Lab 4: Time-dependent transitions in compartmental models

Sample SOLUTIONS

Recap from last lab

Before we think about the SIR model, which is where we started this discussion, let's think about a slightly simpler model - the SI model. Recall, this is the model we said would be appropriate for HIV, since infectious individuals remain infectious throughout their lives (assuming no ART). The differencing equations for this model would be something like:

$S_t = S_{t-1} - \lambda_I S_{t-1}$

$I_t = I_{t-1} + \lambda_I S_{t-1}$

Write code to simulate from an SI model. You can assume the following:

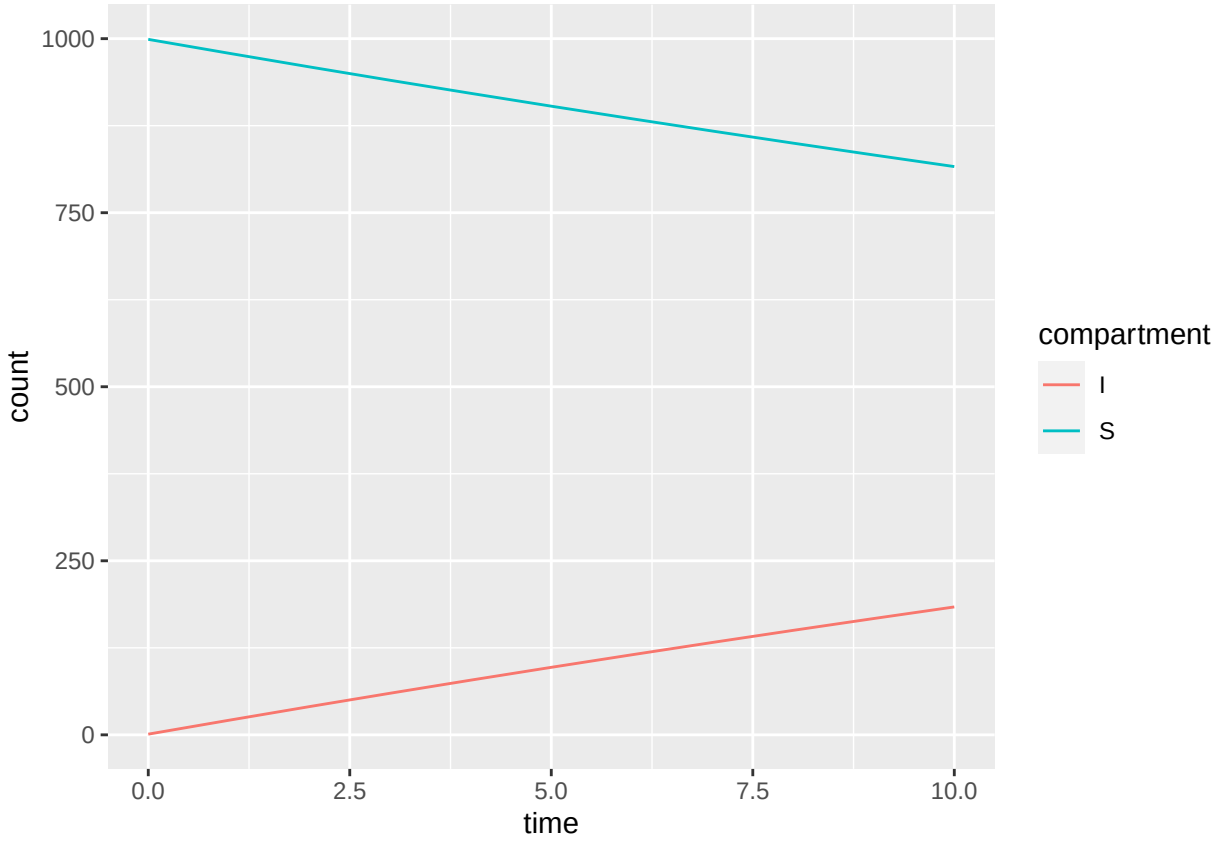
```
N <- 1000
lambdaI <- 0.02
S0 <- 999
I0 <- 1
```

You will need to define storage for S and I before writing your for loop. Assume that we observe this process for 10 time points.

```
S <- rep(NA, length(1:11))
I <- rep(NA, length(1:11))

S[1]  <- S0
I[1]  <- I0
for (t in 2:11){
  S[t] <- S[t-1]-lambdaI*S[t-1]
  I[t] <- I[t-1]+lambdaI*S[t-1]
}
```

Combine your S and I vectors into a data frame, along with a vector for time (1:10) and use ggplot to plot these two curves (one for S and one for I). In this data frame, you will want three columns, one for time, one to denote the compartment (S or I) and one to hold the count values for each of S and I. The data frame should have 20 rows and 3 columns. (22 rows if you include time 0).

```
SI_df <- data.frame(time = rep(0:10, 2),
                    compartment = rep(c("S", "I"), each=11),
                    count = c(S, I))

library(ggplot2)

ggplot(data=SI_df, aes(x=time, y=count)) +
  geom_line(aes(color=compartment))
```



Time-dependent transitions: $S \to I$

Last lab, we assumed that the parameter governing transition from $S \to I$ (λ_I) did not depend on time. More specifically, we assumed that it stayed the same, regardless of the proportion of infectious individuals in the population we are studying. Common sense tells us that the chances of getting sick should go up as more people get infected in a population, since then a susceptible person is more likely to come into contact with an infectious person (and then possibly become infectious themselves). This means we would like to modify our model to make λ_I depend on time. We are going to call this $\lambda_I(t)$, which will govern the $S \to I$ transition at time t (here $t = 0, 1, ..., 10$).

