project

adem medyouni

2025-04-29

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
mydata <- read.csv("data.csv", header=TRUE)
# Reads the CSV file using semicolon separator
attach(mydata)
typeof(mydata)# Makes the columns of the dataframe available by name
```

```
## [1] "list"
```

```
mydata                                    # Displays the content of the dataframe
```

```
##     total_bill   tip gender smoker  day   time size
## 1        16.99  1.01 Female     No  Sun Dinner    2
## 2        10.34  1.66   Male     No  Sun Dinner    3
## 3        21.01  3.50   Male     No  Sun Dinner    3
## 4        23.68  3.31   Male     No  Sun Dinner    2
## 5        24.59  3.61 Female     No  Sun Dinner    4
## 6        25.29  4.71   Male     No  Sun Dinner    4
## 7         8.77  2.00   Male     No  Sun Dinner    2
## 8        26.88  3.12   Male     No  Sun Dinner    4
## 9        15.04  1.96   Male     No  Sun Dinner    2
## 10       14.78  3.23   Male     No  Sun Dinner    2
## 11       10.27  1.71   Male     No  Sun Dinner    2
## 12       35.26  5.00 Female     No  Sun Dinner    4
## 13       15.42  1.57   Male     No  Sun Dinner    2
## 14       18.43  3.00   Male     No  Sun Dinner    4
## 15       14.83  3.02 Female     No  Sun Dinner    2
## 16       21.58  3.92   Male     No  Sun Dinner    2
## 17       10.33  1.67 Female     No  Sun Dinner    3
## 18       16.29  3.71   Male     No  Sun Dinner    3
## 19       16.97  3.50 Female     No  Sun Dinner    3
## 20       20.65  3.35   Male     No  Sat Dinner    3
## 21       17.92  4.08   Male     No  Sat Dinner    2
## 22       20.29  2.75 Female     No  Sat Dinner    2
## 23       15.77  2.23 Female     No  Sat Dinner    2
```

```
## 24       39.42  7.58   Male     No  Sat Dinner    4
## 25       19.82  3.18   Male     No  Sat Dinner    2
## 26       17.81  2.34   Male     No  Sat Dinner    4
## 27       13.37  2.00   Male     No  Sat Dinner    2
## 28       12.69  2.00   Male     No  Sat Dinner    2
## 29       21.70  4.30   Male     No  Sat Dinner    2
## 30       19.65  3.00 Female     No  Sat Dinner    2
## 31        9.55  1.45   Male     No  Sat Dinner    2
## 32       18.35  2.50   Male     No  Sat Dinner    4
## 33       15.06  3.00 Female     No  Sat Dinner    2
## 34       20.69  2.45 Female     No  Sat Dinner    4
## 35       17.78  3.27   Male     No  Sat Dinner    2
## 36       24.06  3.60   Male     No  Sat Dinner    3
## 37       16.31  2.00   Male     No  Sat Dinner    3
## 38       16.93  3.07 Female     No  Sat Dinner    3
## 39       18.69  2.31   Male     No  Sat Dinner    3
## 40       31.27  5.00   Male     No  Sat Dinner    3
## 41       16.04  2.24   Male     No  Sat Dinner    3
## 42       17.46  2.54   Male     No  Sun Dinner    2
## 43       13.94  3.06   Male     No  Sun Dinner    2
## 44        9.68  1.32   Male     No  Sun Dinner    2
## 45       30.40  5.60   Male     No  Sun Dinner    4
## 46       18.29  3.00   Male     No  Sun Dinner    2
## 47       22.23  5.00   Male     No  Sun Dinner    2
## 48       32.40  6.00   Male     No  Sun Dinner    4
## 49       28.55  2.05   Male     No  Sun Dinner    3
## 50       18.04  3.00   Male     No  Sun Dinner    2
## 51       12.54  2.50   Male     No  Sun Dinner    2
## 52       10.29  2.60 Female     No  Sun Dinner    2
## 53       34.81  5.20 Female     No  Sun Dinner    4
## 54        9.94  1.56   Male     No  Sun Dinner    2
## 55       25.56  4.34   Male     No  Sun Dinner    4
## 56       19.49  3.51   Male     No  Sun Dinner    2
## 57       38.01  3.00   Male    Yes  Sat Dinner    4
## 58       26.41  1.50 Female     No  Sat Dinner    2
## 59       11.24  1.76   Male    Yes  Sat Dinner    2
## 60       48.27  6.73   Male     No  Sat Dinner    4
## 61       20.29  3.21   Male    Yes  Sat Dinner    2
## 62       13.81  2.00   Male    Yes  Sat Dinner    2
## 63       11.02  1.98   Male    Yes  Sat Dinner    2
## 64       18.29  3.76   Male    Yes  Sat Dinner    4
## 65       17.59  2.64   Male     No  Sat Dinner    3
## 66       20.08  3.15   Male     No  Sat Dinner    3
## 67       16.45  2.47 Female     No  Sat Dinner    2
## 68        3.07  1.00 Female    Yes  Sat Dinner    1
## 69       20.23  2.01   Male     No  Sat Dinner    2
## 70       15.01  2.09   Male    Yes  Sat Dinner    2
## 71       12.02  1.97   Male     No  Sat Dinner    2
## 72       17.07  3.00 Female     No  Sat Dinner    3
## 73       26.86  3.14 Female    Yes  Sat Dinner    2
## 74       25.28  5.00 Female    Yes  Sat Dinner    2
## 75       14.73  2.20 Female     No  Sat Dinner    2
## 76       10.51  1.25   Male     No  Sat Dinner    2
## 77       17.92  3.08   Male    Yes  Sat Dinner    2
```

```
## 78       27.20  4.00   Male     No Thur  Lunch    4
## 79       22.76  3.00   Male     No Thur  Lunch    2
## 80       17.29  2.71   Male     No Thur  Lunch    2
## 81       19.44  3.00   Male    Yes Thur  Lunch    2
## 82       16.66  3.40   Male     No Thur  Lunch    2
## 83       10.07  1.83 Female     No Thur  Lunch    1
## 84       32.68  5.00   Male    Yes Thur  Lunch    2
## 85       15.98  2.03   Male     No Thur  Lunch    2
## 86       34.83  5.17 Female     No Thur  Lunch    4
## 87       13.03  2.00   Male     No Thur  Lunch    2
## 88       18.28  4.00   Male     No Thur  Lunch    2
## 89       24.71  5.85   Male     No Thur  Lunch    2
## 90       21.16  3.00   Male     No Thur  Lunch    2
## 91       28.97  3.00   Male    Yes  Fri Dinner    2
## 92       22.49  3.50   Male     No  Fri Dinner    2
## 93        5.75  1.00 Female    Yes  Fri Dinner    2
## 94       16.32  4.30 Female    Yes  Fri Dinner    2
## 95       22.75  3.25 Female     No  Fri Dinner    2
## 96       40.17  4.73   Male    Yes  Fri Dinner    4
## 97       27.28  4.00   Male    Yes  Fri Dinner    2
## 98       12.03  1.50   Male    Yes  Fri Dinner    2
## 99       21.01  3.00   Male    Yes  Fri Dinner    2
## 100      12.46  1.50   Male     No  Fri Dinner    2
## 101      11.35  2.50 Female    Yes  Fri Dinner    2
## 102      15.38  3.00 Female    Yes  Fri Dinner    2
## 103      44.30  2.50 Female    Yes  Sat Dinner    3
## 104      22.42  3.48 Female    Yes  Sat Dinner    2
## 105      20.92  4.08 Female     No  Sat Dinner    2
## 106      15.36  1.64   Male    Yes  Sat Dinner    2
## 107      20.49  4.06   Male    Yes  Sat Dinner    2
## 108      25.21  4.29   Male    Yes  Sat Dinner    2
## 109      18.24  3.76   Male     No  Sat Dinner    2
## 110      14.31  4.00 Female    Yes  Sat Dinner    2
## 111      14.00  3.00   Male     No  Sat Dinner    2
## 112       7.25  1.00 Female     No  Sat Dinner    1
## 113      38.07  4.00   Male     No  Sun Dinner    3
## 114      23.95  2.55   Male     No  Sun Dinner    2
## 115      25.71  4.00 Female     No  Sun Dinner    3
## 116      17.31  3.50 Female     No  Sun Dinner    2
## 117      29.93  5.07   Male     No  Sun Dinner    4
## 118      10.65  1.50 Female     No Thur  Lunch    2
## 119      12.43  1.80 Female     No Thur  Lunch    2
## 120      24.08  2.92 Female     No Thur  Lunch    4
## 121      11.69  2.31   Male     No Thur  Lunch    2
## 122      13.42  1.68 Female     No Thur  Lunch    2
## 123      14.26  2.50   Male     No Thur  Lunch    2
## 124      15.95  2.00   Male     No Thur  Lunch    2
## 125      12.48  2.52 Female     No Thur  Lunch    2
## 126      29.80  4.20 Female     No Thur  Lunch    6
## 127       8.52  1.48   Male     No Thur  Lunch    2
## 128      14.52  2.00 Female     No Thur  Lunch    2
## 129      11.38  2.00 Female     No Thur  Lunch    2
## 130      22.82  2.18   Male     No Thur  Lunch    3
## 131      19.08  1.50   Male     No Thur  Lunch    2
```

