

# Formation Logiciel R



Institut Pasteur  
de Nouvelle-Calédonie

1

4<sup>°</sup> séance

**PRÉSENTER DES DONNÉES  
AVEC LE LOGICIEL R**

2

# Présentation des données

**Tableaux**  
**Graphiques**  
**Diagrammes**  
**Cartes**

3

## Tableaux

- Tableaux brut de données
  - Matrice de données
- Tableaux de fréquence
  - Pour les données regroupées en classe: variables qualitatives et variables quantitatives discrètes

4

# Tableau brut de données

VALEURS TEXTE	1	2	3	4	5	6	7	8	9	10
	ANNEE	AGE	SEXE	PLVT	COM_NOSO	PENG	PENG2	OXA	OXA2	VANCO
061013-0249	2006				NOSO	I	1	I	1	S
020415-112	2002	32	F	pus pro	COMM	R	1	R	1	S
010115-030	2001	40	M	pus pro	COMM	R	1	S	0	S
050428-0113	2005	14	M	pus pro	COMM	R	1	S	0	S
041230-066	2004	32	F	pus pro	COMM	R	1	S	0	S
030107-237	2003	63	M	pus pro	COMM	I	1	S	0	S
030109-180	2003	25	M	pus pro	COMM	I	1	S	0	S
021209-124	2002	58	M	pus pro	COMM	I	1	S	0	S
010206-246	2001	69	M	pus pro	COMM	R	1	S	0	S
010305-152	2001	39	F	pus pro	COMM	R	1	S	0	S
020527-139	2002	44	F	pus pro	COMM	R	1	S	0	S
031107- 285	2003	1	F	pus pro	COMM	R	1	S	0	S
020816-206	2002	28	F	pus pro	COMM	I	1	S	0	S
020508-216	2002	3	F	pus pro	COMM	R	1	S	0	S
050426-0247	2005	30	F	pus pro	COMM	R	1	S	0	S
010105-017	2001	9	F	pus pro	COMM	R	1	S	0	S
010921-177	2001	52	M	pus pro	COMM	R	1	R	1	S
031023- 231	2003	50	F	pus pro	COMM	R	1	S	0	S
030805- 218	2003	39	F	va	COMM	R	1	S	0	S
050408-0131	2005	38	F	va	COMM	R	1	S	0	S
040217-116	2004	39	F	va	COMM	S	0	S	0	S
050801-120	2005	38	F	va	COMM	R	1	S	0	S
021003-132	2002	38	F	va	COMM	I	1	S	0	S
011220-087	2001	37	F	va	COMM	R	1	R	1	S
040920-005	2004	40	F	va	COMM	R	1	S	0	S
050921-0032	2005	39	F	va	COMM	R	1	S	0	S
020118-086	2002	37	F	va	COMM	R	1	S	0	S
011213-120	2001	36	F	va	COMM	R	1	S	0	S

5

# Tableau de fréquence

Tableau 1 : Répartition d'une série de 30 sujets en fonction de leur niveau d'étude

Niveau scolaire	Effectifs	Fréquence (%)	Effectif cumulé	Fréquence cumulée (%)
Primaire	11	36,7	30	100,0
Secondaire	6	20,0	19	56,7
Supérieur	13	43,3	13	43,3
Total	30	100,0	-	-

6

# Les graphiques

7

## Diagrammes

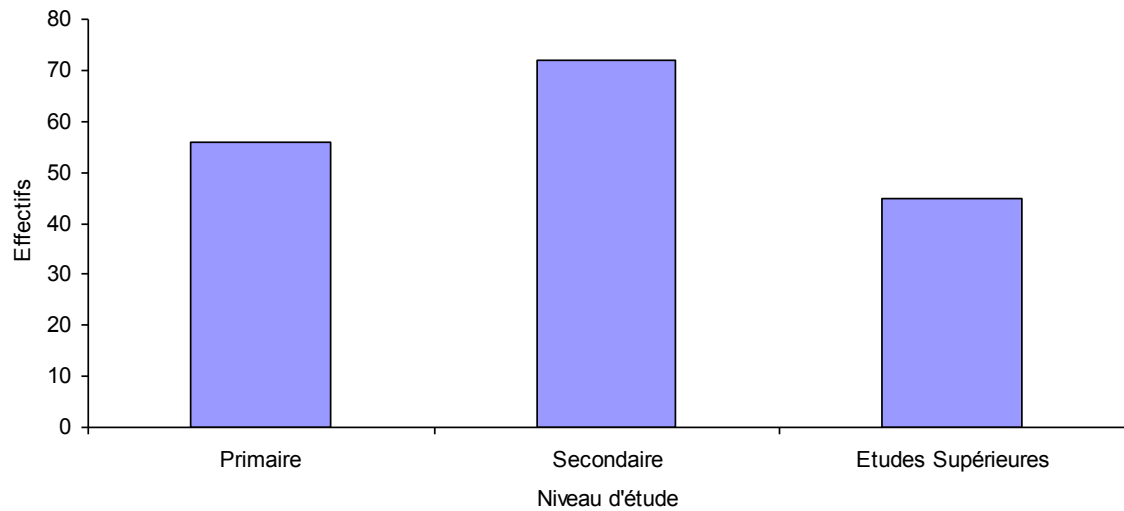
**Représentation des variables exprimées  
en classe distribuées sur une seule  
coordonnée**

- Diagramme en barres
- Diagramme en barres horizontales
- Camembert

8

# Diagramme en barre

Distribution des personnels de l'IP de Vendrepi en fonction du niveau d'étude en 2006

