The Analysis of Coronavirus (COVID-19) New Confirmed Cases Comparison Between Top 10 Countries and Taiwan

Micky Lee (Wen Chi Lee)
UCLA Extension
Introduction to Data Science
Instructor: Daniel D. Gutierrez
March 22, 2020

Contents

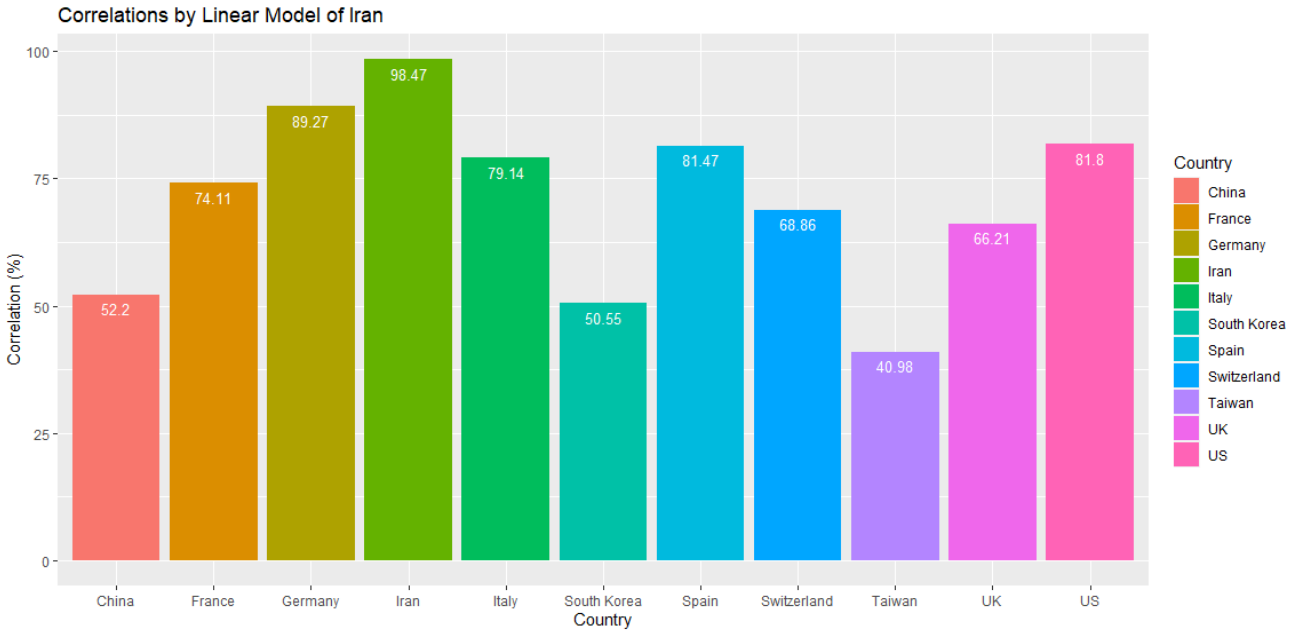
1. Project Description

Since the novel coronavirus (COVID-19) had spread all over the whole world, I want to find out how the new confirmed cases changed in the top 10 confirmed cases countries and my country, Taiwan. I use multiple linear regression to predict and fit between 11 countries, including China, Italy, Spain, Germany, Iran, the United States, France, South Korea, Switzerland, United Kindom, and Taiwan. Then I compare the trends of these countries and try to explain the differences and why the result comes out. The dataset I used collected confirmed, deaths, and recovered data from 179 countries from January 22, 2020, to March 18, 2020.

