

# Segmentation des départements français selon leur offre culturelle et leur contexte socio-économique



ROUSSEAU Nora et BOUDAMOUS Lyna



# Sommaire

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Presentation des bases de données . . . . .	2
1.1.1	Base de données quantitative . . . . .	2
1.1.2	Base de données qualitative . . . . .	2
<b>2</b>	<b>Partie 1 : ACP</b>	<b>3</b>
2.1	Matrice de corrélation . . . . .	3
2.2	Etude de l'inertie . . . . .	3
2.3	Etude des variables . . . . .	4
2.4	Contributions des variables . . . . .	5
2.5	Etude des individus . . . . .	5
2.6	Contributions des individus . . . . .	6
<b>3</b>	<b>Partie 2 : ACM</b>	<b>7</b>
3.1	Etude de l'inertie . . . . .	7
3.2	Nuage des modalités sur les différents plans . . . . .	7
3.3	Nuage des individus sur les différents plan . . . . .	9
3.4	Contribution des modalités . . . . .	10
3.5	Contribution des individus . . . . .	10
3.6	Projection des Variables Supplémentaire sur les Axes Culturels . . . . .	11
3.6.1	Représentations des départements suivant leur salaire moyen net par an . . . . .	11
3.6.2	Représentations des départements suivant leur taux de chômage . . . . .	12
3.6.3	Représentations des départements suivant leur part de population urbaine . . . . .	12
3.6.4	Représentations des départements suivant leur dépenses culturelles par habitant . . . . .	13
3.6.5	Représentations des départements suivant leurs régions . . . . .	13
<b>4</b>	<b>Partie 3 : Clustering</b>	<b>14</b>
4.1	Analyse de l'inertie dans le clustering . . . . .	14
4.2	Dendrogramme et Répartition des Clusters . . . . .	14
4.3	Visualisation des clusters sur les différents plans factoriels . . . . .	15
4.4	Clustering Hiérarchique . . . . .	15
4.5	Tableau des données regroupées par clusters selon les variables . . . . .	16
4.6	Tableau des données regroupées par clusters selon la distribution des individus . . . . .	17
4.7	Tableau des résultats du test du chi-deux . . . . .	18
4.8	Tableau des catégories descriptives du Cluster 1 . . . . .	18
4.9	Tableau des catégories descriptives du Cluster 2 . . . . .	19
4.10	Tableau des catégories descriptives du Cluster 3 . . . . .	19
4.11	Tableau des catégories descriptives du Cluster 4 . . . . .	19
<b>5</b>	<b>Conclusion</b>	<b>20</b>



# 1 Introduction

L'accès à la culture, en tant qu'élément clé de la vie sociale et économique, représente un enjeu majeur dans la construction des territoires. Dans ce contexte, nous avons choisi, dans le cadre de notre Master 1 en économie de l'entreprise et des marchés à l'Université de Tours, d'analyser l'offre culturelle en France. En participant au challenge d'Open Data University, initié par l'un de nos professeurs dans son cours de Datamining, nous avons créé un **Dashboard** (tableau de bord) pour visualiser la répartition des différentes offres culturelles en France. Cette première étape nous a permis de poser les bases de notre analyse en observant la diversité de l'offre culturelle à l'échelle nationale.

Dans cette deuxième partie de notre étude, nous allons approfondir l'analyse en examinant comment l'offre culturelle varie d'un département à l'autre, en tenant compte de plusieurs facteurs socio-économiques. Cette variabilité certaine peut avoir un impact majeur sur l'accès à la culture pour les populations locales. En effet, l'intensité de l'offre culturelle dans un département peut être étroitement liée à des facteurs comme le niveau de revenus, le taux de chômage, ou encore la répartition de la population urbaine et rurale. Ainsi, il devient pertinent de mieux comprendre comment ces facteurs interagissent pour définir des profils culturels distincts à travers le pays.

L'objectif de cette étude est de mieux comprendre les disparités culturelles à travers le pays et ceci en répondant à la problématique suivante : **Comment caractériser et classer les départements en fonction de leur offre culturelle et de leur contexte socio-économique ?**

Pour répondre à cette question, nous allons analyser les départements aux travers de plusieurs variables tel que le nombre des différents équipements culturels pour 100 000 habitants (cinémas, musées, bibliothèques...), ainsi que des indicateurs socio-économiques comme le salaire moyen net par an, le taux de chômage, la part de la population urbaine et les dépenses culturelles des départements par habitant. Ces données nous permettront de réaliser une analyse en profondeur à travers des techniques statistiques telles que l'Analyse en Composantes Principales (ACP), l'Analyse des Correspondances Multiples (ACM) et des méthodes de Clustering pour mieux comprendre les profils des départements.

## 1.1 Présentation des bases de données

### 1.1.1 Base de données quantitative

Cette base de données quantitative comprend diverses variables relatives aux départements, qui seront nos individus pour cette étude. Les variables incluent le nombre de cinémas, bibliothèques, musées, spectacles (théâtres, opéras, et scènes) et d'établissements culturels (centres culturels et conservatoires) pour 100 000 habitants par département. Elle contient également des données sur le salaire moyen net annuel, qui reflète le niveau de vie des habitants des départements, le taux de chômage, qui donne une indication sur la situation économique locale, la part de la population urbaine, ainsi que les dépenses culturelles par habitant, mesurant l'engagement des départements dans la culture, et enfin la région de chaque département.

	A	B	C	D	E	F	G	H	I	J	K
1	Département	Ciné_100k_h	Biblio_100k_h	Spectacles_100k_h	Etablissements_cult_100k_h	Musées_100k_h	Salaire_moyen_net_an	Taux_chômage_p	Part_pop_urbaine_p	Dépenses_culturelles_par_h	Région
2	Ain	3,2	36,9	0,8	0,9	2,6	27420	6	67		19 Auvergne-Rhône-Alpes
3	Aisne	2,7	24,5	1,8	1,1	3,7	25236	11,7	53,2		18 Hauts-de-France
4	Allier	2,3	70,3	2,6	0,9	3,2	24552	8,9	58,3		23 Auvergne-Rhône-Alpes
5	Alpes-Maritim	26,7	86,7	4,2	7,7	22,6	28980	8,7	95,9		19 Provence-Alpes-Côte d
6	Alpes-de-Haut	1,3	8,6	0,2	0,1	0,9	25428	9,8	61,9		34 Provence-Alpes-Côte d
7	Ardennes	1,5	30,8	1,2	0,3	3	25140	9,9	57,3		14 Grand Est
8	Ardèche	7,1	82,5	1,4	1,4	3,2	25524	9	63,3		18 Auvergne-Rhône-Alpes
9	Ariège	7	52,8	1,9	0,6	3,8	24348	10,6	55,5		57 Occitanie
10	Aube	1,3	48,2	0,9	1,6	4,4	25068	10,7	61,3		24 Grand Est

### 1.1.2 Base de données qualitative

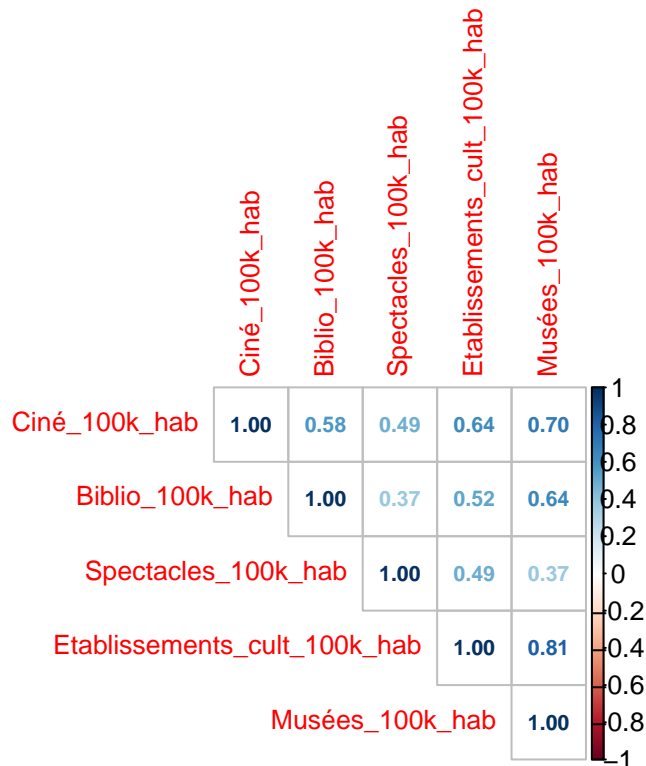
Pour effectuer l'ACM, nous avons transformées les variables quantitatives en variables qualitatives en les regroupant par intervalle. Par exemple, les nombres de cinémas, bibliothèques, spectacles... ont été regroupés en catégories (0-2, 2-4, ...) en fonction de la dispersion des données, et les indicateurs socio-économiques ont été regroupés en catégories du type "Moyen supérieur", "Bas", "Élevé"... en se basant sur les références nationales.

	A	B	C	D	E	F	G	H	I	J	K
1	Département	Ciné_100k_h	Biblio_100k_h	Spectacles_100k_h	Etablissements_cult_100k_h	Musées_100k_h	Salaire_moyen_net_an	Taux_chômage_p	Part_pop_urbaine_p	Dépenses_culturelles_par_h	Région
2	Ain	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inférieur (25-30k) Bas	Urba Moyenne	Faible (10-20€)		Auvergne-Rhône-Alpes
3	Aisne	0-5 C	15-30 B	0-2 S	1-2 EC	2-4 M	Moyen inférieur (25-30k) Élevé	Urba Moyenne	Faible (10-20€)		Hauts-de-France
4	Allier	0-5 C	50-80 B	2-4 S	0-1 EC	2-4 M	Faible (<25k) Moyen	Urba Moyenne	Moyenne (20-50€)		Auvergne-Rhône-Alpes
5	Alpes-Maritim	>15 C	>80 B	>4 S	>2 EC	>4 M	Moyen inférieur (25-30k) Moyen	Urba Très forte	Faible (10-20€)		Provence-Alpes-Côte d'A
6	Alpes-de-Hau	0-5 C	<15 B	0-2 S	0-1 EC	0-2 M	Moyen inférieur (25-30k) Moyen	Urba Moyenne	Moyenne (20-50€)		Provence-Alpes-Côte d'A
7	Ardennes	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inférieur (25-30k) Moyen	Urba Moyenne	Faible (10-20€)		Grand Est
8	Ardèche	5-10 C	>80 B	0-2 S	1-2 EC	2-4 M	Moyen inférieur (25-30k) Moyen	Urba Moyenne	Faible (10-20€)		Auvergne-Rhône-Alpes
9	Ariège	5-10 C	50-80 B	0-2 S	0-1 EC	2-4 M	Faible (<25k) Élevé	Urba Moyenne	Élevée (>50€)		Occitanie
10	Aube	0-5 C	30-50 B	0-2 S	1-2 EC	>4 M	Moyen inférieur (25-30k) Élevé	Urba Moyenne	Moyenne (20-50€)		Grand Est

Pour les analyses qui vont suivre, nous avons choisi de traiter les variables liées à l'offre culturelle comme variables principales, tandis que les variables socio-économiques ainsi que la région ont été intégrées en tant que variables supplémentaires.

