

Analyse des correlations 1

Nous allons dans cette partie mettre en avant les liens existant entre les variables liées à la taille du contenu/sujet du mail et le moment de la journée où il a été envoyé

```
mails_anova1_1 <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/anova1_1.csv")
mails_anova1_2 <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/anova1_2.csv")
mails_anova2 <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/dataset_anova_2.csv")
mails_afc1_1_non_transmit <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/dataset_afc1_1.csv")
mails_afc1_2 <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/dataset_afc1_2.csv")
mails_afc2 <- read.csv("/Users/pierreperrin/Desktop/IG4/semestre7/Projet_DataScience/dataset_afc_2.csv")
```

load library

```
library(FactoMineR)
```

Première analyse : corrélation entre taille du mail (en caractère) et moment de la journée où il est envoyé

Première anova à un facteur: "content_size" et "moment"

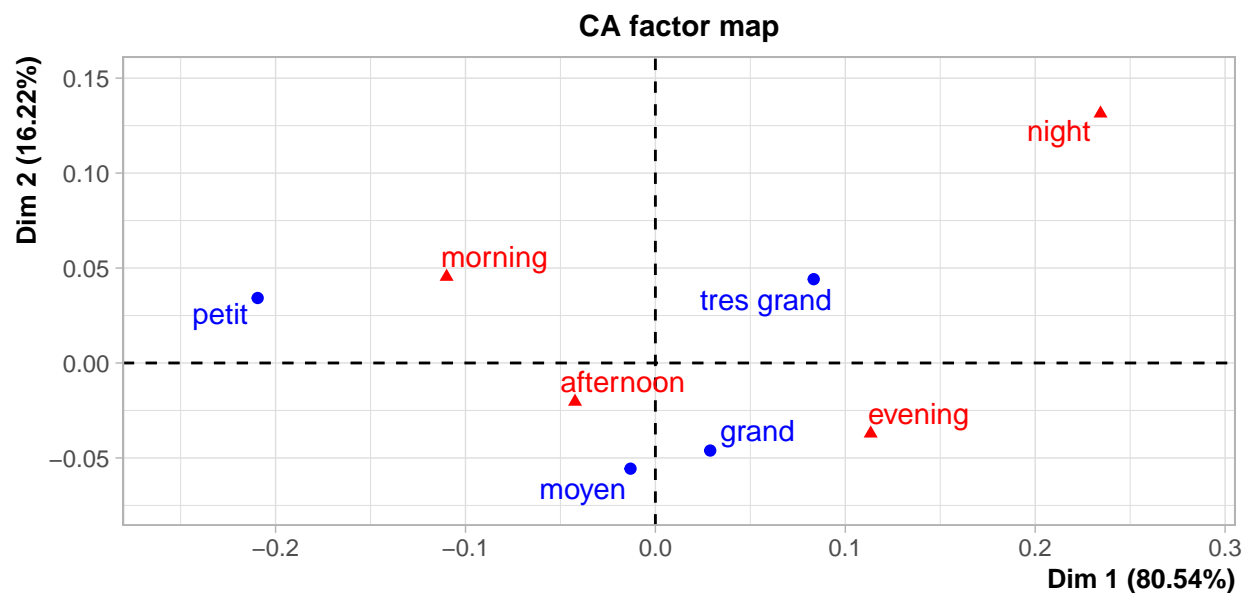
```
summary(aov(mails_anova1_1$content_size~mails_anova1_1$Moment))
```

```
##               Df    Sum Sq   Mean Sq F value Pr(>F)
## mails_anova1_1$Moment      3 3.083e+10 1.028e+10   190.3 <2e-16 ***
## Residuals                99996 5.401e+12 5.401e+07
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

le moment de la journée suivant lequel on envoie le mail influe sur la taille du mail.

AFC pour voir le lien entre le moment de la journée et la taille (petit, moyen, grand, très grand) :

```
table(mails_afc1_1_non_transmit[,1], mails_afc1_1_non_transmit[,2]) -> afc_1
afc_moment_content_size <- CA(afc_1)
```



```
summary(afc_moment_content_size)
```

```
##
## Call:
## CA(X = afc_1)
##
## The chi square of independence between the two variables is equal to 1105.969 (p-value = 2.390135e-1)
##
## Eigenvalues
##           Dim.1   Dim.2   Dim.3
## Variance      0.010   0.002   0.000
## % of var.     80.542  16.220   3.238
## Cumulative % of var. 80.542  96.762 100.000
##
## Rows
##           Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr   cos2   Dim.3
## grand      |      0.920 |  0.029  1.930  0.218 | -0.046 24.420  0.556 | -0.029
## moyen      |      0.880 | -0.013  0.342  0.040 | -0.056 30.666  0.730 |  0.031
## petit      |      7.731 | -0.209 72.329  0.974 |  0.034  9.585  0.026 | -0.005
## tres grand |      3.388 |  0.083 25.399  0.780 |  0.044 35.329  0.218 |  0.004
##
##           ctr   cos2
## grand      49.571  0.225 |
## moyen      48.245  0.229 |
## petit       0.929  0.001 |
## tres grand  1.254  0.002 |
```

