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.
- Since July 18th, 2020, people must wear a mask to cover their mouth and nose in all closed public areas (e.g. stores, banks, gaz station).

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:  2020-07-18
```

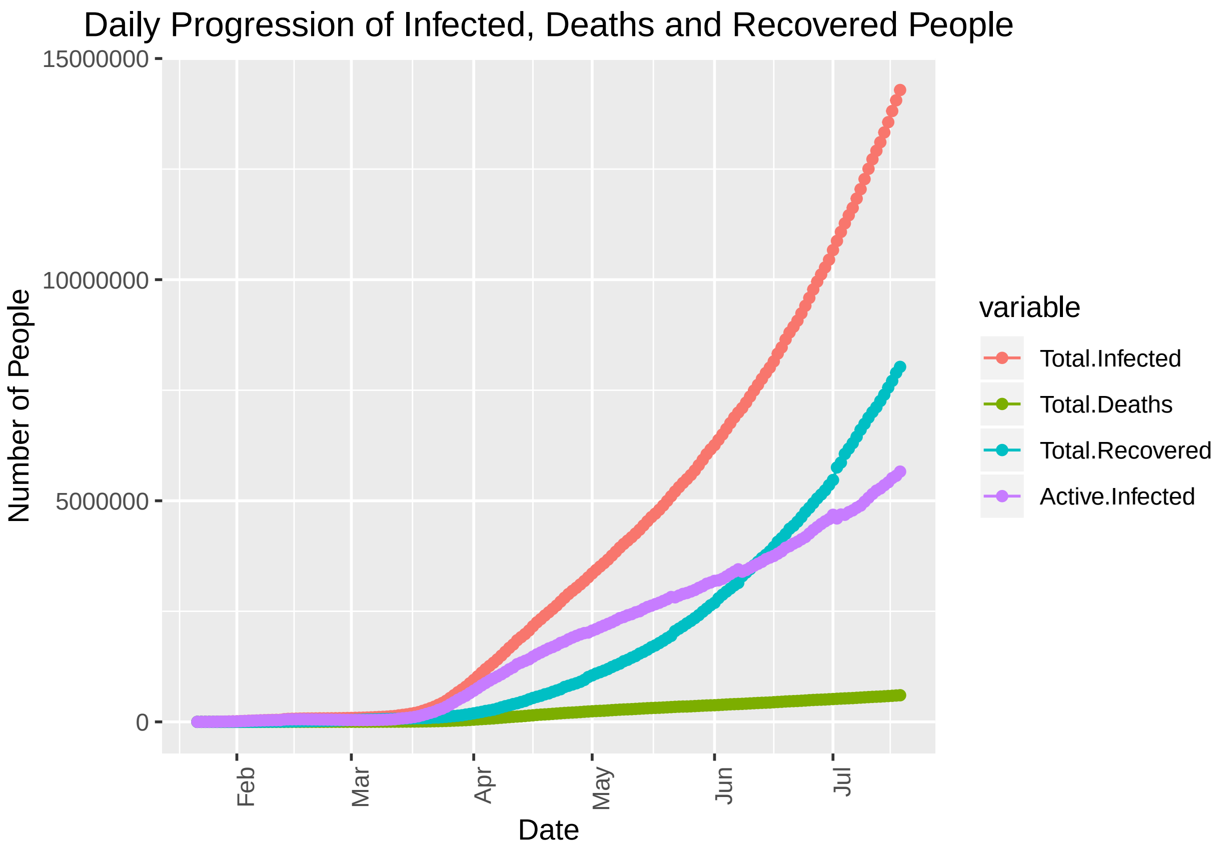
3 Dataset Exploration

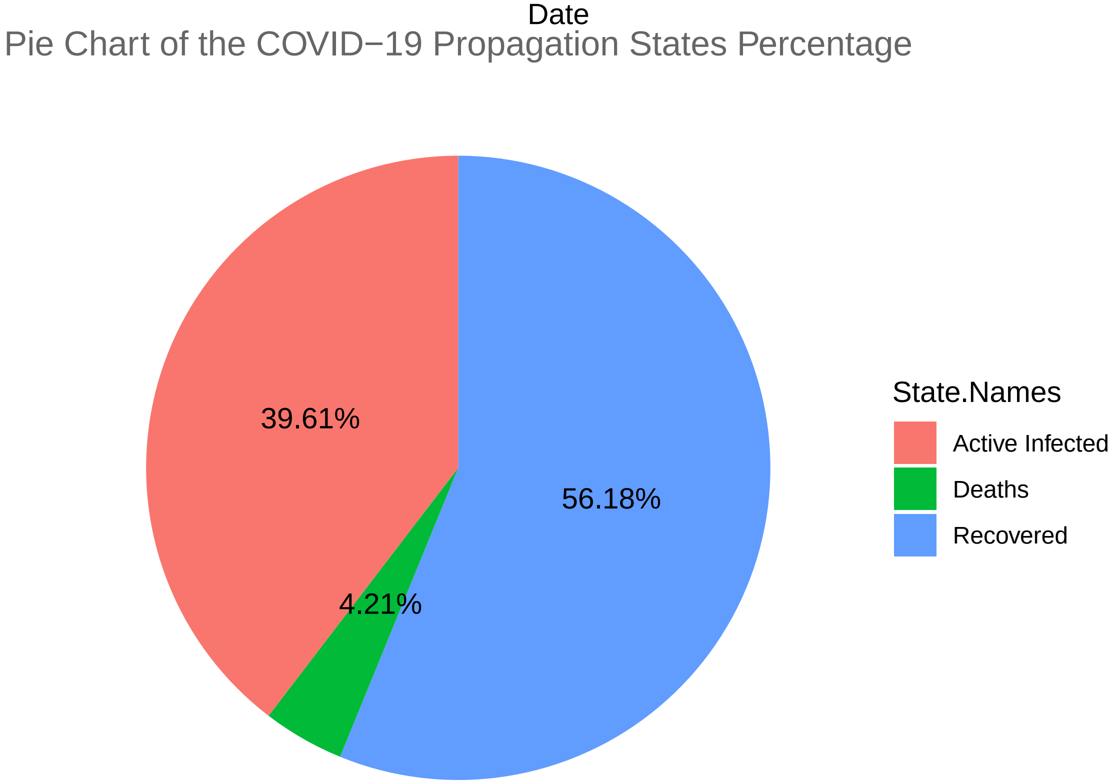
3.1 Worldwide Propagation

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
2020-07-05	11455621	534166	6179006	4742449
2020-07-06	11622965	537963	6302626	4782376
2020-07-07	11833815	544070	6447656	4842089
2020-07-08	12045624	549389	6605607	4890628
2020-07-09	12273853	554847	6740124	4978882
2020-07-10	12506434	560158	6879521	5066755
2020-07-11	12722764	565055	7005299	5152410
2020-07-12	12915440	569009	7116957	5229474
2020-07-13	13108223	572824	7257369	5278030
2020-07-14	13329678	578484	7399474	5351720
2020-07-15	13560803	583977	7559252	5417574
2020-07-16	13813333	589776	7711525	5512032
2020-07-17	14055299	596518	7894890	5563891
2020-07-18	14288689	602138	8027451	5659100



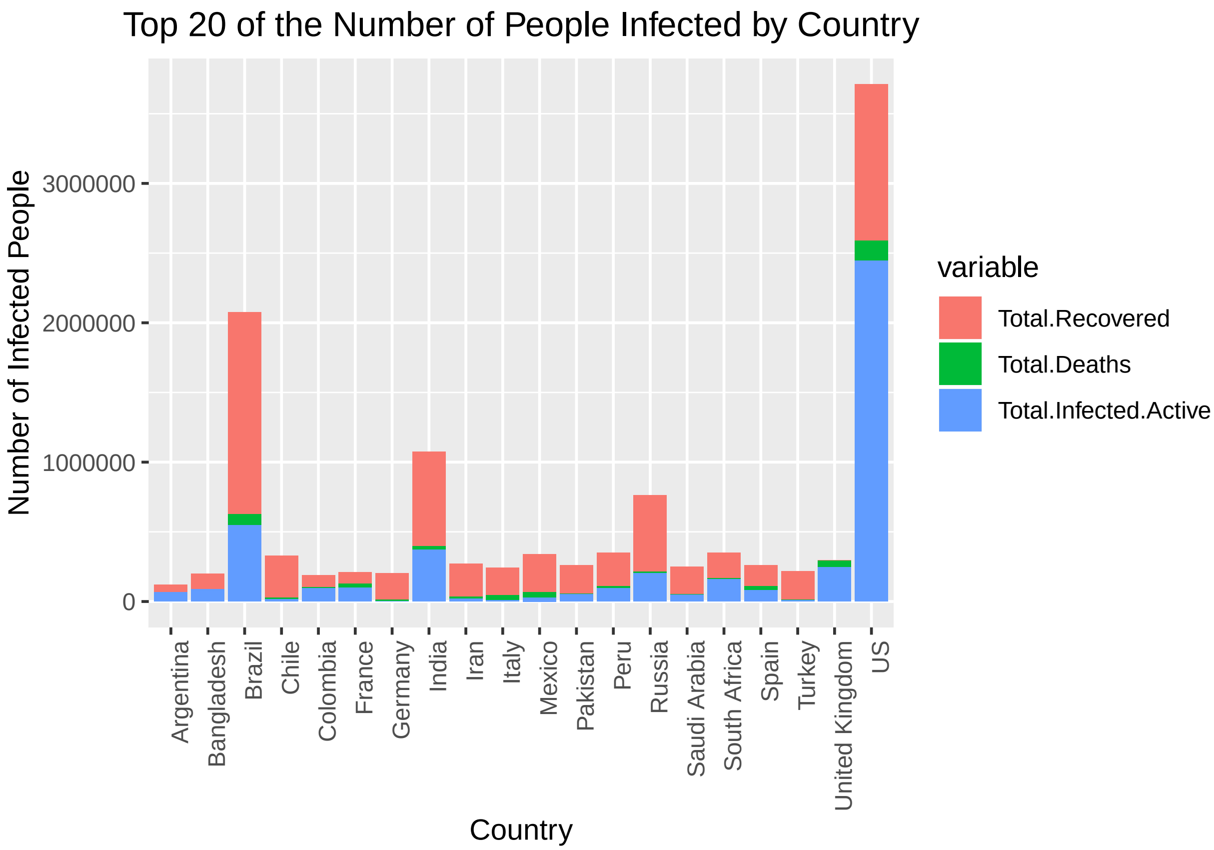


