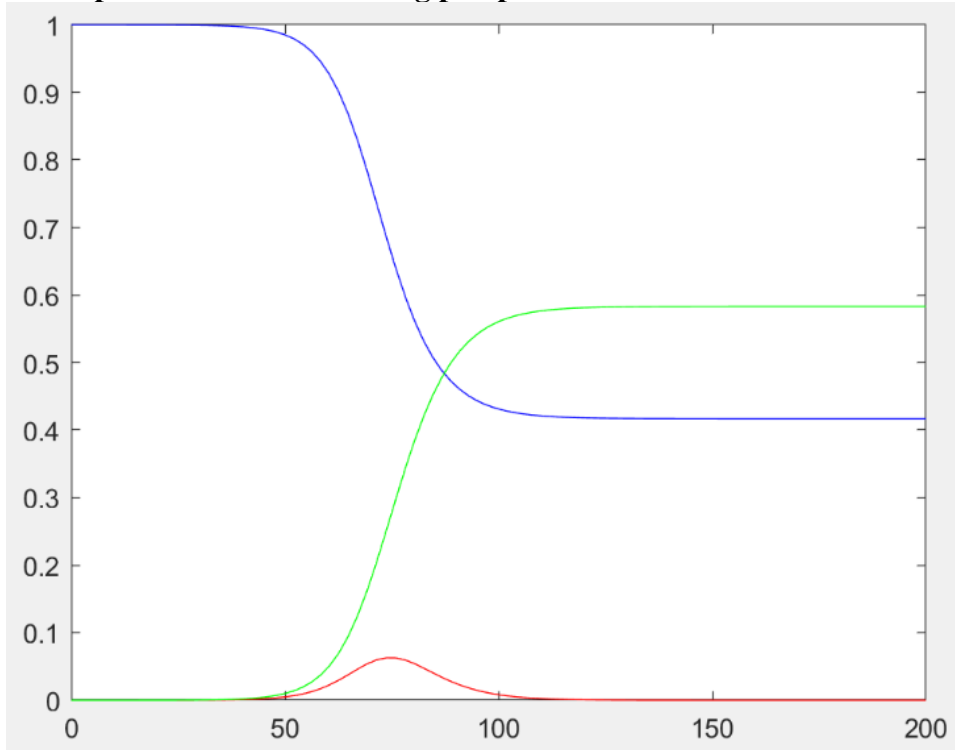


Investigation 2.1

If we guess that each infected would make a possibly infecting contact every two days, then b would be $1/2$. We emphasize, though, that this is just a guess. Plot graphs of $s(t)$, $i(t)$, and $r(t)$ on the same set of axes for $k = 1/3$ and $b = 1/2$. Discuss the key features of your graph, including the maximum number of people infected at a given time, the time required for less than, say, 1% of the population to be infected, and any other features that you think are important from a modeling perspective.

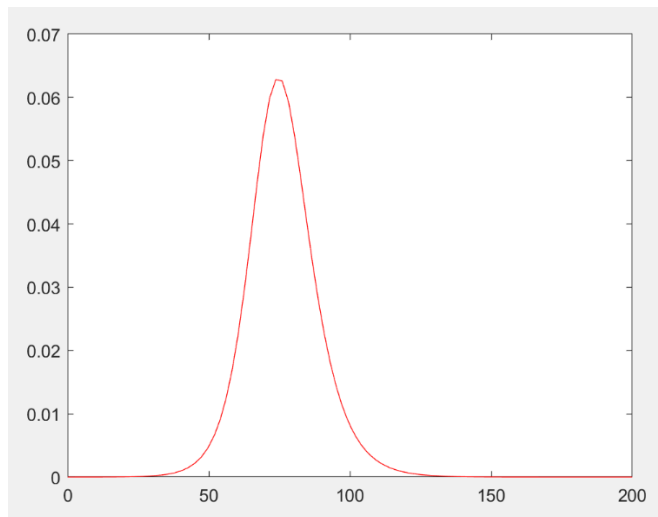


The maximum number of people infected is 0.0628, or 6.28% of the population. This occurs on day 73.7609. From the start of the timeline, the time required for less than 1% of the population to be infected is 98.7051 days. Another key feature of this graph that is important to note is that the sum of the 3 lines is always 1 (these 3 proportions are all supposed to sum to 1). Eventually, around day 110, these 3 lines all stabilize, and there seem to no longer be any changes in the proportions of each of these groups (the disease is gone - there is no one left to infect the remaining susceptible individuals).

