

Crime types and Frequency Analysis of Chicago based on Historical Data.

TEAM

Divya Kampalli ,

Bhavya Sirobhushanam ,

Venkateshwar Reddy Kasturi

2023-12-03

Abstract:

This project delves into the analysis and comparison of historical crime patterns in Chicago, offering a unique perspective by also predicting future crime rates. The insights derived from this study are particularly beneficial for immigrants, providing them with crucial information to make informed decisions about their places of residence. Additionally, this analysis serves as a valuable resource for tourists, students, and travelers, guiding them in planning their visits during times deemed safer based on crime trends. Understanding these patterns not only helps in personal decision-making but also contributes to a broader awareness of safety in urban environments.

A specific question or set of questions that the project seeks to address:

1. What are the historical crime patterns in Chicago, including crime types, locations, and trends over time?
2. How have these crime patterns contrasted over different periods?
3. Can we forecast future crime rates in Chicago based on historical data and trends?
4. What information and recommendations can be provided to immigrants to help them select safer places to reside in Chicago?
5. How can tourists, students, and travelers be supported with information to schedule trips during safer periods in the city?

A proposed methodology/approach to the analysis that will be performed.

Data Collection:

The crime data was obtained from the Chicago Police Department's CLEAR (Citizen Law Enforcement Analysis and Reporting) system, which is available for public download. The data contains information on the date, time, location, and type of crime for each reported incident from 2017 to present.

The key R packages used for the analysis are `data.table`, `ggplot2`, `sf`, `dplyr`, `stringr`, `ggpubr`, `lubridate`, `tseries`, `Metrics`, `gridExtra`, `forecast`, and `padr`. These packages enabled data import, manipulation, visualization, time series analysis, spatial mapping, and modelingData Processing and Cleaning

1. Initial Data Exploration:

Utilizing commands like ``head(DT)`` and ``summary``, we gained an initial understanding of the data structure and content.

2. Data Extraction:

We focused on the last five years to ensure relevance and accuracy in our analysis.

3. Data Refinement:

- Renamed variables for clarity and consistency.
- Identified and removed duplicates based on the 'Case Number', ensuring data integrity.
- Addressed missing values through a methodical approach:
- Replaced NAs in key columns like latitude, longitude, and location descriptions.

- Utilized related records to fill in missing information for District, Ward, and Community **Area**.
- Carefully removed records where missing values couldn't be logically replaced.

Data Analysis and Insights

1. Temporal Analysis:

- Utilized `boxplot. stats` to understand the distribution of crime across different years.

2. Geographical Distribution:

- Identified the number and codes of districts and community areas in the dataset.
- Removed data with illogical community codes (e.g., 0).

3. Crime Pattern Identification:

- Extracted the top 5 most common crimes, providing a focused view of prevalent criminal activities.

Data processing and pipeline - cleaning, imputing, transformation, outlier detection, etc.

Data Cleaning:

- Remove duplicate records, if any.
- Handle missing values in the dataset, especially in columns critical for analysis (e.g., crime type, location, date).
- Check for and correct any data entry errors, such as typos or inconsistent formatting.

Data Imputation:

- For missing values in the dataset, impute them with appropriate values. For example, you can use the mean, median, or mode for numerical features and the most frequent category for categorical features.

Data Transformation:

- Convert date and time information into a standardized format.
- Encode categorical variables such as crime type, location, and neighborhood using one-hot encoding or label encoding.
- Scale numerical features, like the number of incidents or crime rates, to ensure they have similar scales.

Outlier Detection:

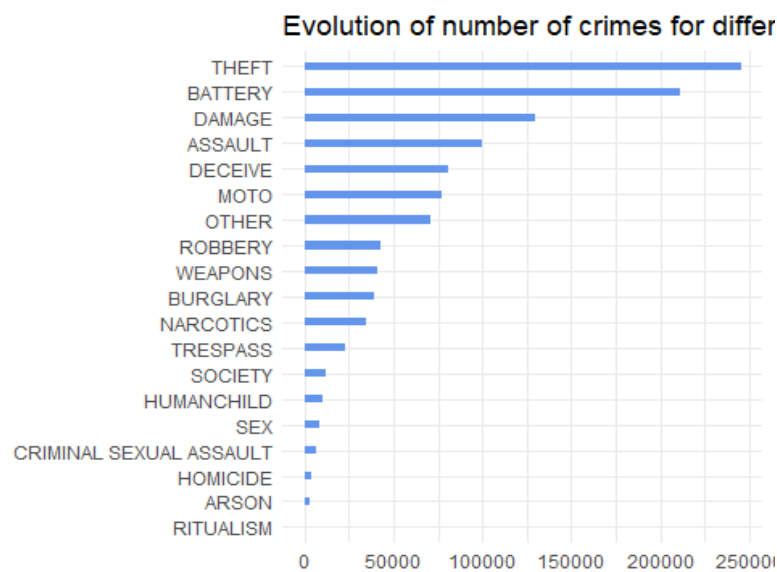
- Identify and handle outliers in the dataset, especially for numerical features that may indicate unusually high or low crime rates in specific areas or time periods.

Feature Engineering:

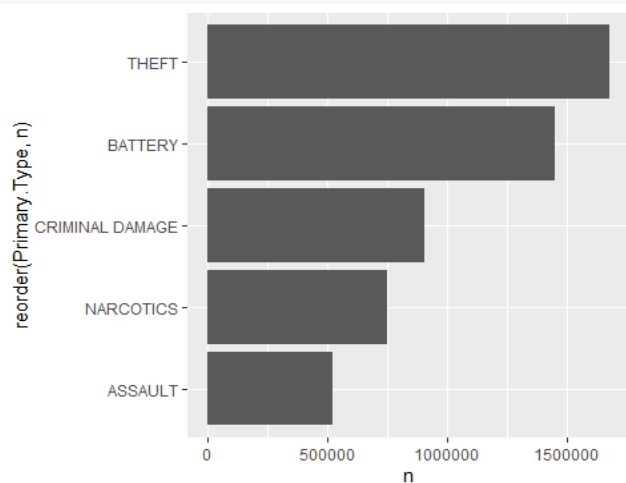
- Create new features that can be insightful for analysis. For example, you might derive features like "day of the week" or "season" from the date information to examine temporal patterns.
- Aggregate data to create features such as "crime rate per capita" or "crime density in neighborhoods."