The distribution of the number of people infected, dead or recovered seems to be exponential. This would make sense because if we assume that all people will be infected one day or another, we expect that the curve will describe a sigmoid curve. The reasons behind this is explained in the **Propagation 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	3711413	140119	1122720	2448574
Brazil	2074860	78772	1447408	548680
India	1077781	26816	677423	373542
Russia	764215	12228	545909	206078
South Africa	350879	4948	182230	163701
Peru	349500	12998	238086	98416
Mexico	338913	38888	271239	28786
Chile	328846	8445	299449	20952
United Kingdom	295632	45358	1413	248861
Iran	271606	13979	235300	22327
Pakistan	263496	5568	204276	53652
Spain	260255	28420	150376	81459
Saudi Arabia	248416	2447	194218	51751
Italy	244216	35042	196806	12368
Turkey	218717	5475	201013	12229
France	211943	30155	79371	102417
Germany	202426	9091	187200	6135
Bangladesh	202066	2581	110098	89387
Colombia	190700	6516	85836	98348
Argentina	122524	2220	52607	67697
Canada	111875	8892	98436	4547
Qatar	106308	154	103023	3131
Iraq	90220	3691	58492	28037
Egypt	87172	4251	27868	55053
China	85314	4644	80018	652

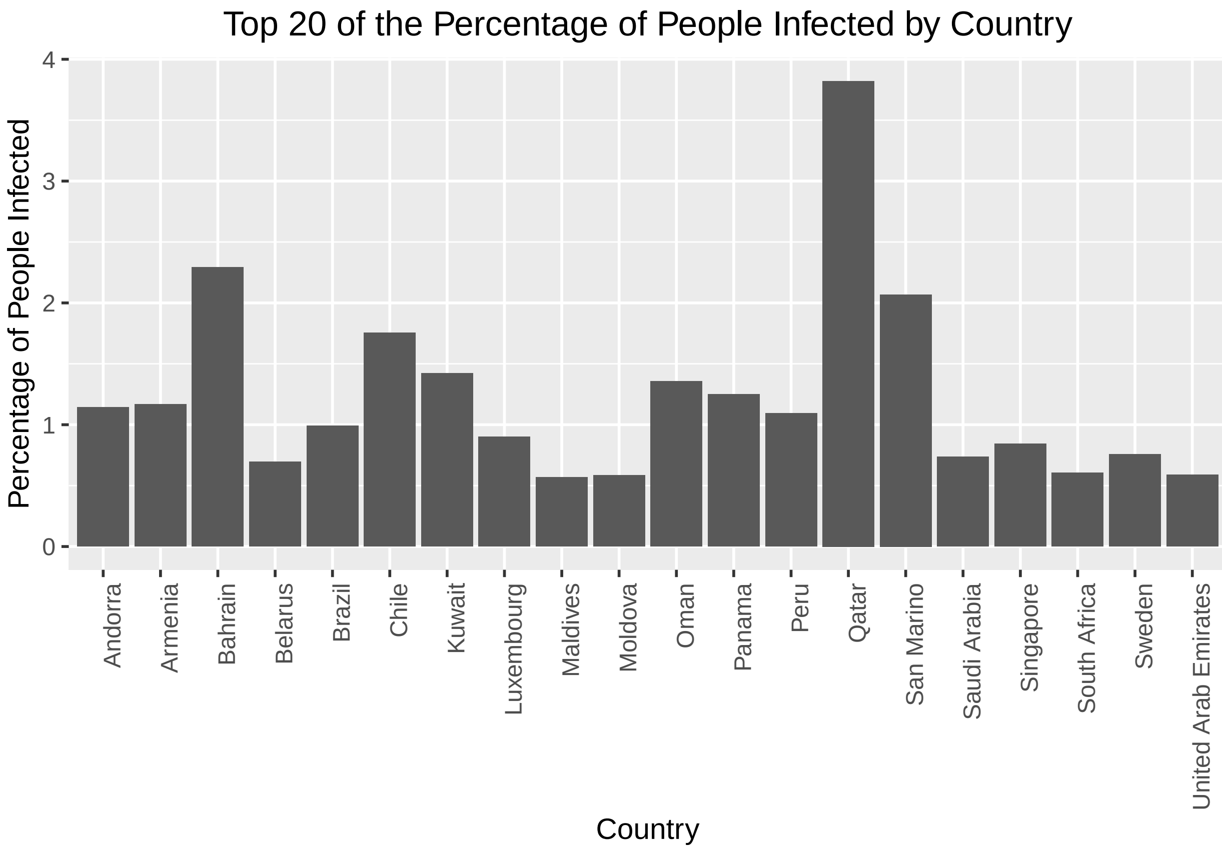


3.1.3 Countries Worst Ratio of Infected and Death People

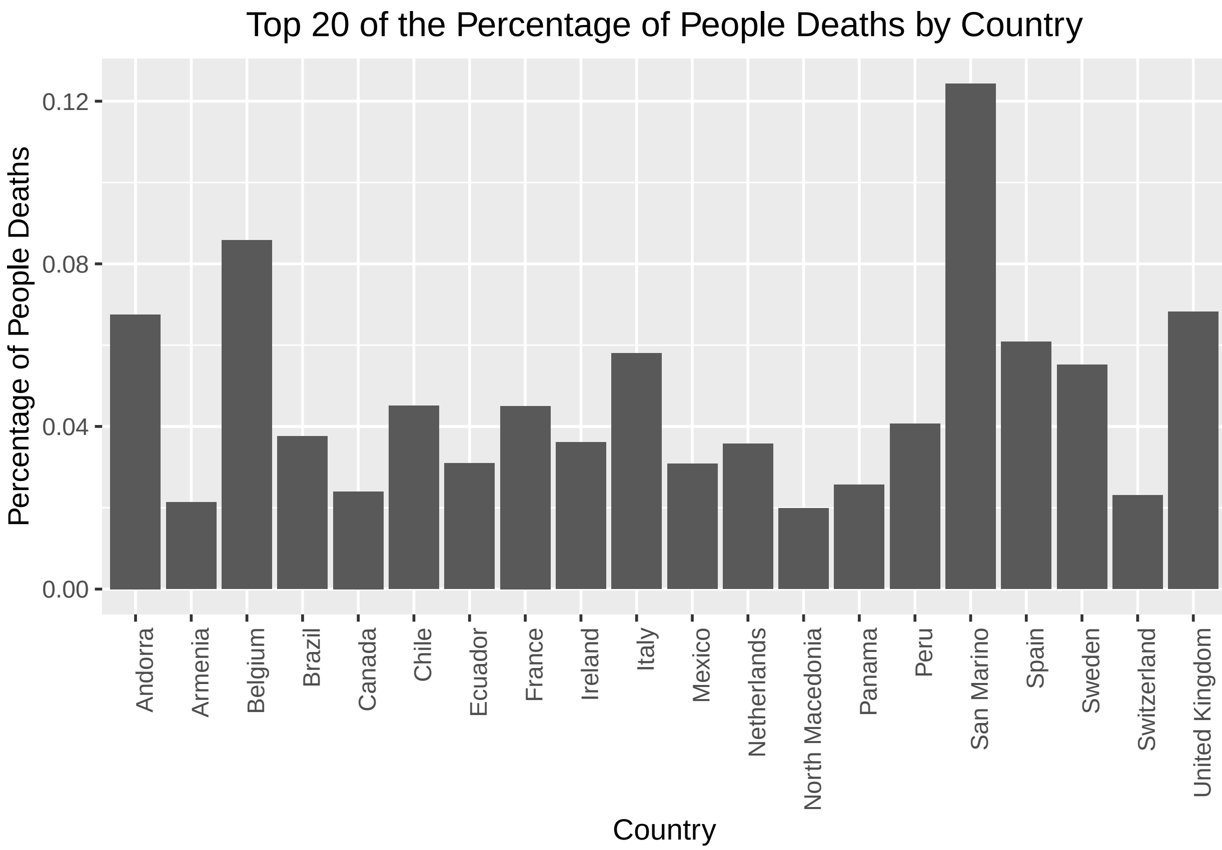
The objective is to know which countries have the worst ratio of deaths and infected people over their population.

