

Andrew Alford Homework 4

1. The built-in PlantGrowth data set contains three different groups, each representing a different plant food diet (you may need to type `data(PlantGrowth)` to activate it). The group labeled “ctrl” is the control group, while the other two groups are each a different type of experimental treatment. Run the `summary()` command on PlantGrowth and explain the output. Create a histogram of the ctrl group. As a hint about R syntax, here is one way that you can access the ctrl group data:
`PlantGrowth$weight[PlantGrowth$group=="ctrl"]` Also create histograms of the trt1 and trt2 groups. What can you say about the differences in the groups by looking at the histograms?

In this question there are a number of different tasks. I will break them down one by one in order from first to last. The first task is to do the `summary()` function on the PlantGrowth data set. After loading the data set and calling the function, I received the following output:

```
> 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
> |
```

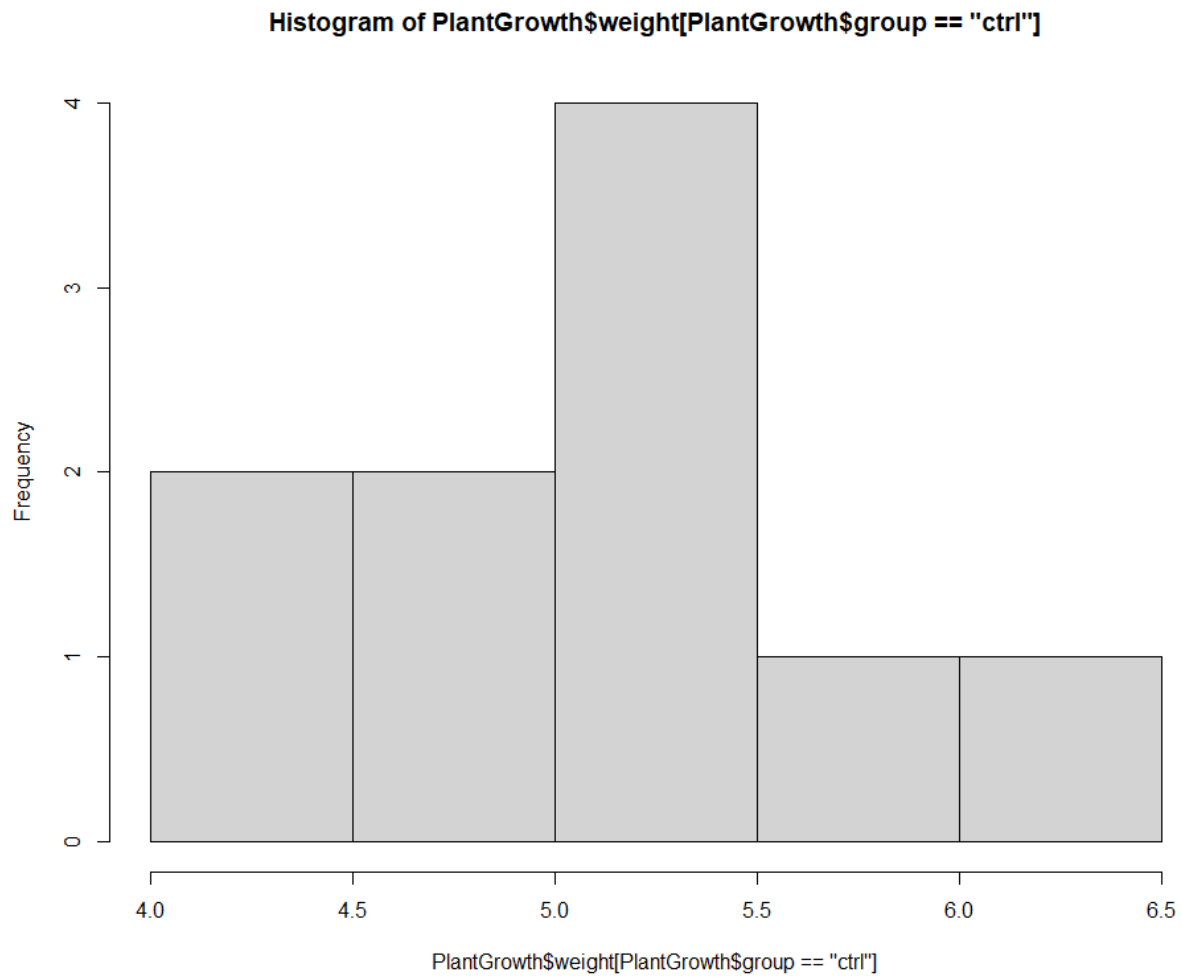
There are a lot of different values in this output that we can interpret. The first is min, which has the value of 3.590. This means that the minimum weight for all plants was

