Roger IL Grande
EM-622 Midterm Project

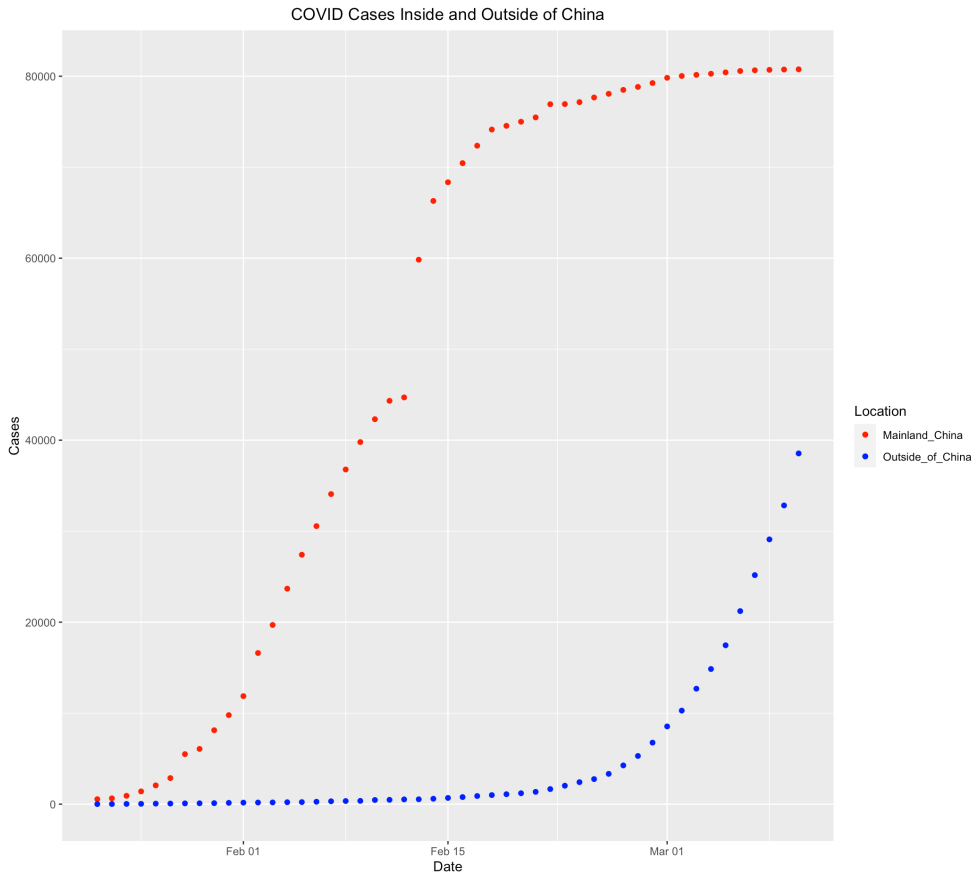
Project Description

During this early period of the Coronavirus pandemic, many of the cases were appearing in Mainland China before spreading to the rest of the world. My analysis will focus on comparing confirmed cases in China with cases outside of China. This analysis provides estimates based on available data, and there are surely variations due to differences in COVID reporting and records worldwide. Some questions that these visualizations collectively answer are:

- Did COVID begin spreading significantly within China before it began to spread around the world?
- When did COVID begin to spread outside of China's borders?
- Which regions of the world first began to report recoveries?
- After cases appeared internationally, where were they distributed?
- Were the number of deaths increasing at this early stage of the pandemic?

