République Algérienne Démocratique et Populaire
الجمهوريــة الجزائريــة الديمقراطيــة الشـعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليــم العــالي و البحــث العلمــي

ECOLE NATIONALE SUPÉRIEURE D'INFORMATIQUE

المدرسة الوطنية العليا للإعلام الآلي
Ecole nationale Supérieure d'Informatique

Rapport TP ANAD

2 ème année Cycle Supérieur (2CS)

Option : Systèmes Informatiques (SQ)

Thème :

Étude d'une société américaine

Réalisé par :

BILAL BELLI

Promotion 2022/2023

1. Table de matière

2. Descriptive sur les données

Les données étudiées sont issues d'un sondage général (social) dans une société américaine (USA).

A partir de ces données on essaie de trouver la relation entre le nombre d'années d'étude d'un individu à son avenir professionnel (profession, classe de la profession, situation familiale).

Le nombre d'individus pour cette étude est 1499.

J'ai traité 6 variables, qui ont les modalités comme suit:

Âge	Sexe	Niveau Education	Class de travail	Occupation	situation
17 à 90	Male Female	Preschool Some-college Prof-school 1st-4th 5th-6th 7th-8th 9th 10th 11th 12th Assoc-acdm Assoc-voc HS-grad Bachelors Masters Doctorate	Federal-gov Local-gov State-gov Private self-emp-inc self-emp-not-inc NA	Adm-clerecal Armed-Forces Craft-repair Other-service Prof-specialty Machine-op-inspct Exec-managerial Tech-support Transport-moving Protective-serv Farming-fishing Sales Priv-house-serv Handlers-cleaners NA	Husband Not-in-family Other-relative Own-child Unmarried Wife

3. Nettoyage et insertion des données sous l'environnement R studio

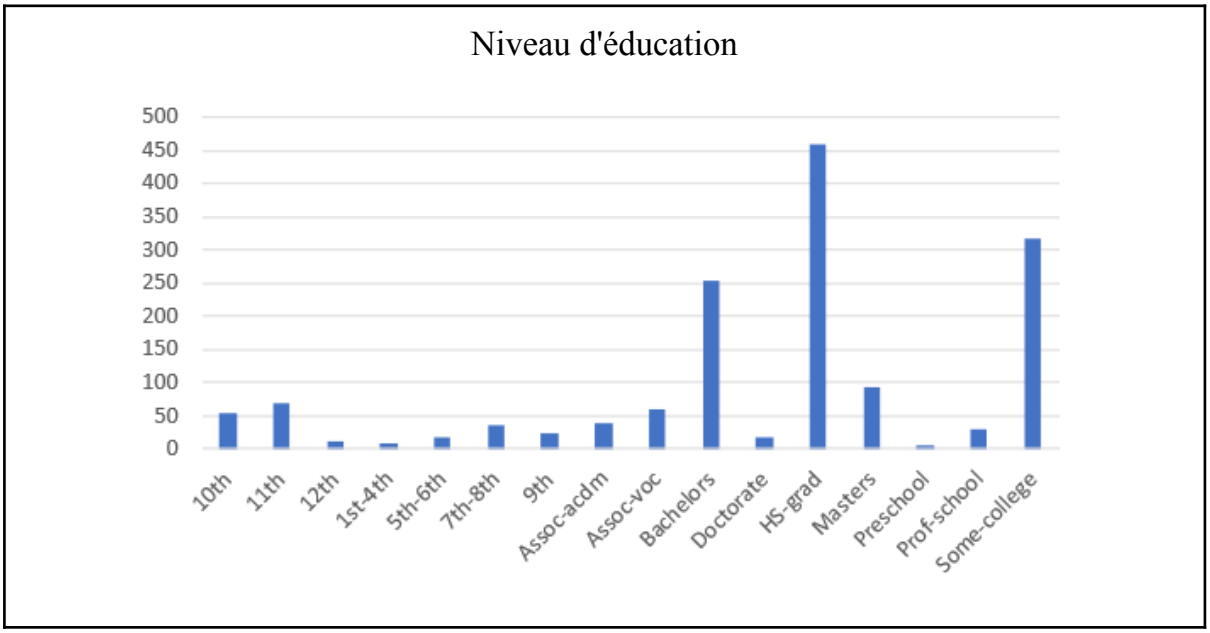
J'ai ignoré les deux premiers variables, car je m'intéresse dans mon étude à l'impact du niveau d'éducation sur la profession au futur.

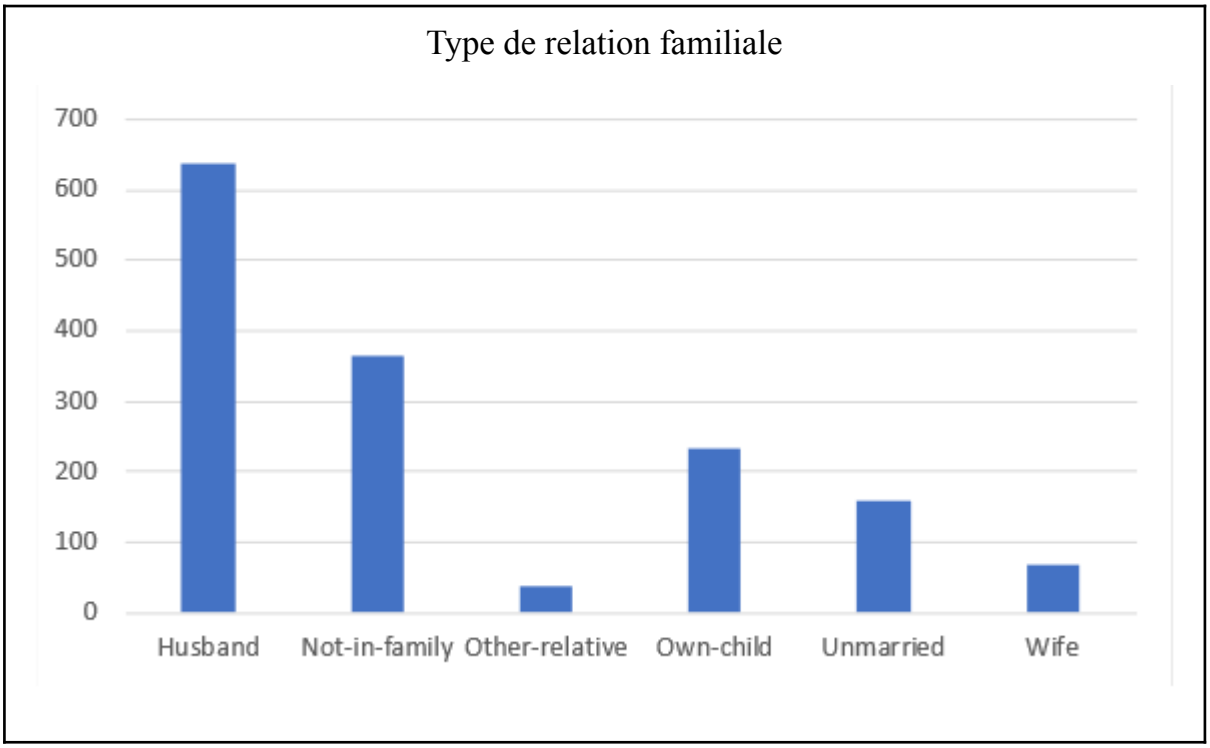
```
library("FactoMineR")
library("ggplot2")
library("factoextra")
```

