

Manipulation de données avec R

Mamadou Mbodj, Novembre 2024

Contact : mamadou.mbodj@ansd.sn

Cours : Traitement de données

(ENSAE 2024)

Classe : ISE 3



Plan du cours

1. Installation et prise en main de R et RStudio
2. Importation et manipulation de variables et d'individus
3. Visualisation de données
4. Un peu de programmation
5. Exercice pratique de traitement de données avec R

❖ Installation de R

R est un logiciel de statistique distribué librement par le CRAN (Comprehensive R Archive Network) à l'adresse suivante : <http://cran.r-project.org>. L'installation de R varie selon le système d'exploitation (Windows, Mac OS X ou Linux) mais les fonctionnalités sont exactement les mêmes et les programmes sont portables d'un système à l'autre. L'installation de R est très simple, il suffit de suivre les instructions après téléchargement.

❖ Installation de RStudio

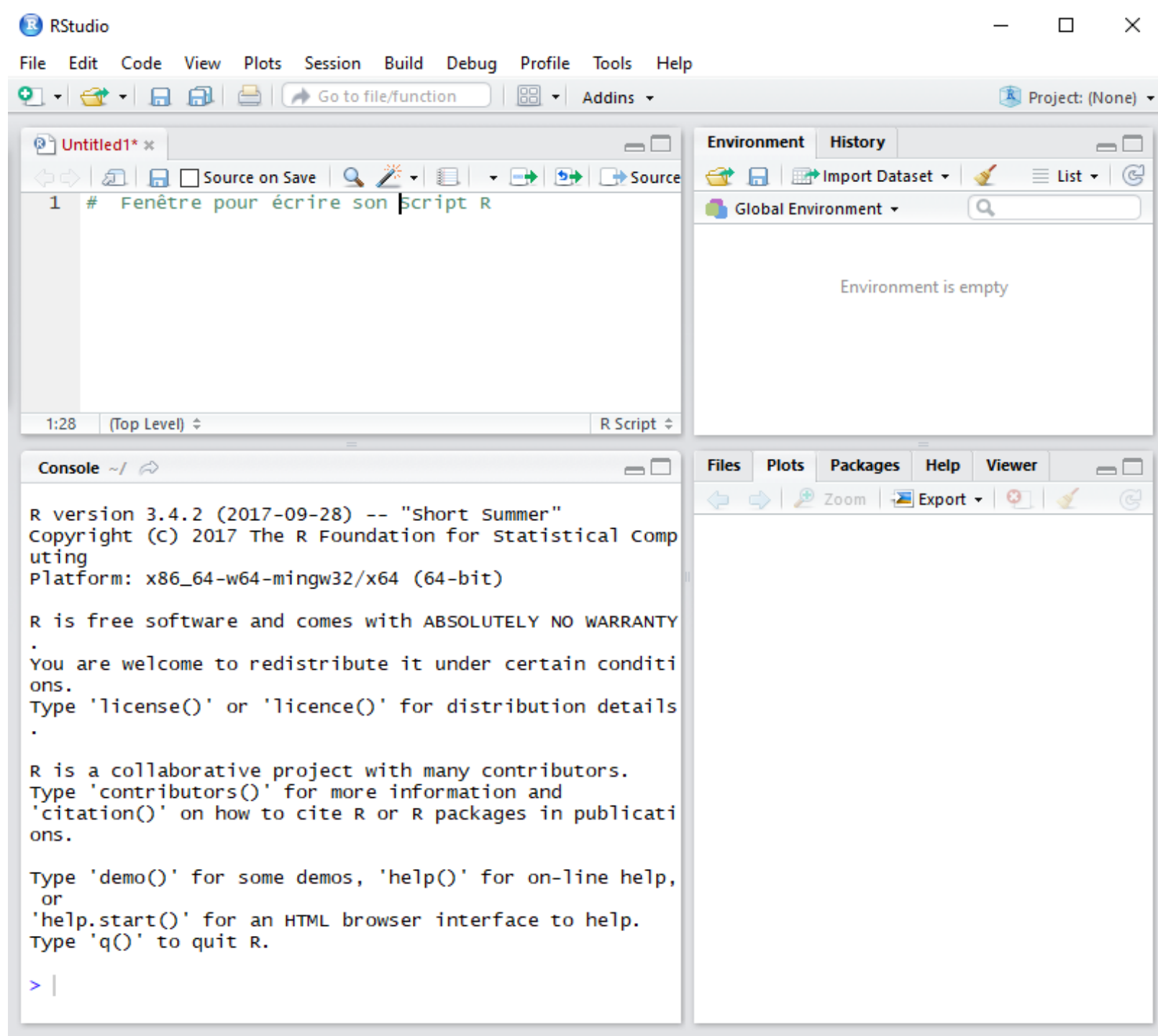
L'utilisation du logiciel R est facilitée par l'utilisation d'un environnement de développement intégré (IDE). Nous préconisons et utiliserons dans ce cours l'environnement RStudio. Son installation suppose d'avoir préalablement installé R mais se fait sans difficulté après téléchargement à l'adresse suivante : <https://posit.co/downloads/>.

