1. **DEFINITION**

**ANOVA (ANALYSIS OF VARIANCE**) is a statistical test used to analyze the difference between the means of more than two groups.

**A TWO-WAY ANOVA** is used to estimate how the mean of a quantitative variable changes according to the levels of two categorical variables.It is used to determine how two independent variables, in combination, affect a dependent variable.

A **quantitative variables** represent measurable quantities The values which these variables can take can be ordered in a logical and natural way i.e. size of shoes, weight of a person, and prices of food products.

A **categorical variables** represent types of categories of variables. It is usually takes on values that are names or labels i.e. color of ball (red, green, blue), or breed of dog (i.e. shepherd, terrier, collie).

A **level** is an individual category within the categorical variable i.e. in a categorical variable fertilizer type, fertilizer types 1, 2 and 3 are levels.

**TWO-WAY ANOVA TEST HYPOTHESES**

* **Null Hypotheses**
* There is no difference in group means at any level of the first independent variable.
* There is no difference in group means at any level of the second independent variable.
* The effect of one independent variable does not depend on the effect of the other independent variable
* **Alternative Hypotheses**
* There is difference in group means at any level of the first independent variable.
* There is difference in group means at any level of the second independent variable.
* There is an interaction effect between planting density and fertilizer type on average yield.

1. **PARAMTERS / REQUIREMENTS**

**When to use two-way ANOVA?**

A two-way ANOVA can be used when the collected data on quantitative dependent variable is at multiple levels of two categorical independent variables.

**What are the assumptions of the two-way ANOVA?**

* **Homogeneity (homoscedasticity) of Variance.** The variation around the mean for each group being compared should be similar among all groups.
* **Independence of Observations.** Independent variables should not depend on the other independent variables.
* **Normally-Distributed Dependent Variable.** The values of the dependent variable should follow a bell curve or are normally distributed.

1. **HOW TO DO TWO WAY ANOVA MANUALLY**

**Example Problem:**

Suppose you are researching which type of fertilizer and planting density produces the greatest crop yield in a field experiment. You assign different plots in a field to a combination of fertilizer type (1, 2, or 3) and planting density (1 = low density and 2 = high density), and measure the final crop yield in bushels per acre at harvest time.

From here we can use the two-way ANOVA to find out if fertilizer type and planting density have an effect on average crop yield.

**Solutions:**

1. Identify the dependent and independent variables and their categories and levels:

|  |  |  |
| --- | --- | --- |
| **Dependent Variable** | **:** | Crop Yield |
| **Independent Variable 1** | **:** | Fertilizer |
| **Independent Variable 2** | **:** | Planting Density |
| **Categorical Variables** | **:** | Fertilizer Type |
| Planting Density Type |
| **Levels of Fertilizer** | **:** | Fertilizer Type 1, 2 and 3 |
| **Levels of Planting Density** | **:** | Planting Density Type 1 and 2 |

1. State the null and alternative hypotheses:

**Null Hypotheses**

* There is no difference in average yield for any fertilizer type.
* There is no difference in average yield for any planting density
* The effect of one independent variable on average yield does not depend on the effect of the other independent variable (no interaction effect).

**Alternative Hypotheses**

* There is difference in average yield by fertilizer type
* There is difference in average yield by planting density type
* There is an interaction effect between planting density and fertilizer type on average yield.

1. Using the datasets containing yield from combination of fertilizer and planting density, we calculated means of fertilizers, planting densities and their combinations as illustrated below.



1. Calculate the sum of squares of fertilizer, planting density and their combinations:

We first let,

1. Sum of Squares of Fertilizer
2. Sum of Squares of Planting Density
3. Sum of Squares of Within of Fertilizer and Planting Density:

1. Sum of Squares of Total:
2. Sum of Squares of Interaction
3. Degrees of Freedom Fertilizers and Planting Density:
4. Mean Square of Fertilizers and Planting Densities:
5. Summary Table



Note that F value for Fertilizer is calculated as its MS over MS of Within, F value for Density is calculated as its MS over MS of Within, and the F value for Interaction is calculated as the product of their respective MS over MS of Within

1. Decision

Reject the null hypotheses.

1. Interpretation:

Fertilizer type and planting density type has effect on the average yield of crop and there is an interaction effect between fertilizer type and density type on average yield of crop.

