Project Report

Weapon Utilization & Drugs/Alcohol consumption in Criminal Behavior: Analyzing the Likelihood, Crime Categories, Respondent Impact, and Locational Trends, & examining the Nexus Between Drug/Alcohol Influence and Weapon Utilization in Criminal Acts.

ICPSR 34650

National Crime Victimization Survey, [United States], 2012

United States. Bureau of Justice Statistics

Subject

Regression Time Series Analysis (16:954:596:0)

Team Members

Kiran Reddy Nemalidinne (220008729)

Bhavana Vemula (226009099)

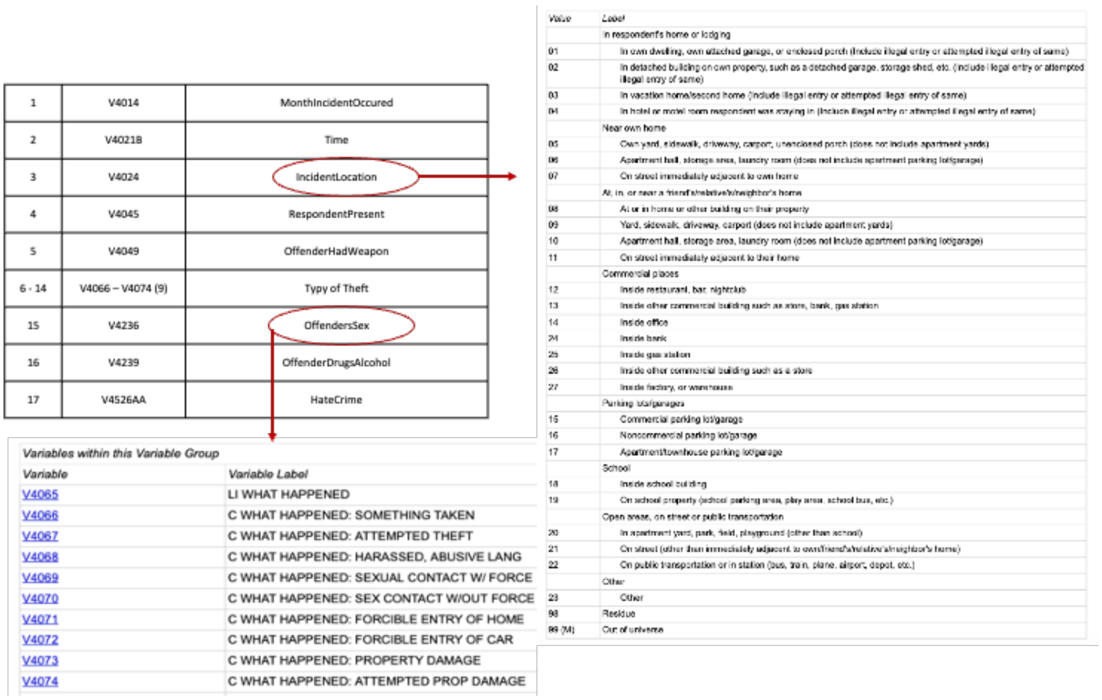
Nikhil chekuri (227006792)

Chittha Bhuvan Reddy D (226007958)

Description of Dataset

The National Crime Victimization Survey (NCVS) is an ongoing survey since 1973, gathering detailed data on personal and household victimization in the U.S. It aims to provide comprehensive information on unreported crimes, uniform crime measures, and trends over time and areas. Covering personal crimes (like assault, robbery) and property crimes (such as burglary, theft), the survey includes data on crime type, location, victim-offender relationship, and victim demographics, along with information on crime reporting and offender characteristics.

We selected specific variables from the National Crime Victimization Survey for their relevance in understanding crime dynamics. Variables like Month and Time of Incident, Incident Location, and whether the Respondent was Present provide contextual details. The inclusion of Offender Had Weapon, Type of Theft, Offender's Sex, Drug/Alcohol Influence, and Hate Crime status allows for a deeper analysis of the nature of the crimes and offender characteristics, crucial for identifying patterns and trends in criminal behavior.



