VISUALISATION AVANCEE DES DONNEES DANS R AVEC GRAPHIQUES INTERACTIFS (Formation complète)

Section 2 : Apprendre à Créer des graphiques de corrélation

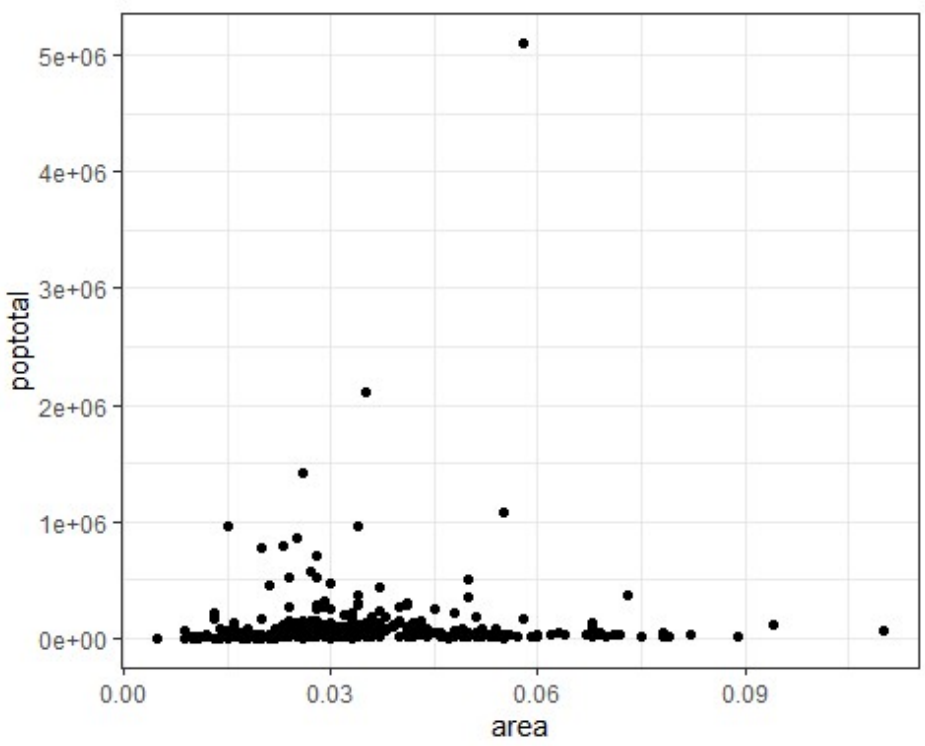
Les graphiques que nous allons construire dans cette section aident à examiner dans quelle mesure deux ou plusieurs variables sont corrélées.

Nuages de points

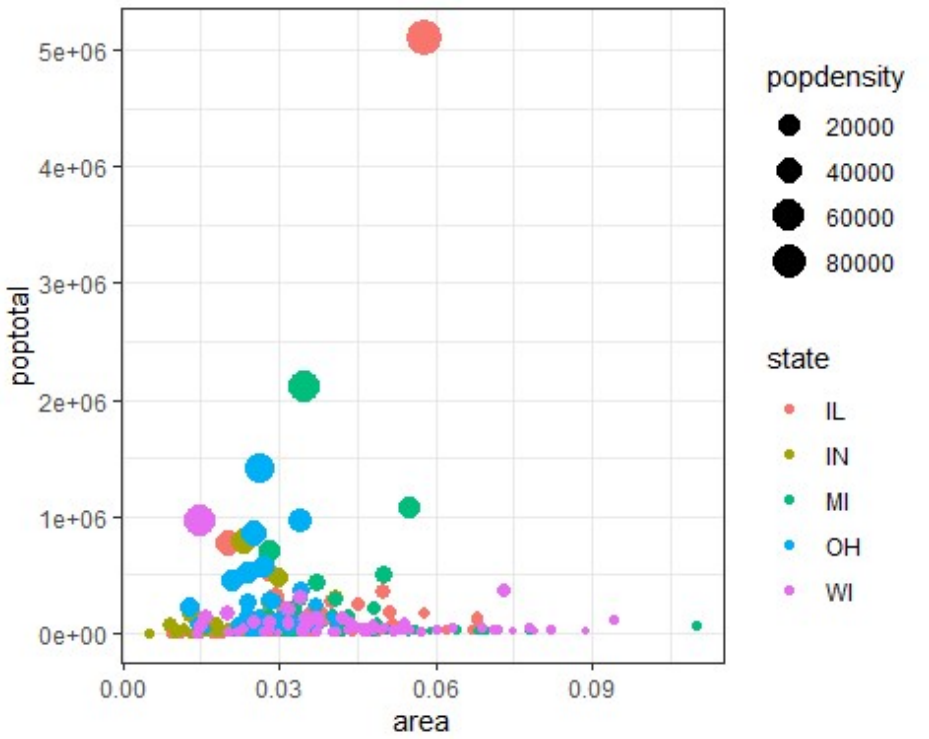
```
theme_set(theme_bw())

data("midwest", package = "ggplot2")

head(midwest)
```

```
## # A tibble: 6 × 28
##     PID county   state  area poptotal popdensity popwhite popblack
popamerindian
##   <int> <chr>    <chr> <dbl>    <int>      <dbl>    <int>    <int>
<int>
## 1   561 ADAMS    IL    0.052    66090      1271.    63917     1702
98
## 2   562 ALEXAND… IL    0.014    10626       759      7054     3496
19
## 3   563 BOND     IL    0.022    14991       681.    14477      429
35
## 4   564 BOONE    IL    0.017    30806      1812.    29344      127
46
## 5   565 BROWN    IL    0.018     5836       324.     5264      547
14
## 6   566 BUREAU   IL    0.05     35688       714.    35157       50
65
## # ℹ 19 more variables: popasian <int>, popother <int>, percwhite <dbl>,
## #   percblack <dbl>, percamerindan <dbl>, percasian <dbl>, percother
<dbl>,
## #   popadults <int>, perchsd <dbl>, percollege <dbl>, percprof <dbl>,
## #   poppovertyknown <int>, percpovertyknown <dbl>, percbelowpoverty <dbl>,
## #   percchildbelowpovert <dbl>, percadultpoverty <dbl>,
## #   percelderlypoverty <dbl>, inmetro <int>, category <chr>
```

```
gg<- ggplot(data= midwest, aes(x=area, y= poptotal))+ geom_point()

gg
```

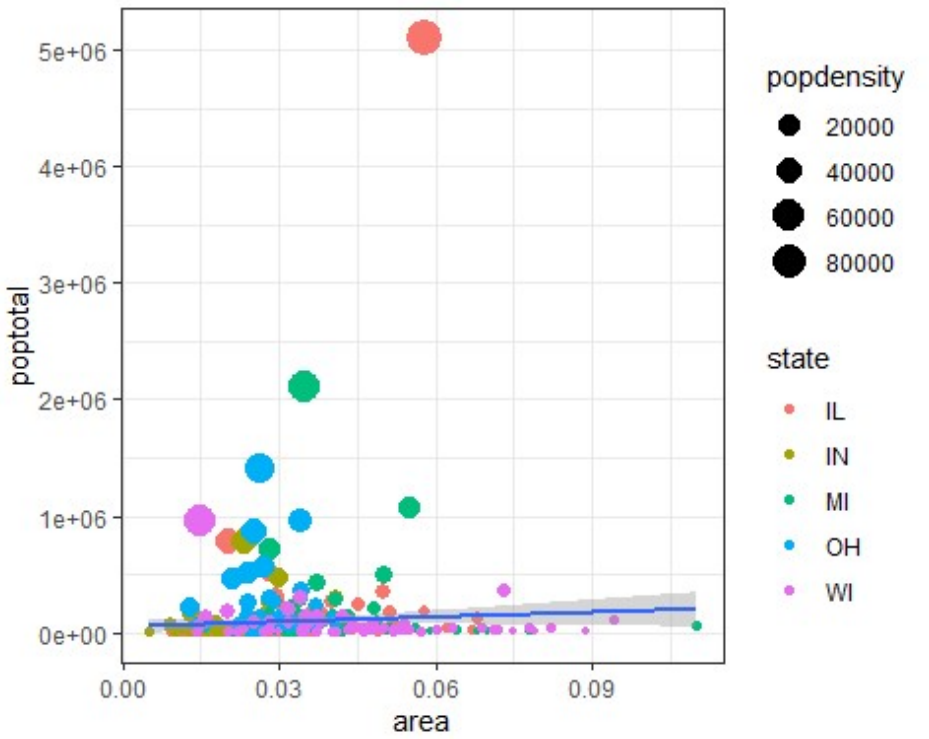



```
gg<- ggplot(data= midwest, aes(x= area, y = poptotal))+ geom_point(aes(col=
state, size = popdensity))

gg
```



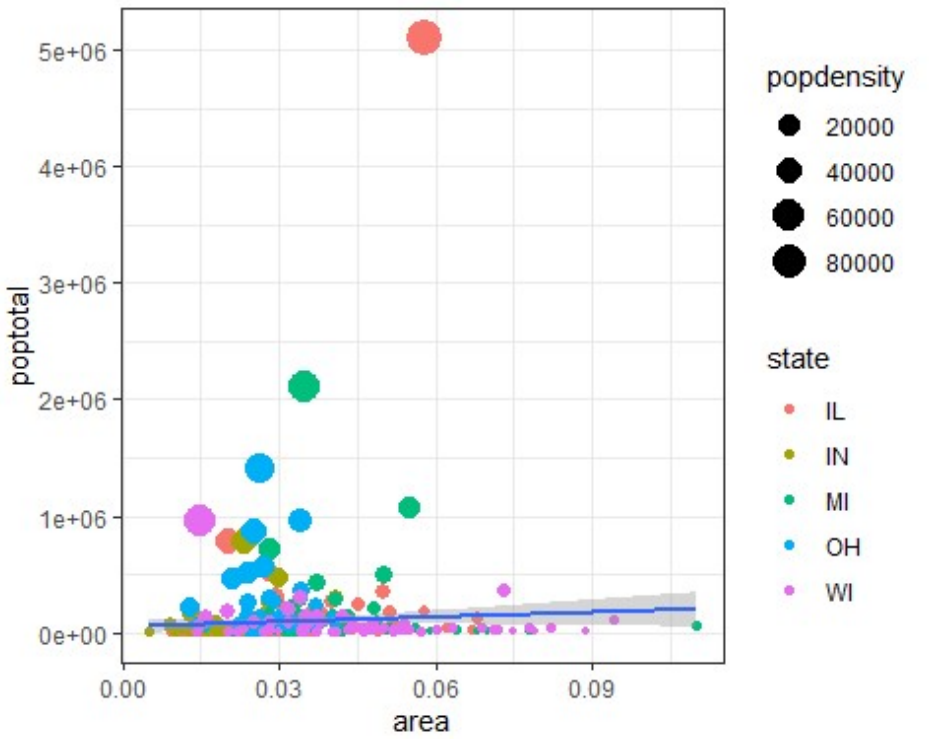

```
gg<- ggplot(data= midwest, aes(x= area, y = poptotal))+ geom_point(aes(col=
state, size = popdensity)) + geom_smooth(method = "lm", se = T)

gg
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



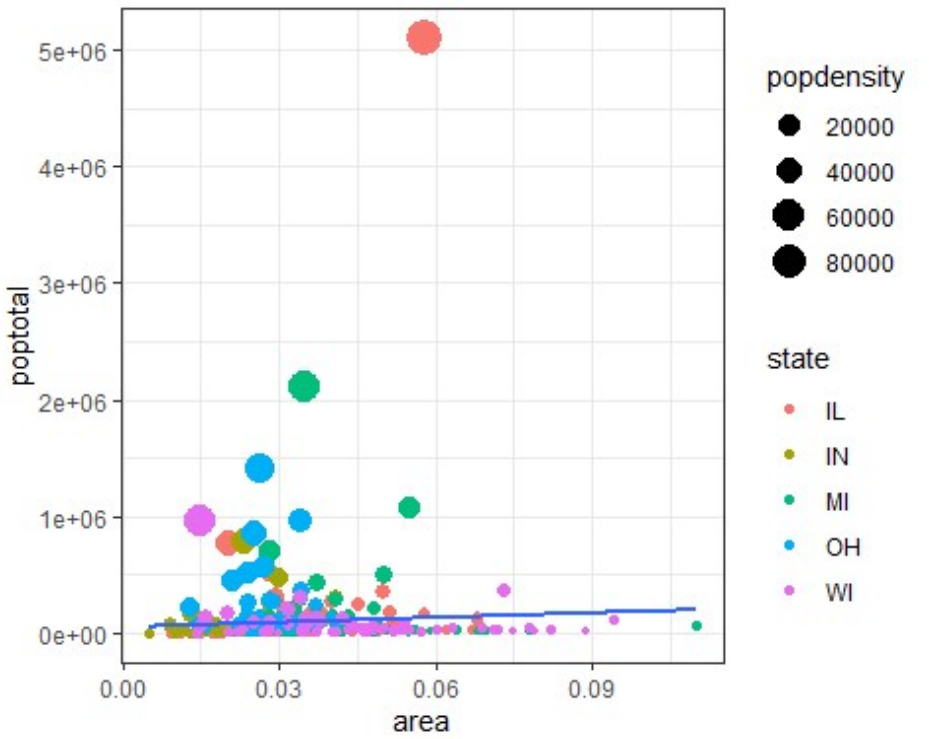

```
gg<- ggplot(data= midwest, aes(x= area, y = poptotal))+ geom_point(aes(col=
state, size = popdensity)) + geom_smooth(method = "lm", se = T)

gg
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

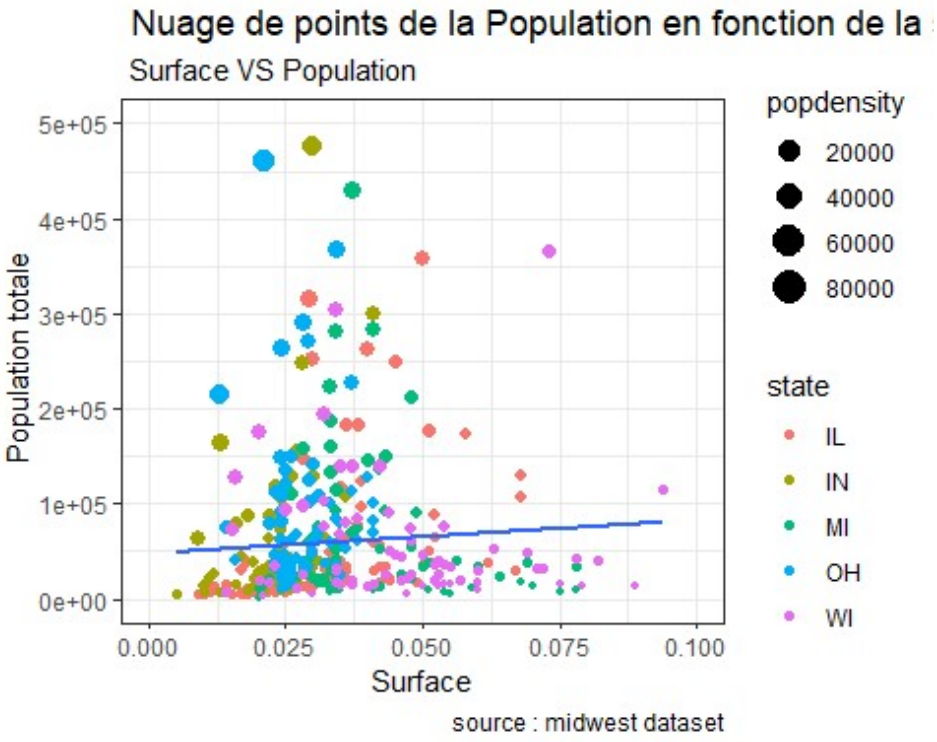



```
gg<- ggplot(data= midwest, aes(x= area, y = poptotal))+ geom_point(aes(col=
state, size = popdensity)) + geom_smooth(method = "lm", se = F)


gg
```

```
## `geom_smooth()` using formula = 'y ~ x'
```




```
gg<- ggplot(data= midwest, aes(x= area, y = poptotal))+ geom_point(aes(col=
state, size = popdensity)) + geom_smooth(method = "lm", se = F) +
  xlim(c(0, 0.1)) +
  ylim(c(0, 500000)) +
  labs(subtitle = " Surface VS Population",
       y = " Population totale",
       x = "Surface",
       title = " Nuage de points de la Population en fonction de la surface
du comté",
       caption = "source : midwest dataset")

gg
```




```
ggplotly(gg)
```

```
## Error in parse(text = input): tentative d'utilisation de nom de variable
de longueur nulle
```

Jitter plot : Nuage de points spécial

```
data(mpg, package = "ggplot2")

head(mpg)
```

