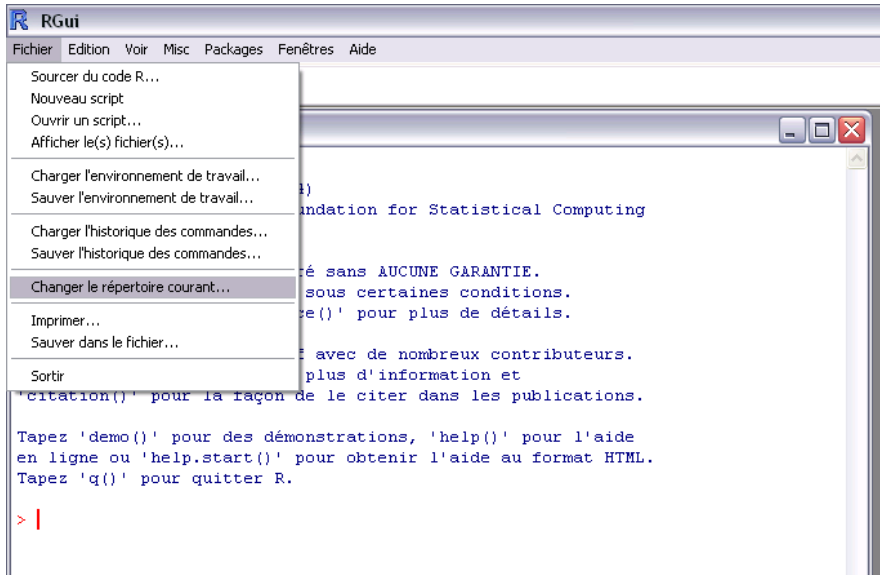


Analyse en composantes principales avec R

1-Changer le répertoire courant

Pour se mettre dans le répertoire qui contient le fichier de données



2-Charger le fichier de données

```
>T=read.table('budget.txt',header=TRUE)
```

T est le tableau des données du fichier *budget.txt* (Vous pouvez choisir un autre nom que *T*) ; *header=TRUE* signifie que la 1^{ère} ligne est l'intitulé des colonnes. Il faut connaître le nom des colonnes pour définir le modèle ci-dessous. Donc bien vérifier que dans le fichier *.txt* des données chaque colonne a un nom.

3-Charger le package nécessaire

Pour effectuer notre ACP sur les données budget, nous allons utiliser le package FactoMineR.

Dans le menu Packages, charger FactoMineR.

Si vous ne le trouvez pas, allez dans le même menu et mettez à jour les packages, puis appelez la bibliothèque correspondante.

```
> library(FactoMineR)
```

4-Effectuer l'ACP

```
> budget.pca<-PCA(T[,1:10])
```

L'intervalle [1 :10] est là pour indiquer qu'il ne faut tenir compte que des 10 premières variables. Les 4 dernières colonnes ne servent qu'à distinguer les individus.

Vous verrez apparaître les deux représentations graphiques des variables et des individus dans le premier plan principal.

5-Affichages supplémentaires

Pour afficher les **valeurs propres** et les **pourcentages d'inertie cumulée** associés, tapez :

> **budget.pca\$eig**

Pour afficher les coordonnées, les contributions et les corrélations des variables, tapez :

>**budget.pca\$var**

Pour les individus, tapez :

>**budget.pca\$ind**

6-Description des données

Présentation des données :

Il s'agit d'une enquête (ONU 1967) sur les budgets-temps (temps passé dans différentes activités au cours de la journée). Le tableau suivant comprend 10 variables numériques et 4 variables catégorisées. Les 10 variables numériques sont: le temps passé en: Profession, Transport, Ménage, Enfants, Courses, Toilette, Repas, Sommeil, Télé, Loisirs.

Les 4 variables catégorisées sont: Le sexe (1=Hommes 2=Femmes), l'activité (1=Actifs 2=Non Act. 9=Non précisé), l'état civil (1=Célibataires 2=Mariés 9=Non précisé), le Pays (1=USA 2=Pays de l'Ouest 3=Pays de l'Est 4=Yougoslavie).

Le code suivant est utilisé pour identifier les lignes: H: Hommes, F: Femmes, A: Actifs, N: Non Actifs(ves), M: Mariés, C: Célibataires, U: USA, W: Pays de l'Ouest sauf USA, E : Est sauf Yougoslavie, Y: Yougoslavie

Les temps sont notés en centièmes d'heures. La première case en haut à gauche du tableau (HAU) indique que les Hommes Actifs des USA passent en moyenne 6 heures et 6 minutes (6 heures + 10/100 d'heure, soit 6 heures et 6mn) en activité PROFESSIONNELLE. Le total d'une ligne (sur ces 10 variables numériques) est 2400 (24 heures).

7-Autres exemples

Vous pouvez appliquer la même méthode sur le fichier **notesCPI.txt** pour retrouver les résultats du cours ou sur le fichier **elections.txt** et commencer à analyser les résultats.