STAT436 Homework 4

Shiny app: https://stevenren.shinyapps.io/homework_4/

Data: https://uwmadison.box.com/shared/static/ni36m8upor79n1ij1wgjblcslkasbm5c.csv

Internet Access Over Time

Introduction

The initial dataset sourced from Kaggle spans from 1980 to 2020, encompassing various countries' statistics on cellular subscriptions, internet usage, and broadband subscriptions. This visualization serves as a valuable tool for comprehending the evolution of global internet accessibility among populations over the years.

What is the essential question that your visualization is supposed to inform?

The essential question that my visualization is how does the growth of internet access change over time globally from the years 1980 to 2020.

How do aspects of your design support exploration of the essential question? Were there trade-offs you had to make so that certain features were more clearly visible?

My design utilizes time series plots along with grouping countries by regions of the world for a more relevant comparison within regions, while also having a feature of comparing two regions' time series plots side by side. One tradeoff I made was to include countries that had missing data, such as missing years with would result in N/A values, or years where the data is suddenly 0 when the years before or after all had real values. If I decided to clean up those values, the only reasonable solution would be to remove the country completely from the visualization, which I felt like is worse than having missing values on the plot.

What are your key findings? How do they relate to your prior understanding?

From the visualizations we can see that cellular subscriptions became popular starting from the year 1990 for the Americas and Europe region, while the other regions Asia, African, and Oceania saw the growth starting from the year 1995 mostly. As for the percentage of people with access to internet, we can see that starting from the year 1995, Americas, Asia, and Europe begin their large growth in numbers, while Africa and Oceania takes until the year 2000 for most countries to begin their growth. We can conclude that looking at regions alone, some regions can be behind in internet growth by between 5-10 years than the leading regions.

How did you create the visualizations? Were there any data preparation steps?

I primarily used the plotly library so that the plots can be interactive, users can zoom in or select specific countries by double clicking on the legend. A region is chosen by default as "Americas" and can be changed, an optional region can be chosen to compare with the first one side by side, and finally a data column can be chosen to specifically visualize, a definition of the data column is shown under the plots.

To prepare the data, I had to remove all the data under the "Region" category from the Code column because the data source does not clearly define which countries fall under the regions. I also removed rows with OWID_KOS or OWID_WRL because their codes are not the standard 3 letter country.