

Etude des biens immobiliers à Saragota en Californie

BOUSSENGUI François - LEGER Aline - NIVELLE Jean-Baptiste

Table des matières

1	Présentation de la base de données	2
1.1	Description de la base	2
2	Caractéristiques des biens immobiliers	3
2.1	Matrice des corrélations	3
2.2	Etude des inerties	3
2.3	Etude des variables	4
2.4	Etude des individus	5
3	Mise en lumière d'un effet de taille	7
3.1	Effectifs et contributions	7
3.2	Etude des inerties	8
3.3	Représentations	8
4	Catégorisation des biens immobiliers	9
4.1	Etudes des inerties	9
4.2	Etude des individus et des modalités des variables	10
4.3	Représentation simultannée des individus et des variables	11
4.4	Approfondissement	12
5	Classification des biens immobiliers	13
5.1	Répartitions des biens en classes	13
5.2	Etude des inerties	13
5.3	Visualisation des groupes	14
5.4	Description des classes	14
5.5	Méthode de partitionnement	16
5.6	Comparaison des groupes réalisés avec k-means et classification hiérarchique	16
6	Annexes	17
6.1	Analyse descriptive des variables	17
6.2	Mise en relation des variables - Statistiques inférentielles	21
6.3	Analyse en composantes principales	23
6.4	Analyse factorielle des correspondances	25
6.5	Analyse des correspondances multiples	26

1 Présentation de la base de données

1.1 Description de la base

Cette base de données contient des informations sur **1728** biens immobiliers présents à Saratoga en Californie. Nous avons **18** variables qui nous apportent de l'information sur ces biens. Ces variables sont :

- **Prix** : prix du bien immobilier (\$) ;
- **Lot.size** : taille du terrain (pied carré) ;
- **Vue sur mer** : variable indiquant si le bien a vue sur la mer ou non ;
- **Age** : âge du bien immobilier ;
- **Valeur du terrain** : valeur du terrain où se situe le bien (\$) ;
- **Nouvelle construction** : variable indiquant si le bien est une nouvelle construction ou non ;
- **Air central** : variable indiquant si le logement a l'air centralisé ;
- **Type de combustible** : variable indiquant le type de combustible présent ;
- **Chauffage** : variable indiquant le type de chauffage présent dans le logement ;
- **Egouts** : variable indiquant si les égouts sont privés ou publics ;
- **Surface habitable** : Superficie totale disponible pour l'aménagement d'un loyer ;
- **Pct.college** : pourcentage des personnes du voisinage ayant obtenu un diplôme du supérieur ;

- **Chambres** : nombre de chambres dans le bien immobilier ;
- **Salle de bain** : nombre de salles de bains dans le bien immobilier ;
- **Pièces** : nombre de pièces du bien immobilier ;
- **Cheminées (x2)** : nombre de cheminées présentes dans le logement ;

De plus, cette base de données contient **0** valeurs manquantes.

Les biens immobiliers de la ville de Saragota présentent des caractéristiques différentes. Ces caractéristiques peuvent influencer le prix des ces biens. Il est donc intéressant d'analyser les données à notre disposition afin de pouvoir établir des groupes de biens, identifier leurs caractéristiques et étudier comment celles-ci peuvent influencer leur prix.

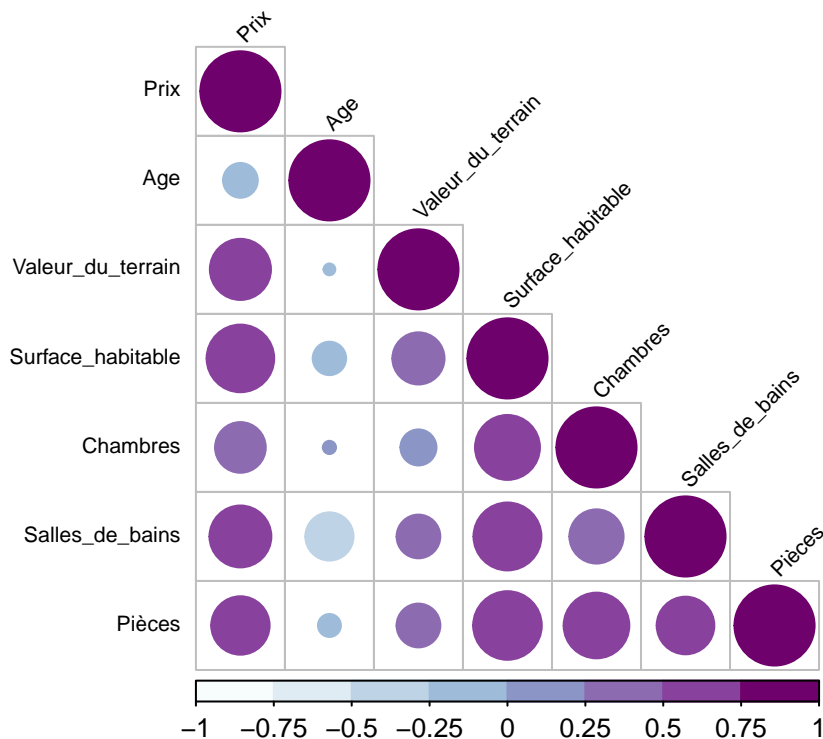
Pour ce faire, au cours de notre étude nous allons procéder à des analyses factorielles ainsi qu'à des méthodes de classification non supervisée.