Key Findings

"Evolution of number of crimes for different types



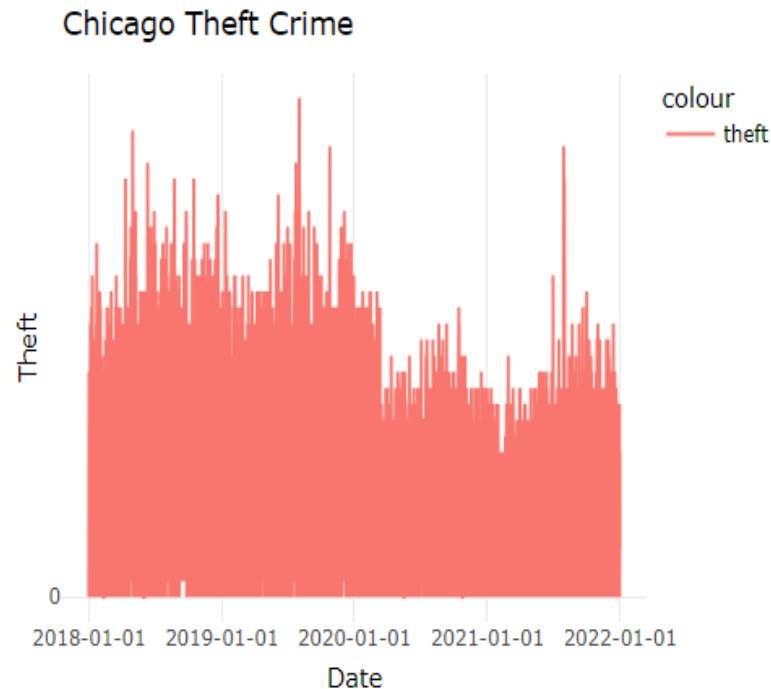
Top 5 Crimes in Chicago



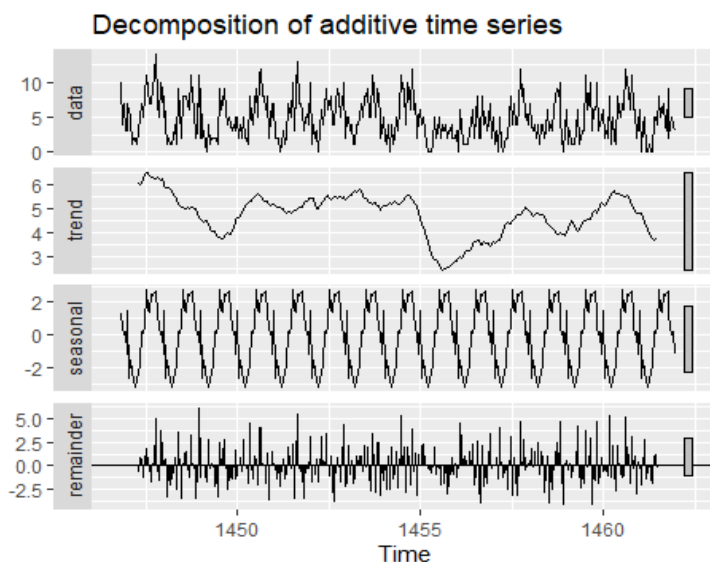
From the above plot, we can see that "THEFT" is the highest occurring crime. Let's do Time series Analysis on THEFT.

Time series analysis for Theft Create Prediction Time Frame We have considered only 5 years data so that it will be accurate.

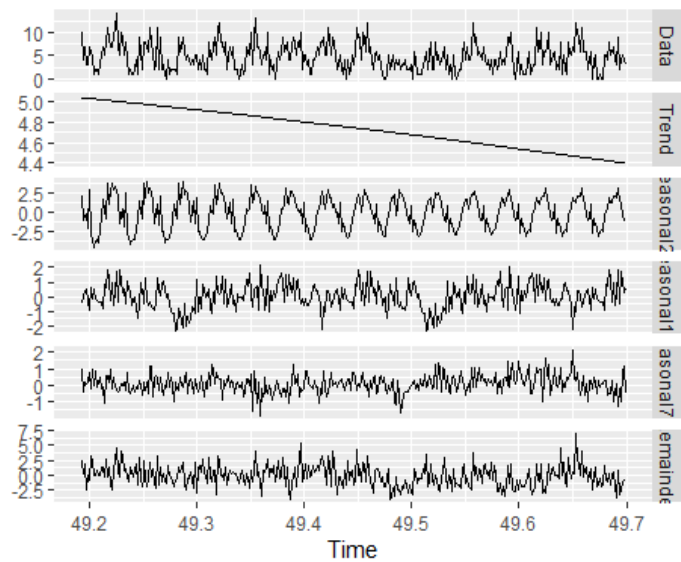
To create a time-series model, we need to create a time-series object from our train data. Time-series object will be based on theft as it is the one that we are going to predict, we set the frequency to be 24 as it is total hour of reported crime for 1 day.



Now, we will use autoplot to see the trend and sesonality



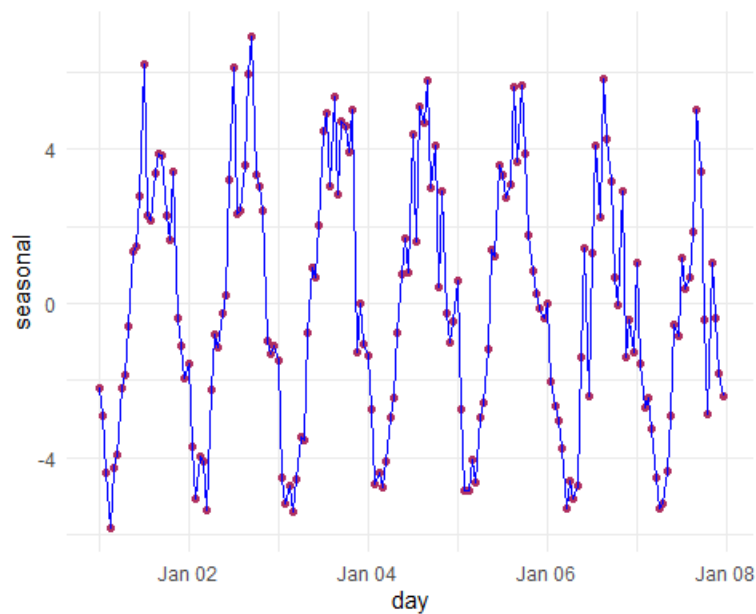
Upon examining the plot, it becomes evident that the trend still reveals certain patterns, resembling a seasonal nature. This suggests the existence of additional seasonality patterns that haven't been captured by the current visualization. To address this, we aim to construct a Multi-Seasonal Time Series Object. Creating and Decompose MSTs Object



From the plot above, we can see the trend of the Theft Crime is already going smooth. The Theft Crime trend itself is decreasing in the last years.

