A4 - Q1 (Ram Yoogesh Gopu - 20867060)

READING THE DATA

```r
lizards <- read.csv("AnolisLizards.csv", header = TRUE)
```

Printing it

```r
lizards
```

```
##       species perch_height_ft perch_diameter_inches count
## 1      sagrei          ">= 3"              "<= 4.25"    15
## 2      sagrei          ">= 3"               "> 4.25"    18
## 3      sagrei           "< 3"              "<= 4.25"    48
## 4      sagrei           "< 3"               "> 4.25"    84
## 5 angusticeps          ">= 3"              "<= 4.25"    21
## 6 angusticeps          ">= 3"               "> 4.25"     1
## 7 angusticeps           "< 3"              "<= 4.25"     3
## 8 angusticeps           "< 3"               "> 4.25"     2
```

(A)

Archipelago of bimini, expecially bahaman fauna which contains all the lizzrd is the target population

Various species of lizards form the units.

(B)

(i)

Units are the lizards in the population. Study are the entire species of lizards (4 species), from the bahaman fauna.

(ii)

The variates are Species, perch_height_ft, perch_diameter_inches, count. Respose variate is species. Explonatory variates are perch_height_ft and perch_diameter_inches

(iii)

On a general note, sampling plan provides an outline on which the research is conducted. It gives us information on which category we should survey, what should be the sample size, and also how to choose the respondents from the population. So for our problem description, the sample plan is the collection of data about the lizards (Sagrei and angusticeps) and theier relevant details(perch_height_ft and perch_diameter_inches) which was recorded between 9:00AM to 4:40PM during the period of november.

(iv)

once we elicit the data about the lizards, namely perch_height_ft and perch_diameter_inches which are further divided into categorical variables (perch_height_ft " >=3 or <3" and perch_diamter_inches "<=4.25 or >4.25")

(v)

This is observational since we just observe the details of the lizards

(C)

(i)

```
lizardsTable <- xtabs(count ~ perch_diameter_inches + perch_height_ft + species, data = lizards)
lizardsTable
```

```
## , , species = angusticeps
##
##                     perch_height_ft
## perch_diameter_inches  "< 3"  ">= 3"
##           "<= 4.25"        3      21
##           "> 4.25"         2       1
```

```
## 
## , , species = sagrei
## 
##                      perch_height_ft
## perch_diameter_inches  "< 3"  ">= 3"
##             "<= 4.25"     48      15
##             "> 4.25"      84      18
```

```
lizardsTable[,1,]
```

```
##                      species
## perch_diameter_inches angusticeps sagrei
##             "<= 4.25"           3     48
##             "> 4.25"            2     84
```

```
lizardsTable[2,,]
```

```
##                species
## perch_height_ft angusticeps sagrei
##         "< 3"             2     84
##         ">= 3"            1     18
```

