kia theft assignment

Bobga-Herman Gwanvoma

2025-01-20

Load the Data

```r
# Load necessary libraries
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.1
```

```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.4.1
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(tidyr)
```

```
## Warning: package 'tidyr' was built under R version 4.4.1
```

```r
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.4.1
```

```r
library(treemap)
```

```
## Warning: package 'treemap' was built under R version 4.4.2
```

```r
library(scales)
library(ggforce)
```

```
## Warning: package 'ggforce' was built under R version 4.4.2
```

```r
# Load datasets
vice_news_data <- read_excel("C:/Users/bobi/Documents/DSC 640/w5&6/Motherboard VICE News Kia Hyundai Th
```

```
## New names:
## * `` -> `...1`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...15`
## * `` -> `...16`
## * `` -> `...18`
## * `` -> `...19`
## * `` -> `...21`
## * `` -> `...22`
## * `` -> `...24`
## * `` -> `...25`
## * `` -> `...27`
## * `` -> `...28`
## * `` -> `...30`
## * `` -> `...31`
## * `` -> `...33`
## * `` -> `...34`
## * `` -> `...36`
## * `` -> `...37`
## * `` -> `...39`
## * `` -> `...40`
## * `` -> `...42`
## * `` -> `...43`
## * `` -> `...45`
## * `` -> `...46`
## * `` -> `...48`
## * `` -> `...49`
## * `` -> `...51`
## * `` -> `...52`
## * `` -> `...54`
## * `` -> `...55`
## * `` -> `...57`
## * `` -> `...58`
## * `` -> `...60`
## * `` -> `...61`
## * `` -> `...63`
## * `` -> `...64`
## * `` -> `...66`
## * `` -> `...67`
## * `` -> `...69`
## * `` -> `...70`
## * `` -> `...72`
## * `` -> `...73`
```

