COVID-19 in South Korea:
Data Visualization

Kathy JaYoung Byun
Rhea Li
Jody Zhu

The current pandemic situation is full of unknowns. Some countries including South Korea are keeping detailed records of those affected, and with the aid of data visualizations, many observations about the large datasets can be made. In this project, we attempt to answer the following three questions:

1. What will be the number of cases in 1 week or in 2 weeks?
2. Which provinces will see the fastest increase or decrease?
3. Which demographics of Koreans are most likely to be confirmed or deceased?

For part one, our analysis is only preliminary and not heavily rooted in mathematics; the main goal is to provide better descriptions of our data with visualizations.

Number of Cases

A huge problem COVID-19 can create is an overwhelming number of patients needing medical care but not enough hospitals and equipment to service them. Having even a rough estimate can improve physical and mental preparations in the coming weeks or months.

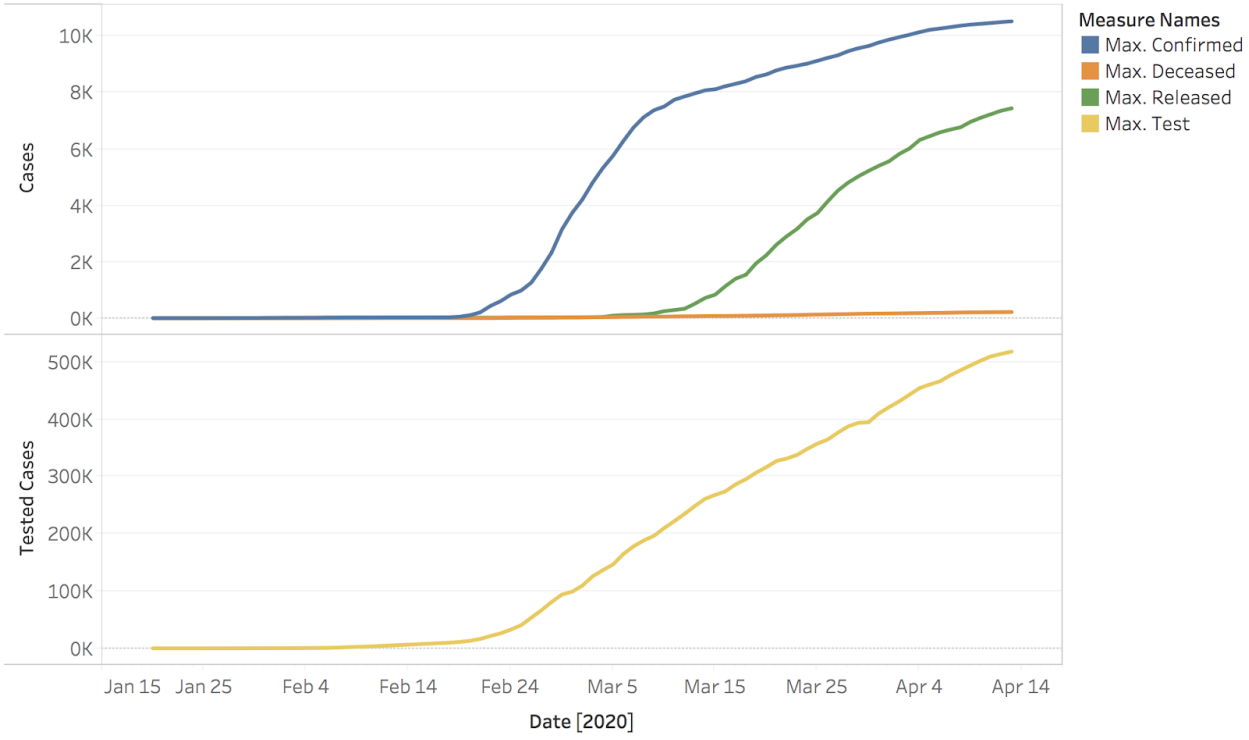

Figure 1: Line graph: Number of confirmed, deceased, released, and tested cases over time

We can see from Figure 1 that as the number of tests conducted goes up, the total number of confirmed cases goes up as well. This increase in testing volume possibly indicates greater government effort put into combatting COVID-19. As of mid-April, the total number of confirmed

cases is plateauing after 2 months of exponential growth. Estimates for the next 2 weeks should see little change and stay between 10,000 and 11,000.

