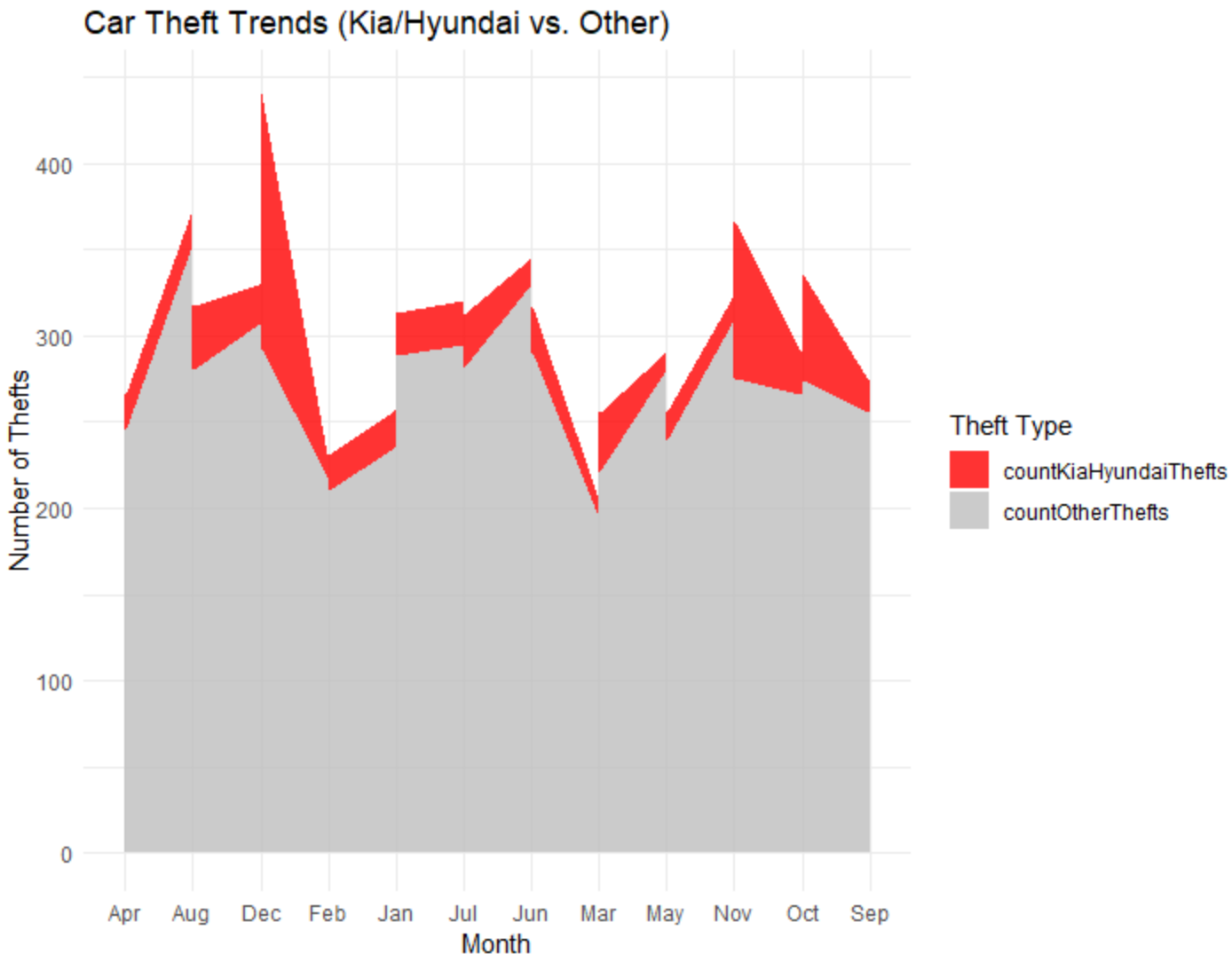
The owner of a private, Denver CO car dealership recently heard that Kia & Hyundai cars are extremely vulnerable to theft and has asked their analyst to verify whether the claims are legitimate. If sales are expected to drop for those brands, the owner would seek to restructure their brand portfolio with more reputable brands. The analyst opts to fashion an expedient, visualization-based analysis to determine whether the claim is worth investigating further.

