

*Appendix 1. Variable description. A. Statistic values.*

```
> # Statistics of the variables
> for (i in c(1:ncol(finches))) {
+   print(colnames(finches)[i])
+   print(summary(finches[,i]))
+   if (colnames(finches)[i] != "Sex" & colnames(finches)[i] != "Survival") {
+     print(paste("Var:",var(finches[,i])))
+     print(paste("SD:",sd(finches[,i])))
+   }
+ }
```

```
[1] "Sex"
```

```
female   male   NA's
      19      47      34
```

```
[1] "First_adult_year"
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1973   1975   1975   1975   1976   1976
```

```
[1] "Var: 1.07828282828283"
```

```
[1] "SD: 1.03840398125336"
```

```
[1] "Last_Year"
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1977   1977   1978   1978   1978   1982
```

```
[1] "Var: 1.84838383838384"
```

```
[1] "SD: 1.35955280823653"
```

```
[1] "Weight_.g."
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
13.00  15.00  16.24  16.35  17.44  21.24
```

```
[1] "Var: 2.85445903030303"
```

```
[1] "SD: 1.68951443625174"
```

```
[1] "Wing_.mm."
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
64.00  67.00  68.19  68.54  70.25  74.01
```

```
[1] "Var: 5.8352861010101"
```

```
[1] "SD: 2.41563368518699"
```

```
[1] "Tarsus_.mm."
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
17.05  18.49  19.13  19.19  20.00  21.06
```

```
[1] "Var: 0.734847474747475"
```

```
[1] "SD: 0.857232450825023"
```

```
[1] "Beak_Length_.mm."
```

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
8.70  10.20  10.80  10.79  11.25  12.73
```

```
[1] "Var: 0.670374333333333"
```

```
[1] "SD: 0.818763905734329"

[1] "Beak_Depth_.mm."
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 7.500  8.795   9.305   9.392 10.100 11.210
[1] "Var: 0.814485090909091"
[1] "SD: 0.902488277435829"

[1] "Beak_Width_.mm."
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 7.400  8.200   8.600   8.641  9.055 10.000
[1] "Var: 0.345983474747475"
[1] "SD: 0.588203599740324"

[1] "Age"
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 2.00   3.00   3.00   3.96   5.00 10.00
[1] "Var: 2.32161616161616"
[1] "SD: 1.52368505985199"

