**Spatial Analysis of Confirmed COVID-19 Cases in Mainland China Post-Wuhan Travel Quarantine**

Group 8: Christian Geofroy (1003193723), Gloria Chien Yi Lim (1003557741), Ziran Jeffrey Zhou (1002425477).

Instructor: Professor Jue Wang

GGR494H5 W - Special Topics in GIS

April 3, 2020

Winter 2020

**Abstract**

In December 2019, a strand of Coronavirus broke out in Wuhan City, Hubei China. This virus has since spread across the world and is classified as a pandemic by the World Health Organization. At the epicenter of the breakout in Wuhan, China, a strict quarantine was put in place on January 23 to contain and restrict the spread of the virus. Getis-Ord Gi\* hot spot analysis was performed on the number of confirmed cases from January 22 to March 7 2020 to determine which provinces in China became hotspots, and if there is any spatial relationship to the quarantine that was put in place in Wuhan, Hubei province, where the outbreak started. The provinces that have contiguity edges with Hubei did become hot spots which would suggest that the quarantine was effective in stopping the spread of the virus to the entire country and limiting it to the nearby province which are highly interconnected in terms of travel and business with each other.

Keywords: Covid-19, Quarantine, China, Hot Spot, Transmission

Contribution of each member:

Gloria Lim: Helped write introduction, methodology, create map & figures. Data pre-processing, report formatting and edits.

Christian Geofroy: Helped write abstract, results, presentation slides

Ziran Jeffrey Zhou: Helped write methodology, results, discussion. Data pre-processing, data processing and analysis, time-lapse animations, hot spot analysis.

**Introduction**

On January 30th, 2020, the World Health Organization (WHO) declared the Coronavirus disease (COVID-19) outbreak as a Public Health Emergency of International Concern (PHEIC) [1]. It first came to the attention of the WHO when a virus outbreak of an unknown cause was reported in Wuhan City, Hubei Province, China in December 2019, which led to a city-wide quarantine on the 23rd of January [2]. The earliest reported infected individuals showed symptoms as early as December 8th, where its origin was traced to the Wuhan South China Seafood Market [1]. This viral outbreak eventually became an international concern and was identified as a new coronavirus, which belong to a large family of viruses which causes a relatively less-severe disease such as the common cold, compared to severe diseases such as MERS and SARS [3]. While it was confirmed that the virus was spreading through human-to-human transmission, the historical spatio-temporal data on the COVID-19 confirmed cases can uncover the relationships between environment and health, taking into account factors such as demographic, environmental, behavioral, socioeconomic, genetic, and infectious risk factors [4]. The impact of an epidemic can be significant for a country especially with extended restrictions on movement as shown historically in SARS, MERS in RoK and Ebola in West Africa [5]. According to some analysts, it was estimated that the impact of this coronavirus outbreak could range from a 0.8% reduction of real GDP if the epidemic is controlled within 3 months, to a 1.9% reduction of GDP if the epidemic lasts for 9 months [5]. The investigation of the rate of transmission of this virus is important to provide insight on quarantine implications and estimate the duration it takes before it is safe to continue activities of trade and commerce, as well as inform strategies that can be taken to protect the economic conditions of a country. The aim of this study is to explore the spatial-temporal distribution of the confirmed coronavirus cases. The factors that will be taken into account include the daily reported number of confirmed cases in 31 provinces and quarantine measures, in order to investigate the process of the virus transmission within the provinces in Mainland China. The findings of this study will help identify the association between a travel quarantine implementation and the resulting rate of coronavirus transmission, which will provide insight as to whether the quarantine was an effective precaution.

