

PROJET - MASTER 1 (MQME)

OBJECTIF PRINCIPAL : Analyse de données Titanic

ANNEE : 2025-2026

AUTEUR: Perrin Kibiti

1. Introduction

Le Titanic coule lors de son premier voyage entre l'Angleterre et les États-Unis dans la nuit du 14 au 15 avril 1912. Une grande partie des personnes à bord perdent la vie, tandis qu'une minorité seulement parvient à survivre. Dans ce travail, nous disposons d'informations sur les passagers à savoir leur classe de voyage (équipage, première, deuxième, troisième), leur sexe (homme ou femme), leur âge (enfant ou adulte) et le fait qu'ils soient survivants ou décédés. L'idée est de regarder comment ces caractéristiques sont liées à la survie, et de voir quels types de passagers ont le plus (ou le moins) de chances d'avoir survécu au naufrage. Pour se faire, nous allons procéder au traitement des données, puis à une analyse adaptée en tenant compte des données auxquelles nous avons affaire.

2. Pré-traitement des données

Objectif : comprendre la structure du fichier puis recoder les modalités des variables.

```
# Chargement des packages
```

```
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  
  
library(FactoMineR)  
library(factoextra)  
  
## Loading required package: ggplot2  
  
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

Importation des données

```
# Import des données
```

```
titanic <- read.csv("C:/Users/gpkib/Box/Gael Box/Université de Lille/Analyse de données I (6ECTS)/Proje
```

```
# Aperçu
```

```
str(titanic)
```

```
## 'data.frame':    2201 obs. of  4 variables:  
##   $ Classe: int  1 1 1 1 1 1 1 1 1 ...  
##   $ Age   : int  1 1 1 1 1 1 1 1 1 ...  
##   $ Sexe  : int  1 1 1 1 1 1 1 1 1 ...  
##   $ Survie: int  1 1 1 1 1 1 1 1 1 ...
```

```
head(titanic)
```

```

##   Classe Age Sexe Survie
## 1     1   1     1     1
## 2     1   1     1     1
## 3     1   1     1     1
## 4     1   1     1     1
## 5     1   1     1     1
## 6     1   1     1     1

# Vérification des valeurs manquantes
sum(is.na(titanic))

```

```
## [1] 0
```

Aucune valeur manquante n'est identifiée dans cette base de données.

Recodage des variables

```

#Classe
titanic$Classe <- factor(titanic$Classe)

levels(titanic$Classe)[levels(titanic$Classe) == "0"] <- "équipage"
levels(titanic$Classe)[levels(titanic$Classe) == "1"] <- "première classe"
levels(titanic$Classe)[levels(titanic$Classe) == "2"] <- "seconde classe"
levels(titanic$Classe)[levels(titanic$Classe) == "3"] <- "troisième classe"

# Age
titanic$Age <- factor(titanic$Age)

levels(titanic$Age)[levels(titanic$Age) == "0"] <- "enfant"
levels(titanic$Age)[levels(titanic$Age) == "1"] <- "adulte"

# Sexe
titanic$Sexe <- factor(titanic$Sexe)

levels(titanic$Sexe)[levels(titanic$Sexe) == "0"] <- "femme"
levels(titanic$Sexe)[levels(titanic$Sexe) == "1"] <- "homme"

# Survie
titanic$Survie <- factor(titanic$Survie)

levels(titanic$Survie)[levels(titanic$Survie) == "0"] <- "décédé"
levels(titanic$Survie)[levels(titanic$Survie) == "1"] <- "survivant"

str(titanic)

## 'data.frame': 2201 obs. of 4 variables:
## $ Classe: Factor w/ 4 levels "équipage","première classe",...: 2 2 2 2 2 2 2 2 2 ...
## $ Age   : Factor w/ 2 levels "enfant","adulte": 2 2 2 2 2 2 2 2 2 ...
## $ Sexe  : Factor w/ 2 levels "femme","homme": 2 2 2 2 2 2 2 2 2 ...
## $ Survie: Factor w/ 2 levels "décédé","survivant": 2 2 2 2 2 2 2 2 2 ...

```

Analyse descriptive

```
summary(titanic)
```

```
##          Classe      Age       Sexe     Survie
## équipage :885   enfant: 109   femme: 470   décédé :1490
## première classe :325   adulte:2092   homme:1731   survivant: 711
## seconde classe :285
## troisième classe:706
```

Sur les 2201 personnes recensées, la majorité sont des adultes (2092), des hommes (1731), et la plupart n'ont pas survécues (1490) pas au naufrage de titanic.

3. Analyse factorielle

Choix de la méthode factorielle

Les quatre variables de notre base de données sont qualitatives (Classe, Age, Sexe, Survie). Nous avons donc plus de deux variables qualitatives, alors la méthode factorielle adaptée est une **analyse des correspondances multiples (ACM)**.

Application de l'ACM

```
# Analyse des correspondances multiples
res.mca <- MCA(titanic, graph = FALSE)

# Résultats
res.mca

## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 2201 individuals, described by 4 variables
## *The results are available in the following objects:
##
##      name           description
## 1  "$eig"         "eigenvalues"
## 2  "$var"         "results for the variables"
## 3  "$var$coord"   "coord. of the categories"
## 4  "$var$cos2"    "cos2 for the categories"
## 5  "$var$contrib" "contributions of the categories"
## 6  "$var$v.test"  "v-test for the categories"
## 7  "$var$eta2"    "coord. of variables"
## 8  "$ind"         "results for the individuals"
## 9  "$ind$coord"   "coord. for the individuals"
## 10 "$ind$cos2"    "cos2 for the individuals"
## 11 "$ind$contrib" "contributions of the individuals"
## 12 "$call"        "intermediate results"
## 13 "$call$marge.col" "weights of columns"
## 14 "$call$marge.li" "weights of rows"

eig <- res.mca$eig
eig

##      eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.4450795          29.671965          29.67196
## dim 2  0.3050437          20.336249          50.00821
## dim 3  0.2500060          16.667067          66.67528
```

```

## dim 4  0.2050373          13.669154          80.34443
## dim 5  0.1785152          11.901011          92.24544
## dim 6  0.1163183          7.754555          100.00000

```

Choix du nombre de composantes

Règle du coude

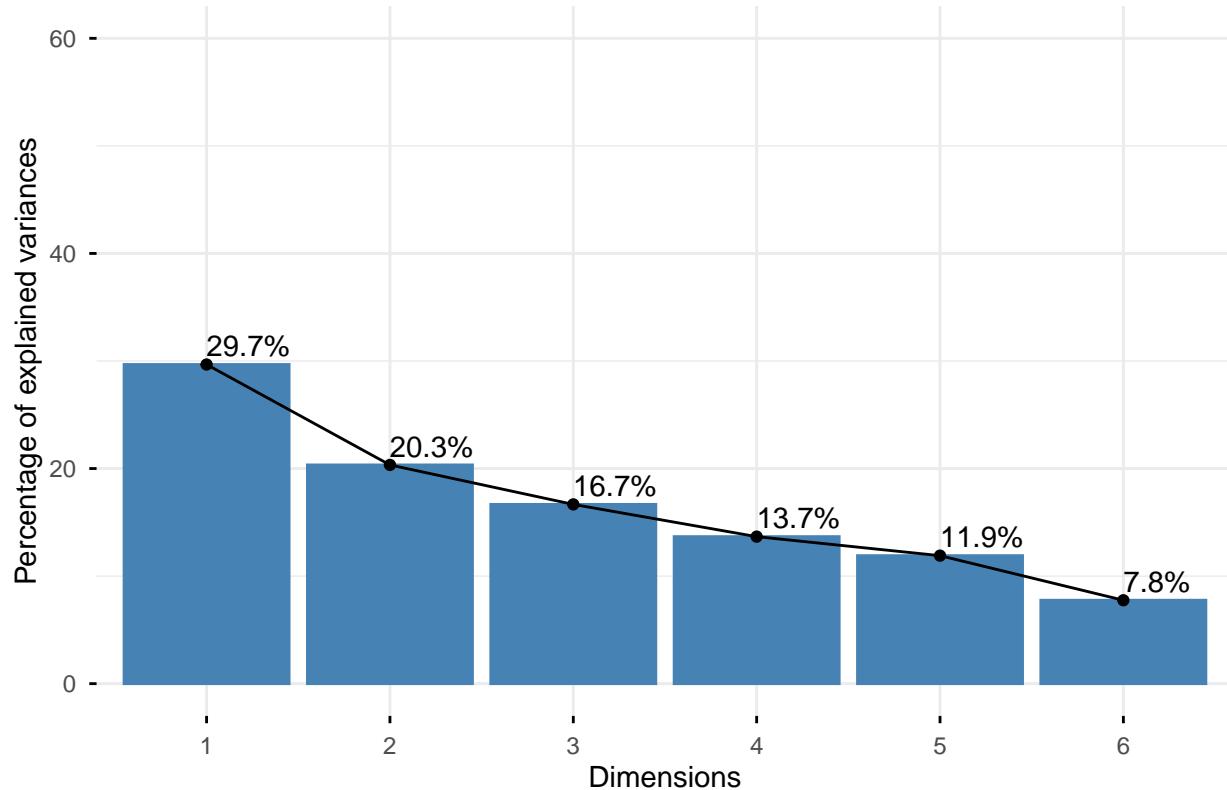
```
fviz_eig(res.mca, addlabels = TRUE, ylim = c(0, 60))
```

```

## Warning in geom_bar(stat = "identity", fill = barfill, color = barcolor, :
## Ignoring empty aesthetic: `width`.

```

Scree plot



Nous remarquons la formation d'un coude au niveau de la cinquième composante (11,9%), ce qui nous conduit au choix des 4 premiers axes pour la suite notre analyse d'après la règle de Coude.

Règle de Kaiser : Choix des composantes dont les valeurs propres sont supérieures à la moyenne

Cette moyenne est calculée de la manière suivante : $1/p$, avec $p = 4$.

