Modeling Population Turnover

In previous models, we have assumed a closed population such that no one dies and no babies are born. In terms of epidemic dynamics, this meant that as more people became infected and recovered, the amount of susceptible persons continually decreased. As a result, R_{eff}, the effective reproduction number, also decreased. Diseases with short infectious periods move through a population more quickly than the population is able to renew itself, so this assumption is mostly fine. However, if dealing with a longer timescale, it is necessary to incorporate births and deaths into the SIR model. The structure of the model can be visualized as this:

The differential equations we can use to create this model include:

$$\frac{dS}{dt} = -\lambda S - \mu S + bN \tag{1}$$
$$\frac{dI}{dt} = \lambda S - \gamma I - \mu I \tag{2}$$
$$\frac{dR}{dt} = \gamma I - \mu R \tag{3}$$

$\lambda = \beta \frac{I}{N}$, so we can rewrite the equations as:

$$\frac{dS}{dt} = -\beta \frac{I}{N} S - \mu S + bN \tag{4}$$
$$\frac{dI}{dt} = \beta \frac{I}{N} S - \gamma I - \mu I \tag{5}$$
$$\frac{dR}{dt} = \gamma I - \mu R \tag{6}$$

Developing a Model for Slow Population Turnover

We assume that all individuals across compartments are subject to the same mortality rate (ex. no distinction between ages), and that everyone who dies is removed from the population and does not contribute to transmission. In this case, we ignore disease-induced mortality and instead focus on deaths from all other causes. We also assume all newborns in the population are susceptible.

We will first look at acute disease epidemic introduced to a fully susceptible human population. The infection and recovery rates will be 0.4 and 0.2 days^{-1} respectively; these can be converted into years^{-1} by multiplying by 365. We can assume the average human lifespan of 70 years and calculate a background mortality rate of $\mu = \frac{1}{70}$years^{-1} or $\frac{1}{70 \times 365}$days^{-1}. We will assume that the population size stays constant over time at 1,000,000, so we set the birth rate b to μ.

We will now create a model to observe over 400 years, to see the disease dynamics over multiple generations. We will use a time step of 1 day.

```
library(deSolve)
library(reshape2)
library(ggplot2)
```

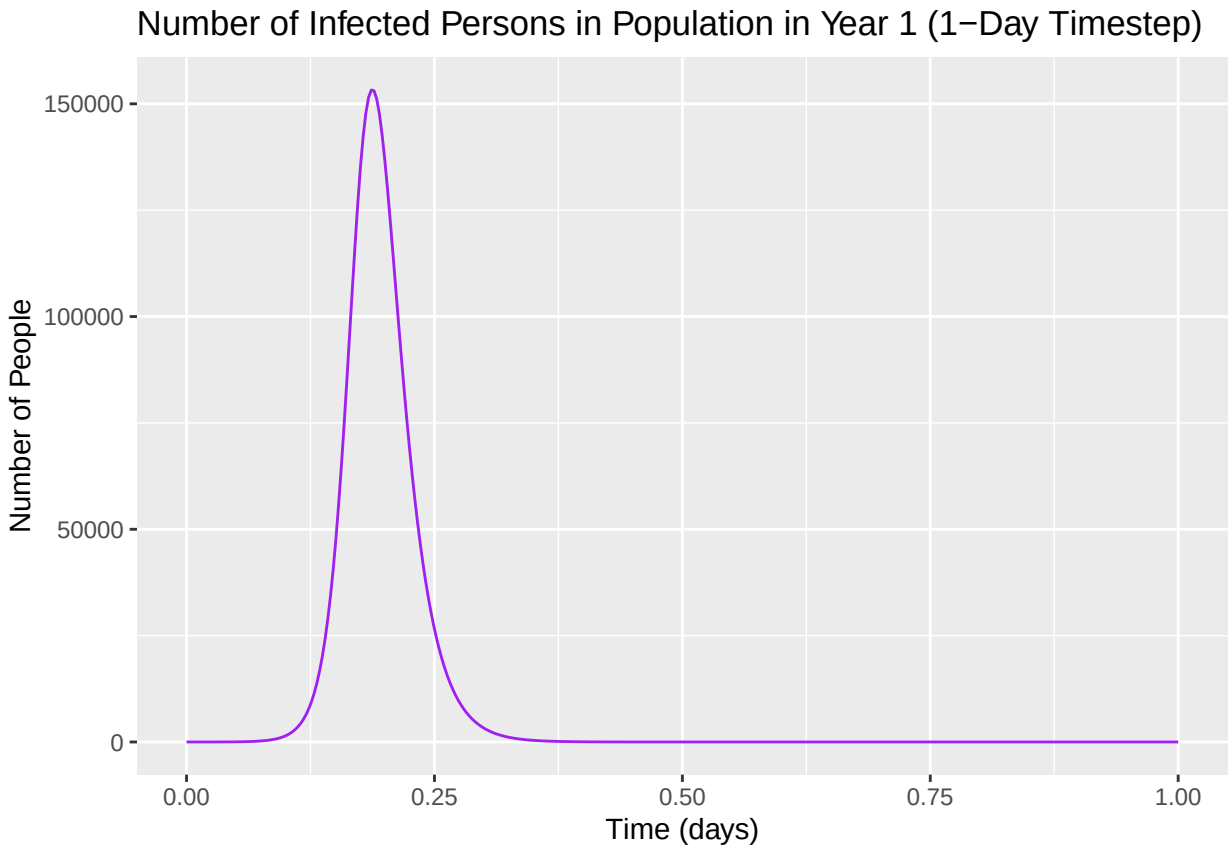
```r
initial_number_susceptible <- 1000000 - 1
initial_number_infected <- 1
initial_number_recovered <- 0

initial_state_values <- c(S = initial_number_susceptible,
                          I = initial_number_infected,
                          R = initial_number_recovered)

times <- seq(from = 0, to = 400, by = 1/365)

birth_rate_days = 1/70 # set birth_rate equal to mu to ensure 0 growth in population

parameters <- c(beta = 0.4 * 365, gamma = 0.2 * 365, mu = birth_rate_days, b = birth_rate_days)
```

```r
SIR_model <- function(time, state, parameters) {

    with(as.list(c(state, parameters)), {

      N = S + I + R

      lambda <- beta * I / N

      dS <- -lambda * S - mu * S + b * N

      dI <- lambda * S - gamma * I - mu * I

      dR <- gamma * I - mu * R

      return(list(c(dS, dI, dR)))
    })
}
```

```r
output1 <- as.data.frame(ode(y = initial_state_values,
                             times = times,
                             func = SIR_model,
                             parms = parameters))
```

```r
ggplot(data = output1,
       aes(x = time, y = I)) +
  geom_line(color = "purple") +
  xlab("Time (days)") +
  ylab("Number of People") +
  labs(title = "Number of Infected Persons in Population in Year 1 (1-Day Timestep)") +
  xlim(c(0,1))
```

