

Projet analyse de données : Les défis de la femme travailleuse

Encadré par :

- Mr. BARRY Khalidou
- Mr. HARTI Mostapha

Réalisé par :

- CHAOUI Marwa
- TAGMOUTI Ghita
- BELHAJ Ikram
- ABDELLAH Qatre en nada
- EL IDRISSE Yassine
- LAKHAL Khalid

Sommaire :

1 – Introduction

2 – Enquete et la collecte des données

3 – Langage R

4 – Tanagra

5 – Régression Linéaire

6 – Analyse en composantes principales (ACP)

7 – Analyse factorielle des correspondances (AFC)

8 – Conclusion

Introduction

Dans le contexte actuel du Maroc, l'équilibre entre vie professionnelle et familiale des femmes est un enjeu crucial. Cette étude examine diverses variables telles que l'âge, le secteur d'emploi et la satisfaction professionnelle pour mieux comprendre les défis et les opportunités auxquels sont confrontées ces femmes. En analysant ces données, l'objectif est de dégager des tendances significatives afin d'éclairer les politiques et les pratiques visant à soutenir leur intégration professionnelle et leur épanouissement personnel, contribuant ainsi à une société plus équitable et inclusive.

Cette étude vise également à sensibiliser davantage l'opinion publique et les décideurs sur l'importance de créer un environnement favorable pour les femmes sur le marché du travail. En mettant en lumière les défis spécifiques auxquels elles sont confrontées et en identifiant les leviers d'action potentiels, elle aspire à encourager des initiatives politiques et sociales qui favorisent l'égalité des chances et la promotion de la diversité dans tous les secteurs de l'économie. En renforçant la reconnaissance et la valorisation du rôle des femmes dans la société, cette démarche contribue à créer un climat propice à leur épanouissement professionnel et personnel, tout en favorisant un développement durable et inclusif pour l'ensemble de la population marocaine.

Enquete et la collecte des données

➔ Nous avons entrepris une enquête approfondie portant sur les femmes actives, en utilisant une méthodologie mixte. Pour ce faire, nous avons combiné l'utilisation de deux logiciels : Sphinx, pour la distribution du questionnaire sur support papier, et Google Forms, pour sa diffusion en ligne. Dans le cadre de cette enquête, un total de 27 questions a été posé, couvrant divers aspects liés à la vie professionnelle des femmes. Nous avons reçu un total de 55 réponses, représentant ainsi une base de données substantielle pour notre analyse. Toutefois, pour garantir la qualité des données, nous avons procédé à un nettoyage préliminaire du fichier initial. Cela a impliqué la suppression des valeurs aberrantes et la substitution des valeurs manquantes par la moyenne de chaque question, assurant ainsi la cohérence et la fiabilité de notre ensemble de données.

La femme Travailleuse

1. Niveau d'éducation ?
☐ 1. moins que Bac
☐ 2. Bac
☐ 3. Bac+1
☐ 4. Bac+2
☐ 5. Bac+3
☐ 6. Bac+4
☐ 7. Bac+5
☐ 8. plus que Bac+5

2. Secteur emploi
☐ 1. Privé
☐ 2. Public
☐ 3. Hybride

3. Situation familiale ?
☐ 1. Célibataire
☐ 2. Marié
☐ 3. Veuf

4. Âge ?

5. Combien d'heures travaillez-vous par semaine en moyenne ?

6. Sur une échelle de 1 à 10, évaluez votre satisfaction globale par rapport à votre carrière jusqu'à présent.
La réponse doit être comprise entre 1 et 10.

7. Disposez-vous d'options de travail flexible dans votre emploi actuel (Télétravail, horaires flexibles) ?
☐ 1. Oui
☐ 2. Non

8. Combien d'années avez-vous passées dans votre emploi actuel ?

9. Sur une échelle de 1 à 10, à quel point estimez-vous votre niveau de stress lié au travail ?
La réponse doit être comprise entre 1 et 10.

10. Avez-vous déjà été confronté(e) à des situations de discrimination au travail en raison de votre genre ?
☐ 1. Oui
☐ 2. Non

11. Quel est votre revenu annuel brut provenant de votre travail estimée ?
☐ 1. moins de 1500
☐ 2. 1500-3000
☐ 3. 3000-5000
☐ 4. 5000-7000
☐ 5. 7000 et plus

12. Combien de fois par mois avez-vous des conversations ou des négociations sur votre salaire ou vos conditions de travail avec votre employeur ?

13. Combien de formations professionnelles avez-vous suivies au cours de la dernière année ?

14. Dans quelle mesure estimez-vous que les opportunités de développement professionnel soient égales pour les hommes et les femmes dans votre secteur d'activité ?
☐ 1. Faible
☐ 2. Moyen
☐ 3. Forte

15. À quelle fréquence participez-vous à des activités de réseautage professionnel en dehors des heures de travail ?

16. Nombre d'enfants ?
☐ 1. 0
☐ 2. 1-3
☐ 3. 3 et plus

17. Combien d'heures de sommeil obtenez-vous en moyenne par nuit ?

18. Comment les politiques familiales (congé parental, garde d'enfants) influencent-elles votre décision de travailler ?
☐ 1. Faible
☐ 2. Moyen
☐ 3. Forte

19. Combien de jours de congé maladie avez-vous pris au cours de l'année dernière ?

20. Quel pourcentage de vos revenus est consacré aux frais de garde d'enfants ?
La réponse doit être comprise entre 0 et 100.

21. Quels facteurs, selon vous, encouragent ou dissuadent les femmes de travailler ?

22. Combien de fois par semaine pratiquez-vous une activité de relaxation pour gérer le stress lié au travail ?

23. Combien de fois par mois ressentez-vous des conflits entre vos responsabilités professionnelles et familiales ?

24. Combien de temps consacrez-vous aux tâches ménagères et familiales chaque semaine ?

25. Dans quelle mesure les responsabilités familiales influencent-elles votre capacité à équilibrer votre vie professionnelle et votre vie personnelle ?
☐ 1. Faible
☐ 2. Moyen
☐ 3. Forte

26. À quelle fréquence recevez-vous des opportunités d'avancement professionnel dans votre emploi actuel ?
☐ 1. 0
☐ 2. 1-2
☐ 3. 2 et plus

27. Avez-vous d'autres commentaires ou suggestions concernant cette enquête sur la vie professionnelle des femmes ?

La femme Travailleuse

Connectez-vous à Google pour enregistrer votre progression. En savoir plus

* Indique une question obligatoire

1. Niveau d'éducation?

☐ moins que Bac
☐ Bac
☐ Bac+1
☐ Bac+2
☐ Bac+3
☐ Bac+4
☐ Bac+5
☐ Plus que Bac+5

2. Secteur emploi?

☐ Privé
☐ Public
☐ Hybride

3. Situation familiale?

☐ Célibataire
☐ Marié
☐ Veuf

4. Âge? *

Sélectionner

5. Combien d'heures travaillez-vous par semaine en moyenne? *

Sélectionner

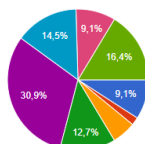
6. Sur une échelle de 1 à 10, évaluez votre satisfaction globale par rapport à votre carrière jusqu'à présent? *

1 2 3 4 5 6 7 8 9 10

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐



1. Niveau d'éducation?
55 réponses



- moins que Bac
- Bac
- Bac+1
- Bac+2
- Bac+3
- Bac+4
- Bac+5
- Plus que Bac+5

 Copier

Signification de chaque variable :

niveau d'education : ne
 secteur d'emploi : se
 situation familiale : sf
 heure de travaille par semaine : hts
 satisfction globale : sg
 travaille flexible :tf
 annees emploi actuel : aea
 niveau de stress lie au travail : st
 discrimination au travail : disc
 revenue mensielle : rm
 negociation salaire : negs
 formation pro : fp
 egalite pro : ep
 acctivite de reseautage pro : arp
 nombre enfants : enfants
 heures sommeils / nuit : sn
 impact politique sur l'emploi : ipe
 congee maladie annee / jour : cmj
 pourcentage de revenue consacre aux frais de garde d'enfants : prfe
 acctivites de relaxation par semaine : ars
 conflits entre responsab prof et familiale / mois : conf
 heures des taches menagers/semaine : hms
 frequence d'opportunités pro : foap

Celibataire 0 Marié 1 Divorcé 2 Veuf 3

-secteur d'emploi : public => 0	privé => 1	hybrid => 2
-travail flexible : oui => 0	non => 1	
-situation de discrimination :	oui => 0	non => 1
-egalite professionnelle : faible => 0	moyen => 1	forte => 2
-l'impact de politique : faible => 0	moyen => 1	forte => 2

Langage R

Le langage R est un outil essentiel en analyse de données, offrant des fonctionnalités avancées pour explorer, manipuler et visualiser des données. Dans notre enquête sur la femme travailleuse, R nous permet d'analyser efficacement les données recueillies, en identifiant des tendances, des corrélations et des facteurs influents. Cela nous aide à prendre des décisions éclairées pour promouvoir l'égalité des sexes sur le lieu de travail.



Tanagra

Tanagra, un logiciel d'analyse de données, offre un outil précieux pour explorer les tendances des travailleuses au Maroc. En intégrant Tanagra dans notre présentation, on peut extraire des insights significatifs et formuler des recommandations pertinentes pour promouvoir l'égalité des sexes sur le marché du travail marocain.



Régression Linéaire

la Régression Linéaire : Une méthode statistique pour modéliser la relation entre une variable dépendante et une ou plusieurs variables indépendantes. En résumé, elle permet de prédire la valeur d'une variable en fonction des valeurs d'autres variables.

1) Première étape :

Préparation des données et ajustement du modèle : Dans cette étape, nous chargeons les données nécessaires et ajustons un modèle de régression linéaire à l'aide de la fonction `lm()`. Nous utilisons toutes les variables disponibles dans notre modèle initial, appelé `modelef`.

2) 2ème étape:

Évaluation du modèle : Nous évaluons ensuite la performance du modèle complet et modèle incomplet en examinant les coefficients de régression, l'intervalle de confiance et les statistiques de résumé.

```
> modelef$coefficients
      (Intercept)      NEBac+1      NEBac+2      NEBac+3      NEBac+4      NEBac+5
-3.954486e-01    -1.143132e+00    -1.882905e-01    8.307748e-03    -2.095923e-01    -2.899546e-01
Nemoins que Bac NEPlus que Bac+5      SE      SF      Age      HTS
-4.009138e-01    -1.693140e-01    -1.154266e-02    1.361114e-02    9.431166e-03    -4.957110e-04
      SG      TF      AEA      ST      RM      NEGS
2.574649e-02    -1.516828e-02    8.177911e-03    8.398423e-02    -5.337834e-06    1.896459e-02
      FP      EP      ARP      enfants      SN      IPE
8.294866e-02    5.977149e-02    -3.664973e-02    -3.962626e-02    4.122529e-02    1.736378e-01
      CMJ      PRFE      ARS      Conf      HMS      FOAP
5.241993e-03    -1.324068e+00    1.696406e-02    2.172442e-03    -1.127695e-02    -1.672359e-01
```

Les coefficients de régression : Dans un modèle de régression linéaire, chaque variable indépendante est associée à un coefficient de régression. Ce coefficient représente l'effet de cette variable sur la variable dépendante, en supposant que toutes les autres variables restent constantes. Par exemple, si le coefficient pour l'âge est 0.5, cela signifie que pour chaque augmentation d'un an, la variable dépendante augmentera de 0.5, toutes choses étant égales par ailleurs.

