Factors Influencing Homicides in the United States

Submitted by:

Leighann Irving

Brittany Milligan

Shradha Pewekar

Kavin Veerapandian

Executive Summary:

The number of homicides, both solved and unsolved, that occur each year in the United States can be quite staggering. It was seeing these statistics that drove our interest to determine additional factors leading to homicide. To pursue this analysis, data was sourced and combined from: Murder Accountability Project (MAP), United States Department of Justice Bureau of Alcohol, Tobaccos, Firearms and Explosives (ATF), Substance Abuse and Mental Health Services Administration (SAMHSA), Federal Bureau of Investigation, Uniform Crime Reporting (FBI: UCR), United States Census Bureau, and The State Science & Technology Institute (SSTI).

The additional factors we decided to study and include in our dataset include: number of homicides, number of guns, number of drug users, number of alcohol abusers, number of law enforcement employees (term used interchangeably with "police" in report), and real GDP per capita. Each of the variables was given by state and year. In order to normalize all the data sourced, each of the variables was converted to per capita.

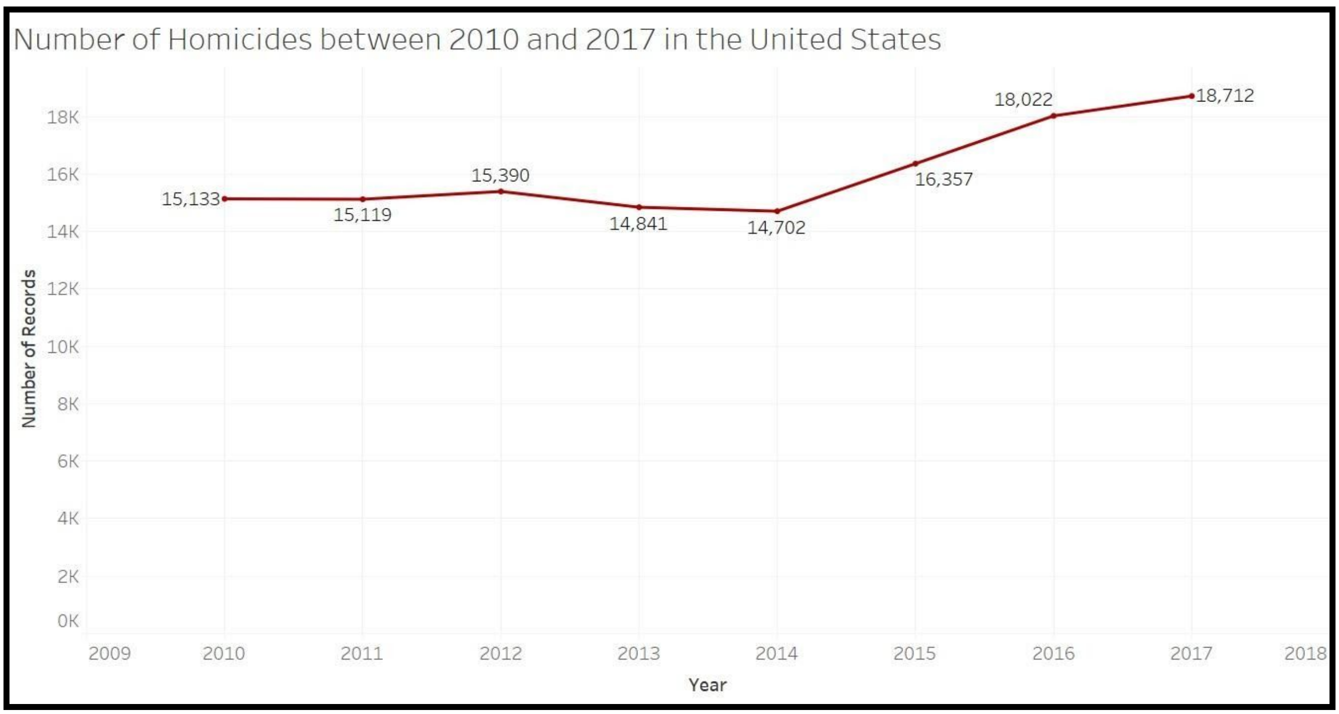
Once the data was combined into a single dataset, exploratory analysis and visualizations of the data was completed. Following this, hypotheses of the results we expected from our models were created and regression analysis on the data was conducted. In total, three models were created to test the hypotheses and the model that best fit the data and was most practical was determined.

Table of Contents:

Problem Definition and Significance:

The United States does not have a designated agency requiring states to report the number of homicides, whether they are solved or remain unsolved. The Federal Bureau of Investigation (FBI) does create a Supplemental Homicide Report (SHR) from the states that report their homicide information. However, it is voluntary to report this information to the FBI, so some states such as Alabama and Florida, do not report their homicide information to the FBI. This is where the Murder Accountability Project (MAP) was created to better educate the public in regards to unsolved homicides, as well as improve the accounting of these unsolved homicides. MAP used Freedom of Information laws to get states that do not voluntarily report to the FBI to report directly to them and therefore have a more complete dataset of homicides.

The number of homicides from 2010 to 2017 has steadily increased in the United States, as can be seen in the chart below. On average, only about 70% of these homicides are solved[3]. This means that many offenders walk away free and no justice is brought to the family or friends of the victim(s). This project will attempt to discover factors that influence the number of homicides, those that reduce the number of homicides, and those that have no effect on homicides. Efforts could be made to reduce factors increasing homicides and increase factors reducing homicides when they are better understood and studied.



Prior Literature:

Studies attempting to find significant factors influencing homicides have been conducted for years. During the 1990s, there was a significant drop in violent crime, and therefore homicides as well, and many studies have focused on this time period to try and determine why. Typically these studies focus on just a few factors and how they influence homicide, rather than a wide array of factors. For instance, some of the most commonly looked at factors in regards to homicides are drug and alcohol use. In a study by William Wieczorek, John Welte and Ernest Abel, it was found that the majority of homicide offenders engaged in excessive alcohol consumption or a combination of excessive alcohol and drug consumption immediately prior to committing the violent act[(5)].

Individuals involved in only excessive drug use were least likely to commit homicide when compared to excessive alcohol use or a combination of excessive alcohol and drug use[(5)]. Many of the drug related homicides were a result of systemic violence: territorial disputes among dealers, distribution of drugs, retribution for bad debts or drugs, or dealers being robbed [(5)]. This coincides with the ever-changing drug market and demand for drugs. For instance, part of the decline in homicides during the 1990s can be attributed to the decline in crack cocaine[(1)]. Crack cocaine is a cheaper illegal drug that allows lower income individuals to purchase it. This in turn created a greater and higher demand for crack cocaine and therefore more sellers needed to be recruited[(1)]. These new sellers were young and as the population of both the buyers and sellers increased in age, demand for the drug decreased, bringing down the number of drug related homicides during the 1990s[(1)].

