**Business Forecasting**

**Time Series Analysis – Crime Data(Cary,NC)**

The project aims at analyzing , forecasting overall crime rate as well as different categories of crimes. The dataset contains comprehensive Crime and Safety data from the Cary Police Department, covering reported incidents in the Town of Cary for the past 10 years and the current year. Extracted from the department's RMS application, the data includes information on Part I crimes such as arson, larceny, and aggravated assault, excluding sensitive cases like sexual assaults and juvenile-related incidents to protect victim identities. Utilizing the National Incident-Based Reporting System (NIBRS), the dataset provides dynamic, daily-updated records with details such as crime category, location, time, geographic coordinates, and neighborhood identifiers. This data supports the analysis of trends, seasonality, and spatial patterns in crime, aiding in the development of predictive models and community safety strategies.

The dataset contains detailed information on reported crimes, categorized by type, location, and time. It includes columns such as:

Date: The date of the occurrence.

Time: The time of the occurrence.

Latitude and Longitude: Geographical coordinates of the incident.

Offense Category: The type of crime, such as larceny, vandalism, or burglary.

Violent/Property Crime Indicator: Classification of crimes as violent or property-related.

Domestic: Indicates if the crime was domestic in nature.

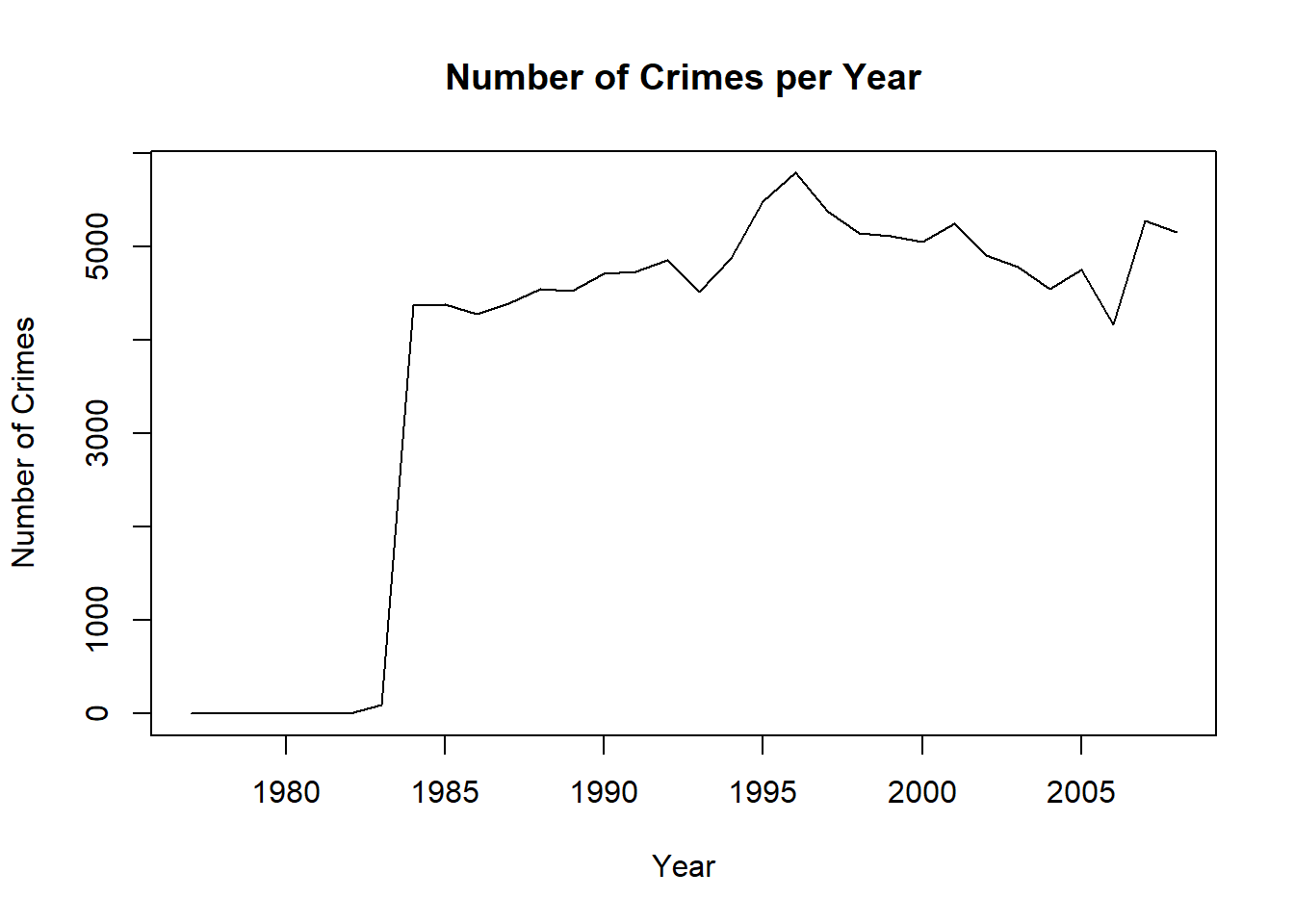
Total Incidents: Number of incidents reported for a specific category.

The data spans multiple years and includes seasonality, trend, and random components. By analyzing these elements, we aim to uncover patterns and develop reliable forecasting models to predict future crime occurrences.

Using advanced time-series methods such as ARIMA, STL decomposition, and Holt-Winters, the analysis aims to identify the most accurate model for predicting crime trends in various categories, thus enabling actionable insights.

Accurately forecasting crime trends is vital for law enforcement, policymakers, and communities. By understanding the future trajectory of specific crime categories, stakeholders can proactively address challenges, optimize resource allocation, and implement preventive measures. This analysis supports data-driven decision-making, enhancing public safety and fostering community well-being.

The timeseries after plotting has unnecesary data untill the year 2000, therefore we would not use the data untill 2000 in out analysis.



The timeseries after cutting the data now looks to be a lot more even with specifically the information we need which would help us in our forecasting.

A graph showing the number of crimes per year

Description automatically generated

Exploratory Data Analysis is crucial in understanding data, it would help us with better understanding the data we are going to work with. Below are a few EDA plots giving us an idea about the classification of crimes, details about each crime and their occurrences. A pie chart with numbers and a few pies

Description automatically generatedA graph of crime

Description automatically generatedA graph of a bar

Description automatically generated with medium confidenceA map of a crime density

