## Introduction

The dataset that I worked with was compiled from multiple studies to create one large dataset of variables including plant species, flower symmetry type, pollinator visitation while observed, the amount of "sharing" done by pollinators, or how many of the visited pollinators visited other plants, and what proportion of sharing was between different symmetry types.The data was compiled for a study that was primarily interested in the affect of symmetry on the number of potential pollinator species that commonly visited. The code below shows that the large data set was compiled from 67 separate studies. Each study goal was somewhat different ranging from modeling on pollinator activity, population ecology, floral structures, and other topics. I chose this large compiled dataset because I wanted to evaluate weather certain flower types and genuses experienced much higher pollination rates across ecosystems than others, and how floral symmetry impacts pollinator preferences. My driving question for this project was to determine if certain genuses were far more likely to be visited by pollinators than other genuses. This would be interesting to explore as it could be used to further narrow down focus to study groups that have unusually high or low visitation by pollinators. The groups that experience very high or very low pollination rates could be subject to current directional selection, or have interesting relationships with their primary pollinators.

```
refrenceData <- read.csv("C:/Users/lizzy/OneDrive/Desktop/refrences.csv")
length(unique(refrenceData$Reference))
```

```
## [1] 67
```

## Summary of data to be analyzed

The original dataframe is 347 KB, and has 4079 rows of data. Each row is a separate plant being observed, not necessarily a new species. Looking at the data the largest challenge is that my analysis takes place on the genus level, but there is no separate column for genus, instead the scientific name with species and genus is grouped together in one column. This can easily be fixed by using a splitting functions that many packages have. I chose to use the splitting function in stringr to separate the genus and species name and get just the genus in the correct order as a vector of strings.

## Methods

My original intent was to run a two-way ANOVA on the available data, evaluating if there is a significant connection between floral symmetry, plant genus, and pollinator visitation rate. Then I would complete post-hoc testing to determine what groups were significantly different. I had to amend this plan, as running post-hoc tests on a dataset with so many groups, hundreds of genuses, would not be practical and would take many hours for the test to run if the test could even be completed on my computer. To amend this, I instead decided to create a random sample of the data.

This random sample would then serve as the new data set for ANOVA analysis of genus, symmetry, and pollinator visitation. The final method used was random selection of 20 genus groups for analysis. Then, one way ANOVA was completed to determine if genus had an impact on pollinator visitation, and a two-way ANOVA was done to determine if there was any association between genus and symmetry and their affect on pollinator visitation.

This code is the manipulation of data I did to set up for the final testing

```
plantpolinatordata <- read.csv("C:/Users/lizzy/OneDrive/Desktop/plantdata.csv")

plantPolSpecies <- plantpolinatordata$plant
```

```r
#helps to separate strings without me writing my own function
library(stringr)
```

```
## Warning: package 'stringr' was built under R version 4.0.5
```

```r
numberSamples <- length(plantPolSpecies)
plantPolGenus <- rep(" ", 2736) # pre allocation of empty strings


plantPolGenus <- word(plantPolSpecies, 1) #this grabs just the genus from the vector of full scientific

plantpolinatordata$Genus <- plantPolGenus

plantPolGenusAndVis <- plantpolinatordata[ , c( "Genus", "n.poll", "symmetry") ] #uses genus list to ma

#then, because the data set it too large to run TUKEY test on all genus groups, I selected 20 random ge
randomSample <- sample(plantPolGenus, 20)
randomSample
```

```
##  [1] "Gentiana"      "Linum"         "Avicennia"     "Stellaria"
##  [5] "Discaria"      "Kalimeris"     "Northea"       "Marcetia"
##  [9] "Chamaecrista"  "Echium"        "Agauria"       "Aristotelia"
## [13] "Eupatorium"    "Gymnocalycium" "Oxalis"        "Chamerion"
## [17] "Lespedeza"     "Trifolium"     "Passiflora"    "Senecio"
```

```r
#subsetting the data by the selected genus
plantPolGenus20 <- subset(plantPolGenusAndVis,  Genus == "Plantago" | Genus == "Psidium" | Genus == "Pa
```