There are multiple ways to accomplish this task, including a differential equations-based approach (but that is not what we are going to do here - see Math 333 if you are interested in learning more about this from a differential equations perspective). **In this class, we are going to take a probabilistic approach to these models.**

$\lambda_I(t)$ as a transition probability

From now on, we are going to think about $\lambda_I(t)$ as the probability of transitioning from $S \to I$ at time t. This probability should depend on two things: (1) contacting an infectious individual before time t and (2) becoming infectious at time t given contact with an infectious individual prior to time t. Thus, we can think of this as:

$$\lambda_I(t) = P(\text{contact infectious individual before time } t \text{ AND become infectious at time } t)$$

How can we write this joint probability as a product of a conditional and a marginal probability?

Recall, $P(A \cap B) = P(A|B)P(B)$

$$\begin{aligned}\lambda_I(t) &= P(\text{contact infectious individual before time } t \text{ AND become infectious at time } t) \\ &= \underbrace{P(\text{become infectious at time } t|\text{contact infectious individual before time } t)}_{\text{from epidemiology literature}} \\ &\times P(\text{contact infectious individual before time } t)\end{aligned}$$

How do you think we should approximate the probability of coming in contact with an infectious person at time t?

One modeling choice:

$$P(\text{contact infectious individual before time } t) = \frac{I_{t-1}}{N}$$

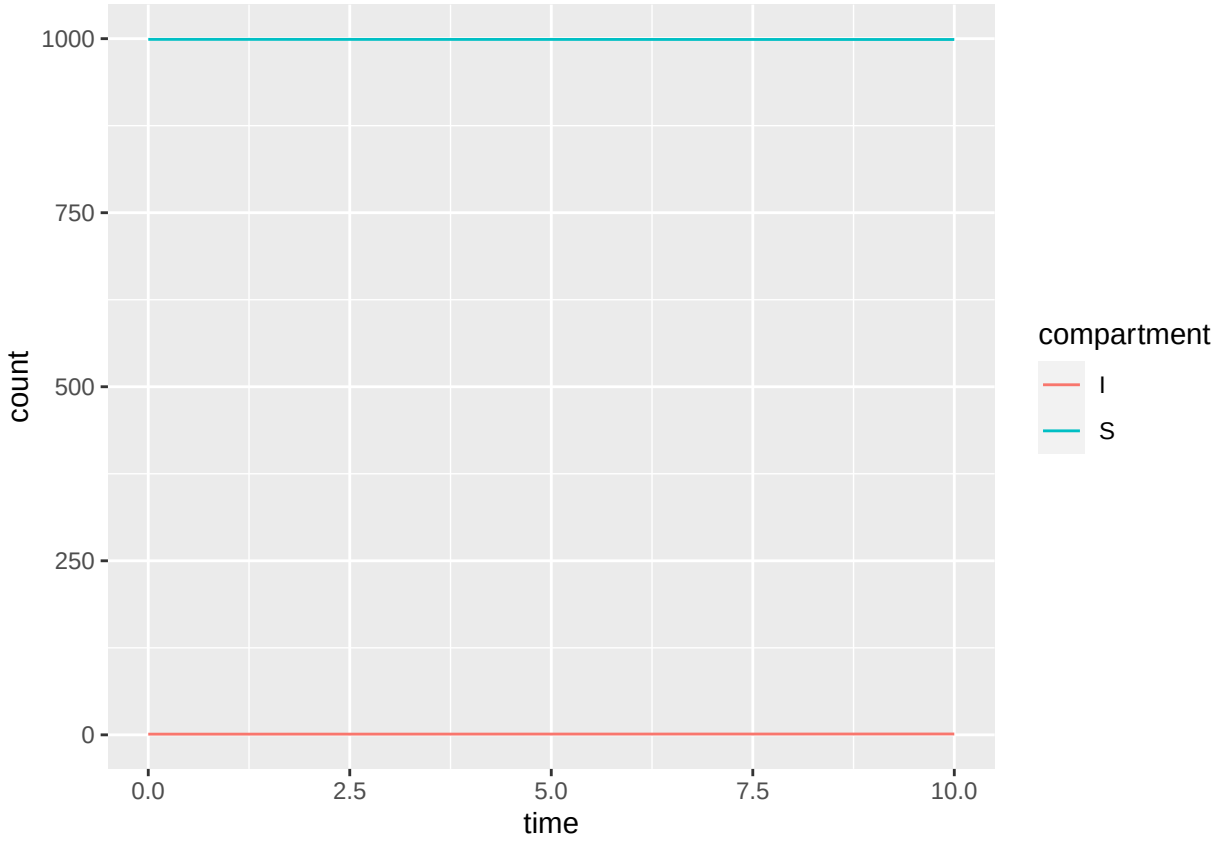
Incorporating $\lambda_I(t)$ into our for loop.

