

```
library(ggplot2) # Sert à importer ggplot2 pour avoir plusieurs plots
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(coin)
```

```
## Loading required package: survival
```

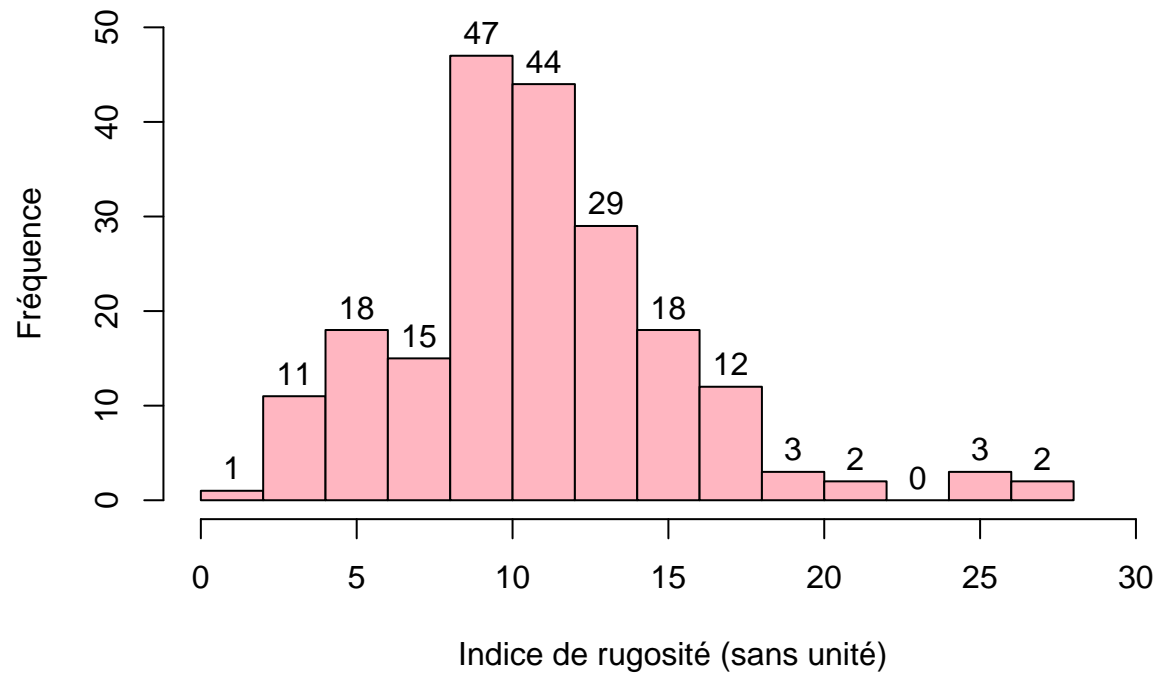
```
#initialisation
source("charger.R")
mondata <- charger(2212435)
```

Phase 1:

Partie a)

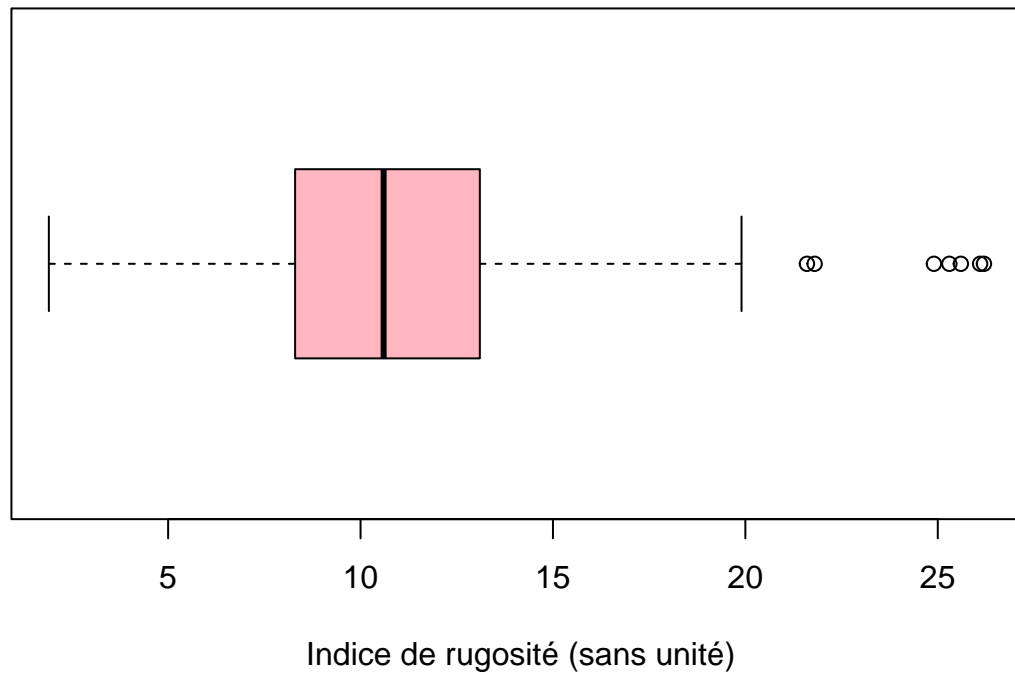
```
#Histogramme
h <- hist(mondata$IR, main = paste("Histogramme de l'indice de rugosité"), col = "lightpink", xlab = "I
text(h$mids,h$counts,labels=h$counts, adj=c(0.5, -0.5))
```