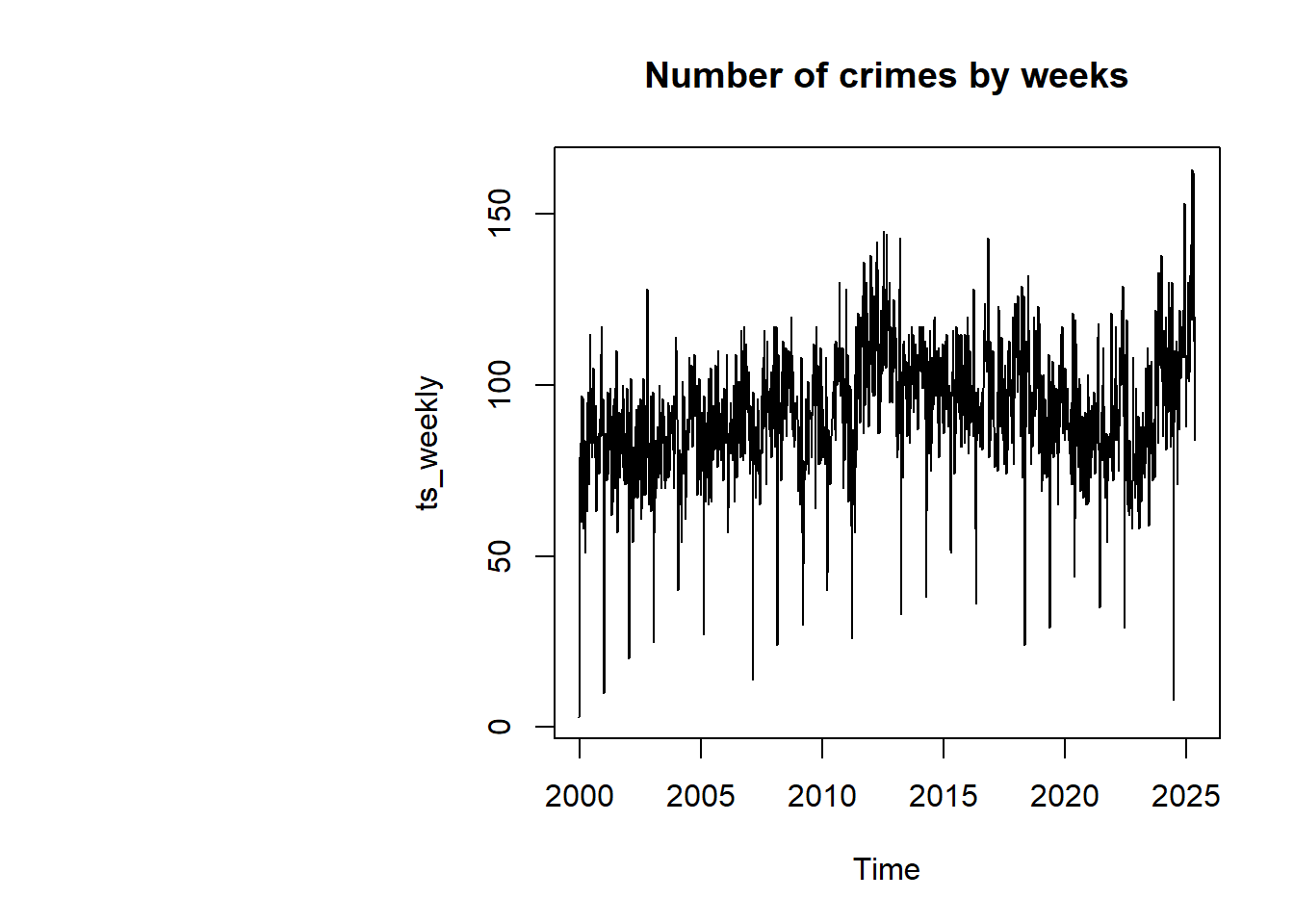
Description automatically generated

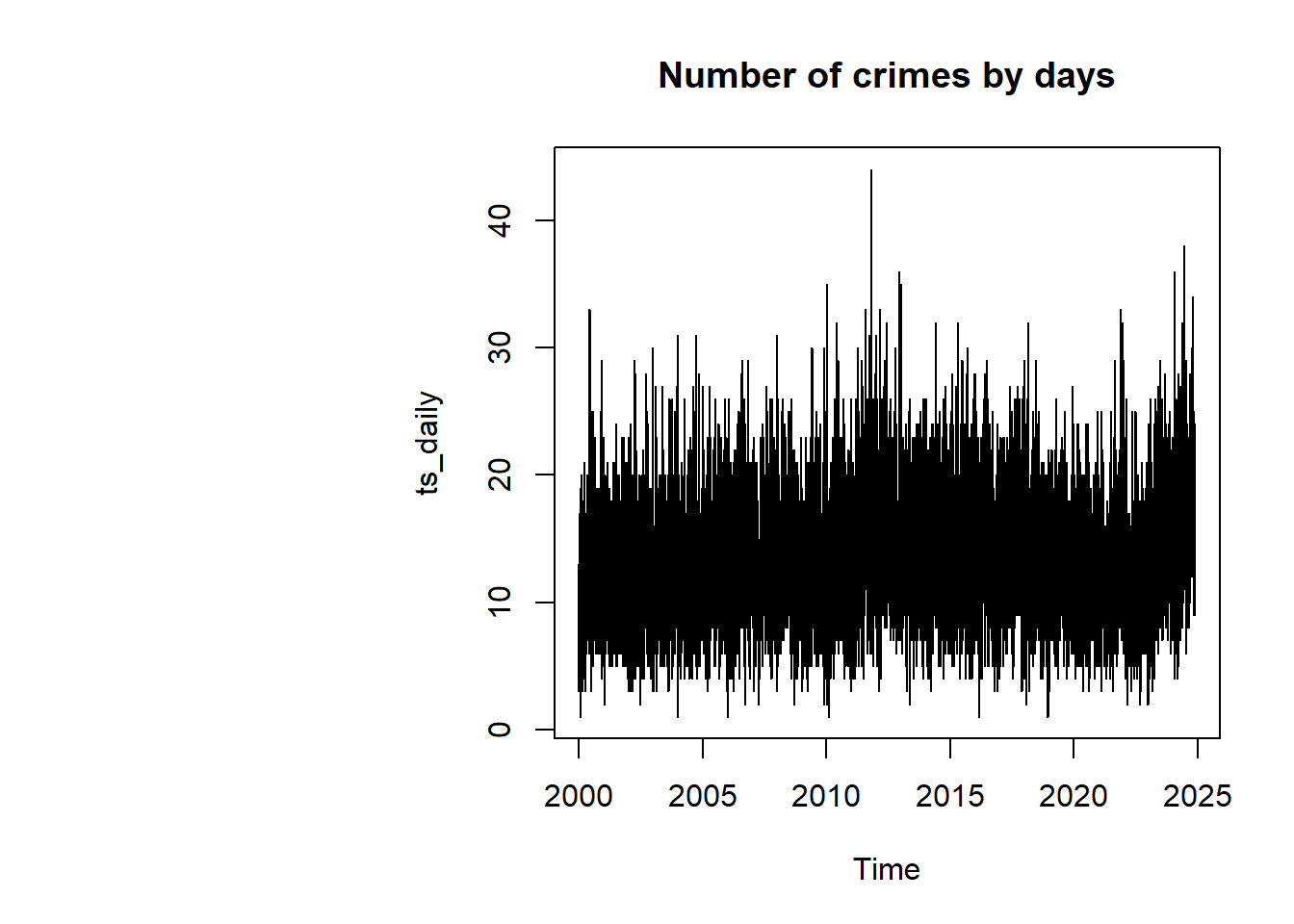
A graph with yellow and black squares

Description automatically generatedA graph with green and black squares

