Crime in Austin

Tessa Mitchell

April 21, 2021

Contents

1 Introduction

This project was created to discover statistics and trends about the crime in Austin, Texas. I used R (and packages such as ggplot2) to analyze the dataset. The goal of this project was to find interesting trends and statistics about crime in Austin by analyzing crime trends over different periods of time, and studying the locations of crimes in the city. I began with several basic questions, but began to study certain topics deeper as I made discoveries. In this report, I will detail the frequency of time over hours, days, and years. We will look at the census tracts that crime happen within the city, and the types of buildings in which crimes occur.

During the course of this report, we will examine patterns among different types of crime. Most of the definitions of these crimes are self-explanatory; however, the difference between theft, robbery, and burglary can be a bit confusing. Using the legal definitions for these crimes for Texas [**3**], I will define these terms below:

- Burglary: Without consent of the owner, entering a building or structure with the intent of committing a felony, theft or assault
- Robbery: During the course of committing a theft, intentionally, knowingly or recklessly causing bodily injury to another, or intentionally or knowingly threatening or placing another in fear of imminent bodily injury or death
- Theft: Stealing that doesn't fall under the category of Burglary or Robbery

In summary, burglary does not involve an act of violence, even if violence was initially intended. in most burglaries, the victim is not present. Robbery, on the other hand, does not require the unlawful entering of a building. Auto theft is also included in the categories in the data, indicating any Theft where a car is stolen.

The main dataset, titled "Austin Crime Dataset 2002 - 2019", comes from Kaggle.com. [**1**] Despite the misleading title, it contains crime data from the beginning of 2003 through part of 2019. It includes 2,124,418 instances of crime with 27 variables. These variables include the occurred date and time (*Occurred.Date* and *Occurred.Time* respectively, with an additional combined variable of *Occurred.Date.Time*), a description of the location (*Location.Type*), the type of crime (*Category.Description*), and the census tract where the crime occurred (*Census.Tract*).

The secondary dataset contains the population per census tract in Texas, as calculated by the American Community Survey. The data was retrieved from the IPUMS National Historical Geographic Information System. [**2**]

This report was created using LaTeX through Overleaf.com.

2 Data Cleaning

We began by reading the initial csv file retrieved from Kaggle into a data frame labeled *crimes*.

In the original csv file for the crime data, the *Location* variable consisted of latitude and longitude separated by a comma. The *read.table()* function in R cannot differentiate between this comma and the commas separating variables. Therefore, all of the column names were offset by one. The initial data cleaning involved adjusting the column names to correct this issue. While doing this, we also changed the *Category.Description* and *Location.Type* variables from character variables to factor variables.

Next I worked with the *Occurred.Date.Time* variable to put them in a more usable format. I created new variable *year* as a substring of *Occurred.Date.Time* representing the year of the instance. We created the *time* variable as a substring of the *Occurred.Date.Time* variable. This time was initially represented as "hour:minute:second" followed by "AM" or "PM". Using the *strptime()* function, we reformatted the time variable as "hour:minute" in 24 hour time. From this variable, I was able to create two new variables *hour* and *minute* as substrings of time.

In order to easily work with the dates, I converted *Occurred.Date* variable to Date objects using *as.Date()* and stored it in a new variable *as_date*. Next, by inputting *as_date* into the *weekdays()* function, I created a variable called *day_of_week*, indicating with day of the week a crime took place.

Next, I wanted to be able to compare the crime between certain days of the year. I created a new variable *as_date2000*, which is the same as *as_date* but with all years changed to 2000. This is used to account for leap years. For example, in a leap year, day 61 would be March 1, but in a non-leap year, day 61 would be March 2. By changing the years to 2000, I was able to avoid these errors. This allowed me to then convert the date object to a day of the year, with 1 representing January first and 366 representing December 31. This was accomplished using the *strftime()* function.

As I began to graph the data, I created several other data frames by filtering the original data frame. I also removed elements that contain values used as default values for certain variables. These data frames are detailed in their respective sections in the Analysis portions of the report. These smaller data frames are specific to their sections, and any data frames with the same names in other sections may not contain the same data.

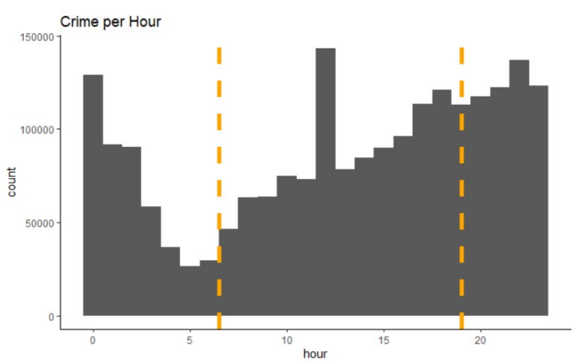
Next we will discuss the results of 6 different ways we analyzed the data. First we will look at crime over a period of time: by hour of the day, day of the week, day of the year, and over multiple years. Then we will study the location of crimes: first by looking at the type of location, such as a home or a store, and then by studying the census tract and area of the city in which the crimes happened.

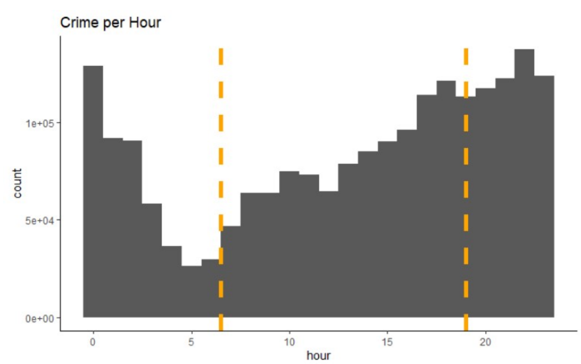
3 Analysis - Crime Over Time

In this section, we will analyze which hours and days crime is most likely to happen in Austin, and how crime has changed over the years. Some of these graphs will lead to the discover of meaningful trends, while others may not provide much insight. However, the lack of a pattern where a pattern was expected can also be a meaningful conclusion, so I decided to include these results in this report.

3.1 Crime per Hour of the Day

First, I made a simple histogram of crime per hour of the day. I noticed a large spike in crime at noon (below, on the left). By making another histogram (which is not included here) of crimes from 12pm to 1pm, I could clearly see an unrealistically large spike of crimes at exactly 12:00. This suggests that 12:00 is used as a default value when the exact time of a crime is unknown. Therefore, I made a second histogram which excluded crimes at 12:00 (below, on the right). However, this histogram may have its own inaccuracies as well, since we can never know the true value of crime at noon from this dataset. Therefore, I have included both graphs.