[1] "Survival"
  Mode  FALSE  TRUE
logical  50    50
```

---

### ***Appendix 1. Variable description. B. Normality check.***

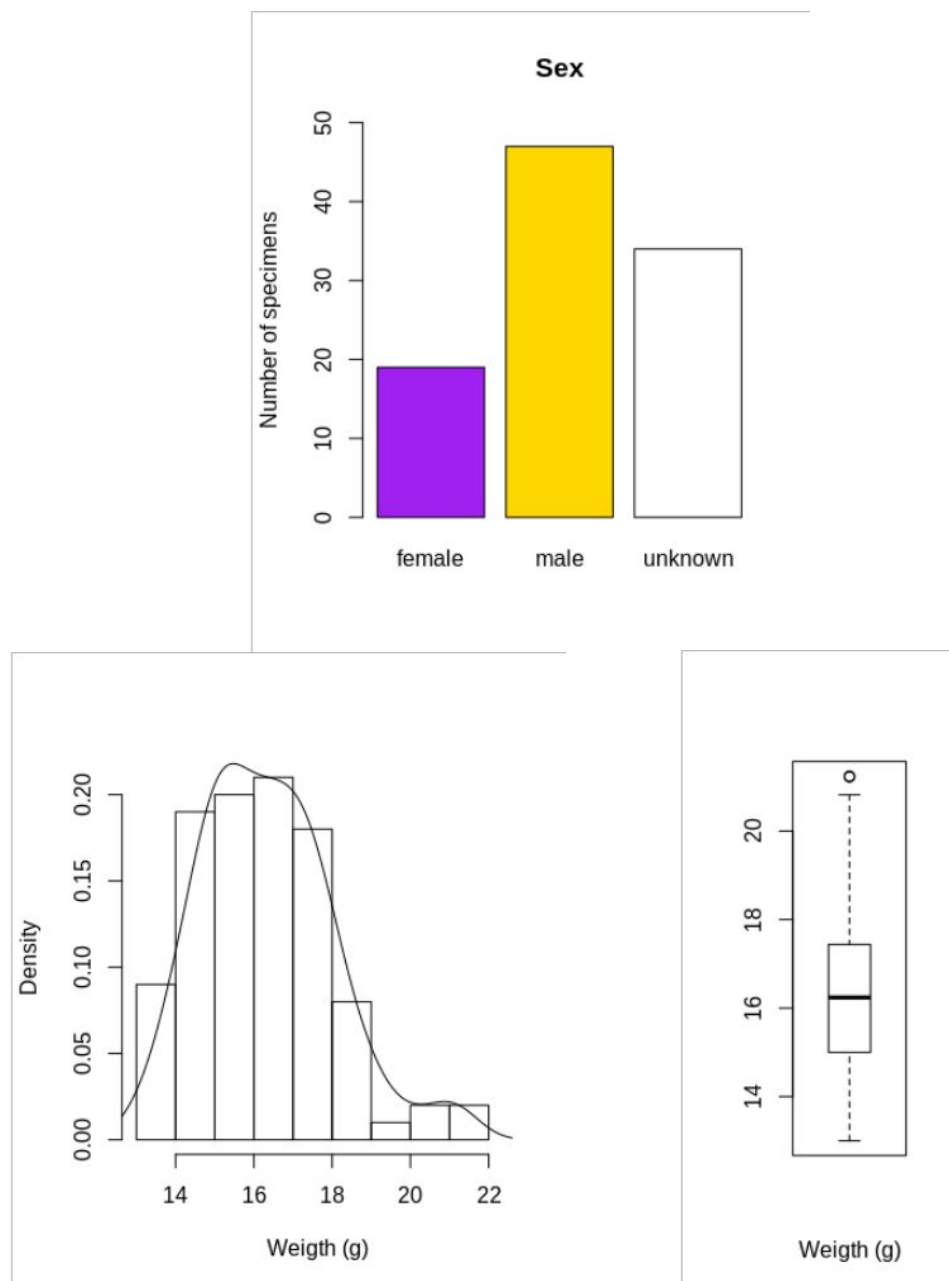
```
> # Plot of the rest of the variables and normality check
> nam_units <- c("Weight (g)", "Wing (mm)", "Tarsus (mm)", "Beak length (mm)",
"Beak depth (mm)", "Beak width (mm)", "Age (years)")
> for (i in c(4:10)){
+   hist(finches[,i], prob=TRUE, xlab=nam_units[i-3], main="",
xlim=c(min(finches[,i]), max(finches[,i]+1)))
+   lines(density(finches[,i]))
+   norm_result <- shapiro.test(finches[,i])[["p.value"]]
+   if (norm_result < 0.05){
+     cat(paste(colnames(finches)[i], "is not normally distributed.\n p-value:",
norm_result, "\n"))
