COVID-19 Analysis

Gabriel Lapointe

March 29, 2020

Contents

1 Overview

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 Questions

The COVID-19 is not fully understood and many questions have to be answered:

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?

1.3 Objective

The objective of this research is to understand the propagation of the COVID-19 in countries and more precisely in Canada. It means to identify factors that appear to impact the transmission rate of 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

2.1 Datasets Source

The datasets are taken from CSV files prepared by the John Hopkins University Center for Systems Science and Engineering.

2.2 Datasets Overview

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
- 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

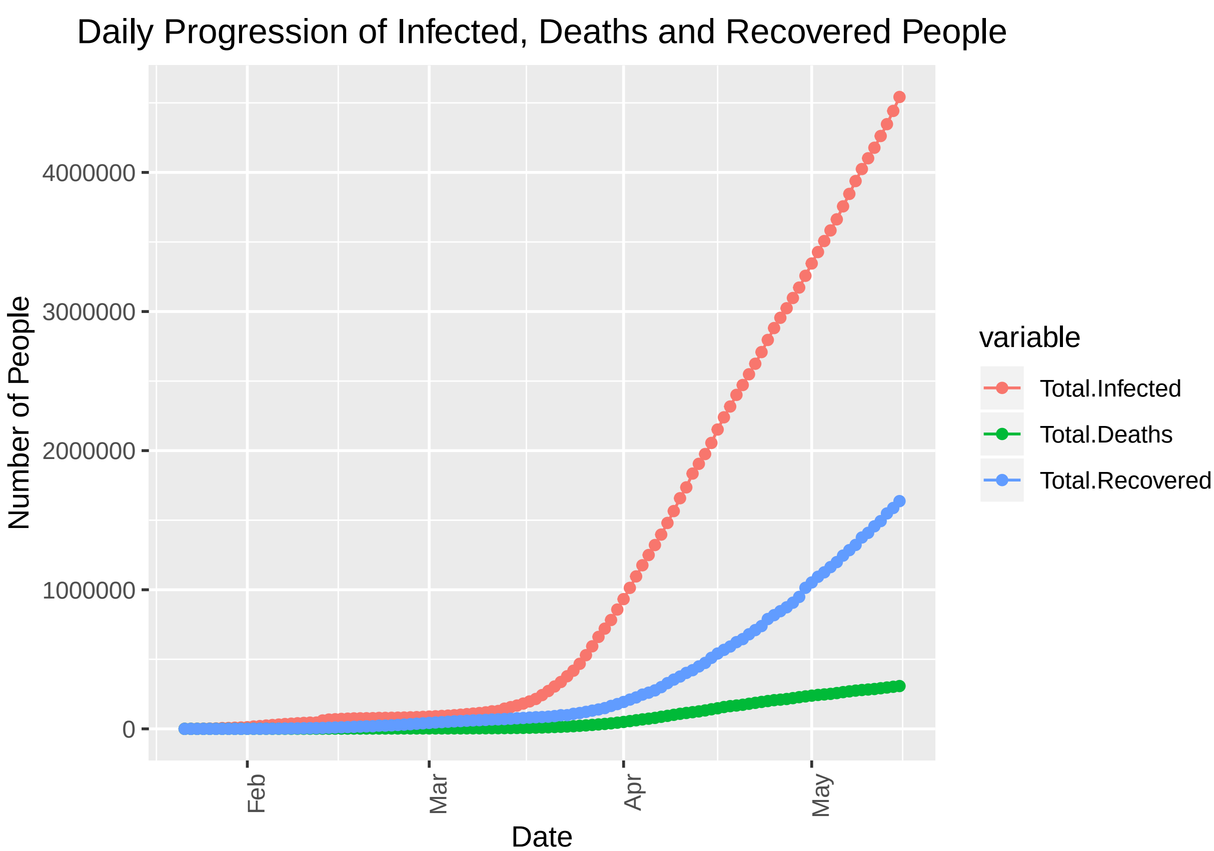
3 Dataset Exploration

3.1 World Overall

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
2020-05-02	3427584	243813	1093137
2020-05-03	3506729	247470	1125236
2020-05-04	3583055	251537	1162724
2020-05-05	3662691	257239	1198832
2020-05-06	3756069	263855	1245413
2020-05-07	3845718	269567	1284741
2020-05-08	3938064	274898	1322050
2020-05-09	4024009	279311	1375624
2020-05-10	4101699	282709	1408980
2020-05-11	4177502	286330	1456209
2020-05-12	4261747	291942	1493414
2020-05-13	4347018	297197	1548547
2020-05-14	4442163	302418	1587893
2020-05-15	4542347	307666	1637067