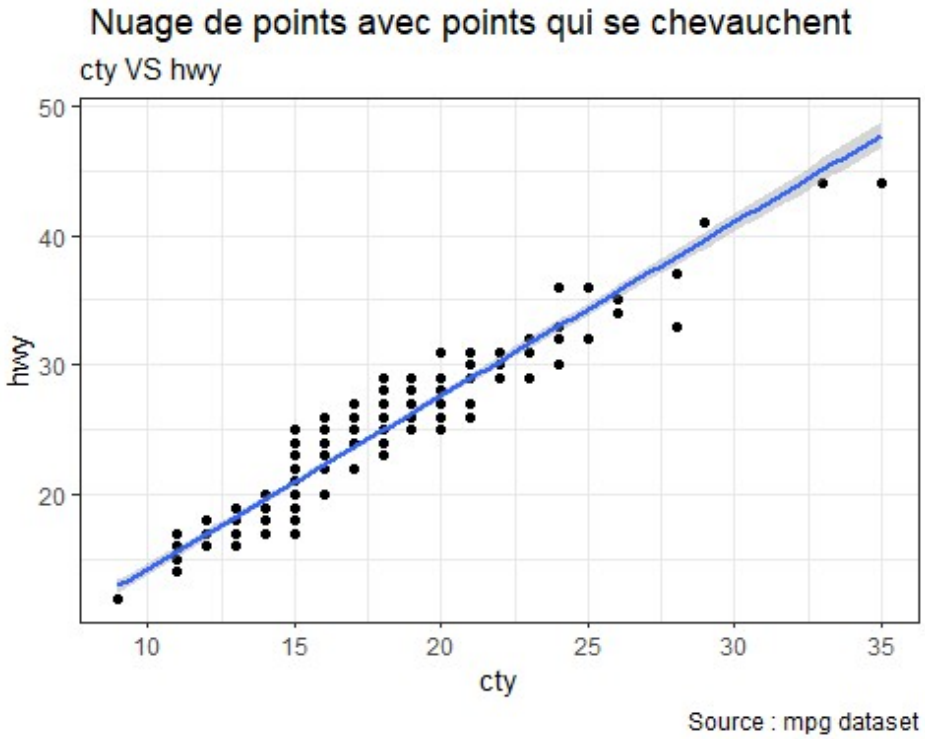
```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl
class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p
compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p
compa…
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p
compa…
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p
compa…
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p
compa…
```

```
## 6 audi          a4      2.8  1999     6 manual(m5) f        18    26 p
compa…
```

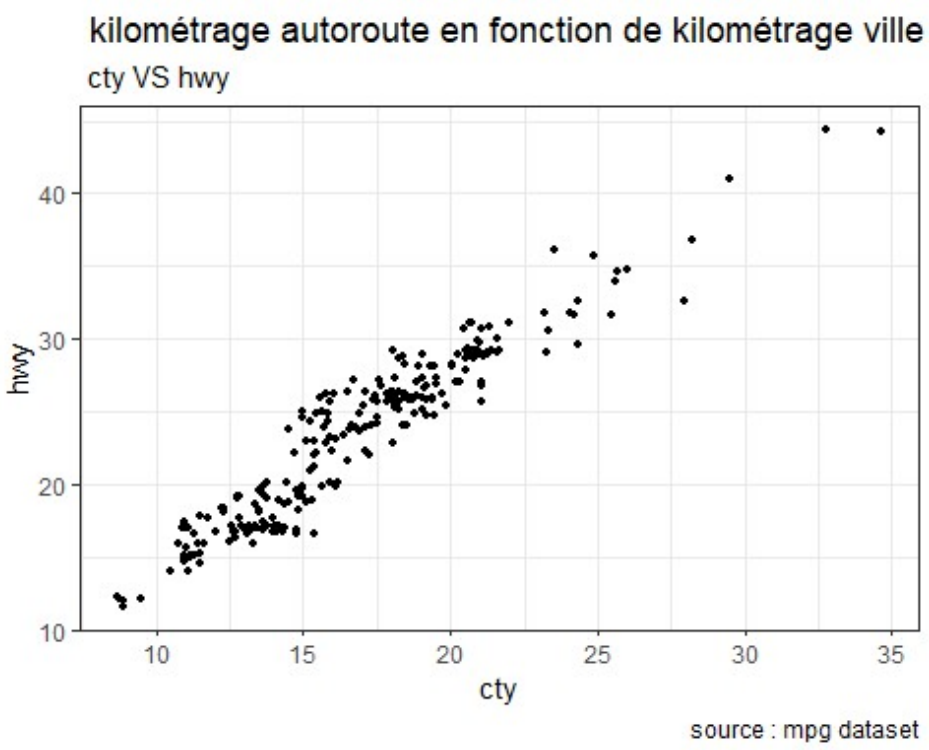
```
str(mpg)
```

```
## tibble [234 × 11] (S3: tbl_df/tbl/data.frame)
##  $ manufacturer: chr [1:234] "audi" "audi" "audi" "audi" ...
##  $ model       : chr [1:234] "a4" "a4" "a4" "a4" ...
##  $ displ       : num [1:234] 1.8 1.8 2 2 2.8 2.8 3.1 1.8 1.8 2 ...
##  $ year        : int [1:234] 1999 1999 2008 2008 1999 1999 2008 1999 1999
2008 ...
##  $ cyl         : int [1:234] 4 4 4 4 6 6 6 4 4 4 ...
##  $ trans       : chr [1:234] "auto(l5)" "manual(m5)" "manual(m6)"
"auto(av)" ...
##  $ drv         : chr [1:234] "f" "f" "f" "f" ...
##  $ cty         : int [1:234] 18 21 20 21 16 18 18 18 16 20 ...
##  $ hwy         : int [1:234] 29 29 31 30 26 26 27 26 25 28 ...
##  $ fl          : chr [1:234] "p" "p" "p" "p" ...
##  $ class       : chr [1:234] "compact" "compact" "compact" "compact" ...
```

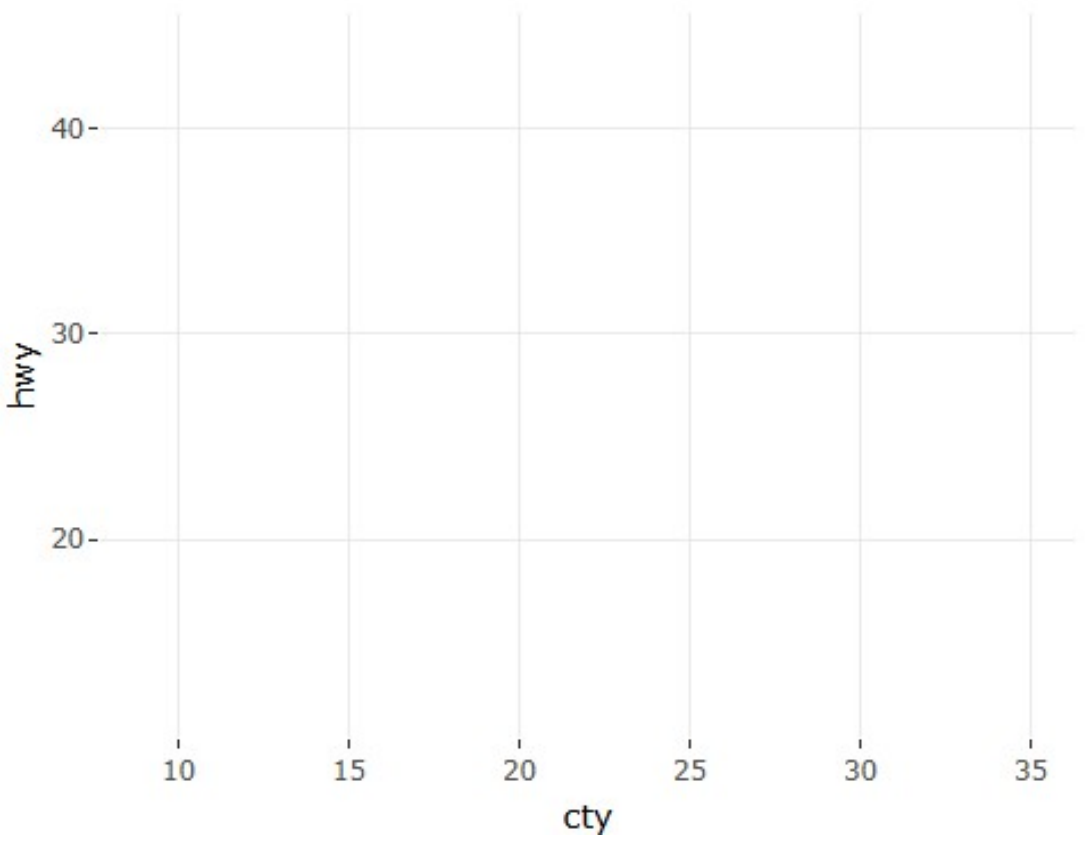
```
g<- ggplot(mpg, aes( x = cty, y = hwy))

g + geom_point() +
  geom_smooth(method = "lm") +
labs(subtitle = "cty VS hwy",
      title = " Nuage de points avec points qui se chevauchent ",
      caption = " Source : mpg dataset")
```



```
g + geom_jitter(width = 0.5, size = 1) +
 labs( subtitle = " cty VS hwy",
       title = " kilométrage autoroute en fonction de kilométrage ville",
       caption = "source : mpg dataset")
```

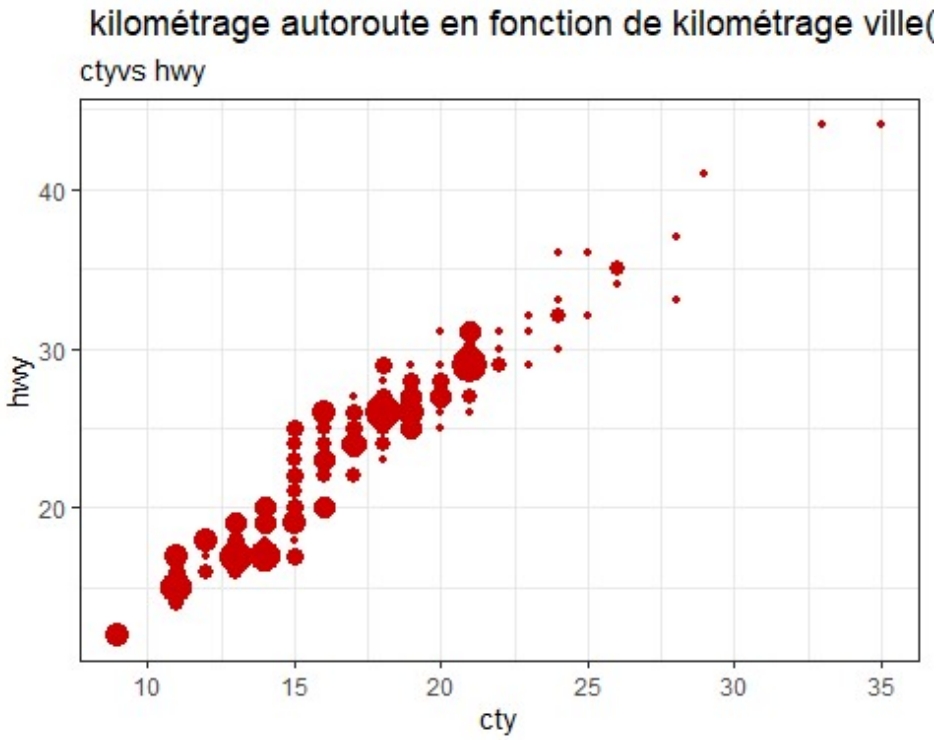



```
ggplotly(g)

##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc05733210f/widgetcc035c75369.html screenshot completed
```



Graphique de décomptes (*Counts charts*)

```
g <- ggplot(mpg, aes(cty, hwy))

g <- g + geom_count(col= ' red3', show.legend = F) +
  labs( subtitle = 'ctyvs hwy',
title = " kilométrage autoroute en fonction de kilométrage ville( count
chart)")

g
```




```
ggplotly(g)
```

```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

Graphique à bulle (*Bubble plot*)

```
mpg_select <- mpg[mpg$manufacturer %in% c("audi","ford","honda", "hyundai"),
]
print(unique(mpg_select$manufacturer))
```

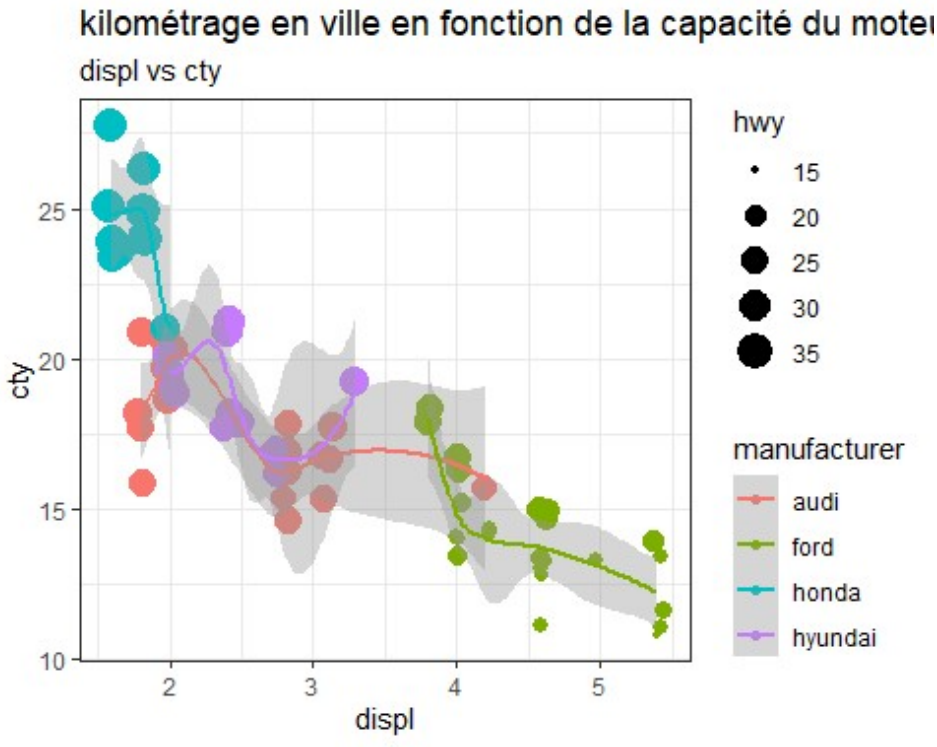
```
## [1] "audi"    "ford"    "honda"   "hyundai"
```

```
head(mpg_select)
```

```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl
class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p
compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p
compa…
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p
compa…
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p
compa…
```

```
## 5 audi          a4      2.8  1999     6 auto(l5)   f         16    26 p
compa…
## 6 audi          a4      2.8  1999     6 manual(m5) f         18    26 p
compa…
```

```
g <- ggplot(mpg_select, aes(displ, cty)) +
  labs( subtitle = "displ vs cty",
        title = "kilométrage en ville en fonction de la capacité du moteur")

g <- g + geom_jitter(aes(col = manufacturer, size = hwy)) +
  geom_smooth(aes(col = manufacturer, method ="lm", se = F))
g
```



Histogramme marginal

```
mpg_select <- mpg[mpg$hwy >= 35 & mpg$cty > 27, ]
g <- ggplot(mpg, aes(cty, hwy)) +
  geom_count(col = "tomato3", show.legend = F) +
  geom_smooth(method ="lm", se = F)

ggMarginal(g, type = "histogram", fill = " transparent")
```

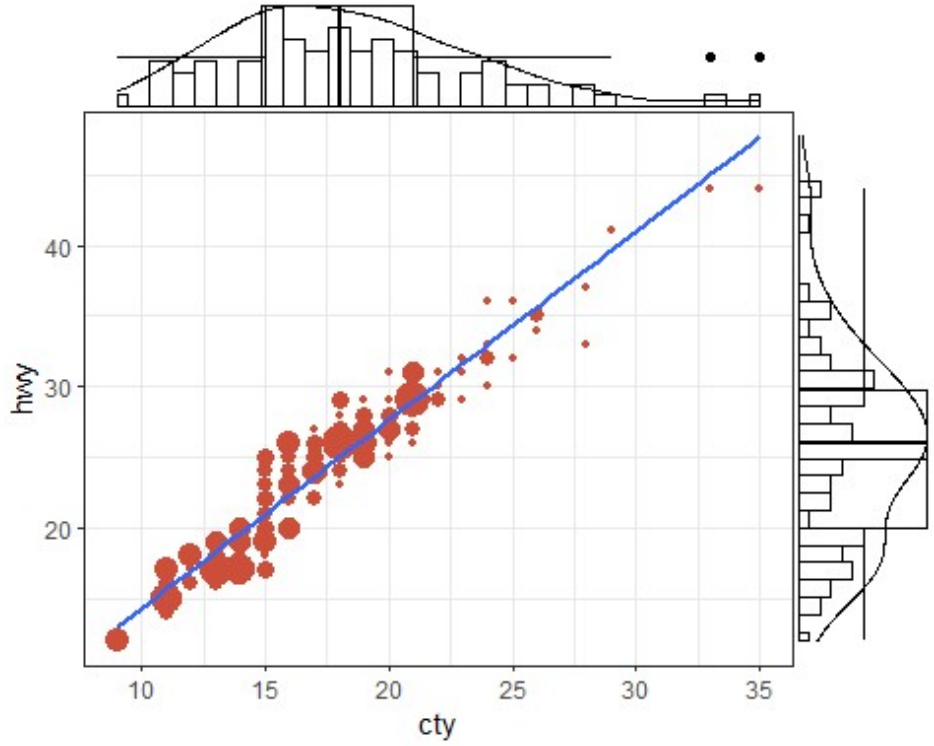
```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

```
ggMarginal(g, type = "boxplot", fill = " transparent")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

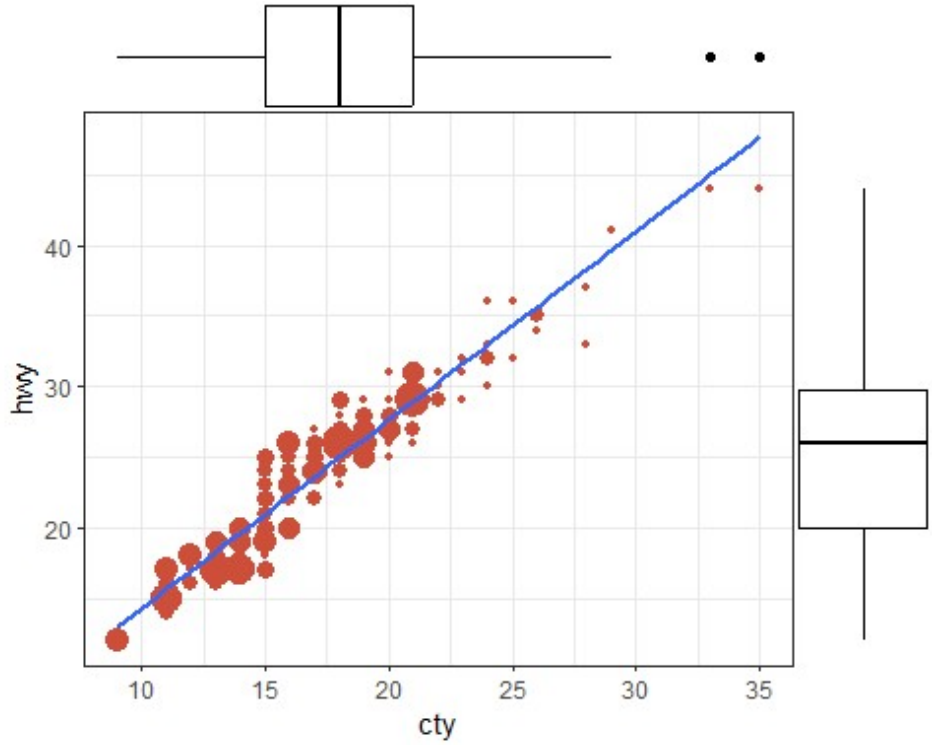
```
ggMarginal(g, type = "density", fill = " transparent")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



