

# MT5762 PROJECT 1

ANDREAS MICHALIS LOUCAS

200029834

#### QUESTION 1 [Which 5 elements did you choose, and why?]

The 5 elements chosen to undertake the project are [Sc,Ti,Se,Sm,Tm]. These five elements have been chosen because at a first glimpse the groups for each element looked similar. We can visualize this characteristics from the [BOXPLOT\[1-5\]](#).

#### QUESTION 2 [Do the data indicate differences in the elemental composition of Cannabis leaves grown in different soil types?]

From the [BOXPLOT\[1-5\]](#) we cannot determine an obvious answer if there are differences in elemental composition since some soils data overlap with each other.

Our data has to fulfil the F-distribution assumptions to achieve precise results. Data must follow independence, normality and have constant spread between them.

For the normality assumption we need to test hypothesis **H<sub>0</sub>**: element follows normal distribution and **H<sub>1</sub>**: element does not follow normal distribution. We tested the hypothesis using a [Shapiro-Wilk](#) test and elements [Sc,Ti,Sm,Tm] follow normality, but [Se] does not and we will use an alternative test on this element that does not need normality the [Kruskal-Wallis](#) test. Results can be found in [TABLE\[SHAPIRO\]](#).

For constant spread assumption we need to test hypothesis **H<sub>0</sub>**:  $\sigma(\text{mb})^2 = \sigma(\text{bhb})^2 = \sigma(\text{pm})^2 = \sigma(\text{nth})^2$  and **H<sub>1</sub>**: at least 1 variance is not equal with at least 1 other variance. We tested the hypothesis using a [Levene's Test](#) and element [Sc,Ti,Se] satisfies it, but elements [Sm,Tm] do not satisfy it, so we will use an alternative test that does not assume equal standard deviations the [Welch's Anova](#) test. Results can be found in [TABLE\[LEVENE\]](#).

The alternative tests and [ANOVA](#) hypothesis for each of our 5 elements is, **H<sub>0</sub>**:  $\mu(\text{mb}) = \mu(\text{bhb}) = \mu(\text{pm}) = \mu(\text{nth})$  and **H<sub>1</sub>**: exists at least 1  $\mu$  not equal with at least 1 other  $\mu$ .

The results from the tests can be found in [TABLE\[ANOVA\]](#). From our results we reject **H<sub>0</sub>** for all 5 elements. Our conclusion is that elemental composition of Cannabis leaves is different depending on different soil types. The means that influence the rejection of **H<sub>0</sub>** can be discovered from [TABLE\[MEANS1-5\]](#). If the range between the lower and upper boundary does not contain zero, then our means between the two groups are significantly different.

#### QUESTION 3 [Are some of the elements related to one another in terms of their levels in the sampled leaves?]

I applied the correlation coefficient measure 'p' to construct my 5 pair elements. From my results I have chosen the pairs that have the largest positive linear relationship (large value of 'p'). Our pairs for the examination are [Sc-Ti,Sm-Tm,Sm-Sc,Sm-Ti,Sc-Tm]. Results of the correlation coefficient for these pairs can be found in [TABLE\[CORRELATION\]](#). When two elements are positive linear related, this implies that as the element on the x-axes increases so does the element on the y-axes, elements influence each other. All our pairs are positive linear related, but none had a value larger than 0.8. This means our data did not fit the linear slope perfectly. We now explore different soil groups between the element pairs. We can observe from the [GRAPH\[1-5\]](#) mb group soil points are close and tight together and show some relation between our pairs, but the other soil groups are spread out and produce clutter to our graph. Except from the pair Sm-Sc, the 3 soil groups look tight and all grouped together, maybe this is why it had the highest correlation value of 0.765. In the Sm-Tm pair we notice although the data is positive related, as nth for Sm increases, nth for Tm decreases, relationships of elemental composition between different pairs respond differently. If we look closely for soil group pm, we detect that for all pairs the data is spread and does not show a sign of relation. We can conclude some elements influence other elements in the leaves composition, but the relationship differs from soil groups. These relations can be withdrawn by examining [GRAPH\[1-5\]](#).

#### QUESTION 4 [Results of this experiment ultimately allow the determination of what soil the plants were grown in, just from the elemental composition of the leaves?]

When considering this experiment on a global scale, I think this cannot be achieved for the given reasons. Firstly, we only consider four soil types in this experiment in one country. In reality there are thousands of different soil types when considering all the globe[1]. Secondly, each country has different weather and climate circumstances which is an important variable for plant growth[2]. Thirdly, our element samples that we considered in this experiment did not fulfil our F-distribution assumptions fully and alternative tests were applied to some elements. Fourthly, we tested all the Hypothesis considering a 5% type-I-error, this implies that there is a 5% probability we accepted the wrong hypothesis. Lastly, there are different kinds of cannabis plants and they can be chemically strengthen which can change the elemental composition. These variables might influence and change the element composition in the cannabis leaves. Answer and Assumptions based on [References](#).

