.

COVID-19 in South Korea

Alex Petrallo

August 11, 2020

1 Introduction

1.1 Background

In the past six months the novel corona virus pandemic has affected most every country in the world. Several countries have dealt with the problem in different ways, but South Korea is viewed as the most successful. South Korea was able to 'flatten the curve' of cases without shutting down major business or having as strict lockdown measures as other wealthy nations.

1.2 The Problem

By analyzing the COVID-19 data we can gain some insight to as to how South Korea was able to be successful in handling the pandemic.

1.3 The Interest

The insights that I will be uncovering in this data will be useful for future epidemics were leaders will attempt to minimize the spread. This is information will also be good for the individual to better understand how they can protect themselves during a pandemic.

2 Data Acquisition and cleaning

2.1 Data Sources

I found a few csv files from one data set on Kaggle that provided data on COVID-19 in South Korea. The data set includes information on the patient details, location, date, testing data, and case status. The dataset can be found: here.

2.2 Data Cleaning

The availability of multiple data tables in the dataset allowed me to compare a lot of different aspects of the pandemic. Due to having a more encompassing set of data, I was able to annotate the curve of cases plot with data from the policy table that may show insight as to how the curve is affected by the government's actions. When comparing different data, such as death counts and cases count, it makes sense to normalize these numbers so accurate

comparison can take place. Some rows and columns had to be dropped due to Nan's, or otherwise unneeded information.

2.3 Feature Selection

The main patient data set had 5165 samples with 14 features. I used this data to get charts of gender ratios, age ratios and number of deaths per positive case. I used the date data to plot the overall accumulation of cases. So far, the data set is accurate up to June 30, 2020. Since then, South Korea has had 78 deaths with 5165 cases. Some other main attributes of the data considered are gender, age, sector, and state of the patient.

3 Exploratory Data Analysis

3.1 What is this section about????****

The analysis of this disease should start with how the Corona virus has been transmitted in Korea. Figure 1 shows how many of each type of infection case have occurred. The most common way this disease was spread was via direct contact with a known patient. The second most likely was infection from people from other countries. The rest seem to be due to a lack of social distancing as these places are social places or work environment. South Korea probably saw the overseas data and in response started to preventing immigration form certain countries before restricting all travel into South Korea.

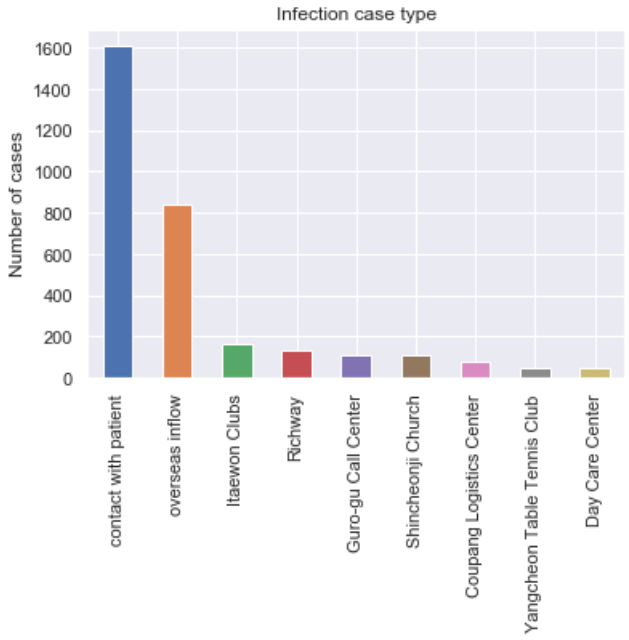

Figure 1: Number of Infections by type of contraction case.

3.2 Data on sex of patients

Some interesting stats I found on the sex and gender of patients is that although men were less likely to contract COVID-19, it was much more likely to be fatal for them. Women were 1.2 times as likely to contract the virus and men were 1.7 times are likely to die if they contracted the virus. The fatality rate in women is 0.0126% and in men is 0.0257%. Figures 2 and 3 show this data represented in bar graphs so you see the relation between the genders.

