Dallas vs. Denver Crime Patterns

STAT 3350.001 Project

Group 5

Sriram Avinash, Pham Thompson, Igor Joslin

Table of contents

Introduction

Crime is a pervasive issue in cities across the world, and the United States is no exception. Two cities in particular, Dallas and Denver, have been grappling with rising crime rates in recent years. Our group has analyzed datasets containing records of crimes in these cities in order to gain insights into the underlying causes of crime, inform policy decisions, and develop evidence-based solutions for crime prevention.

Dallas, Texas is known for its diverse population and thriving economy, but also faces challenges with high levels of crime. The Dallas Police Department has implemented various strategies to combat crime, but there is a need for comprehensive analysis to identify patterns and trends. Similarly, Denver, Colorado has been experiencing a rise in crime rates, which requires a closer examination of the factors contributing to this trend. Interestingly, the chances of becoming a victim of a violent crime in Texas are 1 in 220, while it is 1 in 116 in Dallas. The likelihood of being a victim of a violent crime in Colorado is 1 in 208, while it is 1 in 103 in Denver. Despite having similar populations, both cities differ in their rates of violent crime. This underscores the importance of analyzing crime data in order to identify the underlying factors contributing to crime rates.

The Dallas dataset comprises more than 60,000 crime incidents with 18 variable attributes, including incident number, date, time, year of arrest, geographical coordinates, age, gender, race of the arrestee, arrest city, zip code, state, arrest day, location, weapon involved, nature of offense (drug or non-drug related), and type of drug involved (if applicable). The Denver crime dataset contains over 342,000 crime incidents with 20 variable attributes. We also implemented a second Denver dataset, which focused on poverty percentages and comprises 1249 incidents with 9 variable attributes. Similarly, the second Dallas dataset goes into poverty percentages and comprises 3403 incidents with 29 variable attributes.

Through our analysis, we aim to shed light on the underlying causes of crime in Dallas and Denver and to provide evidence-based solutions for crime prevention. By identifying patterns and trends, policymakers and law enforcement agencies can better allocate resources and implement targeted interventions to reduce crime rates and promote safer communities.

Data Cleaning

In our overall data cleaning process, we began by uploading all the .csv and shapefile files (.shp, .shx, .dbf) that we needed for our analysis. Then, we proceeded to remove columns that were not relevant to our analysis from both the Dallas and Denver datasets. Specifically, in the Dallas crime dataset, we removed the Incident Number, Arrest Time, Arrest Location, and Arrestee Sex columns, whereas in the Denver crime dataset, we removed the Incident Id, Offense Id, Offense Code, and Offense Code Extension, First and Last Occurred Date columns. This is because we did not have any use for these columns due to the nature of the questions we proposed in the beginning of the project.

During the coding process, we encountered a value within the Dallas crime dataset that was considered empty and therefore was not observable by the data. This initially was causing some skew within our results as it did not have any observable features such as an incident number (which proves its uniqueness as a crime) or other values associated with it. As a result, we decided it was best to remove this column to ensure the accuracy of our analysis. Additionally, another issue we encountered while creating the data visualization was within the Denver crime dataset. We needed to convert the Reported Dates column into a date format, so we used the as.Date() function and formatted it specifically to the "%m/%d/%Y" format, but this only applied for select problems.

In some cases, we also added columns to the datasets to make it easier to access and generalize the data for our analysis. For example, we created an Arrest Year column in the Denver dataset to get a count of the number of arrests within a year, which helped us with our analysis. Overall, this helped us simplify the datasets and create more generalizations within our datasets.

Lastly, throughout the coding process, we also used subsetting techniques to narrow down the data and make it specific to the problems we were trying to address. This helped us to clean and prepare the datasets for further analysis and insights.

Questions of Interest and Findings

1. **Has crime in Dallas and Colorado increased or decreased over time?**

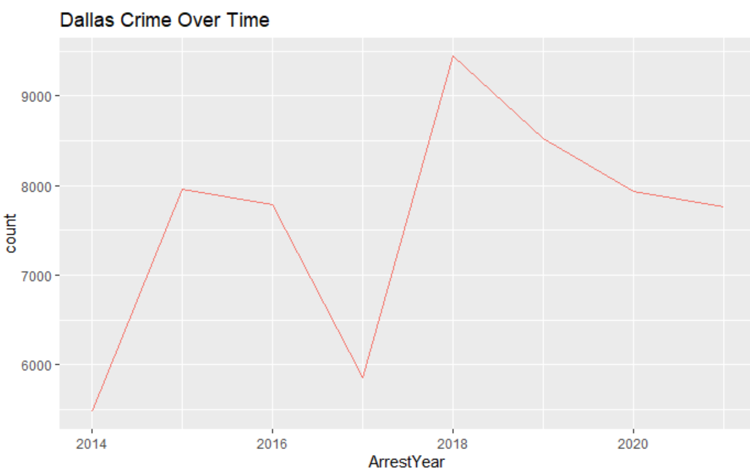


Figure 1.1

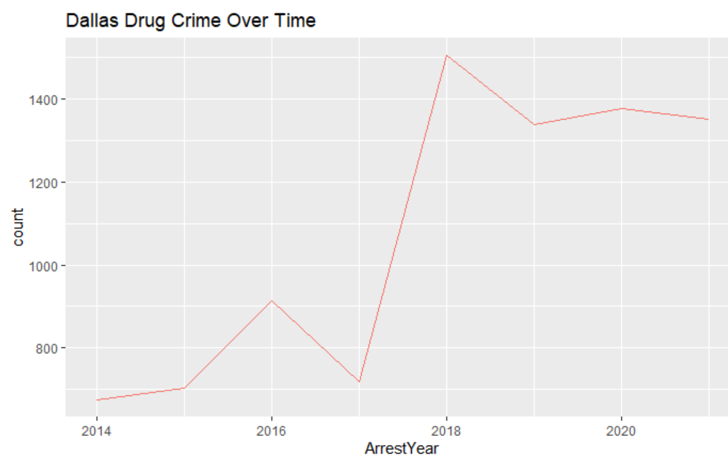


Figure 1.2

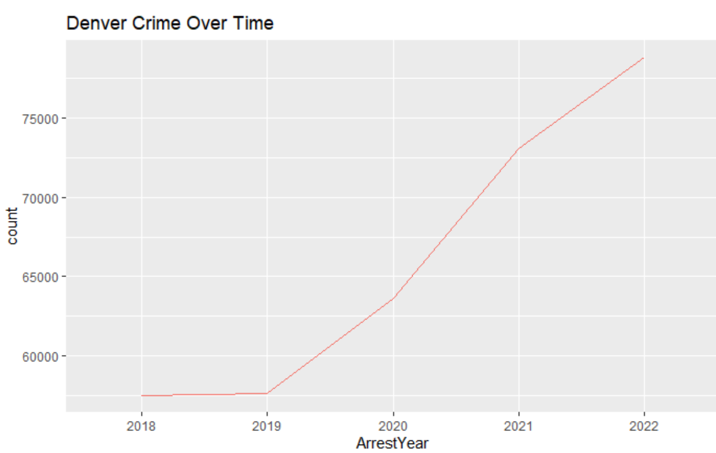


Figure 1.3

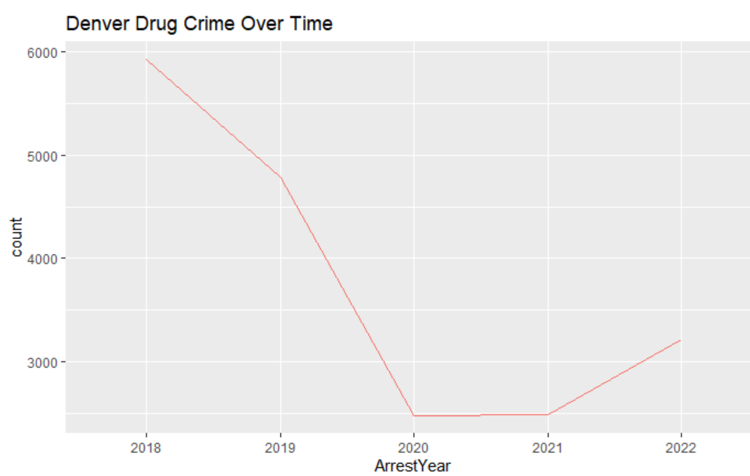


Figure 1.4

To answer this question, we need to analyze the trends in violent crime rates over time in both Dallas and Denver. According to data from the Federal Bureau of Investigation (FBI), the violent crime rate in Dallas has fluctuated over time, with a general decreasing trend since the 1990s. For example, in 1990, the violent crime rate in Dallas was 1,671.7 incidents per 100,000 residents, while in 2020, it was

957.6 incidents per 100,000 residents. This represents a decrease of approximately 42% over the past 30 years. Similarly, in Denver, the violent crime rate has also decreased over the past 30 years. In 1990, the violent crime rate in Denver was 1,439.9 incidents per 100,000 residents, while in 2020, it was 692.6 incidents per 100,000 residents. This represents a decrease of approximately 52% over the past 30 years.

As per our plots, we can see that general Dallas crime has fluctuated over the time period of the dataset, with an overall general higher count of cases in the 7 year span (figures 1.1). The incidence of drug-related crime in Dallas has exhibited a significant increase from 2017 to 2018, followed by a consistent decline from 2018 to 2020, as depicted in Figure 1.2. In contrast, the trend for general crime in Denver shows a consistent increase, as evident in Figure 1.3. However, despite the overall rise in crime incidents, there has been a substantial decline in drug-related offenses, particularly after the year 2020. We believe this may have been caused due to the effects of COVID-19 as people were less likely to go outside, therefore resulting in a general dip in reported cases. But, the same cannot be said about Dallas drug crime as the drug crime counts in Dallas were higher than usual during the 2018-2021 span. Therefore, we were able to notice some substantial shifts in how the nature of general and drug crime occurred in both cities from the respective spans of the datasets.

Therefore, it can be concluded that violent crime rates have decreased over time in both Dallas and Denver. However, it is important to note that there may be fluctuations in crime rates from year to year, and that other factors such as changes in law enforcement, COVID-19, strategies or demographic shifts may also have impacted crime rates.

2. **How does the percentage of population (in 2019 and 2020) Dallas compare to Denver with respect to drug/narcotic violations?**

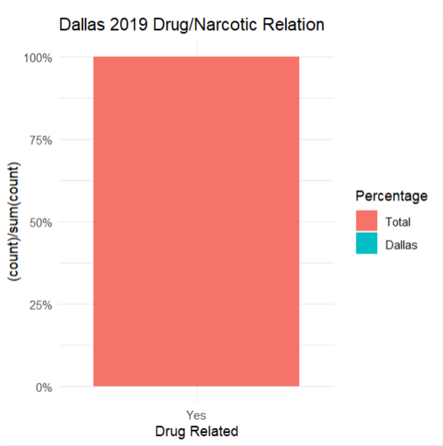


Figure 2.1

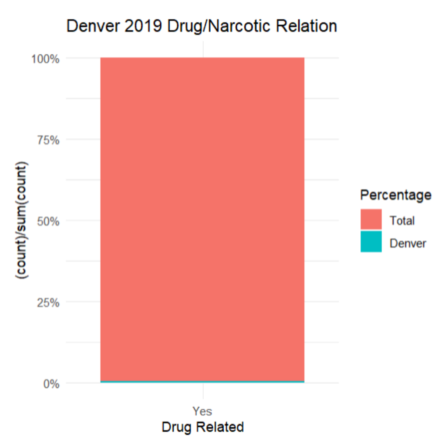
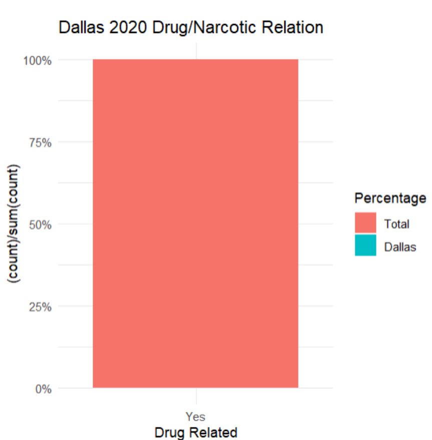


Figure 2.2



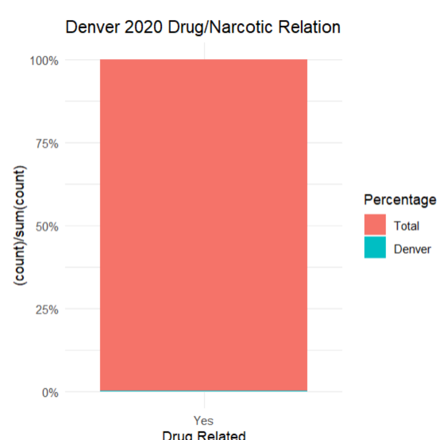
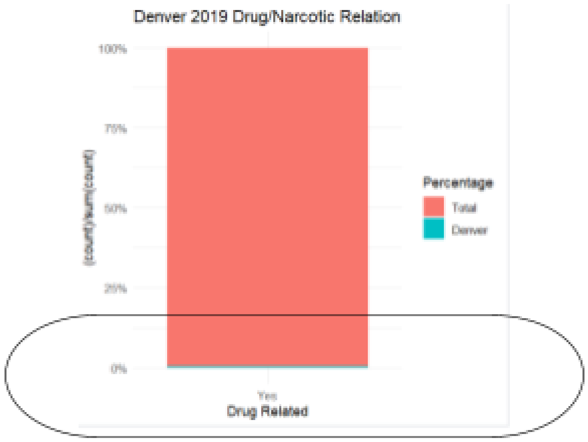


Figure 2.3 Figure 2.4

Important Note:



If you look closer you should barely see drug/narcotic relations because of it's very small percentage compared to total population.

The percentage of population in Dallas and Denver with respect to drug/narcotic violations was analyzed using data from 2019 and 2020. Four charts were used to depict the trends in drug/narcotic violations in these two cities.

