

Projet de statistique

KIEMDE Alain

2023-05-08

1-Analyse Univarier de chaque variable

Importation des donnees

```
data1=read.table("autos2.csv", dec=".", sep=";", header=TRUE)
attach(data1)
data1
```

##	Modele	CYL	PUISS	LONG	LARG	POIDS	V.MAX	FINITION	PRIX	R.POID.PUIS
## 1	Alfasud TI	1350	79	393	161	870	165	B	30570	11.01
## 2	Audi 100	1588	85	468	177	1110	160	TB	39990	13.06
## 3	Simca 1300	1294	68	424	168	1050	152	M	29600	15.44
## 4	Citroen GS Club	1222	59	412	161	930	151	M	28250	15.76
## 5	Fiat 132	1585	98	439	164	1105	165	B	34900	11.28
## 6	Lancia Beta	1297	82	429	169	1080	160	TB	35480	13.17
## 7	Peugeot 504	1796	79	449	169	1160	154	B	32300	14.68
## 8	Renault 16 TL	1565	55	424	163	1010	140	B	32000	18.36
## 9	Renault 30	2664	128	452	173	1320	180	TB	47700	10.31
## 10	Toyota Corolla	1166	55	399	157	815	140	M	26540	14.82
## 11	Alfetta-1.66	1570	109	428	162	1060	175	TB	42395	9.72
## 12	Princess-1800	1798	82	445	172	1160	158	B	33990	14.15
## 13	Datsun-200L	1998	115	469	169	1370	160	TB	43980	11.91
## 14	Taunus-2000	1993	98	438	170	1080	167	B	35010	11.02
## 15	Rancho	1442	80	431	166	1129	144	TB	39450	14.11
## 16	Mazda-9295	1769	83	440	165	1095	165	M	27900	13.19
## 17	Opel-Rekord	1979	100	459	173	1120	173	B	32700	11.20
## 18	Lada-1300	1294	68	404	161	955	140	M	22100	14.04

Imposition de la structure des variables qualitatives nominales

```
data1$FINITION=as.ordered(FINITION)
```

Imposition de la structure des variables quantitatives continues

```
data1$CYL=as.double(CYL)
data1$PUISS=as.double(PUISS)
data1$LONG=as.double(LONG)
data1$LARG=as.double(LARG)
data1$POIDS=as.double(POIDS)
data1$V.MAX=as.double(V.MAX)
data1$PRIX=as.double(PRIX)
data1$R.POID.PUIS=as.double(R.POID.PUIS)
```

Tableaux des frequences et des effectifs pour les variables qualitatives(FINITION)

```
TabEffec_FINITION=table(FINITION)
TabFreq_FINITION=table(FINITION)/length(FINITION)
TabEffec_FINITION
```

```
## FINITION
##  B  M TB
##  7  5  6
```

```
TabFreq_FINITION
```

```
## FINITION
##          B          M          TB
## 0.3888889 0.2777778 0.3333333
```

Tableaux des frequences et des effectifs pour les variables quantitatives

```
TabEffec_CYL=table(CYL)
TabFreq_CYL=table(CYL)/length(CYL)
TabEffec_PUISS=table(PUISS)
TabFreq_PUISS=table(PUISS)/length(PUISS)
TabEffec_LONG=table(LONG)
TabFreq_LONG=table(LONG)/length(LONG)
TabEffec_LARG=table(LARG)
TabFreq_LARG=table(LARG)/length(LARG)
TabEffec_POIDS =table(POIDS)
TabFreq_POIDS =table(POIDS)/length(POIDS)
TabEffec_V.MAX=table(V.MAX)
TabFreq_V.MAX=table(V.MAX)/length(V.MAX)
TabEffec_PRIX=table(PRIX)
TabFreq_PRIX=table(PRIX)/length(PRIX)
TabEffec_R.POID.PUIS=table(R.POID.PUIS)
TabFreq_R.POID.PUIS=table(R.POID.PUIS)/length(R.POID.PUIS)
TabEffec_CYL
```

```
## CYL
## 1166 1222 1294 1297 1350 1442 1565 1570 1585 1588 1769 1796 1798 1979 1993 1998
##    1    1    2    1    1    1    1    1    1    1    1    1    1    1    1    1
## 2664
##    1
```

```
TabFreq_CYL
```

```
## CYL
##      1166      1222      1294      1297      1350      1442      1565
## 0.05555556 0.05555556 0.11111111 0.05555556 0.05555556 0.05555556 0.05555556
##      1570      1585      1588      1769      1796      1798      1979
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      1993      1998      2664
## 0.05555556 0.05555556 0.05555556
```

```
TabEffec_PUISS
```

```
## PUISS
## 55 59 68 79 80 82 83 85 98 100 109 115 128
##  2  1  2  2  1  2  1  1  2  1  1  1  1
```

TabFreq_PUISS

```
## PUISS
##      55      59      68      79      80      82      83
## 0.11111111 0.05555556 0.11111111 0.11111111 0.05555556 0.11111111 0.05555556
##      85      98     100     109     115     128
## 0.05555556 0.11111111 0.05555556 0.05555556 0.05555556 0.05555556
```

TabEffec_LONG

```
## LONG
## 393 399 404 412 424 428 429 431 438 439 440 445 449 452 459 468 469
##   1   1   1   1   2   1   1   1   1   1   1   1   1   1   1   1   1
```

TabFreq_LONG

```
## LONG
##      393      399      404      412      424      428      429
## 0.05555556 0.05555556 0.05555556 0.05555556 0.11111111 0.05555556 0.05555556
##      431      438      439      440      445      449      452
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      459      468      469
## 0.05555556 0.05555556 0.05555556
```

TabEffec_LARG

```
## LARG
## 157 161 162 163 164 165 166 168 169 170 172 173 177
##   1   3   1   1   1   1   1   1   3   1   1   2   1
```

TabFreq_LARG

```
## LARG
##      157      161      162      163      164      165      166
## 0.05555556 0.16666667 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      168      169      170      172      173      177
## 0.05555556 0.16666667 0.05555556 0.05555556 0.11111111 0.05555556
```

TabEffec_POIDS

```
## POIDS
## 815 870 930 955 1010 1050 1060 1080 1095 1105 1110 1120 1129 1160 1320 1370
##   1   1   1   1   1   1   1   1   2   1   1   1   1   1   2   1   1
```

TabFreq_POIDS

```
## POIDS
##      815      870      930      955      1010      1050      1060
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      1080      1095      1105      1110      1120      1129      1160
## 0.11111111 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.11111111
##      1320      1370
## 0.05555556 0.05555556
```

TabEffec_V.MAX

```
## V.MAX
## 140 144 151 152 154 158 160 165 167 173 175 180
##   3   1   1   1   1   1   3   3   1   1   1   1
```

```
TabFreq_V.MAX
```

```
## V.MAX
##      140      144      151      152      154      158      160
## 0.16666667 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.16666667
##      165      167      173      175      180
## 0.16666667 0.05555556 0.05555556 0.05555556 0.05555556
```

```
TabEffec_PRIX
```

```
## PRIX
## 22100 26540 27900 28250 29600 30570 32000 32300 32700 33990 34900 35010 35480
##      1      1      1      1      1      1      1      1      1      1      1      1      1
## 39450 39990 42395 43980 47700
##      1      1      1      1      1
```

```
TabFreq_PRIX
```

```
## PRIX
##      22100      26540      27900      28250      29600      30570      32000
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      32300      32700      33990      34900      35010      35480      39450
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      39990      42395      43980      47700
## 0.05555556 0.05555556 0.05555556 0.05555556
```

```
TabEffec_R.POID.PUIS
```

```
## R.POID.PUIS
##  9.72 10.31 11.01 11.02  11.2 11.28 11.91 13.06 13.17 13.19 14.04 14.11 14.15
##      1      1      1      1      1      1      1      1      1      1      1      1      1
## 14.68 14.82 15.44 15.76 18.36
##      1      1      1      1      1
```

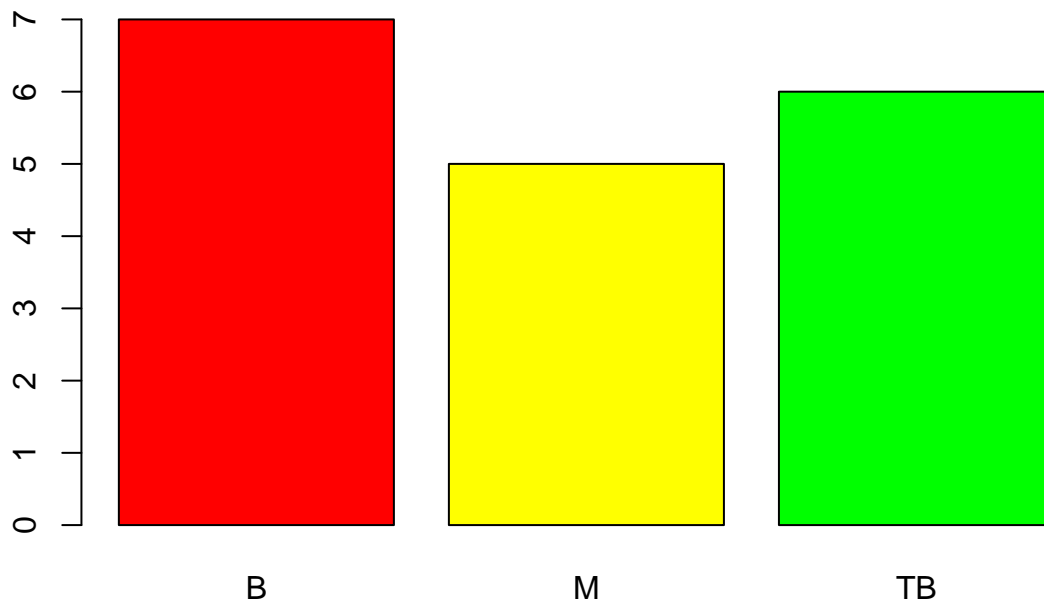
```
TabFreq_R.POID.PUIS
```

```
## R.POID.PUIS
##      9.72      10.31      11.01      11.02      11.2      11.28      11.91
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      13.06      13.17      13.19      14.04      14.11      14.15      14.68
## 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556 0.05555556
##      14.82      15.44      15.76      18.36
## 0.05555556 0.05555556 0.05555556 0.05555556
```

Representation graphiques des variables qualitatives

```
col=c("red", "yellow","green")
barplot(table(FINITION),main="Repartition des Modeles de voitures en fonction de la finition",col=col)
```

Repartition des Modeles de voitures en fonction de la finition

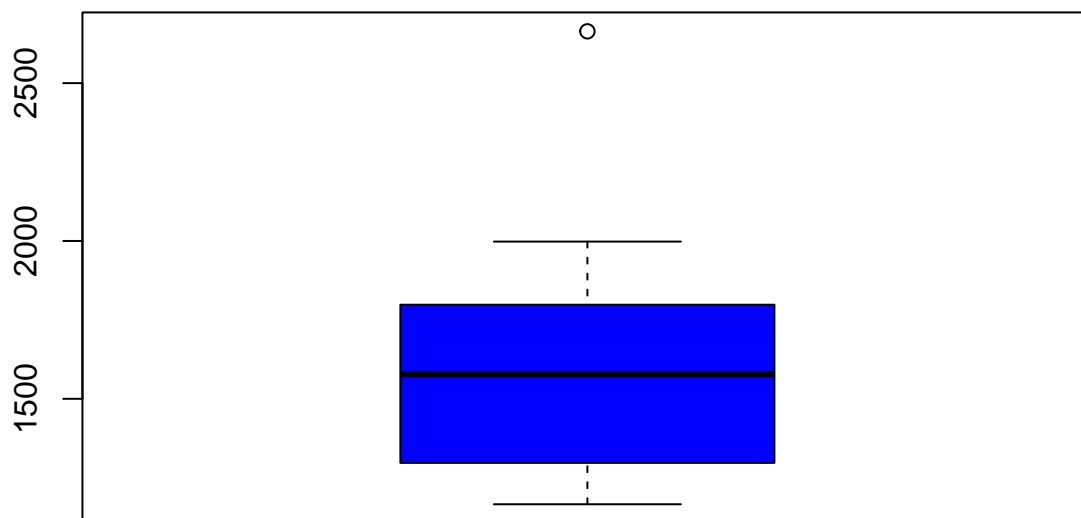


Le graphique montre que les niveaux de finition “B”, “TB” et “M” sont représentés dans l’ensemble de données, mais que le niveau de finition “B” est le plus courant. Cela indique que la plupart des voitures de l’ensemble de données ont un niveau de finition élevé. Cependant, il y a encore des voitures qui ont un niveau de finition inférieur, représentées par les niveaux “TB” et “M”. Il est intéressant de noter que la différence entre les niveaux de finition “B” et “TB” n’est pas très importante, ce qui suggère que les voitures avec ces deux niveaux de finition ont une qualité de finition assez similaire. Cependant, le niveau de finition “M” est le moins fréquent, indiquant que les voitures avec ce niveau de finition sont moins courantes dans l’ensemble de données. En somme, le graphique montre une répartition des niveaux de finition des voitures dans l’ensemble de données et permet d’avoir une idée de la qualité de finition générale des voitures.

