



MASS SHOOTING

GROUP 5

ZAHRA ABBAS
BENEDICT FERRO
PRACHI DEDHIA
NILAVA MUKHERJEE
MICHAEL ROSSI
PAN CHEN

Introduction

- ▶ Mass shooting has always been a widely discussed topic. More than any other forms of crime, the intensity of this form of crime is worrisome.
- ▶ Mass shootings have a devastating impact on our communities from the victims killed, to the surviving witnesses, to the public at large. In order to prevent such tragic violence in the future, we must understand how and why these incidents unfold.
- ▶ We are trying to find the general trend and reasons or factors that might be responsible behind these horrible incidents

MASS SHOOTING

- ▶ PART A: Descriptive Analysis of the Data
 - Describing the source of data
 - Population of Interest
 - Variables Captured
- ▶ PART B: Analyzing the association between the variables
 - Initial Expectations
 - Identifying the dependent and Independent variables
 - Scatter Plots and Correlation Analysis
 - Multiple Regression Analysis

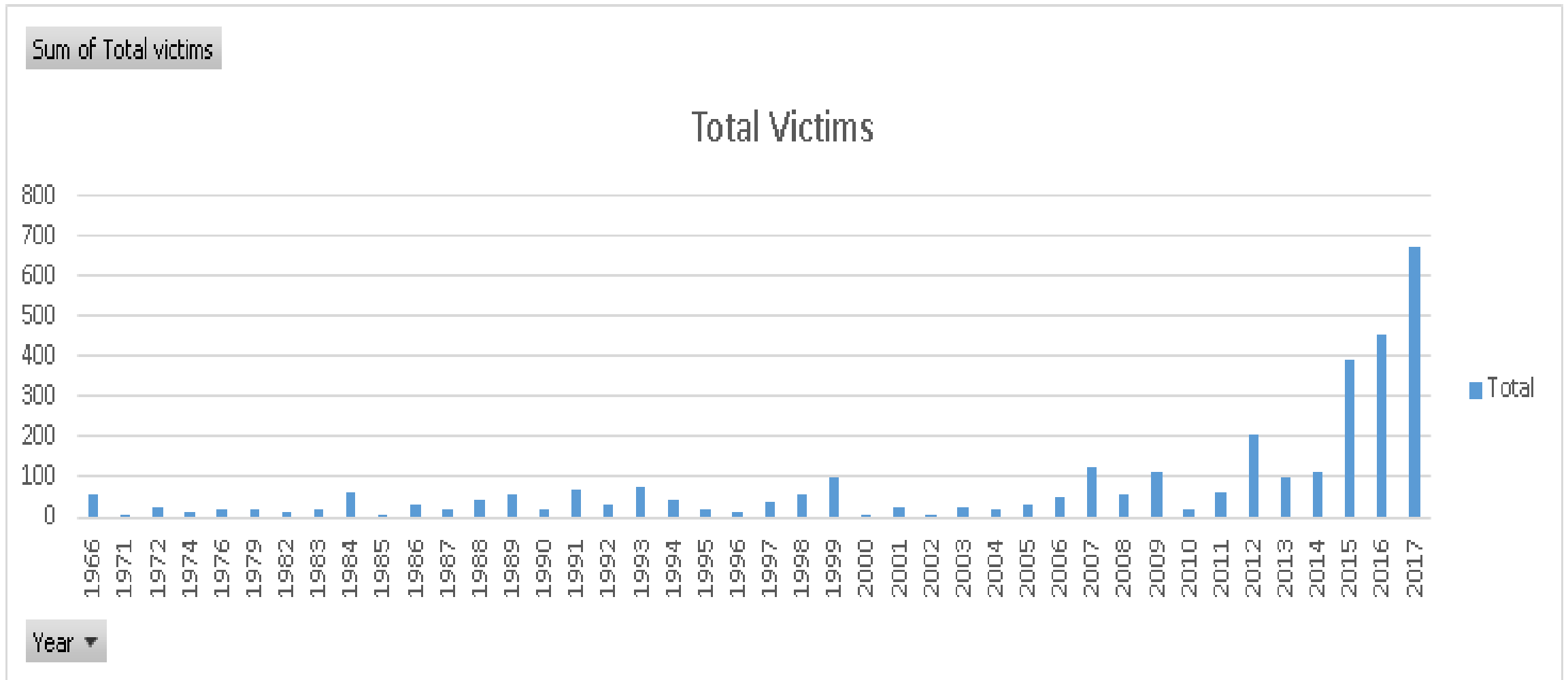
PART A: Descriptive Analysis of the Data

- ▶ Data Source and Collection method: Secondary method of data was used for this project. The data set was obtained from Kaggle.
(<https://www.kaggle.com/zusmani/us-mass-shootings-last-50-years>)
- ▶ Data Valid as of: November 1, 2017
- ▶ Data Set Size: 323 major shootings that took place in United States in the last 50 years.
- ▶ Population of Interest: General Citizens, Shooters, Police force, Gun control law makers
- ▶ Data Type: This is a cross sectional time series data. This is not a random sample since it includes the information on all the shooting incidents that have taken place since the last 50 years in United States

