Projet Data Science: Analyse des données

Anova qui montre que le thème et le moment de la journée ont un effet sur la taille du sujet

```
library(FactoMineR)
library(dplyr)
```

```
##
## 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
```

```
# Anova qui montre que le thème et le moment de la journée ont un effet sur la taille du sujet

dataset <- read.csv("C:/Users/imadl/Desktop/IG4/ProjetDataScience/df3.csv")
df <- subset(dataset,select = c(Moment, subject_size,theme))
df$Moment <- factor(df$Moment)
df$theme <- factor(df$theme)
anov <- aov(df$subject_size~df$Moment*df$theme)
summary(anov)
```

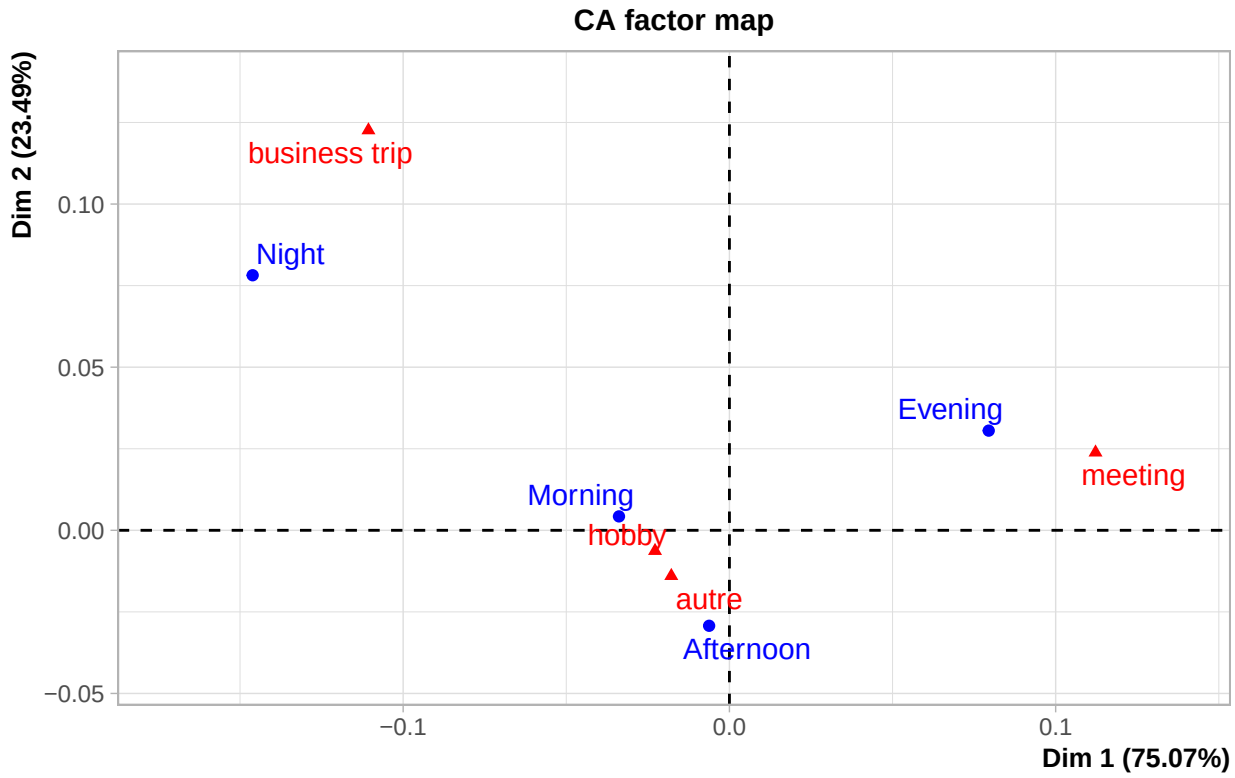
```
##                       Df   Sum Sq Mean Sq F value Pr(>F)
## df$Moment              3   348873  116291   264.3 <2e-16 ***
## df$theme               3   710517  236839   538.2 <2e-16 ***
## df$Moment:df$theme     9    56251    6250    14.2 <2e-16 ***
## Residuals          99984 43996663     440
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AFC pour mettre en évidence une correlation entre le thème du mail et le moment de la journée

```
df2 <- subset(dataset,select = c(Moment,theme))%>%
 mutate(Moment = factor(Moment, levels = c(1,2,3,4), labels = c('Morning', 'Afternoon','Evening','Night
cont <- table(df2)
print(cont)
```

```
##            theme
## Moment      autre business trip hobby meeting
##   Morning   17330          1327  2557    4155
##   Afternoon 30309          1909  4678    7507
##   Evening   15916          1137  2448    5109
##   Night      3806           431   629     752
```

