

Algae Phylogeny

Rachel Schwartz

May 24, 2017

The R Software

If you are using your own computer you will need the software R and the friendly user interface RStudio if you do not have them already. First download R from <https://cran.r-project.org/> then click on the file to install it. The download RStudio from <https://www.rstudio.com/products/rstudio/download3/> and install it.

Open RStudio and make a new project. Select File - New Project then New Directory and Empty Project and give it a name. You now have four windows open: 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.

Load data

Last week you calculated the diversity index for each plot within each of four site types. This allows you to understand the variation in diversity among site types.

- Load class data for the diversity of each plot by pulling from google sheets.
- assign the web address of the data to a variable
- 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 above
- view your data by clicking on the item in the Data window on the right side (mine is called `diversitydata` - yours may be named something else) - this will open a new tab above so you can look at the data

Plot results

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

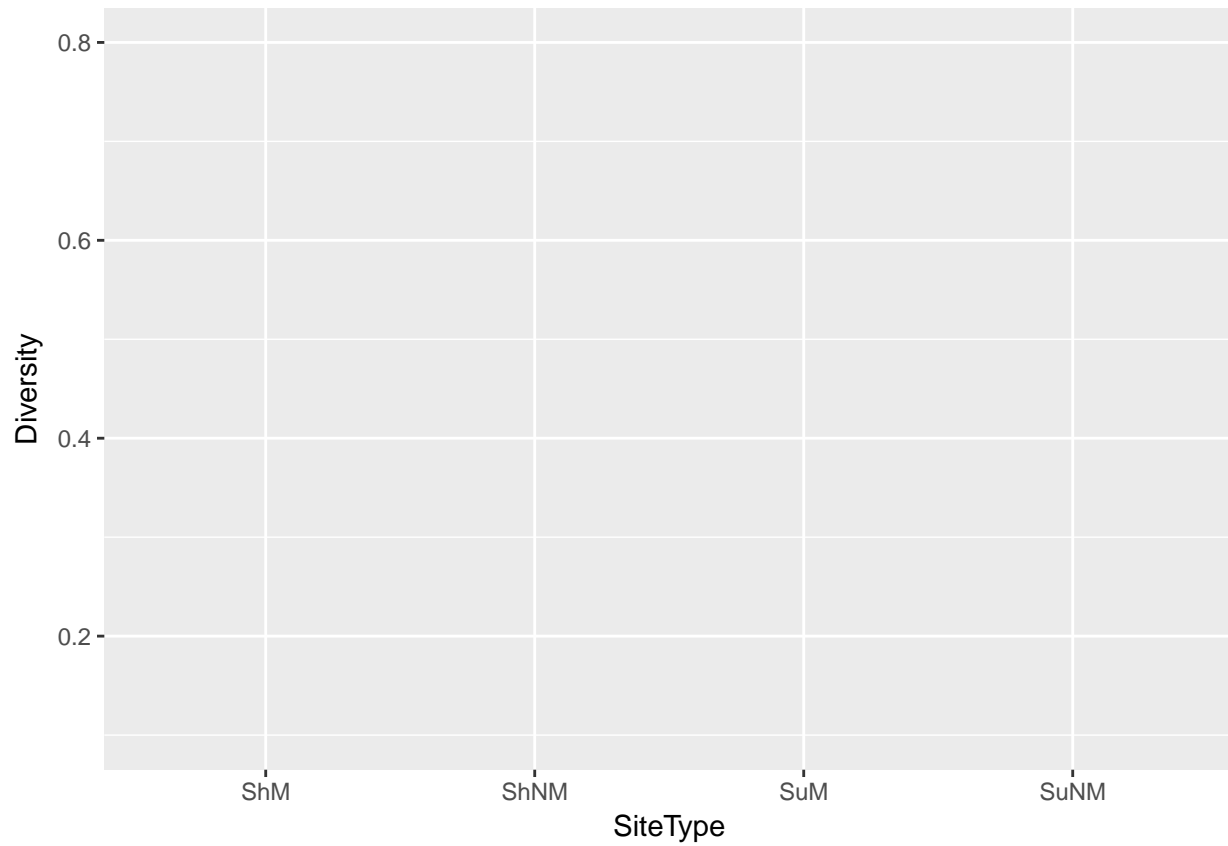
- load the library that allows us to the plotting commands

```
library(ggplot2)
```

