CS7DS4 / CSU44065 Data Visualization 2019-20

Assignment 3

Student Name: Zhu Yuan

Student No: 18303945

Declaration: "I have read, and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year, found at http://www.tcd.ie/calendar.

I have also completed the Online Tutorial on avoiding plagiarism 'Ready Steady Write', located at http://tcd-ie.libguides.com/plagiarism/ready-steady-write."

1 Introduction

In this report, I am going to show the covid_19 spread in West Europe, including Belgium, France, Ireland, Luxembourg, Monaco, Netherlands and the United Kingdom. Because I am living in Dublin right now and I think it is good for us to know about what is happening in the neighbourhoods countries.

West Europe is close countries in regional aspect, and the situation in each country will play an important role on virus spread among these countries.

During the last month is the most serious time in Europe, so I am going to focus on the virus spread in recent time, one month, from 3.19 to 4.19.

2 Description

a) Data
b) Work Environment
c) Approach
d) Novelty
e) Strengths and Weaknesses

a) Data source: Akshare

It is an open project in Github(https://github.com/jindaxiang/akshare/tree/e112727e5c0c7107ccad138c6fc63c455dd122f3), I fell really appreciate for sharing this Real-time data about covid_19. I got the data until 2020.04.19 to do the project but it is flexible I can always use the interface to update my project. Thank you again for sharing!

b) Work Environment

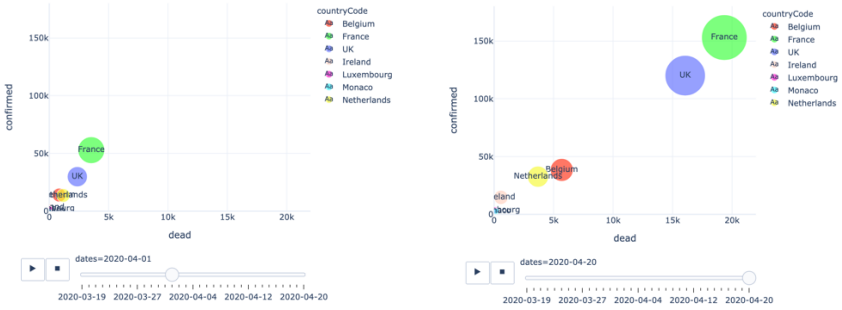
- Mac system
- Anaconda（Python 3.7）
- Jupyter Notebook
- Libraries: akshare, pandas, plotly, datetime(pip install library to use)

c) Approach

Firstly, I get the data form Akshare, then I add one more column date which is another date type I will use later. After that, I pick the West Europe countries data form the data set and replace the simple country code with full country name. After I get West Europe data, I just want to use the data from 2020-03-19 so I get the data in recent month and sort them by date.

Because some countries may not record the data during this time so I fill in null values with 0.

✧ The first chart is a bubble chart.



We can see that the most serious region is in France, from around 10k to over than 150k confirmed people during one month, the UK also has a big number of confirmed patients and significant speed in this period. While Netherlands and Belgium have similar speeds of confirmed patients. Ireland does not have so many patients compare to France and UK but slightly more serious than Luxembourg and Monaco respectively.

- ✧ The second and this chart are bar charts

We can see there is a clear contrast the virus spread in these countries in the different axis system.

d) Novelty

My idea is I just focus on West Europe countries in latest month. And show the data in 3 styles.

The bubble chart is easy and clear to see the daily change and comparison among the countries. The two bar charts can see the contrast respectively in x axis system and y axis system.

e) Strengths and Weaknesses
Users are able to see the covid_19 data in animation, 3 different systems, and get some more straight and clear information. But some data is too small to show off in bubble chart.

3 Citing third party resources

Libraries: akshare, pandas, plotly, datetime.
Data source: Akshare, an open project in Github(https://github.com/jindaxiang/akshare/tree/e112727e5c0c7107ccad138c6fc63c455dd122f3)

References

[1] https://blog.csdn.net/lemonbit/article/details/104624376

[2] https://towardsdatascience.com/see-the-coronavirus-for-yourself-88ce06b88f5e

[3] https://github.com/jindaxiang/akshare/tree/e112727e5c0c7107ccad138c6fc63c455d2f3

My YouTube Link: https://www.youtube.com/watch?v=WKiP-yO0-70