2 Caractéristiques des biens immobiliers

Dans cette partie, nous allons tenter de synthétiser l'information apportée par les variables quantitatives de notre jeu de données en déformant le moins possible la réalité de nos données. Pour cela, nous allons utiliser l'analyse en composantes principales.

Les variables actives dans notre analyse sont: “Prix”, “Age”, “Valeur_du_terrain”, “Surface_habitable”, “Chambres”, “Salles_de_bain” et “Pièces”. Une variable qualitative sera mise en variable supplémentaire : “Lits”.

2.1 Matrice des corrélations

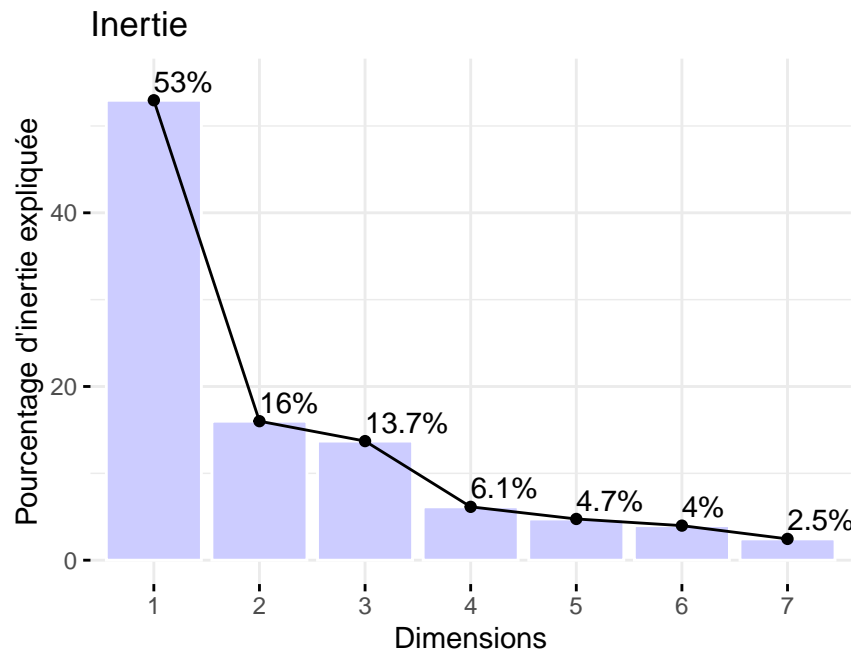


Globalement, les variables corrélées entre elles sont celles qui font référence aux pièces. Aussi, la variable “Pièces” est corrélée positivement (0.73) avec la variable “Surface habitable”. De plus, la variable “Prix” entretient une corrélation élevée avec la variable “Surface habitable” (0.71) et une corrélation un peu plus faible avec “Valeur du terrain” (0.58) et “Pièces” (0.53).

2.2 Etude des inerties

TABLE 1 – Inerties

	1	2	3	4	5	6	7
Inerties	3.70739	1.120435	0.9598722	0.4300206	0.3322842	0.2784664	0.1715315
Inerties relatives %	52.96271	16.006220	13.7124604	6.1431511	4.7469173	3.9780908	2.4504504
Inerties relatives cumulées	52.96271	68.968930	82.6813904	88.8245415	93.5714589	97.5495496	100.0000000



Nous avons décidé de conserver les deux premières dimensions qui cumulent 69% de l'inertie totale, c'est-à-dire que 69% de l'information du tableau des données est contenue dans les deux premiers axes.