1. **HOW TO DO TWO WAY ANOVA IN R**

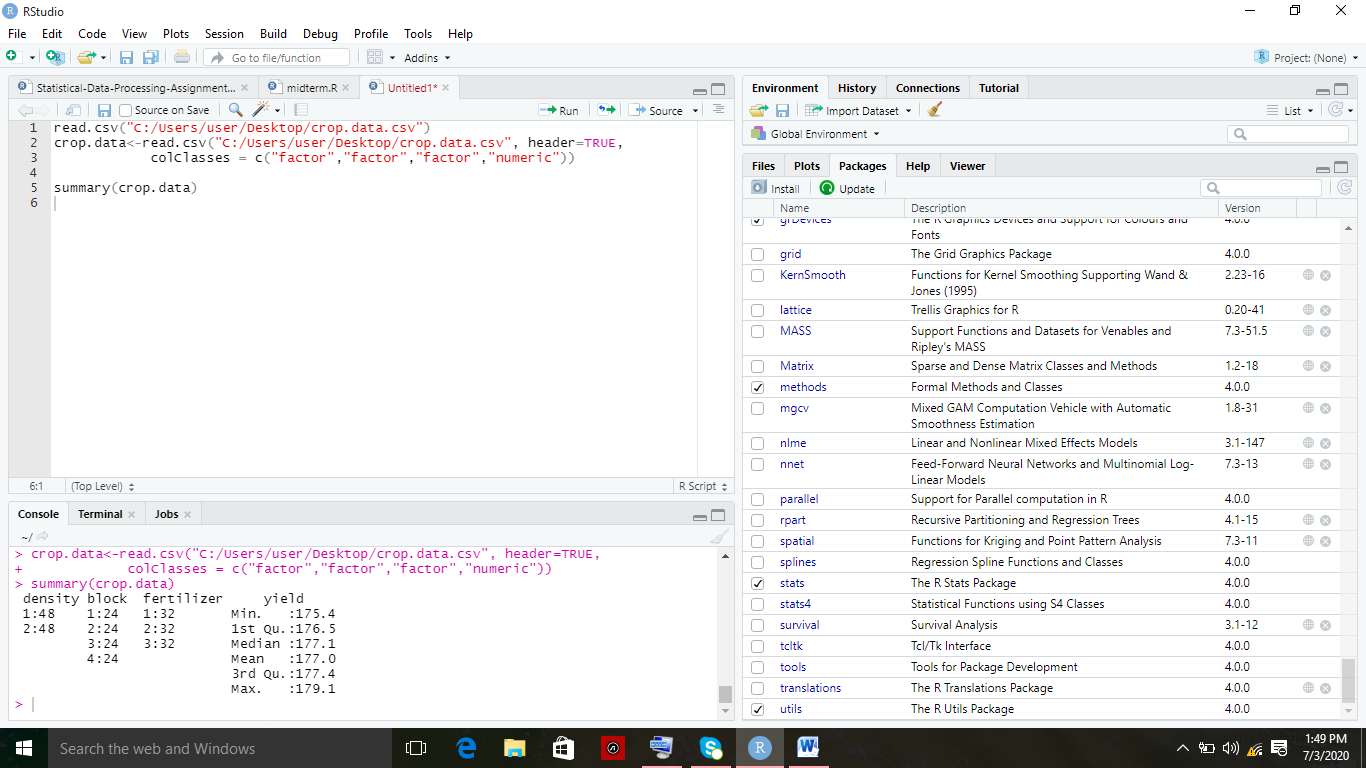
**TWO-WAY ANOVA TEST** is used to evaluate simultaneously the effect of two grouping variables (A and B) on a response variable. The grouping variables are also known as factors. The different categories (groups) of a factor are called levels. The number of levels can vary between factors. The level combinations of factors are called cell.

* When the sample sizes within cells are equal, we have the so-called **balanced design**. In this case the standard two-way ANOVA test can be applied.
* When the sample sizes within each level of the independent variables are not the same (case of **unbalanced designs**), the ANOVA test should be handled differently.

To read csv file, use the function csv () as follows:

read.csv("C:/Users/user/Desktop/crop.data.csv")

