An exploration into the origin and spread of COVID-19:
Author: John Zhang
3/13/2020

Background:

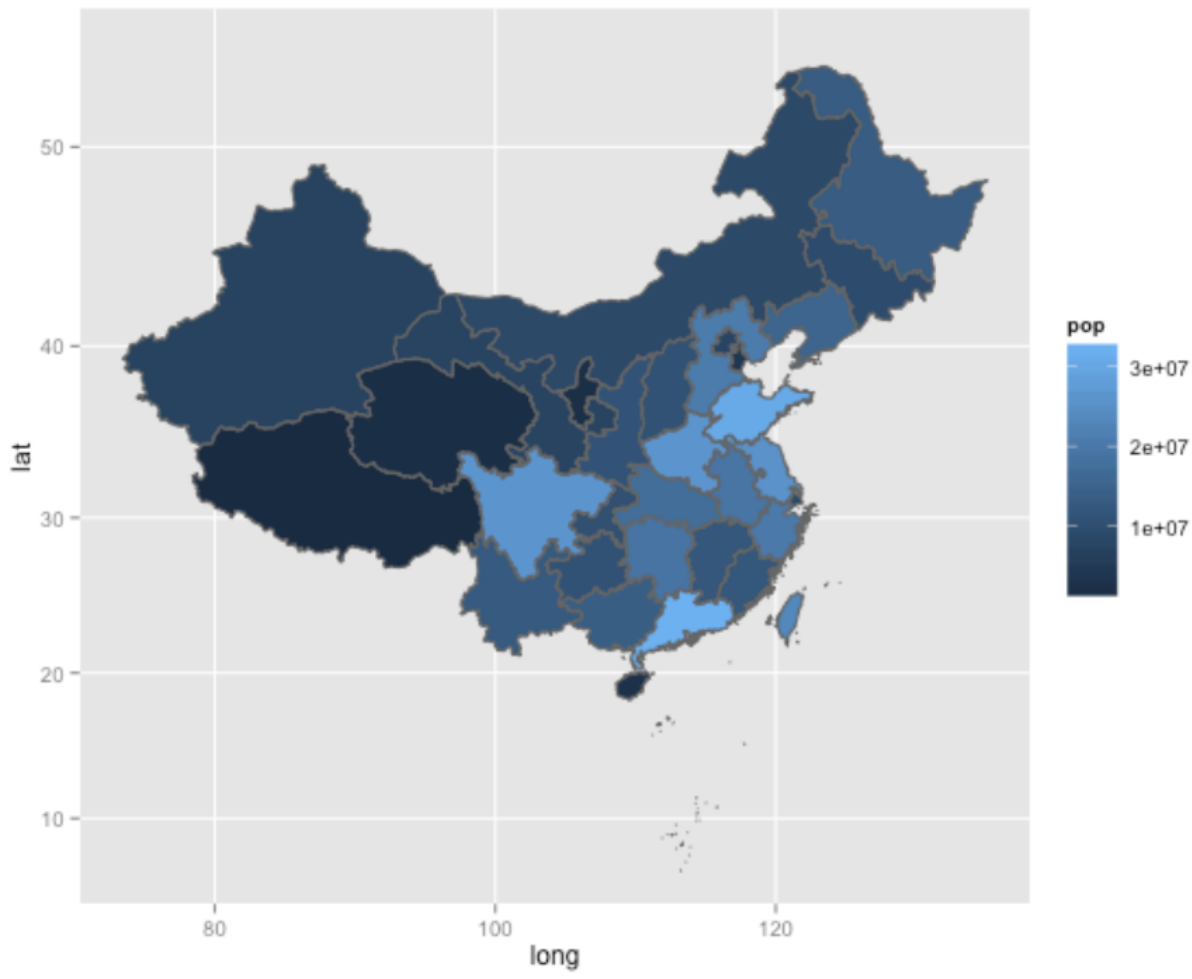
2019-nCoV (COVID-19), is a respiratory disease that was first detected in the Hubei Province of China. The epicenter experienced an outbreak in December 2019, of the outbreak started approximately in December 2019. Since its inception, it has spread to more than 100 locations internationally, including the United States. [1] The virus, COVID-19, spreads from human to human contact. This transmission does not have to be through physical contact. For example, droplets produced by an infected person's coughs or sneezes may be aerosolized and travel 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 the last two months (February and March of 2020) , COVID-19 has jumped continents from China to Europe, and from there to North and South America. The global outbreak of the disease raised its classification from epidemic to pandemic on March 11, 2020. While COVID-19 is considered a pandemic, it affects different demographics of people differently. Two important factors that should be considered are 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 susceptibility between different age and gender demographics.

Visualizations:

China Visualization

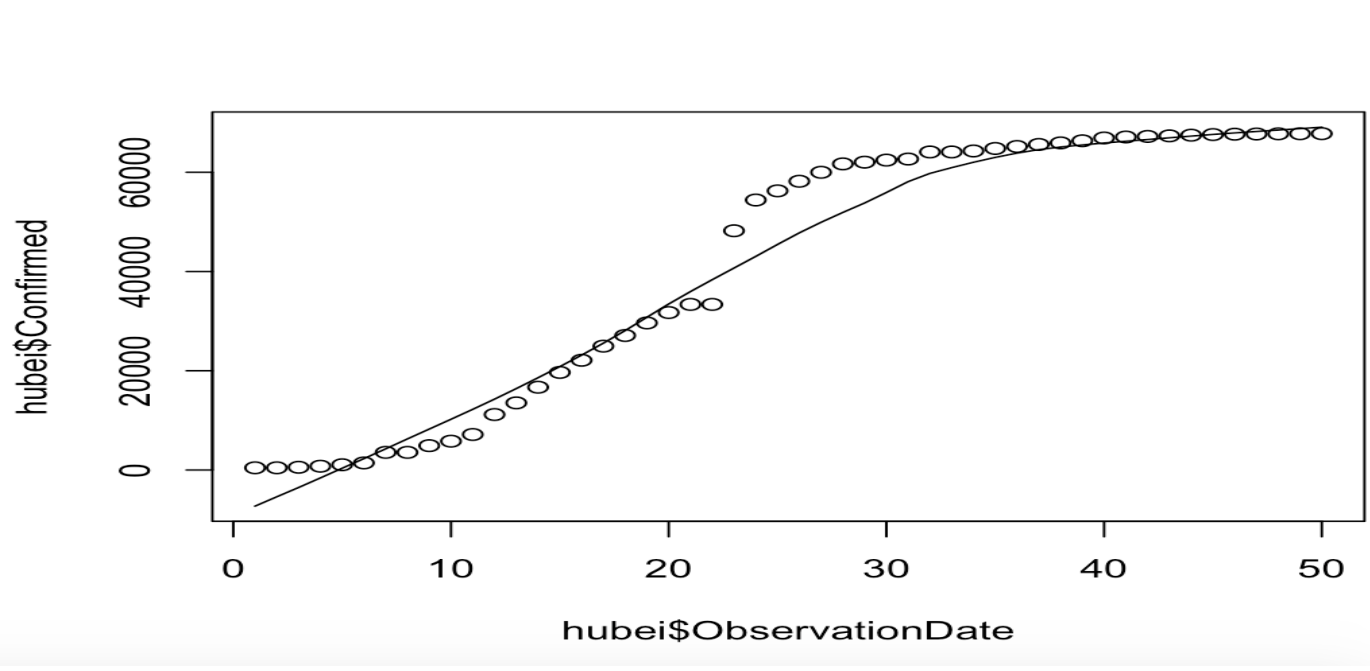


(1A)

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.