Description automatically generatedA graph showing a line of a line

Description automatically generatedA graph with numbers and lines

Description automatically generated



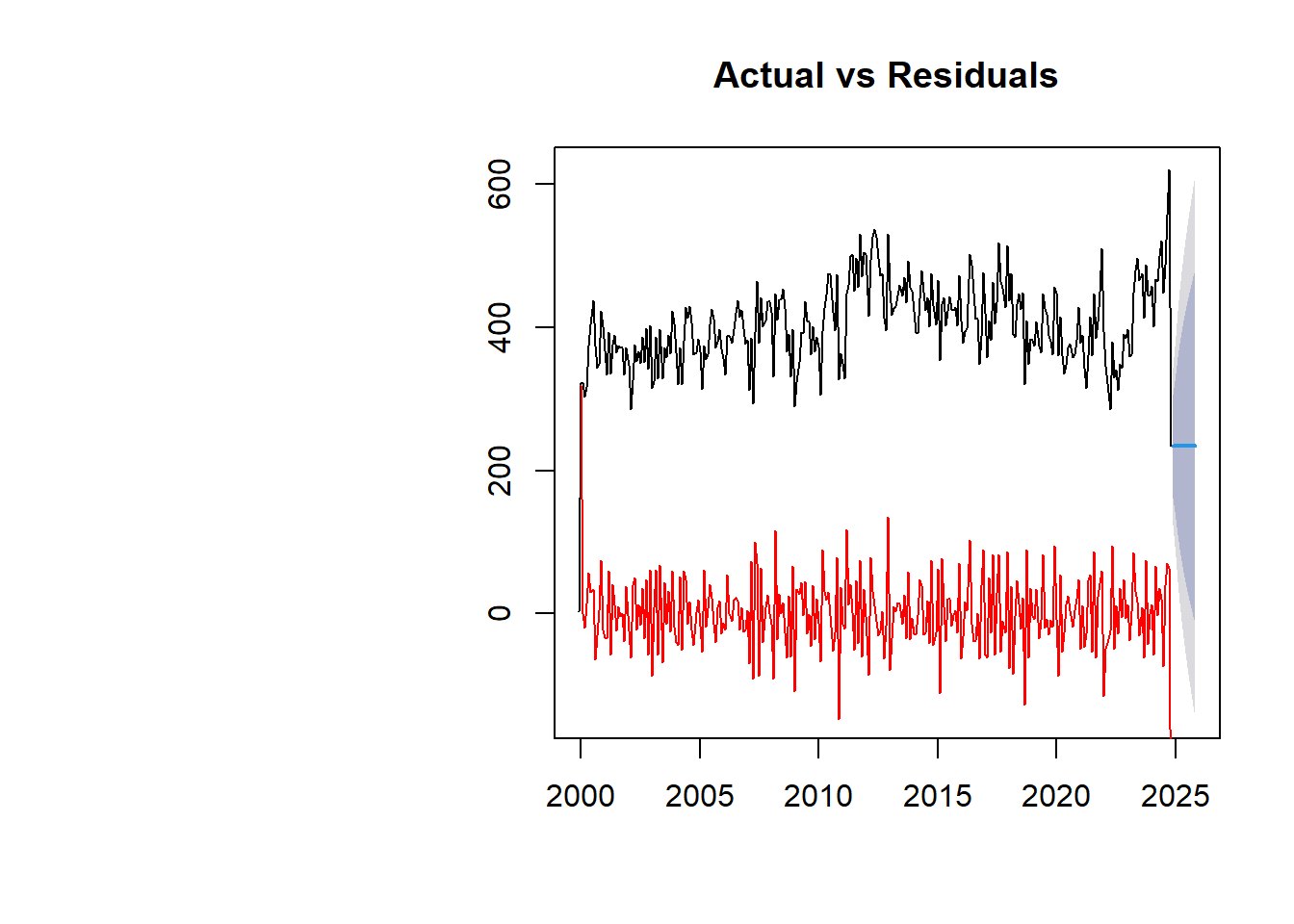
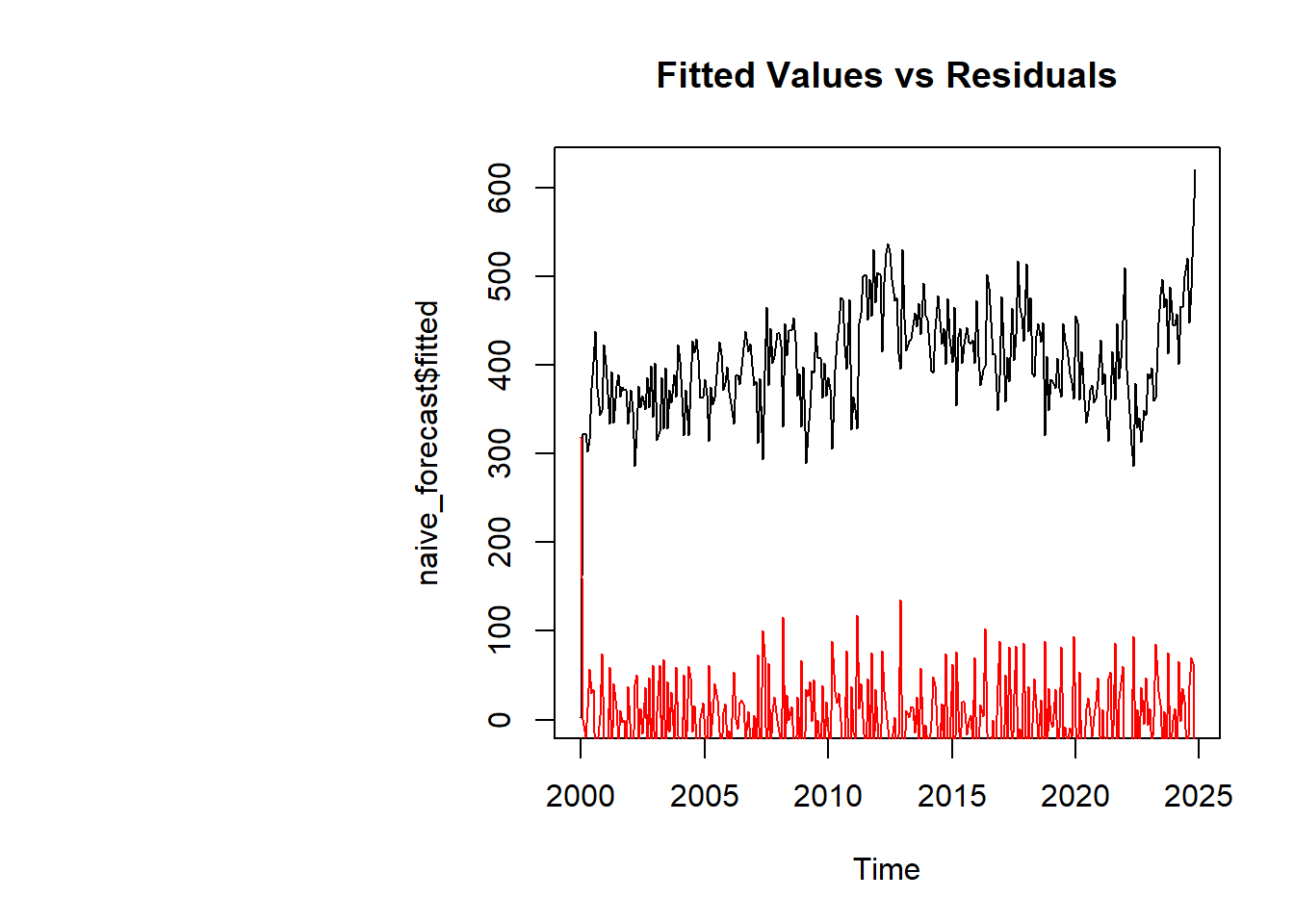
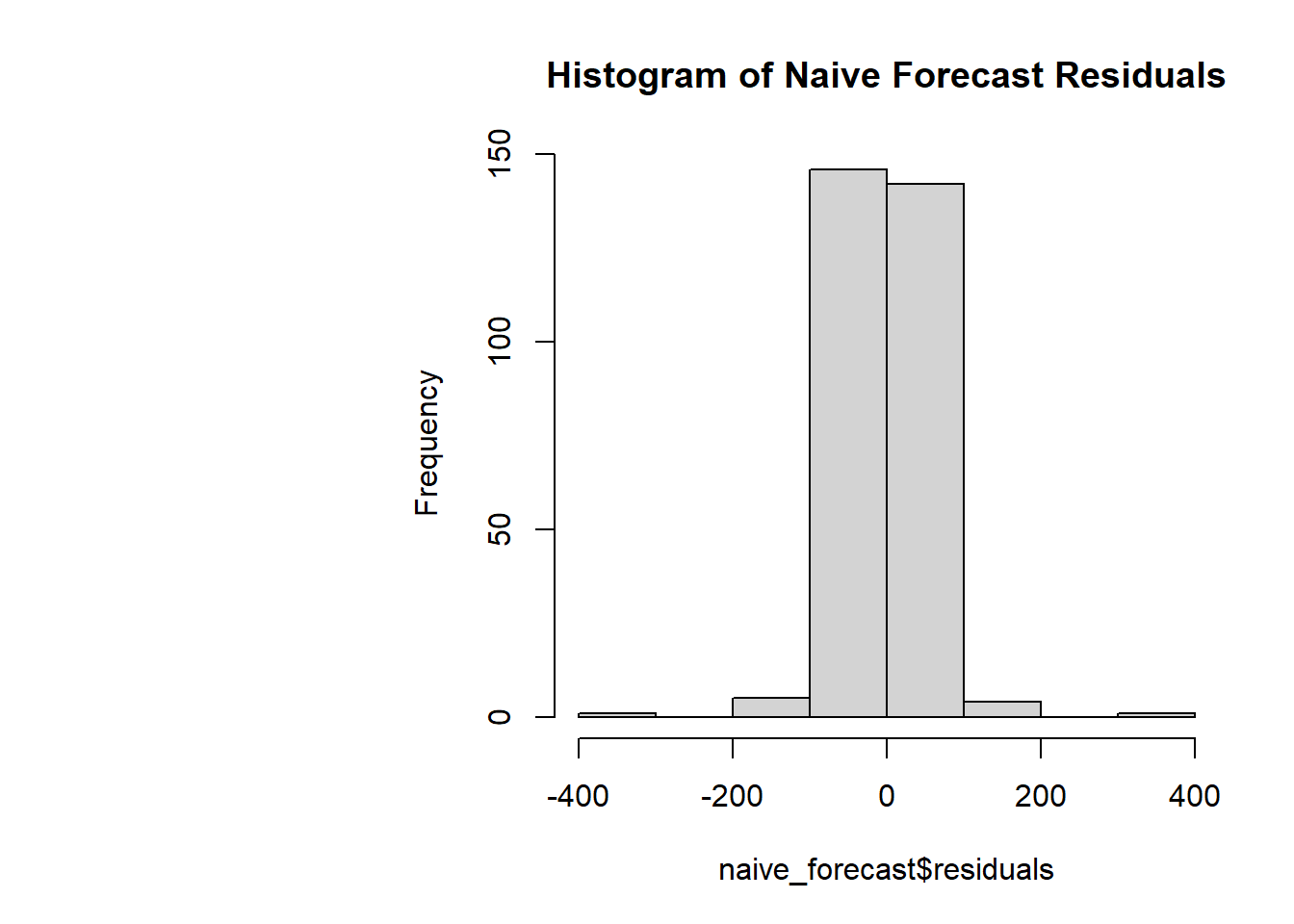
