Assignment2

Jen Rose

2025-10-25

QAC380 Assignment 2 Descriptive Statistics and Plots

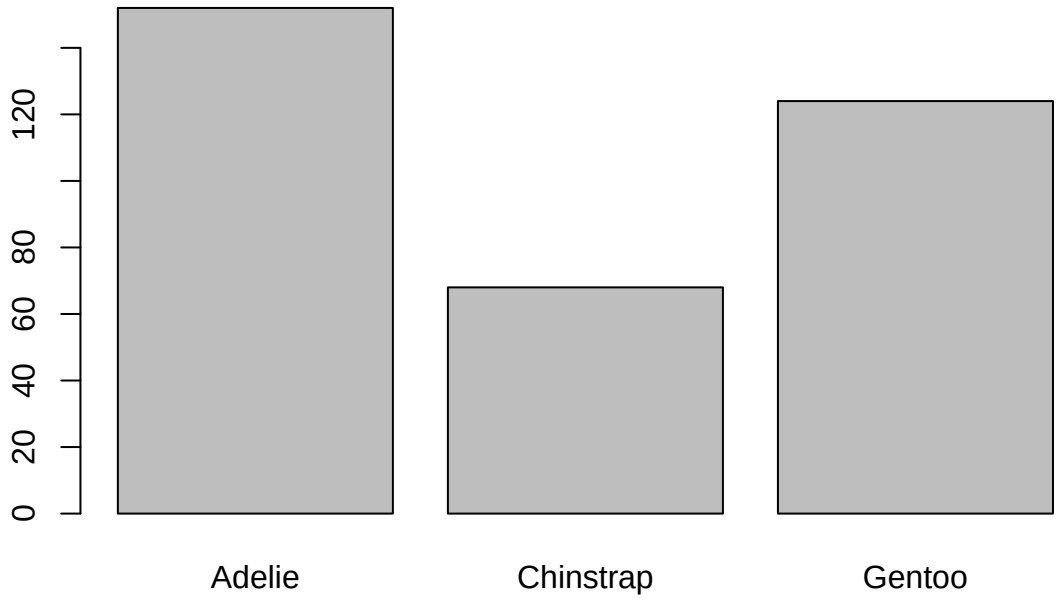
```
# load packages
library(readr)
library(dplyr)
library(ggplot2)
library(Hmisc)
library(descr)

# call in the data set
penguins <- read_csv("C:/Users/Jen Rose/Documents/GitHub/assignment2-JR/penguins.csv")

# data structure
str(penguins)
```

```
## spc_tbl_ [344 x 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ species          : chr [1:344] "Adelie" "Adelie" "Adelie" "Adelie" ...
##  $ island           : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...
##  $ bill_length_mm   : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...
##  $ bill_depth_mm    : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...
##  $ flipper_length_mm: num [1:344] 181 186 195 NA 193 190 181 195 193 190 ...
##  $ body_mass_g      : num [1:344] 3750 3800 3250 NA 3450 ...
##  $ sex              : chr [1:344] "male" "female" "female" NA ...
##  $ year             : num [1:344] 2007 2007 2007 2007 2007 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   species = col_character(),
##   ..   island = col_character(),
##   ..   bill_length_mm = col_double(),
##   ..   bill_depth_mm = col_double(),
##   ..   flipper_length_mm = col_double(),
##   ..   body_mass_g = col_double(),
##   ..   sex = col_character(),
##   ..   year = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
# descriptives
describe(penguins$body_mass_g)
```

