

# Assignment 8

SiruiZHAO

2023-03-08

My GitHub (<https://github.com/siruiZHAO/Metabarcoding-Community-Ecology.git>)

## Load Library

```
library(dplyr)
library(ggplot2)
library(ggpubr)
library(ape)
library(ggtree)
library(reshape2)
library(vegan)
```

## Import/inspect Data

```
#data import
mydata <- read.csv("../Data/FloristicSurvey.csv")

#check data
str(mydata)
```

```
## 'data.frame':    30 obs. of  44 variables:
## $ Quadrate      : chr  "7o3" "7o1" "7o2" "7i3" ...
## $ Population    : int   7 7 7 7 7 7 3 3 3 3 ...
## $ Location      : chr  "o" "o" "o" "i" ...
## $ Rosettes      : int   0 0 0 14 3 0 14 0 0 0 ...
## $ Bolting       : int   0 0 0 8 18 3 3 12 26 0 ...
## $ Budding       : int   0 0 0 157 184 122 11 23 19 0 ...
## $ Bud_Flw       : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Flower        : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Flw_Sil       : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Sillique      : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Claytonia_virginiana : int  38 29 57 22 54 44 0 0 0 0 ...
## $ Anemone_hepatica : int   8 7 3 0 0 0 0 0 0 0 ...
## $ Grass_tuft    : int  14 22 6 10 3 0 0 0 0 0 ...
## $ Trillium_grandifolium : int  9 30 8 0 0 0 0 0 0 0 ...
## $ Erythronium.trout.lily. : int  25 10 5 16 18 7 0 0 0 0 ...
## $ Acer_saccharum : int   5 10 74 0 2 0 66 60 26 71 ...
## $ Dicentra_cucularia : int   0 22 23 0 0 0 0 0 0 0 ...
## $ Bloodroot     : int   0 4 0 0 1 0 0 0 0 0 ...
## $ Gallium_aparine : int   0 27 8 17 22 21 2 0 0 0 ...
## $ Ulmus_america : int   0 3 0 88 150 133 0 0 0 0 ...
## $ Unknown_1     : int   0 1 0 0 0 0 0 0 0 0 ...
## $ Unknown_2     : int   0 0 0 5 4 4 0 0 0 0 ...
## $ Unknown_3.ranunculaceae.: int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_4     : int   0 0 0 0 0 0 1 0 0 0 ...
## $ Dryopteris_marginalis : int  0 0 0 0 0 0 0 1 0 0 ...
## $ Ostrich_fern  : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Plantago_lanceolata : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Violet        : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Rhamnus_frangula : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Raspberry     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_5     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_6     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Solidago_canadensis : int  0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_7     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Dandelion     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ grass         : int   0 0 0 0 0 0 0 0 0 0 ...
## $ vicia_cracca  : int   0 0 0 0 0 0 0 0 0 0 ...
## $ herb_robert   : int   0 0 0 0 0 0 0 0 0 0 ...
## $ thorny_ash    : int   0 0 0 0 0 0 0 0 0 0 ...
## $ rhamnus_cathartica : int  0 0 0 0 0 0 0 0 0 0 ...
## $ rhubarb       : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_8     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ Unknown_9     : int   0 0 0 0 0 0 0 0 0 0 ...
## $ maianthemum_racemosum : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
mydata$Population <- as.character(mydata$Population)
```

## Analysis

## Data Selection

```
#how the communities of plants differ across sample locations
#select species and make Quadrate as row names
data1 <- mydata %>%
  select(11:44)
row.names(data1) <- mydata$Quadrate
```

## Binary Distance Matrix

```
BinDat<-data1
for(row in 1:nrow(data1)){
  for(col in 1:ncol(data1)){
    if(BinDat[row,col]>0){
      BinDat[row,col]<-1
    }
  }
}
head(BinDat)
```

