

Metabarcoding and Community Ecology

Alyssa Green

2023-03-08

GitHub Repository (<https://github.com/alyssagreen02/BIOL432-Metabarcoding.git>)

Question 1

Load Data

```
FS=read.csv("FloristicSurvey.csv")
```

Load Packages

```
library(ggplot2)
library(ape)
library(ggtree)
```

```
## ggtree v3.6.2 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
##
## LG Wang, TTY Lam, S Xu, Z Dai, L Zhou, T Feng, P Guo, CW Dunn, BR
## Jones, T Bradley, H Zhu, Y Guan, Y Jiang, G Yu. treeio: an R package
## for phylogenetic tree input and output with richly annotated and
## associated data. Molecular Biology and Evolution. 2020, 37(2):599-603.
## doi: 10.1093/molbev/msz240
##
## Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods
## for mapping and visualizing associated data on phylogeny using ggtree.
## Molecular Biology and Evolution. 2018, 35(12):3041-3043.
## doi:10.1093/molbev/msy194
```

```
##  
## Attaching package: 'ggtree'
```

```
## The following object is masked from 'package:ape':  
##  
## rotate
```

```
library(vegan)
```

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

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

```
## This is vegan 2.6-4
```

```
library(dplyr)
```

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

```
## The following object is masked from 'package:ape':  
##  
## where
```

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

```
## The following objects are masked from 'package:base':  
##  
## intersect, setdiff, setequal, union
```

Binary Distance Matrix

```
BinDat=FS
for (row in 1:nrow(FS)) {
  for (col in 1:ncol(FS)) {
    if(BinDat[row,col]>0){
      BinDat[row,col]=1
    }
  }
}

BinDat
```

##	Quadrat	Population	Location	Rosettes	Bolting	Budding	Bud_Flw	Flower	Flw_Sil
## 1	1	1	1	0	0	0	0	0	0
## 2	1	1	1	0	0	0	0	0	0
## 3	1	1	1	0	0	0	0	0	0
## 4	1	1	1	1	1	1	0	0	0
## 5	1	1	1	1	1	1	0	0	0
## 6	1	1	1	0	1	1	0	0	0
## 7	1	1	1	1	1	1	0	0	0
## 8	1	1	1	0	1	1	0	0	0
## 9	1	1	1	0	1	1	0	0	0
## 10	1	1	1	0	0	0	0	0	0
## 11	1	1	1	0	0	0	0	0	0
## 12	1	1	1	0	0	0	0	0	0
## 13	1	1	1	1	1	1	0	0	0
## 14	1	1	1	1	1	1	0	0	0
## 15	1	1	1	1	1	1	0	0	0
## 16	1	1	1	0	0	0	0	0	0
## 17	1	1	1	0	0	0	0	0	0
## 18	1	1	1	0	0	0	0	0	0
## 19	1	1	1	0	0	0	0	0	0
## 20	1	1	1	0	0	0	0	0	0
## 21	1	1	1	0	0	0	0	0	0
## 22	1	1	1	1	1	1	0	0	0
## 23	1	1	1	1	1	1	0	0	0
## 24	1	1	1	0	1	1	0	0	0
## 25	1	1	1	0	0	0	0	0	0
## 26	1	1	1	0	0	0	0	0	0
## 27	1	1	1	0	0	0	0	0	0
## 28	1	1	1	1	1	1	0	0	0
## 29	1	1	1	1	1	1	0	0	0
## 30	1	1	1	1	1	1	0	0	0

##	Sillique	Claytonia_virginiana	Anemone_hepatica	Grass_tuft
## 1	0		1	1
## 2	0		1	1
## 3	0		1	1
## 4	0		1	0
## 5	0		1	0
## 6	0		1	0
## 7	0		0	0
## 8	0		0	0
## 9	0		0	0
## 10	0		0	0
## 11	0		0	0
## 12	0		0	0
## 13	0		0	1
## 14	0		0	1
## 15	0		0	0
## 16	0		0	0
## 17	0		0	0
## 18	0		0	0
## 19	0		0	0
## 20	0		0	0