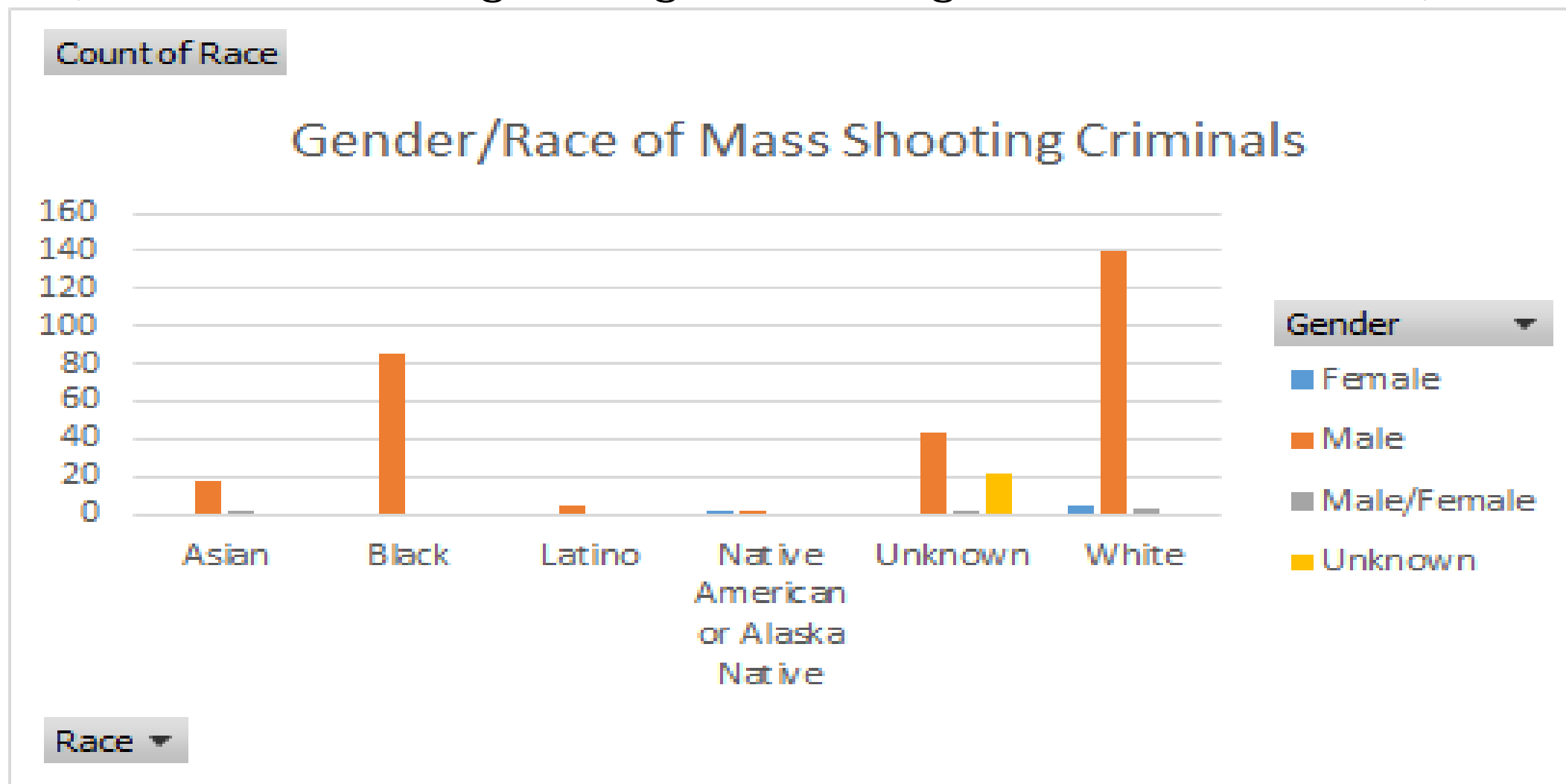
Definition of the Variables

Variable	Variable Type	Description
Title	Categorical - Nominal	Name and the place where the shooting incidents have taken place
City	Categorical – Nominal	Cities where the shootings have take place
State	Categorical – Nominal	States where the shootings have taken place
Date	Quantitative – Discrete	Date on which the shootings occurred
Incident Area	Categorical – Nominal	The area/ place where mass shooting took place
Open/Close Location	Categorical – Ordinal	Whether the location was an open public area or a closed location
Target	Categorical – Ordinal	The category of the people affected in the mass shooting
Cause	Categorical – Ordinal	The reason behind why the shooter conducted the shooting
Fatalities	Quantitative – Discrete	Number of deaths caused due to mass shooting incident
Injured	Quantitative – Discrete	Number of people being injured because of the mass shooting
Total Victims	Quantitative – Discrete	Total number of victims affected in the mass shooting (Fatalities + Injured)
Policemen Killed	Quantitative – Discrete	The number of policemen killed in the mass shooting incident
Age	Quantitative – Discrete	The age of the shooter
Employed	Categorical – Ordinal	Whether the shooter was employed or not
Employed at	Categorical – Nominal	If the shooter was employed, where was he/she employed
Mental Health Issue	Categorical – Ordinal	Whether the shooter has ay mental illness or not
Race	Categorical – Ordinal	The race of the shooter
Gender	Categorical – Ordinal	The gender of the shooter
Latitude	Quantitative - Continuous	The geographic angle at which the shooting took place
Longitude	Quantitative - Continuous	The geographic angle at which the shooting took place

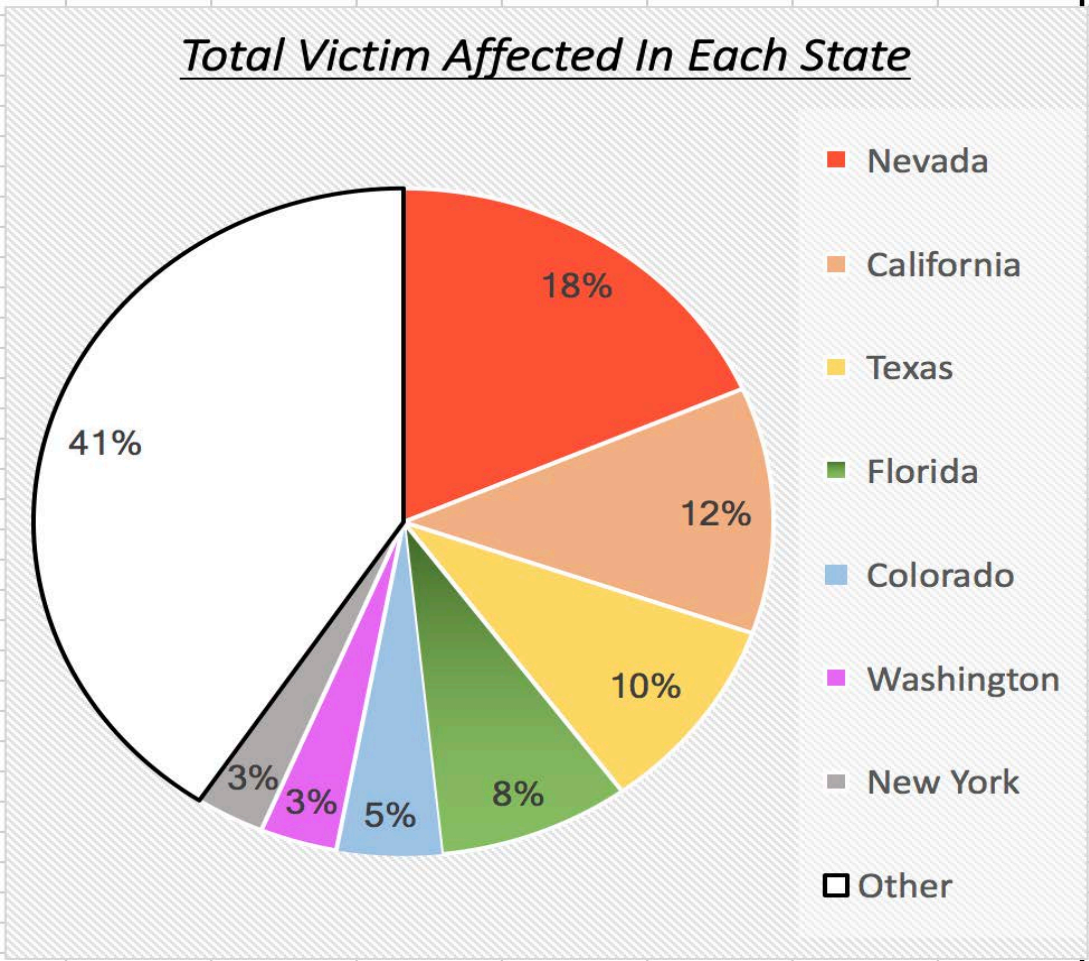
Increasing Trend of the Total Number of Victims from 1966 to 2017