This code is the ANOVA testing, both one way for evaluation of genus affect, and two way to see if there is an interaction between symmetry and genus

```r
# Does plant genus affect the visitation rate of pollinators to flowers?

# Null : The mean pollinator visitation between plant genuses are not significantly different.
# Alternate: The mean pollinator visitation between at least one of the genuses is differnt.

PolGenaov <- aov(n.poll ~ Genus, data = plantPolGenus20)
summary(PolGenaov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Genus        14   1078    77.0   2.862 0.000683 ***
## Residuals   170   4573    26.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#p value of 0.000232

TukeyHSD(PolGenaov)
```

```
##   Tukey multiple comparisons of means
```

```
##     95% family-wise confidence level
##
## Fit: aov(formula = n.poll ~ Genus, data = plantPolGenus20)
##
## $Genus
##                                   diff        lwr       upr     p adj
## Agrimonia-Aechmea          2.086957e+00  -7.584189 11.758102 0.9999883
## Aphloia-Aechmea            3.166957e+00  -1.990988  8.324901 0.7241809
## Argyranthemum-Aechmea      4.086957e+00  -5.584189 13.758102 0.9806866
## Celmisia-Aechmea           4.170290e+00  -2.186962 10.527542 0.6227301
## Cerastium-Aechmea          3.718535e+00  -1.815907  9.252978 0.5826645
## Colea-Aechmea             -1.913043e+00 -11.584189  7.758102 0.9999961
## Eugenia-Aechmea            8.695652e-02 -10.871568 11.045481 1.0000000
## Fraxinus-Aechmea          -1.913043e+00 -15.073806 11.247719 0.9999999
## Homalanthus-Aechmea        1.753623e+00  -6.430071  9.937318 0.9999892
## Mimusops-Aechmea          -7.365729e-01  -6.446513  4.973368 1.0000000
## Paragenipa-Aechmea         8.695652e-02  -4.860743  5.034656 1.0000000
## Plantago-Aechmea           5.869565e-01  -7.596738  8.770651 1.0000000
## Psidium-Aechmea           -8.463768e-01  -6.771157  5.078403 0.9999999
## Vaccinium-Aechmea          6.953623e+00   1.028843 12.878403 0.0068634
## Aphloia-Agrimonia          1.080000e+00  -8.533664 10.693664 1.0000000
## Argyranthemum-Agrimonia    2.000000e+00 -10.623360 14.623360 0.9999998
## Celmisia-Agrimonia         2.083333e+00  -8.223597 12.390264 0.9999948
## Cerastium-Agrimonia        1.631579e+00  -8.189226 11.452384 0.9999996
## Colea-Agrimonia           -4.000000e+00 -16.623360  8.623360 0.9989111
## Eugenia-Agrimonia         -2.000000e+00 -15.634787 11.634787 0.9999999
## Fraxinus-Agrimonia        -4.000000e+00 -19.460395 11.460395 0.9998926
## Homalanthus-Agrimonia     -3.333333e-01 -11.856832 11.190165 1.0000000
## Mimusops-Agrimonia        -2.823529e+00 -12.744294  7.097235 0.9996730
## Paragenipa-Agrimonia      -2.000000e+00 -11.502520  7.502520 0.9999914
## Plantago-Agrimonia        -1.500000e+00 -13.023498 10.023498 1.0000000
## Psidium-Agrimonia         -2.933333e+00 -12.979286  7.112620 0.9995620
## Vaccinium-Agrimonia        4.866667e+00  -5.179286 14.912620 0.9398866
## Argyranthemum-Aphloia      9.200000e-01  -8.693664 10.533664 1.0000000
## Celmisia-Aphloia           1.003333e+00  -5.266128  7.272794 0.9999997
## Cerastium-Aphloia          