## 21	0	0	0	0
## 22	0	1	0	0
## 23	0	0	0	0
## 24	0	0	0	1
## 25	0	1	0	0
## 26	0	0	0	0
## 27	0	1	0	0
## 28	0	0	0	0
## 29	0	1	0	0
## 30	0	1	0	0

##	Trillium_grandifolium	Erythronium.trout.lily.	Acer_saccharum
----	-----------------------	-------------------------	----------------

## 1	1	1	1
## 2	1	1	1
## 3	1	1	1
## 4	0	1	0
## 5	0	1	1
## 6	0	1	0
## 7	0	0	1
## 8	0	0	1
## 9	0	0	1
## 10	0	0	1
## 11	0	0	1
## 12	0	0	1
## 13	0	0	1
## 14	0	0	1
## 15	0	0	1
## 16	0	0	1
## 17	0	0	1
## 18	0	0	1
## 19	0	0	0
## 20	0	0	0
## 21	0	0	0
## 22	0	1	0
## 23	0	0	0
## 24	0	1	0
## 25	0	1	1
## 26	0	1	1
## 27	0	1	1
## 28	0	0	1
## 29	0	0	0
## 30	0	0	1

##	Dicentra_cucularia	Bloodroot	Gallium_aparine	Ulmus_americanana	Unknown_1
----	--------------------	-----------	-----------------	-------------------	-----------

## 1	0	0	0	0	0
## 2	1	1	1	1	1
## 3	1	0	1	0	0
## 4	0	0	1	1	0
## 5	0	1	1	1	0
## 6	0	0	1	1	0
## 7	0	0	1	0	0
## 8	0	0	0	0	0
## 9	0	0	0	0	0
## 10	0	0	0	0	0

## 11	0	0	0	0	0
## 12	0	0	0	0	0
## 13	0	0	1	1	0
## 14	0	0	1	1	0
## 15	0	0	1	1	0
## 16	0	0	1	0	0
## 17	0	0	1	1	0
## 18	0	0	0	0	0
## 19	0	0	0	0	0
## 20	0	0	0	0	0
## 21	0	0	0	0	0
## 22	0	0	1	0	0
## 23	0	0	1	0	0
## 24	0	0	0	0	0
## 25	0	0	1	1	0
## 26	0	0	1	1	1
## 27	0	0	1	1	0
## 28	0	0	0	0	0
## 29	0	0	0	0	0
## 30	0	0	1	0	1

Unknown_2 Unknown_3.ranunculaceae. Unknown_4 Dryopteris_marginalis

## 1	0	0	0	0
## 2	0	0	0	0
## 3	0	0	0	0
## 4	1	0	0	0
## 5	1	0	0	0
## 6	1	0	0	0
## 7	0	0	1	0
## 8	0	0	0	1
## 9	0	0	0	0
## 10	0	0	0	0
## 11	0	0	0	0
## 12	0	0	0	0
## 13	0	0	1	0
## 14	0	0	0	0
## 15	0	0	1	0
## 16	0	0	0	0
## 17	0	0	0	0
## 18	0	0	0	0
## 19	0	0	0	0
## 20	0	0	0	0
## 21	0	0	0	0
## 22	0	0	1	0
## 23	0	0	1	0
## 24	0	0	1	0
## 25	0	0	1	0
## 26	0	0	1	0
## 27	0	0	1	0
## 28	0	0	0	0
## 29	0	1	0	0
## 30	0	1	0	0

Ostrich_fern Plantago_lanceolata Violet Rhamnus_frangula Raspberry Unknown_5

## 1	0	0	0	0	0	0
## 2	0	0	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	0	0	0	0
## 5	0	0	0	0	0	0
## 6	0	0	0	0	0	0
## 7	0	0	0	0	0	0
## 8	0	0	0	0	0	0
## 9	0	0	0	0	0	0
## 10	0	0	0	0	0	0
## 11	0	0	0	0	0	0
## 12	0	0	0	0	0	0
## 13	1	0	0	0	0	0
## 14	0	1	1	1	1	0
## 15	1	0	1	1	0	1
## 16	1	0	0	0	0	1
## 17	1	0	0	0	0	1
## 18	1	0	0	1	1	1
## 19	0	0	0	0	0	0
## 20	0	0	0	0	0	0
## 21	0	0	0	0	0	0
## 22	0	0	0	0	0	1
## 23	0	0	0	1	0	1
## 24	0	0	1	1	1	0
## 25	0	0	0	0	0	0
## 26	0	0	0	0	0	0
## 27	0	0	0	0	1	1
## 28	0	0	0	0	0	0
## 29	1	0	1	0	0	0
## 30	0	0	0	0	1	1

