

# A8\_metabar

Qiuhan Zhang

2023-03-08

My Repository (<https://github.com/qiuhan1008/Assignment8.git>)

```
#import dataset
library(readr)
Mydata<- read_csv("FloristicSurvey.csv")
```

```
## Rows: 30 Columns: 44
## — Column specification —————
## Delimiter: ","
## chr (2): Quadrate, Location
## dbl (42): Population, Rosettes, Bolting, Budding, Bud_Flw, Flower, Flw_Sil, ...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#change population to character
Mydata$Population <- as.character(Mydata$Population)
head(Mydata)
```

```
## # A tibble: 6 × 44
##   Quadr...1 Popul...2 Locat...3 Roset...4 Bolting Budding Bud_Flw Flower Flw_Sil Silli...5
##   <chr>    <chr>    <chr>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 7o3      7        o        0        0        0        0        0        0        0
## 2 7o1      7        o        0        0        0        0        0        0        0
## 3 7o2      7        o        0        0        0        0        0        0        0
## 4 7i3      7        i        14        8       157        0        0        0        0
## 5 7i2      7        i         3       18      184        0        0        0        0
## 6 7i1      7        i         0        3      122        0        0        0        0
## # ... with 34 more variables: Claytonia_virginiana <dbl>, Anemone_hepatica <dbl>,
## #   Grass_tuft <dbl>, Trillium_grandifolium <dbl>,
## #   `Erythronium(trout lily)` <dbl>, Acer_saccharum <dbl>,
## #   Dicentra_cucularia <dbl>, Bloodroot <dbl>, Gallium_aparine <dbl>,
## #   Ulmus_americanus <dbl>, Unknown_1 <dbl>, Unknown_2 <dbl>,
## #   `Unknown_3(ranunculaceae)` <dbl>, Unknown_4 <dbl>,
## #   Dryopteris_marginalis <dbl>, Ostrich_fern <dbl>, ...
```

## Bray-Curtis dissimilarity

```
#install.packages("vegan")
library(vegan)
```

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

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

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

```
Subdata <- subset(Mydata[, 11:44])
row.names(Subdata) <- Mydata$Quadrat
```

```
## Warning: Setting row names on a tibble is deprecated.
```

```
S_dist<- vegdist(Subdata, method = "bray", binary = F)
SDistMat <- as.matrix(S_dist)
library(reshape2)
SDat <- melt(SDistMat)
library(ggplot2)
ggplot(SDat, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradientn(colours = rainbow(4), alpha(0.8)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0.5))
```

