Final Analysis by Aaron Mendelson

Suburban areas play an integral role in the development of sustainable cities; however, developers often do not consider sustainability in the construction of subdivisions and the subsequent adoption of homeowner's association (HOA) covenants. While there are multiple actions homeowners can take to contribute to personal sustainability on the plot-by-plot level, these actions are not always adopted or supported by greater neighborhood norms.

The current literature provides assessments of individual sustainability indicators at the homeowner and neighborhood level as well as multi-indicator sustainability assessments of cities and larger metropolitan areas but lacks such multi-indicator analyses at the homeowner and neighborhood level. This study assesses the relationship among multiple sustainability indicators of homeowner behavior including recycling habits, lawn care, tree planting, and home gardening, and compares these behaviors between neighborhoods with HOAs and those without. Data metrics were collected from twelve neighborhoods in Greenville, South Carolina through on-site observation, analysis of Google Earth images, and qualitative assessment of HOA covenants.

Use the data collected by the researcher to extract information required for their study. The data consist of 1,616 observations of homes in Greenville, SC and the seven variables recorded for each. Basic descriptions of the variables and other important information can be found below.

- **Neighborhood:** This reports which of the 12 neighborhoods in Greenville, SC each observed home is located. There are no missing values for this variable.
- **Lot Number:** This reports the lot number of the homes. There are no missing values for this variable.
- **HOA:** This reports whether or not the homes are part of a homeowners' association (1 = yes; 2 = no). There are no missing values for this variable.
- **Recycle:** This reports the recycling status of the homes (1 = both recycling bin and trash bin present at the home; 2 = only a trash bin was present). There are 274 missing values for this variable. These missing values correspond to neighborhoods with no curb side pick up (i.e., Brownstone, Edgewood, Glastonbury, and Fox Springs).
- **Lawn Care:** This represents a likert variable on a scale of 1 to 4 (1 = excellent; 2 = good; 3= poor; 4 = none). This measure maps onto how artificially managed the lawn is where 1 means the lawns were highly artificially managed with presumed regular chemical application and 4 means the lawns were naturally managed (i.e., not managed at all). There are no missing values for this variable.
- **Trees:** This represents the number of trees in the front yard. There are 14 missing values for this variable.
- **Garden:** This represents whether a garden was present in the front or back of the homes (1 = yes; 2 = no). There are no missing values for this variable.

The data can be accessed in `R` as follows.

```
> dat.HOA<-read.table("https://cipolli.com/students/data/Exam1Data.txt",
+                header=T,sep=",")
```

Solution: The study asks two questions, determine whether or not neighborhoods that have an HOA practice sustainability better than neighborhoods without it, and whether the sustainability factors effect each other in any way. First we provide some supplementary data collected by (Google (2019), HOA (2019), and On-Site Observation) before we do an analysis of both questions, the analysis was conducted in R (R Core Team (2019)):

```
> library("ggplot2")
> library("xtable")
> library(syuzhet)
> library(tm)
> dat.HOA<-read.table("https://cipolli.com/students/data/Exam1Data.txt",
+                     header=T,sep=",")
> ## Label the Data for easier graphing ##
> dat.HOA$HOA <-factor(dat.HOA$HOA, labels = c("Yes","No"))
> dat.HOA$Garden <-factor(dat.HOA$Garden, labels = c("Yes","No"))
> dat.HOA$LawnCare <- factor(dat.HOA$LawnCare, labels = c("Excellent", "Good","Poor","None"))
> ################################
> ##### Supplementary Data
> ################################
> tab<- table(droplevels(dat.HOA$HOA), droplevels(dat.HOA$Neighborhood))
> data1<- data.frame(tab)
```

We will use the package "xtable" (Dahl et al. (2019)), to give the LaTeXcode for the data frame generated above, we will use this package for all tables in the analysis:

```
> xtable(data1, type = latex, file = "supplementary.tex")
```

The LaTeXcode we get from this package results in the following table:

We will use the "ggplot2" package (Wickham (2016)) for all the graphs in the analysis:

```
> p1 <- ggplot(data = data1, aes(x = data1$Var2, fill = data1$Var1))+
+   geom_bar(aes(y=data1$Freq),stat = "identity")+
+   xlab("Neighborhood")+
+   ylab("Number of Homes in Study")+
+   ggtitle("Number of Homes per Neighborhood", subtitle = "Divided by HOA Status")+
+   coord_flip()+
+   scale_fill_discrete(name = "HOA Status")
>
```

Neighborhood Name	HOA Status	Count
Brownstone Crossing	Yes	57
Brownstone Crossing	No	0
Buxton	Yes	0
Buxton	No	102
Croftstone Acres	Yes	0
Croftstone Acres	No	207
Edgewood at Paris	Yes	63
Edgewood at Paris	No	0
Fox Springs	Yes	0
Fox Springs	No	33
Glastonbury Village	Yes	125
Glastonbury Village	No	0
Half Mile Lake	Yes	376
Half Mile Lake	No	0
Liberty Park	Yes	0
Liberty Park	No	77
Northcliff	Yes	285
Northcliff	No	0
Partridge Ridge	Yes	153
Partridge Ridge	No	0
Timberlake	Yes	0
Timberlake	No	94
Windermere	Yes	0
Windermere	No	44

Table 1: This table displays the neighborhoods that have an HOA and the neighborhoods that don't.

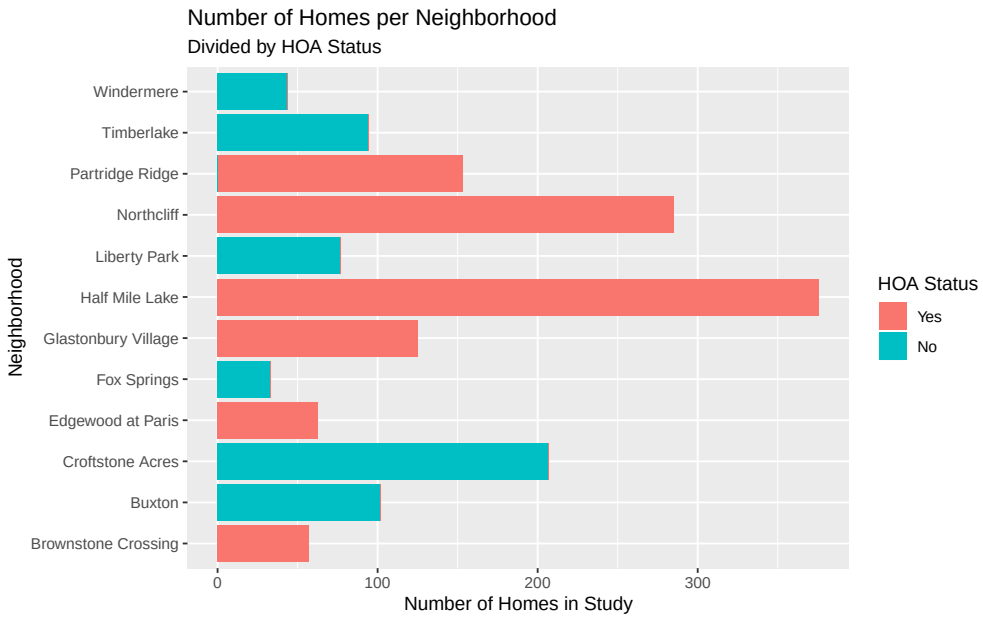

Figure 1: Graph showing how many homes per neighborhood were sampled.

```
> tab <- table(droplevels(dat.HOA$HOA))
> data2<-data.frame(tab)
```

```
> xtable(data2, type = latex, file = "population.tex")
```

HOA Status	Count
Yes	1059
No	557

Table 2: This table displays how many homes sampled were HOA or not.

```
> p2 <- ggplot(data = data2, aes(x = data2$Var1))+
+   geom_bar(aes(y=data2$Freq),stat = "identity")+
+   xlab("HOA Status")+
+   ylab("Number of Homes in Study")+
+   ggtitle("Number of HOA Homes")+
+   geom_hline(yintercept = 0)
```

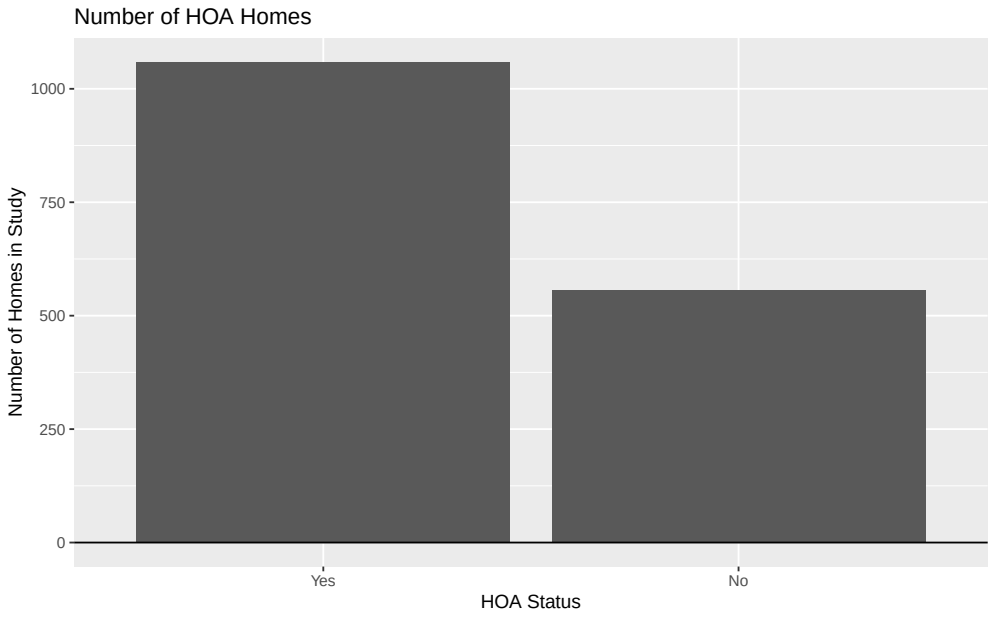

Figure 2: Graph showing how many homes were a part of a HOA.

We can now see the make up of the data, and see by Figure and Table 1 that each neighborhood is either entirely HOA or entirely not. To see any difference in the practices of HOA vs. Non-HOA homes we cannot look on a neighborhood-by-neighborhood basis, but we must look at it from a whole. In Figure and Table 2 we can see that there are nearly double the amount of HOA houses compared to Non-HOA houses, giving us a larger sample size to work with.