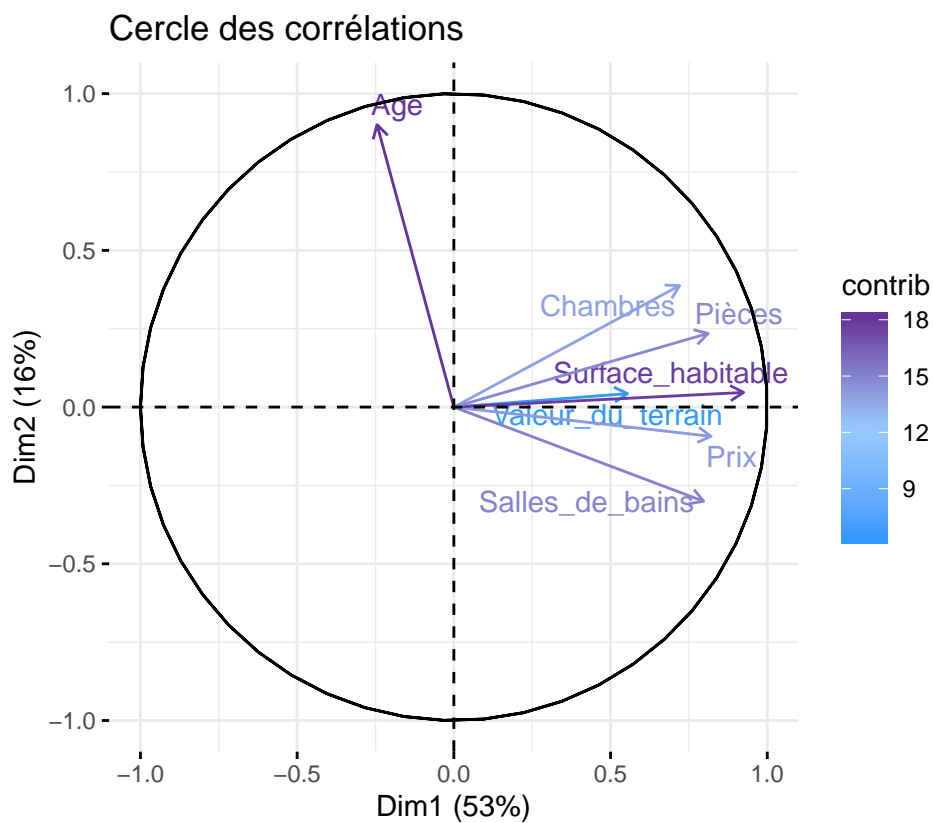
2.3 Etude des variables

TABLE 2 – Caractéristiques des variables

	F1	F2	QualF1	QualF2	QualF1F2	ContribF1	ContribF2
Prix	0.821	-0.093	0.674	0.009	0.682	18.170	0.778
Age	-0.245	0.901	0.060	0.812	0.872	1.621	72.449
Valeur_du_terrain	0.554	0.043	0.307	0.002	0.309	8.272	0.165
Surface_habitable	0.924	0.047	0.854	0.002	0.857	23.047	0.195
Chambres	0.721	0.388	0.519	0.150	0.669	14.003	13.416
Salles_de_bains	0.797	-0.301	0.635	0.090	0.726	17.134	8.064
Pièces	0.811	0.235	0.658	0.055	0.713	17.753	4.933

Concernant la dimension 1, les variables qui contribuent le plus sont : “Surface habitable”, “Prix”, “Pièces”, “Salles de bains” et “Chambres”.

Concernant la dimension 2, une seule variable contribue fortement, il s'agit de la variable “Age”.