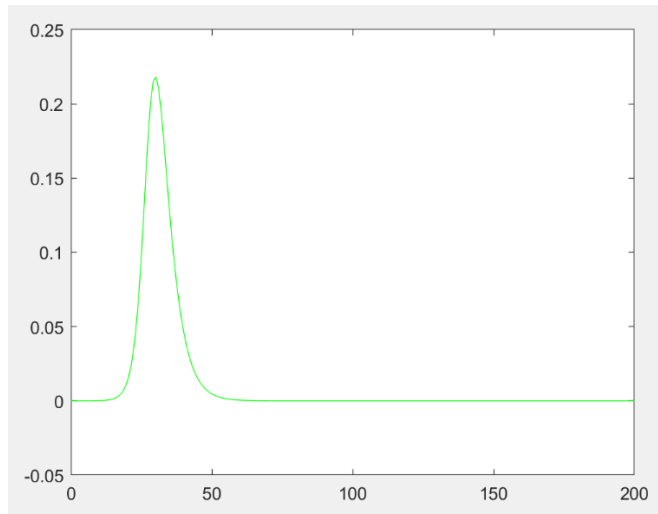
Investigation 2.2

In general, when working on modeling problems, it's important to analyze and understand how our results depend on the numerical values of the parameters involved in the model. One way to do this is to hold all of the parameters fixed except one, and vary that parameter across a range of reasonable values, where “reasonable” is defined (and justified) appropriately in terms of the physical context of your model. In this investigation, we will explore the role that the value of b plays in the behavior of the SIR model. Graph $i(t)$ for $k = 1/3$ and several (at least 4) different values of b between 0.5 and 2. Discuss your results. In your report, include both the individual graphs (i.e. each on its own set of axes) as well as all of your graphs from this problem on a single set of axes, with each value of b clearly identified so that the behavior that results from varying b can be observed.

