**IST 718**

**Big Data Analytics**

**Group Project: Update**

Nash Delcamp

Allison Roth

Keith Kunz

**Problem Specification:**

Suicide is thought to be a preventable cause of death and the reasons behind such an act may vary widely. There are many factors that could contribute to suicide rates including socioeconomic, political, geospatial or medical.

The goal of this analysis is to identify factors that contribute to the suicide rate in an attempt to better understand what drives the number of suicides and, potentially, how to reduce the unnecessary deaths.

**Observations:**

Through exploratory analysis, areas with the highest suicide rates have been identified (Lithuania, Russia, Sri Lanka, and Hungary). Plots and summary statistics have highlighted the skewness of suicide rates – most observations are very low – as well as some records with much greater suicide rates than the rest of the data. These records could provide greater insight into factors relating to suicide rates.

**Analysis:**

Next steps involve creating k-means clusters, multinomial regression models to predict suicide rates, and time series plots and forecasts.

**Recommendation:**

A dataset containing gun ownership for 2017 has been cleaned but still needs to be merged with the suicides dataset. The immediate next step is identifying which predictors to use in building models. Another step is to gather more data to further improve the analysis. This may be obtained from social media or available data sets.

# Works Cited

*Costs of Suicide*. (2019). Retrieved May 13, 2019, from Suicide Prevention Resource Center: https://www.sprc.org/about-suicide/costs/