



Reporting d'une analyse multivariée

SOMMAIRE

Introduction	2
1. Construction de la base de données	2
2. Méthodologie : Choix de l'analyse Factorielle.....	2
3. Analyse et interprétation des résultats	3
3.1 Synthèse des résultats de l'ACP	3
3.2 Interprétation des axes principaux	3
3.3 Projection des communes.....	3
3.4 Réponse à la problématique	4
3.5 Limites et pistes d'amélioration	4
4. Clustering des communes	4
Conclusion.....	5
Annexes	6

Professeur : Antoine ROLLAND

Etudiant : Franklin BECK, Théo FILLON, Muriel ZOUZZOU

BUT SCIENCE DES DONNEES – Parcours VCOD / EMS

Introduction

Dans le cadre de la SAE 4.02 du BUT Science des Données, nous avons mené une étude exploratoire autour de la thématique suivante : la typologie des communes françaises en fonction des équipements disponibles. L'objectif de cette analyse est de mieux comprendre les disparités entre les communes en matière d'accès à certains équipements, et d'illustrer comment ces différences peuvent être décrites et classées à l'aide d'outils statistiques multivariés.

Nous avons choisi de concentrer notre étude sur un angle spécifique : les équipements liés à la pratique du sport et aux soins médicaux associés. Plus précisément, notre analyse porte sur les communes françaises ayant entre 10 000 et 20 000 habitants. Cette taille intermédiaire permet de comparer des territoires ni trop petits pour manquer d'équipements, ni trop grands pour être statistiquement hors norme. Notre problématique est la suivante :

Les communes françaises de taille moyenne présentent-elles des différences notables dans leur niveau d'équipement sportif et de présence de professionnels de santé spécialisés selon leur typologie (urbaine, rurale, périurbaine) ?

Cette étude repose sur une base de données construite à partir de fichiers mis à disposition par l'INSEE, notamment les bases sur la population communale et les équipements publics. L'analyse est conduite à des fins d'aide à la décision territoriale, notamment pour éclairer les politiques publiques de répartition des ressources ou de développement des infrastructures.

1. Construction de la base de données

Pour répondre à la problématique, nous avons constitué une base de données adaptée à partir de plusieurs sources INSEE. Nous avons d'abord sélectionné les communes françaises dont la population se situe entre 10 000 et 20 000 habitants, en utilisant le fichier de typologie communale (grille_densite_7_niveaux_2024.xlsx), ce qui a abouti à un échantillon d'environ 550 communes.

Ensuite, nous avons extrait, depuis les fichiers « Ensemble-com-2021 », les données relatives aux équipements sportifs (stades, gymnases, piscines...) ainsi qu'aux professionnels de santé spécialisés en lien avec l'activité physique (kinésithérapeutes, cardiologues, etc.).

Les données ont été harmonisées, renommées et croisées via le code INSEE pour permettre les jointures nécessaires. Enfin, les variables ont été standardisées afin de rendre les données comparables et aptes à l'analyse multivariée. La base finale permet ainsi de comparer les communes sur la base de leurs équipements sportifs et médicaux, tout en tenant compte de leur typologie urbaine.

2. Méthodologie : Choix de l'analyse Factorielle

Afin de répondre à notre problématique, nous avons appliqué une Analyse en Composantes Principales (ACP), une méthode statistique multivariée particulièrement adaptée pour explorer un jeu de données comportant de nombreuses variables corrélées. L'objectif est ici de réduire la dimension des données tout en conservant l'essentiel de l'information, pour mieux visualiser les relations entre communes et équipements.

Plusieurs configurations ont été testées afin d'optimiser la qualité de l'analyse :

- Une première ACP incluant la variable **population totale (PTOT)** a montré que celle-ci dominait très largement la variance, masquant les structures plus fines liées aux équipements.
- Nous avons ensuite réalisé une seconde ACP **sans la population totale**, ce qui a permis de mieux faire ressortir les relations entre les variables d'équipements sportifs et médicaux.
- Enfin, nous avons exploré une **transformation logarithmique (log1p)** des données pour atténuer les effets de dispersion, sans gain significatif sur l'interprétation.