3.1.1 Do sunrise and sunset affect crime?

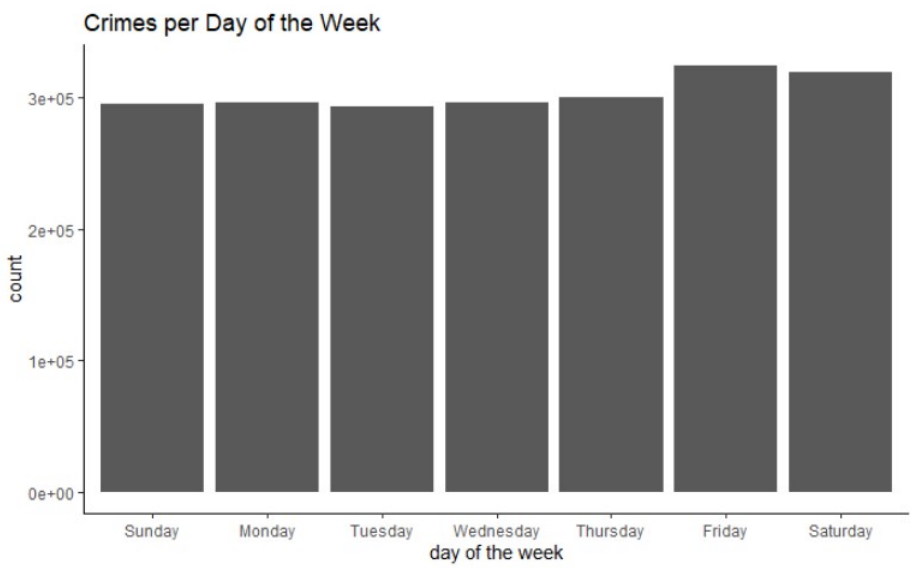
The vertical lines on the above graphs represent rough estimates for sunrise and sunset. Although I expected crime to be a lot higher after sunset, this appears to not be the case. Sunrise and sunset do not seem the cause quick increases or decreases in crime.

3.1.2 How does crime frequency change throughout the day? Which hours of the day are the least/most dangerous?

From the graph, we can see a clear pattern of crime frequency throughout the day. At 5am, crime is the lowest. This is likely because most people are asleep at this time of day. After 5am, crime gradually increases until 11pm, when it reaches its maximum, and then begins to decrease again until 5am the next day. We can see from the graph that the time of day has a large impact of the amount of crime; 11pm has over 5 times as much crime as 5am.

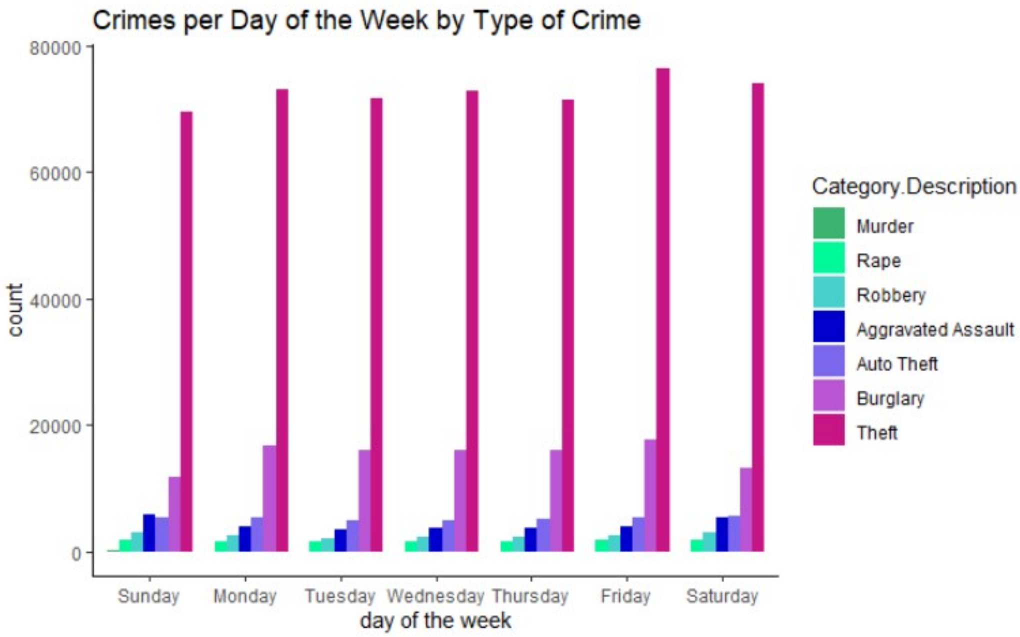
3.2 Crime per Day of the Week

Next, I wanted to examine the way that the day of the week affects crime. I began by making a bar chart showing the frequency of crime for each day of the week.



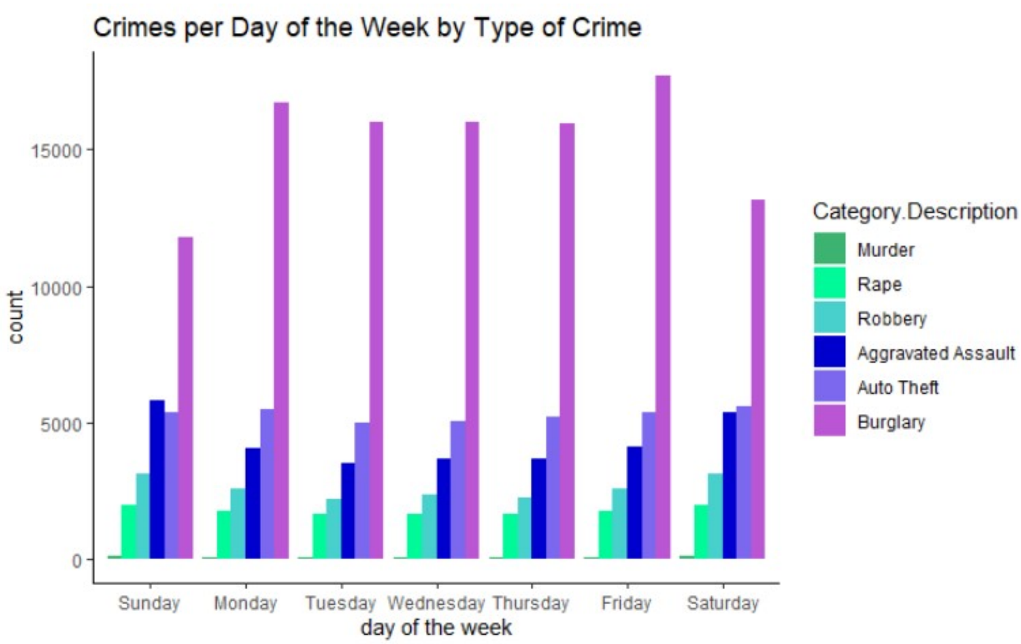
3.2.1 Is crime more likely to happen on the weekends?

In this bar chart, we see a small increase in crime on Friday and Saturday, but the change is small and these results are not too significant. However, in order to gain some more insight, I decided to split the bar chart by type of crime.



Theft

We can see in the bar chart below that theft follows a similar pattern to the original bar chart of total crime, with a slight increase on Friday and Saturday. This is expected, since, as shown by this chart, theft contributes to a very large portion of the total crime in Austin. Therefore, in order to get a better view of the other types of the crime, we will remove theft from the bar chart.