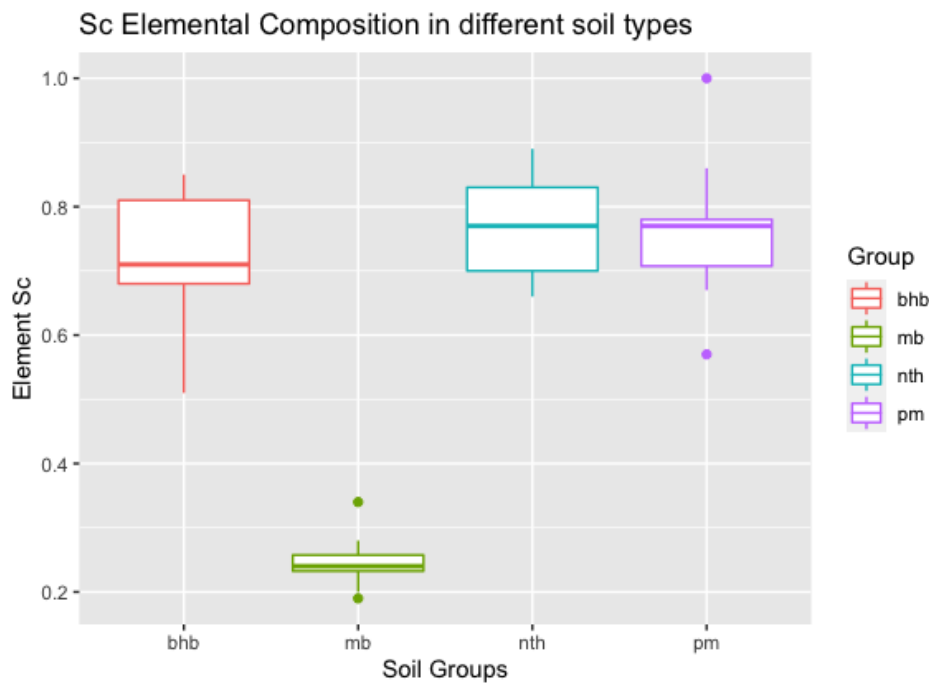
# APEENDIX

All the Graphs and Table results can be obtained by running the R-Code attached in the bottom of the report [APPENDIX-CODE](#) .

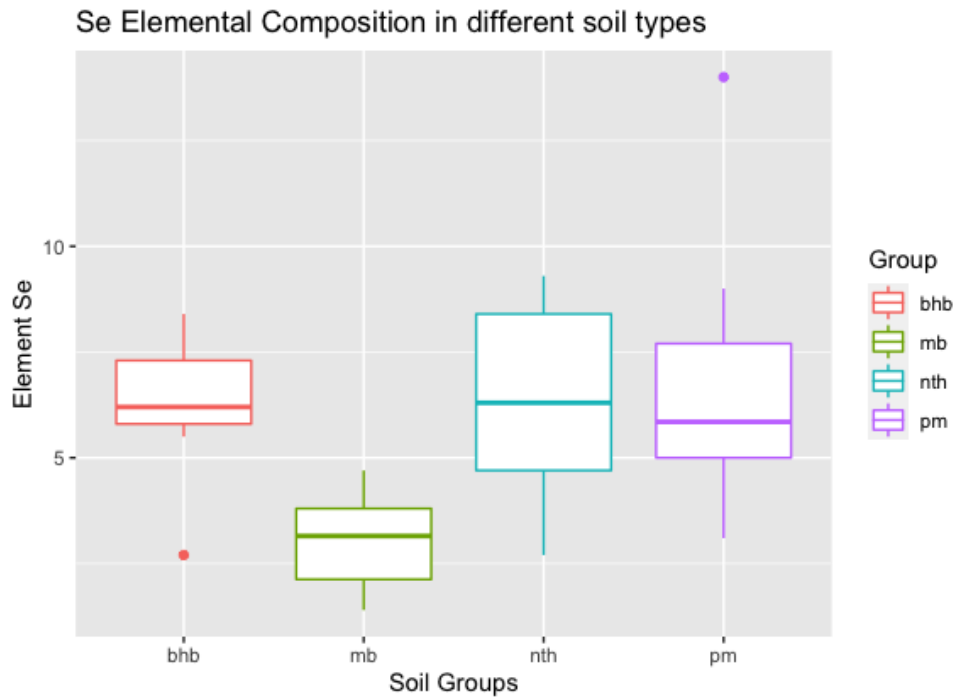
## GRAPHS-PLOTS

### BOXPLOTS [1-5]

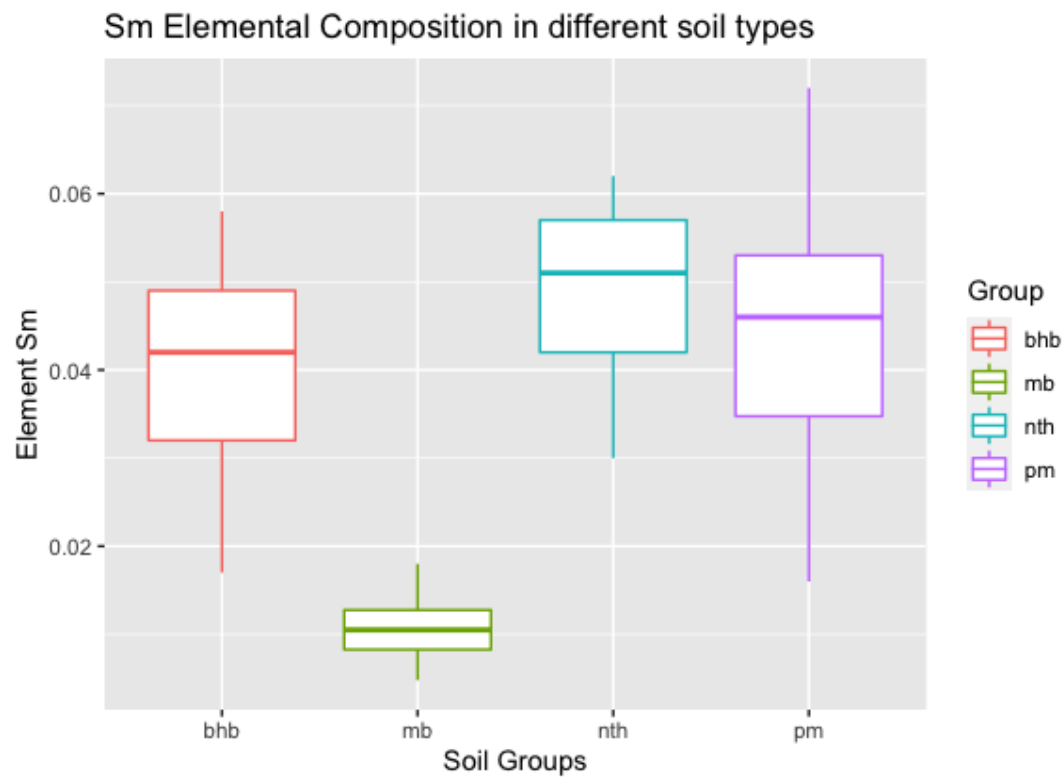
BOXPLOT[1]