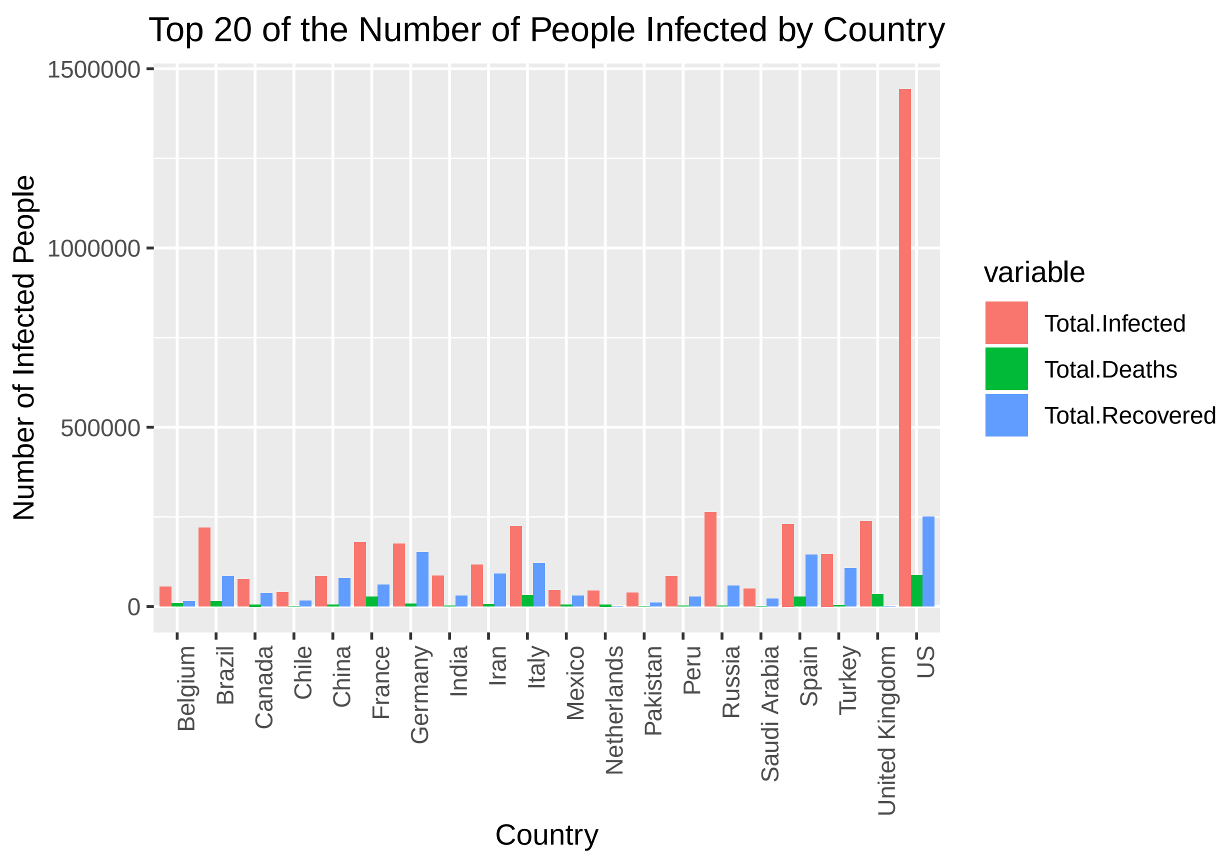

```
Number of countries: 188
```

```
Total infected: 4542347

Total deaths: 307666

Total recovered: 1637067

Total actives: 2597614

Percentage of deaths over infected: 6.773283

Percentage of recovered over infected: 36.04011
```

3.1.2 Countries Breakdown

The objective is to show in which countries there are the most infected people until today.

Country.Region	Total.Infected	Total.Deaths	Total.Recovered
US	1442824	87530	250747
Russia	262843	2418	58226
United Kingdom	238004	34078	1047
Spain	230183	27459	144783
Italy	223885	31610	120205
Brazil	220291	14962	84970
France	179630	27532	60562
Germany	175233	7897	151597
Turkey	146457	4055	106133
Iran	116635	6902	91836
India	85784	2753	30258
Peru	84495	2392	27147
China	84038	4637	79281
Canada	75959	5679	36908
Belgium	54644	8959	14301
Saudi Arabia	49176	292	21869
Mexico	45032	4767	30451
Netherlands	43880	5662	159
Chile	39542	394	16614
Pakistan	38799	834	10880