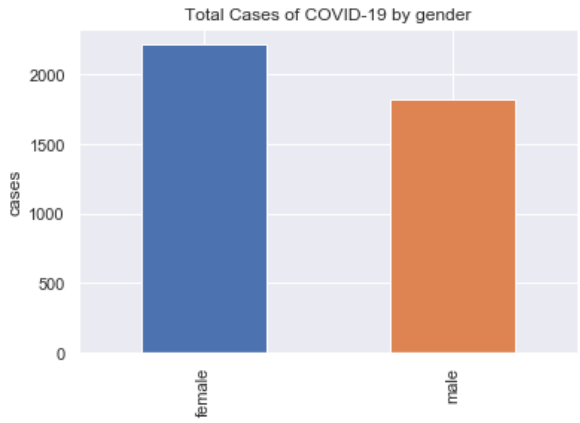

Figure 2: More cases of women contracted COVID-19

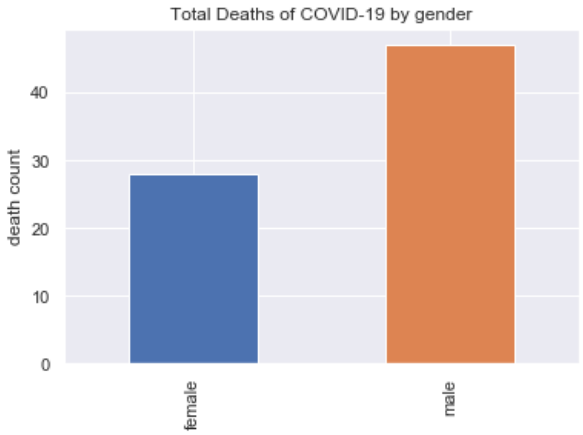

Figure 3: still more men died overall

There can be many reasons that more women contracted the virus; ie) their prominent role in healthcare as well as primary caregivers in a household. Specifically, as of 2015, 89% of nurses were women thus showing their prominent role in healthcare. This inevitably caused them to be on the front-lines and contracting more cases. Biologically, women also tend to have stronger immune systems then. Furthermore, men in South Korea have been more likely to die of lung cancer or other pulmonary issues, this could be a biological sex issue in that men's bodies are weaker, while also being a gendered issue in that men are more likely to drink in smoke, especially in Asian countries.

3.3 Data on age of patients

I also used the data to see how different age group are affected by corona virus. I did an initial plot to see how many people from each group contracted the virus. Figure 4 shows the highest number of cases were in the 20s-50s age group. This makes sense as these would be the groups to be most active and involved in things outside, whereas older people and young children may have an easier time staying home. Although the 70+ age groups were some of the lowest groups to contract the virus, they were the most likely to pass due to contraction. Figure 5 shows the plot of death rates per age group. Zero young people (younger than 30) died from COVID-19 in South Korea.