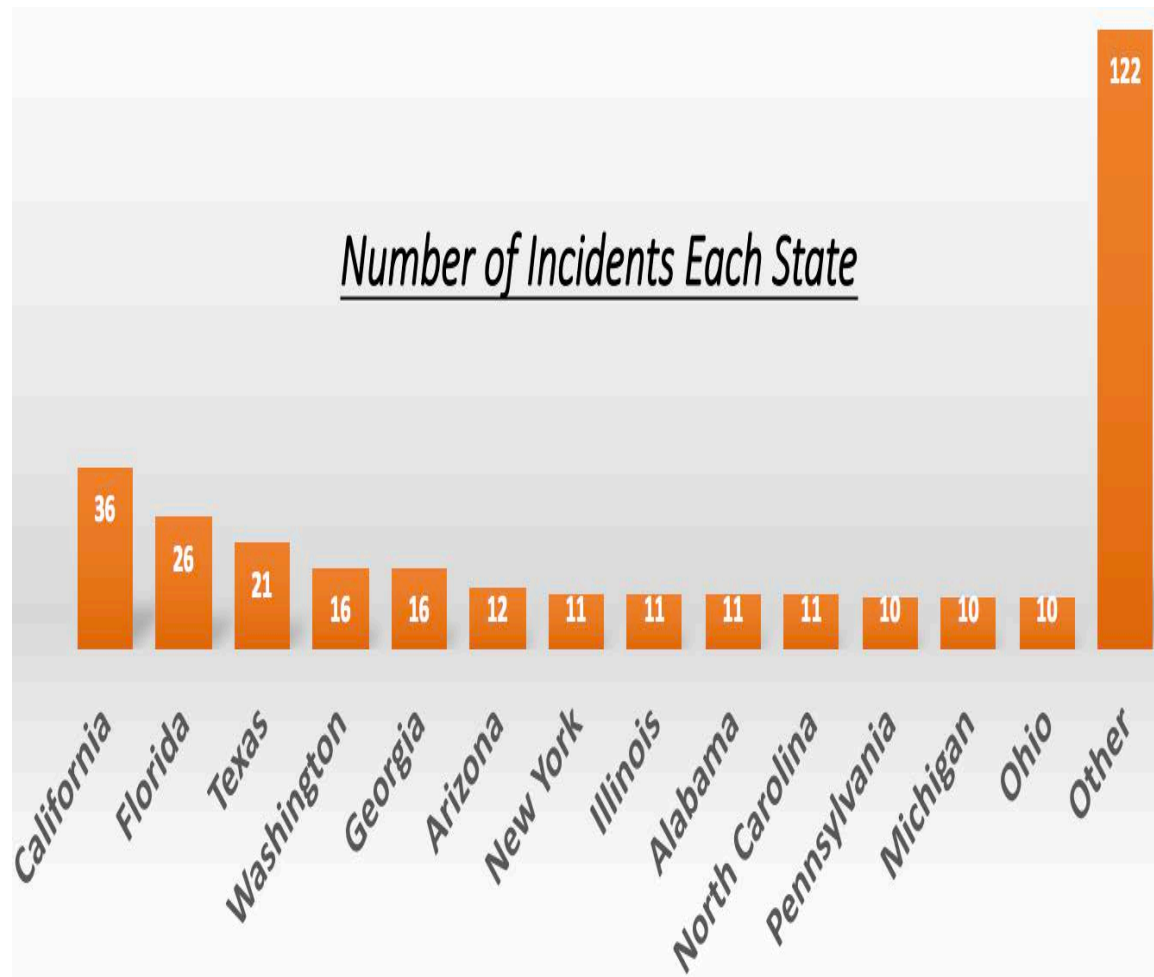
Breaking Down Mass Shooting Criminals by Gender and Race (White Males being the highest amongst the other shooters)



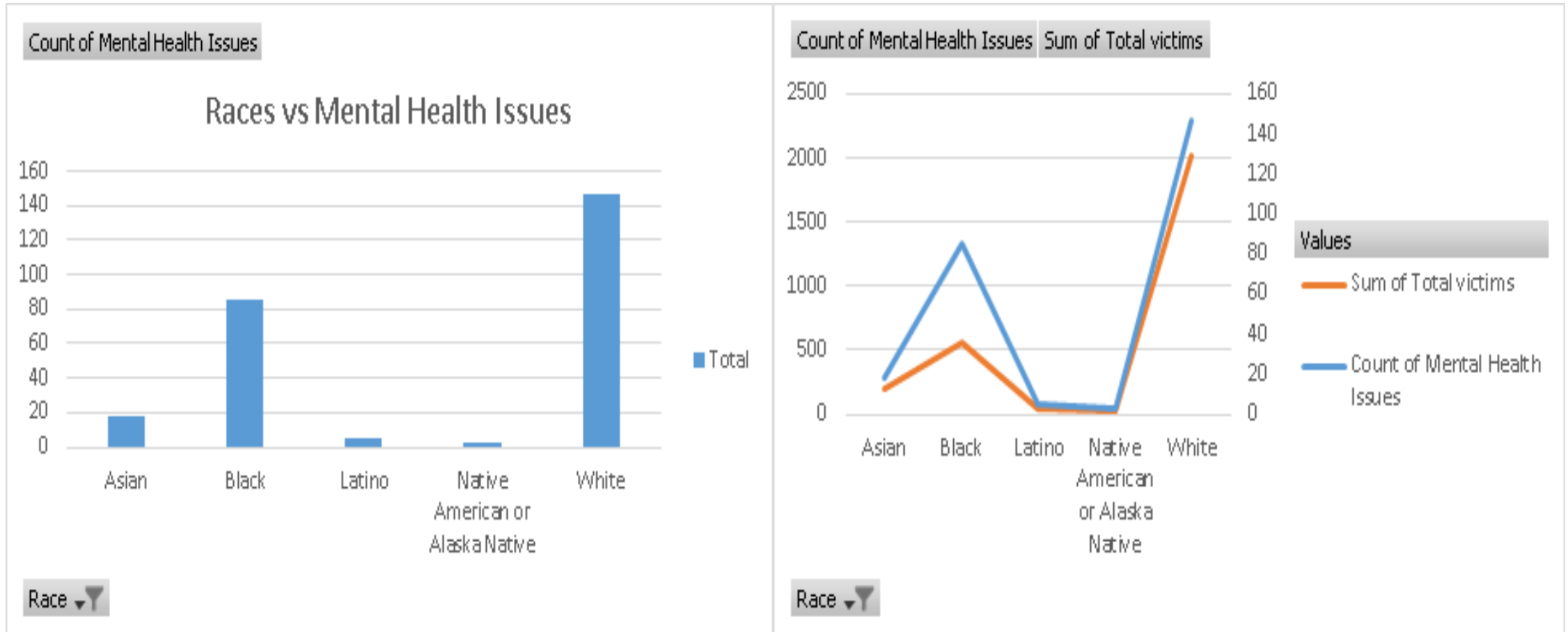
Total Number of Victims Affected – State
(Nevada observed the highest number of victims affected)



