Unveiling the Rat Saga in the busy streets of NYC: A Data-Driven Exploration

Saichand Ghantasala

2024-02-24

Introduction:

New York City, a vibrant metropolis, is home not only to millions of people but also to a notorious urban wildlife resident – rats. For years, the city has grappled with the challenge of coexisting with these creatures, leading to detailed records of rat sightings. In this exploration, we delve into the dataset spanning from 2010 to September 16th, 2017, uncovering the intriguing patterns of rat encounters across the boroughs and communities of NYC.

```
# loading required libraries
library(tidyverse)
library(ggplot2)
library(lubridate)

# Read the data
rat_sightings <- read_csv("data/A1_sightings.csv", na = c("", "NA", "N/A"))
```

Our journey begins with the meticulous task of summarizing the extensive dataset. Using R and tools like dplyr, we aggregated the data, focusing on key variables such as borough, year, and type of structure. This step allowed us to distill the raw information into meaningful insights, providing a foundation for the visual storytelling that follows.

```
# Parse the "Created Date" column to datetime format
rat_sightings$Created_Date <- mdy_hms(rat_sightings$`Created Date`)

# Date-related variables
rat_sightings <- rat_sightings %>%
  mutate(
    sighting_year = year(Created_Date),
    sighting_month = month(Created_Date),
    sighting_day = day(Created_Date),
    sighting_weekday = wday(Created_Date, label = TRUE)
  )
```

```
# Removing duplicate rows
rat_sightings <- rat_sightings %>% distinct()

# Checking for missing values in key columns
sum(is.na(rat_sightings$sighting_year))
```

```
## [1] 0
```

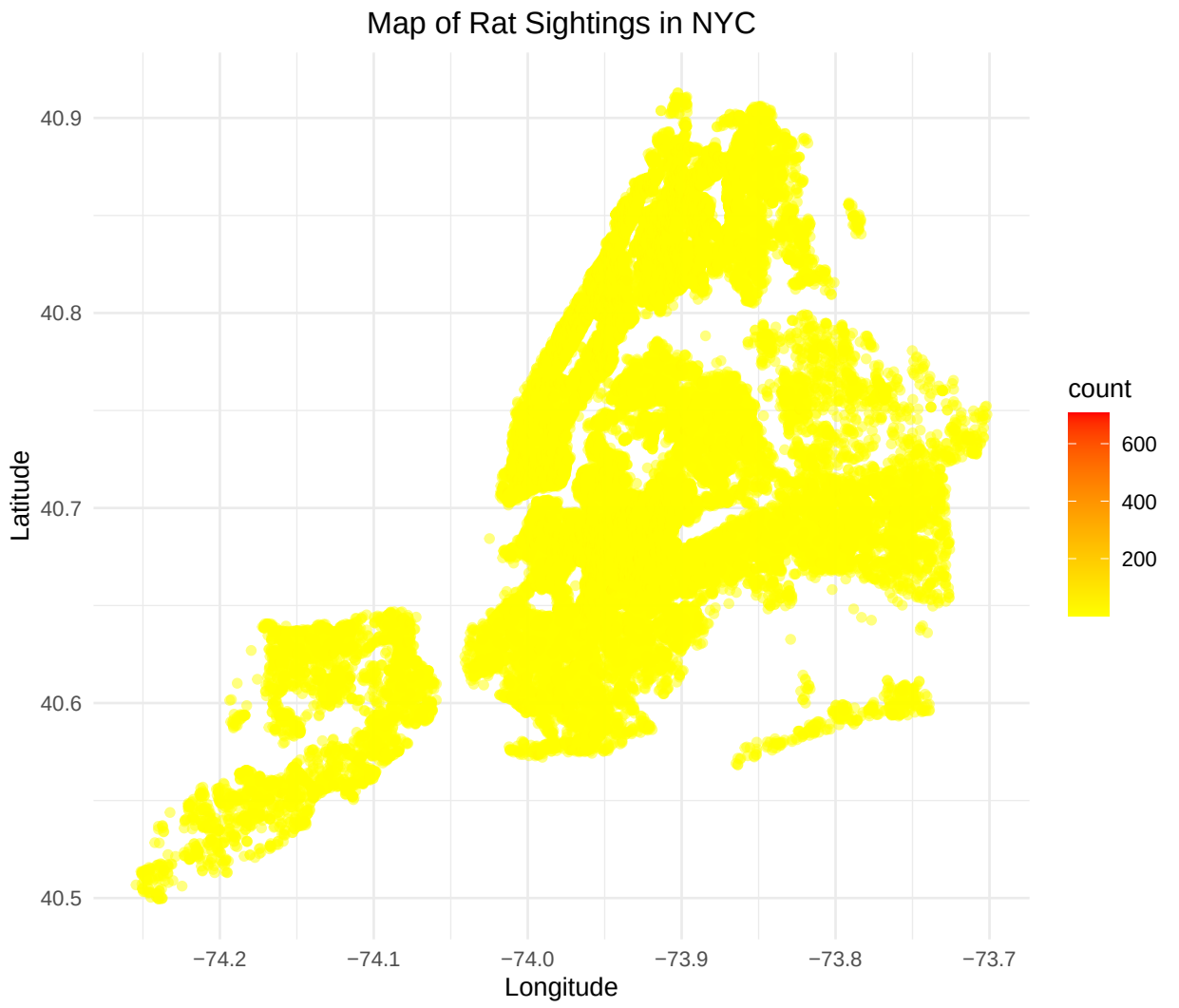
```
sum(is.na(rat_sightings$Borough))
```

```
## [1] 0
```