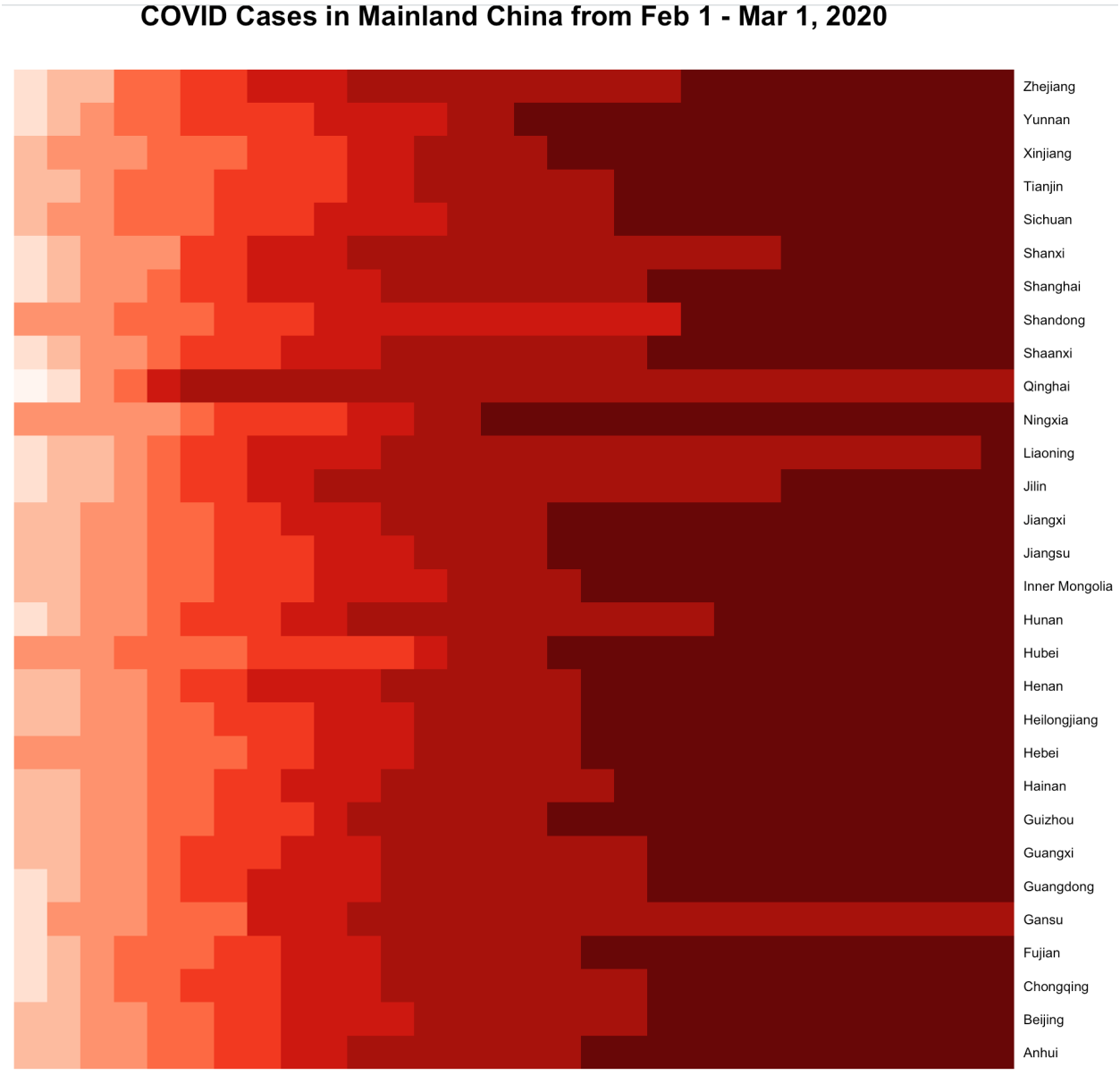
(1) Scatter Plot – Number of confirmed cases in Mainland China vs everywhere else in the world combined (lines 24-79)

This plot seeks to identify when exactly COVID began to spread internationally. The data is separated out between total cases in China and total cases everywhere outside of China.