* `` -> `...75`
* `` -> `...76`
* `` -> `...78`
* `` -> `...79`
* `` -> `...81`
* `` -> `...82`
* `` -> `...84`
* `` -> `...85`
* `` -> `...87`
* `` -> `...88`
* `` -> `...90`
* `` -> `...91`
* `` -> `...93`
* `` -> `...94`
* `` -> `...96`
* `` -> `...97`
* `` -> `...99`
* `` -> `...100`
* `` -> `...102`
* `` -> `...103`
* `` -> `...105`
* `` -> `...106`
* `` -> `...108`
* `` -> `...109`
* `` -> `...111`
* `` -> `...112`
* `` -> `...114`
* `` -> `...115`
* `` -> `...117`
* `` -> `...118`
* `` -> `...120`
* `` -> `...121`
* `` -> `...123`
* `` -> `...124`
* `` -> `...126`
* `` -> `...127`
* `` -> `...129`
* `` -> `...130`
* `` -> `...132`
* `` -> `...133`
* `` -> `...135`
* `` -> `...136`
* `` -> `...138`
* `` -> `...139`
* `` -> `...141`
* `` -> `...142`
* `` -> `...144`
* `` -> `...145`
* `` -> `...147`
* `` -> `...148`
* `` -> `...150`
* `` -> `...151`
* `` -> `...153`
* `` -> `...154`

```
## * `` -> `...156`
## * `` -> `...157`
## * `` -> `...159`
## * `` -> `...160`
## * `` -> `...162`
## * `` -> `...163`
## * `` -> `...165`
## * `` -> `...166`
## * `` -> `...168`
## * `` -> `...169`
## * `` -> `...171`
## * `` -> `...172`
## * `` -> `...174`
## * `` -> `...175`
## * `` -> `...177`
## * `` -> `...178`
## * `` -> `...180`
## * `` -> `...181`
## * `` -> `...183`
## * `` -> `...184`
## * `` -> `...186`
## * `` -> `...187`
## * `` -> `...189`
## * `` -> `...190`
## * `` -> `...192`
## * `` -> `...193`
## * `` -> `...195`
## * `` -> `...196`
## * `` -> `...198`
## * `` -> `...199`
## * `` -> `...201`
## * `` -> `...202`
## * `` -> `...204`
## * `` -> `...205`
## * `` -> `...207`
## * `` -> `...208`
## * `` -> `...210`
## * `` -> `...211`
```

```r
kia_hyundai_thefts <- read.csv("C:/Users/bobi/Documents/DSC 640/w5&6/kiaHyundaiThefts.csv")
car_thefts_map <- read.csv("C:/Users/bobi/Documents/DSC 640/w5&6/carTheftsMap.csv")
milwaukee_data <- read.csv("C:/Users/bobi/Documents/DSC 640/w5&6/KiaHyundaiMilwaukeeData.csv")

# View structure of datasets
str(vice_news_data)
```

```
## tibble [46 x 211] (S3: tbl_df/tbl/data.frame)
## $ ...1              : POSIXct[1:46], format: NA "2019-12-01" ...
## $ Denver            : chr [1:46] "Kia/Hyundais" "48.0" "21.0" "28.0" ...
## $ ...3              : chr [1:46] "All" "615.0" "519.0" "402.0" ...
## $ ...4              : chr [1:46] "Percent" "0.07804878049" "0.04046242775" "0.06965174129"
## $ El Paso           : chr [1:46] "Kia/Hyundais" "13.0" "9.0" "5.0" ...
## $ ...6              : chr [1:46] "All" "103.0" "95.0" "64.0" ...
## $ ...7              : chr [1:46] "Percent" "0.1262135922" "0.09473684211" "0.078125" ...
```

```
##  $ Portland            : chr [1:46] "Kia/Hyundais" "13.0" "12.0" "10.0" ...
##  $ ...9                : chr [1:46] "All" "592.0" "559.0" "498.0" ...
##  $ ...10               : chr [1:46] "Percent" "0.02195945946" "0.02146690519" "0.02008032129"
##  $ Atlanta             : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...12               : chr [1:46] "All" NA NA NA ...
##  $ ...13               : chr [1:46] "Percent" NA NA NA ...
##  $ Chicago             : chr [1:46] "Kia/Hyundais" "46.0" "36.0" "40.0" ...
##  $ ...15               : chr [1:46] "All" "767.0" "669.0" "653.0" ...
##  $ ...16               : chr [1:46] "Percent" "0.05997392438" "0.05381165919" "0.06125574273"
##  $ Virginia Beach      : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...18               : chr [1:46] "All" NA NA NA ...
##  $ ...19               : chr [1:46] "Percent" NA NA NA ...
##  $ Louisville          : chr [1:46] "Kia/Hyundais" "14.0" "18.0" "10.0" ...
##  $ ...21               : chr [1:46] "All" NA "301.0" "277.0" ...
##  $ ...22               : chr [1:46] "Percent" NA "0.05980066445" "0.03610108303" ...
##  $ San Diego           : chr [1:46] "Kia/Hyundais" "25.0" "14.0" "16.0" ...
##  $ ...24               : chr [1:46] "All" "430.0" "423.0" "411.0" ...
##  $ ...25               : chr [1:46] "Percent" "0.05813953488" "0.03309692671" "0.03892944039"
##  $ Sacramento          : chr [1:46] "Kia/Hyundais" "2.0" "9.0" "4.0" ...
##  $ ...27               : chr [1:46] "All" NA "192.0" "157.0" ...
##  $ ...28               : chr [1:46] "Percent" NA "0.046875" "0.02547770701" ...
##  $ Dallas              : chr [1:46] "Kia/Hyundais" "21.0" "26.0" "23.0" ...
##  $ ...30               : chr [1:46] "All" NA "1060.0" "874.0" ...
##  $ ...31               : chr [1:46] "Percent" NA "0.02452830189" "0.02631578947" ...
##  $ Fort Worth          : chr [1:46] "Kia/Hyundais" "12.0" "13.0" "4.0" ...
##  $ ...33               : chr [1:46] "All" "352.0" "340.0" "310.0" ...
##  $ ...34               : chr [1:46] "Percent" "0.03409090909" "0.03823529412" "0.01290322581"
##  $ Bakersfield         : chr [1:46] "Kia/Hyundais" "9.0" "13.0" "12.0" ...
##  $ ...36               : chr [1:46] "All" "360.0" "393.0" "323.0" ...
##  $ ...37               : chr [1:46] "Percent" "0.025" "0.03307888041" "0.03715170279" ...
##  $ Omaha, NE           : chr [1:46] "Kia/Hyundais" "11.0" "19.0" "13.0" ...
##  $ ...39               : chr [1:46] "All" "277.0" "291.0" "266.0" ...
##  $ ...40               : chr [1:46] "Percent" "0.03971119134" "0.06529209622" "0.04887218045"
##  $ Lubbock, TX         : chr [1:46] "Kia/Hyundais" "6.0" "2.0" "3.0" ...
##  $ ...42               : chr [1:46] "All" "130.0" "101.0" "100.0" ...
##  $ ...43               : chr [1:46] "Percent" "0.04615384615" "0.0198019802" "0.03" ...
##  $ Fremont, CA         : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...45               : chr [1:46] "All" "39.0" "63.0" "43.0" ...
##  $ ...46               : chr [1:46] "Percent" NA NA NA ...
##  $ San Antonio         : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...48               : chr [1:46] "All" NA NA NA ...
##  $ ...49               : chr [1:46] "Percent" NA NA NA ...
##  $ Irving, Texas       : chr [1:46] "Kia/Hyundais" "2.0" "3" "7" ...
##  $ ...51               : chr [1:46] "All" "58.0" "78" "71" ...
##  $ ...52               : chr [1:46] "Percent" "0.03448275862" "0.03846153846" "0.0985915493" .
##  $ Madison, WI         : chr [1:46] "Kia/Hyundais" "4.0" "4.0" "1.0" ...
##  $ ...54               : chr [1:46] "All" "83.0" "75.0" "45.0" ...
##  $ ...55               : chr [1:46] "Percent" "0.04819277108" "0.05333333333" "0.02222222222"
##  $ Frisco, TX          : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...57               : chr [1:46] "All" "12.0" "15.0" "8.0" ...
##  $ ...58               : chr [1:46] "Percent" NA NA NA ...
##  $ Lincoln, NE         : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...60               : chr [1:46] "All" "28.0" "62.0" "51.0" ...
##  $ ...61               : chr [1:46] "Percent" NA NA NA ...
```