(a) Modify the code that you used in the last lab to simulate a deterministic result from the SI model, where the transition $S \to I$ depends on time, as we just discussed ($\lambda_I(t)$). You may assume that the probability of becoming infectious given a contact is $p^c(t) = 0.02$.

```
S <- rep(NA, length(1:11))
I <- rep(NA, length(1:11))
pc_t <- 0.02
lambdaI <- rep(NA, length(1:11))

S[1]  <- S0
I[1]  <- I0
for (t in 2:11){
  lambdaI[t] <- pc_t*I[t-1]/N
  S[t] <- S[t-1]-lambdaI[t]*S[t-1]
  I[t] <- I[t-1]+lambdaI[t]*S[t-1]
}
```

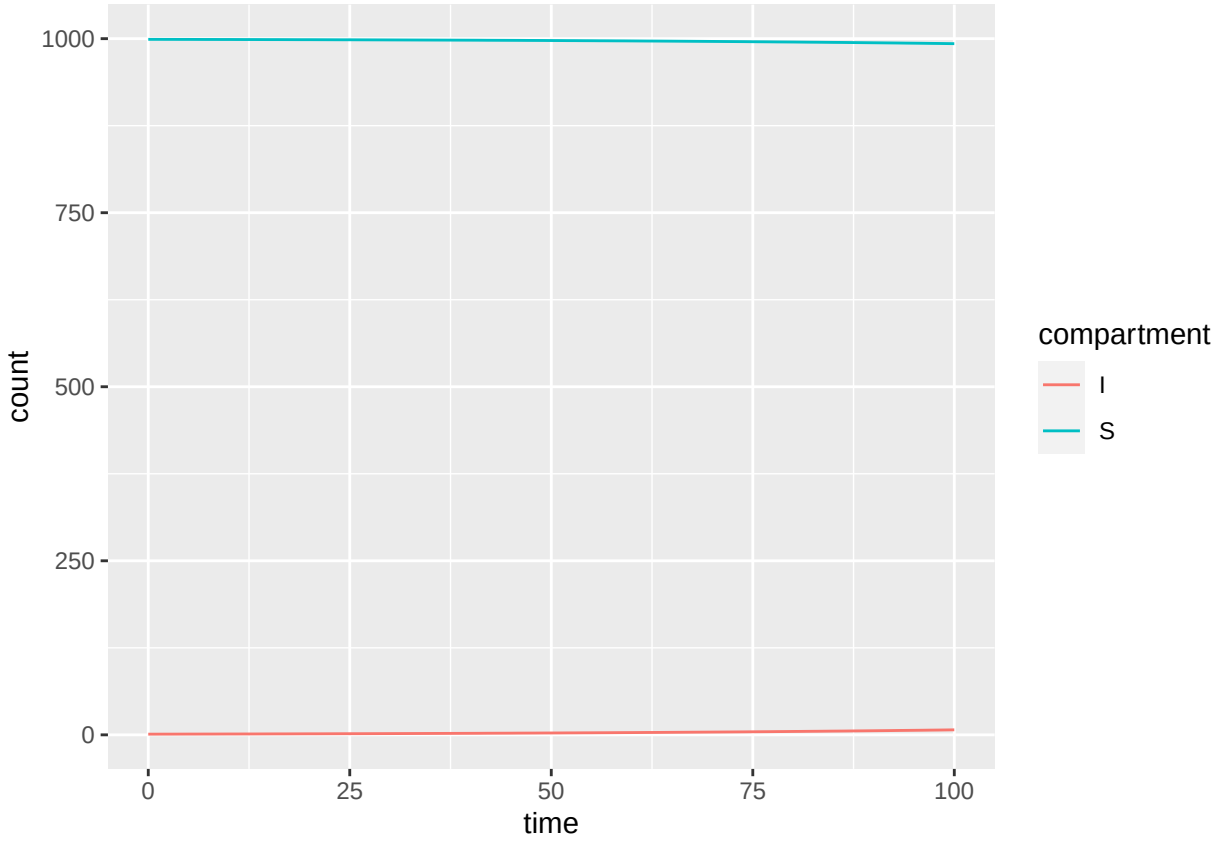
(b) Combine your S and I vectors into a data frame, along with a vector for time (1:10) and use ggplot to plot these two curves (one for S and one for I). In this data frame, you will want three columns, one for time, one to denote the compartment (S or I) and one to hold the count values for each of S and I. The data frame should have 20 rows and 3 columns. (22 rows if you include time 0).

```
SI_df_t <- data.frame(time = rep(0:10, 2),
                      compartment = rep(c("S", "I"), each=11),
                      count = c(S, I))

library(ggplot2)

ggplot(data=SI_df_t, aes(x=time, y=count)) +
  geom_line(aes(color=compartment))
```