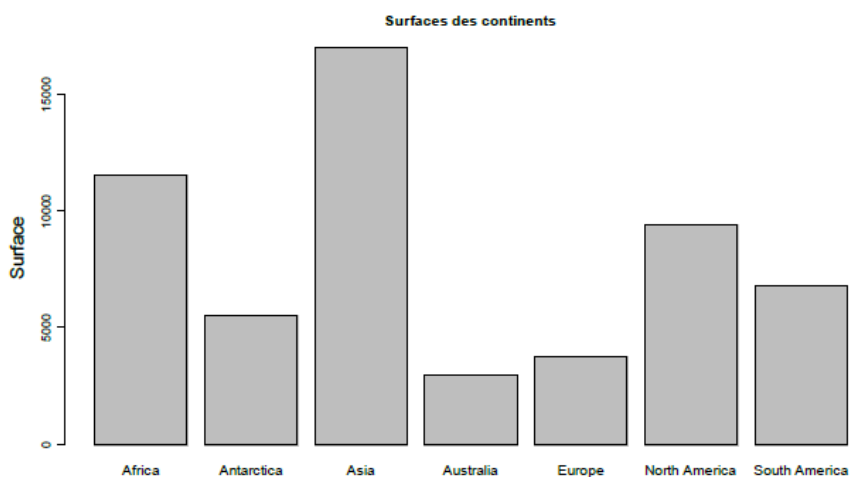


**Pour les variables qualitatives ordinales ou nominales**

9

## Les diagrammes en bâton (barplot) 1

◆ On représente une valeur (effectif, moyenne,...) d'individus pour chaque niveau de la variable qualitative

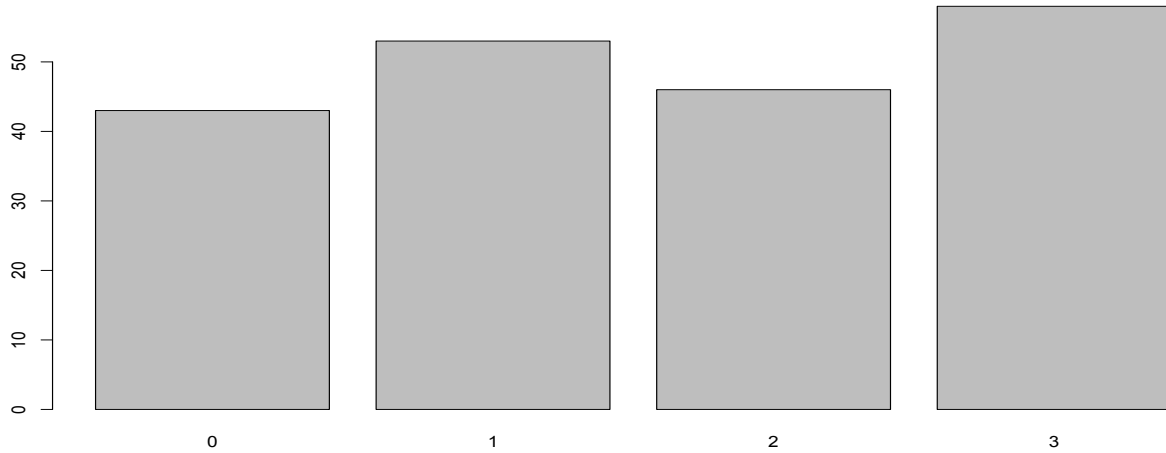


10

# Fonction plot()

- Pour des données catégorielles → Diagramme en Barre

```
>plot(as.factor(Mydata$NIVSCOL))
```

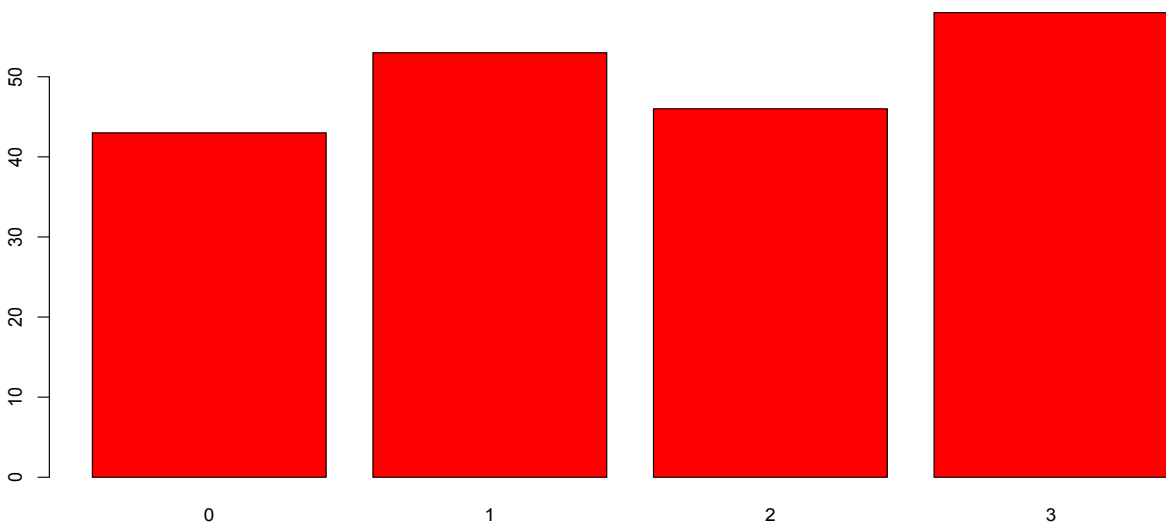


11

# Fonction plot()

- Pour des données catégorielles → Diagramme en Barre

```
>plot(as.factor(Mydata$NIVSCOL),col="red ")
```