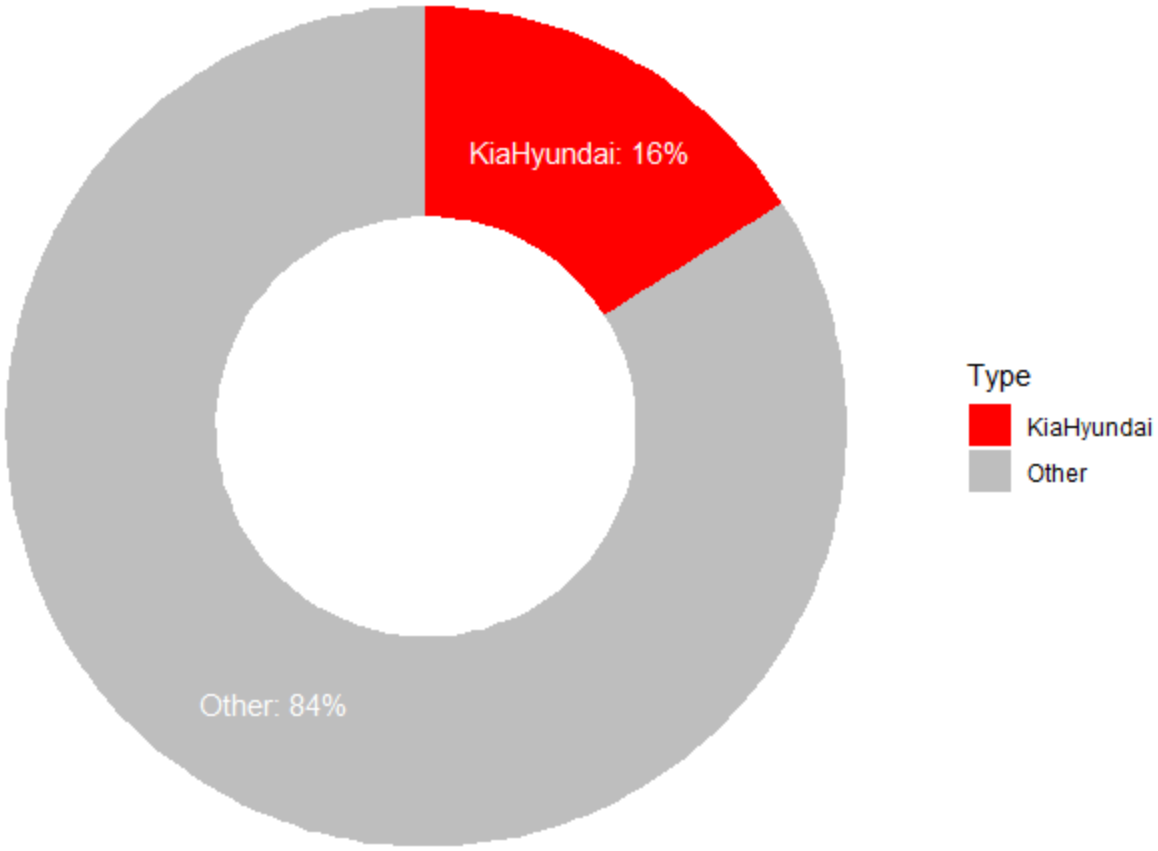
The principal ethical concern of the analyst is to provide a thorough analysis to quell their boss' concerns or to guide further analysis into reducing their brand dependency on Kia/ Hyundai. If the analysis is faulty, a significant portion of revenue may be affected, which forces the dealership to cut costs, resulting in potential layoffs and reduced automobile supply in the area. Despite the high risk, the rumor mills that churn media attention is typically exaggerated, and the analyst's other responsibilities are also important to the dealership operation; therefore, the secondary ethical concern of the analyst is to avoid overindulging the concern via an extensive analysis into an unjustified premise. Both concerns are addressed with a simple, preliminary analysis with minimized balance that stands up to scrutiny without wasting valuable time.

