Using R in field sampling -- BIO 104

Linda Forrester

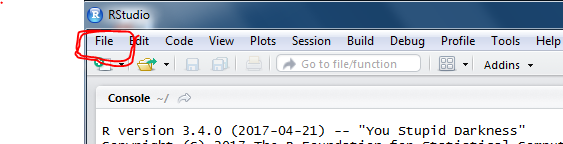
## Starting with the R Software

You will be using **R** for data analysis this semester. Here are some great reasons to use **R**: <http://www.datacarpentry.org/R-ecology-lesson/00-before-we-start.html>

**Open RStudio.** The RStudio program is on the desktop of the lab computers. See [instructions](Rsetup.html) for accessing R via the web or downloading to your computer.

However, to make it easier for starting, we have an R Studio server. **You can access R Studio at** [**https://celsrs.uri.edu/**](https://celsrs.uri.edu/)

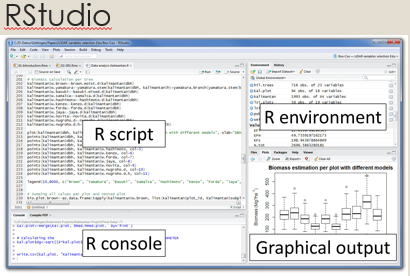
You will get your login information from your TA. Password is the same as username.

Open RStudio and make a new project. -- Select from top left corner: **File - New Project...** **New Directory** and **Empty Project**

\*\*Directory name: LindaForrester\*\* (use your OWN name :)  
 Leave the Subdirectory as the default: ~

click on Console bar to Open the Script window 

Making a new project will give you four windows open in R: a *script* to help you keep track of the commands you are entering, a *console* where those commands are run, an *environment* that show what data you have, and a *file* window. 

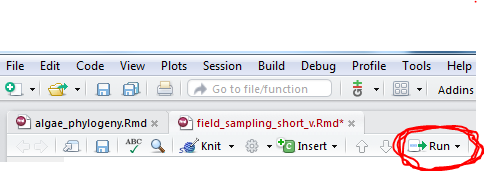
## How to get your data into R so you can use it

Last week you calculated the diversity index for each plot within each of four site types. A diversity index allows you to understand the variation in diversity among site types. The following lines of code will allow you to get your data (the class data about the diversity of each plot) into R directly from google sheets.

**Copy and paste the lines of code (from the grey box below) into the SCRIPT window of RStudio.** Read the following explanation so you understand what these lines of code mean.

* First you must make a necessary "package" accessible to R. (***Packages are like books from the library; packages are pre-written bundles of code that can perform the tasks we want.***) Make a package accessible to your R script by typing library 'gsheet'.
* Assign the web address of the data (the "https" address) to a variable (here our variable is called "url" = "uniform resource locator" = an address on the web)  
  \*\* NOTE: the <- symbols make an arrow. \*\* You name the variable on the left of the arrow. You describe the variable on the right of the arrow.
* Assign the data itself to a variable (in this case we named it diversitydata)
* gsheet2tbl is a command that gets the data from the website so that you can access it in RStudio

library(gsheet)  
url<-'https://docs.google.com/spreadsheets/d/1\_FtL2E935fvANST8JMP50QVQzIVoaZgXyNpLZnVRW-k'  
diversitydata <- gsheet2tbl(url)

**Run these lines using the run button** (top left of the SCRIPT window in RStudio) 

* you will need to run each line separately (have the cursor on each line and click RUN on every line)
* the line was run correctly if you get the prompt (>) in the console window at the bottom of RStudio
* view your data by clicking on the item in the Data window on the right side (my data is called diversitydata - yours may be named something else) - this will open a new tab above so you can look at the data

## Keep track of what your commands do

As you work you are typing commands that you probably won't remember next week. **Use the # symbol to explain your lines of code.** For example, rather than the plain code above your script might look like the following:

#load the package that allows use to access google sheets   
library(gsheet)  
  
#assign the website address for the data to a variable using an arrow <-  
# "variable" <- "website address for data"  
url<-'https://docs.google.com/spreadsheets/d/1\_FtL2E935fvANST8JMP50QVQzIVoaZgXyNpLZnVRW-k'  
  
#load the data from google sheets  
# "varible" <- "command to work on" (data from last line)  
diversitydata <- gsheet2tbl(url)

**Use the # symbol to add comments to your script as you enter the commands below**

## Plot results

### First, load the package "ggplot2" of info you need

To compare the diversity values among site types we need to plot the values for each site type.

