# COVID-19 Analysis

*Gabriel Lapointe*

*March 29, 2020*

## Contents



# 1 Introduction

## 1.1 Context Summary

The COVID-19 is infecting many hundred of thousands people in the world and is very virulant. Among these people, many recovered from the infection while some of them died. These statistics increase day after day and many scientists are working to understand the virus and find out a vaccin to stop the propagation. While the COVID-19 is propagating around the world, safety measures have been enforced in many countries in order to reduce the propagation between infected people and non infected people. However, they discovered that elder people and people with chronic diseases are more at risk to die.

Here are some safety measures applied in Canada:

- Social distancing rule where people have to be at a distance of at least 2 meters between each other. People that are not respecting that rule could get a fine (from 1500$ to 6000$). It also means that gatherings of people are strongly forbidden and is punishable by fine.
- Non-essential services and stores are closed. Services like hospitals, police, gaz stations, drug stores, grocery stores are considered essential services and stores.
- There is a limitation of people that can enter the stores considered as essential. Also, people have to wash their hands with purell when entering the store. In grocery stores, baskets are all desinfected once a client finished his grocery and leave the store.

### 1.2 Problem and Questions

The COVID-19 is not fully understood and many thousands of people are getting infected and die day after day. Some safety measures are in place in many countries, but other problems of psychologic nature may arise from these measures. Are those safety measures really as efficient as we thought? We would say yes because it seems to be the common sense for many people to reduce the propagation. However, other factors might be important to consider and might have more impacts on the propagation than we may think.

Since the virus is not fully understood, many questions have to be answered. Here are some of these questions:

1. In which countries the propagation of the virus slowed down the most quickly?
2. Which countries have the greater ratio of deaths over the population and the total infected people?
3. Which countries have the greater ratio of recovery over the population and the total infected people?
4. What is the age category that is more susceptible to die from the COVID-19 after being infected?
5. What is the age category that got mostly infected.
6. Is there a correlation between the sex of a person and the infection rate, death rate and recovery rate?
7. Which chronic diseases are the most vulnerable against the COVID-19?
8. Does the weather have an impact on the COVID-19 propagation?
9. Do the pollution rates have an impact on the COVID-19 propagation?
10. Does the hospitals capacity have an impact on the number of deaths caused by the COVID-19? Which countries are mostly impacted?
11. Is there a correlation between the density of the population and the propagation velocity of the COVID-19?

### 1.3 Objective

The objective of this analysis is to understand the propagation of the COVID-19 in countries and more precisely in Canada and in the province of Québec. It means to identify factors that appear to impact the propagation velocity of the COVID-19. Understanding these factors will help to understand the propagation of the virus and know how to slow it down quicker.

## 2 Data Preparation

The objective is to gather necessary data in order to answer our questions. The following datasets are used in our analysis:

- Total population by country;
- Worldwide Covid-19 cases prepared by the John Hopkins University Center for Systems Science and Engineering.

The dataset shared by the John Hopkins University Center for Systems Science and Engineering provides the information on the:

- Country or region
- Province or state for some countries

- Latitude
- Longitude
- Date
- Cummulative number of people confirmed with the COVID-19
- Cummulative number of people that died from the COVID-19
- Cummulative number of people that recovered from the COVID-19

```
Number of countries or regions:  188
```

```
Start date:  2020-01-22
```

```
End date:  2020-06-25
```

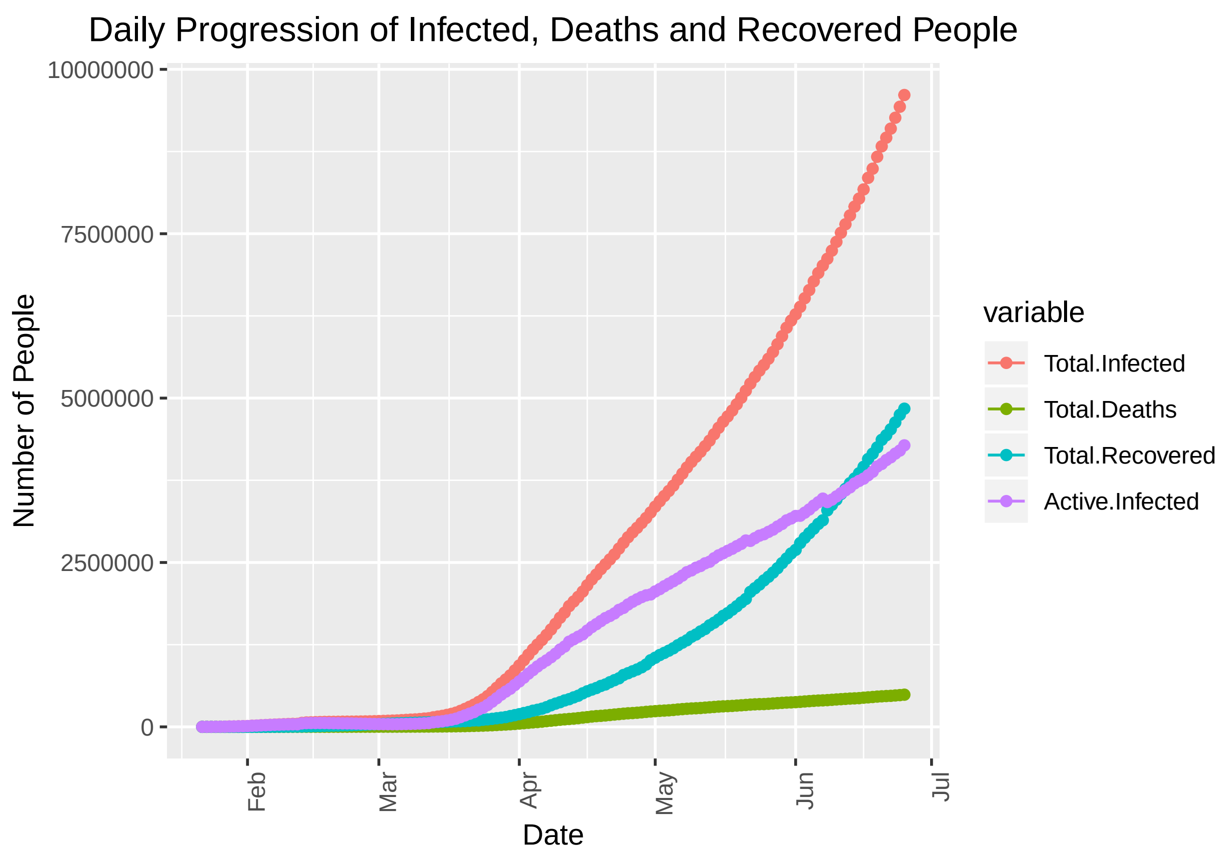
# 3 Dataset Exploration

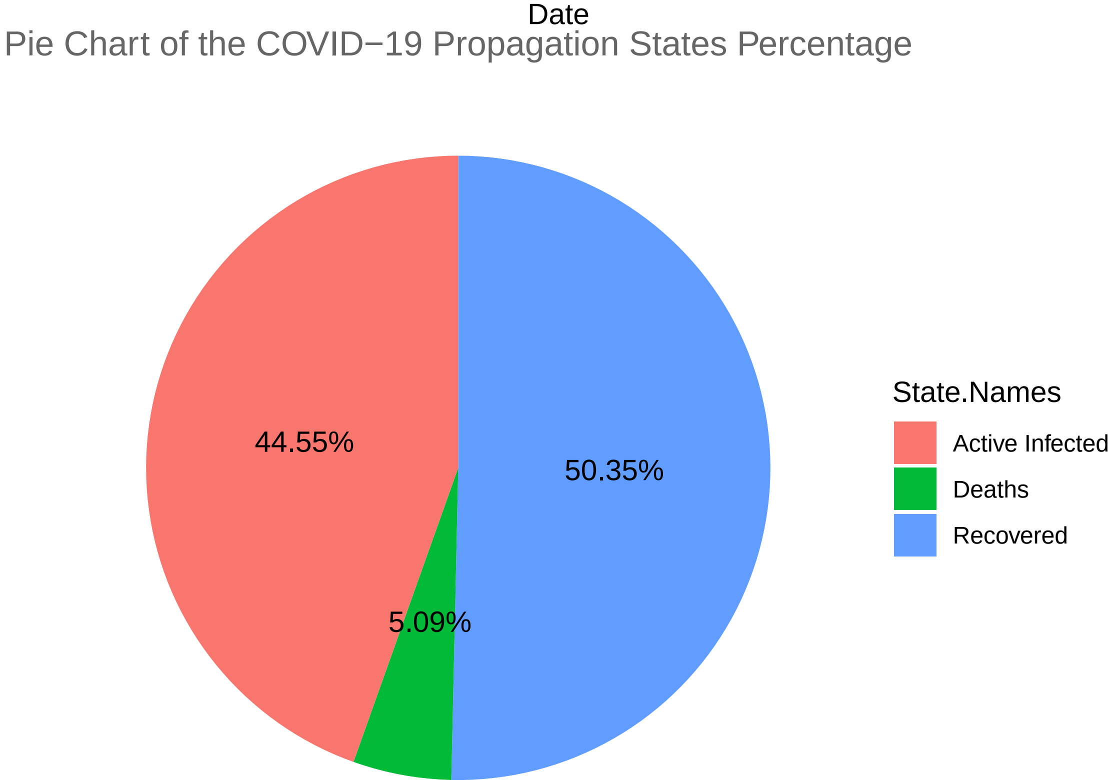
## 3.1 Worldwide Propagation

### 3.1.1 Daily Progression

The objective is to know the distribution of the number of people infected, dead and that recovered from the COVID-19 over days in the world.

| Date | Total.Infected | Total.Deaths | Total.Recovered | Active.Infected |
|---|---:|---:|---:|---:|
| 2020-06-12 | 7644260 | 425780 | 3620412 | 3598068 |
| 2020-06-13 | 7778881 | 430047 | 3706353 | 3642481 |
| 2020-06-14 | 7912426 | 433391 | 3777131 | 3701904 |
| 2020-06-15 | 8034461 | 436899 | 3857338 | 3740224 |
| 2020-06-16 | 8173940 | 443685 | 3955169 | 3775086 |
| 2020-06-17 | 8349950 | 448959 | 4073955 | 3827036 |
| 2020-06-18 | 8488976 | 453979 | 4155099 | 3879898 |
| 2020-06-19 | 8670323 | 460268 | 4250107 | 3959948 |
| 2020-06-20 | 8829186 | 464522 | 4365932 | 3998732 |
| 2020-06-21 | 8960607 | 468583 | 4434628 | 4057396 |
| 2020-06-22 | 9098643 | 472171 | 4526333 | 4100139 |
| 2020-06-23 | 9263935 | 477587 | 4630391 | 4155957 |
| 2020-06-24 | 9431350 | 482758 | 4746118 | 4202474 |
| 2020-06-25 | 9609829 | 489312 | 4838921 | 4281596 |



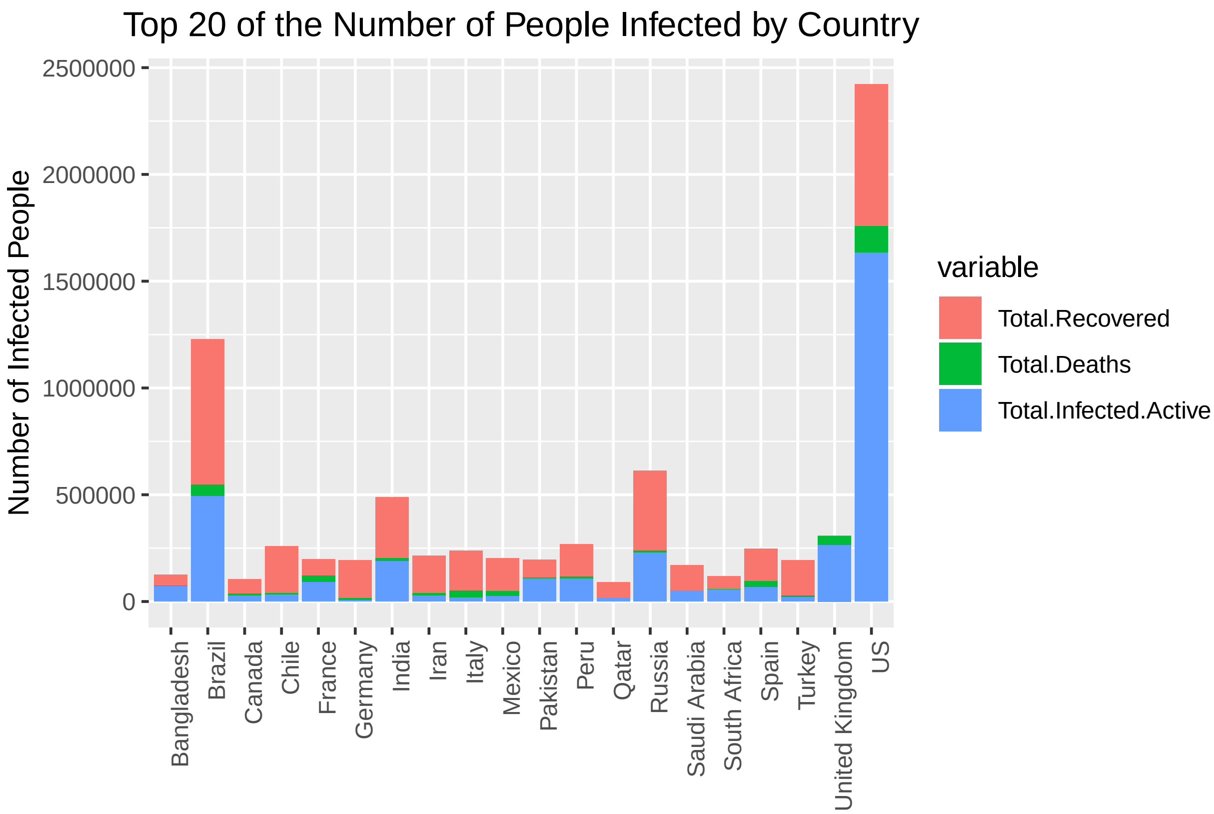




The distribution of the number of people infected, dead or recovered seems to be exponential. This would make sense because if we assume that all people will be infected one day or another, we expect that the curve will describe a sigmoid curve. The reasons behind this is explained in the **Propagation Model section**.

