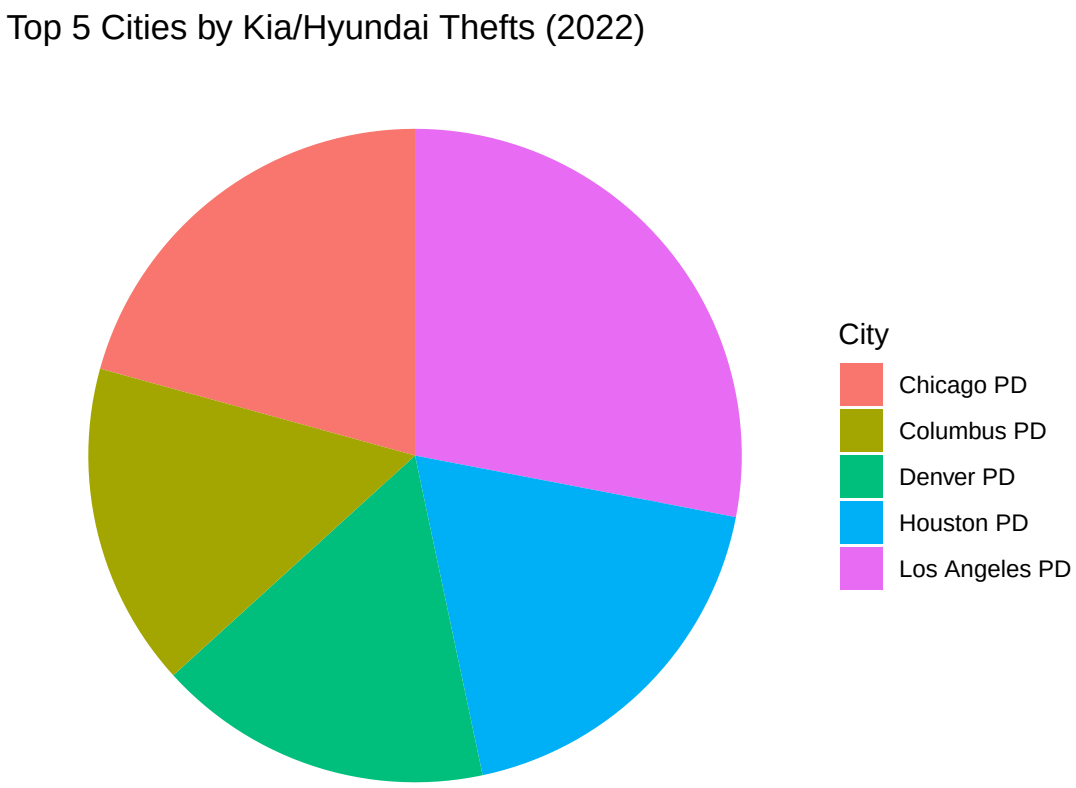
Kia/Hyundai Theft Analysis

Staci McDuffie

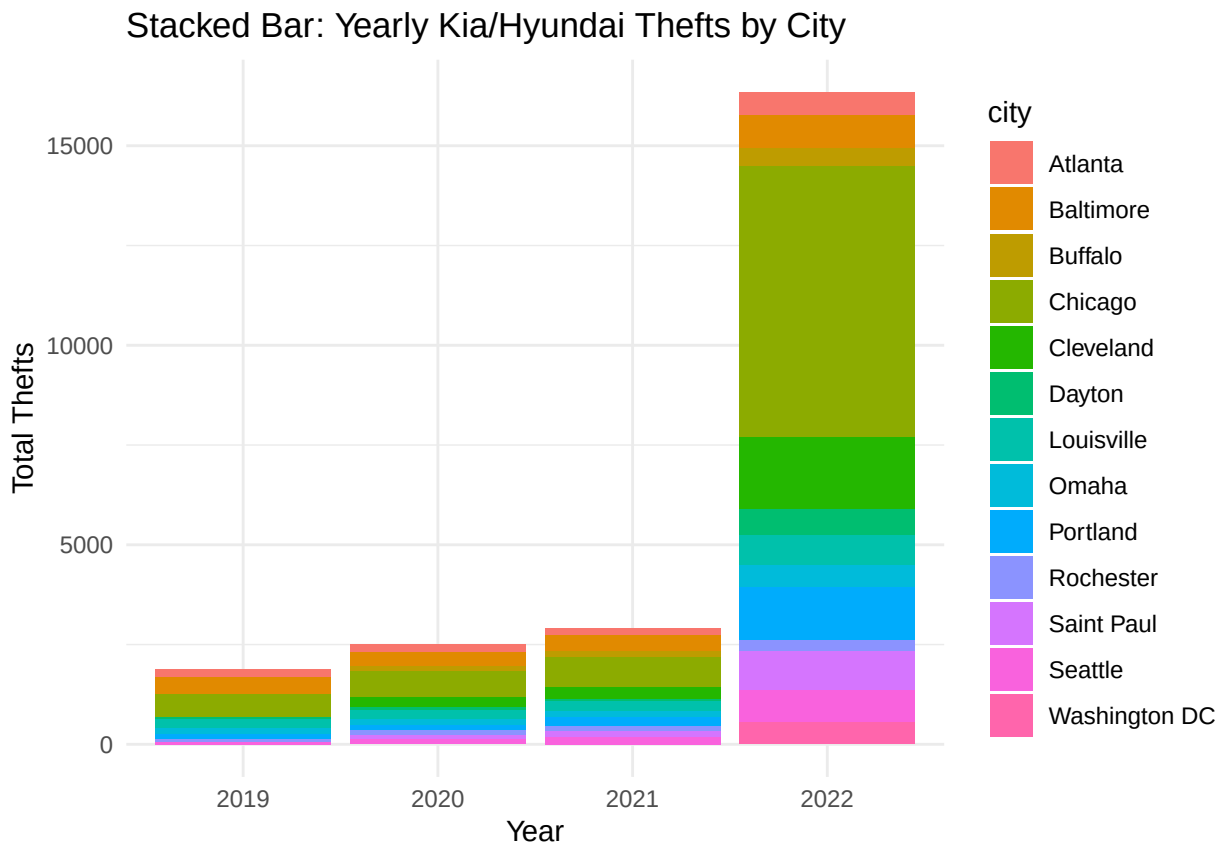
07/11/2025

```
# Required Libraries
library(ggplot2)
library(readr)
library(dplyr)
library(tidyr)
library(treemap)
library(ggthemes)
library(scales)
library(ggpubr)
library(tidyverse)
library(lubridate)
```

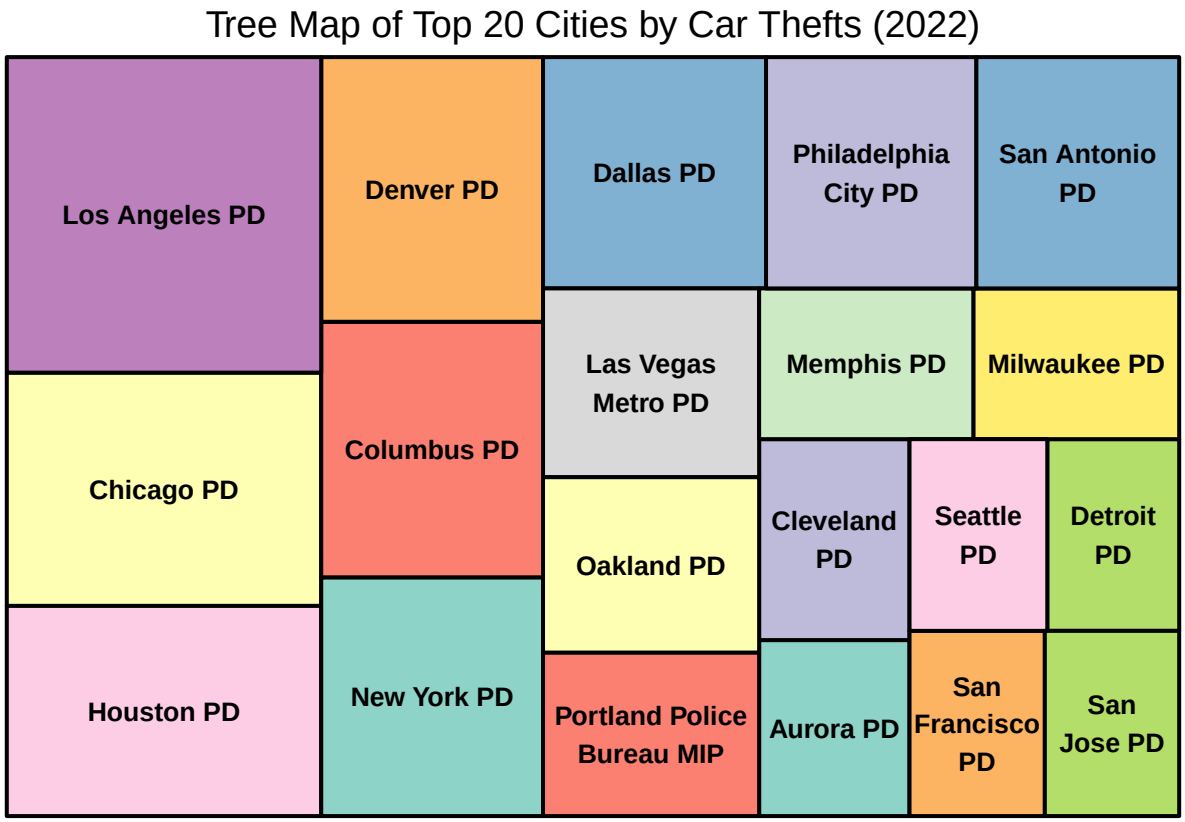
```
carTheftsMap <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/carTheftsMap.csv")

# Summarize by geo_name
city_summary <- carTheftsMap %>%
  group_by(geo_name) %>%
