EDA A4 Q1

Here we consider only a subset of data on two species of *Anolis* lizard: *Sagrei* and *Angusticeps*.

A total of 192 lizards were counted, together with their species and a categorization of their perch habitat.

The counts which fell into the various categories were assembled into the following table.

```
#file <- path_concat(dataDirectory, "AnolisLizards.txt")
#lizards <- read.table('/Users/rudranibhadra/Downloads/AnolisLizards.txt')
#lizards
```

Alternatively we could have read the data in using the ".csv" file.

```
#file <- path_concat(dataDirectory, "AnolisLizards.csv")
lizards <- read.csv('/Users/rudranibhadra/Downloads/AnolisLizards.csv')
#lizards
```

In either case, the data looks like

```
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
```

Of interest is whether the two species of lizard occupy different perch habitats.

a. The **Problem** stage.

b. *(3 marks)* What is the target population here? What are the units making up the population?

The target population is the population of 4 species, Anolis sagrei, distichus, angusticeps and carolinensis.

The units are the anoline lizards.

b. The **Plan** stage.

c. *(3 marks)* What is the study population? What are the units making up the population.

The study population are the lizards which were observed on the Blackland forest areas.

The units are the anoline lizards observed.

ii. *(3 marks)* What are the variates? Which are response? Which are explanatory?

The variates are species, perch_height_ft, perch_diameter_inches and count. species, perch_height_ft and perch_diameter_inches are explanatory variables. count is response variable.

iii. *(3 marks)* Describe the sampling plan.

The activites, perch height and perch diameter of the lizards which were seen during the period November 18 to 30, 1966, between 9:00 A.M. and 4:30 P.M were recorded.

iv. *(2 marks)* Describe the measuring system mentioned.

The perch height and perch diameter of the lizards were recorded. The presence of the lizard on leaves or on the ground rather than on branches was noted. If the lizard was moving, its position when first sighted only was used. If the movement was obvious escape behavior with reference to the observer, the observation was not counted.

v. *(2 marks)* Is this study experimental or observational? Explain your reasoning.

The study is observational as there are no experiments being done on the lizards. Just their activity and presence during certain times of the day are being observed and noted down.

c. The **Data** stage.

d. *(1 mark)* The `lizards` data is a data frame having 8 rows with the counts in one column. Another way to represent this data is as a $2 \times 2 \times 2$ table of counts. Use the `xtabs()` function with an appropriate formula to construct such a table from `lizards`. Assign the table to the variable `lizardsTable`.

Note that `lizardsTable` will not be a `data.frame` but a `table` and an `array`.

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

```
## , , perch_diameter_inches =  "<= 4.25"
##
##              perch_height_ft
## species         "< 3"  ">= 3"
##   angusticeps      3      21
##   sagrei          48      15
##
## , , perch_diameter_inches =  "> 4.25"
##
##              perch_height_ft
## species         "< 3"  ">= 3"
##   angusticeps      2       1
##   sagrei          84      18
```

ii. *(2 marks)* Using `lizardTables` and the `sum` function, programmatically determine how many lizards there were of each species. Show how the same results can be obtained with `margin.table()`.

```
apply(lizardsTable,1,sum)
```

```
## angusticeps      sagrei
##          27         165
```

```
margin.table(lizardsTable,margin=1)
```

```
## species
## angusticeps      sagrei
##          27         165
```

d. The **Analysis** stage.

e. *(10 marks)* Use `xtabs()` with an appropriate formula on the original `lizards` data frame to construct a smaller table that ignores species and just contains the counts at each perch habitat. Call this table `lizardsHabitat`. You will now do a small analysis based only on this table. That is,

- Complete the following:

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

```
##                perch_diameter_inches
## perch_height_ft  "<= 4.25"  "> 4.25"
```

