

Statistiques population traitée 2021

DONNEES

Population totale

```
setwd("~/Documents/2A/Statapp/données")
library(readr)
nat2021us <- read_csv("W:/Documents/2A/Statapp/données/nat2021us.csv", col_types = cols_only((apgar5r = col_guess(),
  dbwt = col_guess(),
  rf_inftr = col_guess(),
  rf_fedrg = col_guess(),
  rf_artec = col_guess(),
  mager = col_guess(),
  mrace6 = col_guess(),
  dmar = col_guess(),
  meduc = col_guess(),
  fagerec11 = col_guess(),
  frace6 = col_guess(),
  feduc = col_guess(),
  priorlive = col_guess(),
  dplural = col_guess()))
```

```
head.matrix(nat2021us)
```

```
## # A tibble: 6 x 14
##   mager mrace6 dmar meduc fagerec11 frace6 feduc priorlive rf_inftr
##   <dbl> <dbl> <dbl> <dbl>      <dbl> <dbl> <dbl>      <dbl> <chr>
## 1    22    10     1     3         3     9     4         0 N
## 2    31    61     2     4         4     5     3         1 N
## 3    29    10     2     4         4     1     4         1 N
## 4    39    10     2     6         7     1     5         0 N
## 5    20    51     2     3         4     5     3         0 N
## 6    29    10     1     3         6     1     9         0 N
## # ... with 5 more variables: rf_fedrg <chr>, rf_artec <chr>,
## #   apgar5r <dbl>, dplural <dbl>, dbwt <dbl>
```

Suppression des observations comportant des valeurs manquantes

```
nat2021us <- subset(nat2021us, (nat2021us$mrace6 != 6 & nat2021us$meduc != 9 &
nat2021us$fagerec11 != 11 & nat2021us$frace6 != 9 & nat2021us$frace6 != 6 & n
at2021us$feduc != 9 & nat2021us$apgar5r != 5 & nat2021us$priorlive != 99 & na
t2021us$dbwt != 9999))
```

Ainsi nettoyée la base de données comporte 2 855399 observations.

Création du groupe de contrôle n'ayant pas recours aux traitements contre l'infertilité

```
control = subset(nat2021us, rf_inftr != "Y")
```

Population cible

De même, on crée une base spécifique à ceux qui ont recours aux traitements

```
pop = subset(nat2021us, rf_inftr == "Y")
```

Cette base comporte 70 843 observations.

Quelques statistiques descriptives

```
table(nat2021us$rf_inftr)
```

```
##  
##      N      U      Y  
## 2782546  2010  70843
```

70843 ont recours à des traitements contre l'infertilité

```
round(table(nat2021us$rf_inftr)/length(nat2021us$rf_inftr),3)
```

```
##  
##      N      U      Y  
## 0.974 0.001 0.025
```

ce qui représente 2.5% des naissances

parmi ceux qui ont eu recours à des traitements contre l'infertilité

```
table(pop$rf_fedrg)
```

```
##  
##      N      U      Y  
## 42343  2657 25843
```

25843 ont recours à des traitements contre l'infertilité sous forme de médicaments ou insémination artificielle

```
round(table(pop$rf_fedrg)/length(pop$rf_fedrg),3) # soit environ 36.5%
```

```
##  
##      N      U      Y  
## 0.598 0.038 0.365
```

```
table(pop$rf_artec)
```

```
##
##      N      U      Y
## 20264 2657 47922

# 47922 sous forme d'assisted reproductive technology (ART) (IVF ou autres)
round(table(pop$rf_artec)/length(pop$rf_artec),3) # soit 67.6%

##
##      N      U      Y
## 0.286 0.038 0.676

# nombre d'enfant avant cette naissance
round(table(pop$priorlive)/length(pop$priorlive),3)

##
##      0      1      2      3      4      5      6      7      8      9     10     11
## 0.559 0.312 0.086 0.028 0.010 0.004 0.002 0.001 0.000 0.000 0.000 0.000
##      12
## 0.000

# De plus, 56% n'ont pas eu d'enfant auparavant
```

COMPARAISON SELON LES TRAITEMENTS

Les traitements sont classés en deux types :

les médicaments et les techniques d'insémination artificielle correspondant à la variable rf_fedrg pour fertility enhancng drugs

les techniques de procréation médiacale assistée plus poussées (fécondation in vitro) que retranscrit la variable rf_artec

On pourrait étudier les recours à tous ces traitements contre l'infertilité, cependant, dans la mesure où notre but est in fine de déterminer les effets causaux d'un traitement sur la santé des nouveau-nés, nous devons choisir le traitement pour lequel l'outcome potentiel pour les traités s'il n'avaient pas reçu le traitement a le plus de probabilité d'exister. Autrement dit, le type de traitement pour lequel, même si les parents n'y avaient pas eu recours, la mère aurait pu donner naissance, ce qui correspond à la sous-population parmi ceux qui ont recours à des traitements contre l'infertilité, qui a le moins de problème d'infertilité ou des problèmes plus "bénins".

