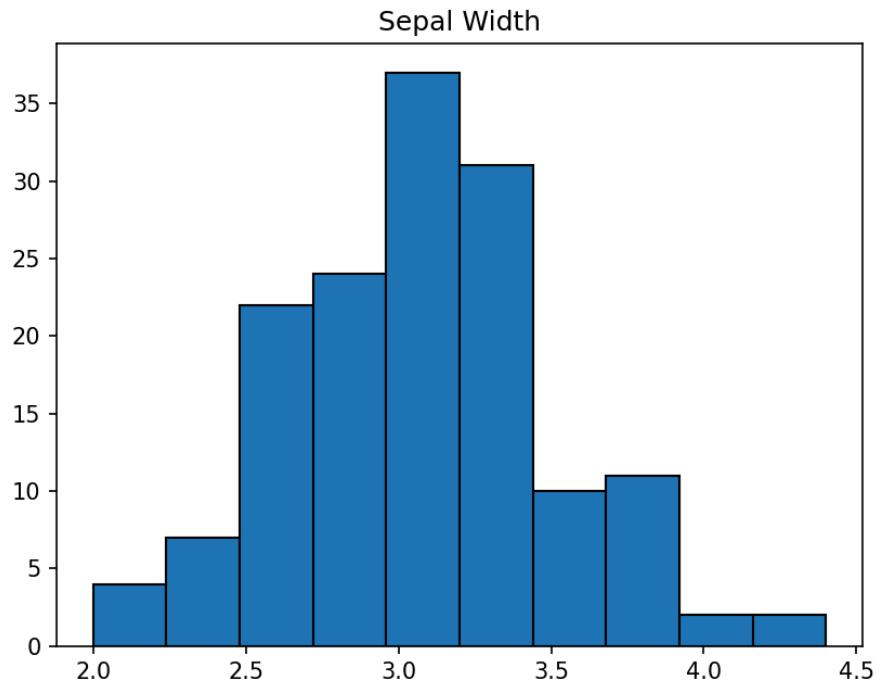


## Question 1

- a. Make a histogram of the variable Sepal.Width.



- b.  
c. Based on the histogram from #1a, which would you expect to be higher, the mean or the median? Why?

Since the histogram appears symmetrical, we can expect the mean and median to be close in value.

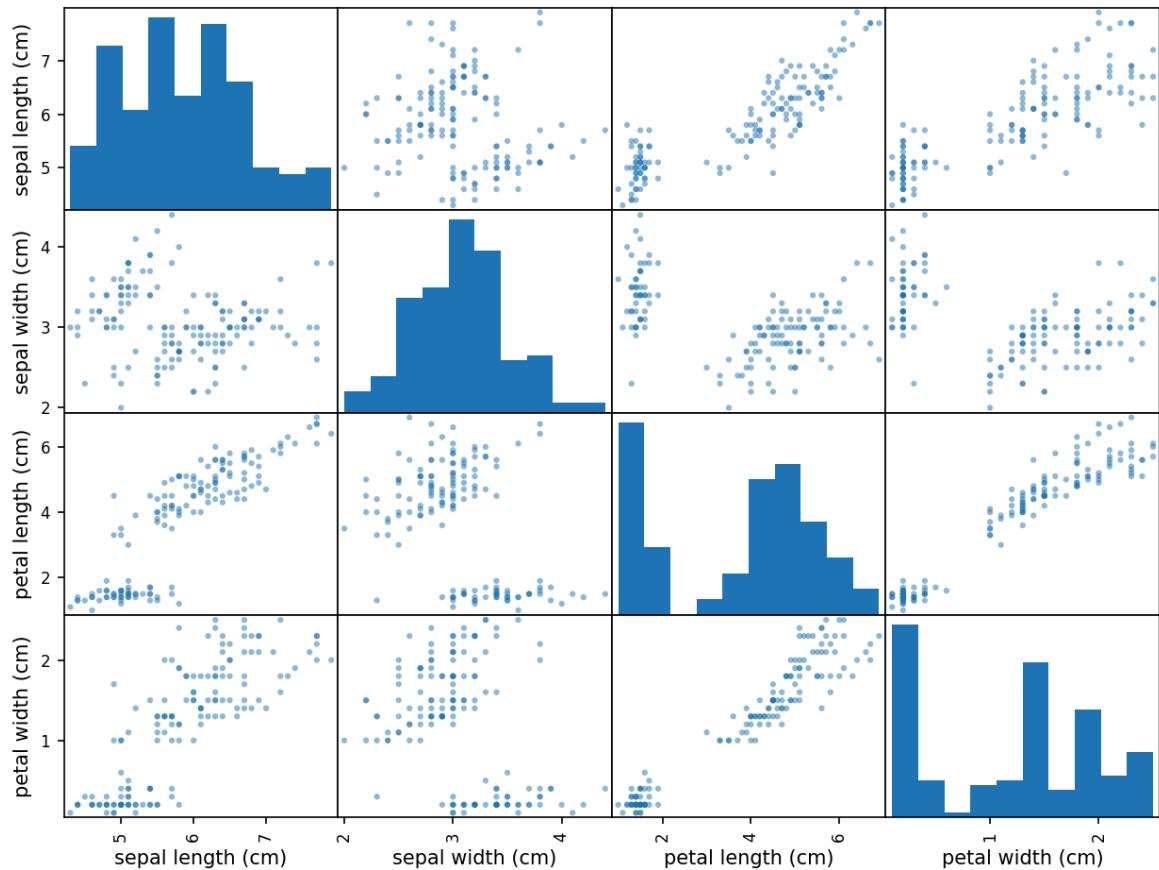
- d. Confirm your answer to #1b by actually finding these values.

```
-----  
Sepal width mean : 3.0573333333333337  
Sepal width median : 3.0  
-----
```

- e. Only 27% of the flowers have a Sepal.Width higher than \_\_\_\_\_ cm.

```
-----  
27 percent of flowers have a sepal width higher than : 3.3  
-----
```

- f. Make scatterplots of each pair of the numerical variables in iris (There should be 6 pairs/plots).