5.515789e-01  -4.881794  5.984952 1.0000000
## Colea-Aphloia             -5.080000e+00 -14.693664  4.533664 0.8872844
## Eugenia-Aphloia           -3.080000e+00 -13.987830  7.827830 0.9997014
## Fraxinus-Aphloia          -5.080000e+00 -18.198580  8.038580 0.9914109
## Homalanthus-Aphloia       -1.413333e+00  -9.529018  6.702352 0.9999992
## Mimusops-Aphloia          -3.903529e+00  -9.515561  1.708503 0.5226867
## Paragenipa-Aphloia        -3.080000e+00  -7.914379  1.754379 0.6694074
## Plantago-Aphloia          -2.580000e+00 -10.695685  5.535685 0.9988719
## Psidium-Aphloia           -4.013333e+00  -9.843814  1.817147 0.5409713
## Vaccinium-Aphloia          3.786667e+00  -2.043814  9.617147 0.6390108
## Celmisia-Argyranthemum     8.333333e-02 -10.223597 10.390264 1.0000000
## Cerastium-Argyranthemum   -3.684211e-01 -10.189226  9.452384 1.0000000
## Colea-Argyranthemum       -6.000000e+00 -18.623360  6.623360 0.9481455
## Eugenia-Argyranthemum     -4.000000e+00 -17.634787  9.634787 0.9995380
## Fraxinus-Argyranthemum    -6.000000e+00 -21.460395  9.460395 0.9912278
## Homalanthus-Argyranthemum -2.333333e+00 -13.856832  9.190165 0.9999947
## Mimusops-Argyranthemum    -4.823529e+00 -14.744294  5.097235 0.9381895
## Paragenipa-Argyranthemum  -4.000000e+00 -13.502520  5.502520 0.9813509
## Plantago-Argyranthemum    -3.500000e+00 -15.023498  8.023498 0.9993178
```

```
## Psidium-Argyranthemum     -4.933333e+00 -14.979286  5.112620 0.9333320
## Vaccinium-Argyranthemum    2.866667e+00  -7.179286 12.912620 0.9996630
## Cerastium-Celmisia        -4.517544e-01  -7.034444  6.130935 1.0000000
## Colea-Celmisia            -6.083333e+00 -16.390264  4.223597 0.7766466
## Eugenia-Celmisia          -4.083333e+00 -15.606832  7.440165 0.9964435
## Fraxinus-Celmisia         -6.083333e+00 -19.718120  7.551454 0.9690535
## Homalanthus-Celmisia      -2.416667e+00 -11.342730  6.509397 0.9998166
## Mimusops-Celmisia         -4.906863e+00 -11.637774  1.824049 0.4396580
## Paragenipa-Celmisia       -4.083333e+00 -10.180995  2.014329 0.5883188
## Plantago-Celmisia         -3.583333e+00 -12.509397  5.342730 0.9879142
## Psidium-Celmisia          -5.016667e+00 -11.930766  1.897432 0.4480066
## Vaccinium-Celmisia         2.783333e+00  -4.130766  9.697432 0.9875981
## Colea-Cerastium           -5.631579e+00 -15.452384  4.189226 0.8106784
## Eugenia-Cerastium         -3.631579e+00 -14.722405  7.459247 0.9984469
## Fraxinus-Cerastium        -5.631579e+00 -18.902706  7.639548 0.9799606
## Homalanthus-Cerastium     -1.964912e+00 -10.324937  6.395112 0.9999663
## Mimusops-Cerastium        -4.455108e+00 -10.415022  1.504805 0.3956421
## Paragenipa-Cerastium      -3.631579e+00  -8.865782  1.602624 0.5271016
## Plantago-Cerastium        -3.131579e+00 -11.491603  5.228445 0.9937838
## Psidium-Cerastium         -4.564912e+00 -10.730962  1.601137 0.4125106
## Vaccinium-Cerastium        3.235088e+00  -2.930962  9.401137 0.8926350
## Eugenia-Colea              2.000000e+00 -11.634787 15.634787 0.9999999
## Fraxinus-Colea             3.241851e-14 -15.460395 15.460395 1.0000000
## Homalanthus-Colea          3.666667e+00  -7.856832 15.190165 