Burglary

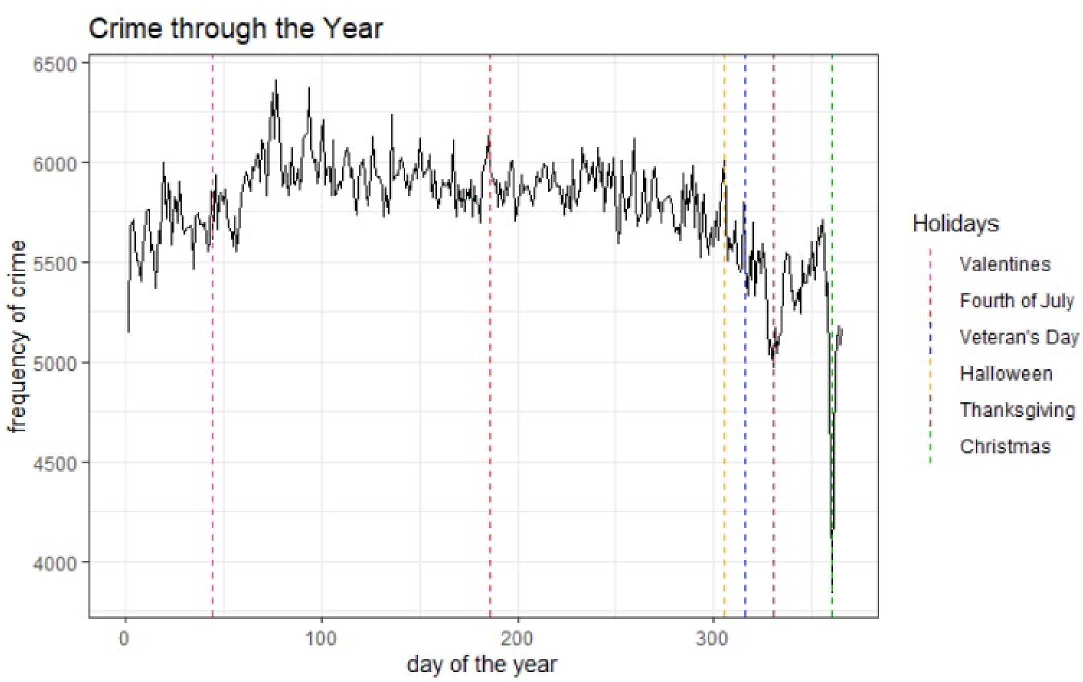
For burglary, we see an interesting and unique pattern, with a large decrease on the weekends, and a small increase on Monday and Friday. This is the only type of crime with a clear decrease in frequency on the weekends.

Aggravated Assault, Robbery, Rape, and Murder

For assault, robbery, rape, and murder - the 4 violent crimes in this list - we see a consistent pattern: a decrease in violent crimes during the middle of the week, and an increase on the weekend. One interesting observation is that aggravated assault surpasses auto theft only on Sundays.

3.3 Crime per Day of the Year

In order to see the pattern of crime throughout the year, I made a line graph of crime by day of the year. The original graph had large spikes at the first of each month, suggesting that this was used as a default value when the exact day of the crime was unknown. There was also a huge drop at the end of February (representing February 29, Leap Day). After removing the first day of each month and leap day from the data, I created the graph shown below.



3.3.1 How does crime change in a year? Doe some seasons have more crime than others?

Looking at the overall shape of the graph, we can see an increase in crime from the beginning of the year until March, when crime slowly decreases for the rest of the year. This shows that Spring and Summer seasons have more crime than Fall and Winter.

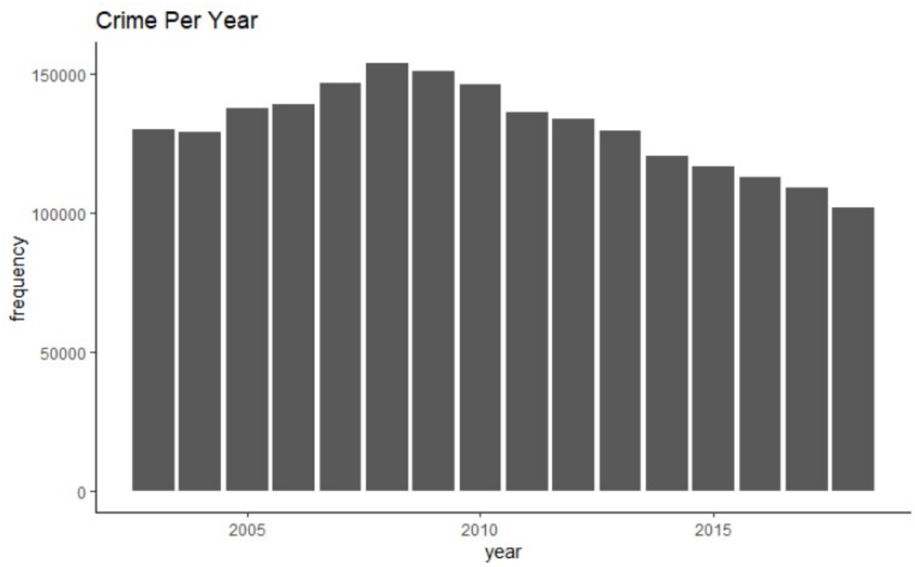
3.3.2 Is there a correlation between holidays and the frequency of crime?

After noticing a large drop in crime on one day towards the end of the year, I decided to start graphing holidays. (This day turned out to be Christmas) The graph above includes a vertical line for each of the 6 most commonly celebrated holidays in the US (For holidays such as Thanksgiving where the exact dates change from year to year, I picked the median day). We can see that not all holidays affect crime the same. During Christmas, there is a large drop in crime, and during the week of Thanksgiving there is also a drop in crime. Other holidays, however, such as Fourth of July and Thanksgiving, correlate with spikes in crime. (Valentines Day and Veterans' Day to not seem to large;y affect crime.) It is interesting to note that Christmas and Thanksgiving, which correlate to decreases in crime, are most likely to be spend with family; holidays such as Fourth of July and Halloween, which correlate with increases in crime, are most likely to be spend with friends. More research would need to be done to see if this is a causation or simply a correlation.

