

Python セミナー練習問題番外編 R で nMDS (nonmetric multidimensional scaling) 解析

Python の `sklearn` でも nMDS は可能ですが、R の方が圧倒的に楽にできます。

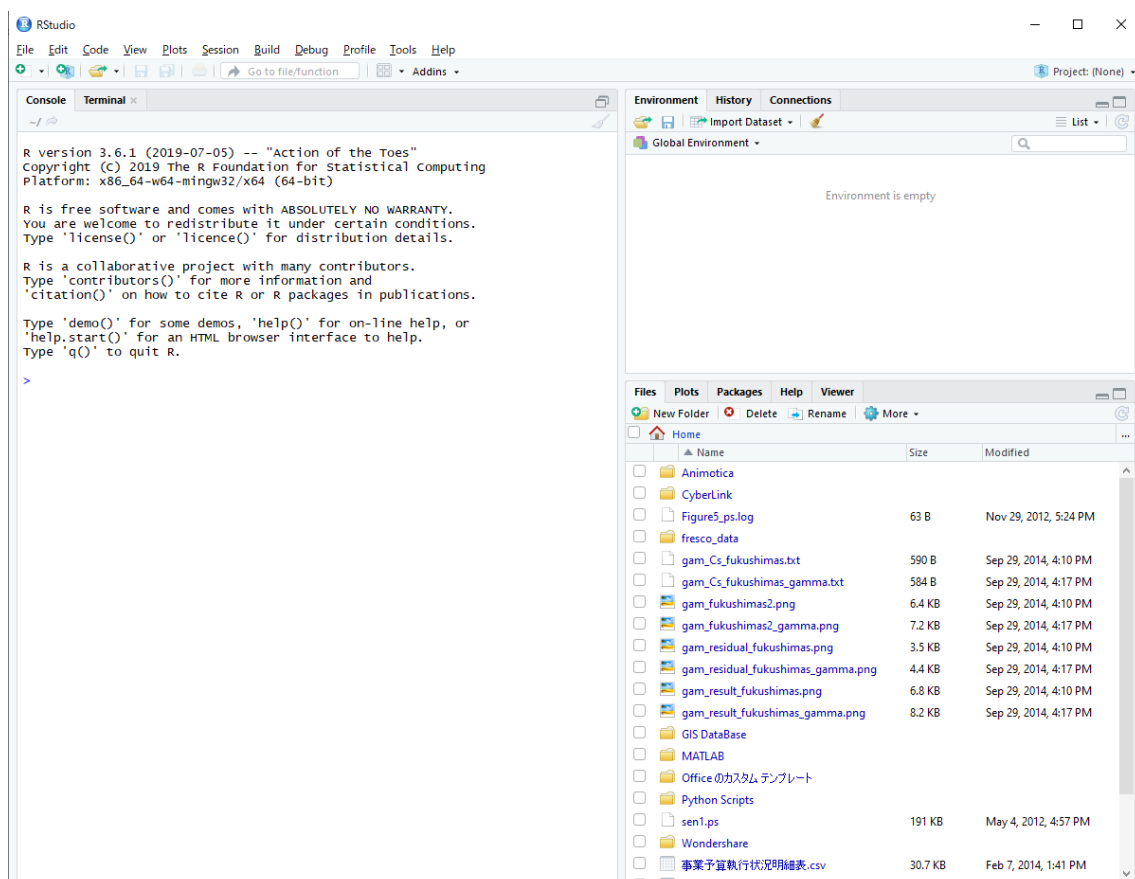
ここでは R で nMDS 解析を行う例を示します。

なお、R を用いた群集組成解析については、水産研究・教育機構の高田宜武さんが詳細な説明を <https://jsnfri.fra.affrc.go.jp/gunshu/> に掲載されているので、参考にしてください。

