

4_2023900021

hrithik.rachakonda@research.iiit.ac.in

2026-02-02

Question 4) Fast and Furious: Heatmap

Sheet 5 (Fast and Furious) contains data pertaining to automobiles enlisting features such as a car's brand, miles per gallon, manufactured year etc. Plot a heatmap for different types of correlations between the features present in the dataset and write about the inferences you can make from it. Also justify as to why one is better than the other. HINT: Make sure you convert all types to numeric and handle missing values.

Answer)

Question 4: Fast and Furious – Heatmap

To analyze relationships between automobile features, three correlation heatmaps were plotted after converting all variables to numeric and handling missing values: **Pearson**, **Spearman**, and **Kendall** correlations.

Key Observations

- **mpg** shows a **strong negative correlation** with **cylinders, cubic inches, horsepower, and weight**, indicating that larger and heavier cars are less fuel-efficient.
- **Cylinders, cubic inches, horsepower, and weight** are **strongly positively correlated**, suggesting these engine-related features increase together.
- **Year** is **positively correlated with mpg** and **negatively correlated with engine size and weight**, showing that newer cars tend to be lighter and more fuel-efficient.
- **Time-to-60** has a **negative correlation with horsepower and engine size**, indicating that more powerful cars accelerate faster.

Comparison of Correlation Methods

- **Pearson correlation** captures **linear relationships** and provides the clearest interpretation for this dataset.
- **Spearman correlation** confirms that the same trends hold in a **rank-based (monotonic)** sense and is less sensitive to outliers.
- **Kendall correlation** is more conservative, showing similar directions but weaker magnitudes.