Représentation graphique des variables quantitatives continues

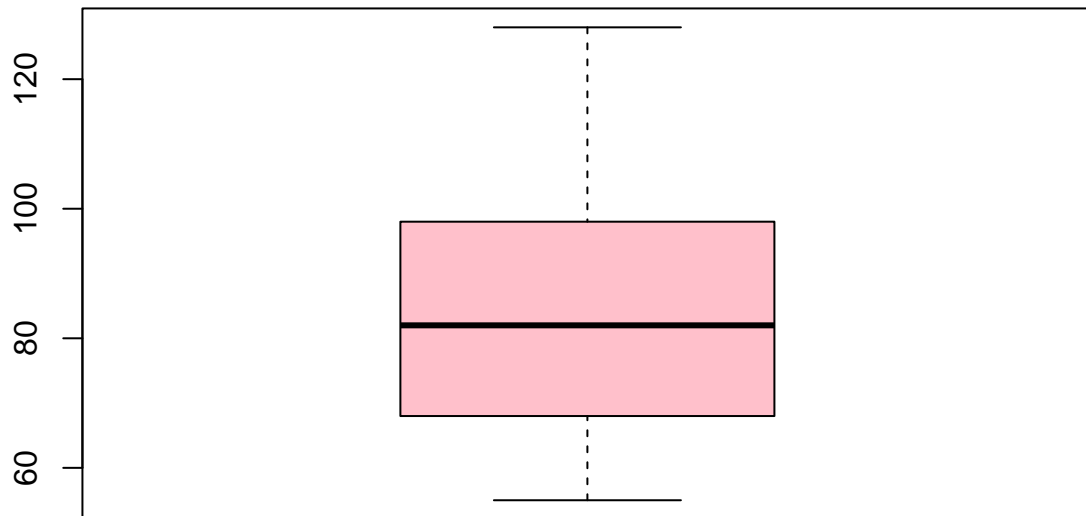
```
boxplot(CYL,main="Boîte à moustaches CYL",col="blue")
```

Boîte à moustaches CYL



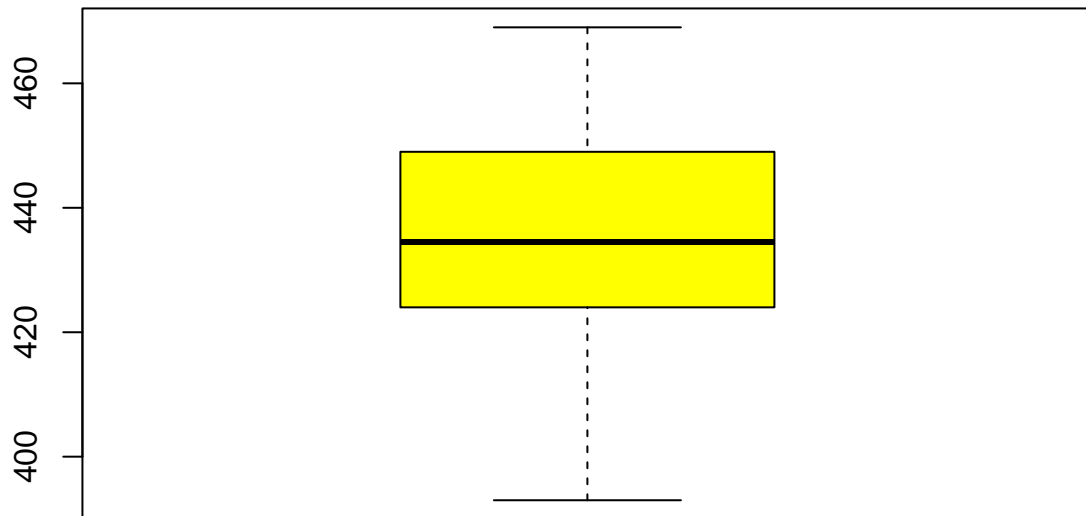
```
boxplot(PUISS,main=" Boîte à moustaches de PUISS",col="pink")
```

Boîte à moustaches de PUISS



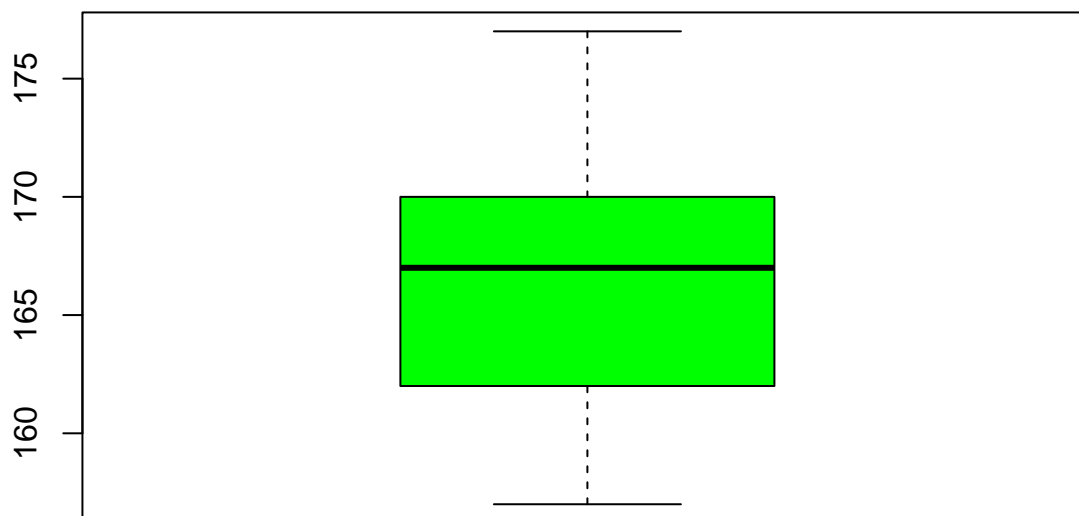
```
boxplot(LONG,main="Boîte à moustaches de LONG",col="yellow")
```

Boîte à moustaches de LONG



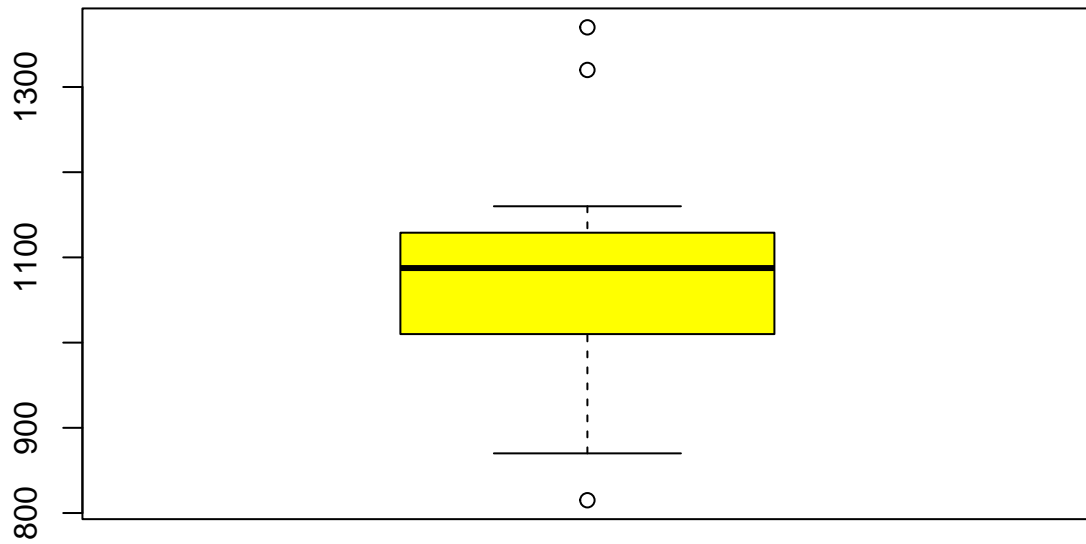
```
boxplot(LARG,main="Boîte à moustaches de LARG",col="green")
```


Boîte à moustaches de LARG



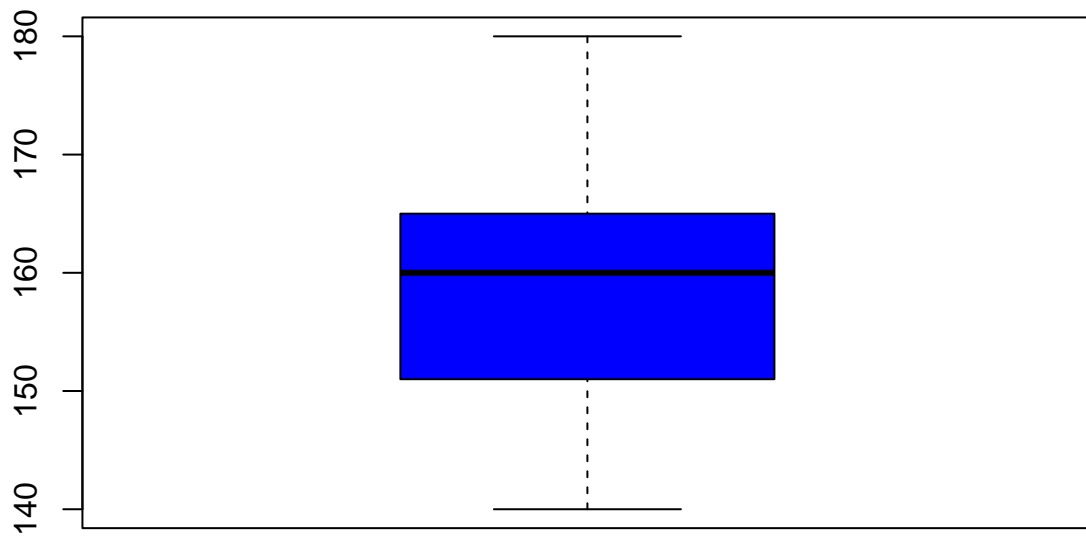
```
boxplot(POIDS,main="Boîte à moustaches de POIDS",col="yellow")
```

