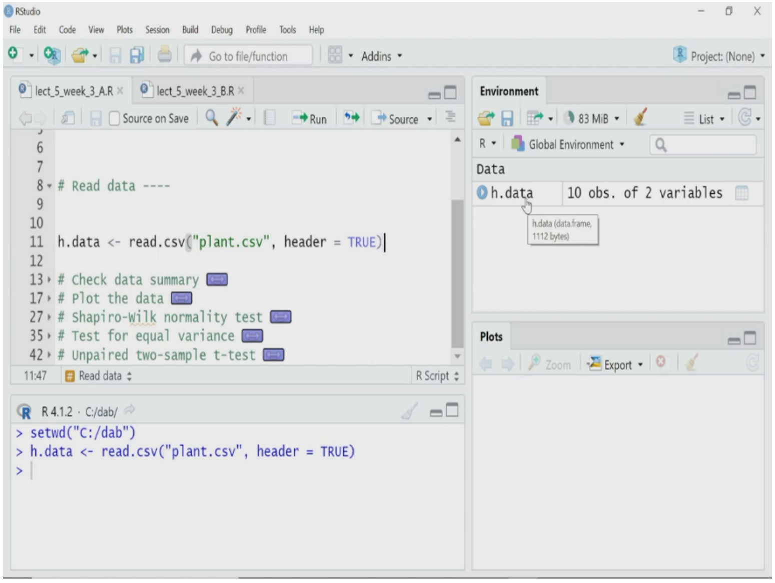
Data Analysis for Biologists
Professor Biplab Bose
Department of Biosciences and Bioengineering
Mehta Family School of Data Science and Artificial Intelligence
Indian Institute of Technology Guwahati
Lecture 18
Statistics using R – t-test and ANOVA

Hello welcome back. In biology, we use diverse type of statistical analysis. Using R you can actually perform those analysis very easily, I will not go in detail of each of those statistical analysis. In this lecture as an example, we will learn how to perform t-test and ANOVA. So, let us take two example, for t-test and ANOVA and perform using R Studio, I am using R Studio but you can use our native R also because these functions that are available by default in native R also. So, let us start and with the t-test and see how to perform that.

(Refer Slide Time: 01:11)



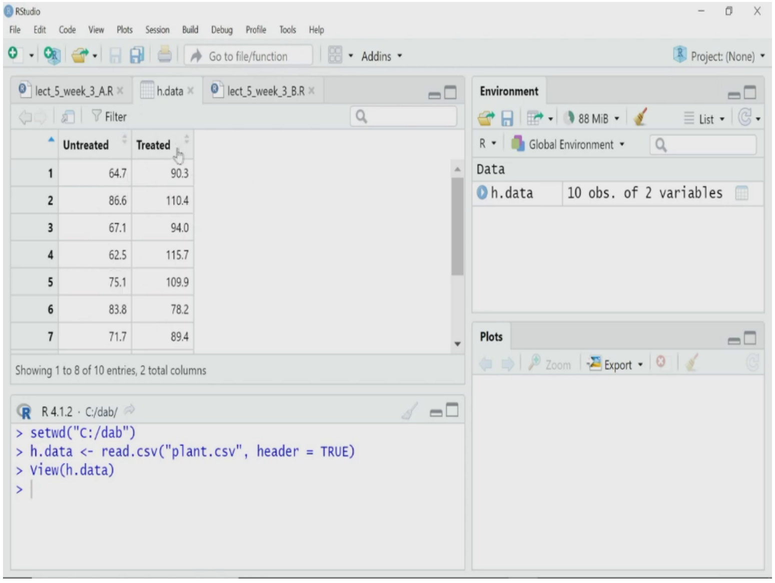
```
# Read data ----


h.data <- read.csv("plant.csv", header = TRUE)

# Check data summary
# Plot the data
# Shapiro-Wilk normality test
# Test for equal variance
# Unpaired two-sample t-test
```

```
> setwd("C:/dab")
> h.data <- read.csv("plant.csv", header = TRUE)
>
```

Data

h.data 10 obs. of 2 variables

h.data (data.frame, 1112 bytes)



	Untreated	Treated
1	64.7	90.3
2	86.6	110.4
3	67.1	94.0
4	62.5	115.7
5	75.1	109.9
6	83.8	78.2
7	71.7	89.4

Showing 1 to 8 of 10 entries, 2 total columns

Data

h.data	10 obs. of 2 variables

```
R 4.1.2 · C:/dab/
> setwd("C:/dab")
> h.data <- read.csv("plant.csv", header = TRUE)
> View(h.data)
>
```

So, I want to perform in this example, I want to perform t-test on a data set where we are treating cells with something suppose some molecule or something else. So, that, we can change the growth behavior of the plant. So, we are measuring the height of the plant as a measure of growth. And we have two samples, one is control and another one is treated one and we want to see whether they are behaving differently or not.

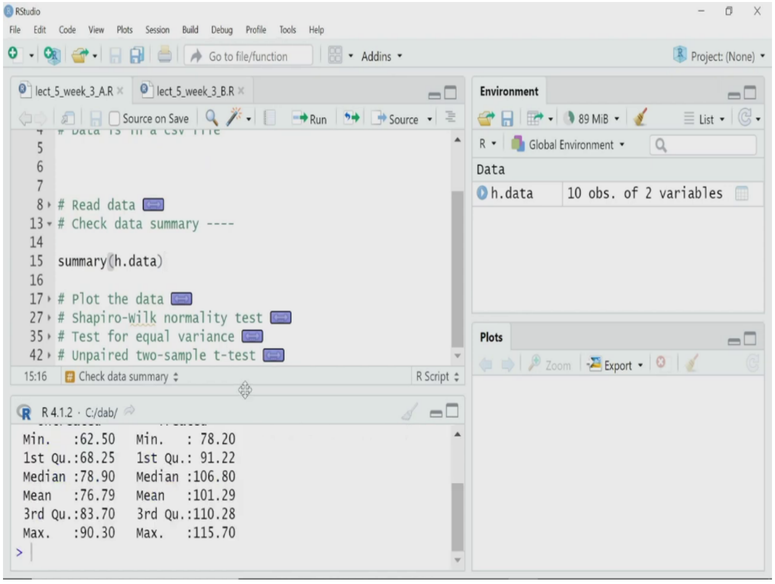
This is a very common example in experiments where you have a control sample, where you have a treated sample and you want to see whether the difference observed between these two is statistically significant or not. In that case, we usually use t-test and I have a data in CSV format. So, I will read that CSV file and then I will perform t-test. So, let us do this step by step. So, the first thing that I have to do is to read the data.

$$\text{h.data} \leftarrow \text{read.csv(“plant.csv”, header = TRUE)}$$

Again, I will use the read dot CSV function, I hope by now you are habituated with this read dot CSV function. My input is, the arguments are plant dot CSV file that has that file, and I am keeping the header equal to TRUE because it has a header and I will assign these to a variable h dot data. So, I have to work on this h dot data. So, let us read it. So, I have read it.

