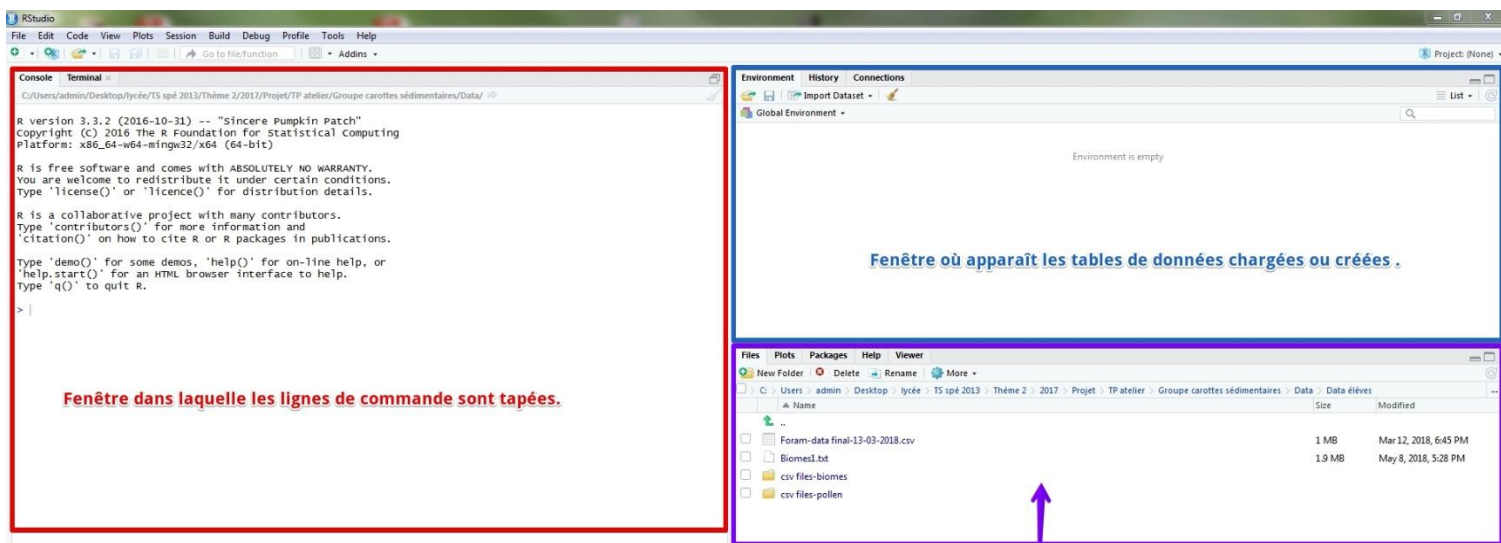


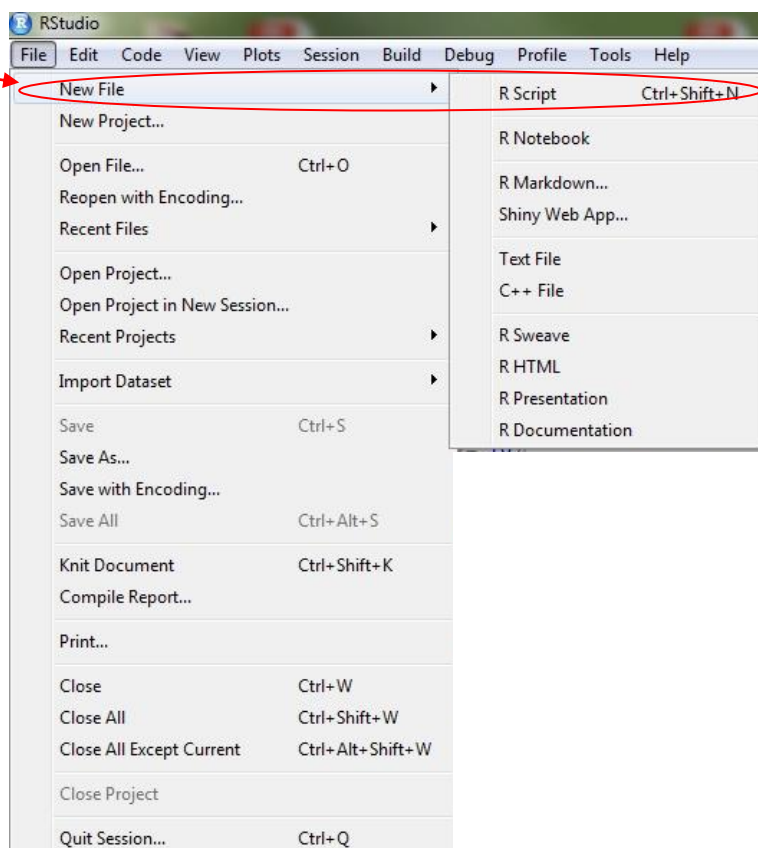
## Protocole pour utiliser Rstudio :

