

Contenus médiatiques et pratiques d'information des Français

Réalisé par

Carfantan Adrien & Hardy Pierre

Université de Tours
Master Économie de l'Entreprises et des Marchés



Année universitaire 2025–2026

Table des matières

1	Introduction	2
1.1	Enjeux et objectifs du projet	2
1.2	Présentation des bases de données	2
2	Partie I – ACP	4
2.1	4
2.2	4
3	Partie II – AFC	8
3.1	Objectif et validation statistique	8
3.2	Choix des axes et potentielle hypothèse	8
3.3	Interprétation du plan factoriel	9
3.4	Conclusion intermédiaire	11
4	Partie III – ACM	12
4.1	Présentation du cadre et des variables utilisés	12
4.2	Résultats factoriels de l'acm	12
4.2.1	Etude des inerties et des axes factoriels	12
4.2.2	Structure du nuage des individus	13
4.2.3	Organisation des modalités dans l'espace factoriel	15
4.2.4	Rôles des variables supplémentaires	17
4.3	Lecture synthétique et limites de l'ACM	19
5	Partie IV – Clustering	20
5.1	Détermination du nombre optimal de cluster	20
5.2	Visualisation du cluster	22
6	Conclusion générale	23
7	Annexe	24
8	Bibliographie	24

1 Introduction

1.1 Enjeux et objectifs du projet

Ce projet a été réalisé dans le cadre du cours de Data Mining, l'enjeu principal de celui ci réside dans la nature même des données étudiées, caractérisées par un nombre élevé de variables et par une forte hétérogénéité de l'information. L'analyse directe de ces données s'avère peu lisible, ce qui rend nécessaire le recours à des méthodes statistiques capables de synthétiser l'information tout en conservant les structures essentielles. L'objectif est ainsi de faire apparaître des régularités, des oppositions et des proximités entre les variables et les individus.

Le projet poursuit plusieurs objectifs. Il s'agit d'abord de mettre en évidence les grandes dimensions qui structurent la couverture thématique de l'information télévisée, afin de mieux comprendre l'organisation des contenus diffusés dans les journaux télévisés. Il vise également à analyser les pratiques d'information des individus, en étudiant les relations entre leurs caractéristiques sociales, leur intérêt pour l'actualité, leurs sources d'information et leur perception de la fiabilité des médias.

Pour répondre à ces objectifs, différentes méthodes d'analyse multivariée ont été mobilisées. Une Analyse en Composantes Principales a d'abord été appliquée afin d'identifier les principaux axes organisant les thématiques de l'information. Une Analyse Factorielle des Correspondances a ensuite permis d'étudier les liens entre les catégories d'âge et l'usage des réseaux sociaux. Enfin, une Analyse des Correspondances Multiples a été réalisée afin d'analyser conjointement plusieurs variables qualitatives et de dégager des profils d'individus en fonction de leurs pratiques et perceptions de l'information.

Cette approche est complétée par une classification des individus fondée sur les résultats de l'ACM, dont l'objectif est de regrouper les répondants en classes homogènes et interprétables. L'ensemble de ces analyses permet ainsi de proposer une lecture synthétique des données et d'apporter des éléments de compréhension sur la diversité des comportements face à l'information et aux médias dans le contexte français contemporain.

1.2 Présentation des bases de données

Le travail réalisé dans ce projet repose sur l'exploitation de deux bases de données distinctes, issues d'organismes publics français, et portant toutes deux sur la thématique de l'information et des médias. Bien que différentes par leur nature et leur structure, ces bases sont complémentaires et permettent d'aborder le sujet sous deux angles : celui de la production de l'information et celui de sa réception par le public.

La première base de données est fournie par l'Institut National de l'Audiovisuel (INA). Elle recense, sur une longue période, les sujets traités dans les journaux télévisés français, en distinguant plusieurs thématiques telles que la politique, l'économie, la santé, la culture, le sport ou encore les faits divers. Chaque observation correspond à un couple date-chaîne et indique le nombre de sujets diffusés pour chaque thème. Cette base est essentiellement quantitative et se prête naturellement à une Analyse en Composantes Principales, dont l'objectif est de dégager les principales dimensions structurant la couverture médiatique des journaux télévisés.

La seconde base de données provient d'une enquête menée par l'ARCOM intitulée « Les Français et l'information ». Elle repose sur des données d'enquête collectées auprès d'un large échantillon de la population française et contient principalement des variables qualitatives. Ces variables décrivent à la fois des caractéristiques socio-démographiques (sexe, âge, niveau d'études, situation professionnelle) et des pratiques liées à l'information, telles que l'intérêt pour l'actualité, les sources utilisées pour se forger une opinion ou encore la perception de la fiabilité de l'information. La richesse et la diversité de ces variables rendent cette base particulièrement adaptée à des méthodes d'analyse factorielle pour données qualitatives, notamment l'Analyse des Correspondances Multiples.

L'utilisation conjointe de ces deux bases de données permet d'adopter une approche globale du phénomène étudié. D'un côté, la base de l'INA offre une vision structurée des contenus informationnels diffusés par les médias audiovisuels ; de l'autre, la base de l'ARCOM permet d'analyser la manière dont ces contenus sont

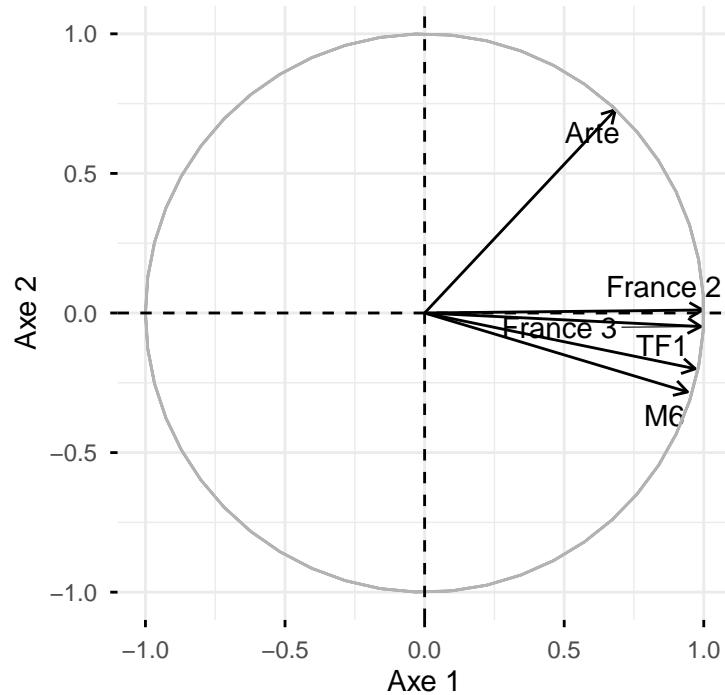
perçus, consommés et évalués par les individus. Ce double regard constitue un atout majeur du projet et justifie le recours à des méthodes statistiques variées afin d'exploiter au mieux l'information contenue dans les données.

2 Partie I – ACP

2.1

2.2

Projection des variables sur le plan factoriel (ACP)



Projection des individus sur le plan factoriel (ACP)

