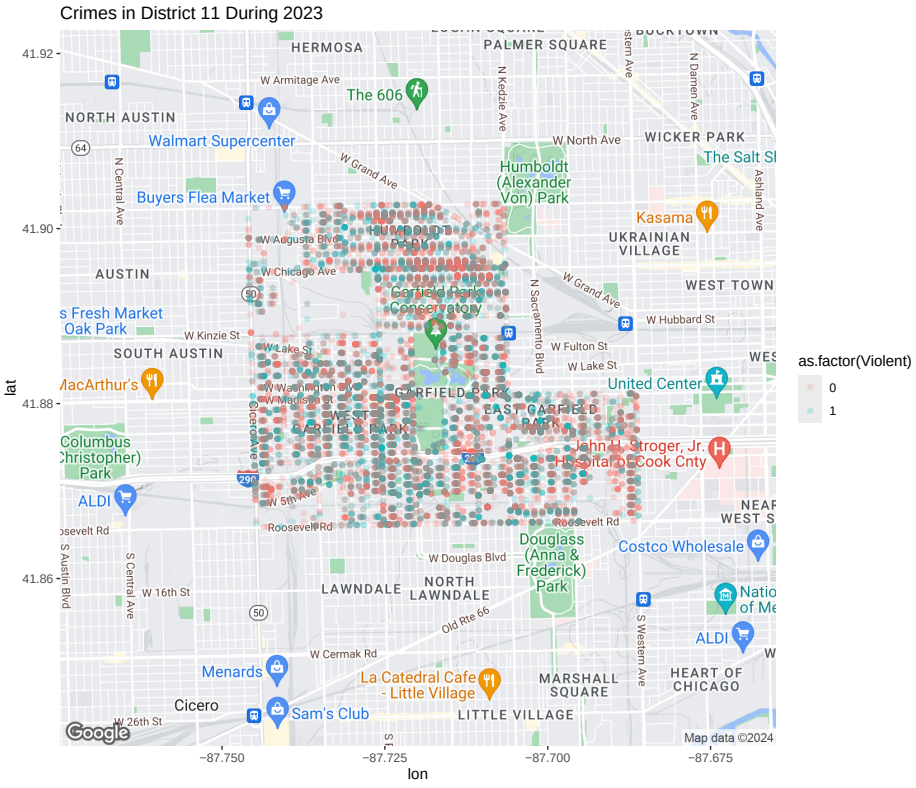
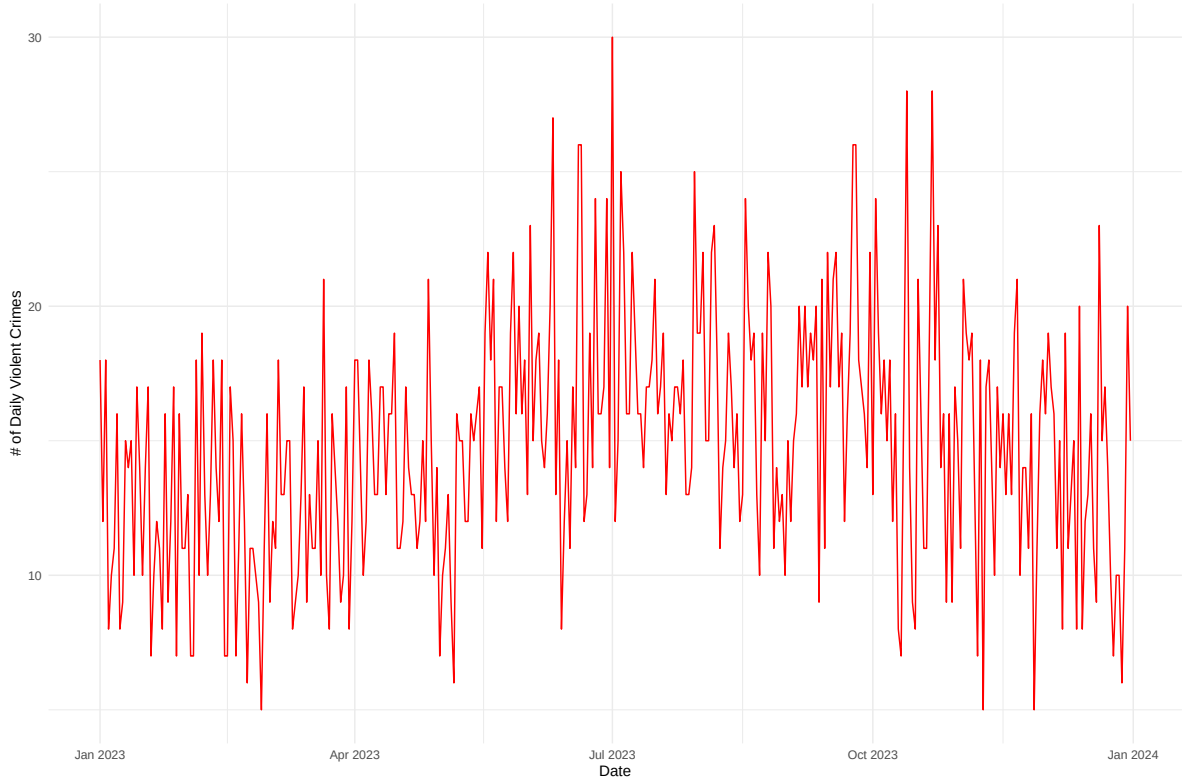
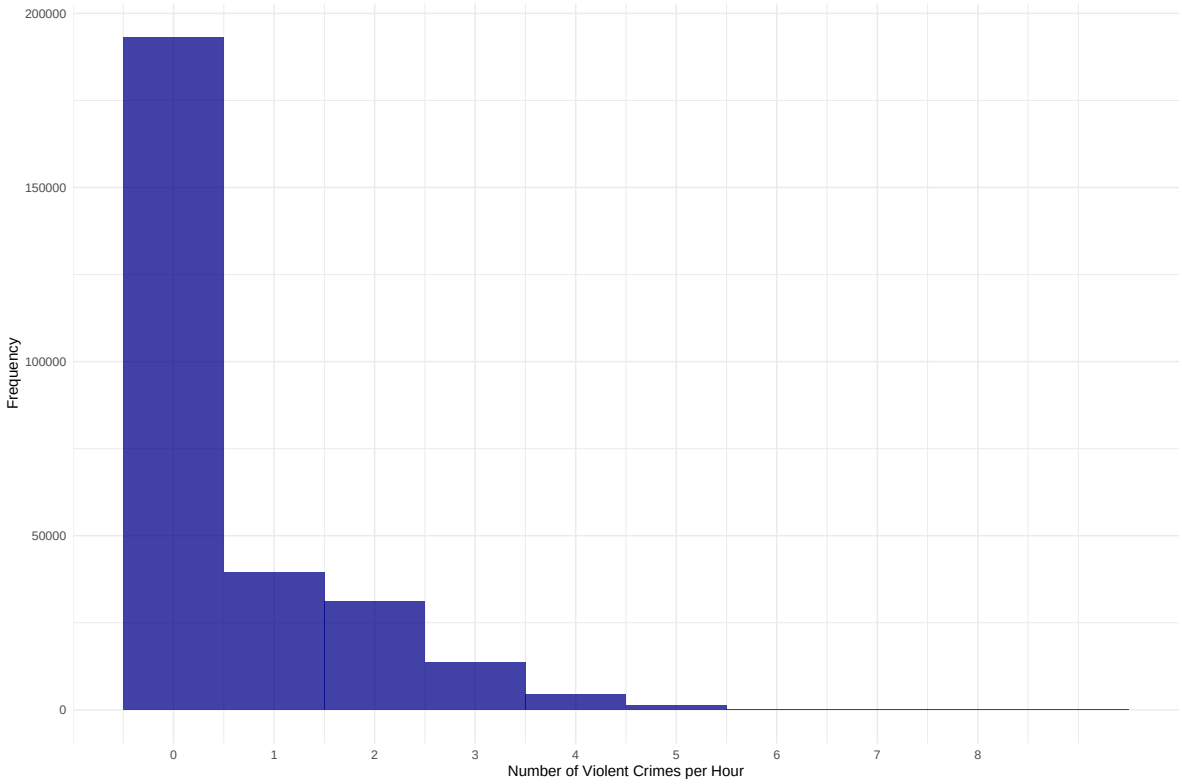
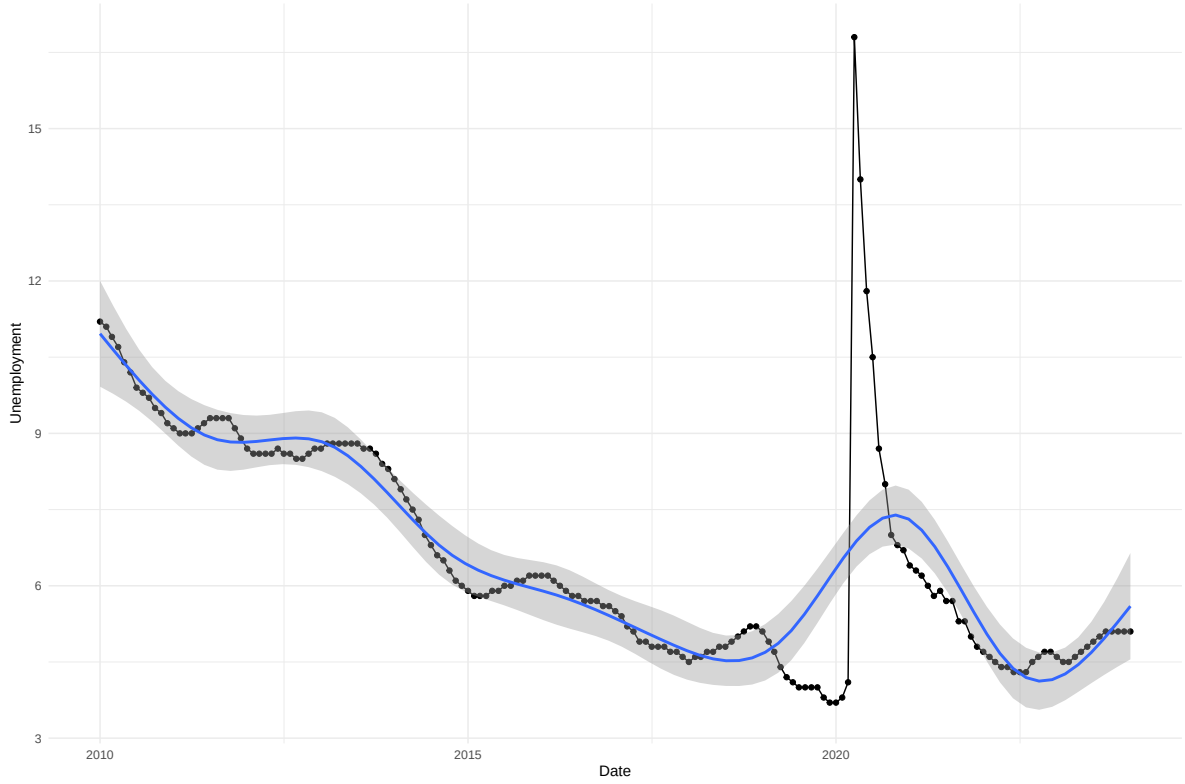
Crimes - EDA

Sam Lee




30
20
10
# of Daily Violent Crimes
Jan 2023
Apr 2023
Jul 2023
Oct 2023
Jan 2024
Date


200000
150000
100000
50000
0
Frequency
0
1
2
3
4
5
6
7
8
Number of Violent Crimes per Hour




```
load("model_zeroinf.RData")

coefficients <- coef(summary(model.zeroinf))$count %>% as.matrix()
```

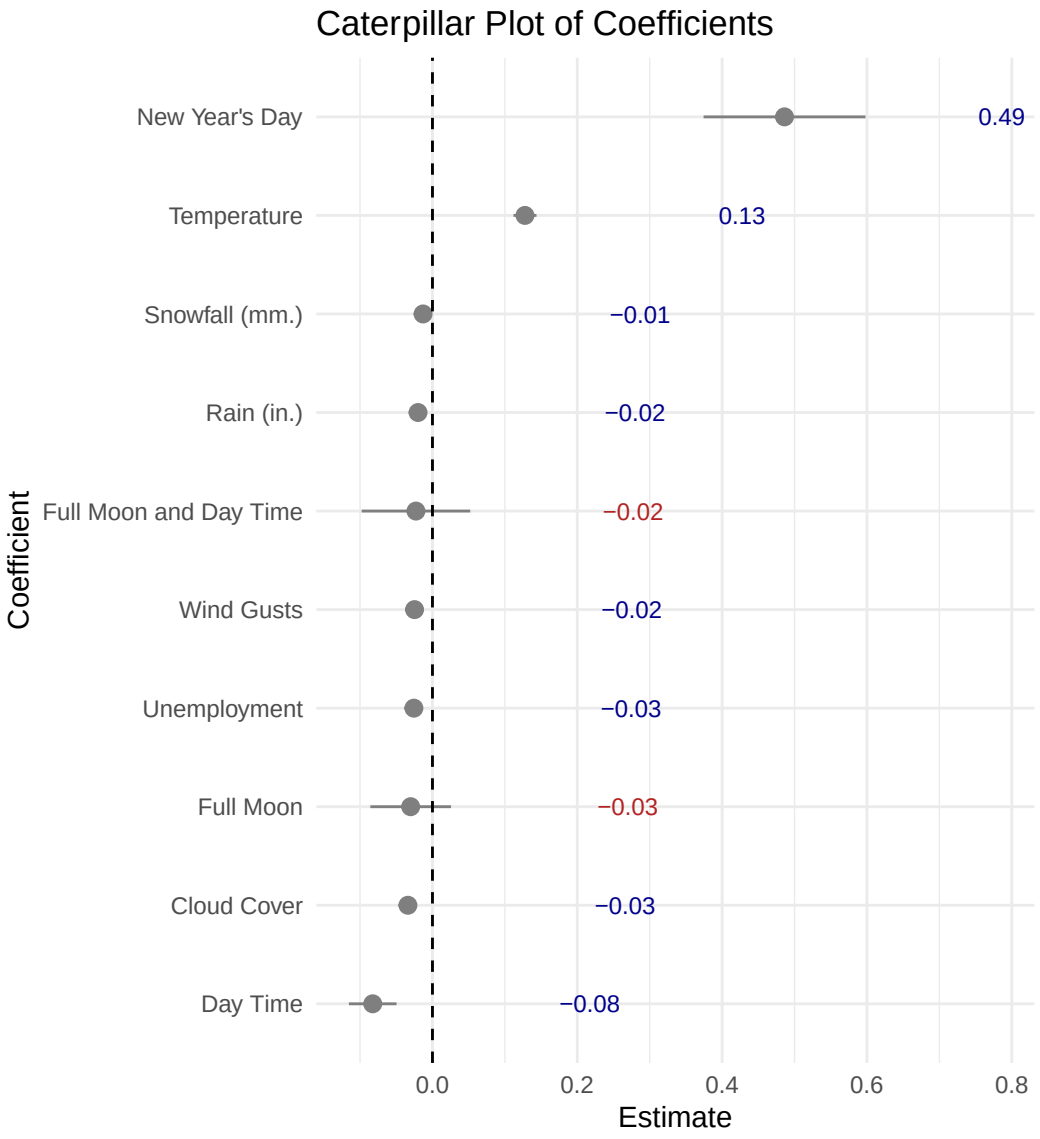
```
Warning in sqrt(diag(object$vcov)): NaNs produced
```

```
effects = c("Full Moon", "Day Time", "Temperature", "Rain", "Unemployment", "New Year's Da
            "Cloud Cover")

# Create a data frame for plotting
plot_data <- data.frame(
  term = rownames(coefficients),