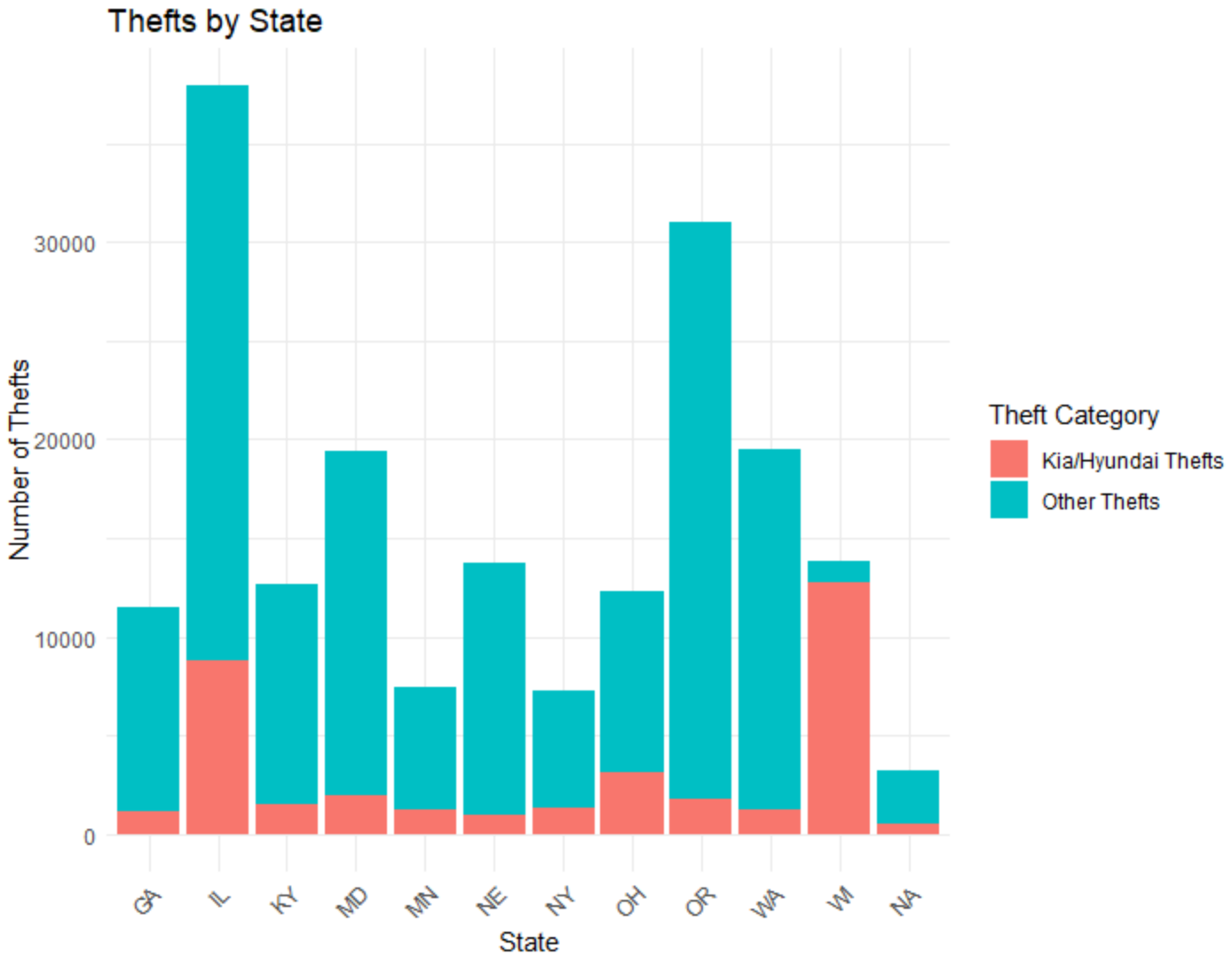
The visualizations utilize each of the four datasets (to limit bias) and provide a comprehensive overview of regionalized car thefts by make and date. Firstly, overall thefts are stacked over time and split according to Kia/Hyndai (K/H hereafter) and other brands. Evidently, K/H thefts during the sample follow the overall trend, with the possibility of greater susceptibility during spikes. The next donut plot shows proportional K/H theft rates over the time vs. other brands, and a small disparity of thefts is present over the average market share of ~ 10% for the same period. The following stacked bar displays thefts by state, detailing the relative proportion of brand thefts by state. WI exhibits disproportional thefts, but the rest of the states have normal variance amounts. Since the analysis differs across states, the next plot focuses on Denver thefts over time and indicates no extraordinary trends for K/H over the rest of the population. The final plot shows proportional K/H over all thefts for each city, revealing that Denver thefts aren't exceptional among other metropolitan regions. Finally, the last plot is an interactive map of car theft trends, indicating that the Denver-region exhibits increased theft trends over time, but so does the rest of the nation.