Country.Region	Population	Percent.Infected
Qatar	2781677	3.8217234
Bahrain	1569439	2.2940681
San Marino	33785	2.0689655
Chile	18729160	1.7557968
Kuwait	4137309	1.4237274
Oman	4829483	1.3563357
Panama	4176873	1.2511992
Armenia	2951776	1.1675005
Andorra	77006	1.1427681
Peru	31989256	1.0925543
Brazil	209469333	0.9905316
Luxembourg	607728	0.9022128
Singapore	5638676	0.8451452
Sweden	10183175	0.7589087
Saudi Arabia	33699947	0.7371406
Belarus	9485386	0.6953117
South Africa	57779622	0.6072712
United Arab Emirates	9630959	0.5888406
Moldova	3545883	0.5864266
Maldives	515696	0.5681642

Country.Region	Population	Percent.Infected



Country.Region	Population	Percent.Deaths
San Marino	33785	0.1243155
Belgium	11422068	0.0857988
United Kingdom	66488991	0.0682188
Andorra	77006	0.0675272
Spain	46723749	0.0608256
Italy	60431283	0.0579865
Sweden	10183175	0.0551793
Chile	18729160	0.0450901
France	66987244	0.0450160
Peru	31989256	0.0406324
Brazil	209469333	0.0376055
Ireland	4853506	0.0361182
Netherlands	17231017	0.0357205
Ecuador	17084357	0.0309172
Mexico	126190788	0.0308168
Panama	4176873	0.0256412
Canada	37058856	0.0239943
Switzerland	8516543	0.0231197
Armenia	2951776	0.0213770
North Macedonia	2082958	0.0198756

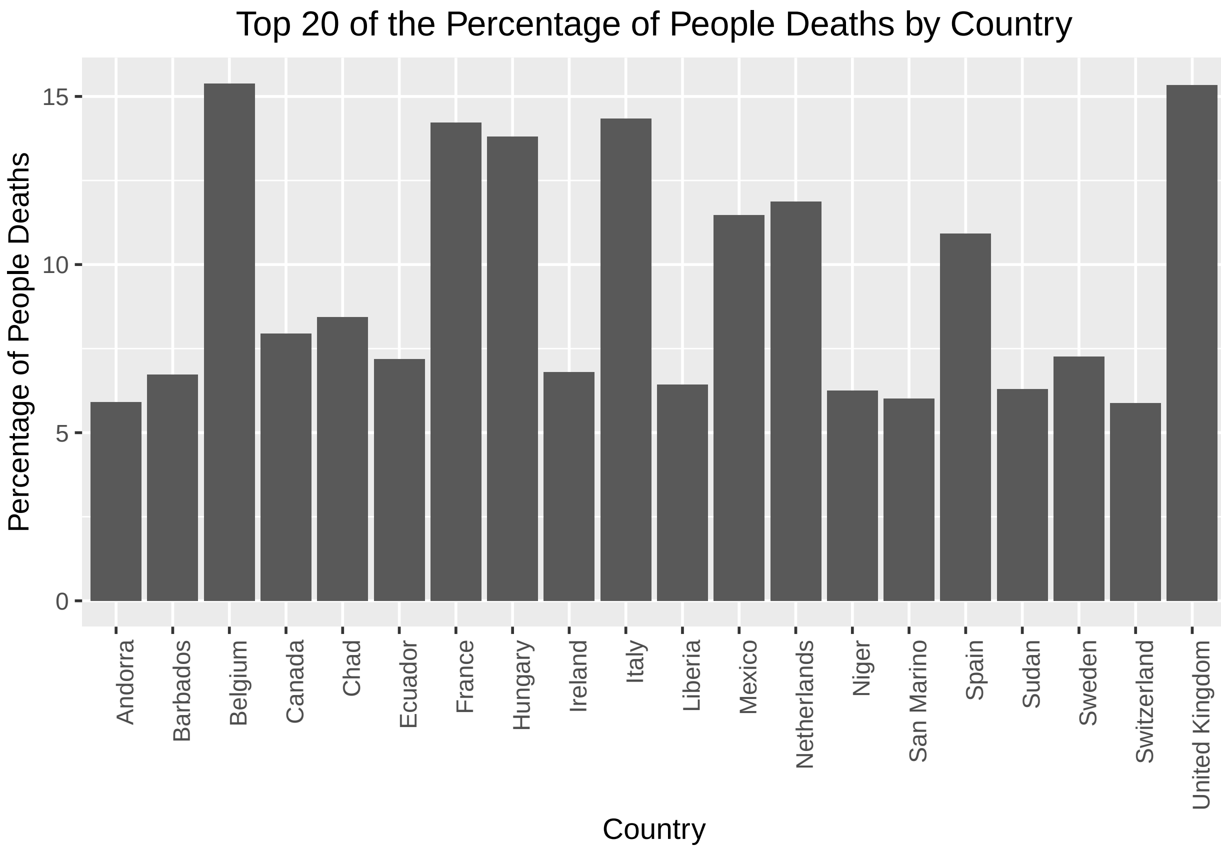


3.1.4 Countries Worst Ratio of Deaths Over Cumulative Infected People

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	63706	9800	15.383166
United Kingdom	295632	45358	15.342723
Italy	244216	35042	14.348773
France	211943	30155	14.227882
Hungary	4315	596	13.812283
Netherlands	51809	6155	11.880175
Mexico	338913	38888	11.474331
Spain	260255	28420	10.920059
Chad	889	75	8.436445
Canada	111875	8892	7.948156
Sweden	77281	5619	7.270869
Ecuador	73382	5282	7.197951
Ireland	25750	1753	6.807767
Barbados	104	7	6.730769
Liberia	1088	70	6.433823
Sudan	10682	673	6.300318
Niger	1104	69	6.250000
San Marino	699	42	6.008584
Andorra	880	52	5.909091
Switzerland	33492	1969	5.879016

Country.Region	Total.Infected	Total.Deaths	Percent.Deaths

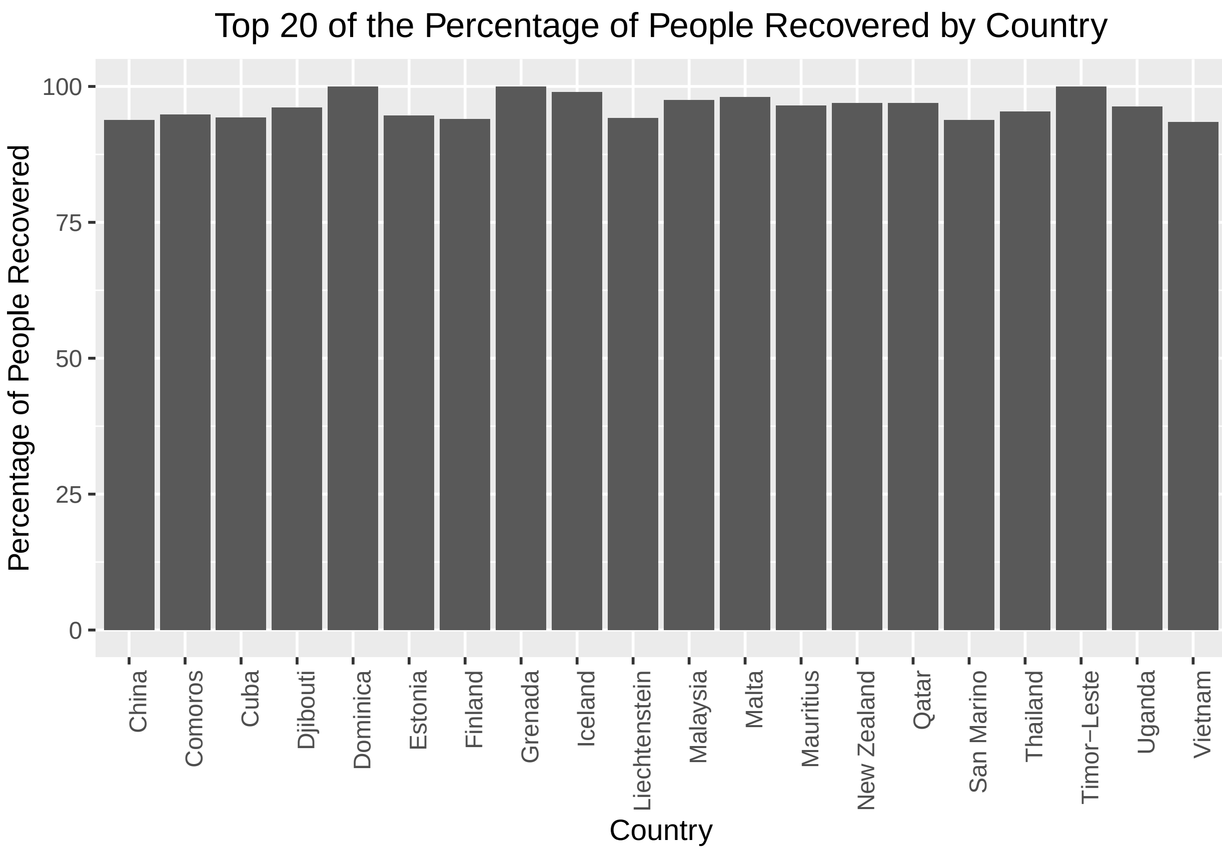


3.1.5 Countries Best Ratio of Recovers Over Cumulative Infected People

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
Iceland	1922	1902	98.95942
Malta	675	662	98.07407
Malaysia	8764	8546	97.51255