The first chart shows the percentage of the population in Dallas involved in drug/narcotic violations in 2019. The data reveals that the percentage of population involved in such violations was less than one percent, indicating a relatively low prevalence of drug/narcotic violations in Dallas during that year. The second chart displays the percentage of the population in Denver involved in drug/narcotic violations in 2019. Similar to Dallas, the data shows that the percentage of population involved in drug/narcotic violations in Denver was also less than one percent in 2019, indicating a comparable prevalence of drug/narcotic violations in both cities during that year. The third chart presents the percentage of the population in Dallas involved in drug/narcotic violations in 2020. The data suggests a slight increase in the percentage of population involved in drug/narcotic violations compared to 2019, but it remains less than one percent. The fourth chart shows the percentage of the population in Denver involved in drug/narcotic violations in 2020. Similar to Dallas, the data shows a slight increase in the percentage of population involved in drug/narcotic violations compared to 2019, but it also remains less than one percent.

Additionally, it should be noted that due to the large population sizes of Dallas and Denver, the low percentages of population involved in drug/narcotic violations may not be visually prominent on the charts. The total population size of these cities may have influenced the scale of the charts, making it challenging to discern the small percentage values

Overall, the analysis of the four charts indicates that the percentage of population involved in drug/narcotic violations in both Dallas and Denver was relatively low, with less than one percent of the population involved in such violations in both 2019 and 2020. These findings suggest that the prevalence of drug/narcotic violations in Dallas and Denver was similar during the analyzed period, with slight increases in both cities in 2020 compared to 2019. Further analysis and research may be needed to understand the underlying factors contributing to drug/narcotic violations in these cities and to inform appropriate strategies for addressing and preventing such violations.

3. How does the legalization of marijuana in Colorado affect the crime rate in Denver compared to Dallas?

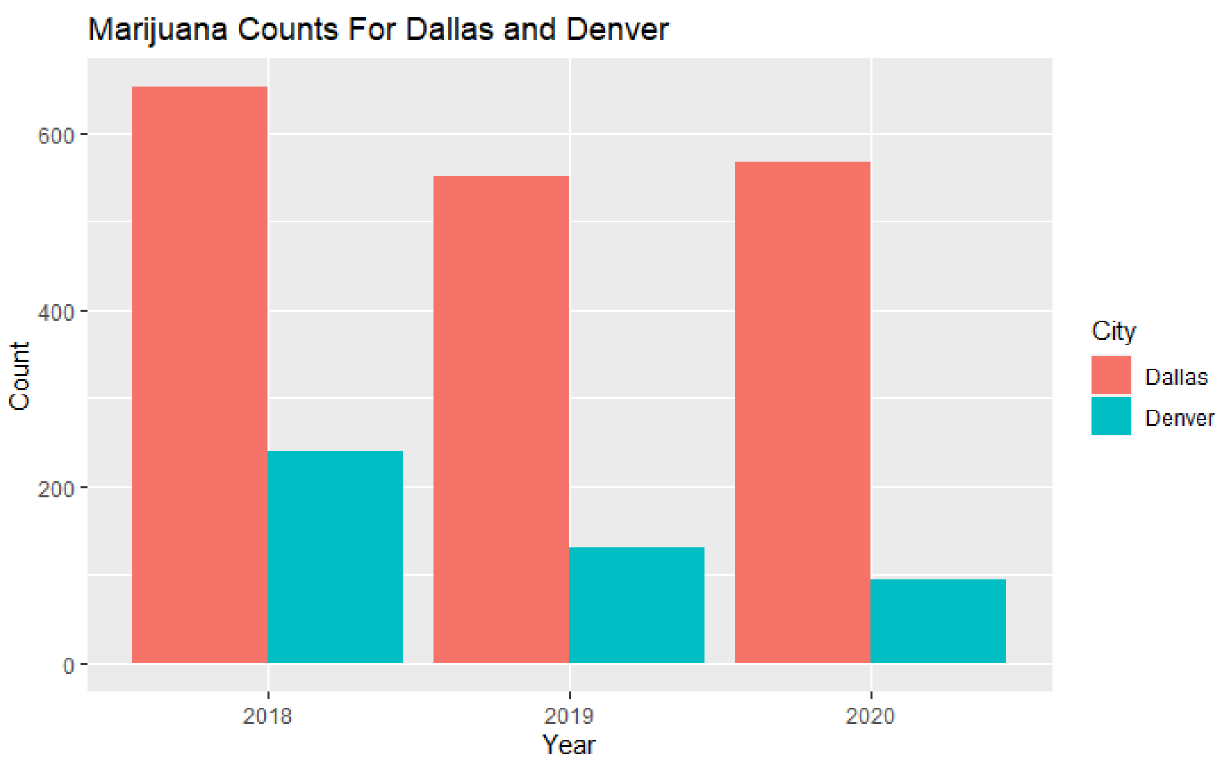


Figure 3: bar plot of marijuana counts for Dallas and Denver.

In January 2014, the legalization of marijuana in Colorado through the passage of Amendment 64 marked a significant shift in drug policy. This change has generated interest in understanding the potential impact of marijuana legalization on crime trends in Colorado, particularly in Denver, the state's capital. To gain insights into this issue, we wanted to do an analysis of crime trends between Denver and Dallas, specifically in Marijuana drug crime cases.

The statistical report presents a comparison of crime incidents from 2018 to 2020 for Dallas and Denver. It shows that Dallas has a significantly higher crime rate compared to Denver, and the crime rate in Dallas does not exhibit a clear trend. In contrast, Denver's crime rate shows a downward trend, with incidents dropping from above 200 in 2018 to below 100 in 2020 (as shown in figure 3). One thing to keep in mind is the potential influence of Covid-19 pandemic in the later stages of 2019 and 2020.

4. **Which locations within Dallas and Denver has the highest incidence of violence, and is it correlated with the location of highest drug crime rate?**

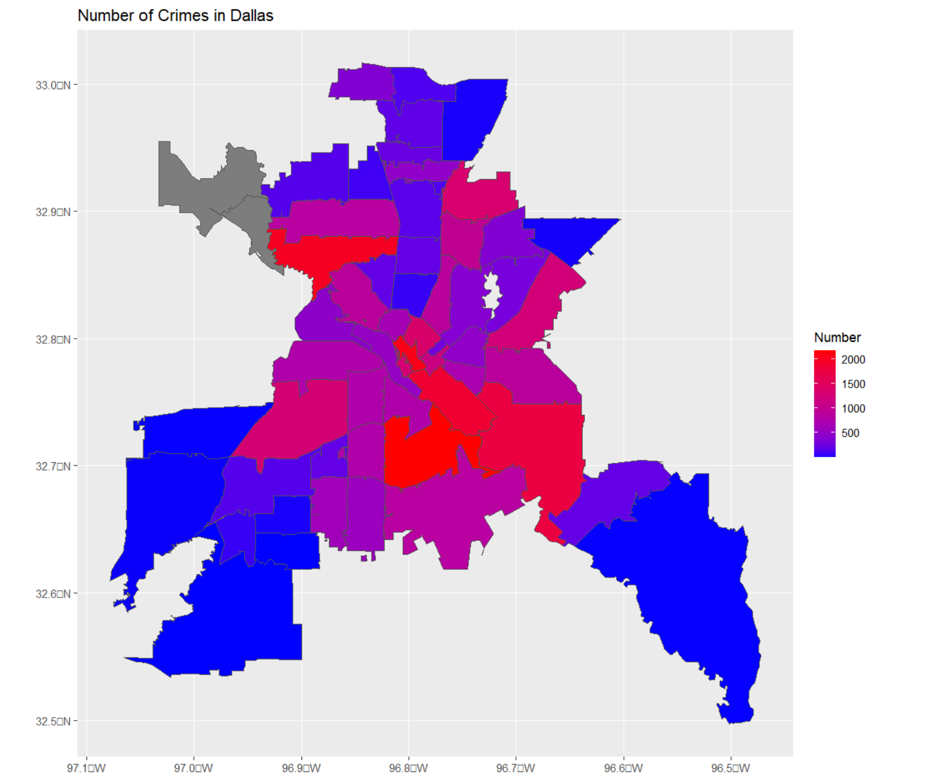
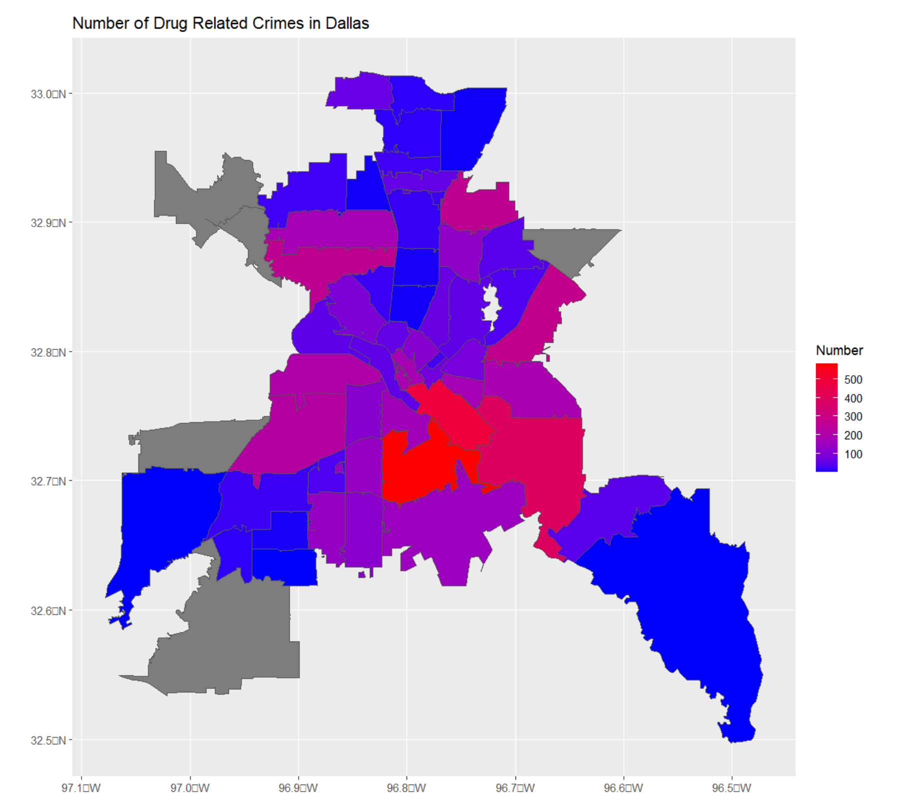


Figure 4.1: Dallas Map



The methodology employed to address this inquiry involved conducting a thorough analysis of crime data for Dallas which correlated violent crime rates with drug crime rates. Understanding which specific areas in Dallas have higher levels of violence and drug-related crimes can help identify potential hotspots that may require targeted interventions and resource allocations. Additionally, investigating the correlation between violence and drug crime rates in different locations within the city can provide valuable information for law enforcement agencies, policymakers, and community organizations to develop evidence-based strategies to address crime and improve public safety**.**

Figure 4.1 presents a cartogram of the frequency of crime-related incidents in Dallas. The data indicates that there is a noticeable trend of increasing crime occurrences as we move closer to the center of Dallas, with the highest peak observed in downtown Dallas. The graph suggests that the concentration of crime incidents tend to be higher in areas with denser populations.. Therefore, we can infer that population density is a contributing factor to crime rates in highly populated areas of Dallas, such as the downtown region.

Factors such as socioeconomic disparities, social disorganization, cultural and ethnic diversity, and the environment. create a space where individuals may be more likely to engage in criminal activities. Additionally, the breakdown of social bonds,weakening of community organizations, and rising tensions between conflicting groups contribute to increasing crime rates. The physical characteristics of densely populated areas, such as poorly lit streets and abandoned buildings provide criminals with opportunities to commit these crimes which creates an atmosphere of fear and insecurity among residents.