Ainsi, on se livre d'abord à une comparaison des caractéristiques des deux sous-populations pour déterminer le traitement que nous allons finalement considérer et ensuite étudier les caractéristiques de cette population plus en détail.

```
summary(subset(pop, rf_fedrg == "Y"))
```

```
##      mager      mrace6      dmar      meduc
## Min.   :16.00   Min.   :10.00   Min.   :1.00   Min.   :1.0
## 1st Qu.:30.00   1st Qu.:10.00   1st Qu.:1.00   1st Qu.:5.0
## Median :33.00   Median :10.00   Median :1.00   Median :6.0
## Mean   :33.25   Mean    :13.89   Mean    :1.05   Mean    :5.8
## 3rd Qu.:36.00   3rd Qu.:10.00   3rd Qu.:1.00   3rd Qu.:7.0
## Max.   :50.00   Max.    :61.00   Max.    :2.00   Max.    :8.0
##
##      fagerec11      frace6      feduc      priorlive
## Min.   : 2.000   Min.   :1.000   Min.   :1.000   Min.   : 0.0000
## 1st Qu.: 5.000   1st Qu.:1.000   1st Qu.:4.000   1st Qu.: 0.0000
## Median : 6.000   Median :1.000   Median :6.000   Median : 0.0000
## Mean    : 5.655   Mean    :1.292   Mean    :5.452   Mean    : 0.6149
## 3rd Qu.: 6.000   3rd Qu.:1.000   3rd Qu.:7.000   3rd Qu.: 1.0000
## Max.    :10.000   Max.    :5.000   Max.    :8.000   Max.    :10.0000
##
##      rf_inftr      rf_fedrg      rf_artec      apgar5r
## Length:25843      Length:25843      Length:25843      Min.   :1.000
## Class :character   Class :character   Class :character   1st Qu.:4.000
## Mode  :character   Mode  :character   Mode  :character   Median :4.000
##                                     Mean    :3.749
##                                     3rd Qu.:4.000
##                                     Max.    :4.000
##
##      dplural      dbwt
## Min.   :1.00   Min.   : 234
## 1st Qu.:1.00   1st Qu.:2730
## Median :1.00   Median :3180
## Mean    :1.19   Mean    :3085
## 3rd Qu.:1.00   3rd Qu.:3545
## Max.    :4.00   Max.    :6435
##
```

```
summary(subset(pop, rf_artec == "Y"))
```

```
##      mager      mrace6      dmar      meduc
## Min.   :18.00   Min.   :10.00   Min.   :1.000   Min.   :1.000
## 1st Qu.:32.00   1st Qu.:10.00   1st Qu.:1.000   1st Qu.:6.000
## Median :35.00   Median :10.00   Median :1.000   Median :6.000
## Mean    :35.66   Mean    :15.81   Mean    :1.045   Mean    :6.064
## 3rd Qu.:39.00   3rd Qu.:10.00   3rd Qu.:1.000   3rd Qu.:7.000
## Max.    :50.00   Max.    :61.00   Max.    :2.000   Max.    :8.000
##
##      fagerec11      frace6      feduc      priorlive
## Min.   : 2.000   Min.   :1.000   Min.   :1.000   Min.   : 0.0000
## 1st Qu.: 5.000   1st Qu.:1.000   1st Qu.:5.000   1st Qu.: 0.0000
## Median : 6.000   Median :1.000   Median :6.000   Median : 0.0000
## Mean    : 6.167   Mean    :1.435   Mean    :5.752   Mean    : 0.6486
## 3rd Qu.: 7.000   3rd Qu.:1.000   3rd Qu.:7.000   3rd Qu.: 1.0000
```

```
## Max. :10.000 Max. :5.000 Max. :8.000 Max. :12.0000
##
## rf_inftr rf_fedrg rf_artec apgar5r
## Length:47922 Length:47922 Length:47922 Min. :1.000
## Class :character Class :character Class :character 1st Qu.:4.000
## Mode :character Mode :character Mode :character Median :4.000
## Mean :3.765
## 3rd Qu.:4.000
## Max. :4.000
##
## dplural dbwt
## Min. :1.000 Min. : 227
## 1st Qu.:1.000 1st Qu.:2778
## Median :1.000 Median :3222
## Mean :1.159 Mean :3127
## 3rd Qu.:1.000 3rd Qu.:3590
## Max. :4.000 Max. :6053
##
```

La différence majeure entre ces deux populations réside dans l'âge de la mère : les mères ayant recours au deuxième type de traitement sont en moyenne plus âgées de deux ans. Les pères sont aussi en moyenne plus âgés mais la différence est moindre, même si la différence tient sans doute aux valeurs extrêmes. En outre, les parents ont en moyenne un niveau d'éducation plus élevé.