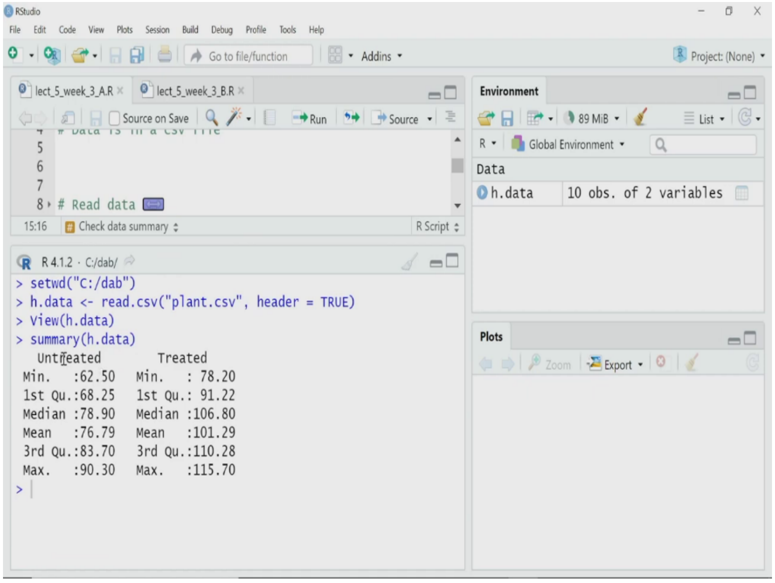
Now, you can see the data here let me click and open. It is a two column data as I said, I have untreated plant, control plant and treated plant, treated with something we do not bother about that. So, how many samples are there? 10, 10 samples are there, it is also written here in the environment that I have 10 observations for two variables. Let me close this data.

(Refer Slide Time: 02:56)



```
# Data is in a csv file



# Read data
# Check data summary ----

summary(h.data)

# Plot the data
# Shapiro-Wilk normality test
# Test for equal variance
# Unpaired two-sample t-test
```

```
Min.   :62.50   Min.   : 78.20
1st Qu.:68.25   1st Qu.: 91.22
Median :78.90   Median :106.80
Mean   :76.79   Mean   :101.29
3rd Qu.:83.70   3rd Qu.:110.28
Max.   :90.30   Max.   :115.70
>
```



```
# Data is in a csv file



# Read data
```

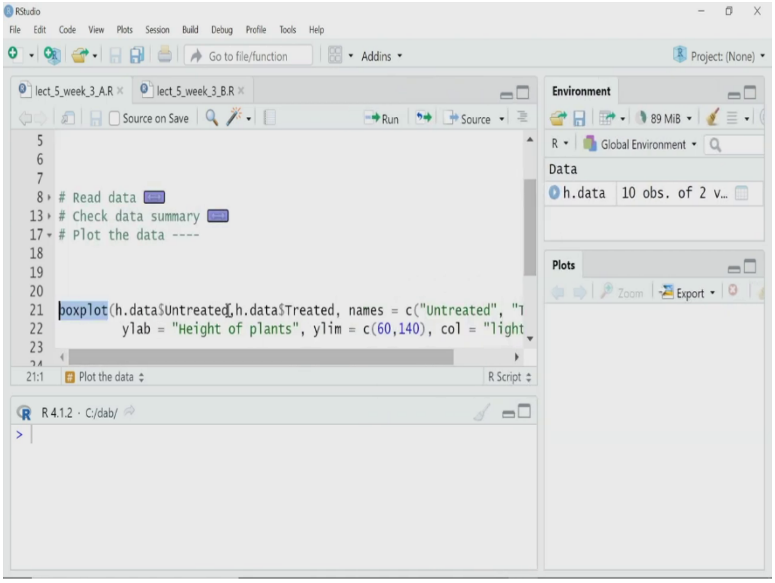
```
> setwd("C:/dab")
> h.data <- read.csv("plant.csv", header = TRUE)
> View(h.data)
> summary(h.data)
   Untreated        Treated
 Min.   :62.50   Min.   : 78.20
 1st Qu.:68.25   1st Qu.: 91.22
 Median :78.90   Median :106.80
 Mean   :76.79   Mean   :101.29
 3rd Qu.:83.70   3rd Qu.:110.28
 Max.   :90.30   Max.   :115.70
>
```

Now, I have read the data I want to do the first thing as I said once you get a data, it is better to check the summary of that. So, I can call my helpful function called summary to calculate that summary, get the summary. And let me check the summary here.

summary(h.data)

So, I have two variables untreated and treated, and the minimum and maximum are given. So, treated height, untreated height varies from 62.5 to 90.3. Whereas for treated you can easily see the minimum is also higher and the maximum is also higher. The mean value for untreated control plant is 76.79, whereas the mean for the treated population is 101.29. So, obviously, it seems on an average the treatment has increased the height of the plant, but we have to do statistical tests for that to be assured that it is actually statistically significant or not.

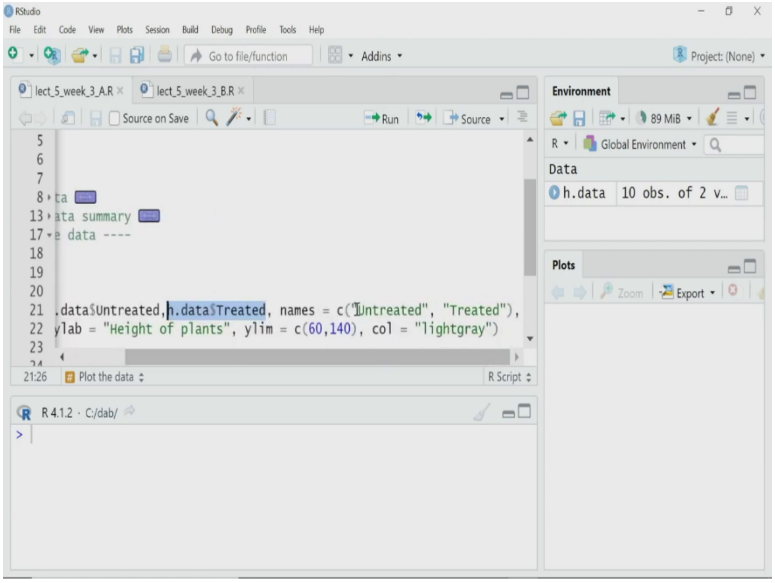
(Refer Slide Time: 03:55)



```
# Read data
# Check data summary
# Plot the data ----



boxplot(h.data$Untreated, h.data$Treated, names = c("Untreated", "T
        ylab = "Height of plants", ylim = c(60,140), col = "light
```

Data
h.data 10 obs. of 2 v...

```
>
```



```
ta
ata summary
e data ----



.data$Untreated, h.data$Treated, names = c("Untreated", "Treated"),
ylab = "Height of plants", ylim = c(60,140), col = "lightgray")
```

Data
h.data 10 obs. of 2 v...

```
>
```

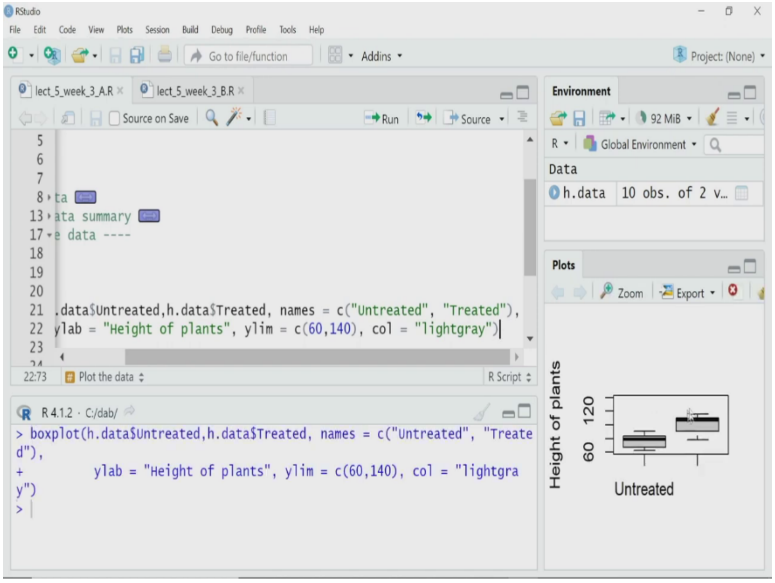
Before we go into that, let me plot these data so that it becomes much more clear that what we are doing, why we are doing a statistical test. Now, if you have these type of two sample two case data, you can actually make a bar plot. But sometimes it is better to draw a box plot. We will have separate lecture where we will discuss how to use R to draw bar plot, then box chart and all these things, but here by default I am using the box plot and we do not go in detail of this box but now just for our today's purpose we will simply draw it.

boxplot(h.data$Untreated, h.data$Treated, names = c(“Untreated”, “Treated”), ylab = “Height of plants”, ylim = c(60,140), col = “lightgray”)

So, what I am doing, I will draw the box plot. Let me increase this plot area. So, to draw blog, a box plot by default, R has some function a function called box plot itself. And it takes multiple arguments. I have used some of those. Let me briefly say what are those. The first two arguments are the data. So, I have two variables. So, this first one is h dot data dollar symbol and untreated that means I am taking the untreated data, this is the first variable.