2. Machine Learning

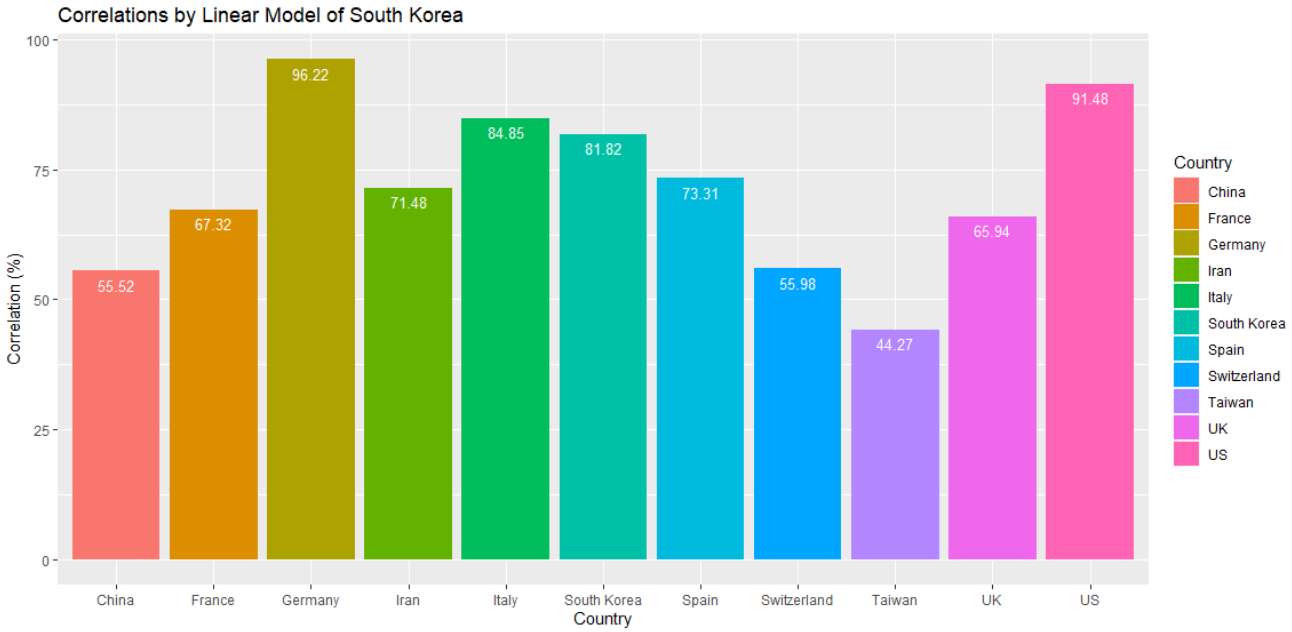
After finished some scatterplots, boxplots, and pairs, I compare them and conclude to train a linear model by Iran and another linear model by South Korea. Iran is currently the third-ranking confirmed cases in the world (until March 18, 2020), and Iran's geographical location is between Asia and Europe. Because the result is diverse in Asia, so I choose South Korea as the other linear model. South Korea has the highest confirmed cases in Asia after China's. Also, South Korea is very close to China, which is the original COVID-19 outbreak region. We have the following visualizations for these countries' relationships.

2.1 Correlation between Iran and other countries



Almost all countries fit the model except China, South Korea, and Taiwan. These three countries have the badest correlations between the model. I notice that they are all in Asia. Taiwan and South Korea both have the first confirmed case on January 22, 2020, which is earlier than other countries. Moreover, both Taiwan and South Korea are very close to China geographically. I will try to explain these situations later.

