

# Epidemiology – Human, Disease Vectors, Water Borne Illnesses

Marc Los Huertos

March 29, 2020

## 1 Introduction

### 1.1 Disease, Pandemics, and Mathematical Epidemiology

COVID-19 began rapidly spreading around the world in late 2019. And by the spring, Italy went into lock down, California declared a state of emergency, schools and universities around the globe suspended in person classes and events, and businesses reduced travel and pushed tele-work policies. All of this was designed to slow the spread of the disease. These efforts are broadly referred to as social distancing.

### 1.2 Disease Spread and Control

The contagious nature of diseases has been appreciated for hundreds, if not thousands of years. However, until 'germ' theory as articulated by Louis Pasteur, the mechanism that imparted infections was highly speculative.

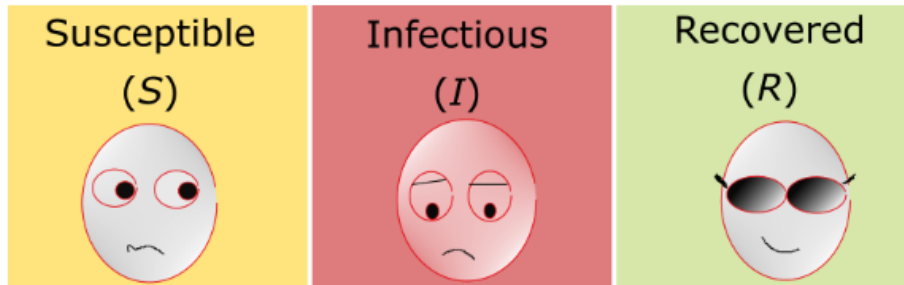
### 1.3 Value of Social Distancing

The idea is to reduce person-to-person contact to make spreading the disease less likely. The effects of this are often illustrated in images such as those in the chart below, where the red plot is flattened to spread out the disease as much as possible. This helps to ensure that there are sufficient resources available for a sick population, which will help improve survival rates.

Flattening the curve to keep infection manageable (Source: Fast.ai). How do we determine the value of such distancing strategies and model this spread?

### 1.4 Modeling Disease – SIR

Mathematical models can simulate the effects of a disease at many levels, ranging from how the disease influences the interactions between cells in a single patient (within-host models) to how it spreads across several geographically separated populations (metapopulation models).



Models simulating disease spread within and among populations, such as those used to forecast the COVID-19 outbreak, are typically based on the Susceptible - Infectious - Recovered (SIR) framework.

The SIR model is a compartmental model for modeling how a disease spreads through a population. It's an acronym for Susceptible, Infected, Recovered. When a disease is introduced to a population, the people move from one of these classes (or compartments) to the next. When they reach the R state, they're no longer able to be infected, depending on your interpretation, they either survived the disease and are now immune or succumbed to the illness and are out of the population.

SIR models are compartmental disease models. "Susceptible", "Infectious", and "Recovered" are compartments, and each individual in the population ( $N$ ) is assigned to one of these compartments. To unpack this a bit further:

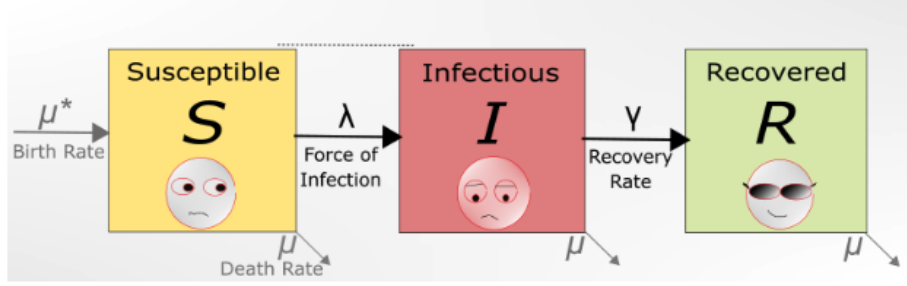
**Susceptible** individuals have no immunity to the disease (immunity can come from prior exposure, vaccination, or a mutation that confers resistance). Therefore, they can become infected. Susceptible individuals can move into the "Infectious" compartment through contact with an infectious person.

**Infectious** people have the disease and can spread it to others. Infectious individuals can move into the "Recovered" compartment by recovering from the illness.

**Recovered** individuals can no longer become infected, typically because they have immunity from a prior exposure. Many SIR-based models assume that a recovered person remains immune, which is often appropriate if immunity is long-lasting (e.g., chicken pox) or the disease is being modeled over a relatively short time period.

Because people move between compartments, the number of people in each compartment changes over time. The SIR model captures population changes in each compartment with a system of ordinary differential equations (ODEs) to model the progression of a disease.

The standard SIR model can be schematically represented as: SIR framework.png



The Kermack-McKendrick model is an SIR model for the number of people infected with a contagious illness in a closed population over time. It was proposed to explain the rapid rise and fall in the number of infected patients observed in epidemics such as the plague (London 1665-1666, Bombay 1906) and cholera (London 1865). It assumes that the population size is fixed (i.e., no births, deaths due to disease, or deaths by natural causes), incubation period of the infectious agent is instantaneous, and duration of infectivity is same as length of the disease. It also assumes a completely homogeneous population with no age, spatial, or social structure.

The model consists of a system of three coupled nonlinear ordinary differential equations,

$$\frac{dS}{dt} = -\beta SI/N \quad (1)$$

$$\frac{dI}{dt} = \beta SI/N - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I, \quad (3)$$

where  $t$  is time,  $S(t)$  is the number of susceptible people,  $I(t)$  is the number of people infected,  $R(t)$  is the number of people who have recovered and developed immunity to the infection.

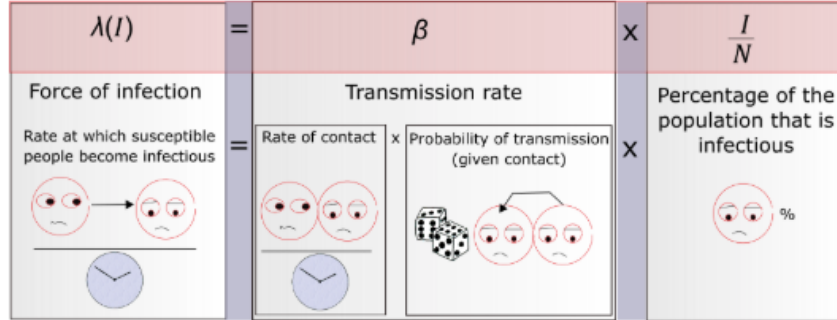
We have three ordinary differential equations in the time domain with three parameters:  $\alpha$ ,  $\beta$ ,  $\gamma$ .