Analyzing Sustainability of HOA Houses

Recycling

We now look at the habits of HOA vs. Non-HOA houses, starting with recycling. We can grab data on homes that recycle (getting rid of those with NA) with:

```
> ################################
> ##### Recycling
> ################################
```

```
> dat.Recycle <- dat.HOA
> dat.Recycle <- dat.Recycle[which(!is.na(dat.Recycle$Recycle)),]
> dat.Recycle$Recycle<-factor(dat.Recycle$Recycle, labels = c("Both Trash/Recycle", "Only Trash"))
> tab <- table(droplevels(dat.Recycle$HOA),droplevels(dat.Recycle$Recycle))
> tab <- round(prop.table(tab,margin=1)*100,4)
> data3 <- data.frame(tab)
```

Note: This removes around 274 values that were listed as NA.

```
> xtable(data3, type = latex, file = "recycling.tex")
```

HOA Status	Recycling Status	Percentage(%)
Yes	Both Trash/Recycle	72.75
Yes	Only Trash	27.25
No	Both Trash/Recycle	74.90
No	Only Trash	25.10

Table 3: This table shows the proportion of the recycling habits of HOA vs. Non-HOA homes.

```
> p3 <- ggplot(data = data3, aes(x=data3$Var2, y = data3$Freq, fill = data3$Var1))+
+   geom_bar(position= "dodge", stat = "identity")+
+   xlab("Recycling Status of House")+
+   ylab("Percentage (%)")+
+   scale_fill_discrete(name = "HOA Status")+
+   ggtitle("Frequency of Recycling Habits", subtitle = "Based on HOA Membership")+
+   geom_hline(yintercept = 0)
```

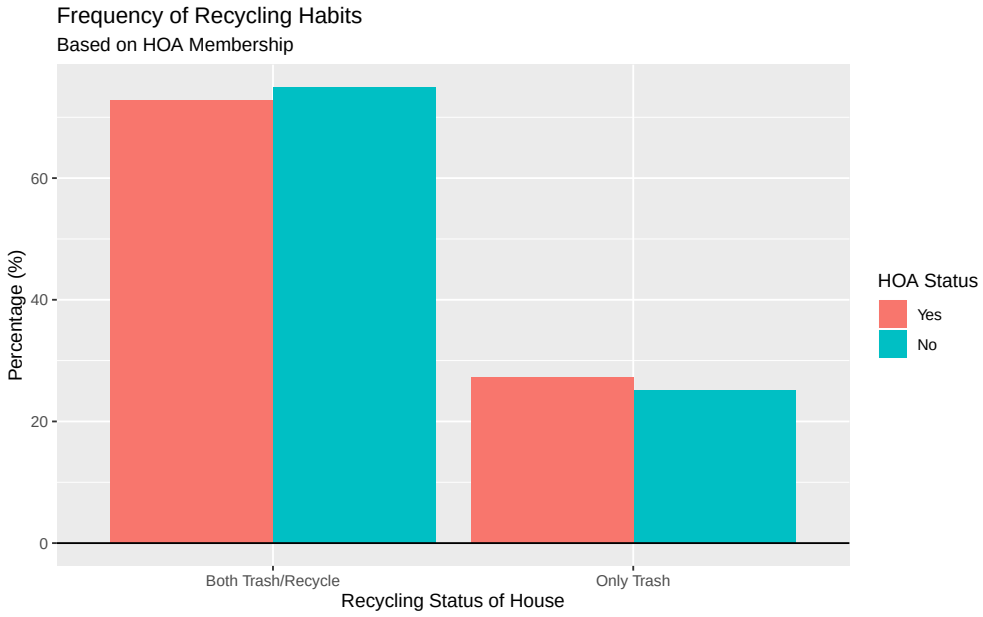

Figure 3: Graph showing the recycling habits of homes based on HOA status.

We want to see if the proportions of homes with both trash and recycling are different across HOA status. We can run a two sample hypothesis test to give more information. The hypotheses we will test are:

$H_0 : p_{HOA} - p_{NHOA} = 0$
$H_a : p_{HOA} - p_{NHOA} \neq 0$
We will make the assumption that each home is independent, checking for the Z distribution to hold with values from tab before its turned into a percentage:

```
> HRecyc <- 614
> HTrash <- 230
> HSum <- HRecyc + HTrash
> NHRecycl <- 373
> NHTrash <- 125
> NHSum <- NHRecycl + NHTrash
```

We can see that for each sample of non-HOA and HOA homes $\hat{p}_i n_i \geq 5$ and $(1 - \hat{p}_i)n_i \geq 5$. We will test with:

```
> prop.test(x = c(HRecyc,NHRecycl), n = c(HRecyc+HTrash,NHRecycl+NHTrash),conf.level = 0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(HRecyc, NHRecycl) out of c(HRecyc + HTrash, NHRecycl + NHTrash)
X-squared = 0.63824, df = 1, p-value = 0.4243
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.07160711  0.02859145
sample estimates:
   prop 1    prop 2
0.7274882 0.7489960
```

Because we have a non-significant p-value, we fail to reject H_0, and indication that the proportions of recycling homes in HOA and non-HOA are the same.

Lawn Care
We now examine the lawn care habits of homes:

```
> ################################
> ##### Lawn Care
> ################################
> dat.lawnCare <- dat.HOA
> tab <- table(droplevels(dat.lawnCare$HOA),droplevels(dat.lawnCare$LawnCare))
> tab <- round(prop.table(tab,margin=1)*100,4)
> data4 <- data.frame(tab)
```

Note: There were no values listed as NA.

```
> xtable(data4, type = latex, file = "lawn care.tex")
```

```
> p4 <- ggplot(data = data4, aes(x=data4$Var2, y = data4$Freq, fill = data4$Var1))+
+   geom_bar(position= "dodge", stat = "identity")+
+   xlab("Lawn Care of House")+
+   ylab("Percentage (%)")+
+   scale_fill_discrete(name = "HOA Status")+
+   ggtitle("Frequency of Lawn Care Habits", subtitle = "Based on HOA Membership")+
+   geom_hline(yintercept = 0)
```

HOA Status	Lawn Care Habit	Percentage(%)
Yes	Excellent	35.60
No	Excellent	19.75
Yes	Good	23.61
No	Good	35.91
Yes	Poor	18.32
No	Poor	32.50
Yes	None	22.47
No	None	11.85

Table 4: This table shows the proportion of the lawn care habits of HOA vs. Non-HOA homes.

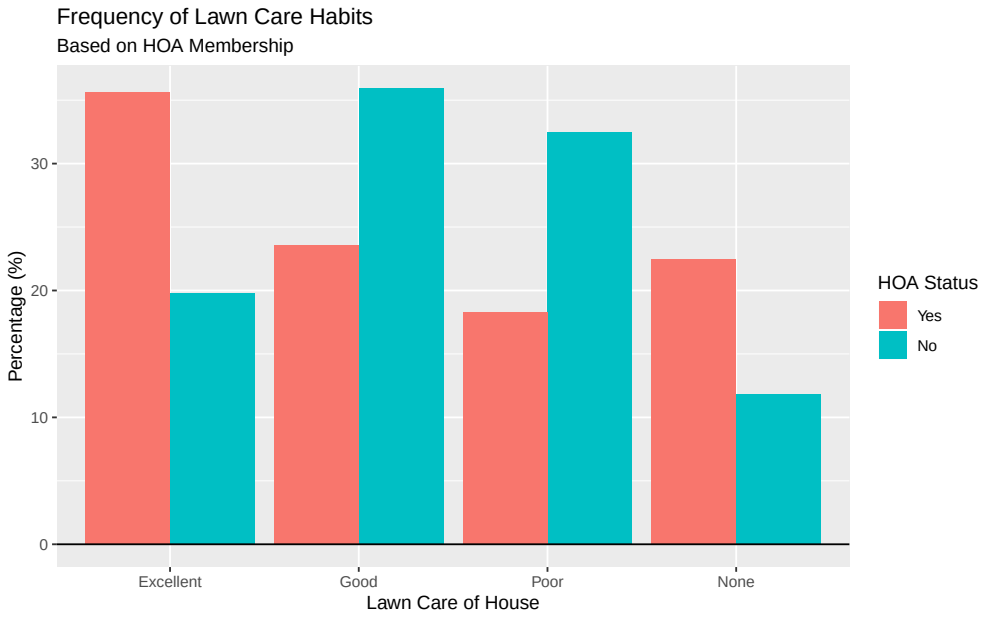

Figure 4: Graph showing the lawn care habits of homes based on HOA status.

By the description of lawn care given in the problem, we know that homes with habits of "Excellent" and "Good" are indicative of the regular use of harmful chemicals, that are bad for the environment. We want to see if the proportion of lawn care habits of "Excellent"/"Good" and "Poor"/"None" are different between HOA homes and non HOA homes. The hypotheses with will be testing are:
$H_0 : p_{HOA} - p_{NHOA} = 0$
$H_a : p_{HOA} - p_{NHOA} \neq 0$
We will make the assumption that each home is independent, checking for the Z distribution to hold with values taken from tab before its turned into a percentage:

```
> HExcellent <- 377
> HGood <- 250
> HPoor <- 194
> HNone <- 238
> HSum <- (HExcellent+HGood+HPoor+HNone)
> NHExcellent <- 110
> NHGood <- 200
> NHPoor <- 181
> NHNone <- 66
> NHSum <- (NHExcellent+NHGood+NHPoor+NHNone)
```

We can see that for each sample of non-HOA and HOA homes $\hat{p_i}n_i \geq 5$ and $(1-\hat{p_i})n_i \geq 5$. We will test with:

```
> prop.test(x = c(HExcellent + HGood, NHExcellent + NHGood),
+           n = c(HSum,NHSum),
+           conf.level = 0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(HExcellent + HGood, NHExcellent + NHGood) out of c(HSum, NHSum)
X-squared = 1.7468, df = 1, p-value = 0.1863
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.01663104  0.08766109
sample estimates:
  prop 1   prop 2
0.592068 0.556553

> prop.test(x = c(HPoor + HNone, NHPoor+NHNone),
+           n = c(HSum, NHSum),
+           conf.level = 0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(HPoor + HNone, NHPoor + NHNone) out of c(HSum, NHSum)
X-squared = 1.7468, df = 1, p-value = 0.1863
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.08766109  0.01663104
sample estimates:
  prop 1   prop 2
0.407932 0.443447
```

For each test, we have a non-significant p-value (> 0.05). Therefore, we fail to reject H_0 and have evidence to show that the proportions are the same. However, if we test to see if the proportion of homes with "Excellent" habits is different across HOA and non-HOA:

```
> prop.test(x = c(HExcellent, NHExcellent),
+           n = c(HSum,NHSum),
+           conf.level = 0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(HExcellent, NHExcellent) out of c(HSum, NHSum)
X-squared = 42.81, df = 1, p-value = 6.033e-11
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1132689 0.2037505
sample estimates:
   prop 1    prop 2
0.3559962 0.1974865
```

A significant p-value (< 0.05), gives us evidence to reject H_0. We say with 95% confidence that the proportion of homes with "Excellent" habits is greater within HOA homes.
By the tests performed, we can conclude that while overall the proportion of home with habits that are damaging to the environment are the same, however there is a greater proportion of HOA homes with "Excellent"

habits than non HOA homes.