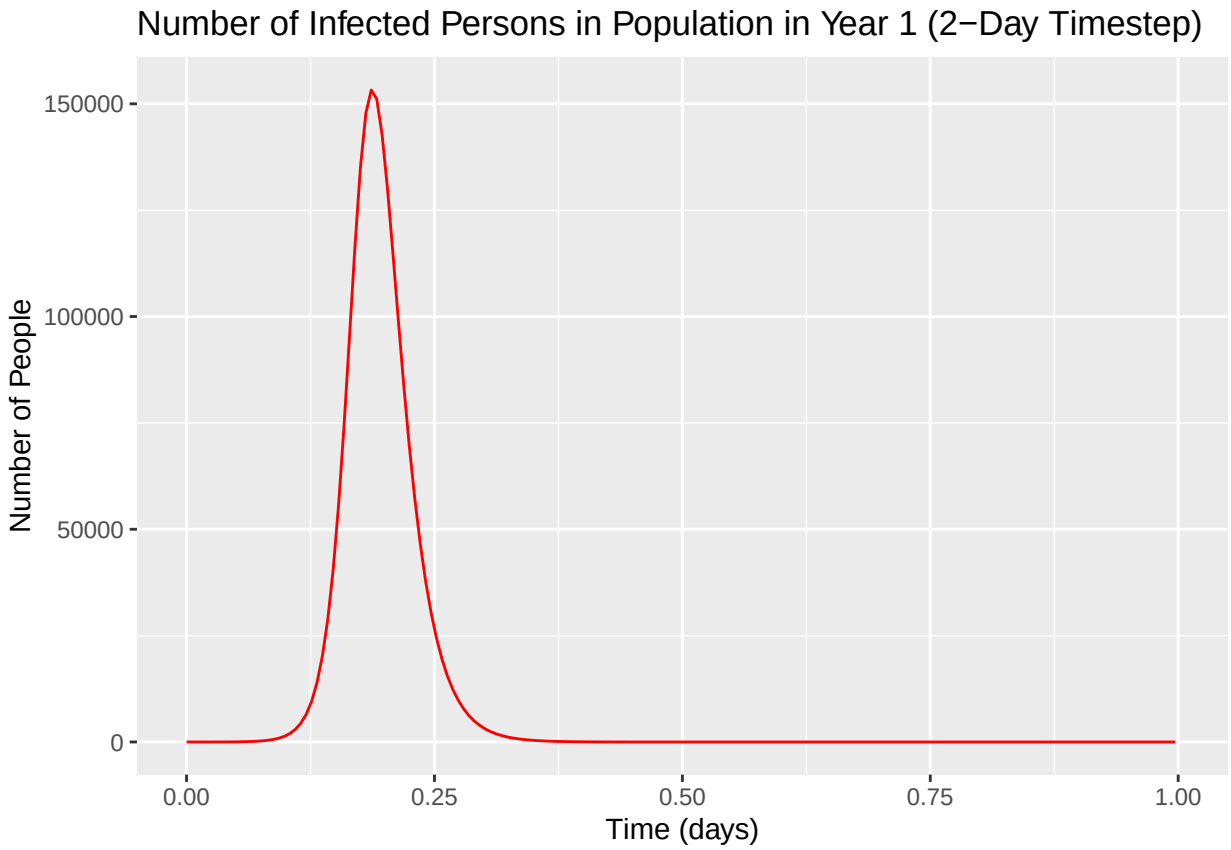
```
## Warning: Removed 145635 row(s) containing missing values (geom_path).
```



We could try a different time step of every 2 days:

```
times2 <- seq(from = 0, to = 400, by = 2/365)

output2 <- as.data.frame(ode(y = initial_state_values,
                             times = times2,
                             func = SIR_model,
                             parms = parameters))

ggplot(data = output2,
       aes(x = time, y = I)) +
  geom_line(color = "red") +
  xlab("Time (days)") +
  ylab("Number of People") +
  labs(title = "Number of Infected Persons in Population in Year 1 (2-Day Timestep)") +
  xlim(c(0,1))
```

```
## Warning: Removed 72818 row(s) containing missing values (geom_path).
```



We see we obtain a very similar result. However, if we try a time step of every 5 days:

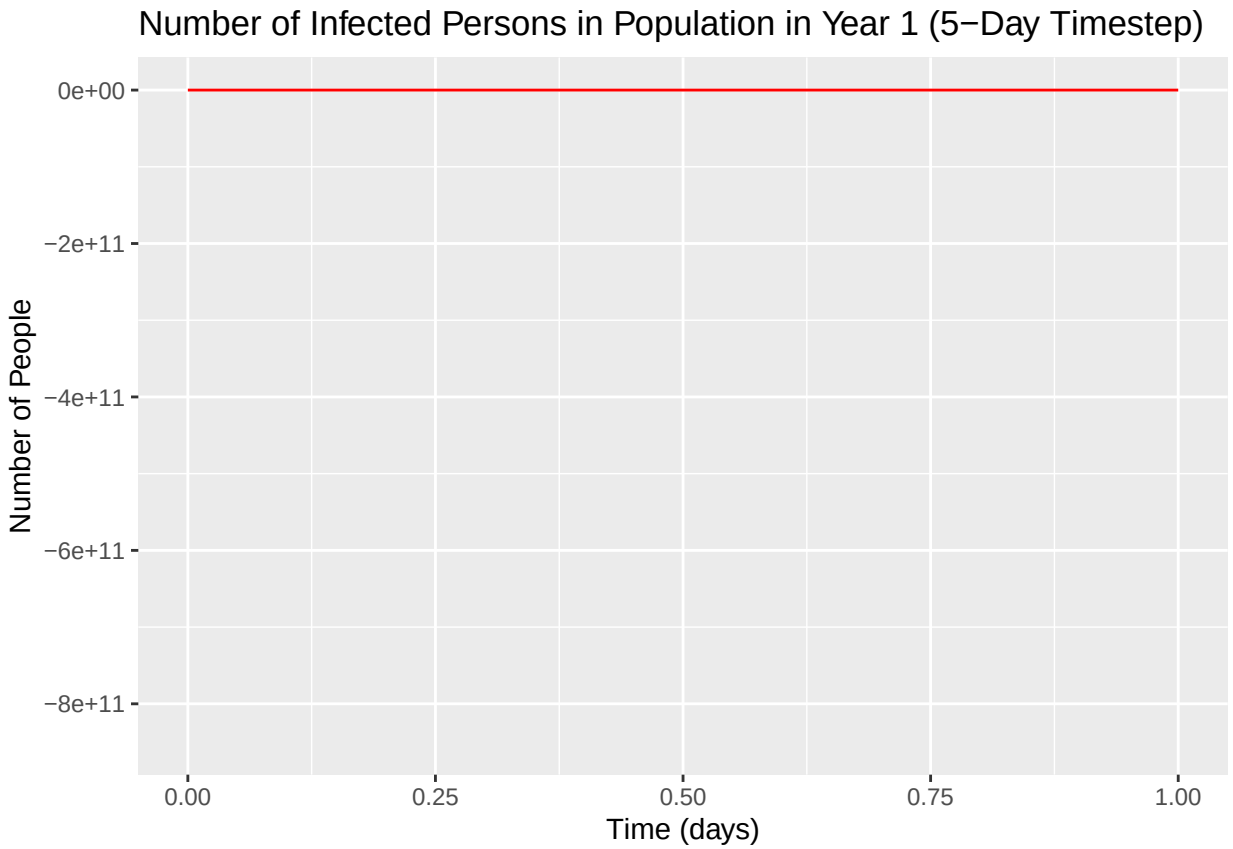
```
times3 <- seq(from = 0, to = 400, by = 5/365)

output3 <- as.data.frame(ode(y = initial_state_values,
                             times = times3,
                             func = SIR_model,
                             parms = parameters))
```

```
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 5.69001e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 5.69001e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 5.69001e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 5.69001e-15
```

```
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 5.69001e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 4.47188e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 4.47188e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 4.47188e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 4.47188e-15
##
## DLSODA-  Warning..Internal T (=R1) and H (=R2) are
##       such that in the machine, T + H = T on the next step
##      (H = step size). Solver will continue anyway.
## In above message, R1 = 71.5756, R2 = 3.7251e-15
##
## DLSODA-  Above warning has been issued I1 times.
##      It will not be issued again for this problem.
## In above message, I1 = 10
##
## DLSODA-  At current T (=R1), MXSTEP (=I1) steps
##       taken on this call before reaching TOUT
## In above message, I1 = 5000
##
## In above message, R1 = 71.5756
##

## Warning in lsoda(y, times, func, parms, ...): an excessive amount of work (>
## maxsteps ) was done, but integration was not successful - increase maxsteps

## Warning in lsoda(y, times, func, parms, ...): Returning early. Results are
## accurate, as far as they go
```

```
ggplot(data = output3,
       aes(x = time, y = I)) +
  geom_line(color = "red") +
  xlab("Time (days)") +
  ylab("Number of People") +
  labs(title = "Number of Infected Persons in Population in Year 1 (5-Day Timestep)") +
  xlim(c(0,1))
```