```
## 132      20.27  2.83 Female     No Thur  Lunch    2
## 133      11.17  1.50 Female     No Thur  Lunch    2
## 134      12.26  2.00 Female     No Thur  Lunch    2
## 135      18.26  3.25 Female     No Thur  Lunch    2
## 136       8.51  1.25 Female     No Thur  Lunch    2
## 137      10.33  2.00 Female     No Thur  Lunch    2
## 138      14.15  2.00 Female     No Thur  Lunch    2
## 139      16.00  2.00   Male    Yes Thur  Lunch    2
## 140      13.16  2.75 Female     No Thur  Lunch    2
## 141      17.47  3.50 Female     No Thur  Lunch    2
## 142      34.30  6.70   Male     No Thur  Lunch    6
## 143      41.19  5.00   Male     No Thur  Lunch    5
## 144      27.05  5.00 Female     No Thur  Lunch    6
## 145      16.43  2.30 Female     No Thur  Lunch    2
## 146       8.35  1.50 Female     No Thur  Lunch    2
## 147      18.64  1.36 Female     No Thur  Lunch    3
## 148      11.87  1.63 Female     No Thur  Lunch    2
## 149       9.78  1.73   Male     No Thur  Lunch    2
## 150       7.51  2.00   Male     No Thur  Lunch    2
## 151      14.07  2.50   Male     No  Sun Dinner    2
## 152      13.13  2.00   Male     No  Sun Dinner    2
## 153      17.26  2.74   Male     No  Sun Dinner    3
## 154      24.55  2.00   Male     No  Sun Dinner    4
## 155      19.77  2.00   Male     No  Sun Dinner    4
## 156      29.85  5.14 Female     No  Sun Dinner    5
## 157      48.17  5.00   Male     No  Sun Dinner    6
## 158      25.00  3.75 Female     No  Sun Dinner    4
## 159      13.39  2.61 Female     No  Sun Dinner    2
## 160      16.49  2.00   Male     No  Sun Dinner    4
## 161      21.50  3.50   Male     No  Sun Dinner    4
## 162      12.66  2.50   Male     No  Sun Dinner    2
## 163      16.21  2.00 Female     No  Sun Dinner    3
## 164      13.81  2.00   Male     No  Sun Dinner    2
## 165      17.51  3.00 Female    Yes  Sun Dinner    2
## 166      24.52  3.48   Male     No  Sun Dinner    3
## 167      20.76  2.24   Male     No  Sun Dinner    2
## 168      31.71  4.50   Male     No  Sun Dinner    4
## 169      10.59  1.61 Female    Yes  Sat Dinner    2
## 170      10.63  2.00 Female    Yes  Sat Dinner    2
## 171      50.81 10.00   Male    Yes  Sat Dinner    3
## 172      15.81  3.16   Male    Yes  Sat Dinner    2
## 173       7.25  5.15   Male    Yes  Sun Dinner    2
## 174      31.85  3.18   Male    Yes  Sun Dinner    2
## 175      16.82  4.00   Male    Yes  Sun Dinner    2
## 176      32.90  3.11   Male    Yes  Sun Dinner    2
## 177      17.89  2.00   Male    Yes  Sun Dinner    2
## 178      14.48  2.00   Male    Yes  Sun Dinner    2
## 179       9.60  4.00 Female    Yes  Sun Dinner    2
## 180      34.63  3.55   Male    Yes  Sun Dinner    2
## 181      34.65  3.68   Male    Yes  Sun Dinner    4
## 182      23.33  5.65   Male    Yes  Sun Dinner    2
## 183      45.35  3.50   Male    Yes  Sun Dinner    3
## 184      23.17  6.50   Male    Yes  Sun Dinner    4
## 185      40.55  3.00   Male    Yes  Sun Dinner    2
```

```
## 186      20.69  5.00   Male     No  Sun Dinner    5
## 187      20.90  3.50 Female    Yes  Sun Dinner    3
## 188      30.46  2.00   Male    Yes  Sun Dinner    5
## 189      18.15  3.50 Female    Yes  Sun Dinner    3
## 190      23.10  4.00   Male    Yes  Sun Dinner    3
## 191      15.69  1.50   Male    Yes  Sun Dinner    2
## 192      19.81  4.19 Female    Yes Thur  Lunch    2
## 193      28.44  2.56   Male    Yes Thur  Lunch    2
## 194      15.48  2.02   Male    Yes Thur  Lunch    2
## 195      16.58  4.00   Male    Yes Thur  Lunch    2
## 196       7.56  1.44   Male     No Thur  Lunch    2
## 197      10.34  2.00   Male    Yes Thur  Lunch    2
## 198      43.11  5.00 Female    Yes Thur  Lunch    4
## 199      13.00  2.00 Female    Yes Thur  Lunch    2
## 200      13.51  2.00   Male    Yes Thur  Lunch    2
## 201      18.71  4.00   Male    Yes Thur  Lunch    3
## 202      12.74  2.01 Female    Yes Thur  Lunch    2
## 203      13.00  2.00 Female    Yes Thur  Lunch    2
## 204      16.40  2.50 Female    Yes Thur  Lunch    2
## 205      20.53  4.00   Male    Yes Thur  Lunch    4
## 206      16.47  3.23 Female    Yes Thur  Lunch    3
## 207      26.59  3.41   Male    Yes  Sat Dinner    3
## 208      38.73  3.00   Male    Yes  Sat Dinner    4
## 209      24.27  2.03   Male    Yes  Sat Dinner    2
## 210      12.76  2.23 Female    Yes  Sat Dinner    2
## 211      30.06  2.00   Male    Yes  Sat Dinner    3
## 212      25.89  5.16   Male    Yes  Sat Dinner    4
## 213      48.33  9.00   Male     No  Sat Dinner    4
## 214      13.27  2.50 Female    Yes  Sat Dinner    2
## 215      28.17  6.50 Female    Yes  Sat Dinner    3
## 216      12.90  1.10 Female    Yes  Sat Dinner    2
## 217      28.15  3.00   Male    Yes  Sat Dinner    5
## 218      11.59  1.50   Male    Yes  Sat Dinner    2
## 219       7.74  1.44   Male    Yes  Sat Dinner    2
## 220      30.14  3.09 Female    Yes  Sat Dinner    4
## 221      12.16  2.20   Male    Yes  Fri  Lunch    2
## 222      13.42  3.48 Female    Yes  Fri  Lunch    2
## 223       8.58  1.92   Male    Yes  Fri  Lunch    1
## 224      15.98  3.00 Female     No  Fri  Lunch    3
## 225      13.42  1.58   Male    Yes  Fri  Lunch    2
## 226      16.27  2.50 Female    Yes  Fri  Lunch    2
## 227      10.09  2.00 Female    Yes  Fri  Lunch    2
## 228      20.45  3.00   Male     No  Sat Dinner    4
## 229      13.28  2.72   Male     No  Sat Dinner    2
## 230      22.12  2.88 Female    Yes  Sat Dinner    2
## 231      24.01  2.00   Male    Yes  Sat Dinner    4
## 232      15.69  3.00   Male    Yes  Sat Dinner    3
## 233      11.61  3.39   Male     No  Sat Dinner    2
## 234      10.77  1.47   Male     No  Sat Dinner    2
## 235      15.53  3.00   Male    Yes  Sat Dinner    2
## 236      10.07  1.25   Male     No  Sat Dinner    2
## 237      12.60  1.00   Male    Yes  Sat Dinner    2
## 238      32.83  1.17   Male    Yes  Sat Dinner    2
## 239      35.83  4.67 Female     No  Sat Dinner    3
```

```
## 240      29.03  5.92   Male     No  Sat Dinner    3
## 241      27.18  2.00 Female    Yes  Sat Dinner    2
## 242      22.67  2.00   Male    Yes  Sat Dinner    2
## 243      17.82  1.75   Male     No  Sat Dinner    2
## 244      18.78  3.00 Female     No Thur Dinner    2
```

```
T=table(names(mydata))
T
```

```
##
##        day     gender       size     smoker       time        tip total_bill
##          1          1          1          1          1          1          1
```