- Choisir l'espace de travail où se trouve les données : cliquer sur Session> Set working directory> choose directory et sélectionner le dossier contenant les fichiers de données

L'interface de R studio est divisée en plusieurs fenêtres ayant des rôles ou apportant des informations différentes :



- Pour écrire un script (lignes de commande) :



- Charger un tableau de données (data) :

Taper la ligne de commande suivante : `data<- read.csv ("nom du fichier", h=T, sep=" ; ")`

Exemple pour la table de données foraminifères : `data<- read.csv("Foram-data.csv", h=T, sep=";")`

- **Exécuter la ligne de commande** : la sélectionner et cliquer sur Run. Son exécution est observée dans la console. Les messages d'erreur s'affichent en rouge.

*Penser à sauvegarder votre script : cliquer sur save as et le nommer*

- **Charger le package plyr pour modifier une table de données :**

Taper la ligne de commande : `library("plyr")`

- **Ajouter à la table de données une colonne zones latitudinales :**

Taper les lignes de commandes suivantes :

```
cat.lat <- c("(a) 90-80°S", "(b) 80-70°S", "(c) 70-60°S", "(d) 60-50°S", "(e) 50-40°S", "(f) 40-30°S", "(g) 30-20°S", "(h) 20-10°S", "(i) 10°S-0°", "(j) 10°N-0", "(k) 20-10°N", "(l) 30-20°N", "(m) 40-30°N", "(n) 50-40°N", "(o) 60-50°N", "(p) 70-60°N", "(q) 80-70°N", "(r) 90-80°N")
min.cat.lat <- seq(-90, 80, by = 10)
max.cat.lat <- seq (-80, 90, by = 10)
cat.lat.dt <- data.frame(cat.lat, min.cat.lat,max.cat.lat)
```

```
vecL <- data$Latitude
data.cat.lat<-lapply(vecL, function(i){
  cat.lat.dt[cat.lat.dt[[2]]<i & cat.lat.dt[[3]]>i ,1
})
data.cat.lat <- unlist(data.cat.lat)
```

`data <- mutate(data, Zone.lat = data.cat.lat)`

- **Sauvegarder la table de données modifiée :**

Taper la ligne de commande : `write.table(data, (data,"nom de la table.csv ou .txt", sep=";")`

Exemple pour la table foraminifères : `write.table(data,"Foram-data1.csv", sep=";")` ou `write.table(data, "Foram-data1.txt", sep=";", quote = F, row.names = F)`

- **Définir des zones temporelles (jeu de données R) :**

Charger la table de données sur laquelle vous voulez travailler.

Taper les lignes de commande suivante :

```
nom <- read.csv("nom de la table.csv ou .txt", h=T, sep=";")
```

```
zoneT1 <- subset(nom de la table, nom de la table $Zone.temps=="X1-X2 ans")
```

Par exemple pour la table de données biomes :

```
biomes <- read.csv("Biomes-data.txt", h=T, sep=";")
```

```
zoneT1 <- subset(biomes, biomes$Zone.temps=="0-10000 ans")
```

```
zoneT2 <- subset(biomes, biomes$Zone.temps=="10000-20000 ans")
```

```
zoneT3 <- subset(biomes, biomes$Zone.temps=="20000-70000 ans")
```

```
zoneT4 <- subset(biomes, biomes$Zone.temps=="70000-130000 ans")
```

Pour explorer les banques et faire de la datavisualisation, on utilise un utilitaire de Rstudio : Rattle

### Protocole pour utiliser Rattle

- Aller dans packages et sélectionner Rattle.
- Taper dans la console : rattle().  
Une nouvelle fenêtre s'ouvre intitulée : Mineur de données R
- **Pour charger les données :**

