Math 189: Midterm Project 2

Romano-British Pottery Data

Working alone, prepare a R Markdown Notebook report based on examining the Romano-British Pottery dataset (available from GitHub). The dataset contains measurements on pottery shards that were collected from five sites in the British Isles. Note: in class we focused on four of the sites, corresponding to Kiln equaling 2, 3, 4, or 5. Here you will also study Kiln site 1 (which corresponds to "Gloucester"). The dataset contains 48 observations on 9 chemical variables:

1. Al2O3: aluminium trioxide
2. Fe2O3: iron trioxide
3. MgO: magnesium oxide
4. CaO: calcium oxide
5. Na2O: natrium oxide
6. K2O: kalium oxide
7. TiO2: titanium oxide
8. MnO: mangan oxide
9. BaO: barium oxide

Introduction

In this assignment, we would like to find out if the group means for 5 different locations according to 9 variables (elemental oxides) are the same. This, in fact would be our null hypothesis. For ANOVA (univariate analysis of variance), we would have 9 different null hypotheses for each of the elements. These would each be that $\mu^{(1)} = \mu^{(2)} = \mu^{(3)} = \mu^{(4)} = \mu^{(5)}$. The alternative would be that at least one of sample means differ. For MANOVA (multivariate analysis of variance) this would be a single null hypothesis including all 9 variables. This would be that $\underline{\mu}^{(1)} = \underline{\mu}^{(2)} = \underline{\mu}^{(3)} = \underline{\mu}^{(4)} = \underline{\mu}^{(5)}$. The alternative would be that at least one of these group means differs. We first demonstrate the normality and means using histogram graphs. We should use normality tentatively throughout this entire assignment. Because the samples are so small, normality and the variances are somewhat off. However, we then use both ANOVA (to demonstrate ability) and then mainly MANOVA to check whether or not the group means are equal. As well, for simplicity the locations are numbered; however, know that location 1 is Gloucester, location 2 is LLanedeyrn, location 3 is Caldicot, location 4 is Isle Thorns, and location 5 is Ashley Rails.

Data

The data displays different chemical compositions (based on 9 oxides) of Romano-British pottery from 5 different locations. This data comes from A. Tubb and N. J. Parker and G. Nickless (1980), The analysis of Romano-British pottery by atomic absorption spectrophotometry. Archaeometry, 22, 153–171. The data includes 48 observations: (22 from location 1, 14 from location 2, 2 from location 3, 5 from location 4, and 5 from location 5). The 9 elements are Aluminum, Iron, Magnesium, Calcium, Sodium, Potassium, Titanium, Manganese, and Barium (all oxides).

Prompt

Is there a significant difference among the 5 group means for these 9 variables? Formulate the appropriate null hypothesis; choose a method of analysis (from the techniques discussed in class, using material up through Lecture 12); discuss the assumptions needed (and justify those assumptions, if possible); perform the analysis using R code (if you use a package, it is your responsibility to ensure their code does what you intend it to do); summarize the results with appropriate displays (figures and/or tables).

##Methods and Analysis

Like in the first project, I prefer to explain this above each block of r-code for clarity and because one block of text would look unprofessional.

Here, we will separate the 5 locations based on the Kiln number. We will also omit the first 3 columns of the table for each of the locations, as these are just identifying numbers and will not be used for this project.

```
"C:\\Users\\Mark's PC\\Desktop\\Math 189\\ma189-main\\Data\\RBPottery.csv" =
  paste0(getwd(), "/RBPottery.csv")
pot = read.csv("C:\\Users\\Mark's PC\\Desktop\\Math 189\\ma189-main\\Data\\RBPottery.csv",
               header = TRUE)
# Please remember that these chemicals are oxides, we will be dropping the O for simplicity.
colnames(pot) = c("No","ID","Kiln","Al","Fe","Mg","Ca","Na","K","Ti","Mn","Ba")
loc1 = pot[pot$Kiln == 1,4:12]
loc2 = pot[pot$Kiln == 2,4:12]
loc3 = pot[pot$Kiln == 3,4:12]
loc4 = pot[pot$Kiln == 4,4:12]
loc5 = pot[pot$Kiln == 5,4:12]
loc1
```

```
##      Al   Fe   Mg   Ca   Na    K   Ti    Mn    Ba
## 1  18.8 9.52 2.00 0.79 0.40 3.20 1.01 0.077 0.015
## 2  16.9 7.33 1.65 0.84 0.40 3.05 0.99 0.067 0.018
## 3  18.2 7.64 1.82 0.77 0.40 3.07 0.98 0.087 0.014
## 4  17.4 7.48 1.71 1.01 0.40 3.16 0.03 0.084 0.017
## 5  16.9 7.29 1.56 0.76 0.40 3.05 1.00 0.063 0.019
## 6  17.8 7.24 1.83 0.92 0.43 3.12 0.93 0.061 0.019
## 7  18.8 7.45 2.06 0.87 0.25 3.26 0.98 0.072 0.017
## 8  16.5 7.05 1.81 1.73 0.33 3.20 0.95 0.066 0.019
## 9  18.0 7.42 2.06 1.00 0.28 3.37 0.96 0.072 0.017
## 10 15.8 7.15 1.62 0.71 0.38 3.25 0.93 0.062 0.017
## 11 14.6 6.87 1.67 0.76 0.33 3.06 0.91 0.055 0.012
## 12 13.7 5.83 1.50 0.66 0.13 2.25 0.75 0.034 0.012
## 13 14.6 6.76 1.63 1.48 0.20 3.02 0.87 0.055 0.016
## 14 14.8 7.07 1.62 1.44 0.24 3.03 0.86 0.080 0.016
## 15 17.1 7.79 1.99 0.83 0.46 3.13 0.93 0.090 0.020
## 16 16.8 7.86 1.86 0.84 0.46 2.93 0.94 0.094 0.020
## 17 15.8 7.65 1.94 0.81 0.83 3.33 0.96 0.112 0.019
## 18 18.6 7.85 2.33 0.87 0.38 3.17 0.98 0.081 0.018
## 19 16.9 7.87 1.83 1.31 0.53 3.09 0.95 0.092 0.023
## 20 18.9 7.58 2.05 0.83 0.13 3.29 0.98 0.072 0.015
## 21 18.0 7.50 1.94 0.69 0.12 3.14 0.93 0.035 0.017
## 22 17.8 7.28 1.92 0.81 0.18 3.15 0.90 0.067 0.017
```

```
loc2
```

```
##      Al   Fe   Mg   Ca   Na    K   Ti    Mn    Ba
## 23 14.4 7.00 4.30 0.15 0.51 4.25 0.79 0.160 0.019
## 24 13.8 7.08 3.43 0.12 0.17 4.14 0.77 0.144 0.020
## 25 14.6 7.09 3.88 0.13 0.20 4.36 0.81 0.124 0.019
## 26 11.5 6.37 5.64 0.16 0.14 3.89 0.69 0.087 0.009
## 27 13.8 7.06 5.34 0.20 0.20 4.31 0.71 0.101 0.021
## 28 10.9 6.26 3.47 0.17 0.22 3.40 0.66 0.109 0.010
## 29 10.1 4.26 4.26 0.20 0.18 3.32 0.59 0.149 0.017
## 30 11.6 5.78 5.91 0.18 0.16 3.70 0.65 0.082 0.015
## 31 11.1 5.49 4.52 0.29 0.30 4.03 0.63 0.080 0.016
## 32 13.4 6.92 7.23 0.28 0.20 4.54 0.69 0.163 0.017
## 33 12.4 6.13 5.69 0.22 0.54 4.65 0.70 0.159 0.015
## 34 13.1 6.64 5.51 0.31 0.24 4.89 0.72 0.094 0.017
## 35 12.7 6.69 4.45 0.20 0.22 4.70 0.73 0.394 0.024
## 36 12.5 6.44 3.94 0.22 0.23 0.81 0.75 0.177 0.019
```

```
loc3
```

```
##      Al   Fe   Mg   Ca   Na    K   Ti    Mn    Ba
## 37 11.6 5.39 3.77 0.29 0.06 4.51 0.56 0.110 0.015
## 38 11.8 5.44 3.94 0.30 0.04 4.64 0.59 0.085 0.013
```

```
loc4
```

```
##      Al   Fe   Mg   Ca   Na    K   Ti    Mn    Ba
## 39 18.3 1.28 0.67 0.03 0.03 1.96 0.65 0.001 0.014
## 40 15.8 2.39 0.63 0.01 0.04 1.94 1.29 0.001 0.014
## 41 18.0 1.50 0.67 0.01 0.06 2.11 0.92 0.001 0.016
## 42 18.0 1.88 0.68 0.01 0.04 2.00 1.11 0.006 0.022
## 43 20.8 1.51 0.72 0.07 0.10 2.37 1.26 0.002 0.016
```

```
loc5
```

```
##      Al   Fe   Mg   Ca   Na    K   Ti    Mn    Ba
## 44 17.7 1.12 0.56 0.06 0.06 2.06 0.79 0.001 0.013
## 45 18.3 1.14 0.67 0.06 0.05 2.11 0.89 0.006 0.019
## 46 16.7 0.92 0.53 0.01 0.05 1.76 0.91 0.004 0.013
## 47 14.8 2.74 0.67 0.03 0.05 2.15 1.34 0.003 0.015
## 48 19.1 1.64 0.60 0.10 0.03 1.75 1.04 0.007 0.018
```