Environnement RStudio

Fenêtre globale scindée en 4 sous-fenêtres distinctes :

- fenêtre de scripts (en haut à gauche),
- la console (en bas à gauche),
- la fenêtre d'environnement et d'historique (en haut à droite),
- la fenêtre des fichiers, graphes, packages et d'aide (en bas à droite).

Pour exécuter du code R écrit dans la fenêtre de scripts, on appuie sur le bouton Run.



❖ Introduction à Rmarkdown : insertion de texte

```
# Titre de niveau 1
## Titre de niveau 2
### Titre de niveau 3
```

```
*Ecrire en italique*
**Ecrire en gras**
```

```
Insérer un lien : <https://cran.r-project.org/>
```

```
Listes :
```

```
* item 1
* item 2
```

```
1. item numero 1
2. item numero 2
```

1. Installation et prise en main de R et RStudio

❖ Introduction à Rmarkdown : insertion de code

```
```${r cars}
summary(cars)
```
```

Option pour ne pas
afficher le code dans le
document final

```
```${r, echo=FALSE}
2+2
```
```

D'autres options

results = 'hide' : Ne pas afficher les résultats de compilation

fig.show = 'hide' : Ne pas afficher les figures

warning = False : Ne pas afficher les warning

Etc.

1. Installation et prise en main de R et RStudio

5

❖ Objets de R

○ Création et affectation

```
> b <- 41.3 # crée l'objet b en lui donnant la valeur 41.3  
> x <- b    # x reçoit la valeur b  
> x = b     # x reçoit la valeur b  
> b -> x    # x reçoit la valeur b
```

○ Affichage

```
> print(x)
```

○ Suppression

```
> rm(x)
```

❖ Objets de R

- Affichage des types des objets

```
> mode(x)
```

```
> is.null(x)
```

```
> is.logical(x)
```

```
> is.numeric(x)
```

```
> is.complex(x)
```

```
> is.character(x)
```

- Conversion du type d'un objet

```
> as.logical(x)
```

```
> as.numeric(x)
```

```
> as.complex(x)
```

```
> as.character(x)
```


2. Importation et manipulation de variables et d'individus

❖ Importation de données

```
> tablo <- read.table("C:/Temp/donnees.csv", sep=";", header=TRUE,  
  dec=".", row.names=1)
```

