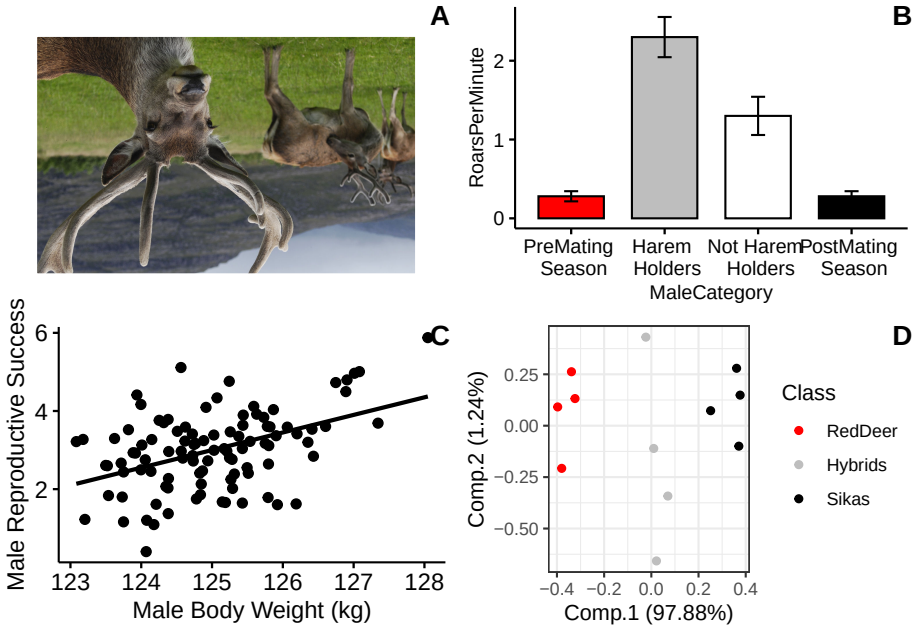
behaviouR: R package and tutorials for online teaching of fundamental concepts in behavior and ecology

Dena J. Clink

2020-08-25

Contents

Welcome



To get started you should download the package from Github using the following code.

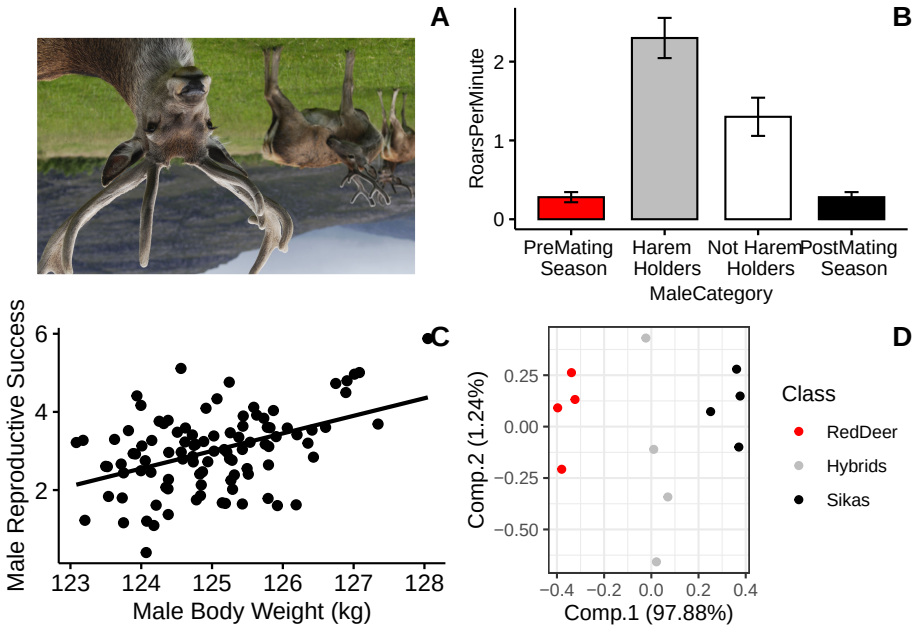
```
# Install devtools from CRAN
install.packages("devtools")

# Install 'behaviouR' package from Gitub
devtools::install_github("https://github.com/DenaJGibbon/behaviouR")
```

Note: If you have never used R before I highly recommend that you check out the primers here: https://rstudio.cloud/learn/primers.

You can navigate using the tabs at the left and/or the arrows.

Computer Lab 1. Data exploration and visualization



Background

Our examples for this lab will focus on roaring behavior in red deer. Red deer on Rum Island, Scotland form harems during the mating season. Harem holding males fight off other males during this time, and many of these males have signs of fighting injuries. Clutton-Brock and colleagues hypothesized that because of the injury associated with fights that red deer males should use honest indicators of each others fighting abilities, and that the main indicator of strength and

fighting ability was the male roars.

Goals of the exercises
The main goal(s) of today's lab are to:
1) teach you about different types of data visualization and types of data.
2) help you to start to become familiar with using R.
3) to get you to think about the ways that scientists analyze data.

Suggested readings:
Clutton-Brock, Tim H., and Steven D. Albon. "The roaring of red deer and the evolution of honest advertisement." *Behaviour* 69.3-4 (1979): 145-170.

Reby D, McComb K. Anatomical constraints generate honesty: acoustic cues to age and weight in the roars of red deer stags. Animal behaviour. 2003 Mar 1;65(3):519-30.

Long, A. M., N. P. Moore, and T. J. Hayden. "Vocalizations in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*), and red× sika hybrids." Journal of Zoology 244.1 (1998): 123-134.

Getting started
First we need to load the relevant packages. Packages contain all the functions that are needed for data analysis. The 'behaviouR' package loads many other packages that have important functions.

```
# First we load the relevant packages
library(behaviouR)
library(ggpubr)
library(ggfortify)
```

Part 1. Categorical data

Categorical variables represent types of data that can be divided into groups. Our first example will census a simulated population of red deer to determine the proportion of individuals in different developmental stages.

```
## Lab 1a. Categorical data
# Here we create a simulated population with four categories (Infant, Juvenile, AdultF
DeerPopulationDF <- data.frame(DevelopmentStage=c('Infant','Juvenile','AdultFemale','Ad
                               NumberOfIndividuals=c(15,50,125,200))

# We then print the object so that we can see the output
DeerPopulationDF
```

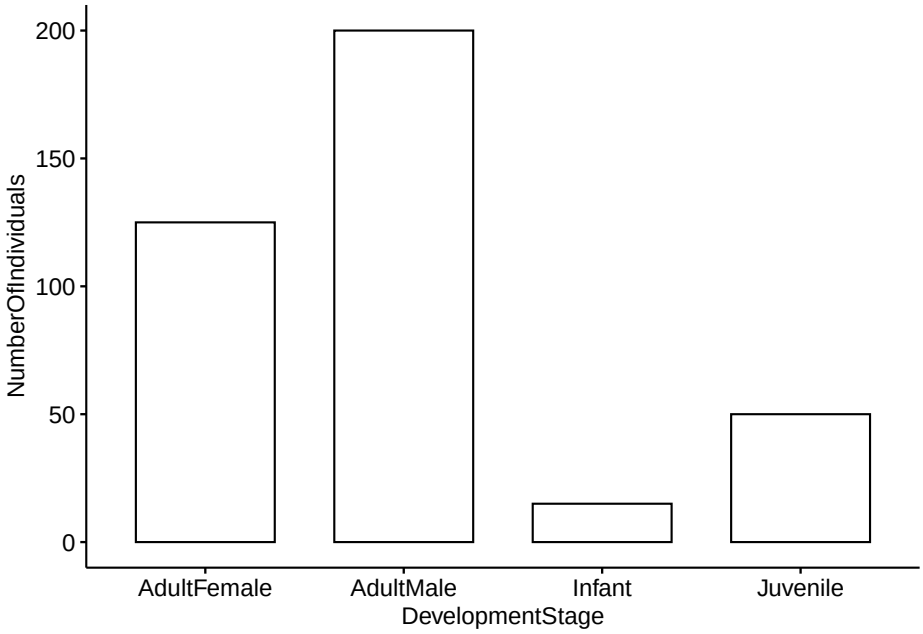
```
##   DevelopmentStage NumberOfIndividuals
## 1           Infant                  15
## 2         Juvenile                  50
## 3      AdultFemale                 125
```

```
## 4        AdultMale                  200
```

Now we want to plot the data. We will use a simple barplot to start.

Throughout these tutorials we will use functions from the 'ggpubr' package for plotting; see https://rpkgs.datanovia.com/ggpubr/ for more details. For more information on the great work the developer is doing see http://www.alboukadel.com/ and http://www.sthda.com/english/.

```
# Now we want to plot the data. We will use a simple barplot to start.
ggbarplot(DeerPopulationDF, x='DevelopmentStage', y='NumberOfIndividuals')
```



Please answer the following questions:
Question 1. How would you interpret this figure? Which category has the most individuals, and which has the least?

Question 2a. This is a simulated popuation, but what do you think the ratio of males to females would mean for male-male competition?

Question 2b. Modify the code above so that our simulated deer population has a more even ratio of males to females.

Part 2. Categorical and continuous data

Our second example will investigate roaring rates in deer as a function of status and breeding season. The categories we will use are: pre-mating season, harem holder, not harem holder and post-mating season.Our outcome variable

(mean number of roars per minute) is continuous; it represents actual numerical measurements.

We will work with a toy dataset that is based on Figure 2 in Clutton-Brock et al. 1979. For each category we have five observations which represent the average number of calls per minute. For each category we have a total of five observations.

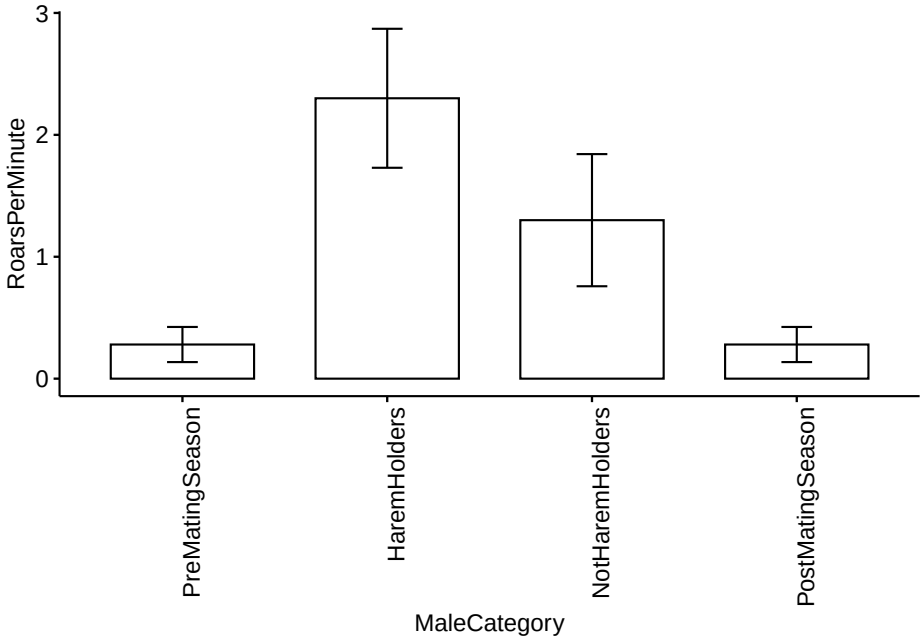
```
## Lab 1b. Categorical and continuous data
# Load the dataset so that we can use it
data("MaleDeerRoarDF")

# We can check the structure of the dataframe by using the command 'head'
head(MaleDeerRoarDF)
```

```
##      MaleCategory RoarsPerMinute
## 1 PreMatingSeason           0.25
## 2 PreMatingSeason           0.50
## 3 PreMatingSeason           0.25
## 4 PreMatingSeason           0.30
## 5 PreMatingSeason           0.10
## 6    HaremHolders           2.50
```

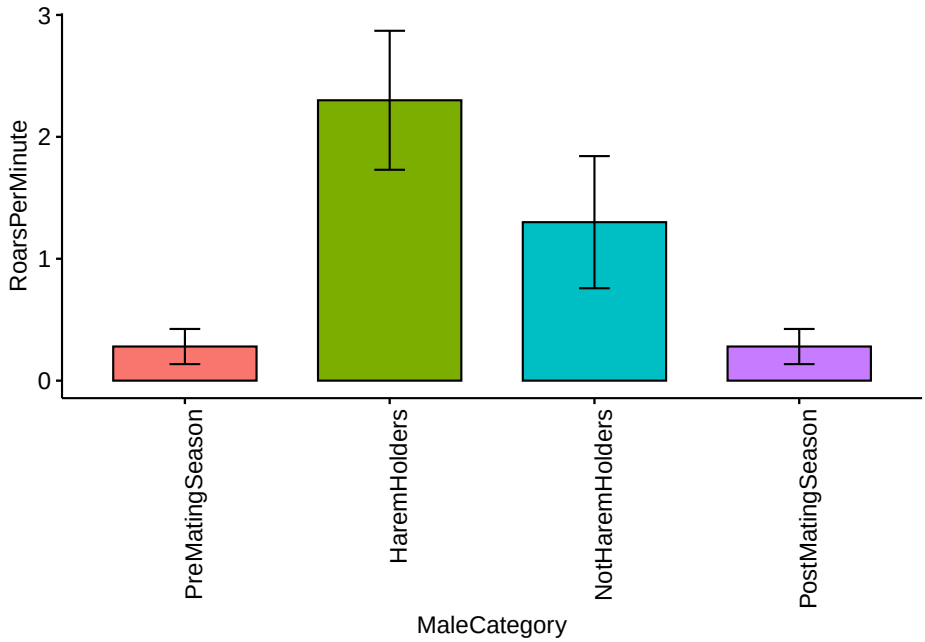
Now we will plot the categorical data. First we will use a barplot with error bars representing the standard deviation. See https://www.biologyforlife.com/interpreting-error-bars.html and https://www.biologyforlife.com/standard-deviation.html for more information about error bars.

```
# Now we will plot the categorical data with standard deviations.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute',
          add = c("mean_sd"),xtickslab.rt = 90)
```



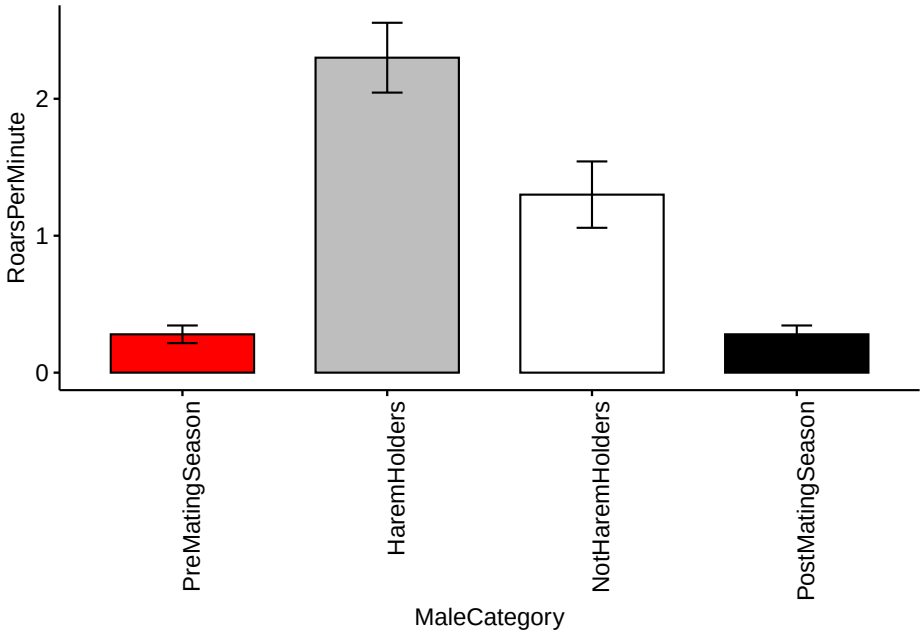
Although it shows us what we need, we can also include some colors in our plot. To do this I added the 'fill = 'MaleCategory' argument, which tells R to color the plot based on male categories.

```
# Now we will plot the categorical data with colors for each category.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute', fill = 'MaleCategory',
          add = c("mean_sd"),xtickslab.rt = 90)+ theme(legend.position = "none")
```



Color for plots is often based on personal preference, below I modified the colors using the 'palette' argument. NOTE: The use of color-blind friendly palettes is becoming increasingly popular among scientists.

```
# Now we will plot the categorical data with user-specified colors for each category.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute', fill = 'MaleCategory',
          palette = c('red','gray','white','black'),
          add = c("mean_se"),xtickslab.rt = 90)+ theme(legend.position = "none")
```



Please answer the following questions:

Question 3. How do you interpret the barplot figure?

Question 4. Visit this site (http://www.stat.columbia.edu/~tzheng/files/Rcolor.pdf) and change the colors of the plot. Hint: change 'palette = c('red','gray','white','black')'to 'palette = c('color1','color2','color3','color4')'.

Part 3. Categorical and continuous data

Continuous data represent numerical measurements, and when both our variables of interest are continuous we can plot them in a different way.

For this example we will consider the relationship between male red deer body weight (in kilograms) and male reproductive success (which is estimated by the number of days that he associated with females during the breeding season).

Below we have a function that will simulate data for us, and we can specify the level of correlation.

We also assign a mean body weight for males in our population. We can also change the male reproductive success value. The starting values that we will use are from Reby & McComb 2003.

```
## Lab 1c. Categorical and continuous data
# This is the function that simulates our data. N is the number of individuals,
# CorrelationCoefficient tells us how correlated our data are,
```

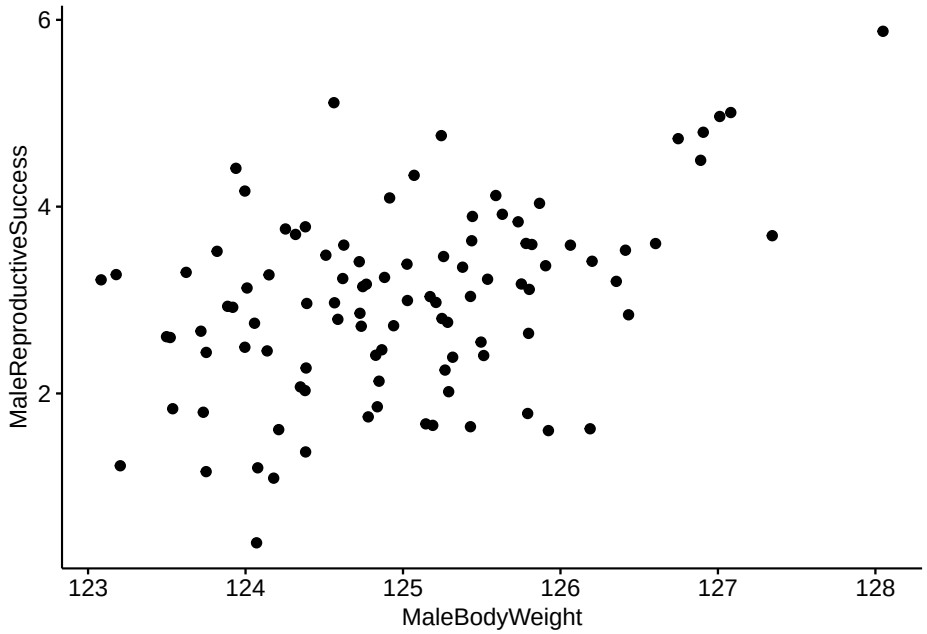
```
# MaleMeanBodyWeight is the mean body weight of males in our population and
# MeanFemalesInHarem is the mean number of females in the harem.
MaleRedDeerDF <- CorrelatedDataSimulationFunction(N=100,
                                                  CorrelationCoefficient= 0.45,
                                                  MaleMeanBodyWeight = 125,
                                                  MaleReproductiveSuccess = 3)

# We can check the output
head(MaleRedDeerDF)
```

```
##   MaleBodyWeight MaleReproductiveSuccess
## 1       125.2576                3.466009
## 2       125.1888                1.658102
## 3       124.7266                2.858253
## 4       124.1372                2.455370
## 5       125.4283                1.643711
## 6       124.3844                2.272255
```

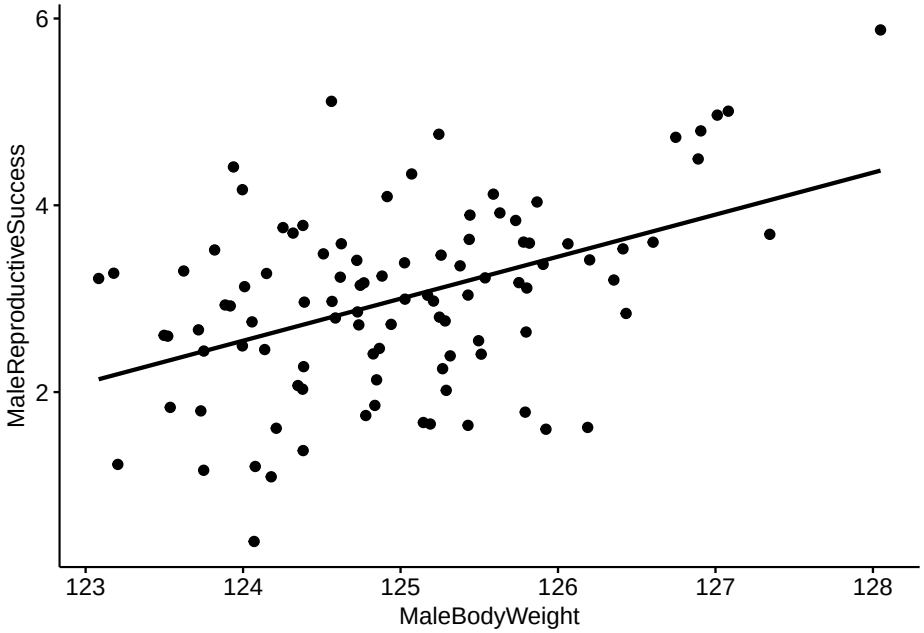
Now we plot the data the data as before, but this time we make a scatterplot

```
# Make a scatterplot of the data.
ggscatter(data=MaleRedDeerDF,x='MaleBodyWeight',y='MaleReproductiveSuccess')
```



It looks like there is a correlation between our two variables, but it would be better visualized with a trend line. We can do that by including add='reg.line'.

```
# Make a scatterplot with a trendline.
ggscatter(data=MaleRedDeerDF,x='MaleBodyWeight',y='MaleReproductiveSuccess',
          add='reg.line')
```



Now we can use R to create a linear model to quantitatively investigate the relationship between our two variables. Linear models describe the relationship of a continuous variable (in this case 'MaleReproductiveSuccess') as a function of predictor variable(s). In the linear model below we are interested in the relationship between MaleBodyWeight and MaleReproductiveSuccess.

```
# Create a linear model where MaleBodyWeight is the independent variable
# and FemalesInHarem is the dependent variable.
MaleDeerModel <- lm(MaleReproductiveSuccess ~ MaleBodyWeight,data=MaleRedDeerDF)
```

We can look at the output of the model and we see that our coefficient for MaleBodyWeight is 0.45. We interpret this to mean that for every 1 kg increase in male body mass, we would expect to see a 0.45 increase in the reproductive success of the male. Since we simulated population so that there would be a correlation 0.45 between our two variables we already knew that this would be the outcome.

```
# We can look at the output of the model
MaleDeerModel
```

```
##
## Call:
## lm(formula = MaleReproductiveSuccess ~ MaleBodyWeight, data = MaleRedDeerDF)
```

```
##
## Coefficients:
##    (Intercept)  MaleBodyWeight
##         -53.25            0.45
```

There are many ways that we can test whether the results of our model are reliable. A common way is through the use of p-values; a somewhat more intuitive way is through the use of model comparison using Akaike information criterion (AIC). AIC provides an estimate of how well the model fits the data.

NOTE: AIC can be used to compare two our more models created using the same dataset, but the relative AIC values are what is important, and you must use the same data for all models.

We will set up two models for this exercise. Our first model will be our 'null'model which does not include MaleBodyWeight as a predictor. Our next model will be our model of interest that does contain MaleBodyWeight as a predictor.

```
# Create a null model and a model with MaleBodyWeight as a predictor.
MaleDeerNull <- lm(MaleReproductiveSuccess ~ 1, data=MaleRedDeerDF)
MaleDeerModel <- lm(MaleReproductiveSuccess ~ MaleBodyWeight,data=MaleRedDeerDF)
```

We can now use a function created to compare models using a modified version of AIC (adjusted for small sample sizes).

```
# Compare models using AIC.
bbmle::AICctab(MaleDeerModel,MaleDeerNull, weights=T)
```

```
##               dAICc df weight
## MaleDeerModel  0.0  3  1
## MaleDeerNull  20.5  2  <0.001
```

NOTE: In this example case the model which contains MaleBodyWeight is ranked higher. We interpret this to mean that there is a reliably positive relationship between MaleBodyWeight and MaleReproductiveSuccess. In other words, as male body weight increases we see an increase in his reproductive success.

Please answer the following questions:

Question 5a. What happens when you change the correlation coefficent from 0.45 to a much smaller number and re-run the code?

Question 5b. What about when you change it to a much bigger number?

Part 4. Multivariate data

In some cases we have multiple measurements of different variables from the same individual; in this case our data would be considered 'multivariate'.

Lets create a toy dataset of red deer, sika deer and red x sika hybrid vocalizations (based on Long et al. 1998). We will simulate four vocalizations per individual and we will measure three aspects of the vocalizations- the duration, the minimum frequency and the maximum frequency of the vocalization. We will then visualize our data using principal component analysis (PCA). PCA is a commonly used data reduction technique.

For more information see https://www.nature.com/articles/nmeth.4346.pdf.

```
## Lab 1d. Multivariate data
# Just as before we load our data
data("DeerSpeciesAcousticFeatures")

# Check the structure
head(DeerSpeciesAcousticFeatures)
```

```
##   Duration MinFrequency MaxFrequency   Class
## 1       15           50          125 RedDeer
## 2       14           49          127 RedDeer
## 3       12           51          126 RedDeer
## 4       13           52          127 RedDeer
## 5       17           35          125 Hybrids
## 6       19           37          126 Hybrids
```

Now we run the PCA. Note: You can only do PCA on numeric data, so we remove the class category.

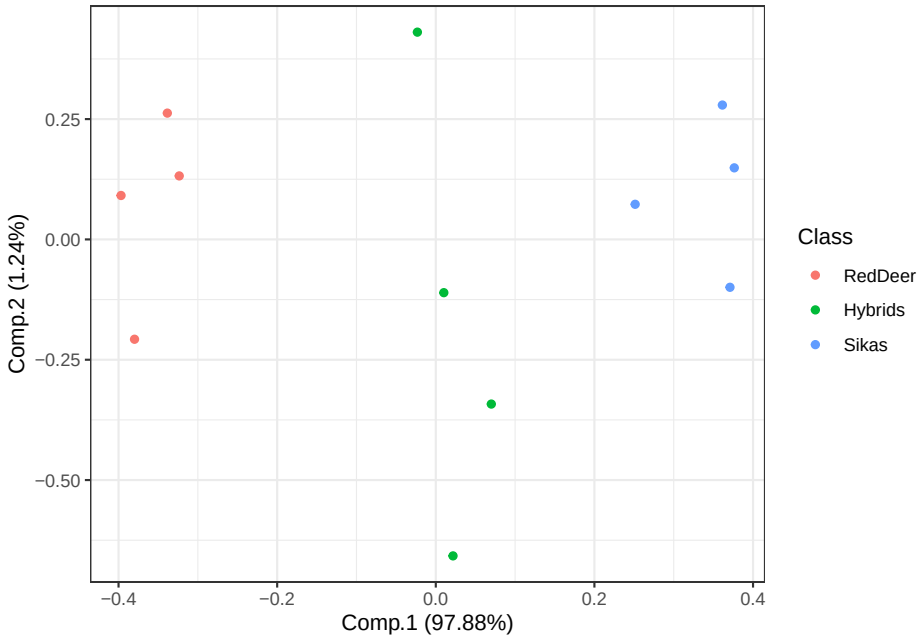
```
# Here is our modified dataset that we will use for PCA
DeerSpeciesAcousticFeatures[,-c(4)]
```

```
##    Duration MinFrequency MaxFrequency
## 1        15           50          125
## 2        14           49          127
## 3        12           51          126
## 4        13           52          127
## 5        17           35          125
## 6        19           37          126
## 7        15           34          123
## 8        16           32          127
## 9        21           20          125
## 10       22           20          126
## 11       23           21          124
## 12       20           25          127
```

```
# Run the PCA using the 'princomp' function
DeerSpeciesAcousticFeaturesPCA <- princomp(DeerSpeciesAcousticFeatures[,-c(4)])
```

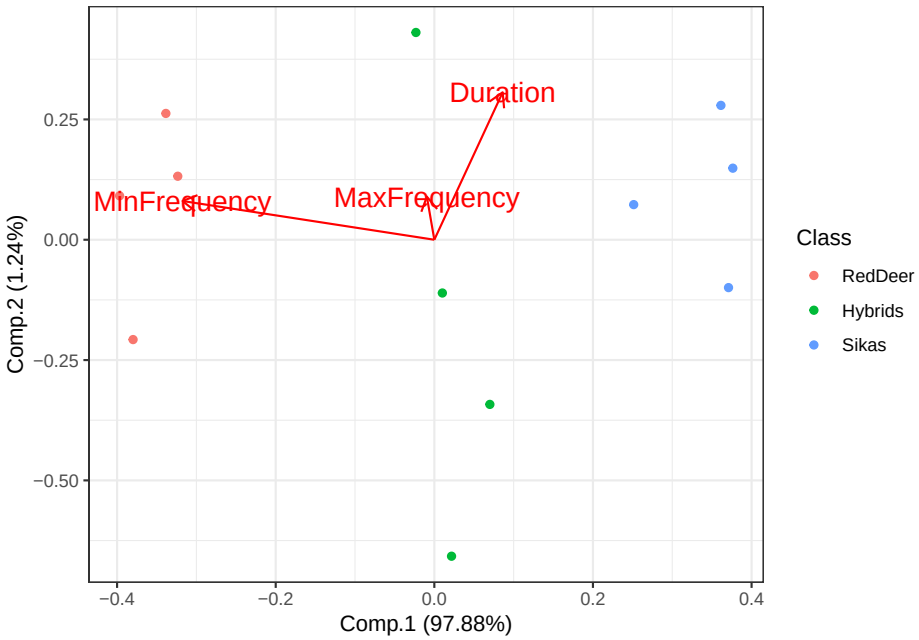
Then we will plot the results NOTE: each point represents one roar and the colors represent the class category.

```
# Plot the results of our PCA.
ggplot2::autoplot(DeerSpeciesAcousticFeaturesPCA, data = DeerSpeciesAcousticFeatures,
         loadings = FALSE)+theme_bw()
```



What if we are interested in which features best distinguish between our groups? We can visualize this using the following code:

```
# Plot the PCA with arrows indicating which features are important for distinguishing
ggplot2::autoplot(DeerSpeciesAcousticFeaturesPCA, data = DeerSpeciesAcousticFeatures,
         loadings = TRUE, loadings.colour = 'red',
         loadings.label = TRUE,
         loadings.label.size = 5)+theme_bw()
```



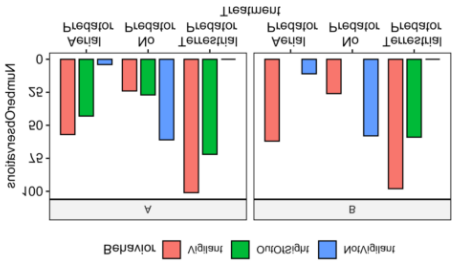
This shows that the feature that best distinguishes between red deer and sika is the minimum frequency of their vocalizations. If you look on the x-axis (Comp.1) you can see that there is a substantial amount of separation between these groups.

