Fiche TD avec le logiciel R : bem2

Des fleurs et des mannequins

A.B. Dufour, J.R. Lobry, D.Chessel

Table des matières

1 Des fleurs

FIGURE 1 – Sir R.A. Fisher (1890-1962)

Les données utilisées ici sont célèbres. Elles ont été collectées par Edgar Anderson [1]. Ce sont les mesures en centimètres des variables suivantes : longueur du sépale (`Sepal.Length`), largeur du sépale (`Sepal.Width`), longueur du pétale (`Petal.Length`) et largeur du pétale (`Petal.Width`) pour trois espèces d'iris : *Iris setosa, I. versicolor et I. virginica.*

Sir R.A. Fisher a utilisé ces données pour construire des combinaisons linéaires des variables permettant de séparer au mieux les trois espèces d'iris [2].

FIGURE 2 – *I.setosa, I.versicolor, I.Virginica*

1.1 Consignes

La manière la plus simple pour se familiariser avec R est de l'utiliser afin de comprendre un jeu de données particulier. Considérons donc les données provenant des iris de Fisher, données sur lesquelles vous pouvez avoir envie de faire une analyse ...de données. Suivez pas à pas les étapes de la session ci-dessous et voyez ce qui se passe. Faites les exercices proposés et n'hésitez pas à utiliser l'aide en ligne de R. Si vous voulez par exemple connaître le contenu de la fonction `hist`, il vous suffit de taper la commande `?hist`. Vous ne comprendrez

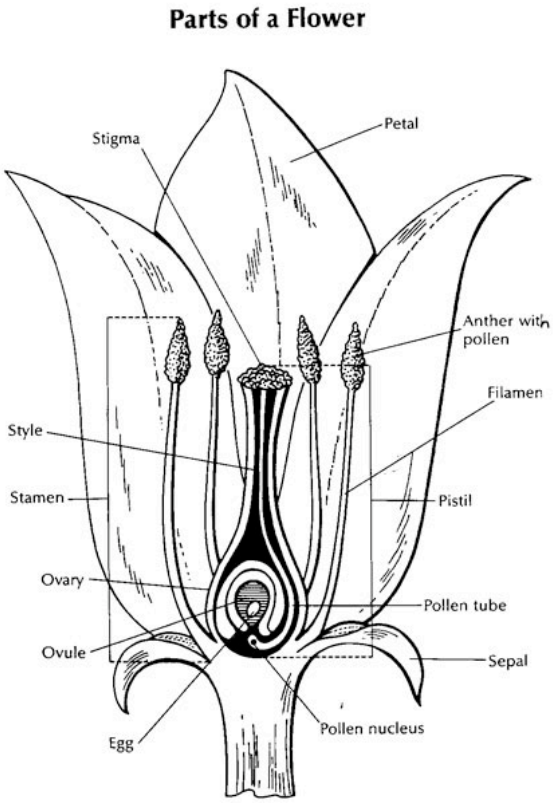


FIGURE 3 – Description d'une fleur

peut-être pas tous les détails mais la meilleure chose à faire est de taper le code et de voir le résultat produit. **Soyez curieux**.

Lorsque vous travaillez sous R, il peut être intéressant de conserver les résultats et les graphiques de vos analyses. Le plus simple, dans un premier temps, est de les enregistrer dans un document de type traitement de texte à l'aide du copier/coller.

1.2 Fichier de données : `iris`

R est un ensemble de bibliothèques de fonctions appelées « packages ». Chaque bibliothèque contient des jeux de données. Pour connaître par exemple les jeux de données de la distribution de base, entrez l'instruction suivante :

```
data()
```

En voici un extrait :

```
airmiles                Passenger Miles on Commercial US Airlines (1937-1960)
airquality              New York Air Quality Measurements
anscombe                Anscombe's Quartet of "Identical" Simple Linear Regressions
attenu                  The Joyner-Boore Attenuation Data
...                     ...
iris                    Edgar Anderson's Iris Data
...                     ...
USArrests               Violent Crime Rates by US State
USJudgeRatings          Lawyers' Ratings of State Judges in the US Superior Court
USPersonalExpenditure   Personal Expenditure Data
uspop                   Populations Recorded by the US Census
VADeaths                Death Rates in Virginia (1940)
volcano                 Topographic Information on Auckland's Maunga Whau Volcano
warpbreaks              The Number of Breaks in Yarn during Weaving
women                   Average Heights and Weights for American Women
```

Notez la présence de `iris`. Pour analyser ces données, il faut les charger en mémoire à l'aide de l'instruction :

```
data(iris)
```

Il existe d'autres procédés pour charger un jeu de données dans le logiciel R mais ce n'est pas l'objet ici.

Exercice. Tapez une à une chacune des instructions ci-dessous et notez le résultat obtenu. Attention, le logiciel R n'est pas indifférent aux majuscules et aux minuscules.

```
iris
dim(iris)
names(iris)
iris$Species
iris$Petal.Length
```

1.3 Représentation des espèces d'iris

La dernière colonne des données `iris` contient le nom des espèces réparties en trois catégories : setosa, versicolor et virginica. Pour accéder à celle-ci, il faut utiliser l'instruction `iris$Species`. On dit que la dernière colonne contient une variable *qualitative* à trois *modalités* appelées `levels` dans R. La fonction `levels()` appliquée à la colonne `iris$Species` donne les modalités de la variable :

```
levels(iris$Species)
```
```
[1] "setosa"     "versicolor" "virginica"
```

Pour résumer l'information contenue dans cette variable, on utilise l'instruction `summary()` :

```
summary(iris$Species)
```
```
    setosa versicolor  virginica
        50         50         50
```

Cette information peut être obtenue en construisant un tableau (`table()`) comptabilisant le nombre d'individus par modalité. Pour ce faire, tapez :

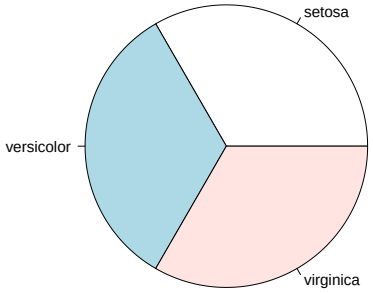
```
table(iris$Species)
```
```
    setosa versicolor  virginica
        50         50         50
```

et comparez avec le résultat précédent.

