Support de cours R / Rstudio / Rmarkdown

```
##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
## Loading required package: usethis
```

1-Premières manip

Je crée un objet data avec 2 colonnes : année et captures I initialize a data object with two columns (year) and (catches)

```
rm(list=ls())

data<-data.frame(
    year=c(2010,2011,2012,2013,2014),                    catches=c(20,50,60,70,70)
    )
```

J'affiche cet objet :

```
head(data)
```

```
##   year catches
## 1 2010      20
## 2 2011      50
## 3 2012      60
## 4 2013      70
## 5 2014      70
```

Quelques manipulations mean, sum, stdev

```
sum(data$catches)
```

```
## [1] 270
```

```
mean(data$catches)
```

```
## [1] 54
```

```
var(data$catches)
```

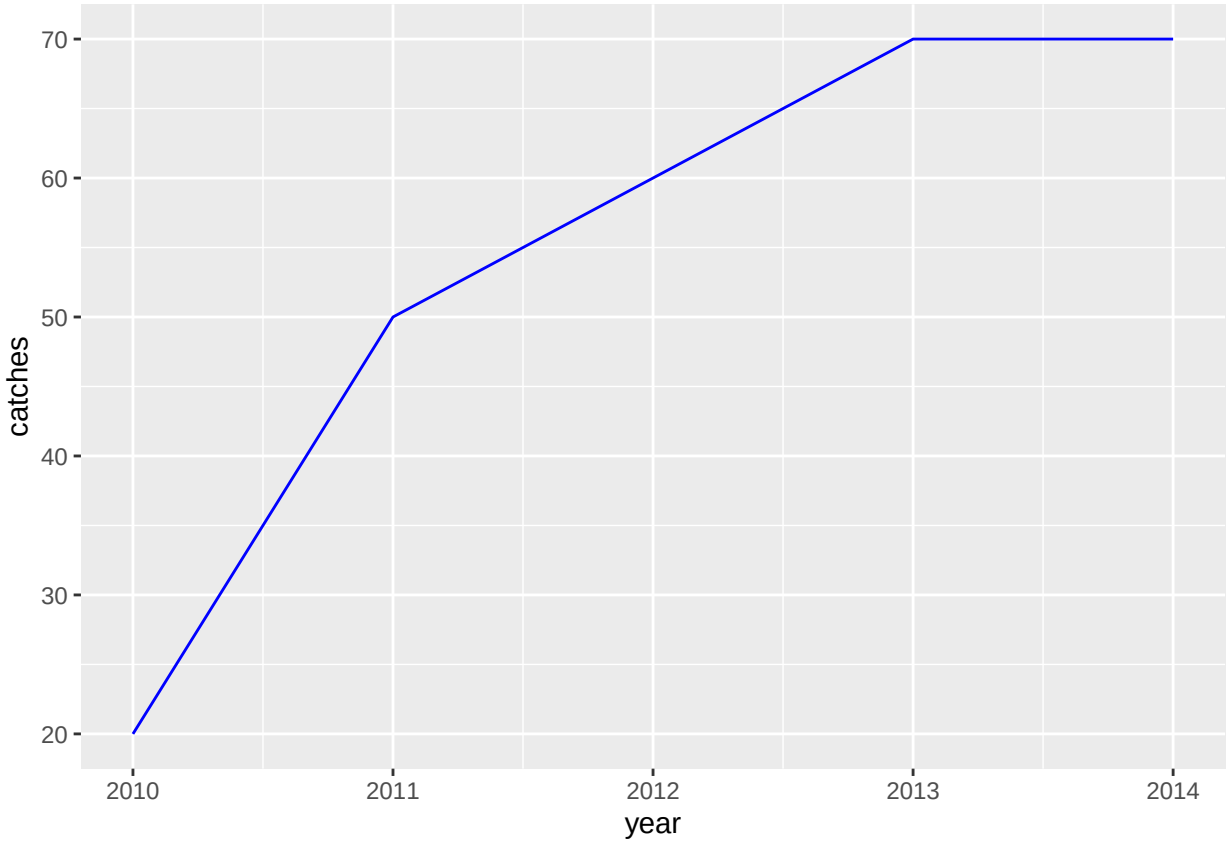
```
## [1] 430
```

```
sd(data$catches)
```

```
## [1] 20.73644
```

Un graphique

```
library(ggplot2)
#Rappel si la librairie n'est pas disponible, utilisez la commande install.packages("ggplot2")

ggplot(data)+ geom_line(aes(x=year,y=catches),color='Blue')
```



2- Allons chercher des données

```
dir()
```

```
##  [1] "index_files"    "index.html"     "index.nb.html"  "index.pdf"
##  [5] "index.Rmd"      "resume.xml"     "tableau_pa.rda" "tableau_pi.csv"
##  [9] "tableau_pi.rda" "test.pdf"       "test.rmd"
```

```
setwd("~/BAS_jerome/34")

dir()
```

```
##  [1] "index_files"    "index.html"     "index.nb.html"  "index.pdf"
##  [5] "index.Rmd"      "resume.xml"     "tableau_pa.rda" "tableau_pi.csv"
##  [9] "tableau_pi.rda" "test.pdf"       "test.rmd"
```

```
#Je charge un fichier et celui-ci cré directement les objets R
load("tableau_pa.rda")
```

```
#Je charge un fichier et le résultat du chargement est utilisé pour créer un nouvel objet
tableau_pi <- read.csv2("tableau_pi.csv")
```

3- Jouons avec le tableau

Premier traitement, on regroupe les captures par annees et on fait un graph

```
library(dplyr)

summary(tableau_pa)
```

```
##   code_pays             annee           mois            zone
##  Length:67218       Min.   :2000   Min.   : 1.000   Length:67218
##  Class :character   1st Qu.:2004   1st Qu.: 3.000   Class :character
##  Mode  :character   Median :2008   Median : 6.000   Mode  :character
##                     Mean   :2009   Mean   : 6.362
##                     3rd Qu.:2012   3rd Qu.:10.000
##                     Max.   :2019   Max.   :12.000
##   port_site         type_pirogue       motorisation    engin_peche
##  Length:67218       Length:67218       Mode :logical   Length:67218
##  Class :character   Class :character   FALSE:18932     Class :character
##  Mode  :character   Mode  :character   TRUE :48286     Mode  :character
##
##
##
##     espece             captures          nb_sorties nb_jour_peche
##  Length:67218       Min.   :      70   Min.   :0    Min.   :   10
##  Class :character   1st Qu.:   14737   1st Qu.:0    1st Qu.:  328
##  Mode  :character   Median :   58816   Median :0    Median : 1020
##                     Mean   :  286799   Mean   :0    Mean   : 2533
##                     3rd Qu.:  203809   3rd Qu.:0    3rd Qu.: 3072
##                     Max.   :50010918   Max.   :0    Max.   :35640
##     zone2              saison          engin_peche2
##  Length:67218       Length:67218       Length:67218
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
names(tableau_pa)
```

```
##  [1] "code_pays"     "annee"         "mois"          "zone"
##  [5] "port_site"     "type_pirogue"  "motorisation"  "engin_peche"
##  [9] "espece"        "captures"      "nb_sorties"    "nb_jour_peche"
## [13] "zone2"         "saison"        "engin_peche2"
```