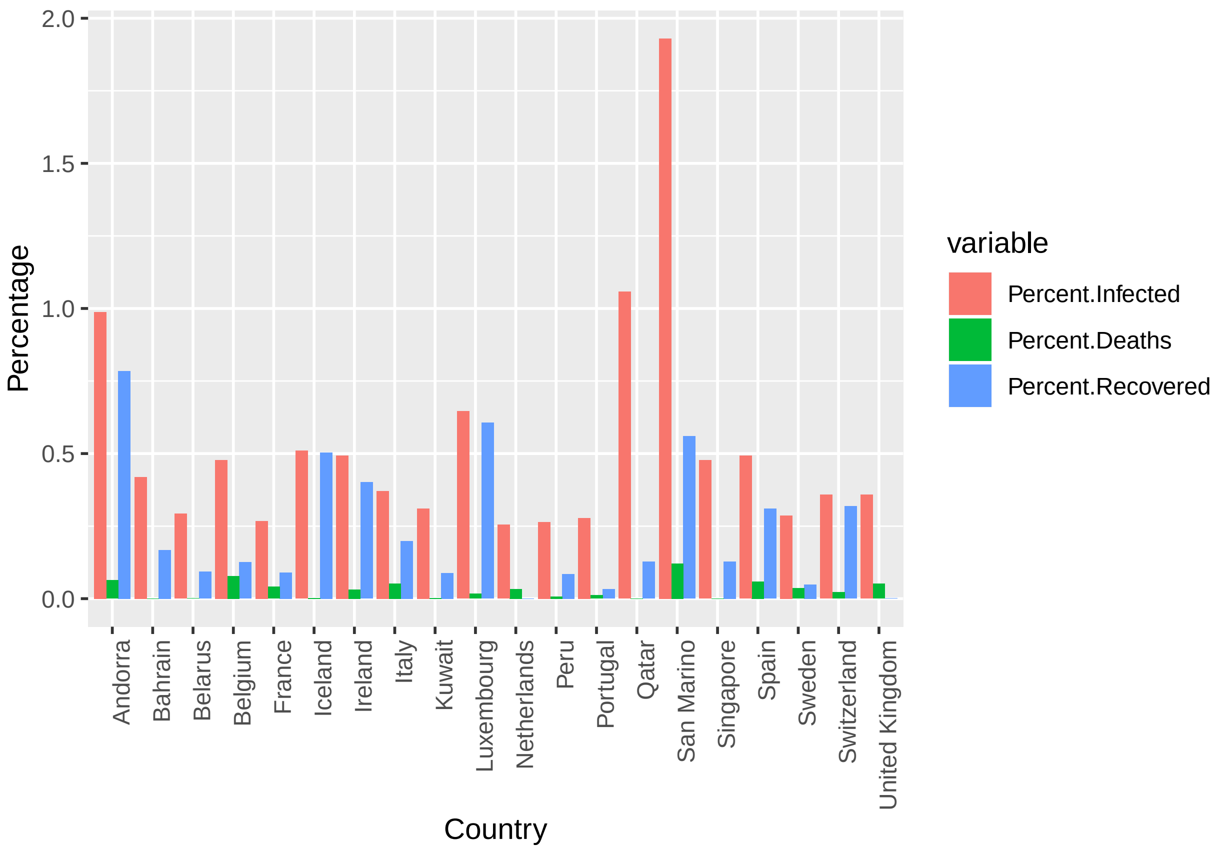


3.1.3 Percentages Over Population

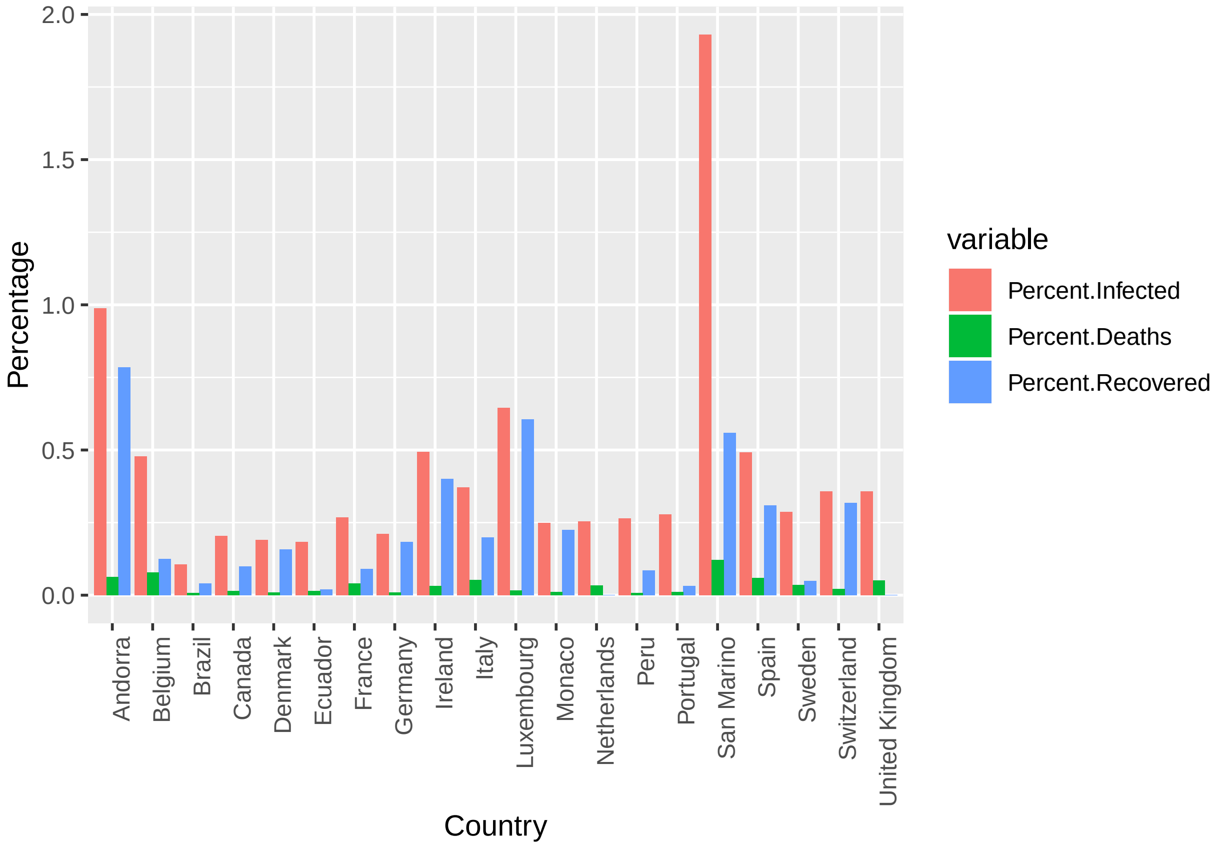
The objective is to know which countries have the worst ratio of deaths and infected people over their population.

Country	Percent.Infected	Percent.Deaths	Percent.Recovered
San Marino	1.9298505	0.1213556	0.5594199
Qatar	1.0578151	0.0005033	0.1274771
Andorra	0.9882347	0.0636314	0.7843545
Luxembourg	0.6455190	0.0171129	0.6058631
Iceland	0.5096529	0.0028283	0.5039963
Ireland	0.4935813	0.0312764	0.4011533
Spain	0.4926467	0.0587688	0.3098703
Belgium	0.4784072	0.0784359	0.1252050
Singapore	0.4769027	0.0003724	0.1285408
Bahrain	0.4194492	0.0007646	0.1682130
Italy	0.3704786	0.0523073	0.1989119
Switzerland	0.3582909	0.0220512	0.3182042
United Kingdom	0.3579600	0.0512536	0.0015747
Kuwait	0.3108301	0.0023203	0.0879799
Belarus	0.2923445	0.0016446	0.0928481
Sweden	0.2868162	0.0358042	0.0488158
Portugal	0.2779971	0.0115739	0.0323680
France	0.2681555	0.0411004	0.0904083
Peru	0.2641356	0.0074775	0.0848629
Netherlands	0.2546571	0.0328593	0.0009228