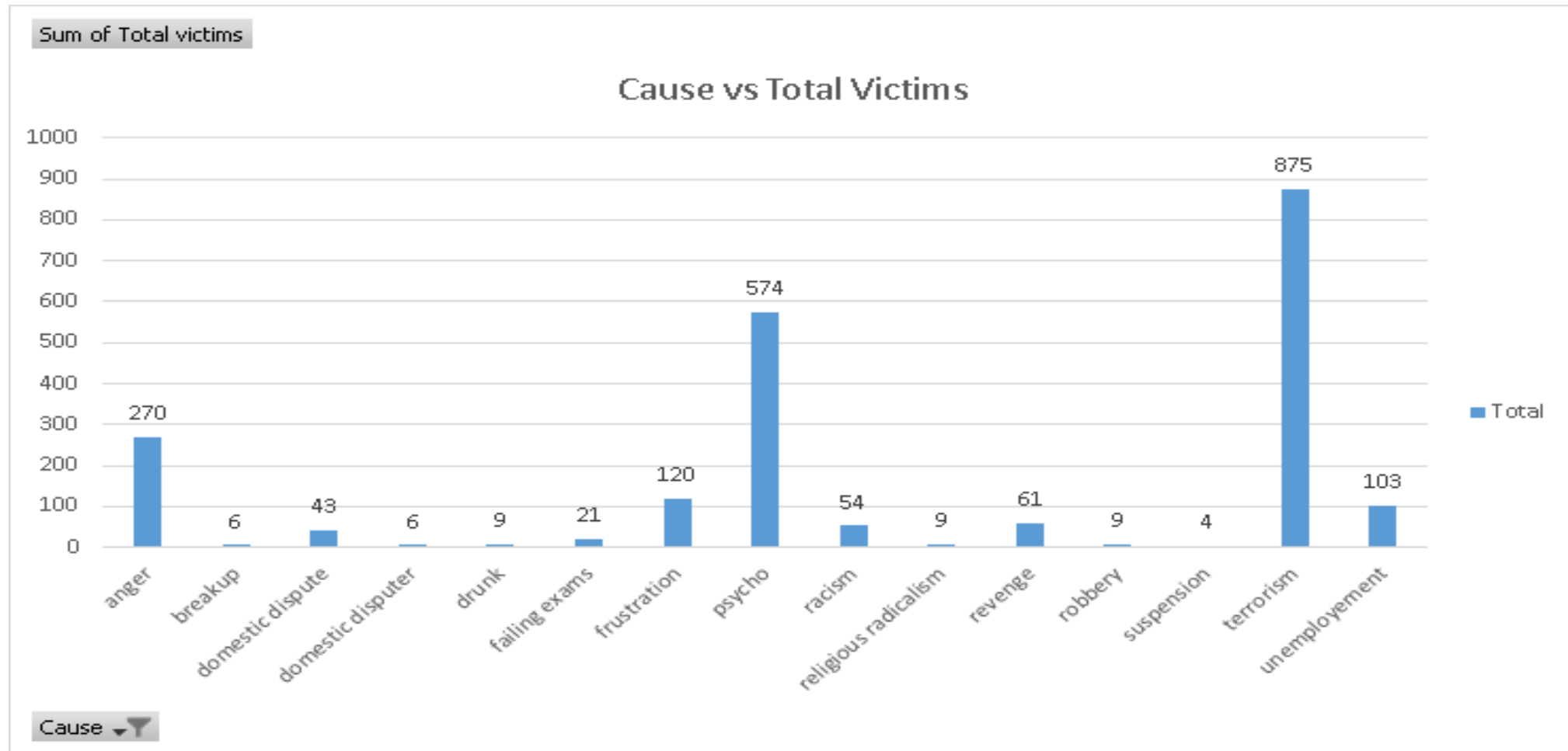
Total Number of Incidents – State
(California had the highest number of mass shooting incidents)



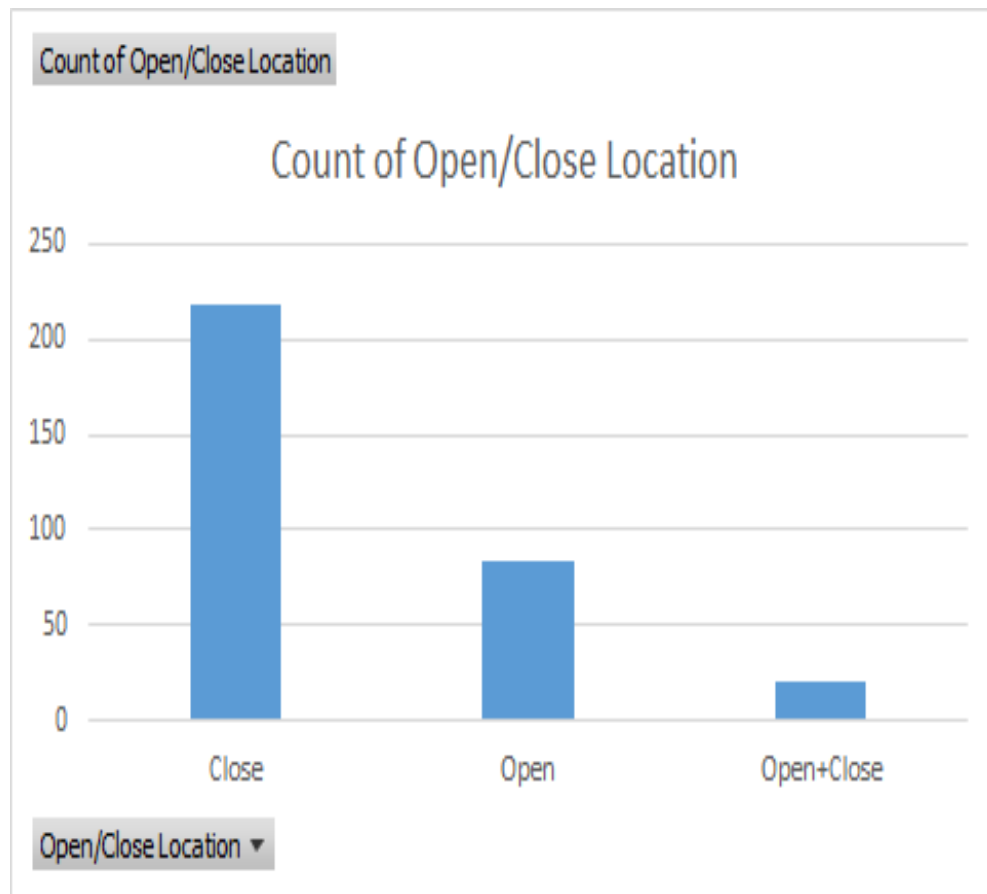
Mental Health Issues Observed Amongst Race (Most Mental Health Issues are from White followed by Black)