Sur la base de ces tests, la version sans PTOT, avec variables standardisées, s'est révélée la plus pertinente. Elle permet de conserver trois axes principaux expliquant ensemble **51,7 % de la variance totale**, un seuil satisfaisant dans une approche exploratoire. Ces axes serviront de fondement à l'analyse des résultats et à la typologie des communes présentée dans la suite du rapport.

3. Analyse et interprétation des résultats

3.1 Synthèse des résultats de l'ACP

L'analyse en composantes principales a été réalisée sur un ensemble de variables standardisées relatives aux équipements sportifs et aux professionnels de santé spécialisés, en excluant volontairement la variable population totale afin d'éviter qu'elle ne domine la structure des données. Cette démarche a permis de mieux mettre en évidence les relations intrinsèques entre les équipements.

Les résultats montrent que les trois premiers axes factoriels expliquent ensemble 51,7 % de la variance totale du jeu de données. Ce niveau de variance expliquée est considéré comme satisfaisant dans un cadre exploratoire, car il permet de dégager des tendances significatives tout en simplifiant l'interprétation.

3.2 Interprétation des axes principaux

- **Premier axe (26,3 % de variance expliquée)** : Cet axe oppose principalement les communes en fonction de la quantité et de la diversité des équipements sportifs disponibles. Les communes bien équipées en infrastructures sportives (stades, piscines, salles de sport) se situent à une extrémité, tandis que celles avec moins d'équipements se trouvent à l'autre.
- **Deuxième axe (16,5 % de variance expliquée)** : Ce facteur semble refléter la présence et la spécialisation des professionnels de santé liés au sport, comme les kinésithérapeutes, cardiologues ou autres spécialistes. Il met en lumière une différenciation des communes selon leur offre en services médicaux spécialisés.
- **Troisième axe (8,9 % de variance expliquée)** : Il met en évidence des nuances plus fines, notamment des équipements moins fréquents ou spécifiques, apportant une couche supplémentaire à la typologie des communes.

3.3 Projection des communes

La projection des communes dans le plan défini par ces trois axes révèle plusieurs groupes distincts, suggérant l'existence de typologies communes :

- Certaines communes se distinguent par un équipement sportif dense et une offre médicale spécialisée importante, souvent situées en zones urbaines ou périurbaines.
- D'autres se caractérisent par une faible présence d'équipements, typiquement des communes plus rurales.
- Un troisième groupe intermédiaire montre des profils mixtes, avec une offre partielle d'équipements.

Cette répartition corrobore l'hypothèse selon laquelle la typologie des communes (urbaine, rurale, périurbaine) impacte significativement leur niveau d'équipement.

3.4 Réponse à la problématique

Les résultats confirment que les communes françaises de taille moyenne présentent des disparités notables en termes d'équipements sportifs et de professionnels de santé spécialisés. L'ACP a permis de mettre en lumière ces différences de manière synthétique et visuelle, facilitant la compréhension des structures sous-jacentes.

L'exclusion de la variable population totale s'est avérée pertinente, car elle a permis de révéler des relations plus fines entre les variables d'équipements, autrement masquées par la dominance de la population.

3.5 Limites et pistes d'amélioration

Cette analyse exploratoire présente certaines limites, notamment le fait que près de la moitié de la variance n'est pas expliquée par les trois premiers axes, ce qui suggère la présence d'autres facteurs non pris en compte.

Par ailleurs, l'exclusion de la population totale, bien que justifiée, pourrait être approfondie en intégrant d'autres variables socio-économiques ou géographiques, telles que le revenu moyen, le taux de chômage, ou la densité de population.

Enfin, une extension possible serait de réaliser un clustering des communes sur la base des composantes principales pour formaliser la typologie, ainsi que d'automatiser cette analyse dans un outil interactif (par exemple avec R Shiny ou Python Streamlit) pour faciliter son exploitation par des décideurs ou gestionnaires territoriaux.

4. Clustering des communes

À la suite de l'ACP, nous avons appliqué une méthode de classification automatique (clustering) sur les coordonnées des communes dans l'espace factoriel. Cette démarche a permis de regrouper les communes en **trois profils types** selon leur niveau d'équipements sportifs et médicaux.