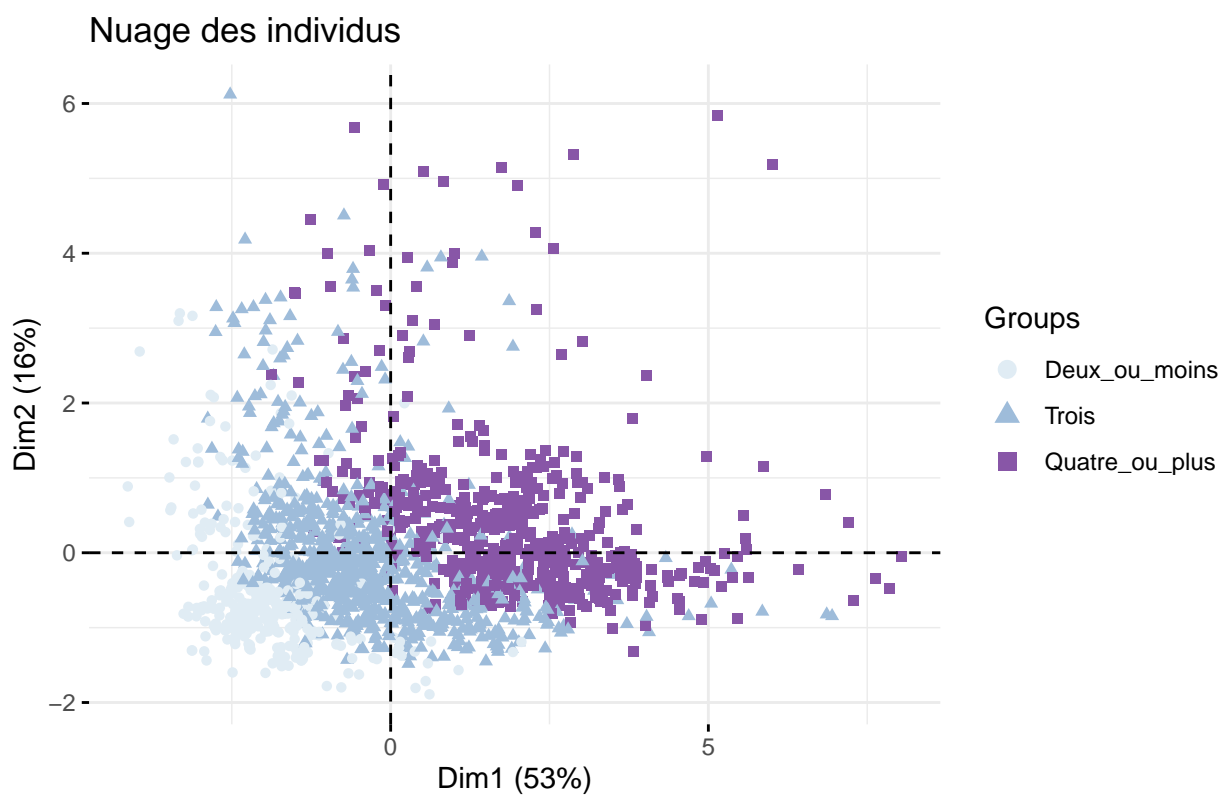


Les variables corrélées positivement avec l'axe 1 dans l'ordre décroissant sont :

“Surface habitable”, “Prix”, “Pièces”, “Salles de bain”, “Chambres”. Puis la variable “Valeur du terrain” moyennement.

La variable corrélée positivement avec l'axe 2 est “Age”.

2.4 Etude des individus



Au vu des variables qui sont corrélées à la dimension 1, nous pouvons dire que les biens immobiliers situés sur la partie positive de F1 sont caractérisés comme ayant une grande surface, un grand nombre de pièces et un prix élevé. Inversement concernant les biens situés sur la partie négative de F1. Ils sont caractérisés par le fait d’avoir une plus petite surface habitable, un nombre de pièces relativement faible et des prix qui ne sont pas hors normes.

Grâce à la variable “Age”, nous pouvons dire que la dimension 2 sépare les biens immobiliers âgés et ceux qui le sont moins. Plus les biens ont une coordonnée élevée sur F2, plus ils sont âgés.

L’ajout de la variable “Lit” en supplémentaire permet de diviser le nuage des individus en trois groupes. Ces trois groupes s’organisent autour de la première bissectrice.

- Le premier groupe faisant référence aux biens ayant deux lits ou moins se trouve dans la partie négative de F1 et F2. Cela peut se traduire par le fait que ces biens sont peu cher, ont peu de pièces, une petite surface habitable et sont peu âgés.
- Le deuxième groupe faisant référence aux biens ayant trois lits sont partagés à la fois dans la partie négative des deux axes et aussi dans leur partie positive. Ces biens ayant trois lits peuvent donc être caractérisé par une surface habitable moyenne, un prix moyen et un âge relativement "moyen".
- Le troisième groupe faisant référence aux biens ayant quatre lits ou plus se trouve dans la partie positive de F1 et à la fois dans la partie positive et négative de F2. Ces biens sont caractérisés comme étant plus grand, plus cher et en général plus âgés que les autres.

2.4.1 Individus les plus contributifs

— Dimension 1 :

	Prix	Age	Valeur_du_terrain	Chauffage	Surface_habitable	Chambres	Salles_de_bains	Pièces
591	620000	14	125100	Air_chaud	451.1	5	4	12

— Dimension 2 :

	Prix	Age	Valeur_du_terrain	Chauffage	Surface_habitable	Chambres	Salles_de_bains	Pièces
63	82800	225	3900	Air_chaud	137.8	3	1	7

Nous avons extrait l’individu qui contribue le plus à chaque dimension. Nous voyons au vu de leurs caractéristiques que nos propos sont confirmés.

3 Mise en lumière d'un effet de taille

Dans cette partie, nous avons décidé de mettre en relation deux variables qualitatives à savoir : “Pièces” et “Chauffage”. Nous avons choisis ces deux variables car nous pensons qu’il existe un lien clair entre ces deux variables.

Pour ce faire, nous avons réalisé une analyse factorielle des correspondances afin de déceler et mettre en évidence la liaison qu’il pourrait exister entre ces deux variables.

