

Assignment 8 - Caroline Tang 20115082

Project Info

Github repository
Github username: carolinetang77
Date: March 14, 2022

Investigating the effects of garlic mustard on the plant community

Load libraries

```
library(dplyr)
```

```
##  
## 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(ggplot2)  
library(ape)  
library(vegan)
```

```
## Loading required package: permute  
  
## Loading required package: lattice  
  
## This is vegan 2.5-7
```

```
library(ggtree)
```

```
## ggtree v3.2.1 For help: https://yulab-smu.top/treedata-book/  
##  
## If you use ggtree in published research, please cite the most appropriate paper(s):  
##  
## 1. Guangchuang Yu. Using ggtree to visualize data on tree-like structures. Current Protocols in Bioinformatics  
## 2. Guangchuang Yu, Tommy Tsan-Yuk Lam, Huachen Zhu, Yi Guan. Two methods for mapping and visualizing  
## 3. Guangchuang Yu, David Smith, Huachen Zhu, Yi Guan, Tommy Tsan-Yuk Lam. ggtree: an R package for visualizing
```

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

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

```
library(reshape2)
```

Load data

```
floralSurvey <- read.csv("./A8_Tang_Caroline_Data.csv")
```

Data wrangling

I created a dataframe of just the number of species in each area, with the quadrat as the row names.

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

Bray-Curtis Dissimilarity

Using the Bray-Curtis dissimilarity metric, we can compare every combination of quadrats based on abundance of each plant species. If garlic mustard has an effect on the plant community, we should see greater differences between quadrats with no garlic mustard and those with garlic mustard.

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

To see the general differences between quadrats, we can create a tile plot

```
#convert the distance to long format
distMat <- as.matrix(dist)
distLong <- melt(distMat)

#create the tile plot
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))
```

