

Assignment 3.1

Mateus Parolin Gomes

Reflection

The readings from “Chapter 1 in Kieran Healy, Data Visualization” provided an engaging reminder of the importance of exploring and summarizing datasets before diving into analysis, specially the emphasis on understanding data structure, spotting outliers, and asking critical question. One example that stood out was the use of clear visual summaries to assess variable distributions, something I often practiced with Python. As someone learning R for the first time, I was struck by how familiar the process feels. Installing R and libraries (or packages) mirrors setting up Python environments. RStudio, much like VSCode, simplifies workflow management with project organization. This course is helping me reframe my analytical skills in a new context. Seeing parallels between R and Python makes the transition smoother and reinforces the value of adaptable problem-solving across languages.

```
# My first plot

# Import Libraries
library(ggplot2)
library(dplyr)

##
## 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

# Load CSV file
cars <- read.csv("./data/cars.csv")

# Avg hwy by manufacturer
avg_hwy_by_manufacturer <- cars %>%
  group_by(manufacturer) %>%
  summarise(avg_hwy = mean(hwy, na.rm = TRUE))

# Create scatterplot
ggplot(data = avg_hwy_by_manufacturer, aes(y = manufacturer, x = avg_hwy)) +
```

```
geom_point(color = "blue", size = 3) +
geom_line(color = "darkblue", linetype = "dashed") +
labs(
  title = "Avg Highway Miles per Gallon by Manufacturer",
  y = "Manufacturer",
  x = "Avg Highway Miles per Gallon"
) +
theme_minimal()
```

`geom_line()`: Each group consists of only one observation.
 ## **i** Do you need to adjust the group aesthetic?