We will first display mean and variance for each location according to each element as is good practice, and these will be used later as well. We then bind these into columns ie. the output will look like: row = element, column = location number.

```
cbind(mean(loc1$Al), mean(loc2$Al), mean(loc3$Al), mean(loc4$Al), mean(loc5$Al))
```

```
##          [,1]     [,2] [,3]  [,4]  [,5]
## [1,] 16.94091 12.56429 11.7 18.18 17.32
```

```r
cbind(mean(loc1$Fe), mean(loc2$Fe), mean(loc3$Fe), mean(loc4$Fe), mean(loc5$Fe))
```

```
##          [,1]     [,2]  [,3]  [,4]  [,5]
## [1,] 7.430909 6.372143 5.415 1.712 1.512
```

```r
cbind(mean(loc1$Mg), mean(loc2$Mg), mean(loc3$Mg), mean(loc4$Mg), mean(loc5$Mg))
```

```
##          [,1]     [,2]  [,3]  [,4]  [,5]
## [1,] 1.836364 4.826429 3.855 0.674 0.606
```

```r
cbind(mean(loc1$Ca), mean(loc2$Ca), mean(loc3$Ca), mean(loc4$Ca), mean(loc5$Ca))
```

```
##           [,1]      [,2]  [,3]  [,4]  [,5]
## [1,] 0.9422727 0.2021429 0.295 0.026 0.052
```

```r
cbind(mean(loc1$Na), mean(loc2$Na), mean(loc3$Na), mean(loc4$Na), mean(loc5$Na))
```

```
##           [,1]      [,2] [,3]  [,4]  [,5]
## [1,] 0.3481818 0.2507143 0.05 0.054 0.048
```

```r
cbind(mean(loc1$K), mean(loc2$K), mean(loc3$K), mean(loc4$K), mean(loc5$K))
```

```
##          [,1]     [,2]  [,3]  [,4]  [,5]
## [1,] 3.105455 3.927857 4.575 2.076 1.966
```

```r
cbind(mean(loc1$Ti), mean(loc2$Ti), mean(loc3$Ti), mean(loc4$Ti), mean(loc5$Ti))
```

```
##           [,1]      [,2]  [,3]  [,4]  [,5]
## [1,] 0.8963636 0.7064286 0.575 1.046 0.994
```

```r
cbind(mean(loc1$Mn), mean(loc2$Mn), mean(loc3$Mn), mean(loc4$Mn), mean(loc5$Mn))
```

```
##            [,1]   [,2]   [,3]   [,4]   [,5]
## [1,] 0.07172727 0.1445 0.0975 0.0022 0.0042
```

```r
cbind(mean(loc1$Ba), mean(loc2$Ba), mean(loc3$Ba), mean(loc4$Ba), mean(loc5$Ba))
```

```
##            [,1]  [,2]  [,3]   [,4]   [,5]
## [1,] 0.01713636 0.017 0.014 0.0164 0.0156
```

```r
cbind(var(loc1$Al), var(loc2$Al), var(loc3$Al), var(loc4$Al), var(loc5$Al))
```

```
##         [,1]     [,2] [,3]  [,4]  [,5]
## [1,] 2.28158 1.896319 0.02 3.152 2.752
```

```
cbind(var(loc1$Fe), var(loc2$Fe), var(loc3$Fe), var(loc4$Fe), var(loc5$Fe))
```

```
##           [,1]      [,2]    [,3]    [,4]    [,5]
## [1,] 0.4256468 0.6171104 0.00125 0.19007 0.54172
```

```
cbind(var(loc1$Mg), var(loc2$Mg), var(loc3$Mg), var(loc4$Mg), var(loc5$Mg))
```

```
##            [,1]     [,2]    [,3]    [,4]    [,5]
## [1,] 0.04161472 1.184225 0.01445 0.00103 0.00403
```

```
cbind(var(loc1$Ca), var(loc2$Ca), var(loc3$Ca), var(loc4$Ca), var(loc5$Ca))
```

```
##            [,1]        [,2]  [,3]    [,4]    [,5]
## [1,] 0.08138983 0.003387363 5e-05 0.00068 0.00117
```

```
cbind(var(loc1$Na), var(loc2$Na), var(loc3$Na), var(loc4$Na), var(loc5$Na))
```

```
##            [,1]       [,2]  [,3]    [,4]    [,5]
## [1,] 0.02554892 0.01503791 2e-04 0.00078 0.00012
```

```
cbind(var(loc1$K), var(loc2$K), var(loc3$K), var(loc4$K), var(loc5$K))
```

```
##            [,1]     [,2]    [,3]    [,4]    [,5]
## [1,] 0.04826407 1.025587 0.00845 0.03133 0.03813
```

```
cbind(var(loc1$Ti), var(loc2$Ti), var(loc3$Ti), var(loc4$Ti), var(loc5$Ti))
```

```
##            [,1]        [,2]    [,3]    [,4]    [,5]
## [1,] 0.04070996 0.003824725 0.00045 0.07053 0.04533
```

```
cbind(var(loc1$Mn), var(loc2$Mn), var(loc3$Mn), var(loc4$Mn), var(loc5$Mn))
```

```
##              [,1]        [,2]      [,3]    [,4]    [,5]
## [1,] 0.0003392554 0.006278115 0.0003125 4.7e-06 5.7e-06
```

```
cbind(var(loc1$Ba), var(loc2$Ba), var(loc3$Ba), var(loc4$Ba), var(loc5$Ba))
```

```
##              [,1]    [,2]  [,3]    [,4]    [,5]
## [1,] 6.694805e-06 1.6e-05 2e-06 1.08e-05 7.8e-06
```

These are assumptions we must make for ANOVA and MANOVA. Where there is the ability to prove these, we will. However, once again, remember the sample sizes are extremely small, so were we to go further and do var.test or qqnorm for every group-element, many of these would look incorrect.

Assumptions

1. The data from group k has common mean μ^k We may simply assume this. We could take separate samples from the groups in an attempt to check that the means would be equivalent. However, the sample size is too small, and this method wouldn't be completely sufficient.

2. Homoskedasticity: The data from all groups have common variance. One method would be to test the variances together using var.test. We will demonstrate one of those tests here. The test for MANOVA is much more rigorous as it's for multivariate analysis and will simply be assumed here. As well, because we assume these variables for each location are normally distributed, we would assume they would have common variance. Please note, some of these tests(using var.test) would fail and we would reject the null hypothesis that the variances are equivalent. However, this is because the sample sizes are extremely small.

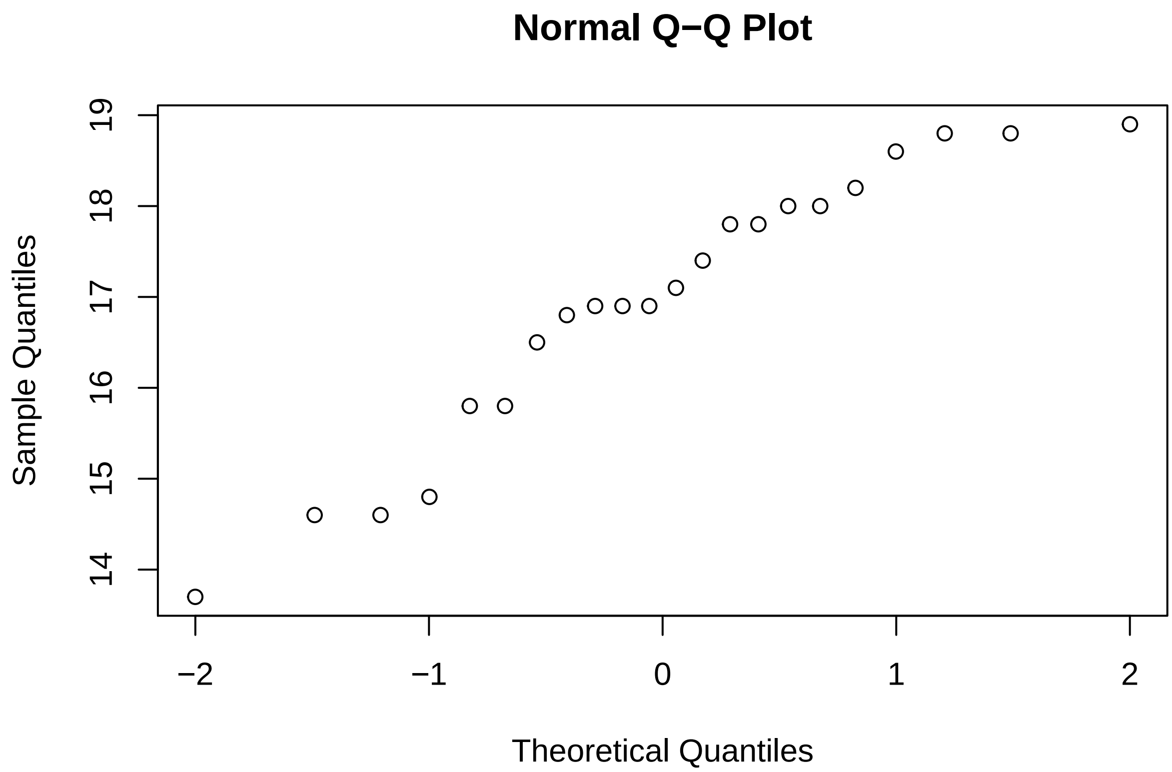
```r
var.test(loc1$Al,loc2$Al,alternative = "two.sided")
```

```
##
##  F test to compare two variances
##
## data:  loc1$Al and loc2$Al
## F = 1.2032, num df = 21, denom df = 13, p-value = 0.7469
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.4103211 3.1255607
## sample estimates:
## ratio of variances
##           1.203163
```