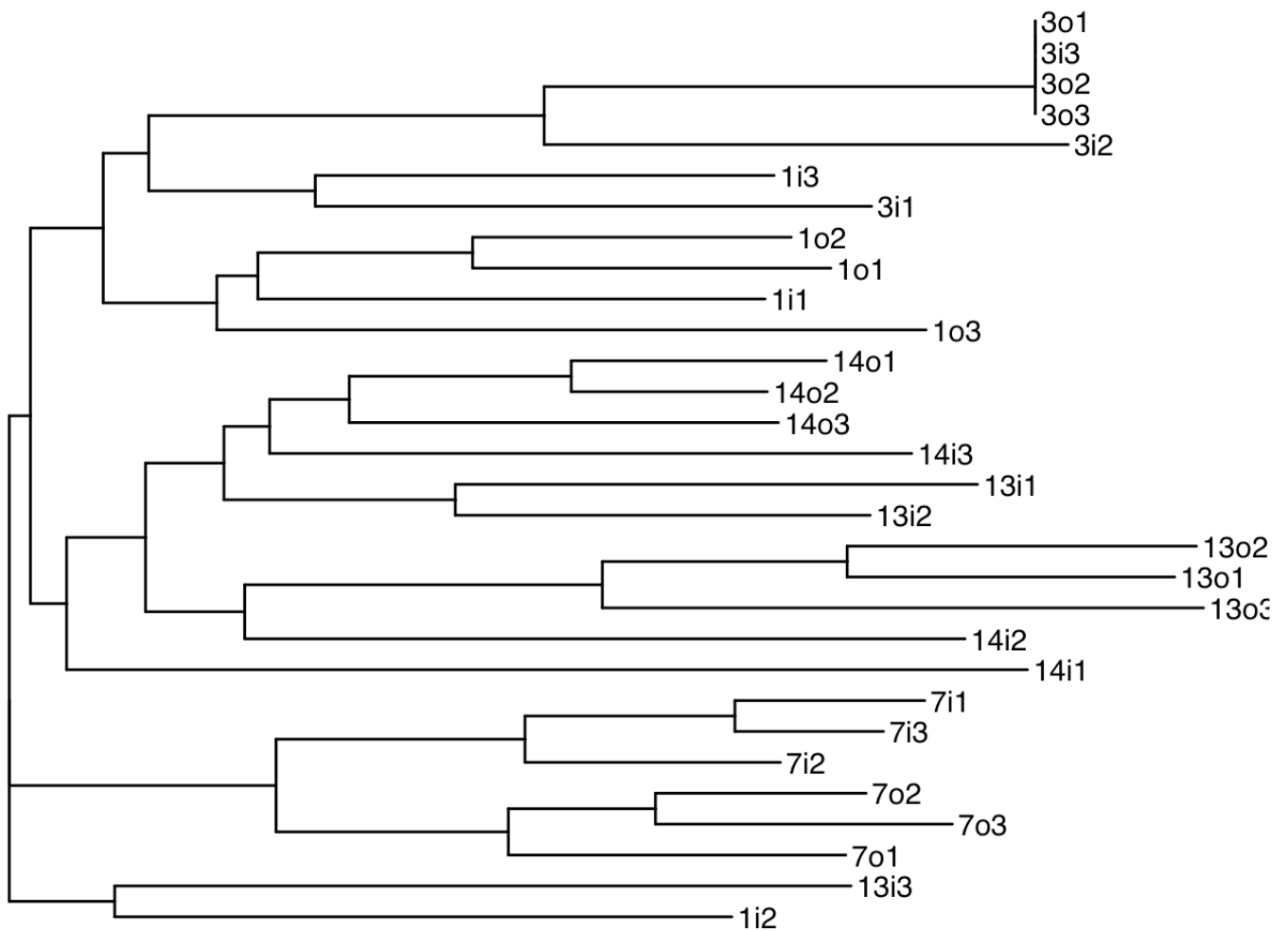
```
##      Claytonia_virginiana Anemone_hepatica Grass_tuft Trillium_grandifolium
## 7o3                      1                1          1                1
## 7o1                      1                1          1                1
## 7o2                      1                1          1                1
## 7i3                      1                0          1                0
## 7i2                      1                0          1                0
## 7i1                      1                0          0                0
##      Erythronium.trout.lily. Acer_saccharum Dicentra_cucularia Bloodroot
## 7o3                      1                1                0          0
## 7o1                      1                1                1          1
## 7o2                      1                1                1          0
## 7i3                      1                0                0          0
## 7i2                      1                1                0          1
## 7i1                      1                0                0          0
##      Gallium_aparine Ulmus_americanana Unknown_1 Unknown_2
## 7o3                      0                0                0          0
## 7o1                      1                1                1          0
## 7o2                      1                0                0          0
## 7i3                      1                1                0          1
## 7i2                      1                1                0          1
## 7i1                      1                1                0          1
##      Unknown_3.rannunculaceae. Unknown_4 Dryopteris_marginalis Ostrich_fern
## 7o3                      0                0                0          0
## 7o1                      0                0                0          0
## 7o2                      0                0                0          0
## 7i3                      0                0                0          0
## 7i2                      0                0                0          0
## 7i1                      0                0                0          0
##      Plantago_lanceolata Violet Rhamnus_frangula Raspberry Unknown_5 Unknown_6
```

