R Crash Course

First thing you should always do before working in R: Clear your workspace! This ensures that your R environment is clean and has no objects or functions defined during previous coding sessions. Think of this as cleaning your whiteboard before you start coding.

```
rm(list = ls())
```

Read in your data

Of course, if you want to use R to analyze your data, it's important to read that data into R! To do so, there's a simple function, read.csv(), that allows you to easily bring your .csv into the R environment. You'll need to know the location of your file to read it in successfully.

```
batdat <- read.csv("bat_data.csv")
```

Once you've read this file in, it should be stored as a dataframe object called "bat.dat"

Load and install packages

R packages are the "tookkits" of the R computing world. They offer users unique tools suited to whatever it is they want to do, from cleaning data, to building beautiful figures, to creating apps, to analyzing biological data. In order to use these toolboxes, you will need to install and then load them into the R environment.

```
#install.packages("ggplot2")
library(ggplot2)
```

Confused? Get help!

The R package you just installed has many tools, called "functions", that allow you to make plots and figures in R. But what should you do if you're not sure how to use one of these functions? One of the most useful tools in the R environment is the '?' tool. When you are unsure what a specific function does, you can ask R! For example, if I am not sure what the function unique() does, I can run the below line of code and R will tell me:

```
?unique
```

Examine your data

One of the most important things to do is, of course, to examine your data. Being familiar with the shape and contents of your dataset is essential, and R has many built-in functions that can help you do that!

```
View(batdat)
```

```
## Warning in system2("/usr/bin/otool", c("-L", shQuote(DSO)), stdout = TRUE):
## running command ''/usr/bin/otool' -L '/Library/Frameworks/R.framework/Resources/
## modules/R_de.so'' had status 1
```

```
unique(batdat$species) # Tells you all the unique forms of the data in the specified col
umn ($species)
```

```
## [1] "MYSE"      "MYLU"      "PESU"      "EPFU"      "SUBSTRATE"
```

```
str(batdat) # Tells you what kind of data is contained in your dataframe
```

```
## 'data.frame':    670 obs. of  11 variables:
##  $ swab_id  : chr  "KL15WI0002" "KL15WI0003" "KL15WI0004" "KL15WI0005" ...
##  $ gd       : int  1 1 0 1 1 1 1 1 0 1 ...
##  $ gdL      : num  7.56e-05 4.79e-01 NA 5.51e-06 3.56e-05 ...
##  $ swab_type: chr  "BAT" "BAT" "BAT" "BAT" ...
##  $ state    : chr  "WI" "WI" "WI" "WI" ...
##  $ site     : chr  "HORSESHOE BAY" "HORSESHOE BAY" "HORSESHOE BAY" "HORSESHOE BAY"
...
##  $ date     : chr  "2/27/15" "2/27/15" "2/27/15" "2/27/15" ...
##  $ species  : chr  "MYSE" "MYLU" "MYLU" "MYLU" ...
##  $ temp     : num  NA NA NA NA NA NA NA NA NA NA ...
##  $ country  : chr  "u.s." "u.s." "u.s." "u.s." ...
##  $ count    : int  3 1110 1110 1110 1110 1110 1110 1110 1110 1110 ...
```

```
head(batdat) # Returns the first several rows of data in your dataframe
```

```
##      swab_id gd        gdL swab_type state          site    date species temp
## 1 KL15WI0002  1 0.00007560       BAT    WI HORSESHOE BAY 2/27/15    MYSE   NA
## 2 KL15WI0003  1 0.47879100       BAT    WI HORSESHOE BAY 2/27/15    MYLU   NA
## 3 KL15WI0004  0         NA       BAT    WI HORSESHOE BAY 2/27/15    MYLU   NA
## 4 KL15WI0005  1 0.00000551       BAT    WI HORSESHOE BAY 2/27/15    MYLU   NA
## 5 KL15WI0006  1 0.00003560       BAT    WI HORSESHOE BAY 2/27/15    MYLU   NA
## 6 KL15WI0007  1 0.00003160       BAT    WI HORSESHOE BAY 2/27/15    MYLU   NA
##   country count
## 1    u.s.     3
## 2    u.s.  1110
## 3    u.s.  1110
## 4    u.s.  1110
## 5    u.s.  1110
## 6    u.s.  1110
```

```
tail(batdat) # Returns the last several rows of data in your dataframe
```

```
##        swab_id gd         gdL swab_type state     site    date   species temp
## 665 KL17WI5460  1 0.000725719     UNDER    WI ST. JOHN 3/12/17 SUBSTRATE  5.3
## 666 KL17WI5461  1 0.000033600      NEAR    WI ST. JOHN 3/12/17 SUBSTRATE  5.3
## 667 KL17WI5462  1 0.001922160     UNDER    WI ST. JOHN 3/12/17 SUBSTRATE  4.2
## 668 KL17WI5463  1 0.000076600      NEAR    WI ST. JOHN 3/12/17 SUBSTRATE  4.2
## 669 KL17WI5464  1 0.010103002     UNDER    WI ST. JOHN 3/12/17 SUBSTRATE  4.9
## 670 KL17WI5465  1 0.000495104      NEAR    WI ST. JOHN 3/12/17 SUBSTRATE  4.9
##     country count
## 665    u.s.    NA
## 666    u.s.    NA
## 667    u.s.    NA
## 668    u.s.    NA
## 669    u.s.    NA
## 670    u.s.    NA
```

```
dim(batdat) # Tells you the size of your dataframe
```

```
## [1] 670  11
```

```
names(batdat) # Tells you the column names in your dataframe
```

```
##  [1] "swab_id"   "gd"        "gdL"       "swab_type" "state"     "site"
##  [7] "date"      "species"   "temp"      "country"   "count"
```

```
nrow(batdat) # Returns the number of rows in your dataframe
```

```
## [1] 670
```

```
ncol(batdat) # Returns the number of columns in your dataframe.
```

```
## [1] 11
```

