

Rworksheet_Apuli#4c

Marian Apuli

2025-12-12

WORKSHEET – R PROGRAMMING (mtcars dataset)

Load packages

```
library(dplyr) library(ggplot2) library(stringr)
```

Prepare mtcars

```
data(mtcars) mtcars$model <- rownames(mtcars) rownames(mtcars) <- NULL
```

PART 1 – EXPLORATION

1. Structure and preview

```
str(mtcars) head(mtcars)
```

2. Categorical variables (common in mtcars)

```
categorical_vars <- c("cyl", "gear", "carb") categorical_vars
```

3. Continuous variables

```
continuous_vars <- c("mpg", "disp", "hp", "drat", "wt", "qsec") continuous_vars
```

4. Extract manufacturer from model name

```
mtcars$manufacturer <- word(mtcars$model, 1)
```

Count manufacturers

```
table(mtcars$manufacturer)
```

PART 2 – QUESTIONS ON MANUFACTURERS & MODELS

5. Manufacturer with most models

```
mtcars %>% count(manufacturer, sort = TRUE)
```

6. Model with most variations

```
mtcars %>% count(model, sort = TRUE)
```

7. Unique models per manufacturer

```
mtcars %>% group_by(manufacturer) %>% summarize(unique_models = n_distinct(model))
```

PART 3 – GRAPHS

8. Histogram of MPG

```
ggplot(mtcars, aes(mpg)) + geom_histogram(binwidth = 3, fill = "steelblue") + ggtitle("Histogram of MPG")
```

9. Histogram of QSEC

```
ggplot(mtcars, aes(qsec)) + geom_histogram(binwidth = 1.5, fill = "orange") + ggtitle("Histogram of QSEC")
```

10. Scatterplot: displacement vs mpg

```
ggplot(mtcars, aes(disp, mpg, color = manufacturer)) + geom_point(size = 3) + ggtitle("Displacement vs MPG")
```

11. Top 10 cars by MPG

```
mtcars %>% slice_max(mpg, n = 10) %>% ggplot(aes(reorder(model, mpg), mpg, fill = mpg)) + geom_col() + coord_flip() + ggtitle("Top 10 Cars by MPG")
```

12. Group by gear and cyl

```
mtcars %>% group_by(gear, cyl) %>% summarize(count = n())
```

13. Scatterplot: cylinders vs displacement

```
ggplot(mtcars, aes(cyl, disp, color = factor(gear))) + geom_point(size = 3) + ggtitle("Cylinders vs Displacement")
```

14. Displacement vs horsepower (with regression line)

```
ggplot(mtcars, aes(disp, hp)) + geom_point(size = 3) + geom_smooth(method = "lm") + ggtitle("Displacement vs Horsepower")
```

15. Number of cars per gear

```
mtcars %>% count(gear) %>% ggplot(aes(factor(gear), n, fill = factor(gear))) + geom_col() + ggtitle("Number of Cars Per Gear")
```

PART 4 – TRAFFIC DATA (SAFE PLACEHOLDER)

NOTE: Replace “traffic.csv” with your file name.

This code will NOT break even if file is missing.

```
if (file.exists("traffic.csv")) {  
  traffic <- read.csv("traffic.csv")  
  # Number of rows and columns nrow(traffic) ncol(traffic)  
  # Structure str(traffic)  
  # Select intersection and violations traffic_sel <- traffic %>% select(intersection, violations)  
  head(traffic_sel)  
  # Boxplot ggplot(traffic_sel, aes(intersection, violations)) + geom_boxplot() + coord_flip() + ggtitle("Violations by Intersection")  
  # Summary summary(traffic)  
  # Total violations per intersection traffic %>% group_by(intersection) %>% summarize(total_violations =  
  # sum(violations)) %>% ggplot(aes(reorder(intersection, total_violations), total_violations)) + geom_col(fill = "skyblue") + coord_flip() + ggtitle("Total Violations Per Intersection")  
} else { print("traffic.csv not found — please place the file in your working directory.") }
```

PART 5 – EXTRA ANALYSIS

16. Line plot: model vs weight

```
ggplot(mtcars, aes(model, wt, group = 1)) + geom_line() + theme(axis.text.x = element_text(angle = 90))  
+ ggtitle("Model vs Weight")
```

17. Correlation matrix

```
cor(mtcars[, c("mpg", "disp", "hp", "wt", "qsec")])
```