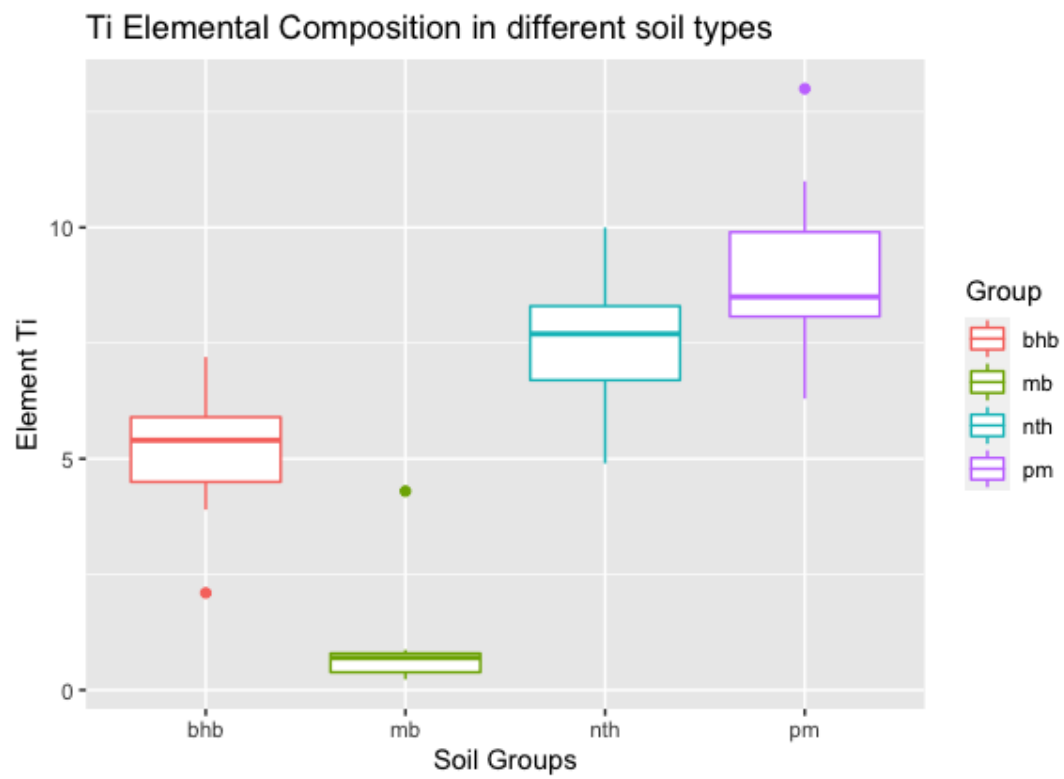
BOXPLOT[2]



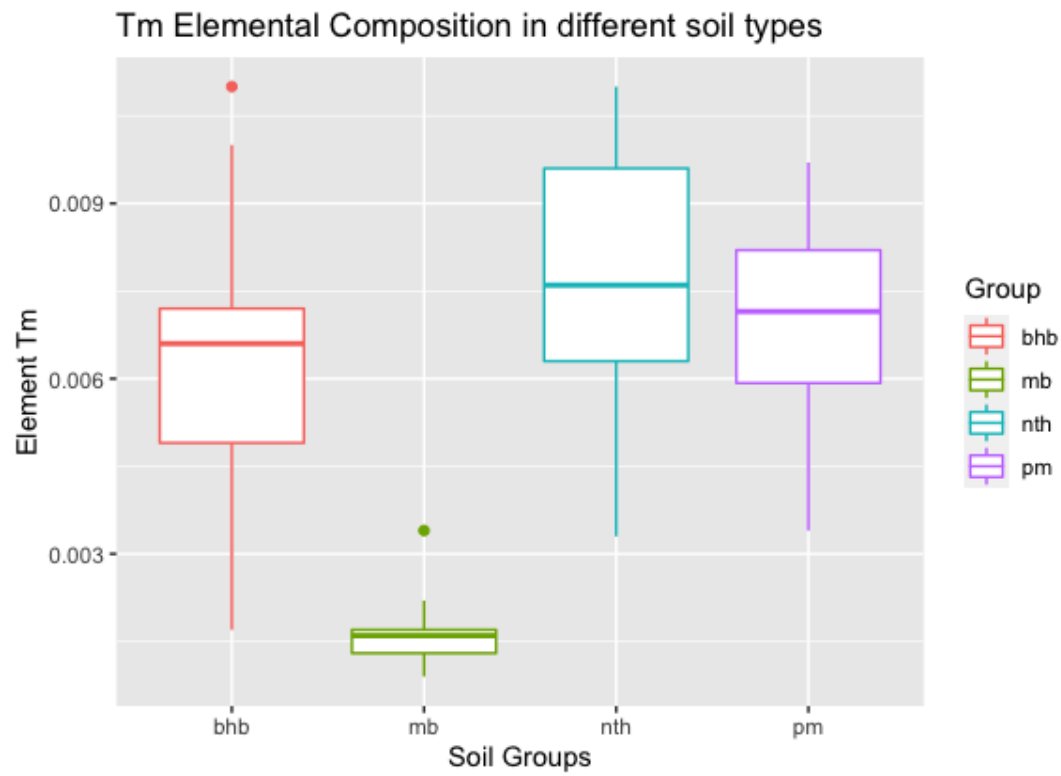
BOXPLOT[3]



BOXPLOT[4]