Add a column to your dataframe

Often times, we want to add additional data to the dataframe we have read into the R environment. This is fairly straightforward, as long as we remember some simple notation. The money symbol ($) is how we tell R exactly which column we are referring to in a dataframe, and parentheses following a word always refers to a function. Below, I will add a column to the batdat dataframe in which I log-transform my fungal load column, gd:

```
batdat$log.loads<-log10(batdat$gdL)
```

Subset data

Another very useful thing to do in R is subsetting data. This allows us to filter down a dataframe to only look at the parts we are interested in. For example, in the batdat dataframe, what if we wanted to make a whole new dataframe that just contains the data from Myotis septentrionalis?

```
MYSE.dat = subset(batdat, species=="MYSE") #a factor/character, so need == and quotes
head(MYSE.dat)
```

```
##        swab_id gd         gdL swab_type state          site    date species
## 1   KL15WI0002  1 0.000075600       BAT    WI HORSESHOE BAY 2/27/15    MYSE
## 15  KL15WI0016  0          NA       BAT    WI HORSESHOE BAY 2/27/15    MYSE
## 16  KL15WI0017  0          NA       BAT    WI HORSESHOE BAY 2/27/15    MYSE
## 21  KL15WI0022  1 0.000119813       BAT    WI HORSESHOE BAY 2/27/15    MYSE
## 62  KL15WI0154  0          NA       BAT    WI    BEAR CREEK  3/4/15    MYSE
## 135 KL15WI0608  1 0.002703159       BAT    WI HORSESHOE BAY 11/7/15    MYSE
##     temp country count log.loads
## 1     NA    u.s.     3 -4.121478
## 15    NA    u.s.     3        NA
## 16    NA    u.s.     3        NA
## 21    NA    u.s.     3 -3.921496
## 62   4.8    u.s.     0        NA
## 135  9.7    u.s.     1 -2.568128
```

```
dim(MYSE.dat)
```

```
## [1] 12 12
```

As you can see, this new dataframe only contains 12 observations from Myotis septentrionalis. What if we wanted to filter numerical data? For example, what if we only wanted data from bats that were roosting at 6 degrees celsius or warmer?

```
warm.temps = subset(batdat, temp>6) #a number, so no quotes
head(warm.temps)
```

```
##       swab_id gd         gdL swab_type state       site    date species temp
## 45 KL15WI0137  0          NA       BAT    WI BEAR CREEK  3/4/15    EPFU  6.1
## 48 KL15WI0140  0          NA       BAT    WI BEAR CREEK  3/4/15    EPFU  6.3
## 90 KL15WI0423  0          NA       BAT    WI   ST. JOHN 3/11/15    PESU  7.2
## 91 KL15WI0424  0          NA       BAT    WI   ST. JOHN 3/11/15    PESU  7.2
## 92 KL15WI0425  0          NA       BAT    WI   ST. JOHN 3/11/15    PESU  7.6
## 93 KL15WI0426  1 0.000101664       BAT    WI   ST. JOHN 3/11/15    MYLU  6.5
##    country count log.loads
## 45    u.s.     9        NA
## 48    u.s.     9        NA
## 90    u.s.    12        NA
## 91    u.s.    12        NA
## 92    u.s.    12        NA
## 93    u.s.    82 -3.992833
```

```
dim(warm.temps)
```

```
## [1] 258  12
```

As you can see here, this new dataframe contains 258 observations, only from bats that were found roosting at 6C or warmer. Often times, you will see code that looks a little different, but accomplishes the same thing by using brackets [ ]. Brackets are a way of telling R exactly which part of a dataframe you are referring to. We can subset our batdat dataframe in the same way we did above, but using the bracket system:

```
MYSE.dat = batdat[batdat$species=="MYSE",]
warm.temps = batdat[batdat$temp>6,]
```

Summarize data using aggregate()

