

# Workbook 1 - Epidemiology

MAM5220 - Statistical Techniques for Computational Biology

Note to see code/images: <https://github.com/sap218/R/tree/master/mam5220/w1>

## Question 2

a)

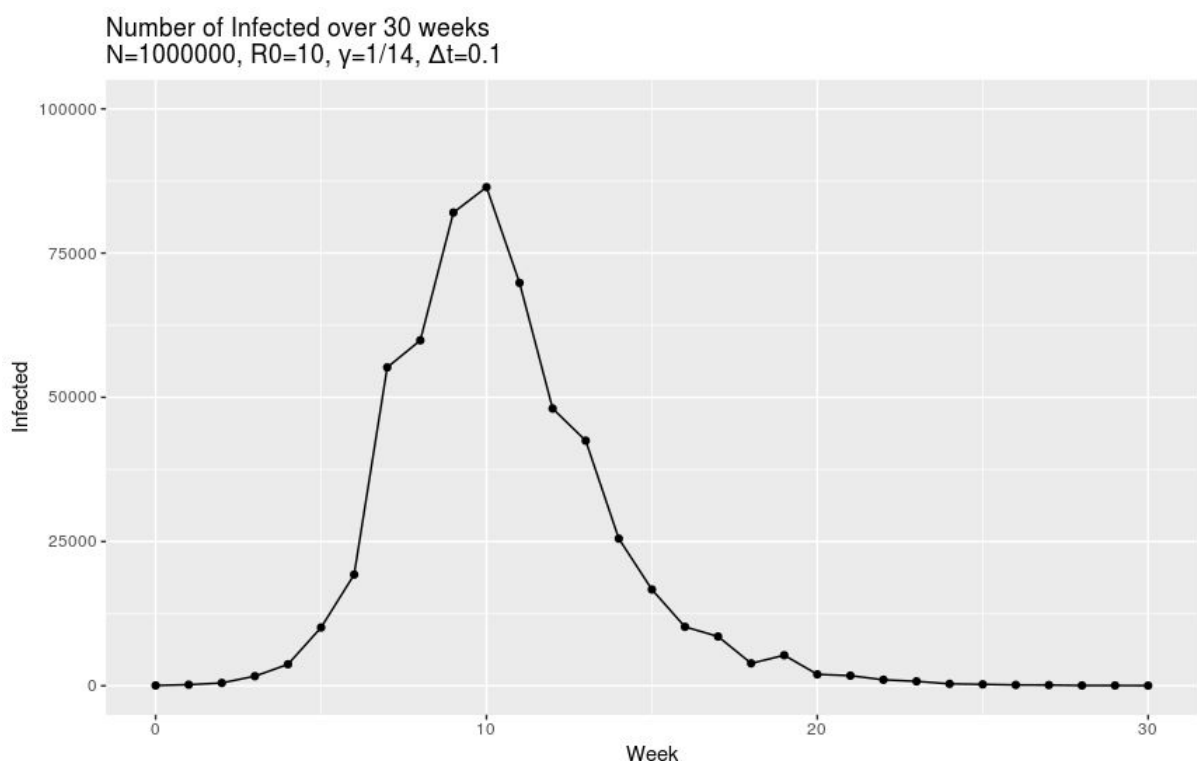


Figure 1: As we can observe, the infected count is exponential until week 10, which is the peak of infection count. The number of infected then decreases until around week 20 when it steadies out. An outlier could include the sudden increase count at week 19: this case stands apart from the overall pattern. The outbreak can be described as “Point Source Outbreak”: the outbreak was a short duration and did not persist over time.<sup>1</sup>

Overall the distribution is approximately bell-curve shape with a fatter right-tail..