+   }
+   else{
+     cat(paste(colnames(finches)[i], "is normally distributed.\n p-value:",
norm_result, "\n"))
+   }
+   boxplot(finches[,i], xlab=nam_units[i-3])
+ }
```

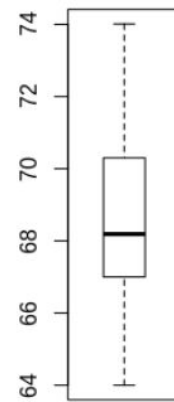
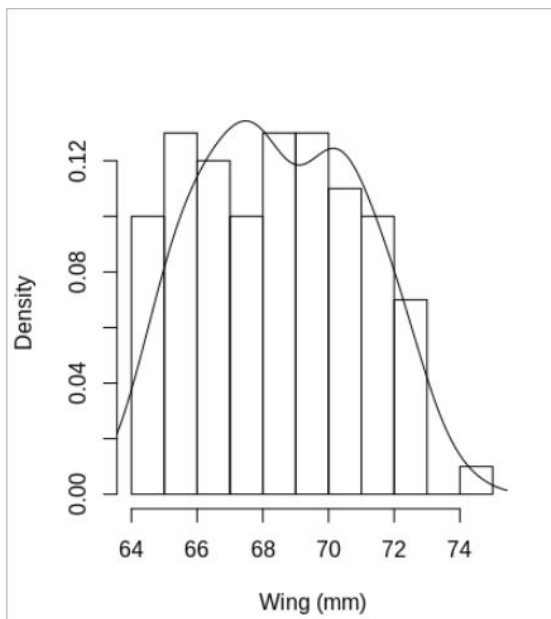
```
Weight_.g. is not normally distributed.
p-value: 0.0146118600730412
```

Wing\_.mm. is not normally distributed.  
p-value: 0.0200195939889433  
Tarsus\_.mm. is normally distributed.  
p-value: 0.310237043894075  
Beak\_Length\_.mm. is normally distributed.  
p-value: 0.596506847488164  
Beak\_Depth\_.mm. is normally distributed.  
p-value: 0.262203149916022  
Beak\_Width\_.mm. is normally distributed.  
p-value: 0.40382358599545  
Age is not normally distributed.  