(ii)

```
sagrei <- sum(lizardsTable[,,2])
angusticeps <- sum(lizardsTable[,,1])

print("Sagrei sum")
```

```
## [1] "Sagrei sum"
```

```
print(sagrei)
```

```
## [1] 165
```

```
print("anguticeps sum")
```

```
## [1] "anguticeps sum"
```

```
print(angusticeps)
```

```
## [1] 27
```

```
# 1 and 2 is gonna provide the count of diameter and height
margin.table(lizardsTable, 3)
```

```
## species
## angusticeps      sagrei
##          27         165
```
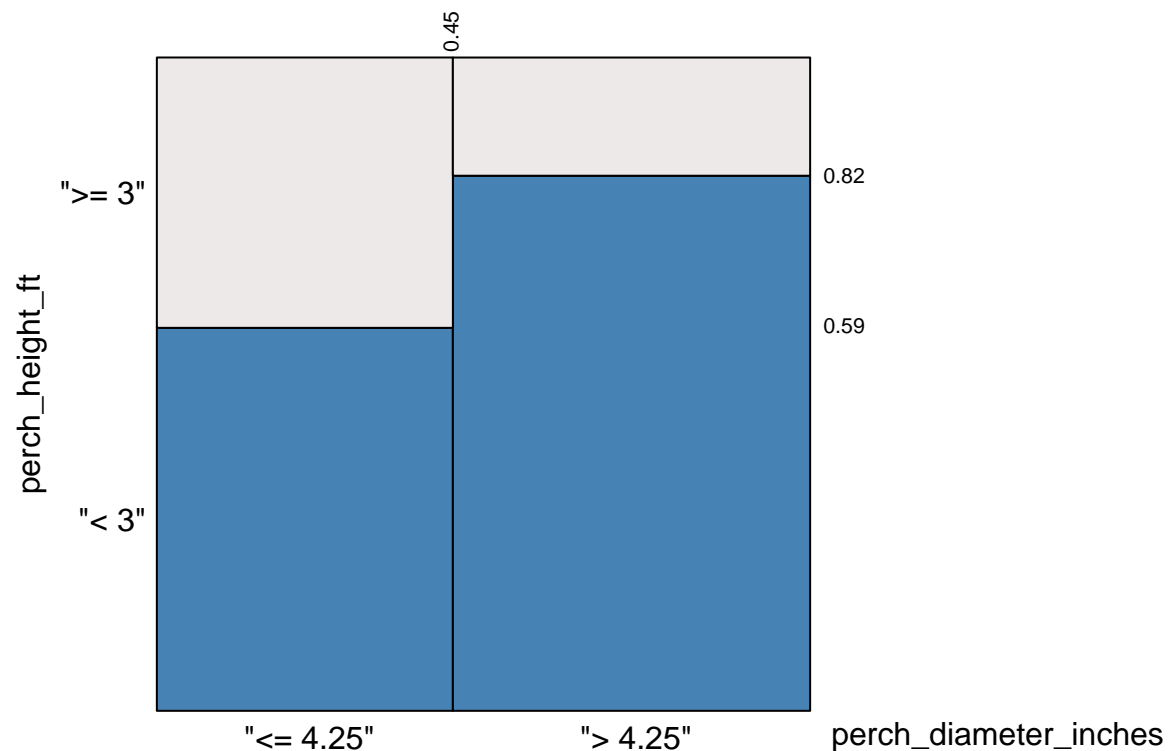
(d)

(i)

```r
library(eikosograms)
lizardsHabitat <- xtabs(count ~perch_diameter_inches + perch_height_ft, data = lizards)
lizardsHabitat
```

```
##                   perch_height_ft
## perch_diameter_inches  "< 3"  ">= 3"
##           "<= 4.25"      51      36
##           "> 4.25"       86      19
```

```r
eikos("perch_height_ft", "perch_diameter_inches" , data = lizardsHabitat)
```



```r
summary(lizardsHabitat)
```

```
## Call: xtabs(formula = count ~ perch_diameter_inches + perch_height_ft,
##     data = lizards)
## Number of cases in table: 192
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 12.62, df = 1, p-value = 0.0003817
```

From the graph we can infer that, the lizards whose height is > 4.25 will mostly have height <3. lizards which have height < 4.25 will mostly be less than 3ft. Also the data is more skewed towards the lizards whose height is greater than 4.25