Tree Planting

We now move over to the tree count of homes:

```
> ################################
> ##### Tree Planting
> ################################
> library(RVAideMemoire)
> library(rcompanion)
> dat.tree <- dat.HOA
> dat.tree <- dat.tree[which(!is.na(dat.tree$Trees)),]
```

Note: This removes 14 values that were listed as NA.

```
> summary(dat.tree[which(dat.tree$HOA == "Yes"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   2.947   4.000  35.000

> round(sd(dat.tree[which(dat.tree$HOA == "Yes"),]$Trees),4)

[1] 2.6682

> summary(dat.tree[which(dat.tree$HOA == "No"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   5.000   5.896   8.000  40.000

> round(sd(dat.tree[which(dat.tree$HOA == "No"),]$Trees),4)

[1] 5.1724
```

Numerical summaries for the amount of trees of HOA(top) versus non-HOA(bottom) homes, along with the standard deviation.

```
> data5 <- data.frame(tab)
> p5 <- ggplot(data = dat.tree, aes(x=dat.tree$Trees, fill = dat.tree$HOA))+
+   geom_density(alpha = 0.2)+
+   xlab("Trees on Property")+
+   ylab("Density")+
+   scale_fill_discrete(name = "HOA Status")+
+   ggtitle("Density of Trees", subtitle = "Based on HOA Membership")+
+   geom_hline(yintercept = 0)
```

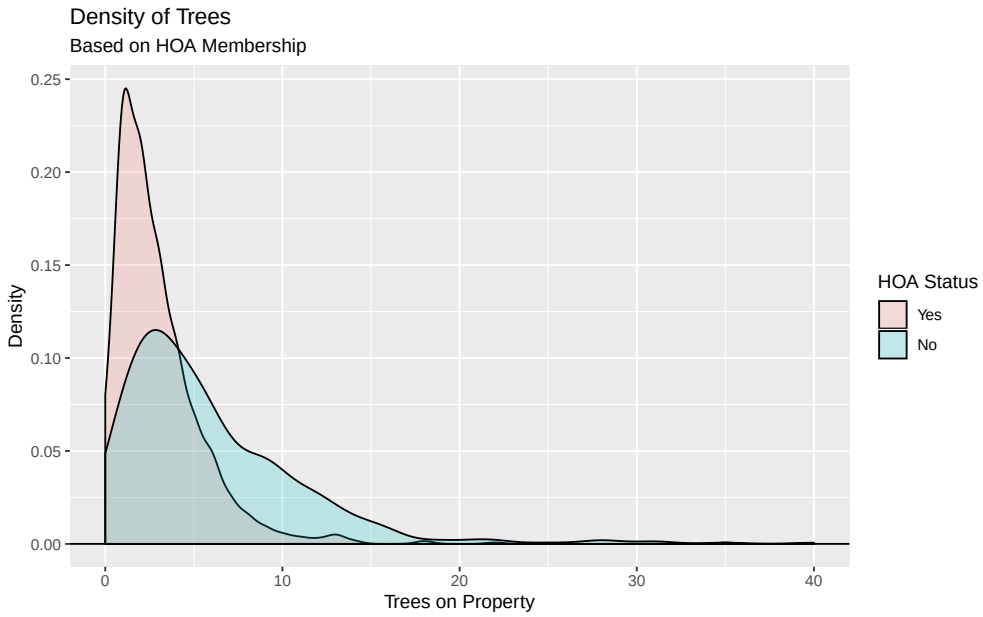

Figure 5: Graph showing the amount of trees owned by homes based on HOA status.

Based on the skew of the data, we will test to see if there is a difference in medians of trees on a property across HOA status. The hypotheses we will test are:
$H_0 : M_{HOA} - M_{NHOA} = 0$
$H_a : M_{HOA} - M_{NHOA} \neq 0$

Using the RVAideMemoire and "rcompanion" (Mangiafico (2019)) packages:

```
> mood.medtest(Trees ~ HOA, data = dat.tree)

        Mood's median test

data:  Trees by HOA
X-squared = 138.67, df = 1, p-value < 2.2e-16
```

A low p-value gives us evidence to reject H_0, we will use a bootstrap interval to give us a better picture of the magnitude of the difference, and which direction it is, using "simpleboot" (Peng (2019)) and "boot" (**?**) packages:

```
> boot.median<-function(data,indices){
+   d<-data[indices]# allows boot to select sample
+   return(median(d))
+ }
> library("simpleboot")
> boot<-two.boot(sample1=dat.tree$Trees[which(dat.tree$HOA=="Yes")],
+                sample2=dat.tree$Trees[which(dat.tree$HOA=="No")],
+                FUN=boot.median, R=1000)
> library("boot")
> (ci<-boot.ci(boot.out=boot,conf=0.95,type="perc"))

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
```

```
boot.ci(boot.out = boot, conf = 0.95, type = "perc")

Intervals :
Level     Percentile
95%   (-3, -2 )
Calculations and Intervals on Original Scale
```

Based on the interval, we have 95% confidence that the difference of the medians is in between (-3, -2), meaning that the median noumber of trees on a non HOA home is higher.

Gardening

We now examine whether garden status is effected by HOA membership:

```
> ################################
> ##### Gardening
> ################################
> dat.garden <- dat.HOA
> tab <- table(droplevels(dat.garden$HOA), droplevels(dat.garden$Garden))
> tab <- round(prop.table(tab,margin=1)*100,4)
> data6 <- data.frame(tab)

> xtable(data6, type = latex, file = "garden.tex")
```

HOA Status	Garden	Percentage(%)
Yes	Yes	7.74
No	Yes	10.05
Yes	No	92.26
No	No	89.95

Table 5: This table shows the proportion of the garden status of HOA vs. Non-HOA homes.

```
> p6 <- ggplot(data = data6, aes(x=data6$Var2, y = data6$Freq, fill = data6$Var1))+
+   geom_bar(position= "dodge", stat = "identity")+
+   xlab("Garden on Property")+
+   ylab("Percentage (%)")+
+   scale_fill_discrete(name = "HOA Status")+
+   ggtitle("Frequency of Garden Status", subtitle = "Based on HOA Membership")+
+   geom_hline(yintercept = 0)
```

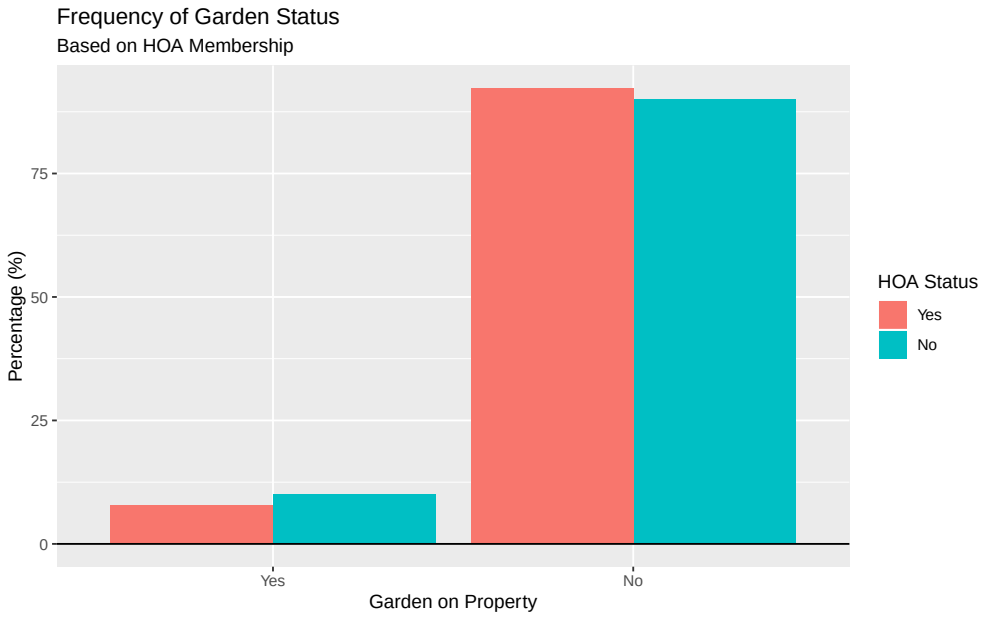

Figure 6: Graph showing whether homes have a garden based on HOA status.

We want to test to see if the proportion of homes without a garden is different across HOA status. The hypotheses we will be test are:
$H_0 : p_{HOA} - p_{NHOA} = 0$
$H_a : p_{HOA} - p_{NHOA} \neq 0$
We will make the assumption that each home is independent, checking for the Z distribution to hold with:

```
> HYes <- 82
> HNo <- 977
> HSum <- (HYes+HNo)
> NHYes <- 56
> NHNo <- 501
> NHSum <- (NHYes+NHNo)
```

We can see that for each sample of non-HOA and HOA homes $\hat{p}_i n_i \geq 5$ and $(1 - \hat{p}_i)n_i \geq 5$. We will test with:

```
> prop.test(x = c(HNo,NHNo), n = c(HSum,NHSum), conf.level = 0.95)

        2-sample test for equality of proportions with continuity correction

data:  c(HNo, NHNo) out of c(HSum, NHSum)
X-squared = 2.2082, df = 1, p-value = 0.1373
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.007974726  0.054188847
sample estimates:
   prop 1    prop 2
0.9225685 0.8994614
```

A non-significant p-value means that we fail to reject H_0, an indication that the proportions are the same.

Conclusion

We tested the various sustainability indicators to see if HOA status had any effect on them. For factors like

recycling and gardening were not affected, we found that HOA status had an effect on the lawn care habits and tree count of the homes. For recycling habits, we found that while homes had the same proportion of habits "Excellent"/"Good" that are damaging to the enviroment, there is a greater proportion of HOA homes with lawn care habits of "Excellent", indicating greater damage to the enviroment being done. For tree count, we found that non-HOA homes have a greater median tree count, saying with 95% confidence that there is a difference of 2-3 trees in favor of non-HOA.

Analyzing Sustainability Factors Relationships

We now look into the relationships of the sustainability factors with each other, within HOA and not.

Number of Trees Based on Garden Status

Within HOA:

First, we examine the tree count of a home owner based off their garden status:

```
> dat.tree2 <- dat.HOA
> dat.tree2 <- dat.tree2[which(!is.na(dat.tree2$Trees)),]
```

Note: This removes 14 missing values.

We now want to only look into the data for HOA homes:

```
> dat.tree2H <- dat.tree2[which(dat.tree2$HOA == "Yes"),]
```

Note: The sample size for this set is 1,056.

```
> summary(dat.tree2H[which(dat.tree2H$Garden == "Yes"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.500   2.988   4.000  10.000

> round(sd(dat.tree2H[which(dat.tree2H$Garden == "Yes"),]$Trees),4)

[1] 2.312

> summary(dat.tree2H[which(dat.tree2H$Garden == "No"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   2.944   4.000  35.000

> round(sd(dat.tree2H[which(dat.tree2H$Garden == "No"),]$Trees),4)

[1] 2.6971
```

Above we see the numerical summary of the number of trees on a property with a garden(top) vs. without a garden(bottom) for HOA homes.

```
> p7 <- ggplot(data = dat.tree2H, aes(x = dat.tree2H$Garden, y=dat.tree2H$Trees))+
+   geom_violin(fill = "lightblue")+
+   geom_boxplot(width = 0.25)+
+   xlab("Garden Status")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees",subtitle = "Based on Garden Status (For HOA Homes)")
```

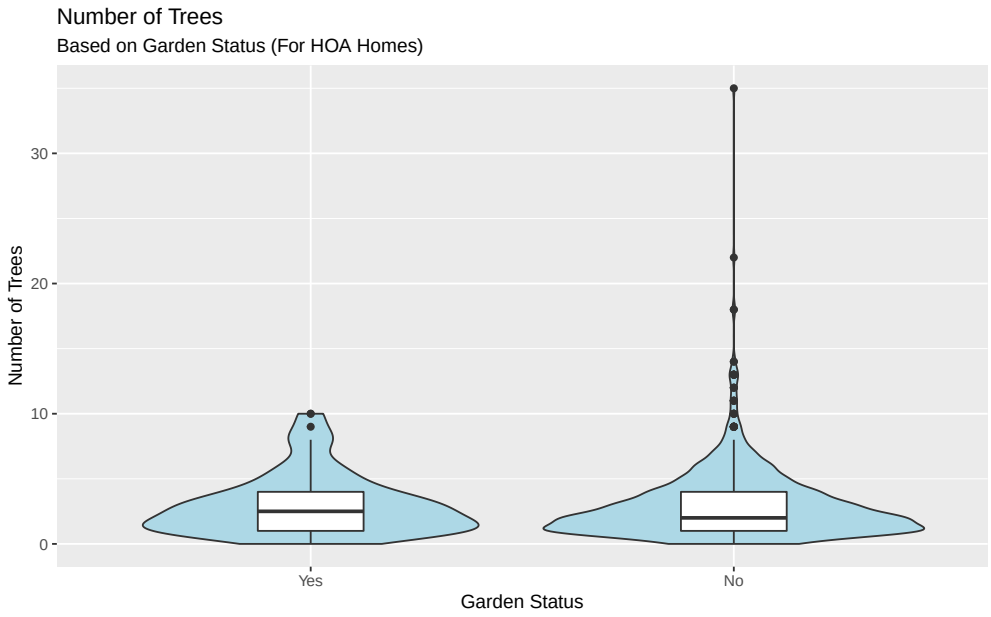

Figure 7: Graph showing how many trees a homeowner has based on garden status for HOA homes.

