

Graphiques dans R

Galharret Jean-Michel

Département MSC

Ressources disponibles à l'adresse : https://galharret-github.io/cours_ONIRIS.html

Ouvrez la base de données iris

```
data("iris")
```

On peut faire un résumé de ce jeu de données au moyen de la fonction summary()

```
summary(iris)
```

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
Min. :4.300	Min. :2.000	Min. :1.000	Min. :0.100
1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.:0.300
Median :5.800	Median :3.000	Median :4.350	Median :1.300
Mean :5.843	Mean :3.057	Mean :3.758	Mean :1.199
3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.:1.800
Max. :7.900	Max. :4.400	Max. :6.900	Max. :2.500

Species
setosa :50
versicolor:50
virginica :50

Cas univarié :

On veut uniquement représenter une variable.

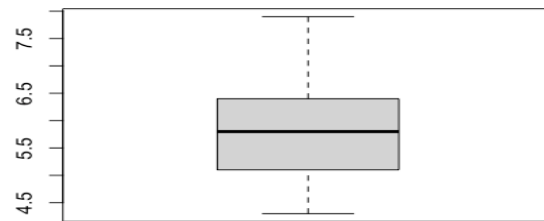
Graphiques pour les variables quantitatives

On utilise principalement les trois types de graphiques suivants :

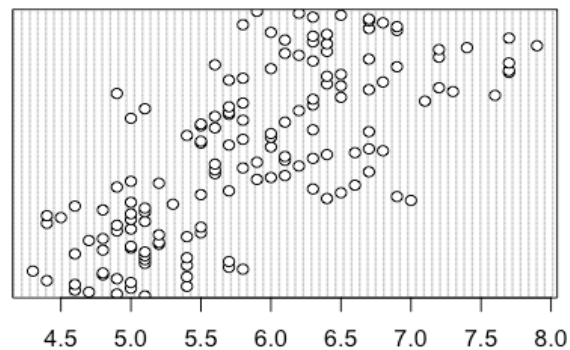
- boxplot (boîte à moustache),
- dotchart (points par points),
- hist (histogramme).

On va s'intéresser uniquement à la longueur des Sépales et réaliser les trois types de graphique :

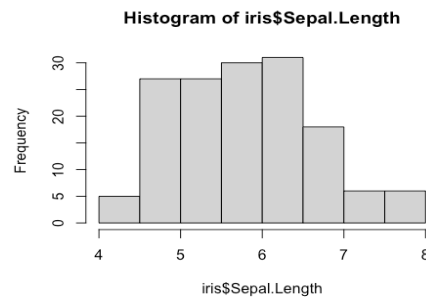
```
boxplot(iris$Sepal.Length)
```



```
dotchart(iris$Sepal.Length)
```

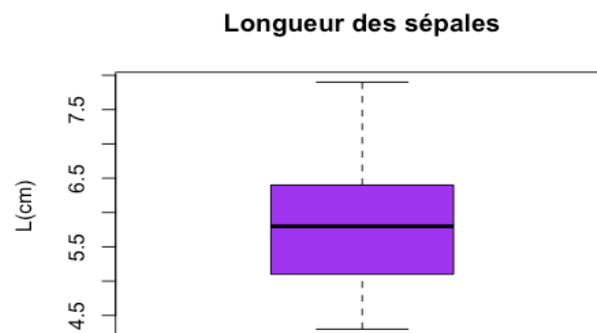


```
hist(iris$Sepal.Length)
```

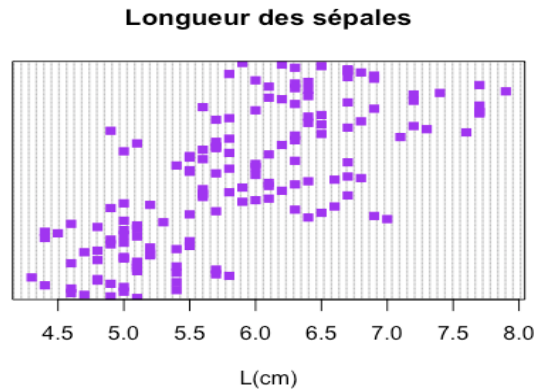


Les graphiques dans R offrent de très nombreux paramètres (titre principal, titre des axes ...)

```
boxplot(iris$Sepal.Length,main="Longueur des sépales",ylab="L(cm)",col="purple")
```