(ii)

```
eikos("perch_height_ft", "perch_diameter_inches", data = lizardsTable[,,2])
```



```
summary(lizardsTable[,,2])
```
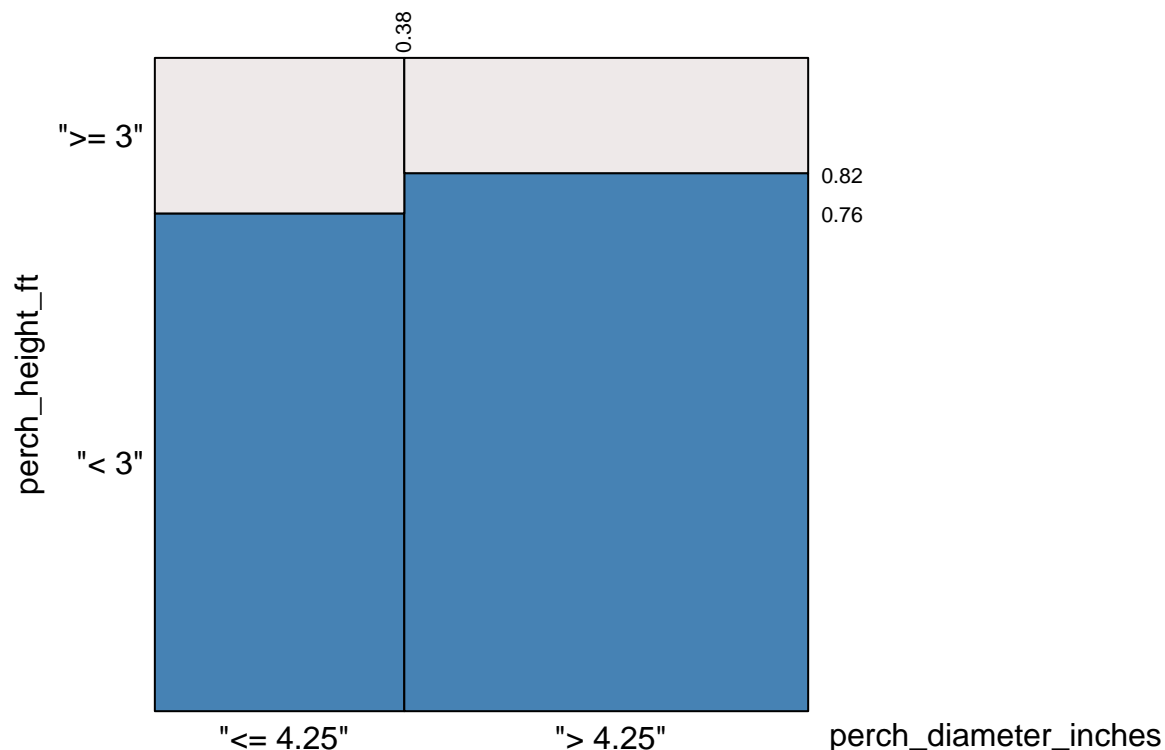
```
## Number of cases in table: 165
## Number of factors: 2
## Test for independence of all factors:
##   Chisq = 0.9244, df = 1, p-value = 0.3363
```
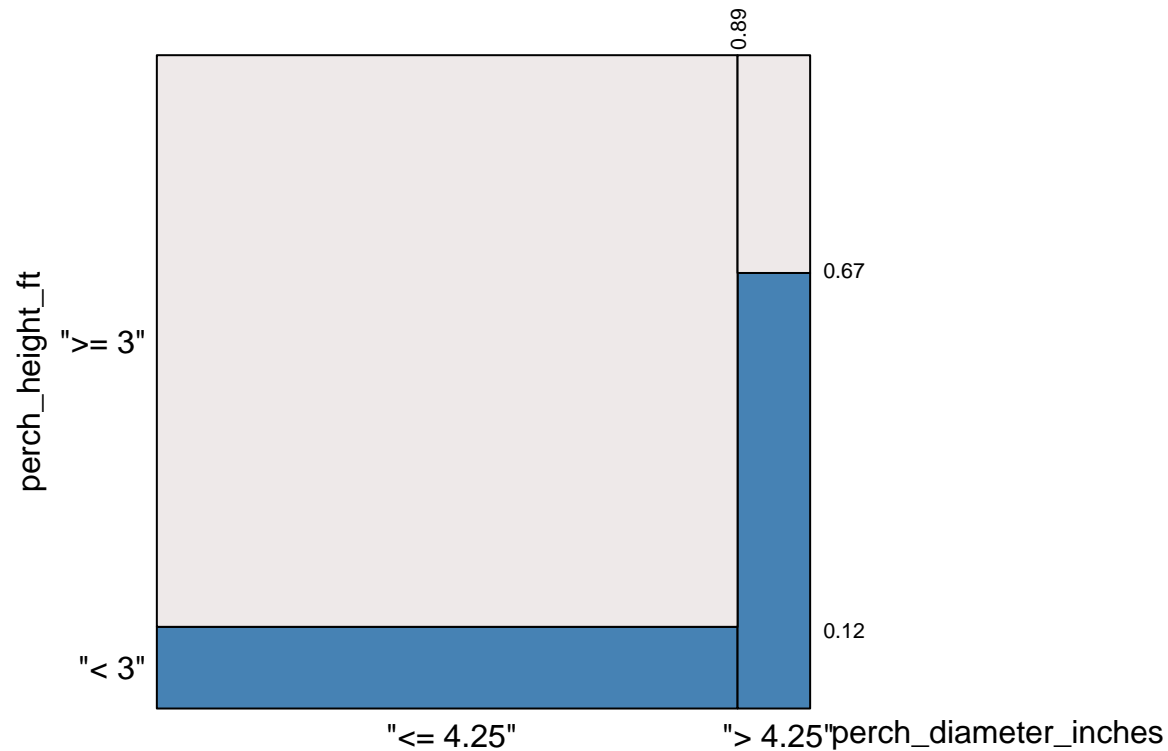
```
chisq.test(lizardsTable[,,2], simulate.p.value = TRUE)
```

```
##
##   Pearson's Chi-squared test with simulated p-value (based on 2000
##   replicates)
##
## data:  lizardsTable[, , 2]
## X-squared = 0.92437, df = NA, p-value = 0.4238
```