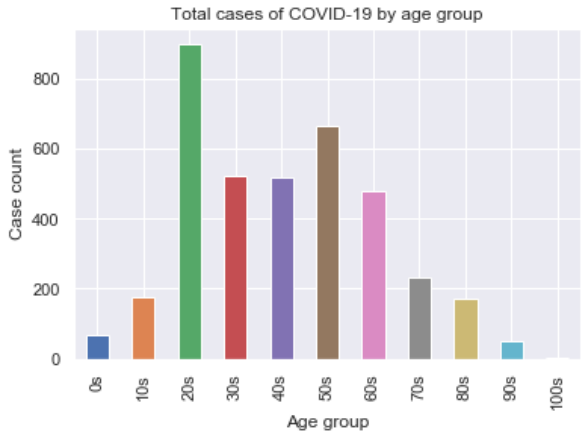

Figure 4: More cases of women contracted COVID-19

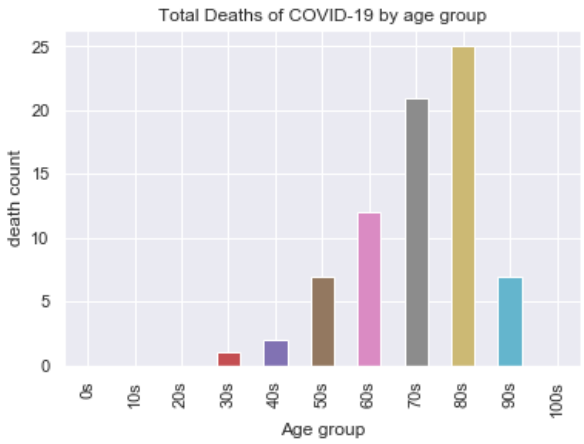

Figure 5: still more men died overall

I then took those two plots and combined them into a stacked bar graph (Figure 6 showing percentage of total cases per age group as well as percentage of total deaths per age group. The chart shows how deadly this virus is for older people. Although making only 3.83% of total cases, the 80s age group had a third of all deaths. The death rate for people in their 80s is 0.147%. The average life expectancy for South Korea in 2018 was 82.7 years old. This puts the people in the 80s age group already more likely to pass, although unfortunate they are taken out by this virus.

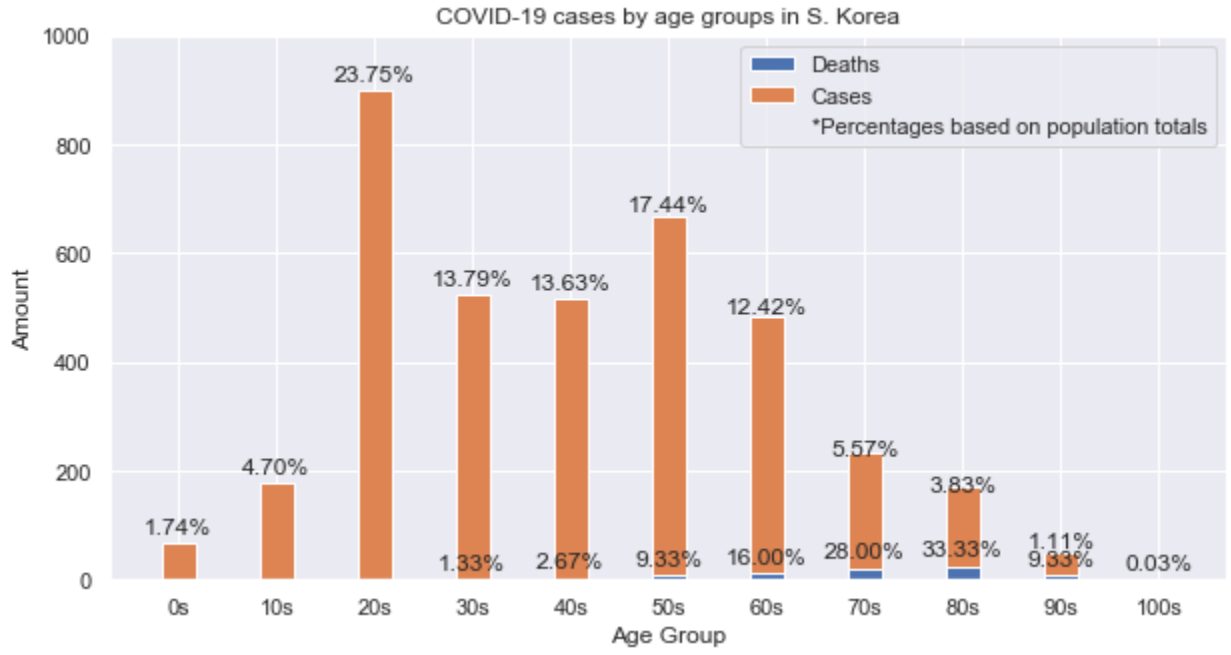

Figure 6: Stacked plot of cases and deaths of age groups.

3.4 Ratios within the testing

Figure 7 shows a plot of the ratio between positive and total tests administered that day. I believe that this can help show that although there are a lot more cases, testing has increased and would therefore affect these numbers. It is also important to consider that the breadth of testing also could mean over-testing or just testing a lot of negative people for cautionary reasons. In the early part of the pandemic, mostly people who were showing the need for the test were taken care of in this way.