3.590. The second value is the 1st Quartile of 4.550 which means that 25% of the observations fall below this weight. The median value is 5.155, which separates the higher half of plant weights from the lower half in the data set. Then we have the mean, which is the average weight of plants across the data set. The third quartile is similar to the first quartile in that 75% of values in the data set fall below 5.530. Finally, the max value is the heaviest plant observed in the data. From the group column, it displays the number of observations in each group, of which there are three: ctrl, trt1, and trt2. There are ten observations in each group, making thirty in total.

Next, the question asks us to make a histogram of the control group. This is done simply with the syntax provided in the question, where one simply needs to wrap it in a `hist()` function to create the histogram. In total, the syntax is as follows:

```
hist(PlantGrowth$weight[PlantGrowth$group=="ctrl"])
```

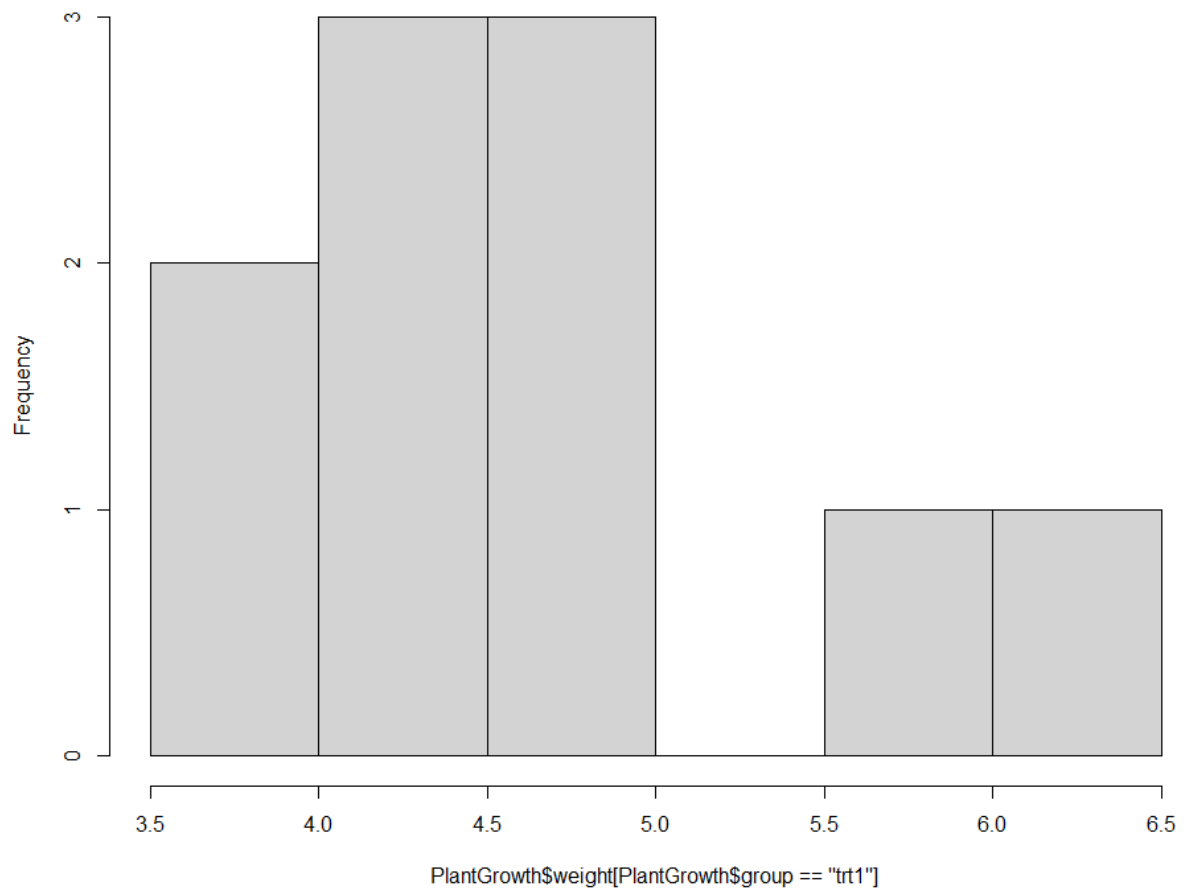
The output from this function looks like this:



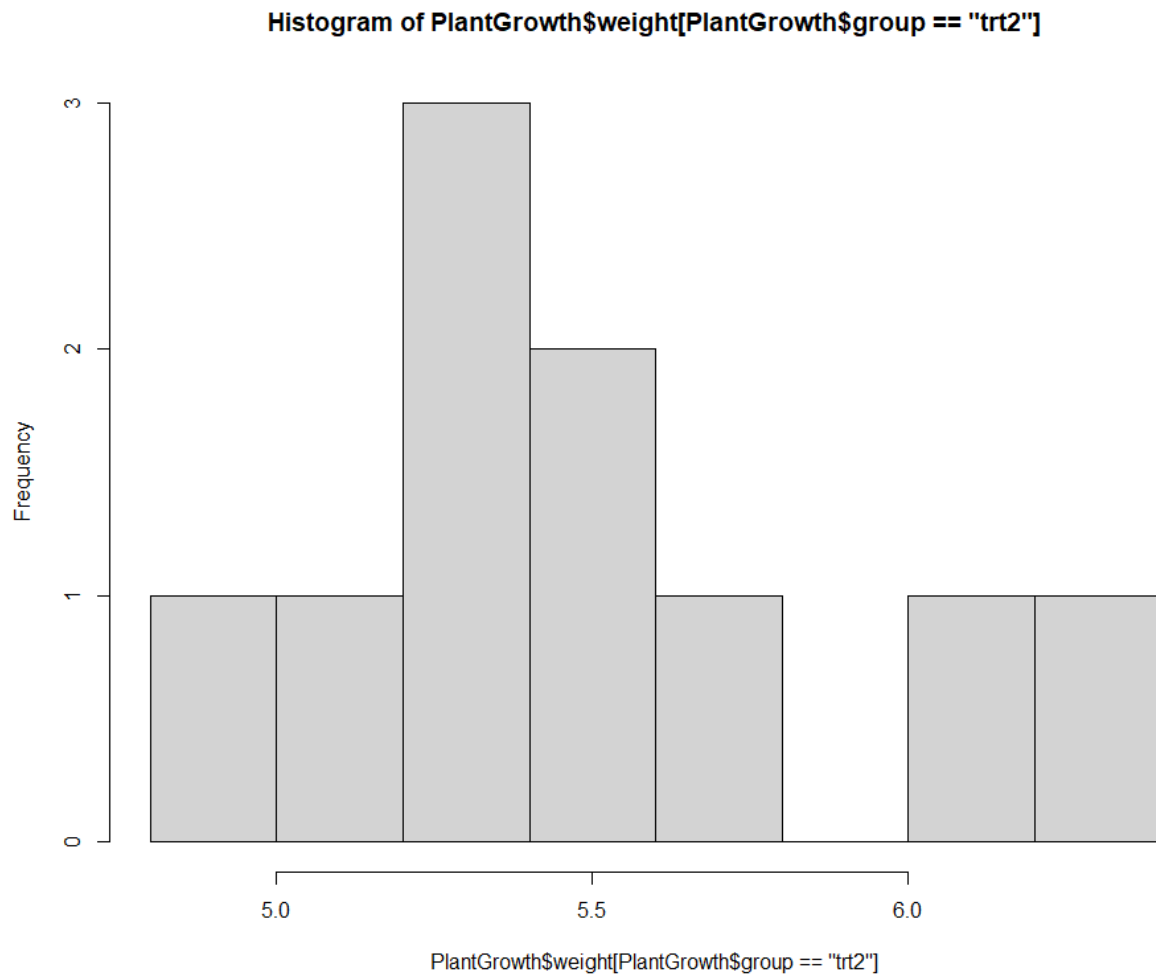
To create the other two histograms needed in the question, one simply alters the group selected in the syntax from “ctrl” to “trt1” or “trt2” to create a histogram of those groups.

