Statistics Homework

Nicole Baker

2023-04-04

# Statistics Homework

## Question 1: My Data

##### The data that I currently have from my project revolves around collections made from traps this past field season. I have collected buprestid beetles in the genus *Chrysobothris* (and a handful of other related genera) and identified them to species.

* The genus *Chrysobothris* contains multiple pest species (ex. Flatheaded Appletree Borer) that commonly attack hardwoods like maple in nurseries and in the landscape.
* Damages caused by *Chrysobothris* can have a huge economic impact, however, the pests in this genus are relatively understudied with respect to distribution in Alabama, emergence windows and flight periods, and taxonomy.
* Characterizing the phenology of these beetles is key for well-informed control strategies. Additionally, developing more effective trapping methods can help improve beetle collection, as well as reduce pressure on the usual target hosts.

#### Part 1: Basic trapping

A set of ten prism panel traps were placed in three Locations across Alabama in February 2022 in order to collect information on basic *Chrysobothris* phenology: Hunter Trees in Alpine AL, Kreher Preserve in Auburn AL, and the Ornamental Horticulture Research Station in Mobile AL. I collected data on:

* Location of collection
* Date of collection
* Species collected
* Male:Female collections

Based on preliminary work done with *Chrysobothris* in other states, we expected to see peak emergences from May to August.

#### Part 2: Testing a new trap type

In Mobile, a trial with PVC pole traps was installed towards the end of May, when standard traps began to collect an influx of *Chrysobothris*. 6 different size diameters were set up in clusters of 6. I collected data on:

* Size of pole collected from
* Date collected
* Species collected
* Male:Female collections

Because these beetles bore into the trunks of trees, we hypothesize that pole traps will perform better than standard panels. we also hypothesized that smaller caliper poles (1/2 inch - 1 inch) would collect more beetles, since young nursery trees are often prime targets of attack.

#### Data assessment

In order to understand trapping dynamics, we will be graphing emergence windows, characterizing the present species across Alabama, and comparing sex ratios. analysis between sex and collection date could also be useful, since females are the ones laying eggs.

In order to determine if pole trap sizes have and effect on trapping, the number of individuals collected on each size will be conducted using an lm. In order to determine whether an effect of species and sex might also influence which size is most appealing, an ANOVA will be done with the six sizes against the number of individuals collected, species, and sex.

Bar graphs will be useful for data visualization.

## Question 2 & 3: Data Analysis and Discussion

First, I will make sure I have my packages uploaded:

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(lme4)

## Loading required package: Matrix  
##   
## Attaching package: 'Matrix'  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, pack, unpack

library(emmeans)  
library(multcomp)

## Loading required package: mvtnorm  
## Loading required package: survival  
## Loading required package: TH.data  
## Loading required package: MASS  
##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select  
##   
##   
## Attaching package: 'TH.data'  
##   
## The following object is masked from 'package:MASS':  
##   
## geyser

Now I will make sure my data file is loaded in appropriately. For this class I will be focusing on analyzing the data from part 2 of my experiment: pole size and *Chrysobothris* collections.

