# Statistical Exercises on Cat Data

## Ruben

## April 1, 2025

## Exercise 1: Familiarization with the dataset

- **(1)** What are the variable names?

    - "Sex", "Bwt", "Hwt"

- **(2)** How many variables and observations does the `cats` dataset contain?

    - There are 144 observations and 3 variables.

- **(3)** Display the first 10 observations of the table. What is the sex and heart weight of cat number 6?

    - Sex: Female, Weight: 2.1 kg, Heart weight: 7.6 g

```
> attach(cats)
> Bwt
  [1] 2.0 2.0 2.0 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.1 2.2 2.2 2.2 2.2 2.2 2.2
 [19] 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.4 2.4 2.4 2.4 2.5 2.5
 [37] 2.6 2.6 2.6 2.7 2.7 2.7 2.9 2.9 2.9 3.0 3.0 2.0 2.0 2.1 2.2 2.2 2.2 2.2
 [55] 2.2 2.2 2.2 2.2 2.3 2.4 2.4 2.4 2.4 2.4 2.5 2.5 2.5 2.5 2.5 2.5 2.5 2.5
 [73] 2.6 2.6 2.6 2.6 2.6 2.6 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.7 2.8 2.8 2.8
 [91] 2.8 2.8 2.8 2.8 2.9 2.9 2.9 2.9 2.9 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0 3.0
[109] 3.1 3.1 3.1 3.1 3.1 3.1 3.2 3.2 3.2 3.2 3.2 3.2 3.3 3.3 3.3 3.3 3.3 3.4
[127] 3.4 3.4 3.4 3.4 3.5 3.5 3.5 3.5 3.5 3.6 3.6 3.6 3.6 3.7 3.8 3.8 3.9 3.9
```

## Exercise 2: Characterization of statistical series

- Empirical mean of the Bwt variable:

- **Mean** = 2.723611

- Median and quartiles:

  - Median: 2.7
  - 1st Quartile (Q1): 2.3
  - 3rd Quartile (Q3): 3.025

- Variance:

  - Variance = 0.2355225

- Standard deviation:

  - Standard deviation = 0.4853066

- Interquartile range:

  - IQR = 0.725

- Comments: The descriptive statistics show an empirical mean close to the median, indicating a relatively symmetric distribution. The interquartile range (0.725) and standard deviation (0.4853) suggest moderate dispersion around the mean. The variance (0.2355) confirms a certain homogeneity in the data. A graphical analysis (histogram or box plot) could help detect outliers.

```
> summary(cats)
    Sex                  Bwt               Hwt
 Length:144        Min.   :2.000    Min.   : 6.30
 Class :character  1st Qu.:2.300    1st Qu.: 8.95
 Mode  :character  Median :2.700    Median :10.10
                   Mean   :2.724    Mean   :10.63
                   3rd Qu.:3.025    3rd Qu.:12.12
                   Max.   :3.900    Max.   :20.50
```

# Exercise 3: Box plots

The box plots confirm the results of the `summary()`, offering a better understanding of the distribution and possible outliers.
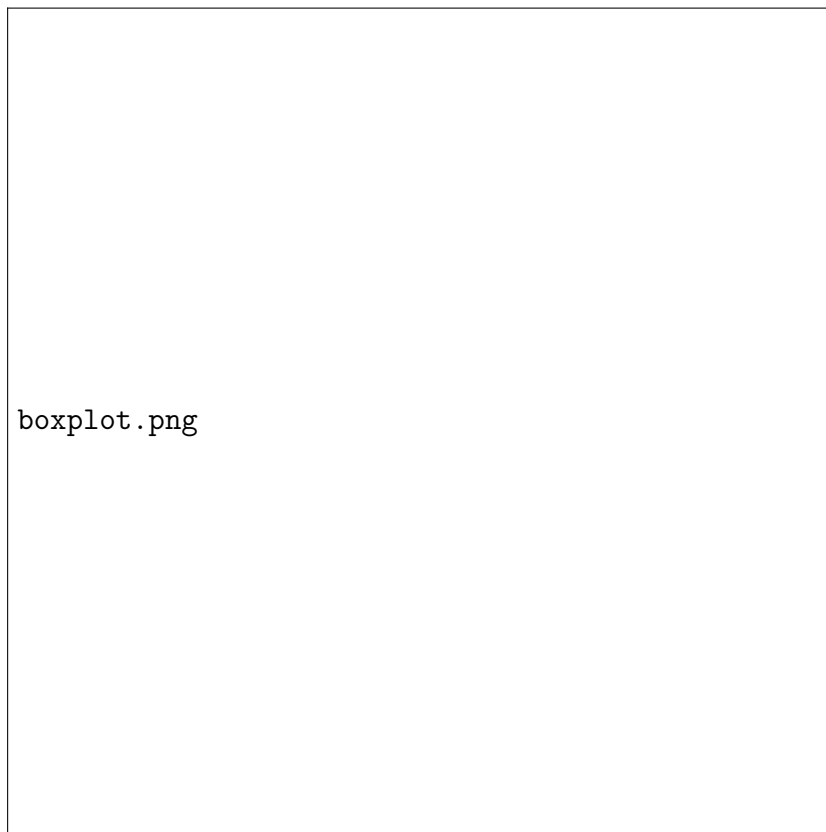
boxplot.png