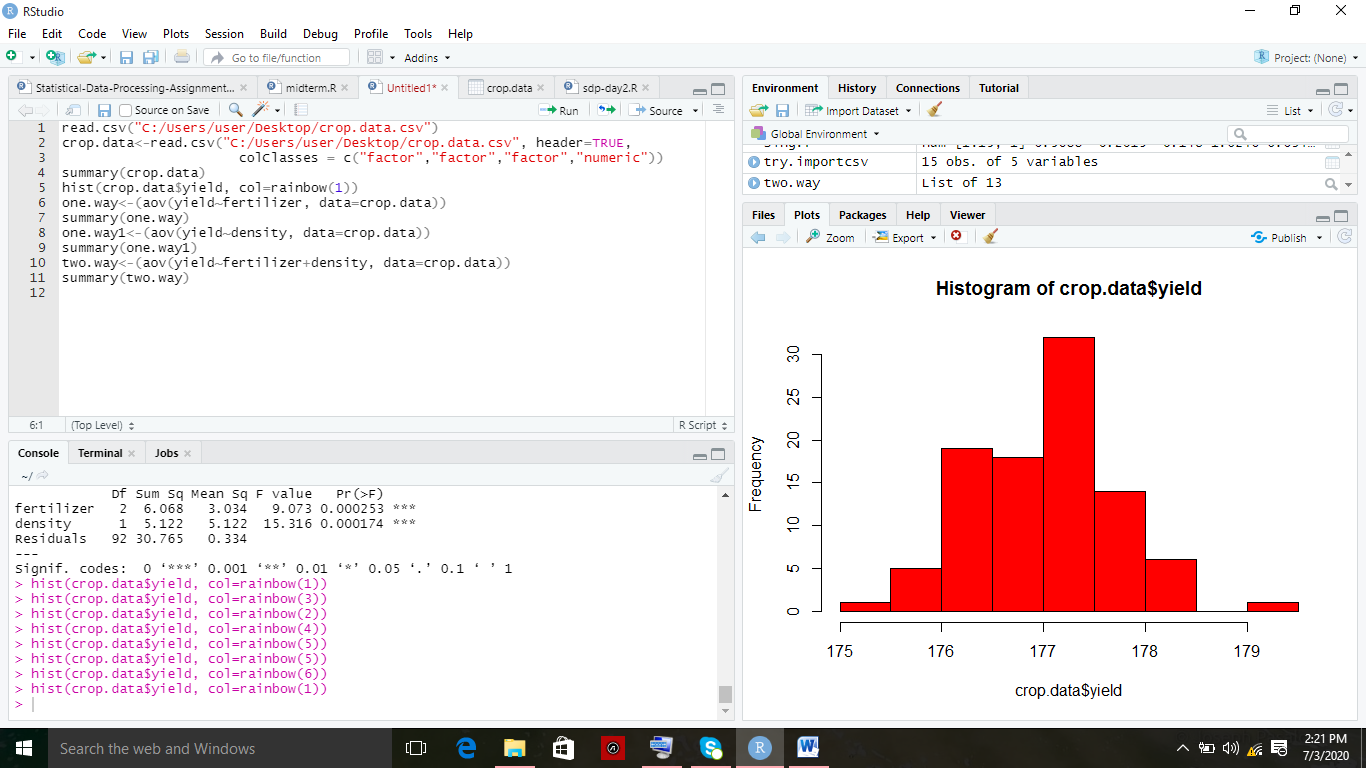
crop.data<-read.csv("C:/Users/user/Desktop/crop.data.csv", header=TRUE, colClasses = c("factor","factor","factor","numeric"))



t

to check normality, type

hist(crop.data$yield)

