Week 3 Assignment

William Ho

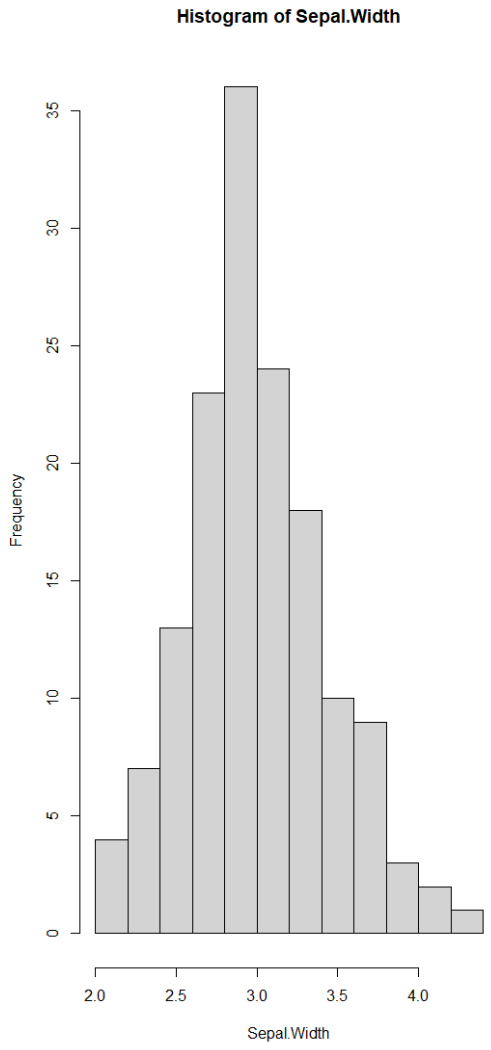
9/14/2025

1a.

Code:

```
#1a.Make a histogram of the variable Sepal.Width.
attach(iris)
hist(Sepal.Width)
```

Histogram:



1b. Based on the histogram in 1a, the histogram is right-skewed, meaning that **mean > median**. This is infered from the "tail" of the histogram extending more to the right. This can indicate that there are some high value outliers that are skewing the mean.

1c.

Based on the output, the mean is greater than the median, confirming that this histogram is right-skewed.

Code:

```
mean(Sepal.Width)
median(Sepal.Width)
```

Output:

```
> mean(Sepal.Width)
[1] 3.057333
> median(Sepal.Width)
[1] 3
>
```

1d.

Only 27% of the flowers have a sepal width higher than 3.3 cm.

Code:

```
quarts=quantile(Sepal.Width,c(.73))
```

Output:

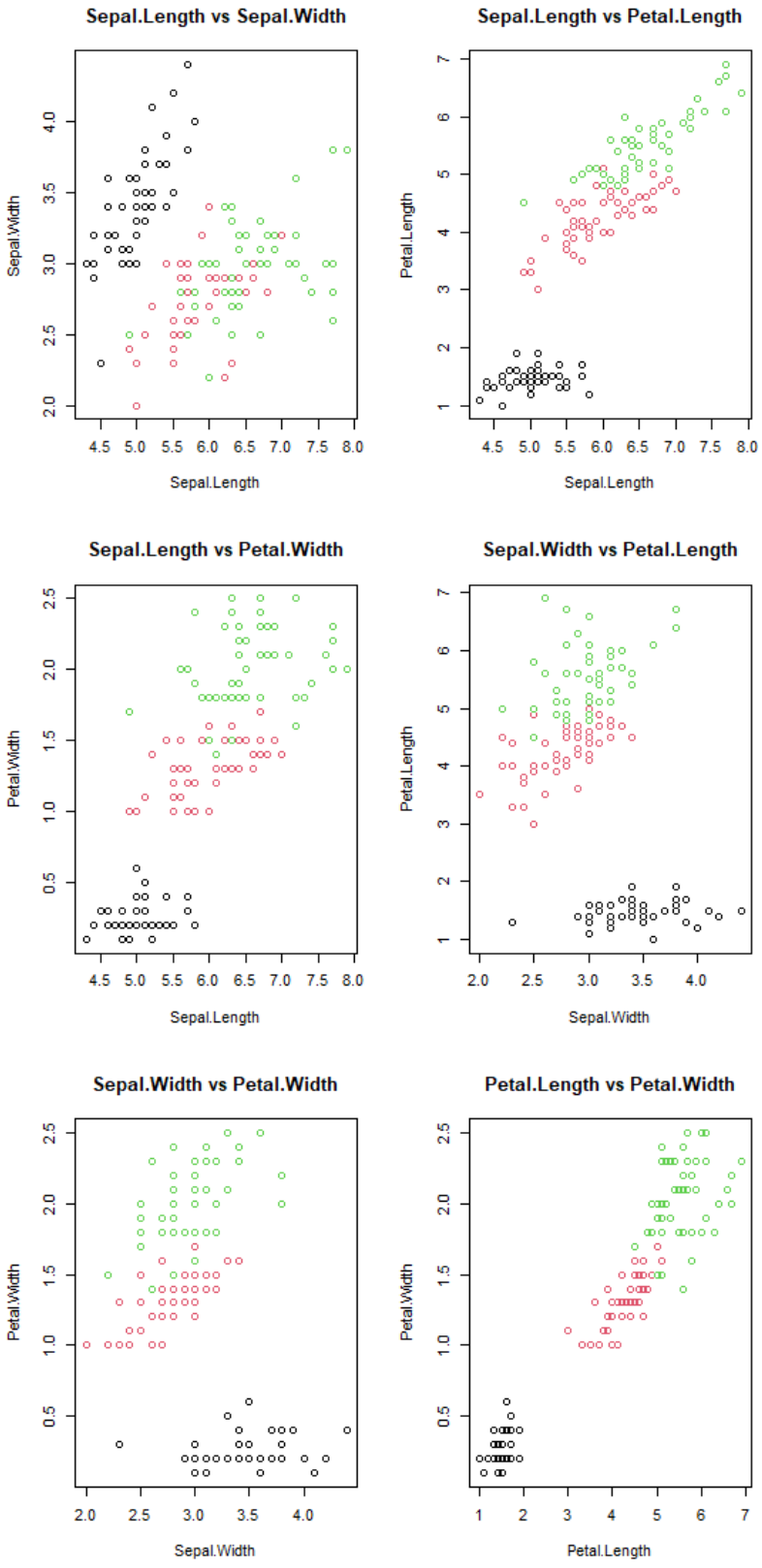
```
> quarts=quantile(Sepal.Width,c(.73))
> quarts
73%
3.3
>
```