## 2 Partie 1 : ACP

### 2.1 Matrice de corrélation

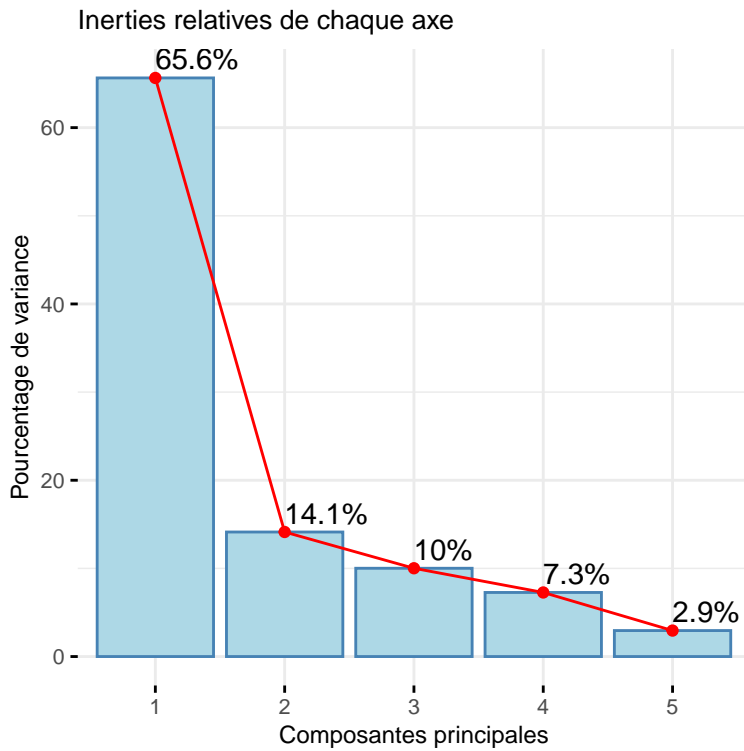


On remarque que les différentes offres culturelles sont toutes corrélées positivement entre elles. On observe une forte corrélation positive (0.81) entre Etablissements\_cult\_100k\_hab et Musées\_100k\_hab, indiquant que les départements avec plus d'établissements culturels ont également plus de musées, il en est de même pour les cinémas et les musées qui ont une corrélation de 0.7.

On voit que les spectacles présentent des corrélations plus faibles avec les autres offres culturelles, notamment avec les bibliothèques et les musées (0.37), suggérant que l'offre de spectacles est moins liée à la présence de ses équipements culturels.

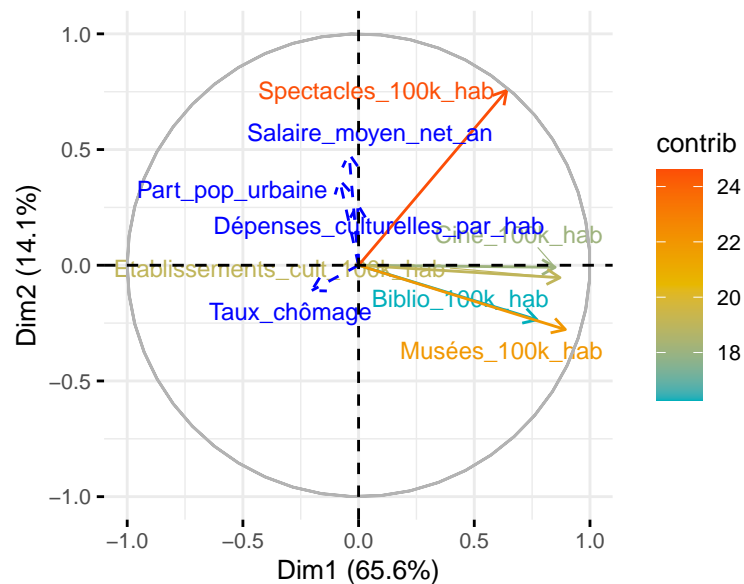
### 2.2 Etude de l'inertie

	comp 1	comp 2	comp 3	comp 4	comp 5
eigenvalue	3.282	0.706	0.501	0.363	0.147
percentage of variance	65.639	14.128	10.016	7.269	2.948
cumulative percentage of variance	65.639	79.767	89.783	97.052	100.000



On conserve les deux premiers axes car ils représentent 79.77 % de la variance totale des données, ce qui signifie qu'ils capturent une grande majorité de l'information présente dans les données.

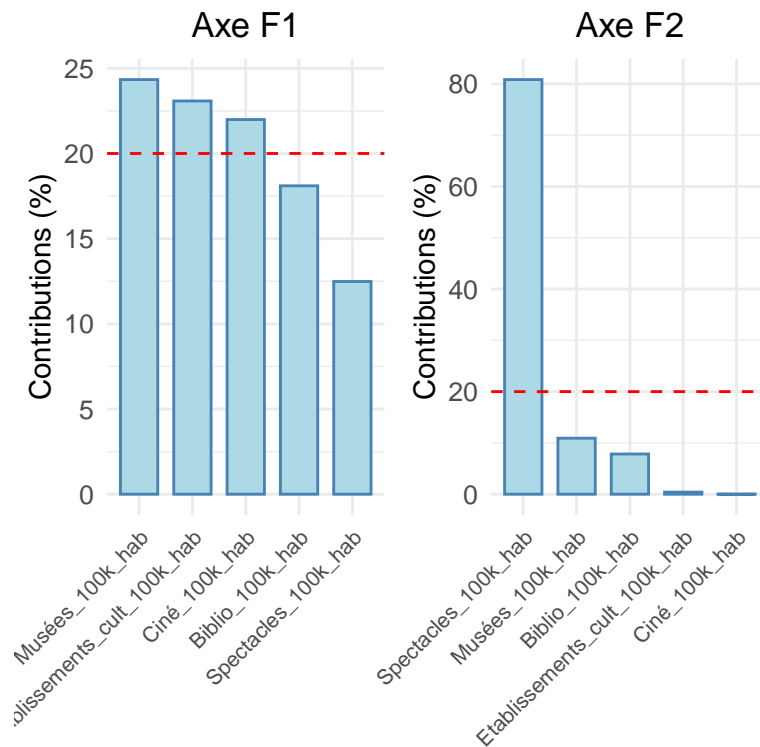
### 2.3 Etude des variables



L'axe  $F1$  explique 65.6 % de la variance, tandis que l'axe  $F2$  en explique 14.1 %. Les variables qui contribuent le plus à la formation des axes  $F1$ ,  $F2$  sont Spectacles\_100k\_hab et Musées\_100k\_hab.

On observe que la variable supplémentaire Salaire\_moyen\_net\_an est associée à l'axe  $F2$ , comme en témoigne la longueur de sa flèche. Cela suggère que  $F2$  est liée à des caractéristiques des départements où le salaire moyen net annuel est élevé, bien que cela ne signifie pas que cet axe soit uniquement défini par cette variable.

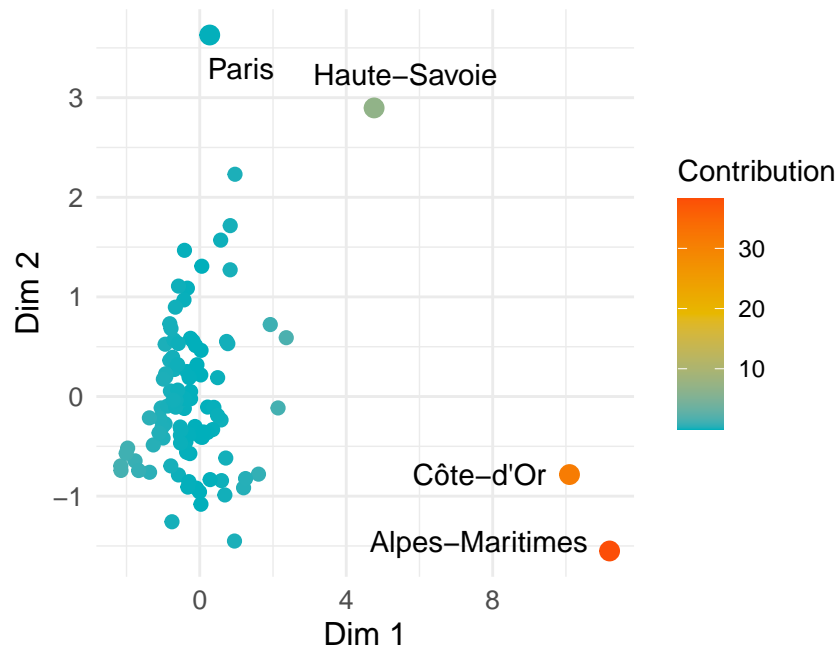
## 2.4 Contributions des variables



On remarque que les variables Musées\_100k\_hab, Etablissements\_cult\_100k\_hab et Ciné\_100k\_hab sont les plus contributives à l'axe  $F_1$ , indiquant qu'elles jouent un rôle important dans la construction de cet axe.

Pour l'axe  $F_2$ , Spectacles\_100k\_hab a la contribution la plus importante, ce qui suggère que cette variable est déterminante dans la formation de cet axe, avec un impact majeur sur la variance expliquée par  $F_2$ .

## 2.5 Etude des individus



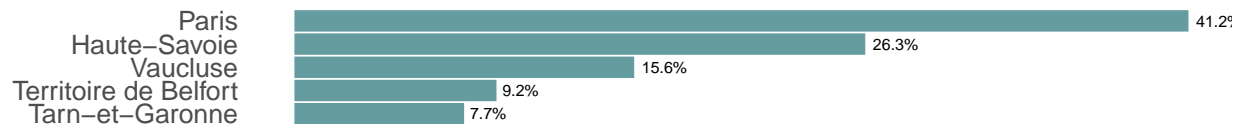
On voit que Haute-Savoie, Côte-d'Or et Alpes-Maritimes contribuent fortement à l'axe  $F_1$ , caractérisée par les variables Musées\_100k\_hab, Etablissements\_cult\_100k\_hab et Ciné\_100k\_hab.

Paris et Haute-Savoie se distinguent sur l'axe  $F_2$ , principalement défini par Spectacles\_100k\_hab, ce qui indique une forte offre de spectacles dans ces départements.

