Bilal Tariq's Final Project Proposal - Revised

Purpose

For my Final Project, I intend to look at data from COVID - 19 through a different lens. In particular, I will be analyzing 3 very different things:

1) How did the spread of COVID-19 look geographically by cases per country each year, and what was the testing rate of each country?

2) How did misinformation spread across the globe and in particular what were the sentiments of the misinformation spread.

3) Was there any correlation with the testing rate of each country and the amount of misinformation that country had? If so, what could have been the motive?

This will show us how over the course of COVID-19 spreading, how a chaos ensued within each nation and what exactly the sentiments were of the people. Incorporating unique and interesting spatial and word visualisations (as I will mention below) will go to crystallise these questions.

Data

For my data, I will be taking 3 CSV files. The first two are datasets by the European Center of Disease Prevention and Control. In particular, I will be downloading their dataset on data on testing for COVID-19 by week and country, and secondly I will be downloading the dataset on Data on the daily number of new reported COVID-19 cases and deaths worldwide.

https://www.ecdc.europa.eu/en/covid-19/data

Finally, I will be taking a look at a dataset published by UNESCO which has a dataset on the spread of misinformation within each country and has the titles of the misinformation said i.e titles of tweets and/or articles. This would then allow me to perform and LDA text analysis on the titles of misinformation and correlate it with the testing and the rise of COVID cases from that area.

https://www.unesco.org/en/world-media-trends/esoc-covid-19-misinformation-dataset

I believe my data is very well-researched already from these reliable sources, as they're from world renown organisations working towards publishing data for the greater good of the public.

Visualisations

To go above and beyond for my final project, I aim to produce visualisations that are simultaneously informative, appealing and attractive.

Visualisation #1: Leaflet Shiny App

My first visualisation would be an interactive and dynamic Shiny World Map made with leaflet. Leaflet would allow the user to scroll and drag the map, seeing more details with every movement. The visualisation would have markers placed for each country for both testing and the number of cases in that country. The user will also be able to adjust the content of the map such that it shows the number of cases. The app will also have a drop down to select which country the user wants to target and will give a detailed breakdown of the respective country's monthly Covid cases and testing rates.

Visualisation # 2: LDA Text Analysis

For the second visualization, I will perform text analysis on the dataset of misinformation spread provided by UNESCO. This will also be done with the Shiny Platform. There will be two components to this:

1) There will be a wordcloud of some infamous bigrams that have been used in this misinformation process. This wordcloud will have a drop down besides it which will also be used to select any of the countries from which the user wants to see a wordcloud from. Furthermore, it will also have a year drop down from 2020 - 2023.

2) LDA Text analysis. I will use the titles in the misinformation dataset to generate certain topics that are frequently brought up - this will allow us to gain an insight into the overall types of misinformation that is being presented.

The proposed timeline for this Project will be:

1. Data Wrangling and manipulation and add to shiny app(24th April)
2. Work on Visualisation 1 using Leaflet (27th April)
3. Finish Visualisation 2 part 1 and prototype part 2 (29th April)
4. Completed Project (4th May)