3.1 Effectifs et contributions

TABLE 3 – Tableau des effectifs (pièces en fonction du type de chauffage)

	Nombre de pièces										
	10	11	12	2	3	4	5	6	7	8	9
Electrique	10	2	2	0	33	54	52	47	57	35	13
Air_chaud	96	54	57	2	36	104	133	166	197	176	100
Eau_chaude	35	16	18	0	11	23	38	34	51	47	29

TABLE 4 – Effectifs attendus - Modèle d'indépendance

	Nombre de pièces										
	10	11	12	2	3	4	5	6	7	8	9
Electrique	24.9	12.7	13.6	0.4	14.1	31.9	39.4	43.6	53.8	45.5	25.1
Air_chaud	91.5	46.7	50.0	1.3	51.9	117.4	144.7	160.2	197.9	167.4	92.1
Eau_chaude	24.6	12.6	13.5	0.3	14.0	31.6	39.0	43.2	53.3	45.1	24.8

Nous pouvons remarquer que les effectifs sous le modèle d'indépendance sont différents des nos effectifs.

Après avoir calculé le tableau des fréquences à partir du tableau des effectifs, on effectue le test du Khi2 dessus. On obtient les résultats suivants :

TABLE 5 – Test du khi2

Valeurs	
Statistique	105.81
DDL	20
P-value	1.13e-13
Alpha	0.05

Au vue des résultats du test, nous pouvons dire qu’au seuil de 5% nous pouvons rejeter Ho. Il n’y a donc pas indépendance entre les deux variables. Il existe donc une liaison entre ces deux variables que l’analyse factorielle par correspondances va nous permettre d’identifier.

TABLE 6 – Contributions au Khi2

	Nombre de pièces											Somme
	10	11	12	2	3	4	5	6	7	8	9	
Electrique	8.9052902	9.0230874	9.8851720	0.3530093	25.2429933	15.222549	4.0587901	0.2656818	0.1862044	2.438690	5.8064882	81.38796
Air_chaud	0.2242963	1.1383066	0.9944505	0.3804153	4.8701374	1.533685	0.9407706	0.2073765	0.0037527	0.444822	0.6742003	11.41221
Eau_chaude	4.3535067	0.9277042	1.5335499	0.3495370	0.6357861	2.356090	0.0243106	1.9470288	0.0996213	0.080883	0.7050132	13.01303
Somme	13.4830932	11.0890982	12.4131724	1.0829616	30.7489169	19.112324	5.0238714	2.4200872	0.2895784	2.964395	7.1857017	105.81320

3.2 Etude des inerties

TABLE 7 – Inerties

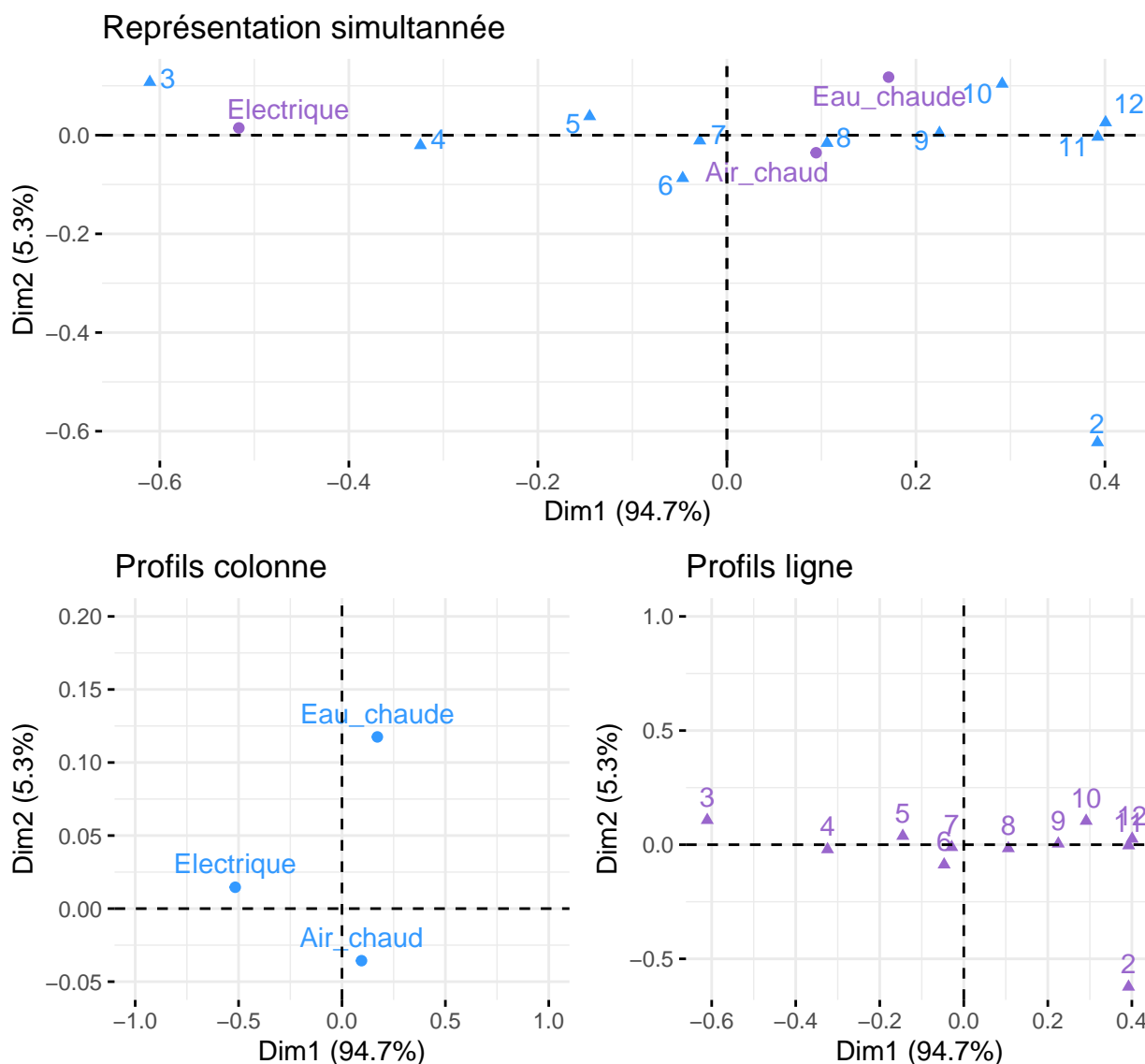
	Valeurs propres	Inerties relatives (%)	Inerties relatives cumulées (%)
Dim.1	0.0579591	94.651018	94.65102
Dim.2	0.0032754	5.348982	100.00000