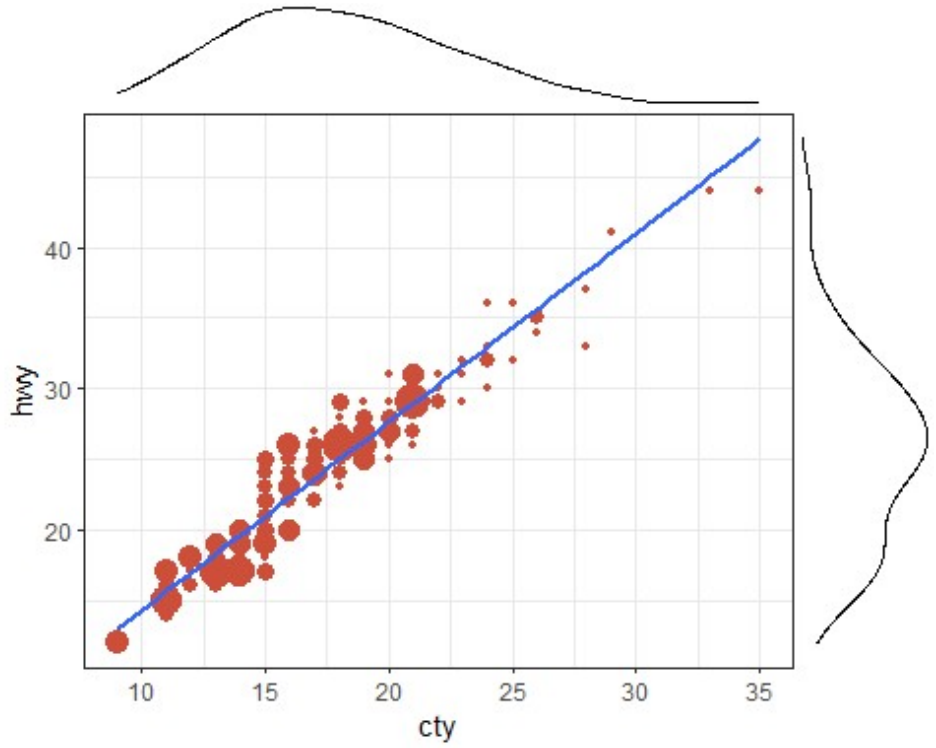

```
mpg_select <- mpg[mpg$hwy >= 35 & mpg$cty > 27, ]
g <- ggplot(mpg, aes(cty, hwy)) +
  geom_count(col = "tomato3", show.legend = F) +
  geom_smooth(method ="lm", se = F)
ggMarginal(g, type = "boxplot", fill = " transparent")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```




```
mpg_select <- mpg[mpg$hwy >= 35 & mpg$cty > 27, ]
g <- ggplot(mpg, aes(cty, hwy)) +
  geom_count(col = "tomato3", show.legend = F) +
  geom_smooth(method ="lm", se = F)
ggMarginal(g, type = "density", fill = " transparent")
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```



Matrice de corrélation

```
head(mtcars)
```

```
##                   mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

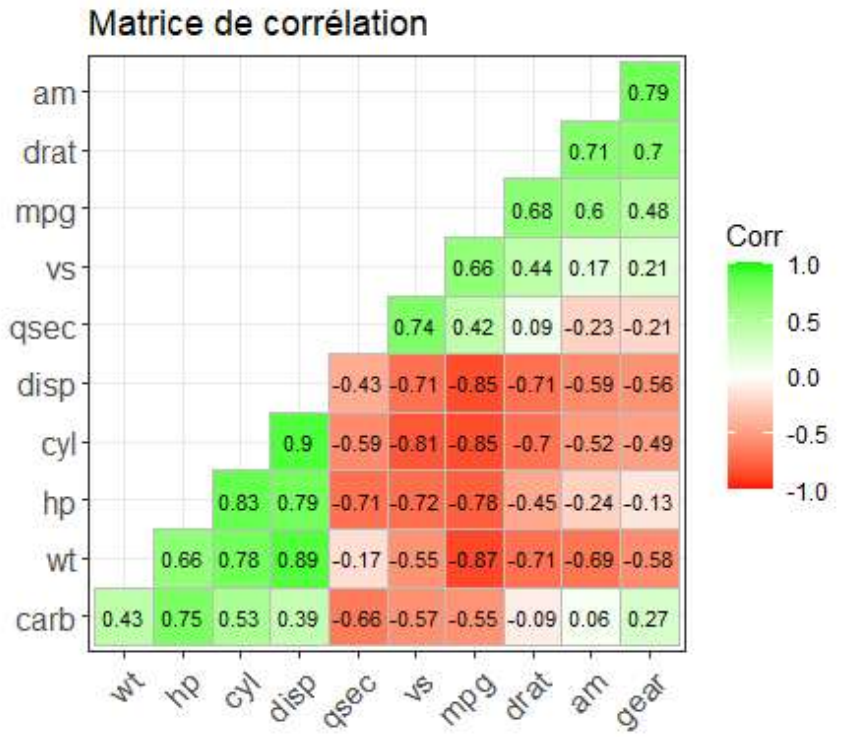
```r
corr <- round(cor(mtcars),2)
corr
```

```
##        mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
## mpg   1.00 -0.85 -0.85 -0.78  0.68 -0.87  0.42  0.66  0.60  0.48 -0.55
## cyl  -0.85  1.00  0.90  0.83 -0.70  0.78 -0.59 -0.81 -0.52 -0.49  0.53
## disp -0.85  0.90  1.00  0.79 -0.71  0.89 -0.43 -0.71 -0.59 -0.56  0.39
## hp   -0.78  0.83  0.79  1.00 -0.45  0.66 -0.71 -0.72 -0.24 -0.13  0.75
## drat  0.68 -0.70 -0.71 -0.45  1.00 -0.71  0.09  0.44  0.71  0.70 -0.09
## wt   -0.87  0.78  0.89  0.66 -0.71  1.00 -0.17 -0.55 -0.69 -0.58  0.43
## qsec  0.42 -0.59 -0.43 -0.71  0.09 -0.17  1.00  0.74 -0.23 -0.21 -0.66
## vs    0.66 -0.81 -0.71 -0.72  0.44 -0.55  0.74  1.00  0.17  0.21 -0.57
## am    0.60 -0.52 -0.59 -0.24  0.71 -0.69 -0.23  0.17  1.00  0.79  0.06
## gear  0.48 -0.49 -0.56 -0.13  0.70 -0.58 -0.21  0.21  0.79  1.00  0.27
## carb -0.55  0.53  0.39  0.75 -0.09  0.43 -0.66 -0.57  0.06  0.27  1.00
```

```r
cor_plot <- ggcorrplot(corr  = corr,
                       hc.order = TRUE,
                       type = "lower",
                       lab = TRUE,
                       lab_size = 3,
                       colors = c("red", "white","green"),
                       title = "Matrice de corrélation",
                       ggtheme = theme_bw)
```

```
## Warning: `aes_string()` was deprecated in ggplot2 3.0.0.
## ℹ Please use tidy evaluation idioms with `aes()`.
## ℹ See also `vignette("ggplot2-in-packages")` for more information.
## ℹ The deprecated feature was likely used in the ggcorrplot package.
##   Please report the issue at
<https://github.com/kassambara/ggcorrplot/issues>.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```r
cor_plot
```



Section 3 : Apprendre à créer des graphiques de déviation

Dans cette section, les graphiques que nous construiront serviront à comparer la variation des valeurs entre un petit nombre d'éléments (ou catégories) par rapport à une référence fixe.

Diagramme à barres divergentes

```
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

```
mtcars$car_name <- rownames(mtcars)
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
```

```
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##                           car_name
## Mazda RX4                  Mazda RX4
## Mazda RX4 Wag          Mazda RX4 Wag
## Datsun 710                Datsun 710
## Hornet 4 Drive        Hornet 4 Drive
## Hornet Sportabout Hornet Sportabout
## Valiant                     Valiant
```

```
mtcars$mpg_z <- round((mtcars$mpg - mean(mtcars$mpg))/ sd(mtcars$mpg), 2)

print(head(mtcars))
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##                           car_name mpg_z
## Mazda RX4                  Mazda RX4  0.15
## Mazda RX4 Wag          Mazda RX4 Wag  0.15
## Datsun 710                Datsun 710  0.45
## Hornet 4 Drive        Hornet 4 Drive  0.22
## Hornet Sportabout Hornet Sportabout -0.23
## Valiant                     Valiant -0.33
```

```
print(summary(mtcars$mpg_z))
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -1.610000 -0.772500 -0.150000  0.000625  0.450000  2.290000
```

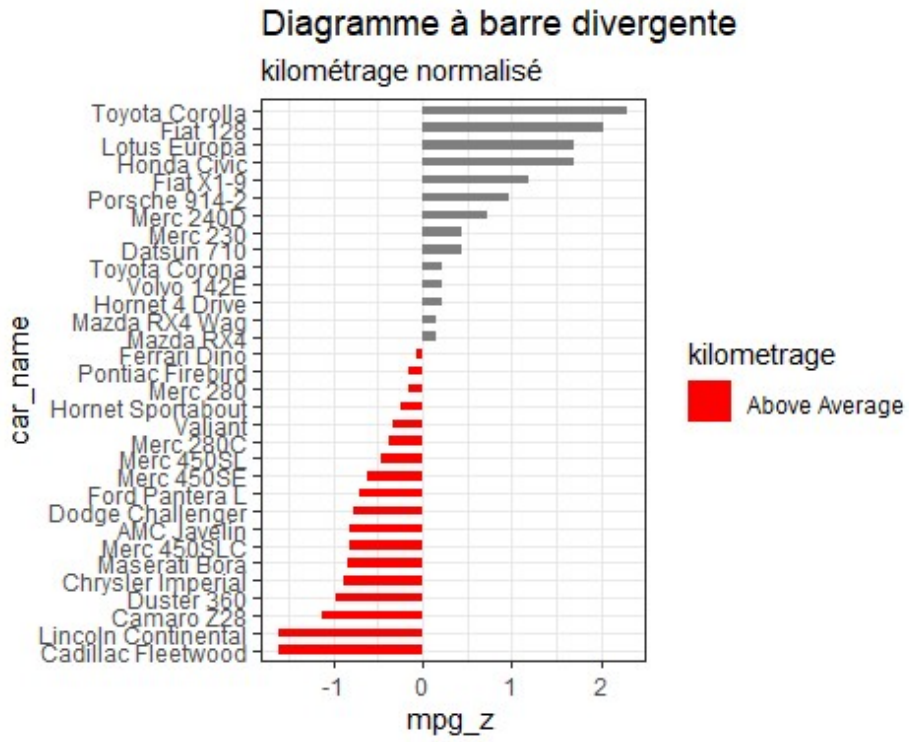
```
print(sd(mtcars$mpg_z))
```

```
## [1] 0.9997611
```

```
mtcars$mpg_type <- ifelse(mtcars$mpg_z < 0, "below", "above")
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##                           car_name mpg_z mpg_type
## Mazda RX4                  Mazda RX4  0.15    above
## Mazda RX4 Wag          Mazda RX4 Wag  0.15    above
## Datsun 710                Datsun 710  0.45    above
```

```
## Hornet 4 Drive       Hornet 4 Drive  0.22    above
## Hornet Sportabout Hornet Sportabout -0.23    below
## Valiant                     Valiant -0.33    below

mtcars <- mtcars[order(mtcars$mpg_z), ]
mtcars$car_name <- factor(mtcars$car_name, levels = mtcars$car_name)

print(head(mtcars))

##                      mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Cadillac Fleetwood  10.4   8  472 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8  460 215 3.00 5.424 17.82  0  0    3    4
## Camaro Z28          13.3   8  350 245 3.73 3.840 15.41  0  0    3    4
## Duster 360          14.3   8  360 245 3.21 3.570 15.84  0  0    3    4
## Chrysler Imperial   14.7   8  440 230 3.23 5.345 17.42  0  0    3    4
## Maserati Bora       15.0   8  301 335 3.54 3.570 14.60  0  1    5    8
##                                car_name mpg_z mpg_type
## Cadillac Fleetwood   Cadillac Fleetwood -1.61    below
## Lincoln Continental Lincoln Continental -1.61    below
## Camaro Z28                   Camaro Z28 -1.13    below
## Duster 360                   Duster 360 -0.96    below
## Chrysler Imperial     Chrysler Imperial -0.89    below
## Maserati Bora             Maserati Bora -0.84    below

 Diag <- ggplot(mtcars, aes(x = car_name, y = mpg_z)) +
  geom_bar(stat = "identity", aes(fill = mpg_type), width = 0.5) +
scale_fill_manual( name = "kilometrage",

labels = c('Above Average', 'Below Average'),

values = c("avove" = "blue", "below" = "red")) + labs(subtitle = "kilométrage
normalisé", title = "Diagramme à barre divergente") +
  coord_flip()
Diag
```




```
ggplotly(Diag)
```

```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc011f758a3/widgetcc05f644d84.html screenshot completed
```