```
# Moyenne des valeurs propres
```

```
moyenne_vp <- 1/4
```

```
moyenne_vp
```

```
## [1] 0.25
```

```
eig
```

```
##      eigenvalue percentage of variance cumulative percentage of variance
```

```

## dim 1 0.4450795          29.671965          29.67196
## dim 2 0.3050437          20.336249          50.00821
## dim 3 0.2500060          16.667067          66.67528
## dim 4 0.2050373          13.669154          80.34443
## dim 5 0.1785152          11.901011          92.24544
## dim 6 0.1163183          7.754555          100.00000

```

Nous constatons que la moyenne des valeurs propres est égale à 0.25 et seulement les 3 premiers axes ont des valeurs propres supérieures à 0.25. Selon la loi de Kaiser, nous n'allons considérer que les 3 axes.

Seuil de variance expliquée

D'après le résultat, nous retenons 4 composantes comme suit :

Composante 1 : 29,7% de variance de la variabilité du nuage des individus est expliquée.

Composante 2 : 20,3% de variance de la variabilité du nuage des individus est expliquée.

Composante 3 : 16,7% de variance de la variabilité du nuage des individus est expliquée.

Composante 4 : 11,9% de variance de la variabilité du nuage des individus est expliquée.

Ainsi 80% (soit $29,7\% + 20,3\% + 16,7\% + 11,9\%$) de la variance cumulée est expliquée par les quatre premières composantes lesquelles par conséquent expliquent quasiment la majorité de l'information engendrée par la base de données. Par conséquent d'après les 2 règles (Coude et le seuil de la variance expliquée), nous gardons 4 composantes pour la suite de notre analyse.

4. Interprétation des axes factoriels

4.1. Contributions des modalités aux axes

```

var <- res.mca$var
print(var$contrib, 2)

##
## Dim 1   Dim 2   Dim 3   Dim 4   Dim 5
## équipage 12.27 7.7e+00 2.7e-01 20.354 10.76
## première classe 11.01 1.8e+01 1.2e+01 0.044 44.14
## seconde classe 3.08 6.8e-01 7.4e+01 4.032 3.74
## troisième classe 0.31 3.0e+01 1.4e+01 13.198 4.26
## enfant 4.71 3.5e+01 2.8e-03 44.744 8.90
## adulte 0.25 1.8e+00 1.5e-04 2.331 0.46
## femme 29.75 1.4e-03 1.8e-03 2.987 7.54
## homme 8.08 3.8e-04 5.0e-04 0.811 2.05
## décédé 9.87 2.0e+00 9.5e-04 3.715 5.87
## survivant 20.68 4.2e+00 2.0e-03 7.784 12.29

print(var$coord, 2)

##
## Dim 1   Dim 2   Dim 3   Dim 4   Dim 5
## équipage -0.737 -0.4827 0.0823 0.644 -0.437
## première classe 1.152 -1.2314 -0.8900 -0.050 1.461
## seconde classe 0.651 0.2525 2.3921 -0.505 0.454
## troisième classe 0.131 1.0701 -0.6591 -0.581 -0.308
## enfant 1.302 2.9426 0.0238 2.722 1.133
## adulte -0.068 -0.1533 -0.0012 -0.142 -0.059
## femme 1.575 0.0089 -0.0093 -0.339 -0.502
## homme -0.428 -0.0024 0.0025 0.092 0.136
## décédé -0.509 0.1902 -0.0038 -0.212 0.249

```

```

## survivant      1.068 -0.3987  0.0079  0.445 -0.521
print(var$cos2, 2)

##           Dim 1   Dim 2   Dim 3   Dim 4 Dim 5
## équipage    0.3652 1.6e-01 4.6e-03 0.27920 0.128
## première classe 0.2299 2.6e-01 1.4e-01 0.00043 0.370
## seconde classe 0.0631 9.5e-03 8.5e-01 0.03799 0.031
## troisième classe 0.0081 5.4e-01 2.1e-01 0.15935 0.045
## enfant       0.0883 4.5e-01 3.0e-05 0.38608 0.067
## adulte        0.0883 4.5e-01 3.0e-05 0.38608 0.067
## femme          0.6734 2.2e-05 2.3e-05 0.03115 0.068
## homme          0.6734 2.2e-05 2.3e-05 0.03115 0.068
## décédé         0.5440 7.6e-02 2.9e-05 0.09431 0.130
## survivant      0.5440 7.6e-02 2.9e-05 0.09431 0.130

```

Dim 1 : Les modalités femme (29,75%) et survivant (20,68%) contribuent le plus à la création du premier axe tandis que première classe et équipage contribuent moyennement.

Dim 2 : Les modalités première classe (18,35%) contribue le plus à la création du deuxième axe.

Dim 3 : La modalité seconde classe (7,40%) contribue le plus à la création du troisième axe.

Dim 4 : Les modalités enfant (44,74%) et équipage sont celles qui contribuent le plus à la création du quatrième axe.

4.2 Qualité de représentation des modalités

Nous constatons que femme et homme ont une qualité meilleure de représentation sur l'axe 1, suivis de décédé et survivant. Cependant, les qualités de représentation ne sont pas bonnes sur les axes 2 et 3. Par contre, enfant, adulte et équipage sont moyennement représentées sur l'axe 4, comparativement aux autres.

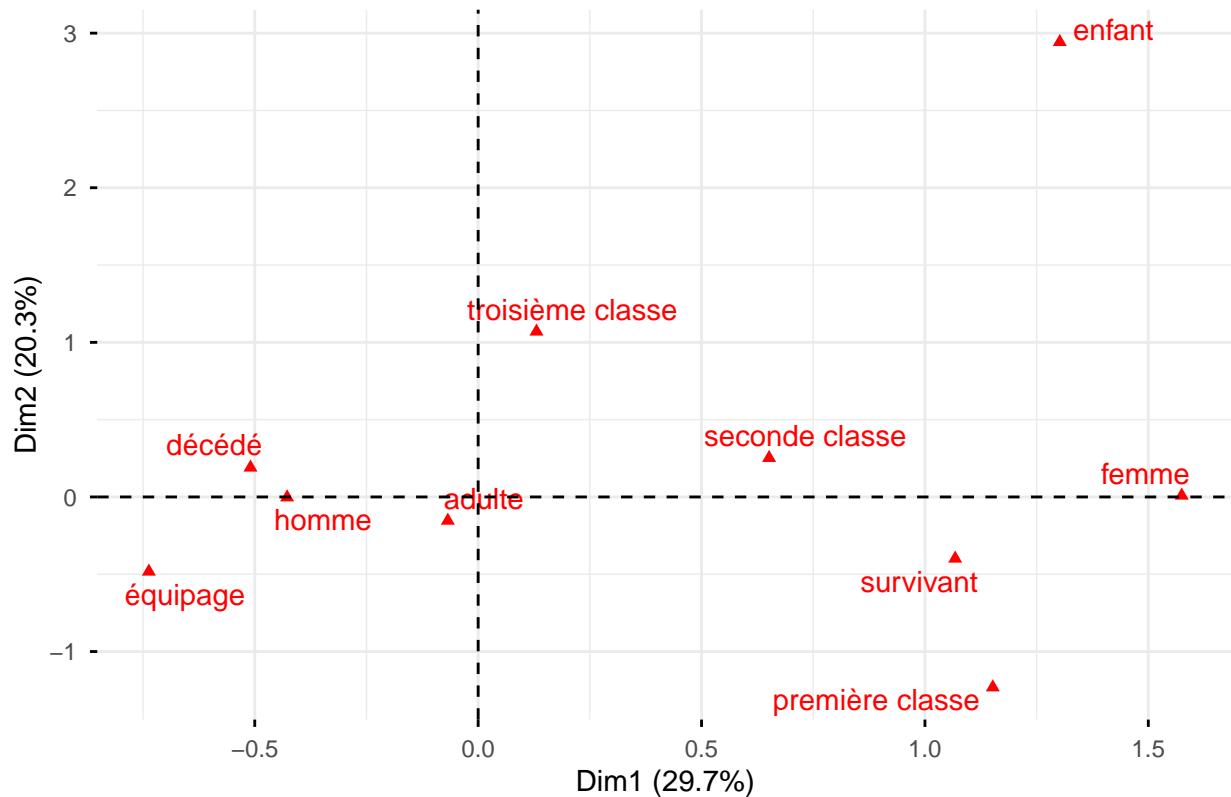
Visualisation sur le premier plan (1,2)