Les risques de complications devenant plus fréquents avec l'âge de la mère, on pourrait en déduire que la sous-population ayant recours au premier type de traitement est plus propice à notre étude, étant donnés les critères évoqués. Pourtant, la fertilité des mères dans les deux groupes ne semble pas être si différentes : le nombre d'enfants que la mère a eu auparavant étant proche pour les deux sous-population. Cela est sans doute dû à un effet d'âge : depuis plusieurs décennies on observe une tendance des mères à avoir des enfants plus tard de sorte que l'âge des mères à la naissance de leur premier enfant est proche de celui de la première sous-population, 33 ans. Ainsi, si la plupart des mères n'ont pas eu d'enfants auparavant dans les deux sous-populations, cela révèle bien soit une tendance à avoir des enfants encore plus tard chez la deuxième sous-population plus éduquée, soit des difficultés à avoir des enfants et donc des problèmes d'infertilité. En effet, l'infertilité est définie par l'OMS comme "l'absence de grosse malgré des rapports sexuels non protégés pendant une période d'au moins 12 mois". Ainsi, si pour avoir recours à des traitements contre l'infertilité, il faut d'abord confirmer ce problème, ce qui prend un an, puis même en cas où la mère tombe peu après enceinte, la naissance n'a lieu qu'environ 9 mois après, ceci impliquant une différence de presque 2 ans entre l'âge de la mère lorsque l'enfant est désiré et l'enfant est né, soit la différence entre l'âge des mères à la naissance dans les deux sous-populations.

Création de la population "traitée"

```
traites = subset(pop, rf_fedrg == "Y")
```

CARACTERISTIQUES DE LA POPULATION CIBLE

Age des parents

Age de la mère :

12 10 – 12 years

13 13 years

...

49 49 years 5

0 50 years and over

Age du père :

01 Under 15 years

02 15-19 years

03 20-24 years

04 25-29 years

05 30-34 years

06 35-39 years

07 40-44 years

08 45-49 years

09 50-54 years

10 55-98 years

11 Not stated

```
summary(select(traites,mager))
```

```
##      mager
##  Min.    :16.00
##  1st Qu.:30.00
##  Median :33.00
##  Mean    :33.25
##  3rd Qu.:36.00
##  Max.    :50.00
```

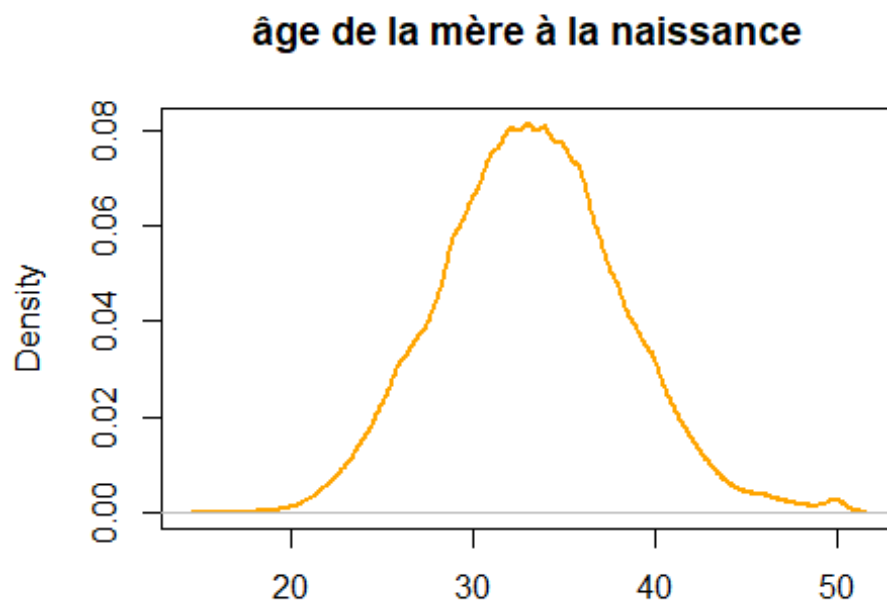
```
summary(select(traites,fagerec11))
```

```
##      fagerec11
##  Min.       : 2.000
##  1st Qu.:  5.000
##  Median    :  6.000
##  Mean      :  5.655
##  3rd Qu.:  6.000
##  Max.      :10.000
```

```
round(table(traites$fagerec11)/length(traites$fagerec11),3)
```

```
##
##      2      3      4      5      6      7      8      9     10
## 0.000 0.019 0.137 0.330 0.305 0.137 0.045 0.017 0.010

plot(density(traites$mager), lwd = 2, col = "orange", xlab = "", main = "âge
de la mère à la naissance")
```