2.2 Correlation between South Korea and other countries



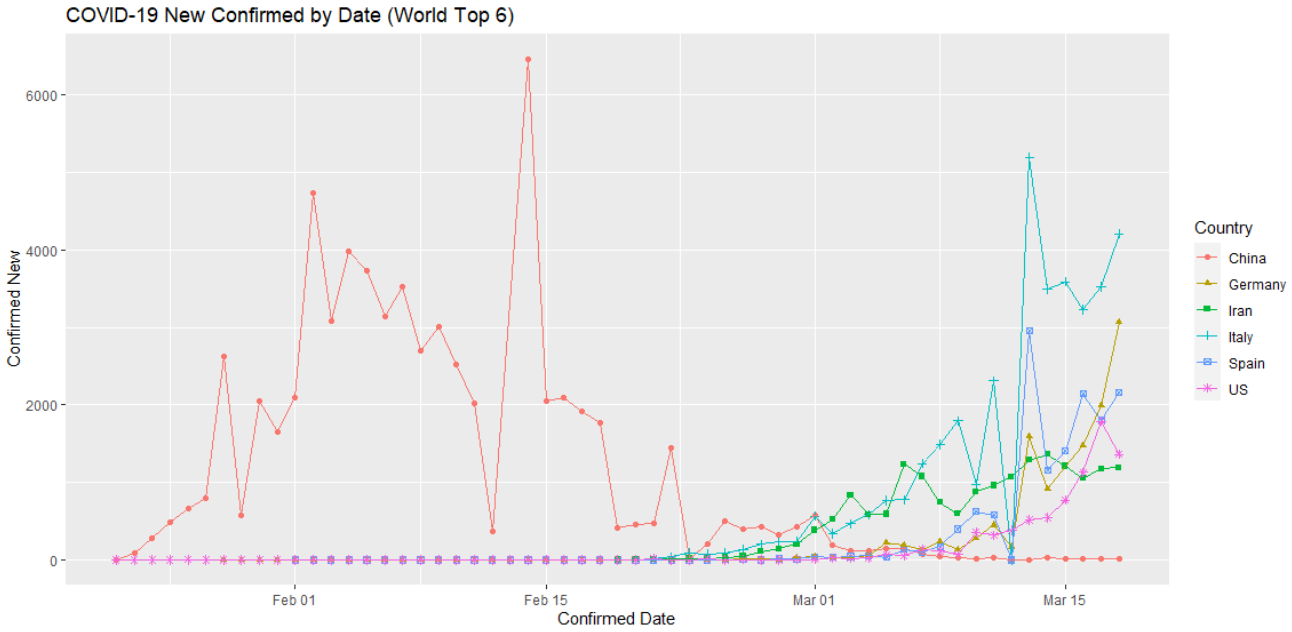
The result is shocking. We can see that China and Switzerland have about 55% correlation with the linear model. Even though South Korea is near China, China's correlation is still so low. The reason why Switzerland has such a weak relationship is that it only has 23-days data, which might not be enough to predict. However, Taiwan, also close to South Korea, has the lowest 44% correlation with the linear model. I will explain the reasons for China and Taiwan later.

