

Using R in field sampling – BIO 104

Starting with the R Software

You will be using **R** for data analysis this semester. Here are some great reasons to use **R**:

- You write down the instructions (code) to do your data analysis so you will remember how it's done
- Instantly rerun your analysis
- Make beautiful plots
- Analyze your data in every possible way

More great information on R is at http://www.datacarpentry.org/R-ecology-lesson/00-before-we-start.html#why_learn_r

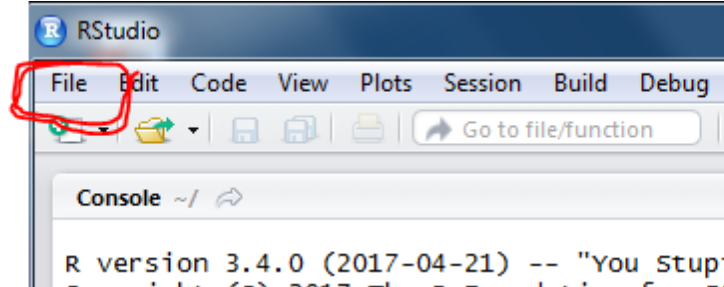
Using RStudio

Log in to the R Studio server at <https://celsrs.uri.edu/> . This will enable you to run the R software to analyze your data.

You will get your login information from your TA. Your 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: FieldSampling
- Select from top left corner:
 - File - New File - R Script
- Leave the Subdirectory as the default: ~



Making a new project and script 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.

Make the tools you need available in R.

The tools you need are found in “packages”. Packages are like books from the library; packages are pre-written bundles of code that can perform the tasks we want.

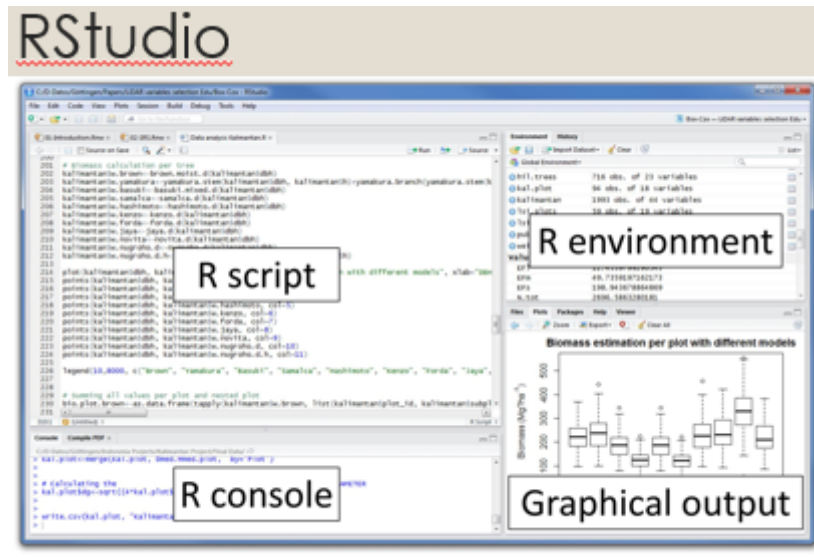


Figure 1:

```
library(gsheet)
```

Type the code (from the grey boxes) into the SCRIPT window of RStudio. Run each line using



the Run button (top right of the SCRIPT window in RStudio)

Tell R where to get your data.

Assign the web address of the data (the “https” address) to a variable (here our variable is called “url” for “uniform resource locator”. This is the term for an address on the web (you can name your variable anything though). You assign a number or word or data to a variable using `<-` (which makes something that looks like an arrow). The variable name goes on the left of the arrow and the information goes on the right.

```
url <- 'https://docs.google.com/spreadsheets/d/15_Oba64U8UsV9CdqrBYNSTTS5YpTARmRrBKDiQwEL1k'
```

Assign the data to a variable

In this case the data is contained in the variable `diversitydata`. The function `gsheet2tbl` gets the data from the website so that you can access it in R.

```
diversitydata <- gsheet2tbl(url)
```

- you will need to run each line separately (put the cursor on each line and click run)

- 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/15_Oba64U8UsV9CdqrBYNSTTS5YpTARmRrBKDiQwELlk'