### 3.1.2 Countries With Highest Number of People Infected

The objective is to show in which countries there are the most infected people until today. Since there are many countries and because the list may be huge enough, we only display the 20 countries with the greatest number of infected people.

| Country.Region | Total.Infected | Total.Deaths | Total.Recovered | Total.Infected.Active |
|---|---:|---:|---:|---:|
| US | 2422299 | 124410 | 663562 | 1634327 |
| Brazil | 1228114 | 54971 | 679524 | 493619 |
| Russia | 613148 | 8594 | 374557 | 229997 |
| India | 490401 | 15301 | 285637 | 189463 |
| United Kingdom | 309455 | 43314 | 1361 | 264780 |
| Peru | 268602 | 8761 | 151225 | 108616 |
| Chile | 259064 | 4903 | 219327 | 34834 |
| Spain | 247486 | 28330 | 150376 | 68780 |
| Italy | 239706 | 34678 | 186725 | 18303 |
| Iran | 215096 | 10130 | 175103 | 29863 |
| Mexico | 202951 | 25060 | 152362 | 25529 |
| France | 197885 | 29755 | 75475 | 92655 |
| Pakistan | 195745 | 3962 | 84168 | 107615 |
| Germany | 193371 | 8940 | 176764 | 7667 |
| Turkey | 193115 | 5046 | 165706 | 22363 |
| Saudi Arabia | 170639 | 1428 | 117882 | 51329 |
| Bangladesh | 126606 | 1621 | 51495 | 73490 |
| South Africa | 118375 | 2292 | 59974 | 56109 |
| Canada | 104463 | 8567 | 66869 | 29027 |
| Qatar | 91838 | 106 | 74544 | 17188 |

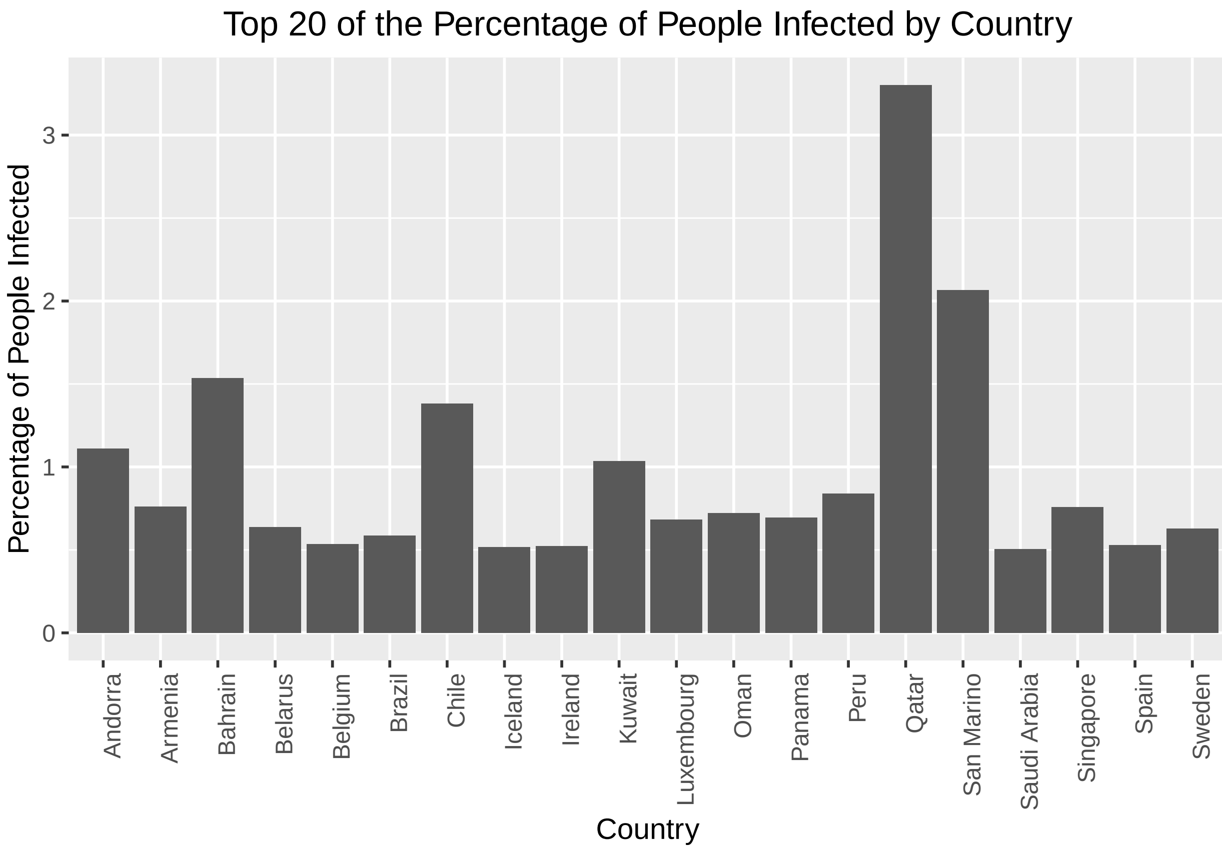



### 3.1.3 Countries Worst Ratio of Infected and Death People

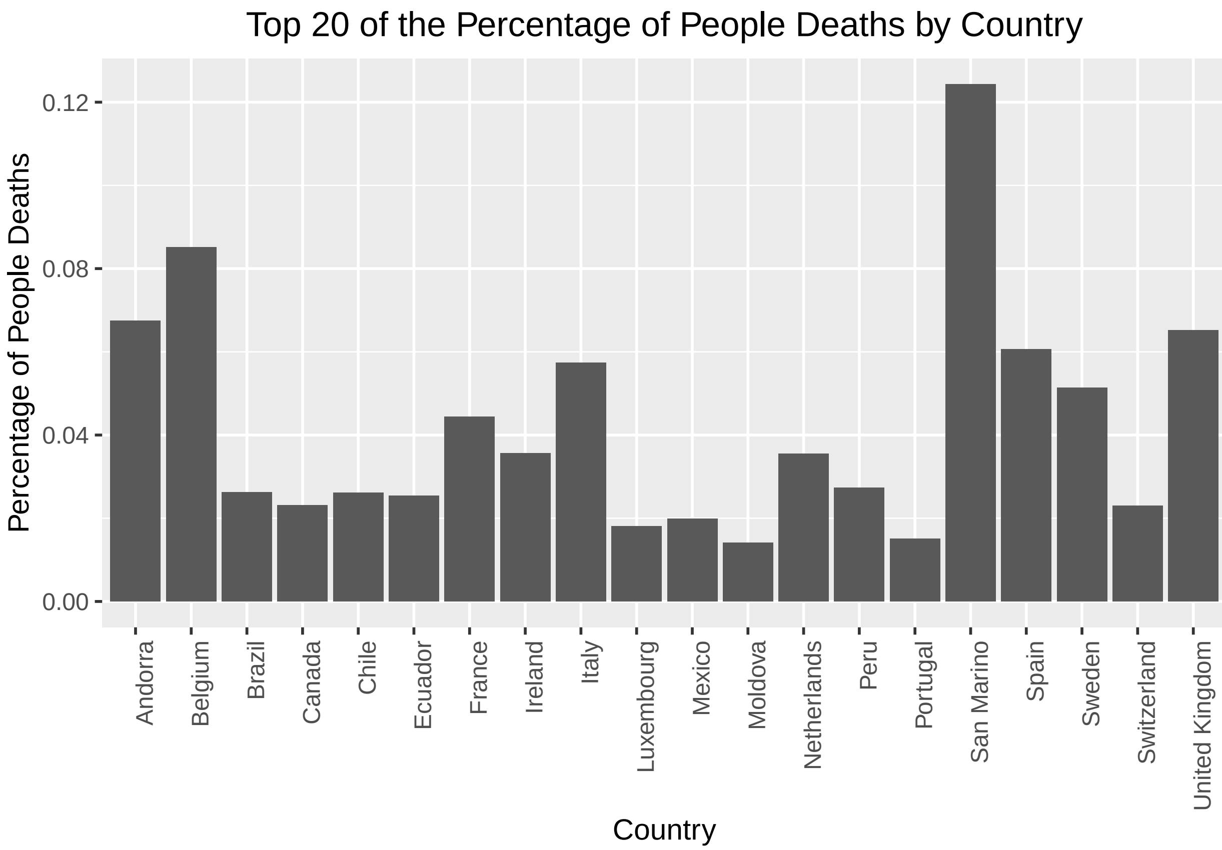
The objective is to know which countries have the worst ratio of deaths and infected people over their population.

| Country.Region | Population | Percent.Infected |
|---|---|---|
| Qatar | 2781677 | 3.3015336 |
| San Marino | 33785 | 2.0660056 |
| Bahrain | 1569439 | 1.5343699 |
| Chile | 18729160 | 1.3832121 |
| Andorra | 77006 | 1.1103031 |
| Kuwait | 4137309 | 1.0341988 |
| Peru | 31989256 | 0.8396632 |
| Armenia | 2951776 | 0.7618464 |
| Singapore | 5638676 | 0.7579084 |
| Oman | 4829483 | 0.7226861 |
| Panama | 4176873 | 0.6951851 |
| Luxembourg | 607728 | 0.6830358 |
| Belarus | 9485386 | 0.6365793 |
| Sweden | 10183175 | 0.6274075 |
| Brazil | 209469333 | 0.5862978 |
| Belgium | 11422068 | 0.5341152 |
| Spain | 46723749 | 0.5296792 |
| Ireland | 4853506 | 0.5234360 |
| Iceland | 353574 | 0.5175720 |
| Saudi Arabia | 33699947 | 0.5063480 |

| Country.Region | Population | Percent.Infected |
|---|---|---|




| Country.Region | Population | Percent.Deaths |
|---|---|---|
| San Marino | 33785 | 0.1243155 |
| Belgium | 11422068 | 0.0851510 |
| Andorra | 77006 | 0.0675272 |
| United Kingdom | 66488991 | 0.0651446 |
| Spain | 46723749 | 0.0606330 |
| Italy | 60431283 | 0.0573842 |
| Sweden | 10183175 | 0.0513592 |
| France | 66987244 | 0.0444189 |
| Ireland | 4853506 | 0.0355825 |
| Netherlands | 17231017 | 0.0355115 |
| Peru | 31989256 | 0.0273873 |
| Brazil | 209469333 | 0.0262430 |
| Chile | 18729160 | 0.0261784 |
| Ecuador | 17084357 | 0.0254209 |
| Canada | 37058856 | 0.0231173 |
| Switzerland | 8516543 | 0.0229905 |
| Mexico | 126190788 | 0.0198588 |
| Luxembourg | 607728 | 0.0181002 |
| Portugal | 10281762 | 0.0150655 |
| Moldova | 3545883 | 0.0141573 |

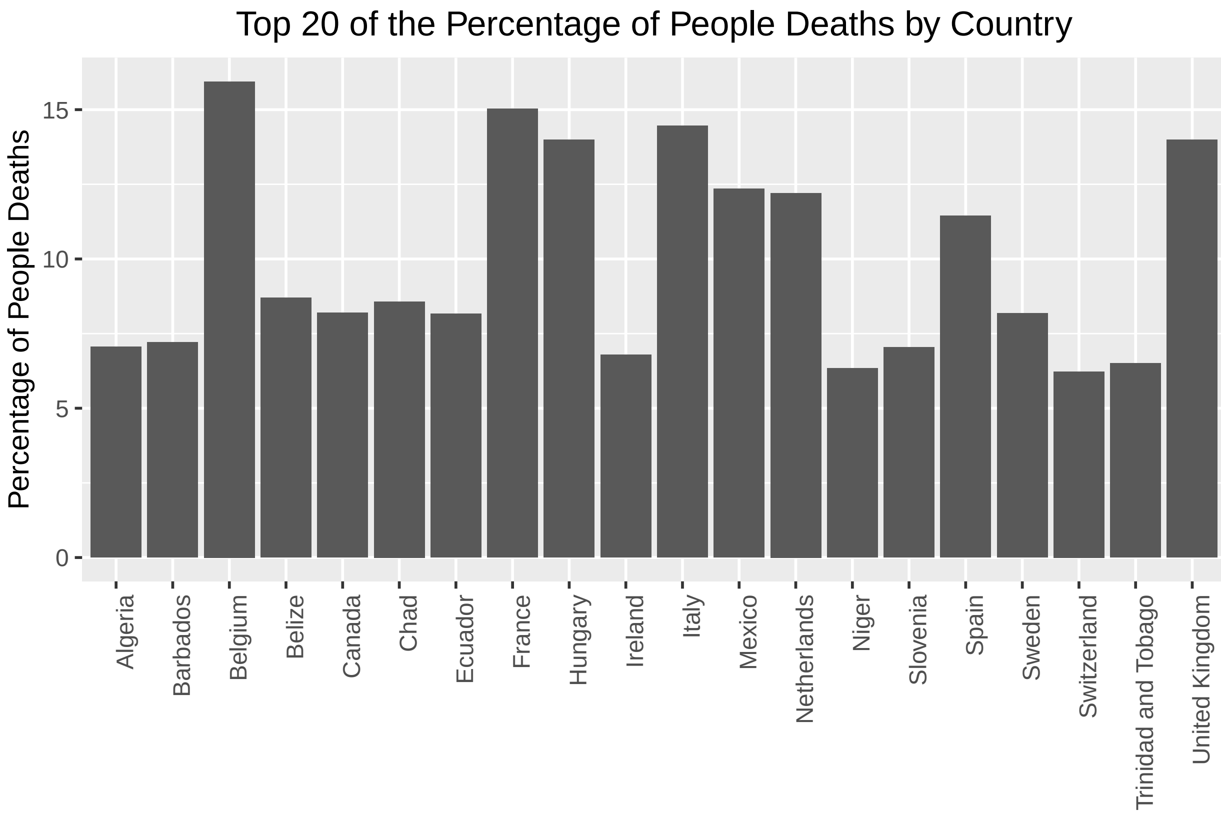



#### 3.1.4 Countries Worst Ratio of Deaths Over Cumulative Infected People

The objective is to know which countries (top 20) have the worst ratio of dead people over the cumulative infected people.

| Country.Region | Total.Infected | Total.Deaths | Percent.Deaths |
|---|---|---|---|
| Belgium | 61007 | 9726 | 15.942433 |
| France | 197885 | 29755 | 15.036511 |
| Italy | 239706 | 34678 | 14.466889 |
| United Kingdom | 309455 | 43314 | 13.996866 |
| Hungary | 4123 | 577 | 13.994664 |
| Mexico | 202951 | 25060 | 12.347808 |
| Netherlands | 50122 | 6119 | 12.208212 |
| Spain | 247486 | 28330 | 11.447112 |
| Belize | 23 | 2 | 8.695652 |
| Chad | 863 | 74 | 8.574739 |
| Canada | 104463 | 8567 | 8.200990 |
| Sweden | 63890 | 5230 | 8.185945 |
| Ecuador | 53156 | 4343 | 8.170291 |
| Barbados | 97 | 7 | 7.216495 |
| Algeria | 12445 | 878 | 7.055042 |
| Slovenia | 1547 | 109 | 7.045895 |
| Ireland | 25405 | 1727 | 6.797874 |
| Trinidad and Tobago | 123 | 8 | 6.504065 |
| Niger | 1056 | 67 | 6.344697 |
| Switzerland | 31428 | 1958 | 6.230113 |

| Country.Region | Total.Infected | Total.Deaths | Percent.Deaths |
|---|---|---|---|