❖ Statistiques sommaires

```
> summary(tablo)
```

| taille | poids | pointure | sexe |
|---------------|---------------|---------------|------|
| Min. :175.5 | Min. :72.00 | Min. :40.00 | F:1 |
| 1st Qu.:176.8 | 1st Qu.:75.00 | 1st Qu.:41.50 | M:3 |
| Median :178.0 | Median :78.00 | Median :42.50 | |
| Mean :179.2 | Mean :76.67 | Mean :42.25 | |
| 3rd Qu.:181.0 | 3rd Qu.:79.00 | 3rd Qu.:43.25 | |
| Max. :184.0 | Max. :80.00 | Max. :44.00 | |
| NA's : 1.0 | NA's : 1.00 | | |

2. Importation et manipulation de variables et d'individus

❖ Manipuler des variables

- Transformation de numérique en facteur

```
> Xqual <- factor(X)
> Xqual
[1] 10 10 10 12 12 13 13 13 13
Levels: 10 12 13
> summary(Xqual)
10 12 13
 3  2  4
```

- Transformation de numérique en catégories

```
> decoupe <- quantile(X, probs=seq(0,1,length=4))
> Xqual <- cut(X, breaks=decoupe, include.lowest=TRUE)
> table(Xqual)
Xqual
[-0.983,-0.544]  (-0.544,0.311]  (0.311,1.69]
               5                5                5
```

2. Importation et manipulation de variables et d'individus

- ❖ Manipuler des individus
 - Append et Merge

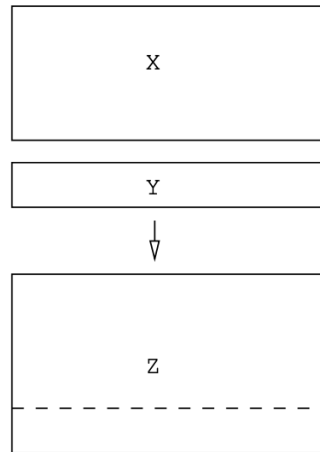


FIGURE 2.2 – Concaténation par ligne : **rbind**(X,Y).

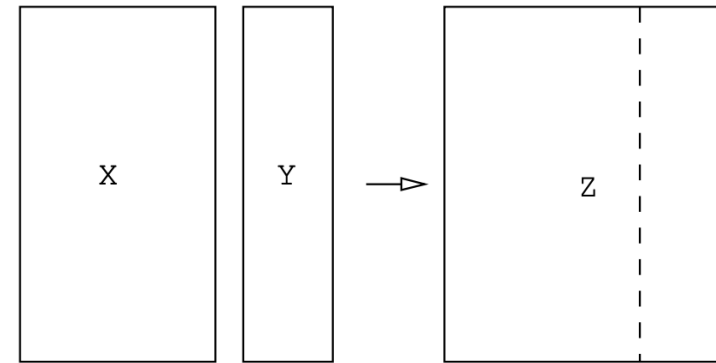


FIGURE 2.3 – Concaténation par colonne : **cbind**(X,Y).

2. Importation et manipulation de variables et d'individus

❖ Tableau croisé

```
> table(don$tension,don$laine)
      Ang Mer Tex
Faible   2   3   0
Forte    1   0   4
```