1	V4014	MonthIncidentOccured
2	V4021B	Time
3	V4024	IncidentLocation
4	V4045	RespondentPresent
5	V4049	OffenderHadWeapon
6 - 14	V4066 – V4074 (9)	Typy of Theft
15	V4236	OffendersSex
16	V4239	OffenderDrugsAlcohol
17	V4526AA	HateCrime

Variables within this Variable Group

Variable	*Variable Label*
V4065	LI WHAT HAPPENED
V4066	C WHAT HAPPENED: SOMETHING TAKEN
V4067	C WHAT HAPPENED: ATTEMPTED THEFT
V4068	C WHAT HAPPENED: HARASSED, ABUSIVE LANG
V4069	C WHAT HAPPENED: SEXUAL CONTACT W/ FORCE
V4070	C WHAT HAPPENED: SEX CONTACT W/OUT FORCE
V4071	C WHAT HAPPENED: FORCIBLE ENTRY OF HOME
V4072	C WHAT HAPPENED: FORCIBLE ENTRY OF CAR
V4073	C WHAT HAPPENED: PROPERTY DAMAGE
V4074	C WHAT HAPPENED: ATTEMPTED PROP DAMAGE

Value	*Label*
	In respondent's home or lodging
01	In own dwelling, own attached garage, or enclosed porch (Include illegal entry or attempted illegal entry of same)
02	In detached building on own property, such as a detached garage, storage shed, etc. (Include illegal entry or attempted illegal entry of same)
03	In vacation home/second home (Include illegal entry or attempted illegal entry of same)
04	In hotel or motel room respondent was staying in (Include illegal entry or attempted illegal entry of same)
	Near own home
05	Own yard, sidewalk, driveway, carport, unenclosed porch (does not include apartment yards)
06	Apartment hall, storage area, laundry room (does not include apartment parking lot/garage)
07	On street immediately adjacent to own home
	At, in, or near a friend's/relative's/neighbor's home
08	At or in home or other building on their property
09	Yard, sidewalk, driveway, carport (does not include apartment yards)
10	Apartment hall, storage area, laundry room (does not include apartment parking lot/garage)
11	On street immediately adjacent to their home
	Commercial places
12	Inside restaurant, bar, nightclub
13	Inside other commercial building such as store, bank, gas station
14	Inside office
24	Inside bank
25	Inside gas station
26	Inside other commercial building such as a store
27	Inside factory, or warehouse
	Parking lots/garages
15	Commercial parking lot/garage
16	Noncommercial parking lot/garage
17	Apartment/townhouse parking lot/garage
	School
18	Inside school building
19	On school property (school parking area, play area, school bus, etc.)
	Open areas, on street or public transportation
20	In apartment yard, park, field, playground (other than school)
21	On street (other than immediately adjacent to own/friend's/relative's/neighbor's home)
22	On public transportation or in station (bus, train, plane, airport, depot, etc.)
	Other
23	Other
98	Residue
99 (M)	Out of universe

We have selectively distilled our dataset to 17 essential variables from a vast array of over 900, through rigorous preprocessing techniques that include outlier detection to identify and address anomalies, normalization to scale the data uniformly, and imputation to fill in gaps where data was missing. Additionally, our data cleaning efforts have removed any duplicates, ensuring every one of the 8,746 observations used is unique and contributes meaningfully to the integrity of our analysis of crime trends. This comprehensive preparation of the dataset paves the way for a highly reliable and insightful statistical exploration.

Key Questions Addressed through Data Analysis

- What is the likely of an offender utilizing a weapon during a crime?
- Which categories of crime most frequently involve the use of weapons by offenders?

- Is there an increased likelihood of an offender brandishing a weapon when a respondent is present?
- In which types of locations are weapons most used by offenders?
- Are offenders under the influence of drugs more likely to use weapons?
- What are the most places where Drug/Alcohol crimes are being held?