An article written by Steven Levitt discusses factors that are considered significant in regards to homicide and crime. He argues that an increase in the number of police employees is effective in reducing violent crime and homicides[(2)]. Throughout the United States, the average increase in police officers was roughly 14 percent during the 1990s and was the largest increase compared to previous decades[(2)]. Based on the findings of the article, it was found that this increase in police officers accounted for roughly a 5 to 6 percent decrease in crime[(2)]. When it comes to fighting and deterring crime, police employees are the first line of defense. So, increasing the presence of police officers, should in fact lead to a decrease in crime and homicides as Levitt argues. The reverse is also true though, where an increase in crime would lead to an increase in police officers. Levitt focused on the 1990s however, which saw a drastic increase in police officers and a decrease in all types of crime all over the country.

Criminologists and economists differ in their opinions on how the economy influences homicides. Criminologists argue that in poor economic times, individuals are more likely to commit violent crimes[(4)]. In a poor economy, people may be unable to buy items they need, so instead they may choose to steal it. This in turn could lead to an increase in violent crime and homicides because the individual will take drastic measures to acquire what they need and the person being offended will take drastic measures to protect their personal property. Economists however, argue that better economic times lead to an increase in crime because individuals “show off” their valuables[(4)]. Someone seeing these valuables could be more inclined to take them because the opportunity is there. The article by John Roman suggests there is no significant relationship between the economy and crime because crime is episodic[(4)]. Individuals will commit crime and violent crime whether the economy is good or bad.

Data Sourcing and Preparation:

Having accurate and reliable data for any analysis project is highly important. Therefore, only reliable and credible sources were used to create the dataset for this project. Data was sourced and combined from: Murder Accountability Project (MAP), United States Department of Justice Bureau of Alcohol, Tobaccos, Firearms and Explosives (ATF), Substance Abuse and Mental Health Services Administration (SAMHSA), Federal Bureau of Investigation, Uniform Crime Reporting (FBI: UCR), United States Census Bureau, and The State Science & Technology Institute (SSTI) for the years of 2010, 2012, and 2017.

Initially, the idea for this project came after working with the data MAP has put together. MAP had the number of homicides broken down by state (including the District of Columbia) and by year. So, the number of homicides per state used in this project was taken from the MAP homicide dataset. Since more information was needed to better determine what factors influence homicides than what the MAP homicide dataset was able to provide, other sources were needed to compile a new collective dataset.

Guns are the number one type of weapon used in homicides and therefore needed to be included in the dataset for this project. Unfortunately, MAP did not have any data on the number of guns by state or year, so a new source needed to be found in order to add this factor to our dataset. It was found that the ATF provides data on the number of guns per state and by year. Using state and year as matching points, the number of guns was added to our dataset. The original files for the three years focused on in this project were PDFs and had to be converted into .csv format utilizing R Studio.

Both drugs and alcohol have played critical roles in influencing homicides as was found by the previous literature and needed to be considered as factors in this project. For both the number of drug users and the number of alcohol abusers, data was sourced from SAMHSA. Again, the data was in a PDF format and was converted into a .csv format utilizing R Studio. The numbers provided for both drug users and alcohol abusers were in thousands, so in our dataset we created two additional columns where each respective number was multiplied by 1,000 to avoid mistakes in interpreting results.

Since prior literature saw a positive impact of law enforcement employees decreasing the number of homicides in the 1990s, the number of law enforcement employees for 2010, 2012, and 2017 should be included in the dataset for this project. It was found that the Uniform Crime Reporting division of the FBI tracks and records the number of law enforcement employees by year and state. This data was contained in an Excel file and was easily combined to our dataset using year and state as matching points.

Looking at the economy was the last important factor in regards to how it impacts homicides needed to be sourced and added to the dataset. It was imperative to determine if the economy played a significant role in reducing or contributing to homicides or if it had no effect at all. SSTI sourced GDP per capita from the Bureau of Economic Analysis (BEA) and combined ten years worth of data into a single Excel file. Again, this data was easily added by matching state and year in our dataset.

To normalize all factors except GDP, data for population by state and year needed to be added to the dataset. The data collected for GDP was already per capita. For population, data was sourced directly from the United States Census Bureau. All three years were contained in a single Excel file and were able to be combined to our dataset utilizing Excel and matching to our dataset using state and year as matching points.

Within R Studio, subsets were created for per capita factors in an effort to preserve the original data collected and without having to create additional columns in the dataset file. Additionally, subsets were created within R Studio for the years 2010, 2012, and 2017 to compare each of the years. Stargazer was used to compare each of the years in a single model beside each other. This allowed the results to be easily interpreted for all three years. For all models, the dependent variable used was the number of homicides per capita. Independent variables used include: guns per capita, alcohol abusers per capita, drug users per capita, law enforcement employees per capita, and real GDP per capita.

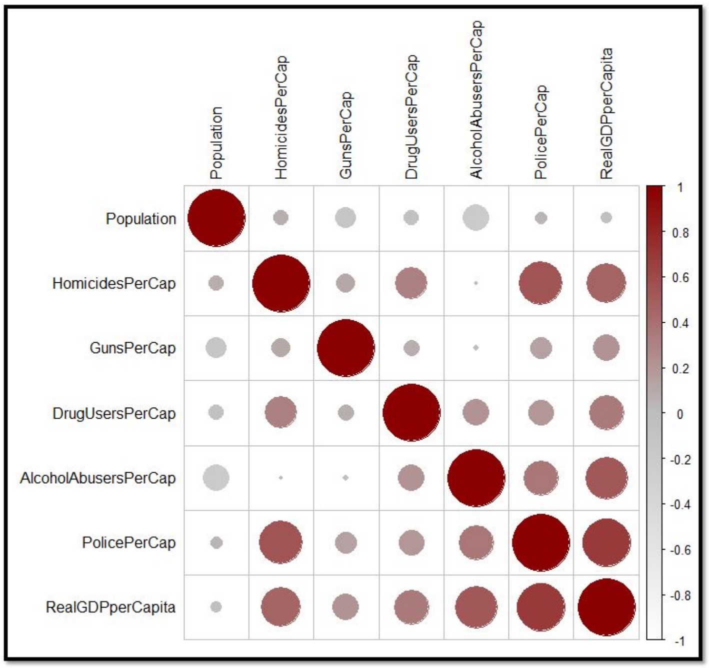
Predictor Variables:

Based on prior literature and intuition, the following hypotheses are expected to be proven true in this project:

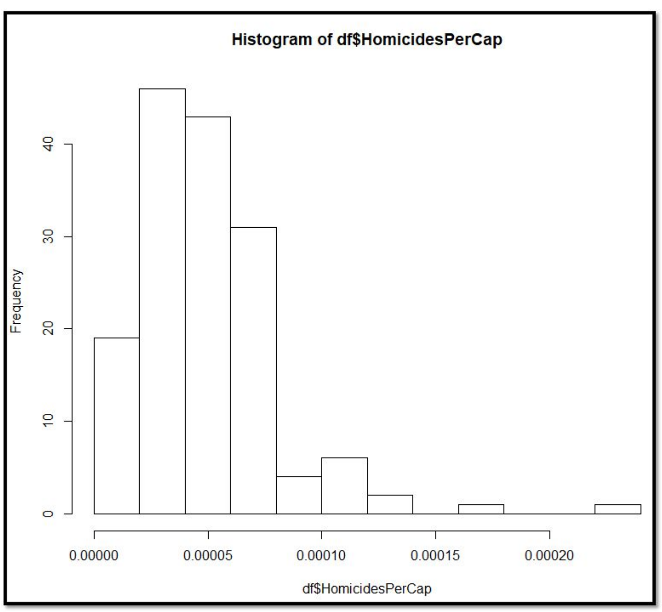
- An increase in law enforcement employees per capita will contribute to a decrease in the number of homicides per capita for the years of 2010, 2012, and 2017.