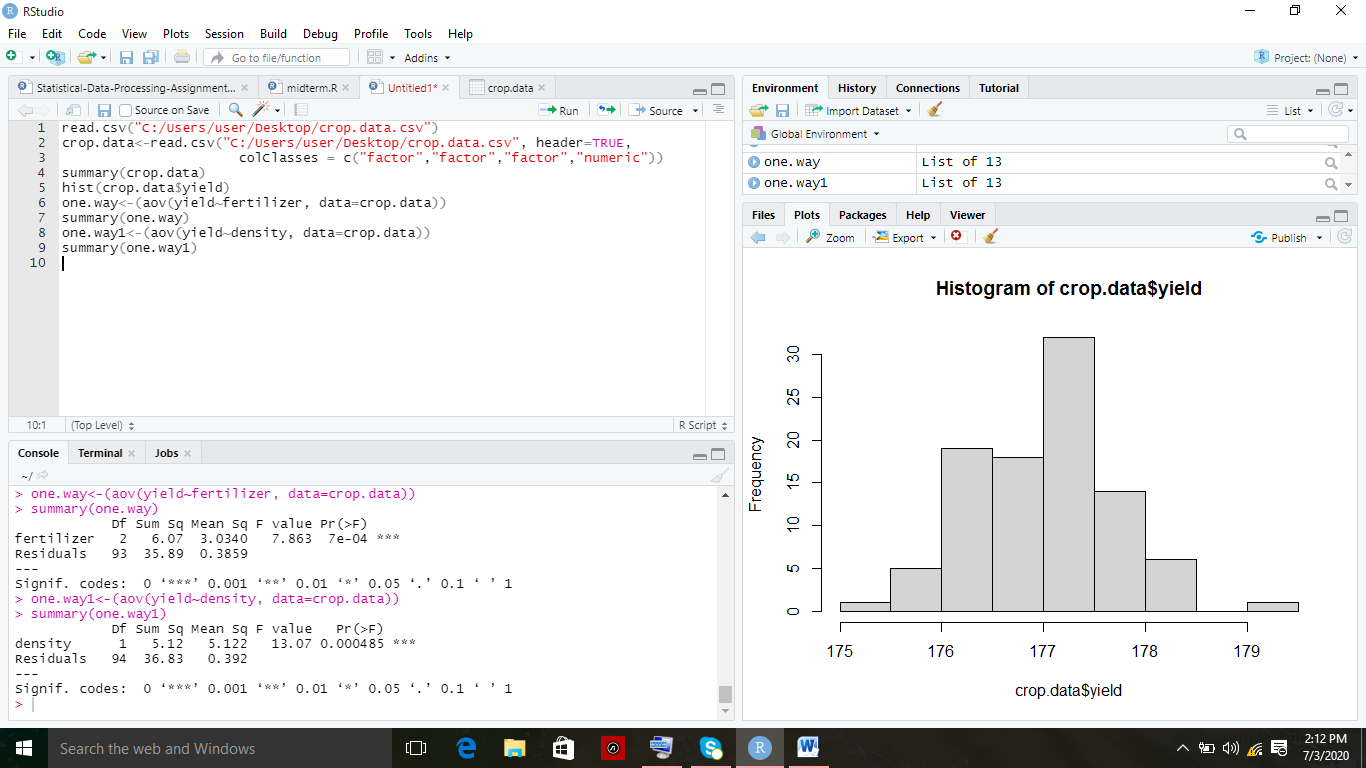


one.way<-(aov(yield~fertilizer, data=crop.data))

summary(one.way)

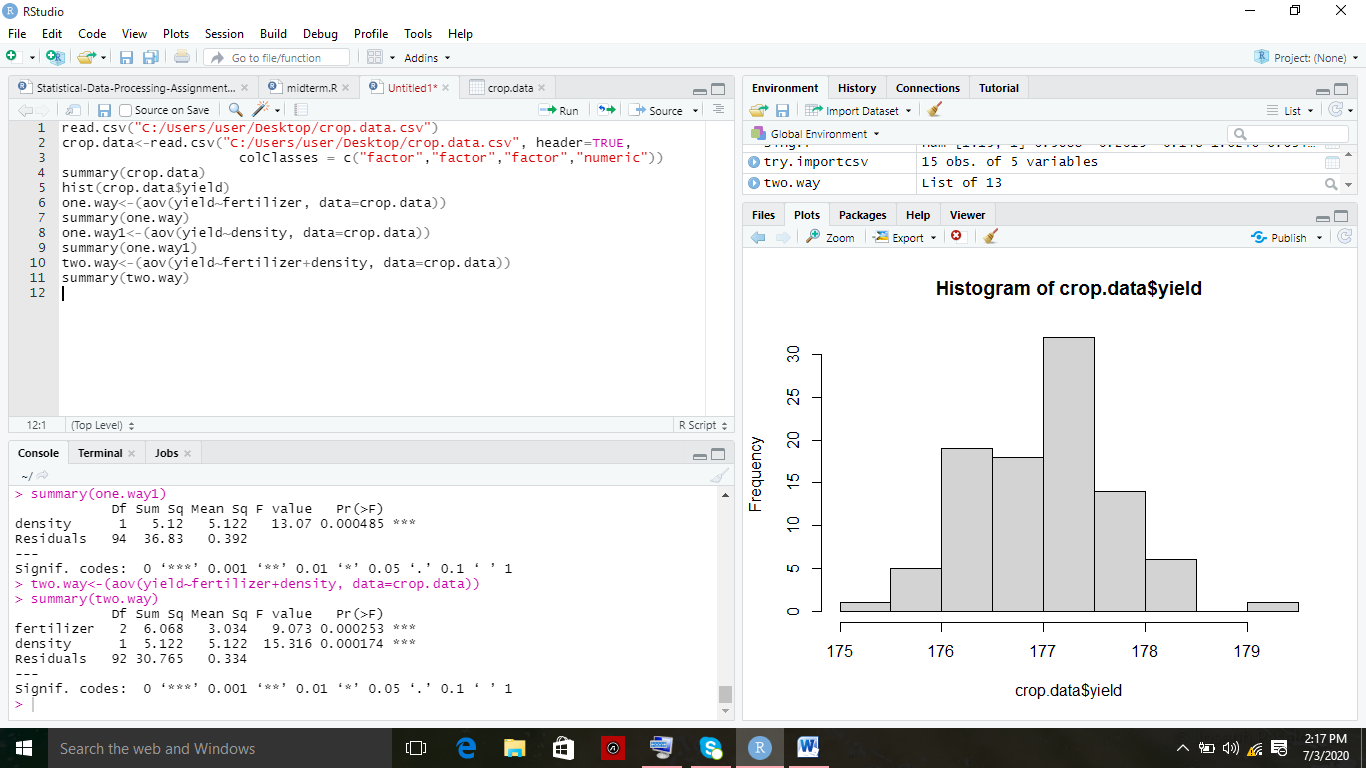
one.way1<-(aov(yield~density, data=crop.data))