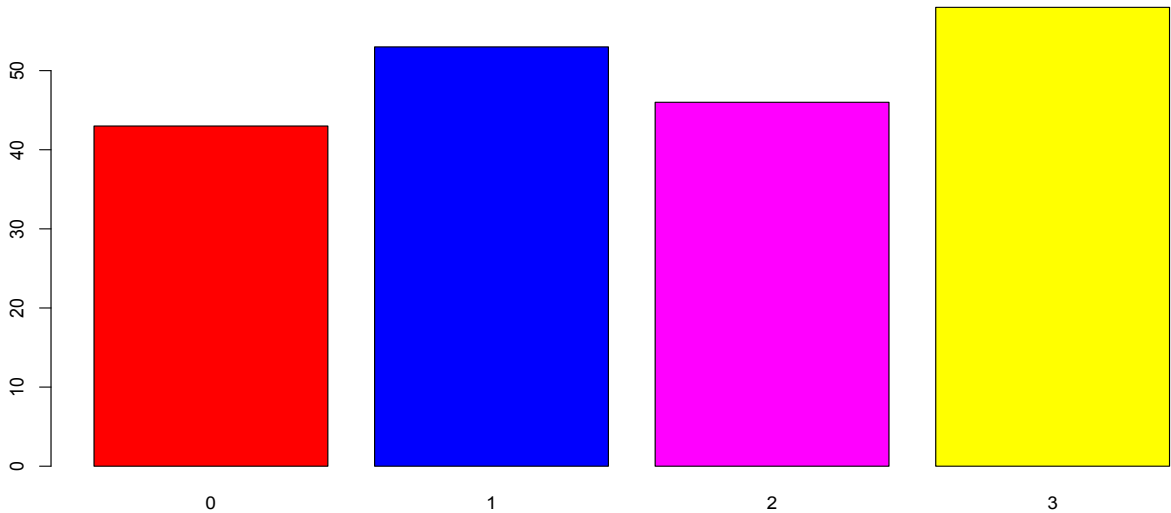


12

# Fonction plot()

- Pour des données catégorielles → Diagramme en Barre

```
>plot(as.factor(Mydata$NIVSCOL),col=c("red ","blue", "magenta","yellow"))
```