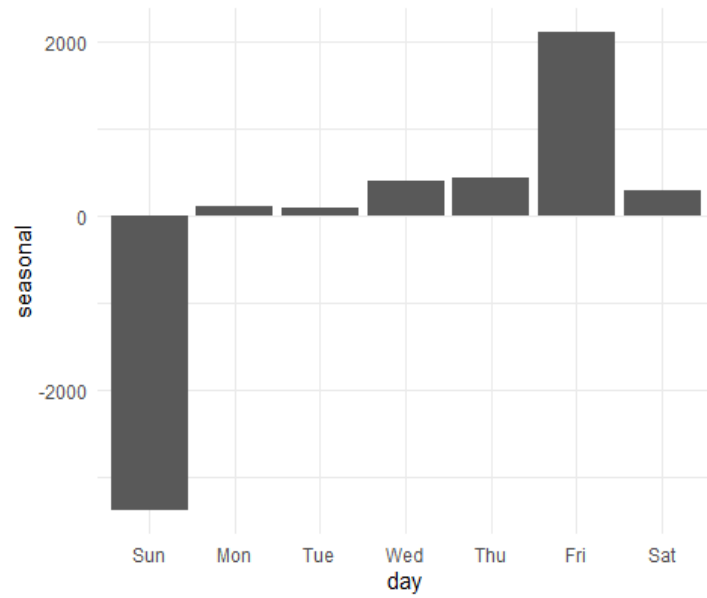
#Seasonality Analysis

Hourly Seasonality



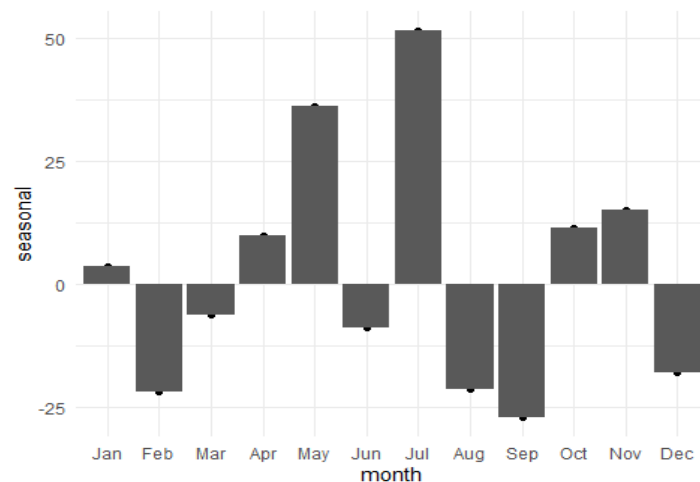
From the above graph, we can see that thefts are occurring more during midday and falling during night.

Daily Seasonality



From the Daily Seasonality Graph, we can see that theft count increases from wednesday and reaches its peak on friday and will fall. The least number of thefts are on sunday.

Monthly seasonality



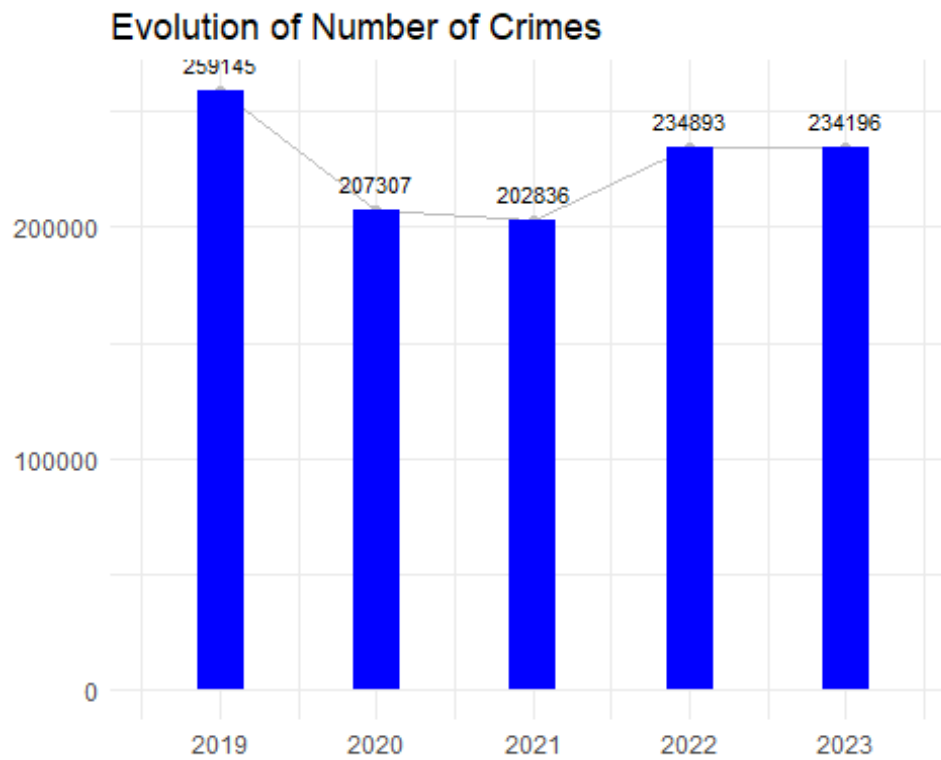
From the above graph, we can see that the Highest number of crimes occurred in July.

In summary We have successfully predicted the frequency of theft crimes based on our analysis. It is reasonable to conclude that theft crime will probably start to rise at 10 a.m., peak at 5 p.m. (after business hours), and then continue to rise until 12 a.m. More crimes occur on Fridays. The actual crime is more likely to occur between June and October.

Similarly, we can do Time series analysis for each crime.