And then I am extracting the treated data, this is my second variable, because I want two data to be plotted in the same plot. And then I have to name those. So, I am using C function to create a vector or list and the names are untreated and treated. And the variable I have assigned as names and the y-axis or the vertical axis, which is detected by this y lab y label, it will be height of plants, that is what I have specified. I have given limit for the values from where to where I want to plot because otherwise the plot may look bit wrong and then I want to use the light gray as color.

(Refer Slide Time: 05:56)

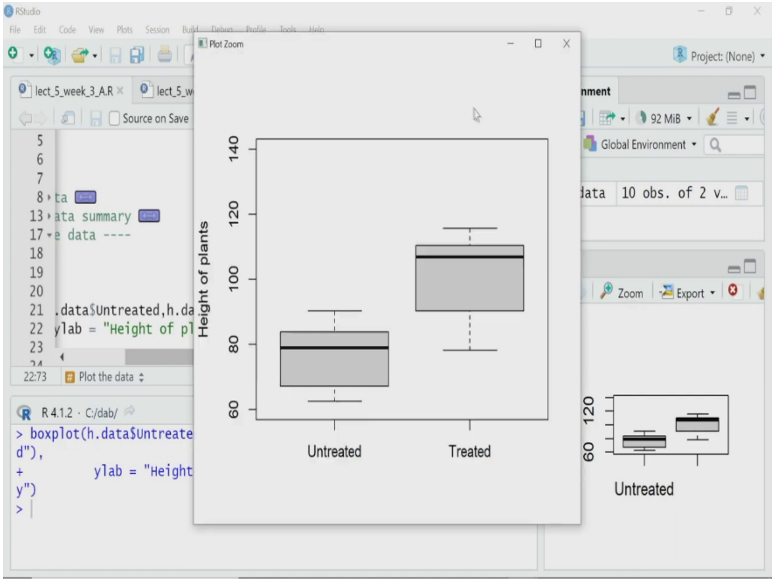


```
ta
ata summary
e data ----



.data$Untreated,h.data$Treated, names = c("Untreated", "Treated"),
ylab = "Height of plants", ylim = c(60,140), col = "lightgray")
```

Data

h.data	10 obs. of 2 v...

```
> boxplot(h.data$Untreated,h.data$Treated, names = c("Untreated", "Treate
d"),
+         ylab = "Height of plants", ylim = c(60,140), col = "lightgra
y")
> 
```



```
ta
ata summary
e data ----


.data$Untreated,h.da
ylab = "Height of pl
```

22:73 Plot the data

```
data   10 obs. of 2 v…
```

R 4.1.2 · C:/dab/

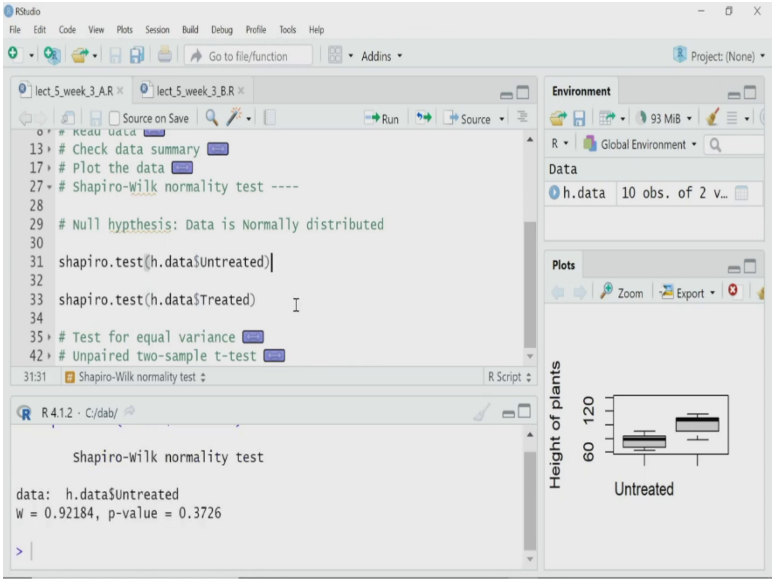
```
> boxplot(h.data$Untreate
d"),
+         ylab = "Height
y")
> 
```

Let us draw it. It will be clear to you why I am using these arguments. So, I have drawn it is bit distorted. So, let me zoom it. So, here we have. So, if you remember I have named this the horizontal axis for these two variables untreated and treated. And I have named or labeled the vertical axis by y lab for height of plants and the box plot has been drawn.

And in box plot these dark line horizontal lines here in both the boxes are the median values. Now, looking at these data, you can easily see and if you remember the mean values that we have calculated for untreated and treated samples, obviously there is difference and possibly it shows that the treated sample has a higher height.

So, that mean treatment of this plant with that particular treatment is increasing the height of the plant. But there is also overlap, is not it, you can see this region we have overlap. So, I have to do statistical tests to check whether this difference between the means of treated and untreated samples is really statistically significant or not. So, I will do t-test, let us go to that to understand and perform it.

(Refer Slide Time: 07:07)



```
# Read data
# Check data summary
# Plot the data
# Shapiro-Wilk normality test ----

# Null hypthesis: Data is Normally distributed

shapiro.test(h.data$Untreated)

shapiro.test(h.data$Treated)

# Test for equal variance
# Unpaired two-sample t-test
```

```
	Shapiro-Wilk normality test

data:  h.data$Untreated
W = 0.92184, p-value = 0.3726

>
```

Data
h.data 10 obs. of 2 v...

Now, I have to perform t-test, what type of t-test I will perform, I will perform unpaired t-test, I will perform unpaired two sample t-test actually. So, before you perform any t-test, we have to remember that t-test is a parametric test. That means inherently it, we have a belief that this data that I am analyzing is normally distributed, is an assumption basic assumption for this t-test. So, I have to check whether my data itself is normally distributed or not.

We have a lecture on t-test earlier. So, you may look into it, there is another issue here. That is that the both the sample here should have equal variance otherwise I cannot perform t-test with equal variance. We have some other options for that. So, what I will do before we jump start the t-test, I have to check whether my data is normally distributed or not. And whether both these treated untreated sample has equal variance or not.

Let us start with checking the normality. Normality of a data set can be checked by different method what I will use here is called the Shapiro-wilk method and R has an inbuilt function to calculate that. What this function will do? This function is called shapiro.test. And it will take that data as an argument and it will perform a statistical test which itself has a null hypothesis, if you remember, t-test, ANOVA all these things will have null hypothesis and alternate hypothesis.

