**An analysis of rate of infection and cause of Coronavirus:**

**Author: John Zhang**

**3/13/2020**

Background:

The 2019-nCoV (coronavirus disease 2019), is a respiratory disease that was first detected in the Hubei Province of China. The beginning of the outbreak started approximately in December 2019. However, since its inception it has spread to more than 100 locations internationally, including the United States. 1 The disease,COVID-19 is believed to spread from human to human contact. This transmission does not have to be through physical contact, for example, droplets produced when an infected person coughs or sneezes may make their way through the air and into another host’s respiratory tract. Because the virus is both transmitted through the air and through physical contact, its infection rate is relatively high. In last two months, COVID-19 has jumped continents from China to Europe, and from there to now North and South America infecting enough people for the World Health Organization (WHO) to declare it a global “pandemic”. While COVID-19 is considered a pandemic, it affects different demographics of people differently. The most meaningful breakdown of the demographics of people affected is through the factor age and gender.

Abstract:

Using R visualization and packages including (ggplot, sf, tidyverse, and map data), this paper will attempt to conceptualize the spread of COVID-19 throughout China and the United States. Furthermore, the visualizations will also analyze the breakdown of affect between different age demographics.

Visualizations:

**China Visualization**

**-------------------------------------------------------------------------------------------------------------------------------**

A close up of a map

Description automatically generated

1. From Zhen Zhang’s code base, this visualization utilizes R’s Shapefile, “bou2\_4p.shp” to depict China’s population density in regard to location. As depicted by this visualization, China’s most heavily populated areas are in the south east regions. This makes sense as China’s capital, Beijing is located there and also other big cities.

A close up of a map

Description automatically generated

* COVID-19 originated in the Hubei Province, Hubei is the ninth most popualted region in China. Consisting of Wuhan, Wuchang, Hanyang, Xinzhou, and Huangpi with a population of approximately 57,237,7407.

A close up of a map

Description automatically generated

* Graph visualizes the region of Hubei. Starting from the first recorded observation. (Jan 22, 2020). In 50 days, Hubei’s Province had over 60,000 people confirmed with the virus.

**First Observations Recorded (Jan 22, 2020) (Approximately 1 month into outbreak)**

-------------------------------------------------------------------------------------------------------------------------------

A screenshot of a cell phone

Description automatically generated

* Starting from the first recorded observation, we see that the Hubei Province already had the most recorded confirmed cases. However, the rest of the regions were approximately around 0 – 30 confirmed positives.
* To check the spread of the virus without the axis being influenced by the epicenter of the outbreak we create a similar visualization without the Hubei Province.

A screenshot of a cell phone

Description automatically generated

* Without the Hubei Region we see confirm that indeed, that most of the provinces have only recorded between 0 – 10 counts of confirmed cases, however there were some provinces already in the 10 – 30 range.