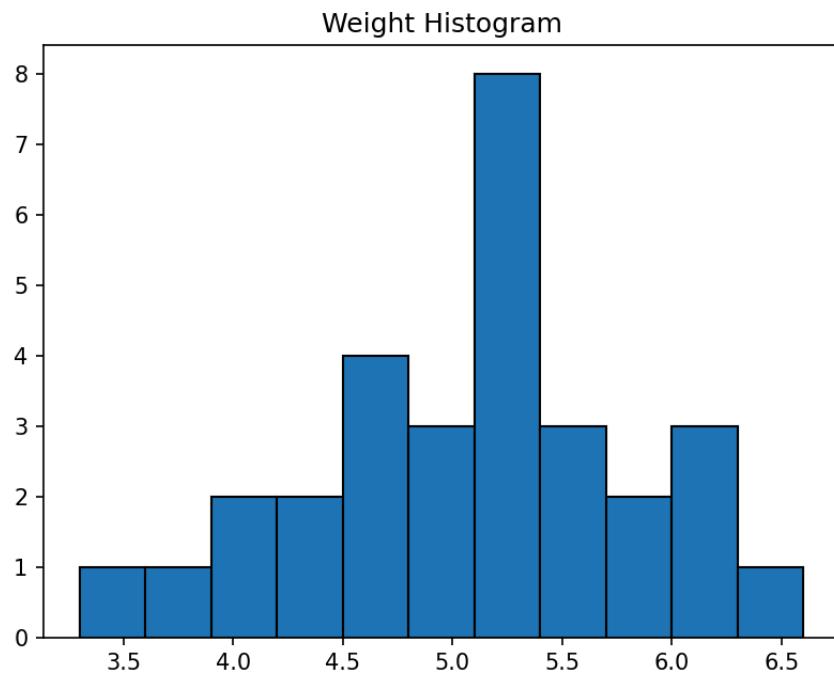
- g. Based on #1e, which two variables appear to have the strongest relationship? And which two appear to have the weakest relationship?

The strongest relationship appears to be between petal length and petal width due to the linear pattern.

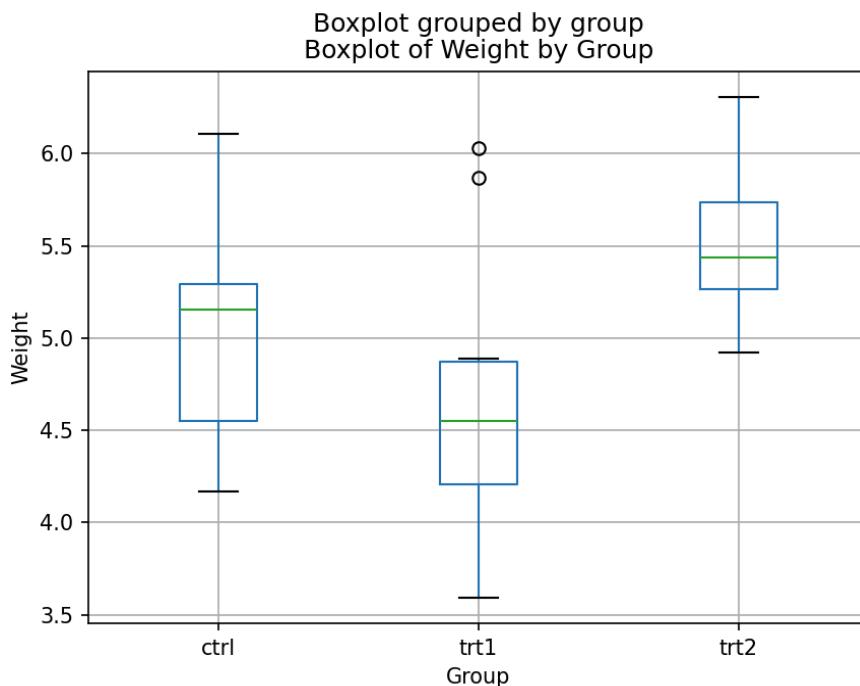
On the other hand the weakest relationship appears to be between sepal length and sepal width.

## Question 2

- a. Make a histogram of the variable weight with breakpoints (bin edges) at every 0.3 units, starting at 3.3.



- b. Make boxplots of weight separated by group in a single graph.



- c. Based on the boxplots in #2b, approximately what percentage of the "trt1" weights are below the minimum "trt2" weight?

There were 2 points from trt1 above the trt2 minimum. There are 10 trt1 data points so approximately 80% of the trt1 weights are below the minimum trt2 weight.

- d. Find the exact percentage of the "trt1" weights that are below the minimum "trt2" weight.

d.

Percentage of trt1 weights below minimum trt2 weight: 80.00%

- e. Only including plants with a weight above 5.5, make a barplot of the variable group. Make the barplot colorful using some color palette (in R, try running ?heat.colors and/or check out <https://www.r-bloggers.com/palettes-in-r/>).

