



Rapport d'Analyse

Les Défis de la Recherche de Stages

Réalisé par : Zeineb Jeljli, Houayda Chawali, Mazen Ben Brahim

École Supérieure des Sciences Appliquées et Informatiques (ESSAI)

Encadré par : Mr Ghazi Belmufti

May 04, 2025

Résumé

Ce rapport présente une analyse statistique des données collectées via un questionnaire intitulé « Les défis de la recherche de stages : Partagez votre expérience ! », conçu par les auteurs pour explorer les obstacles rencontrés par les étudiants et jeunes professionnels. À travers des analyses en composantes principales (ACP), des correspondances multiples (ACM) et une classification hiérarchique, ce travail vise à identifier les principaux défis et à proposer des pistes d'amélioration.

Table des matières

1	Introduction	3
1.1	Objectifs du Questionnaire	3
1.2	Structure du Questionnaire	3
1.3	Données Collectées	4
1.4	Méthodologie d'Analyse	4
2	Statistiques Descriptives	4
3	Bloc 1: Préparation et Compétences(ACP)	5
3.1	Objectifs de l'ACP	5
3.2	Importer et explorer le jeu de données	6
3.3	Choix du nombre d'axes à retenir	6
3.4	Interprétation de la carte des variables	7
3.5	Interprétation de la carte des individus	8
3.6	Lien avec les variables qualitatives supplémentaires	9
3.7	Code et Méthodologie	9
4	Bloc 2: Ressources et Accompagnement(ACP)	10
4.1	Objectifs de l'ACP	10
4.2	Importer et explorer le jeu de données	11
4.3	Choix du nombre d'axes à retenir	11
4.4	Interprétation de la carte des variables	12
4.5	Interprétation de la carte des individus	13
4.6	Lien avec les variables qualitatives supplémentaires	14
4.7	Code et Méthodologie	14
5	Bloc 3: Contraintes et Obstacles(ACM)	15
5.1	Objectifs de l'ACM	15
5.2	Importer et explorer le jeu de données	16
5.3	Choix du nombre d'axes à retenir	16
5.4	Interprétation de la carte des modalités	17
5.5	Interprétation de la carte des individus	18
5.6	Lien avec les variables qualitatives supplémentaires	19
5.7	Code et Méthodologie	19
6	Classification Hiérarchique - Bloc 1 : Profils d'Étudiants selon Préparation et Compétences	20
6.1	Objectifs de la Classification	20
6.2	Classification Hiérarchique Ascendante (CAH) avec agnes	20
6.3	Choix du Nombre de Classes	21
6.4	Description des Classes (CAH)	23
6.5	Représentation des Classes sur le Plan Factoriel (CAH)	23
7	Conclusion	24

1 Introduction

Ce rapport présente une analyse statistique des données collectées via un questionnaire en ligne intitulé « Les défis de la recherche de stages : Partagez votre expérience ! », administré sur Google Forms. Ce questionnaire a été conçu par nous, Zeineb Jelji, Houayda Chawali et Mazen Ben Brahim. Il vise à explorer les obstacles rencontrés par les étudiants et jeunes professionnels lors de la recherche de stages, dans le but d'identifier les axes clés de ces défis pour mieux comprendre leurs attentes et difficultés.

1.1 Objectifs du Questionnaire

Le questionnaire a été structuré pour répondre à plusieurs objectifs :

- Identifier les perceptions des étudiants concernant leur préparation et leurs compétences pour la recherche de stages (Bloc 1).
- Évaluer leur accès aux ressources et à l'accompagnement dans ce processus (Bloc 2).
- Explorer les contraintes et obstacles pratiques auxquels ils font face (Bloc 3).
- Regrouper les étudiants en profils homogènes pour mieux comprendre les différences dans leurs expériences (Classification).

1.2 Structure du Questionnaire

Le questionnaire est composé de trois grandes parties, après une section initiale recueillant des informations démographiques :

- **Informations générales** : Cette section collecte des données démographiques sur les répondants, notamment leur sexe (**Sexe**, Q1 : Homme/Femme), leur formation/spécialité (**Formation**, Q2 : Informatique, Génie civil, etc.), leur année d'études (**Année**, Q3 : Licence, Master, etc.), leur statut actuel (**Statut**, Q4 : Étudiant/Recherche d'emploi), et leur disponibilité pour un stage (**Disponibilité**, Q5 : 1-3 mois, 3-6 mois, 6+ mois).
- **Bloc 1 - Préparation et Compétences Requises pour les Stages (Q6)** : Cette section comprend 15 questions sur une échelle de Likert (1 à 5, de « Pas du tout d'accord » à « Tout à fait d'accord »), évaluant la préparation des étudiants et leurs compétences techniques. Exemples de variables : **Maîtrise_compétences** (« Je maîtrise les compétences techniques requises pour les stages dans mon domaine »), **Préparé_entretiens** (« Je me sens bien préparé(e) pour les entretiens de stage »), **Difficulté_ciblage** (« Je trouve difficile de cibler les entreprises qui correspondent à mon profil »).
- **Bloc 2 - Ressources et Accompagnement (Q7)** : Cette section comprend 15 questions sur une échelle de Likert (1 à 5), explorant l'accès aux ressources et le soutien perçu. Exemples de variables : **Notif_adapt.** (« Je reçois suffisamment de notifications pour des offres de stage adaptées »), **Privil_exp** (« Les entreprises privilégient les candidats ayant déjà de l'expérience »), **Diff_repér** (« Je trouve difficile de repérer des offres de stage pertinentes pour mon profil »).

- **Bloc 3 - Contraintes et Obstacles (Q8)** : Cette section comprend 8 questions binaires (Oui/Non), identifiant les contraintes pratiques. Exemples de variables : **Dur_près** (« Est-ce dur de trouver un stage près de chez toi ? »), **Frais_transp** (« Les frais de transport te posent-ils problème ? »), **Stage_nonpayé** (« Peux-tu faire un stage non payé sans problème ? »).

1.3 Données Collectées

Le questionnaire a été rempli par 61 répondants, majoritairement des étudiants, issus de diverses formations et années d'études. Le tableau de données final contient 61 lignes (individus) et 38 colonnes, correspondant aux variables suivantes :

- 4 variables démographiques qualitatives : **Sexe**, **Formation**, **Année**, **Disponibilité**.
- 15 variables quantitatives (échelle de Likert, 1 à 5) pour le Bloc 1.
- 15 variables quantitatives (échelle de Likert, 1 à 5) pour le Bloc 2.
- 8 variables qualitatives binaires (Oui/Non) pour le Bloc 3.

1.4 Méthodologie d'Analyse

Les données ont été analysées en plusieurs étapes :

1. **Analyse en Composantes Principales (ACP)** : Appliquée aux Blocs 1 et 2 pour explorer les relations entre les variables quantitatives et identifier des profils d'étudiants (Sections 3 et 4).
2. **Analyse des Correspondances Multiples (ACM)** : Réalisée sur le Bloc 3 pour analyser les relations entre les contraintes (Section 5).
3. **Classification Hiérarchique** : Effectuée sur le Bloc 1 pour regrouper les étudiants en classes homogènes, en utilisant une CAH classique (Section 6).

2 Statistiques Descriptives

Avant de procéder aux analyses multivariées, une exploration des caractéristiques démographiques des 61 répondants a été réalisée. Les répartitions par sexe, formation et année d'études sont illustrées ci-dessous.

Les graphiques (Figure 1) montrent que :

- **Sexe** : La répartition est équilibrée avec 29 femmes et 32 hommes.
- **Formation** : La majorité des répondants (46 étudiants) sont en Informatique/Développement, suivis de quelques étudiants en Finance/Gestion (5), Génie civil (3), Génie mécanique (3), Médecine/Santé (2), et Réseau/Télécom (2).
- **Année d'études** : La majorité (43 étudiants) sont en cycle d'ingénieur, suivis de 10 étudiants en licence et 8 en master.

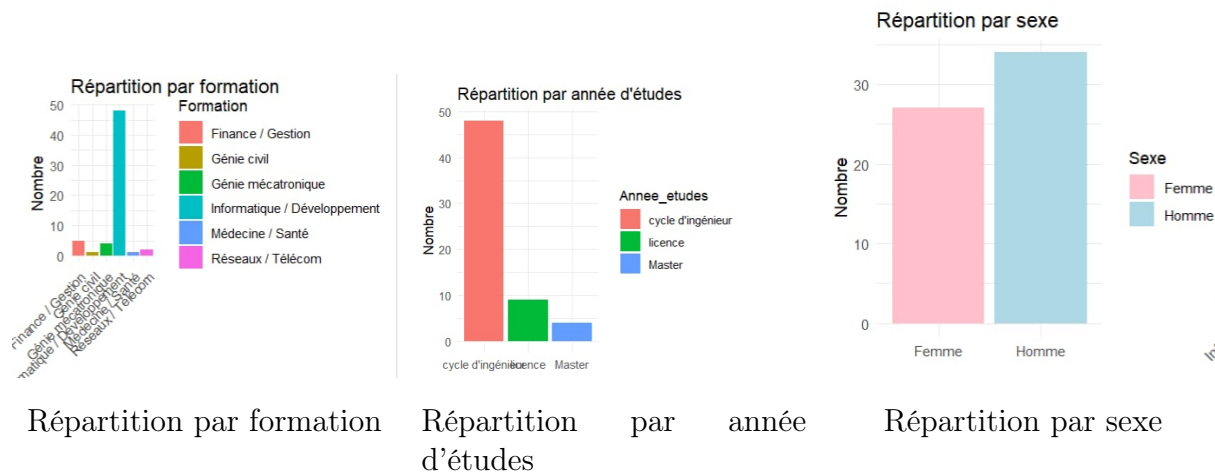


Figure 1: Répartitions démographiques des répondants selon le sexe, la formation et l'année d'études.