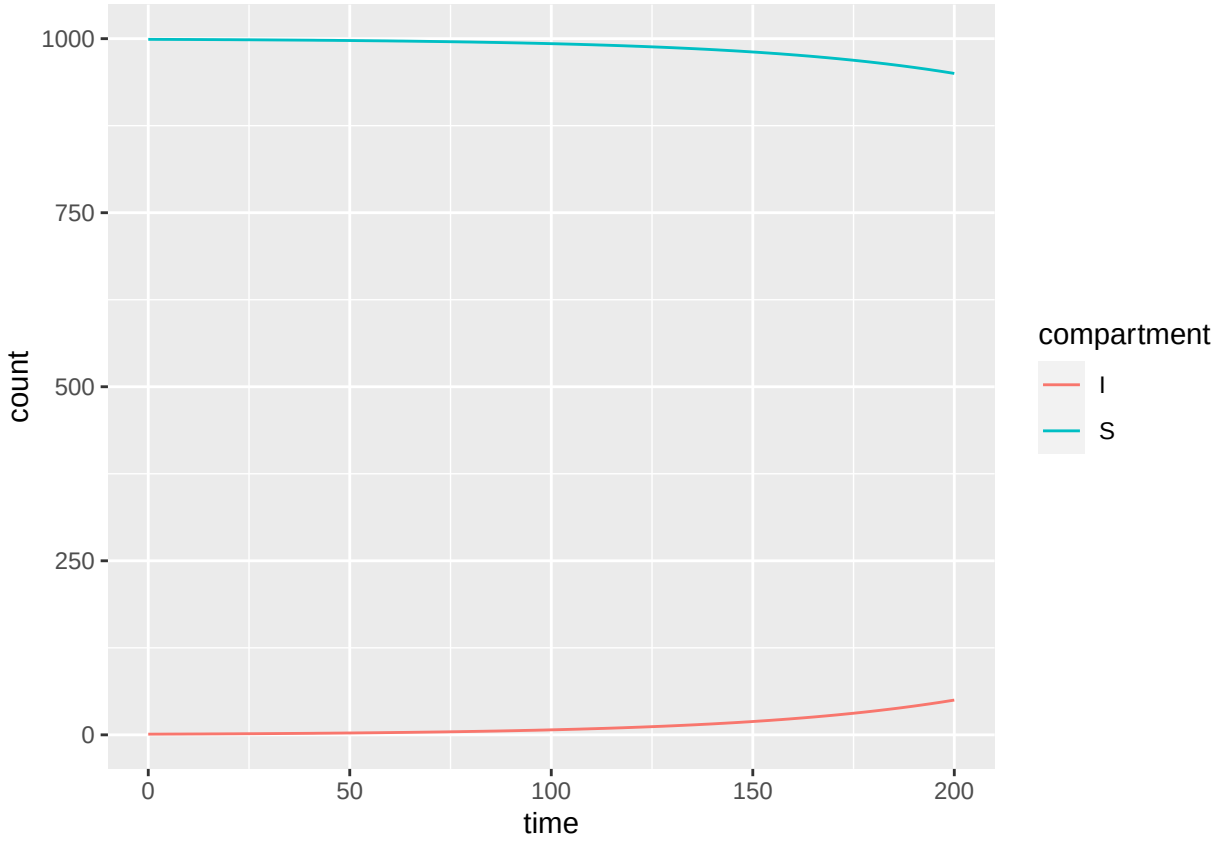


(c) Repeat (a) and (b) for different lengths of time (try more than 10 time points, like 20, 30, etc.) How do your results change?

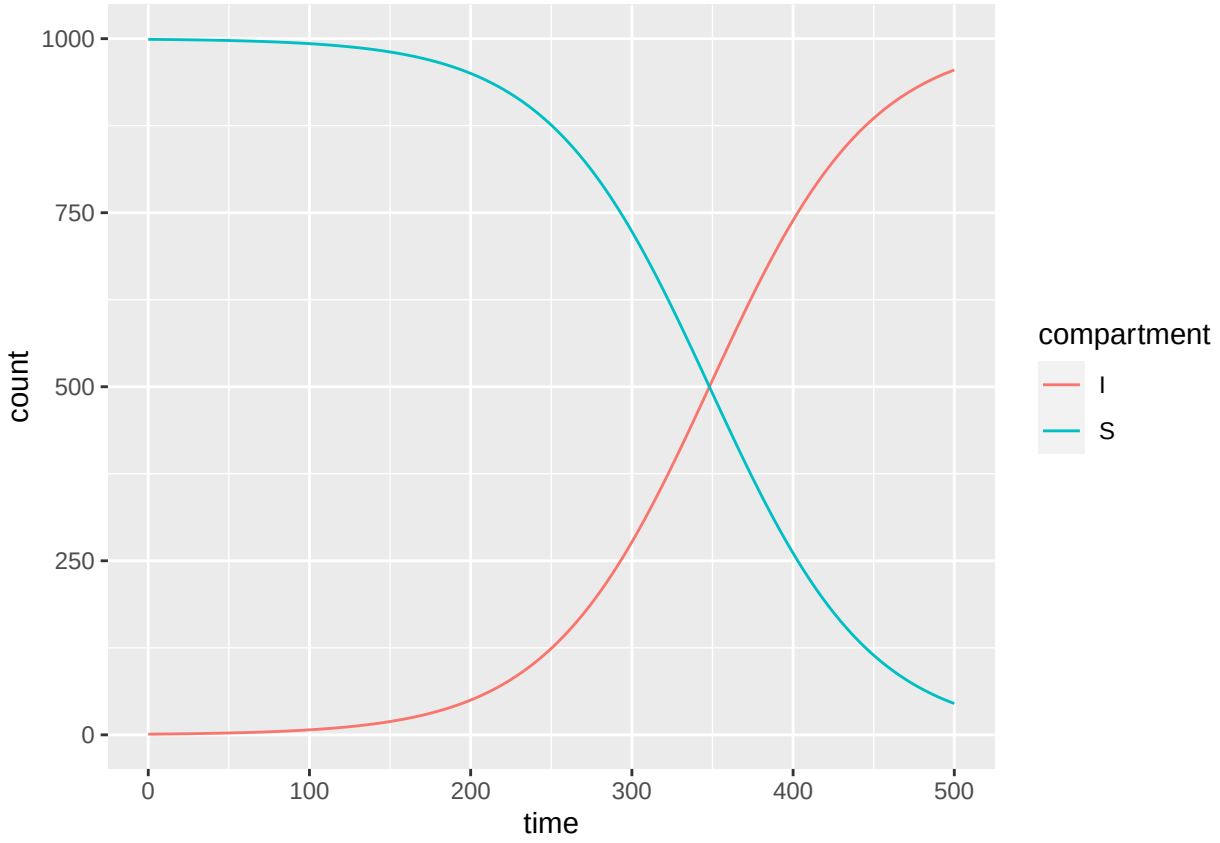
```
## 100 time points
maxTime <- 100

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  pc_t <- 0.02
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_100 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_100, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```




