**Aaron Nelson Capstone**

**Executive Summary**

Embarking on the exploration of the enduring idea that life's only certainties are death and taxes, this project seeks to unveil whether there's a significant connection between individuals' tax contributions and their life expectancy. I am particularly intrigued by the prospect of understanding if there are variations in life expectancy between states with different tax burdens. Through the examination of various tax categories—federal, state, property, and sales, etc.—the aim is to determine how taxes might influence life expectancy. Beyond these considerations, the project also plans to delve into the intricate dynamics of how government spending from tax revenues might impact life expectancy, especially in critical areas like healthcare and social programs. Acknowledging the inherent complexity, this effort is driven by the desire to comprehensively grasp the relationship between taxes and life expectancy in a straightforward and accessible manner.

**Motivation**

My motivation for this project stems from a curiosity sparked by the commonly heard phrase "death and taxes." Repeated over the course of my life, like many others, I had simply accepted these concepts as inevitable certainties. However, upon reflecting on this expression, I began to question the underlying assumption. Taxes, ostensibly, are intended to enhance our quality of life. Therefore, the logical connection would suggest that an increase in taxes could lead to an improvement in quality of life, subsequently increasing life expectancy. If this assumption doesn't hold true, it raises the intriguing possibility that we might be figuratively taxed to death. In such a scenario, understanding how tax dollars are being spent becomes crucial to unraveling this complex relationship.

**Data Question**

 What is the relationship, if any, between taxes and regional mortality?

 Do higher or lower taxes influence life expectancy?

 How do specific types of taxes, such as federal, state, property, and sales taxes, affect life expectancy?

 In what ways does the government's allocation of tax money impact life expectancy?

**Minimum Viable Product (MVP)**

1. Create a storyboard using Tableau that shows if there is a correlation between different types of taxes, government revenue and expenditure, and regional life expectancy.
2. As best as possible, show correlation between taxes paid, government revenue, government expenditure, and regional mortality.
3. Include in the storyboard data a breakdown that shows what percent of tax money is spent on various categories such as medical services, road, police, housing, etc.
4. Have an interactive storyboard to allow users to see all reported data as is pertains to specifics regions.

**Schedule (through 1/4/2023)**

1. Get the Data (11/24/2023)
2. Clean & Explore the Data (12/4/2023 -12/08/2023)
3. Create Presentation of your Analysis (12/20/2023)
4. Internal demos (1/2/2024)
5. Demo Day!! (1/04/2024)

**Data Sources**

<https://www.census.gov/data/datasets/2019/econ/local/public-use-datasets.html>

<https://ghdx.healthdata.org/record/ihme-data/united-states-life-expectancy-by-county-race-ethnicity-2000-2019>

<https://www.irs.gov/statistics/soi-tax-stats-county-data-2019>

<https://www.kaggle.com.com/datasets/danofer/zipcodes-county-fips-crosswalk>

https://www.statsamerica.org/downloads/default.aspx

**Known Issues and Challenges**

* Data from <https://www.census.gov/data/datasets/2019/econ/local/public-use-datasets.html> and [https://www.irs.gov/statistics/soi-tax-stats-county-data-2019 are .xlsx](https://www.irs.gov/statistics/soi-tax-stats-county-data-2019%20are%20.xlsx) files. Minor alterations are required while in .xlsx file format prior to importing into python to simplify usability as python has problem handling columns with identical names.
* Income tax returns are typically broken down by tax brackets. Tax information reported at [SOI Tax Stats County Data 2019 | Internal Revenue Service (irs.gov)](https://www.irs.gov/statistics/soi-tax-stats-county-data-2019) aggregates these brackets and provides only total information for region. This is important because government spending is based on total tax revenue, regardless of sources and all citizens should enjoy government services equally. This decision however, limits the ability to analyze how life expectancy based on income bracket.
* Because tax rates change on a yearly basis, and are different between regions, this analysis has to be conducted using data from a single year. While there will be small changes affecting that year, because taxes are always changing, this still works better than looking at rates over multiple years where changes might be more drastic.
* Because of covid and its effect on mortality from 2020 onward, a tax year prior to 2020 was selected. For this reason, 2019 was elected as the data was both most complete as well as most recent.