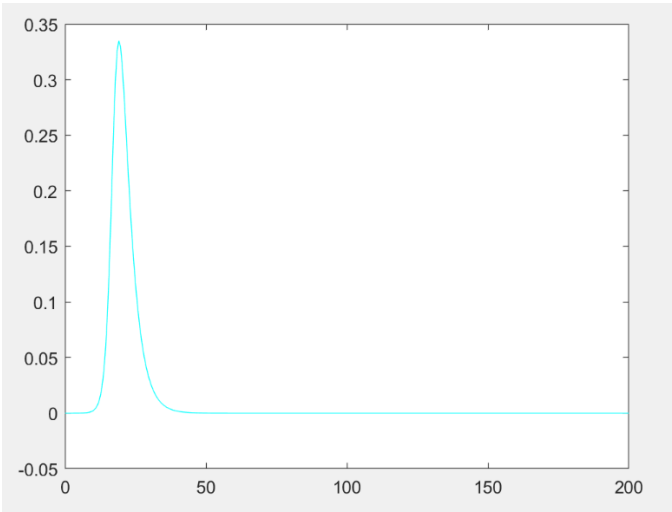
$b=0.5$



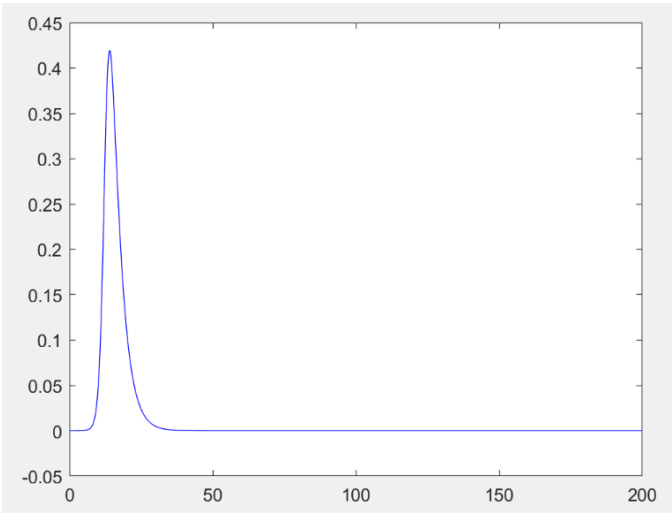
$b=0.8$



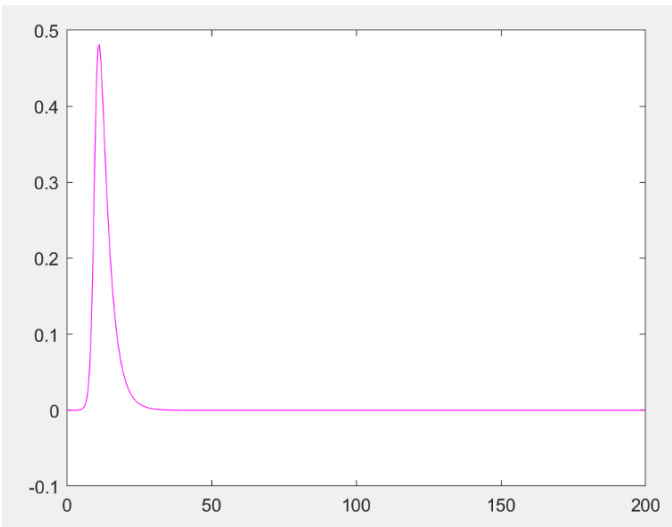
$b=1.1$



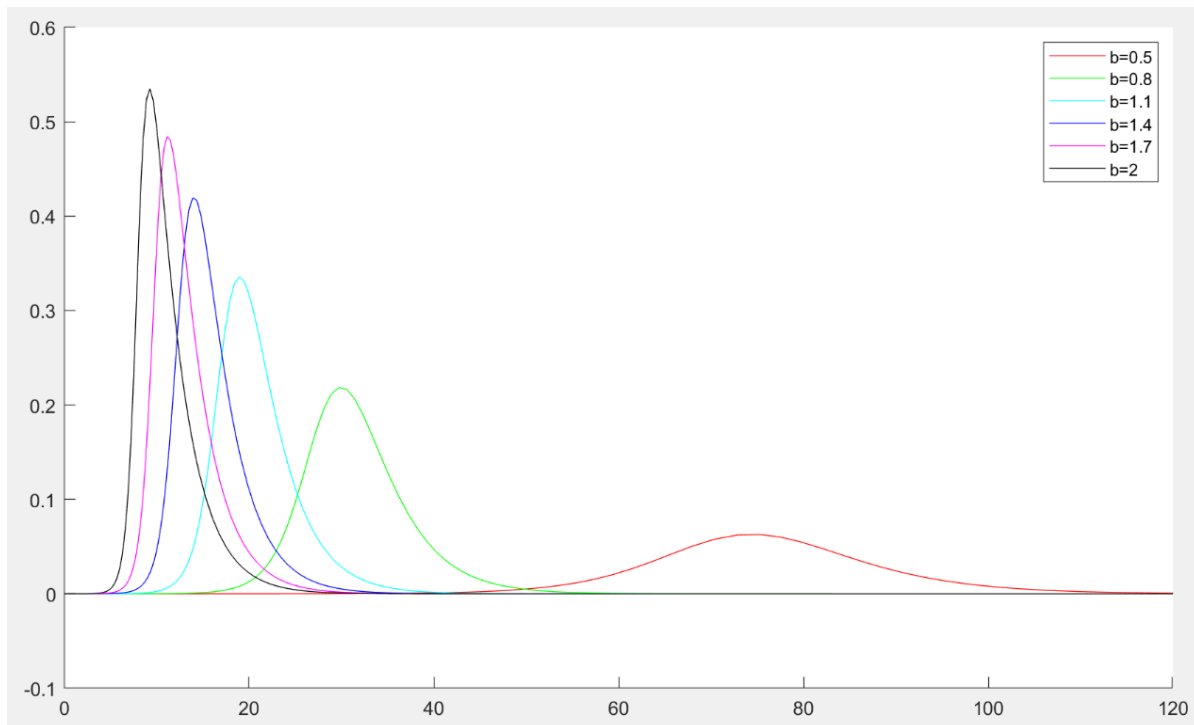
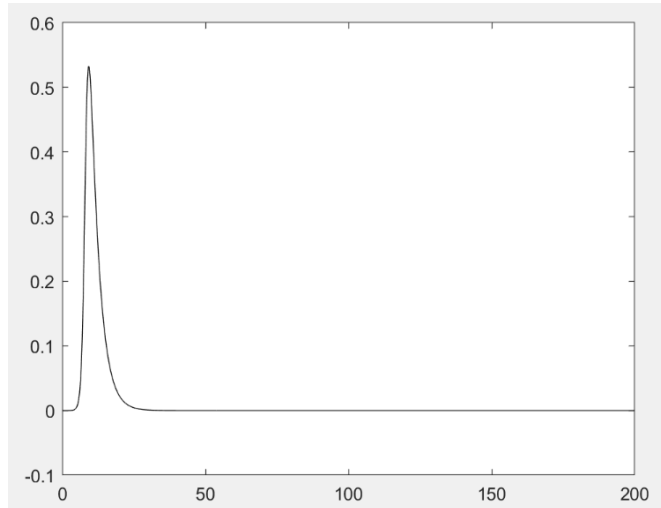
$b=1.4$