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

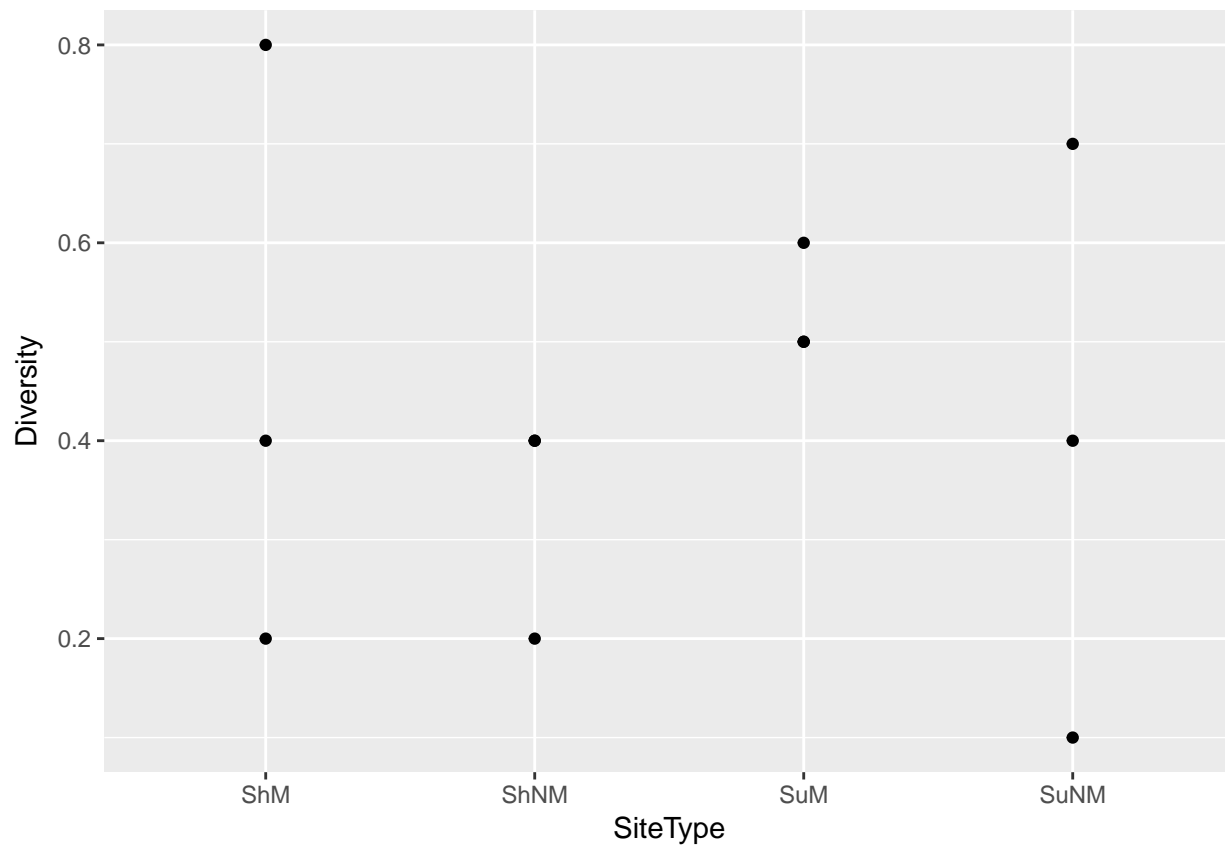
```
ggplot(diversitydata, aes(x=SiteType,y=Diversity))
```