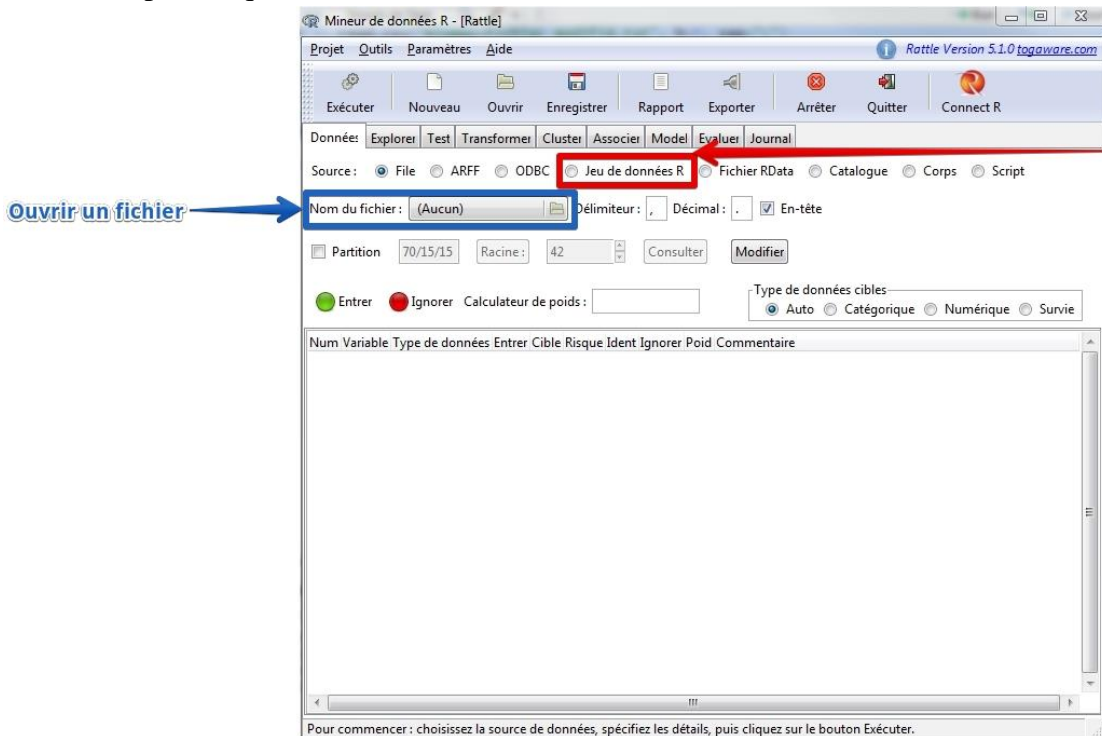
Vous pouvez charger les tables de données présentes dans le bureau de R (ou working directory = votre dossier contenant les tables au format.txt, .csv, .xlsx) :

- Changer le délimiteur en tapant un point-virgule (;) et le décimal en tapant une virgule (,) ou un point(.) pour les fichiers .csv.
- Cliquer sur le dossier en face de la commande « Nom de fichier » et sélectionner dans la fenêtre qui s'ouvre le fichier.
- Cliquer sur Exécuter.

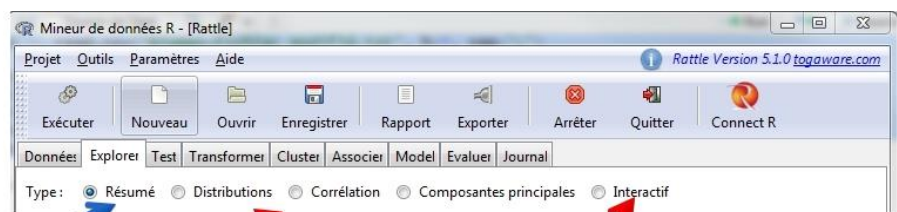
### Cette commande doit être utilisée à chaque nouvelle tâche.

Vous pouvez aussi travailler directement sur des tables de données créées dans la console R et présentes dans « Global Environnement » (exemple : les zones temporelles T1, T2...) :

- Cliquer sur « Jeu de données R »
- Sélectionner dans le menu déroulant « Nom des données » celles que vous voulez explorer puis cliquer sur « Exécuter ».



- **Pour explorer les données :**



Afficher les informations sur la banque de données utilisée (variables...)  
Indispensable pour effectuer un tri des données ensuite

Trier les données pour les traiter et faire une datavisualisation (graphiques, boxplot...)

- Pour trier les données :

**Menu déroulant permettant de choisir des groupes de données catégoriques (= données non numériques)**

**Zone où sont affichées toutes les variables (=données numériques) de la banque de données**

Num Variable	Graphique en boîtes	Histogramme	Cumulatives	Benford	Pairs	Min; Médiane/Moyen; Max
1 site_id	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	6.00; 58.00/53.42; 99.00
3 Sample.depth	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.00; 128.00/177.30; 999.00
4 original_age	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	0.00; 5887.05/5483.83; 9998.30
5 CLAM_min95	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-9999.00; -9999.00/-9076.09; 14791.21
6 CLAM_max95	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-9999.00; -9999.00/-9041.76; 15150.91
7 CLAM_best	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-9999.00; -9999.00/-9061.54; 14970.71
8 CLAM_acc_rate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	-9999.00; -9999.00/-9440.31; 392.55

Num Variable	Graphique en barres	Graphique en points	Mosaïque	Pairs	Niveaux
2 site_name	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	51
9 CLAM_age_model_type	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10
21 Zone.lat	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	10
23 Zone.long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	9
24 Zone.temps	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	5

- Pour visualiser les données (= datavisualisation) :

Cocher les données choisies en face de la représentation désirée (graphique en boîtes (box-plot) ou histogramme) afin de les sélectionner puis cliquer sur exécuter.

La visualisation se fait au niveau de l'interface Rstudio en cliquant sur Plots.

### Signification de ce que représente une box-plot

