PHYS129L FINAL PROJECT

COVID-19 Epidemic Simulator

Stanley Norris

Department of Physics

UCSB

(Dated: June 9, 2020)

Abstract

The program written for this final project is an epidemic simulator which is based off the SIR model developed by Kermack and McKendrick. The epidemic simulator will be applied to 100 of the most populous cities and towns in Great Britain, and the spread of the disease in Great Britain will be tracked for all of these cities. The SIR model in this program is essentially a set of three (ordinary) differential equations for each place which splits the total population into three categories corresponding to S, I and R (Susceptible, Infected and Recovered). The simulator allows the user to input a few parameters which will be applied to these differential equations, such as the infection rate and the recovery rate. These rates will then be used on each city in Great Britain, where the infection rate describes how fast individuals in a population go from S to I and the recovery rate describes how fast individuals go from I to R. The program will solve the equations, then calculate the most optimised location of a hand-sanitizing manufacturing facility based on the data of the cities in Great Britain. The user is also given an option to choose a port from which to receive the ethanol required in the fabrication of hand sanitizer. The Great Britain data is acquired online, and has the population and coordinates of the places in the file. The program will then begin an animation displaying the location of the factory on the map of Great Britain, as well as red dots whose size is proportional to the number of infected people at that location. The model will be ran over 250 days. There is no hardware or software required to run the Python program written for this final project that isn't included in the standard Raspberry Pi updated for Physics 129. This is a feature that was esteemed valuable in the production of the program as accessibility is key to the impact that any program will have on a broader audience. The program does however require an internet connection, which is deemed reasonable as the user wouldn't even have access to the program without it.

I. RESULTS

All that is required to produce the correct output is to have a working Internet connection and to run the file in Python. The first action in the program is to fetch an image of the map of Great Britain as well as the data file named "GBPlaces.csv". The map of Great Britain was produced by installing a library called Basemap, and producing the map by using Basemap for python2.7. It was difficult indeed to install Basemap, and only worked with Python2.7, for this reason the image was uploaded online to be used by the main program. The data wasn't subject to these requirements, but is also fetched online for convenience. These files were uploaded to a GitHub repository created specifically for this final project, and demonstrates the power of being able to access and edit projects online from the command line.

Once the file is run, the user will first be prompted to input the infection rate. The infection rate will determine how fast the people of Great Britain will infected, and must be between 0 and 1, same applies for the recovery rate except that it defines how fast a population recovers from the disease. The data file contains the name, type, population, latitude and longitude of the 100 most populated cities in Great Britain. This data is organised in the program such that 1 member of the population for each place is initially infected, and the rest are susceptible to the disease. These are the equations which underpin the SIR model:

$$\frac{dS}{dt} = -\frac{\beta SI}{N} \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

The equations of the SIR model are then applied with these conditions and those inputted by the user, and the number of individuals in each category for all places in the file is tracked. The resulting data is then organised and then a minor calculation is performed on it, the program determines when the number of Infected people in Great Britain reaches its maximum and what that value is. The user is then given the choice of 4 of the biggest ports in Great Britain to partner with in the synthesis of the hand-sanitizer to be shipped to the people of Great Britain. The factory location optimization that follows essentially fulfils the requirement that the factory should be as close as possible to as many infected people as possible as well as the partner port. This can be done in the simplest way by minimising the sum of the product of distance and population over the 100 places in Great Britain. The port is then added with an arbitrary weighting such that it counts as much as a fairly big city. It is important to note that the choice of weighting of the number of infected people, their proximity to the factory and port location are all treated arbitrarily. The program could be written in many different ways, the characteristics of the best factory location can be made to be very complex. The program goes through many iterations trying to find the minimum (it actually then takes 1/minimum [the inverse] and finds a maximum. But this quantity is less intuitive so the minimum will be discussed even though they are equivalent things in this context), it consists of a random location being created somewhere around Great Britain and then the values of the weighting function are checked in surrounding areas to check for local minima. The program then begins another iteration and eventually outputs the location of the place at which the global mimimum is found. The program is merely intended to demonstrate what can be achieved with a small amount of data, and that the list of parameters to optimize is inexhaustible. The program could be re-written fairly easily, with the parameters for the optimization chosen very carefully to produce an applicable result. If there is research commissioned by the UK government as to what the policies should be to contain the next disease, this program could in theory be used. The project was created with the following scenario in mind: The UK government must create a manufacturing facility in order to ship hand-sanitizer to the people of Great Britain, but what is the best location? If the correct data and demands are given then this can be implemented by modifying this program. All that is required for this application is data that precisely allows one to build a weighting function, for example the cost of fuel, manpower, vehicles, packaging, etc... for every possible shipping route. The model could further be tuned by fitting functions to real data of the 2020 COVID-19 outbreak to determine infection rate, mortality rate, recovery rate and the initial number of infected people in each city. After the optimization, an animation is ran which allows the user to visualise the spread of the disease and gives a good intuition of the situation. It is important to give this kind of intuitive live animation for the sake of communication to the layman. Data must be presented clearly and pleasantly to be accessible to as many people as possible, regardless of their previous knowledge in the relevant discipline. The animation features the map of Great Britain with red circles at every location in the data file which have a size that is proportional to the number of infected people at the location.