3.4 Crime from Year to Year

3.4.1 Has crime increased/decreased from 2003 to 2018?

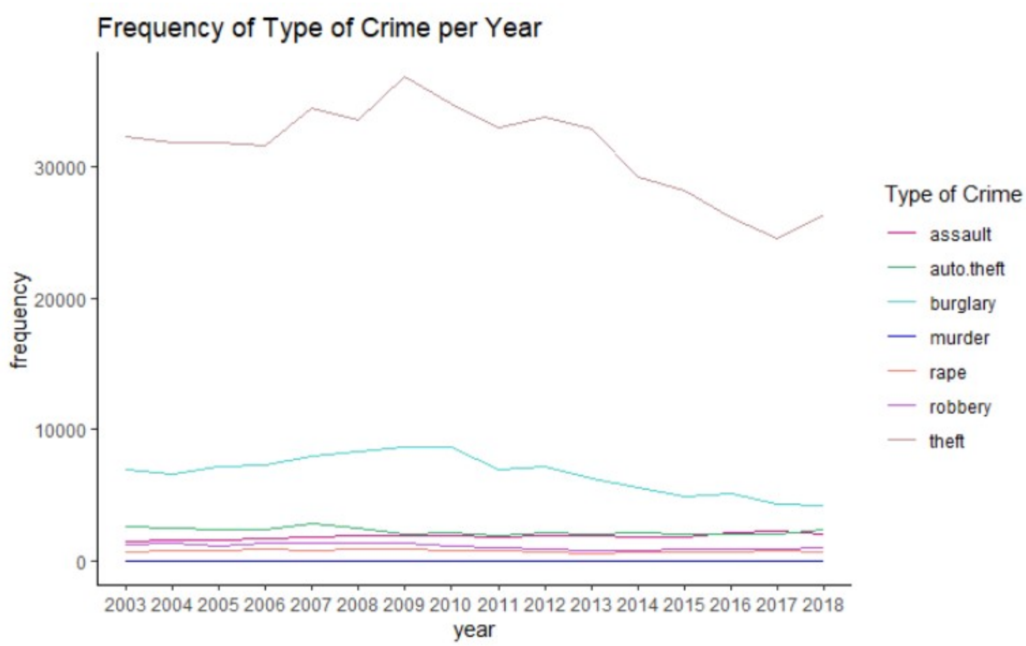
In order to examine crime from year to year, I created a bar graph of the crime from 2003 to 2018 (shown below). (Since the 2019 data only spanned the beginning part of the year, I excluded it from this graph.) From this plot, we can see that in general, crime increased from 2003 to 2008, and decreased from 2008 to 2018.

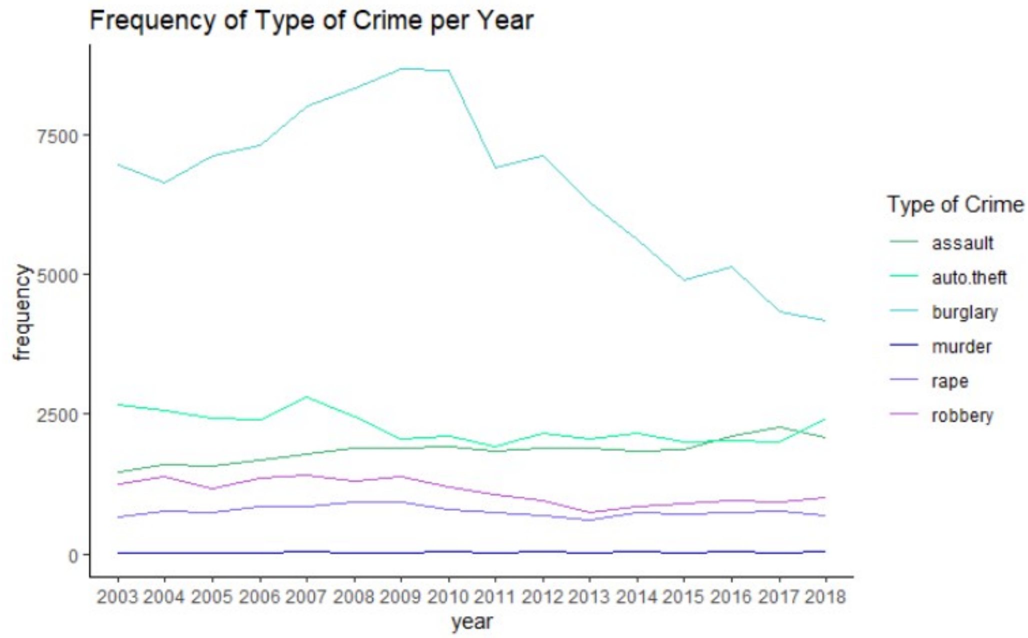


3.4.2 Which types of crime have contributed to the overall changes in crime over the year?

Theft

Next I decided to split up the graph by type of crime. In this first graph below, we can see that theft follows a similar trend to the graph of total crime. Since theft is clearly the most common crime, we can conclude that the decrease in theft has a large contribution.





Burglary, Auto Theft, and Robbery

In order to better view the trend of the other crimes, I recreated the same graph with theft removed. This graph shows that burglary, auto theft, and robbery have also decreased since 2008. These crimes are similar in nature to theft, so it makes sense that they would follow similar trends.

Rape, Murder, and Assault

Rape and murder have both stayed relatively constant over the years. Assault, however, does not seem to follow the pattern of any the other crimes; assault has continually increased from 2003 to 2017, and dropped slightly in 2018.

4 Analysis - Crime By Location

4.1 Crime per Type of Location

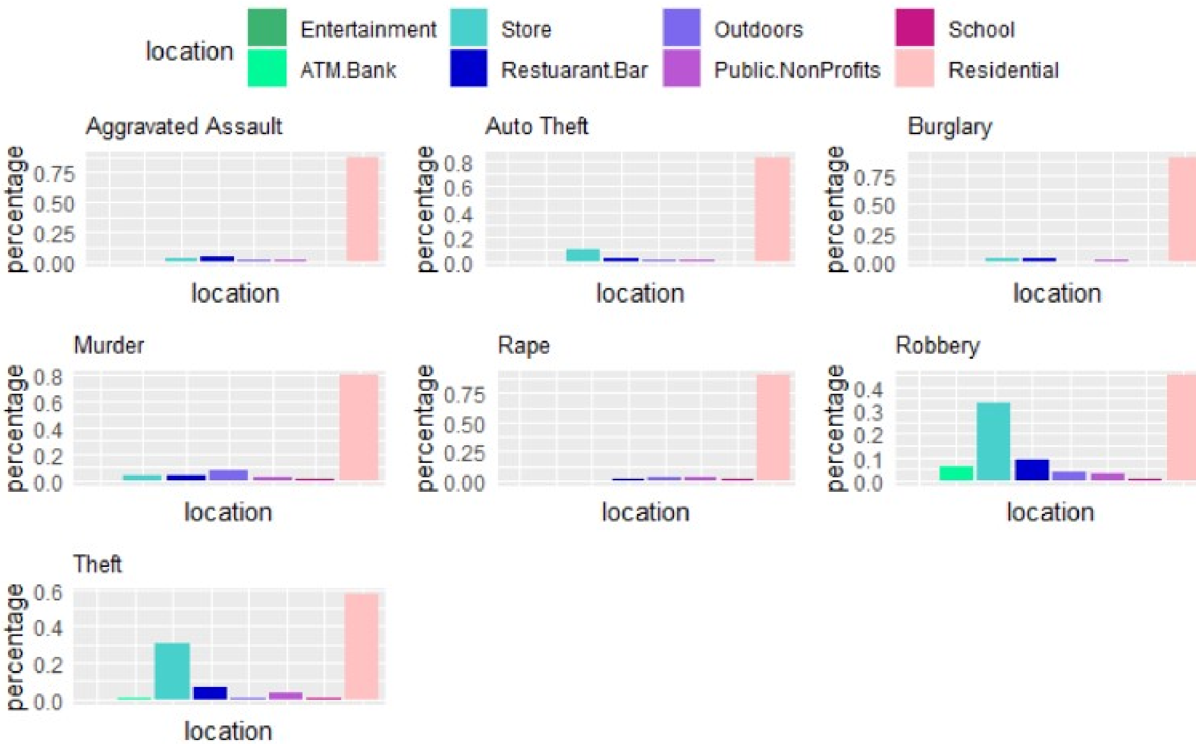
4.1.1 Which types of locations have the most crime?