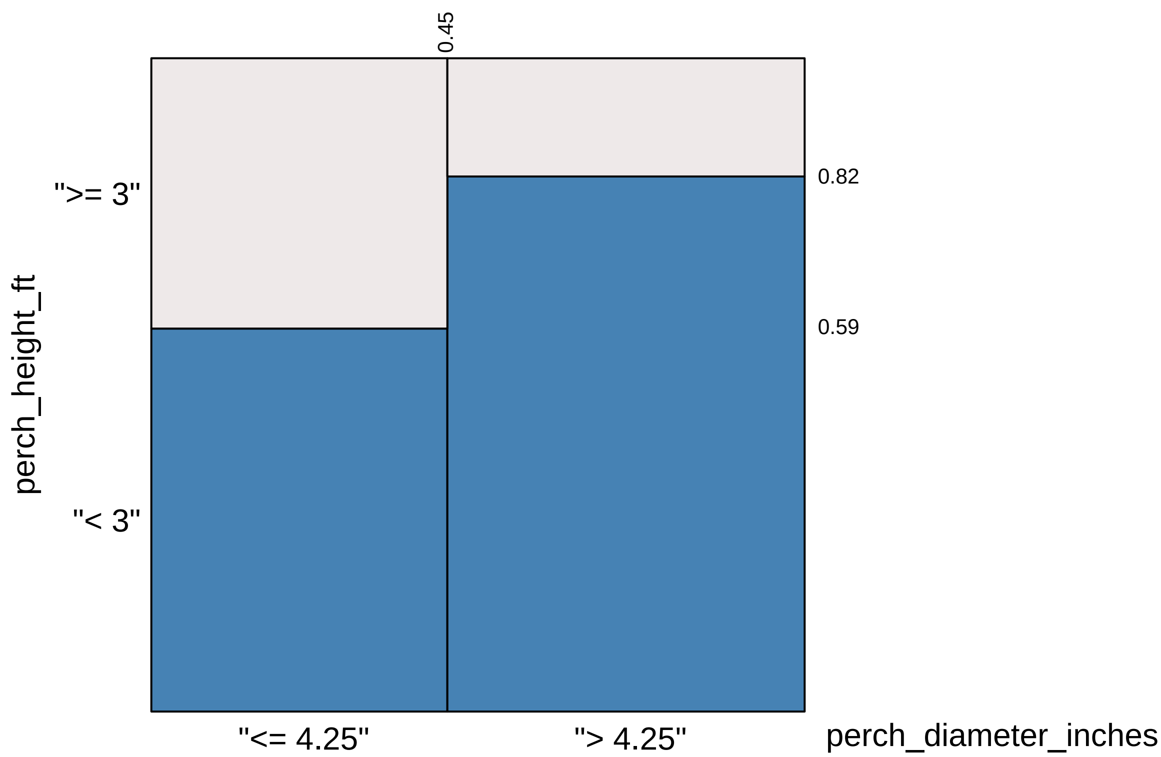
```
##          "< 3"          51         86
##          ">= 3"         36         19
```

- Construct an eikosogram of perch height versus perch diameter using the `lizardsHabitat` table.
- Call `summary(lizardsHabitat)`.
- Show your code.

Based on all of the above results, what can you conclude about the distribution of the perch of these lizards?

What does this say about the preferred habitat of these lizards? Explain.

```r
library(eikosograms)
eikos(perch_height_ft~ perch_diameter_inches, data = lizardsHabitat)
```




```r
summary(lizardsHabitat)
```

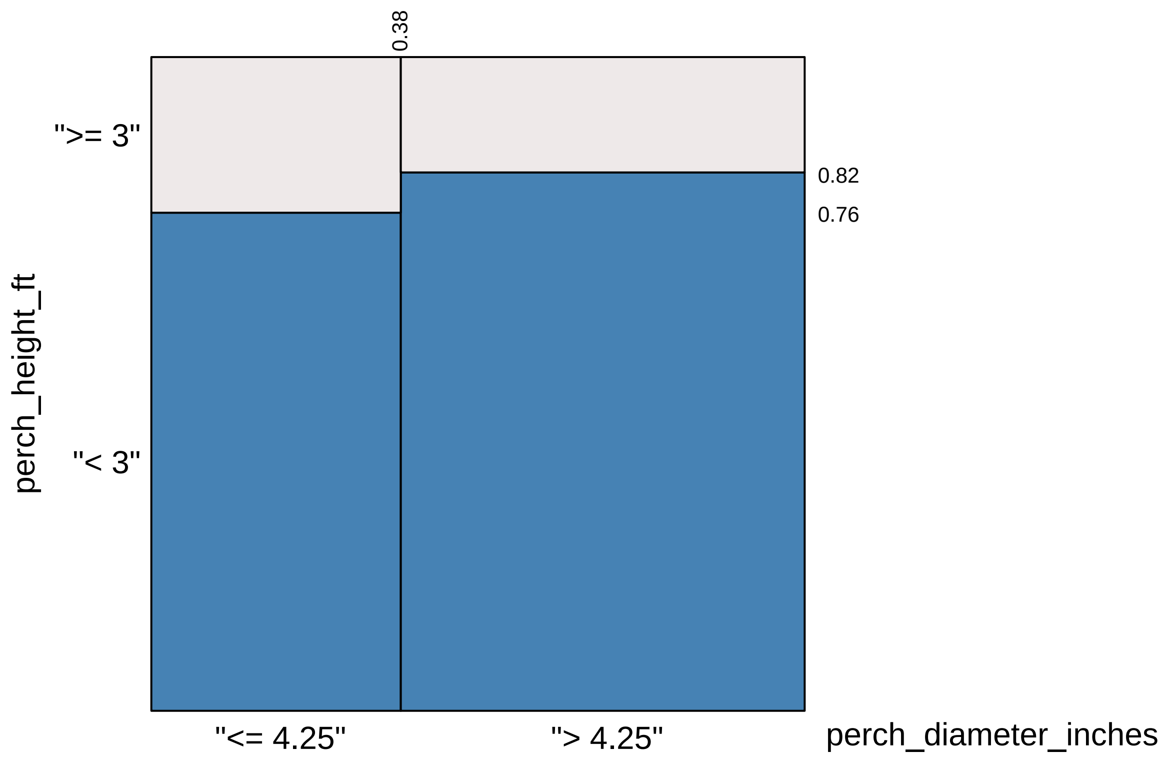
```
## Call: xtabs(formula = count ~ (perch_height_ft + perch_diameter_inches),
##     data = lizards)
## Number of cases in table: 192
## Number of factors: 2
## Test for independence of all factors:
##  Chisq = 12.62, df = 1, p-value = 0.0003817
```