<sup>1</sup> [http://www.med.uottawa.ca/sim/data/public\\_health\\_epidemic\\_curves\\_e.htm](http://www.med.uottawa.ca/sim/data/public_health_epidemic_curves_e.htm)

b)

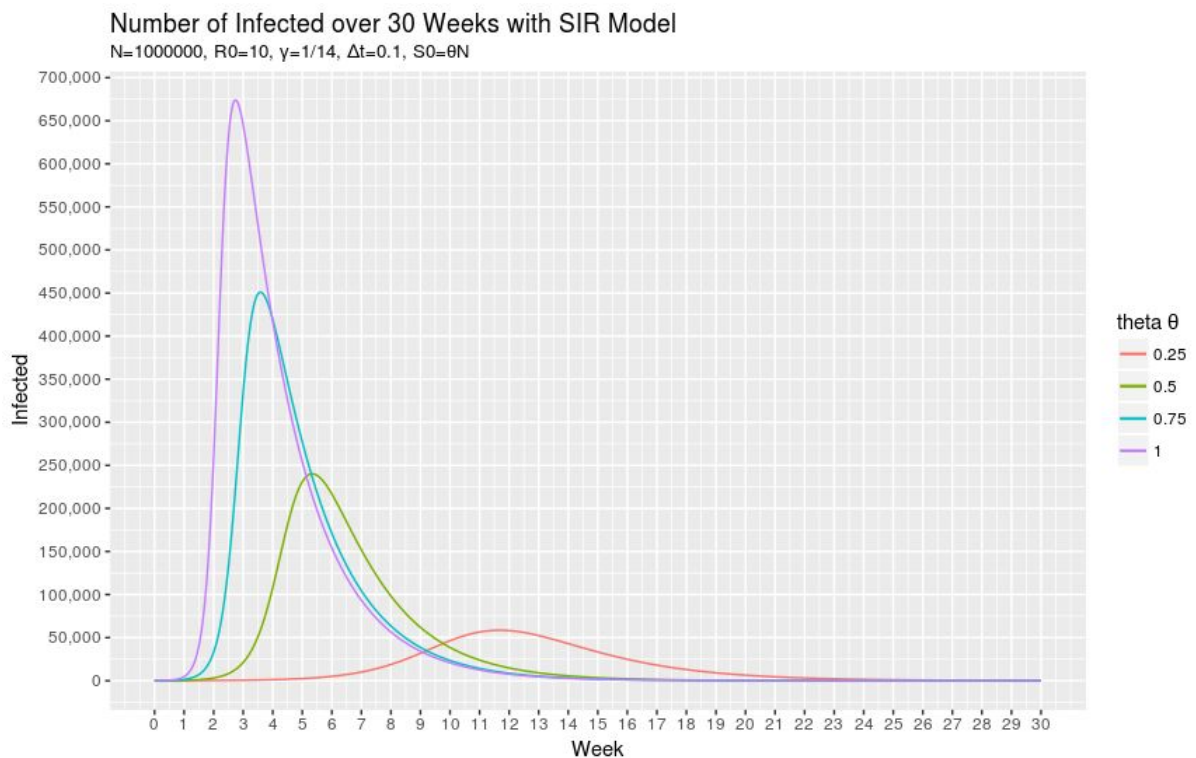


Figure 2: When altering the number of susceptibles by  $S_0=\theta N$  we can see how the infected count differs. As  $\theta$  tends towards 1, we are essentially infecting the whole population, whilst as  $\theta$  tends towards 0, fewer people are infected overall due to the lower number of starting susceptibles.

c)

```
theta <- c(0.25, 0.5, 0.75, 1)
> log(errorSS(epidemiology.data = epidemiology_data, theta = theta[1]))
[1] 23.21358
> log(errorSS(epidemiology.data = epidemiology_data, theta = theta[2]))
[1] 25.97571
> log(errorSS(epidemiology.data = epidemiology_data, theta = theta[3]))
[1] 27.52209
> log(errorSS(epidemiology.data = epidemiology_data, theta = theta[4]))
[1] 28.32531
```

d)

