The Most Common Crime Cases Happened between Daytime and Nighttime across Different Location Types in Austin

Catherine Zhong (xz8222)

2023-10-28

Introduction

The Austin Crime Reports dataset is a valuable resource for understanding and analyzing crime incidents in Austin, Texas. This dataset contains records of incidents that the Austin Police Department responded to and documented in reports, making it an essential tool for understanding criminal activities in the area and the law enforcement's response to them.

The dataset comprises 27 variables, but for this analysis, I've chosen to focus on a subset of key variables, including the Highest Offense Description, its unique code, Occurred Date and Time, and the general description of the premise where the incident occurred. These selected variables are crucial for categorizing and analyzing the crimes effectively. I've also narrowed my analysis to crimes that occurred during a specific time frame, starting from September 26, 2023, to October 21, 2023.

Whenever I am making my travel plan to a new city, I would usually search for the crime map to determine which area I should go and not go, as well as a safe area to choose my hotel. Therefore, I started to think about how these crime maps are generated and what are the works needed to be done behind the scene. Here is a link to Austin Crime Map (https://www.neighborhoodscout.com/tx/austin/crime)

Given the dataset's nature, it can be employed for various purposes, such as crime trend analysis, geographic analysis, time-based insights, and data-driven decision making to improve public safety and reduce criminal activities in Austin. In summary, the Austin Crime Reports dataset is a valuable resource for addressing crime-related issues in the region and making informed decisions related to crime prevention and public safety.

In this project, I would like to investigate the relationship of the number of cases happened among the top 10 offense types between the occurred times (Daytime and Nighttime) among the top 5 specified location types that crimes took place from 09/26/2023 to 10/21/2023. Each row in the dataset represents an independent crime report. For example, a theft incident occurred on September 26, 2023, at 12 am in a residence/home area.

I expect that there would be more crimes occurred during nighttime in general, while the family disturbance would be the most common in the residence/home area. Moreover, I would like to explore what location types is the most dangerous, which means they have the highest number of crime occurred and how are each type of crime related to the location types?

```r
# Read the Crime Report dataset in csv file
library(readr)
crime <- read_csv('Crime_Reports.csv',show_col_types = FALSE)

# Show the dataset
crime
```

```
## # A tibble: 5,510 × 5
##    Highest Offense Desc…¹ `Highest Offense Code` `Occurred Date` `Occurred Time`
##    <chr>                                   <dbl> <chr>                     <dbl>
##  1 THEFT                                     600 09/26/2023                    0
##  2 AGG ROBBERY/DEADLY WE…                    300 09/26/2023                 2036
##  3 DEL CONTROLLED SUB/NA…                   1804 09/26/2023                 1504
##  4 CRUELTY TO ANIMALS                       2717 09/26/2023                 1907
##  5 CRIMINAL MISCHIEF                        1400 09/26/2023                  738
##  6 CRASH/FAIL STOP AND R…                   3604 09/26/2023                 1224
##  7 BURGLARY OF VEHICLE                       601 09/26/2023                 1639
##  8 CRIMINAL MISCHIEF                        1400 09/26/2023                 1237
##  9 MAIL THEFT                               8503 09/26/2023                 1300
## 10 COUNTERFEITING                           1006 09/26/2023                 1455
## # ℹ 5,500 more rows
## # ℹ abbreviated name: ¹`Highest Offense Description`
## # ℹ 1 more variable: `Location Type` <chr>
```

Methods

I started with 5,510 rows and 5 columns and end up with 2,428 rows and 6 columns. Since I am aiming to analyze the relationship between top 10 offense types and the occurred time among the top 5 specified location types that crimes took place, those offense types and location types that are less frequent are dropped, with a total of 3,082 observations. This may cause a potential issue of not analyzing the data comprehensively, and the results ultimately used for crime prevention may not be accurate enough. However, because there are so many types of crime, it takes too much efforts and inputs to analyze and prevent all of them. Here, we will focus on the most common types and get results in the most efficient way.

My dataset is tidy because each variable have its own column, each observation have its own row, and each value have its own cell. This is also because each observation represents an independent incident report.

