Exercise 7

In [1]:
```
str(PlantGrowth)
summary(PlantGrowth)
?PlantGrowth
```

```
'data.frame':   30 obs. of  2 variables:
 $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
 $ group : Factor w/ 3 levels "ctrl","trt1",..: 1 1 1 1 1 1 1 1 1 1 ...
     weight       group   
 Min.   :3.590   ctrl:10  
 1st Qu.:4.550   trt1:10  
 Median :5.155   trt2:10  
 Mean   :5.073            
 3rd Qu.:5.530            
 Max.   :6.310            
```

PlantGrowth {datasets} R Documentation

Results from an Experiment on Plant Growth

Description

Results from an experiment to compare yields (as measured by dried weight of plants) obtained under a control and two different treatment conditions.

Usage

`PlantGrowth`

Format

A data frame of 30 cases on 2 variables.

[, 1]	weight	numeric
[, 2]	group	factor

The levels of `group` are 'ctrl', 'trt1', and 'trt2'.

Source

Dobson, A. J. (1983) *An Introduction to Statistical Modelling*. London: Chapman and Hall.

Examples

```
## One factor ANOVA example from Dobson's book, cf. Table 7.4:
require(stats); require(graphics)
boxplot(weight ~ group, data = PlantGrowth, main = "PlantGrowth data",
        ylab = "Dried weight of plants", col = "lightgray",
```

```
        notch = TRUE, varwidth = TRUE)
anova(lm(weight ~ group, data = PlantGrowth))
```

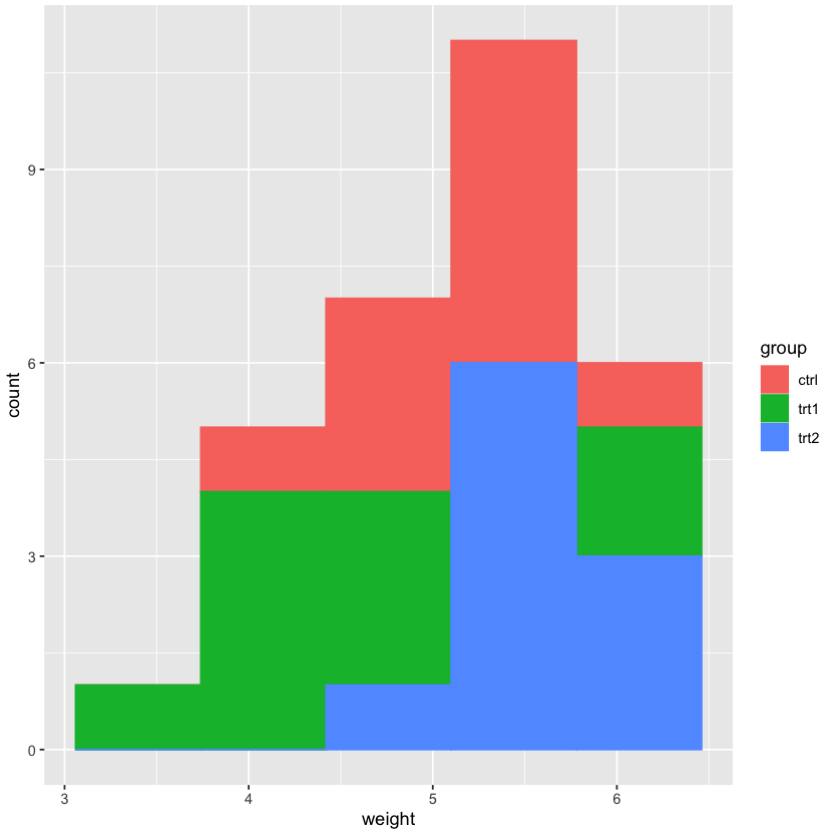
[Package *datasets* version 4.0.3]