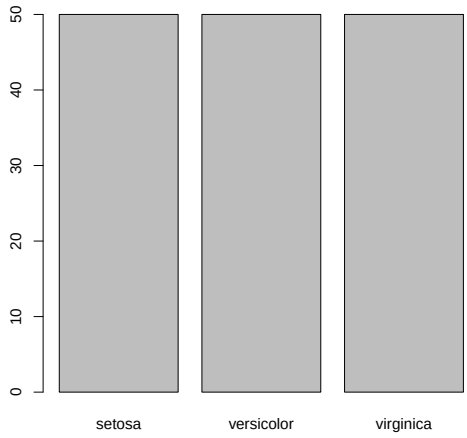
Le logiciel R permet de réaliser d'excellents graphiques. Lorsqu'une instruction graphique est lancée, une nouvelle fenêtre « device » est ouverte. Les

représentations graphiques classique liées aux variables qualitatives sont la représentation en secteurs ou camembert (`pie()`), la représentation en bâtons (`barplot()`), et la représentation de Cleveland (`dotchart()`). Entrez les instructions suivantes :

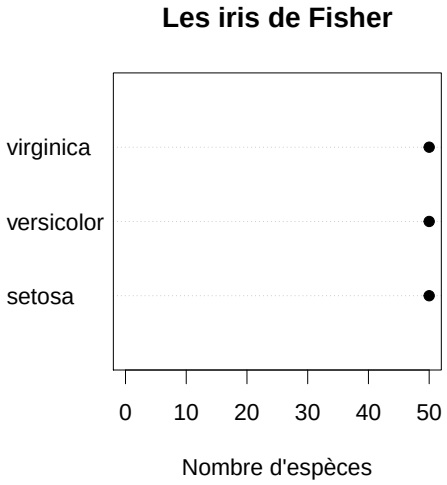
```
pie(table(iris$Species))
```




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




```
x <- unclass(table(iris$Species))
dotchart(x, pch = 19, cex = 1.5, xlim = c(0, max(x)),
  main = "Les iris de Fisher", xlab = "Nombre d'espèces")
```



Il existe un paramètre permettant de découper la fenêtre graphique : `par(mfrow = c(nl, nc))` ou `par(mfcol = c(nl, nc))`. `nl` définit le nombre de graphiques en lignes et `nc` définit le nombre de graphiques en colonnes. `mfrow` signifie que l'ordre d'entrée des graphiques s'effectue selon les lignes et `mfcol` signifie que l'ordre d'entrée des graphiques s'effectue selon les colonnes. Supposons que nous voulions représenter six graphiques dans une fenêtre en deux lignes et trois colonnes.
La première instruction conduit à entrer les graphiques selon l'ordre :

1	2	3
4	5	6

La seconde instruction conduit à entrer les graphiques selon l'ordre :

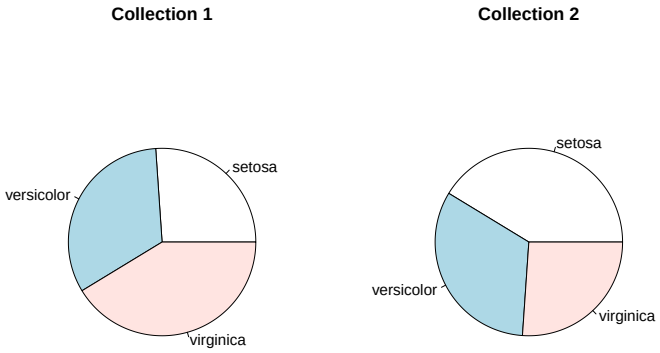
1	3	5
2	4	6

Exercice. Deux botanistes se sont également intéressés aux iris et ont collecté les espèces suivantes.

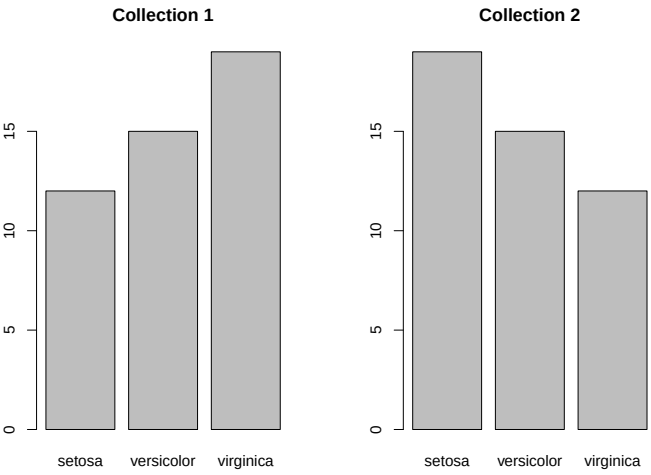
```
collection1 <- rep(c("setosa", "versicolor", "virginica"), c(12, 15, 19))
collection2 <- rep(c("setosa", "versicolor", "virginica"), c(19, 15, 12))
```

En utilisant la commande `par(mfrow = c(1, 2))`,

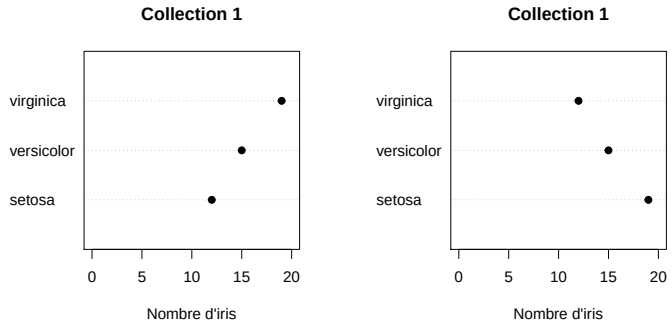
1. construire les camemberts liés à ces deux nouvelles distributions et commenter ;



2. construire les représentations en bâtons de ces deux nouvelles distributions et commenter ;



3. construire les représentations de Cleveland de ces deux nouvelles distributions et commenter ;



4. discuter des avantages et des inconvénients de ces trois types de représentations.