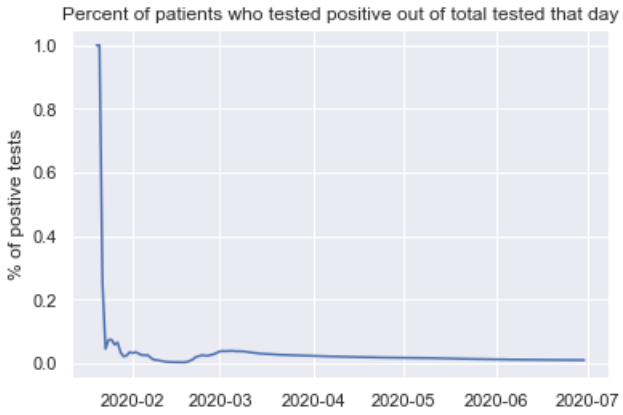

Figure 7: Ratio of positive tests to total per day.

Figure 8 can help show the necessity of test by showing the relationship of positive test and death tolls. The plot shows the percentage of COVID-19 deaths out of total tests that day.

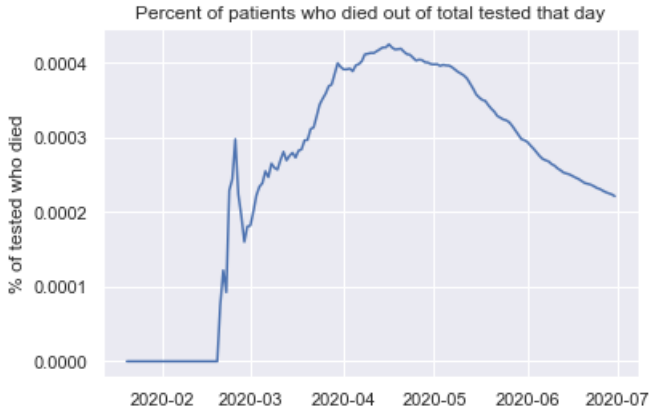

Figure 8: Ratio of deaths to total deaths per day.

Another interesting ratio I found was the relationship between death and total cases that day. I normalized the data so that a comparison can be made and at times the rates of change are different even though you'd expect them to be pretty similar. The plot shows

that the growth of total deaths starts slower than total cases does early on. However, later on death tolls grow faster than the cases do.

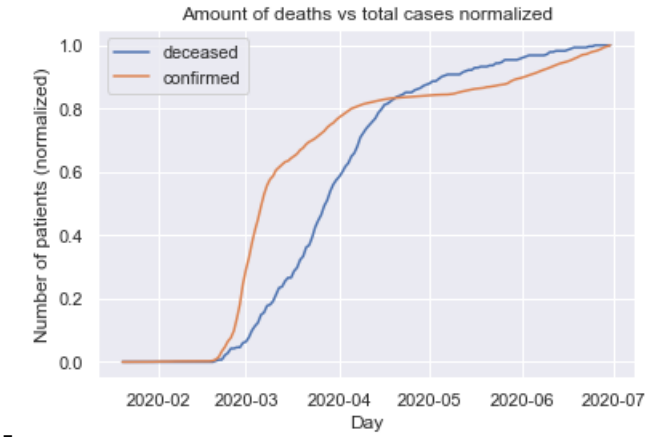

Figure 9: Deaths vs Total Cases Normalized.

3.5 The Curve

Figure 10 shows a plot of the total COVID-19 cases in Korea up to June 27th. I also added some annotated points on the curve that indicate days when certain policies were put in place to foster a discussion on how policy affected the curve. One of the more notable points is the Mandatory 14 Day Quarantine that started on April 1. The curve seems to flatten by the end of the month for a little bit. This could also be an impact of the Special Immigration Screenings coming from all countries two weeks prior. There also seems to be a steady trend of increased cases after school gets out which may be a social indicator for people to visit each other more or just the start of summer time lead people to be more social. Finally, when the bars and clubs open again, the already increasing slope of the curve, it just makes it worse it seems.

