

# Beyond Metabarcoding- Community Ecology

Logan Wistead

2022-03-16

## Project Info

Setup Github username: LoganWistead  
Github repository

## Introduction

The data set FloristicSurvey.csv contains the results of a floristic survey at the Queen's University Biological Station (QUBS). This data set investigates garlic mustard, which is an invasive species that is known to easily and rapidly spread and cause damage to other plants and organisms.

This report will address two important biological questions:

1. What effect (if any) does garlic mustard have on the plant community?
2. What has a stronger effect on plant communities: the presence/absence of garlic mustard (in/out) or sampling population?

## Load libraries

```
library(devtools)
```

```
## Loading required package: usethis
```

```
devtools::install_github('GuangchuangYu/ggtree', force = TRUE)
```

```
## Downloading GitHub repo GuangchuangYu/ggtree@HEAD
```

```
##
```

```
## * checking for file '/private/var/folders/v4/psz5kttt17kq7lc_3x0f_z7wm0000gn/T/RtmpjSERmV/remotes5e0d'
```

```
## * preparing 'ggtree':
```

```
## * checking DESCRIPTION meta-information ... OK
```

```
## * checking for LF line-endings in source and make files and shell scripts
```

```
## * checking for empty or unneeded directories
```

```
## * building 'ggtree_3.3.1.tar.gz'
```

```
library("ggtree")
```

```
## ggtree v3.3.1 For help: https://yulab-smu.top/treedata-book/
##
## If you use the ggtree package suite in published research, please cite
## the appropriate paper(s):
##
## Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam.
## ggtree: an R package for visualization and annotation of phylogenetic
## trees with their covariates and other associated data. Methods in
## Ecology and Evolution. 2017, 8(1):28-36. doi:10.1111/2041-210X.12628
##
## Guangchuang Yu. Using ggtree to visualize data on tree-like structures.
## Current Protocols in Bioinformatics. 2020, 69:e96. doi:10.1002/cpbi.96
##
## S Xu, Z Dai, P Guo, X Fu, S Liu, L Zhou, W Tang, T Feng, M Chen, L
## Zhan, T Wu, E Hu, Y Jiang, X Bo, G Yu. ggtreeExtra: Compact
## visualization of richly annotated phylogenetic data. Molecular Biology
## and Evolution. 2021, 38(9):4039-4042. doi: 10.1093/molbev/msab166
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      intersect, setdiff, setequal, union
```

```
library(ape)
```

```
## Warning: package 'ape' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'ape'
```

```
## The following object is masked from 'package:ggtree':
```

```
##
```

```
##      rotate
```

```
library(vegan)
```

```
## Loading required package: permute
```

```
## Warning: package 'permute' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'permute'
```

```
## The following object is masked from 'package:devtools':
```

```
##
```

```
##      check
```

```
## Loading required package: lattice
```

```
## This is vegan 2.5-7
```

```
library(ggtree)
library(reshape2)
library(ggplot2)
```

## Load data

```
floralSurvey <- read.csv("./A8_WISTEARD_LOGAN_Data.csv")
```

## preparing the data

reordering the columns and removing unneeded ones

```
floralSubset <- floralSurvey %>%
  select(Claytonia_virginiana:maianthemum_racemosum)
row.names(floralSubset) <- floralSurvey$Quadrat
```