Please answer the following questions:

Question 6. In your own words, explain why we would want to use PCA for data visualization? What can we learn about the data when it is visualized this way?

Computer Lab 2. Activity Budgets and Ethograms

A



B

Background
In this lab you will continue to become familiar with the ways that we visualize and analyze behavioral data. For this lab I assume students have conducted Field Labs 1 and 2, so you should be familiar with how focal behavioral data are collected. Field Lab 1 data collection involves observation of animals in your own backyard and the creation of an ethogram. For Field Lab 2 data students learn how to collect focal data and scan data using meerkat videos following the methods outlined in: Hammond 2019, Vigilance behaviour in meerkats, ASAB Education.

Goals of the exercises
The main goal(s) of today's lab are to:
1) Enter and visualize your ethogram data
2) Calculate meerkat activity budgets 3) Compare inter-observer reliability from

the meerkat data

Getting started
First we need to load the relevant packages for our data analysis. Packages contain all the functions that are needed for data analysis.

```
library(behaviouR)
library(ggpubr)
```

Part 1. Enter and visualize your ethogram data

Here we will do an exploratory analysis of the ethograms that you created in Field Lab 1. First we will use some simulated (or made up) data to create our ethogram. In R we call the objects that contain our data 'dataframes'.

```
EthogramDF <-
  data.frame(Behavior=c('Resting','Locomotion','Foraging','Calling','Playing','Other')
                  TimesObserved=c(5,7,3,21,1,0))
```

If you run the object 'EthogramDF' you should see your table:

```
EthogramDF
```

```
##     Behavior TimesObserved
## 1    Resting             5
## 2 Locomotion             7
## 3   Foraging             3
## 4    Calling            21
## 5    Playing             1
## 6      Other             0
```

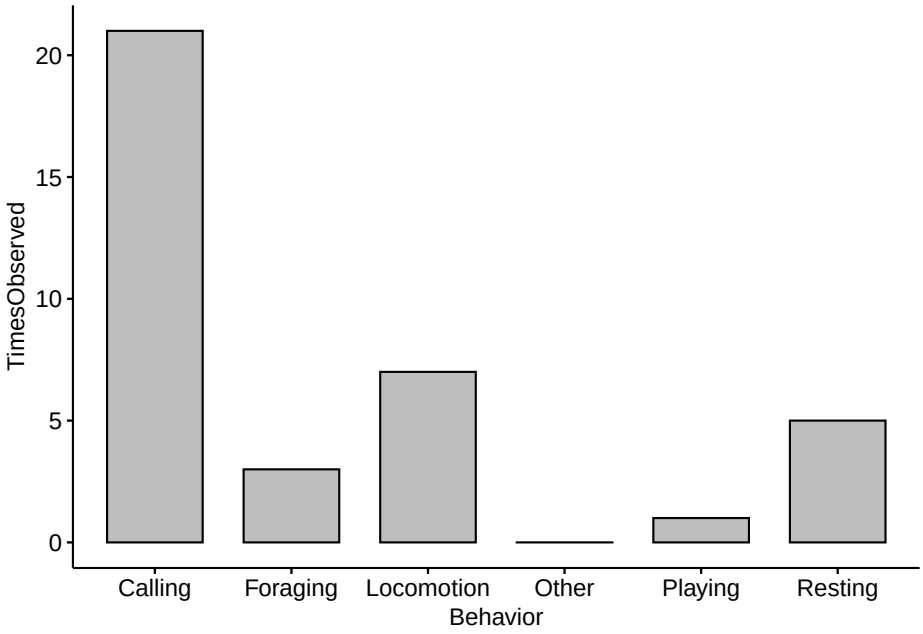
A common command to check data structure is 'str'. Here we can see our dataframe has two variables:'Behavior' and 'TimesObserved'.

```
str(EthogramDF)
```

```
## 'data.frame':    6 obs. of  2 variables:
##  $ Behavior     : Factor w/ 6 levels "Calling","Foraging",..: 6 3 2 1 5 4
##  $ TimesObserved: num  5 7 3 21 1 0
```

Now we can make a barplot to visualize our results using the 'ggpubr' package

```
ggbarplot(data=EthogramDF, x='Behavior', y='TimesObserved',fill='grey')
```



In the space below I want you to add in your own ethogram categories and create a plot based on the ethogram you created in Field lab 1.

Change the Behavior categories (i.e. Category1) to the categories you used in your ethogram and the TimesObserved values to the actual values you recorded

```
EthogramDFupdated <- data.frame(Behavior=c('Category1','Category2','Category3',
                                           'Category4','Category5','Category6'),
                                TimesObserved=c(1,2,3,4,5,6))
```

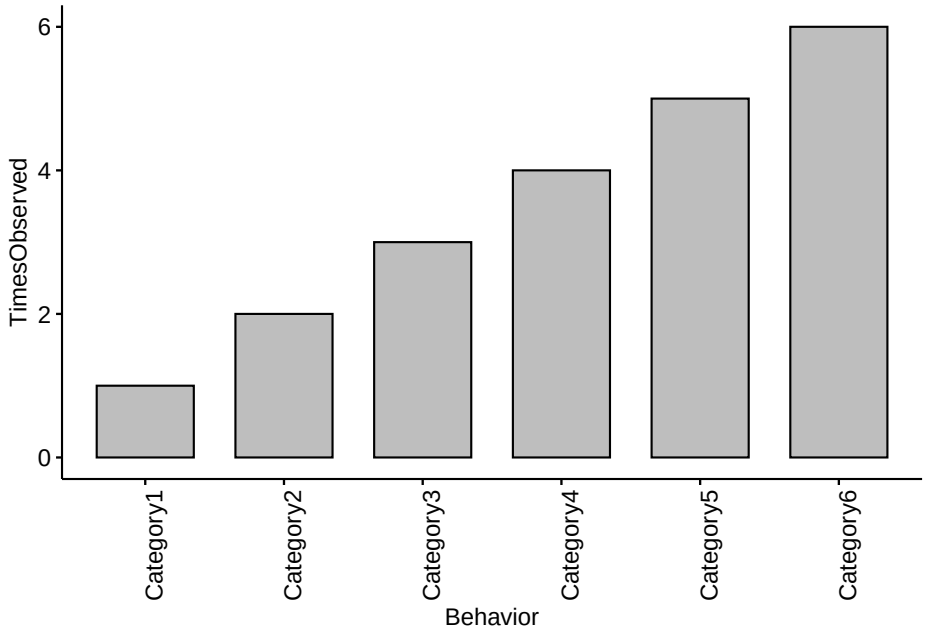
If you run the object you should see your dataframe

```
head(EthogramDFupdated)
```

```
##    Behavior TimesObserved
## 1 Category1             1
## 2 Category2             2
## 3 Category3             3
## 4 Category4             4
## 5 Category5             5
## 6 Category6             6
```

NOTE: Your plot will look different than this as you will enter your own data above!

```
ggbarplot(data=EthogramDFupdated, x='Behavior', y='TimesObserved',fill='grey',xtickslab.rt = 90)
```



Question 1
Change the code to reflect the categories you used in your ethogram and change the TimesObserved values to the actual values you recorded. NOTE: You may need to add more categories if your ethogram had more than six. What trends did you notice in terms of behaviors observed?

Part 2. Calculate meerkat activity budgets.

Our previous ethograms only included counts of different behaviors that we observed, but we did not standardize our results in any way. Activity budgets give an indication of how much time an animal spends doing each activity; these are generally expressed as percentages.

First we need to import the data. If you are using RStudio Cloud you can import your data by clicking the 'upload' button and find the datasheet saved on your computer. You can then use the following lines to read your data into R.

```
MeerkatFocalData <- read.csv('MeerkatFocalData.csv')
MeerkatScanData <- read.csv('MeerkatScanData.csv')
```

We also have sample data saved in the package so we can load that using the following code

```
data("MeerkatFocalData")
data("MeerkatScanData")
```

NOTE: If you are using your own data you will want to keep the file names the exact same or R won't be able to find the data. Let's check the structure of our data.

```
str(MeerkatFocalData)
```

```
## 'data.frame':    28 obs. of  6 variables:
##  $ StartTimeMin            : int  0 1 2 3 3 4 4 5 5 5 ...
##  $ StartTimeSec            : int  30 20 40 0 30 0 20 0 30 50 ...
##  $ BehaviorCode            : Factor w/ 5 levels "Drinking","Eating",..: 2 4 1 1 2 3 2 3 3 3 ..
##  $ TimeSeconds             : int  30 80 160 180 210 240 260 300 330 350 ...
##  $ SecondsEngagedinBehavior: int  30 50 80 20 30 30 20 40 30 20 ...
##  $ Partner                 : Factor w/ 2 levels "A","B": 1 1 1 1 1 1 1 1 1 1 ...
```

Then we calcuate the total number of seconds for each behavior. There are many different ways that we can summarize our data using R, but we will use the package 'dplyr'.

Here we take the sum for each unique behavior and return it

```
MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))
```

```
## # A tibble: 5 x 2
##   BehaviorCode TotalSeconds
##   <fct>               <int>
## 1 Drinking              200
## 2 Eating                150
## 3 Foraging              350
## 4 Locomotion            250
## 5 Sentry                240
```

Then we need to save the output as an R object

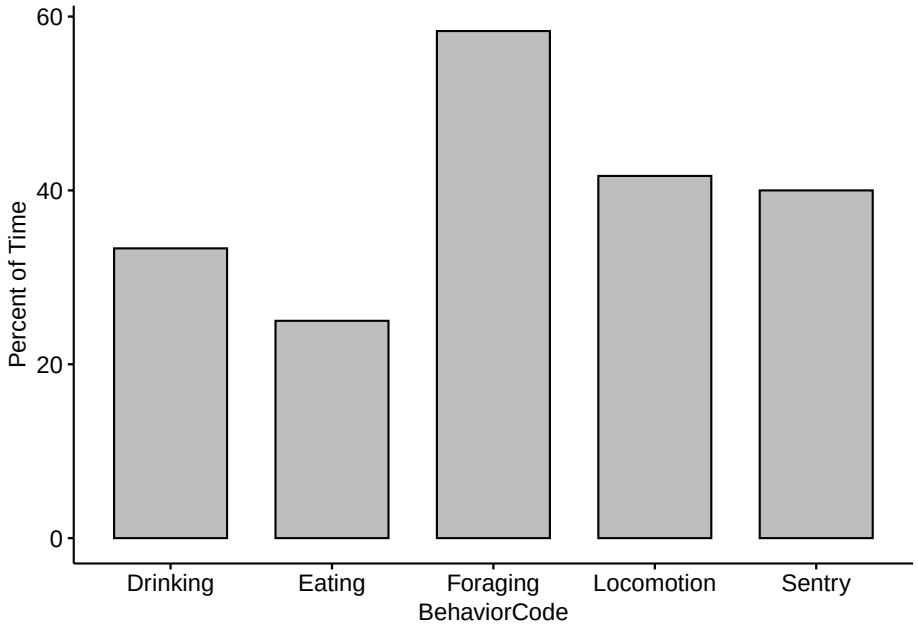
```
MeerkatFocalDataSummary <- MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))
```

Now we need to standardize for our total observation time. We divide by 600 because our video was 10 minutes (or 600 seconds) long.

```
MeerkatFocalDataSummary$ActivityBudget <- MeerkatFocalDataSummary$TotalSeconds/600*100
```

Now let's plot our results! NOTE: The example here is using fake data.

```
ggbarplot(data=MeerkatFocalDataSummary,x='BehaviorCode',y='ActivityBudget',fill='grey'
```



Now lets see if there were differences between you and your partner. We need to save the output as an R object, but this time including the sums by partner.

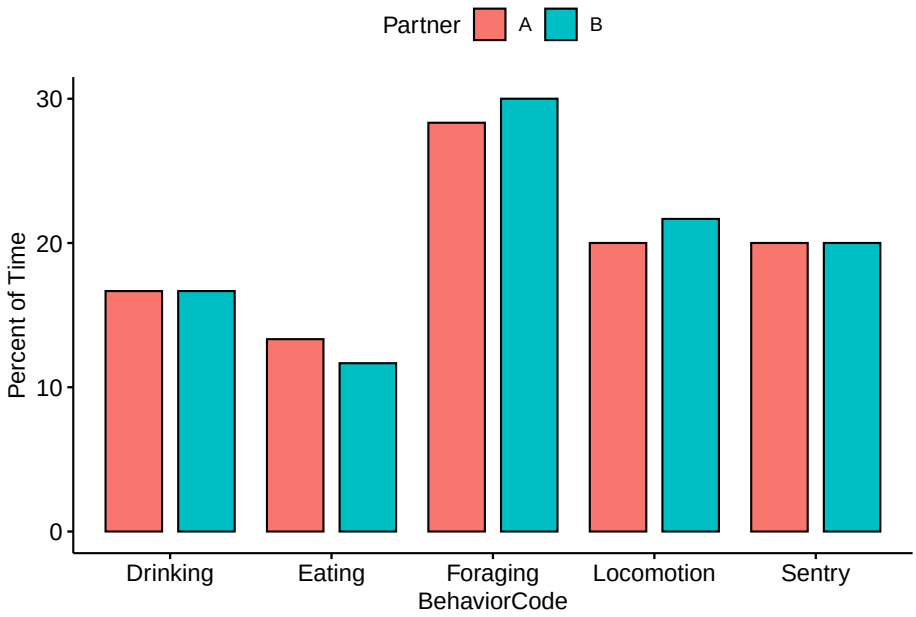
```
MeerkatFocalDataPartner <- MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode,Partner) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))
```

Again, we divide by 600 because our video was 10 minutes (or 600 seconds) long. We then multiply by 100 so that we can report in percentages.

```
MeerkatFocalDataPartner$ActivityBudget <- MeerkatFocalDataPartner$TotalSeconds/600*100
```

Now we compare our data with our partner's. Note that there are two new lines in the code below. The fill argument tells R which category or factor to use to color the bars. The position argument tells R to place the bars side-by-side.

```
ggbarplot(data=MeerkatFocalDataPartner,x='BehaviorCode',y='ActivityBudget', fill='Partn
          position = position_dodge(0.9))+ylab('Percent of Time')
```



Question 3.
Were there any noticeable differences between you and your partner's activity budgets?

Part 3. Scan sampling and inter-observer reliability.

Load your data as you did before, but this time upload the 'MeerkatScanData.csv' file.

Check the structure of our data. If we use 'head' it will return the first few rows of our dataframe.

```
head(MeerkatScanData)
```

```
##   Time Vigilant NotVigilant OutOfSight  Treatment Partner
## 1   10        1           3          2 NoPredator       A
## 2   20        1           5          0 NoPredator       A
## 3   30        2           3          1 NoPredator       A
## 4   40        2           4          0 NoPredator       A
## 5   50        1           4          1 NoPredator       A
## 6   60        2           4          1 NoPredator       A
```

For this analysis the 'Time' column isn't needed so we can remove it; It is the first column so we use 1 in the code below.

```
MeerkatScanData <-MeerkatScanData[,-c(1)]
```

R deals with blanks in dataframes by converting them to 'NA'. This is useful for some cases, but we want to convert our blanks to zeros using the following code.

```
MeerkatScanData[is.na(MeerkatScanData)] <- 0
```

Here we take the sum for each unique behavior and return it.

```
MeerkatScanData %>%
  dplyr::group_by(Treatment) %>%
  dplyr::summarise(Vigilant = sum(Vigilant),
                   OutOfSight=sum(OutOfSight),
                   NotVigilant=sum(NotVigilant))
```

```
## # A tibble: 3 x 4
##   Treatment           Vigilant OutOfSight NotVigilant
##   <fct>                  <int>      <dbl>       <int>
## 1 AerialPredator           119         73          15
## 2 NoPredator                50         64         119
## 3 TerrestrialPredator      199        131           0
```

We need to save the output as an R object

```
MeerkatScanDataSummary <- MeerkatScanData %>%
  dplyr::group_by(Treatment,Partner) %>%
  dplyr::summarise(Vigilant = sum(Vigilant),
                   OutOfSight=sum(OutOfSight),
                   NotVigilant=sum(NotVigilant))
```

There are many functions in R that allow us to change the format of our data. This one changes the format to one that is more easily used by 'ggpubr'

```
MeerkatScanDataSummaryLong <- reshape2::melt(MeerkatScanDataSummary, id.vars=c("Treatme
str(MeerkatScanDataSummaryLong)
```

```
## 'data.frame':    18 obs. of  4 variables:
##  $ Treatment: Factor w/ 3 levels "AerialPredator",..: 1 1 2 2 3 3 1 1 2 2 ...
##  $ Partner  : Factor w/ 2 levels "A","B": 1 2 1 2 1 2 1 2 1 2 ...
##  $ variable : Factor w/ 3 levels "Vigilant","OutOfSight",..: 1 1 1 1 1 1 2 2 2 2 ..
##  $ value    : num  57 62 24 26 101 98 43 30 27 37 ...
```

Now we change the column names so they are better for plotting

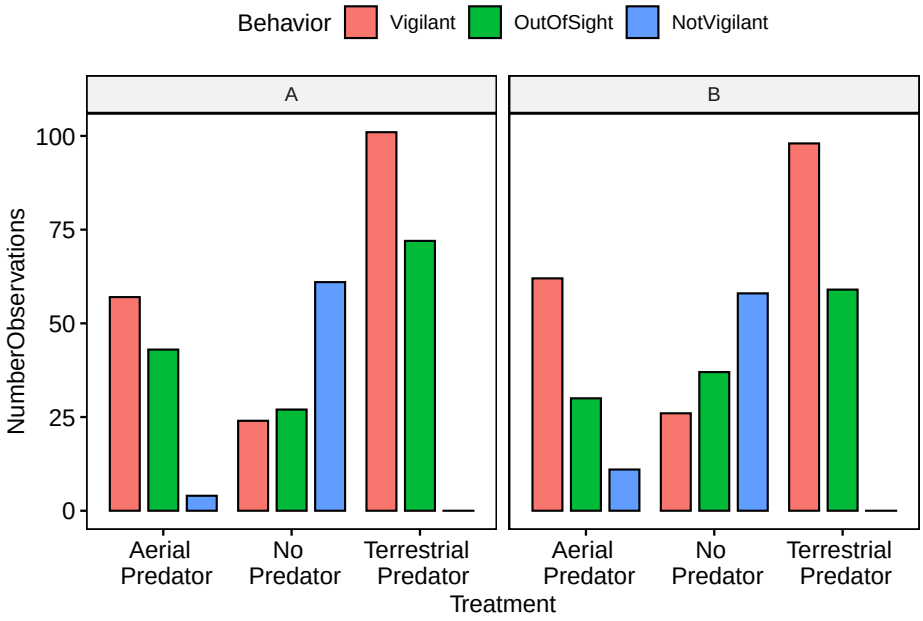
```
colnames(MeerkatScanDataSummaryLong) <- c('Treatment','Partner','Behavior','NumberObse

MeerkatScanDataSummaryLong$Treatment <- plyr::revalue(MeerkatScanDataSummaryLong$Treatm
                                                      c(AerialPredator='Aerial \n Predator',
```

```
                                                    NoPredator='No \n Predator',
                                                    TerrestrialPredator='Terrestrial \n Predator'))
```

And now we can visually compare our results with our partners. Note that in order to separate the plots by partner we use facet.by = 'Partner'.

```
ggbarplot(data=MeerkatScanDataSummaryLong,x='Treatment',y='NumberObservations', fill='Behavior',
          position = position_dodge(0.9),facet.by = 'Partner')
```



Now we want to calculate inter-observer reliablity. NOTE: that you will want to change the A and B to the names you used in the datasheet

```
Partner1Data <- subset(MeerkatScanData,Partner=='A')
Partner1Data <- Partner1Data[,-c(5)] # We remove the 'Partner' column so that we only focus on so

Partner2Data <- subset(MeerkatScanData,Partner=='B')
Partner2Data <- Partner2Data[,-c(5)] # We remove the 'Partner' column so that we only focus on so
```

Reliability is often and most simply expressed as a correlation coefficient. A correlation of 1 means a perfect positive association between two sets of measurements, whereas a correlation of zero means the complete absence of a linear association. Reliability is generally calcuated for each category of behavior.

```
# Here we calculate reliability for vigilance
VigilantCorrelation <- cor(Partner1Data$Vigilant,Partner2Data$Vigilant)

# Here we calculate reliability for vigilance
NotVigilantCorrelation <- cor(Partner1Data$NotVigilant,Partner2Data$NotVigilant)
```

```r
# Here we calculate reliability for out of sight
OutOfSightCorrelation <-  cor(Partner1Data$OutOfSight,Partner2Data$OutOfSight)

# Print the results
cbind.data.frame(VigilantCorrelation,NotVigilantCorrelation,OutOfSightCorrelation)
```

```
##   VigilantCorrelation NotVigilantCorrelation OutOfSightCorrelation
## 1           0.9027009              0.9347273             0.7840471
```

NOTE: The correlation values between you and your partner will be different!

Question 4.
What was the reliability (or correlation coefficent) between you and your partner for each of the different behavioral categories? What do you think lead to these differences?

Computer Lab 3. Analyzing Acoustic Data

```
## NULL
```

```
## [1] "processing sound event 1 out of 13"
## [1] "processing sound event 2 out of 13"
## [1] "processing sound event 3 out of 13"
## [1] "processing sound event 4 out of 13"
## [1] "processing sound event 5 out of 13"
## [1] "processing sound event 6 out of 13"
## [1] "processing sound event 7 out of 13"
## [1] "processing sound event 8 out of 13"
## [1] "processing sound event 9 out of 13"
## [1] "processing sound event 10 out of 13"
## [1] "processing sound event 11 out of 13"
## [1] "processing sound event 12 out of 13"
## [1] "processing sound event 13 out of 13"
```

Background
In this lab you will become familiar with the ways that we analyze acoustic data. You will start by analyzing focal recordings and soundscapes from Southeast Asia, and you will then upload the data you collected for Field Lab 3 and do some prelminary analysis.

Goals of the exercises
The main goal(s) of today's lab are to:
1) Create your own spectrograms
2) Learn how we use PCA to visualize differences in acoustic data
3) Upload and visualize your own data

Getting started
First we need to load the relevant packages for our data analysis. Packages contain all the functions that are needed for data analysis. First we load the required libraries

```
library(behaviouR)
library(ggfortify)
library(ggplot2)
```

We need to do some data prep to work through the examples

```
# First you can check where your working directory is; this is where R will store the
getwd()

# Now we will create a file at this location called 'FocalRecordings'
dir.create(file.path('FocalRecordings'), showWarnings = FALSE)

# Now we will load the sound files that were in the R package
githubURL <-'https://github.com/DenaJGibbon/behaviouRdata/raw/master/data/FocalRecordi
FocalRecordings <- get(load(url(githubURL)))

# Let's check the structure
head(FocalRecordings)

# Now we will save the recordings to the new folder you created
for(a in 1:length(FocalRecordings)){
  FileName <- FocalRecordings[[a]][1][[1]]
  WaveFile <- FocalRecordings[[a]][2][[1]]
  tuneR::writeWave(WaveFile, paste("FocalRecordings/",FileName,sep=''))
}
```

Part 1. Loading a sound file and making a spectrogram

First we want to check which files are in the folder There are five gibbon female recordings and five great argus recordings.

```
list.files("FocalRecordings")
```

```
##  [1] "FemaleGibbon_1.wav" "FemaleGibbon_2.wav" "FemaleGibbon_3.wav"
##  [4] "FemaleGibbon_4.wav" "FemaleGibbon.wav"   "GreatArgus_1.wav"
##  [7] "GreatArgus_2.wav"   "GreatArgus_3.wav"   "GreatArgus_4.wav"
## [10] "MaleSolo_1.wav"     "MaleSolo_2.wav"     "MaleSolo_3.wav"
## [13] "MaleSolo_4.wav"
```

Now let's read in one of the files for further investigation. We do that with the function 'readWave'. The .wav stands for 'Waveform Audio File' and is a common format that scientists use to save their sound files.

```
GibbonWaveFile <- tuneR::readWave("FocalRecordings/FemaleGibbon_1.wav")
```

Now if we run the object it will return some information

```
GibbonWaveFile
```

```
##
## Wave Object
##  Number of Samples:      572054
##  Duration (seconds):     12.97
##  Samplingrate (Hertz):   44100
##  Channels (Mono/Stereo): Mono
##  PCM (integer format):   TRUE
##  Bit (8/16/24/32/64):    16
```

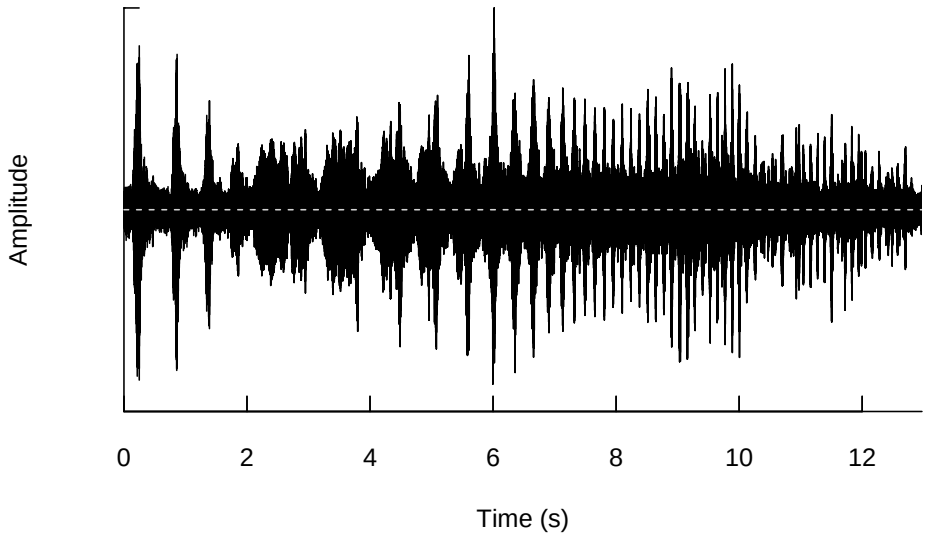
What we see is that the soundfile is ~13 seconds long, was recorded at a sampling rate of 44100 samples per second and has a total of 572054 samples. We can check if that makes sense using the following equation (duration* sampling rate).

```
seewave::duration(GibbonWaveFile)* GibbonWaveFile@samp.rate
```

```
## [1] 572054
```

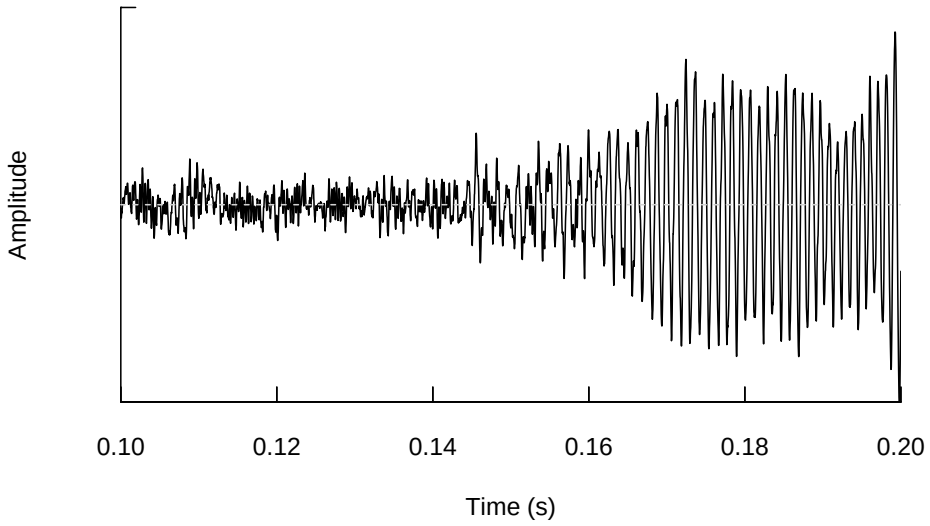
Now we can plot the waveform from our sound file using the following code:

```
seewave::oscillo(GibbonWaveFile)
```



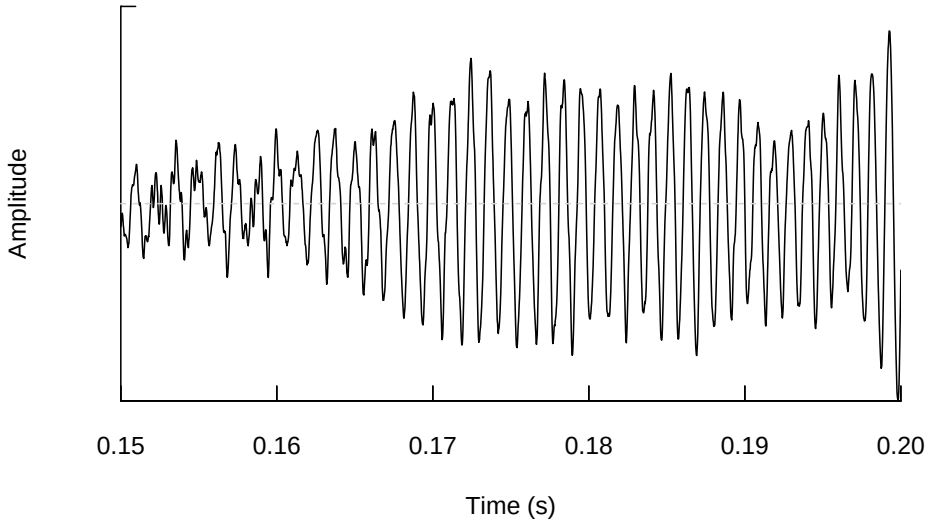
We can also zoom in so that we can actually see the shape of the wave.

```
seewave::oscillo(GibbonWaveFile, from=0.1, to=0.2)
```



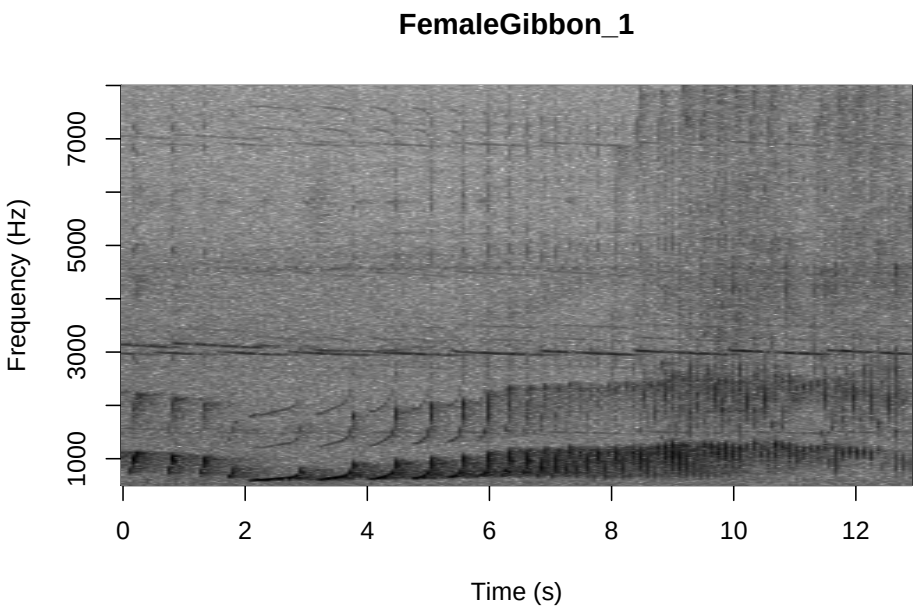
And if we zoom in again we see the wave shape even more.

```
seewave::oscillo(GibbonWaveFile, from=0.15, to=0.2)
```



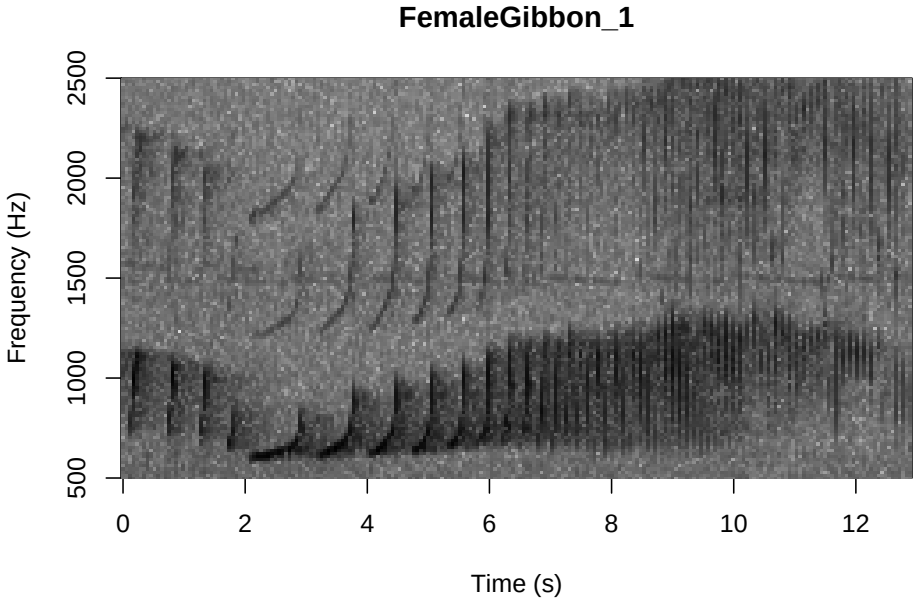
As we learned in class this week, a common way for scientists to study animal sounds is through the use of spectrograms. Here we will plot a spectrogram of a single gibbon call.

```
SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav")
```



There are many different things you can change when creating a spectrogram. In the above version of the plot there is a lot of space above the gibbon calls, so we will change the frequency range to better visualize the gibbons.