(1B)

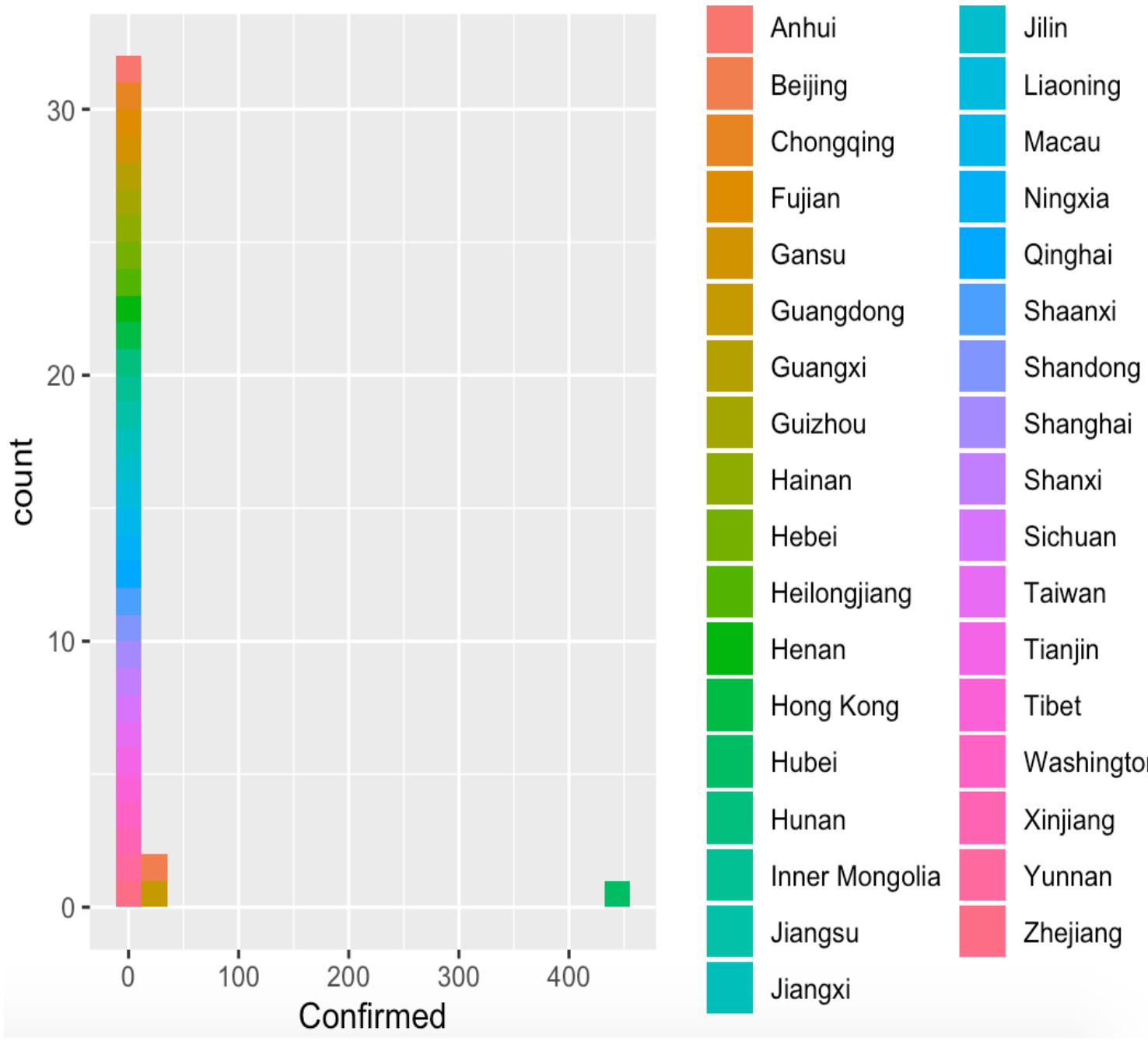
- COVID-19 originated in the Hubei Province, Hubei is the ninth most populated region in China. Consisting of Wuhan, Wuchang, Hanyang, Xinzhou, and Huangpi with a population of approximately 57,237,740[7].



1(C)

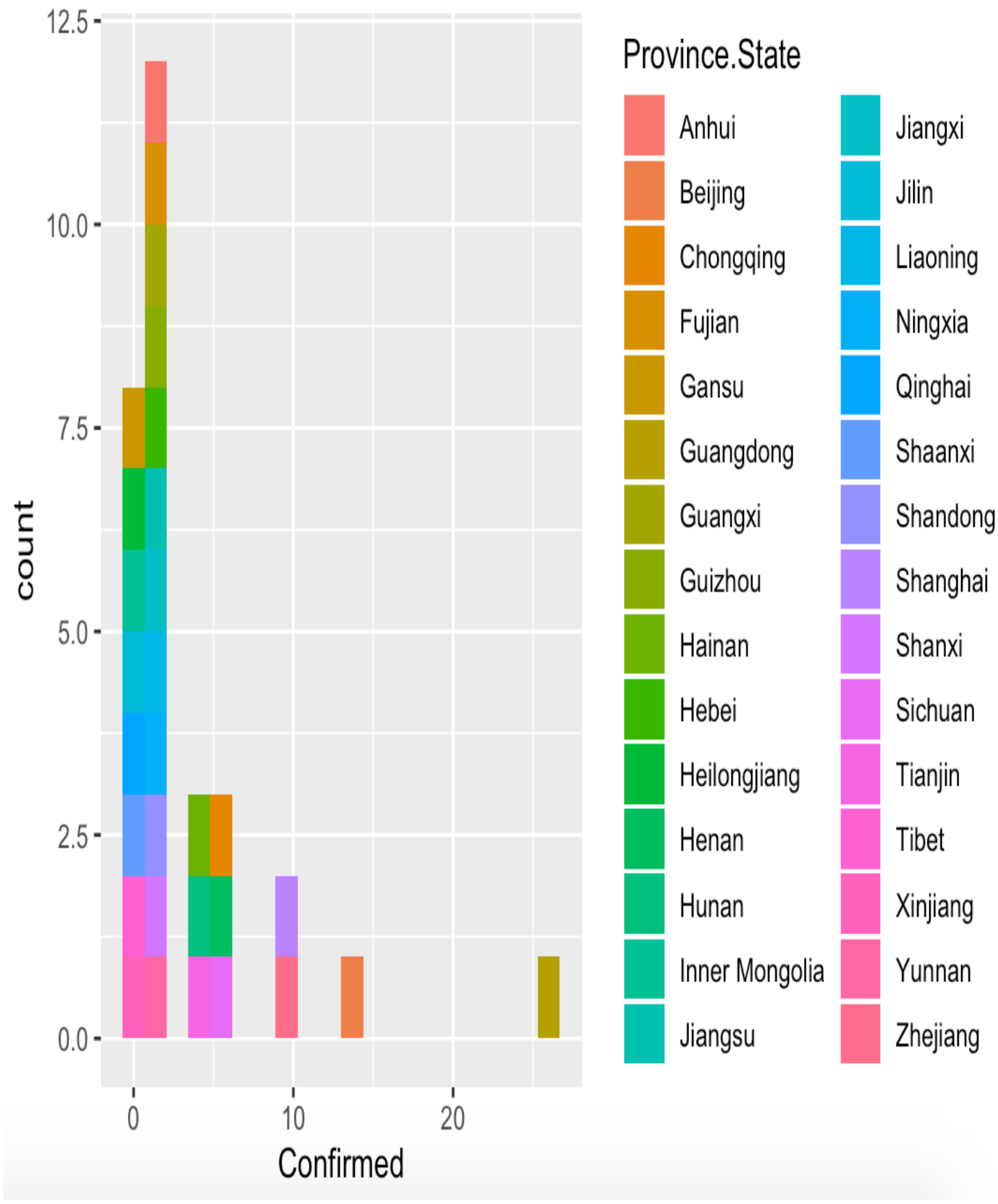
- Graph visualizes the number of cases reported in the region of Hubei. Since the first recorded observation. (Jan 22, 2020). In 50 days, Hubei's Province had over 60,000 people confirmed with the virus.

