

# ACP\_Caractéristiques\_Entreprises\_Non\_Agricoles

Michel TEVOEDJRE

2025-05-15



```
options(repos = c(CRAN = "https://cloud.r-project.org"))
```

```
packages <- c("haven", "dplyr", "FactoMineR", "factoextra", "s  
new.packages <- packages[!(packages %in% installed.packages())]  
if(length(new.packages)) install.packages(new.packages)  
lapply(packages, library, character.only = TRUE)
```

```
## Warning: le package 'haven' a été compilé avec la version R
```

```
## Warning: le package 'dplyr' a été compilé avec la version R
```

```
##
```

```
## Attachement du package : 'dplyr'
```

```
## Les objets suivants sont masqués depuis 'package:stats':
```

```
##
```

```
##      filter, lag
```

```
## Les objets suivants sont masqués depuis 'package:base':
```

```
##
```

# Introduction

Ce document détaille les étapes de préparation des données et la réalisation d'une Analyse en Composantes Principales (ACP) sur les données SEN2018.

# Chargement des bibliothèques et des données

```
library(haven)
library(dplyr)
library(FactoMineR)
library(factoextra)
library(stringr)

install.packages("rstudioapi")
senegaldata <- dirname(rstudioapi::getSourceEditorContext())$path
ma_base <- read_dta(paste0(senegaldata, "/s10_2_me_SEN2018.dta"))
```

# Inspection des variables

```
var_types_labels <- data.frame(  
  Variable = names(ma_base),  
  Type = sapply(ma_base, function(x) class(x)[1]),  
  Label = sapply(ma_base, function(x) attr(x, "label"))  
)  
kable(head(var_types_labels), caption = "Aperçu des variables",  
  kable_styling(bootstrap_options = c("striped", "hover", "condensed"))
```

Table 1: Aperçu des variables

	Variable	Type	Label
vague	vague	numeric	Vague
grappe	grappe	numeric	grappe
menage	menage	numeric	Identifiant du ménage
s10q12a_1	s10q12a_1	numeric	10.12a.Lister entreprises possédées par
s10q12a_2	s10q12a_2	character	10.12a.Lister entreprises possédées par

# Sélection et nettoyage des variables

```
# Suppression des lignes où s10q17b est 1 à 5
ma_base <- ma_base %>% filter(!(s10q17b %in% 1:5))
vars_quant <- ma_base %>%
  select(where(is.numeric)) %>%
  select(where(~ mean(is.na(.)) < 0.3)) %>%
  select(where(~ sd(., na.rm = TRUE) > 0)) %>%
  select(-any_of(c("vague", "grappe", "menage")))
```

# Imputation des valeurs manquantes

```
vars_imputed <- vars_quant %>%  
  mutate(across(everything(), ~ ifelse(is.na(.), mean(., na.rm = TRUE), .)))
```



# Vérification des données

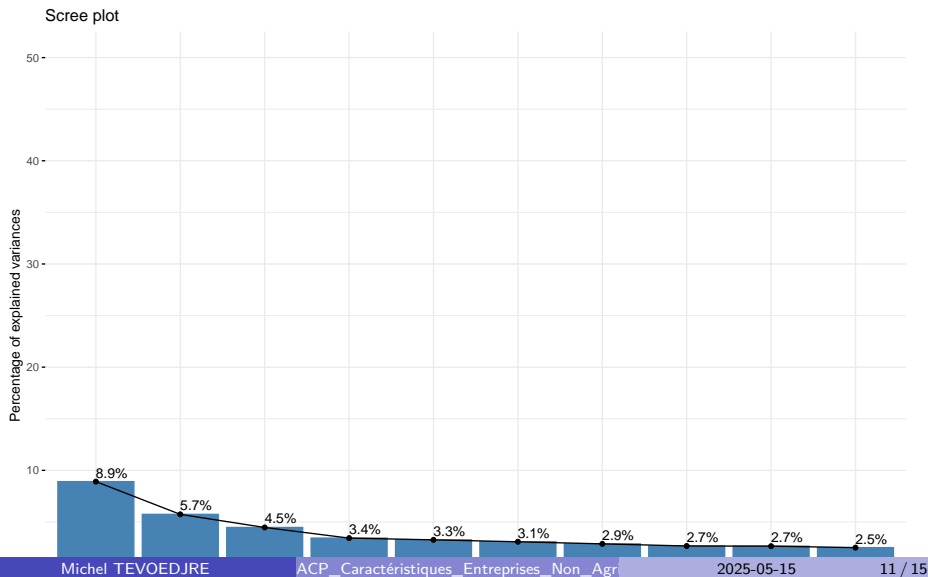
```
stopifnot(!any(is.na(vars_imputed)))  
stopifnot(all(sapply(vars_imputed, sd) > 0))
```