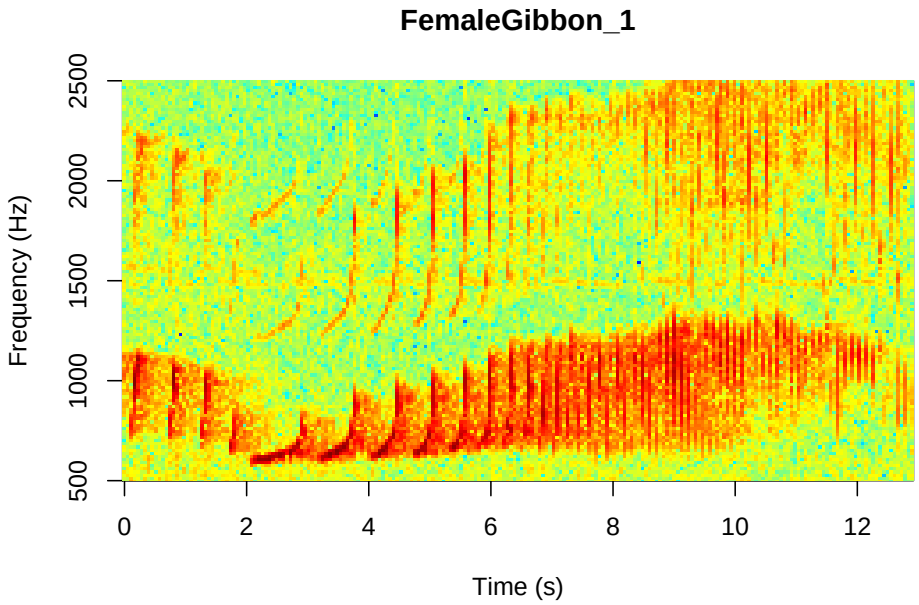
```
SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav",
                  min.freq = 500,max.freq=2500)
```



In the code we are using there is also the option to change the color of the

spectrogram adding the following code:

```
SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav",
                  min.freq = 500,max.freq=2500,
                  Colors = 'Colors')
```



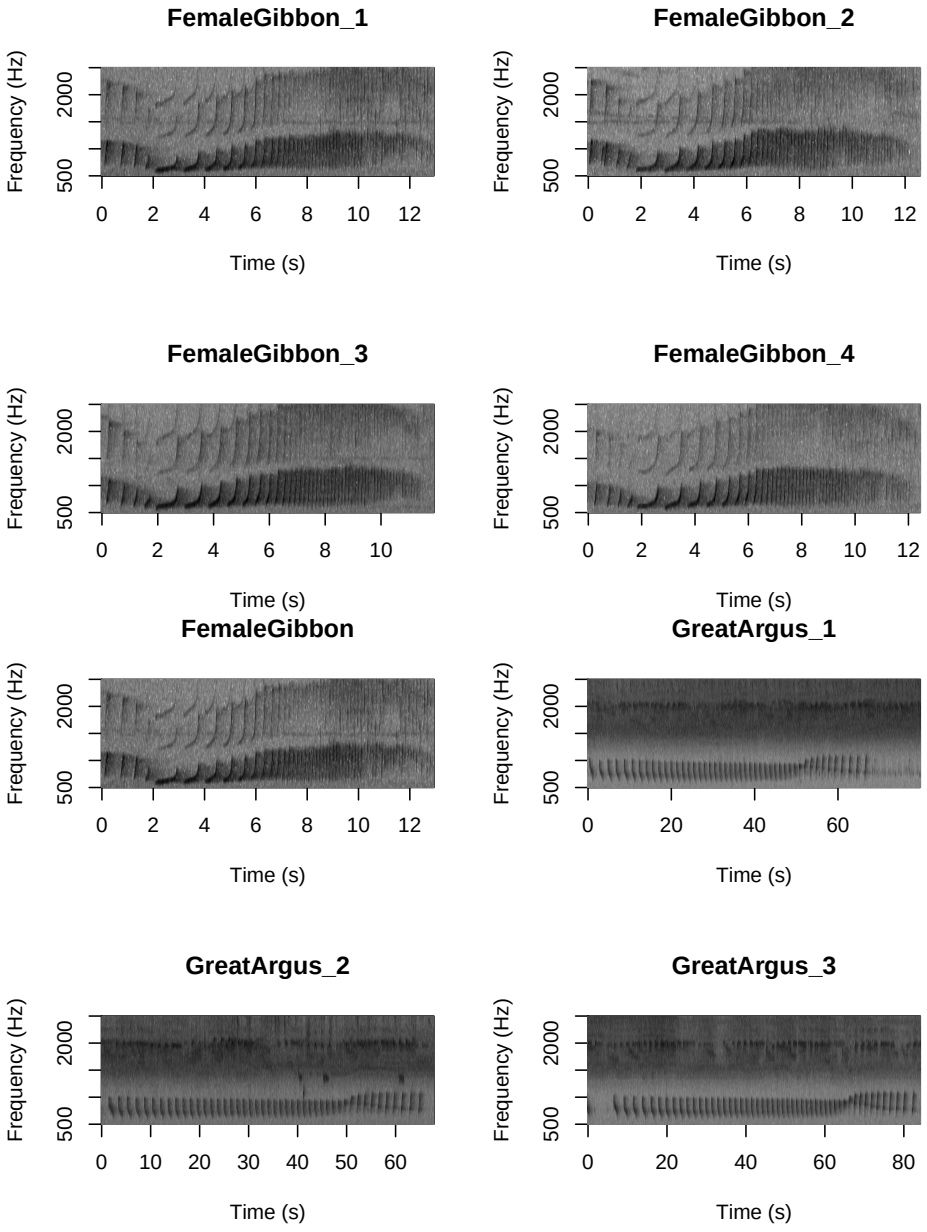
NOTE: Changing the colors doesn't change how we read the spectrograms,and different people have different preferences.

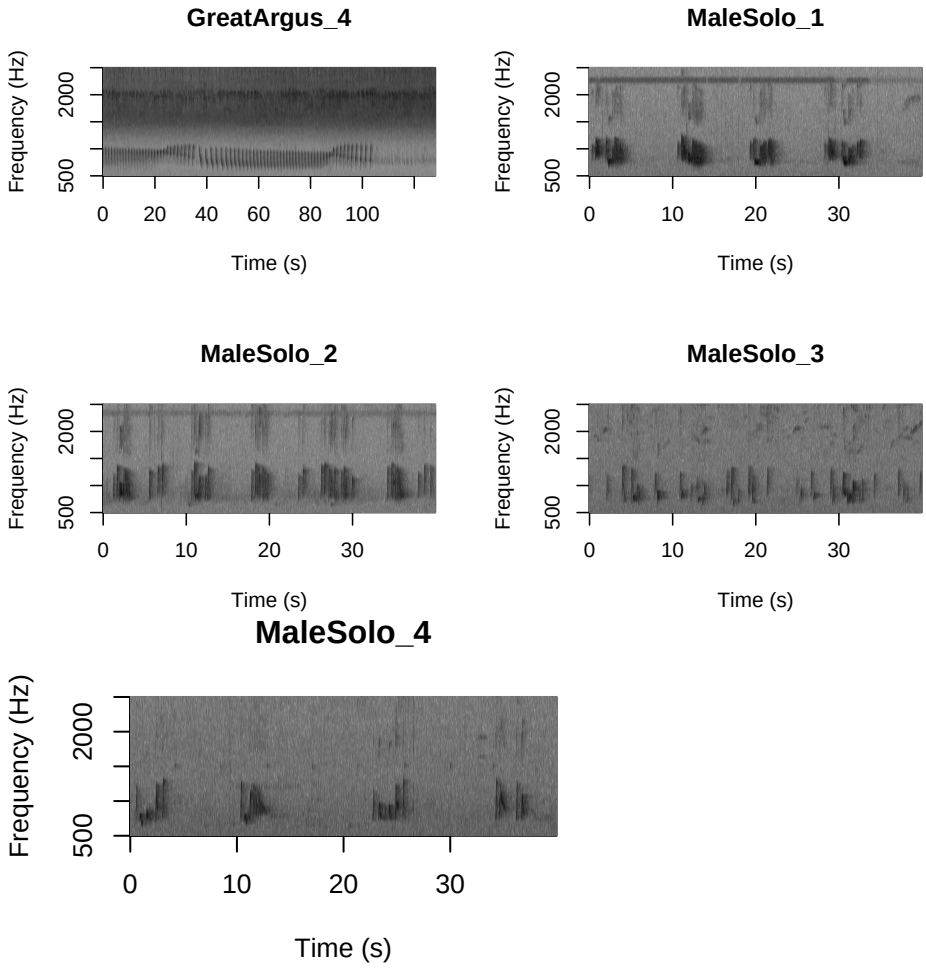
Now we can plot all of the spectrograms in our FocalRecordings file at once.

NOTE: Use the navigation buttons on the plots tab to move back and forth between the different spectrograms.

```
# We can tell R to print the spectrograms 2x2 using the code below
par(mfrow=c(2,2))

# This is the function to create the spectrograms
  SpectrogramFunction(input.dir = "FocalRecordings",min.freq = 500,max.freq=2500)
```


FemaleGibbon_1
FemaleGibbon_2
FemaleGibbon_3
FemaleGibbon_4
FemaleGibbon
GreatArgus_1
GreatArgus_2
GreatArgus_3
Frequency (Hz)
Time (s)



Question 1. What differences do you notice about gibbon and great argus spectrograms?

Now that you are doing looking at the spectrograms we will clear the space. You can always re-run the code to look at the again.

```
graphics.off()
```

Part 2. Visualizing differences in gibbon and great argus calls

A major question in many different types of behavioral studies is whether there are differences between groups. This is also the case for acoustic data, where

researchers may be interested if there are differences between sexes, populations, individuals, or vocalizations emitted in different contexts. Here we will work through a simple example where we will investigate differences in gibbon and argus calls.

The first step that we need to do is called feature extraction. There are many different ways that scientists do this, but the overarching idea is that sound data contains too much redundant information to be used on a model. Computers just don't have enough power to crunch all those numbers, so we need to identify a meaningful set of features that is smaller than using the while waveform. This is also the case for image processing.

We are going to use a feature extraction method called 'Mel-frequency cepstral coefficients'. I will not expect you to know any more about them, apart from the fact they are very useful for human speech and bioaccoustics applications.

Below is the function to extract features from our focal recordings

```
FeatureDataframe <- MFCCFunction(input.dir = "FocalRecordings")
```

In this new dataframe each row represents one gibbon or great argus call. Let's check the dimensions of this new dataframe.

```
dim(FeatureDataframe)
```

```
## [1]  13 178
```

This shows us that for each of the 10 calls there are 178 features. This is in contrast to a sound file. Let's check again to remind ourselves.

```
GibbonWaveFile
```

```
##
## Wave Object
##  Number of Samples:      572054
##  Duration (seconds):     12.97
##  Samplingrate (Hertz):   44100
##  Channels (Mono/Stereo): Mono
##  PCM (integer format):   TRUE
##  Bit (8/16/24/32/64):    16
```

This sound file is comprised of 572054 numbers. Now let's check the first row of our new dataframe which contains the new features for the same gibbon call.

```
# This isolates the first row.
FeatureDataframe[1,]
```

```
##          Class         1         2         3        4         5         6
## 1 FemaleGibbon -4.914312 -16.40678 -4.672827 20.59783 -4.632383 -6.999979
##          7         8        9       10        11        12        13        14
## 1 4.438287 -3.773641 -5.61846 10.73634 0.9757686 -1.163231 -6.800348 -17.70036
```

```
##          15       16       17       18       19        20        21        22
## 1 -2.564371 5.702963 8.604349 5.529339 2.561141 -2.507278 -8.423425 -6.229291
##          23        24        25        26      27       28       29       30
## 1 -3.115276 -7.588447 -23.27116 -3.826297 17.3082 8.632006 -1.60833 2.771364
##          31        32        33        34        35        36       37       38
## 1 -3.018685 -13.47489 -5.147027 -6.892077 -12.00958 -21.92954 9.776812 14.48623
##           39       40       41        42      43      44        45        46
## 1 0.08553639 2.275835 4.157477 -15.22412 -8.4252 8.55771 -7.545372 -20.37116
##          47       48        49       50       51        52        53       54
## 1 -9.496601 20.42913 0.1681313 2.009451 4.554415 -16.20235 -2.036649 9.715715
##          55        56       57        58       59        60       61       62
## 1 -1.521847 -4.390747 -23.6537 -1.739798 13.12778 -5.910114 6.569014 6.026896
##          63       64       65        66        67        68       69       70
## 1 -9.226613 -1.68419 4.550749 -5.170068 -10.94638 -20.08074 2.986301 9.424681
##          71       72      73        74         75       76       77        78
## 1 -6.254541 10.38071 -5.0744 -4.867023 -0.4075844 -2.65999 2.784278 -11.88797
##          79       80       81        82       83        84      85       86
## 1 -13.21431 5.279703 8.139582 -5.553872 5.027628 -3.739047 -3.1116 4.341732
##          87      88        89       90        91        92       93       94
## 1 -1.209793 1.00652 -1218.033 -1112.45 -830.5594 -462.4005 25.70725 66.06332
##          95        96       97       98       99       100      101      102
## 1 -5.735883 -36.24788 53.94479 37.60848 3.826036 -1222.493 -1046.01 -770.889
##         103      104      105     106       107       108       109       110
## 1 -393.6294 80.33639 49.16433 5.84342 -92.54861 -126.7318 -115.9133 -78.65571
##         111       112       113       114      115      116      117       118
## 1 -1222.167 -1044.567 -793.9816 -401.6897 87.24976 31.49047 -2.96363 -118.2727
##         119       120       121       122       123       124       125
## 1 -154.4816 -83.69734 -33.83878 -1229.136 -1086.172 -808.4772 -399.4776
##        126      127      128       129     130      131      132       133
## 1 53.89249 11.00474 11.99423 -57.18338 1.68353 85.31662 102.8748 -1239.725
##        134       135       136      137      138       139       140       141
## 1 -1088.64 -803.1513 -479.0124 44.22369 73.85069 -22.86258 -84.05236 -3.588271
##         142      143       144       145       146       147      148      149
## 1 -2.478177 9.528495 -1240.362 -1068.622 -782.7229 -440.9671 63.08211 80.21284
##         150       151       152    153       154       155       156       157
## 1 -46.79594 -82.71886 -40.87362 -75.92 -49.31079 -1285.428 -1091.577 -842.4694
##         158      159     160       161       162      163      164     165
## 1 -460.0004 72.63099 32.0981 -40.61593 -20.85975 31.29035 11.68475 66.2166
##         166       167       168       169     170      171       172       173
## 1 -1225.674 -1065.805 -823.5072 -452.4384 74.0775 24.51663 -33.48887 -13.24513
##        174      175      176      177
## 1 25.57118 3.053292 26.88251 12.97175
```

```
# This tells us how many features we have.
ncol(FeatureDataframe[1,])
```

```
## [1] 178
```

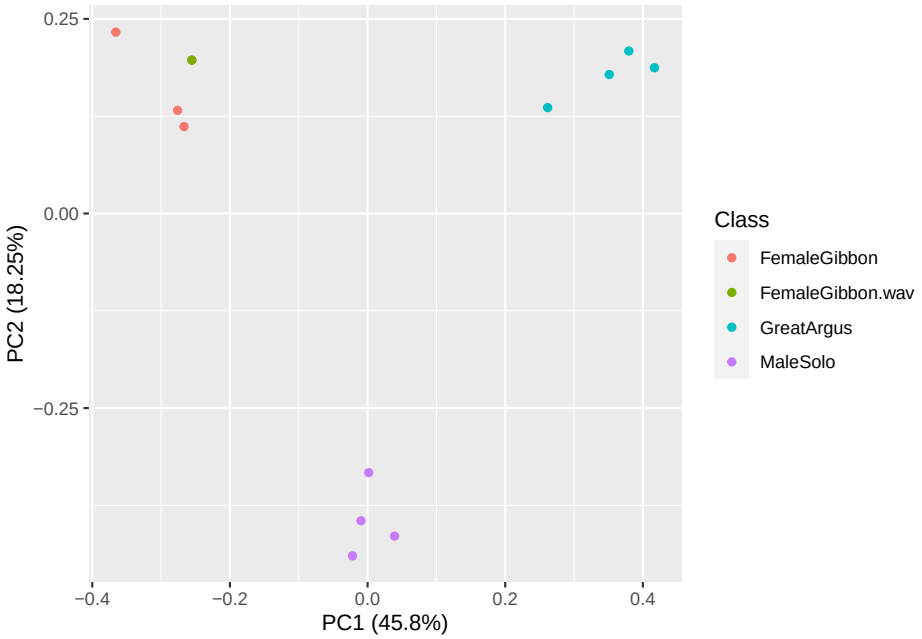
OK, now we are ready to do some plotting. If you remember from our earlier lab, the data structure we have here is multivariate, as each call has multiple values associated with it. So we will use the same approach (principal component analysis) to visualize our gibbon and great argus calls.

First we run the principal component analysis

```
pca_res <- prcomp(FeatureDataframe[,-c(1)], scale. = TRUE)
```

Now we visualize our results

```
ggplot2::autoplot(pca_res, data = FeatureDataframe,
         colour = 'Class')
```



What we see in this plot is strong clustering, with the points from each class of calls more close to the other points than those of different calls.

Question 2. How do we interpret the clustering in this PCA plot?

Part 3. Soundscapes

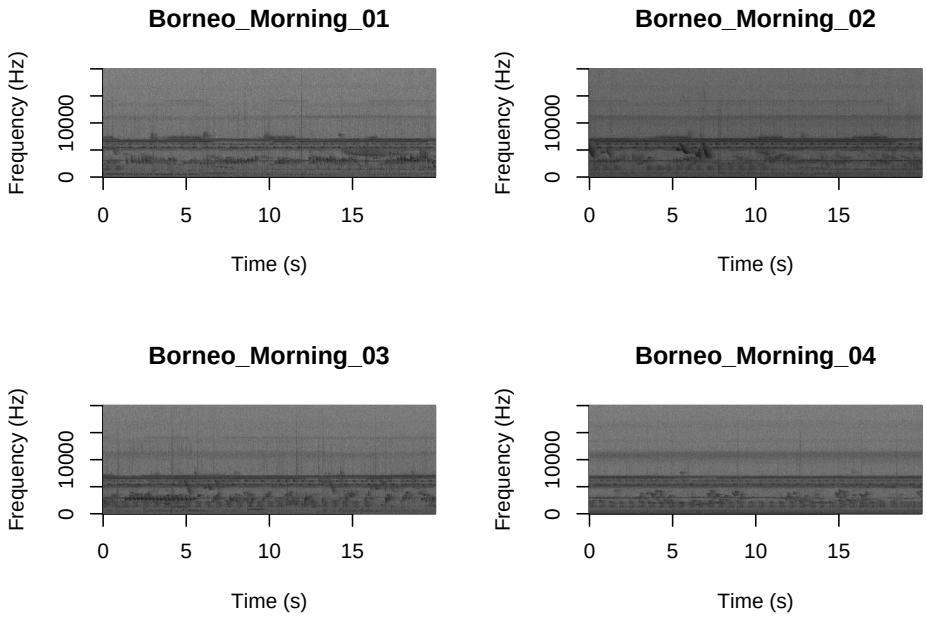
Now we will do something similar to our investigation of the focal recordings, but using a soundscape example. Here we have 20-sec recordings taken from two different locations (Borneo and Sulawesi) and two different times (06:00 and 18:00). Let's check what is in the folder.

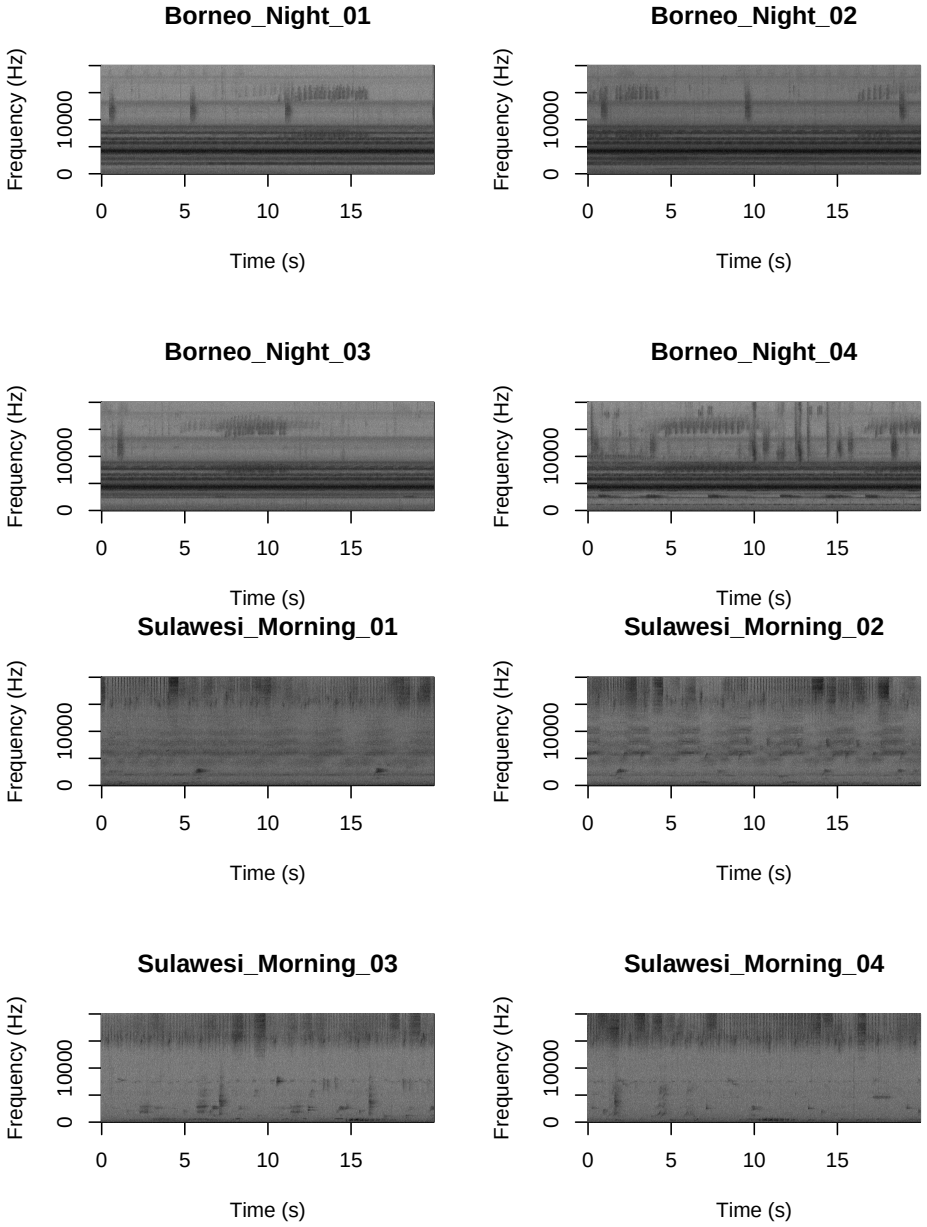
```
list.files("SoundscapeRecordings")
```

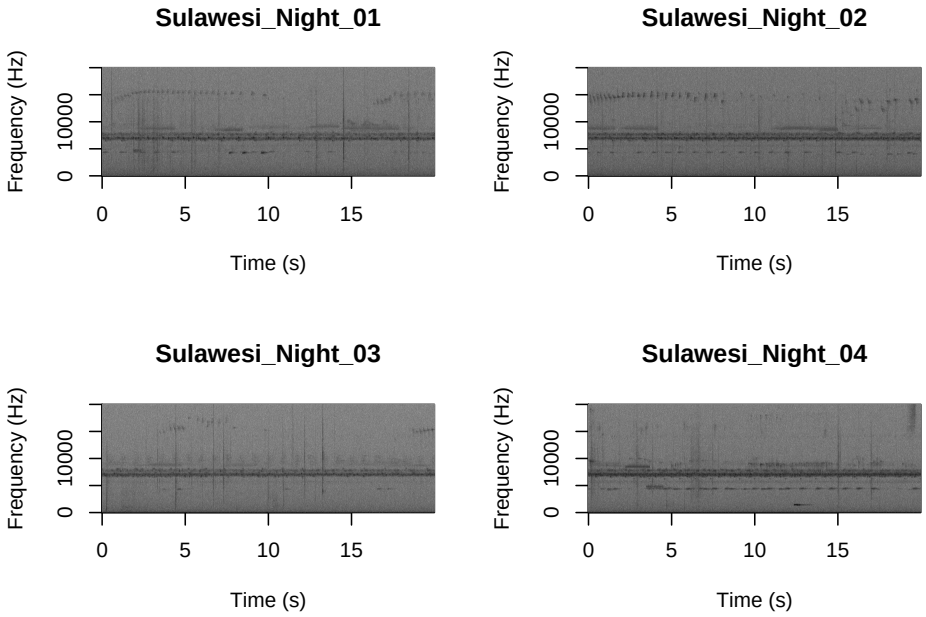
```
##  [1] "Borneo_Morning_01.wav"   "Borneo_Morning_02.wav"
##  [3] "Borneo_Morning_03.wav"   "Borneo_Morning_04.wav"
##  [5] "Borneo_Night_01.wav"     "Borneo_Night_02.wav"
##  [7] "Borneo_Night_03.wav"     "Borneo_Night_04.wav"
##  [9] "Sulawesi_Morning_01.wav" "Sulawesi_Morning_02.wav"
## [11] "Sulawesi_Morning_03.wav" "Sulawesi_Morning_04.wav"
## [13] "Sulawesi_Night_01.wav"   "Sulawesi_Night_02.wav"
## [15] "Sulawesi_Night_03.wav"   "Sulawesi_Night_04.wav"
```

Now we will create spectrograms for each recording

```
par(mfrow=c(2,2))
SpectrogramFunctionSite(input.dir = "SoundscapeRecordings",
                        min.freq = 0,max.freq=20000)
```