Based on the skewness of the data, we will test to see if there is a difference in the median of trees in a HOA home that has a garden versus one that doesn't. The hypotheses we will be testing are:
$H_0 : M_{Garden} - M_{NGarden} = 0$
$H_a : M_{Garden} - M_{NGarden} \neq 0$

```
> mood.medtest(Trees~Garden, data = dat.tree2H)

        Mood's median test

data:  Trees by Garden
X-squared = 0.50291, df = 1, p-value = 0.4782
```

A non significant p-value means that we fail to reject H_0, an indication that the medians are the same.

Within Non-HOA

We now want to grab the non-HOA data:

```
> dat.tree2N <- dat.tree2[which(dat.tree2$HOA == "No"),]
```

Note: The sample size for this set is 546.

```
> summary(dat.tree2N[which(dat.tree2N$Garden == "Yes"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   2.000   4.000   4.554   6.000  15.000

> round(sd(dat.tree2N[which(dat.tree2N$Garden == "Yes"),]$Trees),4)

[1] 2.8662

> summary(dat.tree2N[which(dat.tree2N$Garden == "No"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   5.000   6.049   9.000  40.000
```

```
> round(sd(dat.tree2N[which(dat.tree2N$Garden == "No"),]$Trees),4)

[1] 5.3539
```

Above we see the numerical summary of the number of trees on a property with a garden(top) vs. without a garden(bottom) for non-HOA homes.

```
> p13 <- ggplot(data = dat.tree2N, aes(x = dat.tree2N$Garden, y=dat.tree2N$Trees))+
+   geom_violin(fill = "lightblue")+
+   geom_boxplot(width = 0.25)+
+   xlab("Garden Status")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees",subtitle = "Based on Garden Status (For non-HOA Homes)")
```

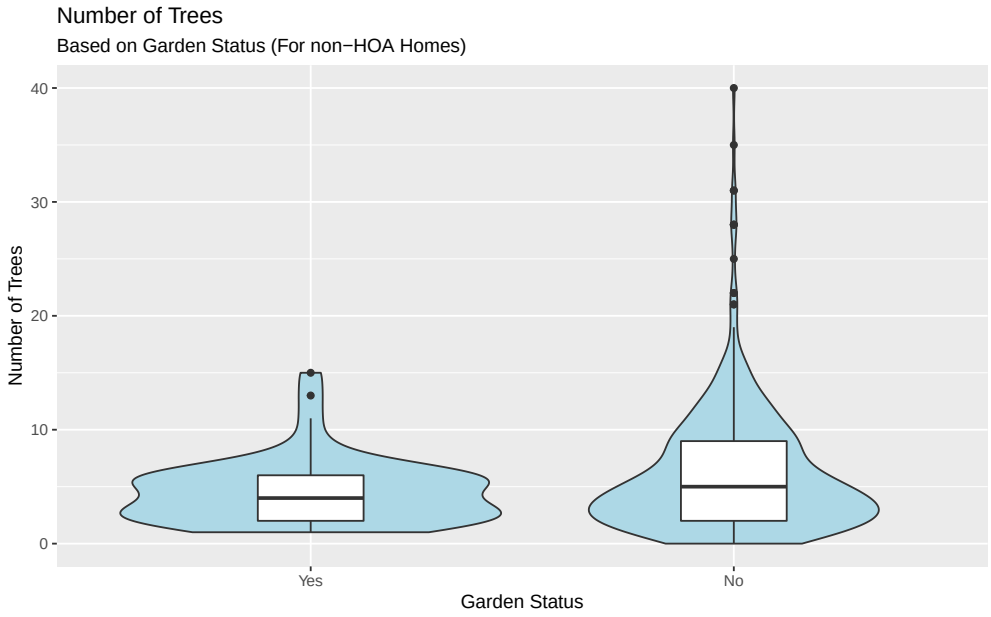

Figure 8: Graph showing how many trees a homeowner has based on garden status for non-HOA homes.

Based on the skewness of the data, we will test to see if there is a difference in the median of trees in a non HOA home that has a garden versus one that doesn't. The hypotheses we will be testing are:
$H_0 : M_{Garden} - M_{NGarden} = 0$
$H_a : M_{Garden} - M_{NGarden} \neq 0$

```
> mood.medtest(Trees~Garden, data = dat.tree2N)

        Mood's median test

data:  Trees by Garden
X-squared = 0.50352, df = 1, p-value = 0.478
```

A non significant p-value means that we fail to reject H_0, an indication that the medians are the same.

Conclusion

In looking at the relationship of tree count based on garden status within HOA and non-HOA homes, we find that the median number of trees are the same for homes with or without garden. This result hold true for both HOA and non-HOA homes.

Number of Trees based on Lawn Care Habits

Taking a look into the relationship of lawn care habits and the number of trees gives the following results:

Within HOA

```
> summary(dat.tree2H[which(dat.tree2H$LawnCare == "Excellent"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   2.739   4.000  18.000

> round(sd(dat.tree2H[which(dat.tree2H$LawnCare == "Excellent"),]$Trees),4)

[1] 2.3678

> summary(dat.tree2H[which(dat.tree2H$LawnCare == "Good"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.500   3.032   4.000  18.000

> round(sd(dat.tree2H[which(dat.tree2H$LawnCare == "Good"),]$Trees),4)

[1] 2.3845

> summary(dat.tree2H[which(dat.tree2H$LawnCare == "Poor"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   3.077   4.000  22.000

> round(sd(dat.tree2H[which(dat.tree2H$LawnCare == "Poor"),]$Trees),4)

[1] 2.8593

> summary(dat.tree2H[which(dat.tree2H$LawnCare == "None"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   3.081   4.000  35.000

> round(sd(dat.tree2H[which(dat.tree2H$LawnCare == "None"),]$Trees),4)

[1] 3.1887
```

Above we see the numerical summary of the number of trees for a HOA homes with lawn care habits described as ”Excellent”, ”Good”, ”Poor”, and ”None”.

```
> p8 <- ggplot(data = dat.tree2H, aes(x = dat.tree2H$LawnCare, y=dat.tree2H$Trees))+
+   geom_violin(fill = "lightblue",
+               trim = FALSE,
+               alpha = 0.5)+
+   geom_boxplot(width = 0.25)+
+   xlab("Lawn Care Habits")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees", subtitle="Based on Lawn Care Habits (For HOA Homes)")
```

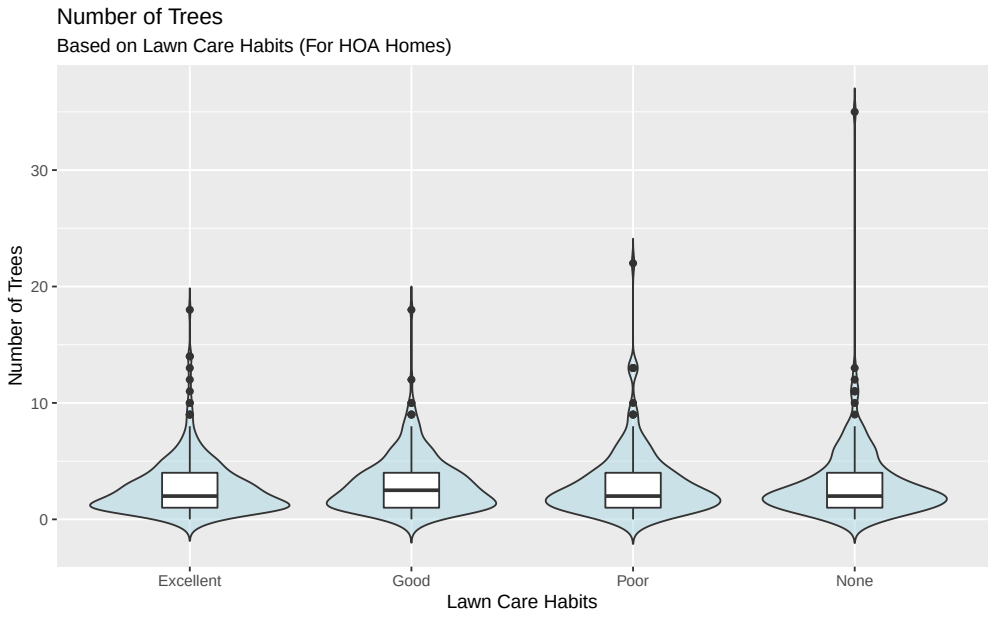

Figure 9: Graph showing the density of trees in HOA homes based on lawn care habits.

We want to see if there is a difference in medians across the different lawn care habits. The hypotheses we will be testing are:
$H_0 : M_1 - M_2 - M_3 - M_4 = 0$
H_a : The medians are different
We test this with:

```
> mood.medtest(Trees~LawnCare, data = dat.tree2H)

        Mood's median test

data:  Trees by LawnCare
X-squared = 2.7752, df = 3, p-value = 0.4276
```

A high p-value means we fail to reject H_0, an indication that the medians are the same.

