Day 2 – Exercises Name:

1. Import the data set “cuckoos.csv”

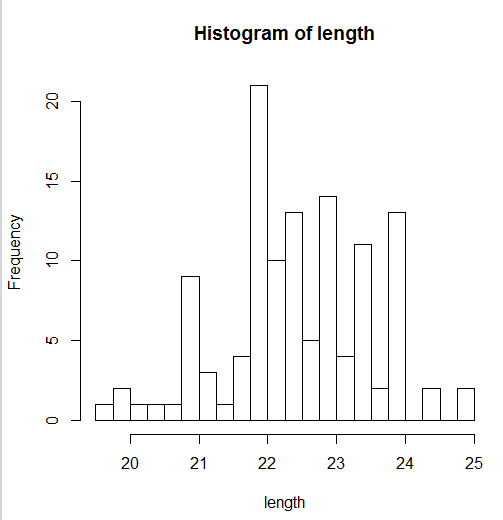
You can use the command:

read.csv("https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/DAAG/cuckoos.csv")

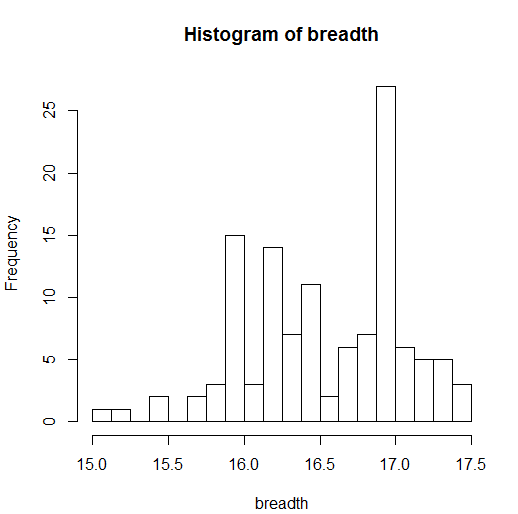
This data consists of measurements for 120 eggs lain in the nests of 6 species of birds who act as hosts for cuckoos. It contains a vector with egg length (mm, numeric), breadth (mm, numeric), species (factor), and sample id (numeric).

Import the “moments” package.

1. Plot a histogram of egg length. Make the bin breaks every 0.25 mm from 19.5 mm to 25 mm.



1. Plot a histogram of egg breadth. Make the bin breaks every 0.125 mm from 15 to 17.5 mm.



1. Find the skewness value for egg length and breadth. What do the results mean?

skewness(length)= -0.1934344

skewness(breadth)= -0.3598363

the distribution of length and breath is negatively(left) skewed with breadth skewed more negatively(left)

1. Find the kurtosis value for egg length and breadth. What to the results mean?

kurtosis(length)= 2.95649;

kurtosis(breadth)=2.557463

the distribution curve of length is closer to 3(closer to normal distribution)