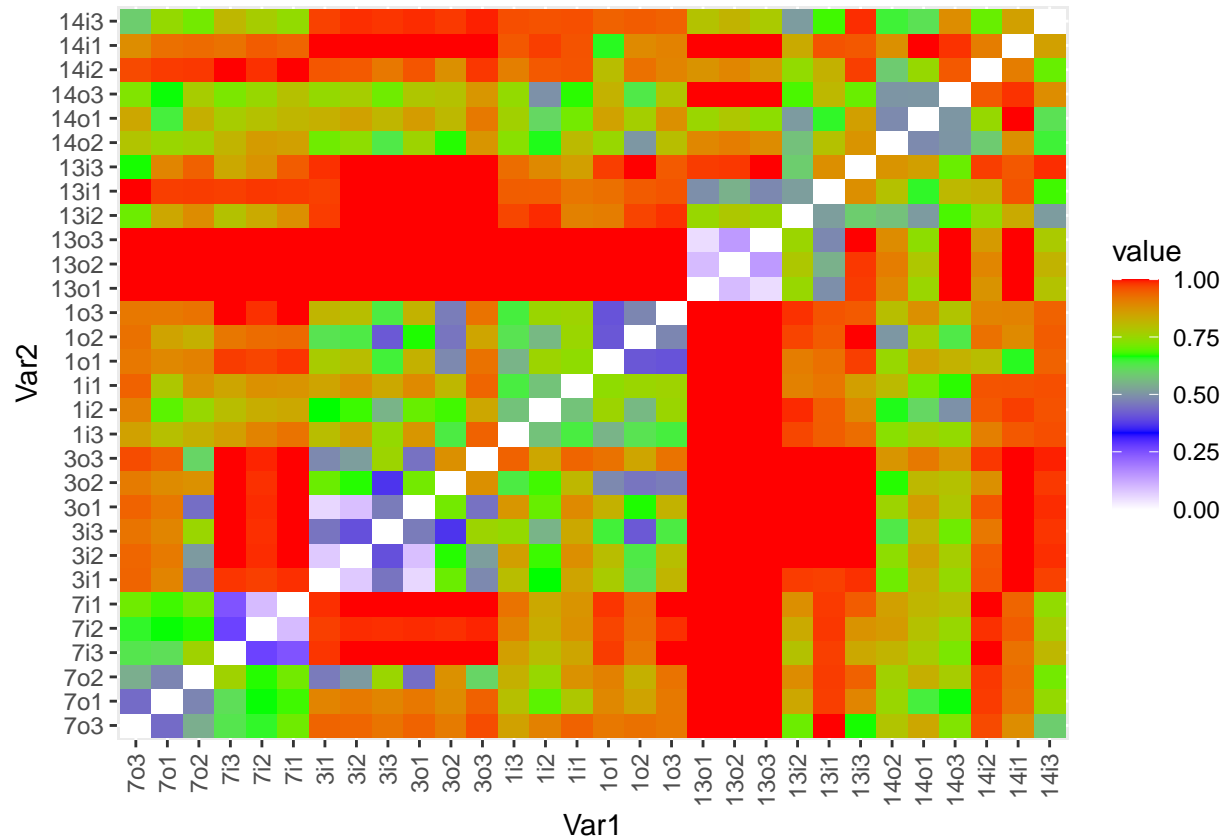


Figure 1. A tile plot illustrating the differences in species abundance between each pair of quadrates

Based on the figure above, sites in the same population and same presence/absence of garlic mustard are the most similar. However, this may be more easily seen through a tree.

Neighbour joining method

I used the neighbour joining method to create the tree, so that more similar sampling sites will be grouped together.

```
floralNJ <- nj(dist)

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

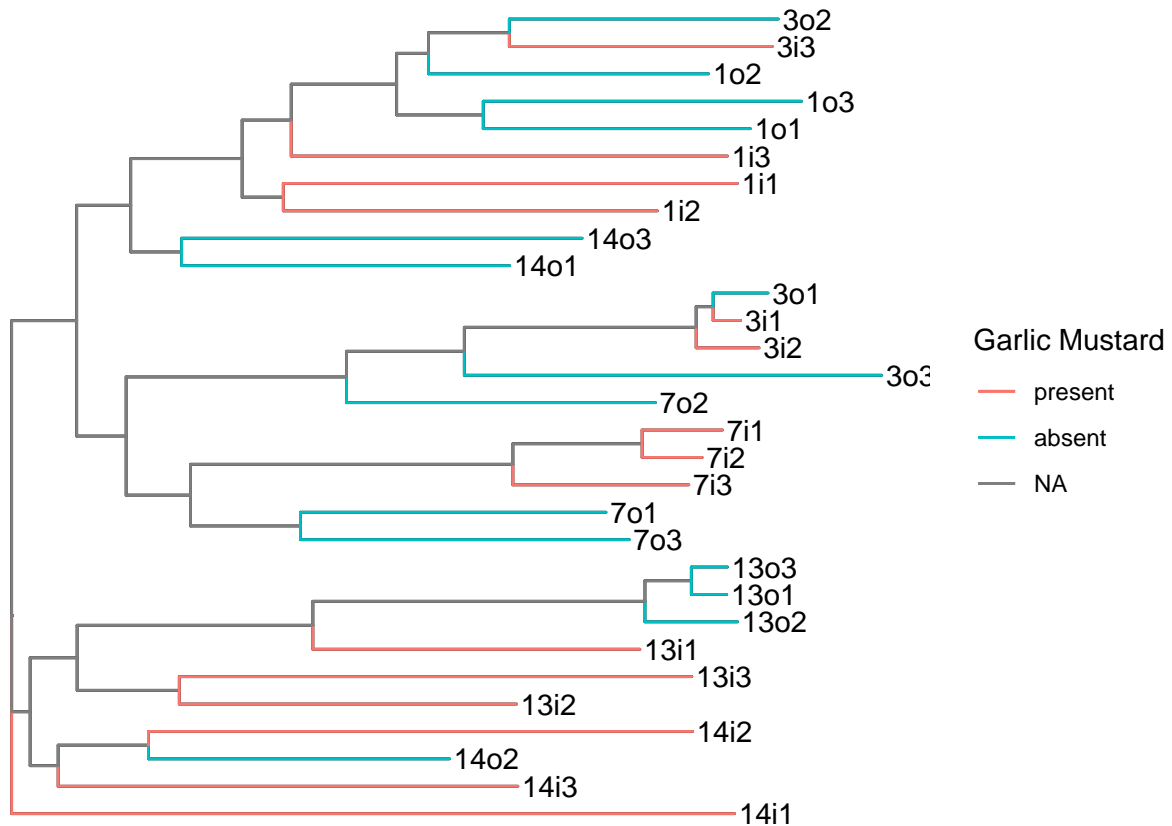


Figure 2. A tree illustrating the similarities between different sampling sites, colour-coded by presence or absence of garlic mustard.

Based on the figure, it appears that sampling sites in the same population are more similar to one another, and the effect of garlic mustard is only similar within the same populations. In addition to tree-based clustering, we can use Non-Metric Dimensional Scaling (NMDS) as a clustering method.

NMDS

