SWRK 730
Assignment 1 - Thematic Maps
Ziyu Zhao

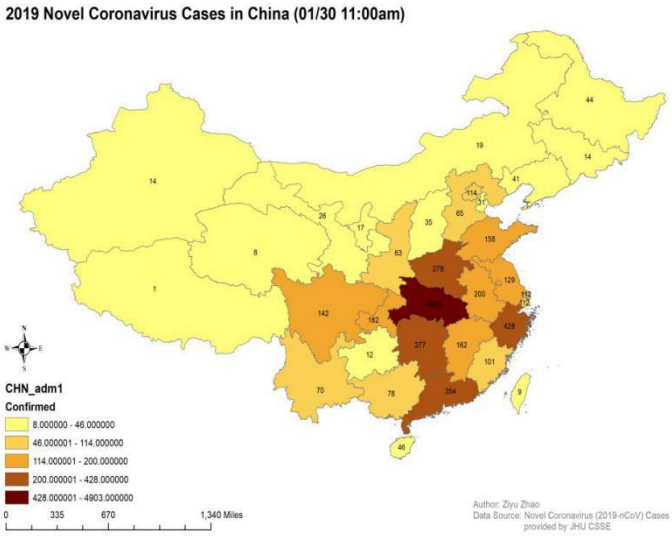
This series of maps illustrates the number and death/recovery rates of patients diagnosed of 2019 Novel Coronavirus in China by 11:00 am (Beijing Time), Jan. 30th. By exploring different combinations of classification rules, symbols, and layers, the author aims at presenting the epidemic situation with the greatest clarity and as much as meaningful details as possible.
Data Source: Novel Coronavirus (2019-nCoV) Cases, provided by JHU CSSE

I. Depicting Number of Diagnoses with Color

The first groups of maps are built on two layers to present complete territory of China with areas in dispute such as Taiwan and Southern Tibet included according to people's general perception in China Mainland. The major difference that sets the maps apart from each other is the classification rules used to define the breaks. The criteria used for define breaks were natural breaks (Jenks), quantile, standard deviation, and manually set breaks.

According to the actual data, the ideal graphic representations of the distribution of 2019-nCoV confirmed cases should emphasize the following two features: (1) the distribution is highly centered in Wuhan, Hubei, the city where the disease first broke out; (2) illustrate different degrees of severity of the outbreak with simple, readable and intuitive classifications.

Map 1: Number of Confirmed Diagnoses (Color: Natural Break)

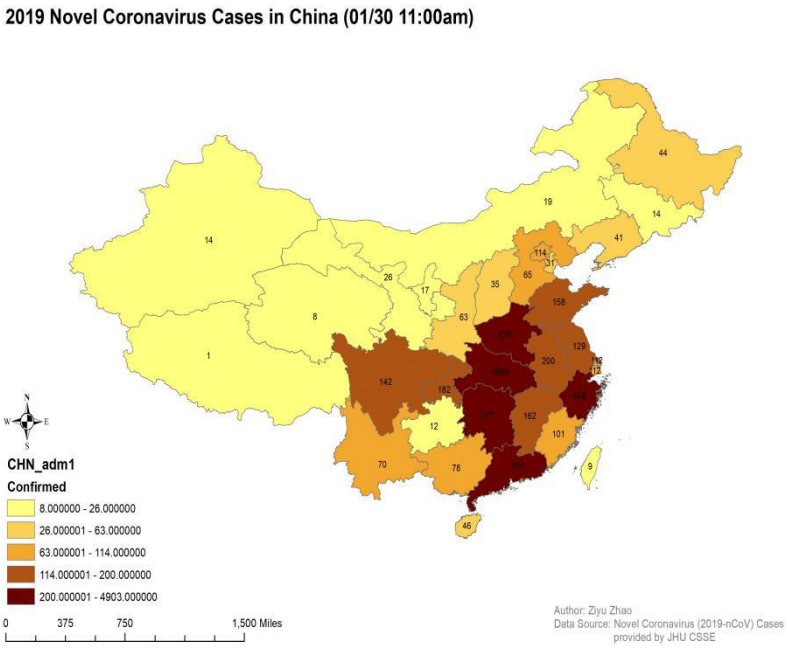


Using natural breaks (Jenks), the map shades different provinces distinctly with gradual yellow to brown while giving Hubei Province a proper emphasis with. The problem with this map is that irregular break margins and the sizes of groups are quite counter-intuitive, adding unnecessary difficulty to map reading and interpreting.

