

# Soybean Analysis

Efoli Matthew

11/08/2021

## Activities

1. Create a new Rstudio Project
2. Create and save a new analysis.Rmd file
3. Open the analysis.Rmd file and start working on it

4. Import the data from a google sheet (library gsheet). Give a name to the dataframe object ([https://docs.google.com/spreadsheets/d/16BdPdmy6GYsrkahFiEM\\_y5Hy13D3psKfGJayBjVdNw/edit?usp=sharing](https://docs.google.com/spreadsheets/d/16BdPdmy6GYsrkahFiEM_y5Hy13D3psKfGJayBjVdNw/edit?usp=sharing) )

## 5. Check the structure of the data set (glimpse())

```
## function (... , list = character(), package = NULL, lib.loc = NULL, verbose = getOption("verbose"),  
##      envir = .GlobalEnv, overwrite = TRUE)
```

## 6. Save the data.frame as data.csv in the project directory

```
#write.csv(data, 'data.csv')
```

## 7. Open the data in the data.csv file and assign to a new dataframe

```
soybean <- data.frame(read.csv("data.csv", header = T))  
head(soybean)
```

```
##   X study year   location state block      treat  
## 1 1    424 2020 CAMPO VERDE    MT      1 Aproach Prima  
## 2 2    424 2020 CAMPO VERDE    MT      2 Aproach Prima  
## 3 3    424 2020 CAMPO VERDE    MT      3 Aproach Prima  
## 4 4    424 2020 CAMPO VERDE    MT      4 Aproach Prima  
## 5 5    424 2020 CAMPO VERDE    MT      1      Ativum  
## 6 6    424 2020 CAMPO VERDE    MT      2      Ativum  
##                                     ai sev  yld  
## 1                                picoxystrobin + cyproconazole 28 3060  
## 2                                picoxystrobin + cyproconazole 28 3145  
## 3                                picoxystrobin + cyproconazole 47 3104  
## 4                                picoxystrobin + cyproconazole 50 3072  
## 5 epoxiconazole + fluxapiroxade+pyraclostrobin 38 3699  
## 6 epoxiconazole + fluxapiroxade+pyraclostrobin 35 3498
```

## 8. Start exploring the data. First, subset the trials and create four different dataframes, one for each trial (there are four trials)

The Soybean data set is subset into 4 different data frames by trials.

```
## [1] "424" "425" "426" "427"

## 'data.frame': 112 obs. of 10 variables:
## $ X : int 1 2 3 4 5 6 7 8 9 10 ...
## $ study : int 424 424 424 424 424 424 424 424 424 424 ...
## $ year : int 2020 2020 2020 2020 2020 2020 2020 2020 2020 2020 ...
## $ location: chr "CAMPO VERDE" "CAMPO VERDE" "CAMPO VERDE" "CAMPO VERDE" ...
## $ state : chr "MT" "MT" "MT" "MT" ...
## $ block : Factor w/ 4 levels "1","2","3","4": 1 2 3 4 1 2 3 4 1 2 ...
## $ treat : Factor w/ 7 levels "Approach Prima",...: 1 1 1 1 2 2 2 2 3 3 ...
## $ ai : chr "picoxystrobin + cyproconazole" "picoxystrobin + cyproconazole" "picoxystrobin + cyproconazole" ...
## $ sev : int 28 28 47 50 38 35 45 40 55 70 ...
## $ yld : int 3060 3145 3104 3072 3699 3498 3420 3387 2658 2994 ...
```

## 9. Produce plots to visualize the response variables

We have two response variables namely; yield and severity. We would present the yield and severity response variables using bar charts.

### Data Visualization for Trial 1

The table shows the soybean mean yield, standard deviation by treatment for trial 1. The mean of the blocks is estimated as the yield for each treatment.

```
## Loading required package: plyr

## -----

## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)

## -----

##

## Attaching package: 'plyr'

## The following object is masked from 'package:ggpubr':
##
## mutate

## The following objects are masked from 'package:dplyr':
##
## arrange, count, desc, failwith, id, mutate, rename, summarise,
## summarize
```

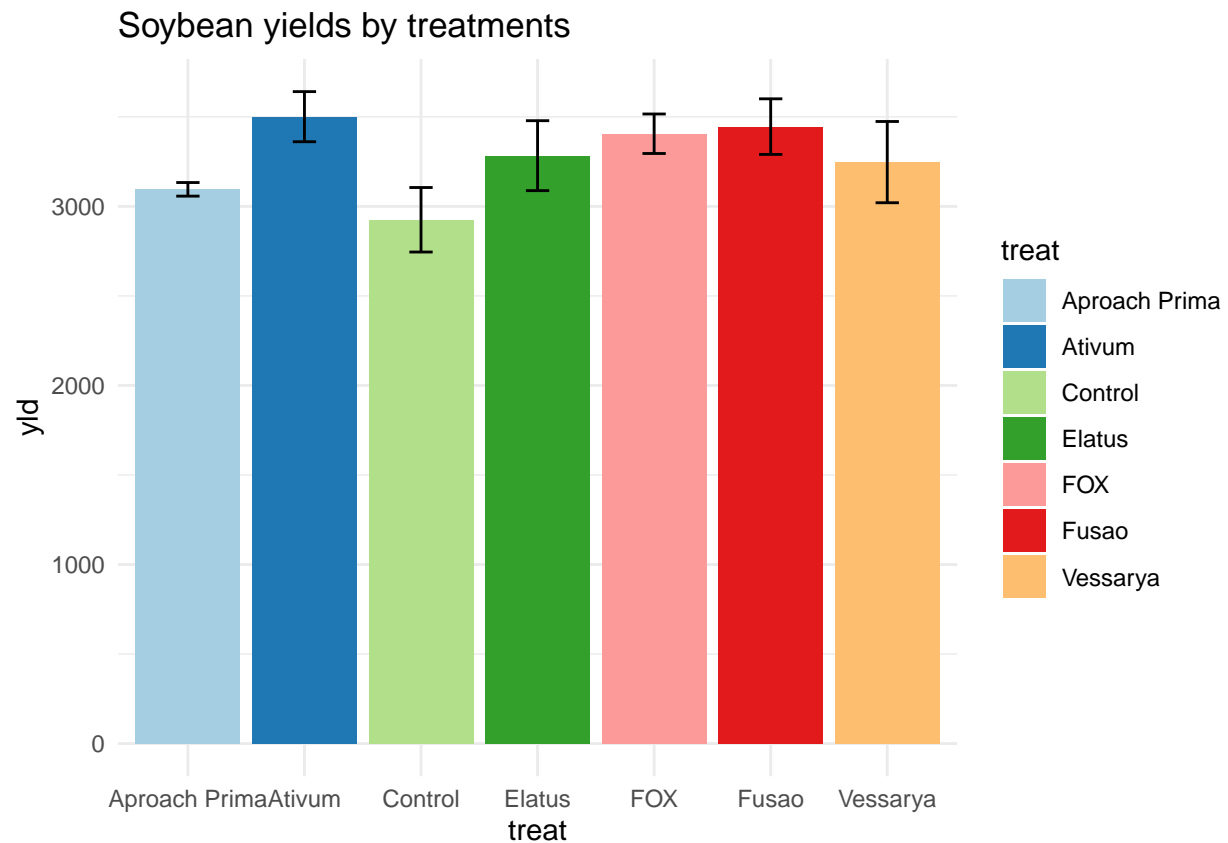
##		treat	yld	sd
## 1	Approach	Prima	3095.25	38.01206
## 2		Ativum	3501.00	139.96428
## 3		Control	2925.25	180.15063
## 4		Elatus	3283.25	195.30382
## 5		FOX	3405.75	110.29166
## 6		Fusao	3445.50	155.22135
## 7		Vessarya	3247.25	226.84264

From the above table, Soybean treated with treatments, Ativum, Fusao, FOX have the highest average yield; 3501, 3445.50 and 3405.75 respectively.