Boîte à moustaches de POIDS



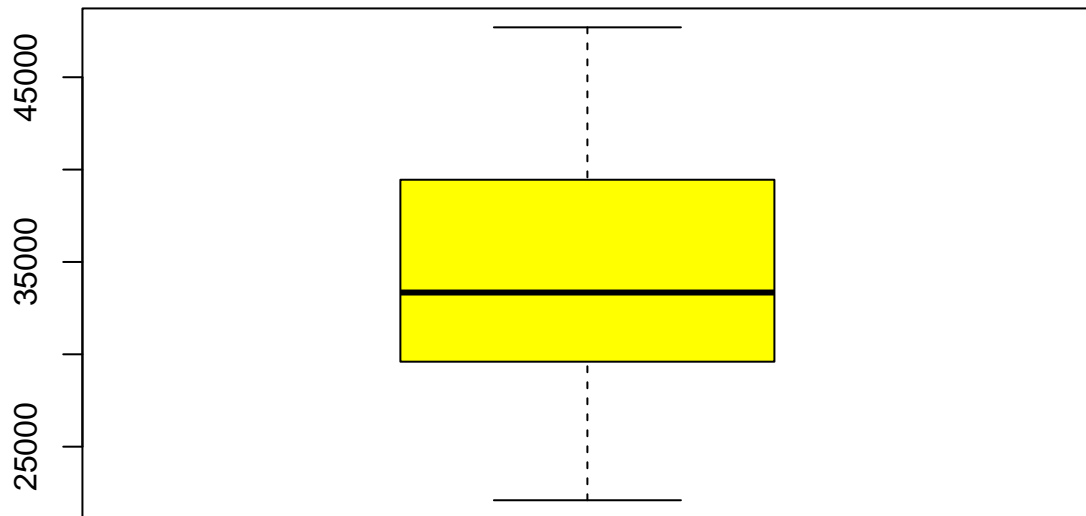
```
boxplot(V.MAX,main="Boîte à moustaches de V.MAX",col="blue")
```

Boîte à moustaches de V.MAX



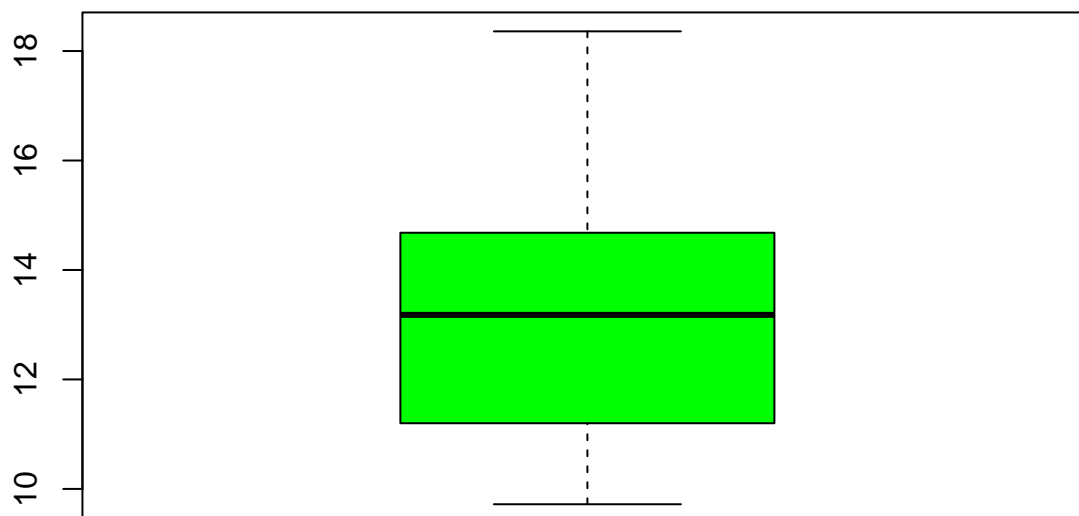
```
boxplot(PRIX,main="Boite à moustache de PRIX",col="yellow")
```

Boite à moustache de PRIX



```
boxplot(R.POID.PUIS,main="Boite à moustache de R.POID.PUIS",col="green")
```

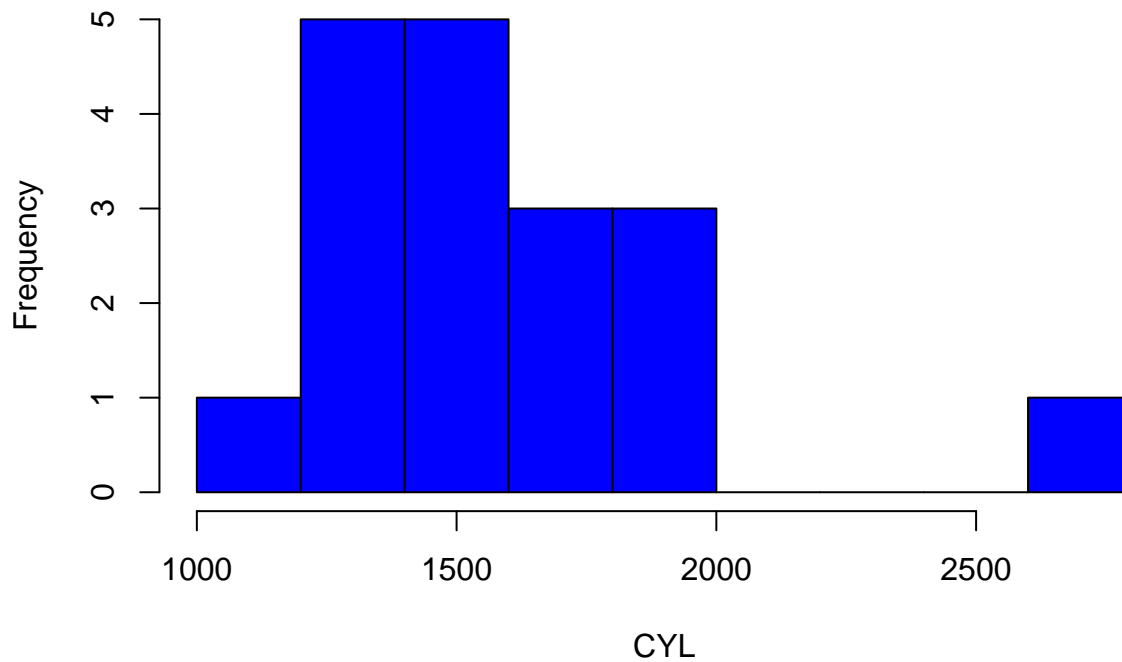
Boite à moustache de R.POID.PUIS



NB: Les variables CYL ET POIDS présentent uniquement des valeurs extrêmes aux regards des boîtes à moustaches générées en dessus à cause de la présence des petits cercles ronds.

```
hist(CYL,main="histogramme de la variable CYL",col="blue")
```

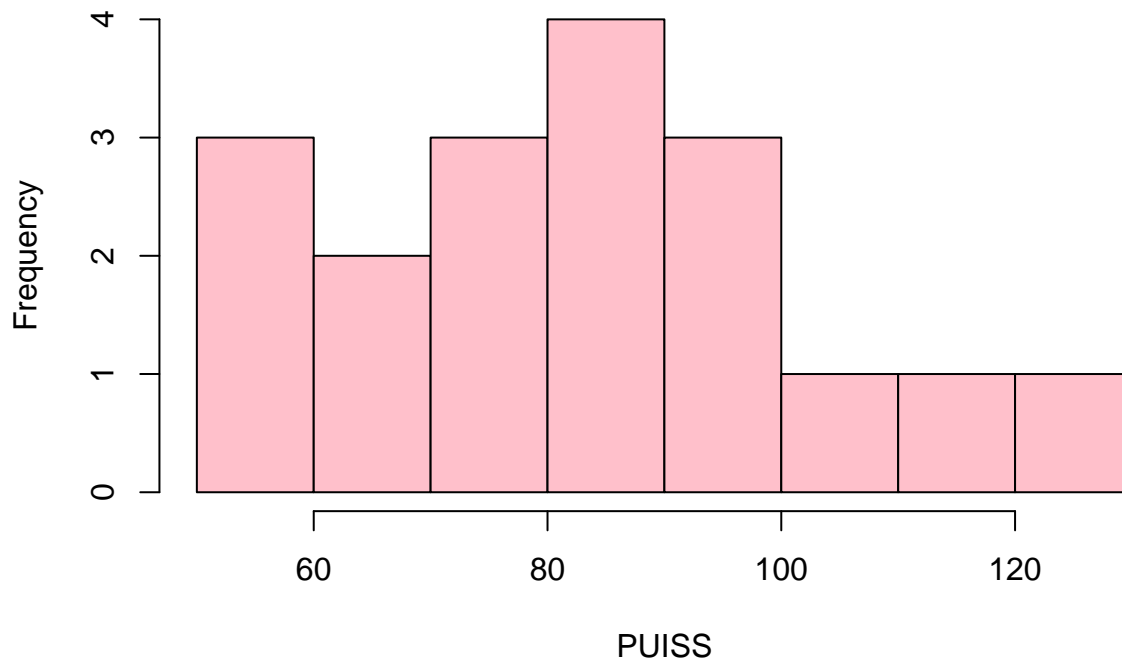
histogramme de la variable CYL



En interprétant l'histogramme, on peut observer que la plupart des valeurs de CYL se situent entre environ 1200 et 2000 cm³. Il y a peu de valeurs inférieures à 1000 cm³ et peu de valeurs supérieures à 3000 cm³. Cela suggère que les voitures dans cet échantillon ont tendance à avoir une cylindrée de moteur relativement similaire, avec quelques exceptions à la marge.

```
hist(PUISS,main=" histogramme de la variable PUISS",col="pink")
```

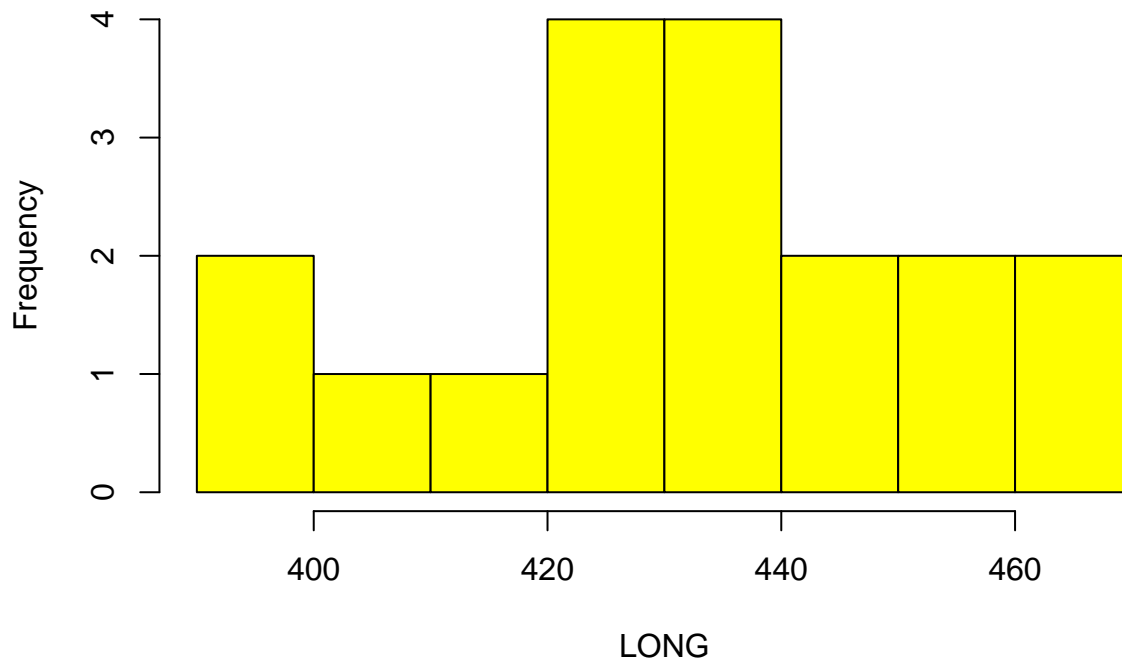
histogramme de la variable PUISS



L'histogramme montre combien de voitures ont une certaine puissance. Plus la barre est haute, plus il y a de voitures avec cette puissance. On peut voir que la plupart des voitures ont une puissance d'environ 80 à 100 chevaux, mais il y a aussi quelques voitures avec une puissance très élevée. Le graphique est symétrique, ce qui signifie que les puissances sont réparties à peu près également des deux côtés de la moyenne. En gros, cet histogramme nous donne une idée de la puissance moyenne des voitures dans l'échantillon.

```
hist(LONG,main="histogramme de la variable LONG",col="yellow")
```