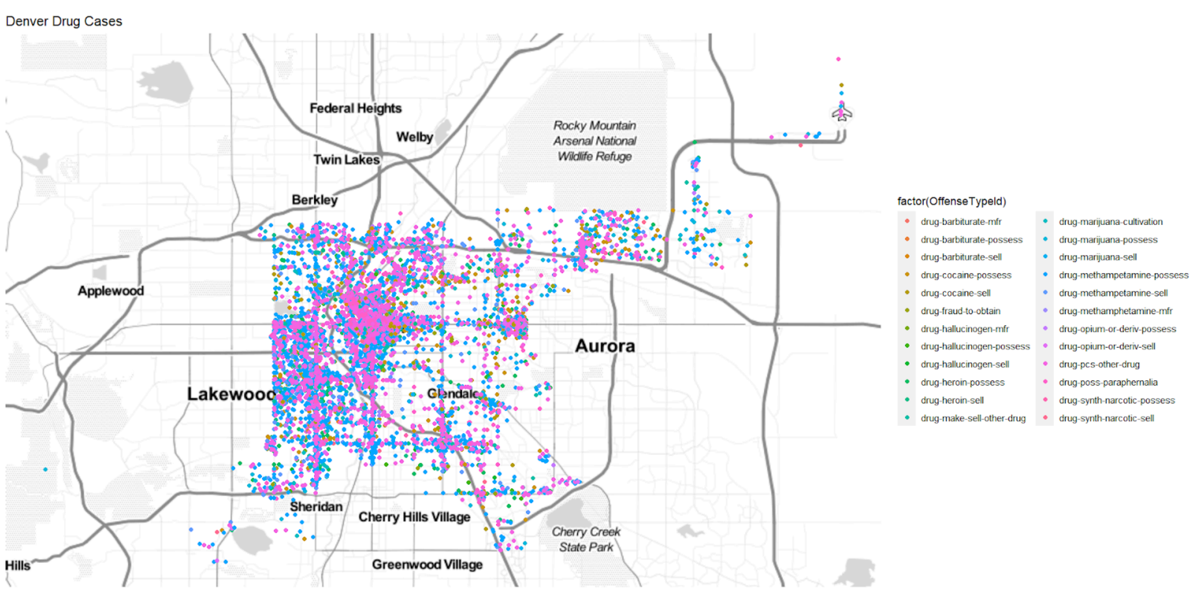
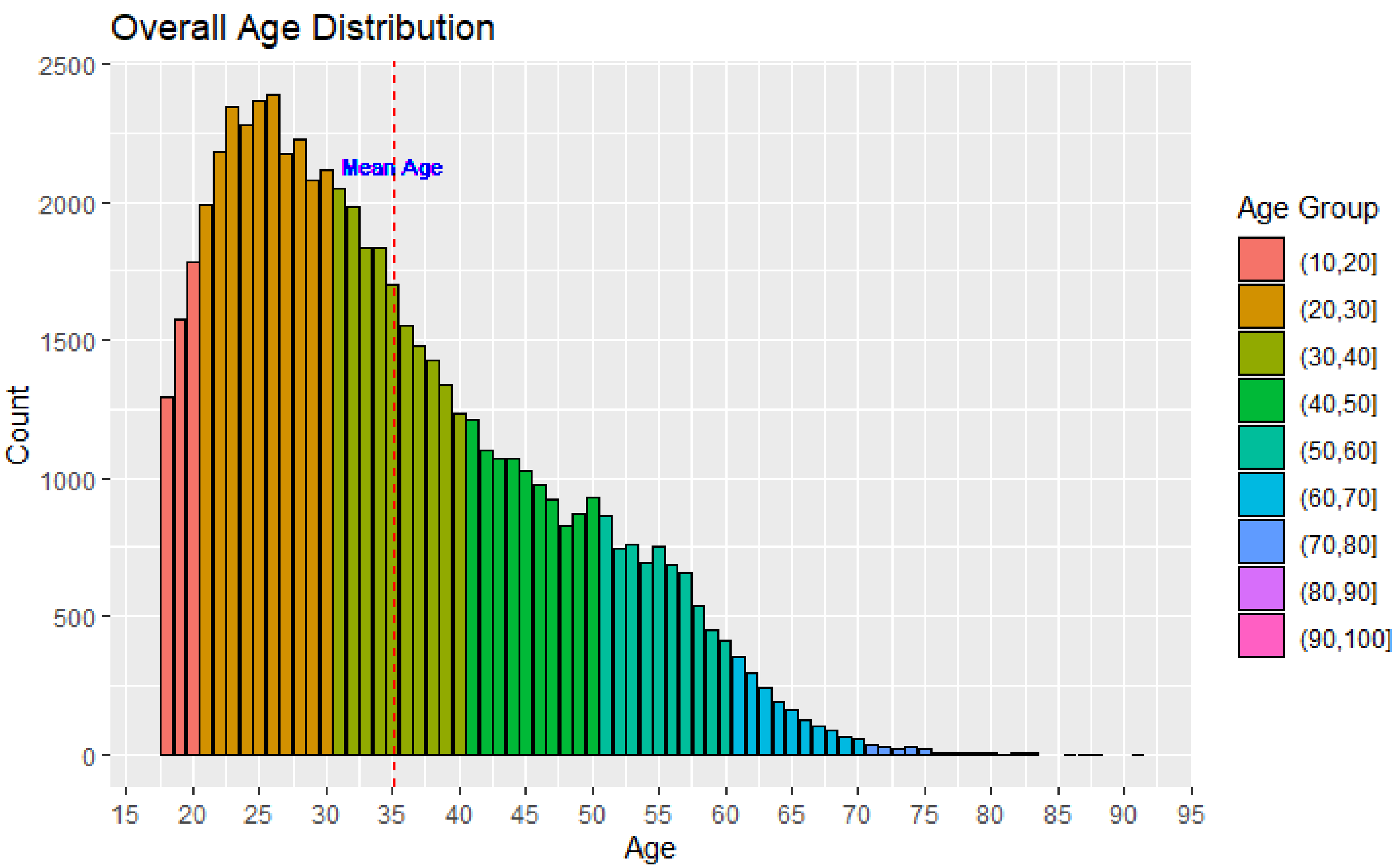
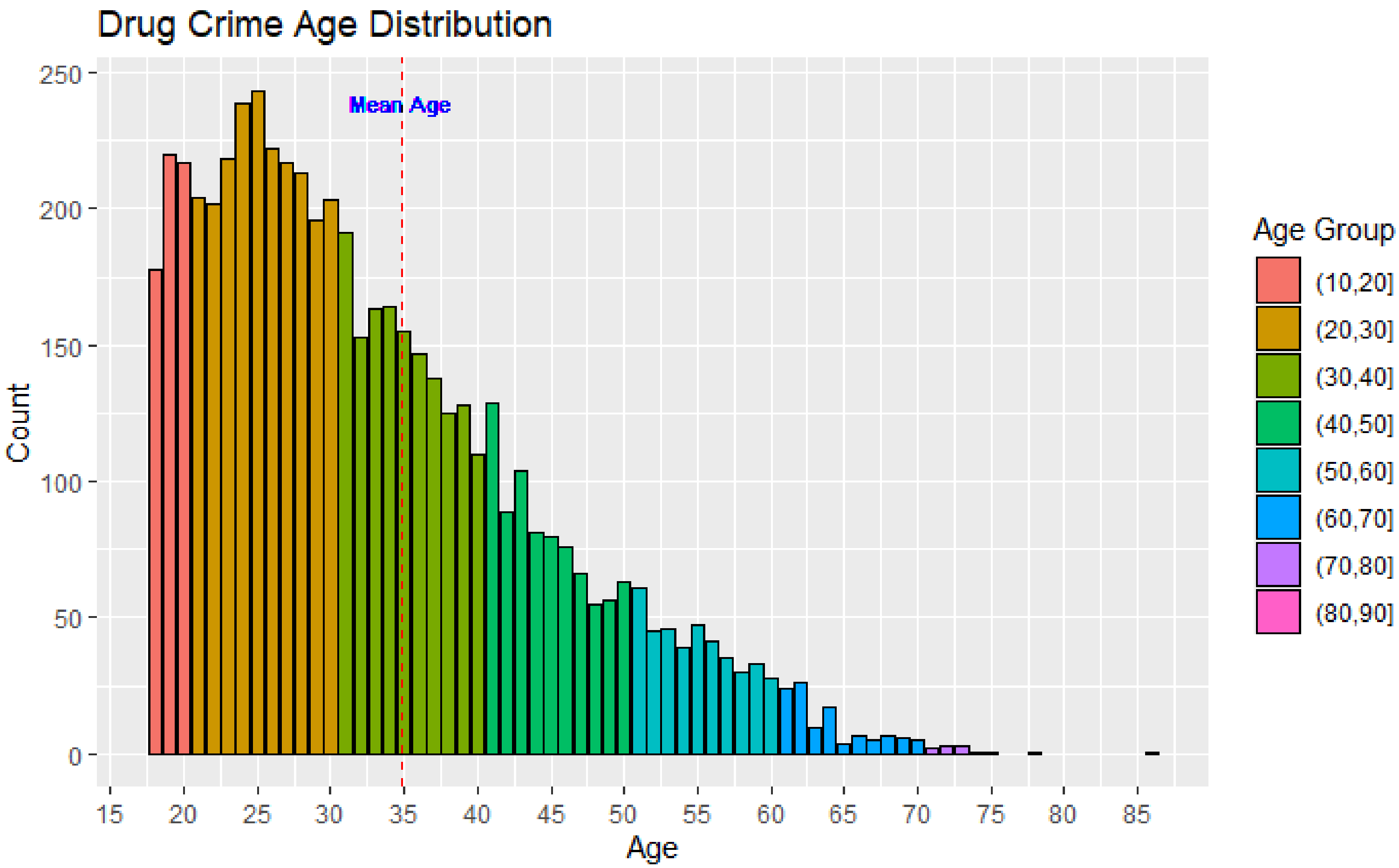


Figure 4.2: Denver Map

Figure 4.2 shows the points where drug related crime incidents occurred in Denver. At a quick glance we see that the points tend to pack around the top left-hand corner and spread outwards towards the top right and bottom left (creating a 'L' shape). Interestingly, this particular area of Denver corresponds to the city's minority neighborhoods and areas with lower percentages of individuals educated beyond highschool. Using this information, we can infer a potential correlation between the location of crime incidents, the occurrence of high drug-related crimes in Denver's minority neighborhoods, and areas with lower education levels. The concentration of drug-related crime incidents in this specific 'L' shaped area may indicate underlying social and economic factors that contribute to the prevalence of drug-related crimes. Further investigation and potential targeted interventions in these communities will aid in addressing the root causes of crime and promote community safety.

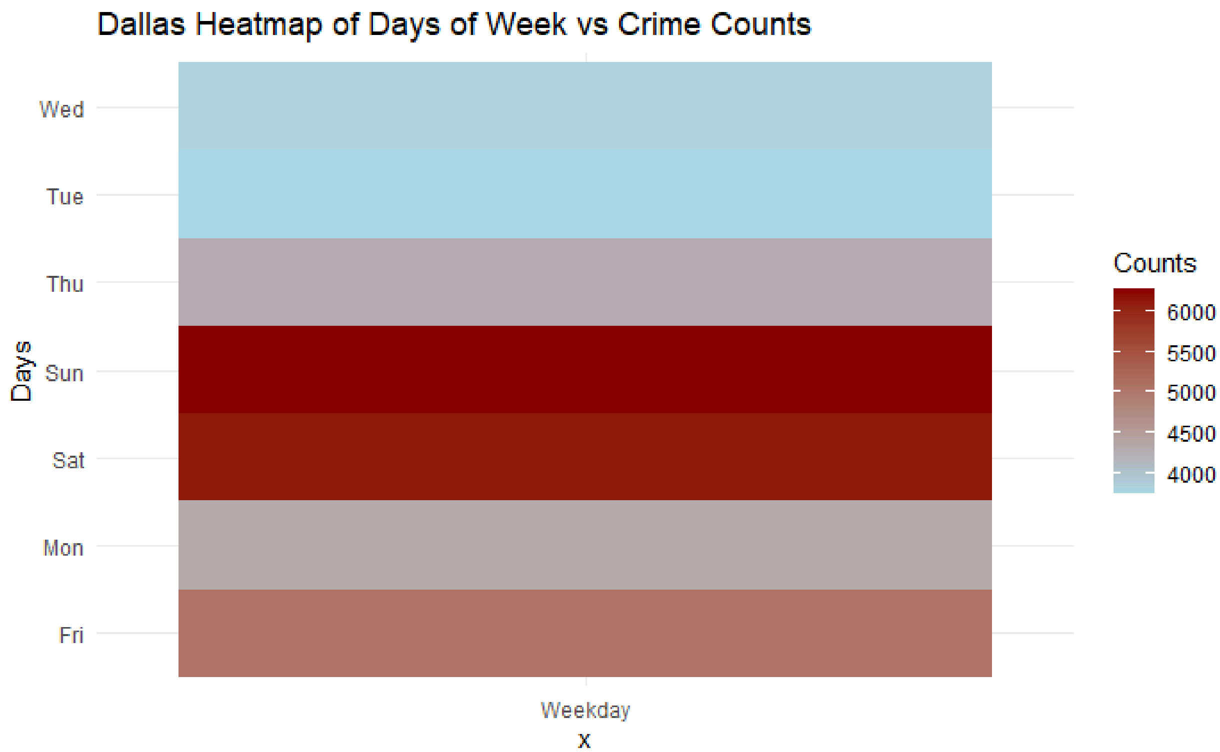
5. **Which age group tends to commit more crime?**



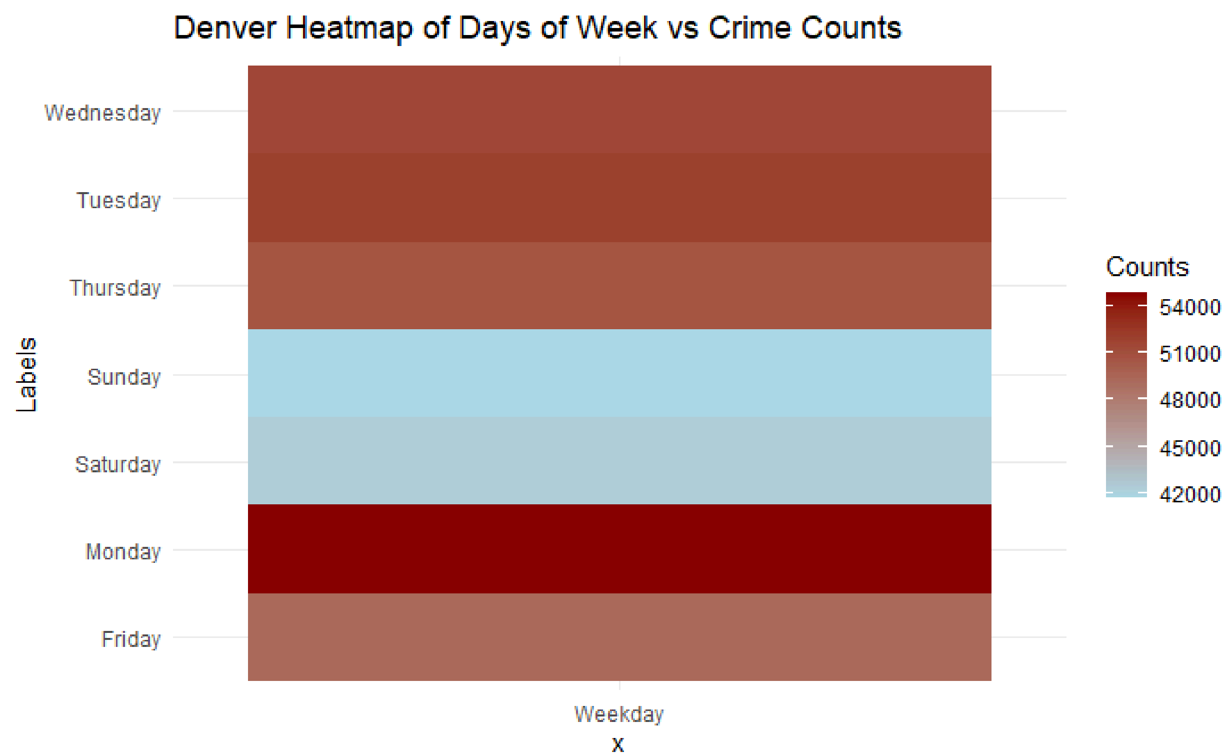


The analysis of crime data in Dallas relating to general crime reveals interesting trends in terms of the age groups that tend to commit more crimes in the two categories (Dallas in general and Dallas drug crime). In Dallas, the age group that exhibits a higher tendency to engage in criminal activities is between the ages of 20 to 30, with a peak occurring around 27 years. Similarly, in the Dallas drug crime plot, it can be seen that the 20 to 30 year old demographic is having the highest reported frequency. However, the peak value is lower, hovering around 25 years old. These findings highlight the significance of age as a factor influencing crime trends in both categories, with a notable difference in the specific age group that exhibits higher criminal involvement. Further research could explore the underlying reasons behind these patterns and inform targeted interventions and prevention strategies aimed at addressing crime among these age groups and categories in Dallas drug crime.