```
typeof(names(mydata))
```

```
## [1] "character"
```

```
names(mydata)[2]
```

```
## [1] "tip"
```

```
mydata$tip
```

```
##   [1]  1.01  1.66  3.50  3.31  3.61  4.71  2.00  3.12  1.96  3.23  1.71  5.00
##  [13]  1.57  3.00  3.02  3.92  1.67  3.71  3.50  3.35  4.08  2.75  2.23  7.58
##  [25]  3.18  2.34  2.00  2.00  4.30  3.00  1.45  2.50  3.00  2.45  3.27  3.60
##  [37]  2.00  3.07  2.31  5.00  2.24  2.54  3.06  1.32  5.60  3.00  5.00  6.00
##  [49]  2.05  3.00  2.50  2.60  5.20  1.56  4.34  3.51  3.00  1.50  1.76  6.73
##  [61]  3.21  2.00  1.98  3.76  2.64  3.15  2.47  1.00  2.01  2.09  1.97  3.00
##  [73]  3.14  5.00  2.20  1.25  3.08  4.00  3.00  2.71  3.00  3.40  1.83  5.00
##  [85]  2.03  5.17  2.00  4.00  5.85  3.00  3.00  3.50  1.00  4.30  3.25  4.73
##  [97]  4.00  1.50  3.00  1.50  2.50  3.00  2.50  3.48  4.08  1.64  4.06  4.29
## [109]  3.76  4.00  3.00  1.00  4.00  2.55  4.00  3.50  5.07  1.50  1.80  2.92
## [121]  2.31  1.68  2.50  2.00  2.52  4.20  1.48  2.00  2.00  2.18  1.50  2.83
## [133]  1.50  2.00  3.25  1.25  2.00  2.00  2.00  2.75  3.50  6.70  5.00  5.00
## [145]  2.30  1.50  1.36  1.63  1.73  2.00  2.50  2.00  2.74  2.00  2.00  5.14
## [157]  5.00  3.75  2.61  2.00  3.50  2.50  2.00  2.00  3.00  3.48  2.24  4.50
## [169]  1.61  2.00 10.00  3.16  5.15  3.18  4.00  3.11  2.00  2.00  4.00  3.55
## [181]  3.68  5.65  3.50  6.50  3.00  5.00  3.50  2.00  3.50  4.00  1.50  4.19
## [193]  2.56  2.02  4.00  1.44  2.00  5.00  2.00  2.00  4.00  2.01  2.00  2.50
## [205]  4.00  3.23  3.41  3.00  2.03  2.23  2.00  5.16  9.00  2.50  6.50  1.10
## [217]  3.00  1.50  1.44  3.09  2.20  3.48  1.92  3.00  1.58  2.50  2.00  3.00
## [229]  2.72  2.88  2.00  3.00  3.39  1.47  3.00  1.25  1.00  1.17  4.67  5.92
## [241]  2.00  2.00  1.75  3.00
```

```
T1=table(mydata$gender)
T1
```

```
##
## Female   Male
##     87    157
```

```r
# 1. Importer les données
mydata <- read.csv("data.csv", header = TRUE)

# 2. Calcul du nombre d'observations
n <- nrow(mydata)

# 3. Calcul du nombre de classes (J)
J_sturges <- ceiling(1 + (10 * log10(n) / 3))  # Règle de Sturges
J_yule <- ceiling(2.5 * n^(1/4))               # Règle de Yule

# Affichage des deux méthodes
cat("Nombre de classes selon Sturges :", J_sturges, "\n")
```

```
## Nombre de classes selon Sturges : 9
```

```r
cat("Nombre de classes selon Yule    :", J_yule, "\n")
```

```
## Nombre de classes selon Yule    : 10
```

```r
# 4. Choisir un nombre de classes (par exemple Sturges)
J <- J_sturges

# 5. Répartition en classes pour total_bill
min_tb <- min(mydata$total_bill, na.rm = TRUE)
max_tb <- max(mydata$total_bill, na.rm = TRUE)
amplitude_tb <- (max_tb - min_tb) / J
cat("Amplitude total_bill :", amplitude_tb, "\n")
```

```
## Amplitude total_bill : 5.304444
```

```r
classes_tb <- cut(mydata$total_bill, breaks = J)
table_tb <- table(classes_tb)
print(table_tb)
```

```
## classes_tb
## (3.02,8.37] (8.37,13.7]   (13.7,19]   (19,24.3] (24.3,29.6] (29.6,34.9]
##           8          59          73          44          25          19
## (34.9,40.2] (40.2,45.5] (45.5,50.9]
##           7           5           4
```

```r
# 6. Répartition en classes pour tip
min_tip <- min(mydata$tip, na.rm = TRUE)
max_tip <- max(mydata$tip, na.rm = TRUE)
amplitude_tip <- (max_tip - min_tip) / J
cat("Amplitude tip :", amplitude_tip, "\n")
```

```
## Amplitude tip : 1
```

```r
classes_tip <- cut(mydata$tip, breaks = J)
table_tip <- table(classes_tip)
print(table_tip)
```

```
## classes_tip
## (0.991,2]     (2,3]     (3,4]     (4,5]     (5,6]     (6,7]     (7,8]     (8,9]
##        78        68        57        23        11         4         1         1
##    (9,10]
##         1
```

```r
freq_tb_rel <- prop.table(table_tb)
freq_tb_cum <- cumsum(freq_tb_rel)

# Résumé complet
stat_tb <- data.frame(
  Classe = names(table_tb),
  Effectif = as.numeric(table_tb),
  Fréquence = round(as.numeric(freq_tb_rel), 3),
  Fréquence_cumulée = round(as.numeric(freq_tb_cum), 3)
)

print(stat_tb)
```

```
##        Classe Effectif Fréquence Fréquence_cumulée
## 1 (3.02,8.37]        8     0.033             0.033
## 2 (8.37,13.7]       59     0.242             0.275
## 3   (13.7,19]       73     0.299             0.574
## 4   (19,24.3]       44     0.180             0.754
## 5 (24.3,29.6]       25     0.102             0.857
## 6 (29.6,34.9]       19     0.078             0.934
## 7 (34.9,40.2]        7     0.029             0.963
## 8 (40.2,45.5]        5     0.020             0.984
## 9 (45.5,50.9]        4     0.016             1.000
```

```r
#pour tip:
freq_tip_rel <- prop.table(table_tip)
freq_tip_cum <- cumsum(freq_tip_rel)

# Résumé
stat_tip <- data.frame(
  Classe = names(table_tip),
  Effectif = as.numeric(table_tip),
  Fréquence = round(as.numeric(freq_tip_rel), 3),
  Fréquence_cumulée = round(as.numeric(freq_tip_cum), 3)
)

print(stat_tip)
```

```
##      Classe Effectif Fréquence Fréquence_cumulée
## 1 (0.991,2]       78     0.320             0.320
## 2     (2,3]       68     0.279             0.598
## 3     (3,4]       57     0.234             0.832
```

```
## 4     (4,5]       23     0.094             0.926
## 5     (5,6]       11     0.045             0.971
## 6     (6,7]        4     0.016             0.988
## 7     (7,8]        1     0.004             0.992
## 8     (8,9]        1     0.004             0.996
## 9    (9,10]        1     0.004             1.000
```