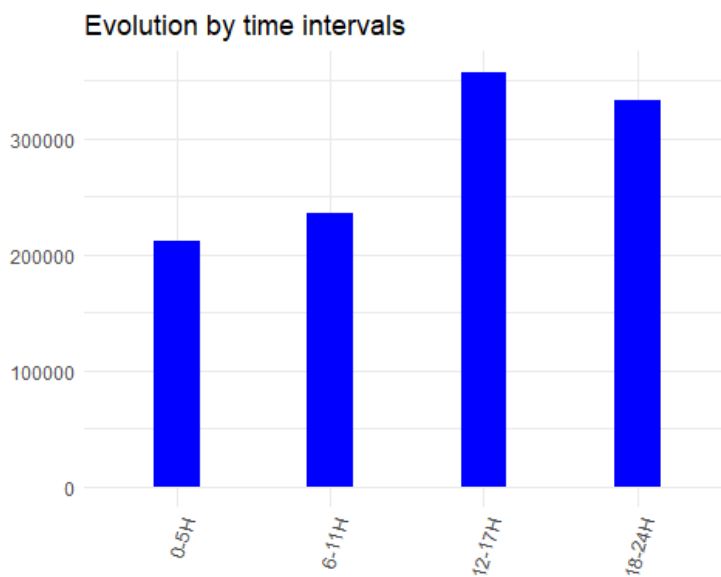
#Analysis and Visualisation

Plotting Number of Crimes versus Year



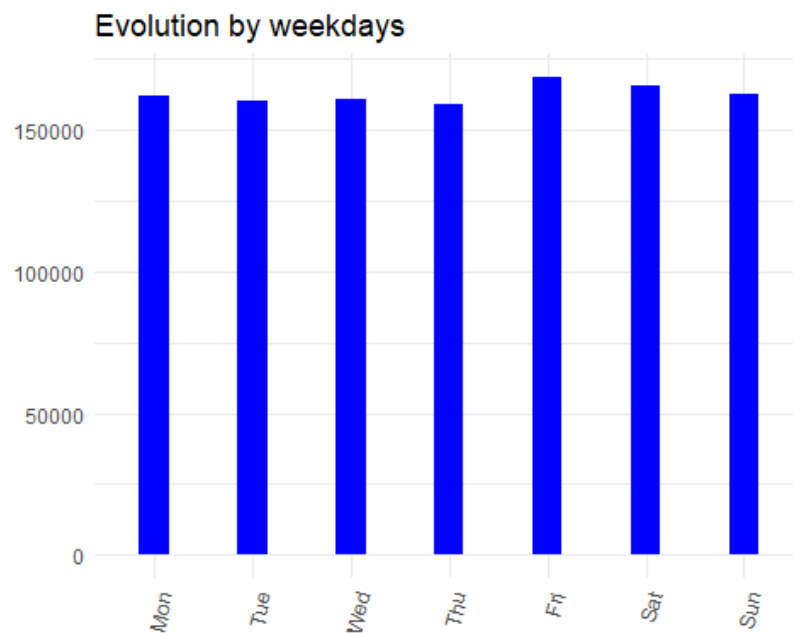
The number of cases decreased from 2019 to 2020 and the trend increased after 2021.

By time intervals



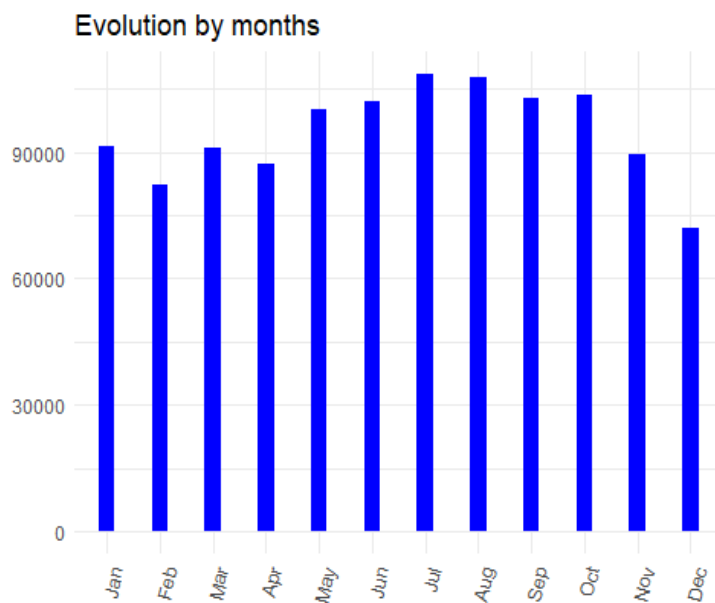
From the above plot, we can see that the most number of crimes are happening from 12-5 and the trend decreases later.

By WeekDays



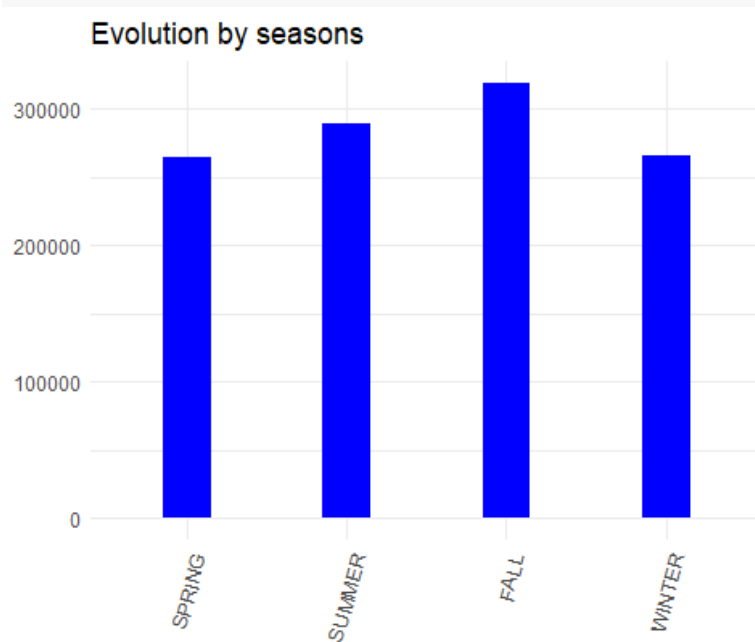
By the above graph we can see that the trend is almost the same across all the days. There is a slight increase of crimes on Fridays and Saturdays.

By Months



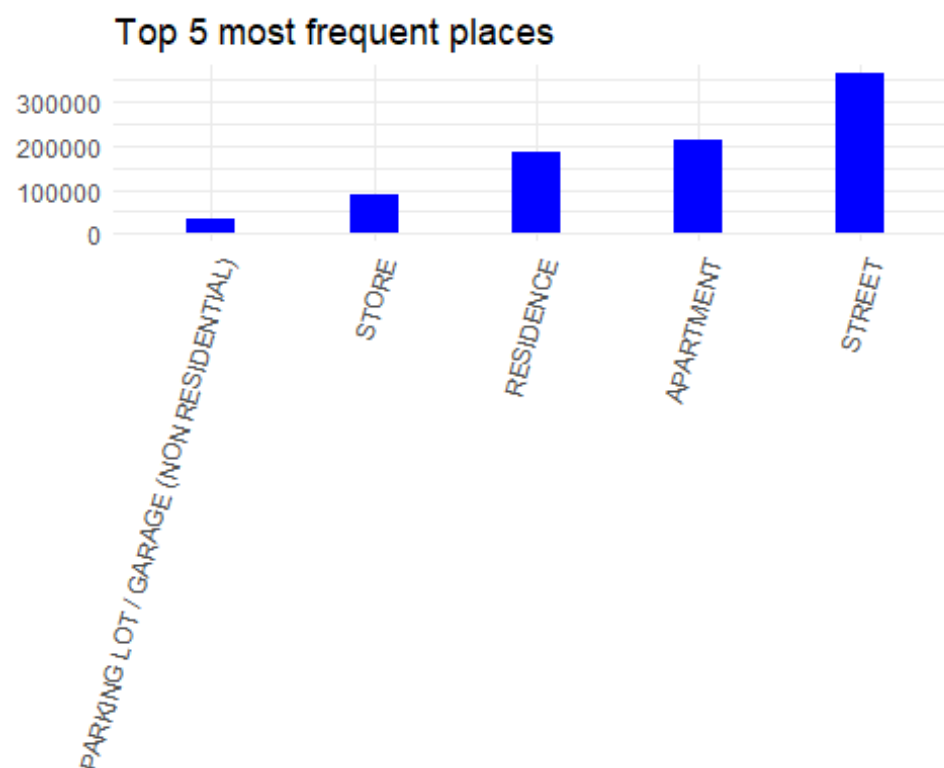
Crimes were more likely to happen in June to August and are less likely to happen in December, February.