$b=1.7$



$b=2.0$

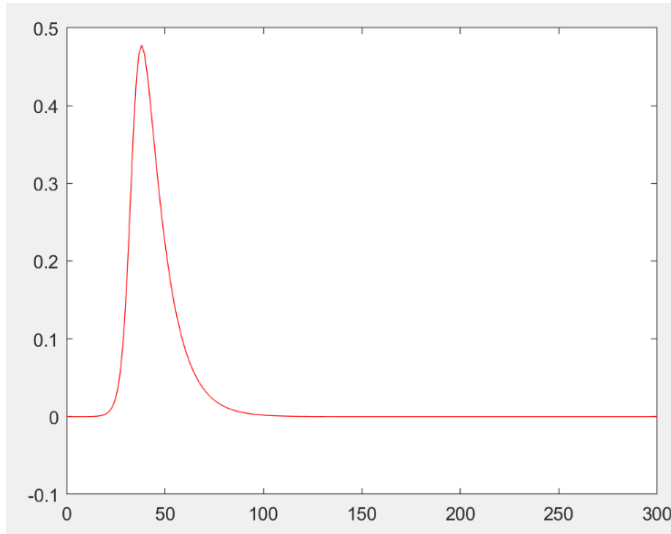


We can see from all of the plots above, especially the combined plot at the end, that as b gets larger, the peak occurs sooner in time, and is also much higher. When $b = 0.5$, the peak occurs around day 75 at less than 10% of the population, whereas when $b = 2$, the peak occurs around day 10 at more than 50% of the population. This makes sense because b represents the number of contacts per day that an infected person is assumed to have with susceptible people. If an infected person has contact with less than 1 susceptible person per day on average, it will take a while for the disease to spread, and there would be enough recovery time between infected people that there would never be a large proportion of the population infected at a time. However, on the other end of our range of b -values, if each infected person comes in contact with 2 susceptible people every day (and $k = \frac{1}{3}$ is held constant, which means each person is infected/infectious for 3 days), then each infected person can infect 6 susceptible people, and the disease will spread much faster and there will be a large number of people infected at a time.

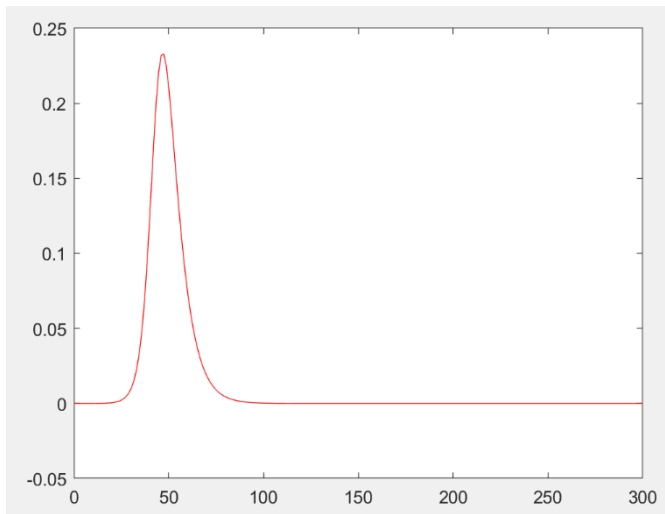
Investigation 2.3

In this investigation, we will explore the role that the value of k plays in the behavior of the SIR model. Graph $i(t)$ for $b = 1/2$ and several (at least 4) different values of k between 0.1 and 0.6. Discuss your results. In your report, include both the individual graphs (i.e. each on its own set of axes) as well as all of your graphs from this problem on a single set of axes, with each value of k clearly identified so that the behavior that results from varying k can be observed.