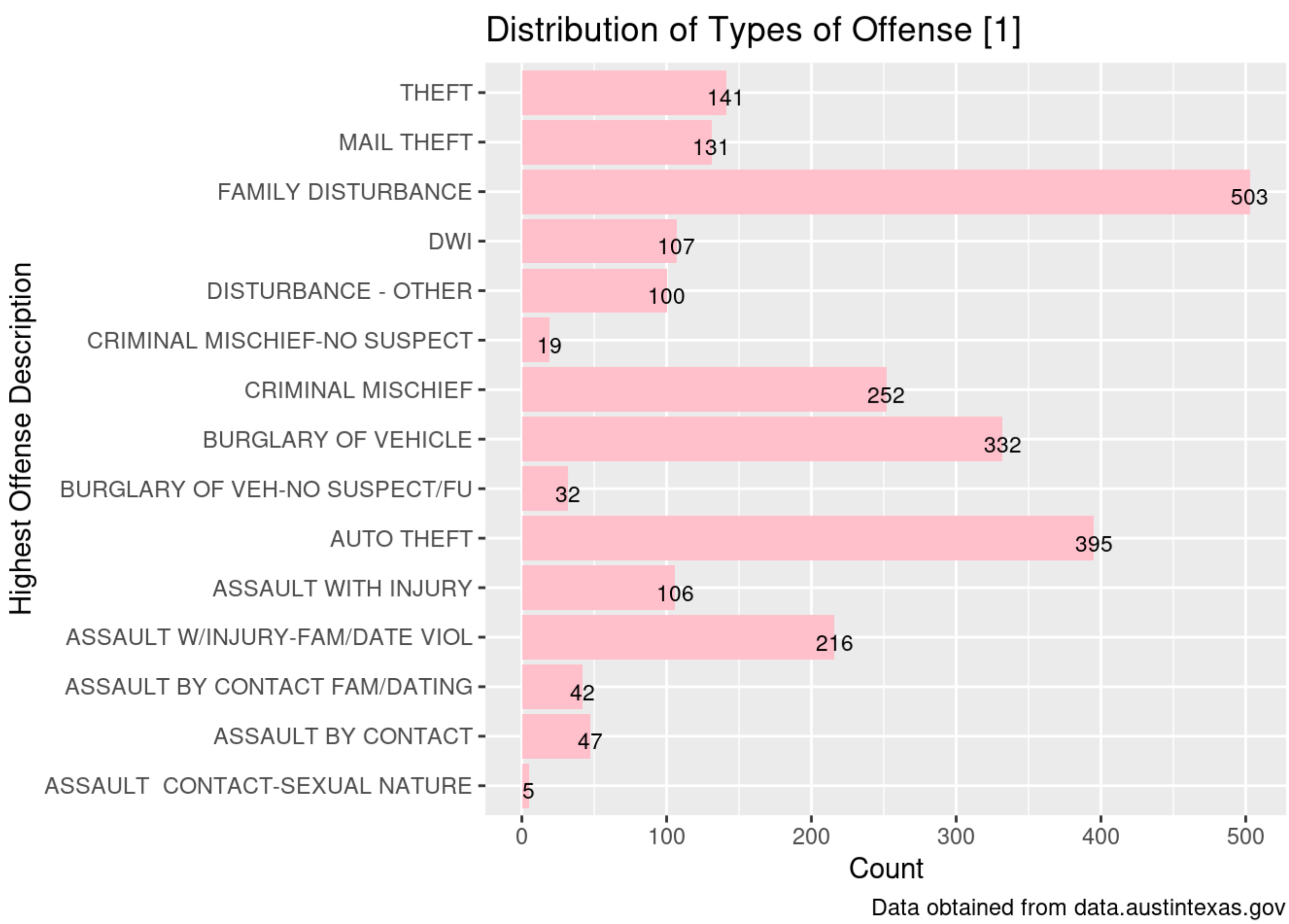
```r
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──
## ✔ dplyr     1.1.2     ✔ purrr     1.0.2
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.3     ✔ tibble    3.2.1
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ── Conflicts ────────────────────────────────────────── tidyverse_conflicts() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts
to become errors
```