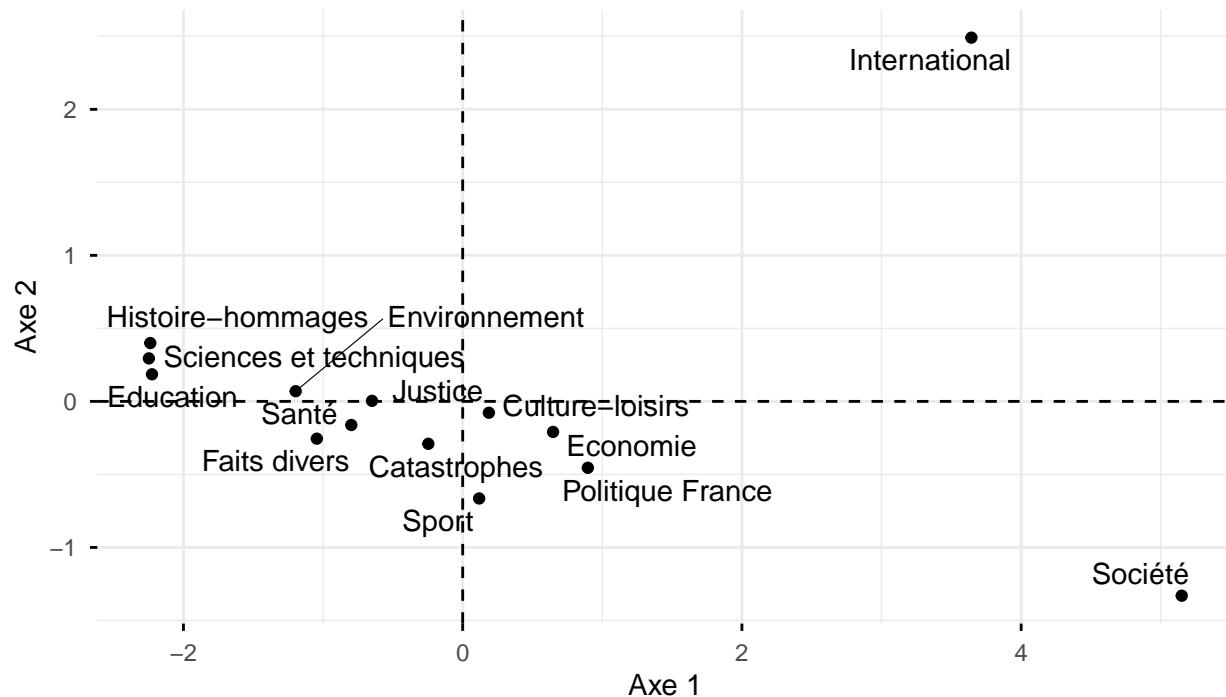
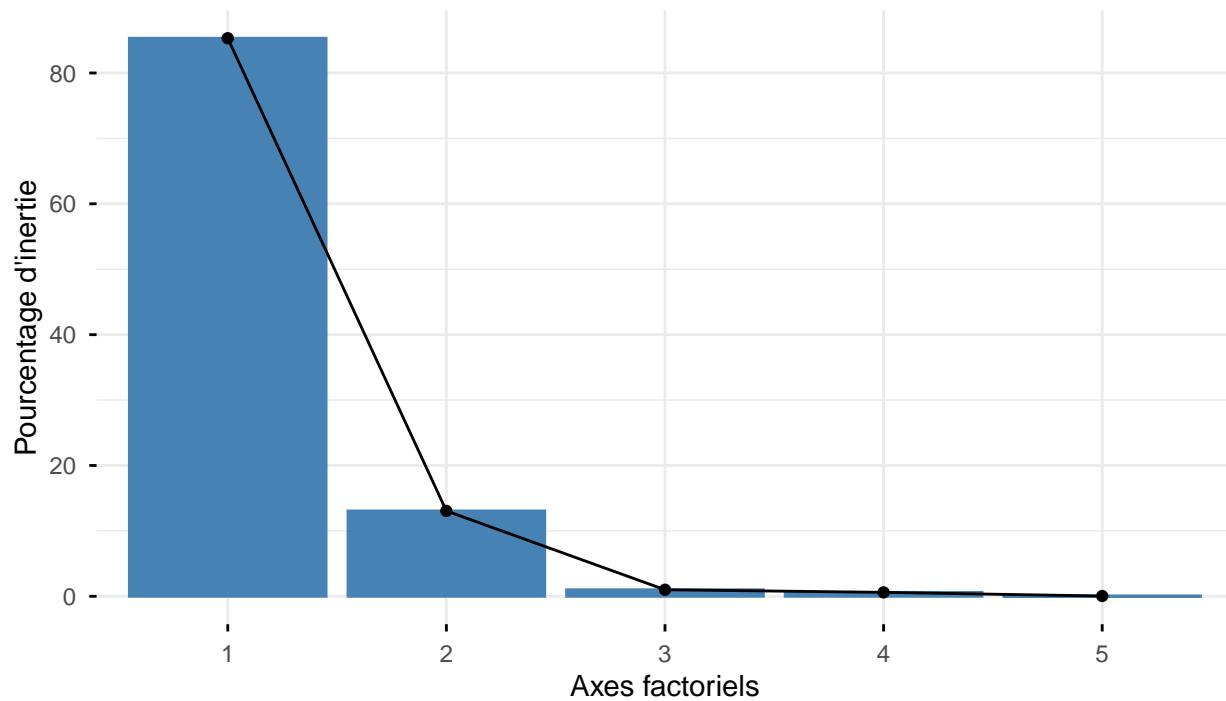
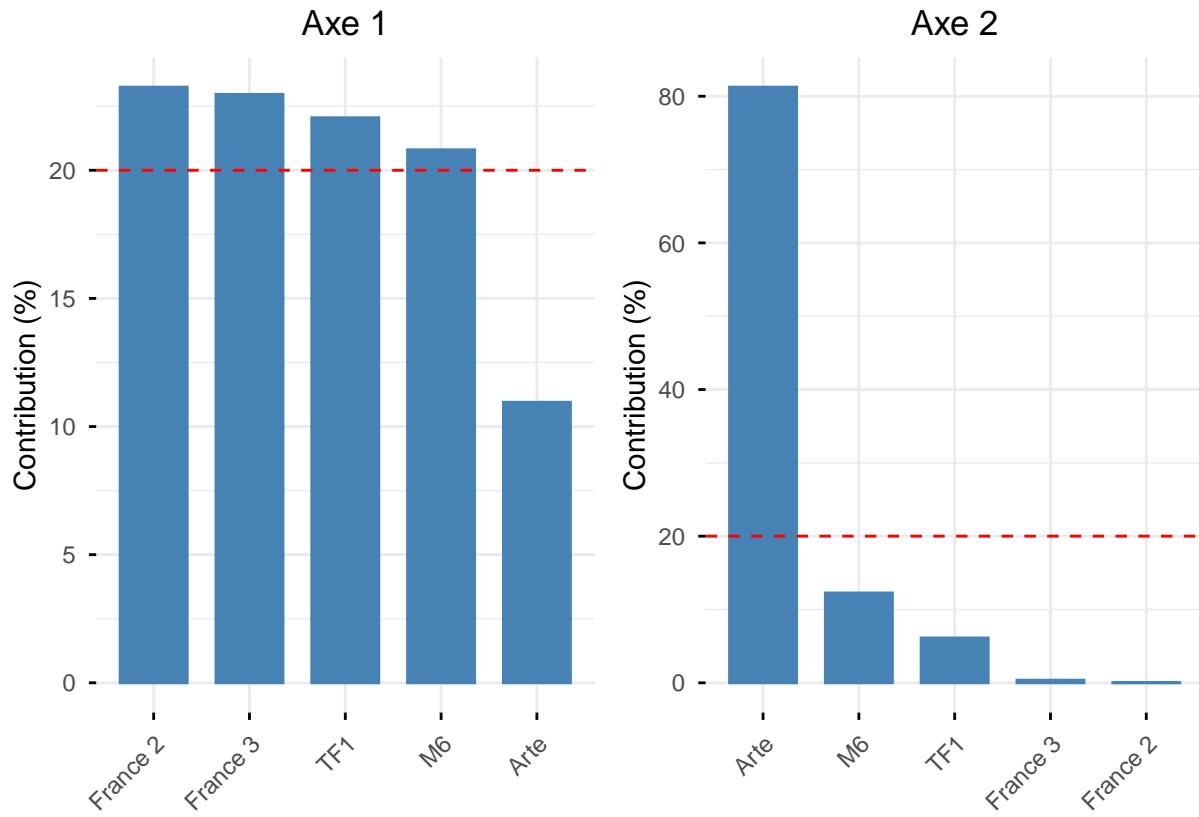
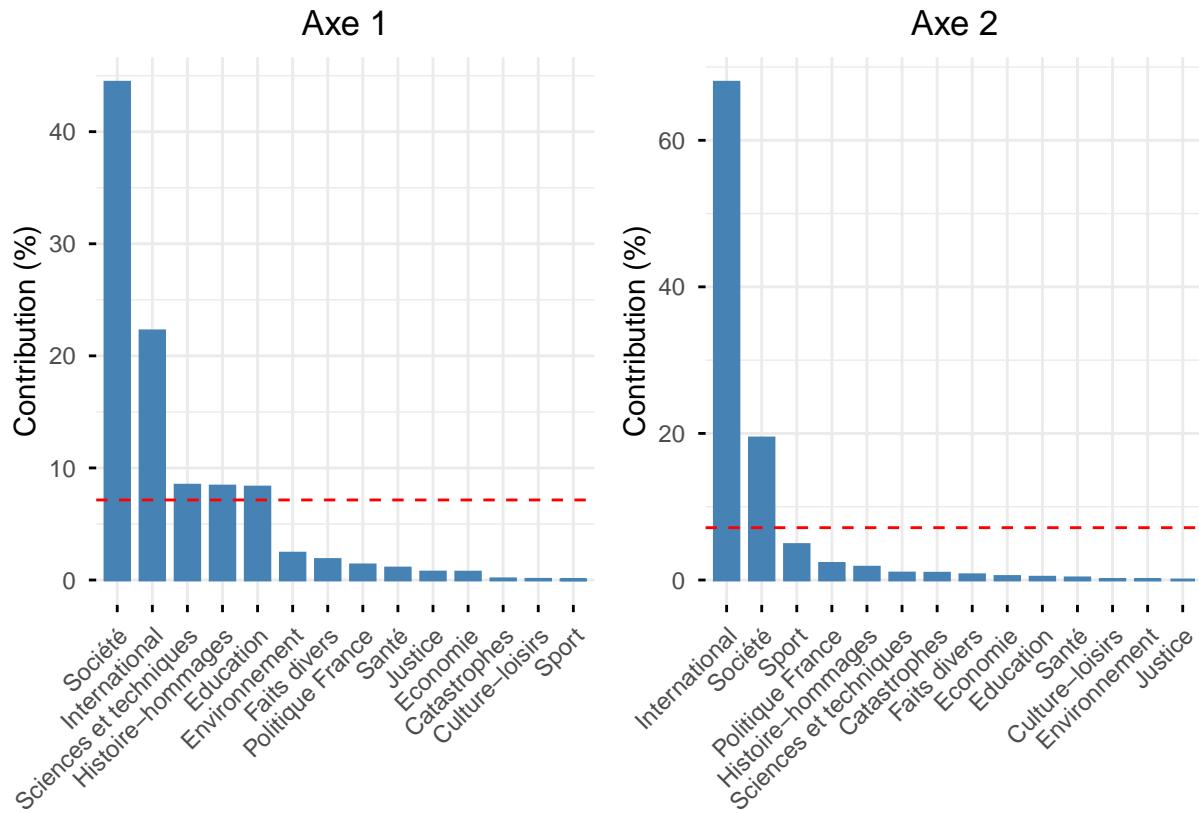


Diagramme des valeurs propres de l'ACP







3 Partie II – AFC

3.1 Objectif et validation statistique

Afin de compléter l'analyse précédente, cette partie s'intéresse aux liens entre les catégories d'âge et le réseau social le plus utilisé pour ce faire nous allons utiliser la base de données de l'ARCOM. L'analyse factorielle des correspondances (AFC) constitue un outil particulièrement adapté, puisqu'elle permet d'analyser de manière synthétique et graphique la relation entre deux variables qualitatives : les classes d'âge et le réseau social le plus utilisé.

Cette analyse représente une première étape dans l'étude des pratiques d'information des individus. Elle vise à identifier l'existence de logiques générationnelles dans l'usage des réseaux sociaux, qui seront ensuite approfondies dans la Partie III à l'aide d'une Analyse des Correspondances Multiples, intégrant un ensemble plus large de variables qualitatives.

Avant toute interprétation, un test du khi-deux d'indépendance est réalisé.