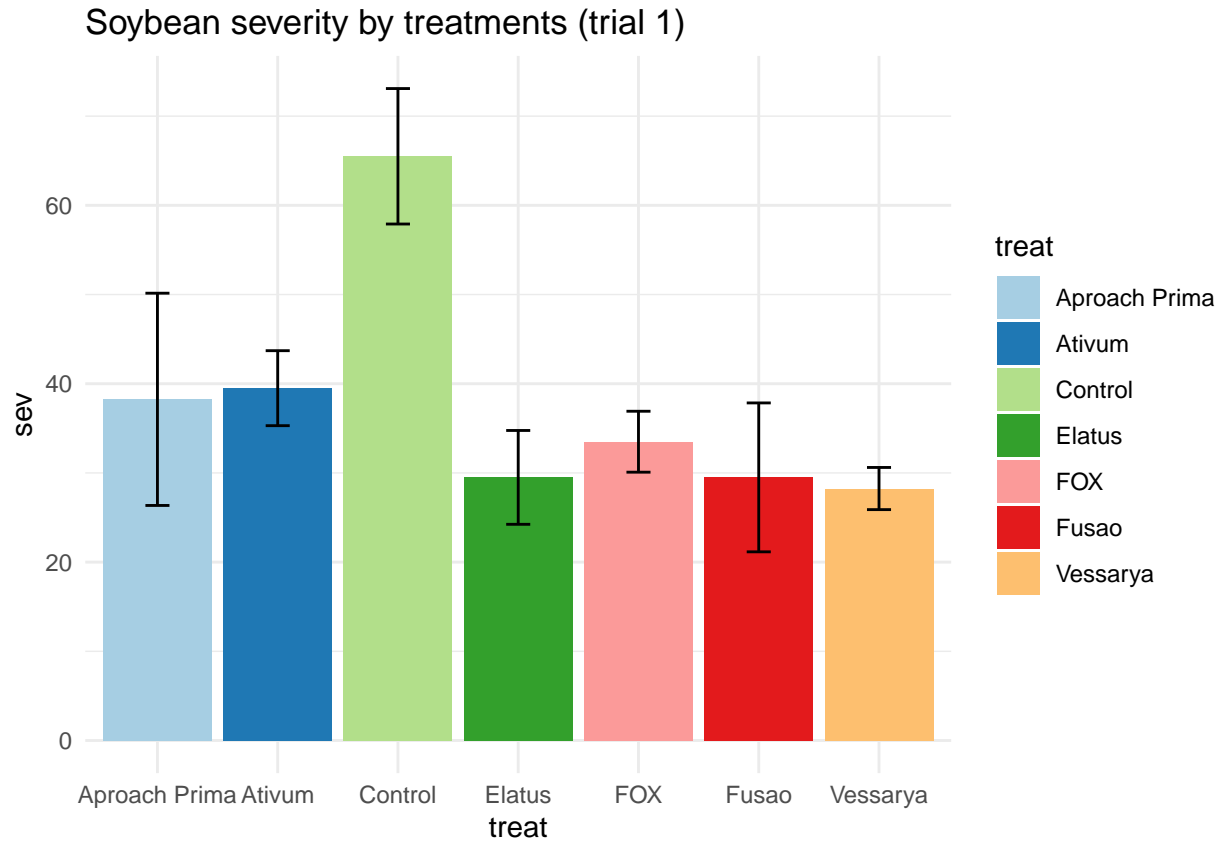
The table shows the soybean mean severity, standard deviation by treatment for the trial 1. The mean of the blocks is estimated as the yield for each treatment.

##		treat	sev	sd
## 1	Approach	Prima	38.25	11.898879
## 2		Ativum	39.50	4.203173
## 3		Control	65.50	7.593857
## 4		Elatus	29.50	5.259911
## 5		FOX	33.50	3.415650
## 6		Fusao	29.50	8.346656
## 7		Vessarya	28.25	2.362908

From the above table, the control, and Soybean treated with treatments, Ativum, Aproach Prima have the highest average severity; 65.50, 39.50, 38.25 respectively.



From the above chart, Soybean treated with treatments, Ativum, Fusao, FOX have the highest average yield; 3501, 3445.50 and 3405.75 respectively.



The chart above shows that the control, and Soybean treated with treatments, Ativum, Aproach Prima have the highest average severity; 65.50, 39.50, 38.25 respectively.

## Data visualization for trial 2

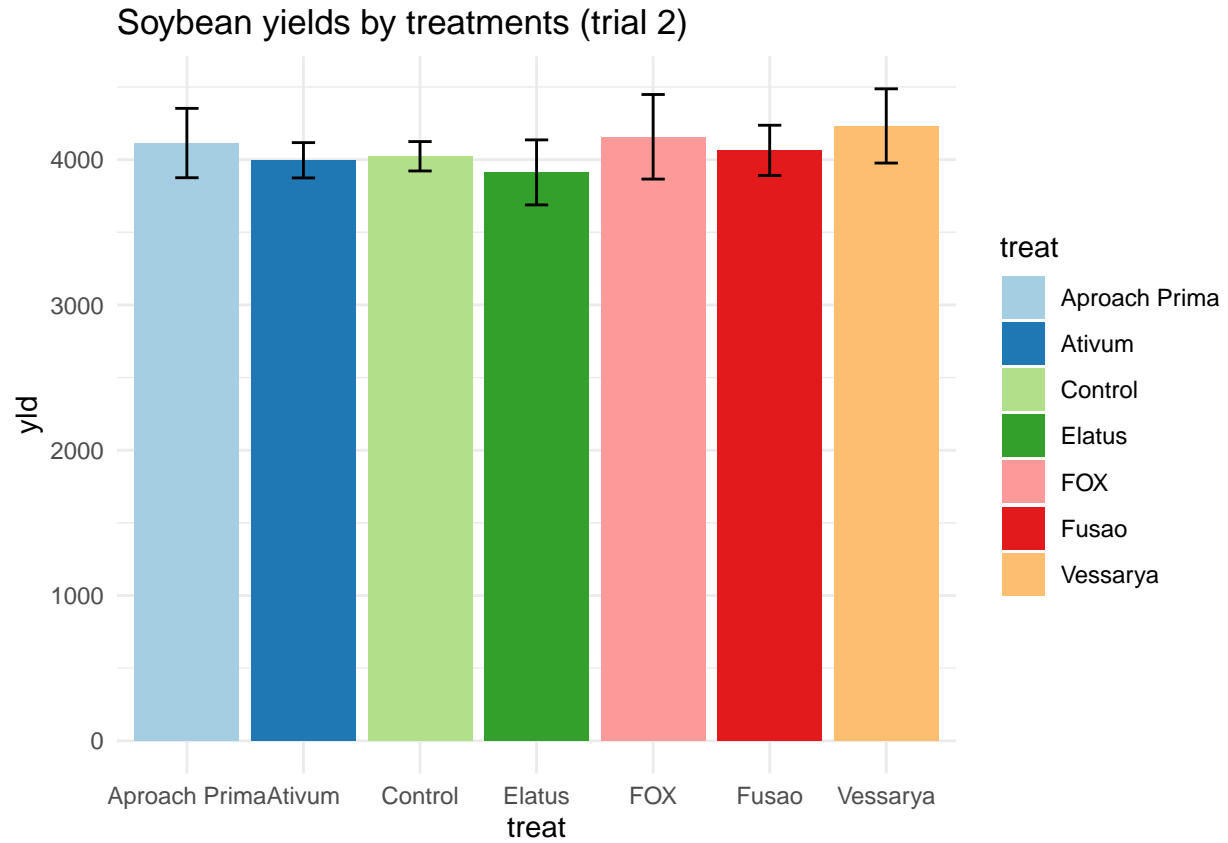
```
##          treat    yld    sd
## 1 Aproach Prima 4114.25 238.6872
## 2          Ativum 3995.75 121.5000
## 3          Control 4023.25 101.1645
## 4          Elatus 3912.25 223.6878
## 5           FOX 4157.25 291.1877
## 6          Fusao 4063.75 172.9824
## 7          Vessarya 4232.00 255.5113
```