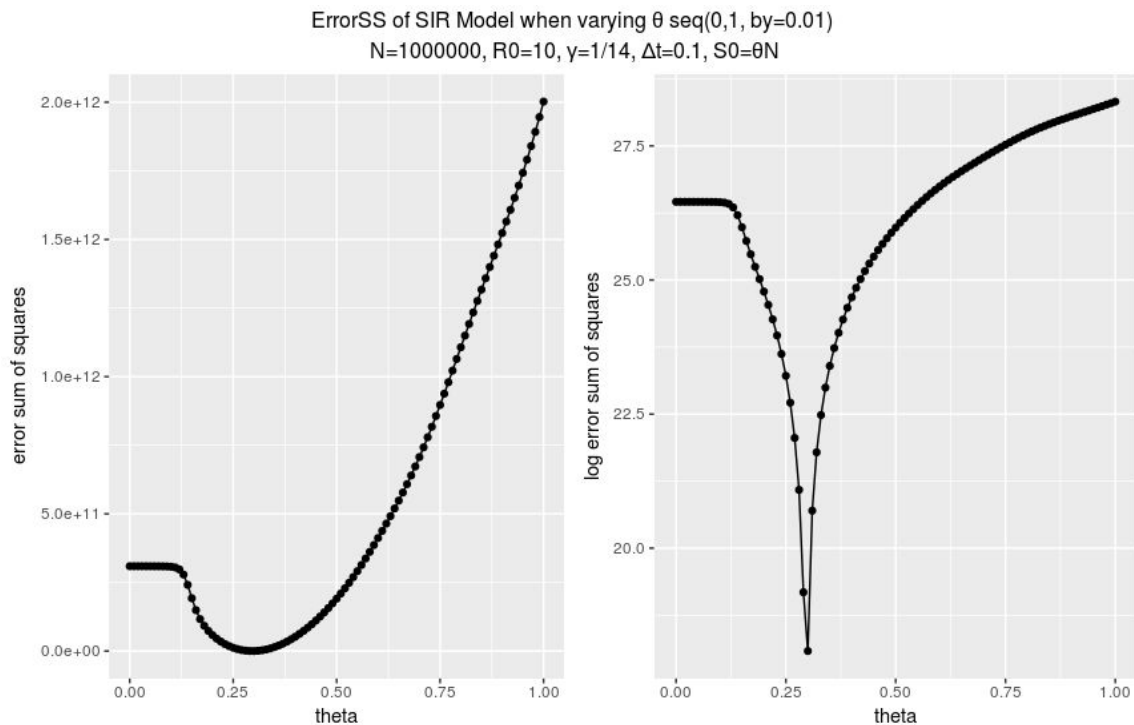


Figure 3: Continuing on from question 2, when extended the SIR model by considering error sum of squares, we can see the lowest 'dip' when log transformed. This shows the optimal value of theta: 0.3 (explained in 2e)

e)

```
> m <- which.min(theta)
> print(m)
[1] 31
> best.theta <- (theta.x[m])
> print(best.theta)
[1] 0.3
> lm <- log(errorSS(epidemiology.data = epidemiology_data, theta = best.theta))
> print(lm)
[1] 18.08245
```

0.3 is the optimal value for theta, so the optimal value for  $S_0 = \theta * N = 300,000$