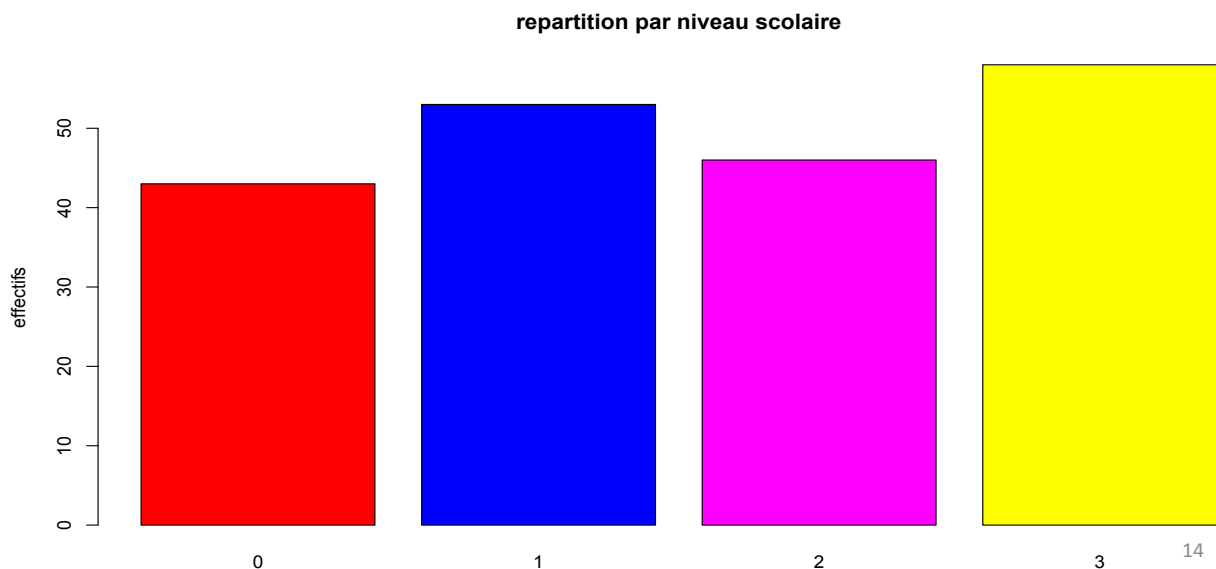


13

# Fonction plot()

- Pour des données catégorielles → Diagramme en Barre

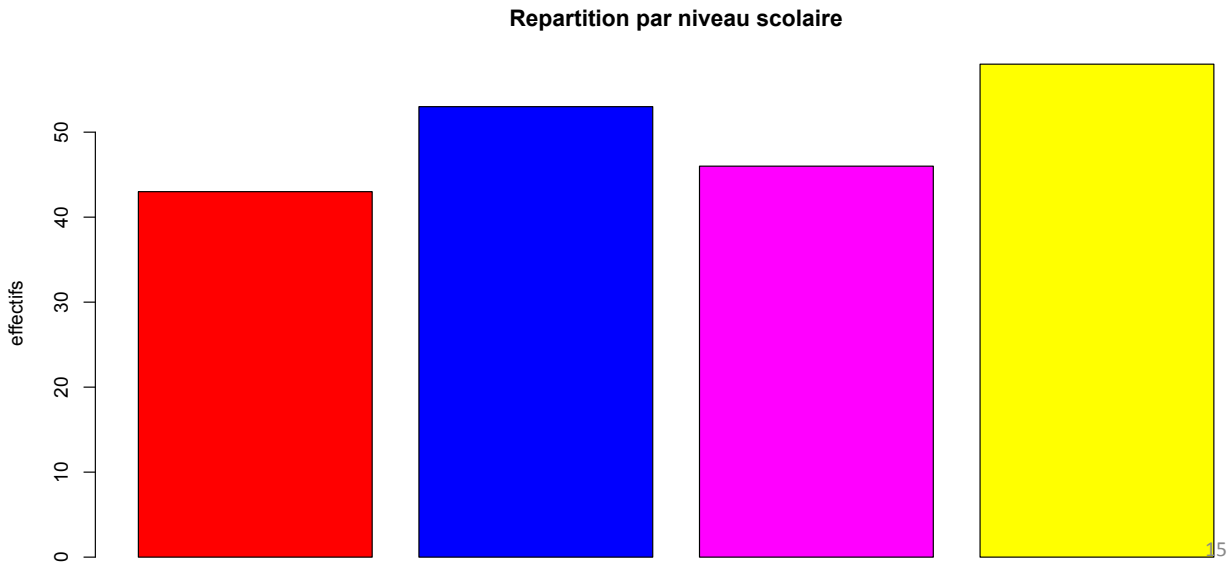
```
>plot(as.factor(Mydata$NIVSCOL),  
col=c("red ","blue", "magenta","yellow"), main="repartition par niveau  
scolaire",ylab="effectifs")
```



14

# Fonction plot()

```
plot(as.factor(Mydata$NIVSCOL), col=c("red","blue","magenta",  
"yellow"), main="Repartition par niveau scolaire",  
ylab="effectifs", xaxt="n")
```



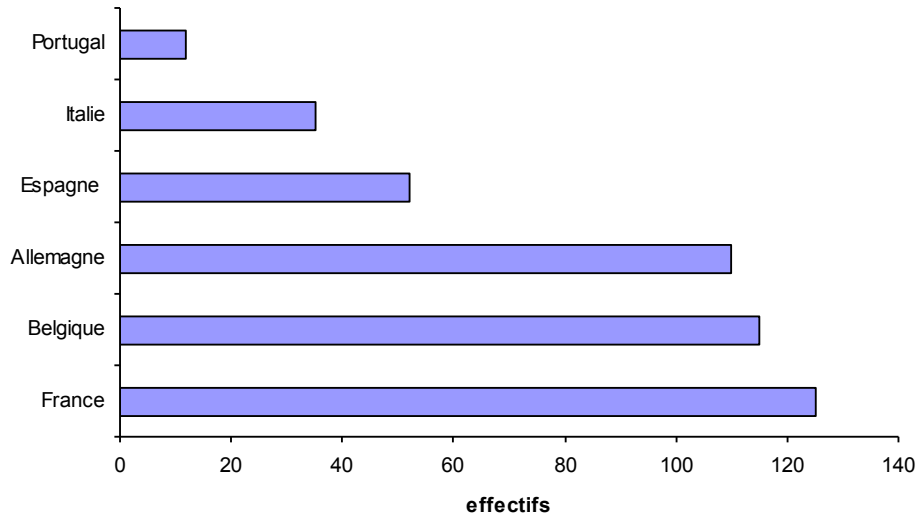
# Fonction plot()

- `plot(as.factor(Mydata$NIVSCOL),col=c('red','blue','magenta','yellow'),xaxt='n')`
- `axis(1,1:4,c('Non scolaire','Primaire','Secondaire','Universitaire'))`
- `plot(as.factor(Mydata$NIVSCOL),col=c('red','blue','magenta','yellow'),xaxt='n')`
- `axis(1,c(0.7,1.9,3.1,4.3),c('Non scolaire','Primaire','Secondaire','Universitaire'))`



# Diagramme en barre horizontale

Distribution des cas de grippe diagnostiqués en 2010 par pays



Pour les variables qualitatives nominales

17

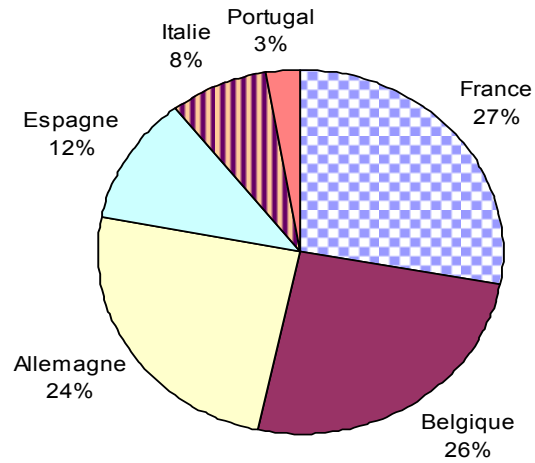
## Fonction plot()

- `plot(as.factor(Mydata$NIVSCOL),col=c('red','blue','magenta','yellow'),horiz=T,yaxt='n' ),xlim=c(0,60))`
- `axis(2,at=c(0.6,1.9,3.0,4.2),c('Non  
scolarise','Primaire','Secondaire','Universitaire'),cex.axis=0.5)`

18

# Camembert

Distribution des diagnostics de Paludisme à l'hôpital de Vendrepi en 2006 en fonction du pays d'origine

