XP Booster
March 23

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 the total volume of avocado's sold in California retail locations. Draw and describe in full detail (remember unusual features).
2. Using ggplot add color fill to the distribution based on the type of avocado. What do you notice about the total volume of avocados sold for conventional vs. organic.
3. Using a pairwise boxplot compare the total volume of conventional avocados sold in California and Boston.
4. Create a scatter plot showing average price against total volume for organic Californian avocados. Does there appear to be an association?

Solution:

```
  ### Load Packages
library(ggplot2)
library(tidyverse)

### Load data
avo <- read.csv('/data/datasets/avo.csv')

#### Visualize distribution for avocado total volume in California (What is
    strange here?)
avo %>% filter(region=='California') %>% ggplot(aes(x=Total.Volume)) + geom_
    histogram(fill='green', col='darkgreen') +
  labs(x='Total volume sold at location', title='Distribution for volume of
    avocados sold')

#### Coloring by type of avocado (you can see that organic are sold in much
    less volume?)
avo %>% filter(region=='California') %>% ggplot(aes(x=Total.Volume, fill=type
    )) + geom_histogram() +
  labs(x='Total volume sold at location', title='Distribution for volume of
    avocados sold')

#### Compare conventional avocados in California to Boston
avo %>% filter(region == 'California' | region == 'Boston') %>% filter(type==
    'conventional') %>%
  ggplot(aes(y=Total.Volume, fill=region)) + geom_boxplot()

### Create a scatter plot of average price against total volume
```

```
avo %>% filter(type == 'organic' & region == 'California') %>% ggplot(aes(x=
    Total.Volume, y=AveragePrice))+
  geom_point()
```