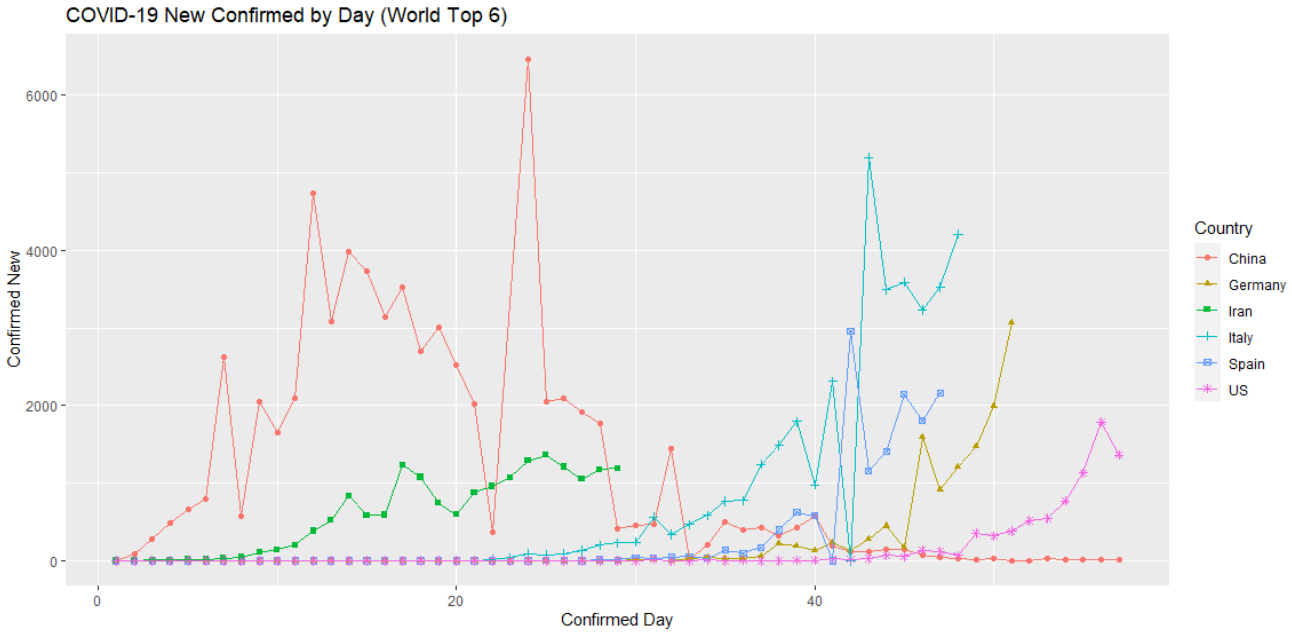
3. Trend comparison

I compare the trend of new confirmed cases of each country in three different categories.

We can compare the trends of new confirmed cases between these countries. Considering the visualizations should be easy to understand, so I separate the comparison into three categories with two different variables. The first variable is **Observation Date**, which is mapping the exact date when new confirmed cases happened. The second variable is **Confirmed Day**, which indicates how many days since the first confirmed cases. Through comparing these two variables, we may found something useful.