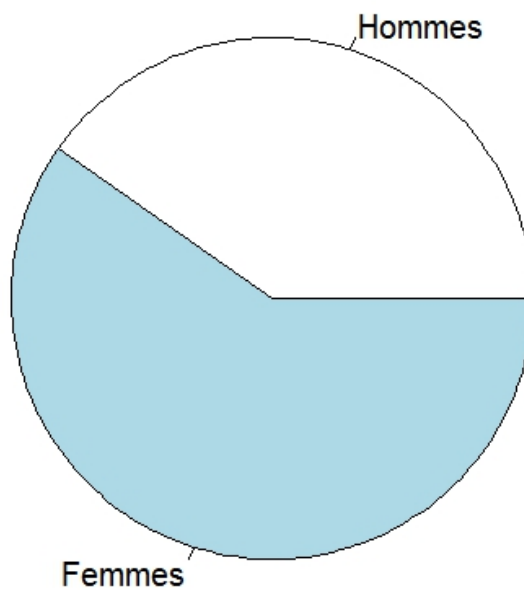


Pour les variables qualitatives nominales

19

# Les camemberts

répartition par sexe



20

# Les camemberts

- Les données doivent être préparées

```
>pie(Mydata$SEXE) # mauvais
```

```
>pie(summary(as.factor(Mydata$SEXE))) # correct
```

- On ajoute un nom "explicite" pour chaque secteur

```
>nom<-c("Homme", "Femme")
```

```
>pie(summary(MyData$SEXE),label=nom)
```

- On ajoute un titre principal

```
>titre<- "Répartition par sexe"
```

```
>pie(summary(MyData$SEXE),label=nom,  
main=titre)
```

21

## Graphiques

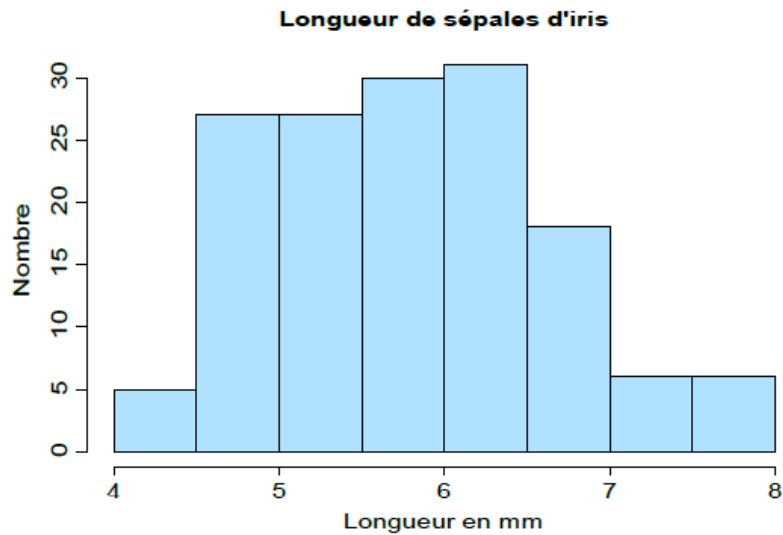
### Représentation des variables quantitatives en utilisant des coordonnées rectangulaires

- Histogramme
- Polygone de fréquence

22

# LES HISTOGRAMMES

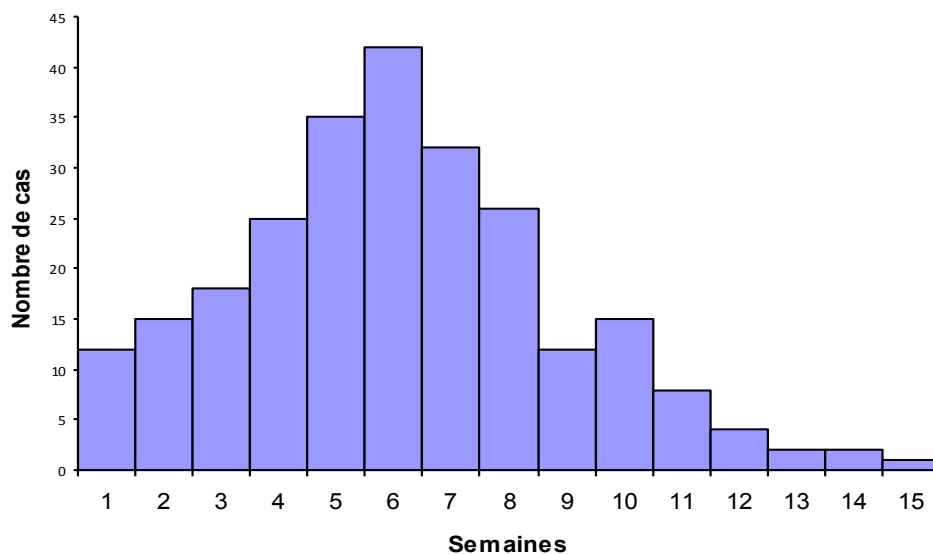
◆ Les variables quantitatives sont découpées en classes d'amplitudes égales. On regroupe toutes les observations qui tombent à l'intérieur d'une même classe.



