Lab 1, Introduction to R
Marina Galvao – ▓▓▓▓▓
January 21st, 2022

Q1:
The Titanic dataset is about the ship Titanic. It describes information about the trip, including the
class each passenger was in, their respective names, ages, location where they embarked from,
and the home destination which they were to return to. The Titanic dataset also includes
passengers' sex, and whether they survived the sinking of the ship or not.

Q2:
The Titanic contained three passenger classes (1st, 2nd, and 3rd)

Q3:
When cross referencing survival with gender, the data is evidence that individuals on the Titanic
likely followed the policy of "women first". This is because 18% of the passengers that died
were women, whereas 82% were men. Furthermore, 68% of the passengers that survived were
female, and 32% male.

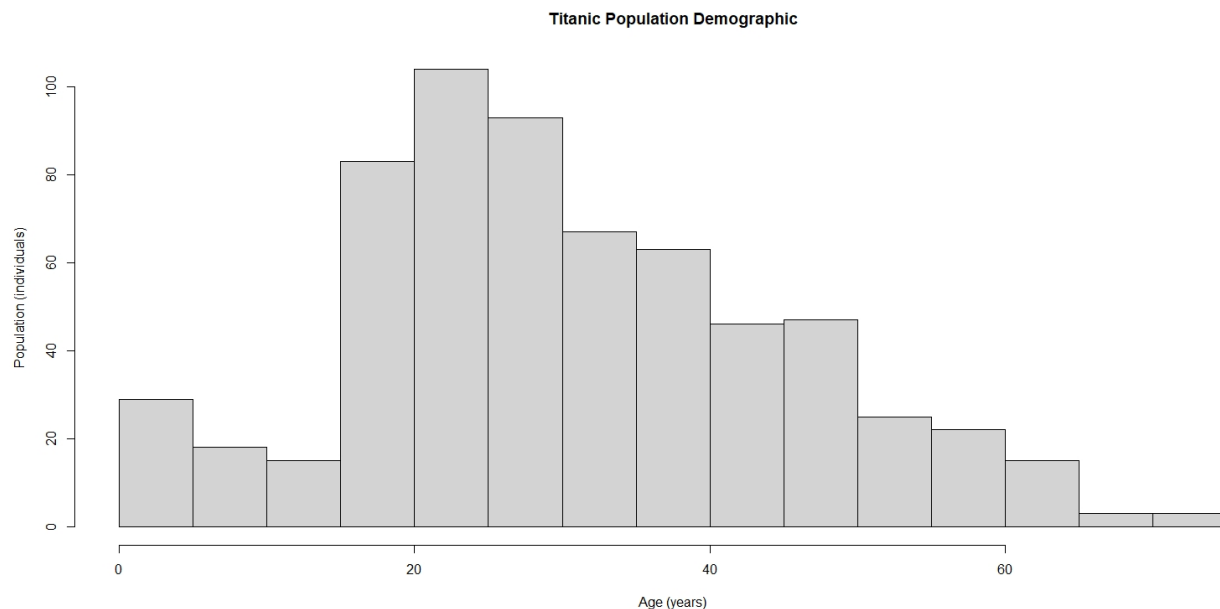|	No	Yes
Female	0.18	0.68
Male	0.82	0.32

Q4:

*Figure 1 (Plot 1), A frequency plot which describes the population demographic of the Titanic. Age is shown in the x-axis (in
years), and population is shown in the y-axis (as number of individuals). The data begins to skew to the right after the age of 20.
The trend shows that a large portion of individuals in the Titanic were between 15 and 30 years of age. The number of people
below 15 years old and beyond 55 years old is minimal in comparison to the rest.*
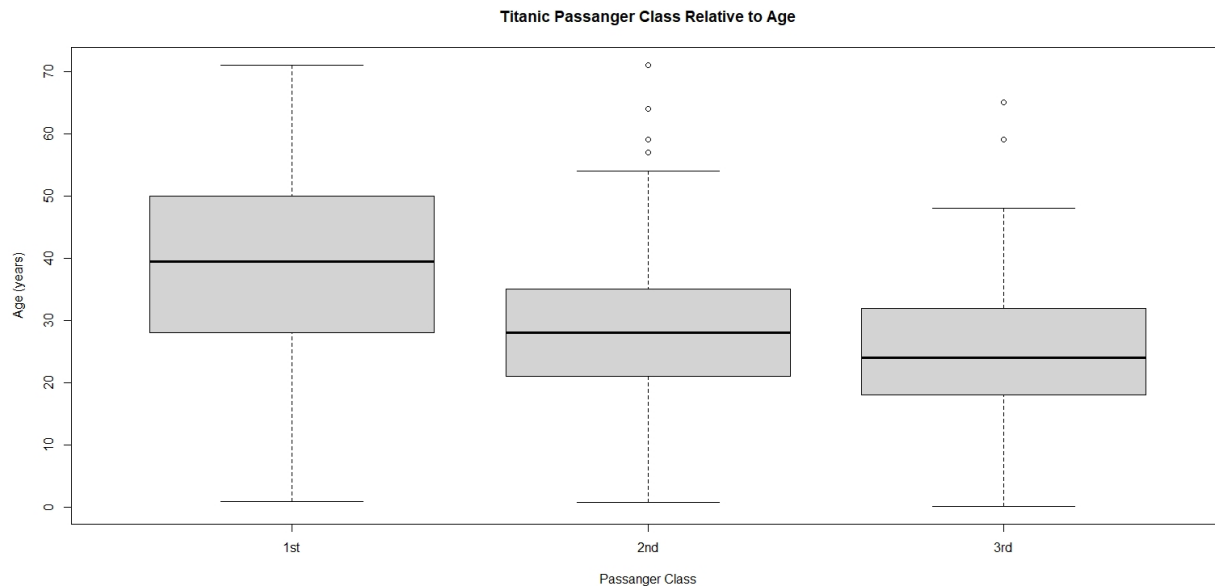
Titanic Passanger Class Relative to Age