```

fviz_mca_var(
  res.mca,
  repel = TRUE
)

```

Variable categories – MCA



On constate ce qui suit :

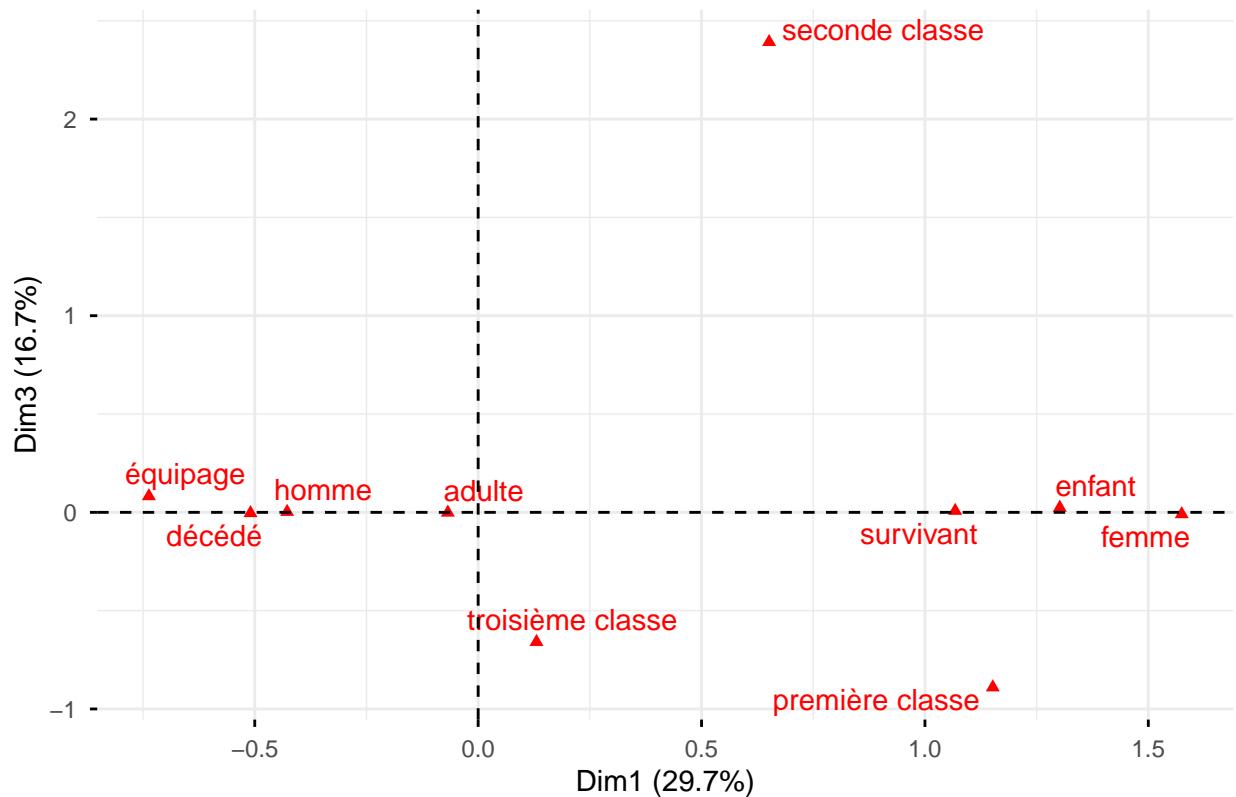
À gauche de l'axe 1, se trouvent les modalités équipage, homme et décédé, ce qui signifie que l'on observe davantage d'hommes de l'équipage décédés. Tandis qu'à droite, nous avons les modalités femme, survivant, première classe et seconde classe se situant dans le même secteur du plan, ce qui indique une plus forte présence de survivants issus de la première et de la seconde classe. Par conséquent, l'axe 1 oppose un profil **“femmes, passagers de première et deuxième classe, plutôt survivants”** à un profil **“hommes d’équipage, davantage décédés”**.

En haut à droite (côté positif), l'axe 2 est structuré par les modalités enfant et troisième classe, ce qui signifie que les enfants et les passagers de troisième classe sont fortement associés à cet axe. En bas (côté négatif), on trouve des modalités comme survivant et adulte, proches de l'origine, ce qui indique que les adultes ne se distinguent pas vraiment sur ce second axe. Par conséquent, l'axe 2 met surtout en avant l'opposition **“enfants souvent en troisième classe”** avec le profil **“première classe”**, enfant étant très éloigné.

Visualisation sur le plan (1,3)

```
fviz_mca_var(
  res.mca,
  repel = TRUE, axes=c(1,3)
)
```

Variable categories – MCA

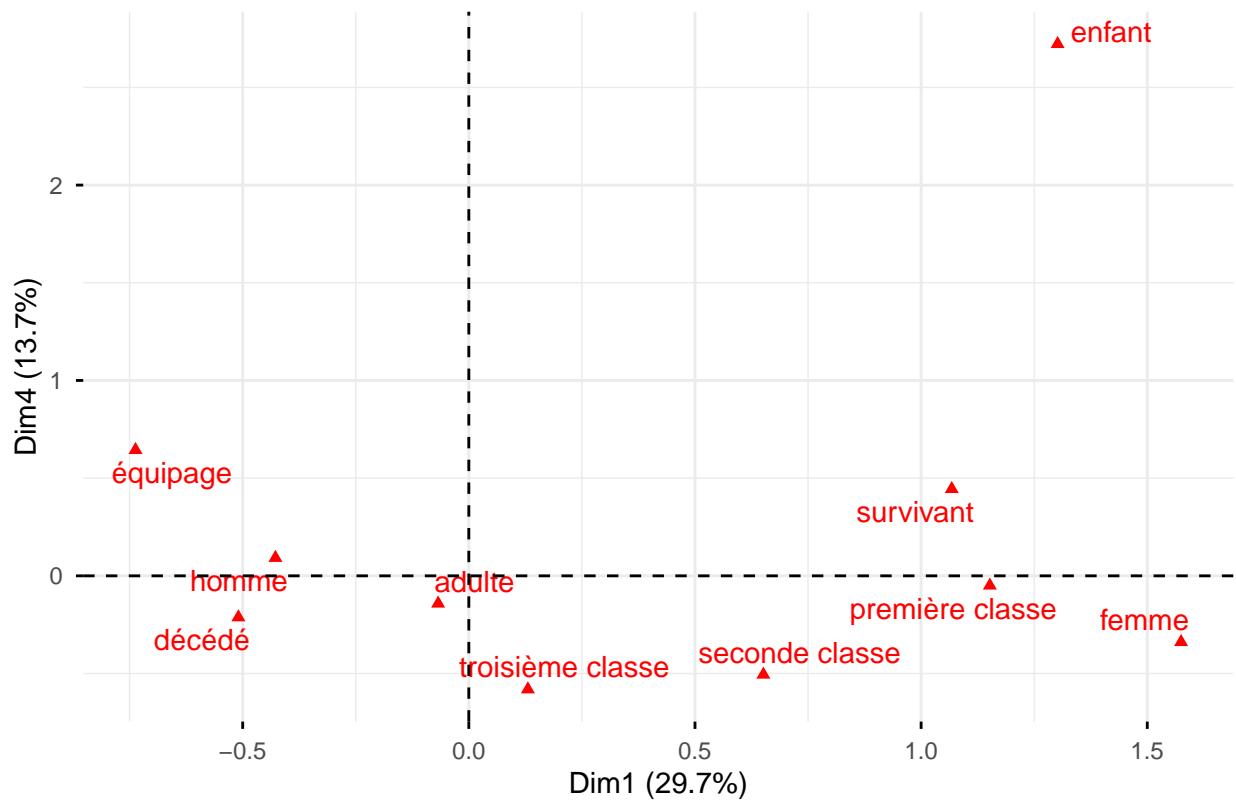


Il est clair de constater que l'axe 3 oppose seconde classe à première et troisième classe car les autres modalités ne contribuent pas vraiment à la création de cet axe. Les hommes de l'équipage sont décédé tandis que les femmes et enfants de la première et seconde classe ont survécu.

Visualisation sur le plan (1,4)