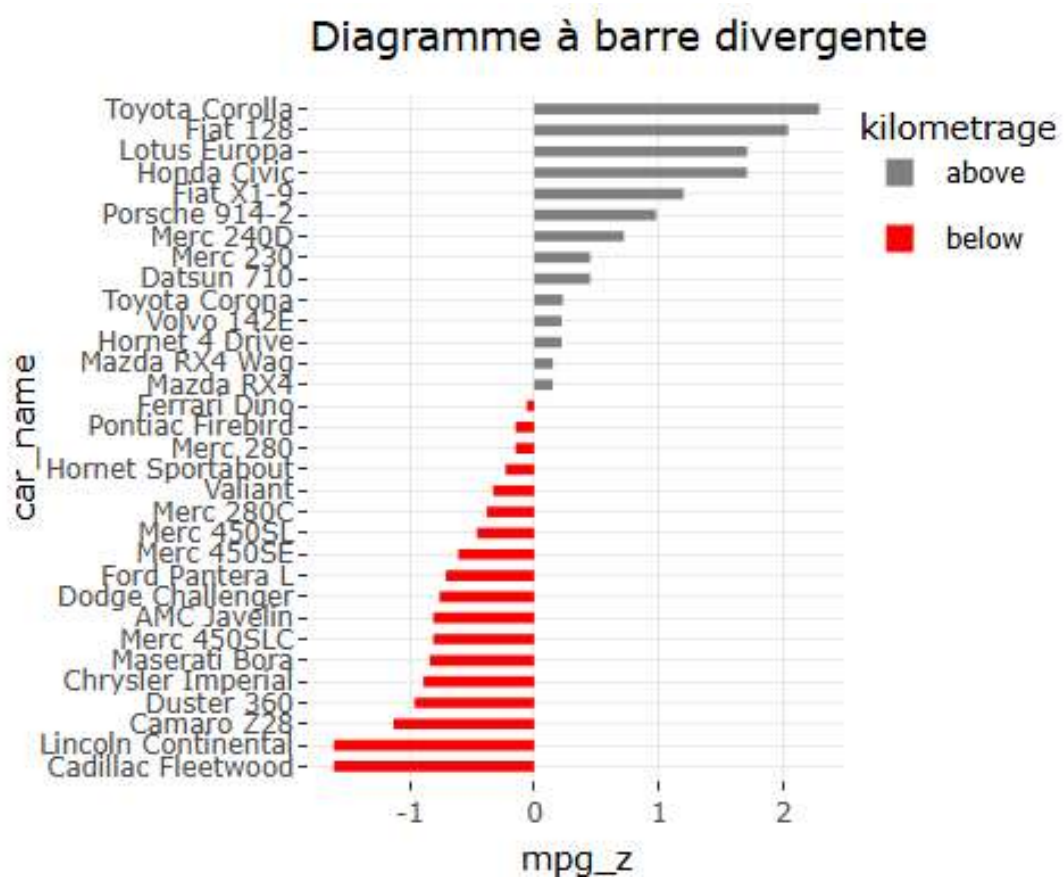
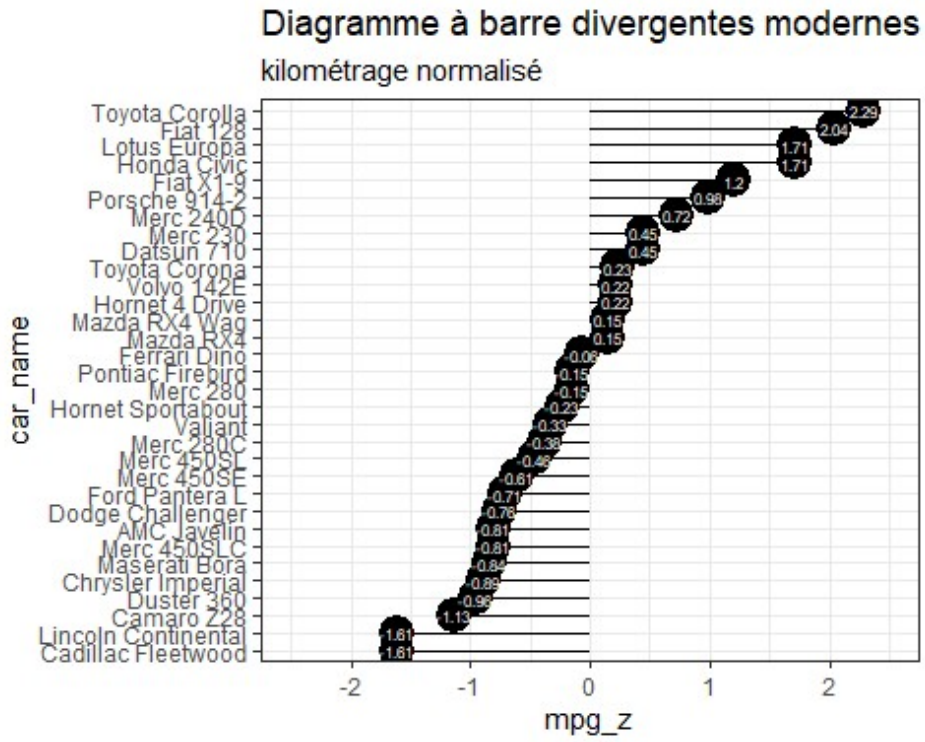


Diagramme à barres divergentes moderne

```
bdm <- ggplot(mtcars, aes(car_name, mpg_z), labels = mpg_z) + geom_point(stat
= "identity", fill = "black", size = 6) + geom_segment(aes(y = 0,
                              x = car_name,
                              yend = mpg_z,
                              xend = car_name,),
                              color = "black") +
  geom_text(color = "white", size = 2, aes(label = mpg_z)) +
  labs(title = "Diagramme à barre divergentes modernes",
       subtitle = "kilométrage normalisé") +
  ylim(-2.5, 2.5) +
  coord_flip()
```

```
## Warning in fortify(data, ...): Arguments in `...` must be used.
## ✖ Problematic argument:
## • labels = mpg_z
## ℹ Did you misspell an argument name?
```

```
bdm
```

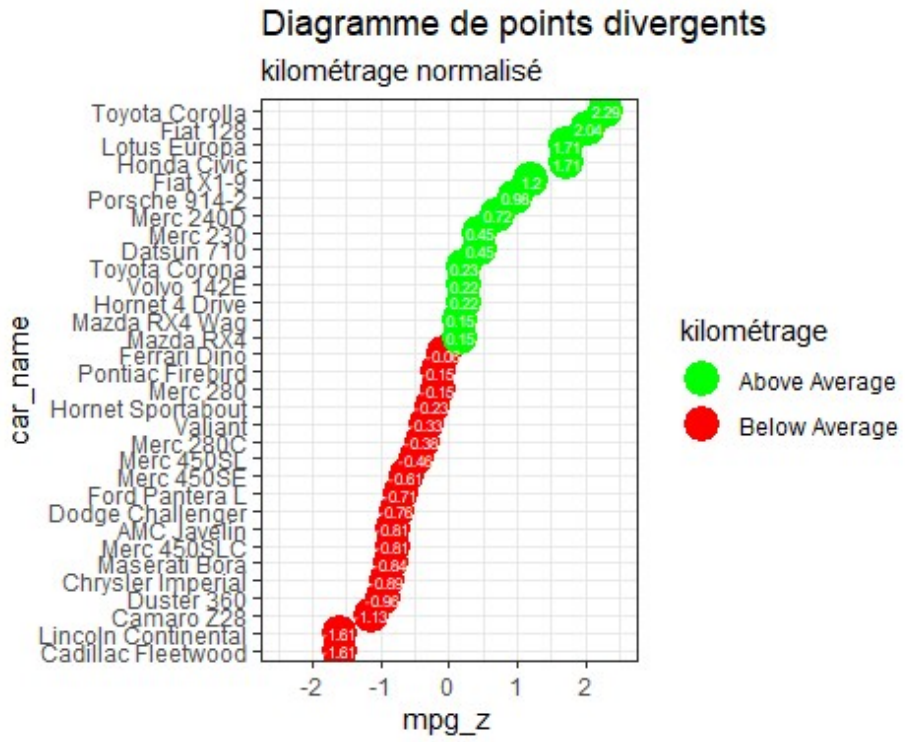



```
ggplotly(bdm)
```

```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

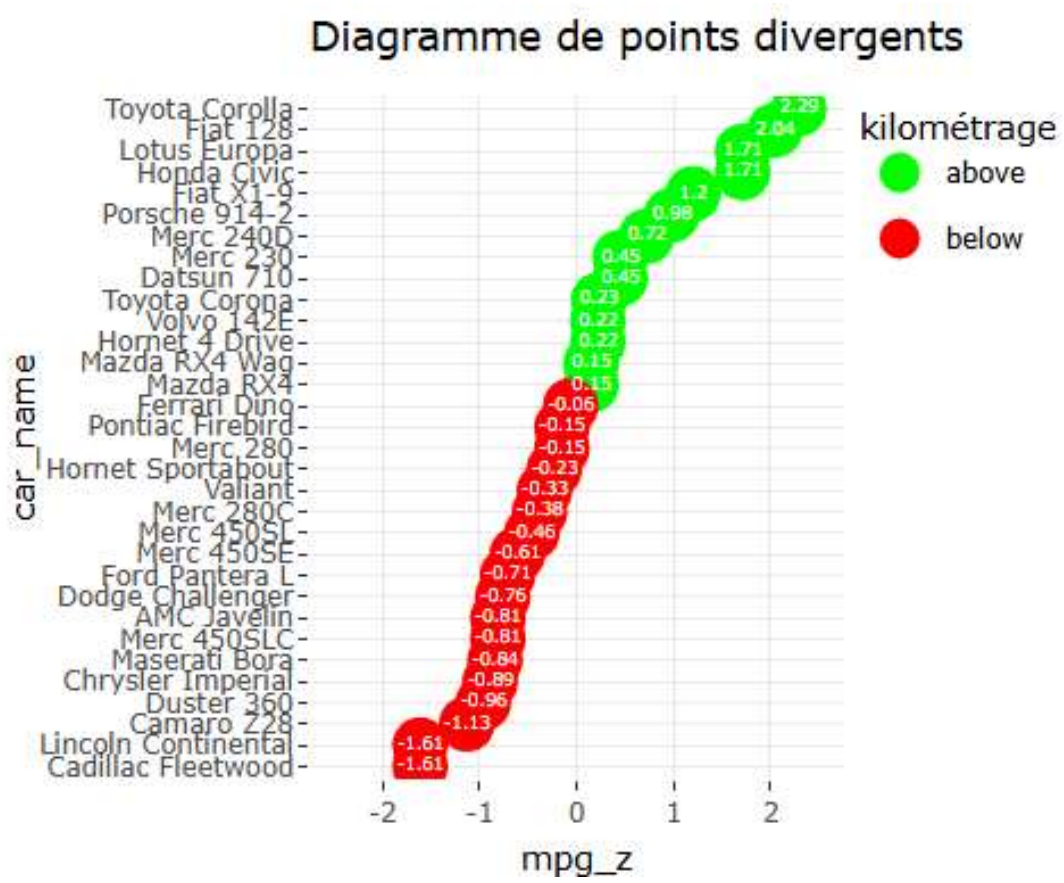
Diagramme de points divergents

```
bpd <- ggplot( mtcars, aes(car_name, mpg_z, label = mpg_z)) + geom_point(stat
= "identity", aes(col = mpg_type), size = 6) + scale_color_manual(name =
"kilométrage",
            labels = c("Above Average", "Below Average "),
            values = c("above" = "green", "below" = "red")) + geom_text(color
= "white", size = 2) +
  labs(title = "Diagramme de points divergents",
       subtitle = "kilométrage normalisé") +
  ylim(-2.5, 2.5) +
  coord_flip()
bpd
```




```
ggplotly(bpd)
```

```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc04ec91845/widgetcc01b7b1087.html screenshot completed
```



Graphiques en aires

```
data("economics", package = "ggplot2")
head(economics)
```

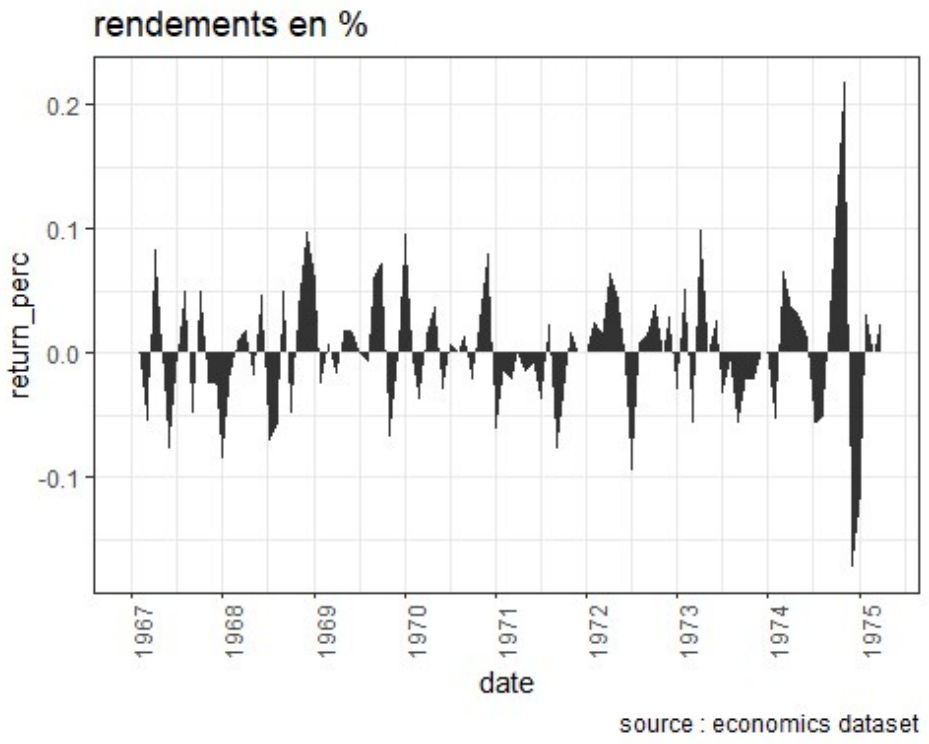
```
## # A tibble: 6 × 6
##   date         pce    pop psavert uempmed unemploy
##   <date>     <dbl>  <dbl>   <dbl>   <dbl>    <dbl>
## 1 1967-07-01  507. 198712    12.6     4.5     2944
## 2 1967-08-01  510. 198911    12.6     4.7     2945
## 3 1967-09-01  516. 199113    11.9     4.6     2958
## 4 1967-10-01  512. 199311    12.9     4.9     3143
## 5 1967-11-01  517. 199498    12.8     4.7     3066
## 6 1967-12-01  525. 199657    11.8     4.8     3018
```

```
economics$return_perc <-c(0, diff(economics$psavert)/economics$psavert[-
length(economics$psavert)])

brks <- economics$date[seq(1, length(economics$date), 12)]
lbls <- lubridate::year(economics$date[seq(1, length(economics$date), 12)])
head(economics)
```

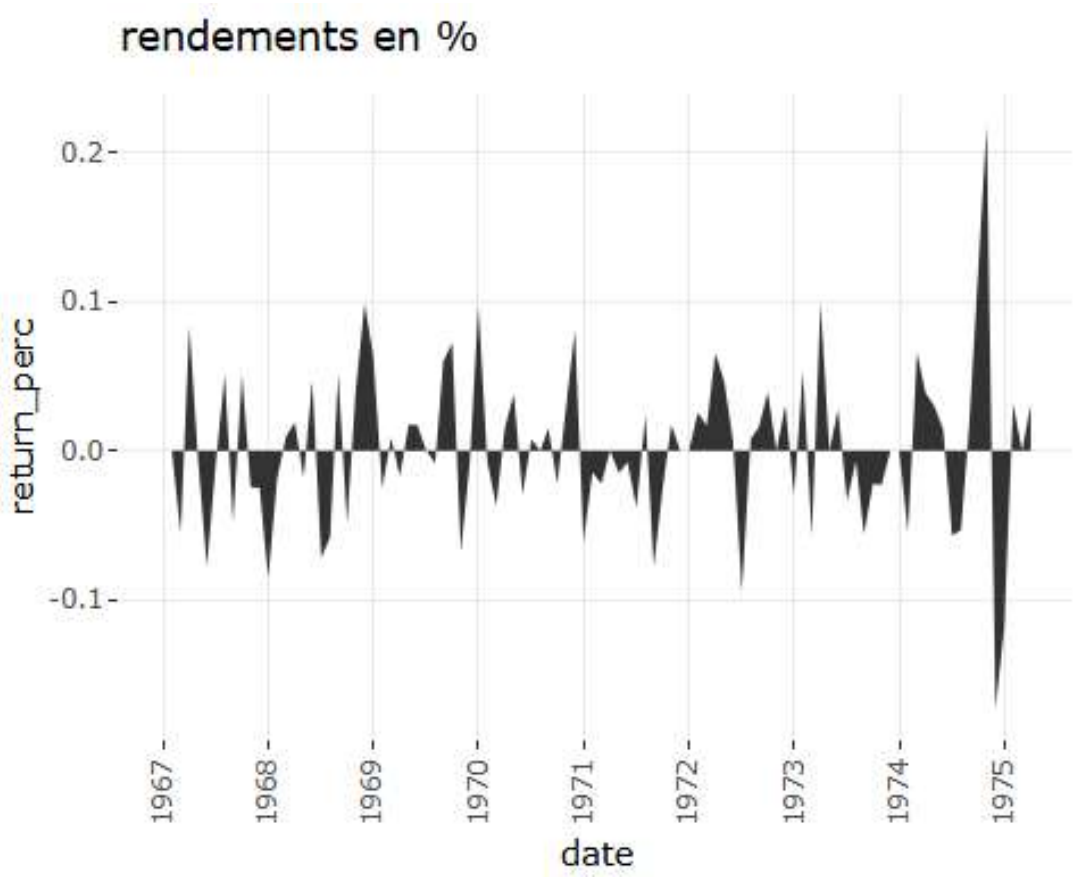
```
## # A tibble: 6 × 7
##   date         pce    pop psavert uempmed unemploy return_perc
```

```
##   <date>     <dbl>  <dbl>   <dbl>   <dbl>    <dbl>       <dbl>
## 1 1967-07-01  507. 198712    12.6     4.5     2944     0
## 2 1967-08-01  510. 198911    12.6     4.7     2945     0
## 3 1967-09-01  516. 199113    11.9     4.6     2958    -0.0556
## 4 1967-10-01  512. 199311    12.9     4.9     3143     0.0840
## 5 1967-11-01  517. 199498    12.8     4.7     3066    -0.00775
## 6 1967-12-01  525. 199657    11.8     4.8     3018    -0.0781
```

```
air_plot <- ggplot(economics[1:100, ], aes(date, return_perc)) +
  geom_area() +
  scale_x_date(breaks = brks, labels = lbls) +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title = "rendements en %",
       caption = "source : economics dataset")

air_plot
```




```
ggplotly(air_plot)

##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc0140eeb5/widgetcc
06d2f6520.html screenshot completed
```



Section 4 : Apprendre à créer des graphiques de classement

Les graphiques qui vont suivre sont appelés graphiques de classement. Ils sont utilisés pour comparer la position ou les performances de plusieurs éléments les uns par rapport aux autres.

Graphique à barres ordonnées

```
head(mpg)
```

```
## # A tibble: 6 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl
class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p
compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p
compa…
## 3 audi         a4      2    2008     4 manual(m6) f        20    31 p
```

```
compa…
## 4 audi         a4      2    2008     4 auto(av)   f        21    30 p
compa…
## 5 audi         a4      2.8  1999     6 auto(l5)   f        16    26 p
compa…
## 6 audi         a4      2.8  1999     6 manual(m5) f        18    26 p
compa…
```

