

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	9
3.1	Objectif et validation statistique	9
3.2	Choix des axes et potentielle hypothese	9
3.3	Interprétation du plan factoriel	10
3.4	Conclusion intermédiaire	12
4	Partie III – ACM	13
4.1	13
4.2	13
5	Partie IV – Clustering	14
5.1	14
5.2	14
6	Conclusion générale	15
7	Bibliographie	16

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.

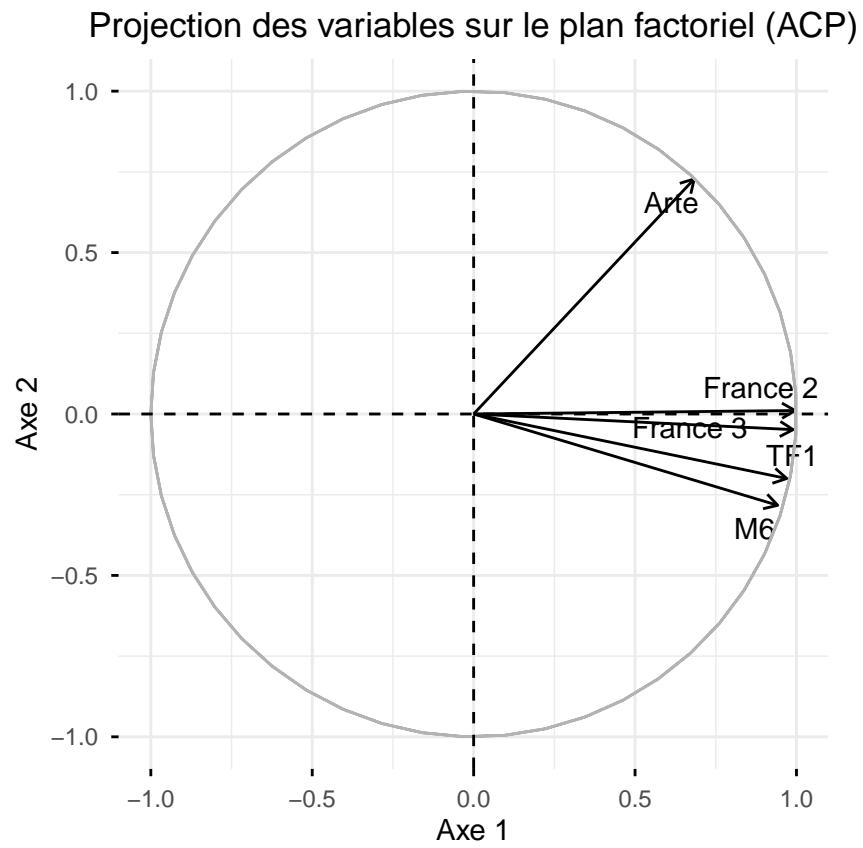
L'utilisation conjointe de ces deux bases de données permet d'adopter une approche globale du phénomène étudié. D'un côté, la base de l'INA offre une vision structurée des contenus informationnels diffusés par les médias audiovisuels ; de l'autre, la base de l'ARCOM permet d'analyser la manière dont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