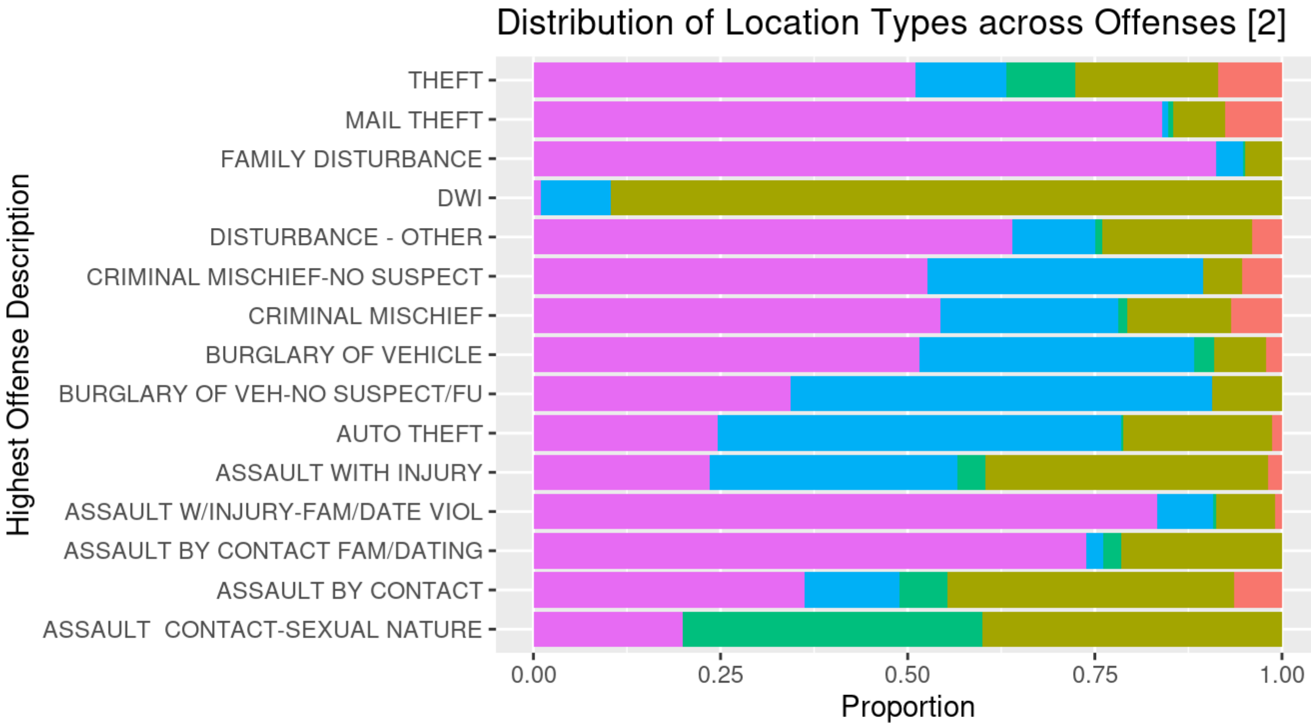
```
location <- crime %>%
  # seperate each unique location type
  group_by(`Location Type`) %>%

  # get the frequency of each location type
  summarize(count = n()) |>

  # get the top 6 specified location type and discard the 'unknown' category in location
type
  slice_max(n = 6, count)%>%
  filter(`Location Type` != "OTHER / UNKNOWN")

code <- crime %>%
  filter(`Location Type` %in% location$`Location Type`) %>%
  # seperate each unique Highest Offense Code
  group_by(`Highest Offense Code`) %>%

  # get the frequency of each Highest Offense Code
  summarize(count = n()) |>

  # get the top 10 Highest Offense Code
  slice_max(n = 10, count)

# filter out the top 5 specified location type and the top 10 Highest Offense Code in th
e 'crime' dataset
New <- crime %>%
  filter(`Location Type` %in% location$`Location Type` &
         `Highest Offense Code` %in% code$`Highest Offense Code`)

# Creat a new variable called 'period' to seperate the occurred time into daytime and ni
ghttime
New <- New %>%
  mutate(period = ifelse(`Occurred Time` >= 600 & `Occurred Time` < 1800, 'Daytime', 'Ni
ghttime'))

# Show the dataset
New
```

```
## # A tibble: 2,428 × 6
##    Highest Offense Desc…¹ `Highest Offense Code` `Occurred Date` `Occurred Time`
##    <chr>                                   <dbl> <chr>                     <dbl>
##  1 THEFT                                     600 09/26/2023                    0
##  2 BURGLARY OF VEHICLE                       601 09/26/2023                 1639
##  3 CRIMINAL MISCHIEF                        1400 09/26/2023                 1237
##  4 MAIL THEFT                               8503 09/26/2023                 1300
##  5 ASSAULT W/INJURY-FAM/…                    900 09/26/2023                 1138
##  6 AUTO THEFT                                700 09/26/2023                   30
##  7 THEFT                                     600 09/26/2023                 1500
##  8 ASSAULT  CONTACT-SEXU…                    902 09/26/2023                 1900
##  9 FAMILY DISTURBANCE                       3400 09/26/2023                 1803
## 10 FAMILY DISTURBANCE                       3400 09/26/2023                 1702
## # ℹ 2,418 more rows
## # ℹ abbreviated name: ¹`Highest Offense Description`
## # ℹ 2 more variables: `Location Type` <chr>, period <chr>
```

