

# Geospatial Accuracy in Paleolithic Sites of the Central Mountain Ranges (Northwestern Iberia): A Statistical Study

Mikel Díaz-Rodríguez

2024-07-23

## Introduction

This document contains the code to reproduce the analyses presented in the chapter titled “Geospatial Accuracy in Paleolithic Sites of the Central Mountain Ranges (Northwestern Iberia): A Statistical Study.” The procedure is outlined below.

## Import data from csv files

```
sample1 <- read.csv("csv/sample1.csv", header=TRUE, sep=";", stringsAsFactors=F, dec=",")
sample2 <- read.csv("csv/sample2.csv", header=TRUE, sep=";", stringsAsFactors=F, dec=",")
```

## Calculate the difference in distances

```
distances <- sample1$distance
```

## Statistical summary

```
summary(distances)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0      0.0     11.5    116.2   182.3    679.4
```

## Check normality with Shapiro test

```
shapiro.test(sample1$distance)
```

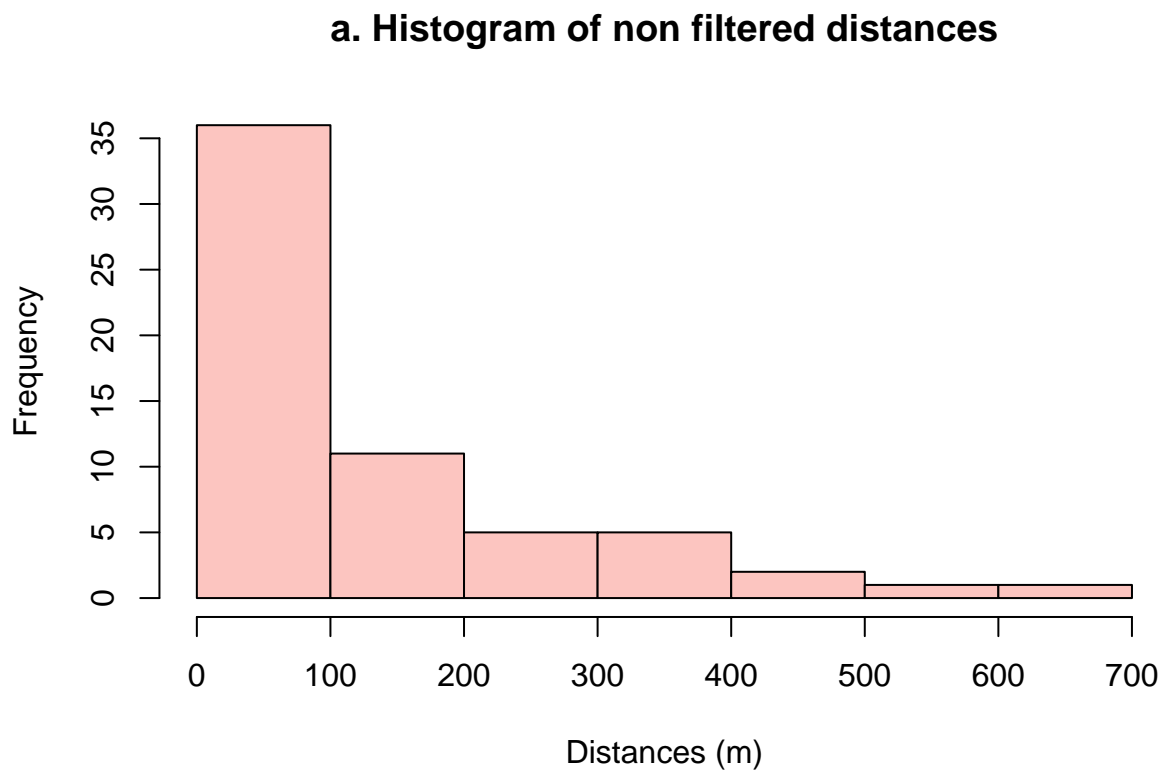
```
##
##  Shapiro-Wilk normality test
##
## data:  sample1$distance
## W = 0.76797, p-value = 1.929e-08
```