```
##
## Columns
##          Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr   cos2   Dim.3
## afternoon |      1.137 | -0.042  7.572  0.693 | -0.020  8.712  0.161 | -0.019
## evening   |      3.854 |  0.113 32.599  0.880 | -0.037 17.350  0.094 |  0.019
## morning   |      3.427 | -0.110 27.231  0.827 |  0.045 23.067  0.141 |  0.022
## night     |      4.501 |  0.234 32.599  0.754 |  0.131 50.871  0.237 | -0.026
##          ctr   cos2
## afternoon 39.723 0.146 |
## evening  23.660 0.026 |
## morning  26.260 0.032 |
## night    10.357 0.010 |
```

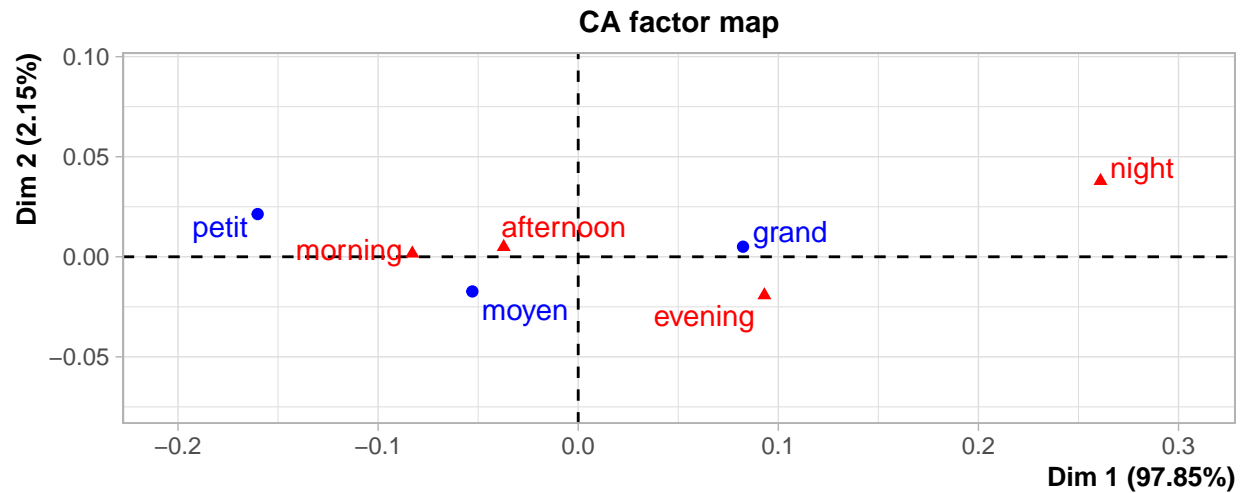
Deuxième analyse : corrélation entre taille du sujet (en caractère) et moment de la journée où il est envoyé
 Deuxième anova à un facteur: “subject_size” et “moment”

```
summary(aov(mails_anova1_2$subject_size~mails_anova1_2$Moment))
```

```
##                Df    Sum Sq Mean Sq F value Pr(>F)
## mails_anova1_2$Moment      3    348873    116291    259.8 <2e-16 ***
## Residuals                99996    44763430         448
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

AFC pour voir le lien entre le moment de la journée et la taille (petit, moyen, grand, très grand) du sujet :

```
table(mails_afc1_2[,1], mails_afc1_2[,2]) -> afc_2
afc_moment_subject_size <- CA(afc_2)
```



```
summary(afc_moment_subject_size)
```

```
##
## Call:
## CA(X = afc_2)
##
## The chi square of independence between the two variables is equal to 849.9776 (p-value = 2.440513e-
##
## Eigenvalues
##               Dim.1   Dim.2
## Variance        0.008   0.000
## % of var.       97.846   2.154
## Cumulative % of var. 97.846 100.000
##
## Rows
##           Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr   cos2
## grand      |    3.485 |  0.082 41.751 0.996 |  0.005  7.080 0.004 |
## moyen      |    1.041 | -0.053 11.307 0.903 | -0.017 55.077 0.097 |
## petit      |    3.973 | -0.160 46.941 0.983 |  0.021 37.844 0.017 |
##
## Columns
##           Iner*1000   Dim.1   ctr   cos2   Dim.2   ctr   cos2
## afternoon |    0.626 | -0.037  7.400 0.983 |  0.005  5.846 0.017 |
## evening   |    2.221 |  0.093 25.615 0.959 | -0.019 49.655 0.041 |
## morning   |    1.744 | -0.083 20.963 1.000 |  0.002  0.379 0.000 |
## night     |    3.908 |  0.261 46.022 0.979 |  0.038 44.120 0.021 |
```