# Standardisation et ACP

```
vars_scaled <- scale(vars_imputed)
res_acp <- PCA(vars_scaled, graph = FALSE)
```

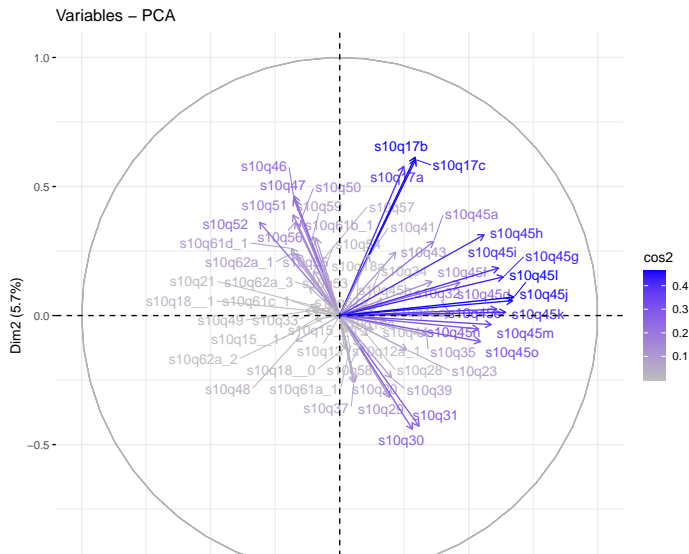
# Visualisation des résultats : Valeurs propres

```
fviz_eig(res_acp, addlabels = TRUE, ylim = c(0, 50))
```



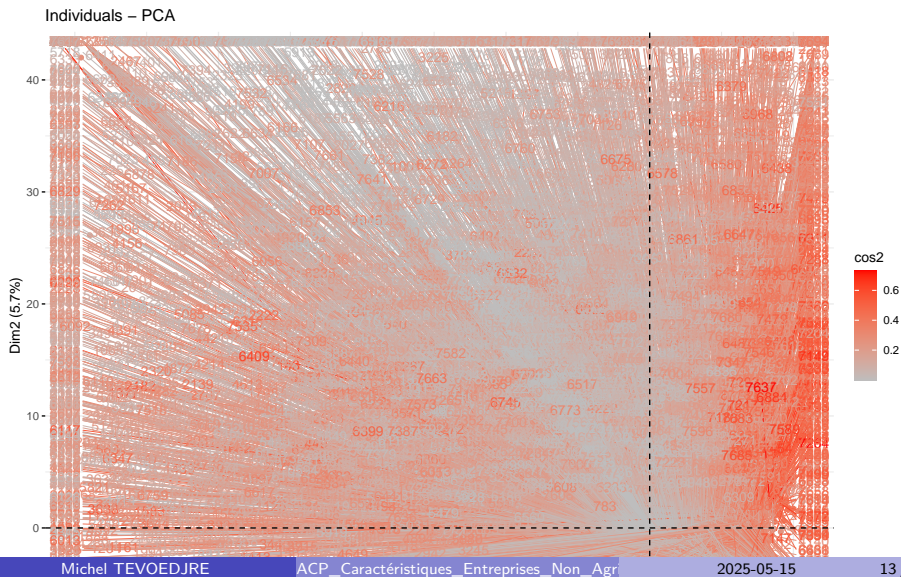
# Visualisation : Variables

```
fviz_pca_var(res_acp, col.var = "cos2", gradient.cols = c("gre
```



# Visualisation : Individus

```
fviz_pca_ind(res_acp, col.ind = "cos2", gradient.cols = c("gre
```



## Résultats : Coordonnées des individus

```
coord_individus <- as.data.frame(res_acp$ind$coord)
coord_individus <- cbind(id = rownames(coord_individus), coord_individus)
kable(head(coord_individus), caption = "Coordonnées des individus",
      kable_styling(bootstrap_options = c("striped", "hover", "condensed")))
```

Table 2: Coordonnées des individus (ACP)

id	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1	-0.4639700	-0.5195670	0.5101594	-0.7105956	-1.0087238
2	-0.4657207	-0.4884308	0.9440010	0.6633286	-1.4757352
3	0.6198622	-0.8730473	1.0302181	0.6610029	-1.3749322
4	1.0130290	-0.6209831	1.3811877	-0.1206627	-0.4089295
5	0.3081650	0.1360724	1.1604901	-0.0358335	-1.3660937
6	-2.6941145	2.6273728	1.8128393	2.5736232	-1.2238856

## Résultats : Contribution des variables

```
contrib_variables <- as.data.frame(res_acp$var$contrib)
contrib_variables <- cbind(variable = rownames(contrib_variables),
kable(head(contrib_variables), caption = "Contribution des variables",
kable_styling(bootstrap_options = c("striped", "hover", "condensed"))
```

Table 3: Contribution des variables aux axes principaux

	variable	Dim.1	Dim.2	Dim.3	Dim.4
s10q12a_1	s10q12a_1	0.0132642	0.1170097	0.0867358	3.86144
s10q13	s10q13	0.0000001	0.0000002	0.0000015	0.00031
s10q15__0	s10q15__0	0.0000002	0.0000012	0.0000041	0.00060
s10q15__1	s10q15__1	0.1602752	0.0057632	0.0288940	0.06562
s10q17a	s10q17a	1.0807747	9.1149327	14.7957609	1.26859
s10q17b	s10q17b	1.4976836	10.2552546	13.5722228	1.26351