As seen in Figure 4

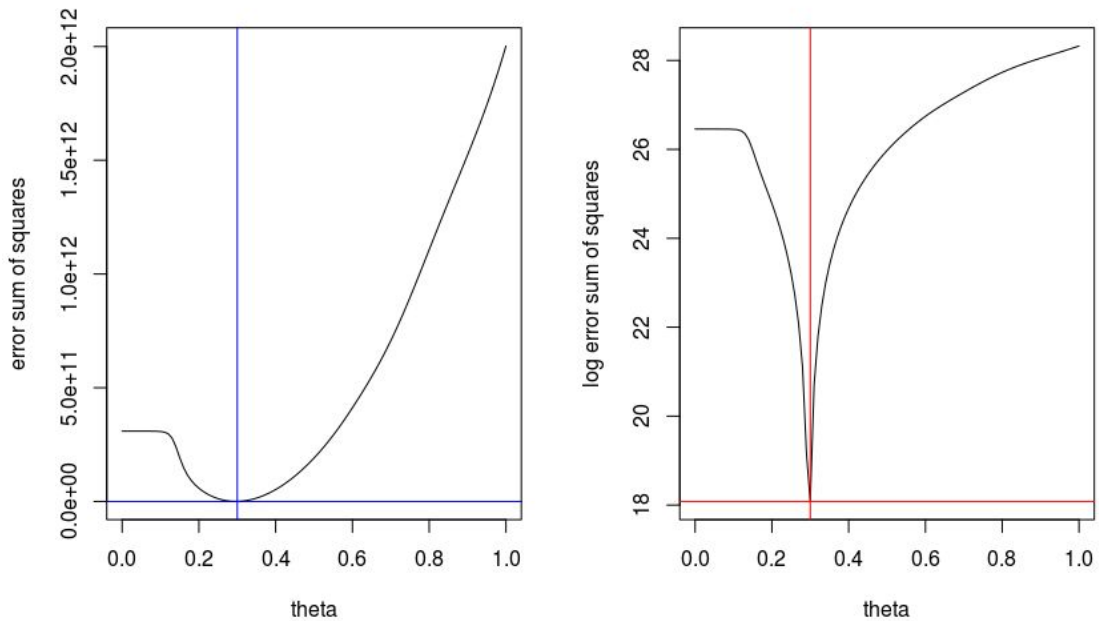


Figure 4: Very similar to figure 3, yet lines are plotted thinner and the optimal value of theta is much clear with the vertical and horizontal lines plotted.

f)

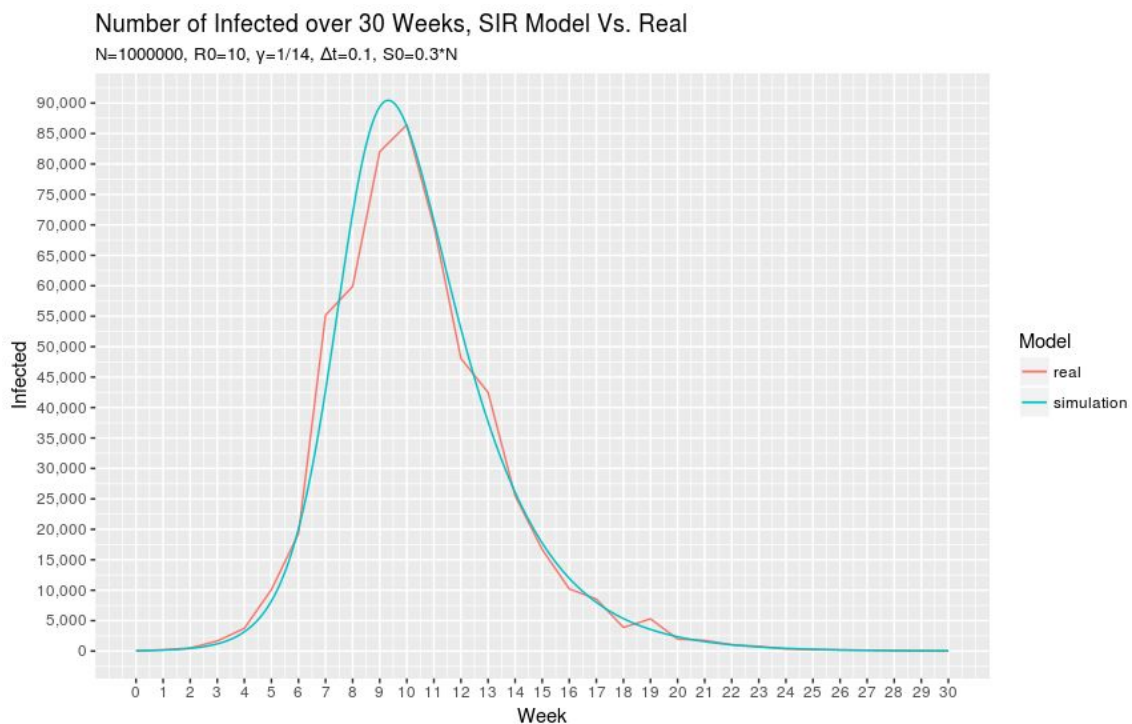


Figure 5: We can observe the real data and the simulated information plotted together. By using 0.3 as the optimal value for theta, we can plot a somewhat accurate curve.