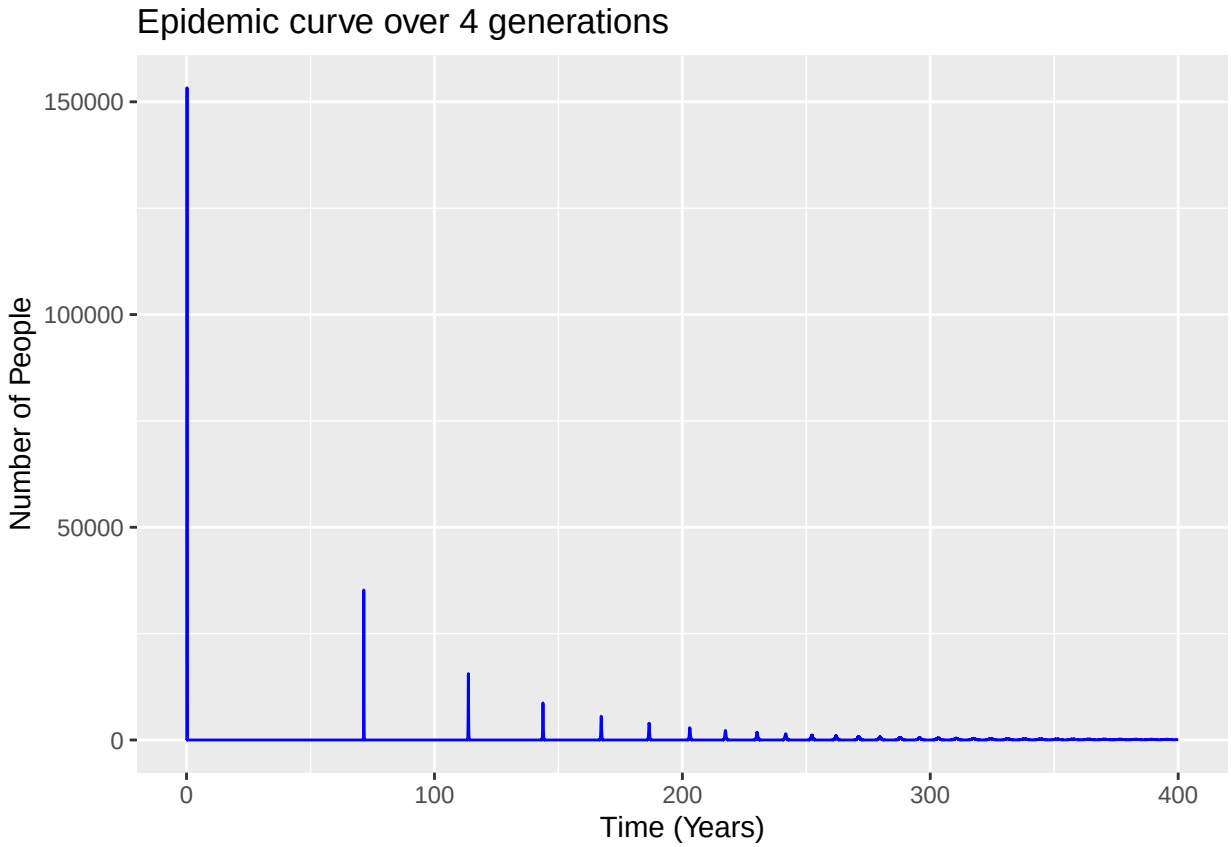
```
## Warning: Removed 5153 row(s) containing missing values (geom_path).
```



We see that we obtain many errors trying to use a 5-day time step. The average infectious period is $\frac{1}{\gamma} = \frac{1}{0.2} = 5$ days. A time step of 5 days is too long, because the new recoveries at each time step are greater than the number of infected individuals. Because the number of newly recovered individuals ($\gamma \times I$) is subtracted from the number of individuals currently infected, this eventually leads to negative values in the number of infected people, or I.

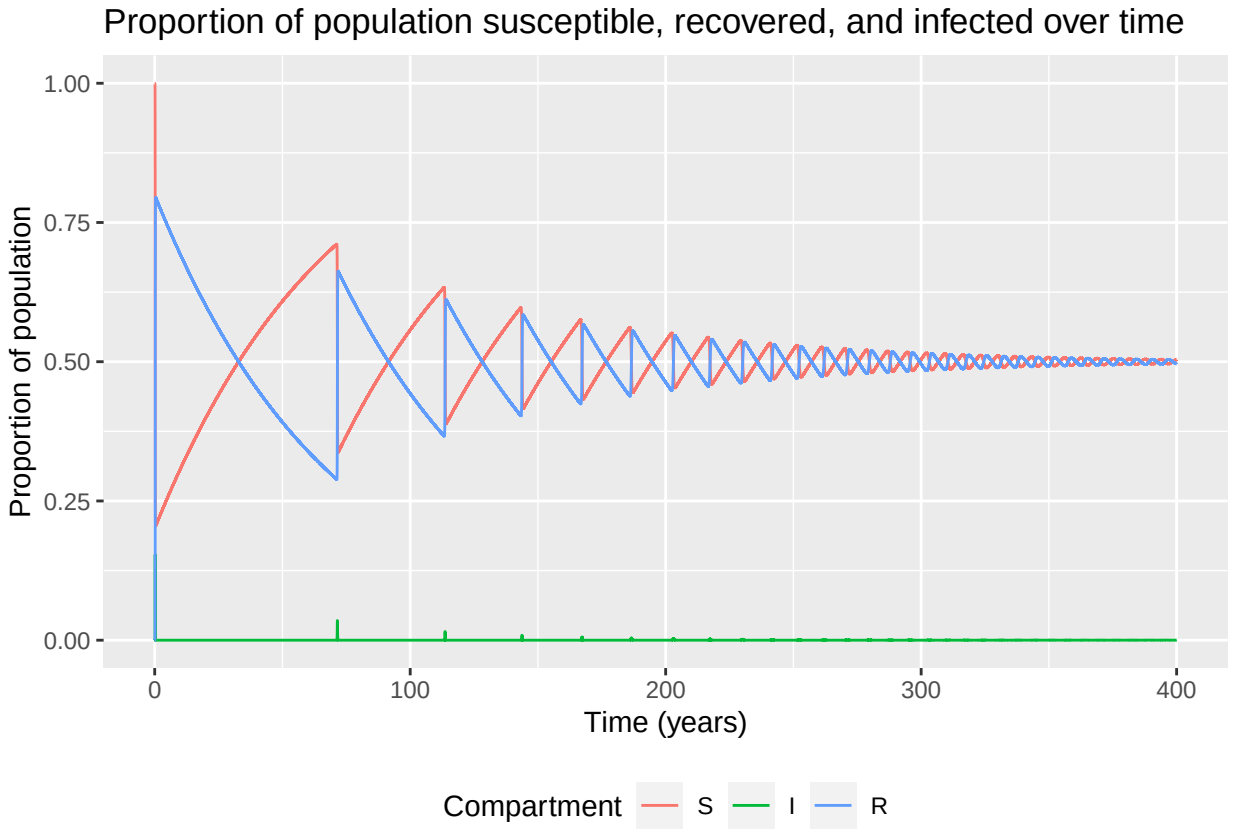
We'll continue using our 1-day time step and will model this epidemic over the full 400 years now.

```
output_long1 <- melt(as.data.frame(output1), id = "time")

ggplot(data = output1,
       aes(x = time, y = I)) +
  geom_line(color = "blue") +
  xlab("Time (Years)") +
  ylab("Number of People") +
  labs(title = "Epidemic curve over 4 generations")
```