```
#gender:
T_gender <- table(mydata$gender)
prop.table(T_gender)
```

```
##
##    Female      Male
## 0.3565574 0.6434426
```

```
#day:
T_day <- table(mydata$day)
prop.table(T_day)
```

```
##
##        Fri        Sat        Sun       Thur
## 0.07786885 0.35655738 0.31147541 0.25409836
```

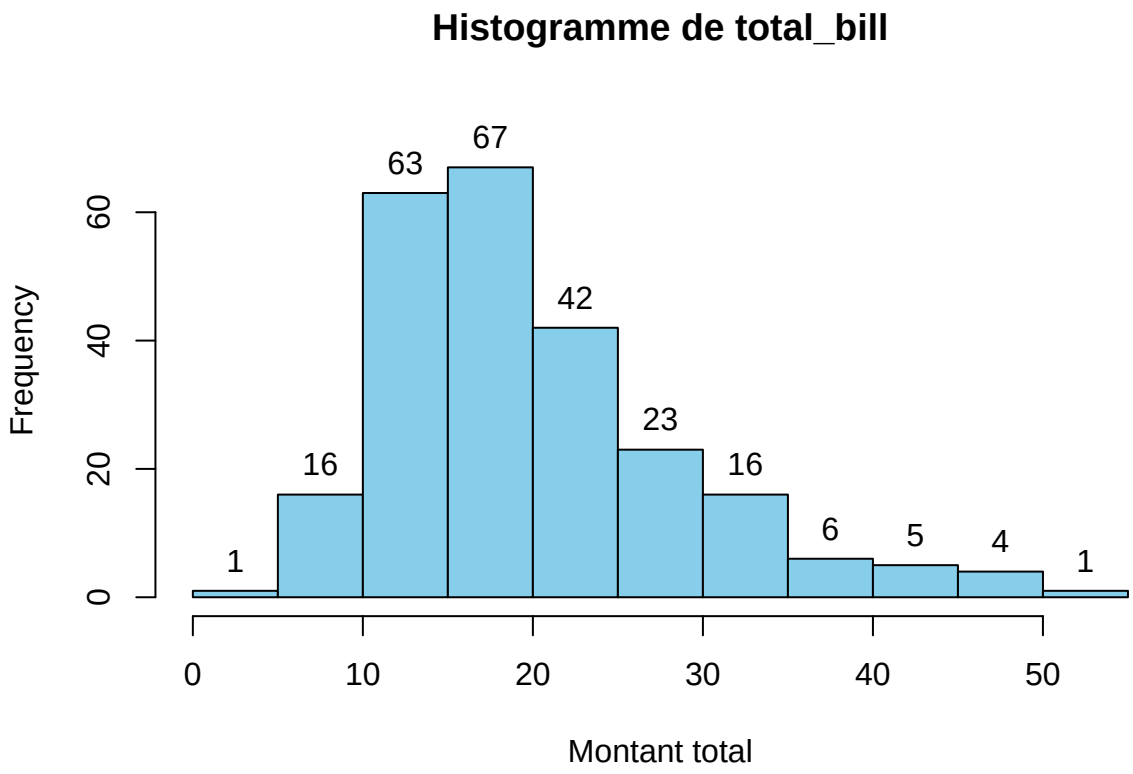
```
#time:
T_time <- table(mydata$time)
prop.table(T_time)
```

```
##
##    Dinner     Lunch
## 0.7213115 0.2786885
```

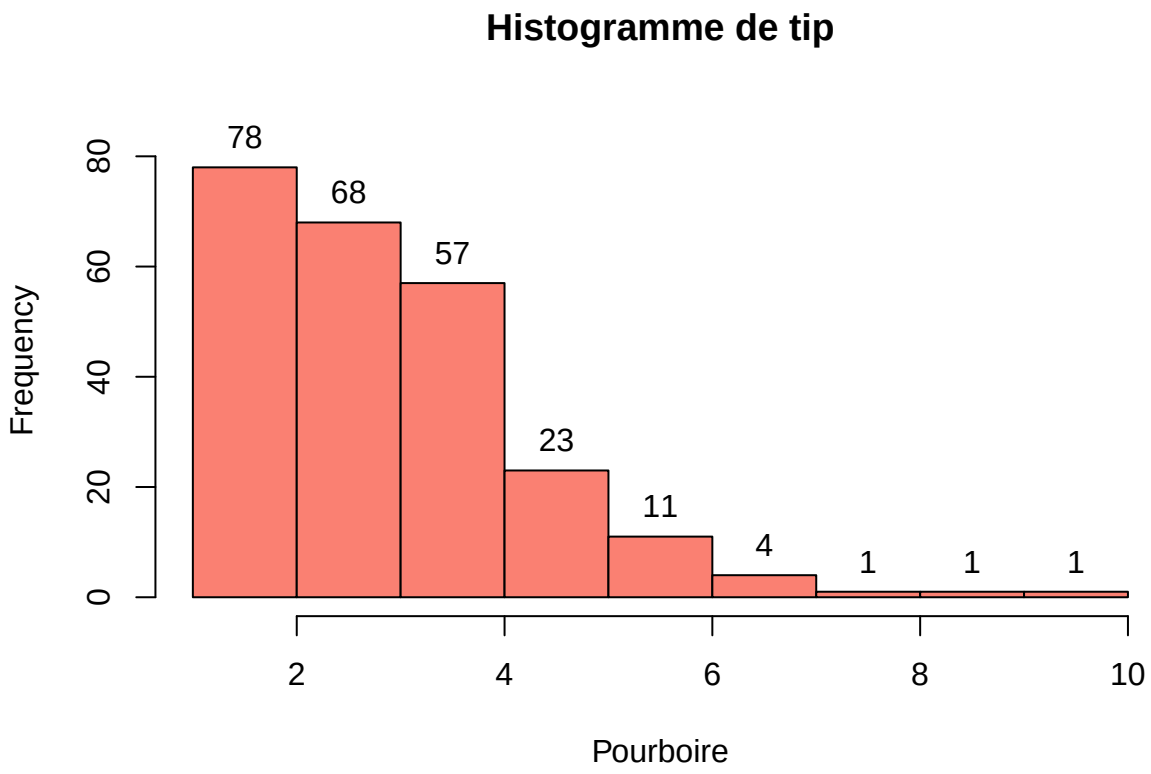
```
#size:
T_size <- table(mydata$size)
prop.table(T_size)
```

```
##
##          1          2          3          4          5          6
## 0.01639344 0.63934426 0.15573770 0.15163934 0.02049180 0.01639344
```

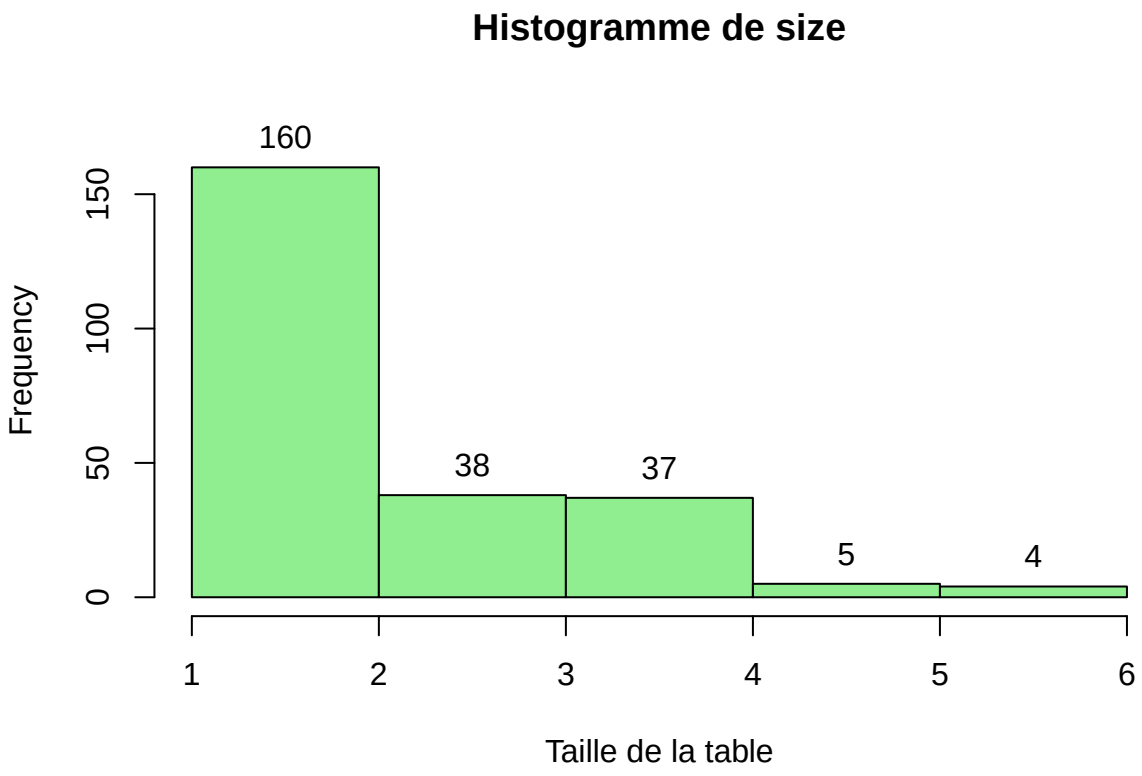
```
# 1. Histogramme - total_bill
h1 <-hist(mydata$total_bill,
     main = "Histogramme de total_bill",
     xlab = "Montant total",
     col = "skyblue",
     breaks = J,
     ylim = c(0, max(hist(mydata$total_bill, breaks = J, plot = FALSE)$counts) * 1.1))
text(h1$mids, h1$counts, labels = h1$counts, pos = 3)
```



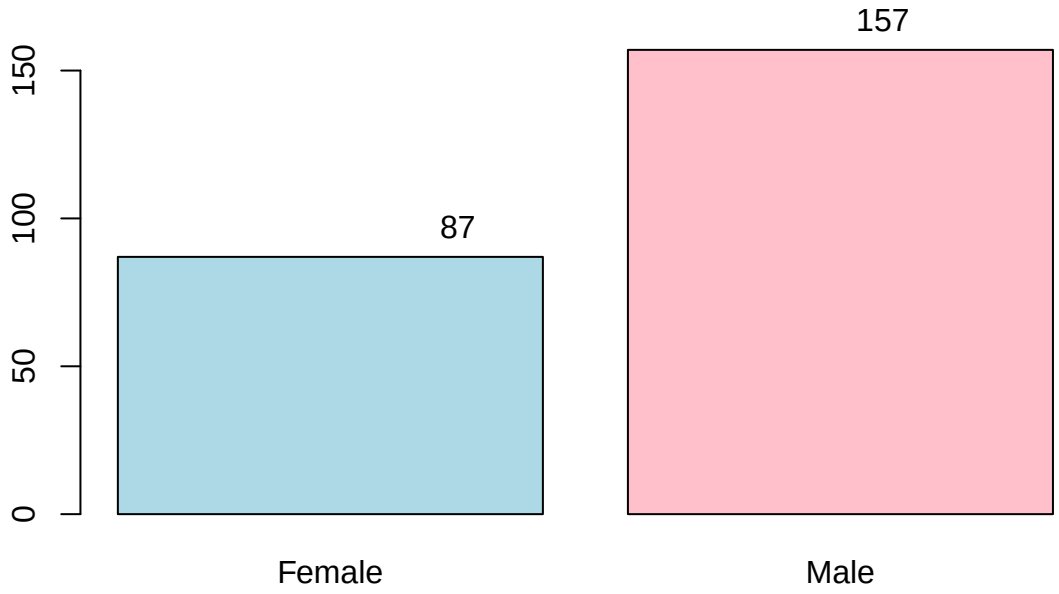