3 Bloc 1: Préparation et Compétences(ACP)

Les données analysées dans cette section concernent le Bloc 1 du questionnaire (Q6), qui évalue les perceptions des étudiants relatives à leur préparation et leurs compétences pour la recherche de stages. Ce bloc comprend 15 variables quantitatives sur une échelle de Likert (1 à 5), telles que **Maîtrise_compétences** (« Je maîtrise les compétences techniques requises pour les stages dans mon domaine »), **Préparé_entretiens** (« Je me sens bien préparé(e) pour les entretiens de stage »), et **Difficulté_ciblage** (« Je trouve difficile de cibler les entreprises qui correspondent à mon profil »). Le tableau de données contient 61 lignes correspondant aux individus (étudiants) et 38 colonnes, dont les 15 premières variables quantitatives du Bloc 1, ainsi que quatre variables démographiques qualitatives supplémentaires (**Sexe**, **Formation**, **Année**, **Disponibilité**). L'objectif de cette analyse est d'explorer la variabilité entre les individus, d'identifier des profils d'étudiants, et d'analyser les relations linéaires entre les variables.

3.1 Objectifs de l'ACP

L'ACP réalisée sur ce bloc répond à plusieurs questions :

- *Étude des individus* : Deux étudiants sont proches s'ils ont des profils similaires en termes de préparation et de compétences. Peut-on identifier des groupes d'étudiants avec des caractéristiques communes ? Existe-t-il une opposition entre certains profils ?
- *Étude des variables* : Quelles sont les corrélations entre les variables du Bloc 1 ? Peut-on résumer ces perceptions par un petit nombre de dimensions ?
- *Lien entre individus et variables* : Les groupes d'étudiants identifiés peuvent-ils être caractérisés par des variables spécifiques ?

3.2 Importer et explorer le jeu de données

Les données ont été chargées et explorées à l'aide des packages `FactoMineR` et `factoextra`. La matrice des corrélations, visualisée via un corrrgramme (Figure 2), révèle des corrélations significatives entre certaines variables, notamment entre `Maîtrise_compétences`, `Confiance_entretiens`, et `Préparé_entretiens` (corrélations positives fortes, $r > 0,75$), ainsi qu'entre `Difficulté_mise_pratique` et `Difficulté_mise_valeur` ($r \approx 0,5$). Ces relations suggèrent la pertinence d'une ACP.

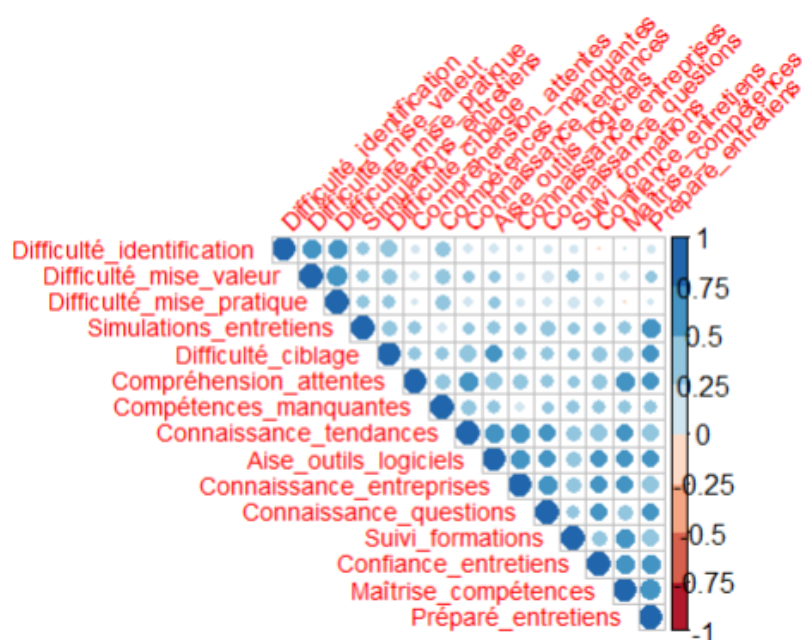


Figure 2: Corrgramme des variables du Bloc 1, montrant les corrélations entre les variables quantitatives.

3.3 Choix du nombre d'axes à retenir

L'analyse des valeurs propres indique que la première dimension explique 26,27% de la variance totale, la deuxième 9,12%, et la troisième 6,92%, les trois premières dimensions cumulées expliquant 42,31% de la variance. Le scree plot (Figure 3) illustre cette répartition. Trois critères ont été utilisés :

- *Critère de Kaiser* : Quatre axes ont des valeurs propres supérieures à 1.
- *Taux d'inertie cumulé* : Les trois premiers axes expliquent 42,31% de la variance, offrant une synthèse significative.
- *Critère du coude* : Le scree plot montre un coude après le deuxième axe.

Les deux premiers axes sont retenus, représentant 42,31% de l'inertie totale.

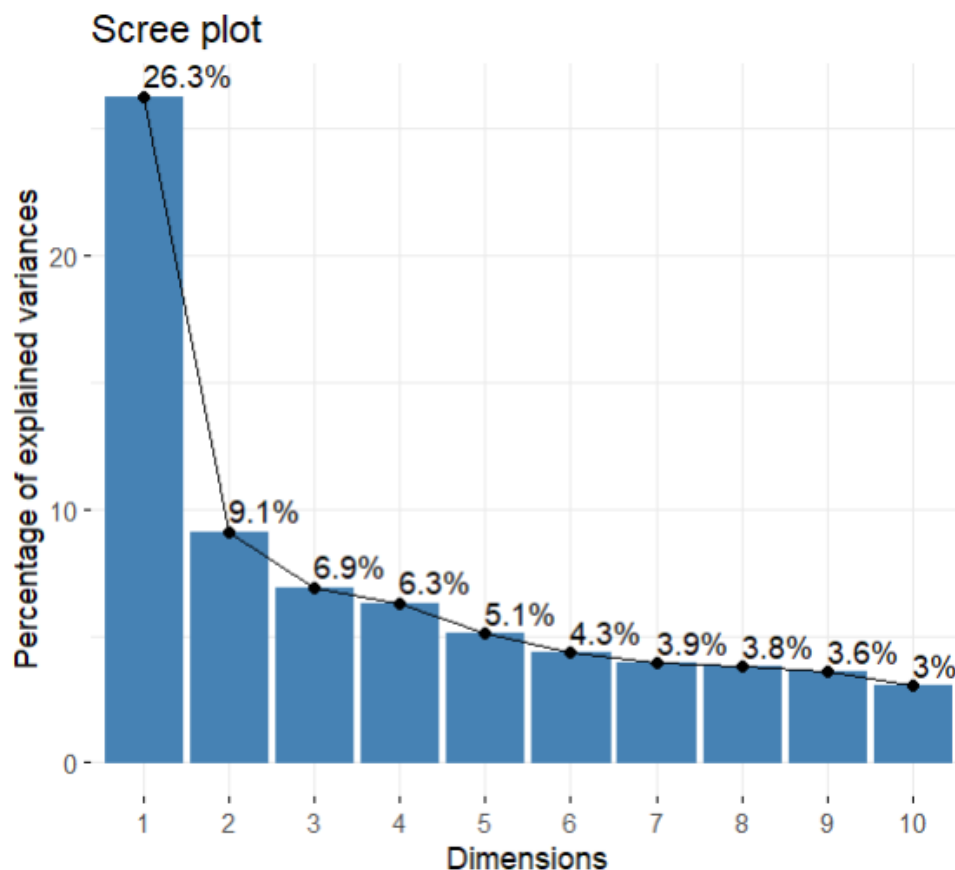


Figure 3: Scree plot des valeurs propres pour l'ACP du Bloc 1, montrant la répartition de la variance expliquée par chaque dimension.

3.4 Interprétation de la carte des variables

La carte des variables (Figure 4), basée sur les coordonnées et les \cos^2 , montre les relations entre les 15 variables du Bloc 1 sur les trois premiers axes. La première dimension (26,27%) est fortement corrélée avec *Maîtrise_compétences* (0,66), *Préparé_entretiens* (0,70), *Confiance_entretiens* (0,65), et *Aise_outils_logiciels* (0,71), représentant un axe de « Compétence et préparation ». La deuxième dimension (9,12%) est associée à *Difficulté_mise_pratique* (0,47), *Difficulté_mise_valeur* (0,37), et *Difficulté_identification* (0,48), indiquant un axe de « Difficultés pratiques ».