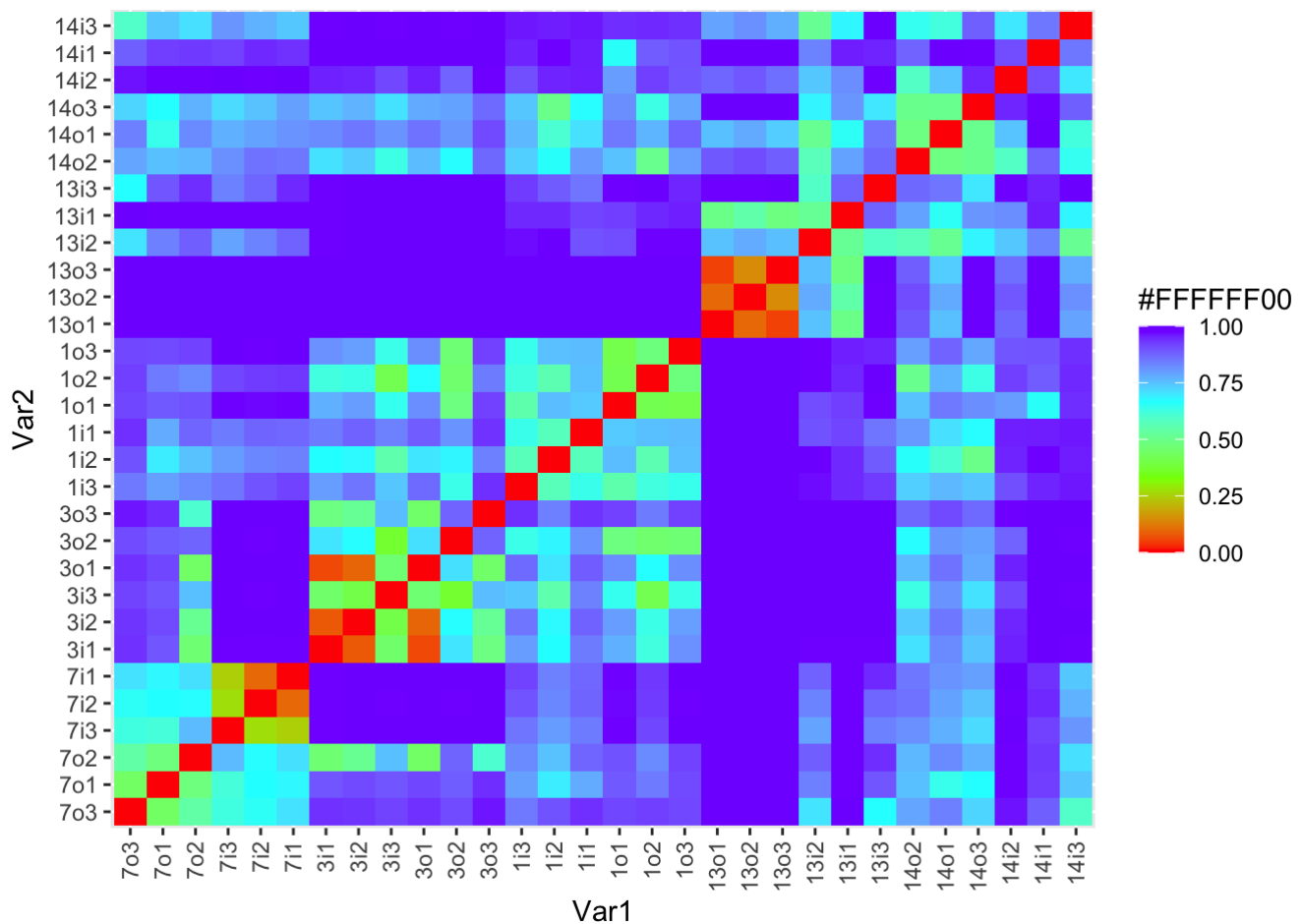


Fig.1. The plot demonstrates the degree of similarity / dissimilarity in the abundance of species across sample locations at QUBS. Color ranges from red to purple represent values from 0 to 1. From 13o1 - 13o3, there's no presence of garlic mustard.