p-value: 1.1660097045966e-09

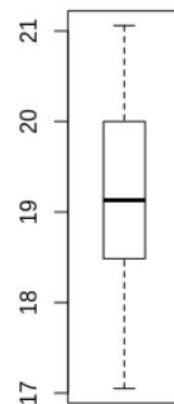
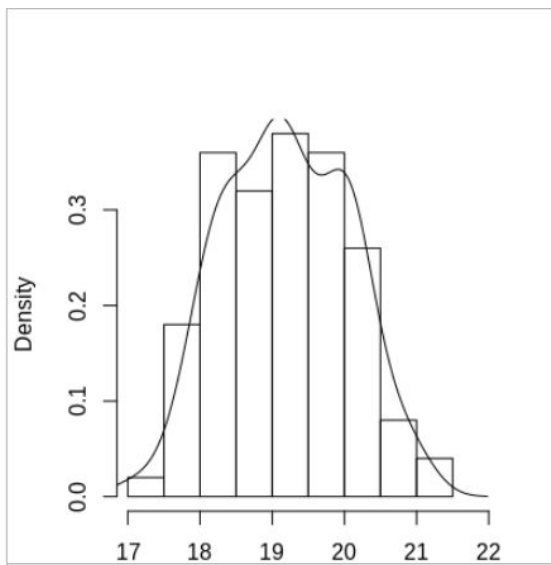
---

*Appendix 1. Variable description. C. Distribution plots.*

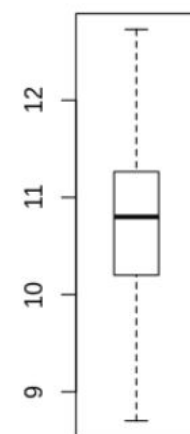
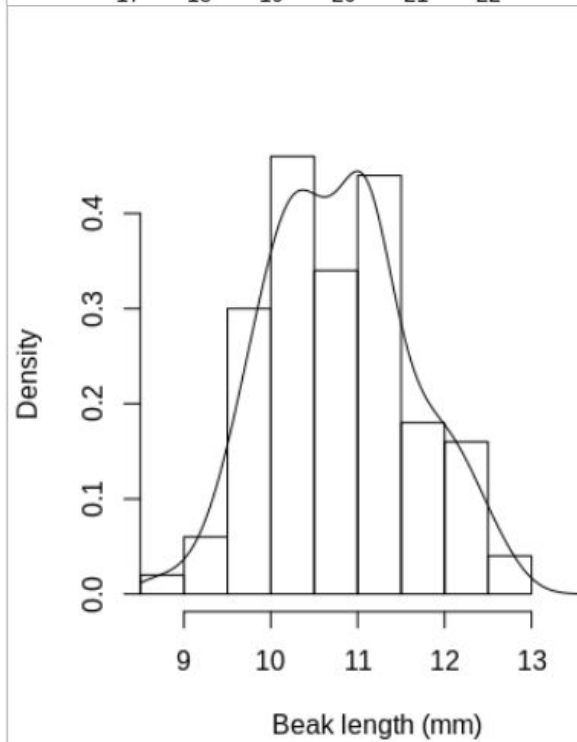




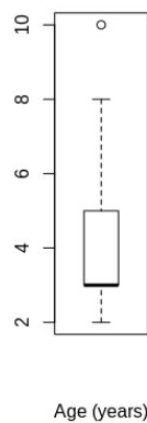
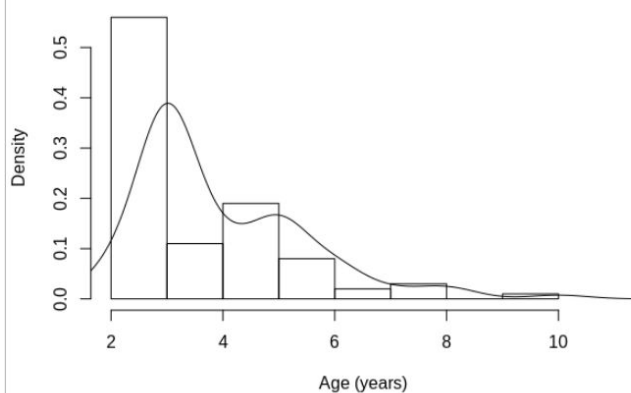
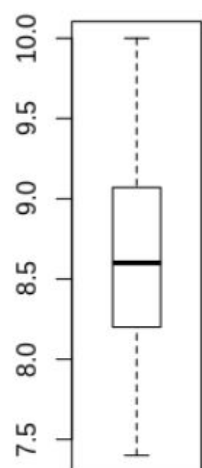
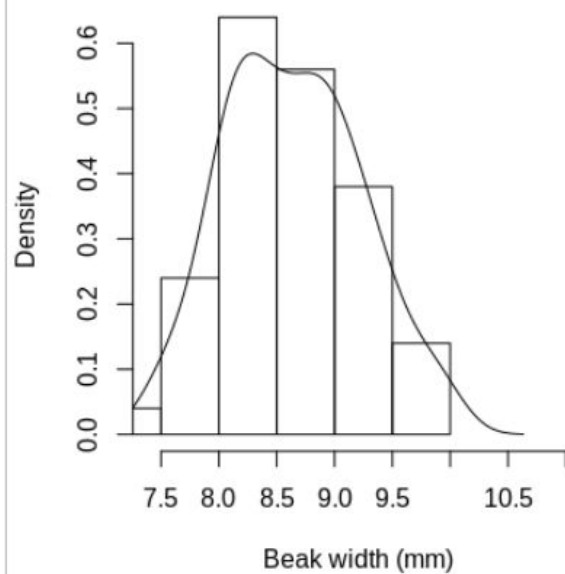
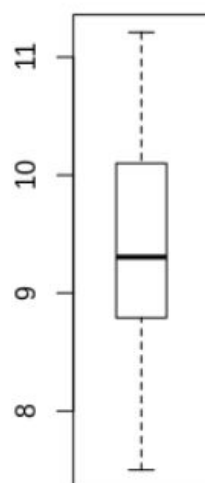
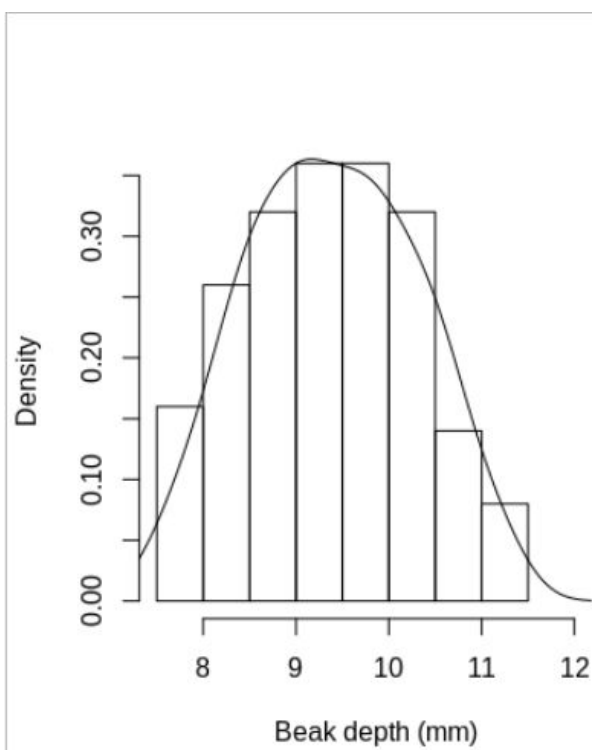
Wing (mm)



Tarsus (mm)



Beak length (mm)



*Appendix 2. Correlation matrix, scatter plots, divided box plots, divided histograms.*

(next page)

*Appendix 3. Hypothesis testing results.*

