Introduction to R markdown

20015367

07/10/2025

# Firstly check you have any packages required in the session, the only packaged required in this session is ggplot2.

# Run the following code, which double-checks that you have the packages installed,

# and loads them into the R session.

if(!require(ggplot2)){install.packages("ggplot2")}

## Loading required package: ggplot2

library(ggplot2)

### Task 1 Introduction to R markdown

This is an R markdown document. Firstly, knit this document to Word, see what happens!

For this exercise we will use data embedded within R, mtcars, start by exploring the structure of the dataset and the variables

str(mtcars)

## 'data.frame': 32 obs. of 11 variables:  
## $ mpg : num 21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...  
## $ cyl : num 6 6 4 6 8 6 8 4 4 6 ...  
## $ disp: num 160 160 108 258 360 ...  
## $ hp : num 110 110 93 110 175 105 245 62 95 123 ...  
## $ drat: num 3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...  
## $ wt : num 2.62 2.88 2.32 3.21 3.44 ...  
## $ qsec: num 16.5 17 18.6 19.4 17 ...  
## $ vs : num 0 0 1 1 0 1 0 1 1 1 ...  
## $ am : num 1 1 1 0 0 0 0 0 0 0 ...  
## $ gear: num 4 4 4 3 3 3 3 4 4 4 ...  
## $ carb: num 4 4 1 1 2 1 4 2 2 4 ...

names(mtcars)

## [1] "mpg" "cyl" "disp" "hp" "drat" "wt" "qsec" "vs" "am" "gear"  
## [11] "carb"

In this workshop we will concentrate on using mpg (miles per gallon); cyl (number of cylinders) and hp (horsepower).

First, find out more about the dataset using the help file, e.g. with the command **?mtcars**

Next, look at the first 6 rows of data using the command **head**

head(mtcars)

## mpg cyl disp hp drat wt qsec vs am gear carb  
## Mazda RX4 21.0 6 160 110 3.90 2.620 16.46 0 1 4 4  
## Mazda RX4 Wag 21.0 6 160 110 3.90 2.875 17.02 0 1 4 4  
## Datsun 710 22.8 4 108 93 3.85 2.320 18.61 1 1 4 1  
## Hornet 4 Drive 21.4 6 258 110 3.08 3.215 19.44 1 0 3 1  
## Hornet Sportabout 18.7 8 360 175 3.15 3.440 17.02 0 0 3 2  
## Valiant 18.1 6 225 105 2.76 3.460 20.22 1 0 3 1

# Examples of embedding a calculated value into sentences:

The mean mpg for cars in this dataset is 20.090625.

Edit the statement so that it only prints to 2 decimal places:

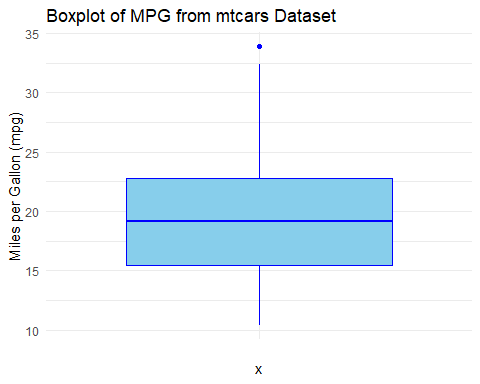
The mean mpg for cars in this dataset is 20.09.

# Question. What other summary statistics may you wish to report within the text of a document?

# Example graphics

Creating a boxplot of miles per gallaon (using geom\_boxplot)

ggplot(mtcars, aes(x = "", y = mpg)) +  
 geom\_boxplot(fill = "skyblue", color = "blue") +  
 labs(title = "Boxplot of MPG from mtcars Dataset",  
 y = "Miles per Gallon (mpg)") +  
 theme\_minimal()

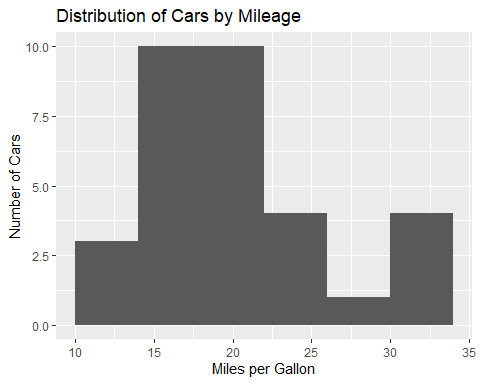


In this code:

* ggplot(mtcars, aes(x = ““, y = mpg)): This sets up the basic ggplot object, specifying the dataset (mtcars) and aesthetics (mapping the empty string”” to the x-axis and the ‘mpg’ variable to the y-axis).
* geom\_boxplot(fill = “skyblue”, color = “blue”): This adds the boxplot layer to the ggplot object, specifying the fill and border color.
* labs(title = “Boxplot of MPG from mtcars Dataset”, y = “Miles per Gallon (mpg)”): This sets the plot title and y-axis label.
* theme\_minimal(): This applies a minimal theme to the plot for better aesthetics.

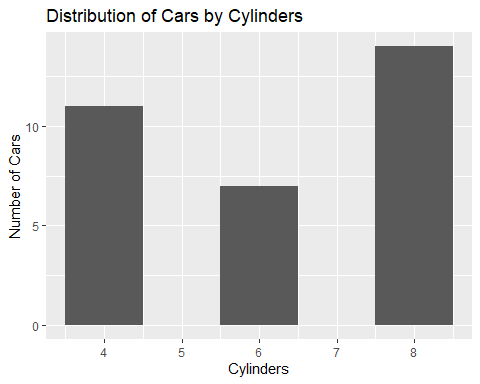
Creating a histogram (using geom\_histogram), of miles per gallon and number of cars:

ggplot(mtcars,aes(mpg))+geom\_histogram(binwidth = 4)+xlab("Miles per Gallon")+ylab("Number of Cars")+ggtitle("Distribution of Cars by Mileage")



Editing this to show the histogram for the number of cylinders (edit the binwidth to be equal to 1).

ggplot(mtcars,aes(cyl))+geom\_histogram(binwidth = 1)+xlab("Cylinders")+ylab("Number of Cars")+ggtitle("Distribution of Cars by Cylinders")



### Task 2. Dewis

This is the return of the boxplot and histrogram Dewis exercise, but now we include the data for you to do some analyses within R.

As per instructions in Dewis, can you create a new R markdown file, and knit a word document with your answers?

### Task 3. Reproducible research

List ‘good practice’ or standards of reproducible research that you have evidenced in this document (and/or in your R markdown file for the Dewis question).

What other ‘good practice’ or standards of reproducible research might be appropriate?

# Question: When might you use the ‘echo = false’ command?

### Task 4. Data Analyses plan

Consider the potential research questions that could arise with the mtcars dataset.

Prepare a data analyses plan to address these research questions.

Open a new R markdown file and prepare a polished report with the analyses as per your statistical analyses plan.