```r
AFC <- CA(cont)
```



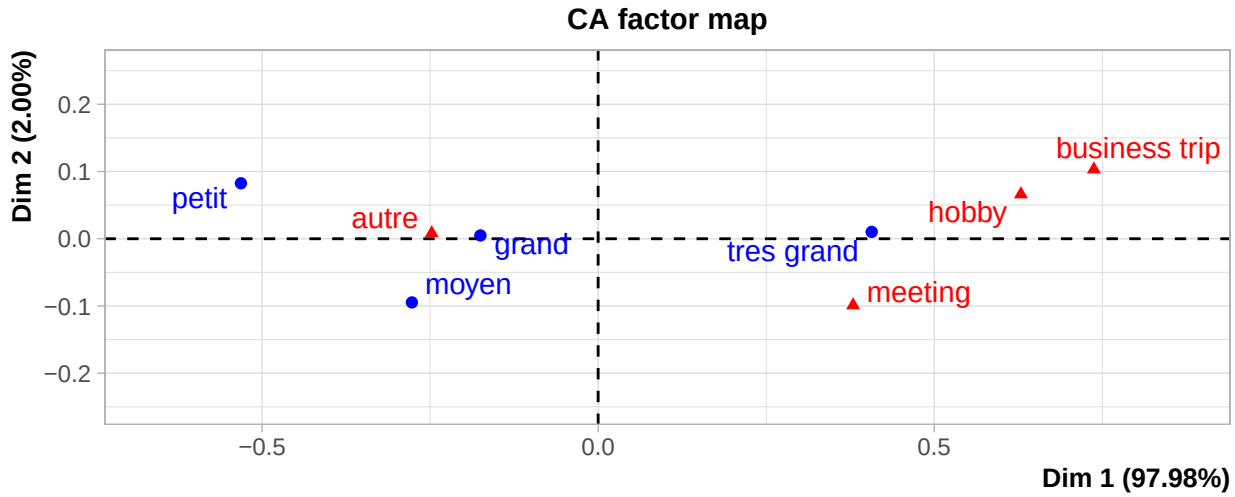

```r
summary(AFC)
```

```
##
## Call:
## CA(X = cont)
##
## The chi square of independence between the two variables is equal to 407.8655 (p-value =  2.87199e-8
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3
## Variance               0.003   0.001   0.000
## % of var.             75.072  23.487   1.441
## Cumulative % of var.  75.072  98.559 100.000
##
## Rows
##                 Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## Morning       |     0.333 | -0.034  9.493  0.872 |  0.004  0.481  0.014 |
## Afternoon     |     0.406 | -0.006  0.561  0.042 | -0.029 39.690  0.936 |
## Evening       |     1.785 |  0.079 50.766  0.871 |  0.031 23.984  0.129 |
## Night         |     1.554 | -0.146 39.180  0.772 |  0.078 35.845  0.221 |
##                Dim.3    ctr   cos2
## Morning       -0.012 64.657  0.114 |
## Afternoon      0.005 15.346  0.022 |
```

```
## Evening         0.001  0.640  0.000 |
## Night           0.014 19.357  0.007 |
##
## Columns
##                  Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## autre          |     0.352 | -0.018  6.978  0.606 | -0.014 13.754  0.374 |
## business trip  |     1.311 | -0.111 19.206  0.449 |  0.123 75.405  0.551 |
## hobby          |     0.109 | -0.023  1.751  0.491 | -0.006  0.428  0.038 |
## meeting        |     2.306 |  0.112 72.065  0.957 |  0.024 10.412  0.043 |
##                  Dim.3    ctr   cos2
## autre          -0.003 11.907  0.020 |
## business trip -0.003  0.585  0.000 |
## hobby           0.022 87.508  0.471 |
## meeting         0.000  0.000  0.000 |
```

AFC pour mettre en évidence une correlation entre la taille du mail et le thème

```
dataset_quali <- read.csv("C:/Users/imadl/Desktop/IG4/ProjetDataScience/df9.csv")
df3 <- subset(dataset_quali,select = c(size_mail_quali,theme))
cont <- table(df3)
AFC2 <- CA(cont)
```




```
summary(AFC2)
```

```
##
```

```
## Call:
## CA(X = cont)
##
## The chi square of independence between the two variables is equal to 13612.59 (p-value =  0 ).
##
## Eigenvalues
##                        Dim.1   Dim.2   Dim.3
## Variance               0.133   0.003   0.000
## % of var.             97.983   2.001   0.016
## Cumulative % of var.  97.983  99.984 100.000
##
## Rows
##                  Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## grand         |     7.472 | -0.175  5.587  0.997 |  0.005  0.216  0.001 |
## moyen         |    15.982 | -0.277 10.727  0.895 | -0.095 61.399  0.105 |
## petit         |    42.502 | -0.531 31.112  0.976 |  0.083 36.758  0.024 |
## tres grand    |    70.169 |  0.407 52.574  0.999 |  0.010  1.628  0.001 |
##                 Dim.3    ctr   cos2
## grand         -0.008 69.884  0.002 |
## moyen          0.003  9.228  0.000 |
## petit          0.005 17.432  0.000 |
## tres grand     0.001  3.456  0.000 |
##
## Columns
##                  Iner*1000    Dim.1    ctr   cos2    Dim.2    ctr   cos2
## autre         |    41.315 | -0.248 30.941  0.999 |  0.008  1.680  0.001 |
## business trip |    26.649 |  0.737 19.587  0.980 |  0.103 18.786  0.019 |
## hobby         |    41.244 |  0.629 30.577  0.989 |  0.066 16.581  0.011 |
## meeting       |    26.918 |  0.379 18.895  0.936 | -0.099 62.953  0.064 |
##                 Dim.3    ctr   cos2
## autre          0.000  0.018  0.000 |
## business trip  0.016 56.823  0.000 |
## hobby         -0.009 42.530  0.000 |
## meeting        0.001  0.629  0.000 |
```