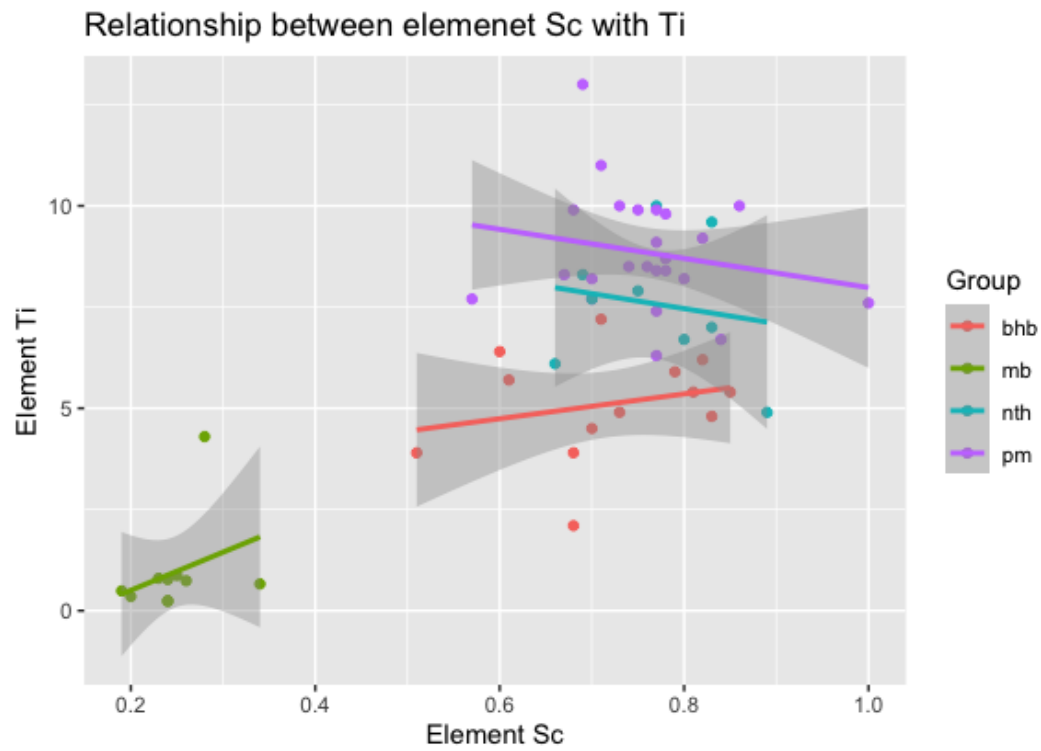


### BOXPLOT[5]

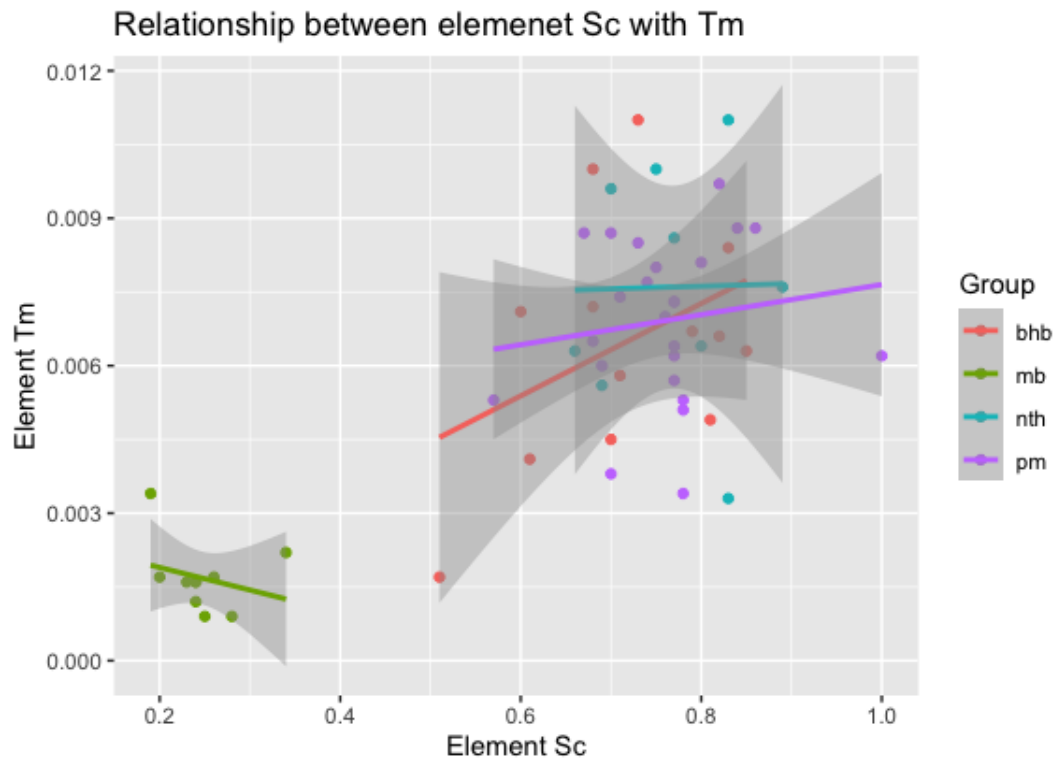