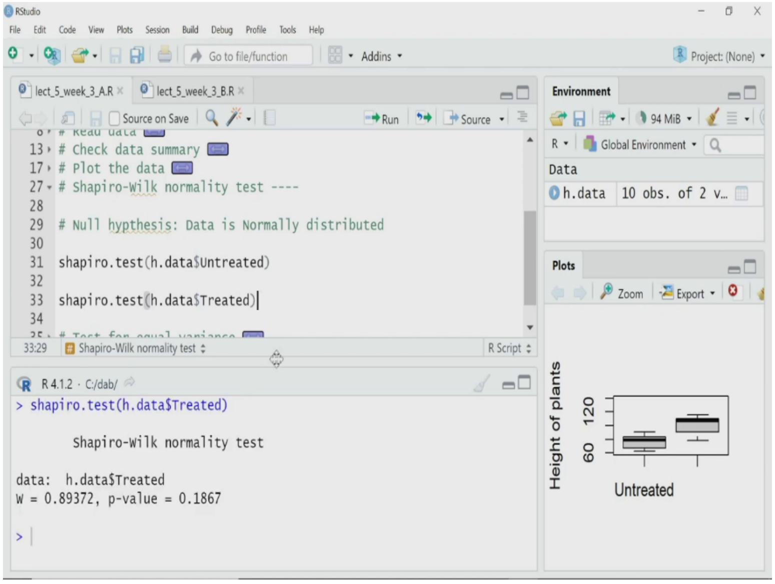
In Shapiro test normality test the null hypothesis is that the data is normally distributed. So, let us perform the normality test Shapiro test on both of my variables. So, what I will do, I will take one variable at a time and I will perform that, that is why for the first Shapiro test,

```
shapiro.test(h.data$Untreated)
```

So, I am using that dollar sign then untreated I am writing. So, it will perform the test only on that untreated variable, the first column of my data, and the second line here is doing the same test on the treated sample. So, let me execute 1 by 1. So, the result is here in the console. It is saying that the p value, focus on the p value is 0.3726. It is quite high, is not it?

So, that means the p value is very high. That means I cannot reject the null hypothesis. As p is very high, p value is very high. I cannot reject the null hypothesis. That means that data is normally distributed because that is my null hypothesis. I am happy my untreated sample has normal distribution.

(Refer Slide Time: 09:59)



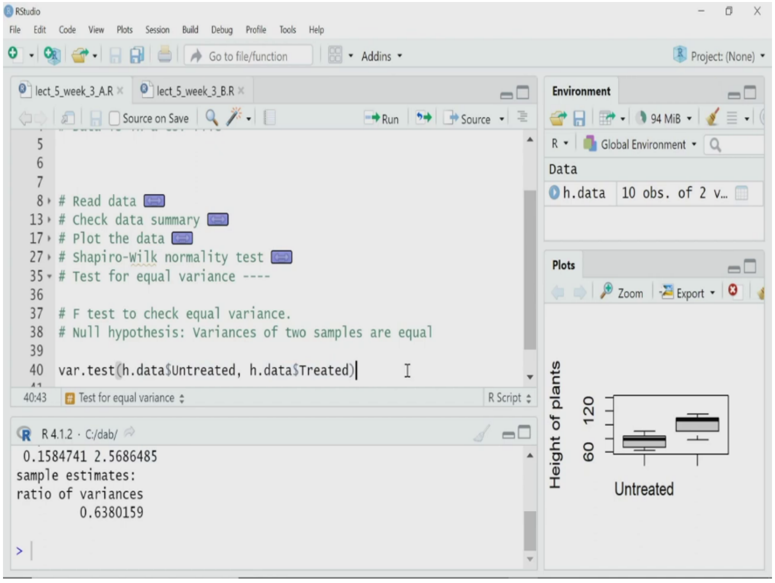
```
# Read data
# Check data summary
# Plot the data
# Shapiro-Wilk normality test ----

# Null hypthesis: Data is Normally distributed

shapiro.test(h.data$Untreated)

shapiro.test(h.data$Treated)

# Test for equal variance
```

```
> shapiro.test(h.data$Treated)

        Shapiro-Wilk normality test

data:  h.data$Treated
W = 0.89372, p-value = 0.1867

>
```

Data
h.data 10 obs. of 2 v...

Let me execute that for the next one the other data other variables the treated variable. Here again, I get the p value is 0.1867. So, again it is high. So, I can easily conclude that I have to accept the null hypothesis, I cannot reject it. So, if I accept the null hypothesis, I have to accept that the data is normally distributed. So, I am happy, both of my variables are normally distributed. Now, I will check whether they have equal variance or not.

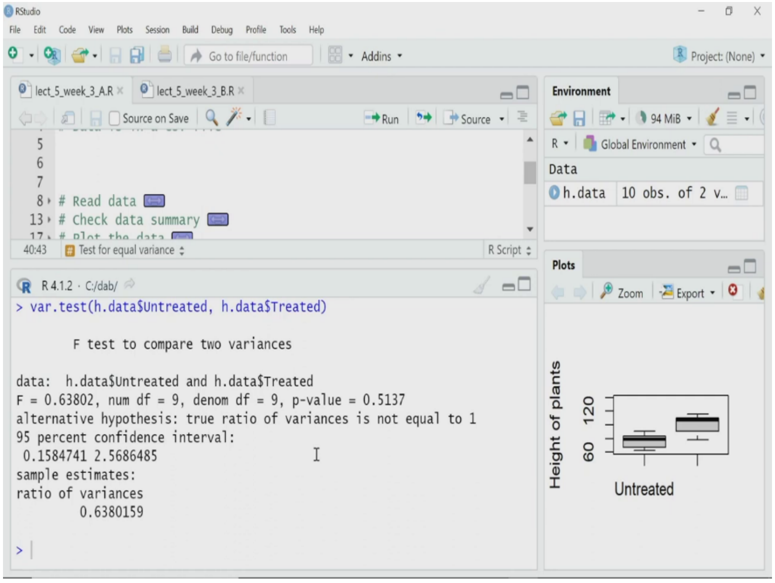
(Refer Slide Time: 10:29)



```
# Read data
# Check data summary
# Plot the data
# Shapiro-Wilk normality test
# Test for equal variance ----

# F test to check equal variance.
# Null hypothesis: Variances of two samples are equal

var.test(h.data$Untreated, h.data$Treated)
```

```
 0.1584741 2.5686485
sample estimates:
ratio of variances
         0.6380159

>
```

Data
h.data 10 obs. of 2 v...



```
# Read data
# Check data summary
# Plot the data
```

```
> var.test(h.data$Untreated, h.data$Treated)

	F test to compare two variances

data:  h.data$Untreated and h.data$Treated
F = 0.63802, num df = 9, denom df = 9, p-value = 0.5137
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1584741 2.5686485
sample estimates:
ratio of variances
         0.6380159

>
```

To do that, I will use again a inbuilt function. So, what I will do, I will use var dot test function, var dot test function. And what it will be doing is that it will perform an F test to check whether my two variable untreated and treated variables they have equal variance or not. And again, as it statistical test, I should have a null hypothesis and an alternate hypothesis, the null hypothesis by default in this case is that variances of two samples are equal.

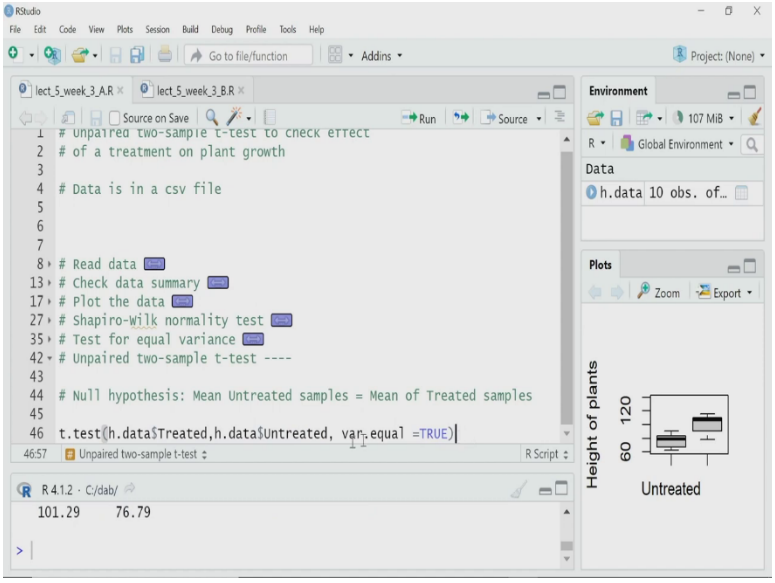
var.test(h.data$Untreated, h.data$Treated)

I am considering that the variances of both the samples are equal and I am using var.test function to perform it, it will be taking these two variable untreated and treated as my argument. So, here I perform that. Let me enlarge this console. It is it has given lots of information, but the most important information for me is here is the p value and it is saying 0.5137. So, this p value is very high that means, I cannot reject my null hypothesis.