## Calculate Bray-Curtis dissimilarity matrix

```
dist <- vegdist(floralSubset, method = "bray", binary = FALSE)
```

Visualize the matrix

```
distMat <- as.matrix(dist)
distLong <- melt(distMat)
ggplot(data = distLong, mapping = aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradientn(colours = c("white", "blue", "green", "red")) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

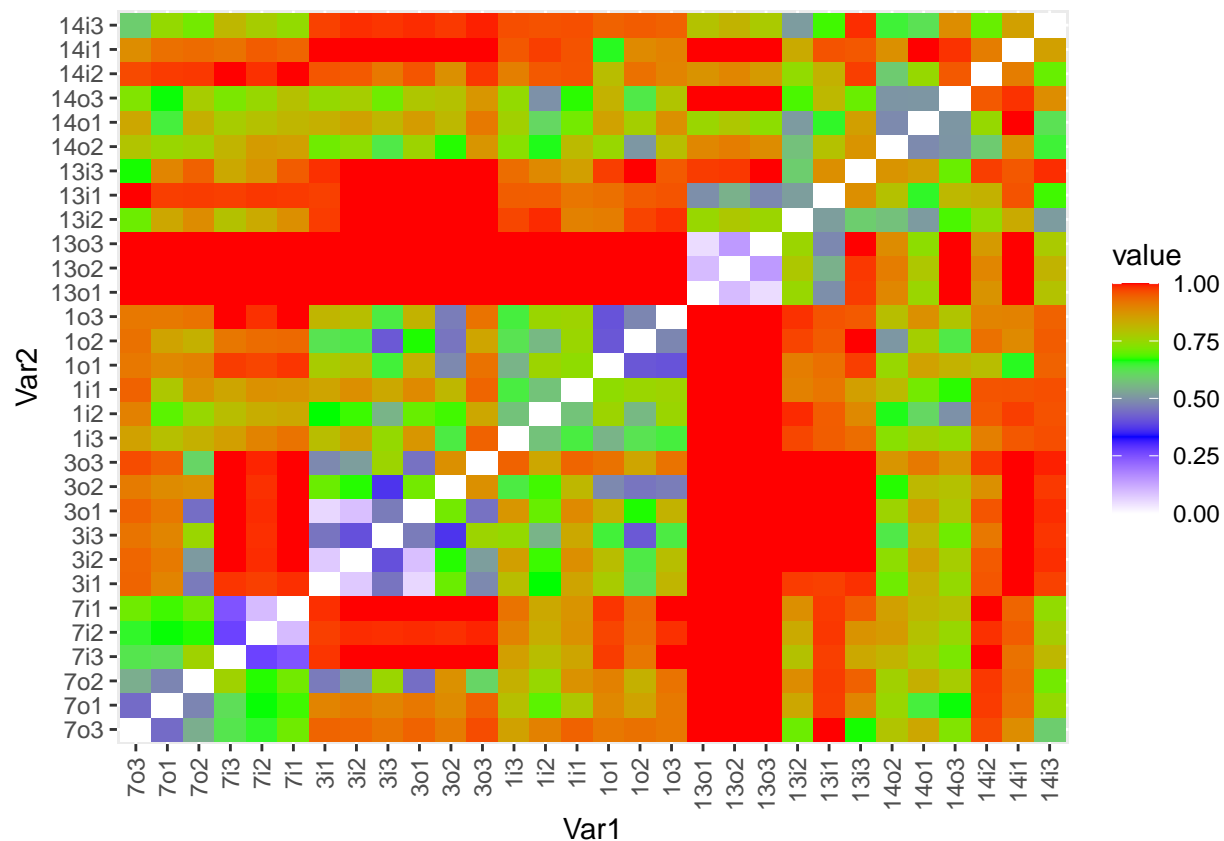


Figure 1. Distance matrix using each Quadrate with Bray-Curtis dissimilarity method.

#Neighbour-Joining Cluster our samples by their similarity. We can use the neighbor-joining method:

```
floralNJ <- nj(dist)
ggtree(floralNJ, layout = "rectangular") %<+% floralSurvey +
  geom_tree(mapping = aes(colour = Location)) +
  geom_tiplab() +
  scale_colour_discrete(name = "Garlic Mustard", labels = c("present", "absent"))
```

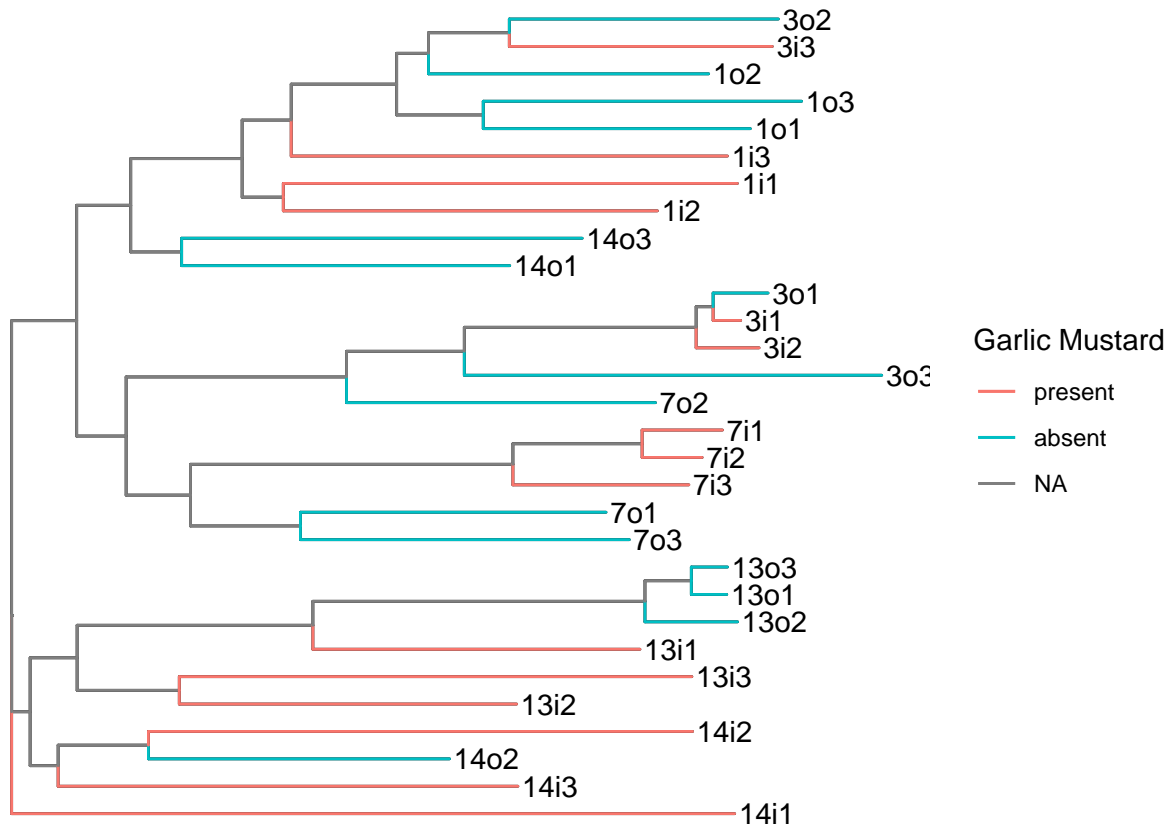


Figure 2. This tree indicates what similarities can be found within differences in sampling sites

#NMDS an algorithm for plotting points so that two points that are similar end up close together on the graph.

```
set.seed(13)
floralnmds <- metaMDS(dist, k = 2)

## Run 0 stress 0.1461135
## Run 1 stress 0.153444
## Run 2 stress 0.153444
## Run 3 stress 0.141439
## ... New best solution
## ... Procrustes: rmse 0.06682265 max resid 0.3092955
## Run 4 stress 0.1489959
## Run 5 stress 0.1489955
## Run 6 stress 0.1560521
## Run 7 stress 0.161178
## Run 8 stress 0.1414389
## ... New best solution
## ... Procrustes: rmse 0.0001063247 max resid 0.0003944835
## ... Similar to previous best
## Run 9 stress 0.141439
## ... Procrustes: rmse 0.0004805731 max resid 0.001783307
## ... Similar to previous best
## Run 10 stress 0.153444
## Run 11 stress 0.1589506
```

```
## Run 12 stress 0.1560521
## Run 13 stress 0.1461135
## Run 14 stress 0.2101893
## Run 15 stress 0.1489955
## Run 16 stress 0.206437
## Run 17 stress 0.1589506
## Run 18 stress 0.1665459
## Run 19 stress 0.1676016
## Run 20 stress 0.1414388
## ... New best solution
## ... Procrustes: rmse 0.0002399126 max resid 0.0008810417
## ... Similar to previous best
## *** Solution reached
```

```
pdat <- data.frame(NMDS1 = floralnmds$points[,1],
                  NMDS2 = floralnmds$points[,2],
                  Quadrate = row.names(floralSubset))
pdat <- merge(pdat, floralSurvey, by = "Quadrate", all.x = T, all.y = T)
```

NMDS plot