  summarise(total_thefts = sum(countCarThefts2022, na.rm = TRUE)) %>%
  arrange(desc(total_thefts)) %>%
  slice(1:5)

# Pie chart of top 5 cities
ggplot(city_summary, aes(x = "", y = total_thefts, fill = geo_name)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Top 5 Cities by Kia/Hyundai Thefts (2022)", fill = "City") +
  theme_void()
```



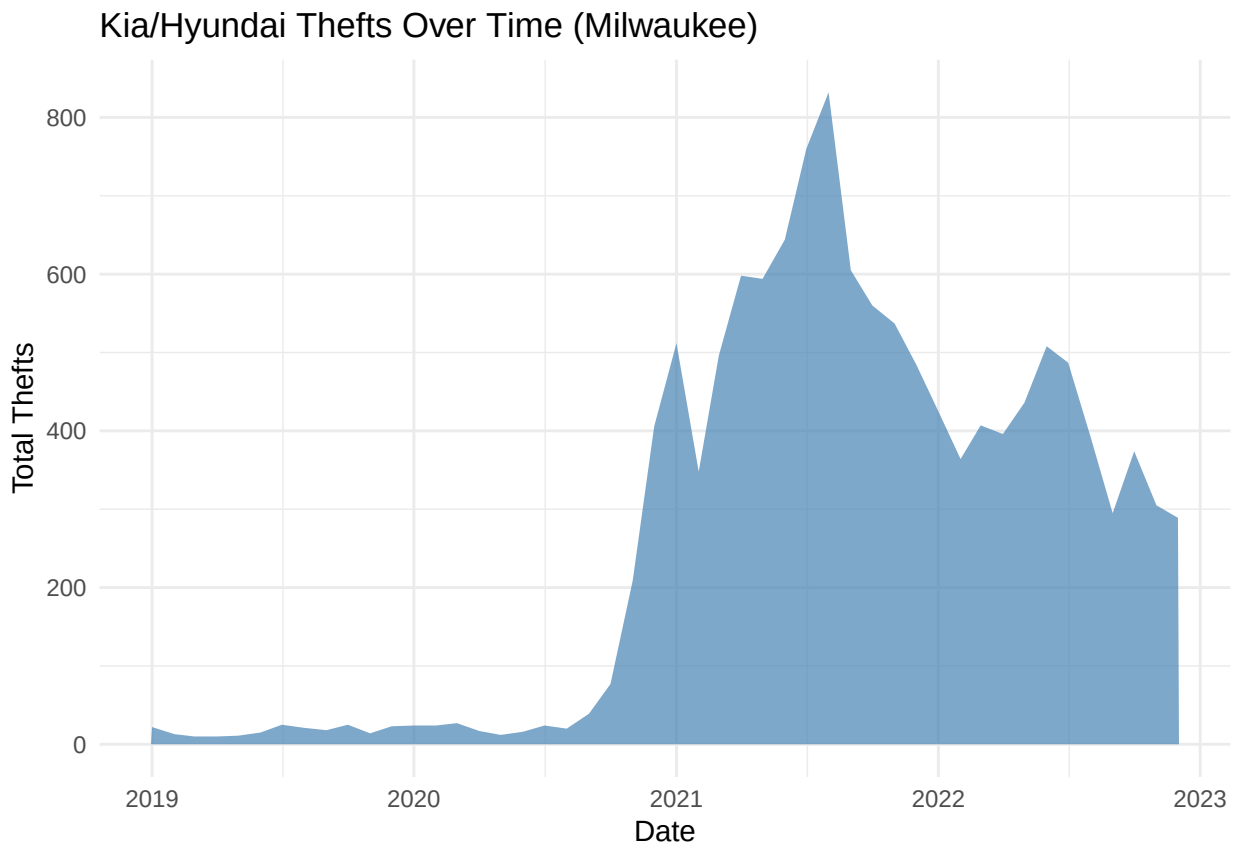

```
kiaHyundaiThefts <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/kiaHyundaiThefts.csv")

stacked_data <- kiaHyundaiThefts %>%
  group_by(year, city) %>%
  summarise(thefts = sum(countKiaHyundaiThefts, na.rm = TRUE)) %>%