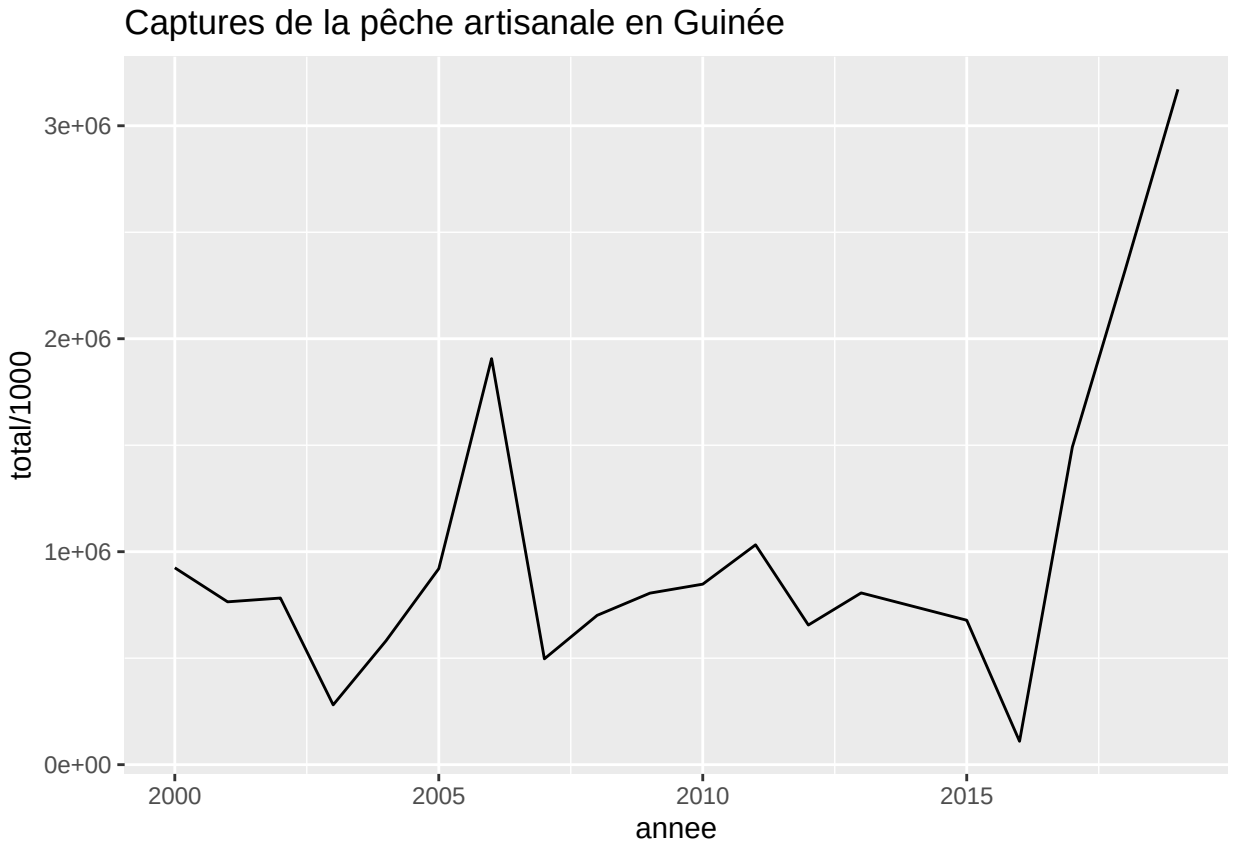
```
captures_total <- tableau_pa %>% group_by (annee) %>% dplyr::summarise(total=sum(captures))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(captures_total)
```

```
## # A tibble: 6 x 2
##   annee      total
##   <int>      <dbl>
## 1  2000 924417680.
## 2  2001 764527099.
## 3  2002 782421295.
## 4  2003 280780383.
## 5  2004 581391187.
## 6  2005 920381500.
```

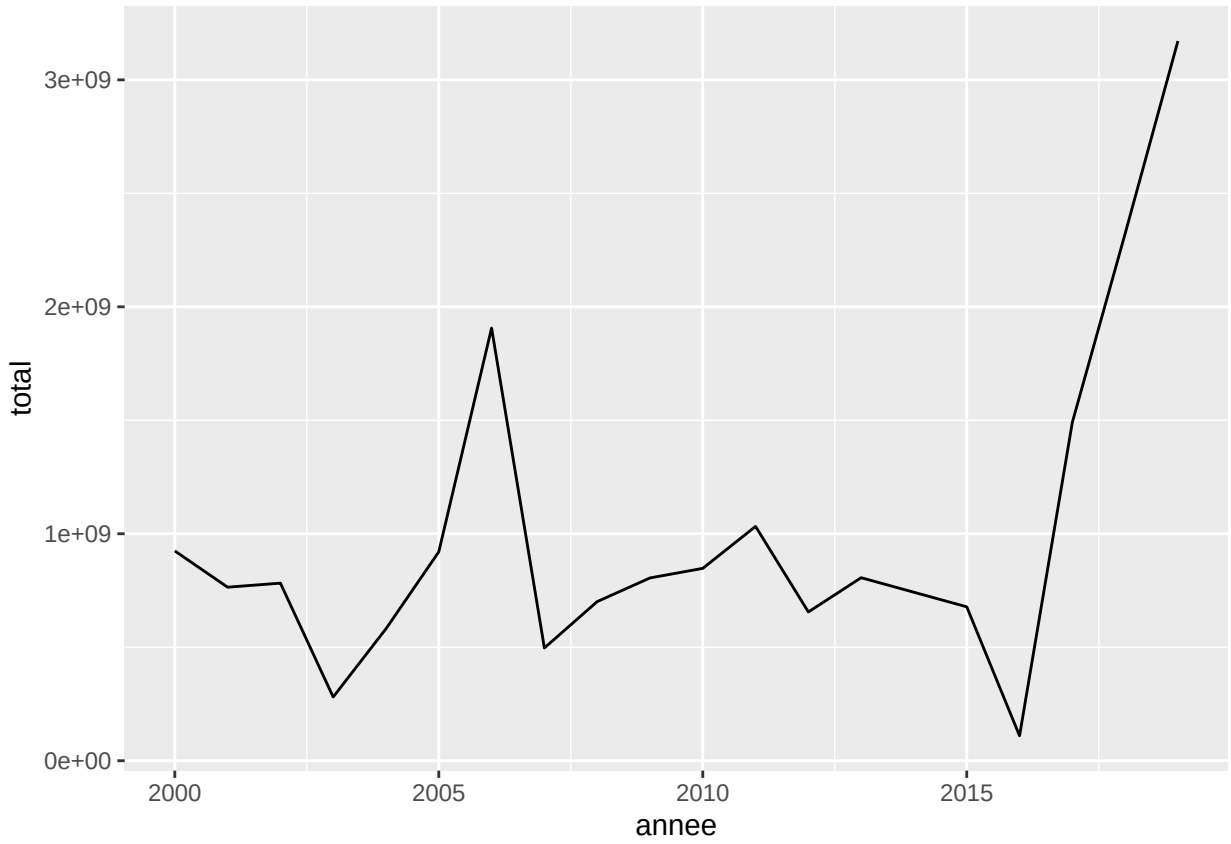
```
ggplot(captures_total)+ geom_line(aes(x=annee,y=total/1000)) +
  ggtitle("Captures de la pêche artisanale en Guinée")
```



