Michael Hart

mjhart@uiowa.edu

Lab 8: Time Series Data Visualization

SLIS/IGPI 6155 – Information Visualization

Due: Next Class Meeting

Description and Purposes

- This lab exercise will provide hands on experience dealing with time-based data.
- Tool: Tableau (or other tools you are most comfortable with)

Deliverables

- This document with corresponding parts filled.

Task Description

Please choose at a dataset related to COVID-19 with time element. A lot of COVID-19 datasets are available for download, including Tableau COVID-19 Data Hub and Visualization (https://www.tableau.com/covid-19-coronavirus-data-resources). This Tableau site provides not only visualizations but also related datasets for you to explore on your own.

For this lab exercise, you are asked to explore one or two datasets from Tableau COVID-19 site (You may also choose from other sources if you prefer). You may choose data sets about a specific country (or drill down to state/province, and county level), some countries, or the whole world for comparison. You may look at their confirmed cases, mortality rate, recovery data, case/mortality increasing rate, etc. You are expected to produce 3 types (out of the 7 types introduced in class) of time-based visualizations using Tableau, or another tool that you prefer.

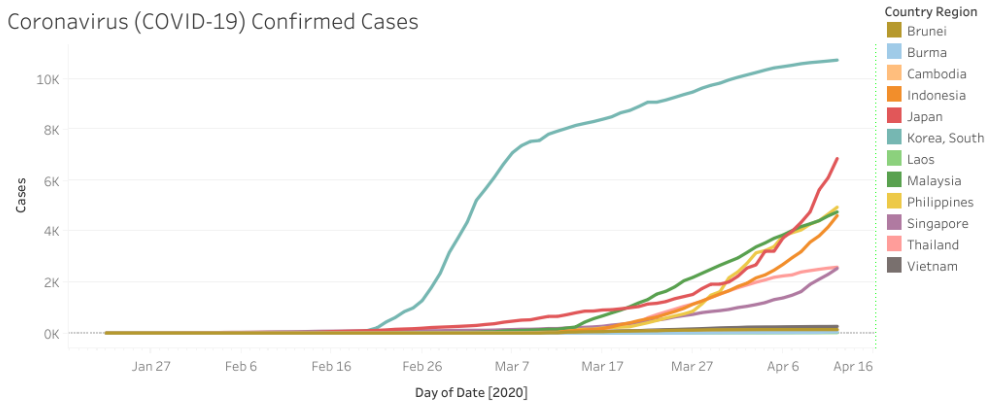
Please answer the following questions:

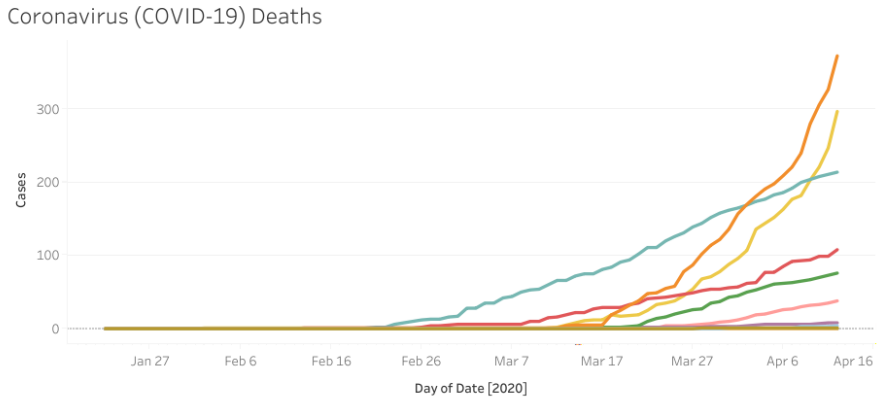
(1) A brief introduction to the dataset you chose.
I chose the dataset for East-Asia and Southeast Asia to compare the rates of the spread of the virus. I excluded China from this dataset since it is the presumed source of the original outbreak. My visualizations will display the total cases in each country, as well as the total deaths in each country.

(2) What are your main questions that you want to answer by using this dataset? 2-3 questions will be ideal.
    (1) Which Asian countries had the highest/lowest number of cases?
    (2) Do wealthier, more developed countries have a greater or fewer number of cases?