23

## Histogramme

Evolution des cas de grippe à Vendrepville en 2004



**Pour les variables quantitatives discrètes**

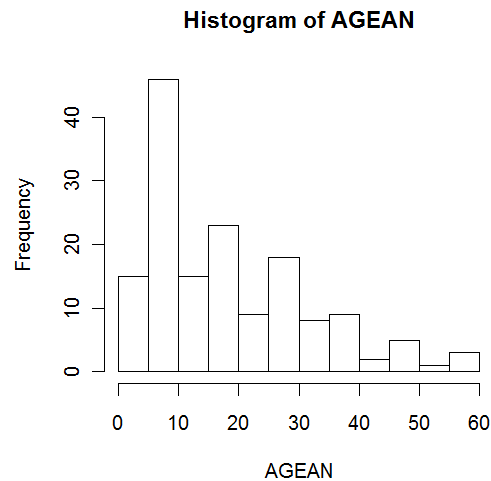
24

# Les histogrammes

## Fonction hist()

- Pour des données quantitatives

```
>hist(Mydata$AGENAN)
```



25

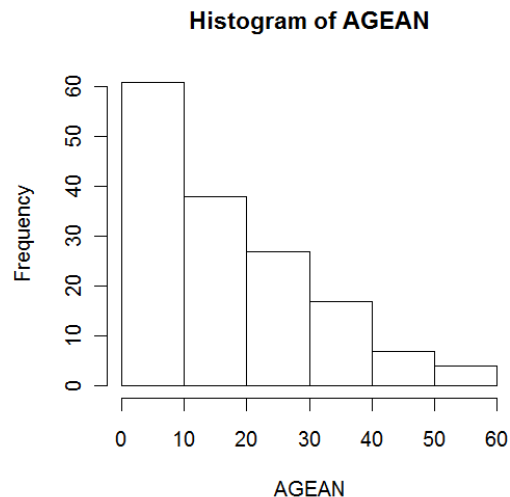
- `hist(Mydata$AGENAN, main="Repartition par tranche d'âge de la population TDR, Niamey, Juillet 2013", xlab="tranches d'âge", ylab="effectifs", col="pink")`

26

## Les histogrammes (2)

- On peut choisir le nombre de barres

```
>hist(Mydata$AGENAN,breaks=c(0,15,30,45,60,75))
```



27

## Les histogrammes (3)

- Pour avoir les bornes et les effectifs des classes

```
>hist(Mydata$AGENAN, nclass=5, plot=F)
```

```
$breaks
```

```
[1] 0 10 20 30 40 50 60
```

```
$counts
```

```
[1] 61 38 27 17 7 4
```

```
$intensities
```

```
[1] 0.039610382 0.024675325 0.017532468 0.011038961 0.004545455 0.002597403
```

```
$density
```

```
[1] 0.039610382 0.024675325 0.017532468 0.011038961 0.004545455 0.002597403
```

```
$mids
```

```
[1] 5 15 25 35 45 55
```

```
$xname
```

```
[1] "AGEAN"
```

```
$equidist
```

```
[1] TRUE
```

```
attr("class")
```

```
[1] "histogram"
```

28

# Les histogrammes (4)

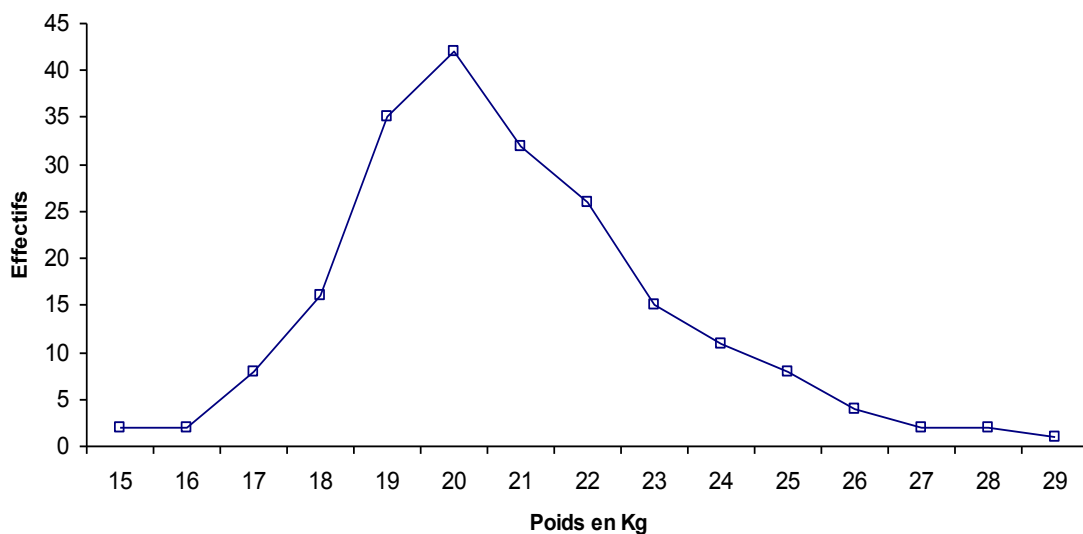
- Choisir ce dont on a besoin :

```
>a<- hist(Mydata$AGENAN, nclass=5, plot=F)
  names(a)
[1] "breaks" "counts" "intensities" "density" "mids"
[6] "xname" "equidist"
> bornes<-a$breaks
>bornes
[1] 0 10 20 30 40 50 60
> Effectifs<-a$counts
>Effectifs
[1] 61 38 27 17 7 4
```