On rajoute du filtre et un enchainement

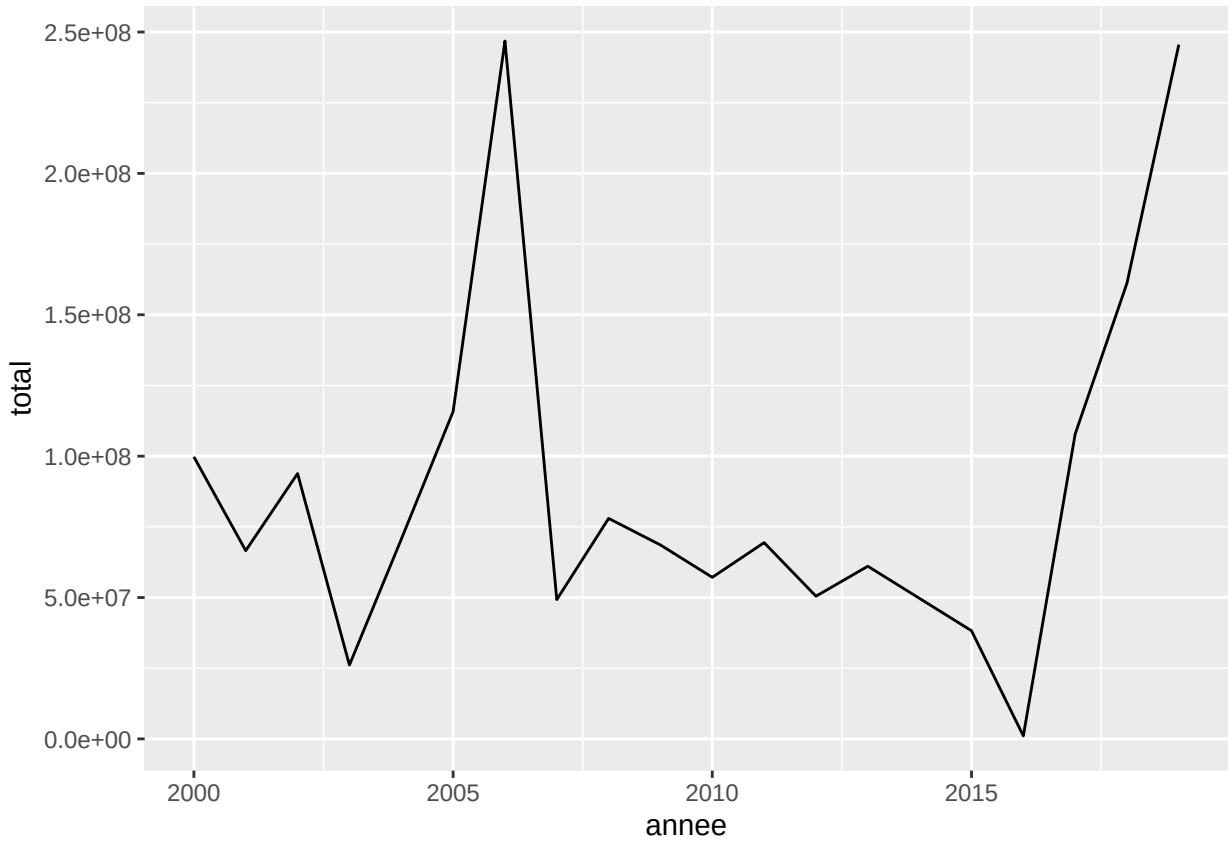
```
tableau_pa %>% group_by (annee) %>% dplyr::summarise(total=sum(captures)) %>% ggplot()+geom_line(aes(x=a
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```




```
tableau_pa %>% filter(espece=='BOBO')%>% group_by (annee) %>% dplyr::summarise(total=sum(captures)) %>%
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```



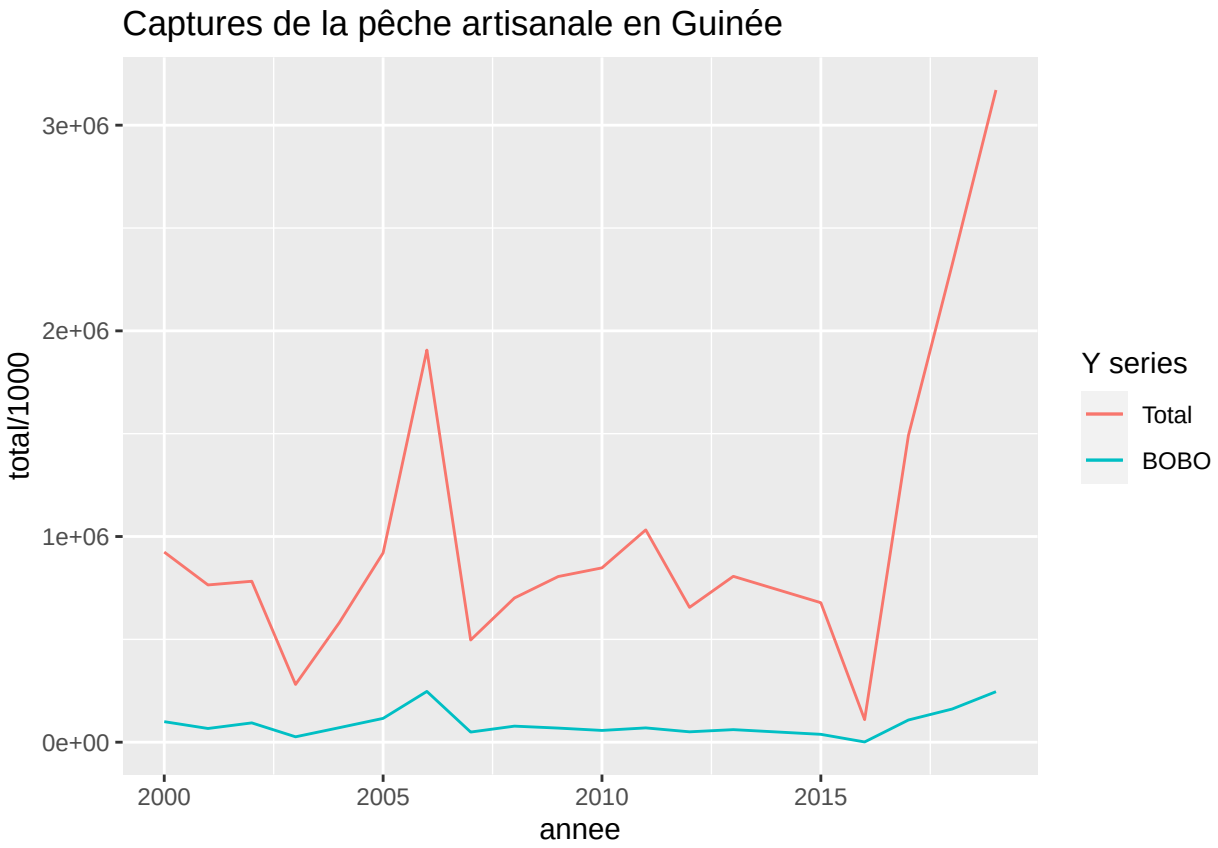

```
espece_traitement<-'BOBO'
#espece_traitement<-'DORADES DIVERSES'

captures_esp_traitement <-tableau_pa %>% filter(espece==espece_traitement)%>% group_by (annee) %>% dply
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(captures_esp_traitement)
```

```
## # A tibble: 6 x 2
##   annee      total
##   <int>      <dbl>
## 1  2000  99746974.
## 2  2001  66580573.
## 3  2002  93841430.
## 4  2003  26120424.
## 5  2004  70591801.
## 6  2005 115756277.
```

```
ggplot(captures_total)+ geom_line(aes(x=annee,y=total/1000,colour='1')) +
  geom_line(data=captures_esp_traitement ,aes(x=annee,y=total/1000,colour='2'))+
    scale_color_discrete(name = "Y series", labels = c("Total", espece_traitement)) +
  ggtitle("Captures de la pêche artisanale en Guinée")
```



Special dédicace pour Didier, le tableau croisé

```
library(tidyr)

tableau_pa %>% group_by(annee,espece) %>% dplyr::summarise(total=sum(captures)) %>%
pivot_wider(names_from=espece,values_from=total)
```

```
## `summarise()` regrouping output by 'annee' (override with `.groups` argument)
```