0.9988607
## Mimusops-Colea             1.176471e+00  -8.744294 11.097235 1.0000000
## Paragenipa-Colea           2.000000e+00  -7.502520 11.502520 0.9999914
## Plantago-Colea             2.500000e+00  -9.023498 14.023498 0.9999875
## Psidium-Colea              1.066667e+00  -8.979286 11.112620 1.0000000
## Vaccinium-Colea            8.866667e+00  -1.179286 18.912620 0.1500316
## Fraxinus-Eugenia          -2.000000e+00 -18.296688 14.296688 1.0000000
## Homalanthus-Eugenia        1.666667e+00 -10.956693 14.290027 1.0000000
## Mimusops-Eugenia          -8.235294e-01 -12.002965 10.355906 1.0000000
## Paragenipa-Eugenia         8.881784e-16 -10.810000 10.810000 1.0000000
## Plantago-Eugenia           5.000000e-01 -12.123360 13.123360 1.0000000
## Psidium-Eugenia           -9.333333e-01 -12.224010 10.357343 1.0000000
## Vaccinium-Eugenia          6.866667e+00  -4.424010 18.157343 0.7374447
## Homalanthus-Fraxinus       3.666667e+00 -10.909534 18.242867 0.9999232
## Mimusops-Fraxinus          1.176471e+00 -12.168798 14.521739 1.0000000
## Paragenipa-Fraxinus        2.000000e+00 -11.037350 15.037350 0.9999999
## Plantago-Fraxinus          2.500000e+00 -12.076200 17.076200 0.9999994
## Psidium-Fraxinus           1.066667e+00 -12.371926 14.505260 1.0000000
## Vaccinium-Fraxinus         8.866667e+00  -4.571926 22.305260 0.6131951
## Mimusops-Homalanthus      -2.490196e+00 -10.967423  5.987031 0.9995311
## Paragenipa-Homalanthus    -1.666667e+00  -9.650380  6.317047 0.9999923
## Plantago-Homalanthus      -1.166667e+00 -11.473597  9.140264 1.0000000
## Psidium-Homalanthus       -2.600000e+00 -11.223397  6.023397 0.9993715
## Vaccinium-Homalanthus      5.200000e+00  -3.423397 13.823397 0.7489840
## Paragenipa-Mimusops        8.235294e-01  -4.595903  6.242962 0.9999999
## Plantago-Mimusops          1.323529e+00  -7.153697  9.800756 0.9999998
## Psidium-Mimusops          -1.098039e-01  -6.433848  6.214240 1.0000000
## Vaccinium-Mimusops         7.690196e+00   1.366152 14.014240 0.0039969
## Plantago-Paragenipa        5.000000e-01  -7.483714  8.483714 1.0000000
## Psidium-Paragenipa        -9.333333e-01  -6.578672  4.712005 0.9999996
## Vaccinium-Paragenipa       6.866667e+00   1.221328 12.512005 0.0039804
```

```
## Psidium-Plantago          -1.433333e+00 -10.056730  7.190063 0.9999996
## Vaccinium-Plantago         6.366667e+00  -2.256730 14.990063 0.4173260
## Vaccinium-Psidium          7.800000e+00   1.281325 14.318675 0.0051365
```

```
# Is there an interaction between the genus and symmetry of flowers on the number of visitations to flo

# Null : There is no interaction between the symmetry and genus of flowers on pollination visitation.
# Alternate: There is an interaction between the symmetry and genus of flowers on pollination visitatio


Polaov <- aov(n.poll ~ Genus + symmetry, data = plantPolGenus20)
summary(Polaov)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Genus        14   1064   75.98   2.795 0.000909 ***
## symmetry      1      0    0.44   0.016 0.899356
## Residuals   168   4567   27.19
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 1 observation deleted due to missingness
```