```
fviz_mca_var(
  res.mca,
  repel = TRUE, axes=c(1,4)
)
```

Variable categories – MCA

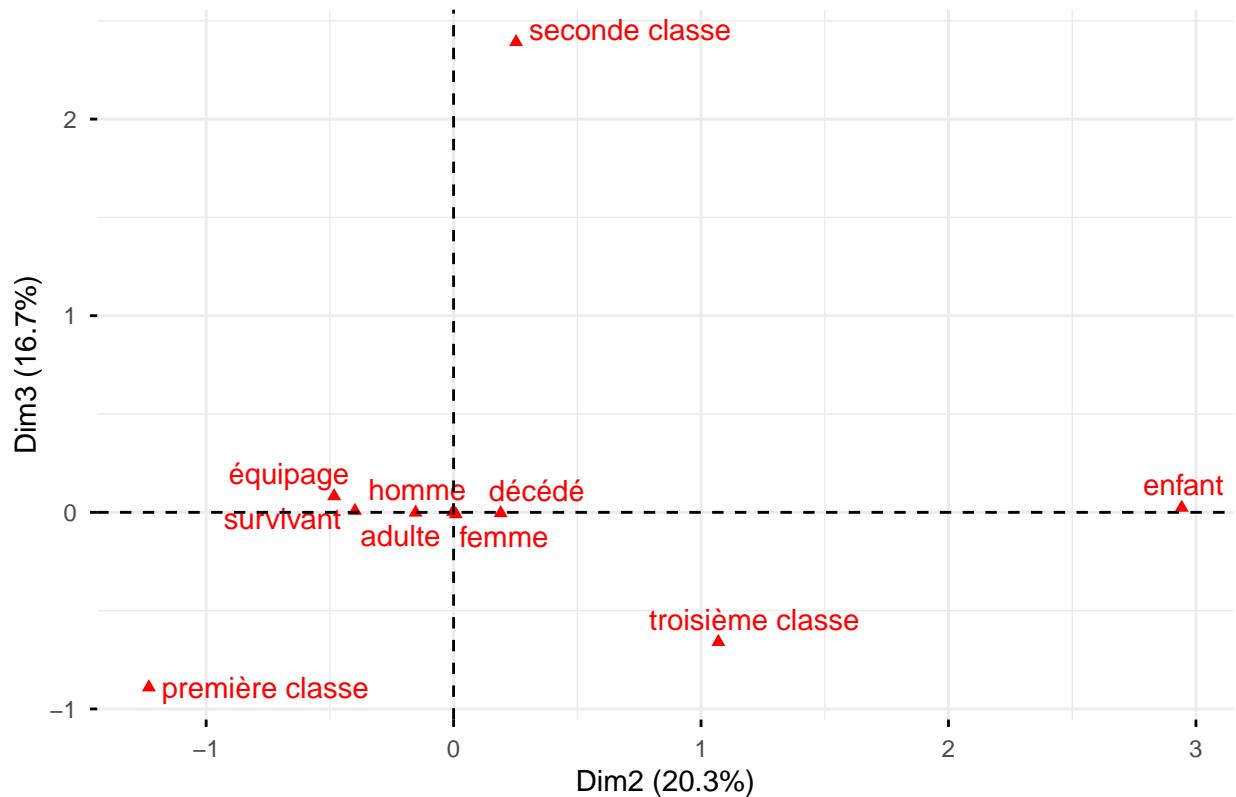


L'axe 4 oppose un groupe très spécifique constitué des enfants et de l'équipage à la masse des passagers standards de première et troisième classe. Nous constatons que les hommes de l'équipage sont décédés tandis que les femmes et les enfants de la première et seconde classe ont survécu.

Visualisation sur le plan (2,3)

```
fviz_mca_var(
  res.mca,
  repel = TRUE, axes=c(2,3)
)
```

Variable categories – MCA

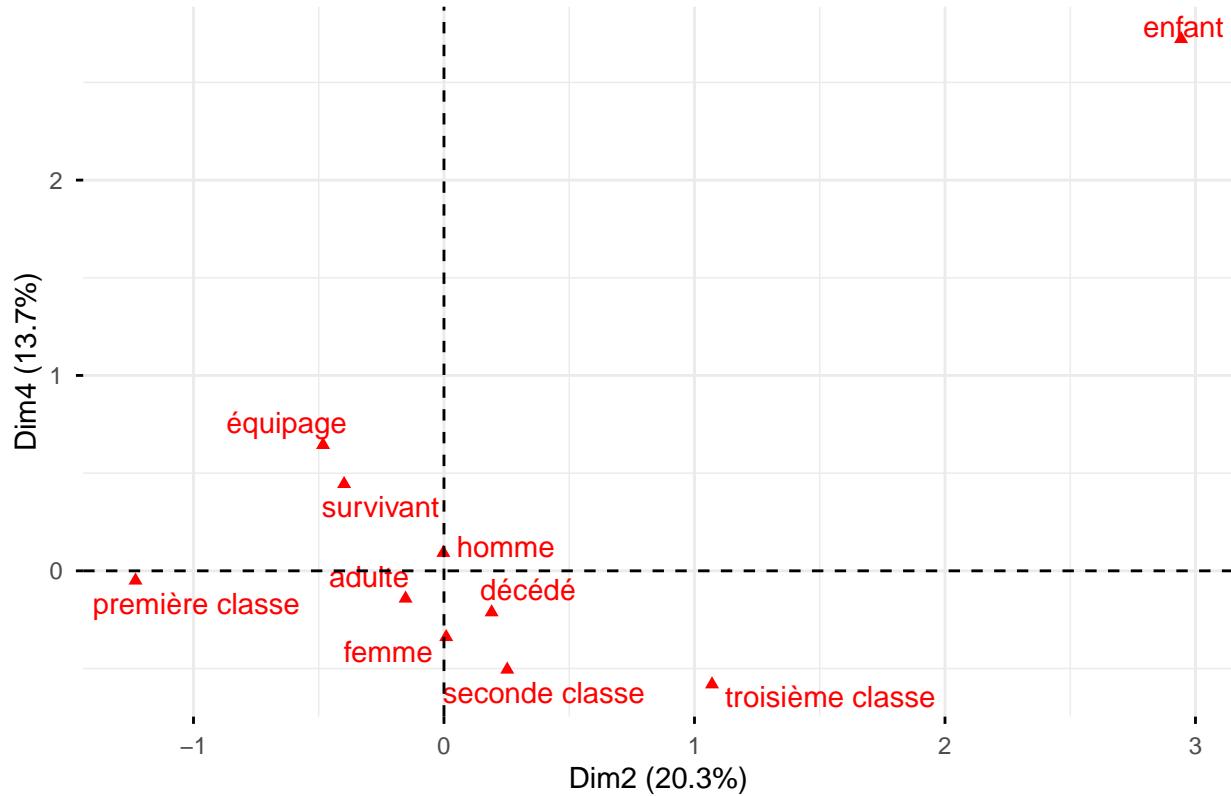


Ce graphique confirme les interprétations des axes 2 et 3. D'un côté l'axe 2 sépare surtout les enfants des adultes, de l'autre l'axe 3 oppose seconde classe à première et troisième classe. Nous constatons que la première classe a survécu avec une partie de l'équipage. Par ailleurs, de nombreux enfants de la troisième classe sont décédés.

Visualisation sur le plan (2,4)

```
fviz_mca_var(
  res.mca,
  repel = TRUE, axes=c(2,4)
)
```

Variable categories – MCA



Nous constatons que une fois de plus que la modalité enfant est atypique.

4.3 Scores des individus

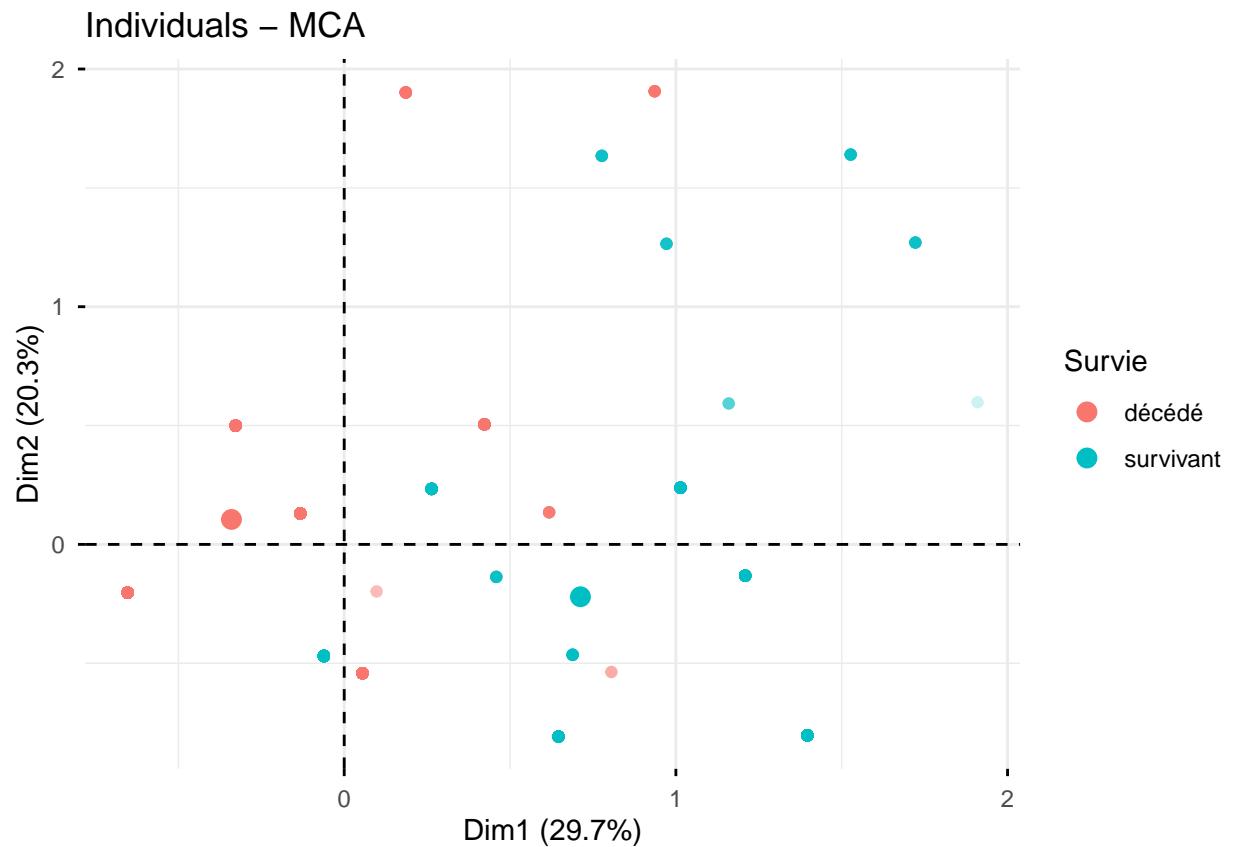
```
ind <- res.mca$ind
head(ind$coord,10)