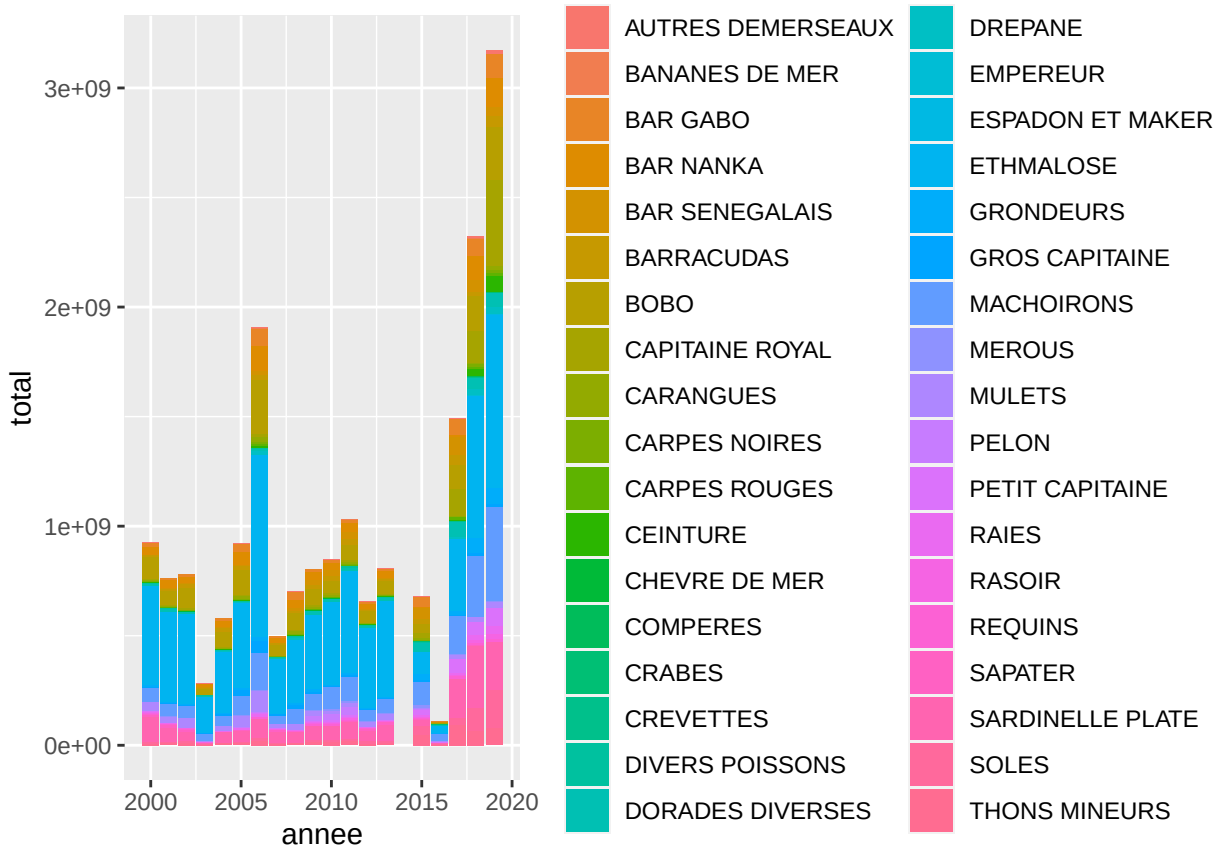
```
## # A tibble: 19 x 37
## # Groups:   annee [19]
##    annee `AUTRES DEMERSE~ `BANANES DE MER` `BAR GABO` `BAR NANKA`
##    <int>            <dbl>            <dbl>      <dbl>       <dbl>
##  1  2000         1958935.          651073.  16723836.   35608156.
##  2  2001         2205657.          710719.   9548055.   35155154.
##  3  2002         2700369.          143052.  11356888.   29789263.
##  4  2003          341386.              NA    3800357.    9446956.
##  5  2004         1926733.          359500   13065639.   24968817.
##  6  2005         2746010.          498500   36566833.   57353399.
##  7  2006         5706035.          954500   78024785.  115056247.
##  8  2007         1225311.          173589.   7120487.   20160878.
##  9  2008         2271000.          632000   35532864.   37606968.
## 10  2009         3383070.          718058.  14015446.   32512889.
## 11  2010         2791911.          524370.  12792740.   32646576.
## 12  2011         3819446.          614933.  13168768.   40785222.
## 13  2012         1808727.          496852.   7489550.   25912776.
```

```
## 14  2013         2255535.          628723.   9138087.   31520028.
## 15  2015         2185220.              NA   46630889.    5935891.
## 16  2016               NA              NA    6573918.     196496.
## 17  2017         2446375.              NA   73958414.    7497843.
## 18  2018         7339382.         5759989.  77572803.  108715961.
## 19  2019        18352139.          180000  110234928.  129170513.
## # ... with 32 more variables: `BAR SENEGALAIS` <dbl>, BARRACUDAS <dbl>,
## #   BOBO <dbl>, `CAPITAINE ROYAL` <dbl>, CARANGUES <dbl>, `CARPES
## #   NOIRES` <dbl>, `CARPES ROUGES` <dbl>, CEINTURE <dbl>, `DIVERS
## #   POISSONS` <dbl>, `DORADES DIVERSES` <dbl>, DREPANE <dbl>, EMPEREUR <dbl>,
## #   ETHMALOSE <dbl>, GRONDEURS <dbl>, `GROS CAPITAINE` <dbl>, MACHOIRONS <dbl>,
## #   MEROUS <dbl>, MULETS <dbl>, PELON <dbl>, `PETIT CAPITAINE` <dbl>,
## #   RAIES <dbl>, RASOIR <dbl>, REQUINS <dbl>, SAPATER <dbl>, `SARDINELLE
## #   PLATE` <dbl>, SOLES <dbl>, `THONS MINEURS` <dbl>, `CHEVRE DE MER` <dbl>,
## #   CRABES <dbl>, COMPERES <dbl>, CREVETTES <dbl>, `ESPADON ET MAKER` <dbl>
```

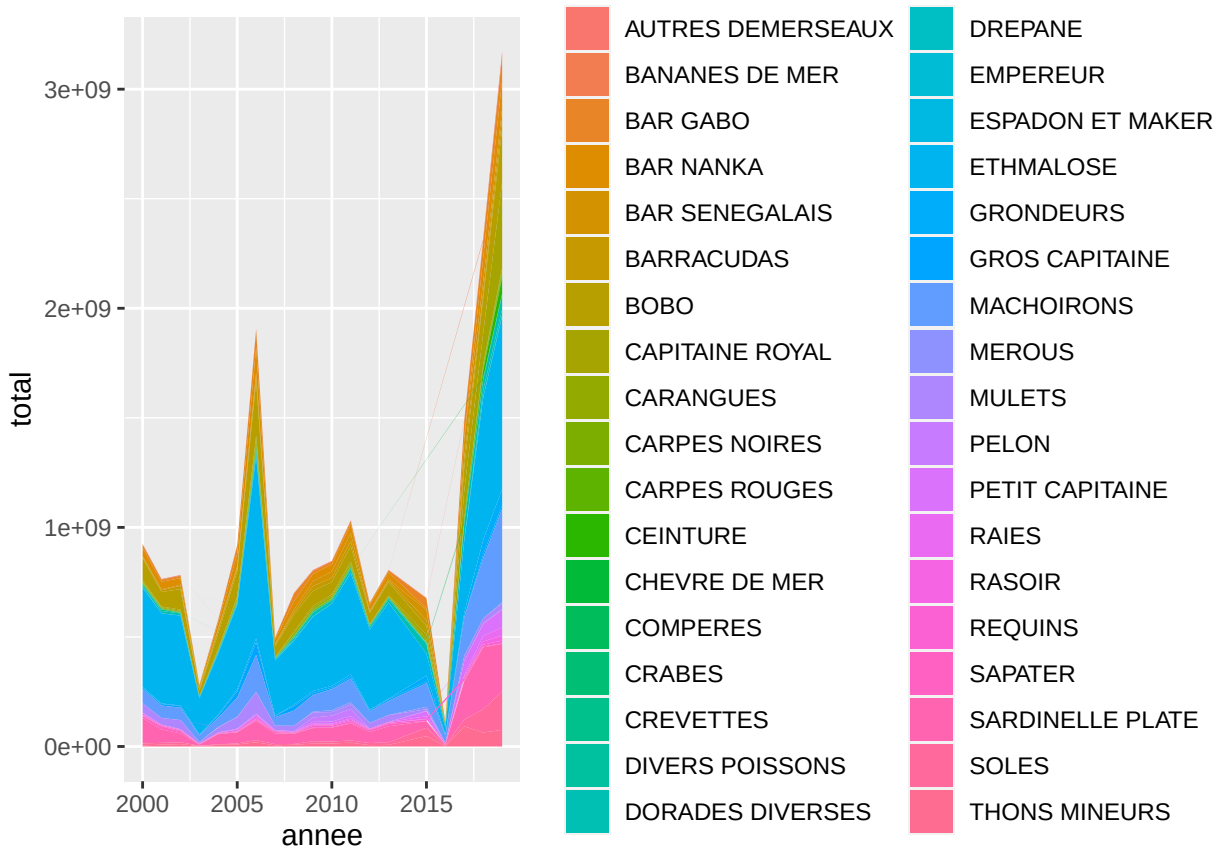
```
data<-tableau_pa %>% group_by(annee,espece) %>% dplyr::summarise(total=sum(captures))
```

```
## `summarise()` regrouping output by 'annee' (override with `.groups` argument)
```

```
ggplot(data)+geom_col(aes(x=annee,y=total,fill=espece))
```




```
ggplot(data)+geom_area(aes(x=annee,y=total,fill=espece))
```




```
#Avec plusieurs items en ligne
data<-tableau_pa %>% group_by(annee,espece,type_pirogue) %>% dplyr::summarise(total=sum(captures)) %>%
pivot_wider(names_from=espece,values_from=total)
```

```
## `summarise()` regrouping output by 'annee', 'espece' (override with `.groups` argument)
```