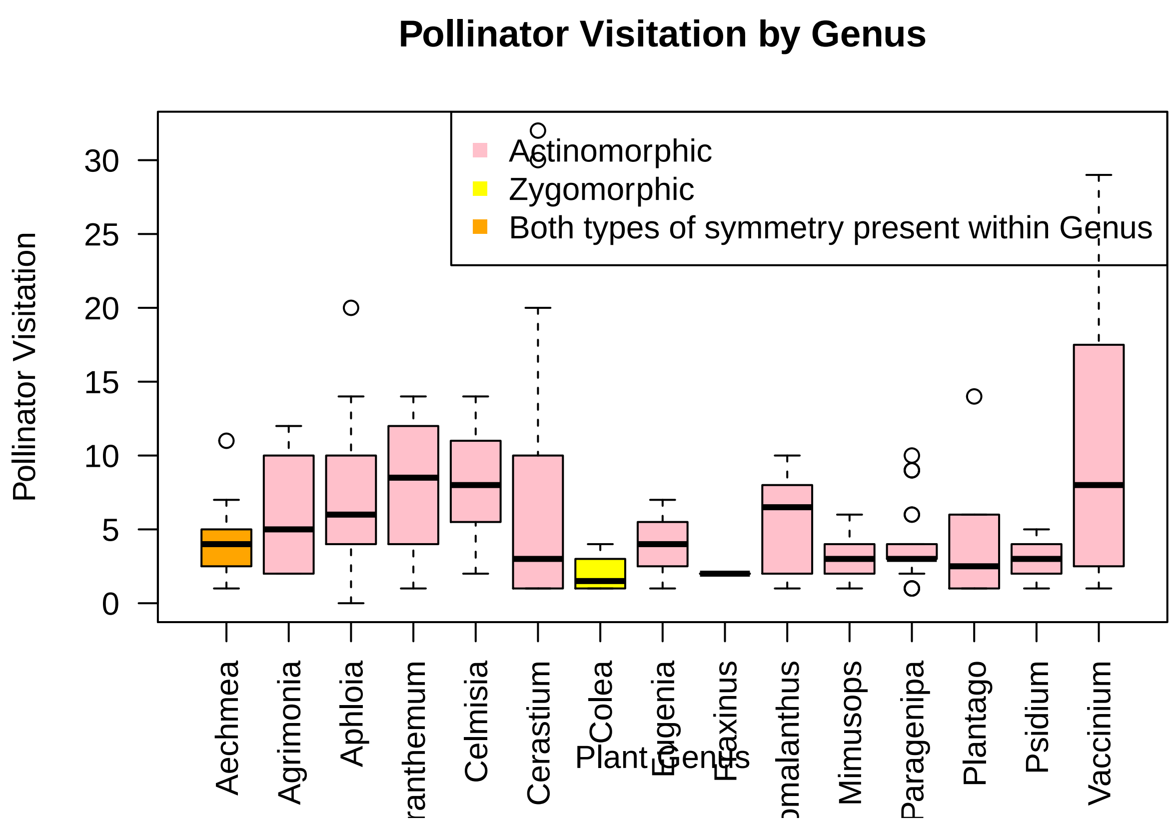
## Results

The one way ANOVA testing showed that there was a significant difference in pollinator visitation between genus groups. With a p value of 0.000232, being less than the alpha of 0.05, the null hypothesis that all the genuses have the same mean can be rejected. This means that the alternate must then be accepted, that there is a difference between means. This means of the genuses randomly selected, there is a difference in the visitation rate of pollinators based on genus. The results of the two way ANOVA testing showed that symmetry type did not correlate to a significant difference in pollinator visitation. While the p-value for the genus variable confirmed again that mean pollinator visitation differed by plant genus by being lower than the alpha, the p-value based on symmetry was not below alpha, so the null that there was no association between genus and symmetry impact on pollinator visitation is accepted. This result is interesting, as plants of a genus often have the same flower symmetry, so my initial belief was that there would be some interaction between symmetry and genus. This was not the case. This may be for two reasons. Not all species within a genus have the same symmetry type, as then these two variables would be completely dependent on each other. I think that the result of seeing no interaction between symmetry type and genus is more likely due to the flawed way that I generated the random sample of genus groups. I used all genus groups, randomized a sample of twenty, instead of stratifying the sample by splitting the sample into genuses with actinomorphic flowers and genuses with zygomorphic flowers. I did not stratify random sample from the original data set, because the majority of genuses sampled where actinomorphic, and because there were genuses in the dataset with different species, some of which were actinomorphic and some of which were zygomorphic. This means I could not aggregate the data based on symmetry type or genus. Instead, I believed that in sampling randomly from the entire data set I would get a relatively even distribution of actinomorphic and zygomorphic flowers. This was not the case. I believe that the result of seeing no association between symmetry type, genus, and pollinator visitation is at least partially due to having few genuses with solely zygomorphic flowers. I believe the method design could be changed to either a larger random sample, or to be stratified to ensure that sampling includes enough genuses from each symmetry type. Another issue with the methods is that because genus and symmetry type are not completely independent from each other, the assumption of independence needed for ANOVA may not be met. I think that additionally another step could be taken to clean the data, I should have eliminated genus groups with fewer than a certain number of species and samples, to ensure that the mean was a true reflection of the genus and not an observation from a small group of species within the genus. Because I did not originally remove genuses with fewer than two observations, I had to then remove them when they showed up in my random sample. The variance between the groups was also very high, because the amount of species sampled in the original dataset for each genus was not equal. Because the dataset was compiled from several other studies, I would classify the data found on the species within each genus to be samples of convince, because the species sampled in the data were the ones that had been sampled in previous studies. This creates a fair amount of bias and potential for error in the original dataset, which may also be present in the subset data from the random sample I created. Overall, it seems that from the limited sample that my computer could process in a realistic amount of time, there is some difference in the pollination of flowers based on plant genus.

### Plots

```
colors <- c("orange", "pink", "pink", "pink", "pink", "pink", "yellow","pink", "pink", "pink", "pink",

boxplot(n.poll ~ Genus, data = plantPolGenus20, xlab = "Plant Genus", ylab = "Pollinator Visitation", ma
legend("topright", legend= c("Actinomorphic", "Zygomorphic", "Both types of symmetry present within Genu
```




From this plot it is recognizable that few zygomorhpic gensus groups were selected, and that some of the genuses randomly selected had very limited data. Additionally, the most different mean was seen in a genus that had limited data (Machitus), which could have made a large impact on the results of the one way ANOVA.

### Refrences


Hadley Wickham (2019). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.0. (https://CRAN.R-project.org/package=stringr)

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL (http://www.R-project.org/.)

Yoder, Jeremy B.; Gomez, Giancarlo; Carlson, Colin J. (2020), Zygomorphic flowers have fewer potential pollinator species, Dryad, Dataset, (https://doi.org/10.5061/dryad.gxd2547j3)