5

After performing chisq on species sageri we can see that the p-value is 0.423. This explains that both the perch height and perch diameter are independent. Also, the height of the lizards are more likely to be <3. But its advisable to choose the lizards > 4.25 in diameter

(iii)

```r
eikos("perch_height_ft",  "perch_diameter_inches", data = lizardsTable[,,1])
```



```r
summary(lizardsTable[,,1])
```

```
## Number of cases in table: 27
## Number of factors: 2
## Test for independence of all factors:
##   Chisq = 5.185, df = 1, p-value = 0.02278
##   Chi-squared approximation may be incorrect
```
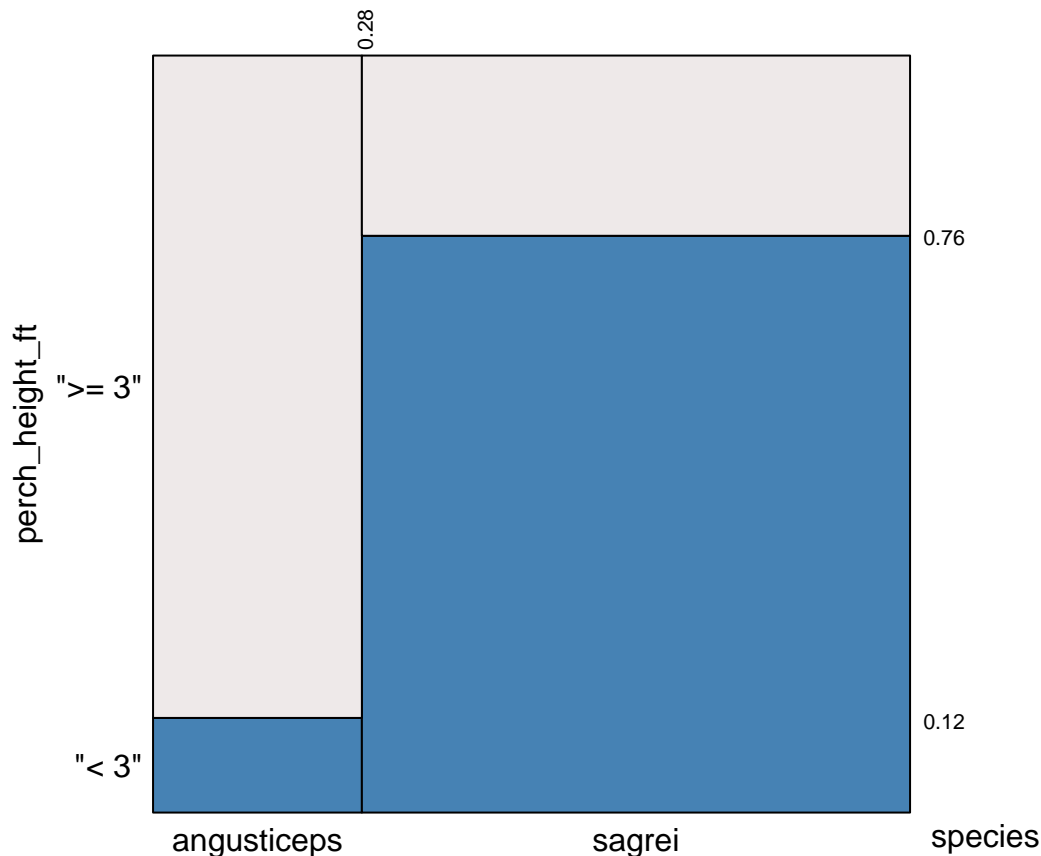
```r
chisq.test(lizardsTable[,,1], simulate.p.value = TRUE)
```

```
##
##   Pearson's Chi-squared test with simulated p-value (based on 2000
##   replicates)
##
## data:  lizardsTable[, , 1]
## X-squared = 5.1852, df = NA, p-value = 0.07246
```

This graph is more skewed towards the angusticeps lizards whose height >=3 and diameter >4.25. P values are kind of absurd because the pvalue is 0.02, which signifies that both the height and diameter are related to each other. Whereas p-value 0.07 (Which is just above the signifance level 0.5) defines that both the height and the diameter are indpendent of each other. So the interpretations doesnt agree.