## 2.6 Contributions des individus



Contribution à l'axe F1



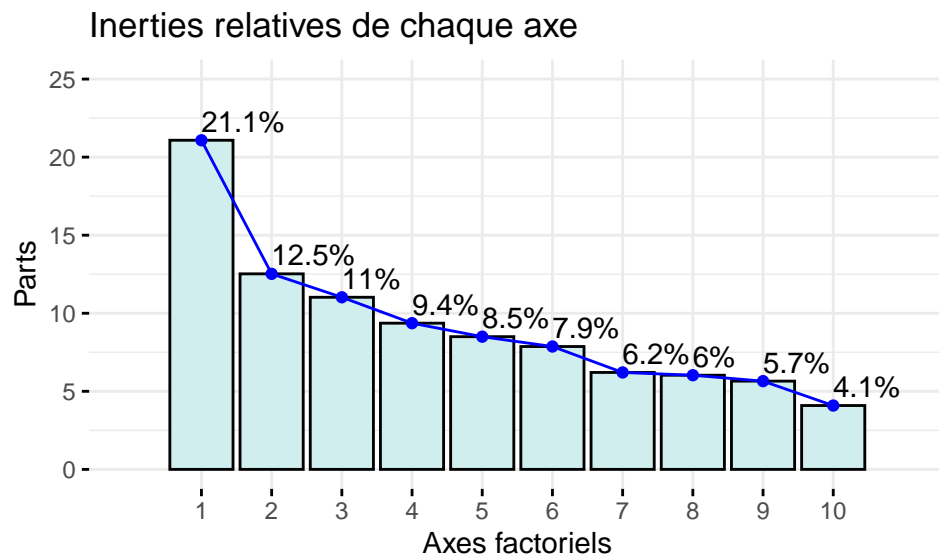
Contribution à l'axe F2

Les départements qui contribuent le plus à l'axe  $F_1$ , sont les Alpes-Maritimes, la Côte-d'Or, la Haute-Savoie, les Alpes-de-Haute-Provence et le Lot ; tandis que pour l'axe  $F_2$ , ce sont Paris, les Hauts-de-Seine, la Haute-Savoie, le Vaucluse, le Territoire de Belfort et le Tarn-et-Garonne.

## 3 Partie 2 : ACM

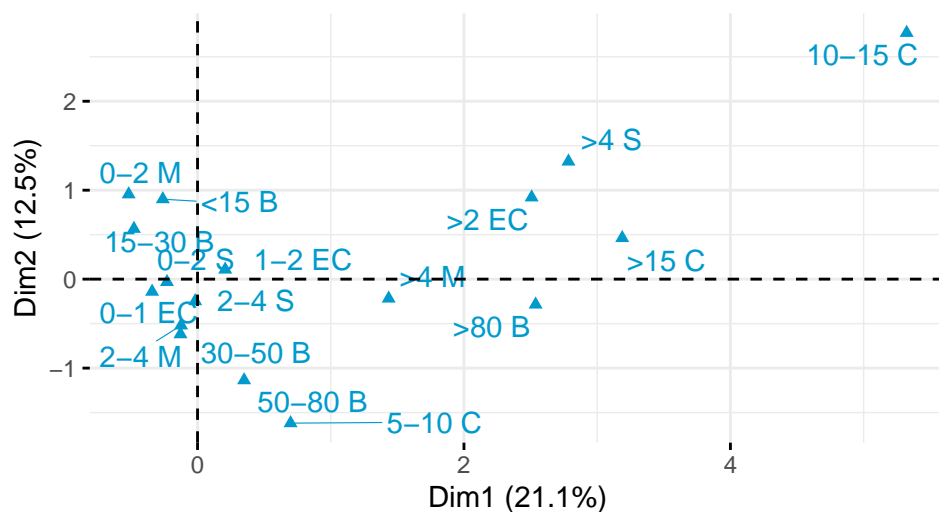
### 3.1 Etude de l'inertie

	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
Valeur propre	0.55	0.33	0.29	0.24	0.22	0.20	0.16	0.16	0.15	0.11	0.08	0.07	0.05
Part d'inertie	21.08	12.52	11.02	9.36	8.50	7.87	6.21	6.03	5.65	4.09	3.23	2.56	1.87
Part cumulée	21.08	33.60	44.63	53.99	62.49	70.36	76.57	82.60	88.25	92.34	95.57	98.13	100.00



On décide de conserver les 4 premiers axes car ils expliquent 53.99 % de l'inertie totale, soit une part significative de la variabilité des données. Ce choix permet de capturer plus de la moitié de l'information présente dans les données tout en maintenant une bonne interprétabilité.

### 3.2 Nuage des modalités sur les différents plans



Les Axes 1 et 2 représentent 33.6 % de l'inertie totale, ce qui constitue la majeure partie de l'information présente dans les données.



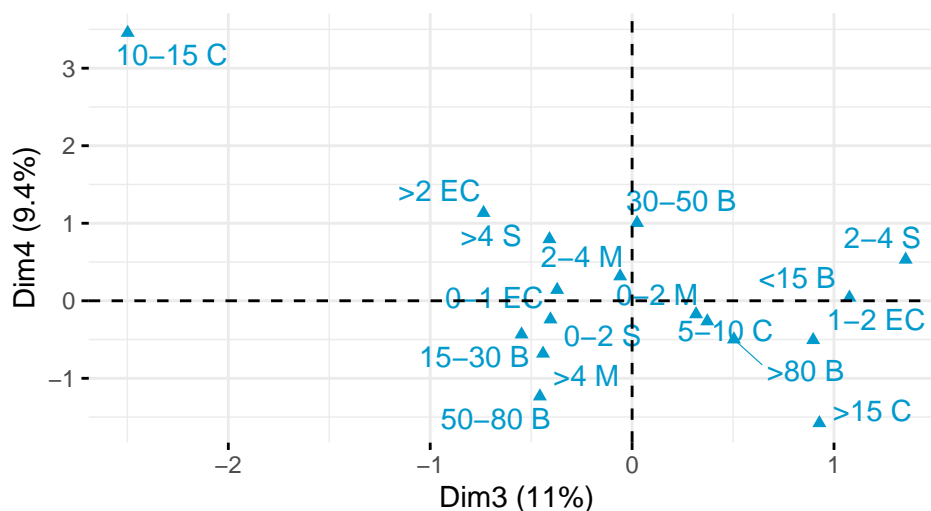
On peut voir que les modalités qui contribuent le plus à l'Axe  $F_1$  sont 10-15 C (Cinémas), >15 C, >4 S (Spectacles), >80 B (Bibliothèques) et >2 EC (Etablissements Culturels) .

Sur l'Axe  $F_2$ , se sont 10-15 C, >4 S, 5-10 C et 50-80 B les modalités qui contribuent le plus.

Les autres modalités proches de l'origine ont une contribution moins importante aux axes et ont des profils plus similaires.

Les modalités 2-4 M (Musées) et 30-50 B sont proches sur le graphique, ce qui indique qu'elles sont associées à des départements ayant des caractéristiques similaires contrairement à 30-50 B et 10-15 C qui sont opposés sur le graphique, il en est de même pour les modalités >4 S et 0-1 EC.

La modalité 10-15 C est isolée, ce qui suggère qu'elle correspond à des départements spécifiques qui se distinguent par leurs profils culturels.



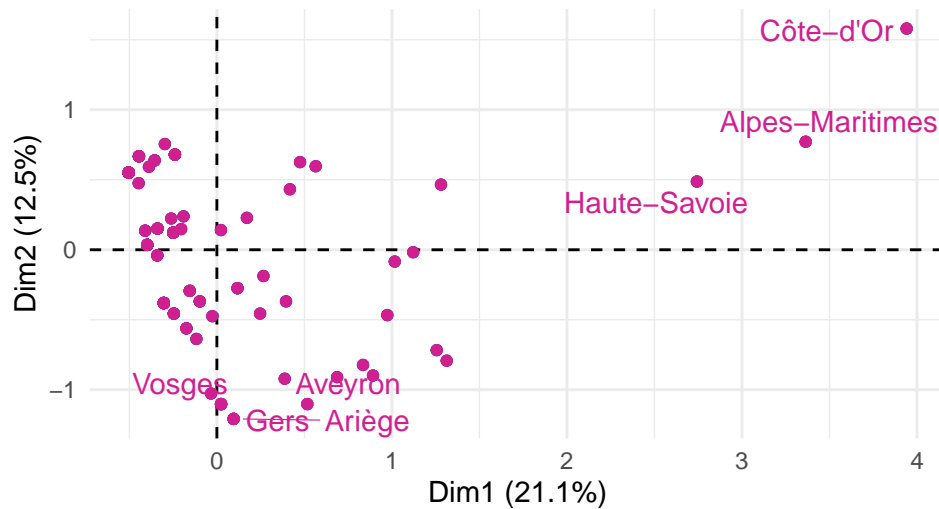
Les Axes 3 et 4 représentent 20,4 % de l'inertie totale, offrant ainsi une contribution importante à la compréhension des variations restantes dans les données, bien que leur part soit moins significative que celle des deux premiers axes.

On peut apercevoir que 10-15 C, 2-4 S, et <15 B sont les modalités qui contribuent le plus à l'Axe  $F_3$  ; tandis que sur l'Axe  $F_4$  se sont 10-15 C, >2 EC, >15 C et 50-80 B.

0-2 M et 5-10 C sont proches sur le graphique, ce qui indique qu'elles sont associées à des départements ayant des caractéristiques plutôt similaires. On remarque que >80 B et 10-15 C sont opposés sur le graphique ce qui indique que ces 2 modalités sont rarement associés dans un même département.

On voit que sur le plan  $F_3, F_4$ , la modalité 10-15 C est également isolée comme sur le plan  $F_1, F_2$  ; ce qui indique que cette modalité correspond à des départements particuliers.

### 3.3 Nuage des individus sur les différents plan

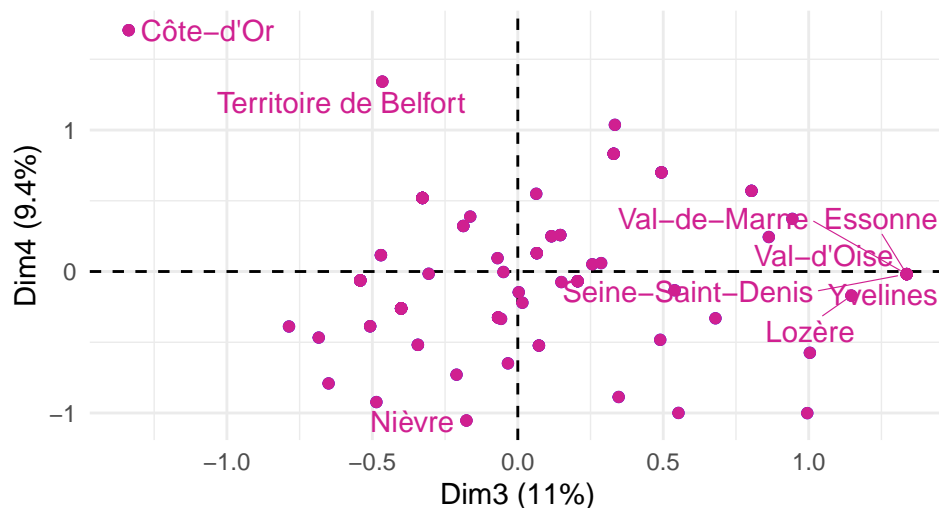


Les Axes 1 et 2 représentent 33.6 % de l'inertie totale, ce qui constitue la majeure partie de l'information présente dans les données.

On peut voir que les individus qui contribuent le plus à l'Axe  $F_1$  sont les départements Côte d'Or, Alpes-Maritimes et Haute-Savoie ; et sur l'Axe  $F_2$  se sont la Côte d'Or, l'Ariège, le Gers, l'Aveyron, les Vosges, le Tarn-et-Garonne et les Hautes-Pyrénées qui contribuent le plus. Les autres départements proches de l'origine ont une contribution moins importante aux axes et ont des profils plutôt similaires entre eux.