Note:

Le 1er axe apporte presque 95% de l'information.

Etant donnée que la variable “Chauffage” possède que 3 modalités, seul un axe est nécessaire pour séparer ces modalités. Les deux axes cumulent la totalité de l'information. La dimension 2 apporte que très peu d'informations.

3.3 Représentations



Pour l'interprétation, on s'intéresse seulement à la dimension 1. Plus les coordonnées sont élevées dans la dimension 1, plus on trouve des biens avec beaucoup de pièces. On a un effet de taille. La représentation simultannée montre que les biens de petites tailles sont associés à un type de chauffage électrique et que ceux de grandes tailles sont associés aux chauffages à l'air chaud et à l'eau chaude.

En effet, en réalité on remarque que les logements de petites tailles sont chauffés à l'électricité dû à son coût d'utilisation faible. Les deux autres types de chauffage sont habituellement adaptés aux grandes surfaces.

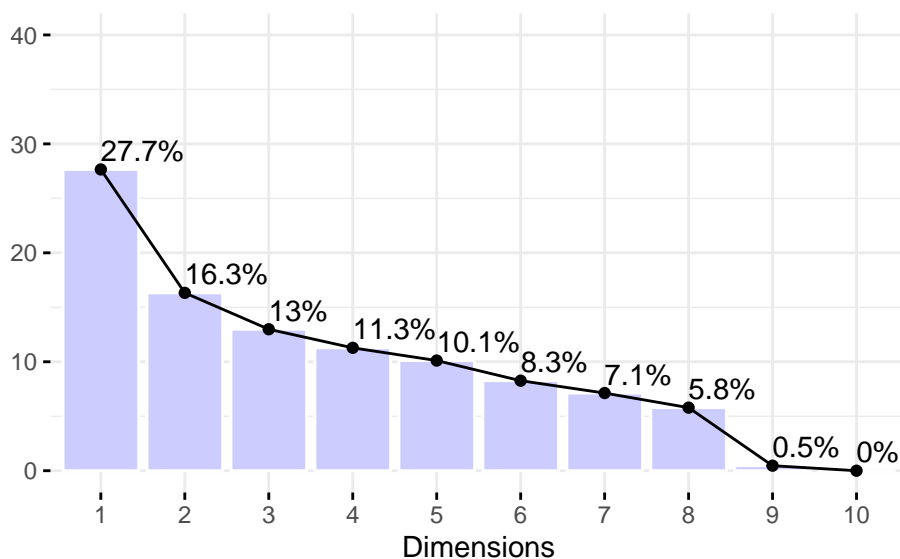
4 Catégorisation des biens immobiliers

Afin de réaliser une catégorisation des biens immobiliers en fonction de leurs caractéristiques, nous avons effectué une analyse des correspondances multiples. Son objectif est de visualiser les profils de biens immobiliers qui sont très proches, c'est-à-dire de même catégorie et ceux qui ne le sont pas.

