

DSC640weeks5and6

2025-07-13

Load required packages

```
library(tidyverse) library(readr) library(ggplot2) library(treemapify) library(scales)
```

Load each dataset

```
milwaukee <- read_csv("KiaHyundaiMilwaukeeData.csv") map_data <- read_csv("CarTheftsMap.csv") thefts <- read_csv("KiaHyundaiThefts.csv") vice <- read_excel("Motherboard VICE News Kia Hyundai Theft Data.xlsx")
```

Preview structure

```
glimpse(milwaukee) glimpse(map_data) glimpse(thefts) glimpse(vice)
```

Remove missing values

```
milwaukee_clean <- na.omit(milwaukee) map_data_clean <- na.omit(map_data) thefts_clean <- na.omit(thefts) vice_clean <- na.omit(vice)
```

```
make_summary <- thefts_clean %>% count(Make) %>% mutate(perc = n / sum(n))
```

```
ggplot(make_summary, aes(x = "", y = perc, fill = Make)) + geom_col(width = 1) + coord_polar(theta = "y") + labs(title = "Share of Thefts by Vehicle Make") + theme_void() # Summarize theft counts by geo_name (state or region) state_summary <- map_data_clean %>% count(geo_name)
```

Create a donut chart

```
ggplot(state_summary, aes(x = 2, y = n, fill = geo_name)) + geom_col(width = 1) + coord_polar("y", start = 0) + xlim(0.5, 2.5) + theme_void() + labs(title = "Vehicle Thefts by Region/State") + theme(legend.position = "right") # Reshape data for stacked bar chart stacked_data <- thefts_clean %>% select(year, countKiaHyundaiThefts, countOtherThefts) %>% pivot_longer(cols = starts_with("count"), names_to = "Make", values_to = "Thefts") %>% mutate(Make = case_when( Make == "countKiaHyundaiThefts" ~ "Kia/Hyundai", Make == "countOtherThefts" ~ "Other" ))
```

Plot stacked bar chart

```
ggplot(stacked_data, aes(x = factor(year), y = Thefts, fill = Make)) + geom_bar(stat = "identity") + labs(title = "Vehicle Thefts by Year and Type", x = "Year", y = "Number of Thefts") + theme_minimal()
```

Make sure Date column exists

```
thefts_clean <- thefts_clean %>% mutate(Date = as.Date(paste(year, month, "01", sep = "-")))
```

Summarize total thefts per month

```
monthly_thefts <- thefts_clean %>% mutate(TotalThefts = countKiaHyundaiThefts + countOtherThefts) %>%
group_by(Date) %>% summarise(Thefts = sum(TotalThefts))
```

Plot area chart

```
ggplot(monthly_thefts, aes(x = Date, y = Thefts)) + geom_area(fill = "steelblue", alpha = 0.7) + labs(title = "Monthly
Vehicle Theft Trend Over Time", x = "Date", y = "Total Thefts") + theme_minimal()
```

Ensure Date column is present

```
thefts_clean <- thefts_clean %>% mutate(Date = ymd(paste(year, month, "01", sep = "-")))
```

Reshape to long format by type

```
stacked_area_data <- thefts_clean %>% select(Date, countKiaHyundaiThefts, countOtherThefts) %>%
pivot_longer(cols = starts_with("count"), names_to = "Make", values_to = "Thefts") %>% mutate(Make =
case_when( Make == "countKiaHyundaiThefts" ~ "Kia/Hyundai", Make == "countOtherThefts" ~ "Other" ))
```

Plot stacked area chart

```
ggplot(stacked_area_data, aes(x = Date, y = Thefts, fill = Make)) + geom_area(position = "stack", alpha = 0.8) +
labs(title = "Monthly Vehicle Thefts by Type", x = "Date", y = "Number of Thefts") + theme_minimal()
```

```
library(treemapify)
```

Create summary of top 10 cities across entire dataset

```
city_summary <- map_data_clean %>% count(geo_name, sort = TRUE) %>% top_n(10, n)
```

Tree map of top cities

```
ggplot(city_summary, aes(area = n, fill = geo_name, label = geo_name)) + geom_treemap() +
geom_treemap_text(colour = "white", place = "centre", size = 10) + labs(title = "Top 10 Cities for Vehicle Thefts") +
theme(legend.position = "none")
```