データは大槌湾で王業浩さんが取得したマイクロプラスチックのデータです。

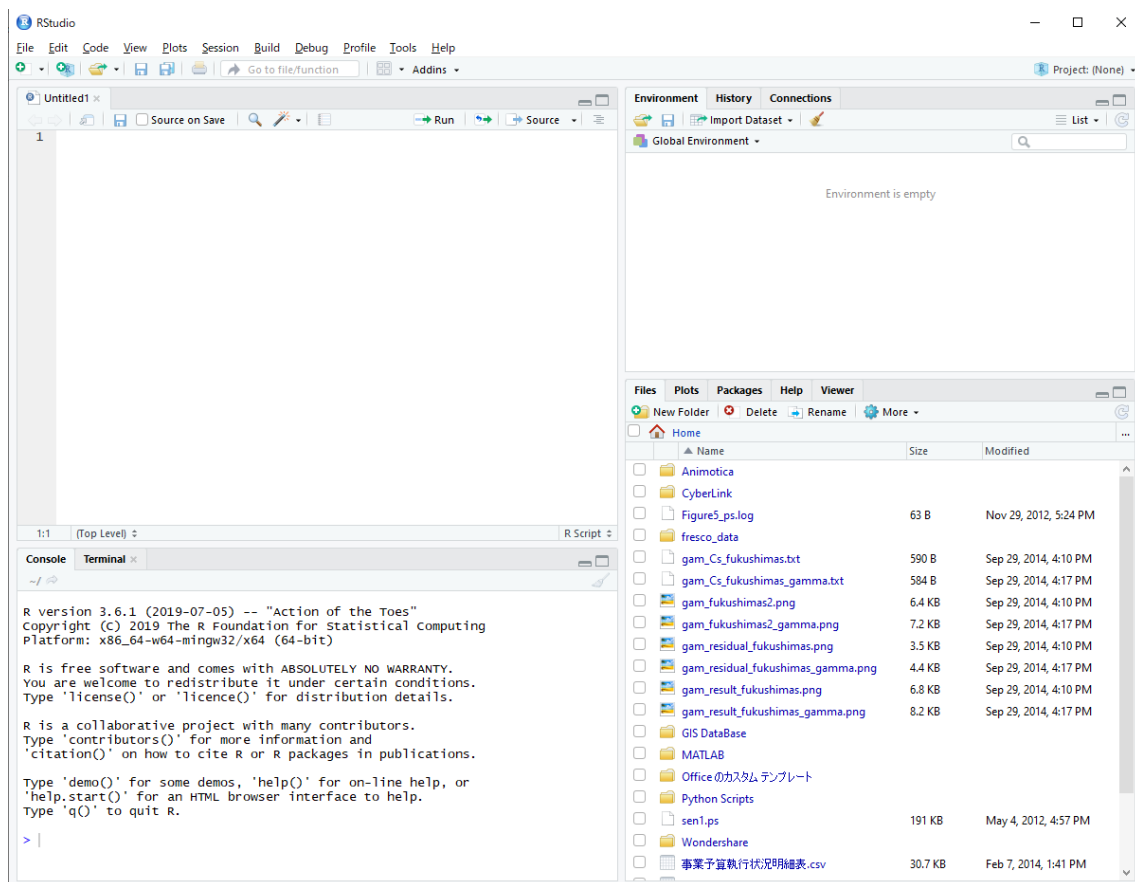
R を使うシステムはたくさんありますが、Anaconda Navigator から RStudio をインストールできるので、ここでは RStudio を使って説明します。