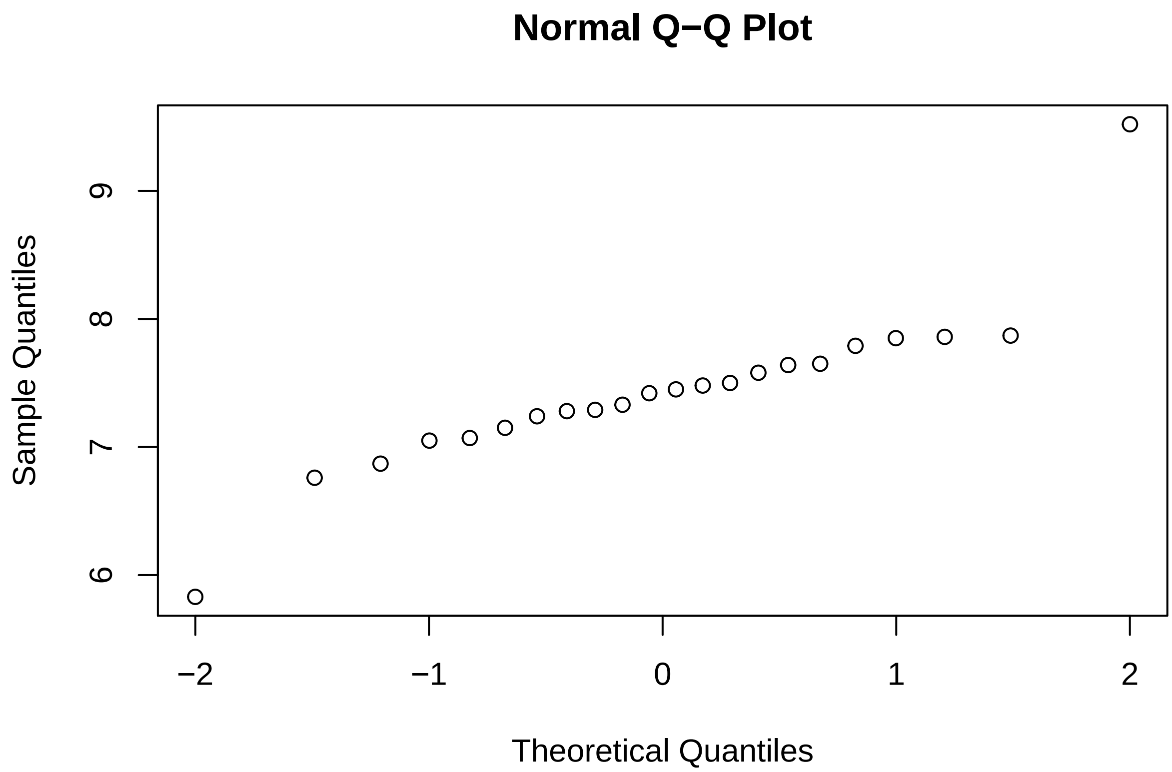
3. The data from each group is independently sampled. This is simply assumed, as trying to calculate independence is fairly rigorous as well.

4. The data are normally distributed. For ANOVA, this would simply be proven by using qqnorm, and as long as the data is relatively linear, we can assume normality. We will demonstrate this for location 1, the rest will be assumed for redundancy. However, for MANOVA, the process is much more difficult to calculate, and we will simply assume this is true.