- $\alpha$  is the inverse of the incubation period ( $1/t_{incubation}$ )
- $\beta$  is the average contact rate in the population or infection rate.
- $\gamma$  is the inverse of the mean infectious period ( $1/t_{infectious}$ ) or recovery rate.

This system is non-linear, however it is possible to derive its analytic solution in closed-form. Other numerical tools include Monte Carlo methods.

First, note that from:

$$\lambda(I) = \beta \frac{I}{N}.$$



$$\frac{dS}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0, \quad (4)$$

it follows that:

$$S(t) + I(t) + R(t) = \text{constant} = N, \quad (5)$$

The key value governing the time evolution of these equations is the so-called epidemiological threshold,

$$R_0 = (\beta S)/\gamma \quad (6)$$

Note that the choice of the notation  $R_0$  is a bit unfortunate, since it has nothing to do with  $R$  used for recovery.  $R_0$  is defined as the number of secondary infections caused by a single primary infection; in other words, it determines the number of people infected by contact with a single infected person before their death or recovery.







## 1.5 Modeling SIR with R

We walk through a SIR epidemiological model and simulate it with R.











```
library(deSolve)
sir <- function(time, state, parameters) {
  with(as.list(c(state, parameters)), {
    dS <- -beta * S * I
    dI <- beta * S * I - gamma * I
    dR <- gamma * I

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







$$\frac{dS}{dt} = -\lambda(I)S$$

$\frac{dS}{dt}$	= -	$\lambda(I)$	X	$S$
Rate of change of the number of susceptible people	= -	Force of infection The rate at which susceptible people become infectious	X	Number of susceptible people
#  — 		 →  — 		# 

$$\frac{dI}{dt} = \lambda(I)S - \gamma I$$

$\frac{dI}{dt}$	=	$\lambda(I)$	X	$S$	-	$\gamma$	X	$I$
Rate of change of the number of infectious people	=	Force of infection Rate at which susceptible people become infectious	X	Number of susceptible people	-	Recovery Rate Rate at which infectious people recover	X	Number of infectious people
#  — 		 →  — 		# 		 →  — 		# 

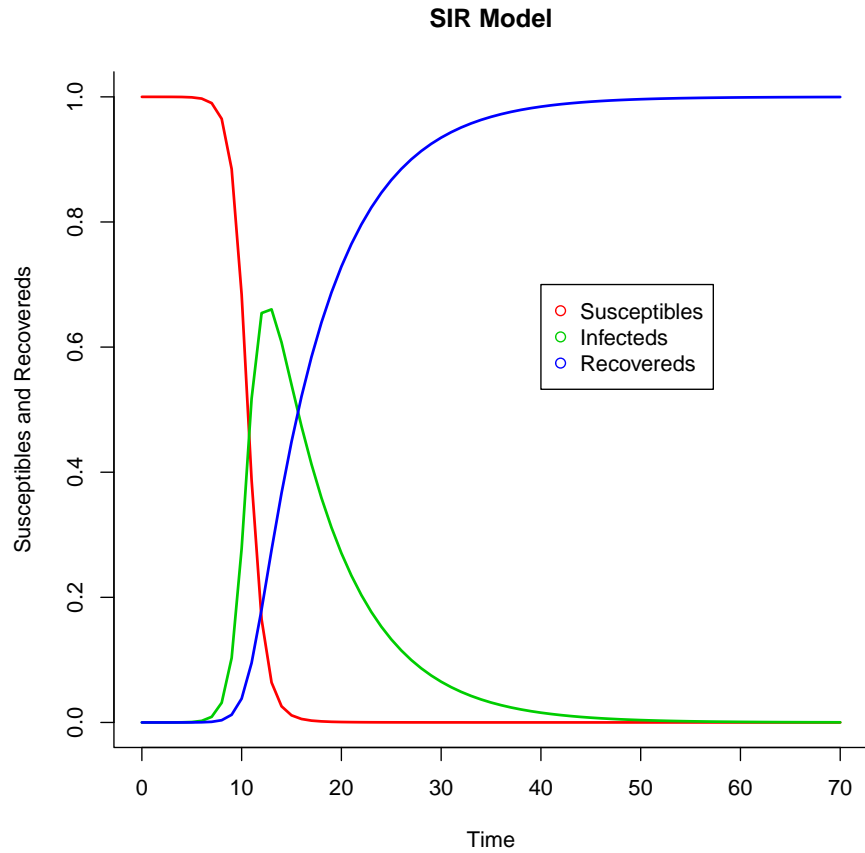
$$\frac{dR}{dt} = \gamma I$$

$\frac{dR}{dt}$	=	$\gamma$	X	$I$
Rate of change of the number of recovered people	=	Recovery rate Rate at which infectious people recover	X	Number of infectious people
#  — 		 →  — 		# 

```

init <- c(S = 1-1e-6, I = 1e-6, 0.0)
parameters <- c(beta = 1.4247, gamma = 0.14286)
times <- seq(0, 70, by = 1)
out <- as.data.frame(ode(y = init, times = times, func = sir, parms = parameters))
out$time <- NULL

```



## 1.6 Modeling Disease – SEIR

The first model is the basic SEIR without social distancing, then we add social distancing to show how the potential effectiveness of these strategies.

An extension of the classic SIR model and simply adds one more equation to show those who are exposed,

$$\frac{dE}{dt} = \beta SI/N - \sigma E \quad (7)$$

$$\frac{dI}{dt} = \sigma E/N - \gamma I \quad (8)$$

Equation 9 is the change in people susceptible to the disease and is moderated by the number of infected people and their contact with the infected. Equation 7 gives the people who have been exposed to the disease. It grows based on the contact rate and decreases based on the incubation period whereby people then become infected.

## 1.7 SEIR with R

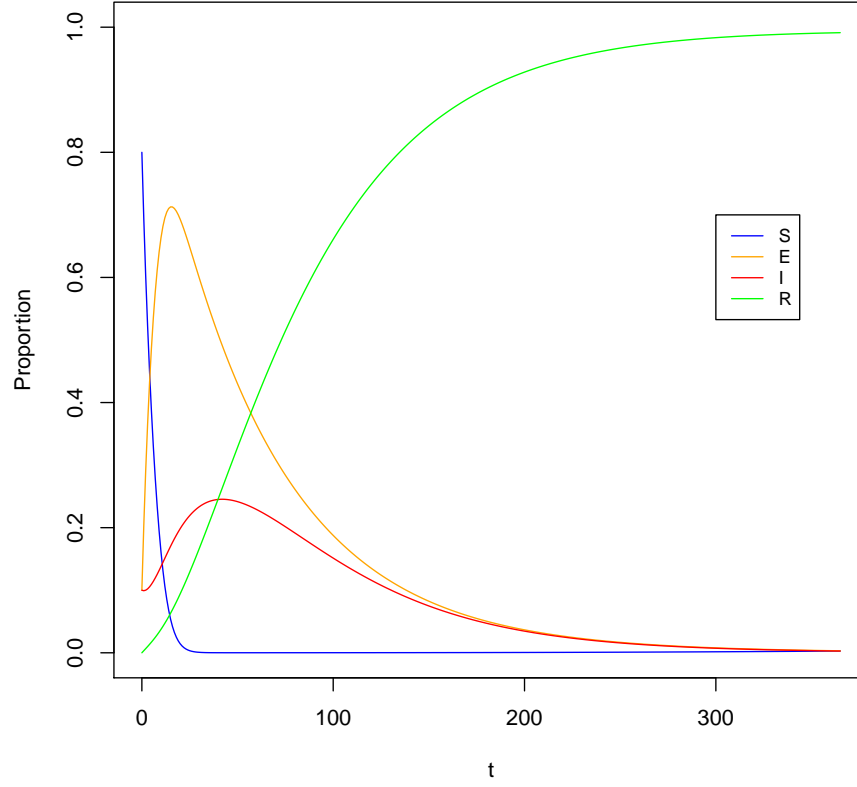
```
# Function to return derivatives of SEIR model
seir_ode<-function(t,Y,par){
  S<-Y[1]
  E<-Y[2]
  I<-Y[3]
  R<-Y[4]

