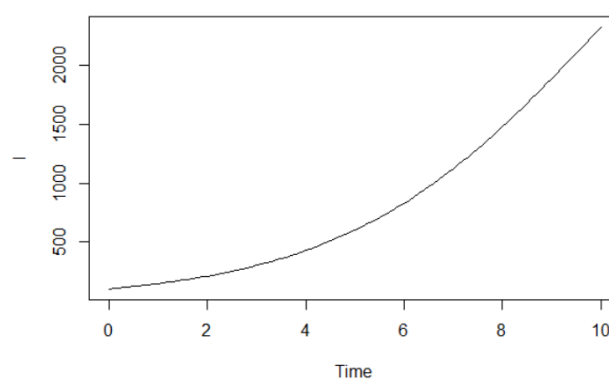


Question 1 (Non-Assessed)

This question introduces the idea of using the Euler method to give an approximate solution to a system of (ordinary) differential equations. The basic idea of the Euler method is to approximate the continuous time derivative (e.g., dS/dt) with a discrete time version, where we take small increments of time.

a) Modify the code given to find the number of susceptible and infectious (which will not necessarily be whole numbers), after 10 time units if we start off with 5000 susceptible, 100 infectious and 900 removed (i.e. $N=6000$), using the following parameter values $\beta = 0.00008$, $\gamma = 0.02$

```
> S0 <- 5000           #Starting value for susceptibles
> I0 <- 100            #Starting value for infectious
> N <- 6000            #Total population size (S0+I0+R)
> beta <- 0.00008      #Infection rate
> gamma <- 0.02        #Recovery rate
>
> delta.t <- 0.1       #Small time increment
> N.time.steps <- 100  #Number of time steps (0.1*100=10 time units)
>
> S <- numeric(N.time.steps+1) #Blank vector to receive S values
> I <- numeric(N.time.steps+1) #Blank vector to receive I values
>
> S[1] <- S0            #Initial value for S
> I[1] <- I0            #Initial value for I
>
> for(i in 1:N.time.steps) {
+   S[i+1] <- S[i]-beta*S[i]*I[i]*delta.t
+   I[i+1] <- I[i]+beta*S[i]*I[i]*delta.t-gamma*I[i]*delta.t
+ }
>
> time.vector <- seq(0, N.time.steps*delta.t, by=delta.t)
>
> plot(time.vector, I, type="l", xlab="Time", ylab="I")
```

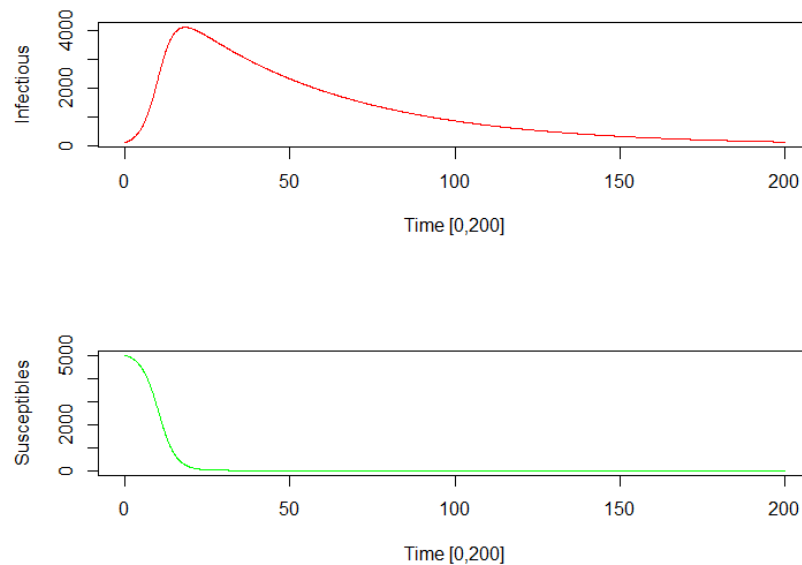


```
> S[101]           #Number of susceptibles after 10 time units
[1] 2611.724
>
> I[101]           #Number of infectious after 10 time units
[1] 2326.782
```

Infectious = 2326.8, Susceptible = 2611.7

As there were 900 people in the removed people already, this means that after 10 time units 161.5 more people have joined the removed group after contracting the disease and 2226.8 people have become infectious after leaving the susceptible pool.

b) Produce side by side plots, with suitable labelling, of the number of susceptible against time and the number of infectious against t over the time interval $[0,200]$



Question 2

This question considers the problem of fitting parameters to a simulated epidemiology data set.

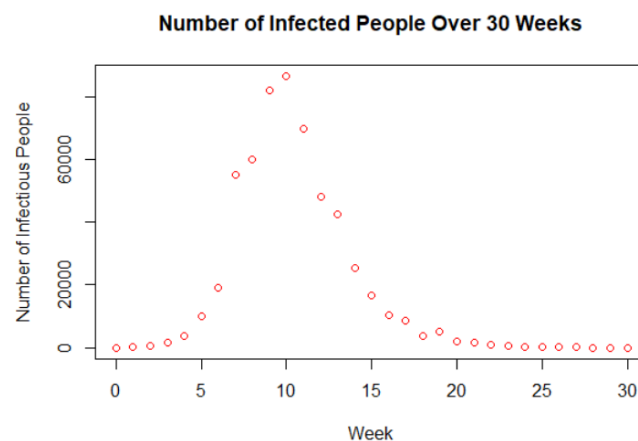
Suppose we have a city of population one million ($N=1000000$). We are given data about the number of infectious people recorded the same day each week over a period of 30 weeks for an epidemic of a disease. We will use a time increment $\Delta t = 0.1$ as previously. (We will use an SIR model without including the birth and death rate).