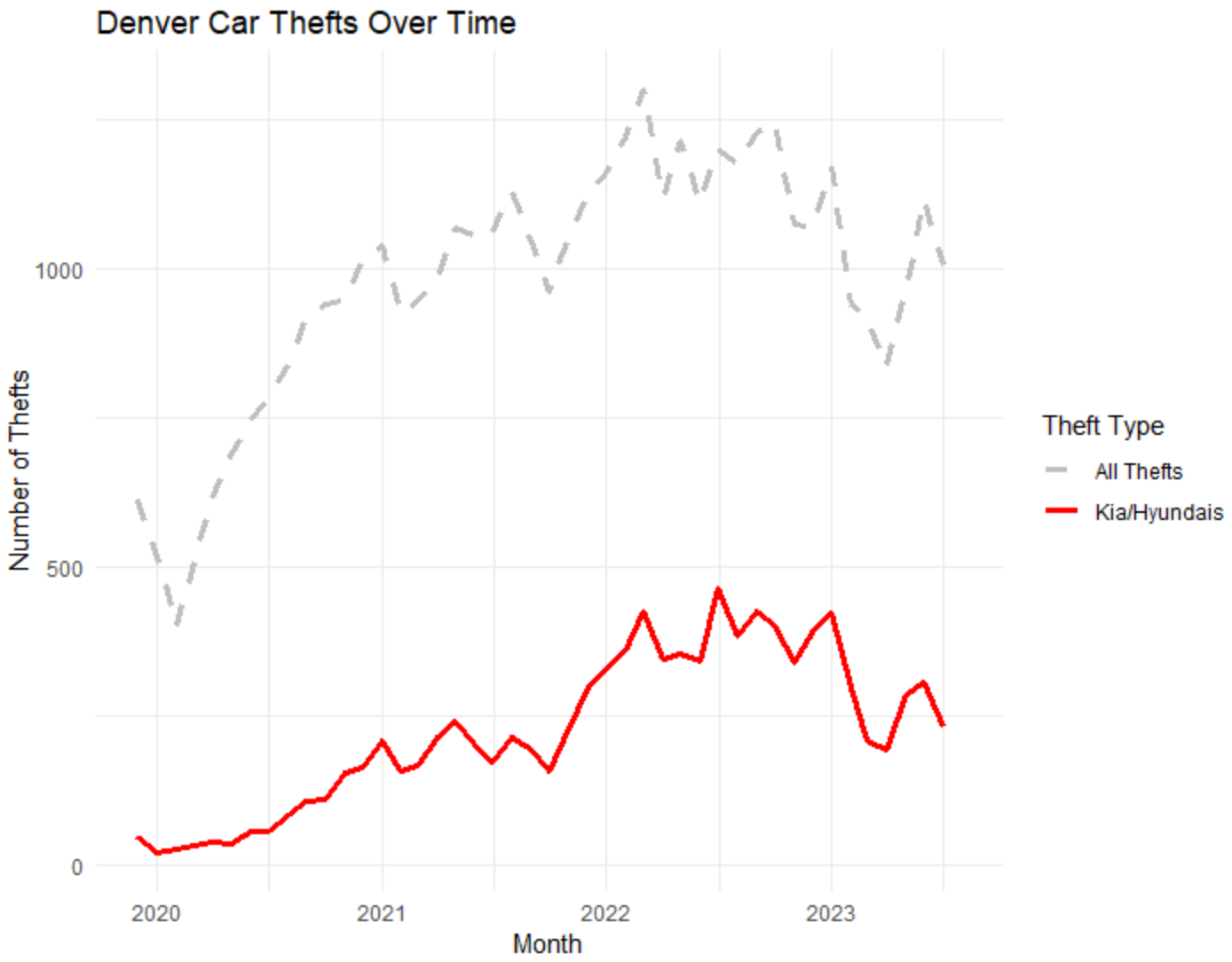
In conclusion, the analyst reports to the owner that claims of extreme vulnerability are exaggerated, but it might serve the dealership to mention the brand damage during vendor to see if the brand has incentives to offer the dealership or new security features planned to mitigate the possible issues. The dealership won't have more leverage given the minor impact of the thefts.


