BIOL 419

Jade Yang & Zhongyu Li

Murder case report

Our dataset is the homicide report from 1980 to 2014. Our goal of the project is to analysis the correlation between the characteristics of victim and perpetrator and the overall solving rate of murders cases among states and different years. For the correlation part, we made a graph showed that the amount of cases (from 1980 to 2014) of different relationship between victim and perpetrator (Figure 1). We found that among all the solved cases, the most abundant relationships are 126018 cases in acquaintance relation and 96593 case in stranger relationship. Within the close relationship, most abundant cases are these in which wife/ girlfriends were reported being killed by husband/ boyfriend. For the future work, we will use more diverse algorithms to analyze side by side correlation of the features for victim and perpetrator (Figure 2). Our hypothesis is that there would be high correlation between victim ethnicity and the perpetrator ethnicity and between the perpetrator race and the victim race.

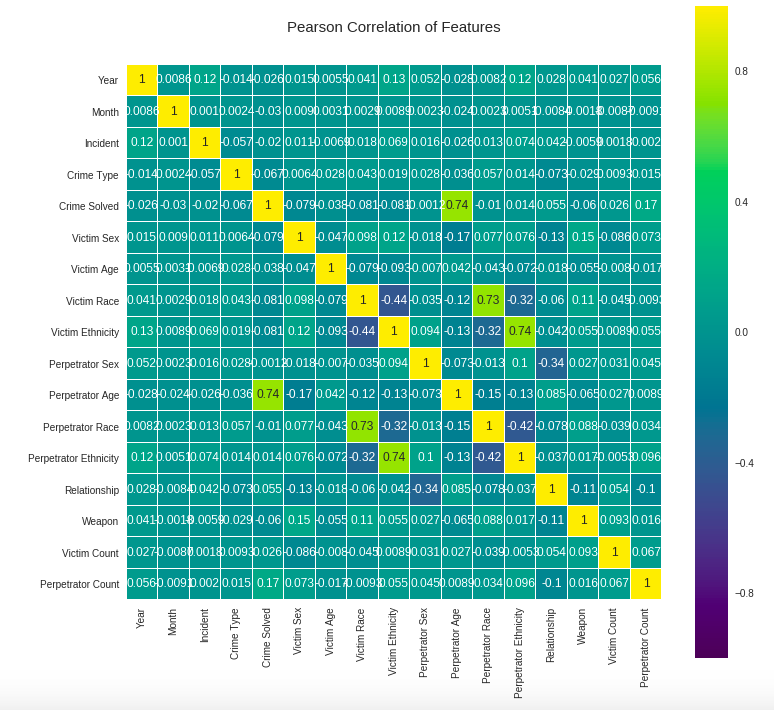
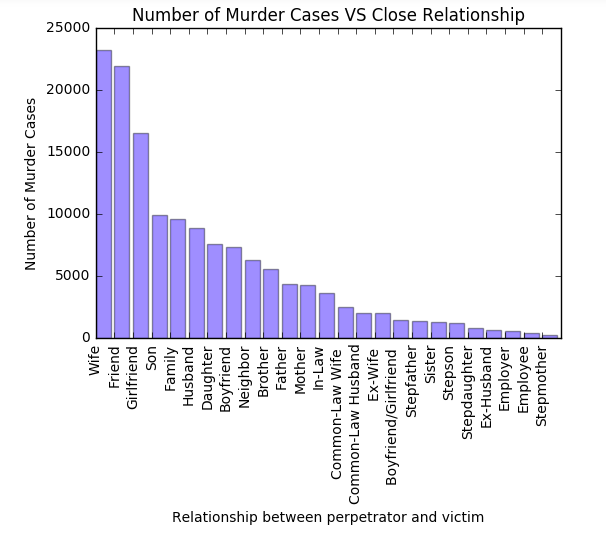
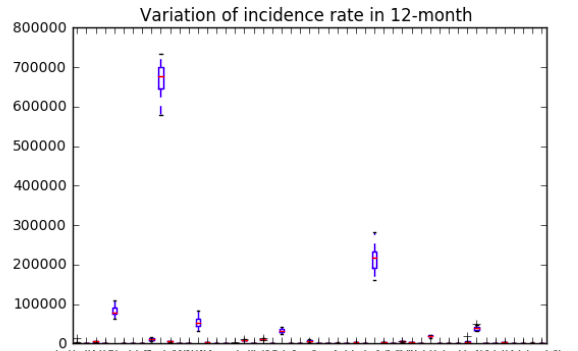


Figure1 Figure 2 https://www.kaggle.com/tomduff/d/murderaccountability/homicide-reports/crime-solving-rate-in-the-us/



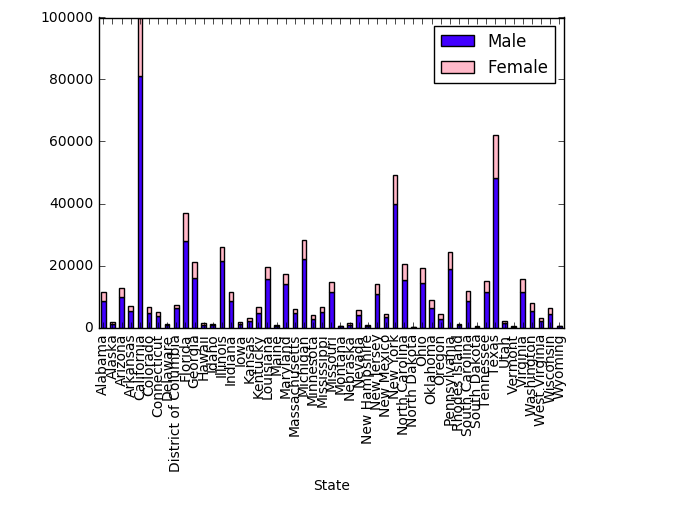
Moreover, to detect potential trends of reported incident rate, we reorganized the database into 12 months (Jan to Dec), including all states and recorded years (1980-2014). However, the boxplot showed very little variance in incident rate across 12-month in the majority of the states, which might indicate that possibility of homicide is not statistically related to a specific month (Figure 3). Hence, we should consider year as the minimum time unit. In addition, from the stacked bar plot, the likelihood of victim being a male is significantly higher than a female. Gender seems to play an important role in murder cases (Figure 3). In the future, we will make graphs of crime solving rate for victims in different age, race, and year. We hypothesize that some races, states, or years tend to have higher case solving rate than others.

Figure 3