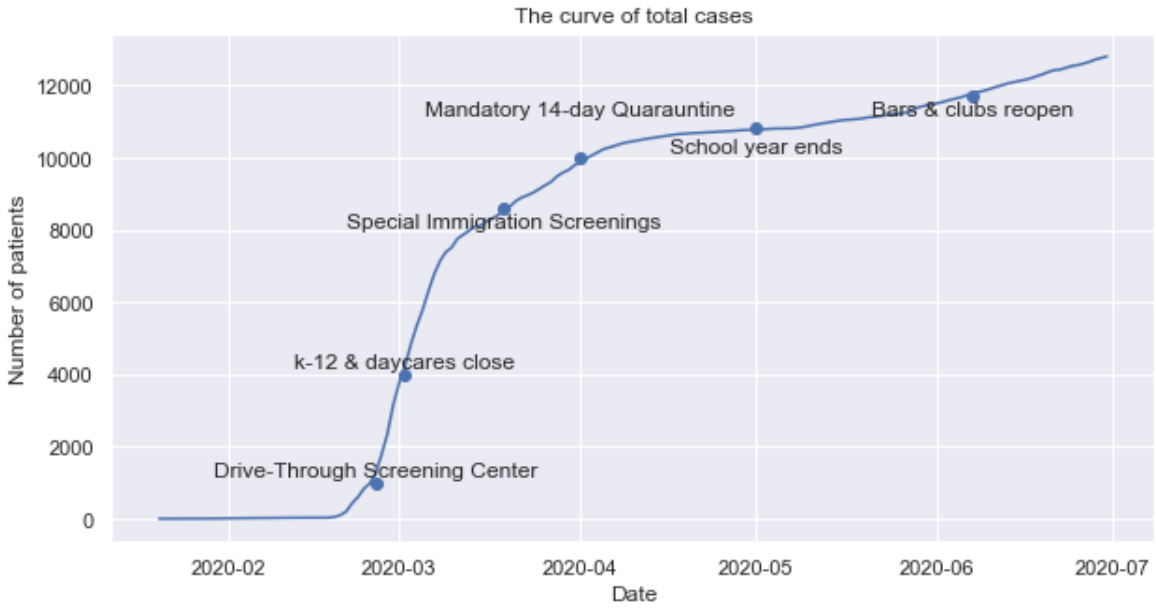

Figure 10: Annotated Plot of total cases.

3.6 The Map

Using Folium and the location data from the data set I was able to create map of Korea plotted with each case per city. The markers were then annotated to provide more information on the click. Figure 11 is the map with the marker Figure 12 shows what pops up when you click on a marker.

