**2. THIS PROJECT CONSISTS OF 2 MAIN PARTS**

**Part A: Descriptive analysis of your variables.**

**A1.Title. Introduction. Explain the objectives of Part A.**

**A2.Provide an explanation of how the data was collected. If you used a survey, show the questions. If you obtained data from the internet, show/explain briefly the sources, and show in a table format variable name, detailed description, and units of measurement.**

**A3.What is your population of interest? Do you think your sample is a random sample? Is it representative of your population of interest?**

**A4.Describe each of the variables. This should include (i) graphical and (ii) numerical descriptive techniques that you learnt in this class, (iii) Pivot tables/charts. Pick only those that are most relevant to describing your data. Please avoid repetitions.**

**Experiment with a variety of descriptive tools that we learnt in class (including 3D maps and time series plots, wherever relevant). Please note that just providing figures and numbers is not enough – discuss your main findings.**

**Pan**

**A5.If any of your variables is qualitative, discuss the difference in descriptive statistics across groups using Pivot charts/tables.**

**A6.Briefly summarize your overall findings.**

**Part B: Analysis of association between your variables.**

**B1.Explain the objectives of Part B.**

**B2.Choose one or two dependent variable(s) (Y-variable(s)).**

Total victims

**B3.For each of your Y-variables, choose three or five variables as the explanatory variables (X-variables).**

Age, Gender?, open/close location, race(???), random target or not

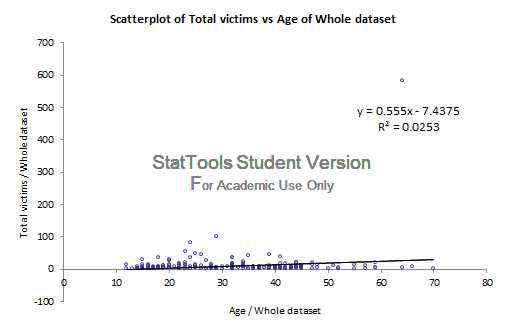
**B4.Explain why and how you expect your explanatory variables to explain/predict your dependent variable(s). Give examples where necessary.**

1. **Age: We want to see if younger killers tend to be cause more or less injuries and/or fatalities.**
2. **Gender: Although the overwhelming majority of the killers are male, we want to see if they tend to kill/injure more or not than female.**
3. **Random target or not:**
4. **Open space or not:**

**B5.Scatterplots and correlation analysis:**

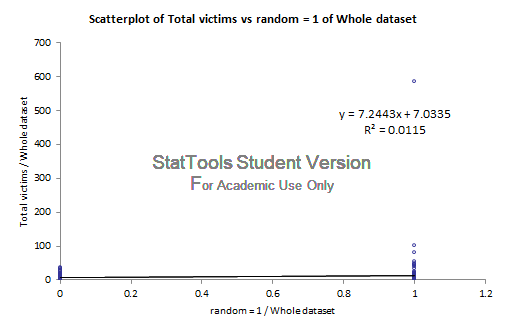
**• Create scatterplots for each of the dependent variables selected in B2 paired with relevant explanatory variables selected in B3. Describe what the scatterplots tell you about the association between your X-variables and the Y-variables.**

**Age of the shooter:**

****

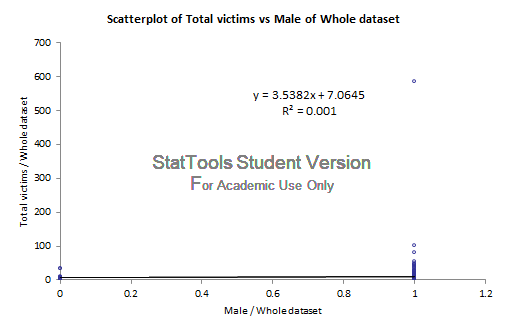
**The scatterplot displayed that the older the killer is, the more people he/she are likely to hurt and kill. However, the R-square are pretty low so it might not be a very good predictor.**