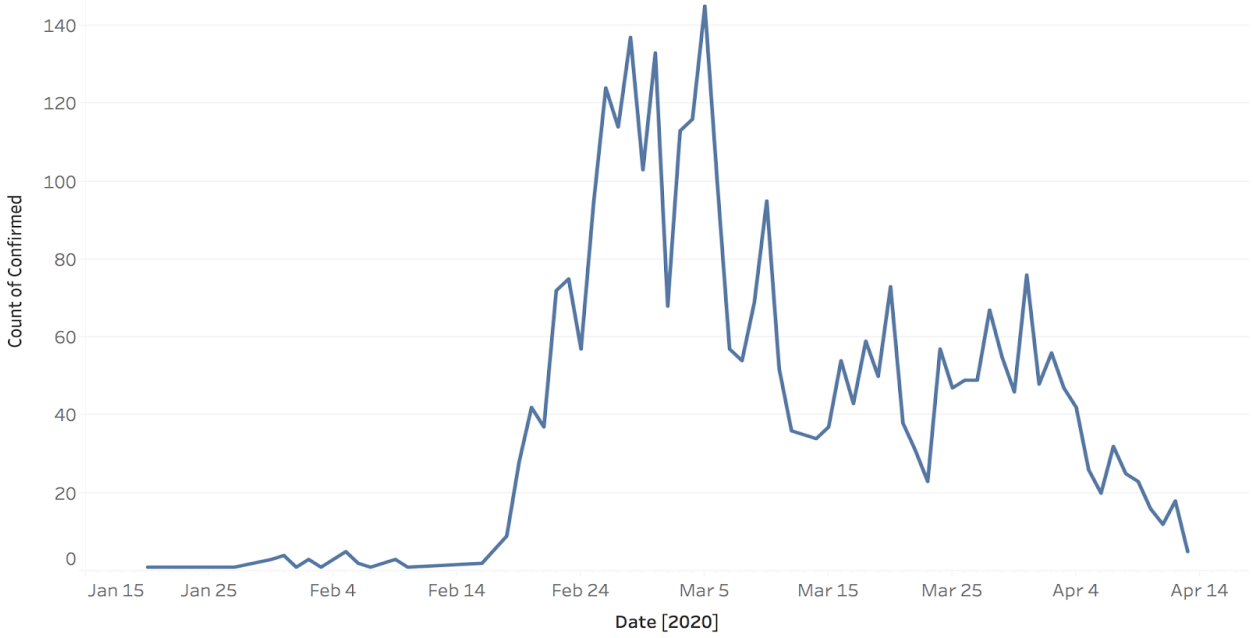

Figure 2: Line graph: Number of new confirmed cases each day

From Figure 2, the number of new cases each day peaks around March 5th and begins decreasing; there is a slowing down trend overall though it remains between 40 to 80 for the remainder of March. From the start of April up until recently, we can see a somewhat steady decrease in the number of newly confirmed cases each day almost to 0. This leads us to the hopeful interpretation that the number of confirmed cases in the upcoming weeks will not grow.

The availability of hospitals and equipment onwards also depends on deceased and released numbers. In Figure 1, the deceased line is close to linear throughout and likely to remain so. Recovered patients or released cases are simply remaining confirmed cases that are not deceased; there should be a continued increase along with more medical care freeing up.

Number of Cases by Province

The nationwide numbers above are composed of those of 9 provinces and 8 special/metropolitan cities, which we will group and loosely call provinces. By breaking down into regions, it will be more informational for individuals in understanding the situation closer to home.

As mentioned above, March 5th marks a turning point for the nation, and Figure 3 (below) depicts the number of weekly confirmed cases for each province starting after this date. There is a decrease for most provinces. A steeper negative slope would indicate a faster decrease rate. Daegu has the fastest decrease while Gyeonggi-do and Seoul are more or less tied in second. Based on the trend, Daegu will have the fastest decrease. Because decrease cannot occur past 0

cases, the long term slopes might not be good predictors, so running regression in part two on the noise at the recent end could provide new and interesting insight.