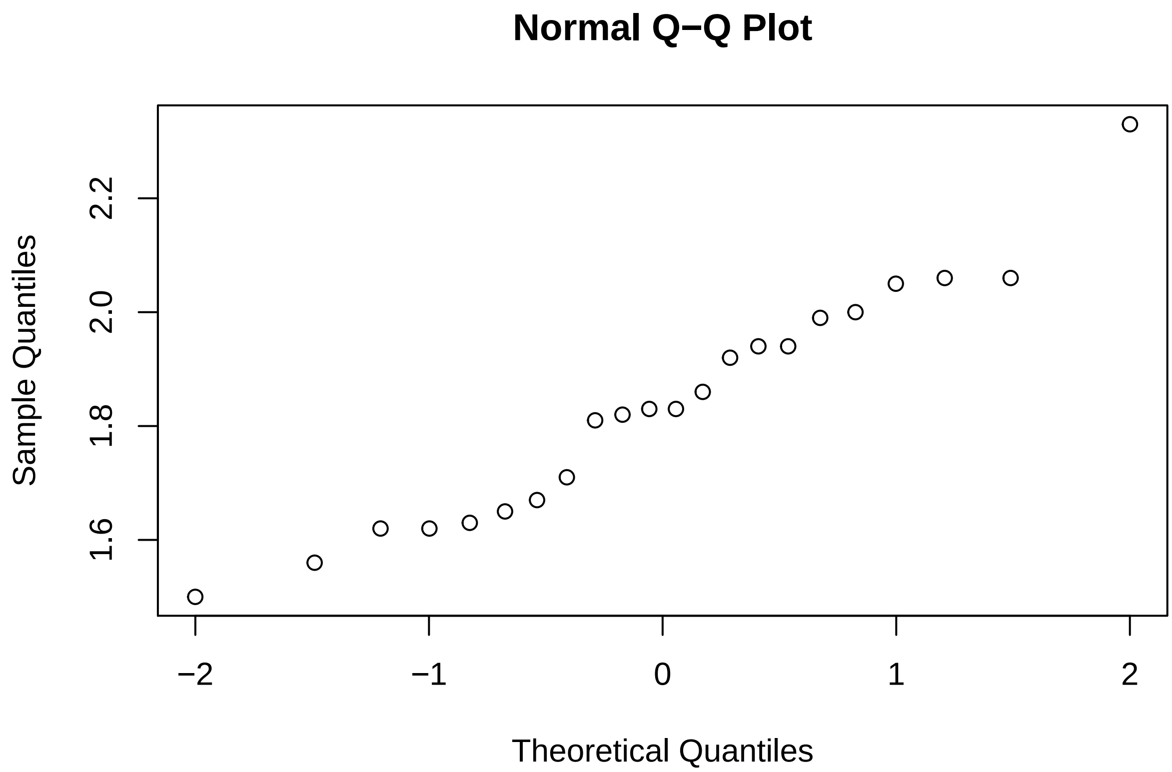
```r
qqnorm(loc1$Al)
```



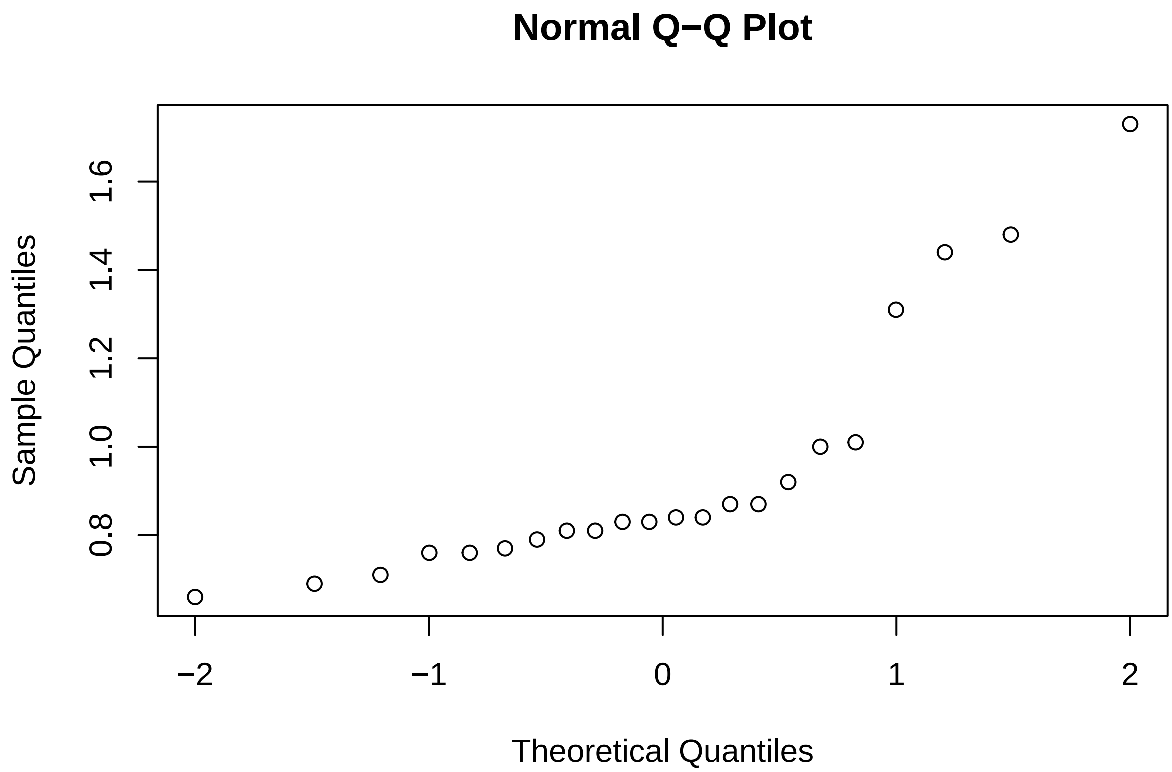

```r
qqnorm(loc1$Fe)
```



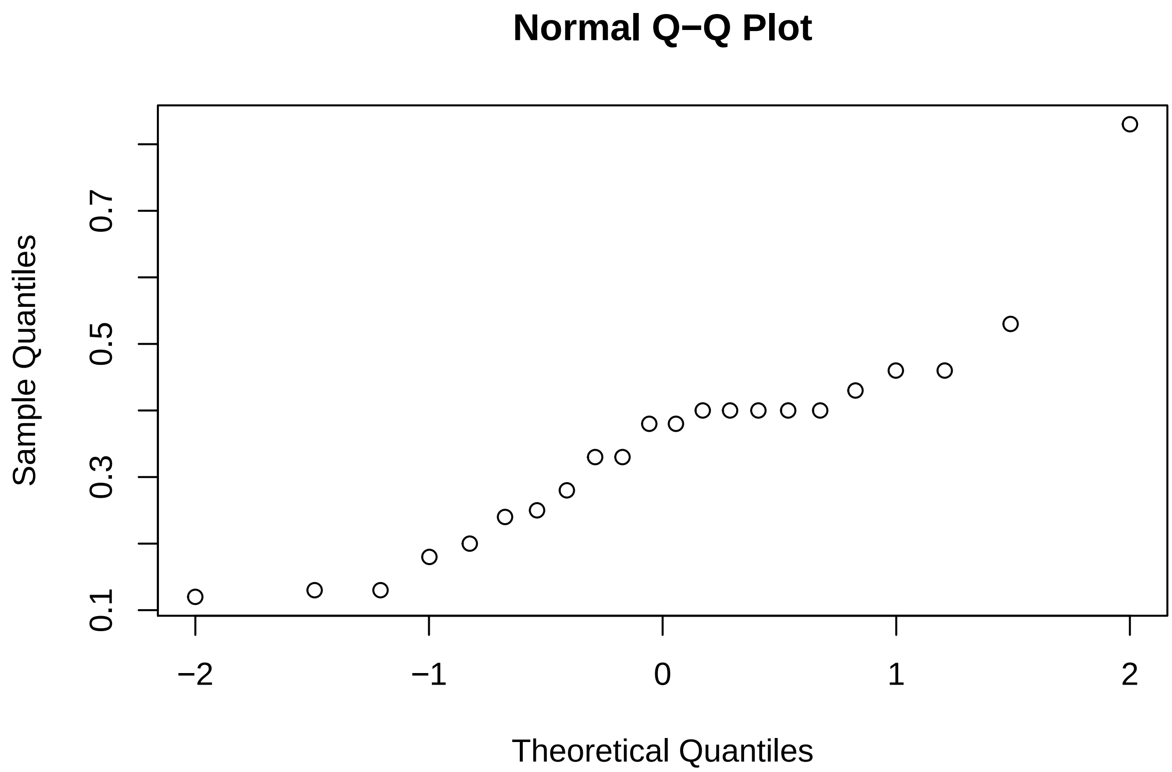

```
qqnorm(loc1$Mg)
```



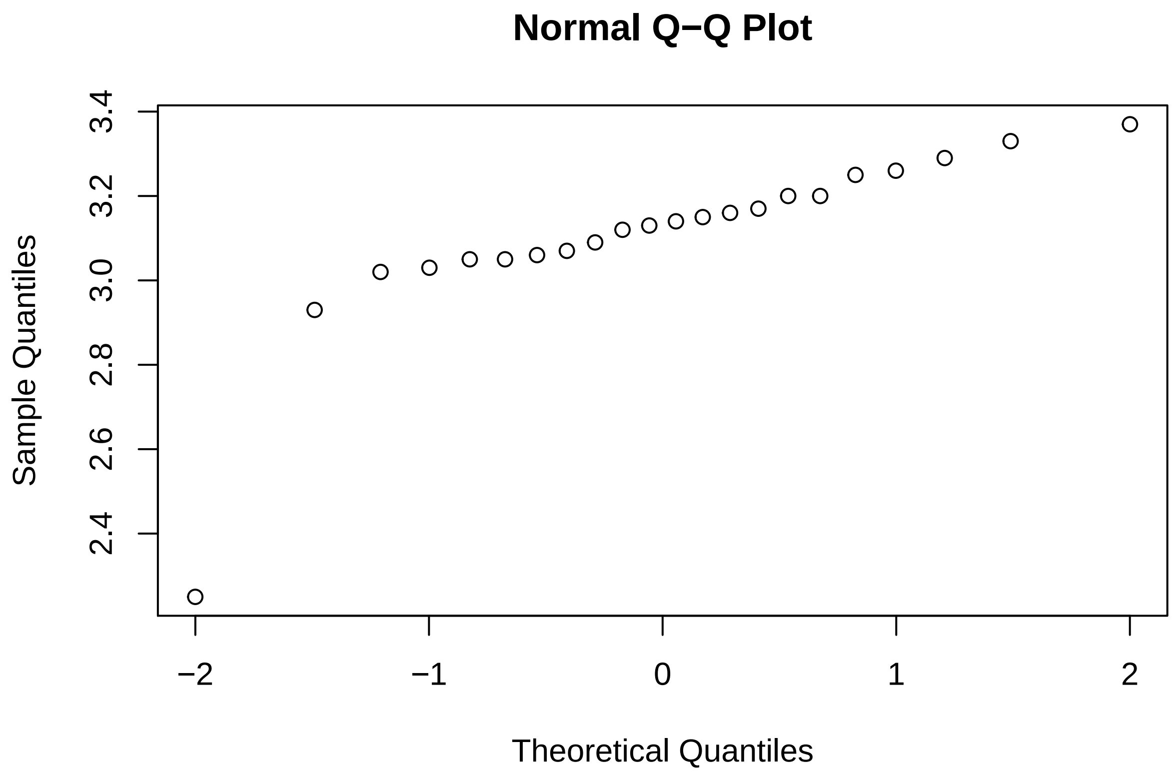

```
qqnorm(loc1$Ca)
```



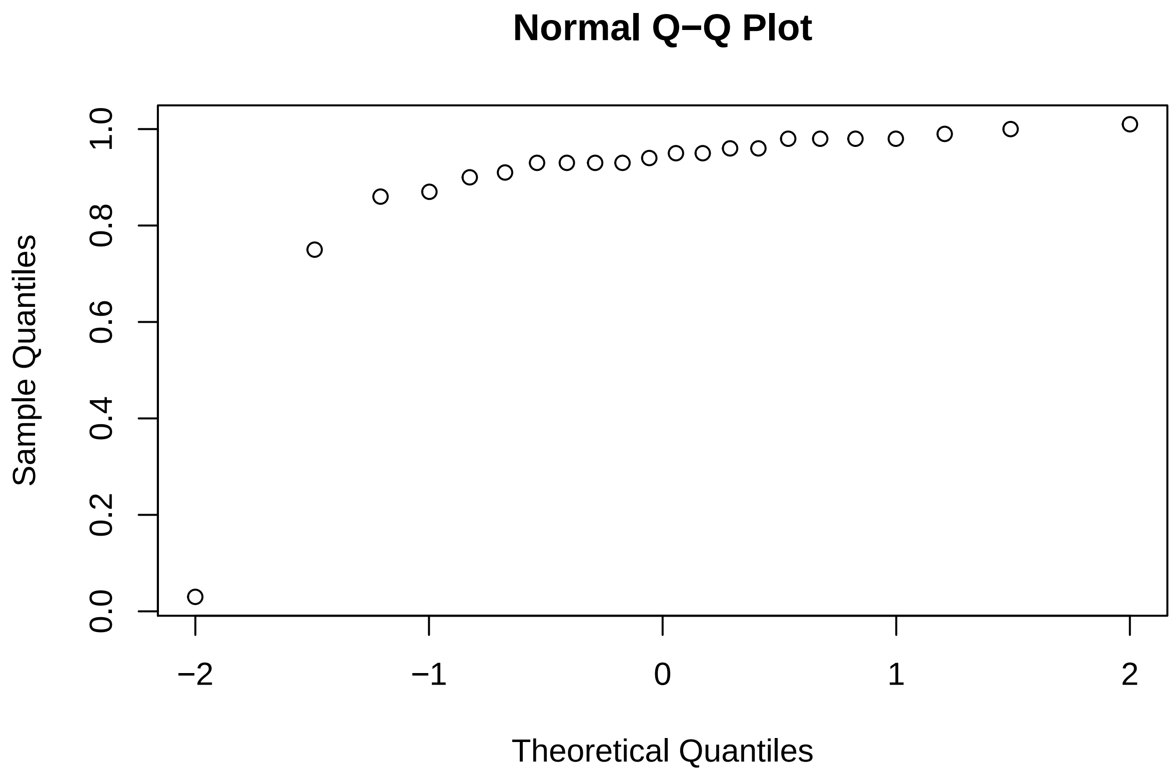

```
qqnorm(loc1$Na)
```



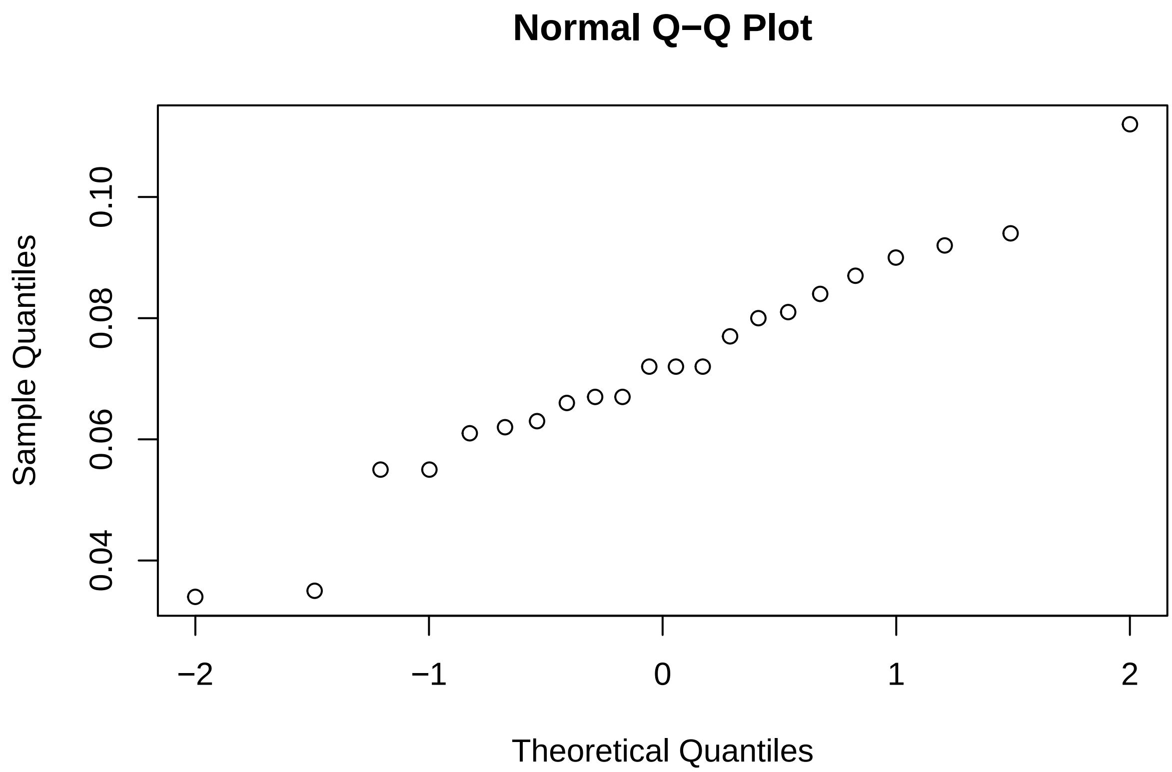

```r
qqnorm(loc1$K)
```



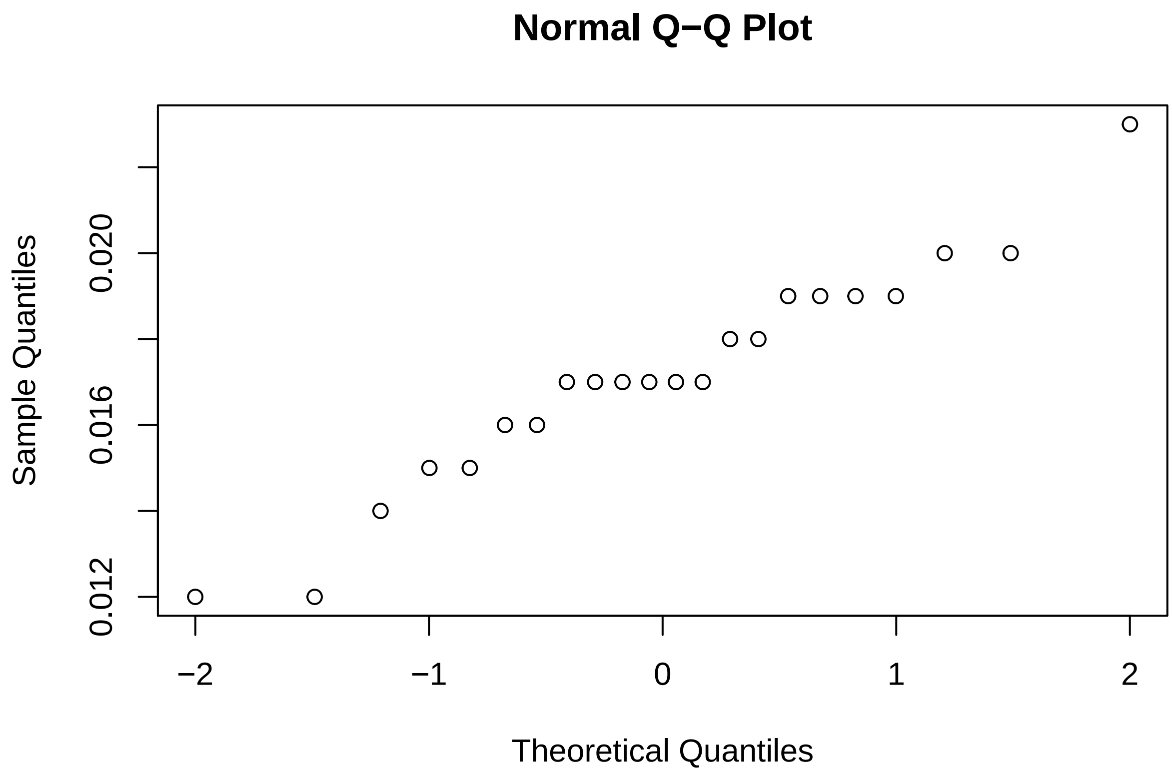

```r
qqnorm(loc1$Ti)
```




```r
qqnorm(loc1$Mn)
```




```r
qqnorm(loc1$Ba)
```



The next 9 graphs will be histogram plots for each of the 9 elements. We will also overlap the 5 locations. The point of this will be to demonstrate the normality of the data and to check whether or not the sample means look simliar. As we will see, all the means for each element look vastly different between the 5 groups. The only two groups that could possibly have means equal to one another would be location 4 and location 5. The min and max functions are just simply to see where to put the xlimits for the graph to look more regular and will be omitted from the remaining 8 r-blocks. Again, please accept these are "approximately" normal.