Visualizing Rat Hotspots: A Map of NYC

To set the stage, we created a visual representation of rat sightings across the city using geographical coordinates. The resulting map highlights the hotspots, revealing areas where rat encounters are more prevalent. The color gradient, ranging from yellow to red, adds depth to the visualization, emphasizing the intensity of sightings at different locations.

```
library(ggmap)

# Summarize the number of sightings at each coordinate
sightings_by_location <- rat_sightings %>%
  group_by(Longitude, Latitude) %>%
  summarize(count = n(), .groups = "drop") %>%
  filter(count > 0)

# Plotting the map
ggplot(sightings_by_location, aes(x = Longitude, y = Latitude, color = count)) +
  geom_point(alpha = 0.5) +
  scale_color_gradient(low = "yellow", high = "red") +
  theme_minimal() +
  labs(title = "Map of Rat Sightings in NYC",
       x = "Longitude",
       y = "Latitude") +
  theme(plot.title = element_text(hjust = 0.5))
```




```
# Save the plot as pdf
ggsave("map_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("map_plot.png", width = 11, height = 8.5, unit = "in")
```

Unraveling Temporal Trends

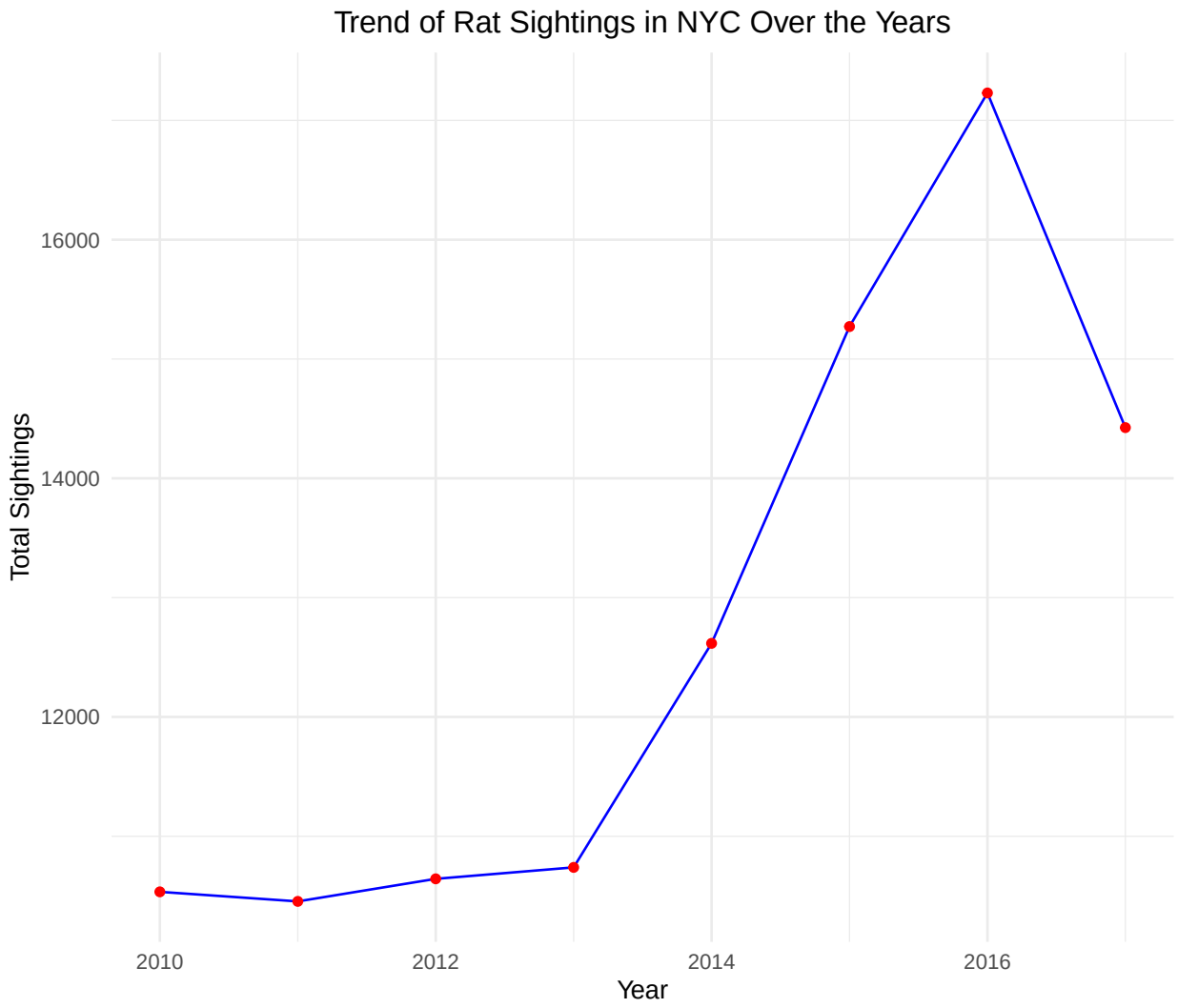
Yearly Evolution of Rat Sightings

Our exploration extends over time, with a focus on the yearly trends of rat sightings. The line chart illustrates the ebb and flow of encounters, offering insights into whether the rat population has undergone significant changes. Notably, we consider the 2014 study, which provided a more accurate estimate of the rat-human ratio, to discern any alignment with the dataset trends.

```
# Summarize the number of sightings per year
yearly_sightings <- rat_sightings %>%
```

```
  group_by(sighting_year) %>%
  summarize(total_sightings = n())
```

```
ggplot(yearly_sightings, aes(x = sighting_year, y = total_sightings)) +
  geom_line(group=1, color="blue") +
  geom_point(color="red") +
  theme_minimal() +
  labs(title = "Trend of Rat Sightings in NYC Over the Years",
       x = "Year",
       y = "Total Sightings") +
theme(plot.title = element_text(hjust = 0.5))
```