Borneo_Night_01
Borneo_Night_02
Borneo_Night_03
Borneo_Night_04
Sulawesi_Morning_01
Sulawesi_Morning_02
Sulawesi_Morning_03
Sulawesi_Morning_04
Frequency (Hz)
0
10000
0
5
10
15
Time (s)



Now that you are doing looking at the figure we will clear the space. You can always re-run the code to look at the again.

```
graphics.off()
```

Now just as above we will do feature extraction of our soundscape recordings. Thiswill convert our dataset into a smaller, much more manageable size.

```
SoundscapeFeatureDataframe <-
  MFCCFunctionSite(input.dir = "SoundscapeRecordings")
```

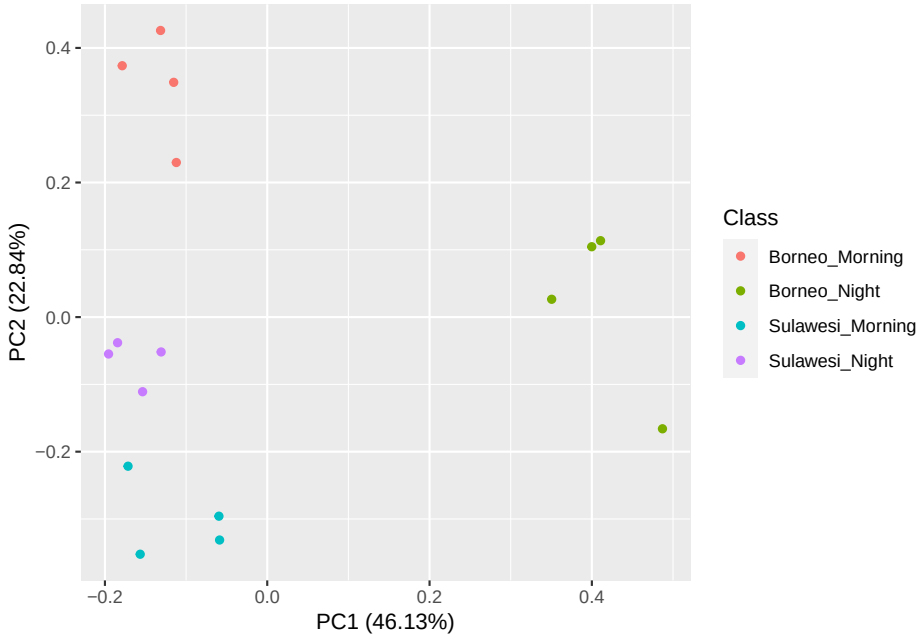
Check the resulting structure of the dataframe to make sure it looks OK.

```
dim(SoundscapeFeatureDataframe)
```

```
## [1]  16 177
```

Now we visualize our results

```
pca_res <- prcomp(SoundscapeFeatureDataframe[,-c(1)], scale. = TRUE)
ggplot2::autoplot(pca_res, data = SoundscapeFeatureDataframe,
         colour = 'Class')
```



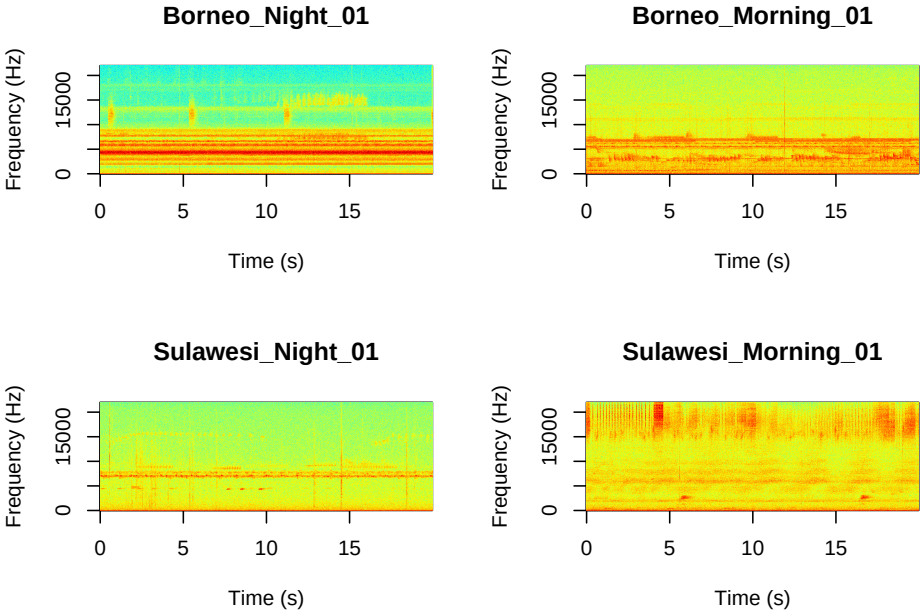
It looks like the Borneo_Night sampling period was much different from the rest! Let's look at a single spectrogram from each sampling location and time.

```
par(mfrow=c(2,2))
SpectrogramSingle(sound.file ="SoundscapeRecordings/Borneo_Night_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'Colors',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Borneo_Morning_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'Colors',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Sulawesi_Night_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'Colors',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Sulawesi_Morning_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'Colors',downsample = FALSE)
```



Question 3. You can listen to example sound files on the Canvas page. Between looking at the spectrograms and listening to the sound files what do you think are the main differences between the different locations and times?

Part 4. Now it is time to analyze the data you collected for this week's field lab.

Upload your sound files into either the 'MyFocalRecordings' folder or the 'MySoundscapeRecordings'folder and run the code below.

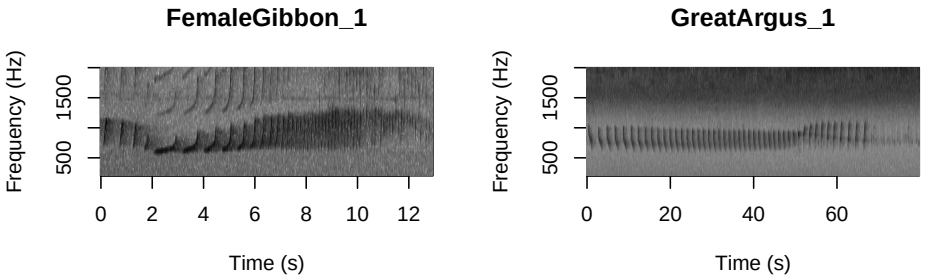
NOTE: Make sure to delete the existing files before you add your own!

0.0.1 Part 4a. Your focal recordings

First lets visualize the spectrograms. You may want to change the frequency settings depending on the frequency range of your focal animals.

```
# We can tell R to print the spectrograms 2x2 using the code below
par(mfrow=c(2,2))

# This is the function to create the spectrograms
SpectrogramFunction(input.dir = "MyFocalRecordings",min.freq = 200,max.freq=2000)
```



Here is the function to extract the features. Again, based on the frequency range of your focal recordings you may want to change the frequency settings.

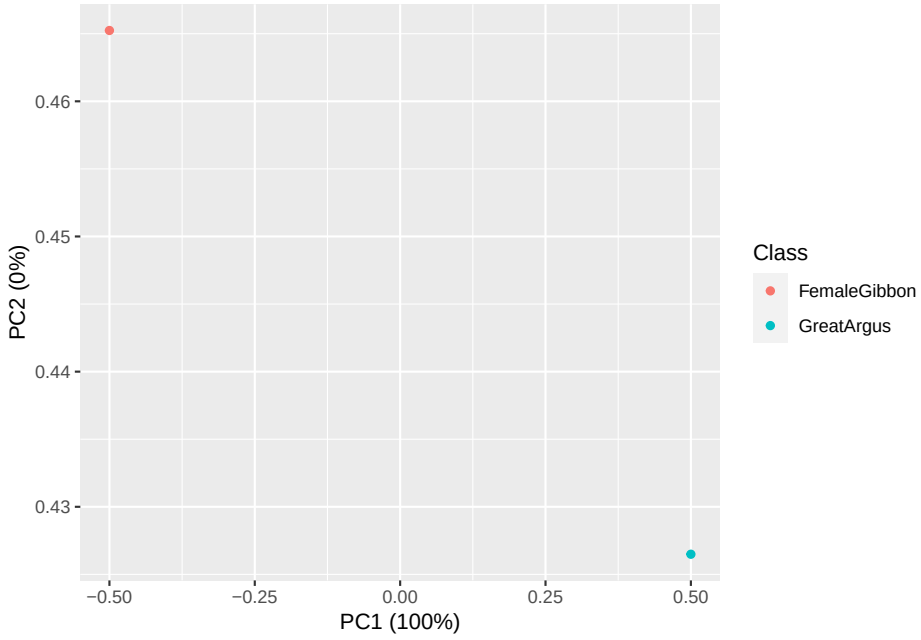
```
MyFeatureDataFrame <- MFCCFunction(input.dir = "MyFocalRecordings",min.freq = 200,max.freq=2000)
```

Then we run the principal component analysis

```
pca_res <- prcomp(MyFeatureDataFrame[,-c(1)], scale. = TRUE)
```

Now we visualize our results

```
ggplot2::autoplot(pca_res, data = MyFeatureDataFrame,
         colour = 'Class')
```

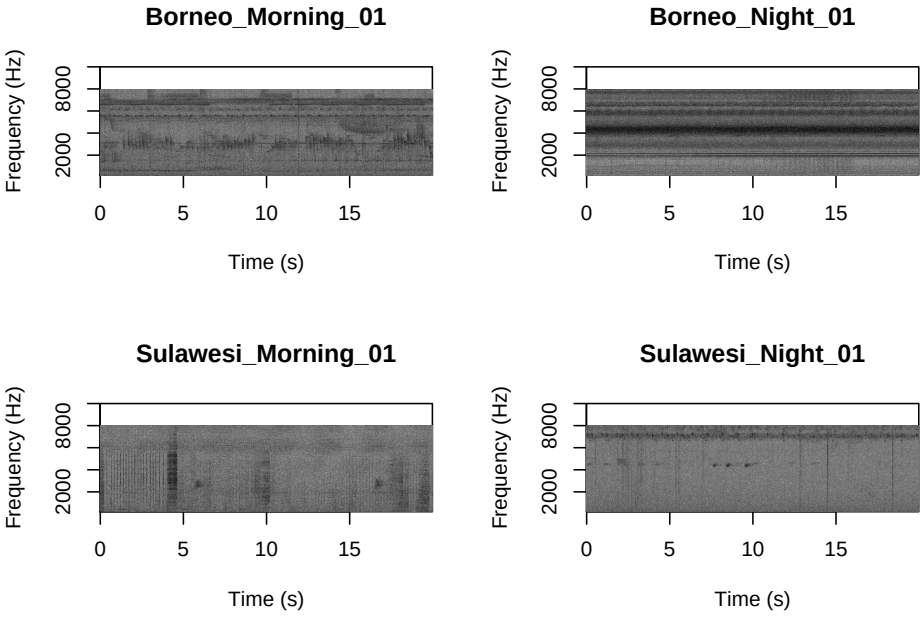


0.0.2 Part 4b. Soundscape recordings

First lets visualize the spectrograms. You probably don't want to change the frequency settings for this.

```
# We can tell R to print the spectrograms 2x2 using the code below
par(mfrow=c(2,2))

# This is the function to create the spectrograms
SpectrogramFunction(input.dir = "MySoundscapeRecordings",min.freq = 200,max.freq=10000)
```



Then we extract the features, which in this case are the MFCCs.

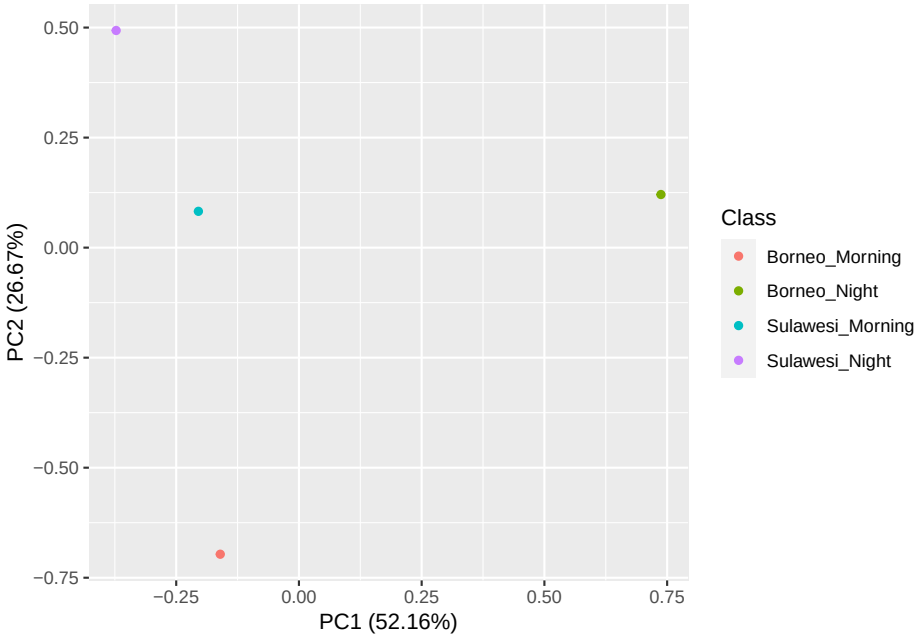
```
MySoundscapeFeatureDataframe <-
  MFCCFunctionSite(input.dir = "MySoundscapeRecordings",min.freq = 200,max.freq=10000)
```

Check the resulting structure of the dataframe to make sure it looks OK.

```
dim(MySoundscapeFeatureDataframe)
```

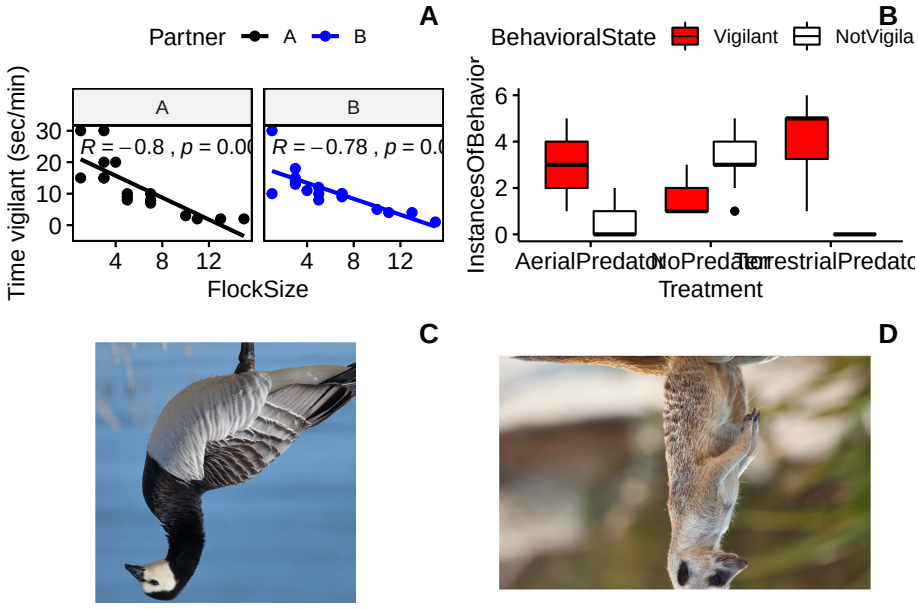
Now we visualize our results

```
pca_res <- prcomp(MySoundscapeFeatureDataframe[,-c(1)], scale. = TRUE)
ggplot2::autoplot(pca_res, data = MySoundscapeFeatureDataframe,
         colour = 'Class')
```



Question 4. Do you see evidence of clustering in either your focal recordings or soundscape recordings?

Computer Lab 4. Vigilance behavior



Background
In this lab we will continue our work investigating vigilance behaviors in geese and meerkats.

Goals of the exercises
The main goal(s) of today's lab are to:
1) Use the data on goose vigilance behavior to investigate the relationship between group size and vigilance behaviors.
2) Analyze the data collected during the meerkat lab to test for differences in vigilant behavior across predator treatment groups.
3) Continue to become familar with the way scientists analyze and interpret

data.

Getting started
First we need to load the relevant packages for our data analysis. Packages contain all the functions that are needed for data analysis. First we load the required libraries

```
library(behaviouR)
library(ggpubr)
```

Part 1: Barnacle goose vigilance

First, load the goose data that you collected using the following code.

```
BarnacleGooseData <- read.csv('BarnacleGooseData.csv')
```

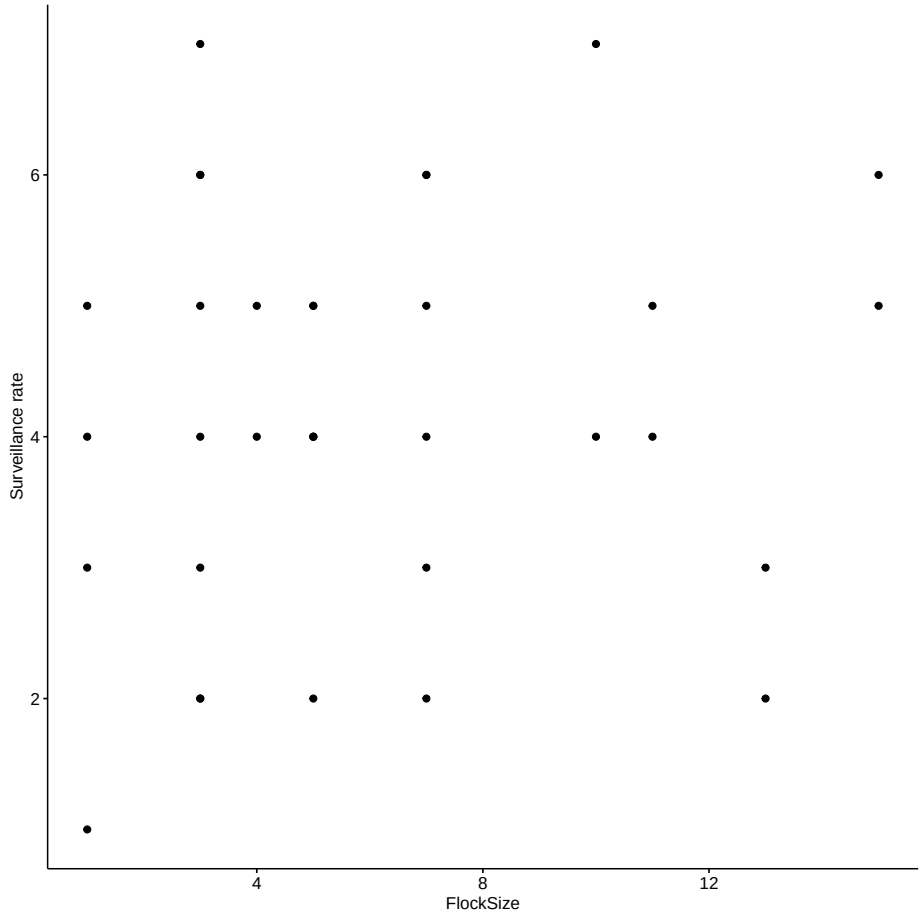
Or you can use the data that comes with the package.

```
data("BarnacleGooseData")
```

Part 1a: Surveillance behavior

First, we will calculate the surveillance rate (or the number of heads up per minute). Let's start by looking at the total number of 'head up' behaviors in our data. To investigate this relationship we will create a scatterplot.

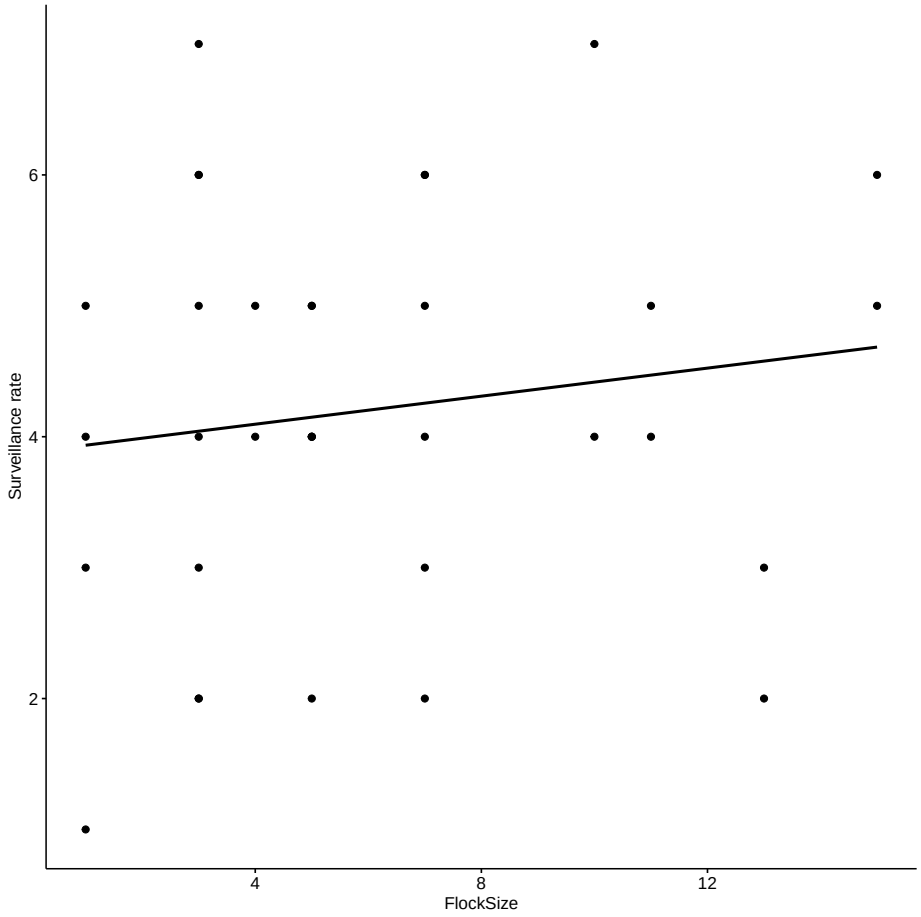
```
# Scatterplot of total number of 'head up' in our data
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp')+ylab('Surveillance rate')
```



Now let's add a trend line to see if there is a relationship between flock size and the total number of 'head up' behaviors

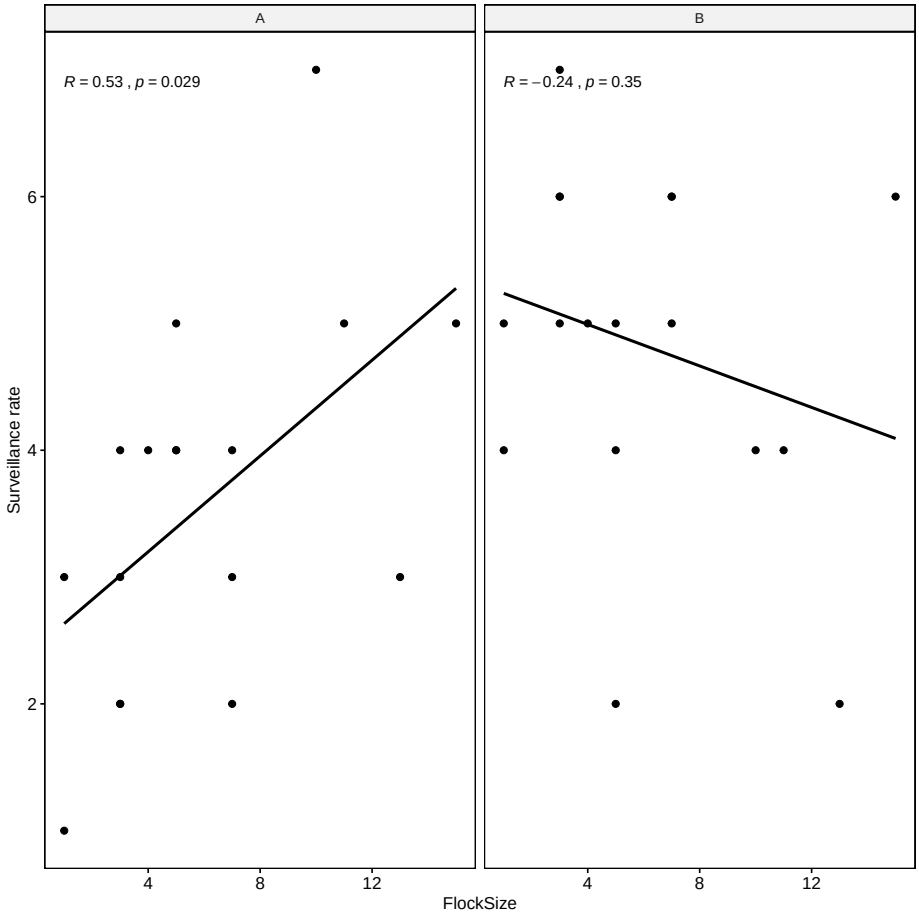
```
# Scatterplot of total number of 'head up' in our data with a trend line.

ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line')+ylab('Surveillance rate')
```



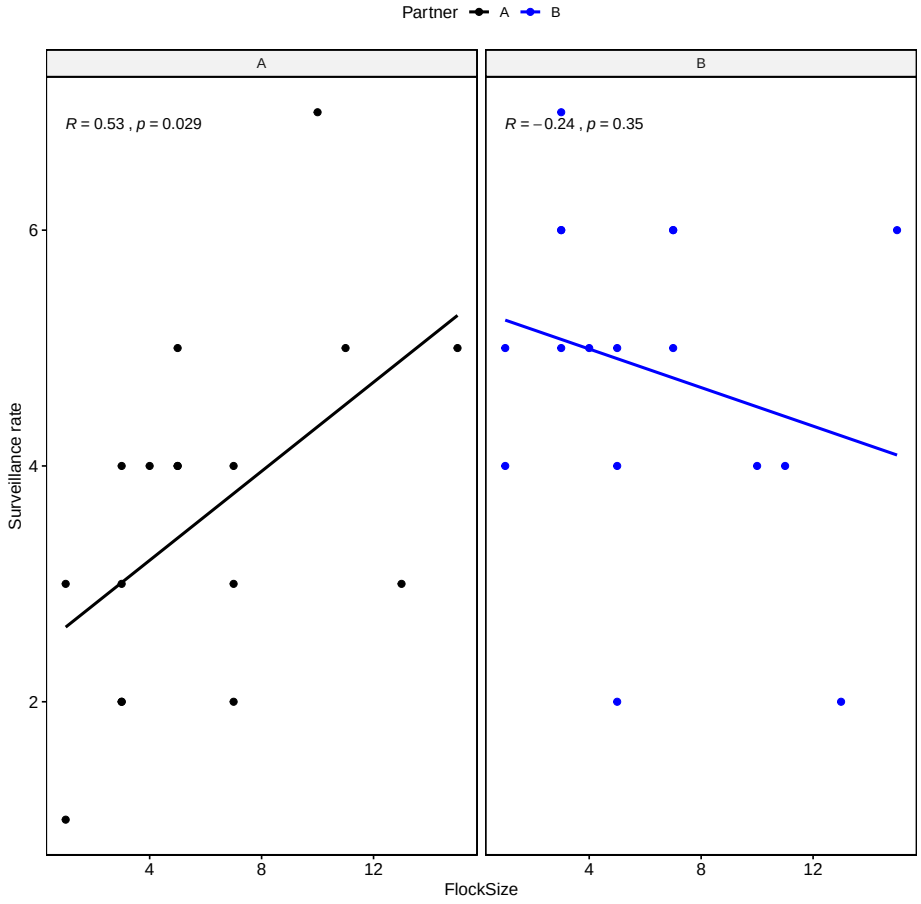
Let's see if there were any differences between you and your partner. We will also add the command 'cor.coef = T' which will give us the correlation coefficient (R) along with an associated p-value.

```
# NOTE: The data here are simulated so your plots should look different
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line', facet.by = 'Partner',
          cor.coef = T)+ylab('Surveillance rate')
```



Let's plot you and your partner's data in different colors.

```
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line', facet.by = 'Partner',
          color = 'Partner', palette =c('black','blue'),
          cor.coef = T)+ylab('Surveillance rate')
```



Question 1: Were there any major differences between you and your partner in terms of the observed relationship between flock size and surveillance rate?

Now we will do model selection using Akaike information criterion (AIC). First we create a null model and then we create a model with flock size as a predictor of total number of heads up. In the model code we specify that we are using a Poisson distribution, as we are dealing with count data instead of a continuous variable.

```
# This is our null model
SurveillanceNullModel <- glm(TotalHeadsUp ~ (1/Partner),family=poisson , data=Barnacle

# This is our model with flock size as a predictor of total number of heads up
SurveillanceModel <- glm(TotalHeadsUp ~ FlockSize + (1/Partner) ,family=poisson, data=

# Then we compare the models using AIC
bbmle::AICtab(SurveillanceNullModel,SurveillanceModel)
```

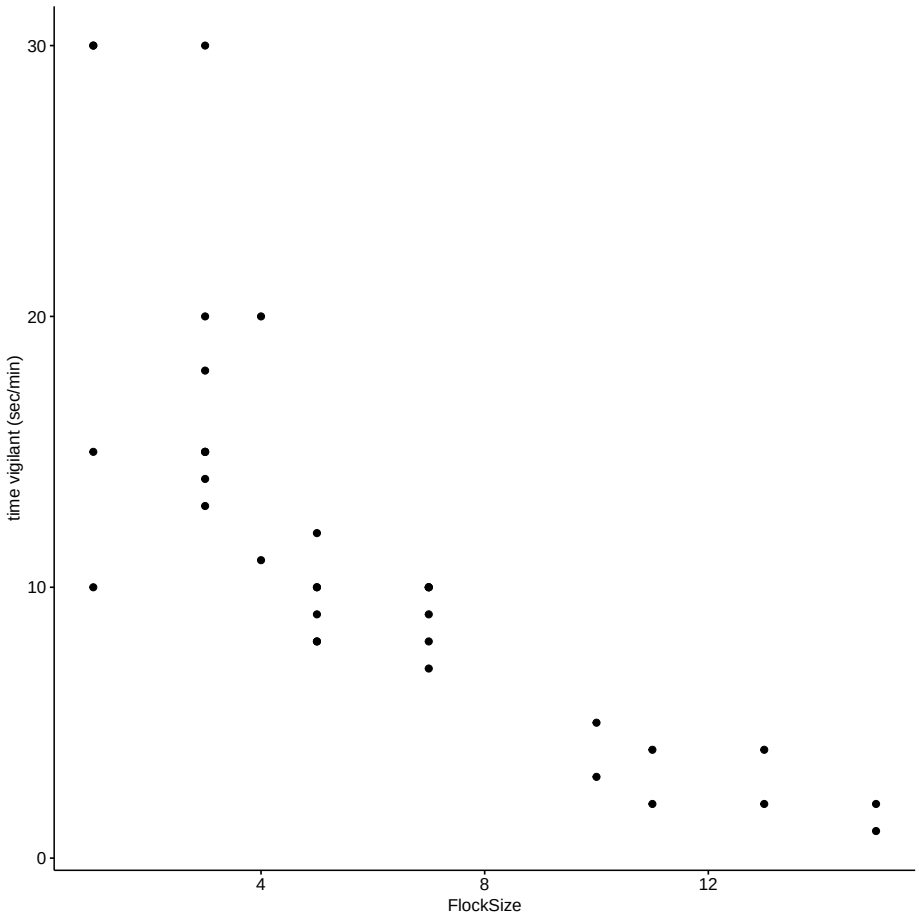
```
##                         dAIC df
## SurveillanceNullModel 0.0  1
## SurveillanceModel     1.6  2
```

Question 2: If the null model is ranked higher than the model with flock size as a predictor, how do would we interpret this finding? What if the model with flock size as a predictor was ranked higher? What were your results?