From the above table, Soybean treated with treatments, Vessarya, FOX, Aproach Prima have the highest average yield; 4232, 4147.25, and 4114.25 respectively.

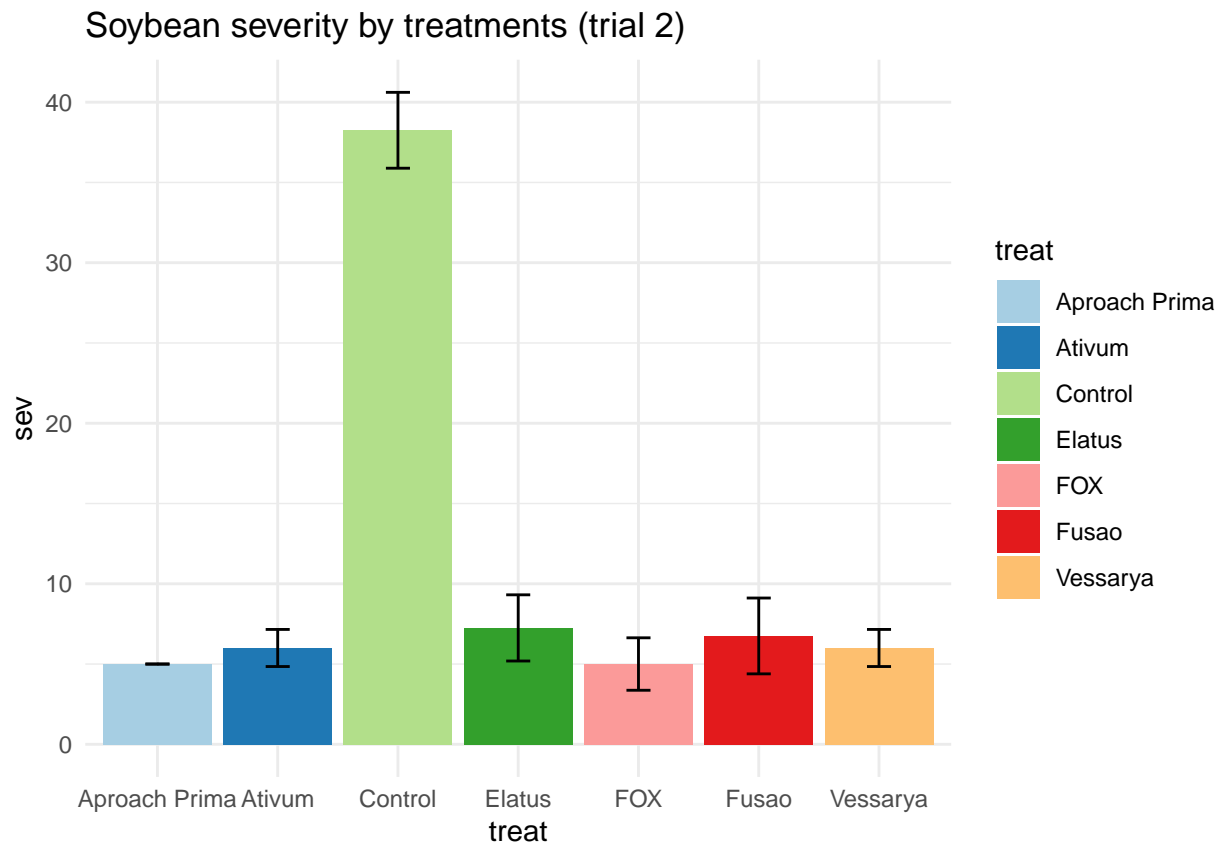
```
##          treat    sev    sd
## 1 Aproach Prima  5.00 0.000000
## 2          Ativum  6.00 1.154701
## 3          Control 38.25 2.362908
## 4          Elatus  7.25 2.061553
## 5           FOX   5.00 1.632993
```

```
## 6      Fusao  6.75 2.362908
## 7      Vessarya 6.00 1.154701
```

The above table shows that Control(38.25) and treatments; Elatus(7.25) and Fusao(6.75) had the highest severity.



From the above table, Soybean treated with treatments, Vessarya, FOX, Aproach Prima have the highest average yield; 4232, 4147.25, and 4114.25 respectively.



The above table shows that Control(38.25) and treatments; Elatus(7.25) and Fusao(6.75) had the highest severity.

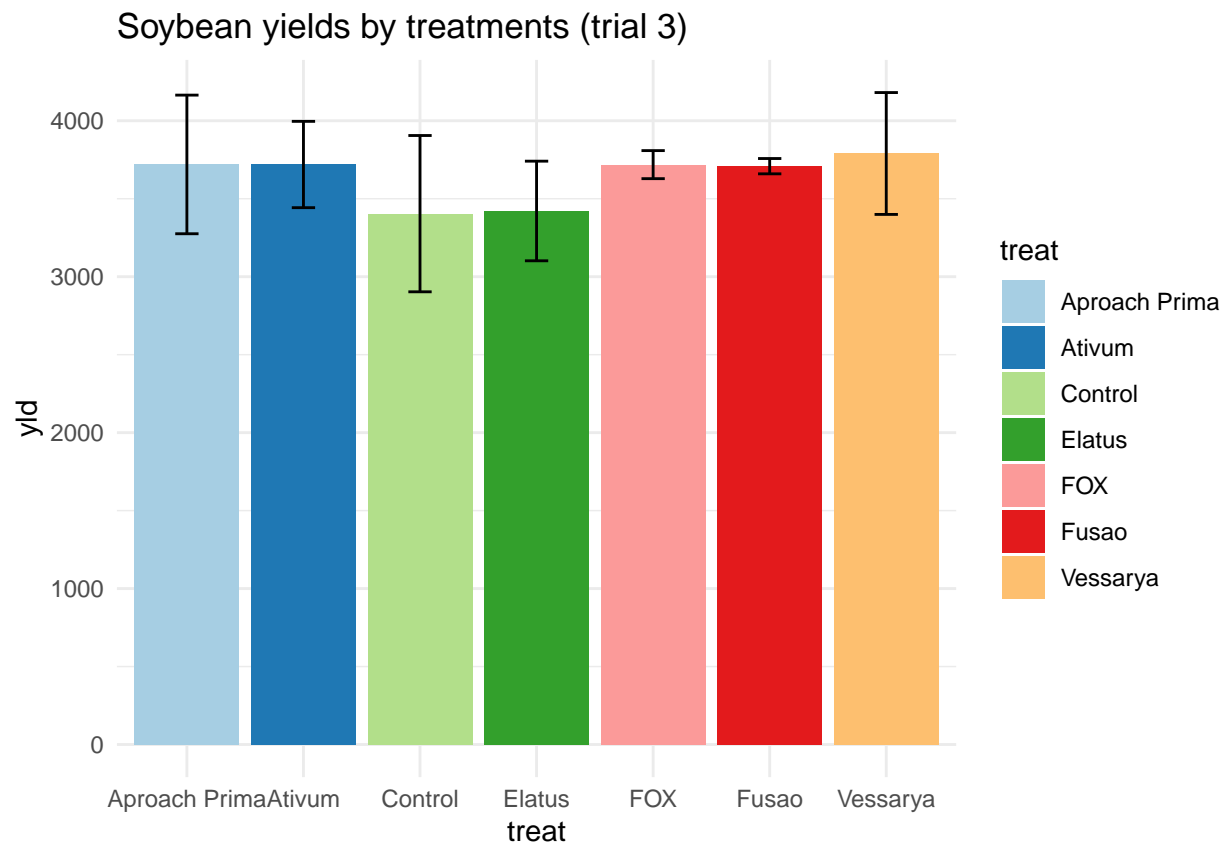
### Data visualization for trial 3

```
##          treat    yld      sd
## 1 Aproach Prima 3720.00 444.60994
## 2      Ativum 3719.50 276.85917
## 3      Control 3404.25 501.13962
## 4      Elatus 3421.50 319.41666
## 5         FOX 3718.75  89.92358
## 6        Fusao 3708.75  49.07392
## 7      Vessarya 3790.25 390.76538
```