##	Unknown_6	Solidago_canadensis	Unknown_7	Dandelion	grass	viccia_cracca
## 1	0	0	0	0	0	0
## 2	0	0	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	0	0	0	0
## 5	0	0	0	0	0	0
## 6	0	0	0	0	0	0
## 7	0	0	0	0	0	0
## 8	0	0	0	0	0	0
## 9	0	0	0	0	0	0
## 10	0	0	0	0	0	0
## 11	0	0	0	0	0	0
## 12	0	0	0	0	0	0
## 13	0	0	0	0	0	0
## 14	0	0	0	0	0	0
## 15	1	0	0	0	0	0
## 16	0	1	0	0	0	0
## 17	0	0	0	0	0	0
## 18	0	0	1	0	0	0
## 19	0	0	0	1	1	0
## 20	0	0	0	1	1	1
## 21	0	0	0	0	1	0

## 22	0	1	0	1	1	0
## 23	0	1	0	1	1	0
## 24	0	1	0	1	0	0
## 25	0	0	0	0	1	0
## 26	0	0	0	0	1	0
## 27	0	0	0	0	0	0
## 28	0	1	0	0	1	0
## 29	0	1	0	0	0	0
## 30	0	0	0	0	1	0
##	herb_robert	thorny_ash	rhamnus_cathartica	rhubarb	Unknown_8	Unknown_9
## 1	0	0	0	0	0	0
## 2	0	0	0	0	0	0
## 3	0	0	0	0	0	0
## 4	0	0	0	0	0	0
## 5	0	0	0	0	0	0
## 6	0	0	0	0	0	0
## 7	0	0	0	0	0	0
## 8	0	0	0	0	0	0
## 9	0	0	0	0	0	0
## 10	0	0	0	0	0	0
## 11	0	0	0	0	0	0
## 12	0	0	0	0	0	0
## 13	0	0	0	0	0	0
## 14	0	0	0	0	0	0
## 15	0	0	0	0	0	0
## 16	0	0	0	0	0	0
## 17	0	0	0	0	0	0
## 18	0	0	0	0	0	0
## 19	0	0	0	0	0	0
## 20	0	0	0	0	0	0
## 21	0	0	0	0	0	0
## 22	1	0	0	0	0	0
## 23	1	1	1	1	0	0
## 24	0	0	0	0	0	0
## 25	1	0	0	0	0	0
## 26	1	0	0	0	0	0
## 27	1	1	0	0	0	0
## 28	1	0	0	0	1	0
## 29	0	0	0	0	0	1
## 30	1	1	0	0	0	0
##	maianthenum_racemosum					
## 1		0				
## 2		0				
## 3		0				
## 4		0				
## 5		0				
## 6		0				
## 7		0				
## 8		0				
## 9		0				
## 10		0				
## 11		0				


```
## 12      0
## 13      0
## 14      0
## 15      0
## 16      0
## 17      0
## 18      0
## 19      0
## 20      0
## 21      0
## 22      0
## 23      0
## 24      0
## 25      0
## 26      0
## 27      0
## 28      0
## 29      0
## 30      1
```

```
BinDat=as.data.frame(BinDat)
```

Make sure all data is numeric

```
#remove quadrat information
FS_nq=FS[,-c(1)]

#Add quadrat as row label
row.names(FS_nq)=FS$Quadrat

#Convert location into numeric
FS_bin_loc=FS_nq

FS_bin_loc$Location[FS_bin_loc$Location == 'i'] <- 1
FS_bin_loc$Location[FS_bin_loc$Location == 'o'] <- 0

#Convert factor to numeric
str(FS_bin_loc$Location)
```

```
## chr [1:30] "0" "0" "0" "1" "1" "1" "1" "1" "1" "0" "0" "0" "1" "1" "1" "0" ...
```

```
FS_bin_loc$Location=as.numeric(FS_bin_loc$Location)
str(FS_bin_loc$Location)
```

```
## num [1:30] 0 0 0 1 1 1 1 1 1 0 ...
```