Next I wanted to study the locations of crime. The original dataset contained a factor variable called *Location.Type* with 46 levels, with each level name containing a list of locations. In order to better analyze the data, I manually recategorized these locations into a new data frame. I grouped the locations as shown below, making sure to include all of the most common locations for crimes. (Some locations, such as tribal lands or military locations, had very small numbers of crime, so they were excluded).

Category	Locations
Entertainment	Amusement Park, Arena, Stadium, Fairgrounds Coliseum, Gambling Facility, Casino, Race Track
ATM / Bank	Bank, Savings and Loan, ATM Separate from Bank
Store	New or Used Car Dealership, Convenience Store, Department Store Discount Store, Grocery Store, Supermarket, Liquor Store Shopping Mall, Specialty Store*, Gas Station, Service Station
Restaurant / Bar	Restaurant, Bar, Night Club
Outdoors	Camp, Campground, Farm Facility, Field, Woods Lake Waterway, Park, Playground, Rest Area, Park Dock, Wharf, Freight, Modal Terminal
Public / NonProfits	Community Center, Drug Store, Doctor's Office, Hospital Government Building, Public Building, Church, Synagogue Temple, Mosque, Shelter Mission, Homeless Shelter
School	College, University, Elementary School Secondary School, Daycare Facility
Residential	Residence, Home, Hotel, Motel

*Specialty stores include niche stores such as a TV store, a mattress store, etc.

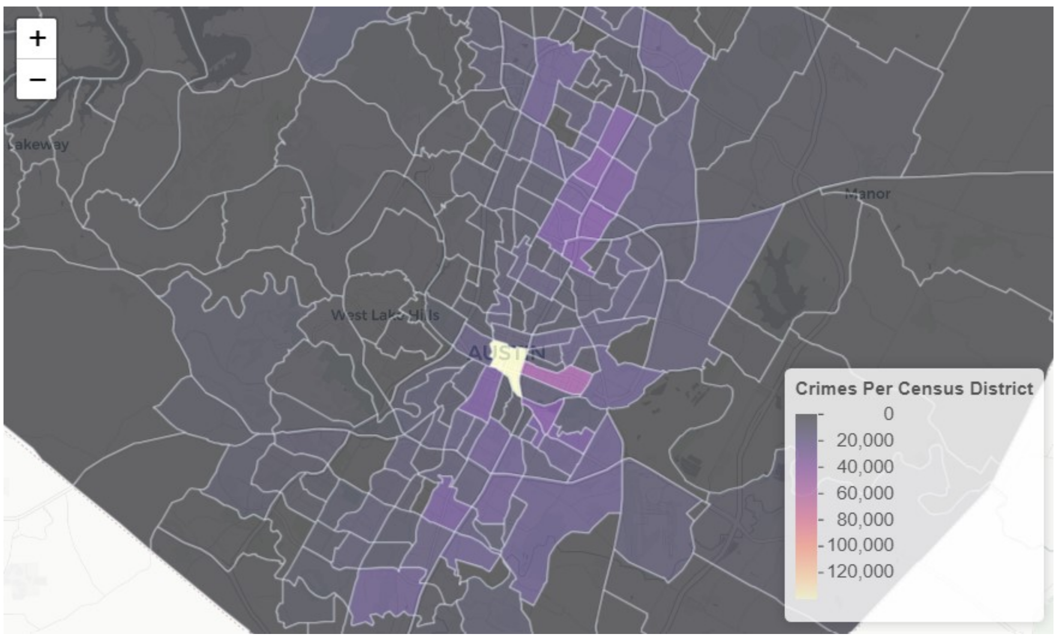
In order to analyze the locations of crimes, I used this new categorization to create a faceted bar plot representing the percentage of occurrences at each location per type of crime, as shown below.



It is clear that for all types of crime, the most common location is in a residential area. This is surprising, since we as a society normally consider our homes to be one of the safest places we can be. The crimes with the lowest percentages of residential locations are robbery and theft, which also commonly happen at stores, restaurants, and bars.

4.2 Map of Crime in Austin

Census tracts are small subdivisions of counties created by the United States Census Bureau. The City of Austin chooses to categorize their crimes by census district, so this is the variable that was included in the dataset. Therefore, the following maps also utilize the census tract system. The shape data for the census tracts was retrieved from the Census Data using an API key (which can be acquired for free online). The key was then entered into R using the *api.key.install()* function, and the shapes were retrieved using the *tracts()* function.



4.2.1 Do urban areas have more crime than rural areas?

The first map (above) shows the frequency of crime in each census district. The yellow maximum in the center of the map represents the center of downtown Austin. It is clear that the urban areas have a lot more crime than the rural areas. However, this may possibly be caused by the fact that the urban areas are a lot more densely populated. Therefore, in order to further investigate this result, I wanted to examine the crime per person for each census district, using the secondary dataset from the American Community Survey.

4.2.2 Is the difference in crime between urban areas and rural areas caused solely by population differences?

The second map (above) represents the crime per person - calculated by dividing the total occurrences of crime in a given census district by the number of residents in the given census district. This graph shows that, even when accounting for population differences, the urban areas still have more crime than the rural areas. We can also note the yellow maximum just north-east of downtown Austin. This area is mostly a commercial district, with few houses. It contains the intersection of two major highways, which may contribute to the large amount of crime.

5 Conclusion

Our goal was to analyze the trends and patterns of crime in Austin, Texas using our dataset from Kaggle. We looked at the data in several different ways to gain insight about our data; first by the time and date of the crime, and then by the location.

While looking at the time of day, we found that crimes are most likely to happen around midnight and least likely to happen at 5am. When analyzing the days of the week, we discovered that most types of crime are slightly more likely to happen on weekends compared to weekdays. We also discovered that crime is much less likely to happen on some holidays such as Thanksgiving and Christmas. We noted a decrease in crime since 2008, with the exception of aggravated assault, which has increased over this time period.

While studying the type of location of crimes, we noted that all crimes are most likely to happen in a residential area, such as a home or an apartment. We also realized that crime is more likely to happen in urban areas than rural areas, even after accounting for the differences in population.

Despite the interesting discoveries described here, there is still much more that could be done with this data. In the future, we could use other statistics from the American Community Survey - the same survey that gave us the population for each census district. We could look for correlations between the average age, average income, or employment rate of a census district and the frequency of crime in that census district. We could also compare the dates of crime to other variables based on date, such as weather. This type of comparison could be used to answer questions such as "Is crime less likely to happen on rainy days?"

References

[1] TS. (2020, March). Austin Crime Dataset 2002-2019, Version 2. Retrieved March 29, 2020 from https://www.kaggle.com/tsaustin/austin-crime-dataset-16-years-of-data

[2] Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 15.0, nghis0001.ds24420195.2019.tract. Minneapolis, MN: IPUMS. 2020. http://doi.org/10.18128/D050.V15.0

[3] Varghese, B. (2020, December 02). The difference between burglary and robbery in Texas. Retrieved April 18, 2021, from https://versustexas.com/blog/difference-between-burglary-and-robbery /: :text=The%20law%3A%20Under%20Texas%20Penal,a%20felony%2C%20theft %20or%20assault.amp;text=Unlike%20robbery%2C%20burglary%20does%20not, or%20an%20act%20of%20violence.