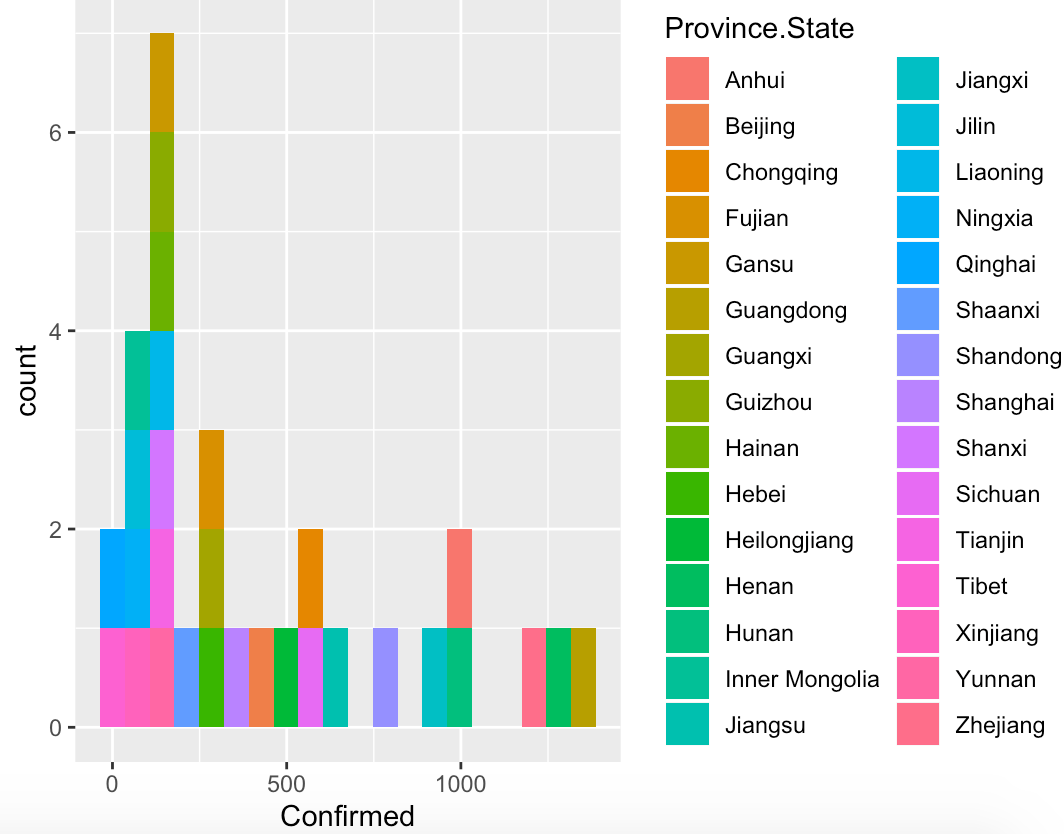
**Last Recorded Observations Recorded (Mar 11, 2020) (Approximately 3 months into outbreak)**

-------------------------------------------------------------------------------------------------------------------------------

A screenshot of a cell phone

Description automatically generated

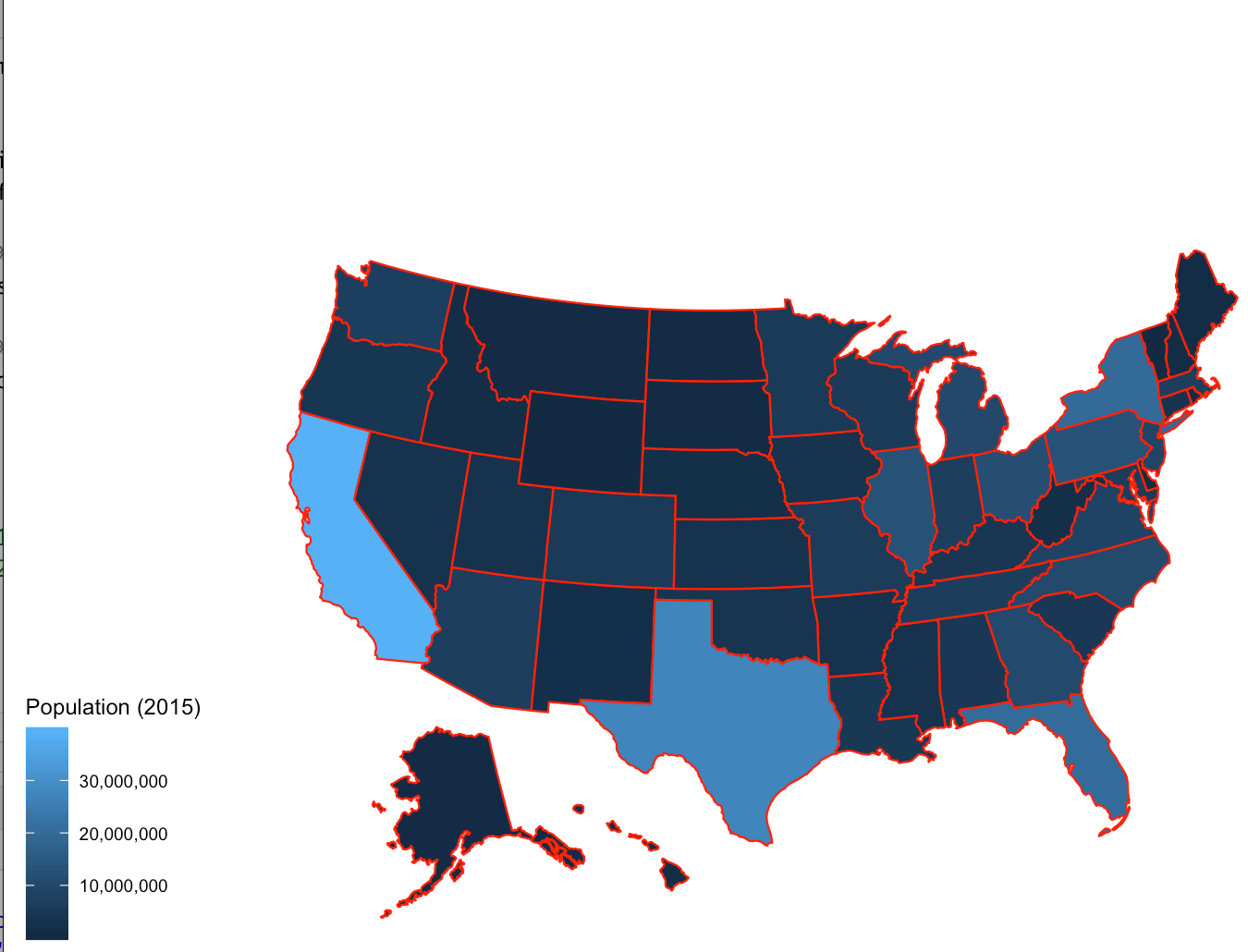
* From this visualization we see that because of the range of confirmed cases, that Hubei as a province has by far contracted the most cases in all of China.
* To visualize what the variance of confirmed sickness in China without Hubei we produce a second visualization.



