Q1 R code and Result

STAT302 - Assignment 1

Deadline: 2025-05-30 at 10PM on Crowdmark

Q1 [8 pts]: R Coding with the Sparrow Dataset

This data is a stratified random sample of 116 Savannah sparrows at Kent Island. The **weight (in grams)** and **wing length (in mm)** were obtained for birds from nests that were reduced, controlled, or enlarged. The data are available in the file `Sparrows` from the R package `Stat2Data`. Fit a simple linear regression model to predict **weight (Y)** from **wing length (X)**.

a. [2 pts] Load in the `Sparrows` data set from the R package `Stat2Data`, using the `str()` function to inspect data structure and further display the first 5 rows with `head()`.
b. [1 pt] Create a scatterplot of weight versus wing length using `ggplot()`, and add a fitted regression line. Refer to the example from the lecture slides.
c. [1 pt] Fit a simple linear regression and display the summary output using `summary()`.
d. [2 pts] Use `par(mfrow=c(1,2))` to display the following two diagnostic plots side-by-side:
   - the residual vs fitted (`which=1`)
   - Normal Q-Q (`which=2`)
e. [2 pts] Use `par(mfrow=c(1,2))` to display the following two diagnostic plots side-by-side:
   - Cook's distance (`which=4`)
   - Residual vs Leverage (`which=5`)

```r
# Remove the # sign to run the code below
# to install Stat2Data package for once
# install.packages('Stat2Data', repos='https://mirror.csclub.uwaterloo.ca/CRAN/')

library(Stat2Data)

# (a) show data structure and display the first 5 rows
data("Sparrows")


# (b) using ggplot to form the scatterplot
library(ggplot2)


# (c) Fit a regression model and output the summary


# (d) display the first diagnostic plots
par(mfrow=c(1,2))
# R code here

par(mfrow=c(1,1))

# (e) display the Cook's distance and Residual vs Leverage plots
par(mfrow=c(1,2))
```

```r
# R code here

par(mfrow=c(1,1))
```