  ungroup()

ggplot(stacked_data, aes(x = factor(year), y = thefts, fill = city)) +
  geom_bar(stat = "identity") +
  labs(title = "Stacked Bar: Yearly Kia/Hyundai Thefts by City",
       x = "Year", y = "Total Thefts") +
  theme_minimal()
```



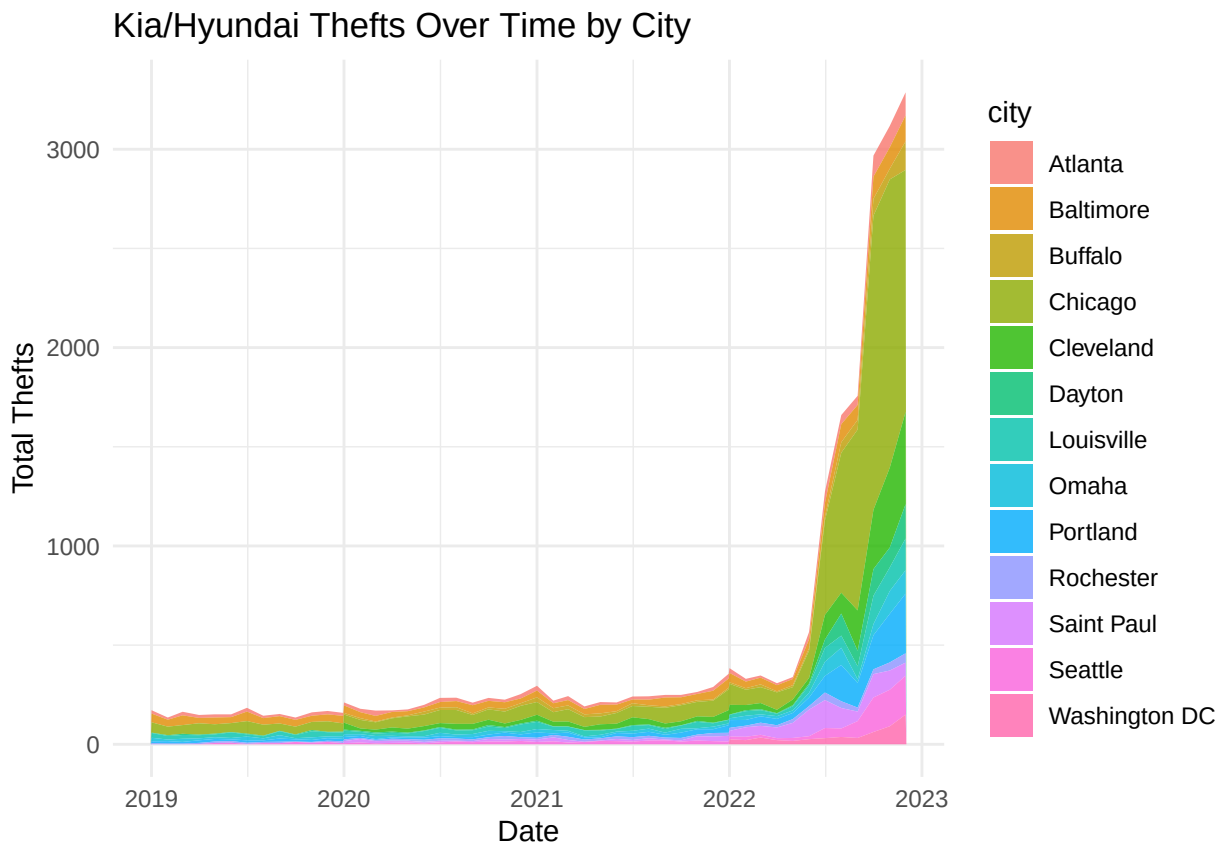

```
carTheftsMap <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/carTheftsMap.csv")

city_counts <- carTheftsMap %>%
  group_by(geo_name) %>%
  summarise(total = sum(countCarThefts2022, na.rm = TRUE)) %>%
  arrange(desc(total)) %>%
  slice(1:20)

treemap(city_counts,
        index = "geo_name",
        vSize = "total",
        title = "Tree Map of Top 20 Cities by Car Thefts (2022)",
        palette = "Set3",
        border.col = "black",
        draw.legend = TRUE,
        fontsize.labels = 10,
        fontcolor.labels = "black")
```




```r
KiaHyundaiMilwaukeeData <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/KiaHyundaiMilwaukeeData.csv

# Convert 3-letter month to numeric
KiaHyundaiMilwaukeeData <- KiaHyundaiMilwaukeeData %>%
  mutate(
    month_num = match(month, month.abb),  # Convert "Jan" → 1, etc.