  beta<-par[1]
  sigma<-par[2]
  gamma<-par[3]
  mu<-par[4]

  dYdt<-vector(length=3)
  dYdt[1]=mu-beta*I*S-mu*S
  dYdt[2]=beta*I*S-(sigma+mu)*E
  dYdt[3]=sigma*E-(gamma+mu)*I

  return(list(dYdt))
}

# Set parameter values
beta<-520/365;
sigma<-1/60;
gamma<-1/30;
mu<-774835/(65640000*365) # UK birth and population figures 2016
init<-c(0.8,0.1,0.1)
t<-seq(0,365)
par<-c(beta,sigma,gamma,mu)
# Solve system using lsoda
sol<-lsoda(init,t,seir_ode,par)
```



## 1.8 SEIR-SEI and Disease Vectors

When diseases are spread by vectors with their own dynamics, for example malaria and mosquitos, we should add additional compartments, namely susceptibility, exposure, and infection compartments for the mosquito. Mosquitos are not born with the malaria parasite, but must be infected from a host, so we can now include additional ODEs:

$$\frac{dS_v}{dt} = -\beta_v S_v I_v / N \quad (9)$$

$$\frac{dI_v}{dt} = \beta_v S_v I_v / N - \gamma I_v \quad (10)$$

$$\frac{dR_v}{dt} = \gamma I_v, \quad (11)$$



The key value governing the time evolution of these equations is the so-called epidemiological threshold and I think this can be used to link these two models, but I don't know how to yet.

$$R_0 = ? \quad (12)$$

```
# Function to return derivatives of SEIR model
seir_ode<-function(t,Y,par){
  # Human Population
  S<-Y[1]
  E<-Y[2]
  I<-Y[3]
  R<-Y[4]

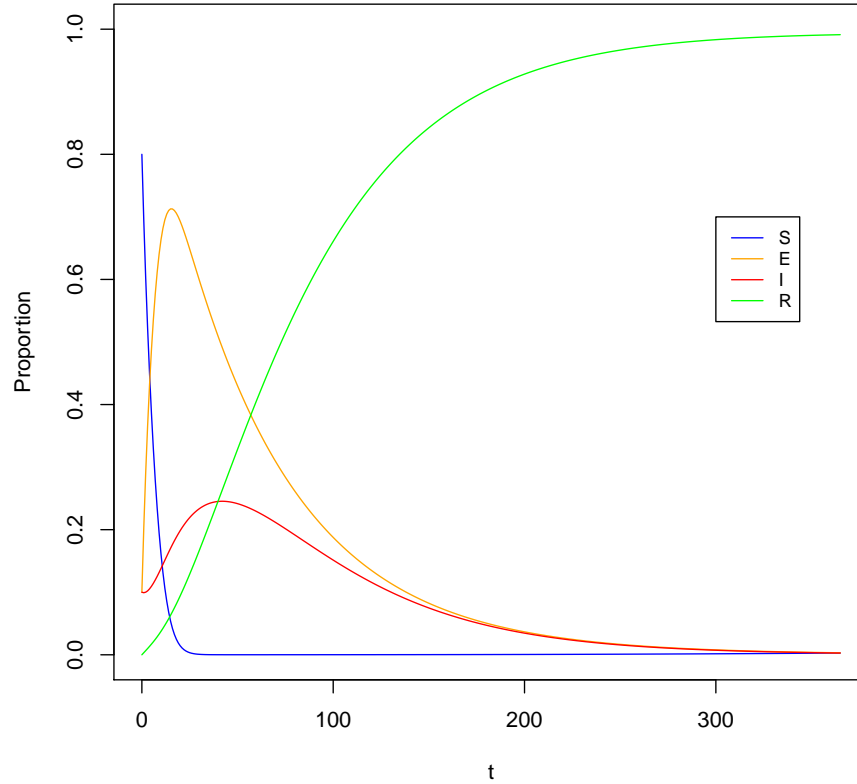
  beta<-par[1]
  sigma<-par[2]
  gamma<-par[3]
  mu<-par[4]

  dYdt<-vector(length=3)
  dYdt[1]=mu-beta*I*S-mu*S
  dYdt[2]=beta*I*S-(sigma+mu)*E
  dYdt[3]=sigma*E-(gamma+mu)*I

  return(list(dYdt))
}

# Set parameter values
beta<-520/365;
sigma<-1/60;
gamma<-1/30;
mu<-774835/(65640000*365) # UK birth and population figures 2016
init<-c(0.8,0.1,0.1)
t<-seq(0,365)
par<-c(beta,sigma,gamma,mu)
# Solve system using lsoda
sol<-lsoda(init,t,seir_ode,par)