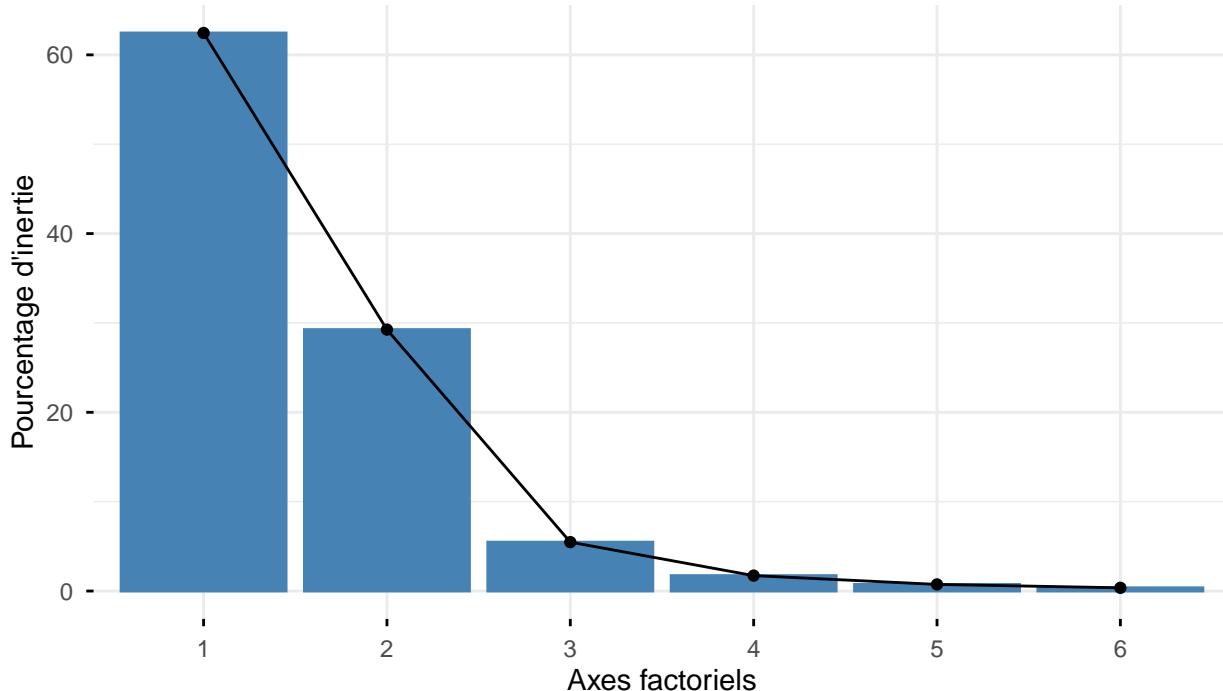
TABLE 1 – Résultats du test du khi-deux

	Statistique_chi2	Degres_de_liberte	p.value
X-squared	679.14	48	2.3e-112

Les résultats du test montrent une statistique de $\chi^2 = 679.14$ pour 48 degrés de liberté, avec une p-value inférieure à $2,3 \times 10^{-112}$. L'hypothèse d'indépendance est donc rejetée, ce qui confirme l'existence d'un lien statistiquement significatif entre l'âge et l'usage des réseaux sociaux, justifiant le recours à l'AFC.

3.2 Choix des axes et potentielle hypothese

Diagramme des valeurs propres de l'AFC

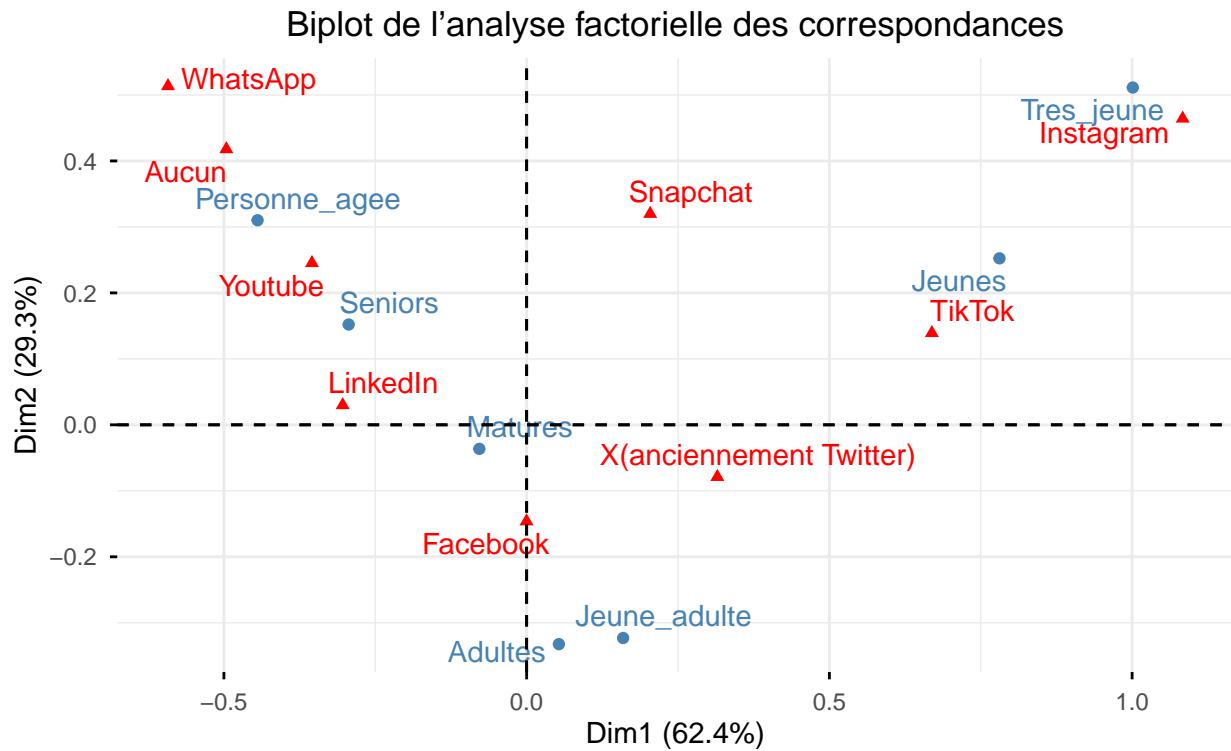


Avant de débuter nous devons rechercher les axes à étudié, l'analyse des valeurs propres indique que :
 l'axe 1 explique environ 62.45 % de l'inertie totale ;
 l'axe 2 explique environ 29.26 % de l'inertie totale.

Les deux premiers axes cumulent ainsi près de 91.71 % de l'inertie, ce qui assure une excellente qualité de représentation du phénomène étudié dans le plan factoriel (Dim 1, Dim 2). L'interprétation sera donc concentrée sur ces deux axes.

Avant d'approfondir l'étude, il est déjà possible de formuler une hypothèse. On peut s'attendre une forte différenciation générationnelle : les personnes les plus jeunes seront majoritairement associées à l'utilisation de réseau plus récent tel que Instagram ou TikTok, tandis que les personnes plus âgées seront associées à des plateformes de communication plus qu'à des réseaux comme WhatsApp soit à une absence d'usage des réseaux sociaux.

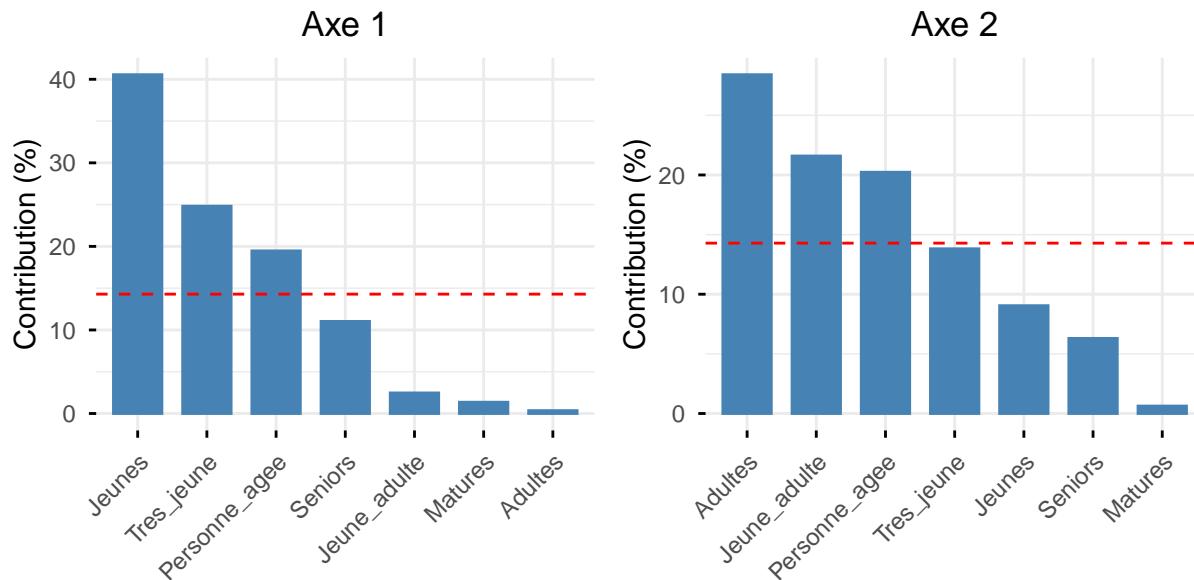
3.3 Interprétation du plan factoriel



Le premier axe met en évidence une opposition générationnelle marquée dans l'usage des réseaux. Les classes d'âge jeunes et très jeunes sont principalement associées à des plateformes récentes telles qu'Instagram, TikTok ou Snapchat, tandis que les classes d'âge les plus élevées se rapprochent davantage de réseaux comme WhatsApp, YouTube voire de l'absence d'utilisation de réseaux sociaux.

Cette opposition est particulièrement visible pour les très jeunes (15–17 ans), proches d'Instagram, et pour les jeunes (18–24 ans), associés à TikTok, alors que les personnes âgées de plus de 70 ans apparaissent fortement liées à la modalité « aucun réseau ».

Le second axe illustre des différences plus fines entre certaines classes d'âge intermédiaires. Il met en évidence que les variations ne se limitent pas à une opposition binaire entre jeunes et âgés : l'âge influence les préférences progressivement, avec des nuances subtiles qui ne deviennent perceptibles qu'en examinant l'ensemble des modalités. Les points proches de l'origine correspondent à des comportements moyens, c'est-à-dire à des groupes d'individus dont le réseau préféré n'est pas particulièrement différencié par rapport aux autres.