- Excessive drug users per capita will contribute to an increase in the number of homicides per capita for the years of 2010, 2012, and 2017.
- Excessive alcohol abusers per capita will contribute to an increase in the number of homicides per capita for the years of 2010, 2012, and 2017.
- An increase in the number of guns per capita will contribute to an increase in the number of homicides per capita for the years of 2010, 2012, and 2017.
- An increase in GDP per capita will contribute to a decrease in the number of homicides per capita for the years of 2010, 2012, and 2017.

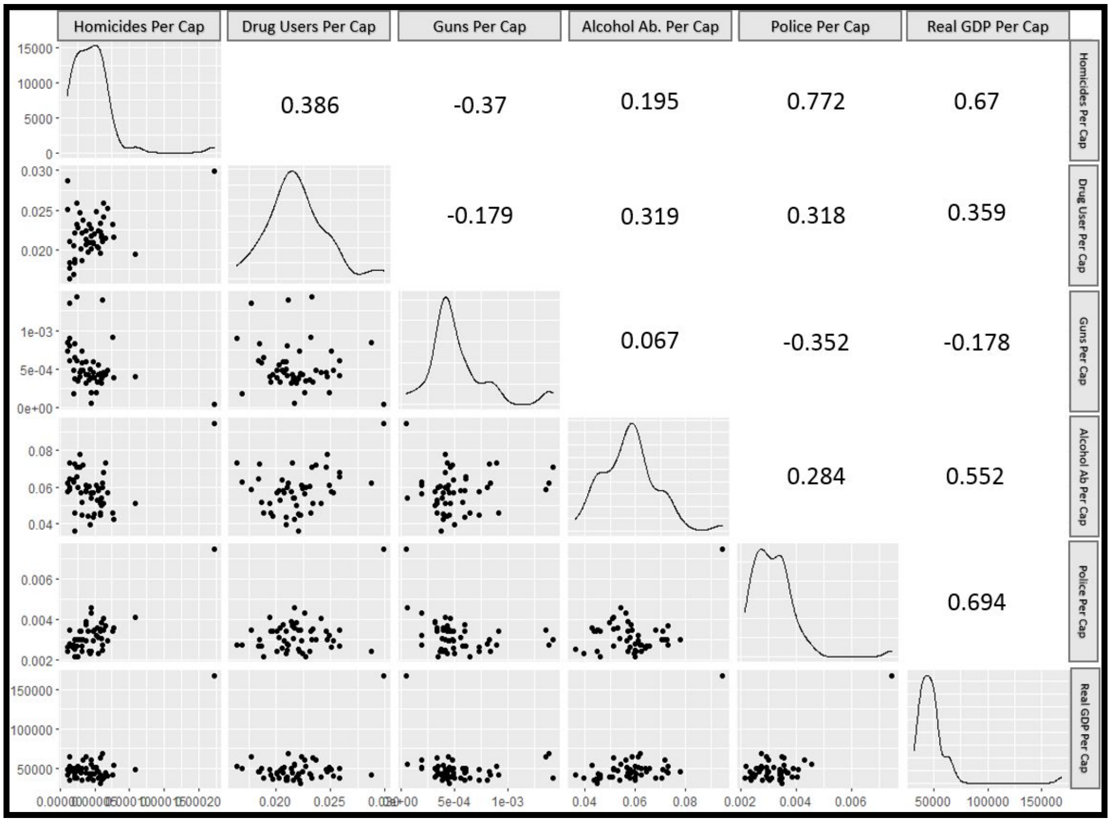
Exploratory Data Analysis and Visualizations:



Initially a correlation plot for all variables was created with all three years combined to see if any of the independent variables were strongly correlated with each other. Based on this correlation plot, real GDP per capita has a high correlation with law enforcement employees per capita. A variance inflation factor (VIF) test would need to be conducted to determine if a model containing both of these variables will not create an issue of multicollinearity. It can also be seen that alcohol abusers per capita is very weakly correlated with homicides per capita.



Next, a histogram for the dependent variable, homicides per capita, was created to help determine if the data was normally distributed. By converting each of the variables to per capita, it helped to normalize the data being used and analyzed. Although this histogram shows there are some outliers in the data, it appears to be relatively normally distributed after being converted to homicides per capita.



Also as part of the exploratory analysis and visualization of the data, scatter plots with correlation were created for each of the variables for each year. This scatter plot gives the results for 2010. Based on this visualization, most of the variables were positively correlated, but a few were negatively correlated. Also, only police per capita and homicides per capita appear to be strongly correlated with each other.

Models:

Based on the core predictors of interest, descriptive analysis and continuous nature of the dependent variable, Homicide,three linear regression models have been formulated and the results analyzed.

<u>Model 1:</u>

```
#Model 1
m1_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + AlcoholAbuserPerCap, data = df2010)
m1_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + AlcoholAbuserPerCap, data = df2012)
m1_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + AlcoholAbuserPerCap, data = df2017)
```

Results:

Homicides Per Capita						
Model 1	2010		2012		2017	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Guns per capita	-0.011	0.010	-0.00004	0.0001	-0.00001	0.0001
Police per capita	0.019	0.005	0.016	0.004	0.016	0.007
Drugs Users per capita	0.002	0.001	0.003	0.001	0.002	0.001
Real GDP per capita	0.000	0.000	0.000	0.000	0.000	0.000
Alcohol Abusers per capita	-0.001	0.0003	-0.001	0.003	-0.002	0.001
Constant	-0.00005	0.00003	-0.00002	0.00003	0.00002	0.00003
Observations	51		51		51	
Adjusted R^2	0.642		0.412		0.193	

Based on the results of Model 1, in 2010, when guns per capita is increased by 100 units, the number of homicides per capita is reduced by one unit while all other variables are held constant. In looking at the results for 2012 and 2017, the marginal effects of guns per capita on homicides per capita are lesser than in 2010, with reduction in the number of homicides per capita of 0.004 and 0.001 respectively.

When looking at police per capita, it can be seen that the marginal effect on homicides is roughly the same for all three years. In 2010, a 100 unit increase in police per capita increased the number of homicides per capita by 1.9 units when keeping all other variables constant. The reason for the increase in homicides per capita rather than a decrease could be attributed to a lag effect. The results of adding more police per capita could take years before seeing positive results.

The effect of drug users per capita has been near constant for three years in Model 1 as well. The only year with a slight difference is 2012 and if there is a 1,000 unit increase in the number of drug users per capita, it would result in a three unit increase in the number of homicides per capita while keeping all other variables constant.

Real GDP per capita has a marginal effect of zero for all three years indicating that in this model, it has no effect on the number of homicides per capita. This is important to note because it aligns with the article written by John Roman mentioned in the Prior Literature section of this report. Individuals will commit crime or homicide regardless of the state of the economy.