GenTrap<-read.csv("GenTrap.csv",na.strings="na")  
GenTrap

## Pole Rep Total Male Female X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9  
## 1 1/2 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 2 1/2 Inch 2 1 1 0 NA NA NA NA NA NA NA NA NA NA  
## 3 1/2 Inch 3 2 0 2 NA NA NA NA NA NA NA NA NA NA  
## 4 1/2 Inch 4 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 5 1/2 Inch 5 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 6 1/2 Inch 6 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 7 3/4 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 8 3/4 Inch 2 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 9 3/4 Inch 3 3 0 3 NA NA NA NA NA NA NA NA NA NA  
## 10 3/4 Inch 4 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 11 3/4 Inch 5 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 12 3/4 Inch 6 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 13 1 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 14 1 Inch 2 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 15 1 Inch 3 4 1 3 NA NA NA NA NA NA NA NA NA NA  
## 16 1 Inch 4 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 17 1 Inch 5 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 18 1 Inch 6 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 19 1.5 Inch 1 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 20 1.5 Inch 2 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 21 1.5 Inch 3 3 2 1 NA NA NA NA NA NA NA NA NA NA  
## 22 1.5 Inch 4 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 23 1.5 Inch 5 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 24 1.5 Inch 6 1 1 0 NA NA NA NA NA NA NA NA NA NA  
## 25 2 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 26 2 Inch 2 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 27 2 Inch 3 2 1 1 NA NA NA NA NA NA NA NA NA NA  
## 28 2 Inch 4 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 29 2 Inch 5 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 30 2 Inch 6 2 0 2 NA NA NA NA NA NA NA NA NA NA  
## 31 3 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 32 3 Inch 2 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 33 3 Inch 3 1 1 0 NA NA NA NA NA NA NA NA NA NA  
## 34 3 Inch 4 2 2 0 NA NA NA NA NA NA NA NA NA NA  
## 35 3 Inch 5 5 1 4 NA NA NA NA NA NA NA NA NA NA  
## 36 3 Inch 6 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 37 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 38 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 39 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 40 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 41 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 42 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 43 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 44 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 45 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 46 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 47 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 48 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 49 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 50 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 51 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 52 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 53 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 54 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 55 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 56 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 57 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 58 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 59 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 60 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 61 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 62 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 63 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 64 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 65 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 66 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 67 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 68 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 69 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 70 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 71 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 72 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 73 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 74 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 75 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 76 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 77 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 78 NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
## 79 NA NA NA NA NA NA NA NA NA NA NA NA NA NA

First, I will make sure that pole size is a factor:

GenTrap$Pole=as.factor(GenTrap$Pole)

Now I will log transform the data so that I can account for high numbers of zeros from my collections:

GenTrap$LogTotal=log(GenTrap$Total+1)  
head(GenTrap)

## Pole Rep Total Male Female X X.1 X.2 X.3 X.4 X.5 X.6 X.7 X.8 X.9  
## 1 1/2 Inch 1 0 NA NA NA NA NA NA NA NA NA NA NA NA  
## 2 1/2 Inch 2 1 1 0 NA NA NA NA NA NA NA NA NA NA  
## 3 1/2 Inch 3 2 0 2 NA NA NA NA NA NA NA NA NA NA  
## 4 1/2 Inch 4 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 5 1/2 Inch 5 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## 6 1/2 Inch 6 1 0 1 NA NA NA NA NA NA NA NA NA NA  
## LogTotal  
## 1 0.0000000  
## 2 0.6931472  
## 3 1.0986123  
## 4 0.6931472  
## 5 0.6931472  
## 6 0.6931472

#### Does Pole Trap Size Matter?

now I can run an ANOVA on my data in order to see whether collections of *Chrysobothris* vary by pole trap size.

results1<-lm(LogTotal~Pole,data=GenTrap)  
summary(results1)

##   
## Call:  
## lm(formula = LogTotal ~ Pole, data = GenTrap)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.71278 -0.51887 0.07192 0.19386 1.11015   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.49929 0.22328 2.236 0.0329 \*  
## Pole1.5 Inch 0.07833 0.31576 0.248 0.8058   
## Pole1/2 Inch 0.14591 0.31576 0.462 0.6473   
## Pole2 Inch 0.09796 0.31576 0.310 0.7585   
## Pole3 Inch 0.21349 0.31576 0.676 0.5042   
## Pole3/4 Inch 0.07833 0.31576 0.248 0.8058   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.5469 on 30 degrees of freedom  
## (43 observations deleted due to missingness)  
## Multiple R-squared: 0.01702, Adjusted R-squared: -0.1468   
## F-statistic: 0.1039 on 5 and 30 DF, p-value: 0.9906

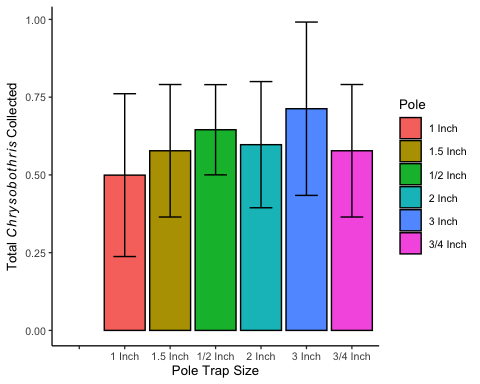
results2<-anova(results1)  
results2

## Analysis of Variance Table  
##   
## Response: LogTotal  
## Df Sum Sq Mean Sq F value Pr(>F)  
## Pole 5 0.1554 0.031077 0.1039 0.9906  
## Residuals 30 8.9736 0.299119