In the data provided, the number of people infected on a particular day each week are shown. The time index increases by 70 each week because there are 70 increments of 0.1 day in a week.

a) Create a plot of number infected against week index with appropriate titles and axis labels. Describe the epidemic curve.

```
> attach(epidemiology_data)
```

```
> plot(Week, No.infected, ylab="Number of Infectious People", main="Number of Infected People Over 30 Weeks", col="red")
```



The disease spreads through the population quite slowly to begin with and appears to have a fairly low transmission rate with the peak of infectious individuals after 10 weeks, with a significant increase in infectious individuals after week 6.

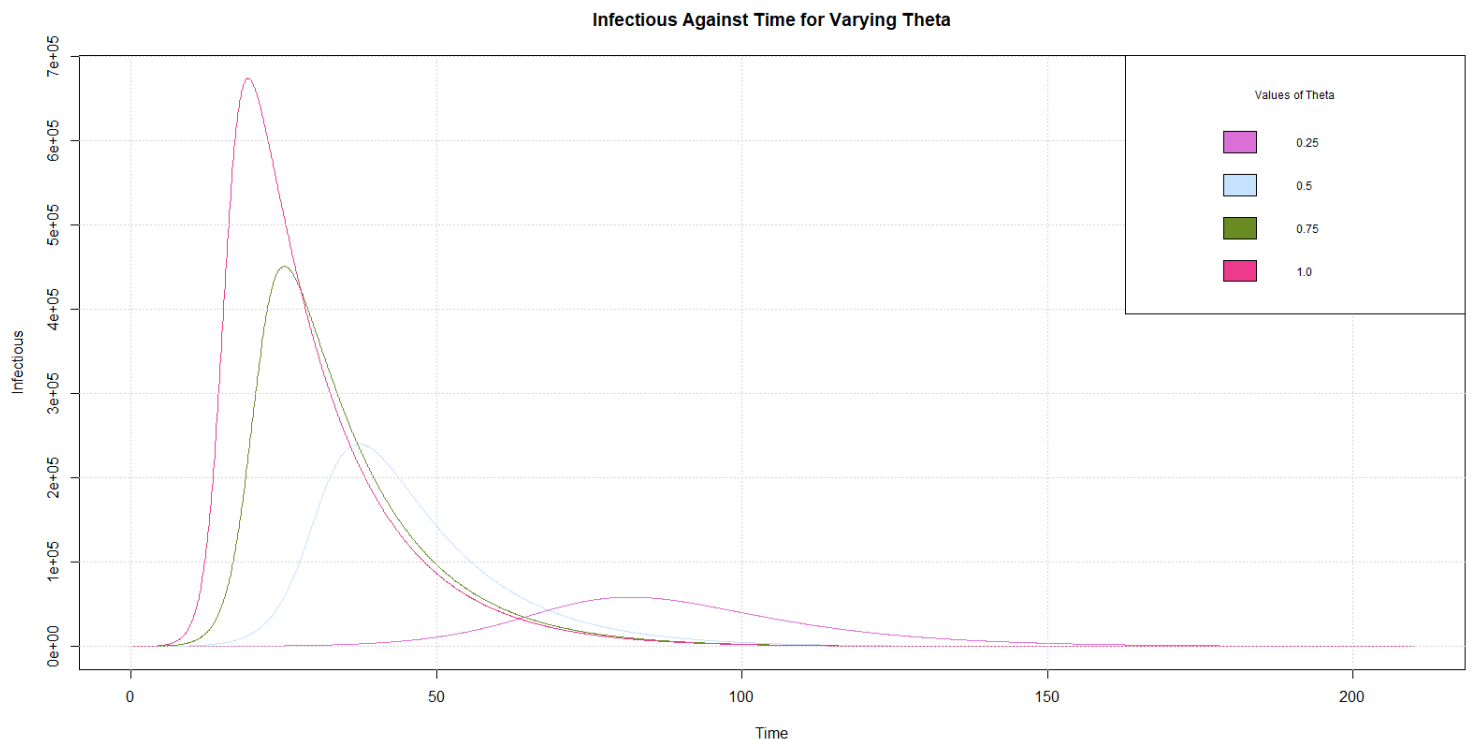
b) We write $S_0 = \theta N$, where θ is a proportion of (i.e. $0 \leq \theta \leq 1$). Modify the code of Question 1, specifying $R_0 = 10$ and mean infectious period over two weeks ($\gamma = \frac{1}{14}$), to produce the graphs of number of infectious against time predicted by the model for $\theta = 0.25, 0.5, 0.75, 1$

```
> infectious.against.time <- function(epidemiology.data, N.time.steps=2100, I0=61, R0=10, N=1000000, theta, delta.t=0.1, gamma=1/14) {
+   S0 <- N*theta
+   beta <- R0*gamma/N
+
+   S <- numeric(N.time.steps+1)      #Empty vector for S values
+   I <- numeric(N.time.steps+1)      #Empty vector for I values
+
+   S[1] <- S0
+   I[1] <- I0
+
+   for(i in 1:N.time.steps) {
+     S[i+1] <- S[i]-beta*S[i]*I[i]*delta.t
+     I[i+1] <- I[i]+beta*S[i]*I[i]*delta.t-gamma*I[i]*delta.t
+   }
+   time.vector <- seq(0, N.time.steps*delta.t, by=delta.t)
+   infectious.against.time.list <- list(time.vector, I)
+   return(infectious.against.time.list)
+ }

> theta.025 <- infectious.against.time(epidemiology.data, theta=0.25)
> theta.05 <- infectious.against.time(epidemiology.data, theta=0.5)
> theta.075 <- infectious.against.time(epidemiology.data, theta=0.75)
> theta.1 <- infectious.against.time(epidemiology.data, theta=1)

> plot(unlist(theta.025[1]), unlist(theta.025[2]), xlab="Time", ylab="Infectious", main="Infectious Against Time for Varying Theta", col="orchid", type="l", ylim=c(0, max(unlist(theta.1[2]))))
>
> lines(unlist(theta.05[1]), unlist(theta.05[2]), col="slategray1", type="l")
>
> lines(unlist(theta.075[1]), unlist(theta.075[2]), col="olivedrab4", type="l")
>
> lines(unlist(theta.1[1]), unlist(theta.1[2]), col="violetred2", type="l")

> grid()
> legend("topright", legend=c("0.25", "0.5", "0.75", "1.0"), fill=c("orchid", "slategray1", "olivedrab4", "violetred2"), title
="Values of Theta", cex=0.75)
```