Lastly, alcohol abusers per capita leads to a decrease in homicides per capita in all three years observed. The amount of reduction in homicides per capita was found to be almost equal across the three years. So, for a 1,000 unit increase in alcohol abusers per capita, a decrease of 1 unit will occur in homicides per capita with all other variables being held constant.

<u>Model 2:</u>

```
#Model 2
m2_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +AlcoholAbuserPerCap + AlcoholAbuserPerCap*PolicePerCap, data = df2010)
m2_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +AlcoholAbuserPerCap + AlcoholAbuserPerCap*PolicePerCap, data = df2012)
m2_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +AlcoholAbuserPerCap + AlcoholAbuserPerCap*PolicePerCap, data = df2017)
```

Results:

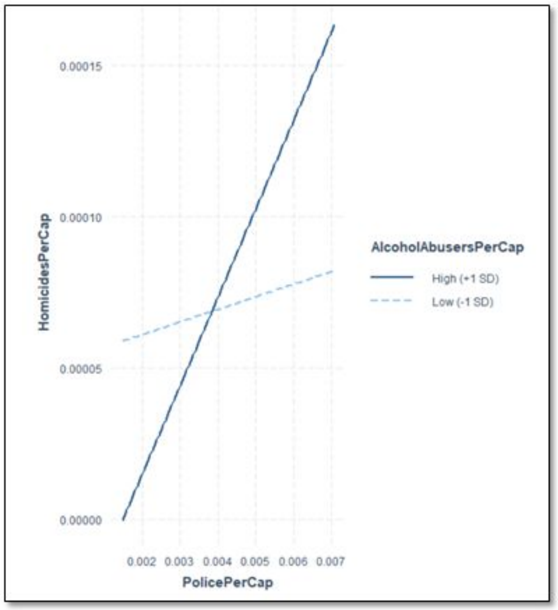
Homicides Per Capita						
Model 2	2010		2012		2017	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Guns per capita	-0.009	0.010	-0.00005	0.0001	-0.00005	0.0001
Police per capita	-0.015	0.014	-0.004	0.012	-0.056	0.018
Drugs Users per capita	0.001	0.001	0.002	0.001	0.002	0.001
Real GDP per capita	0.000	0.000	0.000	0.000	0.000	0.000
Alcohol Abusers per capita	-0.002	0.001	-0.002	0.001	-0.006	0.001
Police per capita * Alcohol Abusers per capita	0.578	0.228	0.315	0.182	1.503	0.358
Constant	0.0001	0.0001	0.0001	0.0001	0.0003	0.0001
Observations	51		51		51	
Adjusted R^2	0.680		0.437		0.411	

As with Model 1, the effect of guns per capita reduces the number of homicides per capita for all three years with the most significant decrease being seen in the year 2010.

In this model, police per capita appears to reduce the number of homicides per capita. The most significant reduction can be seen in 2017, where a 100 unit increase in police per capita leads to a 5.6 unit reduction in homicides per capita.

The results seen for alcohol abusers per capita, drug users per capita, and real GDP per capita are approximately the same as those in Model 1.

Alcohol abusers per capita was chosen as an interaction term with police per capita because heavy drinkers tend to believe that they are invincible and are more likely to do things they otherwise would not do. It was expected that at low levels of alcohol abusers per capita, police per capita would cause a significant decrease in homicides per capita and at high levels of alcohol abusers per capita, police per capita would still cause a decrease in homicides per



capita but to a lesser degree and with a less steep slope. However, the combined effect of police per capita and alcohol abusers per capita leads to an increase in the number of homicides per capita for all three years. The highest increase is seen in 2017, where a one unit increase in alcohol abusers per capita and police per capita leads to a 1.5 unit increase in homicides per capita. In looking at the interaction between alcohol abusers per capita and police per capita on the effects of homicides per capita, it can be seen that at low levels of police, higher levels of alcohol abusers is correlated to lower levels of homicides. However, this changes when higher levels of police per capita is seen, higher levels of alcohol abusers per capita is correlated to higher levels of homicides per capita.

Model 3:

```
#Model 3
m3_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + GunsPerCap*DrugsPerCap, data = df2010)
m3_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + GunsPerCap*DrugsPerCap, data = df2012)
m3_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita + GunsPerCap*DrugsPerCap, data = df2017)
```

Results:

Homicides Per Capita						
Model 3	2010		2012		2017	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Guns per capita	0.001	0.092	-0.002	0.001	-0.004	0.001
Police per capita	0.020	0.005	0.014	0.005	0.012	0.006
Drugs Users per capita	0.002	0.003	0.003	0.002	-0.003	0.002
Real GDP per capita	0.000	0.000	0.000	0.000	0.000	0.000
Guns per capita * Drug Users per capita	-0.697	4.320	0.119	0.059	0.204	0.057
Constant	-0.0001	0.0001	0.00001	0.00004	0.00001	0.00004
Observations	51		51		51	
Adjusted R^2	0.619		0.406		0.291	

In 2010, guns per capita led to an increase in homicides per capita. However, in 2012 and 2017, guns per capita led to a decrease in homicides per capita.

As was seen in Model 1, police per capita leads to an increase in homicides per capita. Again, a lag effect could be the reason why an increase in homicides per capita was seen rather than a decrease per capita.

An interaction term between drug users per capita and guns per capita was added because in states with more drug users per capita, it was expected to see an increase in guns correlate with an increase in the number of homicides. The interaction effect of guns per capita with drug users per capita suggests that the number of homicides per capita goes up significantly by 20.4 units for a 100 unit rise in both guns and drugs per capita. The effect of guns on homicides per capita is higher at higher levels of drug users per capita.

Across all three models, the estimates of real GDP per capita suggest that it has no effect on the number of homicides per capita. This has been the case for all three years in each model.

Our best-fitting model is Model 2, as it had the highest adjusted R^2 on average with a value of 0.509, meaning this model explains approximately 51% of the variation that actually affects homicides per cap.

Quality Checks:

Quality Checks:

Model 1		2010	2012	2017
Multivariate Normality (Shapiro-Wilks Test) *p-value>0.05: data is normal*		p-value = 0.832 PASSES TEST	p-value = 0.669 PASSES TEST	p-value = 0.517 PASSES TEST
Homoscedasticity (Bartlett Test) *p-value>0.05: data is homoscedastic*		p-value = 0.010 FAILS TEST	p-value = 0.682 PASSES TEST	p-value = 0.001 FAILS TEST
Multicollinearity (VIF) *VIF<5: no multicollinearity*	Guns per capita	VIF = 1.202 PASSES TEST	VIF = 1.152 PASSES TEST	VIF = 1.070 PASSES TEST
	Police per capita	VIF = 2.194 PASSES TEST	VIF = 2.211 PASSES TEST	VIF = 1.970 PASSES TEST
	Drug Users per capita	VIF = 1.217 PASSES TEST	VIF = 1.176 PASSES TEST	VIF = 1.603 PASSES TEST
	Real GDP per capita	VIF = 2.643 PASSES TEST	VIF = 3.455 PASSES TEST	VIF = 2.459 PASSES TEST
	Alcohol Abusers per capita	VIF = 1.577 PASSES TEST	VIF = 1.972 PASSES TEST	VIF = 1.875 PASSES TEST