Category 1: World Top 6



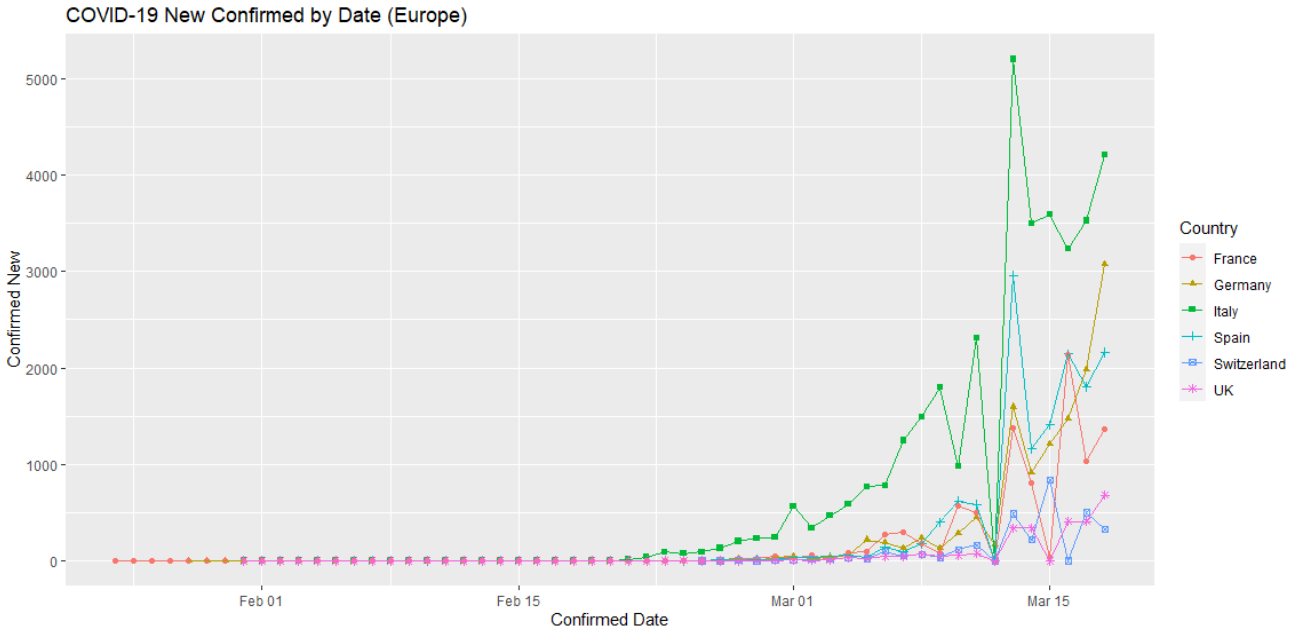


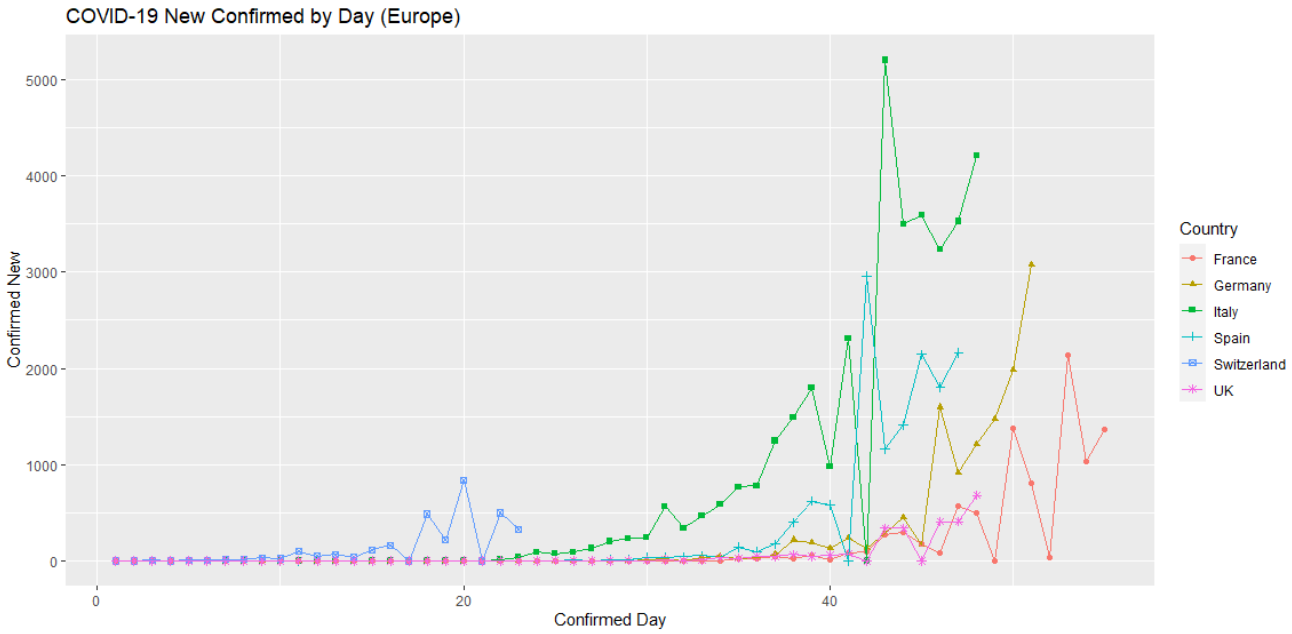
The first plot shows the new confirmed cases by **date**, and the second plot shows the trend by **day**.

In the first plot, we can see that when China's new confirmed cases are going down, the new confirmed cases in other countries are just starting the outbreak of COVID-19. According to the second plot, we can see that after 30 days, Italy, Spain, Germany, and the US, they are all increasing the new confirmed cases daily. But for Iran, it is in the increasing trend after just ten days passed. Let's talk about China's data, it kept average above 2000 confirmed cases daily for at least 16 days, but the number drastically down to only double digits in 10 days, which is believed that the report may be fake. Further, there is a super peak 15133

new confirmed cases eliminated from the data to avoid the range is too wide. According to the media and news during the period after the highest peak, Xi Jinping, the president of China, said the disease is under control. After then, the report of new confirmed cases is drastically down.