A Dual Model Analytical Approach

Categorical Logistic Regression (Prediction variable- OffenderHadWeapon)

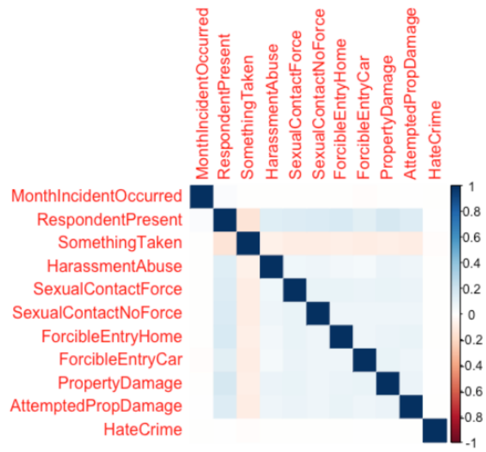
Our initial model employed 'OffenderHadWeapon' as the dependent variable, with all other selected variables serving as predictors to assess their impact on the likelihood of weapon use in criminal incidents. The logistic regression analysis reveals key insights: Incident Locations 16, 21, and 25 (Non-Commercial Parking, inside gas station and on streets) notably increase the probability of weapons being used in crimes. The presence of a respondent significantly influences the likelihood of a weapon-involved incident. The types of crimes where weapons are most reported include 'Something Taken', 'Harassment Abuse', 'Sexual Contact No Force', 'Property Damage', and 'Attempted Property Damage'.

We have optimized our logistic regression model using stepwise selection to predict 'OffenderHadWeapon', reducing the Akaike Information Criterion (AIC) from 10813 to 10804. This refined model indicates significant variables including specific Incident Locations and the presence of a respondent as critical factors in predicting weapon used in crimes. It also highlights certain crime types, such as 'Something Taken', 'Harassment Abuse', and 'Sexual Contact No Force', as commonly associated with weapon involvement. The model's improved fit is substantiated by a lower AIC value, suggesting a more parsimonious model without loss of predictive power.

In our analysis, we evaluated multicollinearity using the Variance Inflation Factor (VIF), confirming that our predictors do not suffer from severe multicollinearity, hence are suitable for the logistic regression model. Furthermore, we constructed a correlation matrix to understand the relationships between various predictors. The correlation matrix revealed that while most variables show little to no multicollinearity, some pairs, like 'RespondentPresent' with 'PropertyDamage' and 'ForcibleEntryHome', exhibit moderate positive correlations, providing insights into the interdependencies within our data. These steps ensure the validity of our regression model and the reliability of the insights drawn from it.