# Plot solution
plot(t,sol[,2],type="l",col="blue",ylim=c(0,1),ylab="Proportion")
lines(t,sol[,3],col="orange")
lines(t,sol[,4],col="red")
lines(t,1-rowSums(sol[,2:4]),col="green")
legend(300,0.7,legend=c("S","E","I","R"),col=c("blue","orange","red","green"), lty=1, cex=0.8)
```



## 2 Model with Spatial Structure

### 2.1 Model Structure

The model relies on spatially explicit movements of 'subjects' that basically bounce around in a box, i.e. petri dish. Each subject is assigned a random location and trajectory (angle) of movement and velocity of movement vector. I based the model on an Washington Post article by Harry Stevens (March 14, 2020). Although his model, written in Java and is probably much cleaner, I wanted to create something that students could use to adjust parameters and evaluate different policy strategies, using a simple visualization as a foundation.

### 2.2 Building a Model

To track subject movements, contact with others, etc I have defined an array to track subject, thus, have to do some backflips to track overlap at each time

step to see if people are sharing space and pathogens.

I'm not a programmer, so this is going to need some TLC as it's pretty inefficient.

## 2.3 Model Parameterization

I have define the following parameters for the model, which will be further described in each section below.

	Model	N	Velocity	tstep	Stationary	Inf_Dist	init_infact	asympt	symp	rec
1	Model1	50.00	5.00	2000.00	0.00	5.00	2.00	24.00	96.00	720.00
2	Model2	100.00	5.00	2000.00	0.00	5.00	2.00	24.00	96.00	720.00
3	Model3	100.00	5.00	3000.00	0.40	5.00	2.00	24.00	96.00	720.00
4	Model4	100.00	5.00	3000.00	0.40	5.00	2.00	60.00	60.00	720.00
5	TBD									720.00
6	Test	5.00	5.00	20.00	0.00	300.00	1.00	24.00	96.00	720.00

But for now, I'd like to see how these align with the equations described above.

## 2.4 Model Functions

First, functions are useful when you have repetitive actions to make – in this case, I have defined functions to move subjects based on their speed and direction of travel.

Since these functions rely on the data array within another function (thus, not globally available, until it is 'returned' at the end of the function), I had to include a reference to the data in the function so that the functions could find the array.

```
#create function to move the subject
move_x = function(data_array, s, t){
  data_array[s, 4, t] * cos(data_array[s,3, t]*pi/180)}
move_y = function(data_array, s, t){
  data_array[s, 4, t] * sin(data_array[s,3, t]*pi/180)}

# check function
# move_y(data_arr, 1, 1); move_x(data_arr, 1,1)

# move function
move = function(data_array, s, t){
  c(data_array[s, 1, t-1]+ move_x(data_array, s, t-1),
    data_array[s, 2, t-1]+ move_y(data_array, s, t-1),
    data_array[s, 3, t-1],
    data_array[s, 4, t-1],
    data_array[s, 5, t-1],
```

```

    data_array[s, 6, t-1])
  }

  # Test functions
  # data_arr[1,,1]
  # move(data_arr, 1, 2)
  # Model1; i = 2

# Contact Infection Function
contact = function(data_array, t, Infect_dist){

  # testfunction
  #data_array=Model1_50; t=88; infect_dist = 5
  pairs = cbind(as.vector(data_array[, "x", t]),
                as.vector(data_array[, "y", t])); pairs
  tmp = as.matrix(dist(pairs, method = "euclidean" )); tmp
  tmp[upper.tri(tmp, diag = TRUE)] <- NA; tmp
  pairs <- which(tmp < Infect_dist ,arr.ind = TRUE);pairs
  tmp2 = cbind(as.numeric(rownames(tmp)[pairs[, 1]]),
               as.numeric(colnames(tmp)[pairs[, 2]])); tmp2
  contacts = dim(tmp2)[1]; contacts
  if(contacts>0){
    for(i in 1:contacts){
      # If one is exposed, susceptible is exposed
      if(data_array[tmp2[i,1],5,t]==2 & data_array[tmp2[i,2],5,t]==1){
        data_array[tmp2[i,2],5,t] = 2
      }
      # If one is exposed, susceptible is exposed
      if(data_array[tmp2[i,2],5,t]==2 & data_array[tmp2[i,1],5,t]==1){
        data_array[tmp2[i,1],5,t] = 2
      }
      # If one is symptomatic, the susceptible is exposed
      if(data_array[tmp2[i,1],5,t]==3 & data_array[tmp2[i,2],5,t]==1){
        data_array[tmp2[i,2], 5, t] = 2
      }
      # If one is symptomatic, the susceptible is exposed
      if(data_array[tmp2[i,1],5,t]==1 & data_array[tmp2[i,2],5,t]==3){
        data_array[tmp2[i,1], 5, t] = 2
      }
    }
  }

  # data_array[, , t]
  return(data_array)
}

```

```
# print(data_array[,5,t])
} # close function

#contact(Model1, 87, 5)
```

Next, I'm going to figure out how to create a function to test for euclidian distances to spread disease, but this will take some more effort ...

## 2.5 Model 1: Modeling Susceptability and Infections with No Behavior Changes and Recovery

For this model, people get instantenously sick and start spreading it around – not very realistic because when people get sick they are usually quarantined and stop bouncing around the environment. Nevertheless this sets up a good base line, where the poputation gets ill pretty quickly.

```
MoveSubjects = function(m=0){
  # m = 0
  N = params[m, 2]
  Velocity = params[m, 3]
  tstep = params[m, 4]
  Stationary = params[m,5];
  Infect_dist=params[m, 6];
  Init_infect=params[m, 7]
  asymp = params[m,8]
  symp = params[m,9]
  rec = params[m,10]

  # Define Data Array
  location = 6 # number of parameters to track
  data_arr = array(dim=c(N, location, tstep))
  dimnames(data_arr)[[2]] <- c("x", "y", "theta", "velocity", "status", "count")
  data_arr[, ,1]
  # Initialize Start Locations
  set.seed(2763)
  subj_x = runif(N, coord_range[1], coord_range[2]); subj_x
  subj_y = runif(N, coord_range[3], coord_range[4]); subj_y

  # Set Up Population Characteristics and Behaviors
  sheltered = sample(N*Stationary); sheltered
  velocity = rep(Velocity,N); velocity[sheltered]=0; velocity
  theta = round(runif(N, 0, 360),0); theta
  SEIR = rep(1, N); SEIR[sheltered] = 0
  SEIR[sample(N, Init_infect)] = 3; SEIR
  count = rep(0, N)
```