summary(one.way1)



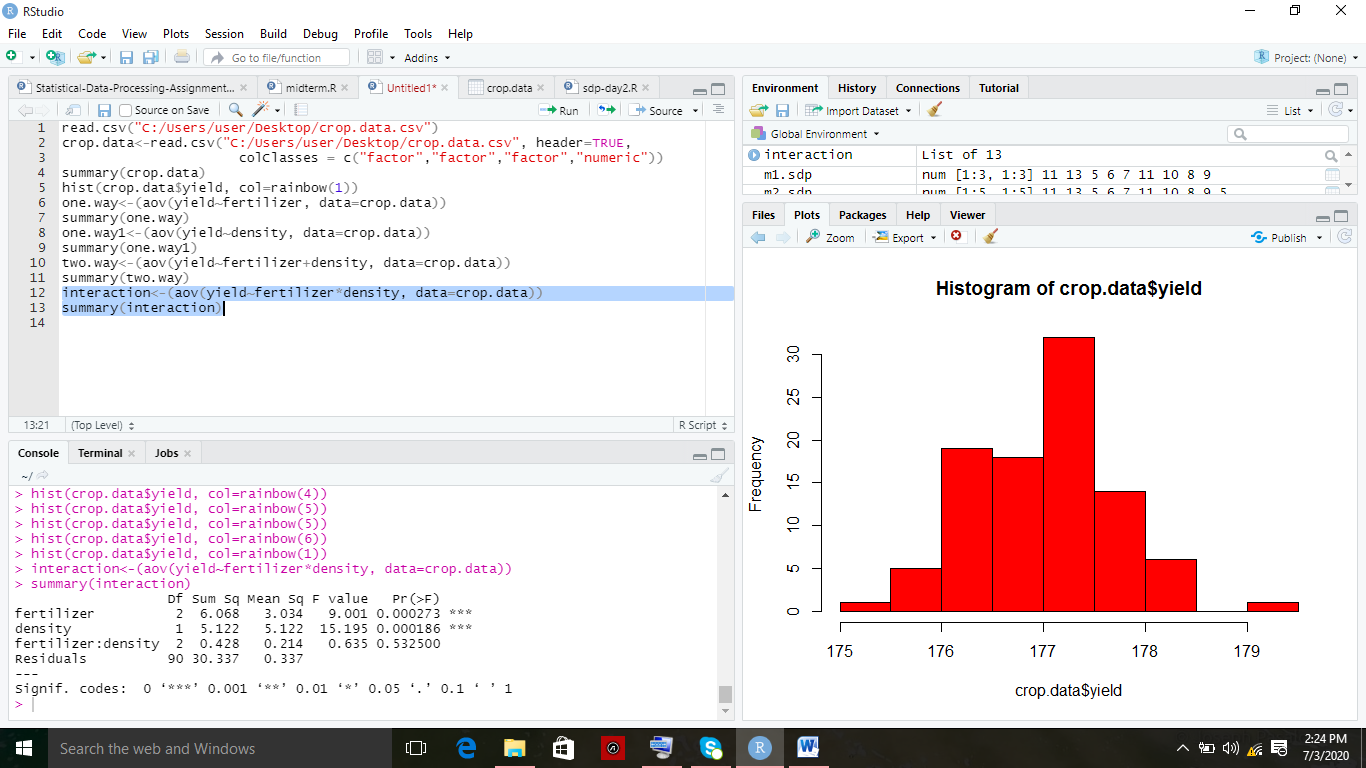
two.way<-(aov(yield~fertilizer+density, data=crop.data))

summary(two.way)