Divide the Plots window in 2 parts

```
par(mfrow = c(1, 2))
```

Create histogram

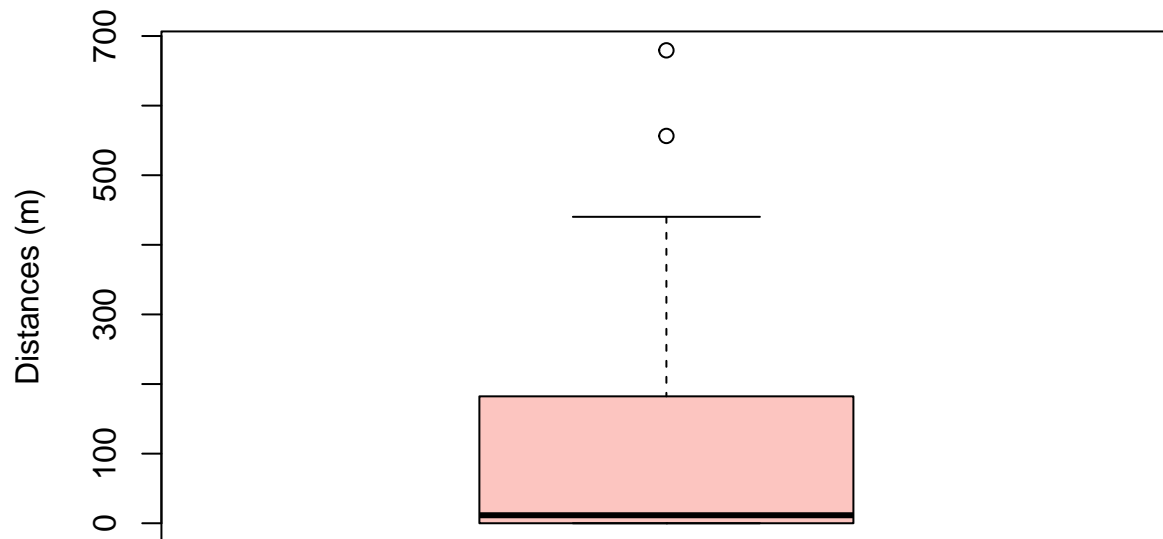
```
hist(sample1$distance, main = "a. Histogram of non filtered distances",  
      xlab = "Distances (m)",  
      col = "#fcc5c0")
```



Create boxplot

```
boxplot(sample1$distance, main = "b. Boxplot of non filtered distances",  
        ylab = "Distances (m)",  
        col = "#fcc5c0")
```

## b. Boxplot of non filtered distances



Exclude distance values = 0

```
filtered_sample <- subset(sample1, distance > 0)
```

Calculate the difference in distances

```
distances2 <- filtered_sample$distance
```

Statistical summary

```
summary(distances2)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	11.5	121.3	182.3	228.6	313.2	679.4

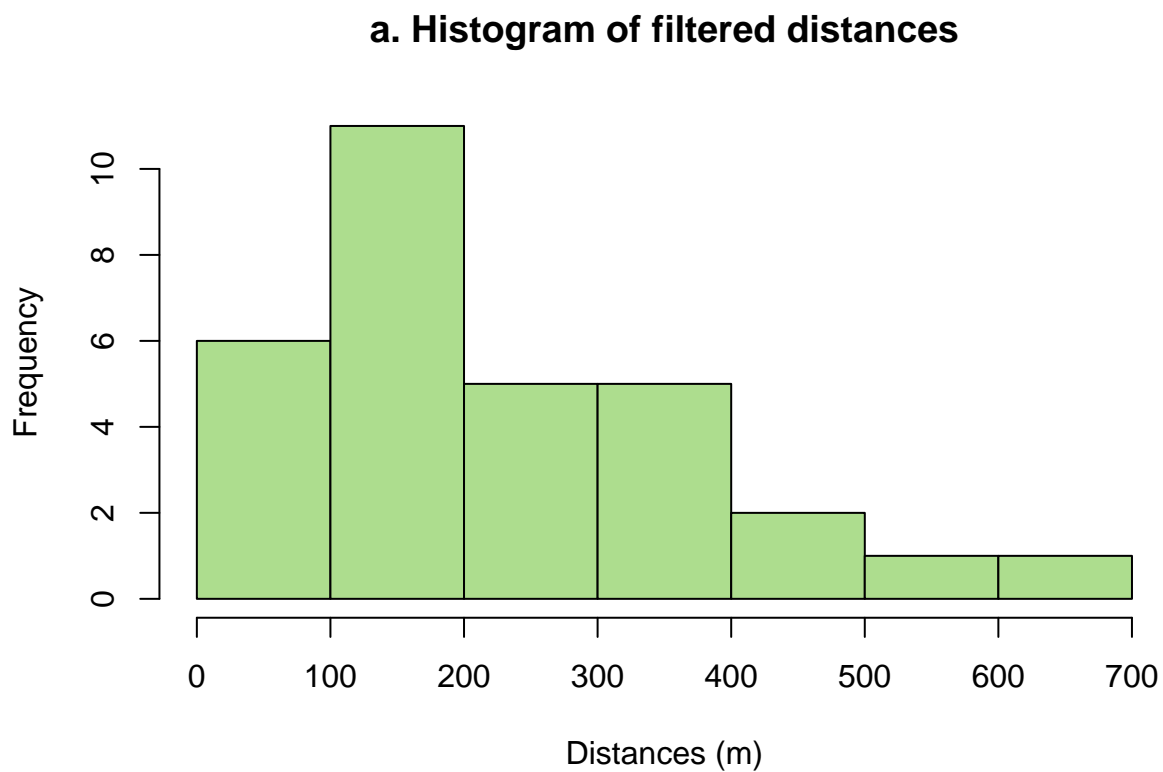
Check normality with Shapiro test