On remarque que le Gers et l'Ariège ont exactement la même position sur le graphique (même point) ce qui signifie que ces deux départements présentent des caractéristiques culturelles très similaires, il en est de même pour les Hautes-Pyrénées et le Tarn-et-Garonne.

L'Ariège et la Côte d'Or sont éloignés l'un de l'autre sur le graphique ce qui indiquent qu'ils ont des profils culturelles très différents.

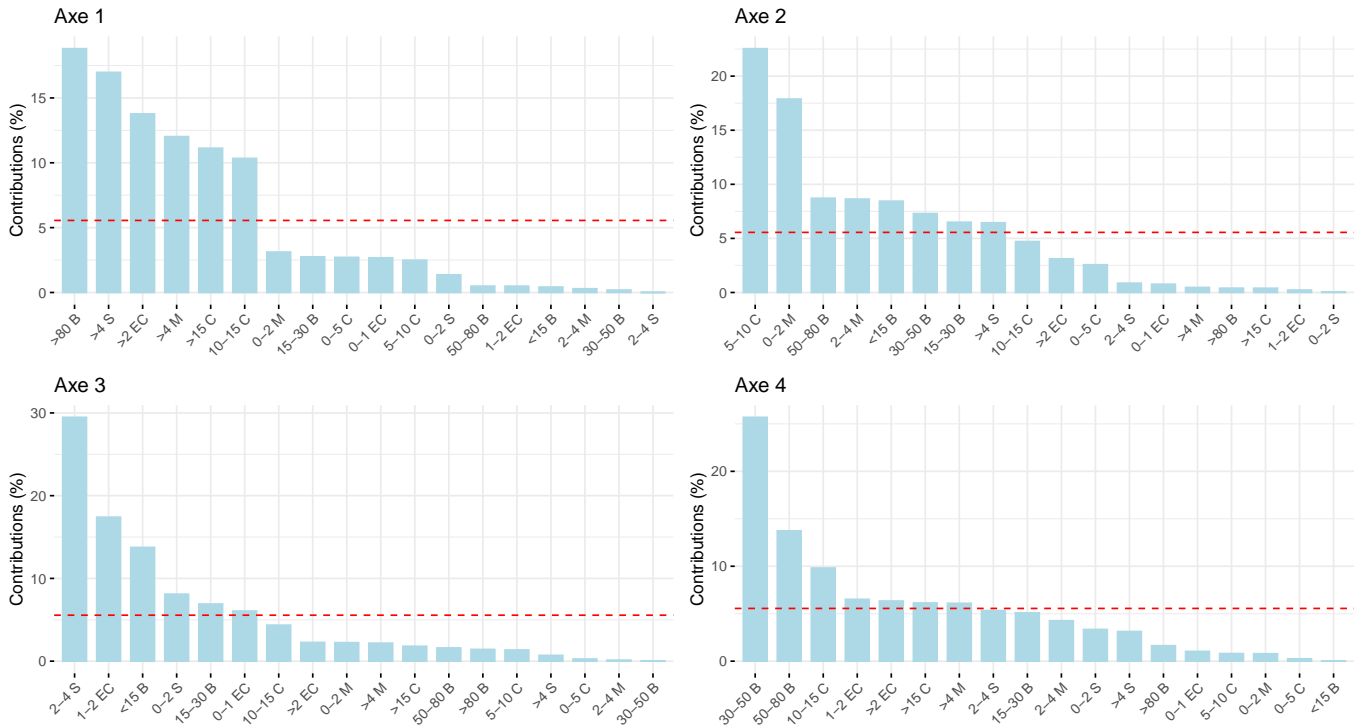


Les Axes 3 et 4 représentent 20,4 % de l'inertie totale, offrant ainsi une contribution importante à la compréhension des variations restantes dans les données, bien que leur part soit moins significative que celle des Axes 1 et 2.

On peut apercevoir que Côte-d'Or, les Yvelines, le Val-d'Oise, le Val-de-Marne, l'Essonne, la Seine-Saint-Denis et la Lozère sont les départements qui contribuent le plus à l'Axe  $F_3$  ; tandis que sur l'Axe  $F_4$  se sont la Côte-d'Or, le Territoire de Belfort et le Nièvre.

On remarque que le Val-de-Marne, les Yvelines, le Val-d'Oise, le Val-de-Marne, l'Essonne, la Seine-Saint-Denis ont des caractéristiques culturelles similaires ; alors que la Lozère et le Territoire de Belfort ont des profils totalement opposés.

### 3.4 Contribution des modalités

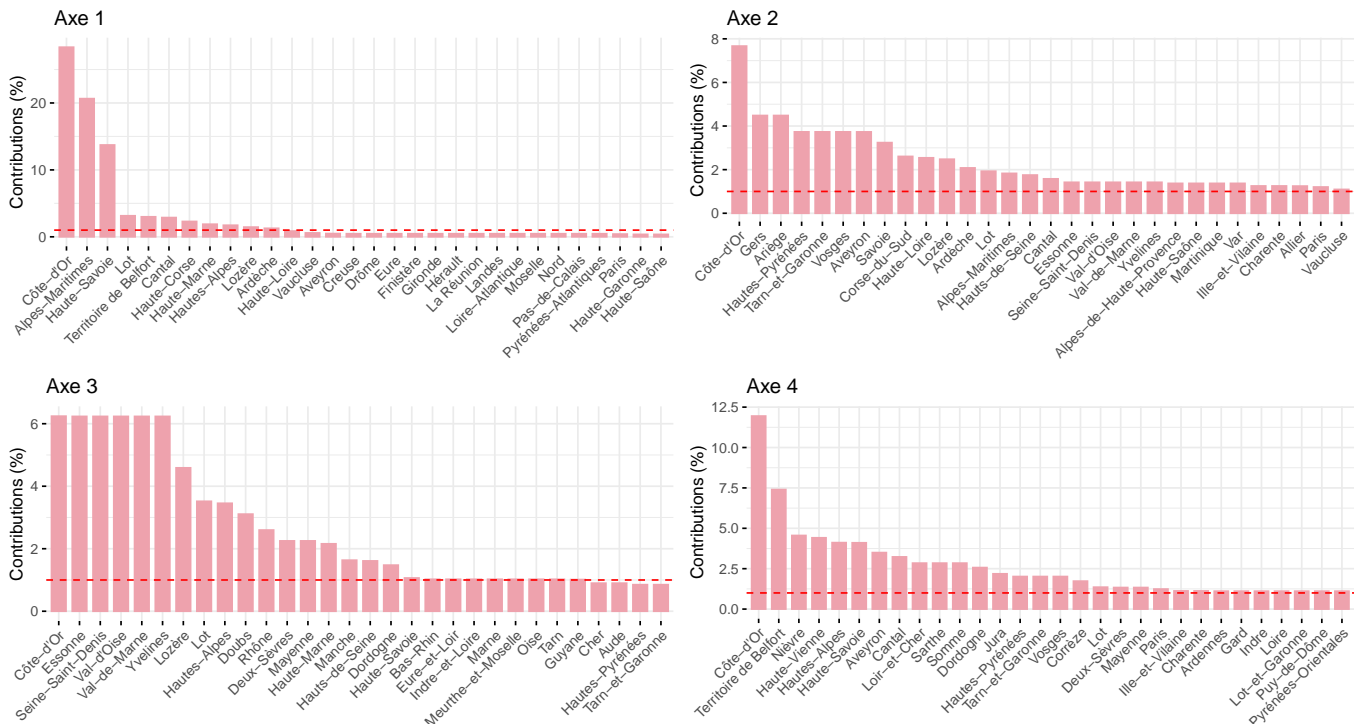


Les modalités qui contribuent le plus à l'axe  $F_1$  sont >80 B (bibliothèques), >4 S (spectacles), >2 EC (établissements culturels), >4 M, >15 C et 10-15 C.

Sur l'axe  $F_2$ , ce sont 5-10 C et 0-2 M qui contribuent le plus.

On observe que 2-4 S, 1-2 EC et <15 B contribuent plus que les autres modalités à l'axe  $F_3$ , alors que sur l'axe  $F_4$  ce sont 30-50 B, 50-80 B et 10-15 C.

### 3.5 Contribution des individus



Les départements qui contribuent le plus à l'axe  $F_1$  sont la Côte-d'Or, les Alpes-Maritimes et la Haute-Savoie on en déduit donc que ces départements sont caractérisés par un nombre très élevé d'équipements culturels pour 100 000 habitants ( $>80$  B,  $>4$  S,  $>2$  EC,  $>4$  M,  $>15$  C et 10-15 C).

Concernant l'axe  $F_2$ , se sont la Côte-d'Or, le Gers, l'Ariège, l'Aveyron, les Vosges, le Tarn-et-Garonne et les Hautes-Pyrénées qui contribuent plus que les autres départements à l'axe  $F_2$  on en conclut donc qu'ils sont associés à un nombre faible de cinémas et de musées pour 100 000 habitants (5-10 C et 0-2 M).

La Côte-d'Or, les Yvelines, le Val-d'Oise, le Val-de-Marne, l'Essonne, la Seine-Saint-Denis et la Lozère sont les départements qui contribuent le plus à l'Axe  $F_3$  ce qui révèle que ces départements ont tendance à avoir une offre culturelle faible (2-4 S, 1-2 EC et  $<15$  B).

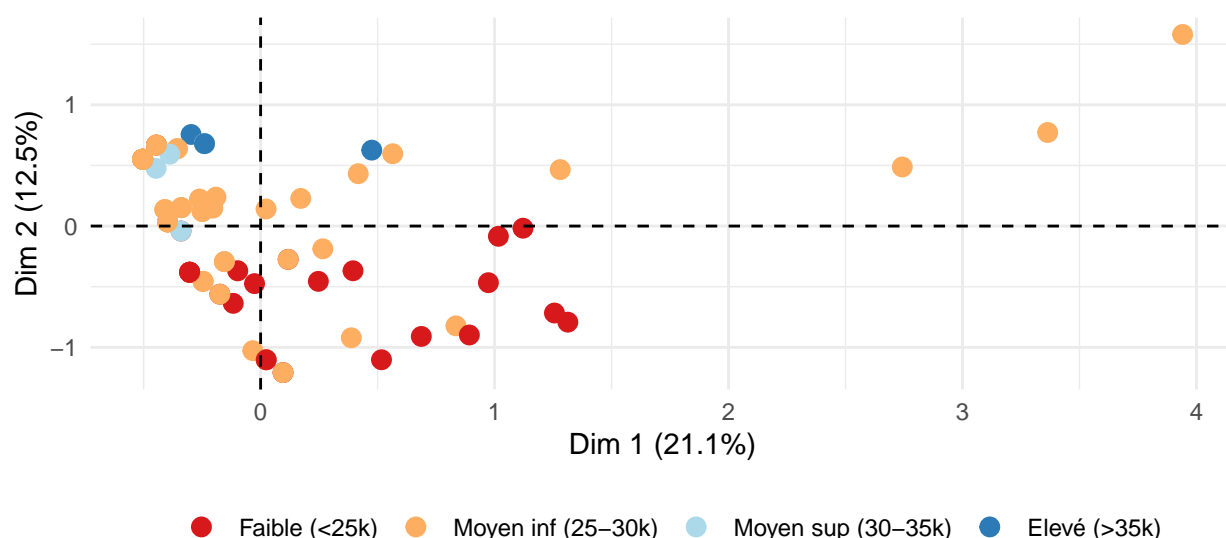
Enfin, on observe que sur l'axe  $F_4$ , la Côte-d'Or et le Territoire de Belfort se distinguent des autres départements par un nombre d'équipements culturels relativement élevée (30-50 B, 50-80 B et 10-15 C).