New Zealand	1553	1506	96.97360
Qatar	106308	103023	96.90992
Mauritius	343	331	96.50146
Uganda	1062	1023	96.32768
Djibouti	5003	4809	96.12233
Thailand	3246	3096	95.37893
Comoros	328	311	94.81707
Estonia	2021	1912	94.60663
Cuba	2445	2304	94.23313
Liechtenstein	86	81	94.18605

Country.Region	Total.Infected	Total.Recovered	Percent.Recovered
Finland	7318	6880	94.01476
San Marino	699	656	93.84835
China	85314	80018	93.79234
Vietnam	382	357	93.45550



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

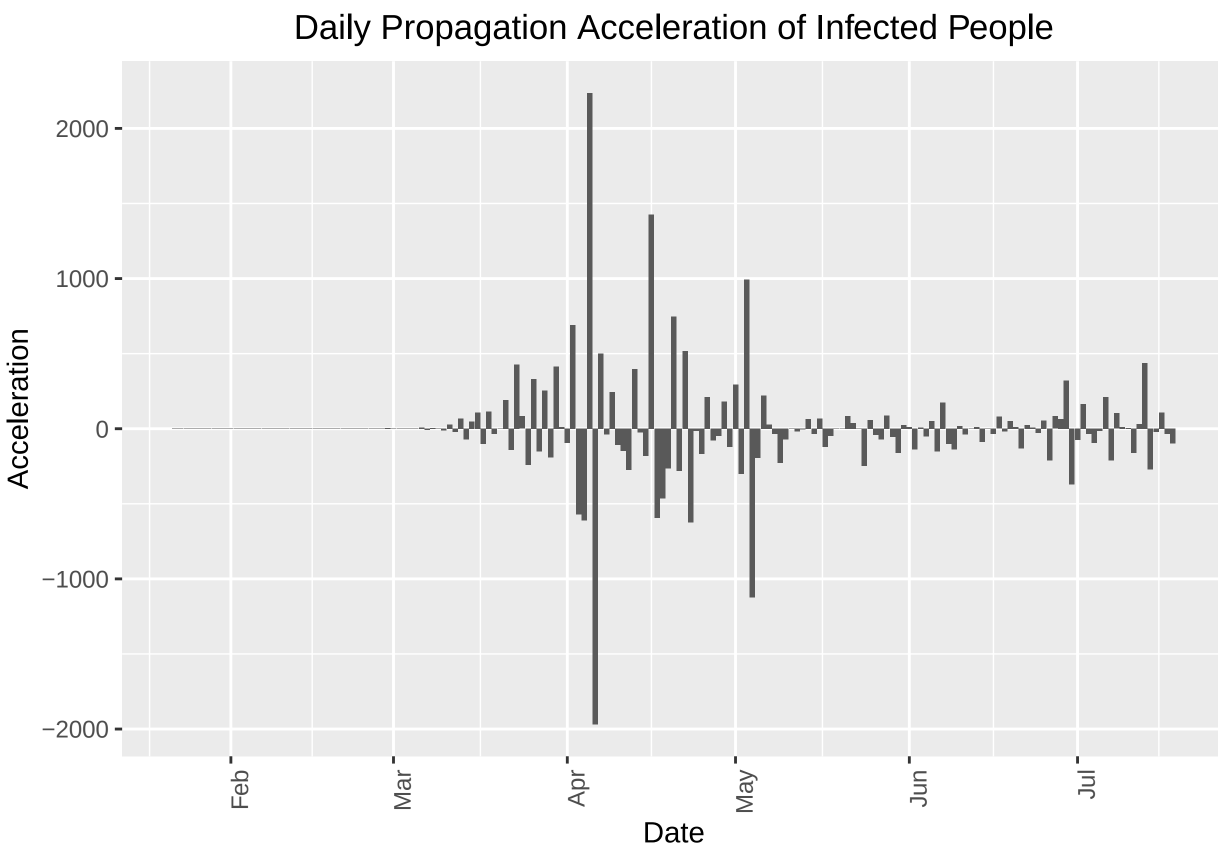
$$\Delta I(t) = \frac{I(t+1) - I(t)}{(t+1) - t} = I(t+1) - I(t)$$

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
2020-07-05	107394	209	-14
2020-07-06	107815	421	212
2020-07-07	108023	208	-213
2020-07-08	108334	311	103
2020-07-09	108656	322	11
2020-07-10	108984	328	6
2020-07-11	109150	166	-162
2020-07-12	109348	198	32
2020-07-13	109984	636	438
2020-07-14	110350	366	-270
2020-07-15	110693	343	-23
2020-07-16	111144	451	108
2020-07-17	111559	415	-36
2020-07-18	111875	316	-99



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. An aberrant value is a value that seems to be out of the "normality" according to the 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. It could be because they forgot to enter all the data received and compensate the day after by adding the current data plus the previous forgotten one.

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 μ 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 $\mu = 1.7653631$ and the variance is $\sigma^2 = 107728.2929509$. Therefore, we have $\Delta^2 I(t) \sim N(1.7653631, 107728.2929509)$ where the positive standard deviation is $\sigma = 328.2198851$.

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, if $k = 2$, non-aberrant propagation accelerations in the Canada have to be exclusively between 2 standard deviations -654.674407 and 658.2051333.

Actually, there are 3.9106145 % 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:

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

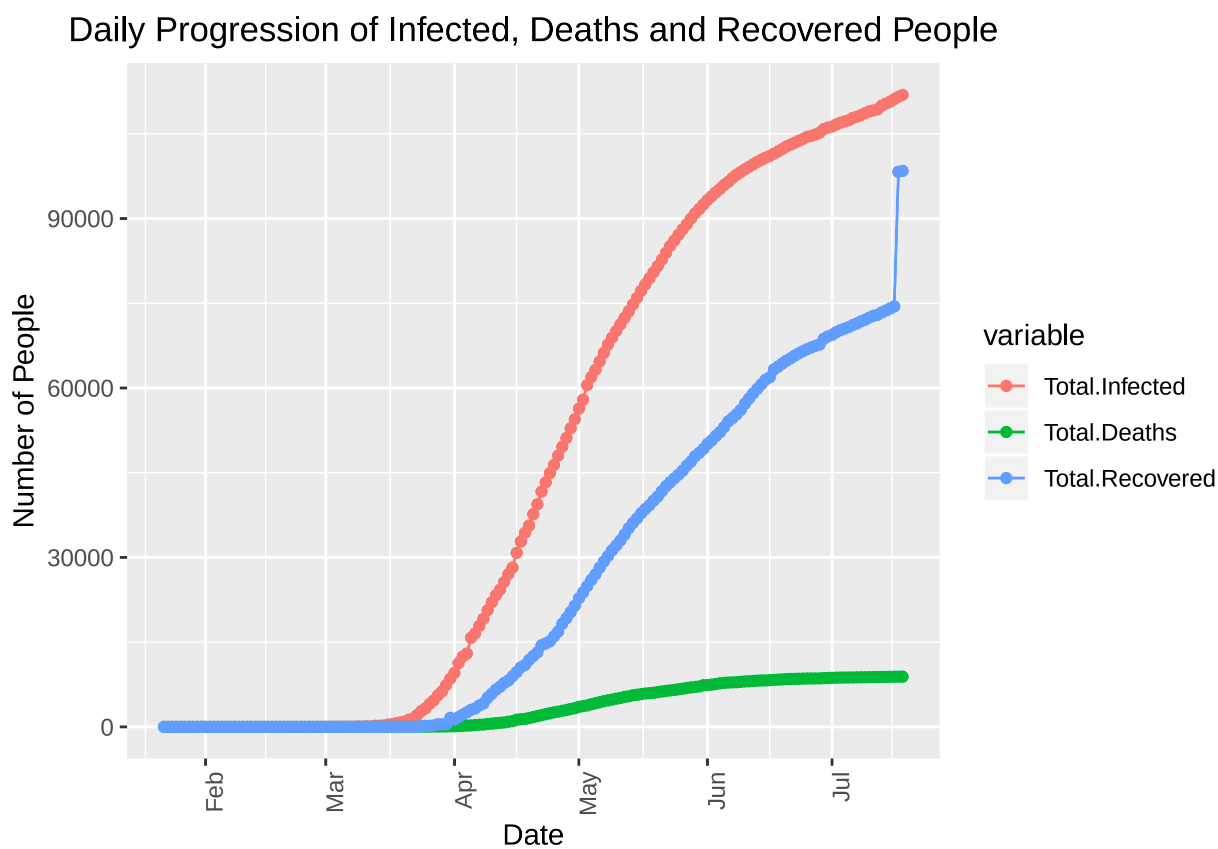
It follows that we have to provide the following parameters:

- the stabilization upper bound ϵ;
- 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.