## 7o3	0	0	0	0	0	0
## 7o1	0	0	0	0	0	0
## 7o2	0	0	0	0	0	0
## 7i3	0	0	0	0	0	0
## 7i2	0	0	0	0	0	0
## 7i1	0	0	0	0	0	0
##	Solidago_canadensis	Unknown_7	Dandelion	grass	viccia_cracca	herb_robert
## 7o3	0	0	0	0	0	0
## 7o1	0	0	0	0	0	0
## 7o2	0	0	0	0	0	0
## 7i3	0	0	0	0	0	0
## 7i2	0	0	0	0	0	0
## 7i1	0	0	0	0	0	0
##	thorny_ash	rhamnus_cathartica	rhubarb	Unknown_8	Unknown_9	
## 7o3	0	0	0	0	0	
## 7o1	0	0	0	0	0	
## 7o2	0	0	0	0	0	
## 7i3	0	0	0	0	0	
## 7i2	0	0	0	0	0	
## 7i1	0	0	0	0	0	
##	maianthenum_racemosum					
## 7o3	0					
## 7o1	0					
## 7o2	0					
## 7i3	0					
## 7i2	0					
## 7i1	0					

## Neighbor-Joining

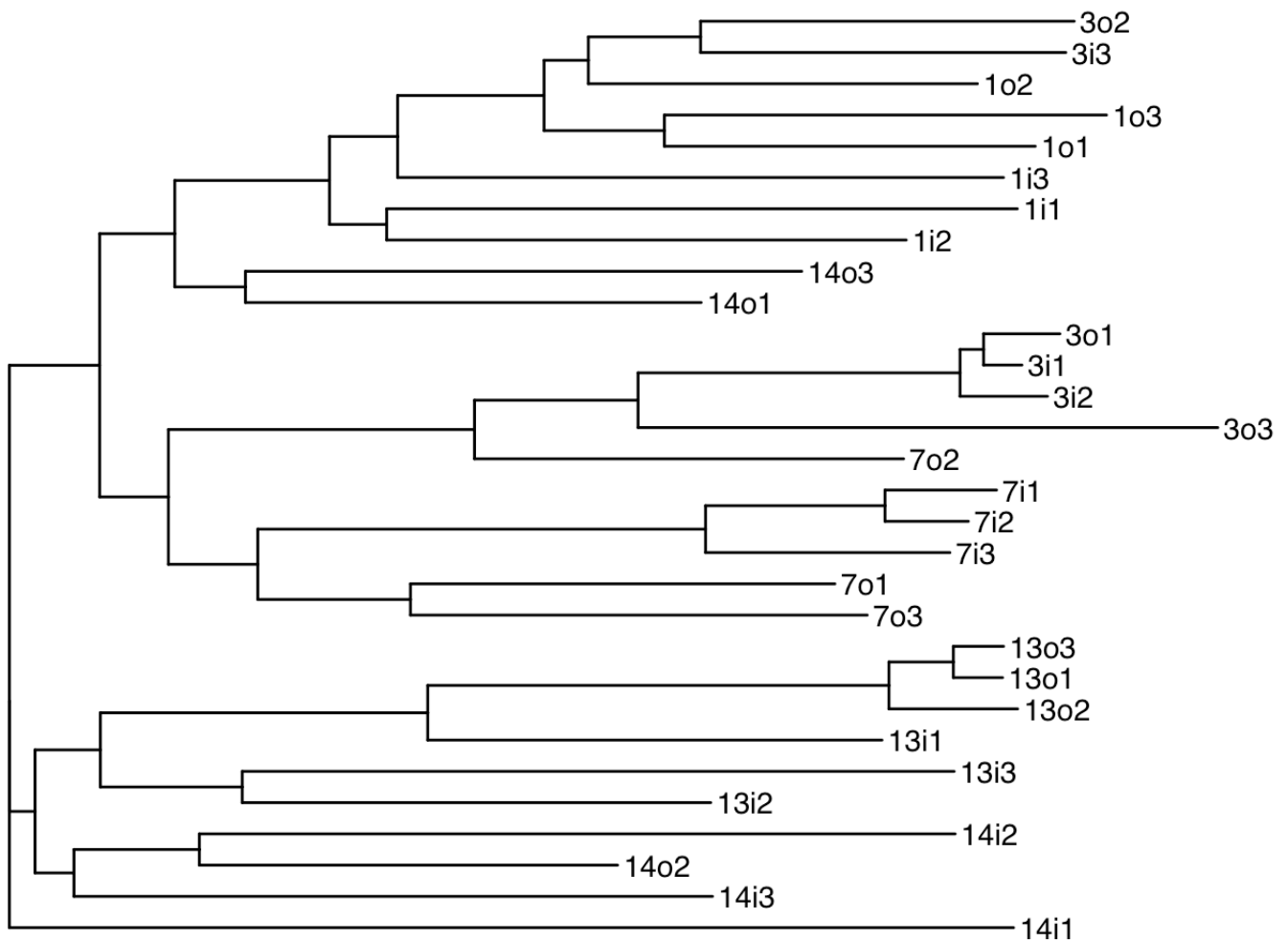
```
#pairwise distance of the binary matrix
BinDist <- dist(BinDat, method = 'binary')

#NJ tree
tree <- nj(BinDist)
ggtree(tree, layout = "rectangular") %<+% mydata +
  geom_tiplab()
```