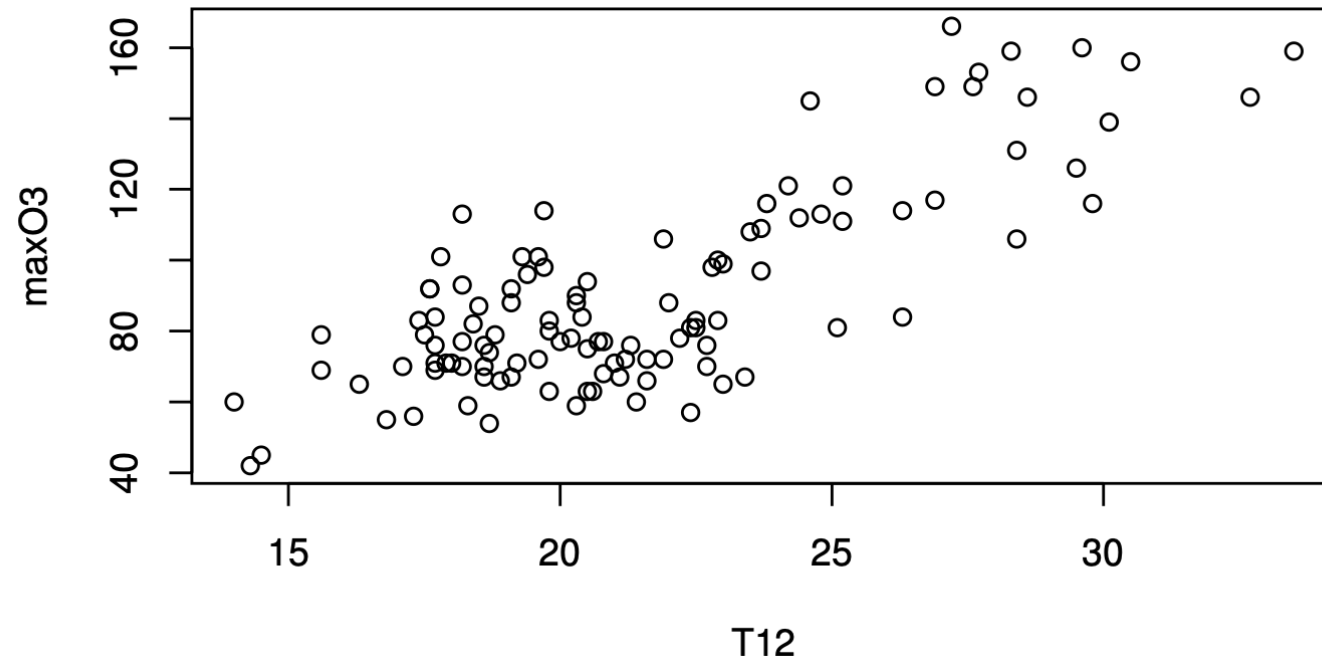
D'autres types de tableaux :

- `prop.table()` #Proportions conjointes
- `prop.table(, 1)` #Proportions en ligne
- `prop.table(, 2)` #Proportions en colonne