Car Theft Trends (Kia/Hyundai vs. Other)
Number of Thefts
400
300
200
100
0
Apr
Aug
Dec
Feb
Jan
Jul
Jun
Mar
May
Nov
Oct
Sep
Month
Theft Type
countKiaHyundaiThefts
countOtherThefts

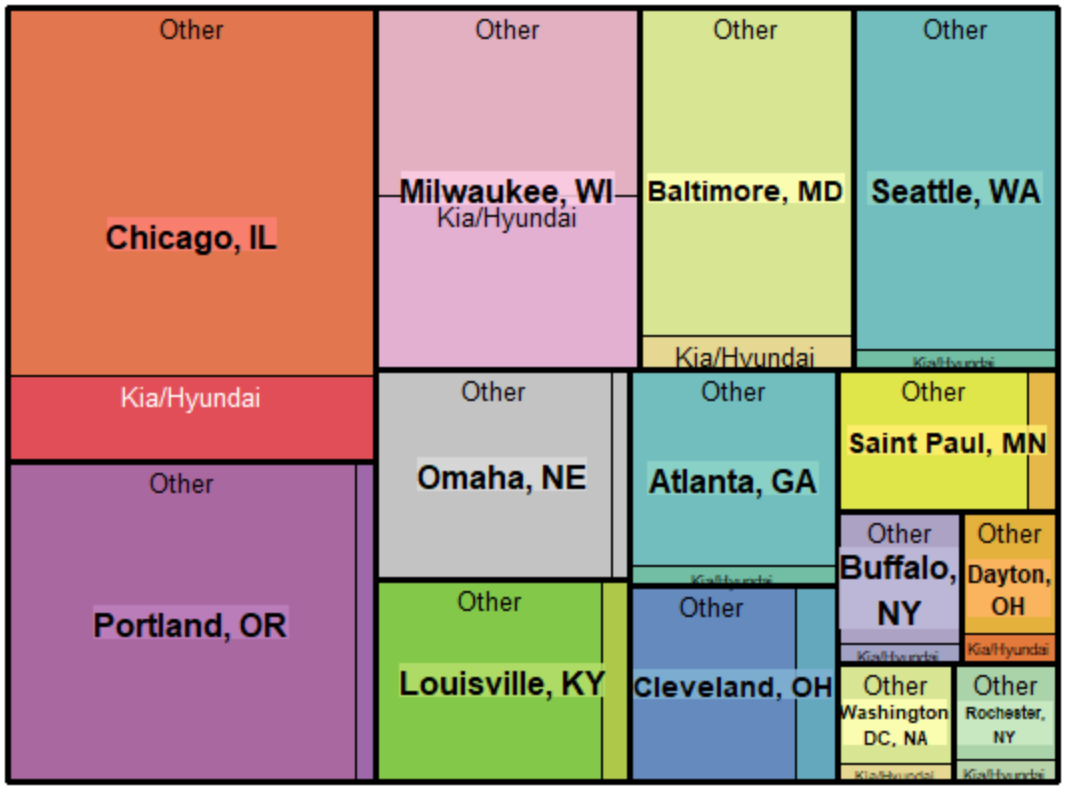
Proportion of Kia/Hyundai vs. Other Thefts

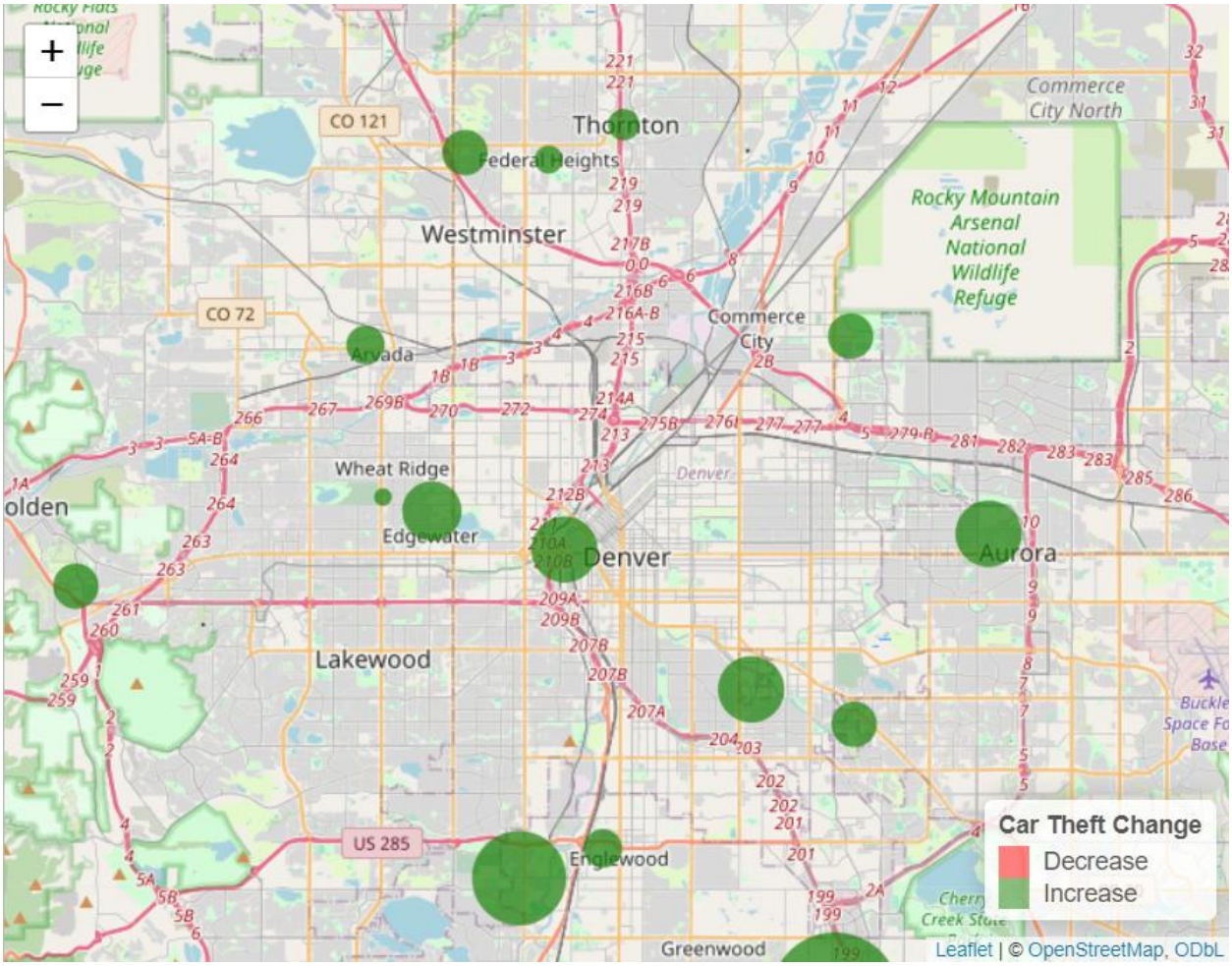



Thefts by State
Number of Thefts
0
10000
20000
30000
GA
IL
KY
MD
MN
NE
NY
OH
OR
WA
WI
NA
State
Theft Category
Kia/Hyundai Thefts
Other Thefts


Denver Car Thefts Over Time
Number of Thefts
1000
500
0
2020
2021
2022
2023
Month
Theft Type
All Thefts
Kia/Hyundais

Car Thefts by Location and Type





R SCRIPT:

```
################################################################################
# SETUP
library(tidyverse)
library(leaflet)
library(readxl)
library(treemap)

# data import
setwd("C:/Users/syver/OneDrive/Documents/05_school/640/data/week 6")
# filter VICE dataset to Denver only per audience
denver_theft <- read_excel("Motherboard VICE News Kia Hyundai Theft Data.xlsx",
                           range = "A2:D50", col_names = TRUE)
carTheftsMap <- read.csv("carTheftsMap.csv")
KiaHyundaiMilwaukeeData <- read.csv("KiaHyundaiMilwaukeeData.csv")
kiaHyundaiThefts <- read.csv("kiaHyundaiThefts.csv")

combined_df = rbind(KiaHyundaiMilwaukeeData, kiaHyundaiThefts) # append df

################################################################################
# THEFT TRENDS STACKED AREA
df_long <- pivot_longer( # reshape df for simplified viz
```

```
  combined_df, cols = c(countKiaHyundaiThefts, countOtherThefts),
  names_to = "TheftType", values_to = "Count")

ggplot(df_long, # plot all thefts vs. KH over time
       aes(x = month, y = Count, fill = TheftType, group = TheftType)) +
  geom_area(alpha = 0.8) +
  scale_fill_manual(
    values = c("countKiaHyundaiThefts" = "red", "countOtherThefts" = "grey")) +
    labs(title = "Car Theft Trends (Kia/Hyundai vs. Other)",
    y = "Number of Thefts", x = "Month", fill = "Theft Type") +
  theme_minimal()