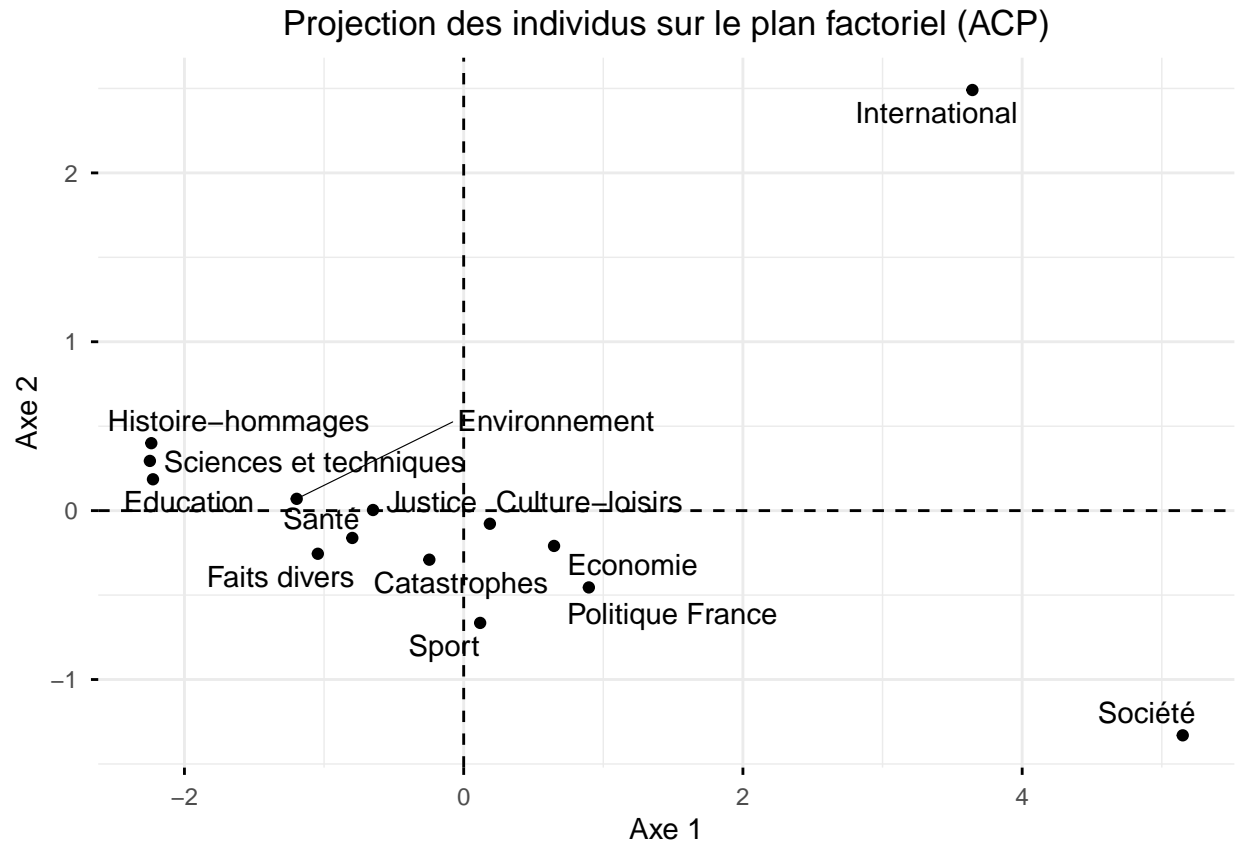
perçus, consommés et évalués par les individus. Ce double regard constitue un atout majeur du projet et justifie le recours à des méthodes statistiques variées afin d'exploiter au mieux l'information contenue dans les données.