1. Based on the previous results, which distribution (length or breadth) appears more normal?

The distribution of length

1. Run a Shapiro-Wilk on egg length and breadth. Do the results support your answer for 1 e.?

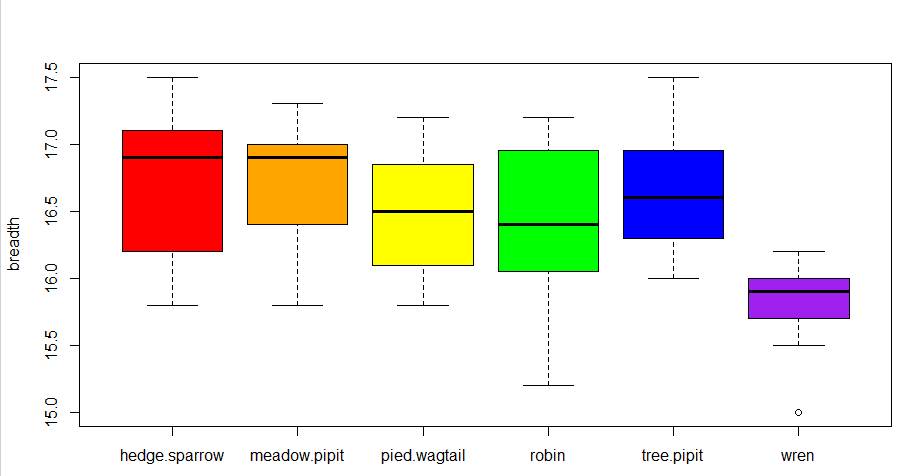
Yes

Length: p-value = 0.1736 >0.05; can’t reject that it is normal distribution

Breadth: p-value = 0.001841<0.05 ; reject Ho, it is not normal distribution at the level α=0.05

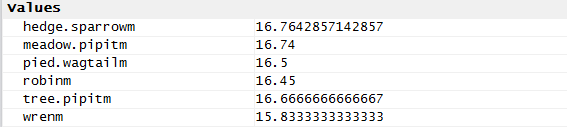
2. Use the same data set that you used for question 1.

1. Create a box plot of the egg breadth of each species. Which species appears to have the most narrow eggs?



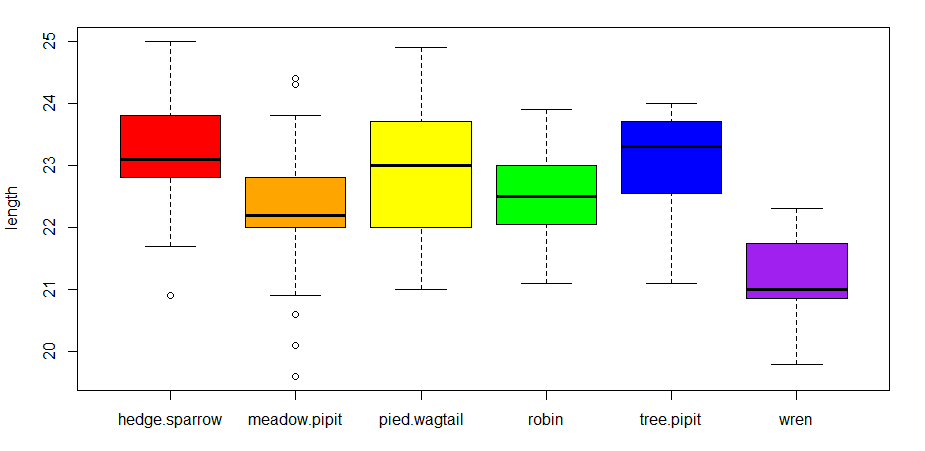
Wren

1. Calculate the mean egg breadth for each species. Do these results support your answer for 1 a.?



Yes

1. Create a box plot of the egg length for each species. Which species has the greatest number of outliers?

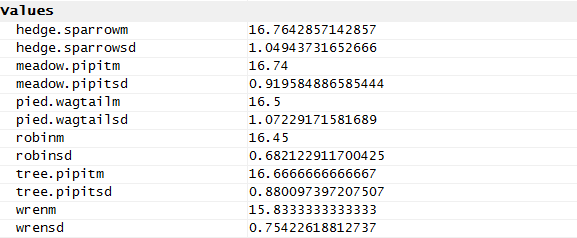


Meadow.pipit

1. Which species would you predict to have the greatest standard deviation for egg length? Why? Which species has the greatest standard deviation for egg length?