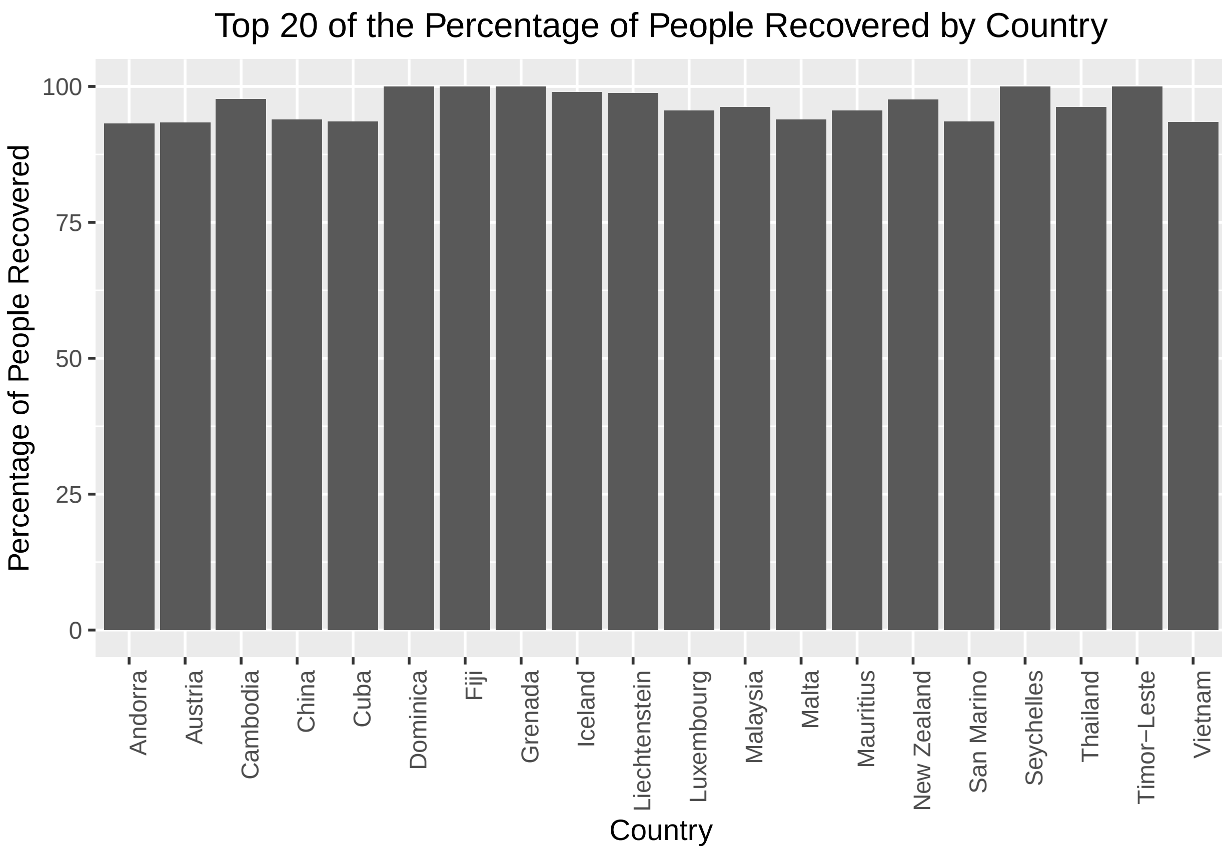



#### 3.1.5 Countries Best Ratio of Recovers Over Cumulative Infected People

The objective is to know which countries (top 20) have the best ratio of recovered people over the cumulative infected people.

| Country.Region | Total.Infected | Total.Recovered | Percent.Recovered |
|---|---|---|---|
| Timor-Leste | 24 | 24 | 100.00000 |
| Grenada | 23 | 23 | 100.00000 |
| Dominica | 18 | 18 | 100.00000 |
| Fiji | 18 | 18 | 100.00000 |
| Seychelles | 11 | 11 | 100.00000 |
| Iceland | 1830 | 1811 | 98.96175 |
| Liechtenstein | 82 | 81 | 98.78049 |
| Cambodia | 130 | 127 | 97.69231 |
| New Zealand | 1520 | 1484 | 97.63158 |
| Thailand | 3158 | 3038 | 96.20013 |
| Malaysia | 8600 | 8271 | 96.17442 |
| Mauritius | 341 | 326 | 95.60117 |
| Luxembourg | 4151 | 3968 | 95.59142 |
| China | 84701 | 79572 | 93.94458 |
| Malta | 668 | 627 | 93.86228 |
| San Marino | 698 | 653 | 93.55301 |

| Country.Region | Total.Infected | Total.Recovered | Percent.Recovered |
|---|---|---|---|
| Cuba | 2321 | 2171 | 93.53727 |
| Vietnam | 352 | 329 | 93.46591 |
| Austria | 17477 | 16320 | 93.37987 |
| Andorra | 855 | 797 | 93.21637 |




## 3.2 Countries With Stable or Decreasing Propagation

The objective is to identify all countries whose propagation is stable or decreasing. In order to find these countries, we have to define what precisely is the meaning of *stable* or *decreasing* propagation.

Let $I(t)$ be the cumulative number of infected people at day $t \in \mathbb{N}$. If we take the difference between $I(t)$ at day $t$ and $t+1$, we get the **propagation velocity**. In mathematical terms, it is expressed as

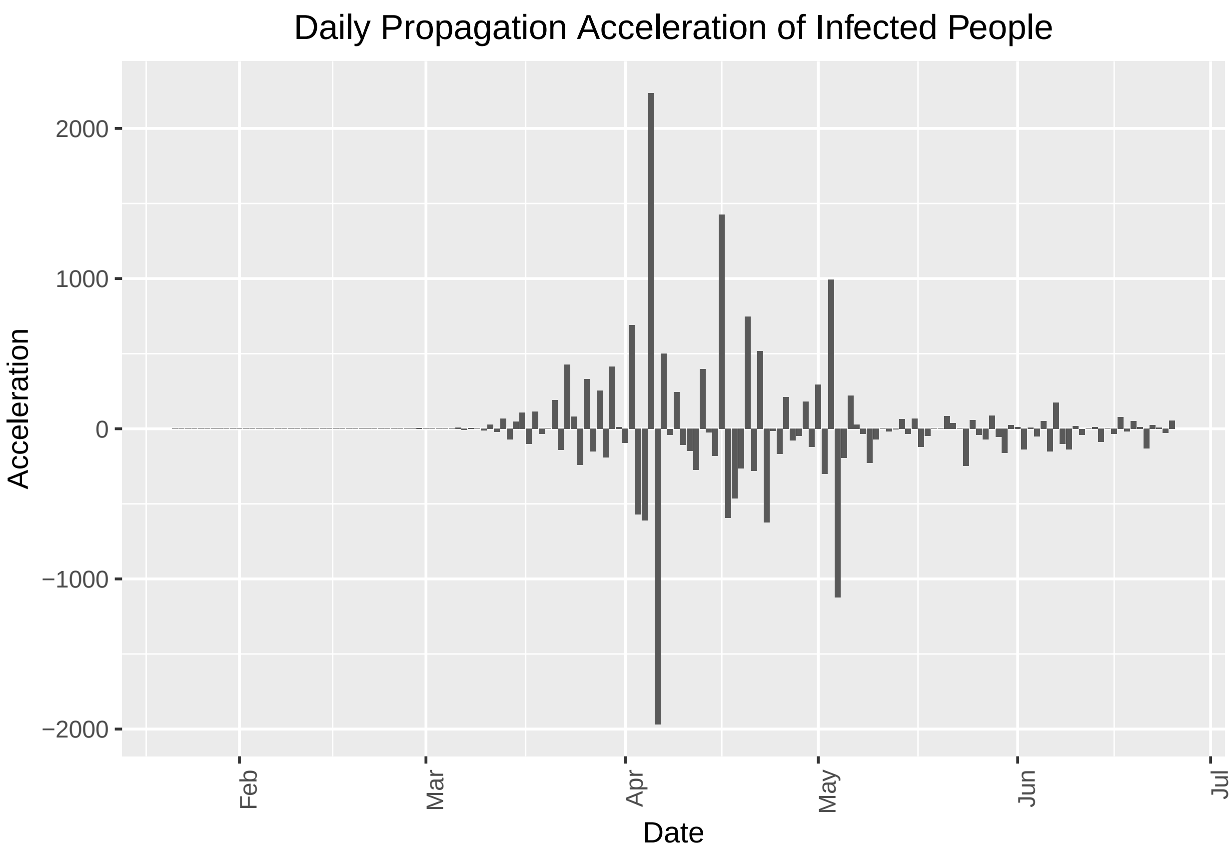
$$\Delta I(t) = \frac{I(t+1) - I(t)}{(t+1) - t} = I(t+1) - I(t)$$

which can be seen as the derivative of $I(t)$ but in a discrete case. The same method is used to define the **propagation acceleration** where we look at the difference between the propagation velocity at day $t$ and $t+1$ expressed as

$$\Delta^2 I(t) = \frac{\Delta I(t+1) - \Delta I(t)}{(t+1) - t} = \Delta I(t+1) - \Delta I(t) = I(t+2) - 2I(t+1) + I(t).$$

The propagation is said **stable** if and only if the acceleration or deceleration is negligible which means that $\Delta^2 I(t) \approx 0$. Therefore, we need to find countries whose propagation acceleration is $\Delta^2 I(t) \lesssim 0$. However, some of the values of $I(t)$ contained in the dataset may be aberrant values. The first step is to determine and remove those aberrant values. The second step is to determine when the propagation is accelerating, stabilized and decelering.

| Date | Total.Infected | infected.delta | infected.acceleration |
|---|---|---|---|
| 2020-06-12 | 99595 | 436 | -3 |
| 2020-06-13 | 100043 | 448 | 12 |
| 2020-06-14 | 100404 | 361 | -87 |
| 2020-06-15 | 100763 | 359 | -2 |
| 2020-06-16 | 101087 | 324 | -35 |
| 2020-06-17 | 101491 | 404 | 80 |
| 2020-06-18 | 101877 | 386 | -18 |
| 2020-06-19 | 102314 | 437 | 51 |
| 2020-06-20 | 102762 | 448 | 11 |
| 2020-06-21 | 103078 | 316 | -132 |
| 2020-06-22 | 103418 | 340 | 24 |
| 2020-06-23 | 103767 | 349 | 9 |
| 2020-06-24 | 104087 | 320 | -29 |
| 2020-06-25 | 104463 | 376 | 56 |




It comes now the following questions:

1. What are the conditions to determine that a propagation velocity $\Delta I(t)$ is categorized as an aberrant value?
2. What is the range of propagation accelerations considerated as approximative to 0 in the expression $\Delta^2 I(t) \approx 0$ and how to determine it?

### 3.2.1 Aberrant Propagation Acceleration Detection Model

The objective is to define what is an aberrant value based on our context and remove them in order to get a better estimation on the propagation acceleration. An aberrant value is a value that seems to be out of the "normality" according to our context. For example, if $\Delta^2(t)$ oscillates normally between $-250$ and 300 and at a day $k$, $\Delta^2(k) = 1500$, then $\Delta^2(k)$ could be considered as an aberrant value.

Since we want to find when the propagation speed is stable or is decreasing over days, the *value* here is the propagation acceleration. Assuming that the propagation accelerations are independent and identically distributed (i.i.d.) between days, the idea is to assume that the propagation accelaration is normally distributed (equivalently $\Delta^2 I(t) \sim N(\mu, \sigma^2)$). The mean $\mu$ is expressed as

$$\mu = \frac{1}{n-2} \sum_{t=1}^{n-2} \Delta^2 I(t)$$

where $n$ is the number of observations in the dataset. Because there are $n$ observations in the dataset, it follows that there are $n-2$ propagation accelerations. The variance is expressed as

$$\sigma^2 = \frac{1}{n-2} \sum_{t=1}^{n-2} (\Delta^2 I(t) - \mu)^2.$$

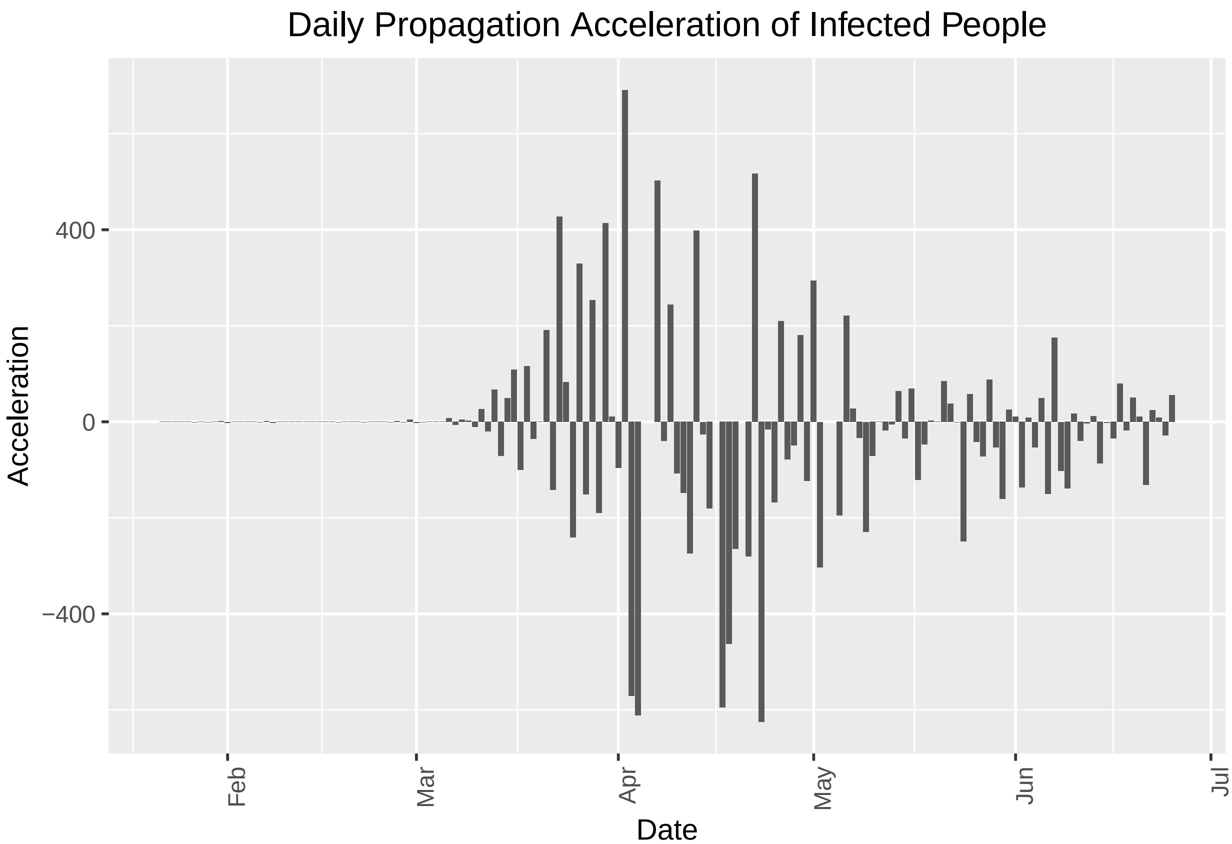
Let's take the data of the Canada as an example. The mean of the propagation accelerations is $\mu = 2.4102564$ and the variance is $\sigma^2 = 118836.1144748$. Therefore, we have $\Delta^2 I(t) \sim N(2.4102564, 118836.1144748)$ where the positive standard deviation is $\sigma = 344.7261442$.