```
shapiro.test(filtered_sample$distance)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  filtered_sample$distance  
## W = 0.9116, p-value = 0.01421
```

Create histogram

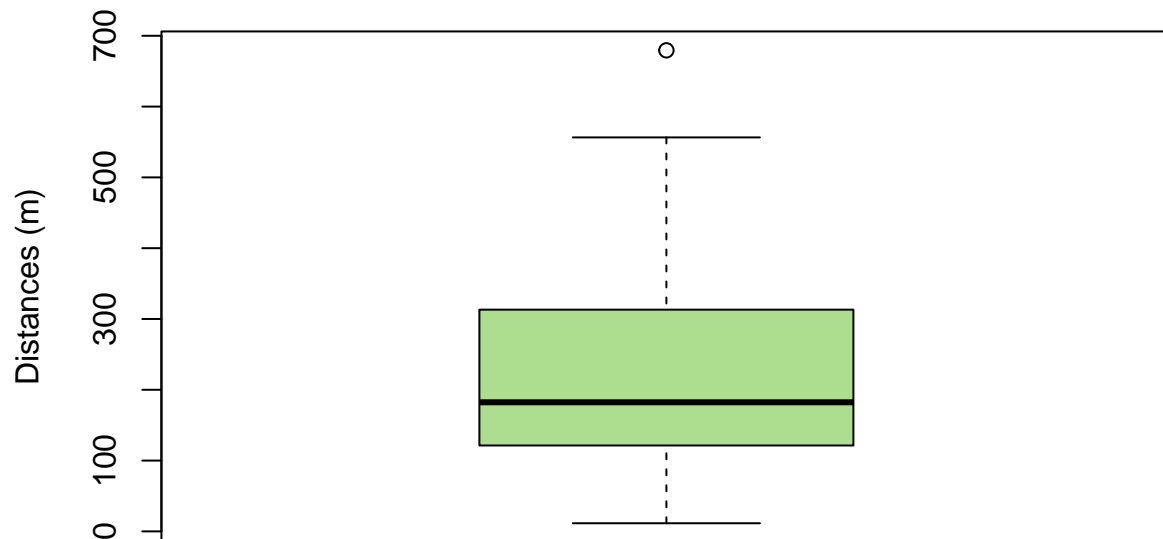
```
hist(filtered_sample$distance, main = "a. Histogram of filtered distances",  
      xlab = "Distances (m)",  
      col = "#add8e")
```



Create boxplot

```
boxplot(filtered_sample$distance, main = "b. Boxplot of filtered distances",  
        ylab = "Distances (m)",  
        col = "#add8e")
```

## b. Boxplot of filtered distances



### Create variables for statistics

```
ALTA1 <- sample1$ALTA
ALTA2 <- sample2$ALTA

HYDROC1 <- sample1$HYDROC
HYDROC2 <- sample2$HYDROC

ASP1 <- sample1$ASP
ASP2 <- sample2$ASP
```

### Statistical summary

```
summary(ALTA1)
```

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	471.0	705.0	725.0	708.3	738.0	759.0

```
summary(ALTA2)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    471.0   720.0   732.0   713.4   742.0   762.0
```

```
summary(HYDROC1)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##     5.34   57.23   68.30   67.51   77.64  108.35
```

```
summary(HYDROC2)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##     5.34   54.76   68.04   65.68   79.51  102.86
```

```
summary(ASP1)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##    18.43  108.44  161.57  178.83  239.04  356.19
```

```
summary(ASP2)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##     9.46   90.00  155.56  170.57  247.62  356.19
```

## Add column

Add 'group' column to each sample to identify where the data comes from.