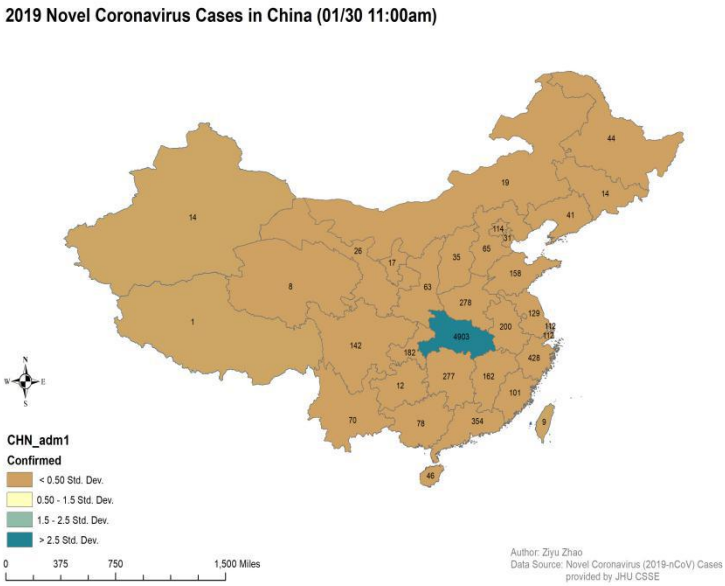
Map 2: Number of Confirmed Diagnoses (Color: Quantile)

The second map uses quantile as the classification method. Compared with the map with natural breaks, one can see a clear increase in the number of places shaded with dark brown, without differentiating between Hubei Province with 4903

cases and Henan Province with 278 cases. Statistically, quantile makes perfect sense here by classifying 278 and 4903 into one same group, given how small the lowest number of diagnoses is. However, in this particular case, quantile might not be the best option as it distorts the severity and may cause misinterpretations.



Map 3: Number of Confirmed Diagnoses (Color: Standard Deviation)



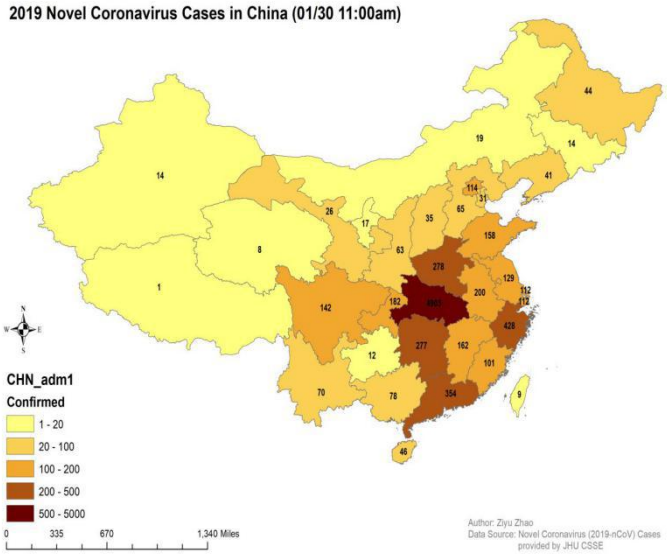
The standard deviation classification failed to present different levels of outbreak severity except for setting Hubei Province apart. While sd breaks emphasized Hubei, they oversimplify the case and are not presenting distributions in other areas detailedly enough.