```
## 200 time points
maxTime <- 200

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  pc_t <- 0.02
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_200 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_200, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```

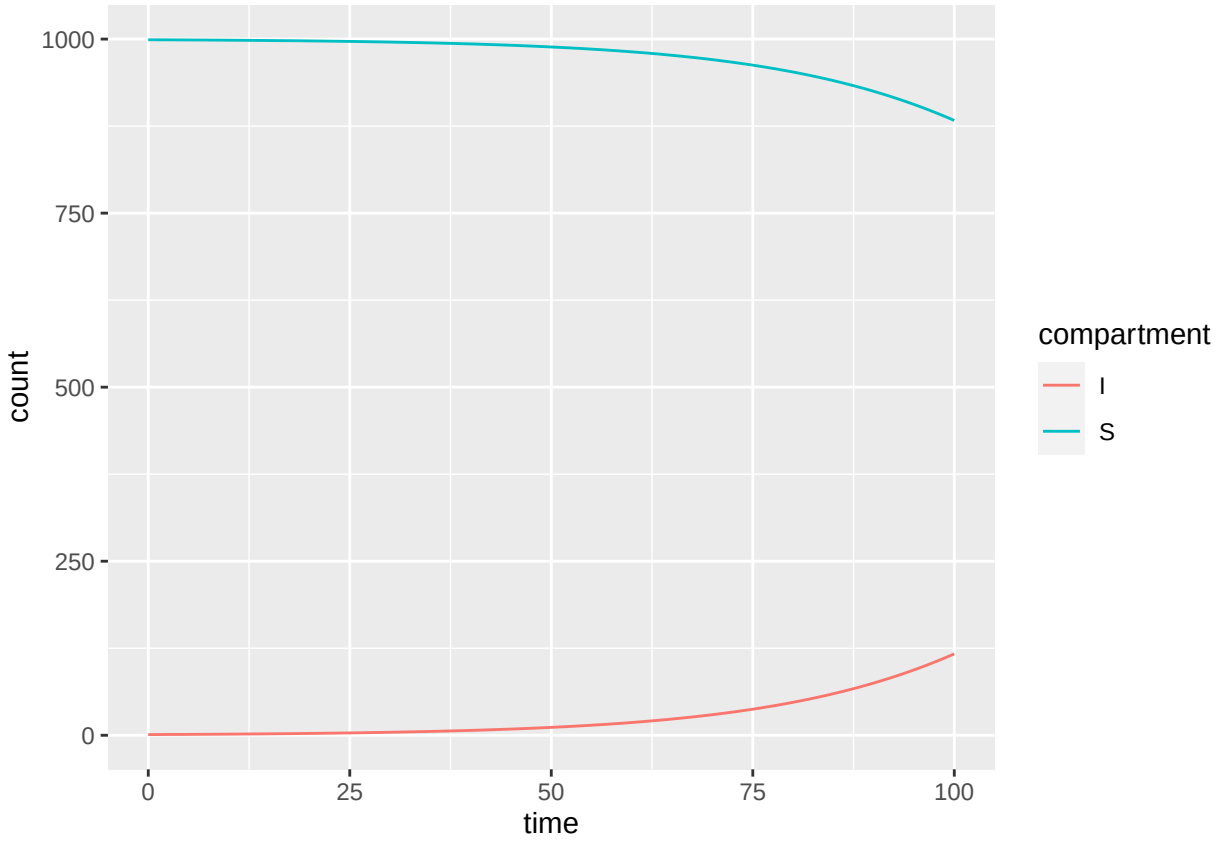



```
  ## 500 time points
maxTime <- 500

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  pc_t <- 0.02