1.4 Représentation de la longueur du pétale

La troisième colonne (`Petal.Length`) du jeu de données `iris` contient la longueur du pétale. Il s'agit d'une variable mesurée qualifiée alors de variable *quantitative*. Pour résumer l'information contenue dans cette variable, on utilise la fonction `summary()` et on obtient le résultat :

```
summary(iris$Petal.Length)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  1.000   1.600   4.350   3.758   5.100   6.900
```

La plus petite (`Min`) longueur de pétale est 1 cm tandis que la plus grande (`Max`) est 6.9 cm. La moyenne (`Mean`) représente la somme des valeurs de la *distribution* divisée par le nombre total d'iris. Elle vaut 3.758 cm.
Si l'ensemble des 150 longueurs de pétale sont classées par ordre croissant, `1st Qu.`, `Median` et `3rd Qu.` sont les trois valeurs qui permettent de couper la distribution en quatre parties égales. On les appelle respectivement premier *quartile*, médiane (ou deuxième quartile) et troisième quartile.

Exercice. Essayons de retrouver ces six valeurs de *paramètres*.

```
min(iris$Petal.Length)
```

```
[1] 1
```

```
max(iris$Petal.Length)
```

```
[1] 6.9
```

```
sum(iris$Petal.Length)
```

```
[1] 563.7
```

```
length(iris$Petal.Length)
```

```
[1] 150
```

```
sum(iris$Petal.Length)/length(iris$Petal.Length)
```

```
[1] 3.758
```

```
sort(iris$Petal.Length)
```

```
  [1] 1.0 1.1 1.2 1.2 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4
 [21] 1.4 1.4 1.4 1.4 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.6 1.6 1.6
 [41] 1.6 1.6 1.6 1.6 1.7 1.7 1.7 1.7 1.9 1.9 3.0 3.3 3.3 3.5 3.5 3.6 3.7 3.8 3.9 3.9
 [61] 3.9 4.0 4.0 4.0 4.0 4.0 4.1 4.1 4.1 4.2 4.2 4.2 4.2 4.3 4.3 4.4 4.4 4.4 4.4 4.5
 [81] 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.6 4.6 4.6 4.7 4.7 4.7 4.7 4.7 4.8 4.8 4.8 4.8 4.9
[101] 4.9 4.9 4.9 4.9 5.0 5.0 5.0 5.0 5.1 5.1 5.1 5.1 5.1 5.1 5.1 5.1 5.2 5.2 5.3 5.3
[121] 5.4 5.4 5.5 5.5 5.5 5.6 5.6 5.6 5.6 5.6 5.6 5.7 5.7 5.7 5.8 5.8 5.8 5.9 5.9 6.0
[141] 6.0 6.1 6.1 6.1 6.3 6.4 6.6 6.7 6.7 6.9
```

```
ordLpetal <- sort(iris$Petal.Length)
ordLpetal
```

```
  [1] 1.0 1.1 1.2 1.2 1.3 1.3 1.3 1.3 1.3 1.3 1.3 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4 1.4
 [21] 1.4 1.4 1.4 1.4 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.6 1.6 1.6
 [41] 1.6 1.6 1.6 1.6 1.7 1.7 1.7 1.7 1.9 1.9 3.0 3.3 3.3 3.5 3.5 3.6 3.7 3.8 3.9 3.9
 [61] 3.9 4.0 4.0 4.0 4.0 4.0 4.1 4.1 4.1 4.2 4.2 4.2 4.2 4.3 4.3 4.4 4.4 4.4 4.4 4.5
 [81] 4.5 4.5 4.5 4.5 4.5 4.5 4.5 4.6 4.6 4.6 4.7 4.7 4.7 4.7 4.7 4.8 4.8 4.8 4.8 4.9
[101] 4.9 4.9 4.9 4.9 5.0 5.0 5.0 5.0 5.1 5.1 5.1 5.1 5.1 5.1 5.1 5.1 5.2 5.2 5.3 5.3
[121] 5.4 5.4 5.5 5.5 5.5 5.6 5.6 5.6 5.6 5.6 5.6 5.7 5.7 5.7 5.8 5.8 5.8 5.9 5.9 6.0
[141] 6.0 6.1 6.1 6.1 6.3 6.4 6.6 6.7 6.7 6.9
```

```
sum(ordLpetal)/length(ordLpetal)
```

```
[1] 3.758
```

```
ordLpetal[38]
```

```
[1] 1.6
```

```
(ordLpetal[75]+ordLpetal[76])/2
```

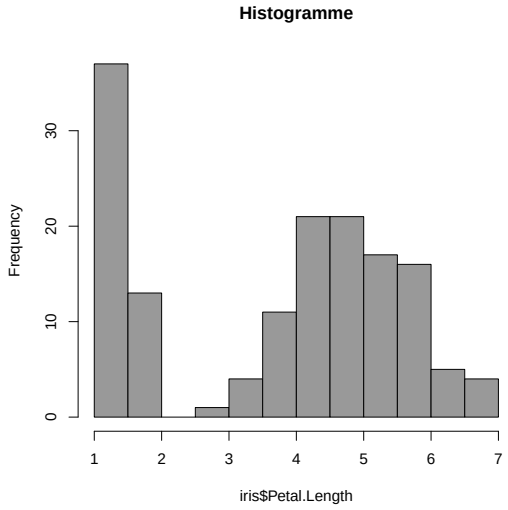
```
[1] 4.35
```

```
ordLpetal[113]
```

```
[1] 5.1
```

Une des représentations adéquates est l'histogramme (`hist()`) :

```
hist(iris$Petal.Length, col = grey(0.6), main="Histogramme")
```



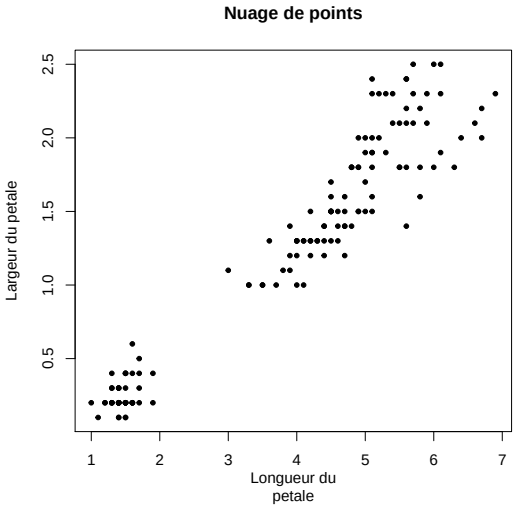
Exercice. Réaliser le même type d'analyse sur chacune des autres variables quantitatives : largeur du pétale, longueur du sépale et largeur du sépale.
Notez que vous n'avez pas toutes les instructions à réécrire en utilisant le système de flèches du clavier. ↑ et ↓ vous permettent de retrouver les fonctions que vous avez utilisées. ← et → vous permettent de vous déplacer dans la fonction et donc, d'en changer certains paramètres.