**Methodology**



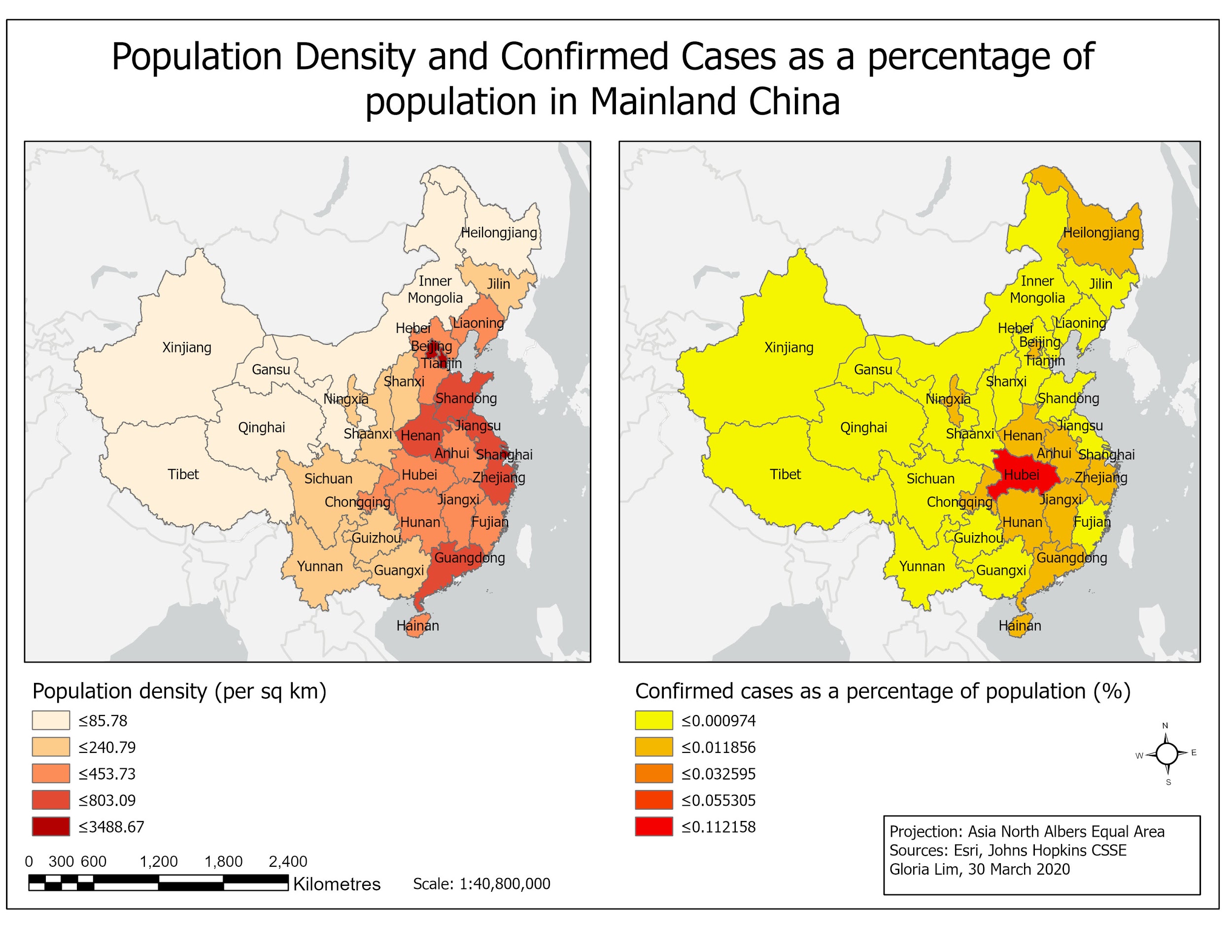
*Figure 1*: Flowchart of space-time data processing and analysis.

As shown in Figure 1, The data used in this analysis is obtained from the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE. The data is preprocessed in excel and was transposed to create a time column. A trendline formula using the logarithm function (LN) was generated in Microsoft Excel, which was used as a simple predictor model to predict the rate of new infections up to 60 days after March 7th, 2020. For example, the trendline for the province of Anhui was: y = 393.53ln(x) - 445.95, with x being the number of days since January 21. Therefore, to get the predicted value 30 days after March 7, x would be equal to 76, since March 7 is 46 days after January 21. Thus, y = 1258, or 1258 predicted cases. This was done for the 31 provinces in Mainland China, see attached file for all formulas. Following the data transposition in excel, we divided the day’s result with the previous day and then minus 1 (e.g. (Day 5/Day 4) - 1) to give us the percentage increase of new cases over the previous day. For example, Day 5 would have a 10% increase in cases over Day 4. The data was then transformed into points in ArcGIS, which was then joined with the China provinces shapefile obtained from ArcGIS Online.