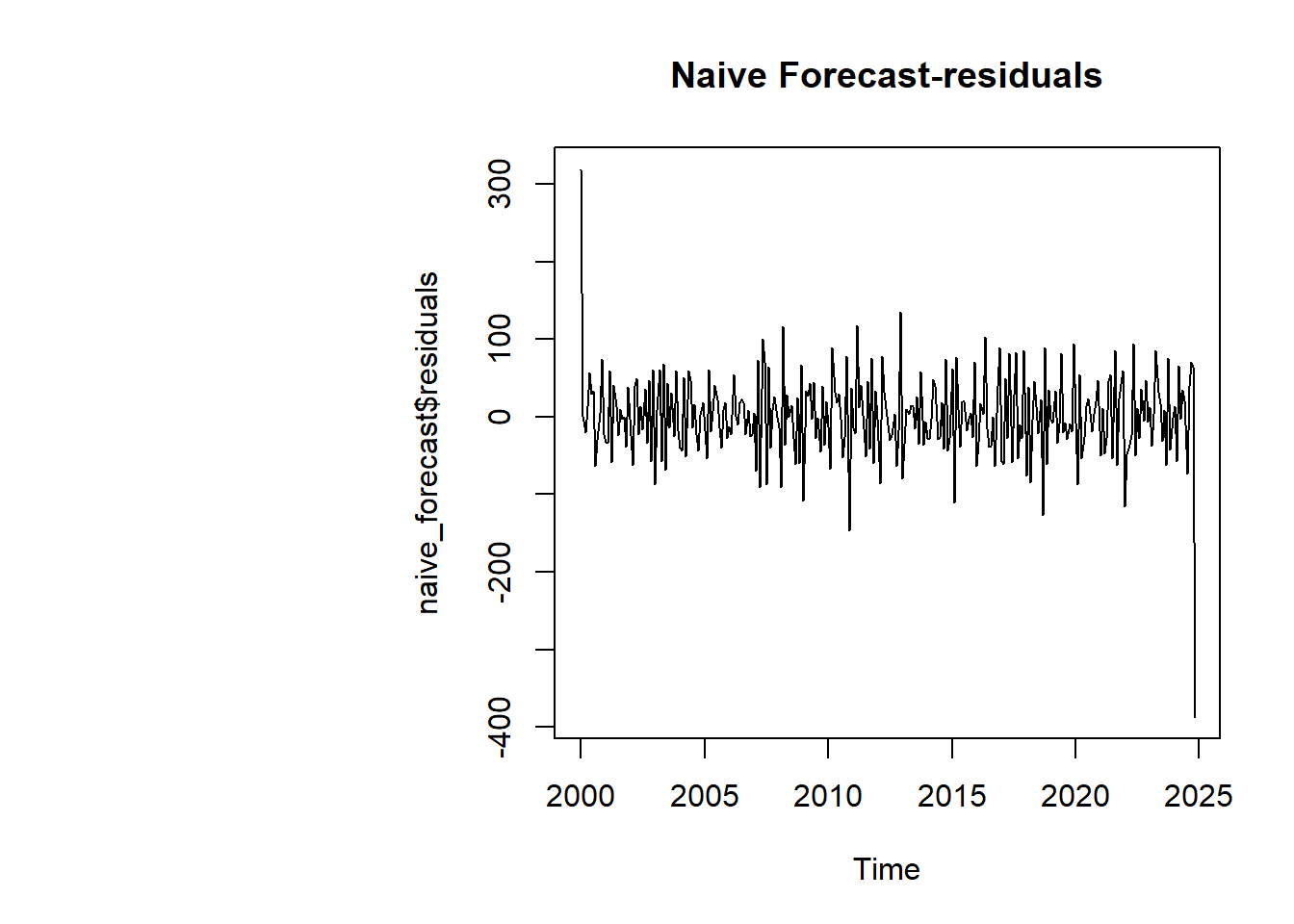
We would consider the monthly aggregate of crimes as this would give us a better forecasting. There appears to be seasonality in the time series as per the ACF plot, indicating the increase of crime in specific months.