#load the data from google sheets
# "variable" <- "command to work on" (data from last line)
diversitydata <- gsheets2tbl(url)
```

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

Plot results

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

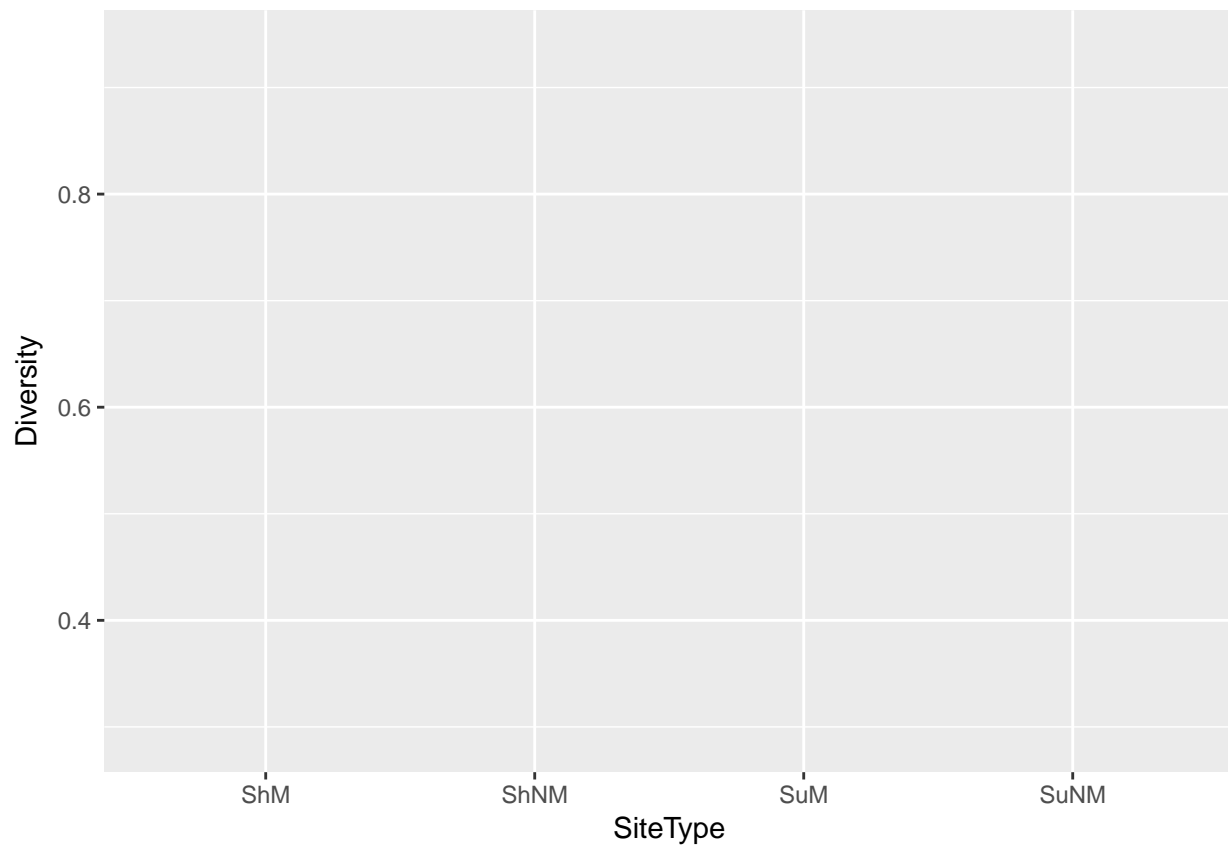
Load the library that allows us to create plots