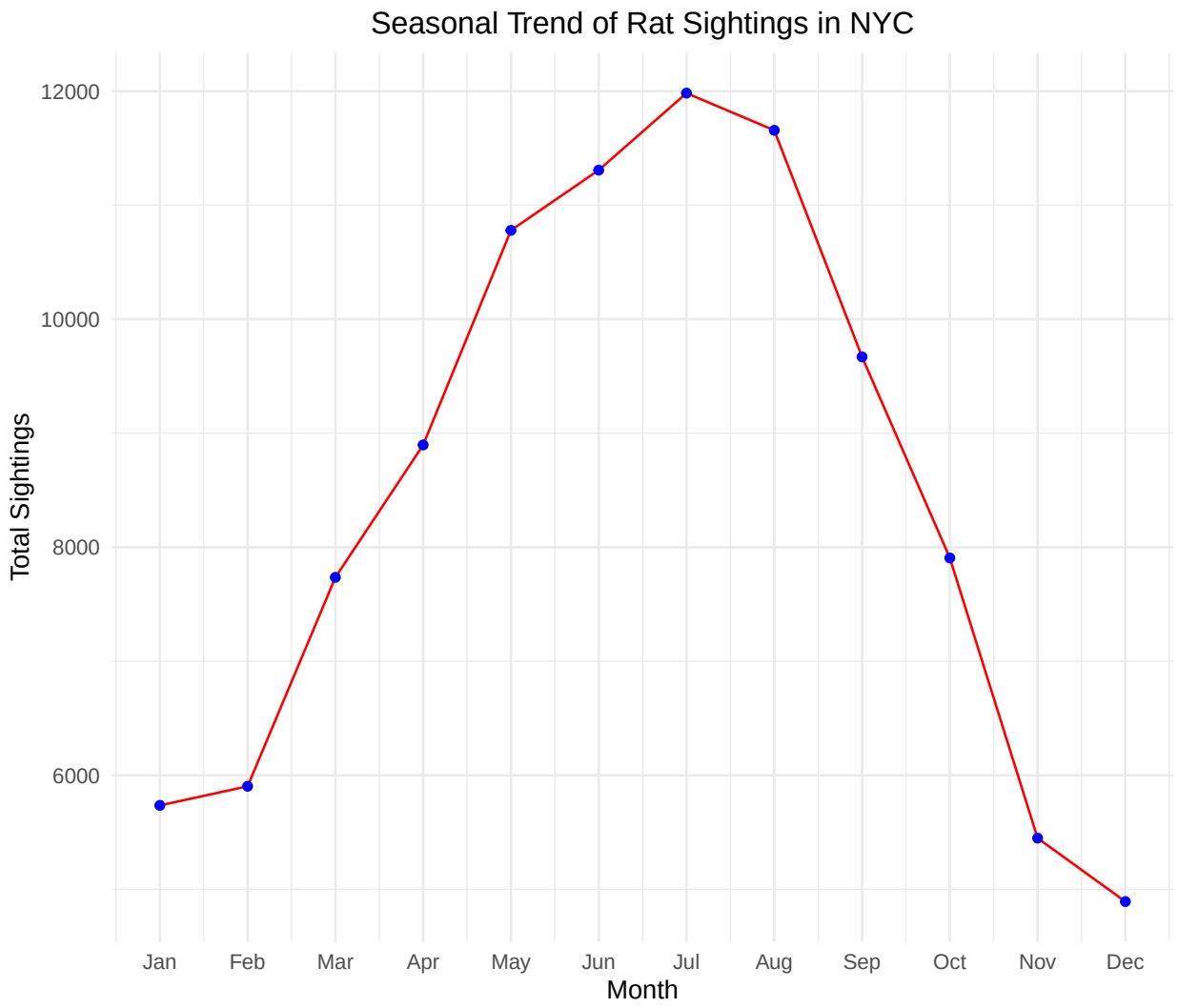



```
# Save the plot as pdf
ggsave("yearly_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("yearly_plot.png", width = 11, height = 8.5, unit = "in")
```

Seasonal Insights into Rat Activity

Seasonal patterns come to the forefront as we examine the monthly and weekly trends of rat sightings. The line chart portrays the cyclical nature of encounters, prompting us to explore the influence of external factors such as weather conditions or city events on rat activity.