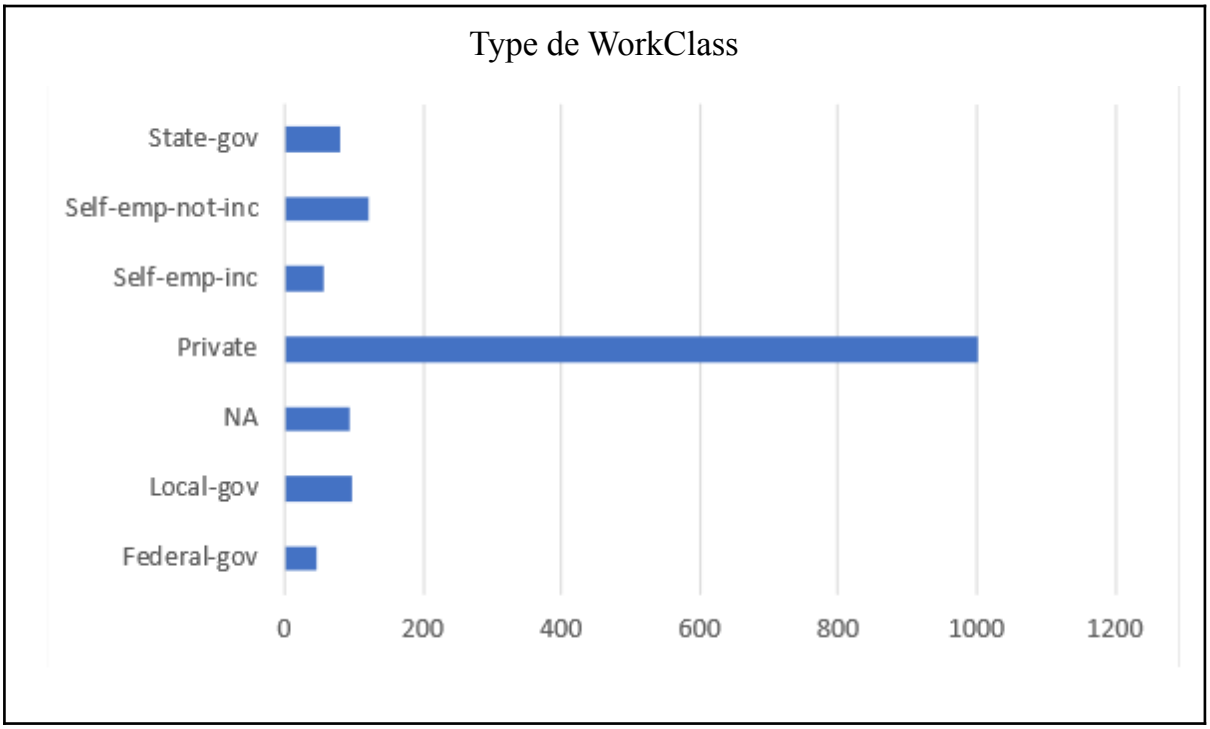
```
library("readxl")
donnes <- read_excel("C:/Users/Hp/OneDrive/Bureau/TP.xls")
donnes <- donnes[,c(3,4,5,6)]
```

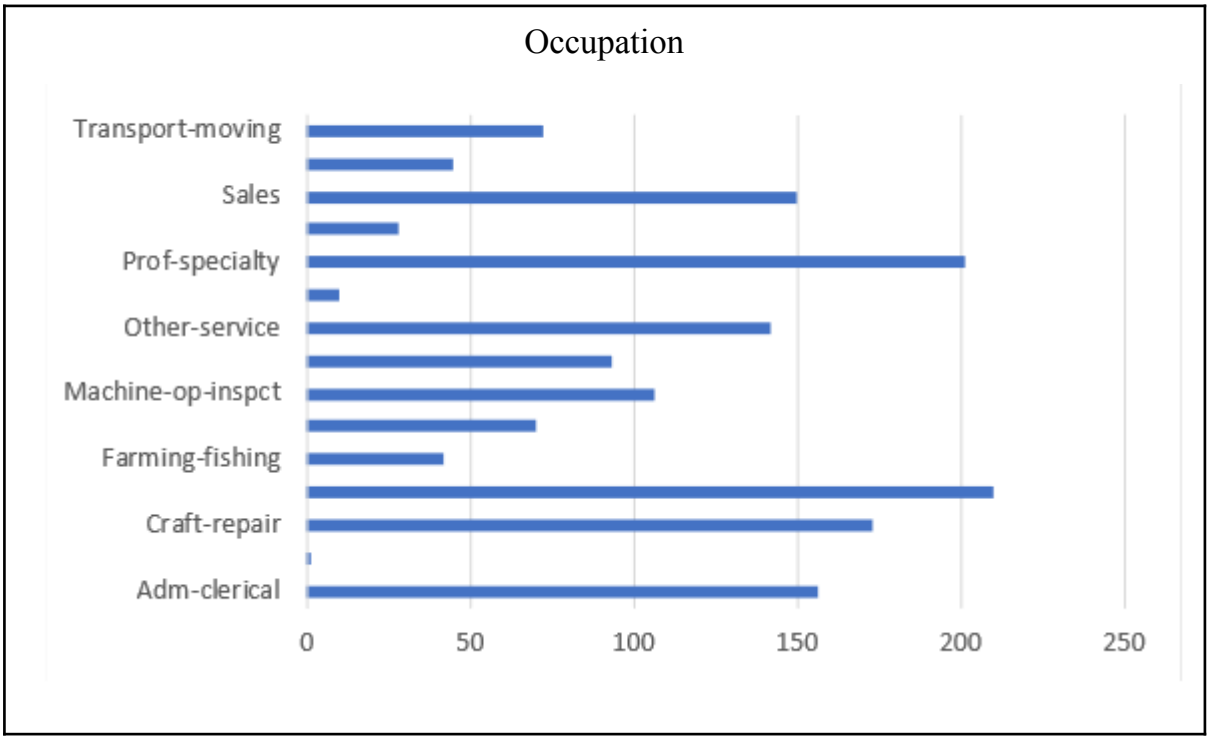
	Education	WorkClass	Occupation	RelationShip
1	11th	Private	Machine-op-inspct	Own-child
2	HS-grad	Private	Farming-fishing	Husband