> confint(modeléf)

	2.5 %	97.5 %
(Intercept)	-2.253566e+00	1.462669e+00
NEBac+1	-2.499063e+00	2.127995e-01
NEBac+2	-1.336210e+00	9.596291e-01
NEBac+3	-1.089079e+00	1.105694e+00
NEBac+4	-1.534711e+00	1.115527e+00
NEBac+5	-1.499178e+00	9.192687e-01
NEMoins que Bac	-1.764563e+00	9.627349e-01
NEPlus que Bac+5	-1.294199e+00	9.555715e-01
SE	-4.130809e-01	3.899956e-01
SF	-2.767100e-01	3.039323e-01
Age	-2.182587e-02	4.068820e-02
HTS	-1.241696e-02	1.142554e-02
SG	-6.416282e-02	1.156558e-01
TF	-3.786136e-01	3.482771e-01
AEA	-3.757920e-02	5.393502e-02
ST	-7.259885e-03	1.752283e-01
RM	-4.303971e-05	3.236404e-05
NEGS	-1.518663e-01	1.897955e-01
FP	1.742117e-02	1.484762e-01
EP	-2.151555e-01	3.346985e-01
ARP	-1.064665e-01	3.316709e-02
enfants	-2.104996e-01	1.312471e-01
SN	-9.883811e-02	1.812887e-01
IPE	-4.979040e-02	3.970661e-01
CMJ	-1.312131e-02	2.360529e-02
PRFE	-2.268352e+00	-3.797828e-01
ARS	-1.337110e-01	1.676391e-01
Conf	-2.273082e-02	2.707571e-02
HMS	-3.592232e-02	1.336842e-02
FOAP	-3.457389e-01	1.126706e-02

Attributes in the equation

Attribute	Coef.	Std-dev	Wald	Signif
constant	-121903173.236334	10000.0000	148603836.4509	0.0000
SE	-111391813.475924	10000.0000	124081361.0946	0.0000
SF	-35550956.907530	10000.0000	12638705.3704	0.0000
Age	1463547.928624	10000.0000	21419.7254	0.0000
HTS	-2943867.966373	10000.0000	86663.5860	0.0000
SG	12549953.421632	10000.0000	1575013.3089	0.0000
TF	26062381.279082	10000.0000	6792477.1794	0.0000
AEA	2434931.017152	10000.0000	59288.8906	0.0000
ST	-1301341.124590	10000.0000	16934.8872	0.0000
RM	3277.955727	10000.0000	0.1074	0.7431
NEGS	-7519182.496389	10000.0000	565381.0541	0.0000
FP	556918.822031	10000.0000	3101.5857	0.0000
EP	40803851.225241	10000.0000	16649542.7481	0.0000
ARP	4432869.052087	10000.0000	196503.2803	0.0000

L'intervalle de confiance : L'intervalle de confiance est une estimation de la plage dans

laquelle la vraie valeur du coefficient de

> summary(modeléf)

```
Call:
lm(formula = DISC ~ NE + SE + SF + Age + HTS + SG + TF + AEA +
    ST + RM + NEGS + FP + EP + ARP + enfants + SN + IPE + CMJ +
    PRFE + ARS + Conf + HMS + FOAP, data = dfemme)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.60728	-0.18065	-0.00538	0.18541	0.62708

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-3.954e-01	9.022e-01	-0.438	0.66492
NEBac+1	-1.143e+00	6.584e-01	-1.736	0.09482
NEBac+2	-1.883e-01	5.574e-01	-0.338	0.73832
NEBac+3	8.308e-03	5.328e-01	0.016	0.98768
NEBac+4	-2.096e-01	6.434e-01	-0.326	0.74732
NEBac+5	-2.900e-01	5.871e-01	-0.494	0.62572
NEMoins que Bac	-4.009e-01	6.621e-01	-0.606	0.55030
NEPlus que Bac+5	-1.693e-01	5.462e-01	-0.310	0.75913
SE	-1.154e-02	1.950e-01	-0.059	0.95326
SF	1.361e-02	1.410e-01	0.097	0.92385
Age	9.431e-03	1.518e-02	0.621	0.53995
HTS	-4.957e-04	5.788e-03	-0.086	0.93243
SG	2.575e-02	4.366e-02	0.590	0.56064
TF	-1.517e-02	1.765e-01	-0.086	0.93219
AEA	8.178e-03	2.222e-02	0.368	0.71590
ST	8.398e-02	4.430e-02	1.896	0.06962
RM	-5.338e-06	1.831e-05	-0.292	0.77301
NEGS	1.896e-02	8.295e-02	0.229	0.82101

NEBac+5	-2.900e-01	5.871e-01	-0.494	0.62572
NEMoins que Bac	-4.009e-01	6.621e-01	-0.606	0.55030
NEPlus que Bac+5	-1.693e-01	5.462e-01	-0.310	0.75913
SE	-1.154e-02	1.950e-01	-0.059	0.95326
SF	1.361e-02	1.410e-01	0.097	0.92385
Age	9.431e-03	1.518e-02	0.621	0.53995
HTS	-4.957e-04	5.788e-03	-0.086	0.93243
SG	2.575e-02	4.366e-02	0.590	0.56064
TF	-1.517e-02	1.765e-01	-0.086	0.93219
AEA	8.178e-03	2.222e-02	0.368	0.71590
ST	8.398e-02	4.430e-02	1.896	0.06962
RM	-5.338e-06	1.831e-05	-0.292	0.77301
NEGS	1.896e-02	8.295e-02	0.229	0.82101
FP	8.295e-02	3.182e-02	2.607	0.01518 *
EP	5.977e-02	1.335e-01	0.448	0.65818
ARP	-3.665e-02	3.390e-02	-1.081	0.28996
enfants	-3.963e-02	8.297e-02	-0.478	0.63707
SN	4.123e-02	6.801e-02	0.606	0.54985
IPE	1.736e-01	1.085e-01	1.601	0.12203
CMJ	5.242e-03	8.916e-03	0.588	0.56186
PRFE	-1.324e+00	4.585e-01	-2.888	0.00789 **
ARS	1.696e-02	7.316e-02	0.232	0.81852
Conf	2.172e-03	1.209e-02	0.180	0.85886
HMS	-1.128e-02	1.197e-02	-0.942	0.35502
FOAP	-1.672e-01	8.667e-02	-1.930	0.06509 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4108 on 25 degrees of freedom
Multiple R-squared: 0.6685, Adjusted R-squared: 0.2841
F-statistic: 1.739 on 29 and 25 DF, p-value: 0.08191

régression se situe. Il est généralement calculé à un niveau de confiance de 95%, ce qui signifie que si l'expérience était répétée de nombreuses fois, 95% des intervalles de confiance calculés contiendraient la vraie valeur du coefficient. Si l'intervalle de confiance pour un coefficient ne comprend pas zéro, cela signifie que l'effet de cette variable sur la variable dépendante est significatif à ce niveau de confiance.

Les statistiques de résumé : Les statistiques de résumé fournissent des informations sur la qualité de l'ajustement du modèle de régression. Elles comprennent le R carré, qui est une mesure de la proportion de la variance de la variable dépendante qui est expliquée par les variables indépendantes. Elles comprennent également la statistique F et la p-value associée, qui sont utilisées pour tester l'hypothèse que tous les coefficients de régression sont nuls. Si la p-value est inférieure à 0.05 (alpha) => on rejette H0 ,ce qui signifie que le modèle de régression est significatif.

⇒ Dans ce cas on constate que tous les paramètres sont nuls sauf : PRFE,FP

Score Attribute	Score_1	
Target size (%)	Score	TP-Rate
0	1.0000	0.0000
5	1.0000	0.0571
10	1.0000	0.1143
15	1.0000	0.2000
20	1.0000	0.2571
25	1.0000	0.2857
30	1.0000	0.3714
35	1.0000	0.4286
40	1.0000	0.4857
45	1.0000	0.5143
50	1.0000	0.5143
55	0.0000	0.5714
60	0.0000	0.6286
65	0.0000	0.6857
70	0.0000	0.6857
75	0.0000	0.7143
80	0.0000	0.8000
85	0.0000	0.8000
90	0.0000	0.8286
95	0.0000	0.9143
100	0.0000	1.0000

3) 3ème étape :

Simplification du modèle : nous simplifions ensuite notre modèle en n'utilisant que les variables les plus significatives. Nous ajustons un nouveau modèle, `modelef2`, en utilisant seulement trois variables : Age, HTS et SG.

```
> modelef2$coefficients
(Intercept)      Age      HTS      SG
1.090321915 -0.006663577 -0.009301485  0.015863680
```

```
modelef2)
      2.5 %      97.5 %
0.43237482  1.7482690132
-0.01945073  0.0061235732
-0.01797331 -0.0006296613
-0.05228987  0.0840172294
```

```
> summary(model2)

Call:
lm(formula = DISC ~ Age + HTS + SG, data = dfemme)

Residuals:
    Min       1Q   Median       3Q      Max
-0.9296 -0.5075  0.1396  0.4107  0.5807

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.090322   0.327731   3.327  0.00163 **
Age          -0.006664   0.006369  -1.046  0.30041
HTS          -0.009301   0.004320  -2.153  0.03604 *
SG           0.015864   0.033948   0.467  0.64228
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4739 on 51 degrees of freedom
Multiple R-squared:  0.1003,    Adjusted R-squared:  0.04733
F-statistic: 1.894 on 3 and 51 DF,  p-value: 0.1423
```

➔ d'après le calcul des statistiques de résumé du model2 (incomplet) => on constate que tous les paramètres sont nuls sauf : HTS

4) Quatrième étape :

Comparaison des modèles : Nous comparons ensuite les performances de nos deux modèles en utilisant l'analyse de la variance (ANOVA) et les critères d'information d'Akaike (AIC) et de Bayes (BIC).

```
> anova(model1, model2)
Analysis of Variance Table

Model 1: DISC ~ NE + SE + SF + Age + HTS + SG + TF + AEA + ST + RM + NEGS +
FP + EP + ARP + enfants + SN + IPE + CMJ + PRFE + ARS + Conf +
HMS + FOAP
Model 2: DISC ~ Age + HTS + SG
Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      25  4.2185      -7.2328 1.6486 0.1077
2      51 11.4513

> anova(model2, model1)
Analysis of Variance Table

Model 1: DISC ~ Age + HTS + SG
Model 2: DISC ~ NE + SE + SF + Age + HTS + SG + TF + AEA + ST + RM + NEGS +
FP + EP + ARP + enfants + SN + IPE + CMJ + PRFE + ARS + Conf +
HMS + FOAP
Res.Df    RSS Df Sum of Sq    F Pr(>F)
1      51 11.4513
2      25  4.2185  26    7.2328 1.6486 0.1077
```

d'après l'utilisation la variance ANOVA pour comparer les deux modèles nous avons trouves que leur p-value sont égaux donc on les recomparent en utilisant les critères AIC et BIC

```
> AIC(model1)
[1] 76.851
> AIC(model2)
[1] 79.77558
```

la valeur de AIC(model1) est plus faible que la valeur de AIC(model2) , donc le modèle 'model1' complet est le meilleur modèle.

5) Cinquième étape :

Sélection du modèle final : Enfin, nous utilisons la méthode de sélection de modèle pas à pas pour déterminer le modèle final le plus approprié.