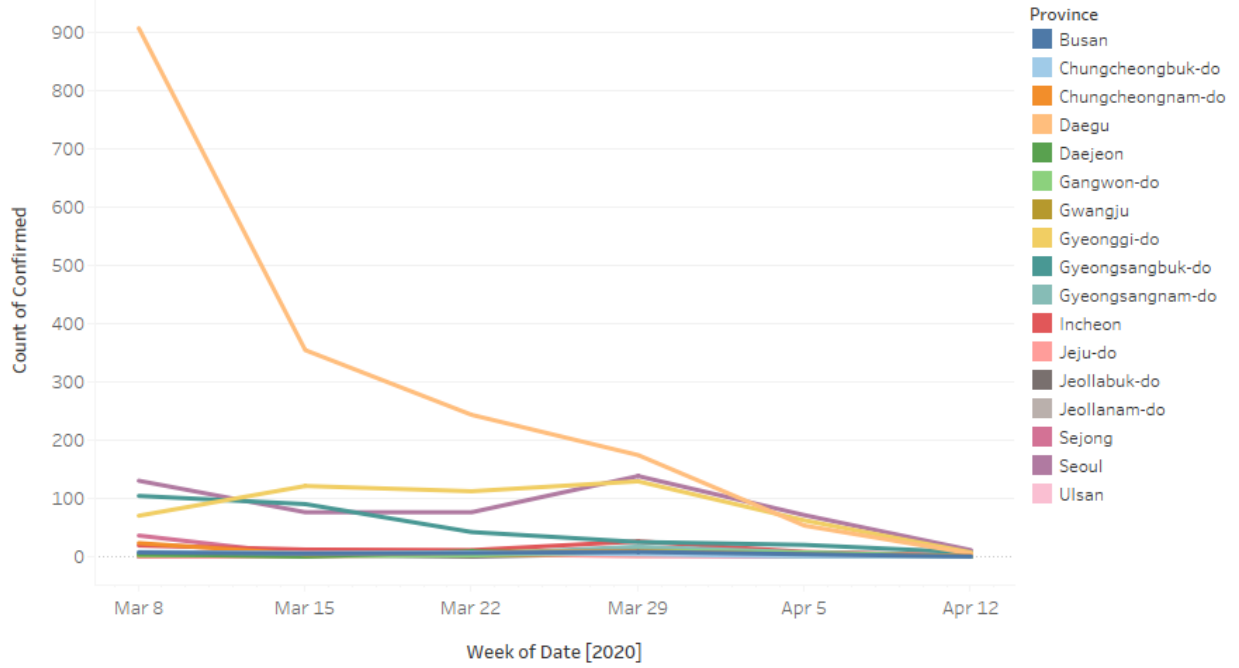

Figure 3: Line graph: Number of new confirmed cases each week by province after March 8

Vulnerable Demographics

Knowing COVID-19's vulnerable populations can lead to further precautions for them. Location, age, and sex are three demographics available in the dataset.

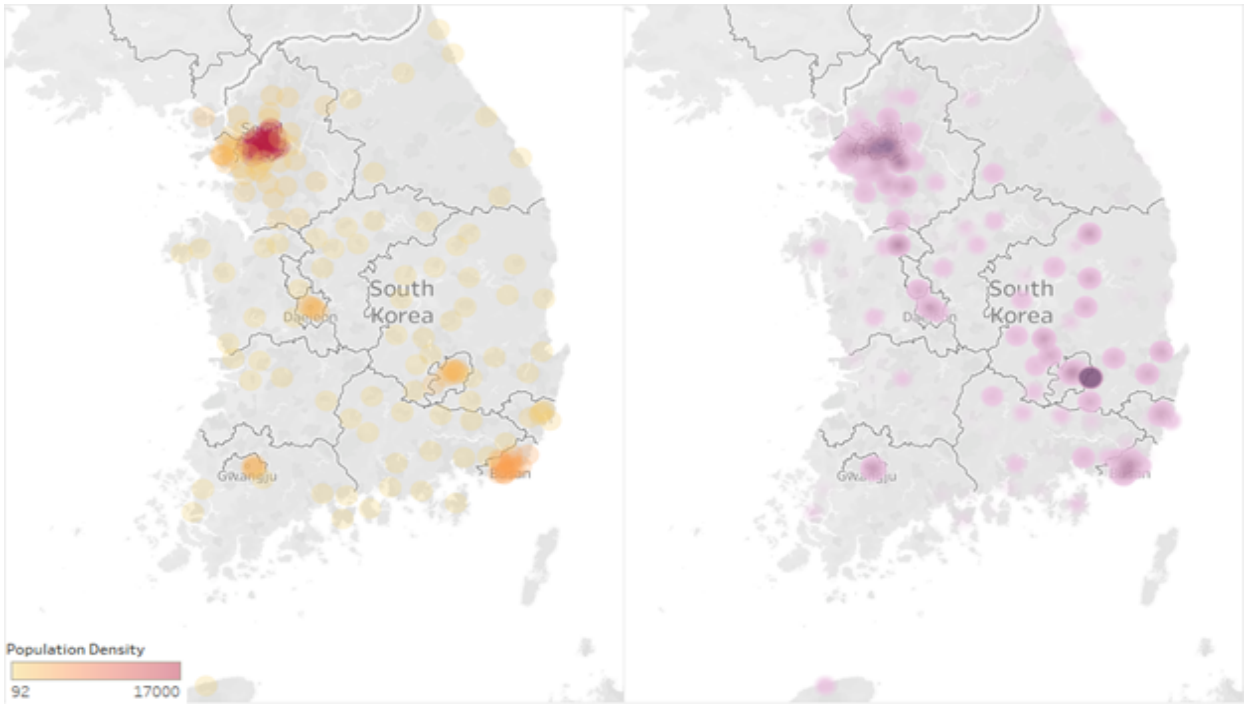

Figure 4: Density maps: Left is population density and right is number of patients

Continuing the discussion of provinces, we examine a possible correlation with population density. The darker red and oranges signify higher population density in the left map of Figure 4, and they seem to match the darker purple patient hotspots in the right map. This means more confirmed cases can occur where more people reside. Information about the deceased is not shown.

