

Simple two sample t test

Learn how to run a simple t test using 2 treatments Variety A and B with 5 plots in each treatment. Please first go to the screen cast on design and layout. To decrease the incidence of error you might like to copy the code directly from this file into your RStudio. Note the data files are `planyield.csv`

Start a new RStudio Project naming it 'Online' and save the data "planyield.csv" into the same fold as the Project file. This will enable you to just load the csv file straight into your RStudio Project from the directory. Type in "`plany <- read.csv("planyield.csv")`" into the "Source on Save" Pane. Highlight the typing and the csv will be added to the project (Figure 1). The "planyield" data has been renamed 'plany' for ease of typing. Notice the data worksheet description '10 obs. Of 3 variables' appears in the Global Environment pane under Data. The syntax or code 'View(plany)' displays the data in the Source pane with the same data description in the top right hand corner (Figure 2).

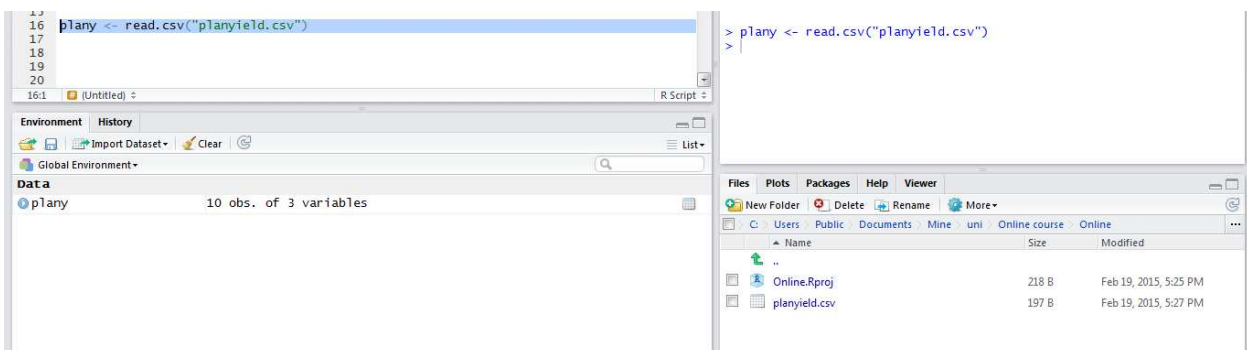


Figure 1 Loading the data into RStudio

plot	variety	Yield
1	VarietyB	194.0
2	VarietyB	205.5
3	VarietyA	190.7
4	VarietyA	233.5
5	VarietyB	199.2
6	VarietyA	203.5
7	VarietyA	216.5
8	VarietyA	222.5
9	VarietyB	172.4
10	VarietyB	184.0

Figure 2 The data loaded into RStudio

Exploration of loaded data

By typing in 'head' RStudio returns (in the Console pane) the first 6 rows of data (Figure 3). Notice the description of 'head' after the hash symbol.

```
> head(planY)## read the first 6 rows of data into the Console Pane
  plot variety Yield
1     1 VarietyB 194.0
2     2 VarietyB 205.5
3     3 VarietyA 190.7
4     4 VarietyA 233.5
5     5 VarietyB 199.2
6     6 VarietyA 203.5
```

Figure 3 The Console Pane after running the 'head' code

Here are some more useful codes to explore you data. To display data use 'names' which returns the names of the columns, 'length' give the length of the columns and 'str' gives the structure of the data in RStudio (Figure 4).

```
> names(planY)## show names of columns
[1] "plot" "variety" "Yield"
> length(planY$Yield)## how many rows to the data
[1] 10
> str(planY)##show structure of data
'data.frame': 10 obs. of 3 variables:
 $ plot : int 1 2 3 4 5 6 7 8 9 10
 $ variety: Factor w/ 2 levels "VarietyA","VarietyB": 2 2 1 1 2 1 1 1 2 2
```