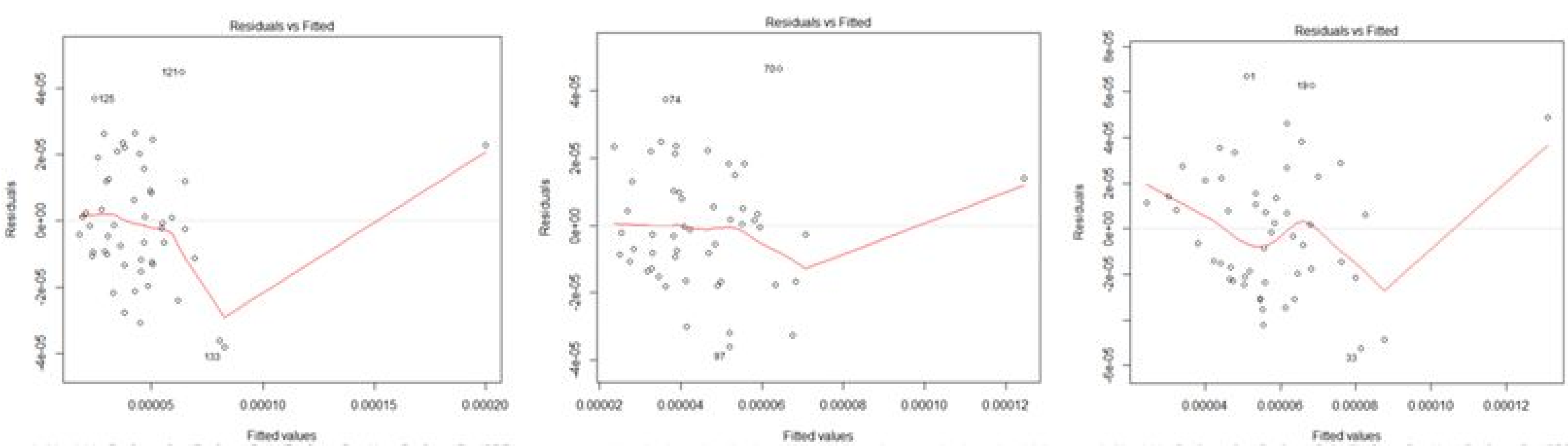
Model 1:

The Shapiro-Wilks test was conducted to test multivariate normality because the sample size was fewer than 2,000. In Model 1, all three years passed the multivariate normality testing as can be seen in the table.

Homoscedasticity was tested using the Bartlett test. Year 2012, was the only year out of the three focused on that passed the homoscedasticity test.

Model 1 also passed the multicollinearity test for each of the variables. The highest variance inflation factors (vif) were seen in real GDP per capita, but none of them exceeded a value of five, which would have suggested there was an issue of multicollinearity.

Residuals vs. Fitted plots - Model 1:



The Residuals vs. Fitted plots for all three years fail the test for linearity. There are outliers in the plots which clearly cause some of the bias. In analyzing the variables used in the model, log transformations were considered and attempted to correct for the non-linearity, but were not successful. The transformation resulted in little to no change in the linearity test.

Quality Checks:				
Model 2		2010	2012	2017
Multivariate Normality (Shapiro-Wilks Test) *p-value>0.05: data is normal*		p-value = 0.275 PASSES TEST	p-value = 0.722 PASSES TEST	p-value = 0.507 PASSES TEST
Homoscedasticity (Bartlett Test) *p-value>0.05: data is homoscedastic*		p-value = 0.001 FAILS TEST	p-value = 0.947 PASSES TEST	p-value = 0.793 PASSES TEST
Multicollinearity (VIF) *VIF<5: no multicollinearity*	Guns per capita	VIF = 1.202 PASSES TEST	VIF = 1.152 PASSES TEST	VIF = 1.070 PASSES TEST
	Police per capita	VIF = 2.194 PASSES TEST	VIF = 2.211 PASSES TEST	VIF = 1.970 PASSES TEST
	Drug Users per capita	VIF = 1.217 PASSES TEST	VIF = 1.176 PASSES TEST	VIF = 1.603 PASSES TEST
	Real GDP per capita	VIF = 2.643 PASSES TEST	VIF = 3.455 PASSES TEST	VIF = 2.459 PASSES TEST
	Alcohol Abusers per capita	VIF = 1.577 PASSES TEST	VIF = 1.972 PASSES TEST	VIF = 1.875 PASSES TEST

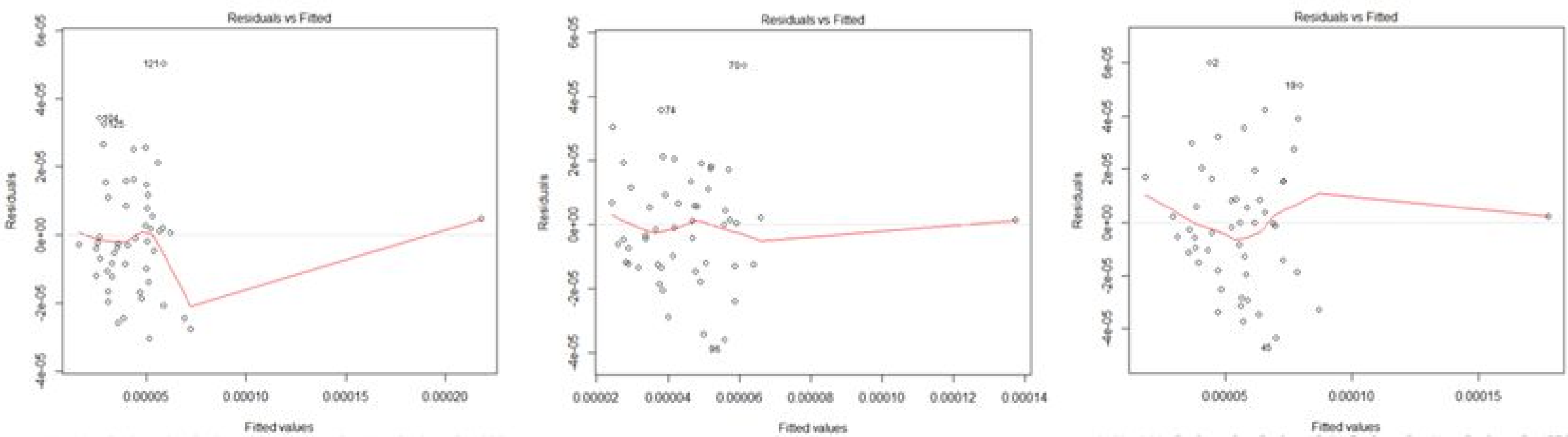
Model 2:

Model 2 passed the multivariate normality test for all three years focused on. Again, the Shapiro-Wilks test was selected to determine whether the data was multivariate normal.

When the Bartlett test was conducted for homoscedasticity, the only year that failed the test was 2010. Both 2012 and 2017 passed the test for homoscedasticity.

The VIF values seen in Model 2 are the same as those in Model 1 because the interaction term of alcohol abusers per capita and police per capita produces inaccurate and abnormally high VIF values. They are highly correlated with their component variables, therefore, Model 2 passes the multicollinearity test.