In [3]:
```
summary(PlantGrowth)
```

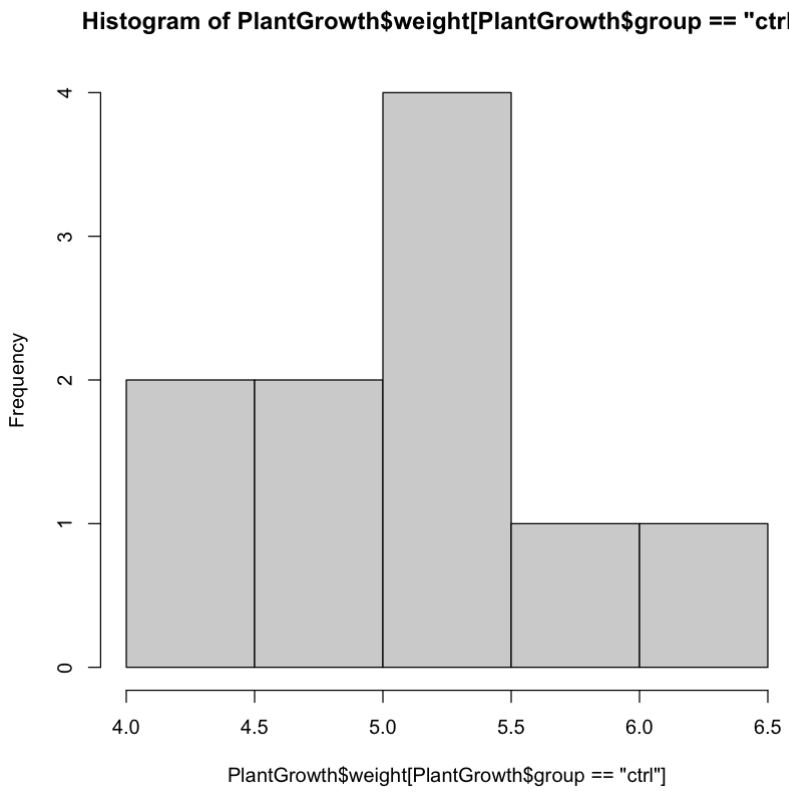
```
    weight       group   
 Min.   :3.590   ctrl:10  
 1st Qu.:4.550   trt1:10  
 Median :5.155   trt2:10  
 Mean   :5.073            
 3rd Qu.:5.530            
 Max.   :6.310            
```

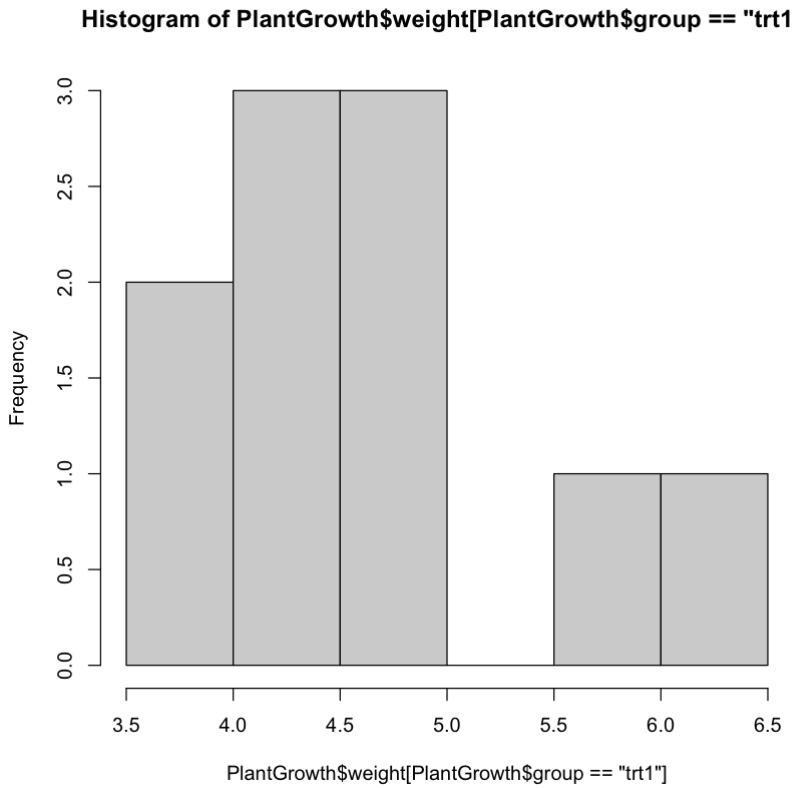
The summary command returns high level statistics of the two columns weight and group in the PlantGrowth dataframe. We can see how the weight column is a numeric column that ranges from 3.59 to 6.31. The group column is a non-numeric column that contains three groups: ctrl=control group, trt1=test1, and trt2=test2 where each group has ten entries.