Using the joined dataset, we created two time lapse animation. One shows the percentage increase from January 22nd, 2020 to March 7th, 2020 which contain actual recorded values. The second animation shows the predicted percentage increase up to 60 days past March 7th. If the percent increase over the previous day continues to be high, i.e. > 25%, then we can suggest that the quarantine measures were ineffective. However, if the percent increase is low i.e. < 25%, we can suggest that the quarantine measures had some effect in controlling the spread of the virus.

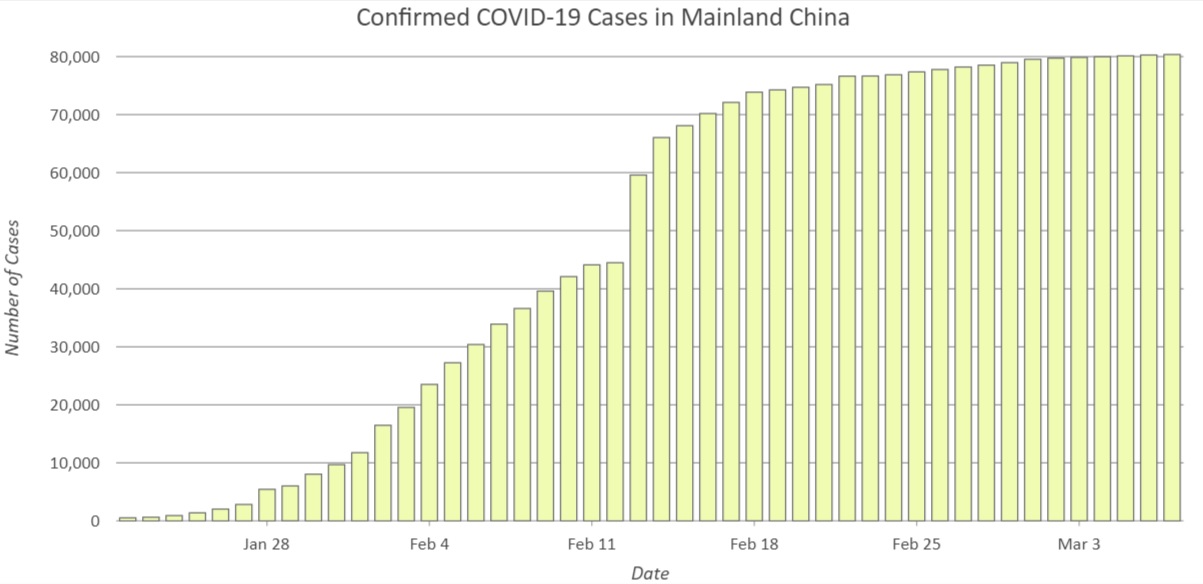
The prediction results can be used to reaffirm whichever result is produced in the previous step.

With the joined dataset, a hot spot analysis via the Getis-Ord Gi\* method was performed using the number of confirmed cases and the conceptualization of spatial relationships set to contiguity edges only. The reason for contiguity edges selection is due to the ability to consistently travel in modern China, especially for work [6]. Therefore, we believe that Covid-19 will spread faster to provinces that are contiguous with Hubei (province where Wuhan city is located) than non-contiguous provinces. The hot spot analysis will serve to determine the effectiveness of the quarantine by determining hot and cold provinces. If hot spots exist in provinces adjacent to Hubei, but not in non-adjacent provinces, we can argue that the quarantine was effective. This is because due to the nature of Tobler’s law [7], near things will be more affected than farther things, as well as the snowball effect, the surrounding provinces will likely see hot spots. However, if non adjacent provinces remain non significant or cold, we can argue that the quarantine was effective.