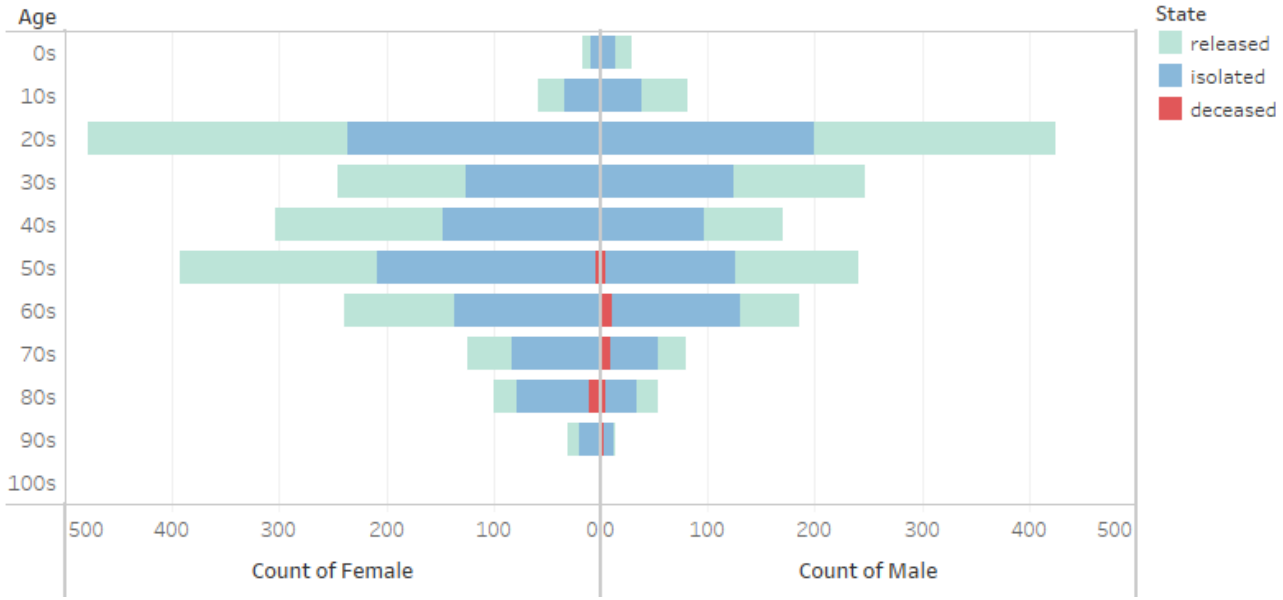

Figure 5: Butterfly stacked bar chart: Number of patients in each state categorized by age and split by sex

In Figure 5, age is depicted by horizontal bars. The number of confirmed cases, which is the summation of cases from all the states, is the total length of the horizontal bars. Currently, the 20s are in the lead followed by the 50s. In general, those who frequently commute to work using public transportation (20s-50s) have greater numbers of confirmed cases than those who are too young or retired. This supports the stay at home campaigns to stop the spreading. Meanwhile, deceased cases are marked in red and only appear for 50s and older; the numbers appear to be growing until it reaches the max in the 80s. This distribution suggests that the elderly face a higher case fatality rate even though fewer of them are infected.

Sex is shown side by side in Figure 5. For the number of confirmed cases, the bars on the right of 0 roughly mirror the shape of the bars on the left. The discrepancies between female and male are small enough where the observations made about age above still hold for both, but it is important to note that there have been more female patients overall and in 8 of the 10 age groups. For the number of deceased, males have slightly more. The fact that males also have fewer confirmed cases in the denominator means that their death rate is higher; however, the significance of the difference cannot be confirmed just from this visualization.

The next step of this project is using tools to reexamine the observations made so far and to better answer the questions.

Bibliography

1. Kim, Jihoo. “Data Science for COVID-19 (DS4C).” *Kaggle*, 14 Apr. 2020, www.kaggle.com/kimjihoo/coronavirusdataset.
2. “Provinces - Republic of Korea.” *Knoema*, knoema.com/atlas/Republic-of-Korea/Provinces-profiles.