1.5 Représentation de la longueur et de la largeur du pétale

Une fois réalisés les graphiques pour chaque variable prise séparément, l'étude peut porter sur la relation entre deux variables. On parle de croisement de deux variables ou d'étude *bivariée*. La représentation graphique liant deux variables quantitatives est le nuage de points. Représentons par exemple la longueur et

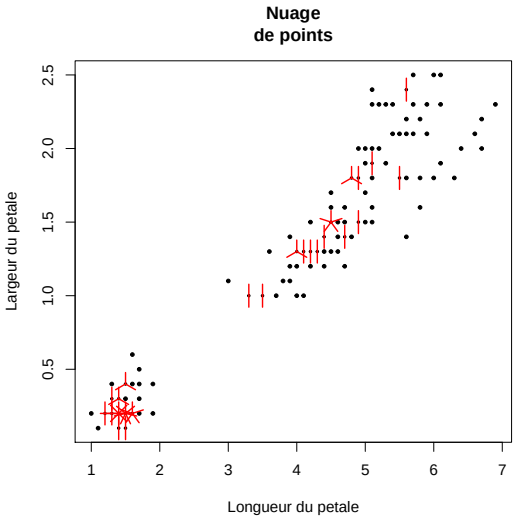
la largeur du pétale pour les 150 iris contenus dans le fichier de données. Commentaire.

```
plot(iris$Petal.Length, iris$Petal.Width, xlab="Longueur du
petale", ylab="Largeur du petale", main="Nuage de points",
pch=20)
```



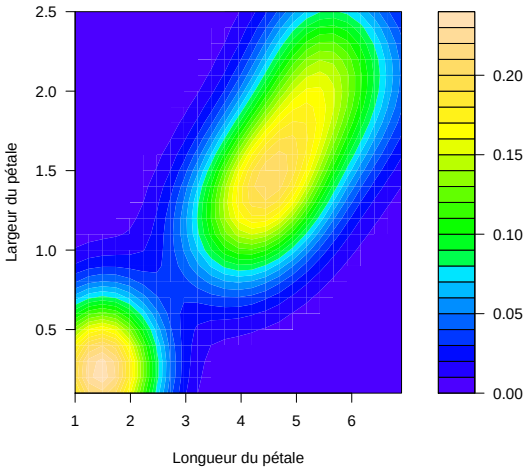
Dans cette représentation graphique, plusieurs individus peuvent être situés sur un même point. La fonction `sunflowerplot()` permet de visualiser ces superpositions.

```
sunflowerplot(iris$Petal.Length, iris$Petal.Width,
xlab="Longueur du petale", ylab="Largeur du petale", main="Nuage
de points",pch=20)
```



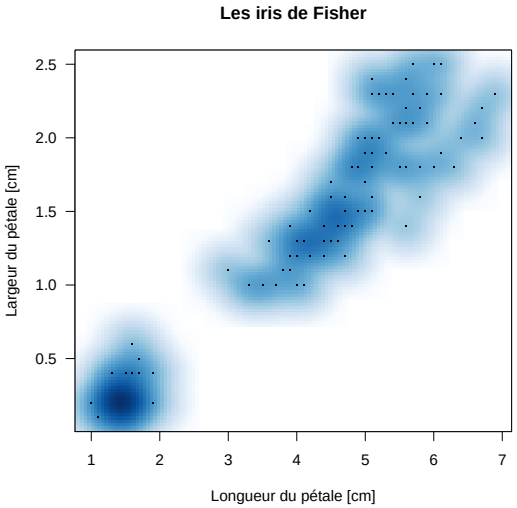
Quand le nombre de points devient trop important, on peut alors représenter la densité des points au lieu des points eux-même, par exemple :

```
library(MASS)
densite <- kde2d(iris$Petal.Length, iris$Petal.Width)
filled.contour(densite, color = topo.colors, xlab = "Longueur du pétale",
ylab = "Largeur du pétale")
```



Ou encore en utilisant la fonction standard `smoothScatter()` de R :

```
smoothScatter(iris$Petal.Length, iris$Petal.Width, las = 1,
  main = "Les iris de Fisher", xlab = "Longueur du pétale [cm]",
  ylab = "Largeur du pétale [cm]")
```

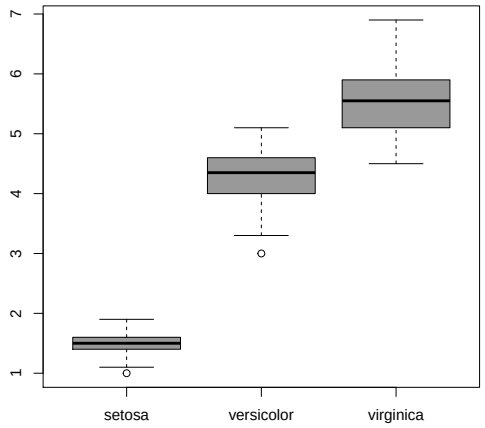


Exercice. Réaliser l'étude du croisement de deux variables quantitatives de votre choix. Il est clair que le sens biologique de l'étude ne doit pas être négligé.

1.6 Représentation de la longueur du pétale selon les différentes espèces

La représentation graphique permettant de lier une variable qualitative et une variable quantitative est la boîte à moustaches (`boxplot()`). Représentons par exemple la longueur des pétales en fonction de l'espèce. Commentaire.

```
boxplot(iris$Petal.Length~iris$Species, col = grey(0.6))
```



Exercice. Choisissez une autre variable quantitative, croisez-la avec la variable espèce d'iris et commentez.