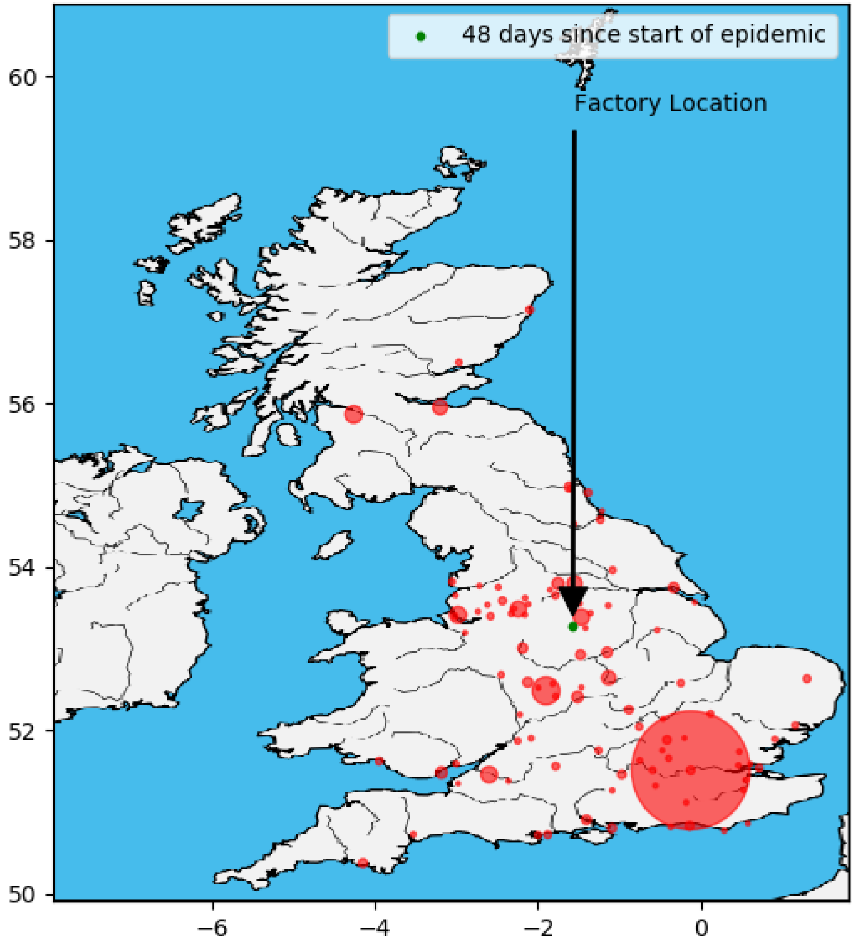

FIG. 1. A screenshot of the animation. $\beta = 0.7$, $\gamma = 0.3$ Port location is London docks, factory has coordinates (53.27919326806212 , -1.570565662125899)

The surprising result is that the factory location isn't further South, for most simulations ran. If the simulation is ran simply on the total population for each city/town rather than the number of infected people at the peak of the disease, the factory always lands in Milton Keynes, which is fairly central but still South compared to to most locations for this configuration. One can see from the data that most places in Great Britain have a population that is at least an order of magnitude smaller than London and Birmingham. This explains why in the case where total population is used for weighting (location of factory is time independent) the factory lands further South, London 'pulls' it towards it more than any other place. The reason that the factory location in Fig.1 is unexpectedly North is because when the virus peaks for Great Britain, London hasn't yet reached its peak. The disease is present for shorter amounts of time in smaller populations, this fact is a product of the equations. When the simulation is ran, it can be seen that the red dots are dominating in the central to Northern regions of Great Britain when the peak of the disease is reached. In order to build a more truthful model, the assumptions made by this model must be removed. For example, this model assumes that at the start of the simulation 1 person in each place is infected. This seems unrealistic, although it is reasonable because the SIR model assumes that the populations are large enough where the equivalent of the thermodynamic limit is taken. The assumption is mildly justified by the fact that if the population is large and mobile, the people will move across the country fast enough that it can be approximated that the disease begins spreading at the same time in all cities.