	MonthIncidentOccurred	RespondentPresent	SomethingTaken	HarassmentAbuse	SexualContactForce	SexualContactNoForce	ForcibleEntryHome	ForcibleEntryCar	PropertyDamage	AttemptedPropDamage	HateCrime
MonthIncidentOccurred	1.000000000	0.020207351	-0.008993612	-0.007323226	-9.708576e-03	-0.0085893370	-0.006521412	-0.01419889	-0.006667727	0.006016552	-3.691214e-03
RespondentPresent	0.020207351	1.000000000	-0.144759854	0.135097622	1.492816e-01	0.1524322027	0.164298443	0.12943823	0.174583780	0.149200025	5.126443e-03
SomethingTaken	-0.008993612	-0.144759854	1.000000000	-0.071163290	-9.769847e-02	-0.0973056495	-0.084427750	-0.09667970	-0.083847821	-0.090866800	-1.865228e-02
HarassmentAbuse	-0.007323226	0.135097622	-0.071163290	1.000000000	6.726800e-02	0.0710689361	0.055483187	0.04453681	0.083530256	0.079447401	-7.300634e-03
SexualContactForce	-0.009708576	0.149281593	-0.097698472	0.067268000	1.000000e+00	0.0915273470	0.090317796	0.08094071	0.094774070	0.080032385	-1.534247e-05
SexualContactNoForce	-0.008589337	0.152432203	-0.097305650	0.071068936	9.152735e-02	1.0000000000	0.078838863	0.07767240	0.078677199	0.076434030	-3.864566e-04
ForcibleEntryHome	-0.006521412	0.164298443	-0.084427750	0.055483187	9.031780e-02	0.0788388630	1.000000000	0.07791491	0.086653733	0.094595658	-3.688359e-03
ForcibleEntryCar	-0.014198890	0.129438233	-0.096679697	0.044536805	8.094071e-02	0.0776724005	0.077914913	1.00000000	0.091749764	0.078529201	4.663150e-03
PropertyDamage	-0.006667727	0.174583780	-0.083847821	0.083530256	9.477407e-02	0.0786771988	0.086653733	0.09174976	1.000000000	0.088441223	8.750513e-03
AttemptedPropDamage	0.006016552	0.149200025	-0.090866800	0.079447401	8.003239e-02	0.0764340298	0.094595658	0.07852920	0.088441223	1.000000000	-2.942721e-03
HateCrime	-0.003691214	0.005126443	-0.018652280	-0.007300634	-1.534247e-05	-0.0003864566	-0.003688359	0.00466315	0.008750513	-0.002942721	1.000000e+00

Further in our analysis, we used the corrplot library to visualize the correlation between selected variables in the dataset. The heatmap generated provides a quick and informative overview, displaying the strength and direction of the relationships among variables, such as the time of the incident, presence of the respondent, nature of the theft, and other factors. Darker colors on the map indicate stronger correlations, which help to quickly identify potential relationships and interactions between different factors associated with crime incidents.

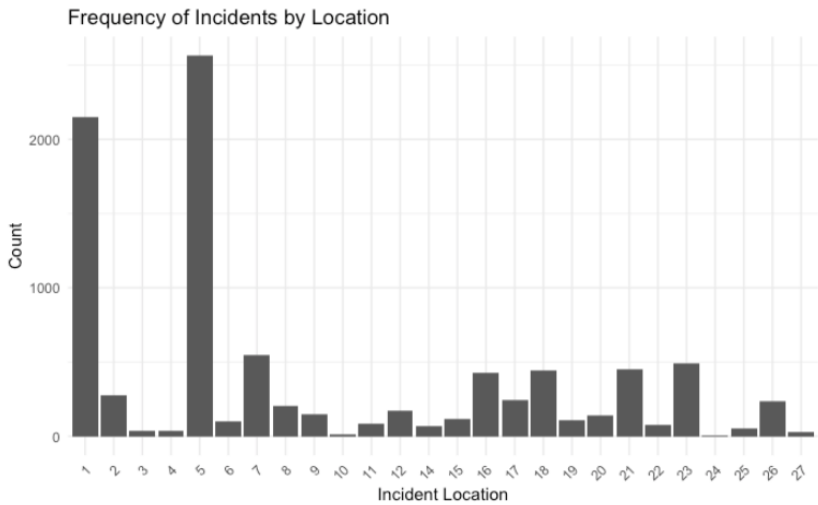


Categorical Logistic Regression (Prediction variable- OffenderDrugsAlcohol)

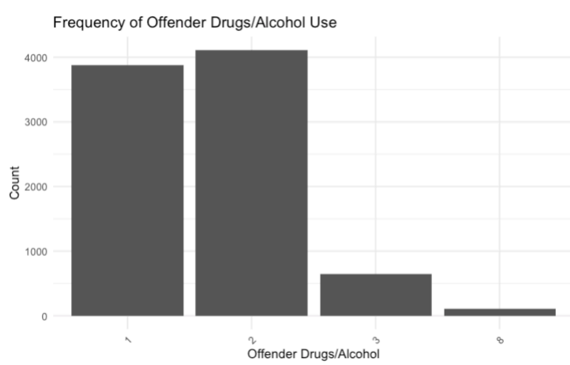
Our second model utilized 'OffenderDrugsAlcohol' as the dependent variable, with all other selected variables serving as predictors to determine their influence on the likelihood of drug or alcohol involvement in criminal incidents. The analysis highlights that specific Incident Locations (particularly 4, 12, 19, and 27 suggests in hotel, restaurant/bar, on school property and inside factory) and certain variables like 'Attempted Theft' significantly correlate with the presence of drugs or alcohol. Notably, the timing of the incident also appears to be a predictor, while the presence of a respondent and other factors like 'Something Taken' and 'Harassment Abuse' showed less influence on the probability of drug or alcohol use by the offender.