```

> # Testing differences between groups
> groups_lab <- c("Sex","Survival")
> quant_lab <- c("Wing_.mm.", "Tarsus_.mm.", "Beak_Length_.mm.",
"Beak_Depth_.mm.", "Beak_Width_.mm.")
> # t-tests for normal data
> for (group in groups_lab){
+   for (quant in quant_lab){
+     cat(paste("\n\nComparing",group,"groups differences in", quant, "\n"))
+     homovar <- ifelse(var.test(finches[,quant] ~ finches[,group], data =
finches)$p.value > 0.05, TRUE, FALSE)
+     if (homovar == TRUE){
+       print("Variance is homogeneous.")
+     }
+     else{
+       print("Variance is not homogeneous.")
+     }
+     t_result <- t.test(finches[,quant] ~ finches[,group], data=finches, var.equal
= homovar)
+     print(t_result)
+   }
+ }

```

Comparing Sex groups differences in Wing\_.mm.  
[1] "Variance is homogeneous."

Two Sample t-test

```

data:  finches[, quant] by finches[, group]
t = -3.2306, df = 64, p-value = 0.001952
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0900941 -0.7286741
sample estimates:
mean in group female    mean in group male
          67.78211          69.69149

```

Comparing Sex groups differences in Tarsus\_.mm.  
[1] "Variance is homogeneous."

Two Sample t-test

```

data:  finches[, quant] by finches[, group]
t = -1.7128, df = 64, p-value = 0.0916
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:

```

```
-0.78911457  0.06060393
sample estimates:
mean in group female    mean in group male
      19.07000          19.43426
```

```
Comparing Sex groups differences in Beak_Length_.mm.
[1] "Variance is homogeneous."
```

Two Sample t-test

```
data:  finches[, quant] by finches[, group]
t = -1.3283, df = 64, p-value = 0.1888
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.7202018  0.1449499
sample estimates:
mean in group female    mean in group male
      10.75684          11.04447
```

```
Comparing Sex groups differences in Beak_Depth_.mm.
[1] "Variance is homogeneous."
```

Two Sample t-test

```
data:  finches[, quant] by finches[, group]
t = -1.5886, df = 64, p-value = 0.1171
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9027744  0.1029983
sample estimates:
mean in group female    mean in group male
      9.239474          9.639362
```

```
Comparing Sex groups differences in Beak_Width_.mm.
[1] "Variance is homogeneous."
```

Two Sample t-test

```
data:  finches[, quant] by finches[, group]
t = -0.8566, df = 64, p-value = 0.3949
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.4538535  0.1814459
sample estimates:
```



mean in group female	mean in group male
8.643158	8.779362

Comparing Survival groups differences in Wing\_.mm.

[1] "Variance is homogeneous."

Two Sample t-test