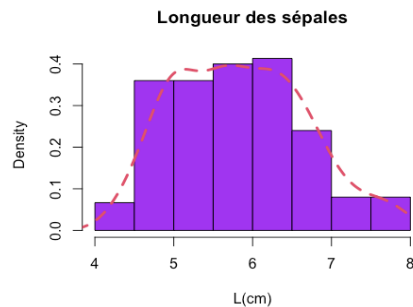


```
dotchart(iris$Sepal.Length,main="Longueur des sépales",xlab="L(cm)",col="purple",pch=15)
```



On peut également superposer plusieurs graphiques grâce à la fonction `lines()` qui permet de dessiner des lignes. Cette fonction devra toujours être précédée d'un graphique. `lwd` = épaisseur du trait, `lty` = type de ligne (continue, pointillés, ...)

```
hist(iris$Sepal.Length,main="Longueur des sépales",xlab="L(cm)",col="purple",prob=T)
lines(density(iris$Sepal.Length),lwd=3,col=2,lty=2)
```

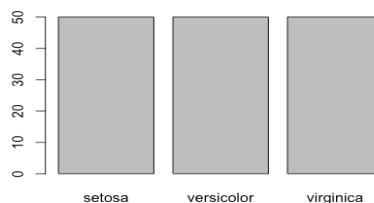


Graphiques pour les variables qualitatives.

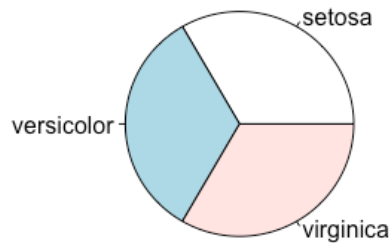
On utilise principalement les deux types de graphiques suivants :

- barplot (diagramme en bâtons),
- pie (camembert ou diagramme circulaire).

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



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



Cas bivarié

On veut représenter une variable en fonction d'une autre.

Une variable quantitative en fonction d'une qualitative

On veut représenter la longueur des sépales en fonction de l'espèce sur un boxplot :

```
boxplot(iris$Sepal.Length~iris$Species)
```

ou bien en points par points (on va utiliser la couleur pour distinguer les espèces) :

```
dotchart(iris$Sepal.Length,col=iris$Species)
legend("bottomright",col=1:3,pch=rep(1,3),legend = levels(iris$Species),cex=0.5)
```

Remarque : Pour la position de la légende c'est une association top ou bottom avec left ou right. Il faut également préciser le type de points avec pch et on peut gérer la taille de la légende avec cex.

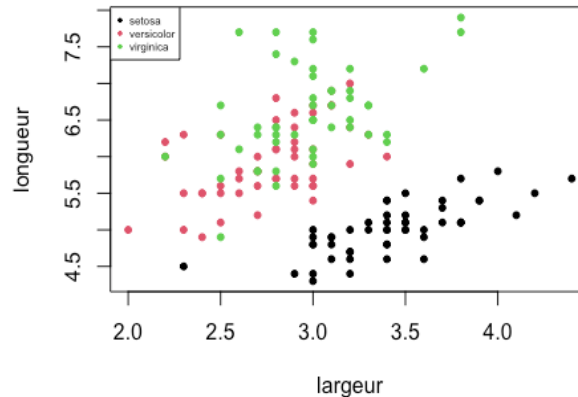
On peut aussi faire plusieurs histogrammes. On va scinder la fenêtre graphique en quatre parties

Une variable quantitative en fonction d'une autre variable quantitative

C'est la fonction plot qui permet de faire ce type de représentation.