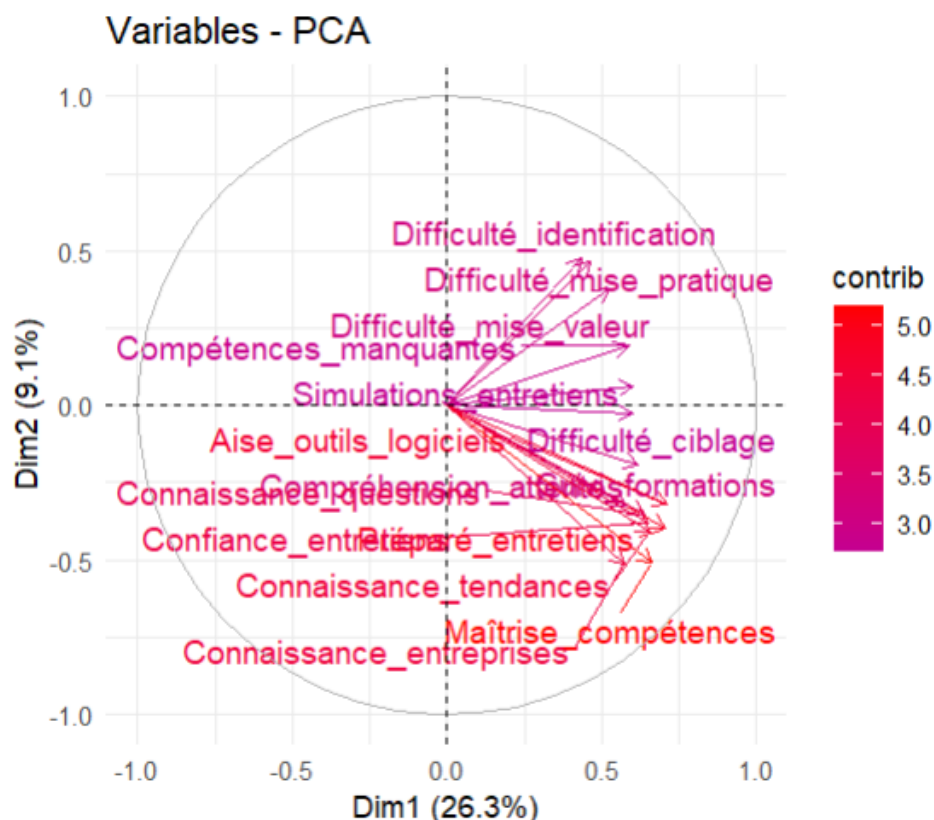


Figure 4: Carte des variables pour l'ACP du Bloc 1, illustrant les corrélations entre les variables sur les trois premiers axes.

3.5 Interprétation de la carte des individus

La carte des individus (Figure 5), colorée selon les \cos^2 , montre une dispersion des étudiants sur les trois premiers axes. Les individus avec des \cos^2 élevés (proches de 0,8, en rouge) sont bien représentés, tandis que ceux avec des valeurs faibles (en bleu, proches de 0,2) le sont moins. La première dimension sépare les étudiants compétents et préparés (côté positif) de ceux moins préparés (côté négatif). La deuxième dimension distingue les étudiants confrontés à des difficultés pratiques (côté positif) de ceux moins affectés (côté négatif).

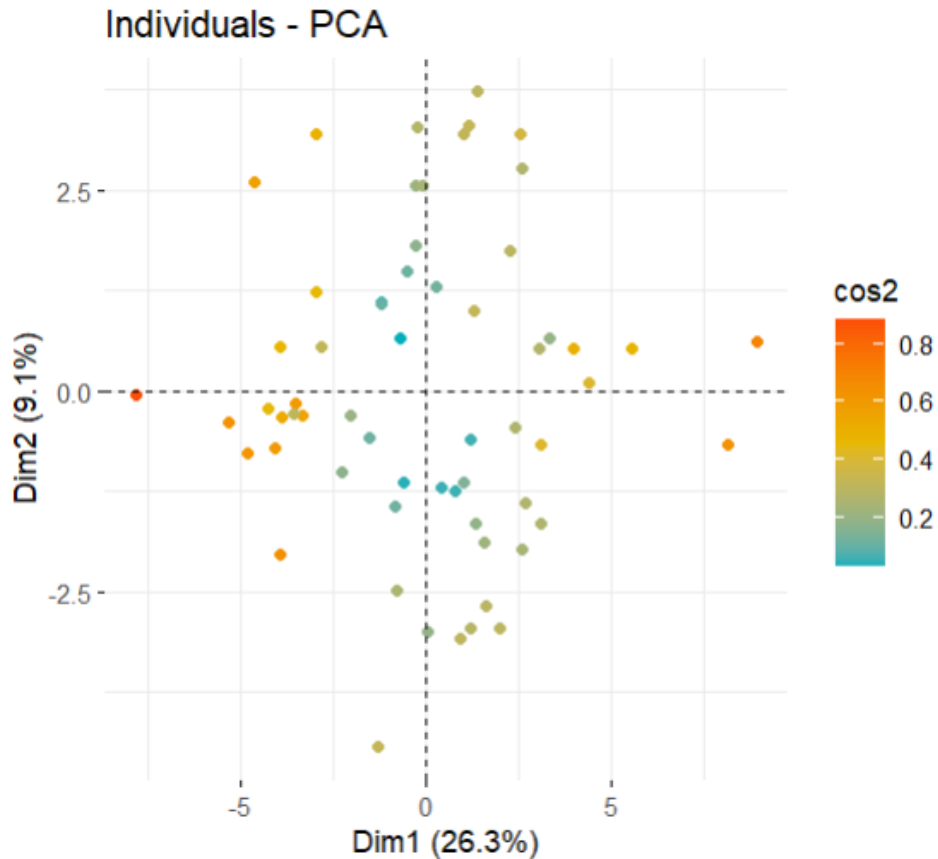


Figure 5: Carte des individus pour l'ACP du Bloc 1, montrant la répartition des étudiants selon les \cos^2 sur les trois premiers axes.

3.6 Lien avec les variables qualitatives supplémentaires

Les variables démographiques (**Sexe**, **Formation**, **Année**, **Disponibilité**) peuvent être explorées sur les trois dimensions pour identifier des tendances. Par exemple, on pourrait examiner si les étudiants en cycle d'ingénieur (Q3) se sentent mieux préparés que ceux en Licence.

3.7 Code et Méthodologie

Le code R utilisé pour l'analyse du Bloc 1 est illustré dans la capture d'écran suivante (Figure 6), incluant toutes les étapes : importation, matrice des corrélations, ACP, valeurs propres, et visualisations.

```

colonnes_likert_bloc1 <- c(
  "Maîtrise_compétences",
  "Suivi_formations",
  "Préparé_entretiens",
  "Simulations_entretiens",
  "Compréhension_attentes",
  "Connaissance_entreprises",
  "Confiance_entretiens",
  "Connaissance_questions",
  "Difficulté_ciblage",
  "Connaissance_tendances",
  "Difficulté_mise_valeur",
  "Difficulté_mise_pratique",
  "Difficulté_identification",
  "Compétences_manquantes",
  "Aise_outils_logiciels"
)

# Sélectionner le bloc pour l'ACP
data_bloc1 <- data[, colonnes_likert_bloc1]

# Explorer les données
head(data_bloc1) # Afficher les premières lignes
str(data_bloc1) # Vérifier la structure des variables
summary(data_bloc1) # Résumé statistique

# Calculer et visualiser la matrice des corrélations pour le bloc
M <- cor(data_bloc1, use = "complete.obs")
corplot(M, type = "upper", order = "hclust",
  col = brewer.pal(n = 8, name = "RdBu"),
  tl.cex = 0.7, tl.srt = 45)

# Exécuter l'ACP avec variables démographiques comme qualitatives supplémentaires
res.pca <- PCA(data,
  scale.unit = TRUE,
  ncp = 5,
  quanti.sup = NULL,
  quali.sup = 1:4, # Sexe, formation, année, disponibilité
  graph = FALSE)

# Afficher les valeurs propres
print(res.pca$eig)

# Visualiser le scree plot
fviz_screplot(res.pca, ncp = 10, addlabels = TRUE)

# Coordonnées et cos2 des variables
print(res.pca$var$coord[colonnes_likert_bloc1, ])
print(res.pca$var$cos2[colonnes_likert_bloc1, ])

# Graphique des variables
fviz_pca_var(res.pca,
  col.var = "contrib",
  select.var = list(name = colonnes_likert_bloc1,
    repel = TRUE) +
  scale_color_gradient2(low = "white", mid = "blue", high = "red", midpoint = 0.6)
  theme_minimal()

# Graphique des individus
fviz_pca_ind(res.pca,
  geom = "point",
  col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

```

Figure 6: Capture d'écran du code R complet pour l'analyse ACP du Bloc 1.

4 Bloc 2: Ressources et Accompagnement(ACP)

Les données analysées dans cette section concernent le Bloc 2 du questionnaire (Q7), qui évalue les perceptions des étudiants relatives aux ressources et à l'accompagnement dans leur recherche de stages. Ce bloc comprend 15 variables quantitatives sur une échelle de Likert (1 à 5), telles que *Notif_adapt.* (« Je reçois suffisamment de notifications pour des offres de stage adaptées »), *Privil_exp* (« Les entreprises privilégient les candidats ayant déjà de l'expérience »), et *Diff_repér* (« Je trouve difficile de repérer des offres de stage pertinentes pour mon profil »). Le tableau de données contient 61 lignes correspondant aux individus (étudiants) et 38 colonnes, dont les 15 variables quantitatives du Bloc 2 (colonnes 20 à 34), ainsi que quatre variables démographiques qualitatives supplémentaires (*Sexe*, *Formation*, *Année*, *Disponibilité*). L'objectif est d'explorer la variabilité des perceptions, d'identifier des profils d'étudiants, et d'analyser les relations entre les variables.