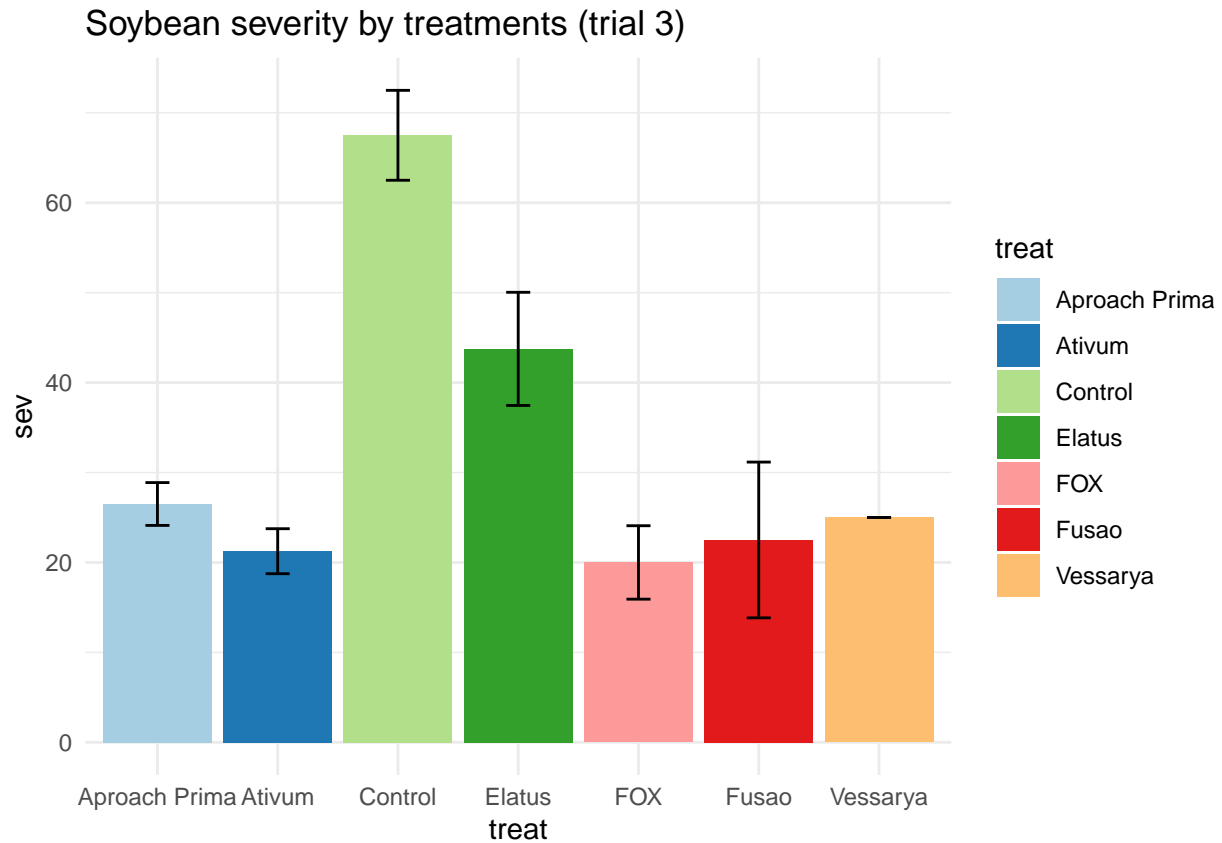
The above table shows that treatments; Vessarya(3790.25), Aproach Prima(3720.00) and Ativum (3719.50) had the highest yield.

```
##          treat    sev      sd
## 1 Aproach Prima 26.50 2.380476
## 2      Ativum 21.25 2.500000
## 3      Control 67.50 5.000000
## 4      Elatus 43.75 6.291529
## 5         FOX 20.00 4.082483
## 6        Fusao 22.50 8.660254
## 7      Vessarya 25.00 0.000000
```

The above table shows that Control(67.50) and treatments; Elatus(43.75) and Aproach Prima(26.50) had the highest severity.



The above chart shows that treatments; Vessarya(3790.25), Aproach Prima(3720.00) and Ativum (3719.50) had the highest yield.



The above chart shows that Control(67.50) and treatments; Elatus(43.75) and Aproach Prima(26.50) had the highest severity.