```
# 2. Histogramme - tip
h2 <-hist(mydata$tip,
     main = "Histogramme de tip",
     xlab = "Pourboire",
     col = "salmon",
     breaks = J,
     ylim = c(0, max(hist(mydata$tip, breaks = J, plot = FALSE)$counts) * 1.1))
text(h2$mids, h2$counts, labels = h2$counts, pos = 3)
```



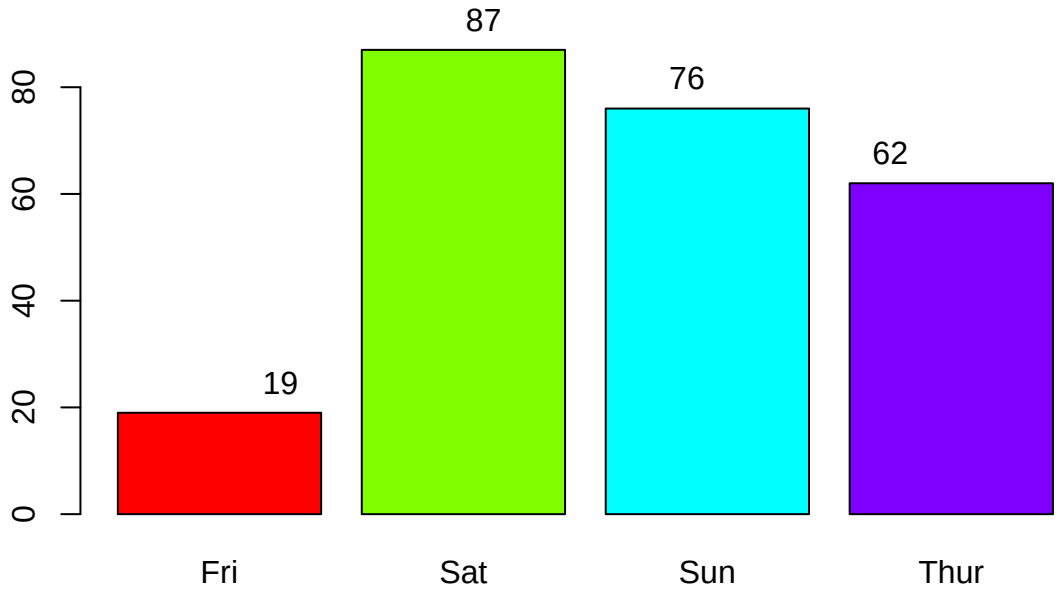

```
# 3. Histogramme - size
nb_size <- length(unique(mydata$size))
h3<-hist(mydata$size,
     main = "Histogramme de size",
     xlab = "Taille de la table",
     col = "lightgreen",
     breaks = nb_size,
     ylim = c(0, max(hist(mydata$size, breaks = nb_size, plot = FALSE)$counts) * 1.1))
text(h3$mids, h3$counts, labels = h3$counts, pos = 3)
```



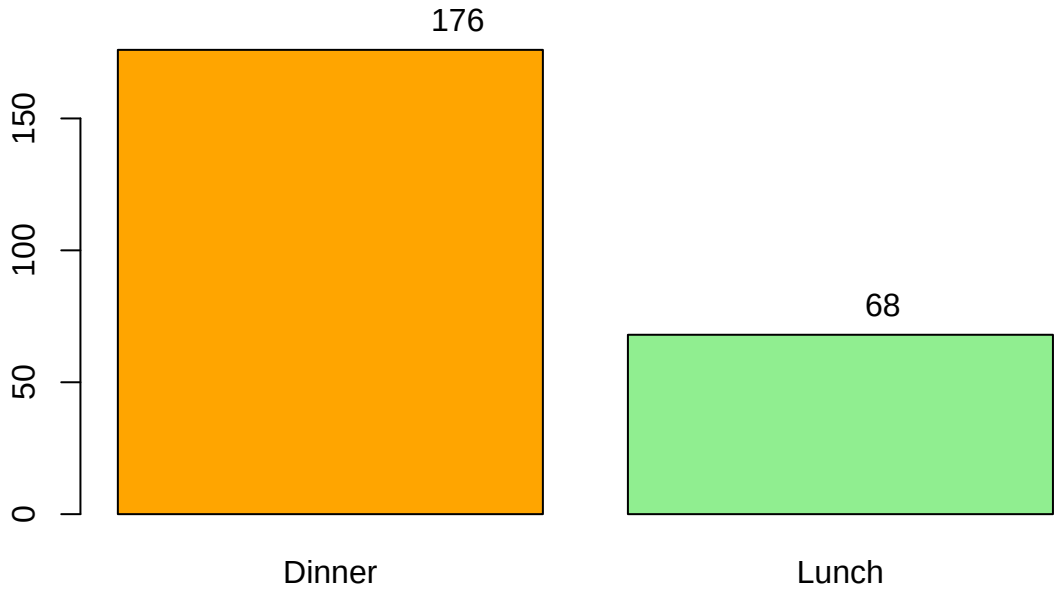

```r
# 4. Diagramme en barres - gender
gender_table <- table(mydata$gender)
barplot(gender_table,
        main = "Genre",
        col = c("lightblue", "pink"),
        ylim = c(0, max(gender_table) * 1.1))
text(x = seq_along(gender_table),
     y = gender_table,
     labels = gender_table,
     pos = 3)
```



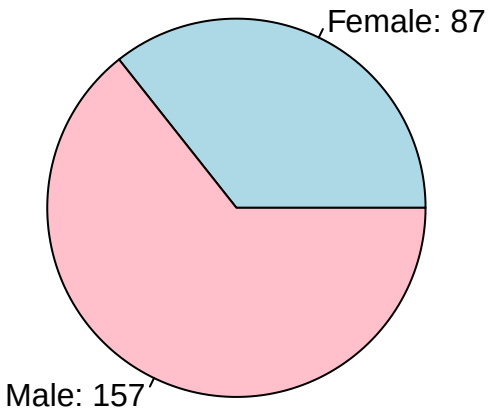