6. **What period of time is more probable to have crime occur in Dallas in terms of day of the week?**

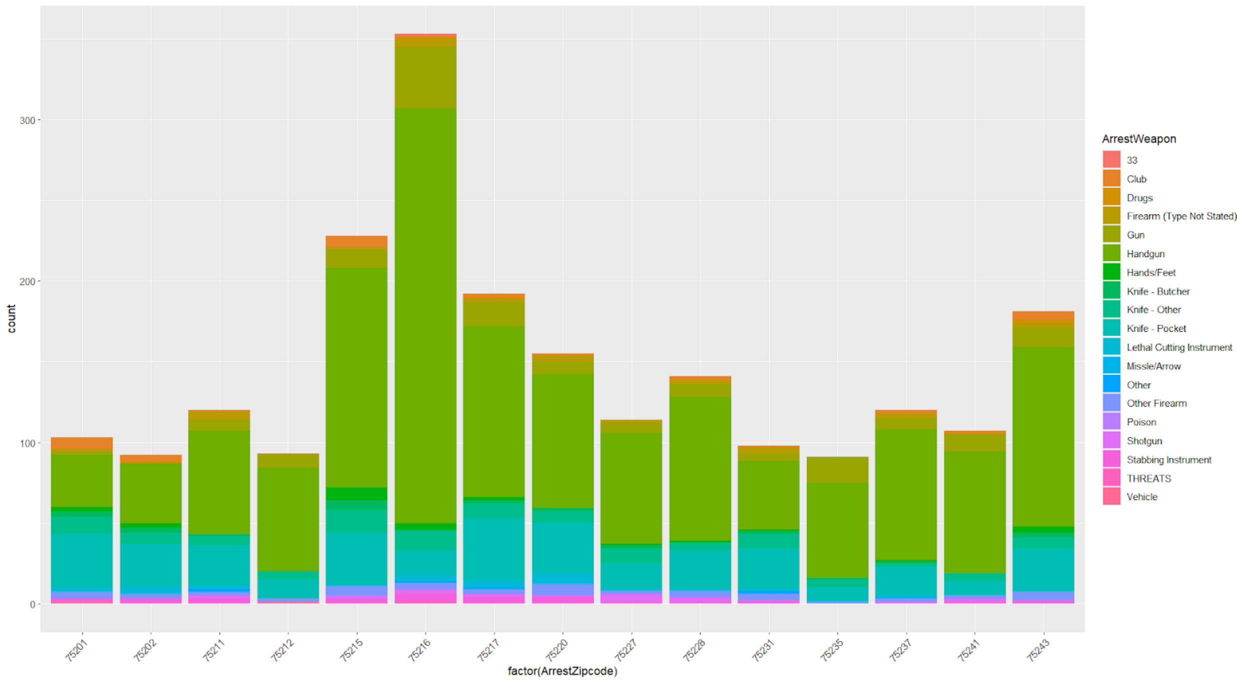


Dallas' data indicates that Saturdays and Sundays are more probable to have higher crime occurrences compared to other days of the week, with crime incidents peaking on Saturdays. Additionally, Fridays also show an increase in crime activity, albeit to a lesser extent than Saturdays and Sundays. These findings highlight the importance of weekends, particularly Saturdays and Sundays, as periods of higher risk for crime in Dallas, with Fridays also exhibiting increased crime activity. The late evening hours are also significant in terms of heightened crime occurrences. Further research could delve into the underlying factors contributing to these patterns and inform strategies for targeted policing, community engagement, and crime prevention efforts during these high-risk periods to enhance public safety in Dallas.



Denver's data indicates that Mondays are more probable to have the highest crime occurrences compared to other days of the week, followed by Wednesdays and Tuesdays, which also exhibit significant crime activity. In contrast, Saturdays and Sundays tend to have the lowest crime occurrences which is the exact opposite of Dallas. These findings suggest that weekdays, particularly Mondays, Wednesdays, and Tuesdays, may be periods of higher risk for crime in Denver, while weekends, specifically Saturdays and Sundays, tend to have lower crime activity. The late afternoon and early evening hours are also significant in terms of heightened crime occurrences given that the analysis shows that crime tends to peak during the late afternoon and early evening hours, particularly between 4:00 PM and 8:00 PM. . Further research could explore the underlying factors contributing to these patterns and inform strategies for targeted policing, community engagement, and crime prevention efforts during these high-risk periods to enhance public safety in Denver.

7. Is there any correlation between different variables such as age, race, location, and nature of offense in relation to crime rate?

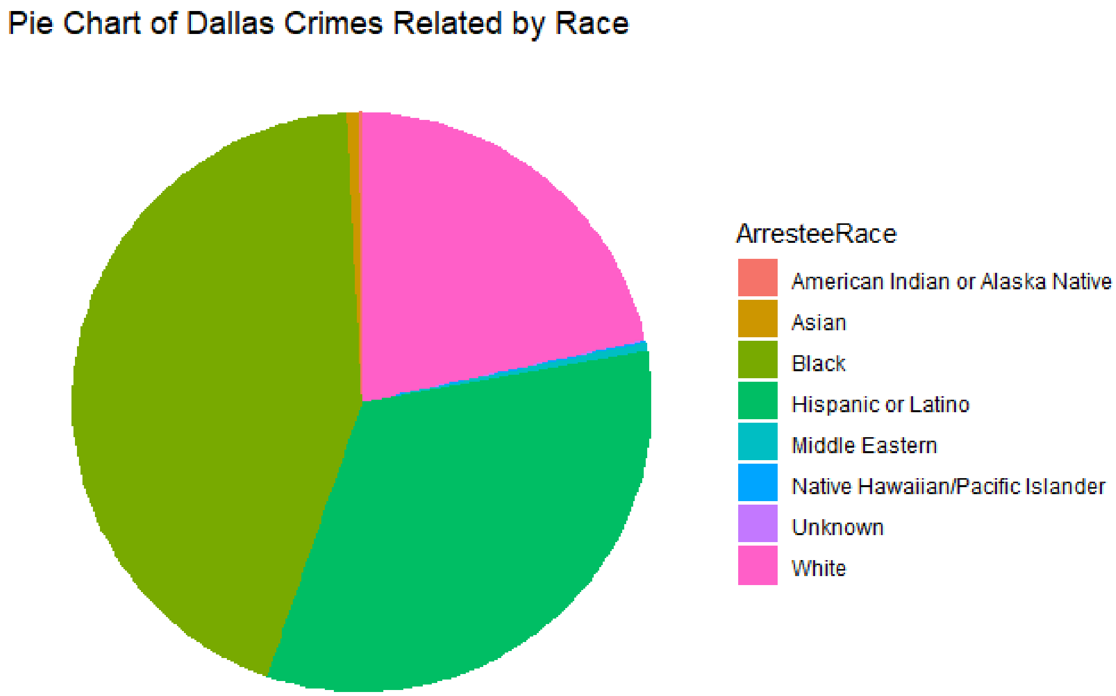


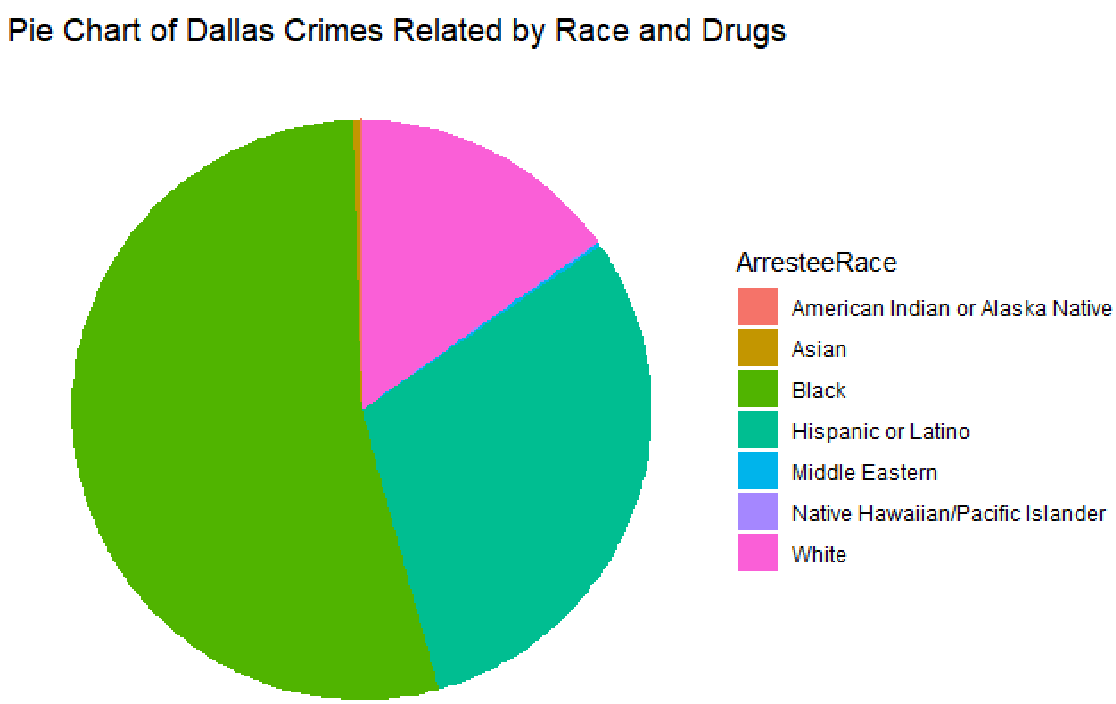
Spearman Correlation: -0.9978603 between number of armed crimes and zip code

Based on the provided figure and the given Dallas crime data set, it can be observed that zip code 75216, which is closer to downtown, has the highest apprehensions related to hand gun offenses, while zip code 75235 has the lowest apprehensions. This suggests that there may be a variation in crime rates based on location, specifically in relation to hand gun offenses, with higher crime rates observed in zip code 75216 and lower crime rates in zip code 75235.

Furthermore, the question of whether there is any correlation between different variables such as age, race, location, and nature of offense in relation to crime rate remains open. The given information about the highest and lowest apprehensions by zip code does not provide direct insights into the correlation between these variables so it would not be possible to determine correlations between these other factors currently. To determine correlations, further analysis and statistical techniques such as regression analysis, correlation analysis, and multivariate analysis may be needed to explore potential relationships between variables such as age, race, location, nature of offense, and crime rates in the given data set. Additional data and more comprehensive analysis would be required to draw any conclusive findings regarding the presence or absence of correlations among these variables in the context of the Dallas crime data set.

8. **Which ethnic group has the highest crime-related cases in Dallas, and is it proportional to the population distribution of ethnic groups?**





From our findings, the ethnic group with the highest number of offenses is the Black community, totaling 2999 offenses, which accounts for 53.8% of the total cases. Hispanic or Latino individuals follow closely behind with 1691 offenses, making up 30.3% of the total cases. The White community comes in third with a total of 843 offenses, representing 15.1% of the cases. Asian individuals accounted for 23 offenses, which is 0.4% of the total cases. Middle Eastern individuals had 12 offenses, accounting for 0.2% of the cases. Native Hawaiian individuals had 3 offenses, making up 0.05% of the cases, and American Indian or Alaska Native individuals had the lowest number of offenses at 2, which represents 0.04% of the total cases.

When comparing crime-related cases to the population distribution of ethnic groups in Dallas, it appears that there may be some disparities. While Blacks make up 53.8% of the crime-related cases, their population distribution in Dallas is not the same percentage (only 24.0%), leading to some disparities. White individuals, unlike the Black community, account for 15.1% of the crime-related cases, while making up 53.8% of Dallas' population. Similar to the white community, Hispanics or Latinos make up 30.3% of the crime-related cases, they comprise 42.0% of Dallas' population. However, it's important to note that crime statistics can be influenced by various factors, such as socioeconomic status, education level, and systemic biases in the criminal justice system, and cannot be solely attributed to an ethnic group's population distribution. Further analysis and contextual information would be required to draw definitive conclusions about the proportional representation of different ethnic groups in crime-related cases in Dallas.