Les parents sont plus jeunes à la naissance de l'enfant sur la population de contrôle

```
summary(select(control,mager))

##      mager
## Min.   :12.00
## 1st Qu.:26.00
## Median :30.00
## Mean   :29.72
## 3rd Qu.:34.00
## Max.   :50.00

summary(select(control,fagerec11))

##      fagerec11
## Min.   : 1.00
## 1st Qu.: 4.00
## Median : 5.00
## Mean   : 4.99
## 3rd Qu.: 6.00
## Max.   :10.00
```

```
round(table(control$fagerec11)/length(control$fagerec11),3)
```

```
##  
##      1      2      3      4      5      6      7      8      9     10  
## 0.000 0.015 0.113 0.233 0.311 0.211 0.081 0.024 0.008 0.004
```

En ce qui concerne les mères : celles qui ont recours au traitement sont plus âgées de 3 à 4 ans en moyenne. Quant aux pères, 63% de la population traitée sont âgés de 30 à 39 ans contre 52% dans le groupe de contrôle. Néanmoins, la différence est surtout flagrante du côté des pères âgés de moins de 30 ans : 16% chez les traités contre 35%.

Ethnie des parents

Race de la mère :

1 White (only)
3 AIAN (American Indian or Alaskan Native) (only)
4 Asian (only) 5 NHOPI (Native Hawaiian or Other Pacific Islander) (only)
6 More than one race

Ethnie du père :

1 White (only)
2 Black (only)
3 AIAN (only)
4 Asian (only)
5 NHOPI (only)
6 More than one race
9 Unknown or Not Stated

```
summary(select(traites,mrace6))
```

```
##      mrace6  
## Min.      :10.00  
## 1st Qu.:10.00  
## Median :10.00  
## Mean    :13.89  
## 3rd Qu.:10.00  
## Max.     :61.00
```

```
summary(select(traites,frace6))
```

```
##      frace6  
## Min.      :1.000  
## 1st Qu.:1.000  
## Median :1.000  
## Mean     :1.292  
## 3rd Qu.:1.000  
## Max.     :5.000
```



```
round(table(traites$frace6)/length(traites$frace6),3)
```

```
##  
##      1      2      3      4      5  
## 0.860 0.063 0.003 0.072 0.002
```

```
round(table(traites$mrace6)/length(traites$mrace6),3)
```

```
##  
##     10     20     30     40     41     51     61  
## 0.847 0.054 0.002 0.066 0.012 0.001 0.018
```

La part des pères blancs est toujours prédominante si on considère toute la population, mais elle l'est moins que chez la population cible et surtout la part des pères noirs est bien plus importante

```
summary(select(control, mrace6))
```

```
##      mrace6  
## Min.   :10.00  
## 1st Qu.:10.00  
## Median :10.00  
## Mean   :14.82  
## 3rd Qu.:10.00  
## Max.   :61.00
```

```
summary(select(control, frace6))
```

```
##      frace6  
## Min.   :1.000  
## 1st Qu.:1.000  
## Median :1.000  
## Mean   :1.379  
## 3rd Qu.:1.000  
## Max.   :5.000
```

```
round(table(control$frace6)/length(control$frace6),3)
```

```
##  
##      1      2      3      4      5  
## 0.765 0.161 0.008 0.062 0.004
```

```
round(table(control$mrace6)/length(control$mrace6),3)
```

```
##  
##     10     20     30     40     41     51     61  
## 0.764 0.134 0.008 0.054 0.015 0.003 0.022
```

Traités : 86 % des pères sont blancs, 7% asiatiques et 6% noirs Contrôle : 77 % des pères sont blancs, 16% noirs et 7% asiatiques Les taux sont similaires pour les mères.