We would start by the Naïve Forecast which assumes future values equal the last observed value.

A graph with numbers and lines

Description automatically generated

The residuals are randomly distributed, shows no pattern. This means that the Naive forecast is being able to capture the seasonality and trend. Since the histogram is normal, it means that the Naive Forecast might be a good method in this case.

A graph of a number of data

Description automatically generated with medium confidence

The Mean Forecast assumes future values equal the mean of past values. Observing the residuals, we can say the Mean Forecast does not effectively capture the pattern. Since the existence of patterns in residuals, it suggests the model’s limitations. A graph with numbers and lines

Description automatically generatedA graph of a graph

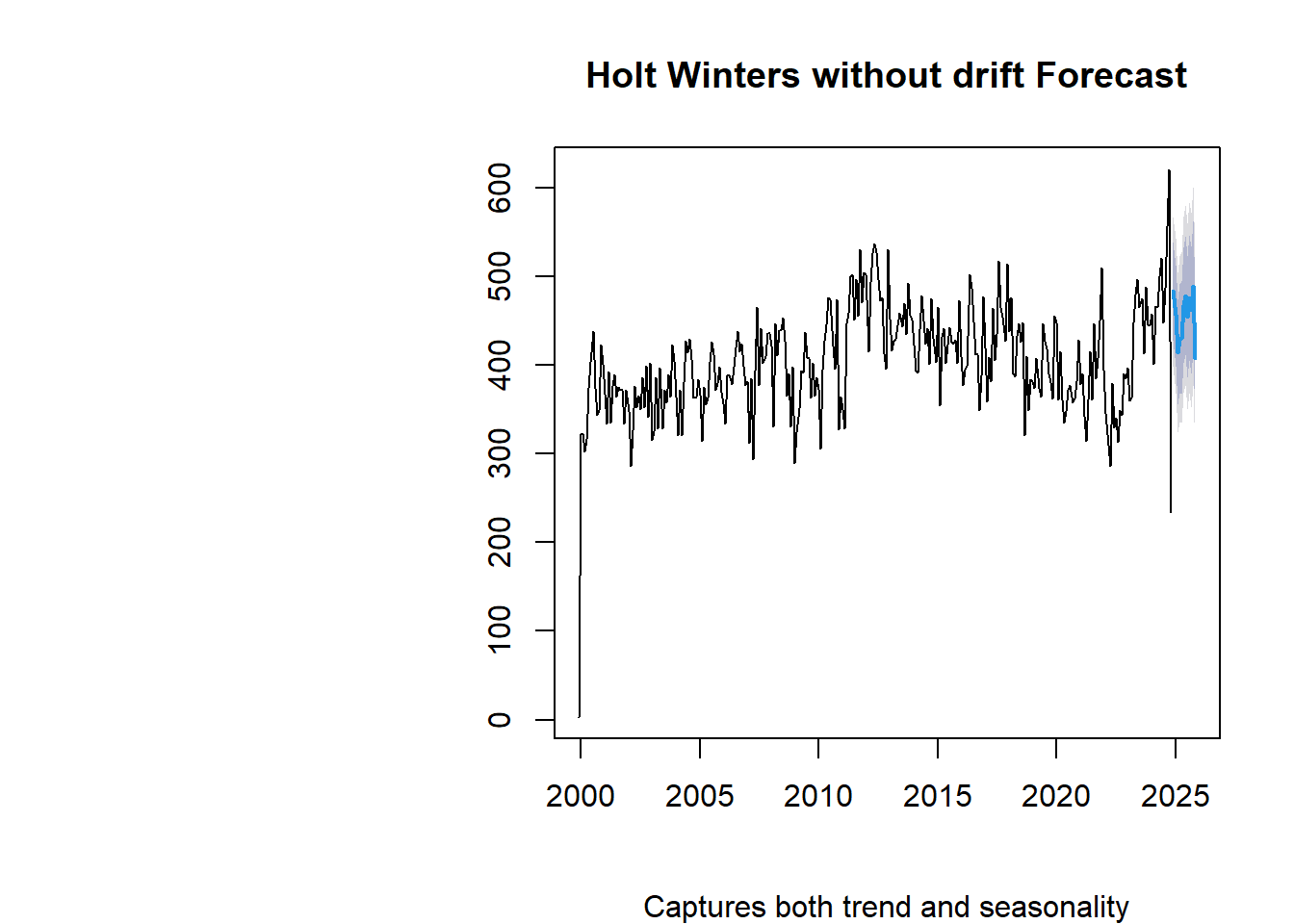
Description automatically generated

A graph with numbers and lines

Description automatically generatedA graph with numbers and lines

Description automatically generatedA graph of a number of percents

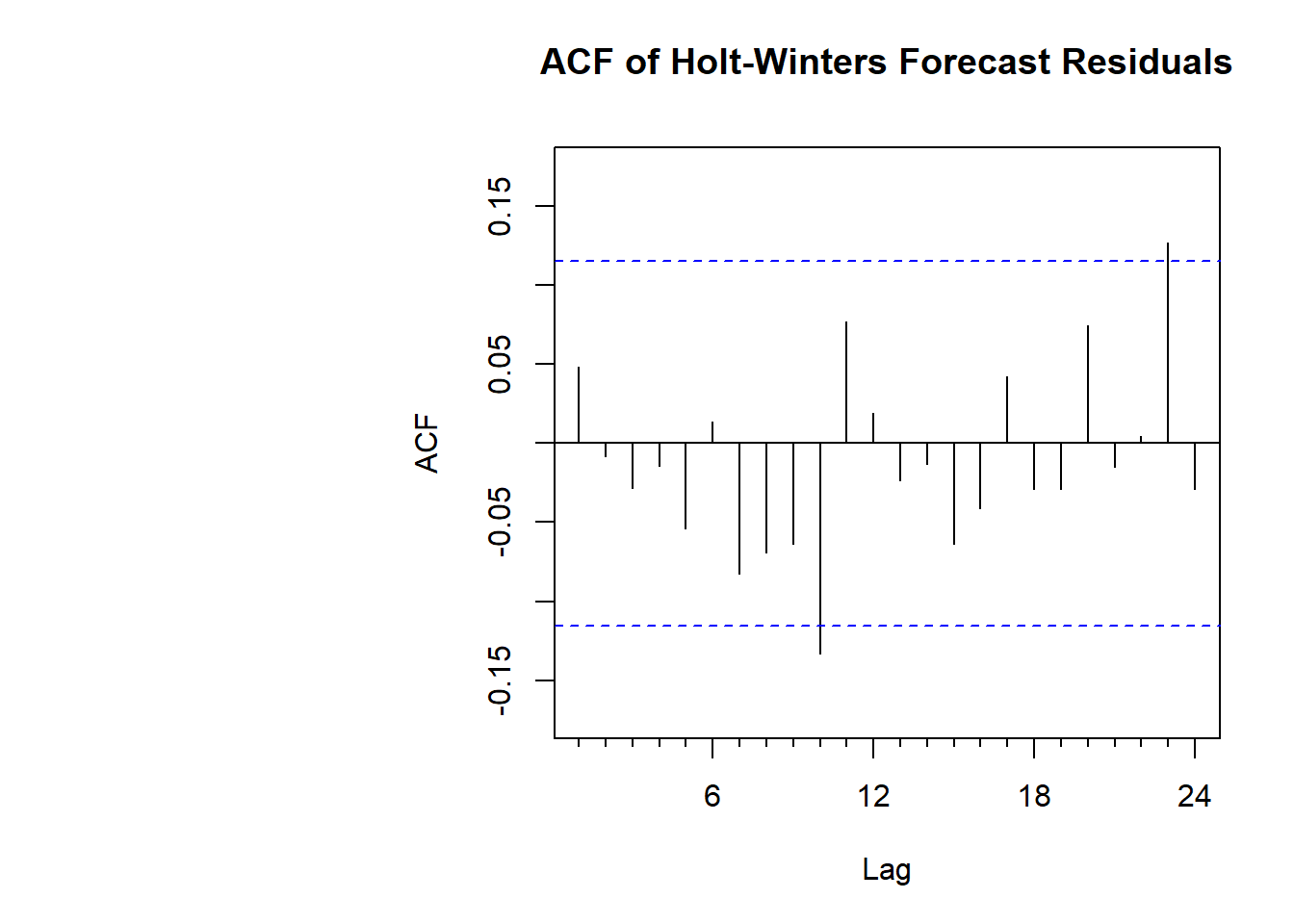
Description automatically generated with medium confidence

The Holt Winters Forecast captures both trend and seasonalityA graph with black lines

Description automatically generated

A graph of a number of data

Description automatically generated with medium confidenceA graph with numbers and lines

Description automatically generatedA graph with numbers and lines

Description automatically generatedA graph with black lines

Description automatically generatedA graph of a number of numbers

Description automatically generatedA graph with numbers and lines

Description automatically generatedA graph with numbers and lines

Description automatically generated Since both the models of Holt-Winters do not show patterns in its residuals, this suggests that the model is capturing the seasonality or trend.