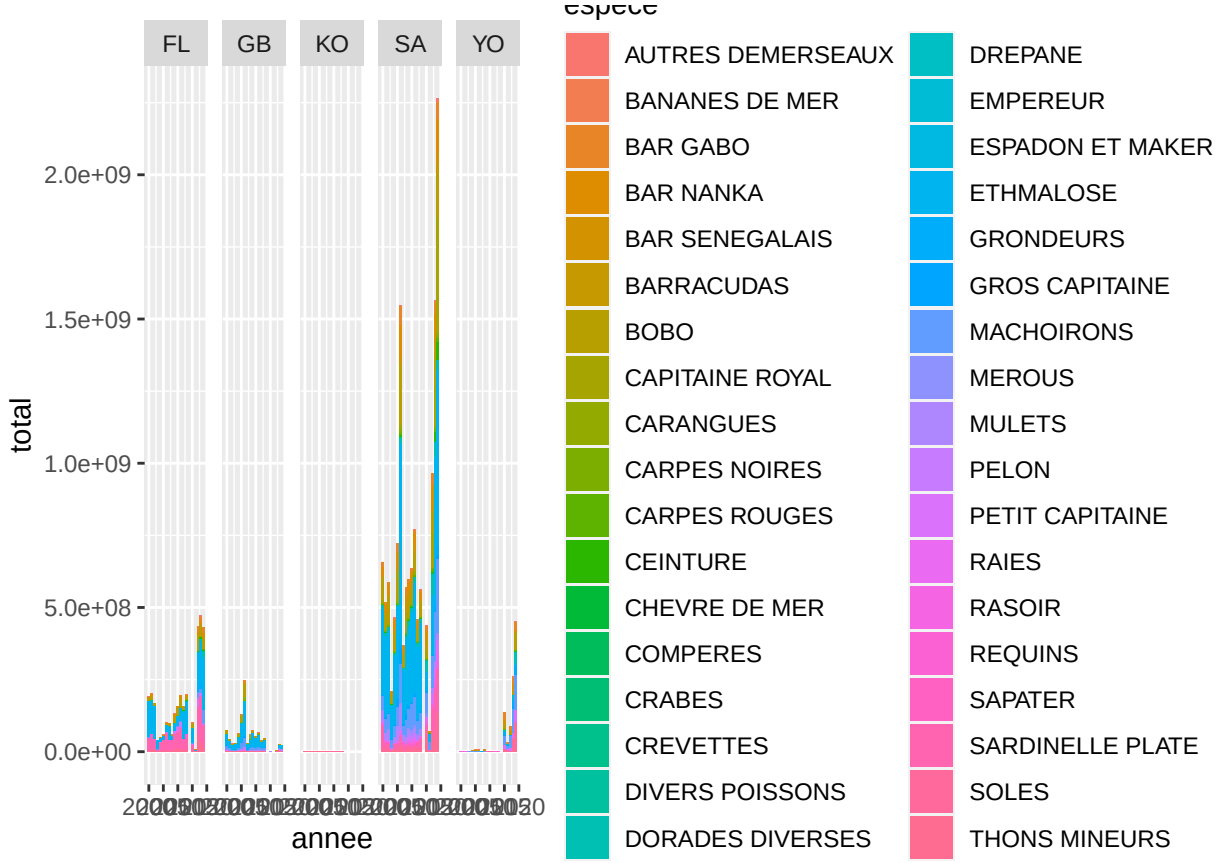
```
head(data)
```

```
## # A tibble: 6 x 38
## # Groups:   annee [2]
##   annee type_pirogue `AUTRES DEMERSE~ `BANANES DE MER` `BAR GABO` `BAR NANKA`
##   <int> <chr>                  <dbl>            <dbl>      <dbl>       <dbl>
## 1  2000 FL                   246513.          630232.    161745.    2079218.
## 2  2000 GB                     1818.              NA     717613.    1014458.
## 3  2000 SA                  1710604.           20841.  15729378.   32211366.
## 4  2000 KO                       NA              NA      63700.     276679.
## 5  2000 YO                       NA              NA      51400.      26434.
## 6  2001 FL                    21599.          695319.    651621.    5345090.
## # ... with 32 more variables: `BAR SENEGALAIS` <dbl>, BARRACUDAS <dbl>,
## #   BOBO <dbl>, `CAPITAINE ROYAL` <dbl>, CARANGUES <dbl>, `CARPES
## #   NOIRES` <dbl>, `CARPES ROUGES` <dbl>, CEINTURE <dbl>, `DIVERS
## #   POISSONS` <dbl>, `DORADES DIVERSES` <dbl>, DREPANE <dbl>, EMPEREUR <dbl>,
## #   ETHMALOSE <dbl>, GRONDEURS <dbl>, `GROS CAPITAINE` <dbl>, MACHOIRONS <dbl>,
## #   MEROUS <dbl>, MULETS <dbl>, PELON <dbl>, `PETIT CAPITAINE` <dbl>,
## #   RAIES <dbl>, RASOIR <dbl>, REQUINS <dbl>, SAPATER <dbl>, `SARDINELLE
## #   PLATE` <dbl>, SOLES <dbl>, `THONS MINEURS` <dbl>, `CHEVRE DE MER` <dbl>,
## #   CRABES <dbl>, COMPERES <dbl>, CREVETTES <dbl>, `ESPADON ET MAKER` <dbl>
```