4. Graphiques sur les modalités de chaque variable









5. Analyse des correspondances multiples sur les variables:

```
res.mca <- MCA(donnes, graph = FALSE)
print(res.mca)
```

```
**Results of the Multiple Correspondence Analysis (MCA)**
The analysis was performed on 1499 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$v.test"    "v-test for the categories"
7  "$ind"           "results for the individuals"
8  "$ind$coord"     "coord. for the individuals"
9  "$ind$cos2"      "cos2 for the individuals"
10 "$ind$contrib"   "contributions of the individuals"
11 "$call"          "intermediate results"
12 "$call$marge.col" "weights of columns"
13 "$call$marge.li"  "weights of rows"
```

a. Les valeurs propres

```
eig.val <- get_eigenvalue(res.mca)
```

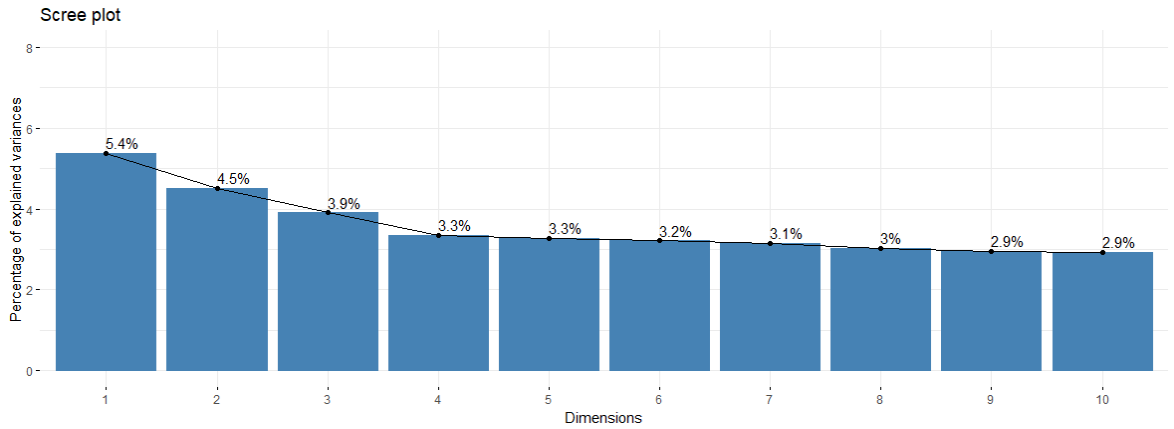
	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	5.368811e-01	5.368811e+00	5.368811
Dim.2	4.518733e-01	4.518733e+00	9.887544
Dim.3	3.902432e-01	3.902432e+00	13.789975
Dim.4	3.348970e-01	3.348970e+00	17.138945
Dim.5	3.273728e-01	3.273728e+00	20.412673
Dim.6	3.210681e-01	3.210681e+00	23.623354
Dim.7	3.141675e-01	3.141675e+00	26.765029
Dim.8	3.031672e-01	3.031672e+00	29.796701
Dim.9	2.947724e-01	2.947724e+00	32.744425
Dim.10	2.919422e-01	2.919422e+00	35.663846
Dim.11	2.846398e-01	2.846398e+00	38.510244
Dim.12	2.822006e-01	2.822006e+00	41.332251
Dim.13	2.783337e-01	2.783337e+00	44.115587
Dim.14	2.708988e-01	2.708988e+00	46.824576
Dim.15	2.689803e-01	2.689803e+00	49.514378
Dim.16	2.639826e-01	2.639826e+00	52.154204
Dim.17	2.603182e-01	2.603182e+00	54.757386
Dim.18	2.574955e-01	2.574955e+00	57.332341
Dim.19	2.533034e-01	2.533034e+00	59.865376

b. Le taux d'inertie représenté par les valeurs propres

J'ai ajusté l'affichage seulement des 10 premières grandes valeurs propres (il y a au total 48 valeurs propres).

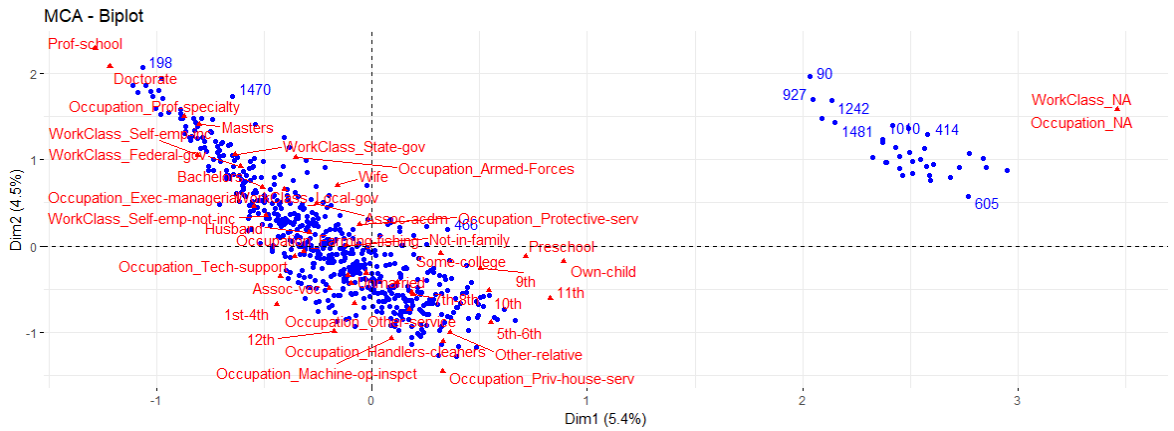
J'ai choisi les deux premiers axes (valeurs propres) pour représenter les résultats de l'analyse.

```
fviz_screeplot (res.mca, addlabels = TRUE, ylim = c (0, 8))
```



c. La représentation des individus et modalités sur le même plan

```
fviz_mca_biplot (res.mca, repel = TRUE,ggtheme = theme_minimal())
```