```

# Initial Locations
data_arr[:,1] = c(subj_x, subj_y, theta, velocity, SEIR, count)
data_arr[:,1]
# t = 2; s=1
for(t in 2:tstep){
  for(s in 1:N){
    # move subjects based on theta and speed
    data_arr[s,,t] = move(data_arr, s, t)
    # coarse corrections when hitting a boundary
    # Min x-boundary
    if(data_arr[s,1,t] < coord_range[1]){
      data_arr[s,3,t-1]=180-data_arr[s,3,t-1]
      data_arr[s,,t] = move(data_arr, s, t)
    }
    #Max x-boundary
    if(data_arr[s,1,t] > coord_range[2]){
      data_arr[s,3,t-1]=180-data_arr[s,3,t-1]
      data_arr[s,,t] = move(data_arr, s, t)
    }
    #Min y-boundary
    if(data_arr[s,2,t] < coord_range[3]){
      data_arr[s,3,t-1]=360-data_arr[s,3,t-1]
      data_arr[s,,t] = move(data_arr, s, t)
    }
    #Max y-boundary
    if(data_arr[s,2,t] > coord_range[4]){
      data_arr[s,3,t-1]=360-data_arr[s,3,t-1]
      data_arr[s,,t] = move(data_arr, s, t)
    }
    # Change Status based on Time Steps
    data_arr[s,6,t] = data_arr[s,6,t] + 1
    if(data_arr[s,5,t]==1 | data_arr[s,5,t]==0) data_arr[s,6,t]=0 # Reset Susceptables
    # Exposed becomes symptomatic
    if(data_arr[s,5,t]==2 & data_arr[s,6,t] > asymp){
      data_arr[s,5,t]=3}
    # Symptomatics move very little
    if(data_arr[s,5,t]==3 & data_arr[s,6,t] > symp){
      data_arr[s,4,t] = Velocity/2}
    # Time to Recovery
    if(data_arr[s,5,t]==3 & data_arr[s,6,t] > rec){
      data_arr[s,5,t] = 4
      data_arr[s,4,t] = Velocity
    }
  } # End of Subj Moving
}

```

```

# Time Step Proceesing
# Test for Euclidean Distances Here!

data_arr <- contact(data_arr, t, Infect_dist)

} # End of Time Step
return(data_arr)
}

```

## 2.6 Model Runs

For after running the models, I visually evaluated the models to make sure there weren't unintended consequences, i.e. usually from coding errors. Below are the models that I ran and the estimated run times.

```

library(tictoc)
#N=3, tstep=4, speed=5, init_infect=1, stationary = 0)
params

tic("Model0")
Test<- MoveSubjects(m=6)
toc() # .02s

tic("Model1")
Model1<- MoveSubjects(m=1)
toc() # 900s; 314s

tic("Model2")
Model2<- MoveSubjects(m=2)
toc() #1800s; 1158s; 1313s

tic("Model3")
Model3 <- MoveSubjects(m=3)
toc() #3308; 2437s; 2599s

tic("Model4")
Model4 <- MoveSubjects(m=4)
toc() #3308s; 2332s; 2616s

```

## 3 Visualizations

### 3.1 Tracking the Disease Transfer

```

#load("/home/CAMPUS/mwl04747/github/beginnersluck.RData")
# Susceptible 1, Exposed-Asymptomatic 2, Symptomatic 3, Recovered 4
SEIRcol = c("steelblue", "orange", "red", "green")

trackresults=function(model,m,tstep){
  results = data.frame(Days = NA, Sh = NA, S=NA, E=NA, I=NA, R=NA)
  for(t in 1:tstep){
    days = t/24
    sh = sum(as.vector(model[,5,t])==0)
    s = sum(as.vector(model[,5,t])==1, sh)
    e = sum(as.vector(model[,5,t])==2)
    i = sum(as.vector(model[,5,t])==3)
    r = sum(as.vector(model[,5,t])==4)
    results[t,] <- data.frame(Days = days, Sh = sh, S=s, E=e, I=i, R=r)
  }
  head(results)
  par(mfrow=c(1,1), mar=c(4, 5, 3, 2)+.1, las=1)
  plot(S~Days, data=results, ty='l', col=SEIRcol[1], ylab="", ylim=c(0,params[m,2]), xlab="Day")
  lines(E~Days, data=results, col=SEIRcol[2], lwd=2)
  lines(I~Days, data=results, col=SEIRcol[3], lwd=2)
  lines(R~Days, data=results, col=SEIRcol[4], lwd=2)
}

m = 6; trackresults(Test, m, params[m,4])

