Assignment 1 - Penguins Dataset

Kristen Durkin

Purpose

Set up R/RStudio/GitHub and demonstrate basic data analysis steps in R using a dataset.

Dataset and Source

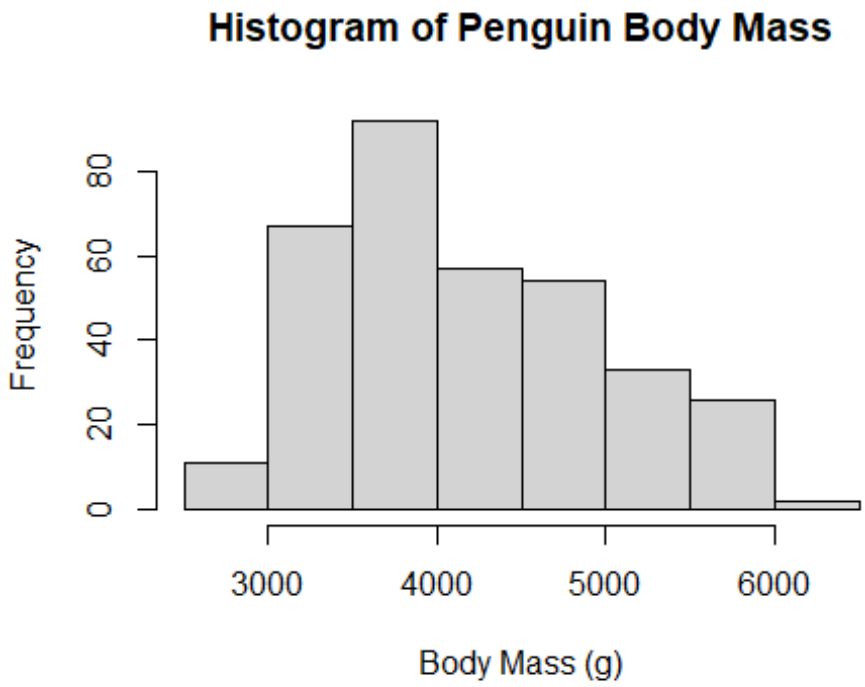
We use the Palmer Penguins dataset (with a mix of quantitative + categorical variables).
Source:
https://raw.githubusercontent.com/allisonhorst/palmerpenguins/main/inst/extdata/penguins.csv

Import dataset

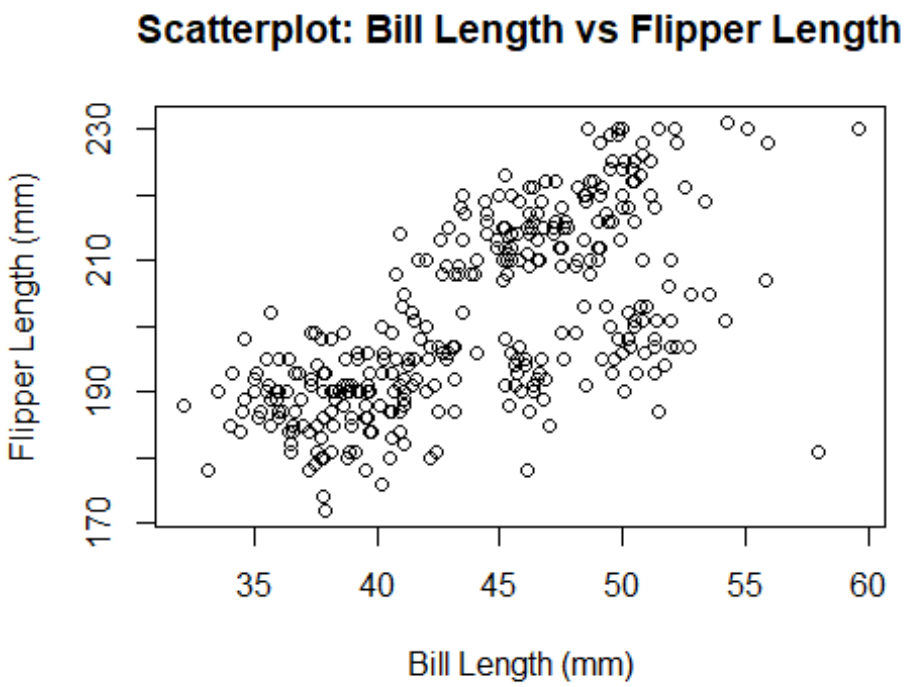
```
penguins <- read.csv(

"https://raw.githubusercontent.com/allisonhorst/palmerpenguins/main/inst/extdata/penguins.csv"
)

## Plot Points
summary(penguins$bill_length_mm) #Quantitative

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   32.10   39.23   44.45   43.92   48.50   59.60       2

summary(penguins$body_mass_g)   #Quantitative

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    2700    3550    4050    4202    4750    6300       2

table(penguins$species)         #Categorical

##
##    Adelie Chinstrap    Gentoo
##       152        68       124

table(penguins$sex)             #Categorical

##
## female   male
##    165    168

# Transform Var
penguins$body_mass_z <- scale(penguins$body_mass_g)
head(penguins$body_mass_z)
```

```
##             [,1]
## [1,] -0.5633167
## [2,] -0.5009690
## [3,] -1.1867934
## [4,]         NA
## [5,] -0.9374027
## [6,] -0.6880121
```

```r
# Quantitative Plot
hist(penguins$body_mass_g,
     main = "Histogram of Penguin Body Mass",
     xlab = "Body Mass (g)")
```




```r
# Scatterplot
plot(penguins$bill_length_mm, penguins$flipper_length_mm,
     xlab = "Bill Length (mm)",
     ylab = "Flipper Length (mm)",
     main = "Scatterplot: Bill Length vs Flipper Length")
```


Scatterplot: Bill Length vs Flipper Length
Flipper Length (mm)
170
190
210
230
35
40
45
50
55
60
Bill Length (mm)