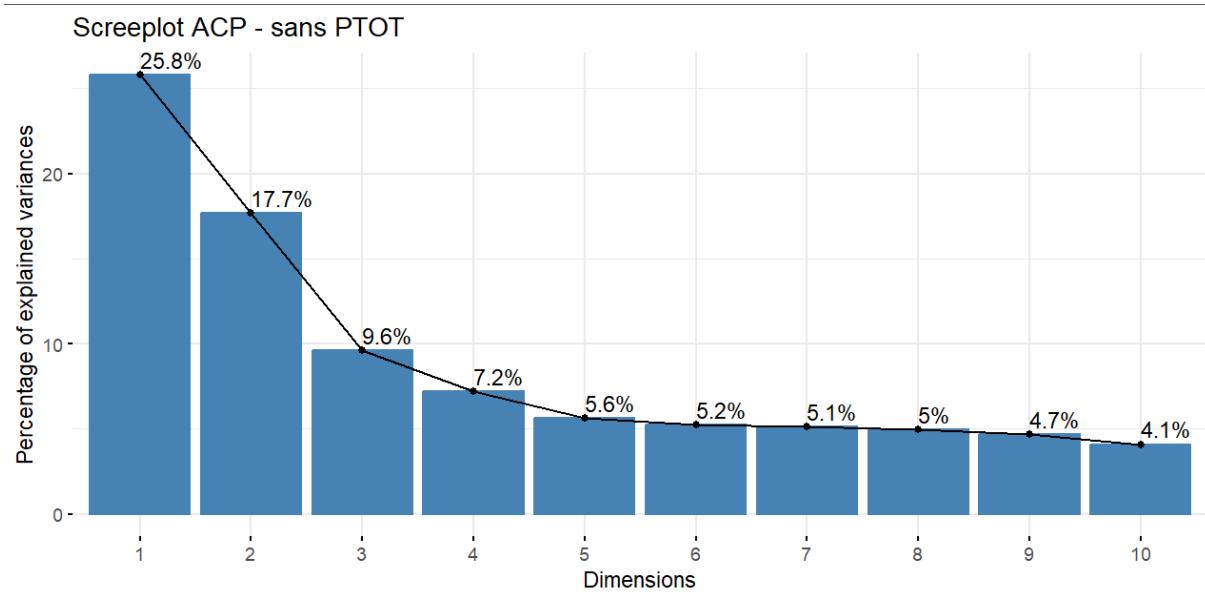
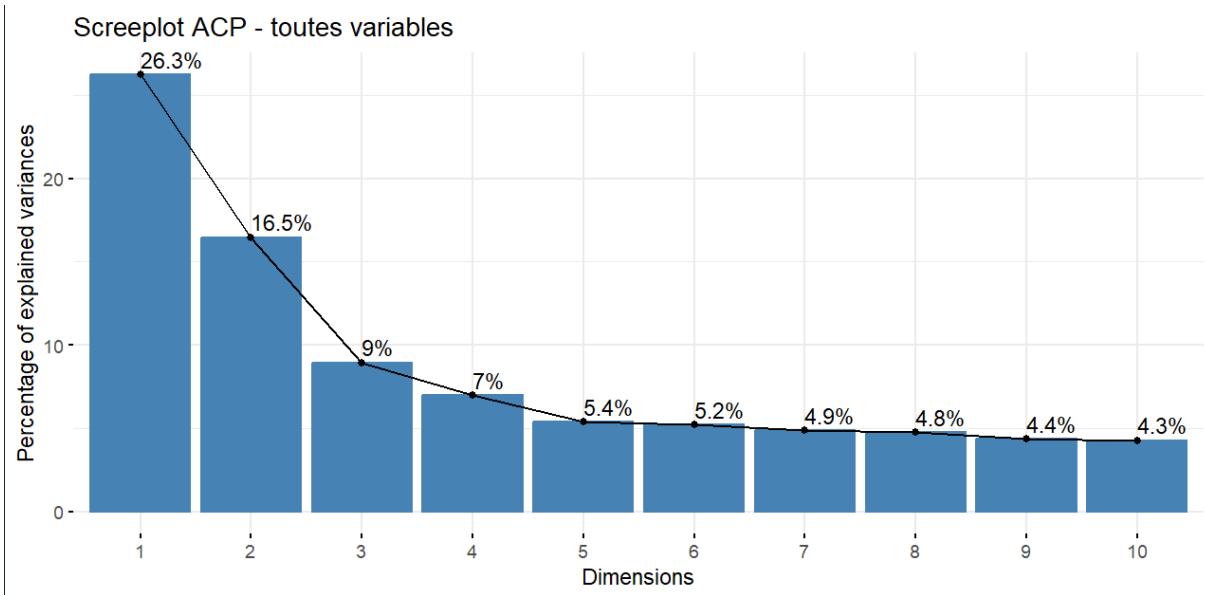
- **Cluster 1** : communes bien dotées, présentant une offre dense et diversifiée en infrastructures sportives et en professionnels de santé. Ces communes sont majoritairement urbaines ou périurbaines.
- **Cluster 2** : communes disposant d'équipements plus spécialisés, notamment sur le plan médical, mais avec une offre sportive plus restreinte.
- **Cluster 3** : communes peu équipées, aussi bien sportivement que médicalement, généralement rurales.