  estimate = coefficients[,1],
  ci_low = coefficients[,1] - 1.96 * coefficients[,2],
  ci_high = coefficients[,1] + 1.96 * coefficients[,2],
  significance = ifelse(coefficients[,4] < 0.05, "Significant", "Not Significant")
)
```

```r
plot_data$term =c("Intercept", "Full Moon", "Day Time", "Unemployment", "Christmas Day", "
  "Independence Day", "Labor Day", "MLK Jr. Day", "Memorial Day", "New Year's Day", "NYE",
  "Thanksgiving Day", "Valentine's Day", "Veterans Day", "Year", "Year^2", "Year^3", "Apri
  "July", "June", "March", "May", "November", "October", "September", "Mon", "Sat", "Sun",

plot_data <- plot_data %>% filter(str_detect(term, paste(effects, collapse = "|"))) %>%
  arrange(desc(estimate))

# plot_data <- plot_data %>% arrange(desc(abs(estimate))) %>% head(30)

# Plot
ggplot(plot_data, aes(x = reorder(term, estimate), y = estimate, ymin = ci_low, ymax = ci_
  geom_pointrange(aes(color = significance), size = 0.5) +
  geom_hline(yintercept = 0, linetype = "dashed", color = "black", size = 0.5) +
  coord_flip() +
  theme_minimal() +
  theme(legend.position = "top") +
  labs(x = "Coefficient", y = "Estimate", title = "Caterpillar Plot of Coefficients") +
  scale_color_manual(values = c("green" = "darkblue", "red" = "firebrick"),
                     labels = c("Significant", "Not Significant")) +
  guides(color = guide_legend(title = "Significance")) +
  geom_text(data = plot_data[plot_data$significance == "Significant", ],
            aes(label = round(estimate, 2), y = estimate + 0.3),
            hjust = 0.5, color = "darkblue", size = 3) +
  geom_text(data = plot_data[plot_data$significance == "Not Significant", ],
            aes(label = round(estimate, 2), y = estimate + 0.3),
            hjust = 0.5, color = "firebrick", size = 3)
```

```
Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
i Please use `linewidth` instead.
```

```
Warning: No shared levels found between `names(values)` of the manual scale and the
data's colour values.
No shared levels found between `names(values)` of the manual scale and the
data's colour values.
```


Caterpillar Plot of Coefficients
New Year's Day
Temperature
Snowfall (mm.)
Rain (in.)
Full Moon and Day Time
Wind Gusts
Unemployment
Full Moon
Cloud Cover
Day Time
0.49
0.13
−0.01
−0.02
−0.02
−0.02
−0.03
−0.03
−0.03
−0.08
Coefficient
0.0
0.2
0.4
0.6
0.8
Estimate