Results



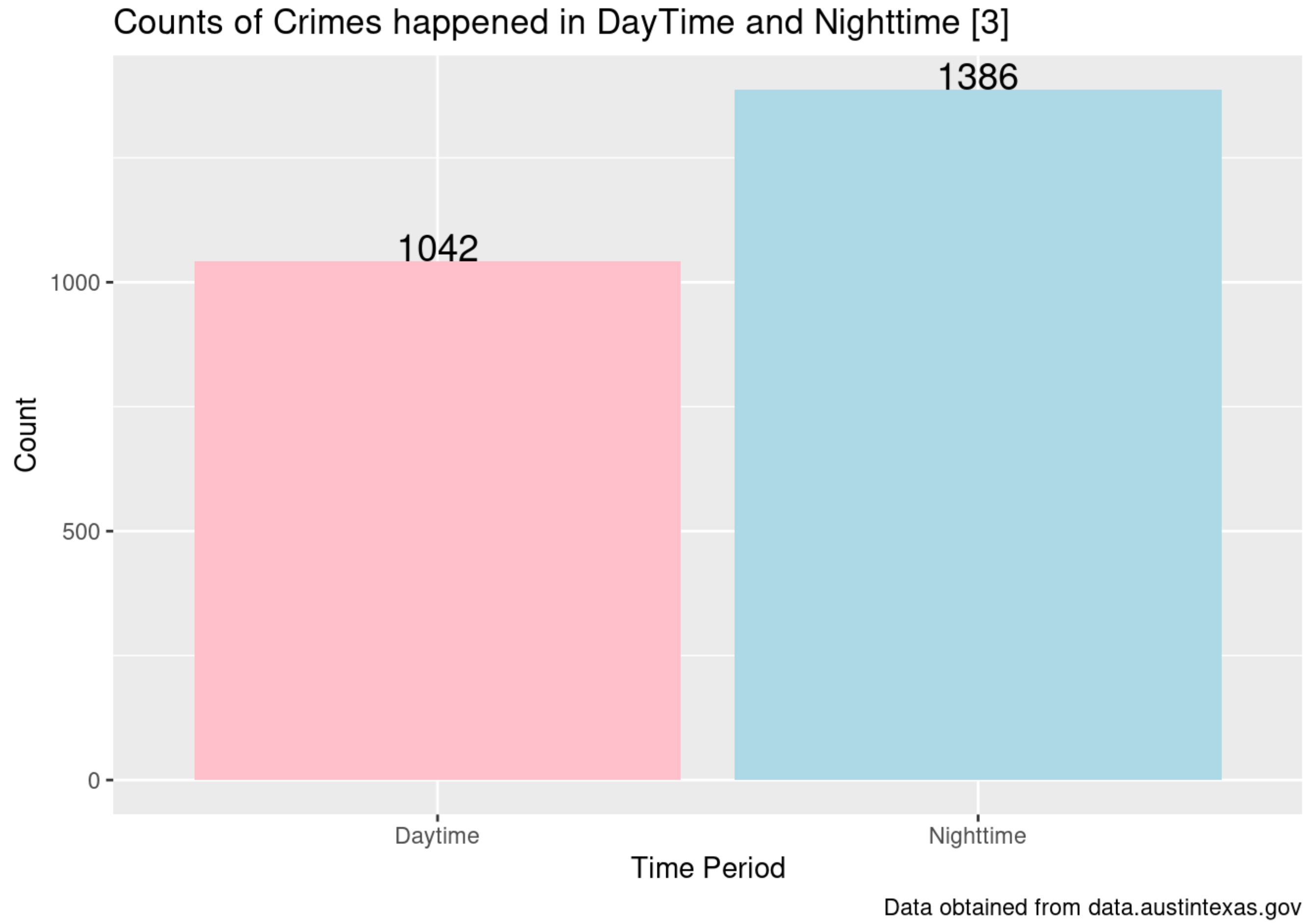
From graph [1], it shows that the most common type of offense is family disturbance, with a frequency of 503 cases in the past month. The second common offense is auto theft and the third common one is the burgalry of vehicle.