```
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
```

```
#format data for plotting
pdat <- data.frame(NMDS1 = floralnmds$points[,1],
                  NMDS2 = floralnmds$points[,2],
                  Quadrate = row.names(floralSubset))

#merge with original data to get plot locations and population IDs
pdat <- merge(pdat, floralSurvey, by = "Quadrate", all.x = T, all.y = T)
```

We can plot the results to see the effect of garlic mustard

```
#plot to see effect of garlic mustard
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()
```

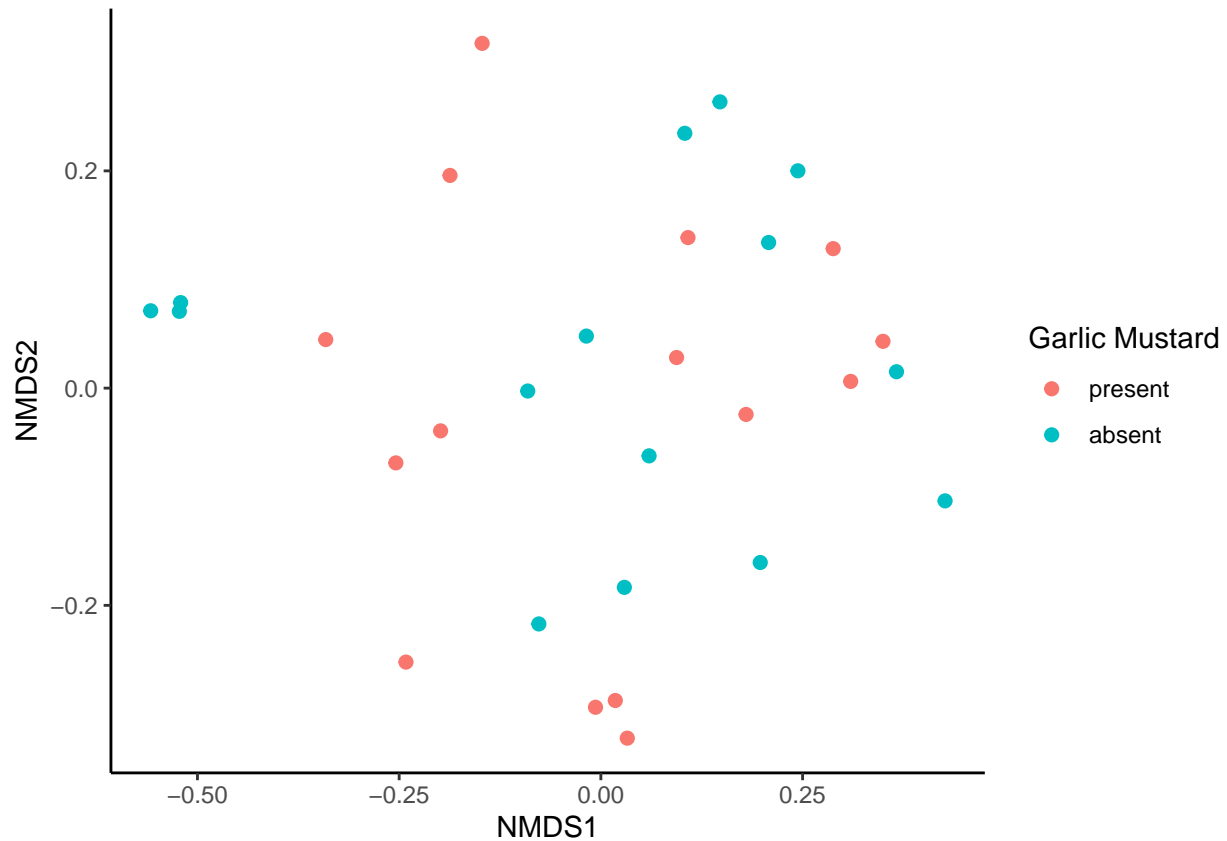


Figure 3. A scatter plot showing the results of the NMDS analysis, colour coded by presence or absence of garlic mustard

Based on the figure above, there doesn't seem to be any noticeable effect of garlic mustard on the plant community as the two groups are not clustered together. However, we can also look at the effects of sampling population, to see if the different populations have a greater influence on the plant community.

