Homework1

Case study 1.

Init

```r
# install.packages("car")
library(readxl)
beak.finch <- read_excel("/Users/kanoalindiwe/Downloads/Projects/playground/Datasets/beak.finch.xlsx")
year.cat <- as.factor(beak.finch$year)
```

1. E

```r
summary(year.cat)
```

```
## 1976 1978
##   89   89
```

2. A/C

```r
tapply(beak.finch$depth, beak.finch$year, summary)["1976"]
```

```
## $`1976`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    6.20    8.90    9.70    9.47   10.20   11.70
```

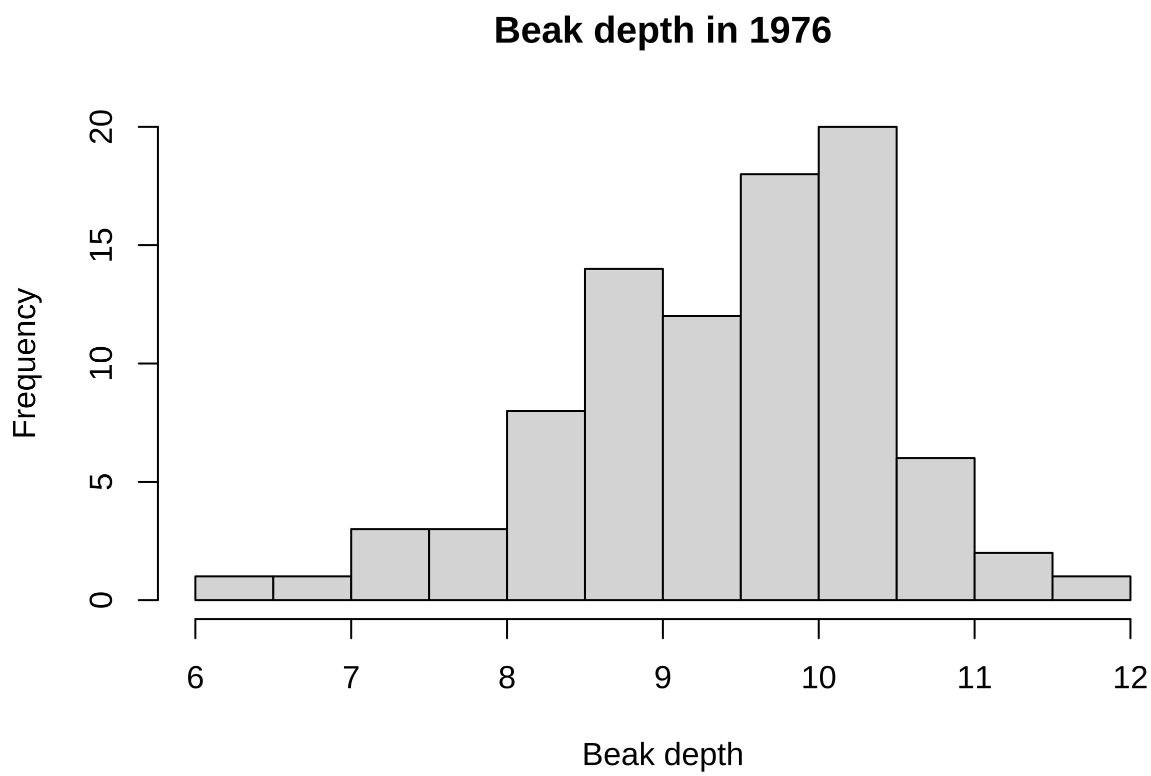
3. B

```r
tapply(beak.finch$depth, beak.finch$year, summary)["1978"]
```