### 3.6 Projection des Variables Supplémentaire sur les Axes Culturels

Dans cette sous-partie, nous allons mettre en relation les résultats trouvés précédemment avec chacune des variables supplémentaires que nous possédons.

Nous choisissons d'utiliser uniquement les deux premiers axes pour représenter les variables supplémentaires, car ces axes capturent déjà une proportion significative de l'inertie (33,60%), offrant une visualisation claire et pertinente des relations entre ces variables sans introduire de complexité inutile.

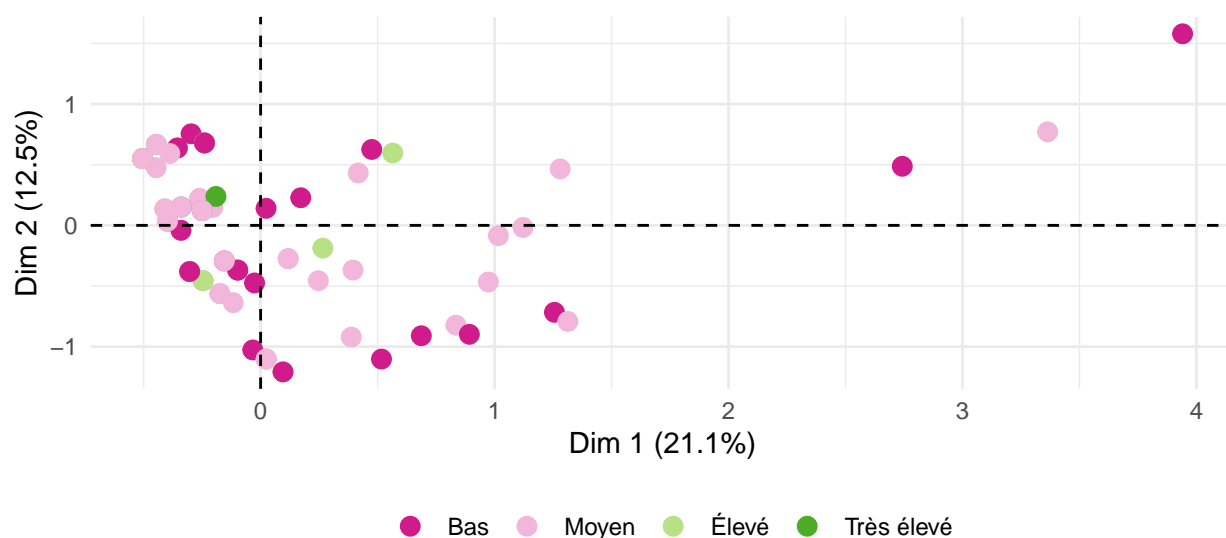
#### 3.6.1 Représentations des départements suivant leur salaire moyen net par an



On a vu précédemment que les départements de la Côte-d'Or, des Alpes-Maritimes et de la Haute-Savoie caractérisées par une offre culturelle riche, sont ceux qui contribuent le plus à l'axe  $F_1$ . On peut voir sur ce graphique que les points les plus éloignés de l'origine sur l'axe  $F_1$  sont associés à la modalité de salaire Moyen inférieur (25-30 k) ce qui suggère que les départements ayant une forte offre culturelle comme la Côte-d'Or tendent à être également liés à des niveaux de salaire légèrement inférieurs à la moyenne.

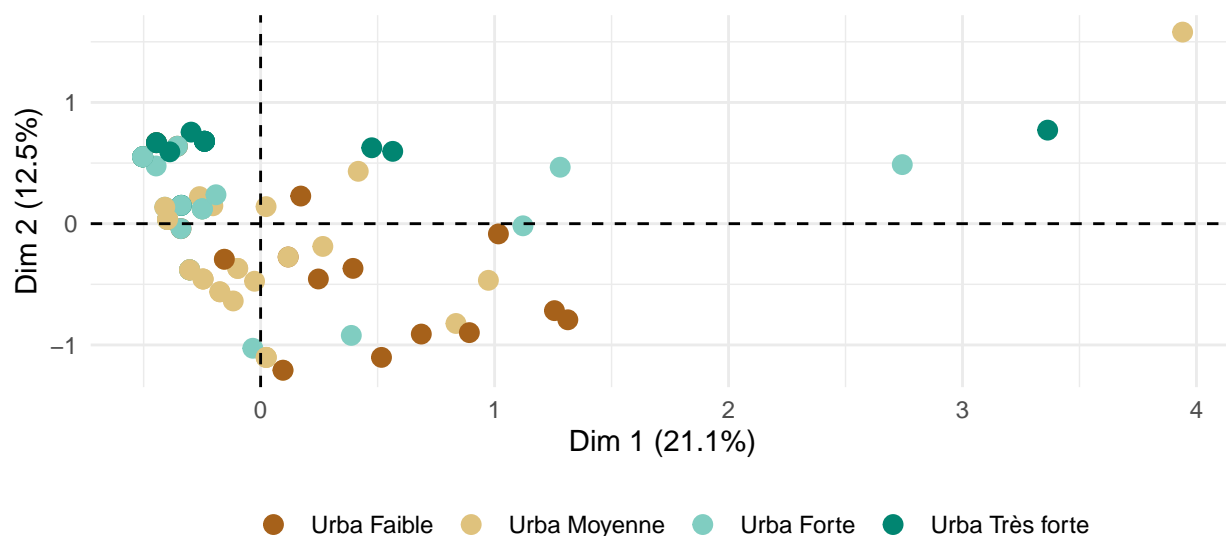
Des départements comme le Gers, l'Ariège, l'Aveyron, les Vosges, le Tarn-et-Garonne et les Hautes-Pyrénées contribuent de manière significative à l'axe  $F_2$ . Ils sont caractérisés par un nombre plus faible de cinémas et de musées pour 100 000 habitants. On observe que les points les plus éloignés de l'origine sur l'axe  $F_2$  sont associés aux modalités de salaire Moyen inférieur (25-30 k) et Faible (<25 k). On peut donc dire que les départements associés à une offre culturelle moins développée en termes de cinémas et de musées sont également liés à des niveaux de salaire moyen net par an plus bas.

### 3.6.2 Représentations des départements suivant leur taux de chômage



Les départements les plus éloignés de l'origine sur l'axe  $F_1$  sont associés à des taux de chômage bas et moyens, de même que ceux situés loin de l'origine sur l'axe  $F_2$ . Cela indique que les départements avec une offre culturelle riche, ainsi que ceux avec un nombre plus faible de cinémas et de musées pour 100 000 habitants, sont associés à des niveaux de chômage plus faibles ou modérés.

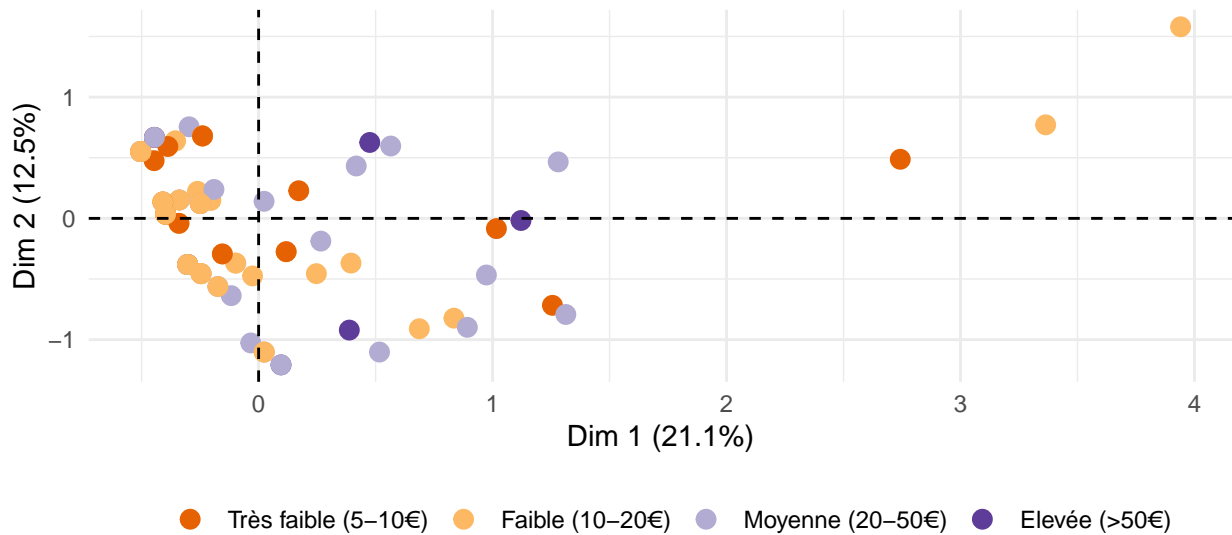
### 3.6.3 Représentations des départements suivant leur part de population urbaine



Le graphique montre que l'offre culturelle est liée à l'urbanisation : les départements avec une forte urbanisation (axe  $F_1$ ) ont une offre culturelle plus riche, tandis que ceux avec une urbanisation plus faible (axe  $F_2$ ) présentent une offre culturelle moins développée. Cette relation souligne qu'une plus grande urbanisation va de pair avec une plus grande densité d'équipements culturels.



### 3.6.4 Représentations des départements suivant leur dépenses culturelles par habitant

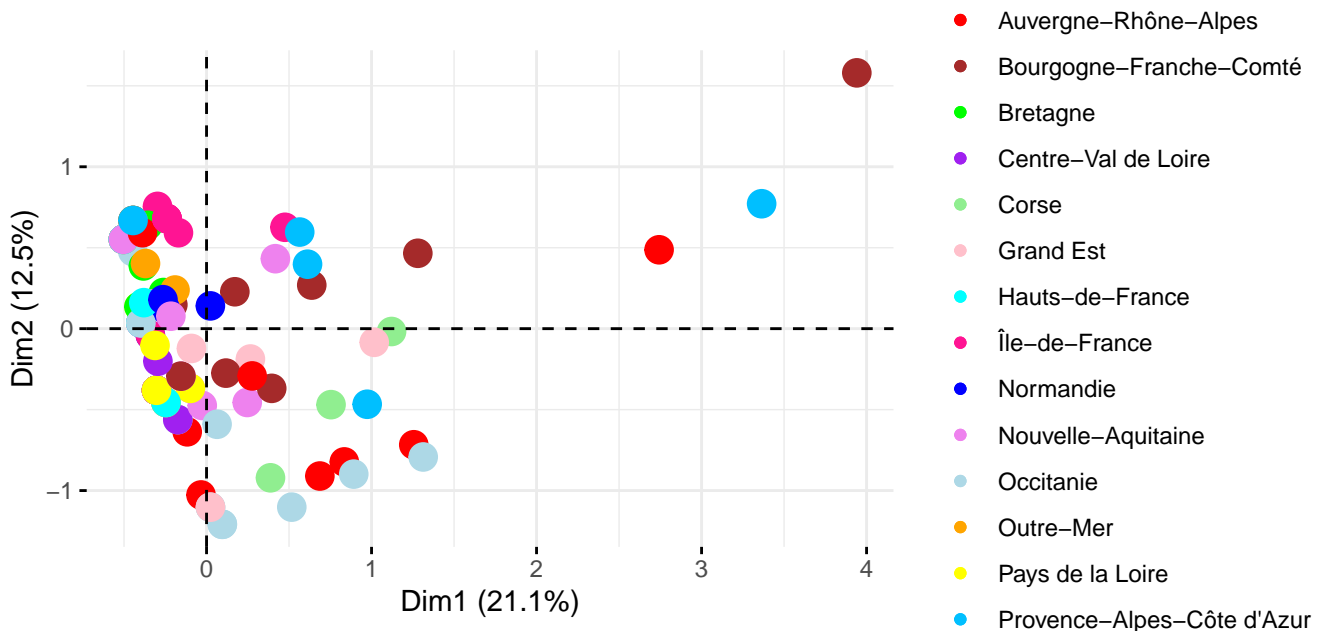


Les départements comme la Côte-d'Or, les Alpes-Maritimes et la Haute-Savoie, qui ont une forte offre culturelle, sont associés à des dépenses culturelles très faibles à faibles (5-20 €) sur l'axe  $F_1$ . Cela suggère que des départements avec une offre culturelle riche n'ont pas nécessairement des dépenses culturelles élevées.