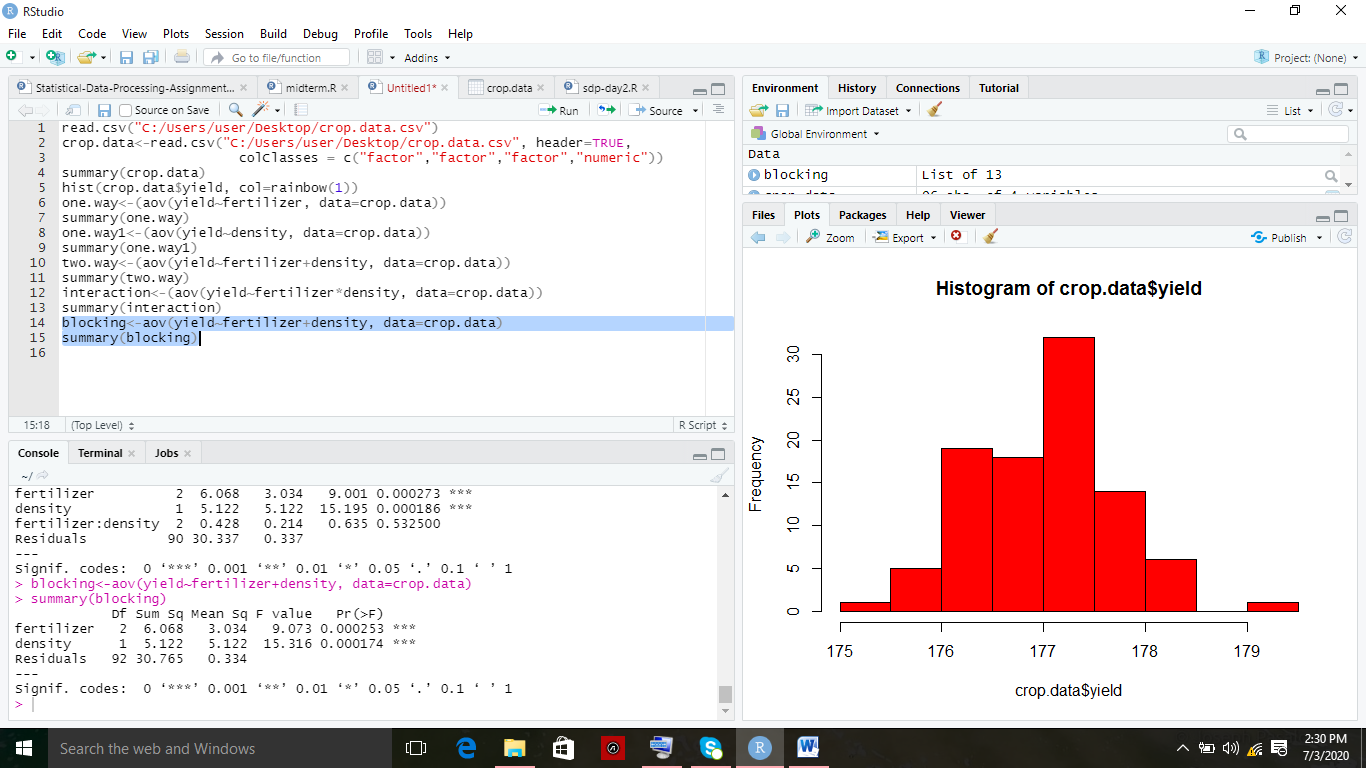
interaction<-(aov(yield~fertilizer\*density, data=crop.data))