9. **What percentage of reported criminals in Dallas were carrying firearms and its association to age?**

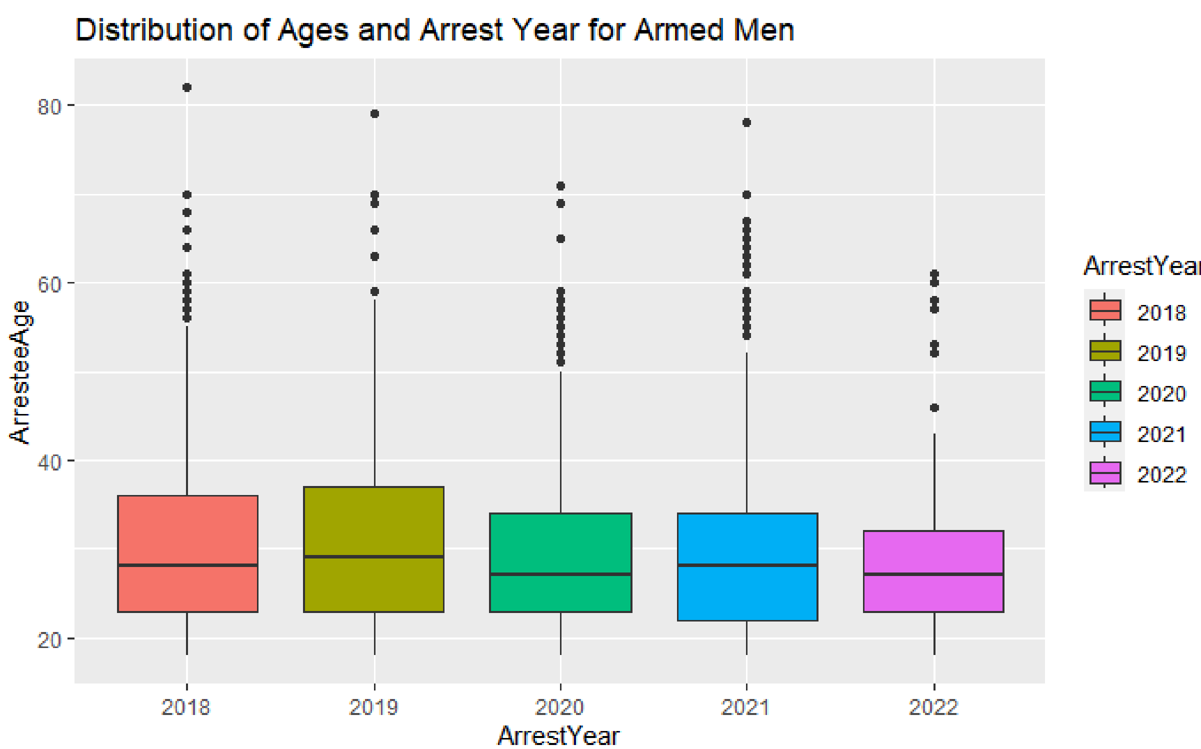


Figure 9.1 Dallas General

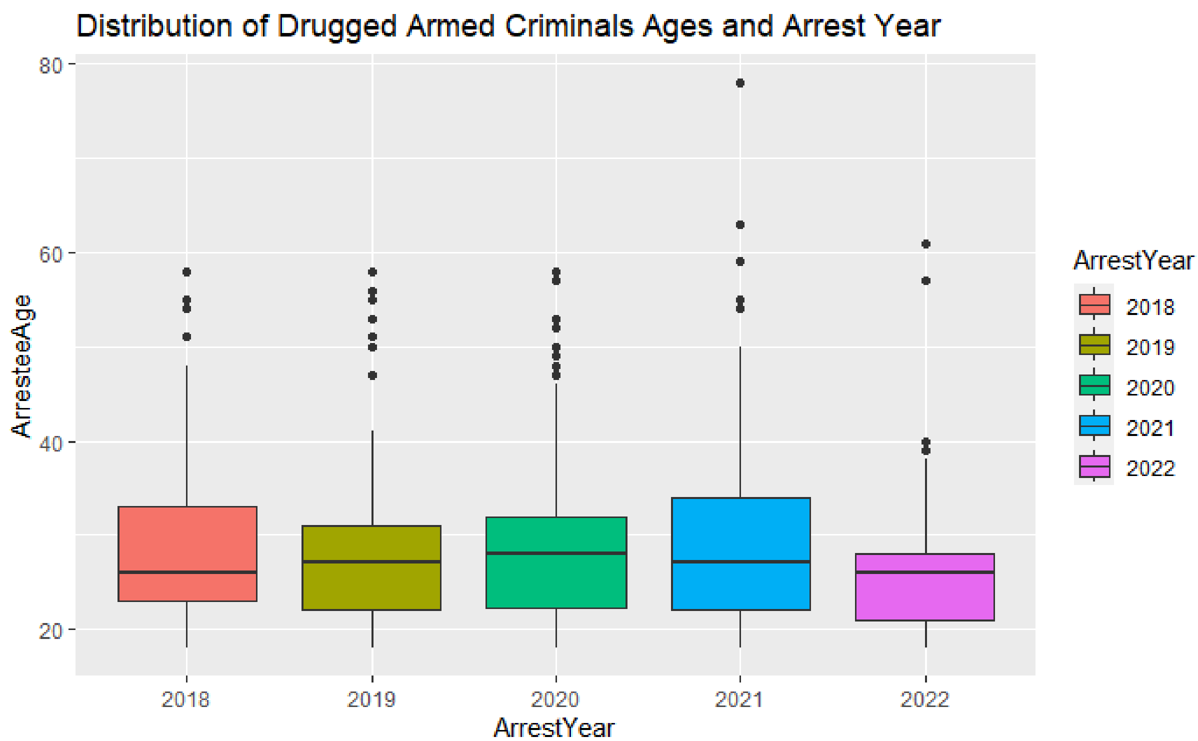


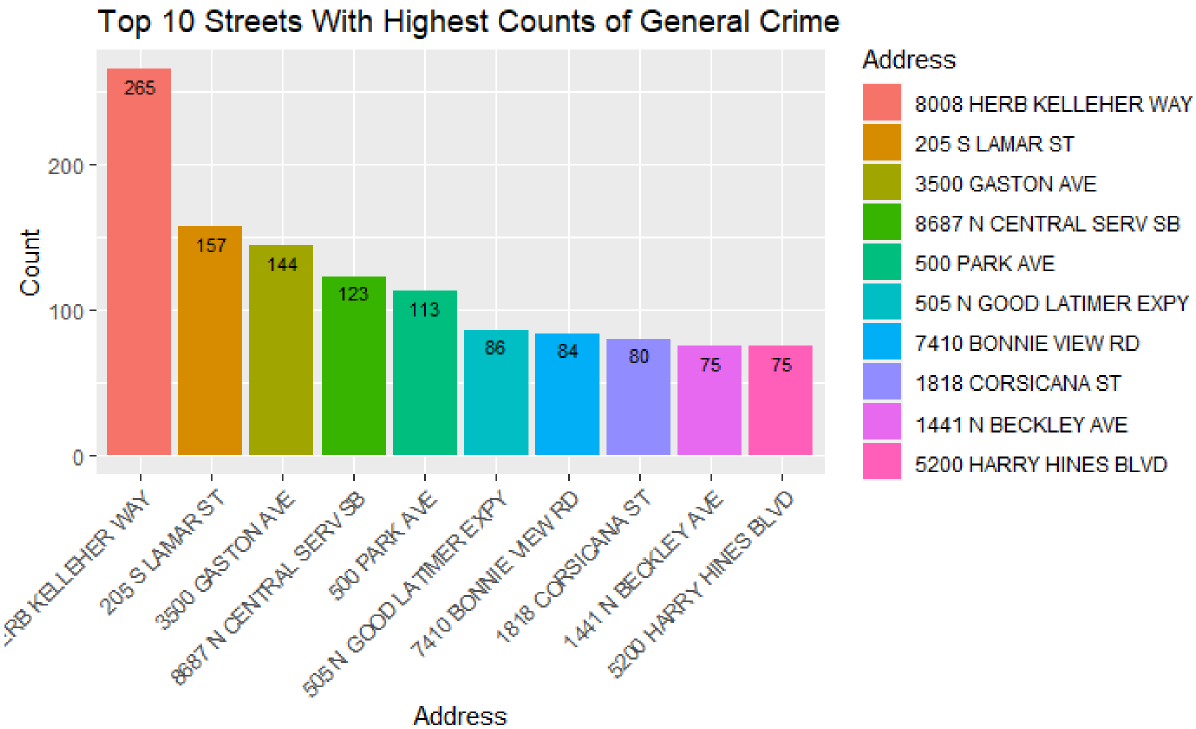
Figure 9.2 Dallas Drug Crime

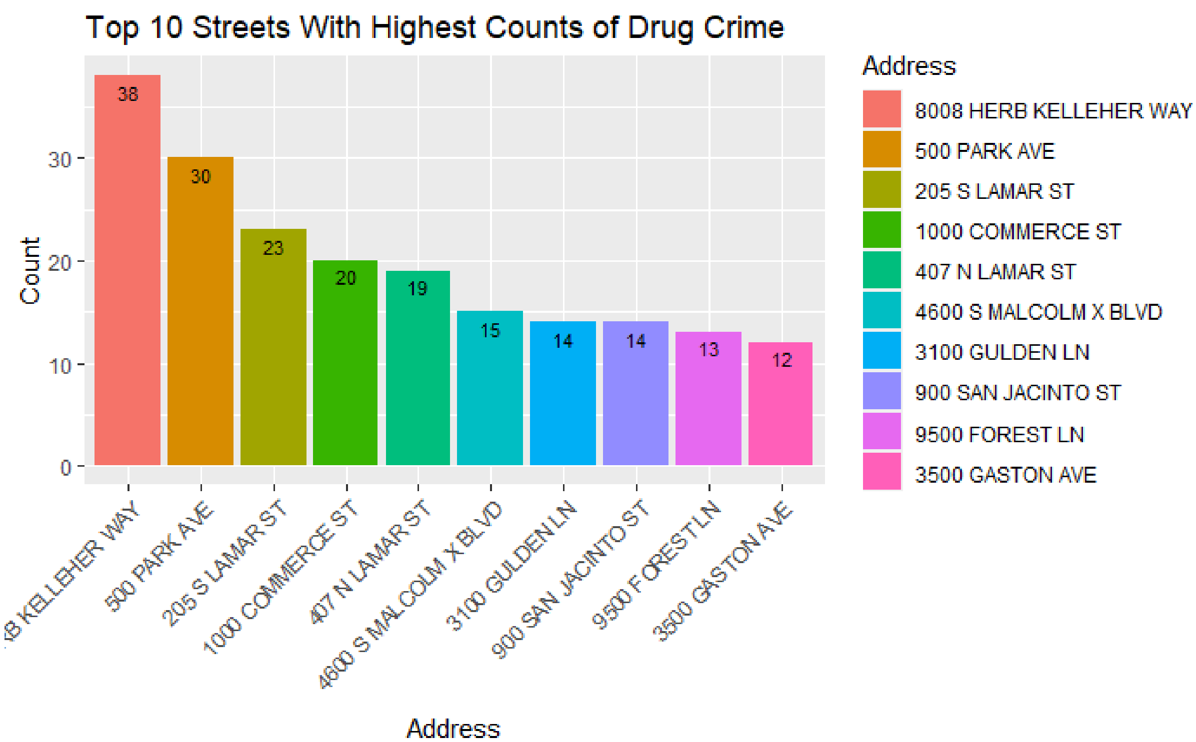
Based on the given data, our objective was to assess the prevalence of firearms possession among criminals. Based on our findings, we observed evidence of firearms possession among a proportion of the criminal population. Specifically, from 2018 to 2020, 6.62% of all reported criminals were found to be carrying firearms. Delving deeper, 1.96% of all reported criminals were found to have possession of both drugs and firearms. Lastly, among reported criminals who were involved in drug-related offenses, 11.84% were found to be carrying some type of firearm.

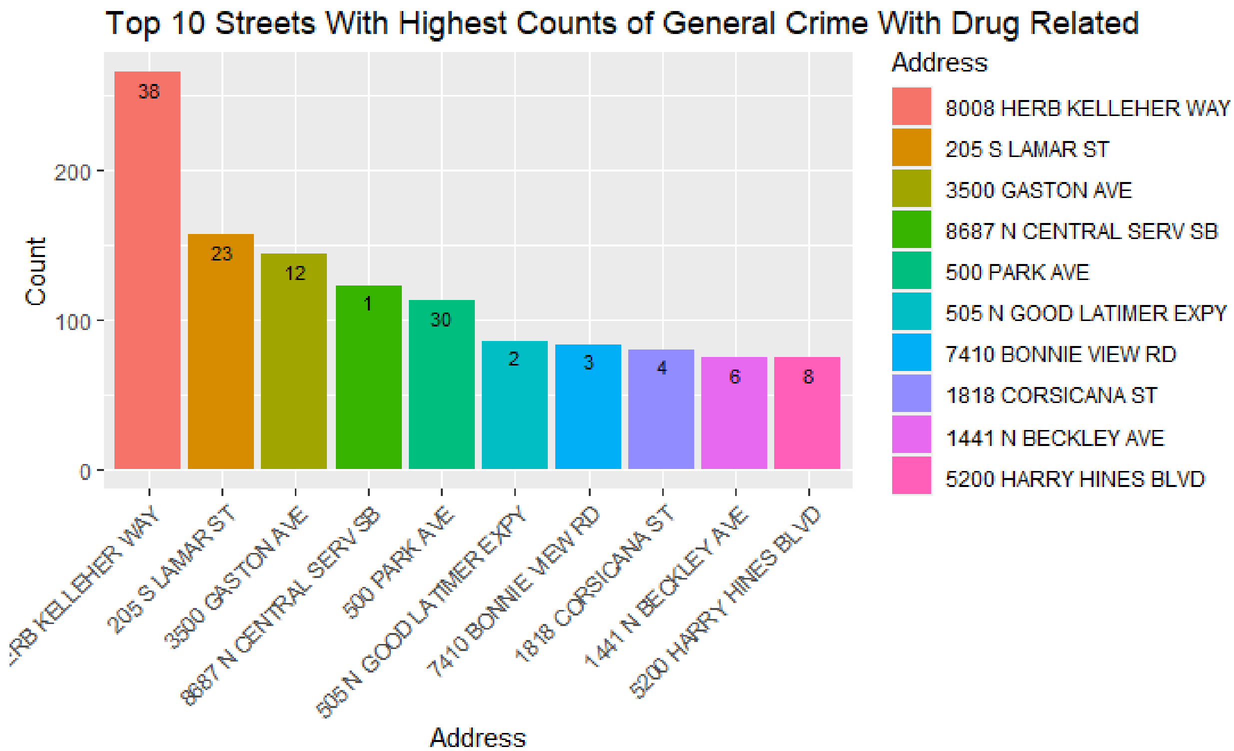
This data suggests that a non-negligible proportion of reported criminals in Dallas were carrying firearms during the given time period. The percentage of criminals carrying firearms was higher among those involved in drug-related offenses, with approximately 11.84% of drug-related criminals carrying firearms. Looking at the data, the first, second, and third quartiles, which represent the 25th, 50th, and 75th percentiles respectively, are lower (in age) in the drug distribution compared to the overall distribution. This suggests that, on average, individuals involved in drug-related crimes tend to be younger compared to those involved in other types of crimes in Dallas.

