P8105 Data Science Final Project

Group members :

Minjie Bao (mb4757), Phoebe Mo (km3624), Jianting Shi (js5095), Pangsibo Shen (ps3194), Yutian Luo (yl4608)

The tentative project title:

COVID-19 Nursing Home Seniors' Mortality Distribution in Republic and Democratic counties in New York State.

The motivation for this project:

Many News articles have reported high mortality rates in senior nursing homes across the U.S. Elderly people are much more vulnerable to COVID-19 compared to their younger counterparts. In this project, we focus on New York State and want to find out if there are uneven seniors' mortality distribution across different counties with different political standpoints.

The intended final products:

A website with research background, statistical analysis, data visualization and research outcomes.

The anticipated data sources:

- US nursing home data:
  https://data.cms.gov/stories/s/COVID-19-Nursing-Home-Data-Archived-Datasets/dyvx-s2qr
- NYC nursing home profile data: https://healthdata.gov/dataset/nursing-home-profile
- Presidential election data for NY

The planned analyses / visualizations / coding challenges:

We have multiple datasets, it will be challenging in the data cleaning part to merge all the datasets together. We are also going to perform a Chi-squared test of independence between influenza vaccination status and Covid-19 infection rate in senior nursing homes.

The planned timeline:

Read reference article and find related datasets	11/10
Data cleaning	11/15
Data exploratory analysis and results	11/15 - 12/1
Building websites and submit	12/5