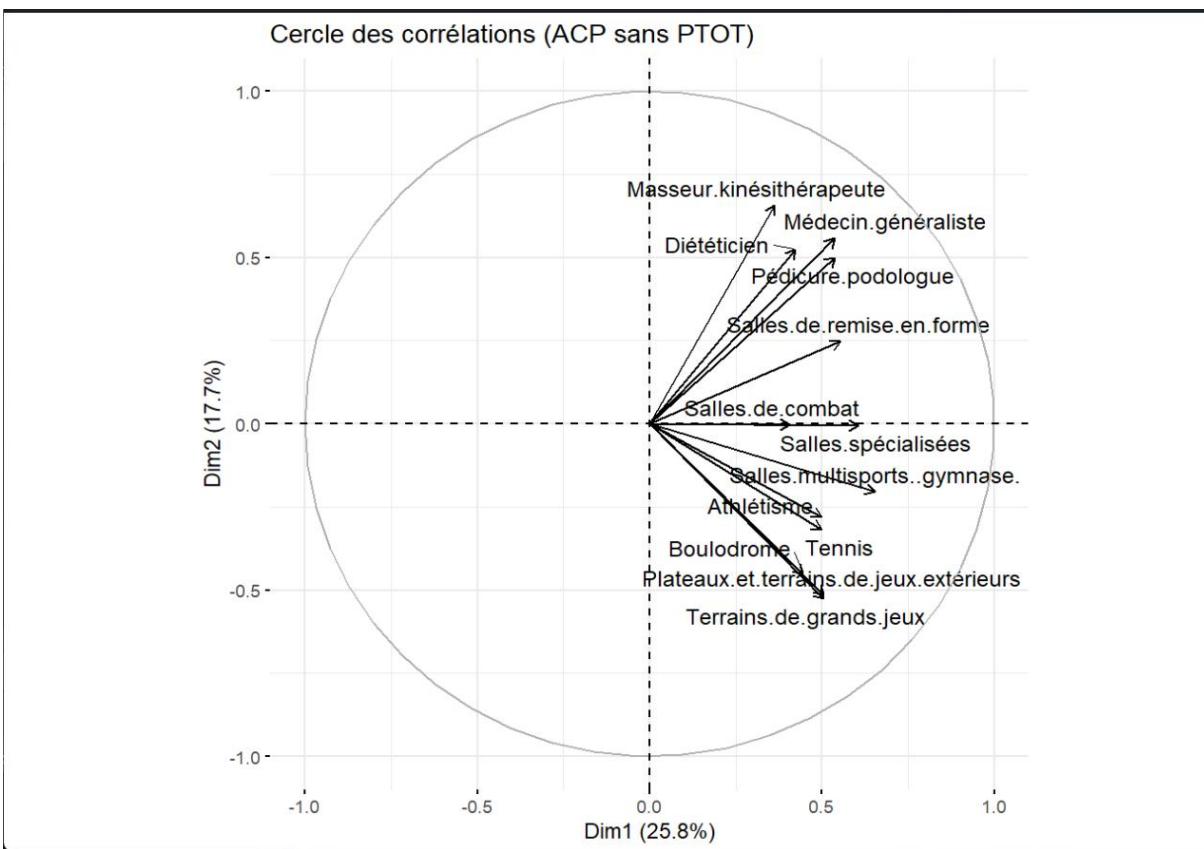
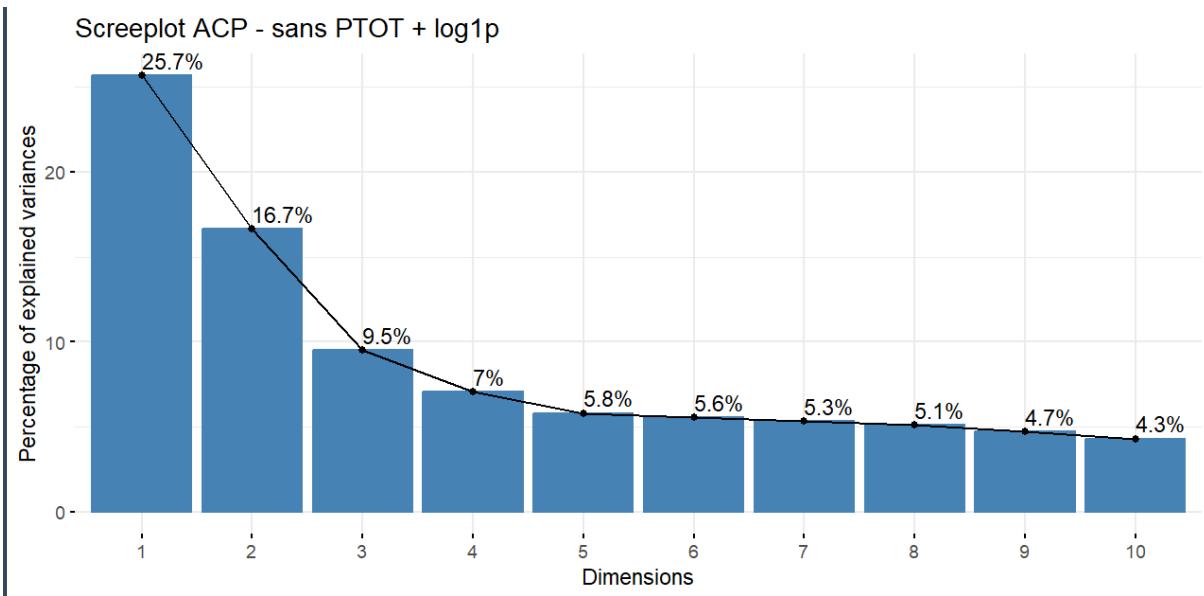
Cette typologie confirme et affine les disparités mises en évidence par l'ACP. Le clustering ajoute ainsi une **lecture structurée et opérationnelle** des résultats, facilitant la prise de décision territoriale par l'identification de besoins spécifiques selon le profil communal.