Therefore, the propagation acceleration is **aberrant** if and only if

$$\Delta^2 I(t) \in ] - \infty, \mu - k\sigma] \cup [\mu + k\sigma, \infty[.$$

where $k \in \mathbb{R}$ is the parameter to provide. For example with $k = 2$, non-aberrant propagation accelerations in the Canada have to be exclusively between 2 standard deviations -687.0420319 and 691.8625448.

Actually, there are 3.8461538 % of the propagation accelerations that are aberrant.




### 3.2.2 Propagation Phases Model

We know that a propagation can accelerate, become stable and decelerate on its curve. The objective is to find a model that determines if the propagation is still accelerating, stabilized or decelering based on a set of points chronologically ordered. The idea is to find the global maximum among the non-aberrant propagation accelerations and define what is considered as an acceleration, a stabilization and a deceleration on the curve.

We define a **propagation cycle** when the 3 propagation phases occur in this order: Acceleration, Stabilization and Deceleration. We saw that a propagation is stable if $\Delta^2(t) \approx 0$. This is equivalent to say that a propagation is **stable** if and only if

$$|\Delta^2(t)| < \epsilon$$

where $\epsilon \in \mathbb{N}_*$. The stabilization phase may take many days and for this reason, we have to extend this definition to $|\Delta^2(t+k)| < \epsilon$ where $t+k \leq n$. However, this is not enough because it may happen that, between days $t$ and $t+k$, the propagation is not stable. But overall, the bar chart will show that the propagation is stable. Therefore, we have to consider that a propagation is stable if the average of accelerations within a range of $k$ days is near 0. In other terms, the **propagation is stable** between days $t$ and $t+k$ if and only if

$$|\mu(\Delta^2(t), k)| = \frac{1}{k} \left| \sum_{i=1}^{k} \Delta^2(t+i) \right| < \epsilon.$$

We define the **propagation acceleration** as an acceleration of the number of infected people between day $t$ and day $t+k$. The same method as the stabilization is used:

$$\mu(\Delta^2(t), k) = \frac{1}{k} \sum_{i=1}^{k} \Delta^2(t+i) \geq \epsilon.$$

We define the **propagation deceleration** as a deceleration of the number of infected people between day $t$ and day $t+k$. The same method as the acceleration is used:

$$\mu(\Delta^2(t), k) = \frac{1}{k}\sum_{i=1}^{k} \Delta^2(t+i) \leq -\epsilon.$$

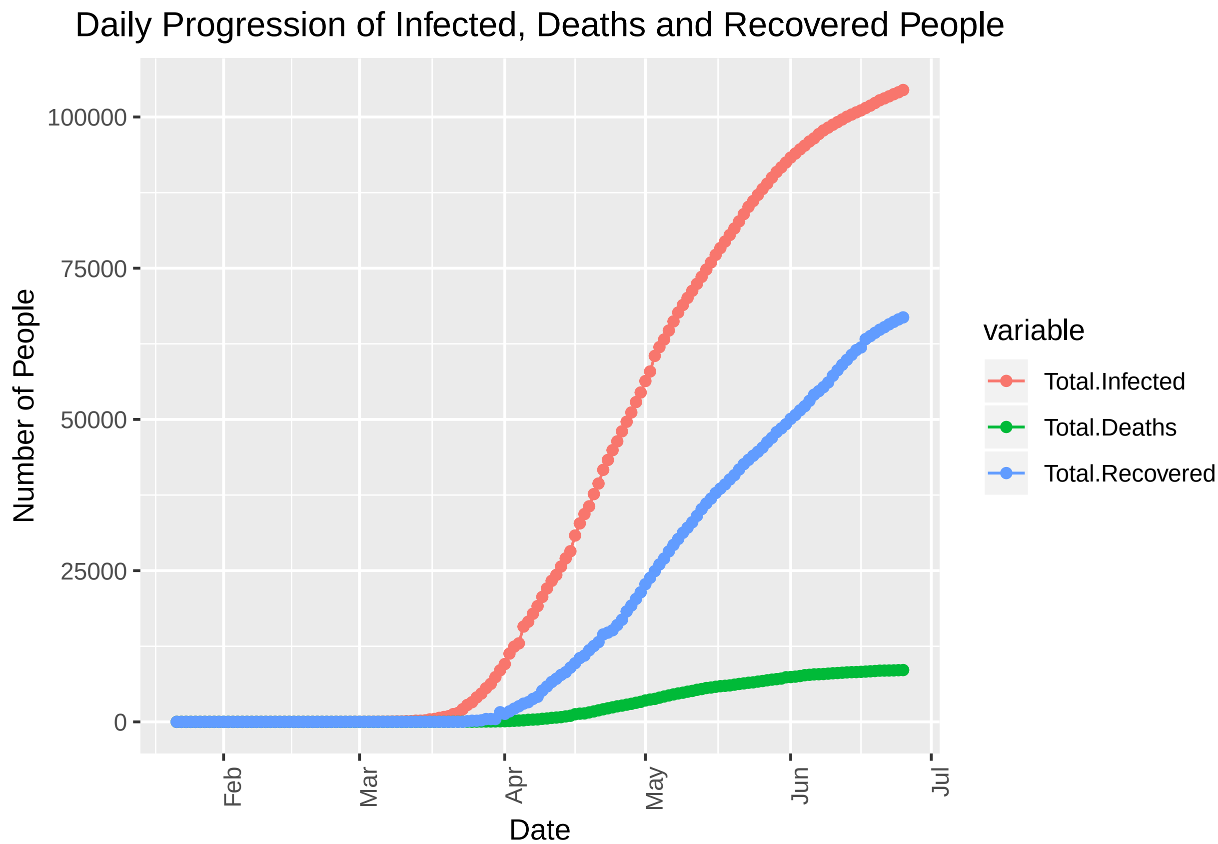
It follows that we have to provide the following parameters:

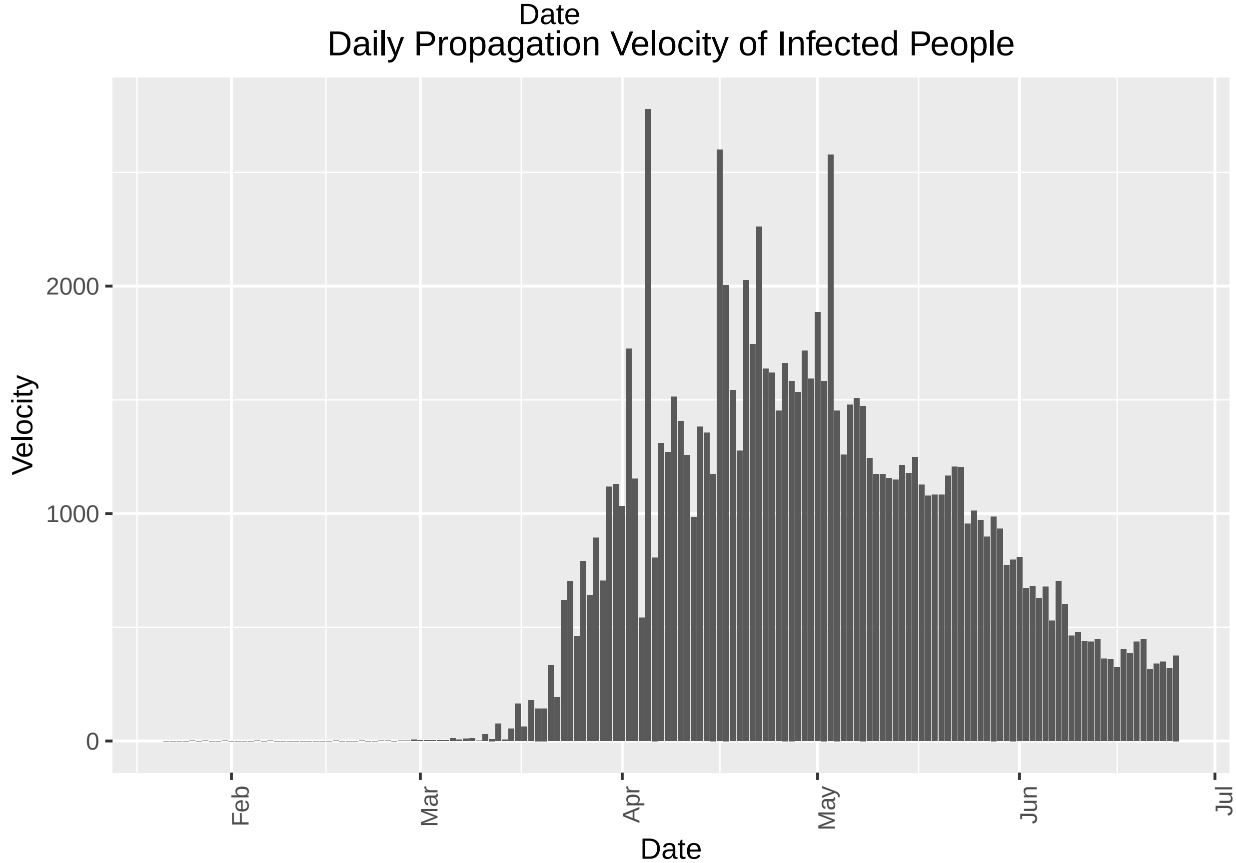
- the stabilization upper bound $\epsilon$;
- the number of days $k$ on which the mean $\mu(\Delta^2(t), k)$ will be calculated.
- The step of days $s$ between the calculation of means.

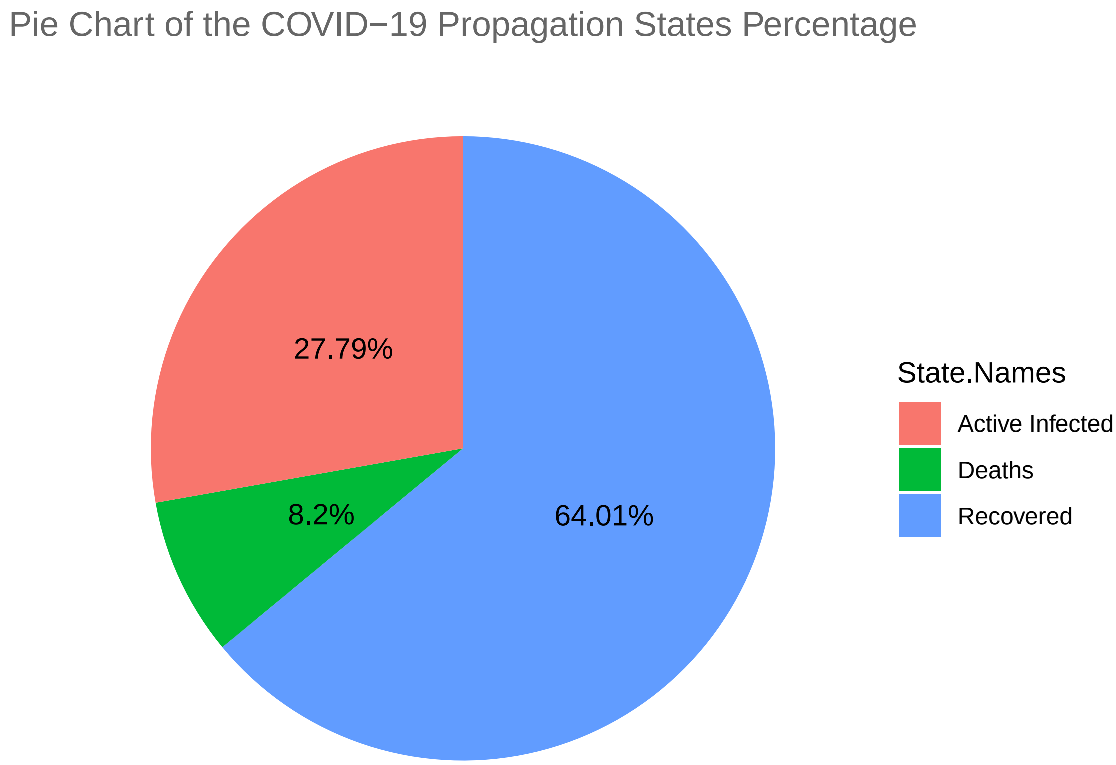
### 3.3 Canada Propagation Overview

The objective is to know the overall propagation in Canada.

| Date | Total.Infected | Total.Deaths | Total.Recovered | infected.delta | deaths.delta | recovered.delta |
|---|---|---|---|---|---|---|
| 2020-06-12 | 99595 | 8125 | 59034 | 436 | 54 | 903 |
| 2020-06-13 | 100043 | 8183 | 59851 | 448 | 58 | 817 |
| 2020-06-14 | 100404 | 8218 | 60668 | 361 | 35 | 817 |
| 2020-06-15 | 100763 | 8228 | 61466 | 359 | 10 | 798 |
| 2020-06-16 | 101087 | 8271 | 61899 | 324 | 43 | 433 |
| 2020-06-17 | 101491 | 8312 | 63280 | 404 | 41 | 1381 |
| 2020-06-18 | 101877 | 8361 | 63782 | 386 | 49 | 502 |
| 2020-06-19 | 102314 | 8408 | 64318 | 437 | 47 | 536 |
| 2020-06-20 | 102762 | 8466 | 64826 | 448 | 58 | 508 |
| 2020-06-21 | 103078 | 8482 | 65249 | 316 | 16 | 423 |
| 2020-06-22 | 103418 | 8494 | 65721 | 340 | 12 | 472 |
| 2020-06-23 | 103767 | 8512 | 66135 | 349 | 18 | 414 |
| 2020-06-24 | 104087 | 8544 | 66533 | 320 | 32 | 398 |
| 2020-06-25 | 104463 | 8567 | 66869 | 376 | 23 | 336 |


Daily Progression of Infected, Deaths and Recovered People
Number of People
100000
75000
50000
25000
0
Feb
Mar
Apr
May
Jun
Jul
Date
variable
Total.Infected
Total.Deaths
Total.Recovered


Daily Propagation Velocity of Infected People
Velocity
2000
1000
0
Feb
Mar
Apr
May
Jun
Jul
Date




## 4 Propagation Model

The COVID-19 is currently a worldwide pandemic virus which is considered very virulant. It means that it propagate from infected people to non-infected people by direct or indirect contacts. For exemple, if an infected person touches an object, a non-infected person touching the same object after a short time is mostly at risk to be infected.