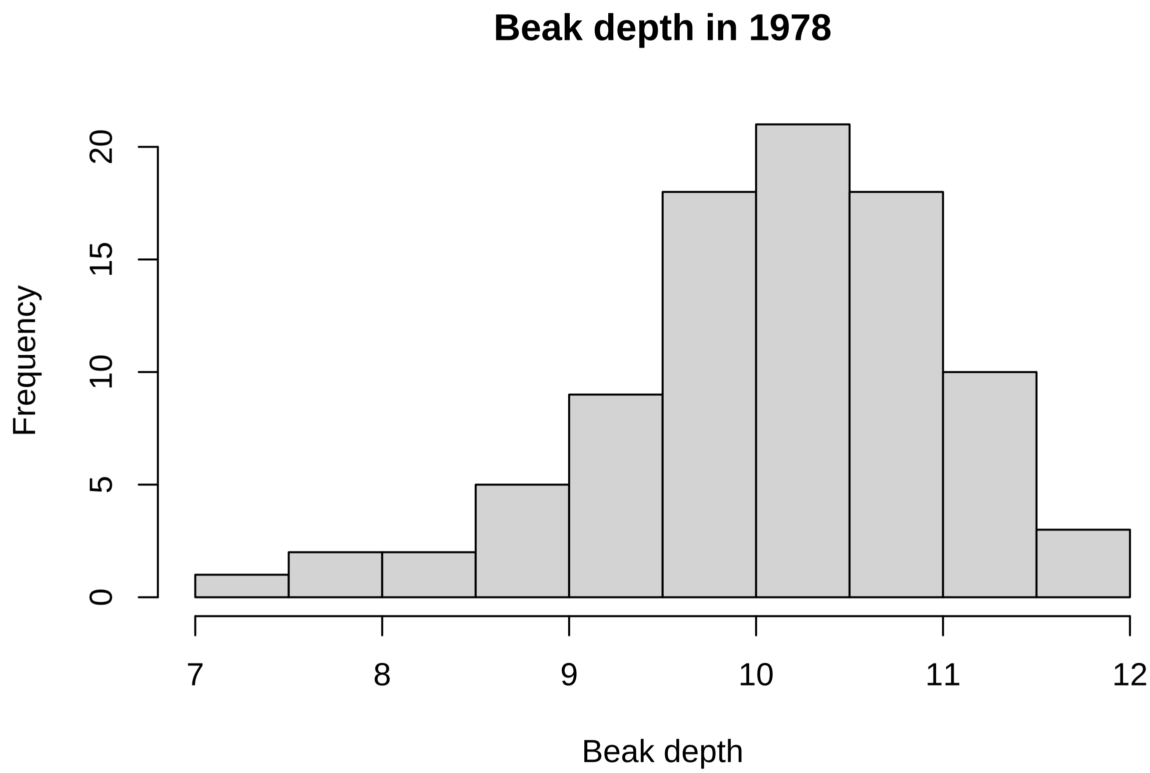
```
## $`1978`
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.10    9.60   10.30   10.14   10.70   11.70
```

4. C

```r
hist(beak.finch$depth[beak.finch$year == 1976],
     main = "Beak depth in 1976",
     xlab = "Beak depth")
```

