with many recovered individuals taken near the end of a control experiment.

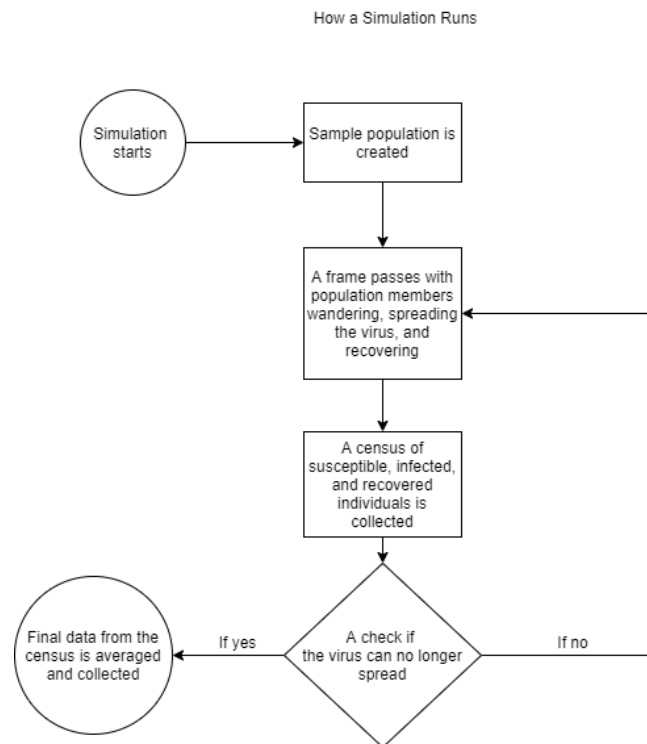


Figure 4. Flowchart of a Simulation. The time line of a single simulation in the build project, for just the control experiment this happened 100 times.

The procedures began with building the used computer simulation while utilizing previous research to have realistic values. Then, a control experiment where no values were changed was ran. After that, 4 experiments were run with each only changing one variable and keeping the other variables the same as the controls. The 4 experiment's passes consisted of them starting with the exact same values as the control experiment, but after that simulation completes a single independent variable was changed by an increment up until a final max or minimum was reached; the control values and the experiment maximums or minimums can be seen in table 1 below. Once all these experiments were completed, the average susceptible, infected, and recovered individuals for each experiment were logged. The 5 experiments completed were then ran 99 times more and their data was averaged overall to avoid sporadic data and to have more accurate results. For example, only looking at the infect radius experiment, it first ran a simulation with an infect radius of 10, an infect chance of 10, an infect period of 240, and with 0 individuals wearing a mask. After that simulation was ran and logged, another simulation was run with an infect radius of 9, an infect chance of 10, an infect period of 240, and 0 masks used. The next simulation would have the same values except the infect radius would be 8, then 7, then 6, and so on until it reached zero where the experiment would be over. This would be done 100 times and all the data for this experiment alone would be averaged and then saved. This process happened to all of the experiments except the control where it was just run 100 times and averaged, there were no changing variables in the control experiment.

	Infect Radius	Infect Chance	Infect Period	Masks Used
Control Value	10	10	240	0
Experiment Max/Min	0	0	60	100
Increment	-1	-1	-10	2

Table 1. Independent Variable Changes. Independent variables and their incrementally changed values used within experimentation.

Data and Results

The following paragraph references figures 5-9 below often. The collected data does meet the hypothesis and isn't out of the ordinary by any means, but the data is still worth while and reinforces ideas already found in previous research. As any of the tested variables decrease, or in the case of masks increase, there is a direct correlation of resulting with less infected individuals. Something to point out is that all the graphs have an aggressive decline in infected people at first, but then flatten out when approaching their maximum/minimum except for the masks experiment, which has a much more static decline of infected people. It would be reasonable to assume that the mask experiment and the infect chance experiment would have nearly identical results as masks directly affected infection chance, but this isn't the case. They do both result in less infected population members, but mask data seem to be more sporadic in their decline and even with an entirely masked population there were still noticeably more infected and recovered individuals than any other experiment, besides the control of course. A possible theory for the sporadic nature of the mask experiment data is that masks depended more on the population movement than the others; another theory for the result in more infected individuals is that masks only provided a -50% of spreading the virus to another person and a -20% chance of spreading when protecting someone, which means that even with a masked infected person exposing a masked susceptible person, the infect chance would be 4, as $10 \times .5 \times .8 = 4$; which is higher than the infect chance minimum which reached 0. Furthermore, one could reasonably assume that infection period would have a more drastic change as the average time it takes to recover and no longer be infectious reduces, but it's downward decline of infected people matches the other experiments well and the increment of 10 frames was randomly selected for simplicity and time sake. In summation, even with all the

minor un-anticipated results, the hypothesis was correct and furthermore the lowered results all had the ability to cause the total infected people to become less than 20%. When looking at this data it should be taken into consideration that the SIR model does have some inaccuracies when emulating real life as people don't wander aimlessly, people have occupations that can be at a higher risk of infection, people's health and ability to fight off viruses can vary, and when people become sick they either distance themselves from others more or feel less able to move about and spread the virus affecting them.