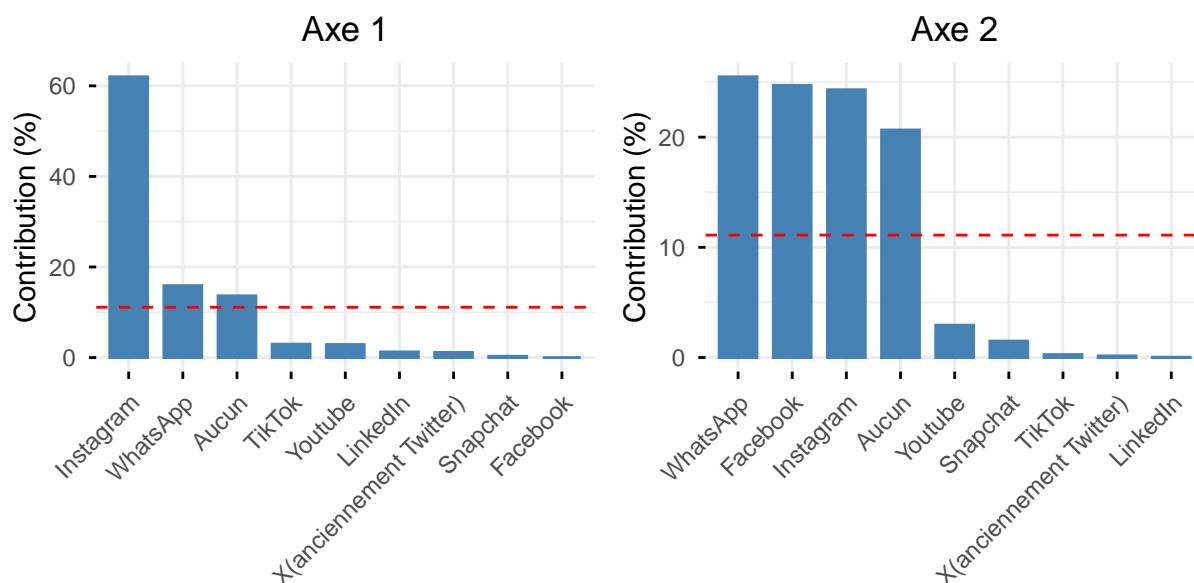


L'analyse des contributions montre que l'axe 1 est principalement structuré par les classes d'âge jeunes. Les catégories Jeunes, Très jeunes et Personnes âgées présentent des contributions nettement supérieures à la moyenne, avec une contribution particulièrement forte des Jeunes (environ 40 %). À l'inverse, les Adultes, Matures et Jeunes adultes contribuent très faiblement à cet axe.

L'axe 1 traduit donc une opposition forte entre les classes d'âge jeunes et le reste de la population, ce qui suggère des pratiques d'information très spécifiques chez les plus jeunes.

L'axe 2 est dominé par les catégories Adultes, Jeunes adultes et Personnes âgées, dont les contributions dépassent largement le seuil moyen. Les Adultes apparaissent comme la modalité la plus structurante de cet axe.

À l'inverse, les Jeunes et les Seniors contribuent moins fortement, tandis que les Matures jouent un rôle quasi marginal. Cet axe met en évidence une différenciation interne aux classes d'âge intermédiaires, plutôt qu'une opposition jeunes / âgés.



L'axe 1 est très largement structuré par Instagram, qui concentre plus de 60 % de la contribution totale. Cette contribution particulièrement élevée indique que cette plateforme est déterminante dans la différenciation des comportements observés sur ce premier axe.

Dans une moindre mesure, WhatsApp et la modalité Aucun contribuent également au-dessus du seuil moyen. Cela suggère une opposition entre des individus fortement utilisateurs d'Instagram et d'autres dont les usages reposent davantage sur des applications de messagerie ou sur une absence de recours aux réseaux sociaux.

Le second axe est principalement structuré par WhatsApp, Facebook, Instagram et la modalité Aucun, dont les contributions dépassent nettement la moyenne. À l'opposé, les plateformes comme YouTube, Snapchat, TikTok, LinkedIn et X contribuent très faiblement à cet axe. Elles jouent donc un rôle secondaire dans l'opposition mise en évidence par la deuxième dimension factorielle.

3.4 Conclusion intermédiaire

L'Analyse Factorielle des Correspondances met en évidence une structuration nette des individus selon l'âge et leur réseau social le plus utilisé. Elle montre que les méthodes factorielles permettent de résumer efficacement des données qualitatives complexes et de visualiser les relations entre les modalités des variables étudiées.

Cette analyse offre un cadre méthodologique solide pour explorer les liens entre caractéristiques démographiques et préférences en matière de réseaux sociaux, en identifiant les différences et similarités entre les groupes. Elle illustre également l'intérêt d'une approche statistique structurée pour dégager des tendances globales et comprendre la manière dont l'âge influence le choix des plateformes.

En conclusion, l'AFC confirme que l'âge constitue un facteur déterminant dans l'utilisation des réseaux sociaux et démontre l'efficacité des méthodes multivariées pour synthétiser et interpréter des données qualitatives hétérogènes.

4 Partie III – ACM

Dans cette partie, nous cherchons à comprendre comment les individus se différencient dans leur rapport à l'information, en fonction de leurs pratiques informationnelles et de leur perception de la fiabilité de l'information, tout en tenant compte de certaines caractéristiques sociodémographiques. L'objectif est d'observer les principales oppositions et similarités entre individus à partir de leurs réponses à l'enquête.

Pour répondre à cette question, nous utilisons une Analyse des Correspondances Multiples (ACM) appliquée aux données de l'enquête « Les Français et l'information » menée par l'ARCOM. Cette méthode permet d'analyser simultanément plusieurs variables qualitatives et de mettre en évidence les axes principaux qui structurent les pratiques et perceptions de l'information, ainsi que la position des individus dans cet espace.

4.1 Présentation du cadre et des variables utilisés

4.2 Résultats factoriels de l'acm

4.2.1 Etude des inerties et des axes factoriels

Représentation des parts d'inerties expliquées par les axes

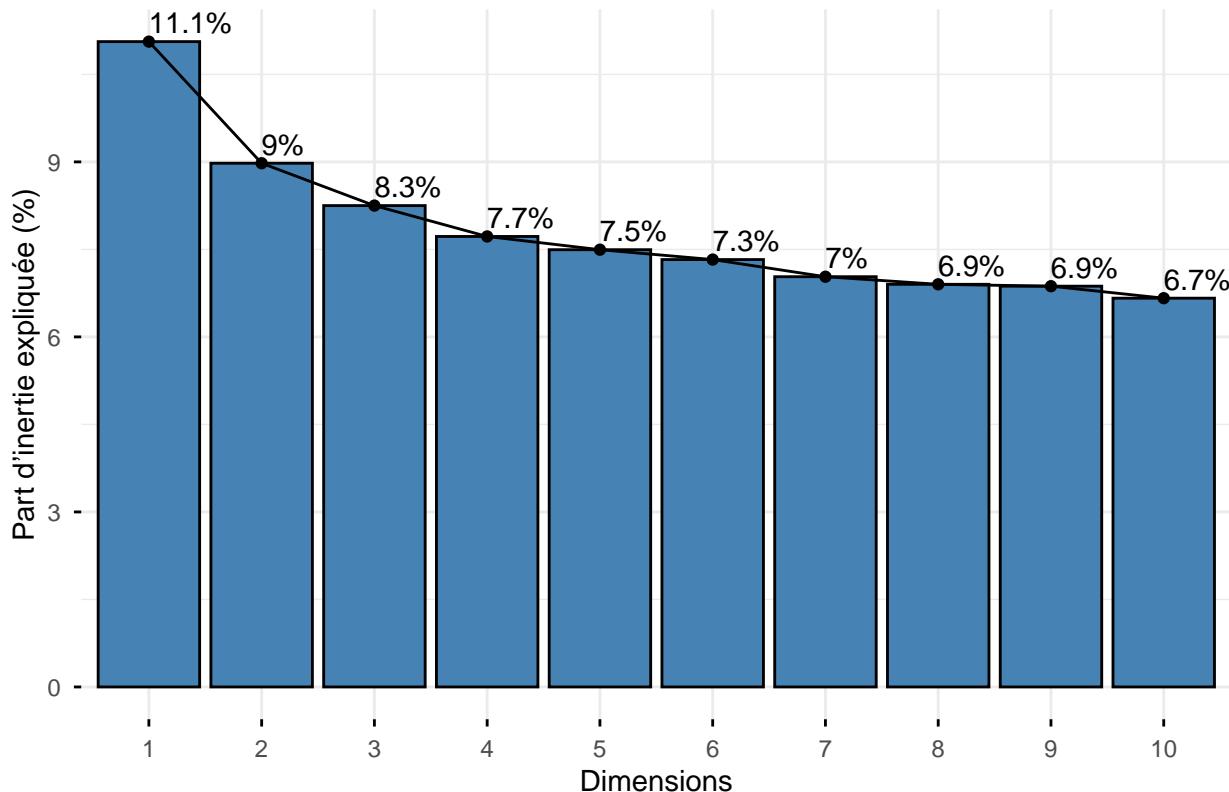
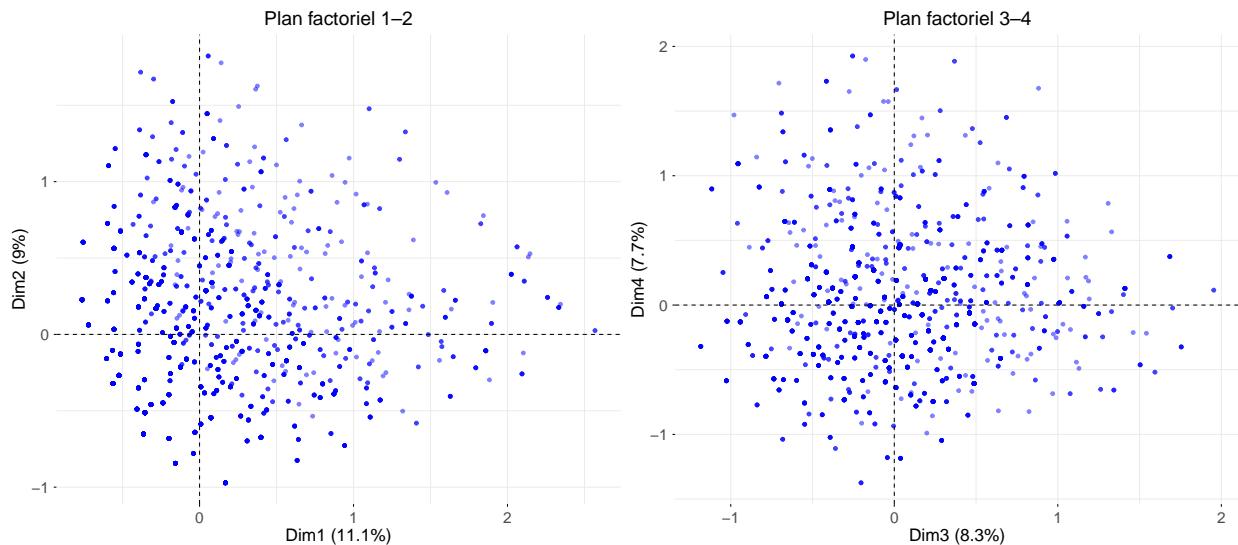