However, it's important to note that while there is a trend of younger age associated with drug crimes, the difference may not be statistically significant. This implies that while there may be a tendency for drug crimes to involve younger individuals, further statistical analysis would be needed to determine if this difference is statistically significant or due to random variation. Additional data and more robust statistical methods may be required to establish the statistical significance of the age differences between overall crimes and drug-related crimes in Dallas.

10. Which streets have the most common crimes and compare them to drug crimes?







Analysis of crime data in Dallas reveals that there are certain streets that have higher counts of general crime compared to others. The top 10 streets with the highest counts of general crime are 8008 Herb Kelleher Way, 205 S Lamar St, 3500 Gaston Ave, 8687 N Central Serv Sb, 500 Park Ave, 505 N Good Latimer Expy, 7410 Bonnie View Rd, 1818 St, 1441 N Beckley Ave, and 5200 Harry Hines. Among these streets, 8008 Herb Kelleher Way has the highest count of general crimes.

On the other hand, when analyzing drug crimes specifically, a different set of streets emerges as the top 10. The streets with the highest counts of drug crimes are 8008 Herb Kelleher Way, 500 Park Ave, 205 S Lamar St, 1000 Commerce St, 407 N Lamar St, 4600 S Malcolm X Blvd, 3100 Gulden Ln, 900 San Jacinto St, 9500 Forest Ln, and 3500 Gaston Ave. Interestingly, while some streets such as 8008 Herb Kelleher Way and 500 Park Ave appear in both lists, the order and frequency of streets differ between general crimes and drug crimes.

Furthermore,an analysis of the relationship between streets with high counts of general crime and drug crimes reveals that there may not be a direct correlation. The percentage of drug crimes among the total instances of crime on each street varies significantly. For example, while 8667 N Central Serv Sb has a relatively low percentage of drug crimes (less than 1%), 500 Park Ave has close to 30% of instances being drug crimes. This suggests that there may be specific hotspots or areas that are more prone to drug crimes, independent of the general crime patterns in those areas.

In conclusion, the analysis of crime data in Dallas indicates that there are streets with higher counts of general crime and drug crimes, but the two sets of streets may not necessarily overlap completely. The variance in the percentage of drug crimes among the total instances of crime on each street suggests that there may be unique factors contributing to drug crime patterns in certain areas. Further research and analysis would be needed to understand the underlying factors and dynamics of crime patterns on specific streets in Dallas.

11. How does the crime rate in low economic areas of Dallas compare to high economic areas, and is it proportional to the socioeconomic status of an area?

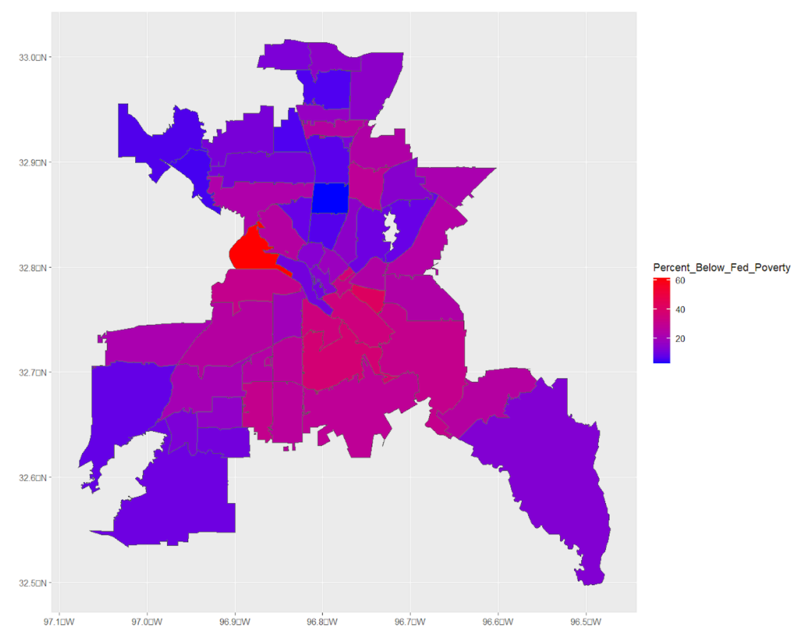


Figure 11.1 Dallas Map

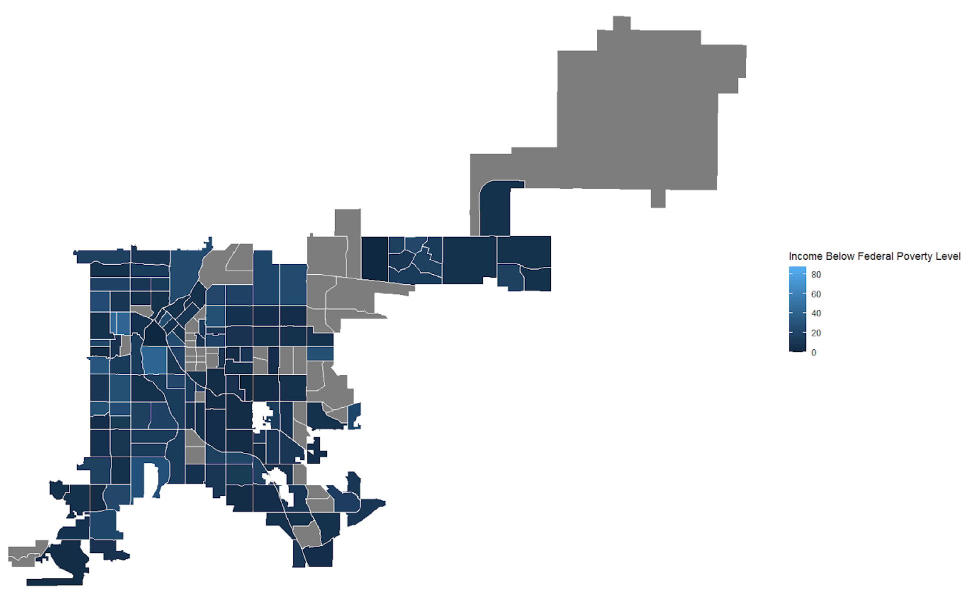


Figure 11.2 Denver Map

Figure 11.1 and 11.2 provide visual representations of the percentage of population below the federal poverty level in Dallas and Denver. In Dallas, the peak income below the poverty level is 61.15455% in zip code 75247, while in Denver, it is 76.8% in census tract 8. These maps highlight that both cities have areas with high levels of poverty, with Dallas showing a gradual increase in poverty towards the center, and Denver having a concentration of poverty in the top left area.

Pulling from previous questions, the data from question 3 also reveals a correlation between the percentage of population below the poverty level and crime rates in the neighborhoods. Areas with higher poverty percentages tend to have higher crime counts. This suggests that there may be a relationship between low economic status and crime in both Dallas and Denver.

Furthermore, the data from the Denver map indicates that areas with higher poverty percentages tend to have higher levels of minorities. This raises questions about the intersectionality of race and socioeconomic status in crime patterns, and whether there may be additional social and structural factors at play.

These findings are interesting and related as they highlight the complex interplay between poverty, race, and crime in Dallas and Denver. Understanding these dynamics can provide important insights for developing effective crime prevention strategies and addressing social inequalities in these communities.

Conclusion

In conclusion, the statistical report comparing crime patterns between Dallas and Denver indicates that the percentage of population involved in drug/narcotic violations was relatively low in both cities, with less than one percent of the population involved in such violations in both 2019 and 2020. The analysis of the four charts suggests that the prevalence of drug/narcotic violations was similar in both cities during the analyzed period, with slight increases in 2020 compared to 2019.

The report also highlights that Dallas has a higher overall Marijuana crime rate compared to Denver , with no clear trend in crime incidents over the analyzed period. In contrast, Denver's crime rate shows a downward trend, with incidents decreasing from above 200 in 2018 to below 100 in 2020. However, it is important to consider potential influences of the COVID-19 pandemic in the later stages of 2019 and 2020 on crime patterns. Possibly from the findings of this study, the city of Dallas can create improvements in how drug laws are approached to reduce the number of drug crimes.

Furthermore, our findings emphasize the role of socioeconomic disparities, social disorganization, cultural and ethnic diversity, and environmental factors in influencing crime patterns in densely populated areas like Dallas and Denver. The findings suggest that there is a correlation between low economic status and crime rates in both cities, with areas having higher poverty percentages tending to have higher crime counts. This underscores the importance of considering socioeconomic status as a key factor when examining crime patterns and developing crime prevention strategies.

Lastly, we conclude by noting that the data analysis indicates a decrease in violent crime rates over time in both Dallas and Denver, but fluctuations may occur from year to year, and other factors such as changes in law enforcement strategies or demographic shifts may also impact crime rates. Further research and analysis are needed to fully understand the underlying mechanisms driving the relationship between poverty and crime in these cities. Nevertheless, the insights from this statistical report can inform policy decisions and interventions aimed at reducing crime rates and addressing social inequalities in low economic areas of Dallas and Denver.