### GRAPH[1-5]

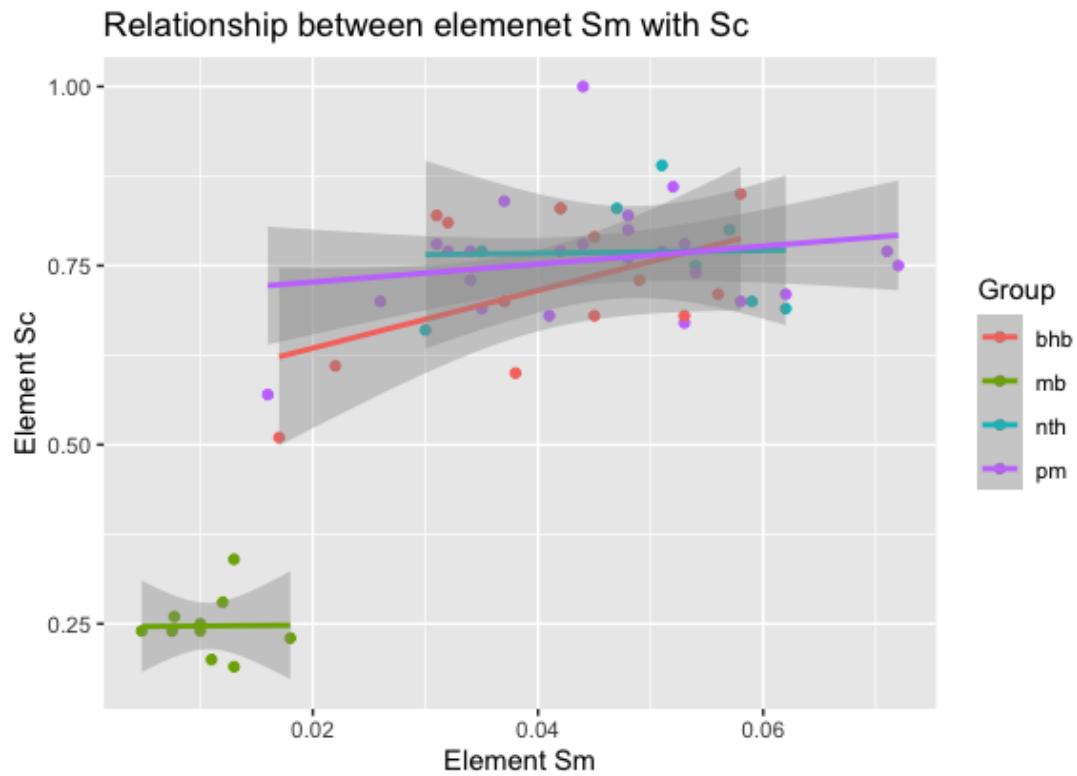
#### GRAPH[1]



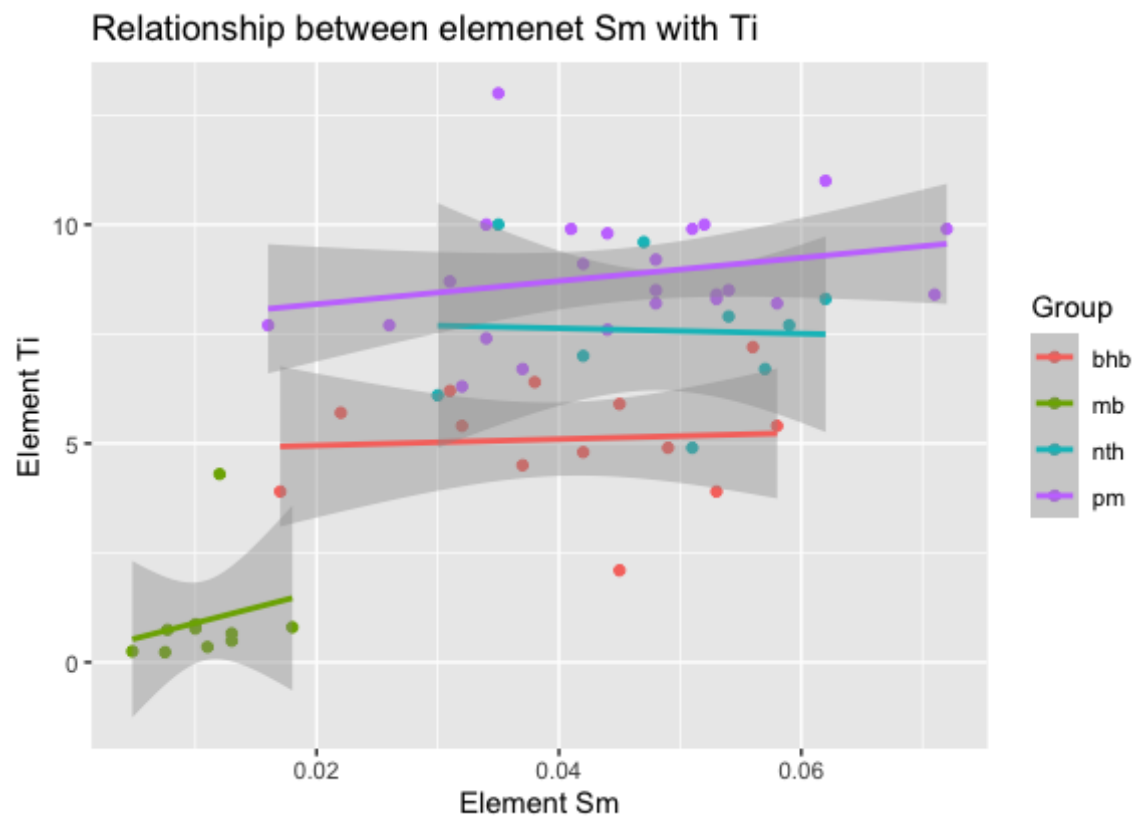
GRAPH[2]