Les départements comme la Côte-d'Or, le Gers, l'Ariège, etc., qui présentent une offre culturelle limitée (moins de musées et de cinémas), sont associés à des dépenses culturelles faibles à moyennes sur l'axe  $F_2$ . Cela montre que dans ces départements, une offre moins développée coïncide avec des dépenses culturelles plus modérées.

Ainsi, on observe que l'offre culturelle plus importante ne se traduit pas nécessairement par des dépenses culturelles par habitant plus élevées.

### 3.6.5 Représentations des départements suivant leurs régions

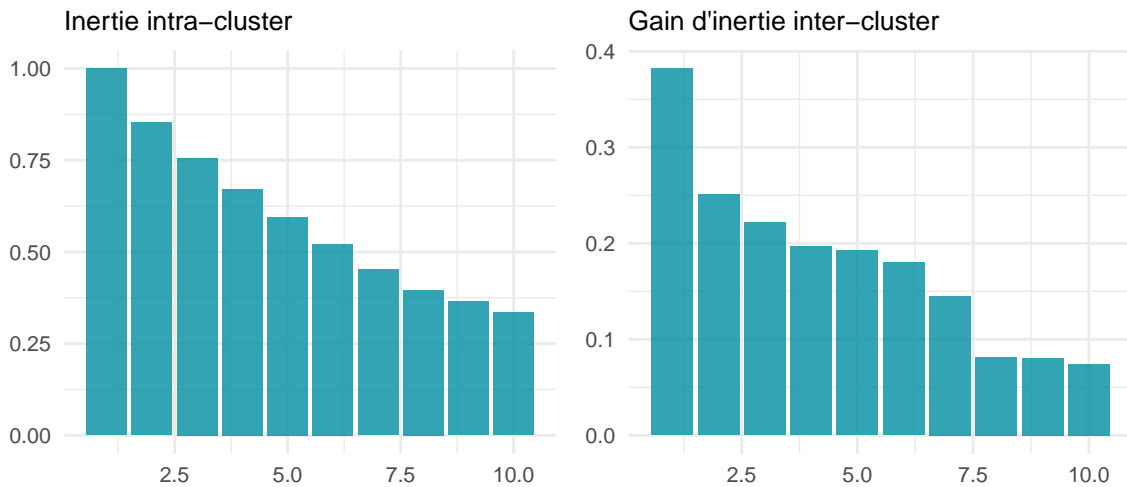


On remarque que les départements d'Ile-de-France sont tous regroupés en haut à gauche du plan  $F_1, F_2$  ce qui signifie qu'ils partagent des caractéristiques communes. En revanche les départements de Provinces-Alpes-Côte d'Azur ou encore ceux de Bourgogne-Franche-Comté qui sont éparpillés, présentent des profils plus diversifiés suggérant des variations plus importantes dans leurs caractéristiques culturelles.

## 4 Partie 3 : Clustering

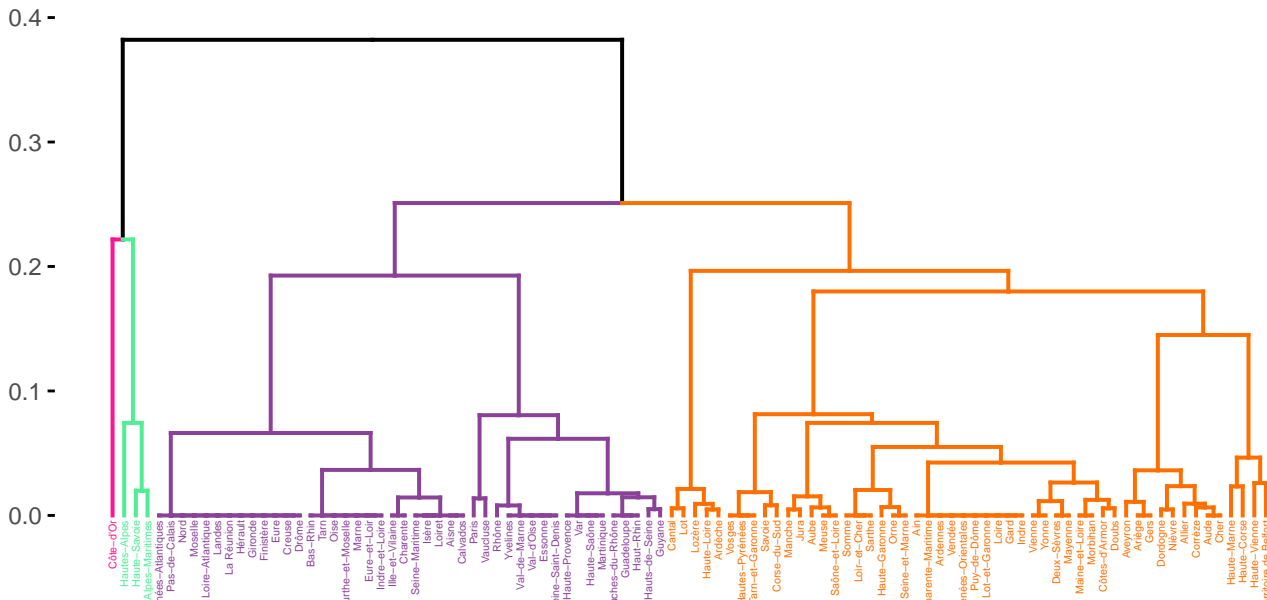
Dans cette section, nous allons procéder à un clustering basé sur les résultats de l'ACM présentée précédemment, dans le but d'identifier des groupes de départements présentant des caractéristiques similaires.

### 4.1 Analyse de l'inertie dans le clustering



On observe que le gain d'inertie inter-cluster reste relativement important jusqu'à 4 clusters, puis commence à diminuer plus lentement, suggérant que l'ajout de clusters au-delà de ce point n'améliore que marginalement la séparation des groupes. Ainsi, couper le dendrogramme en 4 clusters semble être la meilleure solution, car c'est à ce stade que la séparation des clusters est optimale.

### 4.2 Dendrogramme et Répartition des Clusters

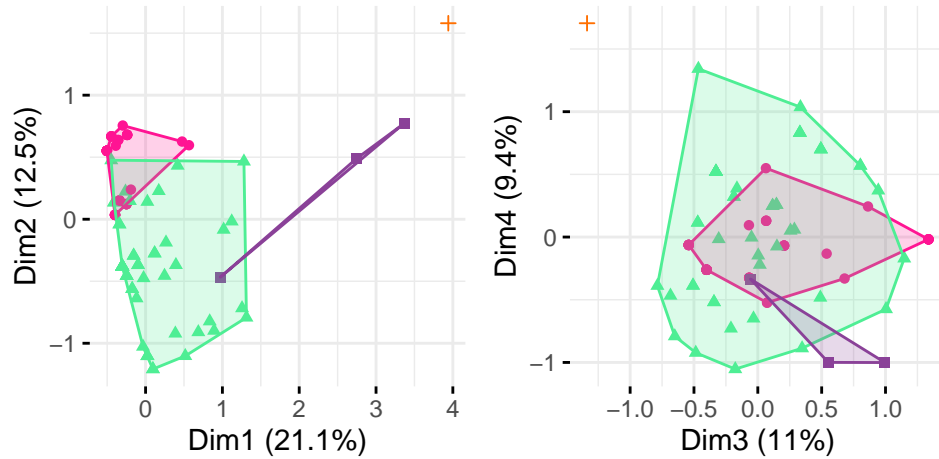


Le dendrogramme montre comment les départements se regroupent progressivement en fonction de leur similarité en termes d'équipements culturels pour 100 000 habitants. Les hauteurs de fusion reflètent l'intensité des différences : une faible hauteur indique des départements très similaires, tandis qu'une hauteur plus élevée signale des différences plus marquées.

- Cluster rose : Ce cluster contient seulement le département de la Côte-d'Or, qui se distingue par des caractéristiques très spécifiques et uniques en termes d'équipements culturels pour 100 000 habitants.

- Cluster vert : Ce cluster regroupe 3 départements, les Alpes-Maritimes et la Haute-Savoie qui montrent des similitudes en terme d'offres culturelles plus fortes entre elles, qu'avec les Hautes-Alpes.
- Cluster violet : Ce cluster comprend deux groupes de départements qui présentent des différences relativement subtiles au sein de chaque groupe. Cependant, les différences entre les deux groupes eux-mêmes sont nettement plus marquées.
- Cluster orange : Ce cluster est composé de nombreux petits groupes d'environ 4-5 départements qui sont très similaires entre eux en termes d'offre culturelle. Cependant, ces petits groupes sont plus éloignés les uns des autres et se rejoignent à des hauteurs plus élevées dans le dendrogramme. Cela indique que, bien que chaque sous-groupe soit homogène, les différences entre ces sous-groupes sont significatives.