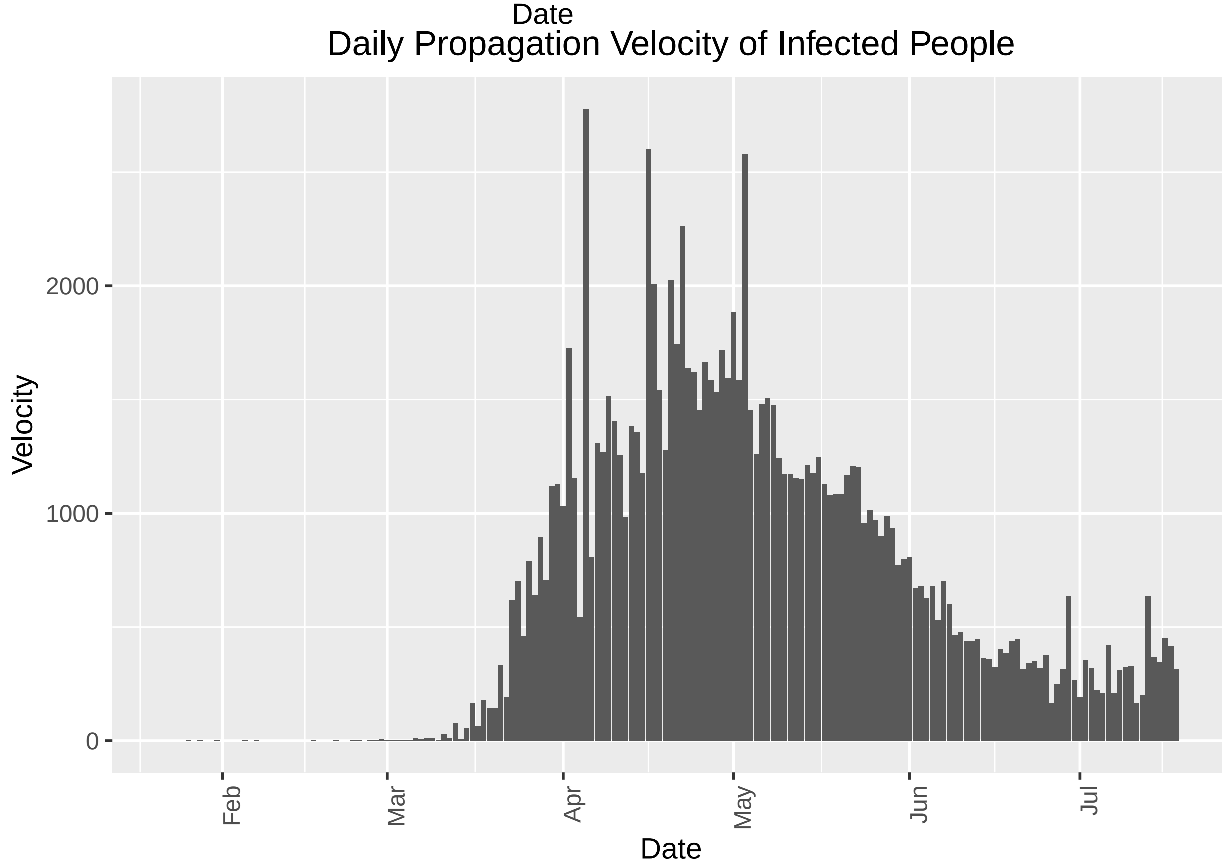
3.3 Canada Propagation Overview

The objective is to know the overall propagation in Canada.

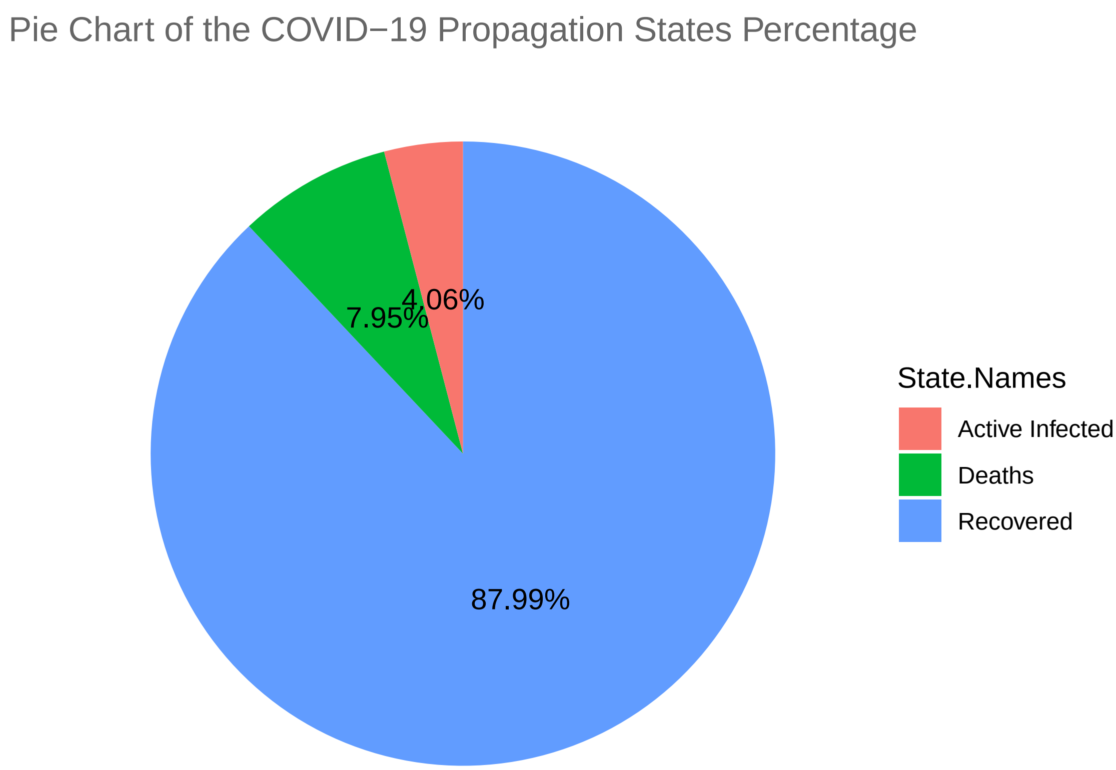
Date	Total.Infected	Total.Deaths	Total.Recovered	infected.delta	deaths.delta	recovered.delta
2020-07-05	107394	8739	70772	209	7	265
2020-07-06	107815	8748	71141	421	9	369
2020-07-07	108023	8765	71418	208	17	277
2020-07-08	108334	8786	71805	311	21	387
2020-07-09	108656	8797	72095	322	11	290
2020-07-10	108984	8811	72466	328	14	371
2020-07-11	109150	8818	72784	166	7	318
2020-07-12	109348	8829	72954	198	11	170
2020-07-13	109984	8836	73381	636	7	427
2020-07-14	110350	8845	73713	366	9	332
2020-07-15	110693	8857	74067	343	12	354
2020-07-16	111144	8875	74433	451	18	366
2020-07-17	111559	8884	98281	415	9	23848
2020-07-18	111875	8892	98436	316	8	155



Daily Progression of Infected, Deaths and Recovered People
Number of People
90000
60000
30000
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).