    date = make_date(year, month_num, 1)
  )

# Group and summarize
area_data <- KiaHyundaiMilwaukeeData %>%
  group_by(date) %>%
  summarise(total_thefts = sum(countKiaHyundaiThefts, na.rm = TRUE), .groups = "drop")

# Plot area chart
ggplot(area_data, aes(x = date, y = total_thefts)) +
  geom_area(fill = "steelblue", alpha = 0.7) +
  labs(title = "Kia/Hyundai Thefts Over Time (Milwaukee)",
       x = "Date", y = "Total Thefts") +
  theme_minimal()
```




```
kiaHyundaiThefts <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/kiaHyundaiThefts.csv")

# Convert month name → number, and create a real date column
kiaHyundaiThefts <- kiaHyundaiThefts %>%
  mutate(
    month_num = match(month, month.abb),        # Convert "Jan" → 1
    date = make_date(year, month_num, 1)        # Build proper date
  )

# Group by city + date
stacked_area_data <- kiaHyundaiThefts %>%
  group_by(date, city) %>%
  summarise(thefts = sum(countKiaHyundaiThefts, na.rm = TRUE), .groups = "drop")

ggplot(stacked_area_data, aes(x = date, y = thefts, fill = city)) +
  geom_area(alpha = 0.8) +
  labs(title = "Kia/Hyundai Thefts Over Time by City",
       x = "Date", y = "Total Thefts") +
  theme_minimal()
```




```
Motherboard_VICE_News_Kia_Hyundai_Theft_Data <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/Motherboard VICE News Kia H

# Step 1: Load the raw file
df_vice_raw <- read_csv(
  "C:/Users/athle/OneDrive/Desktop/DSC 640 Data Presentation & Visualization/Motherboard VICE News Kia H

# Step 2: Extract the first row as column names
raw_header <- df_vice_raw[1, ] |> unlist() |> as.character()

# Step 3: Clean header: replace blank or NA names with unique placeholders
clean_names <- ifelse(is.na(raw_header) | raw_header == "", paste0("X", seq_along(raw_header)), raw_hea

# Step 4: Drop first row and assign clean names
df_vice <- df_vice_raw[-1, ]
colnames(df_vice) <- clean_names

# Step 5: Identify Kia/Hyundai columns (every 3rd column starting at index 2)
kia_indices <- seq(2, length(clean_names), by = 3)
kia_cols <- clean_names[kia_indices]
kia_cols <- kia_cols[!is.na(kia_cols) & kia_cols != ""]  # Remove bad names

# Step 6: Pivot to long format
df_kia_long <- df_vice %>%
  select(all_of(kia_cols)) %>%
  pivot_longer(cols = everything(), names_to = "city", values_to = "thefts") %>%
```

```r
  mutate(thefts = as.numeric(thefts)) %>%
  filter(!is.na(thefts))

# Step 7: Summarize and prepare donut data
donut_data <- df_kia_long %>%
  group_by(city) %>%
  summarise(total_thefts = sum(thefts, na.rm = TRUE)) %>%
  arrange(desc(total_thefts)) %>%
  slice(1:5) %>%
  mutate(
    prop = total_thefts / sum(total_thefts),
    label = paste0(city, " (", round(prop * 100, 1), "%)")
  )

# Step 8: Plot Donut Chart
ggplot(donut_data, aes(x = 2, y = total_thefts, fill = label)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar(theta = "y") +
  xlim(0.5, 2.5) +
  theme_void() +
  labs(title = "Donut Chart: Top 5 Cities by Kia/Hyundai Thefts (VICE News)") +
  theme(legend.position = "right")
```