6. Etude des variables

```
var <- get_mca_var(res.mca)
var
```

```
Multiple Correspondence Analysis Results for variables
 ===================================================
  Name       Description
1 "$coord"   "Coordinates for categories"
2 "$cos2"    "Cos2 for categories"
3 "$contrib" "contributions of categories"
```

a. Les coordonnées des variables

```
# Coordonnées
round(var$coord,2)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
10th	0.54	-0.52	-0.06	-0.26	0.15
11th	0.83	-0.60	-0.06	0.46	-0.36
12th	-0.17	-0.99	0.65	0.78	-0.63
1st-4th	-0.44	-0.68	1.91	-0.17	3.63
5th-6th	0.55	-0.89	0.29	-0.10	-2.87
7th-8th	0.19	-0.56	1.22	-0.02	-0.67
9th	0.50	-0.26	1.02	0.41	-0.25
Assoc-acdm	-0.26	0.49	-0.69	0.66	-0.22
Assoc-voc	-0.43	-0.35	0.36	-0.05	0.85
Bachelors	-0.51	0.68	-0.31	-0.67	0.10
Doctorate	-1.22	2.08	0.20	-0.05	-0.32
HS-grad	0.18	-0.54	0.00	-0.19	-0.28
Masters	-0.80	1.41	-0.63	0.07	-0.22
Preschool	0.71	-0.12	1.69	-0.98	3.92
Prof-school	-1.29	2.29	1.09	1.04	-1.85
Some-college	0.32	-0.09	0.02	0.54	0.61
WorkClass_Federal-gov	-0.61	0.92	-1.37	-1.04	1.00
WorkClass_Local-gov	-0.41	0.65	-0.57	2.24	-0.26
WorkClass_NA	3.46	1.57	0.33	-0.38	-0.16
WorkClass_Private	-0.10	-0.44	-0.12	-0.13	-0.09
WorkClass_Self-emp-inc	-0.81	1.05	1.12	-0.80	-1.30
WorkClass_Self-emp-not-inc	-0.49	0.35	1.67	0.02	0.97
WorkClass_State-gov	-0.64	1.06	-0.69	0.48	0.56
Occupation_Adm-clerical	-0.03	-0.31	-1.30	-0.19	0.74
Occupation_Armed-Forces	-0.35	1.03	-3.35	-2.60	6.04
Occupation_Craft-repair	-0.20	-0.50	1.08	0.04	-0.34
Occupation_Exec-managerial	-0.55	0.46	-0.10	-0.87	-0.07
Occupation_Farming-fishing	-0.32	-0.06	2.32	0.21	3.30
Occupation_Handlers-cleaners	0.33	-1.10	-0.20	0.31	-0.19
Occupation_Machine-op-inspct	0.09	-1.07	0.04	-0.21	-0.86
Occupation_NA	3.46	1.57	0.33	-0.38	-0.16

Occupation_Other-service	0.17	-0.74	-0.59	0.45	0.08
Occupation_Priv-house-serv	0.33	-1.46	-0.43	-0.08	-2.15
Occupation_Prof-specialty	-0.87	1.50	-0.28	0.52	-0.52
Occupation_Protective-serv	-0.06	0.24	-0.47	4.31	0.78
Occupation_Sales	-0.11	-0.34	0.27	-0.40	0.06
Occupation_Tech-support	-0.36	-0.12	-0.48	-0.15	0.66
Occupation_Transport-moving	-0.08	-0.66	0.85	0.18	-0.18
Husband	-0.29	0.15	0.74	0.06	-0.16
Not-in-family	-0.12	0.00	-0.53	-0.38	0.26
Other-relative	0.36	-1.00	-0.31	0.80	-1.10
Own-child	0.89	-0.18	-0.39	0.67	0.31
Unmarried	0.12	-0.42	-0.68	-0.14	-0.40
Wife	-0.16	0.70	-0.96	-0.95	0.54

b. La qualité de représentation des variables

```
# Cos2: qualité de représentation
round(var$cos2,2)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
10th	0.01	0.01	0.00	0.00	0.00
11th	0.03	0.02	0.00	0.01	0.01
12th	0.00	0.01	0.00	0.01	0.00
1st-4th	0.00	0.00	0.02	0.00	0.07
5th-6th	0.00	0.01	0.00	0.00	0.09
7th-8th	0.00	0.01	0.04	0.00	0.01
9th	0.00	0.00	0.02	0.00	0.00
Assoc-acdm	0.00	0.01	0.01	0.01	0.00
Assoc-voc	0.01	0.01	0.01	0.00	0.03
Bachelors	0.05	0.09	0.02	0.09	0.00
Doctorate	0.02	0.06	0.00	0.00	0.00
HS-grad	0.02	0.13	0.00	0.02	0.03
Masters	0.04	0.13	0.03	0.00	0.00
Preschool	0.00	0.00	0.01	0.00	0.06
Prof-school	0.03	0.11	0.02	0.02	0.07
Some-college	0.03	0.00	0.00	0.08	0.10
WorkClass_Federal-gov	0.01	0.03	0.06	0.04	0.03
WorkClass_Local-gov	0.01	0.03	0.02	0.35	0.00
WorkClass_NA	0.79	0.16	0.01	0.01	0.00
WorkClass_Private	0.02	0.39	0.03	0.03	0.02
WorkClass_Self-emp-inc	0.03	0.04	0.05	0.03	0.07
WorkClass_Self-emp-not-inc	0.02	0.01	0.25	0.00	0.08
WorkClass_State-gov	0.02	0.06	0.03	0.01	0.02
Occupation_Adm-clerical	0.00	0.01	0.20	0.00	0.06
Occupation_Armed-Forces	0.00	0.00	0.01	0.00	0.02
Occupation_Craft-repair	0.01	0.03	0.15	0.00	0.01
Occupation_Exec-managerial	0.05	0.03	0.00	0.12	0.00
Occupation_Farming-fishing	0.00	0.00	0.16	0.00	0.31
Occupation_Handlers-cleaners	0.01	0.06	0.00	0.00	0.00
Occupation_Machine-op-inspct	0.00	0.09	0.00	0.00	0.06
Occupation_NA	0.79	0.16	0.01	0.01	0.00
Occupation_Other-service	0.00	0.06	0.04	0.02	0.00
Occupation_Priv-house-serv	0.00	0.01	0.00	0.00	0.03
Occupation_Prof-specialty	0.12	0.35	0.01	0.04	0.04
Occupation_Protective-serv	0.00	0.00	0.00	0.35	0.01
Occupation_Sales	0.00	0.01	0.01	0.02	0.00
Occupation_Tech-support	0.00	0.00	0.01	0.00	0.01
Occupation_Transport-moving	0.00	0.02	0.04	0.00	0.00
Husband	0.06	0.02	0.40	0.00	0.02
Not-in-family	0.00	0.00	0.09	0.05	0.02
Other-relative	0.00	0.03	0.00	0.02	0.03
Own-child	0.15	0.01	0.03	0.08	0.02
Unmarried	0.00	0.02	0.06	0.00	0.02
Wife	0.00	0.02	0.04	0.04	0.01