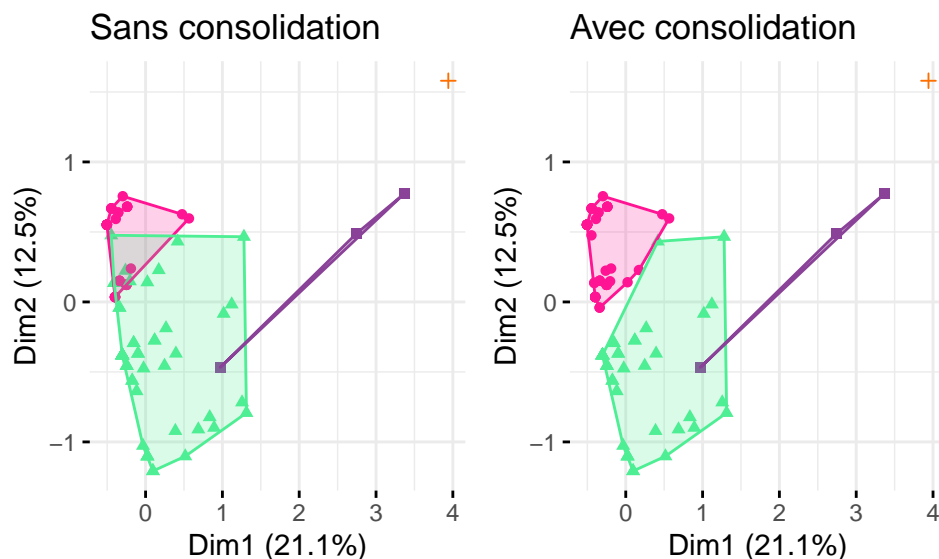
### 4.3 Visualisation des clusters sur les différents plans factoriels



On remarque que sur le plan  $F_1, F_2$ , le cluster orange est représenté par un seul point en haut à droite, ce qui suggère que ce groupe est très concentré et distinct des autres. On voit que les clusters rose et violet chevauchent légèrement le cluster vert, ce qui indique qu'il y a certains départements dans ces deux clusters qui présentent des caractéristiques similaires avec ceux du cluster vert.

Sur le plan  $F_3, F_4$ , tous les clusters se chevauchent, à l'exception du cluster orange, qui reste isolé. Cela suggère que ces axes ne permettent pas de bien séparer les différents clusters, rendant difficile la distinction des groupes. Ce phénomène est probablement dû à la faible proportion de variance capturée par ces deux axes, ce qui les rend moins discriminants.

### 4.4 Clustering Hiérarchique



Avant la consolidation, le cluster rose chevauchait le cluster vert, mais après consolidation, la forme du cluster vert a été ajustée pour éviter tout chevauchement avec le cluster rose, les séparant ainsi complètement. Ce changement améliore la structure des groupes, rendant les frontières entre ces deux clusters plus nettes. Les autres clusters restent inchangés, ce qui montre que la consolidation a permis de mieux distinguer le cluster vert du cluster rose sans affecter la répartition des autres groupes.

## 4.5 Tableau des données regroupées par clusters selon les variables

	Ciné 100k hab	Biblio 100k hab	Spectacle 100k hab	Etabl cult 100k hab	Musée 100k hab	Salaire moyen net an	Taux chômage	Part pop urbaine	Dépenses cult par hab	Région
<b>Cluster 1</b>										
Creuse	0-5 C	15-30 B	0-2 S	0-1 EC	0-2 M	Faible (<25k)	Moyen	Urba Faible	Elevée (>50€)	Nouvelle Aquitaine
Drôme	0-5 C	15-30 B	0-2 S	0-1 EC	0-2 M	Moyen inf (25-30k)	Moyen	Urba Forte	Faible (10-20€)	Auvergne Rhône Alpes
Eure	0-5 C	15-30 B	0-2 S	0-1 EC	0-2 M	Moyen inf (25-30k)	Moyen	Urba Moyenne	Faible (10-20€)	Normandie
Finistère	0-5 C	15-30 B	0-2 S	0-1 EC	0-2 M	Moyen inf (25-30k)	Moyen	Urba Forte	Moyenne (20-50€)	Bretagne
Gironde	0-5 C	15-30 B	0-2 S	0-1 EC	0-2 M	Moyen inf (25-30k)	Moyen	Urba Forte	Très faible (5-10€)	Nouvelle Aquitaine
<b>Cluster 2</b>										
Ain	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inf (25-30k)	Bas	Urba Moyenne	Faible (10-20€)	Auvergne Rhône Alpes
Charente- Maritime	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inf (25-30k)	Moyen	Urba Moyenne	Faible (10-20€)	Nouvelle Aquitaine
Gard	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inf (25-30k)	Élevé	Urba Forte	Faible (10-20€)	Occitanie
Indre	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Faible (<25k)	Moyen	Urba Moyenne	Faible (10-20€)	Centre Val de Loire
Loire	0-5 C	30-50 B	0-2 S	0-1 EC	2-4 M	Moyen inf (25-30k)	Moyen	Urba Forte	Très faible (5-10€)	Auvergne Rhône Alpes
<b>Cluster 3</b>										
Haute- Savoie	>15 C	>80 B	>4 S	1-2 EC	>4 M	Moyen inf (25-30k)	Bas	Urba Forte	Très faible (5-10€)	Auvergne Rhône Alpes
Alpes- Maritimes	>15 C	>80 B	>4 S	>2 EC	>4 M	Moyen inf (25-30k)	Moyen	Urba Très forte	Faible (10-20€)	Provence Alpes Côte d'Azur
Hautes- Alpes	>15 C	50-80 B	2-4 S	1-2 EC	2-4 M	Faible (<25k)	Moyen	Urba Moyenne	Moyenne (20-50€)	Provence Alpes Côte d'Azur
<b>Cluster 4</b>										
Côte-d'Or	10-15 C	>80 B	>4 S	>2 EC	>4 M	Moyen inf (25-30k)	Bas	Urba Moyenne	Faible (10-20€)	Bourgogne Franche Comté

Le cluster 1 (violet) regroupe des départements avec faible activité culturelle pour 100 000 habitants. Ces départements présentent des variations notables dans les dépenses culturelles par habitant, allant de faibles à élevées. Le salaire moyen net par an y est généralement faible ou moyen inférieur, tandis que le taux de chômage est modéré. La population est souvent majoritairement urbaine, ce qui contraste avec la faible offre culturelle.

Le cluster 2 (orange) est caractérisé par une activité culturelle faible, mais avec cette fois-ci, un accès plus important aux bibliothèques et aux musées par rapport au premier cluster. Les départements de ce cluster présentent généralement un revenu moyen net par an globalement moyen inférieur, avec un taux de chômage variant de bas à élevé. La part de la population urbaine est majoritairement importante, ce qui pourrait suggérer une meilleure accessibilité aux infrastructures culturelles. Cependant, les dépenses des départements dans ces infrastructures par habitant sont faibles, ce qui limite l'impact de cette accessibilité.

Le cluster 3 (vert) se distingue par une offre culturelle élevée, comprenant de nombreux cinémas, bibliothèques, spectacles et musées. Les départements de la Haute-Savoie, des Alpes-Maritimes et des Hautes-Alpes, fortement urbanisés, bénéficient d'une large accessibilité à ces équipements culturels. Le salaire moyen net annuel et le taux de chômage y sont généralement modérés, voire faibles, tout comme les dépenses culturelles des départements par habitant.

Le cluster 4 (rose) inclut exclusivement le département de la Côte-d'Or, qui se distingue par une offre culturelle riche. Ce département présente des dépenses culturelles des départements par habitant et un taux de chômage faibles, ainsi

qu'un salaire moyen net inférieur. Par ailleurs, la part de la population urbaine y est moyenne, ce qui peut influencer l'accès aux infrastructures culturelles.

## 4.6 Tableau des données regroupées par clusters selon la distribution des individus

	Ciné 100k hab	Biblio 100k hab	Spectacle 100k hab	Etabl cult 100k hab	Musée 100k hab	Salaire moyen net an	Taux chômage	Part pop urbaine	Dépenses cult par hab	Région
<b>Cluster 1</b>										
Paris	0-5 C	<15 B	>4 S	0-1 EC	2-4 M	Elevé (>35k)	Bas	Urba Très forte	Elevée (>50€)	Île de France
Vaucluse	0-5 C	15-30 B	>4 S	1-2 EC	2-4 M	Moyen inf (25-30k)	Élevé	Urba Très forte	Moyenne (20-50€)	Provence Alpes Côte d'Azur
Essonne	0-5 C	<15 B	2-4 S	1-2 EC	0-2 M	Moyen sup (30-35k)	Bas	Urba Très forte	Faible (10-20€)	Île de France
Seine-Saint-Denis	0-5 C	<15 B	2-4 S	1-2 EC	0-2 M	Moyen inf (25-30k)	Élevé	Urba Très forte	Faible (10-20€)	Île de France
Val-d'Oise	0-5 C	<15 B	2-4 S	1-2 EC	0-2 M	Moyen sup (30-35k)	Moyen	Urba Très forte	Faible (10-20€)	Île de France
<b>Cluster 2</b>										
Territoire de Belfort	0-5 C	30-50 B	>4 S	>2 EC	2-4 M	Moyen inf (25-30k)	Moyen	Urba Forte	Moyenne (20-50€)	Bourgogne Franche Comté
Haute-Corse	5-10 C	<15 B	0-2 S	>2 EC	>4 M	Faible (<25k)	Moyen	Urba Forte	Elevée (>50€)	Corse
Haute-Marne	0-5 C	50-80 B	0-2 S	>2 EC	>4 M	Faible (<25k)	Moyen	Urba Faible	Très faible (5-10€)	Grand Est
Lot	5-10 C	>80 B	2-4 S	1-2 EC	>4 M	Faible (<25k)	Moyen	Urba Faible	Moyenne (20-50€)	Occitanie
Cantal	5-10 C	>80 B	0-2 S	1-2 EC	>4 M	Faible (<25k)	Bas	Urba Faible	Très faible (5-10€)	Auvergne Rhône Alpes
<b>Cluster 3</b>										
Alpes-Maritimes	>15 C	>80 B	>4 S	>2 EC	>4 M	Moyen inf (25-30k)	Moyen	Urba Très forte	Faible (10-20€)	Provence Alpes Côte d'Azur
Haute-Savoie	>15 C	>80 B	>4 S	1-2 EC	>4 M	Moyen inf (25-30k)	Bas	Urba Forte	Très faible (5-10€)	Auvergne Rhône Alpes
Hautes-Alpes	>15 C	50-80 B	2-4 S	1-2 EC	2-4 M	Faible (<25k)	Moyen	Urba Moyenne	Moyenne (20-50€)	Provence Alpes Côte d'Azur
<b>Cluster 4</b>										
Côte-d'Or	10-15 C	>80 B	>4 S	>2 EC	>4 M	Moyen inf (25-30k)	Bas	Urba Moyenne	Faible (10-20€)	Bourgogne Franche Comté

Le premier cluster regroupe des départements très fortement urbanisés, principalement situés en Île-de-France, comme Paris, Essonne,..., ainsi que le Vaucluse en Provence-Alpes-Côte d'Azur. Ces territoires se caractérisent par une offre culturelle relativement faible pour 100 000 habitants avec notamment très peu de cinémas et de bibliothèques. Les salaires moyens nets par an sont globalement bons, variant de moyen inférieur (25-30k) à élevés (>35k), tandis que le taux de chômage oscille entre bas et élevé. Quant aux dépenses culturelles des départements par habitant, elles sont majoritairement faibles (10-20€), mais atteignent des niveaux moyens (20-50€) dans le Vaucluse et élevées (>50€) à Paris.

Le deuxième cluster regroupe des départements tels que le Territoire de Belfort, Haute-Corse, Haute-Marne, Lot, et Cantal, où l'offre culturelle pour 100 000 habitants est très diversifiée. On y trouve un grand nombre d'établissements culturels et de musées, mais relativement peu de cinémas et une offre de spectacles et de bibliothèques inégale. Les salaires moyens nets par an y sont principalement faibles (<25k), tandis que les dépenses culturelles des départements par habitant varient considérablement, allant de très faibles dans le Cantal (5-10€) à élevées en Haute-Corse (>50€). Le taux de chômage est principalement moyen, et la proportion de la population urbaine oscille entre forte et faible.