png(filename =
  "/home/CAMPUS/mwl04747/github/beginnersluck/Epidemiology/Model1.png");
m = 1; trackresults(Model1, m, 2000); dev.off()
png(filename =
  "/home/CAMPUS/mwl04747/github/beginnersluck/Epidemiology/Model2.png");
m = 2; trackresults(Model2, m, 2000); dev.off()
png(filename =
  "/home/CAMPUS/mwl04747/github/beginnersluck/Epidemiology/Model3.png");
m = 3; trackresults(Model3, m, 2000); dev.off()
png(filename =
  "/home/CAMPUS/mwl04747/github/beginnersluck/Epidemiology/Model4.png");
m = 4; trackresults(Model4, m, 2000); dev.off()

```

### 3.2 Static Spatial Plots

Because we are looking at a dynamic system, I don't find the static plot all that useful.



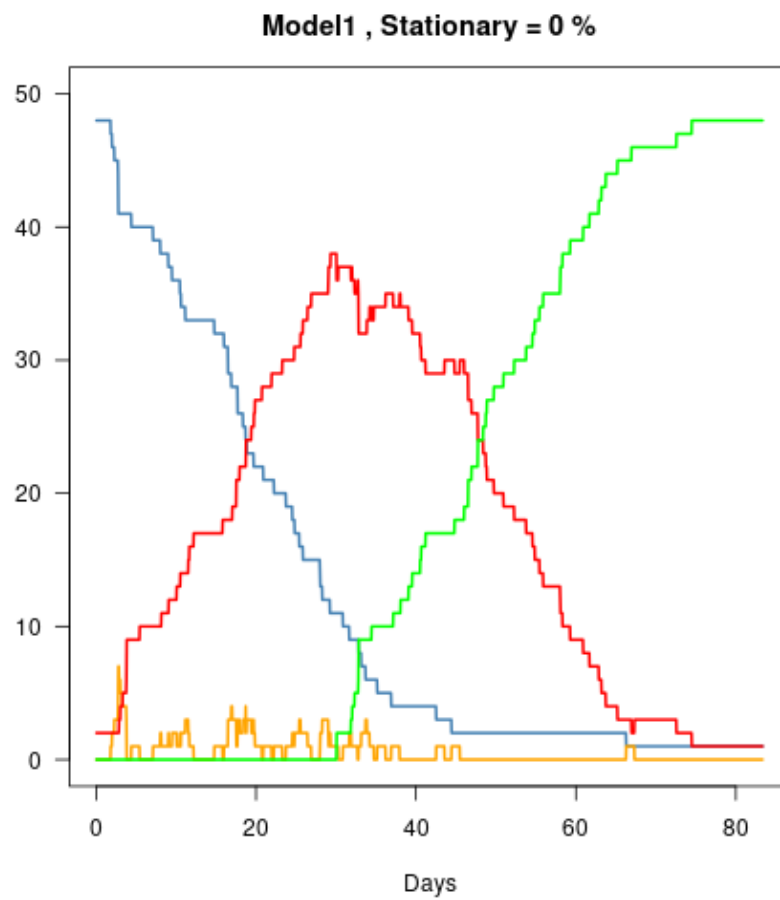