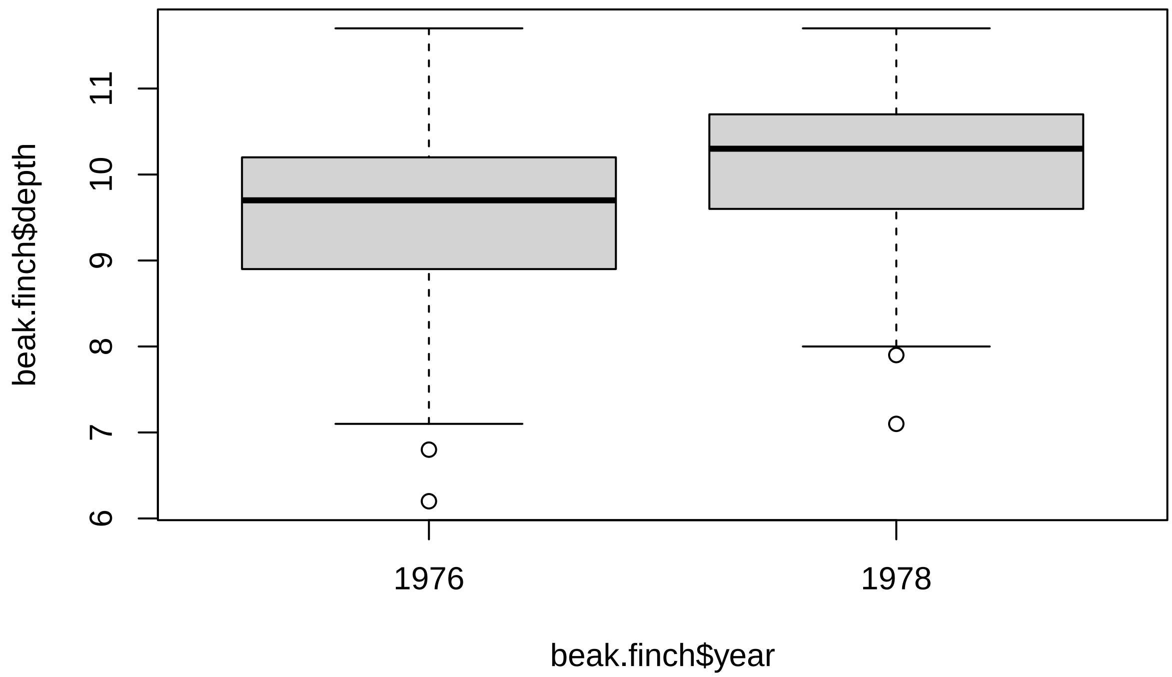



```r
hist(beak.finch$depth[beak.finch$year == 1978],
     main = "Beak depth in 1978",
     xlab = "Beak depth",)
```



5. A (however, the whiskers go to 1.5 times the IQR)

```
boxplot(beak.finch$depth ~ beak.finch$year)
```



6. A

```
t.test(beak.finch$depth ~ beak.finch$year)
```

```
##
##  Welch Two Sample t-test
##
## data:  beak.finch$depth by beak.finch$year
## t = -4.5833, df = 172.98, p-value = 8.739e-06
## alternative hypothesis: true difference in means between group 1976 and group 1978 is not equal to 0
## 95 percent confidence interval:
##  -0.9564436 -0.3806350
## sample estimates:
## mean in group 1976 mean in group 1978
##           9.469663          10.138202
```