based on my statistical analysis, *Chrysobothris* collections are not significantly influenced by pole trap size. Low overall collection totals likely influence this result. In order to visualize the data, I can plot it:

plot1<- ggplot(GenTrap,aes(x=Pole,y=LogTotal,fill=Pole))+  
 stat\_summary(fun=mean,geom="bar",position="dodge",color="black")+  
 stat\_summary(fun.data=mean\_se,geom="errorbar",position="dodge",width=0.5)+  
 theme\_classic(base\_size=10)+  
 labs(x="Pole Trap Size",  
 y=expression('Total'~italic(Chrysobothris)~ 'Collected'))  
plot1

## Warning: Removed 43 rows containing non-finite values (`stat\_summary()`).  
## Removed 43 rows containing non-finite values (`stat\_summary()`).

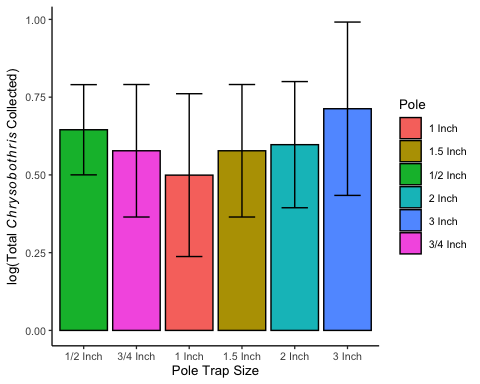


In the above plot, the error bars show the wide variation in collections across the six pole sizes.

Since the x axis is all out of order, here I am fixing it by using the scale\_x\_discrete function:

plot2<-ggplot(GenTrap,aes(x=Pole,y=LogTotal,fill=Pole))+  
 stat\_summary(fun=mean,geom="bar",position="dodge",color="black")+  
 stat\_summary(fun.data=mean\_se,geom="errorbar",position="dodge",width=0.5)+  
 theme\_classic(base\_size=10)+  
 labs(x="Pole Trap Size",  
 y=expression('log(Total'~italic(Chrysobothris)~ 'Collected)'))+  
 scale\_x\_discrete(limits=c("1/2 Inch","3/4 Inch","1 Inch","1.5 Inch","2 Inch","3 Inch"))  
plot2

## Warning: Removed 43 rows containing non-finite values (`stat\_summary()`).  
## Removed 43 rows containing non-finite values (`stat\_summary()`).



#### Male vs Female Collections

Now, I will go ahead and look at males vs females on pole traps, and whether this varies overall and by pole size:

results2<-lm(LogTotal~Male+Female+Pole,data=GenTrap)  
summary(results2)

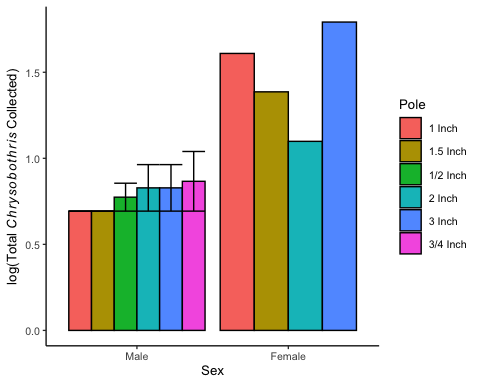
##   
## Call:  
## lm(formula = LogTotal ~ Male + Female + Pole, data = GenTrap)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.099138 -0.023881 -0.007614 0.023255 0.093111   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.387974 0.037301 10.401 1.58e-08 \*\*\*  
## Male 0.337753 0.020090 16.812 1.37e-11 \*\*\*  
## Female 0.298811 0.011882 25.148 2.73e-14 \*\*\*  
## Pole1.5 Inch 0.001036 0.042365 0.024 0.981   
## Pole1/2 Inch 0.019904 0.039485 0.504 0.621   
## Pole2 Inch 0.049953 0.040659 1.229 0.237   
## Pole3 Inch -0.030075 0.042602 -0.706 0.490   
## Pole3/4 Inch 0.030243 0.040934 0.739 0.471   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.05277 on 16 degrees of freedom  
## (55 observations deleted due to missingness)  
## Multiple R-squared: 0.983, Adjusted R-squared: 0.9755   
## F-statistic: 131.8 on 7 and 16 DF, p-value: 6.087e-13