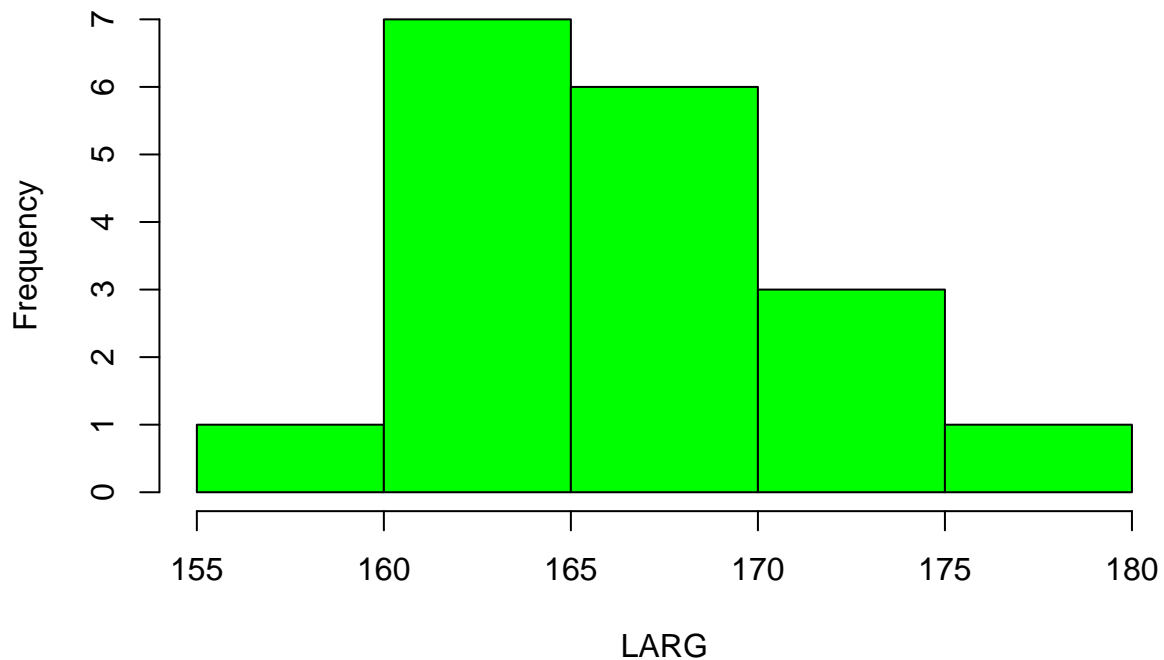
histogramme de la variable LONG



L'histogramme montre la distribution de la longueur des voitures. On peut voir que la majorité des voitures ont une longueur d'environ 420 à 460 cm. Il y a quelques voitures plus courtes, mais également quelques voitures plus longues. Le graphique est asymétrique, ce qui suggère qu'il y a plus de voitures qui ont une longueur dans la plage moyenne. En général, cet histogramme nous donne une idée de la longueur moyenne des voitures dans l'échantillon.

```
hist(LARG,main="histogramme de la variable LARG",col="green")
```

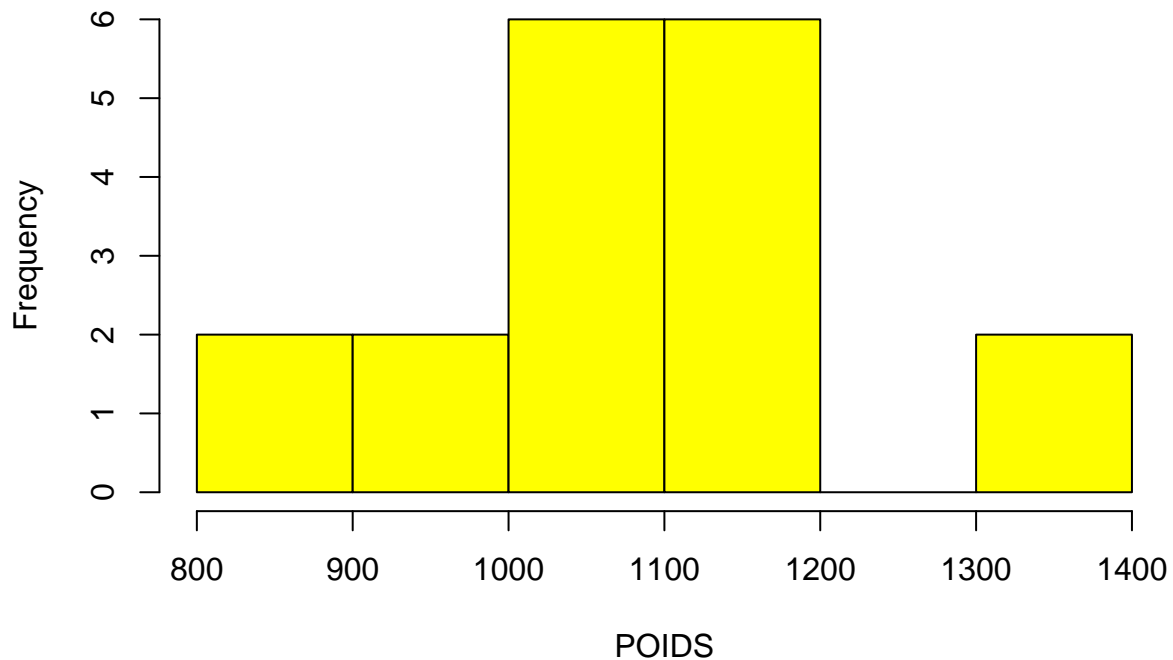

histogramme de la variable LARG



En analysant l'histogramme, on peut observer que la variable LARG a une distribution approximativement normale, avec une légère asymétrie positive. La majorité des voitures ont une largeur située entre 160 et 170 cm. On peut également observer que la distribution est relativement étroite, avec peu de voitures ayant une largeur supérieure à 175 cm ou inférieure à 155 cm..

```
hist(POIDS,main="histogramme de la variable POIDS",col="yellow")
```

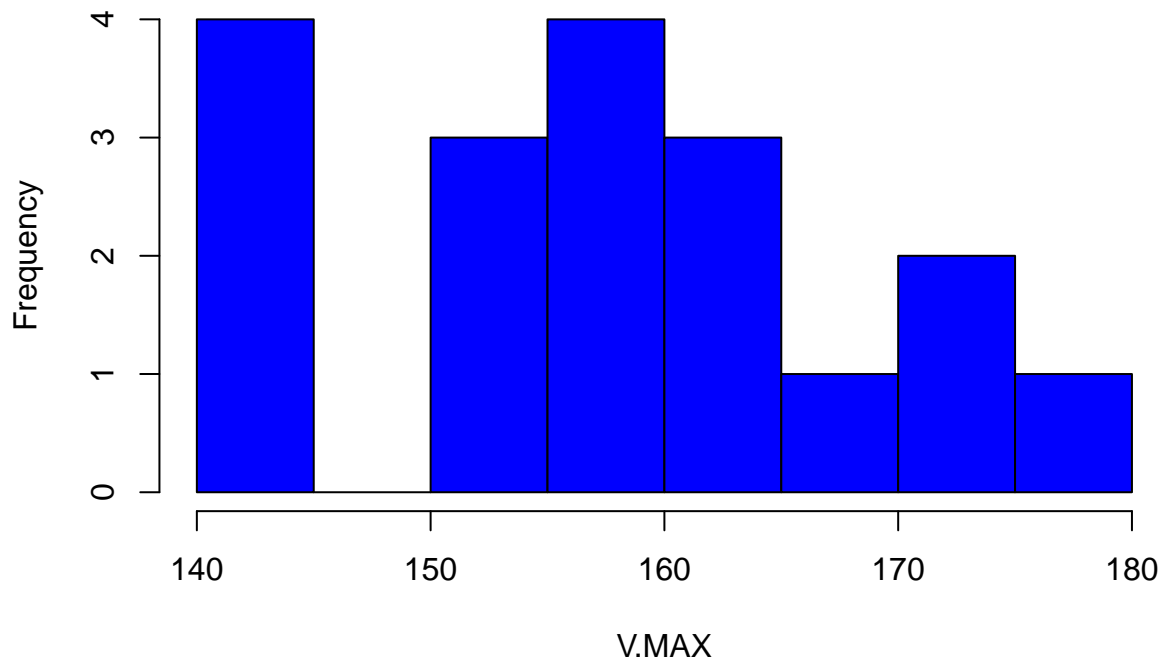
histogramme de la variable POIDS



l'histogramme en dessus presente une concentration plus importante de véhicules dans la plage de poids de 1000 kg à 1200 kg. Il y a également une légère queue vers la droite, indiquant la présence de quelques véhicules plus lourds. La plupart des véhicules semblent peser entre 900 kg et 1400 kg, avec quelques valeurs extrêmes s'étendant jusqu'à environ 1400 kg. Cette variable pourrait être importante pour comprendre les performances de conduite et la consommation de carburant des différents modèles de voitures, en particulier lorsque combinée avec d'autres variables telles que la puissance et la vitesse maximale.

```
hist(V.MAX,main="histogramme de la variable V.MAX",col="blue")
```

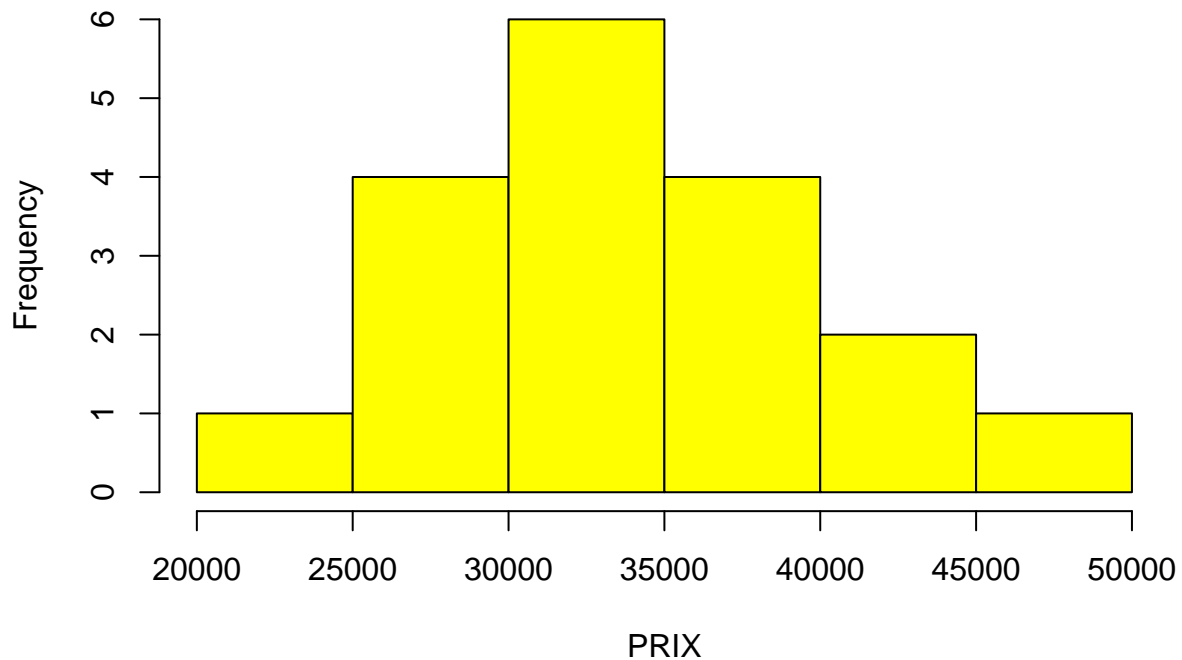
histogramme de la variable V.MAX



D'après l'histogramme de V.MAX montre la distribution de la vitesse maximale des voitures dans l'échantillon. On peut voir que la majorité des voitures ont une vitesse maximale comprise entre 150 et 170 km/h, avec un pic autour de 165 km/h. Cependant, il y a quelques voitures qui ont une vitesse maximale supérieure à 170 km/h. Cela peut indiquer la présence de voitures plus sportives ou de voitures de course dans l'échantillon.

```
hist(PRIX,main="histogramme de la variable PRIX",col="yellow")
```