Figure 2 (Plot 2), A boxplot of Titanic passenger class relative to age. Passenger class is shown in the x-axis, categorizing the passengers into first, second, or third class. Age (in years) is shown on the y-axis. The data shows the maximum age for the 1st class passengers is 70 years, approximately 55 years for the 2nd, and just under 50 years old for the 3rd class. The minimum for all three classes is 0. The median for the 1st class lies at approximately 40 years old, just below 30 years old for the 2nd class, and around 35 years old for the 3rd. The general trend shows that moving from 1st class to 2nd and 3rd, the location of the interquartile range of age decreases as well. Further, the 2nd and 3rd classes possess outliers.
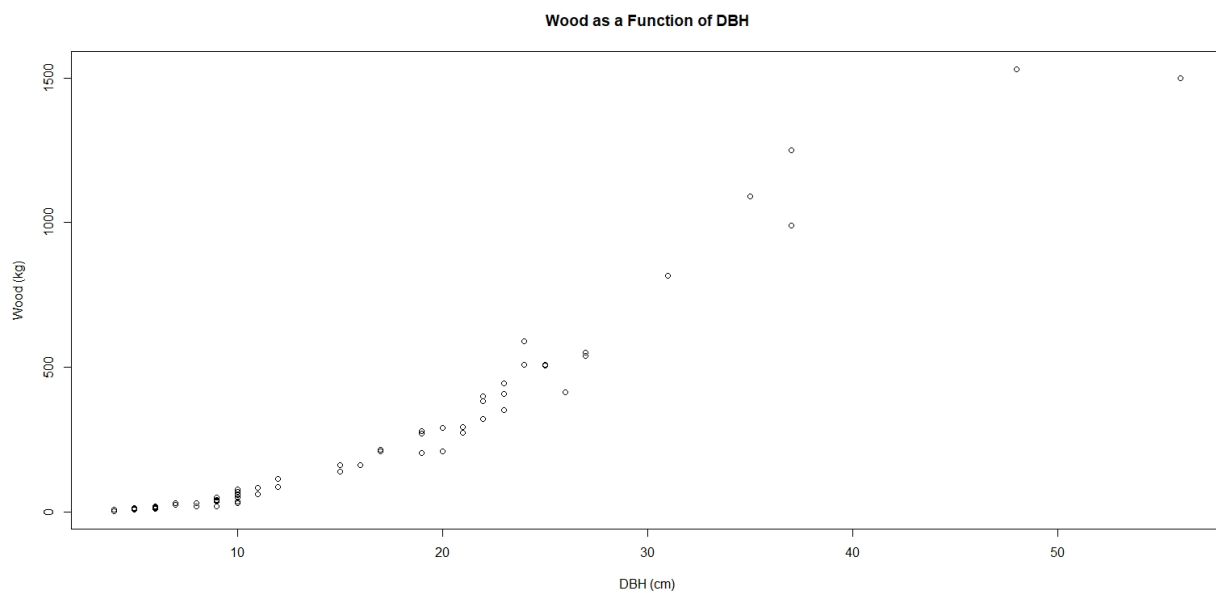
Wood as a Function of DBH

