

# Introduction à R

## Les premières données

Pascal Bessonneau

06/2015

Les iris de Fisher

Les prénoms

# Les iris de Fisher

Les iris de Fisher sont des données très connues dans le milieu des statisticiens.

Ils constituent un jeu de données sur lesquelles on utilise des méthodes de classification notamment.

Ce sont les caractéristiques morphologiques des feuilles pour quelques espèces d'Iris.

# Les iris de Fisher

```
> data(iris)
> class(iris)

## [1] "data.frame"
```

# Les iris de Fisher

```
> summary(iris)

##      Sepal.Length      Sepal.Width
##  Min.       :4.300    Min.       :2.000
##  1st Qu.:5.100    1st Qu.:2.800
##  Median :5.800    Median :3.000
##  Mean     :5.843    Mean     :3.057
##  3rd Qu.:6.400    3rd Qu.:3.300
##  Max.     :7.900    Max.     :4.400
##      Petal.Length      Petal.Width
##  Min.       :1.000    Min.       :0.100
##  1st Qu.:1.600    1st Qu.:0.300
##  Median :4.350    Median :1.300
##  Mean     :3.758    Mean     :1.199
##  3rd Qu.:5.100    3rd Qu.:1.800
##  Max.     :6.900    Max.     :2.500
##           Species
##  setosa      :50
##  versicolor:50
##  virginica   :50
##
```

# Les iris de Fisher

```
> table(iris$Species)

##
##      setosa versicolor  virginica
##          50          50          50

> prop.table(table(iris$Species))

##
##      setosa versicolor  virginica
## 0.3333333 0.3333333 0.3333333

> prop.table(table(iris$Species))*100

##
##      setosa versicolor  virginica
## 33.33333  33.33333  33.33333
```

# Les iris de Fisher

```
> tapply(iris$Sepal.Length, iris$Species, mean)
```

```
##      setosa versicolor  virginica  
##      5.006      5.936      6.588
```

# Les prénoms à Paris

Ce sont les prénoms des nouveaux nés à Paris. Ils viennent de [opendata.paris.fr](https://opendata.paris.fr).

Le but ici est de manipuler et d'extraire les données.



# Les prénoms à Paris

Pour charger le fichier :

```
> prenomes <- read.csv2("data/premoms/liste_des_premoms_2004_a_2012.csv"
+                       stringsAsFactors = F,encoding = "UTF-8")
> p <- fromJSON("data/premoms/liste_des_premoms_2004_a_2012.json")$fiel
>
> colnames(prenomos)

## [1] "Prenoms" "Nombre"  "Sexe"    "Annee"
```

## Les prénoms à Paris

Pour récupérer les prénoms de 2004 :

```
> prenom2004 <- prenom[prenom$Annee==2004,]
```

## Les prénoms à Paris

Quel est le prénom le plus fréquent ?

```
> max(prenoms$Nombre)
```

```
## [1] 398
```

```
> prenoms2004$Prenoms[prenoms2004$Nombre == max(prenoms2004$Nombre)]
```

```
## [1] "Alexandre"
```

## Les prénoms à Paris

Quel est le prénom le moins fréquent ?

```
> prenom2004$Prenoms[prenom2004$Nombre == min(prenom2004$Nombre)]
```

```
##      [1] "Khalil"      "Leana"       "Loubna"
##      [4] "Morgan"      "Natalia"     "Oussama"
##      [7] "Safa"        "Sharon"      "Solenn"
##     [10] "Sylvia"      "Viktor"      "Virgil"
##     [13] "Wandrille"   "Warren"      "Alexane"
##     [16] NA            "Camelia"     "Carl"
##     [19] "Chanel"      "Filipe"      "Halima"
##     [22] "Henry"       "Iban"        "Jawad"
##     [25] "Josh"        "Adil"        "Bahia"
##     [28] "Boubou"      "Clothilde"   "Dana"
##     [31] "Daria"       "Gabriella"   "Harold"
##     [34] "Hasna"       NA            "Latifa"
##     [37] "Louka"       "Mory"        "Nesrine"
##     [40] "Niouma"     "Rami"        "Ramy"
##     [43] "Reda"        "Sebastian"   "Tim"
##     [46] "Wendy"       NA            "Aboubakar"
##     [49] "Adeline"     "Aymane"     "Benoit"
```

# Les prénoms à Paris

Quel est le minimum ?

```
> min(prenoms2004$Nombre)
```

```
## [1] 6
```

## Les prénoms à Paris

Choisissez un prénom et trouver le nombre correspondants :

```
> prenom2004$Nombre[prenom2004$Prenoms=="Pascal"]
```

```
## [1] 11
```

## Les prénoms à Paris

Trouver le prénoms qui ont disparus entre ces années. Cela revient à faire une table et à chercher les prénoms qui apparaissent moins de 8 fois.

## Les prénoms à Paris

```
> tt <- table(prenoms$Prenoms)
> range(tt)

## [1] 1 26
```



## Les prénoms à Paris

Oups y'a une petit problème dans la base de données.

```
> range(prenoms$Annee)
```

```
## [1] 2004 2015
```

```
> head(names(tt)[tt>12])
```

```
## [1] "Adama"      "Alix"       "Amelia"
```

```
## [4] "Andrea"    "Ange"      "Angelina"
```

## Les prénoms à Paris

Oups y'a une petit problème dans la base de données.

```
> pp <- names(tt)[tt>12]
> prenomns_prb <- prenomns[prenomns$Prenoms %in% pp,]
> head(prenomns_prb[order(prenomns_prb$Prenoms),])
```

##	Prenoms	Nombre	Sexe	Annee
## 1268	Adama	5	F	2014
## 2188	Adama	17	M	2008
## 3320	Adama	18	M	2004
## 4817	Adama	15	M	2011
## 5884	Adama	5	M	2012
## 5952	Adama	13	M	2012

# Les prénoms à Paris

```
> head(table(prenoms_prb$Prenoms,prenoms_prb$Annee))
```

```
##
##           2004 2005 2006 2007 2008 2009
## Adama        1    1    1    1    1    1
## Alix         1    1    1    1    1    1
## Amelia       1    1    1    2    0    1
## Andrea       2    2    2    2    2    2
## Ange         1    1    1    1    1    1
## Angelina     2    2    2    2    2    1
##
##           2010 2011 2012 2013 2014 2015
## Adama        1    1    2    1    2    2
## Alix         1    2    2    2    2    2
## Amelia       2    2    2    1    0    1
## Andrea       2    3    3    2    2    2
## Ange         1    1    1    2    2    2
## Angelina     1    2    1    1    1    1
```

## Les prénoms à Paris

```
> prenomns_prb[prenoms_prb$Prenoms=="Andrea" & prenomns_prb$Annee == 2012
```

##		Prenoms	Nombre	Sexe	Annee
##	6307	Andrea	16	F	2012
##	6308	Andrea	35	F	2012
##	6309	Andrea	11	F	2012