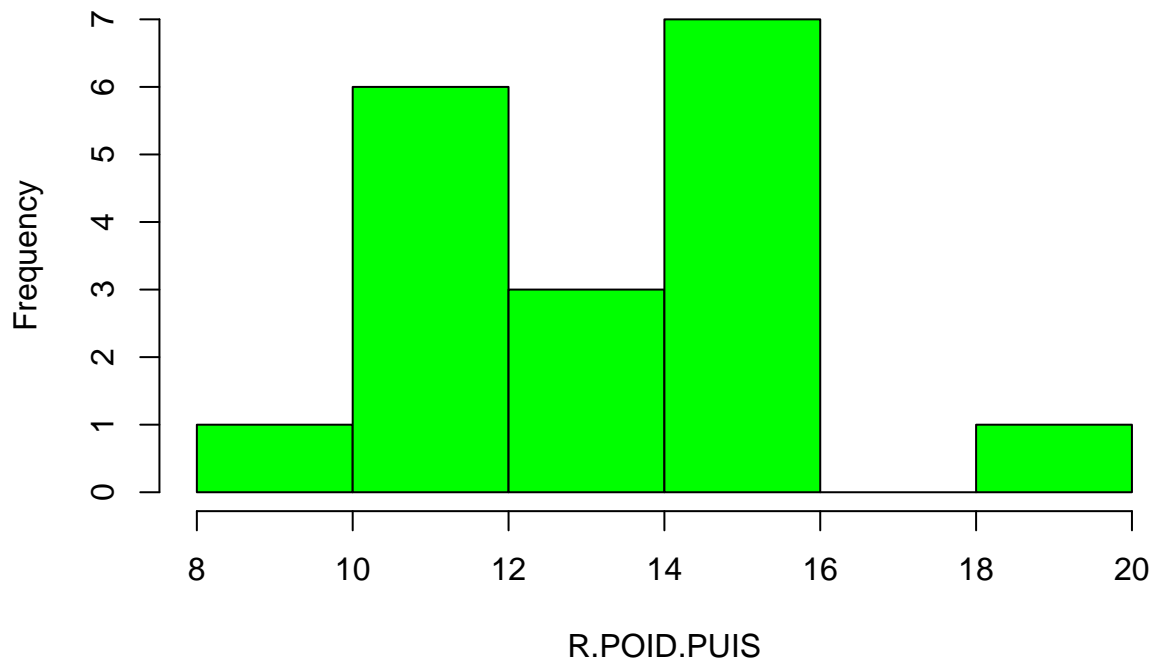
histogramme de la variable PRIX



L'histogramme de la variable PRIX montre que la majorité des voitures coûtent entre 26 000 et 40 000 euros, avec un pic autour de 34 000 euros. Il y a cependant quelques voitures plus chères qui vont jusqu'à près de 48 000 euros. On peut également remarquer que la distribution est légèrement asymétrique à droite, ce qui signifie qu'il y a quelques voitures très chères qui tirent la moyenne vers le haut.

```
hist(R.POID.PUIS,main="histogramme de la variable R.POID.PUIS",col="green")
```

histogramme de la variable R.POID.PUIS



L'histogramme de la variable R.POID.PUIS montre que la majorité des voitures dans l'échantillon ont une valeur de R.POID.PUIS comprise entre 10 et 20, avec un pic autour de 15. Cela suggère que ces voitures ont un poids relativement faible par rapport à leur puissance, ce qui est généralement considéré comme un avantage en termes de performances et d'efficacité. On observe également une petite proportion de voitures ayant une valeur de R.POID.PUIS très élevée, ce qui pourrait indiquer des véhicules plus lourds ou moins puissants par rapport à leur poids.

Les indicateurs de position uniquement pour les variables quantitatives: Mode, Moyenne, médiane, quartiles.

La variable CYL

```
summary(CYL)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1166   1310   1578   1632   1798   2664
```

```
Mode_CYL=names(which.max(table(CYL)))
```

```
Mode_CYL
```

```
## [1] "1294"
```

Interprétation

La valeur minimale (Min) est de **1166**, ce qui signifie que la voiture avec le plus petit nombre de cylindres dans ce jeu de données a une cylindrée de **1166 cm3**. Le premier quartile (1st Qu.) est de **1310**, ce qui signifie que 25% des voitures ont une cylindrée inférieure ou égale à 1310 cm3. La médiane (Median) est de 1578, ce qui signifie que 50% des voitures ont une cylindrée inférieure ou égale à 1578 cm3. La moyenne (Mean) est de 1632, ce qui représente la cylindrée moyenne de toutes les voitures dans ce jeu de données. Le

troisième quartile (3rd Qu.) est de 1798, ce qui signifie que 75% des voitures ont une cylindrée inférieure ou égale à 1798 cm³. La valeur maximale (Max) est de 2664, ce qui signifie que la voiture avec le plus grand nombre de cylindres dans ce jeu de données a une cylindrée de 2664 cm³. Le mode de la variable CYL est 1294, ce qui signifie que la majorité des voitures dans ce jeu de données ont une cylindrée de 1294 cm³. Cela peut être utile pour des analyses comparatives ou pour des choix de modèles de voitures en fonction de la cylindrée souhaitée.

La variable PUISS

```
summary(PUISS)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   55.00  70.75   82.00   84.61   98.00  128.00
```

```
Mode_PUISS=names(which.max(table(PUISS)))
```

```
Mode_PUISS
```

```
## [1] "55"
```

Interprétation

En analysant les résultats, nous pouvons constater que la puissance minimale est de 55 chevaux, tandis que la puissance maximale est de 128 chevaux. En ce qui concerne la répartition des valeurs, nous observons que 25% des voitures ont une puissance inférieure ou égale à 70.75 chevaux, ce qui correspond au premier quartile. De même, le troisième quartile (75%) des voitures ont une puissance inférieure ou égale à 98 chevaux. La moitié des voitures ont une puissance inférieure ou égale à la médiane de 82 chevaux, ce qui indique que la répartition des valeurs n'est pas symétrique. La puissance moyenne des voitures est de 84.61 chevaux. En somme, ces résultats nous permettent de mieux comprendre la distribution des puissances des voitures dans notre jeu de données. Le mode est de 55 chevaux, ce qui signifie que c'est la valeur la plus fréquente pour la puissance des voitures.

La variable LONG

```
summary(LONG)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   393.0  424.0   434.5   433.5   448.0   469.0
```

```
Mode_LONG=names(which.max(table(LONG)))
```

```
Mode_LONG
```

```
## [1] "424"
```

Interpretations

La longueur minimale d'une voiture dans ce jeu de données est de 393 cm, tandis que la longueur maximale est de 469 cm. On peut voir que la majorité des voitures ont une longueur comprise entre le premier quartile (Q1) de 424 cm et le troisième quartile (Q3) de 448 cm. La médiane (valeur au centre) des longueurs des voitures est de 434.5 cm, tandis que la moyenne est de 433.5 cm. Le mode, c'est-à-dire la valeur la plus fréquente de la variable LONG, est de 424 cm. Cela signifie qu'il y a plus de voitures dans le jeu de données ayant une longueur de 424 cm que toute autre longueur.

La variable LARG

```
summary(LARG)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   157.0  162.2   167.0   166.7   169.8   177.0
```

```
Mode_LARG=names(which.max(table(LARG)))
```

```
Mode_LARG
```

```
## [1] "161"
```

Interpretations

La largeur minimale des voitures dans ce jeu de données est de 157 cm. On peut constater que la plupart des voitures ont une largeur comprise entre 162,2 cm et 169,8 cm, car cela correspond à l'intervalle entre le premier quartile (25%) et le troisième quartile (75%). La valeur médiane de la largeur est de 167 cm, ce qui signifie que la moitié des voitures ont une largeur supérieure à cette valeur et l'autre moitié ont une largeur inférieure à cette valeur. La moyenne de la largeur est de 166,7 cm, ce qui est proche de la valeur médiane et indique une distribution relativement symétrique. Le mode de la variable LARG est de 161 cm, ce qui est légèrement inférieur à la médiane et à la moyenne. Cela suggère que la distribution de la largeur est légèrement biaisée vers la gauche, avec une concentration de valeurs plus élevées sur le côté droit de la distribution. La largeur maximale des voitures dans ce jeu de données est de 177 cm.

La variable POIDS

```
summary(POIDS)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      815   1020   1088    1079   1127   1370
```

```
Mode_POIDS=names(which.max(table(POIDS)))
Mode_POIDS
```

```
## [1] "1080"
```

Interpretations

Pour la variable POIDS de notre jeu de données, on peut observer que la voiture la plus légère pèse 815 kg, tandis que la plus lourde pèse 1370 kg. La médiane de cette variable est de 1088 kg, ce qui signifie que la moitié des voitures pèsent moins de 1088 kg et l'autre moitié pèse plus. La moyenne de la variable POIDS est de 1079 kg, ce qui est légèrement inférieur à la médiane. En outre, on peut voir que le premier quartile est de 1020 kg, ce qui signifie que 25% des voitures ont un poids inférieur ou égal à 1020 kg. Le troisième quartile est de 1127 kg, ce qui signifie que 75% des voitures ont un poids inférieur ou égal à 1127 kg. Le mode de cette variable POIDS est de 1080 kg, ce qui signifie que cette valeur est la plus fréquente dans notre jeu de données.

La variable V.MAX

```
summary(V.MAX)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    140.0   151.2   160.0   158.3   165.0   180.0
```

```
Mode_V.MAX=names(which.max(table(V.MAX)))
Mode_V.MAX
```

```
## [1] "140"
```

Interpretations

La variable V.MAX représente la vitesse maximale atteinte par chaque voiture dans le jeu de données. La vitesse minimale est de 140 km/h et la vitesse maximale est de 180 km/h. La vitesse moyenne des voitures est de 158.3 km/h, tandis que la médiane est de 160 km/h. 25% des voitures ont une vitesse maximale inférieure ou égale à 151.2 km/h, tandis que 75% ont une vitesse maximale inférieure ou égale à 165 km/h. Le mode, c'est-à-dire la valeur la plus fréquente, est de 140 km/h. Cela signifie que plus de voitures ont une vitesse maximale de 140 km/h que de toute autre valeur.