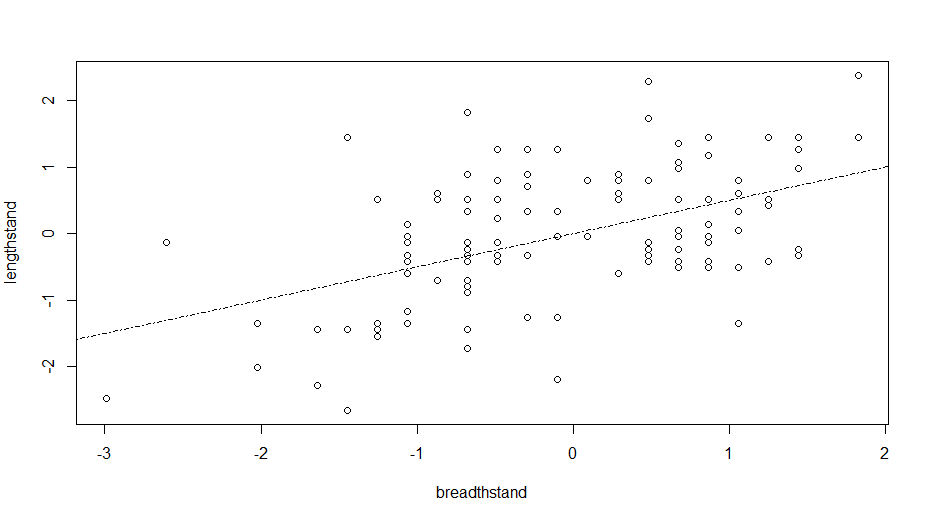
Meadow.pipit

Because it has the most outliers



Pied.wagtail

1. Create a scatter plot of *standardized* egg breadth and width. Based on the output, explain the relationship between the two variables.



The length increases as breath increases, while length doesn’t increase as much as breadth

3-5. Using the data provided below, reproduce following figures.

Data:

Import data from below URL using read.csv.

read.csv("https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/master/csv/boot/cane.csv")

The data frame represents a randomized block design with 45 varieties of sugar-cane and 4 blocks. The aim of the experiment was to classify the varieties into resistant, intermediate and susceptible to a disease called coal of sugar-cane (carvao da cana-de-acucar). This is a disease that is common in sugar-cane plantations in certain areas of Brazil.

For each plot, fifty pieces of sugar-cane stem were put in a solution containing the disease agent and then some were planted in the plot. After a fixed period of time, the total number of shoots and the number of diseased shoots were recorded.

