Bennett, Andrew - Gallo, Mikel - Michelangelo, Arianna - Alvarez, Luis - Bawkenye, Tatiana - Mirabent Rubinat, Guillem

Abstract

We want to understand and model disease spread focusing on the COVID epidemic. We will make use of the 'COVID19_state.csv' data set to explore the real world data, answer several questions and then simulate a simple contagion model to be of use for a potential policy maker.

September 22, 2023

1 Introduction

The project will consist of two sections. The first will be our baseline analysis of the data set and the conclusions we may arrive to. The second section of the report will show the simulation of a basic model of contagion. Making use of the data provided, we will start by doing some exploratory analysis about the number of COVID19 cases and the population. We will answer a few proposed questions that could be of use for a policy maker in this scenario. Our principal goal will be to determine if there is a correlation between the number of cases; and both the amounts of tests and the different subgroups we can identify in the population. Finally, the conclusion of the analysis will be to identify, based on the previous results, subgroups within the population that might be more sensitive to the virus and require special treatment attention.

2 Questions:

2.1 How has been the evolution on the number of cases over time?

Here we can observe that, for every State, the number of positive cases of Covid was growing as the time passed. That being said, the variable 'date' could be affected by the number of tests that each State is realizing. We could solve part of that by doing a ratio of positive cases in relation to the number of tests of each State.