Figure 4 Some helpful code, output and description

Now we will explore some plotting code. To create a histogram the code is simply 'hist' (hist(planY\$Yield) ## show histogram of Yield) (Figure 5) while 'plot' (plot(planY\$Yield) ## show individual values of Yield) gives the individual values for Yield (Figure 6) and 'boxplot' gives a boxplot of Yield (boxplot(planY\$Yield) ##boxplot of yield) (Figure 7). The boxplot has not helped display the differences between the 2 treatments as RStudio has treated the treatment (variety) column as a 'Factor' as it was a text variable. To show the differences between the 2 treatments we have to specify that we want to use 'Variety' as variable by specifying the we want plot Yield broken up into Variety (boxplot(planY\$Yield~planY\$variety) (Figure 8).

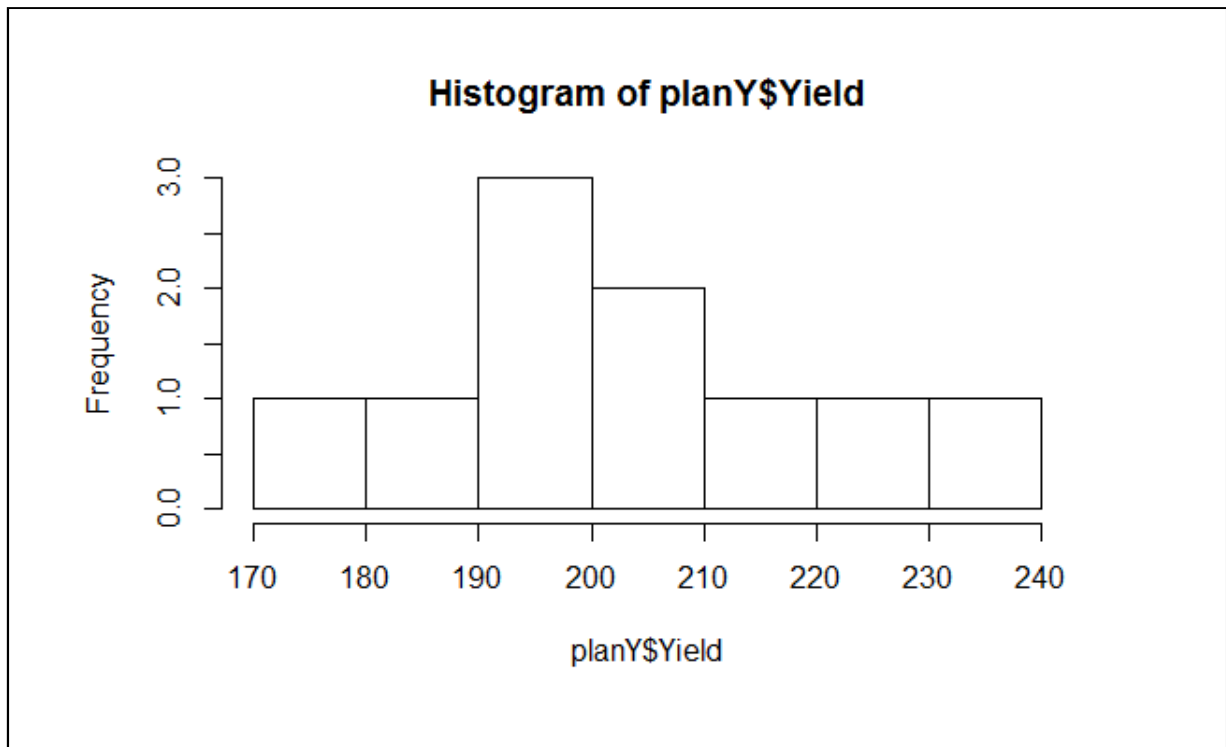


Figure 5 Histogram of Yields

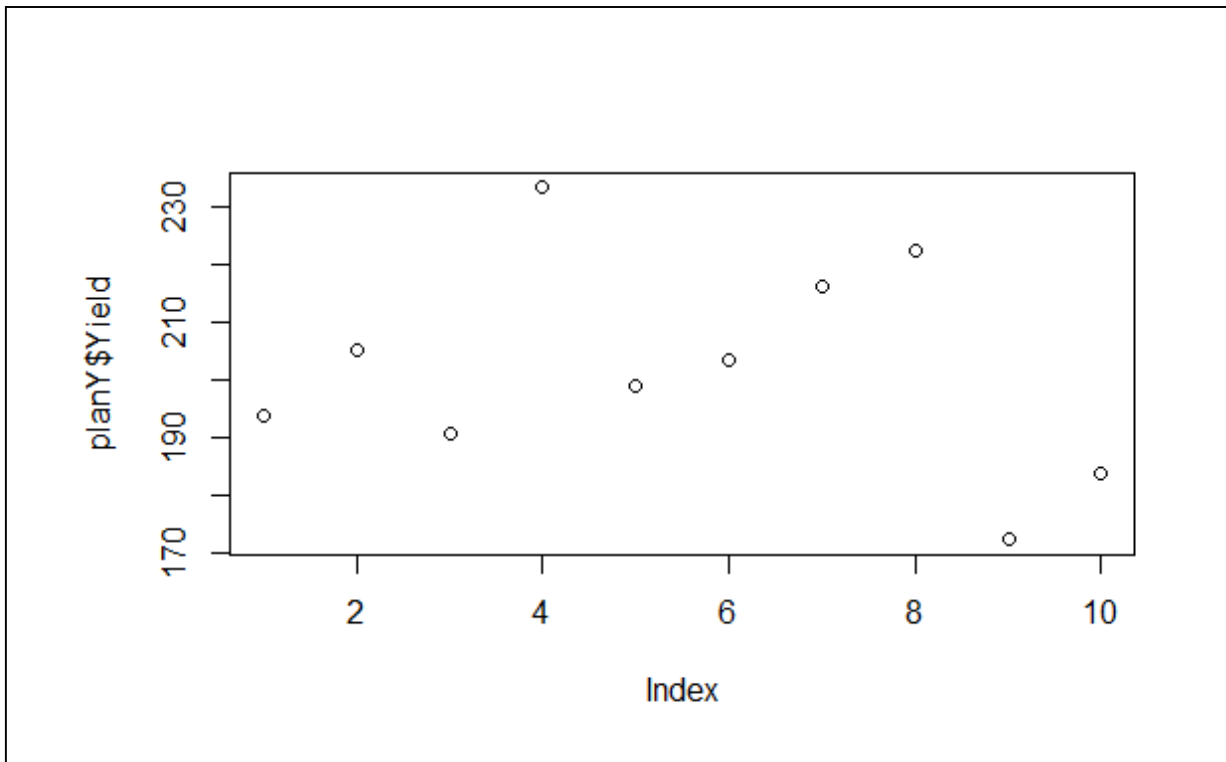


Figure 6 Individual value plot of Yield

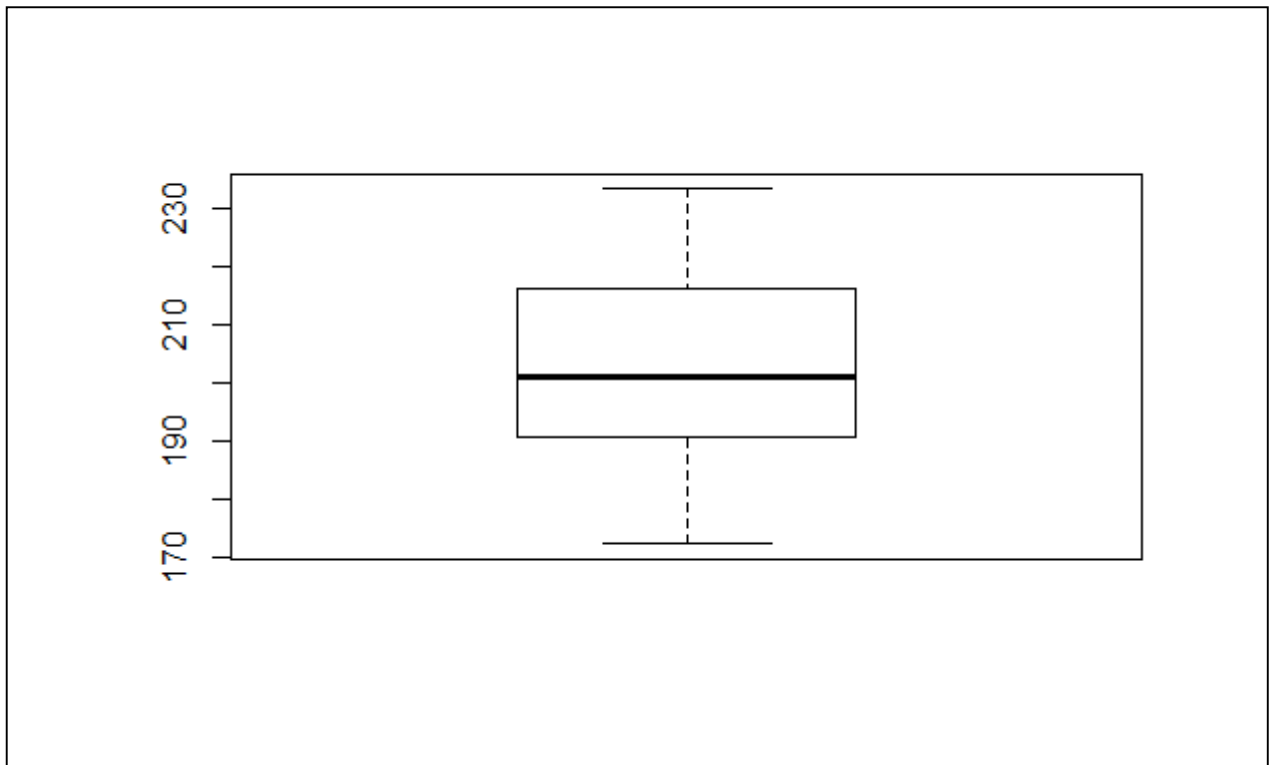


Figure 7 Boxplot of Yield

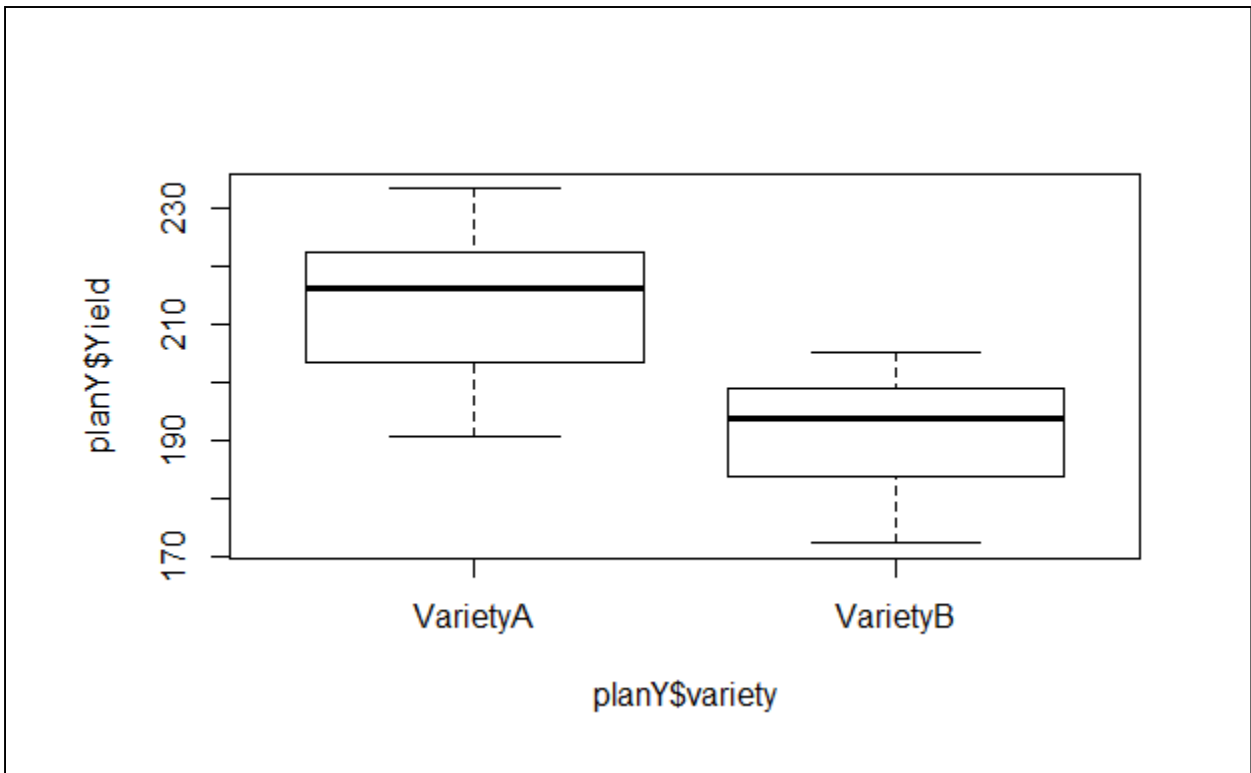


Figure 8 Boxplot of the Yields of both treatments

A simple two sample t test analysis

Now that we can see from the boxplots that there appears to be a difference between the 2 treatments but we would like to check if the difference is significant so we will carry out a simple 2 sample t test using `(t.test(planY$Yield~planY$variety ## testing the difference in treatments)` (Figure 9).

```
> t.test(planY$Yield~planY$variety)## testing the difference in treatments
      welch Two Sample t-test

data:  planY$Yield by planY$variety
t = 2.3585, df = 7.567, p-value = 0.04781
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.278033 44.361967
sample estimates:
mean in group VarietyA mean in group VarietyB
          213.34          191.02
```

Figure 9 The output for the 2 sample t test of variety

What are your conclusions from the test?

Importing a data file.

Left click onto **Import Dataset** select 'from Text File...' (Figure 10) then go to where the data file has been saved and following prompts select the file to import (Figure 11). The 'Import Dataset' will come up (Figure 12). Check that the data looks correct and select 'Yes' for 'Headings' if it has not been selected. The data will appear in the Pane of the 'Source on Save' (Figure 13) note that the name of the data file on the far right of the Pane. Also notice under the name it show '10 observations of 3 variables'.