```
m0=lm(DISC~ 1,data=dfemme)
mc=modelAIC
B=stepAIC(mc,direction="backward")
B$coefficients
summary(B)
A=stepAIC(m0,scope=list(lower=m0, upper=mc),direction="forward")
A$coefficients
summary(A)
```

ces démarches sont utilisées pour trouver le modèle le plus simple qui prédit encore bien la variable dépendante. Cela nous aide à éviter le surajustement, où un modèle complexe s'adapte très bien aux données d'entraînement mais ne généralise pas bien aux nouvelles données.

Analyse en composantes principales

(ACP)

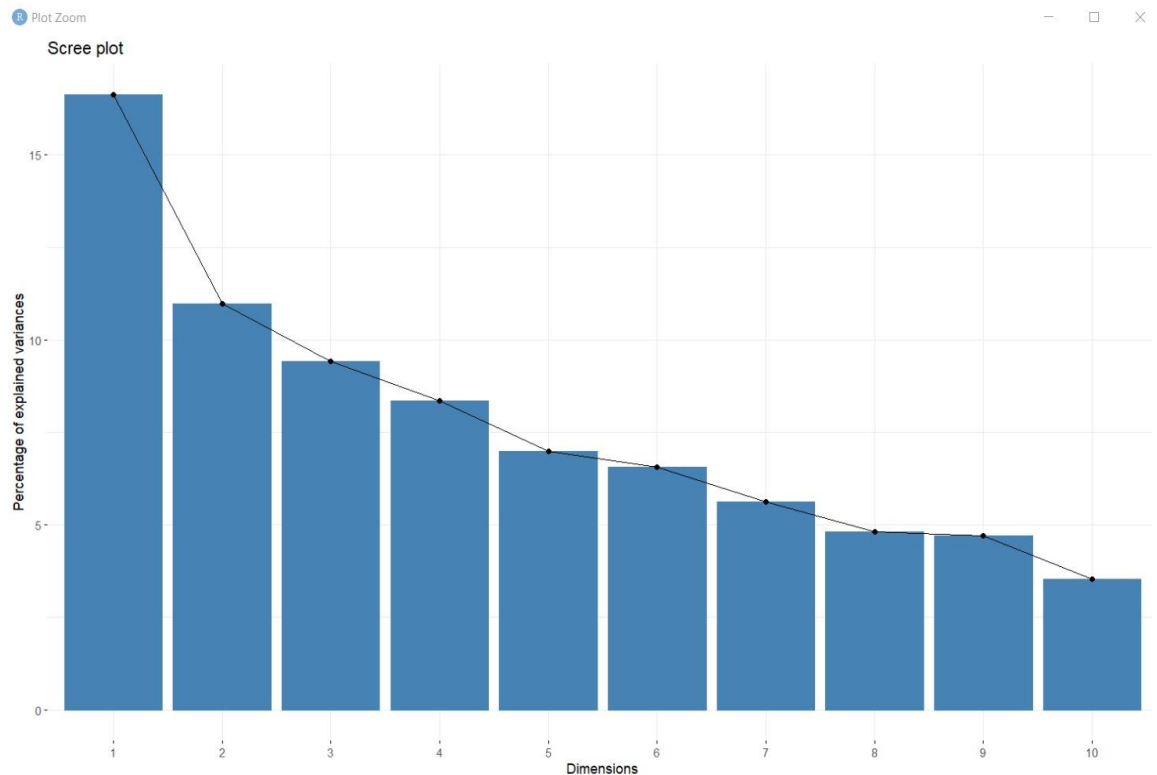
L'Analyse en Composantes Principales (ACP) est une méthode statistique pour réduire la dimensionnalité des données tout en préservant leur structure essentielle. En résumé, elle permet de visualiser et de comprendre les relations entre les variables en les projetant dans un espace de dimensions réduites appelées composantes principales.

1) Première Etape :

- ❖ **le critère de kaiser :** En choisissant de conserver neuf axes principaux, nous nous concentrons sur les aspects les plus importants de nos données, simplifiant ainsi notre analyse tout en préservant les principales tendances et variations.

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	3.99197044	16.6332102	16.63321
comp 2	2.63662313	10.9859297	27.61914
comp 3	2.26253088	9.4272120	37.04635
comp 4	2.00441384	8.3517243	45.39808
comp 5	1.67731450	6.9888104	52.38689
comp 6	1.57617297	6.5673874	58.95427
comp 7	1.34871349	5.6196396	64.57391
comp 8	1.15483436	4.8118099	69.38572
comp 9	1.12862951	4.7026229	74.08835
comp 10	0.84749125	3.5312135	77.61956
comp 11	0.75167166	3.1319652	80.75153
comp 12	0.67854769	2.8272820	83.57881
comp 13	0.61646831	2.5686179	86.14743
comp 14	0.59200982	2.4667076	88.61413
comp 15	0.47516761	1.9798651	90.59400
comp 16	0.44763619	1.8651508	92.45915
comp 17	0.41338603	1.7224418	94.18159
comp 18	0.33111939	1.3796641	95.56125
comp 19	0.30464883	1.2693701	96.83062
comp 20	0.25899049	1.0791270	97.90975
comp 21	0.18669287	0.7778869	98.68764
comp 22	0.16974509	0.7072712	99.39491
comp 23	0.10562403	0.4401001	99.83501
comp 24	0.03959763	0.1649901	100.00000

- ❖ **critère du coude :** nous constatons un changement notable dans la pente de la courbe de l'inertie donc on va admettre 9 axes.



=> Nous pouvons donc admettre qu'à ce point, d'après le deux critère qu'on va admettre 9 axes.

2) 2ème Etape :

2.1) 1^{er} Axe :

❖ Pour les individus : On à 55 individus donc le seuil est : 1,81%

> sort(t1,decreasing = TRUE)

Femme 20	Femme 40	Femme 37	Femme 46	Femme 41	Femme 16	Femme 32
1.263885e+01	1.091292e+01	6.413442e+00	5.246733e+00	4.278958e+00	4.121700e+00	4.027513e+00
Femme 21	Femme 3	Femme 48	Femme 33	Femme 29	Femme 35	Femme 19
3.482820e+00	2.885371e+00	2.793297e+00	2.692823e+00	2.658928e+00	2.449019e+00	2.333446e+00
Femme 7	Femme 43	Femme 42	Femme 49	Femme 28	Femme 18	Femme 10
2.254142e+00	1.972275e+00	1.970828e+00	1.736532e+00	1.713146e+00	1.628821e+00	1.597114e+00
Femme 50	Femme 39	Femme 34	Femme 13	Femme 36	Femme 24	Femme 51
1.573177e+00	1.570657e+00	1.520992e+00	1.401977e+00	1.364382e+00	1.270909e+00	1.227685e+00
Femme 11	Femme 54	Femme 1	Femme 22	Femme 6	Femme 27	Femme 2
1.217200e+00	1.140157e+00	1.083503e+00	9.932175e-01	9.106432e-01	7.692120e-01	7.419114e-01
Femme 12	Femme 17	Femme 4	Femme 53	Femme 55	Femme 25	Femme 52
6.390940e-01	4.524606e-01	4.449594e-01	4.321580e-01	2.535763e-01	2.322969e-01	1.477724e-01
Femme 5	Femme 9	Femme 31	Femme 47	Femme 30	Femme 44	Femme 45
1.363985e-01	1.242084e-01	1.221454e-01	9.807766e-02	9.277520e-02	6.455776e-02	5.625739e-02
Femme 15	Femme 23	Femme 14	Femme 38	Femme 26	Femme 8	
3.435756e-02	3.305411e-02	2.242568e-02	1.648509e-02	2.561472e-03	8.413220e-05	

Donc on constate que tous les contributions qui sont supérieur a 1,81% :

- Femme 20 : 12.63885
- Femme 40 : 10.91292
- Femme 37 : 6.413442
- Femme 46 : 5.246733
- Femme 41 : 4.278958
- Femme 16 : 4.1217
- Femme 32 : 4.027513
- Femme 21 : 3.48282
- Femme 3 : 2.885371
- Femme 48 : 2.793297
- Femme 33 : 2.692823
- Femme 29 : 2.658928
- Femme 35 : 2.449019
- Femme 19 : 2.333446
- Femme 7 : 2.254142
- Femme 43 : 1.972275
- Femme 42 : 1.970828

> indiv\$coord[,1]

Femme 1	Femme 2	Femme 3	Femme 4	Femme 5	Femme 6	Femme 7
-1.54237499	-1.27629484	2.51695740	-0.98840560	0.54724243	1.41399907	-2.22467038
Femme 8	Femme 9	Femme 10	Femme 11	Femme 12	Femme 13	Femme 14
-0.01359115	0.52221628	-1.87259119	1.63476726	-1.18456080	-1.75446808	0.22189515
Femme 15	Femme 16	Femme 17	Femme 18	Femme 19	Femme 20	Femme 21
0.27465416	3.00824490	-0.99670211	-1.89108767	2.26346527	5.26779313	-2.76528885
Femme 22	Femme 23	Femme 24	Femme 25	Femme 26	Femme 27	Femme 28
1.47671673	-0.26939390	-1.67044544	0.71416196	0.07499285	-1.29956505	-1.93942162
Femme 29	Femme 30	Femme 31	Femme 32	Femme 33	Femme 34	Femme 35
-2.41617449	0.45132663	0.51786143	-2.97367495	-2.43152583	-1.82742068	-2.31884149
Femme 36	Femme 37	Femme 38	Femme 39	Femme 40	Femme 41	Femme 42
-1.73078439	3.75249912	-0.19024825	-1.85701625	4.89491793	3.06509557	2.08017245
Femme 43	Femme 44	Femme 45	Femme 46	Femme 47	Femme 48	Femme 49
-2.08093601	0.37648633	-0.35145102	3.39405967	0.46404495	-2.47647281	1.95261427
Femme 50	Femme 51	Femme 52	Femme 53	Femme 54	Femme 55	
1.85850510	-1.64179359	-0.56960220	0.97408371	1.58218537	-0.74615550	

❖ Pour les variables : On a 24 variables donc : $1/24=4.16$

> vare\$contrib[,1]

NE	SE	SF	Age	HTS	SG	TF
1.133108333	3.000309770	6.495825996	16.169973368	0.078634372	2.570189023	3.520120685
AEA	ST	DISC	RM	NEGS	FP	EP
17.137070084	0.653423314	1.134948255	2.643835000	0.273773626	0.014010161	0.524668996
ARP	enfants	SN	IPE	CMJ	PRFE	ARS
0.838731226	15.011980089	0.960829213	3.408485072	2.057095789	12.514531128	0.039726999
Conf	HMS	FOAP				
3.676927153	6.139047969	0.002754379				

Selon la contribution les variables dont les contributions sur le premier axe de l'ACP sont supérieures à 4,16 sont considérées comme ayant une influence significative sur la variation observée le long de cet axe. Voici les variables correspondantes :

- Âge (Age) : 16,17
- Satisfaction globale (SG) : 6,49
- Années d'emploi actuel (AEA) : 17,14
- Impact politique sur l'emploi (IPE) : 4,41
- Pourcentage de revenu consacré aux frais de garde d'enfants (PRFE) : 12,51
- Conflits entre responsabilités professionnelles et familiales par mois (Conf) : 3,68
- Heures des tâches ménagères par semaine (HMS) : 6,14

Selon les coordonnées