3. Visualisation de données

❖ Nuages de points

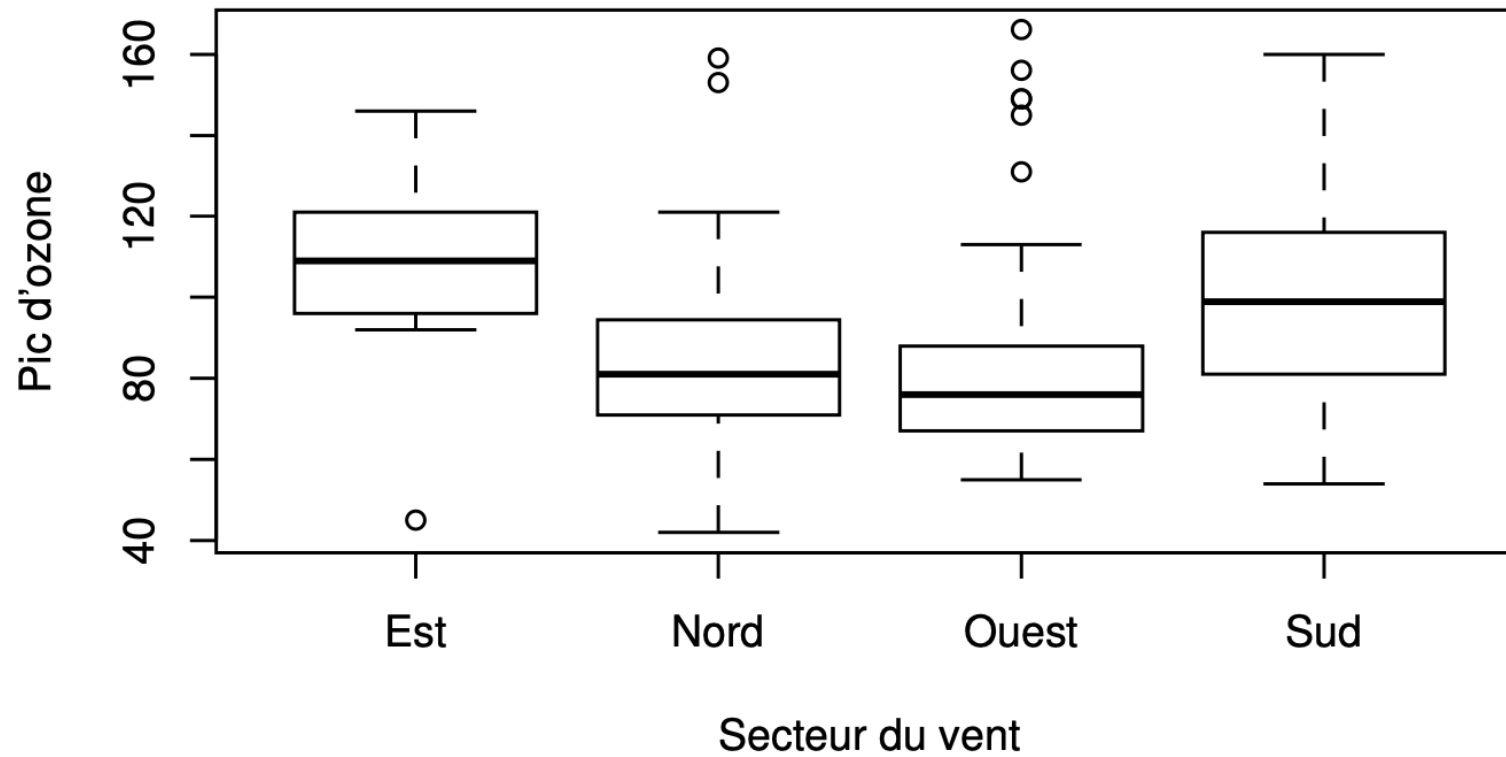
```
> plot(maxO3~T12,data=ozone)
```



3. Visualisation de données

❖ Boxplots

```
> boxplot(maxO3~vent,data=ozone)
```



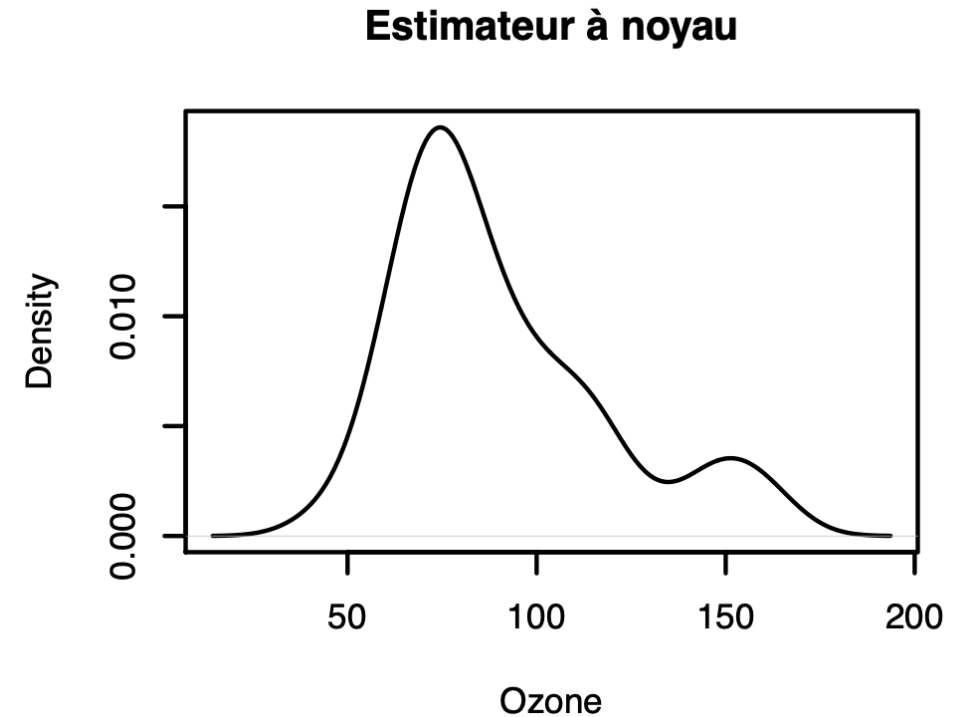