SES can be effective because it uses a weighted average of past values, putting more weight on recent observations without explicitly modeling trend or seasonality. A graph with numbers and lines

Description automatically generated

The plot below is that of a mvoing averages with order 3,6,9. The moving average is similar to the mean forecast, only difference here is that the averages are a sliding window of the order mentiones. The moving average with an order 3 in this case seems to be a good fit.A graph with numbers and lines

Description automatically generated

A graph of different types of lines

Description automatically generated with medium confidence

Since there is a significant overlap between the original timeseries and seasonally independent timeseries, seasonality is not very pronounced. Factors like trend or randomness might have bigger influence on the data.A graph with red and black lines

Description automatically generatedA graph with numbers and lines

Description automatically generatedA graph of a graph of a graph

Description automatically generated with medium confidenceA graph of different types of crime

Description automatically generated with medium confidence

Making data stationary is essential for ARIMA because the model assumes consistent statistical properties (mean, variance, and autocorrelation) over time. Stationarity simplifies analysis by removing trends and seasonality, ensuring reliable parameter estimation and accurate forecasting. Without stationarity, ARIMA may produce spurious results. Techniques like differencing, log transformations, and detrending are commonly used to achieve stationarity, enabling the model to focus on underlying patterns and make robust predictions.

A graph with numbers and lines

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidenceA graph of crime rate forecast

Description automatically generatedA group of graphs showing the results of an experiment

Description automatically generatedA graph of a crime

Description automatically generatedThe ARIMA(2,1,1)(0,0,2)[12] model indicates a seasonal ARIMA with two autoregressive terms (AR), one differencing step (I), and one moving average term (MA), alongside two seasonal moving average terms (sMA) for monthly data. The coefficients provide weights for these components, with associated standard errors indicating the uncertainty of each estimate. Key model evaluation metrics include AIC (3170.87), AICc (3171.16), and BIC (3193.08), where lower values indicate better fit while penalizing complexity. Training set error measures such as RMSE (47.39) and MAE (34.60) highlight the average prediction errors. MAPE (8.85%) reflects a reasonably accurate model, with residuals showing minimal autocorrelation (ACF1 = 0.056), suggesting a good fit for the data with little remaining structure in the errors.

Overall, the forecast suggests a slight decline in crime rates in the near future, followed by minor fluctuations. The trend indicates that crime levels are likely to stabilize and remain close to the current levels by the end of the year.

The RMSE values from different forecasting methods highlight their predictive accuracy for the given time series data. Among the methods, the Moving Average with a window size of 3 (MA3) achieves the lowest RMSE (28.70), suggesting it provides the most accurate forecasts. The Holt-Winters (HW) method also performs well (RMSE: 42.97), capturing trends and seasonality effectively. In contrast, the Naive method has the highest RMSE (55.76), indicating less reliable predictions. These results emphasize the importance of selecting the most suitable method, like MA3, for forecasting in this context.

Below is an analysis, forecast of individual crime categories. A comprehensive analysis of residuals had been performed for each forecasting model, for each crime category. This would help in analysing the specific crime that needs to be paid more attention, the areas where the law enforcement is doing well.

A graph of different types of weather

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generatedA screenshot of a graph

Description automatically generatedA group of graphs showing the weather

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of different types of weather

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA group of graphs showing the results of a weather forecast

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA graph of different types of graphs

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA graph of different types of weather

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA group of graphs showing the results of a simple assault

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generatedA collage of graphs

Description automatically generatedA graph of different types of weather

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA group of graphs showing the results of a forecast

Description automatically generated with medium confidenceA group of graphs showing weather

Description automatically generated with medium confidenceA group of graphs showing different types of data

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of a graph of a graph

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of different types of weather

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA graph of different types of drugs

Description automatically generatedA graph of different types of drugs

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generatedA collage of different graphs

Description automatically generatedA graph of different types of drugs

Description automatically generatedA collage of different graphs

Description automatically generatedA graph of different types of graphs

Description automatically generated with medium confidenceA graph of a person

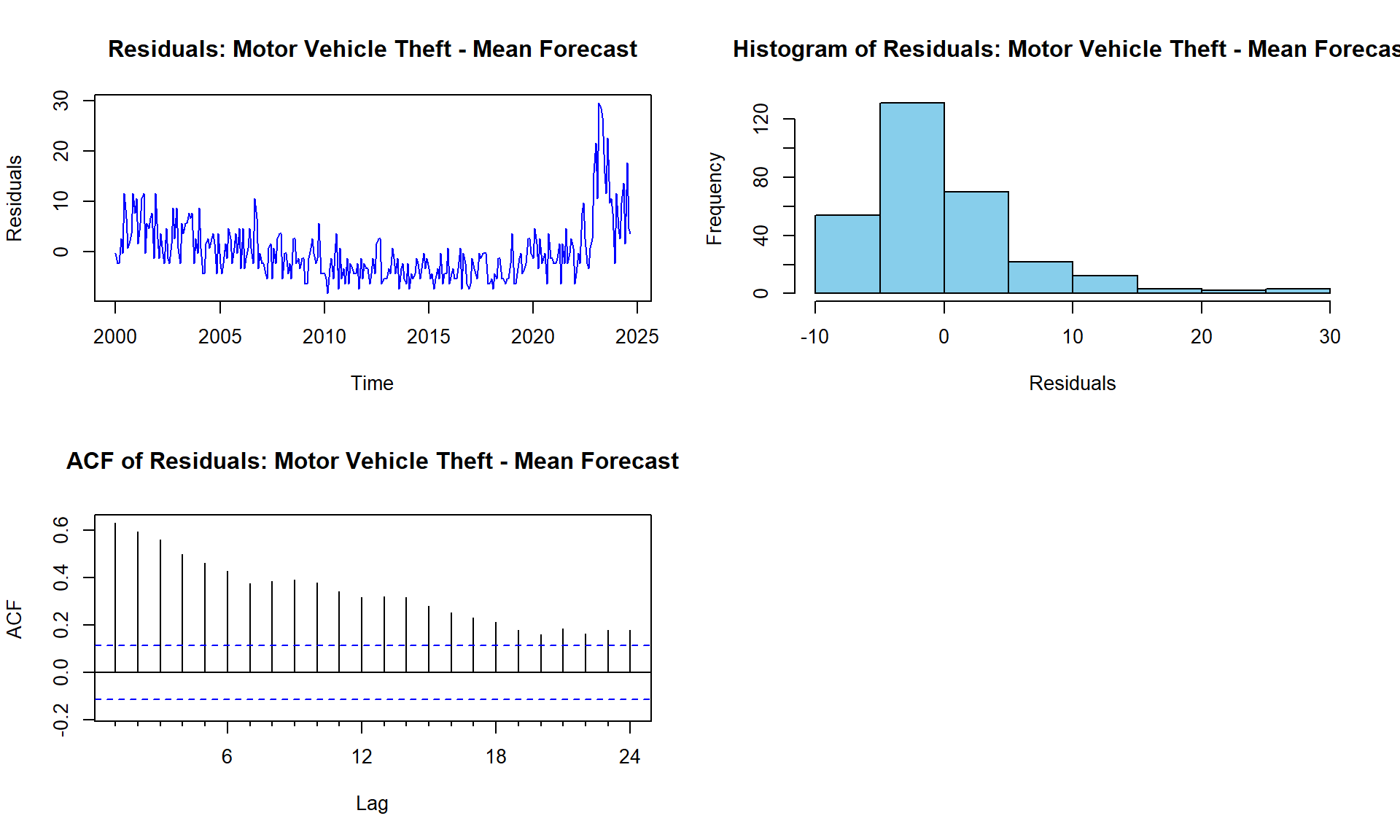
Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA graph of two people