4.1 Objectifs de l'ACP

L'ACP réalisée sur ce bloc répond à plusieurs questions :

- *Étude des individus* : Deux étudiants sont proches s'ils ont des profils similaires en termes de ressources et d'accompagnement. Peut-on identifier des groupes d'étudiants avec des caractéristiques communes ?
- *Étude des variables* : Quelles sont les corrélations entre les variables du Bloc 2 ? Peut-on résumer ces perceptions par un petit nombre de dimensions ?
- *Lien entre individus et variables* : Les groupes d'étudiants identifiés peuvent-ils être caractérisés par des variables spécifiques ?

4.2 Importer et explorer le jeu de données

Les données ont été chargées et explorées à l'aide des packages `FactoMineR` et `factoextra`. La matrice des corrélations, visualisée via un corrgramme (Figure 7), révèle des corrélations significatives entre certaines variables, notamment entre `Plateformes` et `Privil_exp` (corrélations positives fortes), ainsi qu'entre `Diff_démarq` et `Conseils_pr`.

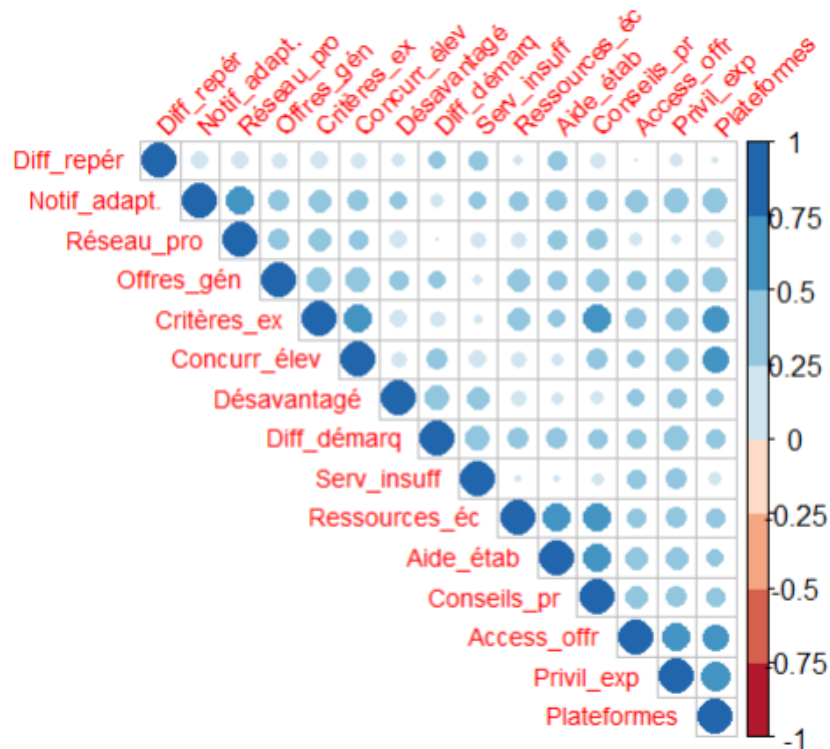


Figure 7: Corrgramme des variables du Bloc 2, montrant les corrélations entre les variables quantitatives.

4.3 Choix du nombre d'axes à retenir

L'analyse des valeurs propres indique que la première dimension explique 26,27% de la variance totale, la deuxième 9,12%, et la troisième 6,92%, les trois premières dimensions cumulées expliquant 42,31% de la variance. Le scree plot (Figure 8) illustre cette répartition. Trois critères ont été utilisés :

- *Critère de Kaiser* : Quatre axes ont des valeurs propres supérieures à 1.
- *Taux d'inertie cumulé* : Les trois premiers axes expliquent 42,31% de la variance, offrant une synthèse significative.
- *Critère du coude* : Le scree plot montre un coude après le deuxième axe.

Les deux premiers axes sont retenus, représentant 42,31% de l'inertie totale.

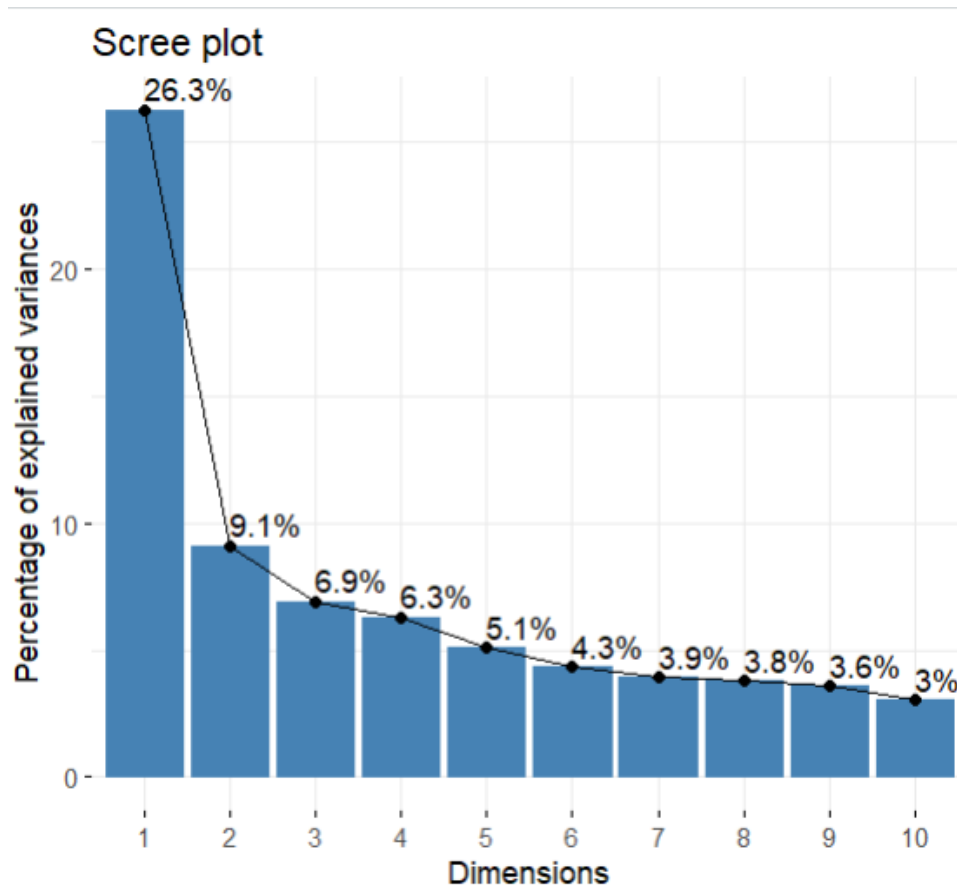


Figure 8: Scree plot des valeurs propres pour l'ACP du Bloc 2, montrant la répartition de la variance expliquée par chaque dimension.

4.4 Interprétation de la carte des variables

La carte des variables (Figure 9), basée sur les coordonnées et les \cos^2 , montre les relations entre les 15 variables du Bloc 2 sur les trois premiers axes. La première dimension (26,27%) est fortement corrélée avec *Notif_adapt.* (0,71), *Privil_exp* (0,69), et *Plateformes* (0,69), représentant un axe de « Utilisation des ressources numériques ». La deuxième dimension (9,12%) est associée à *Conseils_pr* (0,56), *Diff_démarq* (0,55), et *Ressources_éc* (0,46), indiquant un axe de « Soutien et difficultés initiales ».

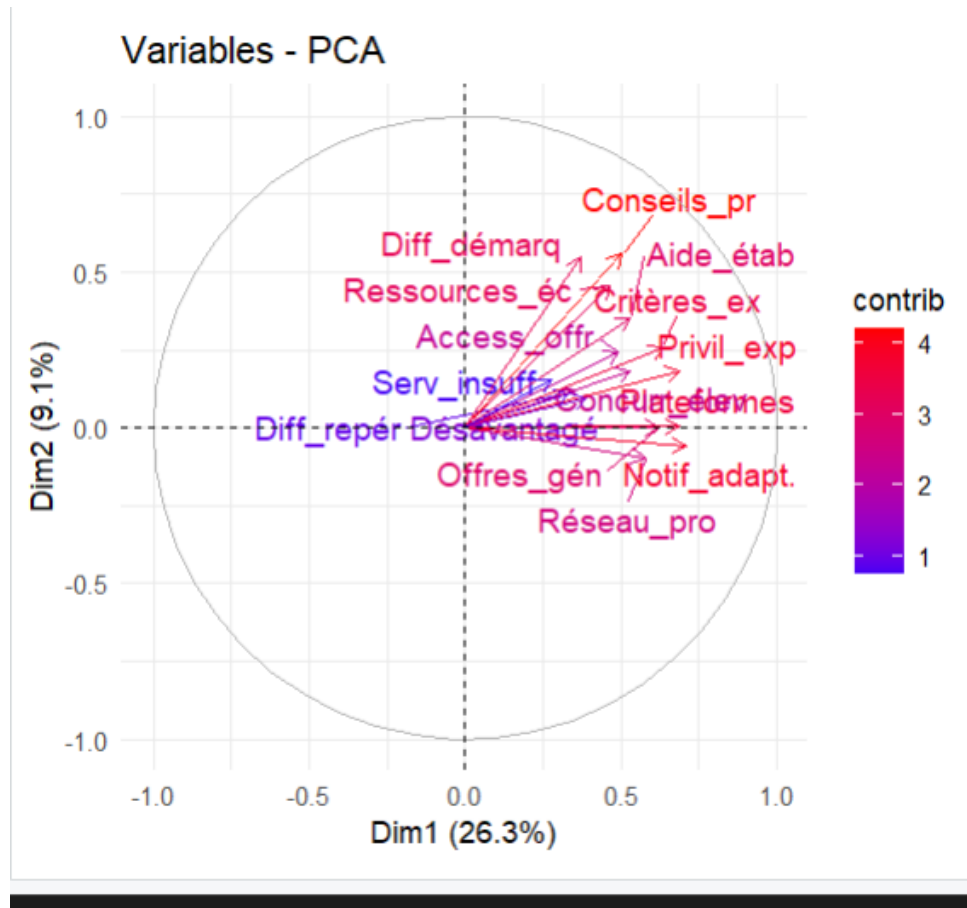


Figure 9: Carte des variables pour l'ACP du Bloc 2, illustrant les corrélations entre les variables sur les trois premiers axes.