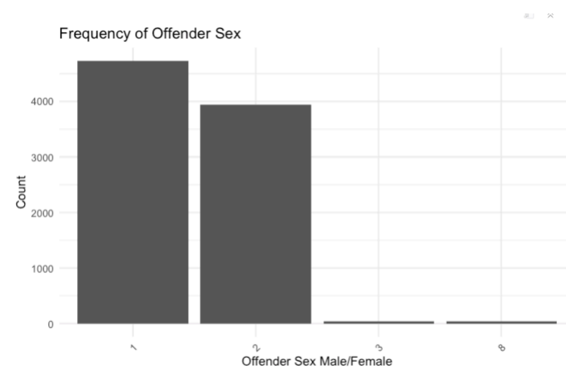
Findings

- **Is there an increased likelihood of an offender brandishing a weapon when a respondent is present:** With analyzing the summary of logistic regression model, As the Respondent Present variable estimate is positive and P-value is significant enough as it is less than the significance value 0.05, we can suggest that RespondentPresent affect the use of weapon. That means, if a person is present with the victim at the time of crime, the offender is most likely to use the weapon.
- **Frequency of incident by locations:** Suggests that the frequency of incidents is highest at the respondent's own home or surrounding areas such as yards and driveways, followed by incidents on the street immediately adjacent to their home. Conversely, locations like schools, noncommercial parking lots, and public transportation exhibit a significantly lower frequency of incidents, indicating that crimes are more prevalent in personal residential spaces than public or semi-public locations.

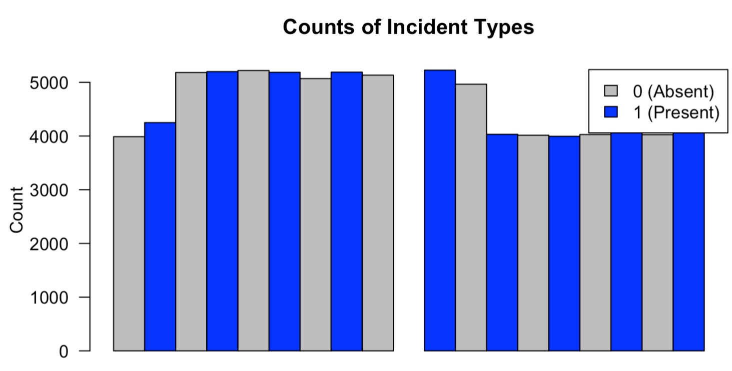


- **Frequencies of Offenders details:** The below depicted bar charts show a higher frequency of crimes where offenders did not use drugs or alcohol compared to those where they did. Regarding offender sex, the frequency of male offenders is higher than that of female offenders. These visualizations provide a clear indication of the prevalence of these characteristics in criminal incidents within the dataset.