Terrorism, Psychotic Behavior and Anger are the Three Main Causes of Mass Shooting



Total Number of Open/Close/Open+Close Locations



Total Number of Victims Affected by Locations (Closed Locations were Main Targets for Mass Shootings)



PART B: Analyzing the association between the variables

► Initial Expectations:

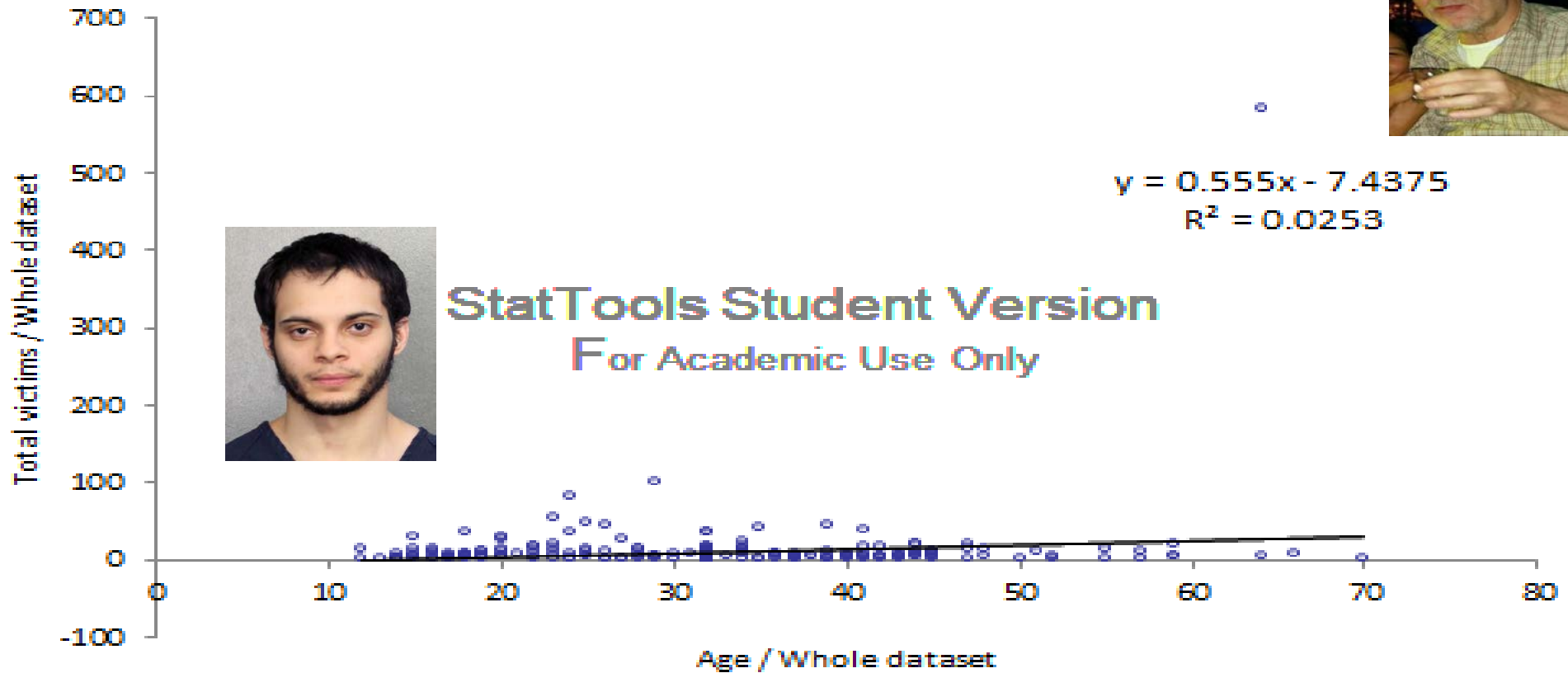
- Age: We want to see if younger shooters tend to cause more injuries and/or fatalities.
- Gender: We want to see if male shooters tend to kill/hurt more people than female shooters.
- Race: We don't have any expectations for this one. We just want to see if a shooter's race makes difference in the number of victim he/she caused
- Mental illness or not: We want to see if a shooter with mental illness would harm more people than the shooters without mental illness
- Open space or not: We want to see if a shooter shot in close space would kill/hurt more people than in open space.

Dependent and Independent Variables

Type of Variable	Variables
Dependent Variable	Total Number of Victims
Independent Variable	Shooter's Age Gender Race Mental Illness (Yes or No) Locations (Open, Close, Open + Close)
Dummy Variables	Locations (Open , Close) Race Gender Mental Health Illness
Interaction Variable	Race*Age

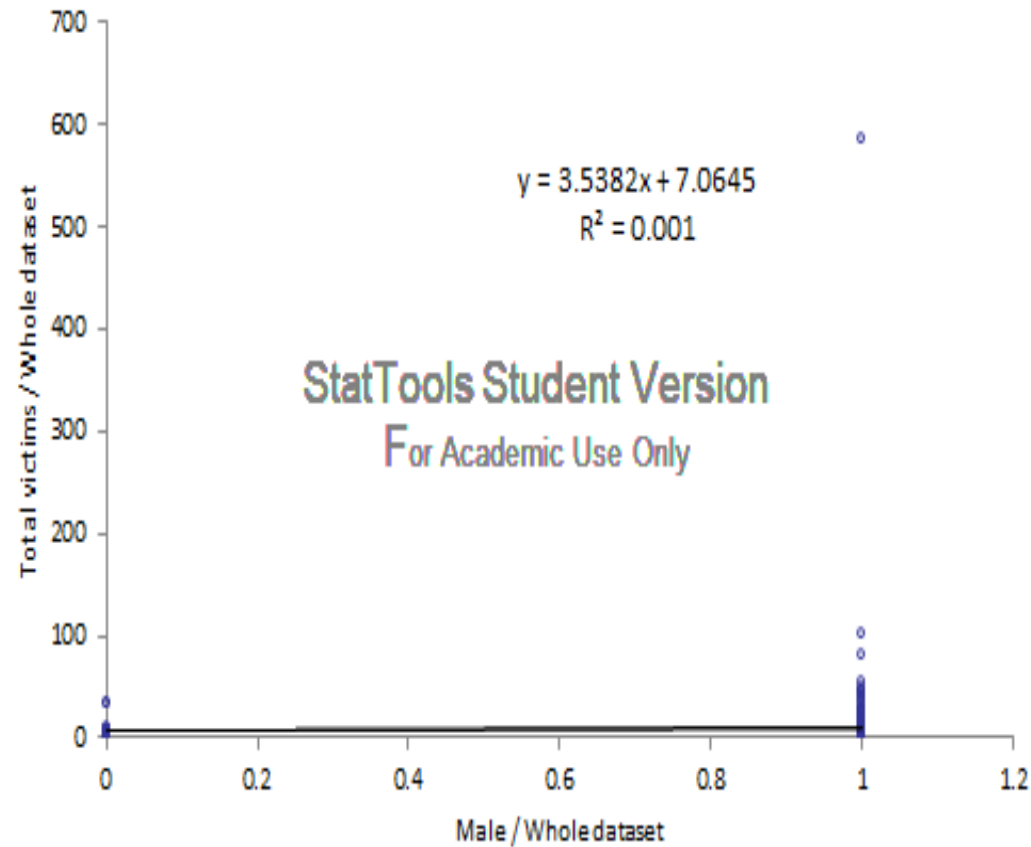
Total Number of Victims vs Age of the shooters