c. La contribution des variables

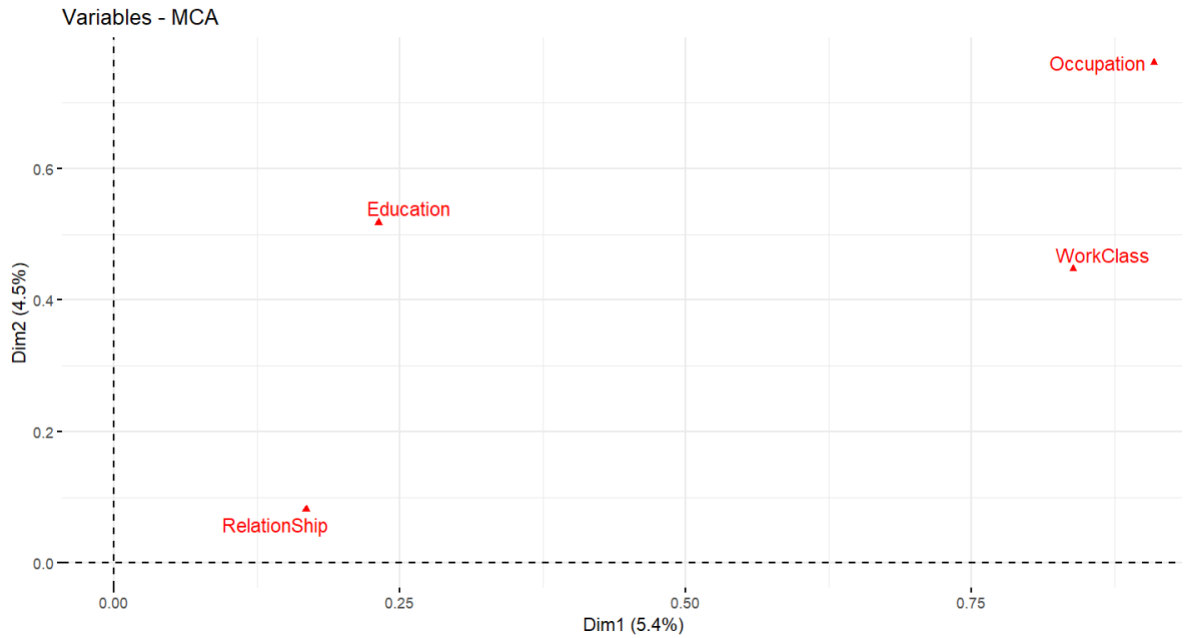
```
# Contributions aux axes
round(var$contrib,2)
```

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5
10th	0.51	0.55	0.01	0.19	0.06
11th	1.49	0.94	0.01	0.75	0.46
12th	0.01	0.47	0.24	0.39	0.26
1st-4th	0.05	0.14	1.25	0.01	5.38
5th-6th	0.16	0.49	0.06	0.01	7.13
7th-8th	0.04	0.43	2.36	0.00	0.84
9th	0.19	0.06	1.07	0.20	0.08
Assoc-acdm	0.08	0.33	0.77	0.83	0.09
Assoc-voc	0.34	0.28	0.34	0.01	2.21
Bachelors	2.04	4.31	1.07	5.76	0.13
Doctorate	0.87	3.04	0.03	0.00	0.10
HS-grad	0.49	4.94	0.00	0.81	1.80
Masters	1.86	6.79	1.56	0.02	0.23
Preschool	0.09	0.00	0.73	0.29	4.69
Prof-school	1.54	5.80	1.53	1.61	5.26
Some-college	1.01	0.10	0.00	4.55	5.99
WorkClass_Federal-gov	0.55	1.49	3.83	2.61	2.45
WorkClass_Local-gov	0.50	1.51	1.36	24.30	0.34
WorkClass_NA	34.63	8.50	0.42	0.68	0.13
WorkClass_Private	0.29	7.11	0.63	0.79	0.45
WorkClass_Self-emp-inc	1.18	2.35	3.11	1.86	5.03
WorkClass_Self-emp-not-inc	0.91	0.54	14.43	0.00	5.83
WorkClass_State-gov	0.99	3.25	1.59	0.91	1.27
Occupation_Adm-clerical	0.00	0.56	11.23	0.28	4.35
Occupation_Armed-Forces	0.00	0.04	0.48	0.34	1.86
Occupation_Craft-repair	0.21	1.57	8.64	0.01	1.01
Occupation_Exec-managerial	1.96	1.61	0.09	7.92	0.05
Occupation_Farming-fishing	0.13	0.01	9.68	0.09	23.30
Occupation_Handlers-cleaners	0.24	3.13	0.13	0.33	0.13
Occupation_Machine-op-inspct	0.03	4.51	0.01	0.23	3.99

Occupation_NA	34.63	8.50	0.42	0.68	0.13
Occupation_Other-service	0.13	2.87	2.11	1.45	0.04
Occupation_Priv-house-serv	0.03	0.78	0.08	0.00	2.35
Occupation_Prof-specialty	4.73	16.65	0.68	2.72	2.73
Occupation_Protective-serv	0.00	0.06	0.26	25.86	0.87
Occupation_Sales	0.06	0.63	0.45	1.17	0.02
Occupation_Tech-support	0.18	0.02	0.45	0.05	1.00
Occupation_Transport-moving	0.01	1.16	2.22	0.12	0.12
Husband	1.68	0.55	14.78	0.11	0.79
Not-in-family	0.15	0.00	4.43	2.57	1.27
Other-relative	0.16	1.41	0.16	1.22	2.35
Own-child	5.71	0.28	1.51	5.16	1.17
Unmarried	0.07	1.03	3.16	0.15	1.28
Wife	0.05	1.22	2.63	2.98	0.99

