# EDA_CharlieGiannini

## Charlie C. Giannini

## 2024-03-01

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(ggplot2)
library(forcats)
load("C:/Users/Scott/Desktop/UnderstandingR/Data/nswodata.RData")
```

## Data Dscription

The data I will be exploring for this project comes from a long term study station monitoring Norther Saw-Whet Owl migration. The Northern Saw-Whet Owl or NSWO is a hand sized owl that because of its secretive ways was once thought to be endangered. Due to extensive banding projects it is now known to be the most numerous owl in north america. The Station at the Big Chico Creek Ecological Reserve has been collecting data during the migration season (October-November) since 2005. I myself have worked at the station for the past 5 years. The main Variable we collect data for are Age, Wing Chord, Mass and Sex. Those data are gathered from the owl. We also take data on Time and date of capture. For the purposes of this project I would like to focus on the the variables Age and Sex. It is well known that more females are caught than males and we use a male sound lure to draw the owls in. We don't know for sure why this is though. One reason could be that males do not migrate because they hold territory year round. Cavity nests can be hard to come by. So we would expect to see mostly first year males and all years of females.It would also be interesting to look at when a majority of the captures took place by date but also by time of night.

## Univariate Exploration

The first variable I want to look at is age. It is separated into three age classes hatch year, second year, and after second year

```r
table(nswo_dataF$AGE)
```

```
##
##       AHY  ASY  ATY BALO BDES   HY   SY   TY
##     5   65  166    6    1    1  576  340    6
```

Immediately I see some problems and that is that there are 9 different ages when there should be only three. so new step one is fixing this.

```r
nswo3 <- nswo_dataF %>% mutate(AGE, recode(AGE, TY="ASY", ATY="ASY")) %>% filter(AGE %in% c("ASY","HY",
```

```r
nswo3$AGE %>% fct_relevel("HY", "SY", "ASY", "AHY") %>% table()
```

```
## .
##  HY  SY ASY AHY
## 576 340 166  65
```

I left the after hatch year on because although it is a bit vauge some of those might be males which will mean something later on. It might make a bit more sense to show it in a proportions table though when we get to the bivariate analysis.

In a bar chart form it looks like this

```r
ggplot(nswo3, aes(x=forcats::fct_infreq(AGE))) + geom_bar() + xlab("AGE")
```
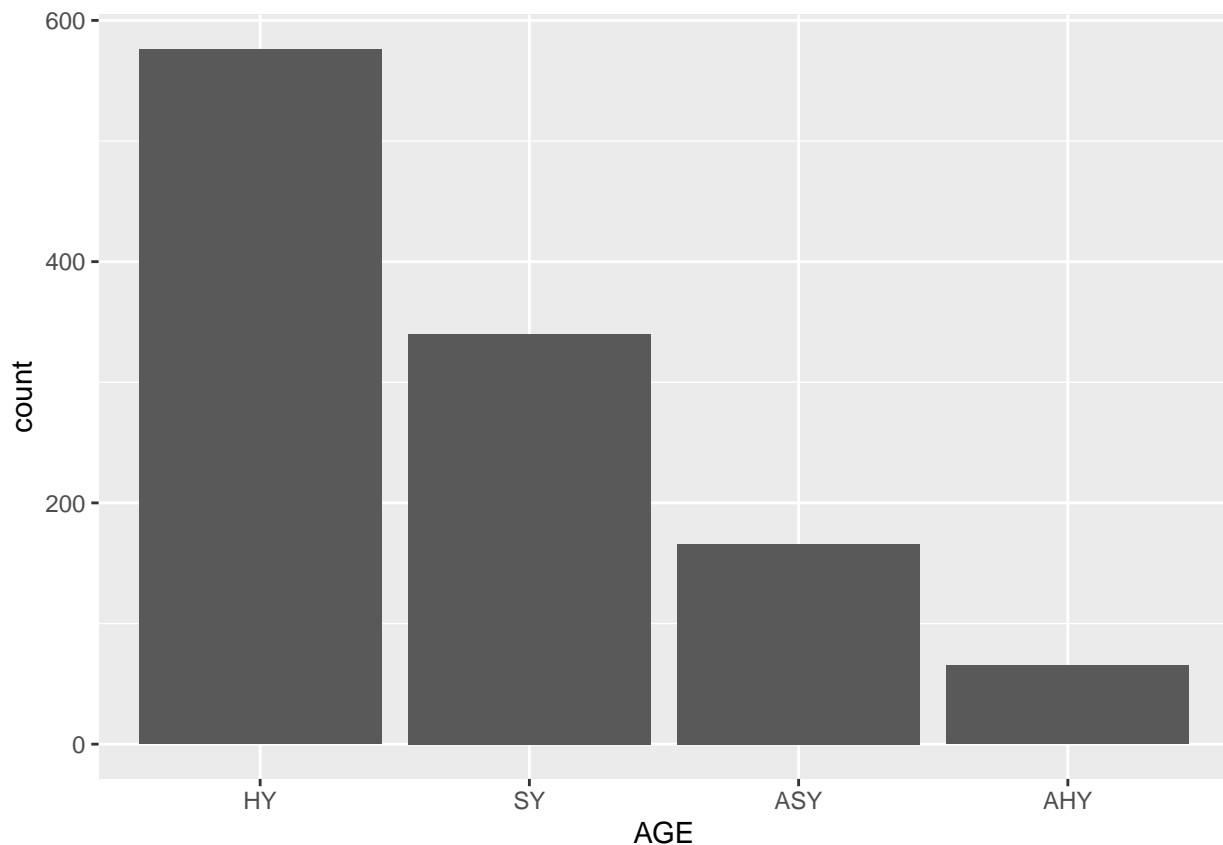
The capture drop off from Hatch Year to after Hatch year makes sense given that the owls in question are very small and easily predated. From this data you could infer some life history information based on mortality from year to year. The incresed number of Hatch year birds is because a lot of them don't make it to the second year.

