US COVID-19 Information and Education Level Analysis

Education level data

Description: The data set includes US education level in each state and was updated in a yearly basis. The data set include the number of populations in a specific category and percent of population in a category. The data set will be used to compare against the COVID-19 information(cases, death, vaccine distributed, vaccine administered).

Format: A data frame with 3157 observations, 10 variables, and 31570 values in 2014 to 2018. Below are some columns that I will be using.

[, 1] State: the location of the observation

[, 2] Percent of adults with less than a high school diploma: The percentage of adults with less than a high school diploma of the location.

[, 3] Percent of adults with a high school diploma only: The percentage of adults with a high school diploma only of the location.

[, 4] Percent of adults completing some college or associate's degree: The percentage of adults completing some college or associate's degree of the location.

[, 5] Percent of adults with a bachelor's degree or higher: The percentage of adults with a bachelor's degree or higher of the location.

Usage: education_data <- read_csv("database/Education.csv")

Source: Kaggle. "Education by State," Accessed Feb 1st, 2024.

https://www.kaggle.com/code/mpwolke/education-by-state-2014-2018

COVID-19 level data

Description: The data set includes US COVID-19 cases and deaths information. The data will be combined with vaccine data set and compare against the education level. The data was updated in a daily basis.

Format: A data frame with 61943 observations, 5 variables, and 309715 values in 2020 to 2023.

Below are some columns that I will be using.

[, 1] Date: the time of the observation

[, 2] State: the location of the observation

[, 3] Cases: The number of populations infected COVID-19 at the specific day

[, 4] Deaths: The number of populations died due to COVID-19

Usage: covid_data <- read_csv("database/us-states.csv"

Source: New York Times. “COVID-19 data,” Accessed Feb 1st, 2024.

https://github.com/nytimes/covid-19-data

Vaccine level data

Description: The data set includes the number of vaccines distributed and vaccines administered across US. The data set also includes a ratio part of data, which indicated the vaccines distributed and vaccines administered per 100k people. The data was reported at 2023.

Format： A data frame with 64 observations, 16 variables, and 1024 values in 2023.
Below are some columns that I will be using.

[, 1] Date: the time of the observation

[, 2] State: the location of the observation

[, 3] doses_distributed: The number of vaccines distributed through the state.

[, 4] doses_administered: The number of vaccines administered through the state.

[, 5] doses_distributed_per_100k: The number of vaccines distributed through the state in every 100k people.

[, 6] doses_ administered _per_100k: The number of vaccines administered through the state in every 100k people.

Usage: vaccine_data <-
read_csv("database/cdc_vaccines_distributed_administered_by_jurisdiction.csv")

Source: Kaggle. “CDC COVID-19 Vaccine Tracker,” Accessed Feb 1st, 2024.

https://www.kaggle.com/datasets/thedevastator/cdc-covid-19-vaccine-tracker