Education des parents

Éducation de la mère :

- 1 8th grade or less
- 2 9th through 12th grade with no diploma
- 3 High school graduate or GED completed
- 4 Some college credit, but not a degree.
- 5 Associate degree (AA,AS)
- 6 Bachelor's degree (BA, AB, BS)
- 7 Master's degree (MA, MS, MEng, MEd, MSW, MBA)
- 8 Doctorate (PhD, EdD) or Professional Degree (MD, DDS,DVM, LLB, JD)
- 9 Unknown

Education du père :

- 1 8th grade or less
- 2 9th through 12th grade with no diploma
- 3 High school graduate or GED completed
- 4 Some college credit, but not a degree.
- 5 Associate degree (AA,AS)
- 6 Bachelor's degree (BA, AB, BS)
- 7 Master's degree (MA, MS, MEng, MEd, MSW, MBA)
- 8 Doctorate (PhD, EdD) or Professional Degree (MD, DDS,DVM, LLB, JD)
- 9 Unknown

```
summary(select(traites,meduc))
```

```
##      meduc
##  Min.    :1.0
## 1st Qu.:5.0
##  Median :6.0
##   Mean   :5.8
## 3rd Qu.:7.0
##   Max.   :8.0
```

```
summary(select(traites,feduc))
```

```
##      feduc
##  Min.    :1.000
## 1st Qu.:4.000
##  Median :6.000
##   Mean   :5.452
## 3rd Qu.:7.000
##   Max.   :8.000
```

```
round(table(traites$meduc)/length(traites$meduc),3)
```

```
##
##      1      2      3      4      5      6      7      8
## 0.003 0.008 0.076 0.122 0.090 0.373 0.247 0.081
```

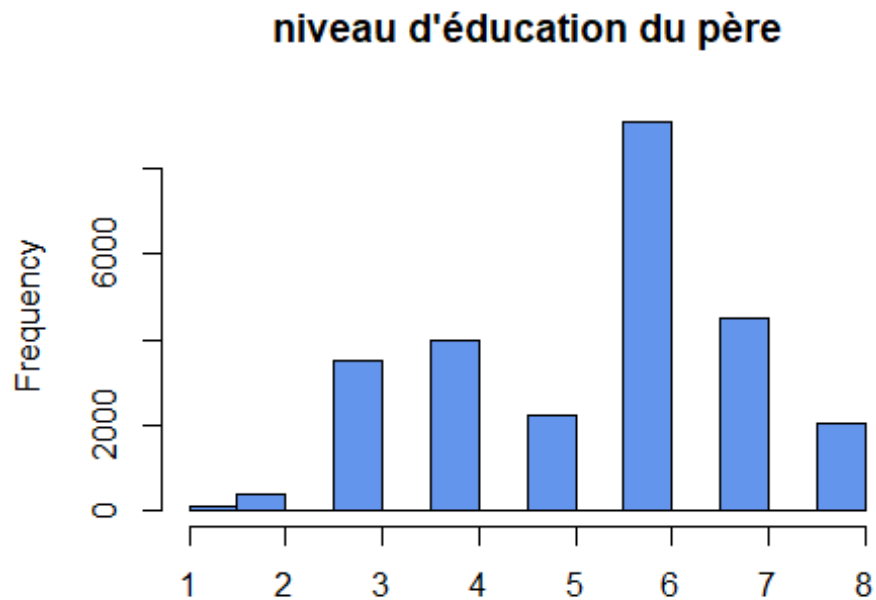
```
round(table(traites$feduc)/length(traites$feduc),3)
```

```
##
```

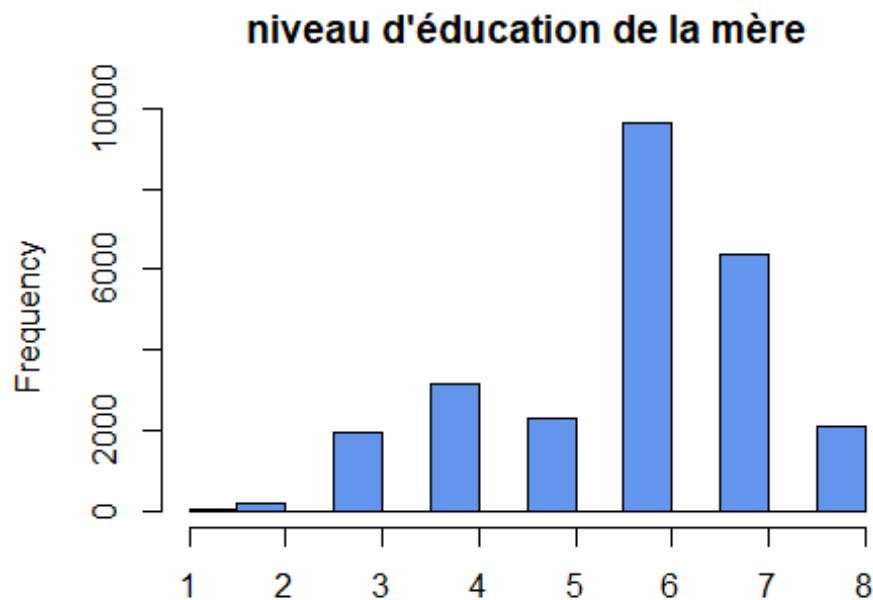
```
##      1      2      3      4      5      6      7      8
```

```
## 0.004 0.014 0.135 0.155 0.087 0.353 0.174 0.079
```

```
hist(traites$feduc, col = "cornflowerblue", xlab = "", main = "niveau d'éducation du père")
```



```
hist(traites$meduc, col = "cornflowerblue", xlab = "", main = "niveau d'éducation de la mère")
```



Sur la population entière, le niveau d'éducation des parents est bien moins élevé : en moyenne les parents qui ont recours au traitement ont un Bachelor's degree, alors que ceux qui n'y ont pas recours ont "Some college credit, but not a degree".