#### Data visualization for trial 4

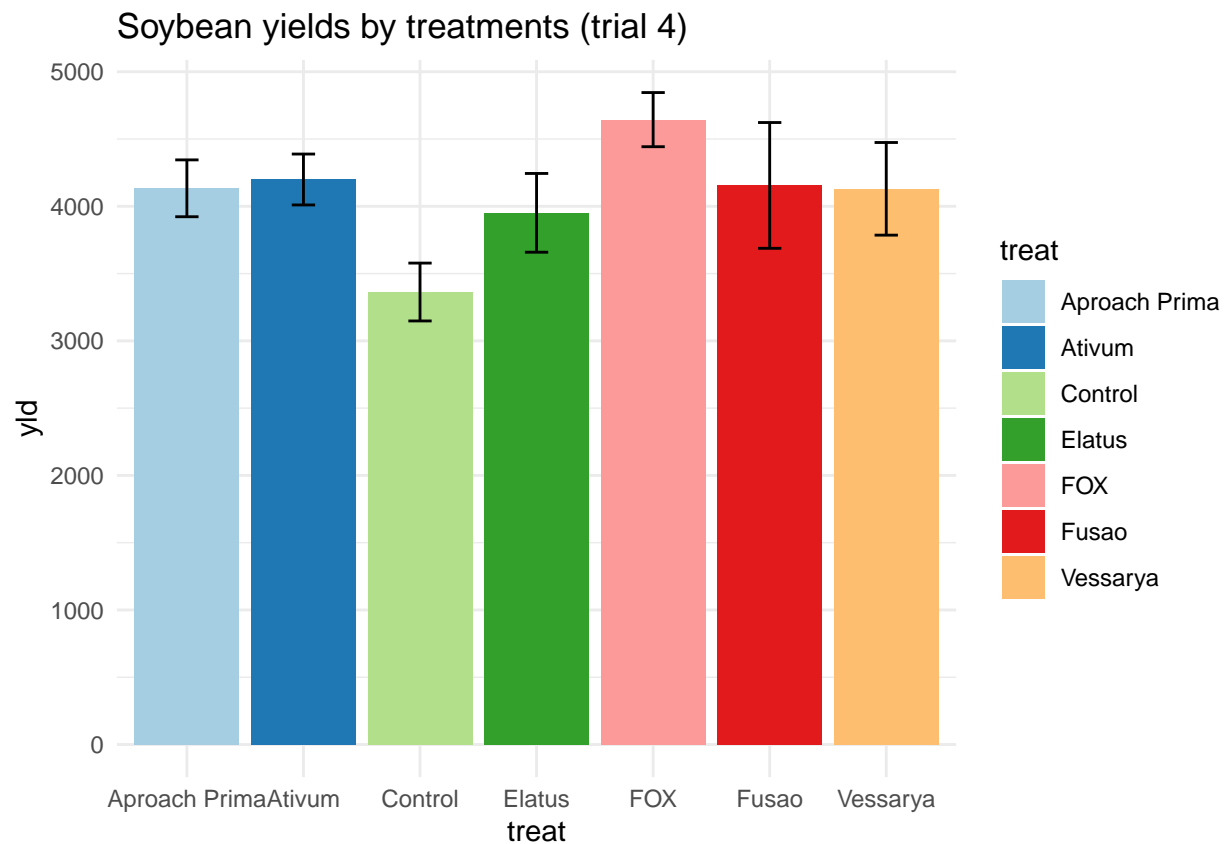
```
##          treat    yld      sd
## 1 Aproach Prima 4133.75 211.1988
## 2      Ativum 4199.25 189.2113
## 3      Control 3363.00 215.0612
## 4      Elatus 3951.50 292.6722
## 5          FOX 4644.25 201.2550
## 6      Fusao 4155.25 467.4551
## 7      Vessarya 4130.00 344.1928
```

The above table shows that treatments; FOX(4644.25), Ativum (4199.25) and Ativum (4155.25) had the highest yield in trial 4.

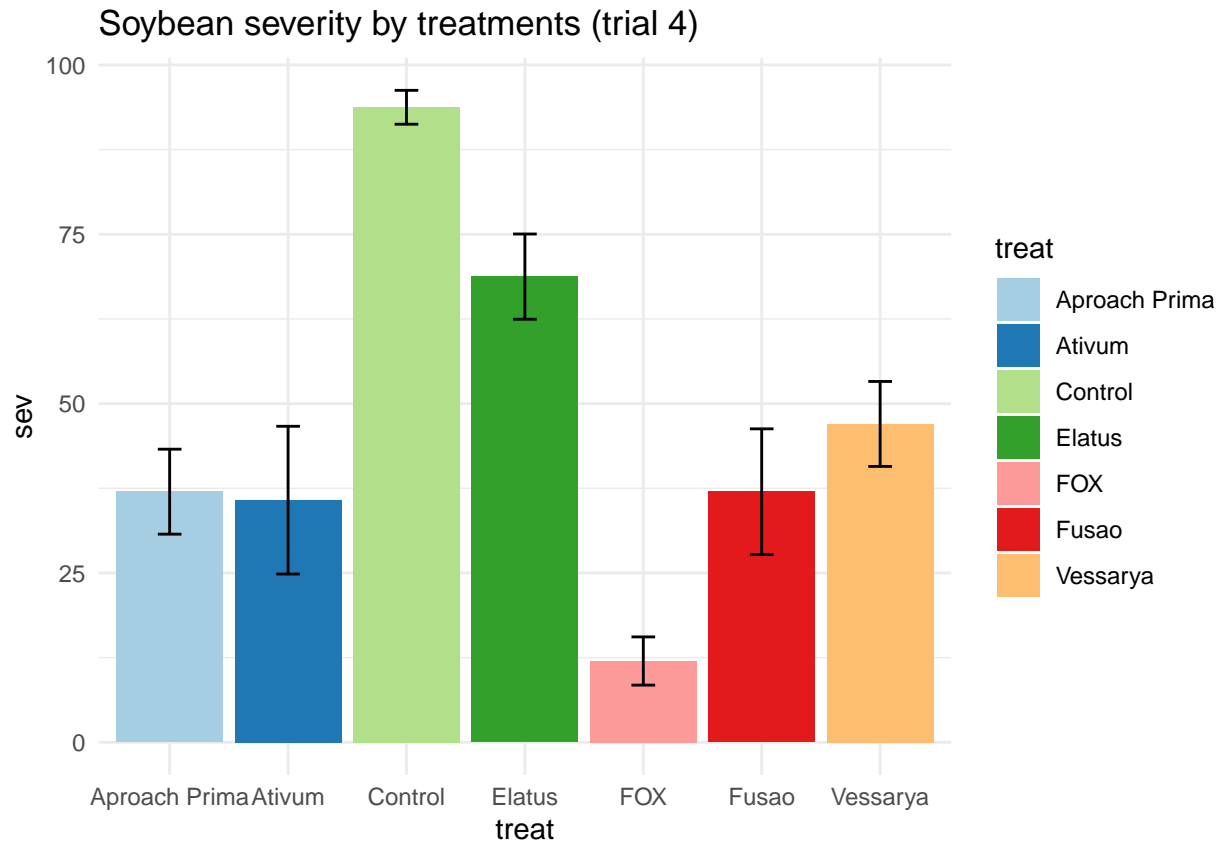
```
##          treat    sev      sd
## 1 Aproach Prima 37.00  6.271629
## 2      Ativum 35.75 10.904892
## 3      Control 93.75  2.500000
## 4      Elatus 68.75  6.291529
## 5          FOX 12.00  3.559026
## 6      Fusao 37.00  9.273618
## 7      Vessarya 47.00  6.271629
```



The above table shows that Control(93.75) and treatments; Elatus(68.75) and Vessarya(47.00) had the highest severity.



The above table shows that treatments; FOX(4644.25), Ativum (4199.25) and Ativum (4155.25) had the highest yield in trial 4.



The above chart shows that Control(93.75) and treatments; Elatus(68.75) and Vessarya(47.00) had the highest severity.

## 10. Test hypotheses on the apparent effects for each trial

The one-way analysis of variance (ANOVA), also known as one-factor ANOVA, is an extension of independent two-samples t-test for comparing means in a situation where there are more than two groups. In one-way ANOVA, the data is organized into several groups base on one single grouping variable (also called factor variable). This tutorial describes the basic principle of the one-way ANOVA test and provides practical anova test examples in R software.

The one-way analysis of variance (ANOVA), also known as one-factor ANOVA, is an extension of independent two-samples t-test for comparing means in a situation where there are more than two groups. In one-way ANOVA, the data is organized into several groups base on one single grouping variable (also called factor variable). This tutorial describes the basic principle of the one-way ANOVA test and provides practical anova test examples in R software.

A one-way anova test will be carried out as it is used to evaluate simultaneously the effect of two grouping variables (A and B) on a response variable. In this case, the grouping variables are block and treats while, the response variables are yield and severity.

### Hypothesis testing for Trial 1

**Hypothesis testing for yield (Trial 1)** Null hypothesis: the soybean yield means of the different treatments are the same.