## Question 3

a)

```
> beta <- R0*gamma/N  
[1] 4e-07
```

b)

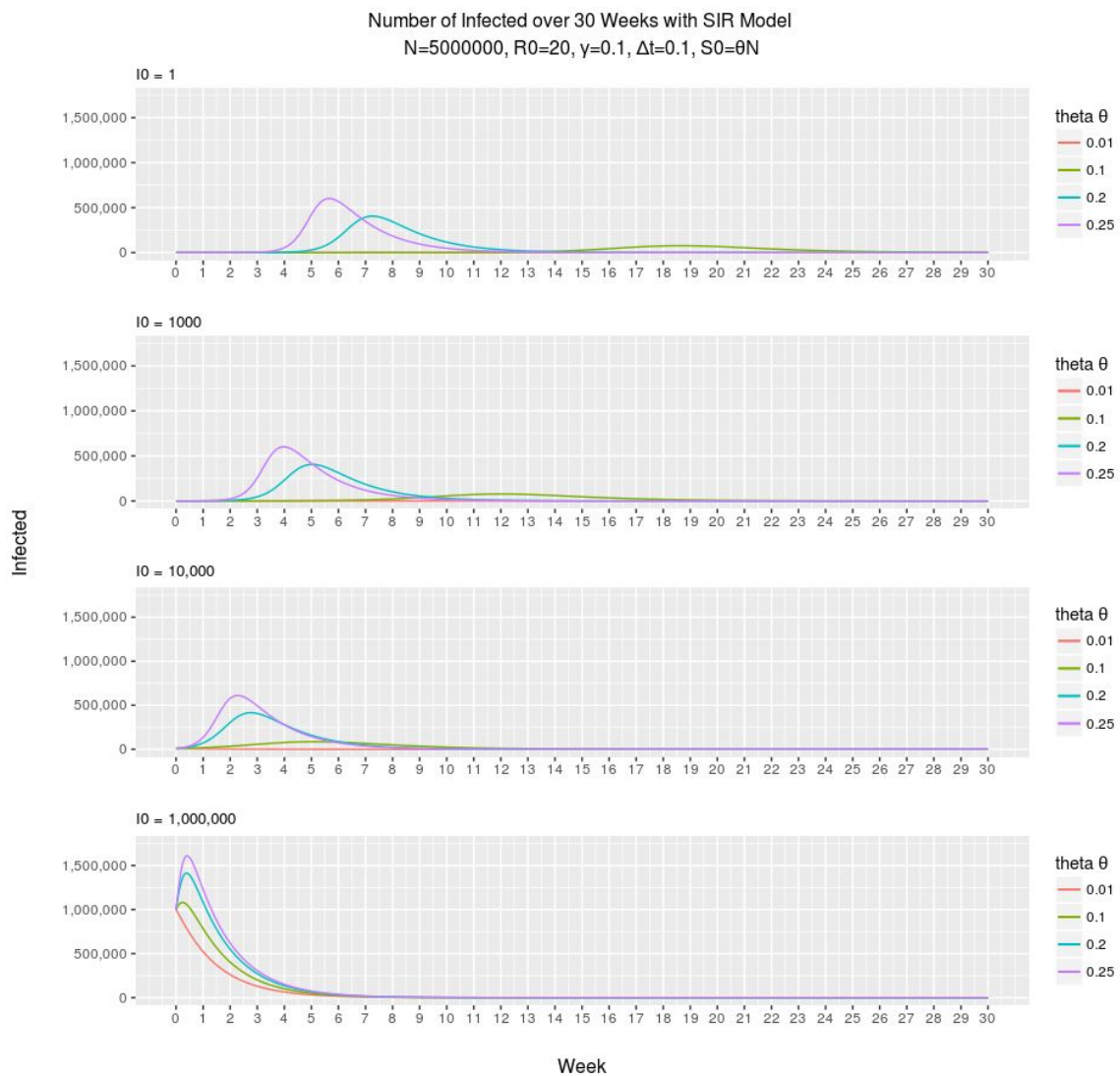


Figure 6: In these plots, we are investigating differences when we change the initial number of infected ( $I_0$ ) plus altering theta. The higher  $I_0$  and higher theta, the plot curves have a positive skew (thicker right-tail): a Leptokurtic distribution<sup>2</sup>.

<sup>2</sup> [http://jukebox.esc13.net/untdeveloper/RM/Stats\\_Module\\_1/Stats\\_Module\\_19.html](http://jukebox.esc13.net/untdeveloper/RM/Stats_Module_1/Stats_Module_19.html)

## Question 4

a)

$$\frac{dS}{dt} = B - \beta SI - \mu S$$

The above equation is used in epidemiology to extend SIR for considering birth/death rates - this equation improves the SIR model.