**Results**

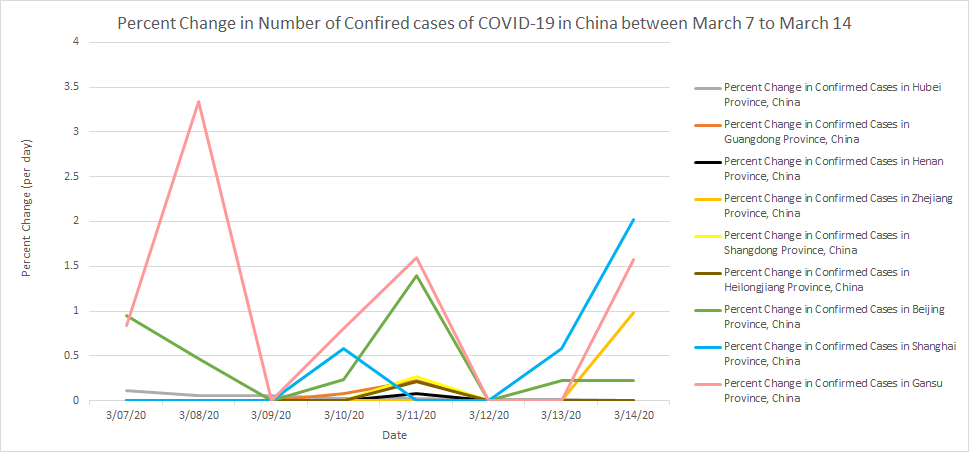
*Figure 2*: Map of population density and percentage of confirmed cases based on population within the province in mainland China both classified with natural breaks consisting of 5 classes.

Figure 2 shows the population density in mainland China and the number of COVID-19 confirmed cases as a percentage of the population in each province. While the population density is greater in the eastern part of China, it is observed that there is a greater percentage of confirmed cases in provinces that are contiguous to Hubei, the epicentre of the outbreak. However, it is not clear if the number of confirmed cases is related to the population density, additional information on the local level will be required to investigate the rate of transmission of COVID-19 in populated areas.

****

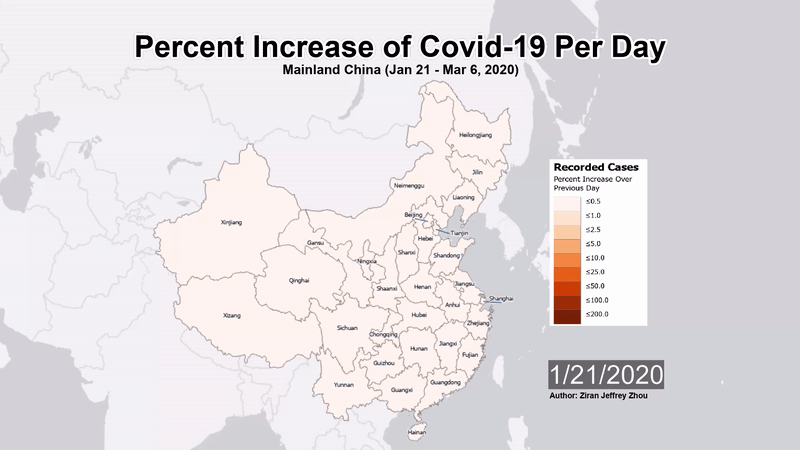
*Figure 3:* Bar chart showing the number of confirmed COVID-19 cases in Mainland China between January 22 to March 7, 2020.

Figure 3 shows the sum of COVID-19 confirmed cases in mainland China over the course of the 46-day period. It can be observed that there was a rapid increase in confirmed cases between January 22 and February 18, after which the curve appears to flatten following the end of February. There is also a notably sharp increase in numbers between February 13 and 14 where the number of confirmed cases rose from around 44,699 to 59,832 cases overnight.

