Group Number: 41

Group Members: Yizhi Fang, Matt Welch, Jiamin Zhong, Jonathan Pickett

Title: Comparison of Campus Crime with Crime in the Surrounding City

Datasets:

1. On-campus arrest data from the 2009-2011 Campus Safety and Security Survey. The data can be downloaded at <https://ope.ed.gov/campussafety/#/datafile/list>. To get the data from 2009-2011, we chose year 2012 because the annual survey contained the previous three years data. The downloaded data folder contains 23 XLS files including on/off campus arrest, public property crime, etc. For this project, we will only use the on campus arrest dataset from 2009-2011. This is our large dataset with 10769 observations and 24 variables. Some of the important variables from this dataset include Institution Name, City, State, Total Population, Weapon Possession, and Drug and Liquor Violations.
2. FBI crime records in 2010. Some of the important variables in this dataset include State, City, Population, Violent Crime, and Property Crime. The Violent Crime is broken down into Manslaughter, Rape, Robbery, and Aggravated Assault. Property crime is further broken down into Burglary, Larceny-theft, and Motor Vehicle Theft. There is one final variable Arson, which we will not include in the dataset due to missing values. To get accurate comparisons, it will be necessary to create new rate variables created by dividing the crime variables by population. This dataset is located at <https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/tables/10tbl08.xls/view>.
3. We are also considering adding an income data set which has median income data for each state (if time permits). It would be be interesting to see how the income levels of each state correlate with crime. This data can be located at:

<https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_05_EST_GCT2001.US13PR&prodType=table>

Our purpose will be to look at how crime in the surrounding city compares to on campus weapon, drug, and alcohol violations across the country in 2010. And if time permits, we would also like to explore the relation between the crime rates and the income levels of each state.

We will read the raw XLS file and then create two permanent SAS datasets accordingly. Then we will clean all the datasets by checking for any missing, improperly formatted, or unreasonable values. We will only use the 2010 variable in the campus crime dataset since the FBI crime records are also in 2010. Eventually, we would like to merge the on-campus arrest dataset from 2009-2011 with the FBI crime records by their common variables State and City. It will be necessary to only include matches since likely there will be many cities without a corresponding university. We will not do any rigorous statistical analyses, except for proc means, freq, and other sorting procedures.