```
cty_mpg <- aggregate(mpg$cty, b = list(mpg$manufacturer), FUN = mean)
head(cty_mpg)
```

```
##     Group.1        x
## 1      audi 17.61111
## 2 chevrolet 15.00000
## 3     dodge 13.13514
## 4      ford 14.00000
## 5     honda 24.44444
## 6   hyundai 18.64286
```

```
colnames(cty_mpg) <- c("make", "mileage")

head(cty_mpg)
```

```
##        make  mileage
## 1      audi 17.61111
## 2 chevrolet 15.00000
## 3     dodge 13.13514
## 4      ford 14.00000
## 5     honda 24.44444
## 6   hyundai 18.64286
```

```
cty_mpg <- cty_mpg[order(cty_mpg$mileage), ]
head(cty_mpg)
```

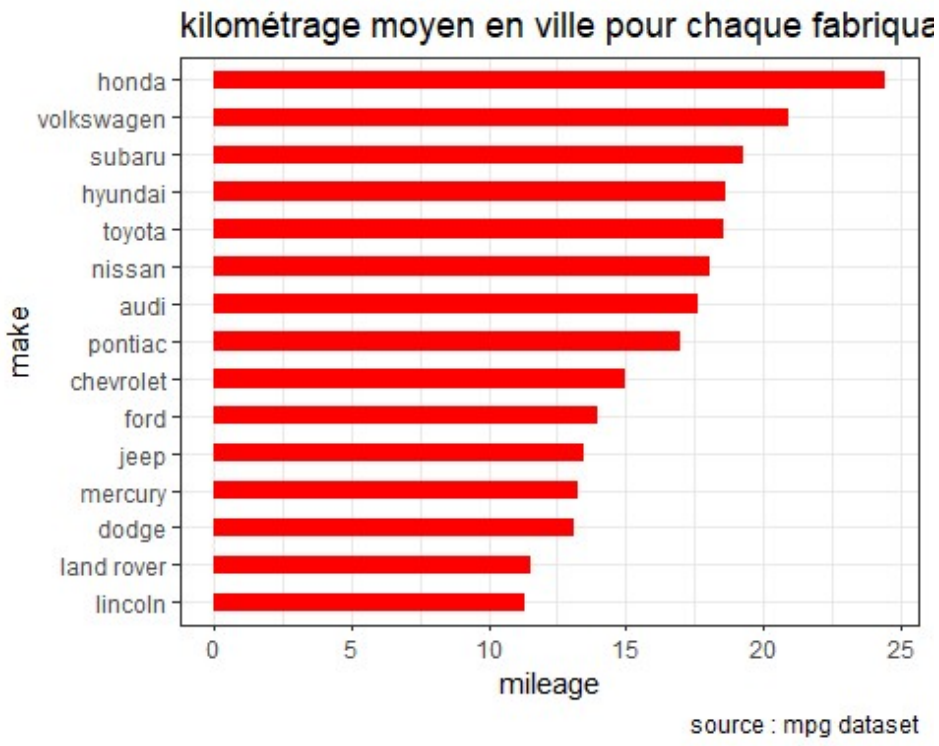
```
##          make  mileage
## 9     lincoln 11.33333
## 8  land rover 11.50000
## 3       dodge 13.13514
## 10    mercury 13.25000
## 7        jeep 13.50000
## 4        ford 14.00000
```

```
cty_mpg$make <- factor(cty_mpg$make, levels = cty_mpg$make)

head(cty_mpg)
```

```
##          make  mileage
## 9     lincoln 11.33333
## 8  land rover 11.50000
## 3       dodge 13.13514
## 10    mercury 13.25000
```

```
## 7        jeep 13.50000
## 4        ford 14.00000
```

```
gbo <- ggplot(cty_mpg, aes(x = make, y = mileage)) +
  geom_bar(stat = "identity", width = 0.5, fill = "red") +
  labs(title = "kilométrage moyen en ville pour chaque fabriquant",
       caption = "source : mpg dataset") +
coord_flip()
gbo
```

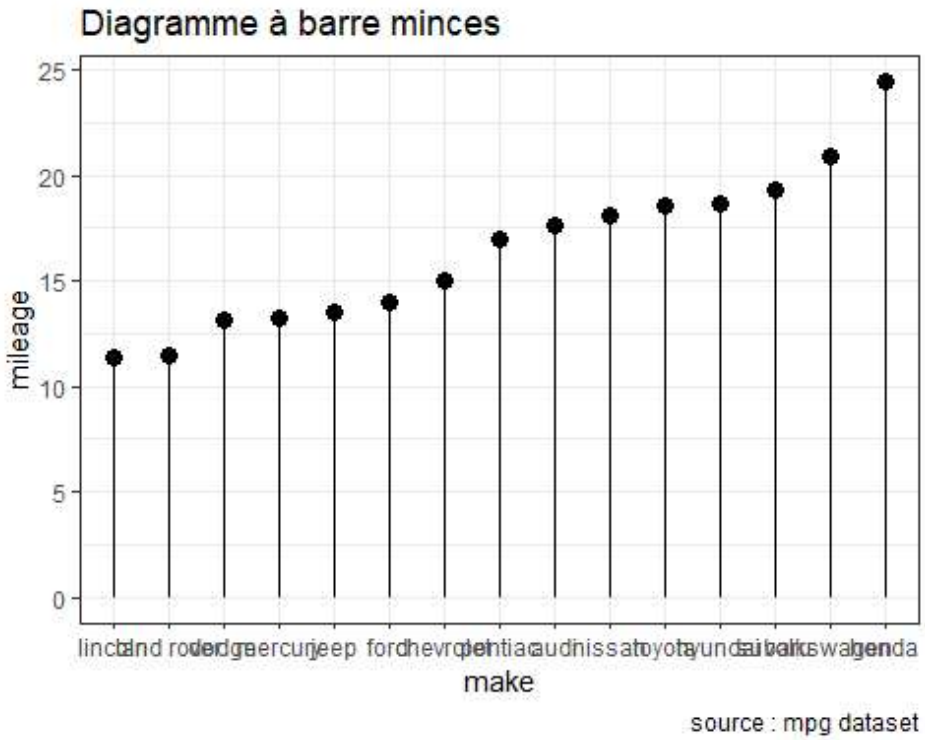



```
ggplotly(gbo)
```

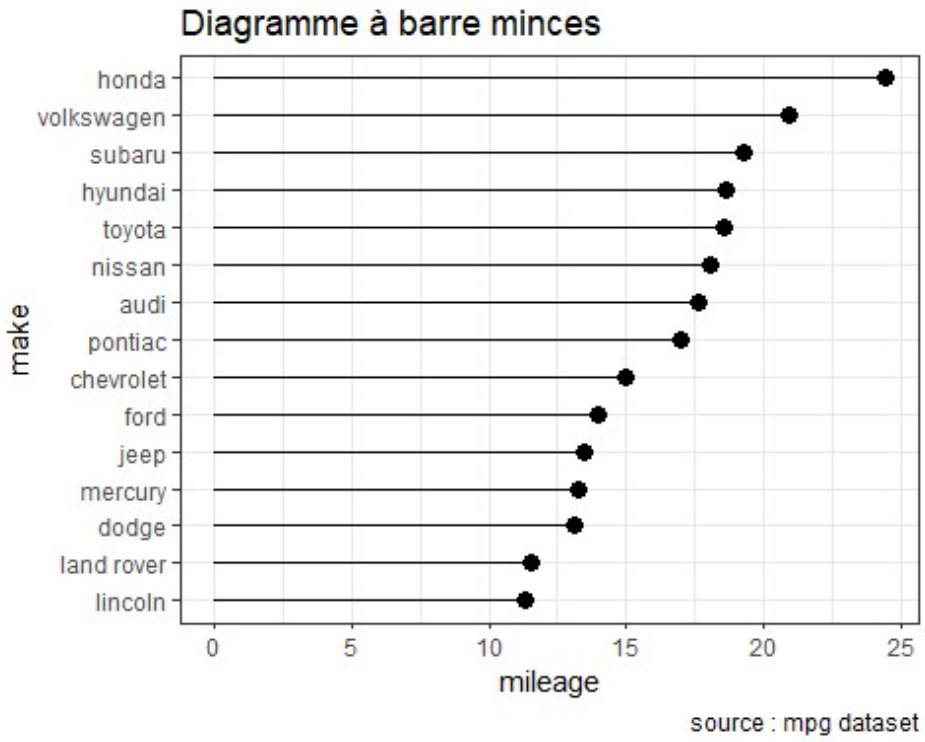
```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

Graphique à barres minces

```
dbarm <- ggplot(cty_mpg, aes(x = make, y = mileage)) +
  geom_point(size = 3) +
  geom_segment(aes(x = make,
                   xend = make,
                   y = 0,
                   yend = mileage)) +
  labs(title = "Diagramme à barre minces",
       caption = "source : mpg dataset")
dbarm
```




```
dbarm <- ggplot(cty_mpg, aes(x = make, y = mileage)) +
  geom_point(size = 3) +
  geom_segment(aes(x = make,
                   xend = make,
                   y = 0,
                   yend = mileage)) +
  labs(title = "Diagramme à barre minces",
       caption = "source : mpg dataset") +
  coord_flip()
dbarm
```




```
ggplotly(dbarm)

##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc0692d3ee8/widgetcc03a641e9d.html screenshot completed
```

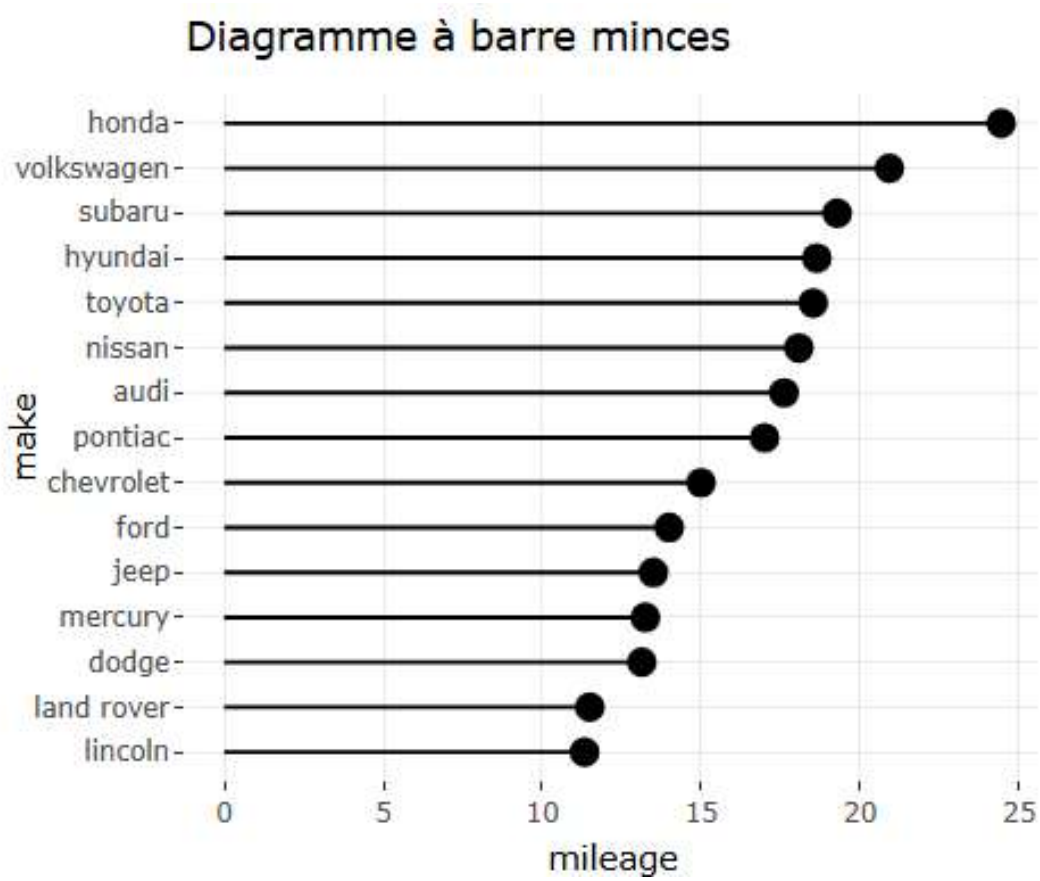
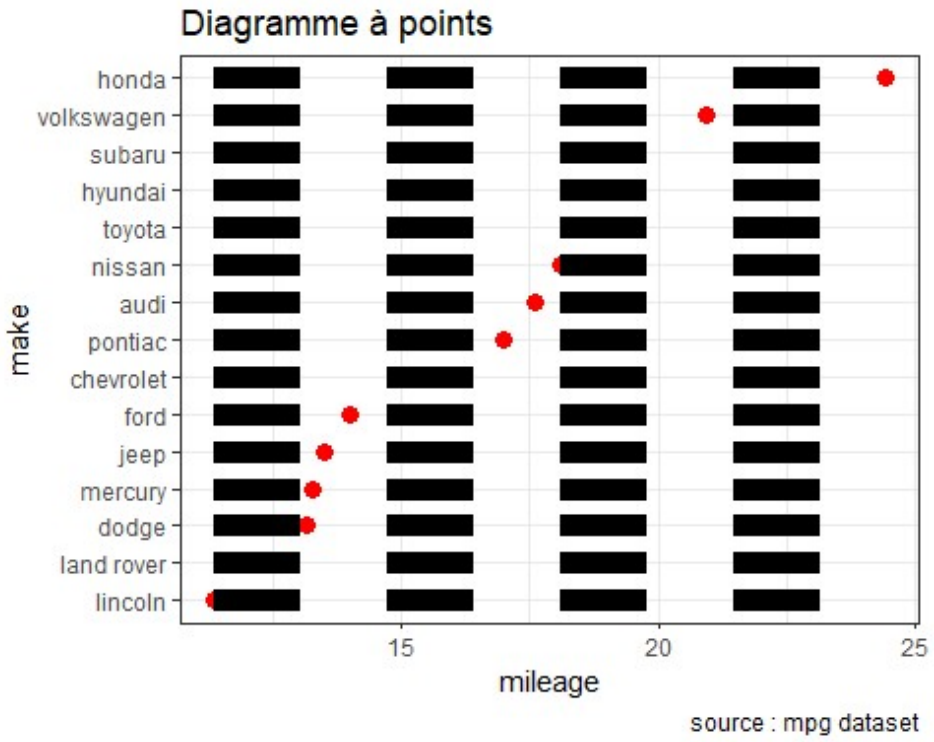


Diagramme à points (*Dot Plot*)

```
dot_plot <- ggplot(cty_mpg, aes(x = make, y = mileage)) +
  geom_point(color = "red", size = 3) +
  geom_segment(aes(x = make,
                   xend = make,
                   y = min(mileage),
                   yend = max(mileage)),
                   linetype = "dashed",
                   size = 4) +
                 labs(title = "Diagramme à points",
                      caption = "source : mpg dataset") +
  coord_flip()
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## ℹ Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
dot_plot
```




```
ggplotly(dot_plot)
```

```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

Section 5 : Apprendre à créer des graphiques de distribution

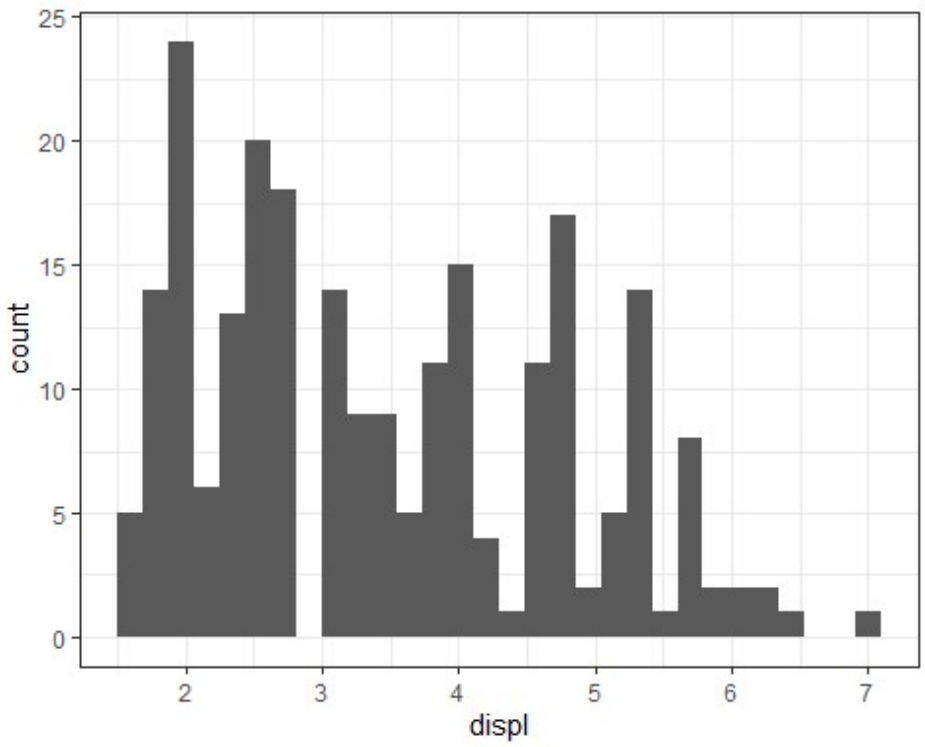
Les graphiques de distribution sont utilisés lorsque vous avez beaucoup de points de données et que vous souhaitez étudier où et comment les points de données sont distribués.

Histogramme

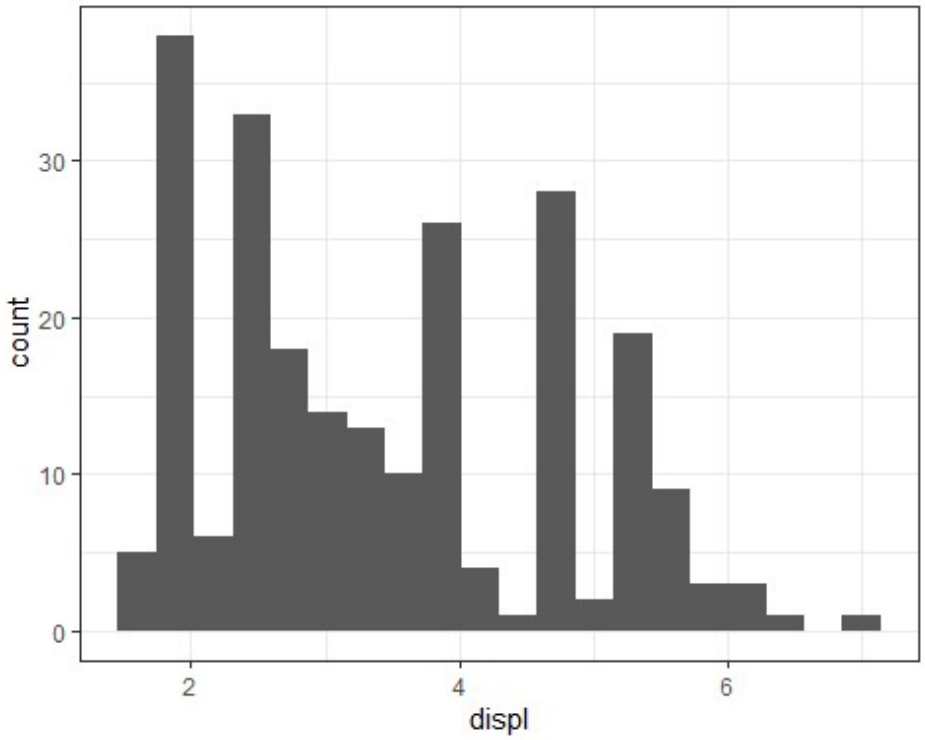
```
head(mpg, 3)
```