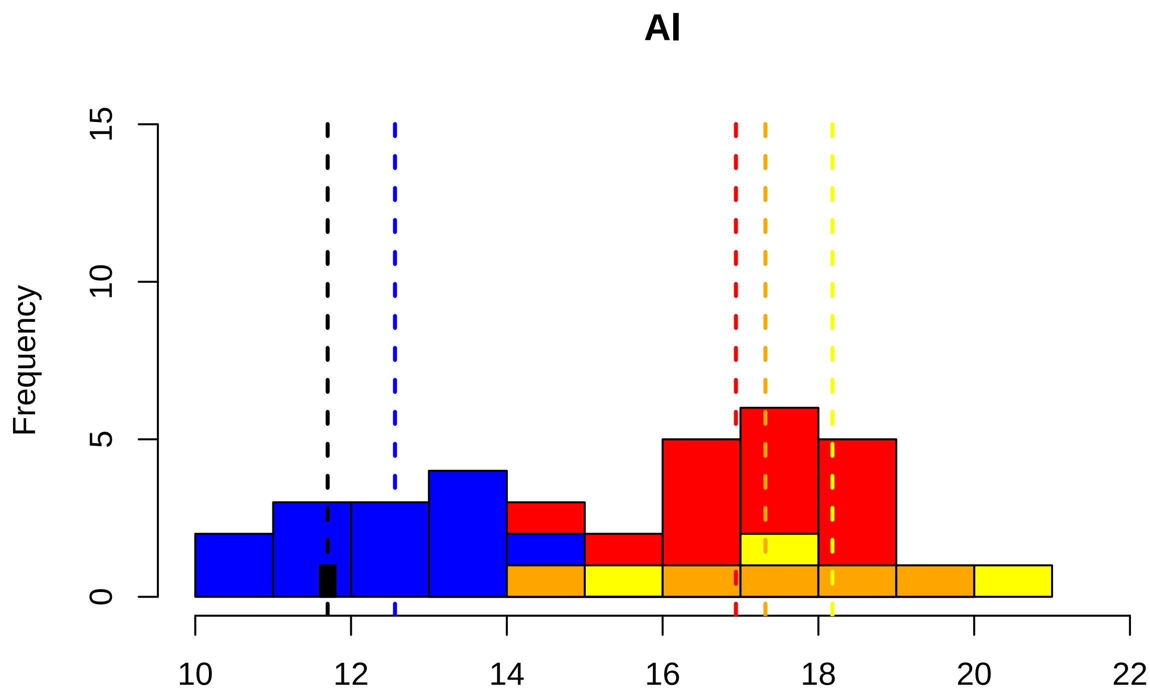
```
min(pot$Al)
```

```
## [1] 10.1
```

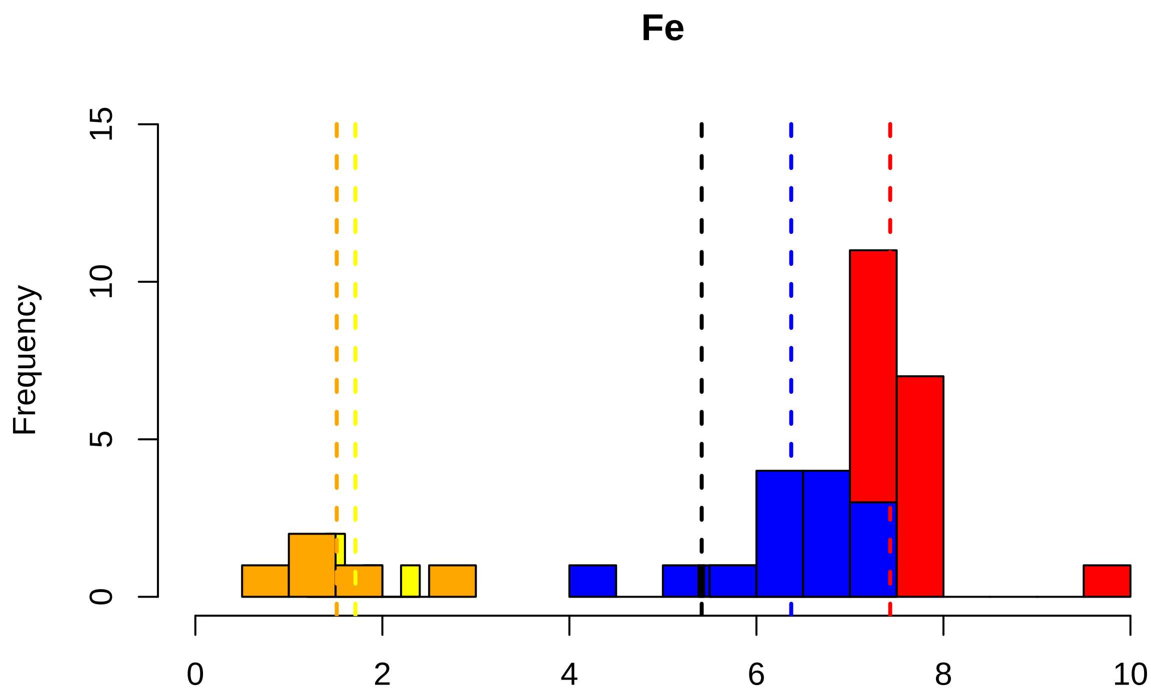
```
max(pot$Al)
```

```
## [1] 20.8
```

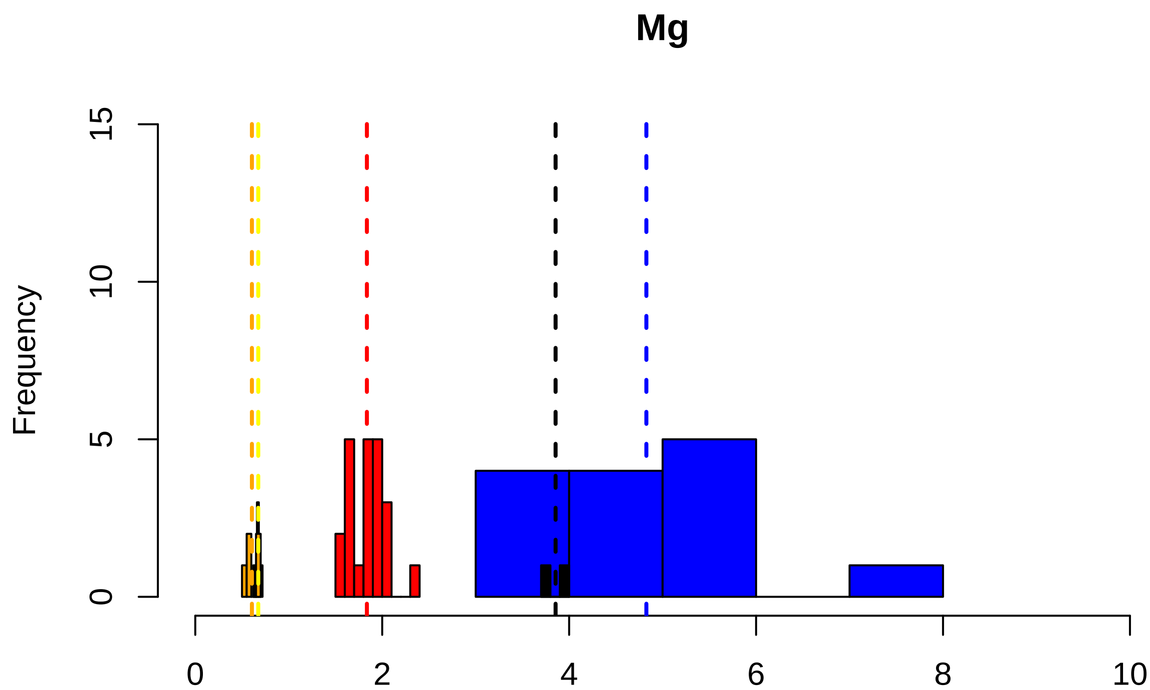
```
hist(loc1$Al,  xlim = c(10,22), ylim = c(0,15), xlab = "", main = "Al", col = "red")
hist(loc2$Al,  xlim = c(10,22), ylim = c(0,15), xlab = "", main = "Al", col = "blue", add = T)
hist(loc3$Al,  xlim = c(10,22), ylim = c(0,15), xlab = "", main = "Al", col = "black", add = T)
hist(loc4$Al,  xlim = c(10,22), ylim = c(0,15), xlab = "", main = "Al", col = "yellow", add = T)
hist(loc5$Al,  xlim = c(10,22), ylim = c(0,15), xlab = "", main = "Al", col = "orange", add = T)
abline(v = mean(loc1$Al), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Al), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Al), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Al), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Al), col = "orange", lwd = 2, lty = 2)
```