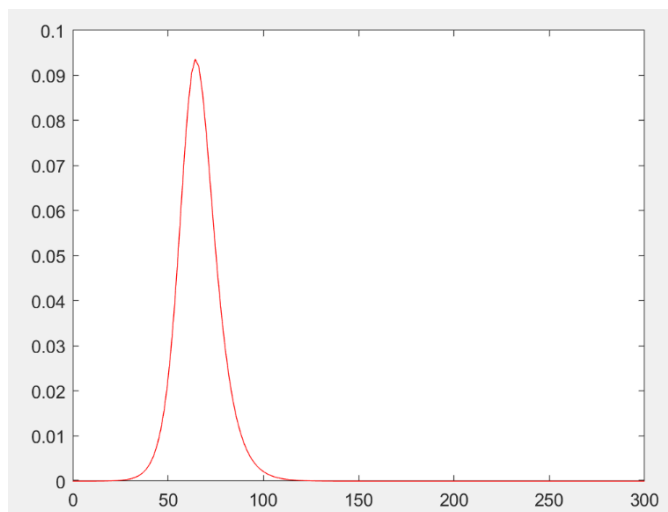
$k=0.1$



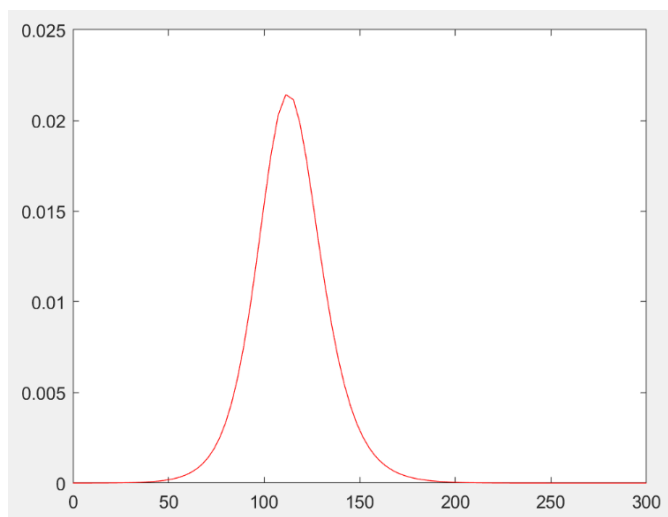
$k=0.2$



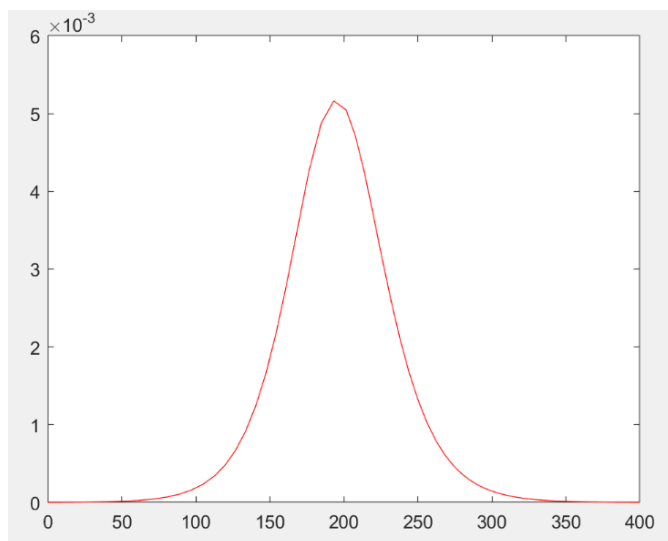
$k=0.3$



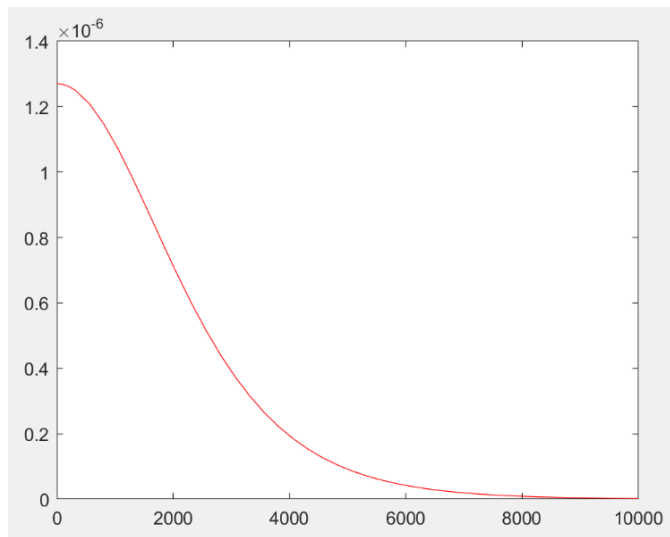
$k=0.4$



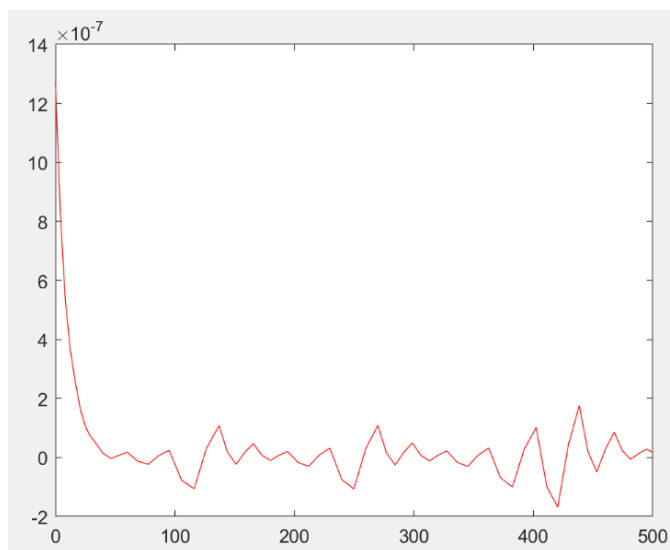
$k=0.45$

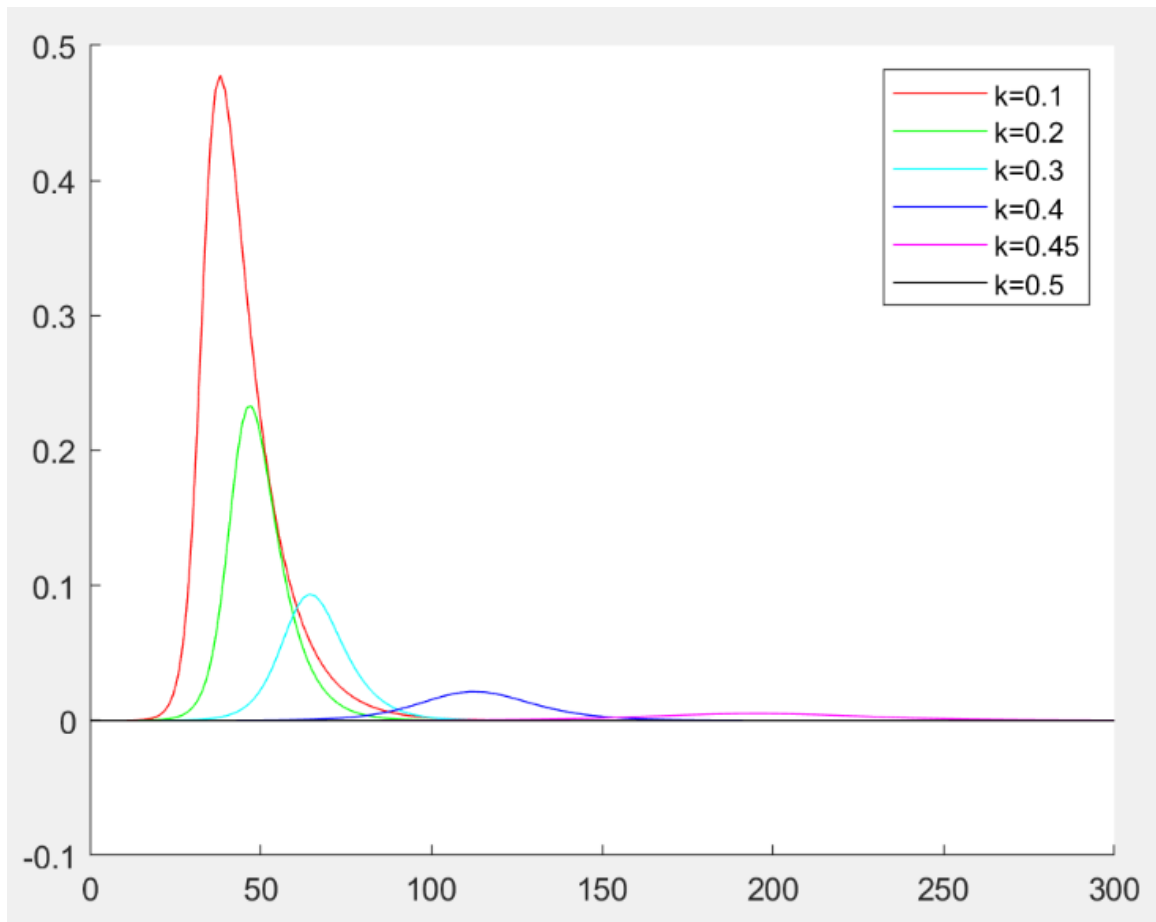


$k=0.5$



$k=0.6$





We recall that k is the rate at which individuals recover from the disease, and we are holding b constant at $\frac{1}{2}$. For example, when $k = 0.5$, it takes people an average of 2 days to recover from the disease (and no longer be infected), so they are most likely only infecting one susceptible person each. As k gets smaller, it takes people a longer amount of time to move from the infected to recovered group. When $k = 0.1$, it takes people an average of 10 days to recover from the disease, whereas with $b = \frac{1}{2}$ being held constant, infected individuals are averaging infecting 5 susceptible people each. Thus, as k gets larger, people are recovering faster, and are spreading the disease to fewer susceptible people, so it takes longer for the disease to spread, and the peak is much lower, so the maximum proportion of the population infected at a single time is much lower.

