

**PROPOSAL OF PROJECT 2**  
**PROBABILITY AND STATISTICAL DATA ANALYSIS**  
**SECI / SCSI 2143**

Source of dataset (URL) :	<a href="https://www.kaggle.com/murderaccountability/homicide-reports">https://www.kaggle.com/murderaccountability/homicide-reports</a>
Dataset description (collected by who and for what purpose?) :	The data was compiled and made available by the Murder Accountability Project, founded by Thomas Hargrove. The dataset includes murder from the FBI's supplementary Homicide Report from 1976 to present and Freedom of Information Act data on more than 22,000 homicides that were not reported to the Justice Department. The purpose of this project is to provide details of each killing and detects the killing activity by perpetrator. This help the unsolved case and prevention of serious crime.
Variables (name of variables and type – categorical / ordinal / interval / ratio) :	<ol style="list-style-type: none"> <li>1. Age: nominal</li> <li>2. Race: nominal</li> <li>3. Gender: nominal</li> <li>4. Ethnicity: ordinal</li> <li>5. Relationship: Categorical</li> <li>6. Victim count: ratio</li> <li>7. Perpetrator count: ratio</li> <li>8. Weapon: categorical</li> <li>9. Type of crime: Categorical</li> </ol>
Description of purpose of study :	Crime in the United States recorded highest remark of homicide according to the report in 2013. The murder rate in US are higher than other four selected developed countries which the homicide rates is 0.8 out of 100,000. The homicide research is used to study the pattern of killer, common weapon used and the relation between the perpetrator and the victim to balance the crime coverage. Thus, the dataset of crime type, weapon used, relationship and other variables were recorded to help the case study.
Specification of target population :	The population of United States of America
Selection of variables (potential variables that will be selected for analysis) :	<ol style="list-style-type: none"> <li>1. Gender</li> <li>2. Relationship</li> <li>3. Weapon</li> <li>4. Type of Crime</li> </ol>
Proposed analysis (potential statistical test analysis related to the variables chosen) :	<ol style="list-style-type: none"> <li>1. Hypothesis testing</li> <li>2. Correlation</li> <li>3. Regression</li> <li>4. Chi Square test of independence</li> <li>5. Goodness of fit test</li> </ol>
Expected outcome for analysis :	Able to analyze the crime pattern of homicide for solved cases.

*Note: Submit this proposal within 2 weeks after briefing done by the lecturer.*