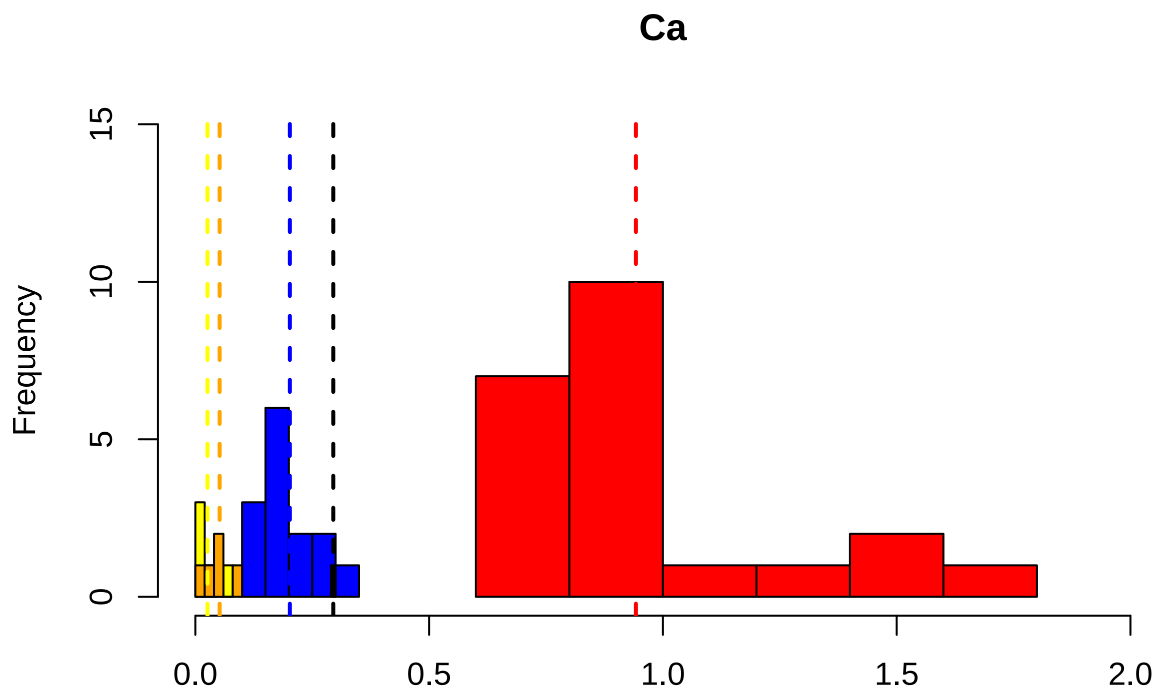

```r
hist(loc1$Fe,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Fe", col = "red")
hist(loc2$Fe,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Fe", col = "blue", add = T)
hist(loc3$Fe,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Fe", col = "black", add = T)
hist(loc4$Fe,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Fe", col = "yellow", add = T)
hist(loc5$Fe,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Fe", col = "orange", add = T)
abline(v = mean(loc1$Fe), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Fe), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Fe), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Fe), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Fe), col = "orange", lwd = 2, lty = 2)
```



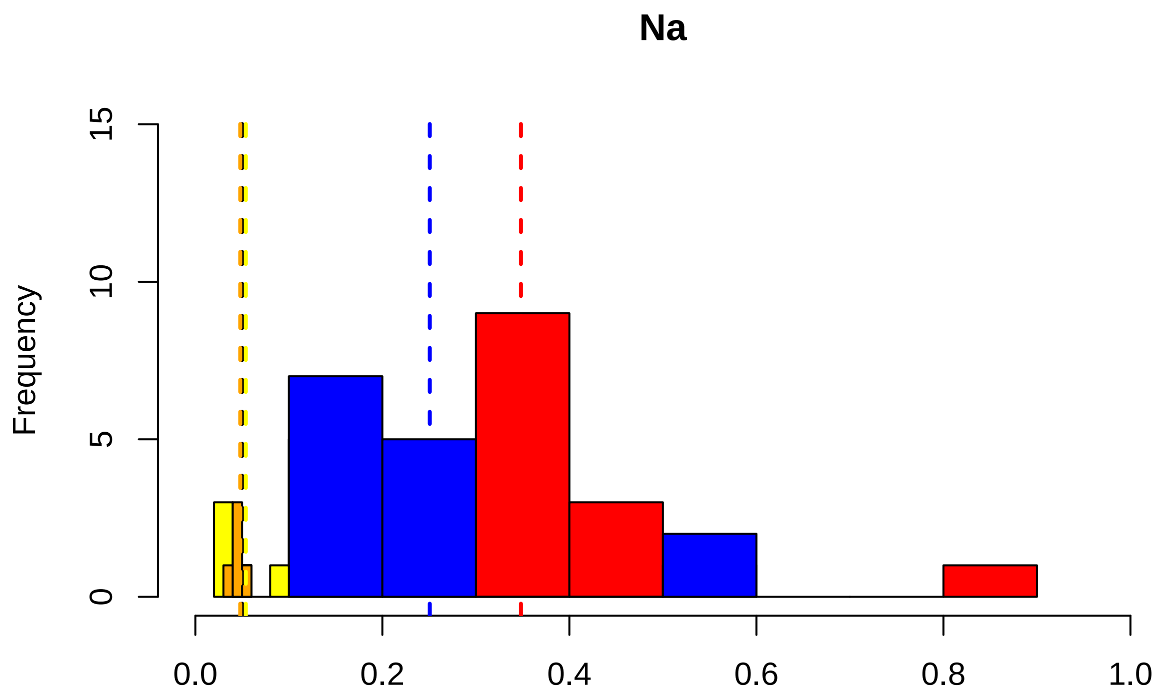

```r
hist(loc1$Mg,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Mg", col = "red")
hist(loc2$Mg,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Mg", col = "blue", add = T)
hist(loc3$Mg,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Mg", col = "black", add = T)
hist(loc4$Mg,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Mg", col = "yellow", add = T)
hist(loc5$Mg,  xlim = c(0,10), ylim = c(0,15), xlab = "", main = "Mg", col = "orange", add = T)
abline(v = mean(loc1$Mg), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Mg), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Mg), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Mg), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Mg), col = "orange", lwd = 2, lty = 2)
```



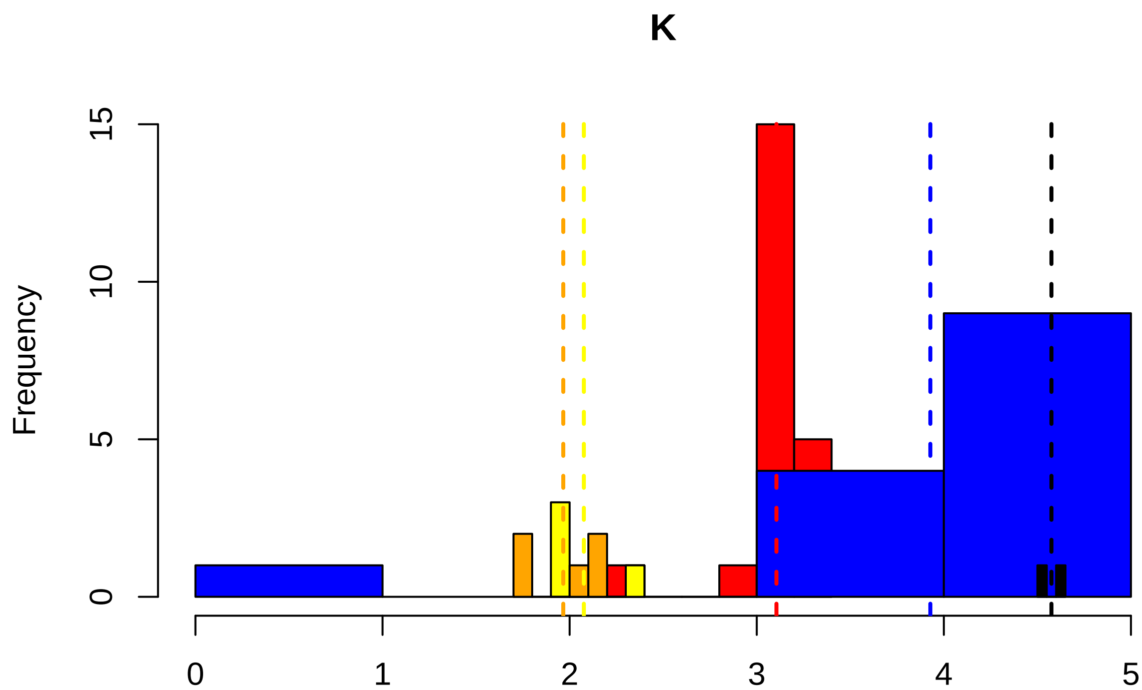

```r
hist(loc1$Ca,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ca", col = "red")
hist(loc2$Ca,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ca", col = "blue", add = T)
hist(loc3$Ca,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ca", col = "black", add = T)
hist(loc4$Ca,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ca", col = "yellow", add = T)
hist(loc5$Ca,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ca", col = "orange", add = T)
abline(v = mean(loc1$Ca), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Ca), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Ca), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Ca), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Ca), col = "orange", lwd = 2, lty = 2)
```