1e.

Code:

```
pairs <- list(
  c("Sepal.Length", "Sepal.Width"),
  c("Sepal.Length", "Petal.Length"),
  c("Sepal.Length", "Petal.Width"),
  c("Sepal.Width", "Petal.Length"),
  c("Sepal.Width", "Petal.Width"),
  c("Petal.Length", "Petal.Width")
)

par(mfrow=c(3,2))

for (i in pairs) {
  x <- iris[[i[1]]]
  y <- iris[[i[2]]]

  plot(x,y,
       xlab = i[1],
       ylab = i[2],
       col = iris$Species,
       main = paste(i[1], "vs", i[2]))
}


par(mfrow=c(1,1))
```

Output:


Sepal.Length vs Sepal.Width
Sepal.Width
Sepal.Length
Sepal.Length vs Petal.Length
Petal.Length
Sepal.Length
Sepal.Length vs Petal.Width
Petal.Width
Sepal.Length
Sepal.Width vs Petal.Length
Petal.Length
Sepal.Width
Sepal.Width vs Petal.Width
Petal.Width
Sepal.Width
Petal.Length vs Petal.Width
Petal.Width
Petal.Length

1f.

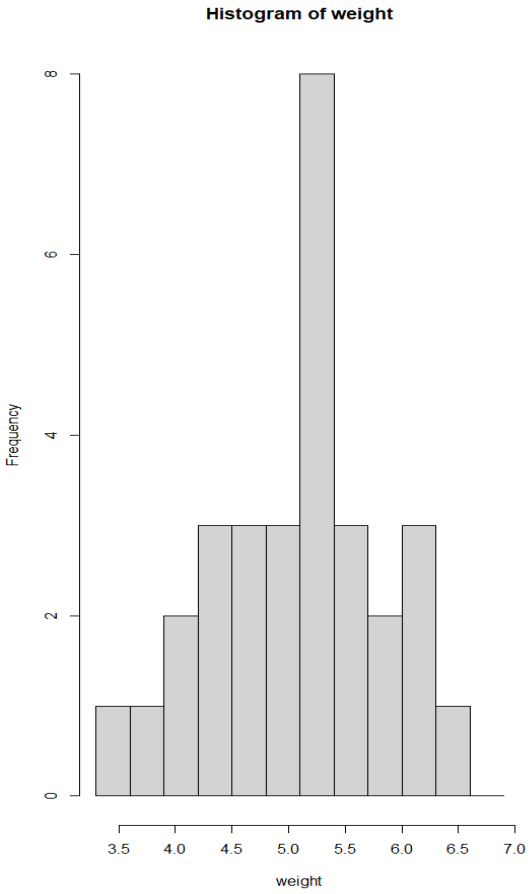
Based on 1e, variables Petal.Length vs Petal.Width and Sepal.Length vs Petal.Length seem to have the strongest relationship, due to a strong positive correlation between them. Variables Sepal.Width vs Petal.Length and Sepal.Width vs Petal.Width seem to have the weakest relationship; there is no negative or positive correlation between them.