```
> vare$coord[,1]
```

NE	SE	SF	Age	HTS	SG	TF
-0.21268133	-0.34608016	0.50922633	0.80343049	-0.05602732	0.32031420	0.37486288
AEA	ST	DISC	RM	NEGS	FP	EP
0.82710747	0.16150686	-0.21285394	0.32487092	0.10454168	0.02364913	0.14472260
ARP	enfants	SN	IPE	CMJ	PRFE	ARS
0.18298061	0.77412777	-0.19584693	0.36887087	0.28656353	0.70680718	0.03982324
Conf	HMS	FOAP				
0.38312119	0.49504442	0.01048589				

=> Interprétation : Dans notre analyse, nous constatons que certaines femmes, comme Femme 3, sont fortement influencées par des variables telles que la satisfaction envers la famille, l'âge avancé et les aspects économiques et administratifs. Par exemple, Femme 3 pourrait être une personne âgée qui accorde une grande importance à sa famille et qui est financièrement stable. En revanche, d'autres femmes, comme Femme 16, semblent être influencées par des caractéristiques opposées, telles qu'une satisfaction moindre envers la famille ou des aspects économiques moins importants. Par exemple, Femme 16 pourrait être une jeune professionnelle concentrée sur sa carrière et moins préoccupée par sa situation familiale. Ces exemples illustrent comment les variables influencent différemment les individus dans notre analyse.

2.2) 1^{er} Axe :

❖ **Pour les individus :** On à 55 individus dans le seuil est : 1,81%

```
> sort(t2,decreasing = TRUE)
Femme 54      Femme 11      Femme 22      Femme 55      Femme 49      Femme 46      Femme 24
19.229477142  7.130925827  7.072552539  6.667277671  5.698746850  5.336602979  5.265715516
Femme 41      Femme 43      Femme 14      Femme 51      Femme 48      Femme 44      Femme 4
4.441866014  4.189656875  3.152475139  2.830368274  2.750391439  2.666473073  2.648220250
Femme 42      Femme 38      Femme 52      Femme 31      Femme 6       Femme 17      Femme 2
2.604336616  1.573383678  1.486477208  1.363243140  1.235530492  1.189533712  1.103688614
Femme 35      Femme 28      Femme 18      Femme 34      Femme 19      Femme 23      Femme 15
1.059355304  1.045798826  0.930012370  0.795427388  0.614076802  0.608322772  0.582163302
Femme 29      Femme 21      Femme 8       Femme 20      Femme 39      Femme 32      Femme 40
0.562324321  0.440230092  0.402584376  0.388131932  0.365190815  0.338079821  0.310560600
Femme 16      Femme 27      Femme 26      Femme 30      Femme 47      Femme 37      Femme 36
0.226796491  0.214851280  0.195138107  0.185908083  0.183702274  0.169271971  0.150690154
Femme 33      Femme 1       Femme 25      Femme 12      Femme 7       Femme 53      Femme 3
0.142203762  0.111948251  0.062718385  0.059537894  0.057836231  0.049027530  0.035183459
Femme 45      Femme 13      Femme 50      Femme 10      Femme 5       Femme 9
0.030978380  0.020171902  0.009901625  0.007280526  0.005292368  0.002359560
```

Donc on constate que tous les contributions qui sont supérieur a 1,81% :

- Femme 54 : 19,229477142
- Femme 11 : 7,130925827
- Femme 22 : 7,072552539
- Femme 55 : 6,667277671
- Femme 49 : 5,698746850
- Femme 46 : 5,336602979
- Femme 24 : 5,265715516
- Femme 41 : 4,441866014
- Femme 43 : 4,189656875
- Femme 14 : 3,152475139
- Femme 51 : 2,830368274
- Femme 48 : 2,750391439
- Femme 44 : 2,666473073
- Femme 4 : 2,648220250
- Femme 42 : 2,604336616
- Femme 38 : 1,573383678
- Femme 52 : 1,486477208
- Femme 31 : 1,363243140
- Femme 6 : 1,235530492


```
> indiv$coord[,2]
Femme 1      Femme 2      Femme 3      Femme 4      Femme 5      Femme 6      Femme 7
0.40291555 -1.26511107  0.22587837 -1.95966765 -0.08760530 -1.33854232  0.28960454
Femme 8      Femme 9      Femme 10     Femme 11     Femme 12     Femme 13     Femme 14
0.76407120  0.05849529 -0.10275116  3.21572079  0.29383404  0.17103256 -2.13811573
Femme 15     Femme 16     Femme 17     Femme 18     Femme 19     Femme 20     Femme 21
-0.91881438 -0.57348695  1.31339014 -1.16131420  0.94366255 -0.75023109  0.79899716
Femme 22     Femme 23     Femme 24     Femme 25     Femme 26     Femme 27     Femme 28
3.20253190  0.93923098 -2.76333838 -0.30158019 -0.53195686  0.55818010  1.23148591
Femme 29     Femme 30     Femme 31     Femme 32     Femme 33     Femme 34     Femme 35
-0.90302299 -0.51922370  1.40602173 -0.70018854  0.45410984  1.07400337  1.23944196
Femme 36     Femme 37     Femme 38     Femme 39     Femme 40     Femme 41     Femme 42
0.46746361  0.49544779  1.51050683  0.72772165  0.67108658 -2.53797944  1.94336301
Femme 43     Femme 44     Femme 45     Femme 46     Femme 47     Femme 48     Femme 49
-2.46487331  1.96640955  0.21195064  2.78187634  0.51613420  1.99711295 -2.87471673
Femme 50     Femme 51     Femme 52     Femme 53     Femme 54     Femme 55
0.11982808  2.02594125 -1.46819757 -0.26664005 -5.28067101 -3.10942184
```

❖ Pour les variables : On a 24 variables donc : $1/24=4.16$