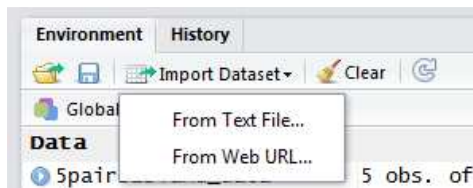


Figure 10

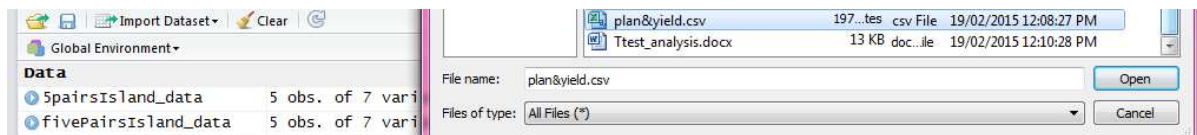


Figure 11 Selecting the csv file to import

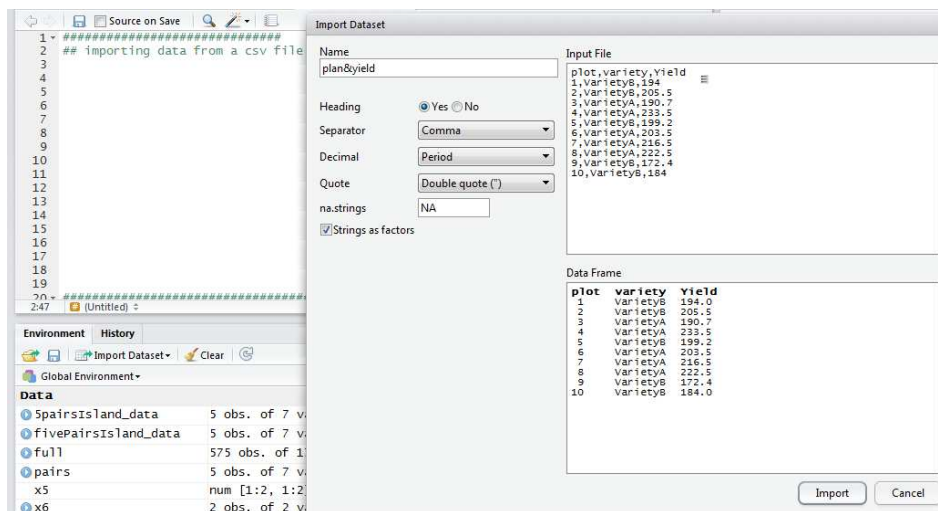


Figure 12 The 'Import Dataset' window

plot	variety	Yield
1	VarietyB	194.0
2	VarietyB	205.5
3	VarietyA	190.7
4	VarietyA	233.5
5	VarietyB	199.2
6	VarietyA	203.5
7	VarietyA	216.5
8	VarietyA	222.5
9	VarietyB	172.4
10	VarietyB	184.0

Figure 13 The data set imported

I was going to try and screencast the Powerpoint.

We can also use your t test PowerPoints

And I would like to also have a simple demo with a t tests, (it a two sample independent t test)

What do you think- you may be able to adapt from this, and add any interpreting questions to assist them.

#####

##

Simple two sample t test

##

Example 2 treatments Variety A and B and 10 plots

in a completely randomised design (CRD)

#####

```
## The steps in this code:
```

```
# set working directory (RStudio project folder)
```

```
# read in data set up from the .csv file
```

```
# with the plan with the yields
```

```
# look at the data
```

```
# remind yourself of the variable names
```

```
# and the number or rows in the variable "Yield"
```

```
#####
```

```
## importing "planyiel.csv" data file using 'Import Dataset'- global Environment Pane
```

```
plany <- read.csv("planyield.csv")##data file needs to be in same folder as Rstudio
```

```
#####
```

```
##
```

```
## Simple two sample t test
```

```
##
```

```
## Example 2 treatments Variety A and B and 10 plots
```

```
## in a completely randomised design (CRD)
```



```
#####
```

```
## The steps in this code:
```

```
# set working directory
```

```
# read in data set up from the .csv file
```

```
# with the plan with the yields
```

```
# look at the data
```

```
# remind yourself of the variable names
```

```
# and the number or rows in the variable "Yield"
```

```
#####
```

```
planY <- read.csv("planyield.csv")
```

```
View(planY)##show the data in table form in another window
```

```
head(planY)## read the data into the Console
```

```
names(planY)## show names of columns
```

```
length(planY$Yield)## how many rows to the data
```

```
hist(planY$Yield)## shows histogram of Yield
```

```
plot(planY$Yield)## shows individual values of Yield
```

```
## look at the structure of the data
```

```
## Is your variety a factor ?
```

```
## In this case it has come in as a factor
```

```
## as it was a string or text variable
```

```
str(planY)
```

```
# Plot all the data
```

```
boxplot(planY$Yield)
```

```
# Plot by variety
```

```
boxplot(planY$Yield~planY$variety)
```

```
## Undertake a t test
```

```
## where y is numeric and x is a binary factor
```

```
## t.test(y~x)
```

```
## The data is not structured, so there are no blocks
```

```
t.test(planY$Yield~planY$variety)
```

```
## What do you conclude ?
```