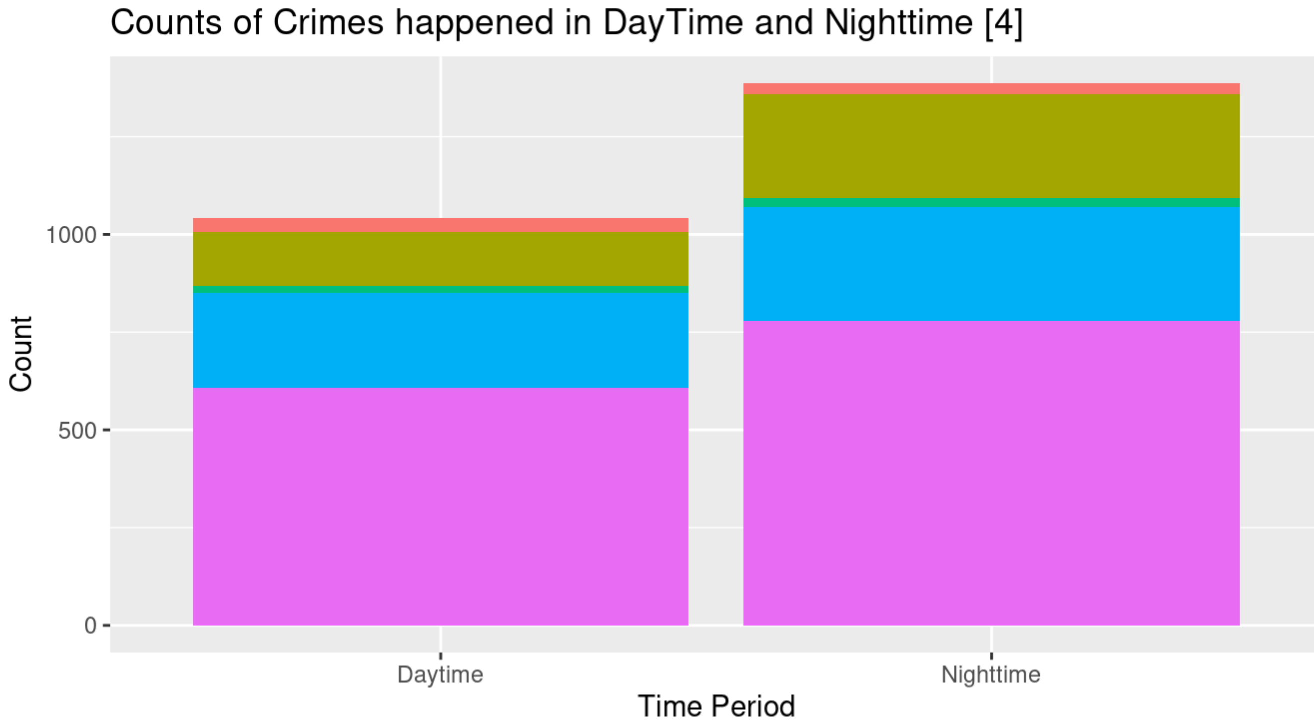
Color codes for Location Types:
Orange: Commercial / Office Building
Greyish-green: HWY/Road/Alley/Street/Sidewalk
Green: Park/Playground
Blue: Parking/Drop lot/Garage
Purple: Residence/Home

Data obtained from data.austintexas.gov

According to graph [2], we can see that the most common location for crimes, including thefts, family and other disturbance, criminal mischiefs, burglary of vehicle, assaults (except for contact-sexual nature and with injury), happened is the residence/home area. In other words, residence/home areas have the highest crime rate. The HWY/Road/Alley/Street/Sidewalk is where DWI and assault with injury/by contact/contact-sexual nature usually take place. More than a half of the burglary of vehicle (no suspect/FU) and auto theft occurred in the Parking/Drop lot/Garage. It also has a significant proportion on assult with injury.