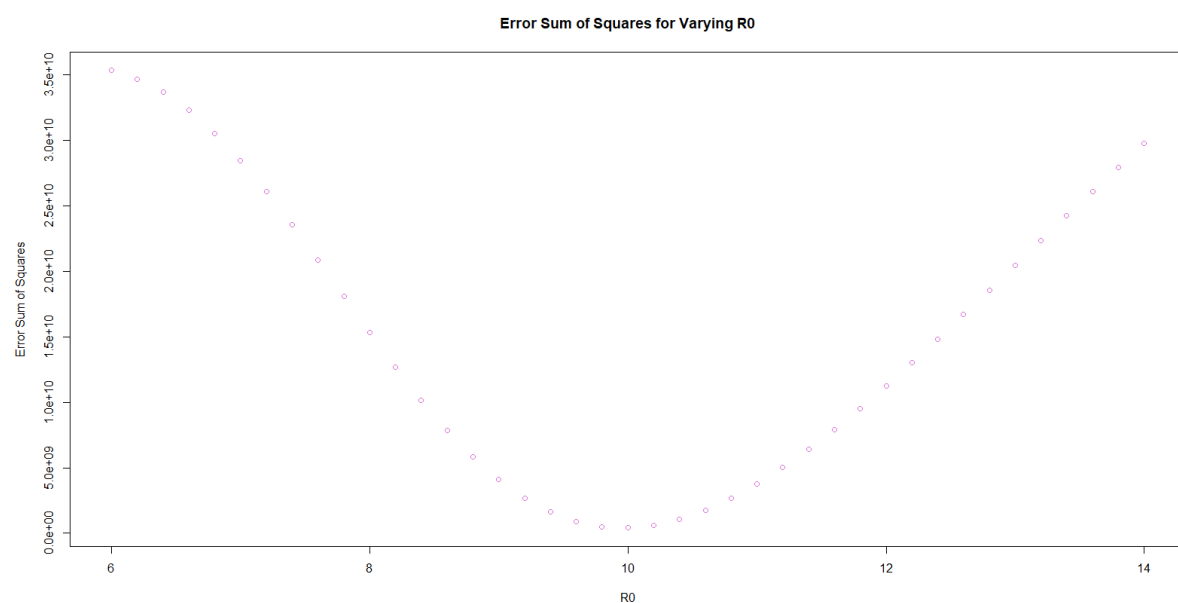


A commonly used technique for measuring how well a particular choice of parameters fits the data is a sum of squares. Suppose we first consider the situation where we consider the mean infectious period fixed ($D=14$ days). We fix $\theta = 0.3$. If we were not told the R_0 value for this epidemic, we may wish to try to infer it from the data.