Within Non-HOA

```
> summary(dat.tree2N[which(dat.tree2N$LawnCare == "Excellent"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.750   5.000   5.889   9.000  28.000

> round(sd(dat.tree2N[which(dat.tree2N$LawnCare == "Excellent"),]$Trees),4)

[1] 4.4874

> summary(dat.tree2N[which(dat.tree2N$LawnCare == "Good"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   5.000   6.467   8.000  40.000

> round(sd(dat.tree2N[which(dat.tree2N$LawnCare == "Good"),]$Trees),4)

[1] 6.3792
```

```
> summary(dat.tree2N[which(dat.tree2N$LawnCare == "Poor"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   4.000   5.017   7.000  25.000

> round(sd(dat.tree2N[which(dat.tree2N$LawnCare == "Poor"),]$Trees),4)

[1] 4.0868

> summary(dat.tree2N[which(dat.tree2N$LawnCare == "None"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   6.500   6.633  10.000  15.000

> round(sd(dat.tree2N[which(dat.tree2N$LawnCare == "None"),]$Trees),4)

[1] 4.4339
```

Above we see the numerical summary of the number of trees for non-HOA homes with lawn care habits described as "Excellent", "Good", "Poor", and "None".

```
> p14 <- ggplot(data = dat.tree2N, aes(x = dat.tree2N$LawnCare, y=dat.tree2N$Trees))+
+   geom_violin(fill = "lightblue",
+               trim = FALSE,
+               alpha = 0.5)+
+   geom_boxplot(width = 0.25)+
+   xlab("Lawn Care Habits")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees", subtitle="Based on Lawn Care Habits (For Non-HOA Homes)")
```

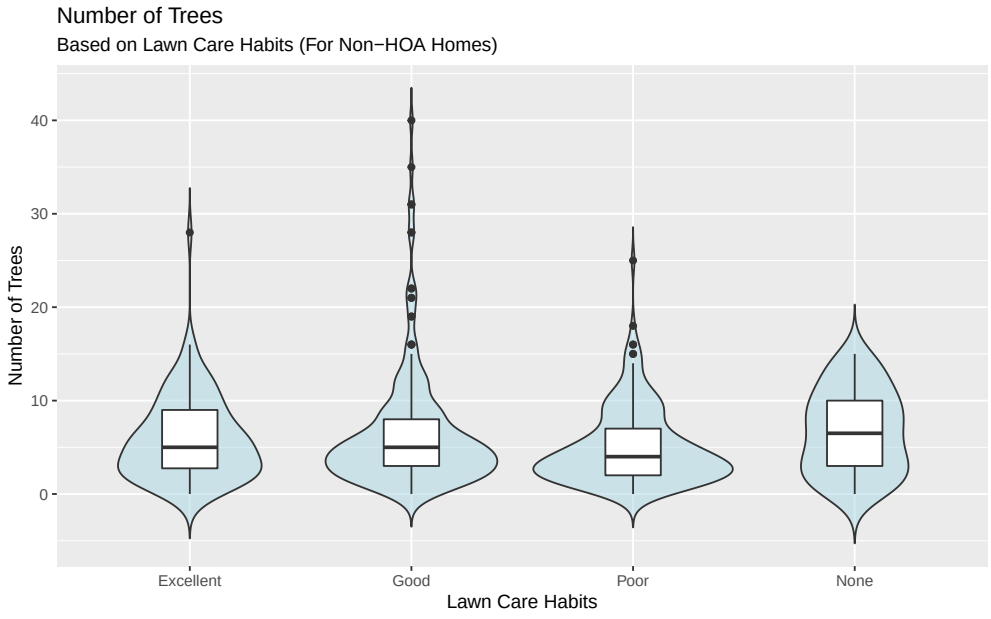

Figure 10: Graph showing the density of trees in non-HOA homes based on lawn care habits.

We want to see if there is a difference in medians across the different lawn care habits. The hypotheses we will be testing are:
$H_0 : M_1 - M_2 - M_3 - M_4 = 0$
H_a : The medians are different
We test this with:

```
> mood.medtest(Trees~LawnCare, data = dat.tree2N)

        Mood's median test

data:  Trees by LawnCare
X-squared = 10.897, df = 3, p-value = 0.01229
```

A p-value less than 0.05 means that we can reject H_0, we will run a pairwise median test to see which median is different:

```
> log<-capture.output({PT=pairwiseMedianTest(Trees~LawnCare,
+                                            data = dat.tree2N,
+                                            method="BH")})
> PT

            Comparison  p.value p.adjust
1 Excellent - Good = 0   0.3904  0.39040
2 Excellent - Poor = 0  0.02144  0.06432
3 Excellent - None = 0   0.2811  0.33730
4      Good - Poor = 0  0.07094  0.10640
5      Good - None = 0  0.05992  0.10640
6      Poor - None = 0 0.005176  0.03106
```

Based on the adjusted p value for "Poor"/"None" being less than 0.05, we can say that the median of trees on a non HOA home is different for homes with lawn care habits of "Poor" and "None".

Conclusion

In looking into whether lawn care status has an effect on the tree count of a home, within HOA homes we found that it didn't make a difference. For non-HOA homes, we found that homes with "Poor" and "None" lawn care habits had a different median number of trees.

Number of Trees based on Recycling Status

Finally, we take a look into whether a homeowners recycling status effects their tree count:

Within HOA

First, we want to remove NAs for both recycling status and tree count:

```
> dat.Recycle2 <- dat.HOA
> dat.Recycle2 <- dat.Recycle2[which(!is.na(dat.Recycle2$Recycle)),]
> dat.Recycle2$Recycle<-factor(dat.Recycle2$Recycle, labels = c("Both Trash/Recycle", "Only Trash"))
> dat.Recycle2 <- dat.Recycle2[which(!is.na(dat.Recycle2$Trees)),]
```

Note: This removes 288 missing values.

We now want to analyze only the HOA data:

```
> dat.Recycle2H <- dat.Recycle2[which(dat.Recycle2$HOA == "Yes"),]
```

Note: The sample size of this set is 844 homes.

```
> summary(dat.Recycle2H[which(dat.Recycle2H$Recycle == "Both Trash/Recycle"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   3.000   3.278   4.000  35.000

> round(sd(dat.Recycle2H[which(dat.Recycle2H$Recycle == "Both Trash/Recycle"),]$Trees),4)

[1] 2.9012

> summary(dat.Recycle2H[which(dat.Recycle2H$Recycle == "Only Trash"),]$Trees)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   1.000   2.000   2.764   4.000  13.000
```

```
> round(sd(dat.Recycle2H[which(dat.Recycle2H$Recycle == "Only Trash"),]$Trees),4)
```

```
[1] 2.2275
```

Above we see the numerical summary of the number of trees on a HOA home with both a trash can and recycling bin versus those with only a trash can.

```
> p11 <- ggplot(data = dat.Recycle2H, aes(x = dat.Recycle2H$Recycle, y=dat.Recycle2H$Trees))+
+   geom_violin(fill = "lightblue",
+               trim = FALSE,
+               alpha = 0.5)+
+   geom_boxplot(width = 0.25)+
+   xlab("Recycling Status")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees", subtitle = "Based on Recycling Status (For HOA Homes)")
```

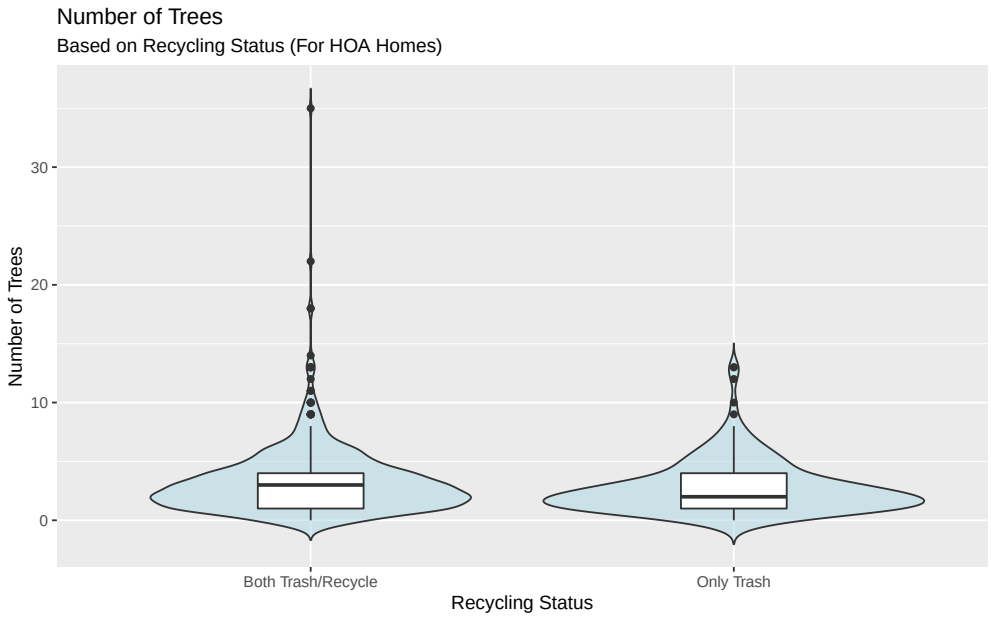

Figure 11: Graph showing how many trees a home has based on recycling status in HOA homes.

We want to test to see if the median of trees on a HOA home differs across recycling status. We are testing:
$H_0 : M_{Recycle} - M_{Trash} = 0$
$H_a : M_{Recycle} - M_{Trash} \neq 0$
We use:

```
> mood.medtest(Trees~Recycle, data = dat.Recycle2H)
```

```
        Mood's median test

data:  Trees by Recycle
X-squared = 6.6962, df = 1, p-value = 0.009662
```

We have a p value less than 0.05, therefore we have evidence to reject H_0. We will take a look at bootstrap confidence interval to determine which direction the difference is in:

```
> boot.median<-function(data,indices){
+   d<-data[indices]# allows boot to select sample
+   return(median(d))
+ }
> library("simpleboot")
> boot<-two.boot(sample1=dat.Recycle2H$Trees[which(dat.Recycle2H$Recycle=="Both Trash/Recycle")],
+                sample2=dat.Recycle2H$Trees[which(dat.Recycle2H$Recycle=="Only Trash")],
+                FUN=boot.median, R=1000)
> library("boot")
> (ci<-boot.ci(boot.out=boot,conf=0.95,type="perc"))

BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 1000 bootstrap replicates

CALL :
boot.ci(boot.out = boot, conf = 0.95, type = "perc")

Intervals :
Level     Percentile
95%   ( 0,  1 )
Calculations and Intervals on Original Scale
```

Based on this result, we can say with 95% confidence that the difference in the median is slightly greater in HOA homes that have both recycling and trash.