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of a vehicle theft

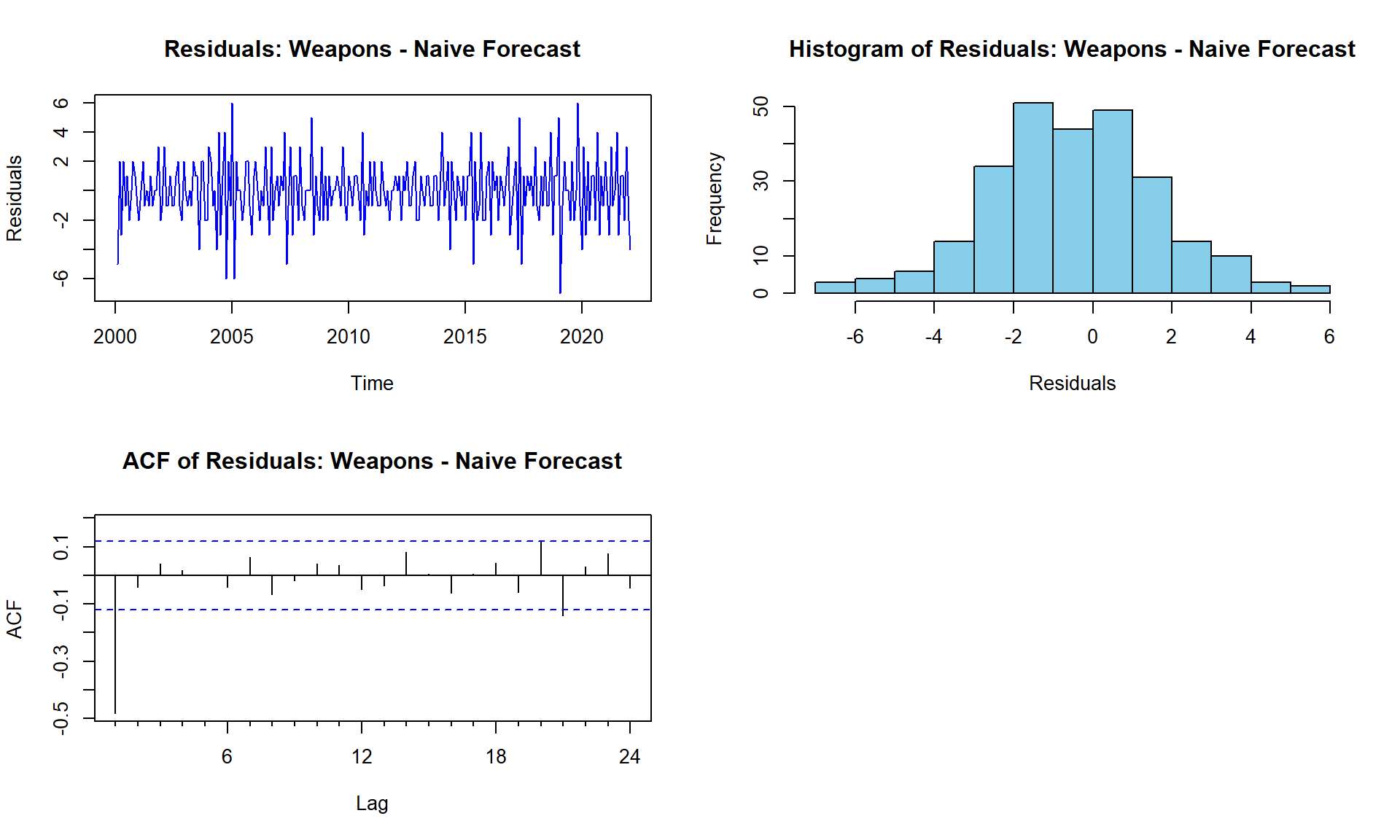
Description automatically generated with medium confidenceA graph of a vehicle

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generatedA screenshot of a graph

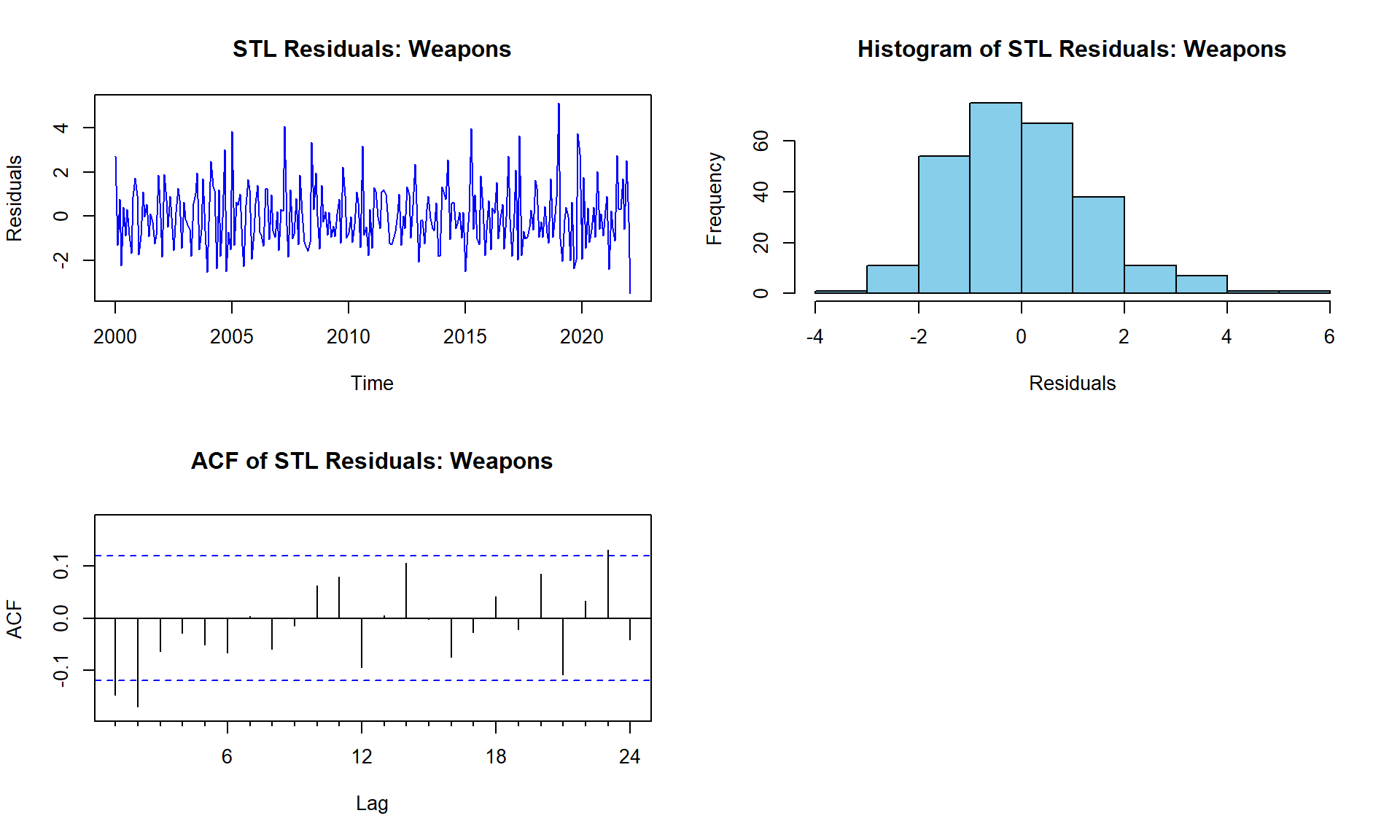
Description automatically generatedA screenshot of a graph