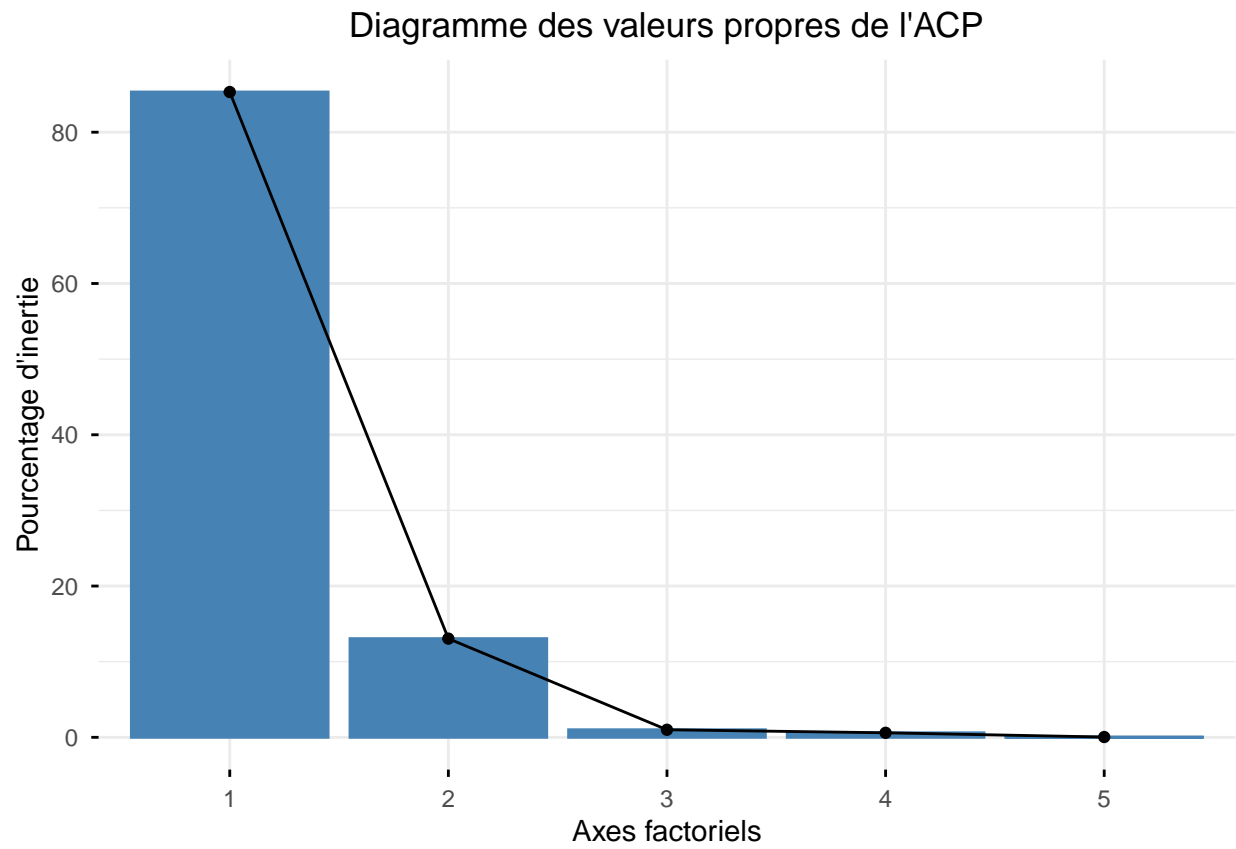
2 Partie I – ACP

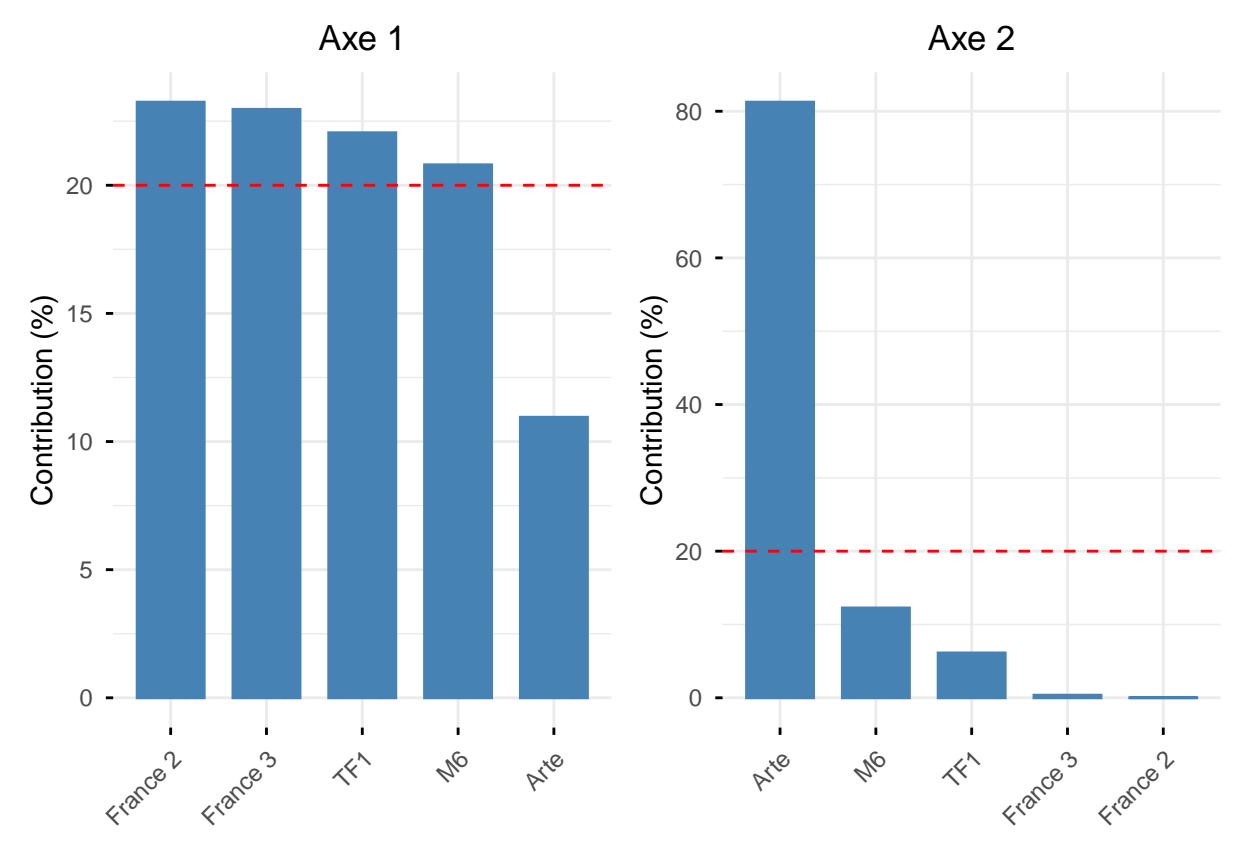
2.1