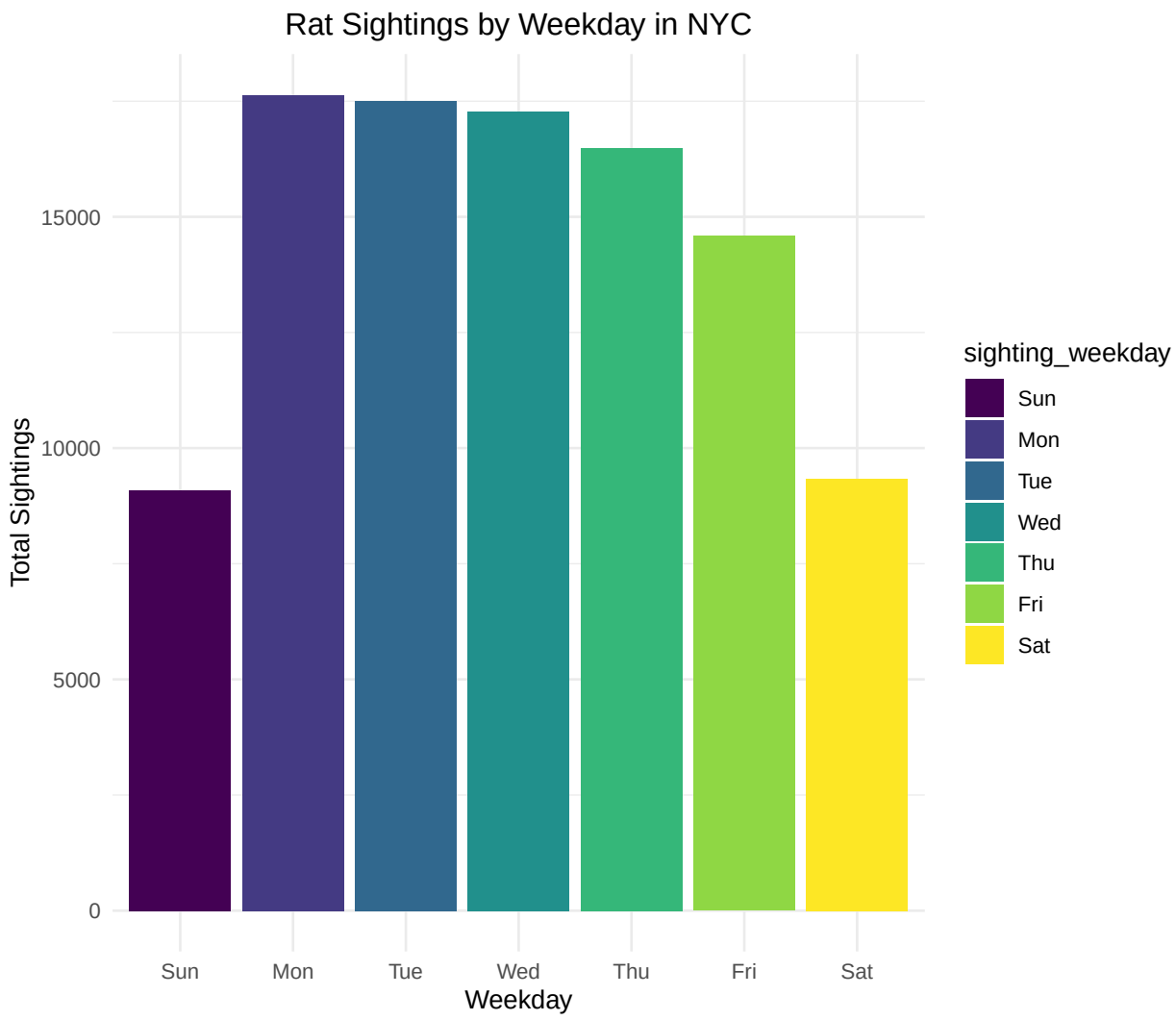
```
# Summarize sightings by month
monthly_sightings <- rat_sightings %>%
  group_by(sighting_month) %>%
  summarize(total_sightings = n())

# Line chart of sightings by month
ggplot(monthly_sightings, aes(x = sighting_month, y = total_sightings)) +
  geom_line(color="red") +
  geom_point(color="blue") +
  scale_x_continuous(breaks = 1:12, labels = month.abb) +
  theme_minimal() +
  labs(title = "Seasonal Trend of Rat Sightings in NYC",
       x = "Month",
       y = "Total Sightings") +
  theme(plot.title = element_text(hjust = 0.5))
```




```r
# Save the plot as pdf
ggsave("monthly_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("monthly_plot.png", width = 11, height = 8.5, unit = "in")
```

```r
# Summarize sightings by weekday
weekday_sightings <- rat_sightings %>%
  group_by(sighting_weekday) %>%
  summarize(total_sightings = n())

# Bar chart of sightings by weekday
ggplot(weekday_sightings, aes(x = sighting_weekday,
                              y = total_sightings, fill = sighting_weekday)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Rat Sightings by Weekday in NYC",
       x = "Weekday",
       y = "Total Sightings") +
```