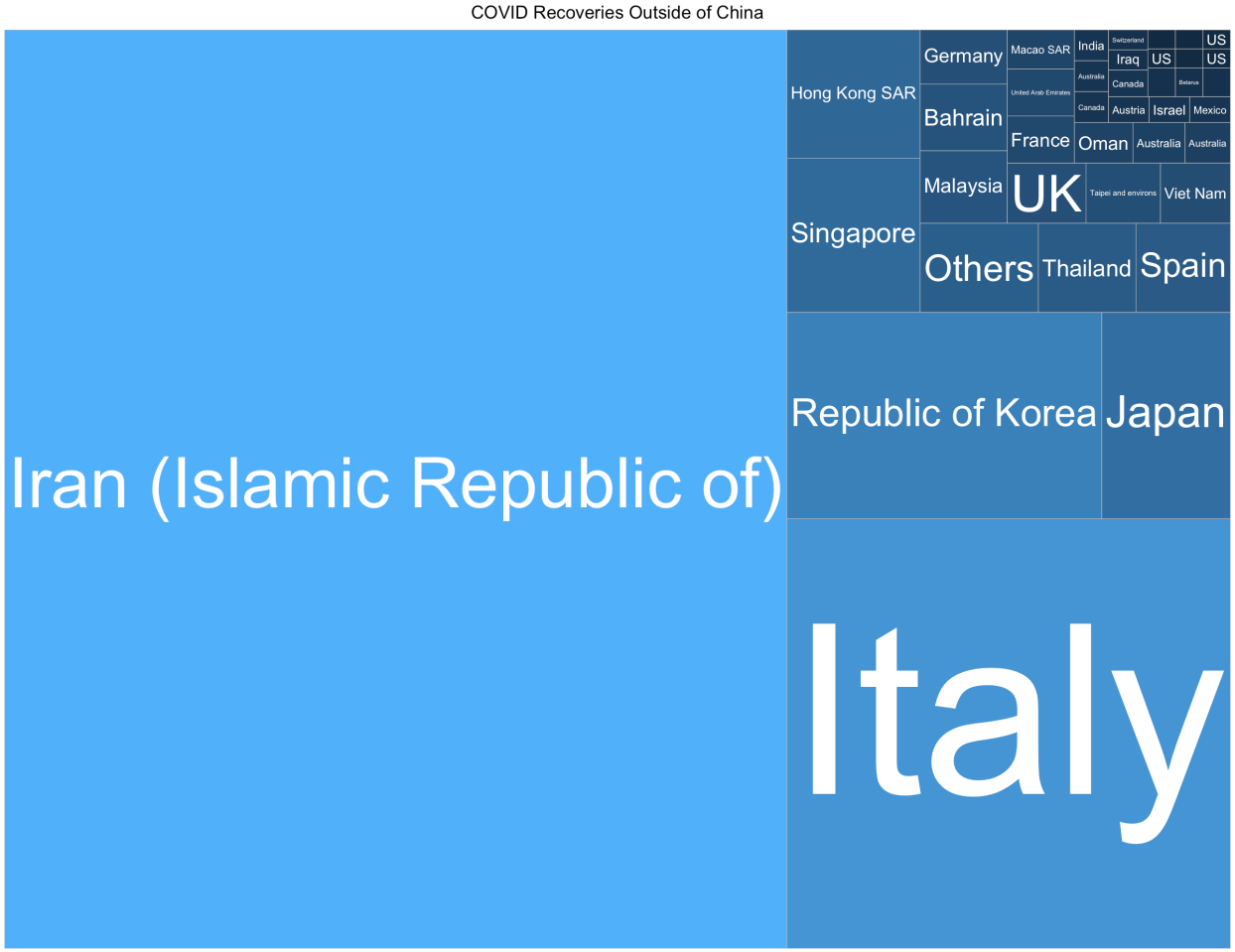
(2) Heat Map – table heat map showing COVID cases from February 1-March 1 within Mainland China (lines 84-110)

This heat map shows how cases universally increased in most regions of China in the final month before cases began to significantly spread internationally. Time progresses to March 1 from left to right.