Figure 3 (Plot 4), A scatterplot which describes the positive correlation between wood weight and DBH. DBH is shown in cm in the x-axis, and wood in kg in the y-axis. The positive relationship appears strongest between 0 and 12 cm in DBH. There are two data points around 1500 kg of wood which do not follow the trend line as strongly as the remaining data points.
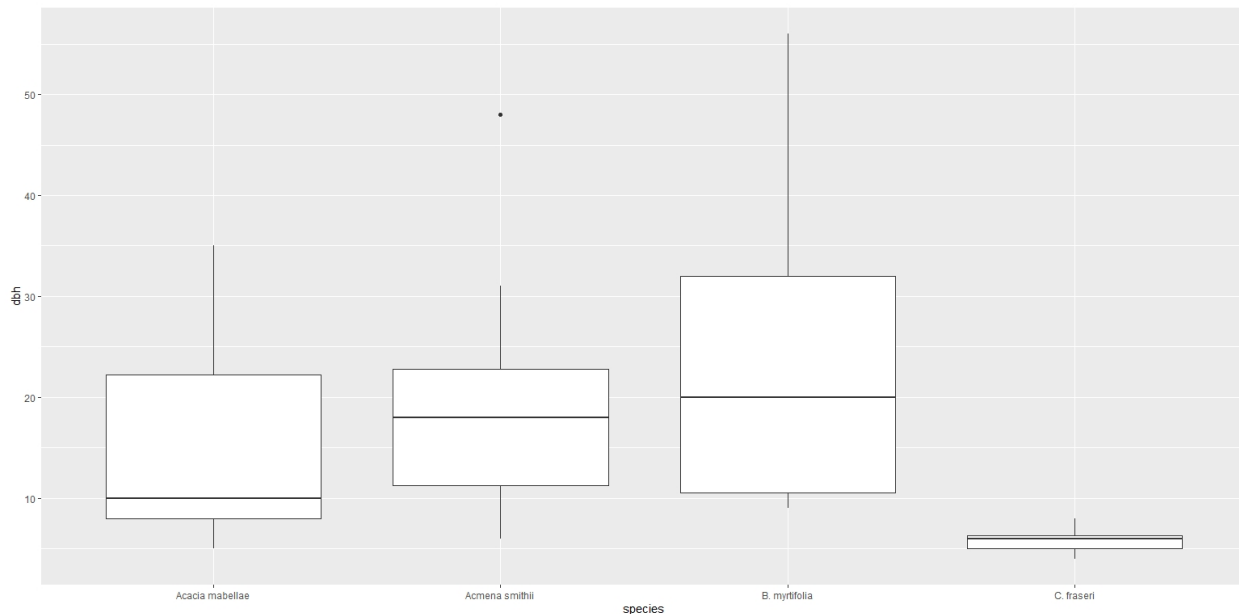
Q5:

Figure 4 (Plot 5), A boxplot containing DBH as a function of species. Species is located on the x-axis and includes A. macabellae, A. smithii, B. myrtifolia, and C. fraseri. DBH (in cm) is provided in the y-axis. As the boxplot shows, B. myrtifolia possesses the greatest interquartile range, followed by A. mabellae, A. smithii and C. fraseri. B. myrtifolia also possesses the highest maximum value, as high as 55cm in DBH. A. mabellae has the second highest maximum value, followed by A. smithii and C. fraseri (lower than 10 cm in DBH). The greatest median can be observed in B. myrtifolia, followed by A. smithii, A. macabellae, and C. fraseri. The only outlier present is observed around 48cm in DBH, under the species A. smithii.
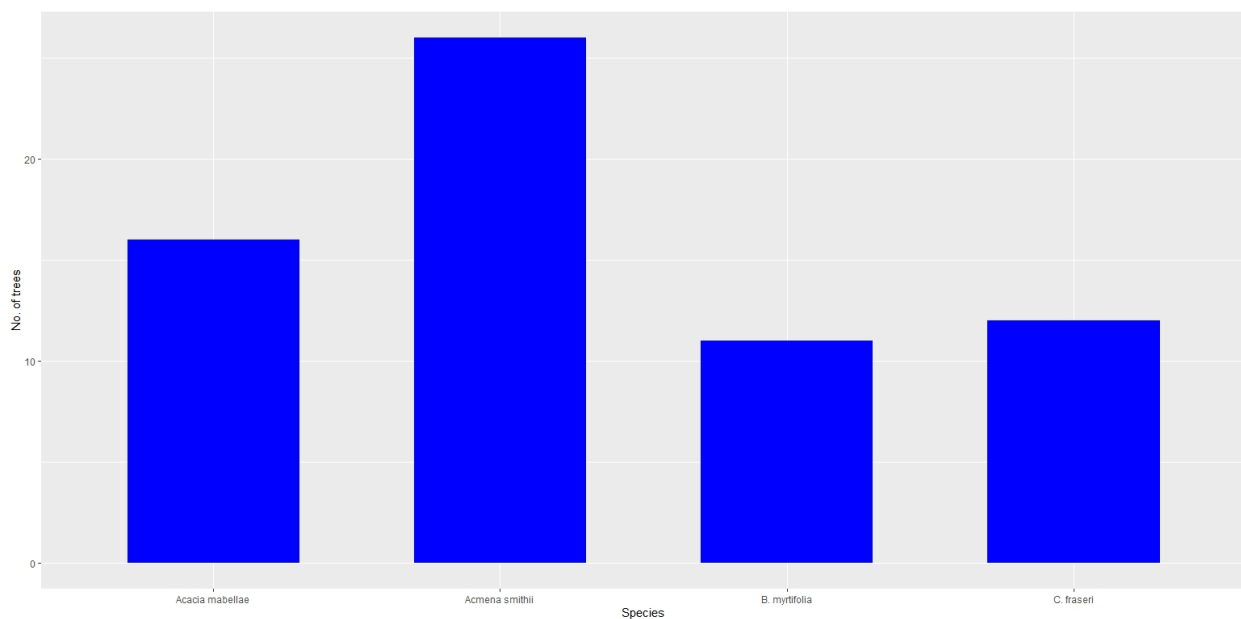