4.5 Interprétation de la carte des individus

La carte des individus (Figure 10), colorée selon les \cos^2 , montre une dispersion des étudiants sur les trois premiers axes. Les individus avec des \cos^2 élevés (proches de 0,8, en rouge) sont bien représentés, tandis que ceux avec des valeurs faibles (en bleu, proches de 0,2) le sont moins. La première dimension sépare les étudiants utilisant intensément les ressources numériques (côté positif) de ceux moins engagés (côté négatif). La deuxième dimension distingue ceux bénéficiant d'un bon soutien (côté positif) de ceux rencontrant des difficultés initiales (côté négatif).

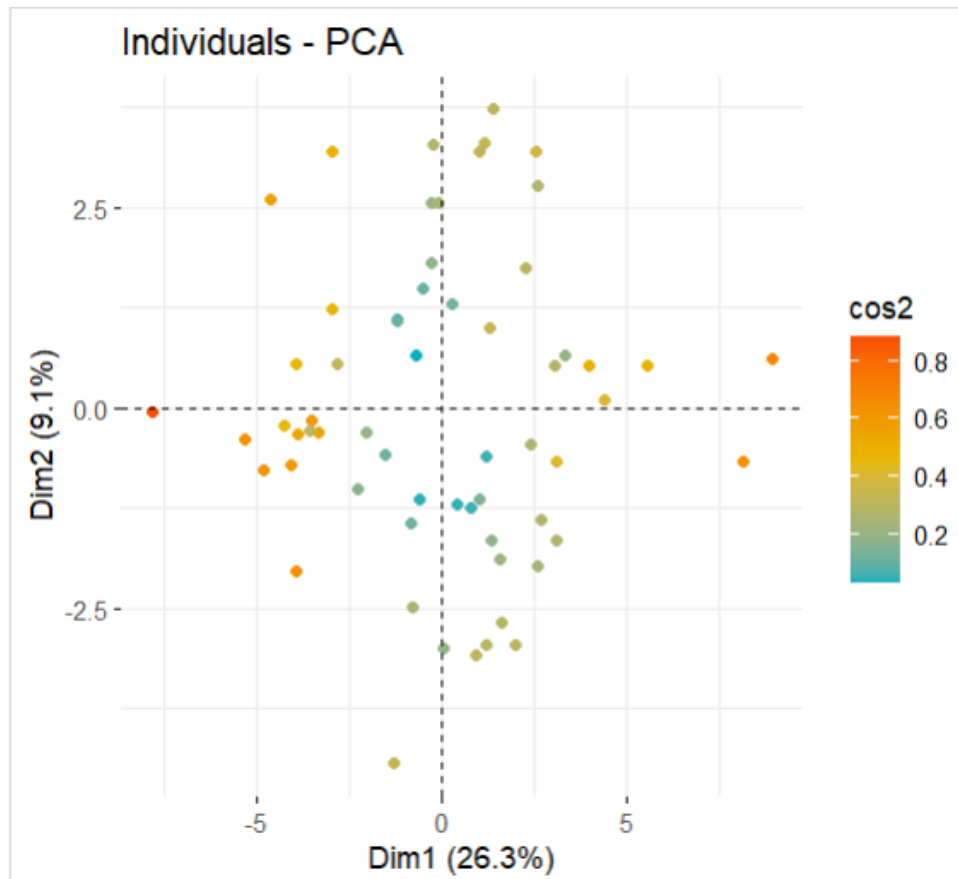


Figure 10: Carte des individus pour l'ACP du Bloc 2, montrant la répartition des étudiants selon les \cos^2 sur les trois premiers axes.

4.6 Lien avec les variables qualitatives supplémentaires

Les variables démographiques (**Sexe**, **Formation**, **Année**, **Disponibilité**) peuvent être analysées sur les trois dimensions pour explorer des tendances spécifiques. Par exemple, on pourrait examiner si les étudiants en Génie mécanique (Q2) perçoivent plus de difficultés pour repérer des offres pertinentes.

4.7 Code et Méthodologie

Le code R utilisé pour l'analyse du Bloc 2 est illustré dans la capture d'écran suivante (Figure 11), incluant toutes les étapes : importation, matrice des corrélations, ACP, valeurs propres, et visualisations.

```

# Sélection des colonnes du bloc 2 (variables Likert du bloc 2)
colonnes_likert_bloc2 <- colnames(data)[20:34]
data_bloc2 <- data[, colonnes_likert_bloc2]

# Étape 1 : Calcul et visualisation de la matrice des corrélations
M2 <- cor(data_bloc2, use = "complete.obs")
corrplot(M2, type = "upper", order = "hclust",
  col = brewer.pal(n = 8, name = "RdBu"),
  tl.cex = 0.7, tl.srt = 45)
# Étape 2 : Exécuter l'ACP avec variables démographiques comme qualitatives supplémentaires
res.pca_bloc2 <- PCA(data,
  scale.unit = TRUE,
  ncp = 5,
  quanti.sup = NULL,
  quali.sup = 1:4, # Sexe, formation, année, disponibilité
  graph = FALSE)

# Étape 3 : Afficher les valeurs propres
print(res.pca_bloc2$eig)
# Étape 4 : Visualiser le scree plot
fviz_screepplot(res.pca_bloc2, ncp = 10, addlabels = TRUE)
# Sauvegarde du scree plot pour le rapport
ggsave("screepplot_bloc2.png", width = 8, height = 6)
# Étape 5 : Coordonnées et cos2 des variables
print(res.pca_bloc2$var$coord[colonnes_likert_bloc2, ])
print(res.pca_bloc2$var$cos2[colonnes_likert_bloc2, ])

# Étape 6 : Graphique des variables
fviz_pca_var(res.pca_bloc2,
  col.var = "contrib",
  select.var = list(name = colonnes_likert_bloc2,
    repel = TRUE) +
  scale_color_gradient2(low = "white", mid = "blue", high = "red", midpoint = 0.6) +
  theme_minimal()
# Étape 7 : Graphique des individus
fviz_pca_ind(res.pca_bloc2,
  geom = "point",
  col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))
# Sauvegarde de la carte des individus pour le rapport
ggsave("individuals_pca_bloc2.png", width = 8, height = 6)

```

Figure 11: Capture d'écran du code R complet pour l'analyse ACP du Bloc 2.

5 Bloc 3: Contraintes et Obstacles(ACM)

Les données analysées dans cette section concernent le Bloc 3 du questionnaire (Q8), qui évalue les contraintes et obstacles rencontrés par les étudiants dans leur recherche de stages. Ce bloc comprend 8 variables qualitatives binaires (*Dur_près*, *Frais_transp*, *Stage_nonpayé*, *Complicqué_étud*, *Obligations*, *Chercher_plus*, *Engagements*, *Contacts_necess*), encodées en 0 (Non) et 1 (Oui). Exemples : *Dur_près* (« Est-ce dur de trouver un stage près de chez toi ? »), *Stage_nonpayé* (« Peux-tu faire un stage non payé sans problème ? »), *Contacts_necess* (« As-tu besoin de contacts pour trouver un stage ? »). Le tableau de données contient 61 lignes correspondant aux individus (étudiants) et 38 colonnes, dont les colonnes correspondant aux 8 variables du Bloc 3. L'objectif est d'explorer les relations entre ces variables qualitatives et d'identifier des profils d'étudiants en fonction des contraintes rencontrées.

5.1 Objectifs de l'ACM

L'ACM réalisée sur ce bloc répond à plusieurs questions :

- *Étude des individus* : Deux étudiants sont proches s'ils partagent des profils similaires en termes de contraintes. Peut-on identifier des groupes d'étudiants avec des caractéristiques communes ?
- *Étude des modalités* : Quelles sont les associations entre les modalités des variables du Bloc 3 ? Peut-on résumer ces relations par un petit nombre de dimensions ?

- *Lien entre individus et modalités* : Les groupes d'étudiants identifiés peuvent-ils être caractérisés par des modalités spécifiques ?

5.2 Importer et explorer le jeu de données

Les données ont été chargées et les 8 variables du Bloc 3 ont été recodées en facteurs avec les niveaux « Non » et « Oui ». Une analyse descriptive montre les répartitions suivantes : `Dur_près` (70,5% Oui), `Frais_transp` (73,8% Oui), `Stage_nonpayé` (70,5% Oui), `Complicé_étud` (82% Oui), `Obligations` (21,3% Oui), `Chercher_plus` (73,8% Oui), `Engagements` (57,4% Oui), `Contacts_necess` (93,4% Oui). Ces proportions élevées pour « Oui » dans la plupart des variables, notamment pour `Contacts_necess` (93,4%), suggèrent que les contraintes, en particulier le besoin de contacts, sont fréquentes, justifiant une ACM pour explorer les associations entre ces modalités.