Histogramme de l'indice de rugosité



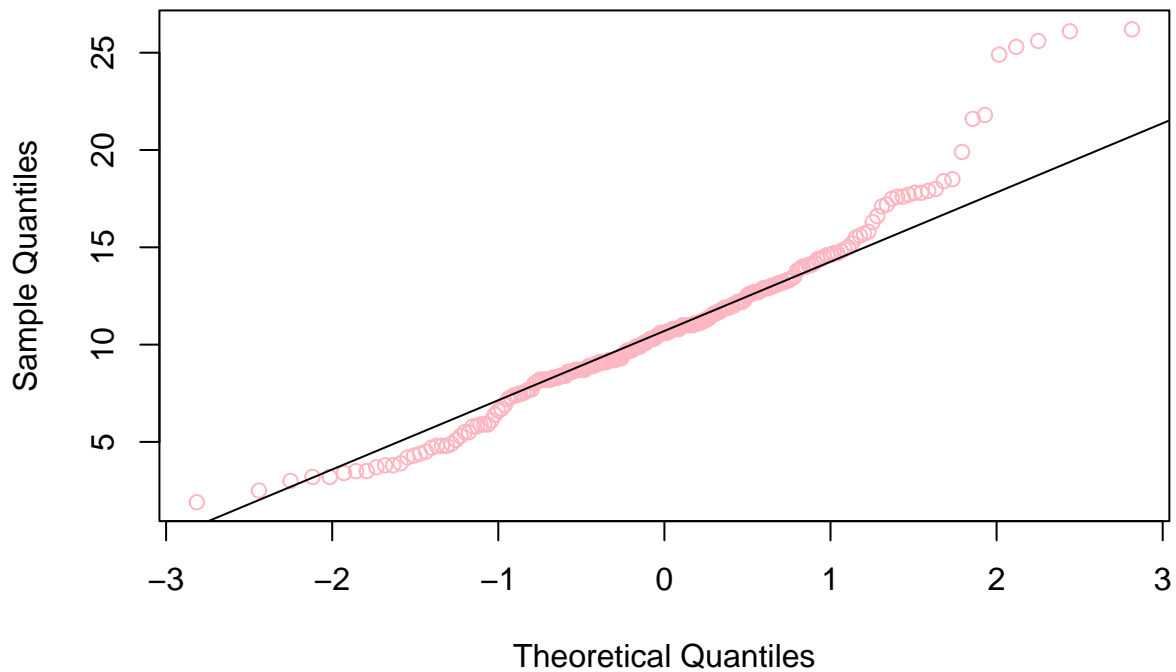
```
#tukey  
boxplot(mondata$IR, main=paste("Diagramme de tukey de l'indice de rugosité"), horizontal=T, xlab =
```

Diagramme de tukey de l'indice de rugosité



```
#droite de Henry  
qqnorm(mondata$IR, col= "lightpink", main = paste("Droite de Henry de l'indice de rugosité"))  
qqline(mondata$IR)
```