Within Non-HOA

We now want to access the only the non-HOA data:

```
> dat.Recycle2N <- dat.Recycle2[which(dat.Recycle2$HOA == "No"),]
```

Note: The sample size of this set is 487 homes.

```
> summary(dat.Recycle2N[which(dat.Recycle2N$Recycle == "Both Trash/Recycle"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   5.000   6.099   8.000  40.000

> round(sd(dat.Recycle2N[which(dat.Recycle2N$Recycle == "Both Trash/Recycle"),]$Trees),4)

[1] 5.4576

> summary(dat.Recycle2N[which(dat.Recycle2N$Recycle == "Only Trash"),]$Trees)

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   3.000   5.000   6.145   9.000  28.000

> round(sd(dat.Recycle2N[which(dat.Recycle2N$Recycle == "Only Trash"),]$Trees),4)

[1] 4.6847
```

Above we see the numerical summary of the number of trees on a non-HOA home with both a trash can and recycling bin versus those with only a trash can.

```
> p14 <- ggplot(data = dat.Recycle2N, aes(x = dat.Recycle2N$Recycle, y=dat.Recycle2N$Trees))+
+   geom_violin(fill = "lightblue",
+               trim = FALSE,
```

```
+                alpha = 0.5)+
+   geom_boxplot(width = 0.25)+
+   xlab("Recycling Status")+
+   ylab("Number of Trees")+
+   ggtitle("Number of Trees", subtitle = "Based on Recycling Status (For Non-HOA Homes)")
```

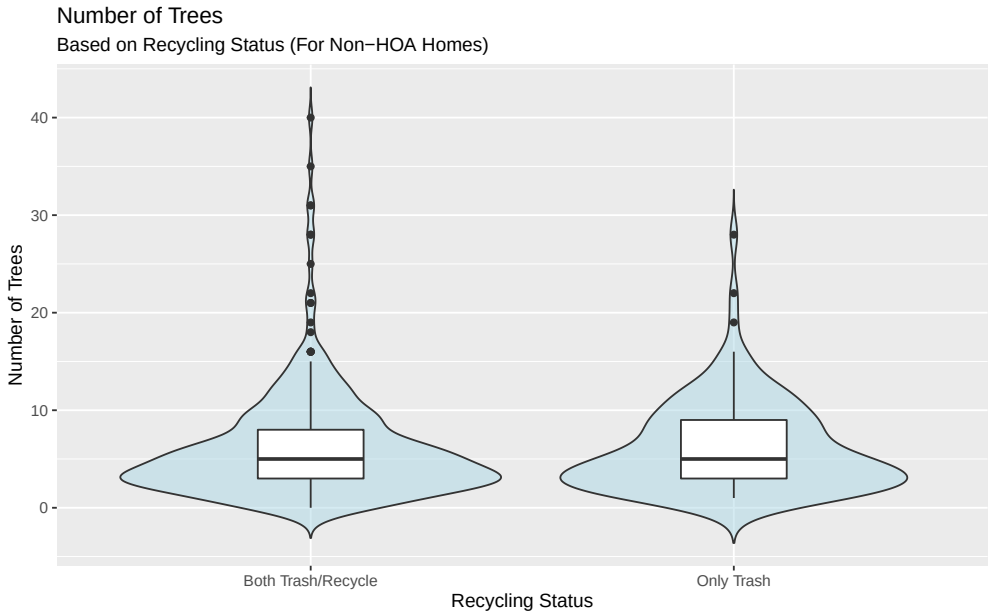

Figure 12: Graph showing how many trees a home has based on lawn care habits in non-HOA homes.

We want to test to see if the median of trees on a non HOA home differs across recycling status. We are testing:
$H_0 : M_{Recycle} - M_{Trash} = 0$
$H_a : M_{Recycle} - M_{Trash} \neq 0$
We use:

```
> mood.medtest(Trees~Recycle, data = dat.Recycle2N)

        Mood's median test

data:  Trees by Recycle
X-squared = 0.012847, df = 1, p-value = 0.9098
```

Due to a high p value, we fail to reject H_0, giving more evidence to say that the medians are the same.

Conclusion

In looking into whether recycling status has an effect on the tree count of a property, within non HOA home we found that there was none. However, with HOA homes we found that the median tree count is different across recycling status, saying with 95% that homes with recycling have a small edge versus HOA homes that don't.

Recycling Status versus Garden Status

We now examine the relationship of a home's garden status and recycling habits:

Within HOA

```
> dat.RecycleH <- dat.Recycle[which(dat.Recycle$HOA == "Yes"),]
```

Note: This set has a sample size of 844 homes.

```
> tabV <- table(droplevels(dat.RecycleH$Recycle), droplevels(dat.RecycleH$Garden))
> tab<- round(prop.table(tabV,margin=1)*100,4)
> data9 <- data.frame(tab)

> xtable(data9, type = latex, file = "garden.tex")
```

Recycling Status	Garden Status	Percentage(%)
Both Trash/Recycle	Yes	7.82
Only Trash	Yes	3.04
Both Trash/Recycle	No	92.18
Only Trash	No	96.96

Table 6: This table shows the frequency of garden status based on a HOA home's recycling status.

```
> p9 <- ggplot(data = data9, aes(x = data9$Var1, y = data9$Freq, fill = data9$Var2))+
+   geom_bar(position = "dodge", stat = "identity")+
+   xlab("Recycling Status")+
+   ylab("Percentage (%)")+
+   scale_fill_discrete(name = "Garden Status")+
+   ggtitle("Frequency of Garden Status", subtitle ="Based on Recycling Status (For HOA Homes)")+
+   geom_hline(yintercept = 0)
```

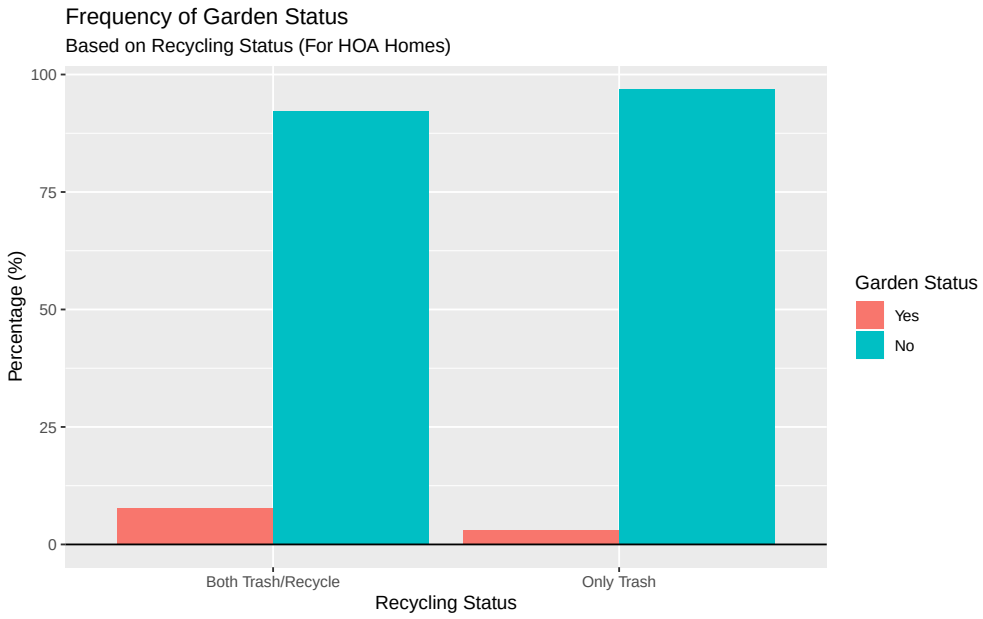

Figure 13: Graph showing the frequency of garden status based off of a HOA home's recycling status.

We want to see if the proportion of HOA homes that don't have a garden is different across recycling status. We will test the following hypotheses:
H_0 : The garden status is not related to recycling status.
H_a : The garden status is related to recycling status.
We will make the assumption that each home is independent:

```
> RecycleYes <- 48
> RecycleNo <- 566
```

```
> RecycleSum <- (RecycleYes+RecycleNo)
> TrashYes <- 7
> TrashNo <- 223
> TrashSum <- (TrashYes+TrashNo)
```

We will test the chi-squared assumptions with:

```
> 2*2 <(RecycleYes + RecycleNo + TrashYes + TrashNo)

[1] TRUE
```

We will run a Pearson's Chi-squared test to look for any corelation between the two variables:

```
> R1 <- c(RecycleYes,RecycleNo)
> R2 <- c(TrashYes,TrashNo)
> tab <- matrix(data = c(R1,R2),
+               nrow = 2,
+               ncol = 2,
+               byrow = TRUE)
> chisq.test(tab)

        Pearson's Chi-squared test with Yates' continuity correction

data:  tab
X-squared = 5.501, df = 1, p-value = 0.01901
```

A p-value less than 0.05 means that we can reject H_0, an indication that there is a relationship between recycling status and garden status, we will run Cramer's V to determine the strength of the relationship:

```
> library(rcompanion)
> cramerV(tab,bias.correct = TRUE)

Cramer V
 0.07898

> cramerV(tab)

Cramer V
 0.08612
```

From these results, we can see that there is a weak to moderate association between recycling habits and garden status in HOA homes.

Within Non-HOA

First, we want to see the data for non-HOA homes:

```
> dat.RecycleN <- dat.Recycle[which(dat.Recycle$HOA == "No"),]
```

Note: The sample size of this set is 498 homes

```
> tab <- table(droplevels(dat.RecycleN$Recycle), droplevels(dat.RecycleN$Garden))
> tab<- round(prop.table(tab,margin=1)*100,4)
> data15 <- data.frame(tab)

> xtable(data15, type = latex, file = "garden.tex")
```

Recycling Status	Garden Status	Percentage(%)
Both Trash/Recycle	Yes	10.99
Only Trash	Yes	6.40
Both Trash/Recycle	No	89.01
Only Trash	No	93.60

Table 7: This table shows the frequency of garden status based on a non-HOA home's recycling status.

```
> p17 <- ggplot(data = data15, aes(x = data15$Var1, y = data15$Freq, fill = data15$Var2))+
+   geom_bar(position = "dodge", stat = "identity")+
+   xlab("Recycling Status")+
+   ylab("Percentage (%)")+
+   scale_fill_discrete(name = "Garden Status")+
+   ggtitle("Frequency of Garden Status", subtitle ="Based on Recycling Status (For a non-HOA Home)")+
+   geom_hline(yintercept = 0)
```

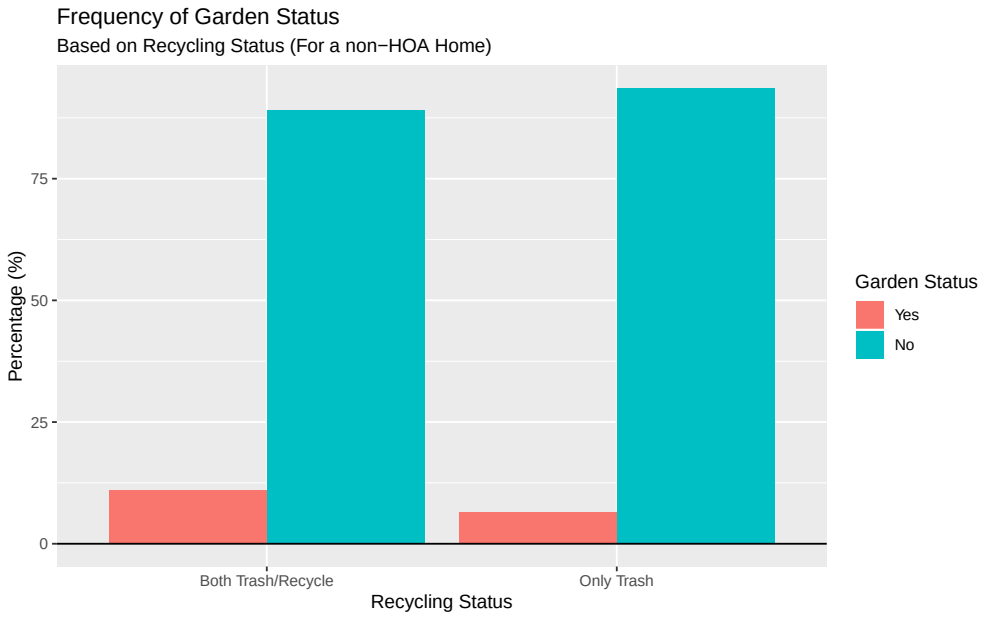

Figure 14: Graph showing the frequency of garden status based off of a non-HOA home's recycling status.