Investigation 2.4

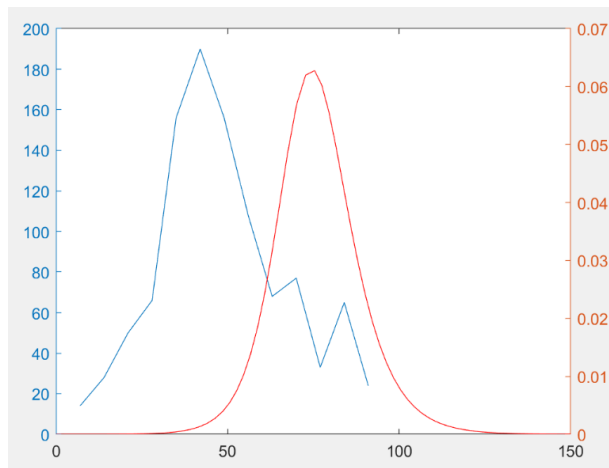
There is a change in the character of the graph of $i(t)$ near one end of the suggested range (0.1 to 0.6) for k (with $b = 1/2$). What is the change, and where does it occur?

When $k \geq 0.5$, we start to see a change in the shape of the graphs. When $k = 0.5$, the number of infected people appears to start out very high and then takes a while to go to 0, but when we look at this graph in perspective with the others, we see that this jump is insignificant. Similarly, when $k=0.6$, there appears to be a very erratic behavior to the graph, but this is really just the number of infected individuals staying around 0. We can recall that k is the rate of transition from the infected to recovered group. At $k = 0.6$, $k > b$, so people are recovering from the disease, on average, faster than they are able to spread the disease to susceptible individuals. When $k = b$, people are recovering from the disease just as fast as they are spreading it, so the number of infected individuals generally does not change – and the number of infected people originally was almost insignificant.