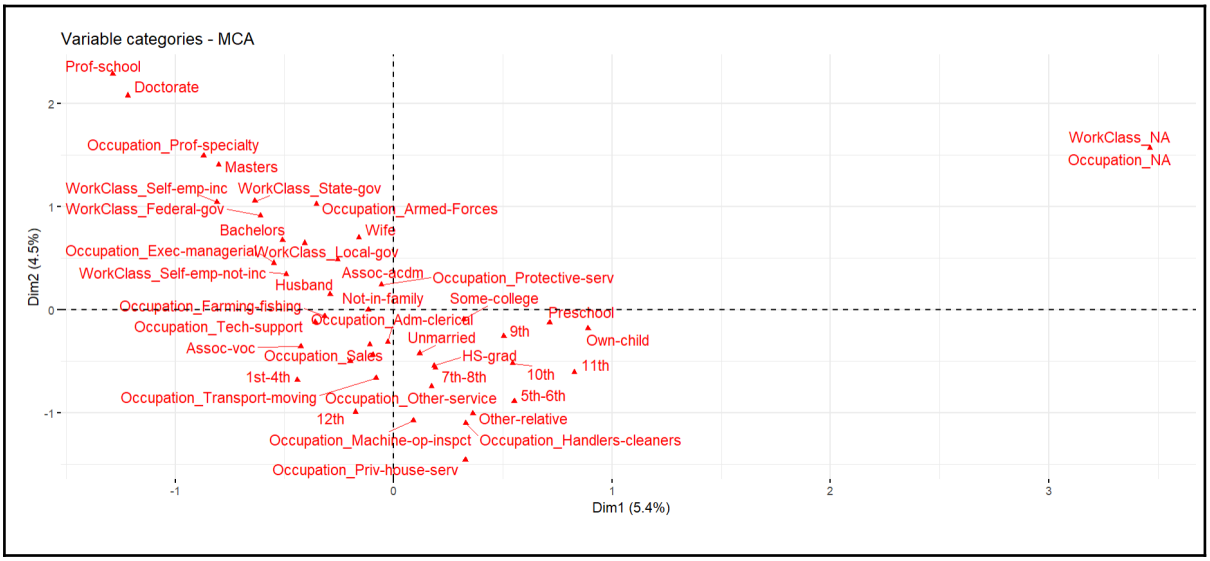
d. représentation des variables

```
fviz_mca_var (res.mca, choice = "mca.cor", repel = TRUE,ggtheme = theme_minimal ())
```



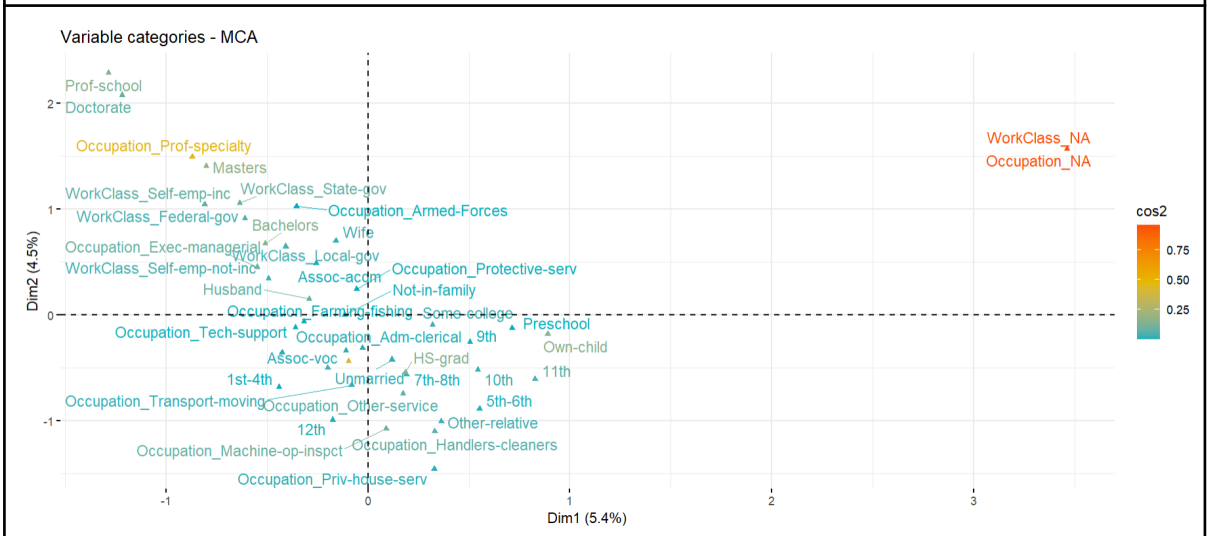
e. représentation des modalités sur le plan

```
fviz_mca_var (res.mca,repel = TRUE,ggtheme = theme_minimal ())
```

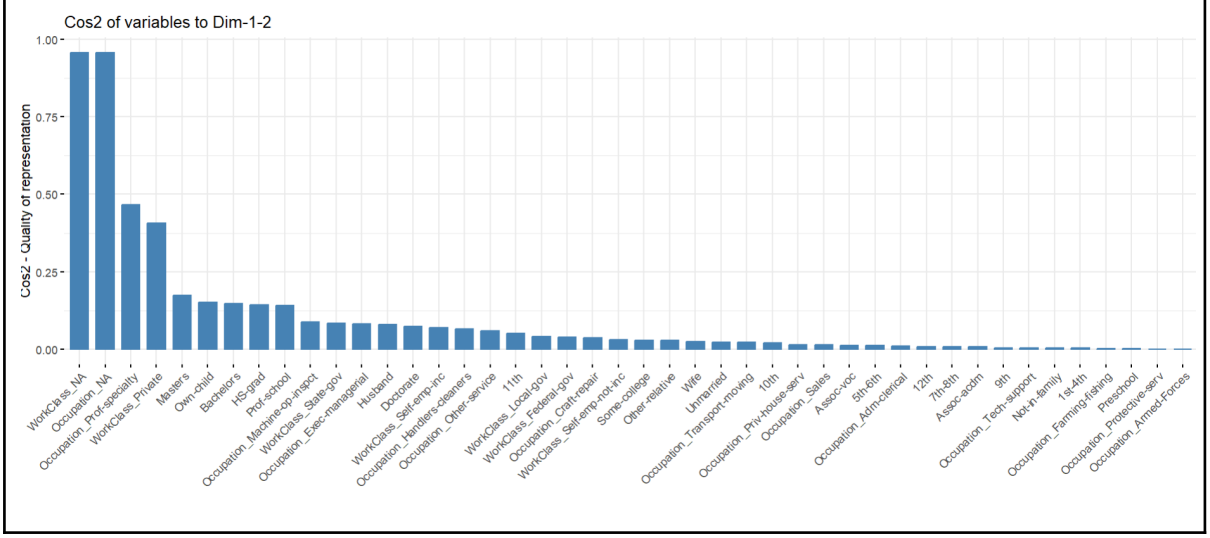


f. qualité de représentation des modalités (sur les deux axes)

fviz_mca_var(res.mca, col.var = "cos2",gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),repel = TRUE, ggtheme = theme_minimal())