Figure 1: Box plot of cat data variables

boxplotbysex.png
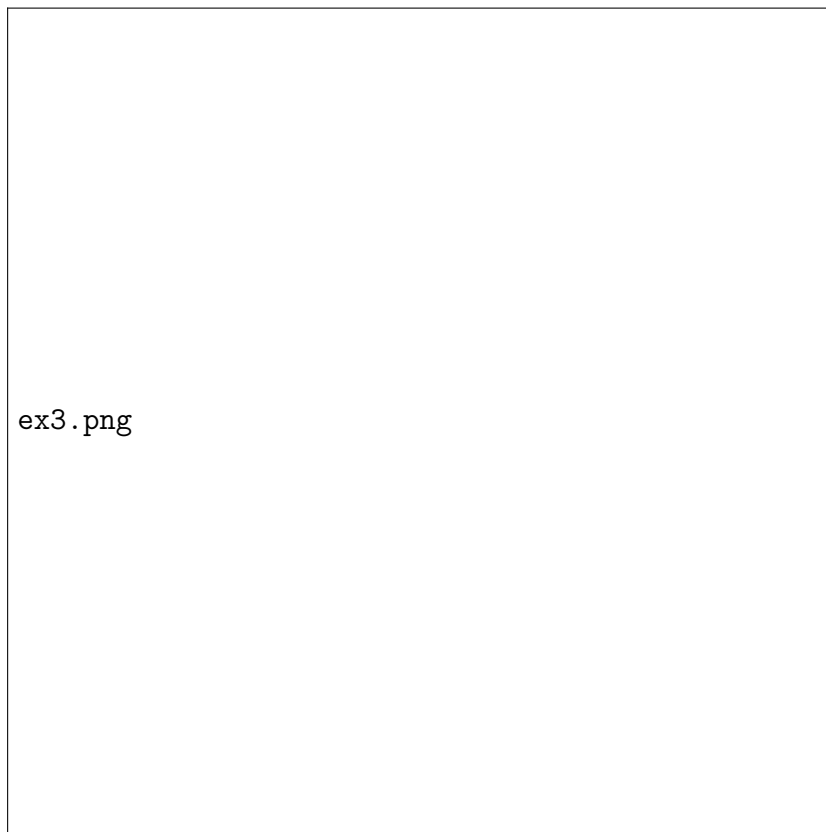
Figure 2: Box plot of cat data by sex
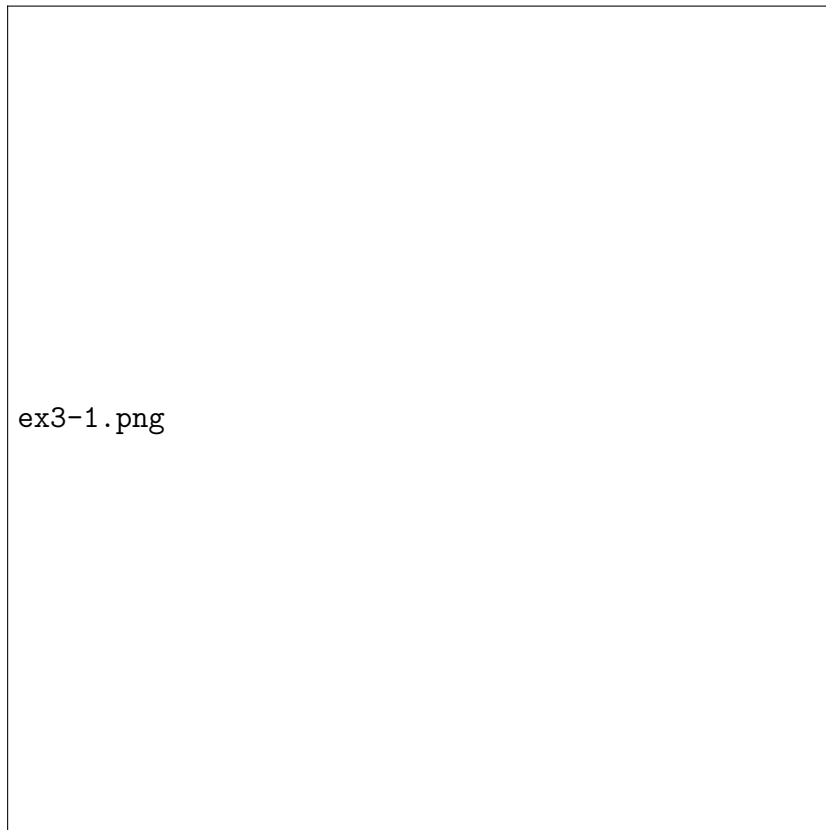
ex3.png