(iv)

```r
eikos("perch_height_ft",  "species" , data = lizardsTable[1,,])
```



```r
summary(lizardsTable[1,,])
```

```
## Number of cases in table: 87
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 29.063, df = 1, p-value = 7.005e-08
```

```r
chisq.test(lizardsTable[1,,], simulate.p.value = TRUE)
```
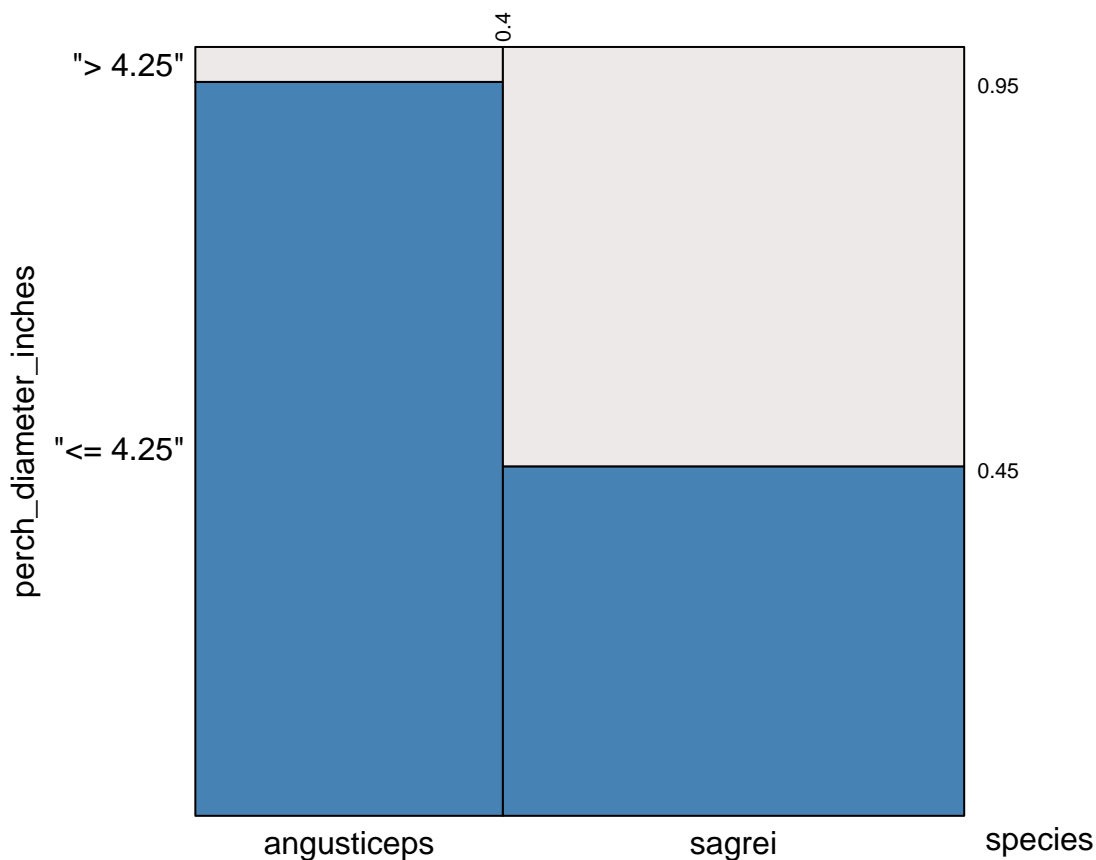
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
```

7

```
##  replicates)
##
## data:  lizardsTable[1, , ]
## X-squared = 29.063, df = NA, p-value = 0.0004998
```

the p-value is **7.005e-08**, explains that both the variates (diameter and species) are not inde-
pedent of each other. The xsquared p-value also proves the same.Also we can say that the
angusticeps lizards are mostly of height $>= 3$ and sargeri lizards are mostly of length < 3.

(V)

```
eikos("perch_diameter_inches", "species", data = lizardsTable[,2,])
```



```
summary(lizardsTable[,2,])
```

```
## Number of cases in table: 55
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 14.594, df = 1, p-value = 0.0001333
```

```
chisq.test(lizardsTable[,2,], simulate.p.value = TRUE)
```

```
## 
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
## 
## data:  lizardsTable[, 2, ]
## X-squared = 14.594, df = NA, p-value = 0.0004998
```

p-value is **0.001** (which is really low) signifies that both the variates (perch_height and species) are not indpendent of each other. x squared p-value is **0.00049** which is still less than **0.5**, also signifies the same that the variates are not independent of each other. Also we can say that the angusticeps lizards are mostly of height $>= 3$ and sargeri lizards are mostly of length < 3.

(E)

As explained previously that the angusticeps are mostly with height $>= 3$ and diameter < 4.25. May be this might be the reason that the lizard is named so.