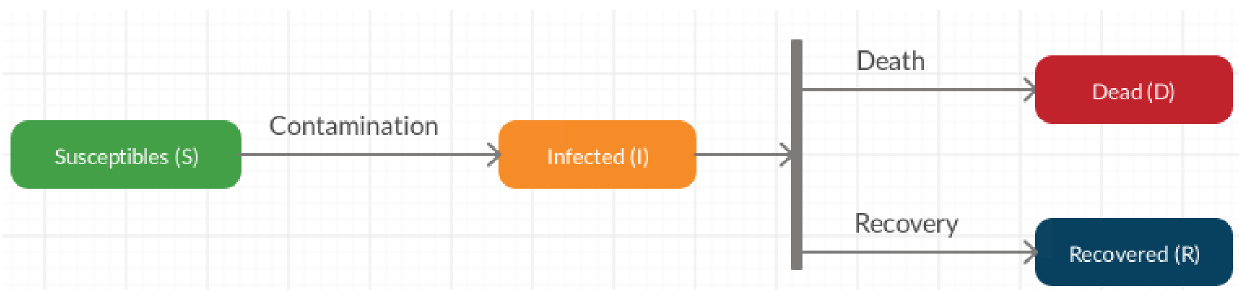
4.1 States and Transitions

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 2020-07-18.

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 β 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 γ that active infected people will recover (transition from *Infected* to *Recovered* state) or a probability of α 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

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

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:

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

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 $\alpha = 0.0028$, $\beta = 0.000000004$ and $\gamma = 0.045$.

Lets see the results of the first iteration:

$$\frac{\partial S(t)}{\partial t} = -0.000000004 \times 37058855 \times 1 = -0.14823542$$
$$\frac{\partial I_a(t)}{\partial t} = 0.000000004 \times 37058855 \times 1 - 0.045 \times 1 - 0.0028 \times 1 = 0.10043542$$
$$\frac{\partial R(t)}{\partial t} = 0.045 \times 1 = 0.045$$
$$\frac{\partial D(t)}{\partial t} = 0.0028 \times 1 = 0.0028$$

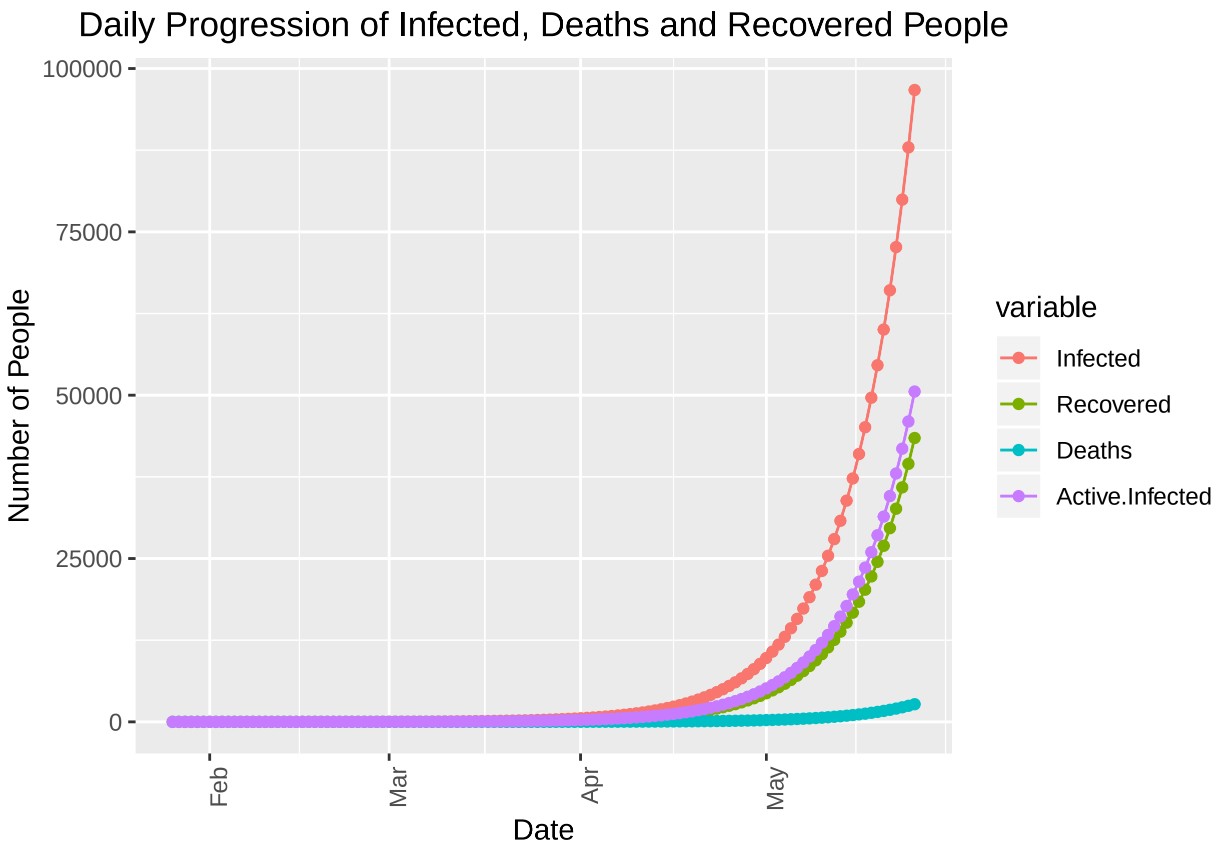
Therefore, we obtain $I(1) = 0.85176458$, $S(1) = 37058854.8517646$, $R(1) = 0.045$ and $D(1) = 0.0028$.

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	37058854	1	0	0	1
2020-02-01	37058853	1	0	0	1
2020-02-02	37058853	1	0	0	1
2020-02-03	37058853	2	0	0	2
2020-02-04	37058852	2	0	0	2
2020-02-05	37058852	2	0	0	2
2020-02-06	37058852	2	0	0	2
2020-02-07	37058851	3	0	0	3
2020-02-08	37058851	3	1	0	2
2020-02-09	37058850	3	1	0	2
2020-02-10	37058850	4	1	0	3
2020-02-11	37058849	4	1	0	3
2020-02-12	37058848	5	1	0	4
2020-02-13	37058848	5	2	0	3
2020-02-14	37058847	6	2	0	4
2020-02-15	37058846	6	2	0	4
2020-02-16	37058845	7	2	0	5
2020-02-17	37058844	8	3	0	5
2020-02-18	37058843	9	3	0	6
2020-02-19	37058841	9	4	0	5
2020-02-20	37058840	10	4	0	6
2020-02-21	37058838	12	4	0	8
2020-02-22	37058836	13	5	0	8
2020-02-23	37058834	14	6	0	8
2020-02-24	37058832	16	6	0	10
2020-02-25	37058830	17	7	0	10
2020-02-26	37058827	19	8	0	11
2020-02-27	37058824	21	9	0	12
2020-02-28	37058821	23	10	0	13
2020-02-29	37058818	25	11	0	14
2020-03-01	37058814	28	12	0	16
2020-03-02	37058810	31	13	0	18
2020-03-03	37058805	34	15	0	19
2020-03-04	37058800	37	16	1	20

Date	Susceptibles	Infected	Recovered	Deaths	Active.Infected
2020-03-05	37058794	41	18	1	22
2020-03-06	37058788	45	20	1	24
2020-03-07	37058781	50	22	1	27
2020-03-08	37058774	55	24	1	30
2020-03-09	37058766	61	27	1	33
2020-03-10	37058756	67	29	1	37
2020-03-11	37058746	74	32	2	40
2020-03-12	37058735	81	36	2	43
2020-03-13	37058723	89	39	2	48
2020-03-14	37058710	98	43	2	53
2020-03-15	37058695	108	48	3	57
2020-03-16	37058679	119	53	3	63
2020-03-17	37058662	131	58	3	70
2020-03-18	37058642	144	64	4	76
2020-03-19	37058620	159	71	4	84
2020-03-20	37058597	175	78	4	93
2020-03-21	37058571	193	86	5	102
2020-03-22	37058542	212	94	5	113
2020-03-23	37058511	233	104	6	123
2020-03-24	37058476	257	114	7	136
2020-03-25	37058438	283	126	7	150
2020-03-26	37058396	311	139	8	164
2020-03-27	37058350	343	153	9	181
2020-03-28	37058299	377	168	10	199
2020-03-29	37058243	415	185	11	219
2020-03-30	37058181	457	204	12	241
2020-03-31	37058113	503	224	13	266
2020-04-01	37058039	553	247	15	291
2020-04-02	37057957	609	272	16	321
2020-04-03	37057866	670	299	18	353
2020-04-04	37057767	737	330	20	387
2020-04-05	37057658	811	363	22	426
2020-04-06	37057537	893	399	24	470
2020-04-07	37057405	983	440	27	516
2020-04-08	37057259	1081	484	30	567
2020-04-09	37057099	1190	532	33	625
2020-04-10	37056922	1310	586	36	688
2020-04-11	37056728	1441	645	40	756
2020-04-12	37056514	1586	710	44	832
2020-04-13	37056279	1745	781	48	916
2020-04-14	37056020	1921	860	53	1008
2020-04-15	37055736	2114	946	58	1110
2020-04-16	37055422	2326	1041	64	1221
2020-04-17	37055078	2559	1146	71	1342
2020-04-18	37054698	2817	1261	78	1478
2020-04-19	37054281	3099	1388	86	1625
2020-04-20	37053821	3411	1528	95	1788
2020-04-21	37053316	3753	1681	104	1968
2020-04-22	37052759	4130	1850	115	2165
2020-04-23	37052147	4545	2036	126	2383
2020-04-24	37051473	5001	2240	139	2622
2020-04-25	37050732	5504	2466	153	2885

Date	Susceptibles	Infected	Recovered	Deaths	Active.Infected
2020-04-26	37049916	6056	2713	168	3175
2020-04-27	37049019	6664	2986	185	3493
2020-04-28	37048031	7333	3286	204	3843
2020-04-29	37046944	8070	3616	225	4229
2020-04-30	37045748	8880	3979	247	4654
2020-05-01	37044432	9771	4378	272	5121
2020-05-02	37042984	10752	4818	299	5635
2020-05-03	37041391	11831	5302	329	6200