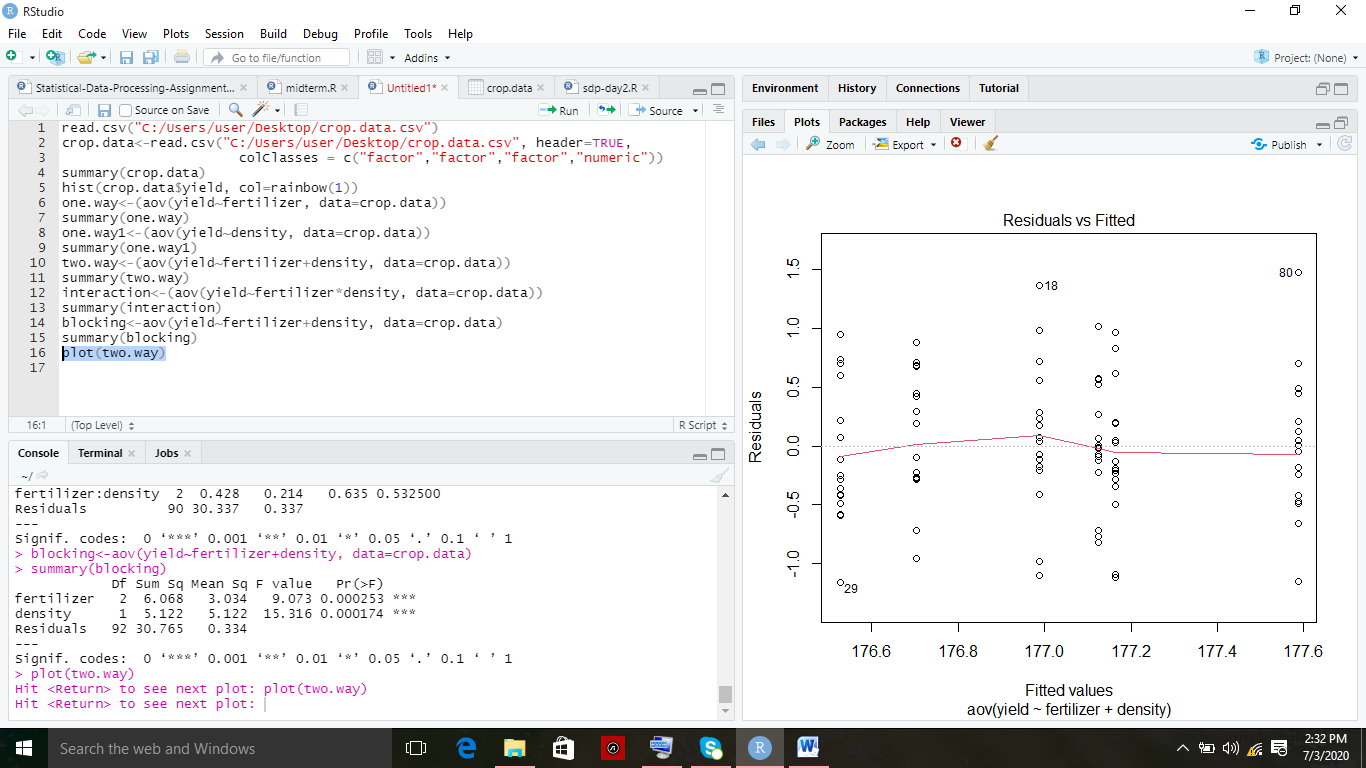
summary(interaction)