Investigation 2.5

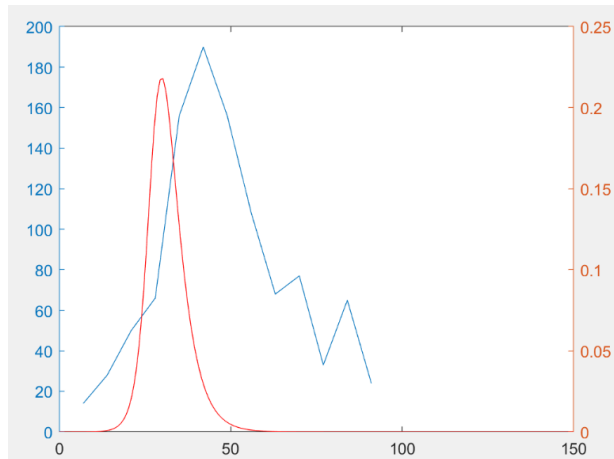
Next, we'll compare our model with the data. Table 2.3 provides the number of deaths each week that could be attributed to the flu epidemic. If we assume that the fraction of deaths among infected individuals is constant, then the number of deaths per week should be roughly proportional to the number of infecteds in some earlier week. Graph the flu death data and $i(t)$ with $k = 1/3$ for several values of b between 0.5 and 2. What value of b produces the most reasonable model (in terms of fitting the shape of the data)? Explain your conclusion.

$b=0.5$



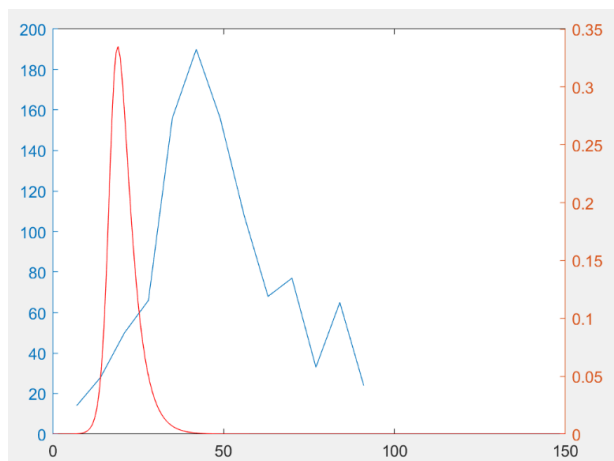
The death peak occurs before the infected peak, so this does not make sense.

$b=0.8$



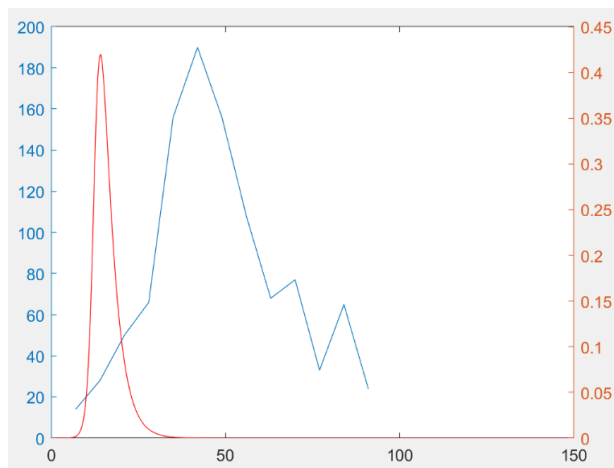
The death peak occurs about a week and a half after the infected peak. This seems like it might be too fast, but we are getting closer.

$b=1.1$



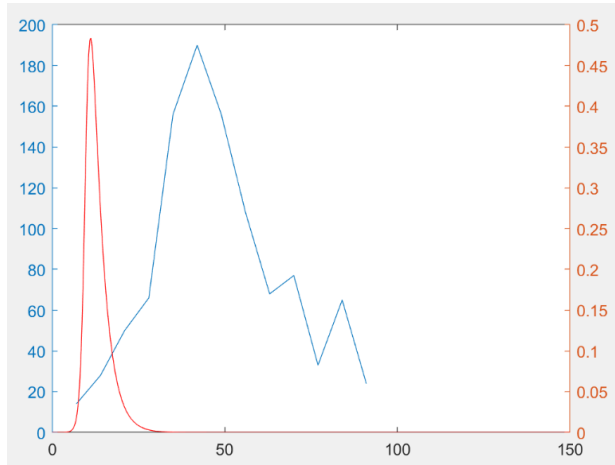
The death peak occurs about three and a half weeks after the infected peak. This might be a little bit too long, but depends on how deadly the disease is.