2020-05-04	37039638	13019	5834	363	6822
2020-05-05	37037709	14325	6420	399	7506
2020-05-06	37035587	15763	7065	439	8259
2020-05-07	37033251	17345	7774	483	9088
2020-05-08	37030682	19085	8555	532	9998
2020-05-09	37027855	21000	9414	585	11001
2020-05-10	37024745	23106	10359	644	12103
2020-05-11	37021323	25424	11399	709	13316
2020-05-12	37017558	27974	12543	780	14651
2020-05-13	37013415	30779	13802	858	16119
2020-05-14	37008858	33864	15187	944	17733
2020-05-15	37003845	37259	16711	1039	19509
2020-05-16	36998330	40993	18387	1144	21462
2020-05-17	36992263	45100	20232	1258	23610
2020-05-18	36985590	49618	22261	1385	25972
2020-05-19	36978249	54587	24494	1524	28569
2020-05-20	36970175	60052	26951	1676	31425
2020-05-21	36961295	66062	29653	1845	34564
2020-05-22	36951528	72671	32626	2030	38015
2020-05-23	36940786	79939	35896	2233	41810
2020-05-24	36928974	87930	39493	2457	45980
2020-05-25	36915986	96715	43450	2703	50562



4.5 Model Representation With Markov Chain

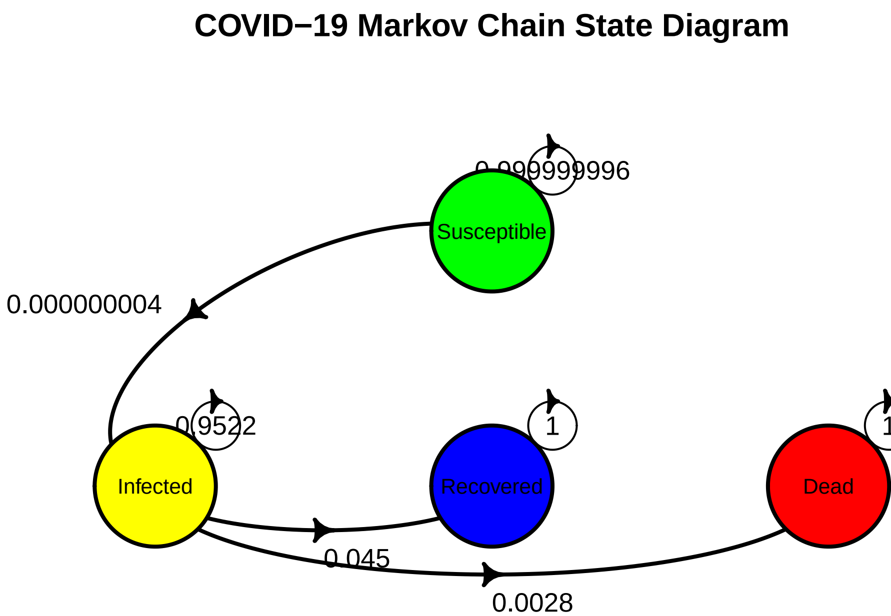
Let $X = (X_t)_{t\geq 0}$ be a sequence of random variables where each random variable is representing a person X at day $t \geq 0$. Let $\mathbb{E} = \{S, I, R, D\}$ the COVID-19 state space. At any day t, a person X_t has to be in a state of $\mathbb{E}$. If $i, j \in \mathbb{E}$, then $\mathbb{P}(X_t \in j | X_{t-1} \in i) = p_{i,j}$. 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 $\alpha = 0.0028$, we obtain

$$\frac{\gamma}{\alpha+\gamma} = \frac{0.045}{0.0478} = 0.9414225941$$

This means that 94.1422594142 % of the population will recover and 5.8577405858 % 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 α is used instead of γ. 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 $D(t)$:

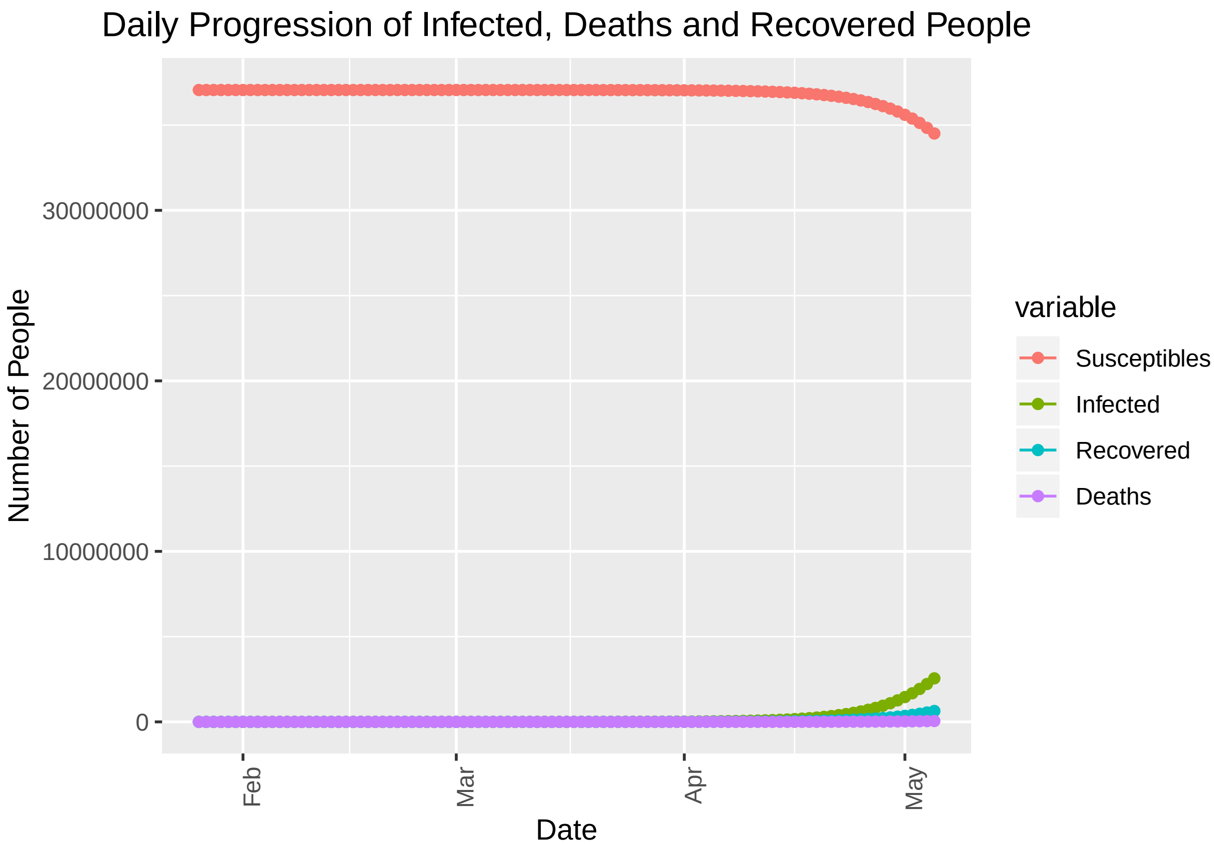
$$D(t) = \frac{-\alpha Ne^{N\beta t}{}_2F_1\left(1; \frac{\alpha}{N\beta}+1; \frac{\alpha}{N\beta}+2; \frac{e^{N\beta t}}{\delta}\right)}{\delta(N\beta+\alpha)} + D_0e^{-\alpha t}.$$

4.6.3 Simulation

Date	Susceptibles	Infected	Recovered	Deaths
2020-01-26	37058855.00	1.000000000	0.0000000000	0.0000000000
2020-01-27	37058854.84	1.159785901	0.2700869545	0.0215009204
2020-01-28	37058854.65	1.345103334	0.3132430431	0.0249364644
2020-01-29	37058854.44	1.560031880	0.3632948664	0.0289209599

Date	Susceptibles	Infected	Recovered	Deaths
2020-01-30	37058854.19	1.809302974	0.4213442657	0.0335421217
2020-01-31	37058853.90	2.098404073	0.4886691408	0.0389016800
2020-02-01	37058853.57	2.433699446	0.5667515820	0.0451176202
2020-02-02	37058853.18	2.822570286	0.6573104968	0.0523267800
2020-02-03	37058852.73	3.273577196	0.7623394498	0.0606878619
2020-02-04	37058852.20	3.796648639	0.8841505493	0.0703849269
2020-02-05	37058851.60	4.403299508	1.0254253455	0.0816314462
2020-02-06	37058850.89	5.106884611	1.1892738630	0.0946750008
2020-02-07	37058850.08	5.922892664	1.3793030643	0.1098027315
2020-02-08	37058849.13	6.869287256	1.5996962536	0.1273476604
2020-02-09	37058848.03	7.966902306	1.8553051683	0.1476960216
2020-02-10	37058846.76	9.239900689	2.1517567849	0.1712957642
2020-02-11	37058845.28	10.716306162	2.4955771915	0.1986664131
2020-02-12	37058843.57	12.428620273	2.8943352532	0.2304105058
2020-02-13	37058841.59	14.414537847	3.3568092330	0.2672268571
2020-02-14	37058839.28	16.717776793	3.8931800364	0.3099259425
2020-02-15	37058836.61	19.389040500	4.5152553344	0.3594477399
2020-02-16	37058833.51	22.487134017	5.2367294973	0.4168824226
2020-02-17	37058829.92	26.080258563	6.0734850628	0.4834943580
2020-02-18	37058825.75	30.247512896	7.0439423741	0.5607499419
2020-02-19	37058820.92	35.080634563	8.1694650857	0.6503498792
2020-02-20	37058815.31	40.686019374	9.4748304632	0.7542666237
2020-02-21	37058808.81	47.187063549	10.9887748298	0.8747877993
2020-02-22	37058801.27	54.726880100	12.7446261681	1.0145665600
2020-02-23	37058792.53	63.471449224	14.7810378027	1.1766799967
2020-02-24	37058782.39	73.613272076	17.1428393146	1.3646968756
2020-02-25	37058770.62	85.375608325	19.8820234201	1.5827562019
2020-02-26	37058756.98	99.017390773	23.0588905387	1.8356583350
2020-02-27	37058741.16	114.838925206	26.7433762470	2.1289706646
2020-02-28	37058722.81	133.188500927	31.0165908411	2.4691501703
2020-02-29	37058701.53	154.470057465	35.9726048992	2.8636855665
2020-03-01	37058676.85	179.152076175	41.7205201521	3.3212621583
2020-03-02	37058648.22	207.777892420	48.3868712493	3.8519530402
2020-03-03	37058615.02	240.977655252	56.1184112940	4.4674408453
2020-03-04	37058576.52	279.482197754	65.0853424670	5.1812749267