Part 1b: Time vigilant (sec/min)

Now we will look at the relationship between the duration (calculated as seconds per minute) that the geese were vigilant as a function of flock size.

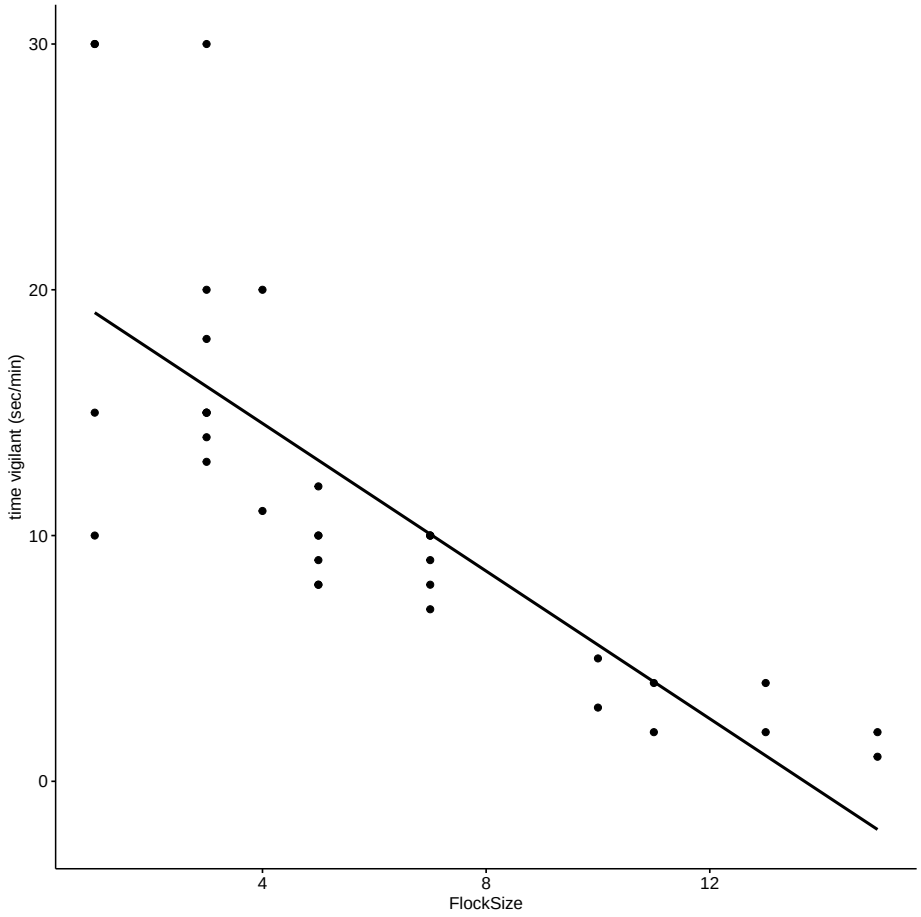
```
# Scatterplot of time vigilant (sec/min) as a function of group size
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TimeSecHeadUp')+ylab('time vigilant (sec/min)')
```



Now let's add a trend line to see if there is a relationship between flock size and the duration of vigilance.

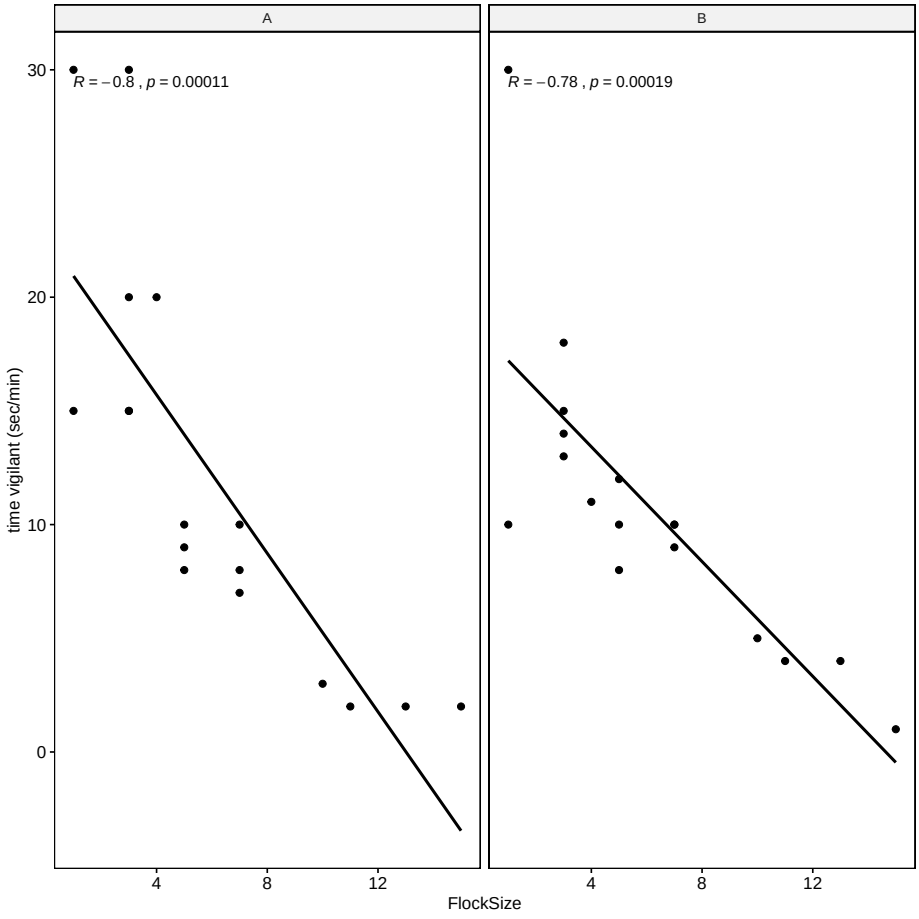
```
# Scatterplot of duration of vigilance behavior in our data with a trend line.

ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line')+
  ylab('time vigilant (sec/min)')
```



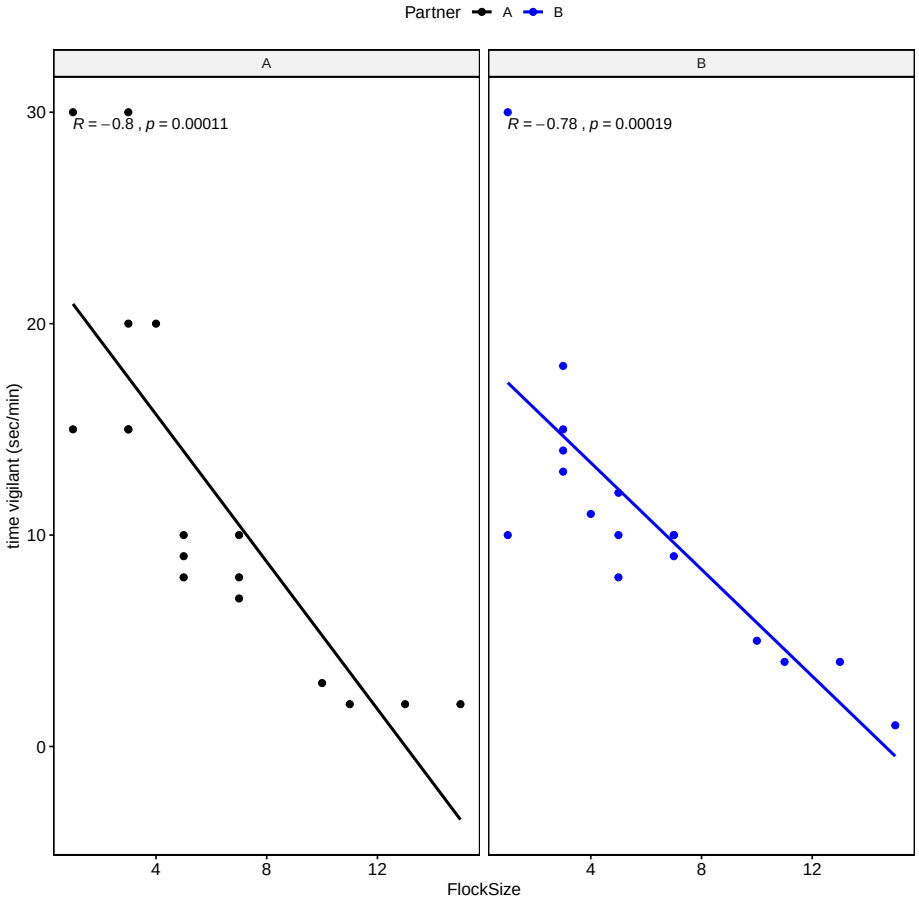
Let's see if there were any differences between you and your partner. We will also add the command 'cor.coef = T' which will give us the correlation coefficient (R) along with an associated p-value.

```
# NOTE: The data here are simulated so your plots should look different
ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line',
          facet.by = 'Partner',cor.coef = T)+
  ylab('time vigilant (sec/min)')
```



Let's plot you and your partner's data in different colors.

```
ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line',
          facet.by = 'Partner', color='Partner', cor.coef = T,
          palette =c('black','blue'))+
            ylab('time vigilant (sec/min)')
```



As before we will create a null model and then a model with flock size as a predictor and compare them using AIC. We will not use a Poisson distribution here because our outcome variable is continuous.

```
# This is our null model
VigilanceNullModel <- lme4::lmer(TimeSecHeadUp ~ (1|Partner), data=BarnacleGooseData)

# This is our model with flock size as a predictor duration of vigilance
VigilanceModel <- lme4::lmer(TimeSecHeadUp ~ FlockSize + (1|Partner) ,data=BarnacleGoos

# Now we compare the models using AIC
bbmle::AICtab(VigilanceNullModel,VigilanceModel)
```

```
##                    dAIC df
## VigilanceModel      0.0 4
## VigilanceNullModel 28.1 3
```

#**Question 3.** How do you interpret the results of your model selection? Was

there a relationship between flock size and duration of vigilance behavior?

Part 2: Meerkat data revisted

Please upload your meerkat scan data to this project and delete the existing datasheet.

```
MeerkatScanData <- read.csv('MeerkatScanData.csv')
```

As before we will turn our NA values to zero

```
MeerkatScanData[is.na(MeerkatScanData)] <- '0'
```

We will remove the time and out of sight columns as we do not need them

```
MeerkatScanData <- dplyr::select(MeerkatScanData,-c(Time,OutOfSight))
```

We need to reformat our data so that we can plot it

```
MeerkatScanDataSummaryLong <- reshape2::melt(MeerkatScanData, id.vars=c("Treatment", "Partner"))
```

Here we add more informative column names

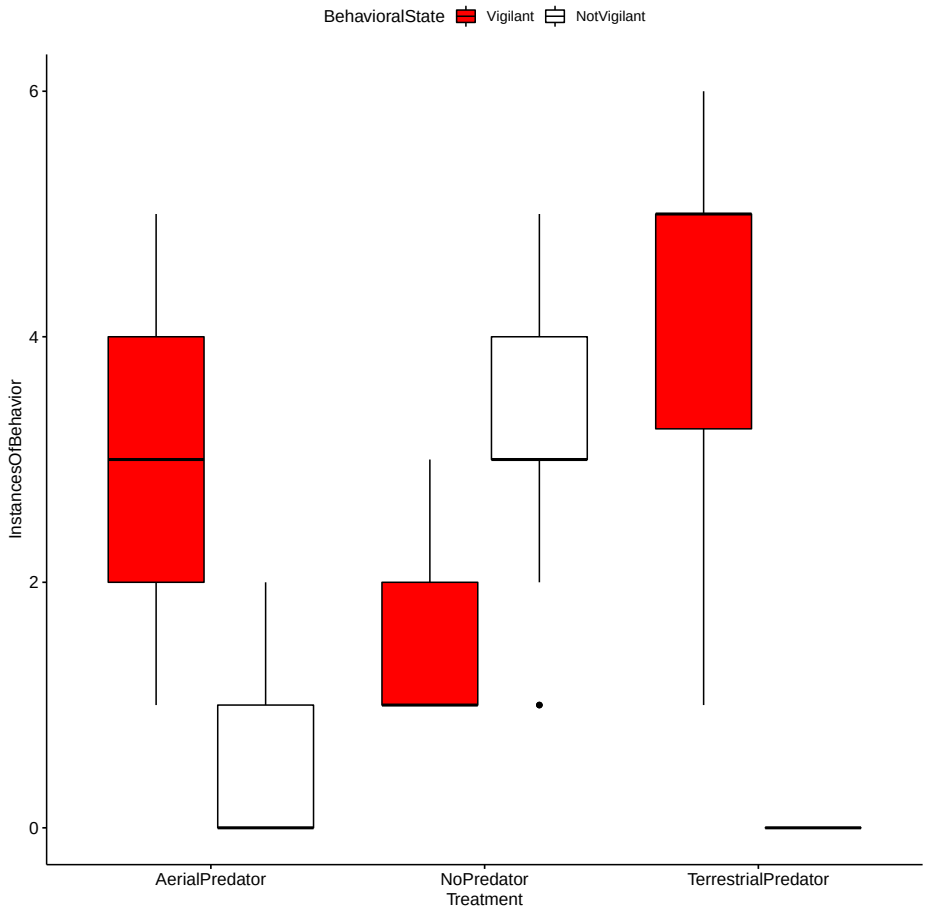
```
colnames(MeerkatScanDataSummaryLong) <- c('Treatment','Partner',
                                          'BehavioralState','InstancesOfBehavior')
```

We need to tell R that our outcome variable is not categorical but numeric

```
MeerkatScanDataSummaryLong$InstancesOfBehavior <-
  as.numeric(MeerkatScanDataSummaryLong$InstancesOfBehavior)
```

Now we plot our data.

```
ggboxplot(MeerkatScanDataSummaryLong,x='Treatment',
          y='InstancesOfBehavior', fill = 'BehavioralState')+scale_fill_manual(values = c('red','
```



Question 4. Based on your inspection of the boxplots, are there any major differences between treatment groups?

Now we will test to see if there were differences in vigilance behaviors across treatments.

```
# First we subset our data so that it only includes the vigilant category
MeerkatScanDataVigilantOnly <- subset(MeerkatScanDataSummaryLong,
                                      BehavioralState=='Vigilant' )

# R can be picky about the format of data, so we use this command to tell R that treat
MeerkatScanDataVigilantOnly$Treatment <-
  as.factor(MeerkatScanDataVigilantOnly$Treatment)

# Here we are reordering the levels of the factors. For our model selection we are int

MeerkatScanDataVigilantOnly$Treatment <-
  factor(MeerkatScanDataVigilantOnly$Treatment, levels = c("NoPredator", "AerialPredat
```

```
                                                        "TerrestrialPredator"))
```

Now as before we will do model selection. Note that because our outcome variable (instances of behavior) is in the form of count data we use a poisson distribution.

```
# This is the null model.
MeerkatVigilanceNullModel <- glm(InstancesOfBehavior ~ 1, family=poisson, data=MeerkatScanDataVig

# This is the model with treatment as a predictor of instances of vigilant behavior
MeerkatVigilanceModel <- glm(InstancesOfBehavior ~ Treatment, family=poisson,data=MeerkatScanData
```

Now we compare the models using AIC

```
bbmle::AICtab(MeerkatVigilanceNullModel,MeerkatVigilanceModel)
```

```
##                            dAIC df
## MeerkatVigilanceModel       0.0 3
## MeerkatVigilanceNullModel 59.2 1
```

Here we will use the summary function to look at the estimates. There is a lot of information here but we want to focus on the 'Estimate'. In particular we are interested in the estimates for 'TreatmentAerialPredator' and 'TreatmentTerrestrialPredator'. The estimate is showing the effect that these variables have on our outcome (instances of behavior), relative to our control (no predator).Therefore positive estimates indicate that there were more vigilance behaviors in aerial and terrestrial predator treatments.

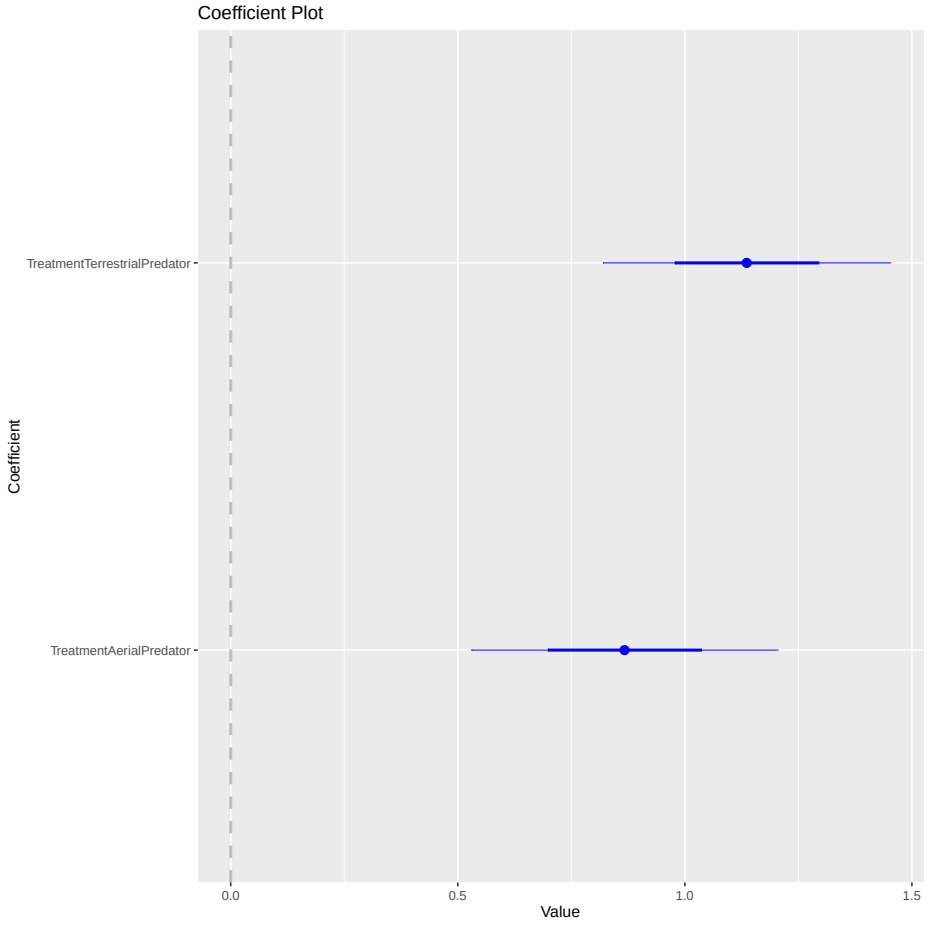
```
summary(MeerkatVigilanceModel)
```

```
##
## Call:
## glm(formula = InstancesOfBehavior ~ Treatment, family = poisson,
##     data = MeerkatScanDataVigilantOnly)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9295  -0.3475  -0.1588   0.3696   1.1825
##
## Coefficients:
##                              Estimate Std. Error z value Pr(>|z|)
## (Intercept)                    0.3285     0.1414   2.323   0.0202 *
## TreatmentAerialPredator        0.8671     0.1685   5.145 2.68e-07 ***
## TreatmentTerrestrialPredator   1.1362     0.1582   7.182 6.86e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##     Null deviance: 100.802  on 117  degrees of freedom
## Residual deviance:  37.587  on 115  degrees of freedom
## AIC: 384.39
##
## Number of Fisher Scoring iterations: 4
```

A common way to visulize results such as these are coefficient plots. Here we are looking at the effect of 'TreatmentAerialPredator' and 'TreatmentTerrestrialPredator' relative to our control group. The reference or group is indicated by the vertical dashed line. So, we can interpret that because the coefficients are positive (and the confidence intervals don't overlap zero) that both terrestrial and aerial treatments lead to an increase in vigilant behaviors.

```
coefplot::coefplot(MeerkatVigilanceModel,intercept=F)
```



For reasons that will not go into here, I am not a fan of p-values or null hypoth-

esis significance testing. There is a nice overview if you want to learn more here: https://doi.org/10.1098/rsbl.2019.0174. But,the model selection approach we used will lead to the same inference as the use of a one-way anova, an approach that you may be more familiar with.

Compute the analysis of variance

```
MeerkatAOV <- aov(InstancesOfBehavior ~ Treatment, data = MeerkatScanDataVigilantOnly)
```

Here this will tell us if there are differences between groups. A significant p-value (< 0.05) indicates there are differences between treatment groups.

```
summary(MeerkatAOV)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Treatment     2  176.0   88.02   87.03 <2e-16 ***
## Residuals   115  116.3    1.01
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

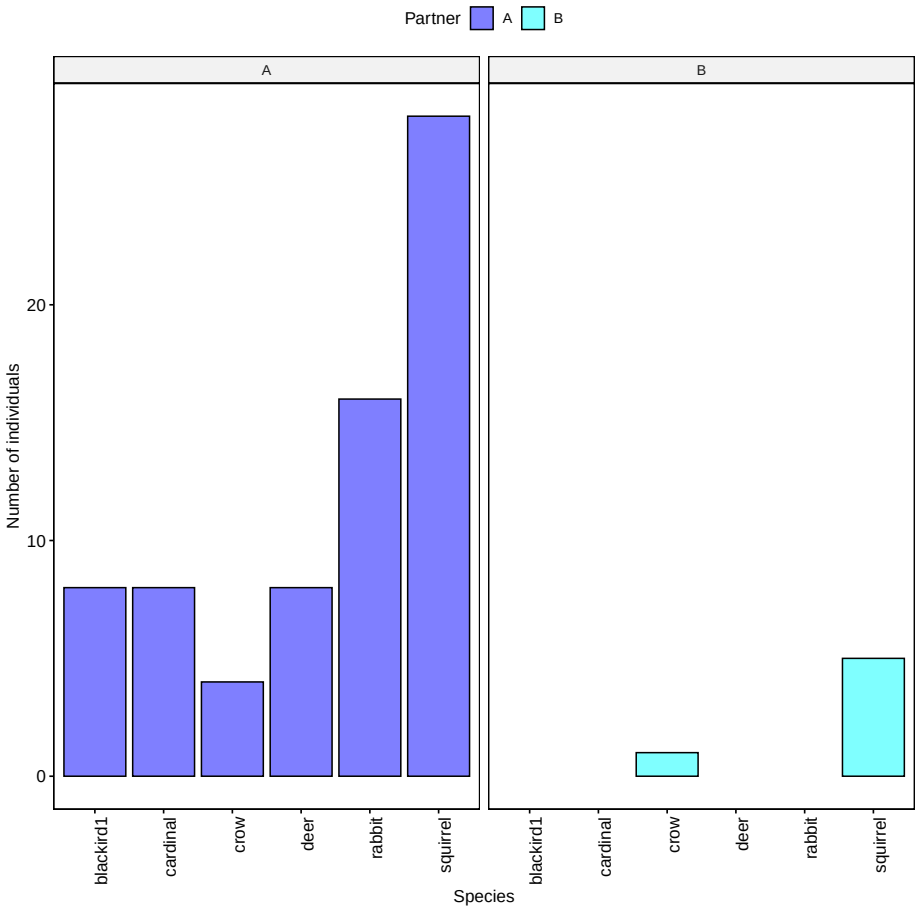
Since the ANOVA test is significant, we can compute Tukey Honest Significant Differences test. Again, a significant p-value (< 0.05) indicates there are differences between treatments.

```
TukeyHSD(MeerkatAOV)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = InstancesOfBehavior ~ Treatment, data = MeerkatScanDataVigilantOnly)
##
## $Treatment
##                                        diff       lwr      upr   p adj
## AerialPredator-NoPredator          1.916667 1.3538432 2.479490 0.0e+00
## TerrestrialPredator-NoPredator     2.937198 2.4058425 3.468554 0.0e+00
## TerrestrialPredator-AerialPredator 1.020531 0.4891758 1.551887 3.8e-05
```

Question 5. Based on your interpretation of the model selection and the coefficient plots were there differences between treatment groups (e.g. control, terrestrial and aerial predators) in meerkat vigilance behavior?

Computer Lab 5. Estimating Population Density and Biodiversity



Background
In lecture this week we learned about population- and community- level interactions, and how they can influence individual and species-specific behavior. Differences in population density can lead to shifts in the proportion of different behavioral phenotypes in a population, and the presence of competitors of other species can shape the behavior of animals on shorter (ecological) and longer (evolutionary) time scales. In this lab you will become familiar with the ways that scientists estimate population density and biodiversity of terrestrial vertebrates.

Goals of the exercises
The main goal(s) of today's lab are to:
1) Estimate population density of focal vertebrates from the field lab.
2) Learn the difference between alpha, beta and gamma diversity.
3) Compare population density and biodiversity across your sampling sites.

Getting started
First we need to load the relevant packages for our data analysis. Packages contain all the functions that are needed for data analysis.

```
library(behaviouR)
```

Then we read in our data

```
CensusData <- read.csv('FieldLab5CensusData.csv')
```

Part 1. Population density estimation.

First let's focus on your data.
NOTE: you need to change 'A' to the name you used!

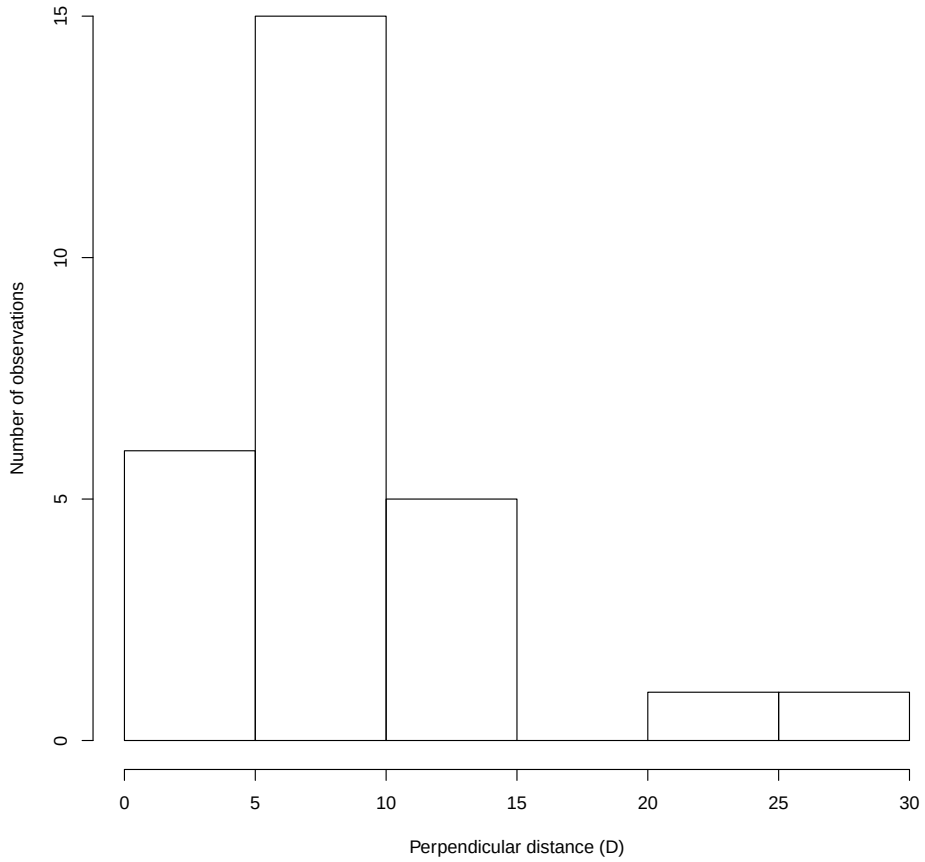
```
MyCensusData <- subset(CensusData, Partner=='A')
```

Now we will subset the data to focus on your focal species for density estimation.
NOTE: in this example we are subsetting the data so that we only have the squirrels; you will need to change the code to subset based on your focal species!

```
MyCensusDataFocal <- subset(MyCensusData, Species=='squirrel')
```

Here we will determine the width of our sampling area by creating a histogram of perpendicular detection distances. We will identify the point where our detections start to dropoff. We will assume that after this point our ability to detect animals past that distance drops substantially, and including observations past this distance could potentially bias our results.

```
hist(MyCensusDataFocal$PerpendicularDistance,xlab='Perpendicular distance (D)',
  ylab='Number of observations',main='')
```



In this example there is a clear break between 15 and 20 meters, so we will only use observations that were within 15 meters.
NOTE: Your data will look different than this!

Change the value (in this example 15) to the cutoff point indicated in your data!

```
CutOffPoint <- 15
```

```
# Here we subset our data so that it only includes detections that were within 15 mete
MyCensusDataFocalAdjusted <- subset(MyCensusDataFocal,PerpendicularDistance <CutOffPoi
```

Now we need to subset by each site.
NOTE: You will need to change to the site names you used

```
MyCensusDataFocalSiteA <- subset(MyCensusDataFocalAdjusted,Site=='SiteA')

MyCensusDataFocalSiteB <- subset(MyCensusDataFocalAdjusted,Site=='SiteB')
```

Now we will calculate the population density based on our two surveys. Change the following to indicate the distance (in meters) of your survey for site A. This will be the actual straight-line distance you walked.

```
SiteACensusDistance <- 500
```

Now we will calculate the area of our census. The sample area (a) is equal to the length of the transect multiplied by twice the width or a= 2wl. We divide by 1000 to convert our answer to square kilometers.

```
SiteACensusArea <- 2*SiteACensusDistance*CutOffPoint/1000
SiteACensusArea
```

```
## [1] 15
```

Now we need to calculate the number of focal animals observed using the following code.

```
NumberFocalAnimalsSiteA <- nrow(MyCensusDataFocalSiteA)
```

Then we can calculate the density by dividing the total number of animals we observed by the area we censused.