Here we can see that perch_height is dependent on perch_diameter. When the perch_diameter is more than 4.25, the perch_height is more likely to be less than 3 as compared to when perch_diameter is less than and equal to 4.25.

This means that the lizards are more likely to be in habitats at a height of <3 when perch diameter is $>$ 4.25.

ii. *(6 marks)* Repeat the analysis of part (i) above but this time on the sub-table of `lizardsTable` (i.e. no `xtabs()` call) of only the counts for `sagrei`. In addition to calling `summary()`, also call `chisq.test()` on the relevant table **but** use a significance level determined by simulation.

```
library(eikosograms)
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)
```

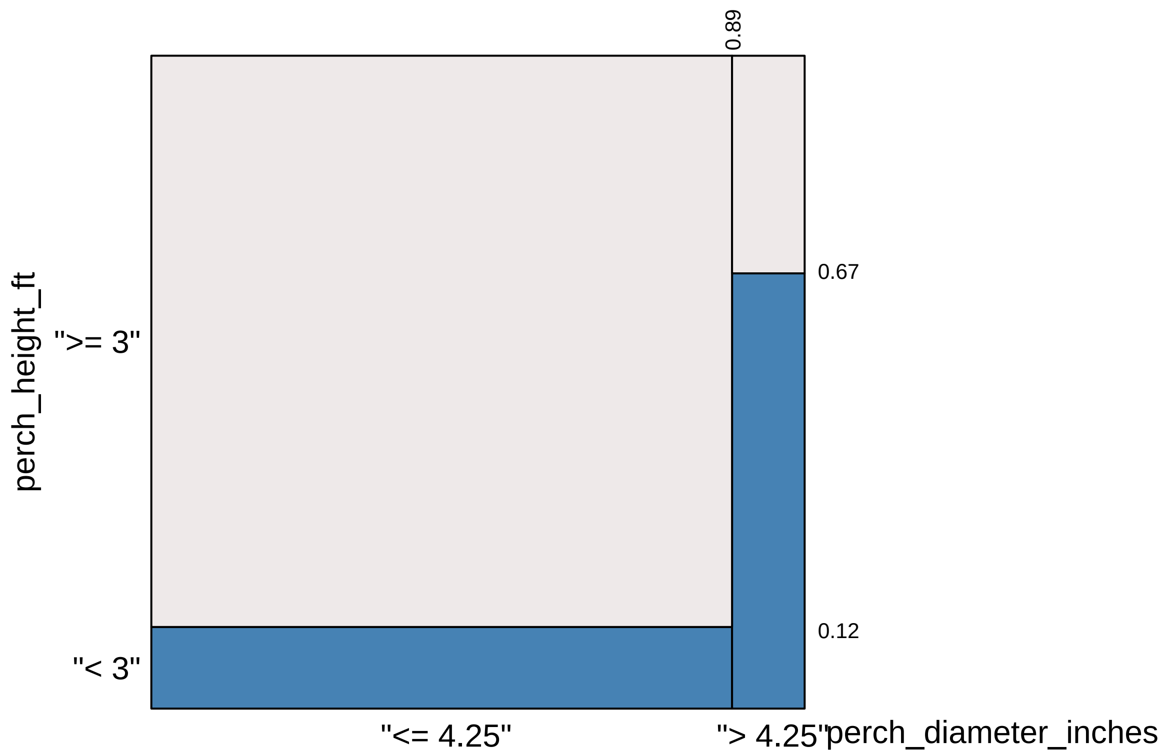
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  lizardsTable[2, , ]
## X-squared = 0.92437, df = NA, p-value = 0.4338
```

iii. *(6 marks)* Construct the sub-table of `lizardsTable` (i.e. no `xtabs()` call) containing only the counts for `angusticeps`. Using this table, construct the eikosogram of perch height versus perch diameter. Also call `chisq.test()` on this table with a significance level determined by simulation.

- Do the interpretations from the eikosogram and the chi-squared tests agree? How do they agree/disagree?

- Explain why this has occurred here.