Country	Percent.Infected	Percent.Deaths	Percent.Recovered



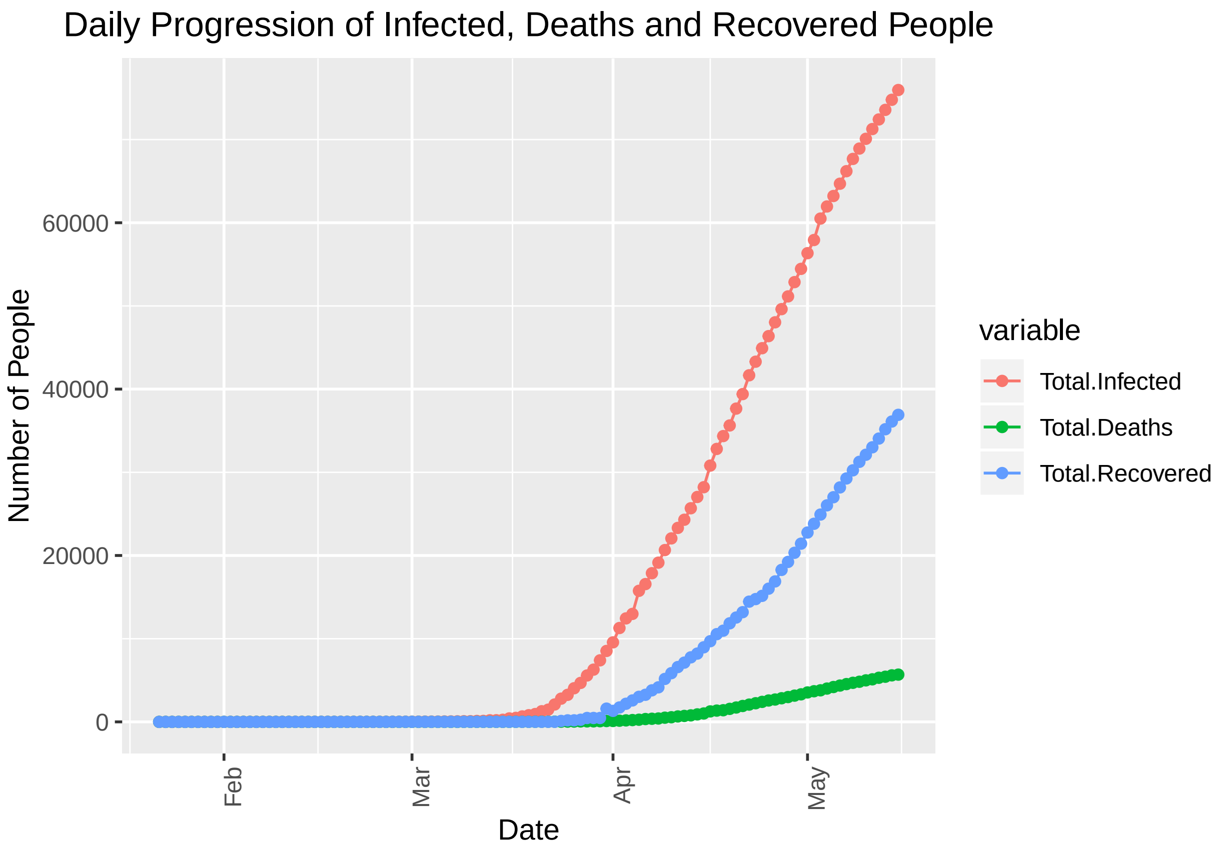
Country	Percent.Infected	Percent.Deaths	Percent.Recovered
San Marino	1.9298505	0.1213556	0.5594199
Belgium	0.4784072	0.0784359	0.1252050
Andorra	0.9882347	0.0636314	0.7843545
Spain	0.4926467	0.0587688	0.3098703
Italy	0.3704786	0.0523073	0.1989119
United Kingdom	0.3579600	0.0512536	0.0015747
France	0.2681555	0.0411004	0.0904083
Sweden	0.2868162	0.0358042	0.0488158
Netherlands	0.2546571	0.0328593	0.0009228
Ireland	0.4935813	0.0312764	0.4011533
Switzerland	0.3582909	0.0220512	0.3182042
Luxembourg	0.6455190	0.0171129	0.6058631
Canada	0.2049686	0.0153243	0.0995929
Ecuador	0.1841860	0.0151835	0.0200944
Portugal	0.2779971	0.0115739	0.0323680
Monaco	0.2481774	0.0103407	0.2249108
Germany	0.2113076	0.0095227	0.1828057
Denmark	0.1895490	0.0092627	0.1579489
Peru	0.2641356	0.0074775	0.0848629
Brazil	0.1051662	0.0071428	0.0405644

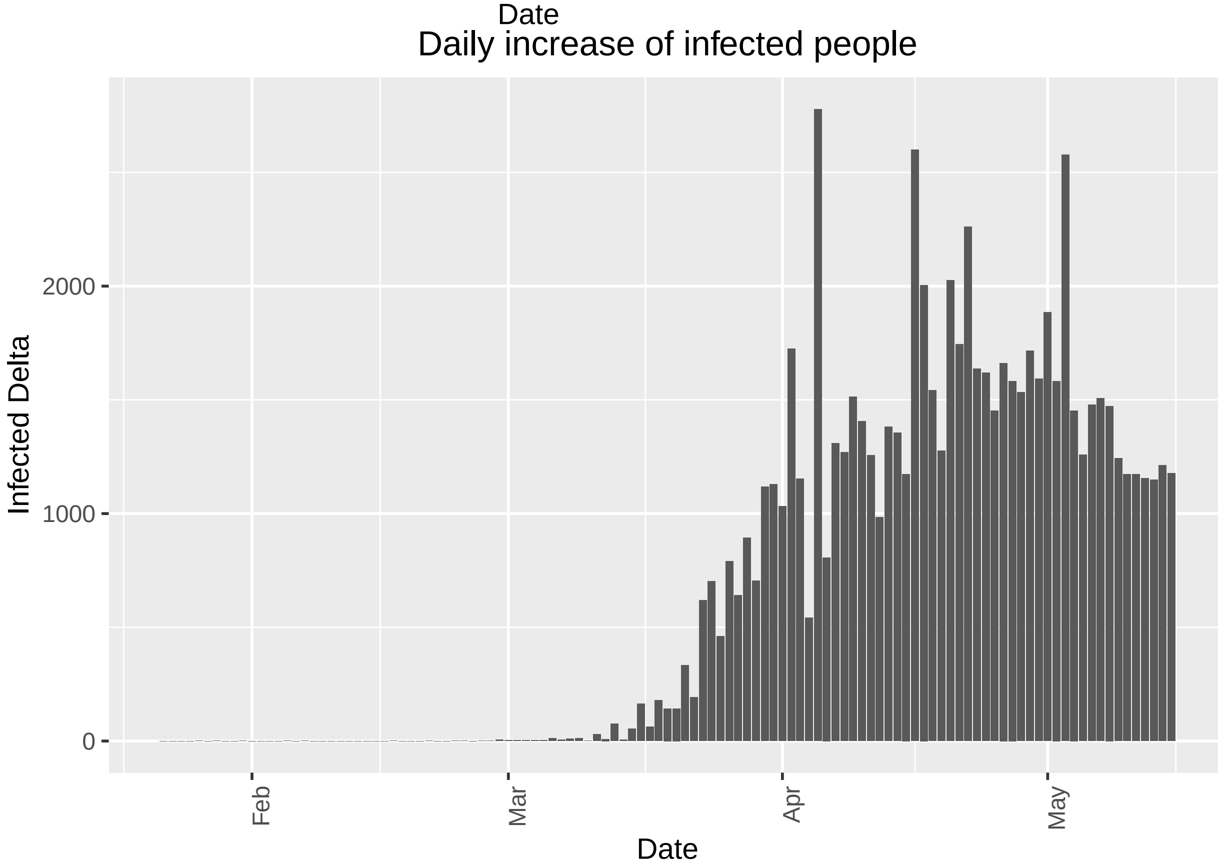


3.2 Canada

The objective is to knwo in which countries the number of infected people start to decrease progressively over days.