5.3 Choix du nombre d'axes à retenir

L'analyse des valeurs propres (`eig`) indique que la première dimension explique 27,1% de la variance totale, la deuxième 18,2%, et la troisième 14,6%, les trois premières dimensions cumulées expliquant environ 59,9% de la variance. Le Scree plot (Figure 12) illustre cette répartition. Trois critères ont été utilisés pour déterminer le nombre d'axes à retenir :

- *Valeurs propres supérieures à la moyenne* : Avec 8 variables et 16 modalités, l'inertie totale est $(16/8) - 1 = 1$, donc la valeur propre moyenne est $1/8 = 0,125$. Les trois premiers axes (0,271, 0,182, 0,146) dépassent ce seuil, justifiant leur retenue.
- *Taux d'inertie cumulé* : Les trois premiers axes expliquent 59,9% de la variance, un taux significatif pour une synthèse approfondie.
- *Critère du coude* : Le Scree plot montre une contribution notable de la troisième dimension, avec une décroissance progressive après cet axe.

Les trois premiers axes sont retenus pour l'interprétation, représentant 59,9% de l'inertie totale.

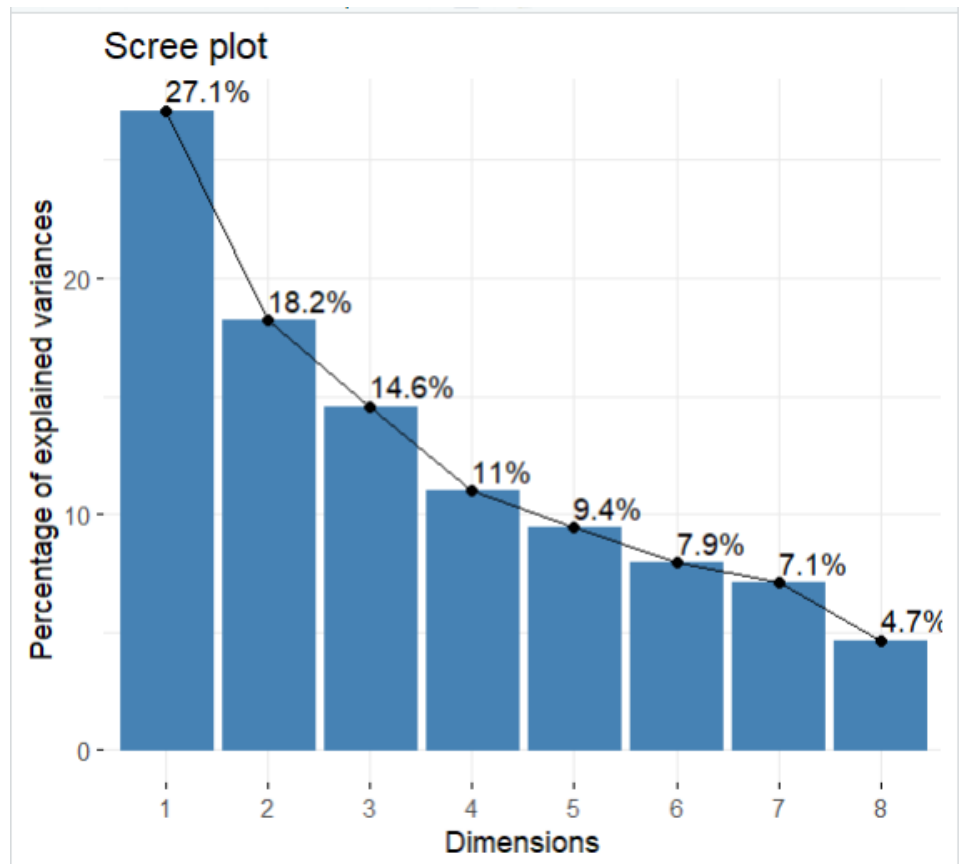


Figure 12: Scree plot des valeurs propres pour l'ACM du Bloc 3, montrant la répartition de la variance expliquée par chaque dimension.

5.4 Interprétation de la carte des modalités

La carte des modalités (Figure 13), basée sur les contributions (`contrib`), montre les relations entre les modalités des 8 variables du Bloc 3 sur les trois premiers axes. La première dimension (27,1%) oppose les modalités **Chercher_plus_Non** (0,476) et **Engagements_Oui** (0,398) à **Chercher_plus_Oui** (-0,476) et **Engagements_Non** (-0,398), selon `dimdesc`, représentant un axe de « Contraintes liées à la recherche prolongée et aux engagements ». La deuxième dimension (18,2%) oppose **Complicé_étud_Non** (0,456) et **Contacts_necess_Non** (0,450) à **Complicé_étud_Oui** (-0,456) et **Contacts_necess_Oui** (-0,450), indiquant un axe de « Difficultés liées aux études et au besoin de contacts ». Les contributions (`contrib`) indiquent que **Chercher_plus** (64,82%) et **Complicé_étud** (67,51%) dominent les axes.

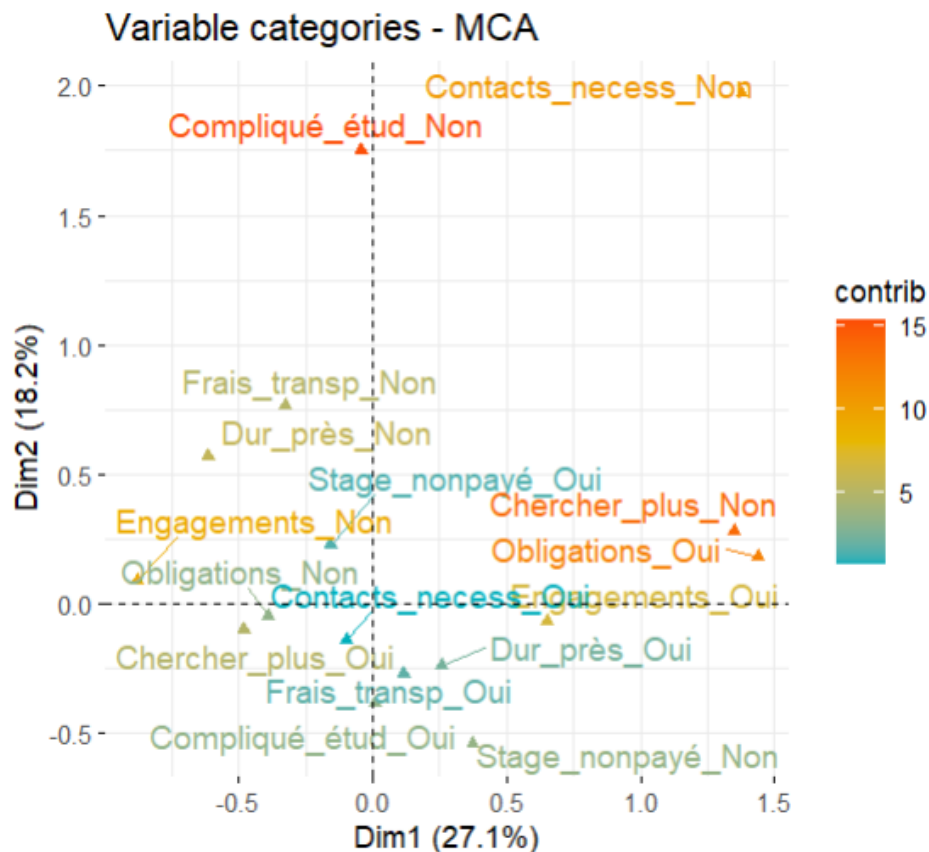


Figure 13: Carte des modalités pour l'ACM du Bloc 3, illustrant les relations entre les modalités sur les trois premiers axes.

5.5 Interprétation de la carte des individus

La carte des individus (Figure 14), colorée selon les \cos^2 , montre une dispersion des étudiants sur les trois premiers axes. Les individus avec des \cos^2 élevés (proches de 0,8, en rouge) sont bien représentés, tandis que ceux avec des valeurs faibles (en bleu, proches de 0,2) le sont moins. La première dimension sépare les étudiants avec une recherche prolongée (côté négatif) de ceux moins engagés (côté positif). La deuxième dimension distingue les étudiants ayant des difficultés liées aux études (côté négatif) de ceux moins affectés (côté positif).