We see that over 150,000 are infected the first year (this is the first epidemic we plotted), with the next spike occurring at around year 70, with significantly less people infected. The pattern follows in later years, where a spike occurs later but the number of infections are much less, until the epidemic comes to a full close by the year 400. Initially, when the disease is introduced to the population, no one is immune and an epidemic of great scale takes place. Once the epidemic peaks in the first year, and then wanes due to a smaller population of susceptible people, it does not resurface till later. This is because the disease has a much shorter duration period than the turnover for a human population. Once the disease has spread through the population and reduced the pool of susceptibles, it takes years for the susceptible pool to replenish through births; this is why there are around 70 year breaks between epidemics, as this is the average human life expectancy. To see the behavior of susceptible and recovered people over time, we can plot the proportion of susceptible, recovered, and infected individuals below:

```
output_long1$proportion <- output_long1$value/sum(initial_state_values)

ggplot(data = output_long1,
       aes(x = time, y = proportion, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (years)") +
  ylab("Proportion of population") +
  labs(color = "Compartment", title = "Proportion of population susceptible, recovered, and infected ov
  theme(legend.position = "bottom")
```



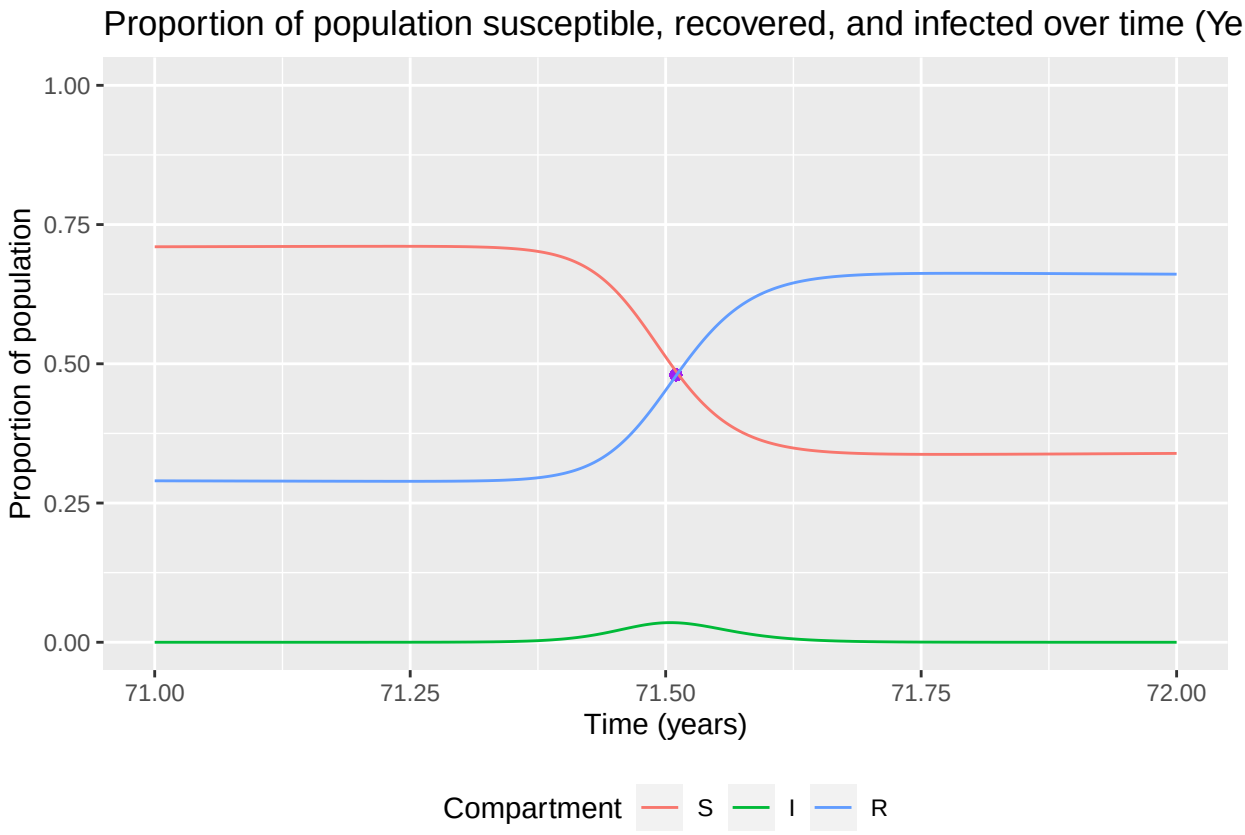
We see there is a relationship between the susceptible and immune, with the proportion of susceptible people being greatest when the proportion of immune people are low, and vice versa. Once the proportion of susceptible individuals in the population is enough for the infection to spread, the number of infected people in the population begins to rise. Just before the epidemic peaks, more susceptibles are removed (are infected) than are added through births, so the susceptible pool begins to decline. The epidemic peaks when the effective reproduction number = 1. In this simple SIR model, the effective reproduction number is proportional to the proportion of susceptibles:

$$R_{eff} = R_0 \frac{S}{N} \tag{7}$$

We see through this equation that the epidemic peaks when the proportion of susceptibles $\frac{S}{N} = \frac{1}{R_0}$. In this example, we know $R_0 = \frac{\beta}{\gamma} = \frac{0.4}{0.2} = 2$. This means that this epidemic peaks when the proportion of susceptibles equals $\frac{1}{R_0} = \frac{1}{2} = 0.5$. In our plot, we can roughly see that the peaks of the number of infectious people correspond with the instances where the proportion of susceptibles equals 0.5 (for example, this can be seen in the second instance of the disease during year 71 below).