################################################################################
# KIA/HYUNDAI THEFT DONUT
df_donut <- combined_df %>%
  summarise(KiaHyundai = sum(countKiaHyundaiThefts), # aggregate then pivot
            Other = sum(countOtherThefts)) %>%
  pivot_longer(cols = everything(), names_to = "Type", values_to = "Total") %>%
  mutate(fraction = Total / sum(Total), # create proportions
         ymax = cumsum(fraction),
         ymin = c(0, head(ymax, n=-1)),
         label_pos = (ymax + ymin) / 2,
         label = paste0(Type, ": ", round(fraction * 100), "%"))

# param & init donut
ggplot(df_donut, aes(ymax=ymax, ymin=ymin, xmax=4, xmin=3, fill=Type)) +
  geom_rect() +
  coord_polar(theta="y") +
  xlim(c(2, 4)) +
  theme_void() +
  geom_text(aes(x=3.5, y=label_pos, label=label), color="white") +
  scale_fill_manual(values=c("KiaHyundai" = "red", "Other" = "gray")) +
  labs(title = "Proportion of Kia/Hyundai vs. Other Thefts")

################################################################################
# STACKED BARS
ggplot(combined_df, aes(x = state)) + # init chart for state thefts by make
  geom_bar(aes(y = countOtherThefts, fill = "Other Thefts"),
           stat = "identity", position = "stack") +
  geom_bar(aes(y = countKiaHyundaiThefts, fill = "Kia/Hyundai Thefts"),
           stat = "identity") +
  theme_minimal() +
  labs(title = "Thefts by State",
       x = "State",
       y = "Number of Thefts",
       fill = "Theft Category") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

################################################################################
# PIE CHART
denver_theft <- denver_theft %>%
  rename(
    date = 1,               # rename col to "date"
```

```
   `KiaHyundais` = `Kia/Hyundais`  # remove special characters
  )
ggplot(denver_theft, aes(x = date)) + # plot KH vs. All for Denver (filtered)
  geom_line(aes(y = KiaHyundais, color = "Kia/Hyundais"), size = 1.2) +
  geom_line(aes(y = All, color = "All Thefts"), size = 1.2, linetype = "dashed") +