```
##  $ Wichita, KS            : chr [1:46] "Kia/Hyundais" "3.0" "4.0" "5.0" ...
##  $ ...63                  : chr [1:46] "All" "182.0" "167.0" "179.0" ...
##  $ ...64                  : chr [1:46] "Percent" "0.01648351648" "0.02395209581" "0.02793296089"
##  $ Plano, TX              : chr [1:46] "Kia/Hyundais" "3.0" "1.0" "1.0" ...
##  $ ...66                  : chr [1:46] "All" "64.0" "42.0" "25.0" ...
##  $ ...67                  : chr [1:46] "Percent" "0.046875" "0.02380952381" "0.04" ...
##  $ Akron, OH              : chr [1:46] "Kia/Hyundais" "9.0" "1.0" "2.0" ...
##  $ ...69                  : chr [1:46] "All" "67.0" "62.0" "42.0" ...
##  $ ...70                  : chr [1:46] "Percent" "0.1343283582" "0.01612903226" "0.04761904762" .
##  $ Buffalo, NY            : chr [1:46] "Kia/Hyundais" "7.0" "7.0" "9.0" ...
##  $ ...72                  : chr [1:46] "All" "107.0" "136.0" "100.0" ...
##  $ ...73                  : chr [1:46] "Percent" "0.06542056075" "0.05147058824" "0.09" ...
##  $ Chula Vista, CA        : chr [1:46] "Kia/Hyundais" "6.0" "3.0" "2.0" ...
##  $ ...75                  : chr [1:46] "All" "77.0" "82.0" "50.0" ...
##  $ ...76                  : chr [1:46] "Percent" "0.07792207792" "0.03658536585" "0.04" ...
##  $ Reno, NV               : chr [1:46] "Kia/Hyundais" NA NA NA ...
##  $ ...78                  : chr [1:46] "All" "82.0" "99.0" "68.0" ...
##  $ ...79                  : chr [1:46] "Percent" NA NA NA ...
##  $ Oxnard, CA             : chr [1:46] "Kia/Hyundais" "1.0" "2.0" "0.0" ...
##  $ ...81                  : chr [1:46] "All" "48.0" "45.0" "28.0" ...
##  $ ...82                  : chr [1:46] "Percent" "0.02083333333" "0.04444444444" "0" ...
##  $ San Francisco          : chr [1:46] "Kia/Hyundais" "11.0" "10.0" "9.0" ...
##  $ ...84                  : chr [1:46] "All" "312.0" "270.0" "285.0" ...
##  $ ...85                  : chr [1:46] "Percent" "0.03525641026" "0.03703703704" "0.03157894737"
##  $ McKinney, TX           : chr [1:46] "Kia/Hyundais" "0.0" "0.0" "0.0" ...
##  $ ...87                  : chr [1:46] "All" "17.0" "16.0" "15.0" ...
##  $ ...88                  : chr [1:46] "Percent" "0" "0" "0" ...
##  $ Arlington, TX          : chr [1:46] "Kia/Hyundais" "8.0" "3.0" "6.0" ...
##  $ ...90                  : chr [1:46] "All" "101.0" "87.0" "75.0" ...
##  $ ...91                  : chr [1:46] "Percent" "0.07920792079" "0.03448275862" "0.08" ...
##  $ Garland, TX            : chr [1:46] "Kia/Hyundais" "3.0" "0.0" "0.0" ...
##  $ ...93                  : chr [1:46] "All" "112.0" "111.0" "78.0" ...
##  $ ...94                  : chr [1:46] "Percent" "0.02678571429" "0" "0" ...
##  $ Riverside County, CA   : chr [1:46] "Kia/Hyundais" "11.0" "10.0" "10.0" ...
##  $ ...96                  : chr [1:46] "All" "594.0" "583.0" "610.0" ...
##  $ ...97                  : chr [1:46] "Percent" "0.01851851852" "0.01715265866" "0.01639344262"
##  $ Stockton, CA           : chr [1:46] "Kia/Hyundais" "3.0" "5.0" "3.0" ...
##  $ ...99                  : chr [1:46] "All" "119.0" "167.0" "143.0" ...
##   [list output truncated]
```

```
str(kia_hyundai_thefts)
```

```
## 'data.frame':    552 obs. of  7 variables:
##  $ month               : chr  "Jan" "Feb" "Mar" "Apr" ...
##  $ year                : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ city                : chr  "Atlanta" "Atlanta" "Atlanta" "Atlanta" ...
##  $ state               : chr  "GA" "GA" "GA" "GA" ...
##  $ countKiaHyundaiThefts: int  17 11 18 15 16 14 19 12 11 12 ...
##  $ countOtherThefts    : int  264 205 181 223 277 220 267 242 234 206 ...
##  $ percentKiaHyundai   : num  0.06 0.051 0.09 0.063 0.055 0.06 0.066 0.047 0.045 0.055 ...
```

```
str(car_thefts_map)
```

```
## 'data.frame':    556 obs. of  9 variables:
##  $ agency_ori           : chr  "MO0490300" "" "TX06802" "MO0530000" ...
##  $ geo_name             : chr  "Carthage PD" "Warren County SO" "Odessa PD" "Laclede County SO" ..
##  $ countCarThefts2019   : chr  "62" "112" "499" "55" ...
##  $ countCarThefts2020   : chr  "58" "94" "464" "74" ...
##  $ countCarThefts2021   : chr  "47" "76" "375" "54" ...
##  $ countCarThefts2022   : chr  "26" "58" "288" "32" ...
##  $ latitude             : num  37.2 41.3 31.9 37.7 37.4 ...
##  $ longitude            : num  -94.3 -80.8 -102.3 -92.5 -89.7 ...
##  $ percentChange2019to2022: num  -0.581 -0.482 -0.423 -0.418 -0.4 ...
```

```r
str(milwaukee_data)
```

```
## 'data.frame':    48 obs. of  7 variables:
##  $ month                : chr  "Jan" "Feb" "Mar" "Apr" ...
##  $ year                 : int  2019 2019 2019 2019 2019 2019 2019 2019 2019 2019 ...
##  $ city                 : chr  "Milwaukee" "Milwaukee" "Milwaukee" "Milwaukee" ...
##  $ state                : chr  "WI" "WI" "WI" "WI" ...
##  $ countKiaHyundaiThefts: int  22 13 10 10 11 15 25 21 18 25 ...
##  $ countOtherThefts     : int  235 218 195 238 280 330 295 352 256 266 ...
##  $ percentKiaHyundai    : num  0.086 0.056 0.049 0.04 0.038 0.043 0.078 0.056 0.066 0.086 ...
```