```
sample1$group <- "Not corrected"
sample2$group <- "Corrected"
```

## Combine the samples into a single dataset

```
combined_sample <- rbind(sample1, sample2)
```

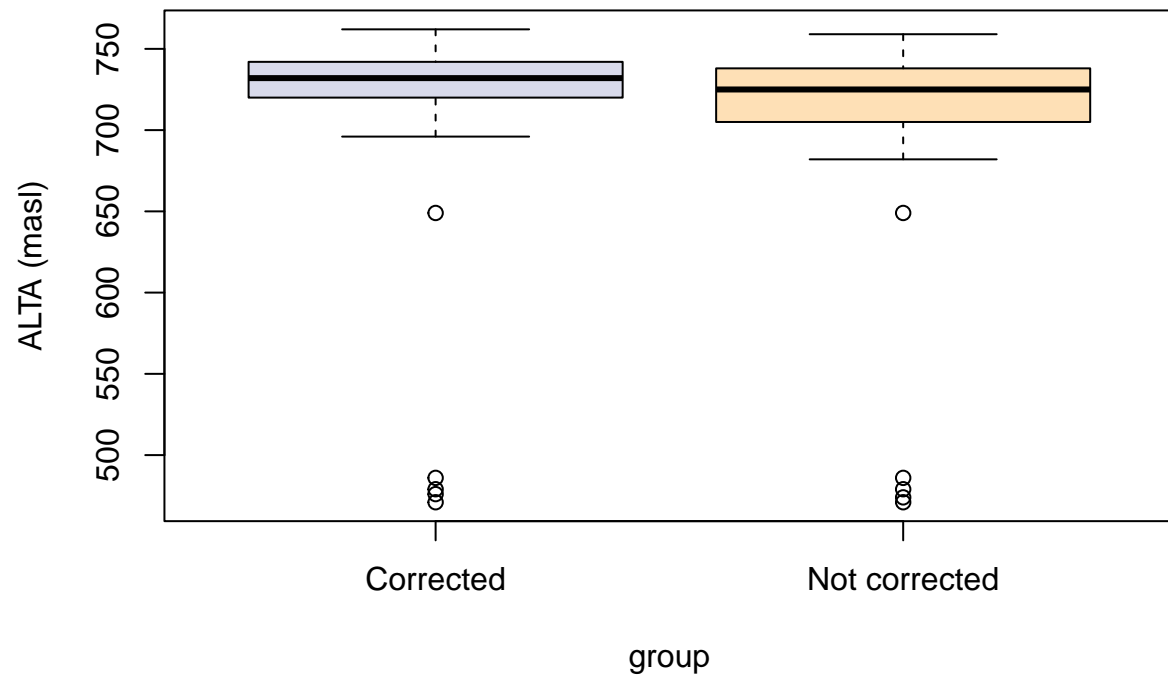
## Divide the Plots window in 3 parts

```
par(mfrow = c(1, 3))
```

## ALTA Boxplot

```
boxplot(ALTA ~ group, data = combined_sample,
        main = "a. ALTA Boxplot",
        ylab = "ALTA (masl)",
        col = c("#d8daeb", "#fee0b6"), # colours to differentiate the samples
        border = "black")
```

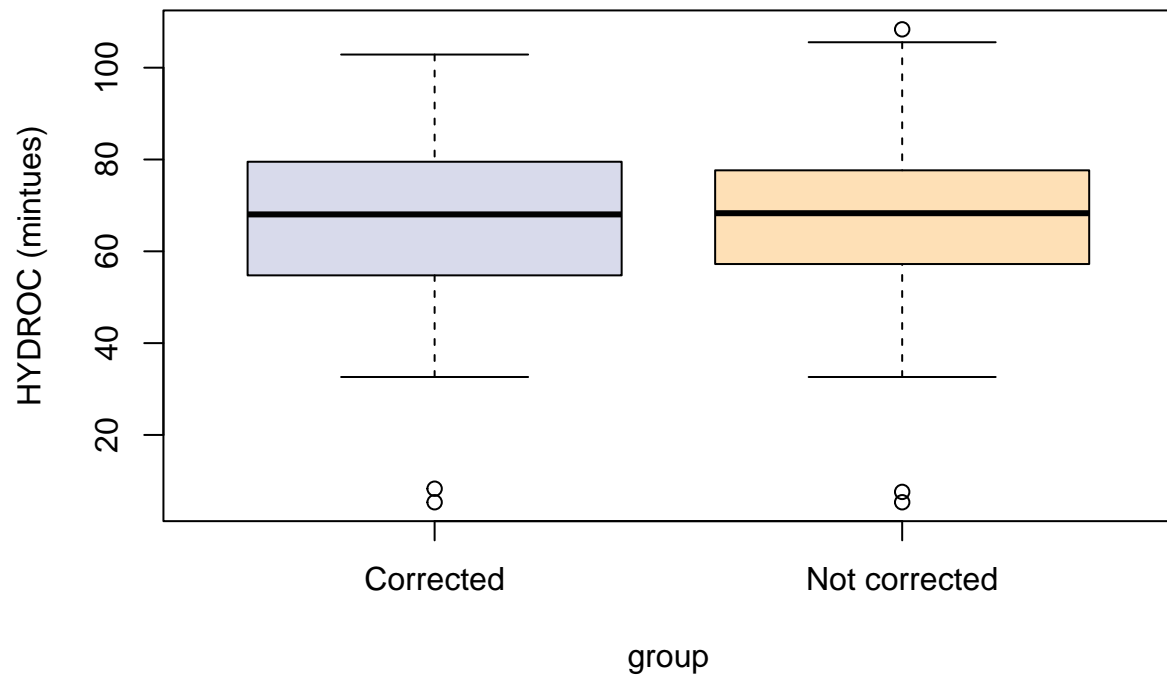
**a. ALTA Boxplot**



**HYDROC Boxplot**