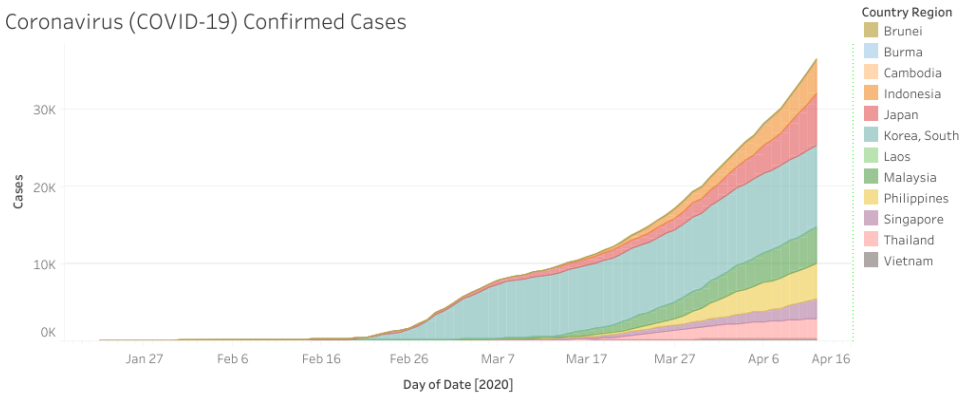
(3) Do poorer, less developed countries have a significantly greater number of deaths per case?

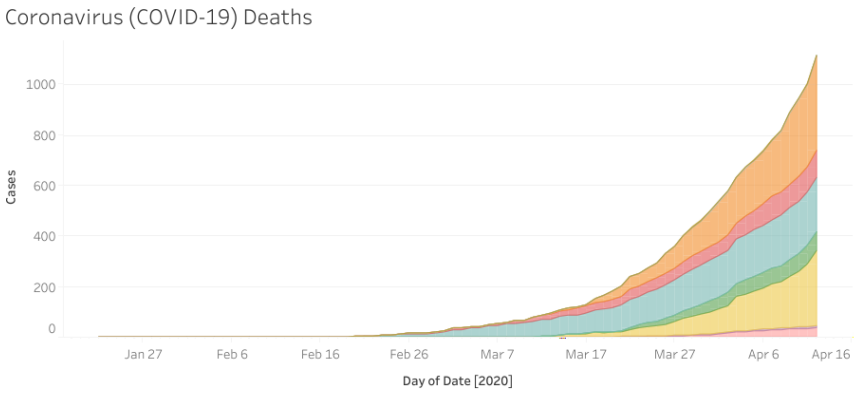
(3) Please provide a screenshot (and a workbook public link if you are using Tableau) for each of the visualizations you've created.