##          Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## 1  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 2  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 3  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 4  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 5  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 6  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 7  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 8  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 9  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 10 0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
```

Au vu du grand nombre d'individus de notre base, il n'est pas intéressant de les associer aux variables (modalités). Cela ne nous servirait d'aucune utilité en termes de visualisations sur les plans factoriels et d'interprétations simultanées. Cependant, nous pourrons effectuer une séparation des groupes avec habillage.

Séparation des groupes avec habillage sur le Plan (1,2)

```
fviz_mca_ind(res.mca,
              axes = c(1,2),
              geom = "point",
              habillage = "Survie",
              alpha.ind = 0.2
            )
```

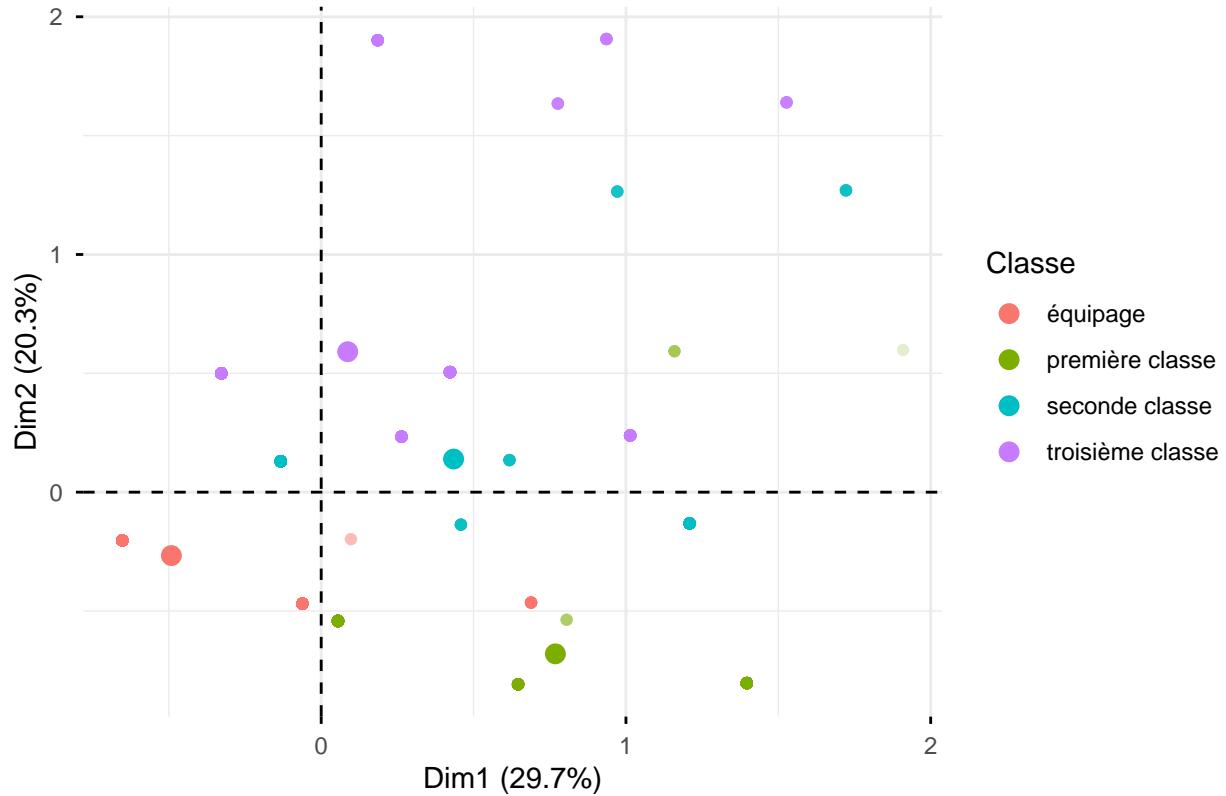


On observe des points rouges (décédés) et bleus (survivants) mélangés sur l'ensemble du graphique. On distingue tout de même une tendance avec davantage d'individus décédés à gauche et de survivants à droite, mais les deux premiers axes ne permettent pas de séparer nettement survivants et décédés au niveau individuel.

Séparation des groupes avec habillage sur le Plan (1,2)

```
fviz_mca_ind(res.mca,
              axes = c(1,2),
              geom = "point",
              habillage = "Classe",
              alpha.ind = 0.2
            )
```

Individuals – MCA



On observe des points de couleurs différentes correspondant aux quatre classes de passagers, répartis sur tout le plan factoriel. On remarque bien quelques zones où certaines classes sont un peu plus présentes (par exemple plusieurs points de troisième classe en haut du graphique), mais globalement les deux premiers axes ne séparent pas nettement les classes : les individus des différentes classes restent largement mêlés.

5. Synthèse des résultats

Au regard de cette étude, nous réalisons que parmi les passagers de titanic, l'analyse fait apparaître une forte opposition entre, d'un côté, les femmes des classes première et seconde, qui ont davantage survécu, et de l'autre les hommes de l'équipage, beaucoup plus souvent décédés. Les enfants constituent un groupe cas atypique car ils sont fréquemment associés à la troisième classe et se distinguent nettement des adultes. La majorité des enfants de la troisième classe n'ont pas survécu.

Vérification de nos interprétations

```
with(titanic, sum(Age == "enfant" & Classe == "troisième classe" & Survie == "décédé"))

## [1] 52

52 enfants de la troisième classe avaient trouvé la mort lors du naufrage de titanic.

with(titanic, sum(Age == "enfant" & Classe == "seconde classe" & Survie == "décédé"))

## [1] 0

with(titanic, sum(Age == "enfant" & Classe == "première classe" & Survie == "décédé"))

## [1] 0
```

Aucun enfant des classes première et seconde n'avait trouvé la mort lors du naufrage de titanic.

6. L'ACM avec Age en quali.sup

Dans un deuxième temps, nous souhaiterions effectuer l'ACM en mettant la variable Age en supérieur afin d'évaluer l'impact de son retrait sur notre analyse. Notre objectif serait de s'assurer que tous les passagers avaient les mêmes chances d'être secourus sans tenir compte de l'âge.

Application de l'ACM

```
# Analyse des correspondances multiples
res.mca2 <- MCA(titanic, graph = FALSE, quali.sup = 2)

# Résultats
res.mca2

## **Results of the Multiple Correspondence Analysis (MCA)**
## The analysis was performed on 2201 individuals, described by 4 variables
## *The results are available in the following objects:
##
##      name              description
## 1  "$eig"            "eigenvalues"
## 2  "$var"             "results for the variables"
## 3  "$var$coord"       "coord. of the categories"
## 4  "$var$cos2"        "cos2 for the categories"
## 5  "$var$contrib"     "contributions of the categories"
## 6  "$var$v.test"      "v-test for the categories"
## 7  "$var$eta2"         "coord. of variables"
## 8  "$ind"              "results for the individuals"
## 9  "$ind$coord"        "coord. for the individuals"
## 10 "$ind$cos2"         "cos2 for the individuals"
## 11 "$ind$contrib"      "contributions of the individuals"
## 12 "$quali.sup"        "results for the supplementary categorical variables"
## 13 "$quali.sup$coord"  "coord. for the supplementary categories"
## 14 "$quali.sup$cos2"   "cos2 for the supplementary categories"
## 15 "$quali.sup$v.test" "v-test for the supplementary categories"
## 16 "$call"              "intermediate results"
## 17 "$call$marge.col"   "weights of columns"
## 18 "$call$marge.li"    "weights of rows"