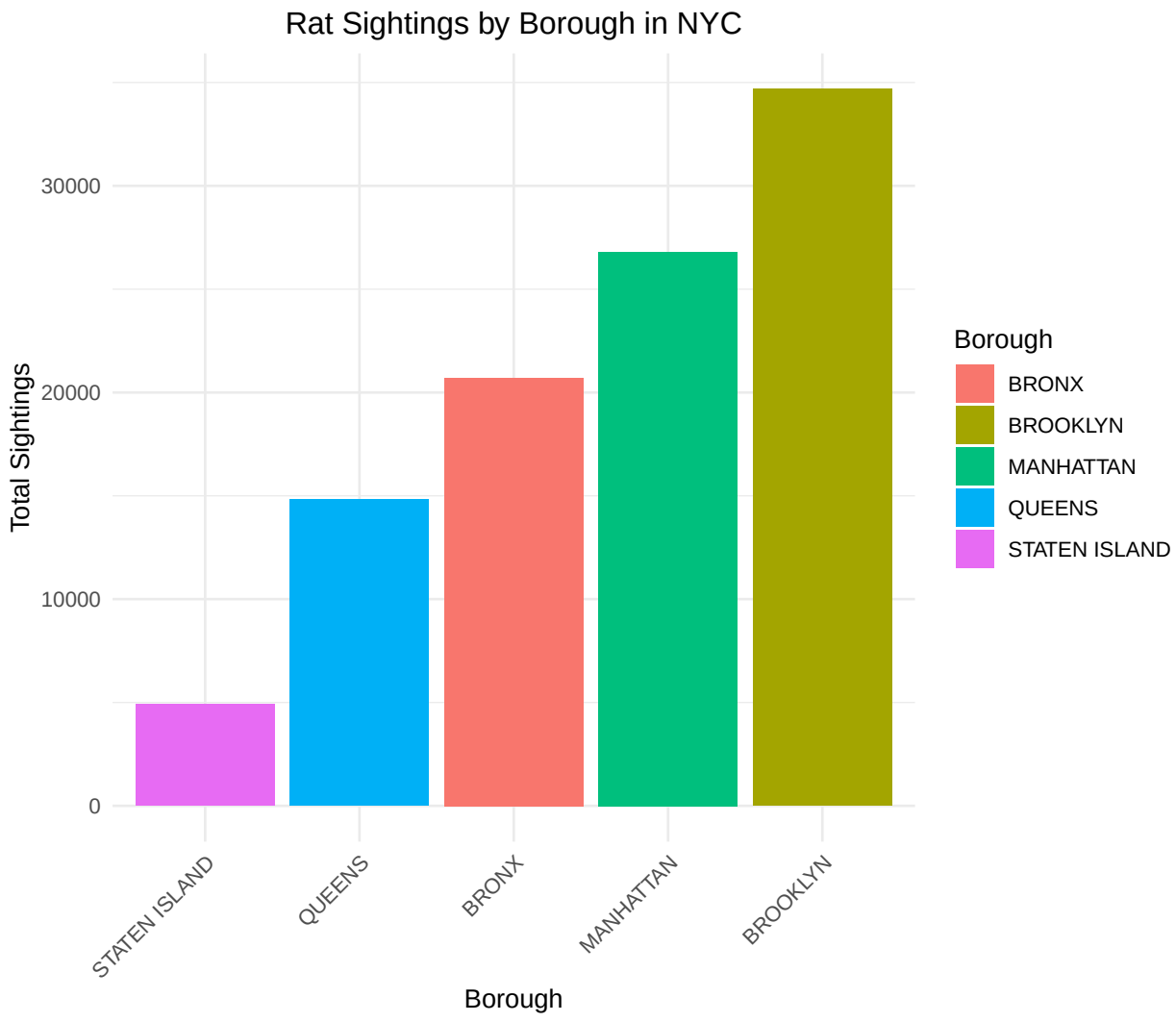
```
  theme(plot.title = element_text(hjust = 0.5))
```




```
# Save the plot as pdf
ggsave("weekly_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("weekly_plot.png", width = 11, height = 8.5, unit = "in")
```

```
# Sightings trend in different borough
borough_sightings <- rat_sightings %>%
  filter(Borough != "Unspecified") %>%
  group_by(Borough) %>%
  summarize(total_sightings = n()) %>%
  arrange(desc(total_sightings))

ggplot(borough_sightings, aes(x = reorder(Borough, total_sightings),
                              y = total_sightings, fill = Borough)) +
```

```
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Rat Sightings by Borough in NYC",
       x = "Borough",
       y = "Total Sightings") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.title = element_text(hjust = 0.5))
```




```
# Save the plot as pdf
ggsave("bar_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("bar_plot.png", width = 11, height = 8.5, unit = "in")
```

Predictive Analysis: Can Rats be Forecasted?

Venturing into the realm of predictive analytics, we seek to answer the intriguing question of whether rat sightings can be forecasted based on historical data. Utilizing time series forecasting models, we project future encounters, offering a glimpse into the potential predictability of rat activity in the city.

```
library(forecast)
library(dplyr)
library(tidyr)
library(purrr)

# Creating a time series object for each borough
borough_ts_list <- rat_sightings %>%
  filter(Borough != "Unspecified") %>%
  group_by(Borough, sighting_year) %>%
  summarize(total_sightings = n(), .groups = "drop") %>%
  group_by(Borough) %>%
  complete(sighting_year = full_seq(sighting_year, period = 1)) %>%
  arrange(Borough, sighting_year) %>%
  group_split(Borough)

# Predictive model for each borough
predictions_list <- borough_ts_list %>%
  map_df(~{
    .x <- .x %>% replace_na(list(total_sightings = 0))

    # Time series object
    ts_data <- ts(.x$total_sightings, start = min(.x$sighting_year), frequency = 1)

    # Model Fitting
    fit <- auto.arima(ts_data)

    # Forecasting for h = 2 years
    future <- forecast(fit, h = 2)


    data.frame(
      Borough = unique(.x$Borough),
      Year = as.numeric(time(future$mean)),
      Forecast = as.numeric(future$mean)
    )
  })

# Forecasted values
print(predictions_list)
```