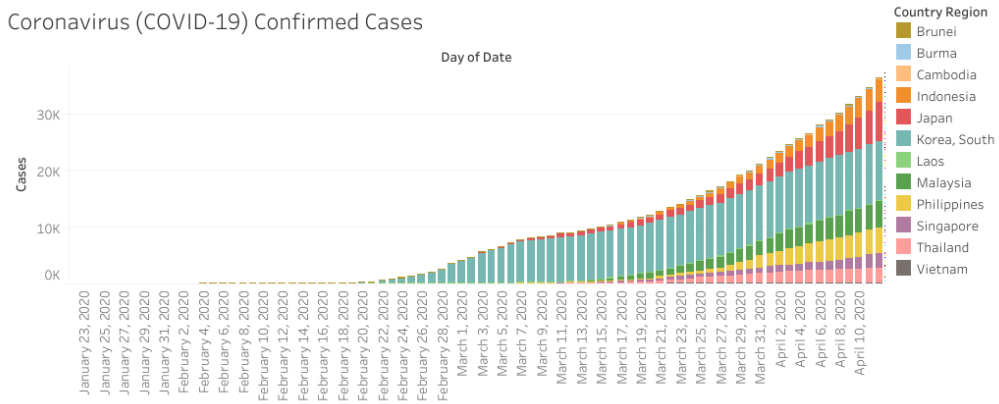


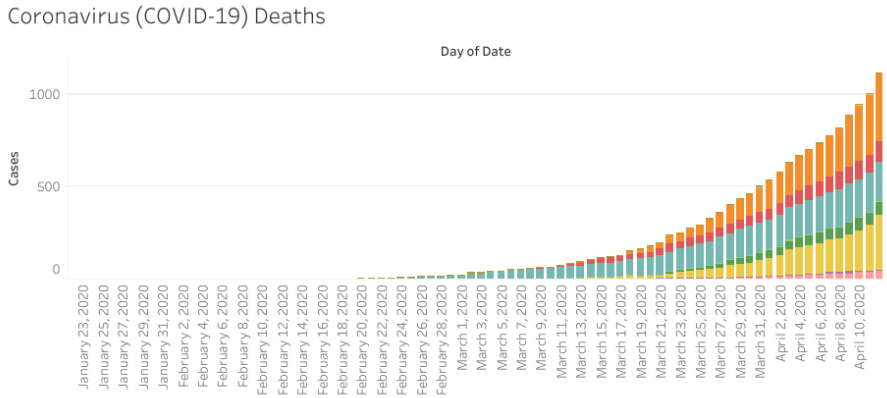
https://public.tableau.com/views/CoronavirusCOVID-19Cases_15868159142680/LineChart?:display_count=y&publish=yes&:origin=viz_share_link





https://public.tableau.com/shared/HN7PD4YS3?:display_count=y&:origin=viz_share_link





https://public.tableau.com/views/CoronavirusCOVID-19Cases_15868159142680/BarChart?:display_count=y&:origin=viz_share_link

(4) Please briefly discuss the insights you can get from the 3 visualizations, in regard to the questions proposed in step (2).

Poorer, less developed countries such as Vietnam, Burma, Cambodia, and Laos have far fewer confirmed cases than the wealthy, developed countries such as South Korea, Japan, and the Philippines. I speculate that this is simply due to the public health infrastructure in poor countries not being equipped to test cases, and that if everyone in those countries were to be tested, the actual number of cases would be much higher. It could also be that there is less travel from China, Korea, and Japan to those poorer countries; Korea and Japan have much closer business/commercial ties to China.

All three visualizations clearly show that although Japan and South Korea have a large number of cases, a glance at the visualizations immediately reveals that those two countries have a smaller proportion of deaths when compared to poorer, less developed countries. Indonesia and the Philippines have a lower number of total cases than Japan/Korea, but have a larger proportion of deaths. This shows that while wealth and technology do not necessarily prevent the spread of the virus, they are very important in saving the lives of patients.

There doesn't seem to be a consistent relationship between geographic proximity and when each countries' outbreak dramatically increased. South Korea's cases rose sharply in early March, but Japan's remained relatively low until early April. On the other hand, the Southeast Asian countries seemed to steadily increase their number of cases at roughly the same schedule as their closest neighbors.

It is important to keep in mind that the number of "confirmed cases" depends entirely on how many people are tested and the accuracy of the testing kits. Also, what is considered a "death caused by the coronavirus" probably varies between cities, states, and countries.

(5) Which of the 3 visualizations is your favorite? Why? The pros and cons of each visualization regarding graphical excellence by Tufte?

I think the stacked area chart is the best because it is easy to compare the relative number of cases between countries at a glance. The stacked area maximizes the data-ink ratio, and doesn't have empty/meaningless white space between the colors representing the different countries. The color is thick and solid enough to clearly contrast between the other colors. Even the countries with a very low number of cases/deaths are easy to pick out, since the color is uniform and solid across the chart.

The line chart is very difficult to read because the lines are so thin, and there is a lot of empty/meaningless white space between the lines that represent the different countries. This makes similar colors such as orange and yellow difficult to distinguish. A line chart like this should probably not be used with more than 5 or so color-coded parameters. My eyes quickly get tired when attempting to analyze the line chart.

The bar chart is decent, and actually looks quite similar to the stacked area chart, but has unnecessary white space between the bars. In this case there is really no reason to use the bar chart instead of the stacked area chart.

I didn't include country GDP as a measure in my visualizations because the relative wealthy/development of most countries is common knowledge, and I didn't want to overly complicate the visualizations. But with some extra work it would be possible to add it in a way that strengthens the points I was trying to make. Also, in retrospect I think having one chart of deaths per confirmed cases would be more useful than putting the confirmed cases chart on top of the deaths chart.