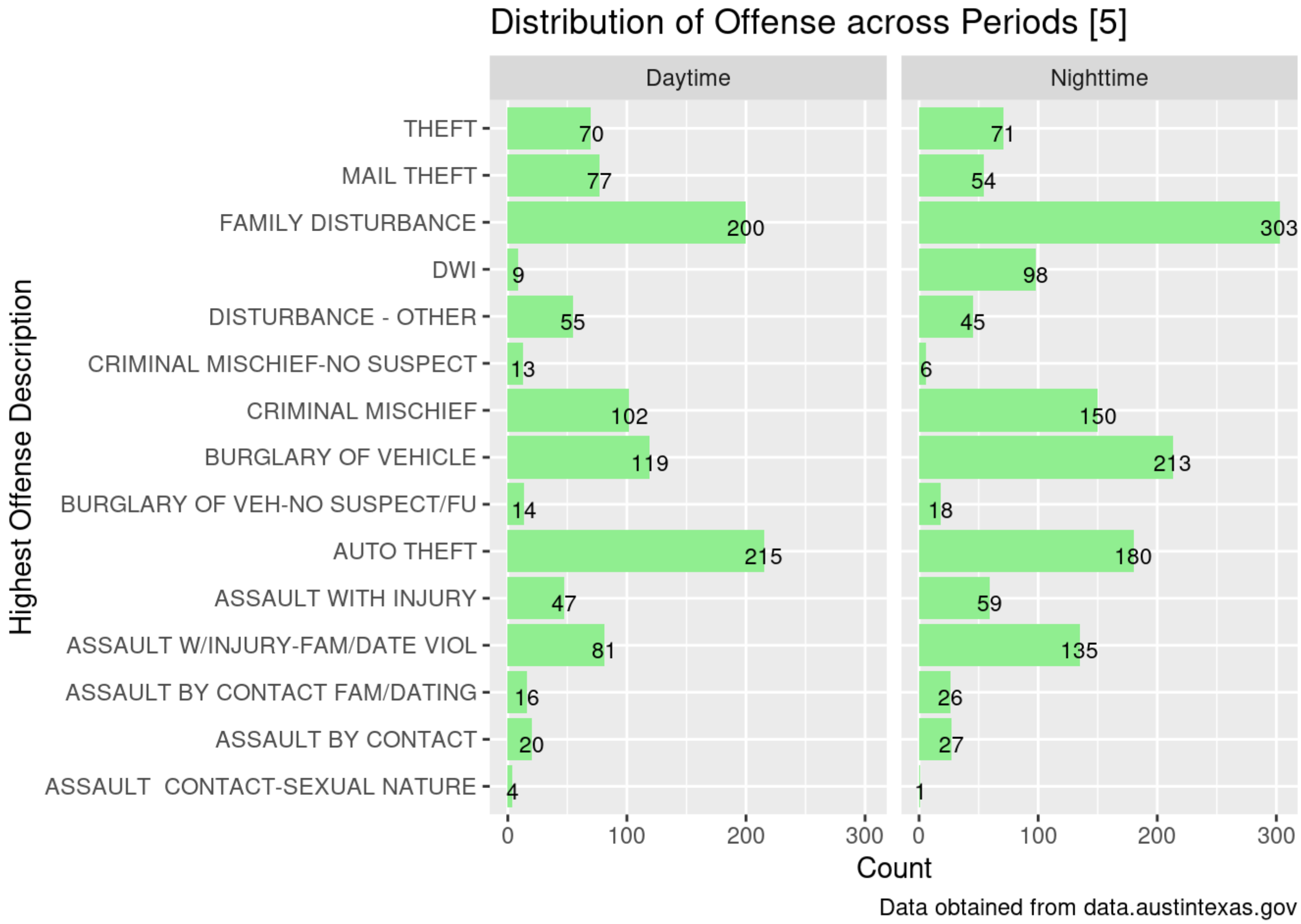
As graph [3] shows, there are more crimes happened at the nighttime with a total number of 1386 cases in the past month. Meanwhile, 1042 cases happened during the daytime is still significantly high.