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()`

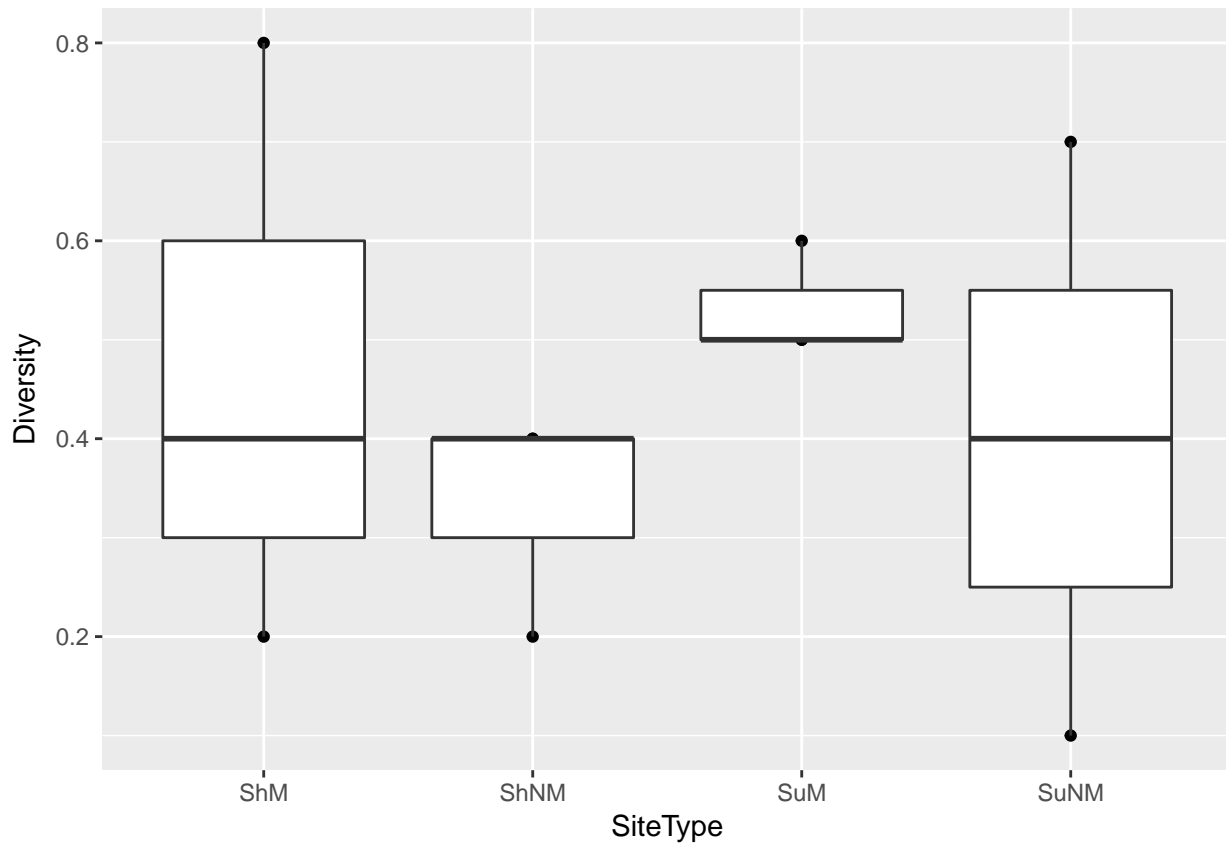
```
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+  
  geom_point()
```