初めて RStudio を立ち上げると下記のような画面になります。



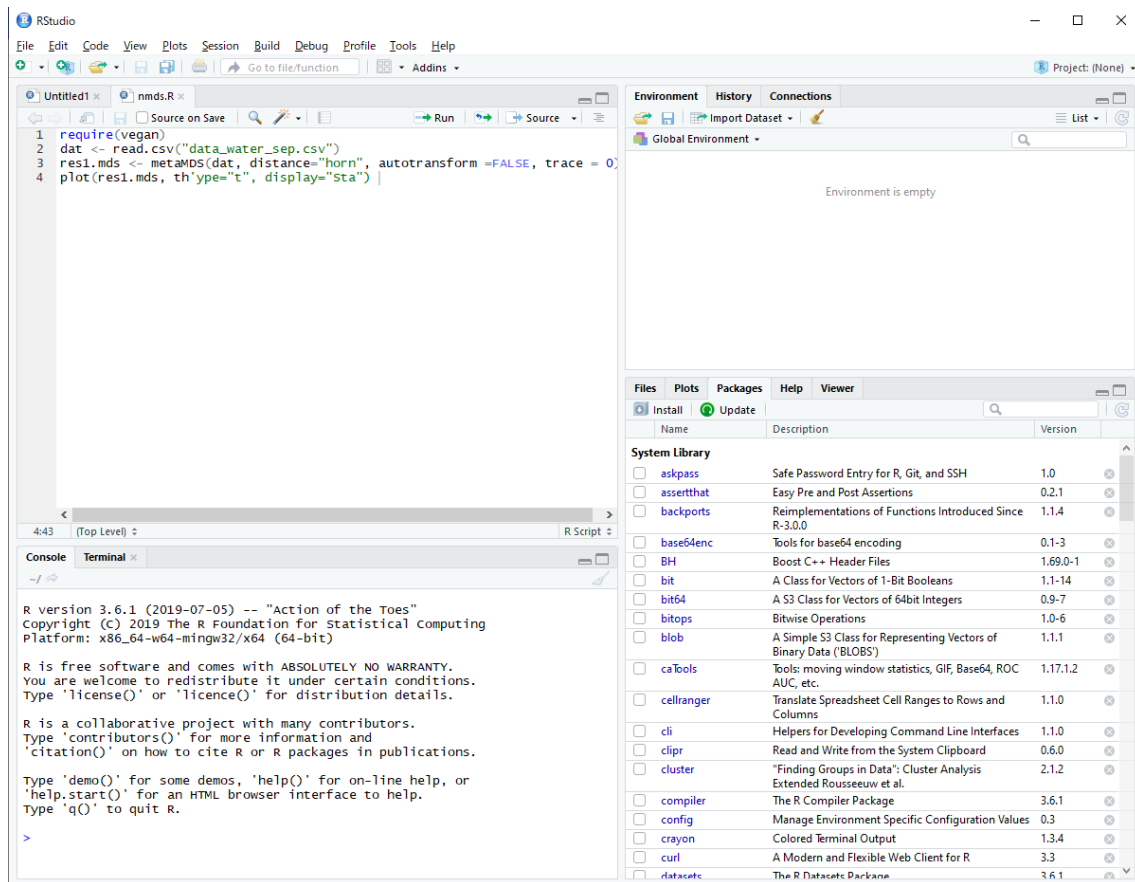
左上の File と Console の間にある緑色の+マークの横の▼を押すと、R Script を選ぶことができます。この R Script にプログラムを入力します。