```
table(nswo3$SEX)
```

```
##
##   F   M   U
## 785 103 259
```

```
ggplot(nswo3, aes(x=SEX)) + geom_bar()
```
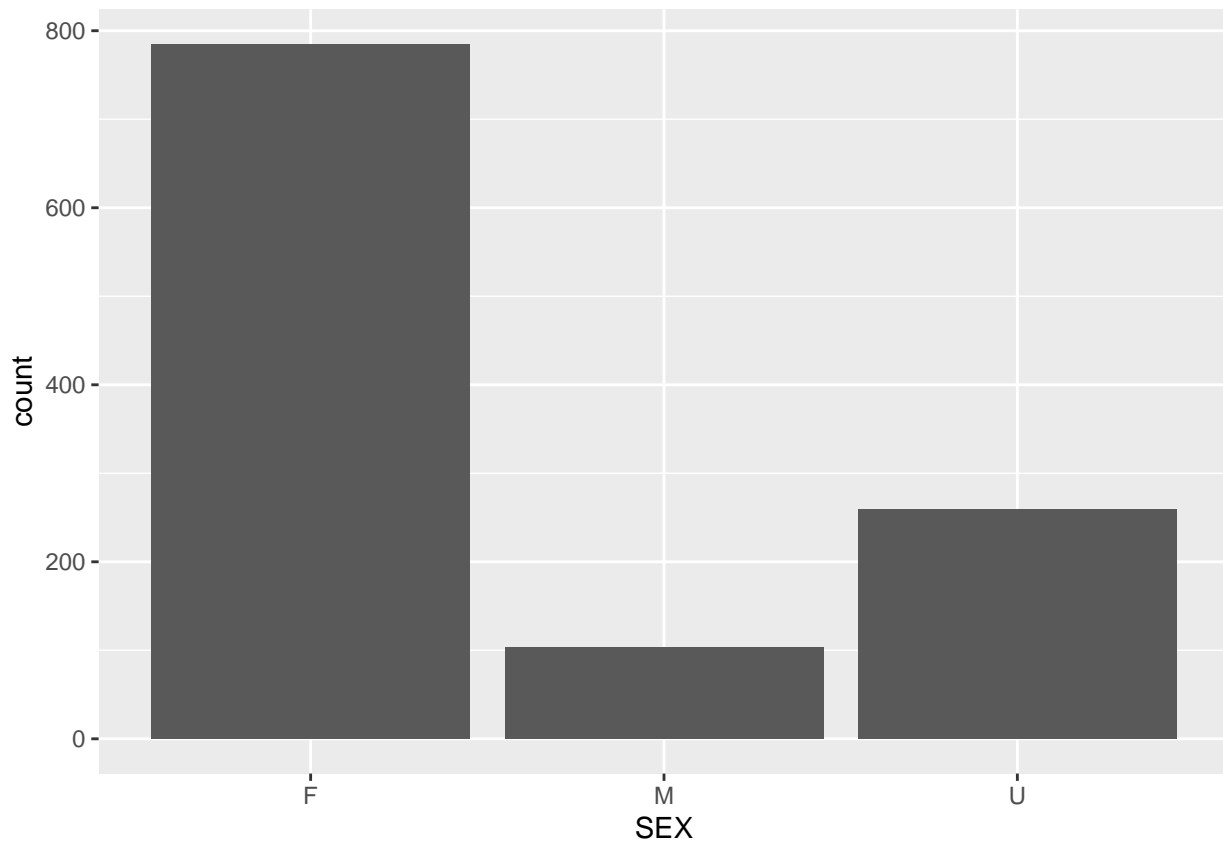


This just goes to show how much the females captured outnumber the males captured. There is a third column, unknown, and this is just due to the way the birds are sexed in the feild and there being some size overlap.

## Bivariate Exploration

In this section I will be looking at the relationship between age and sex to see how the proportion of after hatch year males holds up with that of females.