Category 2: Europe



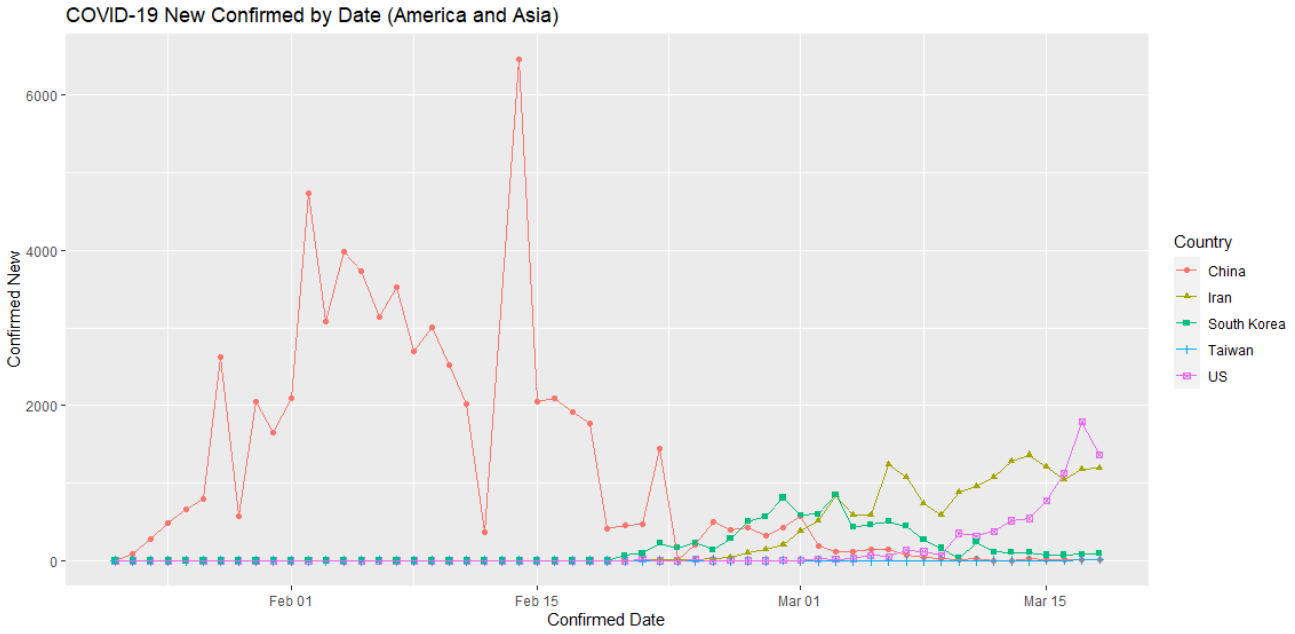


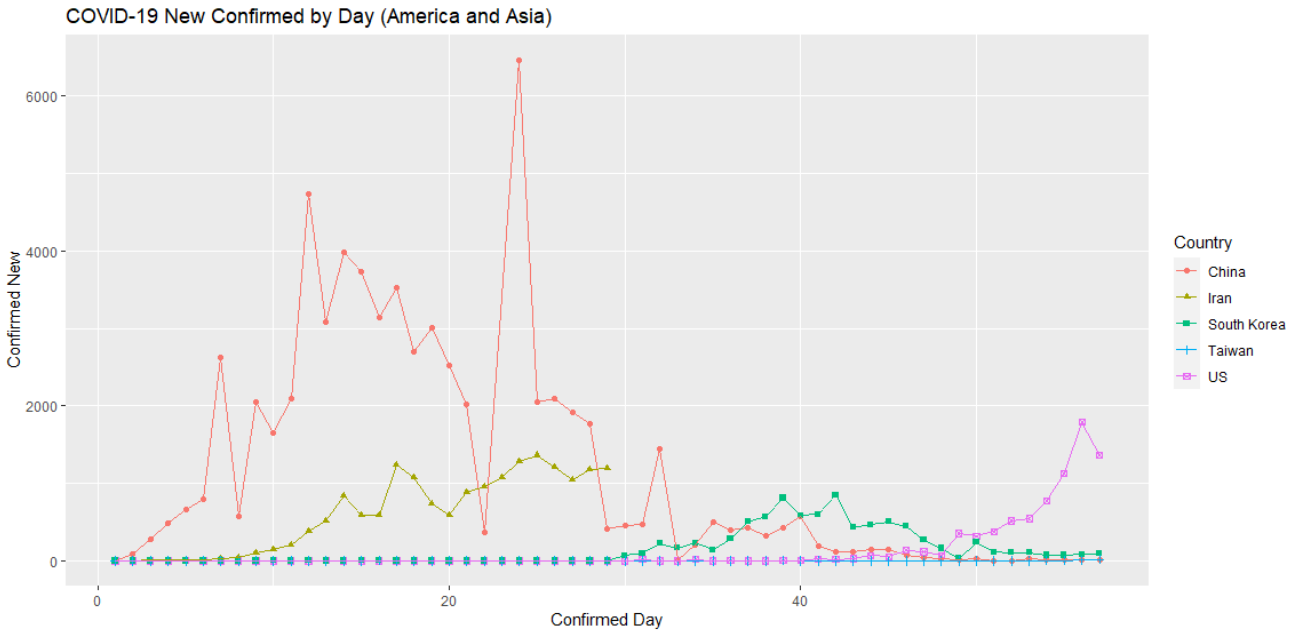
The first plot shows the new confirmed cases by **date,** and the second plot shows the trend by **day**.