## Bray-Curtis dissimilarity

```
Dat_dist <- vegdist(data1, method = "bray", binary = F)
Tree_2 <- nj(Dat_dist)
ggtree(Tree_2, layout = "rectangular") %<+% mydata +
  geom_tiplab()
```



These two tree topology have some differences. And not like the first one, the one in the Bray-Curtis section organize the tree based on how different the group are from each other. So, we would like to use the second one in our analysis.

### NJ tree Group by garlic presence

```
#point out the location and population of garlic mustard
present <- gsub(".*([o|i]).*", "\\1", Tree_2$tip.label)
population <- gsub("(.*)[o|i].*", "\\1", Tree_2$tip.label)

#group tip labels
PreGroups <- split(Tree_2$tip.label, present)
PopuGroups <- split(Tree_2$tip.label, population)

#apply group to plot
PreDTcol <- groupOTU(Tree_2, PreGroups)
PopuDTcol <- groupOTU(Tree_2, PopuGroups)
```

```
#plot
P1 <- ggtree(PreDTcol, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  scale_color_manual(name = "Garlic Mustard Patches",
                    values = rainbow(2),
                    labels = c("Inside", "Outside"))
P2 <- ggtree(PopuDTcol, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  scale_color_manual(name = "Population",
                    values = rainbow(5),
                    labels = c("1", "3", "7", "13", "14"))
figure1 <- ggarrange(P1, P2, labels = c("A", "B"), ncol = 1, nrow = 2)
figure1
```

