INFSCI 510: Data Analysis

Dmitriy Babichencko

Group 8: Mary Rose O’Donnell, Zachary Hoffman, Karina Mehta

For our final project, we have chosen to look at public health (not COVID-19 related) in the United States. Using data sets from the Centers for Disease Control, we would like to find the correlation between physical activity levels, environmental factors, and overall life expectancy of Americans.

We think that our research question is interesting because it concerns something that everyone can benefit from. Knowing about life expectancy and seeing statistics about it displayed in visual format can help influence people to change behaviors that may negatively impact their life expectancy. Since this is a topic that has such a large impact on how people live their lives, we find it to be worth looking into further.

​​Our model would benefit the American general public. The model concerns overall health, and so it would be beneficial for American residents to observe our model and use it as insight in how to possibly improve their health. Our target audience is not limited to a certain age range or demographic, as it pertains to everyone, and will be useful in helping Americans make decisions about their health and future based on the data we present.

Our general plan for data analysis is to combine different datasets we obtain from the CDC. The datasets in particular relate to physical activity, life expectancy, and various environmental factors that could impact overall health. After combining these datasets and creating visualizations, we will then be able to find the correlation between certain factors. We will also observe trends that we come across in our data, further allowing us to build a model that the public can use.