```
ggplot(data = pdat, aes(x = NMDS1, y = NMDS2, colour = Location)) +
  geom_point(size = 2) +
  scale_colour_discrete(labels = c("present", "absent"), name = "Garlic Mustard") +
  theme_classic()
```

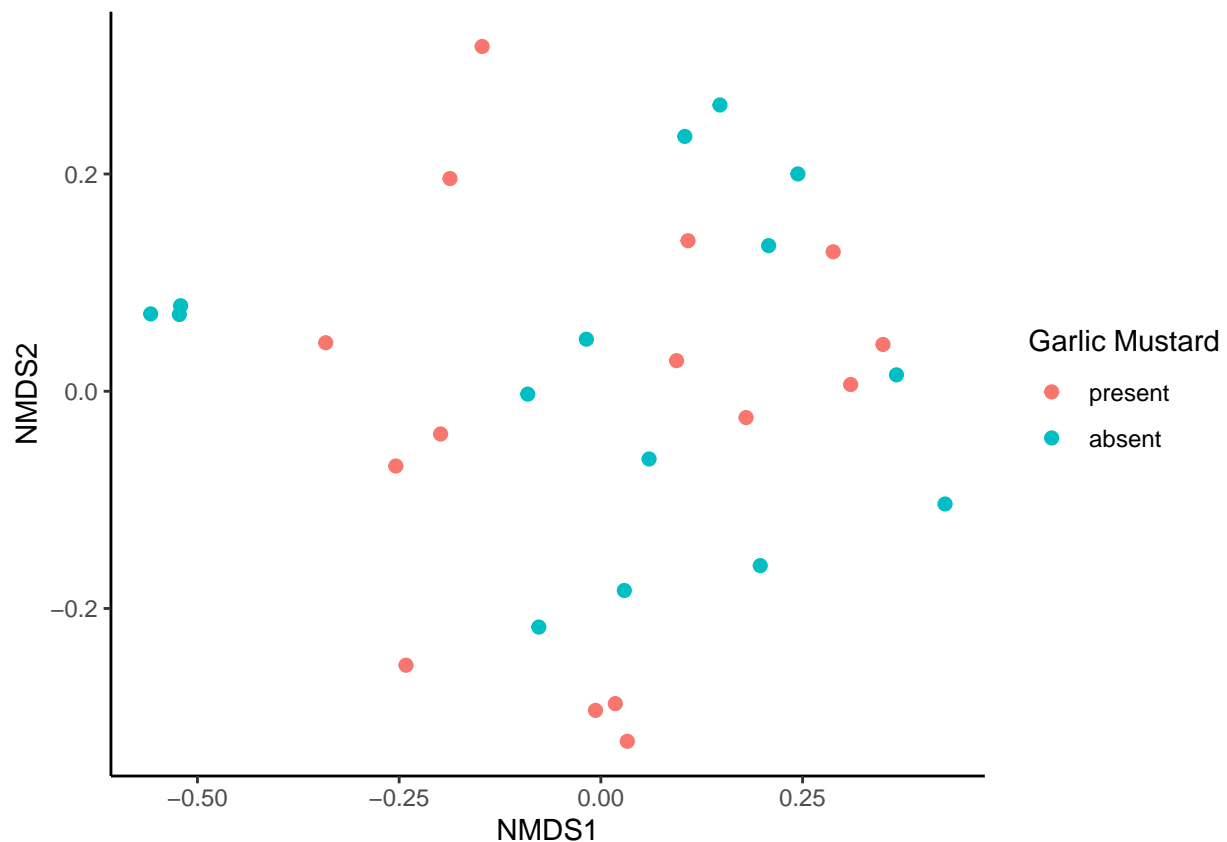


Figure 3.1. Scatter plot to further analyze the NMD results with colour coded populations.