$B$  rate  $B$  births per unit time (assume birth rate constant) - assume population size ( $N$ ) constant:  $B = \mu N$

$\beta SI$  number of susceptibles who become infected per unit time - contact/number of encounters of susceptibles considering those already infected:  $\beta SI$  high if number of infected is high -  $\beta$  transmission rate per capita: proportion of  $SI$  interactions leading to infection -  $SI$  total number of encounters between  $S$  (susceptibles) and  $I$  (infected)

$\mu S$  assume mortality rate  $\mu$  per capita; number of susceptibles who die per unit time

b)

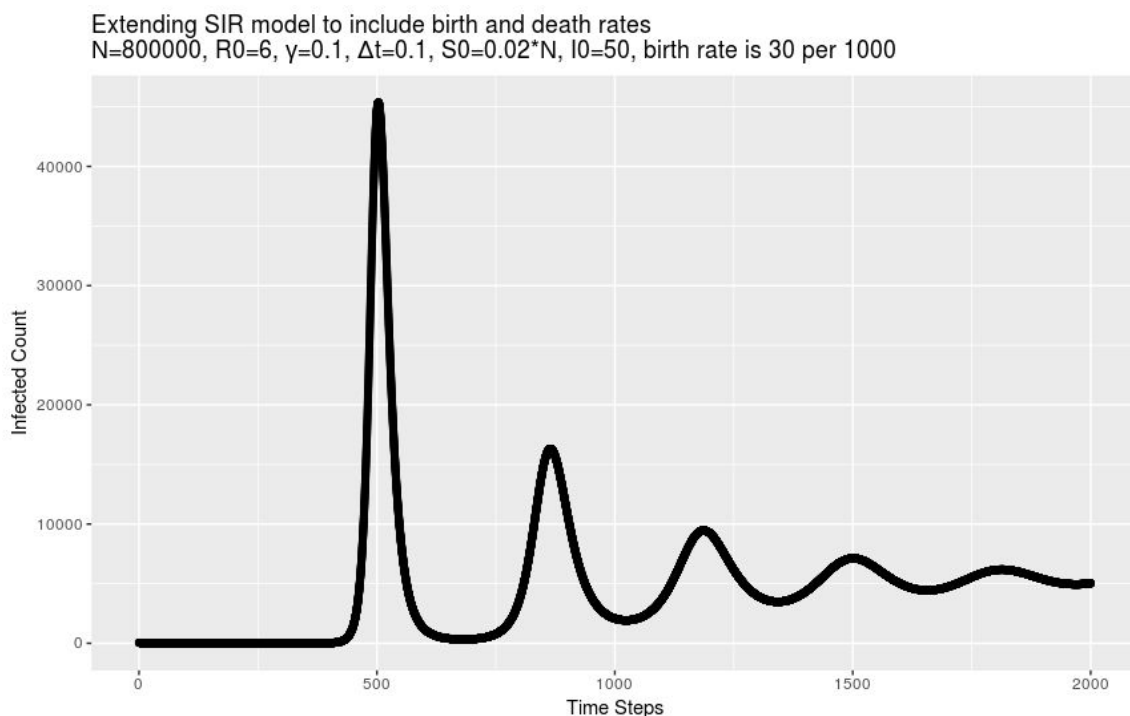


Figure 7: When we consider birth/death rates in the SIR model, it provides us a supply for new susceptibles. We can observe damped oscillations tending towards equilibrium value. We observe that the infection peaks at weeks 500, ~800, ~1200 and 1500.

c)