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



1(D)

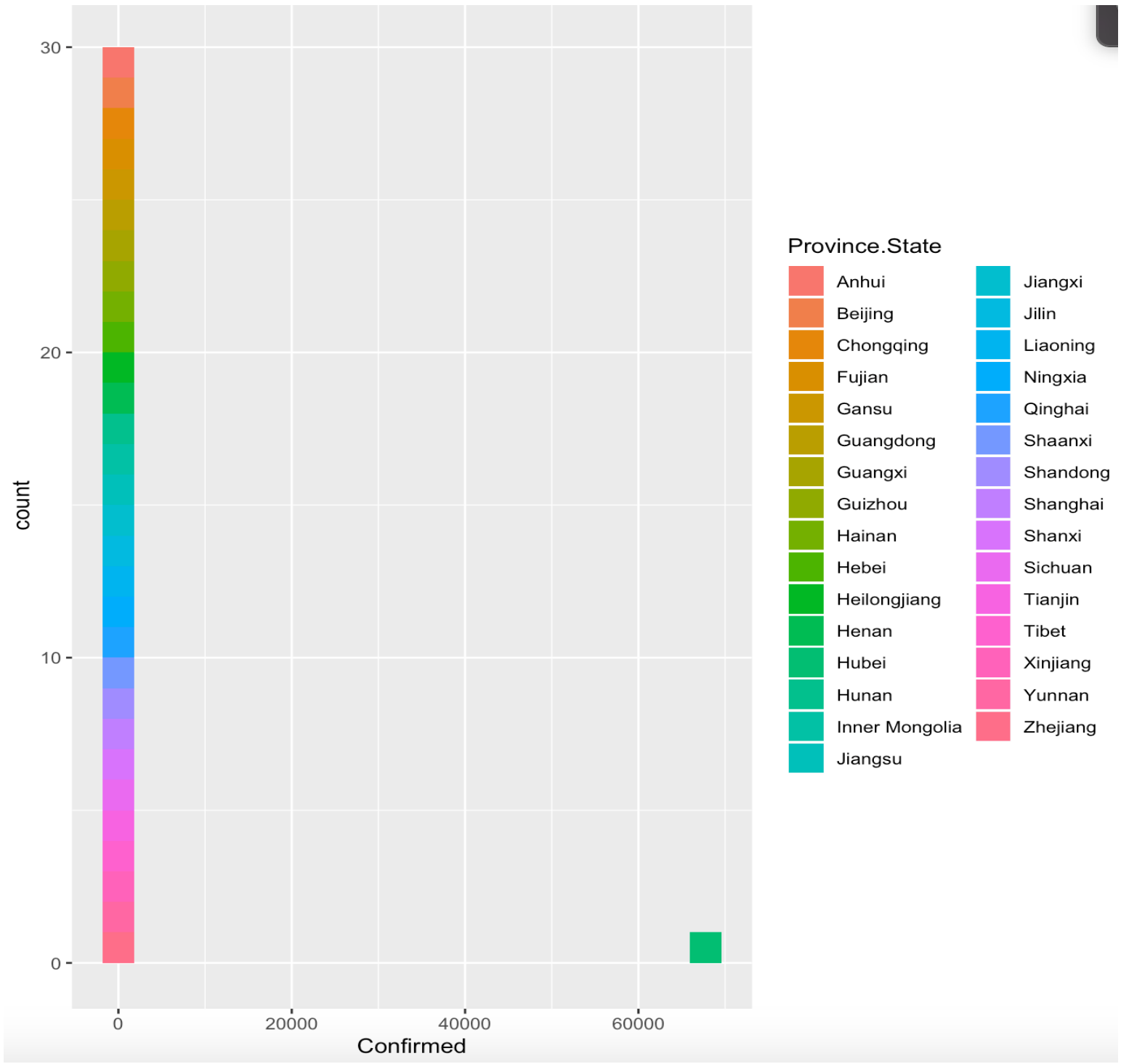
- Starting from the first recorded observation, we see that the Hubei Province already had the most recorded confirmed cases. However, the rest of the regions had approximately around 0 – 30 confirmed positives.
- To check the spread of the virus without the axis being influenced by the epicenter of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



1(E)

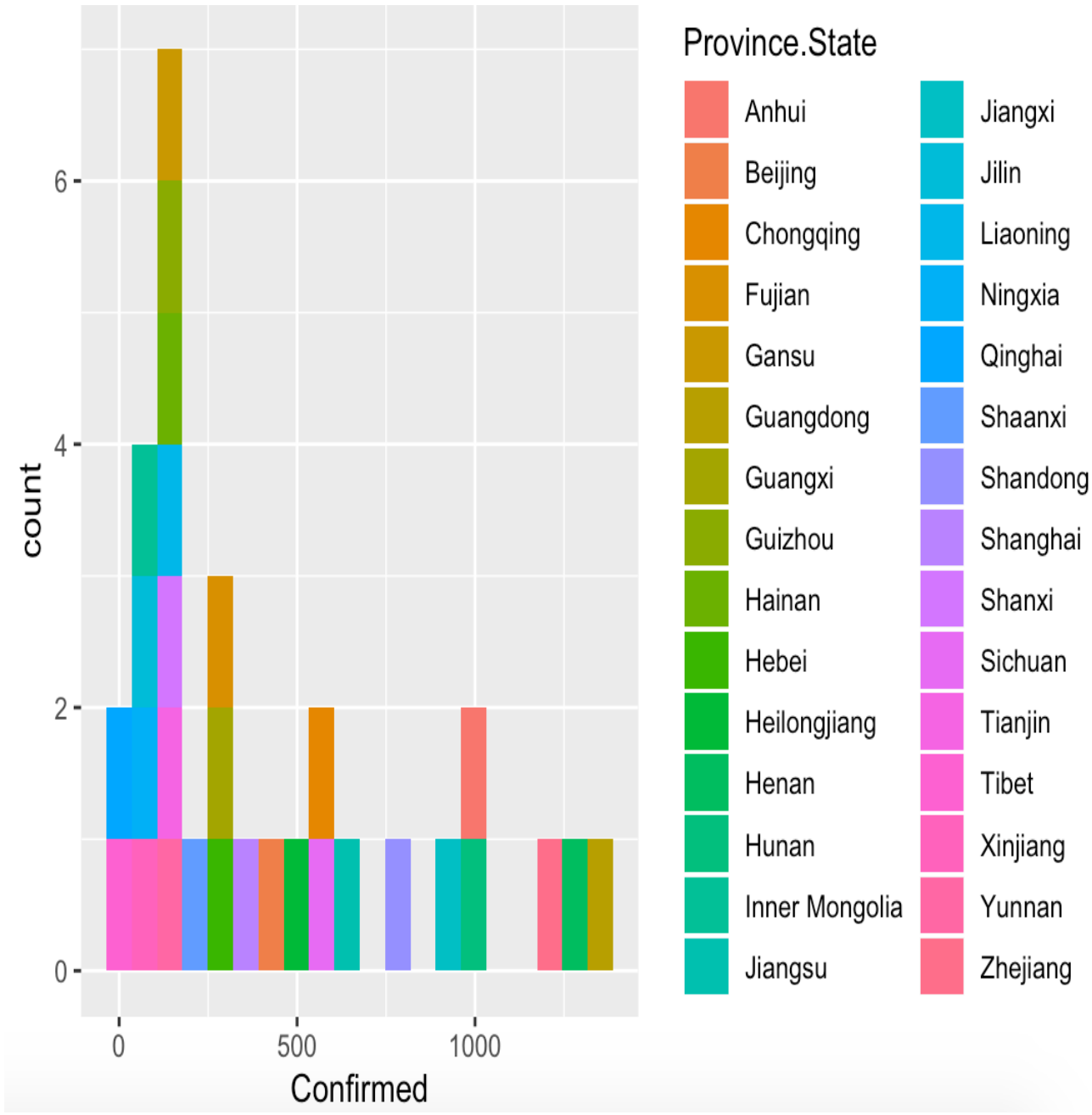
- Without the Hubei Region we see confirm that indeed, that most of the provinces have only recorded between 0 – 10 counts of confirmed cases, however there were some provinces already in the 10 – 30 range.