eig2 <- res.mca2$eig
eig2

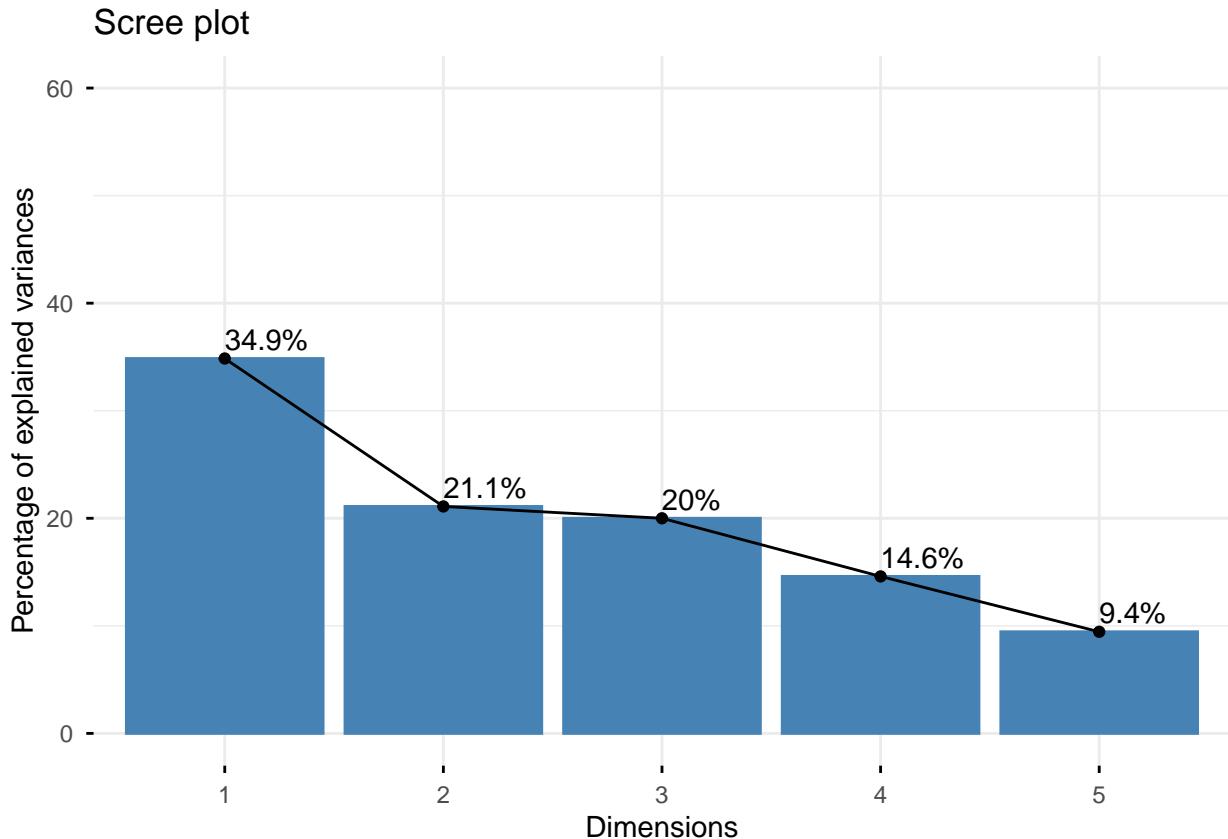
##      eigenvalue percentage of variance cumulative percentage of variance
## dim 1  0.5809291          34.855743           34.85574
## dim 2  0.3516697          21.100181           55.95592
## dim 3  0.3333333          20.000000           75.95592
## dim 4  0.2432665          14.595989          90.55191
## dim 5  0.1574681          9.448087          100.00000
```

Choix des composantes

Règle du coude

```
fviz_eig(res.mca2, addlabels = TRUE, ylim = c(0, 60))

## Warning in geom_bar(stat = "identity", fill = barfill, color = barcolor, :
## Ignoring empty aesthetic: `width`.
```



On remarque la formation d'un coude au niveau de la troisième composante (20%), ce qui nous conduit au choix des 3 premiers axes d'après la règle de Coude.

Règle de Kaiser : Choix des composantes dont les valeurs propres sont supérieures à la moyenne

Cette moyenne est calculée de la manière suivante : $1/p$, avec $p = 3$.

```
# Moyenne des valeurs propres
moyenne_vp <- 1/3
moyenne_vp
```

```
## [1] 0.3333333
```

```
print(eig2, 2)
```

```
##          eigenvalue percentage of variance cumulative percentage of variance
## dim 1      0.58                  34.9                      35
## dim 2      0.35                  21.1                      56
## dim 3      0.33                  20.0                      76
## dim 4      0.24                  14.6                     91
## dim 5      0.16                  9.4                     100
```

Nous constatons que la moyenne des valeurs propres est égale à 0.33 et seulement les 3 premiers axes ont des

valeurs propres supérieures cette valeur. Selon la loi de Kaiser, nous n'allons considérer que les 3 axes.

Seuil de variance expliquée

D'après le résultat, nous retenons 3 composantes :

Composante 1 : 34,9% de variance de la variabilité du nuage des individus est expliquée.

Composante 2 : 21,1% de variance de la variabilité du nuage des individus est expliquée.

Composante 3 : 20% de variance de la variabilité du nuage des individus est expliquée.

Ainsi 76% (soit $34,9\% + 21,1\% + 20\%$) de la variance cumulée est expliquée par les 3 premières composantes lesquelles par conséquent expliquent quasiment la majorité de l'information engendrée par la base de données. Par conséquent d'après les 3 règles (Coude, Kaiser et le seuil de la variance expliquée), nous gardons 3 composantes.

7. Interprétation des axes factoriels

7.1. Contributions des modalités aux axes

```
var2 <- res.mca2$var
print(var2$contrib,2)

## Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## équipage 10.9974 21.37 1.8e-01 19.54 7.704
## première classe 14.0643 16.94 1.2e+01 41.92 0.098
## seconde classe 2.9282 2.19 7.5e+01 5.88 1.342
## troisième classe 0.0066 49.39 1.3e+01 0.97 4.663
## femme 30.7057 2.77 1.6e-26 3.38 41.786
## homme 8.3372 0.75 4.6e-27 0.92 11.346
## décédé 10.6475 2.13 1.9e-27 8.85 10.680
## survivant 22.3133 4.46 1.9e-26 18.55 22.381

print(var2$coord,2)

## Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## équipage -0.690 0.75 6.6e-02 -0.596 0.301
## première classe 1.288 1.10 -9.1e-01 1.439 -0.056
## seconde classe 0.628 -0.42 2.4e+00 0.575 -0.221
## troisième classe 0.019 -1.27 -6.3e-01 -0.148 -0.262
## femme 1.583 -0.37 -2.8e-14 -0.340 0.961
## homme -0.430 0.10 7.7e-15 0.092 -0.261
## décédé -0.524 -0.18 -5.3e-15 0.309 0.273
## survivant 1.097 0.38 2.5e-14 -0.647 -0.572

print(var2$cos2,2)

## Dim 1 Dim 2 Dim 3 Dim 4 Dim 5
## équipage 0.32055 0.377 2.9e-03 0.239 0.06087
## première classe 0.28757 0.210 1.4e-01 0.359 0.00054
## seconde classe 0.05862 0.027 8.6e-01 0.049 0.00728
## troisième classe 0.00017 0.767 1.9e-01 0.010 0.03243
## femme 0.68043 0.037 2.1e-28 0.031 0.25100
## homme 0.68043 0.037 2.2e-28 0.031 0.25100
## décédé 0.57444 0.069 5.9e-29 0.200 0.15618
## survivant 0.57444 0.069 2.9e-28 0.200 0.15618
```

Dim 1 : Les modalités femme (30,70%) et survivant (22,31%) contribuent le plus à la création du premier axe.

Dim 2 : Les modalités troisième classe (49,39%) et équipage (21,37%) contribuent le plus à la création du deuxième axe, suivie de première classe (16,94%) qui contribue moyennement.

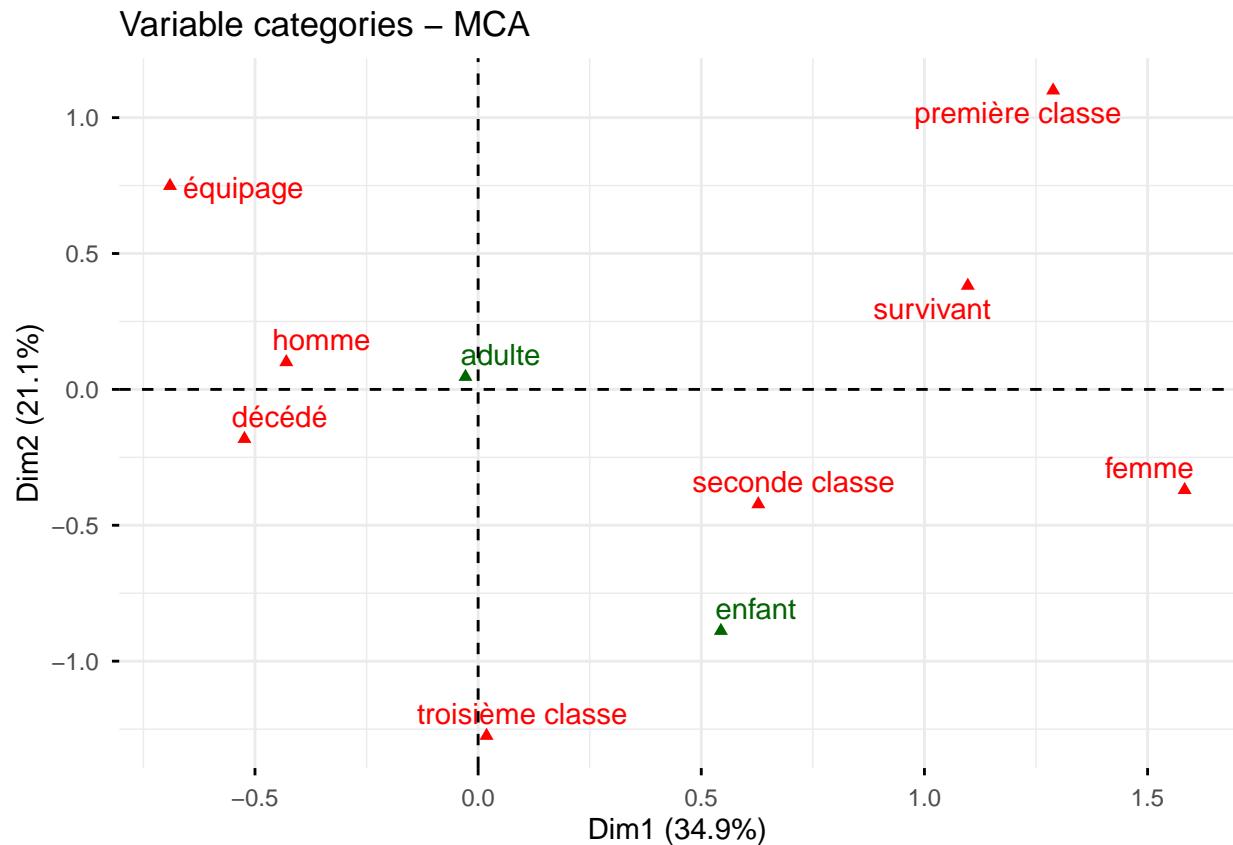
Dim 3 : La modalité première classe (75%) contribue le plus à la création du troisième axe.

Qualité de représentation

Nous constatons que femme et homme sont d'une meilleure qualité de représentation par rapport aux autres sur l'axe 1. Cependant, troisième classe qui a également une meilleure qualité de représentation sur l'axe 2, suivie de la modalité équipage qui est moyennement représentée. Par contre, on constate qu'aucune modalité n'est bien représentée sur l'axe 3.

Visualisation sur le premier plan (1,2)