TABLE 2 – Valeurs propres et inertie cumulée de l'AFC

	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	dim 7	dim 8	dim 9	dim 10	dim 11	dim 12	dim 13	dim 14
eigenvalue	0.31	0.25	0.23	0.22	0.21	0.21	0.20	0.19	0.19	0.19	0.18	0.16	0.14	0.13
percentage of variance	11.06	8.98	8.25	7.72	7.49	7.33	7.03	6.90	6.87	6.66	6.45	5.76	4.89	4.61
cumulative percentage of variance	11.06	20.04	28.29	36.01	43.51	50.83	57.86	64.76	71.63	78.30	84.74	90.50	95.39	100.00

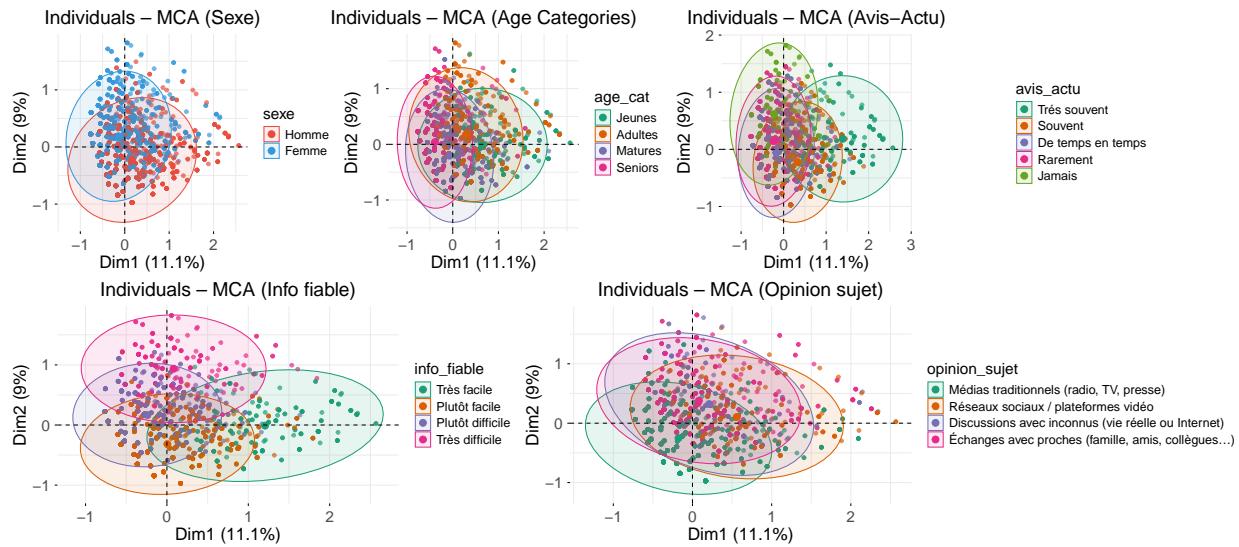
Les résultats montrent que la première dimension explique 11.06 % de l'inertie, tandis que la deuxième en explique 8.98 %. Les deux premiers axes cumulent ainsi environ 20.04 % de l'information totale, ce qui est classique en analyse des correspondances multiples, où l'information est souvent répartie sur plusieurs dimensions étant donné qu'il y a beaucoup de modalités. La décroissance est progressive, sans coude net, donc même si les deux premiers axes restent centraux pour la lecture globale, les dimensions suivantes présentent également des parts d'inertie non négligeables. Il apparaît donc pertinent de ne pas se limiter aux deux premiers axes et de prendre en compte d'autres dimensions afin d'affiner l'interprétation de la structure des données. Nous allons donc porter notre analyse sur les axes 1 et 2 mais également 3 et 4 pour ne pas manquer de capturer des informations dans notre analyse.

4.2.2 Structure du nuage des individus



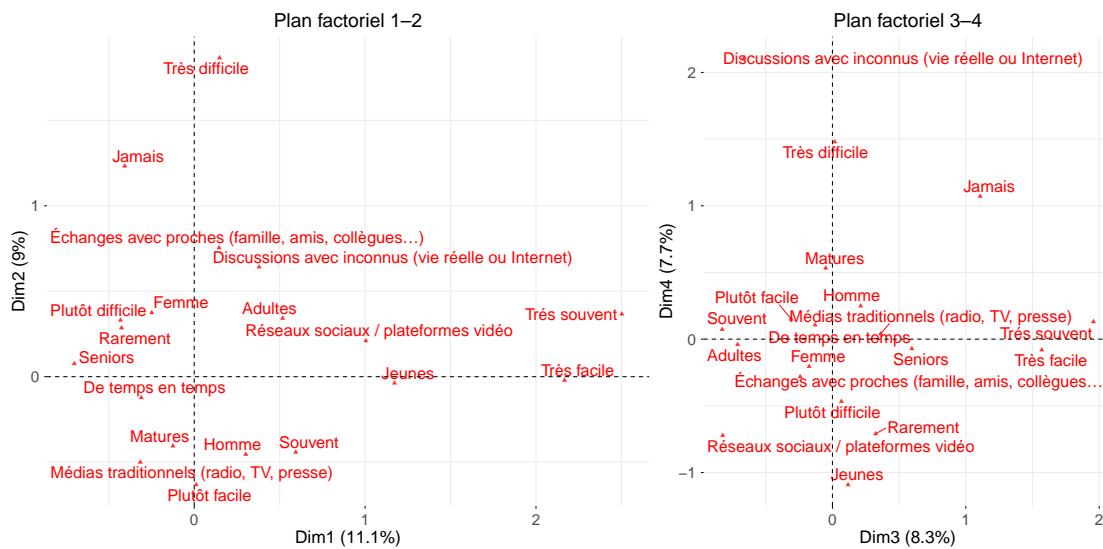
On observe un nuage de point relativement dense autour de l'origine, ce qui indique que la majorité des individus présentent des profils assez proches et similaires au regard des variables retenues. On ne distingue pas de groupes distincts pour l'instant. La dimension 1 montre une dispersion plus importante vers la droite, on remarque que certains individus sont clairement éloignés du centre de cette axe ce qui montre que nous avons des profils plus spécifiques. La dimension 2 nous montre une dispersion verticale plus limitée, cette dimension affine les profils sans créer de rupture franche entre les individus. Donc le nuage ne présente ni clusters évidents ni séparation franche, la structure propose une diversité progressive des profils plutôt qu'une segmentation nette. L'analyse des catégories et des variables supplémentaires est donc indispensable pour donner du sens au position des individus. La projection des individus sur les dimensions 3 et 4 ne fait pas apparaître de structuration nouvelle du nuage. Comme sur le premier plan factoriel, les individus restent majoritairement concentrés autour de l'origine, ce qui confirme que ces axes apportent des informations complémentaires sans remettre en cause l'organisation générale observée.

Code R pour les habillages avec ellipse de dimensions 1 et 2 par catégories (placé 3 graphiques sur une ligne et deux en dessous)



Les ellipses associées aux hommes et aux femmes se recouvrent largement, cela indique une forte proximité des profils informationnels selon le sexe. On observe toutefois une légère tendance des hommes à se projeter davantage du côté positif de la dimension 1. Les catégories d'âge présentent des positions différencierées sur la dimension 1, avec les jeunes et les adultes davantage situés à droite, tandis que les seniors sont plus concentrés à gauche. On observe malgré tout un recouvrement important des ellipses qui traduit une continuité des profils plutôt qu'une rupture générationnelle. Les individus ayant des avis contrastés sur l'actualité occupent des zones partiellement distinctes, notamment sur la dimension 1. Toutefois, l'interpénétration des groupes montre que le degré d'avis personnel ne suffit pas à structurer fortement l'espace des individus. Une opposition plus lisible apparaît entre les individus déclarant un accès « très facile » à une information fiable, plutôt situés à droite, et ceux déclarant un accès « difficile », davantage positionnés à gauche. On remarque également une distinction sur la dimension 2, on voit que les individus qui ont accès à une info fiable facilement sont plutôt situé en bas, alors que à l'inverse pour un accès difficile ils sont plutôt situé en haut. Pour notre dernier habillage on ne distingue pas tellement de séparation encore les ellipses sont partiellement superposés. Les représentations avec habillage confirment l'absence de segmentation nette du nuage des individus. Elles mettent toutefois en évidence des tendances de positionnement associées à l'âge, à l'accès à une information fiable et aux pratiques informationnelles, suggérant l'existence de profils différencier mais non strictement séparés.

4.2.3 Organisation des modalités dans l'espace factoriel

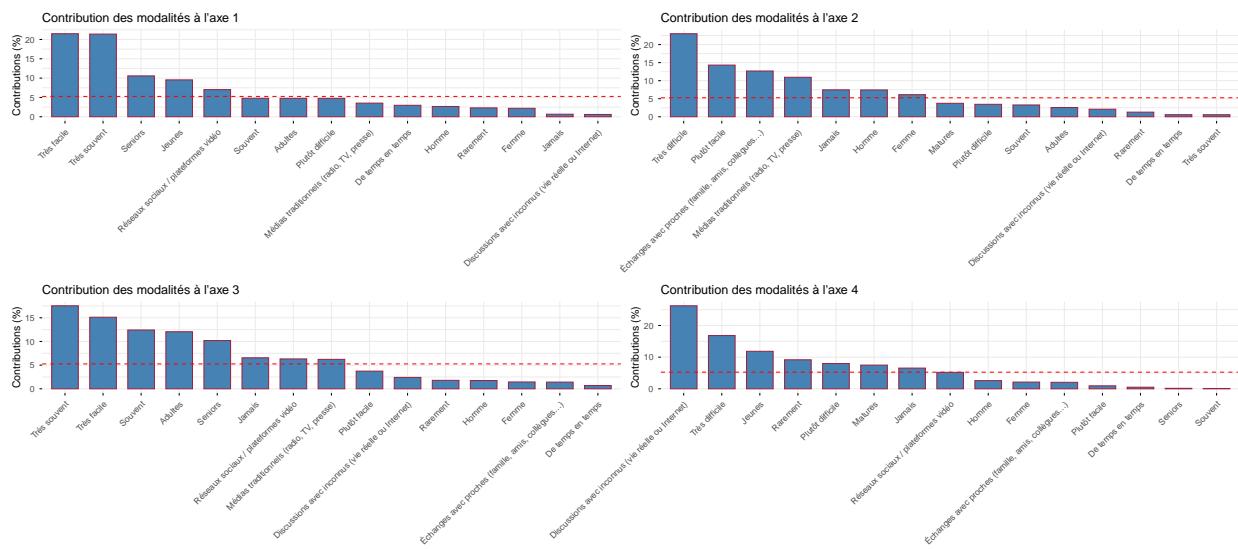


L'analyse des modalités sur les premiers plans factoriels permet de mieux comprendre la manière dont les individus se positionnent face à l'information et aux sources qu'ils mobilisent pour se forger une opinion.