Les variables actives pour l'analyse sont : "Air central", "Type de combustible", "Type de chauffage", "Type d'égoûts", "Cheminées1", "Cheminées2", "Lits". Nous avons rajouté des variables quantitatives supplémentaires qui sont : "Prix" et "Pièces" que nous avons découpées en classe.

4.1 Etudes des inerties

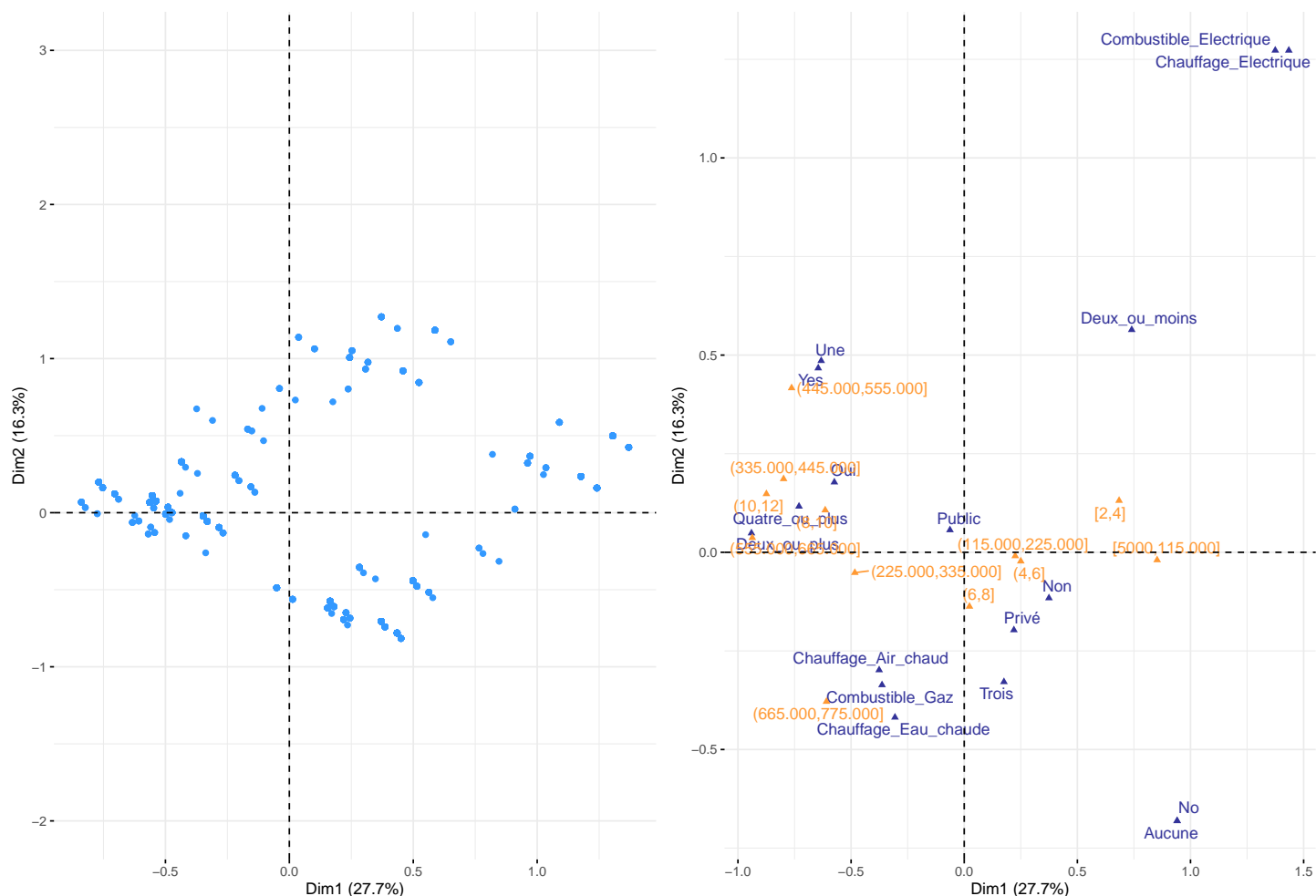
Visualisation des inerties



Les deux premiers axes cumulent 44% de l'inertie totale. Autrement dit, 44 % de l'information du jeu de données est resumée par les deux premières dimensions. Cela est relativement important dans le cadre d'une analyse des correspondances du fait de la diversité des profils d'individus. C'est pourquoi, dans notre analyse nous nous contenterons d'exploiter ces deux axes pour mettre en évidence ce qui se passe dans nos données.

