

COVID-19 and Education Analysis Project

Description

The repo for CSC-324 at Grinnell College by student Haobo Chen This repository contains the following:

- The source code for the implementation of this COVID-19 and Education project, including a Shiny Web App file and an R Markdown file.
- The database folder which includes the three datasets used.
- A description of the three datasets.
- A timelog folder which includes the time log and chart created in R for the project.
- Two screenshot of the project progress.
- A feedback from Vivero Digital Fellows program
- A geographical coordinates dataset

About the App

- Accessing the App: You can either run the source code locally or access the app via this link: <https://haobochen2002.shinyapps.io/CSC-324/>
- App Description: The app consists of two pages: one map page, one plot page, and a description page.
 - Map page: This page consists two US maps with a sidebar. One map is about US education level and the other is about COVID-19 cases in US. From the sidebar, users can select the desired attribute to be displayed on the map, including different education level and COVID-19 cases, deaths and vaccination condition. Additionally, there is also a project statement, a brief introduction of the project and a brief introduction of the attributes.
 - Plot page: This page displays three plots. The first is a scatter plot. Each dot represents a state. The x-axis indicates the education level of each state, with higher numbers indicating higher education level. The y-axis represents the total number of COVID-19-related cases, death and vaccine condition in each state. The choice of attribute can be selected from the sidebar. The second plot is a bar plot. It shows the top ten states with the highest vaccine coverage. The third plot is a bubble plot. It shows the relationship between the number of COVID-19 vaccine distributed and the number of COVID-19 vaccine administered in each state. The size of the bubbles represents the number of COVID-19 deaths in each state.
 - Description page: This page provides a brief introduction of the project, the attributes and the data sources.

Project Purpose

The primary goal of this project is to study the relationship between COVID-19 and education level in US. In other words, I want to see if there is a relationship between the education level of a state and the COVID-19 condition in that state, including the number of cases, deaths and vaccine condition. There are four different education level attributes in the dataset, including the percentage of people with high school diploma, the percentage of people with bachelor degree, the percentage of people with graduate degree and the percentage of people with no high school diploma.

Data Description

The descriptions of the three datasets are provided in the **Data description.pdf** file and the **Geographical coordinates.shp** file. The three datasets are:

- **us_education_data**: Contains the education level of each state in US.
- **us_states**: Contains total COVID-19 cases, deaths for each state in US at 2023-12-31.
- **cdc_vaccines_distributed_administered_by_jurisdiction**: Contains the daily COVID-19 vaccine condition of each state in US
- **US_geographic_dataset**: Provides geographical coordinates for each state in US.

Data Collection

- The dataset is an combination of various datasets, providing insights into COVID-19 and education level accross US.
- It includes symbolic data (city names and states) and numeric data (COVID and education condition).
- The data is collected from Kaggle, New York Times, CDC and U. C. Bureau [1-4].

Users of This Project

This interactive App is designed for:

- Public Health Researchers: To understand the relationship between COVID-19 and education level in US.
- Data Analysts and Economists: To analyze and compare the COVID-19 condition and education level in US.
- Policy Makers: To make policy decisions based on the relationship between COVID-19 and education level in US.

Questions Trying to Answer

- Is there a relationship between the education level of a state and the COVID-19 condition in that state?
- What is the relationship between the education level of a state and the COVID-19 vaccine condition in that state?
- Have COVID-19 vaccinations been effective in preventing deaths from the virus?

Insights

- There is a positive relationship between the education level of a state and the COVID-19 cases and death in all states. The higher the education level, the lower the COVID-19 cases and death are.
- The COVID-19 vaccine has been effective in preventing deaths from the virus. The more vaccine distributed, the more vaccine administered and the less death are.

Improvement Wishlist

- Date Expansion: Include data from different years to show how the COVID-19 condition and education level change over time.

- Additional data: Integrate more factor that may affect the COVID-19 condition and education level, such as the population density, the average income and the average age. So that we can have a more comprehensive understanding of the relationship between COVID-19 and education level.
- User Interface: Add more pages at the dashboard for better visualization.
- Diagram: Add more interactive diagrams, such as pie chart and histogram, to allow users to explore different facets of the data.

Data Tidying and Pre-processing

- The data is read from three CSV files named `education.csv`, `us_states.csv` and `cdc_vaccines_distributed_administered_by_jurisdiction.csv`.
- The `education.csv` file is merged with the `US_geographic_dataset.shp` file to create a new dataset.
- The `us_states.csv` file is merged with the `US_geographic_dataset.shp` file to create a new dataset.
- The `cdc_vaccines_distributed_administered_by_jurisdiction.csv` file is merged with the `US_geographic_dataset.shp` file to create a new dataset.
- The data is then grouped by `state` to calculate averages for various COVID-19 and education metrics.
- The education level is calculated by the percentage of people with high school diploma, the percentage of people with bachelor degree, the percentage of people with graduate degree and the percentage of people with no high school diploma.
- The following R libraries are used for processing and visualizing the data:
- `readr`, `dplyr`, `tidyr`, `magrittr`, `ggplot2`, `lattice`, `wordcloud2`, `shiny`, `leaflet`, `maps`, `sf`, `shinydashboard`.

What-Why-How Analysis Framework

1. What

- Datasets and Attributes:
 - Datasets: The primary datasets are tables of COVID-19 cases, deaths, vaccine condition and education level across different states in US.
 - Attributes: The datasets contain quantitative attributes like COVID-19 cases, deaths, vaccine condition and education level.

2. Why

- The purpose of this app is to provide insights into how education levels across different states may correlate with COVID-19 impacts and vaccine distribution efforts. By analyzing and visualizing these relationships, the application can help policymakers, researchers, and the public understand the potential links between educational attainment and the pandemic's outcomes. This could inform targeted educational or health interventions, policy making, and resource allocation to mitigate the effects of the pandemic and improve public health responses.

3. How

- General: The data is encoded using maps.

- Spatial Position: The primary encoding for the US map. States are positioned based on geographical coordinates.
- Color: Used to represent data magnitude. For instance, darker shades indicate higher cases, deaths, vaccine distributed or vaccine administered.
- Navigate: The Shiny app provides multiple views to allow users to explore different facets of the data.
 - Users can select different attributes (e.g., cases or deaths) and time periods to view different data on the US map.
- Interact:
 - Click over states on the map provides detailed data for that state.

Sources or References

- [1] Kaggle. "Education by State," Accessed Feb 1st, 2024.
<https://www.kaggle.com/code/mpwolke/education-by-state-2014-2018>
- [2] New York Times. "COVID-19 data," Accessed Feb 1st, 2024. <https://github.com/nytimes/covid-19-data>
- [3] Kaggle. "CDC COVID-19 Vaccine Tracker," Accessed Feb 1st, 2024.
<https://www.kaggle.com/datasets/thedevastator/cdc-covid-19-vaccine-tracker>
- [4] U. C. Bureau, "Tiger/line shapefiles," Census.gov, <https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html> (accessed Oct. 10, 2023).

Appendix

- Time Log Chart: The time log chart is included in the timelog folder.

a list of brief descriptions of the 100% requirements

- ☒ Combine various datasets. I used three datasets in this project, including the education level of each state in US, the total COVID-19 cases, deaths for each state in US at 2023-12-31 and the daily COVID-19 vaccine condition of each state in US. I combined these three datasets to create the project.
- ☒ Feedback report of your visit to the Vivero Digital Fellows program I included a feedback from Vivero Digital Fellows program in the repository.