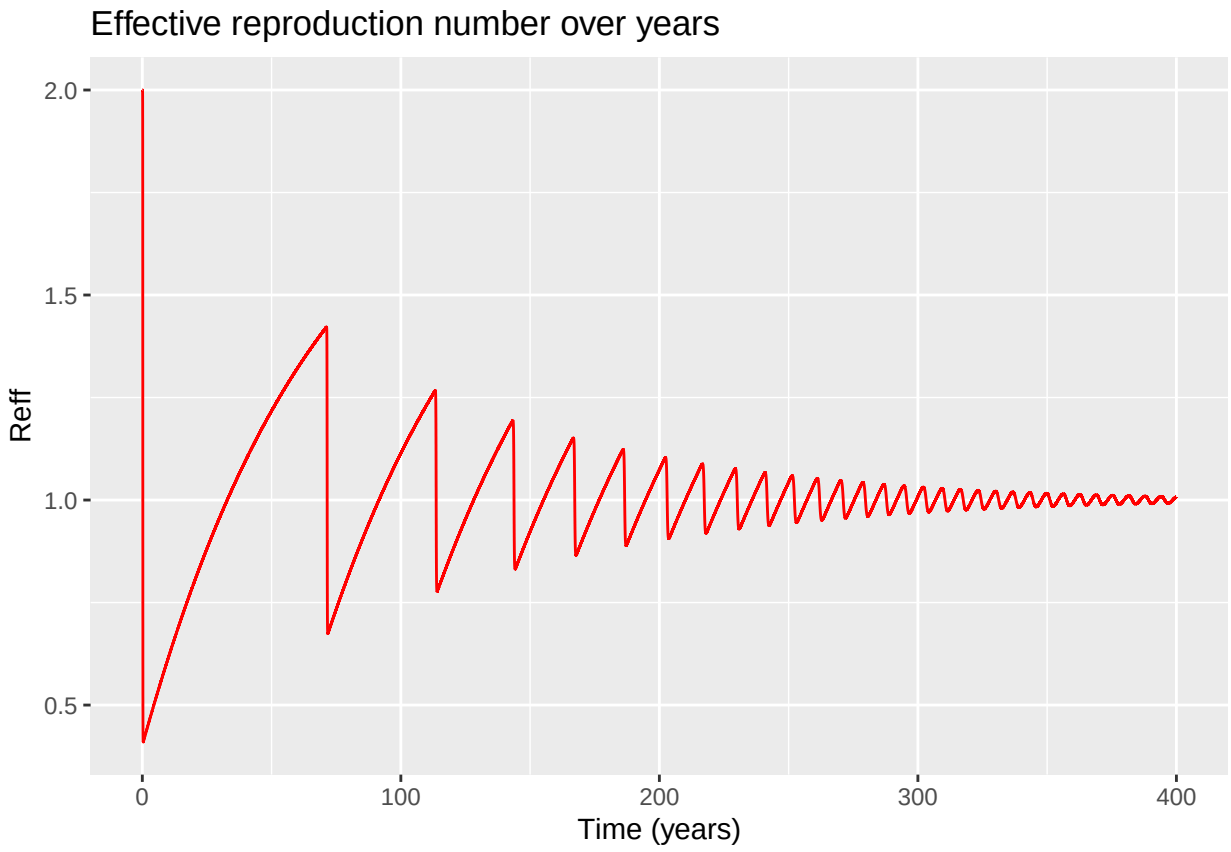
```
ggplot(data = output_long1,
       aes(x = time, y = proportion, group=variable, color=variable)) +
  geom_point(aes(x=71.51,y =0.48), color="purple") +
  geom_line() +
  xlab("Time (years)") +
  ylab("Proportion of population") +
  labs(color = "Compartment", title = "Proportion of population susceptible, recovered, and infected ov
  theme(legend.position = "bottom") +
  xlim(c(71,72))
```

```
## Warning: Removed 436905 row(s) containing missing values (geom_path).
```



After the proportion of susceptibles falls below 0.5, the proportion is too low for each infectious case to cause at least one secondary case on average. Thus, the prevalence of infection decreases and the epidemic ends, only until the pool of susceptibles is replenished as needed through births. We can plot R_{eff} to see if it follows the same pattern as the proportion of people in the susceptible compartment, and we see that it does:

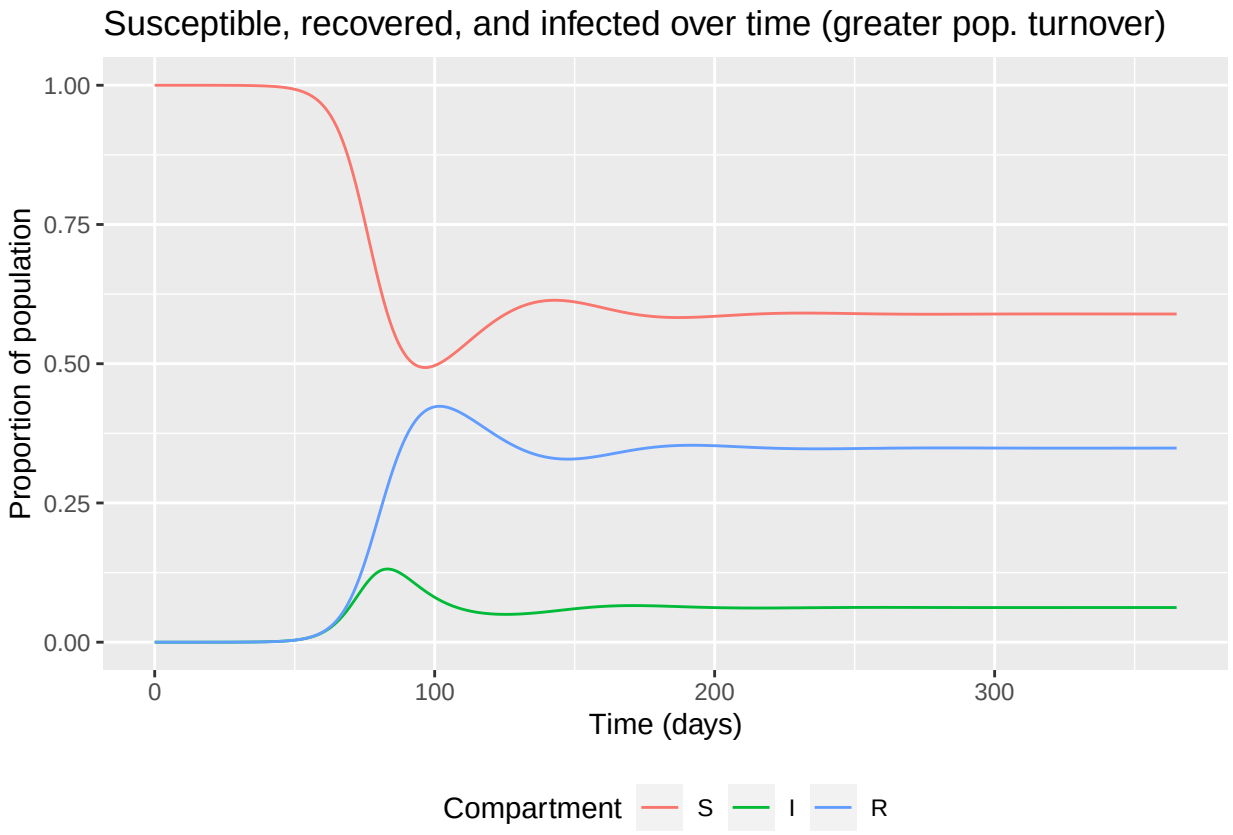
```
output1$reff <- (parameters["beta"]/parameters["gamma"]) * (output1$S / (output1$S + output1$I + output1

ggplot(data = output1,
       aes(x = time, y = reff)) +
  geom_line(color="red") +
  xlab("Time (years)") +
  ylab("Reff") +
  ggtitle("Effective reproduction number over years")
```



Developing a Model for Rapid Population Turnover

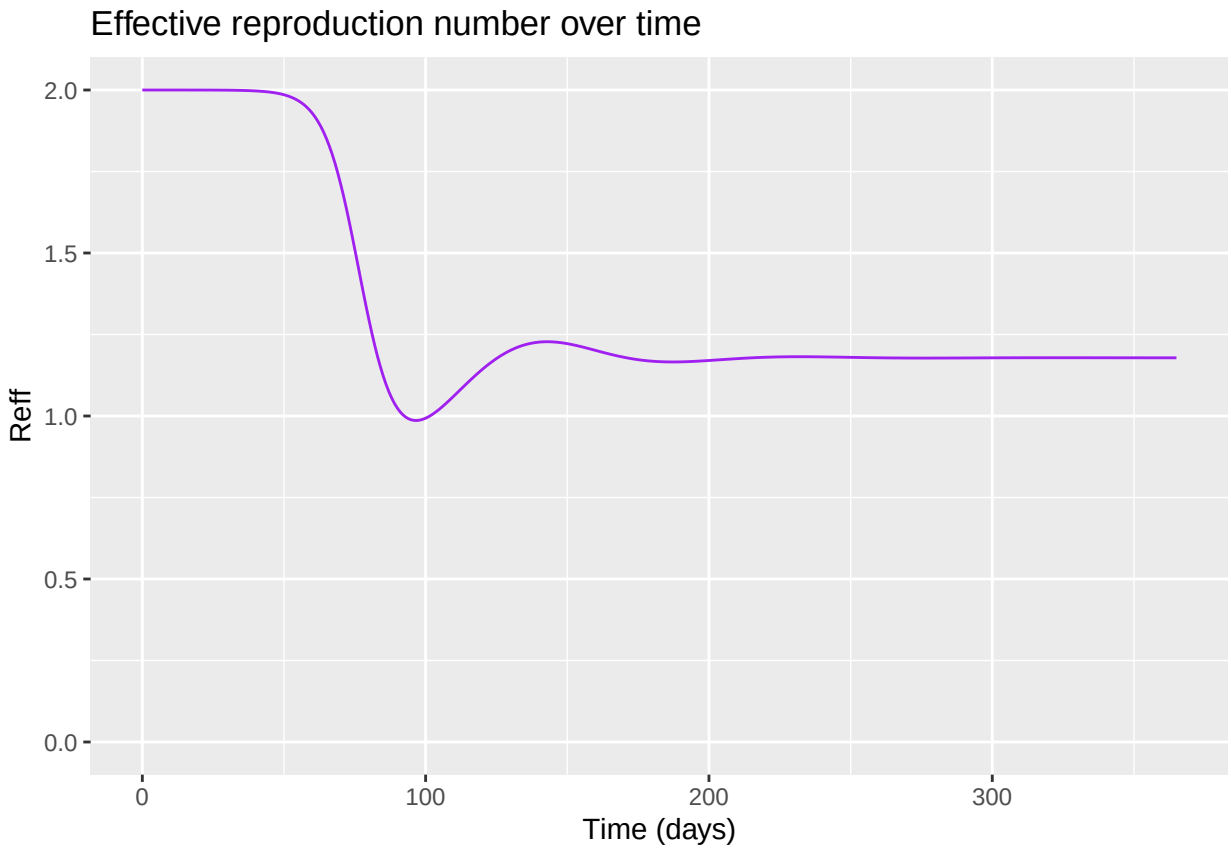
Now, let's try to model a similar acute disease with one major difference: a population with a much faster turnover. We will be using an average lifespan of 4 weeks, as opposed to 70 years. We will model this population over the course of a year with a 1-day time step.

```
parameters4 <- c(beta = 0.4, gamma = 0.2, mu = 1/28, b = 1/28)
times4 <- seq(from = 0, to = 365, by = 1)

output4 <- as.data.frame(ode(y = initial_state_values,
                             times = times4,
                             func = SIR_model,
                             parms = parameters4))

output_long4 <- melt(as.data.frame(output4), id = "time")

output_long4$proportion <- output_long4$value/sum(initial_state_values)

ggplot(data = output_long4,
       aes(x = time, y = proportion, group=variable, color=variable)) +
  geom_line() +
  xlab("Time (days)") +
  ylab("Proportion of population") +
  labs(color = "Compartment", title = "Susceptible, recovered, and infected over time (greater pop. turn
  theme(legend.position = "bottom")
```



