

projet

loua_mehamsadj

30 mars 2017

```
#QUESTION DU SUJET D'ETUDES :  
#Es-ce que la difference de salaire entre les hommes et les femmes a diminué de 1962 à 2012 ?  
#Dans cette étude, nous avons utilisé le geompoint()comme représentation graphique.  
#Parce qu'elle est plus facile à mettre en place pour nos trois variables  
#qui sont : Annee, Sexe et salaire.  
  
# nous chargeons ici les bases données hommes et femmes à partir de mon bureau  
#il faudra changer le chemin d'accès à partir d'un autre pc  
library(readr)  
  
## Warning: package 'readr' was built under R version 3.3.3  
hommes <- read_delim("C:/Users/DELL/Desktop/cours 2016-2017/projet stat/hommes.csv",  
  ";", escape_double = FALSE, col_types = cols(Hommes = col_date(format = "%Y")),  
  trim_ws = TRUE)  
  
femmes <- read_delim("C:/Users/DELL/Desktop/cours 2016-2017/projet stat/femmes.csv",  
  ";", escape_double = FALSE, col_types = cols(Femmes = col_date(format = "%Y")),  
  trim_ws = TRUE)  
  
#nous reprenons nos base de données hommes et femmes générés à partir de fichiers csv  
#après plusieurs modifications sur les fichiers originaux  
#puis nous utilisons rbind  
#pour fusionner les tables et nous rajoutons une colonne sexe  
#pour les différencier dans notre base baserecap  
#enfin nous ne considérons plus les lignes non définies (nd)  
library(reshape2)  
  
## Warning: package 'reshape2' was built under R version 3.3.3  
library(dplyr)  
  
## Warning: package 'dplyr' was built under R version 3.3.3  
##  
## 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  
dfh <- hommes %>% rename(Annee = Hommes) %>% mutate(Sexe = "M")  
dff <- femmes %>% rename(Annee = Femmes) %>% mutate(Sexe = "F")  
  
df <- rbind(dfh, dff)
```

```
baserecap<-melt(df, id.vars = c("Annee", "Sexe") ) %>%
  filter(value != "nd") %>%
  mutate(value = as.integer(value))
```

```
#Nous nous intéressons ici au salaire moyen depuis 1962 jusqu'à 2012
#pour les hommes et les femmes en France
#Nous observons dans un premier temps que le salaire croit en fonction des années
#puis nous remarquons que l'écart du salaire moyen entre les hommes
#et les femmes est plus important aussi au fil des années
#Cependant nous n'avons pas de précision sur le coefficient de corrélation entre
#le salaire moyen de l'homme sur celui de la femme par année
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.3
```

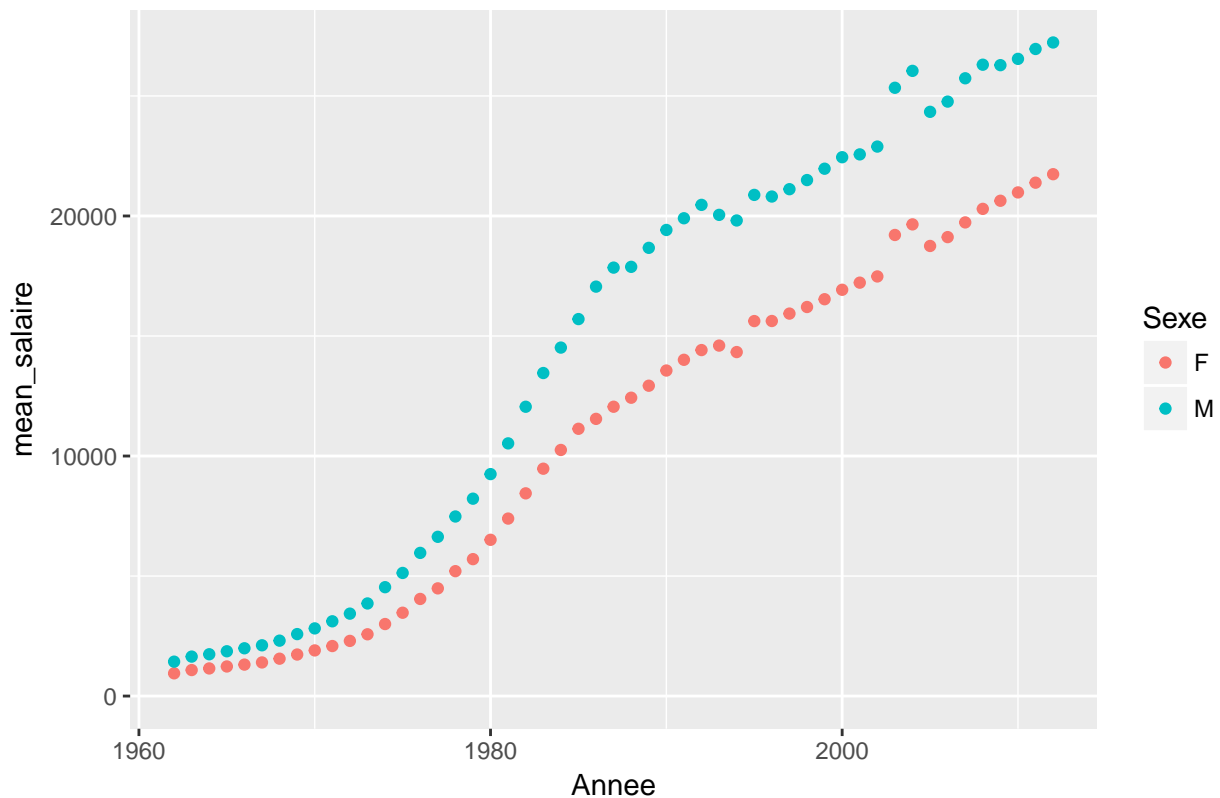
```
library(magrittr);
```

```
## Warning: package 'magrittr' was built under R version 3.3.3
```