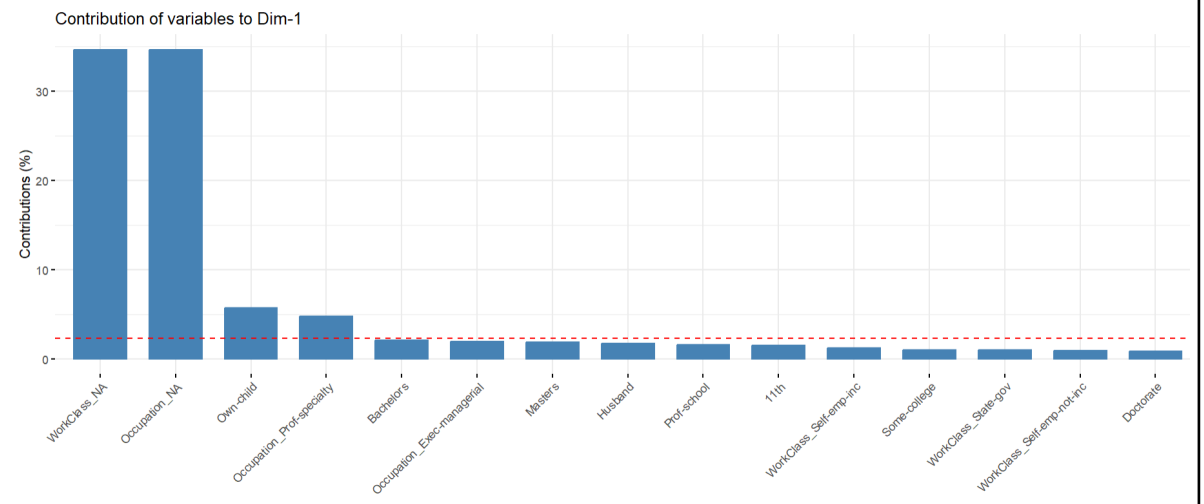


fviz_cos2(res.mca, choice = "var", axes = 1:2)



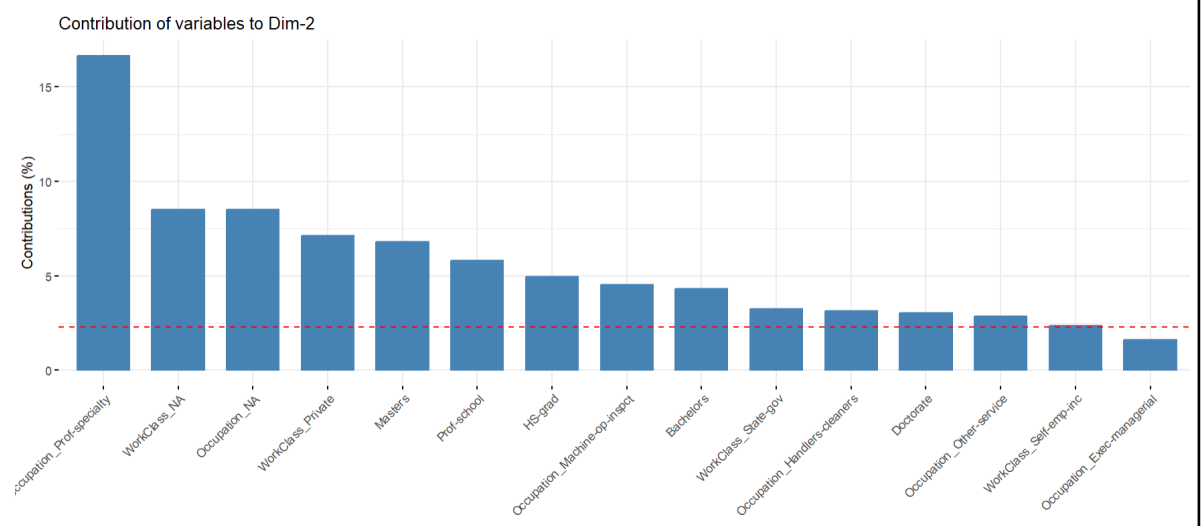
g. contribution des modalités (sur l'axe 1)

```
fviz_contrib (res.mca, choice = "var", axes = 1, top = 15)
```



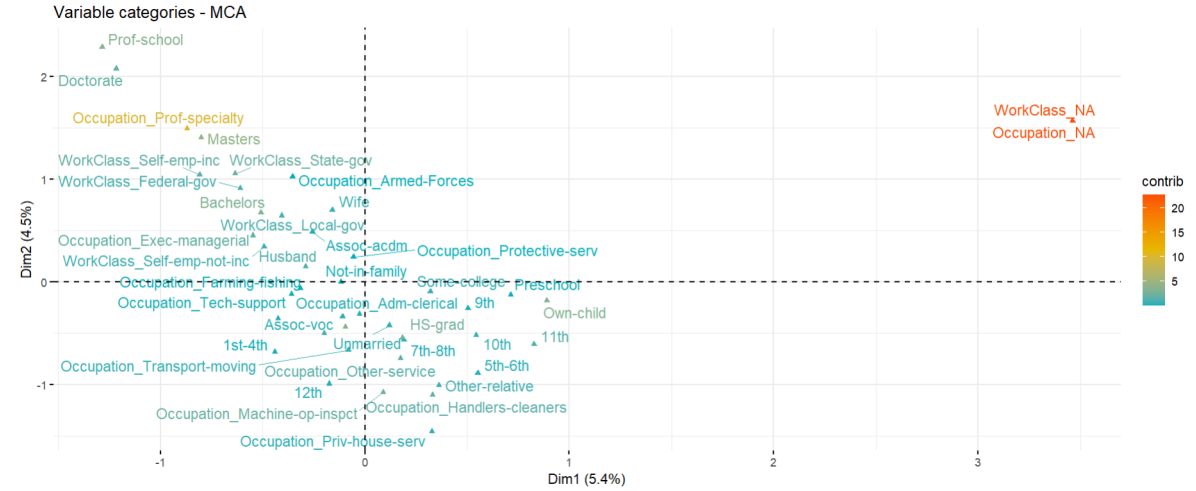
h. contribution des modalités (sur l'axe 2)

```
fviz_contrib (res.mca, choice = "var", axes = 2, top = 15)
```



i. contribution des modalités (sur les deux axes)

```
fviz_mca_var(res.mca, col.var = "contrib",gradient.cols = c("#00AFBB", "#E7B800",
"#FC4E07"),repel = TRUE,ggtheme = theme_minimal())
```