L'écart se voit en particulier à la part de haut niveau d'éducation : plus de 25% des parents de la population traitée ont un master ou un doctorat, tandis que sont plutôt 10% à 15 % des parents du groupe de contrôle.

```
summary(select(control,meduc))
```

```
##      meduc
##  Min.   :1.000
## 1st Qu.:3.000
##  Median :4.000
##   Mean  :4.673
## 3rd Qu.:6.000
##   Max.  :8.000
```

```
summary(select(control,feduc))
```

```
##      feduc
##  Min.   :1.000
## 1st Qu.:3.000
##  Median :4.000
##   Mean  :4.377
## 3rd Qu.:6.000
##   Max.  :8.000
```

```
round(table(control$meduc)/length(control$meduc),3)
```

```
##
```

```
##      1      2      3      4      5      6      7      8  
## 0.022 0.059 0.236 0.185 0.091 0.251 0.120 0.035
```

```
round(table(control$feduc)/length(control$feduc),3)
```

```
##
```

```
##      1      2      3      4      5      6      7      8  
## 0.027 0.072 0.305 0.182 0.076 0.218 0.083 0.036
```

Caractéristiques du couple

Statut marital :

1 Married

2 Unmarried

Enfant(s) encore en vie :

00-30 Number of children still living from previous live births.

99 Unknown or not stated

```
summary(select(traites,dmar))
```

```
##      dmar
```

```
## Min.   :1.00
```

```
## 1st Qu.:1.00
```

```
## Median :1.00
```

```
## Mean   :1.05
```

```
## 3rd Qu.:1.00
```

```
## Max.   :2.00
```

```
## NA's   :1447
```

```
summary(select(traites,priorlive))
```

```
##      priorlive
```

```
## Min.   : 0.0000
```

```
## 1st Qu.: 0.0000
```

```
## Median : 0.0000
```

```
## Mean    : 0.6149
```

```
## 3rd Qu.: 1.0000
```

```
## Max.    :10.0000
```

```
summary(select(traites,dplural))
```

```
##      dplural
```

```
## Min.   :1.00
```

```
## 1st Qu.:1.00
```

```
## Median :1.00
```

```
## Mean :1.19
## 3rd Qu.:1.00
## Max. :4.00

round(table(traites$dmr)/length(traites$dmr),3)

##
## 1 2
## 0.897 0.047

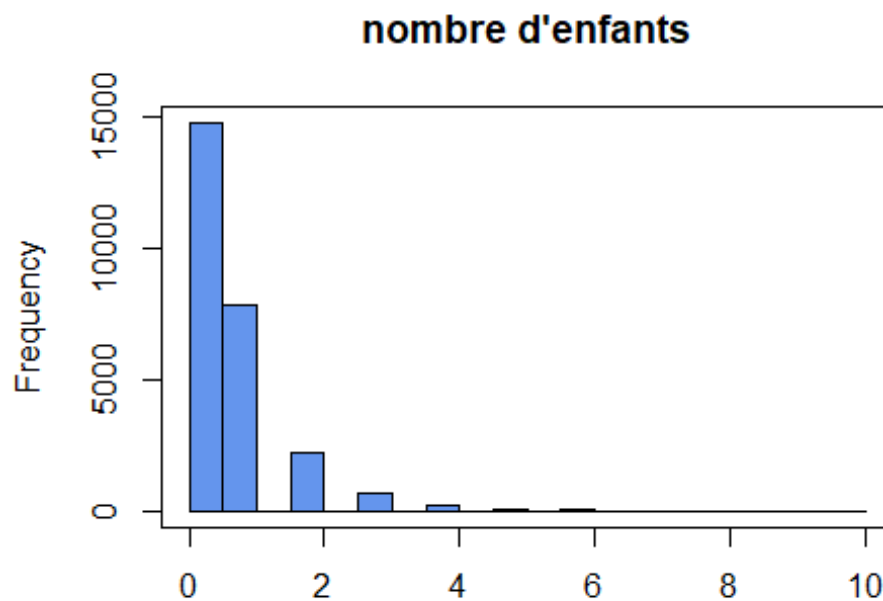
round(table(traites$priorlive)/length(traites$priorlive),3)

##
## 0 1 2 3 4 5 6 7 8 9 10
## 0.572 0.303 0.085 0.026 0.008 0.003 0.001 0.001 0.000 0.000 0.000

round(table(traites$dplur)/length(traites$dplur),3)

##
## 1 2 3 4
## 0.828 0.157 0.014 0.002

hist(traites$priorlive, col = "cornflowerblue", xlab = "", main = "nombre d'
enfants")
box()
```