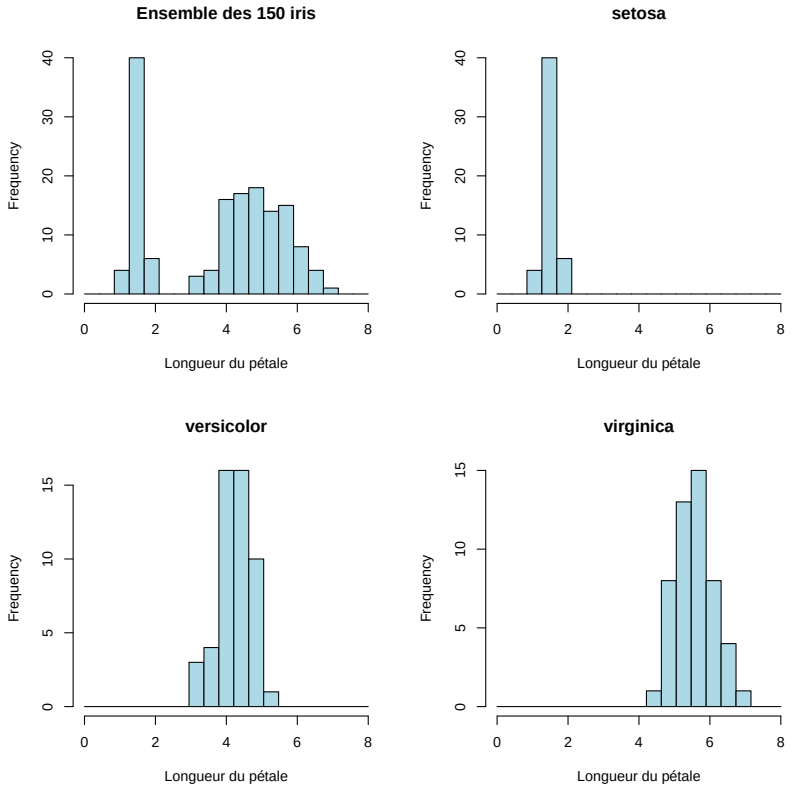
1.7 Pour aller plus loin ...

Le nuage de points comme les boîtes à moustaches montrent que les données morphologiques des iris semblent liées à l'espèce. Il pourrait donc être intéressant de réaliser des graphiques différents pour chacune des modalités *I. setosa*, *I. versicolor* et *I. virginica* ou de superposer l'information espèce dans le graphique des nuages de points. Nous vous proposons ici quelques développements. Libre à vous, de les refaire ou d'en trouver d'autres ...

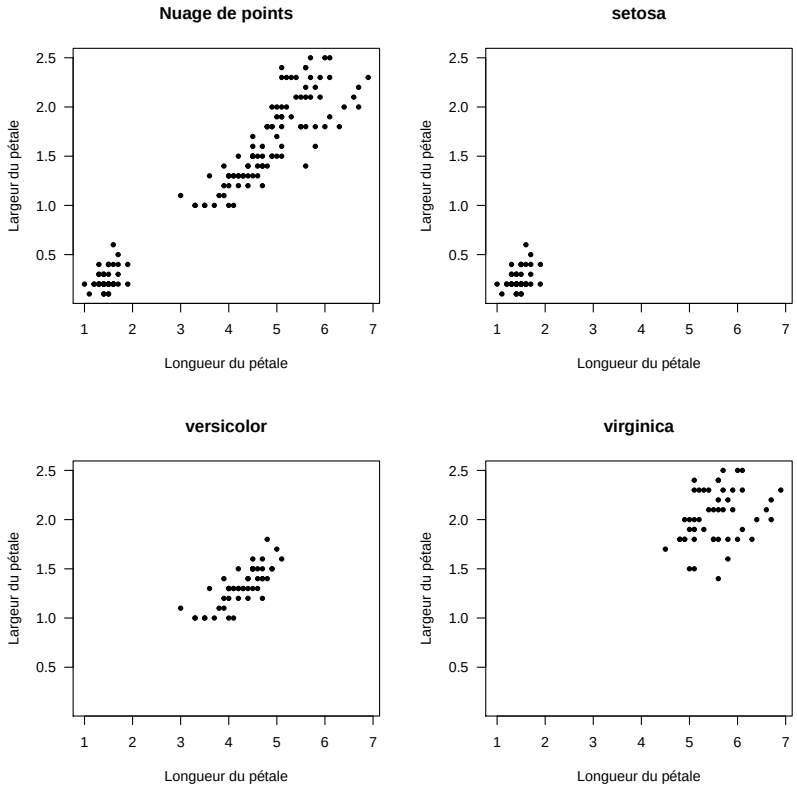
```
summary(iris)
```

```
 Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50
1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
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
```

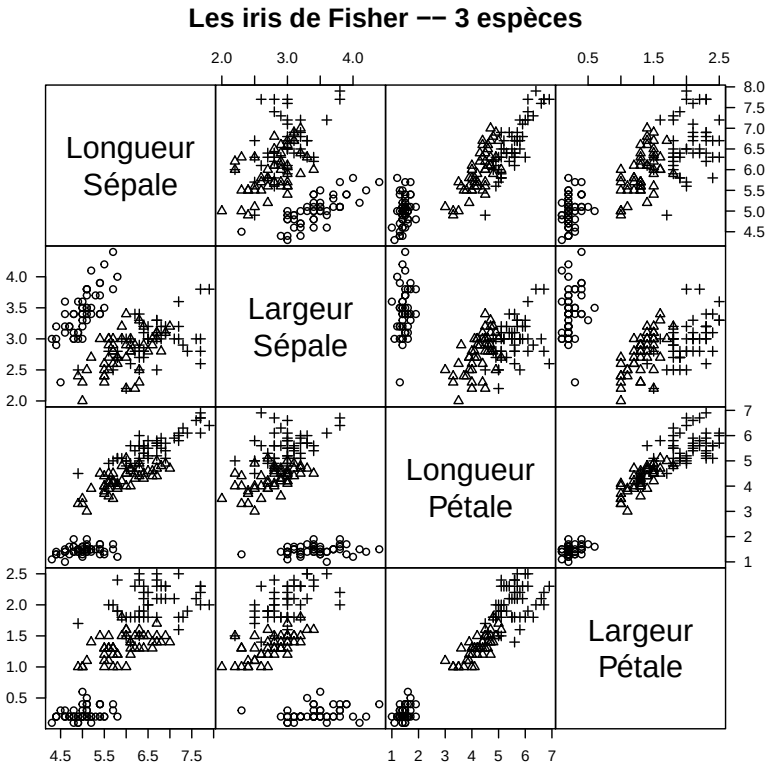
```
par(mfrow = c(2, 2))
brk = seq(from = 0, to = 8, length = 20)
myhist <- function(x, ...) hist(x, xlab = "Longueur du pétale", breaks = brk, col = "lightblue", ...)
with(iris, {
  myhist(Petal.Length, main = "Ensemble des 150 iris")
  for(sp in levels(Species)) myhist(Petal.Length[Species == sp], main = sp)
})
```