Clean and Preprocess the Data

```r
# Clean VICE News Data
vice_clean <- vice_news_data %>%
  select(-starts_with("Unnamed")) %>%  # Remove irrelevant columns
  drop_na()  # Remove rows with missing values

# Clean Kia Hyundai Thefts
kia_hyundai_clean <- kia_hyundai_thefts %>%
  mutate(month_year = paste(month, year, sep = "-")) %>%
  select(city, state, countKiaHyundaiThefts, countOtherThefts, percentKiaHyundai, month_year)

# Clean Car Thefts Map
car_thefts_clean <- car_thefts_map %>%
  rename(geo_location = geo_name) %>%
  mutate(percent_change = as.numeric(percentChange2019to2022)) %>%
  drop_na()

# Clean Milwaukee Data
milwaukee_clean <- milwaukee_data %>%
  mutate(month_year = paste(month, year, sep = "-")) %>%
  select(city, state, countKiaHyundaiThefts, countOtherThefts, percentKiaHyundai, month_year)
```

Visualizations Visualization 1: Pie Chart (Proportion of Kia vs. Hyundai Thefts)

```r
# Aggregate Kia/Hyundai thefts
kia_vs_hyundai <- kia_hyundai_clean %>%
  summarise(total_kia_hyundai = sum(countKiaHyundaiThefts),
            total_other = sum(countOtherThefts))
```

```
# Create pie chart
pie_data <- c(kia_vs_hyundai$total_kia_hyundai, kia_vs_hyundai$total_other)
labels <- c("Kia/Hyundai", "Other Vehicles")

pie(pie_data, labels = labels, col = c("skyblue", "orange"), main = "Proportion of Kia/Hyundai vs. Other
```