$b=1.4$



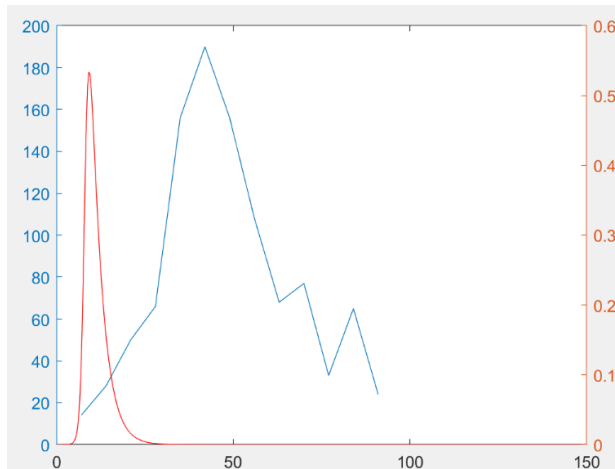
These peaks are about 4 weeks apart.

$b=1.7$



These peaks are about 4.5 weeks apart.

$b=2$



These peaks are about 4.5-5 weeks apart.

Based on the graphs above, we could approximate b to be somewhere between 0.8 and 1.1. The true value of b will be dependent on the average amount of time between infection and death. Anything greater than $b=0.8$ could be plausible, based on how long the disease takes to kill a person. If the average amount of time between infection and death is around 2 weeks, then a b -value between 0.8 and 1.1 would be a good guess, whereas if the average amount of time between infection and death is closer to a month (4-5 weeks) then a b -value around 1.4, 1.7, or 2 would be the best guess. The only value for b that would not be plausible at all is $b=0.5$ (and probably b -value around 0.5) because we can see that in this case, the deaths are occurring before the infections. This does not make sense in the context of this problem, so we can rule this out as a possibility.

Investigation 2.6

Use your numerical solutions from Investigation 2.5 with $k = 1/3$ and your value of b that produced the best fit to the data to compute c both directly as $c = b/k$ and as $c = \ln(s_\infty)/(s_\infty - 1)$. (You'll need to estimate the value of s_∞ from your numerical solutions for $s(t)$ and $i(t)$.) Did you obtain approximately the same value for c computed in these two different ways? Interpret and explain your results.

We will assume from Investigation 2.5 that $b = 1$ (and $k = 1/3$). When we compute c directly, we get $c = b/k = 1 / 1/3 = 3$. When we plot our graph of $s(t)$, $i(t)$, and $r(t)$ for $k = 1/3$ and $b = 1$, we find that $s_\infty = 0.0591744$, so $c = \ln(s_\infty)/(s_\infty - 1) = \ln(0.0591744) / (0.0591744 - 1) = -2.827266263 / -0.9408256 = 3.005090702$. Yes, we obtain approximately the same value for c computed in two different ways. Our error is 0.005090702305 , or 0.17% . This is a very small error, which we can attribute to rounding. We know that c is the contact number, or the number of close contacts per day between an infected person and a susceptible person, times the length of infection. If $b = 1$, then this means that on average, each infected person infects 1 susceptible person every day, and with $k = 1/3$ we are assuming that the length of infection is 3 days, so there are 3 possible days for each infected person to infect a susceptible person. Thus, it does make sense that $c = 3$, because we would expect each infected person to infect 3 susceptible people if they are infecting an average of 1 susceptible person per day and they are infectious for 3 days.

Investigation 2.7

What are the key assumptions of the SIR model? Are these assumptions reasonable and realistic? What modifications do you think should be made to improve the model? Do a literature search on compartmental models in epidemiology and write a short summary of your findings. What other models are used to analyze the spread of infectious diseases?

Assumptions: (1) The SIR model assumes a constant total population, where the population does not change due to births, deaths, or immigration/emigration. Any deaths caused by the disease are accounted for in the "recovered" group. (2) The model does not assume any individuals in the population have immunity through vaccination, natural immunity, or any other method, so becoming infected is the only way to leave the susceptible population. Additionally, recovery or death from the disease is the only way to move from the infected population to the recovered population. The model also assumes that once an individual is recovered, they are not susceptible to contracting the disease again. This is an okay assumption to make when little is known about the dynamic nature of the disease but could introduce more errors if the disease is prone to developing variants.

The assumptions are reasonable and are somewhat realistic when little is known about the change in population size. By not including this dynamic component in the model, it introduces room for error in prediction, but still provides a decent estimate of how the disease will spread in a population. For a disease like influenza, which is seasonal, we would likely want to consider that an individual who has had the flu before is susceptible to contracting the flu again.

The SIR Model is a type of compartmental model in epidemiology. Some models that are used to analyze the spread are SIS endemic, SIR endemic, SIR epidemic, and SEIR. We also may want to consider adding a term in the model which adds individuals from the recovered population back to the susceptible population to account for a lack of permanent immunity. Many of these changes will obviously be subject to the nature of the disease being modeled.