Figure 3: Additional box plot visualization

Figure 4: Additional box plot visualization - variation 1

# Exercise 4: Frequency histogram

The histogram of cat weights (Bwt) shows different behaviors depending on the number of bins used. With only 2 bins, the histogram is too coarse, making it difficult to observe the distribution of the data. With 20 bins, the distribution becomes clearer, revealing the key trends in the data. At 200 bins, the histogram becomes excessively detailed, and at 2000 bins, it resembles a scatter plot, thus losing its interpretability. A good compromise is to use about 10 to 30 bins, with 20 bins being a reasonable choice to effectively visualize the weight distribution.
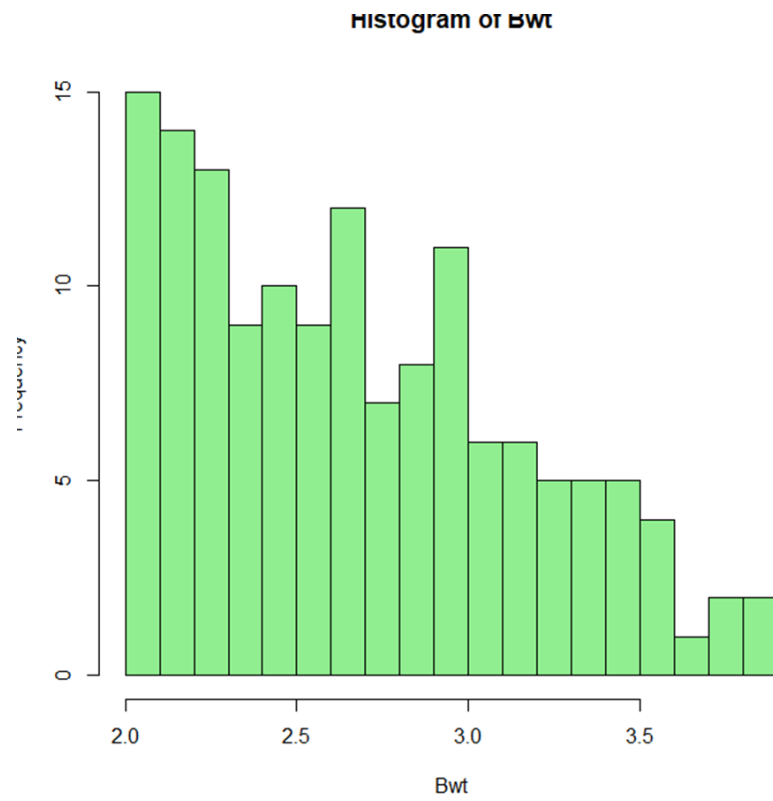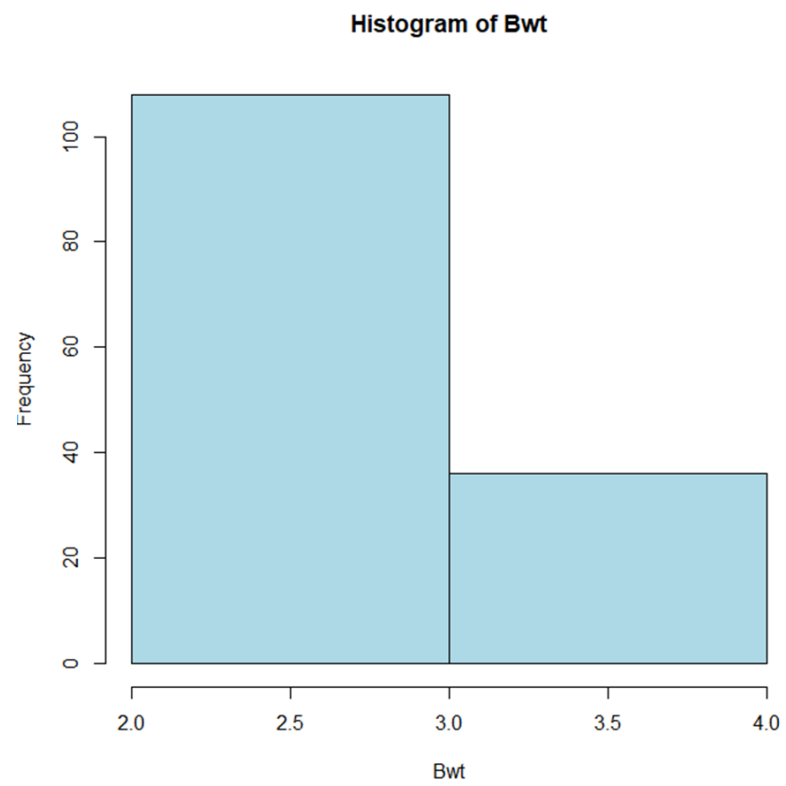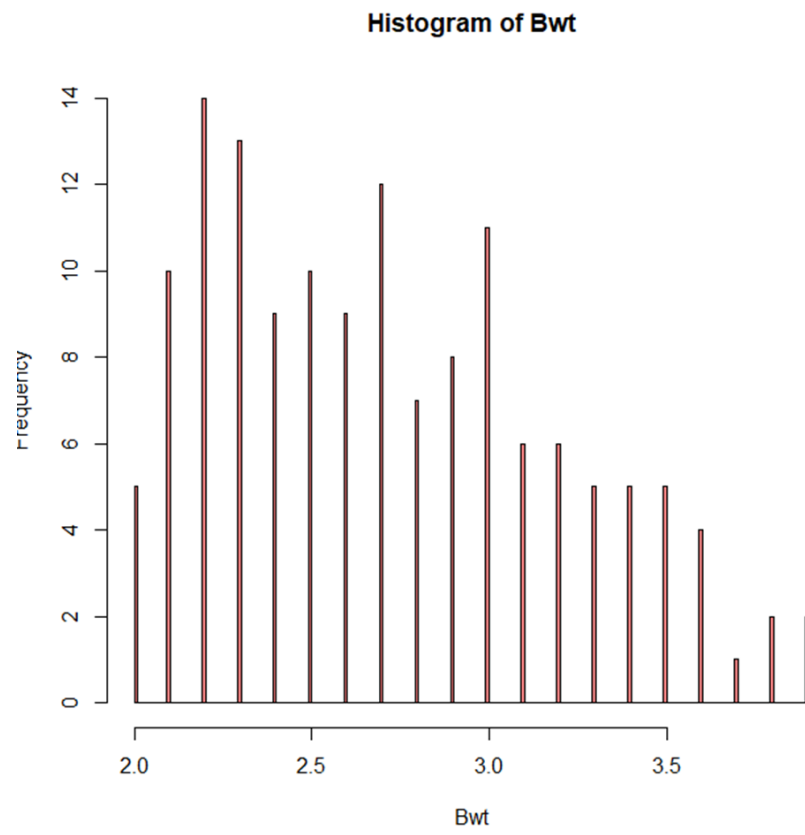
Figure 5: Histogram with 20 bins