Here, we do not witness the same peaks and troughs we observed with a smaller population turnover. In this case, an epidemic occurs after about 80 days of the introduction of the infectious case. After the peak, the prevalence of infection begins to decline; however, it does not decline to 0. In this case, the prevalence of susceptibles, infected, and recovered people reaches a kind of stable equilibrium (approximately 6% of the population remains infected). This illustrates endemicity, where the infection does not die but instead stays at a stable rate within the population. In this case, the same acute disease we saw before can become endemic due to the fast population turnover. The susceptible pool is replenished at a quick rate through new births or introductions, and infectious individuals can spur at least one secondary case on average. The population susceptible remains at around 0.5. We can also plot the effective reproduction number over time to observe similar patterns.

```
output4$reff <- (parameters4["beta"]/parameters4["gamma"]) * (output4$S / (output4$S + output4$I + outpu

ggplot(data = output4,
       aes(x = time, y = reff)) +
  geom_line(color="purple") +
  xlab("Time (days)") +
  ylab("Reff") +
  ggtitle("Effective reproduction number over time") +
  ylim(c(0,2))
```



We see here that R_{eff} over time stabilizes to about 1.17; therefore, the disease continues to survive in the population, but does not spread at a rapid rate. Each primary case, on average, only leads to about 1.17 secondary cases; each infection, on average, only really replaces itself.

Inferences about Population Turnover

Our first example is a parallel to an acute disease such as measles in a human population, where the population turnover is slow relative to the rate of infection. Therefore, as the pool of susceptibles decreased, so did the force of infection of the epidemic. As a result, the effective reproduction number, R_{eff}, reached 1 and then began to decrease. The epidemic then temporarily ceased until the pool of susceptibles could be sufficiently replenished through births (approximately 70 years due to the average human lifespan). In the second example, the population turnover was much quicker, and as a result the infection became endemic to the population. Each infected case, after the peak of the epidemic, was able to spur only about 1.17 secondary cases. Thus, the infection stabilized at a rate of approximately 6% in the population. This parallels pigs with swine flu on a farm that only stay for weeks at a time. However, the SIR model has limitations due to its simplicity. The pre-vaccination era of measles suggests that epidemic oscillations did not flatten out as they did in our hypothetical model. There can be additional factors in disease dynamics.

Additional Factors in Disease Dynamics

Drivers for Epidemic Cycles

Epidemic cycles can occur for a variety of reasons, including but not limited to weather and climate change, changes in social behavior or norms, seasonal transmission (ex. flu transmission in the winter, measles transmission when school is in session), and stochastic effects.

Factors that Drive Population Growth

For a population to grow, the birth rate must be higher than the death rate. In addition, factors such as migration and immigration can affect the growth of a population.

Transmitting Infection from Mother to Child

In this model, we assumed that all babies are born susceptible to the infection. However, neonatal vaccinations and maternal antibodies can give newborns immunity. Some infections can also be transmitted to a newborn from their mothers, such as HIV. To model such a disease, we need to consider two aspects: infected mothers can affect a proportion p of newborns and babies born to uninfected mothers enter the susceptible compartment. We can define the following parameters using this information:

The number of babies infected at birth:

$$births_i = pbI \tag{8}$$

And the number of babies born susceptible to the illness:

$$births_u = (1-p)bI + bS + bR \tag{9}$$

Where p is the proportion of babies born to infected numbers are infected at birth, b is the birth rate, I is the number of infected individuals, S is the number of susceptible individuals, and R is the number of recovered individuals.

From here, we can include $births_u$ and $births_i$ into our differential equations, capturing an addition into the susceptible and infected compartments over time, respectively:

$$\frac{dS}{dt} = -\beta\frac{I}{N}S - \mu S + births_u \tag{10}$$

$$\frac{dI}{dt} = \beta\frac{I}{N}S - \gamma I - \mu I + births_i \tag{11}$$

$$\frac{dR}{dt} = \gamma I - \mu R \tag{12}$$