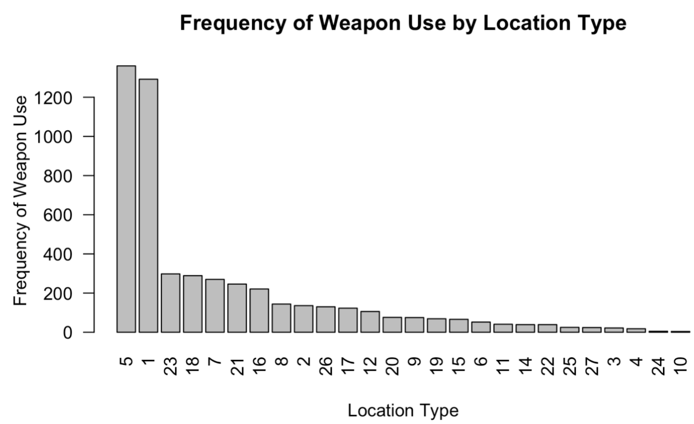




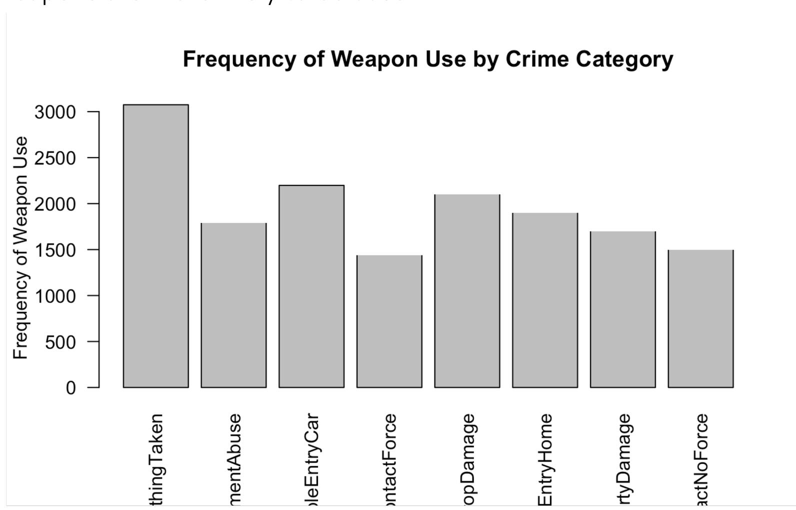
- **Frequencies of incident details:** The bar chart illustrates that the most frequent types of incidents reported involve 'Something Taken' and 'Harassment Abuse', with 'Attempted Theft' and 'Sexual Contact No Force' also being quite common. The presence (indicated in blue) and absence (in grey) of the respondent show that these incidents occur with varying frequency regardless of the respondent's presence.



- **Frequency of weapon use by location:** The frequency of weapon use in criminal incidents is most common in the offender's own yard, sidewalk, driveway, or carport, followed by incidents inside the offender's own dwelling or attached garage. Incidents at schools and other unspecified locations also show a notable occurrence of weapon use. Conversely, weapon use is least reported in apartment common areas such as halls, storage areas, and laundry rooms.

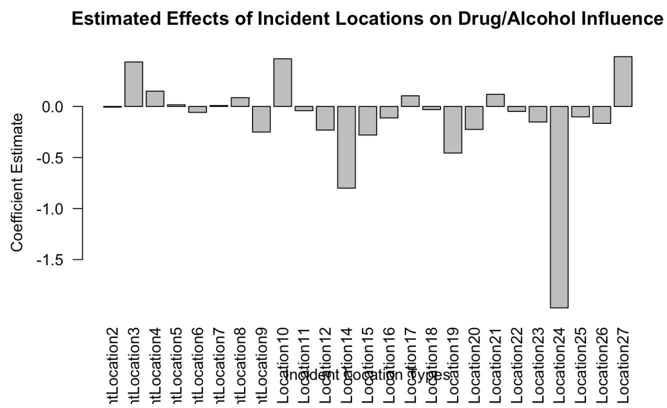


- **Frequency of weapon use by crime category:** Crimes involving theft, such as 'Something Taken', 'Forcible Entry of a Car', and 'Property Damage', are the most likely to involve weapon use by offenders. The likelihood of an offender using a weapon during a crime is relatively high, as indicated by the predictive model, which shows a mean probability of around 59%. This suggests that, in more than half of the incidents, there's a significant chance that a weapon could be involved, underscoring the importance of understanding the contexts in which weapons are more likely to be used.