Now, let me look back what is my null hypothesis, my null hypothesis is the variances of two samples are equal and my F test is saying I cannot reject it that means, my variances of these two variables are equal. So, I can actually proceed and perform t-test. So, the test result also say the same thing here in a different way it tells that the alternate hypothesis is that ratio, true ratio of various variances is not equal to 1.

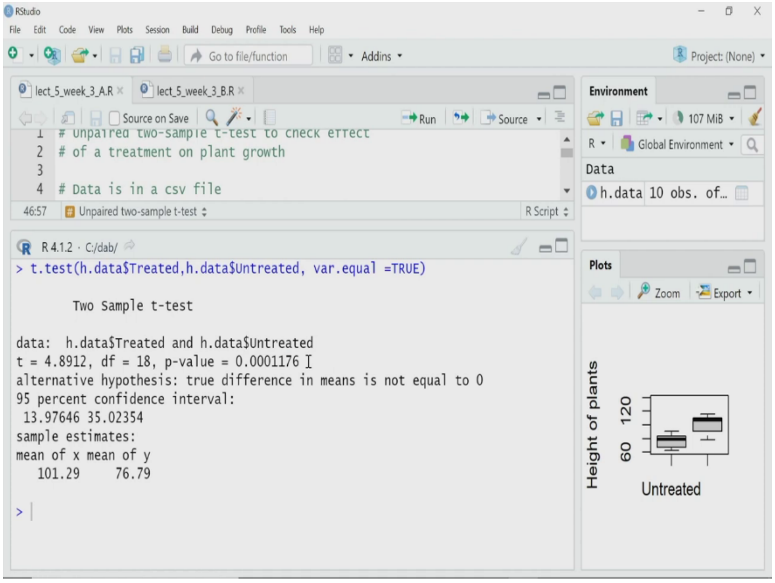
Remember, if the variances are equal, which is my null hypothesis, then that ratio should be 1. So, alternate hypothesis is that the ratio is not equal to 1. And in this case, we are rejecting this alternate hypothesis and we are accepting the null hypothesis because the p value is very high. So, now, I have completed my equal variance test, I have completed my normality test, the data has passed both these tests. So, I can now go for the simple unpaired two sample t-test and R has a function to do that.

(Refer Slide Time: 12:47)



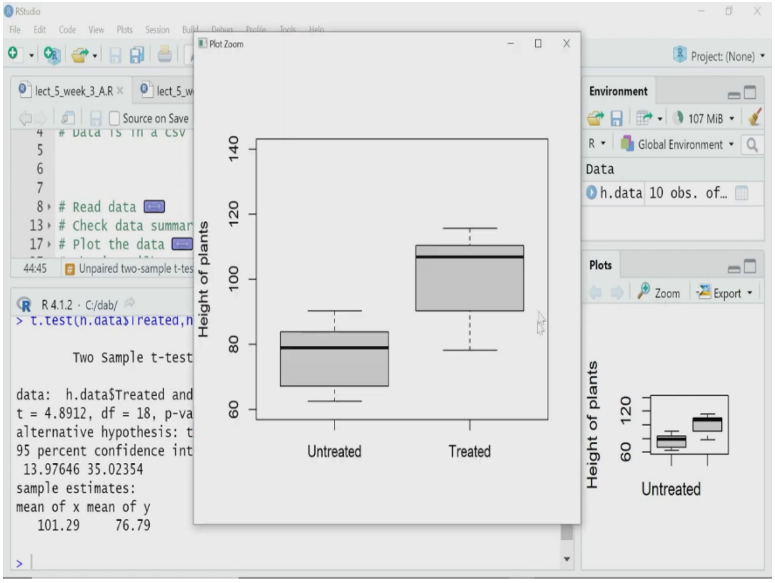
```
# Unpaired two-sample t-test to check effect
# of a treatment on plant growth

# Data is in a csv file



# Read data
# Check data summary
# Plot the data
# Shapiro-Wilk normality test
# Test for equal variance
# Unpaired two-sample t-test ----

# Null hypothesis: Mean Untreated samples = Mean of Treated samples

t.test(h.data$Treated,h.data$Untreated, var.equal =TRUE)
```

```
  101.29      76.79

>
```



```
# Unpaired two-sample t-test to check effect
# of a treatment on plant growth

# Data is in a csv file
```

```
> t.test(h.data$Treated,h.data$Untreated, var.equal =TRUE)

        Two Sample t-test

data:  h.data$Treated and h.data$Untreated
t = 4.8912, df = 18, p-value = 0.0001176
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 13.97646 35.02354
sample estimates:
mean of x mean of y
   101.29     76.79

>
```



```
# Data is in a csv



# Read data
# Check data summar
# Plot the data
```

```
> t.test(h.data$Treated,h

        Two Sample t-test

data:  h.data$Treated and
t = 4.8912, df = 18, p-va
alternative hypothesis: t
95 percent confidence int
 13.97646 35.02354
sample estimates:
mean of x mean of y
   101.29     76.79

>
```

```
Data
h.data  10 obs. of…
```

So, here I will perform the t-test using the inbuilt function called t.test, t.test. So, the null hypothesis here is that the mean of the untreated sample is equal to mean of the treated sample. My belief is they are different. So, that is my alternate hypothesis and I want to reject, I wish I can reject the null hypothesis after this t-test. So, this is my null hypothesis, that both variable treated and untreated has the same mean and I am performing t-test using t.test function.

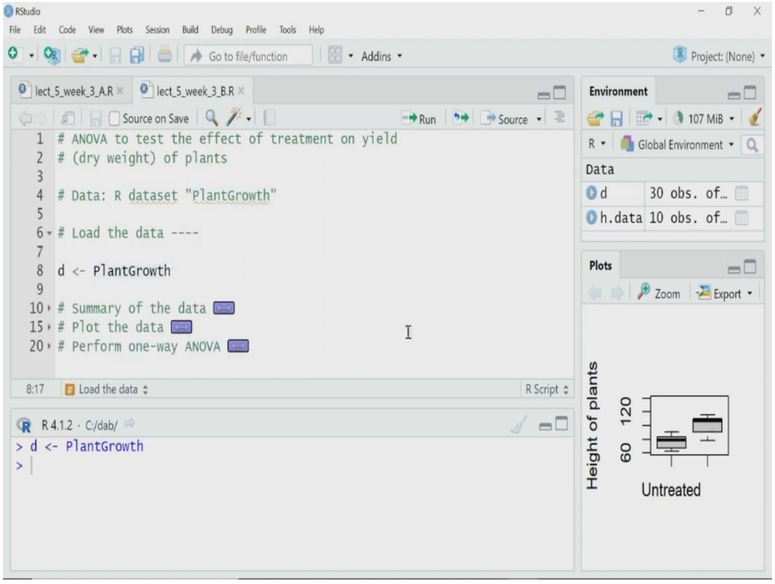
```
t.test(h.data$Treated, h.data$Untreated, var.equal = TRUE)
```

What are the argument? One argument at the end I have written is the variance are equal, variances are equal. So, that is written as var dot equal to TRUE. There are other arguments also I have not used those, but the first and the most important argument is obviously, what are the variables. The first variable is the treated sample, the data of that and the second data is the untreated one. So, I have to perform the t-test on these two data and it has to, it will consider that the variances are equal.