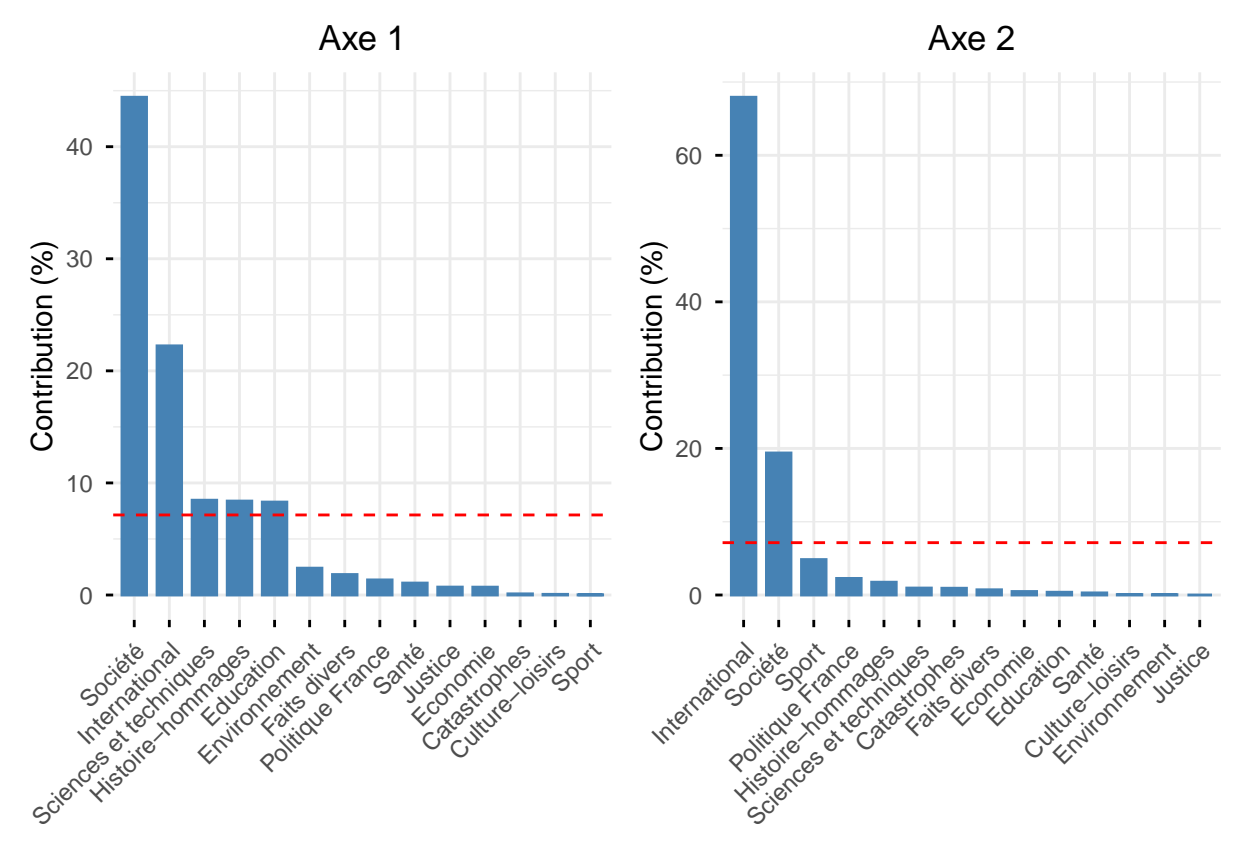
2.2











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.