Sur le plan formé par les dimensions 1 et 2, la dimension 1 oppose clairement deux grands profils. D'un côté, on observe les individus qui déclarent qu'il est très facile d'accéder à une information fiable et qui s'appuient davantage sur les réseaux sociaux et plateformes vidéo pour construire leur opinion. Ces modalités sont principalement associées aux jeunes, ce qui traduit une plus grande aisance avec les supports numériques et une perception plus fluide de l'accès à l'information. À l'inverse, l'autre côté de l'axe regroupe davantage les seniors et les matures, ainsi que le recours aux médias traditionnels (radio, télévision, presse). Ces modalités sont associées à une perception plus nuancée, voire plus difficile, de l'accès à l'information fiable, suggérant un rapport plus prudent et institutionnel à l'actualité.

La dimension 2 apporte un éclairage complémentaire en mettant en avant la question de la difficulté perçue dans l'accès à une information fiable. Les modalités « Très difficile » et le recours aux échanges avec les proches (famille, amis, collègues) apparaissent proches, indiquant que lorsque l'information est perçue comme complexe ou incertaine, les individus ont tendance à s'appuyer davantage sur leur entourage pour se forger une opinion. À l'inverse, les discussions avec des inconnus, dans la vie réelle ou sur Internet, se distinguent, traduisant une autre manière de confronter les points de vue et l'information.

Les dimensions 3 et 4, bien que secondaires, confirment ces tendances générales. La dimension 3 renforce notamment le clivage générationnel déjà observé, avec une opposition entre les jeunes, plus proches des réseaux sociaux, et les profils plus âgés, davantage associés aux médias traditionnels. La dimension 4 met en évidence des modalités plus spécifiques, comme la perception d'un accès très difficile à l'information ou le recours aux discussions avec des inconnus, qui correspondent à des profils moins centraux mais clairement identifiables.



La dimension 1 est principalement structurée par la modalité « Très facile », qui contribue à hauteur d'environ 21,5 % à la construction de l'axe. Cette modalité est fortement liée au fait de s'informer à partir des grands médias (radio, télévision, presse), lesquels représentent eux aussi une contribution importante (21,4 %). À eux deux, ces éléments expliquent plus de 42 % de l'inertie de l'axe, ce qui indique que cette première dimension reflète avant tout un rapport simple et institutionnalisé à l'information, perçue comme facilement accessible pour se forger une opinion.

On observe ensuite l'intervention de facteurs générationnels, avec une contribution non négligeable des seniors (10,6 %) et des jeunes (9,5 %), suggérant des pratiques d'information différenciées selon l'âge. Enfin, les réseaux sociaux et plateformes vidéo contribuent également à la structuration de cette dimension (7,0 %), ce qui met en évidence une opposition progressive entre les sources d'information traditionnelles et les sources numériques dans la construction de l'opinion. La dimension 2 est principalement structurée par la modalité « Très difficile », qui représente à elle seule environ 22 % de la contribution totale à l'axe. Elle est suivie par « Plutôt facile » (14 %) et par les échanges avec les proches (famille, amis, collègues) (13 %). Ces premières modalités montrent que cet axe oppose surtout des perceptions contrastées de la facilité d'accès à une information fiable, en lien avec des sources plus personnelles ou informelles.

On observe également une contribution notable des médias traditionnels (radio, télévision, presse) (11 %), ainsi que des différences selon le sexe (hommes 8 %, femmes 6 %) et certains groupes d'âge comme les matures. Ainsi, la dimension 2 semble traduire un axe de perception subjective de la fiabilité de l'information, opposant un sentiment de difficulté à s'informer à des pratiques reposant davantage sur l'entourage ou sur des sources jugées plus accessibles.

La dimension 3 est principalement structurée par la modalité « Très facile », qui contribue à environ 17 % à la construction de l'axe. Elle est suivie par les adultes (15 %) et les seniors (12 %), ce qui montre que cet axe met fortement en avant un profil d'individus pour lesquels l'accès à l'information est perçu comme simple, en particulier chez les classes d'âge plus élevées.

On observe également une contribution non négligeable des médias traditionnels (radio, télévision, presse) (6 %) et des réseaux sociaux / plateformes vidéo (6 %), ce qui suggère que la dimension 3 distingue moins les types de sources que les profils sociodémographiques associés à un sentiment de facilité ou de familiarité avec l'information. Cette dimension apparaît donc comme un axe secondaire, mais pertinent, qui affine la lecture des pratiques informationnelles en complétant les oppositions mises en évidence par les deux premières dimensions.

La dimension 4 est principalement structurée par la modalité « Discussions avec des inconnus (vie réelle ou Internet) », qui représente environ 26 % de la contribution totale à l'axe. Elle est suivie par la modalité « Très difficile » (17 %) et par les jeunes (12 %), ce qui indique que cet axe renvoie à une forme de

distance ou de difficulté perçue dans l'accès à l'information, associée à des pratiques informationnelles moins institutionnelles et plus incertaines.

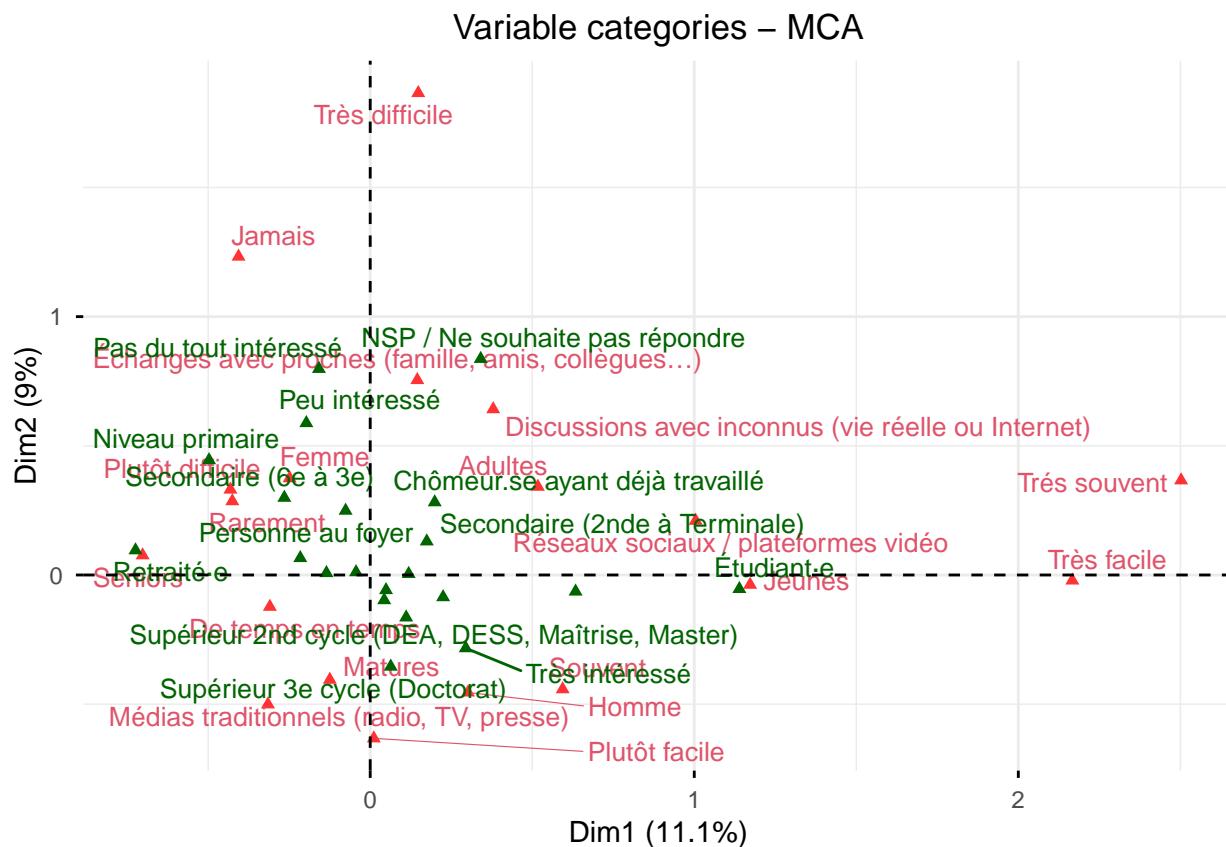
Les modalités « Plutôt difficile » (8 %), les matures (7 %) et le recours aux réseaux sociaux / plateformes vidéo (5 %) viennent compléter la structuration de cet axe. À l'inverse, les modalités liées à un accès jugé facile ou aux échanges avec les proches contribuent très faiblement, suggérant que cette dimension oppose des pratiques d'information complexes et indirectes à des formes plus simples et familières, déjà bien captées par les axes précédents.

L'analyse des modalités dans l'espace factoriel permet de répondre directement à la problématique du projet, en montrant que la manière dont les individus se forgent une opinion dépend à la fois de l'accès perçu à l'information et des sources mobilisées. Les premières dimensions de l'ACM sont principalement structurées par des modalités liées à la facilité ou à la difficulté à accéder à une information fiable, ce qui souligne que cette perception joue un rôle central dans les pratiques informationnelles.

Les résultats montrent également que les types de sources ne sont pas neutres dans la construction de l'opinion. Les médias traditionnels, les réseaux sociaux et les échanges interpersonnels s'organisent différemment dans l'espace factoriel, traduisant des façons distinctes de s'informer. Ces choix de sources sont par ailleurs étroitement liés aux caractéristiques des individus, notamment à l'âge, ce qui met en évidence l'existence de profils informationnels différenciés selon les générations.

4.2.4 Rôles des variables supplémentaires

Les variables supplémentaires permettent de compléter l'interprétation de l'ACM sans intervenir dans la construction des axes. Elles apportent un éclairage utile pour mieux comprendre les profils mis en évidence précédemment.