Alternative hypothesis: At least one treatment means of the soybean yield is not equal to the others.

```
##
## Call:
## lm(formula = yld ~ treat, data = trl1)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##          3095.2           405.8          -170.0           188.0           310.5
##      treatFusao      treatVessarya
##          350.3           152.0

## Analysis of Variance Table
##
## Response: yld
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat       6 1010590  168432  6.5739 0.000507 ***
## Residuals  21  538045    25621
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.01, we can conclude that there are significant differences between the groups highlighted with “\*\*\*” in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean yield.

**Hypothesis testing for severity (Trial 1)** Null Hypothesis: The treatments has no effects on the soybean severity.

Alternate Hypothesis: The treatments has effects on the soybean severity.

```
##
## Call:
## lm(formula = sev ~ treat, data = trl1)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##          38.25           1.25          27.25          -8.75          -4.75
##      treatFusao      treatVessarya
##          -8.75          -10.00

## Analysis of Variance Table
##
## Response: sev
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat       6 4071.2   678.54  14.328 1.825e-06 ***
## Residuals  21  994.5    47.36
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.001, we can conclude that there are significant differences between the groups highlighted with “\*\*\*” in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean severity.

## Hypothesis testing for Trial 2

**Hypothesis testing for yield (Trial 2)** Null hypothesis: the soybean yield means of the different treatments are the same.

Alternative hypothesis: At least one treatment means of the soybean yield is not equal to the others.

```
##
## Call:
## lm(formula = yld ~ treat, data = trl2)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      4114.3          -118.5           -91.0          -202.0           43.0
##      treatFusao      treatVessarya
##      -50.5           117.7

## Analysis of Variance Table
##
## Response: yld
##           Df Sum Sq Mean Sq F value Pr(>F)
## treat      6 273708   45618   1.0235 0.4374
## Residuals 21 936010   44572
```

As the p-value = 0.4374 which is greater than the significance level 0.05, we can conclude that there are no significant differences between the groups highlighted with "\*" in the model summary. we will also accept the null hypothesis; the soybean yield means of the different treatments are the same.

**Hypothesis testing for severity (Trial 2)** Null Hypothesis: The treatments has no effects on the soybean severity.

Alternate Hypothesis: The treatments has effects on the soybean severity.

```
##
## Call:
## lm(formula = sev ~ treat, data = trl2)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      5.000e+00      1.000e+00      3.325e+01      2.250e+00     -8.333e-15
##      treatFusao      treatVessarya
##      1.750e+00      1.000e+00

## Analysis of Variance Table
##
## Response: sev
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat      6 3582.4   597.07  201.42 < 2.2e-16 ***
## Residuals 21   62.2     2.96
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.001, we can conclude that there are significant differences between the groups highlighted with "\*\*\*" in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean severity.

### Hypothesis testing for Trial 3

**Hypothesis testing for yield (Trial 3)** Null hypothesis: the soybean yield means of the different treatments are the same.

Alternative hypothesis: At least one treatment means of the soybean yield is not equal to the others.

```
##
## Call:
## lm(formula = yld ~ treat, data = trl3)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      3720.00          -0.50        -315.75         -298.50         -1.25
##      treatFusao      treatVessarya
##      -11.25          70.25

## Analysis of Variance Table
##
## Response: yld
##           Df Sum Sq Mean Sq F value Pr(>F)
## treat      6  598170   99695   0.8826  0.5246
## Residuals 21 2372067  112956
```

As the p-value = 0.5246 which is greater than the significance level 0.05, we can conclude that there are no significant differences between the groups highlighted with "\*" in the model summary. Therefore, we will also accept the null hypothesis; the soybean yield means of the different treatments are the same.

**Hypothesis testing for severity (Trial 3)** Null Hypothesis: The treatments has no effects on the soybean severity.

Alternate Hypothesis: The treatments has effects on the soybean severity.

```
##
## Call:
## lm(formula = sev ~ treat, data = trl3)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      26.50          -5.25         41.00         17.25         -6.50
##      treatFusao      treatVessarya
##      -4.00          -1.50

## Analysis of Variance Table
##
## Response: sev
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat      6  7305.9  1217.65  50.685 2.045e-11 ***
## Residuals 21   504.5    24.02
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.001, we can conclude that there are significant differences between the groups highlighted with "\*\*\*" in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean severity.

## Hypothesis testing for Trial 4

**Hypothesis testing for yield (Trial 4)** Null hypothesis: the soybean yield means of the different treatments are the same.

Alternative hypothesis: At least one treatment means of the soybean yield is not equal to the others.

```
##
## Call:
## lm(formula = yld ~ treat, data = trl4)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      4133.75          65.50        -770.75        -182.25         510.50
##      treatFusao      treatVessarya
##      21.50          -3.75

## Analysis of Variance Table
##
## Response: yld
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat         6 3496841  582807   6.917 0.0003676 ***
## Residuals    21 1769402   84257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.001, we can conclude that there are significant differences between the groups highlighted with “\*\*\*” in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean yield.

**Hypothesis testing for severity (Trial 4)** Null Hypothesis: The treatments has no effects on the soybean severity.

Alternate Hypothesis: The treatments has effects on the soybean severity.

```
##
## Call:
## lm(formula = sev ~ treat, data = trl4)
##
## Coefficients:
## (Intercept)      treatAtivum      treatControl      treatElatus      treatFOX
##      3.700e+01      -1.250e+00      5.675e+01      3.175e+01      -2.500e+01
##      treatFusao      treatVessarya
##      2.284e-14        1.000e+01

## Analysis of Variance Table
##
## Response: sev
##           Df Sum Sq Mean Sq F value    Pr(>F)
## treat         6 16837.9 2806.31  57.425 6.055e-12 ***
## Residuals    21  1026.2   48.87
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As the p-value is less than the significance level 0.001, we can conclude that there are significant differences between the groups highlighted with “\*\*\*” in the anova table. We will also reject the null hypothesis which states The treatments has no effects on the soybean severity.