Figure 5 (Plot 7), A bar chart of tree species in relation to the number of trees. Tree species included in the survey are A. mabellae, A. smithii, B. myrtifolia, and C. fraseri. Number of trees is present as the added-up number of individuals trees present within that species. The data shows the predominant species is A. smithii (with just over 25 trees), followed by A. mabellae (with just over 15 trees), C. fraseri (with approximately 12 trees), and B myrtifolia (with approximately 11 trees).
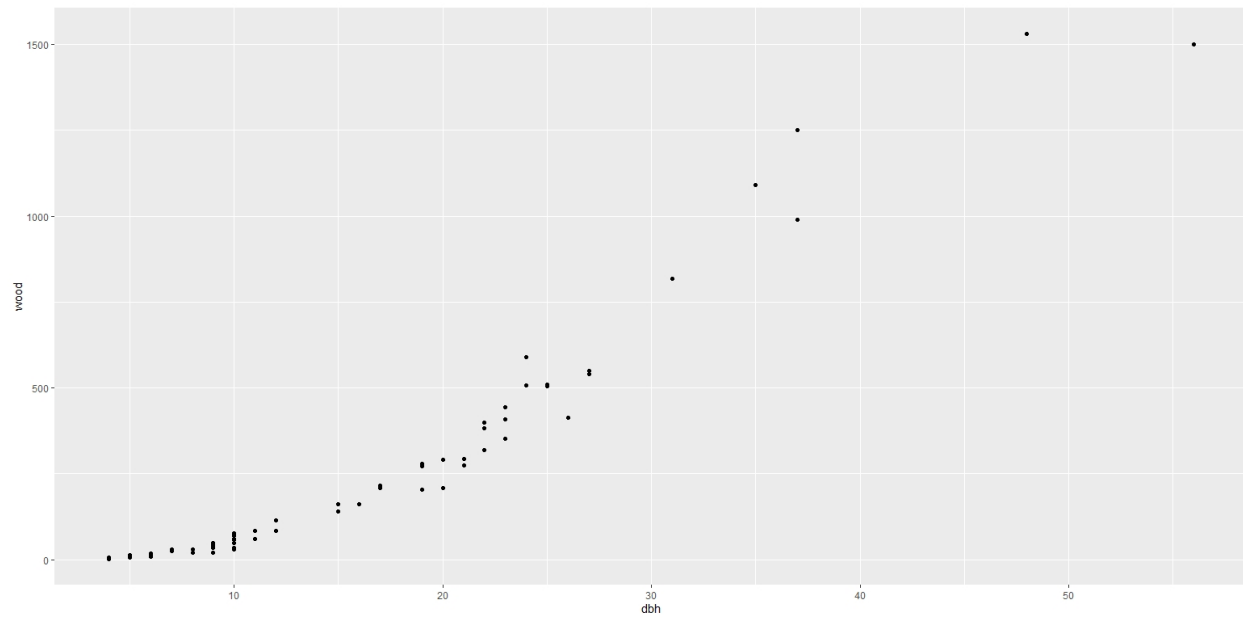
Figure 6 (Plot 8), A scatterplot which shows wood as a function of DBH. Wood is shown in kilograms, DBH is shown in cm. The trend shows a positive relationship between wood and DBH, with two data points located approximate to 1500 kg falling just outside of the general trend line.