6 Code

The code is sorted in the same order as the report, with comments mirroring the subtitles of each section.

```
---
title: "Project v4"
author: "Tessa Mitchell"
date: "4/16/2021"
output: html_document
---

# Reading the Data

```{r}
crimes <- read.table("AustinCrimes.csv", na.strings = "NA", sep = ",",
                     header = TRUE, quote = "", fill = TRUE, row.names = NULL)
```

Cleaning the Data

```{r}
# Shifting all the columns by 1 (due to the error from the
# comma in the location variable)
crimes$Location <- paste(crimes$Longitude, crimes$Location, sp = ", ")
crimes$Longitude <- crimes$Latitude
crimes$Latitude <- crimes$Y.coordinate
crimes$Y.coordinate <- crimes$X.coordinate
crimes$X.coordinate <- crimes$Category.Description
crimes$Category.Description <- factor(crimes$UCR.Category)
crimes$UCR.Category <- crimes$Clearance.Date
crimes$Clearance.Date <- crimes$Clearance.Status
crimes$Clearance.Status <- crimes$Census.Tract
crimes$Census.Tract <- crimes$PRA
crimes$PRA <- crimes$APD.District
crimes$APD.District <- crimes$APD.Sector
crimes$APD.Sector <- crimes$Council.District
crimes$Council.District <- crimes$Zip.Code
crimes$Zip.Code <- crimes$Address
crimes$Address <- crimes$Location.Type
crimes$Location.Type <- factor(crimes$Report.Time)
crimes$Report.Time <- crimes$Report.Date
```
```

```
crimes$Report.Date <- crimes$Report.Date.Time
crimes$Report.Date.Time <- crimes$Occurred.Time
crimes$Occurred.Time <- crimes$Occurred.Date
crimes$Occurred.Date <- crimes$Occurred.Date.Time
crimes$Occurred.Date.Time <- crimes$Family.Violence
crimes$Family.Violence <-crimes$Highest.Offense.Code
crimes$Highest.Offense.Description <- crimes$Incident.Number
crimes$Incident.Number <- crimes$row.names
```

```{r}
# creating variable for year
crimes$year <- as.numeric(substring(crimes$Occurred.Date,
                      unlist(gregexpr("/", crimes$Occurred.Date.Time))[2] + 1))

# creating variable for time
crimes$time <- paste(substring(crimes$Occurred.Date.Time,
                      unlist(gregexpr(" ", crimes$Occurred.Date.Time))[1] + 1,
                      unlist(gregexpr(":", crimes$Occurred.Date.Time))[2] - 1),
                    substring(crimes$Occurred.Date.Time,
                     unlist(gregexpr(" ", crimes$Occurred.Date.Time))[2] + 1),
                    sep = "")

crimes$time <- strptime(crimes$time, format = "%I:%M%p")

crimes$time <- substring(crimes$time, unlist(gregexpr(" ", crimes$time))[1] + 1,
                         unlist(gregexpr(":", crimes$time))[2] - 1)

# creating variable for hour
crimes$hour <- as.numeric(substring(crimes$time, 1,
                                unlist(gregexpr(":", crimes$time))[1] - 1))

# creating variable for minute
crimes$minute <- as.numeric(substring(crimes$time,
                                unlist(gregexpr(":", crimes$time))[2] + 1))

# creating variable for date
crimes$as_date <- as.Date(paste(substring(crimes$Occurred.Date, 1,
                                unlist(gregexpr("/", crimes$Occurred.Date))[2]),
                                crimes$year, sep = ""), "%m/%d/%Y")

# creating variable for date with year 2000 (used to make day_of_year)
crimes$as_date2000 <- as.Date(paste(substring(crimes$Occurred.Date, 1,
```

~~~~
                                unlist(gregexpr("/", crimes$Occurred.Date))[2]),
                                "2000", sep = ""), "%m/%d/%Y")

crimes$day_of_yr <- as.numeric(strftime(crimes$as_date2000, format = "%j"))

# creating variable for day of the week and reordering days in chronological
# order
crimes$day_of_week <- factor(weekdays(crimes$as_date))

crimes$day_of_week <- factor(crimes$day_of_week,
                             levels = c("Sunday", "Monday", "Tuesday",
                                 "Wednesday", "Thursday", "Friday", "Saturday"))
```

# Crimes per Hour of the Day

```{r}
library(ggplot2) # for graphing
library(dplyr) # for filter()

# creating histogram of crime per hour
ggplot(crimes, aes(x = hour)) + geom_histogram(bins = 24) +
  geom_vline(xintercept = 6.5, linetype = "dashed", color = "orange", size = 2) +
  geom_vline(xintercept = 19, linetype = "dashed", color = "orange", size = 2) +
  ggtitle("Crime per Hour") + theme_classic()

# filtering out crimes with default time data of noon
crimes_without_noon <- filter(crimes, !(crimes$hour == 12 & crimes$minute == 0))

# creating new histogram of crime per hour without noon data
ggplot(crimes_without_noon, aes(x = hour)) +
  geom_histogram(bins = 24) +
  geom_vline(xintercept = 6.5, linetype = "dashed", color = "orange", size = 2) +
  geom_vline(xintercept = 19, linetype = "dashed", color = "orange", size = 2) +
  ggtitle("Crime per Hour") + theme_classic()

# filtering out crimes without a category
crimes_with_category <- filter(crimes_without_noon, Category.Description != "")

# creating stacked bar plot of crime per hour - this graph was not used in the
# final report
ggplot(crimes_with_category, aes(x = hour)) +
  geom_bar(aes(y = ..count.., fill = Category.Description)) +
~~~~

```
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred")) +
  ggtitle("Crime per Hour by Type of Crime") + theme_classic()
```

Crimes per Day of the Week

```{r}
# creating bar plot of crime per day of the week
ggplot(crimes, aes(x = day_of_week)) + geom_bar(aes(y = ..count..)) +
  xlab("day of the week") + theme_classic() +
  ggtitle("Crimes per Day of the Week")

# filtering out crimes without a category
crimes_with_category <- filter(crimes, Category.Description != "")

# sorting crimes in decreasing order
crimes_with_category$Category.Description <-
  factor(crimes_with_category$Category.Description,
         levels = c("Murder", "Rape", "Robbery", "Aggravated Assault",
                    "Auto Theft", "Burglary", "Theft"))

# creating grouped bar plot of crime per day of the week separated by type
ggplot(crimes_with_category, aes(x = day_of_week)) +
  geom_bar(aes(y = ..count.., fill = Category.Description),
           position = "dodge") +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred")) +
  xlab("day of the week") + theme_classic() +
  ggtitle("Crimes per Day of the Week by Type of Crime")

# filtering out craft
crimes_without_theft <- filter(crimes_with_category,
                               Category.Description != "Theft")

# creating grouped bar plot of crime per day of the week seperated by type
# without theft
ggplot(crimes_without_theft, aes(x = day_of_week)) +
  geom_bar(aes(y = ..count.., fill = Category.Description),
```

```
          position = "dodge") +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid")) +
  xlab("day of the week") + theme_classic() +
  ggtitle("Crimes per Day of the Week by Type of Crime")
```

```
# Crimes per Day of the Year