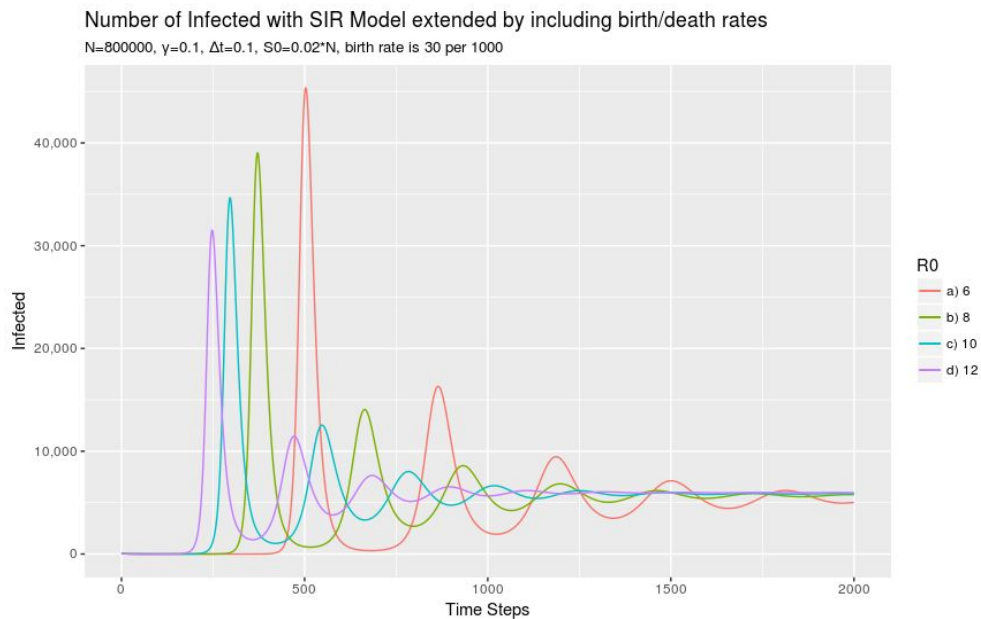


Figure 8: When extending the SIR model to include birth/death rates and altering  $R_0$  (basic reproduction number), the higher the  $R_0$  value the less peaks in the curves plus they reach equilibrium faster. The smaller the  $R_0$ , the infection is worst (infected count higher) yet takes longer.

d)

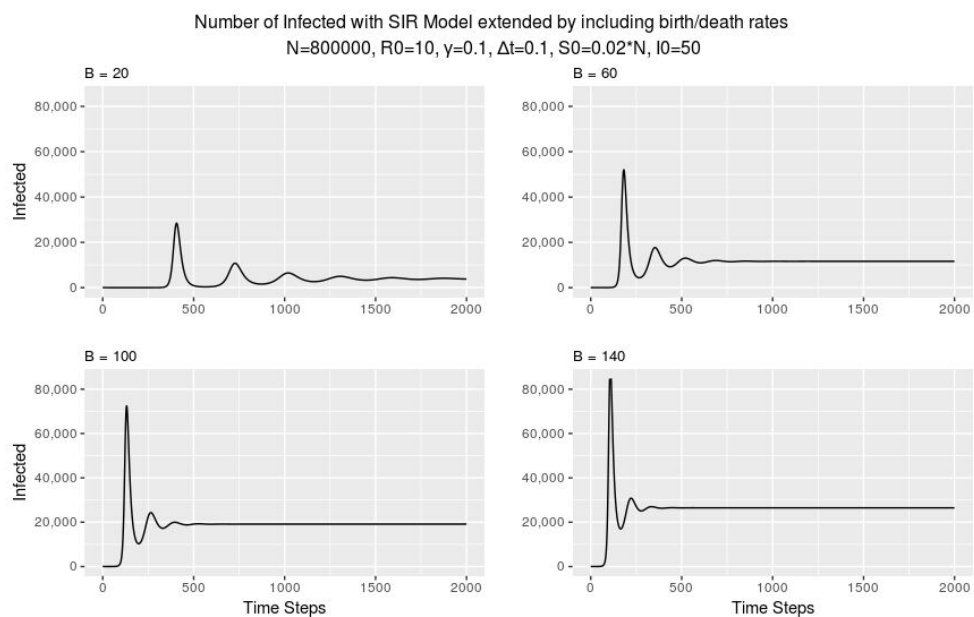


Figure 9: The impact of increasing  $B$  (annual births per 1000) affects the form of the graphs, when  $B$  is higher, the epidemic infects more of the population plus is shorter time overall (faster to infect people plus faster to reach equilibrium).

e)