results3<-anova(results2)  
results3

## Analysis of Variance Table  
##   
## Response: LogTotal  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Male 1 0.64819 0.64819 232.8029 5.916e-11 \*\*\*  
## Female 1 1.90873 1.90873 685.5366 1.452e-14 \*\*\*  
## Pole 5 0.01229 0.00246 0.8826 0.515   
## Residuals 16 0.04455 0.00278   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

here, we can see that experiment-wide male and female collections were significantly different. This was expected, since Females are likely using specific visual cues for egg laying. Since pole traps are mimicking tree hosts, it makes sense that more females would be collected than males overall. Below I have plotted Male and Female collections:

plot3<-GenTrap %>%  
 filter(!is.na(Male)) %>%  
 filter(!is.na(Female)) %>%  
 ggplot(aes(x=Male & Female,y=LogTotal,fill=Pole))+ stat\_summary(fun=mean,geom="bar",position="dodge",color="black")+  
 stat\_summary(fun.data=mean\_se,geom="errorbar",position="dodge")+  
 theme\_classic(base\_size=10)+  
 labs(x="Sex",  
 y=expression('log(Total'~italic(Chrysobothris)~ 'Collected)'))+  
 scale\_x\_discrete(labels=c("Male","Female"))  
plot3



#### Species in Mobile

Finally, I will make a visual depicting totals of various species in Alabama.

First I need to load in the data file with species information:

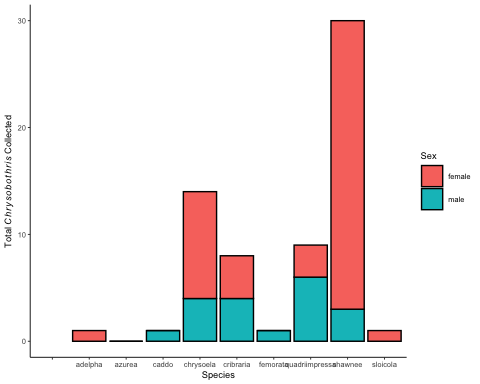
Species<-read.csv("MobileSpecies.csv")  
Species

## Species Sex Total  
## 1 adelpha male 0  
## 2 adelpha female 1  
## 3 azurea male 0  
## 4 azurea female 0  
## 5 caddo male 1  
## 6 caddo female 0  
## 7 chrysoela male 4  
## 8 chrysoela female 10  
## 9 cribraria male 4  
## 10 cribraria female 4  
## 11 femorata male 1  
## 12 femorata female 0  
## 13 quadriimpressa male 6  
## 14 quadriimpressa female 3  
## 15 shawnee male 3  
## 16 shawnee female 27  
## 17 sloicola male 0  
## 18 sloicola female 1  
## 19 NA  
## 20 NA  
## 21 NA  
## 22 NA  
## 23 NA  
## 24 NA  
## 25 NA  
## 26 NA  
## 27 NA  
## 28 NA  
## 29 NA  
## 30 NA  
## 31 NA  
## 32 NA  
## 33 NA  
## 34 NA  
## 35 NA  
## 36 NA  
## 37 NA  
## 38 NA  
## 39 NA  
## 40 NA  
## 41 NA  
## 42 NA  
## 43 NA  
## 44 NA  
## 45 NA  
## 46 NA  
## 47 NA  
## 48 NA  
## 49 NA  
## 50 NA  
## 51 NA  
## 52 NA  
## 53 NA  
## 54 NA  
## 55 NA  
## 56 NA  
## 57 NA  
## 58 NA  
## 59 NA  
## 60 NA  
## 61 NA  
## 62 NA  
## 63 NA  
## 64 NA  
## 65 NA  
## 66 NA  
## 67 NA  
## 68 NA  
## 69 NA  
## 70 NA  
## 71 NA  
## 72 NA  
## 73 NA  
## 74 NA  
## 75 NA  
## 76 NA  
## 77 NA  
## 78 NA  
## 79 NA

