Wenhao Wang

DSCI 510

Final Project Proposal

Title: Education and Immunization: Analyzing the Relationship Between Regional Education Levels and COVID-19 Vaccination Rates in the U.S.

Research Question:

This project will examine how regional education levels correlate with COVID-19 vaccination rates across the United States. The central question is:
Do states or counties with higher educational attainment also report higher COVID-19 vaccination rates?

I will also introduce the urban/ rural classification as the third factor, since the population density and life style is very different in cities and rural areas, I would infer that the vaccination rate could be also different even with similar education level.

Data Source 1: COVID-19 Vaccination Data – CDC

Source: https://data.cdc.gov/Vaccinations

Will use API to collect data from the CDC's public datasets. Includes vaccination rates by county and state.

Data Source 2: Education Attainment Data – U.S. Census Bureau

Source: https://data.census.gov/table/ACSST1Y2023.S1501?q=Education

Scraped using BeautifulSoup from education tables showing the percentage of adults with a high school diploma or higher, and bachelor's degree or higher, by region.

Data Source 3: Urban-Rural Classification – USDA

Source: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/

Provides categorical classifications of counties as urban or rural, which helps to contextualize the education-vaccination relationship by geographic and infrastructural factors. Will download CSV directly.

How Sources Combine:

- All three datasets include geographic identifiers (e.g., FIPS codes, county/state names), allowing for merging by region.
- I will combine the datasets using these shared identifiers, the merged dataset will include:
   - Vaccination rate per region (dependent variable)
   - Education attainment level per region (independent variable 1)
   - Urban/rural classification per region (independent variable 2, categorical)
- I will then employ linear regression method to test how education level and urban/rural status jointly predict vaccination rates.

References:

Centers for Disease Control and Prevention. (2025). COVID-19 Vaccination Data. Retrieved from https://data.cdc.gov/Vaccinations

U.S. Census Bureau. (2025). Educational Attainment Data. Retrieved from https://data.census.gov/table/ACSST1Y2023.S1501?q=Education

U.S. Department of Agriculture. (2025). Rural-Urban Continuum Codes. Economic Research Service. Retrieved from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/