

## Introduction au langage R

On voudrait faire des statistiques sur une base de donnée réelle prise dans le domaine publique. L'université de Berkeley (USA) met à disposition une base de donnée démographique mortality database. Nous avons téléchargé la base de morts en France entre 1899 et 1995 et l'avons transformée au format .CSV pour pouvoir mieux exploiter ces données sous R.

### Exercice 1 - Travailler avec une base de données

1. Charger la base dans RStudio et la mémoriser dans la variable mydata
2. Accéder à la première ligne de la base pour visualiser la structure et les noms des colonnes.
3. Calculer le nombre total de morts enregistrés dans la base.
4. Calculez le nombre total de morts en 1900.
5. Calculez la moyenne d'âge des morts en 1900, 1943, 1960, 1980 et 1990.
6. Modifiez votre programme de manière à que les espérances affichées soient arrondies à deux décimales après la virgule et changez le point (point décimal en notation anglaise) avec une virgule (point décimal en français).
7. Calculez la probabilité de mourir avant la première année de vie pour quelqu'un qui est née en 1899 et comparez-la avec celle de ceux qui sont nées en 1990.

Le langage R met à disposition beaucoup de primitives pour faire du graphisme. Nous rappelons que pour avoir de l'aide sur ces primitives via la fonction help il faut tout d'abord charger la librairie prévue à cet effet :

```
1 library(help="graphics")
```

### Exercice 2 - Un peu de graphisme

1. Dessinez un diagramme à bande qui montre le nombre de morts en 1900 par rapport à leur âge.
2. Ajoutez à votre diagramme le titre "Morts en France en 1900".
3. Ajoutez à ce diagramme une légende qui reporte en abscisse les âges et ordonnées le nombre de morts.
4. Dessinez un diagramme à bande qui montre une comparaison entre les morts en France en 1900 et 1990 par rapport à leur âge. Ajoutez aussi une légenda.
5. A présent, afin de mieux apprécier les différences entre les deux années en question, on souhaite afficher seulement les différences relatives. Colorez aussi les bandes en vert.
6. Si du diagramme précédent on se rend bien compte qu'il y a beaucoup de différence entre les deux années concernant ceux qui sont morts pendant la première année de vie. Justement cette grosse différence rend compliqué l'affichage car toutes les autres différences sont considérablement plus petites. Du coup, essayez d'afficher le même diagramme que précédemment mais en excluant la première année. Que remarquez-vous ?
7. Pour chercher à vérifier l'intuition précédente, reproduisez le même graphe qu'au point précédent mais en excluant les 10 premières années, puis cherchez à mettre en abscisse les années de manière à mieux informer le lecteur. En un deuxième temps, re-calculez aussi les espérances, toujours en excluant la même plage d'années.