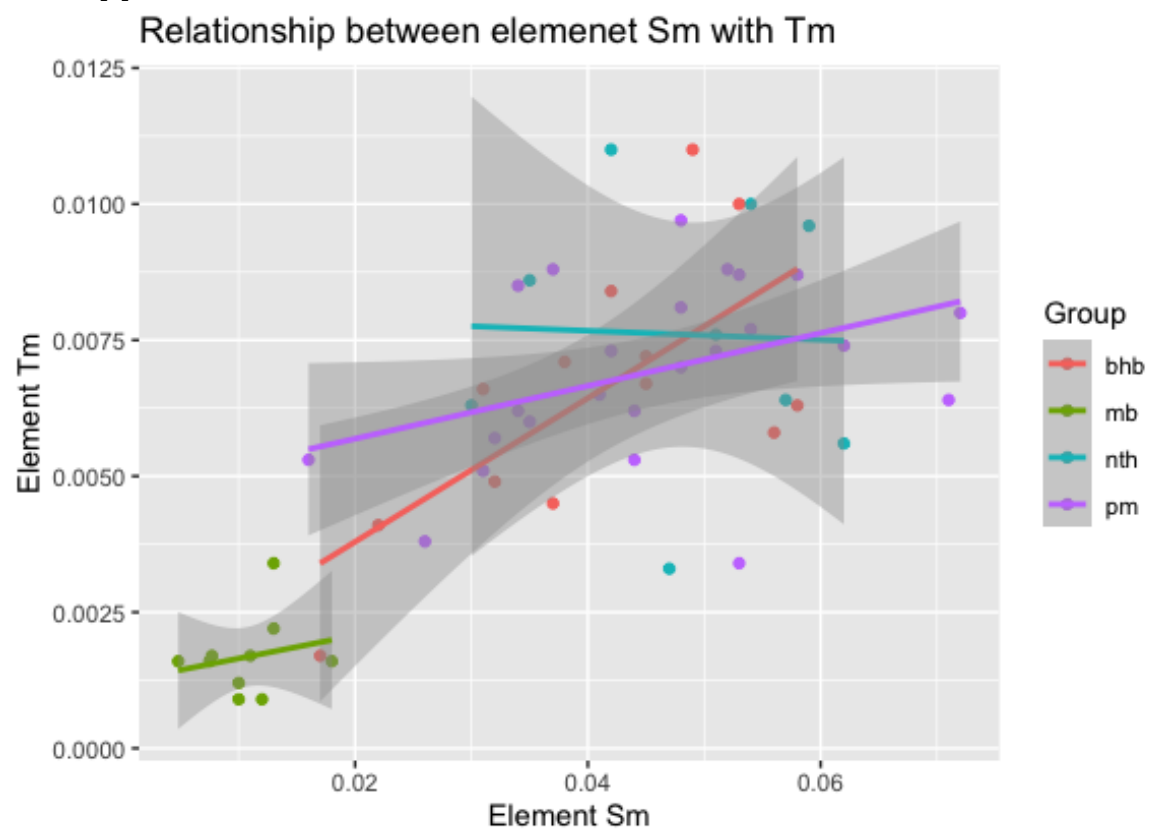
GRAPH[3]



GRAPH[4]



GRAPH[5]



## TABLES

TABLE[SHAPIRO]

| ELEMENT | P-VALUE | Ho = Reject/Accept (Accepting for p-value > 0.05) |
|---------|---------|---|
| Sc      | 0.30    | Large p-value. Accept                             |
| Ti      | 0.10    | Large p-value. Accept                             |
| Se      | 0.008   | Smaller p-value than 0.05. Reject                 |
| Sm      | 0.94    | Large p-value. Accept                             |
| Tm      | 0.81    | Large p-value. Accept                             |

TABLE[LEVENE]

| ELEMENT | P-VALUE | Ho = Reject/Accept (Accepting for p-value > 0.05) |
|---------|---------|---|
| Sc      | 0.13    | Large p-value. Accept                             |
| Ti      | 0.46    | Large p-value. Accept                             |
| Se      | 0.18    | Large p-value. Accept                             |
| Sm      | 0.03    | Smaller p-value than 0.05. Reject                 |
| Tm      | 0.02    | Smaller p-value than 0.05. Reject                 |

TABLE[ANOVA]

| ELEMENT | Pr(>F)       | Testing to P-Value 0.05 | Ho=Reject/Accept                         |
|---------|--------------|-------------------------|--|
| Sc      | -10.56       | Pr(>F) < P-Value        | P-Value small. Reject Ho                 |
| Ti      | -10.56       | Pr(>F) < P-Value        | P-Value small. Reject Ho                 |
| Se      | 0.0005       | Pr(>F) < P-Value        | P-Value small. Reject Ho (Kruskal Test). |
| Sm      | 4.912618e-12 | Pr(>F) < P-Value        | P-Value small. Reject Ho (Welch's Anova) |
| Tm      | 3.384352e-11 | Pr(>F) < P-Value        | P-Value small. Reject Ho (Welch's Anova) |

TABLES[MEAN1-5]

TABLE[MEAN(1)] – ELEMENT Sc

| ELEMENT Sc | DIFF        | LOWER       | UPPER       | RESULT(RANGES ABOVE ZERO?) |
|------------|-------------|-------------|-------------|----------------------------|
| mb-bhb     | -0.46992308 | -0.55963967 | -0.38020648 | FALSE = MEANS DIFFERENT    |
| nth-bhb    | 0.05196581  | -0.04052508 | 0.14445670  | TRUE = MEANS NOT DIFFERENT |
| pm-bhb     | 0.04182692  | -0.03162516 | 0.11527900  | TRUE = MEANS NOT DIFFERENT |
| nth-mb     | 0.52188889  | 0.42388663  | 0.61989115  | FALSE = MEANS DIFFERENT    |
| pm-mb      | 0.51175000  | 0.43146870  | 0.59203130  | FALSE = MEANS DIFFERENT    |
| pm-nth     | -0.01013889 | -0.09350905 | 0.07323127  | TRUE = MEANS NOT DIFFERENT |



**TABLE[MEAN(2)] – ELEMENT Ti**

| ELEMENT Ti | DIFF      | LOWER      | UPPER     | RESULT(RANGES ABOVE ZERO?) |
|------------|-----------|------------|-----------|----------------------------|
| mb-bhb     | -4.154000 | -5.7255426 | -2.582457 | FALSE = MEANS DIFFERENT    |
| nth-bhb    | 2.477778  | 0.8576386  | 4.097917  | FALSE = MEANS DIFFERENT    |
| pm-bhb     | 3.750000  | 2.4633587  | 5.036641  | FALSE = MEANS DIFFERENT    |
| nth-mb     | 6.631778  | 0.42388663 | 8.348458  | FALSE = MEANS DIFFERENT    |
| pm-mb      | 7.904000  | 6.4977331  | 9.310267  | FALSE = MEANS DIFFERENT    |
| pm-nth     | 1.272222  | -0.1881515 | 2.732596  | TRUE = MEANS NOT DIFFERENT |