Date	Total.Infected	Total.Deaths	Total.Recovered	infected.delta	deaths.delta	recovered.delta
2020-05-02	57926	3684	23814	1583	147	1050
2020-05-03	60504	3795	24921	2578	111	1107
2020-05-04	61957	4003	26030	1453	208	1109
2020-05-05	63215	4190	27006	1258	187	976
2020-05-06	64694	4366	28184	1479	176	1178
2020-05-07	66201	4541	29260	1507	175	1076
2020-05-08	67674	4697	30239	1473	156	979
2020-05-09	68918	4823	31262	1244	126	1023
2020-05-10	70091	4991	32109	1173	168	847
2020-05-11	71264	5115	33007	1173	124	898
2020-05-12	72419	5300	34055	1155	185	1048
2020-05-13	73568	5425	35177	1149	125	1122
2020-05-14	74781	5592	36104	1213	167	927
2020-05-15	75959	5679	36908	1178	87	804


Daily Progression of Infected, Deaths and Recovered People
Number of People
60000
40000
20000
0
Feb
Mar
Apr
May
Date
variable
Total.Infected
Total.Deaths
Total.Recovered


Daily increase of infected people
Infected Delta
2000
1000
0
Feb
Mar
Apr
May
Date

3.3 Countries With Stable or Decreasing Propagation

The objective is to identify all countries whose propagation is stable or is 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 number of infected people at day $t \in \mathbb{N}$. If we take the difference between the number of infected people at day t and $t+1$, we get the **propagation velocity**. In mathematical terms, it is expressed as

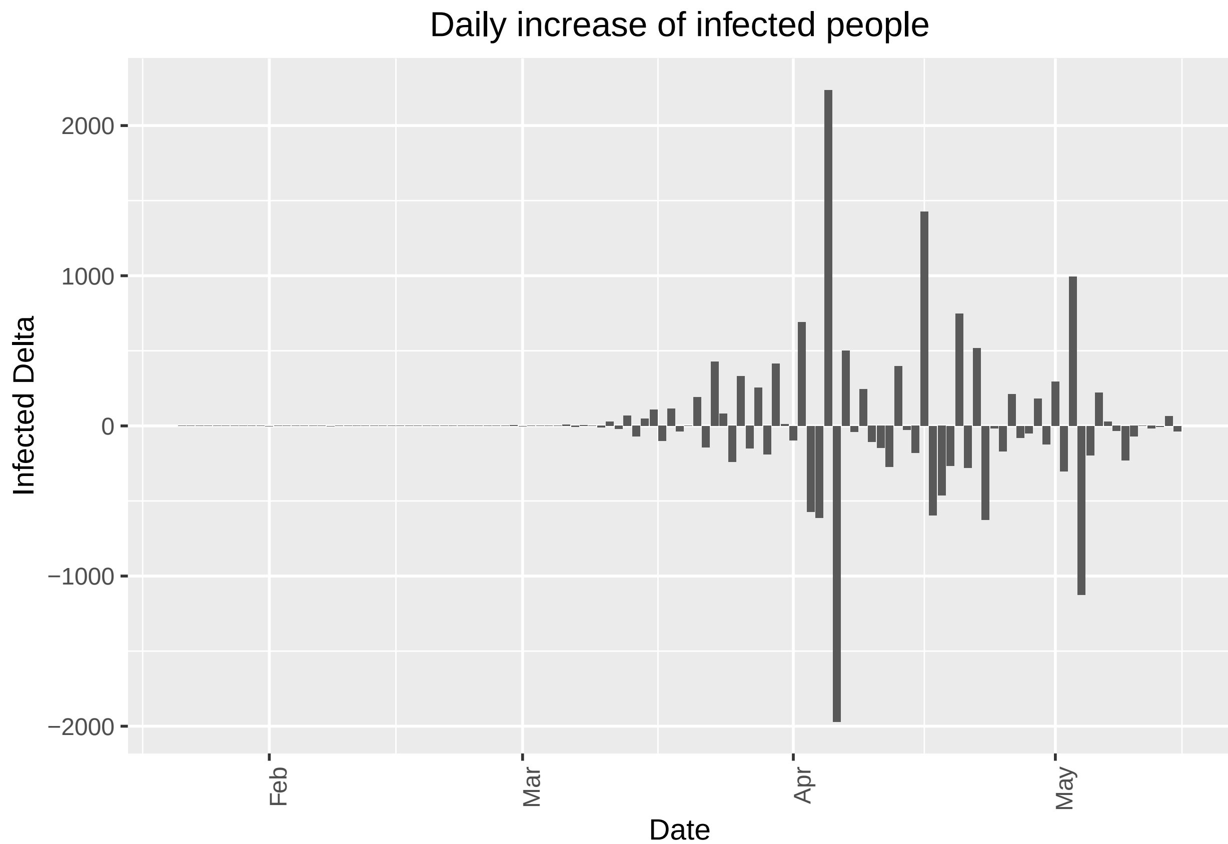
$$\Delta I(t) = \frac{I(t+1) - I(t)}{(t+1) - t} = I(t+1) - I(t)$$

which can be seen as the derivative of $I(t)$ but on a discrete case. The same logic 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** (flatten the curve) if and only if there is neglictable acceleration or deceleration meainng that $\Delta^2 I(t) \approx 0$. Therefore, we need to find countries whose acceleration of the propagation is $\Delta^2 I(t) \leq 0$. However, some of the values of $I(t)$ contained in the dataset may be aberrant values.

Date	Total.Infected	Total.Deaths	Total.Recovered	infected.delta
2020-05-02	57926	3684	23814	-303
2020-05-03	60504	3795	24921	995
2020-05-04	61957	4003	26030	-1125
2020-05-05	63215	4190	27006	-195
2020-05-06	64694	4366	28184	221
2020-05-07	66201	4541	29260	28
2020-05-08	67674	4697	30239	-34
2020-05-09	68918	4823	31262	-229
2020-05-10	70091	4991	32109	-71
2020-05-11	71264	5115	33007	0
2020-05-12	72419	5300	34055	-18
2020-05-13	73568	5425	35177	-6
2020-05-14	74781	5592	36104	64
2020-05-15	75959	5679	36908	-35



It comes now the following questions:

1. What is the range of values considerated as approximative to 0 in the expression $\Delta^2 I(t) \approx 0$? How do we determine that range of values?
2. What are the rules to determine that a value $\Delta I(t)$ is categorized as an aberrant value?