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



1(F)

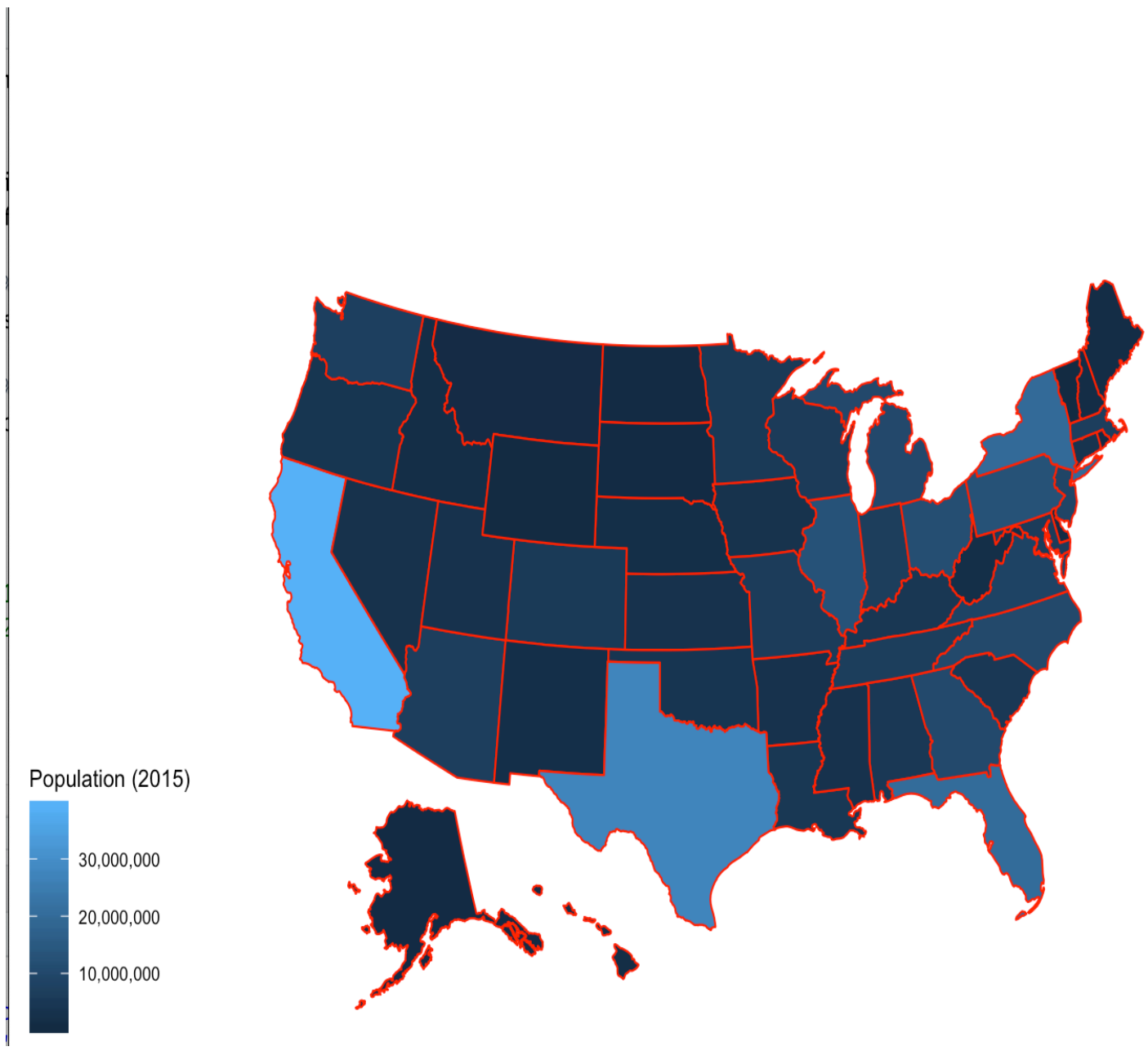
- From this visualization we see that because of the range of confirmed cases, that Hubei as a province has by far contracted the most cases in all of China.
- To visualize what the variance of confirmed sickness in China without Hubei we produce a second visualization.



1(G)