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_100 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_100, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```

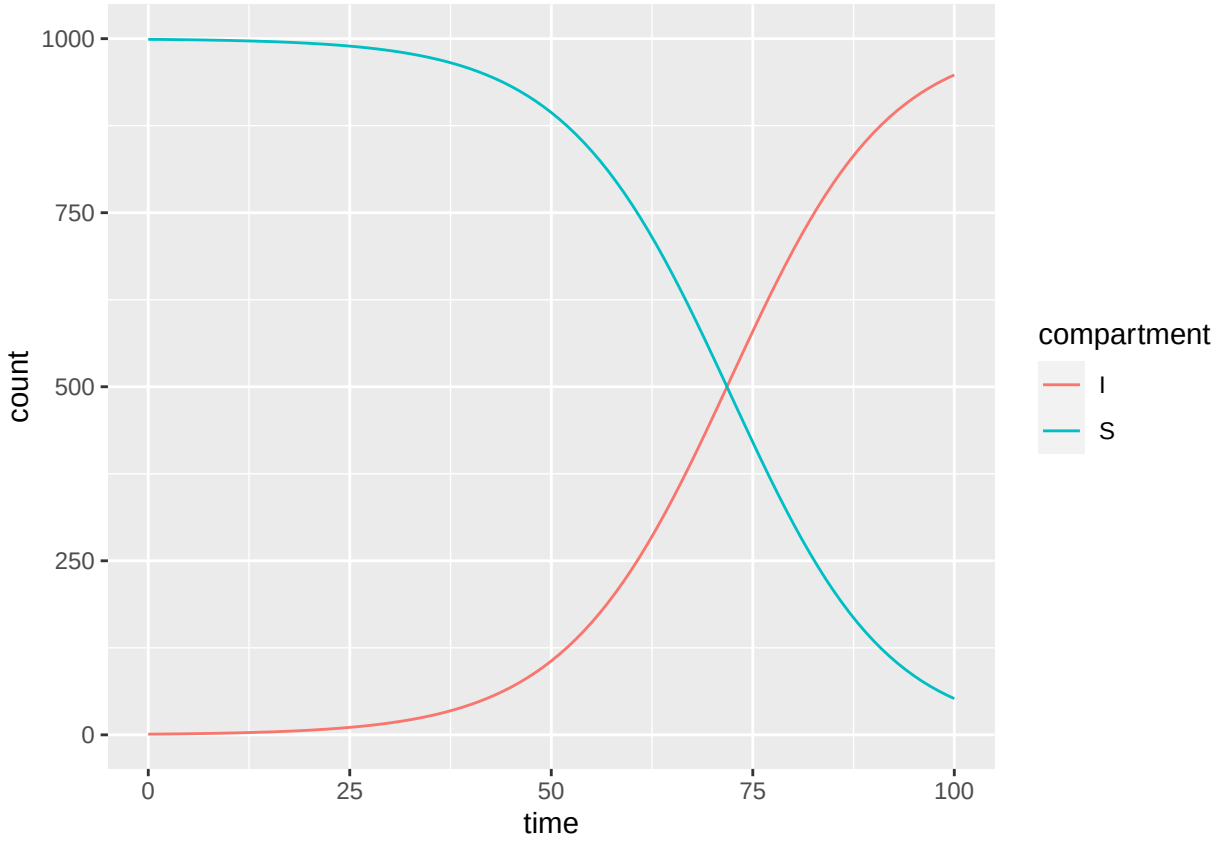


(d) Repeat (a) and (b) for different values of $p^c(t) = \{0.05, 0.10, 0.20\}$. How do your results change? (Keep the number of time points the same for all different values of $p^c(t)$)

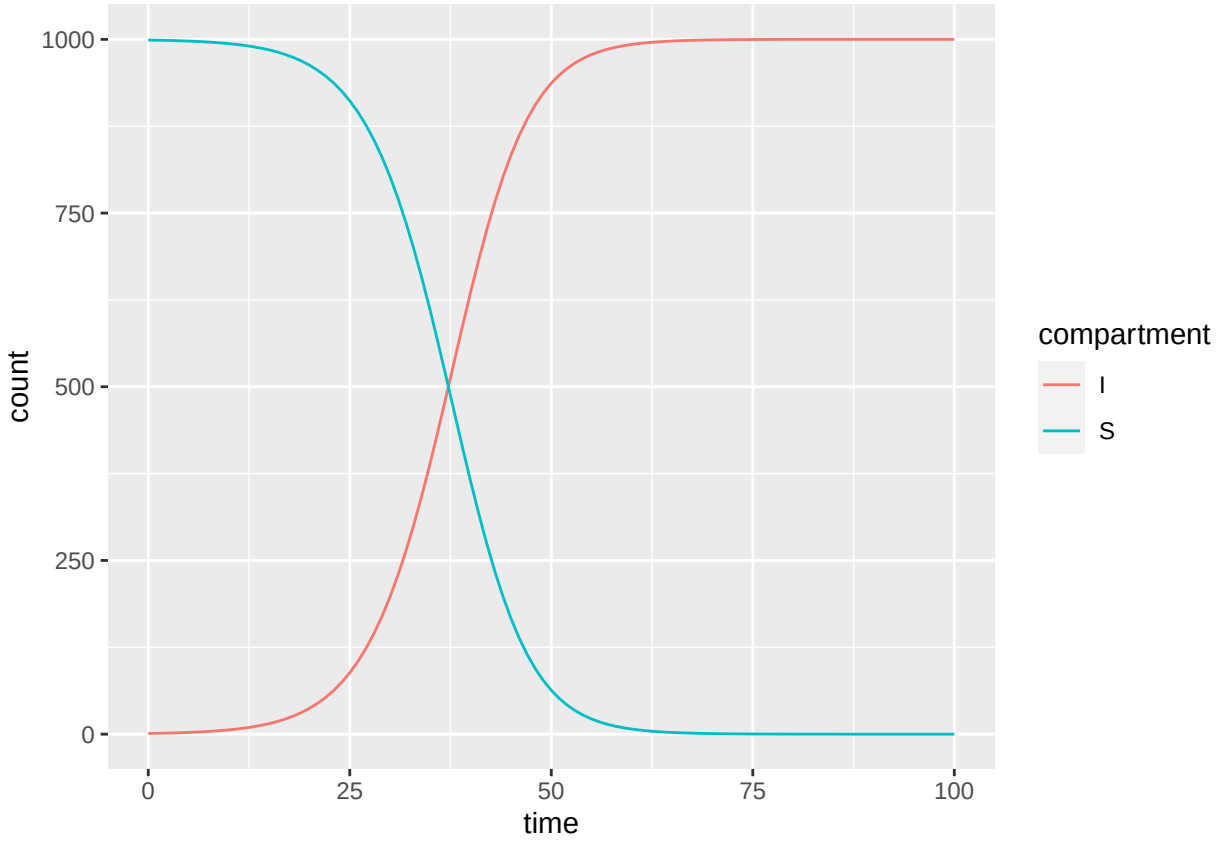
```
maxTime <- 100