7. **A**

8. **A**

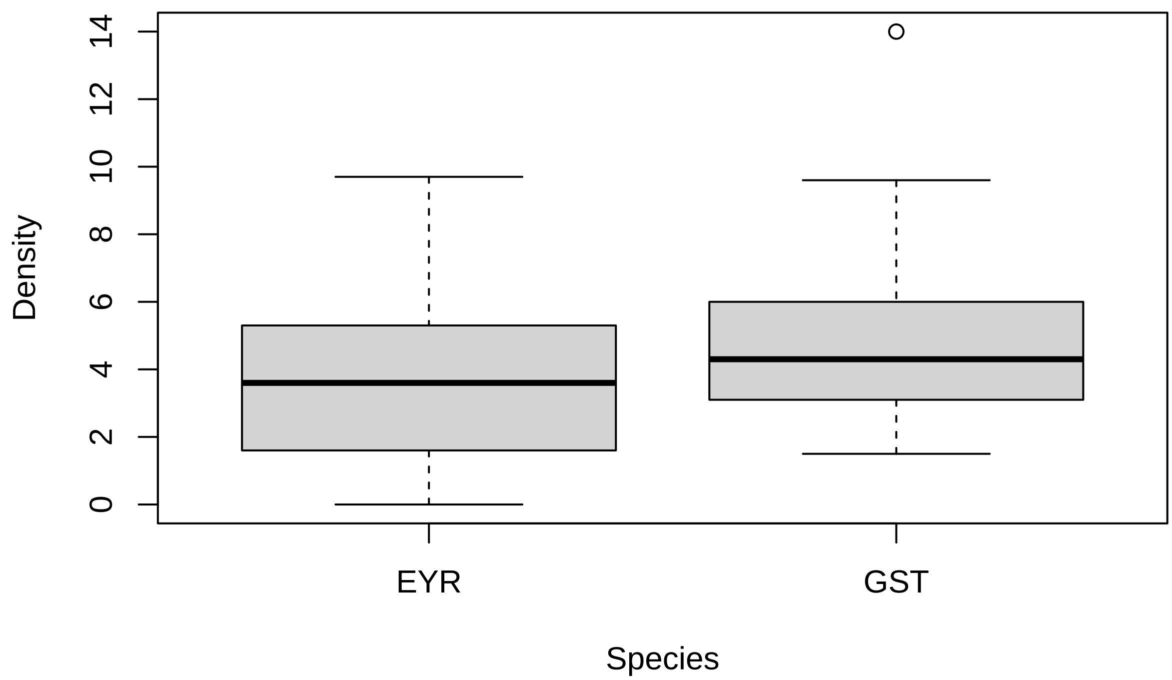
Case study 2.

Init

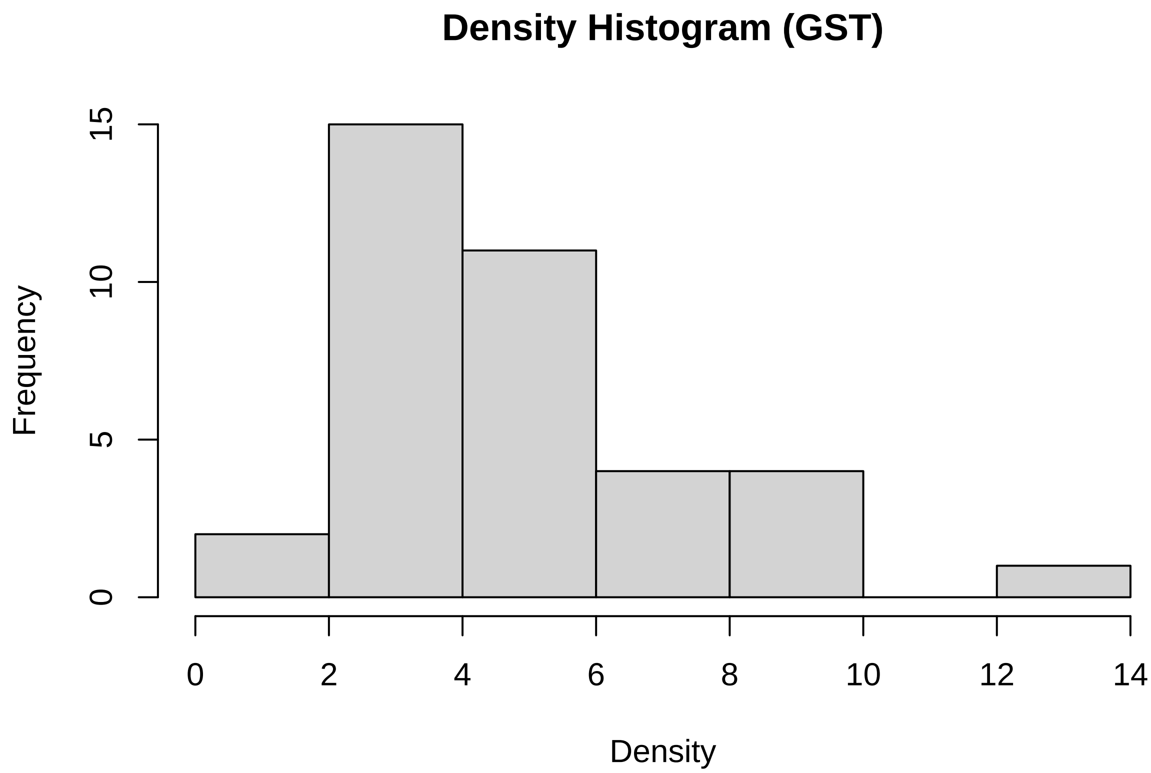
```r
macnally <- read_excel("/Users/kanoalindiwe/Downloads/Projects/playground/Datasets/macnally.xlsx")
```