```
data: finches[, quant] by finches[, group]
t = -3.2706, df = 98, p-value = 0.001482
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.4229847 -0.5930153
sample estimates:
mean in group FALSE mean in group TRUE
      67.7886      69.2966
```

Comparing Survival groups differences in Tarsus\_.mm.

[1] "Variance is homogeneous."

Two Sample t-test

```
data: finches[, quant] by finches[, group]
t = -1.8197, df = 98, p-value = 0.07186
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.64472667  0.02792667
sample estimates:
mean in group FALSE mean in group TRUE
      19.0388      19.3472
```

Comparing Survival groups differences in Beak\_Length\_.mm.

[1] "Variance is homogeneous."

Two Sample t-test

```
data: finches[, quant] by finches[, group]
t = -3.6335, df = 98, p-value = 0.0004474
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.8680152 -0.2547848
sample estimates:
mean in group FALSE mean in group TRUE
      10.5122      11.0736
```

Comparing Survival groups differences in Beak\_Depth\_.mm.

```
[1] "Variance is homogeneous."
```

Two Sample t-test

```
data:  finches[, quant] by finches[, group]
t = -3.2668, df = 98, p-value = 0.0015
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.9046804 -0.2209196
sample estimates:
mean in group FALSE  mean in group TRUE
          9.1110          9.6738
```

Comparing Survival groups differences in Beak\_Width\_.mm.

```
[1] "Variance is homogeneous."
```

Two Sample t-test

```
data:  finches[, quant] by finches[, group]
t = -2.58, df = 98, p-value = 0.01136
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.52225761 -0.06814239
sample estimates:
mean in group FALSE  mean in group TRUE
          8.4930          8.7882
```

```
> # Mann-Wihtney U test for weight
> for (group in groups_lab){
+   cat(paste("\n\nComparing",group,"groups differences in weights\n"))
+   homovar <- ifelse(var.test(finches$Weight_.g. ~ finches[,group], data =
finches)$p.value > 0.05, TRUE, FALSE)
+   if (homovar == TRUE){
+     print("Variance is homogeneous.")
+   }
+   else{
+     print("Variance is not homogeneous.")
+   }
+   mannwhitney_result <- wilcox.test(finches$Weight_.g. ~ finches[,group],
data=finches)
+   print(mannwhitney_result)
+ }
```

Comparing Sex groups differences in weights

```
[1] "Variance is homogeneous."
```

Wilcoxon rank sum test with continuity correction

```
data: finches$Weight_.g. by finches[, group]
```

```
W = 327, p-value = 0.09156
```

```
alternative hypothesis: true location shift is not equal to 0
```

Comparing Survival groups differences in weights

```
[1] "Variance is homogeneous."
```

Wilcoxon rank sum test with continuity correction

```
data: finches$Weight_.g. by finches[, group]
```

```
W = 712.5, p-value = 0.0002102
```

```
alternative hypothesis: true location shift is not equal to 0
```

---

#### ***Appendix 4. Code used for the assignment.***

```
setwd("/home/xoel/github/masters/module_ii/statistical_inference/assignment")
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
# The dataset had to be fixed, since there was blank spaces instead of tabulations  
at different points across the file.
```

```
finches <- read.table("./finches_dataset.txt", header=TRUE, sep="\t")
```

```
# Replace the unknown values for Sex with NA.
```

```
finches$Sex[finches$Sex == "unknown"] <- NA
```

```
finches$Sex <- droplevels(finches$Sex)
```

```
# Check which variables are in the dataset
```

```
colnames(finches)
```

```
# Feature amplification
```

```
Age <- finches$Last_Year - finches$First_adult_year + 1
```

```
Survival = finches$Last_Year > 1977
```

```
finches <- cbind(finches, Age, Survival)
```

```
# Remove Band and Species
```

```
finches <- finches[,c(3:ncol(finches))]
```

```

# Statistics of the variables
for (i in c(1:ncol(finches))) {
  print(colnames(finches)[i])
  print(summary(finches[,i]))
  if (colnames(finches)[i] != "Sex" & colnames(finches)[i] != "Survival") {
    print(paste("Var:", var(finches[,i])))
    print(paste("SD:", sd(finches[,i])))
  }
}