In the first plot, we can see that only Italy has the increasing new confirmed cases in the earlier time, and then almost the same day, all European countries increased new confirmed cases drastically on March 13,

2020. According to the second plot, we realize that the coronavirus just spread in Switzerland for about 23 days, which is fewer than in other countries. The other countries increased a large amount of new confirmed cases after 35~40 days since the first confirmed case.

Category 3: America and Asia





The first plot shows the new confirmed cases by **date,** and the second plot shows the trend by **day**.

In the first plot, we can see that, when the new confirmed cases are going down in China, other countries started to increase the new confirmed cases except Taiwan. According to the second plot, when

other countries had COVID-19 for near 60 days, Iran just happened the disease for less than 30 days. We also notice that China, South Korea, and Taiwan seem to control the outbreak. The trends show that only Iran and the US are still in the increasing trend; other Asian countries are already in the decreasing trend.

Actually, China had the first coronavirus confirmed case in November 2019. Therefore we know the COVID-19 has already existed more than 100 days since then. The peak of the new confirmed case in China is about 70~80 days after the outbreak.

4. Conclusion

4.1 Linear regression result comparison

The linear models trained by Iran and South Korea all suited for European and American countries in our samples. The only two exceptions are China and Taiwan. As we all know, China government ignored and hid the disease at the very beginning time, and then it blocked the information from the world. In the end, the outbreak is boomed in Wuhan city. The confirmed cases and deaths suddenly increased a lot. After that, President Xi Jinping controlled the media, news, and new confirmed cases. It is believed that the government may report the fake data to WHO when after one of Xi's talk to the citizens in China.

Taiwan is in the other situation. Taiwan suffered SARS in 2003. Since then, Taiwan has always prepared for another disease outbreak. So when Taiwan first heard there might be new coronavirus appeared in Wuhan in the early of January 2020, the government soon decided to build the Central Epidemic Command Center (CECC) to manage all the information about COVID-19. CECC arranged many policies to prevent the disease from spreading and monitor people who may be in danger. It turns out to make the incredible few confirmed cases in Taiwan, although Taiwan is the closest country to China in the world. Do not forget that there are two to three million people fly to and back between China and Taiwan every year.

According to the above reason, other countries are not so aware of COVID-19 or China like Taiwan. Even the WHO also declared that this disease is not dangerous. So other countries will not prepare the outbreak because they believe WHO. But people in Taiwan never believe China and know that China actually controls the WHO. Hence, Taiwan prepared for it in advance. The population of Taiwan is 23 million people; meanwhile, the people of South Korea are 51 million. Compare to the only 100 confirmed cases in Taiwan, and there are 8413 confirmed cases in South Korea, it significantly showed why I said Taiwan truly prevents the COVID-19 from spreading.

These explanations tell us why the linear models trained by Iran and South Korea do not suit for China and Taiwan but suit other countries. The reasons are China's data is not precise, and Taiwan's situation is under control without any outbreak. We know that the first confirmed case in Taiwan is the same day as the first confirmed case in South Korea and Japan. But it turns out that South Korea has 8413 confirmed cases, and Japan has 889 confirmed cases, which both are the severe COVID-19 affected area.

4.2. Summary

According to the trends and plots showed above, the top 10 confirmed cases countries have a similar situation since the disease appeared in those countries. The only exception is China in the top 10. By these comparisons, they showed the peak of the new confirmed case would occur after 30 ~ 40 days since the first confirmed case. The linear models produced by Iran and South Korea are almost perfectly suited for other countries. This means the COVID-19 outbreak has the same pattern in almost any country except China and Taiwan. For Taiwan, because it prepares and controls in advance before WHO's announcement. For China, its strange patterns about the disease are obviously controlled by the communist government. Until now, the new confirmed cases are still not transparent in China.