Code

```
# Load Necessary Packages
library(ggplot2)
library(dplyr)
library(stringr)
library(scales)
library(ggmap)
library(tigris)
library(tidycensus)
library(sf)
library(viridis)
library(forcats)
```

```
# Data Preparation
dallas_data <- read.csv("dallas_crime.csv")
denver_data <- read.csv("denver_data.csv")
denver_income <- read.csv("Income_Poverty_(Census_Tracts).csv")
dallas_shp <- st_read("STREETS.shp")
zip_shape <- st_read("DallasZipCodes_2018.shp")
dallas_poverty <- read.csv("indicator_data_download_20230420.csv")

names(dallas_data) <- c("IncidentNum", "ArrestYear", "ArrestDate", "ArrestTime", "ArrestAddress",
"ArrestZipCode", "Latitude", "Longitude", "ArrestCity", "ArrestState", "ArrestDay", "ArrestLocation",
"ArrestWeapon", "ArresteeAge", "ArresteeRace", "ArresteeSex", "DrugRelated", "DrugType")

names(denver_data) <- c("IncidentId", "OffenseId", "OffenseCode", "OffenseCodeExtension",
"OffenseTypeId", "DrugRelated", "FirstOccuranceDate", "LastOccuranceDate", "ReportedDate",
"IncidentAddress", "Geo_x", "Geo_y", "Geo_lon", "Geo_lat", "DistrictId", "PrecinctId",
"NeighborhoodId", "IsCrime", "IsTraffic", "VictimCount")

# Removing unnecessary columns
denver_data <- denver_data[, c(- 1, - 2, - 3, - 4, -7, -8)]
dallas_data <- dallas_data[, c(-1, -4, -12, -16)]
```

```r
# Question 1
dallas_data <- subset(dallas_data, dallas_data$ArrestYear <= 2021)
ggplot(data = dallas_data, aes(x = ArrestYear, color = "magenta")) +
  geom_line(stat = "count") + labs(title = "Dallas Crime Over Time") +
  theme(legend.position = "none")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2017 <- subset(denver_data, ReportedDateStandard >= as.Date("2017-01-01") &
ReportedDateStandard <= as.Date("2017-12-31"))
denver_2017$ArrestYear <- format(denver_2017$ReportedDateStandard, "%Y")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2018 <- subset(denver_data, ReportedDateStandard >= as.Date("2018-01-01") &
ReportedDateStandard <= as.Date("2018-12-31"))
denver_2018$ArrestYear <- format(denver_2018$ReportedDateStandard, "%Y")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2019 <- subset(denver_data, ReportedDateStandard >= as.Date("2019-01-01") &
ReportedDateStandard <= as.Date("2019-12-31"))
denver_2019$ArrestYear <- format(denver_2019$ReportedDateStandard, "%Y")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2020 <- subset(denver_data, ReportedDateStandard >= as.Date("2020-01-01") &
ReportedDateStandard <= as.Date("2020-12-31"))
denver_2020$ArrestYear <- format(denver_2020$ReportedDateStandard, "%Y")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2021 <- subset(denver_data, ReportedDateStandard >= as.Date("2021-01-01") &
ReportedDateStandard <= as.Date("2021-12-31"))
denver_2021$ArrestYear <- format(denver_2021$ReportedDateStandard, "%Y")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2022 <- subset(denver_data, ReportedDateStandard >= as.Date("2022-01-01") &
ReportedDateStandard <= as.Date("2022-12-31"))
denver_2022$ArrestYear <- format(denver_2022$ReportedDateStandard, "%Y")

denver_data <- rbind(denver_2017, denver_2018,
denver_2019,denver_2020,denver_2021,denver_2022)

ggplot(data = denver_data, aes(x = ArrestYear, group = IsCrime, color = "lavender")) + geom_line(stat =
"count") + labs(title = "Denver Crime Over Time") +
```

```
  theme(legend.position = "none")

dallas_data_drugs <- subset(dallas_data, dallas_data$DrugRelated == "Yes")
ggplot(data = dallas_data_drugs, aes(x = ArrestYear, color = "purple")) + geom_line(stat = "count") +
labs(title = "Dallas Drug Crime Over Time") +
  theme(legend.position = "none")

denver_data_drugs <- subset(denver_data, denver_data$DrugRelated == "drug-alcohol")
ggplot(data = denver_data_drugs, aes(x = ArrestYear, group = IsCrime, color = "pink")) +
geom_line(stat = "count") + labs(title = "Denver Drug Crime Over Time") +
  theme(legend.position = "none")
```

```
# Question 2
# Filter to find necessary info
denver_drugs <- filter(denver_data, DrugRelated == "drug-alcohol")
dallas_drugs <- filter(dallas_data, DrugRelated == "Yes")
denver_drugs$ReportedDate <- format(as.Date(denver_drugs$ReportedDate, format =
"%m/%d/%Y"), "%Y")
# Removed values not connect to drugs
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId != "liquor-possession"),]
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId != "liquor-sell"),]
dallas_data[dallas_data == ""] <- NA
denver_drugs$DrugRelated <- str_replace(denver_drugs$DrugRelated, "drug-alcohol", "Yes")


# Changed Values so it would match up with dallas_data
# Joined both data sets
City = c("Denver")
denver_drugs <- cbind(denver_drugs, City)
City = c("Dallas")
dallas_drugs <- cbind(dallas_drugs, City)

# Create ggplot specifically for 2019 for Denver
denver_2019 <- denver_drugs[which(denver_drugs$ReportedDate %in% 2019),]
# Created a separate data set that shows the total population of denver in 2019
den_total_2019 <- as.data.frame(matrix(ncol = 2, nrow = 725508))
names(den_total_2019) <- c("City", "DrugRelated")
den_total_2019$City <- rep(c("Total"), c(725508))
den_total_2019$DrugRelated <- rep(c("Yes"), c(725508))
# Join two data sets
```

```r
den_2019_drugs <- full_join(denver_2019, den_total_2019)
den_2019_drugs$City <- factor(den_2019_drugs$City, levels =
names(sort(table(den_2019_drugs$City), decreasing = TRUE)))
# Create ggplot appear as stacked to give the image of percentage
ggplot(data = den_2019_drugs) + geom_bar(mapping = aes(x = DrugRelated, y =
(..count..)/sum(..count..), fill = City), position = "stack") +
  scale_y_continuous(labels = percent) +
  labs(title = "Denver 2019 Drug/Narcotic Relation", x = "Drug Related", fill = "Percentage") +
  theme_minimal()

# Create ggplot specifically for 2020 for Denver
denver_2020 <- denver_drugs[which(denver_drugs$ReportedDate %in% 2020),]
# Created a separate data set that shows the total population of denver in 2020
den_total_2020 <- as.data.frame(matrix(ncol = 2, nrow = 717630))
names(den_total_2020) <- c("City", "DrugRelated")
den_total_2020$City <- rep(c("Total"), c(717630))
den_total_2020$DrugRelated <- rep(c("Yes"), c(717630))
# Join two data sets
den_2020_drugs <- full_join(denver_2020, den_total_2020)
den_2020_drugs$City <- factor(den_2020_drugs$City, levels =
names(sort(table(den_2020_drugs$City), decreasing = TRUE)))
# Create ggplot appear as stacked to give the image of percentage
ggplot(data = den_2020_drugs) + geom_bar(mapping = aes(x = DrugRelated, y =
(..count..)/sum(..count..), fill = City), position = "stack") +
  scale_y_continuous(labels = percent) +
  labs(title = "Denver 2020 Drug/Narcotic Relation", x = "Drug Related", fill = "Percentage") +
  theme_minimal()

# Create ggplot specifically for 2019 for Dallas
dallas_2019 <- dallas_drugs[which(dallas_drugs$ArrestYear %in% 2019),]
# Created a separate data set that shows the total population of Dallas in 2019
dal_total_2019 <- as.data.frame(matrix(ncol = 2, nrow = 2635603))
names(dal_total_2019) <- c("City", "DrugRelated")
dal_total_2019$City <- rep(c("Total"), c(2635603))
dal_total_2019$DrugRelated <- rep(c("Yes"), c(2635603))
# Join two data sets
dal_2019_drugs <- full_join(dallas_2019, dal_total_2019)
dal_2019_drugs$City <- factor(dal_2019_drugs$City, levels = names(sort(table(dal_2019_drugs$City),
decreasing = TRUE)))
# Create ggplot appear as stacked to give the image of percentage
ggplot(data = dal_2019_drugs) + geom_bar(mapping = aes(x = DrugRelated, y =
```

```
(..count..)/sum(..count..), fill = City), position = "stack") +
  scale_y_continuous(labels = percent) +
  labs(title = "Dallas 2019 Drug/Narcotic Relation", x = "Drug Related", fill = "Percentage") +
  theme_minimal()

# Create ggplot specifically for 2020 for Dallas
dallas_2020 <- dallas_drugs[which(dallas_drugs$ArrestYear %in% 2020),]
# Created a separate data set that shows the total population of Dallas in 2020
dal_total_2020 <- as.data.frame(matrix(ncol = 2, nrow = 2610957))
names(dal_total_2020) <- c("City", "DrugRelated")
dal_total_2020$City <- rep(c("Total"), c(2610957))
dal_total_2020$DrugRelated <- rep(c("Yes"), c(2610957))
# Join two data sets
dal_2020_drugs <- full_join(dallas_2020, dal_total_2020)
dal_2020_drugs$City <- factor(dal_2020_drugs$City, levels = names(sort(table(dal_2020_drugs$City),
decreasing = TRUE)))
# Create ggplot appear as stacked to give the image of percentage
ggplot(data = dal_2020_drugs) + geom_bar(mapping = aes(x = DrugRelated, y =
(..count..)/sum(..count..), fill = City), position = "stack") +
  scale_y_continuous(labels = percent) +
  labs(title = "Dallas 2020 Drug/Narcotic Relation", x = "Drug Related", fill = "Percentage") +
  theme_minimal()

# Calculate % difference of population involved with drugs between Dallas and Denver in 2019
pop_diff_2019 <- round(((length(which(denver_2019$DrugRelated == "Yes")) / 725508) -
(length(which(dallas_2019$DrugRelated == "Yes")) / 2635603))*100, 2)
cat("% difference in drug related crime between Denver and Dallas in 2019: ", pop_diff_2019, "%\n")

# Calculate % difference of population involved with drugs between Dallas and Denver in 2020
pop_diff_2020 <- round(((length(which(denver_2020$DrugRelated == "Yes")) / 717630) -
(length(which(dallas_2020$DrugRelated == "Yes")) / 2610957))*100, 2)
cat("% difference in drug related crime between Denver and Dallas in 2020: ", pop_diff_2020, "%\n")
```

```
# Question 3
dallas_2018 <- subset(dallas_data, dallas_data$ArrestYear == "2018")
dallas_2019 <- subset(dallas_data, dallas_data$ArrestYear == "2019")
dallas_2020 <- subset(dallas_data, dallas_data$ArrestYear == "2020")
dallas_2018_drug <- subset(dallas_2018, dallas_2018$DrugRelated == "Yes")
dallas_2019_drug <- subset(dallas_2019, dallas_2019$DrugRelated == "Yes")
```

```
dallas_2020_drug <- subset(dallas_2020, dallas_2020$DrugRelated == "Yes")

denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2018 <- subset(denver_data, ReportedDateStandard >= as.Date("2018-01-01") &
ReportedDateStandard <= as.Date("2018-12-31"))
denver_2018$ArrestYear <- format(denver_2018$ReportedDateStandard, "%Y")
denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2019 <- subset(denver_data, ReportedDateStandard >= as.Date("2019-01-01") &
ReportedDateStandard <= as.Date("2019-12-31"))
denver_2019$ArrestYear <- format(denver_2019$ReportedDateStandard, "%Y")
denver_data$ReportedDateStandard <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_2020 <- subset(denver_data, ReportedDateStandard >= as.Date("2020-01-01") &
ReportedDateStandard <= as.Date("2020-12-31"))
denver_2020$ArrestYear <- format(denver_2020$ReportedDateStandard, "%Y")
denver_2018_drug <- subset(denver_2018, denver_2018$DrugRelated == "drug-alcohol")
denver_2019_drug <- subset(denver_2019, denver_2019$DrugRelated == "drug-alcohol")
denver_2020_drug <- subset(denver_2020, denver_2020$DrugRelated == "drug-alcohol")

# Marijuana between 2 cities
dal_marijuana_rows_2018 <- grep("Marijuana", dallas_2018_drug$DrugType)
dal_marijuana_rows_2019 <- grep("Marijuana", dallas_2019_drug$DrugType)
dal_marijuana_rows_2020 <- grep("Marijuana", dallas_2020_drug$DrugType)
den_marijuana_rows_2018 <- grep("marijuana", denver_2018_drug$OffenseTypeId)
den_marijuana_rows_2019 <- grep("marijuana", denver_2019_drug$OffenseTypeId)
den_marijuana_rows_2020 <- grep("marijuana", denver_2020_drug$OffenseTypeId)
marijuana_df <- data.frame(c(length(dal_marijuana_rows_2018), length(dal_marijuana_rows_2019),
length(dal_marijuana_rows_2020), length(den_marijuana_rows_2018),
length(den_marijuana_rows_2019), length(den_marijuana_rows_2020)), c("Dallas", "Dallas", "Dallas",
"Denver", "Denver", "Denver"), c("2018", "2019", "2020", "2018", "2019", "2020"))
names(marijuana_df) <- c("Count", "City", "Year")
ggplot(data = marijuana_df, aes(x = Year, y = Count, fill = City)) + geom_bar(stat = "identity", position =
"dodge") + labs(title = "Marijuana Counts For Dallas and Denver")
```

```
# Question 4
# Filter data to only be for 2018 up
dallas_data <- subset(dallas_data, dallas_data$ArrestYear >= 2018)

# Count the number of times an arrest occurred at a zip code
dallas_zip <- dallas_data %>%
  count(dallas_data$ArrestZipCode)
```

```
names(dallas_zip) <- c("ZipCode", "Number")
# Merge the count of arrest to the shape file
merged_zip <- left_join(zip_shape, dallas_zip, by = "ZipCode")
# Create a ggplot that represents low amount of crime with blue and high
amounts of crime with red corresponding with the zip code
ggplot(data = merged_zip) +
  geom_sf(aes(fill = Number)) +
  scale_fill_gradient(low = "blue", high = "red") +
  ggtitle("Number of Crimes in Dallas")
# Count the number of drug related crimes in dallas and grouped each based
on zip code
dallas_zip_drugs <- dallas_data %>%
  group_by(dallas_data$ArrestZipCode, dallas_data$DrugRelated) %>%
  count(dallas_data$DrugRelated)
# Removed from the data frame rows that had "No" and "Unknown" in drug
related cases
indices <- which(dallas_zip_drugs$`dallas_data$DrugRelated` != "No")
dallas_zip_drugs <- dallas_zip_drugs[indices, ]
indices <- which(dallas_zip_drugs$`dallas_data$DrugRelated` != "Unknown")
dallas_zip_drugs <- dallas_zip_drugs[indices, ]
names(dallas_zip_drugs) <- c("ZipCode", "DrugRelated","Number")
# Merge the count of drug related cases with the shapefile
merged_zip <- left_join(zip_shape, dallas_zip_drugs, by = "ZipCode")
# Mapped the areas where drug related crimes occurred with the the
corresponding zip code
ggplot(data = merged_zip) +
  geom_sf(aes(fill = Number)) +
  scale_fill_gradient(low = "blue", high = "red") +
  ggtitle("Number of Drug Related Crimes in Dallas")

# Removed all of the values not pertaining to drugs
denver_drugs <- filter(denver_data, DrugRelated == "drug-alcohol")
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId !=
"liquor-possession"),]
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId !=
"liquor-sell"),]
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId !=
"liquor-manufacturing"),]
denver_drugs <- denver_drugs[which(denver_drugs$OffenseTypeId !=
"liquor-other-viol"),]
# Selected years 2018 and above
denver_drugs$ReportedDate <- format(as.Date(denver_drugs$ReportedDate,
format =  "%m/%d/%Y"), "%Y")
```

```r
denver_drugs <- subset(denver_drugs, denver_drugs$ReportedDate >= 2018)
# Select boundaries which shows Denver
denver_map <- denver_drugs %>%
  filter(
    between(denver_drugs$Geo_lon, -105.3809, -104.3275),
    between(denver_drugs$`Geo_lat`, 39.5718, 39.9571)
  )
# Map Denver and placed points where drug related crimes occured
qmplot(Geo_lon, `Geo_lat`, data = denver_map, maptype = "toner-lite", color
= factor(OffenseTypeId), main = "Denver Drug Cases")
```

```r
# Question 5
# For the entirety of the dallas dataset

# Overall version
age_breaks <- cut(dallas_data$ArresteeAge, breaks = seq(0, 100, by = 10))
age_df <- data.frame(table(age_breaks))
names(age_df) <- c("Breaks", "Count")

ggplot(data = dallas_data) +
  geom_bar(mapping = aes(x = ArresteeAge, fill = cut(ArresteeAge,
                         breaks = seq(0, 100, by = 10))), color = "black", position = "identity") +
  scale_x_continuous(name = "Age", breaks = seq(0, 100, by = 5)) +
  labs(title = "Overall Age Distribution", x = "Age", y = "Count", fill = "Age Group") +
  geom_vline(xintercept = mean(dallas_data$ArresteeAge), color = "red", size = 0.5, linetype =
"dashed") +
  geom_text(aes(x = mean(dallas_data$ArresteeAge), y = 100, label = "Mean Age"), color = "blue", size
= 3, vjust = -30)

# Drugs
dallas_drugs <- subset(dallas_data, dallas_data$DrugRelated == "Yes")

# Drugs version
d_age_breaks <- cut(dallas_drugs$ArresteeAge, breaks = seq(0, 100, by = 10))
d_age_df <- data.frame(table(d_age_breaks))
names(d_age_df) <- c("Breaks", "Count")

ggplot(data = dallas_drugs) +
  geom_bar(mapping = aes(x = ArresteeAge, fill = cut(ArresteeAge,
                         breaks = seq(0, 100, by = 10))), color = "black", position = "identity") +
```

```
  scale_x_continuous(name = "Age", breaks = seq(0, 100, by = 5)) +
  labs(title = "Drug Crime Age Distribution", x = "Age", y = "Count", fill = "Age Group") +
  geom_vline(xintercept = mean(dallas_data$ArresteeAge), color = "red", size = 0.5, linetype =
"dashed") +
  geom_text(aes(x = mean(dallas_data$ArresteeAge), y = 100, label = "Mean Age"), color = "blue", size
= 3, vjust = -20)

# Conclusion
# The age group 20 - 30 has the most crimes in Dallas consistently over the years
```

```
# Question 6
days_order <- c("Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday", "Saturday")
dallas_days_df <- data.frame(table(dallas_data$ArrestDay))
names(dallas_days_df) <- c("Days", "Counts")
# dallas_days_df$Labels <- factor(dallas_days_df$Days, levels = days_order)

# Set the color scale for the heatmap
color_scale <- scale_fill_gradient(low = "#ABCDEF", high = "#FF0000")

# Create a Dallas heatmap
ggplot(dallas_days_df, aes(x = "Weekday", y = Days, fill = Counts)) +
  geom_tile() + scale_fill_gradient(low = "lightblue", high = "darkred") +
  labs(title = "Dallas Heatmap of Days of Week vs Crime Counts") + theme_minimal()

denver_dates <- as.Date(denver_data$ReportedDate, format = "%m/%d/%Y")
denver_data$ArrestDay <- weekdays(denver_dates)

denver_days_df <- data.frame(table(denver_data$ArrestDay))
names(denver_days_df) <- c("Days", "Counts")

# Create a Denver heatmap
ggplot(denver_days_df, aes(x = "Weekday", y = Days, fill = Counts)) +
  geom_tile() + scale_fill_gradient(low = "lightblue", high = "darkred") +
  labs(title = "Denver Heatmap of Days of Week vs Crime Counts") + theme_minimal()

# Dallas Chi-Square Test
dallas_crime_table <- table(dallas_days_df$Days, dallas_days_df$Counts)
dallas_chi_result <- chisq.test(dallas_crime_table)
```

```
print(dallas_chi_result)
# P-value comes to 0.227 so since it is greater than 0.05, there is not enough evidence to disprove the null hypothesis (null hypothesis: There is no association between the crime count and days of week)
```

```
# Question 7
# Selected Dallas data that was 2018 and up, then selected only crimes involving weapons
dallas_weapon <- subset(dallas_data, dallas_data$ArrestYear >= 2018)
dallas_weapon <- subset(dallas_weapon, dallas_weapon$ArrestWeapon !=
"Unarmed")
dallas_weapon <- subset(dallas_weapon, dallas_weapon$ArrestWeapon !=
"None")
# Counted the number of crimes involving some form of weapon
zip_counts <- table(dallas_weapon$ArrestZipCode)
zip_counts <- sort(zip_counts, decreasing = TRUE)
zip_df <- data.frame(zip_counts)
names(zip_df) <- c("Zip", "Count")
# Choose the top 15 crimes involving a weapon
top_15_zip <- head(zip_df, 15)
dallas_weapon <- subset(dallas_weapon, ArrestZipCode %in% top_15_zip$Zip)
# Plotted the data set
ggplot(dallas_weapon) + geom_bar(mapping = aes(x = factor(ArrestZipCode),
fill = ArrestWeapon)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
# Find the spearman correlation
zip_df$Zip <- as.numeric(zip_df$Zip)
correlation <- cor(zip_df$Zip, zip_df$Count, method = "spearman")
print(correlation)
```

```
# Question 8
# Question: Which ethnic group has the highest crime related cases in Dallas?

# Group the data by ethnicity and calculate the total number of crime-related cases for each ethnicity

ethnicity_counts <- data.frame(table(dallas_data$ArresteeRace))
ethnicity_counts <- na.omit(ethnicity_counts)
names(ethnicity_counts) <- c("ArresteeRace", "Total")
ethnicity_counts <- ethnicity_counts[order(-ethnicity_counts$Total), ]
```

```
# Calculate percentage of each group
ethnicity_counts$Percent <- ethnicity_counts$Total / sum(ethnicity_counts$Total) * 100

# Create pie chart
ggplot(ethnicity_counts, aes(x = "", y = Percent, fill = ArresteeRace)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = " Pie Chart of Dallas Crimes Related by Race") +
  theme_void()

# Drug Related
drug_related <- subset(dallas_data, dallas_data$DrugRelated == "Yes")
ethnicity_counts <- data.frame(table(drug_related$ArresteeRace))
ethnicity_counts <- na.omit(ethnicity_counts)
names(ethnicity_counts) <- c("ArresteeRace", "Total")
ethnicity_counts <- ethnicity_counts[order(-ethnicity_counts$Total), ]

# Calculate percentage of each group
ethnicity_counts$Percent <- ethnicity_counts$Total / sum(ethnicity_counts$Total) * 100

# Create pie chart
ggplot(ethnicity_counts, aes(x = "", y = Percent, fill = ArresteeRace)) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y") +
  labs(title = " Pie Chart of Dallas Crimes Related by Race and Drugs") +
  theme_void()
```

```
# Question 9
dallas_armed <- subset(dallas_data, dallas_data$ArrestWeapon != "Unarmed" &
dallas_data$ArrestWeapon != "None")
dallas_handgun <- dallas_armed[grep("gun", dallas_armed$ArrestWeapon), ]
dallas_more_guns <- dallas_armed[grep("Gun", dallas_armed$ArrestWeapon), ]
dallas_guns <- subset(dallas_armed, dallas_armed$ArrestWeapon == "Other Firearm")
dallas_firearms <- rbind(dallas_handgun, dallas_guns, dallas_more_guns)
dallas_firearms$ArrestYear <- factor(dallas_firearms$ArrestYear)

dallas_total_crime <- sum(dallas_data$ArrestState == "Texas")
dallas_gun_crime <- sum(dallas_firearms$ArrestState == "Texas")

# Percent of Firearm Crime
```

```
total_percent <- (dallas_gun_crime / dallas_total_crime) * 100
print(total_percent)

# Association to Drugs
dallas_drug_firearms <- subset(dallas_firearms, dallas_firearms$DrugRelated == "Yes")
dallas_drug_firearms <- dallas_drug_firearms[dallas_drug_firearms$DrugType != "",]

# Filter for arrests with drugs and firearms
dallas_drug_firearms <- subset(dallas_firearms, dallas_firearms$DrugRelated == "Yes")

# Count the number of people who had drugs and guns, handguns, shotguns
drug_firearms_count <- sum(dallas_drug_firearms$ArrestState == "Texas")

# Print the results
cat("Number of people who had drugs and guns, handguns, or shotguns:", drug_firearms_count, "\n")

# Percent of Drug Firearm Crime
drug_percent <- (drug_firearms_count / dallas_total_crime) * 100
print(drug_percent)

dallas_total_drug_crime <- sum(dallas_drugs$ArrestState == "Texas")

# Percent of Drug Firearm Crime of all Drug Crimes
drug2_percent <- (drug_firearms_count / dallas_total_drug_crime ) * 100
print(drug2_percent)

dallas_firearms_zip <- subset(dallas_firearms, select = ArrestZipCode)
dallas_firearms_df <- data.frame(table(dallas_firearms_zip))
names(dallas_firearms_df) <- c("Zipcode", "Count")

ggplot(data = dallas_firearms, aes(x = ArrestYear, y = ArresteeAge)) +
  geom_boxplot(aes(fill = ArrestYear)) + labs(title = "Distribution of Ages and Arrest Year for Armed
Men", xlab = "Year", ylab = "Age")

drugged_firearms <- subset(dallas_firearms, dallas_firearms$DrugRelated == "Yes")

ggplot(data = drugged_firearms, aes(x = ArrestYear, y = ArresteeAge)) +
  geom_boxplot(aes(fill = ArrestYear)) + labs(title = "Distribution of Drugged Armed Criminals Ages and
Arrest Year", xlab = "Year", ylab = "Age")
```

```r
# Question 10
address_counts <- table(dallas_data$ArrestAddress)
address_counts <- sort(address_counts, decreasing = TRUE)
address_df <- data.frame(address_counts)
names(address_df) <- c("Address", "Count")

drug_dallas <- subset(dallas_data, dallas_data$DrugRelated == "Yes")
drug_addy_counts <- table(drug_dallas$ArrestAddress)
drug_addy_counts <- sort(drug_addy_counts, decreasing = TRUE)
drug_address_df <- data.frame(drug_addy_counts)
names(drug_address_df) <- c("Address", "Count")

top_10_addy <- head(address_df, 10)
top_10_drugAddy <- head(drug_address_df, 10)

ggplot(data = top_10_addy, aes(x = Address, y = Count, fill = Address)) + geom_bar(stat = "identity") +
theme(axis.text.x = element_text(angle = 45, hjust = 1)) + labs(title = "Top 10 Streets With Highest
Counts of General Crime") + geom_text(data = top_10_addy, aes(label = Count), size=3, vjust=1.5)

ggplot(data = top_10_drugAddy, aes(x = Address, y = Count, fill = Address)) + geom_bar(stat =
"identity") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + labs(title = "Top 10 Streets With
Highest Counts of Drug Crime") + geom_text(data = top_10_drugAddy, aes(label = Count), size=3,
vjust=1.5)

# Merge Dataframes for Displaying Top 10 Violent Streets with Drug Relations
top_10_merge_df <- merge(top_10_addy, drug_address_df, by = "Address", all.x = TRUE)
top_10_merge_df[is.na(top_10_merge_df)] <- 0
names(top_10_merge_df) <- c("Address", "Count", "DrugCount")

ggplot(data = top_10_merge_df, aes(x = Address, y = Count, fill = Address)) + geom_bar(stat =
"identity") + theme(axis.text.x = element_text(angle = 45, hjust = 1)) + labs(title = "Top 10 Streets With
Highest Counts of General Crime With Drug Related") + geom_text(data = top_10_merge_df, aes(label
= DrugCount), size=3, vjust=1.5)
```

```r
# Question 11
# Removing excess characters so the data will match the data from our Denver Data set
denver_income$Tract_Name <- gsub("(.*),.*", "\\1", denver_income$Tract_Name)
denver_income$Tract_Name <- gsub("(.*),.*", "\\1", denver_income$Tract_Name)

names(denver_income)
```

```r
# Created a Denver map based on tracts
st <- states()
plot(st$geometry)
co_counties <- counties("CO")
plot(co_counties$geometry)
den_tracts <- tracts("CO", "Denver")
plot(den_tracts$geometry)
colnames(denver_income)[4] <- "NAMELSAD"
denver_map_join <- den_tracts %>%
  left_join(denver_income, by = c("NAMELSAD" = "NAMELSAD"))
# Map the percentages of the population in Denver lower than the federal poverty level at each tract
ggplot(denver_map_join, aes(fill = Percent_Poverty_AllPeople_Income_Below_Pov_Level)) +
  geom_sf(color = "white", lwd = 0.2) +
  theme_void() +
  labs(fill  = "Income Below Federal Poverty Level")

# For Dallas
# Since there is multiple percentages per zip code, we  looked at the mean percentage at each zip
dallas_poverty <- dallas_poverty %>%
  group_by(Location) %>%
  summarize(avg_value = mean(Indicator.Rate.Value, na.rm = TRUE))

names(dallas_poverty) <- c("ZipCode", "Percent_Below_Fed_Poverty")
# Joined the shapefile for zip code and data set containing poverty percentages
merged_zip <- left_join(zip_shape, dallas_poverty, by = "ZipCode")
# Mapped the percentages below the federal poverty level in Dallas
ggplot(data = merged_zip) +
  geom_sf(aes(fill = Percent_Below_Fed_Poverty)) +
  scale_fill_gradient(low = "blue", high = "red")
```