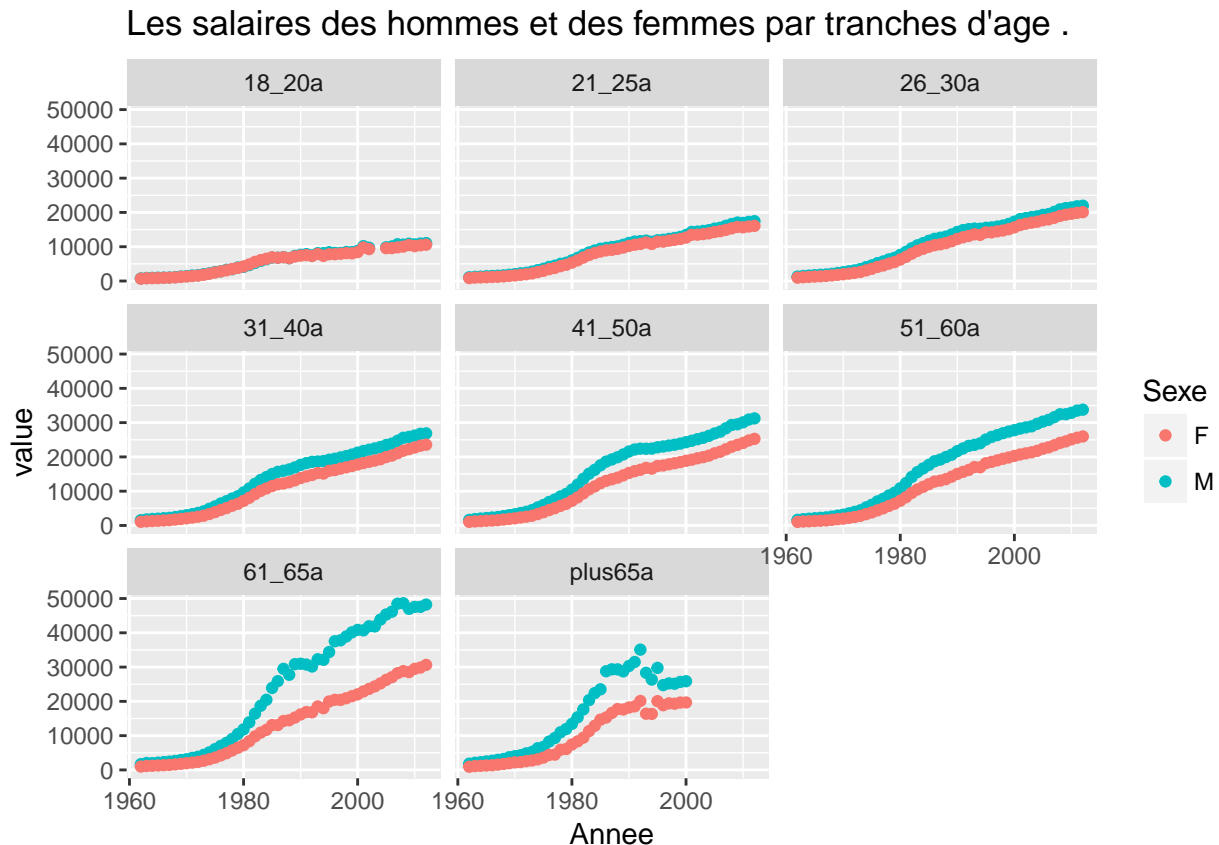
```
p<- baserecap %>% group_by(Annee,Sexe)%>%
  summarize(mean_salaire=mean(value))%>%
  ggplot(aes(x=Annee,y=mean_salaire,color=Sexe)) +
  geom_point() +
  ylim(0, NA)+
  ggtitle("Les salaires moyens des hommes et des femmes de 1962 à 2012")
```