**TABLE[MEAN(3)] – ELEMENT Se (Kruskall Test)**

| ELEMENT Ti | OBBS. DIFF  | CRITICAL.DIFF | RESULT DIFFERENCE |
|------------|-------------|---------------|-------------------|
| bhb-mb     | 24.93846154 | 18.09882      | TRUE              |
| bhb-nth    | 0.09401709  | 18.65848      | FALSE             |
| bhb-pm     | 2.49679487  | 14.81772      | FALSE             |
| mb-nth     | 24.84444444 | 19.77031      | TRUE              |
| mb-pm      | 22.44166667 | 16.19540      | TRUE              |
| nth-pm     | 2.40277778  | 16.81853      | FALSE             |

**TABLE[MEAN(4)] – ELEMENT Sm (Used Anova to identify difference. Results might not be the true outcome since it doesn't fill all the assumptions.)**

| ELEMENT Sm | DIFF         | LOWER        | UPPER        | RESULT(RANGES ABOVE ZERO?) |
|------------|--------------|--------------|--------------|----------------------------|
| mb-bhb     | -0.029684615 | -0.042790326 | -0.016578905 | FALSE = MEANS DIFFERENT    |
| nth-bhb    | 0.008170940  | -0.005340037 | 0.021681917  | TRUE = MEANS NOT DIFFERENT |
| pm-bhb     | 0.004865385  | -0.005864422 | 0.015595191  | TRUE = MEANS NOT DIFFERENT |
| nth-mb     | 0.037855556  | 0.023539483  | 0.052171628  | FALSE = MEANS DIFFERENT    |
| pm-mb      | 0.034550000  | 0.022822588  | 0.046277412  | FALSE = MEANS DIFFERENT    |
| pm-nth     | -0.003305556 | -0.015484185 | 0.008873074  | TRUE = MEANS NOT DIFFERENT |

**TABLE[MEAN(5)] – ELEMENT Tm (Used Anova to identify difference. Results might not be the true outcome since it doesn't fill all the assumptions.)**

| ELEMENT Tm | DIFF          | LOWER        | UPPER        | RESULT(RANGES ABOVE ZERO?) |
|------------|---------------|--------------|--------------|----------------------------|
| mb-bhb     | -0.0048046154 | -0.006923685 | -0.002685546 | FALSE = MEANS DIFFERENT    |
| nth-bhb    | 0.0011153846  | -0.001069212 | 0.003299981  | TRUE = MEANS NOT DIFFERENT |
| pm-bhb     | 0.0004278846  | -0.001307023 | 0.002162793  | TRUE = MEANS NOT DIFFERENT |
| nth-mb     | 0.0059200000  | 0.003605227  | 0.008234773  | FALSE = MEANS DIFFERENT    |
| pm-mb      | 0.0052325000  | 0.003336289  | 0.007128711  | FALSE = MEANS DIFFERENT    |
| pm-nth     | -0.0006875000 | -0.002656669 | 0.001281669  | TRUE = MEANS NOT DIFFERENT |

## TABLE[CORRELATION]

| PAIR OF ELEMENTS | VALUE OF 'p' | Relation                                       |
|------------------|--------------|--|
| Sc with Ti       | 0.757        | Positive Linear Relationship                   |
| Sm with Tm       | 0.733        | Positive Linear Relationship                   |
| Sm with Sc       | 0.765        | Positive Linear Relationship                   |
| Sm with Ti       | 0.675        | Positive Linear Relationship (could be better) |
| Sc with Tm       | 0.737        | Positive Linear Relationship                   |

## ARTICLE REFERENCES

- [1] K-12 Soil Science Teacher Resources, <https://www.soils4teachers.org/lessons-and-activities/teachers-guide/world-soils>
- [2] Does Weather Affect Plant Growth: Effect Of Temperature On Plants, author: Mary H. Dyer, <https://www.gardeningknowhow.com/plant-problems/environmental/temperature-on-plants.htm>
- [3] Organic vs. Chemical Cannabis Fertilizers for Your Grow, Trevor Hennings , December 6, 2016, <https://www.leafly.com/news/growing/organic-vs-chemical-nutrients-growing-cannabis>

## APPENDIX – CODE

```
# Importing package tidyverse to use powerful functions to clean data and create plots.
# Importing package car to use levene's test to check assumptions for ANalysis Of VAriance.
# You need to have all packages (tidyverse, car, pgirmess, onewaytests) installed to run the
# R-Script.
# Tidyverse import overrides two in built functions filter() and log()
  library(tidyverse)
# Car import overrides a dplyr function called recode.
  library(car)
# To use alternative Anova tests you need to download these 2 libraries.
  library(pgirmess)
  library(onewaytests)

# Function that implements the ANOVA test and identifies the different means for H1.
  applyAnovaTest <- function(cannabisData, element){
# Anova Testing
    ANOVA_result <- aov(element ~ Group, data = cannabisData)
# Outputting the results of the anova tests.
    print(summary(ANOVA_result))

# Function that identifies the means that are significantly different. Using Tukeys Method
    difference <- TukeyHSD(ANOVA_result)
# Prefer the table text to see results. Not plotting the data.
    print(difference)

# Return residual values to apply Shapiro Test later for assumption.
```

```

    return(ANOVA_result$residuals)

}

# Function that checks Constant Spread Assumption [Levene Test]. Variances are equal for
# ANOVA.
    applyLeveneTest <- function(cannabisData, element){
        leveneTest(element ~ Group, data = cannabisData)
    }

# Function that checks Normality of Data. [Shapiro-Wilk]. Data comes from Normal
# Distribution.
    applyShapiroTest <- function(residual){
        print(shapiro.test(residual))
    }

# Reading the data out of the CSV files.
    cannabisData <- read.csv("DATA/potplants_MT5762.csv")
# Selecting only the columns i am interested in.
    cannabisData <- cannabisData %>%
        select(Group, Sc, Ti, Se, Sm, Tm)

# Making the groups column a factor type for more control in program.
    cannabisData <- cannabisData %>%
        mutate_if(is.character, as.factor)

#### Commands needed to execute Question 1 and 2.
# Using tidyverse function to plot a boxplot for each element and compare
# the elementl composition in different soil types.
# Sc
    ggplot(cannabisData) +
        geom_boxplot(aes(x = Group, y = Sc, colour = Group)) +
        xlab("Soil Groups") +
        ylab("Element Sc") +
        ggtitle("Sc Elemental Composition in different soil types")
# Ti
    ggplot(cannabisData) +
        geom_boxplot(aes(x = Group, y = Ti, colour = Group)) +
        xlab("Soil Groups") +
        ylab("Element Ti") +
        ggtitle("Ti Elemental Composition in different soil types")
# Se
    ggplot(cannabisData) +
        geom_boxplot(aes(x = Group, y = Se, colour = Group)) +
        xlab("Soil Groups") +
        ylab("Element Se") +
        ggtitle("Se Elemental Composition in different soil types")
# Sm
    ggplot(cannabisData) +