By seasons



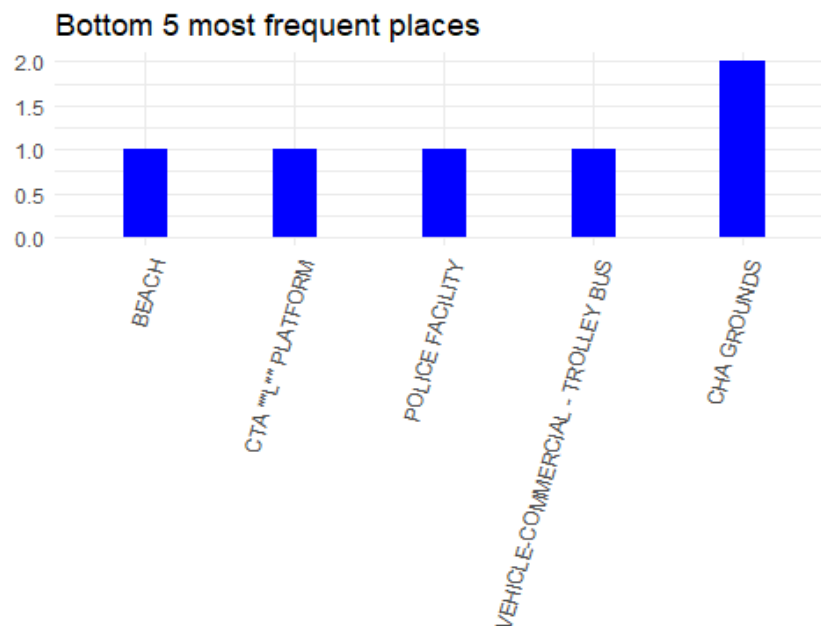
Crimes are more likely to happen in fall and less likely to happen in winter.

Top 5 most frequent Crime areas



Street is the top place where crime can happen. Dont think your apartment/residence is safe. The next place where crimes could take place is apartment followed by residence.

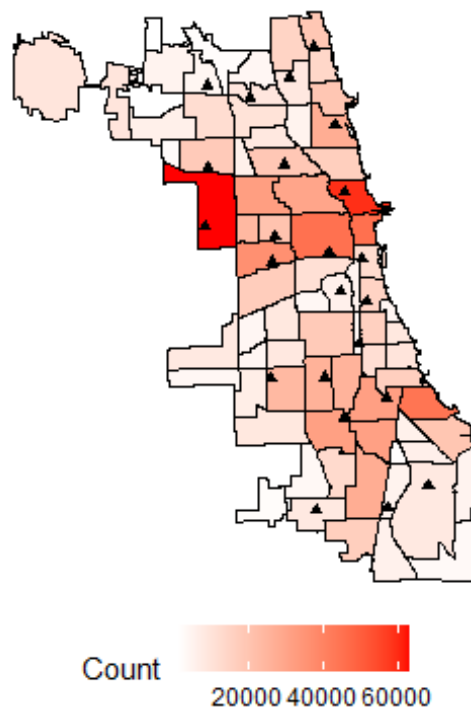
Bottom 5 most frequent Crime areas



Least number of crimes happen in the above places.

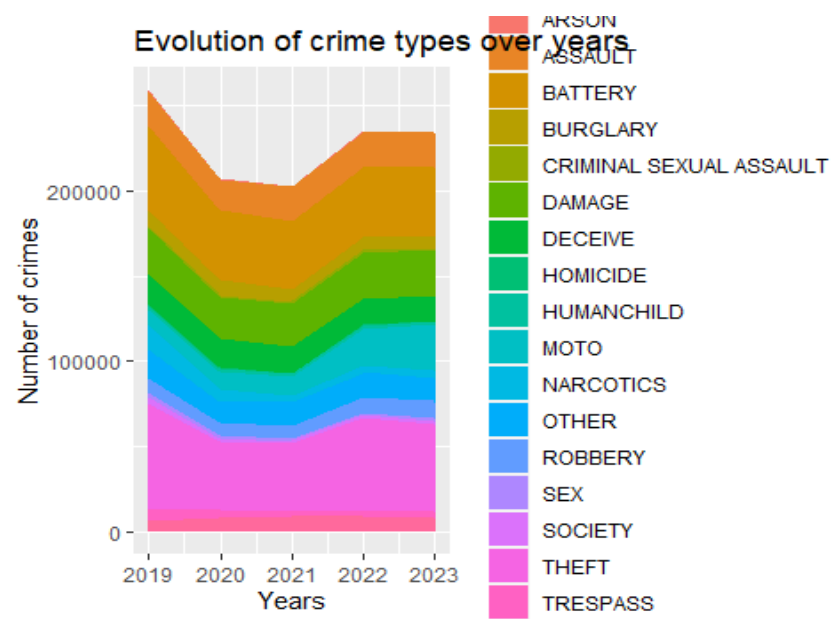
Checking Number of Communities and count of crimes in each Community