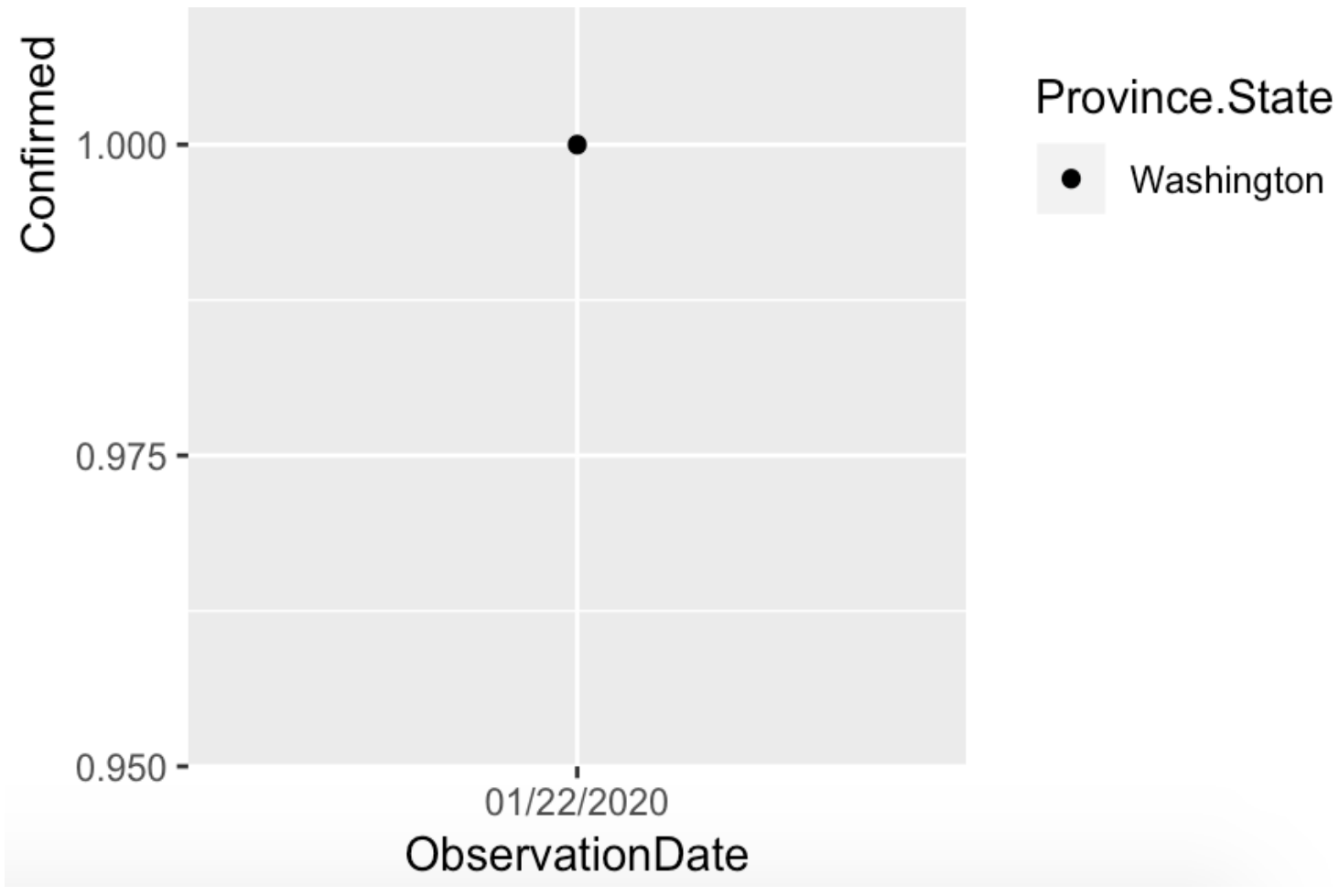
- This visualization shows that within two months, provinces that in the first month had 0 to 10 cases have on average increased to a range between 100 – 500 with many other provinces in the 500 - 1000 range.

United States Visualization



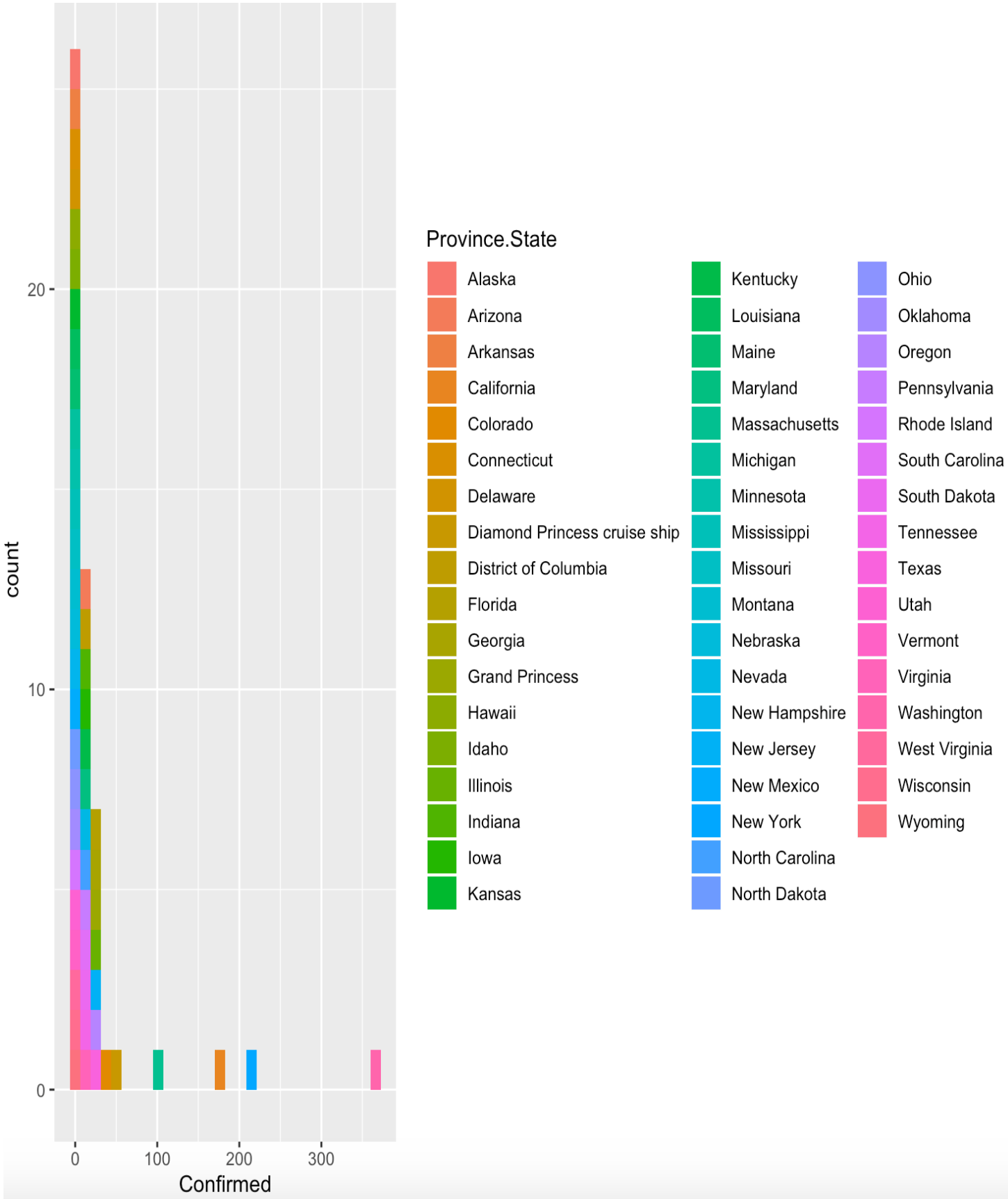
2(A)

- Population Density of the United States. Creation of code attributed to "Lorenzo, Paolo Di. "US Maps Including Alaska and Hawaii [R Package Usmap Version 0.5.0]."