Variables

***n***: The total number of shoots in each plot

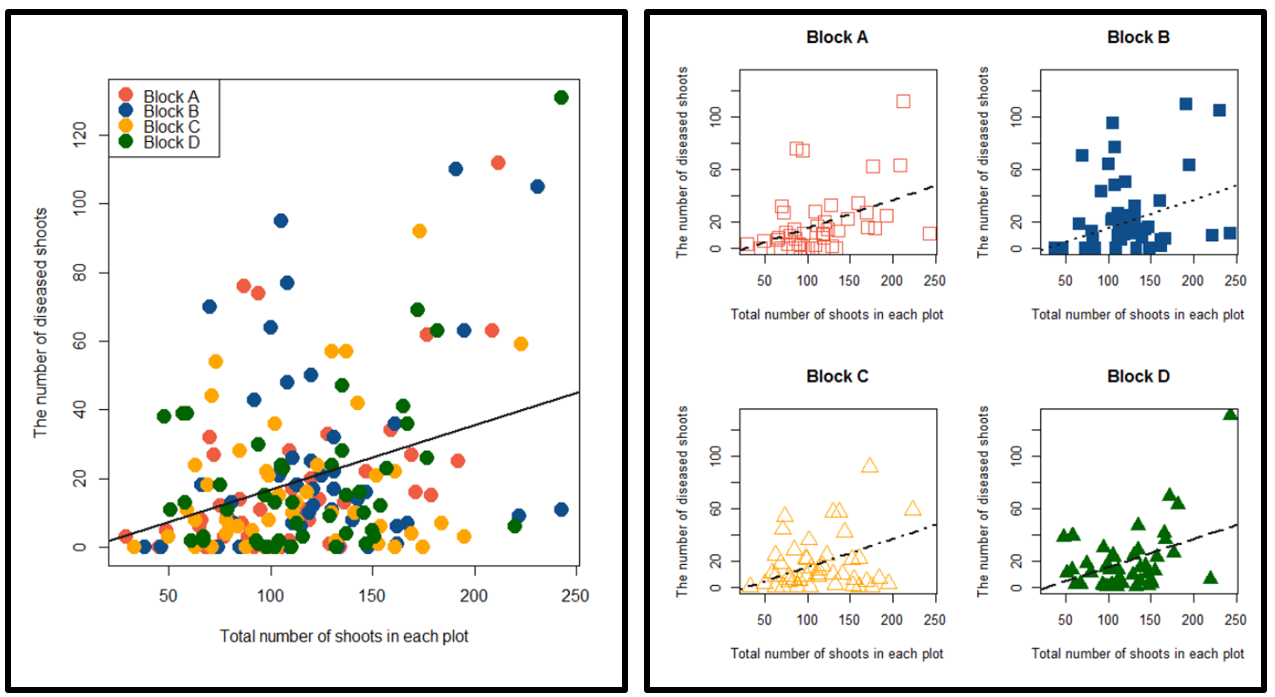
***r***: The number of diseased shoots

***x***: The number of pieces of the stems, out of 50, planted in each plot

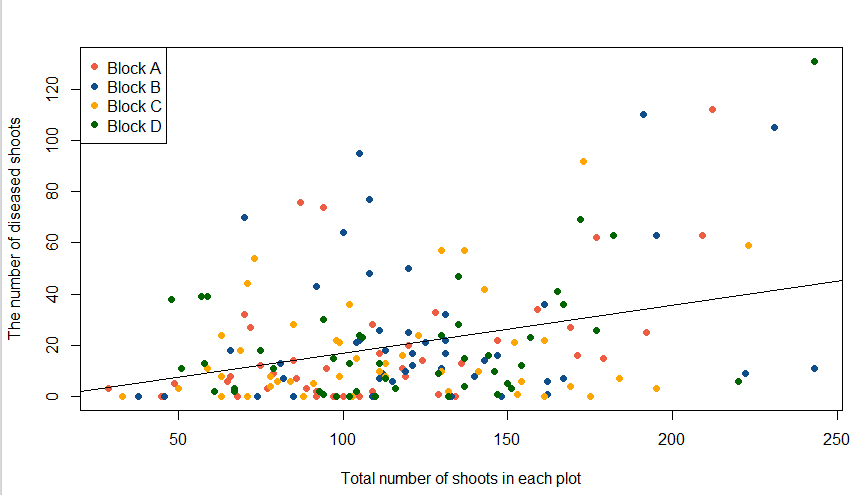
***var***: A factor indicating the variety of sugar-cane in each plot