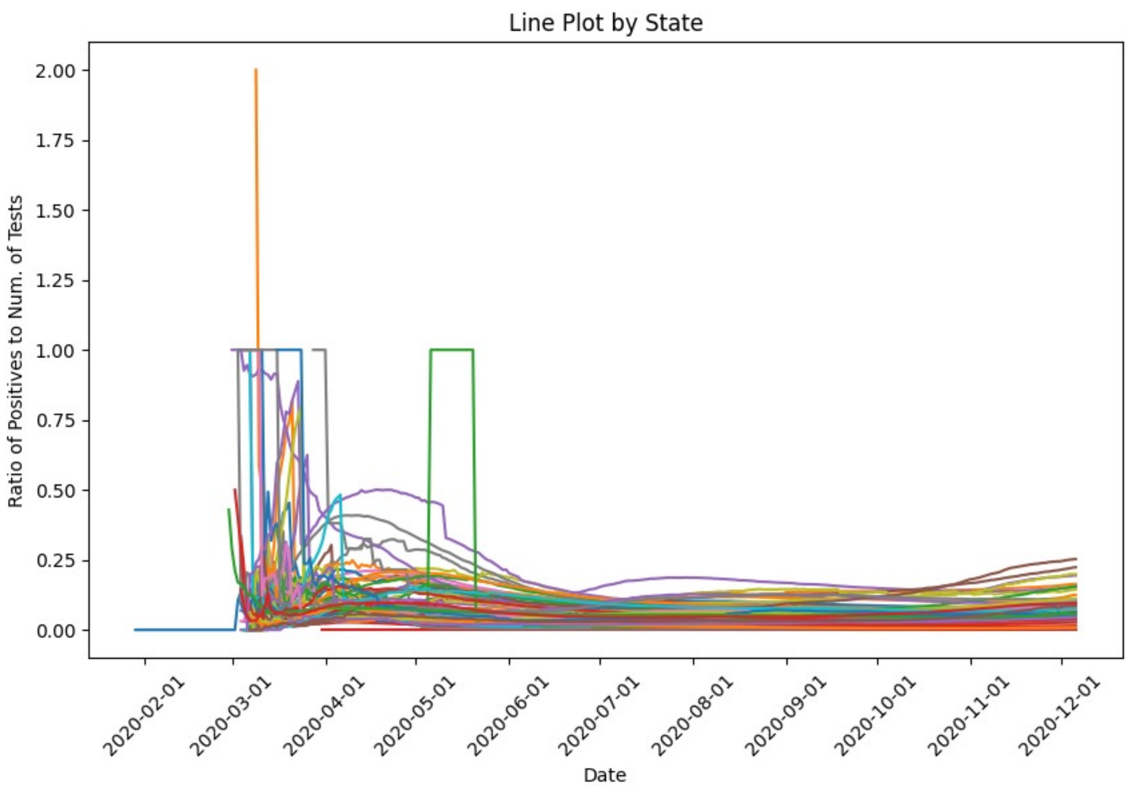


Figure 1: Number of positive cases

Here, on the other hand, we can observe that it doesn't seem as obvious that positive cases increase as time passes. It could as well be that the positive cases had been the same during the period covered (2020) but due to a more intensive testing to the end of the year more cases were discovered. We cannot be as sure anymore that actual Covid cases increased over time during 2020.