p

Les salaires moyens des hommes et des femmes de 1962 à 2012



```
#ici on compare les salaires entre les hommes et les femmes par tranches d'age .
library(ggplot2)
library(magrittr);
# Barplot basique
p<- baserecap %>%
  ggplot(aes(x=Annee, y=value, color=Sexe)) +
  geom_point() +
  facet_wrap(~variable) + ylim(0, NA)+
  ggtitle("Les salaires des hommes et des femmes par tranches d'age . ")
p
```



```
#nous choisissons trois tranches d'age relativement significative :
#les hommes et les femmes
#ayant entre 18 et 20 ans ,entre 31 et 40 ans et entre 61 et 65 ans .
```

```
#pour la tranche d'age des 18 20 ans on peut dire que le salaire est approximativement similaire
#meme si ceux des hommes restent légèrement supérieur sur la majorité de la durée observée
```

```
#pour la tranche d'age des 31 40 ans qui est peut etre la plus importante
#la différence des salaires est nette en faveur des hommes
#on peut aussi émettre plusieurs hypothèses expliquant cette différence
#par exemple à cette tranche d'âge beaucoup de femmes pourraient etre en congé maternité
```

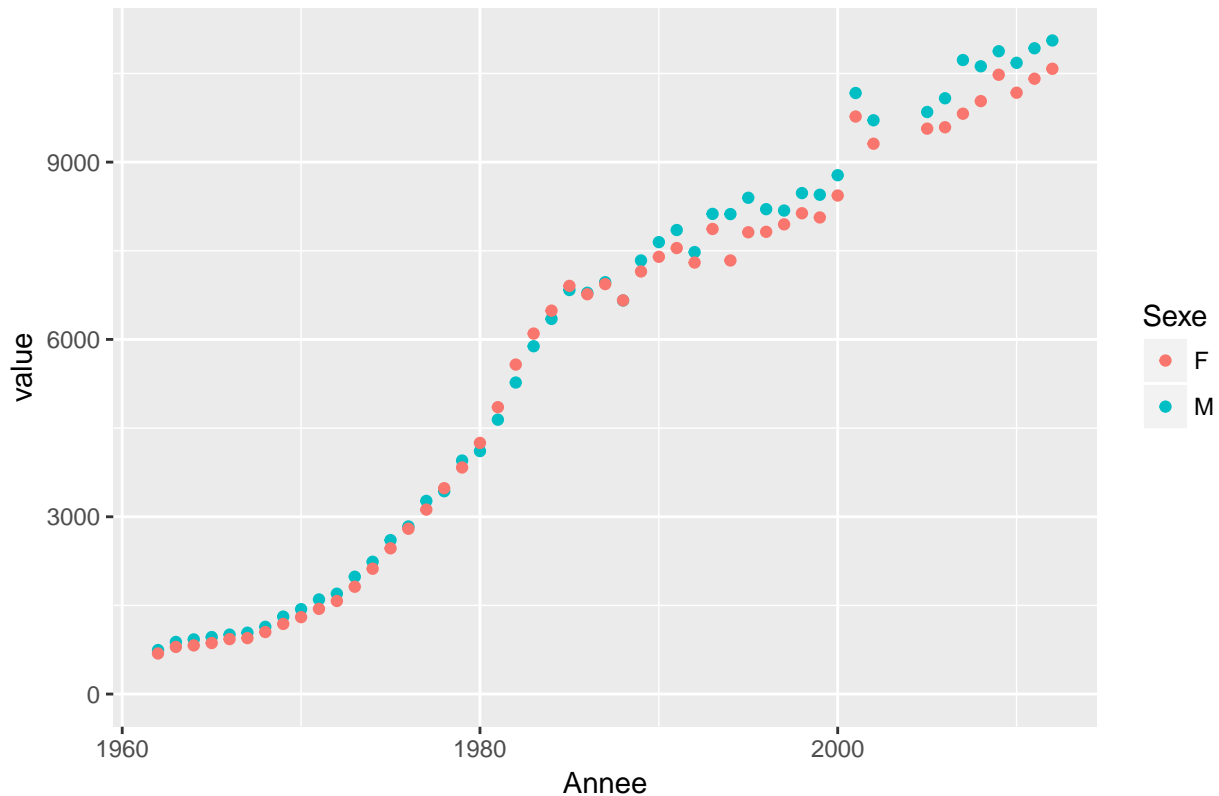
```
#pour la tranche d'age des 60 65 ans les salaires
#sont plus élevés que pour les tranches d'age précédentes
#cependant l'écart entre les salaire par genre
```