```r
library(eikosograms)
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 = 0.08796
```

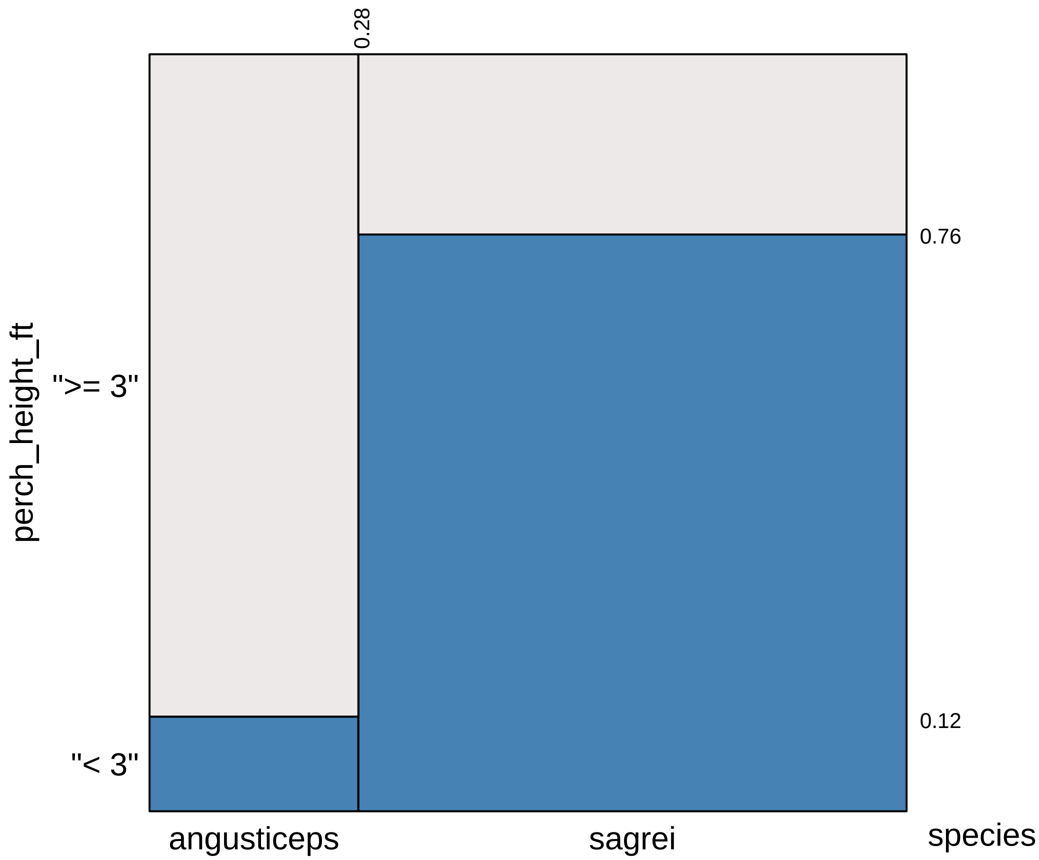
The interpretations from the eikosogram and the chi-squared tests disagree. Even though the Chisq value from the eikosogram and the X-squared value from the Chi-squared test are equal, the p values are different. The p value from the chi-square test is greater than the p value from the eikos summary and also greater than 0.05 which seems evidence for the null hypothesis.

The p values are different since the number of observations in the table are too less to show any evidence against the null hypothesis.

iv. *(4 marks)* Construct the sub-table of `lizardsTable` (i.e. no `xtabs()` call) containing only the counts for lizards on perches that are less than 4.25 inches in diameter. Using this table, construct the eikosogram of perch height versus species. Also call `chisq.test()` on this table with a significance level determined by simulation.

What do you conclude, if anything, about the preference in habitat for the two species? Justify your conclusions.