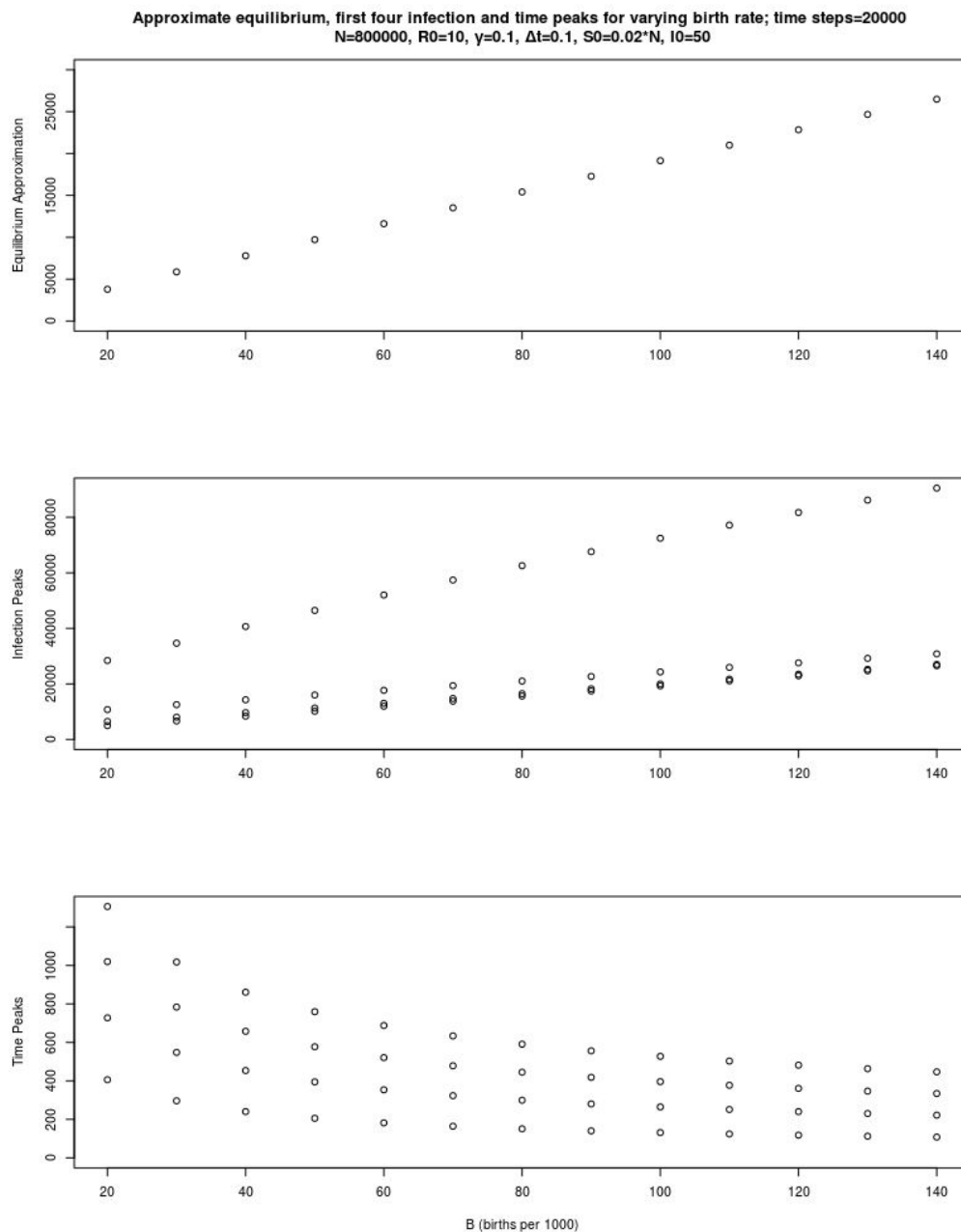


Figure 10: Above the three graphs show the different quantities for various values of B. The first includes the values of B which reach equilibrium faster; the higher the value of B the higher the the equilibrium approximation. For the second plot, we are observing infection peaks, we can see that increasing B, the peaks increase somewhat closely related however the highest peaks for each B value differs greatly: the first infections are always the largest. For the final graph, we can see the highest peaks are at lower time points when we increase B, this follows up from figure 9 - equilibrium is reached faster the higher the value of B so comparing with the first plot we can see the approximation increases.