#est encore plus important que les tranche d'age vues précédemment

```
library(ggplot2)
library(magrittr);
# Barplot basique
p<- baserecap %>% filter(variable == "18_20a") %>%
  ggplot(aes(x=Annee, y=value, color=Sexe)) +
  geom_point() +
  ggtitle(" Les salaires des hommes et des femmes ayant entre 18 et 20 ans par année ") +
  ylim(0, NA)
p
```

Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)): taille de
police inconnue pour le caractère 0x17

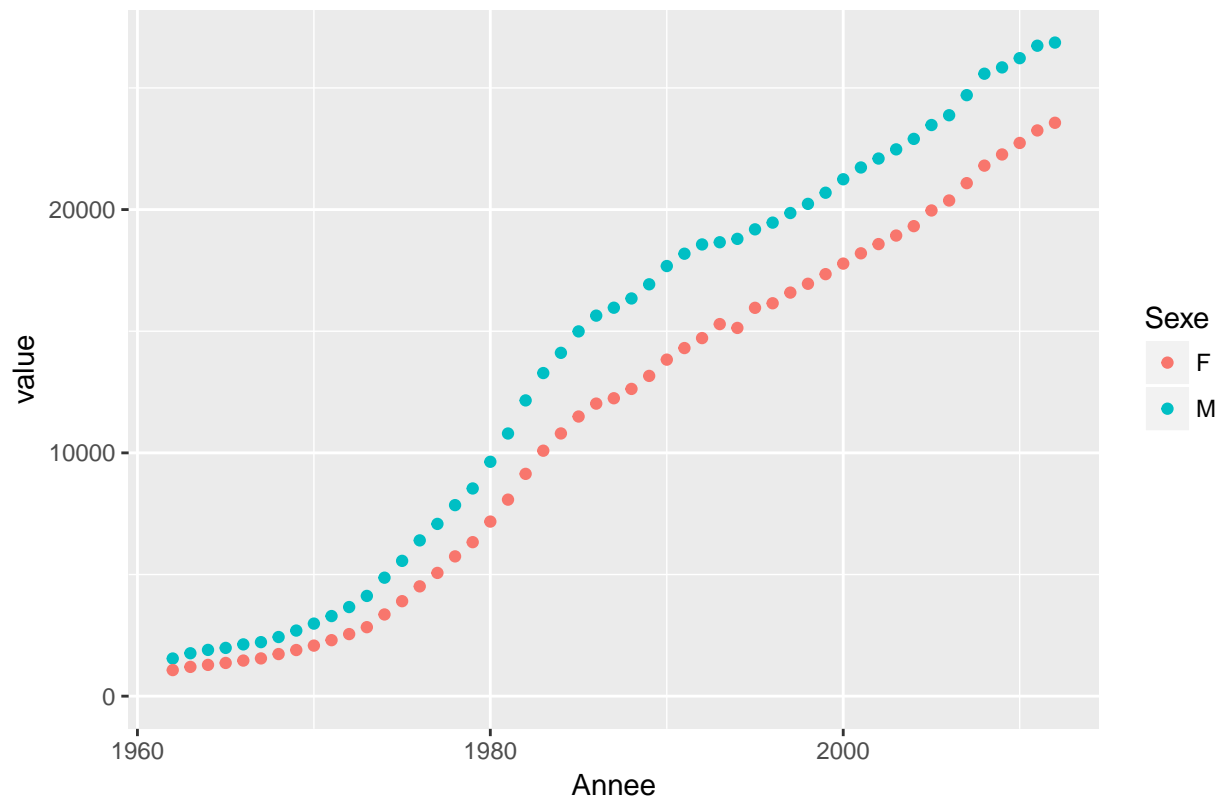
Les salaires des hommes et des femmes ayant entre 18 et 20 ans par an



```
p<- baserecap %>% filter(variable == "31_40a") %>%
  ggplot(aes(x=Annee, y=value, color=Sexe)) +
  geom_point()+
  ggtitle(" Les salaires des hommes et des femmes ayant entre 31 et 40 ans par année ") +
  ylim(0, NA)
p
```

Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x\$label)): taille de
police inconnue pour le caractère 0x17

Les salaires des hommes et des femmes ayant entre 31 et 40 ans par ar

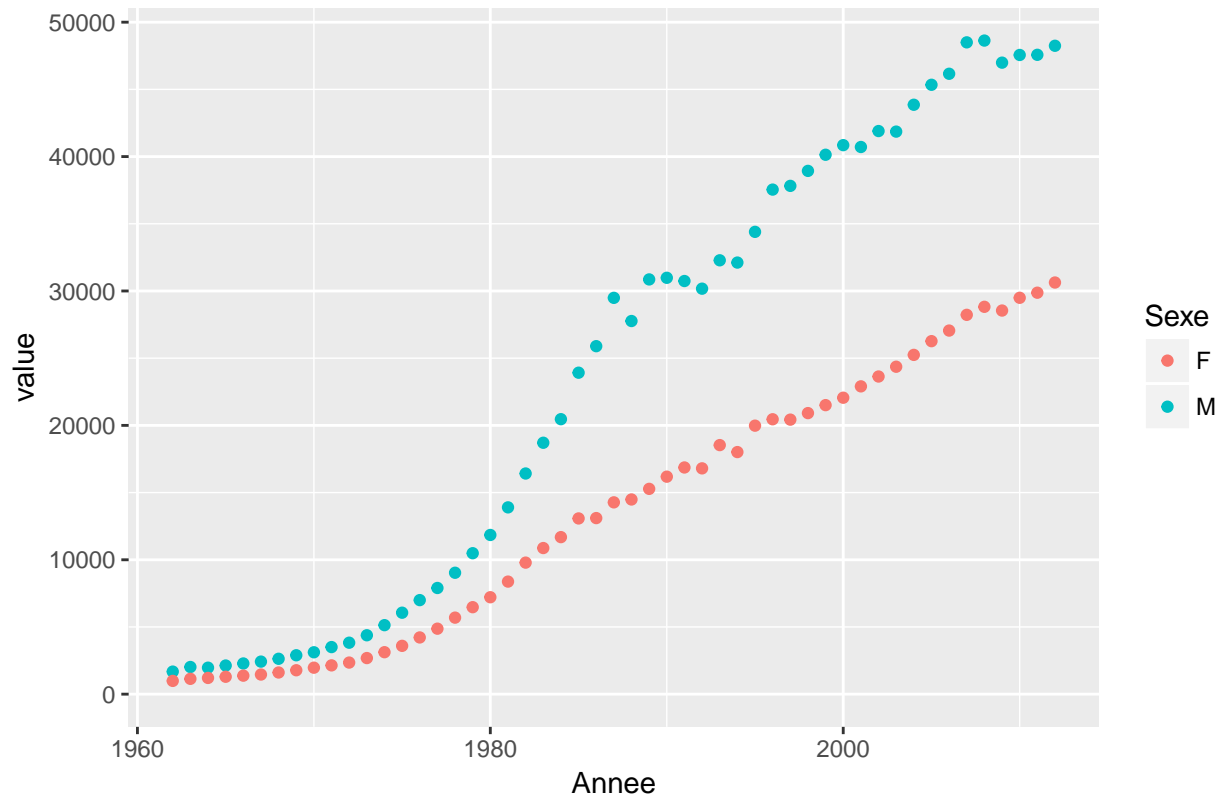


```
p<- baserecap %>% filter(variable == "61_65a") %>%
  ggplot(aes(x=Annee, y=value, color=Sexe)) +
  geom_point() +
  ggtitle(" Les salaires des hommes et des femmes ayant entre 61 et 65 ans par année")+
  ylim(0, NA)
```

p

```
## Warning in grid.Call(L_stringMetric, as.graphicsAnnot(x$label)): taille de
## police inconnue pour le caractère 0x17
```

Les salaires des hommes et des femmes ayant entre 61 et 65 ans par an



*#Notre étude a montré que les salaires des femmes depuis 1962
 #est généralement inférieur à ceux des hommes
 #et ceci n'a visiblement pas changé en 2012
 #Nous observons aussi avec la repartition par tranche d'age
 #que l'expérience en année chez l'homme est plus récompensé que chez la femme .*