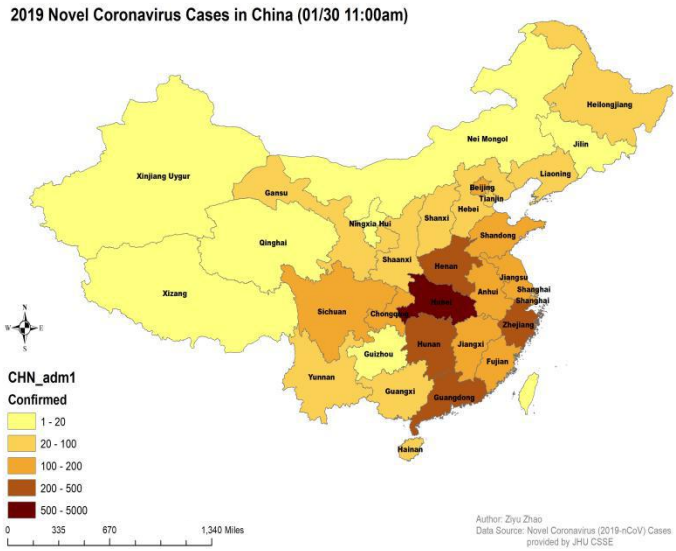
Map 4: Number of Confirmed Diagnoses (Color: Manual Break)

Given the above trying-outs, to fully present both wanted features and achieve the ideal effects, I set manual breaks based on natural breaks. Instead of using random numbers in certain ranges, I chose 10 multiples (such as 20) and 100 multiples (such as 100, 200, 500 and 5000) to set the breaks. The output map looks very similar to the natural break map, only with more readable legends and intuitive classification.

Whether label the map with numbers of diagnoses (left) or with names of the

areas (right) depends on whom the audience is, their familiarity with China's geographical administrative divisions and if they are concerned with numbers in detail or merely general trend of distribution.