```
fviz_mca_var(  
  res.mca2,  
  repel = TRUE  
)
```

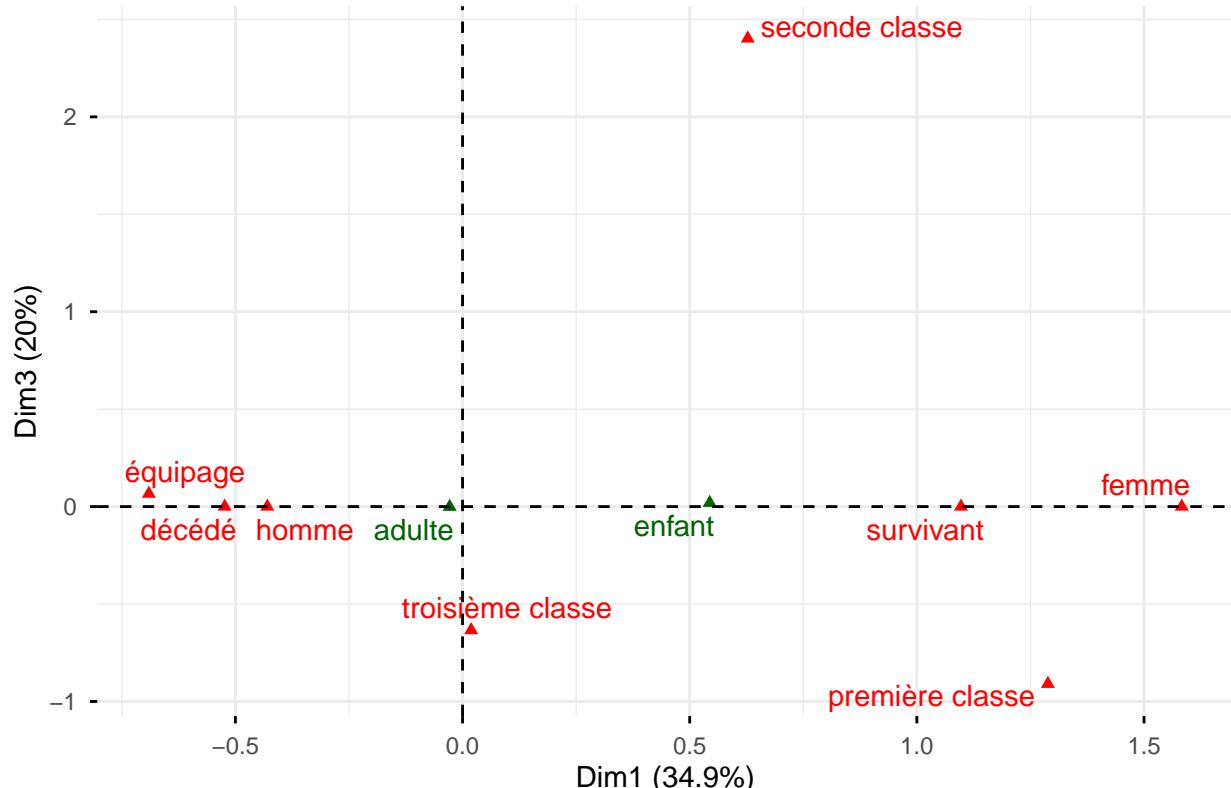


On constate que les modalités équipage, homme et décédé sont étroitement associées entre elles, tandis que survivant, femme, première classe et seconde classe forment un groupe opposé. Cela signifie qu'un grand nombre d'hommes de l'équipage sont décédés, alors que de nombreuses femmes des première et seconde classes ont survécu. L'axe 1 oppose donc clairement les personnes décédées de l'équipage aux survivants issus des première et seconde classes. Tandis que l'axe 2 oppose première classe et équipage à seconde classe et troisième classe.

Visualisation sur le plan (1,3)

```
fviz_mca_var(  
  res.mca2,  
  repel = TRUE, axes=c(1,3)  
)
```

Variable categories – MCA

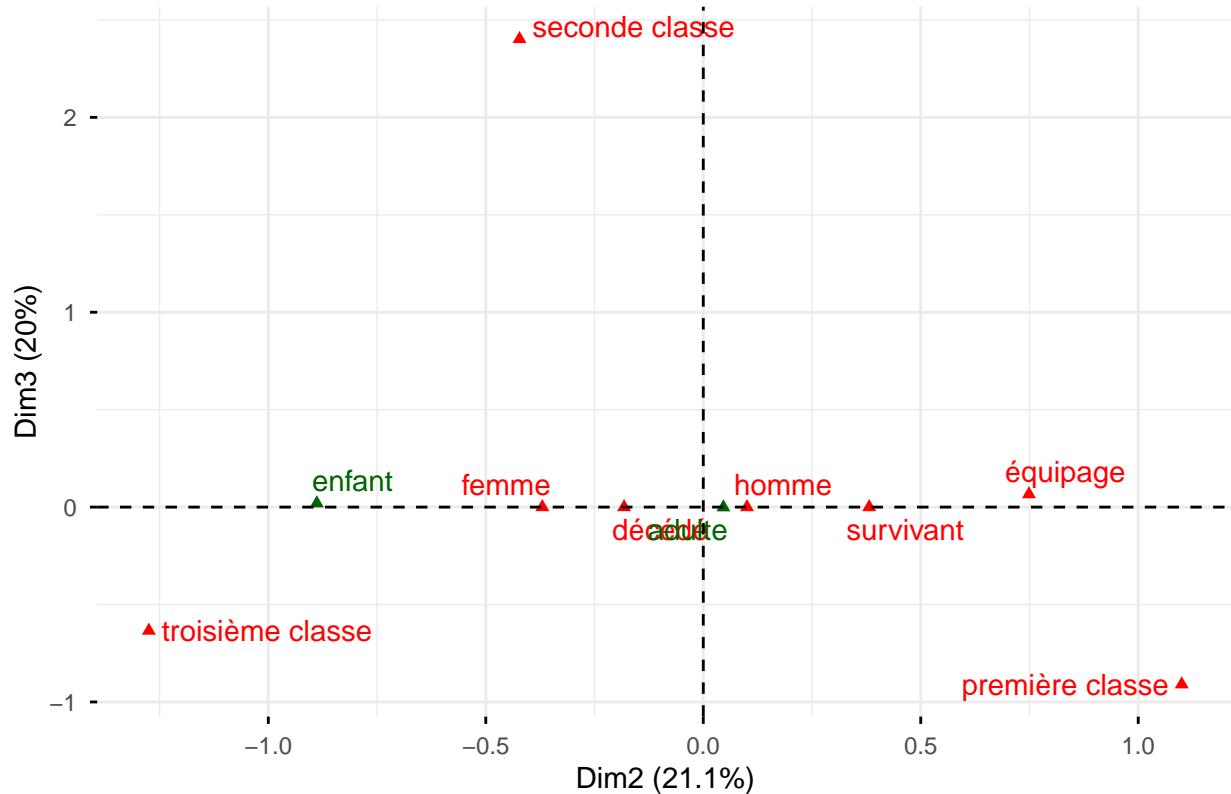


On constate aussi que plusieurs hommes de l'équipage sont décédés, alors que les femmes des première et seconde classes ont survécu. Par conséquent, l'axe 3 oppose seconde classe à première classe et troisième classe.

Visualisation sur le plan (2,3)