Figure 1:  $N = 50$

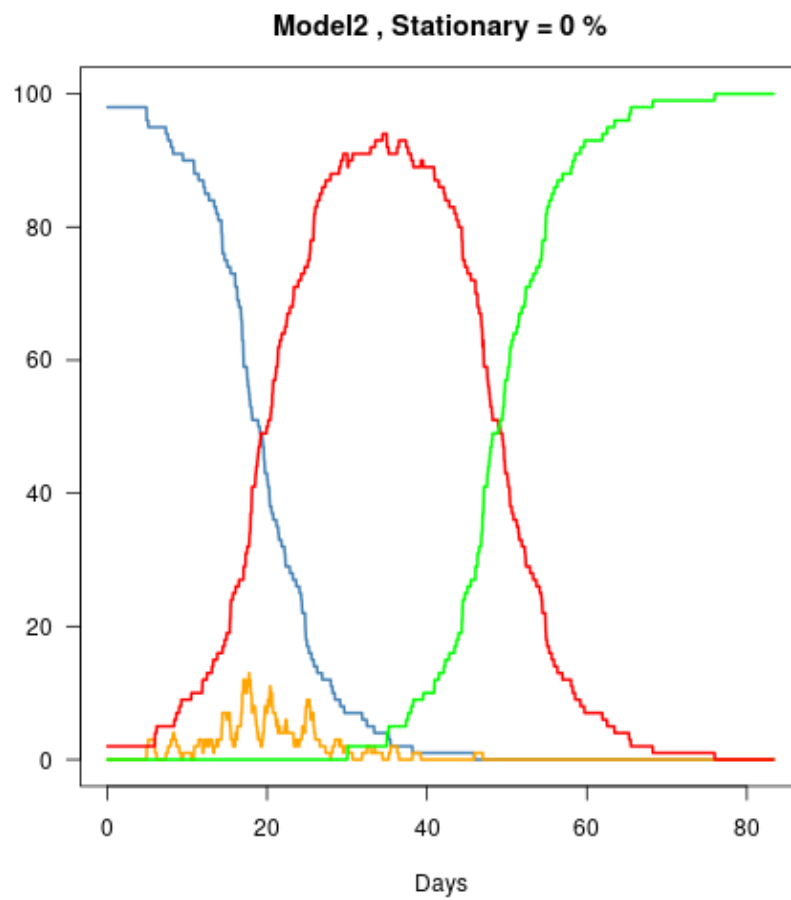


Figure 2:  $N=100$

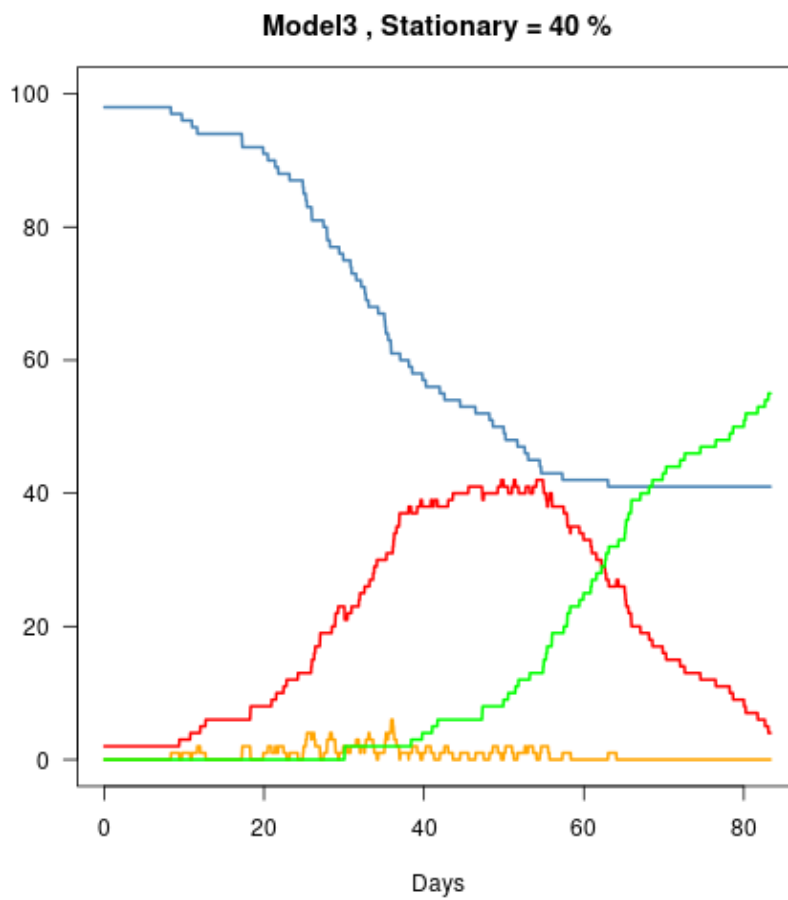


Figure 3: 40% shelter in place

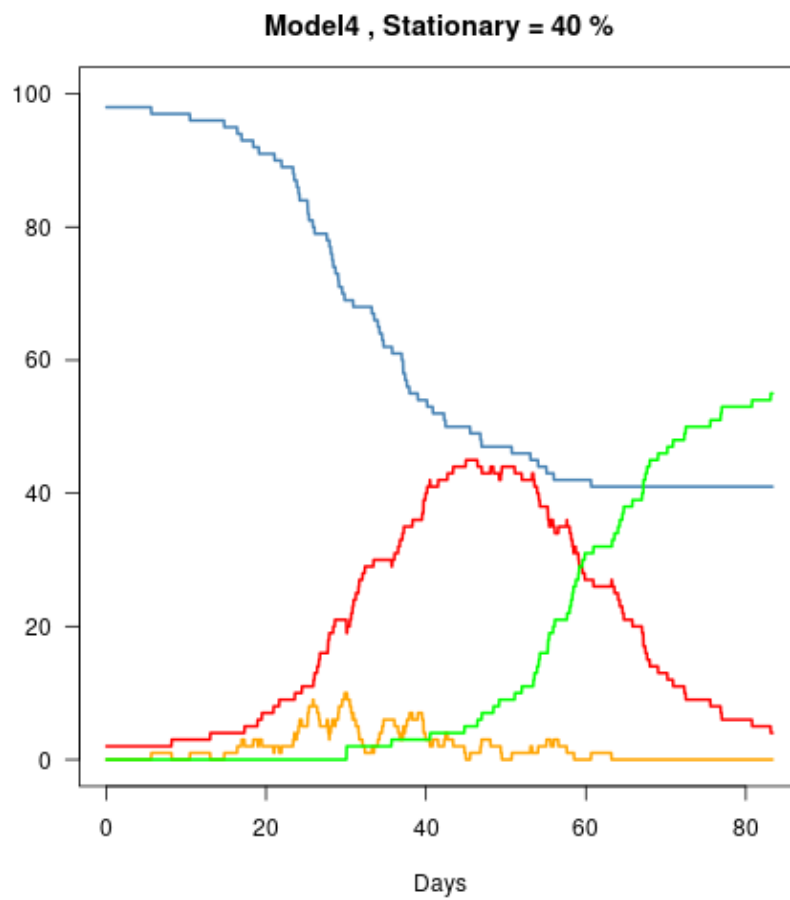


Figure 4: 40% shelter in place, exposed without symptoms.

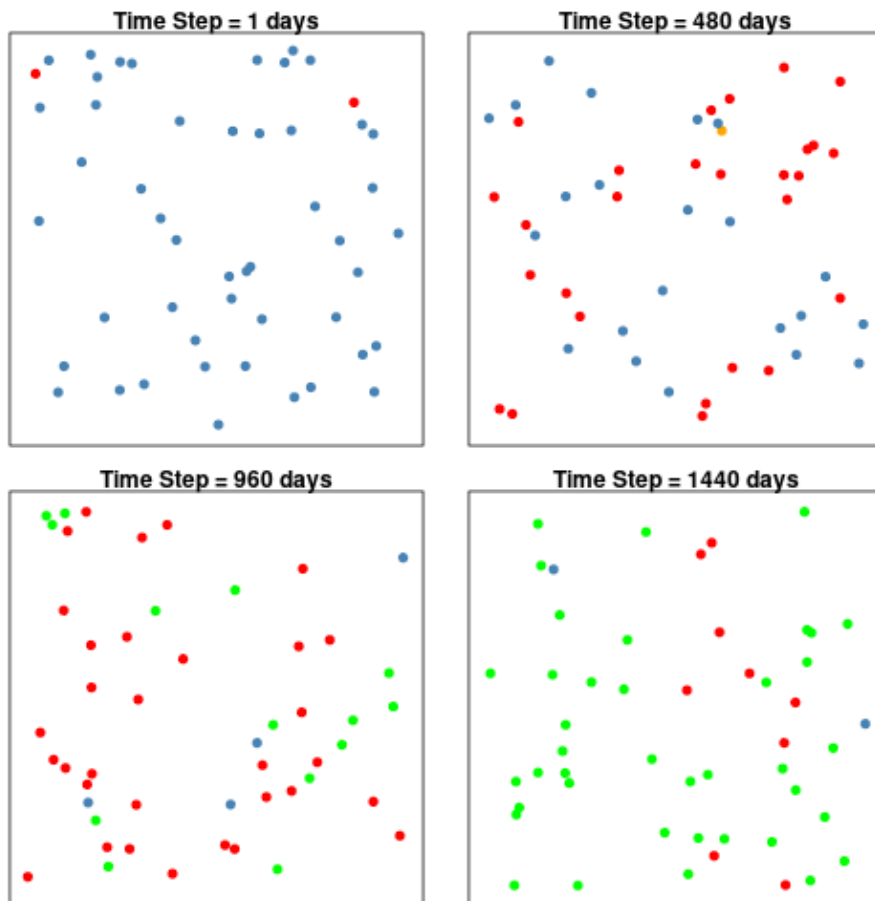


Figure 5:  $N = 50$

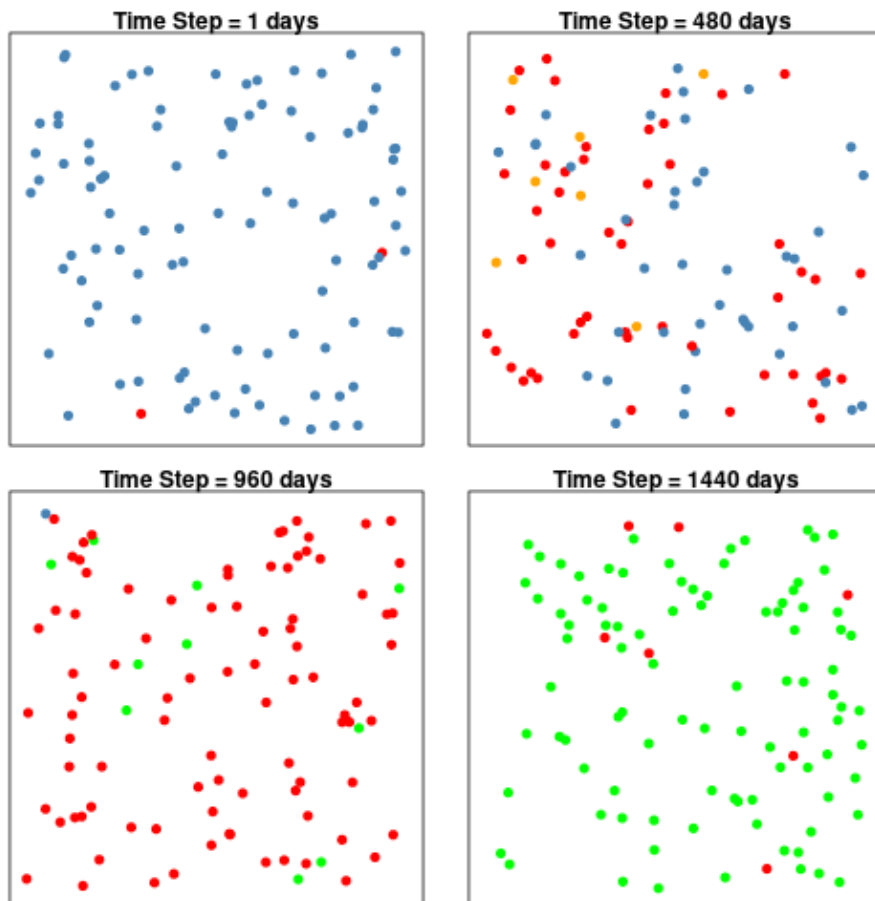


Figure 6:  $N=100$

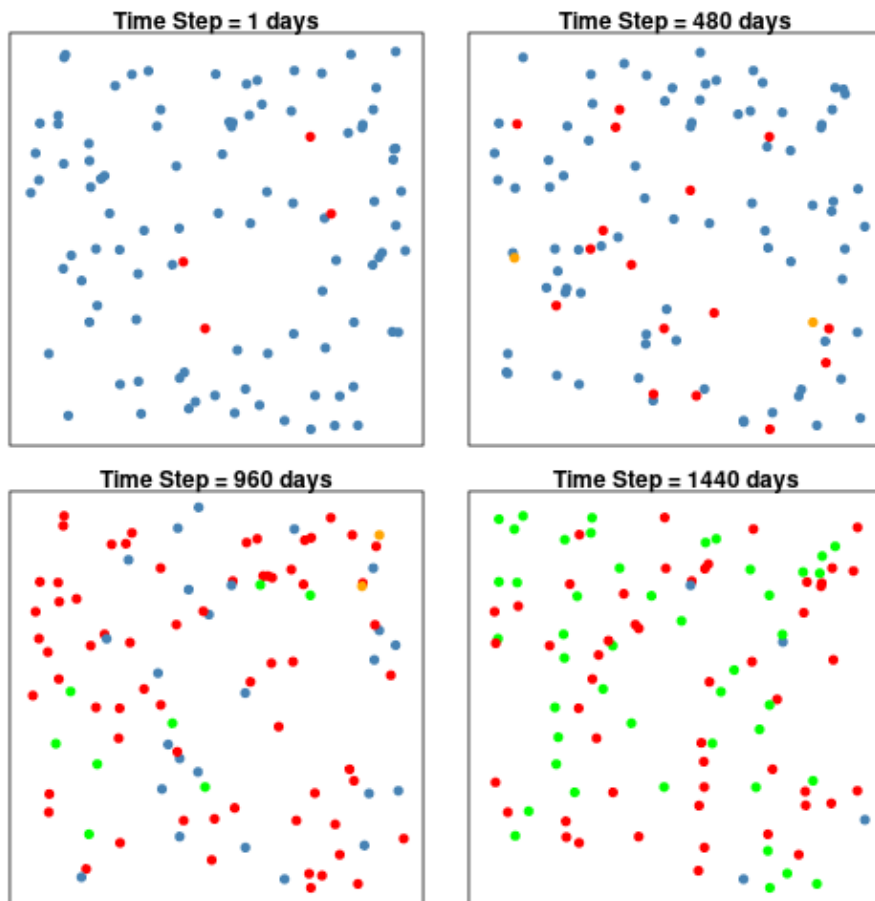


Figure 7: 40% shelter in place

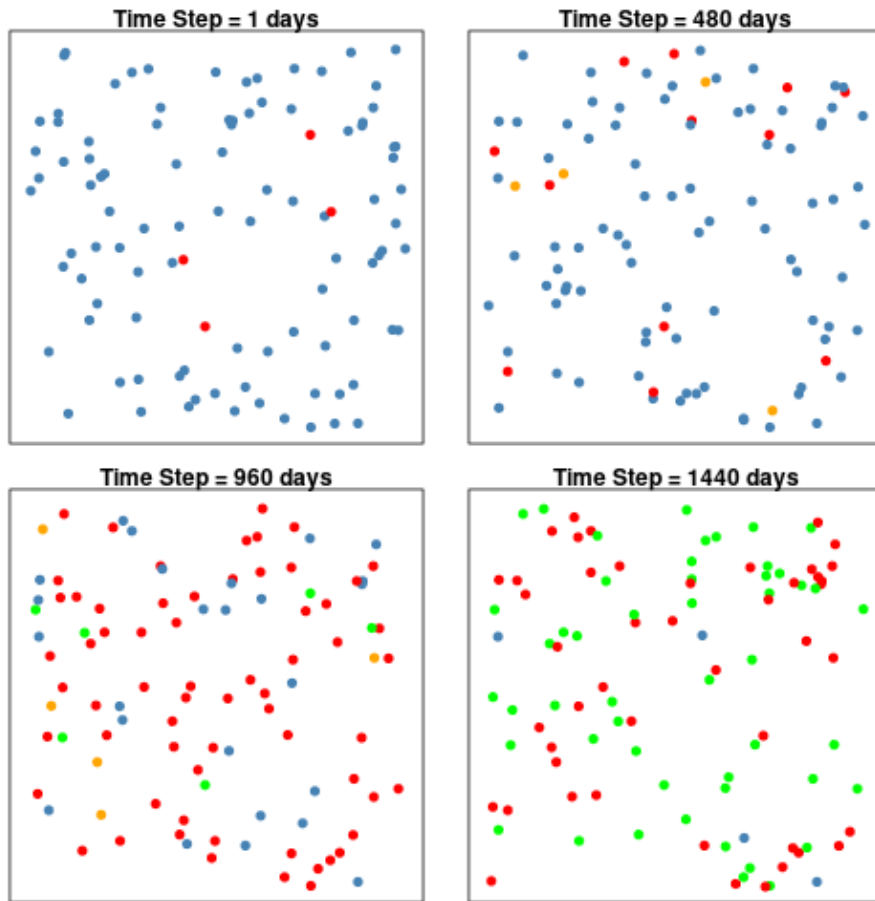


Figure 8: 40% shelter in place, exposed without symptoms.



### 3.3 Animating the Results

Still working on this... I have to export the file to my laptop then use photoshop to make a movie.