```{r}
# creating a frequency table for the data
crimes_per_day <- data.frame(table(as.numeric(crimes$day_of_yr)))
names(crimes_per_day) <- c("date", "frequency")
crimes_per_day$date <- as.numeric(crimes_per_day$date)
crimes_per_day <- filter(crimes_per_day, date != 1)

# creating line plot of frequency of crime through the year
ggplot(crimes_per_day, aes(x = as.numeric(date), y = frequency, group = 1)) +
  geom_line() + xlim(0, 365) + ggtitle("Crime through the Year") +
  xlab("day of the year") + ylab("frequency of crime") + theme_bw()

# filtering out default data of first of each month and leap day
crimes_per_day <- filter(crimes_per_day, date != 1 & date != 32 & date != 60 &
                           date != 61 & date != 92 & date != 122 & date != 153 &
                           date != 183 & date != 214 & date != 245 &
                           date != 275 & date != 306 & date != 336)

# creating line plot of frequency of crime through the year without default
# value or leap day
ggplot(crimes_per_day, aes(x = as.numeric(date), y = frequency, group = 1)) +
  geom_line() + xlim(0, 365) +
  geom_vline(aes(xintercept = 360, color = "green3"), linetype = "dashed") +
  geom_vline(aes(xintercept = 305, color = "orange1"), linetype = "dashed") +
  geom_vline(aes(xintercept = 186, color = "red"), linetype = "dashed") +
  geom_vline(aes(xintercept = 045, color = "violetred1"), linetype = "dashed") +
  geom_vline(aes(xintercept = 0316, color = "blue"), linetype = "dashed") +
  geom_vline(aes(xintercept = 330, color = "brown"), linetype = "dashed") +
  scale_color_identity(name = "Holidays", breaks = c("violetred1", "red",
                                                     "blue", "orange1", "brown",
                                                     "green3"),
                       labels = c("Valentines", "Fourth of July",
                                  "Veteran's Day", "Halloween", "Thanksgiving",
```
```

```
                                "Christmas"), guide = "legend") +
  ggtitle("Crime through the Year") + xlab("day of the year") +
  ylab("frequency of crime") + theme_bw()
```

Crime from Year to Year

```{r}
# filtering out 2019 data
crimes_without_2019 <- filter(crimes, crimes$year != 2019)
ggplot(crimes_without_2019, aes(x = year)) + geom_bar() +
  ylab("frequency") + ggtitle("Crime Per Year") + theme_classic()

# creating datasets for each type of crime
assault_by_year <- filter(crimes_without_2019,
                          Category.Description == "Aggravated Assault")
auto_by_year <- filter(crimes_without_2019,
                       Category.Description == "Auto Theft")
burglary_by_year <- filter(crimes_without_2019,
                           Category.Description == "Burglary")
murder_by_year <- filter(crimes_without_2019,
                         Category.Description == "Murder")
rape_by_year <- filter(crimes_without_2019,
                       Category.Description == "Rape")
robbery_by_year <- filter(crimes_without_2019,
                          Category.Description == "Robbery")
theft_by_year <- filter(crimes_without_2019,
                        Category.Description == "Theft")

# creating frequency variables for each type of crime
crimes_per_year <- data.frame(table(assault_by_year$year))
names(crimes_per_year) <- c("year", "assault.frequency")
crimes_per_year$assault.frequency <- crimes_per_year$assault.frequency

auto_frequency <- data.frame(table(auto_by_year$year))
names(auto_frequency) <- c("year", "auto.frequency")
crimes_per_year$auto.frequency <- auto_frequency$auto.frequency

burglary_frequency <- data.frame(table(burglary_by_year$year))
names(burglary_frequency) <- c("year", "burglary.frequency")
crimes_per_year$burglary.frequency <- burglary_frequency$burglary.frequency

murder_frequency <- data.frame(table(murder_by_year$year))
```

```
names(murder_frequency) <- c("year", "murder.frequency")
crimes_per_year$murder.frequency <- murder_frequency$murder.frequency

rape_frequency <- data.frame(table(rape_by_year$year))
names(rape_frequency) <- c("year", "rape.frequency")
crimes_per_year$rape.frequency <- rape_frequency$rape.frequency

robbery_frequency <- data.frame(table(robbery_by_year$year))
names(robbery_frequency) <- c("year", "robbery.frequency")
crimes_per_year$robbery.frequency <- robbery_frequency$robbery.frequency

theft_frequency <- data.frame(table(theft_by_year$year))
names(theft_frequency) <- c("year", "theft.frequency")
crimes_per_year$theft.frequency <- theft_frequency$theft.frequency

# creating line plot for each type of crime
ggplot(crimes_per_year, aes(x = year, group = 1)) +
  geom_line(aes(y = assault.frequency, color = "assault")) +
  geom_line(aes(y = auto.frequency, color = "auto.theft")) +
  geom_line(aes(y = burglary.frequency, color = "burglary")) +
  geom_line(aes(y = murder.frequency, color = "murder")) +
  geom_line(aes(y = rape.frequency, color = "rape")) +
  geom_line(aes(y = robbery.frequency, color = "robbery")) +
  geom_line(aes(y = theft.frequency , color = "theft")) +
  scale_color_manual("Type of Crime", values = c(assault = "mediumvioletred",
                                                 auto.theft = "mediumseagreen",
                                                 burglary = "mediumturquoise",
                                                 murder = "mediumblue",
                                                 rape = "salmon",
                                                 robbery = "mediumorchid",
                                                 theft = "rosybrown")) +
  ylab("frequency") + ggtitle("Frequency of Type of Crime per Year") +
  theme_classic()

# creating line plot for each type of crime exclusing theft
ggplot(crimes_per_year, aes(x = year, group = 1)) +
  geom_line(aes(y = assault.frequency, color = "assault")) +
  geom_line(aes(y = auto.frequency, color = "auto.theft")) +
  geom_line(aes(y = burglary.frequency, color = "burglary")) +
  geom_line(aes(y = murder.frequency, color = "murder")) +
  geom_line(aes(y = rape.frequency, color = "rape")) +
  geom_line(aes(y = robbery.frequency, color = "robbery")) +
  scale_color_manual("Type of Crime", values = c(assault = "mediumvioletred",
```

```
                                                    auto.theft = "mediumseagreen",
                                                    burglary = "mediumturquoise",
                                                    murder = "mediumblue",
                                                    ape = "salmon",
                                                    robbery = "mediumorchid")) +
  ylab("frequency") + ggtitle("Frequency of Type of Crime per Year") +
  theme_classic()
```