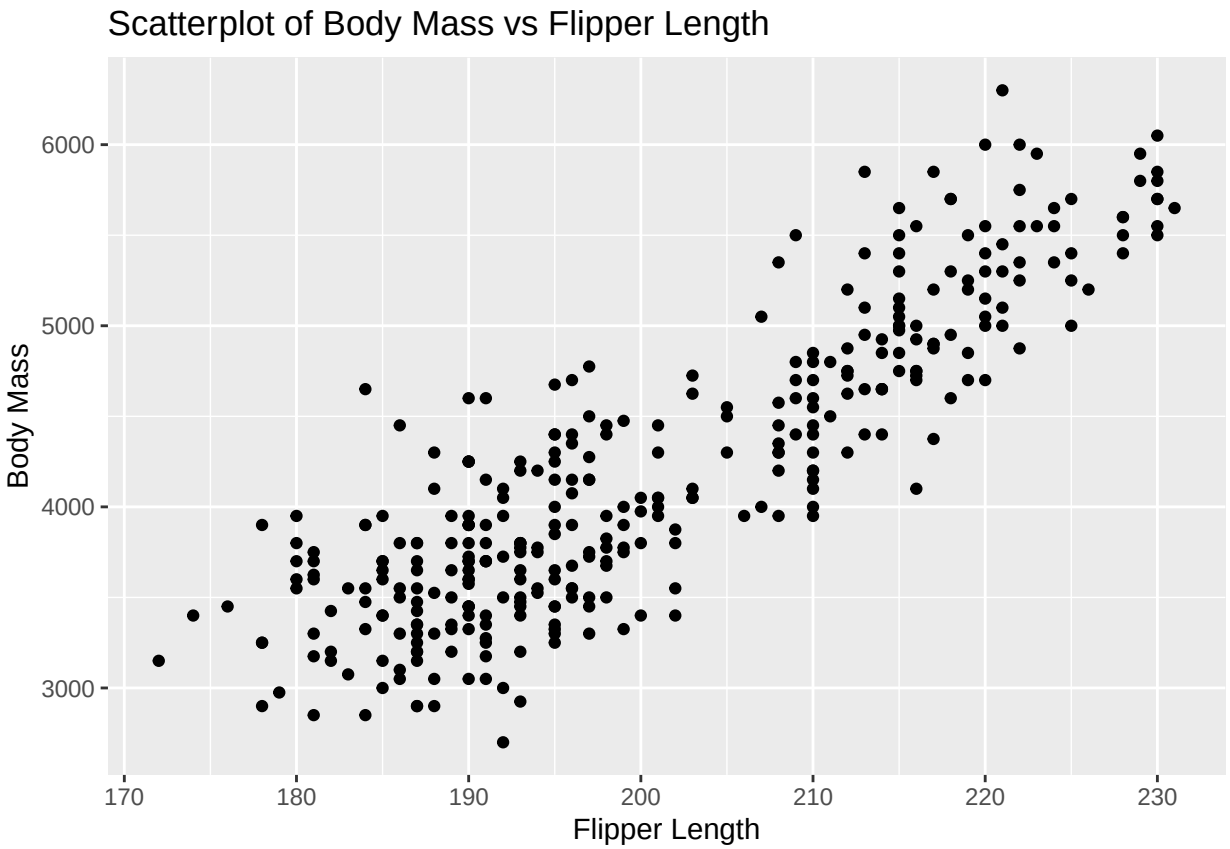
```
## penguins$body_mass_g
##        n  missing distinct     Info     Mean      Gmd      .05      .10
##      342        2       94        1     4202    911.8     3150     3300
##      .25      .50      .75      .90      .95
##     3550     4050     4750     5400     5650
##
## lowest : 2700 2850 2900 2925 2975, highest: 5850 5950 6000 6050 6300
```

```
freq(penguins$species)
```




```
## penguins$species
##           Frequency Percent
## Adelie          152   44.19
## Chinstrap        68   19.77
## Gentoo          124   36.05
## Total           344  100.00
```

```
# scatterplot body mass by flipper length
ggplot(penguins, aes(x=flipper_length_mm, y=body_mass_g)) +
  geom_point() +
  labs(title = "Scatterplot of Body Mass vs Flipper Length",
       x = "Flipper Length",
       y = "Body Mass")
```


Scatterplot of Body Mass vs Flipper Length
Body Mass
6000
5000
4000
3000
170
180
190
200
210
220
230
Flipper Length