Number of crimes per community



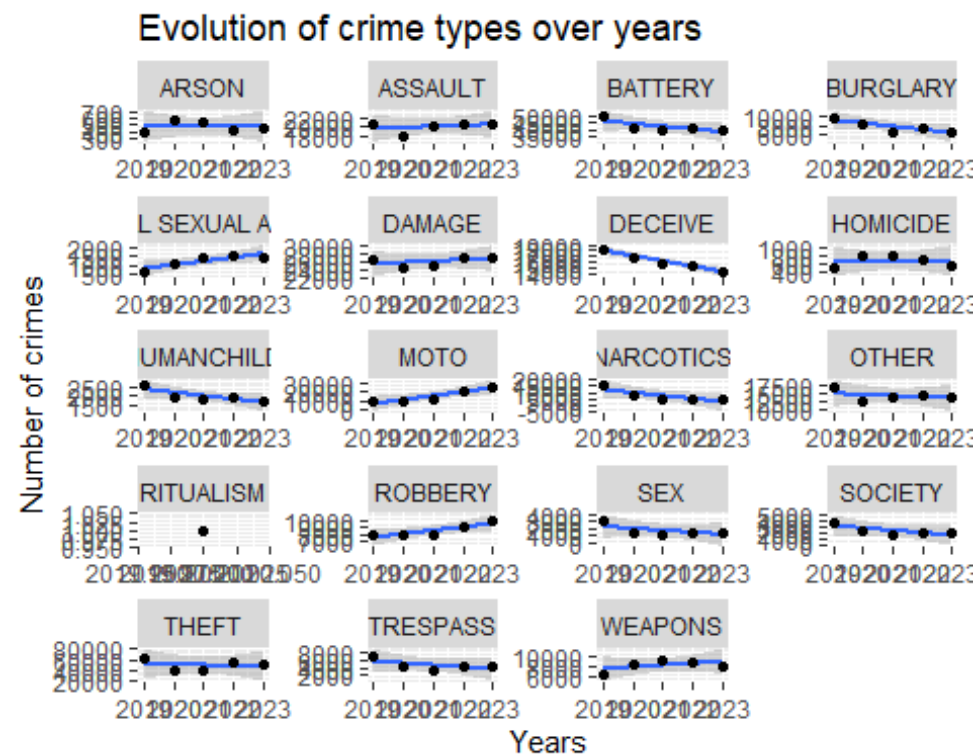
From the above plot, we can see that Austin has highest number of crimes(63537).

Evolution over years

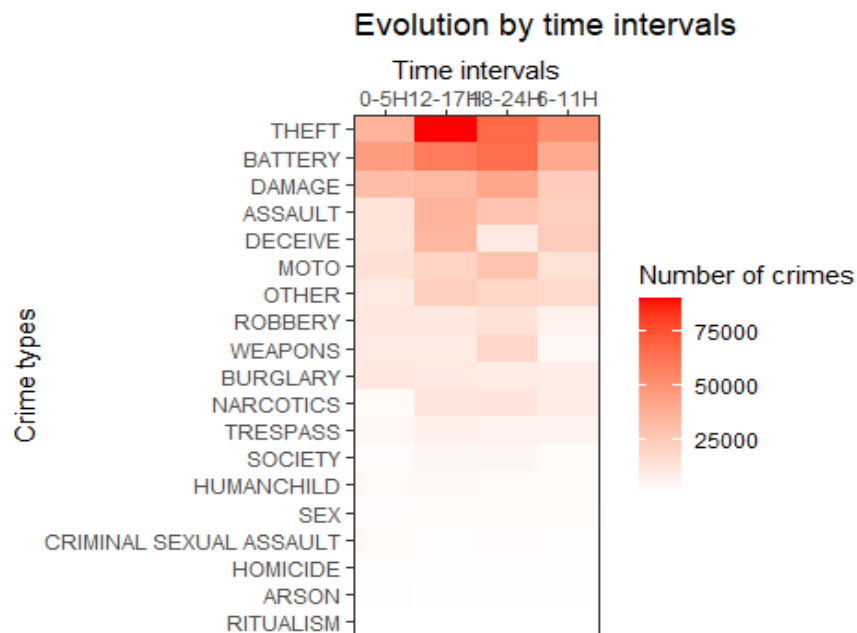


From the above graph we can see the evolution of crime types over the years.

Evolution of crime types over years multiplots

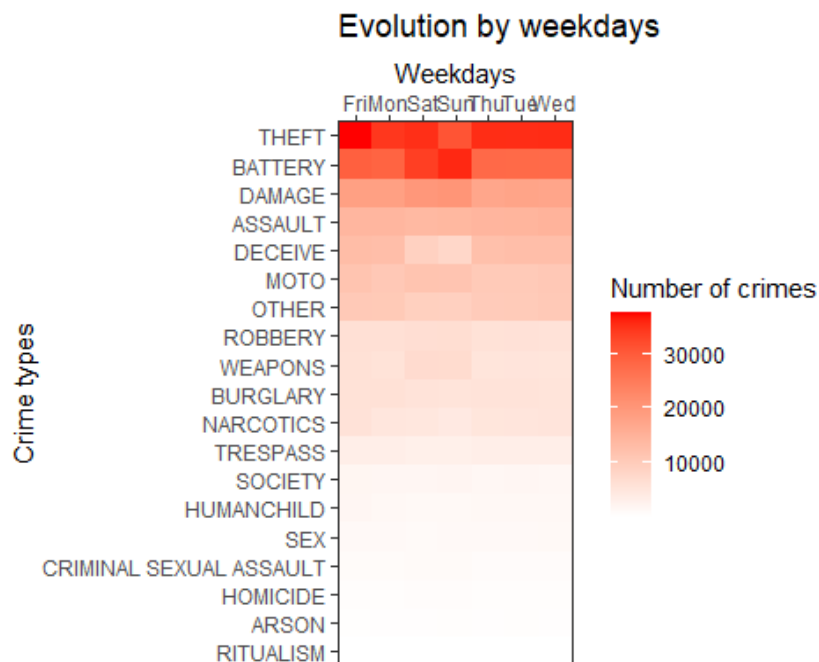


Evolution of crime by time intervals



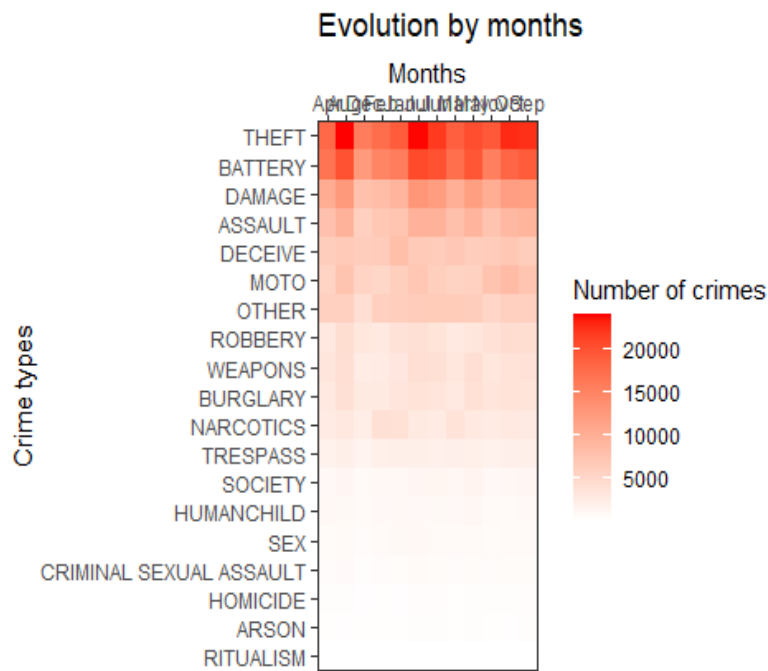
From the above heatmap we can see that theft is the highest at time interval 12pm to 5pm.