3.3.1 Aberrant Values Detection Model

Assuming that the data 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. 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 acceleration is $\mu = 10.2434783$ and the variance is $\sigma^2 = 158939.5016018$. Therefore, we have $\Delta^2 I(t) \sim N(10.2434783, 158939.5016018)$ where the positive standard deviation is $\sigma = 398.6721731$.

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

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

For example, non-aberrant propagation acceleration values in the Canada have to be inclusively between -787.1008679 and 807.5878245.

Actually, there are 4.3478261 % of the values that are aberrant.

3.3.2 Stable Propagation Model

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 with our assumption:

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 Influenza, that they can be infected after a period of time (virus mutation for example).

4.1 States and Transitions

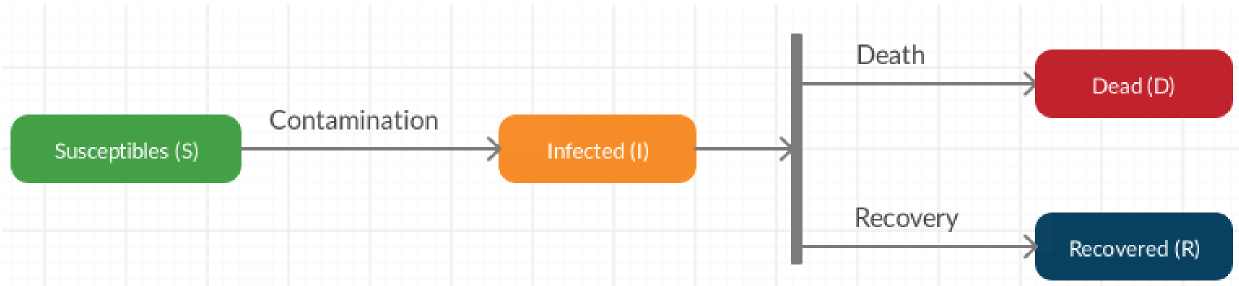
A person can be in one of the following states during the propagation:

- Susceptible (noted $S(t)$): People susceptible to be infected by the COVID-19 at day t.
- Infected (noted $I(t)$): People tested positive to the COVID-19 at day t.

- 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 2020-05-15.

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



The arrows between the states represents 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 disease.
2. An infected person could stay infected the next day or could recover or die. 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 the COVID-19. 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 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, we mean that the probability that they infect others is much 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 the model is not considering the quarantine and social distancing between people which should help to reduce the probabilities of contamination by contacts.
7. All susceptible people will be contaminated one day or another. It does take for account that some of the susceptible people could be immune against the virus, could never be contaminated 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. Thus, the population density is not considered in our model.

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

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

where $S(t)$ will decrease while $I(t) + R(t) + D(t)$ will increase over days. Let $I_0 > 0$ be the initial number of people infected. 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.

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 ($I(t)$), therefore the total number of infected people generated is $\beta S(t)I(t)$. The population will then decrease at this rate.

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

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

Per assumption 2, there is a probability γ that infected people will recover (transition from *Infected* to *Recovered* state) or a probability of α that an 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

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

As the recovered people increase of $\gamma I(t)$ and the dead people increase of $\alpha I(t)$, the number of infected people decrease of $(\gamma + \alpha)I(t)$. We deduce that

$$\frac{\partial I(t)}{\partial t} = \beta S(t)I(t) - (\gamma + \alpha)I(t).$$

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

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

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 $I(0) = 1$, $R(0) = 0$, $D(0) = 0$ and fix the population to $N = 37500000$ people. It follows that $S(0) = 37499999$. 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:

$$\frac{\partial S(t)}{\partial t} = -0.0000000045 \times 37499999 \times 1 = -0.1687499955$$
$$\frac{\partial I(t)}{\partial t} = 0.0000000045 \times 37499999 \times 1 - 0.045 \times 1 - 0.0065 \times 1 = 0.1172499955$$
$$\frac{\partial R(t)}{\partial t} = 0.045 \times 1 = 0.045$$
$$\frac{\partial D(t)}{\partial t} = 0.0065 \times 1 = 0.0065$$

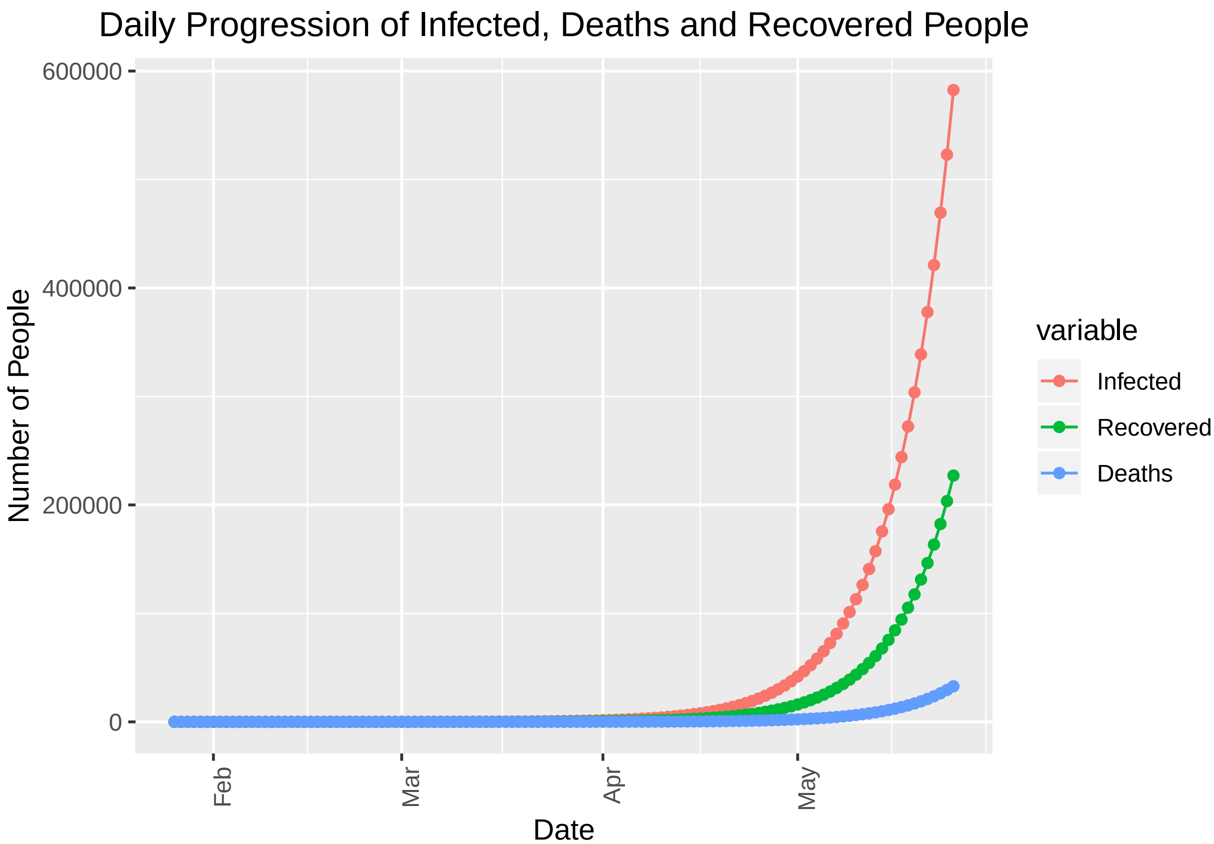
Therefore, we obtain $I(1) = 0.8312500045$, $S(1) = 37499998.83125$, $R(1) = 0.045$ and $D(1) = 0.0065$.