```
table(nswo3$AGE, nswo3$SEX) %>% prop.table(margin=2) %>% round(3)
```

```
##
##         F     M     U
##   AHY 0.062 0.029 0.050
##   ASY 0.159 0.097 0.120
##   HY  0.460 0.699 0.552
##   SY  0.318 0.175 0.278
```

From this table we can see that the males caught are Hatch years 70% of the time and in females the percentage of HY's is 46%. I would say this is a pretty big difference and might be what you would expect to see if Older males stay back to defend breeding territory.

Now here is another look a this but in bar chart format.

```
ggplot(nswo3, aes(x=SEX, fill=AGE)) + geom_bar(position = "dodge")
```
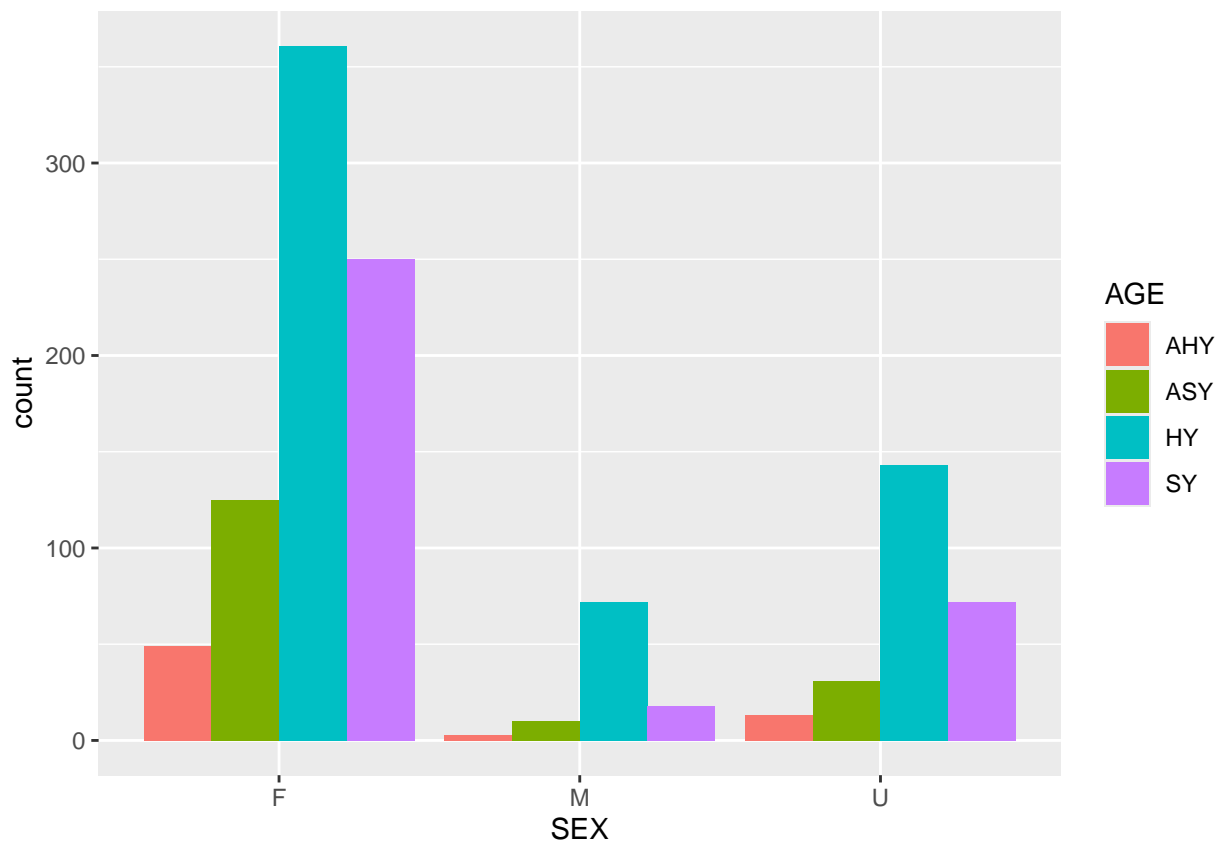


For the purposes of this project I much prefer the table but this graph is also a nice visual because it also makes it clear the descreptancy in the number of observations for each variable. This is always something that should be considered.

##Conclusion

The data does appear to suport my original hypothesis that male owls stay back to defend key nest sites. However, That is only one possible explanation and there could be many other factors at play. It would be an interesting direction to go now that we have evidence that we get mostly hatch year males. This could involve radio telemertry of males durring the non-breeding season.