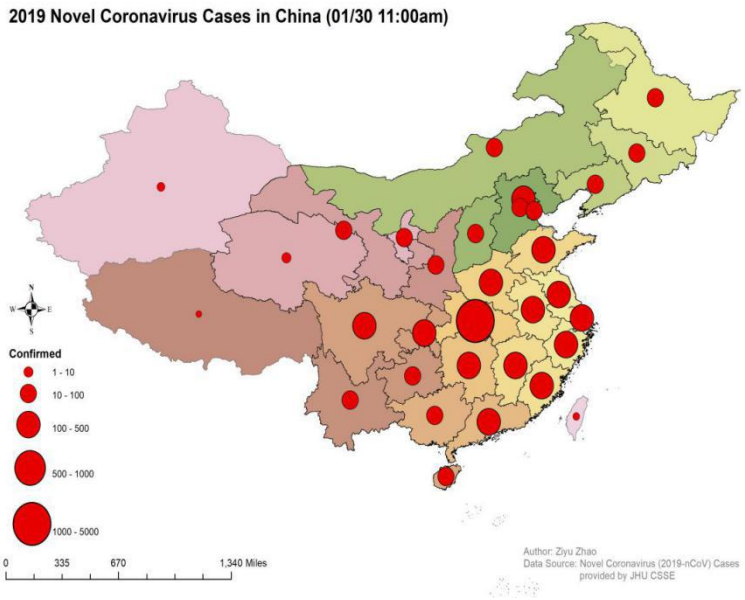




II. Depicting Number of Diagnoses with Symbols

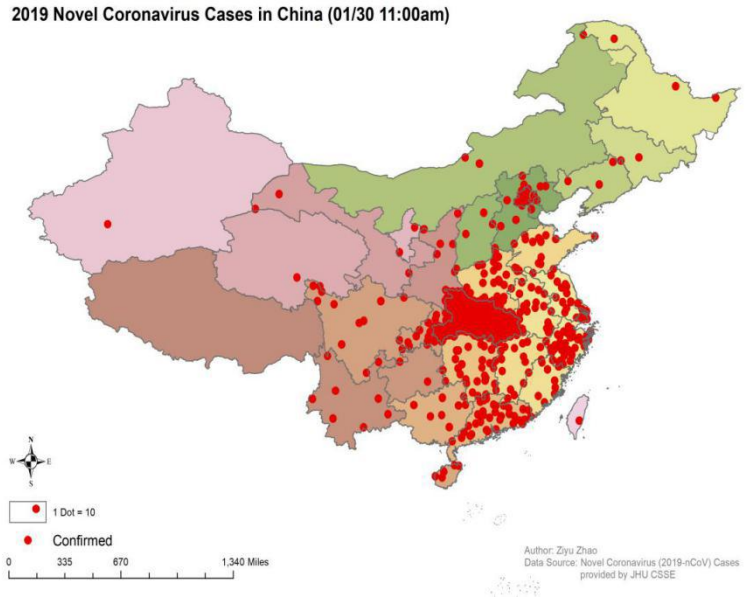
Except using color shades, an extra layer with gradual symbols or density dots can also serve the purpose of representing the number of diagnoses. In this case, the major concern will be whether the symbols can proportionally reflect the data and the spread of 2019-nCoV.

Map 5: Number of Confirmed Diagnoses (Gradual Symbols)



The gradual round red symbols in this map represent the numbers of confirmed cases of the new disease. While the symbol for 4903 in Hubei is larger than the rest symbols, its highlighting effect of the tenfold scale is relatively weaker than expected.

Map 6: Number of Confirmed Diagnoses (Dot Density)



The dot density plot, on the other hand, reflects the severity in Hubei Province with highly concentrated colored random dots that covered almost the entire area. The pattern of dots distribution also shows that the disease is more serious in the southeast area than in the northwest of China. And one can also read vaguely that the outbreak also sub-concentrated in Guangdong, Zhejiang Province, and Beijing. Overall, the expressiveness of maps using color shades outperforms maps using symbols in the case of 2019-nCoV spread in China.

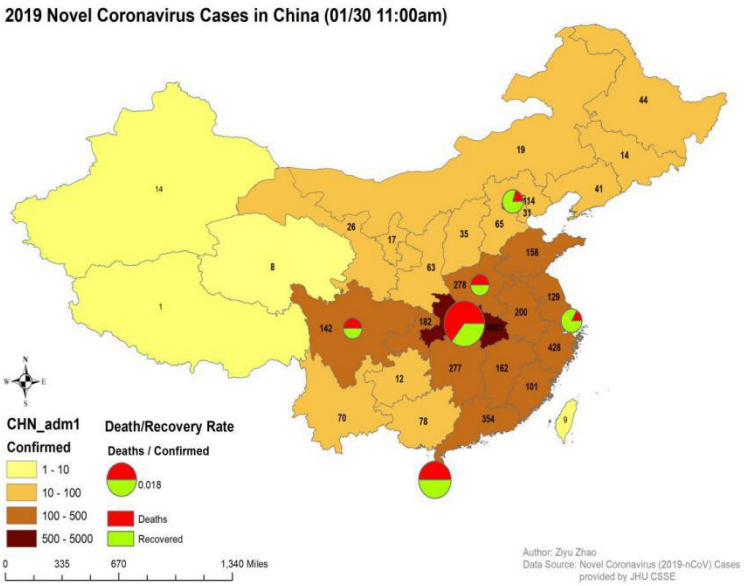
III. Depicting Death/Recovery Rates

Another important variable in the epidemic study is the death/recovery rate. Using an extra layer on top of color-shaded representation of numbers of confirmed diagnoses with manual breaks, I tried gradual symbols, pie charts and bar charts to investigate how death rate varies in different areas in China.