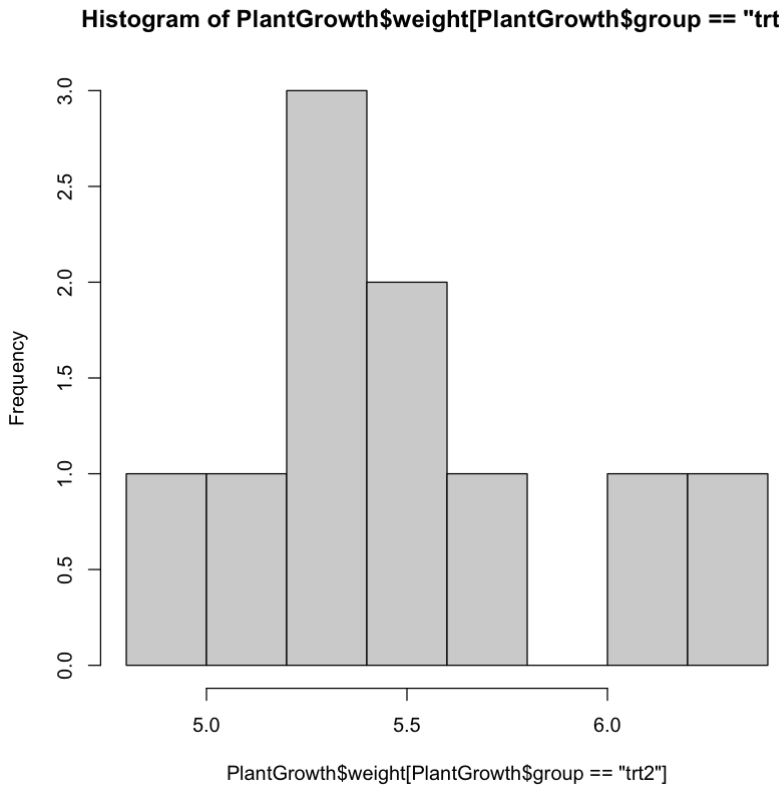
In [16]:
```
library(ggplot2)
ggplot(PlantGrowth, aes(x = weight)) +
  geom_histogram(aes(color = group, fill = group), bins = 5)
```



In [6]:
```
hist(PlantGrowth$weight[PlantGrowth$group=='ctrl'])
hist(PlantGrowth$weight[PlantGrowth$group=='trt1'])
hist(PlantGrowth$weight[PlantGrowth$group=='trt2'])
```


Histogram of PlantGrowth$weight[PlantGrowth$group == "ctrl"]
Frequency
0
1
2
3
4
4.0
4.5
5.0
5.5
6.0
6.5
PlantGrowth$weight[PlantGrowth$group == "ctrl"]