So, let me perform the t-test. Here is the result, the result has lots of thing let us focus first on the p value. The p value is very low 0.0001176. So, even if I consider 0.001 as my level of significance, then also I can easily reject my null hypothesis. So, what is my null hypothesis? My null hypothesis if I go that the mean of treated sample and untreated samples are equal. So, as the p value is very small, I can reject the null hypothesis that means, I will accept the alternate hypothesis where it says that the means are different.

So, that is what it is written here, alternative hypothesis the true difference in mean is not equal to 0. So, we accept that. So, if I again open this figure, I can say that this treated data has the mean which is significantly different from the untreated one, that means treatment has a statistically significant effect on the height of the plant. That is all for the t-test let me fast move into ANOVA, analysis of variance.

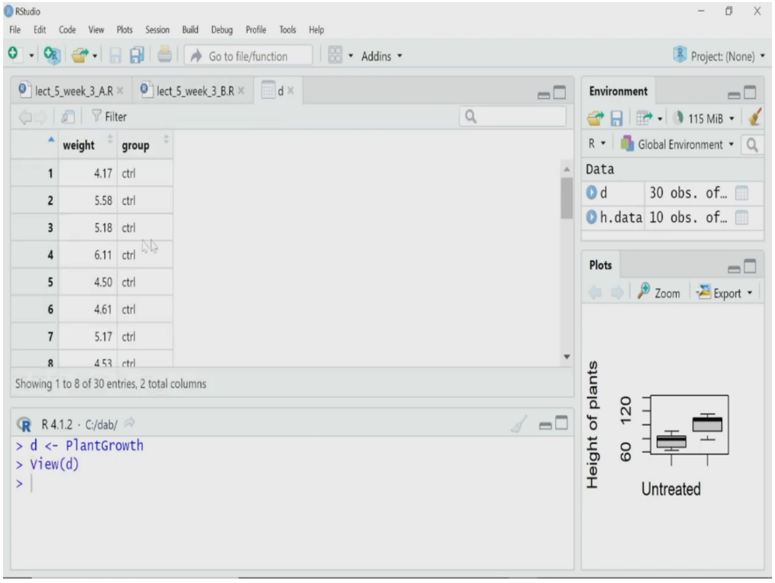
(Refer Slide Time: 15:19)



```
# ANOVA to test the effect of treatment on yield
# (dry weight) of plants

# Data: R dataset "PlantGrowth"

# Load the data ----

d <- PlantGrowth

# Summary of the data
# Plot the data
# Perform one-way ANOVA
```

```
> d <- PlantGrowth
>
```

Data
d 30 obs. of…
h.data 10 obs. of…



	weight	group
1	4.17	ctrl
2	5.58	ctrl
3	5.18	ctrl
4	6.11	ctrl
5	4.50	ctrl
6	4.61	ctrl
7	5.17	ctrl
8	4.53	ctrl

Showing 1 to 8 of 30 entries, 2 total columns

```
> d <- PlantGrowth
> View(d)
>
```

Data
d 30 obs. of…
h.data 10 obs. of…

Again, I have a plant data set. And what is happening, what people have done here in this data set is that suppose you have a control set of plant and then you have two other groups one is treatment 1 and treatment 2, consider those treatment are, maybe treatment same treatment, but with different concentration at different doses. And then you are checking the yield and you are measuring the yield in terms of a dry weight of the plant.

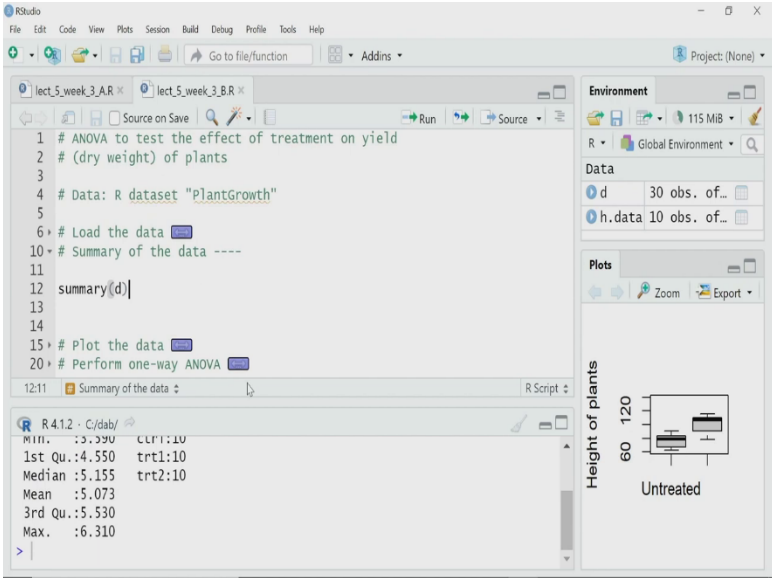
This data set is by default present in your R data set. So, I will be using that. And now remember, in this case, then there are three variables control, three cases three types of samples. So, control, then you have treatment condition 1, treatment condition 2. So, I cannot use t-test, I have to do ANOVA, analysis of variance and I will perform one way variance because there is only one way the thing has been changed, we have changed the treatment condition one type of treatment condition.

So, I will perform one way ANOVA. So, the as I said the data set is a plant growth data set present by default in your R built, you must have in your computer by now if you have installed R. So, let me load that data. So, what I will do I will load these data and assign that data to a variable called d. I have loaded it. So, let us check the data how it is. It is two column data the first one is weight and second one is group.

d ← PlantGrowth

And in the group, we have written control, control, Ctrl, control, control, then treatment condition 1, then treatment 2. So, all these groups are called factors or labels are in one column. Those are labeled there and their corresponding weights are also given the first column. So, now I have to perform ANOVA on this.

(Refer Slide Time: 17:06)

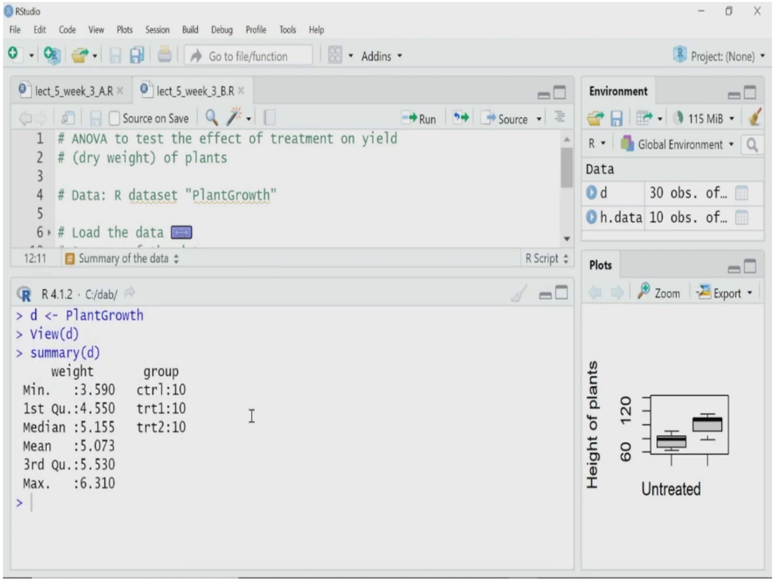


```
# ANOVA to test the effect of treatment on yield
# (dry weight) of plants

# Data: R dataset "PlantGrowth"

# Load the data
# Summary of the data ----

summary(d)


# Plot the data
# Perform one-way ANOVA
```

```
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
>
```