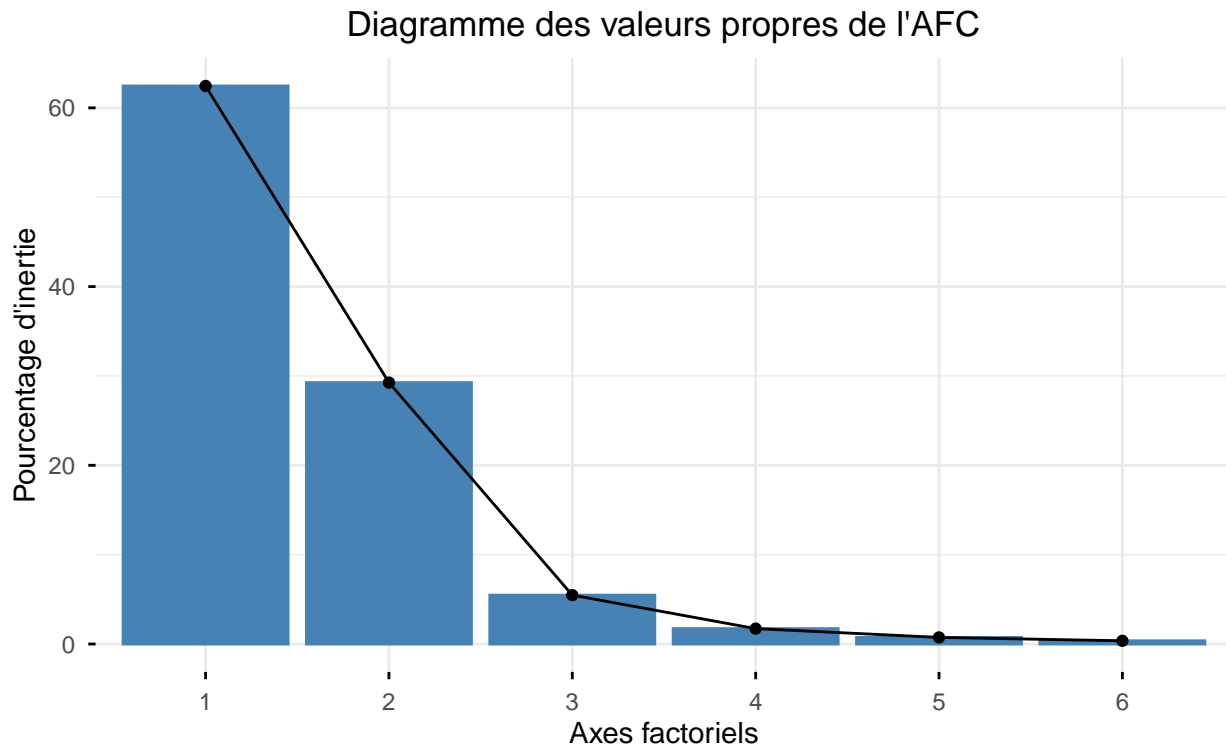
Avant toute interprétation, un test du khi-deux d'indépendance est réalisé.