Calculate Distances

```
#Calculate Bray-Curtis dissimilarity
dist=vegdist(FS_bin_loc, method="bray", binary = F)

#Calculate neighbour joining
tree=nj(dist)
```

Plot Tree

```
ggtree(tree, layout = "rectangular") %<+% FS +
geom_tiplab(aes(colour=Location)) + theme(legend.position = "right")
```

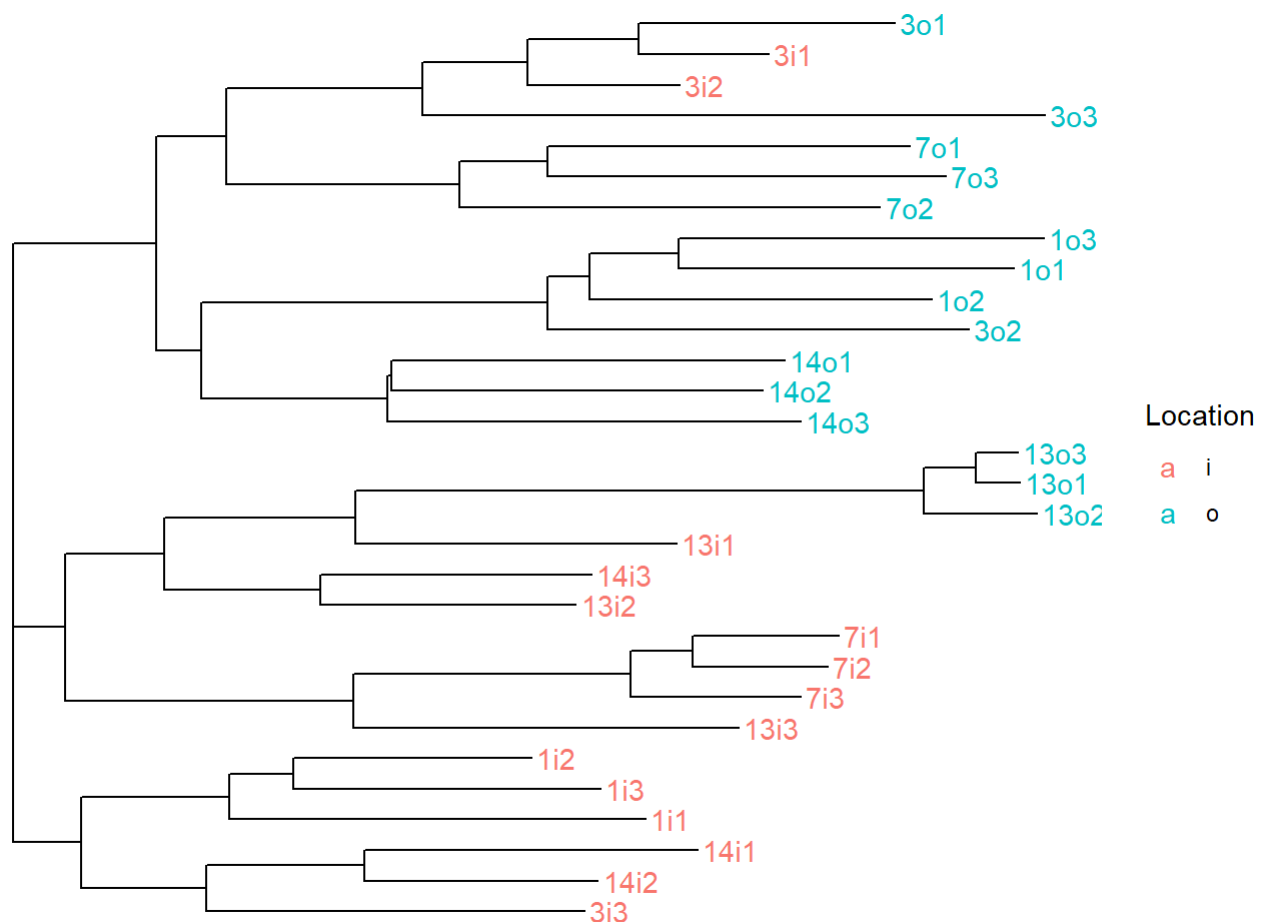


Figure 1: Tree of different quadrates. i indicates that the quadrate patch is inside garlic mustard patches, and o indicates that the quadrate patch is outside the garlic mustard patches. We can see this because the quadrates coloured for i and o are in almost entirely separate clades, except for 13o3, 13o1, 13o2, 3i1, and 3i2. Therefore, this figure demonstrates that garlic mustard has a large effect on the plant community.

Analysis

Based on the results seen in Figure 1, we can conclude that garlic mustard has a large effect on the plant community. This is because the quadrates coloured for i and o are in almost entirely separate clades, except for 13o3, 13o1, 13o2, 3i1, and 3i2. Due to the small number of outlying quadrates, they are likely located in their

positions due to some other property, such as bolding or budding, which, when not accounted for, would likely cause the outliers to move into their respective clades. This finding would further indicate that garlic mustard has a large effect on the plant community.

Question 2

NMDS

```
set.seed(13)
NMDSdat=metaMDS(dist, k=2, trymax = 100)
```

```