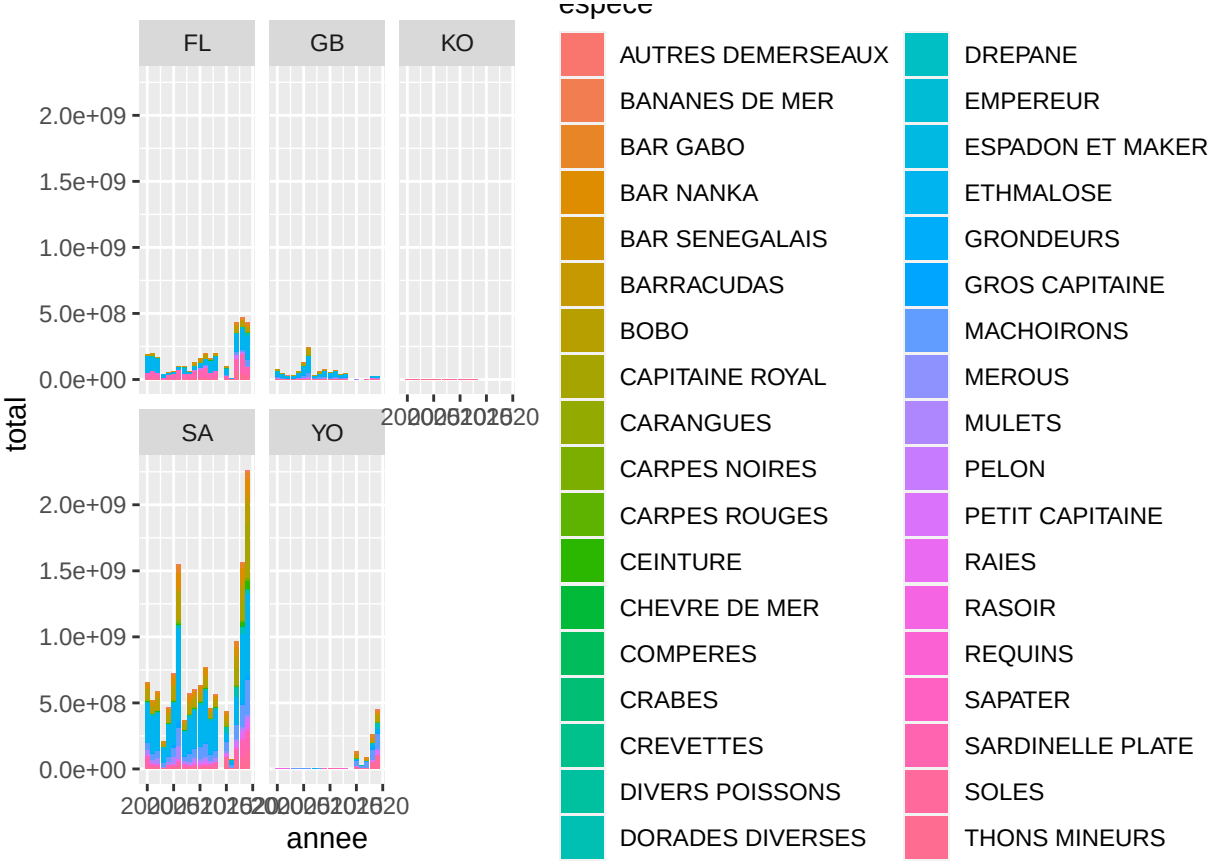
```
data<-tableau_pa %>% group_by(annee,espece,type_pirogue) %>% dplyr::summarise(total=sum(captures))
```

```
## `summarise()` regrouping output by 'annee', 'espece' (override with `.groups` argument)
```

```
ggplot(data)+geom_col(aes(x=annee,y=total,fill=espece))+
  facet_grid(~type_pirogue)
```




```
ggplot(data)+geom_col(aes(x=annee,y=total,fill=espece))+
  facet_wrap(~type_pirogue)
```



Regroupement de strates

De temps en temps, il arrive que l'on veuille ajouter des colonnes qui vont permettre d'analyser les données sur des strates qui correspondent à des regroupement de strates fines (Les mois en saison par exemple)

```
tableau_pa<-tableau_pa %>% mutate(saison = case_when(
        mois %in% 11:12 ~ "SEC",
        mois %in%  1:4  ~ "SEC",
        mois %in%  5:10  ~ "HUMIDE"))

tableau_pa<-tableau_pa %>% mutate(engin_peche2 = substr(engin_peche,1,2))
```

Calcul des cpue

Dans mon tableau j'ai des captures (capture retenue) et des efforts (duree_peche), la cpue est la division de l'un par l'autre

```
tableau_pi<-tableau_pi %>% mutate(cpue=capture_retenue/duree_peche)
```

Je commence à avoir trop de colonne, je veux n'engarder que certaines : la fonction select

```
head(
  tableau_pi %>% select(code_pays,code_navire,licence,annee,mois,jour,engin,tjb2,puissance2,cpue)
)
```

```
##   code_pays code_navire      licence annee mois jour engin   tjb2 puissance2
```

```
## 1       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
## 2       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
## 3       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
## 4       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
## 5       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
## 6       GIN           4 P. DEMERSALE  2001    7    1    NA middle       high
##       cpue
## 1 38.18182
## 2 38.18182
## 3 14.00000
## 4 14.00000
## 5 28.00000
## 6 29.16667
```

```
#Et si je ne veux plus engin et le jour

head(
  tableau_pi %>% select(-c(engin,jour))
)
```

```
##   X code_pays code_navire annee mois numero_operation                  espece
## 1 1       GIN           4  2001    7                1 PSEUDOTOLITHUS ELONGATUS
## 2 2       GIN           4  2001    7                1 PSEUDOTOLITHUS ELONGATUS
## 3 3       GIN           4  2001    7                2 PSEUDOTOLITHUS ELONGATUS
## 4 4       GIN           4  2001    7                2 PSEUDOTOLITHUS ELONGATUS
## 5 5       GIN           4  2001    7                3 PSEUDOTOLITHUS ELONGATUS
## 6 6       GIN           4  2001    7                4 PSEUDOTOLITHUS ELONGATUS
##   capture_retenue capture_rejetee      licence longitude_fin latitude_fin
## 1             126               0 P. DEMERSALE        -14.65        10.02
## 2             126               0 P. DEMERSALE        -14.65        10.02
## 3              28               0 P. DEMERSALE        -14.62        10.00
## 4              28               0 P. DEMERSALE        -14.62        10.00
## 5              56               0 P. DEMERSALE        -14.47        10.03
## 6              56               0 P. DEMERSALE        -14.63        10.03
##   date_debut_maree date_fin_maree duree_peche nombre_operation   sect_cod.x
## 1               NA             NA        3.30                1 10-25-zee-10
## 2               NA             NA        3.30                1 10-25-zee-10
## 3               NA             NA        2.00                1 10-25-zee-10
## 4               NA             NA        2.00                1 10-25-zee-10
## 5               NA             NA        2.00                1 10-25-zee-10
## 6               NA             NA        1.92                1 10-25-zee-10
##   tsect_cod.x intitule f_area   id sect_cod.y tsect_cod.y profondeur
## 1          80  10-25 m  bathy 1830  GRAT-1830          50          9
## 2          80  10-25 m  bathy 1831  GRAT-1831          50          9
## 3          80  10-25 m  bathy 1830  GRAT-1830          50          8
## 4          80  10-25 m  bathy 1831  GRAT-1831          50          8
## 5          80  10-25 m  bathy 1891  GRAT-1891          50          9
## 6          80  10-25 m  bathy 1831  GRAT-1831          50          9
##           geometry   nom_navire    tjb puissance longueur nationalite saison
## 1    c(-14.65, 10) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
## 2    c(-14.65, 10) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
## 3    c(-14.63, 10) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
## 4    c(-14.63, 10) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
## 5 c(-14.45, 10.02) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
## 6 c(-14.62, 10.02) POONG LIM 11 253.75      1600       45  Cor<e9>ens HUMIDE
```

```
##     tjb2 puissance2 longueur2 profondeur2     cpue
## 1 middle       high    middle    very low 38.18182
## 2 middle       high    middle    very low 38.18182
## 3 middle       high    middle    very low 14.00000
## 4 middle       high    middle    very low 14.00000
## 5 middle       high    middle    very low 28.00000
## 6 middle       high    middle    very low 29.16667
```