  scale_color_manual(values = c("Kia/Hyundais" = "red", "All Thefts" = "gray")) +
  labs(title = "Denver Car Thefts Over Time",
       x = "Month", y = "Number of Thefts",
       color = "Theft Type") +
  theme_minimal()
################################################################################
# TREEMAP
df_treemap <- combined_df %>%
  select(city, state, countKiaHyundaiThefts, countOtherThefts) %>%
  tidyr::pivot_longer(
    cols = c(countKiaHyundaiThefts, countOtherThefts),
    names_to = "type",
    values_to = "count"
  ) %>%
  mutate(type = recode(type,
                       countKiaHyundaiThefts = "Kia/Hyundai",
                       countOtherThefts = "Other"))

df_treemap <- df_treemap %>% # concat levels to avoid squished labels
  mutate(location = paste(city, state, sep = ", "))

treemap(df_treemap,
        index = c("location", "type"),
        vSize = "count",
        type = "index",
        palette = "Set3",
        fontsize.labels = c(12, 10),
        align.labels = list(c("center", "center"), c("center", "top")),
        title = "Car Thefts by Location and Type")

################################################################################
# CAR THEFTS INTERACTIVE MAP
leaflet(carTheftsMap) %>%
  addTiles() %>% # init tiles, markers, map to features
  addCircleMarkers(~longitude, ~latitude,
                   radius = ~abs(percentChange2019to2022 * 10),
                   color = ~ifelse(percentChange2019to2022 < 0, "red", "green"),
                   label = ~paste(geo_name, "<br>",
                                  "2019:", countCarThefts2019, "<br>",
                                  "2022:", countCarThefts2022, "<br>",
                                  "Change:", round(percentChange2019to2022, 2)),
                   stroke = FALSE, fillOpacity = 0.7) %>%
  addLegend(position = "bottomright",
            colors = c("red", "green"),
            labels = c("Decrease", "Increase"),
            title = "Car Theft Change")
```