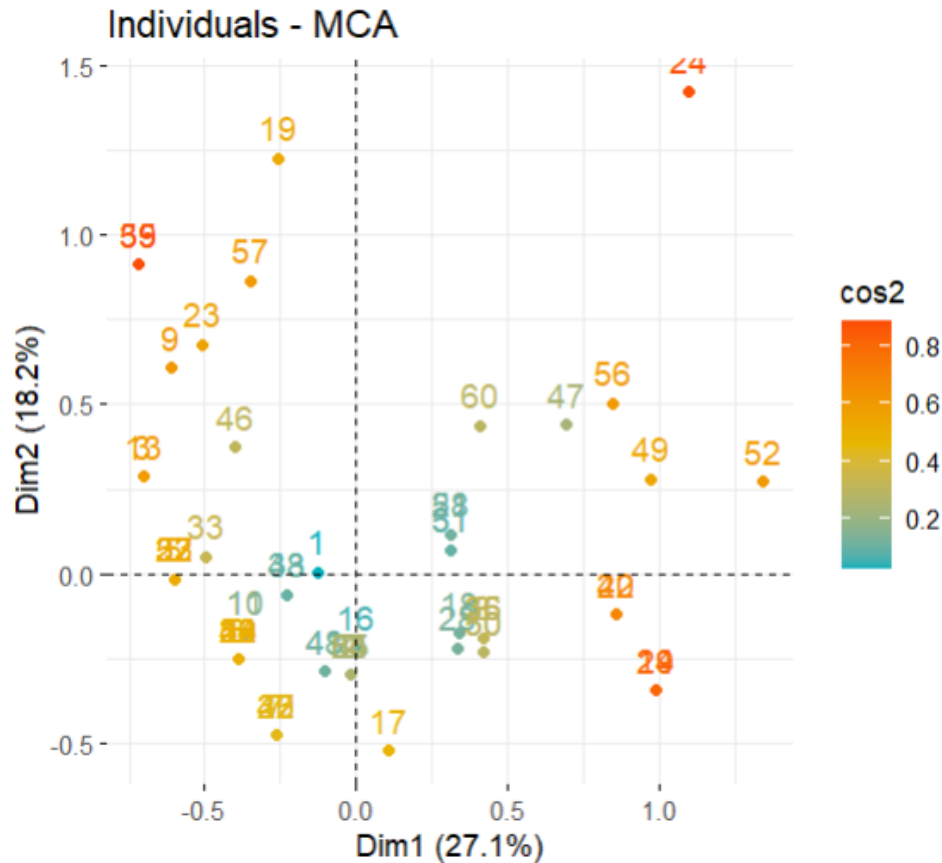


Figure 14: Carte des individus pour l'ACM du Bloc 3, montrant la répartition des étudiants selon les \cos^2 sur les trois premiers axes.

5.6 Lien avec les variables qualitatives supplémentaires

Les variables démographiques (Sexe, Formation, Année, Disponibilité) pourraient être ajoutées comme variables supplémentaires pour enrichir l'interprétation. Par exemple, on pourrait vérifier si les étudiants en Médecine/Santé (Q2) ont un besoin plus marqué de contacts.

5.7 Code et Méthodologie

Le code R utilisé pour l'analyse du Bloc 3 est illustré dans la capture d'écran suivante (Figure 15), incluant toutes les étapes : importation, recodage, ACM, valeurs propres, visualisations, et description des dimensions.

```

# Sélection des colonnes de AI à AP (35 à 43 en index R)
colonnes_acm <- c("Dur_près", "Frais_transp", "Stage_nonpayé", "Complicé_étud", "Obligations", "Chercher_plus", "Engagements",
df_acm <- data[, colonnes_acm]

# Recodage en facteur (variables binaires 0/1)
for(col in colonnes_acm) {
  df_acm[[col]] <- factor(df_acm[[col]],
    levels = c(0, 1),
    labels = c("Non", "Oui"))
  print(paste("Répartition pour", col, ":"))
  print(round(prop.table(table(df_acm[[col]])) * 100, 1))
}

# ACM (Analyse des Correspondances Multiples)
res.mca <- MCA(df_acm, graph = FALSE, ncp = 5)

# Valeurs propres (variance expliquée par chaque dimension)
print(res.mca$eig)

# Graphique des valeurs propres (éboulis)
fviz_screplot(res.mca, addlabels = TRUE)

# Carte des modalités (variables)
fviz_mca_var(res.mca,
  choice = "var.cat",
  repel = TRUE,
  col.var = "contrib",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

# Carte des individus
fviz_mca_ind(res.mca,
  col.ind = "cos2",
  gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"))

# Description des dimensions (corrélations variables-axes)
print(dimdesc(res.mca, axes = 1:2))

```

Figure 15: Capture d'écran du code R complet pour l'analyse ACM du Bloc 3.

6 Classification Hiérarchique - Bloc 1 : Profils d'Étudiants selon Préparation et Compétences

Cette section présente une classification hiérarchique ascendante (CAH) appliquée aux données du Bloc 1 (Q6), qui concernent les perceptions des étudiants sur leur préparation et leurs compétences. L'objectif est d'identifier des groupes d'étudiants ayant des profils similaires, d'abord via une CAH classique avec la méthode **agnes**, puis via une classification hiérarchique sur composantes principales (HCPC) à partir des résultats de l'ACP.

6.1 Objectifs de la Classification

La classification répond aux questions suivantes :

- Peut-on regrouper les étudiants en classes homogènes selon leurs perceptions ?
- Quelles sont les caractéristiques des groupes identifiés ?
- Comment ces groupes se positionnent sur le plan factoriel de l'ACP ?

6.2 Classification Hiérarchique Ascendante (CAH) avec agnes

Une CAH a été réalisée sur les données standardisées du Bloc 1 (15 variables quantitatives) en utilisant la méthode de Ward, qui minimise la perte d'inertie intra-groupe. Le dendrogramme (Figure 16) montre la structure hiérarchique des regroupements, avec un coefficient d'agglomération de 0,87, indiquant une bonne qualité de la classification.

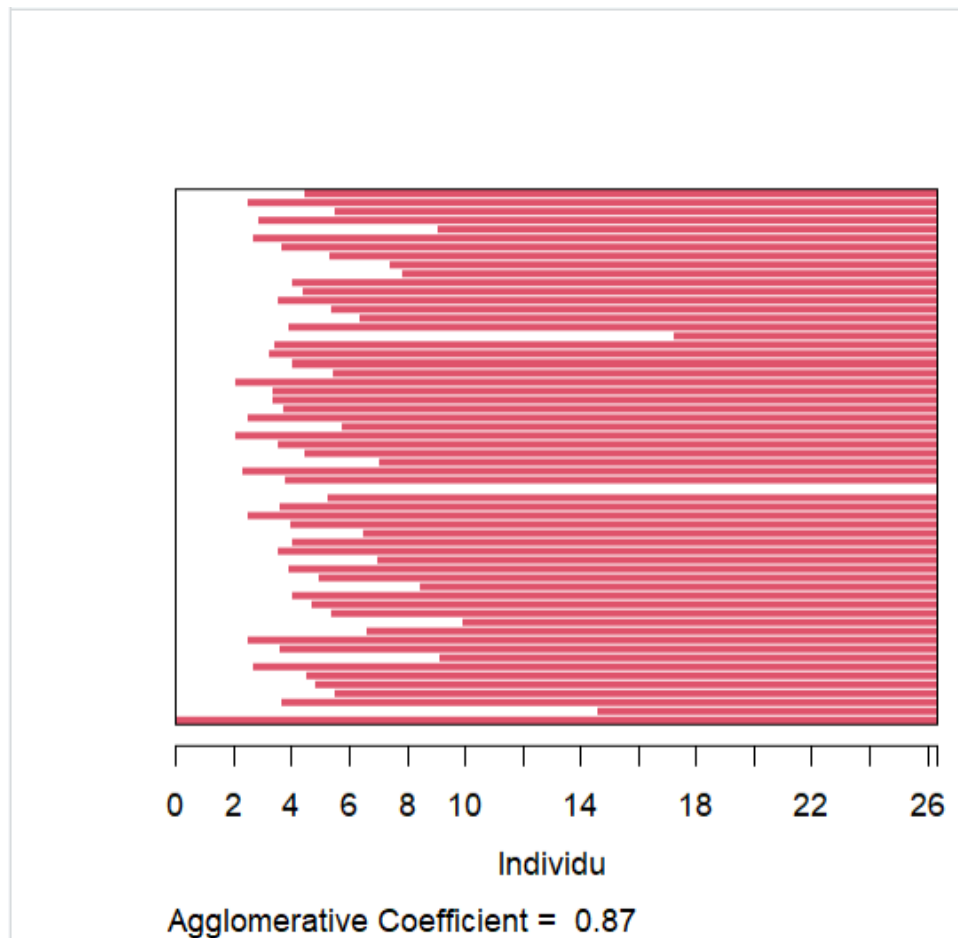


Figure 16: Dendrogramme de la classification hiérarchique (CAH) sur le Bloc 1, avec découpage en 2 classes.

6.3 Choix du Nombre de Classes

Le choix du nombre de classes a été déterminé en examinant l'évolution des hauteurs (Figure 17) et de l'inertie (Figure 18). La courbe des hauteurs montre une diminution progressive, et la courbe de l'inertie en fonction du nombre de classes révèle un coude marqué autour de 2 classes, suggérant une partition en deux groupes. Cette partition est confirmée par le dendrogramme, où une coupure à 2 classes (indiquée en rouge) sépare les étudiants en deux groupes distincts.