La variable PRIX

```
summary(PRIX)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

```
## 22100 29843 33345 34159 38458 47700
```

```
Mode_PRIX=names(which.max(table(PRIX)))  
Mode_PRIX
```

```
## [1] "22100"
```

Interpretations

La variable PRIX de notre jeu de données représente le prix en dollars de chaque voiture. On peut voir que le prix minimum pour une voiture est de 22100 dollars, tandis que le prix maximum est de 47700 dollars. La moitié des voitures ont un prix inférieur ou égal à la médiane de 33345 dollars. La moyenne des prix est de 34159 dollars. Le troisième quartile des prix est de 38458 dollars, ce qui signifie que 75% des voitures ont un prix inférieur ou égal à ce montant. Enfin, le mode de cette variable est de 22100 dollars, ce qui signifie que plus de voitures ont un prix de 22100 dollars que tout autre prix.

La variable R.POID.PUIS

```
summary(R.POID.PUIS)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
##      9.72  11.22   13.18   13.18   14.55   18.36
```

```
Mode_R.POID.PUIS=names(which.max(table(R.POID.PUIS)))  
Mode_R.POID.PUIS
```

```
## [1] "9.72"
```

Interpretations

La variable R.POID.PUIS est la relation entre le poids et la puissance de la voiture, exprimée en kg/CV. On observe que la valeur minimale pour cette variable est de 9.72 kg/CV, tandis que la valeur maximale est de 18.36 kg/CV. La médiane pour cette variable est de 13.18 kg/CV, ce qui signifie que la moitié des voitures ont une relation poids-puissance inférieure à ce nombre et l'autre moitié ont une relation supérieure. La moyenne est également de 13.18 kg/CV, ce qui indique une répartition relativement uniforme des valeurs de la variable autour de cette moyenne. Enfin, le troisième quartile est de 14.55 kg/CV, ce qui signifie que 75% des voitures ont une relation poids-puissance inférieure ou égale à ce nombre. Le mode de cette variable est de 9.72 kg/CV, ce qui est la valeur la plus fréquente dans les données.

Indicateurs de dispersion : variance, écart-type et coefficients de variation

La variable CYL

```
variance_CYL=var(CYL)  
sd_CYL = sqrt(variance_CYL)  
cv_CYL= sd_CYL/mean(variance_CYL)  
variance_CYL
```

```
## [1] 139823.5
```

```
sd_CYL
```

```
## [1] 373.9298
```

```
cv_CYL
```

```
## [1] 0.002674298
```

Interpretations

La variance de la variable CYL est de 139823.5, ce qui signifie que les valeurs de la cylindrée des moteurs de différentes voitures sont assez dispersées. L'écart-type de CYL est de 373.9298, ce qui est relativement

élevé et confirme la dispersion importante des valeurs de la cylindrée des moteurs. Le coefficient de variation (covariance_CYL) est de 0.002674298. Ce chiffre est relativement faible, ce qui peut indiquer que la dispersion des valeurs de CYL est relativement homogène par rapport à la moyenne de la variable.

La variable PUISS

```
variance_PUISS=var(PUISS)
sd_PUISS = sqrt(variance_PUISS)
covariation_PUISS= sd_PUISS/mean(variance_PUISS)
variance_PUISS
```

```
## [1] 415.1928
```

```
sd_PUISS
```

```
## [1] 20.37628
```

```
covariation_PUISS
```

```
## [1] 0.04907667
```

Interpretations

La variance de PUISS est de 415,1928, ce qui indique que les puissances des voitures sont assez dispersées autour de leur moyenne. La racine carrée de la variance, c'est-à-dire l'écart-type de PUISS, est de 20,37628, ce qui signifie que les puissances des voitures varient en moyenne de 20,37628 par rapport à la moyenne. La covariation de PUISS est de 0,04907667, ce qui indique une faible corrélation entre la puissance et les autres variables du jeu de données.

La variable LONG

```
variance_LONG=var(LONG)
sd_LONG = sqrt(variance_LONG)
covariation_LONG= sd_LONG/mean(variance_LONG)
variance_LONG
```

```
## [1] 488.7353
```

```
sd_LONG
```

```
## [1] 22.10736
```

```
covariation_LONG
```

```
## [1] 0.04523381
```

Interpretations

La variable LONG représente la longueur de chaque véhicule dans la liste. La variance_LONG est égale à 488,7353, ce qui indique que les valeurs de la longueur de la voiture sont dispersées autour de la moyenne. La racine carrée de la variance (sd_LONG) est de 22,10736, ce qui est l'écart-type de la variable. Cela montre que la longueur moyenne des voitures diffère d'environ 22,11 cm de la valeur moyenne. Enfin, la covariation_LONG, qui est égale à 0,04523381, est une mesure de la variabilité relative de la longueur des voitures. Il s'agit d'un indicateur de dispersion relative, qui montre que la longueur des voitures dans la liste a une variabilité relativement faible par rapport à la moyenne de cette variable.

La variable LARG

```
variance_LARG=var(LARG)
sd_LARG = sqrt(variance_LARG)
covariation_LARG= sd_LARG/mean(variance_LARG)
variance_LARG
```

```
## [1] 28.23529
```

```
sd_LARG
```

```
## [1] 5.313689
```

```
covariation_LARG
```

```
## [1] 0.1881932
```

Interpretations

La variable LARG mesure la largeur des voitures dans l'échantillon. Les calculs montrent que la variance de la variable est de 28,23529 et l'écart-type est de 5,313689. Le coefficient de variation, qui est le rapport entre l'écart-type et la moyenne de la variable, est de 0,1881932. Cela signifie que la dispersion des valeurs est relativement faible par rapport à la moyenne de la variable. En d'autres termes, la largeur des voitures dans l'échantillon est relativement homogène et il y a peu de voitures ayant une largeur très différente de la moyenne.

La variable POIDS

```
variance_POIDS=var(POIDS)
sd_POIDS = sqrt(variance_POIDS)
covariation_POIDS= sd_POIDS/mean(variance_POIDS)
variance_POIDS
```

```
## [1] 18757.44
```

```
sd_POIDS
```

```
## [1] 136.9578
```

```
covariation_POIDS
```

```
## [1] 0.007301519
```

Interpretations

La variable "POIDS" représente le poids de la voiture. D'après les résultats obtenus, la variance du poids des voitures dans l'échantillon est de 18757.44, ce qui indique une dispersion importante des données autour de la moyenne. La déviation standard (écart-type) du poids est de 136.9578, ce qui montre que les données sont relativement dispersées par rapport à la moyenne. Enfin, la covariation est de 0.007301519, ce qui indique une faible variation relative du poids des voitures dans l'échantillon par rapport à la moyenne. Cela peut suggérer que les voitures de l'échantillon ont un poids relativement homogène.

La variable V.MAX

```
variance_V.MAX=var(V.MAX)
sd_V.MAX = sqrt(variance_V.MAX)
covariation_V.MAX= sd_V.MAX/mean(variance_V.MAX)
variance_V.MAX
```

```
## [1] 147.3889
```

```
sd_V.MAX
```

```
## [1] 12.14038
```

```
covariation_V.MAX
```

```
## [1] 0.08236973
```

Interpretations

La variable V.MAX représente la vitesse maximale de la voiture. La variance de cette variable est de 147,39. Cela signifie que les valeurs de V.MAX sont assez dispersées autour de la moyenne. L'écart-type de V.MAX est de 12,14, ce qui indique que les valeurs individuelles de V.MAX peuvent varier de manière significative par rapport à la moyenne. La covariation de V.MAX est de 0,082, ce qui indique une faible association entre les valeurs de V.MAX et les autres variables de la base de données.

La variable PRIX

```
variance_PRIX=var(PRIX)
sd_PRIX = sqrt(variance_PRIX)
covariation_PRIX= sd_PRIX/mean(variance_PRIX)
variance_PRIX
```

```
## [1] 43185635
```

```
sd_PRIX
```

```
## [1] 6571.578
```

```
covariation_PRIX
```

```
## [1] 0.0001521705
```

Interpretations

La variable PRIX a une variance de $4,318,5635 \times 10^7$ et un écart-type de 6,571.578. Cela indique que les données sont très dispersées et que les valeurs individuelles peuvent varier considérablement par rapport à la moyenne. De plus, la covariation est très faible, ce qui indique qu'il y a très peu de relation entre la variable PRIX et les autres variables de l'ensemble de données.

La variable R.POID.PUIS

```
variance_R.POID.PUIS=var(R.POID.PUIS)
sd_R.POID.PUIS =sqrt(variance_R.POID.PUIS)
covariation_R.POID.PUIS= sd_R.POID.PUIS/mean(variance_R.POID.PUIS)
variance_R.POID.PUIS
```

```
## [1] 4.989547
```

```
sd_R.POID.PUIS
```

```
## [1] 2.233729
```

```
covariation_R.POID.PUIS
```

```
## [1] 0.4476818
```

Interpretations

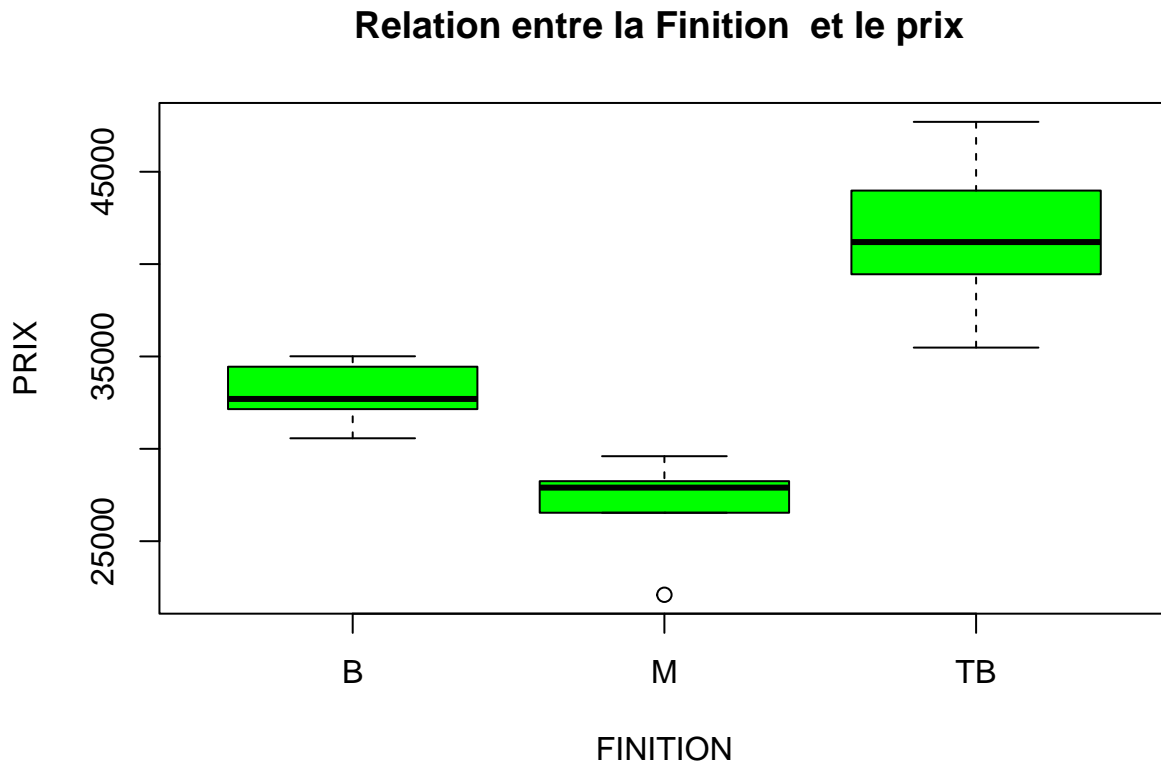
La variable R.POID.PUIS est la relation entre la puissance et le poids du véhicule. La variance de cette variable est de 4.989547, ce qui indique une faible dispersion des données autour de la moyenne. La racine carrée de la variance (ou l'écart type) est de 2.233729, ce qui indique une faible dispersion des données. La covariation de cette variable (0.4476818) est relativement élevée, ce qui indique une forte relation linéaire entre la puissance et le poids du véhicule. En d'autres termes, plus le poids du véhicule est élevé, plus sa puissance est élevée.

2-Analyse bivarier entre la variable PRIX et les autres variables

Analyse Bivarier entre variable quantitative et qualitative

Analyse bivarier entre PRIX et Finition