We want to see if there is a relationship in HOA homes between the garden status and recycling status. We will test the following hypotheses:
H_0 : The garden status is not related to lawn care habits.
H_a : The garden status is related to lawn care habits.
We will make the assumption that each home is independent:

```
> RecycleYes <- 41
> RecycleNo <- 332
> RecycleSum <- (RecycleYes+RecycleNo)
> TrashYes <- 8
> TrashNo <- 117
> TrashSum <- (TrashYes+TrashNo)
```

We will test the chi-squared assumptions with:

```
> 2*2 <(RecycleYes + RecycleNo + TrashYes + TrashNo)

[1] TRUE
```

We will run a Pearson's Chi-squared test to look for any corelation between the two variables:

```
> R1 <- c(RecycleYes,RecycleNo)
> R2 <- c(TrashYes,TrashNo)
> tab <- matrix(data = c(R1,R2),
```

```
+               nrow = 2,
+               ncol = 2,
+               byrow = TRUE)
> chisq.test(tab)

        Pearson's Chi-squared test with Yates' continuity correction

data:  tab
X-squared = 1.7378, df = 1, p-value = 0.1874
```

A large p value (0.1874) means that we fail to reject H_0 an indication that there is no relationship between the two variables.

Conclusion

In looking into whether there is a relationship between garden status and recycling status we found that for non HOA homes there was no relationship between the two variables. However, for HOA homes we found that there was a weak to moderate association between garden status and recycling status.

Lawn Care Habits versus Garden Status

Within HOA

First, we want the data for HOA homes (There were no values listed as NA):

```
> dat.HOAH <- dat.HOA[which(dat.HOA$HOA == "Yes"),]
```

Note: This set has a sample size of 1059

```
> tab <- table(droplevels(dat.HOAH$LawnCare),droplevels(dat.HOAH$Garden))
> tab <- round(prop.table(tab,margin=1)*100,4)
> data10 <- data.frame(tab)

> xtable(data10, type="latex", file = "LawnVgarden.tex")
```

Lawn Care Habit	Garden Status	Percentage(%)
Excellent	Yes	6.90
Good	Yes	8.40
Poor	Yes	7.22
None	Yes	8.82
Excellent	No	93.10
Good	No	91.60
Poor	No	92.78
None	No	91.18

Table 8: This table shows the frequency of garden status based on a HOA home's lawn care habits.

```
> p10 <- ggplot(data = data10, aes(x=data10$Var1, y=data10$Freq, fill = data10$Var2))+
+   geom_bar(position="dodge", stat = "identity")+
+   xlab("Lawn Care Habits")+
+   ylab("Frequency (%)")+
+   ggtitle("Frequency of Garden Status", subtitle = "Based on Lawn Care Habits (For HOA Homes)")+
+   scale_fill_discrete(name = "Garden Status")+
+   geom_hline(yintercept = 0)
```

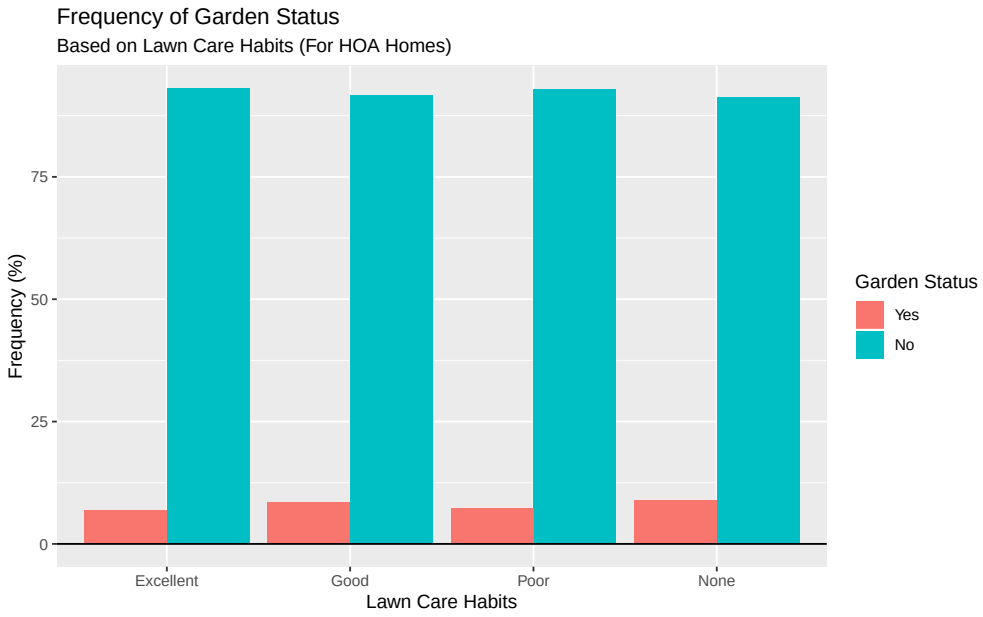

Figure 15: Graph showing the frequency of garden status based off of a HOA home's lawn care habits.

We want to see if garden status effects thelawn care habits within HOA homes. The hypotheses we will be testing are:
H_0 : The garden status is not related to lawn care habits.
H_a : The garden status is related to lawn care habits.
We will make the assumption that each home is independent:

```
> GardenExcellent <- 26
> NGardenExcellent <- 351
> ExcellentSum <- (GardenExcellent+NGardenExcellent)
> GardenGood <- 21
> NGardenGood <- 229
> GoodSum <- (GardenGood+NGardenGood)
> GardenPoor <- 14
> NGardenPoor <- 180
> PoorSum <- (GardenPoor+NGardenPoor)
> GardenNone <- 21
> NGardenNone <- 217
> NoneSum <- (GardenNone+NGardenNone)
```

We will test the chi-squared assumptions with:

```
> 4*2 <(ExcellentSum + GoodSum + PoorSum + NoneSum)

[1] TRUE
```

We will run a Pearson's Chi-squared test to look for any corelation between the two variables:

```
> R1 <- c(GardenExcellent,GardenGood,GardenPoor,GardenNone)
> R2 <- c(NGardenExcellent,NGardenGood,NGardenPoor,NGardenNone)
> tab <- matrix(data = c(R1,R2),
+               nrow = 2,
+               ncol = 4,
+               byrow = TRUE)
> chisq.test(tab)
```

```
        Pearson's Chi-squared test

data:  tab
X-squared = 0.99345, df = 3, p-value = 0.8028
```

We have a p-value that is greater than 0.05, this means that we fail to reject H_0, an indication that there is no relation between garden status and lawn care habits.

Within Non-HOA

We only want to see the non-HOA homes before we start (there were no values listed as NA):

```
> dat.HOAN <- dat.HOA[which(dat.HOA$HOA == "No"),]
```

Note: This set has a sample size of 557 homes.

```
> tab <- table(droplevels(dat.HOAN$LawnCare),droplevels(dat.HOAN$Garden))
> tab <- round(prop.table(tab,margin=1)*100,4)
> data13 <- data.frame(tab)

> xtable(data13, type="latex", file = "LawnVgarden.tex")
```

Lawn Care Habit	Garden Status	Percentage (%)
Excellent	Yes	7.27
Good	Yes	11.50
Poor	Yes	12.71
None	Yes	3.03
Excellent	No	92.73
Good	No	88.50
Poor	No	87.29
None	No	96.97

Table 9: This table shows the frequency of garden status based on a non-HOA home's lawn care habits.

```
> p15 <- ggplot(data = data13, aes(x=data13$Var1, y=data13$Freq, fill = data13$Var2))+
+   geom_bar(position="dodge", stat = "identity")+
+   xlab("Lawn Care Habits")+
+   ylab("Frequency (%)")+
+   ggtitle("Frequency of Garden Status", subtitle = "Based on Lawn Care Habits (For Non-HOA Homes)")+
+   scale_fill_discrete(name = "Garden Status")+
+   geom_hline(yintercept = 0)
```

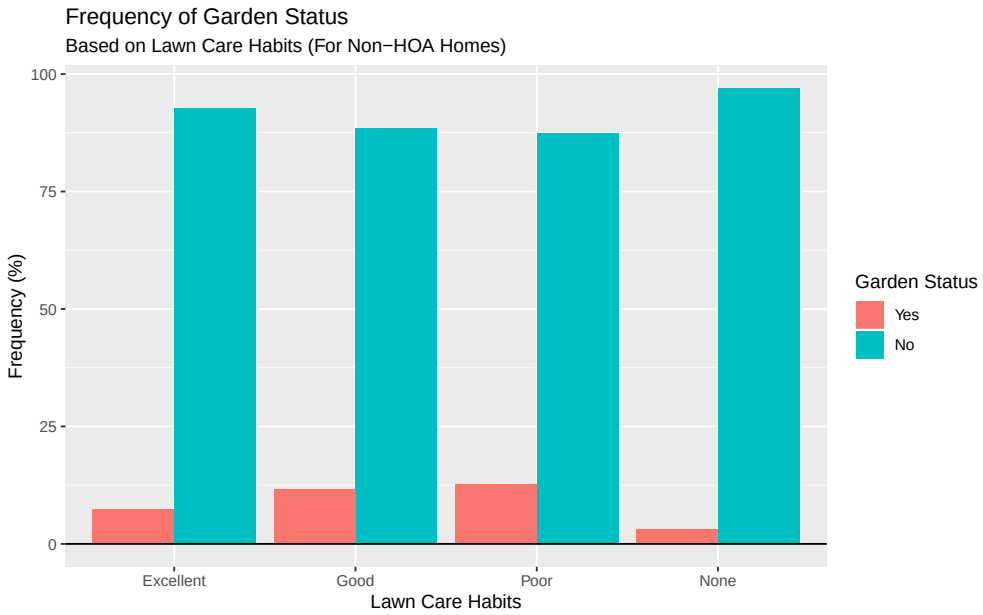

Figure 16: Graph showing the frequency of garden status based off of a non-HOA home's lawn care habits.

We want to see if garden status effects the lawn care habits within HOA homes. The hypotheses we will be testing are:
H_0 : The garden status is not related to lawn care habits.
H_a : The garden status is related to lawn care habits.
We will make the assumption that each home is independent:

```
> GardenExcellent <- 8
> NGardenExcellent <- 102
> ExcellentSum <- (GardenExcellent+NGardenExcellent)
> GardenGood <- 23
> NGardenGood <- 177
> GoodSum <- (GardenGood+NGardenGood)
> GardenPoor <- 23
> NGardenPoor <- 158
> PoorSum <- (GardenPoor+NGardenPoor)
> GardenNone <- 2
> NGardenNone <- 64
> NoneSum <- (GardenNone+NGardenNone)
```

We will check for chi-squared assumptions with:

```
> 4*2 < (ExcellentSum+GoodSum+PoorSum+NoneSum)

[1] TRUE
```

Now running a chi-squared test with:

```
> R1 <- c(GardenExcellent,GardenGood,GardenPoor,GardenNone)
> R2 <- c(NGardenExcellent, NGardenGood,NGardenPoor,NGardenNone)
> tab <- matrix(data = c(R1,R2),
+               nrow = 2,
+               ncol = 4,
+               byrow = TRUE)
> chisq.test(tab)
```

```
        Pearson's Chi-squared test

data:  tab
X-squared = 6.4128, df = 3, p-value = 0.09317
```

We have a p-value that is greater than 0.05, this means that we fail to reject H_0, an indication that there is no relation between garden status and lawn care habits.

Conclusion

In looking into whether lawn care habits effected the garden status of a home, we showed that the habits had no effect on the garden status of home, both within HOA and non HOA homes.