Droite de Henry de l'indice de rugosité



```
#Shapiro  
shapiro.test(mondata$IR)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  mondata$IR  
## W = 0.95471, p-value = 4.275e-06
```

```
#tableau statistiques descriptives
```

```
summary_table <- mondata %>%  
  summarise(  
    moyenne = mean(mondata$IR),  
    ecart_type_2 = sd(mondata$IR),  
    q1 = quantile(mondata$IR, 0.25),  
    q2 = quantile(mondata$IR, 0.5),  
    q3 = quantile(mondata$IR, 0.75),  
    Interval_Confiance_Min = t.test(mondata$IR)$conf.int[1],  
    Interval_Confiance_Max = t.test(mondata$IR)$conf.int[2]  
  )
```

```
knitr::kable(summary_table, caption = "Tableau des statistiques descriptives sur l'indice de rugosité")
```

Table 1: Tableau des statistiques descriptives sur l'indice de rugosité

moyenne	ecart_type_2	q1	q2	q3	Interval_Confiance_Min	Interval_Confiance_Max
10.85854	4.517969	8.3	10.6	13.1	10.23638	11.48069

Explications:

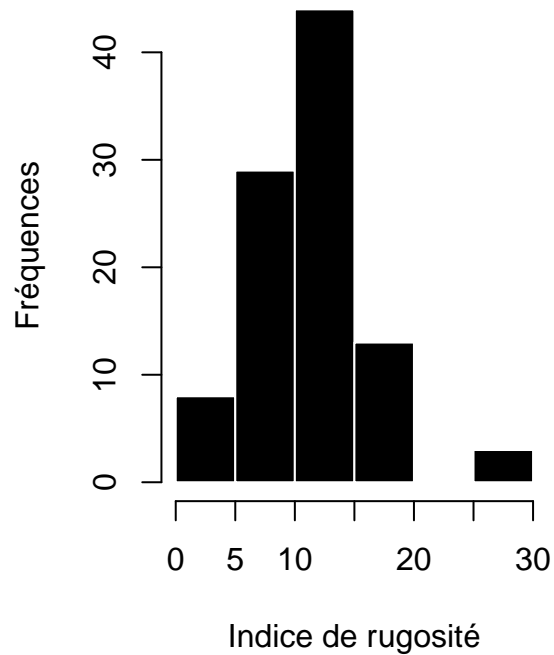
L'histogramme de l'indice de rugosité révèle des caractéristiques significatives sur la distribution de cette mesure au sein de notre ensemble de données. Deux classes, $8 < x < 10$ et $10 < x < 12$, émergent nettement avec les fréquences les plus élevées, suggérant que la majorité des observations d'IR se concentrent dans ces plages. De plus, la dispersion des données n'est pas uniforme, indiquant des concentrations spécifiques plutôt qu'une répartition égale. Nous pouvons voir avec l'histogramme et le diagramme de Tukey que les valeurs de l'IR sont assez bien réparties et que les valeurs sont plus étendues vers la droite, ce qui explique pourquoi la médiane (q2) est un peu moins élevée que la moyenne. Nous constatons aussi des données aberrantes dans le Diagramme de Tukey. Ces valeurs aberrantes peuvent avoir un impact sur la moyenne et l'écart-type. C'est aussi cette distribution anormale (visible à droite de la droite d'Henry) qui amène les données à échouer le test de normalité. En effet, le test de Shapiro-Wilk nous donne une valeur p très proche de 0 ($4.275e-06$), ce qui est inférieur au seuil de 0.05.

partie b)

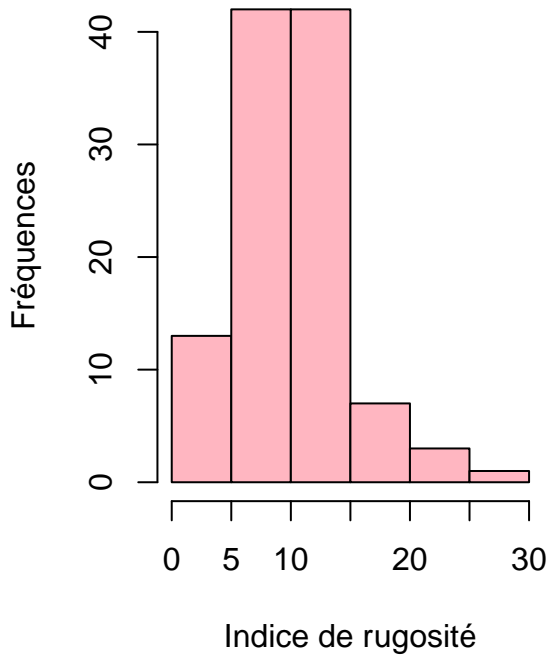
```
# sous elements pour les types de M
matiere_A <- filter(mondata, M == 0)
matiere_B <- filter(mondata, M == 1)

layout(matrix(1:2,1,2)) # permet de diviser la sortie graphique en deux
hist(main = paste("Histogramme Matériau 0"), matiere_A$IR, col="black",border="white",xlab="Indice de r
hist(matiere_B$IR, col="lightpink",border="black",
main=paste("Histogramme Matériau 1"),xlab="Indice de rugosité",ylab="Fréquences")
```