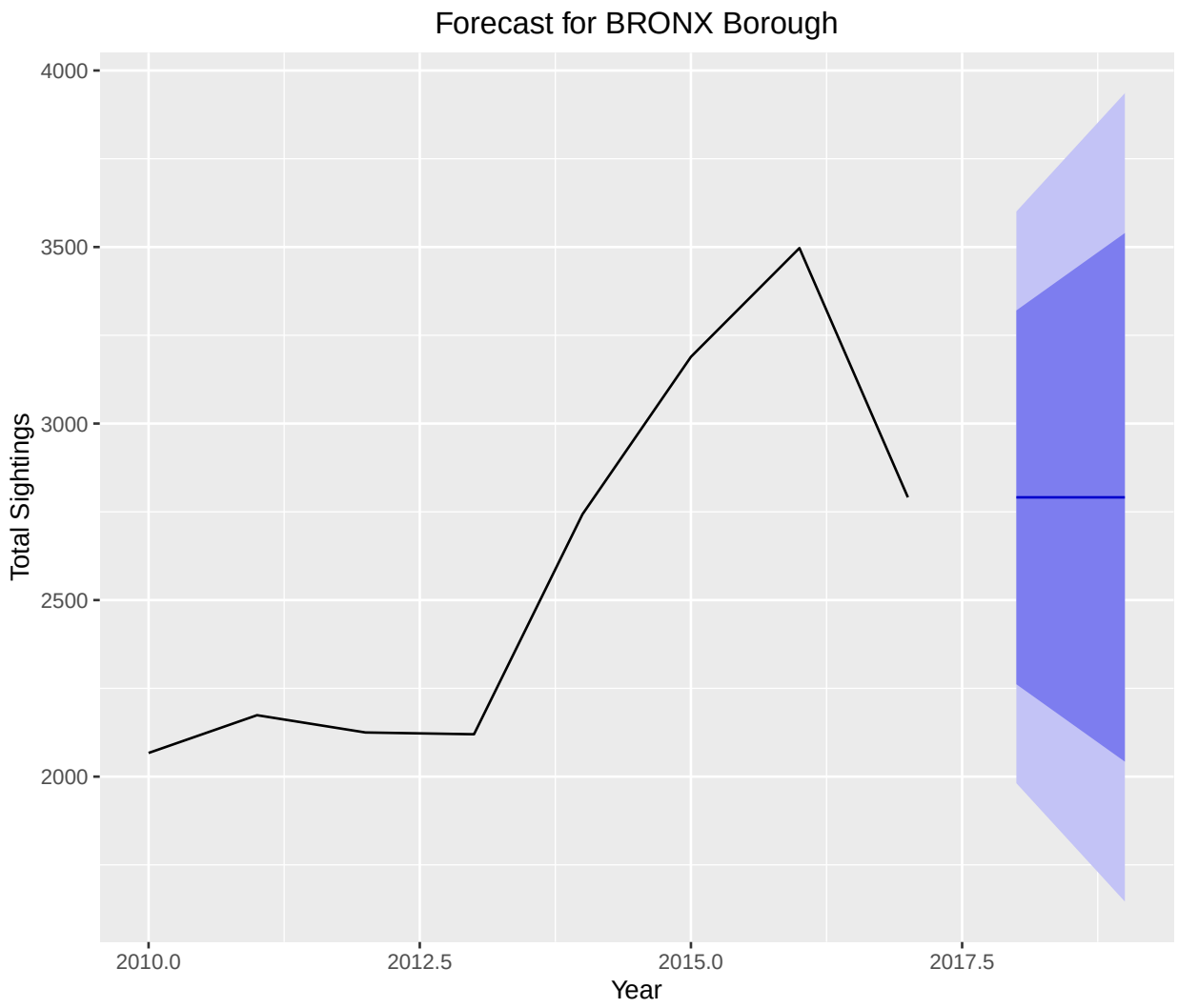
```
##          Borough Year Forecast
## 1          BRONX 2018     2791
## 2          BRONX 2019     2791
## 3       BROOKLYN 2018     5493
## 4       BROOKLYN 2019     5493
## 5      MANHATTAN 2018     3326
## 6      MANHATTAN 2019     3326
## 7         QUEENS 2018     2199
```