Proportion of Kia/Hyundai vs. Other Vehicle Thefts
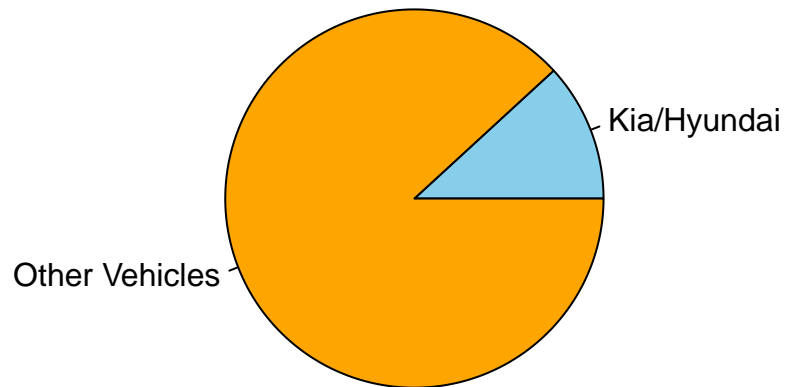
Visualization 2: Donut Chart (Thefts by Time Period)

```
library(ggplot2)

# Summarize data for donut chart
time_summary <- milwaukee_clean %>%
  group_by(month_year) %>%
  summarise(total_thefts = sum(countKiaHyundaiThefts))

# Donut chart
ggplot(time_summary, aes(x = 2, y = total_thefts, fill = month_year)) +
  geom_bar(stat = "identity", width = 1, color = "white") +
  coord_polar(theta = "y") +
  xlim(1, 2.5) +
  theme_void() +
  ggtitle("Thefts Over Time (Donut Chart)")
```

Thefts Over Time (Donut Chart)

month_year

Apr–2019	Jan–2019	May–2019
Apr–2020	Jan–2020	May–2020
Apr–2021	Jan–2021	May–2021
Apr–2022	Jan–2022	May–2022
Aug–2019	Jul–2019	Nov–2019
Aug–2020	Jul–2020	Nov–2020
Aug–2021	Jul–2021	Nov–2021
Aug–2022	Jul–2022	Nov–2022
Dec–2019	Jun–2019	Oct–2019
Dec–2020	Jun–2020	Oct–2020
Dec–2021	Jun–2021	Oct–2021
Dec–2022	Jun–2022	Oct–2022
Feb–2019	Mar–2019	Sep–2019
Feb–2020	Mar–2020	Sep–2020
Feb–2021	Mar–2021	Sep–2021
Feb–2022	Mar–2022	Sep–2022