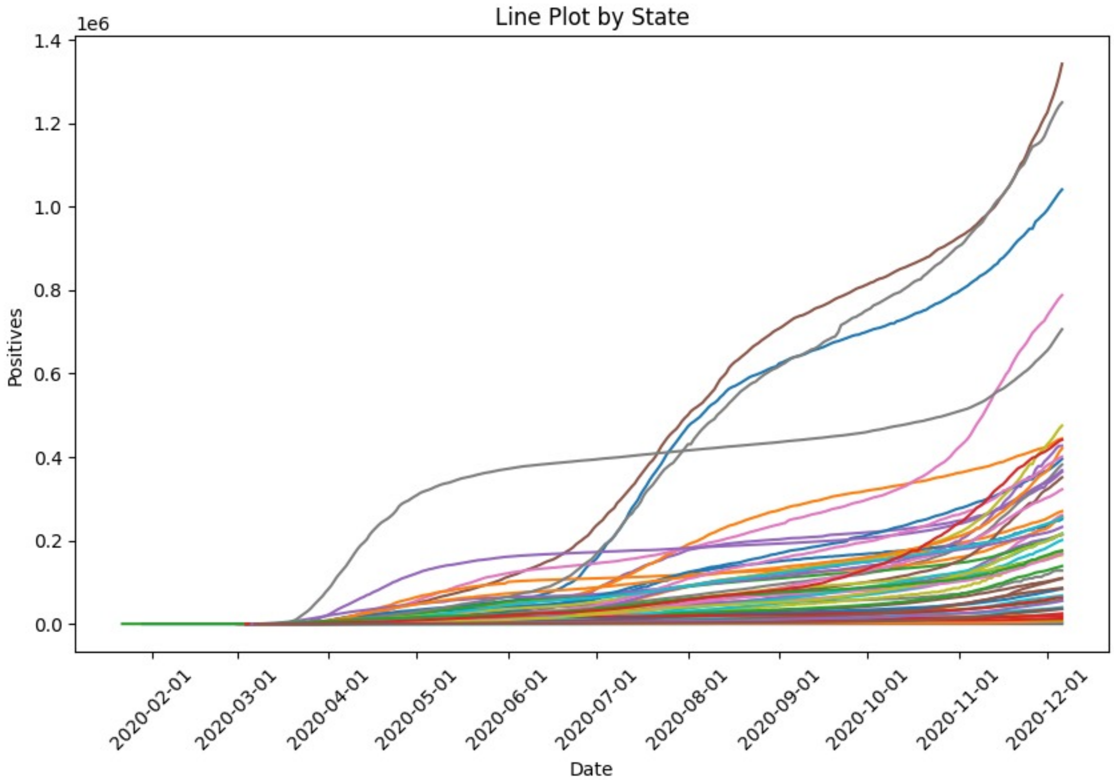


Figure 2: Ratio of positive cases

2.2 Do tests seem to have some effect?

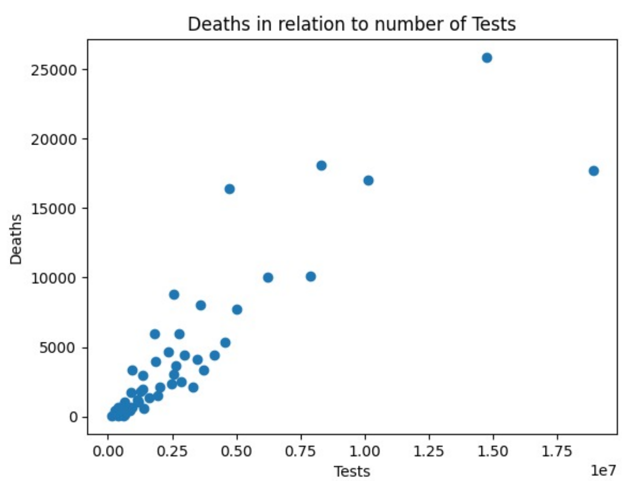


Figure 3: Total Deaths

Here we can observe that the resulting plot seems to show a strong correlation between Tests and Deaths, but we might be missing an important bias. It is entirely possible that States with more Testing have also more Deaths due to both being explained by a higher population. Therefore, in the next steps we are going to project the same plot but with variables being converted to a ratio between the variable itself and the population of the State.

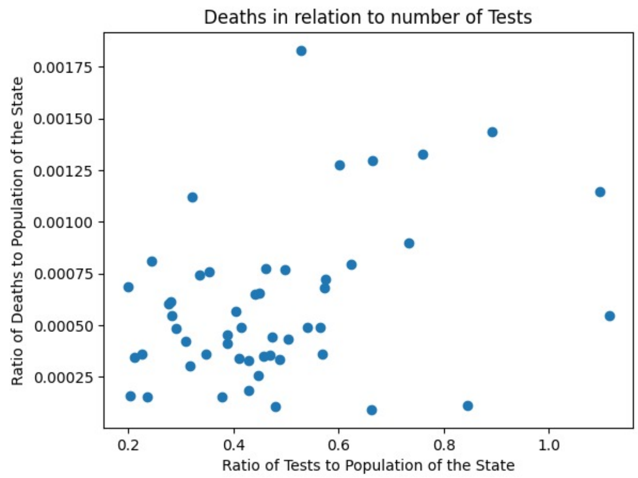


Figure 4: Ratio of Deaths to population

In this second case we can observe that the apparent correlation shown before has mostly disappeared. It no longer seems that more tests lead to a higher death count, following our hypothesis of the total population.

2.3 What are the subgroups that suffered most?

To understand the effects of COVID on different groups, we calculated three variables: the number of infected individuals as a share of the whole state population, the number of deaths as a share of the whole state population, and the death rate among infected individuals. We analyzed the following categories: rural and urban citizens, people of different age and gender.

First of all, we studied the relationship between the level of urbanization and COVID mortality rate in individual states. We plotted, and calculated $Corr = .372$ between the level of urbanization by state and the ratio of $deaths/infected$. A higher level of urbanization is correlated with a higher level of deaths among those infected.

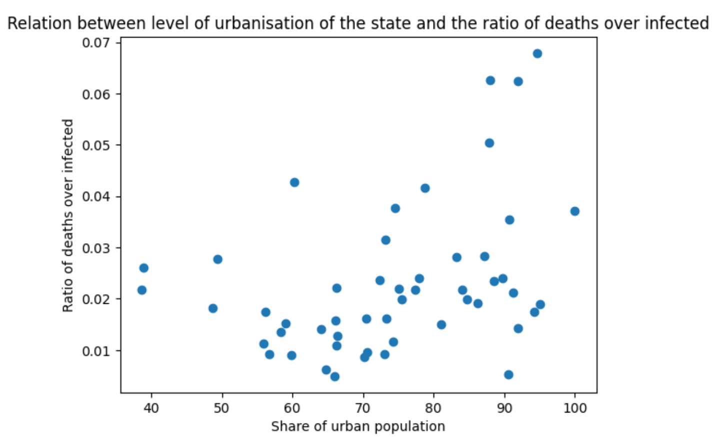


Figure 5: Mortal rate among infected to share of urban population

Nevertheless, the estimations could be biased due to various factors. We expect that the higher the level of urbanization, the greater the chances of contagion due to more crowded places. At the same time, we could think, that higher level of urbanisation would lead to a lower percentage of deaths over the whole population of the state; due to availability of ICU beds, better infrastructure, and other perks cities can offer.