**Histogram of Bwt**



Figure 6: Histogram with 100 bins

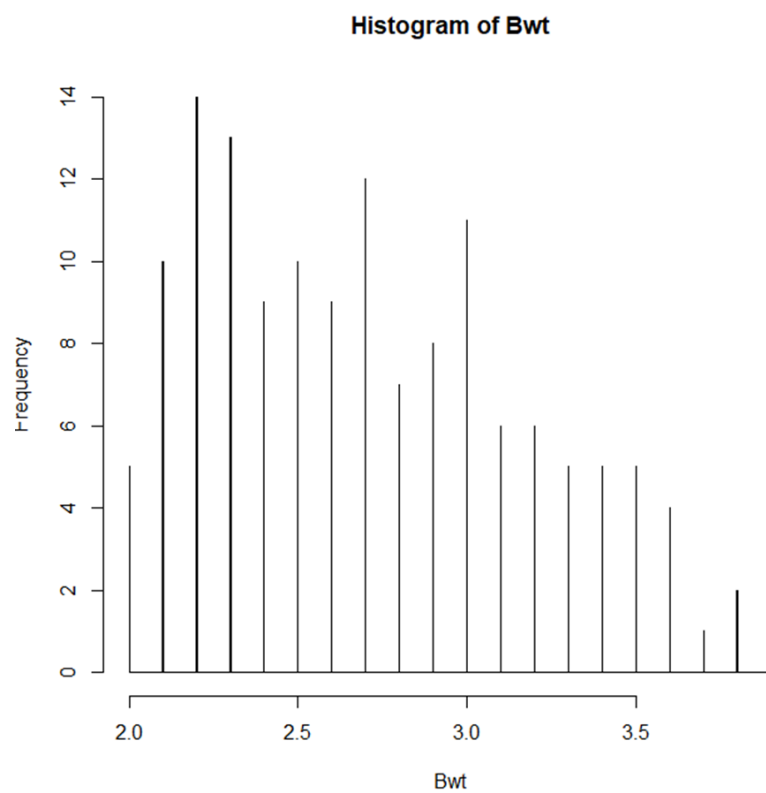Figure 7: Histogram with 1000 bins

9

## Histogram of Bwt



Figure 8: Histogram with 2000 bins

# Exercise 5: Qualitative variables

- **Frequency table for sex:**

```
> table(Sex)
Sex
F  M
47 97
```

- **Frequency table for heart weight (Hwt):**

```
> table(Hwt)
Hwt
6.3  6.5    7  7.1  7.2  7.3  7.4  7.6  7.7  7.9    8  8.1  8.2  8.3  8
      1    2    1    1    2    3    1    2    1    5    1    1    1    2
```
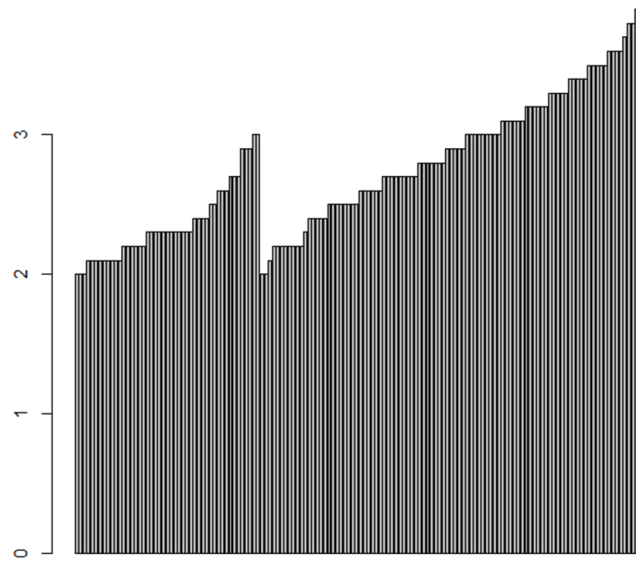
- **Barplot for Bwt:**