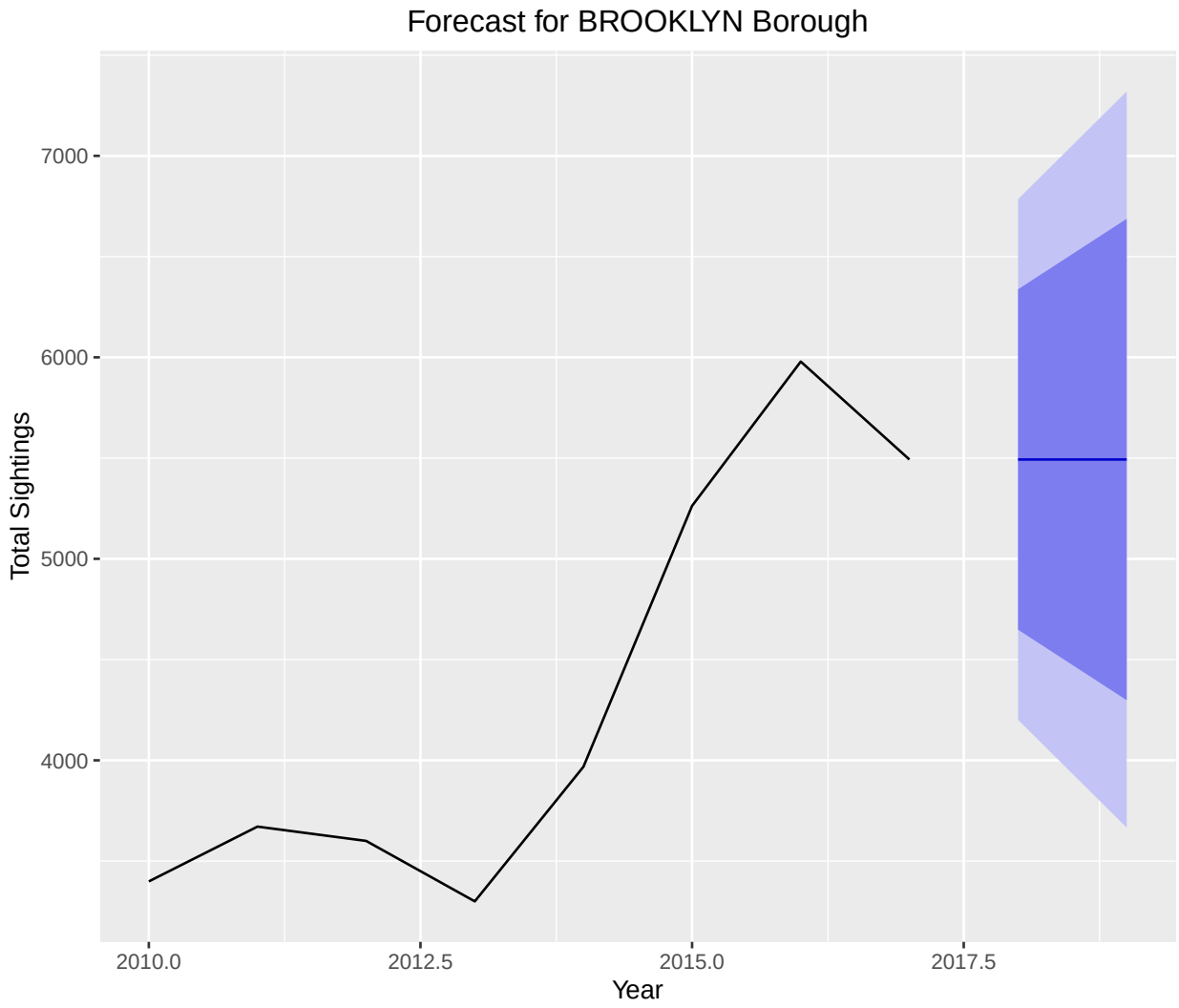
```
## 8         QUEENS 2019     2199
## 9  STATEN ISLAND 2018      615
## 10 STATEN ISLAND 2019      615
```

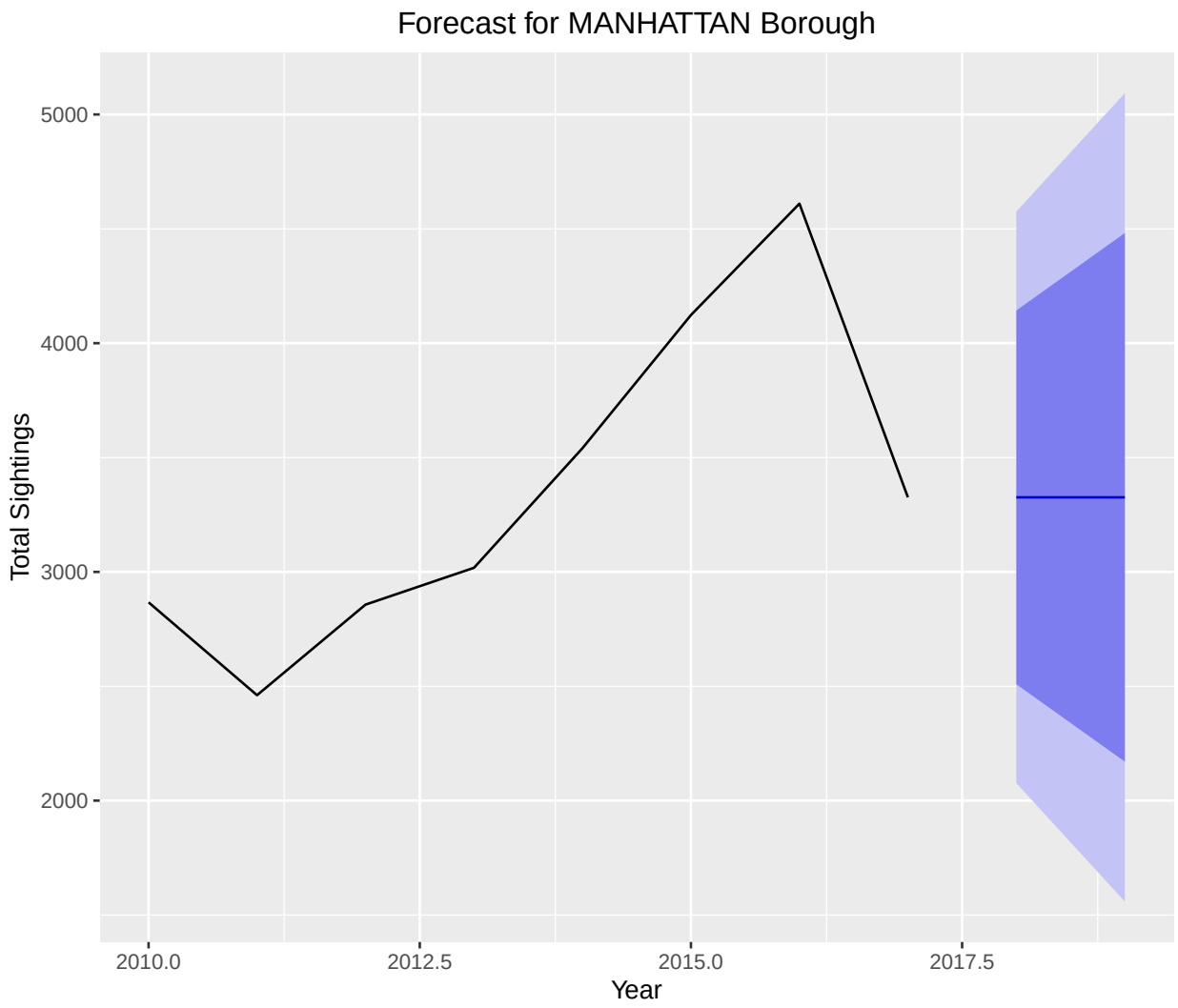
Borough Battle: Rat Sightings Across NYC