##NJ Tree grouped presence of garlic mustard

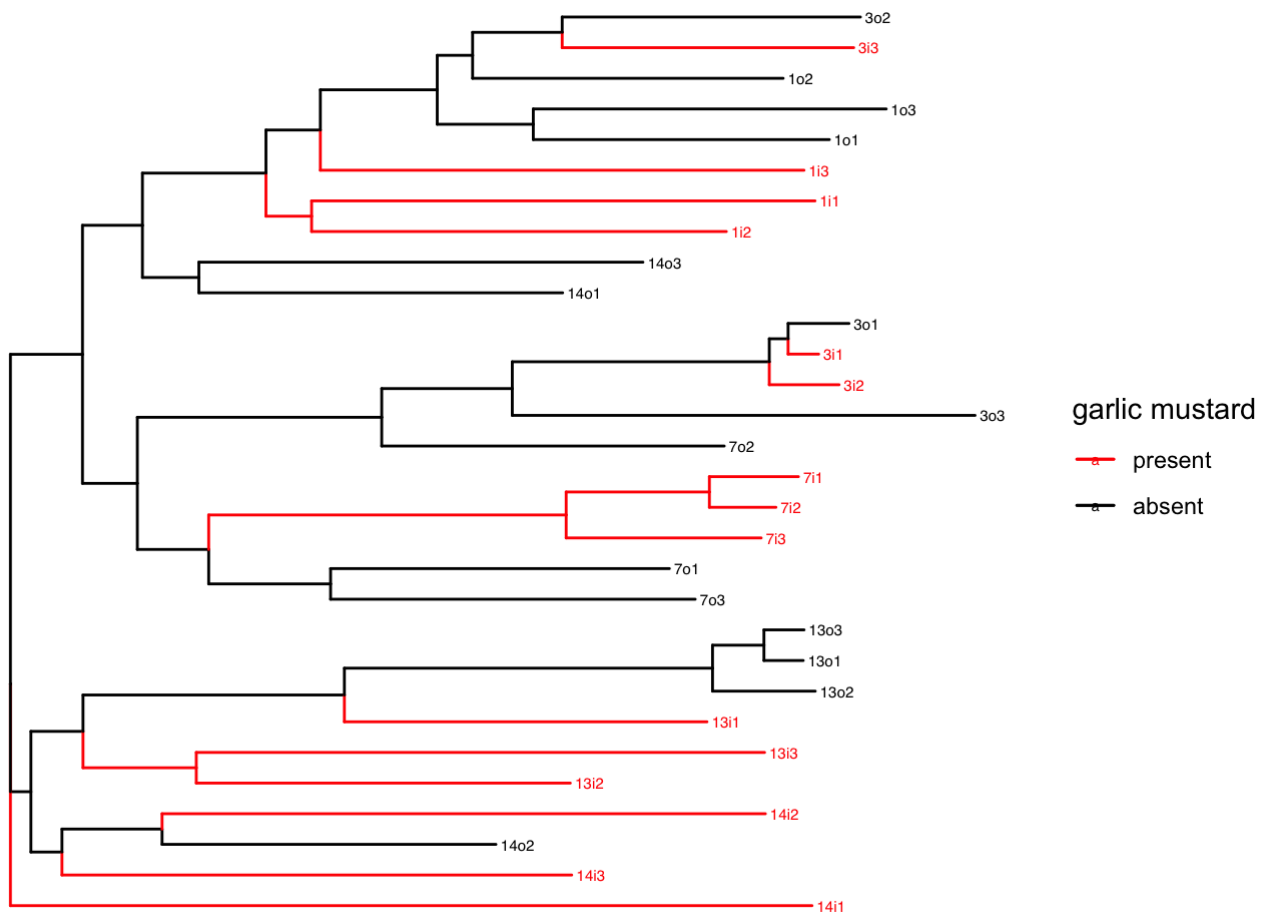
```
library(ape)
library(ggtree)
```

```
## ggtree v3.7.1.002 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
##
## 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
```

```
STreeNJ <- nj(S_dist)
# use gsub to highlight the presentence of garlic mustard as to see if it affects the sp
# ecies abundance.
Pres <- gsub(".*([o|i]).*", "\\1", STreeNJ$tip.label)
# group 'tip.labels' by their corresponding treatment
TreatGroups <- split(STreeNJ$tip.label, Pres)
# use the groupOTU function to apply the grouping information for plotting
SDTcol <- groupOTU(STreeNJ, TreatGroups)
# plot
ggtree(SDTcol, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  theme(legend.position="right") +
  scale_colour_manual(name = "garlic mustard",
                      values = c("red", "black"),
                      labels = c("present", "absent"))
```



```
## NJ Tree grouped by populations
```

Fig.2. The neighbour-joining tree clustered sample depends on the species abundance. The presence of garlic mustard were labeled by red and absent is in black.

```
# use gsub to highlight the presentence of garlic mustard as to see if it affects the species abundance.
Pop <- gsub("(.*)[o|i].*", "\\1", STreenNJ$tip.label)
# group 'tip.labels' by their corresponding treatment
PopGroups <- split(STreenNJ$tip.label, Pop)
# use the groupOTU function to apply the grouping information for plotting
SDTcol2 <- groupOTU(STreenNJ, PopGroups)
# plot
ggtree(SDTcol2, layout = "rectangular", aes(color = group)) +
  geom_tiplab(size = 2) +
  theme(legend.position="right") +
  scale_colour_manual(name = "population",
    values = rainbow(5),
    labels = c("1", "3", "7", "13", "14"))
```