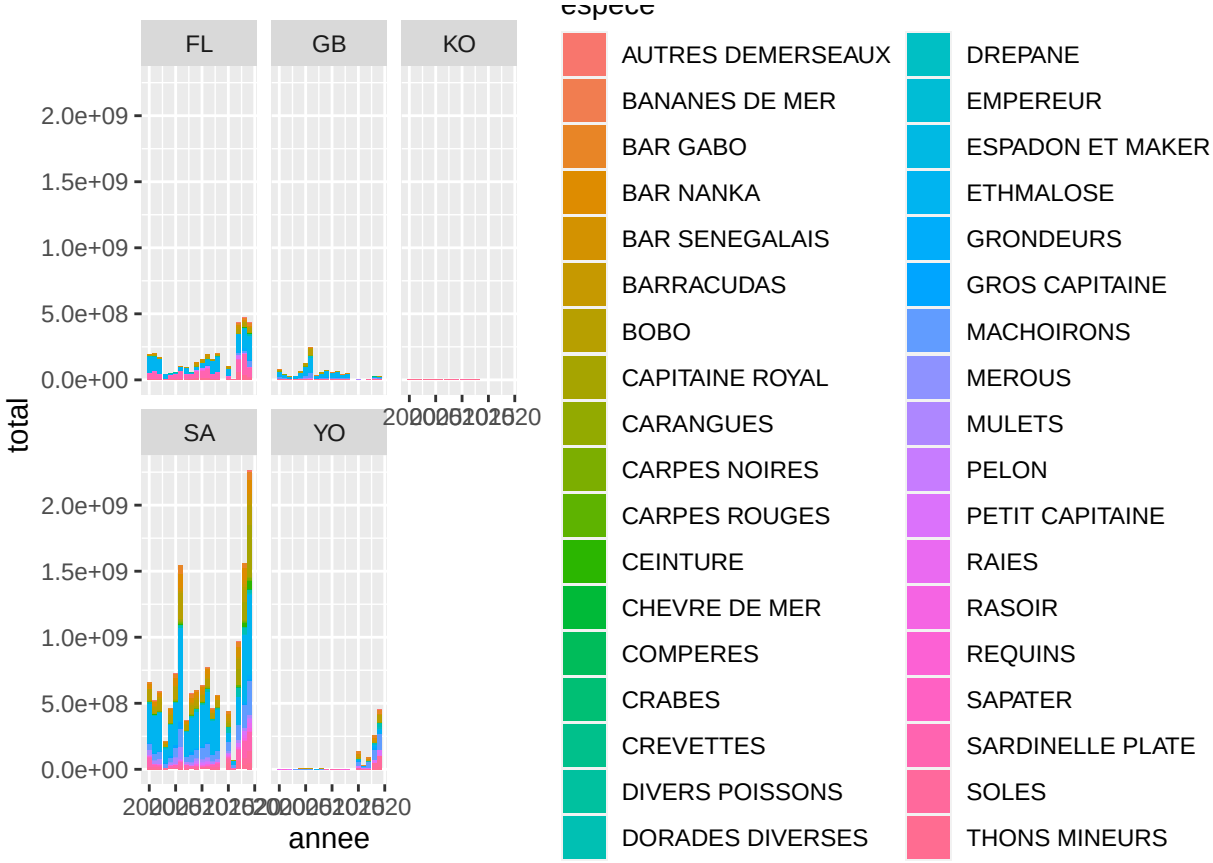
On enchaine et on refait plus proprement le graphique en barres empilées

Je veux afficher les 5 premières espèces et mettre le reste en autres

Donc il faut d'abord repérer ces 5 espèces, renommer le reste en Autre et faire le graphe

Pour le repérage, * je fais une somme des captures par espece(mutate) * ensuite on tri la somme par ordre décroissant (arrange) * je ne prend que les lignes de 1 à 5 (slice) * je ne prend que la colonne espece (select)

```
ggplot(data)+geom_col(aes(x=annee,y=total,fill=espece))+
  facet_wrap(~type_pirogue)
```



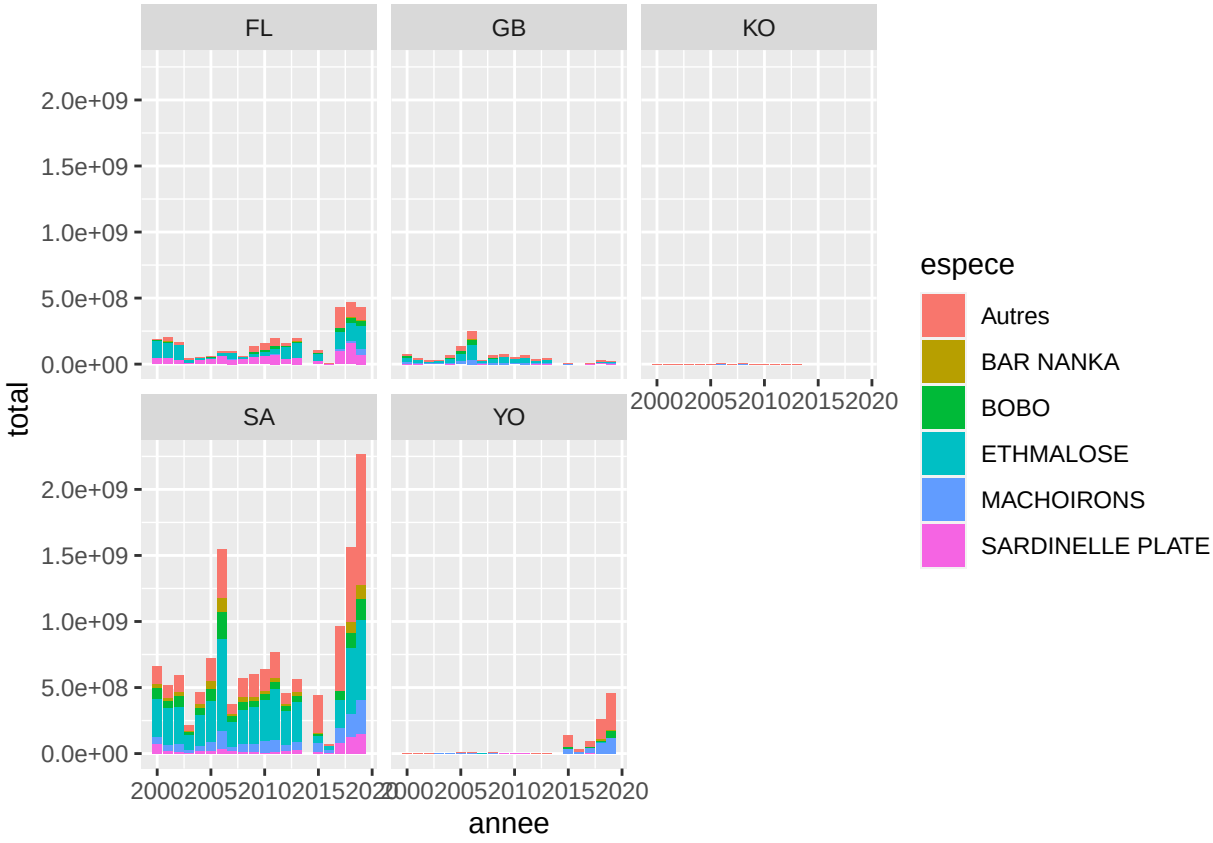

```
liste_5_premiers<-
  data %>%
  group_by(espece) %>% summarize(total=sum(total)) %>% arrange(desc(total)) %>% slice(1:5) %>% select(es
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
head(liste_5_premiers)
```

```
## # A tibble: 5 x 1
##   espece
##   <chr>
## 1 ETHMALOSE
## 2 MACHOIRONS
## 3 BOBO
## 4 SARDINELLE PLATE
## 5 BAR NANKA
```