The objective is to define a model that represents the propagation of the COVID-19 based on assumptions. However, we should take a look on variables that could have an impact on the propagation of the virus. Here is a list of some of those variables:

1. *Population density*: Countries with high population density should be more at risk because the contact between people is much easier hence more at risk to propagate the virus.
2. *Age of people*: Elder people are mostly to die after being infected by the virus because they are more fragile than younger people.
3. *People with chronic diseases*: People with chronic diseases like heart disease, lung disease, kidney disease, cancer, Alzheimer, diabetes, asthma and many others are more at risk to die after being infected.
4. *Births and deaths*: Since the propagation of the virus is a long time period, during this peiod, some people will die from any other causes than the COVID-19 which will decrease the population. In the other case, some women will give birth which will increase the population.
5. *Safety measures*: During the pandemic, many countries adopted safety measures in order to help reducing the contamination between people.
6. *Number of COVID-19 tests*: Since these tests are expensive, they are limited. Coutries that are part of the third world coutries will have less tests than the other countries. Therefore, the number of tests should at least be function of the country.

7. *Number of infected people not tested*: Some people that want to be tested because they might be infected by the COVID-19 are not tested because they are not considered as *essential*. By essential, we mean that the probability they infect others is much greater than other people that do not interact with people in their work (examples of essential people: police officers, nurses, docters). Other people could be infected and prefer to stay at home without asking to be tested.
8. *Infected error factor*: It may happen that a person has been tested positive to the COVID-19 but is not infected at all. Thus, errors when testing could happen.
9. *Recovery error factor*: Errors could happen when people are considered to have recovered but in fact, they did not recover yet. They identified them as recovered too soon.
10. *Death error factor*: It may happen that a person has not died from the COVID-19 but is counted as being dead because of the COVID-19.

We did not consider the immunity against the COVID-19 because we are incertain if there are people that will never be infected by the COVID-19 because they are immune. The same incertainty holds for people recovering from the COVID-19. We do not know if they are immune for the rest of their life or it is like the Influenza; they can be infected after a period of time (virus mutation for example).

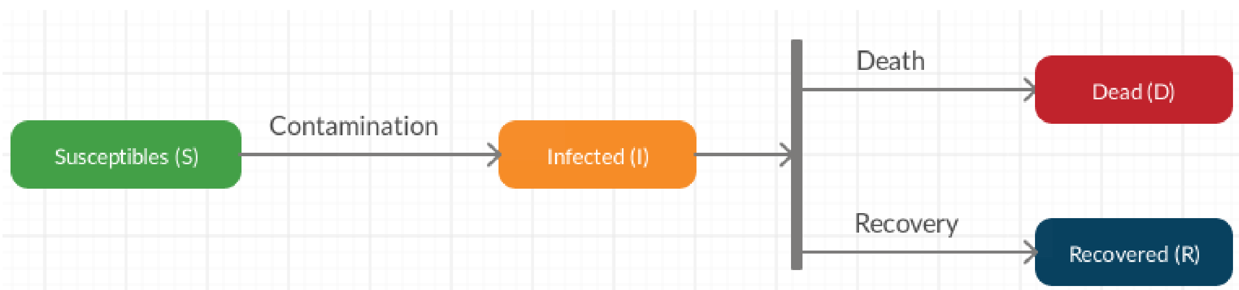
## 4.1 States and Transitions

A person can be in one of the following states during the propagation:

- Susceptible (noted $S(t)$): People susceptible to be infected by the COVID-19 at day $t$.
- Infected (noted $I_a(t)$): People tested positive to the COVID-19 at day $t$ but did neither recovered nor died yet.
- Dead (noted $D(t)$): People died from the COVID-19 at day $t$.
- Recovered (noted $R(t)$): People recovered from the COVID-19 at day $t$.

Initially, all people in the population are in the *Susceptible* state. Then, it needs at least one person to start the propagation of the virus. This starts at day 1 (2020-01-22) given in our dataset and ends actually on 2020-06-25.

Here is a state diagram describing the interaction between the states:




The arrows between the states represent the transitions between 2 states.

## 4.2 Assumptions

The following assumptions are made to simplify the model:

1. A person taken randomly in the population has the same probabilities to be infected than any other person taken randomly in the population (it follows a uniform distribution). Therefore, any people have the same probabilities (homogenous population) to be infected without considering their age or if they have a chronical diseases.

2. An infected person could stay infected, recover or die the next day. There is no error made when a person is categorized as infected by the COVID-19. It means that it is not possible for a person to be in the *Infected* state and then transits back to the *Susceptible* state because an error has been made.
3. Every person that recovers from the COVID-19 are immune against it. It means that once a person recovered, that person cannot be infected anymore. Therefore, there is no transition between the *Recovered* state and the *Susceptible* state or *Infected* state.
4. During the propagation of the virus, there are neither births nor deaths (demography is ignored). The initial population is fixed to a constant $N$.
5. We know that there are more people infected than what the dataset is providing. Many circumstances make that these people have not been tested yet against the COVID-19. For simplicity, we assume that the dataset provides the right values of infected people. It means that we will not add an additional estimator to estimate the number of susceptible people that are in fact infected but not tested yet. However, we know that this assumption is not representative of the reality because the number of tests is limited. Indeed, some people that want to be tested because they have the COVID-19 symptomes are not tested because they are not considered as *essential*. By essential people, we mean that the probability that they infect others is greater than other people that do not interact with people in their work (e.g. police officers, nurses, docters). These tests are also expensive which also explain why they are limited.
6. There are no safety measures taken during the pandemy. It means that there is no quarantine and social distancing between people, and any other safety measures.
7. All the population will have been infected one day. It does **not** take for account that some of the susceptible people could be immune against the virus, could never be infected or have not been tested but got infected by the COVID-19 and recovered.
8. The density of the population is independent of the propagation of the COVID-19.

## 4.3 SIR Epidemic Model

According to the assumptions, there are 3 transition phases between states on which our model is based:

- From *Susceptible* to *Infected* between days $t$ and $t+1$
- From *Infected* to *Recovered* between days $t$ and $t+1$
- From *Infected* to *Dead* between days $t$ and $t+1$

For exemple, if at day $t = 1$ there are 2 infected people and at day $t = 2$, there are 5 infected people and 1 dead, then between days $t = 1$ and $t = 2$, there are 4 people that transited from the *Susceptible* state to the *Infected* state and 1 person transited from the *Infected* state to the *Dead* state.

Per assumption 4, let the initial fixed population noted $N$ be

$$N = S(t) + I_a(t) + R(t) + D(t)$$

where $S(t)$ will decrease while $I_a(t)+R(t)+D(t)$ will increase over days. On the first day of the propagation, there has to have at least one person infected in order to prapagate the virus to susceptible people. Generally, at this initial state, there are neither recovered nor dead people because they have to be infected before. However, it depends on the initial values given in the dataset. It may happen that the data have been gathered later like it is for our dataset.

We introduce $I_c(t)$ the **cumulative number of infected people** at day $t$ because our dataset provides this feature. It means that $N = S(t) + I_c(t)$.

Per assumption 1, each infected person can be in contact with susceptible people and has the probability $\beta$ to infect each of them. Therefore, each infected person generates $\beta S(t)$ infected people every day. This is true for all infected people ($I_c(t)$), therefore the total number of infected people generated is $\beta S(t) I_c(t)$. The population will then decrease at this rate.

The transition between the suceptible state and the infected state is represented by the equation

$$\frac{\partial S(t)}{\partial t} = -\beta S(t) I_c(t).$$

Per assumption 2, there is a probability $\gamma$ that active infected people will recover (transition from *Infected* to *Recovered* state) or a probability of $\alpha$ that an active infected person will die (transition from *Infected* to *Dead* state) fromt day $t-1$ to $t$. The transitions between the *Infected* state and the *Recovered* state or *Dead* state are given by

$$\begin{aligned}\frac{\partial R(t)}{\partial t} &= \gamma I_a(t)\\ \frac{\partial D(t)}{\partial t} &= \alpha I_a(t).\end{aligned}$$

We know that the number of susceptible people decreses when they become infected. It follows that the number of infected people increases by the same value. Therefore, we have that

$$\frac{\partial I_c(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} = \beta S(t) I_c(t).$$

We did not remove the deaths and recovered people from the infected ones because in our case, the number of infected people is cumulative ($I_c$). However, to fit with our state diagram, we have to consider the **active** infected people ($I_a(t)$). This means that these people are infected by the COVID-19 but did neither recovered or died yet. Therefore, we have to subtract the deaths and recovered from the number of infected people:

$$\frac{\partial I_a(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} - \frac{\partial R(t)}{\partial t} - \frac{\partial D(t)}{\partial t} = \beta S(t) I_c(t) - (\gamma + \alpha) I_a(t).$$

We have the following equations that represent our state transition model:

$$\begin{aligned}\frac{\partial S(t)}{\partial t} &= -\beta S(t) I_c(t)\\ \frac{\partial I_c(t)}{\partial t} &= \beta S(t) I_c(t)\\ \frac{\partial I_a(t)}{\partial t} &= \beta S(t) I_c(t) - (\gamma + \alpha) I_a(t)\\ \frac{\partial R(t)}{\partial t} &= \gamma I_a(t)\\ \frac{\partial D(t)}{\partial t} &= \alpha I_a(t)\end{aligned}$$

Let $R_0 = \frac{\beta}{\gamma+\alpha}$ be the number of infected people over the revovered and dead ones where $0 < \gamma + \alpha \leq 1$. We expect that $R_0 > 1$ will increase during the rising part of the propagation (contamination phase). Then, we expect that $R_0$ will decrease over the days and be nearer to 0 because the contamination phase will slow down while the recovery and death phases will increase faster. Finally, all phases will stabilize slowly to $R_0 = 1$.

## 4.4 Example

The example is based on the data we have for the Canada in this dataset. The first infected person appears to be on 2020-01-26.

Thus, let $I_c(0) = 1$, $R(0) = 0$, $D(0) = 0$ and fix the population to $N = 37058856$ people. It follows that $S(0) = 37058855$. Lets also fix the model parameters to $\alpha = 0.0065$, $\beta = 0.0000000045$ and $\gamma = 0.045$.

Lets see the results of the first iteration:

$$\frac{\partial S(t)}{\partial t} = -0.0000000045 \times 37058855 \times 1 = -0.1667648475$$
$$\frac{\partial I_a(t)}{\partial t} = 0.0000000045 \times 37058855 \times 1 - 0.045 \times 1 - 0.0065 \times 1 = 0.1152648475$$
$$\frac{\partial R(t)}{\partial t} = 0.045 \times 1 = 0.045$$
$$\frac{\partial D(t)}{\partial t} = 0.0065 \times 1 = 0.0065$$

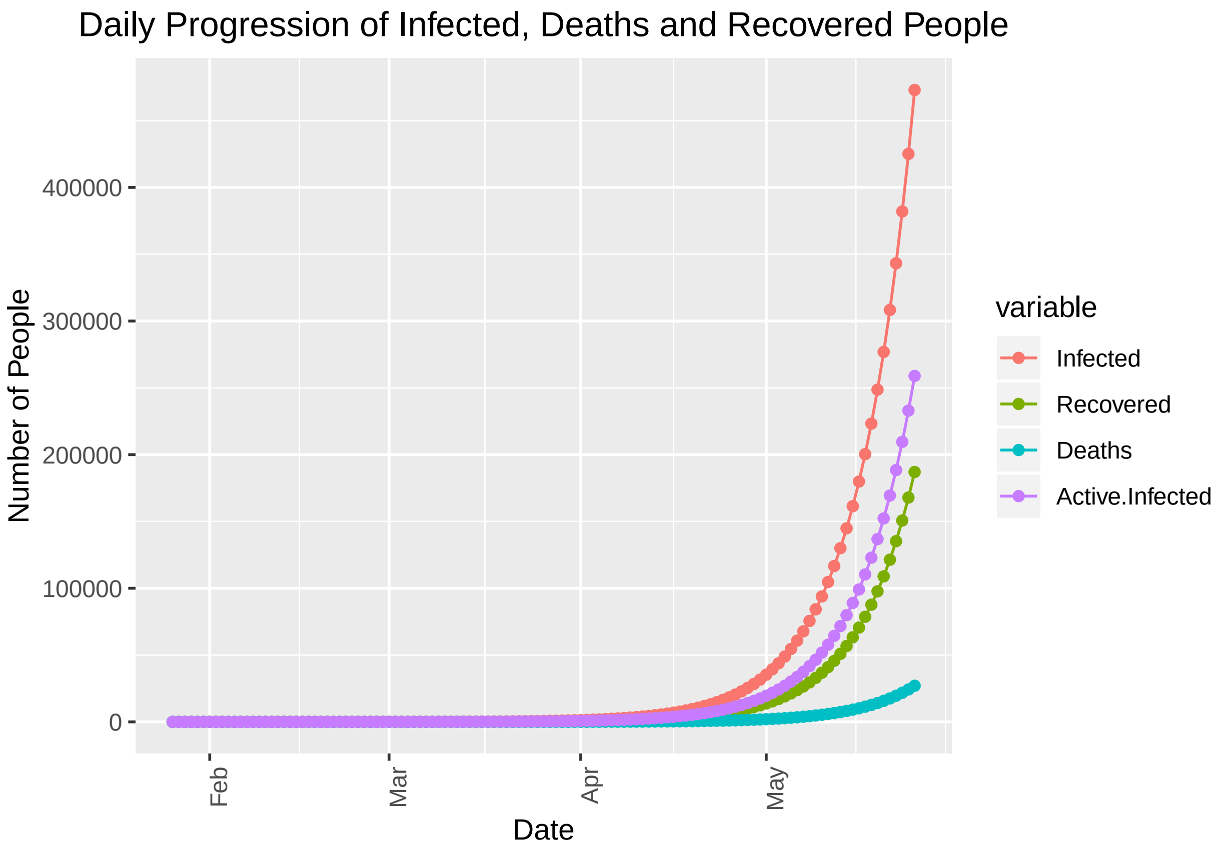
Therefore, we obtain $I(1) = 0.8332351525$, $S(1) = 37058854.8332352$, $R(1) = 0.045$ and $D(1) = 0.0065$.