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



Avant de debuter nous devons rechercher les axes a etudié, l'analyse des valeurs propres indique que :

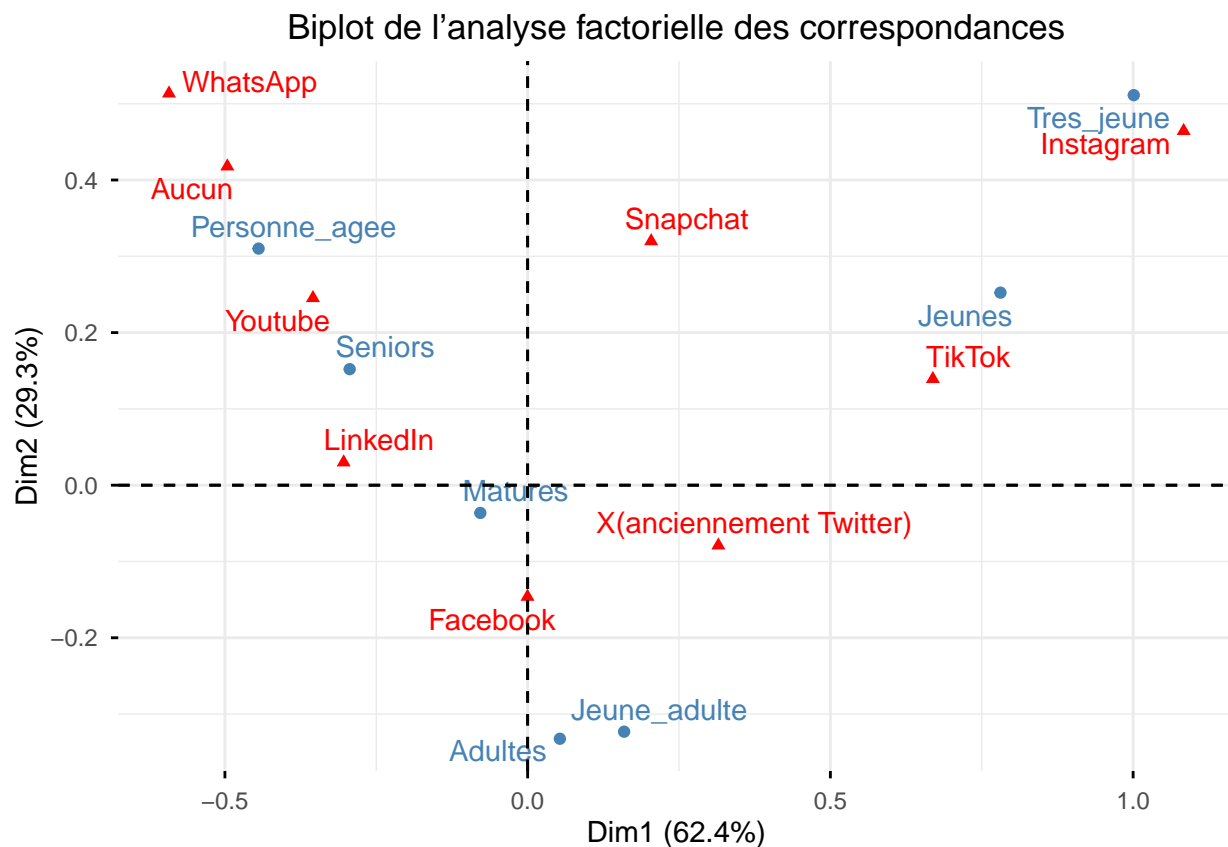
l'axe 1 explique environ 62.45 % de l'inertie totale ;