Visualization 3: Stacked Bar Chart (Thefts by Neighborhood)

```
# Stacked bar chart
kia_hyundai_clean %>%
  group_by(city, state) %>%
  summarise(Kia_Hyundai_Thefts = sum(countKiaHyundaiThefts),
            Other_Thefts = sum(countOtherThefts)) %>%
  gather(key = "Type", value = "Count", -city, -state) %>%
  ggplot(aes(x = city, y = Count, fill = Type)) +
  geom_bar(stat = "identity") +
  theme_minimal() +
  labs(title = "Thefts by Neighborhood", x = "City", y = "Number of Thefts")
```

```
## `summarise()` has grouped output by 'city'. You can override using the
## `.groups` argument.
```
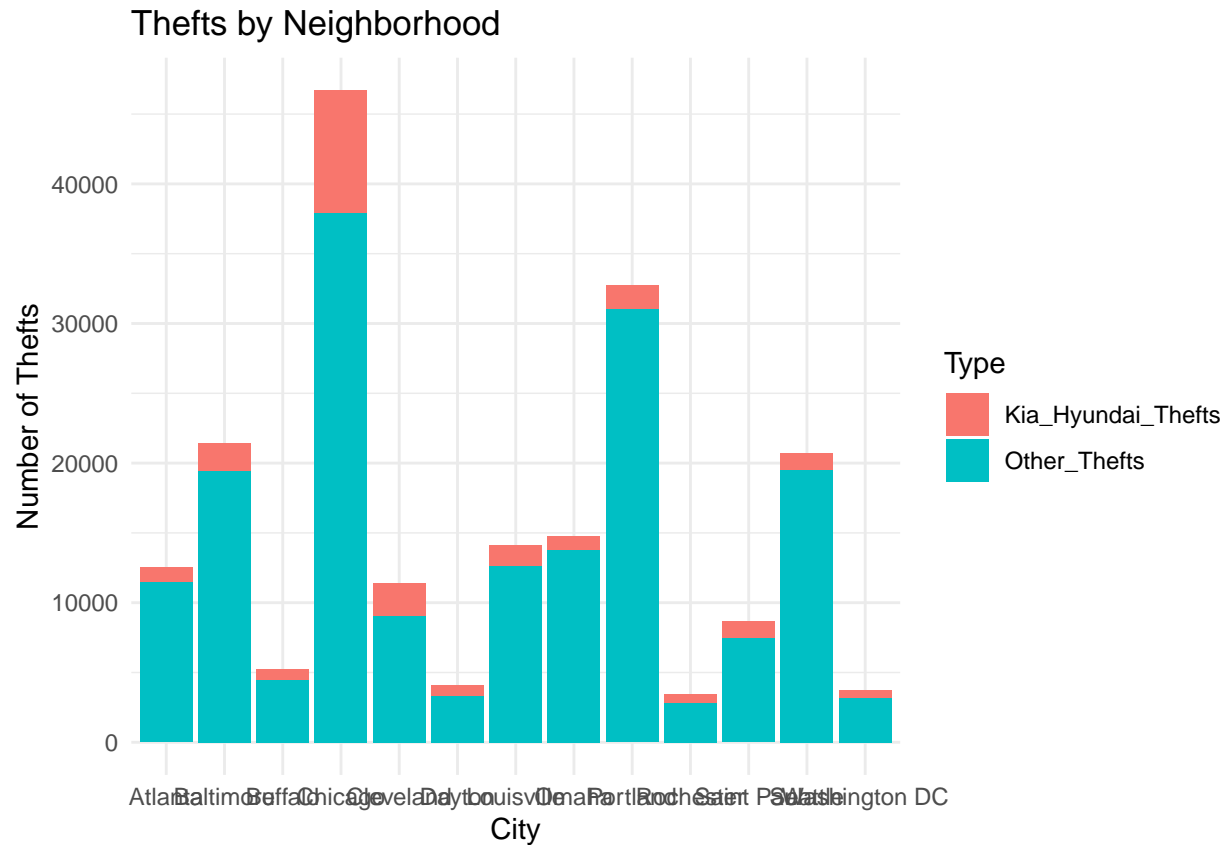
Visualization 4:Tree Map: Theft Distribution by Vehicle Type

```r
# Load necessary library
library(treemap)

# Data summarization
vehicle_summary <- milwaukee_clean %>%
  group_by(city) %>%
  summarise(Total_Thefts = sum(countKiaHyundaiThefts))

# Create treemap
treemap(vehicle_summary,
        index = "city",
        vSize = "Total_Thefts",
        title = "Theft Distribution by Vehicle Type",
        palette = "Blues")
```
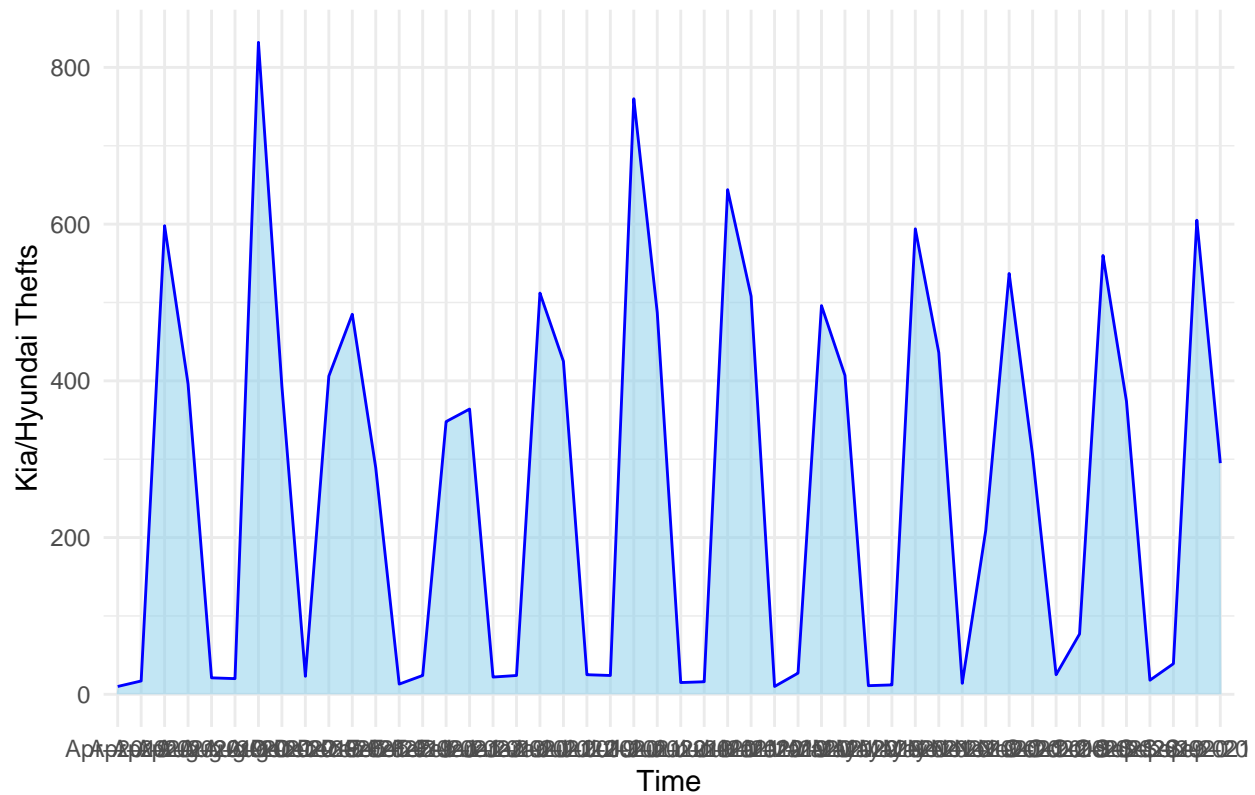
Theft Distribution by Vehicle Type
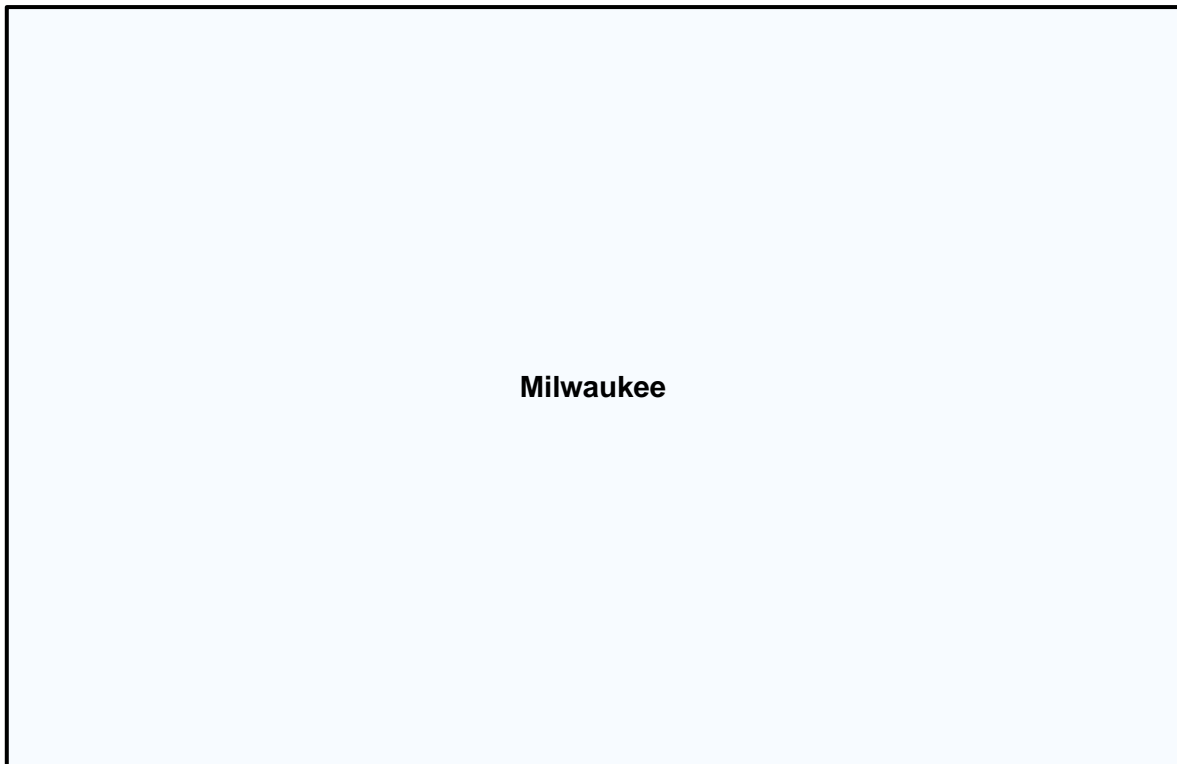
Milwaukee

