Multidimensional Scaling

Dmitry Kondrashov & Stefano Allesina

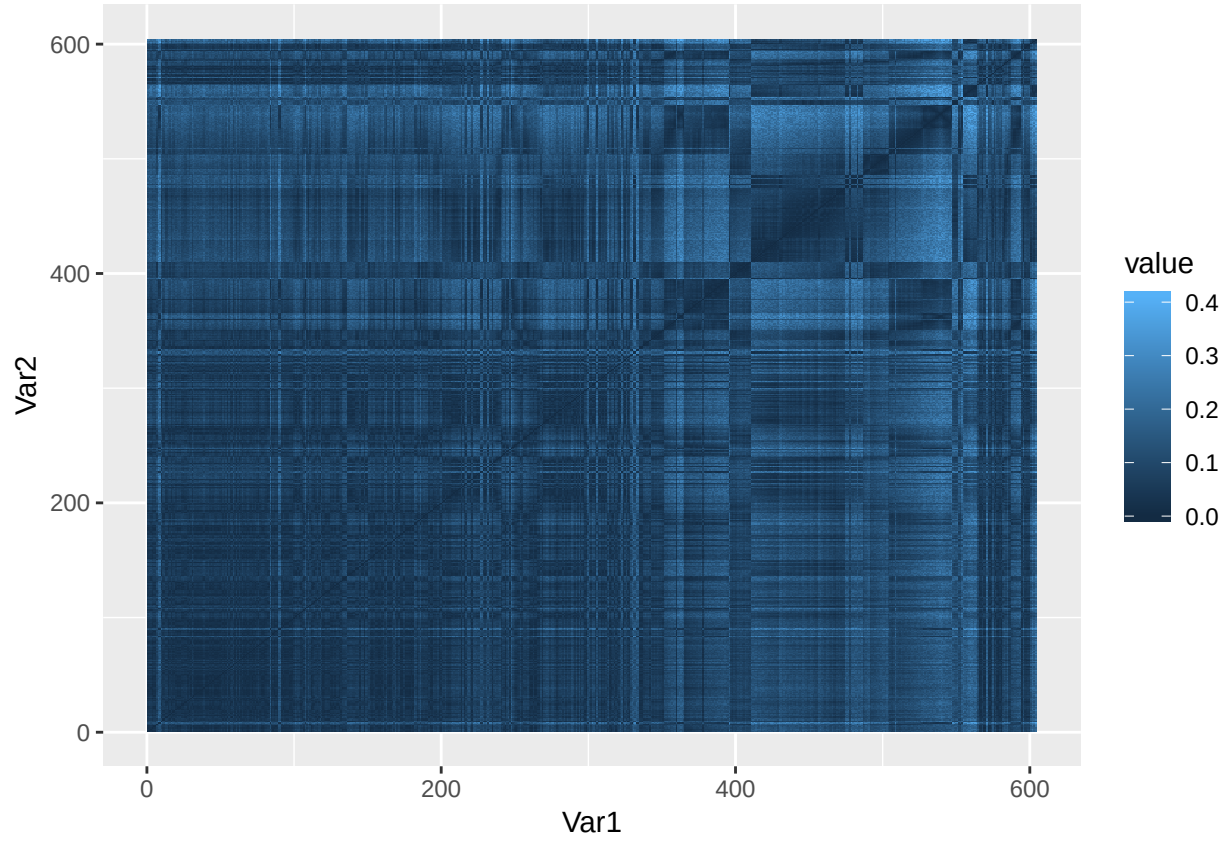
Fundamentals of Biological Data Analysis – BIOS 26318