Moins de couples sont mariés et plus ont déjà eu des enfants auparavant mais surtout, beaucoup moins de naissance de jumeaux

```
summary(select(control,dmar))
```

```
##      dmar
## Min.   :1.00
## 1st Qu.:1.00
## Median :1.00
## Mean   :1.31
## 3rd Qu.:2.00
## Max.   :2.00
## NA's   :307294
```

```
summary(select(control,priorlive))
```

```
##      priorlive
## Min.   : 0.000
## 1st Qu.: 0.000
## Median : 1.000
## Mean   : 1.106
## 3rd Qu.: 2.000
## Max.   :23.000
```

```
summary(select(control,dplural))
```

```
##      dplural
## Min.   :1.000
## 1st Qu.:1.000
## Median :1.000
## Mean   :1.029
## 3rd Qu.:1.000
## Max.   :4.000
```

```
round(table(control$dmar)/length(control$dmar),3)
```

```
##
##      1      2
## 0.613 0.277
```

```
round(table(control$priorlive)/length(control$priorlive),3)
```

```
##
##      0      1      2      3      4      5      6      7      8      9     10     11
## 0.386 0.330 0.168 0.069 0.026 0.011 0.005 0.002 0.001 0.001 0.000 0.000
##     12     13     14     15     16     17     18     19     20     21     22     23
## 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000
```

```
round(table(control$dplural)/length(control$dplural),3)
```

```
##
##      1      2      3      4
## 0.971 0.028 0.001 0.000
```

Part des parents mariés : 90% VS 61%

Part des mères dont c'est le premier enfant : 58% VS 39%

Part des mères ayant déjà deux enfants ou plus : 13% VS 28%

Part des jumeaux : 16% VS 3%

Santé de l'enfant

Score Apgar :

1 A score of 0-3

2 A score of 4-6

3 A score of 7-8

4 A score of 9-10

5 Unknown or not stated

Poids à la naissance :

0227-8165 Number of grams

9999 Not stated birth weight

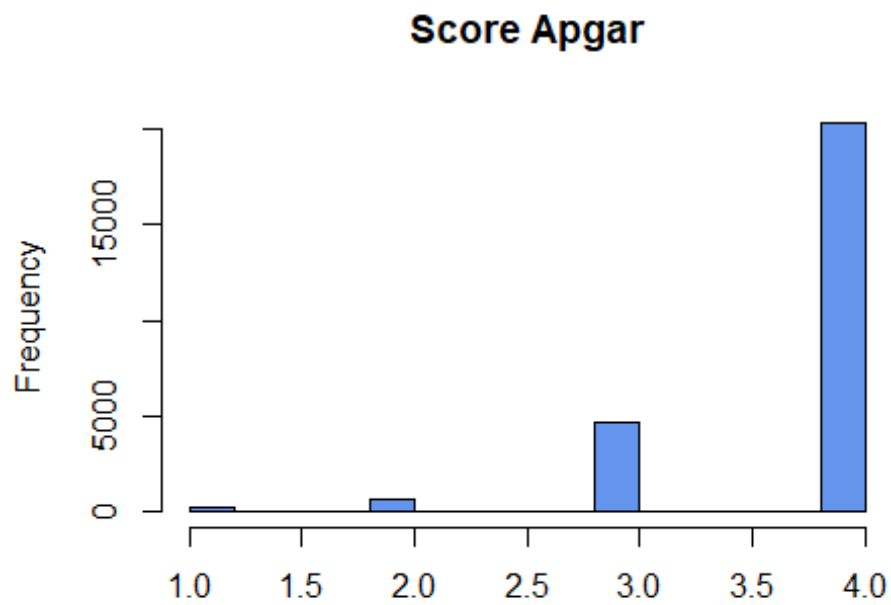
```
summary(select(traites, apgar5r))
```

```
##      apgar5r
##  Min.    :1.000
## 1st Qu.:4.000
##  Median :4.000
##   Mean   :3.749
## 3rd Qu.:4.000
##   Max.   :4.000
```