Evolution of crime by weekdays



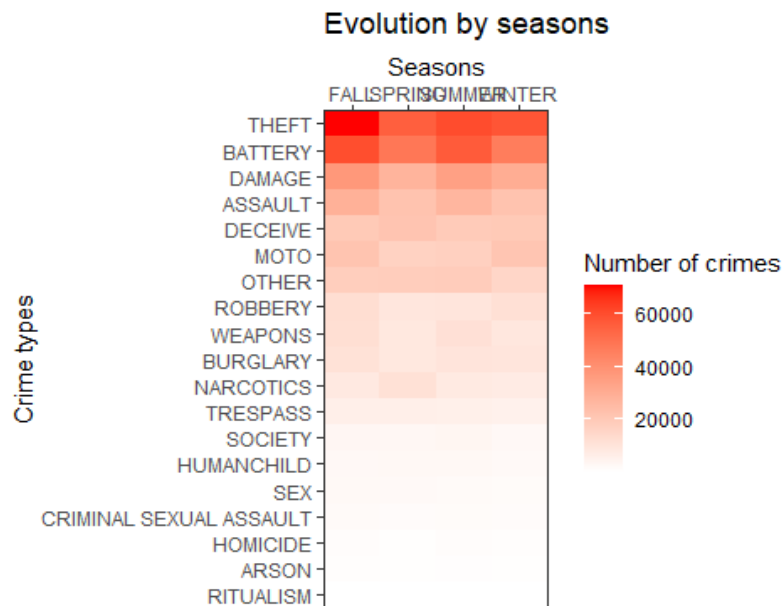
From the above heatmap, we can see that theft is high in almost all days except Sunday and it is contrary for battery.

Evolution of crime by months



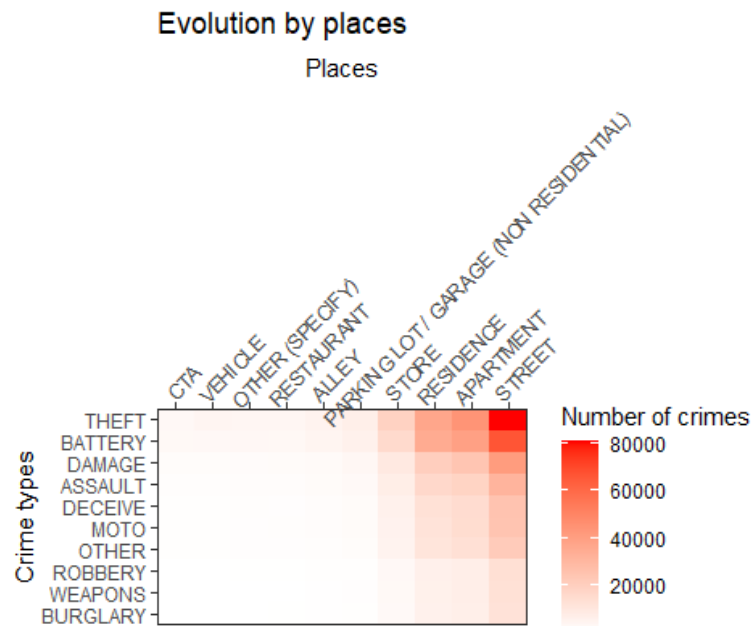
From the above heatmap, we can see that the highest number of thefts are in July.