```
ggplot(data = pdat, aes(x = NMDS1, y = NMDS2, colour = factor(Population), shape = Location)) +
  geom_point(size = 2) +
  scale_colour_discrete(name = "Population") +
  theme_classic()
```

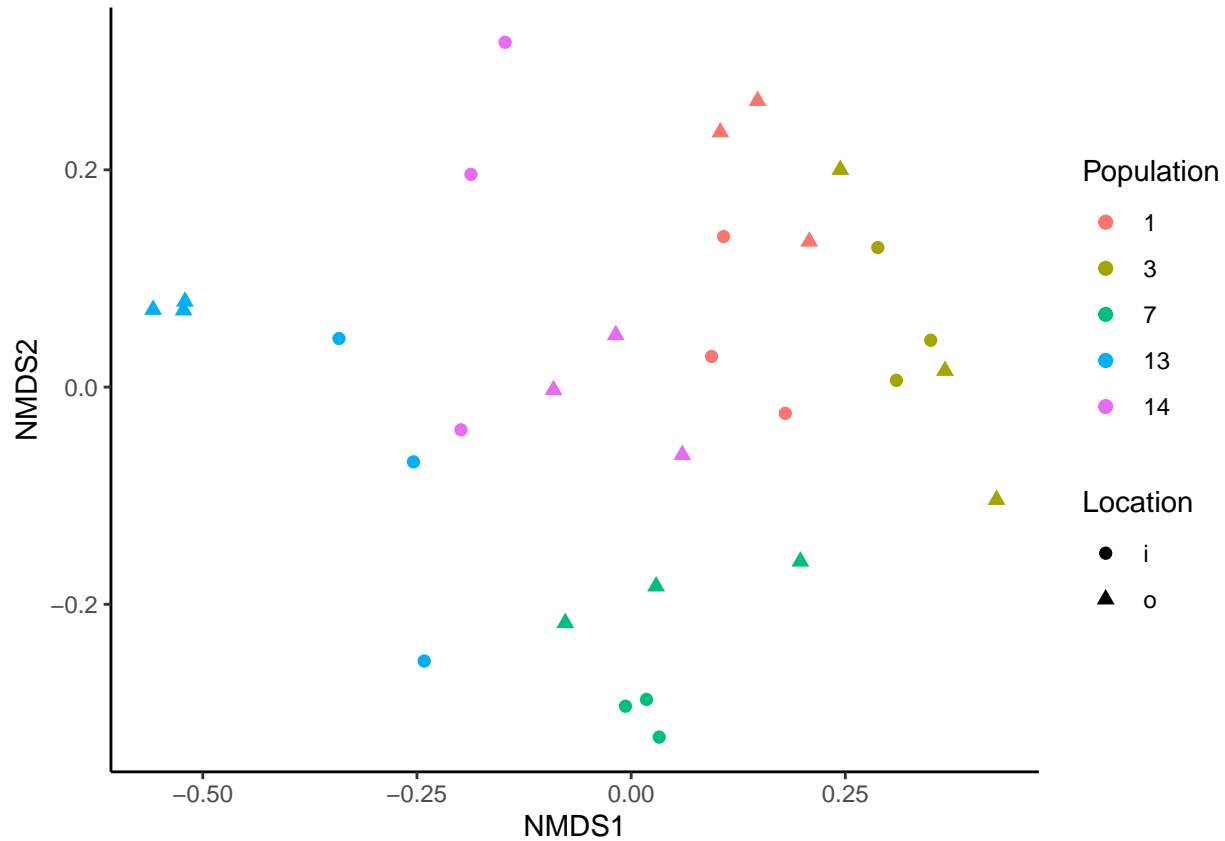


Figure 3.2. Attempting to try and find more clustering within populations

## Discussion

Based on Figure 1. and the similarity between the samples, it is likely that sites in the same presence or absence have them most similarity. This likely does not have an effect on the plant community. A neighbor joining tree may display these findings more clearly.

Based on Figure 2. and the neighbor joining method, the sample sites from within the same population are most similar to one another. While the effects of garlic mustard are only similar within the same population

Figure 3.1. and the NMDS, it appears that the sampling populations have greater effects on the plant communities than presence or absence of garlic mustard. Based on Figure 3.1. there was still no clustering to make a clear conclusion. Figure 3.2. shows that there may be sites affected within the same populations as there was more evidence of clustering.