すると左上に、Script 用の Window が増えます。



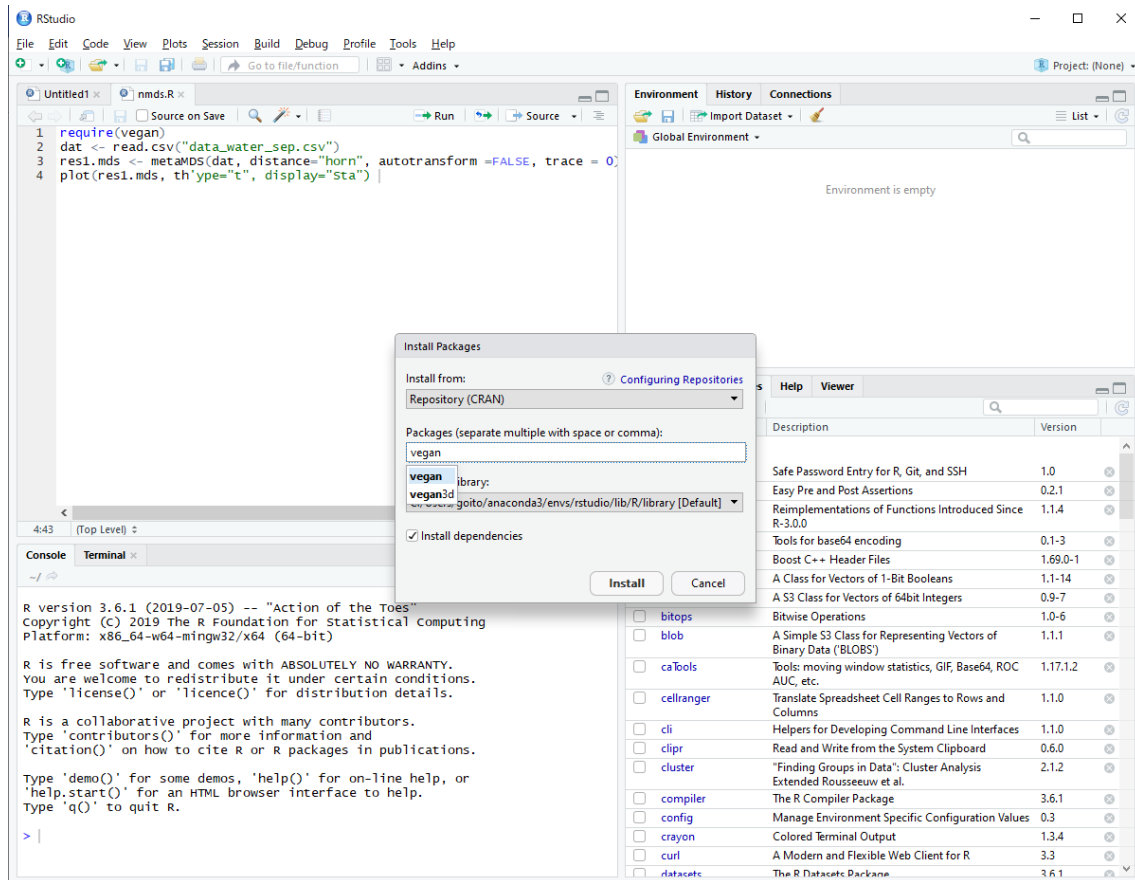
R Script がすでにあり読み込むときは、Code の下にあるフォルダーマークをクリックすると既存の R Script を読み込むことができます。