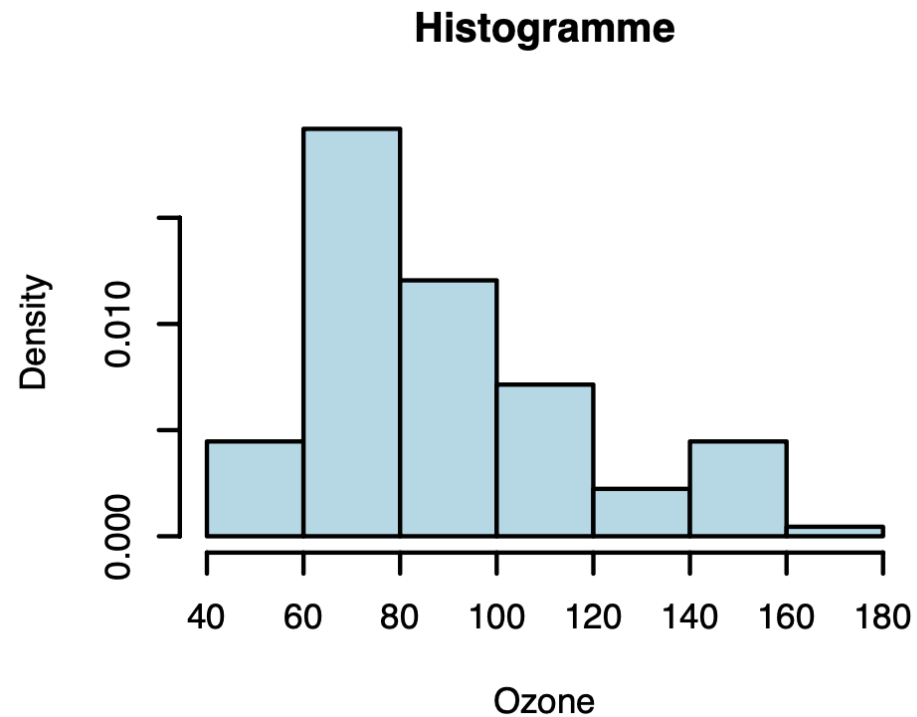
3. Visualisation de données

13

❖ Histogrammes

```
> hist(ozone[, "maxO3"], main="Histogramme", prob=TRUE, xlab="Ozone",  
       col="lightblue")
```

```
> plot(density(ozone[, "maxO3"]), main="Estimateur à noyau", xlab="Ozone")
```