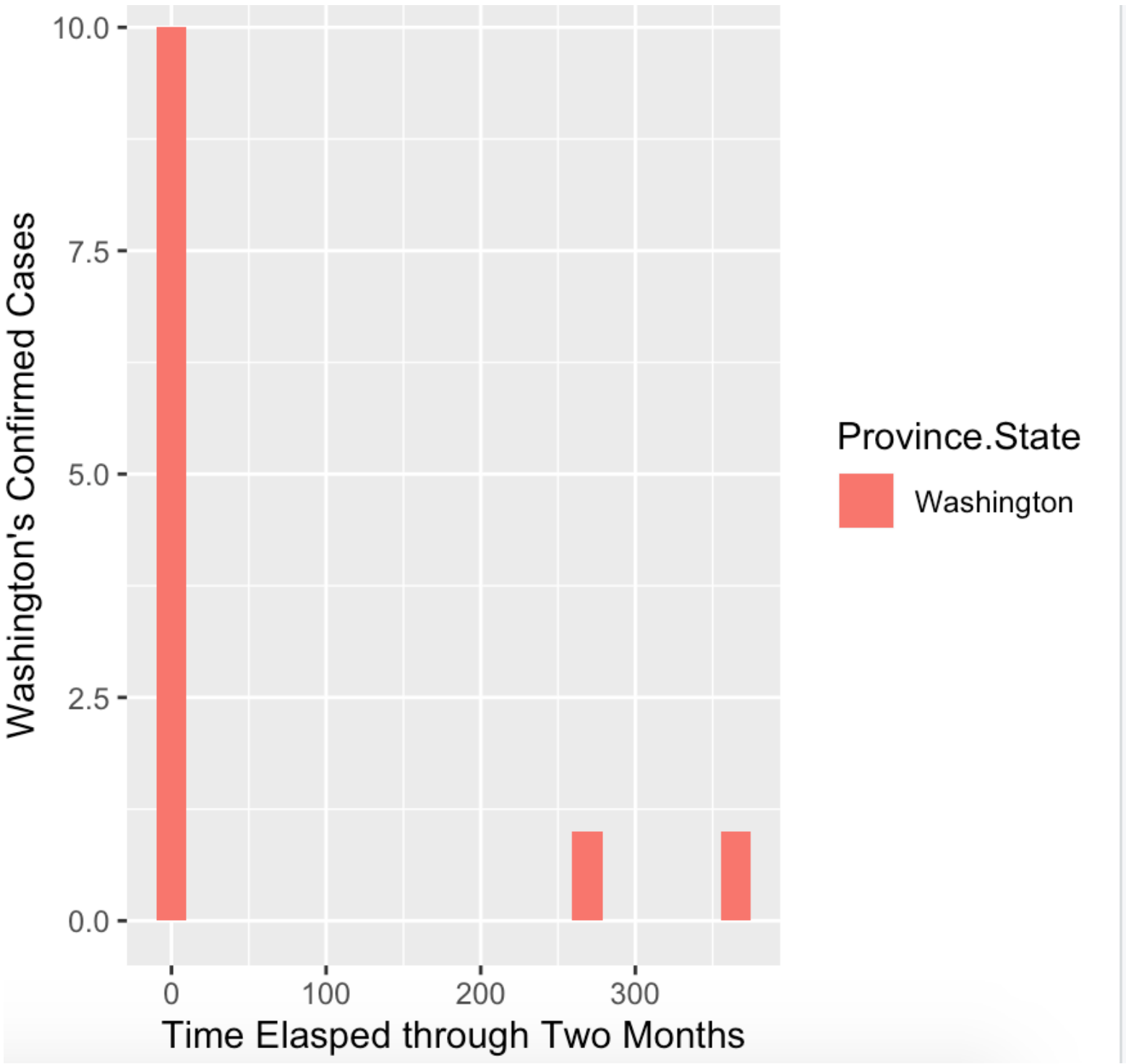
2(B)

- On Jan 22, 2020 there was one recorded case in the United States in Washington State.



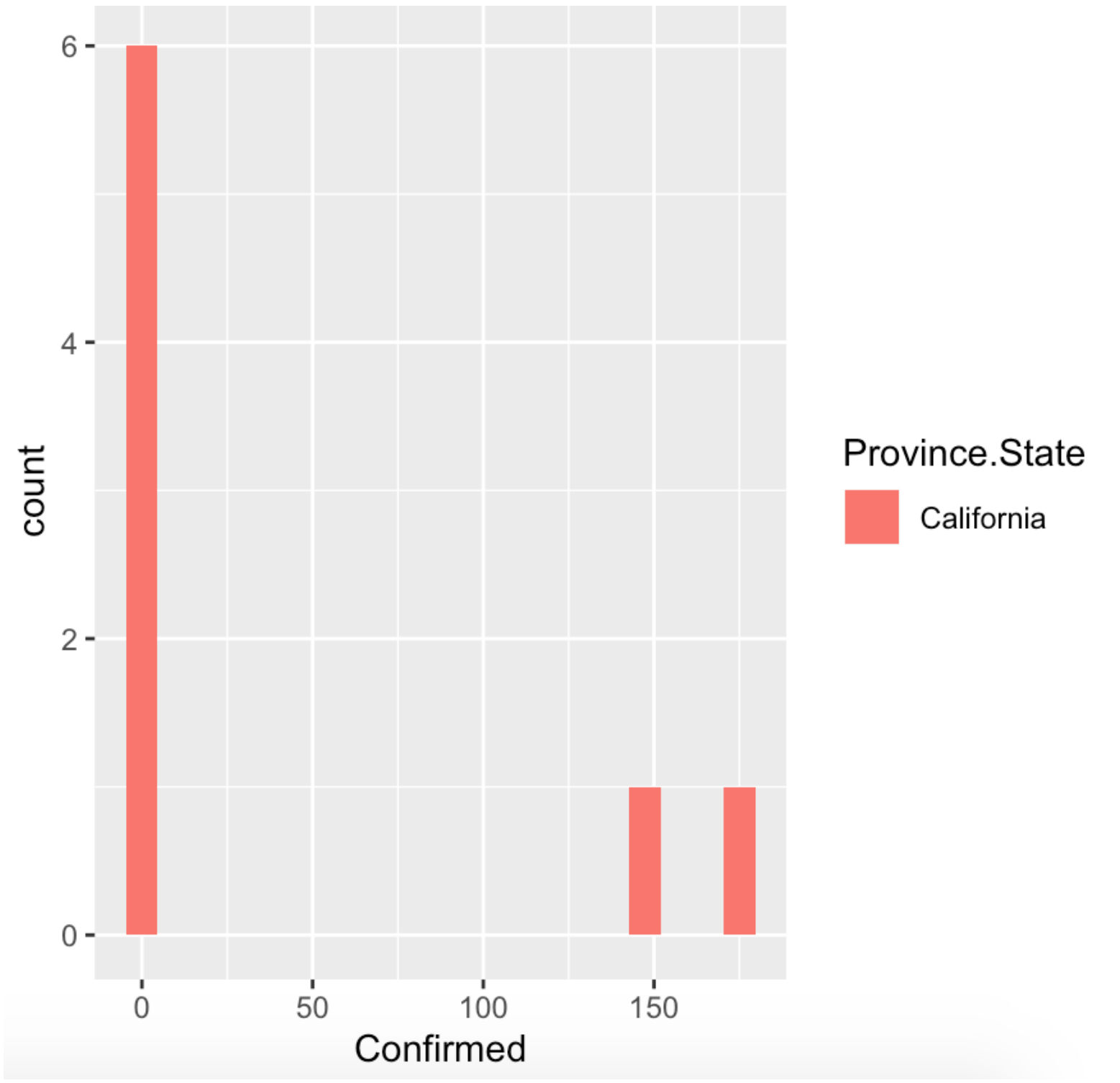
2(C)

- 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.