```
## # A tibble: 3 × 11
##   manufacturer model displ  year   cyl trans      drv     cty   hwy fl
class
##   <chr>        <chr> <dbl> <int> <int> <chr>      <chr> <int> <int> <chr>
<chr>
## 1 audi         a4      1.8  1999     4 auto(l5)   f        18    29 p
compa…
## 2 audi         a4      1.8  1999     4 manual(m5) f        21    29 p
compa…
```

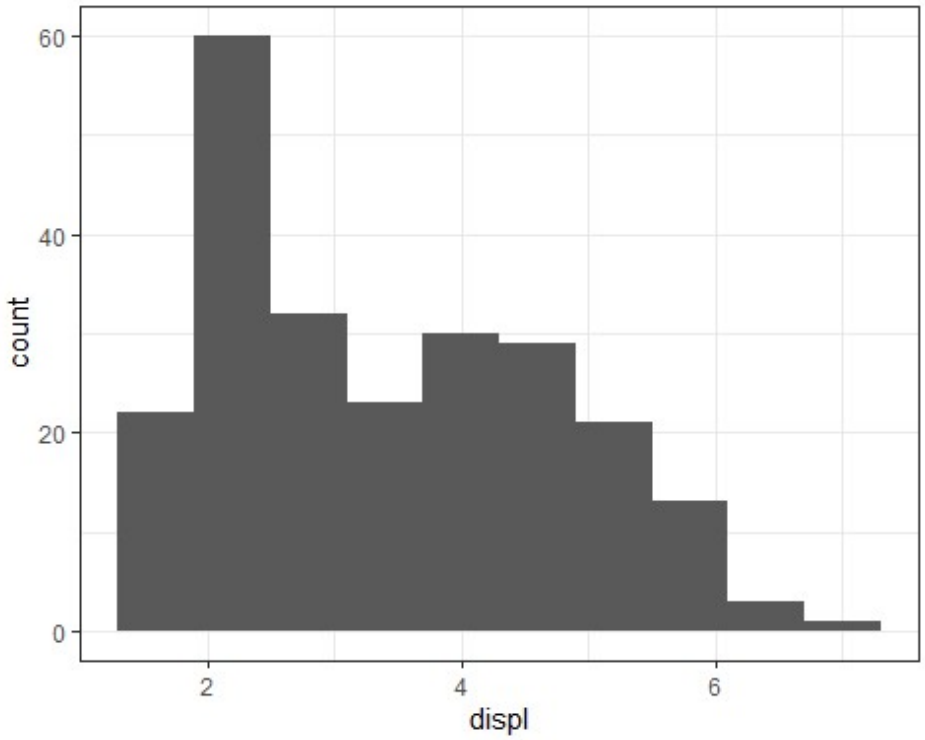
```
## 3 audi          a4      2    2008     4 manual(m6) f        20    31 p
compa…

ho <- ggplot(mpg, aes(displ)) +
  geom_histogram()
ho
```




```
ho <- ggplot(mpg, aes(displ)) +
  geom_histogram(bins = 20)
ho
```




```
ho <- ggplot(mpg, aes(displ)) +
  geom_histogram(bins = 10)
ho
```



```
ggplotly(ho)
```

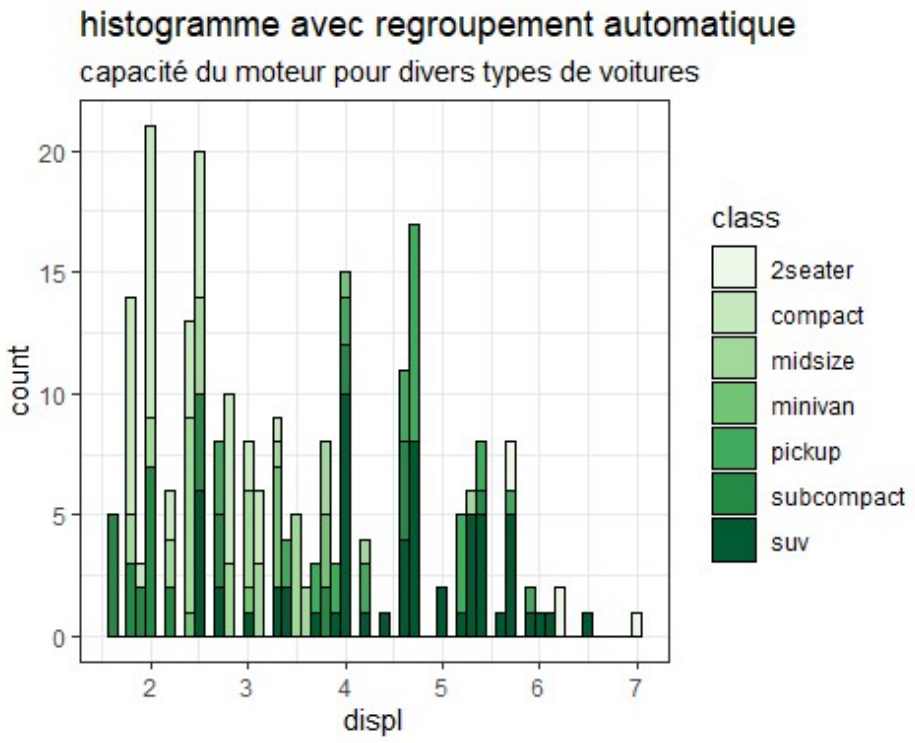
```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

```
h1 <-ggplot(mpg, aes(displ)) +
  scale_fill_brewer(palette = "spectral") +
  geom_histogram(aes(fill = class),
                 binwidth = 0.1,
                 col = "black",
                 size = 0.1) +
  labs(title = "histogramme avec regroupement automatique",
       subtitle = "capacité du moteur pour divers types de voitures")
```

```
## Warning: Unknown palette: "spectral"
```

```
## Warning in geom_histogram(aes(fill = class), binwidth = 0.1, col =
"black", :
## Ignoring unknown parameters: `size`
```

```
h1
```



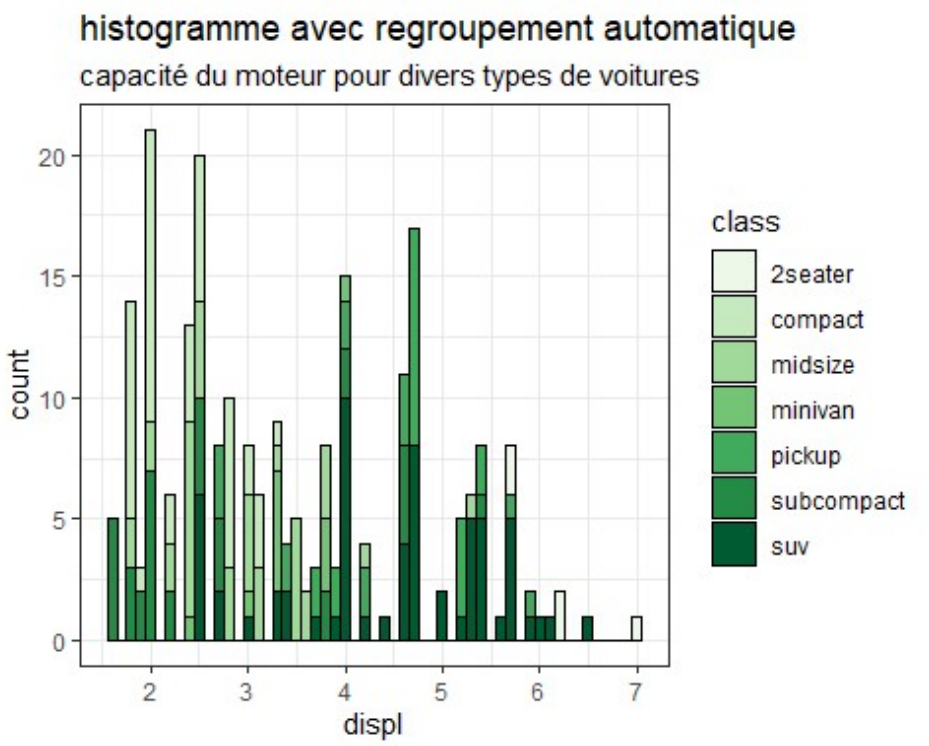

```
h2 <-ggplot(mpg, aes(displ)) +
  scale_fill_brewer(palette = "spectral") +
  geom_histogram(aes(fill = class),
                 bins = 5,
                 binwidth = 0.1,
                 col = "black",
```

```
                  size = 0.1) +
  labs(title = "histogramme avec regroupement automatique",
       subtitle = "capacité du moteur pour divers types de voitures")
```

```
## Warning: Unknown palette: "spectral"
```

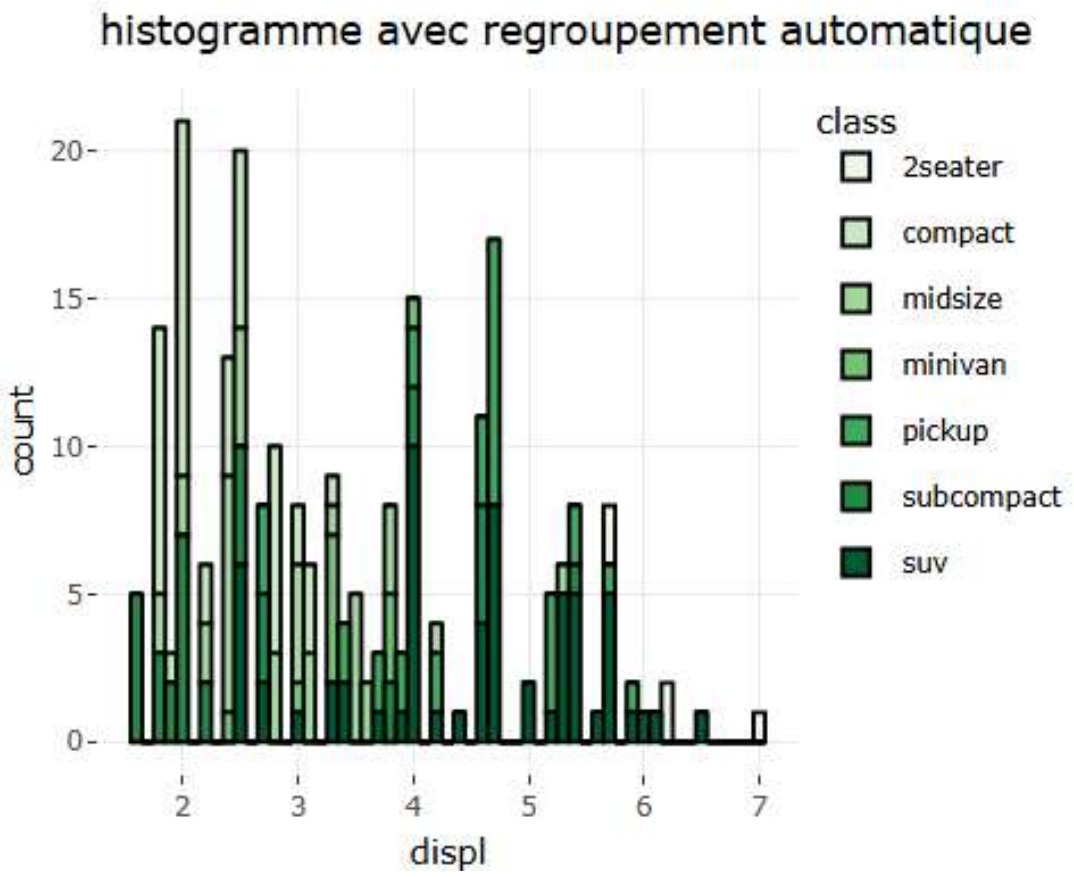
```
## Warning in geom_histogram(aes(fill = class), bins = 5, binwidth = 0.1, col
=
## "black", : Ignoring unknown parameters: `size`
```

```
h2
```

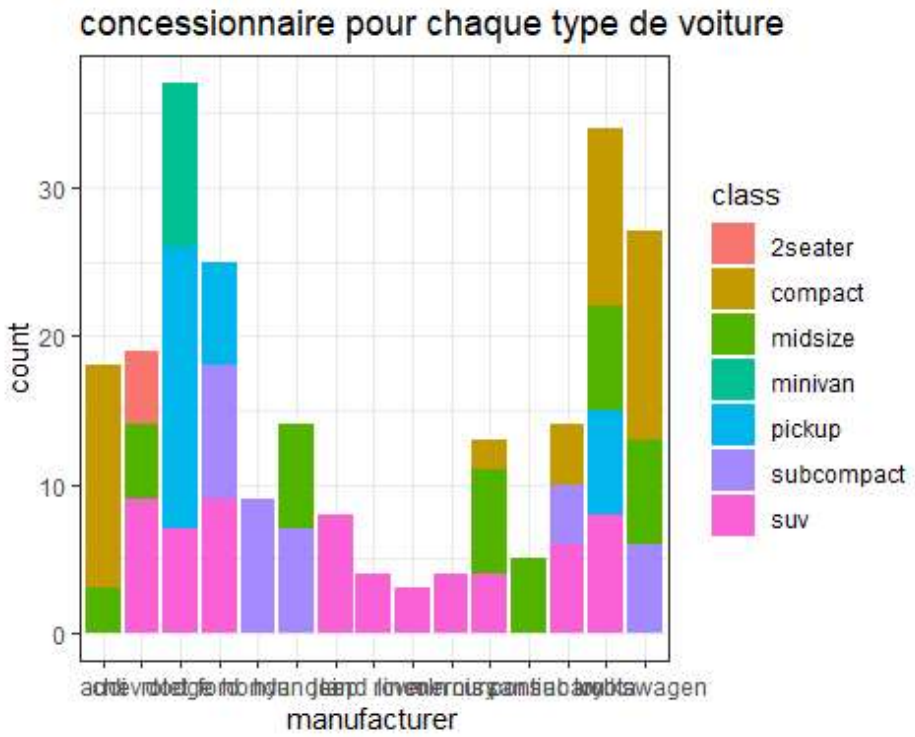



```
ggplotly(h2)
```

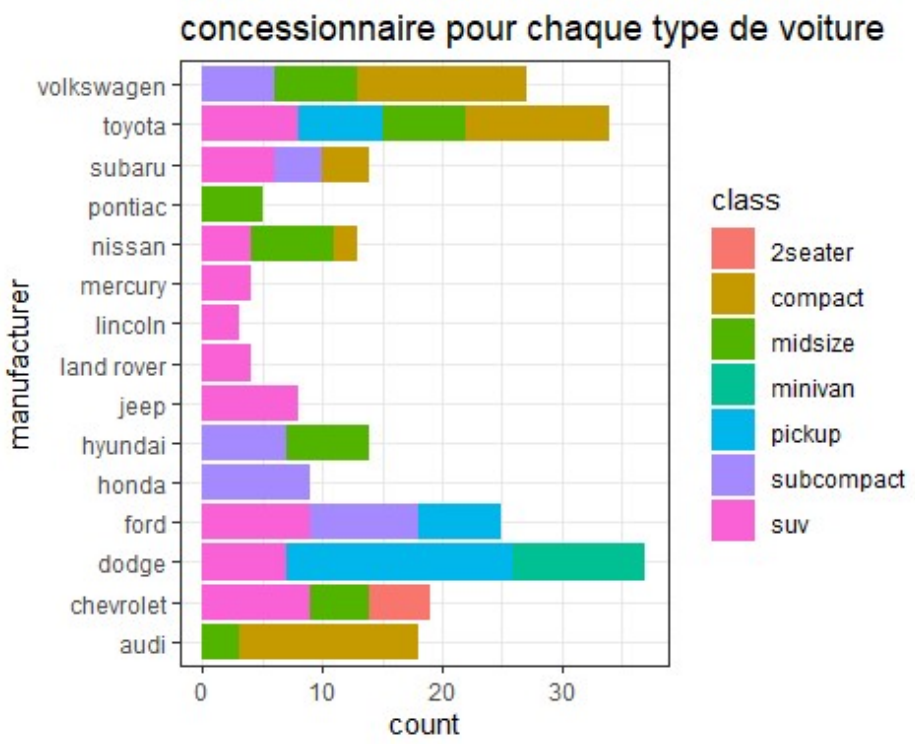
```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc0506321a2/widgetc
c06fe570ca.html screenshot completed
```




```
hc <- ggplot(mpg, aes(manufacturer)) +
  geom_bar(aes(fill = class)) +
  labs(title = "concessionnaire pour chaque type de voiture")
hc
```




```
hc <- ggplot(mpg, aes(manufacturer)) +
  geom_bar(aes(fill = class)) +
  labs(title = "concessionnaire pour chaque type de voiture") +
  coord_flip()
hc
```