Crimes per Type of Location

```{r}
library(forcats)
library("ggpubr")

# renaming location levels to simpler names to work with
levels(crimes$Location.Type) <- c("", "Abandoned Structure", "Amusement Park",
                                  "Arena", "ATM", "Car Dealership", "Bank",
                                  "Bar", "Camp", "Church",
                                  "Commercial Building", "Community Center",
                                  "Construction Site", "Convenience Store",
                                  "Daycare", "Department Store", "Dock",
                                  "Hospital", "Farm", "Field", "Casino",
                                  "Gas Station", "Public Building",
                                  "Grocery Store", "Hotel", "Industrial Site",
                                  "Jail", "Lake", "Liquor Store", "Military",
                                  "Other", "Park", "Parking Lot", "Storage",
                                  "Home", "Rest Area", "Restuarant", "College",
                                  "School", "Alternative Schools",
                                  "Homeless Shelter", "Mall", "Specialty Store",
                                  "Road", "Transportation", "Tribal Lands")

# re-categorizing locations
crimes$location <- fct_collapse(crimes$Location.Type,
                                Store = c("Car Dealership", "Convenience Store",
                                          "Department Store", "Grocery Store",
                                          "Liquor Store", "Mall",
                                          "Specialty Store", "Gas Station"),
                                Restuarant.Bar = c("Restuarant", "Bar"),
                                School = c("College", "School", "Daycare",
                                           "Alternative Schools"),
                                Outdoors = c("Camp", "Farm", "Field", "Lake",
                                             "Park", "Rest Area", "Dock"),
```

```
                                Transportation = c("Parking Lot", "Road",
                                                   "Transportation"),
                                Entertainment = c("Amusement Park", "Arena",
                                                  "Casino"),
                                ATM.Bank = c("ATM", "Bank"),
                                Public.NonProfits = c("Community Center",
                                                      "Hospital",
                                                      "Public Building",
                                                      "Church",
                                                      "Homeless Shelter"),
                                Residential = c("Home", "Hotel"),
                                other_level = NULL)

# filtering locations to include only most common locations
crimes_collapsed_location <- filter(crimes, location == "Entertainment" |
                                      location == "ATM.Bank" |
                                      location == "Store" |
                                      location == "Restuarant.Bar" |
                                      location == "Outdoors" |
                                      location == "Public.NonProfits" |
                                      location == "School" |
                                      location == "Residential")

# create a data frame for each type of crime
assault <- filter(crimes_collapsed_location,
                  Category.Description == "Aggravated Assault")
auto <- filter(crimes_collapsed_location,
               Category.Description == "Auto Theft")
burglary <- filter(crimes_collapsed_location,
                   Category.Description == "Burglary")
murder <- filter(crimes_collapsed_location,
                 Category.Description == "Murder")
rape <- filter(crimes_collapsed_location,
               Category.Description == "Rape")
robbery <- filter(crimes_collapsed_location,
                  Category.Description == "Robbery")
theft <- filter(crimes_collapsed_location,
                Category.Description == "Theft")

# creating a bar plot for each type of crime
assault_graph <- ggplot(assault, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
```

```
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Aggravated Assault") + xlab("location") + ylab("percentage")

auto_graph <- ggplot(auto, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Auto Theft") + xlab("location") + ylab("percentage")

burglary_graph <- ggplot(burglary, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Burglary") + xlab("location") + ylab("percentage")

murder_graph <- ggplot(murder, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumspringgreen", "mediumturquoise",
                               "mediumblue", "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Murder") + xlab("location") + ylab("percentage")
```

```
rape_graph <- ggplot(rape, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Rape") + xlab("location") + ylab("percentage")

robbery_graph <- ggplot(robbery, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Robbery") + xlab("location") + ylab("percentage")

theft_graph <- ggplot(theft, aes(x = location, fill = location)) +
  geom_bar(aes(y = (..count..)/sum(..count..))) +
  scale_fill_manual(values = c("mediumseagreen", "mediumspringgreen",
                               "mediumturquoise", "mediumblue",
                               "mediumslateblue", "mediumorchid",
                               "mediumvioletred", "rosybrown1")) +
  guides(location = FALSE) +
  theme(legend.position = "none",
        plot.title = element_text(size = 10, face = "plain"),
        axis.text.x = element_blank(), axis.ticks = element_blank()) +
  ggtitle("Theft") + xlab("location") + ylab("percentage")

# combining the graphs into one graphic with a combined legend
ggarrange(assault_graph, auto_graph, burglary_graph, murder_graph, rape_graph,
          robbery_graph, theft_graph, ncol = 3, nrow = 3, common.legend = TRUE)
```

Map of Crime in Austin

```
```{r}
# importing necessary libraries
library(tigris)
library(acs)
library(leaflet)

# getting shape data from API key
api.key.install(key="6cc6b5c19367cd65d921e1886ec7ea752d0b9025")
travis <- tracts("TX", "Travis", cb = TRUE)
geo <- geo.make(state = c("TX"),
              county = c("Travis"), tract="*")

# creating frequency table for crimes per census district
tract_crimes <- as.data.frame(table(crimes$Census.Tract))
names(tract_crimes) <- c("tract", "frequency")

# joining shape and crime data
tract_data <- geo_join(travis, tract_crimes, "NAME", "tract")
tract_data[is.na(tract_data)] <- 0

# formatting shape data
popup <- paste0("GEOID: ", tract_data$NAME, "<br>", "Crimes ",
                tract_data$frequency)

# creating color palette
pal <- colorNumeric(
  palette = "magma",
  domain = tract_data$frequency
)

# creating map
map <-leaflet() %>%
  addProviderTiles("CartoDB.Positron") %>%
  addPolygons(data = tract_data,
              fillColor = ~pal(frequency),
              color = "aliceblue",
              fillOpacity = 0.6,
              weight = 1,
              smoothFactor = 0.3,
              popup = popup) %>%
  addLegend(pal = pal,
            values = tract_data$frequency,
```
```

```
            position = "bottomright",
            title = "Crimes Per Census District",)

map
```

```{r}
# importing necessary libraries
library("sf")

# reading population data file
population <- read.table(file = "census_tracts_2019.csv", sep = ",",
                         header = TRUE)

# formatting population data
population$TRACTA <- as.character(as.numeric(population$TRACTA) / 100)
population <- filter(population, population$COUNTY == "Travis County")
merged_data <- merge(tract_crimes, population, by.x = "tract", by.y = "TRACTA")
merged_data$population <- merged_data$ALUBE001
merged_data$crimes_per_person <- round(as.numeric(merged_data$frequency /
                                                    merged_data$population),
                                       digits = 0)

# joining shape and crime data
tract_data2 <- geo_join(travis, merged_data, "NAME", "tract")
tract_data2[is.na(tract_data2)] <- 0
tract_data2$crimes_per_person[which(
  !is.finite(tract_data2$crimes_per_person))] <- 0

# formatting shape data
popup <- paste0("GEOID: ", tract_data2$NAME, "<br>", "Crimes ",
                tract_data2$crimes_per_person)

# creating color palette
palette <- colorNumeric(
  palette = "magma",
  domain = tract_data2$crimes_per_person
)

# creating map per population
map <-leaflet() %>%
  addProviderTiles("CartoDB.Positron") %>%
  addPolygons(data = tract_data2,
```

```
              fillColor = ~palette(crimes_per_person),
              color = "aliceblue",
              fillOpacity = 0.6,
              weight = 1,
              smoothFactor = 0.3,
              popup = popup) %>%
  addLegend(pal = palette,
            values = tract_data2$crimes_per_person,
            position = "bottomright",
            title = "Crimes Per Person",)

map
```