```
boxplot(HYDROC ~ group, data = combined_sample,  
  main = "b. HYDROC boxplot",  
  ylab = "HYDROC (mintues)",  
  col = c("#d8daeb", "#fee0b6"),  
  border = "black")
```

## b. HYDROC boxplot

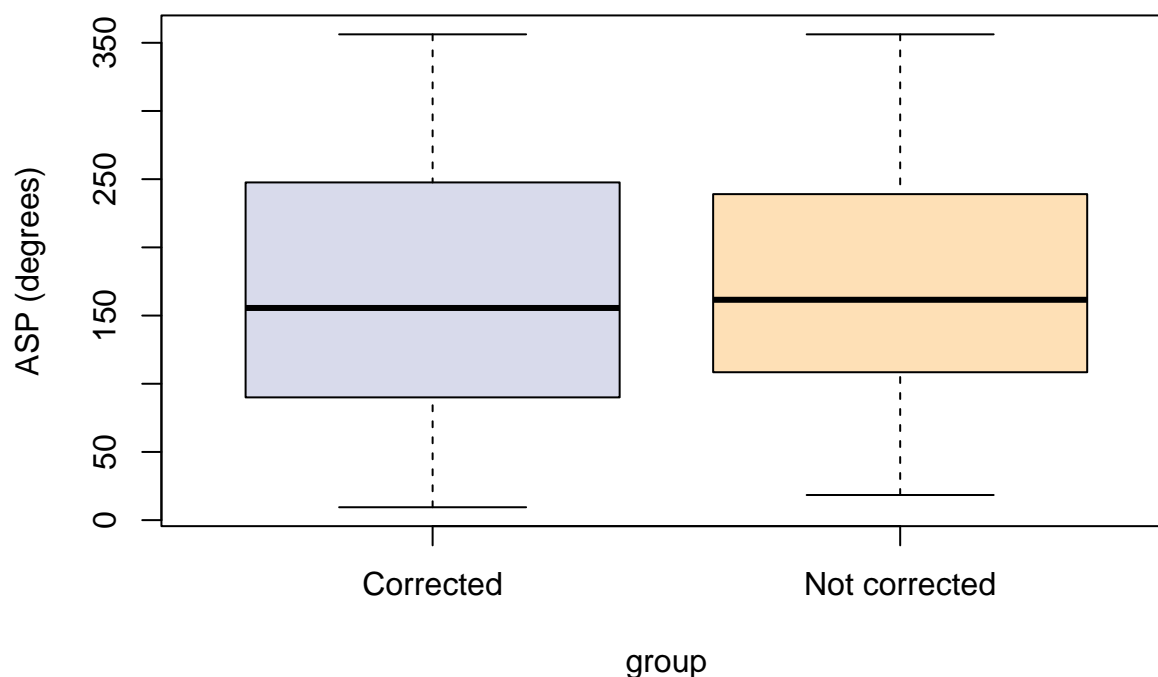


## ASP Boxplot