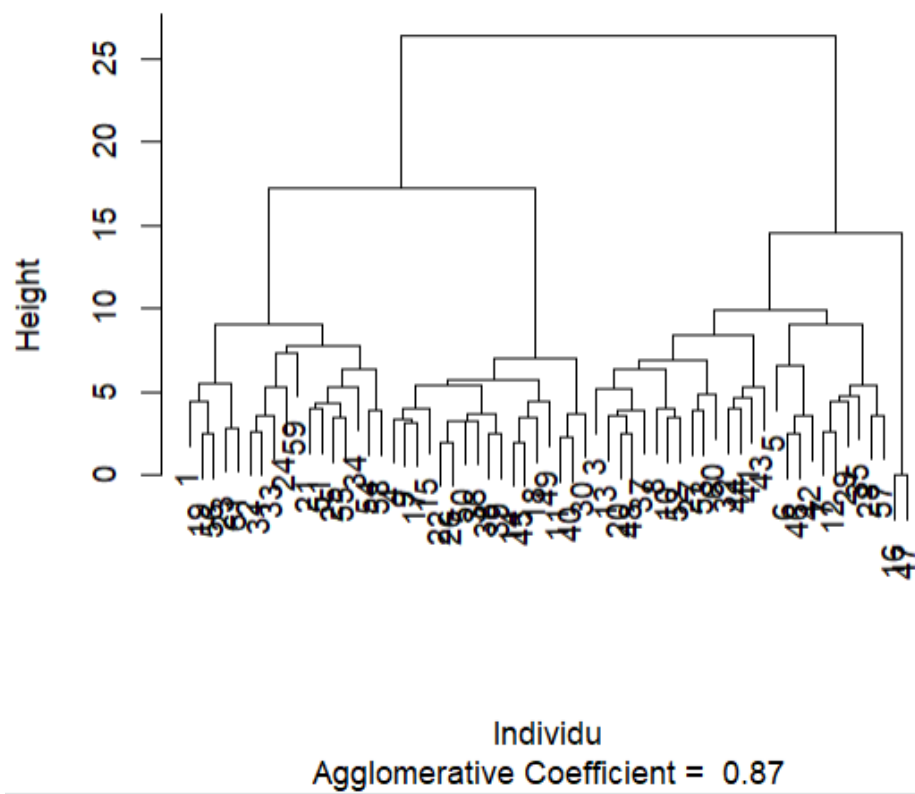


Figure 17: Courbe des hauteurs pour la CAH sur le Bloc 1, illustrant l'évolution des distances lors des fusions.

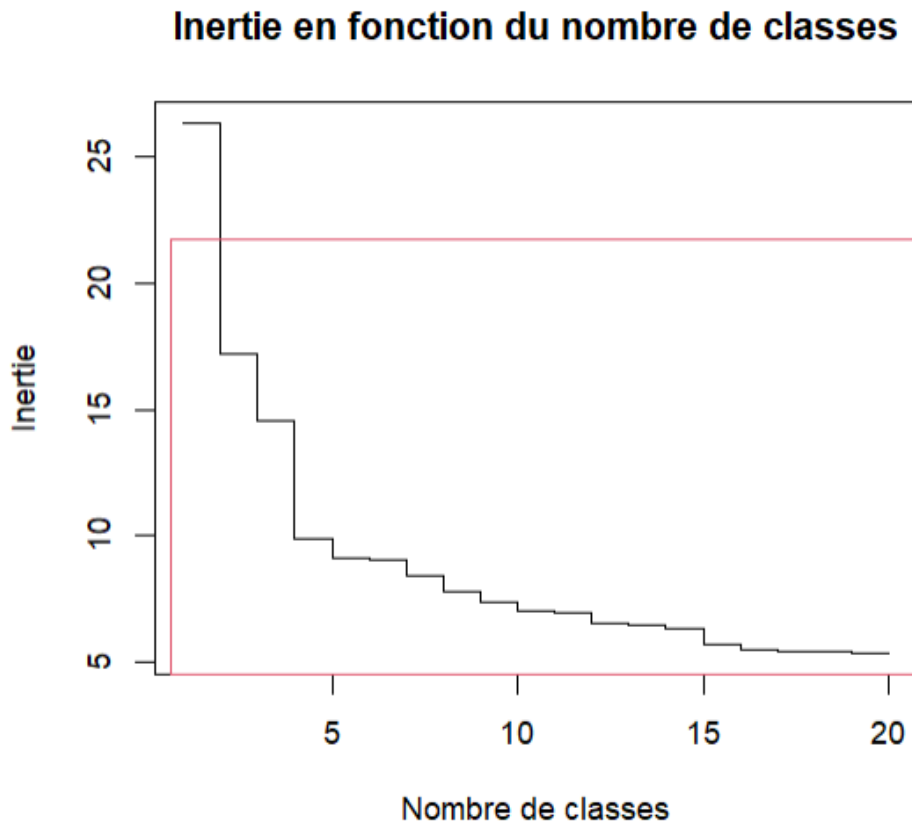


Figure 18: Courbe de l’inertie en fonction du nombre de classes pour la CAH sur le Bloc 1, montrant un coude à 2 classes.

6.4 Description des Classes (CAH)

Les deux classes ont été décrites à l’aide de la fonction `catdes`. La classe 1 (34 étudiants) est caractérisée par des valeurs faibles sur les variables liées à la préparation et aux compétences, comme `Maîtrise_compétences` (moyenne 2,12, v.test -6,59), `Confiance_entretiens` (moyenne 2,26, v.test -5,69), et `Préparé_entretiens` (moyenne 2,03, v.test -5,52). Ces étudiants rencontrent également plus souvent des contraintes de proximité (`Dur_près`, moyenne 0,82, v.test 2,26). La classe 2 (27 étudiants) regroupe des étudiants plus confiants et mieux préparés, avec des moyennes élevées pour `Maîtrise_compétences` (4,30, v.test 6,59), `Confiance_entretiens` (4,00, v.test 5,69), et `Préparé_entretiens` (3,67, v.test 5,52). Ils sont moins affectés par les contraintes de proximité (`Dur_près`, moyenne 0,56, v.test -2,26).

6.5 Représentation des Classes sur le Plan Factoriel (CAH)

Les classes ont été projetées sur le plan factoriel de l’ACP (Figure 19). La classe 1 (en bleu) se positionne majoritairement sur le côté négatif de la première dimension (26,3% de variance), associée à des étudiants moins préparés et compétents (ex. étudiants 30, 50, 60). La classe 2 (en jaune) est située sur le côté positif, regroupant des étudiants mieux préparés (ex. étudiants 2, 10, 52). Les ellipses de confiance confirment une séparation nette entre les deux groupes.

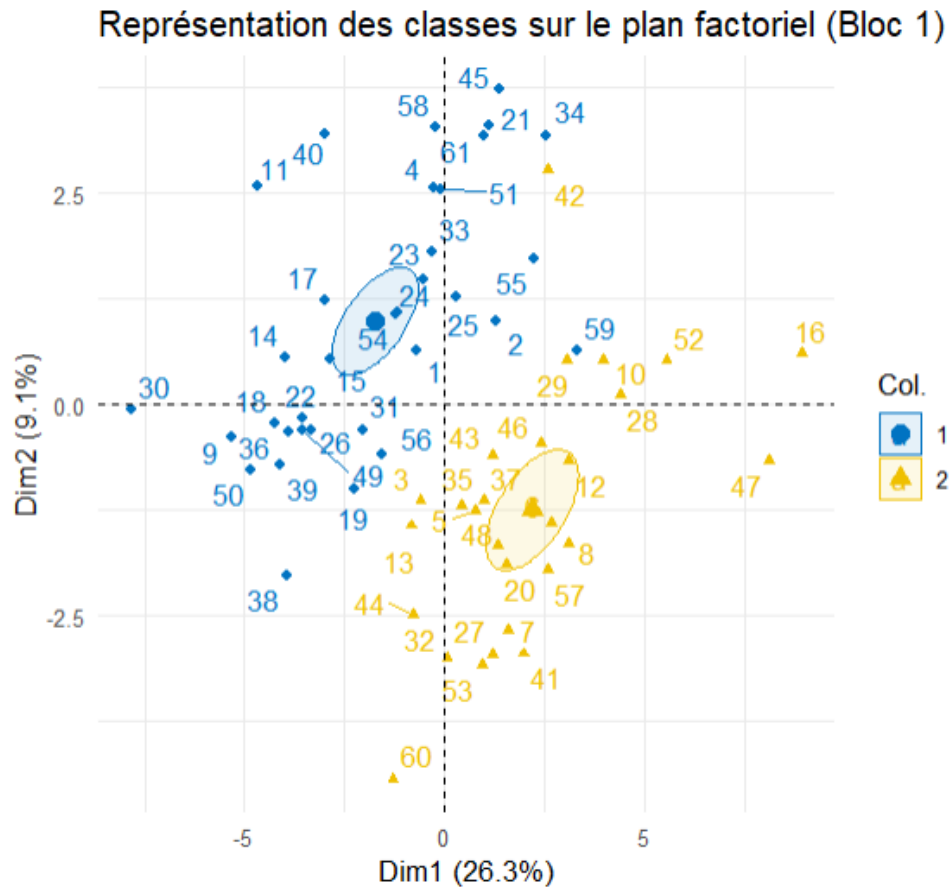


Figure 19: Représentation des classes CAH sur le plan factoriel de l'ACP (Bloc 1), avec ellipses de confiance.

7 Conclusion

Cette analyse a permis d'identifier les principaux défis rencontrés par les étudiants dans leur recherche de stages, notamment des lacunes en préparation et compétences (Bloc 1), un accès limité aux ressources et un soutien insuffisant (Bloc 2), ainsi que des contraintes pratiques comme la proximité et les frais (Bloc 3). Les analyses en composantes principales (ACP) et des correspondances multiples (ACM) ont révélé des profils variés, tandis que la classification hiérarchique a distingué des groupes d'étudiants selon leur niveau de préparation et leurs contraintes. Ces résultats suggèrent la nécessité d'améliorer l'accompagnement académique, d'optimiser les plateformes de recherche de stages, et de proposer des solutions pour réduire les obstacles financiers et géographiques. Des études complémentaires pourraient explorer les différences selon les formations ou les années d'études pour affiner ces recommandations.