TP2

2022-10-10

Plots and Given Names

We red the prenoms.csv doc and we plot these information: – The number of births by year:

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.7      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
date <- read.csv("prenoms.csv",header=TRUE)
```

– The number of births by year:

```
annee <- date %>% group_by(Année) %>%
  summarise(nbr = sum(Nombre))
ggplot(annee,aes(Année,nbr))+geom_point()
```
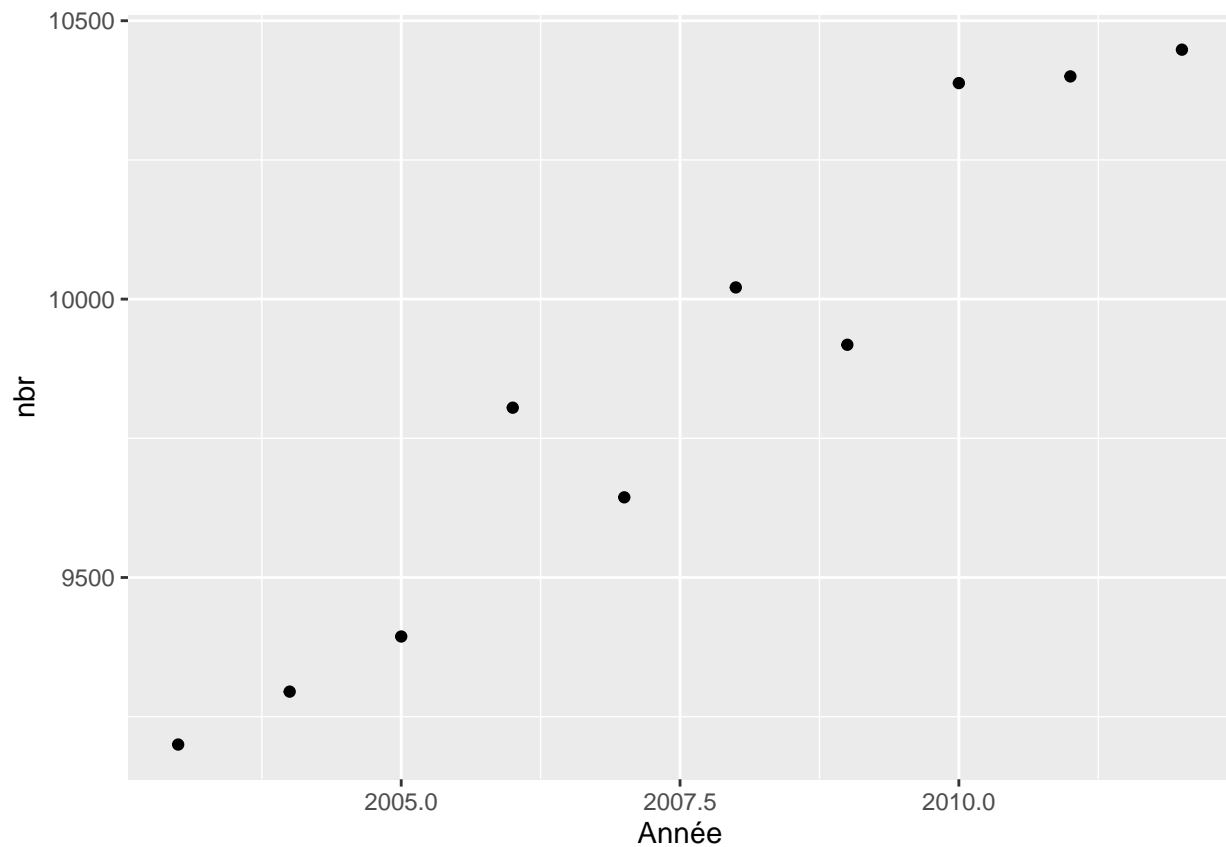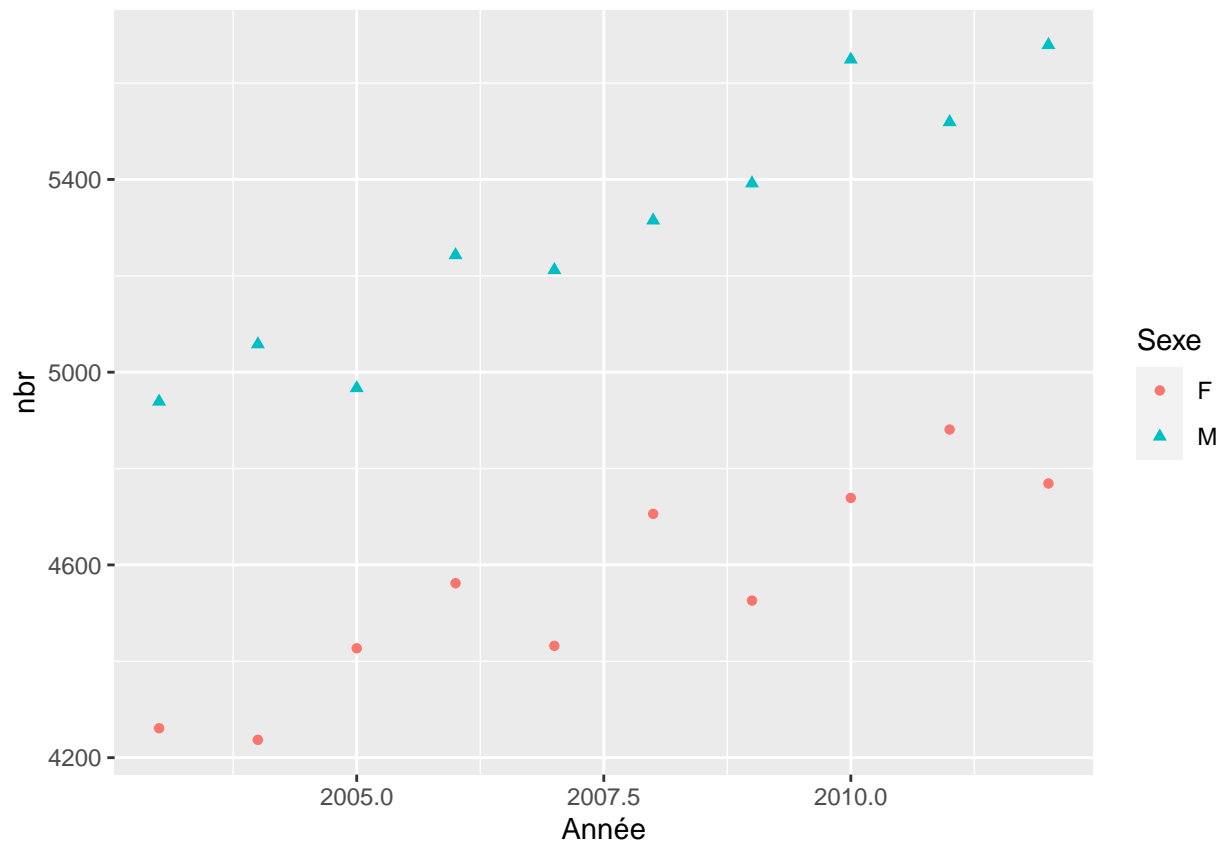
The number of male/female births by year