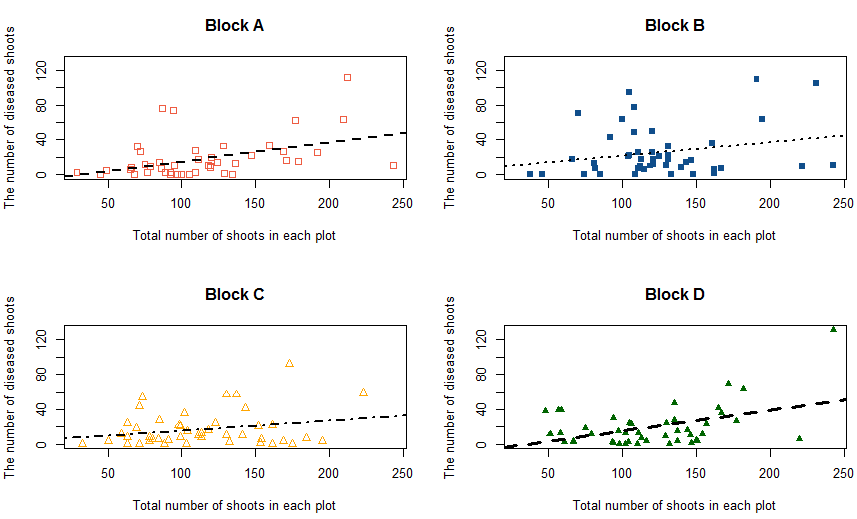
***block***: A factor for the blocks

3. Scatter plot

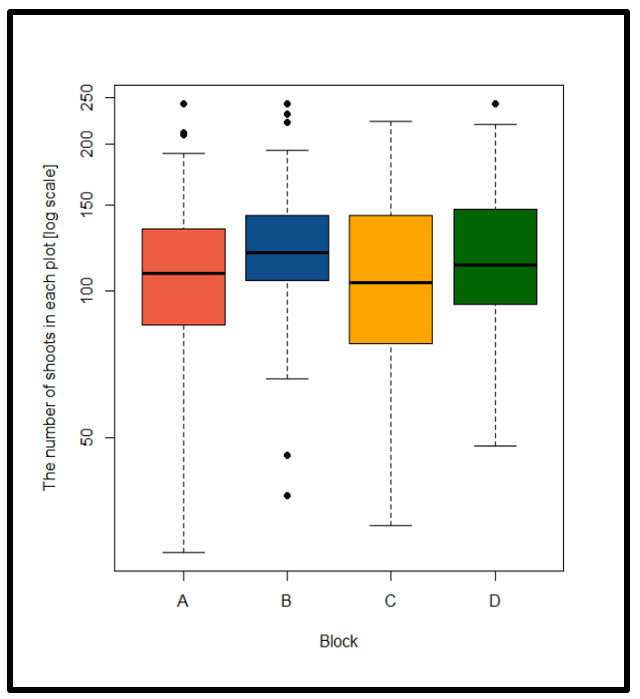


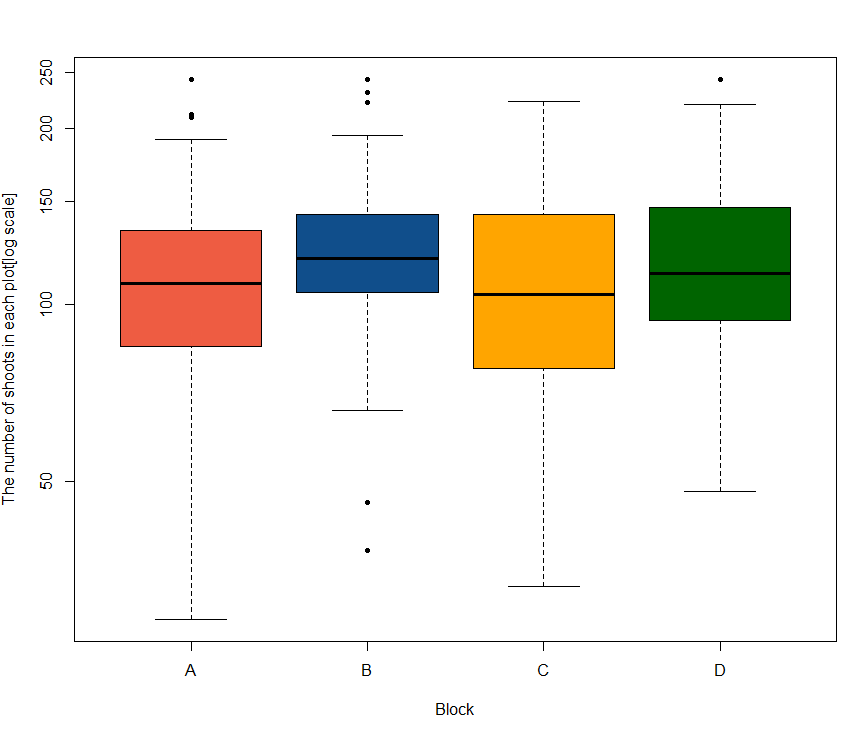
Color code used: “tomato2”, “dodgerblue4”, “orange”, “dark green”. You can use any four distinct colors you prefer to use.