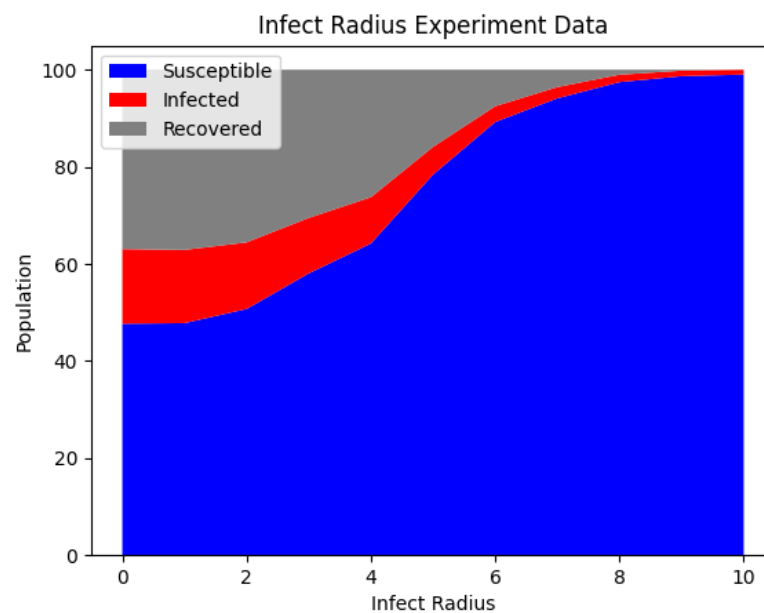


Figure 5. Infect Radius Experiment Data. Data collected from the infect radius experiment. The experiment ran a simulation for each variable condition, which has 10 different variable conditions all changing the infect radius of the simulation. Those 10 simulations were run 100 times and averaged causing this data to consist of 1,000 simulations.

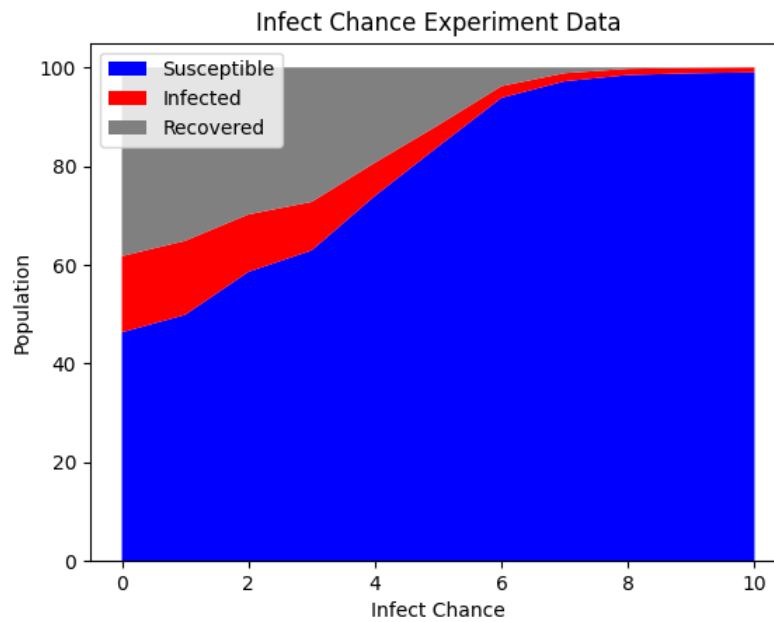


Figure 6. Infect Chance Experiment Data. Data collected from the infect chance experiment. The experiment ran a simulation for each variable condition, which has 10 different variable conditions all changing the infect chance of the simulation. Those 10 simulations were run 100 times and averaged causing this data to consist of 1,000 simulations.