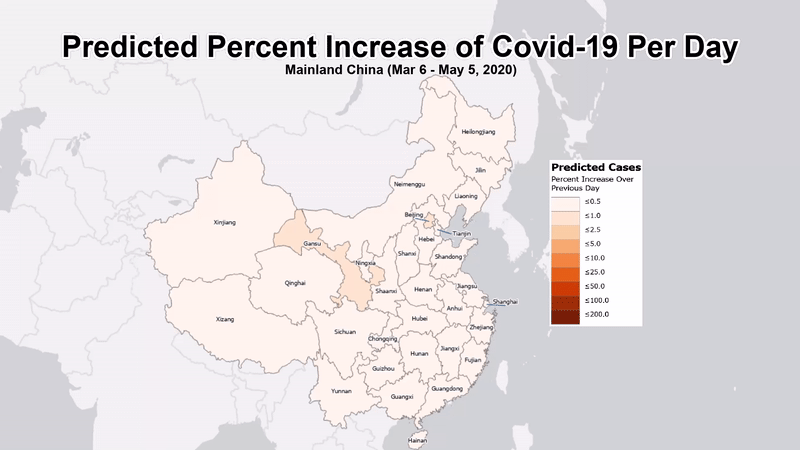
****

*Figure 4:* Line graph showing the actual percent change in number of cases compared to predicted percent change in the animation - March 7 to March 14, 2020

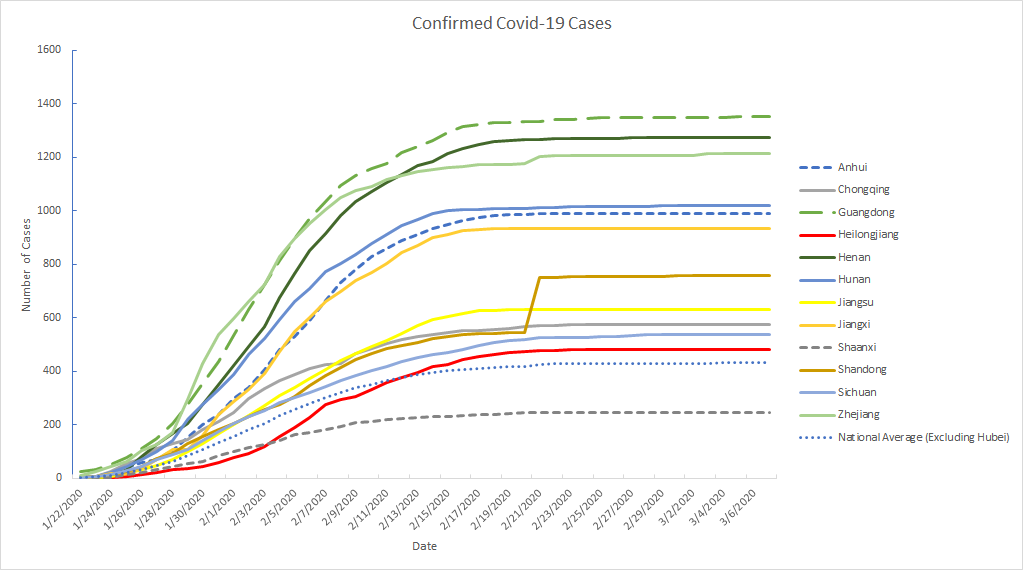
Figure 4 shows the actual percent increases for the number of confirmed cases of COVID-19 one week past our animation. This data uses the John Hopkins dataset of confirmed cases and shows the actual data compared to the animation that was created which was only a prediction based off of the data from January 22 to March 7. Note that the percent change per day is still extremely low compared to the weeks in February. The time lapse animation from January 22 to March 7 (Figure 5) shows a percentage increase of no more than 10% for the majority of provinces after February 15. Additionally, with the exception of Gansu, there are no percentage increases greater than 5% after February 21. The prediction animation (Figure 6) suggests that there will be no further percentage increases greater than 25% after March 7 if the current conditions remain the same. In the hot spot analysis using Getis-Ord Gi\* (Figure 7), hot spots with 99% confidence exist in the province of Hubei and all provinces that are contiguous to it. Non-contiguous provinces are either non-significant or have varying confidences of cold spots.