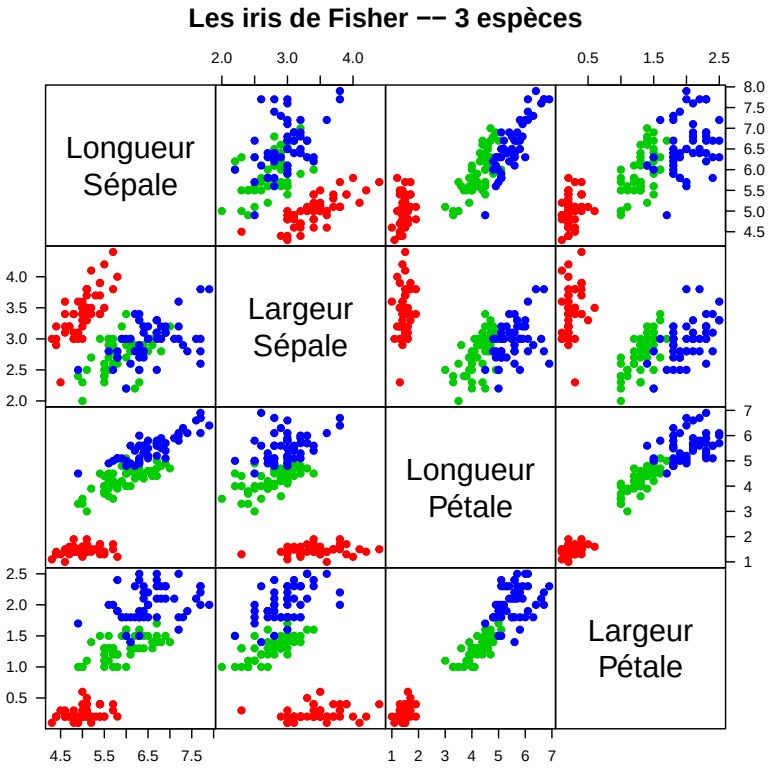

```
par(mfrow = c(2, 2))
with(iris, {
  myplot <- function(x, y, ...) plot(x, y, xlab = "Longueur du pétale",
    ylab = "Largeur du pétale", pch = 20, las = 1, xlim = range(Petal.Length),
    ylim = range(Petal.Width), ...)
  x <- Petal.Length ; y <- Petal.Width
  myplot(x, y, main = "Nuage de points")
  for(sp in levels(Species)) myplot(x[Species == sp], y[Species == sp], main = sp)
 })
```




```
typepoint <- (1:3)[as.integer(iris$Species)]
pairs(iris[1:4], main = "Les iris de Fisher -- 3 espèces",
  las = 1, gap = 0, pch = typepoint,
  labels = c("Longueur\nSépale", "Largeur\nSépale", "Longueur\nPétale",
  "Largeur\nPétale"))
```




```
couleurs <- c("red", "green3", "blue")[as.integer(iris$Species)]
pairs(iris[1:4], main = "Les iris de Fisher -- 3 espèces",  col = couleurs,
  las = 1, gap = 0, pch = 19, labels = c("Longueur\nSépale", "Largeur\nSépale",
  "Longueur\nPétale", "Largeur\nPétale"))
```



En guise de conclusion, soulignons le fait que les représentations graphiques sont une étape fondamentale dans la connaissance des données et que le logiciel R est un *excellent* outil. Les représentations graphiques sont là pour éclairer la nature des données et non pour souligner votre côté artistique. Chaque information ajoutée à un graphe, comme par exemple une couleur, doit contribuer à cet éclairage.

2 Des mannequins

MADRID (AFP) lundi 11 septembre 2006, 10h40 : « **Les mannequins trop maigres interdites au grand défilé de la mode madrilène.** Les mannequins trop maigres n'auront pas le droit de défiler au grand rendez-vous de la mode madrilène, la Pasarela Cibeles, du 18 au 22 septembre, en raison du mauvais exemple donné aux jeunes Espagnoles dont l'obsession des kilos incite à l'anorexie. Les médias espagnols ont rapporté que le gouvernement régional de Madrid, qui cofinance l'événement, avait exclu 30 à 40% des top-modèles, trop maigres, ayant participé à la dernière édition. En collaboration avec la Société espagnole d'endocrinologue et de nutrition (SEEN), il a imposé un critère strict : aucun mannequin présentant un indice de masse corporelle inférieur à 18 (56 kilos pour 1m75) ne pourra défiler. »

La masse corporelle d'un individu est mesurée à partir du poids et de la taille. L'indice le plus utilisé est celui proposé par Adolphe Quételet (1796-1874) [3], astronome et mathématicien belge (voir `http://statbel.fgov.be/`

info/quetelet_fr.asp).