We measured $Corr = .329$ to see, that higher level of urbanisation is related to the higher level of deaths over infected. Thus, we had to reject our hypothesis, that urban states are better equipped.

Besides, the correlation between the number of ICU beds over population and urbanisation level of the state was insignificant.

We also plotted level of urbanisation and number of infected as a share of total population of the state to see its relationship Corr = .328. It proved our hypothesis, that people get easily infected in the more crowded places.

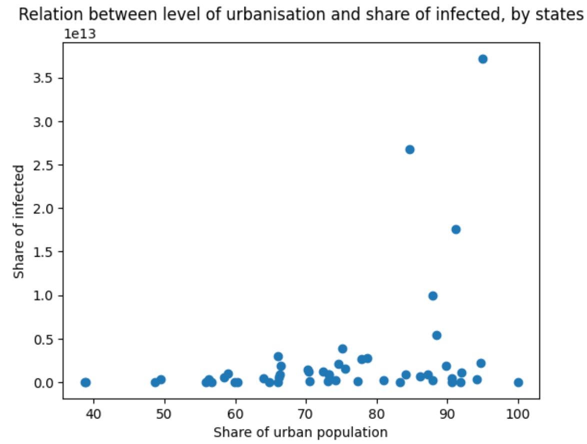


Figure 6: Ratio of infected to share of urban population

We then decided to work with age groups. Let us name each age group:

1. Age: $(0-25)$
2. Age: $(26-54)$
3. Age: $(+55)$

Data showed, that the middle aged group is the biggest in terms of proportion for most states.

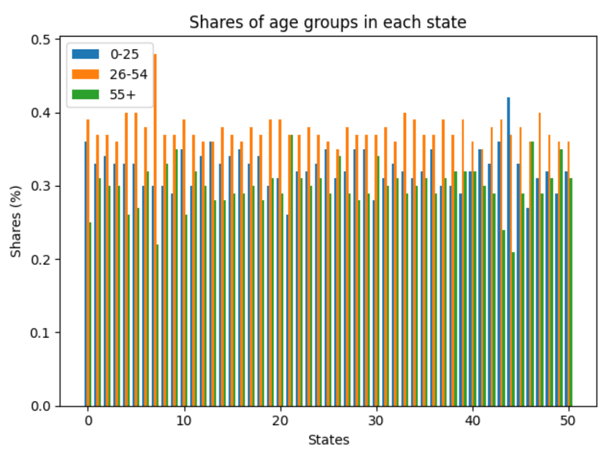


Figure 7: Groups distribution by States

We looked into correlations of age groups shares with the levels of both death rates and infected cases.

1. $Corr(Infected) = .067, \quad Corr(Deaths/Infected) = -.471$
2. $Corr(Infected) = .193, \quad Corr(Deaths/Infected) = .263$
3. $Corr(Infected) = -.226, \quad Corr(Deaths/Infected) = .19$

For the young group, the higher the group share is, the lower the *dead/infected* ratio is in the state. In the older and middle aged groups, the case seems to be the opposite. Still, two last of the correlation coefficients are very low and nothing can be inferred from such analysis.

The main problem with this line of thought lies in the fact, that we don't know the ratio of deaths/infected over population for particular age groups separately. Having this in mind, we know that since the share of middle age group is constantly higher than for the rest, we assume, that it represents most of the effect. In this case, we need to come with some age index for every state. To solve this, we create an index to show the age composition of the state. This index will go from [1 : 3] depending on the weighting by the share of each group in the population of the state. The age index will then reflect if a state is relatively "younger" or "older" for comparison.