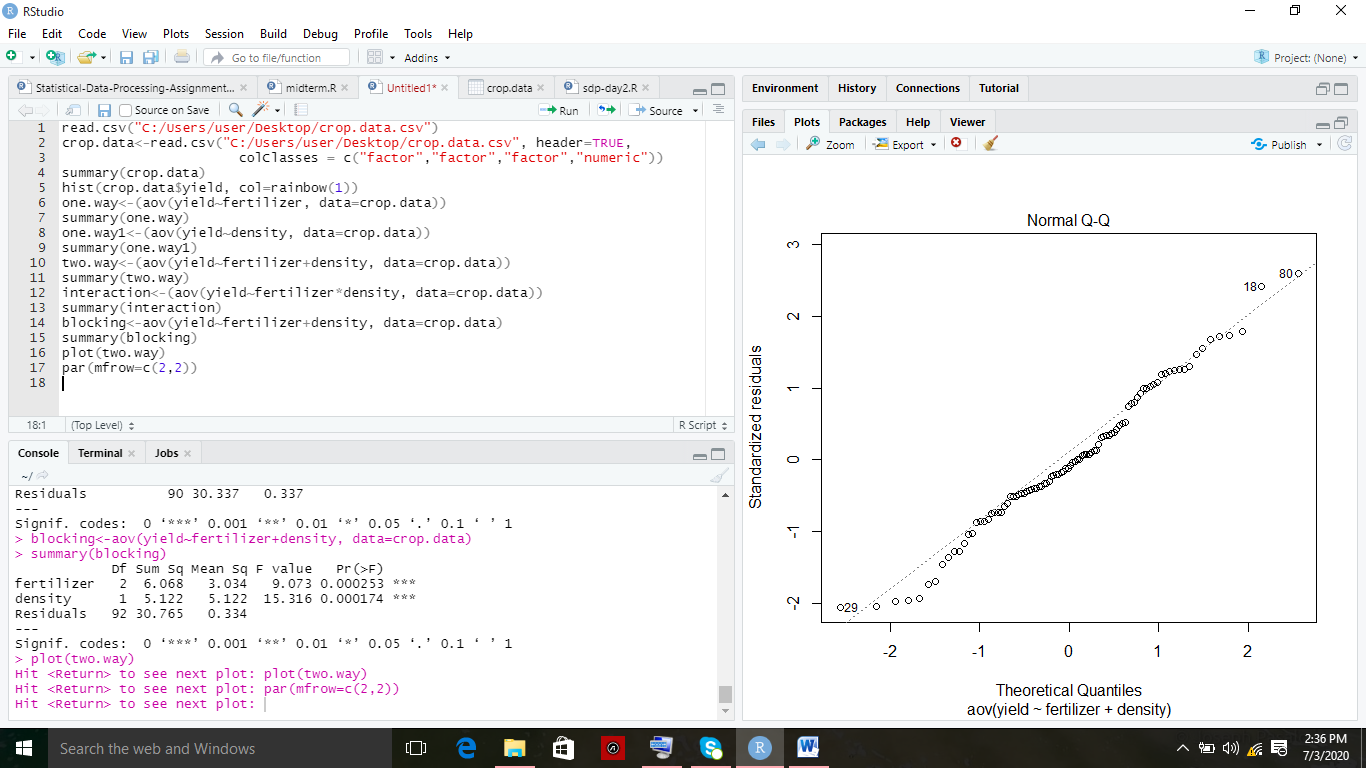


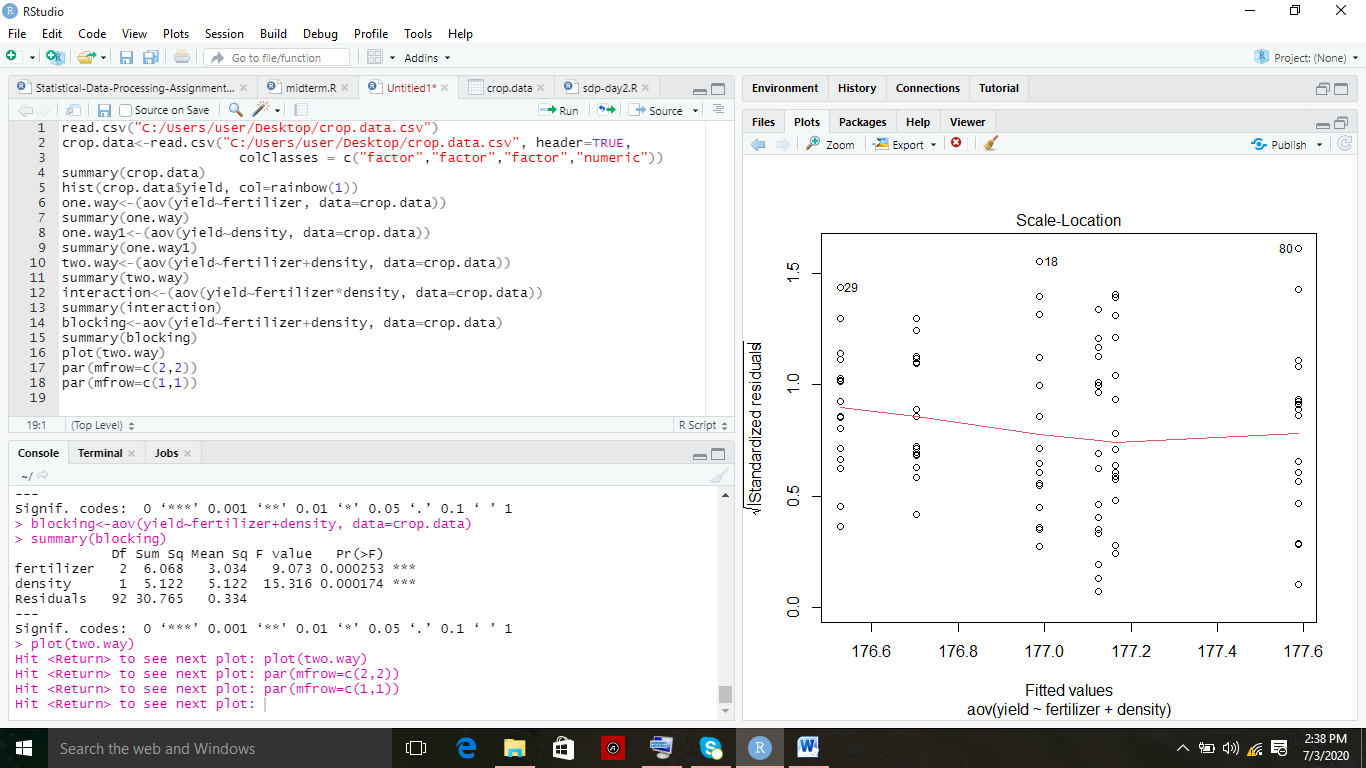
blocking<-aov(yield~fertilizer+density, data=crop.data)

summary(blocking)









plot(two.way)

par(mfrow=c(2,2))

par(mfrow=c(1,1))