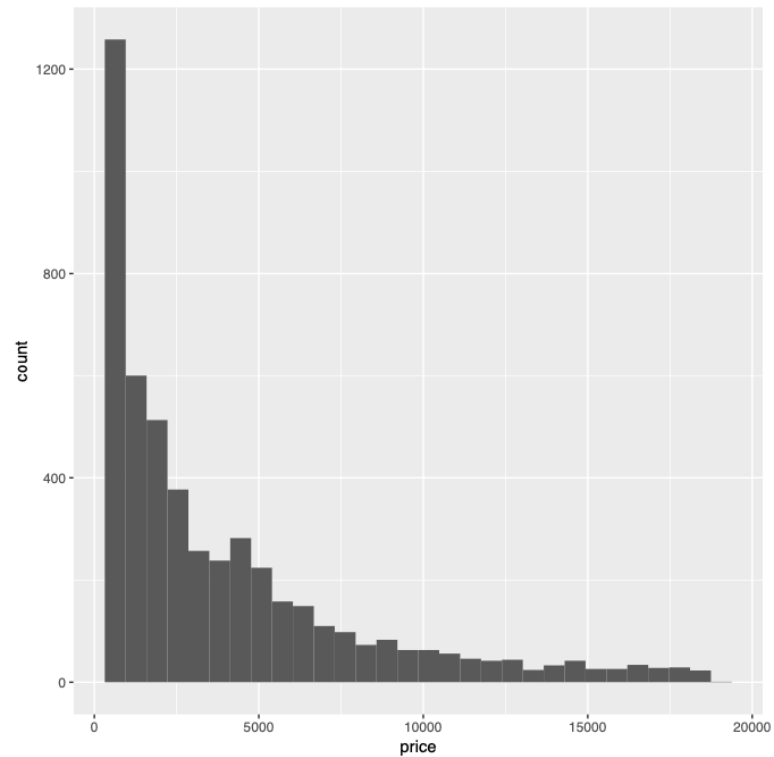
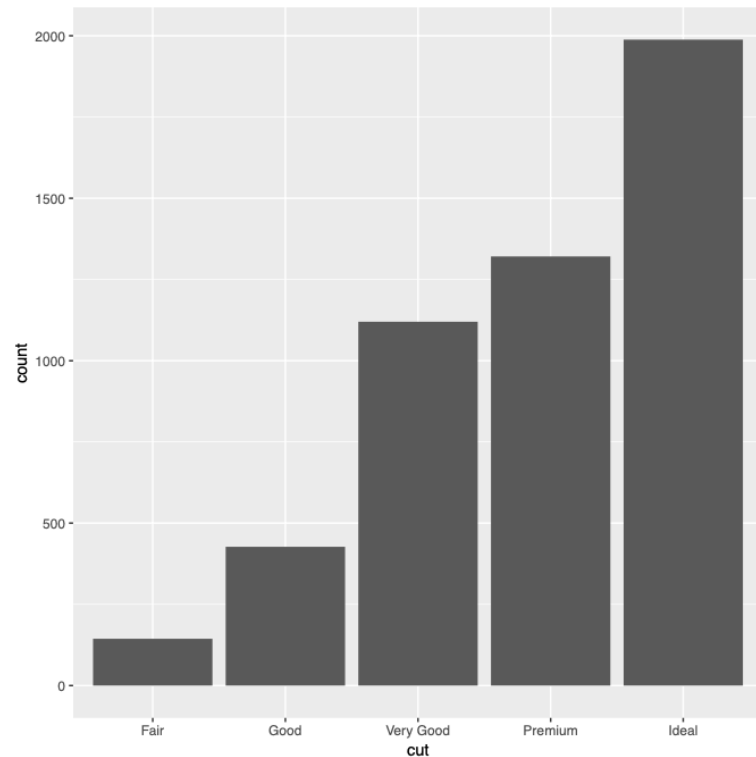