```
library(ggplot2)
```

Create the base layer of your plot

- use 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

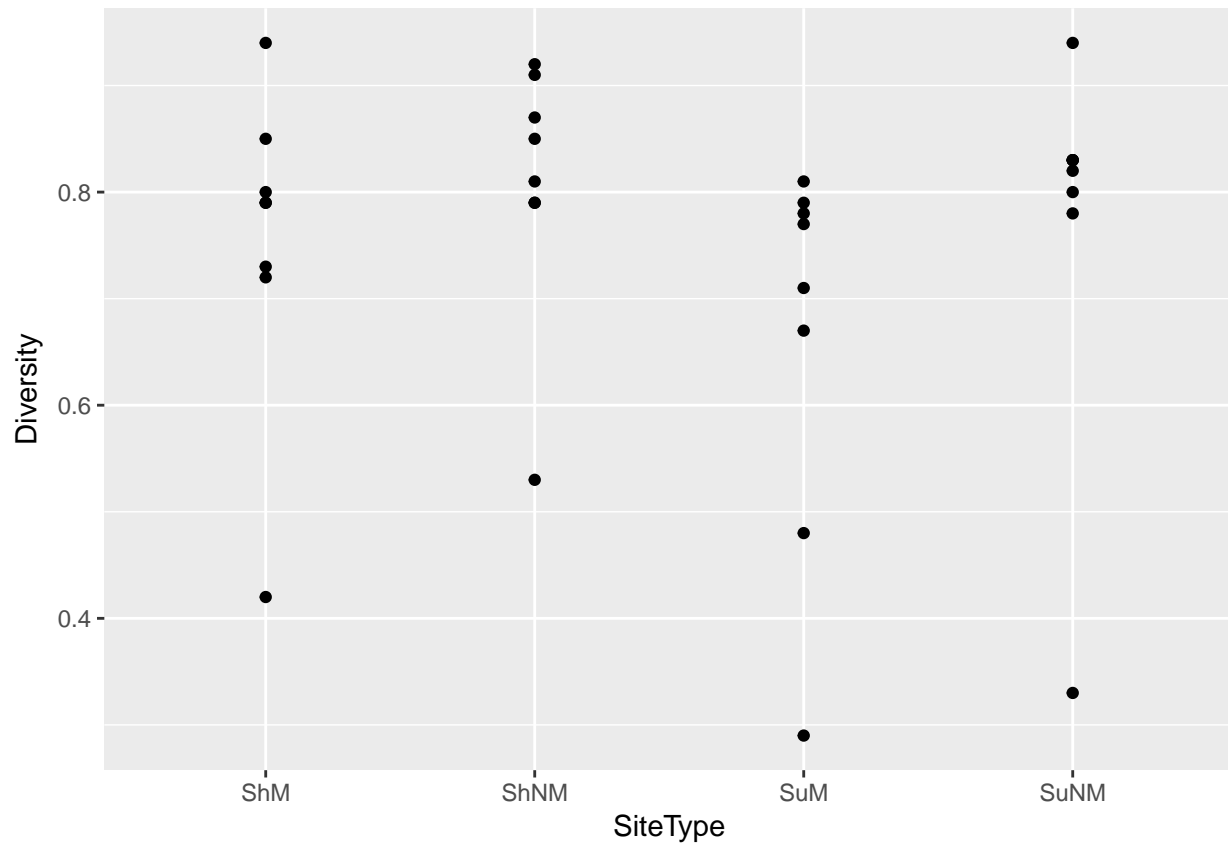
```
ggplot(diversitydata, aes(x=SiteType,y=Diversity))
```



Add data 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()`

```
#insert comment here  
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
  geom_point()
```

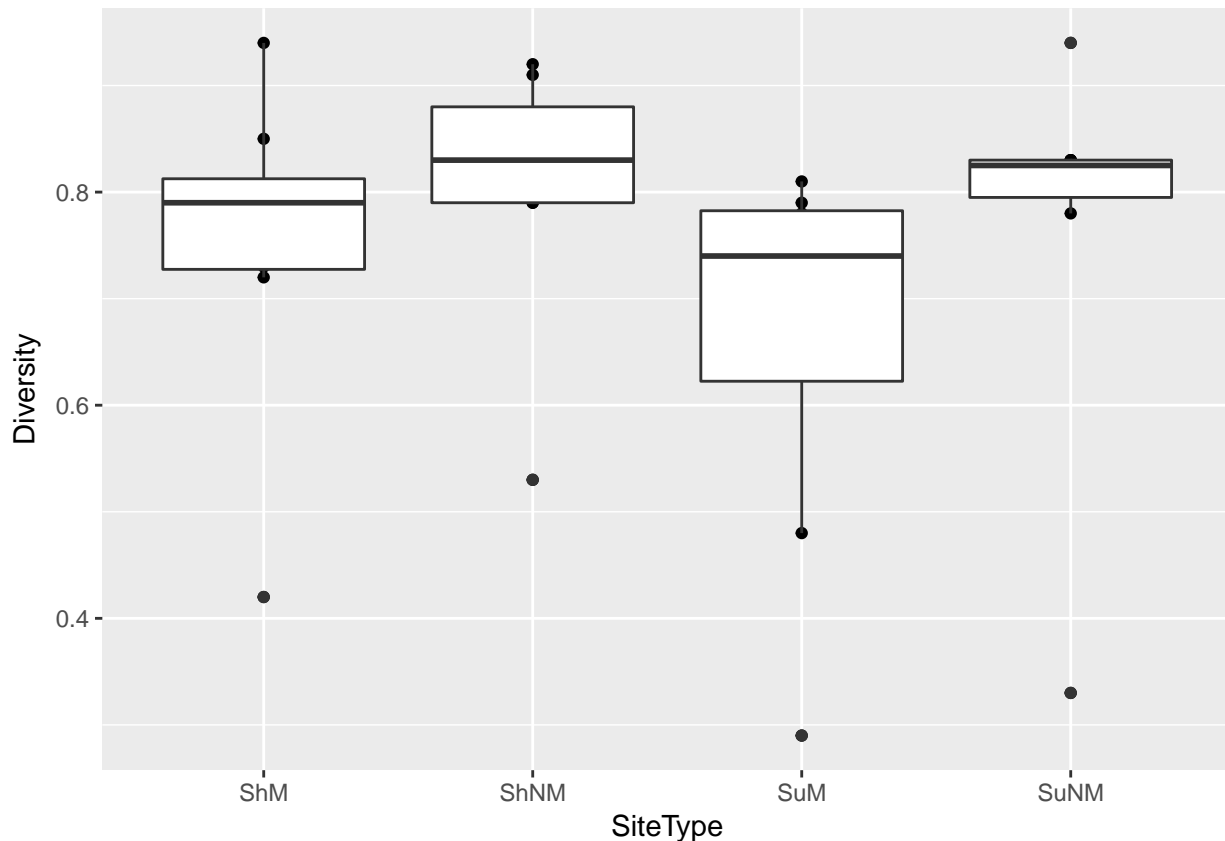


Add real comments to your script

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()`
- note the `+` must come at the end of a line or the command will terminate and not plot another layer