1. EYR density is approximately normal, GST density is not normally distributed.

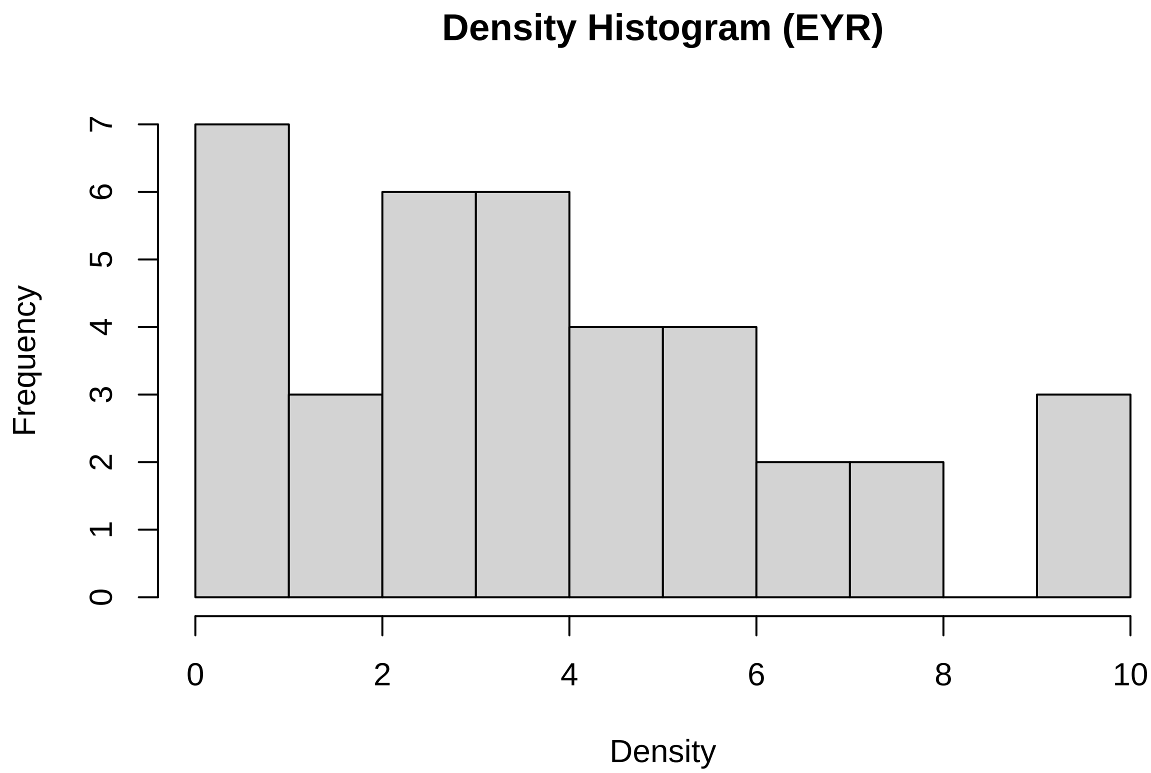
```r
boxplot(macnally$Density ~ macnally$Species, xlab="Species", ylab="Density")
```




```r
hist(macnally$Density[macnally$Species=="GST"], xlab="Density", main="Density Histogram (GST)")
```




```
hist(macnally$Density[macnally$Species=="EYR"], xlab="Density", main="Density Histogram (EYR)")
```