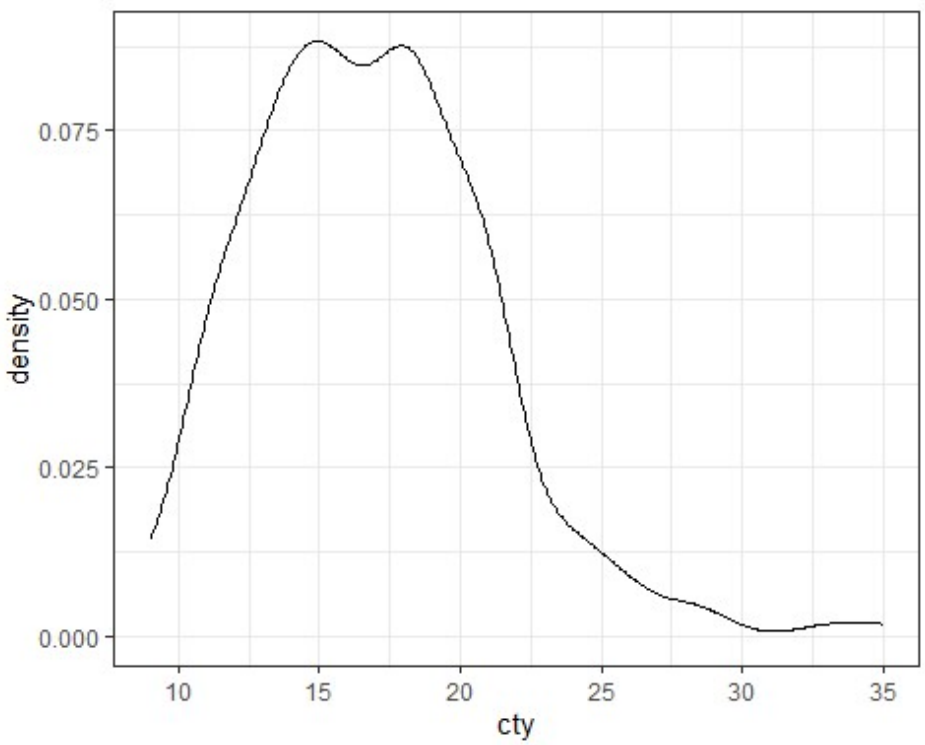



```
ggplotly(hc)
```

```
## Error: Chromote: timed out waiting for response to command
Runtime.evaluate
```

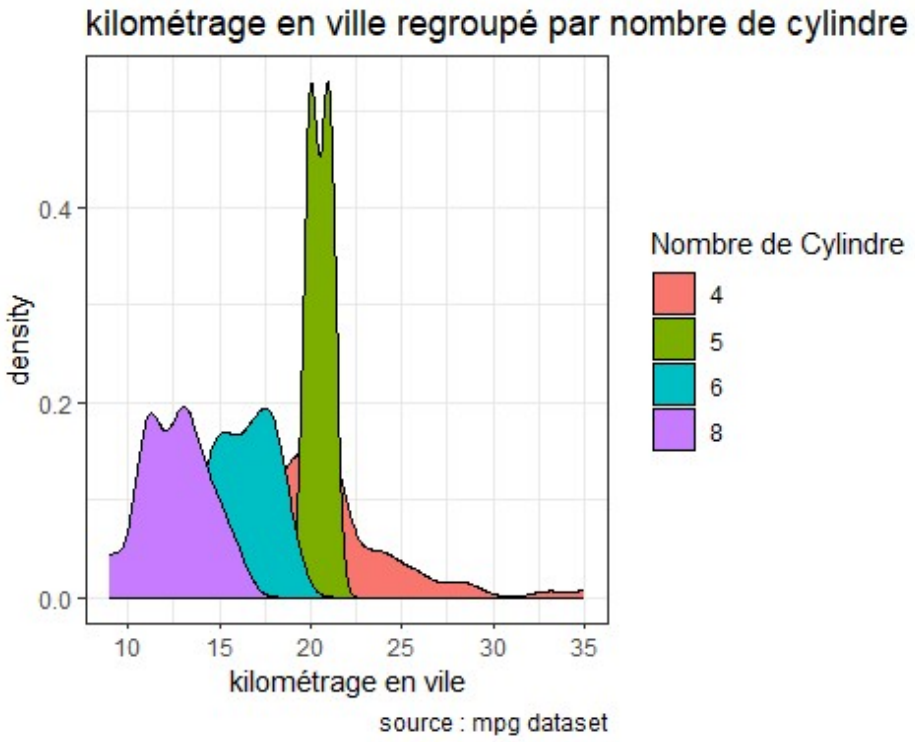
Graphique de densité

```
density_plot <- ggplot(mpg, aes(cty)) +
  geom_density()
density_plot
```



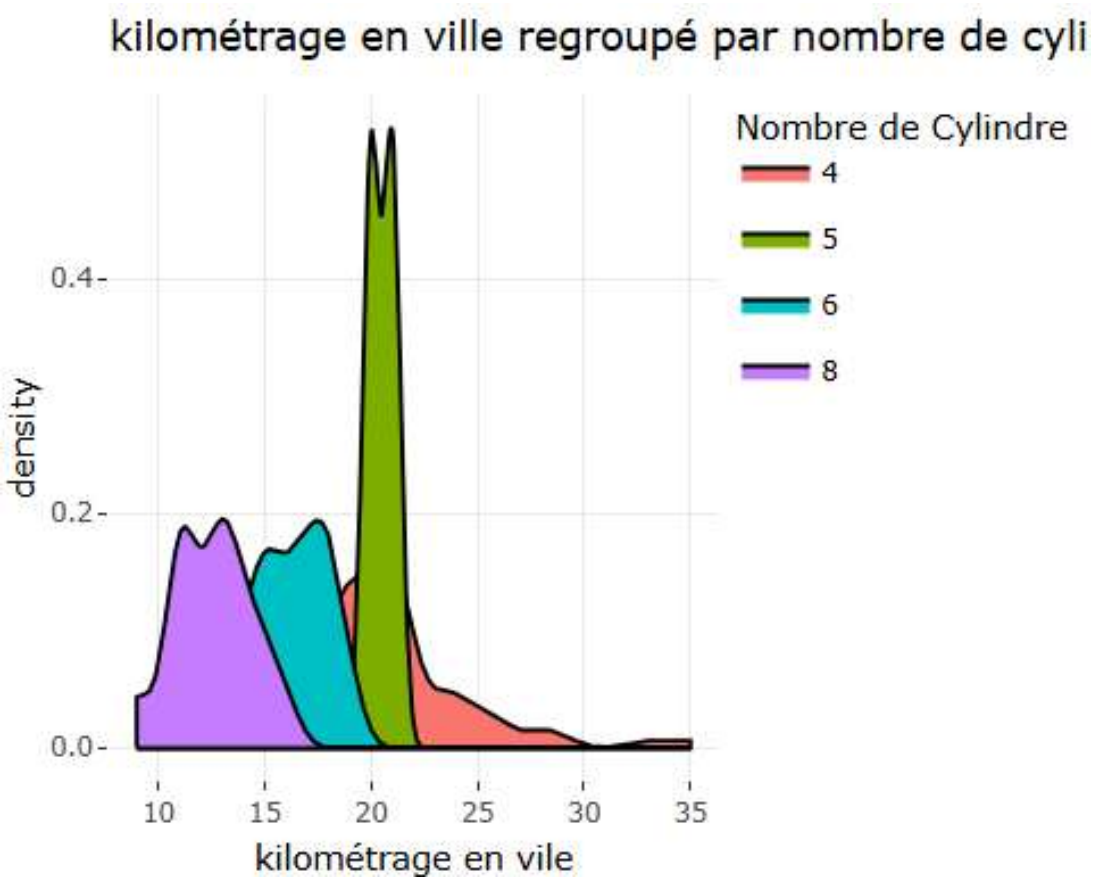
discretisation selon le nombre de cylindre

```
density_plot <- ggplot(mpg, aes(cty)) +
  geom_density(aes(fill = factor(cyl)), alpha = 1) +
labs(title = "kilométrage en ville regroupé par nombre de cylindre",
       caption = " source : mpg dataset",
       x = "kilométrage en vile",
       fill = "Nombre de Cylindre")
density_plot
```



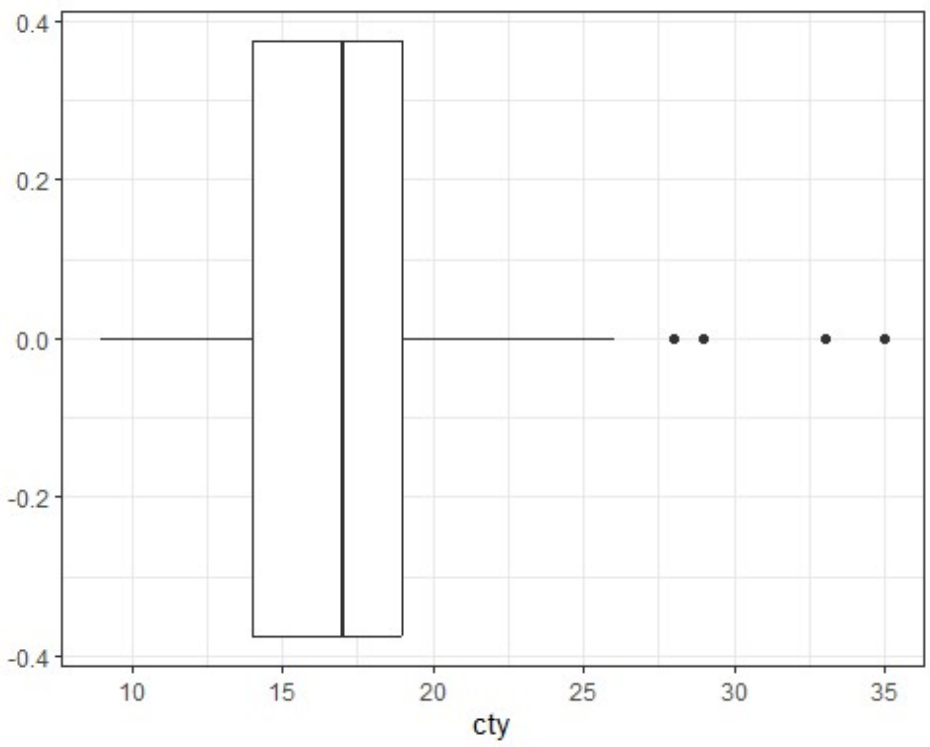

```
ggplotly(density_plot)
```

```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc0727c69cd/widgetcc05ef853cf.html screenshot completed
```

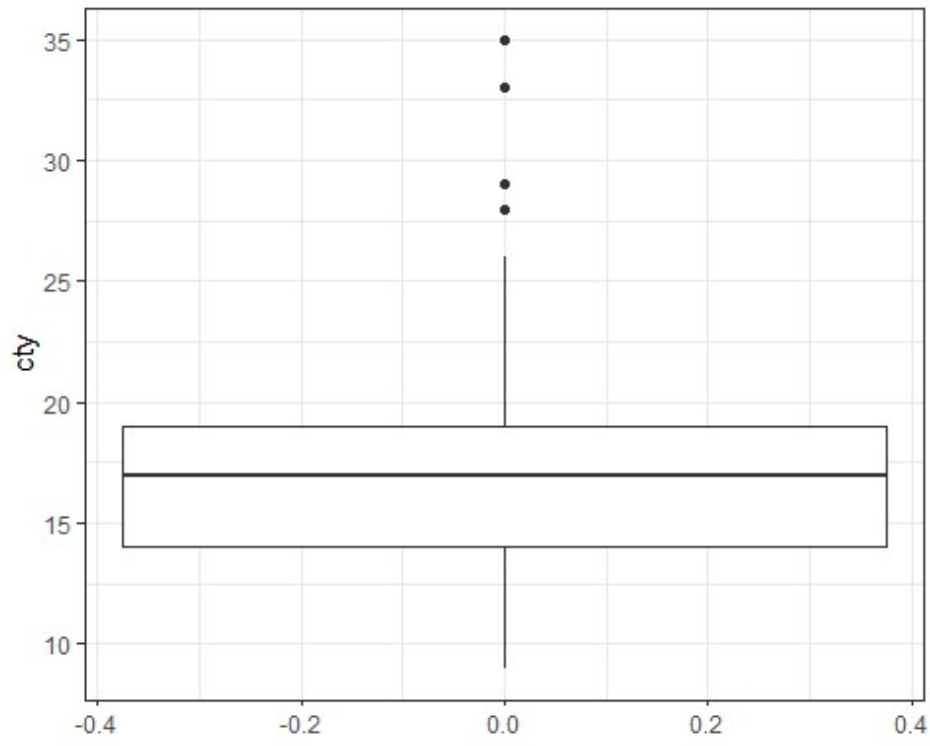


Boîte à moustache

```
box_plot <- ggplot(mpg, aes(cty)) +
  geom_boxplot()
box_plot
```

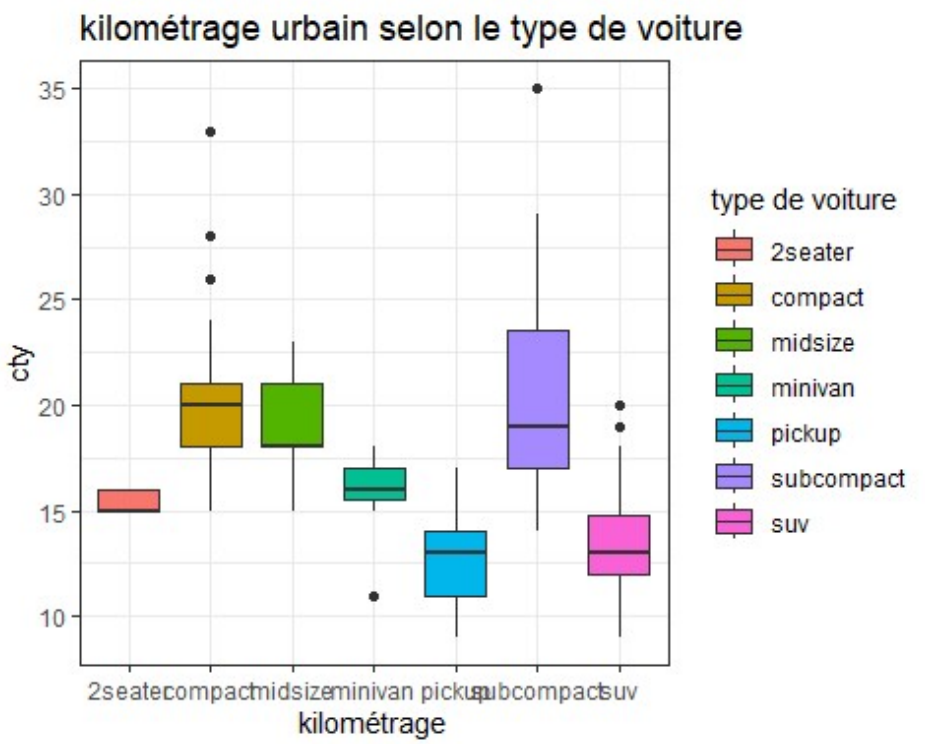



```
box_plot <- ggplot(mpg, aes(y = cty)) +
  geom_boxplot()
box_plot
```



discretisation par classe pour chaque type de voiture

```
box_plot <- ggplot(mpg, aes(x = class, y = cty, fill = factor(class))) +
  geom_boxplot() +
  labs(title = "kilométrage urbain selon le type de voiture",
       x = "kilométrage",
       fill = "type de voiture")
box_plot
```




```
ggplotly(box_plot)
```

```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc04c4045e9/widgetcc036311443.html screenshot completed
```