In order to visualize species totals in Mobile, I will make a plot:

plot4<-ggplot(Species,aes(x=Species,y=Total,fill=Sex))+  
 geom\_bar(stat="identity",color="black")+  
 theme\_classic(base\_size=7)+  
 labs(x="Species",  
 y=expression('Total'~italic(Chrysobothris)~ 'Collected'))  
plot4

## Warning: Removed 61 rows containing missing values (`position\_stack()`).



here you can see the distribution of individuals collected in Mobile in 2022. We can see clearly that the species collected most frequently was *Chrysobothris shawnee*. Of all the species collected across Alabama, the only one not found in Mobile was *Chrysobothris azurea*.

It is surprising that more individuals of *Chrysobothris femorata* were not found, however, this species is extremely similar to *C. shawnee*. *C. shawnee* and *C. femorata* are considered the same species by some resources, however, others categorize them as separate species. Based on taxonomic resources I used from the official Flatheaded Borer project group, I considered them to be 2 separate species.

Additionally, identification of these beetles can be extremely challenging without male genitalia. Since the vast majority of *C. shawnee* individuals I collected were female, it is possible that the more obscure female characteristics between *C. shawnee* and *C. femorata* prevented 100% accurate species identification of my collections.

In order to more accurately determine the identities of these beetles, genetic analysis would be useful. Since comprehensive, accurate genetic records of *Chrysobothris* species are not currently available, collaborators on the Flatheaded Borer research project are currently working on characterizing the genomes of collected *Chrysobothris* species across the southeast.

#### *C.shawnee* Phenology

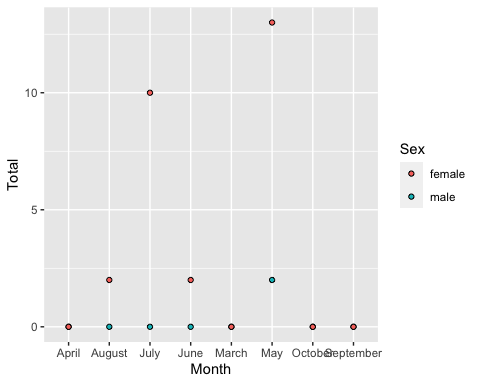
Since *C. shawnee* is the only species I collected a large number of throughout the trapping season, I will use this information to graphically depict basic phelology by grouping collections by month. I need to use a new data file:

ShawneeData<-read.csv("ShawneeMobile.csv")  
ShawneeData

## Month Sex Total  
## 1 March male 0  
## 2 March female 0  
## 3 April male 0  
## 4 April female 0  
## 5 May male 2  
## 6 May female 13  
## 7 June male 0  
## 8 June female 2  
## 9 July male 0  
## 10 July female 10  
## 11 August male 0  
## 12 August female 2  
## 13 September male 0  
## 14 September female 0  
## 15 October male 0  
## 16 October female 0

Now I can make a graph:

plot5<-ggplot(ShawneeData,aes(x=Month,y=Total,fill=Sex))+  
 geom\_point(stat="identity",shape=21,color="black")  
plot5



This shows the number of collected specimens by month, but I am not sure how to connect the data points or make it depict information over time more clearly. This can give some insight to emergence and flight periods of this individual species, however, conclusions are limited due to our collections only being made every couple of weeks - once a month on traps, which give no host information and no clear emergence dates.

#### Overall Conclusion

Overall, my data set consists of exploratory species collections from spring/summer 2022 in order to more clearly understand the presence of *Chrysobothris* species across Alabama. Since these beetles are common, problematic pests of many ornamental, fruit, and nut crops in the southeastern US, understanding the distribution and phenology of them is key for more comprehensive and successful management.

Due to low overall collections, This year in 2023 I will be utilizing alternative sites to maximize trapping. Additionally, I will be transitioning to using all pole style traps, since they more directly mimic potential hosts. Finally, a set of red maples will be used as sacrificial hosts alongside trapping in Mobile. By replacing a tree alongside my traps every two weeks, we can later harvest the wood and rear out any potential borers the following spring. Gaining information that connects these beetles directly to a host is key for understanding their environmental dynamics.