Residuals vs. Fitted Plots - Model 2:



The Residuals vs. Fitted plots for all three years fail the test for linearity. There are outliers in the plots which clearly cause some of the bias. In analyzing the variables used in the model, log transformations were considered and attempted to correct for the non-linearity, but were not successful. The transformation resulted in little to no change in the linearity test.

Quality Checks:				
Model 3		2010	2012	2017
Multivariate Normality (Shapiro-Wilks Test) *p-value>0.05: data is normal*		p-value = 0.767 PASSES TEST	p-value = 0.254 PASSES TEST	p-value = 0.314 PASSES TEST
Homoscedasticity (Bartlett Test) *p-value>0.05: data is homoscedastic*		p-value = 0.023 FAILS TEST	p-value = 0.624 PASSES TEST	p-value = 0.048 FAILS TEST
Multicollinearity (VIF) *VIF<5: no multicollinearity*	Guns per capita	VIF = 1.163 PASSES TEST	VIF = 1.152 PASSES TEST	VIF = 1.064 PASSES TEST
	Police per capita	VIF = 2.161 PASSES TEST	VIF = 2.155 PASSES TEST	VIF = 1.946 PASSES TEST
	Drug Users per capita	VIF = 1.171 PASSES TEST	VIF = 1.171 PASSES TEST	VIF = 1.232 PASSES TEST
	Real GDP per capita	VIF = 2.038 PASSES TEST	VIF = 2.204 PASSES TEST	VIF = 2.296 PASSES TEST

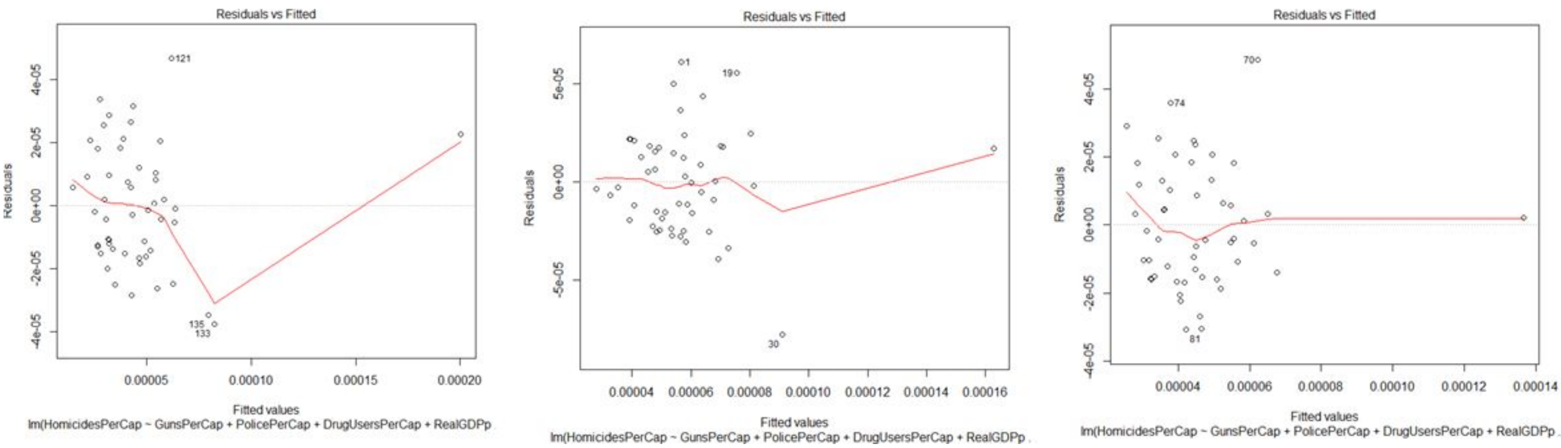
Model 3:

Model 3 passes the multivariate normality test for all three years focused on. As with the previous two models, the Shapiro-Wilks test was also used to test for multivariate normality in Model 3.

The Bartlett test was used to test for homoscedasticity and only the year 2012 passed. Both 2010 and 2017 failed the homoscedasticity test.

In order to obtain realistic VIF numbers to test for multicollinearity, Model 3 was run without the interaction term of guns per capita and drug users per capita. This allowed for a more accurate multicollinearity test since guns per capita and drug users per capita are highly correlated with their component variables. As can be seen in the table, there is no multicollinearity in Model 3.

<u>Residuals vs. Fitted Plots - Model 3:</u>



The Residuals vs. Fitted plots for all three years fail the test for linearity. There are outliers in the plots which clearly cause some of the bias. In analyzing the variables used in the model, log transformations were considered and attempted to correct for the non-linearity, but were not successful. The transformation resulted in little to no change in the linearity test.

Insights and Recommendations:

From analysis of our best-fitting model, Model 2, it was found that:

- The marginal effect of drug users per capita on homicides per capita is an increase of 0.002 units on average. Due to the increase in homicides per capita when drug users per capita increases, it is recommended to help get these offenders clean by offering drug rehabilitation programs.
- Based on the results of the models, GDP per capita has no effect on homicides per capita. Homicide rates remain the same whether the economy is good or poor.
- The marginal effect of police per capita on homicides per capita is a function of the number of alcohol abusers. When police per capita and alcohol abusers per capita increase by 1 unit each, homicides per capita increase by 0.774 units on average.
- Guns per capita has an inverse relationship with homicides per capita in the United States. For instance in 2017, a 100 unit increase in the number of gun users per capita resulted in a 0.005 unit decrease in the number of homicides per capita. This can be attributed to the fact that the U.S. has the highest gun ownership rates compared to other developed countries, but so few of these guns are used to commit homicide when compared to ownership.

References:

Background Information Sources:

1. Blumstein A, Rivara F, Rosenfeld R. The Rise and Decline of Homicide—and Why. Annual Reviews. https://www.annualreviews.org/doi/full/10.1146/annurev.publhealth.21.1.505. Published 2000.

2. Levitt S. Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not. Pubs.aeaweb.org. https://pubs.aeaweb.org/doi/pdf/10.1257/089533004773563485. Published 2004.

3. Murder Accountability Project. Murderdata.org. http://www.murderdata.org/. Published 2019.

4. Roman J. The Puzzling Relationship Between Crime and the Economy. CityLab. https://www.citylab.com/life/2013/09/puzzling-relationship-between-crime-and-economy/6982/. Published 2013.

5. Wieczorek W, Welte J, Abel E. Alcohol, Drugs and Murder: A Study of Convicted Homicide Offenders. Www-sciencedirect-com.ezproxy.lib.usf.edu. https://www-sciencedirect-com.ezproxy.lib.usf.edu/science/article/pii/004723529090002S?via%3Dihub. Published 1990.