References

Texas, H. N. (n.d.). *Healthy North Texas*. Healthy North Texas :: Indicators :: People Living Below Poverty Level :: County : Dallas. Retrieved April 23, 2023, from https://www.healthyntexas.org/indicators/index/view?indicatorId=347&localeId=2631

Dallaszipcodes 2018. City of Dallas GIS Services. (n.d.). Retrieved April 23, 2023, from https://gisservices-dallasgis.opendata.arcgis.com/datasets/DallasGIS::dallaszipcodes-2018/explore?filters=eyJTSEFQRWFyZWEiOlsxNDQxMTQuNjY5OTcyLDE4MTA0MDA2MTUuODNdLCJTSEFQRWxlbiI6WzE2OTcuNTcwNzc4ODcsMjgwMzI0Ljg4OTY0N10sIlppcENvZGUiOls3NTAzOSw3NTM5MF0sIlNoYXBlX19BcmVhIjpbMTg2NzQuNSwyMzc3MDMyMDFdLCJTaGFwZV9fTGVuZ3RoIjpbNjE2LjcxMjk3NzMxMjUxMywxMDExMjYuMDg2MjQ5OTM3XX0%3D&location=32.692336%2C-96.793771%2C12.97&showTable=true

Income/Poverty (census tracts). Colorado Department of Public Health and Environment. (n.d.). Retrieved April 23, 2023, from https://data-cdphe.opendata.arcgis.com/datasets/income-poverty-census-tracts/explore?location=38.926115%2C-105.550600%2C7.86&showTable=true

Sachs, D. (2021, March 22). *This shape explains Denver's past, present and likely its future*. Denverite. Retrieved April 23, 2023, from https://denverite.com/2018/12/21/denver-socioeconomic-map-shape/

Schiller, A. (2023). Denver, CO Crime Rates. *NeighborhoodScout*. https://www.neighborhoodscout.com/co/denver/crime

Schiller, A. (2023a). Dallas, TX Crime Rates. *NeighborhoodScout*. https://www.neighborhoodscout.com/tx/dallas/crime

United States Census Bureau QuickFacts. (n.d.). *U.S. Census Bureau QuickFacts: Dallas city, Texas*. Census Bureau QuickFacts. https://www.census.gov/quickfacts/dallascitytexas