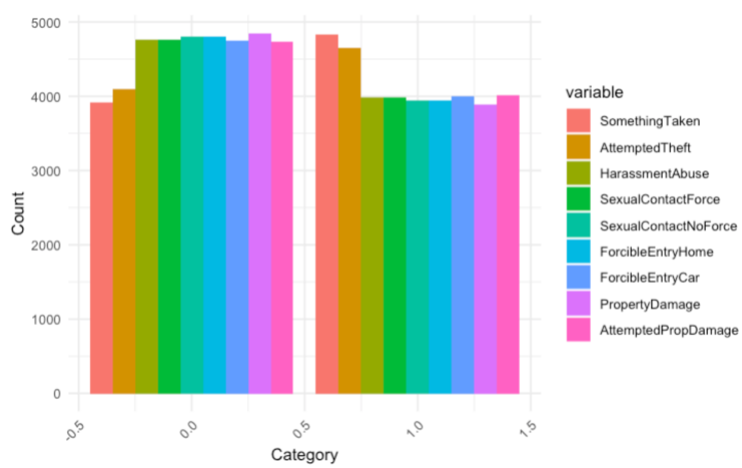


- **Are offenders under the influence of drugs more likely to use weapons:** The analysis shows that the influence of drugs or alcohol on an offender does not significantly increase the likelihood of weapon use during a crime. The p-value for 'OffenderHadWeapon' is above the conventional threshold for significance, and the associated effect size is minimal. This suggests that other factors may play a more pivotal role in determining whether a weapon is used in a crime.

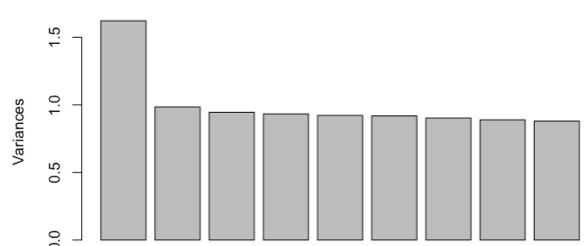
- **What are the most places where Drug/Alcohol crimes are being held:** The bar chart indicates that drug/alcohol-related crimes are most frequently reported in vacation homes or second homes (Location 3), apartment communal areas (Location 10), and industrial settings like factories or warehouses (Location 27). Conversely, such crimes are least likely to occur in more formal and secure environments like offices (Location 14) and banks (Location 24), suggesting that offenders under the influence of drugs or alcohol may prefer less public and less secured locations for committing crimes.



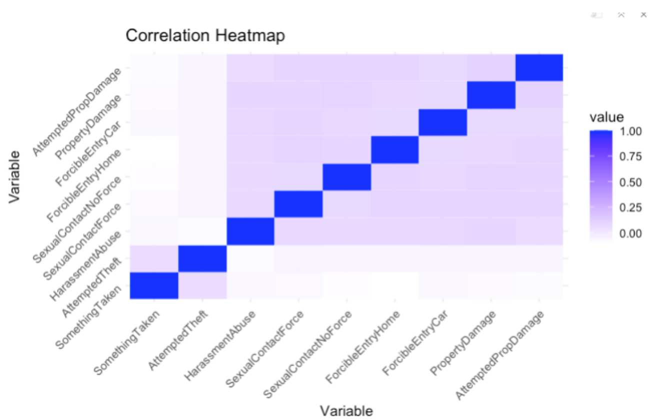
- **Categorical analysis of each type of incident:** The bar graph represents a distribution of various crime categories, showing that incidents of 'Something Taken', 'Harassment Abuse', and 'Attempted Theft' occur with a high frequency. The counts are relatively uniform across these categories, indicating that these types of crimes are commonly reported and may share similar occurrence rates within this dataset.