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 récente 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

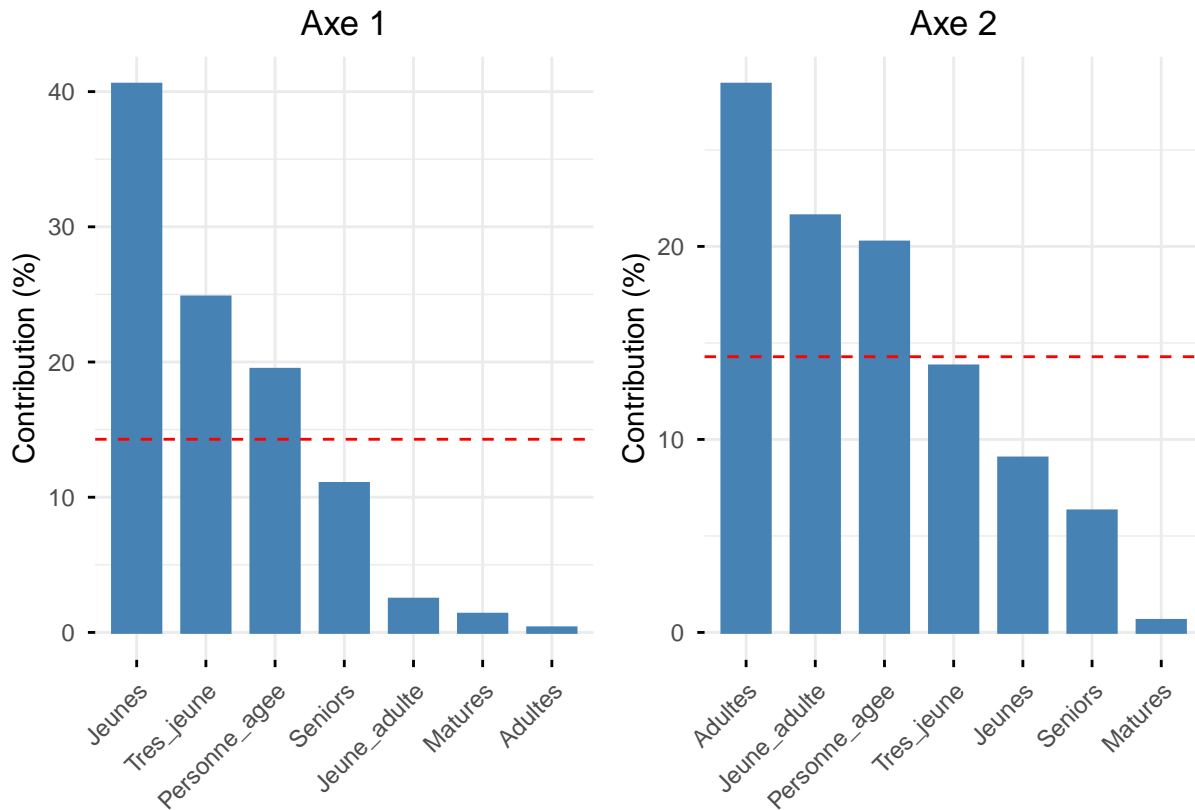


Le premier axe met en évidence une opposition générationnelle marquée dans l'usage des réseaux. Les classes d'âge jeunes et très jeunes sont principalement associées à des plateformes récentes telles qu'Instagram, TikTok ou Snapchat, tandis que les classes d'âge les plus élevées se rapprochent davantage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

rences 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.