```
tri2 <- date %>% group_by(Année, Sexe) %>%
  summarise(nbr=sum(Nombre))
```

```
## `summarise()` has grouped output by 'Année'. You can override using the
## `.groups` argument.
```

```
ggplot(tri2,aes(Année,nbr,color=Sexe,shape=Sexe))+geom_point()
```

Is your name in the dataset ?

```
date[date$Prénom == "Clémence",]
```

```
##          Année     Prénom Nombre Sexe Ordre
## 329      2011 Clémence     47    F    56
## 493      2003 Clémence     21    F   117
## 662      2008 Clémence     42    F    61
## 859      2010 Clémence     44    F    60
## 1016     2012 Clémence     41    F    64
## 1172     2007 Clémence     34    F    80
## 1805     2005 Clémence     37    F    69
## 2418     2006 Clémence     38    F    72
## 3508     2009 Clémence     48    F    52
## 5081     2004 Clémence     36    F    76
```

– Represent the 10 most given names

```
famous <- date %>% group_by(Prénom) %>%
  summarise(nbr = sum(Nombre))
head(famous[order(-famous$nbr),],10)
```

```
## # A tibble: 10 x 2
##     Prénom    nbr
##     <chr>   <int>
##   1 Emma     1259
##   2 Lucas    1183
##   3 Enzo     1121
##   4 Manon     994
##   5 Clément   975
```

```
##  6 Thomas      954
##  7 Nathan      933
##  8 Camille     911
##  9 Louis       910
## 10 Maxime      900
```

– Select for each year the top 5 given names by sex and represent their evolution along years.

– Plot the average number of letters by year – Plot the average number of vowels/consonants by year – How the number of composed names (like Jean-Baptiste or Lou-Ann – Define a "hype" criteria and find the hypest names

```
summary(cars)
```

```
##      speed           dist
##  Min.   : 4.0   Min.   :  2.00
##  1st Qu.:12.0   1st Qu.: 26.00
##  Median :15.0   Median : 36.00
##  Mean   :15.4   Mean   : 42.98
##  3rd Qu.:19.0   3rd Qu.: 56.00
##  Max.   :25.0   Max.   :120.00
```

Including Plots

You can also embed plots, for example:
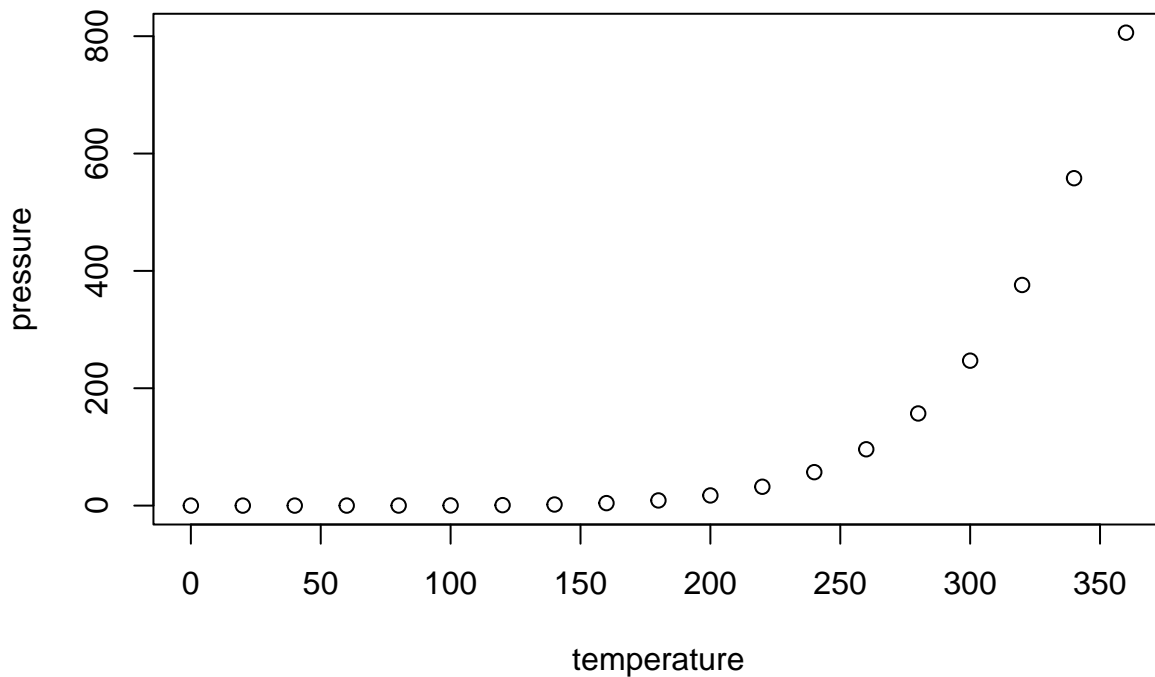
Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.