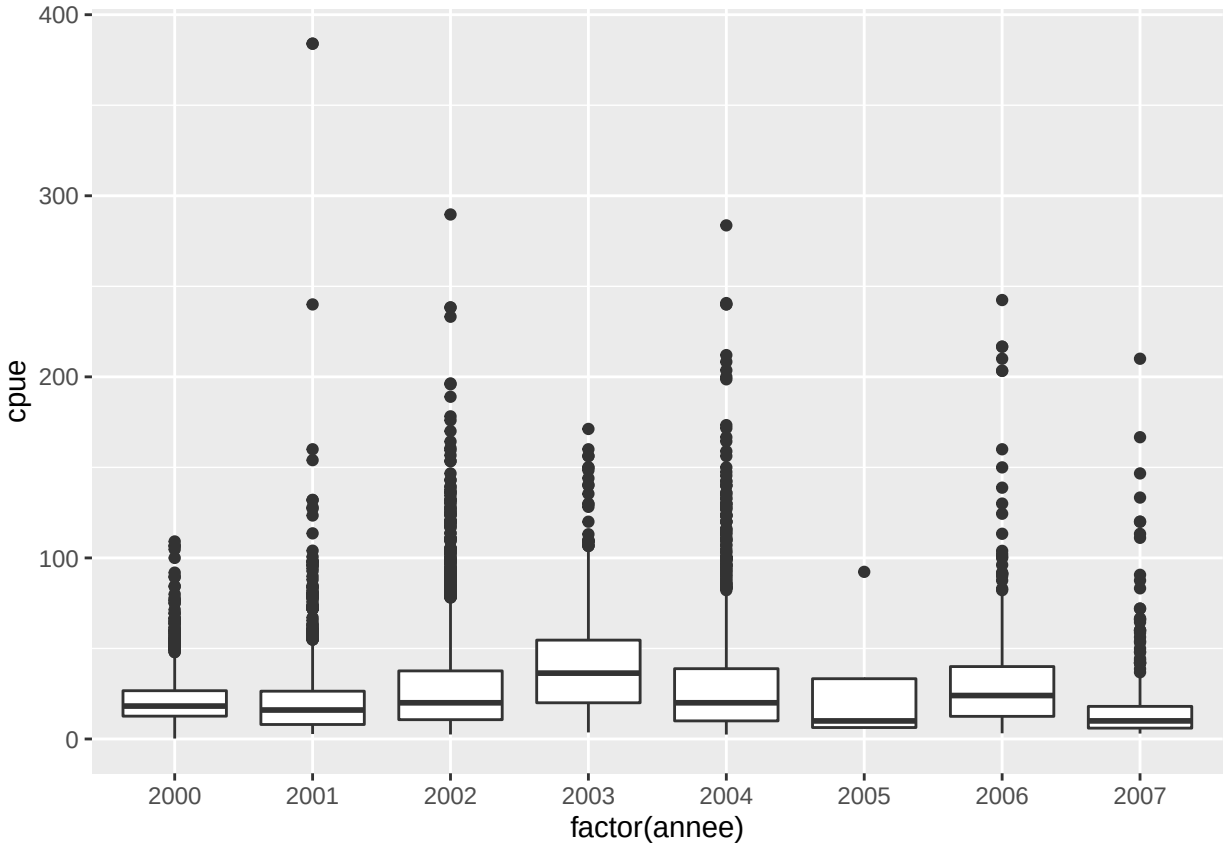
```
data %>% mutate(espece=
                  case_when(
        espece %in% liste_5_premiers$espece ~ espece,
        TRUE ~ 'Autres')) %>%
  ggplot()+geom_col(aes(x=annee,y=total,fill=espece))+
  facet_wrap(~type_pirogue)
```



Un autre type de graphique : les boxplot

Les captures par unités d'effort sont intéressantes à mettre dans un boxplot pour savoir si les données se répartissent de manière homogène autour de la moyenne

```
ggplot(tableau_pi)+geom_boxplot(aes(x=factor(annee),y=cpue))
```

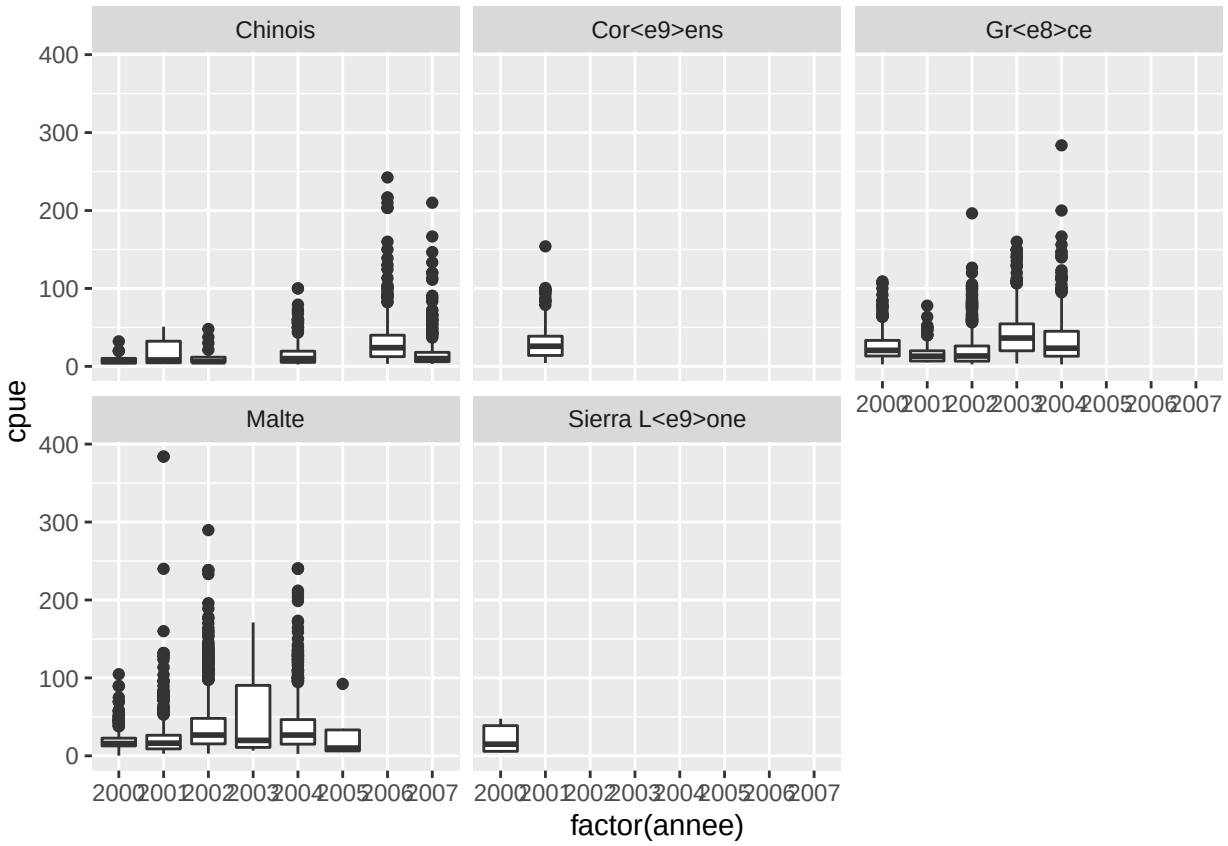


Et on peut regarder ces mêmes données ppar type de licence par exemple

```
names(tableau_pi)
```

```
##  [1] "X"                "code_pays"        "code_navire"      "annee"
##  [5] "mois"             "jour"             "numero_operation" "espece"
##  [9] "capture_retenue"  "capture_rejetee"  "licence"          "longitude_fin"
## [13] "latitude_fin"     "date_debut_maree" "date_fin_maree"   "engin"
## [17] "duree_peche"      "nombre_operation" "sect_cod.x"       "tsect_cod.x"
## [21] "intitule"         "f_area"           "id"               "sect_cod.y"
## [25] "tsect_cod.y"      "profondeur"       "geometry"         "nom_navire"
## [29] "tjb"              "puissance"        "longueur"         "nationalite"
## [33] "saison"           "tjb2"             "puissance2"       "longueur2"
## [37] "profondeur2"      "cpue"
```

```
ggplot(tableau_pi)+geom_boxplot(aes(x=factor(annee),y=cpue))+
  facet_wrap(~nationalite)
```




```
unique(tableau_pi$nationalite)
```

```
## [1] "Cor<e9>ens"      "Gr<e8>ce"        "Malte"           "Sierra L<e9>one"
## [5] "Chinois"
```