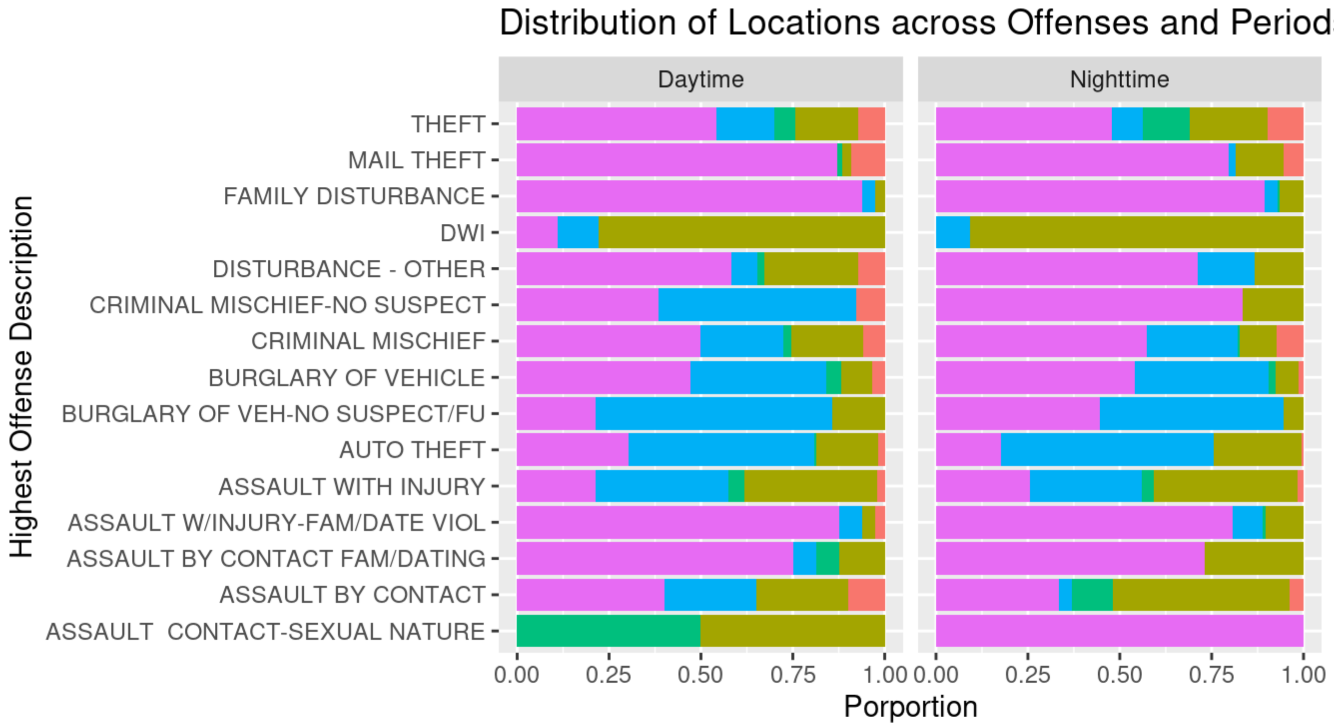
Color codes for Location Types:
Orange: Commercial / Office Building
Greyish-green: HWY/Road/Alley/Street/Sidewalk
Green: Park/Playground
Blue: Parking/Drop lot/Garage
Purple: Residence/Home

Data obtained from data.austintexas.gov

In graph [4], it is obvious that about half of the offenses, over 1000 cases, take place in the residence/home areas, in both time periods. In the mean time, the residence/home areas and the HWY/Road/Alley/Street/Sidewalk has more crime reported during nighttime, while there are no significant differences between 2 periods for the rest of the location types.



graph [5] shows that, during the daytime, the most common offenses happened are auto theft and family disturbance. Futhermore, the number of family disturbance cases are even much higher at the nighttime, than both the number in the day, and other offenses at night. Burglary of vehicle is the second common and auto theft is the third common at night.