Add a boxplot to 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()
```



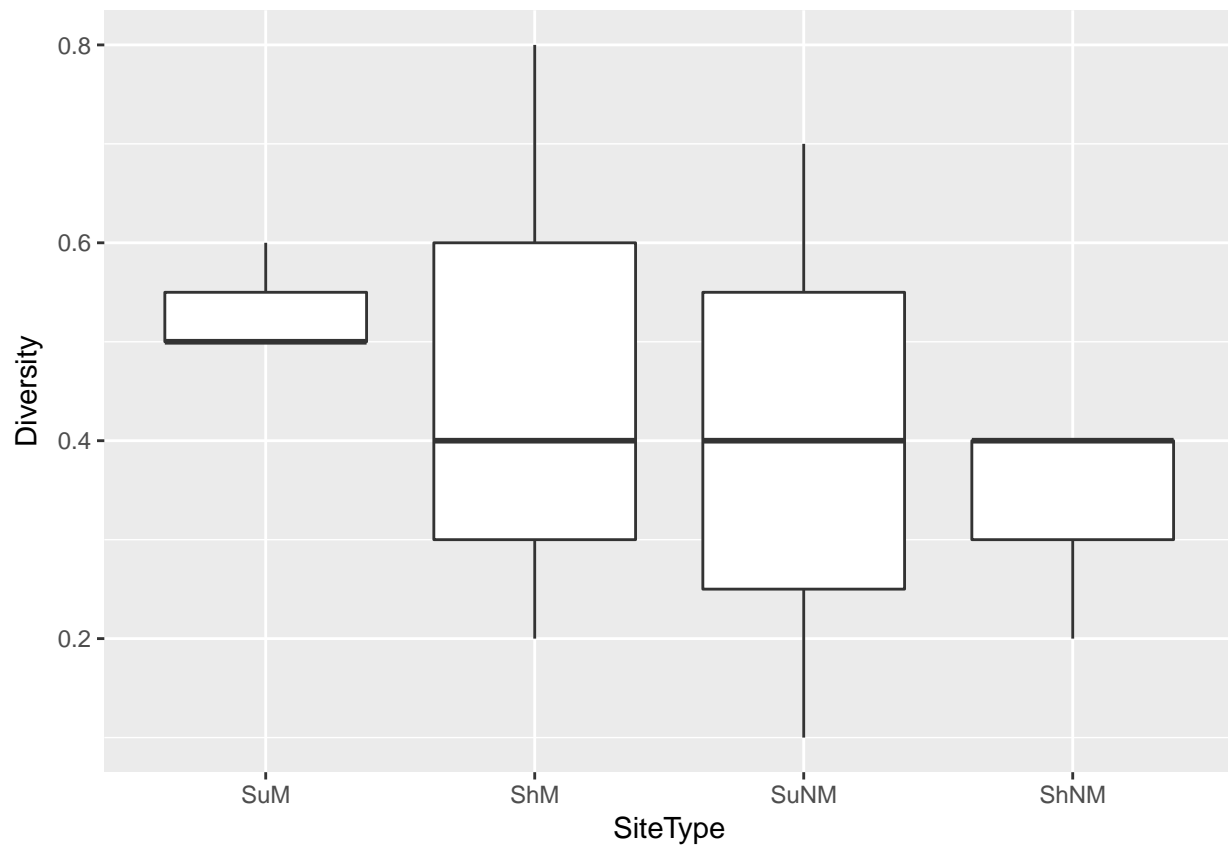
- in order to see both the box and the points plot the box first then the point layer on top
- note: the order of the plotting instructions specify how the plot is made - i.e. first the base, then the points, then the box covering the points
- plot just the boxplot as this is the typical way to visualize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

Reorder data

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

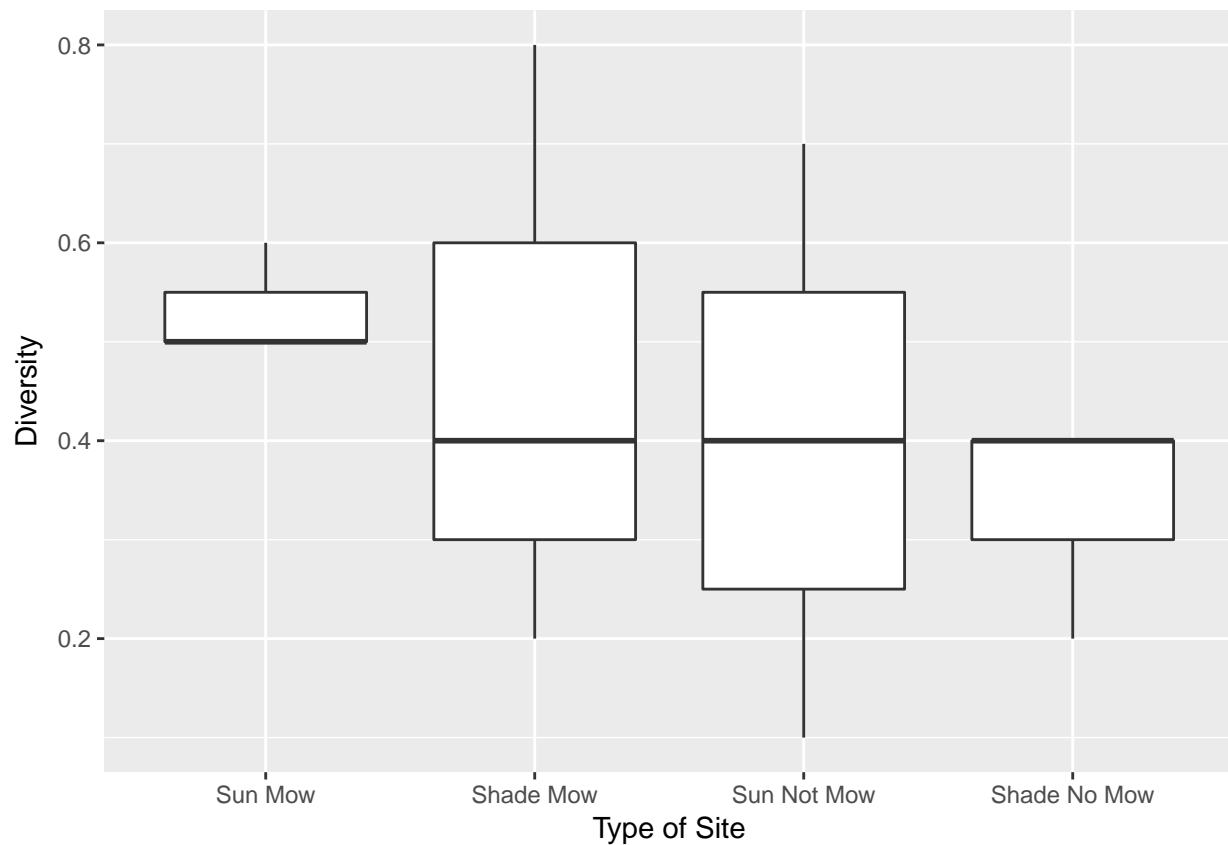
```
ggplot(diversitydata, aes(x=SiteType,y=Diversity))+
  geom_boxplot()+
  scale_x_discrete(limits=c("SuM", "ShM", "SuNM", "ShNM"))
```



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 ="Type of Site")
```



Turn in this final boxplot of the data

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 based on the final plot you made in class.