```
PopulationDensitySiteA <-NumberFocalAnimalsSiteA/SiteACensusArea
PopulationDensitySiteA
```

```
## [1] 0.4666667
```

Now we will calculate the population density based on our second survey. Change the following to indicate the distance (in meters) of your survey for site B.This will be the actual straight-line distance you walked.

```
SiteBCensusDistance <- 500
```

Now we will calculate the area of our census. The sample area (a) is equal to the length of the transect multiplied by twice the width or a= 2wl.We divide by 1000 to convert our answer to square kilometers.

```
SiteBCensusArea <- 2*SiteBCensusDistance*CutOffPoint/1000
SiteBCensusArea
```

```
## [1] 15
```

Now we need to calculate the number of animals using the following code.

```
NumberFocalAnimalsSiteB <- nrow(MyCensusDataFocalSiteB)
```

Then we can calculate the density by dividing the total number of animals we observed by the area we censused.

```
PopulationDensitySiteB <-NumberFocalAnimalsSiteB/SiteBCensusArea
PopulationDensitySiteB
```

```
## [1] 1.133333
```

Now we can compare population density using the following code. The code below asks if the population density of site A was higher than site B?

```
PopulationDensitySiteA > PopulationDensitySiteB
```

```
## [1] FALSE
```

The code below asks if the population density of site B was higher than site A?

```
PopulationDensitySiteB > PopulationDensitySiteA
```

```
## [1] TRUE
```

Question 1. What were the population density estimates (reported as number of individuals per square kilometer) for your two sites? Do your results of the population density estimates match your predictions? Why or why not?

Part 2. Comparing biodiversity.

Alpha diversity.

First, we will estimate the alpha diversity, or the diversity within a particular area or ecosystem. The alpha diversity is simply the number of different species present at each site.

Here we subset by partner and site A (you may need to change these!)

```
MyCensusDataSiteA <- subset(MyCensusData,Partner=='A' & Site=='SiteA')

# What were the unique species present?
unique(MyCensusDataSiteA$Species)
```

```
## [1] cardinal  blackird1 rabbit    deer      squirrel  crow
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

```
# How many unique species were there?
SiteANumberSpecies <- length(unique(MyCensusDataSiteA$Species))
SiteANumberSpecies
```

```
## [1] 6
```

Here we subset by partner and site B (you may need to change these!)

```
MyCensusDataSiteB <- subset(MyCensusData,Partner=='A' & Site=='SiteB')

# What were the unique species present?
unique(MyCensusDataSiteB$Species)
```

```
## [1] squirrel  blackird1 rabbit    deer      crow
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

```
# How many unique species were there?
SiteBNumberSpecies <- length(unique(MyCensusDataSiteB$Species))
SiteBNumberSpecies
```

```
## [1] 5
```

Question 2. Which of your sites had higher species richness (i.e. number of species)?

Beta diversity.

Now we will estimate beta diversity, which estimates changes in species diversity between ecosystems or along environmental gradients.

```
# This code tells us which species both sites have in common
intersect(unique(MyCensusDataSiteA$Species),unique(MyCensusDataSiteB$Species))
```

```
## [1] "blackird1" "rabbit"    "deer"      "squirrel"  "crow"
```

```
# Now we calculate the number of species in common
SpeciesInCommonBothSites <- length(intersect(unique(MyCensusDataSiteA$Species),unique(MyCensusDat
)
SpeciesInCommonBothSites
```

```
## [1] 5
```

To investigate community similarity we will calculate Sørenson's index; a value of 1 means exactly the same number of species a value of 0 means no overlap.

Beta diversity = 2c / S1 + S2
Where c is the number of species the sites have in common, S1 is the number of species at the first site and S2 is the number of species at the second site

```
2*SpeciesInCommonBothSites/ (SiteANumberSpecies+SiteBNumberSpecies)
```

```
## [1] 0.9090909
```

Now we will use your partner's data to estimate alpha diversity

```
MyPartnersCensusDataSiteA <- subset(CensusData,Partner=='B' & Site=='SiteC')
unique(MyPartnersCensusDataSiteA$Species)
```

```
## [1] squirrel crow
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

```
length(unique(MyPartnersCensusDataSiteA$Species))
```

```
## [1] 2
```

```
SiteANumberSpecies <- length(unique(MyPartnersCensusDataSiteA$Species))
SiteANumberSpecies
```

```
## [1] 2
```

```
MyPartnersCensusDataSiteB <- subset(CensusData,Partner=='B' & Site=='SiteD')
unique(MyPartnersCensusDataSiteB$Species)
```

```
## [1] squirrel
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

```
length(unique(MyPartnersCensusDataSiteB$Species))
```

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

```
SiteBNumberSpecies <- length(unique(MyPartnersCensusDataSiteB$Species))
SiteBNumberSpecies
```

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

Question 3. How did the alpha diversity of each of your sites compare with that of your partner?

Gamma diversity

Gamma diversity is the total number of species over a large area or region; there are many different ways that this can be measured. The way we will do it is a bit of an oversimplification by simply comparing the number of species seen during the census at both locations.

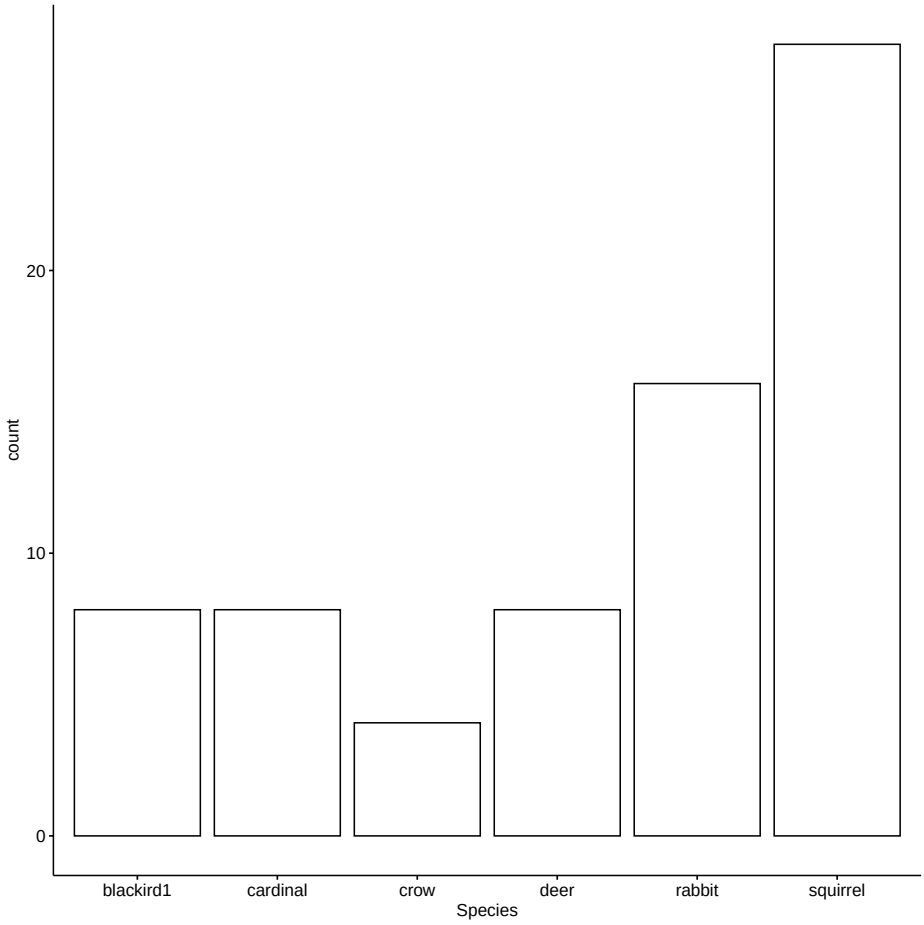
First we subset by the first partner

```
MyCensusData <- subset(CensusData, Partner=='A')
unique(MyCensusData$Species)
```

```
## [1] cardinal  blackird1 rabbit    deer      squirrel  crow
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

Then we plot the results

```
gghistogram(data=MyCensusData, x='Species',stat="count")
```



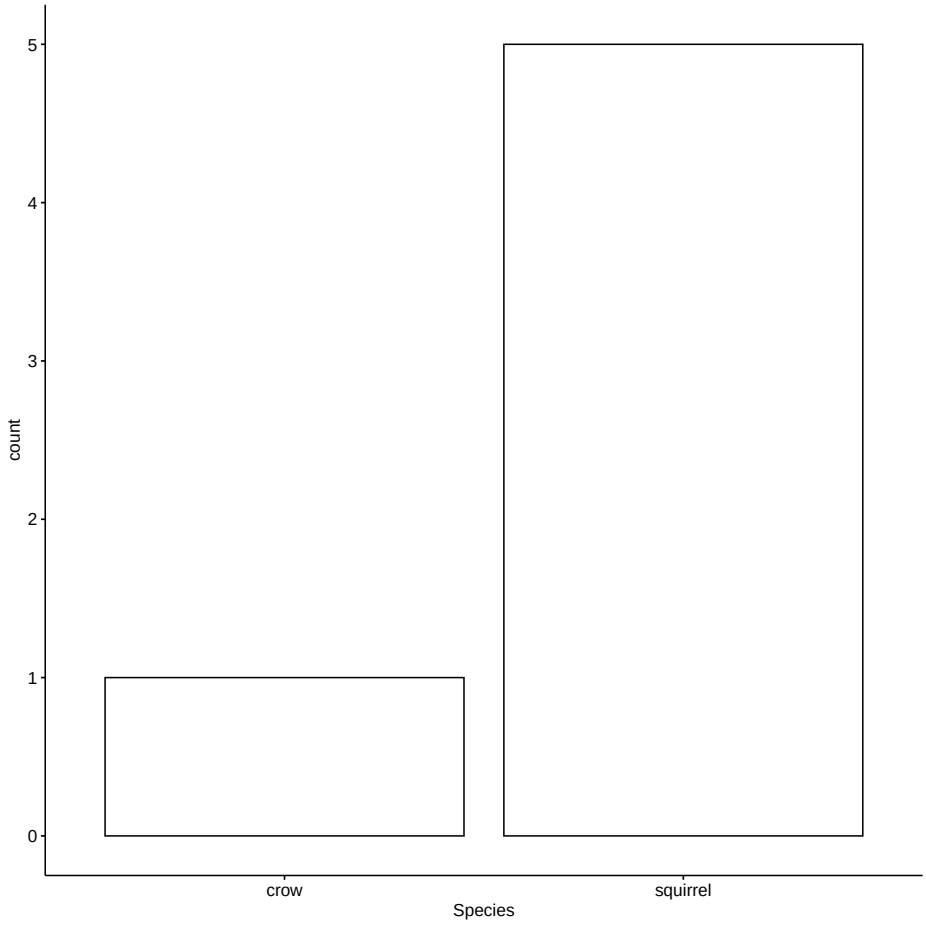
Now we subset by the second partner

```
MyPartnersCensusData <- subset(CensusData, Partner=='B')
unique(MyPartnersCensusData$Species)
```

```
## [1] squirrel crow
## Levels: blackird1 cardinal crow deer rabbit squirrel
```

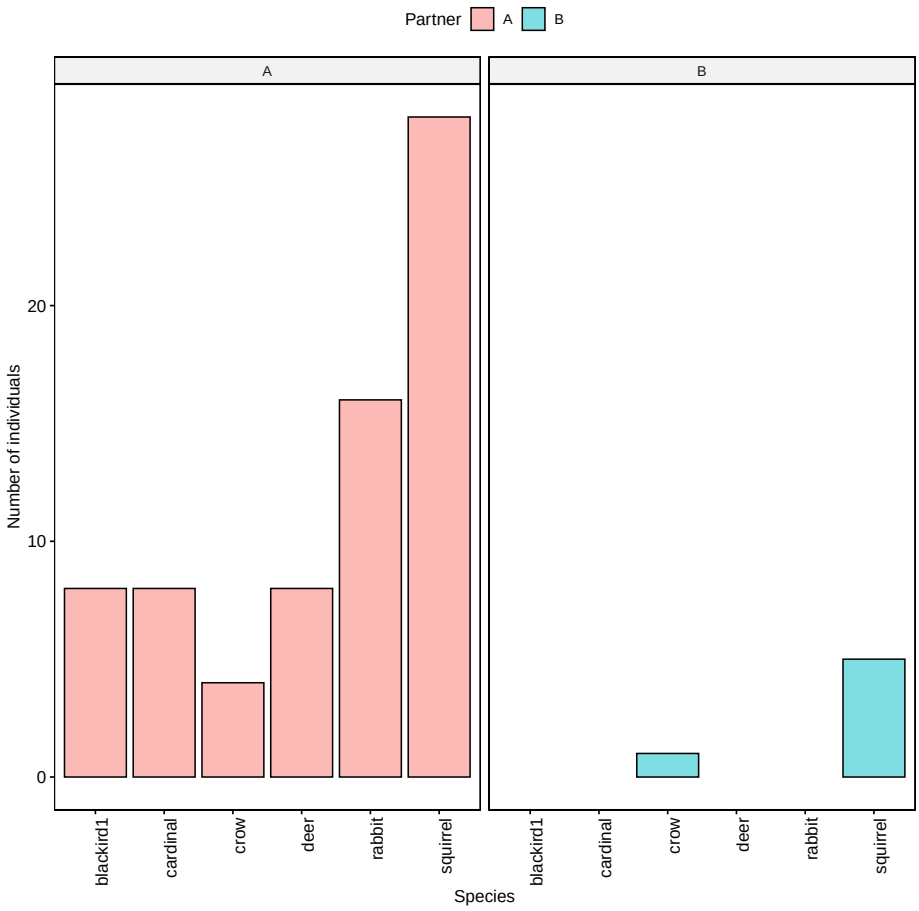
Then we plot the results

```
gghistogram(data=MyPartnersCensusData, x='Species',stat="count")
```



Now we can plot all the data together, separated by partner

```
gghistogram(data=CensusData, x='Species',stat="count",
            facet.by = 'Partner',x.text.angle =90,
            fill='Partner')+xlab('Species')+ylab('Number of individuals')
```



Question 4. How did the gamma diversity of your site compare with that of your partners?

Part 3. Biodiversity indices in the real world.

There are special packages in R that can measure different diversity indices, as biodiversity indices are tools many ecologists use. The package we will use is called 'vegan'.

```
library(vegan)
```

First we need to convert our data into a table that can be used to calculate the indices.

```
BiodiversityTable <- table(CensusData$Site,CensusData$Species)
```

Simpson's Index (D) measures the probability that two individuals randomly

selected from a sample will belong to the same species (or some category other than species). With this index, 1 represents infinite diversity and 0 means no diversity.

```
H <- diversity(BiodiversityTable,index="simpson")
H
```

```
##     SiteA     SiteB     SiteC     SiteD
## 0.7962963 0.6728395 0.4444444 0.0000000
```

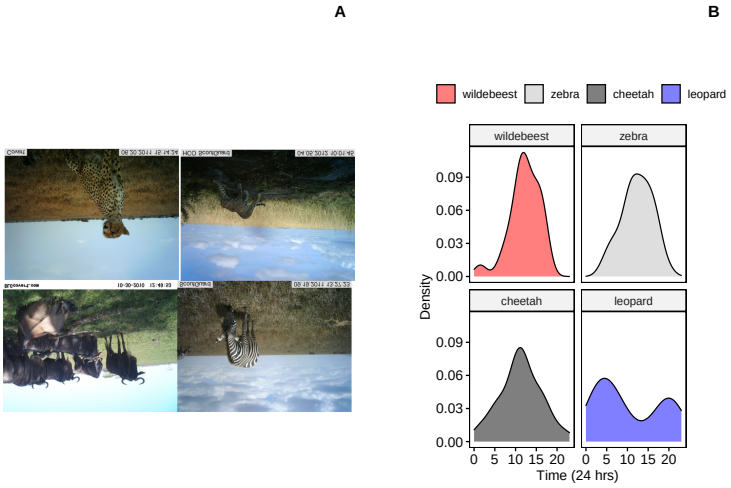
Species evenness refers to how close in numbers each species is in an environment. We can calculate evenness using the following code. The value is constrained between 0 and 1, with communities that have a more even representation of species having values closer to 1.

```
J <- H/log(specnumber(BiodiversityTable))
J
```

```
##     SiteA     SiteB     SiteC     SiteD
## 0.4444214 0.4180587 0.6411978       NaN
```

Question 5. Which of the four sites you analyzed was most diverse? Which was the most even? Why is it important to consider diversity and evenness when studying biodiversity?

Computer Lab 6. Analyzing camera trap data.



Background
In this lab we will use actual camera trap data collected in Serengeti National Park to investigate temporal niche partitioning. The camera trap data that we will use for this field lab comes from the 'Snapshot Serengeti' project (http://lila.science/datasets/snapshot-serengeti). The scientists set out out 225 cameras within a 1,125 km2 area. The cameras have been deployed continuously since 2010, and the researchers recruited citizen scientists to help with classifying images (https://www.zooniverse.org/projects/zooniverse/snapshot-serengeti).

Goals of the exercises
The main goal(s) of today's lab are to:
1) Understand the types of behavioral data that scientists can collect from camera trap photos.
2) Become familiar with one of the ways scientists can compare differences in activity patterns.

3) Think about interspecific interactions of predator-prey and potential competitors in the Serengeti.

Getting started
First we need to load the relevant packages for our data analysis. Packages contain all the functions that are needed for data analysis.

```
library(behaviouR)
library(ggpubr)
```

Part 1: Collect Serengeti camera trap data

The original dataset includes over 2.65M sequences of camera trap images (totaling 7.1M images) from ten field seasons. As you can imagine that is a lot of data! For this lab we will focus on a subset of the data collected over the course of a few seasons.

We will be taking random subsets of the camera trap photos (the current default is 5 per season which you should change for your actual study), and we use 'set.seed' to make sure that our results are reproducible.

```
set.seed(2210)
```

We can use the following code to query the database by season to see which photos are available. The available seasons are 1-4, 6-8,10,11. This function will return a table with all of the available camera trap photos for a particular season. With camera trap data often times photos will be taken in a sequence (e.g. if the animal is moving in front of the camera for a long time). The values in the table below include all photos in a sequence, but the code we use to download the photos only takes the first photo in a sequence. Therefore the sample sizes indicated in this table and the actual sample size may be different.

```
CameraTrapAnnotations(season = 1)
```

```
##
##       aardvark       aardwolf         baboon    batearedfox        buffalo
##            143             88            348            315           1455
##       bushbuck        caracal        cheetah          civet         dikdik
##             36             57            575             27            728
##         eland       elephant   gazellegrants gazellethomsons          genet
##            113           1687           4967          29705             28
##        giraffe      guineafowl           hare     hartebeest    hippopotamus
##           1923           3217            327           2139            340
##    honeybadger    hyenaspotted    hyenastriped         impala         jackal
##            42           2752            112            691            304
##    koribustard        leopard     lionfemale       lionmale       mongoose
##            591             38           1519            665            290
```

```
##    monkeyvervet         ostrich       otherbird       porcupine        reedbuck
##              64             211            2322             114             946
##        reptiles      rhinoceros         rodents   secretarybird          serval
##             357               6             129             198             174
##            topi        warthog       waterbuck         wildcat      wildebeest
##             428            2685              25              48             689
##           zebra         zorilla
##            5592               9
```

For this lab you are going to choose two animals from the camera trap dataset. It could be a pair of animals that are predator and prey or two potential competitors. Once you decide on the two animals you want to compare you will make some predictions about how you think they will differ in their activity patterns.

Now we will use another function to download the camera trap photos and save them locally to our computer. You can change the season by changing the values for season (remember the available seasons are 1-4, 6-8,10,11). You can change the focal animal by changing the 'AnimalID'; make sure that the spelling and case is exactly the same as in the table above. You can change the number of photos per season that you download (the default is 5). When 'create.dir' is true this will create a folder in your current working directory. For this example the folder is called 'CameraTrapPhotoszebra'. If create.dir='FALSE' all the photos will be downloaded directly to your working directory. To find the directory type 'getwd()' into your R console.

```
CombinedAnimalDF <- CameraTrapDataAccess(urlpath= 'https://lilablobssc.blob.core.windows.net/snap
                    season= list(1,2),AnimalID='zebra', NumPhotos= 5,create.dir=TRUE)
```

```
## ERROR : cannot open URL 'https://lilablobssc.blob.core.windows.net/snapshotserengeti-unzipped/
```

Here we isolate only the columns from the dataframe that we need.

```
CombinedAnimalDF <- CombinedAnimalDF[,c("category_id","season","location","filename")]
head(CombinedAnimalDF)
```

```
##       category_id season location                                filename
## 4742        zebra     S1      C06 CameraTrapPhotoszebra/zebra_S1_C06.JPG
## 292         zebra     S1      B05 CameraTrapPhotoszebra/zebra_S1_B05.JPG
## 7513        zebra     S1      D05 CameraTrapPhotoszebra/zebra_S1_D05.JPG
## 4582        zebra     S1      C06 CameraTrapPhotoszebra/zebra_S1_C06.JPG
## 45467       zebra     S2      E02 CameraTrapPhotoszebra/zebra_S2_E02.JPG
## 26166       zebra     S2      D02 CameraTrapPhotoszebra/zebra_S2_D02.JPG
```

The function below will allow you to enter data and look through each photo included in the 'CombinedAnimalDF' spreadsheet and enter the time that the photo was taken. You can change option='Plot' to option='Viewer' to load the photos more quickly, but you will have to expand the photo to see the whole thing. The input file should be the dataframe created using the 'CameraTrap-

DataAccess' function. You can break out of the function at any time by typibg 'break'. It will print out which row of the dataframe you were on when you stopped the function. If you would like to resume where you left off you must change 'rowstart=1' to the row number indicated, and change 'dataframe.cont = FALSE' to 'dataframe.cont = TRUE'.

Here is the function to view photos and annotate data.

```
CombinedAnimalDF_TimeAdded <- CameraTrapDataCollection(inputfile = CombinedAnimalDF,
                                                       rowstart = 1, dataframe.cont = I
```

Now you may want to save your datasheet locally

```
write.csv(CombinedAnimalDF_TimeAdded,'CombinedAnimalDF_TimeAdded.csv',row.names = F)
```

Part 2: Analyze your Serengeti camera trap data

First we load your data sheet with the times added. NOTE: Make sure that your updated datasheet has the exact same name as the file indicated below.

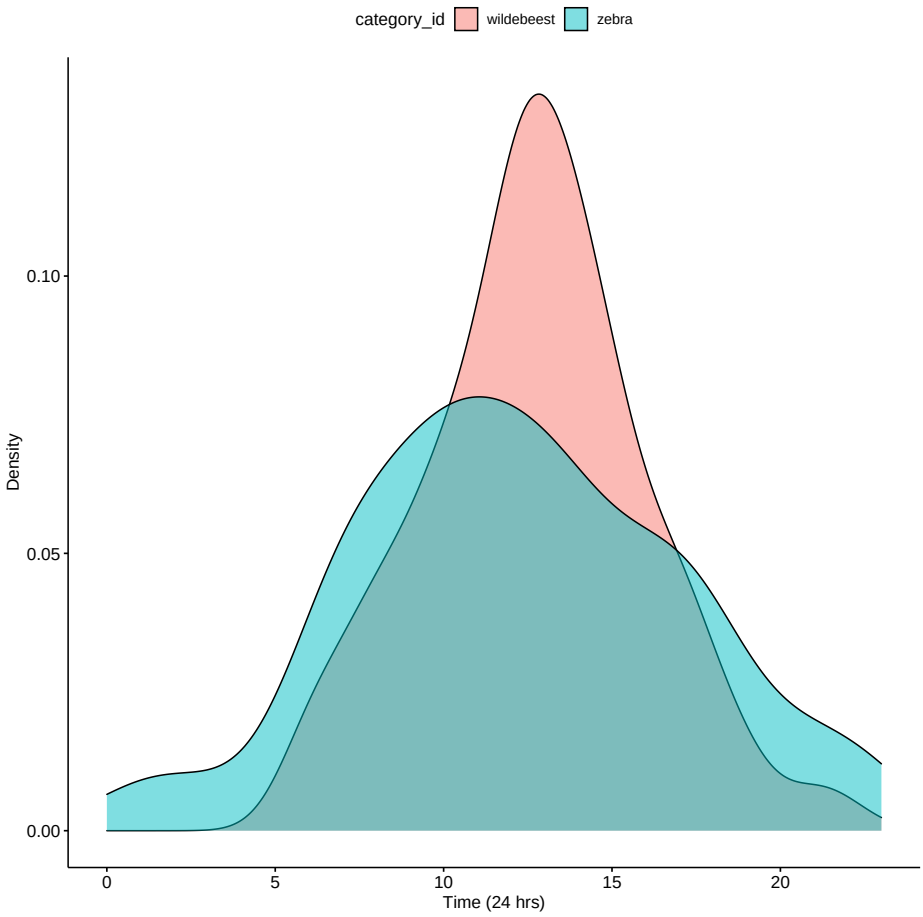
```
# You can read in your datasheet that you saved
CombinedDFTimes <- read.csv('CombinedAnimalDF_TimeAdded.csv')
```

Then we can check the structure

```
head(CombinedDFTimes)
```

Now we can make a density plot that will show the distribution of camera trap photos that were taken over 24-hours. We add the fill = 'category_id' so that we show different distributions for each animal.

```
ggdensity(data=CombinedDFTimes,x='Time',fill = 'category_id')+
  xlab('Time (24 hrs)') +ylab('Density')
```



Question 1. What do you notice about the overlap of the two density curves? Does it look like there is temporal niche partitioning?

Now we can calculate an overlap coefficient which can be used to investigate potential competitive and interaction possibilities between species.The value ranges from ranges from 0 (no overlap) to 1 (complete overlap).

First we subset our data to focus on the first animal in our dataset

```
FirstAnimal <- subset(CombinedDFTimes,category_id=='zebra')
```

Then we subset our data to focus on the second animal in our dataset

```
SecondAnimal <- subset(CombinedDFTimes,category_id=='wildebeest')
```

Now we use the overlap function to calculate the overlap coefficient

```
bayestestR::overlap(FirstAnimal$Time,SecondAnimal$Time)
```

```
## # Overlap
```

```
##
## 0.81
```

Question 2. How do you interpret the overlap coefficient for your data?

Part 3: Focus on your partner's Serengeti camera trap data

Now we will read in our partners data.
NOTE: Make sure that your updated datasheet has the exact same name as the file indicated below.

```
CombinedPartnerDFTimes <- read.csv('CombinedAnimalDF_TimeAddedPartner.csv')
```

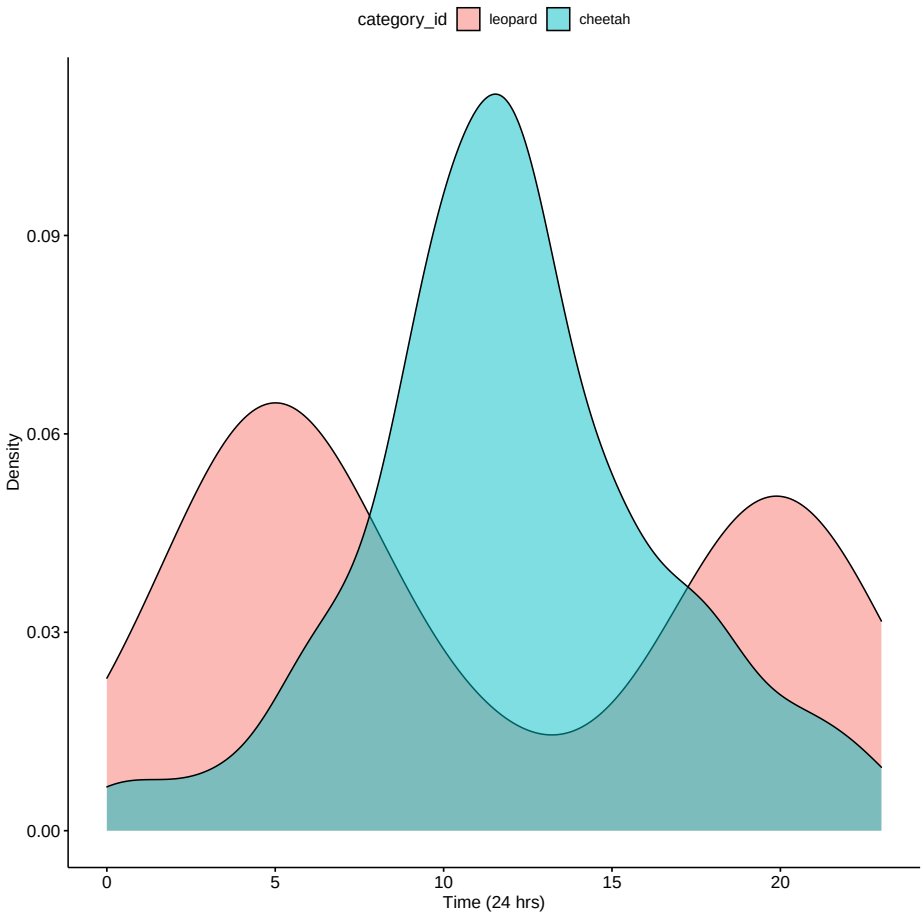
Let's check the structure

```
head(CombinedPartnerDFTimes)
```

```
##   category_id season location                                 filename Time
## 1     leopard     S1      B06 CameraTrapPhotosleopard/leopard_S1_B06.JPG    4
## 2     leopard     S1      D06 CameraTrapPhotosleopard/leopard_S1_D06.JPG   20
## 3     leopard     S1      D06 CameraTrapPhotosleopard/leopard_S1_D06.JPG   20
## 4     leopard     S1      D06 CameraTrapPhotosleopard/leopard_S1_D06.JPG   20
## 5     leopard     S1      E06 CameraTrapPhotosleopard/leopard_S1_E06.JPG    4
## 6     leopard     S1      E06 CameraTrapPhotosleopard/leopard_S1_E06.JPG    4
```

Now we can make a density plot for our partner's data.

```
ggdensity(data=CombinedPartnerDFTimes,x='Time',fill = 'category_id')+
  xlab('Time (24 hrs)') +ylab('Density')
```



Question 3. What do you notice about the overlap of the two density curves for your partner's data? Does it look like there is temporal niche partitioning?

Now we can calculate an overlap coefficient which can be used to investigate potential competitive and interaction possibilities between species.The value ranges from ranges from 0 (no overlap) to 1 (complete overlap).