Evolution of crime by Seasons

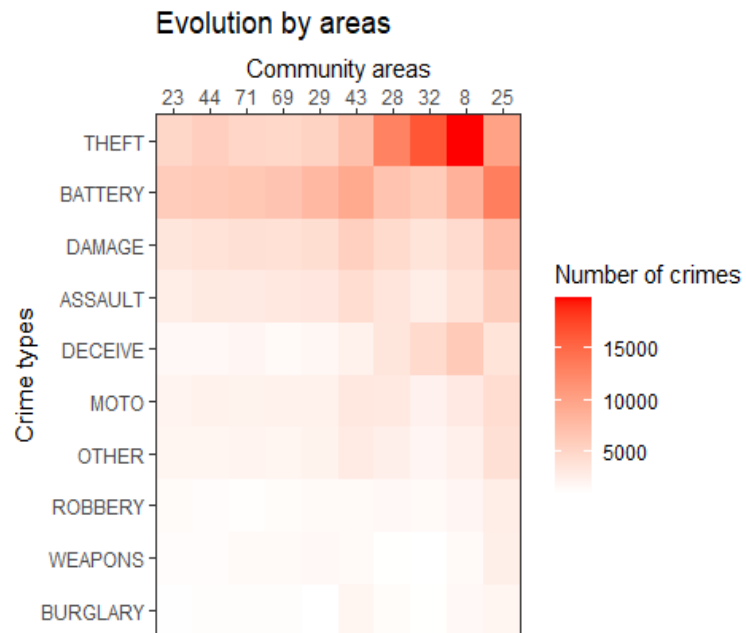


From the above heatmap, we can see that the highest number of thefts are in Fall

Top 10 most frequent places for Crimes

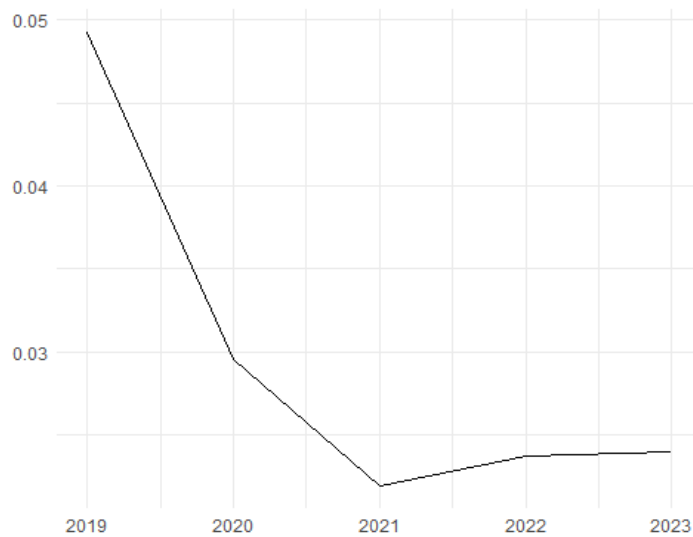


Top10 most dangerous community areas



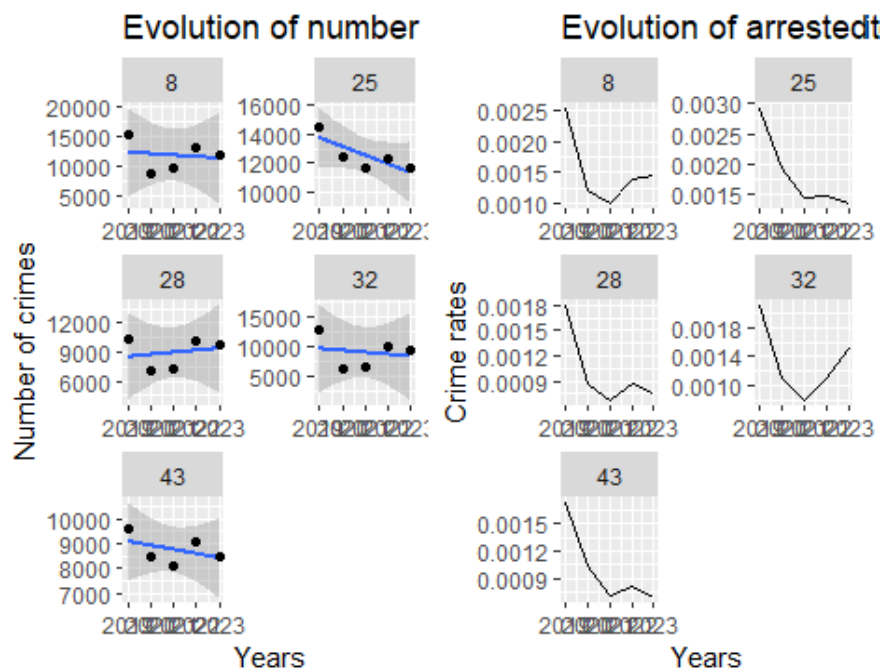
#Analysis of Arrest Rate

Arrest rate over years

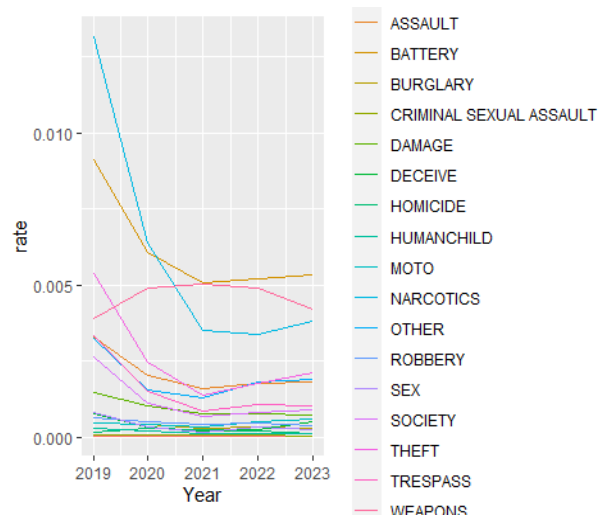


There is a steady decrease of arrest rate from 2019 to 2020 and after 2021 it slowly increased.

Number of arrests in top 5 dangerous areas and arrest rate in that area



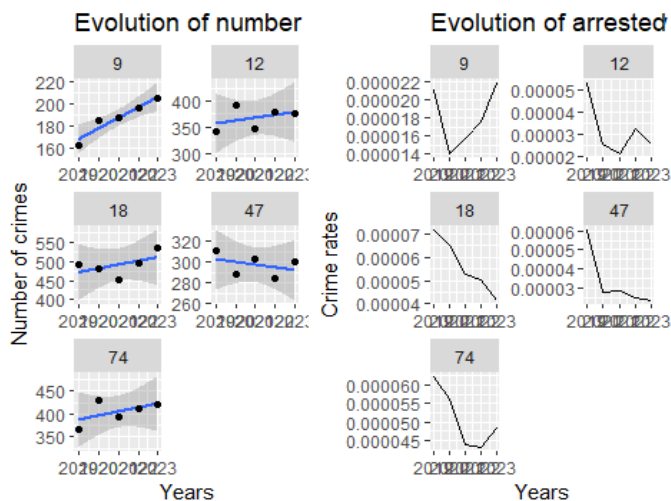
Arrest trend per crime



There is a steady decrease in the number of crimes even in the most dangerous communities. But there was also significant reduction in the arrest rate. This shows the police inefficiency.

Top 5 areas with least crime rate and arrest rate

"Evolution of arrested crime rates in different community areas over years")



From the above graph we can see that the crime rate is steadily decreasing in the top 5 safe areas. Except in 18 and 47 where we have steady decrease in the arrest rate, all the other areas have pretty good arrest rate.

From this we can conclude that The best areas to live in Chicago with good arrest rate 1. Edison Prak 2. Forest Glen 3. Mount Greenwood.

Conclusion:

we can conclude that theft is the most occurring crime where as it is decreasing steadily throughout the years.

Public transportation areas, downtown retail spaces remain hotspots for crime.

The most safer areas where crime is less and arrest rate is high are Edison Park, Forest Glen and Mount Greenwood.

Most dangerous neighborhood is Austin and Crime rate is high in Fall and on Fridays. This comprehensive analysis and modeling of Chicago crime data has provided crucial insights into the city's evolving crime profile, trends, patterns, and predictors over the past two decades.

The exploratory analysis illuminated key temporal and spatial trends across different crime types. Statistical modeling quantified influence of socioeconomic factors. Interactive visualizations mapped vulnerability across Chicago's landscape based on empirical data.

- Crime has declined, but further reduction requires localized prevention strategies beyond policing
- Public transportation areas, downtown retail spaces remain crime hotbeds needing focus
- Socioeconomic disparity translates strongly into crime rate differences across a few miles

However, significant scope remains for deeper community-level analysis and solution design. The models built are also limited by data available for the study period.

References:

<https://data.cityofchicago.org/Public-Safety/Violence-Reduction-Victims-of-Homicides-and-Non-Fa/gumc-mgzz>

<https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Community-Areas-current-/cauq-8yn6>

<https://data.cityofchicago.org/Public-Safety/Police-Stations-Map/gkur-vufi>

<https://wetlands.io/maps/Crime-Analysis-Using-R.html>