Group-A - COVID-19 Insights

Data Visualization Final Project Proposal

Project Title: Effects of COVID-19 on the working population in the U.S.

Group Members:
Shauna Han (shh2145@columbia.edu),
Toby Law (kl3343@columbia.edu),
Wenyi Li (wl2770@columbia.edu),
Yu Wang(yw3760@columbia.edu)

GitHub repository: https://github.com/QMSS-G5063-2022/Group_A_COVID-19_insights

Abstract: We would like to visualize how the COVID-19 pandemic has affected the working population of the United States. For this project, we plan to focus on several indicative features of the working population: state/industry-level trends in unemployment rate, differences in daily movement patterns of the population, changes in wage and working hours in different industries, and sentiment within text data related to unemployment. We will focus on the United States at both national and state levels.
We aim to generate five visualizations using techniques such as ggplot2, leaflet, tmap, web scraping, text analysis packages in R/Python, and more.

Visualizations:

1. **Unemployment rate**
   - Figure 1: A line chart to capture changes in average unemployment rates over time. We plan to first plot a national average, and also include separate line plots so as to contrast trends between different industries (static plot, with annotations indicating COVID-19 waves or lockdown periods)
   - Figure 2: An interactive map showing all 50 states, with a timeline toggle, plus individual values for each state. Data ranging from Jan 2019 - Dec 2021, either monthly or quarterly values
   - National data: https://data.bls.gov/timeseries/LNS14000000?years_option=all_years
   - State-level data: https://www.bls.gov/lau/

2. **Average weekly wage and working hours - how wage and working hours have been impacted by the COVID-19 pandemic**
   - Datasets: https://data.bls.gov/cew/apps/data_views/data_views.htm#tab=Tables - data finder for weekly wage by industry
     https://www.bls.gov/web/empsit/ceseeb4a.htm - data table for weekly wage and working hours, segregated by industry
   - Main figure: average weekly wage and working hours across time, aggregated by industries (before and at the height of the COVID-19 pandemic)

- Datasets on hand range from 2017 - 2021, with yearly records
- Line chart to show how wage and working hours have changed due to effects of the pandemic, and whether there are differences between industries.
- Auxiliary figure: changes in weekly working hours across time (U.S. national level, timeframe: 2018 - 2021). Dataset on world Employment and Social Outlook https://www.ilo.org/wesodata/?chart=Z2VuZGVyPVsiVG90YWwiXSZ1bml0PSJOdW1iZXIiJnNlY3Rvcj1bXSZ5ZWFyRnJvbT0xOTkxJmluY29tZT1bXSZpbmRpY2F0b3I9WyJ1bmVtcGxveW1lbnQiLCJ0b3RhbExhYm91ckZvcmNlIiwid29ya2luZ0hvdXJzIl0mc3RhdHVzPVtdJnJlZ2lvbj1bIldvcmxkIl0mY291bnRyeT1bIlVuaXRlZCBTdGF0ZXMiXSZ3b3JraW5nUG92ZXJ0eT1bXSZ5ZWFyVG89MjAyMyZ2aWV3Rm9ybWF0PSJDaGFydCImYWdlPVsiQWdlMTVwbHVzIl0mbGFuZ3VhZ2U9ImVuIg%3D%3D

3. **Movement of population**
    - Map-based visualization: since cities have gone into lockdown during the pandemic, people have switched to working from home and offices have closed down, their day-to-day movement ranges and patterns have drastically changed. We aim to visualize differences in people's movement patterns across time before and during the pandemic.
    - Data for movement ranges of Facebook users consenting to share their location: https://dataforgood.facebook.com/dfg/tools/movement-range-maps
    - Geospatial data for the US, for mapping purposes: https://gadm.org/download_country.html

4. **Sentiment analysis for unemployment**
    - Mine the information posted on Twitter about unemployment using hashtags
    - Perform text mining to generate word clouds of common phrases contained by those posts, and to conduct positive/negative sentiment analysis. We can compare the word clouds and sentiment composition for posts timestamped before and after the emergence of COVID-19.
    - Time range: December 2019 to Dec 2021
    - Useful resource → https://ieeexplore.ieee.org/document/7456920; https://towardsdatascience.com/how-to-navigate-analytics-job-search-during-covid-19-b6d2159fac8e

Any significant hurdles/doubts that we can foresee:
We each need a better grasp on data visualization techniques as well as text mining techniques at the time of writing. There could be unforeseen challenges that we encounter during data cleaning, dealing with missing or inaccurate data etc.