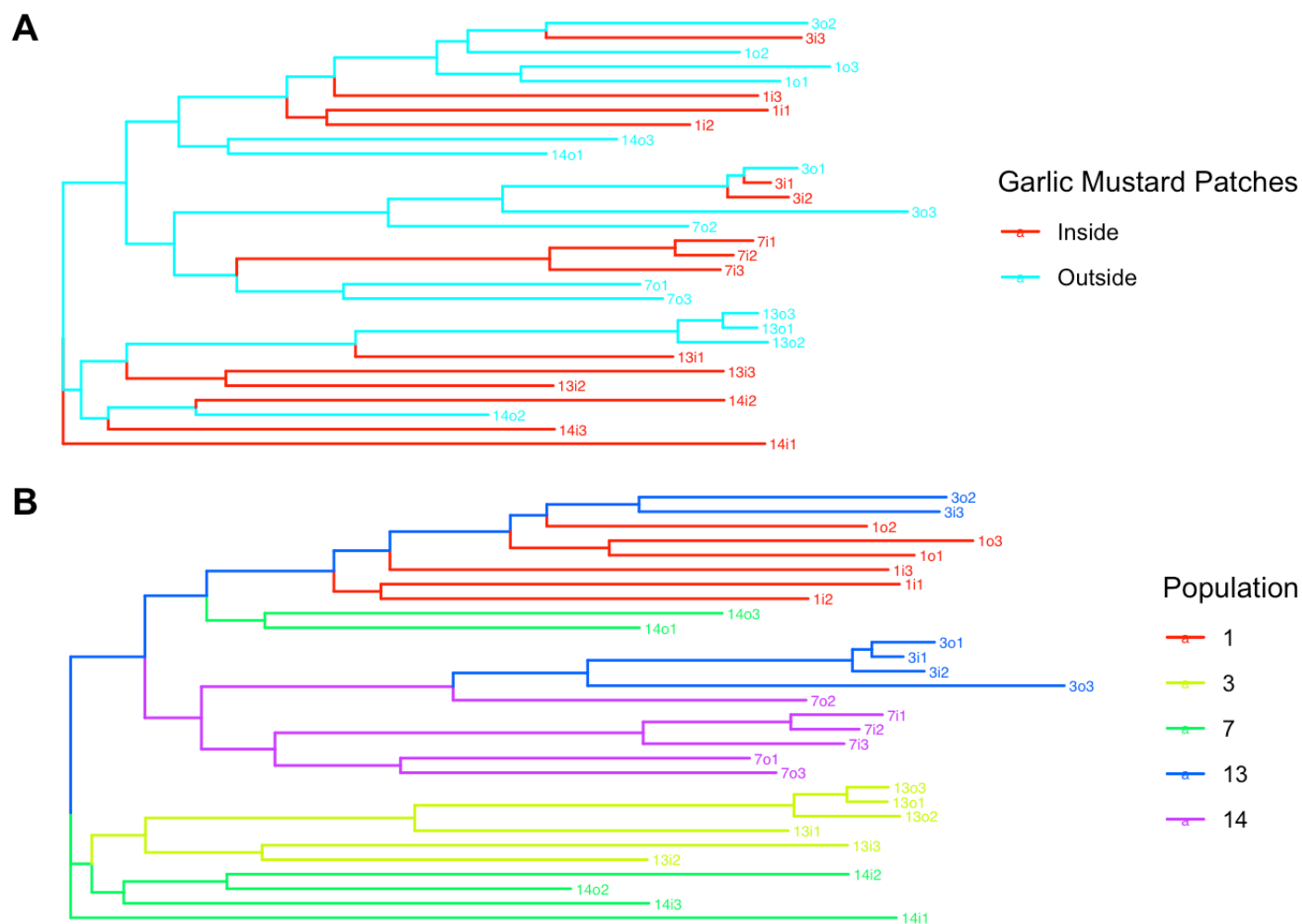


Figure 1: Neighbor-Joining Tree that clusters samples based on the similarity of species abundance. A) Different colors represent whether quadrat measurements are taken inside (red) or outside (blue) of garlic mustard patches. B) Different colors represent a specific sampling location at QUBS, determined by presence of garlic mustard nearby.

## NMDS analysis

```
set.seed(20)
NMDSdat <- metaMDS(Dat_dist, k = 2, trymax = 100)
```

```
## Run 0 stress 0.1461135
## Run 1 stress 0.1414388
## ... New best solution
## ... Procrustes: rmse 0.06689963  max resid 0.309254
## Run 2 stress 0.1666366
## Run 3 stress 0.1602287
## Run 4 stress 0.1414388
## ... Procrustes: rmse 8.157245e-05  max resid 0.0002927585
## ... Similar to previous best
## Run 5 stress 0.1561092
## Run 6 stress 0.1666369
## Run 7 stress 0.1461135
## Run 8 stress 0.1414389
## ... Procrustes: rmse 0.000167074  max resid 0.0006280928
## ... Similar to previous best
## Run 9 stress 0.1534441
## Run 10 stress 0.153444
## Run 11 stress 0.1461136
## Run 12 stress 0.1489965
## Run 13 stress 0.1560522
## Run 14 stress 0.1489971
## Run 15 stress 0.1665458
## Run 16 stress 0.1461135
## Run 17 stress 0.1589506
## Run 18 stress 0.1461135
## Run 19 stress 0.1489954
## Run 20 stress 0.1414389
## ... Procrustes: rmse 0.0001979677  max resid 0.0007305364
## ... Similar to previous best
## *** Best solution repeated 3 times
```

```
#create data for plotting
PDat <- data.frame(NMDS1 = NMDSdat$points[,1],
                  NMDS2 = NMDSdat$points[,2],
                  Quadrate = row.names(data1))

#add species labels
PDat <- merge(PDat, mydata, by = "Quadrate", all.x = T, all.y = F)
```

```
#plot
qplot(x = NMDS1, y = NMDS2, color = Population, shape = Location,
      alpha = I(0.6), data = PDat) +
  theme_bw()
```



```
## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
```