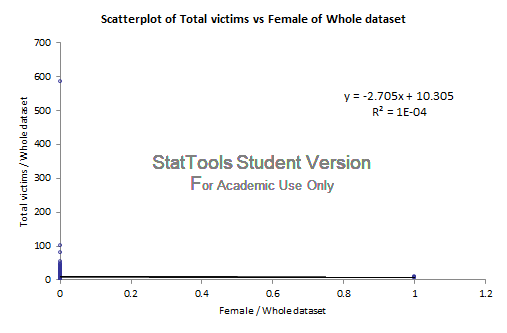
**Random Target or not:**

****

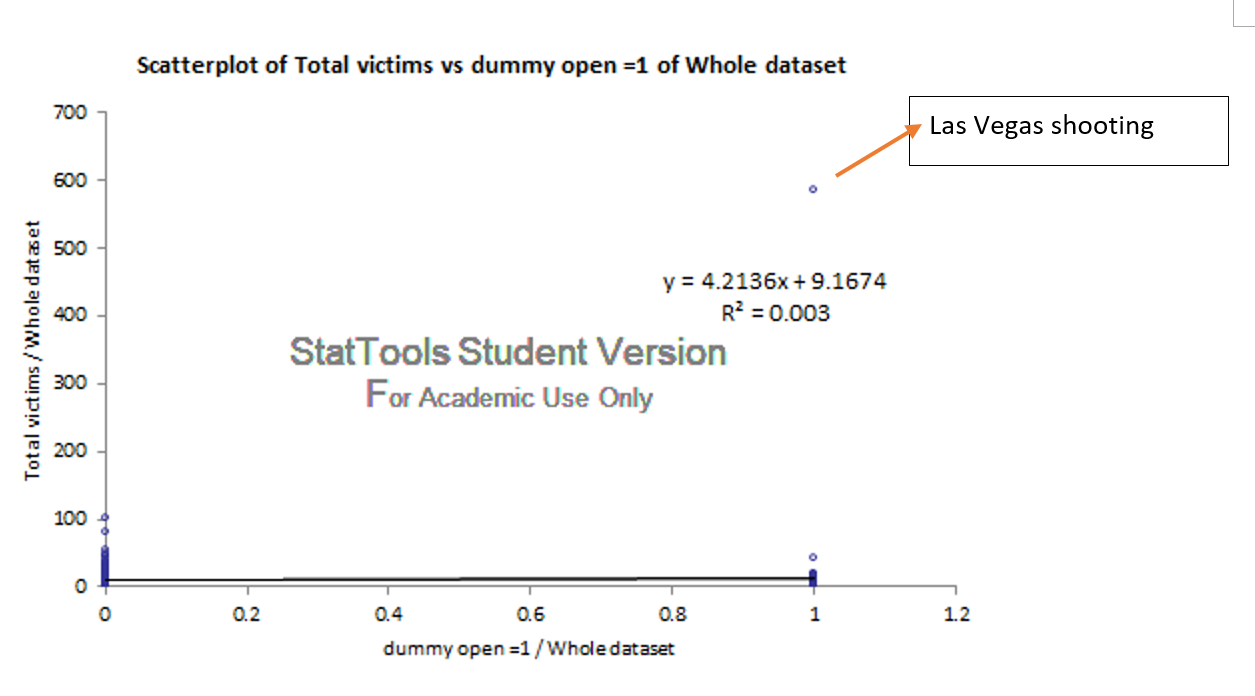
Intercept: If it’s a non-random shootout(dummy=0), then the total victim is predicted to be 7.0335  
 Slope: It is predicted that, random shoot-outs had 7.2443 more victims per incident...Why the hell am i doing this?

**Gender of the shooter:**

****

****

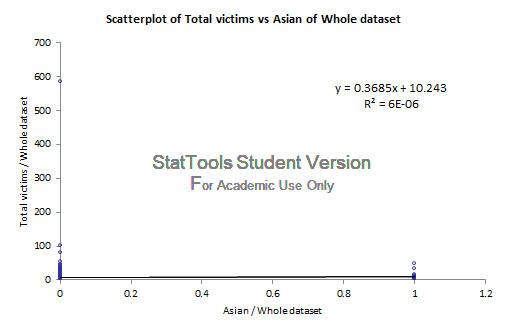
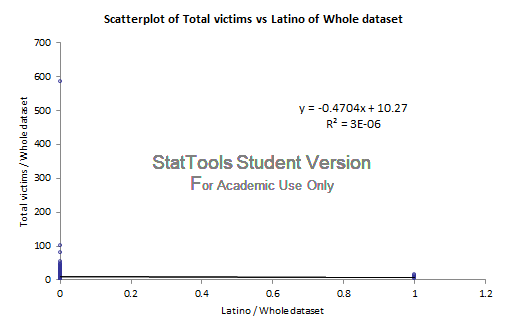
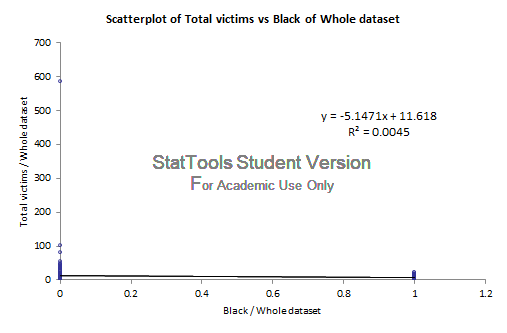
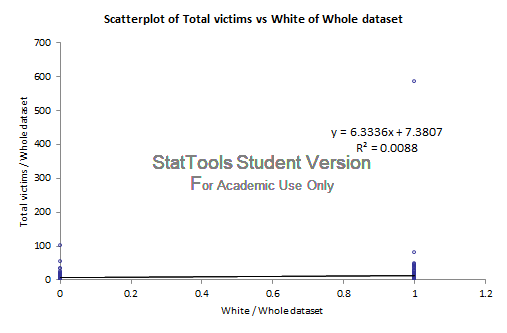
**Open or not:**

****

We cannot simply exclude that extremely high number on the top-right corner as an outlier, because it was not a data error. On the contrary, it’s very worth highlighting that this horrible incident was Las Vegas strip shooting which took place 2 months ago which resulted in that many victims.

Intercept: If the shootout took place in a closed area (dummy=0), then the total victim is predicted to be 9.1674  
 Slope: It is predicted that, shoot-outs that took places in the open area had 4.2136 more victims per incident.

**Race:**

****

**• Conduct corresponding correlation analysis and discuss the results.**

****

**• Compare the results with your predictions in part B4. Note that they may not necessarily match your expectations that you had.**

**B6.Conduct a multiple regression analysis to evaluate the relationship between the dependent variables and explanatory variables. You may use either an OLS or a Logistic model – whichever is relevant. Your models must include as explanatory variables:**

**• Numerical variables (continuous)**

**• Dummy variable(s)**

**• Interaction variable(s) (FUCK THIS)**

**B7.Do not forget to give full interpretations to the following: • Coefficients**

**• *p*-values (these will be discussed later during the semester)**

**• Evaluate your overall model - good!**

**• Predictions/forecasting.**

**B8.Conclusions. Summarize your findings and discuss the business/economic implications of your statistical results. Who may benefit from your findings? How may companies (or universities, or other relevant entities, depending on your project topic) use your results?**