On représente la longueur des sépales en fonction de leur largeur (width). Comme précédemment on peut colorer les points en fonction de l'espèce et ajouter une légende.

```
plot(iris$Sepal.Width,iris$Sepal.Length,col=iris$Species,pch=20,
     xlab="largeur",ylab="longueur")
legend("topleft",legend=levels(iris$Species),col=1:3,pch=rep(20,3),cex=0.5)
```



Une variable qualitative en fonction d'une autre variable qualitative

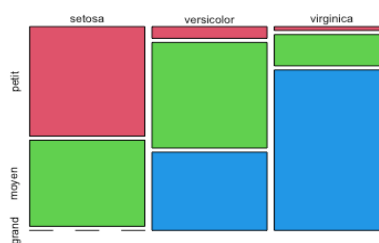
On va créer trois classes d'iris (colonne nommée class) en fonction de la longueur du sépale : petit (inférieur à 5), moyen (entre 5 et 6) et grand (au delà de 6). On utilise la fonction cut de R :

```
iris$class<-cut(iris$Sepal.Length,breaks = c(0,5,6,10))
iris$class<-factor(iris$class,labels=c("petit","moyen","grand"))
```

On représente alors la table de contingence de la classe définie ci-dessus en fonction de l'espèce.

```
mosaicplot(table(iris$Species,iris$class),color = 2:4,main="répartition des c  
lasses en fonction des espèces")
```

répartition des classes en fonction des espèces



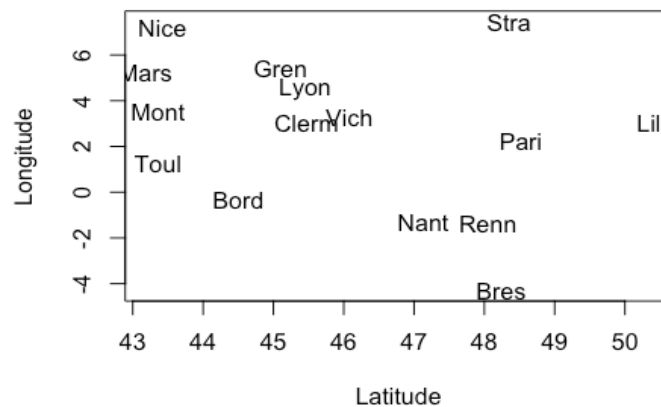
Exercices

Exercice 1 :

Télécharger le fichier temperat1.csv sur CONNECT puis ouvrez le dans RStudio (utiliser l'interface graphique), la première colonne correspondra aux noms des lignes du data.frame.

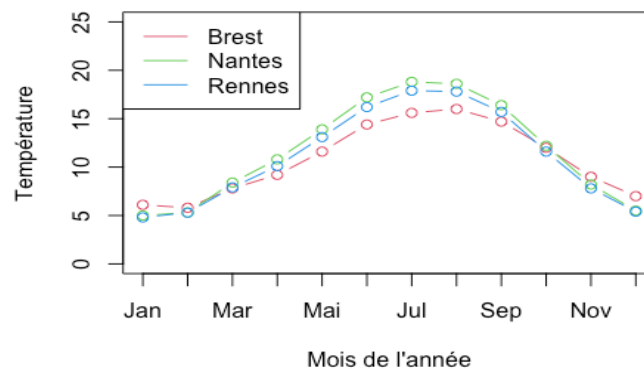
Premier travail : On veut reproduire le graphe suivant :

Villes en fonction de leur latitude et longitude.



aide : pour ne pas avoir de points sur le graphe on utilise `type="n"` et on utilise la fonction `text()` [voir l'aide de R].

Deuxième travail : Représenter les températures moyennes pour les trois villes : Nantes, Brest et Rennes.



Pour mettre les mois en labels sur l'axe des abscisses on utilisera l'argument `xaxt="n"` dans `plot` et ensuite on utilisera la fonction `axis()` voir l'aide de R. L'argument `type="b"` pourra être utilisé dans `plot` et dans `lines` pour dessiner les points et les courbes.

Exercice 2 :

Télécharger sur Connect le fichier `bordeaux1.csv` et ouvrez le dans RStudio à l'aide de l'interface graphique. Les années seront utilisées pour nommer les lignes.

Il s'agit des données des vins de Bordeaux issues de "Méthodes statistiques en gestion" (M. Tenenhaus, Dunod, 1994).

On cherche à étudier l'éventuelle relation entre la qualité du vin et les variables météorologiques. Les variables sont les suivantes : TEMPERATURE : somme des températures moyennes journalières, SOLEIL : durée d'insolation, CHALEUR : nombre de

jours de grande chaleur, PLUIE : hauteur des pluies, Y (QUALITE DU VIN) : 1 (Bon), 2 (Moyen), 3 (Médiocre).