```
boxplot(ASP ~ group, data = combined_sample,  
  main = "c. ASP Boxplot",  
  ylab = "ASP (degrees)",  
  col = c("#d8daeb", "#fee0b6"),  
  border = "black")
```



### c. ASP Boxplot



### Statistical tests

```
t.test(ALTA ~ group, data = combined_sample)
```

```
##
## Welch Two Sample t-test
##
## data: ALTA by group
## t = 0.43077, df = 120, p-value = 0.6674
## alternative hypothesis: true difference in means between group Corrected and group Not corrected is not equal to 0
## 95 percent confidence interval:
## -18.33519 28.53191
## sample estimates:
## mean in group Corrected mean in group Not corrected
## 713.3770 708.2787
```

```
wilcox.test(ALTA ~ group, data = combined_sample)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: ALTA by group
## W = 2110, p-value = 0.2022
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(HYDROC ~ group, data = combined_sample)
```

```
##
## Welch Two Sample t-test
##
## data: HYDROC by group
## t = -0.52474, df = 119.51, p-value = 0.6007
## alternative hypothesis: true difference in means between group Corrected and group Not corrected is not equal to 0
## 95 percent confidence interval:
## -8.758692 5.088856
## sample estimates:
## mean in group Corrected mean in group Not corrected
## 65.67918 67.51410
```

```
wilcox.test(HYDROC ~ group, data = combined_sample)
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: HYDROC by group
## W = 1776.5, p-value = 0.669
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(ASP ~ group, data = combined_sample)
```

```
##
## Welch Two Sample t-test
##
## data: ASP by group
## t = -0.46896, df = 119.37, p-value = 0.64
## alternative hypothesis: true difference in means between group Corrected and group Not corrected is not equal to 0
## 95 percent confidence interval:
## -43.10435 26.59616
## sample estimates:
## mean in group Corrected mean in group Not corrected
## 170.5743 178.8284
```

```
wilcox.test(ASP ~ group, data = combined_sample)
```

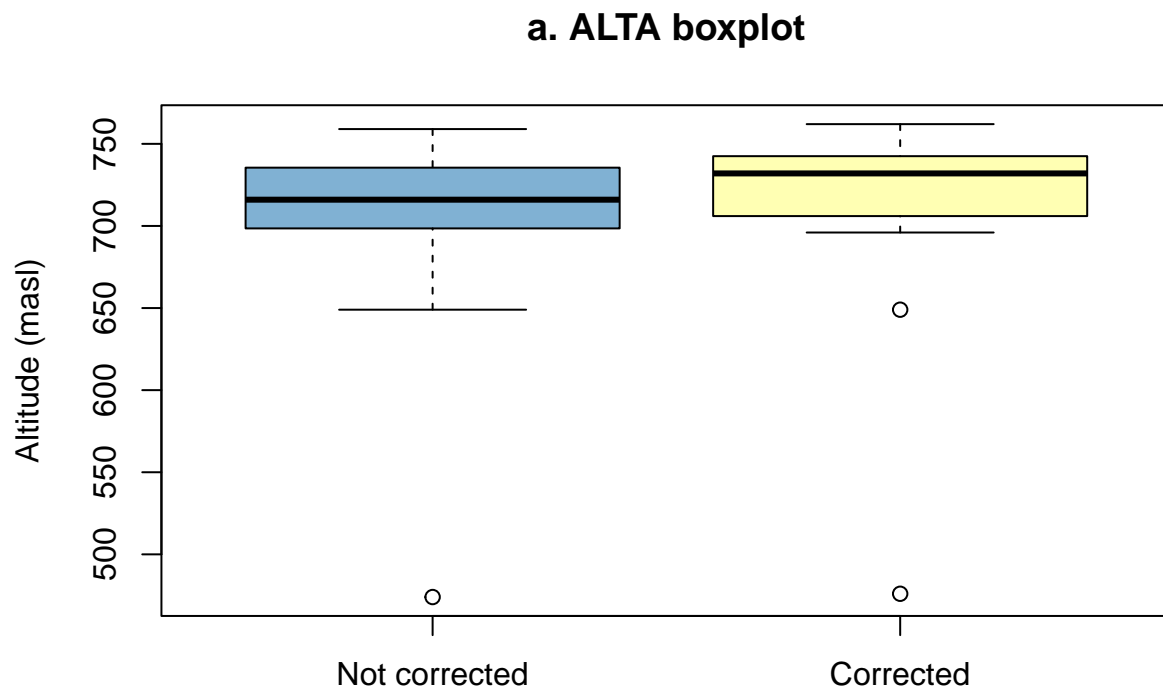
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: ASP by group
## W = 1771, p-value = 0.6486
## alternative hypothesis: true location shift is not equal to 0
```

