

NRSG 741 - Homework 2

Melinda Higgins

02/06/2019

Due Date is 13 February 2019

This homework is meant to further your `dplyr` and `ggplot2` skills.

Abalones Dataset from UCI Repository

For this homework, you will keep working with the `abalone` dataset from the UCI data repository at <https://archive.ics.uci.edu/ml/datasets/abalone>.

Use tools within the `dplyr` package as much as possible to answer the following questions.

Question 1: What kind of R object is the `abalone` dataset?

Question 2: How many observations are in the `abalone` dataset?

Question 3: For shucked weight, how many abalones weigh more than 0.8 grams?

Question 4: How many abalones have shucked weights larger than their whole weight? (*HINT: create a new variable using `mutate` and then filter*)

Create a subset containing only infants `sex == "I"`

Question 5: How many infants are in this subset?

Show off your `dplyr` skills with `group_by()`

Question 6: What is the average whole weight for each abalone sex (get whole weight means for females “F”, males “M” and infants “I” separately)?

Question 7: Get the means for the abalone length and height by sex?

Test your graphing skills using `ggplot2`

Using the `abalone` dataset, create the following graphics/figures using `ggplot()` and associated `geom_xxx()` functions.

Question 8: Create a histogram of abalone whole weight

What do you notice about the distribution (any outliers or skewness)?

Question 9: Create side-by-side boxplots of the number of rings by gender

HINT use `geom_boxplot` with $x = \text{sex}$ and $y = \text{rings}$

Question 10: Create a scatterplot of the whole weight on the X axis and shucked weight on the Y axis and color the points by sex