```
boxplot(PRIX ~ FINITION,main="Relation entre la Finition et le prix ",col="green")
```



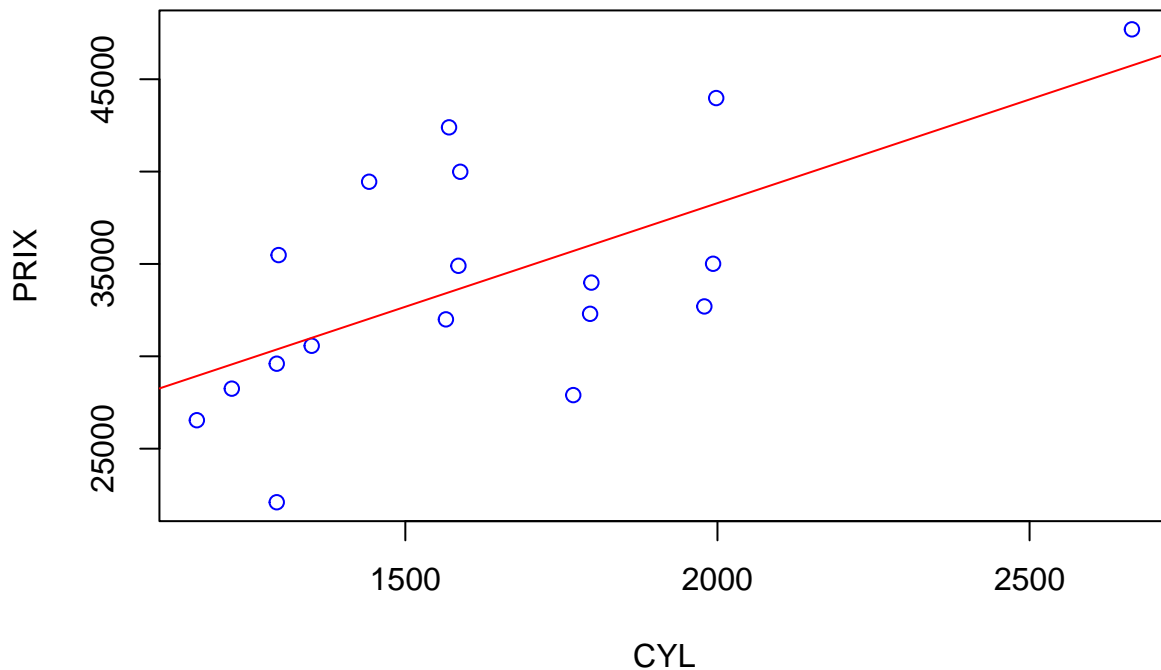
D'après le boxplot il existe également une relation entre le prix et la Finition. L'analyse de ce graphique en boîte montre que le niveau de finition "B" a une gamme de prix plus large et des valeurs plus élevées pour la médiane, le quartile supérieur et les points aberrants que les autres niveaux de finition. Cela suggère que les voitures avec un niveau de finition "B" sont généralement plus chères que les autres niveaux de finition. En revanche, les niveaux de finition "TB" et "M" ont une plage de prix plus étroite et des valeurs plus faibles pour la médiane, le quartile supérieur et les points aberrants, indiquant que ces niveaux de finition sont généralement associés à des voitures moins chères.

Analyse Bivarier entre variable quantitative et quantitative

Analyse bivarier entre *PRIX* et *CYL*

```
plot(PRIX ~ CYL,main="Nuages de points entre Prix et CYL",col="blue")
abline(lm(PRIX ~ CYL),col='red')
```

Nuages de points entre Prix et CYL



```
cord=round(cor(CYL,PRIX),2)
cord
```

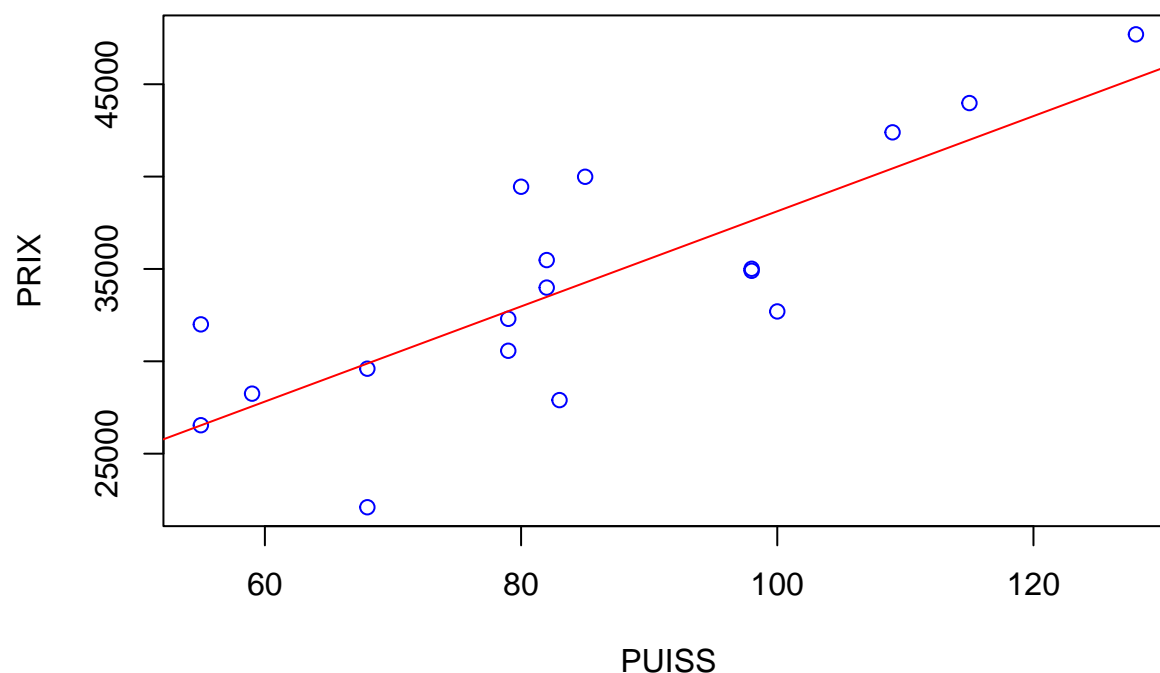
```
## [1] 0.64
```

D'après le nuage de point il existe une relation positive entre le prix et CYL .La valeur de cord qui en ressort est de 0.64 cette valeur indique une corrélation positive moyenne entre le nombre de cylindres et le prix des voitures. En d'autres termes, on peut dire que dans cette base de données, les voitures avec un nombre plus élevé de cylindres ont tendance à être plus chères que les voitures avec un nombre de cylindres moins élevé. La corrélation n'est pas très forte, mais elle est suffisante pour montrer une certaine tendance.

Analyse bivarier entre PRIX et PUISS

```
plot(PRIX~PUISS,main="Nuages de points entre Prix et CYL",col="blue")
abline(lm(PRIX~PUISS),col='red')
```

Nuages de points entre Prix et CYL



```
cord=round(cor(PRIX,PUISS),2)
cord
```

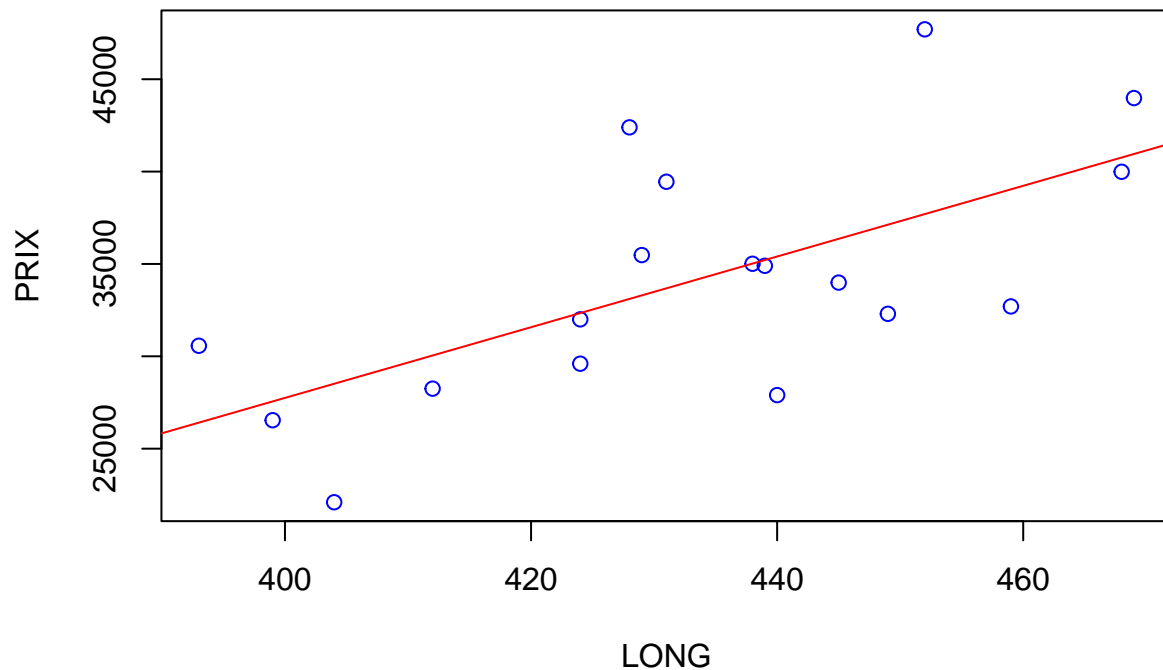
```
## [1] 0.8
```

D'après le nuage de point il existe une relation positive entre le prix et PUISS .La corrélation calculée entre "PRIX" et "PUISS" est de 0.8, ce qui indique une forte corrélation positive entre ces deux variables. Cela suggère que, en général, les voitures plus puissantes ont tendance à coûter plus cher que les voitures moins puissantes.

Analyse bivarier entre PRIX et LONG

```
plot(PRIX~LONG,main="Nuages de points entre Prix et LONG",col="blue")
abline(lm(PRIX~LONG),col='red')
```

Nuages de points entre Prix et LONG



```
cord=round(cor(PRIX,LONG),2)
cord
```

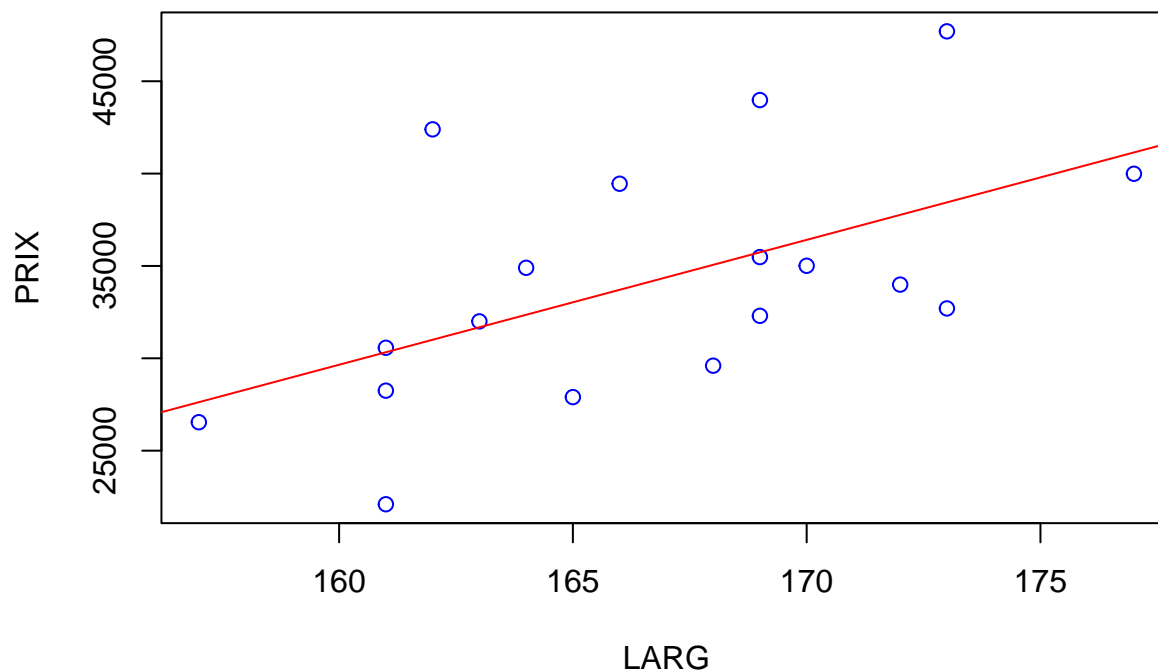
```
## [1] 0.64
```

D'après le nuage de point il existe une relation positive entre le prix et LONG. Le résultat de cord indique que la corrélation entre le prix et la longueur des voitures est de 0.64, arrondi à deux décimales. Cela signifie qu'il y a une corrélation positive modérée entre ces deux variables : en général, plus la voiture est longue, plus son prix est élevé. Cependant, la corrélation n'est pas très forte, ce qui suggère qu'il y a d'autres facteurs qui peuvent influencer le prix des voitures.

Analyse bivarier entre PRIX et LARG

