# Lab 7 : Prepare the Avocado Data

1. Open a blank R file and insert comments at the top with your name, and the weirdest pet name you ever have heard.
2. Read the data from the .csv file into a tibble and display the data.
3. Make the capitalization of the column names consistent by renaming the type, year, and region columns to Type, Year, and Region.
4. Add a column named SmallPercent that contains the percentage of small bags out of the number of total bags.
5. Use the select() function to display the Date, SmallBags, TotalBags, and SmallPercent columns.
6. Group and summarize the data to display this tibble:

# A tibble: 54 × 3

Region Count AveragePrice

<chr> <int> <dbl>

1 Albany 338 1.56

2 Atlanta 338 1.34

3 BaltimoreWashington 338 1.53

4 Boise 338 1.35

...

1. Group and summarize the data to display this tibble:

# A tibble: 432 × 6

# Groups: Region, Year [216]

Region Year Type Count TotalBags AveragePrice

<chr> <dbl> <chr> <int> <dbl> <dbl>

1 Albany 2015 conventional 52 662366. 1.17

2 Albany 2015 organic 52 57289. 1.91

3 Albany 2016 conventional 52 759091. 1.35

4 Albany 2016 organic 52 79209. 1.72

5 Albany 2017 conventional 53 699561. 1.53

6 Albany 2017 organic 53 135944. 1.75

7 Albany 2018 conventional 12 245241. 1.34

8 Albany 2018 organic 12 41553. 1.53

9 Atlanta 2015 conventional 52 2935926. 1.05

10 Atlanta 2015 organic 52 61065. 1.71

...

1. Add a column named PriceGroup that puts the rows into 10 bins with an approximately equal number of values.
2. To confirm that the previous step worked correctly, display the AveragePrice and PriceGroup columns. Then, display the number of values in each bin.