Group Name: Spike Proteins
Patricia Zito, Matt Ludwig,
Sean Maclatchy, Emma Dyer
CS 12200
2 February 2020

Interactive Chicago Area Life Expectancy Calculator

Project Description

Our project will allow users to input personal information such as age, gender, race, income, and the Chicago neighborhood that they live in and receive an estimate of their life expectancy based on these factors. Once given this initial output, users will be able to interact with sliding scales to adjust their input factors and see how their life expectancy may change as they alter their original input factors.

Visualization

Users of the final product will be prompted to input a variety of information such as age, sex, income and locational data. The Matplot library and the Seaborn library will be used to display a plot with income as the independent variable and life-expectancy as the dependent variable. The fitted-model will take in additional parameters that the users has inputted and will be depicted on the plot. On the bottom of the plot a slider will allow the user to change their income level and see in real time how income changes could affect their estimated life-expectancy. The feasibility of adding sliders for non-continuous variables such as age and race will be explored with the possibility of changing the displayed independent variable on the graph. In addition, a map of Chicago centered on the user's location will display the nearest health centers within a specified radius. This will likely be achieved with the Matplotlib, Cartopy, and Geopandas libraries.

Approximate Methodology

We can use linear regression to get our best estimates for this problem. Some of the variables we might consider are listed above in the "Project Description" section. We can start by comparing simple linear regression models then see how considering more variables impacts the accuracy and potential overfitting. Perhaps truer to real life, we would consider interaction effects between certain variables in an attempt to improve the overall model.

Project Goal

Through this project, we hope to provide an engaging platform for individuals to observe how many societal factors may affect one's life expectancy. We hope to further elucidate factors that may provide greater contextualization for what recent research and reporting has termed the "life expectancy" gap across the city of Chicago.[1,2]

Data Sources (preliminary, may expand)

1. Life Expectancy by Zip Code
    - Centers for Disease Control Neighborhood Life Expectancy Project (2010-2015)
    - This data is grouped by 2010 census tracts and will be linked to corresponding Chicago neighborhoods during our data processing phase
2. Census Income Tables
    - United States Census Bureau

[1] "Chicago's lifespan gap: Streeterville residents live to 90. Englewood residents die at 60. Study finds it's the largest divide in the U.S." *The Chicago Tribune*. June 6, 2019. https://www.chicagotribune.com/business/ct-biz-chicago-has-largest-life-expectancy-gap-between-neighborhoods-20190605-story.html

[2] City Health Dashboard. NYU Langone Health. Metric Detail Life Expectancy Chicago, IL. https://www.cityhealthdashboard.com/il/chicago/metric-detail?metric=837&metricYearRange=2010-2015%2C+6+Year+Modeled+Estimate&dataRange=city.

    - This data is grouped by 2010 census tracts and will be linked to corresponding Chicago neighborhoods during our data processing phase
    - Will be used in life expectancy calculations and heat map visualizations
3. Primary Care Facilities (Chicago)
    - Chicago Data Portal
    - Will be used to plot the locations of primary care facilities on maps of Chicago
4. Life Expectancy Race/Ethnicity (Chicago)
    - Chicago Data Portal
    - Will be used in life expectancy calculations and in heat map visualizations

Timeline

4th Week (January 30th - February 5th)

1. Assemble group, put together proposal
2. Friday Feb. 5th proposal presentation

5th Week (February 6th - February 14th)

1. Data processing
    a. Collect and clean data from desired data sources
2. Begin statistical model testing and evaluation
    a. Review literature on life expectancy models
    b. Begin testing linear regression models and hierarchical regression models
3. Choose the specific visualization tools that we will be using from python libraries
    a. Seaborn
    b. Python widgets (sliders for user interface)

6th Week (February 15th - February 21st)

1. Complete any residual data processing
2. Statistical modeling
    a. Evaluation
3. Begin testing visualizations with selected libraries
    a. Heat maps
    b. Maps with primary care facility locations
    c. Sliding scale widgets

7th Week (February 22nd - Feb 28th)

1. Begin building interactive user interface
2. Finalize statistical models
    a. Begin writing documentation for our models
3. Continue work on visualizations

8th Week (March 1st - March 7th)

1. Continue building user interface
2. Continue work on visualizations
3. Continue statistical methods documentation

9th Week (March 8th - due date March 12th 4:00PM)

1. Finalize project and prepare for final deliverable, due the 12th at 4:00PM