## P(Infection|Contact) = 0.05
pc_t <- 0.05

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_0.05 <- data.frame(time = rep(0:maxTime, 2),
                           compartment = rep(c("S", "I"), each=(maxTime+1)),
                           count = c(S, I))

  ggplot(data=SI_df_0.05, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```



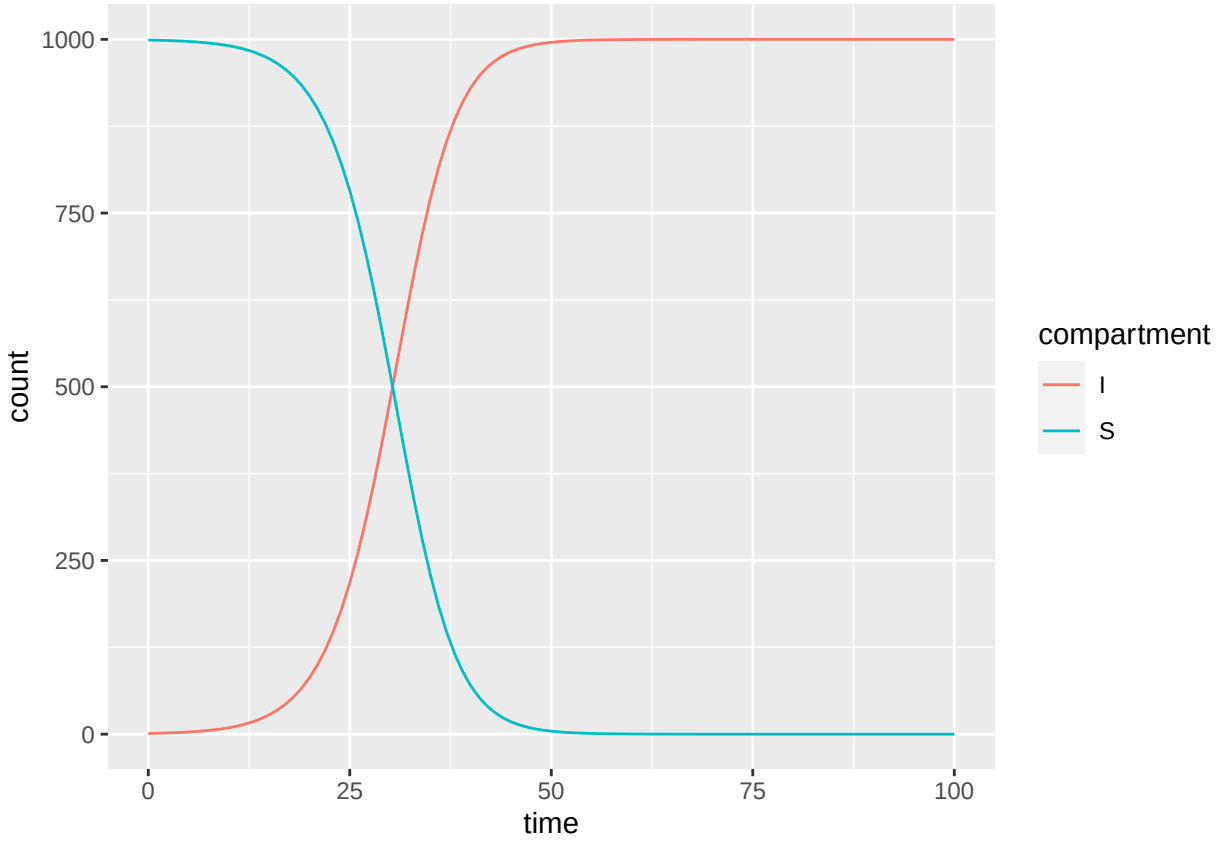

```
## P(Infection|Contact) = 0.1
pc_t <- 0.1

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_0.1 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_0.1, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```



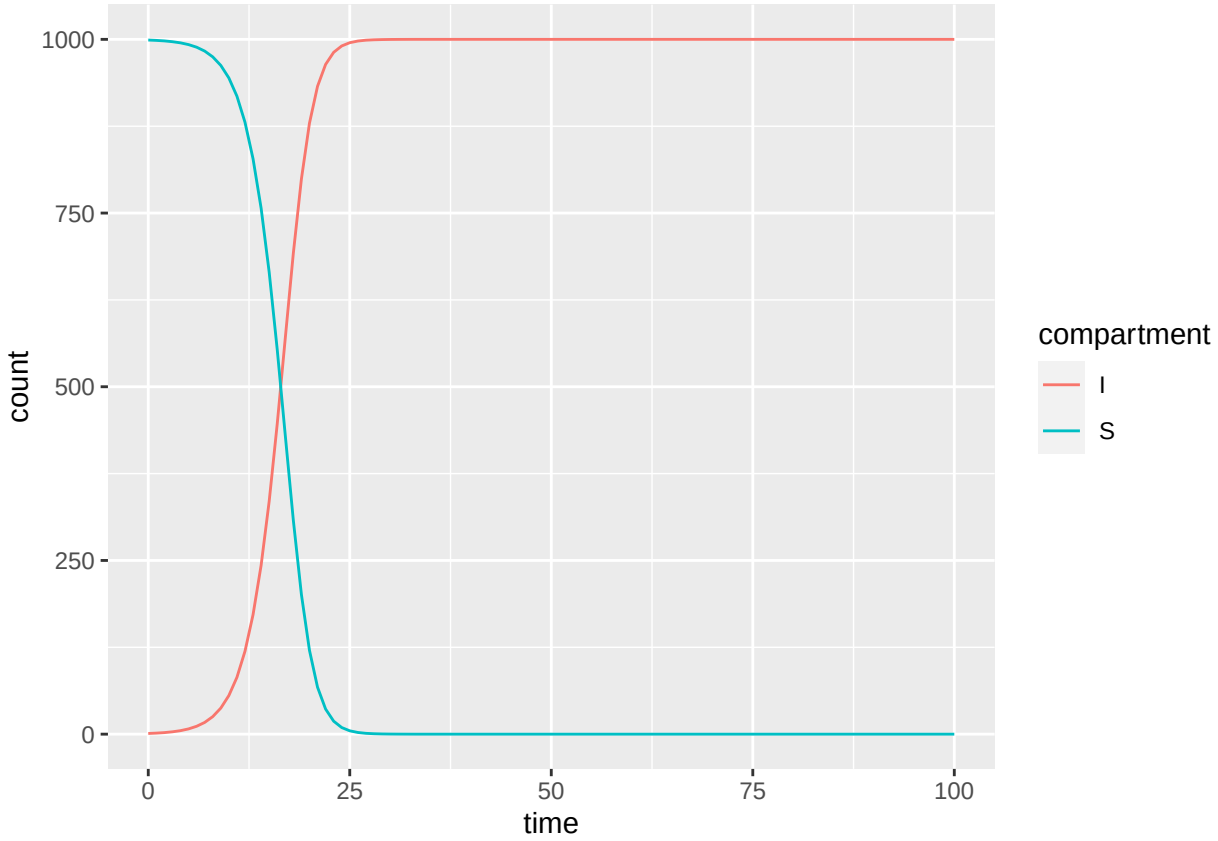