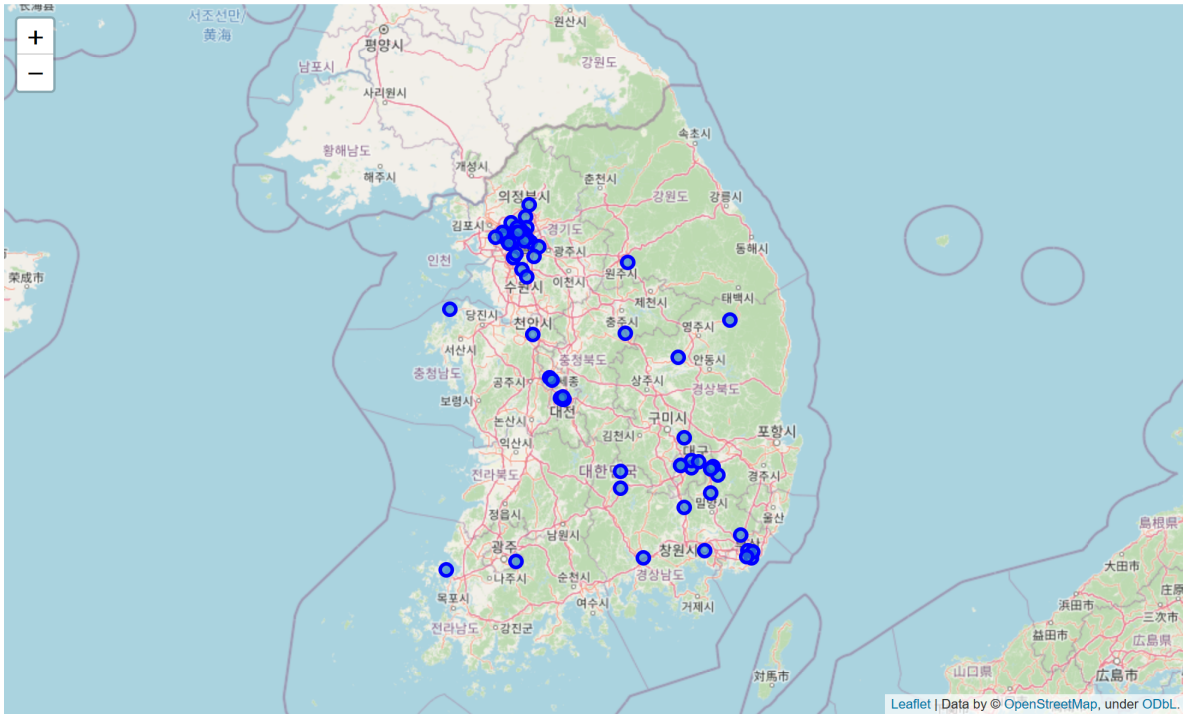

Figure 11: Map of Korea with markers at case locations.

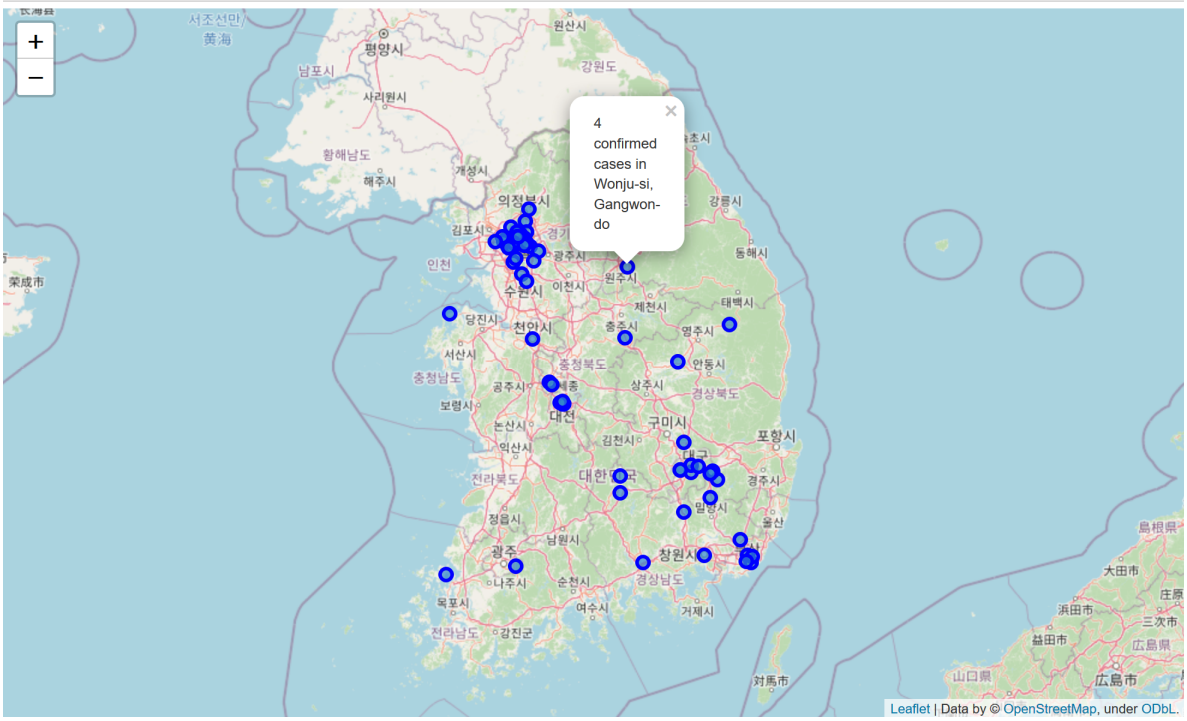

Figure 12: Map of Korea with markers at case locations.

Conclusion

There were many factors that affected how the COVID-19 pandemic played out. It seems the biggest indicators of success in this was age. Younger people simply didn't die although they contracted more. Furthermore it is clear from the data that social distancing is crucially important. Hopefully the insights from this data can be used to make moire educated decision on how to handle an outbreak of a deadly disease.