2a.

Code:

```
head(PlantGrowth)
attach(PlantGrowth)
hist(weight,breaks=seq(3.3,7,0.3))
```

Output:

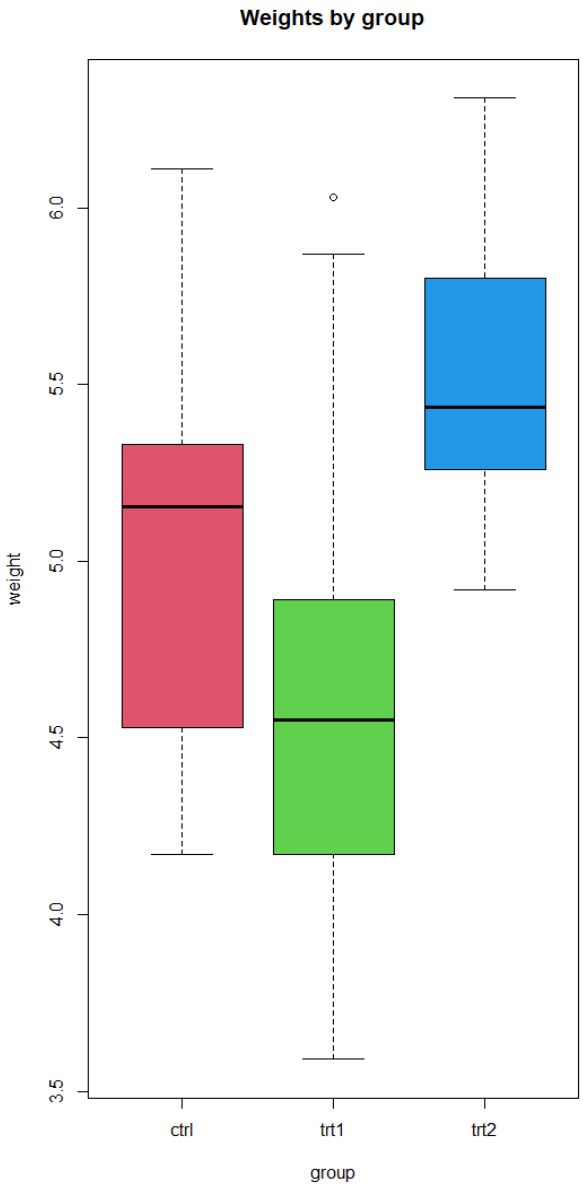


2b.

Code:

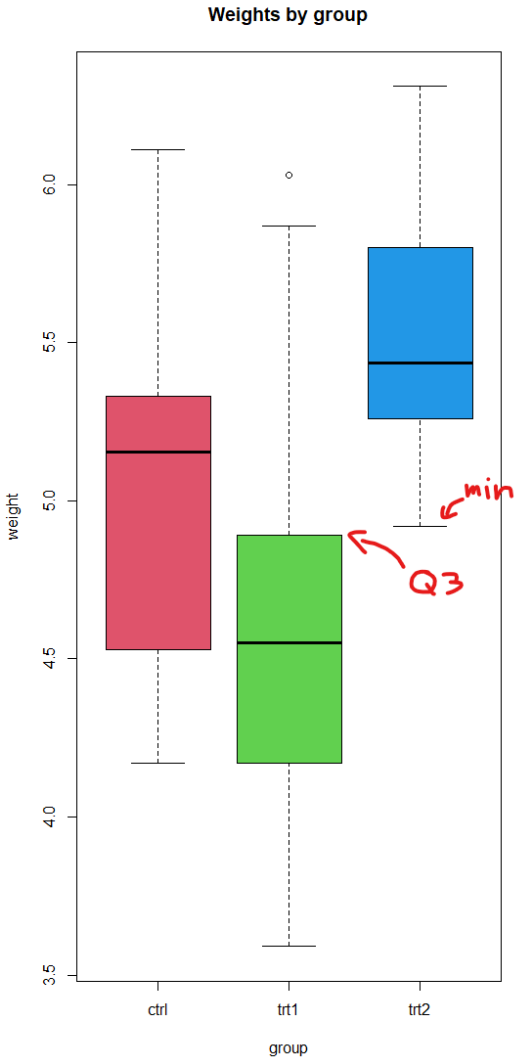
```
boxplot(weight~group,col=2:4,
        main="weights by group")
```

Output:



2c.

Based on the boxplots in 2b, about 75% of the “trt1” weights are below the minimum “trt2” weights. Below I have pointed out why around 75% of the “trt1” weights are below the minimum “trt2” weights.



2d.