```
> vare$contrib[,2]
NE      SE      SF      Age      HTS      SG      TF      AEA
10.2881500 2.3931554 2.7594232 4.4623090 8.6478367 16.6369970 4.0543325 3.5853386
ST      DISC     RM      NEGS     FP      EP      ARP      enfants
6.8881705 2.2059309 9.4045749 1.3712707 0.3710116 3.4825420 0.7746201 0.7826991
SN      IPE      CMJ      PRFE     ARS     Conf     HMS     FOAP
2.2186949 1.4662113 0.1625247 0.2705411 6.2038408 9.4829210 1.4110405 0.6758637
```

Selon la contribution Les variables dont les contributions sur le premier axe de l'ACP sont supérieures à 4,16 sont considérées comme ayant une influence significative sur la variation observée le long de cet axe. Voici les variables correspondantes :

-NE : 10,2881500

-HTS : 8,6478367

-SG : 16,6369970

-ST : 6,8881705

-RM : 9,4045749

-ARS : 6,2038408

-Conf : 9,4829210

Selon les coordonnées

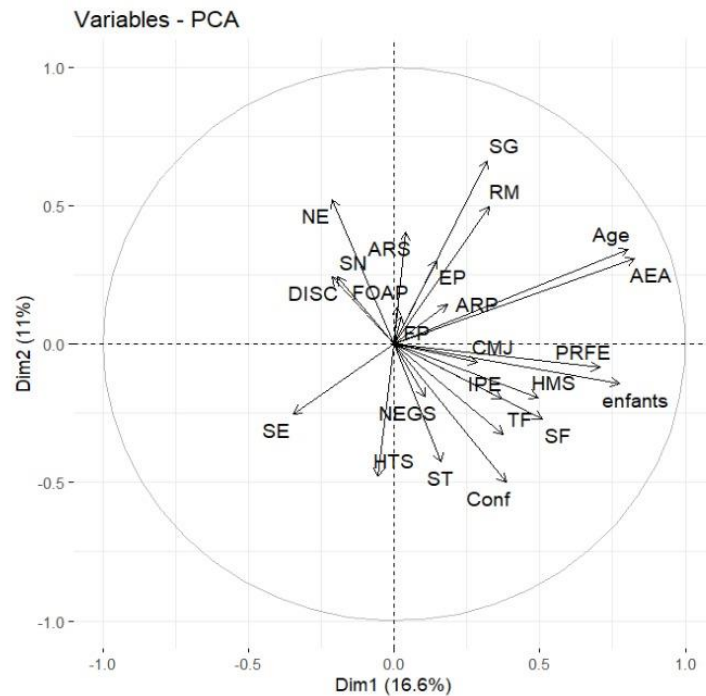
```
> vare$coord[,2]
      NE      SE      SF      Age      HTS      SG      TF
0.52082602 -0.25119413 -0.26973244 0.34300768 -0.47750483 0.66231028 -0.32695178
      AEA      ST      DISC      RM      NEGS      FP      EP
0.30746035 -0.42616323 0.24116816 0.49795903 -0.19014531 0.09890489 0.30302064
      ARP      enfants      SN      IPE      CMJ      PRFE      ARS
0.14291190 -0.14365523 0.24186488 -0.19661756 -0.06546116 -0.08445798 0.40444023
      Conf      HMS      FOAP
-0.50002889 -0.19288291 0.13349149
```

=>**Interprétation :** Parmi les individus dont les contributions sont élevées, on observe deux groupes distincts en fonction des variables qui les influencent le plus. Dans un groupe, des individus comme Femme 54, Femme 11 et Femme 22 sont fortement influencés par des facteurs tels que le secteur d'emploi, la satisfaction globale au travail et l'impact des politiques. Cela suggère qu'ils peuvent être particulièrement sensibles à leur environnement professionnel, leur satisfaction au travail et les politiques gouvernementales en matière d'emploi. Dans l'autre groupe, des individus comme Femme 46, Femme 24 et Femme 41 sont moins influencés par ces facteurs, mais peuvent être davantage touchés par des aspects tels que la flexibilité du travail, la discrimination au travail et le niveau de stress lié au travail.

3) 3eme et 4eme Etape :

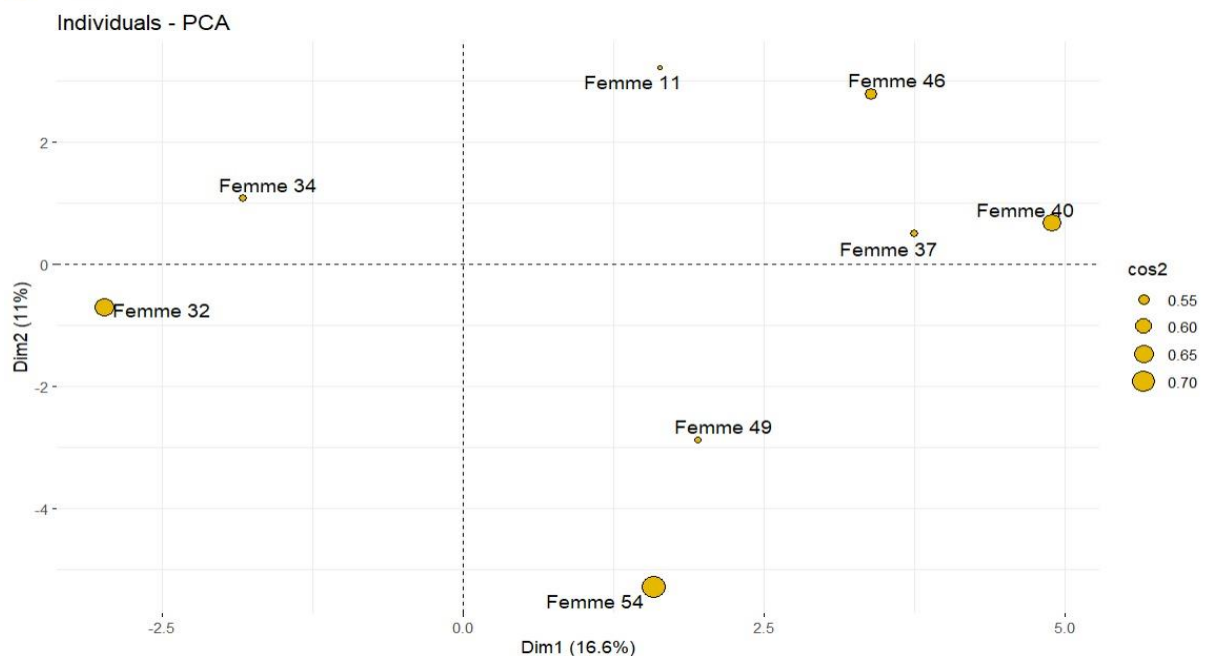
Pour les variables :

Parmi les variables qui dépassent le seuil de 0,5 en termes de qualité de représentation, nous identifions l'AEA, l'Age, les enfants, le SG, le PRFE, le Conf, le RM et le SF. Ces variables jouent un rôle crucial dans la caractérisation de la composante principale, ce qui signifie qu'elles contribuent de manière significative à la structure des données observées. Leur forte qualité de représentation indique qu'elles capturent efficacement la variation présente dans l'ensemble des données, fournissant ainsi des informations essentielles pour l'interprétation et la compréhension des tendances sous-jacentes.

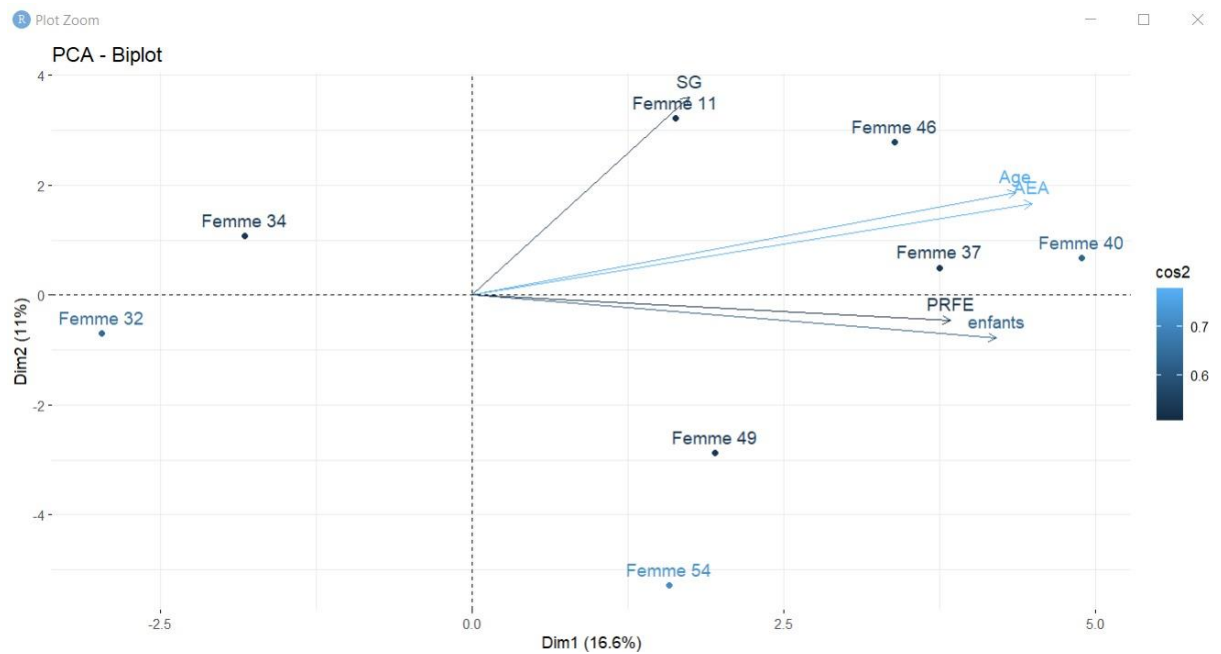


Pour les individus :

Parmi les individus dont la qualité de représentation dépasse le seuil de 0,5, nous n'observons que les femmes telles que Femme 54, Femme 32, Femme 40, Femme 46, Femme 34 et Femme 37 sont bien représentées dans la composante principale. Cela signifie que ces femmes contribuent de manière significative à la structure des données observées et capturent efficacement la variation présente dans l'ensemble des données.



4) Synthèse :



Lors de la cinquième étape de l'analyse, une conclusion importante a émergé : un seul groupe de femmes de 37, 40 et 46. Ce groupe est caractérisé par une influence significative de l'âge, du nombre d'années d'emploi actuel, du pourcentage de revenu consacré aux frais de garde d'enfants et du nombre d'enfants. Ces variables semblent jouer un rôle majeur dans la situation de ce groupe spécifique de femmes, suggérant qu'elles sont confrontées à des défis ou des dynamiques similaires, indépendamment de leur âge précis. Cette observation souligne l'importance de considérer non seulement l'âge, mais également d'autres facteurs tels que la situation professionnelle et les responsabilités familiales lors de l'analyse des données démographiques et socio-économiques.

Analyse factorielle des correspondances (AFC)

I. Introduction

L'Analyse Factorielle des Correspondances (AFC) est une méthode pour analyser des données catégorielles, en révélant les associations entre les variables et en les représentant dans un espace de dimensions réduites. Elle aide à comprendre les relations entre différentes catégories de variables. Dans votre projet sur les femmes travailleuses, l'AFC pourrait éclairer les caractéristiques influençant leur situation professionnelle.

	moins que Bac C	moins que Bac M	moins que Bac V	Bac C	Bac+1 V	Bac+1 C	Bac+2 M	Bac+2 C	Bac+3 C	Bac+3 M	Bac+3 D	Bac+4 M	Bac+4 V	Bac+5 M	Bac+5 C	Plus que Bac+5 M	Plus que Bac+5 C
HTS	48	37	54	14	10	52	26	38	39	35	20	39	44	38	28	33	35
TF	0	1	1	0	0	1	1	1	0	1	1	1	1	0	0	0	0
AEA	6	20	6	2	32	3	17	5	6	10	3	17	2	5	3	9	7
ST	9	9	8	7	5	8	8	8	6	7	9	7	8	6	5	8	8
DISC	0	1	0	1	0	0	1	1	1	1	1	0	1	0	1	1	1
SN	6	7	6	7	8	7	7	8	7	6	6	7	6	7	7	7	7
ENFANTS	0	4	4	0	3	0	3	0	0	1	2	2	3	1	0	1	0
CONF	5	11	20	9	2	21	2	8	4	9	18	6	10	3	4	8	3
PRFE	0	0,4	0,2	0	0,05	0	0,14	0	0	0,14	0,44	0,43	0,3	0,04	0	0,08	0

II. Etape 1 : Le test Khi 2 et lancement du AFC

Code :

```
chisq<-chisq.test(FT)
print(chisq)

res.AFC<-CA(FT, graph = FALSE)
summary(res.AFC)
```

Résultats :

```
> print(chisq)

Pearson's Chi-squared test

data: FT
X-squared = 479.62, df = 128, p-value < 2.2e-16
```

- ➔ Le test de χ^2 sert à connaître l'indépendance des variables.
- ➔ On a $\alpha = 0,05 > p\text{-value} = 2,2e-16$ Donc on rejette l'hypothèse 0 et les variables ne sont pas indépendantes.

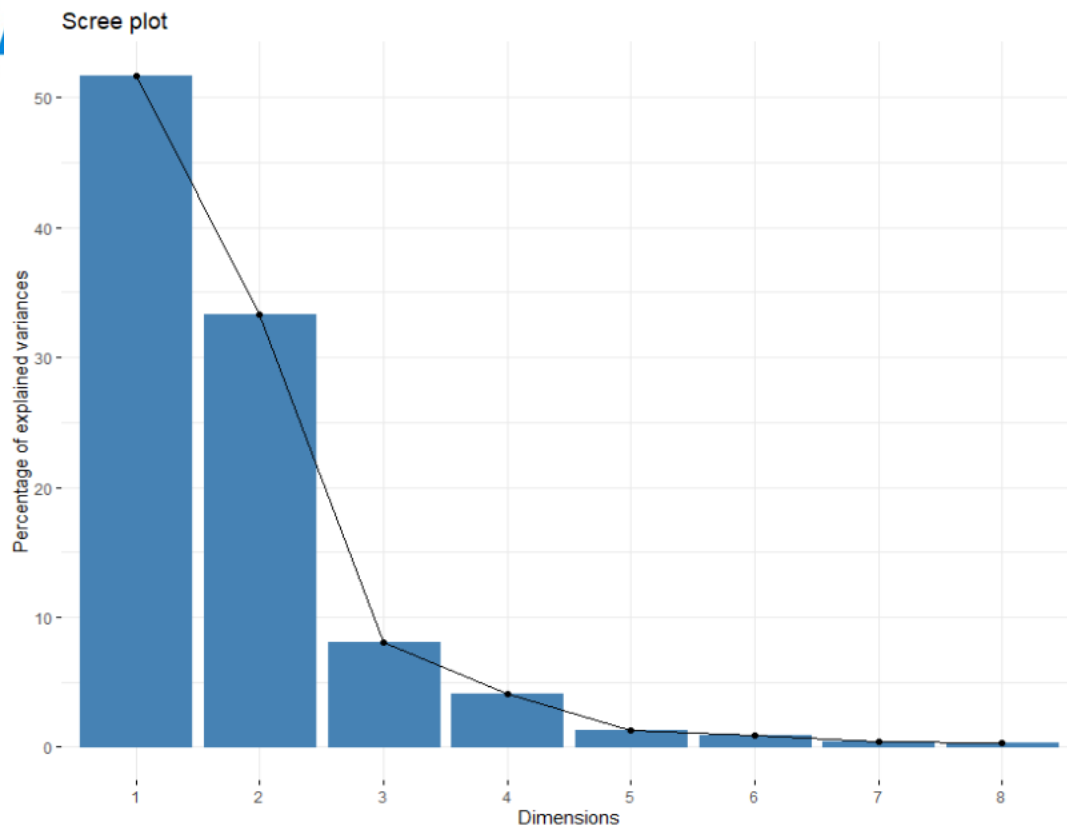
III. Etape 2 : Choix du nombre d'axes

Code :