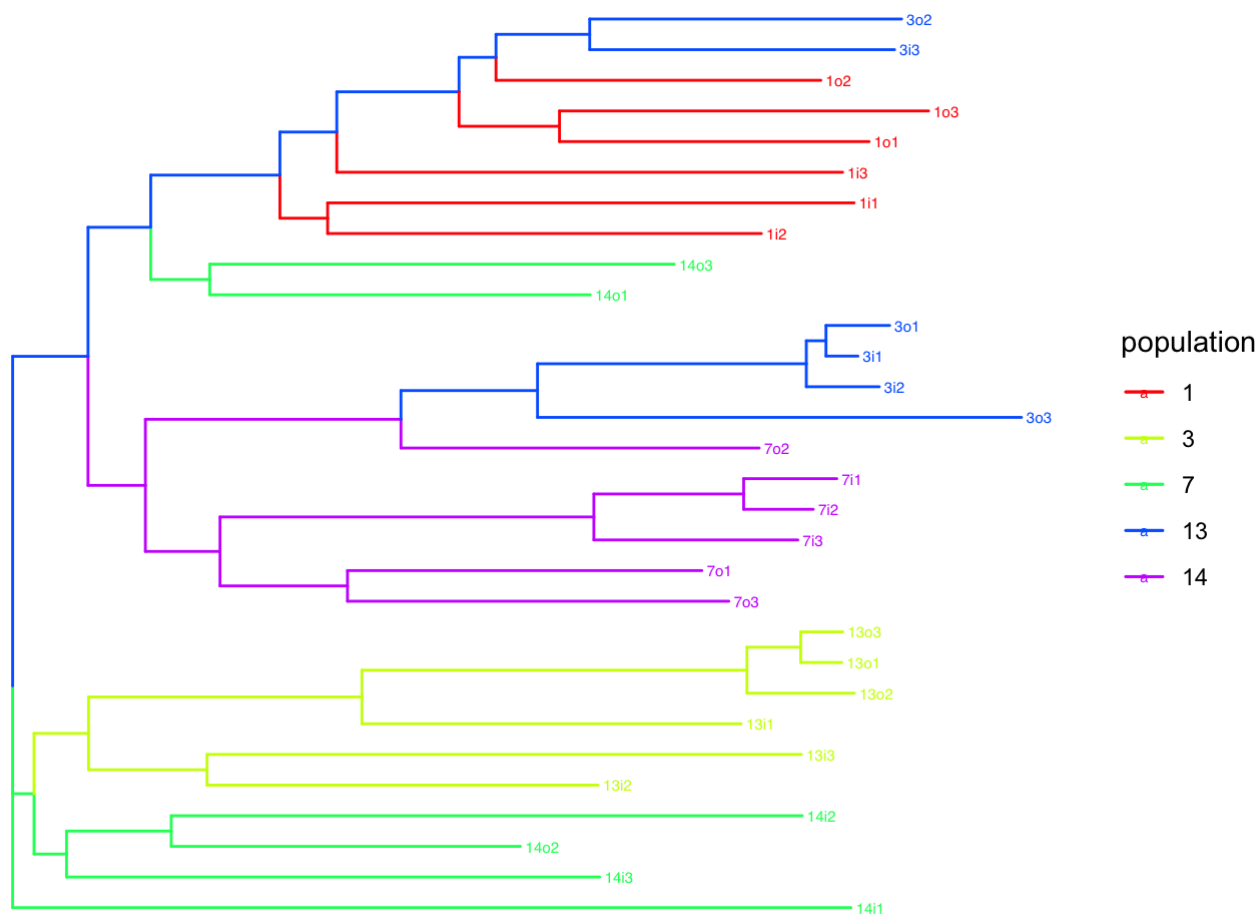


Figure 3. This tree clusters samples based on the similarity of their species abundance. The 5 sampling populations are colorered.

## NMDS

```
set.seed(20)
NMDSdat <- metaMDS(S_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 8.157246e-05 max resid 0.0002927582
## ... 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 0.0001979678 max resid 0.0007305365
## ... Similar to previous best
## *** Best solution repeated 3 times
```

```
SDat2 <- data.frame(NMDS1 = NMDSdat$points[,1],
                    NMDS2 = NMDSdat$points[,2],
                    Quadrate = row.names(Subdata))
SDat2 <- merge(SDat2, Mydata, by = "Quadrate", all.x = T, all.y = F)
qplot(x = NMDS1, y = NMDS2, colour = Population, shape = Location, alpha = I(0.6), data =
SDat2) +
  theme_bw()
```

