ZOO 800

Homework Week 4

Submission instructions

Submit a single URL to a public GitHub repository on Canvas. Solutions to Objectives 1-3 should be included in a single R script.

Problem

Antarctic researchers think that smaller penguins are more susceptible to climate change, but they need some help summarizing their data.

The penguin data set they have sent you can be found in the package: "palmerpenguins" and the data frame object is called “penguins”. Install the package and read in the dataset.

Objective 1

a) Create a function to convert a continuous variable into a binary variable (e.g., high/low, yes/no, 1/0, etc.). The function should allow the user to specify the breakpoint by which the data are divided in two, as well as specify the labels that both groups are assigned
b) Use your function to convert body mass into a binary variable of ‘small’ and ‘large’ penguins

Objective 2

Now that they’ve thought more about it, the researchers think that there might be a goldilocks size for a penguin to be…

a) Generalize your function from Objective 1 to be able to accommodate any number of break points. This new function should allow the user to specify how many categories the data is broken into, the corresponding breakpoints, and the labels for each category
b) Use your function to convert body mass into a categorical variable of ‘small’, ‘medium’, and ‘large’ penguins

Objective 3

The researchers forgot to mention there are multiple species included in this dataset! Obviously, what constitutes a 'small' or 'large' penguin depends on the species.

a) Use the quantile function (or something similar) to determine sensible breakpoints for each species
b) Modify the functions in Objective 2 to discretize body mass conditional on species (HINT: this will likely require a ***for*** loop or ***if*** statements inside the function)
c) Use your function to convert body mass into a categorical variable with three levels with different breakpoints for each species

Objective 4

The researchers want to visualize their summarized data.

a) Modify your function from Objective 3 so that the output is a boxplot, grouped by species and size category. Bonus points for making something really pretty

Objective 5 (bonus)

The researchers have collected new data from some other species of penguins that are completely different sizes to those presently included.

a) try including the quantile() function within your custom function so that it can determine sensible breakpoints rather than them being manually specified beforehand