Data Sources:

- **Guns 2010 (Page 29):** https://www.atf.gov/file/56646/download
- **Drug Users and Alcohol Abusers 2010:** http://www.samhsa.gov/data/sites/default/files/NSDUHStateEst2010-2011_v2/279/ChangeTabs/NSDUHsaeChangeTabs2011.pdf
- **Law Enforcement Employees 2010:** https://ucr.fbi.gov/crime-in-the-u.s/2010/crime-in-the-u.s.-2010/tables/10tbl77.xls/view
- **Population 2010 (Annual Estimates of the Resident Population for Selected Age Groups by Sex: April 1, 2010, to July 1, 2018):** https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html#par_textimage_785300169
- **GDP 2010:** https://ssti.org/blog/useful-stats-capita-gdp-state-2008-2017
- **Guns 2012 (Page 14):** https://www.hsdl.org/?abstract&did=709348
- **Drug Users and Alcohol Abusers 2012 (Pages: 36-37 & 32-33):** https://www.samhsa.gov/data/sites/default/files/NSDUHStateEst2011-2012/CountTabs/Web/NSDUHsaeCountTabs2012.pdf
- **Law Enforcement Employees 2012:** https://ucr.fbi.gov/crime-in-the-u.s/2012/crime-in-the-u.s.-2012/tables/77tabledatadecpdf/table_77_full_time_law_enforcement_employess_by_state_2012.xls
- **Population 2012 (Annual Estimates of the Resident Population for Selected Age Groups by Sex: April 1, 2010 to July 1, 2018):**

https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html#par_text image_785300169

- **GDP 2012:** https://ssti.org/blog/useful-stats-capita-gdp-state-2008-2017
- **Guns 2017 (Page 16):** https://www.atf.gov/file/118216/download
- **Drug Users and Alcohol Abusers 2017 (Pages 40 && 44):** https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHsaeTotal2017A/NSDUHsaeTotals2017.pdf
- **Law Enforcement Employees 2017:** https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/tables/table-77
- **Population 2017 (Annual Estimates of the Resident Population for Selected Age Groups by Sex: April 1, 2010 to July 1, 2018):** https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html#par_textimage_785300169
- **GDP 2017:** https://ssti.org/blog/useful-stats-capita-gdp-state-2008-2017