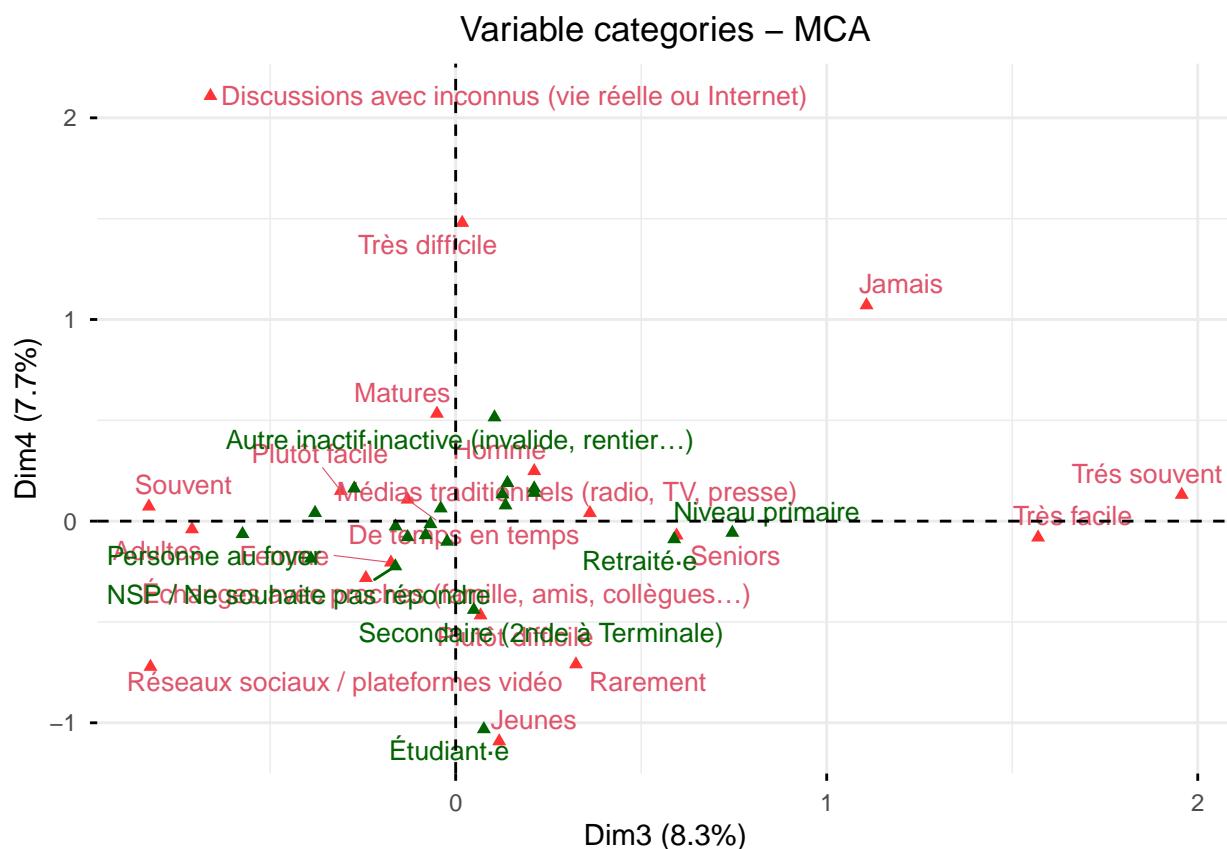
Sur la partie droite de la dimension 1, on observe une forte proximité entre la modalité « Très facile », les

jeunes, les étudiants et les individus se déclarant très intéressés par l'actualité. Ce regroupement suggère un profil d'individus pour lesquels l'accès à l'information est perçu comme simple, fluide et naturel. Ces individus sont également proches des réseaux sociaux et plateformes vidéo, ce qui indique que le numérique joue un rôle central dans leur manière de s'informer et de construire une opinion. Il existe donc ici une corrélation claire entre jeunesse, intérêt pour l'actualité, usage des supports numériques et facilité perçue à s'informer.

À l'opposé, sur la gauche de l'axe, se regroupent les modalités « Plutôt difficile » et « Très difficile », associées davantage aux seniors, aux personnes inactives (retraités, personnes au foyer) et à des niveaux d'études plus faibles. Ce pôle traduit un rapport plus distant à l'information, marqué par une perception de complexité et par un moindre engagement. Les médias traditionnels apparaissent également plus proches de ce groupe, suggérant une pratique informationnelle plus classique, mais aussi potentiellement moins adaptée aux formes contemporaines de circulation de l'information.

La dimension 2 met davantage en évidence le degré d'intérêt pour l'actualité et les modes de socialisation de l'information. Les individus situés en haut de l'axe sont plus proches des modalités « Peu intéressé » ou « Pas du tout intéressé », ainsi que des discussions avec des inconnus, ce qui peut traduire une relation plus diffuse ou indirecte à l'information. À l'inverse, les individus situés plus bas apparaissent davantage liés aux échanges avec les proches et à un intérêt plus affirmé, suggérant une construction de l'opinion ancrée dans des cercles sociaux familiers.

Les variables supplémentaires ne créent pas de nouvelles oppositions, mais confirment et précisent la structure mise en évidence par les variables actives. Elles montrent que les pratiques d'information s'organisent selon des logiques générationnelles, scolaires et sociales, où l'âge, le niveau d'études et l'intérêt pour l'actualité jouent un rôle central.



La dimension 3 est principalement structurée par l'opposition entre les discussions avec des inconnus (dans la vie réelle ou sur Internet) et les médias traditionnels (radio, télévision, presse). Cet axe semble traduire

une distinction entre des formes d'accès à l'information institutionnelles et des formes plus relationnelles ou informelles.

Du côté des discussions avec des inconnus, on retrouve davantage les jeunes et les étudiants, souvent associés aux réseaux sociaux et plateformes vidéo. Ce regroupement suggère un rapport à l'information plus interactif, reposant sur l'échange, la confrontation des points de vue et la circulation horizontale des contenus. L'information ne se limite plus à une transmission descendante, mais s'inscrit dans des espaces de discussion élargis.

À l'inverse, les médias traditionnels apparaissent plus proches des seniors et des catégories plus éloignées des usages numériques. Cette position reflète des pratiques informationnelles plus stabilisées, fondées sur des sources reconnues et institutionnelles, mais potentiellement moins interactives.

La dimension 3 met donc en évidence une pluralité des modes de construction de l'opinion, qui dépasse la simple opposition numérique / traditionnel et renvoie à des manières différentes de s'approprier l'information.

La dimension 4 est davantage structurée par la perception subjective de la difficulté à se forger une opinion. On observe une opposition entre les modalités « Très facile » et « Très difficile », sans qu'elle ne corresponde à un clivage social aussi marqué que sur les axes précédents.

Les individus situés du côté de la facilité perçue sont plus proches des profils jeunes, étudiants et des usages numériques, ce qui suggère une plus grande aisance dans la navigation entre les sources et dans l'évaluation de l'information. À l'inverse, les modalités associées à la difficulté apparaissent davantage liées à des profils plus âgés ou moins insérés dans les pratiques numériques contemporaines.

Cependant, les recouvrements importants entre les modalités montrent que cette dimension renvoie avant tout à un ressenti individuel, influencé par l'expérience personnelle, le degré de familiarité avec l'information et les outils numériques, plutôt qu'à une opposition sociale nette.

Pris ensemble, les axes 3 et 4 ne bouleversent pas la structure globale révélée par l'ACM, mais en affinent la lecture. Ils montrent que, au-delà des grandes oppositions liées à l'âge, aux sources d'information et à l'intérêt pour l'actualité, existent des différences plus subtiles dans la manière dont les individus interagissent avec l'information et évaluent leur propre capacité à se forger une opinion.

Ces dimensions confirment ainsi l'idée d'un continuum de pratiques, allant d'un rapport actif, interactif et confiant à l'information, vers un rapport plus distant ou plus incertain, sans segmentation rigide de la population.

4.3 Lecture synthétique et limites de l'ACM

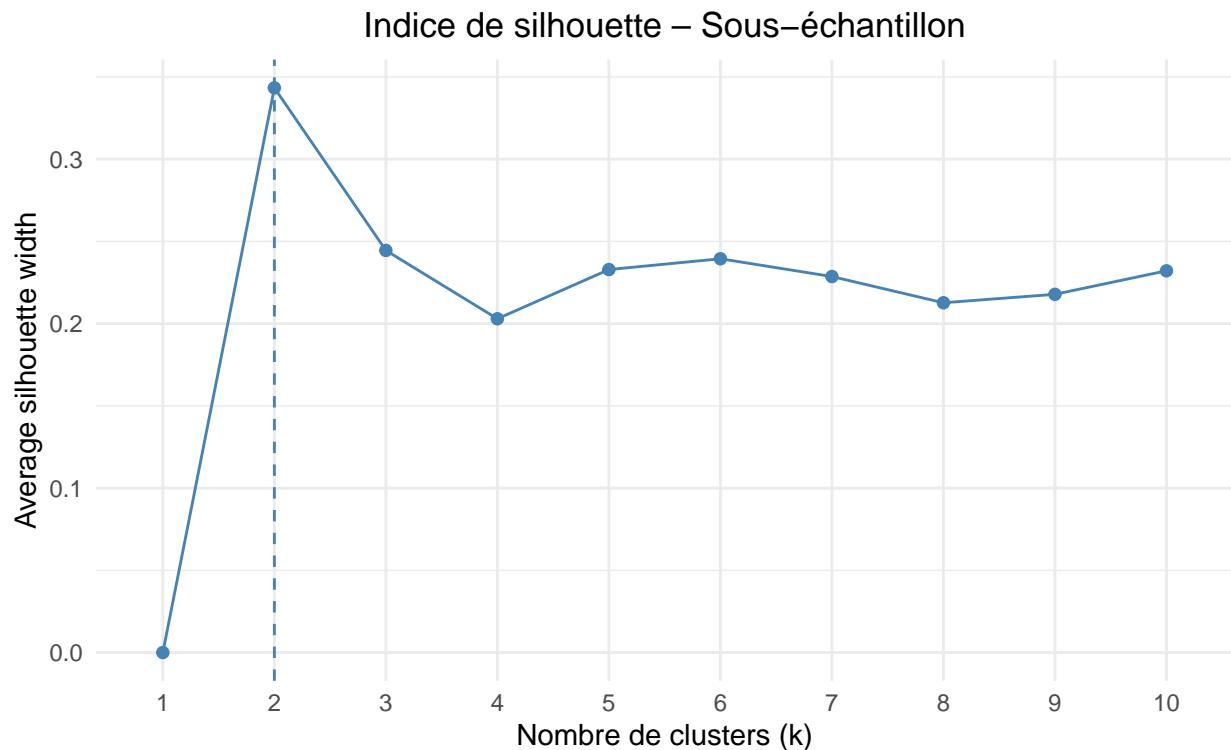
A mettre en annexe : Habillage dimension 3 et 4