また、R では様々な Package を使いますが、Package がインストールされていない場合があります。右下のウィンドウで Packages をクリックすると現在インストールされている Package 一覧を見ることができます。



必要な Package がない場合は、右下のウィンドウの Install をクリックし、インストールしたい Package 名を入れて Install ボタンをクリックするとインストールできます。

ちなみに nMDS が入っている vegan という Package はデフォルトでは入っていませんので、自分で Install する必要があります。



では、実際にスクリプトを作って走らせてみましょう。

```

require(vegan)
fig1 <- "nmds.png"
png(fig1, width = 600, height = 600)
dat <- read.csv("data_water_sep.csv", row.names=1)
dat_plastic <- dat[,5:8]
dat_env <- dat[,1:3]
res1.mds <- metaMDS(dat_plastic, distance="bray", autotransform =FALSE, trace = 0,
k = 2)
plot(res1.mds, type="t", display="site", xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
stress = format(res1.mds["stress"],digits=2)
text(c(-0.8), c(1.0), labels=paste("stress=",c(stress)), col="red")
par(new=T)
plot(res1.mds, type="t", display="species", xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
envNMDS<-envfit(res1.mds,dat_env)

```

```
par(new=T)
plot(envNMDS,p.max=0.05, xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
dev.off()
```

と Script ウィンドウに入力してください。

Run ボタンをクリックすると実行できます。

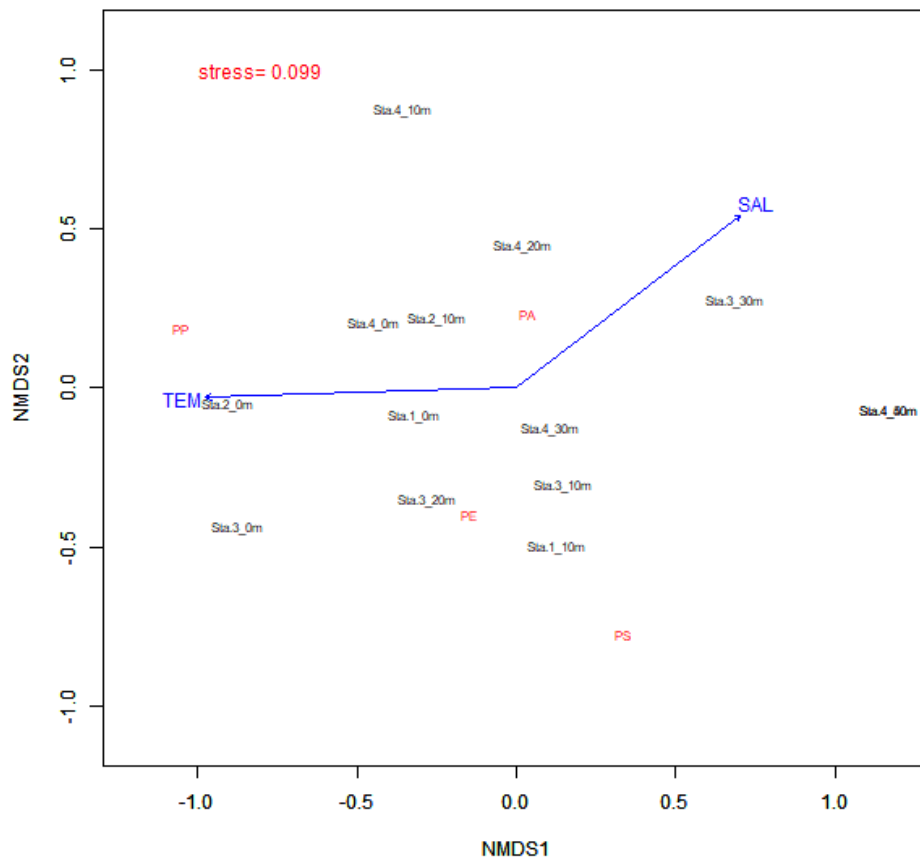
しかし、このまま Run をクリックしても"data_water_sep.csv"が無いとエラーが表示されます。

これは現在 Work directory 上に"data_water_sep.csv"がないためです。

Session ボタンを押し、Set Work Directory ⇒ Choose Directory から "data_water_sep.csv"があるディレクトリーを設定してください。

それから Run をクリックすると実行できます。

結果として、下記のような図が作成されます。



それではコマンドの説明をします。

```
require(vegan)
```

Package `vegan` を読んでいます。

```
fig1 <- "nmds.png"
```

```
png(fig1, width = 600, height = 600)
```

`nmds.png` という図ファイルの用意をします。

```
dat <- read.csv("data_water_sep.csv", row.names=1)
```

```
dat_plastic <- dat[,5:8]
```

```
dat_env <- dat[,1:3]
```

1 列目を行の名前として `data_water_sep.csv` を読み込みます。

プラスチック 4 種のデータが入っているデータと、環境要素が入っているデータに分離します。

```
res1.mds <- metaMDS(dat_plastic, distance="bray", autotransform =FALSE, trace = 0,  
k = 2)
```

プラスチック 4 種のデータを用いて `nMDS` を実施します。

種組成の場合は、距離は **Bray-Curtis** を用いることが一般的です。

`Autotransform` は前述の高田さんの解説ページでは `FALSE` にしてありますが、`vegan` のホームページでは `community data` がない場合に `FALSE` にすると記載されています。今回のデータではどちらで実施しても変化はありませんでした。

`trace` で関数をトレースする度合いを指定します。高田さんの説明ファイルによると `trace=0` としてありますが、`vegan` のホームページでは `trace=2` 以上だと情報が沢山表示されると書かれています。実際に `trace=2` にすると繰り返し回数上限の 20 回までの詳しい内容が表示され、`trace` の値を増やすと表示される内容が増えます。ただし結果は変化ありません。`trace=0` では情報が表示されないというだけのようです。

`k` でいくつの `nMDS` に投影するかを決めます。ここが大きく主成分分析と違う点です。主成分分析では種の数だけモードが存在しますが、`nMDS` では主観で決めてしまいます。口述する `Stress` が小さければ `k` の値に無理がなかったことになります。

```
plot(res1.mds, type="t", display="site", xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
```

`display="site"` を指定すると、それぞれのサンプルが `nMDS` 平面上でどのような位置にあるのかをプロットします。

```
stress = format(res1.mds["stress"], digits=2)
```

```
text(c(-0.8), c(1.0), labels=paste("stress=",c(stress)), col="red")
```

ストレスの値を表示します。ストレスが小さいほどちゃんと投影ができていることになり
ます。ストレスの値が大きい場合は、繰り返し計算回数を決めている `try` や `trymax` を 20
よりも大きくしてみると良いかと思います。

```
par(new=T)
```

重ね書きを指定します。

```
plot(res1.mds, type="t", display="species", xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
```