```
#etape 2
#valeurs propres /Variances
eig.val <-get_eigenvalue(res.AFC)
eig.val
res.AFC$eig
fviz_eig(res.AFC)
```

Résultats :

```
> res.AFC$eig
      eigenvalue percentage of variance cumulative percentage of variance
dim 1 0.177876880           51.6996071           51.69961
dim 2 0.114458469           33.2671558           84.96676
dim 3 0.027730430            8.0598014           93.02656
dim 4 0.014174893            4.1199081           97.14647
dim 5 0.004272533            1.2418043           98.38828
dim 6 0.002955287            0.8589490           99.24723
dim 7 0.001426316            0.4145563           99.66178
dim 8 0.001163668            0.3382181          100.00000
```



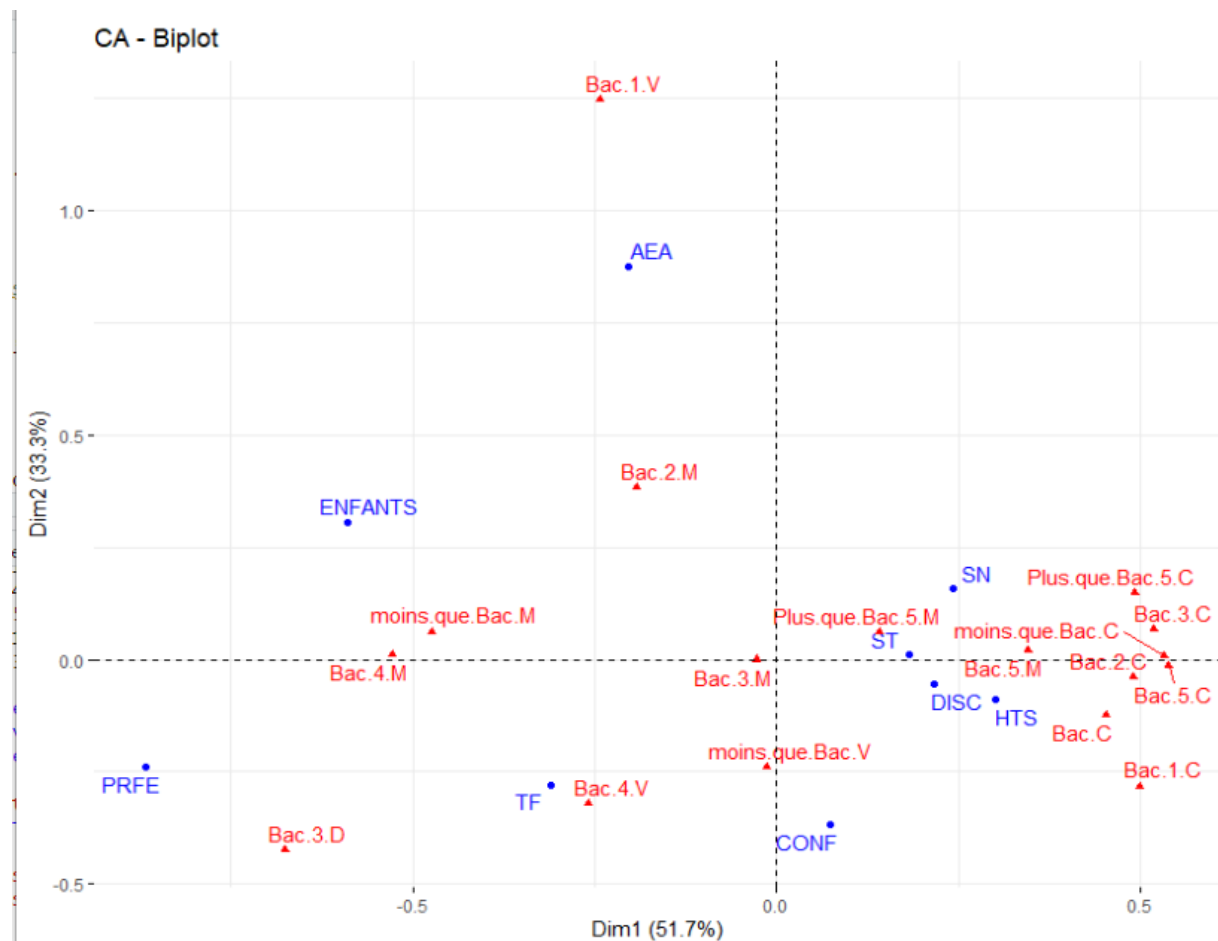
➔ D'après le cumule et le graphe, on va prendre 2 axes avec un taux d'inertie total de 86,96%.

IV. Etape 3 : Interprétation des axes, des contributions et des qualités

➔ On relance l'AFC sur les deux axes et on isole les lignes et les colonnes :

```
#etape 3 : interpretation des axes contributions et qualités
#repe1 = TRUE pour éviter le chevauchement de texte
res.AFC<-CA(FT, graph = TRUE, axes = c(1,2), ncp = 2)
fviz_ca_biplot(res.AFC, repe1 = TRUE)
ligne <- get_ca_row(res.AFC)
colo=get_ca_col(res.AFC)
```

Resultats :

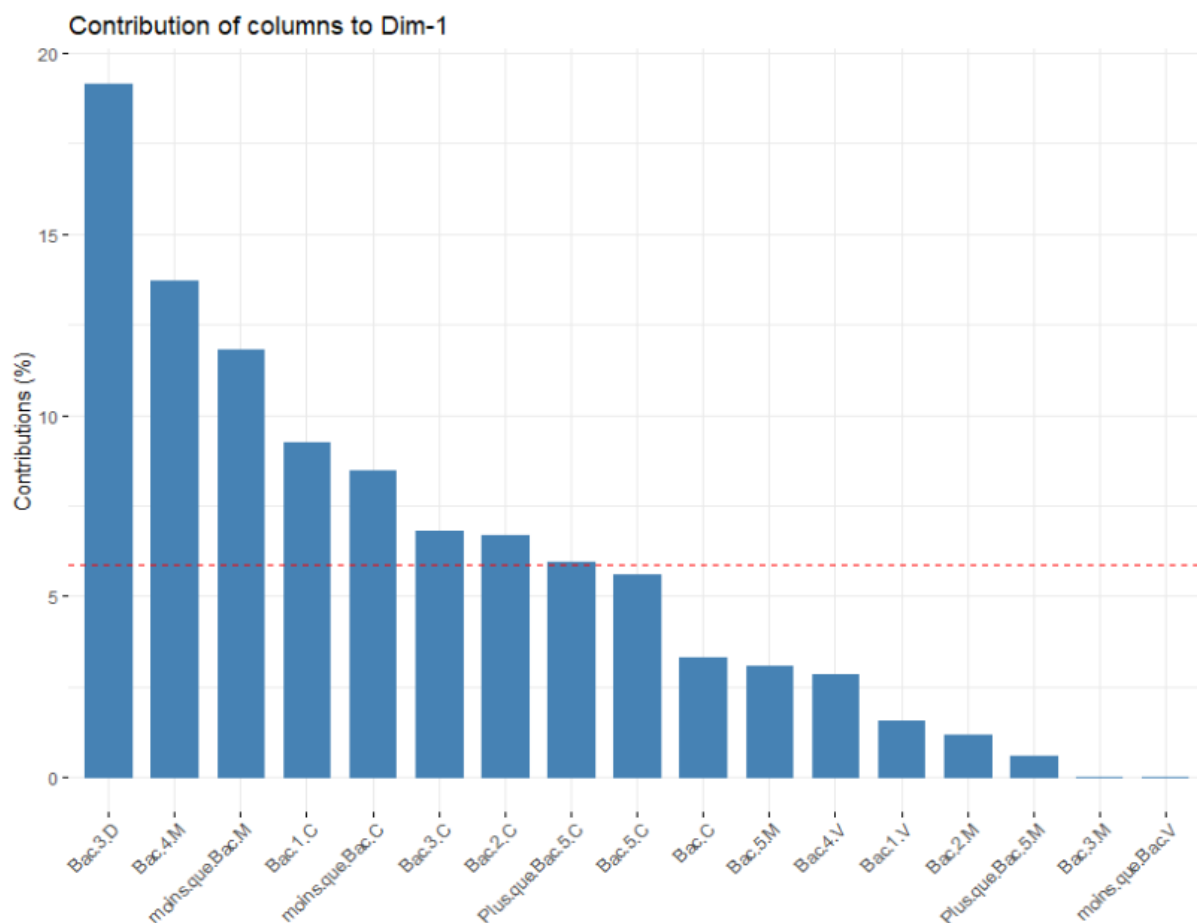


→ Puis on interprète l'AXE 1 :

- Pour les colonnes :

```
#contributions des colonnes à la dimension 1
fviz_contrib(res.AFC, choice="col", axes = 1, top = 17)
sort(colo$contrib[,1], decreasing = TRUE)
colo$cos2[,1]
colo$coord[,1]
```


Résultats :



→ On a pris une moyenne de 6,68

→ Les contributions qui sont supérieure à cette moyenne :

Bac+3D, Bac+4M, moins_que_BacM, Bac+1C, moins_que_BacC, Bac+3C et Bac+2C

→ Toutes ces variables ont une bonne qualité c-à-d supérieure à 55%

```
> sort(colos$contrib[,1], decreasing = TRUE)
Bac.3.D      Bac.4.M      moins.que.Bac.M      Bac.1.C      moins.que.Bac.C
19.14861071  13.71474529  11.79596088      9.24709085  8.46218066
Bac.3.C      Bac.2.C      Plus.que.Bac.5.C      Bac.5.C      Bac.C
6.81571783   6.68004539   5.94501339      5.60866857  3.30847491
Bac.5.M      Bac.4.V      Bac.1.V
3.07403497   2.83708917   1.54622767
Bac.3.M      moins.que.Bac.V
0.02646084   0.00860885