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

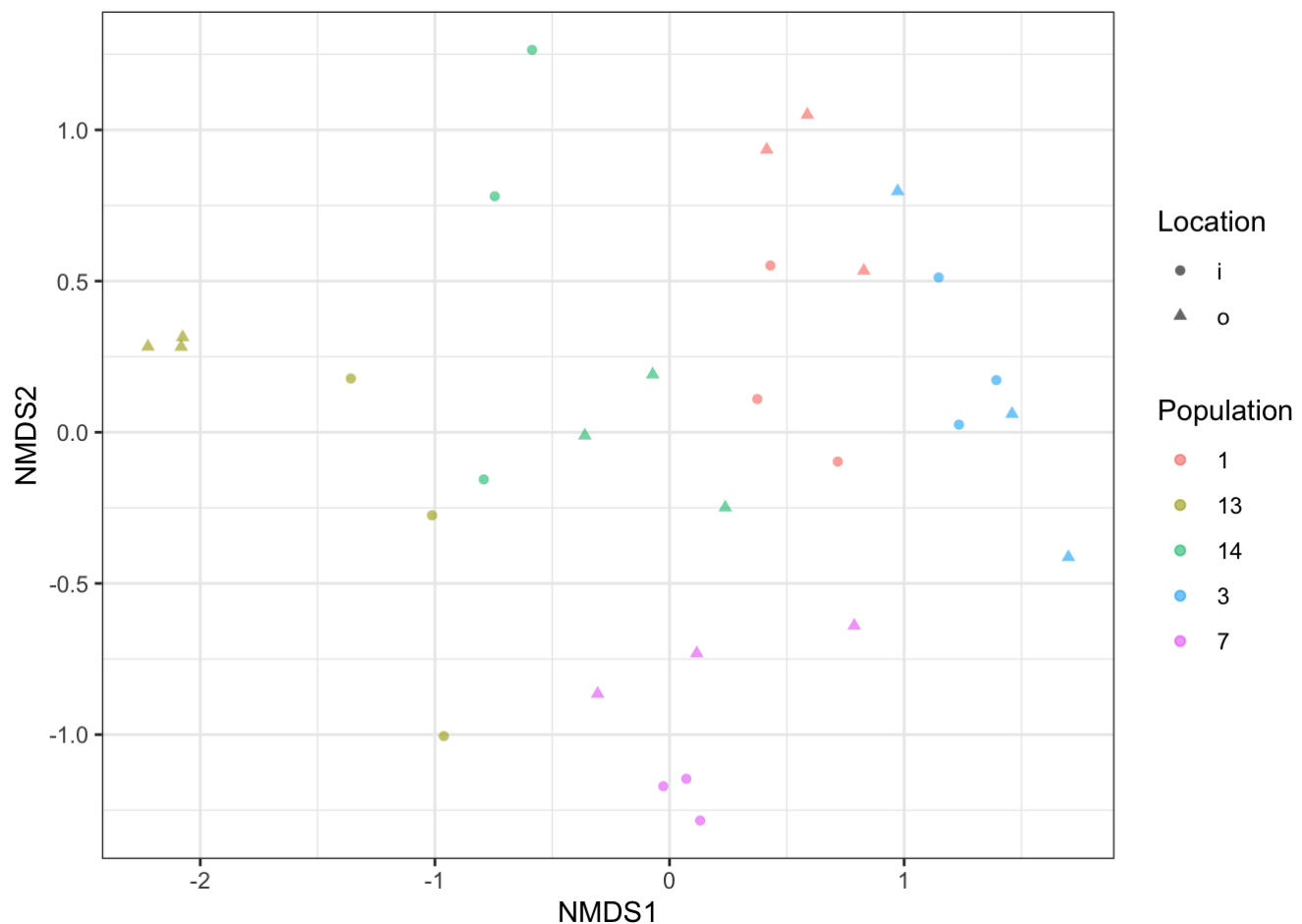


Fig.4.

A NMDS plot showed the variation in species abundance across different populations and the presence of garlic mustard. Different colors indicate different populations, different shapes stand for different locations.

## Report

Garlic mustard (*Alliaria petiolata*) is a native herb to Europe, but an invasive species in North America – it brought negative impact on the local ecological interactions. Based on the distance matrix visualization, the plant species abundance of the communities is different across sample locations especially the population at location 23 where no garlic mustard was present (Fig.1.). The neighbor joining tree (Fig.2.) reveal that the presence of garlic mustard will potentially drive the species abundance differentiation. Fig.3 shows the location can also influence biodiversity. Fig.4 shows that the sampling location has a stronger effect on plant community compared with the introduction of garlic mustard since the points from one population stay together.