Ryan Burczak
Independent Project: Data Modeling
Summer 2020

COVID-19 Pandemic Simulator

During the months of May and June 2020, I worked on a project to simulate the spread of a pandemic within a population. It's a relevant project because the novel coronavirus (SARS-CoV-2), which causes COVID-19, entered the US earlier this year. COVID-19 is now a global pandemic.

What does the simulator do?

This simulator models a hypothetical microcosm of the public, with people moving around within a defined area and engaging in their everyday life. A small percentage of the population is initially infected with a highly infectious virus. As the people, represented by dots, move around and interact with one another, the viral infection spreads. When a person is infected, they stay in place for 14 days and, based on their health, either recover or die before the recovery period is finished. The total number of infections, recoveries and deaths are tallied and recorded every day.

A healthy person is represented by a green dot. A person infected with the virus is represented by a red dot. A person who has recovered from the coronavirus is represented by a yellow dot. A person who has died is represented by a black dot.

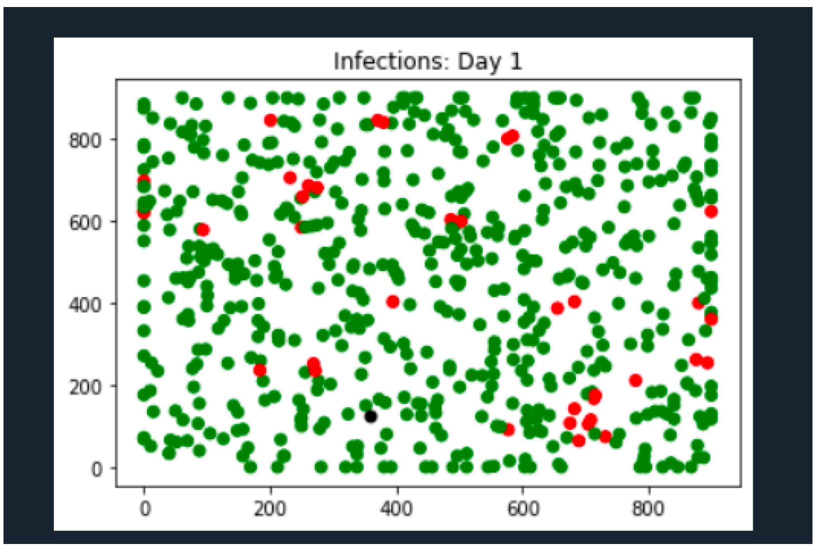
The user is allowed to set parameters, such as the dimensions of the field of study, the number of people in the study, the length (in days) of the study, the radius of infection (how far can this virus travel from one person), how far people can move in a day, etc. For the people, 85% of them move "locally," or within a maximum distance set by the user. The other 15% move "globally," or at most a particular percentage of the dimension(s) of the field of study. The user is not allowed to change constants; for example, the number of days a person roams around asymptomatic before stopping in place (2 days), and the total time to recovery (14 days).

Social distancing guidelines are not considered for this project. Therefore, healthy dots (green and yellow) do not stay far away from infected dots (red). Additionally, it has been recently discovered that just because someone recovers from the virus and develops immunity does not mean they cannot become re-infected. This has not been taken into account yet; as a result, the yellow dots cannot become red again.