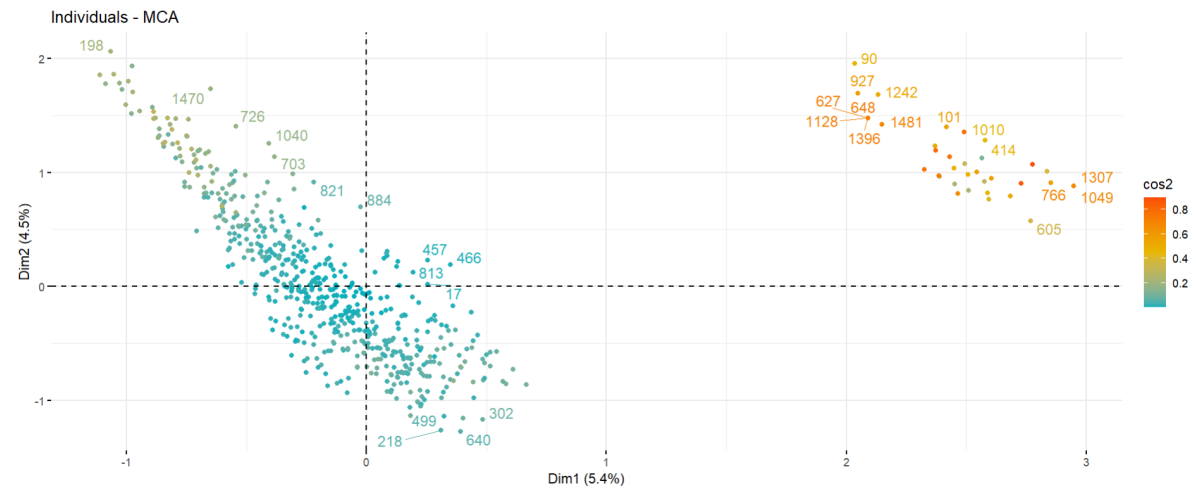
7. Etude des individus

```
ind <- get_mca_ind (res.mca)
ind
```

```
Multiple Correspondence Analysis Results for individuals
 ===================================================
  Name       Description
1 "$coord"   "Coordinates for the individuals"
2 "$cos2"    "Cos2 for the individuals"
3 "$contrib" "contributions of the individuals"
```

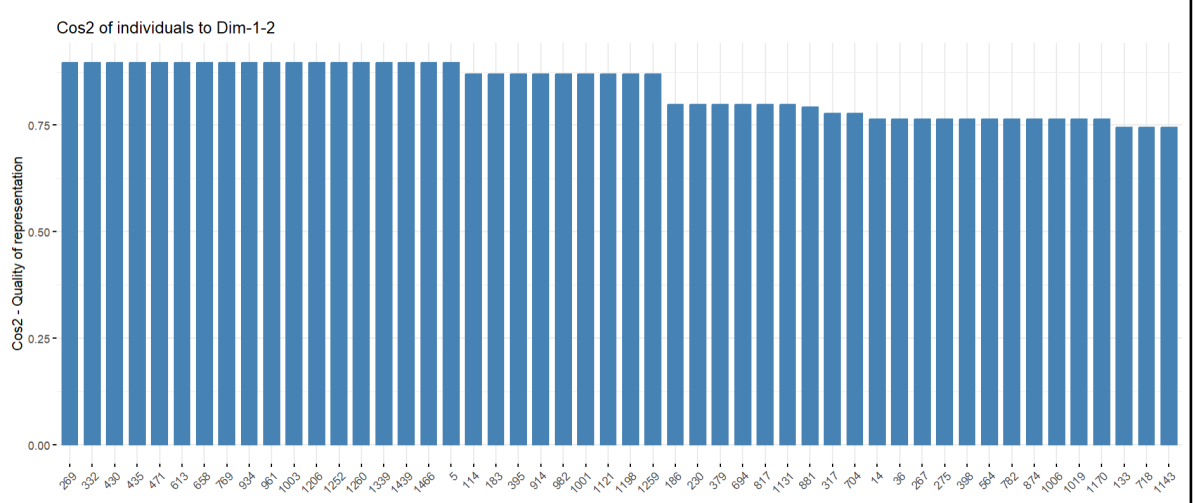
a. représentation des individus sur le plan

```
fviz_mca_ind(res.mca, col.ind = "cos2", gradient.cols = c("#00AFBB", "#E7B800",
"#FC4E07"), repel = TRUE,ggtheme = theme_minimal())
```



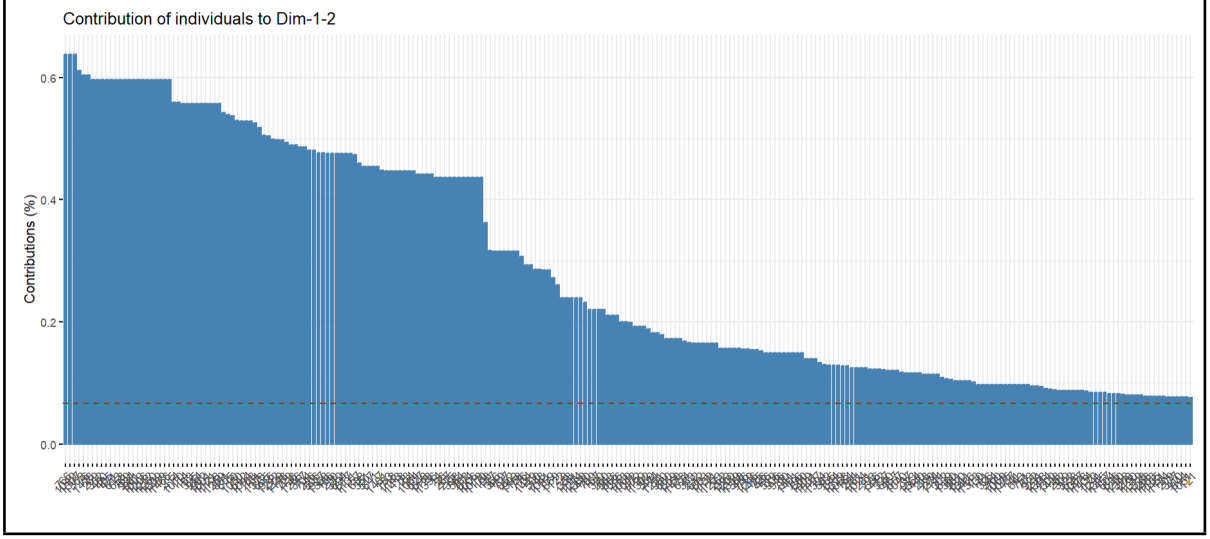
b. qualité de représentation pour quelques individus

```
# Cos2 des individus
fviz_cos2(res.mca, choice = "ind", axes = 1:2, top = 50)
```



c. contribution de 250 individus (ordonnés selon la plus grand contribution)

```
# Contribution des individus aux dimensions
fviz_contrib(res.mca, choice = "ind", axes = 1:2, top = 250)
```



8. Source

https://raw.githubusercontent.com/selva86/datasets/master/adult_test.csv