it would be nice to do this in r and then embedd in a pdf..

## 4 SEIRS with vital dynamics

You can also add vital dynamics to an SEIRS model, where  $\mu$  and  $\nu$  represent the birth and death rates, respectively. To maintain a constant population, assume that  $\mu = \nu$ . In steady state  $\frac{dI}{dt} = 0$ . The ODE then becomes:

$$\frac{dS}{dt} = \mu N - \frac{\beta SI}{N} + \xi R - \nu S \quad (13)$$

$$\frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E - \nu E \quad (14)$$

$$\frac{dI}{dt} = \sigma E - \gamma I - \nu I \quad (15)$$

$$\frac{dR}{dt} = \gamma I - \xi R - \nu R \quad (16)$$

where  $N = S + E + I + R$  is the total population.

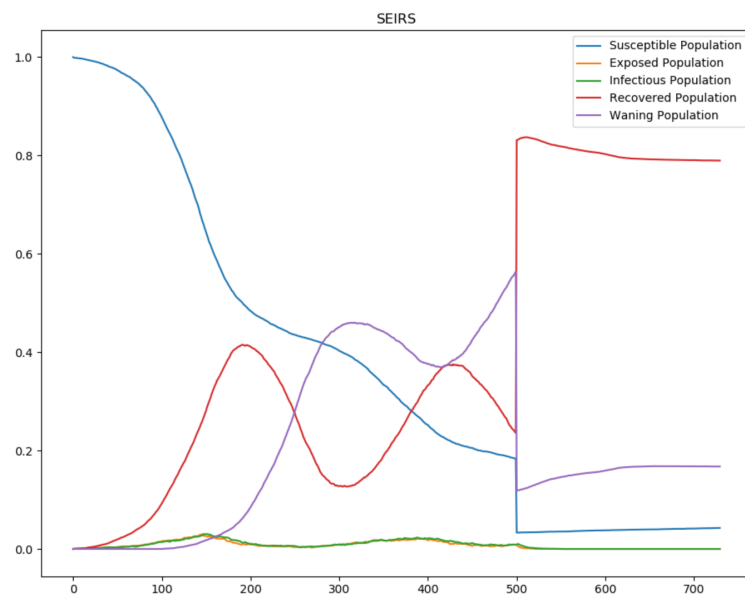


Figure 9: SEIRS