Lets simulate our model for 100 days to see the propression of each state over the days.

Date	Susceptibles	Infected	Recovered	Deaths
2020-01-26	37499999	1	0	0
2020-01-27	37499998	1	0	0
2020-01-28	37499998	1	0	0
2020-01-29	37499998	1	0	0
2020-01-30	37499998	1	0	0
2020-01-31	37499997	1	0	0
2020-02-01	37499997	1	0	0
2020-02-02	37499997	2	0	0
2020-02-03	37499996	2	0	0
2020-02-04	37499996	2	0	0
2020-02-05	37499996	3	0	0
2020-02-06	37499995	3	0	0
2020-02-07	37499994	3	1	0
2020-02-08	37499994	4	1	0
2020-02-09	37499993	4	1	0
2020-02-10	37499992	5	1	0
2020-02-11	37499991	5	1	0
2020-02-12	37499990	6	2	0
2020-02-13	37499989	7	2	0
2020-02-14	37499988	8	2	0
2020-02-15	37499987	9	3	0
2020-02-16	37499985	10	3	0
2020-02-17	37499983	11	4	0
2020-02-18	37499982	12	4	0
2020-02-19	37499979	14	5	0

Date	Susceptibles	Infected	Recovered	Deaths
2020-02-20	37499977	15	5	0
2020-02-21	37499974	17	6	0
2020-02-22	37499971	19	7	1
2020-02-23	37499968	22	8	1
2020-02-24	37499964	24	9	1
2020-02-25	37499960	27	10	1
2020-02-26	37499955	31	11	1
2020-02-27	37499950	34	12	1
2020-02-28	37499944	38	14	2
2020-02-29	37499938	43	16	2
2020-03-01	37499930	48	18	2
2020-03-02	37499922	54	20	2
2020-03-03	37499913	60	22	3
2020-03-04	37499903	67	25	3
2020-03-05	37499891	75	28	4
2020-03-06	37499879	84	31	4
2020-03-07	37499864	94	35	5
2020-03-08	37499848	105	40	5
2020-03-09	37499831	117	44	6
2020-03-10	37499811	131	50	7
2020-03-11	37499789	146	55	8
2020-03-12	37499764	164	62	9
2020-03-13	37499736	183	69	10
2020-03-14	37499705	204	78	11
2020-03-15	37499671	228	87	12
2020-03-16	37499632	255	97	14
2020-03-17	37499589	285	109	15
2020-03-18	37499541	319	122	17
2020-03-19	37499487	356	136	19
2020-03-20	37499427	398	152	22
2020-03-21	37499360	444	170	24
2020-03-22	37499285	497	190	27
2020-03-23	37499201	555	212	30
2020-03-24	37499107	620	237	34
2020-03-25	37499002	693	265	38
2020-03-26	37498885	774	296	42
2020-03-27	37498755	865	331	47
2020-03-28	37498609	966	370	53
2020-03-29	37498445	1080	414	59
2020-03-30	37498263	1206	462	66
2020-03-31	37498060	1348	517	74
2020-04-01	37497832	1506	577	83
2020-04-02	37497578	1682	645	93
2020-04-03	37497294	1880	721	104
2020-04-04	37496977	2100	805	116
2020-04-05	37496622	2346	900	130
2020-04-06	37496226	2621	1005	145
2020-04-07	37495784	2929	1123	162
2020-04-08	37495290	3272	1255	181
2020-04-09	37494737	3656	1403	202
2020-04-10	37494121	4084	1567	226
2020-04-11	37493431	4563	1751	252

Date	Susceptibles	Infected	Recovered	Deaths
2020-04-12	37492661	5098	1956	282
2020-04-13	37491801	5696	2186	315
2020-04-14	37490840	6364	2442	352
2020-04-15	37489766	7110	2728	394
2020-04-16	37488567	7943	3048	440
2020-04-17	37487227	8874	3406	492
2020-04-18	37485730	9914	3805	549
2020-04-19	37484057	11076	4251	614
2020-04-20	37482189	12374	4750	686
2020-04-21	37480102	13824	5307	766
2020-04-22	37477770	15443	5929	856
2020-04-23	37475166	17252	6624	956
2020-04-24	37472256	19273	7400	1068
2020-04-25	37469006	21531	8267	1194
2020-04-26	37465376	24052	9236	1334
2020-04-27	37461321	26869	10319	1490
2020-04-28	37456791	30015	11528	1665
2020-04-29	37451732	33528	12878	1860
2020-04-30	37446081	37452	14387	2078
2020-05-01	37439770	41834	16073	2321
2020-05-02	37432722	46728	17955	2593
2020-05-03	37424851	52193	20058	2897
2020-05-04	37416061	58295	22407	3236
2020-05-05	37406245	65108	25030	3615
2020-05-06	37395286	72714	27960	4038
2020-05-07	37383050	81206	31232	4511
2020-05-08	37369389	90684	34886	5039
2020-05-09	37354139	101264	38967	5628
2020-05-10	37337117	113071	43524	6286
2020-05-11	37318119	126245	48612	7021
2020-05-12	37296919	140944	54293	7842
2020-05-13	37273263	157341	60636	8758
2020-05-14	37246872	175629	67716	9781
2020-05-15	37217435	196022	75619	10922
2020-05-16	37184605	218756	84440	12197
2020-05-17	37148000	244095	94284	13618
2020-05-18	37107196	272328	105269	15205
2020-05-19	37061722	303777	117524	16975
2020-05-20	37011058	338796	131194	18950
2020-05-21	36954632	377775	146439	21152
2020-05-22	36891809	421142	163439	23607
2020-05-23	36821894	469368	182391	26345
2020-05-24	36744121	522969	203512	29396
2020-05-25	36657648	582509	227046	32795



