

## XP Booster

March 24

For this challenge you will be working in GROUPS OF TWO. You will be using the avocado dataset (avocado.csv). The variables you will be using are as follows:

- Total.Volume: The total volume of avocados sold at a location over several years.
  - AveragePrice: The average price of an avocado sold at each location.
  - type: whether avocados are conventional or organic type.
  - region: The region that a given store is located
1. Visualize the distribution for total volume of avocados sold. Be sure to colour your distribution 'green'. Draw and describe the distribution below. What is strange here?
  2. Visualize the distribution for total volume of *conventional* avocados sold. Draw and describe in detail below.
  3. Using a pairwise boxplot, compare the total volume of avocados sold in California vs. Boston.
  4. Using a pairwise boxplot, compare the average price of avocados sold in California vs. Boston.

### Solution:

```
1  ### Load Packages
2  library(ggplot2)
3  library(tidyverse)
4
5  ### Load data
6  avo <- read.csv('/data/datasets/avo.csv')
7
8  ##### Visualize distribution for avocado total volume in California (What is
9         strange here?)
10 dat <- avo %>% filter(region=='California')
11 hist(dat$Total.Volume, col='green', xlab='Total volume sold', main='Distribution
12      for total volume sold')
13
14 ##### Visualize distribution for avocado total volume in California (What is
15      strange here?)
16 dat <- avo %>% filter(region=='California') %>% filter(type=='conventional')
17 hist(dat$Total.Volume, col='green', xlab='Total volume sold', main='Distribution
18      for total volume sold for conventional avocados')
19
20 ##### Compare conventional avocados in California to Boston
21 cali <- avo %>% filter(region == 'California') %>% filter(type=='conventional')
22 bost <- avo %>% filter(region == 'Boston') %>% filter(type=='conventional')
23 boxplot(cali$Total.Volume, bost$Total.Volume)
24
25 ##### Compare conventional avocados in California to Boston
26 cali <- avo %>% filter(region == 'California') %>% filter(type=='conventional')
27 bost <- avo %>% filter(region == 'Boston') %>% filter(type=='conventional')
28 boxplot(cali$AveragePrice, bost$AveragePrice)
```