`display="species"`を指定すると、種が nMDS 平面上でどのような位置にあるのかをプロッ
トします。

```
envNMDS<-envfit(res1.mds,dat_env)
```

環境要素が nMDS 平面でどのように対応するのか、重回帰分析と似たような計算をします。

```
par(new=T)
```

```
plot(envNMDS,p.max=0.05, xlim = c(-1.1,1.1), ylim = c(-1.1,1.1))
```

回帰した際の p 値が 0.05 以下の環境要素のみを nMDS 平面上にプロットします。

```
dev.off()
```

デバイスを `off` にして図を出力します。

```
*****
```

では、もう一度出力された図を見てみましょう。

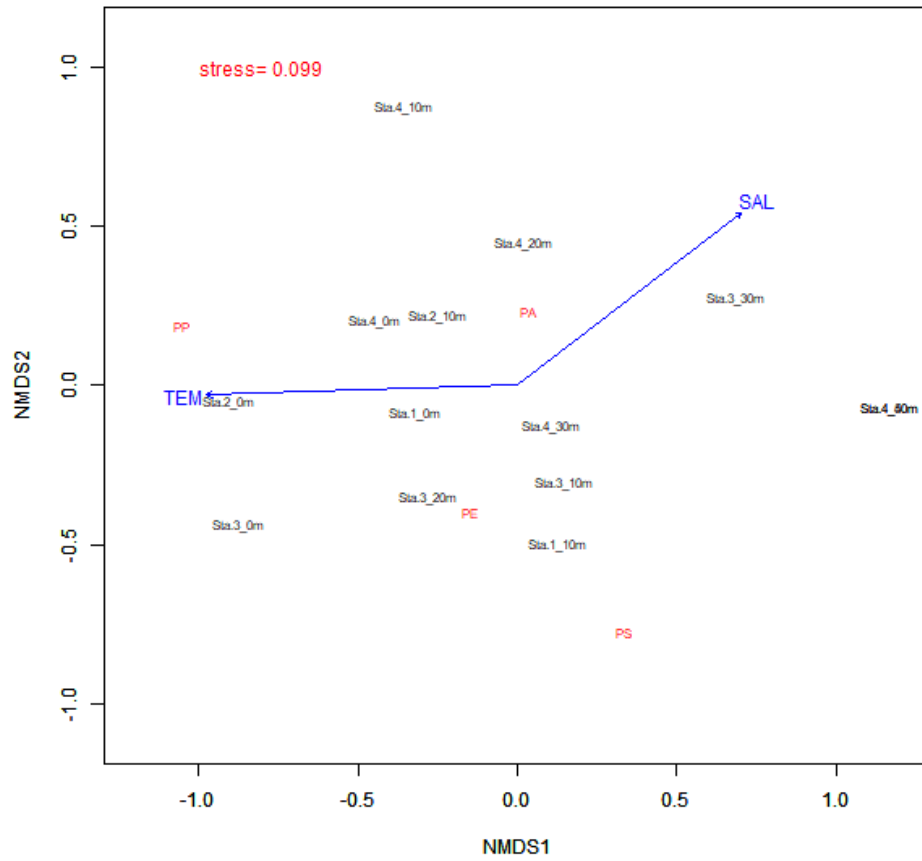
PP (ポリプロピレン) が nMDS1 軸上の負の方向に大きく出ています。この方向に Sta.2
の 0m があるので、Sta.2 の 0m が PP 主体であることがわかります。

同様に 0m のデータが Sta.1 から 4 すべてが nMDS1 の負側になるので、PP は表層に多い
ことがわかります。

そして PP は、水温が高く、塩分が低いところに多いという結果を示しています。塩分が低
いということは外洋由来ではなく、河川由来であることが示唆されます。

nMDS2 軸上では PA (ポリアミド) が正側に、PS (ポリスチレン) , PE (ポリエチレン)
が負側に位置しています。Sta1 の 10m が PS 主体、Sta3 の 20m が PE 主体、Sta4 の 20m
が PA 主体であることが予想されます。

PE が塩分が低いところに多いという結果になるため、PE も河川由来であることが示唆さ
れます。



皆さんも自分のデータを用いて nMDS 解析をしてみましょう。

文責：伊藤進一