5 Partie IV – Clustering

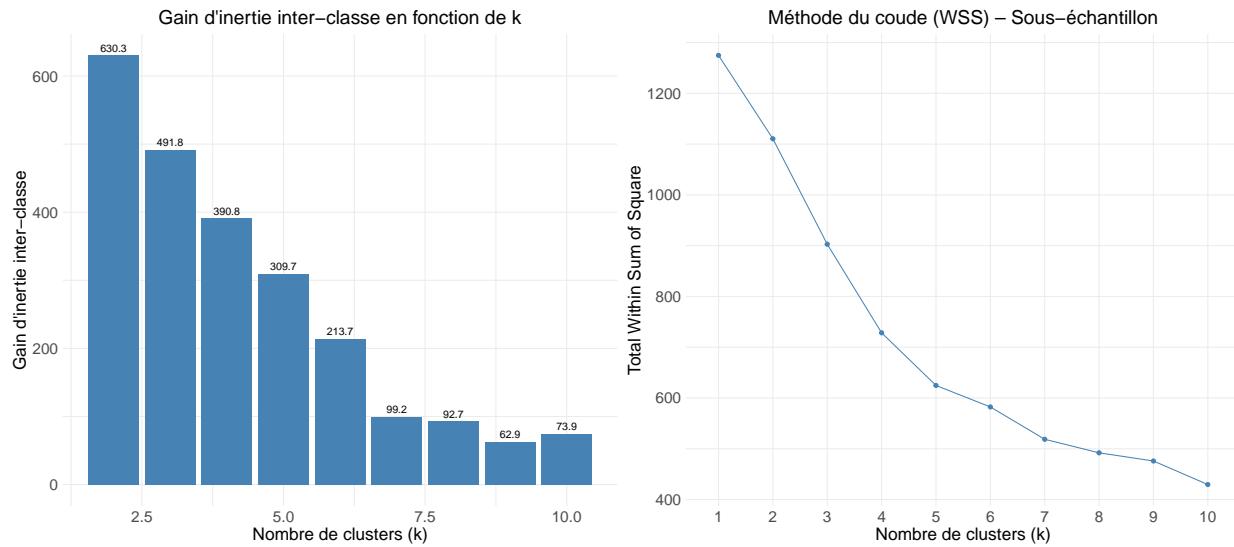
Afin de prolonger l'analyse factorielle des correspondances multiples, une classification des individus a été réalisée à partir des coordonnées factorielles issues de l'ACM. Cette démarche vise à identifier des groupes homogènes d'individus, partageant des profils similaires

5.1 Détermination du nombre optimal de cluster

Le choix du nombre de clusters constitue une étape centrale de la classification. Il repose ici sur une analyse conjointe de plusieurs critères complémentaires



L'indice de silhouette mesure la qualité du regroupement en comparant la proximité des individus à leur propre cluster par rapport aux autres clusters. Le graphique de l'indice de silhouette met en évidence un maximum pour $k = 2$, suggérant un découpage très naturel de la population en deux grands groupes. Toutefois, une telle partition apparaît trop grossière au regard des objectifs de l'étude, car elle ne permet pas de rendre compte de la diversité des profils observés dans les données.



Le graphique du gain d'inertie inter-classe permet d'évaluer l'amélioration apportée par l'augmentation du nombre de clusters. On observe une augmentation marquée du gain jusqu'à $k = 6$, puis un ralentissement net au-delà. Cela signifie qu'à partir de ce seuil, l'ajout de clusters supplémentaires apporte une amélioration marginale limitée en termes de structuration de l'information.

Ce résultat suggère l'existence d'un compromis optimal autour de six clusters, permettant de capter l'essentiel de l'hétérogénéité entre les individus sans sur-segmentation.

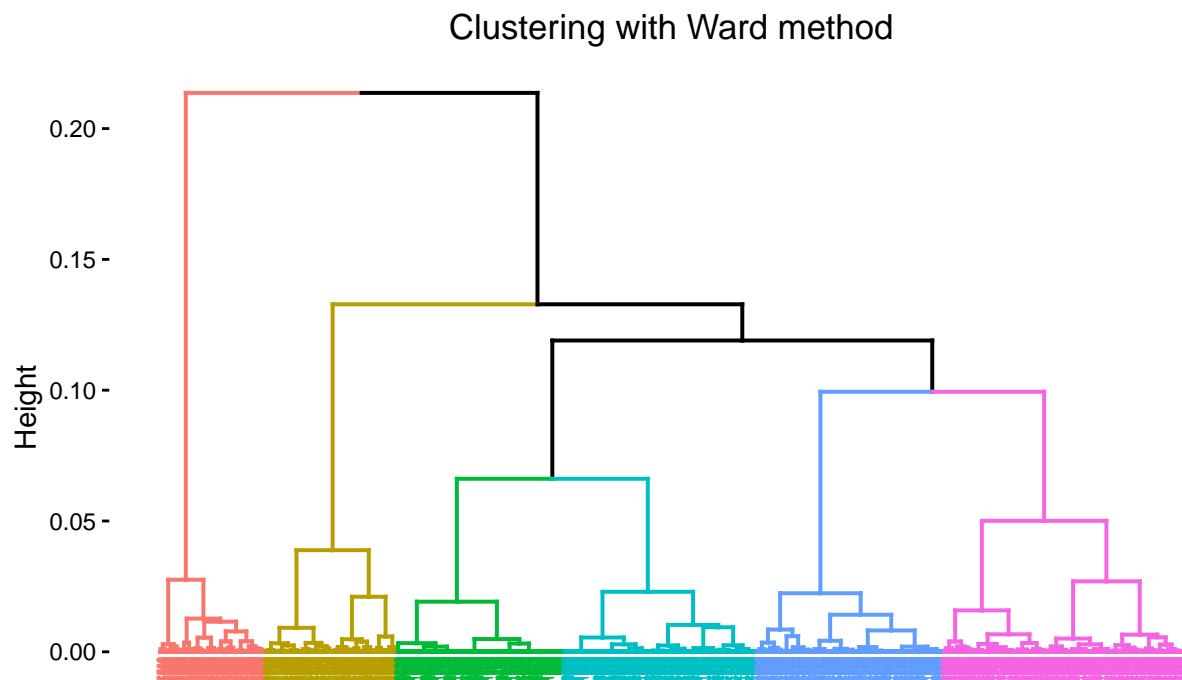
La méthode du coude, fondée sur la somme des carrés intra-classes (WSS), confirme cette analyse. La courbe présente un ralentissement lorsqu'on rajoute un cluster visible au niveau de $k = 6$, indiquant qu'au-delà de ce nombre de clusters, la réduction de l'inertie intra-classe devient faible. Ce point correspond à un bon équilibre entre homogénéité interne des groupes et lisibilité globale de la classification.

En croisant les résultats :

- de l'indice de silhouette,
- du gain d'inertie inter-classe,
- et de la méthode du coude (WSS),

le choix de six clusters apparaît comme le plus pertinent. Il permet de dépasser une opposition trop simplifiée en deux groupes tout en conservant une structure stable, interprétable et statistiquement robuste.

5.2 Visualisation du cluster

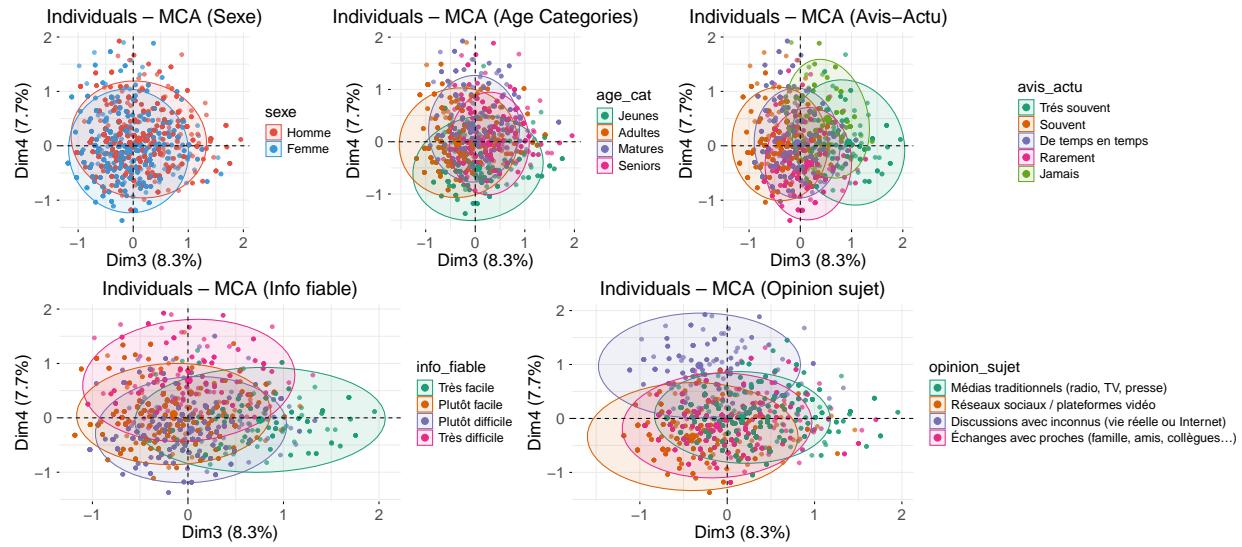


Le dendrogramme met en évidence six branches distinctes identifiées par des codes couleurs. La ligne de coupure horizontale suggérée par cette coloration se situe à une hauteur d'environ 0.05, niveau auquel les groupes présentent une homogénéité interne forte tout en restant bien différenciés les uns des autres.

On observe également que les six clusters possèdent des largeurs de base relativement comparables, ce qui suggère une répartition équilibrée de la population. Aucun groupe n'apparaît comme marginal ou démesurément large.

6 Conclusion générale

7 Annexe



8 Bibliographie

Autorité de Régulation de la Communication Audiovisuelle et Numérique (ARCOM). (2024). Les Français et l'information – 2024 [jeu de données]. <https://defis.data.gouv.fr/datasets/662bd340e32d60bee0ca828a>

Institut National de l'Audiovisuel (INA). (2024). Classement thématique des sujets de journaux télévisés (janvier 2000 - décembre 2020) [jeu de donnée]. <https://defis.data.gouv.fr/datasets/57fe1446c751df182f79df72>