## Run 0 stress 0.177062
## Run 1 stress 0.1878424
## Run 2 stress 0.177062
## ... Procrustes: rmse 3.111086e-05 max resid 0.0001234516
## ... Similar to previous best
## Run 3 stress 0.1822645
## Run 4 stress 0.1812268
## Run 5 stress 0.1744617
## ... New best solution
## ... Procrustes: rmse 0.06201532 max resid 0.2113024
## Run 6 stress 0.1822645
## Run 7 stress 0.1822645
## Run 8 stress 0.1744617
## ... Procrustes: rmse 4.452285e-05 max resid 0.0001904364
## ... Similar to previous best
## Run 9 stress 0.177062
## Run 10 stress 0.1744396
## ... New best solution
## ... Procrustes: rmse 0.002791267 max resid 0.01220943
## Run 11 stress 0.1822645
## Run 12 stress 0.177062
## Run 13 stress 0.1812268
## Run 14 stress 0.2795768
## Run 15 stress 0.1844849
## Run 16 stress 0.181227
## Run 17 stress 0.1822646
## Run 18 stress 0.1812265
## Run 19 stress 0.1744617
## ... Procrustes: rmse 0.002795924 max resid 0.0122497
## Run 20 stress 0.1882835
## Run 21 stress 0.1744396
## ... New best solution
## ... Procrustes: rmse 2.122644e-06 max resid 5.307102e-06
## ... Similar to previous best
## *** Best solution repeated 1 times
```

Create data for plotting

```
row.names(FS)=FS$Quadrat
row.names(FS)
```

```
## [1] "7o3" "7o1" "7o2" "7i3" "7i2" "7i1" "3i1" "3i2" "3i3" "3o1"
## [11] "3o2" "3o3" "1i3" "1i2" "1i1" "1o1" "1o2" "1o3" "13o1" "13o2"
## [21] "13o3" "13i2" "13i1" "13i3" "14o2" "14o1" "14o3" "14i2" "14i1" "14i3"
```

```
PDat=data.frame(NMDS1=NMDSdat$points[,1],
                NMDS2=NMDSdat$points[,2],
                Quadrat=row.names(FS))
```

Add labels

```
PDat=merge(PDat, FS, by="Quadrat", all.x=T, all.y =F)
```

Plot

```
PDat$Population=as.factor(PDat$Population)
qplot(x=NMDS1, NMDS2, alpha=I(0.6), data = PDat, shape=Location, colour=Population)+theme_bw()
```