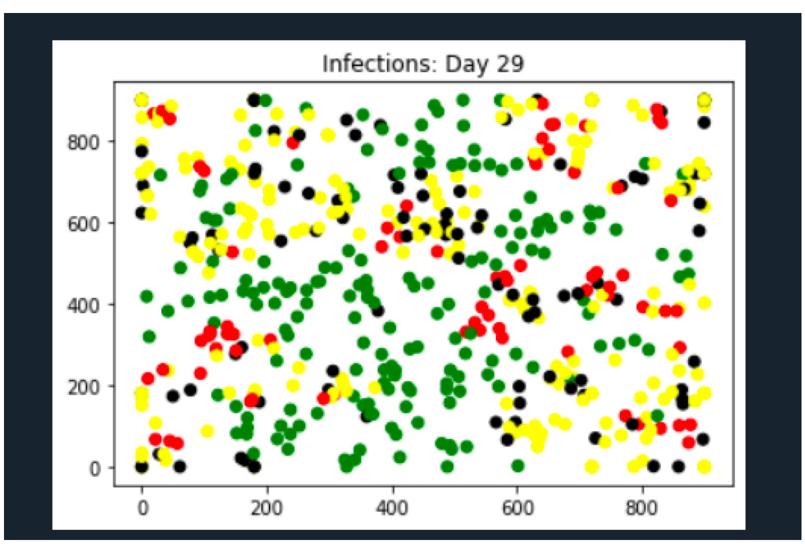
Sample Results

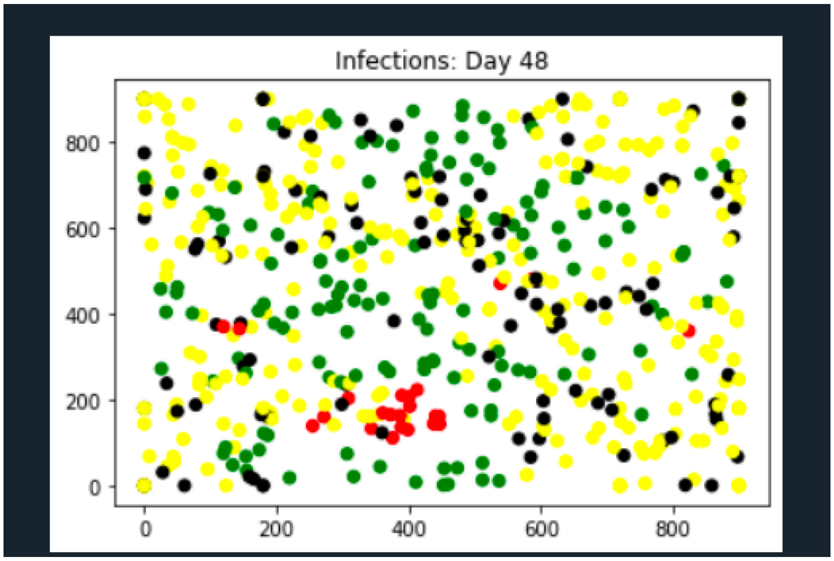
The images below illustrate one simulation of the spread of a virus over 100 days. Parameters used in this simulation are as follows:

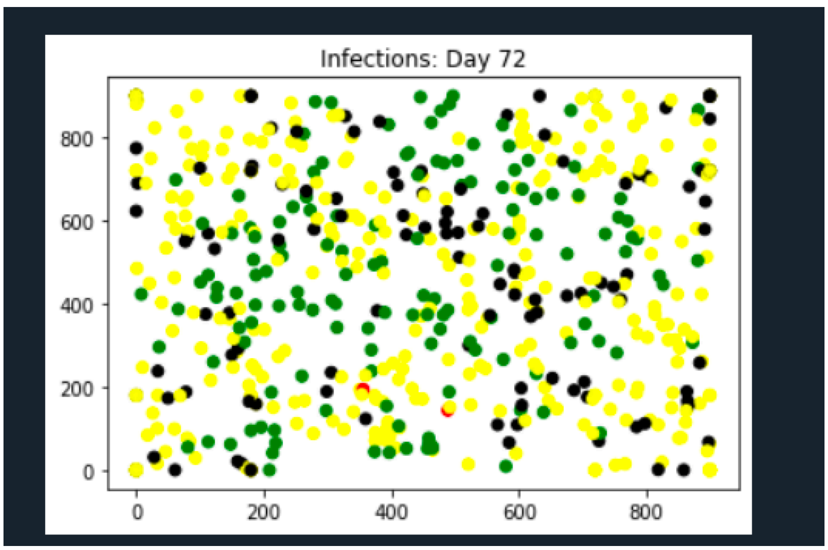
- Period of infection = 14 days
- Days before an infected person quarantines (when they're asymptomatic) = 2 days
- To start, roughly 4% of the people have coronavirus
- If anyone comes within 20 units of an infected person, they get infected too.
- There are 600 people on a 900x900 grid
- The simulation happens over 100 days
- People's state of health is decided by a normal distribution curve; their health quality (from 1 being the absolute worst and 100 being the absolute best) determines their likelihood of dying from the virus over their 14-day sick period.