Appendix:

```
library(purrr)
library(tidyr)
library(ggplot2)
library(corrplot)
library(stargazer)
library("car")

setwd("C:/Users/shrad/Desktop/BAIS/QMB 6304/Project/R")
d <- read.csv("HomicideDataFinal2.csv")

df$GunsPerCap                   <- df$NumGuns / df$Population
df$HomicidesPerCap          <- df$NumHomicides / df$Population
df$DrugUsersPerCap          <- df$NumDrugUsers / df$Population
df$AlcoholAbusersPerCap     <- df$NumAlcoholAbusers / df$Population
df$PolicePerCap                 <- df$NumLawEnforcementEmployees /  df$Population

df2 <- subset(df, select = c(“State”, “Year”, “Population”, “HomicidesPerCap”, “GunsPerCap”,
“DrugUsersPerCap”, “AlcoholAbusersPerCap”, “PolicePerCap”, “RealGDPPerCapita”))

df2017 <- subset(df2, d$Year == "2017")
df2012 <- subset(df2, d$Year == "2012")
df2010 <- subset(df2, d$Year == "2010")

#Correlation Plot

df2_numeric <- df2[ ,3:length(df2)]
df2_cor <- cor(df2_numeric)
corrplot(df_2cor, tl.col = “black”, col = colorRampPalette (c(“white”, “grey”, “darkred”))(200))

#Histogram
hist(df$HomicidesPerCap)

#Scatter Plot with Correlation
install.packages("GGally")
library("GGally")
ggpairs(df2010[4:9])

#Model 1

m1_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
        AlcoholAbuserPerCap , data = df2010)
m1_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
        AlcoholAbuserPerCap , data = df2012)
m1_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
```

```
        AlcoholAbuserPerCap , data = df2017)

stargazer(m1_2010, m1_2012, m1_2017, type="text")

                         Dependent variable:
                     ----------------------------
                         HomicidesPerCap
                         (1)          (2)          (3)
-----------------------------------------------------------
GunsPerCap               -0.011   -0.00004    -0.00001
                         (0.010)   (0.0001)    (0.0001)

PolicePerCap             0.019***  0.016***   0.016**
                         (0.005)    (0.004)    (0.007)

DrugUsersPerCap          0.002      0.003**     0.002*
                         (0.001)    (0.001)     (0.001)

RealGDPperCapita         0.000**     0.000       0.000
                         (0.000)    (0.000)    (0.000)

AlcoholAbuserPerCap -0.001*   -0.001**   -0.002**
                         (0.0003)   (0.0003)   (0.001)

Constant                 -0.00005  -0.00002   0.00002
                         (0.00003) (0.00003) (0.00003)

-----------------------------------------------------------
Observations                51          51          51
R2                        0.678       0.471       0.274
Adjusted R2               0.642       0.412       0.193
Residual Std. Error (df = 45)  0.00002   0.00002   0.00003
F Statistic (df = 5; 45)       18.936*** 8.010***   3.397**
============================================================
Note:                    *p<0.1; **p<0.05; ***p<0.01

# Multivariate normality

shapiro.test(m1_2010$res)
        Shapiro-Wilk normality test
        data:  m1_2010$res
        W = 0.98666, p-value = 0.8322

shapiro.test(m1_2012$res)
        Shapiro-Wilk normality test
        data:  m1_2012$res
        W = 0.98295, p-value = 0.6693
```

```
shapiro.test(m1_2017$res)
        Shapiro-Wilk normality test
        data:  m1_2017$res
        W = 0.97947, p-value = 0.5165

# Homoscedasticity

bartlett.test(list(m1_2010$res, m1_2010$fit))
        Bartlett test of homogeneity of variances
        data:  list(m1_2010$res, m1_2010$fit)
        Bartlett's K-squared = 6.6953, df = 1, p-value = 0.009666

bartlett.test(list(m1_2012$res, m1_2012$fit))
        Bartlett test of homogeneity of variances
        data:  list(m1_2012$res, m1_2012$fit)
        Bartlett's K-squared = 0.16786, df = 1, p-value = 0.682

bartlett.test(list(m1_2017$res, m1_2017$fit))
        Bartlett test of homogeneity of variances
        data:  list(m1_2017$res, m1_2017$fit)
        Bartlett's K-squared = 11.314, df = 1, p-value = 0.0007694

# Multicollinearity

vif(m1_2010)
GunsPerCap   PolicePerCap   DrugUsersPerCap   RealGDPperCapita   AlcoholAbuserPerCap
  1.202137       2.194166          1.217112           2.642839              1.576641

vif(m1_2012)
GunsPerCap   PolicePerCap   DrugUsersPerCap   RealGDPperCapita   AlcoholAbuserPerCap
  1.151988       2.210575          1.175501           3.455021              1.972371

vif(m1_2017)
GunsPerCap   PolicePerCap   DrugUsersPerCap   RealGDPperCapita   AlcoholAbuserPerCap
  1.069557       1.970293          1.603364           2.458676              1.875230

# Residual vs. Fitted

plot(m1_2010$res ~ m1_2010$fit)
plot(m1_2012$res ~ m1_2012$fit)
plot(m1_2017$res ~ m1_2017$fit)


#Model 2

m2_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
```

```
        AlcoholAbusersPerCap + AlcoholAbuserPerCap*PolicePerCap , data = df2010)
m2_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
        AlcoholAbusersPerCap + AlcoholAbuserPerCap*PolicePerCap , data = df2012)
m2_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap  + RealGDPperCapita +
        AlcoholAbusersPerCap + AlcoholAbuserPerCap*PolicePerCap , data = df2017)

stargazer(m2_2010, m2_2012, m2_2017, type="text")

                                        Dependent variable:
                                    ----------------------------
                                          HomicidesPerCap
                                       (1)        (2)        (3)
----------------------------------------------------------------------------
GunsPerCap                           -0.009   -0.00005 -0.00005
                                     (0.010)  (0.0001) (0.0001)

PolicePerCap                         -0.015    -0.004   -0.056***
                                     (0.014)   (0.012)  (0.018)

DrugUsersPerCap                       0.001     0.002    0.002*
                                     (0.001)   (0.001)  (0.001)

RealGDPperCapita                      0.000    -0.000   -0.000**
                                     (0.000)   (0.000)  (0.000)

AlcoholAbuserPerCap                 -0.002***  -0.002** -0.006***
                                     (0.001)   (0.001)  (0.001)

PolicePerCap:AlcoholAbuserPerCap  0.578**    0.315*   1.503***
                                     (0.228)   (0.182)  (0.358)

Constant                             0.0001    0.0001   0.0003***
                                    (0.0001)  (0.0001)  (0.0001)
----------------------------------------------------------------------------------------
Observations                           51        51        51
R2                                    0.719     0.505     0.481
Adjusted R2                           0.680     0.437     0.411
Residual Std. Error (df = 44)        0.00002   0.00002   0.00003
F Statistic (df = 6; 44)           18.742***  7.473*** 6.806***
========================================================
Note:                   *p<0.1; **p<0.05; ***p<0.01


# Multivariate normality

shapiro.test(m2_2010$res)
        Shapiro-Wilk normality test
```

```
        data:  m2_2010$res
        W = 0.97229, p-value = 0.2746

shapiro.test(m2_2012$res)
        Shapiro-Wilk normality test
        data:  m2_2012$res
        W = 0.98411, p-value = 0.722

shapiro.test(m2_2017$res)
        Shapiro-Wilk normality test
        data:  m2_2017$res
        W = 0.97924, p-value = 0.5069

# Homoscedasticity

bartlett.test(list(m2_2010$res, m2_2010$fit))
        Bartlett test of homogeneity of variances
        data:  list(m2_2010$res, m2_2010$fit)
        Bartlett's K-squared = 10.519, df = 1, p-value = 0.001182

bartlett.test(list(m2_2012$res, m2_2012$fit))
        Bartlett test of homogeneity of variances
        data:  list(m2_2012$res, m2_2012$fit)
        Bartlett's K-squared = 0.0044083, df = 1, p-value = 0.9471

bartlett.test(list(m2_2017$res, m2_2017$fit))
        Bartlett test of homogeneity of variances
        data:  list(m2_2017$res, m2_2017$fit)
        Bartlett's K-squared = 0.068935, df = 1, p-value = 0.7929

# Multicollinearity

# Used vif numbers from previous model due to interaction term

# Residual vs. Fitted

plot(m2_2010$res ~ m2_2010$fit)
plot(m2_2012$res ~ m2_2012$fit)
plot(m2_2017$res ~ m2_2017$fit)

library(interactions)

interact_plot(m2_2017, pred = DrugUsersPerCap, modx = GunsPerCap, modxvals="plus-minus", modx.labels=
c("Low (-1 SD)", "High (+1 SD)"))

#Model 3
```

```
m3_2010 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +
        + DrugUsersPerCap*GunsPerCap, data = df2010 )
m3_2012 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +
        + DrugUsersPerCap*GunsPerCap, data = df2012 )
m3_2017 <- lm(HomicidesPerCap ~ GunsPerCap + PolicePerCap + DrugUsersPerCap + RealGDPperCapita +
        + DrugUsersPerCap*GunsPerCap, data = df2017 )

stargazer(m3_2010, m3_2012, m3_2017, type="text")

                                    Dependent variable:
                                  ----------------------------
                                      HomicidesPerCap
                                     (1)       (2)        (3)
 ---------------------------------------------------------------------------
GunsPerCap                          0.001   -0.002**  -0.004***
                                   (0.092)   (0.001)   (0.001)

PolicePerCap                       0.020***  0.014***   0.012*
                                   (0.005)   (0.005)   (0.006)

DrugUsersPerCap                     0.002    0.0003    -0.003*
                                   (0.003)   (0.002)   (0.002)

RealGDPperCapita                    0.000    -0.000*   -0.000
                                   (0.000)   (0.000)   (0.000)

GunsPerCap:DrugUsersPerCap         -0.697    0.119**  0.204***
                                   (4.320)  (0.059)   (0.057)

Constant                           -0.0001  0.00001  0.0001**
                                   (0.0001) (0.00004) (0.00004)

 ----------------------------------------------------------------------------------
Observations                          51        51         51
R2                                  0.657     0.465      0.362
Adjusted R2                         0.619     0.406      0.291
Residual Std. Error (df = 45)      0.00002   0.00002   0.00003
F Statistic (df = 5; 45)          17.235*** 7.829***  5.114***
===================================================================
 Note:                     *p<0.1; **p<0.05; ***p<0.01

# Multivariate normality

shapiro.test(m3_2010$res)
        Shapiro-Wilk normality test
        data:  m3_2010$res
        W = 0.98513, p-value = 0.7673
```

```
shapiro.test(m3_2012$res)
        Shapiro-Wilk normality test
        data:  m3_2012$res
        W = 0.97147, p-value = 0.2543

shapiro.test(m3_2017$res)
        Shapiro-Wilk normality test
        data:  m3_2017$res
        W = 0.97374, p-value = 0.3142

# Homoscedasticity

bartlett.test(list(m3_2010$res, m3_2010$fit))
        Bartlett test of homogeneity of variances
        data:  list(m3_2010$res, m3_2010$fit)
        Bartlett's K-squared = 5.1347, df = 1, p-value = 0.02345

bartlett.test(list(m3_2012$res, m3_2012$fit))
        Bartlett test of homogeneity of variances
        data:  list(m3_2012$res, m3_2012$fit)
        Bartlett's K-squared = 0.24014, df = 1, p-value = 0.6241

bartlett.test(list(m3_2017$res, m3_2017$fit))
        Bartlett test of homogeneity of variances
        data:  list(m3_2017$res, m3_2017$fit)
        Bartlett's K-squared = 3.9028, df = 1, p-value = 0.04821

# Multicollinearity

# ran model without interaction term to acquire accurate vif numbers

vif(m3_2010)
GunsPerCap      PolicePerCap      DrugUsersPerCap      RealGDPperCapita
  1.163454        2.160837          1.171184              2.037509

vif(m3_2012)
GunsPerCap      PolicePerCap      DrugUsersPerCap      RealGDPperCapita
  1.151984        2.154600          1.170821              2.203640

vif(m3_2017)
GunsPerCap      PolicePerCap      DrugUsersPerCap      RealGDPperCapita
  1.064479        1.945712          1232210               2.296420

# Residual vs. Fitted

plot(m3_2010$res ~ m3_2010$fit)
```

```
plot(m3_2012$res ~ m3_2012$fit)
plot(m3_2017$res ~ m3_2017$fit)
```