Modelling the Pandemic

Sociodemographic predictors of COVID-19 impact in Chicago neighborhoods

Group: Bored Grad Yacht Club

Christopher Owen
cowen20@uic.edu
https://github.com/antennarius

Kazi Shahrukh Omar
komar3@uic.edu
https://github.com/komar41

Abdul Rafey Siddiqui
asiddi73@uic.edu
https://github.com/rafeyyyyy

Nguyen Hoa Pham
npham30@uic.edu
https://github.com/nhpham27

Gautam Kushwah
gkushw2@uic.edu
https://github.com/gautam-kushwah

Group repository: https://github.com/uic-cs418/cs418-spring22-bored-grad-yacht-club

Problem

Our "big idea": Use sociodemographic factors to predict the impact of Covid-19 in different geographic regions.

How we chose this problem:

- Existing CCVI model is only semi-quantitative
  - → Only uses rank-based data
  - → Predictive variables weighted equally
- Our aim is to create a more accurate model

What we want to answer:

- Can sociodemographic factors predict COVID-19 impact?
- Which of these factors are most important?

Why this is important:

- Improve COVID-19 resource and vaccination distribution
- Understand relationship between COVID-19 impact and other social disparities

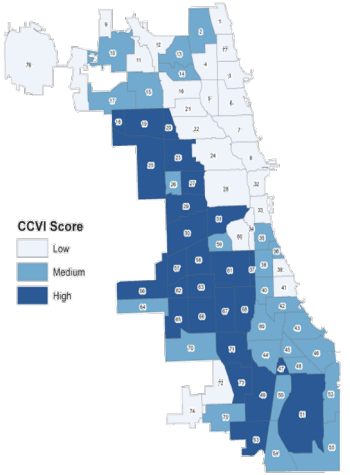

Fig 1: COVID-19 vulnerability index (CCVI score) in Chicago neighborhoods. Darker blue indicates more vulnerable areas.

Initial hypothesis: COVID-19 impact has strong correlation to factors like population density, age, health status and healthcare system.

Data

Socio-demographic data:

→ Chicago Metropolitan Agency for Planning (CMAP):

- ❑ Organized by neighborhood
- ❑ Contains 77 rows (one for each neighborhood)
- ❑ CSVs at https://datahub.cmap.illinois.gov/dataset/community-data-snapshots-raw-data
- ❑ Same data but with different geographical granularity is published by "US Census Bureau"
  - ❍ Number of rows depends on granularity (up to 2189)
  - ❍ CSVs at https://censusreporter.org/

COVID-19 data:

→ Cook County Medical Examiner data:

- ❑ Latitude, longitude, address, for COVID-19 related deaths
- ❑ 15,100 rows
- ❑ CSVs at https://datacatalog.cookcountyil.gov/Public-Safety/Medical-Examiner-Case-Archive-COVID-19-Related-Dea/3trz-enys

→ Chicago Data Portal:

- ❑ Weekly test results, case numbers, hospitalizations by zip code
- ❑ 727 rows total (1 per zip code per week since March 3, 2020)
- ❑ https://data.cityofchicago.org/browse?limitTo=datasets&sortBy=alpha&tags=covid-19

Solution

Our Approach:

- Visualize cleaned data
- Apply feature selection methods
- Find good set of features to predict Covid-19 impact on neighborhoods
- Build and compare different ML models

Desired end result:

- Visualize correlation between Covid-19 and Sociodemographic factors
- Accurate ML model to predict Covid-19 impact

Techniques to be used:

- EDA: NumPy, Pandas, Matplotlib, Plotly
- ML framework: scikit-learn

Scope of the project & Next Steps:

- Currently, we plan to do our exploration in Chicago neighborhoods.
- In future, we plan to extend our exploration to other major US cities.

Plan for progress report:

- Cleaned dataset, EDA, ML model.