4.5 Model Representation With Markov Chain

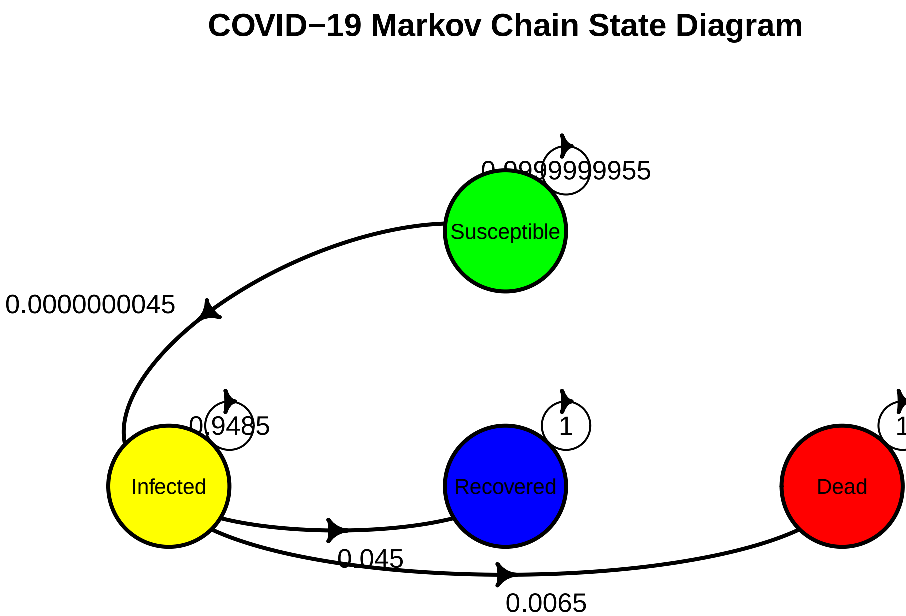
Let $X = (X_t)_{t \geq 0}$ be a sequence of random variables where each random variable is representing a person X at day $t \geq 0$. Let $\mathbb{E} = \{S, I, R, D\}$ the COVID-19 state space. At any day t, a person X_t has to be in a state of $\mathbb{E}$. If $i, j \in \mathbb{E}$, then $\mathbb{P}(X_t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

The objective is to solve the first-order differential equations defined for the 3 transitions in order to find $S(t)$, $I(t)$ and $R(t)$.

From the first equation, we obtain

$$I(t) = \frac{-1}{\beta S(t)}\frac{\partial S(t)}{\partial t}.$$

Then, substituting $I(t)$ in the third equation gives

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

We know that $\frac{1}{S(t)}\frac{\partial S(t)}{\partial t} = \frac{\partial \ln S(t)}{\partial t}$. Thus, we have that

$$\frac{\partial R(t)}{\partial t} = \frac{-\gamma}{\beta}\frac{\partial \ln S(t)}{\partial t}.$$

Integrating on both sides gives

$$R(t) = \frac{-\gamma}{\beta}\ln S(t) + S_0$$

where S_0 is the integration constant representing the initial population. We deduce that $\ln S(t) + S_0 = \frac{-\beta}{\gamma}R(t)$. It follows that

$$S(t) = S_0 e^{\frac{-\beta}{\gamma}R(t)}.$$

We know that $\frac{\partial I(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} - \frac{\partial R(t)}{\partial t}$. After integrating on t, we obtain

$$I(t) = -S(t) - R(t) + I_0$$

where I_0 is the integration constant. From $R(t)$ and $I(t)$, we deduce that

$$I(t) - I_0 = \frac{\gamma}{\beta}\ln S(t) - S(t)$$

We have $I(t)$ in function of $S(t)$ but we want to know what is $S(t)$ in function of $I(t)$. Applying the exponential function on both sides, we get

$$e^{I(t)-I_0} = e^{\frac{\gamma}{\beta}\ln S(t) - S(t)}$$

With I_0 as an arbitraty constant and using the logarithm property where $a\ln x = \ln x^a$, we have

$$I_0 e^{I(t)} = e^{\ln S(t)^{\frac{\gamma}{\beta}} - S(t)}.$$

After simplifications and usage of the exponential properties, we obtain

$$S^{\frac{\gamma}{\beta}}(t)e^{-S(t)} = I_0 e^{I(t)}.$$

We raise the expression to $\frac{\beta}{\gamma}$ on both sides to get

$$S(t)e^{\frac{-\beta}{\gamma}S(t)} = (I_0 e^{I(t)})^{\frac{\beta}{\gamma}}.$$

We multiply both sides by $\frac{-\beta}{\gamma}$ to get

$$\frac{-\beta}{\gamma}S(t)e^{\frac{-\beta}{\gamma}S(t)} = \frac{-\beta}{\gamma}(I_0 e^{I(t)})^{\frac{\beta}{\gamma}}.$$

We obtain the form $we^w = z$ where $w = \frac{-\beta}{\gamma}S(t)$ and $z = \frac{-\beta}{\gamma}(I_0 e^{I(t)})^{\frac{\beta}{\gamma}}$. We can use the W Lambert function stating that for any $w, z \in \mathbb{C}$, we have $we^w = z \Leftrightarrow w = W(z)$. Therefore, we obtain

$$S(t) = \frac{-\gamma}{\beta}W\left(\frac{-\beta}{\gamma}(I_0 e^{I(t)})^{\frac{\beta}{\gamma}}\right).$$