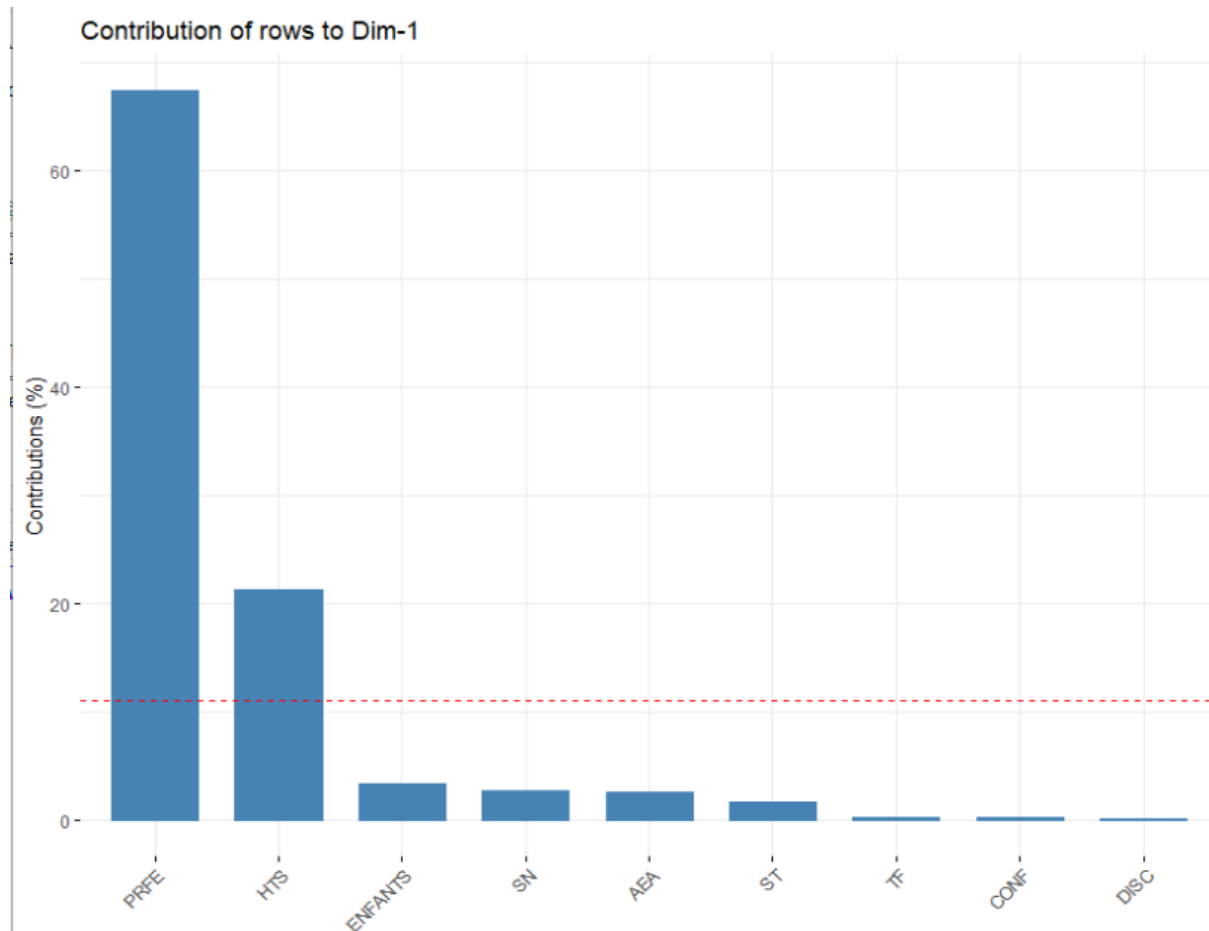
> colos$cos2[,1]
moins.que.Bac.C      moins.que.Bac.M      moins.que.Bac.V      Bac.C      Bac.1.V
0.831156281      0.974647195      0.001500802      0.346576584      0.035073297
Bac.1.C      Bac.2.M      Bac.2.C      Bac.3.C      Bac.3.M
0.561397615      0.169099067      0.906202223      0.870924240      0.065447542
Bac.3.D      Bac.4.M      Bac.4.V      Bac.5.M      Bac.5.C
0.670114905      0.860351197      0.342466645      0.627881671      0.857484052
Plus.que.Bac.5.M      Plus.que.Bac.5.C
0.583260358      0.794968636

> colos$coord[,1]
moins.que.Bac.C      moins.que.Bac.M      moins.que.Bac.V      Bac.C      Bac.1.V
0.53249609      -0.47433604      -0.01339338      0.45287158      -0.24286849
Bac.1.C      Bac.2.M      Bac.2.C      Bac.3.C      Bac.3.M
0.49922898      -0.19198290      0.48995512      0.51793668      -0.02794816
Bac.3.D      Bac.4.M      Bac.4.V      Bac.5.M      Bac.5.C
-0.67568317      -0.52796496      -0.25884095      0.34510842      0.53827029
Plus.que.Bac.5.M      Plus.que.Bac.5.C
0.14163853      0.49158955
```

- Pour les lignes :

```
#contributions des lignes à la dimension 1
fviz_contrib(res.AFC, choice="row", axes = 1, top = 9)
sort(ligne$contrib[,1], decreasing = TRUE)
ligne$cos2[,1]
ligne$coord[,1]
```

Résultats :



- ➔ On a pris une moyenne de 21,37
- ➔ Les contributions qui sont supérieures à cette moyenne :
PRFE et HTS
- ➔ Toutes ces variables ont une bonne qualité c-à-d supérieure à 55%

```

> #contributions des lignes à la dimension 1
> fviz_contrib(res.AFC, choice="row", axes = 1, top = 9)
> sort(ligne$contrib[,1], decreasing = TRUE)
      PRFE      HTS      ENFANTS      SN      AEA      ST      TF      CONF
67.3689422 21.3729583  3.3787871  2.7506101  2.5781466  1.6728135  0.3513420  0.3188539
      DISC
0.2075462
> ligne$cos2[,1]
      HTS      TF      AEA      ST      DISC      SN      ENFANTS      CONF
0.78027400 0.15956002 0.05098340 0.44621221 0.05758587 0.42253296 0.48971494 0.01663768
      PRFE
0.91863090
> ligne$coord[,1]
      HTS      TF      AEA      ST      DISC      SN      ENFANTS      CONF
0.2997074 -0.3111250 -0.2044086 0.1814386 0.2162978 0.2424804 -0.5908346 0.0743565
      PRFE
-0.8674501
> |

```

Résultats Finale :

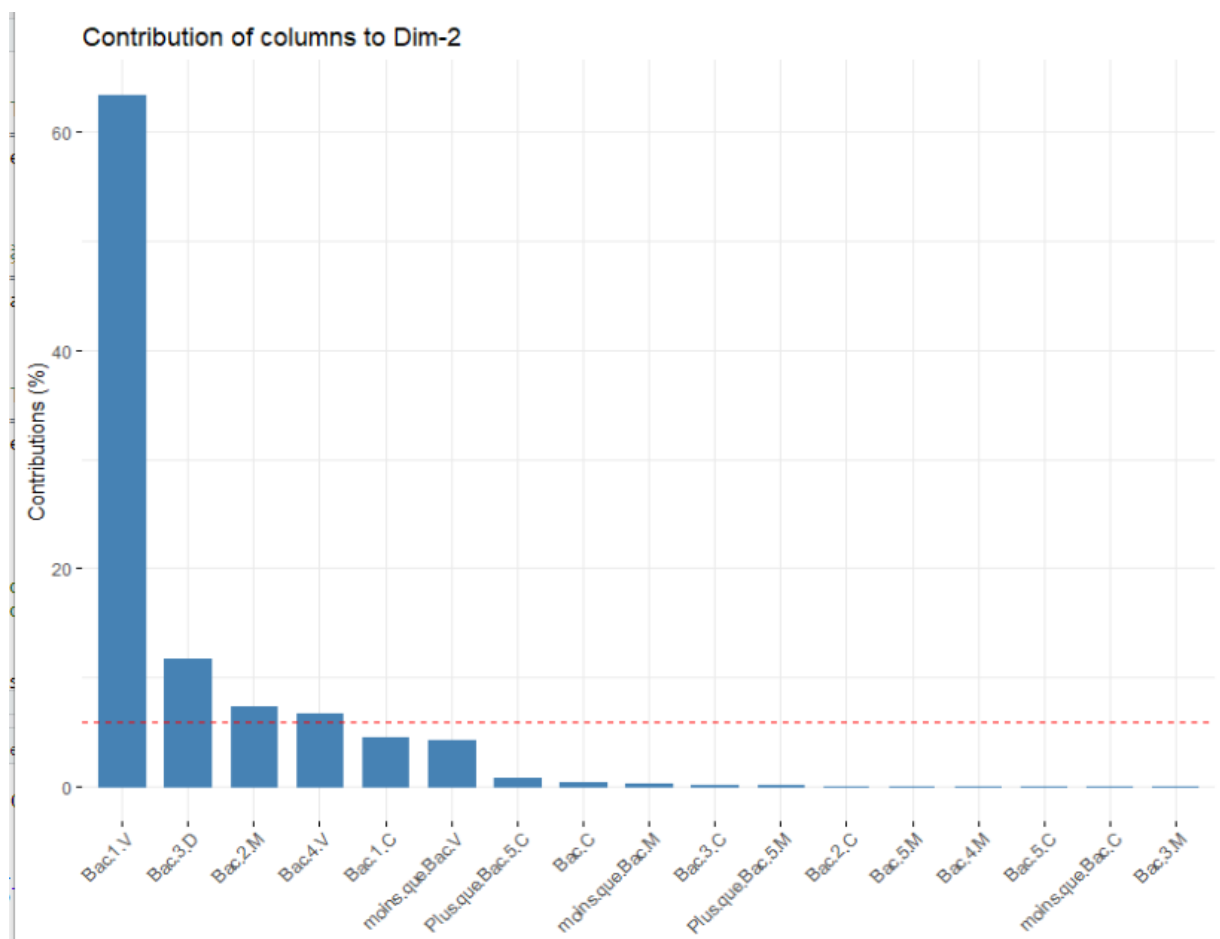
- ➔ Les femmes qui ont Bac+3 Divorcée, Bac+4 et moins que Bac Mariée sont ceux qui dépensent le plus grand pourcentage de leurs salaires pour la garde de leurs enfants.
- ➔ Les femmes qui ont Bac+1, moins que Bac, Bac+3 et Bac+2 tous célibataires sont ceux qui travaillent le plus d'heures par semaine.

→ Puis on interprète l'AXE 2 :

- Pour les colonnes :

```
#contributions des colonnes à la dimension 2
fviz_contrib(res.AFC, choice="col", axes = 2, top = 17)
sort(colo$contrib[,2], decreasing = TRUE)
colo$cos2[,2]
colo$coord[,2]
```

Résultats :



→ On a pris une moyenne de 6,73

→ Les contributions qui sont supérieures à cette moyenne :

Bac+3D, Bac+4V, Bac+1V et Bac+2M

→ Toutes ces variables ont une bonne qualité c-à-d supérieure à 55% sauf Bac+3D et BAC+4V

```
> fviz_contrib(res.AFC, choice="col", axes = 2, top = 17)
> sort(colo$contrib[,2], decreasing = TRUE)
```

Bac.1.V	Bac.3.D	Bac.2.M	Bac.4.V	Bac.1.C
6.334400e+01	1.174493e+01	7.305709e+00	6.734999e+00	4.580132e+00
moins.que.Bac.V	Plus.que.Bac.5.C	Bac.C	moins.que.Bac.M	Bac.3.C
4.299208e+00	8.373265e-01	3.843111e-01	3.143637e-01	1.800858e-01
Plus.que.Bac.5.M	Bac.2.C	Bac.5.M	Bac.4.M	Bac.5.C
1.743472e-01	6.664544e-02	1.526641e-02	9.884624e-03	5.742317e-03
moins.que.Bac.C	Bac.3.M			
3.047971e-03	7.459219e-08			