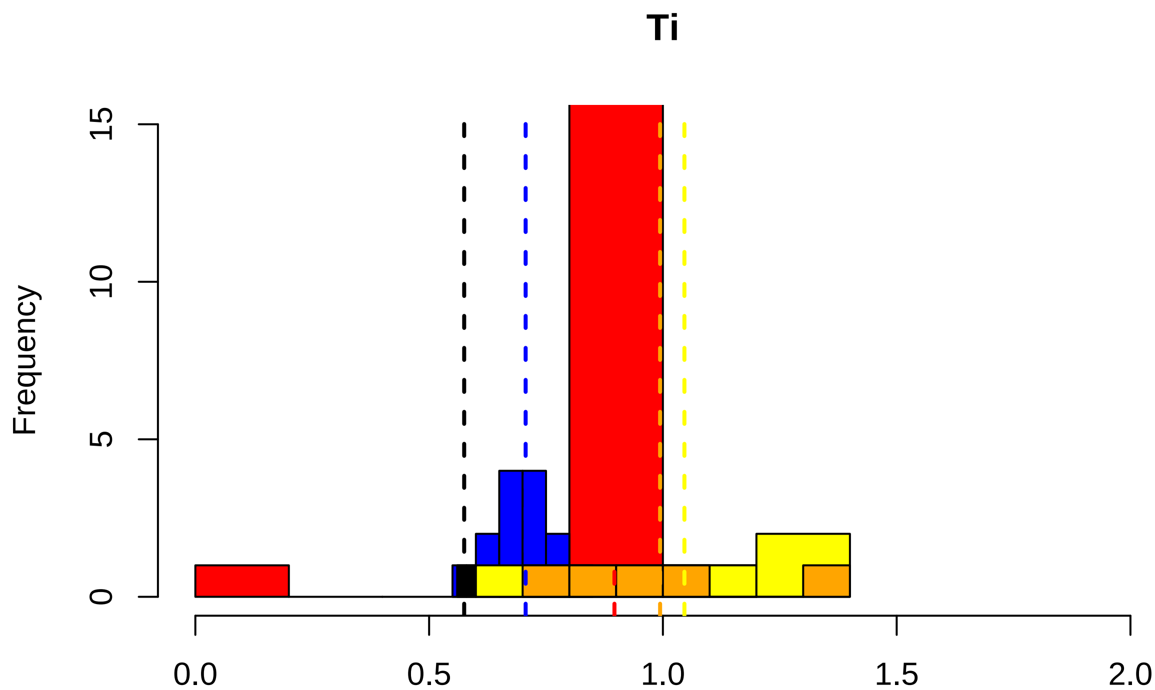

```
hist(loc1$Na,  xlim = c(0,1), ylim = c(0,15), xlab = "", main = "Na", col = "red")
hist(loc2$Na,  xlim = c(0,1), ylim = c(0,15), xlab = "", main = "Na", col = "blue", add = T)
hist(loc3$Na,  xlim = c(0,1), ylim = c(0,15), xlab = "", main = "Na", col = "black", add = T)
hist(loc4$Na,  xlim = c(0,1), ylim = c(0,15), xlab = "", main = "Na", col = "yellow", add = T)
hist(loc5$Na,  xlim = c(0,1), ylim = c(0,15), xlab = "", main = "Na", col = "orange", add = T)
abline(v = mean(loc1$Na), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Na), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Na), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Na), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Na), col = "orange", lwd = 2, lty = 2)
```



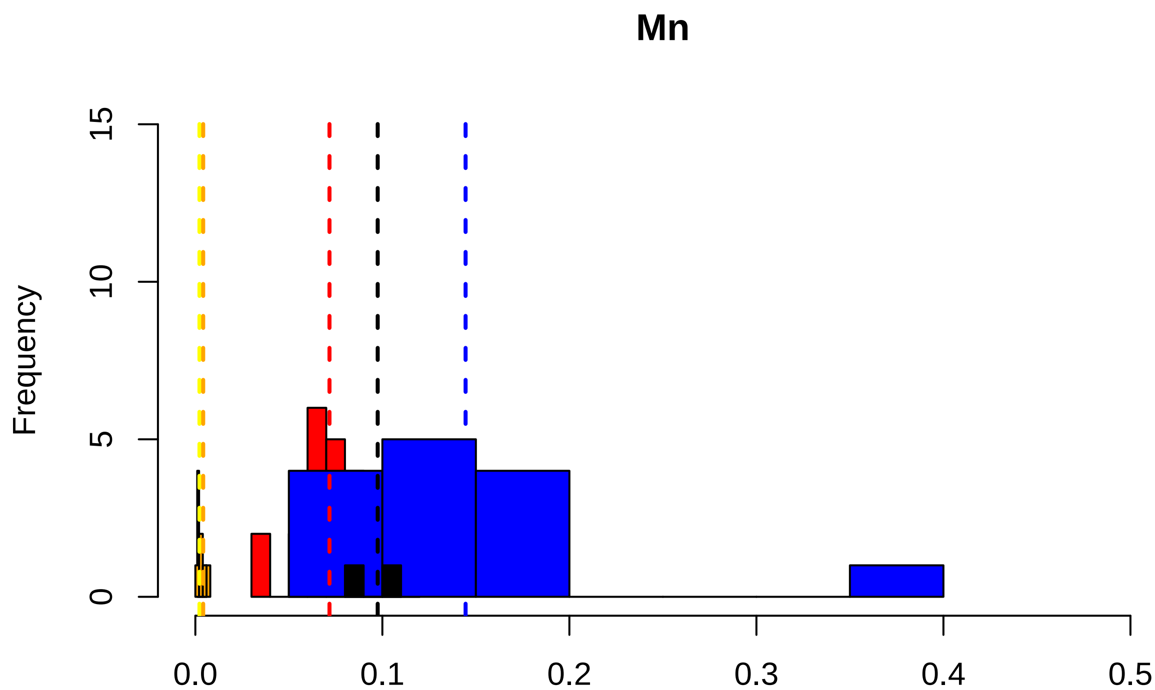

```
hist(loc1$K,  xlim = c(0,5), ylim = c(0,15), xlab = "", main = "K", col = "red")
hist(loc2$K,  xlim = c(0,5), ylim = c(0,15), xlab = "", main = "K", col = "blue", add = T)
hist(loc3$K,  xlim = c(0,5), ylim = c(0,15), xlab = "", main = "K", col = "black", add = T)
hist(loc4$K,  xlim = c(0,5), ylim = c(0,15), xlab = "", main = "K", col = "yellow", add = T)
hist(loc5$K,  xlim = c(0,5), ylim = c(0,15), xlab = "", main = "K", col = "orange", add = T)
abline(v = mean(loc1$K), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$K), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$K), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$K), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$K), col = "orange", lwd = 2, lty = 2)
```



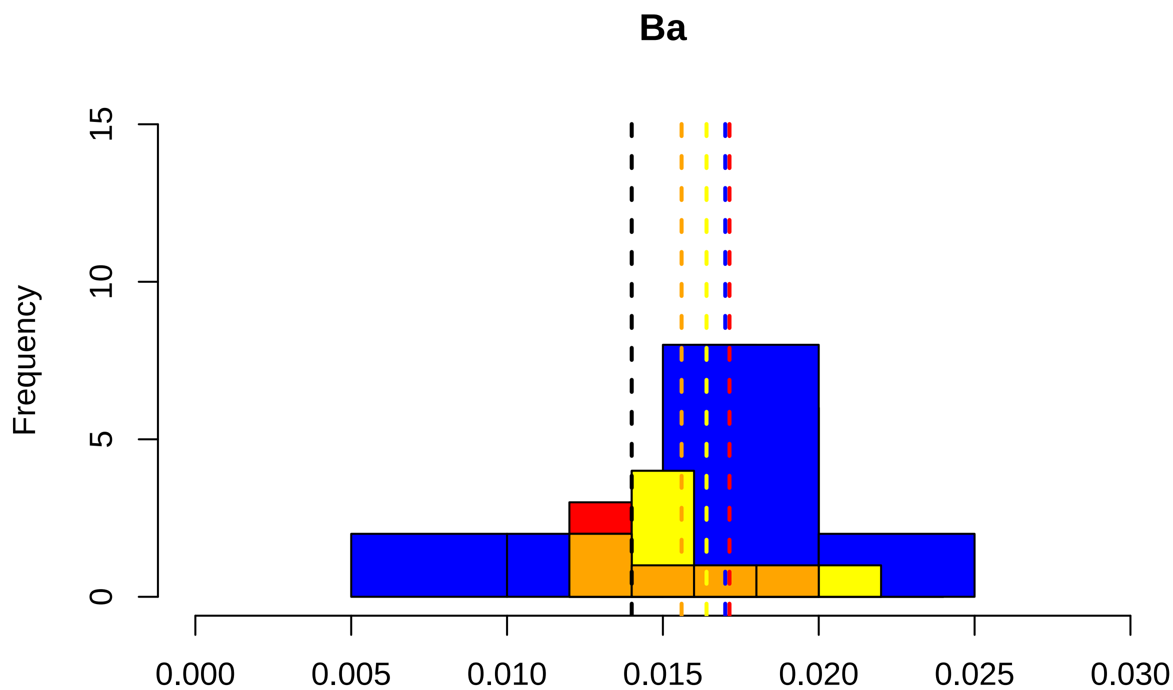

```r
hist(loc1$Ti,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ti", col = "red")
hist(loc2$Ti,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ti", col = "blue", add = T)
hist(loc3$Ti,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ti", col = "black", add = T)
hist(loc4$Ti,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ti", col = "yellow", add = T)
hist(loc5$Ti,  xlim = c(0,2), ylim = c(0,15), xlab = "", main = "Ti", col = "orange", add = T)
abline(v = mean(loc1$Ti), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Ti), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Ti), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Ti), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Ti), col = "orange", lwd = 2, lty = 2)
```




```r
hist(loc1$Mn,  xlim = c(0,.5), ylim = c(0,15), xlab = "", main = "Mn", col = "red")
hist(loc2$Mn,  xlim = c(0,.5), ylim = c(0,15), xlab = "", main = "Mn", col = "blue", add = T)
hist(loc3$Mn,  xlim = c(0,.5), ylim = c(0,15), xlab = "", main = "Mn", col = "black", add = T)
hist(loc4$Mn,  xlim = c(0,.5), ylim = c(0,15), xlab = "", main = "Mn", col = "yellow", add = T)
hist(loc5$Mn,  xlim = c(0,.5), ylim = c(0,15), xlab = "", main = "Mn", col = "orange", add = T)
abline(v = mean(loc1$Mn), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Mn), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Mn), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Mn), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Mn), col = "orange", lwd = 2, lty = 2)
```




```r
hist(loc1$Ba,  xlim = c(0,.03), ylim = c(0,15), xlab = "", main = "Ba", col = "red")
hist(loc2$Ba,  xlim = c(0,.03), ylim = c(0,15), xlab = "", main = "Ba", col = "blue", add = T)
hist(loc3$Ba,  xlim = c(0,.03), ylim = c(0,15), xlab = "", main = "Ba", col = "black", add = T)
hist(loc4$Ba,  xlim = c(0,.03), ylim = c(0,15), xlab = "", main = "Ba", col = "yellow", add = T)
hist(loc5$Ba,  xlim = c(0,.03), ylim = c(0,15), xlab = "", main = "Ba", col = "orange", add = T)
abline(v = mean(loc1$Ba), col = "red", lwd = 2, lty = 2)
abline(v = mean(loc2$Ba), col = "blue", lwd = 2, lty = 2)
abline(v = mean(loc3$Ba), col = "black", lwd = 2, lty = 2)
abline(v = mean(loc4$Ba), col = "yellow", lwd = 2, lty = 2)
abline(v = mean(loc5$Ba), col = "orange", lwd = 2, lty = 2)
```