```r
# 5. Diagramme en barres - day
day_table <- table(mydata$day)
barplot(day_table,
        main = "Jour de la semaine",
        col = rainbow(length(day_table)),
        ylim = c(0, max(day_table) * 1.1))
text(x = seq_along(day_table),
     y = day_table,
     labels = day_table,
     pos = 3)
```



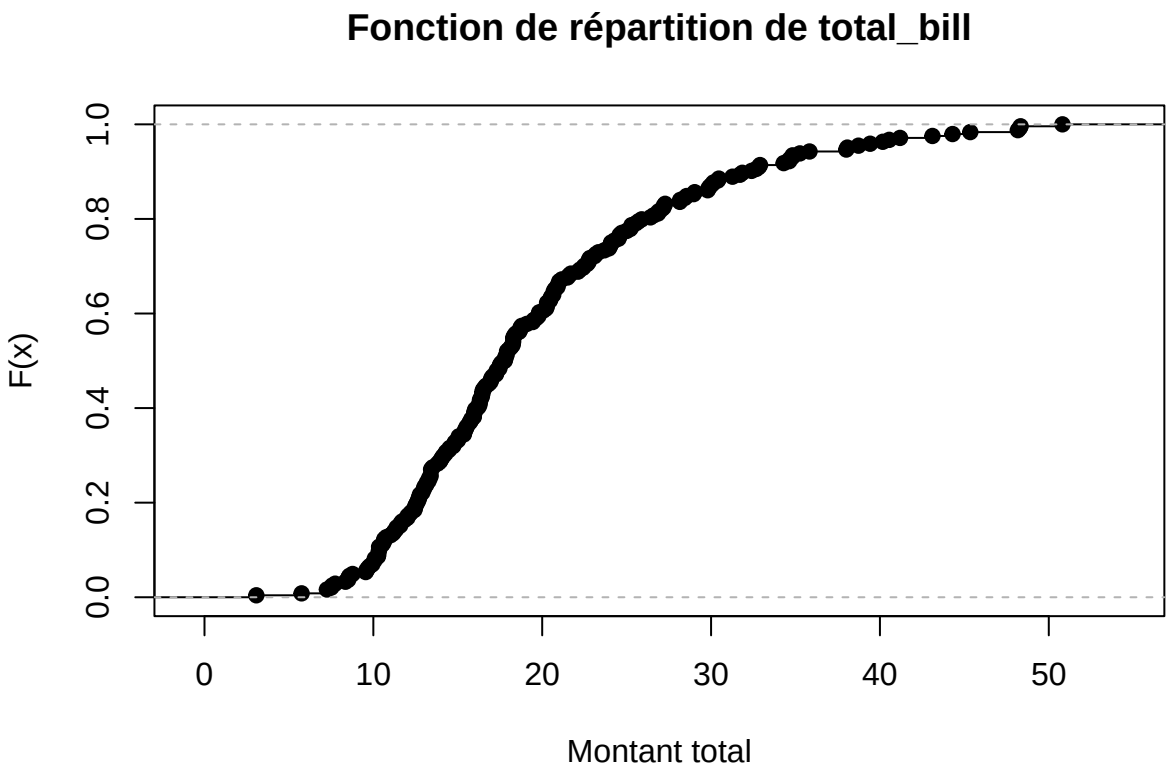

```r
# 6. Diagramme en barres - time
time_table <- table(mydata$time)
barplot(time_table,
        main = "Moment de la journée",
        col = c("orange", "lightgreen"),
        ylim = c(0, max(time_table) * 1.1))
text(x = seq_along(time_table),
     y = time_table,
     labels = time_table,
     pos = 3)
```




```r
# Diagramme en secteurs - gender
# Diagramme en secteurs - gender avec les numéros (fréquences)
gender_table <- table(mydata$gender)
labels <- paste(names(gender_table), ": ", gender_table, sep = "")  # Créer les étiquettes avec les fré
pie(gender_table, main = "Répartition par genre", col = c("lightblue", "pink"), labels = labels)
```




```r
# Fonction de répartition - total_bill
total_bill_cdf <- ecdf(mydata$total_bill)
plot(total_bill_cdf, main = "Fonction de répartition de total_bill", xlab = "Montant total", ylab = "F(
```




```r
# Charger les librairies nécessaires
library(e1071)  # pour skewness et kurtosis
```

```
## Warning: package 'e1071' was built under R version 4.4.3
```

```r
# Fonction pour calculer et afficher les paramètres statistiques pour une variable donnée
calculate_stats <- function(variable) {
  # Paramètres de position
  mean_var <- mean(variable, na.rm = TRUE)          # Moyenne
  median_var <- median(variable, na.rm = TRUE)      # Médiane

  # Paramètres de dispersion
  var_var <- var(variable, na.rm = TRUE)            # Variance
  sd_var <- sd(variable, na.rm = TRUE)              # Écart-type
  range_var <- range(variable, na.rm = TRUE)        # Plage (min, max)
  IQR_var <- IQR(variable, na.rm = TRUE)            # Intervalle interquartile

  # Paramètres de forme
  skewness_var <- skewness(variable, na.rm = TRUE) # Asymétrie
  kurtosis_var <- kurtosis(variable, na.rm = TRUE) # Aplatissement

  # Affichage des résultats
  cat("Paramètres de position:\n")
  cat("Moyenne : ", mean_var, "\n")
  cat("Médiane : ", median_var, "\n")
```

```
  cat("Paramètres de dispersion:\n")
  cat("Variance : ", var_var, "\n")
  cat("Écart-type : ", sd_var, "\n")
  cat("Plage : ", range_var, "\n")
  cat("Intervalle interquartile : ", IQR_var, "\n")
  cat("Paramètres de forme:\n")
  cat("Asymétrie : ", skewness_var, "\n")
  cat("Aplatissement : ", kurtosis_var, "\n")
  cat("\n")
}

# Application aux variables quantitatives
cat("### Paramètres pour total_bill ###\n")
```

```
## ### Paramètres pour total_bill ###
```

```
calculate_stats(mydata$total_bill)
```

```
## Paramètres de position:
## Moyenne :  19.78594
## Médiane :  17.795
## Paramètres de dispersion:
## Variance :  79.25294
## Écart-type :  8.902412
## Plage :  3.07 50.81
## Intervalle interquartile :  10.78
## Paramètres de forme:
## Asymétrie :  1.119318
## Aplatissement :  1.135065
```

```
cat("### Paramètres pour tip ###\n")
```

```
## ### Paramètres pour tip ###
```

```
calculate_stats(mydata$tip)
```

```
## Paramètres de position:
## Moyenne :  2.998279
## Médiane :  2.9
## Paramètres de dispersion:
## Variance :  1.914455
## Écart-type :  1.383638
## Plage :  1 10
## Intervalle interquartile :  1.5625
## Paramètres de forme:
## Asymétrie :  1.447482
## Aplatissement :  3.495977
```

```
cat("### Paramètres pour size ###\n")
```

```
## ### Paramètres pour size ###
```

```r
calculate_stats(mydata$size)
```

```
## Paramètres de position:
## Moyenne :  2.569672
## Médiane :  2
## Paramètres de dispersion:
## Variance :  0.9045908
## Écart-type :  0.9510998
## Plage :  1 6
## Intervalle interquartile :  1
## Paramètres de forme:
## Asymétrie :  1.430128
## Aplatissement :  1.633712
```

```r
# Fonction pour les variables qualitatives (catégorielles)
calculate_cat_stats <- function(variable) {
  # Fréquence des modalités
  freq_var <- table(variable)

  # Affichage des résultats
  cat("Fréquence des modalités : \n")
  print(freq_var)

  # Asymétrie et aplatissement ne sont pas calculés pour les variables catégorielles
  cat("\n")
}

# Application aux variables qualitatives
cat("### Fréquence pour gender ###\n")
```

```
## ### Fréquence pour gender ###
```

```r
calculate_cat_stats(mydata$gender)
```

```
## Fréquence des modalités :
## variable
## Female   Male
##     87    157
```

```r
cat("### Fréquence pour day ###\n")
```

```
## ### Fréquence pour day ###
```

```r
calculate_cat_stats(mydata$day)
```

```
## Fréquence des modalités :
## variable
##  Fri  Sat  Sun Thur
##   19   87   76   62
```

```r
cat("### Fréquence pour time ###\n")
```

```
## ### Fréquence pour time ###
```

```r
calculate_cat_stats(mydata$time)
```

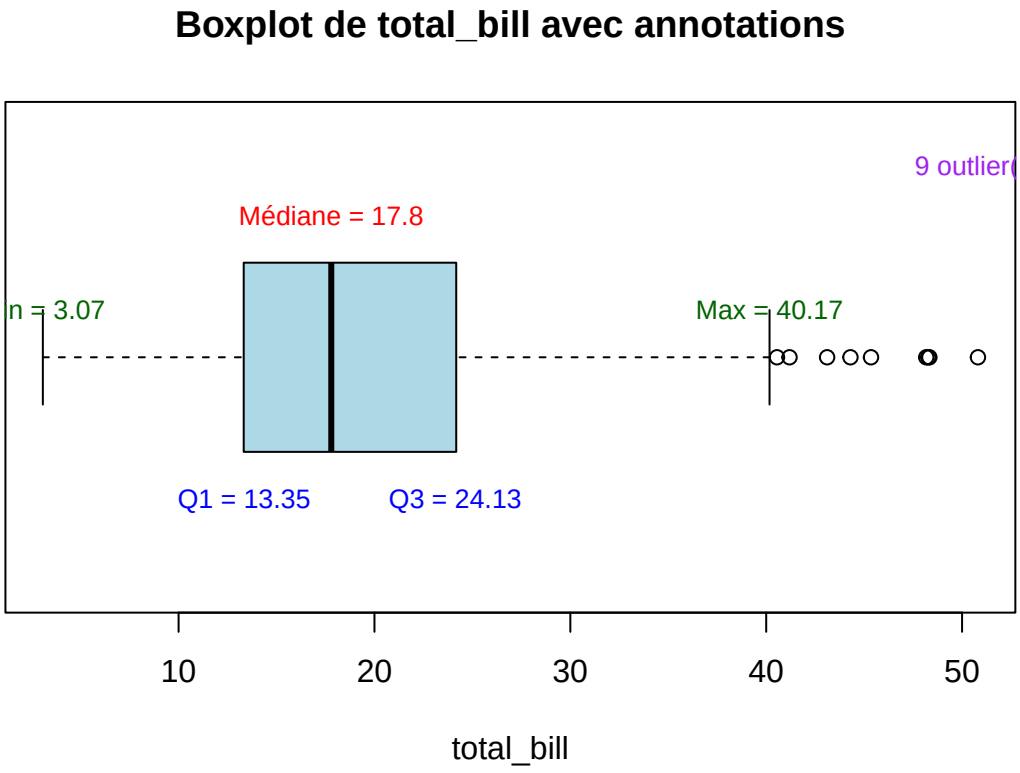
```
## Fréquence des modalités :
## variable
## Dinner  Lunch
##    176     68
```