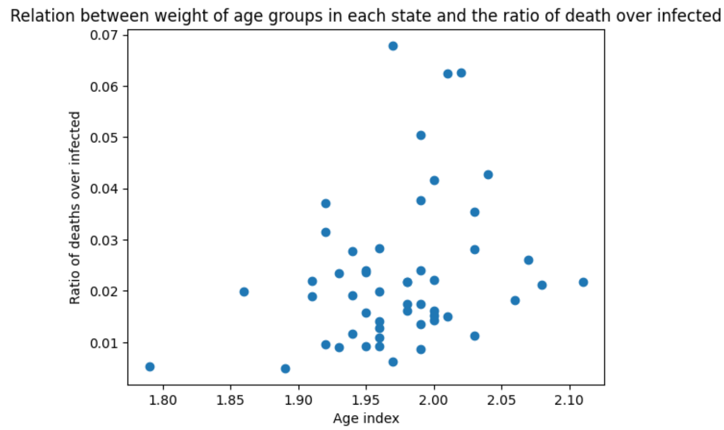


Figure 8: Ratio of deaths from infected to state age index

For "older" states the death rate of infected is higher. The shares of infected/deaths in the whole population for the "older" states are lower. Anyway, the correlations are quite weak, so we can not prove, there are particular relations.

2.4 If we were to do confinement measures by subgroups of populations, on which groups should we focus to stop the contagion early?

According to our analysis, the most vulnerable groups are those living in highly populated urban areas. Furthermore, despite the fact that younger people are more susceptible to infection, middle-aged and older groups are associated with higher death rates among those infected. Therefore, confinement measures should be implemented accordingly.

3 Model:

The model has been designed to showcase the time that will take, for a group of people (10), to spread an infectious disease to the rest of the population (1000) for a determined probability of infection, and an infection rate.

3.1 Assumptions

- Infection rate: $3\frac{People}{Day}$
- Probability of infection: 0.05

We also did some additional assumptions that will come into play in later stages of the model:

- Adjusted propagation (Proportionally decreasing the susceptible population to the increase of the infected base)
- SIR Model: Recovered (6 days after infected, 90%) , dead(10%), Susceptible
- Hospital average capacity index : 2.96 beds per 1000 people: 338 total available beds.

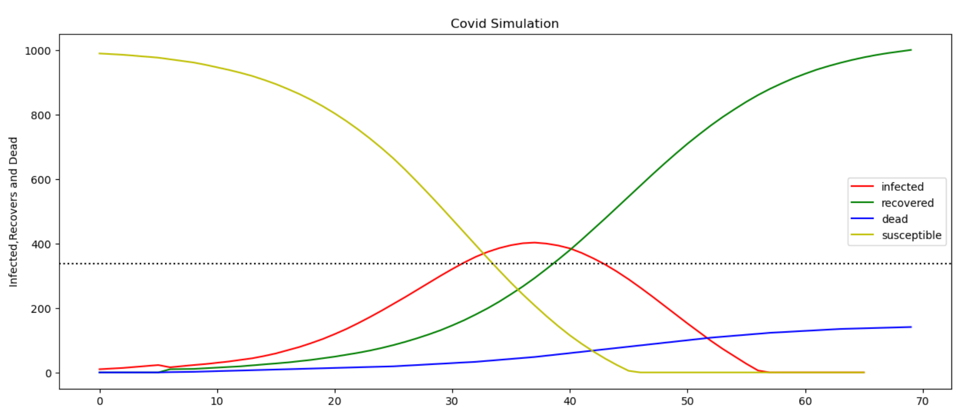
3.2 Methodology Procedure

The SIR model shows the transition of people across multiple conditions/states resulting in a more accurate representation of the effects of a infectious disease in a given population.