29

## Polygone de Fréquence

Distribution du poids des enfants de 10 ans de Vendrepville en 2006



Pour les variables quantitatives continues

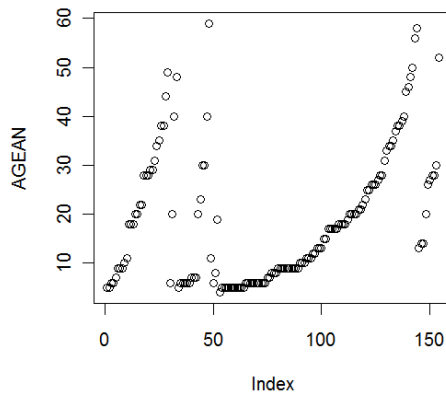
```
plot(Mydata$POIDS,type='l')
```

30

# Fonction plot()

- Pour des données numériques  
→ nuage de point

```
>plot(Mydata$AGENAN)
```



31

- `Mydata$DTEXAM<-as.Date(Mydata$DTEXAM, "%d/%m/%Y")`
- `Mydata$SEMAINE<-as.character(Mydata$DTEXAM,'%U')`
- `Mydata$ANNEE<-as.character(Mydata$DTEXAM,'%Y')`
- `Mydata$INDIC<-paste(Mydata$ANNEE,Mydata$SEMAINE)`
  
- `table(Mydata$INDIC)`
- `plot(table(Mydata$INDIC))`
- `plot(table(Mydata$INDIC),type='l')`

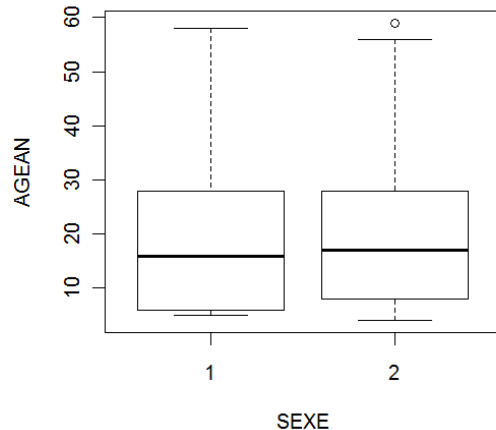
32



# Boites à moustache

- **Boite simple**
  - `boxplot(Mydata$AGENAN)`
- **Pour des données numériques par catégorie → Boite à moustache**

```
>boxplot(Mydata$AGENAN~Mydata$SEXE)
# affichage de "boites à moustache" de l'âge par Sexe
Pour rappel:
> by(AGENAN,SEXE,mean)
SEXE: 1
[1] 18.32258
-----
SEXE: 2
[1] 19.31522
```



33

# Boites à moustache

- **Une boite par niveau de facteur**
  - `Boxplot(Mydata$AGENAN~Mydata$SEXE)`
- **Pour rendre le graphique plus lisible, ajouter :**
  - `main="Age Moyen",`
  - `sub= "Surveillance sentinelle",`
  - `ylab= "Effectifs",`
  - `xlab= "sexe"`
- **Représenter les individus sur l'échelle à gauche**
  - `rug(MyData$AGENAN, side=2)`
- **Exporter le graphique sous différents formats. Menus File/Save as/...**

34

# Ajouter un titre

- Argument : **main**

- 1° Méthode

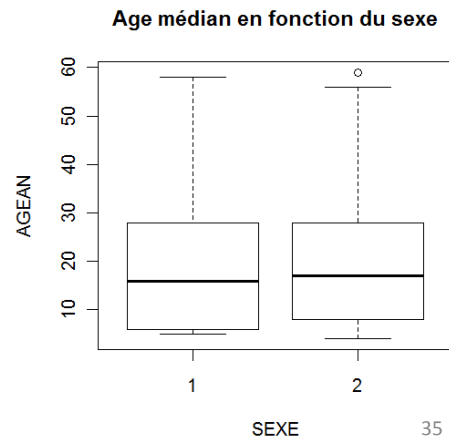
```
> plot(Mydata$AGENAN~Mydata$SEXE, main="Age médian en fonction du sexe")
```

- 2° Méthode

```
>Titre<- "Age médian en fonction du sexe")
```

```
> plot(Mydata$AGENAN~Mydata$SEXE, main=Titre)
```

NB: la variable SEXE aura été factorisée, sinon utiliser la fonction `as.factor()`



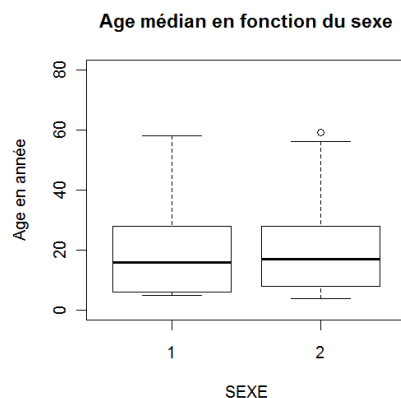
# Axe des Ordonnées

- Arguments :

- **ylab** : label de l'axe

- **ylim** : limites inf et sup de l'échelle

```
>boxplot(Mydata$AGENAN~ Mydata$SEXE, main="Age médian en fonction du sexe", ylab="Age en année", ylim=c(0, 80))
```



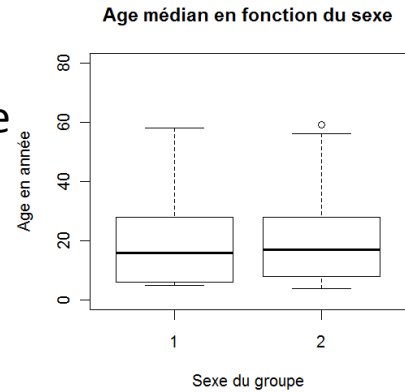
# Axe des Abscisses

- Arguments :