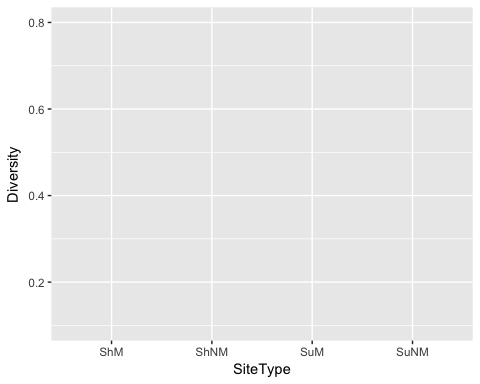
* load the library that allows us to the plotting commands (copy, paste and run lines in grey boxes. Include "# comments".) (***Remember: Packages are like books from the library; packages are pre-written bundles of code that can perform the tasks we want.***) Make a package accessible to your R script by typing library(packageName).

library(ggplot2)

### Second, create the base layer of your plot

* using the command ggplot
* enter the name of the variable containing your data
* aes specifies the "aesthetics" of the plot (i.e. how it looks), including the x and y variables
* specify the variable names based on viewing the data
* run this plot command

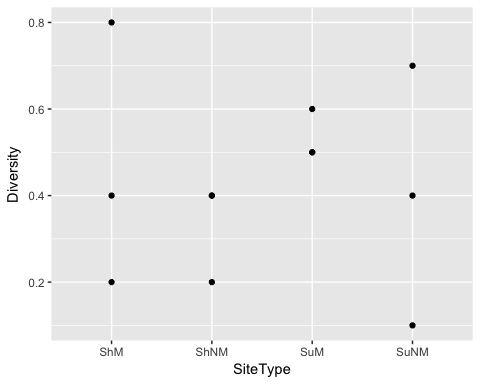
ggplot(diversitydata, aes(x=SiteType,y=Diversity))



### Third, add points to your plot

* now that you have set up your plot add your points using geom\_point()
* put a + between the ggplot command and geom\_point()

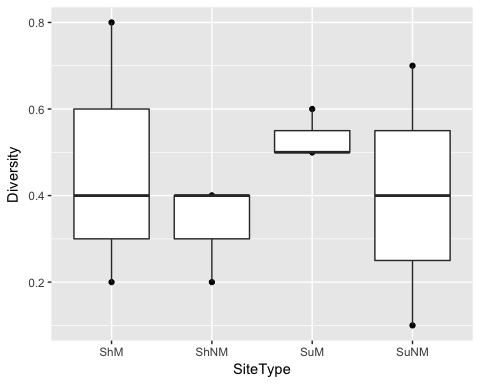
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
 geom\_point()



### Fourth, add a boxplot on top of your plot

* a boxplot summarizes your data by plotting
* the median (middle) value as a horizontal line
* the middle half of the data as a box
* the remaining top and bottom quarters of the data as lines above and below the box
* add a boxplot using geom\_boxplot()
* put a + between the geom\_point() command and geom\_boxplot()

ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
 geom\_point()+  
 geom\_boxplot()



* in order to see both the box and the points plot the box first then the point layer on top
* note: the order of the plotting instructions specifies how the plot is made - i.e. first the base, then the points, then the box covering the points
* plot just the boxplot as this is the typical way to visualize the data

### SAVE your work...

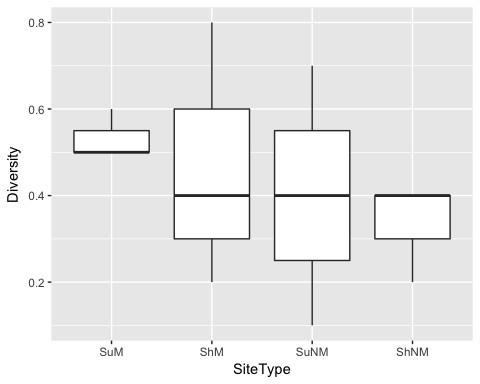
Save your script by clicking File Save in the script window. Your work will save to the R-Server that we are using for this lab. You will be able to access your work on this R Server from other computers, as long as you know the web address and your username and password.