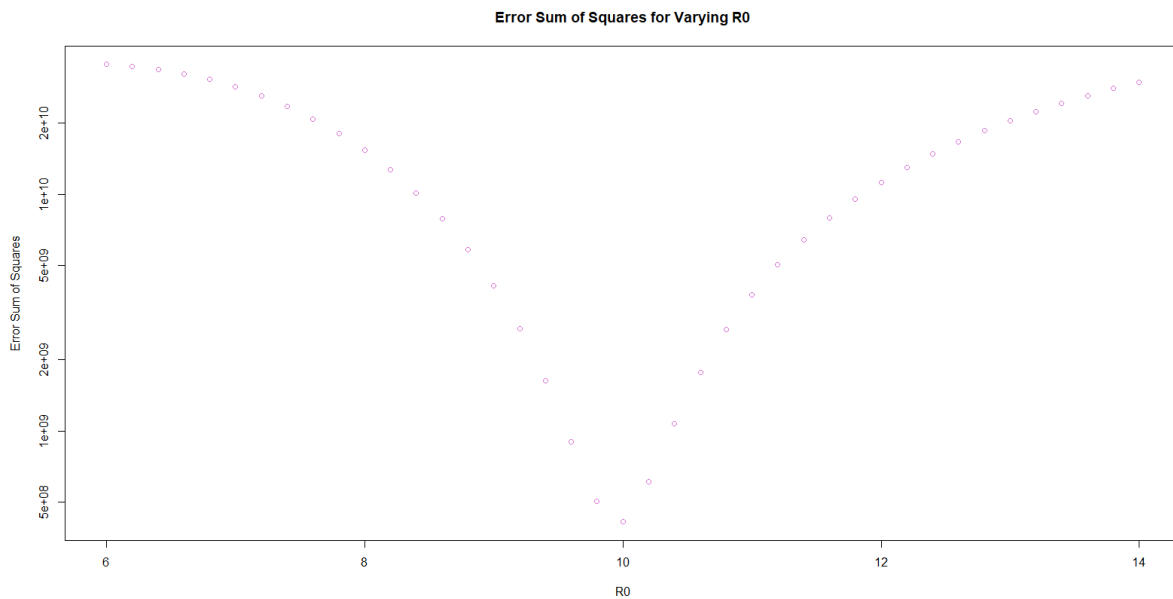
c) Calculate the \log_e (error sum of squares) for R_0 taking values in the vector (6, 6.2, 6.4, ..., 13.6, 13.8, 14.0) and present your results visually, commenting on the form of the plot. Explain why taking logarithms is still valid in this case.

```
> errorSS <- function(epidemiology.data, N=1000000, R0, D, theta, N.time.steps=2100, delta.t=0.1) {
+   gamma <- 1/D
+   beta <- R0*gamma/N
+   S0 <- theta*N
+   I0 <- epidemiology.data$No.infected[1]
+   sample.index <- epidemiology.data$Time.index
+
+   S <- numeric(N.time.steps+1)
+   I <- numeric(N.time.steps+1)
+
+   S[1] <- S0
+   I[1] <- I0
+
+   for (i in 1:N.time.steps) {
+     S[i+1] <- S[i]-beta*S[i]*I[i]*delta.t
+     I[i+1] <- I[i]+beta*S[i]*I[i]*delta.t-gamma*I[i]*delta.t
+   }
+
+   errorSS <- sum((epidemiology.data$No.infected-I[sample.index])^2)
+   return(errorSS)
+ }
>
> R0.vector <- seq(6, 14, 0.2)
> loge.values <- numeric(length(R0.vector))
> for(i in 1:length(R0.vector)) {
+   loge.values[i] <- errorSS(epidemiology.data, R0=R0.vector[i], D=14, theta=0.3)
+ }

> plot(R0.vector, loge.values, xlab="R0", ylab="Error Sum of Squares", main="Error Sum of Squares for Varying R0", col="orchid")
```



```
> plot(R0.vector, loge.values, xlab="R0", ylab="Error Sum of Squares", main="Error Sum of Squares for Varying R0", col="orchid", log="y")
```



Through taking logarithms on the y axis as seen in the second graph, we remove the error sum of squares being equal to 0.

Suppose we wish now to infer **both** R_0 and D (the mean infectious period) from the data. Allow R_0 to vary over the same vector of values as in part (c), and allow D to vary over the values in the vector (10, 10.2, 10.4, ..., 17.6, 17.8, 18). (In order to store the error sums of squares obtained you will need a matrix).

d) Report the minimum \log_e (error sum of squares) obtained along with the corresponding R_0 and D values.

```
> D.vector <- seq(10, 18, 0.2)

> loge.values.d <- numeric(length(R0.vector))
>
> for (i in 1:length(R0.vector)) {
+   loge.values.d[i] <- errorSS(epidemiology.data, R0=R0.vector[i], D=D.vector[i], theta=0.3)
+ }
>
> min(loge.values.d)
[1] 394213638
>
> which.min(loge.values.d)
[1] 20
>
> R0.vector[20]
[1] 9.8
>
> D.vector[20]
[1] 13.8
```

The minimum \log_e is 394213638 which corresponds to an R_0 value of 9.8 and a mean infectious period of 13.8 days.