Conclusion

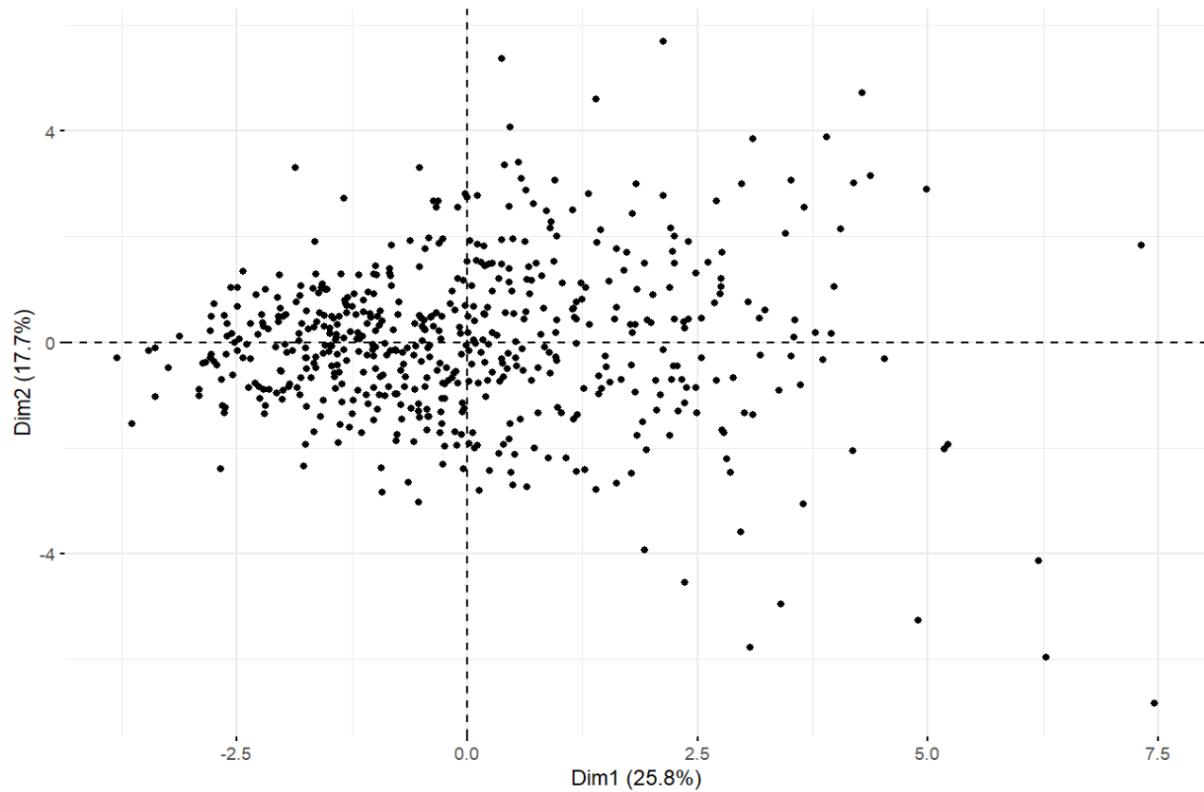
Cette étude a mis en lumière les disparités d'équipements sportifs et médicaux entre les communes françaises de taille moyenne. L'analyse factorielle, enrichie par un clustering, a permis de dégager des profils types cohérents selon la typologie territoriale. Ces résultats peuvent orienter les politiques d'aménagement et de répartition des ressources, en ciblant plus finement les besoins des territoires.

Annexes

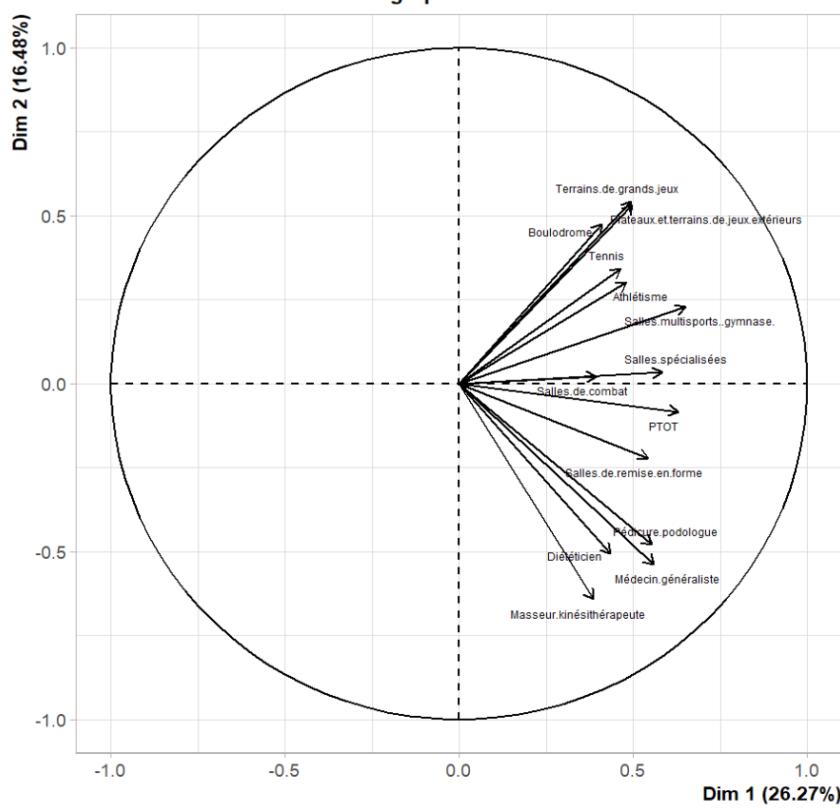




Projection des communes (ACP sans PTOT)



PCA graph of variables



Typologie des communes (Clustering sur ACP)