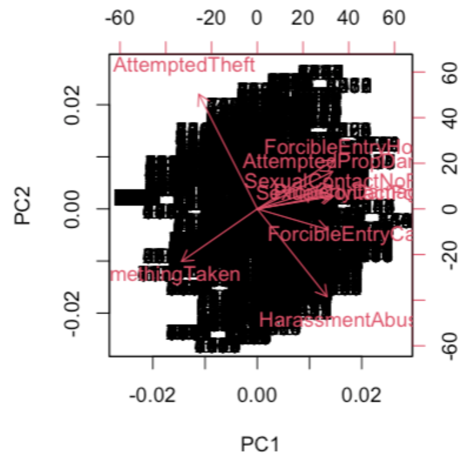


- Is there a discernible pattern or correlation among the various categories of incidents in the dataset, such as a relationship between victims of attempted theft and their likelihood of experiencing harassment abuse or sexual harassment at the same time?



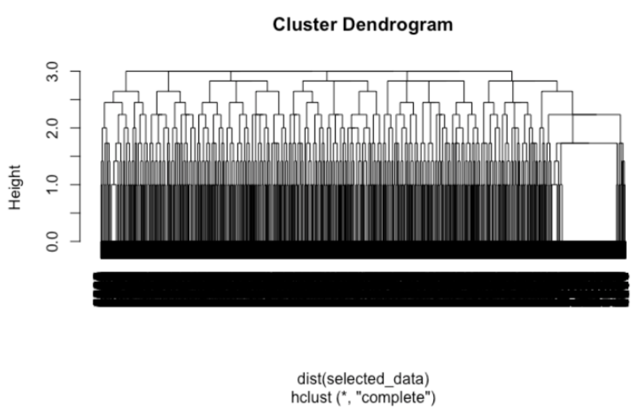
Based on the visualizations, there appears to be a pattern or correlation among the various categories of incidents in the dataset. The strong correlation among certain variables indicates that incidents such as Forcible Entry Car, Attempted Theft, Sexual Contact Force, Forcible Entry Home, and Property Damage are likely to co-occur. This could suggest that when a crime of one nature occurs, it might increase the likelihood of another crime happening in conjunction, potentially due to the circumstances or environment in which these incidents take place.

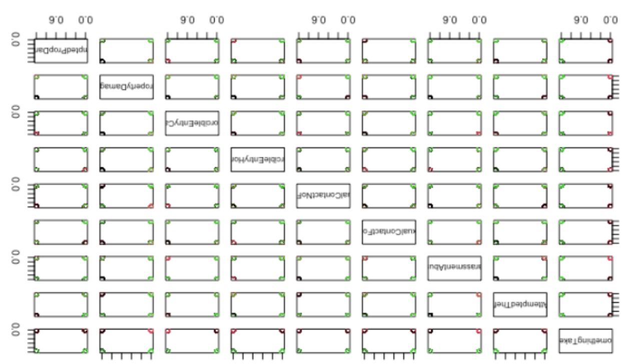




The correlation heatmap further reinforces this observation, showing darkened squares where there is a higher correlation between variables. For instance, if there's a strong blue square correlating Forcible Entry Car and Property Damage, this could imply that when a car is forcibly entered, there is often accompanying property damage.

In summary, the data suggest that certain types of incidents are not isolated and may indeed be predictive of other crime types occurring simultaneously. This kind of analysis can be vital for law enforcement and community safety programs to identify patterns and potentially prevent crime by addressing underlying factors that contribute to these correlations.





For further hierarchical analysis we performed K-means cluster algorithm on the data, The scatterplot matrix provides a visual representation of the distribution and relationship between pairs of variables in the dataset, with each plot showing the correlation (or lack thereof) between two variables. The histograms on the diagonal represent the distribution of a single variable. The cluster dendrogram is a visual summary of the hierarchical clustering analysis, where each branch represents a possible cluster. It provides insight into the data structure and how closely related different observations are to one another.

From the dendrogram, it looks like there is a significant cluster that separates well from the others at a certain height, which suggests that there may be a distinct grouping within the data.

The End