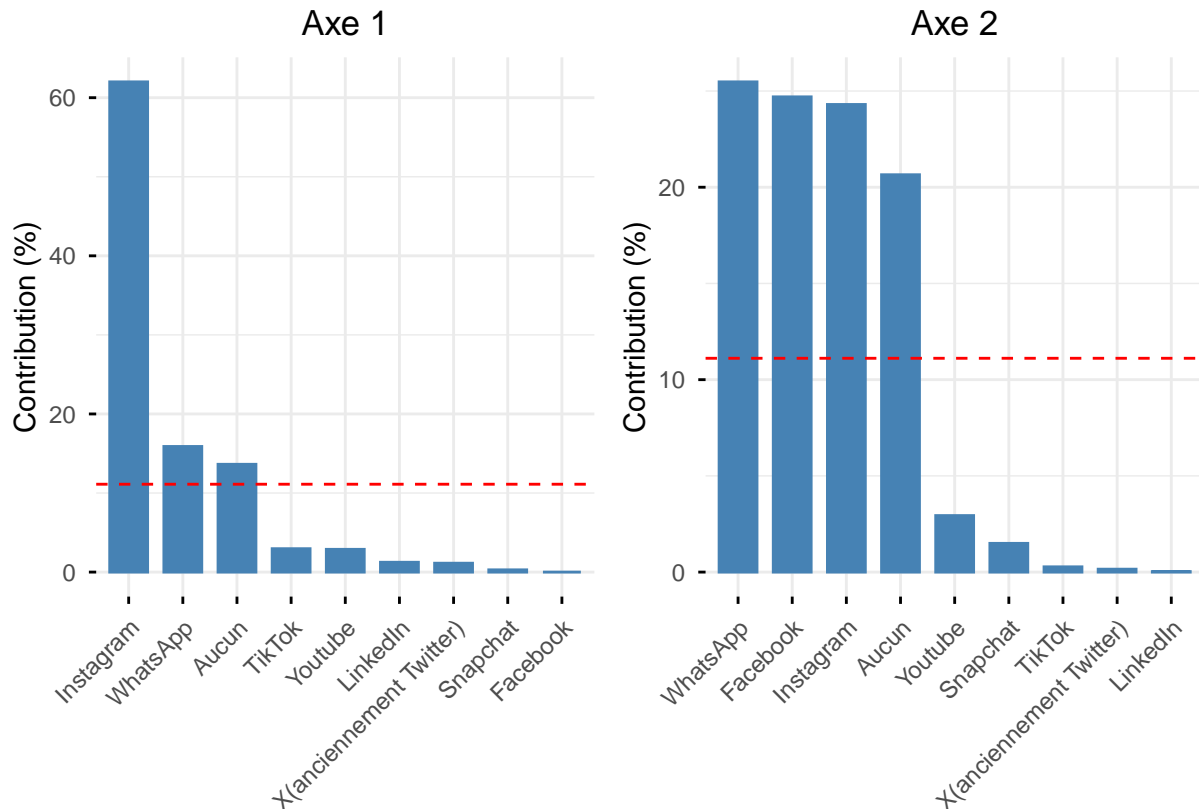


L'analyse des contributions montre que l'axe 1 est principalement structuré par les classes d'âge jeunes. Les catégories Jeunes, Très jeunes et Personnes âgées présentent des contributions nettement supérieures à la moyenne, avec une contribution particulièrement forte des Jeunes (environ 40 %). À l'inverse,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

4.1

4.2

5 Partie IV – Clustering

5.1

5.2

6 Conclusion générale

7 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>