

## TP2 : Représentations graphiques

### Exercice 1: fichier de données iris

Le fichier de données iris contient des données collectées par Edgar Anderson, données relatives à des mesures sur des plantes.

1. Charger le fichier de données iris disponible dans le logiciel R et décrire les données
2. tapez les différences instructions suivantes et identifier le résultat obtenu :

```
iris
dim(iris)
names(iris)
str(iris)
iris$Petal.Length
iris$Species
level(iris$Species)
summary(iris$Species)
table(iris$Species)
```

3. Quelles sont les représentations graphiques obtenues par :

```
pie(table(iris$Species))
barplot(table(iris$Species))
```

Ces représentations sont-elles conformes à ce que l'on souhaite?  
Comment corriger éventuellement?

4. Deux botanistes se sont également intéressés aux iris et ont collectés les données suivantes :

```
collection1<-rep(c("setosa","versicolor","virginica"),c(15,19,12))
collection2<-rep(c("setosa","versicolor","virginica"),c(22,27,17))
```

Sur une même fenêtre graphique, représenter ces deux nouvelles distributions et commenter.

5. Maintenant, on s'intéresse à la troisième colonne du fichier.  
Quelle est la nature de cette variable?  
Quelle type de représentation graphique appropriée pourrait-on faire afin de la résumer?  
La faire.
6. Faire de même avec les trois autres variables du fichier

### Exercice 2: Données brutes ou regroupement par classes

On considère le jeu de données suivant, relatif au taux d'hémoglobine dans le sang mesuré en g/L chez des adultes supposés en bonne santé.

Femmes	105	110	112	112	118	119	120	120	125	126
	127	128	130	132	133	134	135	138	138	138
	138	142	145	148	148	150	151	154	154	158
Hommes	141	144	146	148	149	150	150	151	153	153
	153	154	155	156	156	160	160	160	163	164
	164	165	166	168	168	170	172	172	176	179

1. Faire une représentation graphique adaptée de ces données brutes.
2. A présent, on considère les classes  $]104; 114]$ ,  $]114; 124]$ ,  $]124; 134]$ ,  $]134; 144]$ ,  $]144; 154]$ ,  $]154; 164]$ ,  $]164; 174]$ ,  $]174; 184]$ .  
Faire une représentation graphique prenant en compte ces classes. Comparez avec ce qui précède.
3. Sur les données brutes puis sur les données regroupées, tracer la courbe des fréquences cumulées.

### Exercice 3: Influence du nombre de classes et du nombre de données

On considère le jeu de données donnees1.txt.

1. Charger ce jeu de données dans le logiciel R
2. Faire un histogramme de ces données et comparer le résultat lorsque l'on augmente le nombre de classes. Que constatez-vous?
3. Prélever aléatoirement 500 observations du jeu de données initial et faire de même. Constatez-vous la même chose? Pourquoi?

### Exercice 4: Fonction de répartition empirique

On considère de nouveau le jeu de données précédent.

1. On considère dans un premier temps les classes dont la plus petite limite de classe est -4.6, la plus grande 4.6 et les limites intermédiaires étant les éléments d'une suite arithmétique de pas 0.5 et de premier terme -4.6.  
Tracer la courbe des fréquences cumulées associées à ces classes.
2. Faire de même à partir des classes définies par la fonction hist lorsque l'on force l'argument breaks à valoir 100.
3. Tracer la fonction de répartition  $F$  empirique définie sur un jeu de données  $(x_i)_{1 \leq i \leq n}$  par :

$$\forall t \in \mathbb{R}, F(t) = \frac{1}{n} \sum_{i=1}^n 1_{x_i \leq t}$$

$$\text{où } 1_{x_i \leq t} = \begin{cases} 1 & \text{si } x_i \leq t \\ 0 & \text{sinon} \end{cases} .$$

Que constatez-vous?