Data

d	30 obs. of…
h.data	10 obs. of…

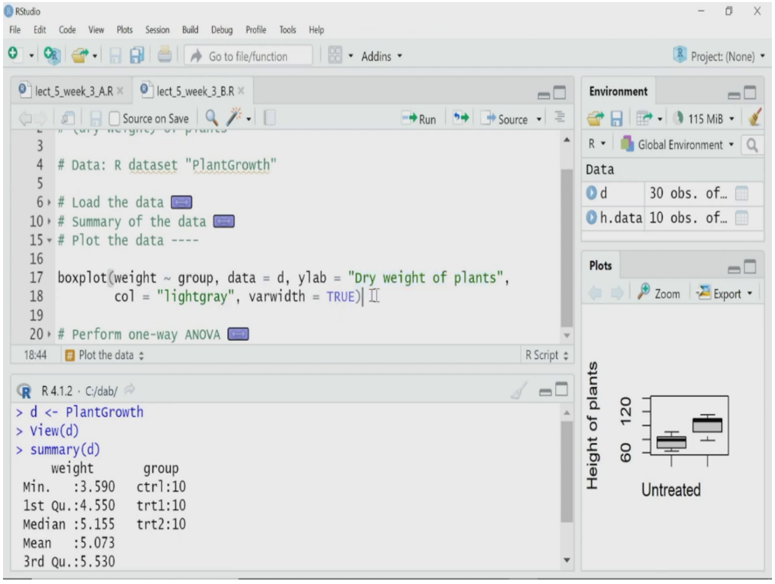


```
# ANOVA to test the effect of treatment on yield
# (dry weight) of plants

# Data: R dataset "PlantGrowth"

# Load the data
```

```
> d <- PlantGrowth
> View(d)
> summary(d)
     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            
> 
```

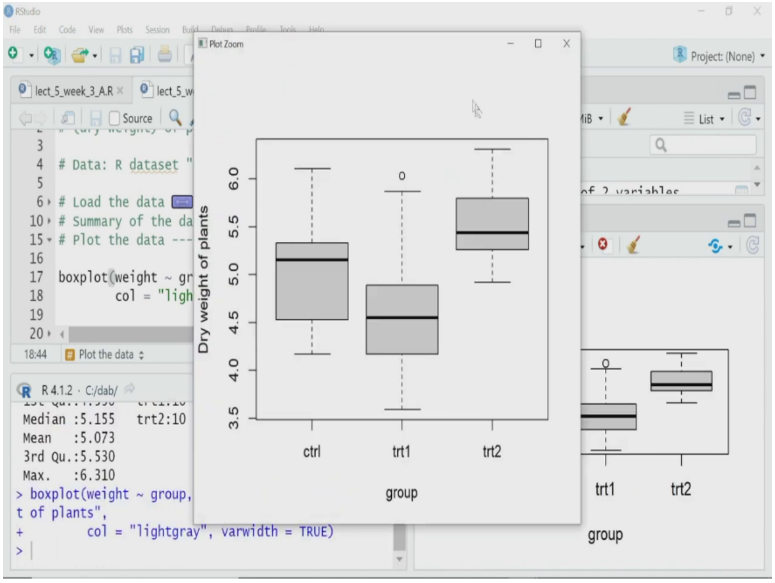
To do so, before I move let us look into the summary of that data. So, I will use that again the summary function, summary(d). Now, I have two columns. So, the first column is the weight, the minimum weight is 3.59, maximum is 6.3, mean, median other things are given. And the group variable we have three types of factors or three types of sample control, there are 10 numbers of them. Treatment 1, 10 numbers of them. Treatment 2, 10 numbers of them. Before I move into ANOVA just like in t-tests, it will be good if I visualize it.

(Refer Slide Time: 17:48)



```
# Data: R dataset "PlantGrowth"

# Load the data
# Summary of the data
# Plot the data ----

boxplot(weight ~ group, data = d, ylab = "Dry weight of plants",
        col = "lightgray", varwidth = TRUE)

# Perform one-way ANOVA
```

```
> d <- PlantGrowth
> View(d)
> summary(d)
     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            
```



```
# Data: R dataset "

# Load the data
# Summary of the da
# Plot the data ---

boxplot(weight ~ gr
        col = "ligh
```

```
Median :5.155   trt2:10
Mean   :5.073
3rd Qu.:5.530
Max.   :6.310
> boxplot(weight ~ group,
t of plants",
+         col = "lightgray", varwidth = TRUE)
> 
```

And again, I will use that box plot. So, now, I am plotting using the same box plot function.

boxplot(weight ~ group, data = d, ylab = “Dry Weight of plants”, col = “lightgray”, varwidth = TRUE)

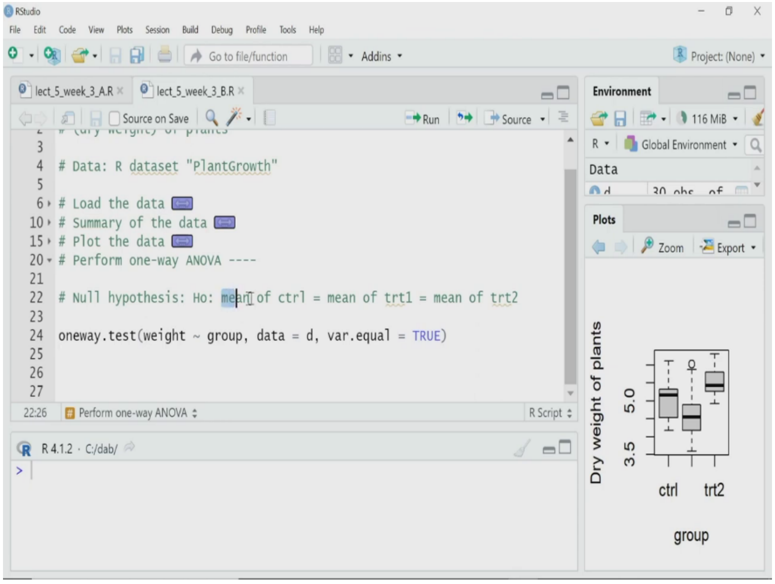
As I said we will discuss how to do box plot and bar plot separately in another video. Here, what we are saying is that here I have written as, the first argument I have written as weight and then I have given this tilde and written as group. So, it means the R understand that group weight is the response variable or dependent variable and group is the independent variable.

So, if I have weight as dependent variable a response so it should be in my vertical axis, whereas a group is the independent variable, so it should be in the horizontal axis. So, that is how it will understand. The data is that the d variable, y label means vertical axis label, I want to write dry weight of plants, and I want to keep the width of each of these box as equal so that is why, so not equal, I can want to change it depending on the size of the sample.

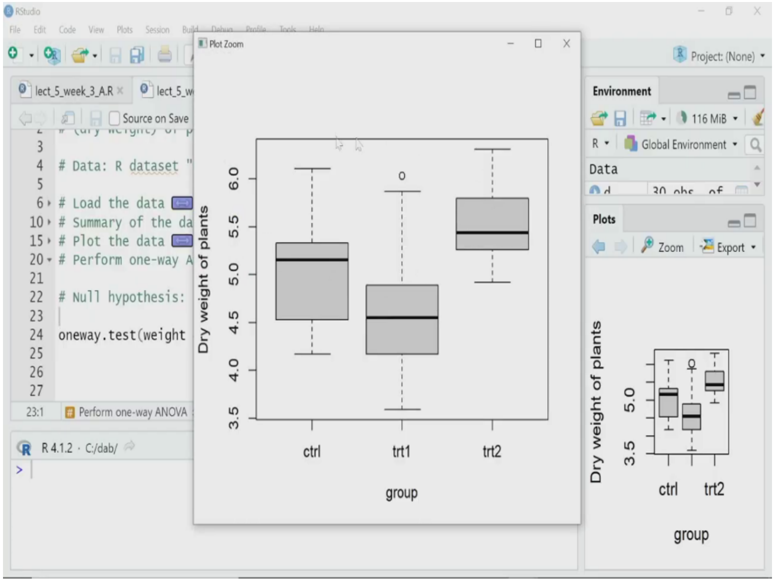
So, I have kept width equal to TRUE, variable width equal to TRUE and the color I want light gray. So, let me draw this, let us zoom, here is that data. So, you can see very nicely this box plot actually represent the mean and behavior as well as the dispersion of the data in each of the categories. We have quite a bit dispersion here.