Here are the other group histograms

Histogram of PlantGrowth\$weight[PlantGrowth\$group == "trt1"]



↑↑↑ Trt1 ↑↑↑



↑↑↑ Trt2 ↑↑↑

Given that there are only ten observations in each group, which is a very small sample size, one should not extrapolate too much on larger implications for the hypothetical plants in this data set. However, there are notable differences. For example, the control group has a stronger central tendency, with four observations of between 5.0 and 5.5. The trt1 group has a slightly greater deviation than the control group, with some

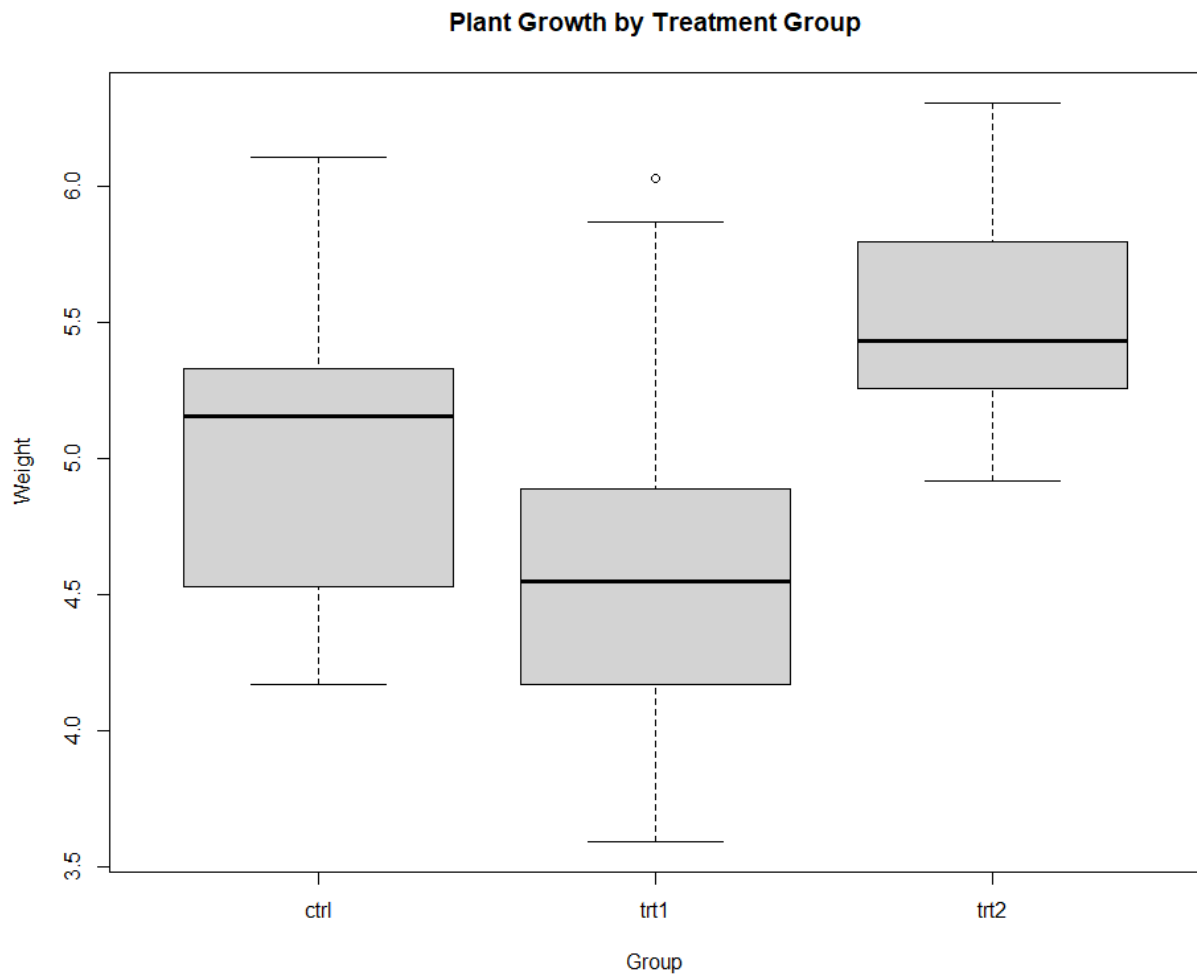
observations as low as 3.5, but still as high as 6.5. The trt2 group has a small deviation of only 5.0 and 6.0.