```r
# 1. Importer les données
mydata <- read.csv("data.csv", header = TRUE)

# 2. Vérifier les colonnes
print(names(mydata))  # Affiche les noms des colonnes
```

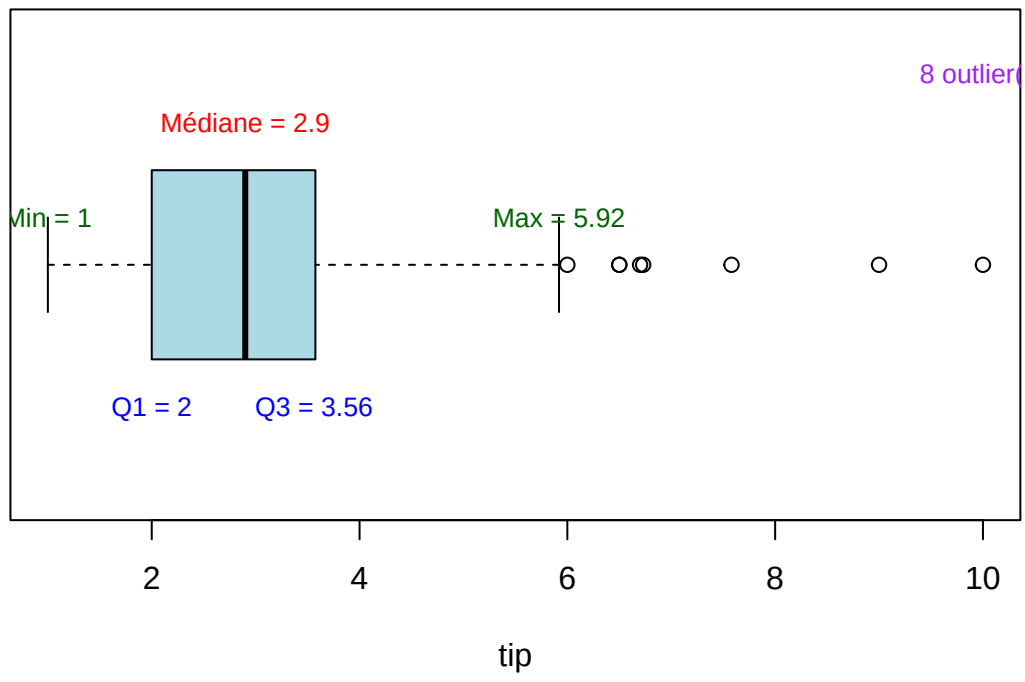
```
## [1] "total_bill" "tip"        "gender"     "smoker"     "day"
## [6] "time"       "size"
```

```r
# Fonction pour ajouter des annotations au boxplot
annotated_boxplot <- function(data, variable, title) {
  # Calcul des statistiques
  stats <- boxplot.stats(data[[variable]])
  q <- quantile(data[[variable]], probs = c(0.25, 0.5, 0.75))

  # Création du boxplot
  boxplot(data[[variable]],
          main = title,
          horizontal = TRUE,
          col = "lightblue",
          xlab = variable)

  # Ajout des annotations
  text(x = q[2], y = 1.3, paste("Médiane =", round(q[2], 2)),
       col = "red", cex = 0.8)
  text(x = q[1], y = 0.7, paste("Q1 =", round(q[1], 2)),
       col = "blue", cex = 0.8)
  text(x = q[3], y = 0.7, paste("Q3 =", round(q[3], 2)),
       col = "blue", cex = 0.8)
  text(x = stats$stats[1], y = 1.1, paste("Min =", round(stats$stats[1], 2)),
       col = "darkgreen", cex = 0.8)
  text(x = stats$stats[5], y = 1.1, paste("Max =", round(stats$stats[5], 2)),
       col = "darkgreen", cex = 0.8)

  # Affichage des outliers si existants
  if(length(stats$out) > 0) {
    text(x = max(stats$out), y = 1.4,
         paste(length(stats$out), "outlier(s)"),
         col = "purple", cex = 0.8)
  }
```

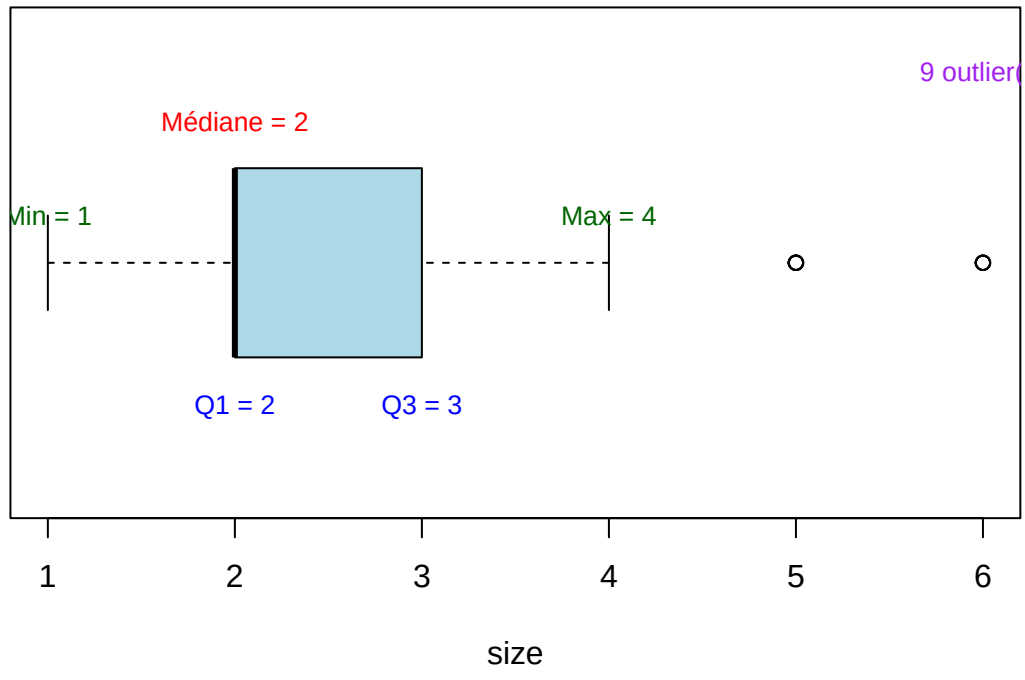
```r
}

# 3. Générer les boxplots annotés
if ("total_bill" %in% names(mydata)) {
  annotated_boxplot(mydata, "total_bill", "Boxplot de total_bill avec annotations")
}
```




```r
if ("tip" %in% names(mydata)) {
  annotated_boxplot(mydata, "tip", "Boxplot de tip avec annotations")
}
```




```
# 4. Boxplot pour 'size' si la colonne existe
if ("size" %in% names(mydata)) {
  annotated_boxplot(mydata, "size", "Boxplot de size avec annotations")
} else {
  print("La colonne 'size' est absente du jeu de données.")
}
```




```
# 5. Sauvegarder les graphiques
png("boxplots_annotés.png", width = 800, height = 600)
par(mfrow = c(1, 3))  # 1 ligne, 3 colonnes pour afficher côte à côte

if ("total_bill" %in% names(mydata)) {
  annotated_boxplot(mydata, "total_bill", "Total Bill")
}

if ("tip" %in% names(mydata)) {
  annotated_boxplot(mydata, "tip", "Tip")
}

if ("size" %in% names(mydata)) {
  annotated_boxplot(mydata, "size", "Size")
} else {
  plot.new()
  text(0.5, 0.5, "Size non disponible", cex = 1.2)
}

dev.off()
```

```
## pdf
##   2
```

```r
# 1. Calcul des variances marginales, des moyennes marginales, des écarts-types marginaux et des quanti