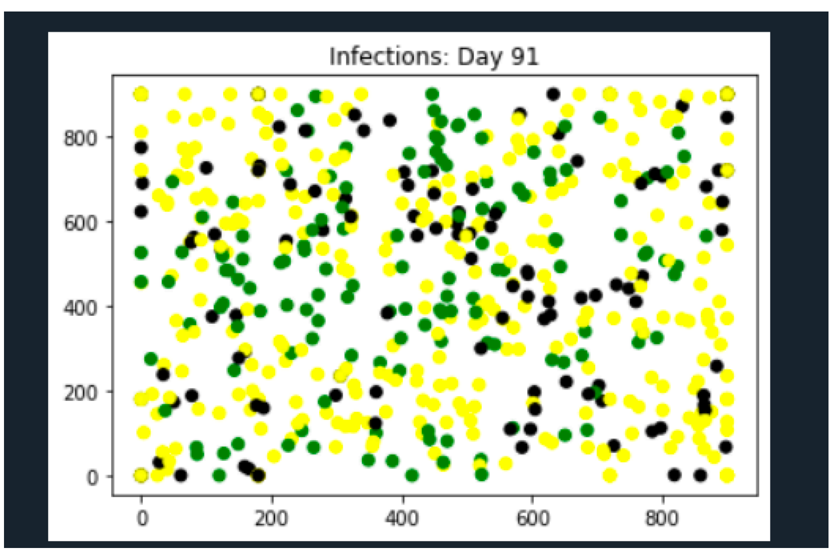


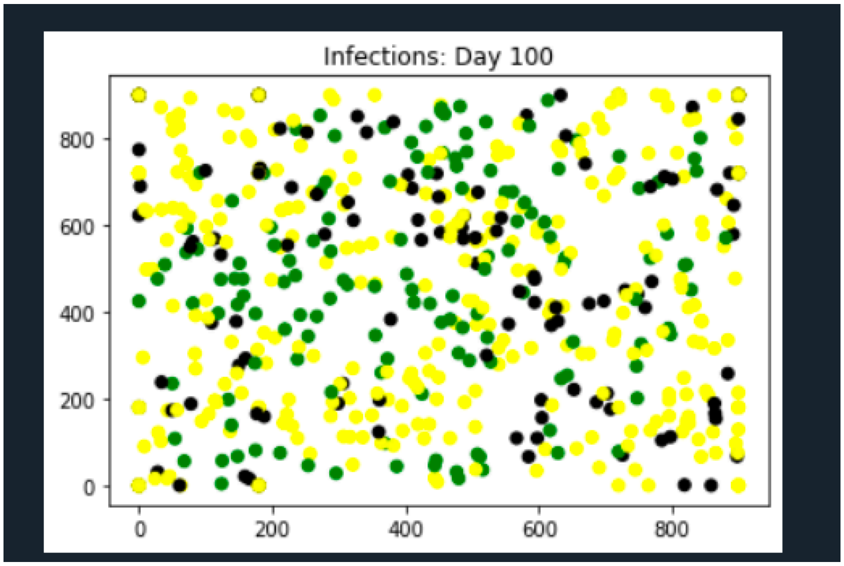
Green – healthy   Red – infected   Yellow – recovered   Black – dead




Infections: Day 48
0
200
400
600
800
0
200
400
600
800


Infections: Day 72
0
200
400
600
800
0
200
400
600
800


Infections: Day 91
0
200
400
600
800
0
200
400
600
800



What did I get out of this project?
I wrote the simulation in Python because of its expanding use in Data Science. The code was written in a program called Spyder, which allows for easy editing and running of Python code. In taking on this project, I learned how to code animation, which is an important skill for modeling and simulating data. This program simply envisions a hypothetical scenario; it does not take in any concrete data.

Next Steps
The inclusion of real-time data is a logical next step to extend this project. There is much more data on the spread of the COVID-19 virus available now that wasn't available when I began this project in May. The Johns Hopkins University Covid-19 Daily Reports database is available on GitHub. With this data, the model can be improved by using realistic data for variables such as rates of infection, recovery and death. Additional information, such as incorporating social distancing guidelines, reduced infection transmission due to mask usage, reinfection rates, and the presence of super-spreaders within the population can also be incorporated.

Acknowledgements
Thanks to Dr. Shrirang Karandikar of AlgoAsylum for mentoring me throughout this project.