1. Balance the increasing number of infections with recovery and death to yield a more realistic evolution of the disease.
2. Estimate the potential volume of dead people.
3. Estimate the time for the infection to end.
4. Explain the stages of a disease

In this dynamic, we introduce a few factors to consider and further develop the analysis. The implementation of a new policy (Vaccination) on day 3, as requested on the instructions. The practical way to view this is to think of a lower infection probability. The objective will be to measure the impact (replacing probability from 0.05 for 0.01) of this new vaccine on the spread of the population. For which, we will make a comparison of the two curves and calculate the delta. Finally we will analyze the slope curves from each model and determine the dates in which the spread was at its most critical state. Moreover, we will make recommendations/suggestions of when should we implement the corrective actions to prevent the infection of the whole population.

3.3 Simulations



As we can observe in the graph above (No vaccine), we can expect to see a 70% increase in the number of infected people by day 3. Given that the speed of the spread is very aggressive, we will consider it to be already a potential epidemic by day 3 (potential considering that less than 1% of the population got infected).

SIR - Model (dotted black line is the avg number of beds in hospitals in spain (capacity = 2.93 beds per 1000 people) This graph shows how people transition between states, which allow us to better model the stages of the disease propagation.

On the other hand, we can see that there's an inflection point on day 2, which could indicate that by that time the disease starts to become more difficult to control.

Taking into account both the recovered and dead population, we've been able to determine 3 stages of infection:

1. **Initiation:** First 10 days we can see that the infection base starts to slowly infect our susceptible population, already signs of a high rate of contagion. This stage could be described as a period of recognition and investigation.

2. **Acceleration:** Day 10 to Day 30 the disease spreads at an exponentially rate, to the point of reaching the maximum capacity of hospitals (338 beds).

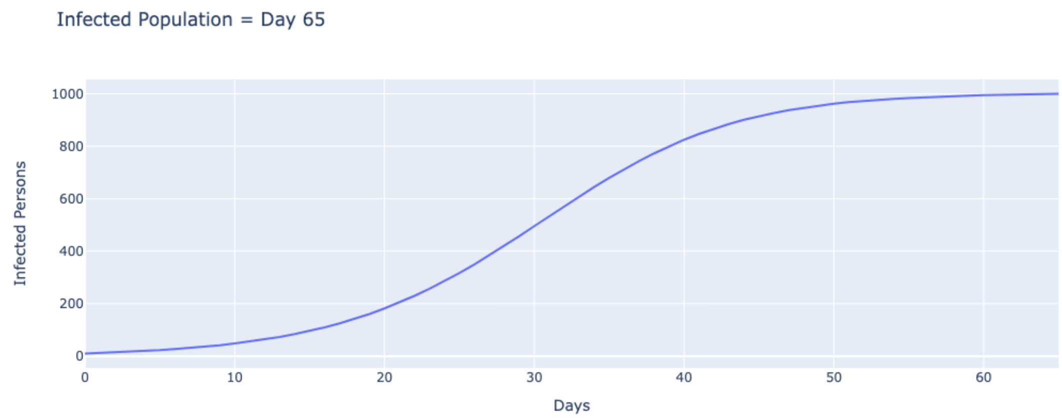
3. **Critical:** With not enough beds available at hospitals and number of infected people at it's peak, between day 30 and day 43 the risk of death increased considerably, yielding at an increasing number of infected during that period.

4. **Deceleration:** After day 42 we see that the infection cases start to decline. As more people start to recover, hospitals regain capacity and dead also start to increase at a lower rate, which concludes in a progressive decline of the disease in the population. Nevertheless, This stage is delicate if there are no preventive policies.

3.4 Questions

3.4.1 How many days will it take for the full population to be infected?

Here we can see that (based on our simple infection model), the disease should fully infect our population (1000) by day 65.