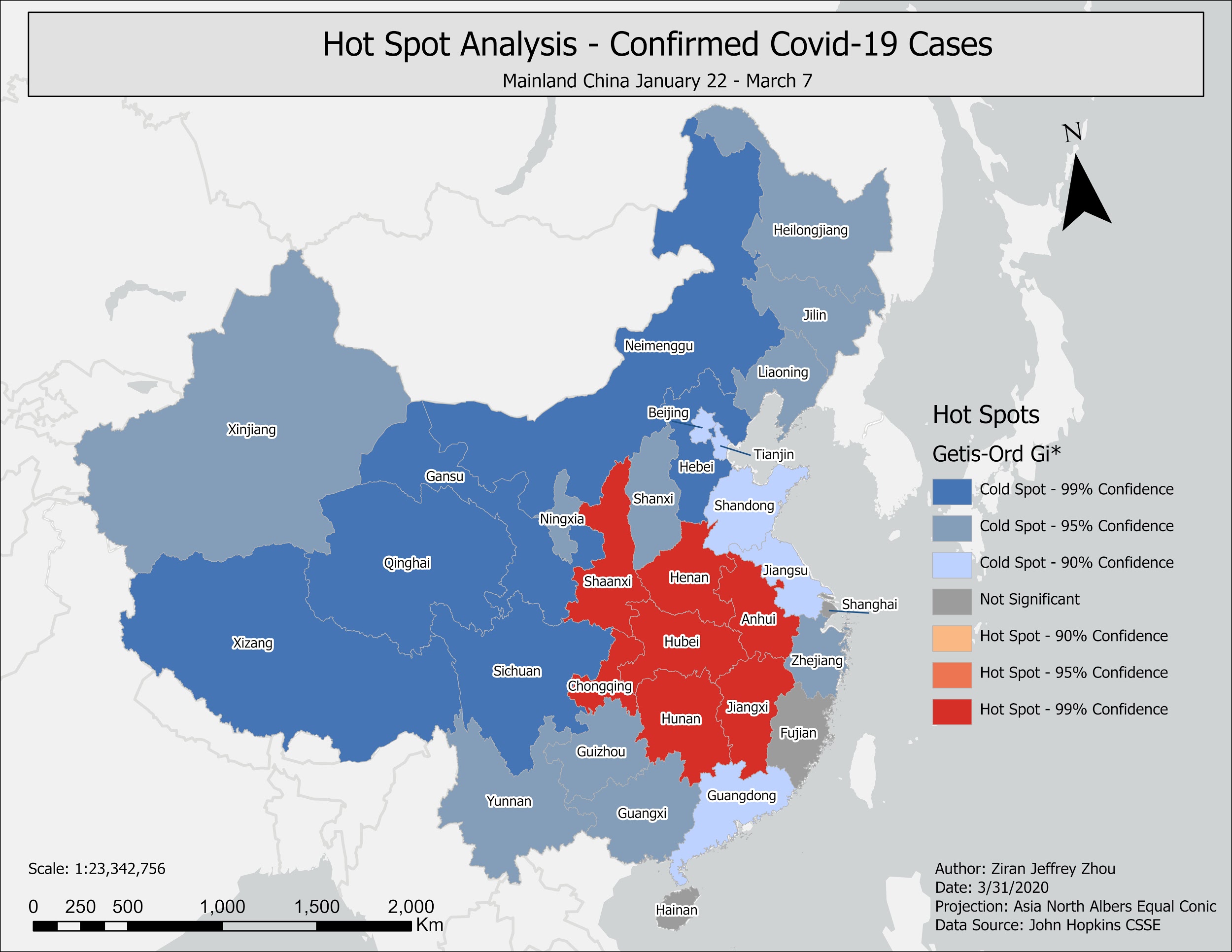
*Figure 5: Gif of a time lapse animation of percent increase of COvid-19 cases per day, from January 22 to March 7. For higher resolution, see attached files.*