Question 3

The basic reproduction number, R_0 , specific to a particular infectious disease, is given by

$$R_0 = \frac{\beta N}{\gamma}$$

and its interpretation is the *number of secondary infections caused by a **single** infective introduced into a population made up entirely of susceptible individuals ($S_0=N$).*

Given the following parameters –

$$N = 5000000$$

$$R_0 = 20$$

$$\gamma = 0.1$$

$$\Delta t = 0.1$$

a) Calculate β , the infection rate, using the R_0 value

$$R_0 = \frac{\beta N}{\gamma} \quad \leftrightarrow \quad \beta = \frac{R_0 * \gamma}{N}$$

```
> N <- 5000000           #Total population size
> R0 <- 20                #Reproduction number
> gamma <- 0.1            #Recovery rate
> delta.t <- 0.1          #Small time increment
>
> beta <- (R0*gamma)/N
> beta
[1] 4e-07
```

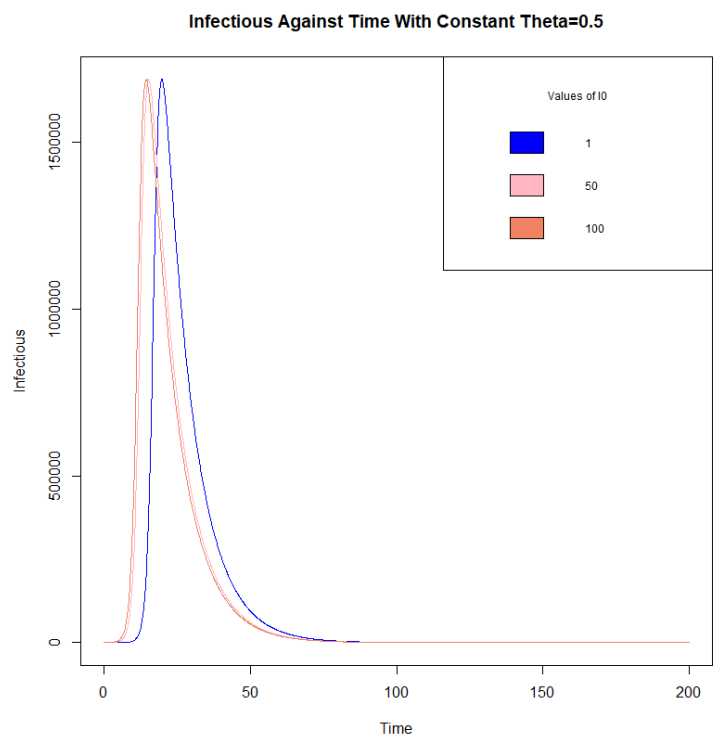
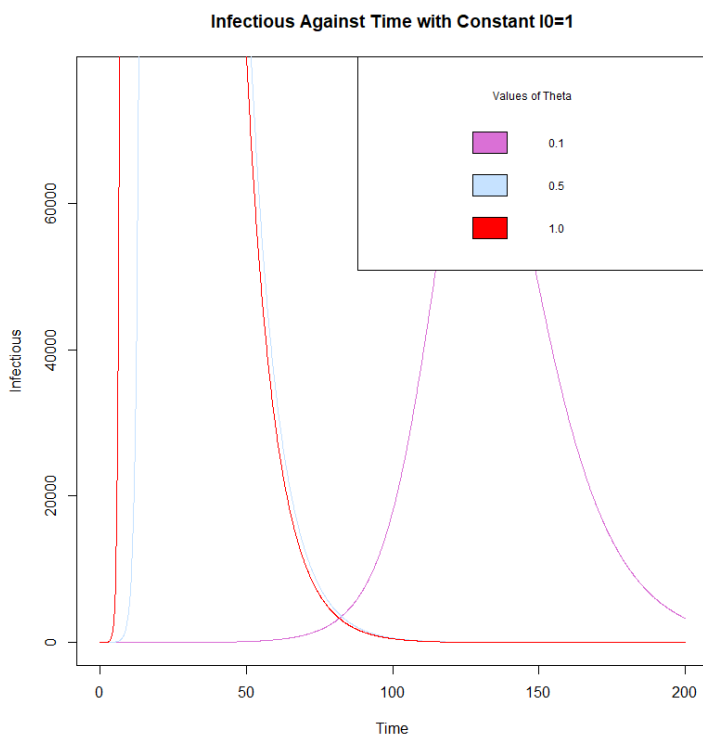
Therefore the infection rate is 0.0000004 which means of the total encounters between the infected and susceptible groups, a proportion of 0.0000004 result in infection.