Scatterplot of Total victims vs Age of Whole dataset



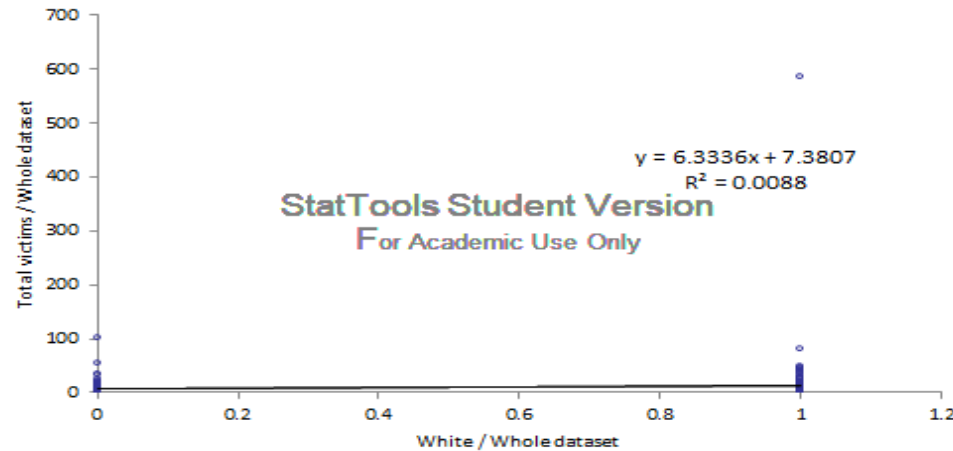
Correlation between Total number of Victims and Gender of the Shooter

Scatterplot of Total victims vs Male Shooter

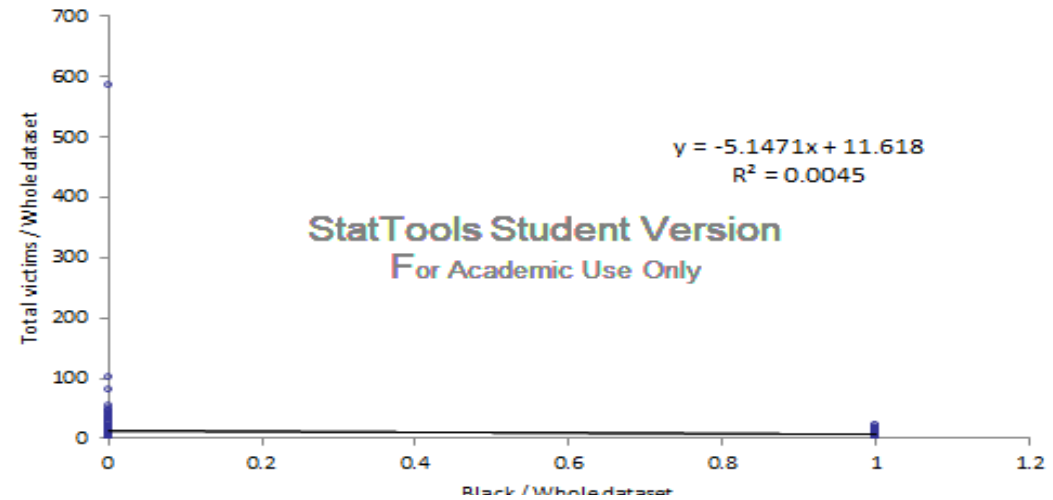


Correlation between Total number of Victims and Race of the Shooter

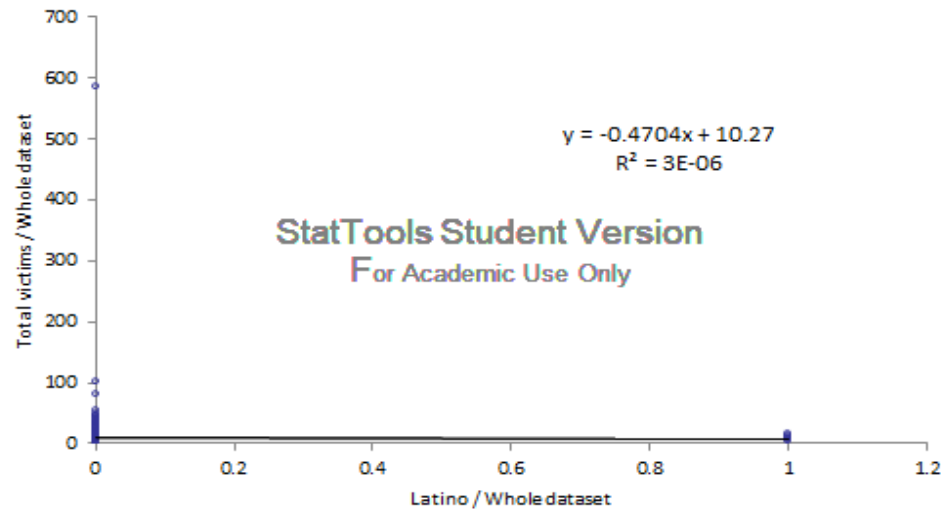
Scatterplot of Total victims vs White of Whole dataset



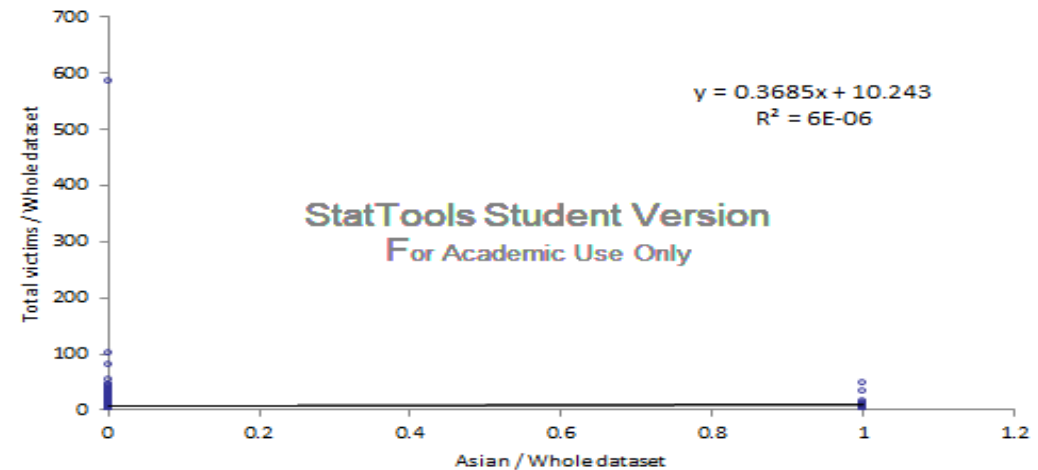
Scatterplot of Total victims vs Black of Whole dataset



