**Stat 627 Group Project Plan**

**Project Title:** Global Homicides by Gender

**Group Members:** Barbara Broussard, Adrian Bogart, Rebecca Rogers

**Deliverable:** Poster and Presentation (Option A)

**Topic:** Our project will look at how to use machine learning models to predict future homicide rates by gender using regional or country intentional homicide rates/counts between 2000 – 2021.

**Questions of Interest:**

Regression: Can a machine learning application predict future death rates by gender when accounting for country/region and year?

Classification: Can a machine learning application predict the gender of the victim using country, region, homicide rates, and year?

**Approaches:** The project team will use linear regression, logistic regression, and/or KNN to analyze the data. To tune each model, the team will use LOOCV, jackknife, and bootstrap as the cross-validation methods.

**Group Responsibilities:**

* Whole Group: Identify two research questions and identify relevant variables. Search databases to find a dataset on global homicides and gender. This may require the combination of a few datasets. Peer review models, analyze respective models, and write summaries of results with plots or diagrams.
* Barbara: Merge datasets into a master dataset. Use linear and logistic regression to analyze data. Write a paragraph on methodology and lead bibliography and sourcing. Write project recommendations. Lead consolidation of code.
* Adrian: Clean data from the master dataset. Create a regression MLA model and use LOOCV or another cross-validation method to validate the model. Use a jackknife or bootstrap method to eliminate bias in the model. Write an executive summary. Lead physical creation of the poster.
* Rebecca: Use a logistic model or KNN to predict whether the individual victim will be male or female. Use a jackknife or bootstrap method to eliminate bias in the model. Lead digital creation of poster and final submission of the project.

**Data:** UN Office on Drugs and Crime – Intentional Homicide Statistics

<http://data.un.org/DocumentData.aspx?id=444>

**Variables:**

* Gender
* Country
* Region
* Year
* Numbers of intentional homicides (per 100,000 population) (by sex/gender)
* Rates of intentional homicides (per 100,000 population) (by sex/gender)
* Situational Context *[might add more control variables on related crime rates]*