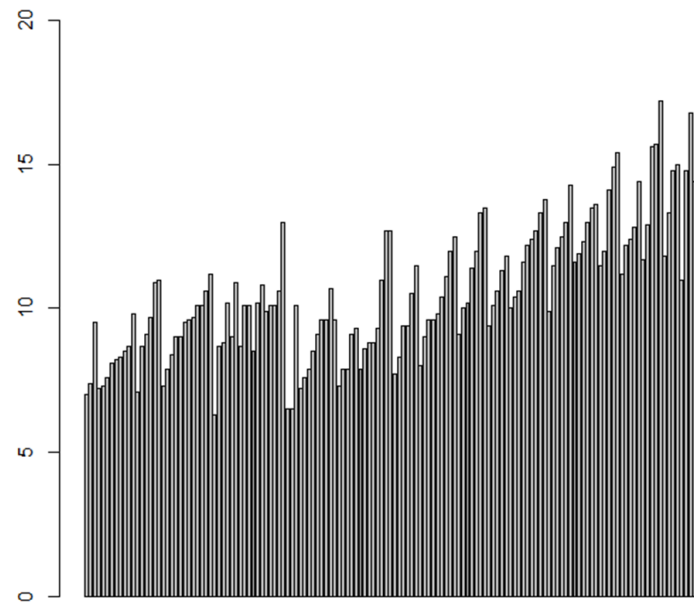Figure 9: Barplot for body weight (Bwt)

- **Barplot for Hwt:**

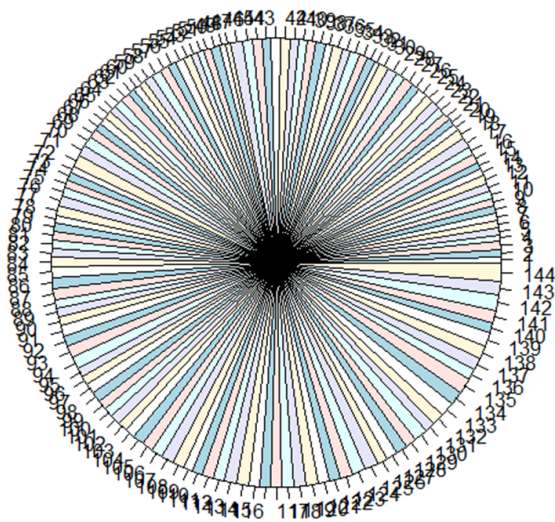Figure 10: Barplot for heart weight (Hwt)

- **Pie chart for Hwt:**

Figure 11: Pie chart for heart weight (Hwt)