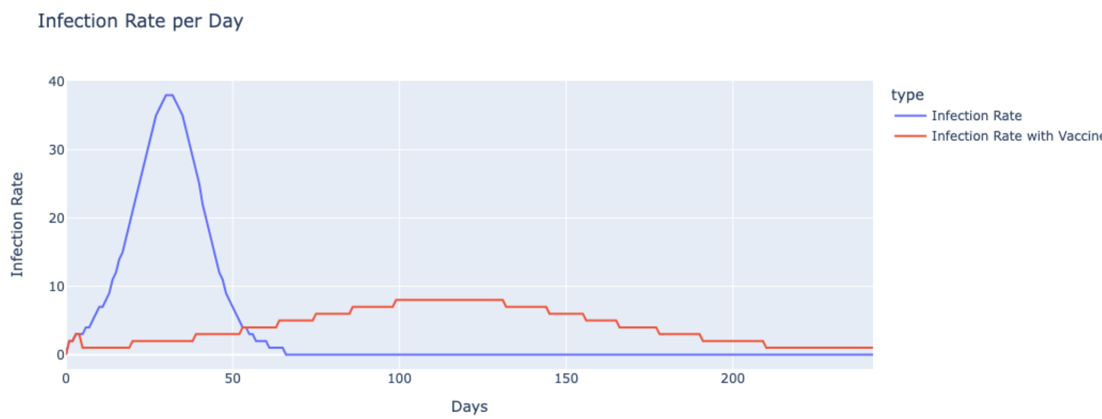
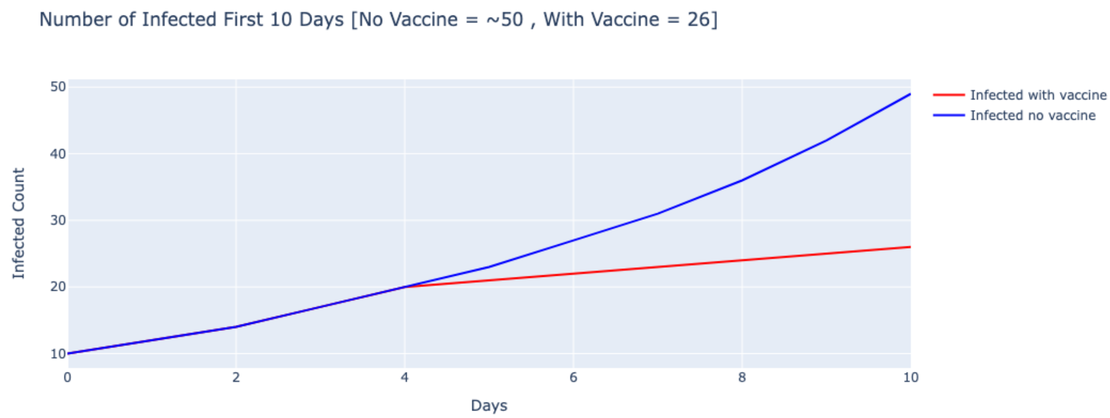


Figure 9: Mortal rate among infected to share of urban population

At an average rate of 4.07 people infected per day, it will take 65 days for a population of 1,000 people to be infected. The infection rate with and without the vaccine can be observed in the Infection Rate per Day graph. The infection rate without the vaccine begins an immediate exponential trend then maximizes at 38 infections per day from day 30 to 32. After day 32 the infections begin to taper as the remaining infection free population dwindles.

3.4.2 How many infections result in the 10th day if we introduce a vaccine at period 3?



At an average rate of 3.2 people infected per day for the first 10 days, there will be 26 people infected by day 10 with a vaccine in effect. Without a vaccine there is 49 people infected on day 10 which is a 47 percent decrease in the number of infections. With a vaccine the effects are evident as early as day 4 where the infection rate tapers from 4 infections in one day to 1 infection in a day. The infection rate grows more slowly and holds the maximal infection rate of 8 infections per day in a 32 day window from day 99 to day 131. The infection rate then begins to taper back to zero.

3.5 Results and Conclusions

We have seen the development of the disease outbreak for our model scenario. We can observe that the introduction of the vaccine, naturally lowers the contagion rate and therefore relieves of stress the medical system as a whole. From our simulations, with the previously mentioned assumptions; the scenario did not become endemic and both deaths and positive cases could be contained. Modifying this base parameters will consequently lead to different results that may be worth studying as well.