The exploration wouldn't be complete without dissecting the data across the diverse boroughs of NYC. The bar chart unveils the varying prevalence of rat sightings, allowing us to identify which boroughs are more susceptible to these encounters. This spatial breakdown is crucial for targeted interventions and urban planning initiatives.

```
library(forecast)
library(ggplot2)
library(dplyr)
library(tidyr)
library(purrr)

# Filtering for specific boroughs
borough_ts_list <- rat_sightings %>%
  filter(Borough %in% c("MANHATTAN", "BROOKLYN", "BRONX")) %>%
  group_by(Borough, sighting_year) %>%
  summarize(total_sightings = n(), .groups = "drop") %>%
  group_by(Borough) %>%
  complete(sighting_year = full_seq(sighting_year, period = 1)) %>%
  arrange(Borough, sighting_year) %>%
  group_split(Borough)

# Forecasting
forecast_plots <- map(borough_ts_list, function(borough_data) {
  borough_data <- borough_data %>% replace_na(list(total_sightings = 0))
  ts_data <- ts(borough_data$total_sightings, start = min(borough_data$sighting_year),
                frequency = 1)
  fit <- auto.arima(ts_data)
  future <- forecast(fit, h = 2)

  # Generate plot
  plot <- autoplot(future) +
    labs(title = paste("Forecast for", unique(borough_data$Borough), "Borough"),
         x = "Year", y = "Total Sightings") +
    theme(plot.title = element_text(hjust = 0.5))

  return(plot)
})

# Printing the plots
for(plot in forecast_plots) {
  print(plot)
}
```


Forecast for BRONX Borough
Total Sightings
4000
3500
3000
2500
2000
2010.0
2012.5
2015.0
2017.5
Year


Forecast for BROOKLYN Borough
Total Sightings
7000
6000
5000
4000
2010.0
2012.5
2015.0
2017.5
Year




```
# Save the plot as pdf
ggsave("forecasting_plot.pdf", width = 11, height = 8.5, unit = "in")

# Save the plot as png
ggsave("forecasting_plot.png", width = 11, height = 8.5, unit = "in")
```

Conclusion:

In this data-driven narrative, we've journeyed through the hidden realms of rat sightings in the heart of NYC. From the geographical distribution of encounters to temporal trends and predictive insights, the visualizations provide a comprehensive understanding of the dynamic relationship between the city and its furry inhabitants. As we conclude, we reflect on the potential implications for public health, urban planning, and the ongoing saga of coexistence in the bustling streets of the Big Apple.

Appendix

Reflection on Graphical Excellence and Visualization Principles

Original Graphic Critique

The original graphic (map) provided for NYC rat sightings was neither clear nor insightful. It failed to communicate the data effectively, suffering from aesthetic issues (tackiness and inconsistency), substantive issues (ineffective use of data), and perceptual issues (potential to confuse or mislead viewers) without revealing any actual conclusion of sightings

The Story of the New Graphics

My new graphics aim to tell a story of the rat sightings in NYC by presenting a clear, honest, and insightful narrative. I used line charts to show trends over time, and bar charts to compare boroughs. This approach allows us to impart the most information with the least ink in the smallest space, providing multivariate analysis while maintaining clarity and efficiency.

Principles of CRAP (Contrast, Repetition, Alignment, Proximity)

By applying the CRAP principles, I enhanced the visual appeal and readability of my graphics. I used contrast to differentiate between data points, repetition to maintain a coherent style across graphics, alignment to organize the information logically, and proximity to group related items, ensuring that our graphics are both aesthetically good and functionally effective.

Healy's and Cairo's Principles

My visualizations adhere to Healy's principles of good practice by being honest and reproducible, avoiding the pitfalls of default settings, and using data effectively. I also embraced Cairo's five qualities of great visualizations by ensuring that my graphics are truthful, functional, beautiful, insightful, and enlightening, thus achieving graphical excellence as described by Tufte.

By adhering to these principles, I ensured that my visualizations not only present the data in an interesting and statistically sound manner but also communicate complex ideas with precision and efficiency. I strive to tell the truth about the data, providing the viewer with a clear and accurate understanding of rat sightings in NYC