```
#insert comment here
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+
  geom_point()+
  geom_boxplot()
```



Adjust comments as needed

- 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**. Your work will save to the server that we are using for this lab. You will be able to access your work on this server from other computers, as long as you know the web address and your username and password.

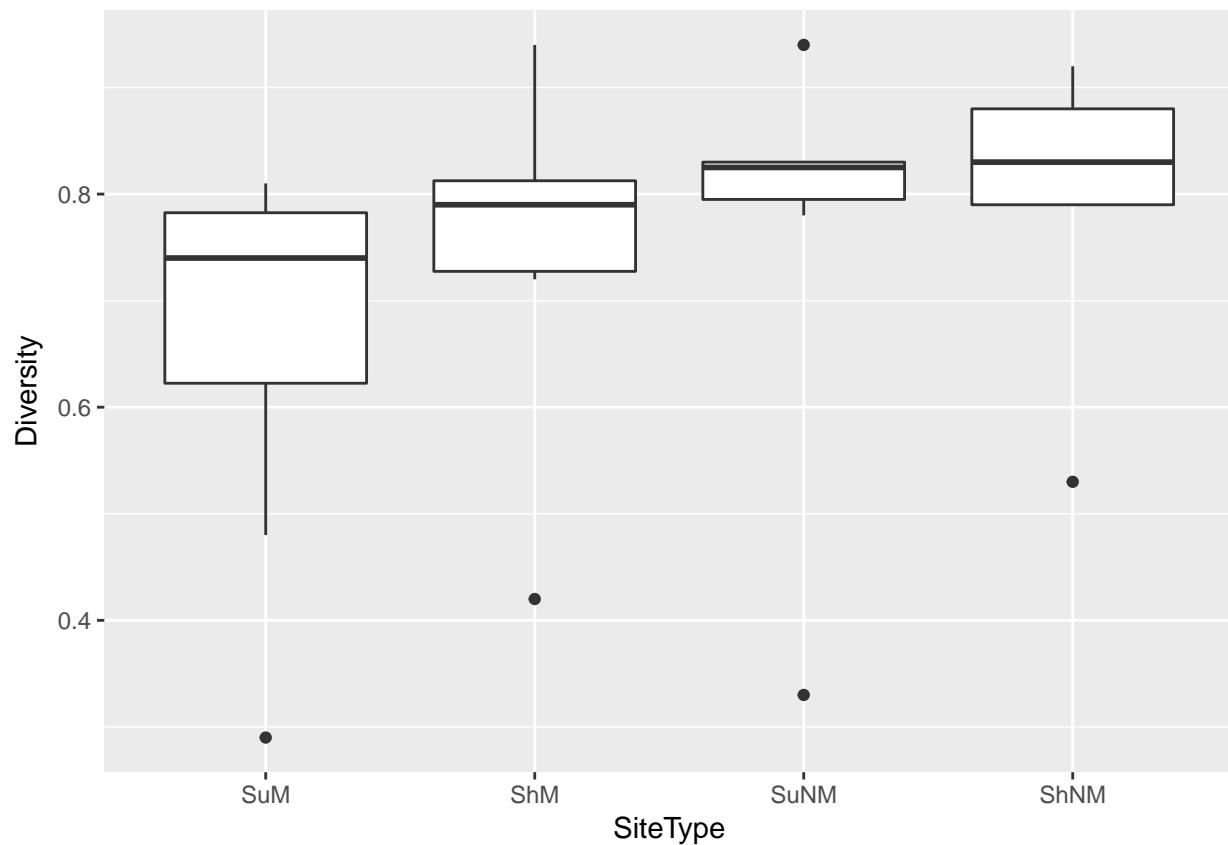