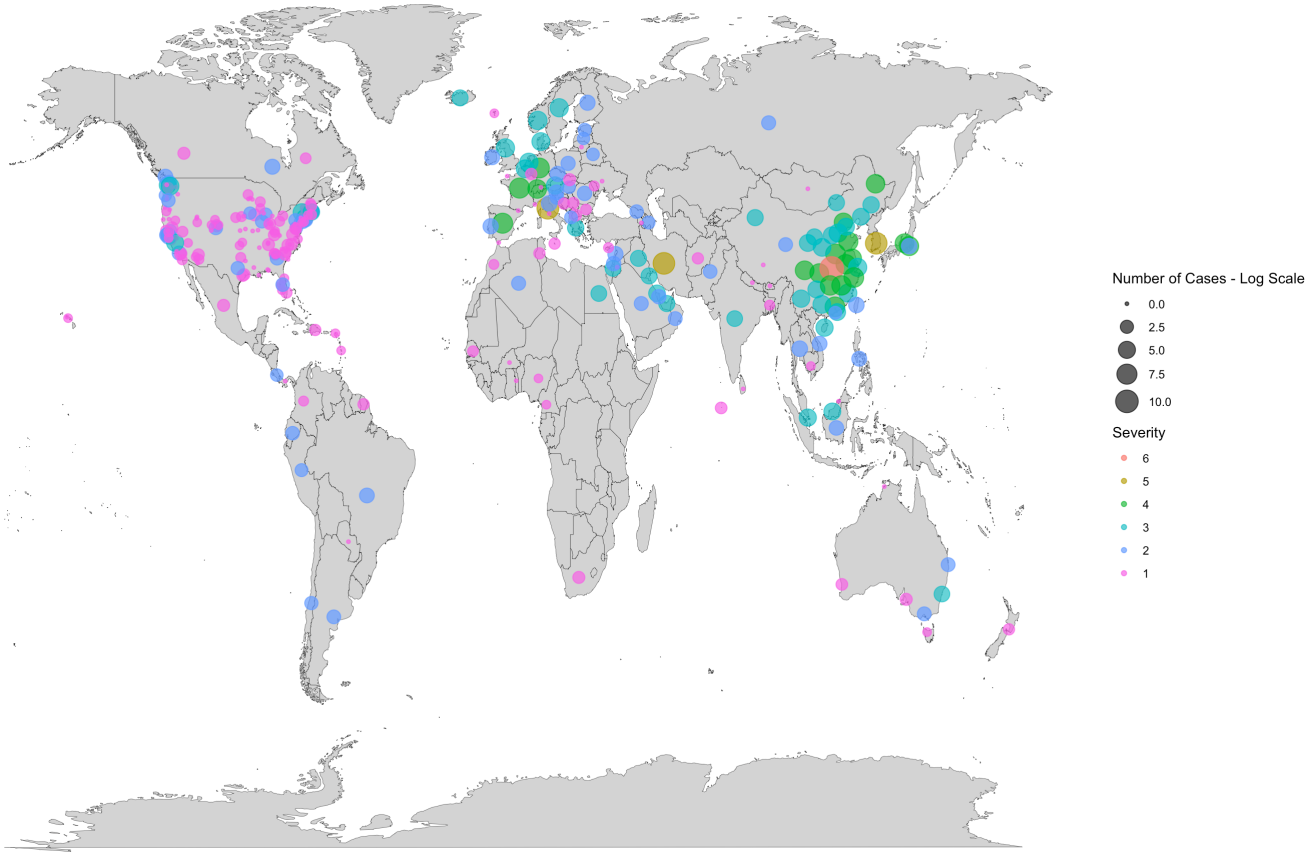
(3) Tree Map - Where the most/least COVID recoveries were by 3/10/2020 outside of Mainland China (lines 115-134)

This tree map provides a visualization for where in the world recoveries were reported by 3/10/2020, and where most of those recoveries were. The color scale is generated from a log normalized number of recoveries for each region to better emphasize a large difference in numbers (the max in Iran was almost 3000 and the minimums were under 20), and the size of each section is determined directly from the number of recoveries data.



(4) Geographic Map – Map of the world on 3/10/20 showing distribution of total confirmed cases (lines 139-163)

This is a fantastic way to visualize where the most confirmed cases were happening on March 10, 2020. There is a clear epicenter around Asia where the other plots show cases first began to rise. There are smaller numbers of confirmed cases the farther geographically they are from that orange epicenter. The size of the points for visualization purposes was based on a normalized log scale of number of cases, and the severity colors broke up this log scale further into six levels.



(5) Choice Plot – Scatterplot of total worldwide cases vs deaths (lines 168-223)

This scatter plot begins to explore deaths during the early stage of the pandemic. It is interesting to note that the rapid change in the rate of deaths shown around March 1, 2020 will not be reflected in this data set, as deaths obviously lag behind confirmed cases. This plot uses merged data from multiple data sets.