Scatterplot of Total victims vs Latino of Whole dataset

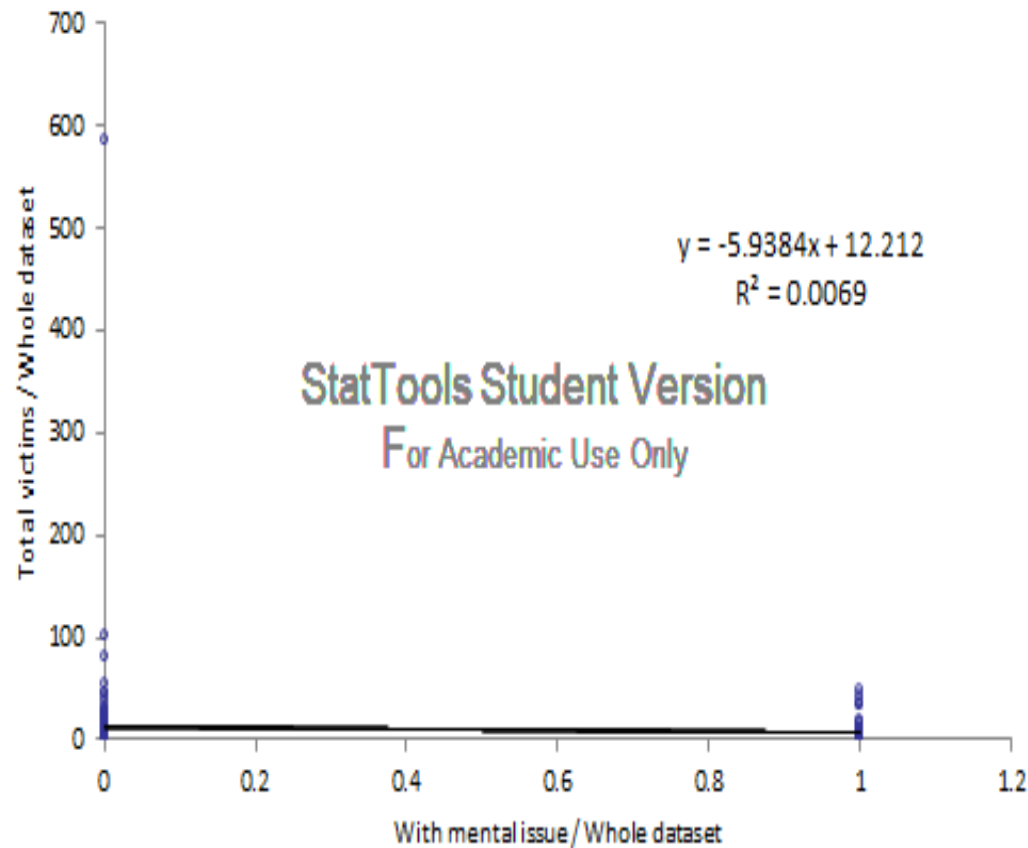


Scatterplot of Total victims vs Asian of Whole dataset

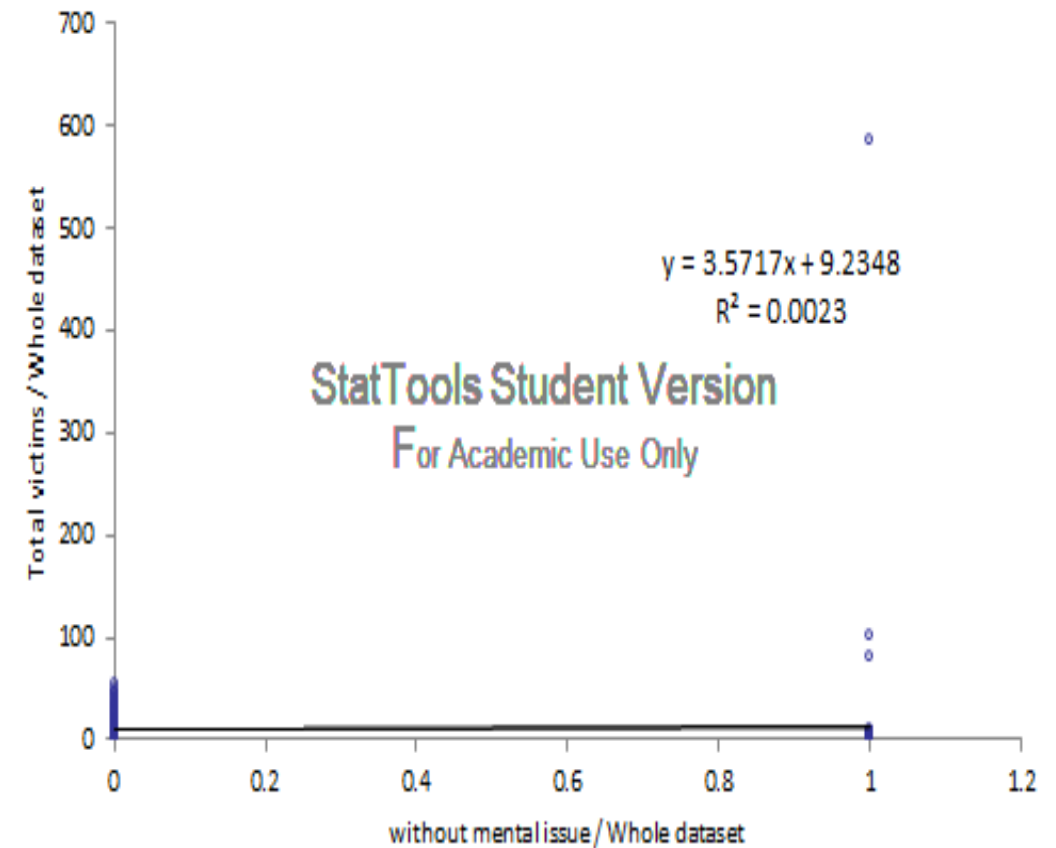


Correlation between the total number of victims and the mental health of the shooter