Plotting Part II

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 sites 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

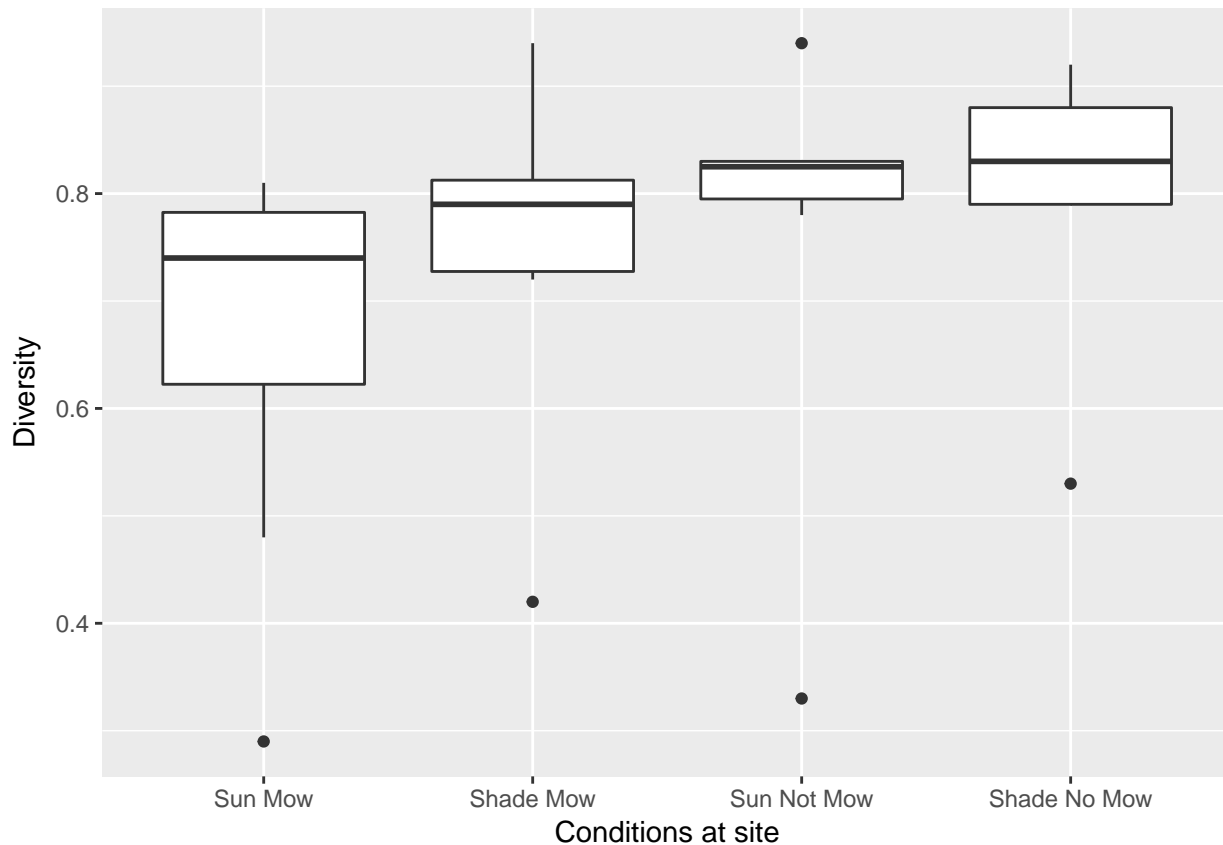
```
#insert comment here
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+
  geom_boxplot()+
  scale_x_discrete(limits=c("SuM", "ShM", "SuNM", "ShNM"))
```



Add labels to your plot

- 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")
```