b) Let $S_0 = \theta N$. Investigate the effects of altering θ and I_0 on the time course of infectious against time. Choose a range of values for θ in the interval $(0,1]$ and consider a range of possible values of I_0 , including $I_0 = 1$. (It is sensible to vary θ and I_0 independently, i.e. holding the other fixed). You may wish to plot multiple graphs on the same axes to save space and to help with visualization.

```
> infectious.against.time <- function(N.time.steps=2000, I0, R0=20, N=5000000, theta, delta.t=0.1, gamma=0.1) {
+   S0 <- N*theta
+   beta <- R0*gamma/N
+
+   S <- numeric(N.time.steps+1)      #Empty vector for S values
+   I <- numeric(N.time.steps+1)      #Empty vector for I values
+
+   S[1] <- S0
+   I[1] <- I0
+   #Initial value for S
+   #Initial value for I
+
+   for(i in 1:N.time.steps) {
+     S[i+1] <- S[i]-beta*S[i]*I[i]*delta.t
+     I[i+1] <- I[i]+beta*S[i]*I[i]*delta.t-gamma*I[i]*delta.t
+   }
+   time.vector <- seq(0, N.time.steps*delta.t, by=delta.t)
+   infectious.against.time.list <- list(time.vector, I)
+   return(infectious.against.time.list)
+ }

> theta <- seq(0.1, 1, 0.1)
> I0 <- seq(1, 100, 10)

> constantI.1 <- infectious.against.time(I0=1, theta=0.1)
> constantI.2 <- infectious.against.time(I0=1, theta=0.5)
> constantI.3 <- infectious.against.time(I0=1, theta=1)
>
> constantT.1 <- infectious.against.time(I0=1, theta=0.5)
> constantT.2 <- infectious.against.time(I0=50, theta=0.5)
> constantT.3 <- infectious.against.time(I0=100, theta=0.5)
>
> par(mfrow=c(1,2))
>
> plot(unlist(constantI.1[1]), unlist(constantI.1[2]), type="l", xlab="Time", ylab="Infectious", main="Infectious Against Time with Constant I0=1", col="orchid")
> lines(unlist(constantI.2[1]), unlist(constantI.2[2]), type="l", col="slategray1")
> lines(unlist(constantI.3[1]), unlist(constantI.3[2]), type="l", col="red")
>
> legend("topright", legend=c("0.1", "0.5", "1.0"), fill=c("orchid", "slategray1", "red"), cex=0.75, title="Values of Theta")
```



Question 4

Here we extend the model to include births and deaths. We assume a constant birth rate, B births per unit time, all of whom enter the S pool (the susceptibles). We also assume a constant per capita mortality rate, μ . The equations become –

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

a) What do the three parts of the right hand side of the first equation represent?

- $+B$
This part represents the influx of births into the susceptible pool
- $-\beta SI$
This part represents the rate at which the individuals in the susceptible pool get infected and become infectious
- $-\mu S$
This part represents the rate at which individuals in the susceptible pool die due to the mortality rate μ

Therefore, this equation reads as the rate of change of susceptibles with time is equal to the birth rate minus the rate at which people get infected and leave for the infectious pool minus the amount of ‘natural deaths’ that occur in the susceptible pool that aren’t caused by the disease.

b) The basic reproduction number for diphtheria may be taken as $R_0=6$. Suppose we wish to model a diphtheria epidemic in a city of population 800,000. Suppose we know that $\gamma = 0.1$ and the birth rate is 30 per 1000 population per year (so the daily rate, $B=30 \times (800,000/1000)/365=65.75$). (We assume that the birth rate=overall death rate, i.e. $B = \mu N$). Using $\Delta t = 0.1$ as before, an initial number of infected equal to 50 and an initial number of susceptibles, $S_0 = 0.02N$, produce a plot showing how the number of infectious people varies over time. (Make sure the time axis extends far enough to capture multiple peaks in the time course). Comment on the form of the plot.