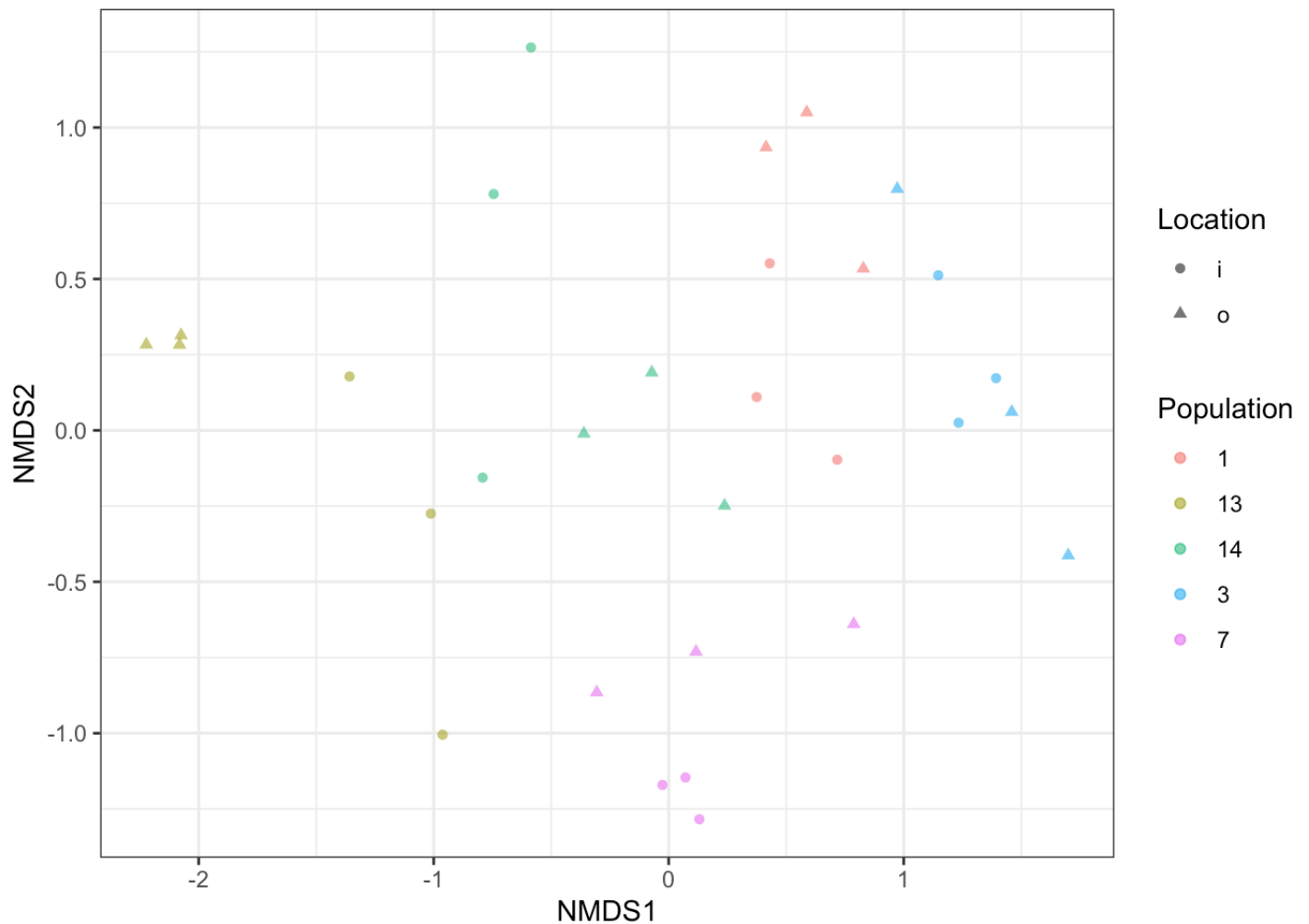


Figure 2: Non\_Metric Multidimensional Scaling to visualize similarity and differences among species abundance.

## Report

Garlic mustard is an invasive plant in ON, CA. it can reduce the biodiversity and aesthetic value of natural areas. In this study, the data collected from 5 different locations (Populations) at QUBS, which were determined by the presence of garlic mustard nearby, and grouped by the location, which indicates whether Quadrature measurements are taken 'i' for 'inside' or 'o' for 'outside' of garlic mustard patches.

During our analysis, we select the 34 species present at all sites and Quadrature code. The data were grouped by "Population" and "Location" to compare if the presence of garlic mustard influences the species' abundance. The results show that the presence of garlic mustard can drive the differentiation of species abundance in the local communities (Figure 1(A)(B)). In the NMDS analysis, data in different "Locations" impact the local communities more, which can be determined by the distance between each point (Figure 2).

Thanks!