```
plot(PRIX~LARG,main="Nuages de points entre Prix et LARG",col="blue")
abline(lm(PRIX~LARG),col='red')
```

Nuages de points entre Prix et LARG



```
cord=round(cor(PRIX,LARG),2)
cord
```

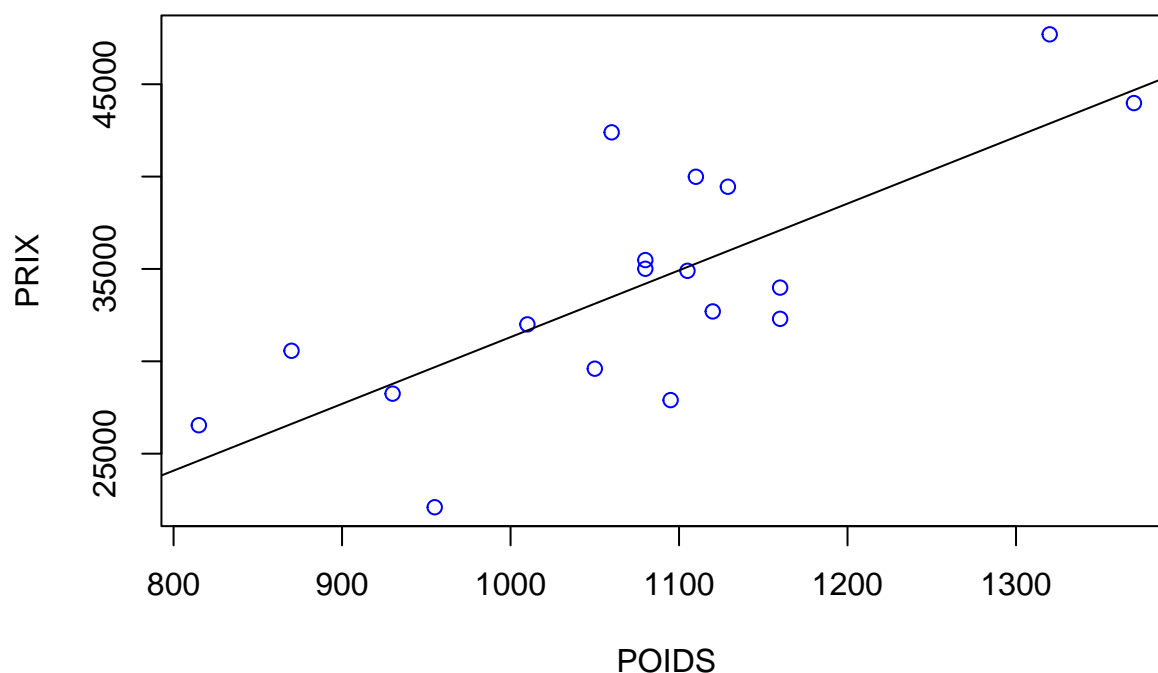
```
## [1] 0.55
```

D'après le nuage de point il existe une relation entre le PRIX et LARG. Le résultat `cord` indique que la corrélation entre la variable "PRIX" et la variable "LARG" est de 0.55, arrondi à deux décimales. Cela suggère une corrélation modérée positive entre le prix et la largeur de la voiture. En d'autres termes, les voitures plus larges ont tendance à avoir des prix plus élevés, bien que la relation ne soit pas très forte.

Analyse bivariée entre PRIX et POIDS

```
plot(PRIX ~ POIDS, main="Nuages de points entre Prix et POIDS", col="blue")
abline(lm(PRIX ~ POIDS), col='black')
```


Nuages de points entre Prix et POIDS



```
cord=round(cor(PRIX,POIDS),2)
cord
```

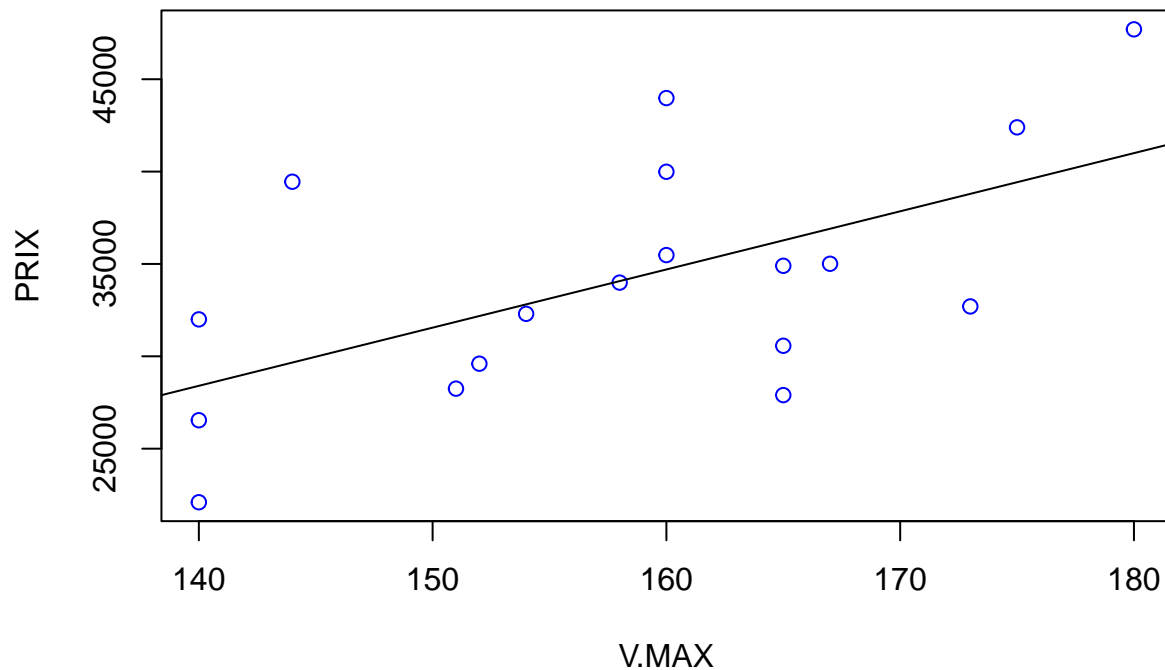
```
## [1] 0.75
```

D'après le nuage de point il existe une relation entre le PRIX et POIDS .Le résultat $\text{cord} = 0.75$. Cela signifie qu'il y a une corrélation positive modérément forte entre le prix et le poids des voitures. En d'autres termes, on peut s'attendre à ce que les voitures plus lourdes soient généralement plus chères que les voitures plus légères.

Analyse bivarier entre PRIX et V.MAX

```
plot(PRIX ~ V.MAX,main="Nuages de points entre Prix et V.MAX",col="blue")
abline(lm(PRIX ~ V.MAX),col='black')
```

Nuages de points entre Prix et V.MAX



```
cord=round(cor(PRIX,V.MAX),2)
cord
```

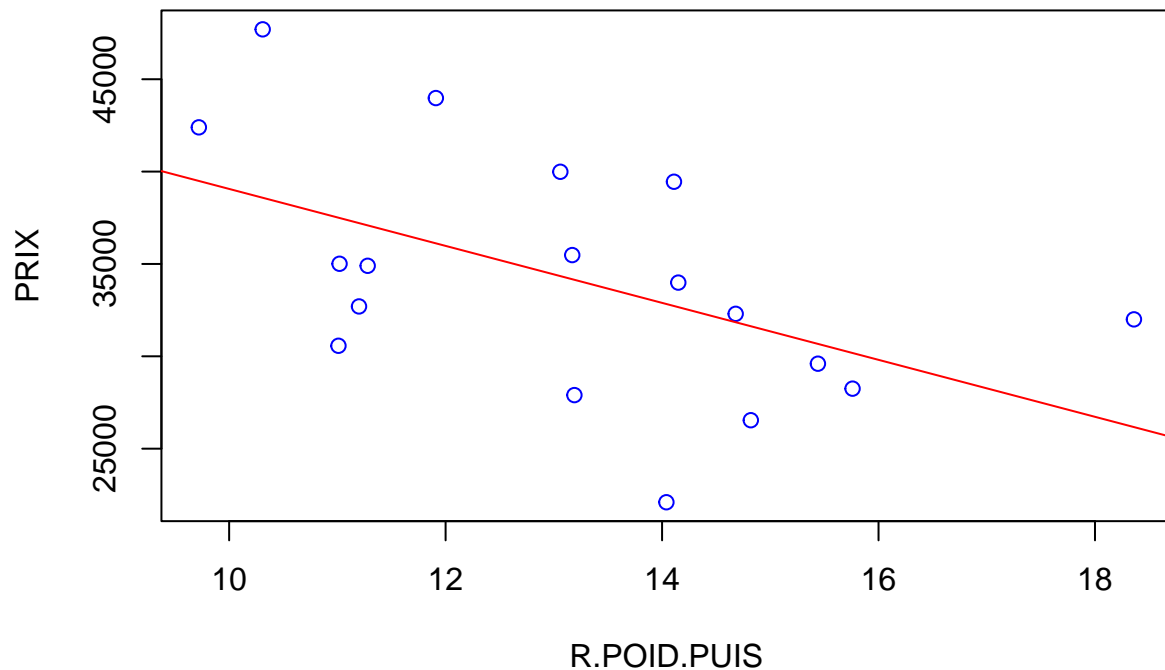
```
## [1] 0.58
```

D'après le nuage de point il existe une relation positive entre le PRIX et V.MAX .Le coefficient de corrélation entre le prix et la vitesse maximale est de 0.58, arrondi à deux décimales. Cela suggère qu'il y a une corrélation positive modérée entre le prix et la vitesse maximale des voitures, ce qui signifie que les voitures plus chères ont tendance à avoir une vitesse maximale plus élevée.

Analyse bivarier entre PRIX et R.POID.PUIS

```
plot(PRIX ~ R.POID.PUIS,main="Nuages de points entre Prix et R.POID.PUIS",col="blue")
abline(lm(PRIX ~ R.POID.PUIS),col='red')
```

Nuages de points entre Prix et R.POID.PUIS



```
cord=round(cor(PRIX,R.POID.PUIS),2)  
cord
```

```
## [1] -0.52
```

D'après le nuage de point il existe une relation négative entre le PRIX et R.POID.PUIS. Le résultat `cord=-0.52` signifie qu'il y a une corrélation négative modérée entre le prix et le rapport poids/puissance du véhicule. Autrement dit, lorsque le rapport poids/puissance d'un véhicule augmente, le prix a tendance à diminuer et vice versa. Cela peut s'expliquer par le fait que les voitures plus légères avec une puissance élevée peuvent être considérées comme plus sportives et donc plus chères, tandis que les voitures plus lourdes avec une puissance élevée peuvent être moins chères en raison de leur poids supplémentaire.