2.1 Questions

1. Construire la variable « indice de Quételet » à partir des données de t3var (première séance de TP).

$$imc = \frac{poids}{taille^2}$$

   La taille est exprimée en mètre et le poids en kilogrammes.
   Cet indice s'appelle aujourd'hui « indice de masse corporelle » (imc) ou « body mass index » (bmi). Il permet de mesurer la corpulence de l'homme adulte. L'organisation mondiale de la santé (OMS) a défini les critères suivants : maigreur (inférieur à 18.5), normal (de 18.5 à 25), risque de surpoids (de 25 à 30), obésité (supérieur à 30). Mais l'indice de Quételet n'a qu'une valeur indicative. Pour déterminer l'existence d'une obésité réelle, il faut faire d'autres mesures destinées à établir exactement la proportion de masse grasse, car c'est l'excès de masse grasse qui représente un facteur de risque.
2. Calculer les paramètres statistiques élémentaires (moyenne, médiane, variance, minimum, maximum, quartiles etc.) de cette nouvelle variable sur l'ensemble des individus et en fonction de chaque sexe.
3. Construire l'histogramme de cette nouvelle variable sur l'ensemble des individus.
4. Comparer graphiquement les indices de masse corporelle chez les hommes et chez les femmes à l'aide d'histogrammes (representer les deux histogrammes dans la même fenêtre).
5. En utilisant la fonction boxplot, représenter l'indice de Quételet en fonction du sexe.
6. Construire les nuages de points de l'indice de Quételet en fonction de la taille, pour chaque sexe. Construire ensuite les nuages de points de l'indice de Quételet en fonction du poids, pour chaque sexe. Commenter.
7. Représenter, sur un seul graphique, le nuage de points de l'indice de Quételet en spécifiant hommes et femmes. Donner une couleur différente pour chaque sexe. Placer les bornes des critères de l'OMS. Commenter.

2.2 Réponses

1. Les données

```
options(digits=4)
t3var <- read.table("http://pbil.univ-lyon1.fr/R/donnees/t3var.txt",h=T)
poi <- t3var$poi
tai <- t3var$tai*0.01
sexe <- factor(t3var$sexe)
imc <- poi/(tai^2)
imc
```

```
 [1] 20.76 19.96 17.24 18.17 17.72 19.29 18.29 20.16 20.31 20.90 19.47 26.77 19.92
[14] 23.04 21.46 20.43 19.71 24.66 22.34 20.48 19.63 18.82 25.01 23.37 18.34 23.04
[27] 21.60 20.16 23.66 20.48 19.25 22.09 23.99 18.07 17.99 22.27 19.95 19.49 22.31
[40] 21.22 23.15 19.05 16.92 19.26 18.38 21.74 22.40 21.56 21.67 20.37 22.41 24.93
[53] 21.97 18.13 24.96 22.22 19.82 21.88 25.66 22.28 23.20 25.96 23.80 20.52 20.89
[66] 21.39
```

2. Paramètres statistiques élémentaires

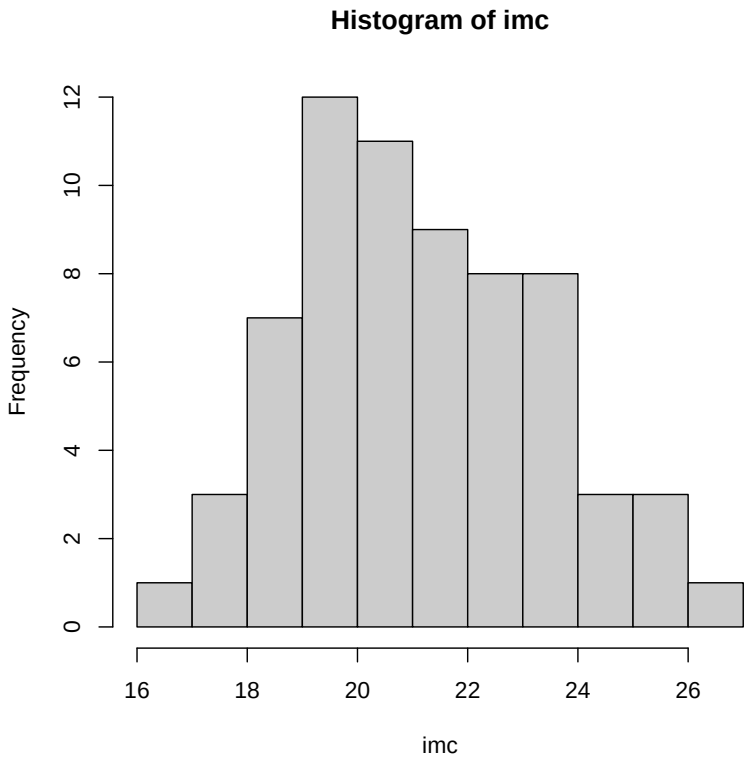
```
summary(imc)
```
```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  16.9    19.5    20.9    21.2    22.4    26.8
```
```
mean(imc)
```
```
[1] 21.16
```
```
var(imc)
```
```
[1] 5.137
```
```
sd(imc)
```
```
[1] 2.267
```
```
summary(imc[sexe=="f"])
```
```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  16.9    18.2    19.5    19.5    20.3    23.7
```
```
summary(imc[sexe=="h"])
```
```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  18.4    20.8    22.0    22.2    23.2    26.8
```
```
mean(imc[sexe=="f"])
```
```
[1] 19.46
```
```
mean(imc[sexe=="h"])
```
```
[1] 22.19
```
```
var(imc[sexe=="f"])
```
```
[1] 2.673
```
```
var(imc[sexe=="h"])
```
```
[1] 3.858
```
```
sd(imc[sexe=="f"])
```
```
[1] 1.635
```
```
sd(imc[sexe=="h"])
```
```
[1] 1.964
```

Une autre manière de calculer les statistiques élémentaires en fonction du sexe :

```
imcs <- split(imc,sexe)
imcs
lapply(imcs,mean)
lapply(imcs,var)
lapply(imcs,sd)
lapply(imcs,summary)
```

3. Histogramme sur l'ensemble des individus

```
hist(imc, col = grey(0.8))
```



4. Représentations graphiques par sexe

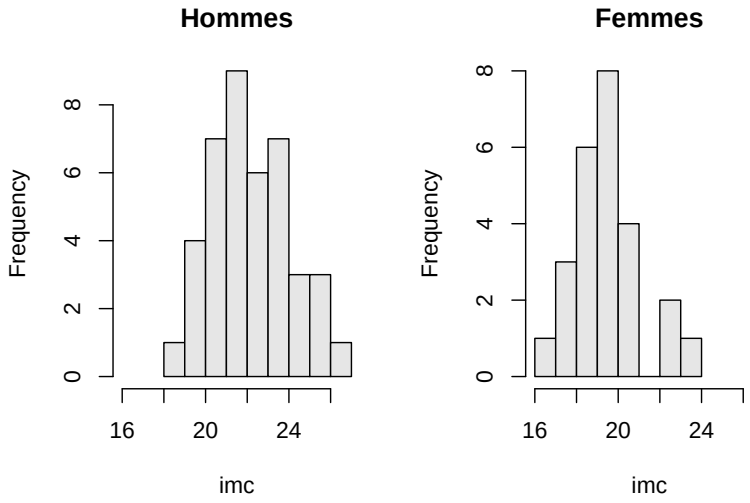
```
range(imc[sexe=="f"])
```

```
[1] 16.92 23.66
```

```
range(imc[sexe=="h"])
```