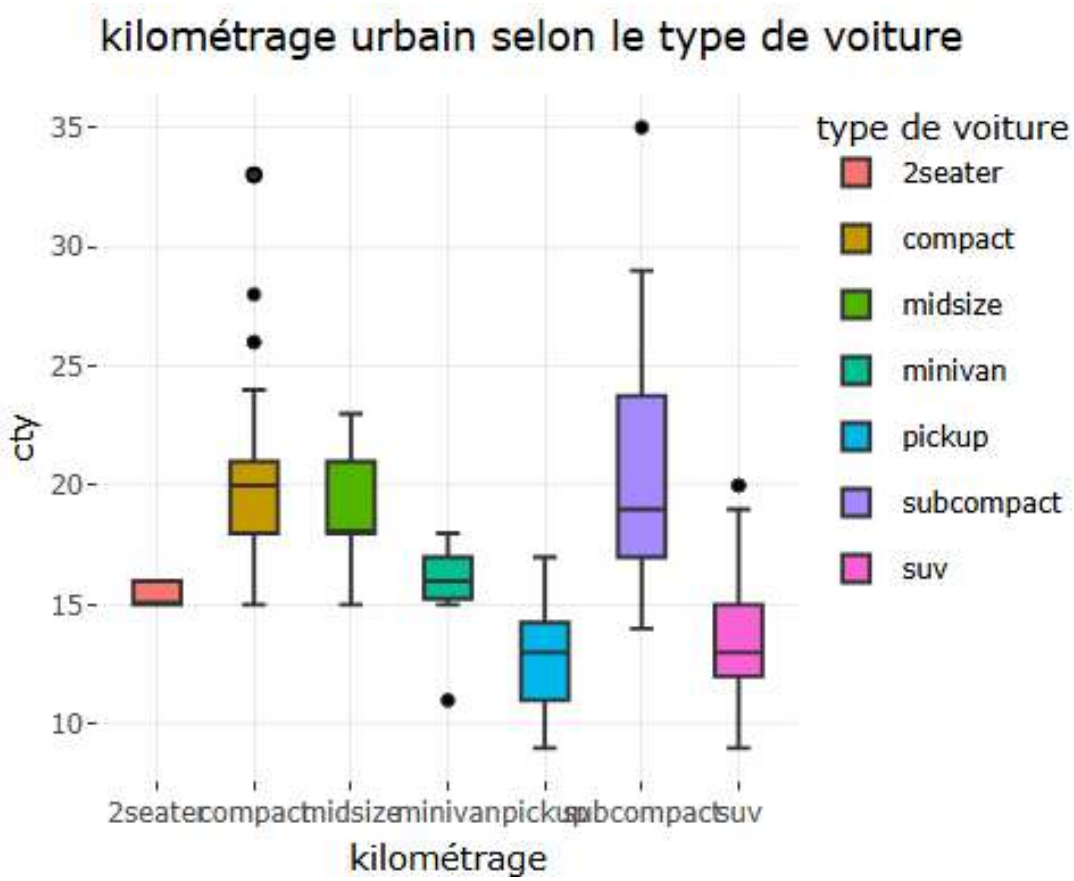
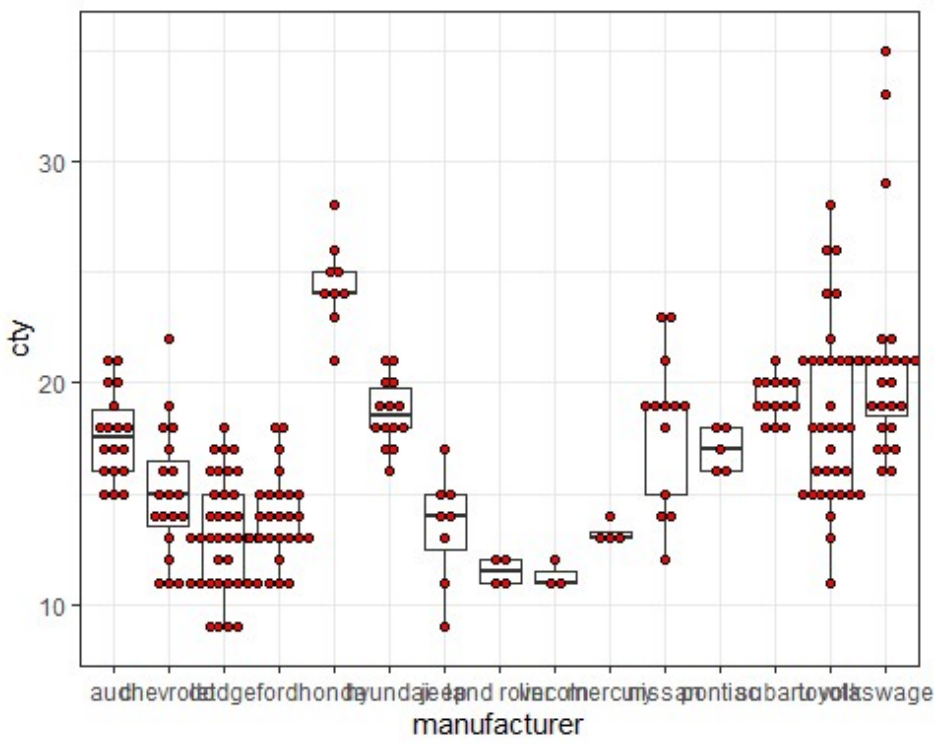


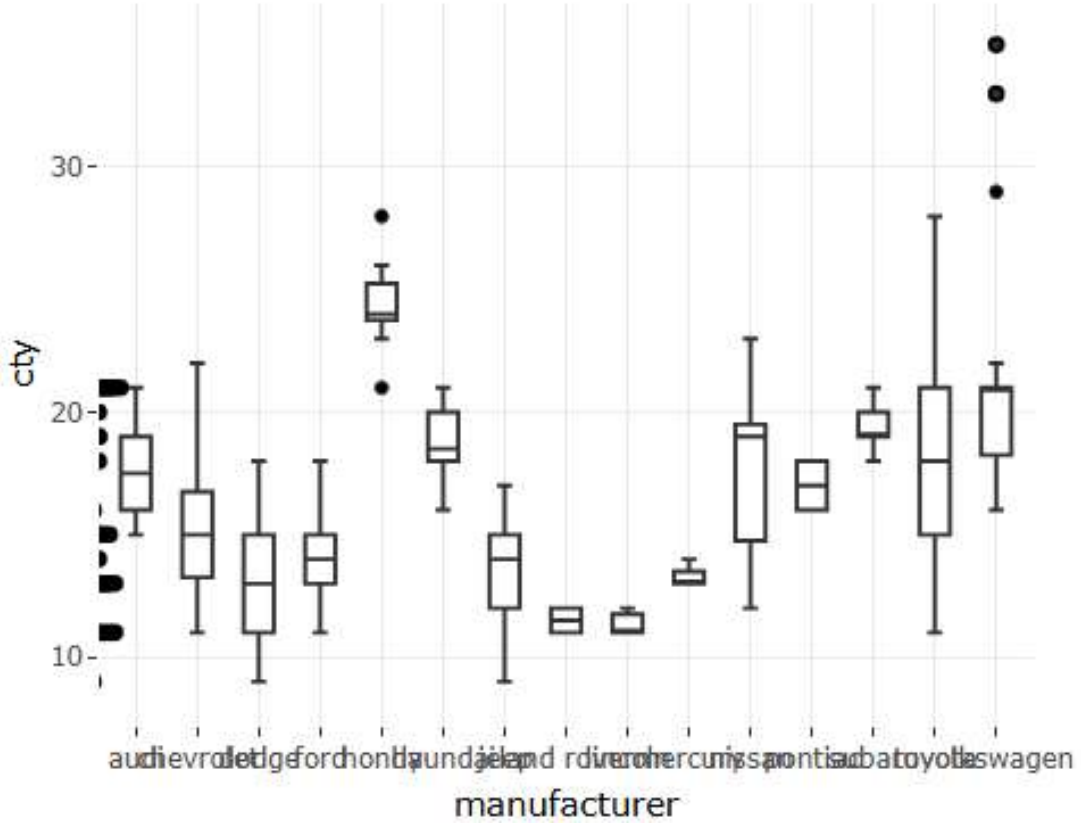
Diagramme à points + Boîte à moustache

```
dot_plot <- ggplot(mpg, aes(manufacturer, cty)) +
  geom_boxplot() +
  geom_dotplot(fill = "red", binaxis = 'y', stackdir = "center", dotsize =
0.5)
dot_plot
```



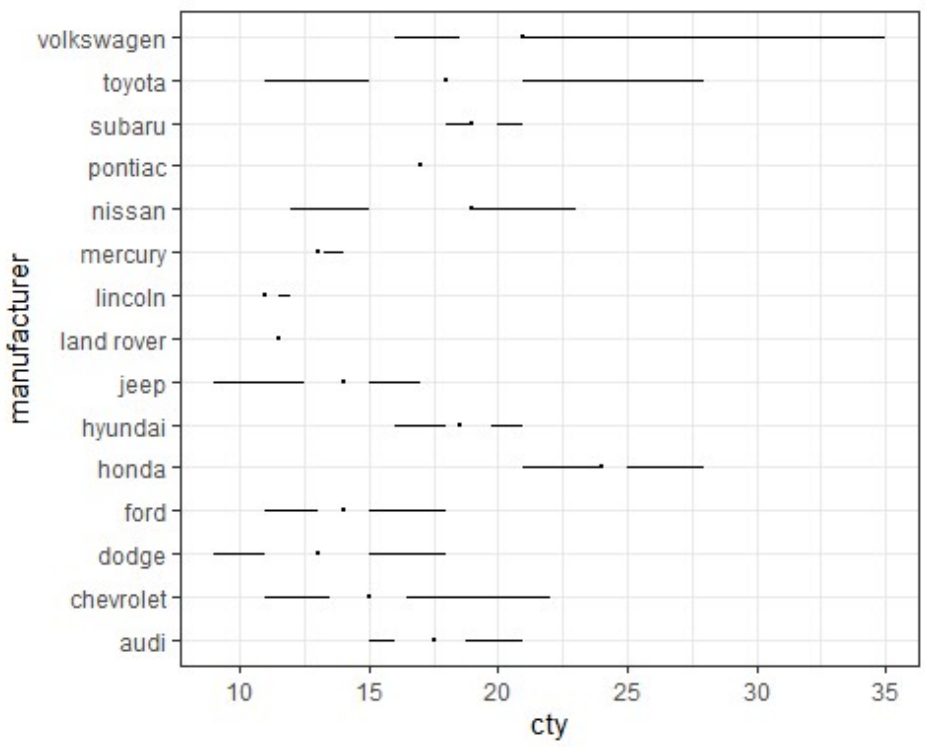

```
ggplotly(dot_plot)
```

```
## Bin width defaults to 1/30 of the range of the data. Pick better value
with
## `binwidth`.
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc030ee60cd/widgetc
c074417163.html screenshot completed
```

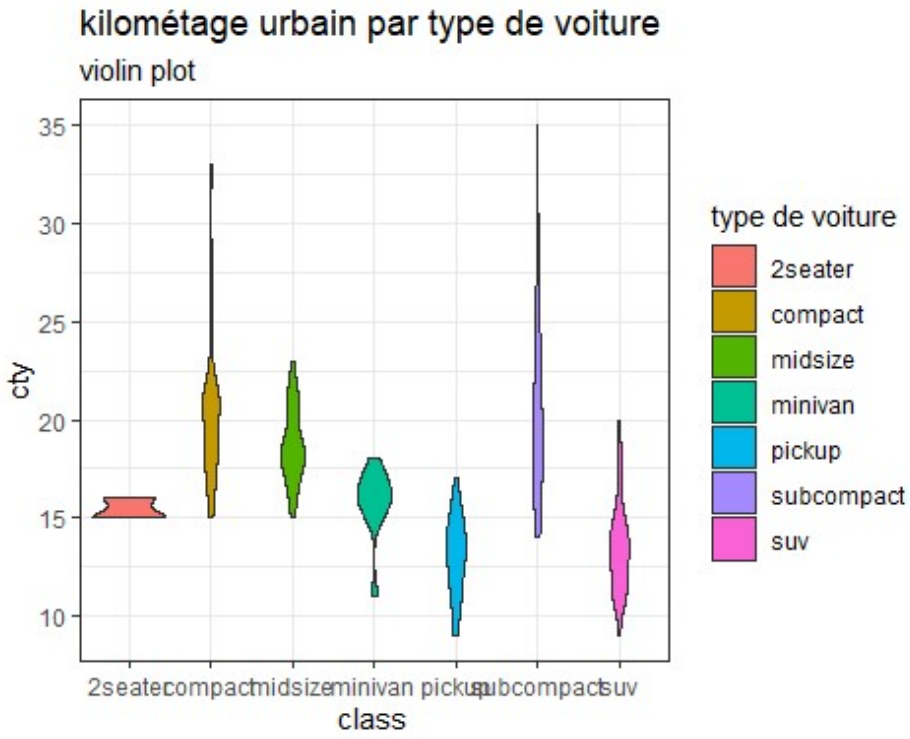


Boîte à moustache de tufte

```
t_box <- ggplot(mpg, aes(manufacturer, cty)) +
  geom_tufteboxplot() +
  coord_flip()
t_box
```

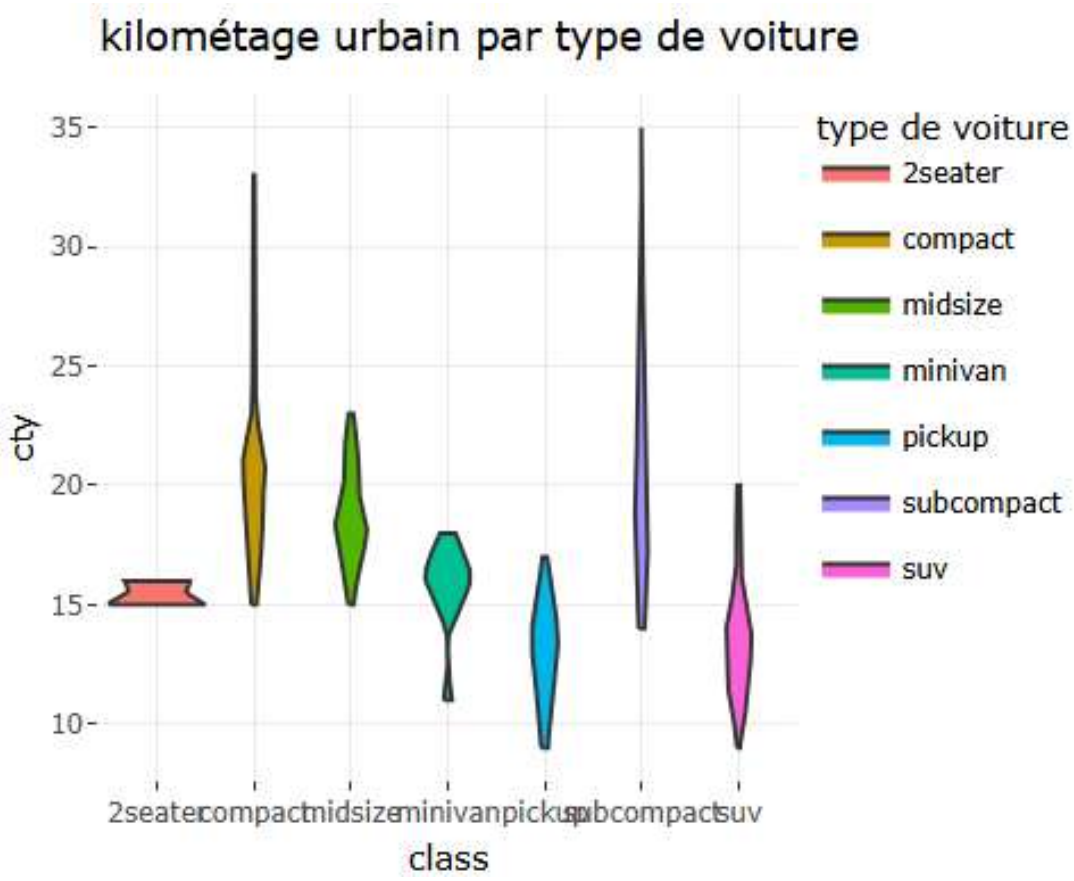



```
## graphique de violon

violin_plot <- ggplot(mpg, aes(class,cty)) +
  geom_violin(aes(fill =(class))) +
  labs(title = "kilométage urbain par type de voiture", subtitle = "violin
plot",
       fill = "type de voiture")
violin_plot
```




```
ggplotly(violin_plot)
```

```
##
file:///C:/Users/lenovo/AppData/Local/Temp/RtmpgBucVo/filecc0794b3875/widgetcc014876e52.html screenshot completed
```



Section 6 : Apprendre à créer des graphiques de composition

Diagramme circulaire

```
df <- as.data.frame(table(mpg$class))
df

##         Var1 Freq
## 1    2seater    5
## 2    compact   47
## 3    midsize   41
## 4    minivan   11
## 5     pickup   33
## 6 subcompact   35
## 7        suv   62

colnames(df) <- c('class', 'Frequence')
df

##        class Frequence
## 1    2seater         5
## 2    compact        47
## 3    midsize        41
## 4    minivan        11
```

```
## 5     pickup        33
## 6 subcompact        35
## 7        suv        62
```

```r
pie <- ggplot(df, aes(x = " ", y = Frequence, fill = factor(class))) +
  geom_bar(stat = "identity") +
  coord_polar(theta = "y", start = 0) +
  labs(title = "Fréquence du typt de voiture",
       fill = "type de voiture")
pie
```