Recycling Status Versus Lawn Care Habits

Within HOA

We can reuse the dat.RecycleH data frame created above:

```
> tab <- table(droplevels(dat.RecycleH$Recycle), droplevels(dat.RecycleH$LawnCare))
> tab<- round(prop.table(tab,margin=1)*100,4)
> data12 <- data.frame(tab)

> xtable(data12, type = "latex", file="LawnVRecycle.tex")
```

Recycling Status	Lawn Care Habit	Percentage (%)
Both Trash/Recycle	Excellent	38.76
Only Trash	Excellent	24.35
Both Trash/Recycle	Good	24.27
Only Trash	Good	20.87
Both Trash/Recycle	Poor	17.26
Only Trash	Poor	23.91
Both Trash/Recycle	None	19.71
Only Trash	None	30.87

Table 10: This table shows the proportion of lawn care habits based on a HOA home's recycling status.

```
> p12 <- ggplot(data = data12, aes(x=data12$Var1, y=data12$Freq, fill = data12$Var2))+
+   geom_bar(position="dodge", stat = "identity")+
+   xlab("Recycling Status")+
+   ylab("Frequency (%)")+
+   ggtitle("Frequency of Lawn Care Habits", subtitle = "Based on Recycling Status (For HOA Homes)")+
+   scale_fill_discrete(name = "Lawn Care Habits")+
+   geom_hline(yintercept = 0)
```

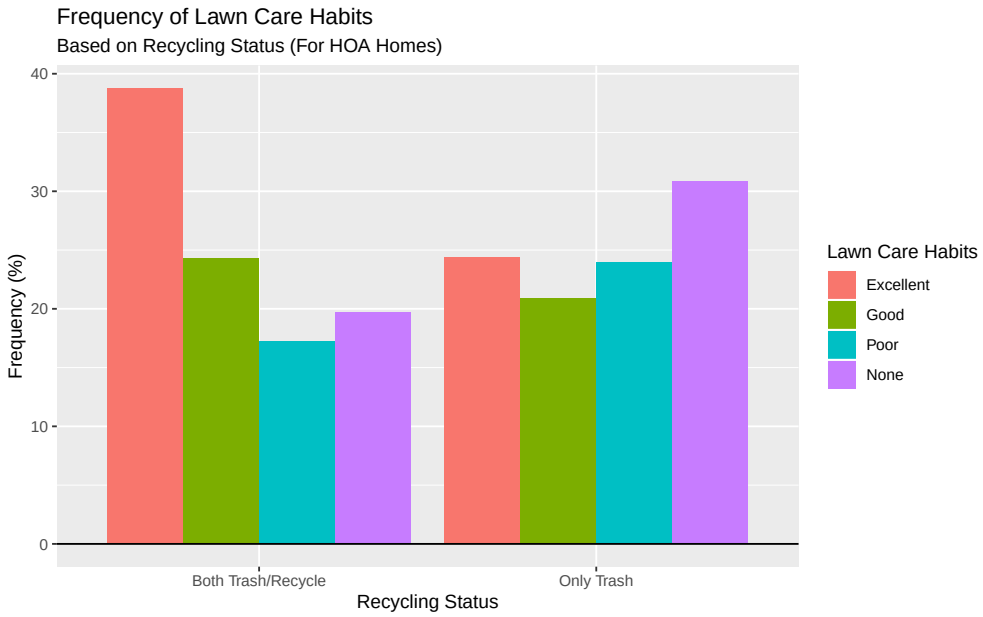

Figure 17: Graph showing the frequency of lawn care habits based off of a HOA home's recycling status.

We are testing to see if recycling status has an effect on lawn care habits within HOA homes. The hypotheses we are testing are:
H_0 : The recycling status is not related to lawn care habits.
H_a : The recycling status is related to lawn care habits.
We will make the assumption that each home is independent:

```
> RecycleExcellent <- 238
> TrashExcellent <- 56
> ExcellentSum <- RecycleExcellent + TrashExcellent
> RecycleGood <- 149
> TrashGood <- 48
> GoodSum <- RecycleGood + TrashGood
> RecyclePoor <- 106
> TrashPoor <- 55
> PoorSum <- RecyclePoor + TrashPoor
> RecycleNone <- 121
> TrashNone <- 71
> NoneSum <- RecycleNone + TrashNone
```

We will test for chi-squared assumptions with:

```
> 4*2 < (ExcellentSum + GoodSum + PoorSum + NoneSum)

[1] TRUE
```

We will test with:

```
> R1 <- c(RecycleExcellent,RecycleGood,RecyclePoor,GardenNone)
> R2 <- c(TrashExcellent,TrashGood,TrashPoor,TrashNone)
> tab <- matrix(data = c(R1,R2),
+               nrow = 2,
+               ncol = 4,
+               byrow = TRUE)
> chisq.test(tab)
```

```
        Pearson's Chi-squared test

data:  tab
X-squared = 171.93, df = 3, p-value < 2.2e-16
```

A p-value less than 0.05 means that we can reject H_0 in favor of H_a, meaning that there maybe a relation between recycling status and lawn care habits. We will run cramer's V to see the strength of the relationship:

```
> cramerV(tab,bias.correct = TRUE)

Cramer V
   0.483

> cramerV(tab)

Cramer V
   0.487
```

From the results of the cramer V tests, we can say that there is a weak association between recycling status and lawn care habits.

Within Non-HOA

```
> tab <- table(droplevels(dat.RecycleN$Recycle), droplevels(dat.RecycleN$LawnCare))
> tab<- round(prop.table(tab,margin=1)*100,4)
> data14 <- data.frame(tab)

> xtable(data14, type = "latex", file="LawnVRecycle.tex")
```

Recycling Status	Lawn Care Habit	Percentage (%)
Both Trash/Recycle	Excellent	21.98
Only Trash	Excellent	12.80
Both Trash/Recycle	Good	34.58
Only Trash	Good	41.60
Both Trash/Recycle	Poor	31.90
Only Trash	Poor	35.20
Both Trash/Recycle	None	11.53
Only Trash	None	10.40

Table 11: This table shows the proportion of lawn care habits based on a non-HOA home's recycling status.

```
> p16 <- ggplot(data = data14, aes(x=data14$Var1, y=data14$Freq, fill = data14$Var2))+
+   geom_bar(position="dodge", stat = "identity")+
+   xlab("Recycling Status")+
+   ylab("Frequency (%)")+
+   ggtitle("Frequency of Lawn Care Habits", subtitle = "Based on Recycling Status (For Non-HOA Homes)")
+   scale_fill_discrete(name = "Lawn Care Habits")+
+   geom_hline(yintercept = 0)
```

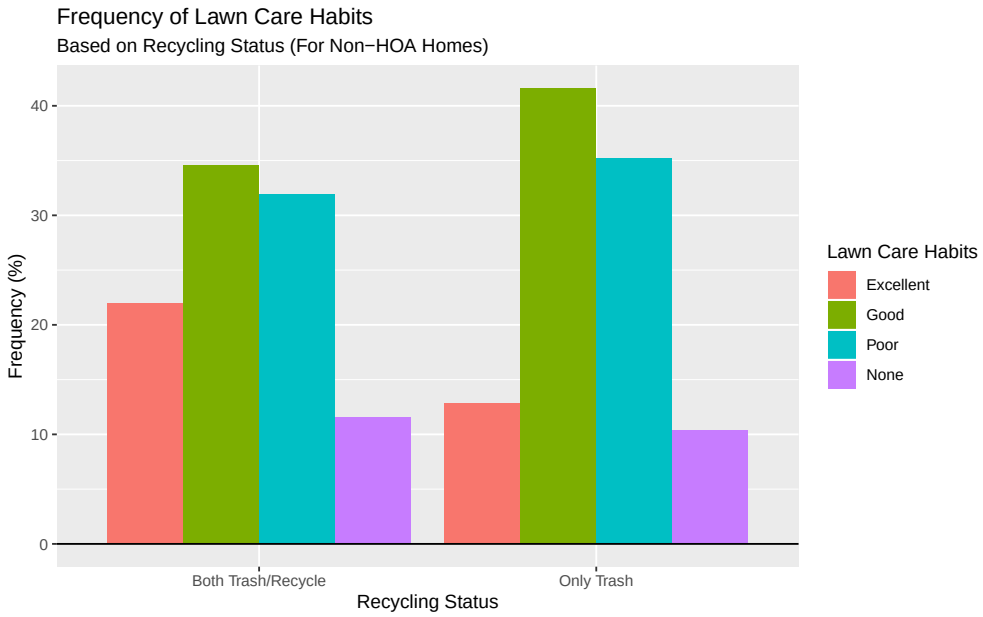

Figure 18: Graph showing the frequency of lawn care habits based off of a non-HOA home's recycling status.

We are testing to see if recycling status has an effect on lawn care habits within non HOA homes. The hypotheses we are testing are:
H_0 : The recycling status is not related to lawn care habits.
H_a : The recycling status is related to lawn care habits.
We will make the assumption that each home is independent:

```
> RecycleExcellent <- 82
> TrashExcellent <- 16
> ExcellentSum <- RecycleExcellent + TrashExcellent
> RecycleGood <- 129
> TrashGood <- 52
> GoodSum <- RecycleGood + TrashGood
> RecyclePoor <- 119
> TrashPoor <- 44
> PoorSum <- RecyclePoor + TrashPoor
> RecycleNone <- 43
> TrashNone <- 13
> NoneSum <- RecycleNone + TrashNone
```

We will test for the assumptions for the chi-squared tests:

```
> (4*2) < (ExcellentSum+GoodSum+PoorSum+NoneSum)

[1] TRUE
```

We will test with:

```
> R1 <- c(RecycleExcellent,RecycleGood,RecyclePoor,RecycleNone)
> R2 <- c(TrashExcellent, TrashGood,TrashPoor,TrashNone)
> tab <- matrix(data = c(R1,R2),
+               nrow = 2,
+               ncol = 4,
+               byrow = TRUE)
> chisq.test(tab)
```

```
        Pearson's Chi-squared test

data:  tab
X-squared = 5.6975, df = 3, p-value = 0.1273
```

We get no warnings, so we see that a p-value greater than 0.05 means that we fail to reject H_0 an indication that there is no correlation.

Conclusion

In looking into whether there is a relationship between lawn care habits and the recycling status of a home, for non HOA homes we found there was no relationship. However, for HOA homes we found that there was a weak association between recycling status and lawn care habits.

Overall Conclusion

In looking into the sustainablity factors relationships to each other within HOA and non-HOA homes, for non-HOA we only found that lawn care habits of "Poor"/"None" had different median trees on the property. However, for HOA homes we found numerous relationships. First, we found that HOA homes that recycled had a slightly larger median tree count than those that didn;t recycle. Additionally, we found a weak to moderate relationship between recyling status and garden status and a weak relationship between recycling and lawn care habits. We find that the sustainability factors tend to have more relationships between each other in HOA homes compared to non-HOA homes.

References

(2019). Qualitative assessment of hoa covenants.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., and Swinton, J. (2019). *xtable: Export Tables to LaTeX or HTML*. R package version 1.8-4.

Google (2019). Google earth images.

Mangiafico, S. (2019). *rcompanion: Functions to Support Extension Education Program Evaluation*. R package version 2.3.7.

Peng, R. D. (2019). *simpleboot: Simple Bootstrap Routines*. R package version 1.1-7.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.