Lets simulate our model for 120 days to see the propression of each state over the days.

| Date | Susceptibles | Infected | Recovered | Deaths | Active.Infected |
|---|---|---|---|---|---|
| 2020-01-26 | 37058855 | 1 | 0 | 0 | 1 |
| 2020-01-27 | 37058854 | 1 | 0 | 0 | 1 |
| 2020-01-28 | 37058854 | 1 | 0 | 0 | 1 |
| 2020-01-29 | 37058854 | 1 | 0 | 0 | 1 |
| 2020-01-30 | 37058854 | 1 | 0 | 0 | 1 |
| 2020-01-31 | 37058853 | 1 | 0 | 0 | 1 |
| 2020-02-01 | 37058853 | 1 | 0 | 0 | 1 |
| 2020-02-02 | 37058853 | 2 | 0 | 0 | 2 |
| 2020-02-03 | 37058852 | 2 | 0 | 0 | 2 |
| 2020-02-04 | 37058852 | 2 | 0 | 0 | 2 |
| 2020-02-05 | 37058852 | 2 | 0 | 0 | 2 |
| 2020-02-06 | 37058851 | 3 | 0 | 0 | 3 |
| 2020-02-07 | 37058851 | 3 | 1 | 0 | 2 |
| 2020-02-08 | 37058850 | 4 | 1 | 0 | 3 |
| 2020-02-09 | 37058849 | 4 | 1 | 0 | 3 |
| 2020-02-10 | 37058849 | 5 | 1 | 0 | 4 |
| 2020-02-11 | 37058848 | 5 | 1 | 0 | 4 |
| 2020-02-12 | 37058847 | 6 | 2 | 0 | 4 |
| 2020-02-13 | 37058846 | 7 | 2 | 0 | 5 |
| 2020-02-14 | 37058844 | 7 | 2 | 0 | 5 |
| 2020-02-15 | 37058843 | 8 | 3 | 0 | 5 |
| 2020-02-16 | 37058842 | 9 | 3 | 0 | 6 |
| 2020-02-17 | 37058840 | 11 | 3 | 0 | 8 |
| 2020-02-18 | 37058838 | 12 | 4 | 0 | 8 |
| 2020-02-19 | 37058836 | 13 | 4 | 0 | 9 |
| 2020-02-20 | 37058834 | 15 | 5 | 0 | 10 |
| 2020-02-21 | 37058831 | 17 | 6 | 0 | 11 |
| 2020-02-22 | 37058828 | 19 | 7 | 1 | 11 |
| 2020-02-23 | 37058825 | 21 | 7 | 1 | 13 |
| 2020-02-24 | 37058822 | 23 | 8 | 1 | 14 |
| 2020-02-25 | 37058818 | 26 | 9 | 1 | 16 |
| 2020-02-26 | 37058813 | 29 | 11 | 1 | 17 |
| 2020-02-27 | 37058808 | 32 | 12 | 1 | 19 |
| 2020-02-28 | 37058803 | 36 | 13 | 2 | 21 |
| 2020-02-29 | 37058797 | 40 | 15 | 2 | 23 |
| 2020-03-01 | 37058790 | 45 | 17 | 2 | 26 |
| 2020-03-02 | 37058782 | 50 | 19 | 2 | 29 |
| 2020-03-03 | 37058774 | 56 | 21 | 3 | 32 |
| 2020-03-04 | 37058765 | 63 | 24 | 3 | 36 |

| Date | Susceptibles | Infected | Recovered | Deaths | Active.Infected |
|---|---|---|---|---|---|
| 2020-03-05 | 37058754 | 70 | 27 | 3 | 40 |
| 2020-03-06 | 37058742 | 78 | 30 | 4 | 44 |
| 2020-03-07 | 37058729 | 87 | 33 | 4 | 50 |
| 2020-03-08 | 37058715 | 97 | 37 | 5 | 55 |
| 2020-03-09 | 37058698 | 108 | 42 | 6 | 60 |
| 2020-03-10 | 37058680 | 121 | 47 | 6 | 68 |
| 2020-03-11 | 37058660 | 135 | 52 | 7 | 76 |
| 2020-03-12 | 37058637 | 151 | 58 | 8 | 85 |
| 2020-03-13 | 37058612 | 168 | 65 | 9 | 94 |
| 2020-03-14 | 37058584 | 187 | 73 | 10 | 104 |
| 2020-03-15 | 37058553 | 209 | 81 | 11 | 117 |
| 2020-03-16 | 37058518 | 233 | 90 | 13 | 130 |
| 2020-03-17 | 37058479 | 260 | 101 | 14 | 145 |
| 2020-03-18 | 37058435 | 290 | 113 | 16 | 161 |
| 2020-03-19 | 37058387 | 324 | 126 | 18 | 180 |
| 2020-03-20 | 37058333 | 361 | 140 | 20 | 201 |
| 2020-03-21 | 37058272 | 403 | 157 | 22 | 224 |
| 2020-03-22 | 37058205 | 449 | 175 | 25 | 249 |
| 2020-03-23 | 37058130 | 501 | 195 | 28 | 278 |
| 2020-03-24 | 37058046 | 559 | 218 | 31 | 310 |
| 2020-03-25 | 37057953 | 624 | 243 | 35 | 346 |
| 2020-03-26 | 37057849 | 696 | 271 | 39 | 386 |
| 2020-03-27 | 37057733 | 776 | 302 | 43 | 431 |
| 2020-03-28 | 37057603 | 865 | 337 | 48 | 480 |
| 2020-03-29 | 37057459 | 965 | 376 | 54 | 535 |
| 2020-03-30 | 37057298 | 1076 | 420 | 60 | 596 |
| 2020-03-31 | 37057118 | 1201 | 468 | 67 | 666 |
| 2020-04-01 | 37056918 | 1339 | 522 | 75 | 742 |
| 2020-04-02 | 37056695 | 1493 | 582 | 84 | 827 |
| 2020-04-03 | 37056446 | 1665 | 650 | 93 | 922 |
| 2020-04-04 | 37056168 | 1857 | 725 | 104 | 1028 |
| 2020-04-05 | 37055858 | 2072 | 808 | 116 | 1148 |
| 2020-04-06 | 37055512 | 2310 | 901 | 130 | 1279 |
| 2020-04-07 | 37055127 | 2577 | 1005 | 145 | 1427 |
| 2020-04-08 | 37054697 | 2874 | 1121 | 162 | 1591 |
| 2020-04-09 | 37054218 | 3205 | 1251 | 180 | 1774 |
| 2020-04-10 | 37053684 | 3574 | 1395 | 201 | 1978 |
| 2020-04-11 | 37053087 | 3986 | 1556 | 224 | 2206 |
| 2020-04-12 | 37052423 | 4446 | 1735 | 250 | 2461 |
| 2020-04-13 | 37051681 | 4958 | 1935 | 279 | 2744 |
| 2020-04-14 | 37050854 | 5530 | 2158 | 311 | 3061 |
| 2020-04-15 | 37049932 | 6167 | 2407 | 347 | 3413 |
| 2020-04-16 | 37048904 | 6878 | 2685 | 387 | 3806 |
| 2020-04-17 | 37047757 | 7670 | 2994 | 432 | 4244 |
| 2020-04-18 | 37046479 | 8554 | 3340 | 482 | 4732 |
| 2020-04-19 | 37045053 | 9539 | 3724 | 538 | 5277 |
| 2020-04-20 | 37043462 | 10638 | 4154 | 600 | 5884 |
| 2020-04-21 | 37041689 | 11864 | 4633 | 669 | 6562 |
| 2020-04-22 | 37039711 | 13231 | 5166 | 746 | 7319 |
| 2020-04-23 | 37037506 | 14755 | 5762 | 832 | 8161 |
| 2020-04-24 | 37035047 | 16454 | 6426 | 928 | 9100 |
| 2020-04-25 | 37032304 | 18349 | 7166 | 1035 | 10148 |

| Date | Susceptibles | Infected | Recovered | Deaths | Active.Infected |
|---|---|---|---|---|---|
| 2020-04-26 | 37029247 | 20462 | 7992 | 1154 | 11316 |
| 2020-04-27 | 37025837 | 22817 | 8913 | 1287 | 12617 |
| 2020-04-28 | 37022035 | 25444 | 9940 | 1435 | 14069 |
| 2020-04-29 | 37017796 | 28373 | 11085 | 1601 | 15687 |
| 2020-04-30 | 37013070 | 31638 | 12361 | 1785 | 17492 |
| 2020-05-01 | 37007800 | 35278 | 13785 | 1991 | 19502 |
| 2020-05-02 | 37001925 | 39337 | 15373 | 2220 | 21744 |
| 2020-05-03 | 36995375 | 43861 | 17143 | 2476 | 24242 |
| 2020-05-04 | 36988073 | 48904 | 19117 | 2761 | 27026 |
| 2020-05-05 | 36979933 | 54525 | 21317 | 3079 | 30129 |
| 2020-05-06 | 36970859 | 60791 | 23771 | 3433 | 33587 |
| 2020-05-07 | 36960746 | 67774 | 26506 | 3828 | 37440 |
| 2020-05-08 | 36949473 | 75556 | 29556 | 4269 | 41731 |
| 2020-05-09 | 36936910 | 84227 | 32956 | 4760 | 46511 |
| 2020-05-10 | 36922910 | 93890 | 36747 | 5307 | 51836 |
| 2020-05-11 | 36907310 | 104655 | 40972 | 5918 | 57765 |
| 2020-05-12 | 36889929 | 116646 | 45681 | 6598 | 64367 |
| 2020-05-13 | 36870565 | 130003 | 50930 | 7356 | 71717 |
| 2020-05-14 | 36848995 | 144877 | 56780 | 8201 | 79896 |
| 2020-05-15 | 36824971 | 161440 | 63300 | 9143 | 88997 |
| 2020-05-16 | 36798219 | 179878 | 70565 | 10192 | 99121 |
| 2020-05-17 | 36768432 | 200401 | 78659 | 11361 | 110381 |
| 2020-05-18 | 36735274 | 223239 | 87677 | 12664 | 122898 |
| 2020-05-19 | 36698371 | 248645 | 97723 | 14115 | 136807 |
| 2020-05-20 | 36657309 | 276902 | 108912 | 15731 | 152259 |
| 2020-05-21 | 36611631 | 308319 | 121373 | 17531 | 169415 |
| 2020-05-22 | 36560835 | 343237 | 135247 | 19535 | 188455 |
| 2020-05-23 | 36504364 | 382030 | 150693 | 21766 | 209571 |
| 2020-05-24 | 36441608 | 425112 | 167884 | 24250 | 232978 |
| 2020-05-25 | 36371895 | 472932 | 187014 | 27013 | 258905 |




### 4.5 Model Representation With Markov Chain

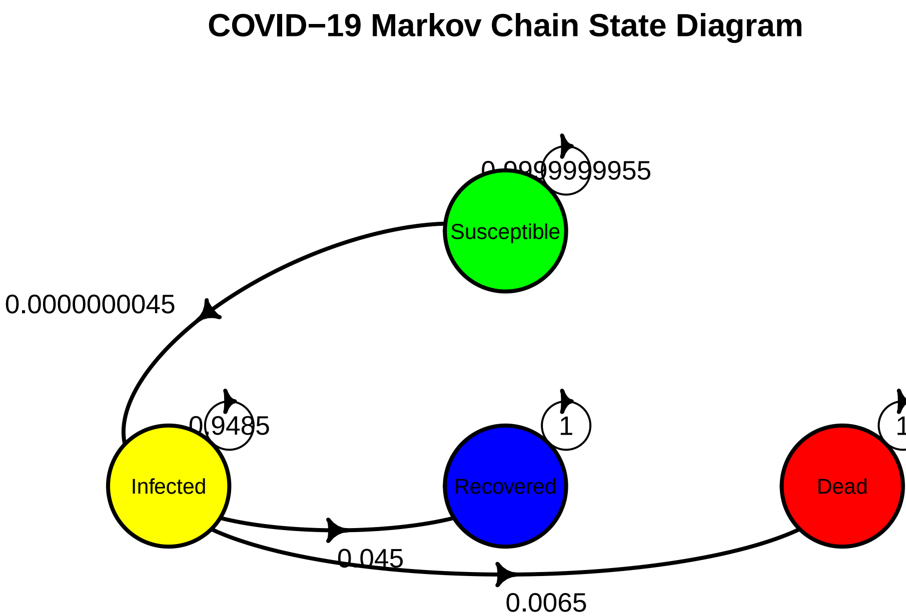
Let $X = (X_t)_{t \geq 0}$ be a sequence of random variables where each random variable is representing a person $X$ at day $t \geq 0$. Let $\mathbb{E} = \{S, I, R, D\}$ the COVID-19 state space. At any day $t$, a person $X_t$ has to be in a state of $\mathbb{E}$. If $i, j \in \mathbb{E}$, then $\mathbb{P}(X_t \in j | X_{t-1} \in i) = p_{i,j}$. A person at day $t = 0$ as to start in a state of $\mathbb{E}$.

The probability that a susceptible person becomes infected at day $t$ is defined as $\mathbb{P}(X_t \in I | X_{t-1} \in S) = \beta$. The susceptible people that do not transit to the infected state is expressed as $\mathbb{P}(X_t \in S | X_{t-1} \in S) = 1 - \beta$. Per assumption 5, it is impossible to transit from the *Infected* state to the *Susceptible* state. Therefore, we have $\mathbb{P}(X_t \in S | X_{t-1} \in I) = 0$.

The assumption 2 states that an infexcted person can transit to the death state or to the recovery state. It means that $\mathbb{P}(X_t \in R | X_{t-1} \in I) = \gamma$ and $\mathbb{P}(X_t \in D | X_{t-1} \in I) = \alpha$. Since the sets of recovered $R$ and deaths $D$ are mutually disjoint, we have that

$$\mathbb{P}(X_t \in R \cup D | X_{t-1} \in I) = \mathbb{P}(X_t \in R | X_{t-1} \in I) + \mathbb{P}(X_t \in D | X_{t-1} \in I) = \gamma + \alpha$$

We deduce that the remaining infected people that will stay in the *Infected* state (will not transit to another state on the next day) from day $t - 1$ to day $t$ is $\mathbb{P}(X_t \in I | X_{t-1} \in I) = 1 - \alpha - \gamma$.