```

```

    geom_boxplot(aes(x = Group, y = Sm, colour = Group)) +
    xlab("Soil Groups") +
    ylab("Element Sm") +
    ggtitle("Sm Elemental Composition in different soil types")
# Tm
    ggplot(cannabisData) +
    geom_boxplot(aes(x = Group, y = Tm, colour = Group)) +
    xlab("Soil Groups") +
    ylab("Element Tm") +
    ggtitle("Tm Elemental Composition in different soil types")

#### Commands needed to execute Question 2.
# Calling Self-declared function to apply ANOVA TEST for each element.
# Once Anova Called we test our assumptions using the Anova residuals to make
# sure our results are at the best level they can be.
# applyLeveneTest = Self-declared function to apply Levene Test and test Constant Spread
# Assumption.
# applyShapiroTest = Self-declared function to apply the Shapiro-Wilk Test, testing
# Normality.
# Sc
    applyLeveneTest(cannabisData, cannabisData$Sc)
    Sc.residuals <- applyAnovaTest(cannabisData, cannabisData$Sc)
    applyShapiroTest(Sc.residuals)
# Ti
    applyLeveneTest(cannabisData, cannabisData$Ti)
    Ti.residuals <- applyAnovaTest(cannabisData, cannabisData$Ti)
    applyShapiroTest(Ti.residuals)

# Cannot Apply Anova test, because it does not follow normality but has equal
# standard deviation we can use Kruskal test on Se. Kruskal assumes
# equal standard deviation.
# Se
    applyLeveneTest(cannabisData, cannabisData$Se)
    Se.residual_anova <- applyAnovaTest(cannabisData, cannabisData$Se)
    applyShapiroTest(Se.residual_anova)
# Fails the shapiro test, indicates does not follow normality. Use Kruskal's
    Se.residuals_kruskal <- kruskal.test(Se ~ Group, data=cannabisData)
    kruskalmc(Se ~ Group, data=cannabisData)

# Cannot Apply Anova test because data does not have constant spread in there groups of
# data.
# Resulting we need to use an alternative Anova test that does not assumes equal
# standard deviations. We will use the Welch's Anova test and we have normality here.
# Sm
    applyLeveneTest(cannabisData, cannabisData$Sm)
# Fails constant spread test. So we check for normality
    Sm.residuals_anova <- applyAnovaTest(cannabisData, cannabisData$Sm)
    applyShapiroTest(Sm.residuals_anova)
# Passes normality. Use welch's Test

```

```

# Fails normality assumption and so we use welch's Anova
Sm.residuals_welch <- welch.test(Sm ~ Group, data=cannabisData)

# Tm
  applyLeveneTest(cannabisData, cannabisData$Tm)
# Fails constant spread test. So we check for normality
  Tm.residuals_anova <- applyAnovaTest(cannabisData, cannabisData$Tm)
  applyShapiroTest(Tm.residuals_anova)
# Passes normality. Use welch's Test
# Fails normality assumption and so we use welch's Anova
  Tm.residuals_welch <- welch.test(Tm ~ Group, data=cannabisData)

```

#### Commands needed to execute Question 3.

# calling R function cor(), finds the correlation coefficient  $\rho$  which  
# which indicates if two datasets are linear related.

# Sc with Ti

```

cor(cannabisData$Sc, cannabisData$Ti)
ggplot(cannabisData,aes(x = Sc, y = Ti, colour = Group)) +
  geom_point() +
  stat_smooth(method = ("lm")) +
  xlab("Element Sc") +
  ylab("Element Ti") +
  ggtitle("Relationship between elemenet Sc with Ti")

```

# Sm with Tm

```

cor(cannabisData$Sm, cannabisData$Tm)
ggplot(cannabisData,aes(x = Sm, y = Tm, colour = Group)) +
  geom_point() +
  stat_smooth(method = ("lm")) +
  xlab("Element Sm") +
  ylab("Element Tm") +
  ggtitle("Relationship between elemenet Sm with Tm")

```

# Sm with Sc

```

cor(cannabisData$Sm, cannabisData$Sc)
ggplot(cannabisData,aes(x = Sm, y = Sc, colour = Group)) +
  geom_point() +
  stat_smooth(method = ("lm")) +
  xlab("Element Sm") +
  ylab("Element Sc") +
  ggtitle("Relationship between elemenet Sm with Sc")

```

# Sm with Ti

```

cor(cannabisData$Sm, cannabisData$Ti)
ggplot(cannabisData,aes(x = Sm, y = Ti, colour = Group)) +
  geom_point() +
  stat_smooth(method = ("lm")) +
  xlab("Element Sm") +
  ylab("Element Ti") +
  ggtitle("Relationship between elemenet Sm with Ti")

```

# Sc with Tm

```

cor(cannabisData$Sc, cannabisData$Tm)

```

```
ggplot(cannabisData,aes(x = Sc, y = Tm, colour = Group)) +  
  geom_point() +  
  stat_smooth(method = "lm") +  
  xlab("Element Sc") +  
  ylab("Element Tm") +  
  ggtitle("Relationship between elemenet Sc with Tm")
```