First we subset our data to focus on the first animal in your partner's data set

```
FirstAnimalPartner <- subset(CombinedPartnerDFTimes,category_id=='cheetah')
```

Then we subset our data to focus on the second animal in our dataset

```
SecondAnimalPartner <- subset(CombinedPartnerDFTimes,category_id=='leopard')
```

Now we use the overlap function to calculate the overlap coefficient

```
bayestestR::overlap(FirstAnimalPartner$Time,SecondAnimalPartner$Time)
```

```
## # Overlap
```

```
##
## 0.44
```

Question 4. What is the overlap coefficient for your partner's data? How do you interpret this?

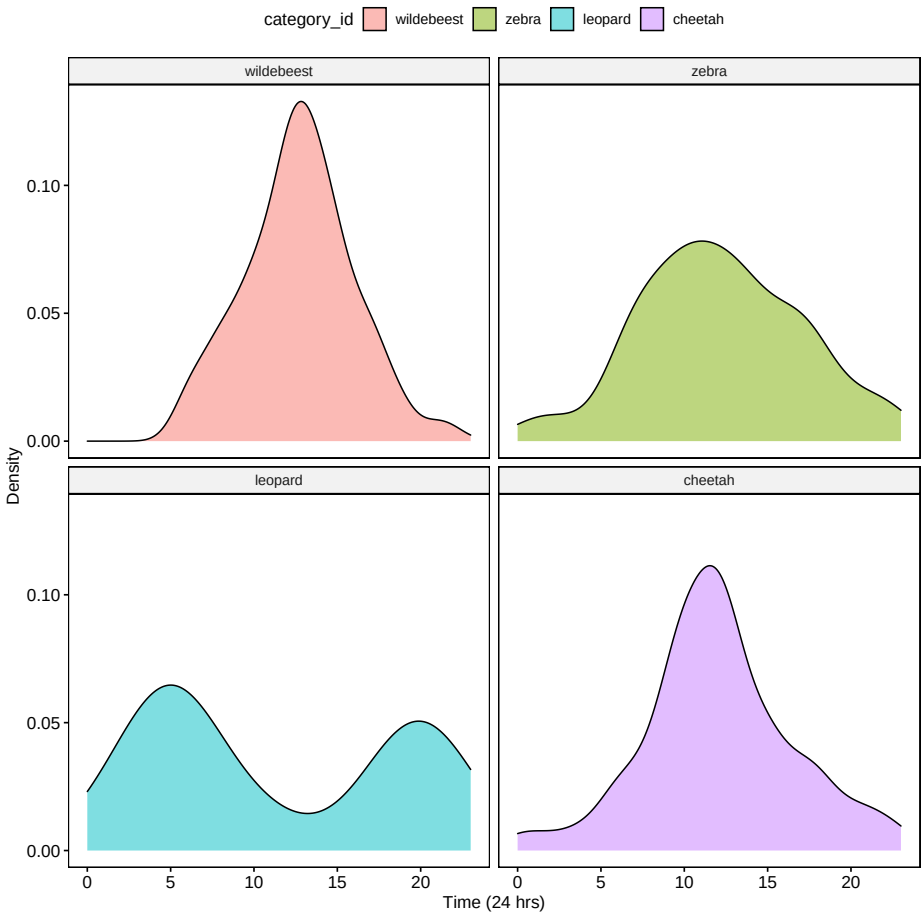
Part 4. Investigating temporal niche partitioning in four different animals

First we combine both datasets

```
AllDataCombined <- rbind.data.frame(CombinedDFTimes,CombinedPartnerDFTimes)
```

Then we plot the diel activity patterns for all four species

```
ggdensity(data=AllDataCombined,x='Time',fill = 'category_id',facet.by = 'category_id')+
  xlab('Time (24 hrs)') +ylab('Density')
```



Question 5. Based on the activity patterns and your understanding of the Serengeti food web how do you interpret these results? Is there evidence of temporal niche partitioning among potential competitors? What about interactions between potential predators and prey?

Field Lab 1: Introduction to how we study behavior

Part 1. Observe behaviors

Choose a safe spot to sit, set a timer for 10 minutes and observe the animals around you. Write down any and all animal behaviors you observe. If you know the species that is great, otherwise you can just classify into major groups (e.g. birds, insects). You and your partner will do this exercise independently. NOTE: If you do not feel that you can do this exercise safely in your current location please contact me and I will give you a virtual option.

Question 1. Now that you have spent 10 minutes observing and taking notes of animal behavior, what do you see as pros/cons of this approach?

Part 2. How we describe behaviors (ethograms)

Now you and your partner will compare notes and create an ethogram that includes broad categories of the behaviors you observed.

Question 2: What were the broad categories you and your partner included in your ethogram?

Part 3. Collect data on a focal species or taxonomic group

Using your newly created ethogram you will now collect data on a focal species or taxonomic group. As before, set a timer for 10 minutes, but this time use your ethogram to collect data. Tally how many times your organism(s) exhibit each behavior in your ethogram.

Question 3: What were the most common behaviors you observed? What about the least common?

Part 4: Develop a research question, hypothesis and prediction

Once you have taken time to make some behavioral observations, or frankly as you are making them, select one particular behavior that interests you the most. Use this behavior to generate a Question -> Hypothesis -> and Prediction about its relevance to your organism's life history. Pick any level of analysis. If you don't have time to do this part in the field, you can complete it afterward. In fact, taking time to think about the behaviors you observed is often very important when coming up with questions, hypotheses, and predictions.

Question 5: What was the behavior you observed? What question did you develop about this behavior? What hypotheses do you have about the function of this behavior?

Part 5. Find a primary literature article related to your hypothesis

Now your job is to find a primary literature article related to your hypotheses and predictions outlined in Part 4.

Write a brief summary that you will share with me and your partner. Include the following information (~1 sentence each):

- Background: What was already known before the study?
- Research question(s): What was the main research question? What hypotheses were they testing?
- Methods: How did they test their hypotheses?
- Results: What were the main findings?
- Discussion/Future Directions: What were their main conclusions? Did they highlight any future directions?

Field Lab 2: Ethograms and activity budgets

NOTE: you may want to look ahead to Computer Lab 2. You will need to enter your data from this lab into the spreadsheets. For the field lab you can take the data anyway that you like, but be prepared to enter and format it for the computer lab!

Part 1. Build an ethogram from meerkat observations

Your first task will be to watch this short video: https://vimeo.com/80600819 (Links to an external site.). While you watch the video, document all of the different behaviors that you observe. After you watch the video, develop an ethogram with your partner including: a) behavioral categories and b) definitions for each category.

Question 1a. Were the behaviors that you and your partner listed different?
Question 1b. Please enter your ethogram (behavioral categories and definitions) here. Make sure to include 'other' and 'out of view' as two of your categories.

Part 2. Focal versus scan sampling

Watch this video for an introduction to focal versus scan sampling.

Part 3. Focal sampling and inter-observer reliability

Now, both you and your partner will watch this 10-minute focal video of one meerkat (Bumble) https://vimeo.com/80602697. You will want to develop a short code that you can use for each behavior category (e.g. drinking could be denoted with a 'D'). You will want to note the start time of each behavior and the behavior code. You can take the initial data using any format you choose (e.g. I used a pen and paper) and you can subsequently enter your data into the spreadsheet.

Question 2. How similar or different were your observations to your partners?

Part 4. Calculating activity budgets using focal data

Activity budgets are used to provide information about how animals spend their time. They are generally presented as the proportion or percent time that an animal spends in a particular activity. They are important for providing baseline data on animal behavior, and can be used for testing hypotheses related to different experimental treatments. To calculate meerkat activity budgets use the following steps:

1. Calculate the number of seconds engaged in each entry on data sheet.
2. Sum up total number of seconds for each behavior
3. Divide the total number of seconds engaged in behavior by the total number of seconds of observation (in this case 600).
4. Then multiply each value by 100 to report your activity budget in percent

Question 3. Enter the results of your activity budget calculations here.

Part 5. Scan sampling and inter-observer reliability

Part 5a. First, watch this clip of meerkats engaging in sentry and vigilance behaviors (refer to the introduction for definitions of these behaviors).

https://vimeo.com/80600820

Part 5b. Now you and your partner will watch three videos of meerkats under different experimental treatments.

The treatments will be: in the presence of no predator, in the presence of a terrestrial predator and in the presence of an aerial predator. Set a timer to go off every 10 seconds, and at each interval record the number of meerkats performing each behavior. We will use a modified ethogram that includes only three categories: vigilant, not vigilant or out of site.

Vigilant: Head raised at or above horizontal plain and eyes open (to include scanning /guarding / raised guarding). Not Vigilant: Eyes closed or head lower than horizontal plane (to include foraging, moving, sleeping, resting). Out of sight: Not visible by the researcher

Question 4. Do you think meerkats will be more or less vigilant in the presence of predators? Will their behavior vary depending on the type of predator?

NOTE: Do not communicate with your partner while watching the videos!

No predator: https://vimeo.com/80600822 (Links to an external site.)

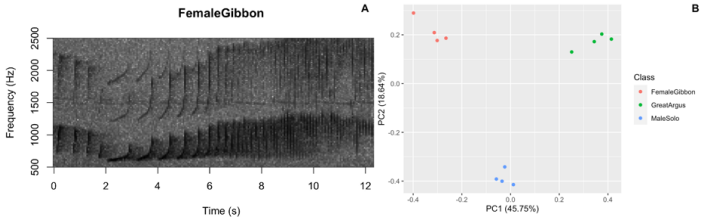
Terrestrial predator: https://vimeo.com/71877438 (Links to an external site.)

Aerial predator: https://vimeo.com/80600821 (Links to an external site.)

Question 5. Were there any differences between you and your partner?

Material for this lab was adapted from: Hammond 2019, Vigilance behaviour in meerkats, ASAB Education. Please cite this source in any labs based on this material.

Field Lab 3: Introduction to bioacoustics



Part 1. Introduction

As we have learned in lecture this week, animals communicate through a variety of modalities including visual, auditory and chemical. In lab this week we will focus on bioacoustics techniques, or the study of animal sounds and their habitats. For this lab you will need access to a smartphone or device with recording capabilities, a place where you can record vocal animals and a notebook to write down your observations.

NOTE: If you do not have access to a recording device and/or do not feel you can do this lab outside safely please contact me for a virtual option.

Recording on your smartphone

Please see the website below for tips to record using your smartphone. It would be better to download one of the free apps listed that records sound files as '.wav'. If this is not possible you can send me your recordings in the format you have and I can convert them for you.

https://support.ebird.org/en/support/solutions/articles/48001064305-smartphone-recording-tips

Tutorial: How we study sounds

Please see the 'How we study sounds' tutorial. Link here: https://rstudio.cloud/project/1381321. You will need to create an RStudio cloud login. To access the tutorial open the 'HowWeStudySoundTutorial.Rmd' file and then click the 'knit' button. It may take a few minutes to load.

Question 1. What were the main differences between the duets of the different primates? *Question 2.* What did you notice about the two soundscapes in Ithaca during different seasons? What about the soundscapes from Borneo and Sulawesi?

Part 2. Focal recordings

Focal recordings are generally done using a hand-held, battery operated recording device in the presence of human observer. Focal recordings are used for testing hypotheses related to behavior of the focal animal(s), and the goal is to get high-quality recordings that can be used for subsequent analysis. But, getting high-quality recordings of your target animal is not easy!

For this part of the lab, you are going to record short (5-10 seconds each) focal recordings of any target species. Coordinate with your partner so that you will have the same taxonomic focus (e.g. two species of bird, two species of frog). Aim to get 3-5 recordings of each species. Once you are done give your recordings standardized informative names following the format below (this will help with our analysis later):

Partner 1: 'BirdSpecies1_a.wav', 'BirdSpecies1_b.wav', 'BirdSpecies1_c.wav'

Partner 2: 'BirdSpecies2_a.wav', 'BirdSpecies2_b.wav', 'BirdSpecies2_c.wav'

To be good scientists, make sure to write down the day, time, location of recordings and any weather notes!

Important: Once you have your recordings please email them to me so that I can make sure they are converted to the right format.

Question 3. What did you and your partner decide to record? What was your sample size? What did you notice about the calling behavior of your target species? *Question 4.* Did you encounter any difficulties during data collection?

Part 3. Soundscapes

Traditionally, studies of animal behavior relied on human observers, but the use of technology (e.g. acoustic recorders and camera traps) has greatly expanded

the spatial and temporal scales that researchers can collect data. Battery-operated, autonomous acoustic recorders can collect data on all vocal animals at the same time, and generally have a farther detection range than camera traps. Soundscapes are the combination of all sounds in a particular environment. Scientists have become increasingly interested in understanding how soundscapes vary over space and time, and in relations to human disturbance.

Part 3a. Study Design

For this part of the lab you and your partner are going to design a small study that will investigate variation in soundscapes at different times (e.g. dawn and dusk) and different locations (e.g. urban versus rural). For each time and location, we will want 3-5 recordings that are the same duration. In an ideal world we would record continuously for 24 hours, but due to space limitations please limit each individual recording to 20 seconds.

Question 5: In the space below briefly outline your study design and predictions. Do you predict there will be more animals calling at a particular time, or at a particular location? How long will you wait in between recordings?

Part 3b. Data collection

Now collect your data! While you are recording, use your notebook write down all the sounds that you hear. If you don't know the species no problem! Once you are done please format your recordings in the following way:

Partner 1: 'Site1_Morning_0800.wav', 'Site1_Morning_0805.wav'

Partner 1: "Site2_Evening_1800.wav', "Site2_Evening_1805.wav'

Important: Once you have your recordings please email them to me so that I can make sure they are converted to the right format.

Question 5. What did you notice during data collection? Were there any differences in the recording times or locations?

Question 6. Do you feel you were able to document all of the sounds you heard in your notebook? Did you encounter any difficulties?

Part 4. Optional acoustic trivia

Here is a link to acoustic trivia put together by the Center for Conservation Bioacoustics: https://www.birds.cornell.edu/ccb/acoustic-trivia/.

Field Lab 4: Vigilance behavior

Part 1. Introduction

Please watch this short video for an introduction to vigilance behavior.

Question 1. In this lab we will be investigating the relationship between group size and vigilance behavior. Please develop a hypothesis and an associated prediction related to predation risk, group size and vigilance behavior. Remember your hypothesis is related to the 'big-picture' question, and the prediction is something that you can actually test.

Part 2. Data collection

In your notebook document the following:

1. Flock size
2. The number of times they are in the 'head up' position
3. The start and stop time of each 'head up' position.

I recommend using a stopwatch for this. Here is a link to the video:

Part 3. Data entry

Please calculate the total number of 'head up' behaviors and the total duration of 'head up' for each flock size and enter into the following data sheet. Then share with your partner.

Part 4. Follow-up questions

Question 2. Before we do any data analysis, please provide a qualitative analysis. Were there any trends that were apparent to you (the human observer)?

Question 3. In your own words, describe the potential costs and benefits of living in groups. Please discuss in terms of predation risk and feeding competition.

Question 4. What are some other factors (besides predation risk and group size) that could influence the vigilance behavior of an individual?

Field lab 5. Estimating population density and biodiversity

Part 1. Background

In lecture this week we learned about population- and community- level interactions, and how they can influence individual and species-specific behavior. Differences in population density can lead to shifts in the proportion of different behavioral phenotypes in a population, and the presence of competitors of other species can shape the behavior of animals on shorter (ecological) and longer (evolutionary) time scales.
For many types of behavioral studies—and for effective conservation of animals—a clear estimate of population density of focal species along with estimates of overall biodiversity of an area are necessary. In the field lab this week you will become familiar with field methods scientists use to estimate population density and biodiversity. There are many different methods and approaches that scientists can use, and the methods we will use in this lab are most appropriate for studies of terrestrial vertebrates.

Part 2. Estimating distance

A crucial part of the approach we will use today is the ability to estimate how far animals are away from you. If we were doing this activity for a scientific study, we would use a range finder to get accurate estimates. But, we can also use the number of steps we take to estimate distance. The best way to do this is to use a tape measure to estimate your stride length (or distance covered in a single step) in meters. If this is not possible, and you have an app on your phone that tracks distance and number of steps you could divide the distance of your last walk (in meters)/number of steps. Or you can visit this website

(http://www.kylesconverter.com/length/steps-to-meters).

Question 1. What did you estimate your stride length to be in meters? Which approach did you use to estimate this?

Part 3. Designing your study

You and your partner will develop testable research questions related to both population density of a focal species, and overall biodiversity. You and your partner do not need to focus on the same species, as depending on your geographic location this may not be possible. You will conduct the census twice in two distinct habitats (it could be as simple as your front and back yard, different streets, the edge versus the center of a park, etc.). Ideally, the length of each census will be 500 meters and in a straight line, but you can do shorter or longer depending on your location and limitations. To control for potential differences in animal circadian rhythms, it would be best to conduct your censuses at the same time (e.g. both in the morning). As always, if you do not feel you can do this activity safely please contact me for a virtual alternative.

Part 3a. Population density

For this section you and your partner will each choose a target focal species. It can be the same species or a different species. Choose one that is relatively common so that you will actually get some data! Before you start data collection please answer the questions below.

Question 2. What is your focal species? Which habitats will you be censusing? Which habitat do you predict will have higher population density of your focal species, and why? How long will your census route be?

Part 3b. Estimating biodiversity

In a perfect world we would be able to inventory all of the animals (insects, reptiles, birds, mammals) in an area. But oftentimes, this is not possible due to the immense diversity of species (particularly in the tropics). Depending on the research question or conservation goals is common to focus on a subset of animals (e.g. birds, mammals etc.) to estimate diversity. For this exercise you will compare biodiversity estimates from four different census routes- the two sites that you chose and the two sites that your partner chose. Before we begin please answer the questions below.

Question 3. What are the four habitat types you and your partner will be censusing? Which habitat do you predict will have the highest biodiversity?

Which do you predict will have the lowest? Will you focus on birds, mammals or a combination?

Part 4. Collecting your data

To be good scientists make sure to note the date, time, weather, the total distance (in meters) of your census route, and any other notes in your notebook. You can keep track of the distance you walked either by using an app on your phone or by counting your steps. For each animal encountered along the transect note the following information:

Species: I assume for your focal species you will be able to identify to the species. If you encounter birds or animals that you cannot identify to species assign them to their own categories (e.g. small black bird, large black bird).

Distance (in meters) along your census line: Did you see the animal at the start of your census? Then the distance would be 0 or 1 meters. If you saw it at the end it would be 500 meters.

Perpendicular distance from you census line: How far (in meters) was the animal away from you when you detected it? You can estimate this by counting the number of steps away the animal is from your census line.

Method of detection: Did you see the animal, hear the animal, step on it?

Question 4. What were your general observations? Do you feel confident that you were able to detect all the animals on the census route? What could have potentially biased your results?

Question 5. How could inter-individual differences in behavior potentially influence the results of your survey?

Part 5. Data analysis

We will be analyzing the data during the computer lab. Be sure that your data are entered into the data sheet following the correct format. You can download the datasheet below.

Field Lab 6: Design a camera trap study in Serengeti National Park.

Part 1. Background

The camera trap data that we will use for this field lab comes from the 'Snapshot Serengeti' project (http://lila.science/datasets/snapshot-serengeti (Links to an external site.)). The scientists set out out 225 cameras within a 1,125 km2 area (see map to the right). The cameras have been deployed continuously since 2010, and they recruited citizen scientists to help with classifying images (https://www.zooniverse.org/projects/zooniverse/snapshot-serengeti (Links to an external site.)).

Part 2. Data format and size

The camera trap photos along with the annotations (i.e. image classifications) are available for anyone to access. The original dataset includes over 2.65M sequences of camera trap images (totaling 7.1M images) from ten field seasons. As you can imagine that is a lot of data! For this lab we will focus on data collected over the course of three seasons. I have done a substantial amount of preprocessing the data so that you can access it for this lab. The table below includes the number of images available for each of the species. Note that annotations are tied to images, but are only reliable at the sequence level. For example, there are some sequences in which two of three images contain a lion, but the third is empty (lions, it turns out, walk away sometimes), but all three images would be annotated as "lion". In the R code to download the photos it will pull out only the first photo in the sequence, but the table below includes all of the photographs in the sequence, so the actual sample size may vary slightly.

Part 3. Temporal niche separation using camera trap data

For this lab you are going to choose two animals from the camera trap dataset. It could be a pair of animals that are predator and prey or two potential competitors. Once you decide on the two animals you want to compare you will make some predictions about how you think they will differ in their activity patterns. You should choose different animals from your partner (so please check with them first).

Question 1. What are the two species you chose to compare? What are your predictions regarding their activity patterns?

Part 4. Data collection

We will be using RStudio to access and download 50 photos of each of the two species you indicated above. See Part I of Computer Lab 6. Analyzing camera trap data for detailed instructions on how to collect your data.

Part 5. Follow-up questions

Question 2. What did you notice during data collection? Does it seem that the data were consistent with your predictions?

Question 3. For this lab we are pooling data from different locations and seasons. How could this influence our results?

Question 4. What other factors (apart from time of day) could be influencing the presence or absence of the animals in your camera trap data?

Question 5. If you were to design a followup study, what additional data would you like to collect?

Appendix 1 R script. Data exploration and visualization

```
# Please use the Lab 1 tutorial located here: https://bookdown.org/djc426/behaviouR-R-package-tu

# Hello and welcome to your first R session!
# Remember any lines that include a # symbol will not be read by R,
# but provides information to the user (you!)

# First we load the relevant packages
library(behaviouR)
library(ggpubr)
library(ggfortify)

## Lab 1a. Categorical data
# Here we create a simulated population with four categories (Infant, Juvenile, AdultFemale and A
DeerPopulationDF <- data.frame(DevelopmentStage=c('Infant','Juvenile','AdultFemale','AdultMale'),
                               NumberOfIndividuals=c(15,50,125,200))

# We then print the object so that we can see the output
DeerPopulationDF

# Now we want to plot the data. We will use a simple barplot to start.
ggbarplot(DeerPopulationDF, x='DevelopmentStage', y='NumberOfIndividuals')


## Lab 1b. Categorical and continuous data
# Load the dataset so that we can use it
data('MaleDeerRoarDF')

# We can check the structure of the dataframe by using the command 'head'
```

```r
head(MaleDeerRoarDF)

# Now we will plot the categorical data with standard deviations.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute',
          add = c("mean_sd"),xtickslab.rt = 90)

# Now we will plot the categorical data with colors for each category.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute', fill = 'MaleCategory',
          add = c("mean_sd"),xtickslab.rt = 90)+ theme(legend.position = "none")

# Now we will plot the categorical data with user-specified colors for each category.
ggbarplot(MaleDeerRoarDF, x='MaleCategory', y='RoarsPerMinute', fill = 'MaleCategory',
          palette = c('red','gray','white','black'),
          add = c("mean_se"),xtickslab.rt = 90)+ theme(legend.position = "none")


## Lab 1c. Categorical and continuous data
# We will simulate a dataset for males of different weight and harem size.

# The function below simulates our data. N is the number of individuals,
# CorrelationCoefficient tells us how correlated our data are,
# MaleMeanBodyWeight is the mean body weight of males in our population and
# MaleReproductiveSuccess is the mean number of females in the harem.
MaleRedDeerDF <- CorrelatedDataSimulationFunction(N=100,
                                                  CorrelationCoefficient= 0.45,
                                                  MaleMeanBodyWeight = 125,
                                                  MaleReproductiveSuccess = 3)

# We can check the output
head(MaleRedDeerDF)

# Make a scatterplot of the data.
ggscatter(data=MaleRedDeerDF,x='MaleBodyWeight',y='MaleReproductiveSuccess')

# Make a scatterplot with a trendline.
ggscatter(data=MaleRedDeerDF,x='MaleBodyWeight',y='MaleReproductiveSuccess',
          add='reg.line')

# Create a linear model where MaleBodyWeight is the independent variable
# and ReproductiveSuccess is the dependent variable.
MaleDeerModel <- lm(MaleReproductiveSuccess ~ MaleBodyWeight,data=MaleRedDeerDF)

# We can look at the output of the model
MaleDeerModel
```

```r
# Create a null model and a model with MaleBodyWeight as a predictor.
MaleDeerNull <- lm(MaleReproductiveSuccess ~ 1, data=MaleRedDeerDF)
MaleDeerModel <- lm(MaleReproductiveSuccess ~ MaleBodyWeight,data=MaleRedDeerDF)

# Compare models using AIC.
bbmle::AICctab(MaleDeerModel,MaleDeerNull, weights=T)

## Lab 1d. Multivariate data
# Just as before we load our data
data("DeerSpeciesAcousticFeatures")

# Check the structure
head(DeerSpeciesAcousticFeatures)

# Here is our modified dataset that we will use for PCA
DeerSpeciesAcousticFeatures[,-c(4)]

# Run the PCA using the 'princomp' function
DeerSpeciesAcousticFeaturesPCA <- princomp(DeerSpeciesAcousticFeatures[,-c(4)])

# Plot the results of our PCA.
ggplot2::autoplot(DeerSpeciesAcousticFeaturesPCA, data = DeerSpeciesAcousticFeatures, colour = 'C
         loadings = FALSE)+theme_bw()

# Plot the PCA with arrows indicating which features are important for distinguishing between gro
ggplot2::autoplot(DeerSpeciesAcousticFeaturesPCA, data = DeerSpeciesAcousticFeatures, colour = 'C
         loadings = TRUE, loadings.colour = 'red',
         loadings.label = TRUE,
         loadings.label.size = 5)+theme_bw()
```

Appendix 2 R script. Activity Budgets and Ethograms

```
# Lab 2:  Activity Budgets and Ethograms

# First we load the required packages
library(behaviouR)
library(ggpubr)

# Reminder: any lines that include a # symbol will not be read by R

# Exercise 1: Enter and visualize ethogram data
# This is a toy ethogram with some simulated data
EthogramDF <- data.frame(Behavior=c('Resting','Locomotion','Foraging','Calling','Playing','Other'
                         TimesObserved=c(5,7,3,21,1,0))

# If you run the object 'Ethogram' you should see your table
EthogramDF

# Now we can make a basic barplot to visualize our results
ggbarplot(data=EthogramDF, x='Behavior', y='TimesObserved',fill='grey')

# In the space below I want you to add in your own ethogram categories and create a plot based on
# ethogram you created in Field lab 1.
# Change Category1, Category2 to the categories you used in your ethogram and the numeric values
# your observed values
EthogramDFupdated <- data.frame(Behavior=c('Category1','Category2','Category3','Category4','Categ
                                TimesObserved=c(1,2,3,4,5,6))

# If you run the object you should see your dataframe
EthogramDFupdated
```

```r
# Now we can make a barplot to visualize our results
ggbarplot(data=EthogramDFupdated, x='Behavior', y='TimesObserved',fill='grey')

# Exercise 2.Calculate meerkat activity budgets. Our previous ethograpm examples only
# we observed, but we did not standardize our results in any way. Activity budgets giv
# how much time an animal spends doing each activity; these are generally expressed as

# First we need to import the data
data('MeerkatFocalData')

# Let's check the structure of our data
str(MeerkatFocalData)

# Then we calcuate the total number of sections for each behavior
# There are many different ways that we can summarize our data using R, but we will us
# 'dplyr'

# Here we take the sum for each unique behavior and return it
MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))

# We need to save the output as an R object
MeerkatFocalDataSummary <- MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))


# Now we need to standardize for our total observation time
# We divide by 600 because our video was 10 minutes (or 600 seconds) long
MeerkatFocalDataSummary$ActivityBudget <- MeerkatFocalDataSummary$TotalSeconds/600*100

# Now let's plot our results.
# NOTE: The example here is using fake data.
ggbarplot(data=MeerkatFocalDataSummary,x='BehaviorCode',y='ActivityBudget',fill='grey'

# Now lets see if there were differences between you and your partner
# We need to save the output as an R object
MeerkatFocalDataPartner <- MeerkatFocalData %>%
  dplyr::group_by(BehaviorCode,Partner) %>%
  dplyr::summarise(TotalSeconds = sum(SecondsEngagedinBehavior))

# We divide by 600 because our video was 10 minutes (or 600 seconds) long
# We then multiply by 100 so that we can report in percentages.
MeerkatFocalDataPartner$ActivityBudget <- MeerkatFocalDataPartner$TotalSeconds/600*100
```

```
# Now we compare our data with our partner's
# Note that there are two new lines in the code below
# The fill argument tells R which category or factor to use to color the bars.
# The position argument tells R to place the bars side-by-side.
ggbarplot(data=MeerkatFocalDataPartner,x='BehaviorCode',y='ActivityBudget', fill='Partner',
          position = position_dodge(0.9))

# Exercise 2c. Scan sampling and inter-observer reliability.
# Read in the data
data('MeerkatScanData')

# Check the structure of our data
# If we use 'head' it will return the first few rows of our dataframe
head(MeerkatScanData)

# For this analysis the 'Time' column isn't needed so we can remove it; It is the first column so
MeerkatScanData <-MeerkatScanData[,-c(1)]

# R deals with blanks in dataframes by converting them to 'NA'. This is useful for some cases,
# but we want to convert our blanks to zeros using the following code.
MeerkatScanData[is.na(MeerkatScanData)] <- 0

# Here we take the sum for each unique behavior and return it
MeerkatScanData %>%
  dplyr::group_by(Treatment) %>%
  dplyr::summarise(Vigilant = sum(Vigilant),
            OutOfSight=sum(OutOfSight),
            NotVigilant=sum(NotVigilant))

# We need to save the output as an R object
MeerkatScanDataSummary <- MeerkatScanData %>%
  dplyr::group_by(Treatment,Partner) %>%
  dplyr::summarise(Vigilant = sum(Vigilant),
                   OutOfSight=sum(OutOfSight),
                   NotVigilant=sum(NotVigilant))


MeerkatScanDataSummaryLong <- reshape2::melt(MeerkatScanDataSummary, id.vars=c("Treatment", "Part
head(MeerkatScanDataSummaryLong)

colnames(MeerkatScanDataSummaryLong) <- c('Treatment','Partner','Behavior','NumberMeerkats')


# And now we can visually compare our results with our partners
ggbarplot(data=MeerkatScanDataSummaryLong,x='Treatment',y='NumberMeerkats', fill='Behavior',
```

```
          position = position_dodge(0.9),facet.by = 'Partner')

# Now we want to calculate inter-observer reliablity
# NOTE: that you will want to change the A and B to the names you used in the datashee
Partner1Data <- subset(MeerkatScanData,Partner=='A')
Partner1Data <- Partner1Data[,-c(5)] # We remove the 'Partner' column so that we only

Partner2Data <- subset(MeerkatScanData,Partner=='B')
Partner2Data <- Partner2Data[,-c(5)] # We remove the 'Partner' column so that we only

# Here we calculate reliability for vigilance
VigilantCorrelation <- cor(Partner1Data$Vigilant,Partner2Data$Vigilant)

# Here we calculate reliability for vigilance
NotVigilantCorrelation <- cor(Partner1Data$NotVigilant,Partner2Data$NotVigilant)

# Here we calculate reliability for out of sight
OutOfSightCorrelation <-  cor(Partner1Data$OutOfSight,Partner2Data$OutOfSight)

# Print the results
cbind.data.frame(VigilantCorrelation,NotVigilantCorrelation,OutOfSightCorrelation)
```