# Variables
total_bill <- mydata$total_bill
tip <- mydata$tip

# Moyennes marginales
mean_total_bill <- mean(total_bill)
mean_tip <- mean(tip)

# Variances marginales
var_total_bill <- var(total_bill)
var_tip <- var(tip)

# Écarts-types marginaux
sd_total_bill <- sd(total_bill)
sd_tip <- sd(tip)

# Quantiles marginaux
quantiles_total_bill <- quantile(total_bill)
quantiles_tip <- quantile(tip)

# Afficher les résultats
cat("Moyenne total_bill:", mean_total_bill, "\n")
```

```
## Moyenne total_bill: 19.78594
```

```r
cat("Moyenne tip:", mean_tip, "\n")
```

```
## Moyenne tip: 2.998279
```

```r
cat("Variance total_bill:", var_total_bill, "\n")
```

```
## Variance total_bill: 79.25294
```

```r
cat("Variance tip:", var_tip, "\n")
```

```
## Variance tip: 1.914455
```

```r
cat("Écart-type total_bill:", sd_total_bill, "\n")
```

```
## Écart-type total_bill: 8.902412
```

```r
cat("Écart-type tip:", sd_tip, "\n")
```

```
## Écart-type tip: 1.383638
```

```r
cat("Quantiles total_bill:", quantiles_total_bill, "\n")
```

```
## Quantiles total_bill: 3.07 13.3475 17.795 24.1275 50.81
```

```r
cat("Quantiles tip:", quantiles_tip, "\n")
```

```
## Quantiles tip: 1 2 2.9 3.5625 10
```

```r
# 2. Calcul de la covariance
covariance <- cov(total_bill, tip)
cat("Covariance entre total_bill et tip:", covariance, "\n")
```

```
## Covariance entre total_bill et tip: 8.323502
```

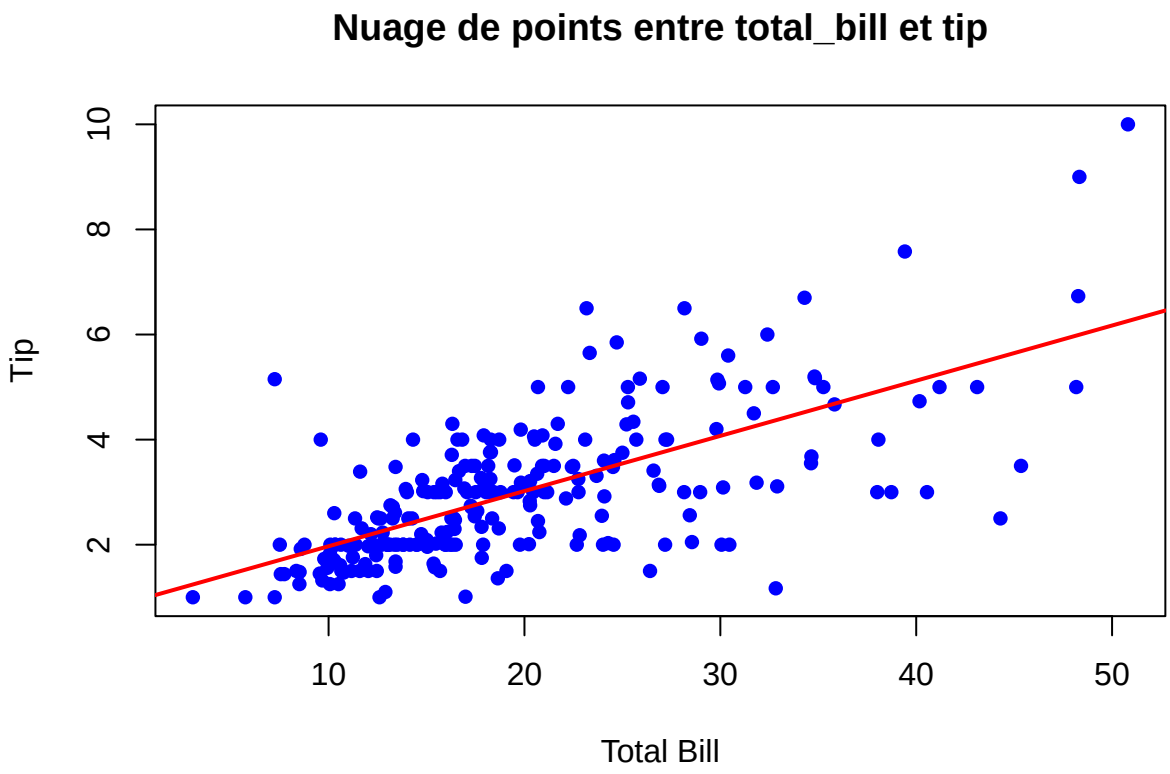
```r
# 3. Calcul du coefficient de corrélation
correlation <- cor(total_bill, tip)
cat("Coefficient de corrélation entre total_bill et tip:", correlation, "\n")
```

```
## Coefficient de corrélation entre total_bill et tip: 0.6757341
```

```r
# 4. Calcul du coefficient de détermination
r_squared <- correlation^2
cat("Coefficient de détermination (R²) :", r_squared, "\n")
```

```
## Coefficient de détermination (R²) : 0.4566166
```

```r
# 5. Représentation graphique du nuage de points
plot(total_bill, tip,
     main = "Nuage de points entre total_bill et tip",
     xlab = "Total Bill",
     ylab = "Tip",
     col = "blue",
     pch = 16)

# 6. Représentation de la droite de régression
model <- lm(tip ~ total_bill)
abline(model, col = "red", lwd = 2)
```




```r
# Résumé du modèle de régression
cat("\nRésumé du modèle de régression:\n")
```

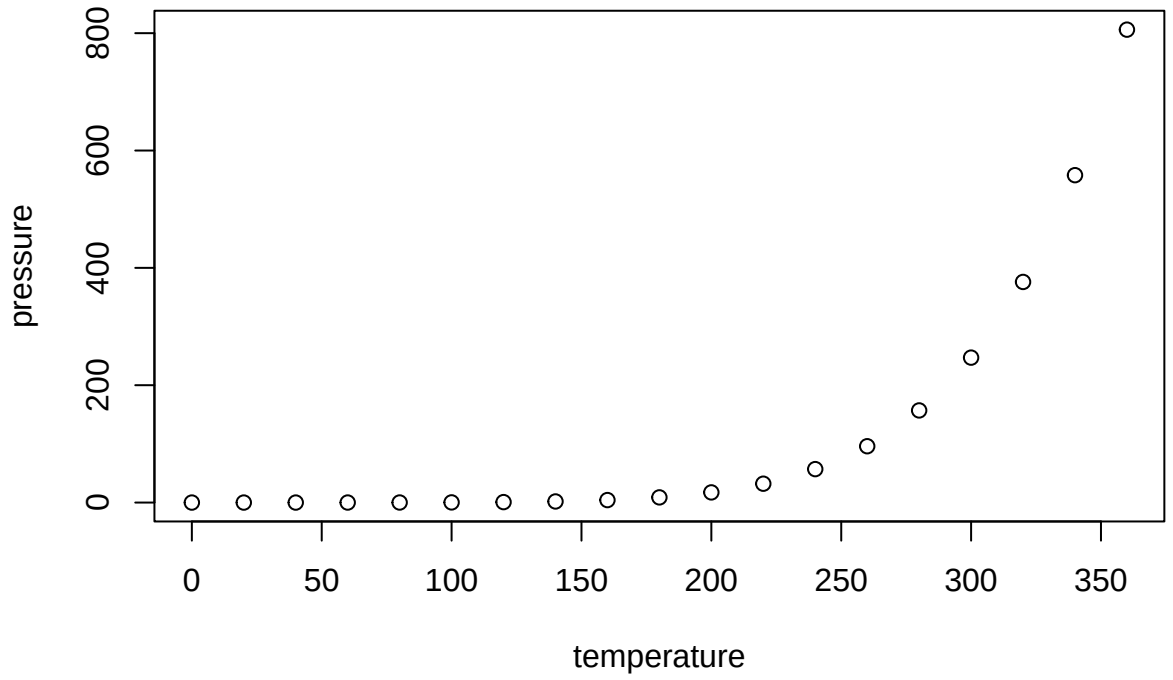
```
##
## Résumé du modèle de régression:
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = tip ~ total_bill)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.1982 -0.5652 -0.0974  0.4863  3.7434
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.920270   0.159735   5.761 2.53e-08 ***
## total_bill  0.105025   0.007365  14.260  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.022 on 242 degrees of freedom
## Multiple R-squared:  0.4566, Adjusted R-squared:  0.4544
## F-statistic: 203.4 on 1 and 242 DF,  p-value: < 2.2e-16
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.