## 11. Check the correlation between severity and yield for each trial.

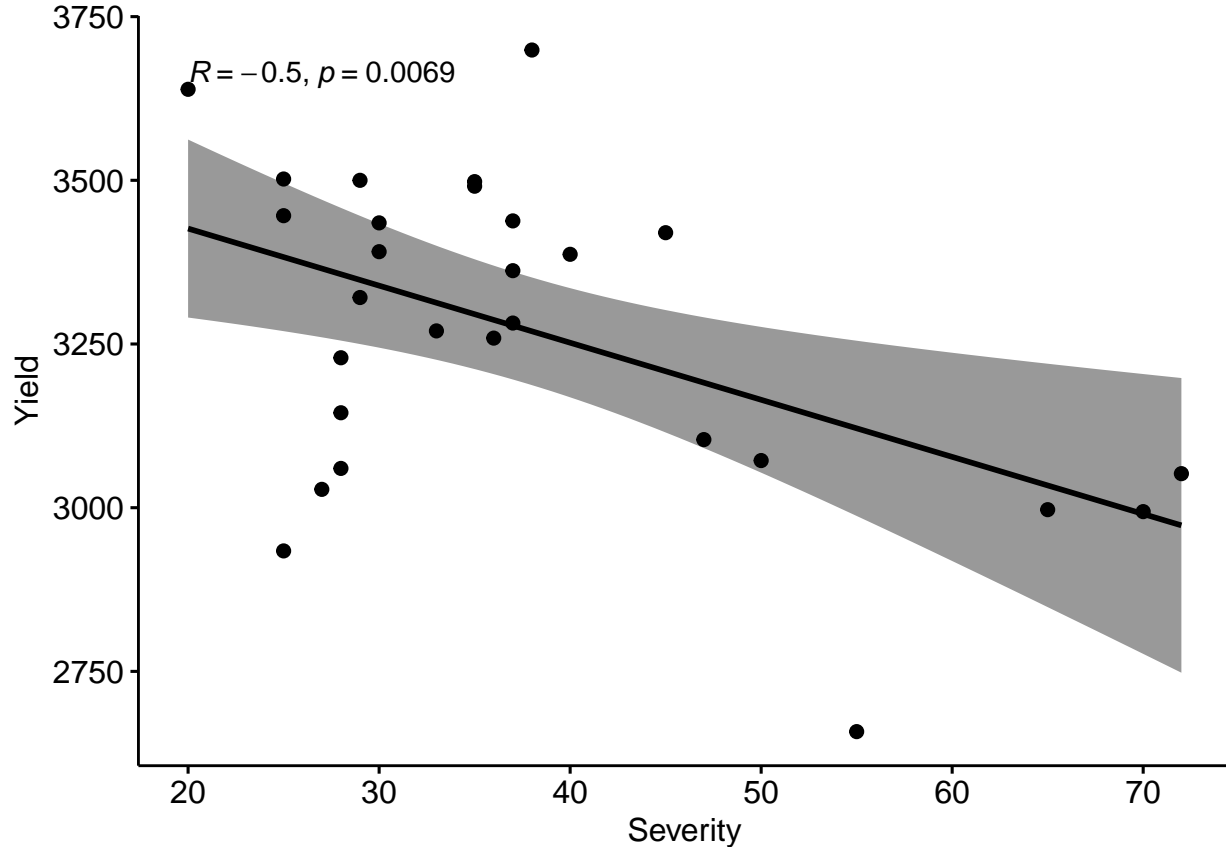
If the p-value is  $< 5\%$ , then the correlation between x and y is significant.

Visualizing the correlation between severity and yield using a scatter plot for Trial 1

```
## [1] -0.4985777

##
## Pearson's product-moment correlation
##
## data: trl1$sev and trl1$yld
## t = -2.9328, df = 26, p-value = 0.006925
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.7349484 -0.1541793
## sample estimates:
## cor
## -0.4985777

## 'geom_smooth()' using formula 'y ~ x'
```



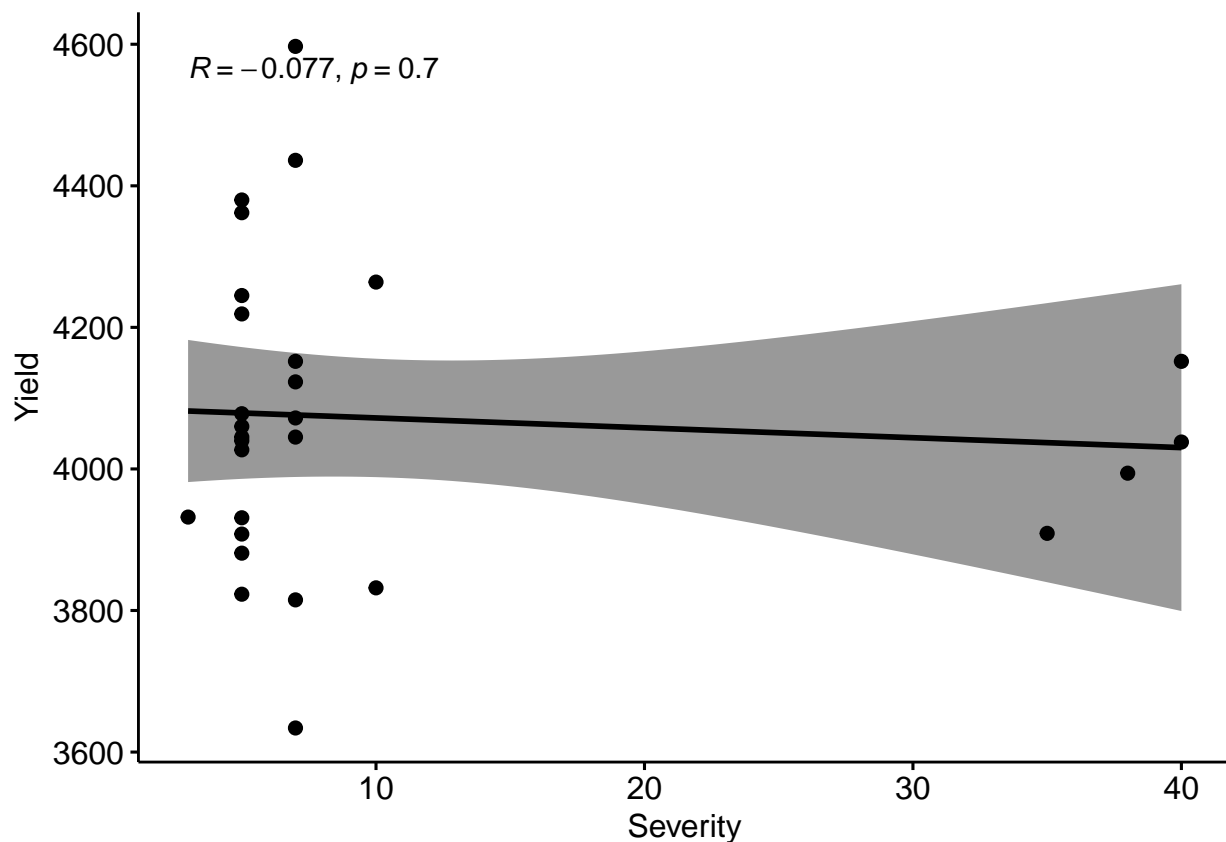
The p-value of the test is 0.006925, which is less than the significance level  $\alpha = 0.05$ . We can conclude that severity and yield are significantly correlated with a correlation coefficient of -0.49837 and p-value of 0.006925. A correlation coefficient of -0.49837 indicates a strong negative correlation : this means that every time severity increases, yield decreases.

### Visualizing the correlation between severity and yield using a scatter plot for Trial 2

```
## [1] -0.07672589

##
## Pearson's product-moment correlation
##
## data: trl2$sev and trl2$yld
## t = -0.39238, df = 26, p-value = 0.698
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.4372857 0.3050840
## sample estimates:
## cor
## -0.07672589

## 'geom_smooth()' using formula 'y ~ x'
```



The p-value of the test is 0.698, which is greater than the significance level  $\alpha = 0.05$ . We can conclude that severity and yield are not significantly correlated with a correlation coefficient of -0.07672589 and p-value of 0.698. There is almost no association between the response variables yield and severity for trial 2.