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

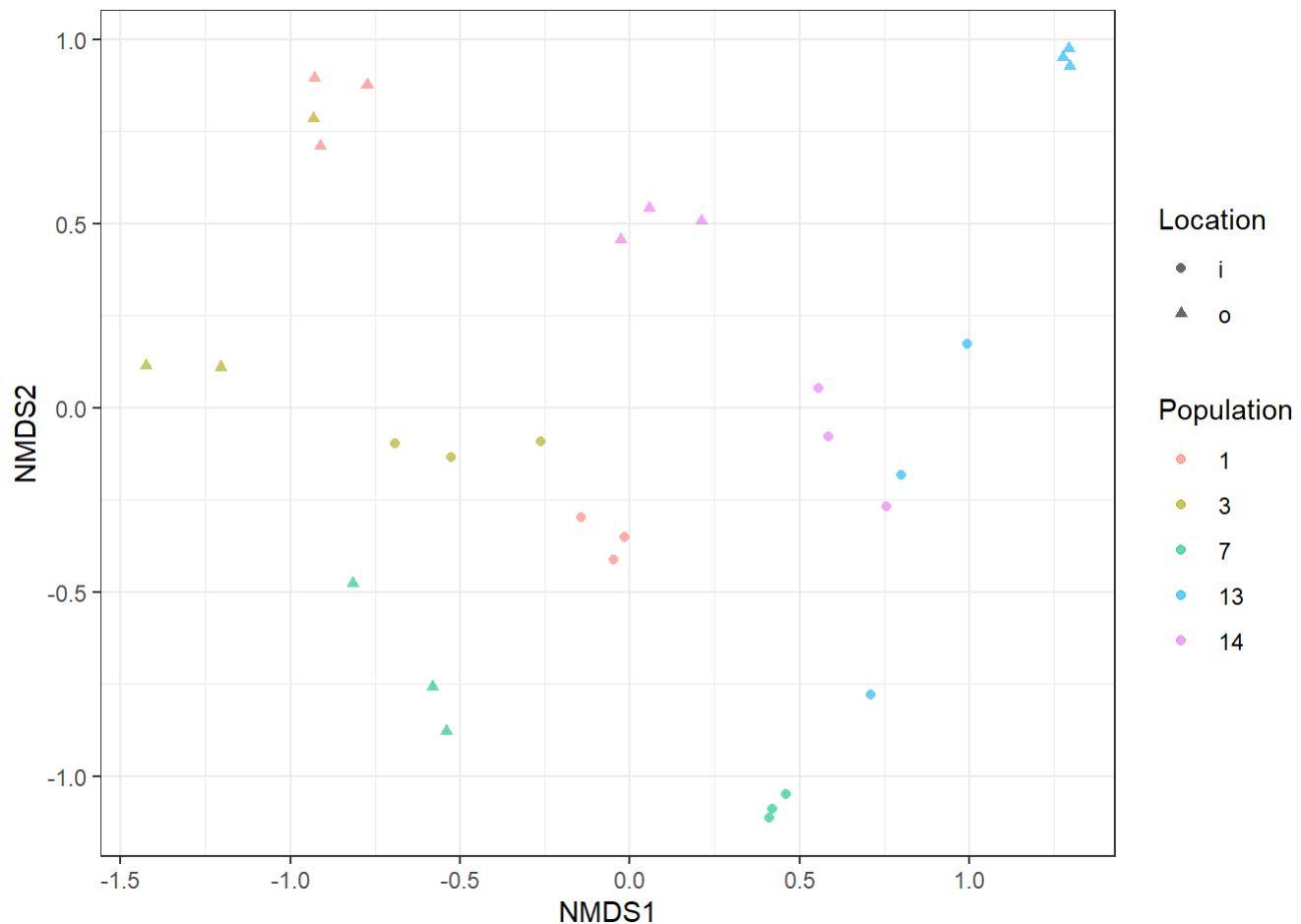


Figure 2: NMDS Plot. i indicates that the quadrat patch is inside garlic mustard patches, and o indicates that the quadrat patch is outside the garlic mustard patches. This figure demonstrates the distinction of samples based on location and population and shows that the locations are well-clustered and divided across the 45-degree line. We can also see that the populations are not as well clustered, with different colours located sporadically across the plot.

Analysis

Based on the results seen in Figure 2, we can conclude that the presence/absence of garlic mustard (location: in(i)/out(o)) has a stronger effect on plant communities than the sampling population. This is seen because the presence/absence of garlic mustard is well clustered and divided across the 45-degree line, with a few deviants. In contrast, populations are not clustered as well, as different colours are located sporadically across the plot.