<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/version/1>

Suicide is an issue that’s most likely been around since the beginning of time, the person closest to you could be suffering in silence and you would never know. The Centers for Disease Control and Prevention, in their weekly Morbidity and Mortality Report, announced an increase of 28.4% in the suicide rate for adults age 35 to 64 years, with the greatest increases above age 45. Specifically, rates of suicides in these groups increased from 13.7 to 17.6 per 100,000 from 1999 to 2010. Notably, this was a reversal of a previous trend towards decreased suicidality throughout the United States. Since this report was released, it hasn’t received a great deal of public attention. With a population that includes the “baby boomer” generation, this finding seems to be a possible commentary on the stresses and ailments affecting adults in 2013, with an estimated 42,000 Americans died by suicide in 2015, each one deeply affecting friends and family. Changes over the past decades in the economy and in the economic safety net leave many with far less security than their parents enjoyed, and economics is just a small trigger in the recent influxes of suicides. Very little data and information is available about people who attempt to take their life, and thus scientific exploration has been hampered. My research objective for this exploratory data analysis is to uncover trends in suicide rates and see what factors/variables correlate the most, with this knowledge we can prevent suicide triggers in the future. It is imperative that we understand and discover if there is any geological, social, or economic triggers that can be changed and lower the rate of suicides committed overall.

It is estimated that around 20% of global suicides are due to pesticide self-poisoning, most of which occur in rural agricultural areas in low- and middle-income countries. Other common methods of suicide are hanging and firearms. Knowledge of the most commonly used suicide methods is important to devise prevention strategies which have shown to be effective, such as restriction of access to means of suicide. Elizabeth Weinberg (MD), staff psychiatrist at Erikson Institute for Education and Research at Riggs, commented, “The best evidence we have available is that effective treatment of those at risk for committing suicide involves a human connection, an attempt to understand the nature of the problem, and the view of the patient as a person who exists in an interpersonal, social, and cultural context, as well as a medical and psychiatric context.” But one suffering from suicidal thoughts and tendencies often like to be alone and will not be surrounded by people all the time.

By engaging in research, we can work toward understanding suicide and suicidal behavior more fully and thereby uncover and pursue the best ways to reach and treat those who struggle with suicide while providing an evidence base that elevates the issue beyond awareness and moves us toward action.

The challenging aspect of this data science project is to determine if there is truly a correlation between suicide rates and the victims’ generation, sex, and year the victims suicide took place. I hypothesize that the generation a victim was born to and their and sex affects the number of suicides per country. The dataset I chose was a dataset that explored suicide rates in various countries between the years 1985 and 2016. Metadata for this dataset include: country, year, sex, age, number of suicides, population, suicides/per 100k population, gdp per capital, and the generation of the person who committed suicide. Dictionaries were created for the objects sex and generation in order to convert them from strings to int objects that could be analyzed and processed via visualizations. I added a number of new visualizations since the last EDA project and discovered some new statistics; for the line graph detailing the number of suicides per year it came as a right skewed graph before cutting off after the year 2015, the boomer generation, like before, continued to be an outlier even in my cluster/scatterplots investigating the relationship (this find goes hand in hand with the research elaborated on earlier from the morbid and mortality report) with them committing 2000 suicides in a brief period. Other than that, the scatterplots displayed patterns of even distribution, specifically the scatterplot entailing the relationship between suicide numbers and the year in which a suicide was taken place. The countries of Saint Kitts and Nevis and Dominica were also outliers as they both displayed, they had 0 suicides between 1985 to 2016, I first assumed it to be null data, but it seems to be just outliers to the rest of the data.

With my prediction objectives I just want to identify whether the number of suicides per year is a regression problem. To do that I utilized the performance metrics of: sum of squared errors, absolute squared errors, r-squared. In order to regularize my models and visualizations I created both a ridge linear model as well as a lasso linear model.