Final Project Proposal

ALY 6015

April 27, 2021

**Group 8**

Molly White

Mark Tomaschek

Miranda Harvey

Charles Einsiedler

**Project Proposal:**

Group 8 is made up of Molly White, Mark Tomaschek, Miranda Harvey, and Charles Einsiedler. We will be using a dataset from The World Health Organization on life expectancy and potential predictor variables. This study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. The data was collected from 193 countries over years 2000 to 2015 from the World Health Organization and from the United Nations. The data set contains 22 columns and 2938 rows. Variables in the dataset include if the country is developed or developing, adult mortality rates, infant deaths, alcohol consumption per capita, expenditure on health as a percentage of GDP per capita, percentage of Hepatitis B immunization coverage amongst 1 year olds, the number of measles cases reported per 1000 people in the population, average BMI, the frequency of deaths under 5 per 1000 population, Polio immunization, government expenditure on health, DTP immunizations among 1 year olds, HIV/AIDS deaths, GDP, population of the country, prevalence of thinness by specified age groups, number of years of schooling, and human development index in terms of income composition of resources.

Questions we are hoping to answer with our analysis are:

* Do Infant and Adult mortality rates affect life expectancy?
* Is there a positive or negative correlation with life expectancy and eating habits, lifestyle, exercise, smoking, and drinking?
* Do densely populated countries tend to have lower life expectancy?
* Is there a significant impact of schooling on the lifespan of humans?
* What is the impact of Immunization coverage on life Expectancy?

We will be using statistical methods learned throughout ALY6015 in order to answer some of the above questions. Initial exploratory data analysis and descriptive statistics will be conducted to understand the life expectancy dataset. Linear regression, chi-square independence testing, and analysis of variance testing are some of the methods that may be used to understand the relationships among the variables in the dataset. The findings from our analysis will be presented in a final report as well as a presentation.

**References:**

*Life Expectancy (WHO)*. https://kaggle.com/kumarajarshi/life-expectancy-who. Accessed 25 Apr. 2021.