The function aggregate() is an extremely useful function to continue exploring your data. It allows you to summarize your dataset by applying a function to subsets of data individually. For example, say that I wanted to know what the mean log fungal load (batdat$log.loads) was on my bats for each species. The aggregate() function allows me to apply the mean() function (finds the average of data) to find a separate average value for each species:

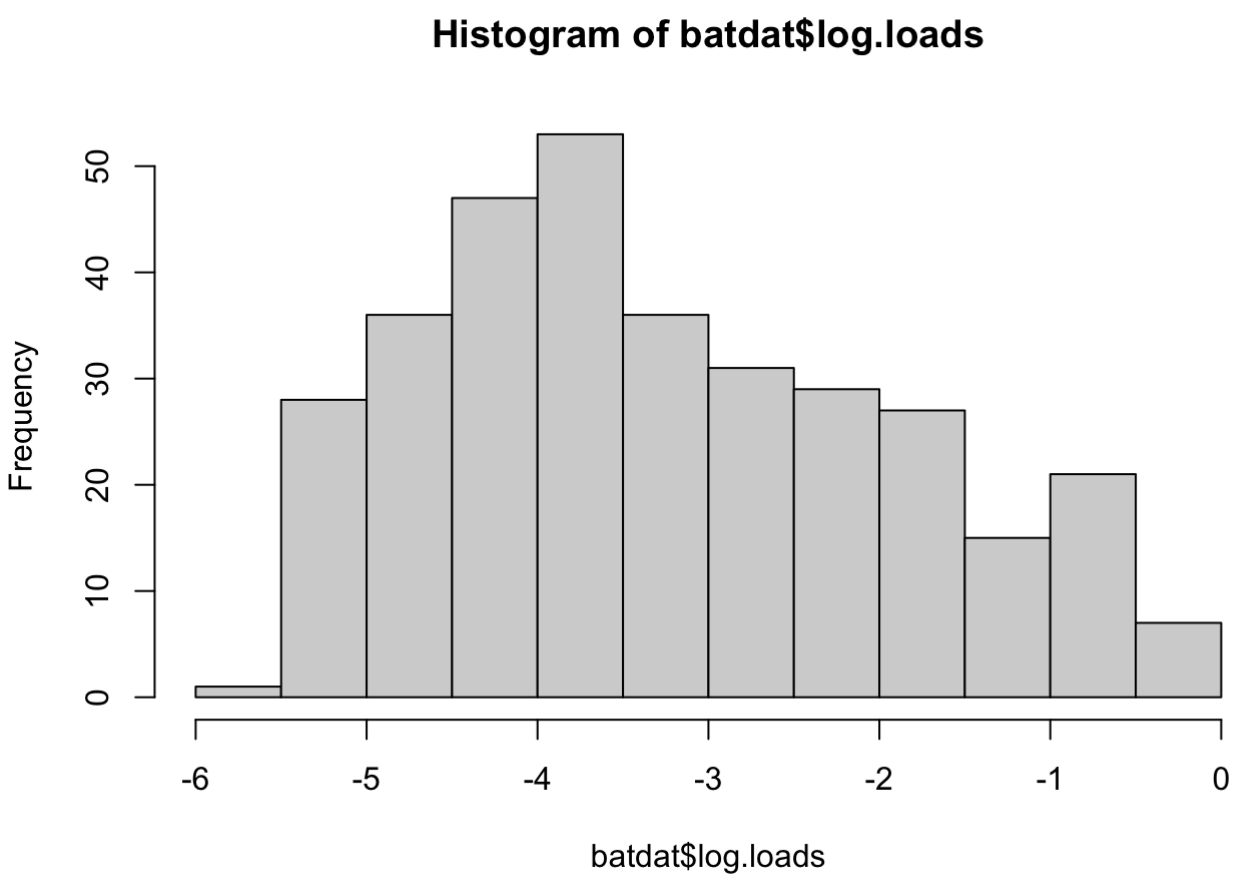
```
bat.summary = aggregate(log.loads~species, FUN=mean, data = batdat) #aggregate data usin
g the mean to give fungal loads by species
bat.summary
```

```
##     species log.loads
## 1      EPFU -3.642464
## 2      MYLU -3.026398
## 3      MYSE -3.688292
## 4      PESU -2.039707
## 5 SUBSTRATE -4.110905
```

Make a histogram of a column in your data

Another useful function to visualize the shape of your data is hist(), which will make a histogram with the data that you feed it. In the example below, we will make a histogram of the raw log fungal loads (batdat$log.loads) contained in our batdat dataframe:

```
hist(batdat$log.loads)
```



Write a .csv file

We have created several new dataframes from our original batdat dataframe. If we want to save these new dataframes as a .csv file, the function write.csv() will come in handy. In the example below, I will write the MYSE.dat dataframe as a new .csv file:

```
write.csv(MYSE.dat, "MYSE_dat.csv", row.names = F)
```