3. Visualisation de données

❖ Librairie ggplot2

```
> ggplot(diamonds2)+aes(x=cut)+geom_bar()
```

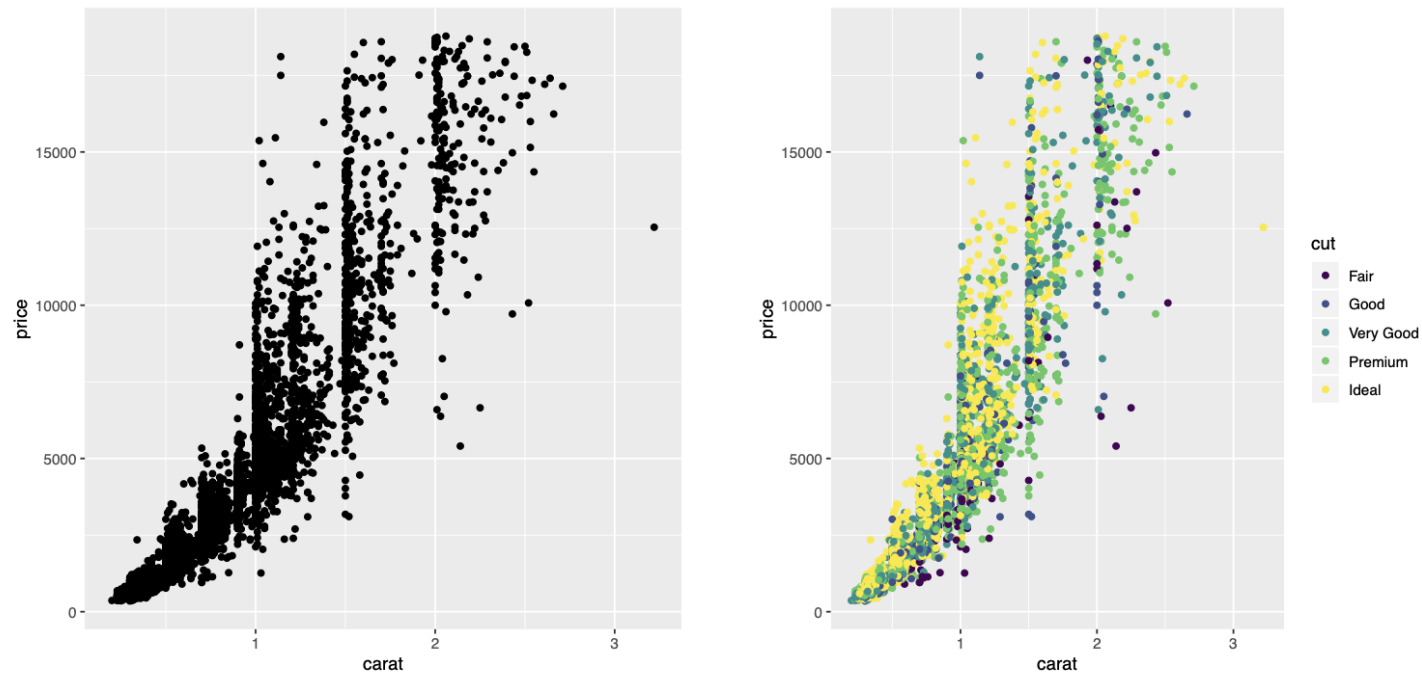
```
> ggplot(diamonds2)+aes(x=price)+geom_histogram()
```



3. Visualisation de données

❖ Librairie ggplot2

```
> ggplot(diamonds2)+aes(x=carat,y=price,color=cut)+geom_point()
```



4. Un peu de programmation

16

❖ Fonctions dans R

- Définition d'une fonction qui calcule une somme de 1 à n

```
som <- function(n) {  
  resultat <- sum(1:n)  
  return(resultat)  
}
```

- Affichage de résultats

```
> som(3)  
[1] 6
```

4. Un peu de programmation

17

❖ Structures de contrôle

○ If_else

```
> if (condition) {  
  expr1  
  expr2  
  ...  
}
```

○ Boucle for / while

```
> vecteur <- c("lundi", "mardi", "mercredi")  
> for (i in vecteur) print(i)  
[1] "lundi"  
[1] "mardi"  
[1] "mercredi"
```

```
> i <- 1  
> while (i<3) {  
  print(i)  
  i <- i+1 }  
[1] 1  
[1] 2
```

5. Exercice pratique de traitement de données avec R

Voir script dans le site du cours : <https://github.com/MMbodj/Traitement-de-donnees-ENSAE-2024>