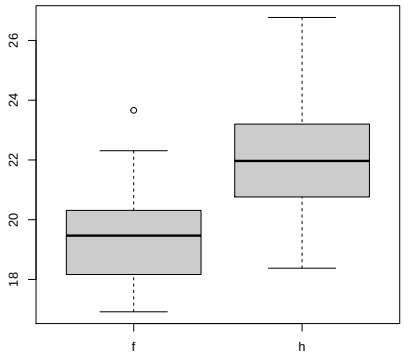
```
[1] 18.38 26.77
```

```
par(mfcol=c(1,2))
par(mar=c(5,4,3,2))
hist(imc[sexe == "h"], main = "Hommes", xlab = "imc", xlim = c(16,27), col = grey(0.9))
hist(imc[sexe == "f"], main = "Femmes", xlab = "imc", xlim = c(16,27), col = grey(0.9))
```



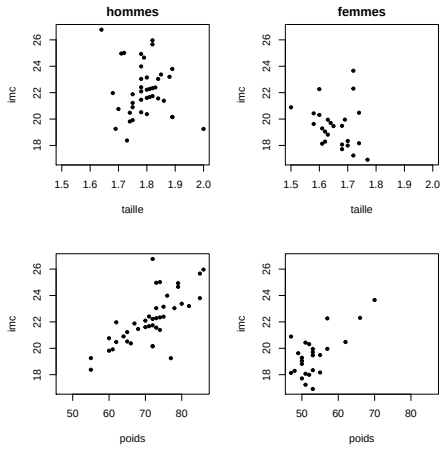
5. Boîte à moustaches

```
boxplot(imc~sexe,col=grey(0.8))
```



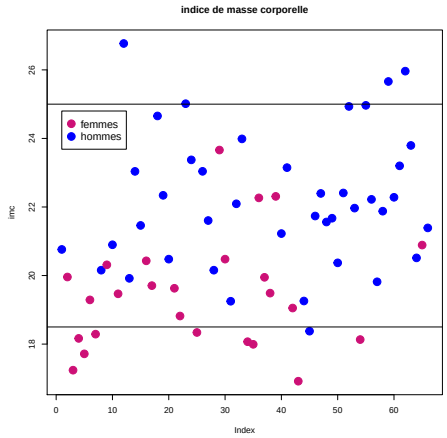
6. Relation entre l'imc, la taille et le poids

```
par(mfcol = c(2,2))
par(mar=c(5,4,2,2))
plot( tai[sexe == "h"], imc[sexe == "h"],ylim=c(min(imc),max(imc)), xlim=c(min(tai),max(tai)),
    pch = 20, ylab = "imc", xlab = "taille", main = "hommes" )
plot(poi[sexe == "h"],imc[sexe == "h"], ylim=c(min(imc),max(imc)), xlim=c(min(poi),max(poi)),
     pch = 20, ylab = "imc", xlab = "poids")
plot( tai[sexe == "f"],imc[sexe == "f"], ylim=c(min(imc),max(imc)), xlim=c(min(tai),max(tai)),
     pch = 20, ylab = "imc", xlab = "taille", main = "femmes" )
plot( poi[sexe == "f"], imc[sexe == "f"],ylim=c(min(imc),max(imc)), xlim=c(min(poi),max(poi)),
     pch = 20, ylab = "imc", xlab = "poids")
```



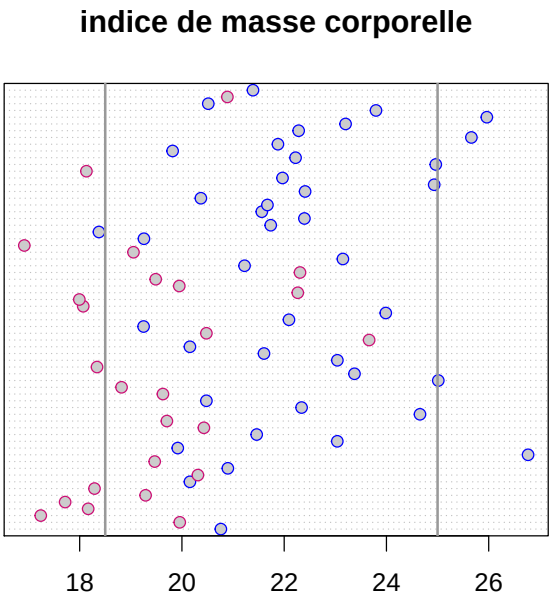
7. Un seul graphique
   Version de base

```
couleur <- ifelse(sexe == "f", "deeppink3", "blue")
plot(imc, pch = 19, cex = 2, col=couleur)
abline(h = 18.5)
abline(h = 25)
title("indice de masse corporelle")
malegende=c("femmes","hommes")
legend(1,24.8,malegende, pch = 21, pt.cex = 2, cex = 1.25,
    col = c("deeppink3","blue"), pt.bg = c("deeppink3","blue"))
```



   Version Graphe de Cleveland

```
dotchart(imc, main = "indice de masse corporelle", pch = 21,
  bg = grey(0.8), cex = 1.25,col = couleur)
abline(v = 18.5, lwd = 2, col = grey(0.6))
abline(v = 25, lwd = 2, col = grey(0.6))
```



Références

[1] Anderson E. The irises of the Gaspe Peninsula. *Bulletin of the American Iris Society*, 59 :2–5, 1935.

[2] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2) :179–188, 1936.

[3] A. Quételet. *Anthropométrie ou mesures des différentes facultés de l'homme*. Editions de Bruxelles, Bruxelles, editions de bruxelles edition, 1870.