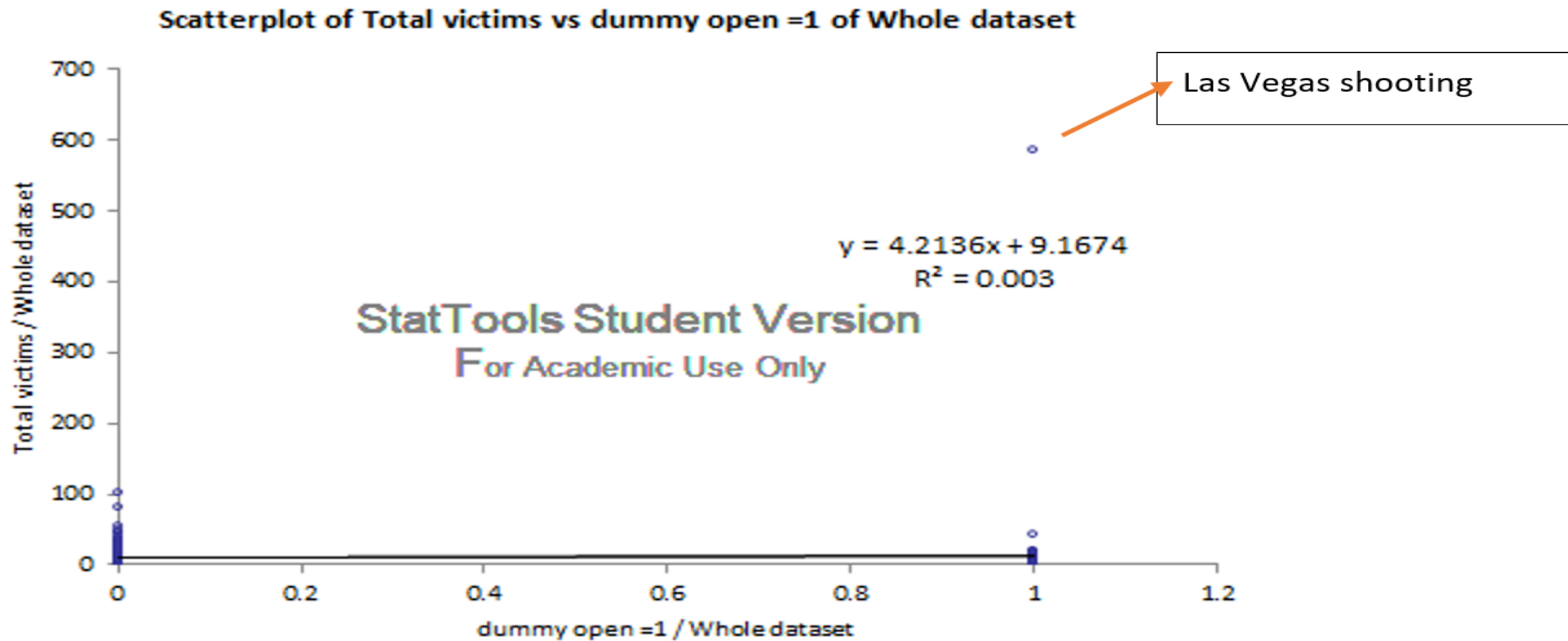
Scatterplot of Total victims vs With mental issue of Whole dataset



Scatterplot of Total victims vs without mental issue of Whole dataset



Correlation between Total number of victims and Location (Open/Close)



Correlation Analysis

	dummy open =1	Total victims	Age	White	Black	Latino	Asian	Male	Female	With mental issue	without mental issue
Linear Correlation	Plot Area	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset	Whole dataset
dummy open =1	1.000										
Total victims	0.055	1.000									
Age	0.050	0.159	1.000								
White	-0.216	0.094	-0.016	1.000							
Black	0.046	-0.067	-0.085	-0.024	1.000						
Latino	-0.074	-0.002	0.003	0.036	-0.075	1.000					
Asian	-0.082	0.003	0.044	-0.005	-0.145	-0.030	1.000				
Male	-0.238	0.031	-0.017	0.150	0.052	0.041	0.033	1.000			
Female	-0.017	-0.010	0.048	0.087	-0.075	-0.016	-0.030	-0.385	1.000		
With mental issue	0.037	-0.083	-0.075	-0.122	0.032	-0.034	-0.055	-0.063	-0.088	1.000	
without mental issue	0.028	0.048	0.037	0.037	0.024	-0.080	-0.035	-0.048	0.031	-0.444	1.000

1. Age and Total number of victims have the highest correlation. However, the correlation not being very high indicates that there may be other important factors that would strongly represent the relationship with Total number of victims.

Multiple Regression Analysis

<i>Multiple Regression for Total victims Summary</i>	Multiple R	R-Square	Adjusted R-square	Std. Err. of Estimate	Rows Ignored	Outliers
	0.2389	0.0571	0.0299	33.15440725	0	0
<i>ANOVA Table</i>	Degrees of Freedom	Sum of Squares	Mean of Squares	F	p-Value	
Explained	9	20820.42415	2313.380461	2.104575583	0.0289	
Unexplained	313	344054.2074	1099.21472			
<i>Regression Table</i>	Coefficient	Standard Error	t-Value	p-Value	Confidence Interval 95%	
					Lower	Upper
Constant	5.148792735	12.28511241	0.419108313	0.6754	-19.02305071	29.32063618
Age	0.109195785	0.375121177	0.291094696	0.7712	-0.628882146	0.847273716
White	-20.0560942	13.9549352	-1.437204395	0.1517	-47.51343423	7.401245837
Black	10.93462317	16.65303291	0.656614517	0.5119	-21.83141842	43.70066476
Latino	23.98068897	67.36098573	0.356002643	0.7221	-108.5569018	156.5182797
Asian	34.24717778	33.07905886	1.035312943	0.3013	-30.83825286	99.33260842
White*Age	0.818174743	0.422169023	1.938026474	0.0535	-0.012473214	1.648822701
Black*age	-0.474561492	0.522752165	-0.907813538	0.3647	-1.503114024	0.55399104
Latino*age	-0.879467014	2.041449224	-0.430805236	0.6669	-4.896165386	3.137231358
Asian*age	-1.045732233	0.956662623	-1.093104515	0.2752	-2.928034843	0.836570376

$$\hat{Y} = 5.15 + 0.11 \cdot \text{age} - 20.01 \cdot \text{White} + 10.93 \cdot \text{Black} + 23.98 \cdot \text{Latino} + 34.25 \cdot \text{Asian} + 0.82 \cdot \text{White} \cdot \text{Age} - 0.46 \cdot \text{Black} \cdot \text{age} - 0.88 \cdot \text{Latino} \cdot \text{age} - 1.05 \cdot \text{Asian} \cdot \text{Age}$$

Conclusions

- ▶ There are other factors which might be more important than the ones we have analyzed from the present data set.
- ▶ The bigger problem here is not related to a particular race, age or mental issues, but factor like gun control laws might be more substantial.