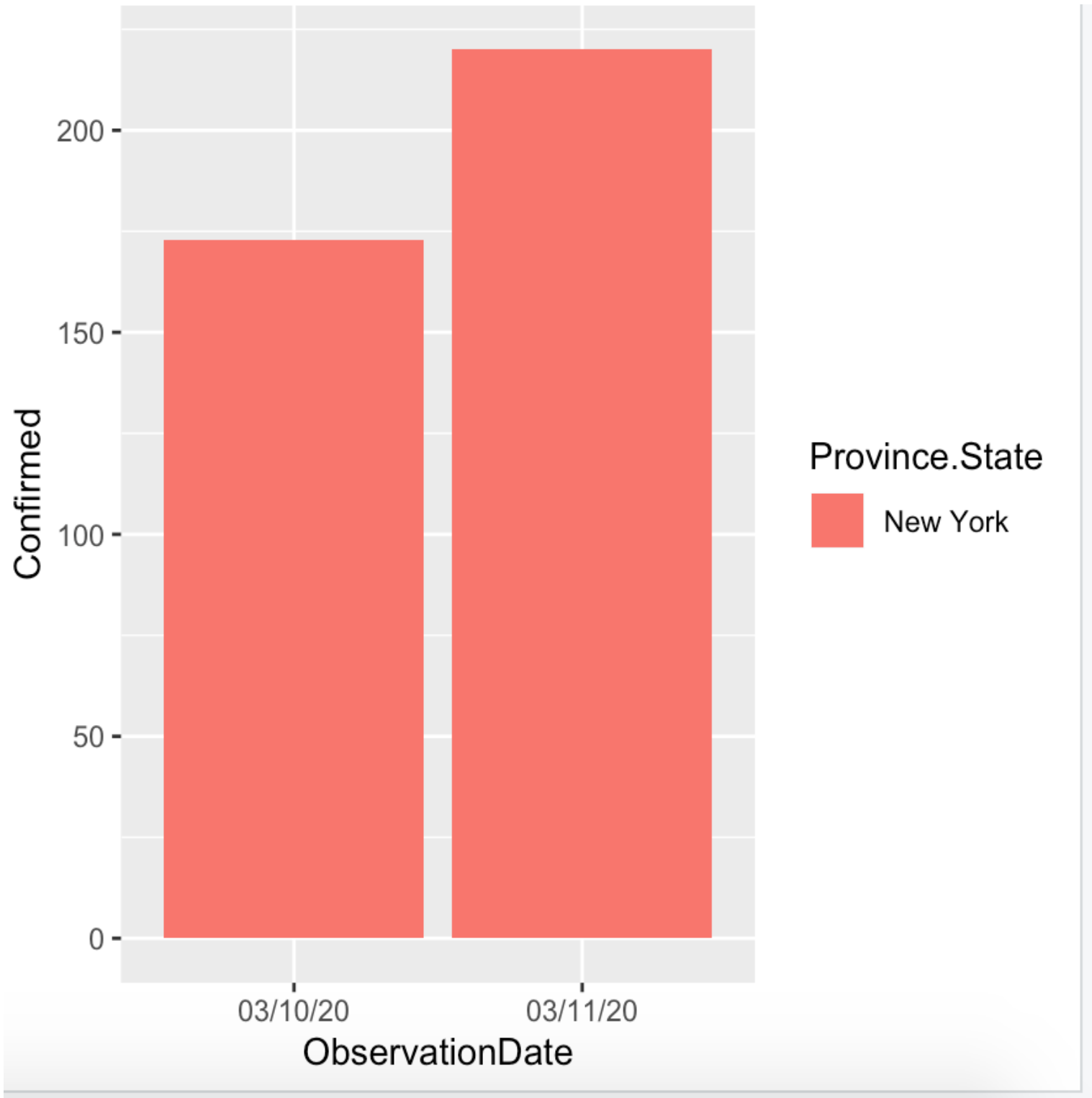
2(D)

- COVID-19 First hit Washington (State) In 2020. Within 2 months the virus went from 10 cases, to over 500. This State has the most recorded confirmed cases.



2(E)

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

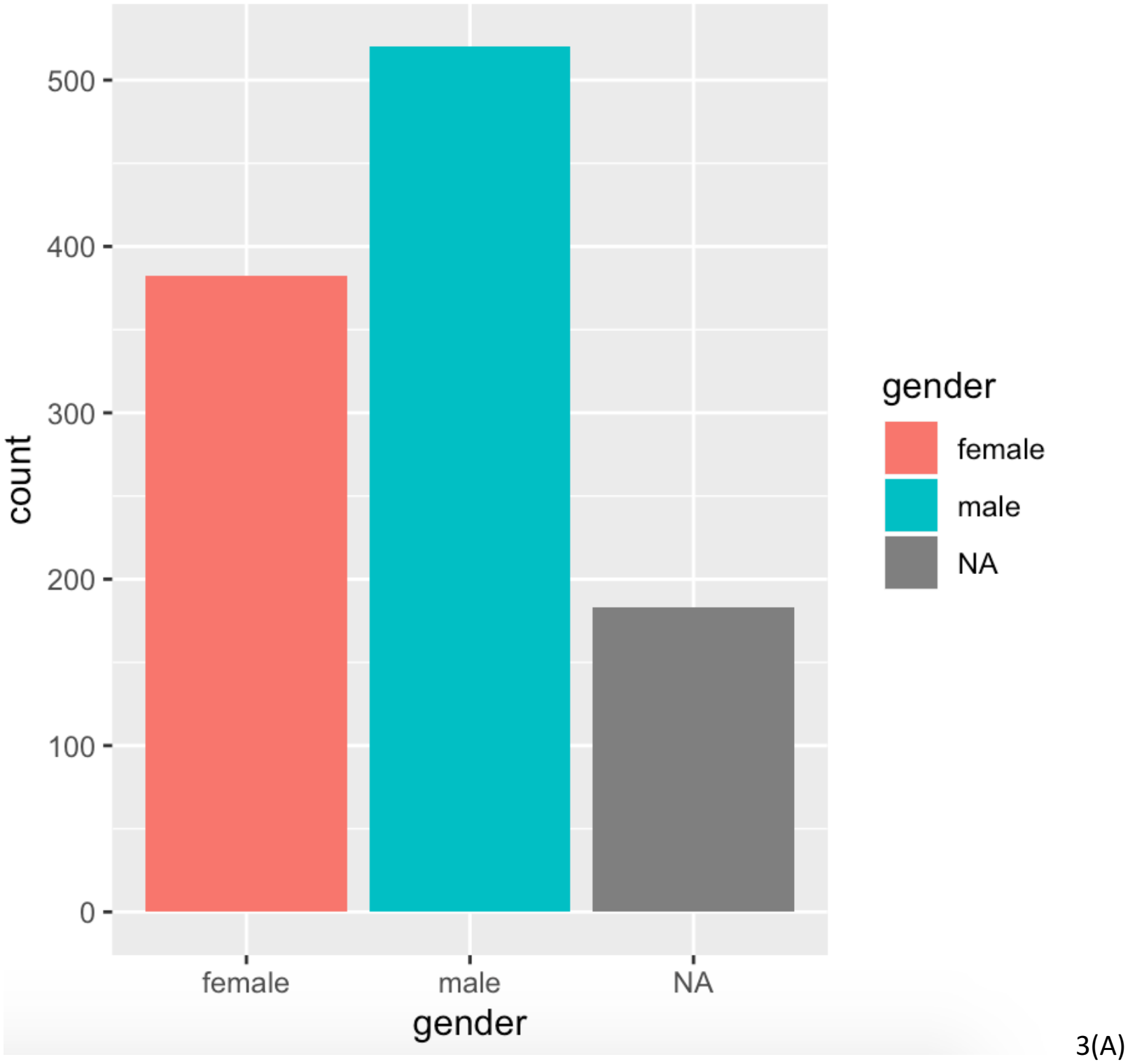


2(F)

- New York has the second most confirmed cases of COVID-19 at approximately 230. Alarmingly, in one day, the virus increased 40% in one day in New York.