```
#plot to see effect of sampling population
ggplot(data = pdat, aes(x = NMDS1, y = NMDS2, colour = factor(Population), shape = Location)) +
  geom_point(size = 2) +
  scale_colour_discrete(name = "Population") +
  theme_classic()
```

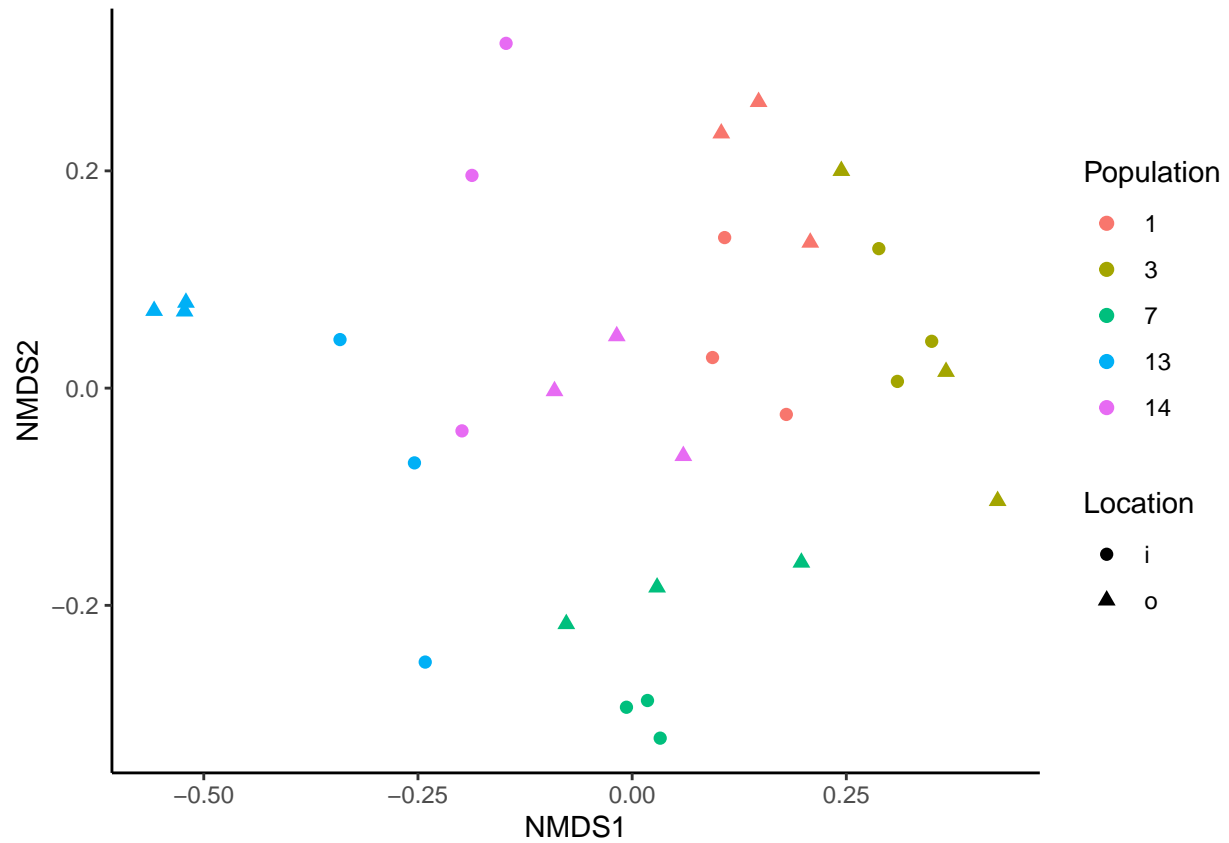


Figure 3. A scatter plot showing the results of the NMDS analysis, colour coded by sampling population. As points from the same populations tend to be closer together, it seems that the sampling population has a greater effect on the plant community than presence or absence of garlic mustard. However, within populations, the sampling sites appear to be somewhat grouped by presence or absence of garlic mustard.