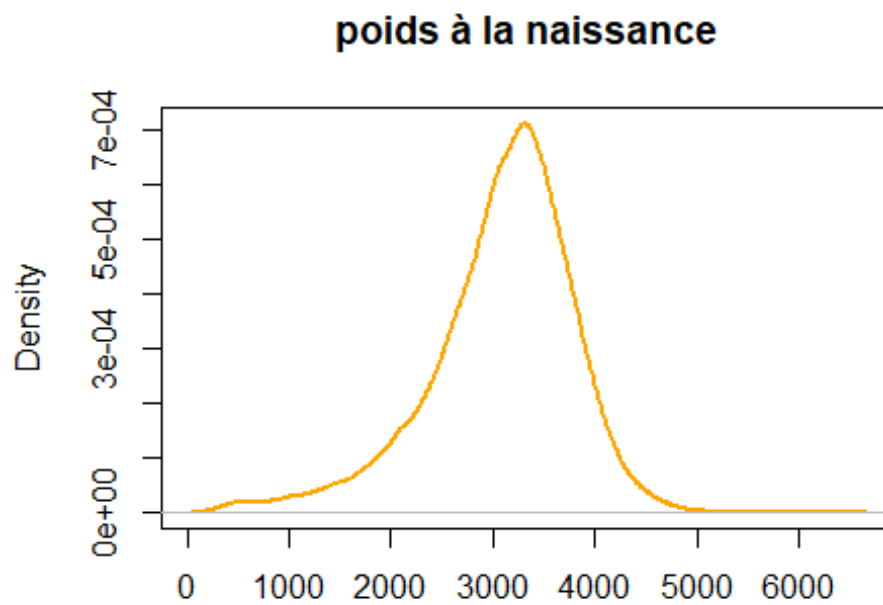
```
summary(select(traites, dbwt))
```

```
##      dbwt
##  Min.    : 234
## 1st Qu.:2730
##  Median :3180
##   Mean   :3085
## 3rd Qu.:3545
##   Max.   :6435
```

```
hist(traites$apgar5r, col = "cornflowerblue", xlab = "", main = "Score Apgar")
```

```
plot(density(traites$dbwt), lwd = 2, col = "orange", xlab = "", main = "poids  
à la naissance")
```



Le poids à la naissance est plus élevé en moyenne mais le score Apgar est comparable

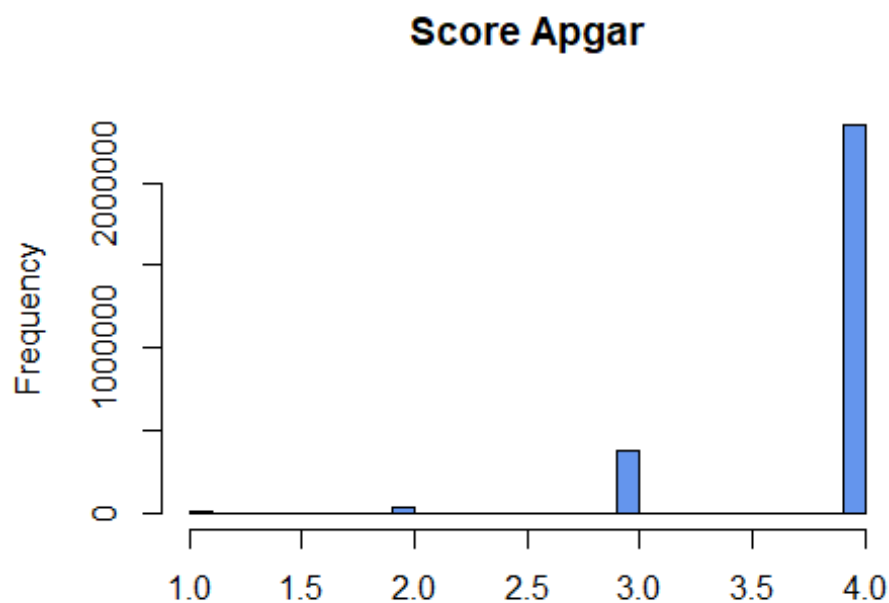
```
summary(select(control, apgar5r))
```

```
##      apgar5r
## Min.   :1.00
## 1st Qu.:4.00
## Median :4.00
## Mean   :3.82
## 3rd Qu.:4.00
## Max.   :4.00
```

```
summary(select(control, dbwt))
```

```
##      dbwt
## Min.   : 227
## 1st Qu.:2977
## Median :3317
## Mean   :3278
## 3rd Qu.:3635
## Max.   :8142
```

```
hist(control$apgar5r, col = "cornflowerblue", xlab = "", main = "Score Apgar")
```



```
plot(density(control$dbwt), lwd = 2, col = "orange", xlab = "", main = "poids à la naissance")
```

poids à la naissance