# Plot of sex distribution
gendertab <- table(finches$Sex, useNA = "ifany")
names(gendertab)[is.na(names(gendertab))] <- "unknown"
barplot(gendertab, ylim=c(0,50), ylab="Number of specimens", col=c("purple",
"gold", "white"), main="Sex")

# Plot of the rest of the variables and normality check
nam_units <- c("Weigth (g)", "Wing (mm)", "Tarsus (mm)", "Beak length (mm)", "Beak
depth (mm)", "Beak width (mm)", "Age (years)")
for (i in c(4:10)) {
  hist(finches[,i], prob=TRUE, xlab=nam_units[i-3], main="",
xlim=c(min(finches[,i]), max(finches[,i]+1)))
  lines(density(finches[,i]))
  norm_result <- shapiro.test(finches[,i])[["p.value"]]
  if (norm_result < 0.05) {
    cat(paste(colnames(finches)[i], "is not normally distributed.\n p-value:",
norm_result, "\n"))
  }
  else {
    cat(paste(colnames(finches)[i], "is normally distributed.\n p-value:",
norm_result, "\n"))
  }
  boxplot(finches[,i], xlab=nam_units[i-3])
}

# Correlation, scatter, density matrix
ggpairs(finches)

# Checking if weight and wing follow normal distributions if divided by sex
norm_result <- shapiro.test(finches$Weight_.g.[which(finches$Sex ==
"male")])[["p.value"]]
norm_result # Normal
norm_result <- shapiro.test(finches$Weight_.g.[which(finches$Sex ==
"female")])[["p.value"]]
norm_result # Not normal

```

```

norm_result <- shapiro.test(finches$Wing_.mm.[which(finches$Sex ==
"male")])["p.value"]
norm_result # Normal
norm_result <- shapiro.test(finches$Wing_.mm.[which(finches$Sex ==
"female")])["p.value"]
norm_result # Normal

#Checking if weight and wing follow normal distributions if divided by survival
norm_result <- shapiro.test(finches$Weight_.g.[which(finches$Survival ==
TRUE)])["p.value"]
norm_result # Not normal
norm_result <- shapiro.test(finches$Weight_.g.[which(finches$Survival ==
FALSE)])["p.value"]
norm_result # Normal

norm_result <- shapiro.test(finches$Wing_.mm.[which(finches$Survival ==
TRUE)])["p.value"]
norm_result # Normal
norm_result <- shapiro.test(finches$Wing_.mm.[which(finches$Survival ==
FALSE)])["p.value"]
norm_result # Normal

# Testing differences between groups
groups_lab <- c("Sex","Survival")
quant_lab <- c("Wing_.mm.", "Tarsus_.mm.", "Beak_Length_.mm.", "Beak_Depth_.mm.",
"Beak_Width_.mm.")

# t-tests for normal data
for (group in groups_lab){
  for (quant in quant_lab){
    cat(paste("\n\nComparing",group,"groups differences in", quant, "\n"))
    homovar <- ifelse(var.test(finches[,quant] ~ finches[,group], data =
finches)$p.value > 0.05, TRUE, FALSE)
    if (homovar == TRUE){
      print("Variance is homogeneous.")
    }
    else{
      print("Variance is not homogeneous.")
    }
    t_result <- t.test(finches[,quant] ~ finches[,group], data=finches, var.equal
= homovar)
    print(t_result)
  }
}

# Mann-Wihtney U test for weight
for (group in groups_lab){

```

```
cat(paste("\n\nComparing",group,"groups differences in weights\n"))
homovar <- ifelse(var.test(finches$Weight_.g. ~ finches[,group], data =
finches)$p.value > 0.05, TRUE, FALSE)
if (homovar == TRUE){
  print("Variance is homogeneous.")
}
else{
  print("Variance is not homogeneous.")
}
mannwhitney_result <- wilcox.test(finches$Weight_.g. ~ finches[,group],
data=finches)
print(mannwhitney_result)
}
```