Turn in this final boxplot of the data

- Save the boxplot by clicking “export” in the “plots” window (lower right). Your work will save to the server that we are using for this lab.

Optional: Look at a subset of data

If you want to show only data from the two sunny sites, then make a new variable (here called ‘SunnyOnly’) containing these data

- First load the `dplyr` library to provide a method of filtering the data.

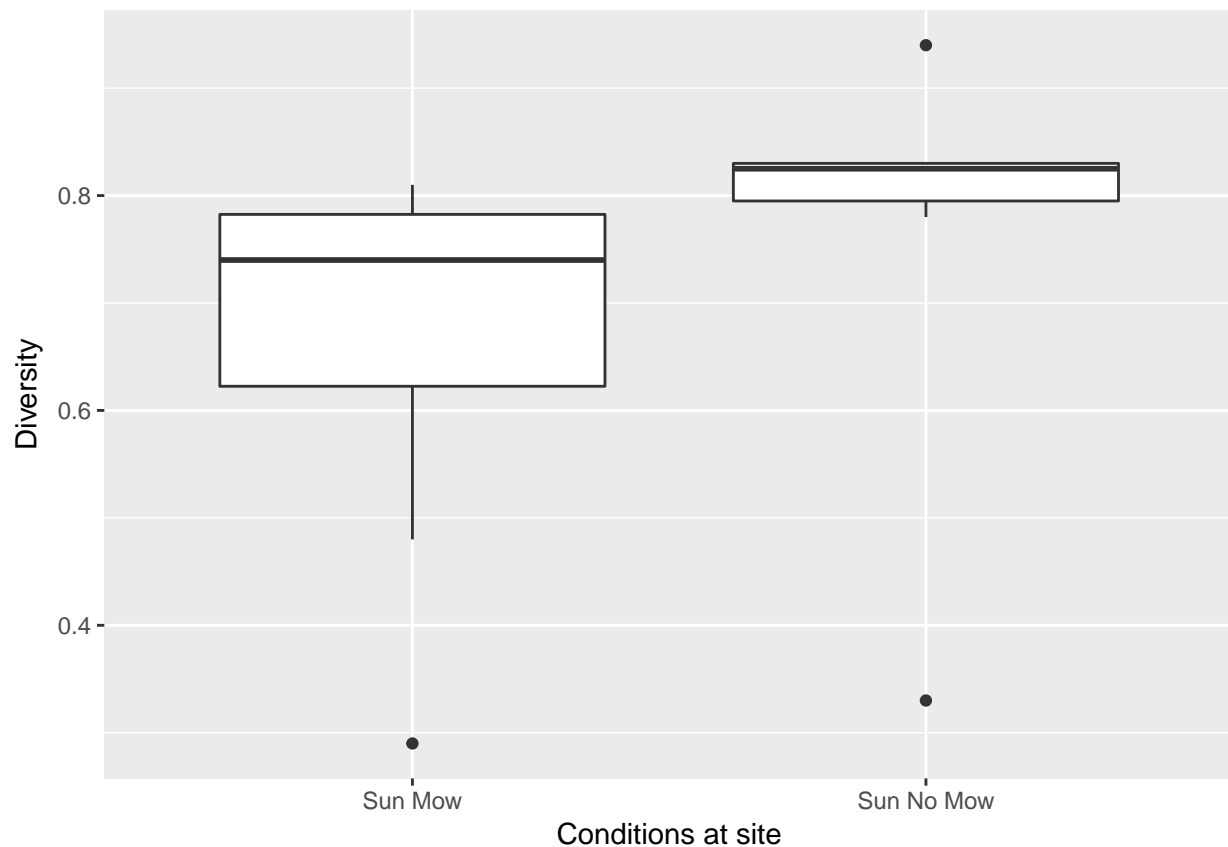
```
library(dplyr)
```

- Then `filter` the data where the value in the SiteType column is either SuNM or SuM (the vertical bar indicates ‘or’)

```
SunnyOnly <- filter(diversitydata, SiteType=="SuNM" | SiteType=="SuM")
```

- Now plot this new dataset

```
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")
```

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 biomes 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