```
fviz_mca_var(  
  res.mca2,  
  repel = TRUE, axes=c(2,3)  
)
```

Variable categories – MCA



Nous observons que l'axe 3 sépare la classe moyenne (seconde classe) des classes extrêmes (première et troisième).

7.2. Scores des individus

```
ind2 <- res.mca2$ind
head(ind$coord, 10)

##           Dim 1      Dim 2      Dim 3      Dim 4      Dim 5
## 1  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 2  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 3  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 4  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 5  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 6  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 7  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 8  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 9  0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
## 10 0.6461159 -0.8083508 -0.4404294  0.1905671  0.6017006
```

Séparation des groupes avec habillage sur le Plan (1,2)

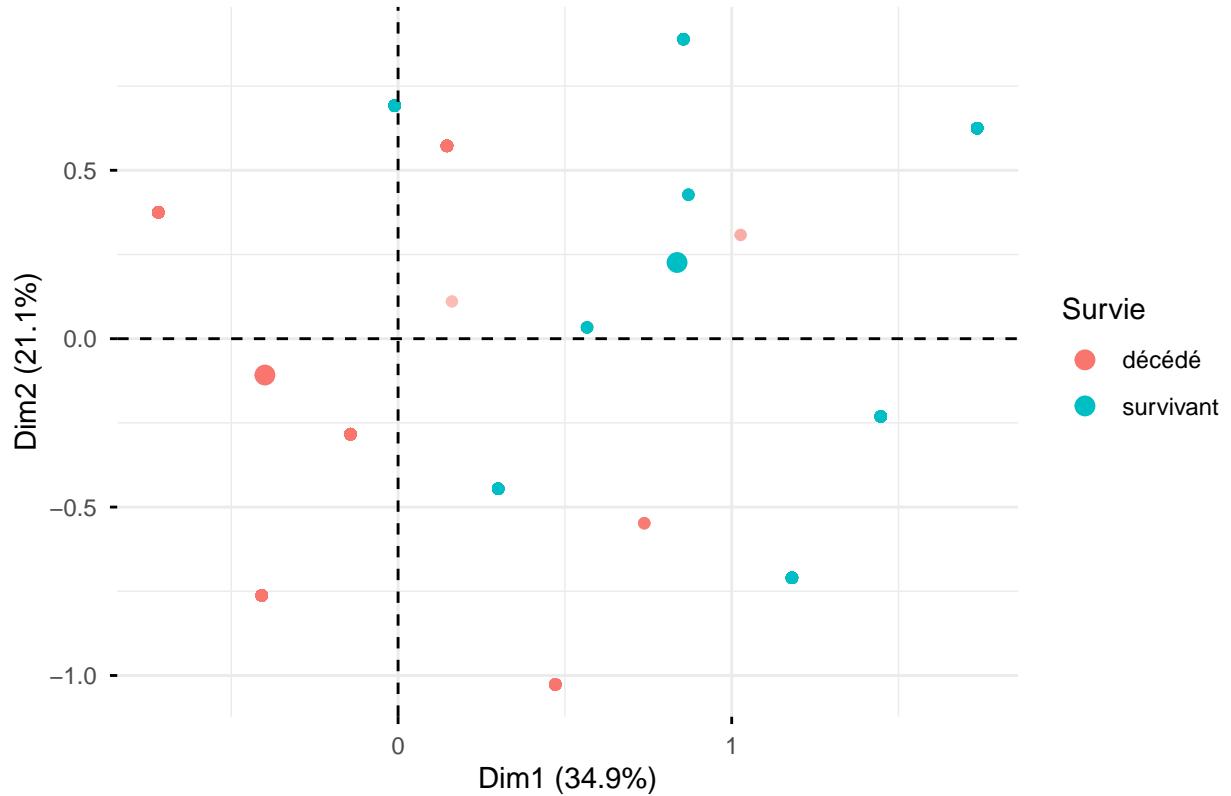
```
fviz_mca_ind(res.mca2,
              axes = c(1,2),
              geom = "point",
              habillage = "Survie",
```

```

alpha.ind = 0.2
)

```

Individuals – MCA



On observe des points décédés et survivants dispersés sur tout le plan factoriel. Comme précédemment, on devine une légère tendance avec davantage de survivants dans la partie droite du graphique, mais les axes 1 et 2 ne permettent pas de distinguer nettement les survivants des décédés.

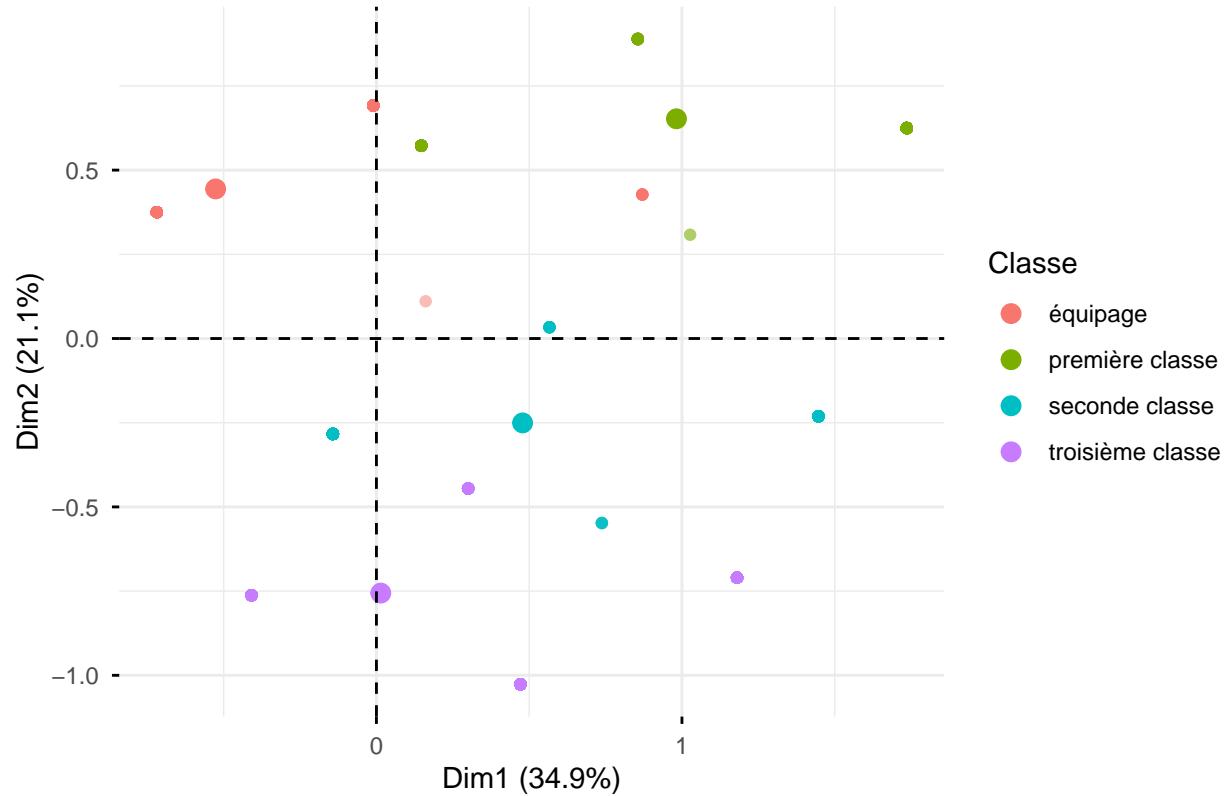
Séparation des groupes avec habillage sur le Plan (1,2)

```

fviz_mca_ind(res.mca2,
              axes = c(1,2),
              geom = "point",
              habillage = "Classe",
              alpha.ind = 0.2
)

```

Individuals – MCA



On voit des points de couleurs différentes correspondant aux quatre classes répartis sur tout le plan factoriel. Comme pour le graphique précédent, quelques regroupements locaux apparaissent, mais les deux premiers axes ne séparent pas nettement les classes. Les passagers des différentes classes restent largement mélangés dans ce plan.

En mettant l'âge en variable supplémentaire, nous avons empêché la modalité "Enfant" d'influencer la construction des axes.

8. Conclusion générale

Pour survivre, il ne suffisait pas d'être de la première classe ou d'être une femme ou un enfant ; c'était une combinaison de plusieurs facteurs. La variable âge a joué un rôle important dans cette étude. Elle a permis de sauver des enfants qui, à cause de leur classe sociale inférieure (troisième classe pour la plupart), auraient normalement eu très peu de chances de s'en sortir. Cependant, les enfants n'ont pas été priorisés comme ça se devrait.