Histogramme Materiau 0



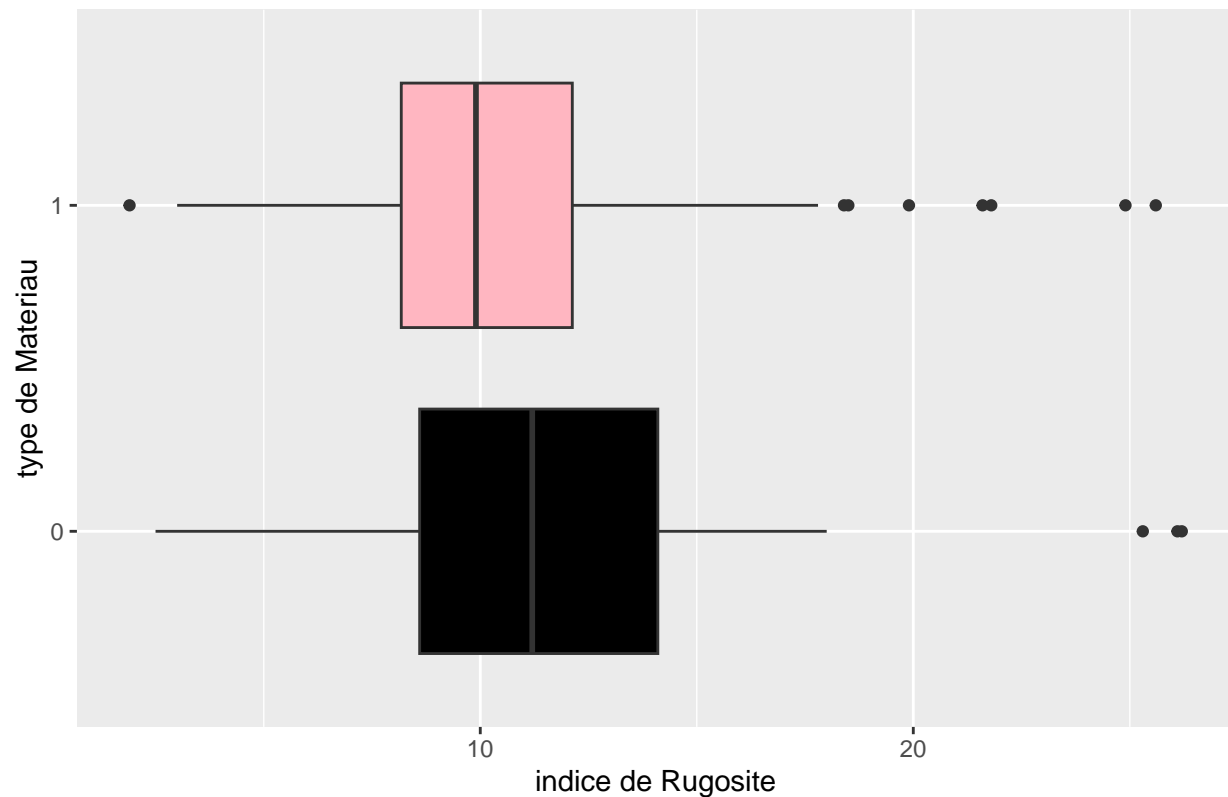
Histogramme Materiau 1



#Diagramme de Tukey

```
ggplot(mondata, aes(y = factor(M), x = IR, fill = factor(M))) + geom_boxplot(fill = c("black", "lightpink"))
```