- **xlab** : label de l'axe

- **xlim** : limites inf et sup de l'échelle

```
>boxplot(Mydata$AGENAN~Mydata$SEXE,  
main="Age médian en fonction du sexe",  
ylab="Age en année", ylim=c(0, 80),  
xlab="Sexe du groupe")
```



```
boxplot(Mydata$AGENAN~ as.factor(Mydata$SEXE),  
main="Age médian en fonction du sexe", ylab="Age  
en année", ylim=c(0, 80), xlab="Sexe du groupe")
```

37

## Avec ggplot2

```
library(ggplot2)
```

```
Graph<-ggplot(Mydata, aes(x=SEXE,y=AGENAN))  
+geom_boxplot()
```

```
Graph<-Graph+ggtitle("Moyenne d'age \n en  
fonction du sexe")+xlab("Sexe")+ylab("Age en  
année")
```

```
# ou Graph<-Graph+labs(title="Moyenne d'age  
\n en fonction du sexe", x="Sexe"+y="Age en  
année")
```

38

## Afficher plusieurs graphiques dans la même fenêtre

- Fonction `par()`
  - Argument `mfrow`
  - Exemple:  
`>par(mfrow=c(3, 2))` #partage la fenêtre graphique en 6 régions identiques  
`>plot(Mydata$VALEURF~Mydata$SEXE)` #répéter plusieurs fois

NB: pour supprimer le découpage de la fenêtre graphique

```
>par(mfrow=c(1, 1))
```

39

## Afficher plusieurs graphiques dans la même fenêtre

- Fonction `layout()`
  - Exemple:  
`>zones<-matrix(c(1,1,1,2,3,4),ncol=2)`  
`>layout(zones)`  
`>layout.show(max(zones))` #affiche les régions de figure  
`>plot(Mydata$VALEURF~Mydata$SEXE)` #à répéter 4 fois

NB: pour supprimer le découpage de la fenêtre graphique

```
>layout(matrix(1))
```

40

- `Zone<-matrix(c(1,1,2,2,3,3),ncol=2,byrow=T)`
- `layout(Zone)`
- `layout.show(max(Zone))`

41

## Gérer plusieurs fenêtres graphiques

`windows()` # ouvre une fenêtre

`dev.cur()` # indique la fenêtre active

`dev.list()` # liste les numéros de fenêtres créées et leur nature

`dev.set(x)` # fixe la fenêtre x comme active

`dev.off()` # ferme la fenêtre active

`dev.off(x)` # ferme la fenêtre numéro x

`graphics.off()` # ferme toutes les fenêtres graphiques

42

# Exporter un graphique

pdf(), jpeg(), bmp(), png(), tiff()...

Pour connaître toute la liste ?device

## Exemple1

```
graphics.off()
```

```
pdf(file="exempleA.pdf") #2 graphiques dans le même fichier
```

```
plot(Mydata$NIVSCOL, main="Répartition par centre", col="red")
```

```
hist(Mydata$AGENAN, nclass=5, main="Répartition par âge", col="blue")
```

```
dev.off() # ferme la fenêtre virtuelle , ce qui permet ensuite d'ouvrir le fichier .pdf
```

## Exemple2

```
graphics.off()
```

```
pdf(file="exempleB%03d.pdf", onefile=FALSE) #2 graphiques dans un fichier différent, finir le  
nom du fichier par %03d pour incrémenter le nom des fichiers
```

```
Plot(Mydata$NIVSCOL, main= "Repartition par niveau scolaire", col="red")
```

```
hist(Mydata$AGENAN, nclass=5, main= "Repartition par âge", col="blue")
```

```
dev.off()
```

43

## Fonctions graphiques de haut niveau

- Réalisation de graphiques complets avec plus ou moins d'information
  - plot()
  - hist()
  - pie()

44

# Arguments des fonctions graphiques

**xlim** #limites inférieures et supérieures de l'échelle sur l'axe des abscisses  
**ylim** #limites inférieures et supérieures de l'échelle sur l'axe des ordonnées  
**xlab** #légende sur l'axe des abscisses  
**ylab** #legend sur l'axe des ordonnées  
**main** # titre du graphique  
**sub** # sous titre du graphique  
**cex.axis** #taille des valeurs pour échelles des axes.(1 valeur par défaut)  
**cex.lab** # taille des caractères pour les légendes des axes (1 valeur par défaut)  
**cex.main** #taille des caractères du titre  
**col** #couleur des symboles et des lignes (col="red")  
    **col.main, col.axis, col.lab, col.sub**  
**bty** # pour tracer le cadre autour de la figure ("l", "o", "7 ",... )

45

## Fonction graphique de bas niveaux

- Pour ajouter des renseignements sur un graphique
- Pour ajouter de nouveaux éléments graphiques
- Se familiariser avec les graphique sous R ,  
demo(graphics)

46

# Paramètres des graphiques

- Fonction par()
- À utiliser avant de construire le graphique
  - adj (ajustement des textes) :  
(0:gauche, 0.5:centré, 1:droite)
  - bg (couleur de fond)
  - col (couleur des éléments du graphique)

47

## Ajout d'éléments sur un graphique

**points()** #ajoute des points

**lines()** #ajoute des points reliés par une ligne

**abline()** #ajoute une ligne sans limite ex: abline(h=3) ligne horizontale d'ordonnée=3

**segments()** #ajoute un segment de droite

**arrows()** #trace une flèche

**legend()** # ajoute une légende

**rug()** # ajoute des marques secondaires sur les axes

**axis()** # ajoute une échelle sur un côté du graphique

**text()** #ajoute du texte

**mtext()** #ajoute du texte dans les marges

**title()** #ajoute du texte dans le titre

48



# Attribut pch