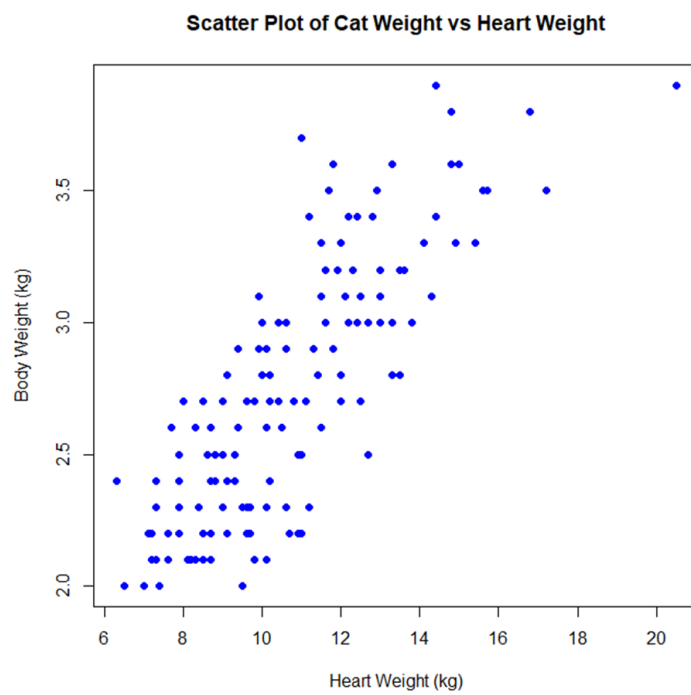# Exercise 6: Relationship between body weight and heart weight
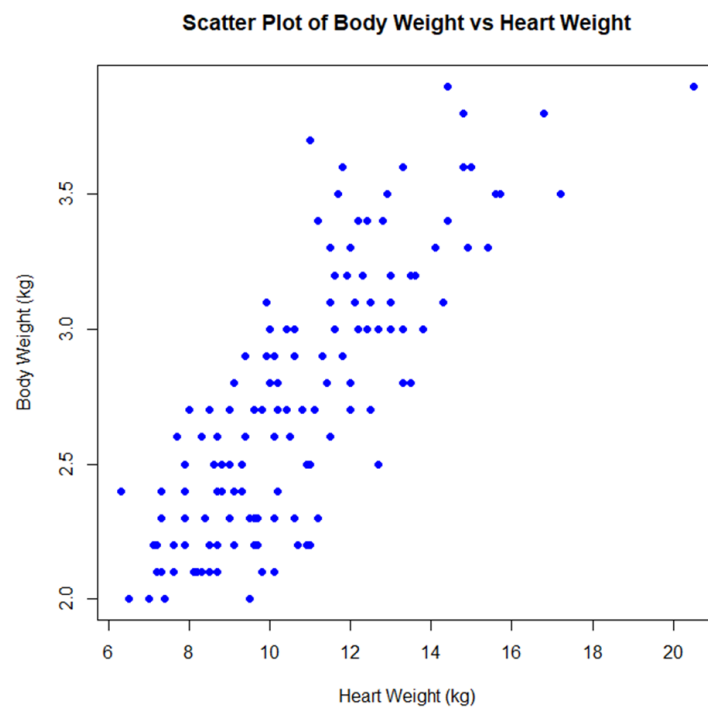
**Scatter Plot of Cat Weight vs Heart Weight**



Figure 12: Scatterplot of body weight (Bwt) vs heart weight (Hwt)

Figure 13: Alternative scatterplot visualization

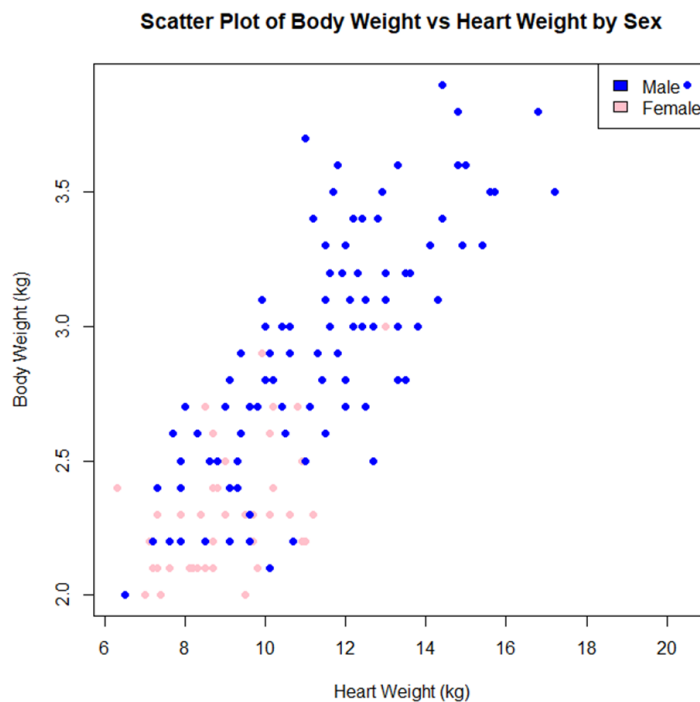**Scatter Plot of Body Weight vs Heart Weight by Sex**



Figure 14: Additional scatterplot visualization

# Exercise 7: Covariance and Correlation

- Covariance between Bwt and Hwt: 0.2950

- Correlation between Bwt and Hwt: 0.5457

```
> cor(cats$Bwt, cats$Hwt)
[1] 0.5456888
```

The positive correlation coefficient of approximately 0.55 indicates a moderate positive linear relationship between a cat's body weight and heart weight. As body weight increases, heart weight tends to increase as well, though the relationship is not perfectly linear.