If the median values of treatment 1 seems lower than the control 1 whereas in treatment 2 the median value is higher, but each of these groups has quite a bit dispersion. So, now the question comes to my mind, is there any effect of this treatment. I have two treatment groups maybe with different doses, but is there any effect of these treatment on the dry weight or the yield of this plant. So, when I mean any effect, I mean, whether any statistically significant effect or not, and to get that I will perform one way ANOVA.

(Refer Slide Time: 20:03)



```
# (dry weight) of plants

# Data: R dataset "PlantGrowth"

# Load the data
# Summary of the data
# Plot the data
# Perform one-way ANOVA ----

# Null hypothesis: Ho: mean of ctrl = mean of trt1 = mean of trt2

oneway.test(weight ~ group, data = d, var.equal = TRUE)
```



How can I do that? Performing one way ANOVA is very easy and you have an inbuilt function to do that the function is called oneway.test.

oneway.test(weight ~ group, data = d, var.equal = TRUE)

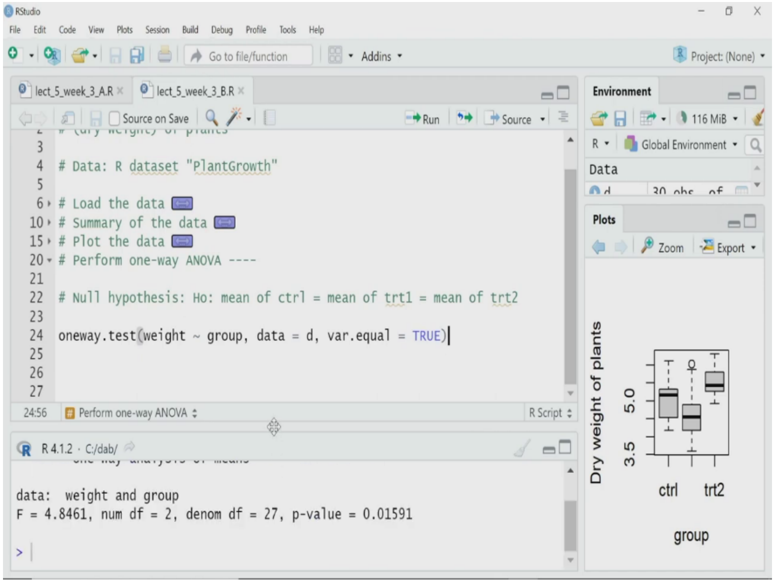
Now, it is a statistical test. So, obviously, there must be null hypothesis and alternate hypothesis. For a null hypothesis, in ANOVA the null hypothesis will be obviously, that all means of all treatment groups should be equal.

That is what is written here is the null hypothesis is mean of control is equal to mean of treatment 1 equal to mean of treatment 2. If I accept the null hypothesis after this test that means, I have to consider that these differences that I see in the plot these little differences that I see are not statistically significant. But if I reject then I have to accept that the difference that I have seen are statistically significant, that means, there is a considerable statistically considerable effect of the treatment on the yield of the plant.

So, what is the function I am using as I said it is one dot test and the arguments are, the arguments are written like this weight tilde group. Again I am telling the function is that the weight is the dependent variable or the response variable. And whereas, the group is the predictor or independent variable. So, I want to see how the weight of each group is varying, and I want to perform the ANOVA for that.

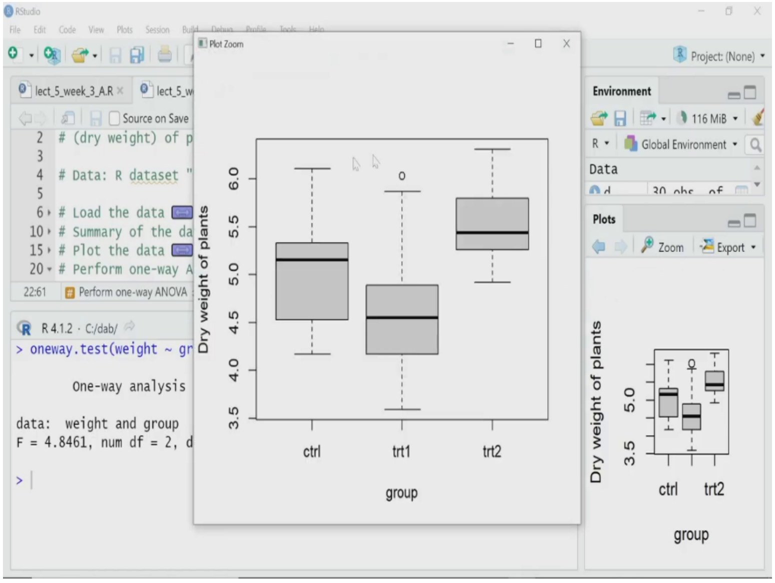
And the data is equal to d and the variances are equal. So, that is why I have considered written is a TRUE, var.equal = TRUE. And specifying that yes, the variances of these three groups of data that is controlled, treated 1, treated 2 they are equal. I have not shown the variance check test here I have performed before I picked up this data, I have also performed the same normality test, data is also normally distributed. Whenever you do these type of parametric tests like ANOVA or t-test, you should perform those the way I have shown what t-test case.

(Refer Slide Time: 22:13)



```
# (dry weight) of plants

# Data: R dataset "PlantGrowth"

# Load the data
# Summary of the data
# Plot the data
# Perform one-way ANOVA ----

# Null hypothesis: Ho: mean of ctrl = mean of trt1 = mean of trt2

oneway.test(weight ~ group, data = d, var.equal = TRUE)
```

```
One-way analysis of means

data:  weight and group
F = 4.8461, num df = 2, denom df = 27, p-value = 0.01591

>
```



```
# (dry weight) of p

# Data: R dataset "

# Load the data
# Summary of the da
# Plot the data
# Perform one-way A
```

```
> oneway.test(weight ~ gr

        One-way analysis

data:  weight and group
F = 4.8461, num df = 2, d

>
```

So, let us perform the ANOVA because, I have already checked they are normally distributed and their variance is equal. So, if I perform the test here is my output in the console, is one single line statement which is very clear, I have to look into, you can look into F value. Obviously, if you want to look into the F table, it is the F test, but I will go directly to the p value. The p value is 0.0159. So, roughly 0.016. So, if I consider the cutoff value as 0.05.

So, then I can say this p value is lesser than that cutoff, 0.5. So, in that case, I have to reject the null hypothesis. So, what is my null hypothesis? My null hypothesis is that the mean of control, mean of treatment 1, mean of treatment 2, all the means of these three treatment conditions are equal. And as the p value for my ANOVA test has come less than my cutoff 0.05. So, I am rejecting that null hypothesis.

That means, that the difference that you see in the mean value of this three treatment group are statistically significant. That is how we perform an ANOVA on this type of data. That is all for this video. In this video, we have learned t-test and ANOVA, t-test and ANOVA will have different versions, different varieties. And R you can actually perform all of them they are different functions, each function have lots of arguments.

And there are, apart from t-test and ANOVA, there are other type of tests also which are available R and used by biologists. This course is not a biostatistics course, that is why we are not going in detail of all of these statistical tests. But I hope with this simple demonstration, you will be able to move forward and learn other statistical tests whenever you need to do those, perform those. That is all for this video. Thank you for learning with me today. See you in the next video.