80% of the weights in group “trt1” are below the minimum value of weights in group “trt2”

Code:

```
#find min value of weights in trt2
min_trt2 <- min(PlantGrowth$weight[PlantGrowth$group=="trt2"])
#data frame of just weights in group trt1
df_trt1 <- PlantGrowth$weight[PlantGrowth$group=="trt1"]
#find sum of trt1 weights that are below min of trt2
sum_below <- sum(df_trt1 < min_trt2)
#calcuate percentage
percent_below <- (sum_below/length(df_trt1))*100
percent_below
```

Output:

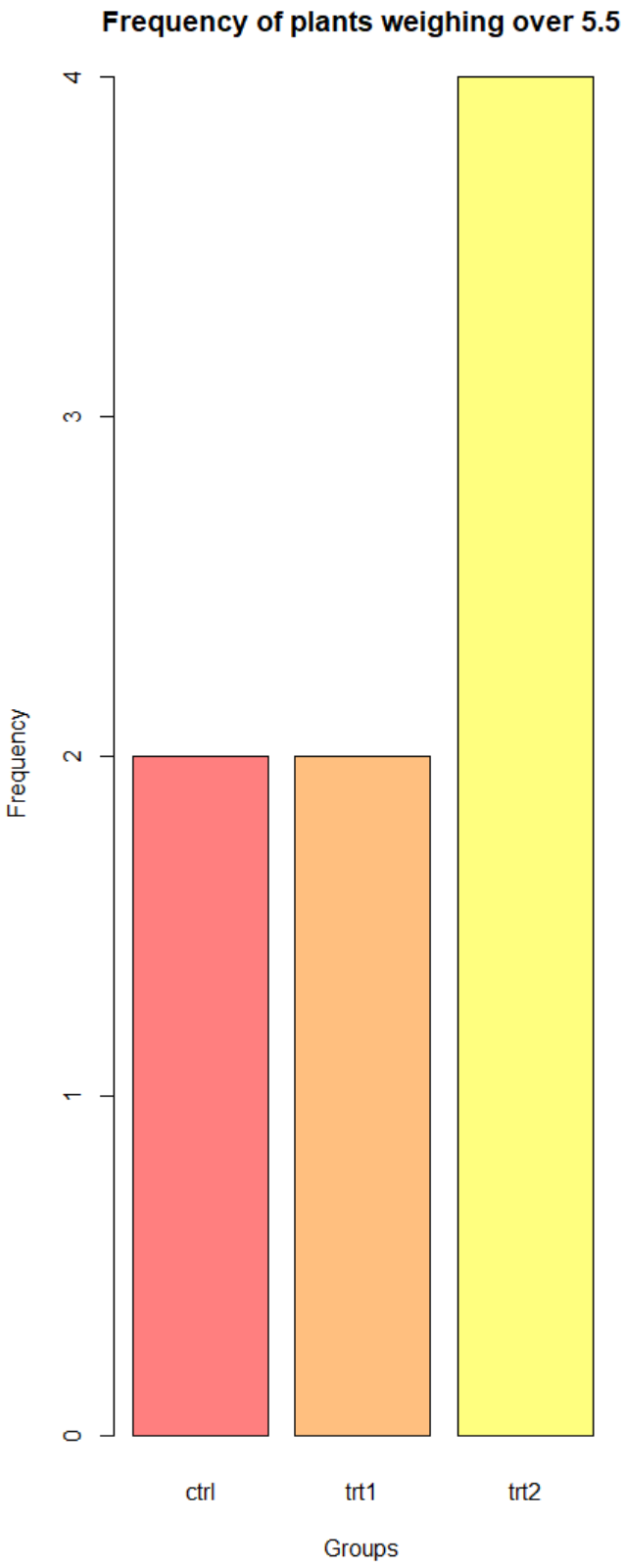
```
> #find min value of weights in trt2
> min_trt2 <- min(PlantGrowth$weight[PlantGrowth$group=="trt2"])
> #data frame of just weights in group trt1
> df_trt1 <- PlantGrowth$weight[PlantGrowth$group=="trt1"]
> #find sum of trt1 weights that are below min of trt2
> sum_below <- sum(df_trt1 < min_trt2)
> #calcuate percentage
> percent_below <- (sum_below/length(df_trt1))*100
> percent_below
[1] 80
>
```

2e.

Code:

```
#dataframe wiht just plants that weigh more than 5.5
df_5.5 <- PlantGrowth$group[PlantGrowth$weight>5.5]
#Create frequency table
group_cat <- factor(df_5.5)
tcat <- table(group_cat)

barplot(tcat,
        main="Frequency of plants weighing over 5.5",
        xlab ="Groups",
        ylab = "Frequency",
        col = heat.colors(n=3,alpha=0.5,rev=FALSE)
        )
```

Output:


Frequency of plants weighing over 5.5
Frequency
0
1
2
3
4
ctrl
trt1
trt2
Groups