- 2. Create a boxplot of the plant growth data, using the model “weight ~ group.” What can you say about the differences in the groups by looking at the boxplots for the different groups?**

The first step in this question is to create the boxplots. To do this, I used the following code:

```
boxplot(weight ~ group, data = PlantGrowth, main = "Plant Growth by Treatment  
Group", xlab = "Group", ylab = "weight")
```

With this code, I got the following plot:



There is a lot of fantastic information to unpack here. I will look at a number of different statistical differences outlined below.

a. Central Tendency:

The median weight for “ctrl” is higher than “trt1” but lower than “trt2”. This indicates that on average, the plants in “trt2” grew the most, while those in “trt1” grew the least.

b. Spread:

The “ctrl” group has the broadest interquartile range, indicating greater variability in plant growth. The “trt2” group has a narrow interquartile

range, indicating less variability in plant growth. “Trt1” is somewhere in the middle between the other two.

c. Outliers:

The “trt1” group has an outlier on the upper side, which means there’s at least one plant in this group that grew exceptionally well compared to others in the same group.

d. Skewness:

All boxes seem relatively symmetric about their median, indicating no significant skewness in the data.

e. Overlap:

The boxes for “ctrl” and “trt2” have significant overlap, indicating similar distributions for these groups. The “trt1” box is distinctly lower, indicating that this group’s plant growth is generally lesser than the other two groups.

Findings:

The “trt2” group seems to perform best overall in terms of plant growth, with the highest median growth. The “trt1” group performs the least well, with the lowest median growth and also the least variability. This suggests that this treatment might be consistently less effective across multiple plants, although it’s important to note that the sample size is very small.

3. Run a *t*-test to compare the means of ctrl and trt1 in the PlantGrowth data.

Report and interpret the confidence interval. Make sure to include a carefully worded statement about what the confidence interval implies with respect to the population mean difference between the ctrl and trt1 groups.

First things first, to create the t-test, I used the following code and got the following output:

Code

```
t.test(PlantGrowth$weight[PlantGrowth$group=="ctrl"],  
PlantGrowth$weight[PlantGrowth$group=="trt1"])
```

Output

```
welch Two Sample t-test  
  
data:  PlantGrowth$weight[PlantGrowth$group == "ctrl"] and PlantGrowth$weight[PlantGrowth$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
```

Using this information, we can interpret significant results. The t-value is 1.1913 and the p-value is 0.2504, which is greater than the significance of 0.05. This means that we fail to reject the null hypothesis, implying that there isn't enough evidence to say that there's a statistically significant difference between the means of the ctrl and trt1 groups. In terms of the confidence interval, the interval for the difference in means between the

two groups is -0.2875162 to 1.0295162. Given that this interval contains 0, it suggests there's a possibility that there's no difference in means between the two groups.

4. Run a t-test to compare the means of ctrl and trt2 in the PlantGrowth data.

Report and interpret the confidence interval.

The nice thing about this question is that we can reuse the same code as above, simply changing the string "trt1" to "trt2". The t-value is -2.134 and the p-value is 0.0479, which is just below the common significance level of 0.05. This means that we reject the null hypothesis, suggesting that there is a statistically significant difference between the means of the ctrl and trt2 groups. The confidence interval means that we are 95% confident that the true population mean difference in plant growth between the ctrl and trt2 groups lies between a decrease of approximately 0.983 and a decrease of approximately 0.005. Since the entire interval lies below zero, it suggests that the trt2 group has a significantly higher mean than the ctrl group.