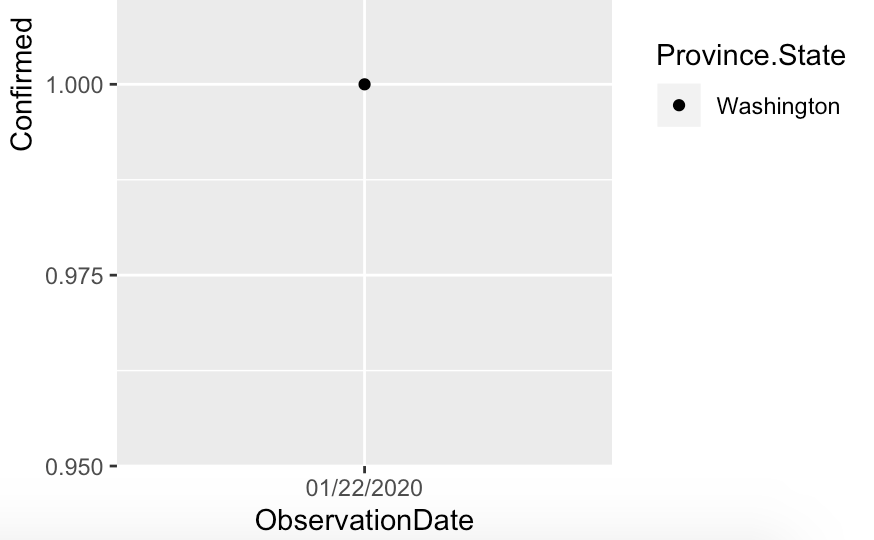
* This visualization shows that within two months, provinces that in the first month had 0 to 10 cases have on average increased to a range between 100 – 500 with many other provinces in the 500 - 1000 range.

**United States Visualization**

**-------------------------------------------------------------------------------------------------------------------------------**



* Population Density of United States.



* On Jan 22, 2020 there was one recorded case in the United States

A screenshot of a video game

Description automatically generated

* By March 10, the virus had spread from one state, Washington to about every US state, with some states being hit the hardest from the virus.

A screenshot of a cell phone

Description automatically generated

* COVID-19 First hit Washington (State) In 2020. Within 3 months the virus went from 10 cases, to over 500. This State has the most recorded confirmed cases.

A screenshot of a cell phone

Description automatically generated

* COVID-19 then spread to California, within two months there were over 200 cases. This state has the second most cases.