**Exclude sites with distance values = 0**

```
sample1 <- sample1[sample1$distance != 0, ]
sample2 <- sample2[sample2$distance != 0, ]
```

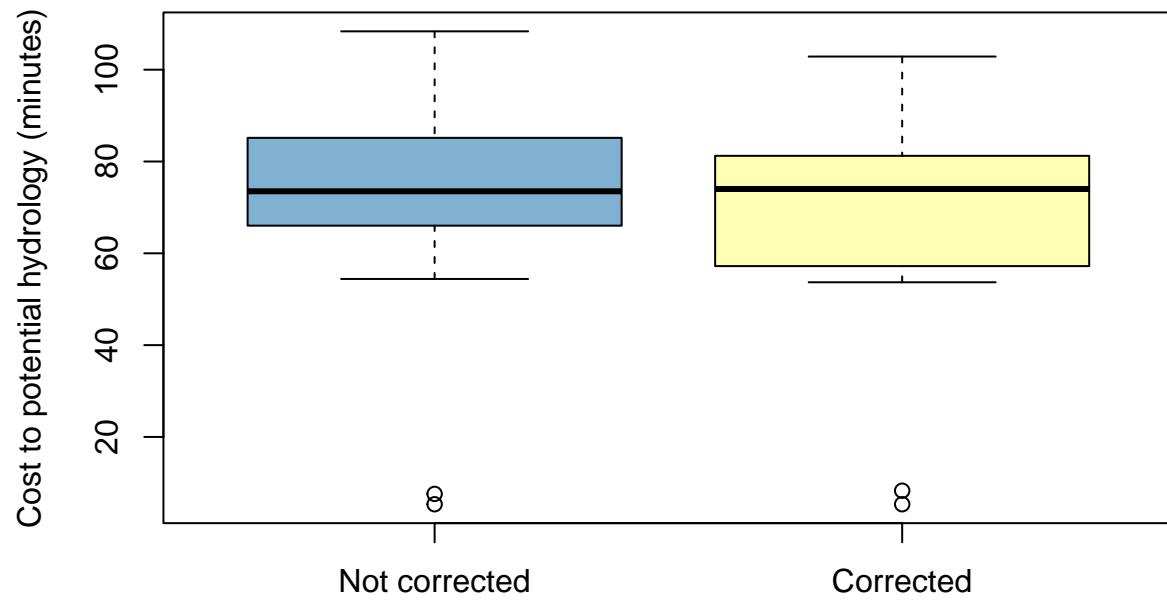
Boxplots to create each variable: ALTA, HYDROC and ASP

```
boxplot(sample1$ALTA, sample2$ALTA,
  main="a. ALTA boxplot",
  ylab="Altitude (masl)",
  col = c("#80b1d3", "#ffffb3"),
  names=c("Not corrected", "Corrected"))
```



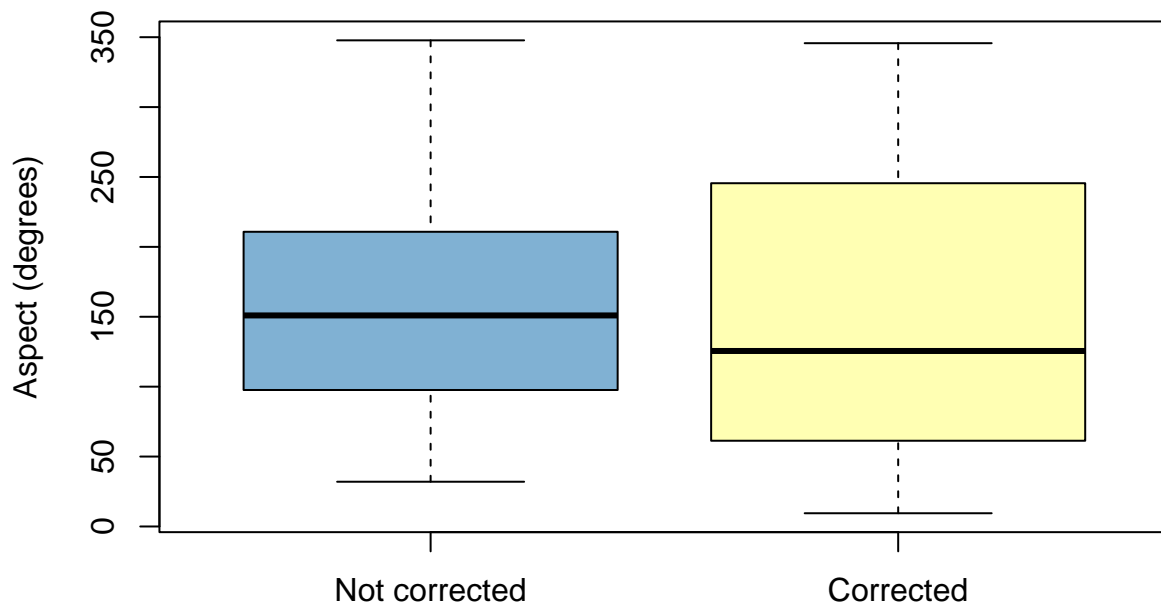
```
boxplot(sample1$HYDROC, sample2$HYDROC,
  main="b. HYDROC boxplot",
  ylab="Cost to potential hydrology (minutes)",
  col = c("#80b1d3", "#ffffb3"),
  names=c("Not corrected", "Corrected"))
```