Appendix 3 R script. Analyzing acoustic data

```
# First we load the required library
library(behaviouR)
library(ggfortify)
library(ggplot2)

# We need to do some data prep to work through the examples

# First you can check where your working directory is; this is where R will store the files local
getwd()

# Now we will create a file at this location called 'FocalRecordings'
dir.create(file.path('FocalRecordings'), showWarnings = FALSE)

# Now we will load the sound files that were in the R package
githubURL <-'https://github.com/DenaJGibbon/behaviouRdata/raw/master/data/FocalRecordings.rda'
FocalRecordings <- get(load(url(githubURL)))

# Let's check the structure
head(FocalRecordings)

# Now we will save the recordings to the new folder you created
for(a in 1:length(FocalRecordings)){
  FileName <- FocalRecordings[[a]][1][[1]]
  WaveFile <- FocalRecordings[[a]][2][[1]]
  tuneR::writeWave(WaveFile, paste("FocalRecordings/",FileName,sep=''))
}


# Part 1. Loading a sound file and making a spectrogram
# First we want to check which files are in the folder
```

```
# There are five gibbon female recordings and five great argus recordings.
list.files("FocalRecordings")

# Now let's read in one of the files for further investigation. We do that with the
# function 'readWave'. The .wav stands for 'Waveform Audio File' and is a common forma
# scientists use to save their sound files.
GibbonWaveFile <- tuneR::readWave("FocalRecordings/FemaleGibbon_1.wav")

# Now if we run the object it will return some information
GibbonWaveFile

# What we see is that the soundfile is ~13 seconds long, was recorded at a sampling
# rate of 44100 samples per second and has a total of 572054 samples.
# We can check if that makes sense using the following equation (duration* sampling ra
seewave::duration(GibbonWaveFile)* GibbonWaveFile@samp.rate

# Now we can plot the waveform from our sound file using the following code:
seewave::oscillo(GibbonWaveFile)


# We can also zoom in so that we can actually see the shape of the wave.
seewave::oscillo(GibbonWaveFile, from=0.1, to=0.2)

# And if we zoom in again we see the wave shape even more.
seewave::oscillo(GibbonWaveFile, from=0.15, to=0.2)

# A common way for scientists to study animal sounds is through the use of spectrogram
# gibbon call.

SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav")

# There are many different things you can change when creating a spectrogram.
# In the above version of the plot there is a lot of space above the gibbon calls,
# so we will change the frequency range to better visualize the gibbons.
SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav",
                  min.freq = 500,max.freq=2500)

# In the code we are using there is also the option to change the color of the spectro
# adding the following code:
SpectrogramSingle(sound.file ="FocalRecordings/FemaleGibbon_1.wav",
                  min.freq = 500,max.freq=2500,
                  Colors = 'Colors')

# NOTE: Changing the colors doesn't change how we read the spectrograms,
# and different people have different preferences.
```

```r
# Now we can plot all of the spectrograms in our FocalRecordings file at once.
# NOTE: Use the navigation buttons on the plots tab to move back and forth between the different
# We change the plot settings to show four plots at a time
par(mfrow=c(2,2))

#NOTE: If you get an error drag the plot window so that it is larger on the screen!
SpectrogramFunction(input.dir = "FocalRecordings",min.freq = 500,max.freq=2500)

# Question 1. What differences do you notice about gibbon and great argus spectrograms?

# Now that you are done looking at the spectrograms we will clear the space. You can always
# re-run the code to look at the again.
graphics.off()

# Part 2. Visualizing differences in gibbon and great argus calls
# A major question in many different types of behavioral studies is whether there
# are differences between groups. This is also the case for acoustic data, where
# researchers may be interested if there are differences between sexes, populations,
# individuals, or vocalizations emitted in different contexts. Here we will work through
# a simple example where we will investigate differences in gibbon and argus calls.

# The first step that we need to do is called feature extraction. There are many different
# ways that scientists do this, but the overarching idea is that sound data contains too much
# redundant information to be used on a model. Computers just don't have enough power to
# crunch all those numbers, so we need to identify a meaningful set of features that is smaller
# than using the while waveform. This is also the case for image processing.

# We are going to use a feature extraction method called 'Mel-frequency cepstral coefficients'.
# I will not expect you to know any more about them, apart from the fact they are very useful
# for human speech and bioaccoustics applications.

# Here is the function to extract the features
FeatureDataframe <- MFCCFunction(input.dir = "FocalRecordings")

# In this new dataframe each row represents one gibbon or great argus call. Let's check the dimer
dim(FeatureDataframe)

# This shows us that for each of the 10 calls there are 178 features. This is in contrast
# to a sound file. Let's check again to remind ourselves.
GibbonWaveFile

# This sound file is comprised of 572054 numbers. Now let's check the first row of our new dataf
# which contains the new features for the same gibbon call.
FeatureDataframe[1,] # This isolates the first row.
ncol(FeatureDataframe[1,]) # This tells us how many features we have.
```

```
# OK, now we are ready to do some plotting. If you remember from our earlier lab,
# the data structure we have here is multivariate, as each call has multiple values
# associated with it. So we will use the same approach (principal component analysis)
# to visualize our gibbon and great argus calls.

# First we run the principal component analysis
pca_res <- prcomp(FeatureDataframe[,-c(1)], scale. = TRUE)

# Now we visualize our results
ggplot2::autoplot(pca_res, data = FeatureDataframe,
         colour = 'Class')

# What we see in this plot is strong clustering, with the points from each class of ca

*Question 2.* How do we interpret the clustering in this PCA plot?

# Part 3. Soundscapes
# Now we will do something similar to our investigation of the focal recordings, but
# using a soundscape example. Here we have 20-sec recordings taken from two different
# locations (Borneo and Sulawesi) and two different times (06:00 and 18:00). Let's
# check what is in the folder

# Now we will create a file at this location called 'SoundscapeRecordings'
dir.create(file.path('SoundscapeRecordings'), showWarnings = FALSE)

# Now we will load the sound files that were in the R package
githubURL <-'https://github.com/DenaJGibbon/behaviouRdata/raw/master/data/SoundscapeRe

SoundscapeRecordings <- get(load(url(githubURL)))

# Let's check the structure
head(SoundscapeRecordings)

# Now we will save the recordings to the new folder you created
for(a in 1:length(SoundscapeRecordings)){
  FileName <- SoundscapeRecordings[[a]][1][[1]]
  WaveFile <- SoundscapeRecordings[[a]][2][[1]]
  tuneR::writeWave(WaveFile, paste("SoundscapeRecordings/",FileName,sep=''))
}

# List all the files in the folder
list.files("SoundscapeRecordings")

# Now we will create spectrograms for each recordings
par(mfrow=c(2,2))
```

```
#NOTE: If you get an error drag the plot window so that it is larger on the screen!

SpectrogramFunctionSite(input.dir = "SoundscapeRecordings",
                    min.freq = 0,max.freq=20000)

# Now that you are doing looking at the figure we will clear the space. You can always
# re-run the code to look at the again.
graphics.off()

# Now just as above we will do feature extraction of our soundscape recordings. This
# will convert our dataset into a smaller, much more manageable size.
SoundscapeFeatureDataframe <-
  MFCCFunctionSite(input.dir = "SoundscapeRecordings")

# Check the resulting structure of the dataframe to make sure it looks OK.
head(SoundscapeFeatureDataframe)

# Now we visualize our results
pca_res <- prcomp(SoundscapeFeatureDataframe[,-c(1)], scale. = TRUE)
ggplot2::autoplot(pca_res, data = SoundscapeFeatureDataframe,
         colour = 'Class')

# It looks like the Borneo_Night sampling period was much different from the rest!
# Let's look at a single spectrogram from each sampling location and time.
par(mfrow=c(2,2))
SpectrogramSingle(sound.file ="SoundscapeRecordings/Borneo_Night_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'BW',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Borneo_Morning_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'BW',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Sulawesi_Night_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'BW',downsample = FALSE)

SpectrogramSingle(sound.file ="SoundscapeRecordings/Sulawesi_Morning_01.wav",
                  min.freq = 0,max.freq=22000,
                  Colors = 'BW',downsample = FALSE)

graphics.off()

# Question 3. You can listen to example sound files on the Canvas page. Between
# looking at the spectrograms and listening to the sound files what do you think
```

```
# are the main differences between the different locations and times?

# Part 4. Now it is time to analyze the data you collected for this week's field lab.
# Upload your sound files into either the 'MyFocalRecordings' folder or the 'MySoundsc
# folder and run the code below.

# NOTE: Make sure to create the folder and add your sound files!
dir.create(file.path('MyFocalRecordings'), showWarnings = FALSE)

### Part 4a. Your focal recordings
# First lets visualize the spectrograms. You may want to change the frequency settings
# We can tell R to print the spectrograms 2x2 using the code below.
par(mfrow=c(2,2))

# This is the function to create the spectrograms
SpectrogramFunction(input.dir = "MyFocalRecordings",min.freq = 200,max.freq=2000)

# Here is the function to extract the features. Again, based on the frequency range of
MyFeatureDataFrame <- MFCCFunction(input.dir = "MyFocalRecordings",min.freq = 200,max.

#Then we run the principal component analysis
pca_res <- prcomp(MyFeatureDataFrame[,-c(1)], scale. = TRUE)

# Now we visualize our results
ggplot2::autoplot(pca_res, data = MyFeatureDataFrame,
         colour = 'Class')


### Part 4b. Soundscape recordings
dir.create(file.path('MySoundscapeRecordings'), showWarnings = FALSE)

# First lets visualize the spectrograms. You probably don't want to change the frequen
# We can tell R to print the spectrograms 2x2 using the code below
par(mfrow=c(2,2))

# This is the function to create the spectrograms
SpectrogramFunction(input.dir = "MySoundscapeRecordings",min.freq = 200,max.freq=10000

# Then we extract the features, which in this case are the MFCCs.
MySoundscapeFeatureDataframe <-
  MFCCFunctionSite(input.dir = "MySoundscapeRecordings",min.freq = 200,max.freq=10000)

# Check the resulting structure of the dataframe to make sure it looks OK.
dim(MySoundscapeFeatureDataframe)
```

```
# Now we visualize our results
pca_res <- prcomp(MySoundscapeFeatureDataframe[,-c(1)], scale. = TRUE)
ggplot2::autoplot(pca_res, data = MySoundscapeFeatureDataframe,
         colour = 'Class')

#*Question 4*. Do you see evidence of clustering in either your focal recordings or soundscape re
```

Appendix 4 R script. Vigilance behavior

```
# First we need to load the packages
library(behaviouR)
library(ggpubr)

## Part 1: Barnacle goose vigilance
# Then we load in our goose data
data('BarnacleGooseData')

# Note you will want to collect your own data and load into R using read.csv
# BarnacleGooseData <- read.csv('FULL FILE PATH TO YOUR DATASHEET.csv')

### Part 1a: Surveillance behavior
# Scatterplot of total number of 'head up' in our data
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp')+ylab('Surveillance rate')

# Now let's add a trend line to see if there is a relationship between flock size and the total n
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line')+ylab('Surveillance rate')

# Let's see if there were any differences between you and your partner
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line', facet.by = 'Partner',
          cor.coef = T)+ylab('Surveillance rate')

# Let's plot you and your partner's data in different colors.
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TotalHeadsUp',add='reg.line', facet.by = 'Partner',
          color = 'Partner', palette =c('black','blue'),
          cor.coef = T)+ylab('Surveillance rate')
```

```
#*Question 1*: Were there any major differences between you and your partner?

# Now we will do model selection using Akaike information criterion (AIC). First we cr
# and then we create a model with flock size as a predictor of total number of heads u
SurveillanceNullModel <- glm(TotalHeadsUp ~ (1/Partner),family=poisson , data=BarnacleG
SurveillanceModel <- glm(TotalHeadsUp ~ FlockSize + (1/Partner) ,family=poisson, data=B

# Then we compare the models using AIC
bbmle::AICtab(SurveillanceNullModel,SurveillanceModel)

#*Question 2*: If the null model is ranked higher than the model with flock size as a p

### Part 1b: Time vigilant (sec/min)
# Now we will look at the relationship between the duration (calculated as seconds per
# that the geese were vigilant as a function of flock size.

# Scatterplot of time vigilant (sec/min) as a function of group size
ggscatter(data=BarnacleGooseData,
          x='FlockSize',y='TimeSecHeadUp')+ylab('time vigilant (sec/min)')

# Scatterplot of duration of vigilance behavior in our data with a trend line.
ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line')+
  ylab('time vigilant (sec/min)')


# Let's see if there were any differences between you and your partner.
# We will also add the command 'cor.coef = T' which will give us the correlation coeff
# along with an associated p-value.

# NOTE: The data here are simulated so your plots should look different
ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line',
          facet.by = 'Partner',cor.coef = T)+
  ylab('time vigilant (sec/min)')


# Let's plot you and your partner's data in different colors.
ggscatter(data=BarnacleGooseData,x='FlockSize',y='TimeSecHeadUp',add='reg.line',
          facet.by = 'Partner', color='Partner', cor.coef = T,
          palette =c('black','blue'))+
  ylab('time vigilant (sec/min)')

# As before we will create a null model and then a model with flock size as a predicto
# using AIC. We will not use a Poisson distribution here because our outcome variable

# This is our null model
```

```
VigilanceNullModel <- lme4::lmer(TimeSecHeadUp ~ (1|Partner), data=BarnacleGooseData)

# This is our model with flock size as a predictor duration of vigilance
VigilanceModel <- lme4::lmer(TimeSecHeadUp ~ FlockSize + (1|Partner) ,data=BarnacleGooseData)

# Now we compare the models using AIC
bbmle::AICtab(VigilanceNullModel,VigilanceModel)

#*Question 3.* How do you interpret the results of your model selection?

## Part 2: Meerkat data revisted
# Please upload your meerkat scan data to this project and delete the existing datasheet.

data('MeerkatScanData')

# Note you will want to collect your own data and load into R using read.csv
# MeerkatScanData <- read.csv('FULL FILE PATH TO YOUR DATASHEET.csv')

# As before we will turn our NA values to zero
MeerkatScanData[is.na(MeerkatScanData)] <- '0'

# We will remove the time and out of sight columns as we do not need them
MeerkatScanData <- dplyr::select(MeerkatScanData,-c(Time,OutOfSight))

# We need to reformat our data so that we can plot it
MeerkatScanDataSummaryLong <- reshape2::melt(MeerkatScanData, id.vars=c("Treatment", "Partner"))

# Here we add more informative column names
colnames(MeerkatScanDataSummaryLong) <- c('Treatment','Partner',
                                          'BehavioralState','InstancesOfBehavior')

# We need to tell R that our outcome variable is not categorical but numeric
MeerkatScanDataSummaryLong$InstancesOfBehavior <-
  as.numeric(MeerkatScanDataSummaryLong$InstancesOfBehavior)

# Now we plot our data
ggboxplot(MeerkatScanDataSummaryLong,x='Treatment',
          y='InstancesOfBehavior', fill = 'BehavioralState')

#**Question 4.** Based on your inspection of the boxplot, are there any major differnces between

# Now we will test to see if there were differences in vigilance behaviors across treatments?

# First we subset our data so that it only includes the vigilant category
MeerkatScanDataVigilantOnly <- subset(MeerkatScanDataSummaryLong,
```

```
                                    BehavioralState=='Vigilant' )

# R can be picky about the format of data, so we use this command to tell R that treat
MeerkatScanDataVigilantOnly$Treatment <-
  as.factor(MeerkatScanDataVigilantOnly$Treatment)

# Here we are reordering the levels of the factors. For our model selection we are int
# in whether we see differences from the control (no predator) and the predator treatm
# so here we are setting the no predator group as our reference group.
MeerkatScanDataVigilantOnly$Treatment <-
  factor(MeerkatScanDataVigilantOnly$Treatment, levels = c("NoPredator", "AerialPredat
                                                           "TerrestrialPredator"))

# Now as before we will do model selection. Note that because our outcome variable (in
# is in the form of count data we use a poisson distribution.
MeerkatVigilanceNullModel <- glm(InstancesOfBehavior ~ 1, family=poisson, data=MeerkatS
MeerkatVigilanceModel <- glm(InstancesOfBehavior ~ Treatment, family=poisson,data=Meerk

# Now we compare the models using AIC
bbmle::AICtab(MeerkatVigilanceNullModel,MeerkatVigilanceModel)

# Here we will use the summary function to look at the estimates. There is a lot of in
# but we want to focus on the 'Estimate'. In particular we are interested in the estim
# for 'TreatmentAerialPredator' and 'TreatmentTerrestrialPredator'. The estimate is sh
# that these variables have on our outcome (instances of behavior), relative to our co
# Therefore positive estimates indicate that there were more vigilance behaviors in ae
# and terrestrial predator treatments.
summary(MeerkatVigilanceModel)

# A common way to visulize results such as these are coefficient plots. Here we are lo
# effect of 'TreatmentAerialPredator' and 'TreatmentTerrestrialPredator' relative to o
# The reference or group is indicated by the vertical dashed line. So, we can interpre
# coefficients are positive (and the confidence intervals don't overlap zero) that bot
# aerial treatments lead to an increase in vigilant behaviors.
coefplot::coefplot(MeerkatVigilanceModel,intercept=F)


# For reasons that will not go into here, I am not a fan of p-values or null hypothesi
# testing. There is a nice overview if you want to learn more here: https://doi.org/10
# the model selection approach will lead to the same inference as the use of a one-way
# you may be more familiar with.

# Compute the analysis of variance
MeerkatAOV <- aov(InstancesOfBehavior ~ Treatment, data = MeerkatScanDataVigilantOnly)
```

```
# Here this will tell us if there are differences between groups
summary(MeerkatAOV)

# Since the ANOVA test is significant, we can compute Tukey Honest Significant Differences test.
TukeyHSD(MeerkatAOV)

#**Question 5.** Based on your interpretation of the model selection and the coefficient plots we
```

Appendix 5 R script. Estimating Population Density and Biodiversity

```
# First we load the necessary libraries
library(behaviouR)
library(ggpubr)

# Then we read in our data
data('CensusData')

# Part 1. Population density estimation.
# First let's focus on your data
# NOTE you need to change 'A' to the name you used
MyCensusData <- subset(CensusData, Partner=='A')

# Now we will subset the data to focus on your focal species for density estimation
# NOTE in this example we subsetting the data so that we only have the squirrels; you will need t
# the code to subset based on your focal species
MyCensusDataFocal <- subset(MyCensusData, Species=='squirrel')

# Here we will determine the width of our sampling area by creating a histogram of
# perpendicular detection distances.

hist(MyCensusDataFocal$PerpendicularDistance,xlab='Perpendicular distance (D)',
  ylab='Number of observations',main='')

# In this example there is a clear break between 15 and 20 meters, so we will
# only use observations that were within 15 meters.
# NOTE: Your data will look different than this!

# Change the value to the cutoff point indicated in your data
```

```
CutOffPoint <- 15

MyCensusDataFocalAdjusted <- subset(MyCensusDataFocal,PerpendicularDistance <CutOffPoi

# Now we need to subset by each site
# NOTE: You will need to change to the site names you used

MyCensusDataFocalSiteA <- subset(MyCensusDataFocalAdjusted,Site=='SiteA')

MyCensusDataFocalSiteB <- subset(MyCensusDataFocalAdjusted,Site=='SiteB')

# Now we will calculate the population density based on our two surveys
# Change the following to indicate the distance (in meters) of your survey for site A

SiteACensusDistance <- 500

# Now we will calculate the area of our census. The sample area (a) is equal to the le
# transect multiplied by twice the width or a= 2wl. We divide by 1000 to convert our a
SiteACensusArea <- 2*SiteACensusDistance*CutOffPoint/1000

# Now we need to calculate the number of animals using the following code
NumberFocalAnimalsSiteA <- nrow(MyCensusDataFocalSiteA)

PopulationDensitySiteA <-NumberFocalAnimalsSiteA/SiteACensusArea
PopulationDensitySiteA


SiteBCensusDistance <- 500

# Now we will calculate the area of our census. The sample area (a) is equal to the le
# transect multiplied by twice the width or a= 2wl
SiteBCensusArea <- 2*SiteBCensusDistance*CutOffPoint/1000

# Now we need to calculate the number of animals using the following code
NumberFocalAnimalsSiteB <- nrow(MyCensusDataFocalSiteB)

PopulationDensitySiteB <- NumberFocalAnimalsSiteB/SiteBCensusArea
PopulationDensitySiteB

# Now we can compare population density using the following code. Was the population d
PopulationDensitySiteA > PopulationDensitySiteB

# Was the population density of site B higher than site A?
PopulationDensitySiteB > PopulationDensitySiteA
```

```
# **Question 1**. What were the population density estimates (reported as number of individuals p
#Do your results of the population density estimates match your predictions? Why or why not?

# Part 2. Comparing biodiversity

# First, we will estimate the alpha diversity, or the diversity within a particular area or ecosy
# The alpha diversity is simply the number of different species present at each site
# Here we subset by partner and site A (you may need to change these!)
MyCensusDataSiteA <- subset(MyCensusData,Partner=='A' & Site=='SiteA')

# What were the unique species present?
unique(MyCensusDataSiteA$Species)

# How many unique species were there?
SiteANumberSpecies <- length(unique(MyCensusDataSiteA$Species))
SiteANumberSpecies

# Here we subset by partner and site B (you may need to change these!)
MyCensusDataSiteB <- subset(MyCensusData,Partner=='A' & Site=='SiteB')

# What were the unique species present?
unique(MyCensusDataSiteB$Species)

# How many unique species were there?
SiteBNumberSpecies <- length(unique(MyCensusDataSiteB$Species))
SiteBNumberSpecies

# What was the species richness for both sites sampled?
unique(c(MyCensusDataSiteA$Species,MyCensusDataSiteB$Species))

# **Question 2**. Which of your sites had higher species richness (i.e. number of species)?

# Now we will estimate beta diversity, which estimates changes in species diversity between
# ecosystems or along environmental gradients

# Beta diversity = 2c / S1 + S2

# This code tells us which species both sites have in common
intersect(unique(MyCensusDataSiteA$Species),unique(MyCensusDataSiteB$Species))

# Now we calculate the number of species in common
SpeciesInCommonBothSites <- length(intersect(unique(MyCensusDataSiteA$Species),unique(MyCensusDat
)
SpeciesInCommonBothSites
```

```
# Scientists are often interested in the community similarity between sites. For this
# we will calculate Sørenson's index; a value of 1 means exactly the same number of sp
# a value of 0 means no overlap
2*SpeciesInCommonBothSites/ (SiteANumberSpecies+SiteBNumberSpecies)

# Now we will use your partner's data
MyPartnersCensusDataSiteA <- subset(CensusData,Partner=='B' & Site=='SiteC')
unique(MyPartnersCensusDataSiteA$Species)
length(unique(MyPartnersCensusDataSiteA$Species))
SiteANumberSpecies <- length(unique(MyPartnersCensusDataSiteA$Species))
SiteANumberSpecies

MyPartnersCensusDataSiteB <- subset(CensusData,Partner=='B' & Site=='SiteD')
unique(MyPartnersCensusDataSiteB$Species)
length(unique(MyPartnersCensusDataSiteB$Species))
SiteBNumberSpecies <- length(unique(MyPartnersCensusDataSiteB$Species))
SiteBNumberSpecies

# What was the species richness for both sites sampled?
unique(c(MyPartnersCensusDataSiteA$Species,MyPartnersCensusDataSiteB$Species))


# **Question 3**. How did the alpha diversity of each of your sites compare with that


# Gamma diversity is total species over a large area or region; there are many differe
# that this can be measured. The way we will do it is a bit of an oversimplification b
# comparing the number of species seen during the census at both locations.

# First we subset by the first partner
MyCensusData <- subset(CensusData, Partner=='A')
unique(MyCensusData$Species)

# Then we plot the results
gghistogram(data=MyCensusData, x='Species',stat="count")

# Now we subset by the second partner
MyPartnersCensusData <- subset(CensusData, Partner=='B')
unique(MyPartnersCensusData$Species)

# Then we plot the results
gghistogram(data=MyPartnersCensusData, x='Species',stat="count")


# Now we can plot all the data together, separated by partner
```

```
gghistogram(data=CensusData, x='Species',stat="count",
            facet.by = 'Partner',fill='Partner',x.text.angle =90)+xlab('Species')+ylab('Number of

# **Question 4**. How did the gamma diversity of your site compare with that of your partners?

# There are special packages in R that can measure different diversity indices, as this is a tool
# many ecologists use. We will use the 'vegan' package.
library(vegan)

# First we need to convert our data into a table thate can be used to calculate the indices.
BiodiversityTable <- table(CensusData$Site,CensusData$Species)

#Simpson's Index (D) measures the probability that two individuals randomly
# selected from a sample will belong to the same species (or some category other
# than species). With this index, 1 represents infinite diversity and 0 means no diversity.
H <- diversity(BiodiversityTable,index="simpson")
H

# Species evenness refers to how close in numbers each species in an environment is. We can calcu
# that using the following code. The value is constrained between 0 and 1, with communities that
# that have a more even representation of species having values closer to 1.
J <- H/log(specnumber(BiodiversityTable))
J

# **Question 5**. Which of the four sites was most diverse? Which was the most even? Why is it
# important to consider diversity and evenness when studying biodiversity?
```

Appendix 6 R script. Analyzing camera trap data

```
# First we load the required libraries
library(behaviouR)
library(ggpubr)

# We will be taking random subsets of the camera trap photos, and we
# use 'set.seed' to make sure that our results are reproducible.
set.seed(2210)

# We can use the following code to query the database by season to see
# which photos are available. The available seasons are 1-4, 6-8,10,11.
# This function will return a table with all of the available camera
# trap photos for a particular season.
CameraTrapAnnotations(season = 1)

# Now we will use another function to download the camera trap photos
# and save them locally to our computer.  You can change the season
CombinedAnimalDF <- CameraTrapDataAccess(urlpath = "https://lilablobssc.blob.core.windows.net/sna
                                         season = list(1, 2), AnimalID = "zebra", NumPhotos = 5,

# Here we isolate only the columns from the dataframe that we need.
CombinedAnimalDF <- CombinedAnimalDF[, c("category_id", "season", "location",
                                         "filename")]

# Here is the function to view photos and annotate data.
CombinedAnimalDF_TimeAdded <- CameraTrapDataCollection(inputfile = CombinedAnimalDF,
                                                       rowstart = 8, dataframe.cont = FALSE, opti

CombinedAnimalDF_TimeAdded

# Optional: You can save your spreasheet using the following code:
```

```
# write.csv(CombinedAnimalDF_TimeAdded,'CombinedAnimalDF_TimeAdded.csv)

# You can then load it back in later using the following code.
# CombinedDFTimes <- read.csv('CombinedAnimalDF_TimeAdded.csv')

# Now we load your data sheet with the times added NOTE: we are using
# sample data here; you will want to use your own data
data("CombinedDFTimes")

# CombinedDFTimes <- CombinedAnimalDF_TimeAdded

# Let's check the structure
head(CombinedDFTimes)

# Now we can make a density plot that will show the distribution of
# camera trap photos that were taken over 24-hours. We add the fill =
# 'category_id' so that we show different distributions for each
# animal.
ggdensity(data = CombinedDFTimes, x = "Time", fill = "category_id") + xlab("Time (24 h
  ylab("Density")

# Question 1. What do you notice about the overlap of the two density
# curves? Does it look like there is temporal niche partitioning?

# Now we can calculate an overlap coefficient which can be used to
# investigate potential competitive and interaction possibilities
# between species.The value ranges from ranges from 0 (no overlap) to 1
# (complete overlap).

# First we subset our data to focus on the first animal in our dataset
FirstAnimal <- subset(CombinedDFTimes, category_id == "zebra")

# Then we subset our data to focus on the second animal in our dataset
SecondAnimal <- subset(CombinedDFTimes, category_id == "wildebeest")

# Now we use the overlap function to calculate the overlap coefficient
bayestestR::overlap(FirstAnimal$Time, SecondAnimal$Time)

# Question 2. How do you interpret the overlap coefficient for your
# data?

data("CombinedPartnerDFTimes")

# Now we will read in our partners data. NOTE that you need to make
# sure the name of the .csv file is exactly as shown below.
```

```r
#CombinedPartnerDFTimes <- read.csv("CombinedAnimalDF_TimeAddedPartner.csv")

# Let's check the structure
head(CombinedPartnerDFTimes)

# Now we can make a density plot for our partner's data.
ggdensity(data = CombinedPartnerDFTimes, x = "Time", fill = "category_id") +
  xlab("Time (24 hrs)") + ylab("Density")

# Question 3. What do you notice about the overlap of the two density
# curves for your partner's data? Does it look like there is temporal
# niche partitioning?

# Now we can calculate an overlap coefficient which can be used to
# investigate potential competitive and interaction possibilities
# between species.The value ranges from ranges from 0 (no overlap) to 1
# (complete overlap).

# First we subset our data to focus on the first animal in your
# partner's data set
FirstAnimalPartner <- subset(CombinedPartnerDFTimes, category_id == "cheetah")

# Then we subset our data to focus on the second animal in our dataset
SecondAnimalPartner <- subset(CombinedPartnerDFTimes, category_id == "leopard")

# Now we use the overlap function to calculate the overlap coefficient
bayestestR::overlap(FirstAnimalPartner$Time, SecondAnimalPartner$Time)

# Question 4. What is the overlap coefficient for your partner's data?
# How do you interpret this?

# Now we will combine your data with your partners
AllDataCombined <- rbind.data.frame(CombinedDFTimes, CombinedPartnerDFTimes)

# And we plot the diel activity patterns for all four species
ggdensity(data = AllDataCombined, x = "Time", fill = "category_id", facet.by = "category_id") +
  xlab("Time (24 hrs)") + ylab("Density")

# Question 5. Based on the activity patterns and your understanding of
# the Serengeti food web how do you interpret these results? Is there
# evidence of temporal niche partitioning among potential competitors?
# What about interactions between potential predators and prey?
```