A screenshot of a cell phone

Description automatically generated

**DEMOGRAPHICS: 1/23/2020 (China)**

**-------------------------------------------------------------------------------------------------------------------------------**

1/23/2020

A screenshot of a cell phone

Description automatically generated

1/23/2020

A screenshot of a cell phone

Description automatically generated

a

1/23/2020

A screenshot of a cell phone

Description automatically generated

Discussion:   
China:

When analyzing the spread of COVID-19, we can compare and contrast insights from how it spread in China to how the future of the infection will spread in the United States. The epicenter of the disease started in the province Hubei, in the city of Wuhan. Hubei is China’s ninth most populated province with a standing population of approximately 57 million residents.7 Approximately one month into the outbreak, surrounding provinces of Hubei had confirmed cases in the range of 10-30 patients. However, Hubei itself, had more than 400 confirmed cases within the first month. Within 50 days, Hubei’s confirmed cases went from approximately 400 cases to around 60,000 cases. The following growth, as depicted in the graphs looks to follow a logarithmic pattern. Suggesting that while the virus in the early stages is able to infect many people, at some point there are factors that limit its further spread. In China, factors that limited the further spread of the virus included; Governmental lockdown of major cities, social distancing, and quarantining large areas simultaneously.

US:

Following its spread in mainland China, corona virus had been spread to the United States. The first state to have a confirmed case being Washington with population 7.536 million. Within two months, on Washington went from between 0-5 confirmed cases to over 300. Following Washington, California then was infected by the outbreak. Its cases growing to over 200 in a few weeks. Surprisingly, while California is in closer physical proximity to Washington State, New York is the second most infected State, while California is third. As evidenced by the New York graph, in the span of one day, the virus had spread from approximately 150 cases to over approximately 230 cases. Corona virus is relatively new to the United States, the first case being reported around January. In comparison by the same time, China had already experienced the outbreak for a month. Therefore, while it is impossible to predict the future spread of the outbreak, we can expect to see a rise in confirmed cases within the coming weeks.

Demographics:

The demographics show that, besides in the youngest age group, from 0 – 10 that the virus has no bias in its infection of male or female, as the proportions for gender split pretty evenly in all other age groups.

Limitations:

Due to the dynamic nature of the virus, by the time this report is finished the numbers will almost certainly changed. The infection rate of this virus is high, more so, there are limitations in testing for this virus, therefore these numbers are under-representative of the true infection number. In the United States, there is a shortage of testing kits, therefore reasonably there is going to be missing and under-representative reports coming in.

Conclusion:

The nature of the virus is highly dynamic. As because the virus may be transmitted through the air and through physical contact many people will become infected in the coming weeks. There are many different factors associated with susceptibility and infection rate however the virus is spreading in a very fast pace. With the new restrictions on travel and other preventative measures being introduced to society, the only way to see if the virus is going to be tapered off is with time and future observation.

References:

1. *2019 Novel Coronavirus (2019-NCoV) Situation Summary.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 11 Mar. 2020,* [*www.cdc.gov/coronavirus/2019-ncov/summary.html*](http://www.cdc.gov/coronavirus/2019-ncov/summary.html)*.*
2. Zhang, Zhen. “R Map Part 2 -- Ggplot.” *Zhen Zhang's Blog*, 16 Oct. 2015, imzhen.com/note/2015/09/16/20150916/.
3. *Lecture 11: Spatial Methods in R*, [www.nathanschiff.com/webdocs/grad\_urban/grad\_urban\_2017/spatial\_methods\_lecture.nb.html](http://www.nathanschiff.com/webdocs/grad_urban/grad_urban_2017/spatial_methods_lecture.nb.html).
4. Srk. “Novel Corona Virus 2019 Dataset.” *Kaggle*, 12 Mar. 2020, [www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset](http://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset).
5. Lorenzo, Paolo Di. “US Maps Including Alaska and Hawaii [R Package Usmap Version 0.5.0].” *The Comprehensive R Archive Network*, Comprehensive R Archive Network (CRAN), cran.r-project.org/web/packages/usmap/.
6. Srk. “Novel Corona Virus 2019 Dataset.” *Kaggle*, 12 Mar. 2020, www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset#time\_series\_covid\_19\_recovered.csv.
7. “Asia.” *Hubei (China): Prefectural Division, Major Cities & Counties - Population Statistics, Maps, Charts, Weather and Web Information*, www.citypopulation.de/en/china/cities/hubei/.