Comme décrit précédemment, les clusters 3 et 4 se caractérisent par une offre culturelle élevée, un salaire moyen net par an et un taux de chômage faible à moyen. Cependant, ils se distinguent par leur degré d'urbanisation : le cluster 3 correspondant aux Alpes-Maritimes, à la Haute-Savoie et aux Hautes-Alpes est globalement fortement urbanisé, tandis que le cluster 4 l'est de manière plus modérée. Concernant les dépenses culturelles des départements par habitant, le cluster 4 qui concerne uniquement la Côte-d'Or présente des dépenses faibles par habitant, tandis que celles du cluster 3 varient entre très faibles et moyennes.



## 4.7 Tableau des résultats du test du chi-deux

	p.value	df
Ciné_100k_hab	1.26e-42	9
Biblio_100k_hab	2.26e-17	12
Musées_100k_hab	1.48e-08	6
Spectacles_100k_hab	7.49e-07	6
Etablissements_cult_100k_hab	1.36e-04	6
Part_pop_urbaine	6.46e-04	9
Salaire_moyen_net_an	1.88e-03	9
Région	2.11e-03	39

On remarque que toutes les variables présentent des p-values inférieures à 0,05, indiquant des différences statistiquement significatives entre les clusters. Cela suggère que les caractéristiques culturelles et socio-économiques varient de manière marquée d'un cluster à l'autre.

Certaines variables, comme le nombre de cinémas et de bibliothèques, affichent des p-values beaucoup plus faibles, ce qui révèle des différences nettes dans la répartition de ces équipements culturels entre les clusters. À l'inverse, des variables telles que le salaire moyen net par an et la région ont des p-values plus élevées, mais restent significatives, ce qui montre que, bien que des différences existent, les salaires restent relativement homogènes entre les clusters. Ainsi, les différences les plus marquées concernent les indicateurs culturels.

## 4.8 Tableau des catégories descriptives du Cluster 1

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Biblio_100k_hab=15-30 B	100.0	62.3	33	6.867e-13	7.18
Musées_100k_hab=0-2 M	96.9	58.5	32	1.565e-10	6.40
Ciné_100k_hab=0-5 C	64.6	100.0	82	1.490e-07	5.25
Biblio_100k_hab=<15 B	94.1	30.2	17	1.146e-04	3.86
Part_pop_urbaine=Urba Très forte	93.8	28.3	16	2.403e-04	3.67
Région=Île-de-France	100.0	15.1	8	4.763e-03	2.82
Salaire_moyen_net_an=Moyen sup (30-35k)	100.0	11.3	6	1.926e-02	2.34
Salaire_moyen_net_an=Moyen inf (25-30k)	61.5	75.5	65	2.215e-02	2.29
Part_pop_urbaine=Urba Forte	71.4	37.7	28	2.316e-02	2.27
Région=Normandie	100.0	9.4	5	3.812e-02	2.07
Région=Auvergne-Rhône-Alpes	25.0	5.7	12	4.583e-02	-2.00
Région=Occitanie	23.1	5.7	13	2.463e-02	-2.25
Etablissements_cult_100k_hab=>2 EC	0.0	0.0	6	9.008e-03	-2.61
Musées_100k_hab=2-4 M	38.5	37.7	52	2.742e-03	-3.00
Biblio_100k_hab=>80 B	0.0	0.0	8	1.690e-03	-3.14
Part_pop_urbaine=Urba Moyenne	32.5	24.5	40	9.277e-04	-3.31
Musées_100k_hab=>4 M	12.5	3.8	16	4.112e-04	-3.53
Biblio_100k_hab=50-80 B	0.0	0.0	11	1.230e-04	-3.84
Ciné_100k_hab=5-10 C	0.0	0.0	14	7.732e-06	-4.47
Salaire_moyen_net_an=Faible (<25k)	15.4	7.5	26	7.355e-06	-4.48
Biblio_100k_hab=30-50 B	12.9	7.5	31	4.822e-08	-5.46

On observe que tous les départements présentant les modalités 15-30 bibliothèques, Île-de-France, Moyen supérieur (30-35k) et Normandie se retrouvent dans le Cluster 1. De plus, tous les départements de ce cluster affichent un nombre de cinémas pour 100 000 habitants compris entre 0 et 5, tandis que la majorité d'entre eux ont un salaire moyen net par an, moyen inférieur (entre 25 et 30k). Ce cluster n'est donc pas caractérisé par des départements ayant une offre culturelle abondante pour 100 000 habitants, comme le révèlent les zéros des modalités >80 B, >2 EC, 50-80 B et 5-10 C.

## 4.9 Tableau des catégories descriptives du Cluster 2

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Biblio_100k_hab=30-50 B	87.1	62.8	31	1.717e-09	6.02
Ciné_100k_hab=5-10 C	100.0	32.6	14	1.774e-06	4.78
Salaire_moyen_net_an=Faible (<25k)	80.8	48.8	26	7.569e-06	4.48
Musées_100k_hab=2-4 M	59.6	72.1	52	5.307e-04	3.46
Biblio_100k_hab=50-80 B	90.9	23.3	11	8.529e-04	3.34
Part_pop_urbaine=Urba Moyenne	62.5	58.1	40	1.560e-03	3.16
Région=Occitanie	76.9	23.3	13	1.073e-02	2.55
Part_pop_urbaine=Urba Faible	68.8	25.6	16	2.869e-02	2.19
Musées_100k_hab=>4 M	68.8	25.6	16	2.869e-02	2.19
Région=Provence-Alpes-Côte d'Azur	0.0	0.0	6	3.044e-02	-2.16
Salaire_moyen_net_an=Moyen sup (30-35k)	0.0	0.0	6	3.044e-02	-2.16
Part_pop_urbaine=Urba Forte	25.0	16.3	28	2.473e-02	-2.25
Salaire_moyen_net_an=Moyen inf (25-30k)	33.8	51.2	65	1.366e-02	-2.47
Région=Île-de-France	0.0	0.0	8	8.880e-03	-2.62
Ciné_100k_hab=0-5 C	35.4	67.4	82	1.328e-03	-3.21
Biblio_100k_hab=<15 B	5.9	2.3	17	4.164e-04	-3.53
Part_pop_urbaine=Urba Très forte	0.0	0.0	16	4.302e-05	-4.09
Musées_100k_hab=0-2 M	3.1	2.3	32	3.813e-09	-5.89
Biblio_100k_hab=15-30 B	0.0	0.0	33	2.553e-11	-6.67

Le cluster 2 regroupe tous les départements ayant la modalité 5-10 cinémas, ainsi que presque tous ceux ayant entre 50 et 80 bibliothèques. On note que 72,1 % des départements de ce cluster possèdent entre 2 et 4 musées. Ce cluster se caractérise donc par une offre culturelle modérée, tandis qu'une grande majorité des départements présentent des revenus moyens nets faibles par an.

## 4.10 Tableau des catégories descriptives du Cluster 3

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Ciné_100k_hab=>15 C	100.0	100.0	3	6.184e-06	4.52
Région=Provence-Alpes-Côte d'Azur	33.3	66.7	6	8.967e-03	2.61
Spectacles_100k_hab=>4 S	33.3	66.7	6	8.967e-03	2.61
Biblio_100k_hab=>80 B	25.0	66.7	8	1.662e-02	2.39
Etablissements_cult_100k_hab=0-1 EC	0.0	0.0	63	4.805e-02	-1.98
Spectacles_100k_hab=0-2 S	0.0	0.0	71	2.260e-02	-2.28
Ciné_100k_hab=0-5 C	0.0	0.0	82	5.046e-03	-2.80

Le cluster 3 se distingue par une forte concentration de cinémas, avec 100% des départements ayant plus de 15 cinémas pour 100 000 habitants, ce qui est une caractéristique exclusive à ce groupe. Cela indique que le cinéma est un facteur clé dans ce cluster. La région Provence-Alpes-Côte d'Azur est également sur-représentée, avec 66,7% des départements du cluster issus de cette région, ce qui souligne un lien géographique fort.

## 4.11 Tableau des catégories descriptives du Cluster 4

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Ciné_100k_hab=10-15 C	100	100	1	0.01	2.58

Le cluster 4 composé uniquement de la Côte-d'Or, est le seul département à présenter une densité de cinémas entre 10 et 15 pour 100 000 habitants. Cette modalité unique caractérise ce cluster. Cela met en évidence la singularité de la Côte-d'Or en termes d'offre cinématographique dans cette analyse.

## 5 Conclusion

L'ensemble des analyses (ACP, ACM et Clustering) a permis d'explorer en profondeur la relation entre les caractéristiques culturelles et socio-économiques des départements français. Le clustering, en particulier, nous a permis de segmenter les départements en quatre groupes distincts, facilitant ainsi l'analyse des tendances spécifiques à chaque groupe.

Plusieurs tendances clés se sont dégagées, notamment l'association complexe entre l'urbanisation et l'offre culturelle. Dans les zones fortement urbanisées, comme Paris ou l'Île-de-France, bien qu'il y ait une grande concentration d'équipements culturels, la densité de population élevée entraîne une offre culturelle relativement faible par 100 000 habitants. Cela met en évidence l'importance de la densité de la population dans l'interprétation de l'offre culturelle. Par conséquent, les départements très urbanisés peuvent sembler moins riches culturellement lorsqu'on mesure l'offre par rapport à la population, même s'ils disposent de nombreux équipements en nombre absolu. À l'inverse, dans des zones moins urbanisées, bien que le nombre total d'équipements soit plus modeste, l'offre culturelle par 100 000 habitants est souvent plus élevée en raison d'une population moins dense. Cette tendance est notamment observée dans des départements comme Paris, où l'offre culturelle par 100 000 habitants est faible malgré sa forte urbanisation, et la Côte-d'Or, qui affiche une offre culturelle plus riche par 100 000 habitants, bien qu'elle soit moyennement urbanisée.

En parallèle, les analyses ont révélé que les départements affichant des taux de chômage plus faibles ou modérés sont généralement associés à une offre culturelle plus riche par 100 000 habitants. Cela suggère qu'une meilleure accessibilité aux équipements culturels pourrait avoir un impact positif sur la dynamique économique locale, contribuant à des conditions économiques plus stables.

D'autre part, une tendance intéressante émerge entre l'offre culturelle et les salaires moyens : les départements à l'offre culturelle pour 100 000 habitant plus faible, souvent fortement urbanisés (comme l'Île-de-France), présentent des salaires moyens net par an plus élevés (25k€ à plus de 35k€), tandis que ceux avec une offre culturelle plus riche tendent à afficher des salaires moyens plus faibles (souvent inférieur à 25k€). Cela suggère que ces départements, bien que culturellement riches, sont souvent moins urbanisés et présentent une économie moins dynamique.

Cependant, malgré ces observations, nous n'avons pas réussi à dégager de tendances claires entre l'offre culturelle et les dépenses culturelles des collectivités territoriales par habitant, ces dernières variant considérablement d'un département à l'autre sans lien évident avec les autres variables.