```r
library(eikosograms)
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)
```

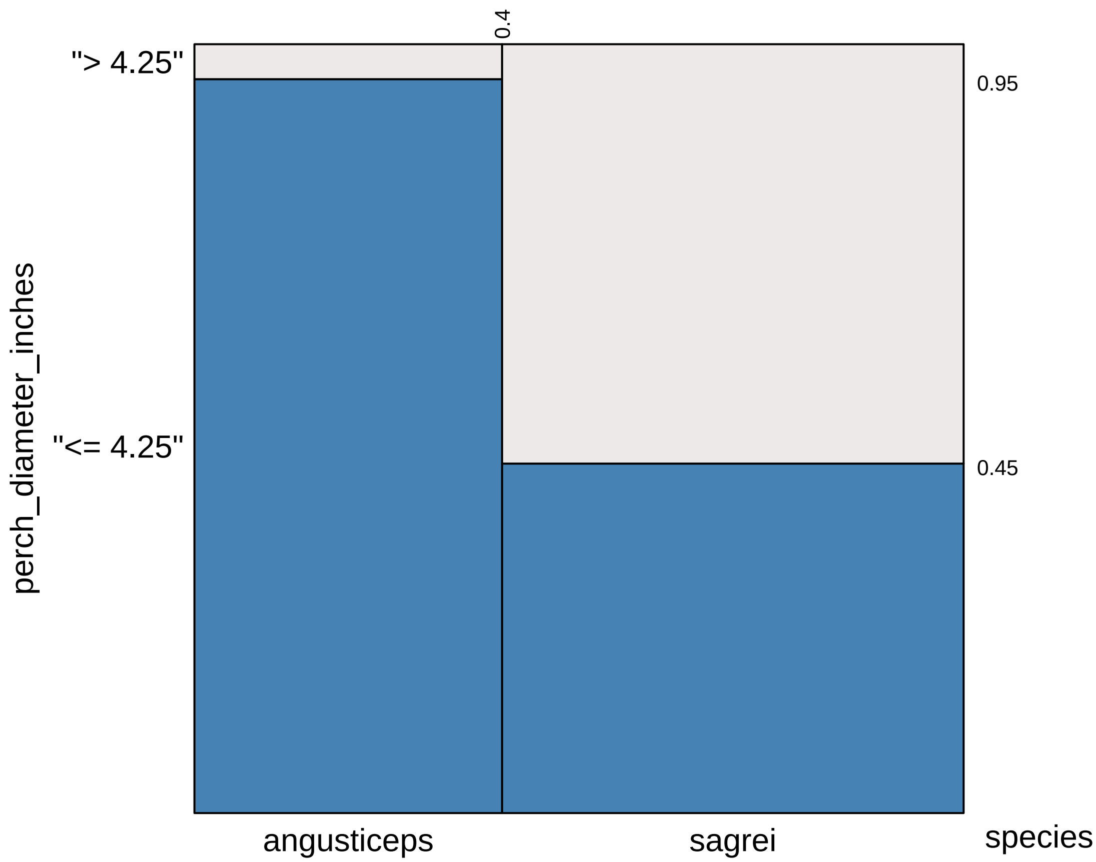
```
##
##  Pearson's Chi-squared test with simulated p-value (based on 2000
##  replicates)
##
## data:  lizardsTable[, , 1]
## X-squared = 29.063, df = NA, p-value = 0.0004998
```

Given the perch diameter less than 4.25 inches, perch height is dependent on the species. From the eikosogram, we can see that the likelihood of perch height is less than 3 is more for lizards belonging to the sagrei species. Whereas, the likelihood of perch height more than or equal to 3 is more for the angusticeps lizards.

This means the most sagrei lizards prefer to stay at a height less than 3 when the diameter is less than 4.25 inches. The angusticeps lizards prefer to stay at a height more than or equal to 3.

v. *(4 marks)* Construct the sub-table of `lizardsTable` (i.e. no `xtabs()` call) containing only the counts for lizards on perches that are three feet or higher in height. Using this table, construct the eikosogram of perch diameter versus species. Also call `chisq.test()` on this table with a significance level determined by simulation.

```
library(eikosograms)
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 = 0.0009995
```

e. The **Conclusions** stage:

f. *(2 marks)* A common name for *anolis angusticeps* is the *Cuban Twig Anole*. In light of your analysis explain why, or why not, this name might have become common?

From all the analysis above , we observe that anolis angusticeps lizards prefer in habitats having perch heights more than 3 feet and perch diameter less than 4.25 inches. So this means they prefer to stay in places which are high and narrow. So thats why they are also called Cuban Twig Anole since twigs are slender branches in trees which grow at a height much above the ground.