Visualization 5: Area Chart (Theft Trends Over Time)

```r
# Area chart
ggplot(milwaukee_clean, aes(x = month_year, y = countKiaHyundaiThefts, group = 1)) +
  geom_area(fill = "skyblue", alpha = 0.5) +
  geom_line(color = "blue") +
  theme_minimal() +
  labs(title = "Theft Trends Over Time", x = "Time", y = "Kia/Hyundai Thefts")
```

Theft Trends Over Time


Visualization 6: Proportion of Thefts by Type and City

```r
# Multi-layer pie chart
multi_layer_data <- milwaukee_clean %>%
  group_by(city) %>%
  summarise(Kia_Hyundai_Thefts = sum(countKiaHyundaiThefts),
            Other_Thefts = sum(countOtherThefts)) %>%
  pivot_longer(cols = c(Kia_Hyundai_Thefts, Other_Thefts), names_to = "Type", values_to = "Count")

ggplot(multi_layer_data, aes(x0 = 0, y0 = 0, r0 = ifelse(Type == "Kia_Hyundai_Thefts", 0.6, 1), r = ifel
  geom_arc_bar(stat = "pie") +
  scale_fill_brewer(palette = "Set3") +
  coord_fixed() +
  labs(title = "Proportion of Thefts by Type and City", fill = "City") +
  theme_void() +
  theme(plot.title = element_text(hjust = 0.5))
```

Proportion of Thefts by Type and City

City

Milwaukee
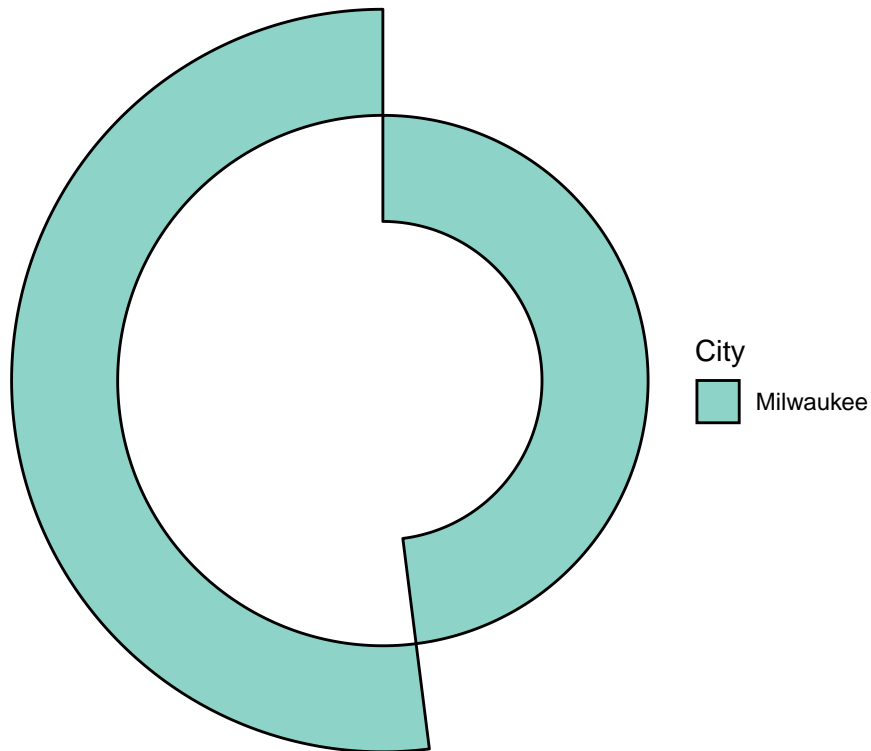
```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/pie_chart.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/donut_chart.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/stacked_bar_chart.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/tree_map.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/area_chart.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
ggsave("C:/Users/bobi/Documents/DSC 640/w5&6/stacked_area_chart.png")
```

Saving 6.5 x 4.5 in image