```
> colo$cos2[,2]
```

moins.que.Bac.C	moins.que.Bac.M	moins.que.Bac.V	Bac.C	Bac.1.V
1.926370e-04	1.671379e-02	4.822755e-01	2.590495e-02	9.245641e-01
Bac.1.C	Bac.2.M	Bac.2.C	Bac.3.C	Bac.3.M
1.789254e-01	6.769586e-01	5.817610e-03	1.480733e-02	1.187166e-07
Bac.3.D	Bac.4.M	Bac.4.V	Bac.5.M	Bac.5.C
2.644789e-01	3.990034e-04	5.231320e-01	2.006478e-03	5.649134e-04
Plus.que.Bac.5.M	Plus.que.Bac.5.C			
1.078361e-01	7.204775e-02			

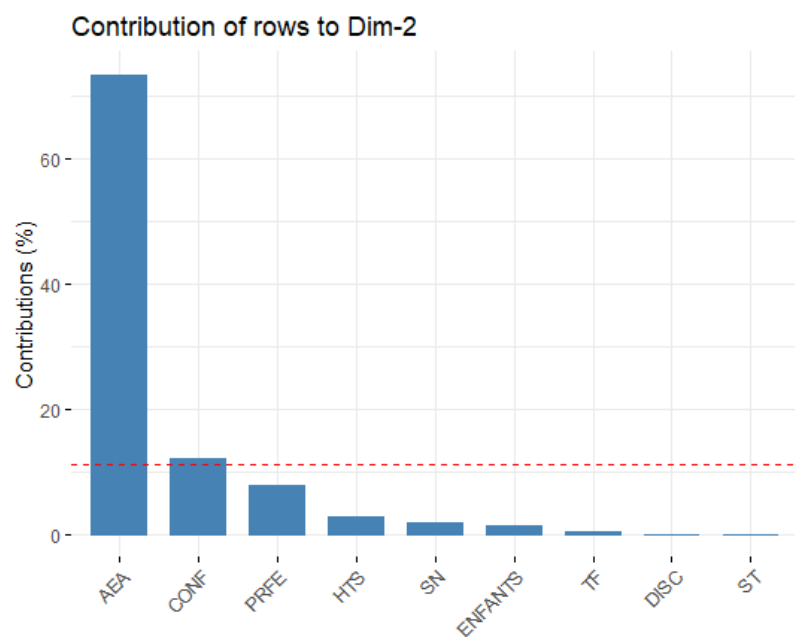
```
> colo$coord[,2]
```

moins.que.Bac.C	moins.que.Bac.M	moins.que.Bac.V	Bac.C	Bac.1.V
8.106715e-03	6.211549e-02	-2.400911e-01	-1.238132e-01	1.246957e+00
Bac.1.C	Bac.2.M	Bac.2.C	Bac.3.C	Bac.3.M
-2.818386e-01	3.841254e-01	-3.925690e-02	6.753436e-02	3.764111e-05
Bac.3.D	Bac.4.M	Bac.4.V	Bac.5.M	Bac.5.C
-4.244866e-01	1.136987e-02	-3.199112e-01	1.950895e-02	-1.381587e-02
Plus.que.Bac.5.M	Plus.que.Bac.5.C			
6.090209e-02	1.479919e-01			

- Pour les lignes :

```
#contributions des lignes à la dimension 1
fviz_contrib(res.AFC, choice="row", axes = 1, top = 9)
sort(ligne$contrib[,1], decreasing = TRUE)
ligne$cos2[,1]
ligne$coord[,1]
```

Résultats :



- ➔ On a pris une moyenne de 12,03
- ➔ Les contributions qui sont supérieures à cette moyenne :
AEA et CONF
- ➔ Toutes ces variables ont une bonne qualité c-à-d supérieure à 55%
sauf CONF

```
> sort(ligne$contrib[,2], decreasing = TRUE)
      AEA      CONF      PRFE      HTS      SN      ENFANTS      TF
73.505464662 12.036525187  7.924893427  2.889472581  1.780804892  1.398524924  0.437446587
      DISC      ST
0.019629363  0.007238378
> ligne$cos2[,2]
      HTS      TF      AEA      ST      DISC      SN      ENFANTS
0.067878073 0.127834311 0.935339326 0.001242406 0.003504578 0.176025800 0.130431101
      CONF      PRFE
0.404138808 0.069534955
> ligne$coord[,2]
      HTS      TF      AEA      ST      DISC      SN      ENFANTS
-0.08839723 -0.27848167 0.87552681 0.00957394 -0.05335955 0.15650727 0.30491925
      CONF      PRFE
-0.36646955 -0.23865777
> |
```

Résultats Finale :

- ➔ Les femmes qui ont Bac+1 Veuf, Bac+2 Mariée ont passé le plus d'années dans leurs emploi actuel

V. Etape 4 : Formulation des groupes

Code :

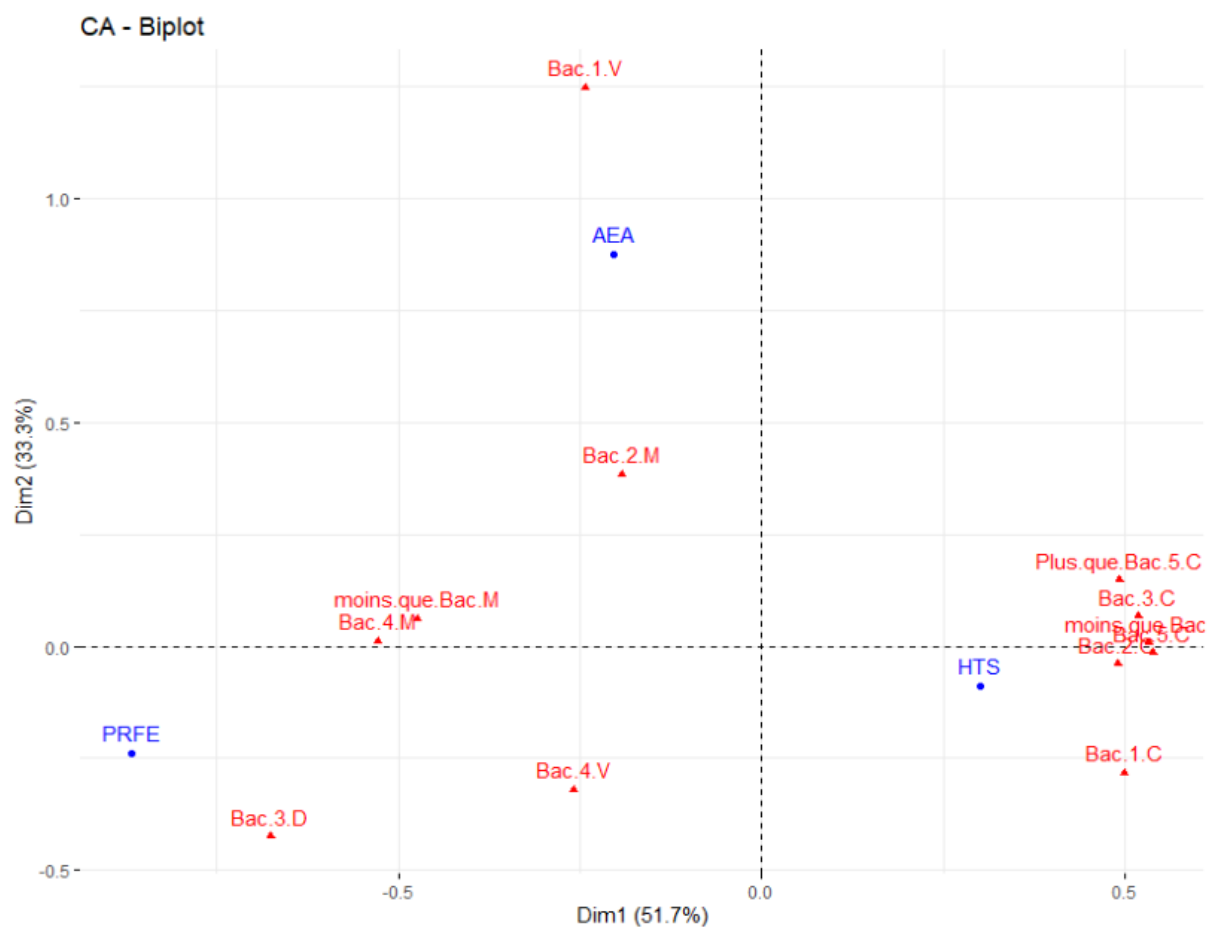
```
#etape 4
sort(ligne$cos2[,1]+ligne$cos2[,2], decreasing = TRUE)
sort(colo$cos2[,1]+colo$cos2[,2], decreasing = TRUE)
fviz_ca_biplot(res.AFC, select.row=list(cos2=0.84), select.col=list(cos2=0.74), axes = 1:2)
```

- ➔ La 1^{ère} ligne dans le code c'est pour connaître la qualité des lignes pour la représentation du dernier graphe
- ➔ La 2^{ème} ligne dans le code c'est pour connaître la qualité des colonnes pour la représentation du dernier graphe

```
> sort(ligne$cos2[,1]+ligne$cos2[,2], decreasing = TRUE)
PRFE      AEA      HTS      ENFANTS      SN      ST      CONF      TF
0.98816585 0.98632272 0.84815208 0.62014604 0.59855876 0.44745461 0.42077648 0.28739433
DISC
0.06109044
> sort(colo$cos2[,1]+colo$cos2[,2], decreasing = TRUE)
moins.que.Bac.M      Bac.1.V      Bac.3.D      Bac.2.C      Bac.3.C
0.99136099      0.95963742      0.93459378      0.91201983      0.88573157
Plus.que.Bac.5.C      Bac.4.V      Bac.4.M      Bac.5.C      Bac.2.M
0.86701638      0.86559862      0.86075020      0.85804897      0.84605769
moins.que.Bac.C      Bac.1.C      Plus.que.Bac.5.M      Bac.5.M      moins.que.Bac.V
0.83134892      0.74032297      0.69109641      0.62988815      0.48377633
Bac.C      Bac.3.M
0.37248153      0.06544766
```

→ Pour les lignes on a pris la qualité supérieure ou égale à 0,7 c'est : **0,84**

→ Pour les colonnes on a pris la qualité supérieure ou égale à 0,7 c'est : **0,74**



Résultats Finale :

→ On a trouvé 3 groupes :

- **Groupe 1** : Bac+1 Veuf, Bac+2 Mariée avec (AEA)
- **Groupe 2** : moins que Bac, Bac+1, Bac+2, Bac+3, Bac+4 et Plus que Bac plus 5 qui ont Célibataire avec (HTS)
- **Groupe 3** : moins que Bac, Bac+4 qui ont Mariée et Bac+3 Divorcée avec (PRFE)

Conclusion

Au nom de notre équipe, nous tenons à exprimer notre sincère gratitude envers nos professeurs, Monsieur Khalidou et Monsieur Harti, pour l'opportunité et l'encadrement qu'ils nous ont offerts lors de cette expérience d'enquête dans la vie réelle. Cette expérience s'est avérée extrêmement enrichissante pour notre parcours académique. Leur guidance et leurs conseils ont été précieux tout au long du processus, nous permettant de développer nos compétences et notre compréhension dans la conduite d'une enquête de manière professionnelle et rigoureuse. Leur dévouement et leur soutien ont grandement contribué à notre réussite, et nous sommes reconnaissants de l'opportunité qui nous a été offerte de bénéficier de leur expertise.