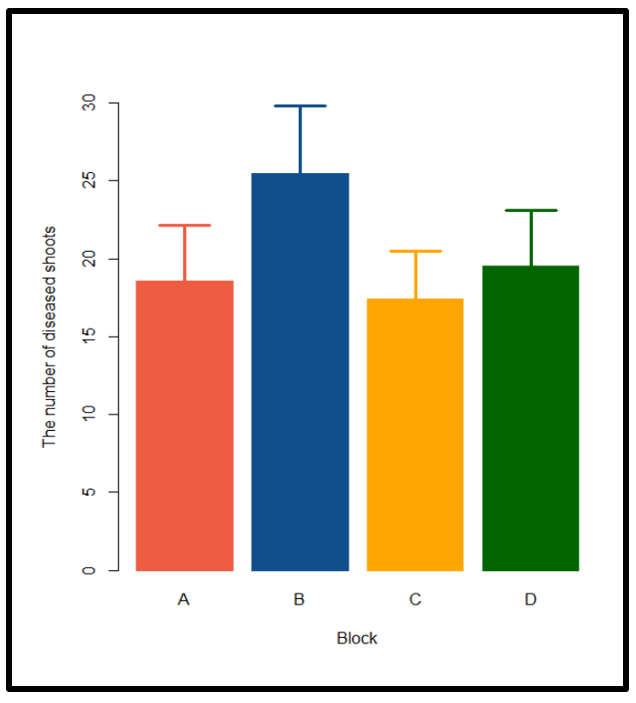


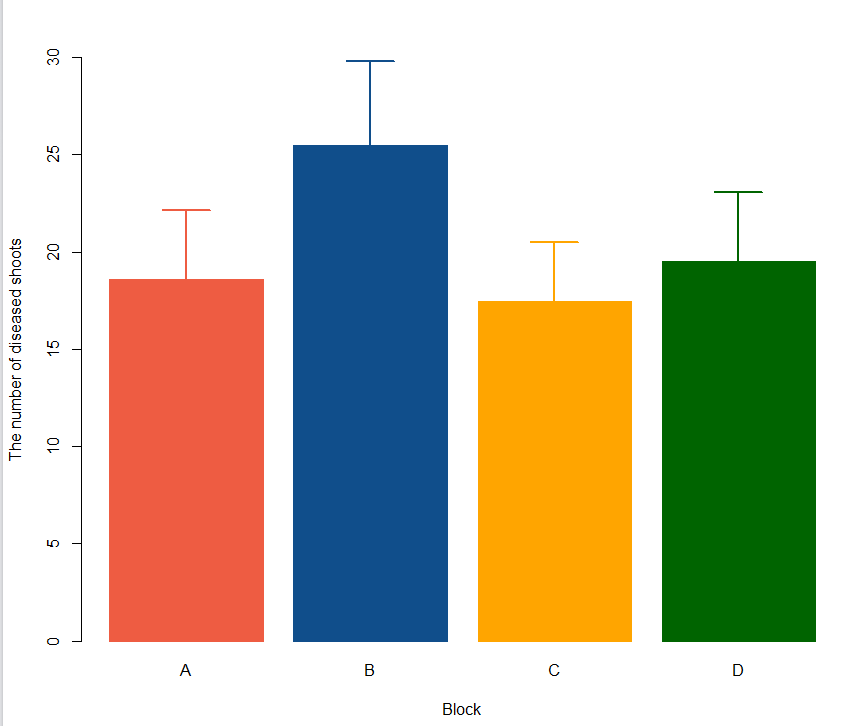
4. Boxplot





5. Bar plot





Use means for each block for making this barplot. Error bars are standard error.

**Name your files: LastName\_Exercise# (e.g. Knott\_Exercise1.docx)**

Answer each question and paste screenshots of any relevant output into this document. Submit your annotated code and answers to Blackboard by the start of class on Thursday.