2. Two sided test: failed to reject null that mean is different from 5.7. Less than test: Successfully rejected null. Greater than test: Failed to reject null.

```r
t.test(macnally$Density[macnally$Species=="GST"], mu=5.7)
```

```
##
##  One Sample t-test
##
## data:  macnally$Density[macnally$Species == "GST"]
## t = -1.9765, df = 36, p-value = 0.0558
## alternative hypothesis: true mean is not equal to 5.7
## 95 percent confidence interval:
##  4.035292 5.721465
## sample estimates:
## mean of x
##  4.878378
```

```r
t.test(macnally$Density[macnally$Species=="GST"], mu = 5.7, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  macnally$Density[macnally$Species == "GST"]
## t = -1.9765, df = 36, p-value = 0.0279
## alternative hypothesis: true mean is less than 5.7
```

```
## 95 percent confidence interval:
##      -Inf 5.580211
## sample estimates:
## mean of x
##  4.878378
```

```r
t.test(macnally$Density[macnally$Species=="GST"], mu = 5.7, alternative = "greater")
```

```
##
##  One Sample t-test
##
## data:  macnally$Density[macnally$Species == "GST"]
## t = -1.9765, df = 36, p-value = 0.9721
## alternative hypothesis: true mean is greater than 5.7
## 95 percent confidence interval:
##  4.176546      Inf
## sample estimates:
## mean of x
##  4.878378
```

3. We fail to reject the null, there forfor we cannot say the data have significantly different means.

```r
eyr <- macnally$Density[macnally$Species == "EYR"]
gst <- macnally$Density[macnally$Species == "GST"]
t.test(eyr, gst, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  eyr and gst
## t = -1.8904, df = 72, p-value = 0.06273
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.39877179  0.06363665
## sample estimates:
## mean of x mean of y
##  3.710811  4.878378
```

```r
t.test(eyr, gst, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  eyr and gst
## t = -1.8904, df = 71.37, p-value = 0.06276
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.39895746  0.06382232
## sample estimates:
## mean of x mean of y
##  3.710811  4.878378
```

4. There is not significant difference in medians.

```
wilcox.test(eyr, gst, alternative = "two.sided")
```

```
## Warning in wilcox.test.default(eyr, gst, alternative = "two.sided"): cannot
## compute exact p-value with ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  eyr and gst
## W = 521.5, p-value = 0.07872
## alternative hypothesis: true location shift is not equal to 0
```

Case study 1 Questions:

1. Which function would you use to generate several basic summary stats for beak depth?
2. Basic summary stats of beak depth for the year 1976 are (hint, use tapply to get summary statistics by year):
3. Basic summary stats of beak depth for the year 1978 are:
4. Use the function "hist" to make histogram of finch beak depth for the years 1976 and 1978 (make separate histograms for each year). An appropriate description of these figures is:
5. Use the function "boxplot" to make a box plot of finch beak depth for the years 1976 and 1978 (show both years in a single figure containing two box plots). An appropriate description of this figure is:
6. Based on your results from the previous steps, choose a suitable statistical test for comparing beak size by year.
7. Use R to run a statistical test for comparing beak size by year. What is your statistical conclusion?
8. How should you describe the biological conclusion from your statistical test? If you chose Wilcox above now run Welch t test

Case study 2 Questions:

1. Examine assumptions of parametric tests by creating boxplots and histograms of both bird species. Be sure to add appropriate labels to each plot. Describe the data distribution.

2. Run a one sample t test for both a 2-sided and one-sided alternative (you specify the direction) comparing the mean for GST against a mu of 5.7. Report your results. Do you accept or reject the Null in each test?

3. Run a two sample t test comparing EYR to GST (first assume equal variances, then run another test assuming unequal variances). Report results for each. Do you accept or reject the Null?

4. Now run an appropriate non-parametric two-sample test