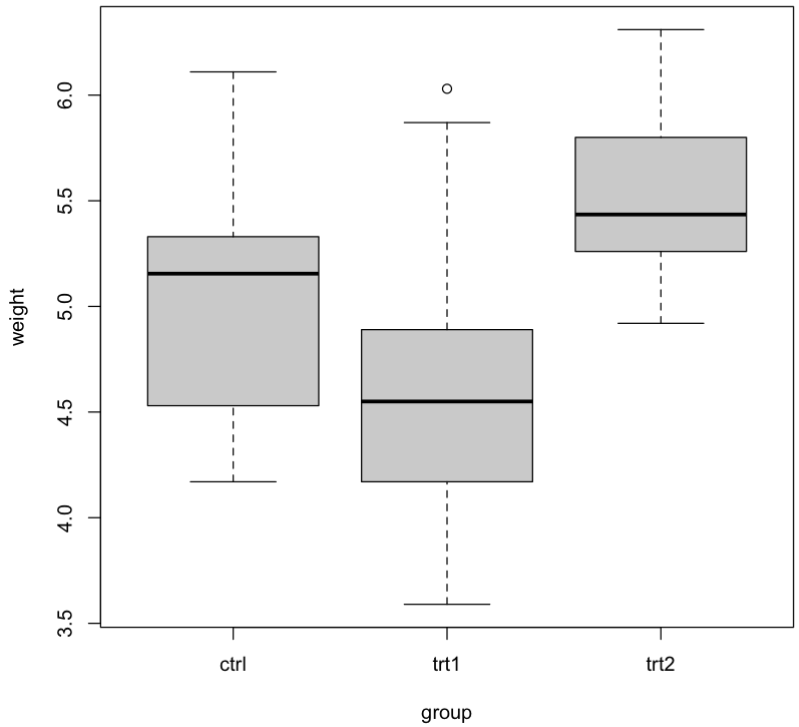

Histogram of PlantGrowth$weight[PlantGrowth$group == "trt1"]
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
PlantGrowth$weight[PlantGrowth$group == "trt1"]



It's hard to make any conclusions from the three histograms as they all have different x axis'. The main difference is that the center of each histogram is different for trt1 compared to the rest. Group trt1 shows a majority of its weight below 5 while other groups are closer to 5.5.

Exercise 8

In [20]:
```
boxplot(weight ~ group, data=PlantGrowth) # Boxplot of mpg, group by am
```



The trt1 has a higher variation than the rest of the groups. It's lower half is the lowest of the 3 groups. Group trt2 has the distribution with the highest center. It all of its data lies past 75% of the weights in group trt1 while at least 50% of its data lies past 75% of the weights in the ctrl group.

Exercise 9

In [21]:
```
t.test(PlantGrowth$weight[ PlantGrowth$group == 'ctrl' ], PlantGrowth$weight[ Pl
```

```
        Welch Two Sample t-test

data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[Pl
antGrowth$group == "trt1"]
t = 1.1913, df = 16.524, p-value = 0.2504
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.2875162  1.0295162
sample estimates:
mean of x mean of y
    5.032     4.661
```

The confidence interval of the t-test above means that out of 100 replications of the study, 95% of the confidence intervals would contain the mean difference. I.e., the mean difference of the weight between crtl and trt1 will be within ~-.29 and ~1 95 out of 100 replications of the study.

The results of the T-Test also means that in a 95% significance level, we cannot reject the null hypothesis (H0 = both distributions have a different mean).

Exercise 10

In [22]:
```
t.test(PlantGrowth$weight[ PlantGrowth$group == 'ctrl' ], PlantGrowth$weight[ Pl
```

```
        Welch Two Sample t-test

data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[Pl
antGrowth$group == "trt2"]
t = -2.134, df = 16.786, p-value = 0.0479
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.98287213 -0.00512787
sample estimates:
mean of x mean of y
    5.032     5.526
```

The mean difference of the weight between crtl and trt12 will be within ~.5 and ~5.5 95 out of 100 replications of the study. The T-Test result also tells us that with a 95% significance level, we reject the null hypothesis. I.e., 95% of the studies will find that the means of the two groups are different from each other.