Here, we will simply conduct univariate analysis of variance for Aluminum. This is simply to demonstrate ability. The F-statistic is found by calculating the Error sum of squares and Treatment sum of squares. We then divide TSS by ESS accounting for degrees of freedom in the equation. We will compare this answer with the answer given by the function (oneway.test) to verify its validity. We will see that the p-value is exceptionally small and the F-statistic exceeds both critical values (normal and adjusted by Bonferroni). Thus, we reject the null hypothesis that the group means for Aluminum are equivalent.

```
oneway.test(Al ~ Kiln, data = pot, var.equal = TRUE)
```

```
##
##  One-way analysis of means
##
## data:  Al and Kiln
## F = 27.608, num df = 4, denom df = 43, p-value = 2.184e-11
```

```
m = 9
alpha = 0.05
n1 = dim(loc1)[1]
n2 = dim(loc2)[1]
n3 = dim(loc3)[1]
n4 = dim(loc4)[1]
n5 = dim(loc5)[1]
N = n1+n2+n3+n4+n5
m1 = mean(loc1[,1])
m2 = mean(loc2[,1])
m3 = mean(loc3[,1])
```

```
m4 = mean(loc4[,1])
m5 = mean(loc5[,1])
mg = mean(pot$Al)
ESS_Al = sum((loc1$Al-m1)^2) + sum((loc2$Al-m2)^2) + sum((loc3$Al-m3)^2) +
 sum((loc4$Al-m4)^2) + sum((loc5$Al-m5)^2)
# Alternative method:
#ESS_Al = ((n1 - 1)*(var(loc1$Al))) + ((n2 - 1)*(var(loc2$Al))) +
#((n3 - 1)*(var(loc3$Al))) + ((n4 - 1)*(var(loc4$Al))) + ((n5 - 1)*(var(loc5$Al)))
TSS_Al = n1*(m1-mg)^2 + n2*(m2-mg)^2 + n3*(m3-mg)^2 + n4*(m4-mg)^2 + n5*(m5-mg)^2
F1 = (TSS_Al/(5-1)) / (ESS_Al/(N-5))
F1
```

```
## [1] 27.6075
```

```
qf(1-alpha,df1=5-1,df2=N-5)
```

```
## [1] 2.588836
```

```
qf(1-alpha/m,df1=5-1,df2=N-5)
```

```
## [1] 4.241675
```

For curiosity, we will conduct an ANOVA test for Barium simply because it looks like the only possibility for group means to be equivalent. In fact, we will find out that we fail to reject the null hypothesis. In fact, for Barium, the sample means for each group seem equivalent. However, because it was fairly close and no other histogram looks like a possibility, we will stop there for ANOVA.

```
oneway.test(Ba ~ Kiln, data = pot, var.equal = TRUE)
```

```
##
##  One-way analysis of means
##
## data:  Ba and Kiln
## F = 0.67001, num df = 4, denom df = 43, p-value = 0.6164
```

```
m = 9
alpha = 0.05
n1 = dim(loc1)[1]
n2 = dim(loc2)[1]
n3 = dim(loc3)[1]
n4 = dim(loc4)[1]
n5 = dim(loc5)[1]
N = n1+n2+n3+n4+n5
m1 = mean(loc1[,9])
m2 = mean(loc2[,9])
m3 = mean(loc3[,9])
m4 = mean(loc4[,9])
m5 = mean(loc5[,9])
mg = mean(pot$Ba)
ESS_Ba = sum((loc1$Ba-m1)^2) + sum((loc2$Ba-m2)^2) + sum((loc3$Ba-m3)^2) +
```

```
 sum((loc4$Ba-m4)^2) + sum((loc5$Ba-m5)^2)
# Alternative method:
#ESS_Ba = ((n1 - 1)*(var(loc1$Ba))) + ((n2 - 1)*(var(loc2$Ba))) + ((n3 - 1)*(var(loc3$Ba))) +
#((n4 - 1)*(var(loc4$Ba))) + ((n5 - 1)*(var(loc5$Ba)))
TSS_Ba = n1*(m1-mg)^2 + n2*(m2-mg)^2 + n3*(m3-mg)^2 + n4*(m4-mg)^2 + n5*(m5-mg)^2
F1 = (TSS_Ba/(5-1)) / (ESS_Ba/(N-5))
F1
```

```
## [1] 0.6700114
```

```
qf(1-alpha,df1=5-1,df2=N-5)
```

```
## [1] 2.588836
```

```
qf(1-alpha/m,df1=5-1,df2=N-5)
```

```
## [1] 4.241675
```

We will now conduct multivariate analysis. This is done by calculating the Error sum of squares and Hypothesis sum of squares. These equations are different from above in the fact that we will be multiplying matrices.

```
mu1 = colMeans(loc1)
mu2 = colMeans(loc2)
mu3 = colMeans(loc3)
mu4 = colMeans(loc4)
mu5 = colMeans(loc5)
mug = (mu1*n1 + mu2*n2 + mu3*n3 + mu4*n4 + mu5*n5)/(n1+n2+n3+n4+n5)
ESS = cov(loc1)*(n1-1) + cov(loc2)*(n2-1) + cov(loc3)*(n3-1) +
  cov(loc4)*(n4-1) + cov(loc5)*(n5-1)
HSS = n1*(mu1-mug) %*% t(mu1-mug) + n2*(mu2-mug) %*% t(mu2-mug) + n3*(mu3-mug) %*%
 t(mu3-mug) + n4*(mu4-mug) %*% t(mu4-mug) + n5*(mu5-mug) %*% t(mu5-mug)
```

We then calculate the test statistic and F-statistic. There are 4 different methods to accomplish this, but 1 will be sufficient. Here, we use Pillai's Trace to find the test statistic (first column), F-statistic (second column), and p-value (third column). We will notice that we reject the null hypothesis that the group means between the 5 locations are equivalent.

```
p = 9
g = 5
pil = sum(diag(HSS %*% solve(ESS+HSS)))
pil_s = min(p,g-1)
pil_m = (abs(p-g+1)-1)/2
pil_r = (N-g-p-1)/2
pil_stat = (2*pil_r + pil_s + 1)*pil/((2*pil_m + pil_s + 1)*(pil_s - pil))
output = c(pil,pil_stat, 1 - pf(pil_stat,df1 = pil_s*(2*pil_m + pil_s + 1),
                                 df2 = pil_s*(2*pil_r + pil_s + 1)))
output
```

```
## [1] 2.226843e+00 5.302536e+00 1.391109e-13
```

Our goal here was to figure out if there was a significant difference between the 5 group means for the 9 variables. The most appropriate test to figure this out was a MANOVA test. First, for visual help, we demonstrated the data in histogram graphs. Already from this, we could've deduced that there was a significant difference between at least one of the group means. From curiosity, which deemed to be well-placed, we checked if the group means were equivalent for Barium. This, in fact, was true. In the grand scheme, were one wanting to check the origin of Romano-British pottery based on chemical composition, they could not base it on Barium oxide. However, using the MANOVA, we know that there is a significant difference between the group means. The main setback in this project was the sample size. Because the sample size was small, let alone sample size based on location, quite a bit of problems occurred. Were one to actually calculate the variance test between variables, we would see many rejections of the null hypothesis that the variances were equivalent, which is an assumption. Also, as easily seen from the graphs, the histograms for each group weren't perfectly normal. This could lead to many errors such as, were the pottery all to be from the same place in that location, it could have been unfairly exposed to certain elements that weren't representative of the location entirely. The group means could have even been equal with an extreme amount of samples (although this is very unlikely). In all, we may conclude that, based on the observations, there is indeed a significant difference between the 5 location means for the 9 elemental oxides. As always, further statistical analysis would be needed to firmly determine and accept the results.