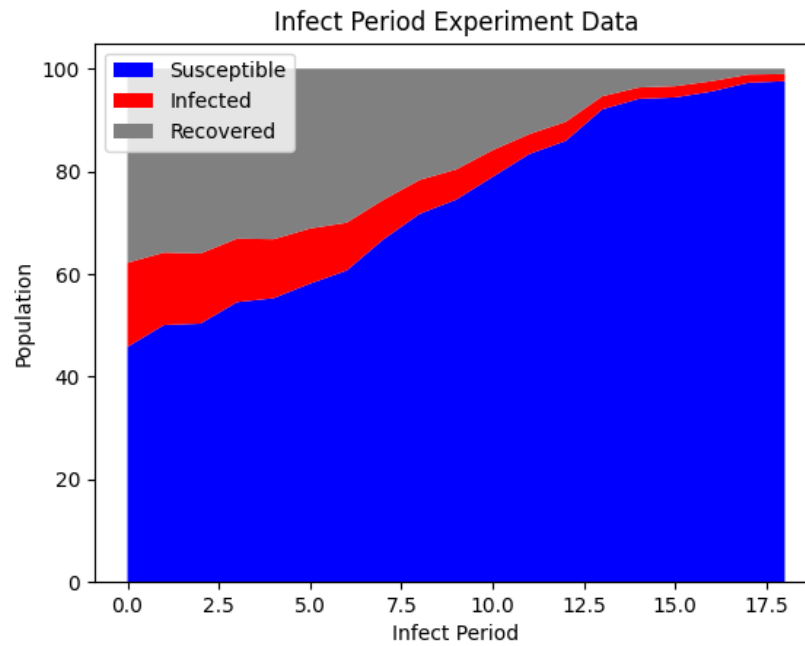


Figure 7. Infect Period Experiment Data. Data collected from the infect period experiment. The experiment ran a simulation for each variable condition, which has 18 different variable conditions, as the infect period starts at the control value of 240 and goes to 60 by increment of 10. Those 18 simulations were run 100 times and averaged causing this data to consist of 1,800 simulations.

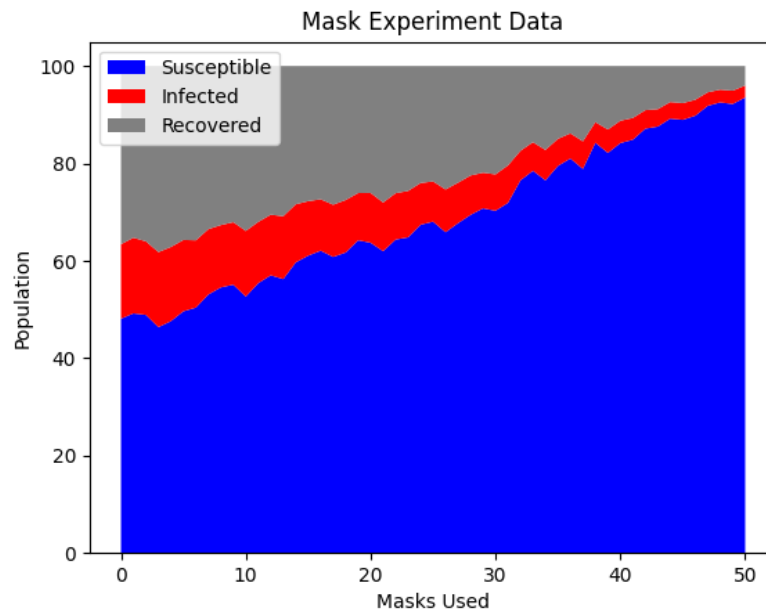


Figure 8. Mask Experiment Data. Data collected from the mask experiment. The experiment ran a simulation for each variable condition, which has 50 different variable conditions, as the number of masks used starts at the control value of 0 and goes to 100 by increment of 2. Those 50 simulations were 100 times and averaged causing this data to consist of 5,000 simulations.

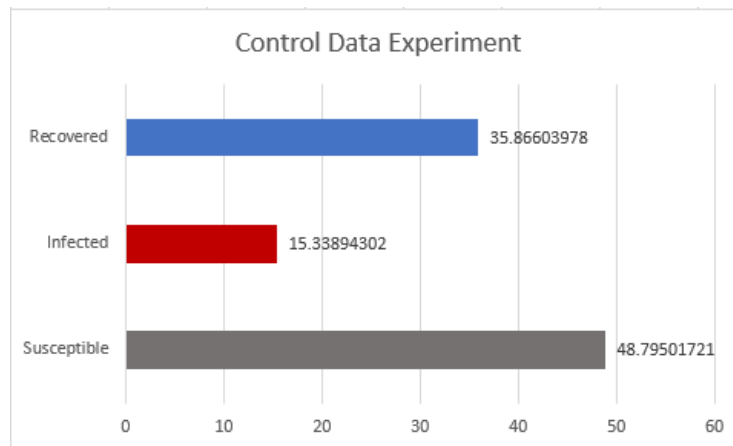


Figure 9. Control Data Experiment. Data from the control experiment. No variables were changed in this experiment, so it was run once to collect its data and then 99 more times and averaged to avoid sporadic data. The added values from this data don't exactly equal 100, the sample population, due to averaging.