*Figure 6: Gif of a time lapse animation of predicted percent increase of Covid-19 cases per day, from March 6 to May 5. For higher resolution, see attached files.*

**

*Figure 5: Graph of confirmed Covid 19 cases of provinces above the national average (excluding Hubei) as well as provinces indicated as hot spots.*

**

*Figure 7: Hot spot analysis of confirmed Covid-19 cases in Mainland China.*

**Discussion & Conclusion**

Based on the trend seen in the percent increases per day, we can suggest that the travel quarantine negatively affected the rate of transmission in some capacity as the percentage of new cases decreased dramatically following February 21. Due to the restrictions of movement, there is a lower risk of contact between healthy individuals and individuals carrying Covid-19, which greatly hinders the spread of the virus. If conditions remain the same, or at least similar to the periods between January 22 and March 7, the prediction model suggests that there will be no large increases in the amount of new Covid-19 cases. Additionally, Covid-19 hot spots appeared in Hubei and all provinces contiguous to it, while non-contiguous provinces were either not significant or cold spots. This suggests that due to the close proximity to Hubei, Covid-19 was spread much faster, and caused a snowball effect in its neighbouring provinces that could not be stopped. However, even though Covid-19 managed to spread to closeby provinces in significant capacity, the quarantine enacted was effective in preventing further mass transmission to even more provinces, as no further hot spots appeared. To further support this claim, Figure 5 shows that only six provinces deemed non-significant or in cold spots are above the national average in the number of Covid-19 Cases. The provinces are Guangdong, Heilongjiang, Jiangsu, Shandong, Sichuan, and Zhejiang. Meanwhile all of the neighbouring provinces to Hubei with the exception of Shaanxi are above the national average in number of Covid-19 cases and are hot spots. In summary, the travel quarantine imposed was an effective measure in preventing further large outbreaks in non-contiguous provinces.

The limitations identified in this study is that the numbers might be under-reported. However, the 2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE has been proven to be the most reliable source as it draws its data from various databases such as the Centre for Disease Control and Prevention (CDC), European CDC (ECDC), China’s National Health Commission (NHC), as well as state and national government health departments. Regardless, results are based on both the amount of tests done per day, availability of test kits, the integrity of reports, as well as people going to get tested. There is also the possibility of asymptomatic individuals not being tested and accounted for. Furthermore, we could not account for the travel patterns of major cities such as Guangzhou and Shanghai, where people commute from further away to work [6], thus potentially spreading Covid-19 further. Future research should account for this factor, which may explain why cold spot provinces have their number of cases above the national average.

**References**

[1] Park, M., Thwaites, R. S., & Openshaw, P. J. M. (2020). COVID-19: Lessons from SARS and MERS. *European Journal of Immunology*, 50(3), 308–311. <https://doi.org/10.1002/eji.202070035>

[2] ncov – CSSE. (2020). Systems.jhu.edu. Retrieved 9 March 2020, from <https://systems.jhu.edu/research/public-health/ncov/>

[3] He, F., Deng, Y., & Li, W. (2020). Coronavirus Disease 2019 (COVID‐19): What we know? *Journal of Medical Virology*, (March), 1–7. <https://doi.org/10.1002/jmv.25766>

[4] Elliott, P., & Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, 112(9), 998–1006. <https://doi.org/10.1289/ehp.6735>

[5] McCloskey, B., & Heymann, D. L. (2020). SARS to novel coronavirus - Old lessons and new lessons. *Epidemiology and Infection*, 1–4. <https://doi.org/10.1017/S0950268820000254>

[6] Garske, T., Yu, H., Peng, Z., Ye, M., Zhou, H., & Cheng, X. et al. (2011). Correction: Travel Patterns in China. *Plos ONE*, *6*(2). doi: 10.1371/annotation/e4781265-2e4f-45a9-85b7-1930e0a16171

[7] What is Tobler's First Law of Geography? - GIS Geography. (2020). Retrieved 30 March 2020, from <https://gisgeography.com/tobler-first-law-of-geography/>