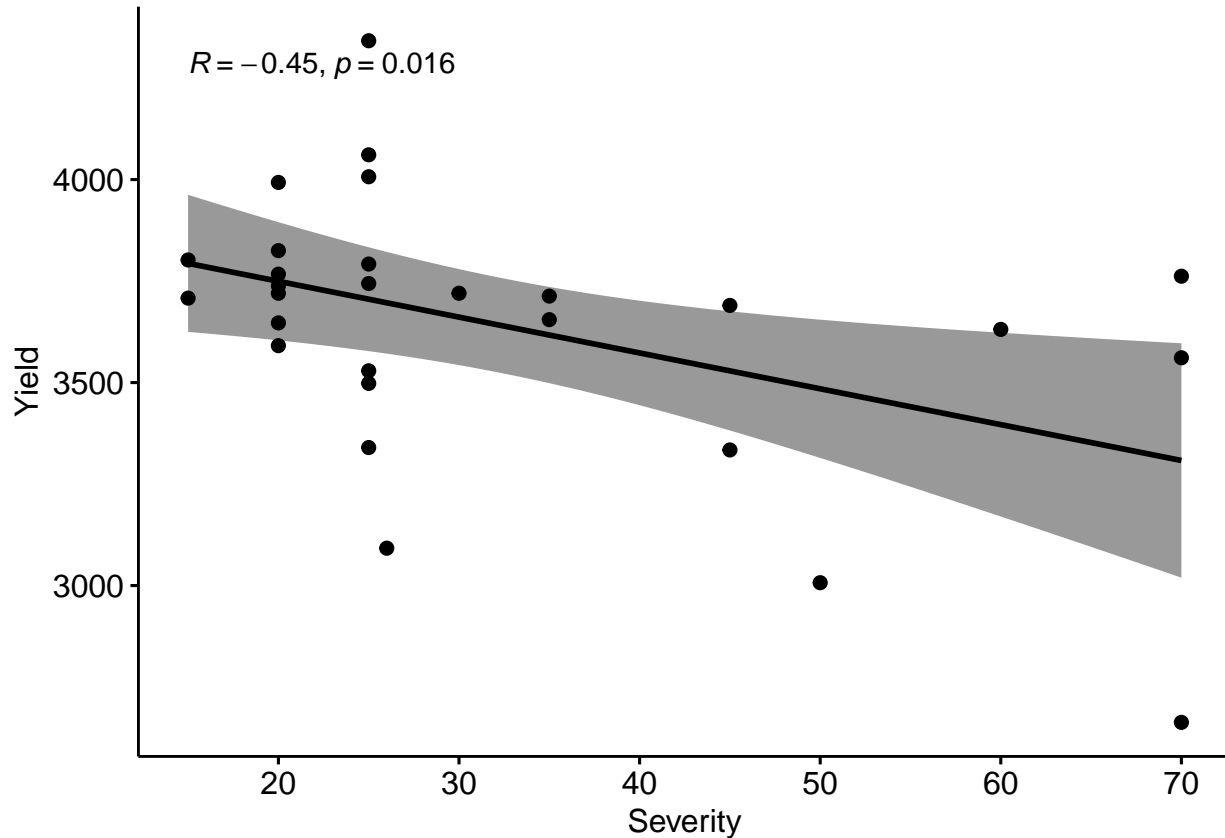


### Visualizing the correlation between severity and yield using a scatter plot for Trial 3

```
## [1] -0.4526989

##
## Pearson's product-moment correlation
##
## data: trl3$sev and trl3$yld
## t = -2.5888, df = 26, p-value = 0.01557
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.70646059 -0.09580208
## sample estimates:
##      cor
## -0.4526989

## 'geom_smooth()' using formula 'y ~ x'
```



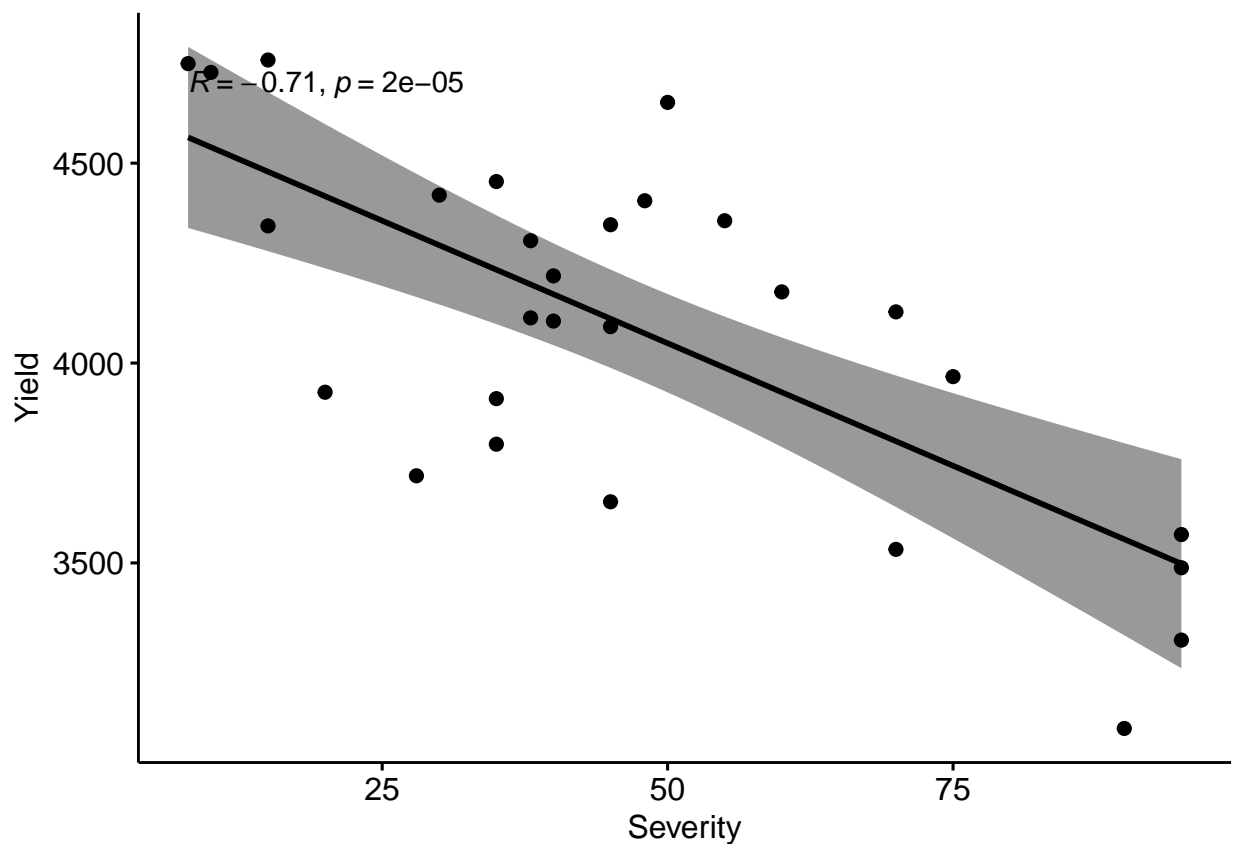
The p-value of the test is 0.01557, which is less than the significance level  $\alpha = 0.05$ . We can conclude that severity and yield are significantly correlated with a correlation coefficient of -0.4526989 and p-value of 0.01557. A correlation coefficient of -0.4526989 indicates a strong negative correlation : this means that every time severity increases, yield decreases for trial 3.

### Visualizing the correlation between severity and yield using a scatter plot for Trial 4

```
## [1] -0.713965
```

```
##
## Pearson's product-moment correlation
##
## data: trl4$sev and trl4$yld
## t = -5.1994, df = 26, p-value = 1.987e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.8583962 -0.4646556
## sample estimates:
## cor
## -0.713965

## 'geom_smooth()' using formula 'y ~ x'
```



The p-value of the test is 1.987e-05, which is less than the significance level  $\alpha = 0.05$ . We can conclude that severity and yield are significantly correlated with a correlation coefficient of -0.713965 and p-value of 1.987e-05. A correlation coefficient of -0.4526989 indicates a strong negative correlation : this means that every time severity increases, yield decreases for trial 3.