Conclusion

The **Pearson correlation heatmap** is the **most suitable** for this dataset, as automobile features exhibit strong linear relationships. Spearman and Kendall serve as supporting checks, but Pearson best highlights the strength and structure of associations among vehicle attributes.

```

# =====
# Q4: Fast and Furious - Correlation Heatmaps
# =====

# install.packages(c("readxl", "dplyr", "ggplot2", "reshape2"))
library(readxl)
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

library(ggplot2)
library(reshape2)

# ---- Load data ----
xlsx_path <- "BRSM_Visualisation_Assignment.xlsx"
data <- read_excel(xlsx_path, sheet = "Fast and Furious")
data <- as.data.frame(data)

# ---- Convert to numeric where possible ----
data_num <- data %>%
  mutate(across(everything(), ~ suppressWarnings(as.numeric(.))))

# ---- Remove columns with all NA ----
data_num <- data_num[, colSums(is.na(data_num)) < nrow(data_num)]

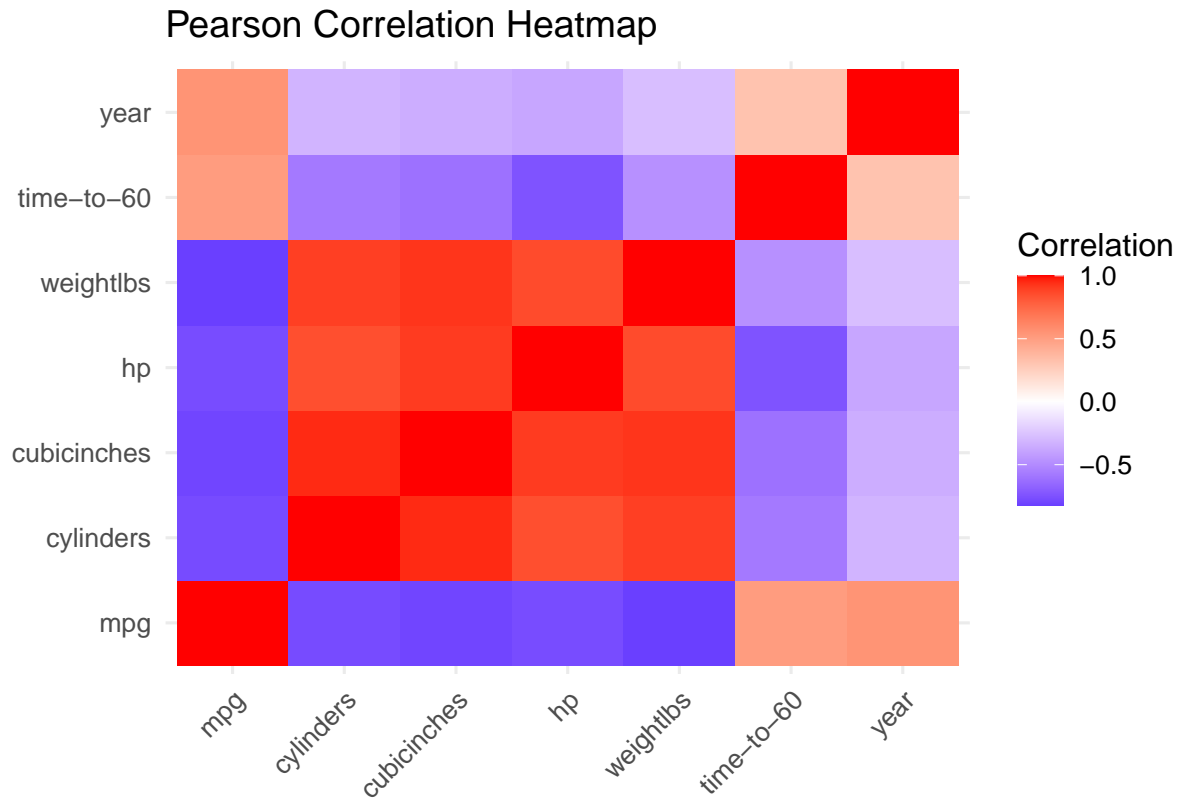
# ---- Handle missing values (median imputation) ----
data_num <- data_num %>%
  mutate(across(everything(), ~ ifelse(is.na(.), median(., na.rm = TRUE), .)))

# ---- Correlation matrices ----
cor_pearson <- cor(data_num, method = "pearson")
cor_spearman <- cor(data_num, method = "spearman")
cor_kendall <- cor(data_num, method = "kendall")

# ---- Helper function for heatmap ----
plot_heatmap <- function(cor_mat, title) {
  melt(cor_mat) %>%
    ggplot(aes(Var1, Var2, fill = value)) +
    geom_tile() +
    scale_fill_gradient2(low = "blue", mid = "white", high = "red", midpoint = 0) +
    labs(title = title, x = "", y = "", fill = "Correlation") +
    theme_minimal(base_size = 12) +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

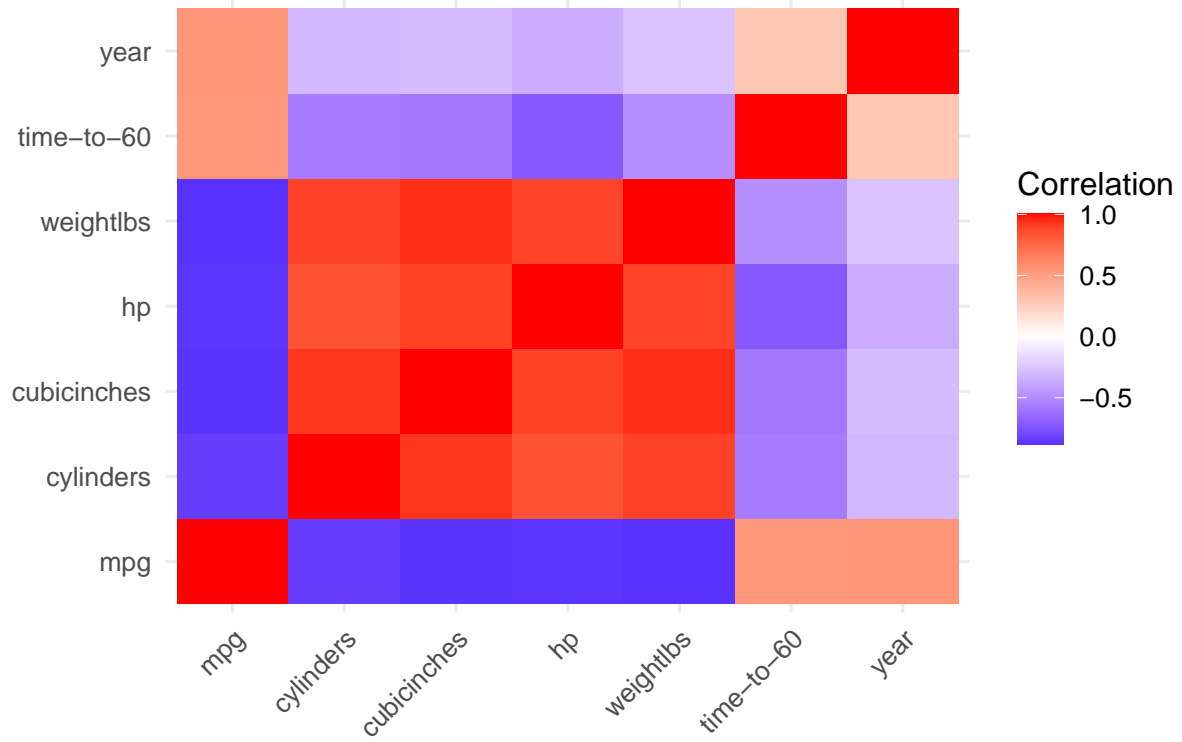
```

```
}
# ---- Plot heatmaps ----
plot_heatmap(cor_pearson, "Pearson Correlation Heatmap")
```



```
plot_heatmap(cor_spearman, "Spearman Correlation Heatmap")
```

Spearman Correlation Heatmap



```
plot_heatmap(cor_kendall, "Kendall Correlation Heatmap")
```

Kendall Correlation Heatmap

