# Lab 4 Analyze Avocado Data

## The data file to use:

avocado.csv

## Tasks

1. Read the data from the CSV file into a tibble. To do that, you can use the read\_csv() function like this (make sure your copy of avocados.csv is in your working directory within RStudio, see the lecture notes).

avocados <- read\_csv("avocado.csv")

1. Display the tibble on the console.
2. Summarize all columns with the summary() function.
3. Display the first four columns of data with the select() function.
4. Summarize the AveragePrice and TotalVolume columns.
5. Display the first three columns of the last five rows of the data set.
6. Add a new column named EstimatedRevenue that stores the result of multiplying the TotalVolume and AveragePrice columns. Then, display the data for these columns to confirm that the column was added and has the correct values.
7. Create a tibble named avocados\_by\_region that’s grouped by region and type and has a third column that includes the average price for each group. Then, display the tibble to view its data.
8. Save your R script file as ***lastname\_lab4*** and submit for grading.

## Questions

Answer these questions using comments within your R script file.

1. How many unique regions are there?
2. What is the average price for each type of avocado (organic and conventional)? Be sure to include just the type and AveragePrice columns in the results.
3. Which region has the lowest average price for *organic* avocados? (Create a tibble that has the correct data. Then, use RStudio to view the tibble and sort by the AveragePrice and type columns.)

## How The Lab Will Be Graded:

|  |  |  |
| --- | --- | --- |
| Lab Component | Point value | Total |
| Steps 1-8 completed and saved in an R script that is submitted on Ecampus. | 9 steps @ 6 points each | 54 |
| Three questions answered | 3 questions @ 2 points each | 6 |
| Total |  | 60 |