The transition matrix (noted $P$) is the following considering the order in $\mathbb{E}$ and the transition between day $t-1$ and day $t$:

$$P = \begin{bmatrix} p_{S,S} & p_{S,I} & p_{S,R} & p_{S,D} \\ p_{I,S} & p_{I,I} & p_{I,R} & p_{I,D} \\ p_{R,S} & p_{R,I} & p_{R,R} & p_{R,D} \\ p_{D,S} & p_{D,I} & p_{D,R} & p_{D,D} \end{bmatrix} = \begin{bmatrix} 1-\beta & \beta & 0 & 0 \\ 0 & 1-\alpha-\gamma & \gamma & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Let's assume that there is an infected person, no deaths and no recovery on the first day. Including that there are $N-1$ susceptible people, this means that if $\mathbf{x}^{(0)}$ is the initial vector, we have

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)}P = \begin{bmatrix} N-1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1-\beta & \beta & 0 & 0 \\ 0 & 1-\alpha-\gamma & \gamma & \alpha \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} (N-1)(1-\beta) & (N-1)\beta + (1-\alpha-\gamma) & \gamma & \alpha \end{bmatrix}$$

On the second day, we have

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)}P = \begin{bmatrix} (N-1)(1-\beta)^2 \\ \beta(N-1)((1-\beta)+(1-\alpha-\gamma)) + (1-\alpha-\gamma)^2 \\ \beta(N-1)\gamma + \gamma((1-\alpha-\gamma)+1) \\ \beta(N-1)\alpha + \alpha((1-\alpha-\gamma)+1) \end{bmatrix}^T$$

For $n$ days, we have $\mathbf{x}^{(n)} = \mathbf{x}^{(0)}P^n$. By induction on $n \geq 1$, one shows that

$$\mathbf{x}^{(n)} = \mathbf{x}^{(0)}P^n = \begin{bmatrix} (N-1)(1-\beta)^n \\ \beta(N-1)\sum_{i=0}^{n-1}(1-\beta)^{n-1-i}(1-\alpha-\gamma)^i + (1-\alpha-\gamma)^n \\ \beta(N-1)\gamma\sum_{i=0}^{n-2}(1-\beta)^i(1-\alpha-\gamma)^{n-2-i} + \gamma\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i \\ \beta(N-1)\alpha\sum_{i=0}^{n-2}(1-\beta)^i(1-\alpha-\gamma)^{n-2-i} + \alpha\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i \end{bmatrix}^T$$

Let's see what is the probability law $\mathbf{x}$. We have to find $\mathbf{x} = \lim_{n\to\infty} \mathbf{x}^{(n)}$. We calculate this limit term by term in $\mathbf{x}^{(n)}$. The first term:

$$\lim_{n\to\infty}(N-1)(1-\beta)^n = 0$$

because $0 < \beta \leq 1$ then $0 \leq (1-\beta) < 1$. For the second term, we have that

$$\lim_{n\to\infty} \beta(N-1)\sum_{i=0}^{n-1}(1-\beta)^{n-1-i}(1-\alpha-\gamma)^i+(1-\alpha-\gamma)^n = \beta(N-1)\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\beta)^{n-1-i}(1-\alpha-\gamma)^i+\lim_{n\to\infty}(1-\alpha-\gamma)^n = 0.$$

Indeed, we get $\lim_{n\to\infty}(1-\alpha-\gamma)^n = 0$ because $0 \leq (1-\alpha-\gamma) < 1$. We also have that $\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\beta)^{n-1-i}(1-\alpha-\gamma)^i = 0$ because $\lim_{n\to\infty}(1-\beta)^{n-1-i} = 0$ since $0 \leq \beta < 1$. For the third term, we have that

$$\beta(N-1)\gamma\lim_{n\to\infty}\sum_{i=0}^{n-2}(1-\beta)^i(1-\alpha-\gamma)^{n-2-i}+\gamma\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i = \gamma\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i = \gamma\sum_{i=0}^{\infty}(1-\alpha-\gamma)^i$$

because using the same properties as for the second term, the first limit is evaluated to 0. For the second limit, we have a geometric series of ratio $(1-\alpha-\gamma) < 1$. Note that it is impossible to have $(1-\alpha-\gamma) = 1$ because it would mean that $\alpha = \gamma = 0$ which contradicts our assumption stating that an infected person at day $t$ has to transit to either the *Recovered* state or the *Dead* state at day $t+k$ where $k \geq 1$.

Therefore, we have that

$$\gamma\sum_{i=0}^{\infty}(1-\alpha-\gamma)^i = \frac{\gamma}{1-(1-\alpha-\gamma)} = \frac{\gamma}{\alpha+\gamma}.$$

For the fourth term, we use the same logic as the third term

$$\beta(N-1)\alpha\lim_{n\to\infty}\sum_{i=0}^{n-2}(1-\beta)^i(1-\alpha-\gamma)^{n-2-i}+\alpha\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i = \alpha\lim_{n\to\infty}\sum_{i=0}^{n-1}(1-\alpha-\gamma)^i = \frac{\alpha}{\alpha+\gamma}.$$

Therefore, the final result is

$$\mathbf{x} = \begin{bmatrix}0 & 0 & \frac{\gamma}{\alpha+\gamma} & \frac{\alpha}{\alpha+\gamma}\end{bmatrix}.$$

We expect that with the SIR model based on our assumptions, all susceptible people will be infected and then, all infected people will either recover or die. In other terms, we expect that at the end of the COVID-19 pandemic, a percentage of the population will recover while the rest of the population will die. Therefore, our vector $\mathbf{x}$ makes sense because $\frac{\gamma}{\alpha+\gamma}+\frac{\alpha}{\alpha+\gamma} = 1$.

If we take back our example where $\gamma = 0.045$ and $\alpha = 0.0065$, we obtain

$$\frac{\gamma}{\alpha+\gamma} = \frac{0.045}{0.0515} = 0.8737864078$$

This means that 87.3786407767 % of the population will recover and 12.6213592233 % will die once the pandemic will end.

## 4.6 SIR Model Solutions

The objective is to solve the first-order differential equations defined for the 3 transitions in order to find $S(t)$, $I_a(t)$ and $R(t)$ in function of $t$. However, it is useful to mention that the propagation is described basically by the 3 following phases:

1. **Initialization** The propagation will start and slowly progress; only a small number of people will be infected at the beginning of the pandemic.
2. **Acceleration** The propagation will increase quickly because every person infected will infect other people that were not infected making the propression increasing exponentially.
3. **Resolution** The worst case is when the majority of the population is infected. Indeed, there are less non-infected people remaining and the progression will have no choice but to slow down until all people of the population are infected.

These phases justify why the sigmoid function is a good choice. This sigmoid function $I_c(t)$ starts at $t = 0$ and ends until all the population is infected. Therefore, we deduce that the bounds of $t$ and $I_c(t)$ are $t \in [0, \infty[$ and $I_c(t) \in [0, N]$ where $I_c(t)$ will increase over days (cumulative function). It also follows that $\lim_{t\to\infty} I_c(t) = N$.

### 4.6.1 Cumulative Infected and Susceptible People Models

We know that when susceptible people are getting infected, the same number of people decreases from susceptible to increase in the infected state from day $t-1$ to day $t$. It means that $I_c(t) = N - S(t)$. Therefore, we have

$$\frac{\partial I_c(t)}{\partial t} = \beta I_c(t)(N - I_c(t)).$$

We have $\frac{\partial I_c(t)}{\partial t} = N\beta I_c(t) - \beta I_c^2(t)$ which is a Bernouilli's differential equation. Dividing the equation by $I_c^2(t)$ gives

$$\frac{\frac{\partial I_c(t)}{\partial t}}{I_c^2(t)} - \frac{N\beta}{I_c(t)} = -\beta.$$

Let $y(t) = -\frac{1}{I_c(t)}$. Then, we have $\frac{\partial y(t)}{\partial t} = \frac{1}{I_c^2(t)}\frac{\partial I_c(t)}{\partial t}$. Replacing in the equation above, we get

$$\frac{\partial y(t)}{\partial t} = -\beta(1 + Ny(t)).$$

Dividing by $1 + Ny(t)$ on both sides gives

$$\frac{\frac{\partial y(t)}{\partial t}}{1 + Ny(t)} = -\beta.$$

Since $\int \frac{1}{1+Ny(t)}dt = \frac{1}{N}\ln(1 + Ny(t))$, we have, after multiplying by $N$ on both sides, that

$$\frac{\partial(\ln(1 + Ny(t)))}{\partial t} = -N\beta.$$

We know that $t \geq 0$, hence we have to integrate this last equation from 0 to $t$ on both sides:

$$\int_0^t \frac{\partial(\ln(1 + Ny(x)))}{\partial x} dx = -N\beta \int_0^t dx$$

which gives:

$$\ln(1 + Ny(t)) - \ln(1 + Ny(0)) = -N\beta t.$$

Using the substraction of logarithms property, we have:

$$\ln\left(\frac{1 + Ny(t)}{1 + Ny(0)}\right) = -N\beta t.$$

Applying the exponential function on both sides, it follows that

$$\frac{1 + Ny(t)}{1 + Ny(0)} = e^{-N\beta t}$$

which is equivalent to

$$y(t) = \frac{(1 + Ny(0))e^{-N\beta t} - 1}{N}.$$

Since $y(t) = -\frac{1}{I_c(t)}$ and then $y(0) = -\frac{1}{I_c(0)}$, we have

$$I_c(t) = \frac{N}{1 - \left(1 - \frac{N}{I_c(0)}\right)e^{-N\beta t}}$$

where $I_c(0) > 0$ because having $I_c(0) = 0$ implies that there are no infected people and then the propagation cannot start. Using the same method to solve $\frac{\partial S(t)}{\partial t} = -\beta S(t)(N - S(t))$, we obtain

$$S(t) = \frac{N}{1 + \left(\frac{N}{S(0)} - 1\right) e^{N\beta t}}$$

where $S(0) > 0$ because having $S(0) = 0$ implies that there is no population initially which is impossible.

We know that $t \in [0, \infty[$ and $S(t), I_c(t) \in [0, N]$. Taking the limit to infinite days, we have for $I_c(t)$ that:

$$\lim_{t\to\infty} \frac{N}{1 - \left(1 - \frac{N}{I_c(0)}\right) e^{-N\beta t}} = N$$

and for the initial day:

$$\lim_{t\to 0} \frac{N}{1 - \left(1 - \frac{N}{I_c(0)}\right) e^{-N\beta t}} = I_c(0).$$

Taking the limit to infinite days, we have for $S(t)$ that:

$$\lim_{t\to\infty} \frac{N}{1 + \left(\frac{N}{S(0)} - 1\right) e^{N\beta t}} = 0$$

and for the initial day:

$$\lim_{t\to 0} \frac{N}{1 + \left(\frac{N}{S(0)} - 1\right) e^{N\beta t}} = S(0).$$

### 4.6.2 Number of Recovered and Deaths Models

We know that $\frac{\partial R(t)}{\partial t} = \gamma I_a(t)$. But $I_a(t) = I_c(t) - R(t) - D(t)$ hence

$$\frac{\partial R(t)}{\partial t} = \gamma(I_c(t) - R(t) - D(t)).$$

However, $R(t)$ and $D(t)$ are the same except that the constant $\alpha$ is used instead of $\gamma$. Let $R_D(t) = R(t)+D(t)$ and $k = \gamma + \alpha$. It comes that

$$\frac{\partial R_D(t)}{\partial t} = k(I_c(t) - R_D(t))$$

which is equivalent to:

$$\frac{\partial R_D(t)}{\partial t} + kR_D(t) = kI_c(t).$$

This is a first-order differential equation. First, let's solve the homogeneous equation

$$\frac{\partial R_D(t)}{\partial t} + kR_D(t) = 0.$$

Hence, we have to solve the separated-variable equation:

$$\frac{\partial R_D(t)}{\partial t} = -kR_D(t).$$

After dividing by $R_D(t)$ on both sides:

$$\frac{\partial R_D(t)}{R_D(t)} = -k\partial t.$$

Then integrating both sides we obtain the general solution:

$$R_D(t) = R_0 e^{-kt}$$

where $R_0 \in \mathbb{R}$ is the integration constant. Suppose that $R_0$ is function of $t$ in $R_D(t)$ where $R_D(t) = R_0(t)e^{-kt}$ and substitute it in the initial equation:

$$\begin{aligned} kI_c(t) &= \frac{\partial R_0(t)e^{-kt}}{\partial t} + kR_0(t)e^{-kt} \\ &= \frac{\partial R_0(t)}{\partial t}e^{-kt} + R_0(t)\frac{\partial e^{-kt}}{\partial t} + kR_0(t)e^{-kt} \\ &= \frac{\partial R_0(t)}{\partial t}e^{-kt} - kR_0(t)e^{-kt} + kR_0(t)e^{-kt} \\ &= \frac{\partial R_0(t)}{\partial t}e^{-kt} \end{aligned}$$

It follows that

$$\frac{\partial R_0(t)}{\partial t} = kI_c(t)e^{kt}.$$

Knowing $I_c(t)$, we have to solve:

$$R_0(t) = kN \int \frac{e^{kt}dt}{1 - \left(1 - \frac{N}{I_c(0)}\right)e^{-N\beta t}}.$$

Let $\delta = 1 - \frac{N}{I_c(0)}$. The previous equation is then written as

$$R_0(t) = kN \int \frac{e^{kt}dt}{1 - \delta e^{-N\beta t}}.$$

The result of this integral is given by:

$$R_0(t) = \frac{-kNe^{t(N\beta+k)}{}_2F_1\left(1; \frac{k}{N\beta}+1; \frac{k}{N\beta}+2; \frac{e^{N\beta t}}{\delta}\right)}{\delta(N\beta+k)} + D_0$$

where ${}_2F_1(\cdot)$ is the Gauss's hypergeometric function and $D_0 \in \mathbb{R}$ is the integration constant. Subtituing $R_0(t)$ in $R_D(t) = R_0(t)e^{-kt}$ gives:

$$R_D(t) = \frac{-kNe^{N\beta t}{}_2F_1\left(1; \frac{k}{N\beta}+1; \frac{k}{N\beta}+2; \frac{e^{N\beta t}}{\delta}\right)}{\delta(N\beta+k)} + D_0e^{-kt}.$$

We can conclude on the following equations for $R(t)$:

$$R(t) = \frac{-\gamma Ne^{N\beta t}{}_2F_1\left(1; \frac{\gamma}{N\beta}+1; \frac{\gamma}{N\beta}+2; \frac{e^{N\beta t}}{\delta}\right)}{\delta(N\beta+\gamma)} + D_0e^{-\gamma t}$$

and for $D(t)$:

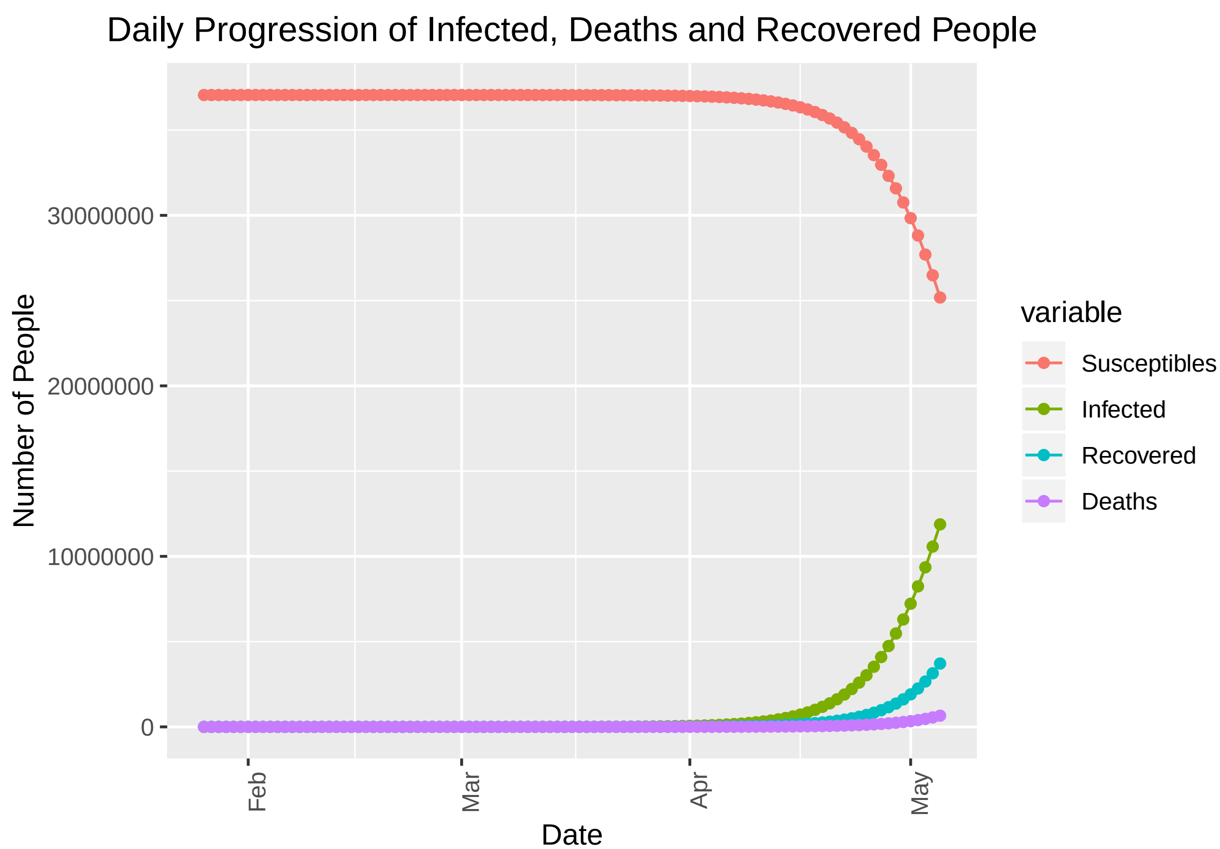
$$D(t) = \frac{-\alpha Ne^{N\beta t}{}_2F_1\left(1; \frac{\alpha}{N\beta}+1; \frac{\alpha}{N\beta}+2; \frac{e^{N\beta t}}{\delta}\right)}{\delta(N\beta+\alpha)} + D_0e^{-\alpha t}.$$

#### 4.6.3 Simulation

| Date | Susceptibles | Infected | Recovered | Deaths |
|---|---|---|---|---|
| 2020-01-26 | 37058855.00 | 1.000000000 | 0.0000000000 | 0.0000000000 |
| 2020-01-27 | 37058854.82 | 1.181476405 | 0.2510635708 | 0.0443228779 |
| 2020-01-28 | 37058854.60 | 1.395886494 | 0.2966256866 | 0.0523664347 |
| 2020-01-29 | 37058854.35 | 1.649206954 | 0.3504562515 | 0.0618697073 |

| Date | Susceptibles | Infected | Recovered | Deaths |
|---|---|---:|---:|---:|
| 2020-01-30 | 37058854.05 | 1.948499097 | 0.4140557942 | 0.0730975997 |
| 2020-01-31 | 37058853.70 | 2.302105698 | 0.4891971536 | 0.0863630898 |
| 2020-02-01 | 37058853.28 | 2.719883547 | 0.5779748972 | 0.1020359533 |
| 2020-02-02 | 37058852.79 | 3.213478208 | 0.6828637071 | 0.1205530719 |
| 2020-02-03 | 37058852.20 | 3.796648639 | 0.8067873618 | 0.1424306107 |
| 2020-02-04 | 37058851.51 | 4.485650724 | 0.9532002365 | 0.1682784067 |
| 2020-02-05 | 37058850.70 | 5.299690401 | 1.1261835942 | 0.1988169680 |
| 2020-02-06 | 37058849.74 | 6.261459031 | 1.3305593508 | 0.2348975578 |
| 2020-02-07 | 37058848.60 | 7.397765916 | 1.5720244862 | 0.2775259234 |
| 2020-02-08 | 37058847.26 | 8.740285606 | 1.8573098477 | 0.3278903319 |
| 2020-02-09 | 37058845.67 | 10.326440825 | 2.1943677727 | 0.3873946925 |
| 2020-02-10 | 37058843.80 | 12.200445626 | 2.5925937601 | 0.4576976909 |
| 2020-02-11 | 37058841.59 | 14.414537847 | 3.0630883704 | 0.5407590251 |
| 2020-02-12 | 37058838.97 | 17.030435237 | 3.6189666539 | 0.6388940321 |
| 2020-02-13 | 37058835.88 | 20.121055821 | 4.2757237332 | 0.7548382279 |
| 2020-02-14 | 37058832.23 | 23.772550471 | 5.0516667299 | 0.8918235603 |
| 2020-02-15 | 37058827.91 | 28.086704337 | 5.9684250768 | 1.0536684991 |
| 2020-02-16 | 37058822.82 | 33.183774068 | 7.0515534380 | 1.2448844765 |
| 2020-02-17 | 37058816.79 | 39.205839912 | 8.3312440467 | 1.4708016432 |
| 2020-02-18 | 37058809.68 | 46.320766130 | 9.8431683140 | 1.7377174464 |
| 2020-02-19 | 37058801.27 | 54.726880100 | 11.6294711708 | 2.0530721717 |
| 2020-02-20 | 37058791.34 | 64.658500548 | 13.7399458586 | 2.4256563405 |
| 2020-02-21 | 37058779.61 | 76.392468968 | 16.2334219180 | 2.8658557471 |
| 2020-02-22 | 37058765.74 | 90.255866287 | 19.1794050631 | 3.3859409620 |
| 2020-02-23 | 37058749.36 | 106.635129826 | 22.6600146557 | 4.0004093751 |
| 2020-02-24 | 37058730.01 | 125.986824665 | 26.7722727847 | 4.7263893103 |
| 2020-02-25 | 37058707.15 | 148.850369583 | 31.6308087594 | 5.5841174785 |
| 2020-02-26 | 37058680.14 | 175.863072216 | 37.3710544045 | 6.5975030762 |
| 2020-02-27 | 37058648.22 | 207.777892420 | 44.1530192267 | 7.7947942549 |
| 2020-02-28 | 37058610.52 | 245.484428810 | 52.1657506831 | 9.2093655394 |
| 2020-02-29 | 37058565.97 | 290.033713205 | 61.6326038851 | 10.8806481434 |
| 2020-03-01 | 37058513.33 | 342.667503809 | 72.8174676280 | 12.8552291160 |
| 2020-03-02 | 37058451.15 | 404.852893149 | 86.0321202986 | 15.1881499563 |
| 2020-03-03 | 37058377.68 | 478.323194808 | 101.6449207061 | 17.9444408974 |
| 2020-03-04 | 37058290.87 | 565.126247700 | 120.0910760946 | 21.2009336258 |
| 2020-03-05 | 37058188.32 | 667.681483047 | 141.8847735565 | 25.0484029664 |
| 2020-03-06 | 37058067.15 | 788.847342934 | 167.6335130124 | 29.5940972336 |
| 2020-03-07 | 37057924.00 | 932.000927129 | 198.0550412880 | 34.9647277812 |
| 2020-03-08 | 37057754.87 | 1101.132084717 | 233.9973593270 | 41.3100010845 |
| 2020-03-09 | 37057555.05 | 1300.954568264 | 276.4623602405 | 48.8067918125 |
| 2020-03-10 | 37057318.96 | 1537.037341885 | 326.6337571054 | 57.6640732145 |
| 2020-03-11 | 37057040.04 | 1815.959693561 | 385.9100789996 | 68.1287422551 |
| 2020-03-12 | 37056710.51 | 2145.494461683 | 455.9436550380 | 80.4925018737 |
| 2020-03-13 | 37056321.18 | 2534.824464056 | 538.6866730933 | 95.0999922122 |
| 2020-03-14 | 37055861.20 | 2994.798135497 | 636.4455970861 | 112.3583974684 |
| 2020-03-15 | 37055317.77 | 3538.231462480 | 751.9454597314 | 132.7487961671 |
| 2020-03-16 | 37054675.74 | 4180.264579013 | 888.4058229004 | 156.8395712371 |
| 2020-03-17 | 37053917.22 | 4938.782890948 | 1049.6305229974 | 185.3022537007 |
| 2020-03-18 | 37053021.09 | 5834.914365787 | 1240.1137030045 | 218.9302416202 |
| 2020-03-19 | 37051962.38 | 6893.616707965 | 1465.1650868421 | 258.6609160908 |
| 2020-03-20 | 37050711.63 | 8144.370588964 | 1731.0579880700 | 305.6017707639 |
| 2020-03-21 | 37049234.00 | 9621.997979614 | 2045.2041786768 | 361.0612832643 |

| Date | Susceptibles | Infected | Recovered | Deaths |
|---|---|---|---|---|
| 2020-03-22 | 37047488.37 | 11367.628009911 | 2416.3604924297 | 426.5853890394 |
| 2020-03-23 | 37045426.16 | 13429.836741674 | 2854.8729218578 | 504.0005743531 |
| 2020-03-24 | 37042990.01 | 15865.991874772 | 3372.9650130813 | 595.4647896410 |
| 2020-03-25 | 37040112.16 | 18743.838824034 | 3985.0785974974 | 703.5276024384 |
| 2020-03-26 | 37036712.63 | 22143.370919417 | 4708.2763582312 | 831.2012666460 |
| 2020-03-27 | 37032696.97 | 26159.033826800 | 5562.7174529005 | 982.0446891911 |
| 2020-03-28 | 37027953.68 | 30902.322802049 | 6572.2194506928 | 1160.2626346563 |
| 2020-03-29 | 37022351.16 | 36504.841225675 | 7764.9222477662 | 1370.8229332094 |
| 2020-03-30 | 37015734.10 | 43121.900171058 | 9174.0724676347 | 1619.5949590064 |
| 2020-03-31 | 37007919.25 | 50936.751680752 | 10838.9502117198 | 1913.5132391590 |
| 2020-04-01 | 36998690.44 | 60165.563087669 | 12805.9639932658 | 2260.7707538698 |
| 2020-04-02 | 36987792.74 | 71063.256203901 | 15129.9443759323 | 2671.0473159828 |
| 2020-04-03 | 36974925.65 | 83930.353519721 | 17875.6723772754 | 3155.7793960341 |
| 2020-04-04 | 36959735.01 | 99120.993600624 | 21119.6852414071 | 3728.4789141854 |
| 2020-04-05 | 36941803.70 | 117052.299349409 | 24952.4099167952 | 4405.1098853727 |
| 2020-04-06 | 36920640.69 | 138215.305147413 | 29480.6837099574 | 5204.5334166648 |
| 2020-04-07 | 36895668.33 | 163187.671137634 | 34830.7323783407 | 6149.0334611458 |
| 2020-04-08 | 36866207.57 | 192648.433523761 | 41151.6886768073 | 7264.9379837242 |
| 2020-04-09 | 36831460.94 | 227395.056389913 | 48619.7494373085 | 8583.3528538849 |
| 2020-04-10 | 36790492.94 | 268363.059743140 | 57443.0870604568 | 10141.0289226621 |
| 2020-04-11 | 36742207.50 | 316648.495285019 | 67867.6523269611 | 11981.3864536308 |
| 2020-04-12 | 36685322.48 | 373533.518864231 | 80184.0302824534 | 14155.7254639565 |
| 2020-04-13 | 36618340.74 | 440515.257066822 | 94735.5403036865 | 16724.6557137953 |
| 2020-04-14 | 36539517.93 | 519338.072057246 | 111927.8061356733 | 19759.7862050375 |
| 2020-04-15 | 36446826.82 | 612029.176593508 | 132240.0626648177 | 23345.7212842193 |
| 2020-04-16 | 36337918.68 | 720937.318295927 | 156238.5146046508 | 27582.4189910266 |
| 2020-04-17 | 36210082.09 | 848773.911654693 | 184592.1194679081 | 32587.9773914204 |
| 2020-04-18 | 36060200.48 | 998655.515634053 | 218091.2347757316 | 38501.9265644983 |
| 2020-04-19 | 35884710.10 | 1174145.897841589 | 257669.6492954694 | 45489.1180073146 |
| 2020-04-20 | 35679560.94 | 1379295.056118555 | 304430.6124284378 | 53744.3198749281 |
| 2020-04-21 | 35440184.55 | 1618671.453753314 | 359677.5873175501 | 63497.6461481213 |
| 2020-04-22 | 35161473.65 | 1897382.350708482 | 424950.5849185396 | 75020.9710669898 |
| 2020-04-23 | 34837780.50 | 2221075.499819260 | 502069.0918480192 | 88635.5076329177 |
| 2020-04-24 | 34462942.30 | 2595913.702825055 | 593182.7886233302 | 104720.7614298913 |
| 2020-04-25 | 34030344.04 | 3028511.958009627 | 700831.4720665249 | 123725.1093531670 |
| 2020-04-26 | 33533030.52 | 3525825.482125952 | 828015.8522111481 | 146178.2981276497 |
| 2020-04-27 | 32963879.78 | 4094976.224296817 | 978281.1971775634 | 172706.2110125271 |
| 2020-04-28 | 32315849.74 | 4743006.262744494 | 1155816.1576200332 | 204048.3143144600 |
| 2020-04-29 | 31582306.51 | 5476549.495484716 | 1365569.5254797610 | 241078.2700313694 |
| 2020-04-30 | 30757436.84 | 6301419.157102739 | 1613388.1816973640 | 284828.2892048339 |
| 2020-05-01 | 29836737.42 | 7222118.586236479 | 1906180.0781810875 | 336517.9048314724 |
| 2020-05-02 | 28817559.54 | 8241296.461109583 | 2252106.7971576559 | 397587.9663790159 |
| 2020-05-03 | 27699671.45 | 9359184.547514625 | 2660811.0555029497 | 469740.8035051383 |
| 2020-05-04 | 26485782.41 | 10573073.588338261 | 3143685.4957421012 | 554987.6785438312 |
| 2020-05-05 | 25181958.53 | 11876897.471322054 | 3714190.2562755286 | 655704.8504986896 |




## 5 Model Parameters Learning