Description automatically generatedA graph of different weapons

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA collage of graphs and charts

Description automatically generatedA screenshot of a graph

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA collage of graphs

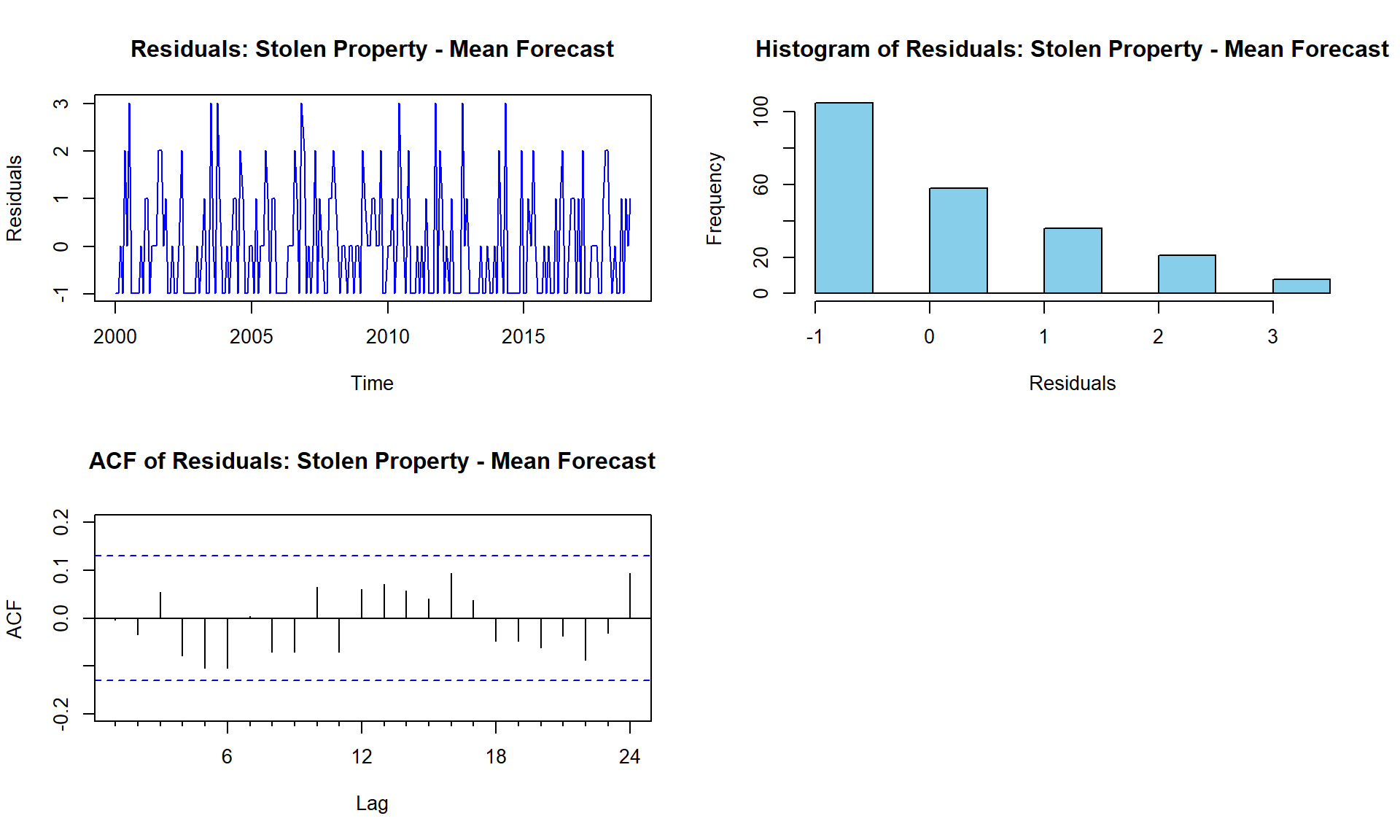
Description automatically generatedA graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generatedA collage of graphs and charts

Description automatically generatedA collage of graphs

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidenceA graph of a graph of a graph

Description automatically generated with medium confidenceA graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generatedA graph of a person and person

Description automatically generated with medium confidenceA collage of graphs

Description automatically generatedA graph of different types of alcohol

Description automatically generated with medium confidenceA graph of alcohol offense

Description automatically generated with medium confidenceA collage of graphs and charts

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidenceA graph of alcohol offense and alcohol offense

Description automatically generatedA graph of alcohol offense and histograms

Description automatically generatedA graph of different types of weather

Description automatically generatedA collage of graphs

Description automatically generatedA group of graphs showing the results of a forecast

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA collage of graphs

Description automatically generatedA collage of graphs and diagrams

Description automatically generatedA graph of different types of weather

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA graph of different types of graphs

Description automatically generated with medium confidenceA collage of graphs and diagrams

Description automatically generatedA graph of different types of graphs

Description automatically generated with medium confidenceA collage of different graphs

Description automatically generatedA graph of different types of waves

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA group of graphs showing the results of a forecast

Description automatically generated with medium confidenceA graph of different types of data

Description automatically generated with medium confidenceA screenshot of a graph

Description automatically generatedA graph of different types of data

Description automatically generated with medium confidence

Based on the individual forecasts for various crime categories, the analysis reveals distinct trends. Larceny, forgery, drugs, missing person, motor vehicle theft, and stolen property initially show a decrease, with stolen property subsequently increasing. Vandalism, burglary, fraud, weapons offenses, and suicide are projected to rise, albeit at varying rates. Simple assault, alcohol-related offenses, DUI, aggravated assault, arson, embezzlement, and other offenses are expected to remain stable, with minor fluctuations in some cases. Traffic violations, excluding DWI, are predicted to either decline slightly or remain steady. These insights provide a nuanced understanding of potential future trends in crime rates across different categories.

The RMSE values highlight the accuracy of different forecasting methods, with lower values indicating better performance. STL decomposition consistently achieves the lowest RMSE across most categories, making it the most reliable method, particularly for crimes like Larceny, Vandalism, and Forgery. ARIMA and Holt-Winters also perform well in capturing trends and seasonality, while Naive and Mean forecasts generally have higher RMSE, indicating lower accuracy. These comparisons guide the selection of effective models for forecasting crime trends.

Here is performing simple regression on latitude and longitude, since the dataset has a lot of categorical data, regression might not be a best strategy, however the linear regression model predicts the longitude (Lon) based on the latitude (Lat) of crime incidents. The scatter plot visualizes the relationship between Lat and Lon, with the regression line (in red) showing the best fit. The summary of the regression model includes coefficients, their significance, and overall model metrics, providing insights into the spatial distribution of crimes.

A blue and red line

Description automatically generated

The analysis and forecasting of crime data in Cary, NC, provide critical insights into crime trends and their spatial and temporal patterns. By leveraging advanced time series methods such as ARIMA, Holt-Winters, and STL decomposition, the project identified distinct trends across various crime categories, enabling accurate predictions and highlighting areas requiring focused attention. The results emphasize the effectiveness of models like STL decomposition and Moving Average in capturing trends and seasonality, while regression analysis offered a spatial perspective of crime incidents. The forecasts indicate stabilization in overall crime rates, with variations in specific categories like larceny, vandalism, and burglary. These findings support data-driven decision-making for law enforcement, policymakers, and community stakeholders, facilitating resource optimization and targeted preventive measures to enhance public safety and community well-being.