Conclusion

In conclusion of the project, the SIR model can be modeled many ways and boiled down to a few simple equations, but in general consist of a sample population, an index case, and a census of the susceptible, infected, and recovered populations.⁶ The idea of simulating virus spread is a common practice in today's world and has even been done with the recent sars-covid-19 pandemic.¹ Even though viruses can vary, masks still have evidence behind them declaring that they do prevent virus spread and have at least a 65% chance reduction in spreading or contracting the sars-covid-19 virus.⁵ Certain aspects, such as occupation, affect a person's chances of being infected greatly; for example in an outbreak of Ebola in New Guiana ended fatally for 02% of the population, but 45% of the countries medical personal passed away due to this outbreak, which is strong evidence indicating that because of the situation some one's job can put them can put them at higher risk of infection.²

The procedures of this project consisted of building the simulation used, then running a control experiment, then running 4 experiments with each incrementally changing either the infect radius, the infect chance, the infect period, and the number of masks used; then logging the average census of susceptible, infected, and recovered population members. The 5 experiments are then run 100 times and their collected data is averaged to avoid sporadic data.

The collected data found that the hypothesis that at least 50% of the sample population would remain susceptible was achieved and that with enough change of any of the independent variables tested 80% of the population could remain susceptible by the end of the simulation, meaning they were never infected. Out of all the independent variables tested, such as infect radius, infect chance, infect period, and masks used; all had rather steep decline in infected individuals as the independent variable changed, except for the use of masks, which had a rather

constant decline of infected people. This data proves previous research on the importance of key variables of viruses and how a population should react to a virus if they want to minimize spread. However, the SIR model used does have some inaccuracies as the population members in this experiment wandered aimlessly, all had the same health, and continued to wander when sick. This doesn't mirror real life as people don't wander aimlessly, some people's occupations are more at risk to infection to certain viruses, people's health and ability to fight off viruses are different, and generally when someone becomes sick they try to isolate themselves more or at least feel less capable of moving about and therefore can spread the virus less.

Even considering these inaccuracies, countless lives could be saved from these projections. Hospitals and governmental bodies could use this data to try and save as many lives as possible as they could predict how many patients to expect, medicine needed, rooms needed, etc. Furthermore, this data could resolve conflict about how much a population should react to a virus as masks showed to be effective, just not as much as a smaller spread radius or infection chance, but these variables are out of everyone's concern it simply is a reality.

Possible future research could try to account for more variables affecting virus spread rather than the 4 tested. Furthermore, the inaccuracies listed earlier, such as failing to incorporate a population members occupation, their health, their movement as it is random in this project, or any attempt of distancing from sick individuals; could strengthen the provided data or at the minimum provide more data to talk about rather than what has already been heavily covered in previous research.

Bibliography

1. D'ebarre, Florence. "SIR Models of Epidemics." SIR models of epidemics – Theoretical Biology | ETH Zurich, 2019.
<https://tb.ethz.ch/education/learningmaterials/modelingcourse/level-1-modules/SIR.html>.
2. Evans, David K, Markus Goldstein, and Anna Popova. "Health-Care Worker Mortality and the Legacy of the Ebola Epidemic." The Lancet Global Health, July 8, 2015.
[https://www.thelancet.com/journals/langlo/article/PIIS2214109X\(15\)00065-0/fulltext](https://www.thelancet.com/journals/langlo/article/PIIS2214109X(15)00065-0/fulltext).
3. Flight, Colette. "History - British History in Depth: Smallpox: Eradicating the Scourge." BBC. BBC, February 17, 2011.
https://www.bbc.co.uk/history/british/empire_seapower/smallpox_01.shtml.
4. Ivanov, Dmitry. "Predicting the Impacts of Epidemic Outbreaks on Global Supply Chains: A Simulation-Based Analysis on the Coronavirus Outbreak (COVID-19/SARS-CoV-2) Case." Transportation Research Part E: Logistics and Transportation Review. Pergamon, March 24, 2020.
<https://www.sciencedirect.com/science/article/abs/pii/S1366554520304300>.
5. UC Davis Health, Public Affairs and Marketing. "UC Davis Experts: Science Says Wearing Masks and Social Distancing Slow COVID-19 (VIDEO)." UC Davis Health, July 6, 2020.
<https://health.ucdavis.edu/health-news/newsroom/uc-davis-experts-science-says-wearing-masks-and-social-distancing-slow-covid-19/2020/07>.

6. Weiss, Howard. "The SIR Model and the Foundations of Public Health." Barcelona :
Departament de Matemàtiques de la Universitat Autònoma de Barcelona, 2013.