```
## P(Infection|Contact) = 0.2
pc_t <- 0.2

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_0.2 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_0.2, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```




```
pc_t <- 0.25

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_0.25 <- data.frame(time = rep(0:maxTime, 2),
                           compartment = rep(c("S", "I"), each=(maxTime+1)),
                           count = c(S, I))

  ggplot(data=SI_df_0.25, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```




```
pc_t <- 0.5

  S <- rep(NA, length(1:(maxTime+1)))
  I <- rep(NA, length(1:(maxTime+1)))
  lambdaI <- rep(NA, length(1:(maxTime+1)))

  S[1]  <- S0
  I[1]  <- I0
  for (t in 2:(maxTime+1)){
    lambdaI[t] <- pc_t*I[t-1]/N
    S[t] <- S[t-1]-lambdaI[t]*S[t-1]
    I[t] <- I[t-1]+lambdaI[t]*S[t-1]
  }

  SI_df_0.5 <- data.frame(time = rep(0:maxTime, 2),
                          compartment = rep(c("S", "I"), each=(maxTime+1)),
                          count = c(S, I))

  ggplot(data=SI_df_0.5, aes(x=time, y=count)) +
    geom_line(aes(color=compartment))
```



(e) In our model, we are assuming one contact (with an infectious person), on average, per time interval (like a week). Do you think this is reasonable? In what situations might this be reasonable, and in what situations would this likely be unreasonable? If you are looking for ideas, you may want to think about transmission mode.

For a respiratory infection that is airborne and thus transferred from person to person through breathing in virus particles, this is probably not a reasonable contact process to assume. In most cases, you would imagine that you would contact more than one infectious person, or possibly the same infectious person more than once in a week, for example. Similarly for insect-borne diseases, without proper insect repellent and clothing, a person may be bitten repeatedly by an infected vector (insect in this case). Sexually-transmitted infections, on the other hand would depend heavily on sexual behavior, so that would influence what we would assume for contacts over a particular time period.

Time permitting:

(f) Suppose instead we were considering a simulation for an SIR model. How would your code change? What other information would you need to carry out this simulation?

We would need to add a compartment (R) and also a parameter (λ_R) that controls the movement from $I \rightarrow R$. We would need an initial value for the number of individuals in R at time 0, and a reasonable value for λ_R. Storage would thus change to accommodate these changes. Our plots would have a third curve for recovered.