2020-03-05	37058531.86	324.139123239	75.4850628584	6.0091696334
2020-03-06	37058480.07	375.931461210	87.5465119912	6.9693502455
2020-03-07	37058420.00	435.999303497	101.5352106973	8.0829541864
2020-03-08	37058350.34	505.664896481	117.7591062951	9.3744963415
2020-03-09	37058269.54	586.461741285	136.5753517445	10.8724087295
2020-03-10	37058175.83	680.168341855	158.3981680141	12.6096664051
2020-03-11	37058067.15	788.847342934	183.7079627455	14.6245133717
2020-03-12	37057941.11	914.890918274	213.0619059502	16.9613044857
2020-03-13	37057794.93	1061.073406600	247.1061955546	19.6714818842
2020-03-14	37057625.39	1230.612351786	286.5902828060	22.8147074328
2020-03-15	37057428.76	1427.239287995	332.3833706982	26.4601761223
2020-03-16	37057200.72	1655.281823953	385.4935486123	30.6881393280
2020-03-17	37056936.24	1919.758827916	447.0899844043	35.5916714636
2020-03-18	37056629.51	2226.490801367	518.5286624748	41.2787189225
2020-03-19	37056273.77	2582.227861346	601.3822344202	47.8744764101
2020-03-20	37055861.20	2994.798135497	697.4746393963	55.5241429814
2020-03-21	37055382.72	3473.279818748	808.9212563287	64.3961184538

Date	Susceptibles	Infected	Recovered	Deaths
2020-03-22	37054827.80	4028.200655205	938.1754718807	74.6857105621
2020-03-23	37054184.23	4671.769204352	1088.0826893253	86.6194344643
2020-03-24	37053437.86	5418.142939135	1261.9429672746	100.4599992481
2020-03-25	37052572.26	6283.739019390	1463.5836671947	116.5120912108
2020-03-26	37051568.41	7287.594503340	1697.4437089700	135.1290812255
2020-03-27	37050404.22	8451.783821140	1968.6712893185	156.7208038504
2020-03-28	37049054.10	9801.902558586	2283.2372142337	181.7625794298
2020-03-29	37047488.37	11367.628009911	2648.0663403518	210.8056777995
2020-03-30	37045672.63	13183.368582782	3071.1900187988	244.4894539454
2020-03-31	37043566.98	15289.016006202	3561.9228974136	283.5554227690
2020-04-01	37041125.18	17730.816436345	4131.0679734761	328.8635828016
2020-04-02	37038293.62	20562.379013193	4791.1544109705	381.4113482189
2020-04-03	37035010.16	23845.843232419	5556.7133576953	442.3555059239
2020-04-04	37031202.77	27653.229705335	6444.5978340604	513.0376810627
2020-04-05	37026788.00	32068.002530682	7474.3537356050	595.0138715702
2020-04-06	37021669.12	37186.875642116	8668.6501164890	690.0887019206
2020-04-07	37015734.10	43121.900171058	10053.7782262218	800.3551501444
2020-04-08	37008853.12	50002.875118793	11660.2302854266	928.2406226617
2020-04-09	37000875.87	57980.129500140	13523.3707418144	1076.5603912262
2020-04-10	36991628.27	67227.730626091	15684.2147834022	1248.5795683384
2020-04-11	36980908.82	77947.180335534	18190.3312471701	1448.0849854568
2020-04-12	36968484.33	90371.668735041	21096.8897997964	1679.4685563340
2020-04-13	36954085.04	104770.963280611	24467.8754431145	1947.8239606392
2020-04-14	36937398.98	121457.019686809	28377.4970804248	2259.0587762606
2020-04-15	36918065.59	140790.409929679	32911.8211518497	2620.0245287698
2020-04-16	36895668.33	163187.671137634	38170.6663015935	3038.6675209567
2020-04-17	36869726.31	189129.686872331	44269.8007863270	3524.2037627994
2020-04-18	36839684.78	219171.218374488	51343.4909984597	4087.3218527768
2020-04-19	36804904.29	253951.706644161	59547.4572075126	4740.4182767560
2020-04-20	36764648.53	294207.465167169	69062.3015873018	5497.8703043255
2020-04-21	36718070.62	340785.375560594	80097.4839935516	6376.3524901161
2020-04-22	36664197.82	394658.181558629	92895.9330147325	7395.2037475713
2020-04-23	36601914.55	456941.446888936	107739.3938038355	8576.8530759342
2020-04-24	36529943.81	528912.194931714	124954.6304182884	9947.3133124048
2020-04-25	36446826.82	612029.176593508	144920.6192063721	11536.7537789340
2020-04-26	36350901.39	707954.610145536	168076.8915954033	13380.1644299039
2020-04-27	36240278.91	818577.093970305	194933.2099398784	15518.1261212465
2020-04-28	36112820.80	946035.199930025	226080.7894337807	17997.7039577117
2020-04-29	35966115.00	1092741.000081394	262205.3131263042	20873.4833844365
2020-04-30	35797453.55	1261402.451166117	304102.0265536573	24208.7718313456
2020-05-01	35603812.85	1455043.149455206	352693.2442802730	28076.9923633891
2020-05-02	35381838.53	1677017.466665823	409048.6537385699	32563.3000164466
2020-05-03	35127837.51	1931018.486969319	474408.8633361310	37766.4564009276
2020-05-04	34837780.50	2221075.499819260	550212.7132185153	43801.0038406055
2020-05-05	34507318.90	2551537.098064071	638128.9499070459	50799.7869081944



5 Model Parameters Learning