Diagramme de Tukey de l'IR par type de Matériau



```
#tableau des statistiques descriptives par groupe
summary_table <- mondata %>%
  group_by(M) %>%
  summarise(
    moyenne = mean(IR),
    ecart_type_2 = sd(IR),
    q1 = quantile(IR, 0.25),
    q2 = quantile(IR, 0.5),
    q3 = quantile(IR, 0.75),
    Interval_Confiance_Min = t.test(IR)$conf.int[1],
    Interval_Confiance_Max = t.test(IR)$conf.int[2]
  )
```

knitr::kable(summary_table, caption = "Tableau des statistiques descriptives sur l'indice de rugosité p

Table 2: Tableau des statistiques descriptives sur l'indice de rugosité pour le matériau 1 et 0

M	moyenne	ecart_type_2	q1	q2	q3	Interval_Confiance_Min	Interval_Confiance_Max
0	11.49381	4.598071	8.600	11.2	14.100	10.567098	12.42053
1	10.28796	4.387849	8.175	9.9	12.125	9.450959	11.12497

Explication:

Tout d'abord, les histogrammes nous permet de constater que la plupart des valeurs pour les deux matériaux se situent dans les mêmes classes (classe 5-10 et classe 10-15), avec des valeurs plus étendues vers la droite. Les histogrammes et le diagramme nous permettent de constater que la moyenne du Matériau 0 est plus élevée que celle du Matériau 1 par leurs distributions. Cette observation est cohérente avec les calculs des statistiques descriptives. En effet, en examinant les moyennes, le Matériau 0 affiche une moyenne de 11.49 et de 10.29 pour Matériau 1, confirmant la tendance constatée. De plus, les quartiles et les intervalles de confiance décrivent la répartition des données, montrant que le Matériau 0 a une distribution plus étendue que le Matériau 1. Cette conclusion est également étayée par l'écart-type, qui est légèrement plus élevé pour le Matériau 0 (4.60) que pour le Matériau 1 (4.39). Les observations du diagramme de Tukey renforce ces résultats, suggérant que le centre du Matériau 0 est inférieur à celui du Matériau 1. En résumé, on peut constater des différences entre les deux matériaux que ce soit de tendance centrale, de dispersion ou de la présence de données aberrantes.

Avant de procéder aux tests d'hypothèses, nous allons effectuer le test de shapiro pour vérifier si l'échantillon d'IR pour les deux types de matières suivent une loi normale:

```
shapiro.test(matiere_A$IR)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  matiere_A$IR
## W = 0.95848, p-value = 0.003767
```

```
shapiro.test(matiere_B$IR)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  matiere_B$IR
## W = 0.93804, p-value = 7.851e-05
```

Tests d'hypothese moyenne égale:

Comme la valeur p est inférieure à 0.05 dans les deux cas ($0.003767 < 0.05$ et $7.851e-05 < 0.05$), on rejette l'hypothèse nulle selon laquelle les échantillons suivent une distribution normale. Nous allons donc utiliser le test de student afin de vérifier si les deux moyennes sont égales. Voici nos hypothèses:

$$H_0 : \mu_1 = \mu_2 \quad H_1 : \mu_1 \neq \mu_2$$

Puisque les variances sont inconnues, nous allons utiliser le test de Student suivant:

$$Z_0 = (\bar{X}_1 - \bar{X}_2) / \sqrt{(S_1^2/n_1) + (S_2^2/n_2)}$$

```
# test d'hypothese:
source("charger.R")
monddata <- charger(2212435)
matiere_A <- filter(monddata, M == 0)
matiere_B <- filter(monddata, M == 1)

t.test(matiere_A$IR,matiere_B$IR)
```



```
##
## Welch Two Sample t-test
##
## data: matiere_A$IR and matiere_B$IR
## t = 1.9157, df = 198.26, p-value = 0.05685
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.03546338 2.44716632
## sample estimates:
## mean of x mean of y
## 11.49381 10.28796
```

Interprétation:

Puisque le p-value est supérieur à 0.05 ($0.05685 > 0.05$) donc nous ne pouvons pas rejeter H_0 à un niveau de confiance de 95%. Il n'y a donc pas suffisamment de preuves statistiques pour affirmer que les moyennes des deux groupes sont égales.