## b. HYDROC boxplot



```
boxplot(sample1$ASP, sample2$ASP,  
        main="c. ASP boxplot",  
        ylab="Aspect (degrees)",  
        col = c("#80b1d3", "#ffffb3"),  
        names=c("Not corrected", "Corrected"))
```

### c. ASP boxplot



### Statistical tests

```
t.test(sample1$ALTA, sample2$ALTA)
```

```
##
##  Welch Two Sample t-test
##
## data:  sample1$ALTA and sample2$ALTA
## t = -0.78518, df = 59.996, p-value = 0.4354
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -35.59014  15.52562
## sample estimates:
## mean of x mean of y
##  709.0645  719.0968
```

```
wilcox.test(sample1$ALTA, sample2$ALTA)
```

```
## Warning in wilcox.test.default(sample1$ALTA, sample2$ALTA): cannot compute
## exact p-value with ties
```

```
##
```

```
## Wilcoxon rank sum test with continuity correction
##
## data: sample1$ALTA and sample2$ALTA
## W = 355.5, p-value = 0.07947
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(sample1$HYDROC, sample2$HYDROC)
```

```
##
## Welch Two Sample t-test
##
## data: sample1$HYDROC and sample2$HYDROC
## t = 0.65394, df = 59.727, p-value = 0.5157
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -7.434841 14.656131
## sample estimates:
## mean of x mean of y
## 72.58000 68.96935
```

```
wilcox.test(sample1$HYDROC, sample2$HYDROC)
```

```
## Warning in wilcox.test.default(sample1$HYDROC, sample2$HYDROC): cannot compute
## exact p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: sample1$HYDROC and sample2$HYDROC
## W = 526.5, p-value = 0.5218
## alternative hypothesis: true location shift is not equal to 0
```

```
t.test(sample1$ASP, sample2$ASP)
```

```
##
## Welch Two Sample t-test
##
## data: sample1$ASP and sample2$ASP
## t = 0.6358, df = 59.297, p-value = 0.5274
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -34.86962 67.35349
## sample estimates:
## mean of x mean of y
## 166.3571 150.1152
```

```
wilcox.test(sample1$ASP, sample2$ASP)
```

```
## Warning in wilcox.test.default(sample1$ASP, sample2$ASP): cannot compute exact
## p-value with ties
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: sample1$ASP and sample2$ASP
## W = 543.5, p-value = 0.3789
## alternative hypothesis: true location shift is not equal to 0
```

## Create variables for statistics

```
ALTA1 <- sample1$ALTA
ALTA2 <- sample2$ALTA

HYDROC1 <- sample1$HYDROC
HYDROC2 <- sample2$HYDROC

ASP1 <- sample1$ASP
ASP2 <- sample2$ASP
```

## Statistical summary

```
summary(ALTA1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  474.0   698.5   716.0   709.1   735.5   759.0
```

```
summary(ALTA2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  476.0   706.0   732.0   719.1   742.5   762.0
```

```
summary(HYDROC1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.34   66.03   73.51   72.58   85.15  108.35
```

```
summary(HYDROC2)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    5.34   57.21   74.01   68.97   81.23  102.86
```

```
summary(ASP1)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   32.01   97.63  150.95  166.36  210.81  347.74
```

```
summary(ASP2)
```

```
##      Min. 1st Qu.  Median      Mean 3rd Qu.      Max.
##      9.46   61.33  125.54  150.12  245.53  345.70
```