Map 7: Number of Confirmed Diagnoses & Death Rate (Gradual Symbols, Color: Manual Break)

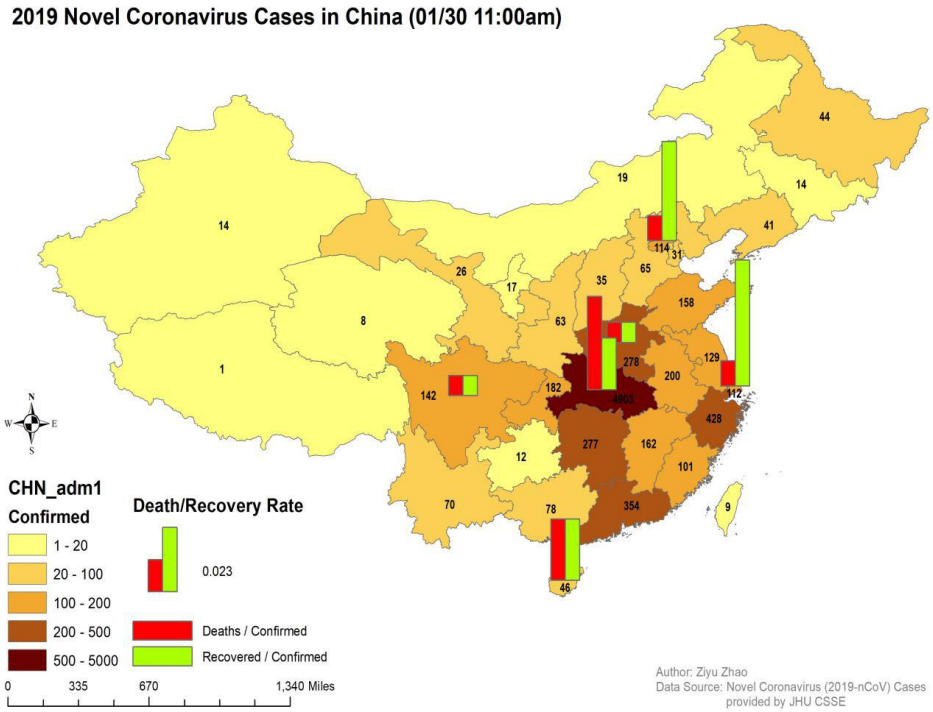
The red triangles in this map represent death rates, the number of death divided by the number of diagnoses. It is a bit unclear what the mortality rates are, except for 3 deaths occur among every 100 patients diagnosed with 2019-nCoV in Hubei.

Map 8: Number of Confirmed Diagnoses & Death Rate (Pie Chart, Color: Manual Break)



Map 8 depicts both death rate and recovery rate using pie charts as the red sectors represent death rates and the green sectors represent the recovery rate. The pie charts are also scaled proportionally according to the sum of deaths and recoveries. The shared problem of Map 7 and 8 is the symbols unclarity and difficulty to read, especially that the differences of sectors' center angles can be very subtle and unlikely to be observed by bare eyes. Also, the meaning of the sectors can be counter-intuitive and confusing without further explanation.

Map 9: Number of Confirmed Diagnoses & Death Rate (Bar Chart, Color: Manual Break)



Map 9 presents death and recovery rates with readable bar charts on an extra layer. The relative relationship, i.e. the comparison, between death rate and recovery rate is quite clearly presented with heights of bars. One can read that the recovery rate is higher in Beijing and Shanghai compared to Hubei Province. The two explanations are: (1) patients in Beijing and Shanghai have more medical resources and care per capita; (2) most of the people got infected in Wuhan and traveled from Hubei to Beijing and Shanghai are young adults, while infected population in Hubei are had more old and weak patients among them.

Personally speaking, Map 9 is the best illustration of both the spread (color shades with manual breaks) and regional death/recovery rates (bars) of 2019-nCoV by 01/30.

IV. Further Tasks

During the process of mapping 2019-nCoV, I found that there is still room for further improvement in several places. With more detailed data of different kinds, it is possible to draw more meaningful maps that might help understand the spread of 2019-nCoV and facilitate future decision making concerning resource distribution and allocation. Future work may include:

(1) Add labels for the top layer with symbols

(2) Include population data

(3) Map distribution of diagnoses to third (city) or fourth (community) administrative level, in regions like Hubei Province

(4) Map the needs of medical resources in infected areas

(5) Map medical resources that can be reassigned for the time needs

(6) Map donations to supervise the operation of sub-branches of Red Cross, such as Hubei Red Cross (which has been accused of poor conduct), and other organizations

(7) .....