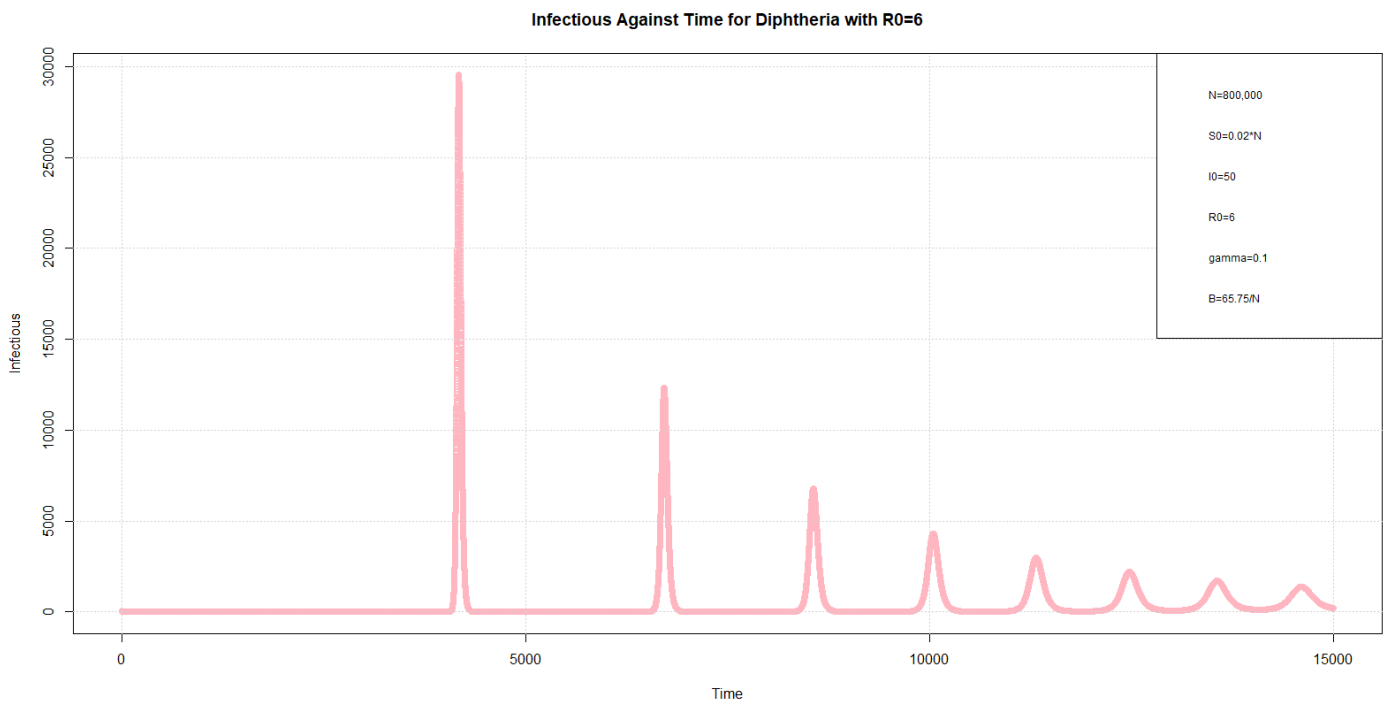
```
> diphtheria.data <- generate.S.I.by.time.vital.dynamics(N.time.steps=1000, S0=S0, I0=I0, R0=R0, gamma=gamma, mu=mu, N=N)
> diphtheria.susceptibles <- diphtheria.data[1]
> diphtheria.infected <- diphtheria.data[2]
> diphtheria.time <- diphtheria.data[3]
+   beta <- R0*gamma/N
+
+   for(i in 1:N.time.steps){
+     S[i+1] <- S[i]+mu*N*delta.t-beta*S[i]*I[i]*delta.t-mu*S[i]*delta.t
+     I[i+1] <- I[i]+beta*S[i]*I[i]*delta.t-gamma*I[i]*delta.t-mu*I[i]*delta.t
+   }
+
+   time.vector <- seq(0, N.time.steps*delta.t, by=delta.t)
+   out <- list(S=S, I=I, time.vector=time.vector)
+   return(out)
+ }
```

```

> N <- 800000                                #Total population size
> gamma <- 0.1                               #Recovery rate
> B <- 65.75                                 #Birth rate
> mu <- B/N                                  #Death rate
>
> I0 <- 50                                   #Initial value for infectious
> S0 <- 0.02*N                               #Initial value for susceptibles
> R0 <- 6                                     #Basic reproduction number for diphtheria
>
> generate.S.I.by.time.vital.dynamics(N.time.steps=2000, S0=S0, I0=I0, R0=R0, gamma=gamma, mu=mu, N=N)

> plot(unlist(diphtheria.data[3]), unlist(diphtheria.data[2]), xlab="Time", ylab="Infectious", main="Infectious Against Time for Diphtheria with R0=6", col="lightpink")
> grid()
> legend("topright", legend=c("N=800,000", "S0=0.02*N", "I0=50", "R0=6", "gamma=0.1", "B=65.75/N"), cex=0.75)