Color codes for Location Types:
Orange: Commercial / Office Building
Greyish-green: HWY/Road/Alley/Street/Sidewalk
Green: Park/Playground
Blue: Parking/Drop lot/Garage
Purple: Residence/Home

Data obtained from data.austintexas.gov

The last visualization is composed of all 3 key variables. It is obvious that the Residence/Home areas have a significant proportion in most of the offense types, regardless the time periods. Notably assault contact-sexual nature happened in both HWY/Road/Alley/Street/Sidewalk and Park/Playground only during the day and Residence/Home only at night.

Discussion

As I expected, there were more crimes occurred during nighttime in general, and the family disturbance would be the most common in the residence/home area. Moreover, this is also the most dangerous location type, with the highest number of crime occurred. Below are more key findings based on the exploratory data analysis, and implications derived from the visualizations:

Common Offenses: Graph [1] reveals that family disturbance is the most common type of offense in the past month, with a frequency of 503 cases. Auto theft and burglary of vehicles follow as the second and third most common offenses. This information can be crucial for the City of Austin to prioritize resources and strategies for addressing these specific crime types.

High Crime Rate in Residence/Home Areas: Graph [2] highlights that the residence/home areas have the highest crime rate, where various types of crimes, including thefts, family disturbances, and burglaries of vehicles, commonly occur. This finding could lead to increased community awareness and neighborhood safety measures.

Location-Specific Insights: The data also provides other valuable location-specific insights, such as the prevalence of DWI and assault on HWY/Road/Alley/Street/Sidewalk and the concentration of burglary of vehicles and auto theft in Parking/Drop lot/Garage. This information can guide law enforcement efforts and community policing initiatives.

Day vs. Night Crime Rates: Graph [3] indicates that more crimes occurred at night (1386 cases) compared to daytime (1042 cases) in the past month. This suggests a need for additional nighttime security measures and patrols to address this higher nighttime crime rate.

Location and Time Period Analysis: Graph [4] emphasizes that residence/home areas have consistently high crime rates in both daytime and nighttime. This indicates the need for community engagement and crime prevention strategies tailored to these areas. The graph also shows that some specific location types, like HWY/Road/Alley/Street/Sidewalk and Park/Playground, are more active during the day.

Offense Type, Time of Day & Locations: Graph [5] provides a detailed breakdown of offense types during the daytime and nighttime. And the last visualization is composed of all 3 key variables. It is obvious that the Residence/Home areas have a significant proportion in most of the offense types, regardless the time periods. Notably assault contact-sexual nature happened in both HWY/Road/Alley/Street/Sidewalk and Park/Playground only during the day and Residence/Home only at night. This is because there are only 5 cases in total, which is the least frequent type of crime.

It is also important to inform the City of Austin about the limitations of the dataset, including potential inconsistencies in the data that I filtered out, some categories, typos, and missing data, since quality of the data and the potential issues is vital for making informed decisions based on this information.

Regarding the ethical implications of the data analysis, the City of Austin should take the potential impacts on specific communities in to consideration, ensuring that law enforcement and public safety efforts are equitable and unbiased. This involves being mindful of potential biases in data and analysis, maintaining transparency in decision-making processes, protecting individual privacy in data collection, and actively engaging the community in discussions about public safety and crime prevention strategies. By addressing these ethical aspects, the city can ensure that its actions are just, transparent, and responsive to the diverse needs of its communities.

Reflection, acknowledgements, and references

Conducting this project has been an insightful experience for me. The most challenging thing was cleaning the data, which took a decent amount of effort and time because I need to think about how I can make this dataset into the easiest version for analysis.

Through this process, I gained valuable experience and get more familiar with cleaning data and creating visualizations to better understand the relationships between variables in R Studio. This project has equipped me with a comprehensive understanding of how to analyze a random dataset and what values can it bring to the society.

I want to express my gratitude to the TAs and Professor Guyot who provided guidance and support throughout the project. Additionally, I acknowledge the Austin Police Department, the data owners, for making the dataset available to the public, which gives me a valuable opportunity to practice while doing something meaningful. The following link is the Austin Crime Reports Dataset (https://data.austintexas.gov/Public-Safety/Crime-Reports/fdj4-gpfu) for reference and background context.