### Optional: Reorder data on the X-axis

If you want to make particular comparisons, such as between mowed sites, you might want to have your data in a particular order.

* plot the two mowed sites on the left and the two unmowed sides on the right side of the plot
* scale\_x\_discrete allows you to change all the information on the x axis
* limits specifies the range of data on the x axis - since the data are categorical not continuous (e.g. from 1-10) you put the items in a list

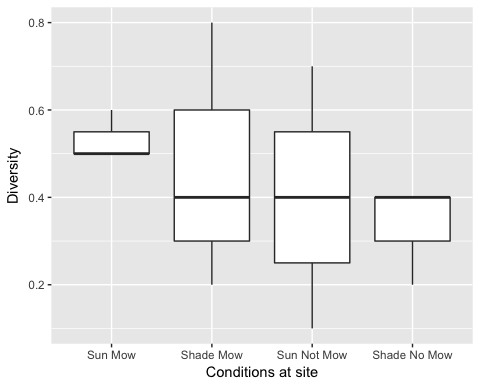
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
 geom\_boxplot()+  
 scale\_x\_discrete(limits=c("SuM", "ShM", "SuNM", "ShNM"))



### Add labels to your plot

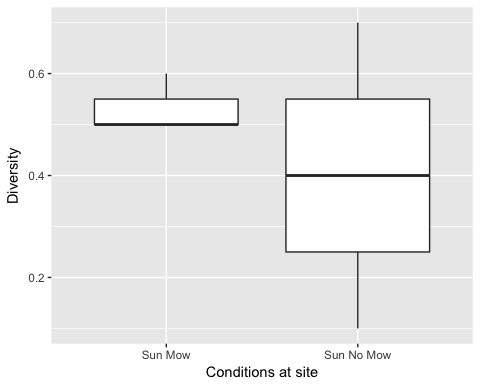
* you can use scale\_x\_discrete to relabel the category names to something more descriptive and to specify the overall x axis name

ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
 geom\_boxplot()+  
 scale\_x\_discrete(limits=c("SuM", "ShM", "SuNM", "ShNM"),   
 labels=c("SuM" = "Sun Mow", "ShM" = "Shade Mow",  
 "SuNM" = "Sun Not Mow", "ShNM" = "Shade No Mow"),   
 name ="Conditions at site")



### If you want to show only data from the two sunny sites, then make a new variable (here called 'SunnyOnly')

SunnyOnly <- subset(diversitydata,SiteType=="SuNM" | SiteType=="SuM")  
ggplot(SunnyOnly, aes(x=SiteType, y=Diversity)) +  
 geom\_boxplot()+  
 scale\_x\_discrete(limits=c("SuM", "SuNM"),  
 labels=c("SuM" = "Sun Mow", "SuNM" = "Sun No Mow"),   
 name ="Conditions at site")



**Turn in this final boxplot of the data**

Save the boxplot by clicking "export" in the "plots" or Graphical output" window. Your work will save to the R-Server that we are using for this lab. You will be able to access your work on this R Server from other computers, as long as you know the web address and your username and password.

## Homework Assignment

* Make a boxplot comparing diversity among terrestrial biomes of the world.
* Data are available at <https://docs.google.com/spreadsheets/d/1KAur7rjQtw3zN1bEQR_xXr0JRS3yn1nnvdrzQu6gWaw>
* Create a new R script
* Load the necessary libraries
* Load your data
* Make a boxplot similar to the Diversity index plot you made in class
* Everyone will be assigned THREE bimomes from 3 countries to plot. In the limits and labels, include only the three biomes you were assigned

**Write a one sentence (or more) statement describing differences in diversity among the biomes you were assigned**