```

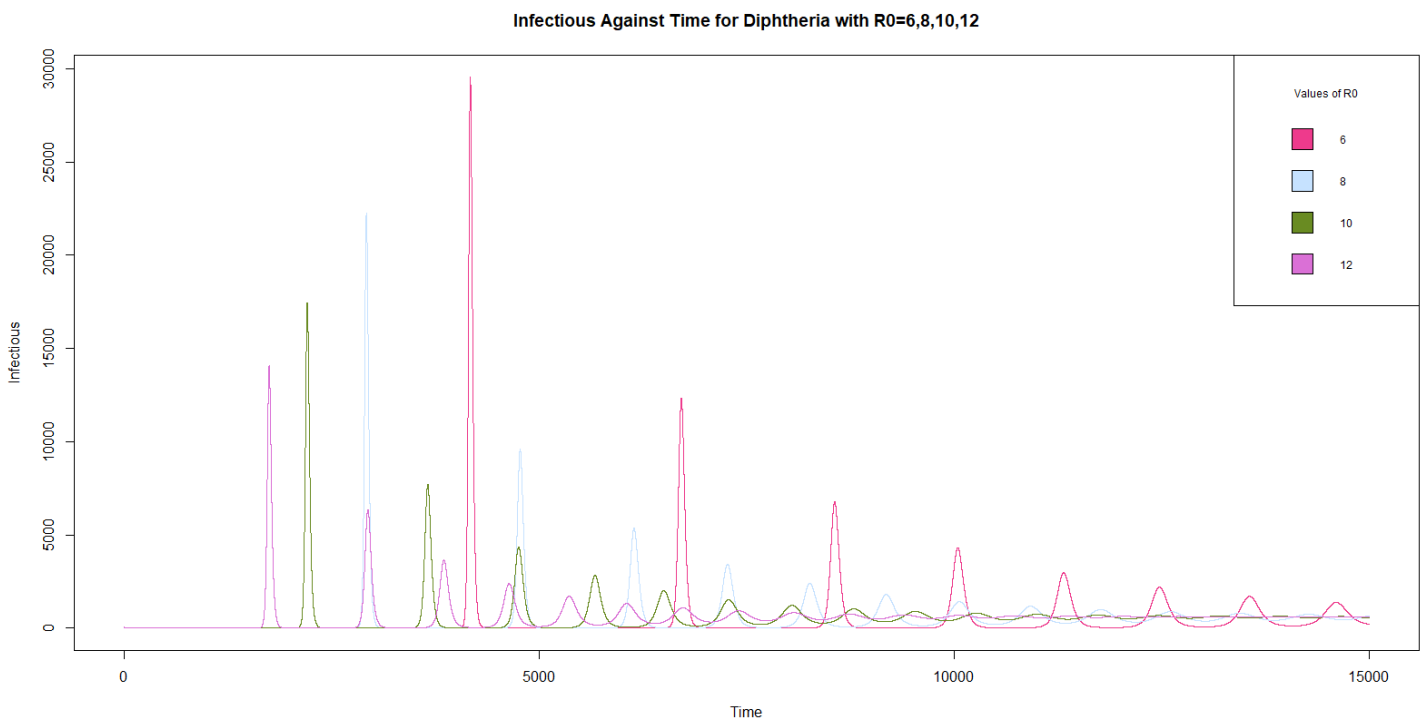


The number of infectious people decreases after time after the first big outbreak at around time 4900, after people have been infected at this point once past the recovery time they enter the removed pool and are immune to the disease which is why the infectious rate gets smaller and smaller.

c) Investigate the effect of changing R_0 over the values [6,8,10,12]. Clearly describe the effect on the graph of number of infectious over time.

```
> diphtheria.susceptibles.6 <- diphtheria.data.6[1]
> diphtheria.infected.6 <- diphtheria.data.6[2]
>
> diphtheria.susceptibles.8 <- diphtheria.data.8[1]
> diphtheria.infected.8 <- diphtheria.data.8[2]
>
> diphtheria.susceptibles.10 <- diphtheria.data.10[1]
> diphtheria.infected.10 <- diphtheria.data.10[2]
>
> diphtheria.susceptibles.12 <- diphtheria.data.12[1]
> diphtheria.infected.12 <- diphtheria.data.12[2]

> plot(unlist(diphtheria.data[3]), unlist(diphtheria.infected.6), xlab="Time", ylab="Infectious", type="l",
      , main="Infectious Against Time for Diphtheria with R0=6,8,10,12", col="violetred2")
> lines(unlist(diphtheria.data[3]), unlist(diphtheria.infected.8),type="l", col="slategray1")
> lines(unlist(diphtheria.data[3]), unlist(diphtheria.infected.10),type="l", col="olivedrab4")
> lines(unlist(diphtheria.data[3]), unlist(diphtheria.infected.12),type="l", col="orchid")
>
> legend("topright", legend=c("6", "8", "10", "12"), fill=c("violetred2", "slategray1", "olivedrab4", "orchid"),
      title="Values of R0", cex=0.75)
```



A higher R_0 means that the disease is more infectious which can clearly be seen in the graph as $R_0=12$ is the first peak to appear on the graph, with the lower R_0 not appearing until much later.

d) We now consider the effect of changing the birth rate, B . We fix R_0 at 10. Firstly allow B to vary over the set $\{100, 200, 300, 600\}$ (where these numbers represent daily birth rates). Produce four graphs of infectious against time and comment on the impact of increasing B on the form of the graphs.

```
> mu1 <- 100/N          #Birth rate = 100
> mu2 <- 200/N          #Birth rate = 200
> mu3 <- 300/N          #Birth rate = 300
> mu4 <- 600/N          #Birth rate = 600

> birthrate.1 <- generate.S.I.by.time.vital.dynamics(N.time.steps=100000, S0=S0, I0=I0, R0=10, gamma=gamm
a, mu=mu1, N=N)
> View(birthrate.1)
> birthrate.1.infected <- birthrate.1[2]
> birthrate.time <- birthrate.1[3]
>
> birthrate.2 <- generate.S.I.by.time.vital.dynamics(N.time.steps=100000, S0=S0, I0=I0, R0=10, gamma=gamm
a, mu=mu2, N=N)
> birthrate.2.infected <- birthrate.2[2]
>
> birthrate.3 <- generate.S.I.by.time.vital.dynamics(N.time.steps=100000, S0=S0, I0=I0, R0=10, gamma=gamm
a, mu=mu3, N=N)
> birthrate.3.infected <- birthrate.3[2]
>
> birthrate.4 <- generate.S.I.by.time.vital.dynamics(N.time.steps=100000, S0=S0, I0=I0, R0=10, gamma=gamm
a, mu=mu4, N=N)
> birthrate.4.infected <- birthrate.4[2]

> par(mfrow=c(2,2))

> plot(unlist(birthrate.time), unlist(birthrate.1.infected), xlab="Time", ylab="Infectio
us", main="Infectious Against Time with Birth Rate = 100", col="orchid", type="l")
>
> plot(unlist(birthrate.time), unlist(birthrate.2.infected), xlab="Time", ylab="Infectio
us", main="Infectious Against Time with Birth Rate = 200", col="slategray1", type="l")
>
> plot(unlist(birthrate.time), unlist(birthrate.3.infected), xlab="Time", ylab="Infectio
us", main="Infectious Against Time with Birth Rate = 300", col="sienna2", type="l")
>
> plot(unlist(birthrate.time), unlist(birthrate.4.infected), xlab="Time", ylab="Infectio
us", main="Infectious Against Time with Birth Rate = 600", col="lightpink", type="l")
```

