**COSC 578 Project 2**

Team Member: Yuming Wang, Ziang He, Jianchen Sun

1. **Source Identification**

For this project, we used three datasets to build our Database System:

* Dataset\_0: [United States COVID-19 Cases, Deaths, and Laboratory Testing (NAATs) by State, Territory, and Jurisdiction](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days)
* Dataset\_1: [Data Table for Vaccinations Equity (SVI)](https://covid.cdc.gov/covid-data-tracker/#vaccination-equity)
* Dataset\_2: [Data Table for Vaccinations Equity (Metro/Non-Metro)](https://covid.cdc.gov/covid-data-tracker/#vaccination-equity)
  1. **Dataset\_0**

This dataset shows the number of COVID-19 cases, deaths and laboratory testing for every 100,000 people over the last 7 days, allowing us to compare areas with different population sizes.

* 1. **Dataset\_1 and Dataset\_2**

These two datasets provide a county-level view of COVID-19 vaccination coverage, Social vulnerability and Metropolitan vs. Non-Metropolitan:

Social vulnerability is measured by CDC Social Vulnerability Index (SVI), which uses U.S. Census data on categories like poverty, housing, and vehicle access to estimate a community’s ability to respond to and recover from disasters or disease outbreaks.

Metropolitan vs. Non-Metropolitan classification is based off an aggregation of the six 2013 National Center for Health Statistics (NCHS) Urban-Rural classifications, where “Metro” counties include Large Central Metro, Large Fringe Metro, Medium Metro, and Small Metro and “Non-Metro” counties include Micropolitan and Non-Core (Rural).

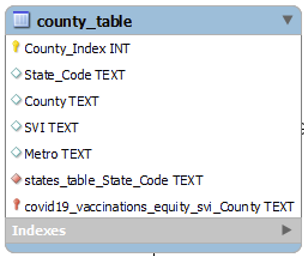
* 1. **Data Preprocessing**

1. **Schema Definition**

Every relation is in 3NF

1. **Functional Dependencies**

**3.1. county\_table**





3.2. covid

1. **Canned Queries**