DEMOGRAPHICS: 1/23/2020 (China)

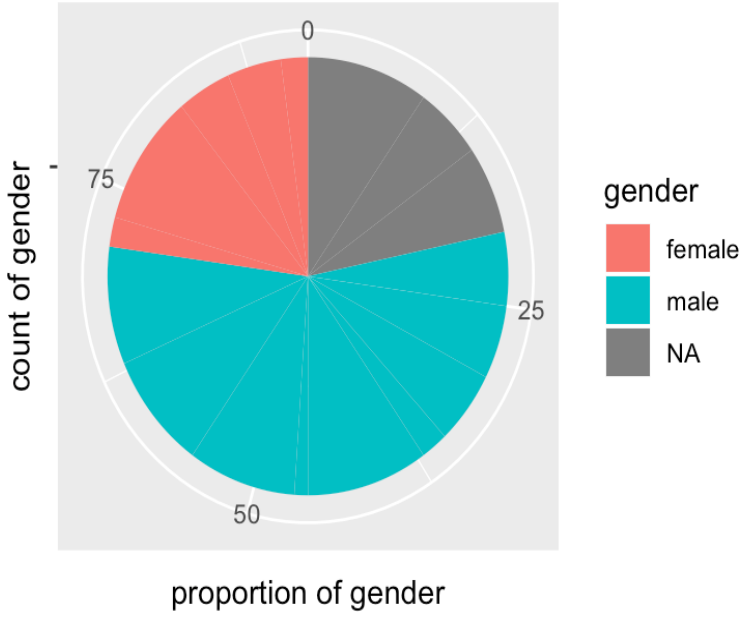
1/23/2020



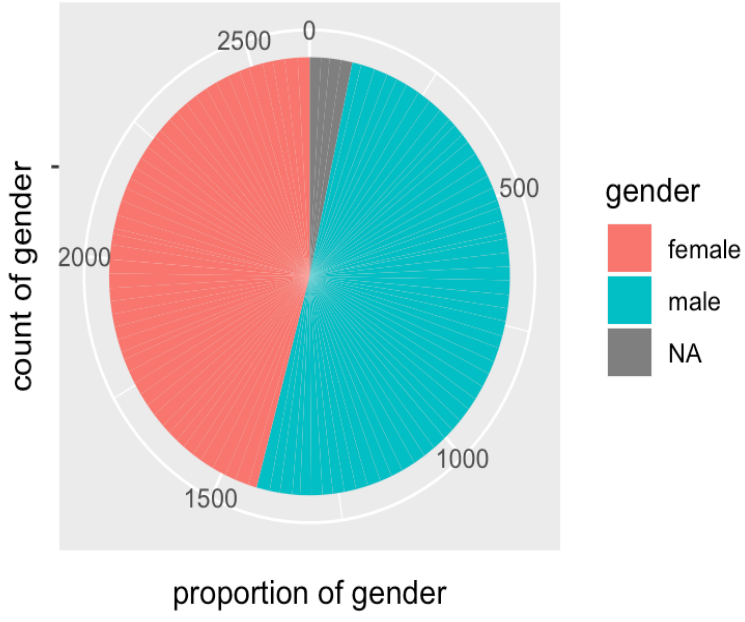
3(A)

- COVID-19 infection breakdown between gender.

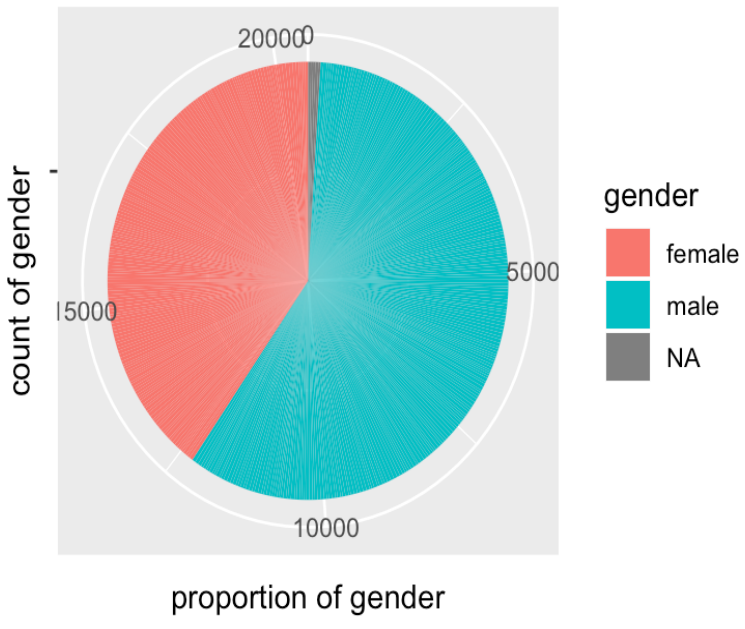
proportion 0 - 10



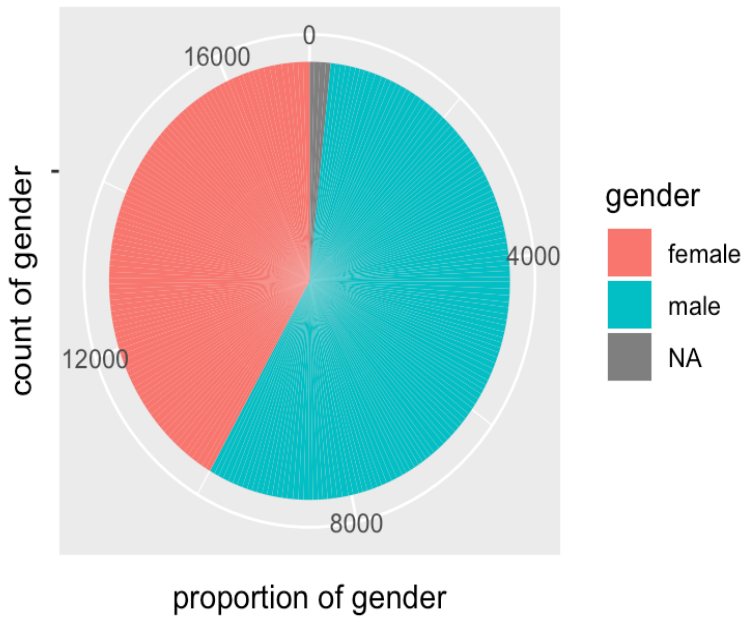
10- 30



30- 60



60 -100



3(B)

- Breakdown of infection between age groups and genders.

Discussion:

China:

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 10-30 cases. 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. This suggests 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 include: governmental lockdown of major cities, social distancing, and quarantining large areas simultaneously.

US:

Following its spread in mainland China, the Coronavirus made its way across the Ocean to the United States. The first state to have a confirmed case is Washington, with a population 7.536 million. Within two months, Washington went from 0-5 confirmed cases to over 300 confirmed cases. Following Washington was California. Its cases growing to over 200 in four 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. COVID-19 was first reported around January in the United States. To speculate what conditions may underlie the trend in the spread of the virus, some factors that

should be considered include population density, the original site of infection, and physical proximity. Because Washington State was the original site of infection, it has the highest number of confirmed cases. This spread may be attributed to victims being asymptomatic but still infectious and spreading it to other people. While New York has the second most amount of confirmed cases, it should be noted that New York State has nine out of the top 10 most densely populated cities in the United States. And while California is quite big as a state, its physical proximity to Washington has made it susceptible to the virus spread as well. From these speculative factors, it is evidenced that the virus spreads well in densely populated areas as well as its original infection site. In addition, without increased precautions to curtail the spread of the virus, the spread will continue at an alarming pace.

Demographics:

Figure 3B illustrates that besides in the youngest age group, 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 have 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. For example, the United States has a immense shortage in testing kits, therefore all data in the United States reflects lower than its true infection values.

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.*

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.

4) Srk. "Novel Corona Virus 2019 Dataset." *Kaggle*, 12 Mar. 2020, 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/.

8) *GADM*, gadm.org/data.html.