

## Social Determinants of Health Project

BMI 6016 Spring 2023

- **Basic Info:** The project title, your names, e-mail addresses, UNIDs, a link to the project repository.
  - **Project Title:** "Social Determinants of Health (SDOH)"
  - **Project Repository:** [Github Link](#)
  - **Project Members:**
    - Wayne Richards
      - UNID: u0553192
      - Email: [wayne.richards@utah.edu](mailto:wayne.richards@utah.edu)
    - Tessa Price
      - UNID: u0731478
      - Email: [tessa.price@utah.edu](mailto:tessa.price@utah.edu)
    - Ayushi Sharma
      - UNID: u1369772
      - Email: [ayushi.sharma@utah.edu](mailto:ayushi.sharma@utah.edu)
    - Jackson Viscomi
      - UNID: u1437829
      - Email: [u1437829@umail.utah.edu](mailto:u1437829@umail.utah.edu)
    - Govinda Dhakal
      - UNID: u0861480
      - Email: [govinda.dhakal@utah.edu](mailto:govinda.dhakal@utah.edu)
- **Background and Motivation.** Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

Social Determinants of Health (SDOH) are the nonmedical factors that influence health outcomes. Often these data are not available at a personal level or recorded within the medical records. Attempts are being made in extracting SDOH data from publicly available resources such as the US Census. Since this is an emerging aspect of Health Informatics, our group felt this topic would be both interesting and relevant given that the comorbidities that were observed during the COVID-19 pandemic contributed significantly to health outcomes (both positive and negative).

- **Project Objectives.** Provide the primary questions you are trying to answer with your data. What would you like to learn and accomplish? List the benefits of how the data could be useful.

Our group wants to understand how SDOH in a person's environment might contribute to, or compound these health outcomes. Our group will attempt to generate SDOH data gathered

from New York City, NY, USA census data and other government sources. We will assess the spatial and temporal quality of this SDOH data and its applications to health outcomes and research, and its applicability to intervention.

1. **Spatial Accuracy**: Refers to the accuracy of the geographic location of the data.
    - We will compare the US Census data with other sources of geographic information such as maps, satellite imagery, and ground surveys. We will also use spatial analysis tools to identify any patterns or anomalies in the data that might indicate inaccuracies.
  2. **Temporal Accuracy**: Refers to the accuracy of the time period for which the data was collected.
    - We will compare the US Census data with other sources of temporal information such as historical records and time series data. We will also use time series analysis tools to identify any trends or patterns in the data that might indicate inaccuracies.
  3. **Completeness**: Refers to the extent to which the data captures all the relevant information for a given geographic location and time period.
    - We will compare the census data with other sources of information such as administrative records and survey data. We will also use descriptive statistics to identify any missing values or outliers in the data.
  4. **Consistency**: Refers to the consistency of the data across different geographic locations and time periods.
    - We will compare the census data from different geographic locations and time periods to identify any discrepancies or inconsistencies in the data.
- **Data**. From where and how are you collecting your data? If appropriate, provide a link to your data sources.

We will be collecting data from various resources: ACS, EPA, USDA, CDC, and US Census (<https://data.census.gov/>). Our project will also utilize the FACETS dataset, which can be found in the following GitHub repository: <https://github.com/mcantor2/FACETS>. The FACETS dataset contains SDOH data collected between 2016-2018 for New York City, NY, USA, and was gathered from US Census data and other government sources. We will use this data as a point of reference.

- **Data Processing**. Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented?

We do not anticipate substantial data cleanup as the data has been cleaned as it was aggregated. We will assess data completeness, accuracy, and consistency as we begin the data exploration and analysis phase. We will compare the US Census data with other sources of data such as administrative records and survey data. We will also compare the US census data from different geographic locations and periods to identify any discrepancies or inconsistencies. We will conduct a data quality assessment using one of the frameworks from module 2 (TBD) to ensure the data meets our intended project objectives.

- **Design.** How will you display your data? Provide some general ideas that you have for the design. Develop **one alternative prototype design for your data**. Describe your designs and justify your choices of visual encodings.

We will display the data using the following methods: Heat maps, and geo heat maps of the different FIPS areas for NYC over the three-year data collection period (broken down by year). Density clusters including density of moving objects if it's appropriate for the variable. Moving objects could be variables affecting *climate environment* or *population distribution and user behavior*. <https://iopscience.iop.org/article/10.1088/1742-6596/1792/1/012056> Maybe use Space-Time Autoregressive Integrated Moving Average or Spatial Multivariate Age-Period-Cohort (APC) Effects. (<https://www.publichealth.columbia.edu/research/population-health-methods/spatiotemporal-analysis>). Joining in historical COVID cases counts and morbidity at the census-tract level would enable us to check our SDOH features for correlation.

- **Must-Have Features.** List the features without which you would consider your project to be a failure.

Our group determined that the following SDOH's were most essential to our projects overall success: "FIPS", "Total Population", Races categorized into one column, "% citizen/non-citizen", "% schooling", "% High School/GED", "% AA/BA", "% limited english", "% in poverty", "Median Household Income", "Unemployment rate", "% Uninsured/Insured", "% Medicare/Medicaid", "Respiratory Hazard Index", "Low access to healthy food", "Park distance", "GINI". Double check with the SVI.

- **Optional Features.** List the features which you consider to be nice to have, but not critical.

Statistics such as sample size, standard deviation/variance and median for each feature would help us better understand the population and the considerations we must take when utilizing aggregated SDOH metrics rolled up to the Census Tract.

These are the features our group determined would be optional for our project: Urban/Rural Flag, % Other Race, % 2 or more Races, % Latino/Hispanic Ethnicity, % Native Born in US, % Foreign Born, % completed high school, no degree, % Some college, no degree, % Masters, professional, doctorate, % Other level of schooling (< High school), % Uninsured, <18 yo, % Uninsured, age 18-64, % Uninsured, over 65, % Any private insurance all ages, % Any public insurance all ages, Neighborhood walkability scale, Percentile of walkability (higher better), Tobacco retailers/1000 population, # of 7 serious crimes /1000 population, Housing violations per 1000 rental units, Voter Turnout, Turnout Quartile

- **Project Schedule.** Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

**Week 1 (02/13/2023)**: We will finalize the collection of data from our various data sources required for generating the SDOH data. Finally, we will check how much programming is required to integrate the various datasets into a single location and file.

*(Independent Work)*

**Week 2 (02/20/2023)**: We will conduct a data quality assessment to ensure that the data is Complete, Correct, Concordant, Current, and Plausible. We will then store the data into tables based on features.

*(Independent Work)*

**Week 3 (02/27/2023)**: We will implement any feedback from our Project Review Meeting.

**Project Review Meeting (03/02/2023)**

*(Independent Work)*

**Week 4 (03/06/2023)**: We will begin to organize the results into a powerpoint presentation that will be used for the final project submission along with the feedback gathered at various project milestone meetings with instructors.

*(Independent Work)*