Contents

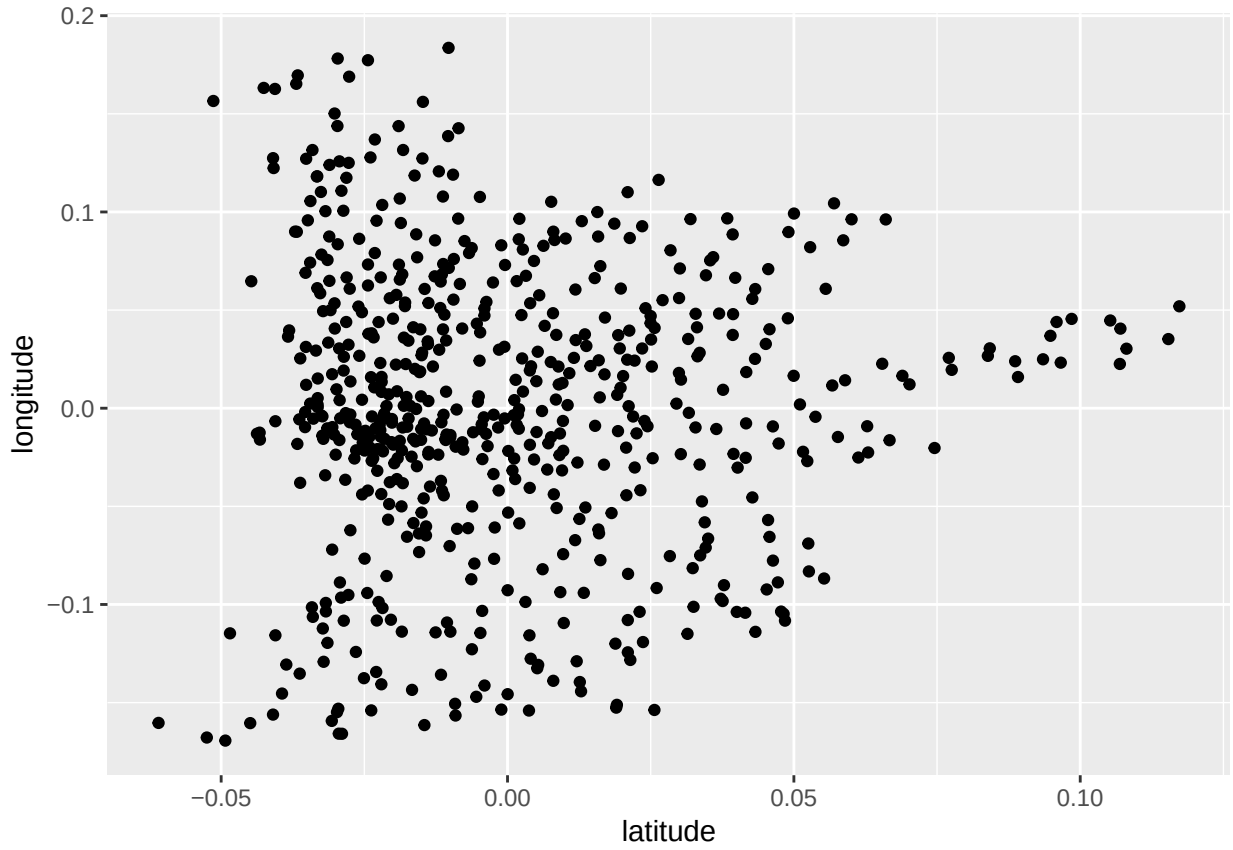
Goal

```
library(tidyverse) # our friend the tidyverse
library(vegan)
#source("general_code/read_xls_from_url.R") # function to read excel from URL
```

```
# load distance matrix
load("data/divvy_stations_distances.RData")
# add some noise to make it more fun
n <- nrow(distance_matrix)
distance_matrix <- distance_matrix * matrix(runif(n * n, min = 0.8, max = 1.2), n, n)
distance_matrix %>% reshape2::melt() %>% ggplot(aes(x = Var1, y = Var2, fill = value)) + geom_tile()
```



```r
# classical MDS
mds_fit <- cmdscale(distance_matrix, k = 2) # k is the dimension of the embedding
mds_fit <- tibble(id = rownames(distance_matrix), longitude = mds_fit[,1], latitude = mds_fit[,2])
mds_fit %>% ggplot() + aes(x = latitude, y = longitude) + geom_point()
```




```r
# aligning coordinates using Procrustes rotation and scaling
# this is the actual location
actual_locations <- read_csv("data/divvy_stations.csv")
new_coord <- procrustes(mds_fit %>% select(latitude, longitude),
                        actual_locations %>% select(latitude, longitude))
```