4.2 Etude des individus et des modalités des variables

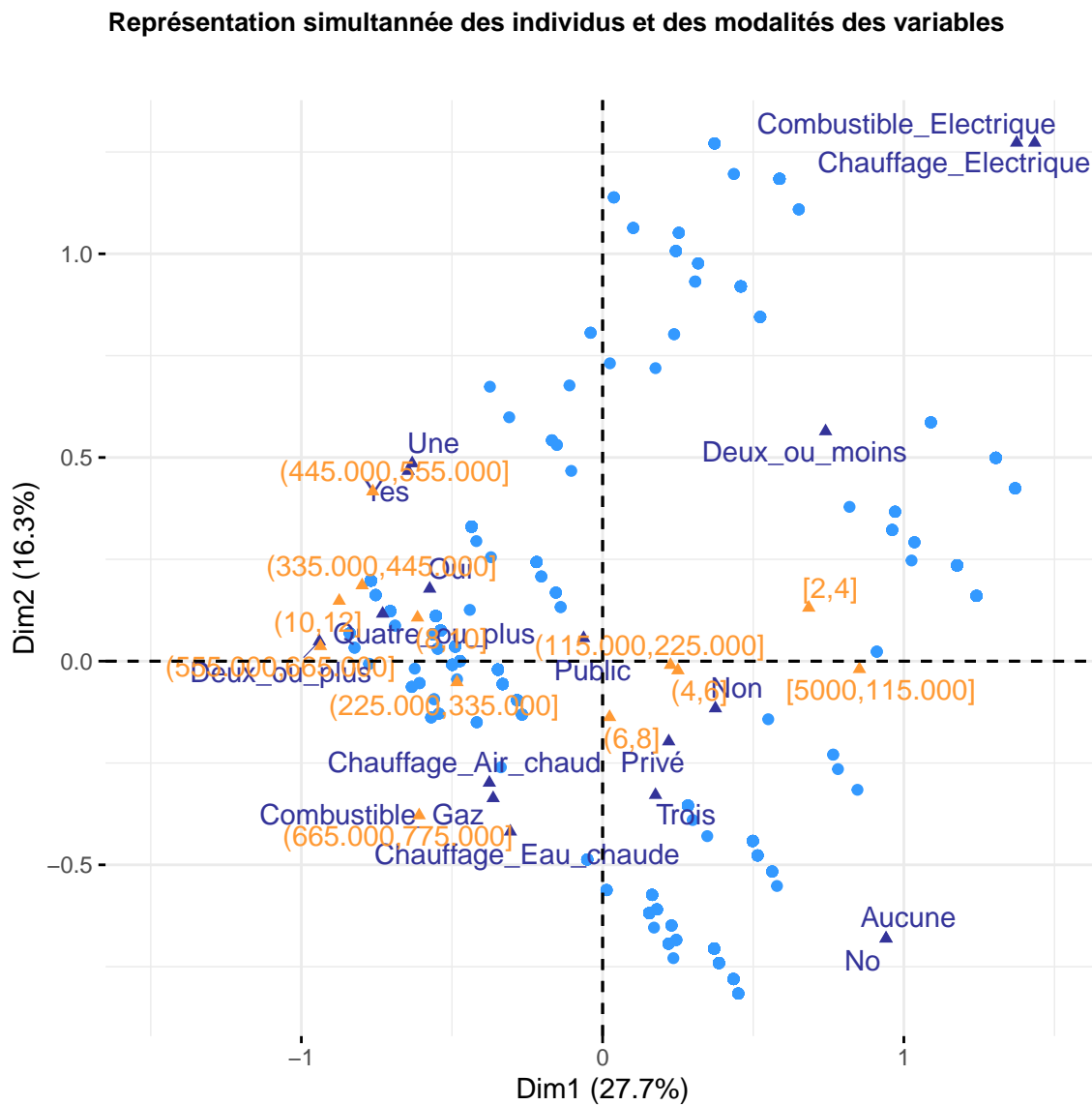
Nuage des individus et des modalités des variables sur les plans F(1,2)



Au vu de la forme de nos nuages, nous allons les interpréter en fonction des bissectrices. Grâce au graphique des modalités des variables nous pouvons dire que la première bissectrice sépare les différents types de chauffage et de combustible. Dans la partie haute du graphique, on trouve les biens chauffés à l'électrique et dans la partie basse les biens chauffés à l'air chaud et à l'eau chaude. Concernant la seconde bissectrice, elle sépare les biens possédant une ou plusieurs cheminées situés dans la partie haute et ceux n'en possédant pas situés dans la partie basse.

Grâce aux variables signalétiques "Prix" et "Pièces", on voit que les biens de plus petite taille qui sont donc les biens le moins cher n'ont pas de cheminées, sont chauffés à l'électrique et ont 3 lits ou moins. Tandis que c'est la situation inverse pour les biens immobiliers de grande taille.

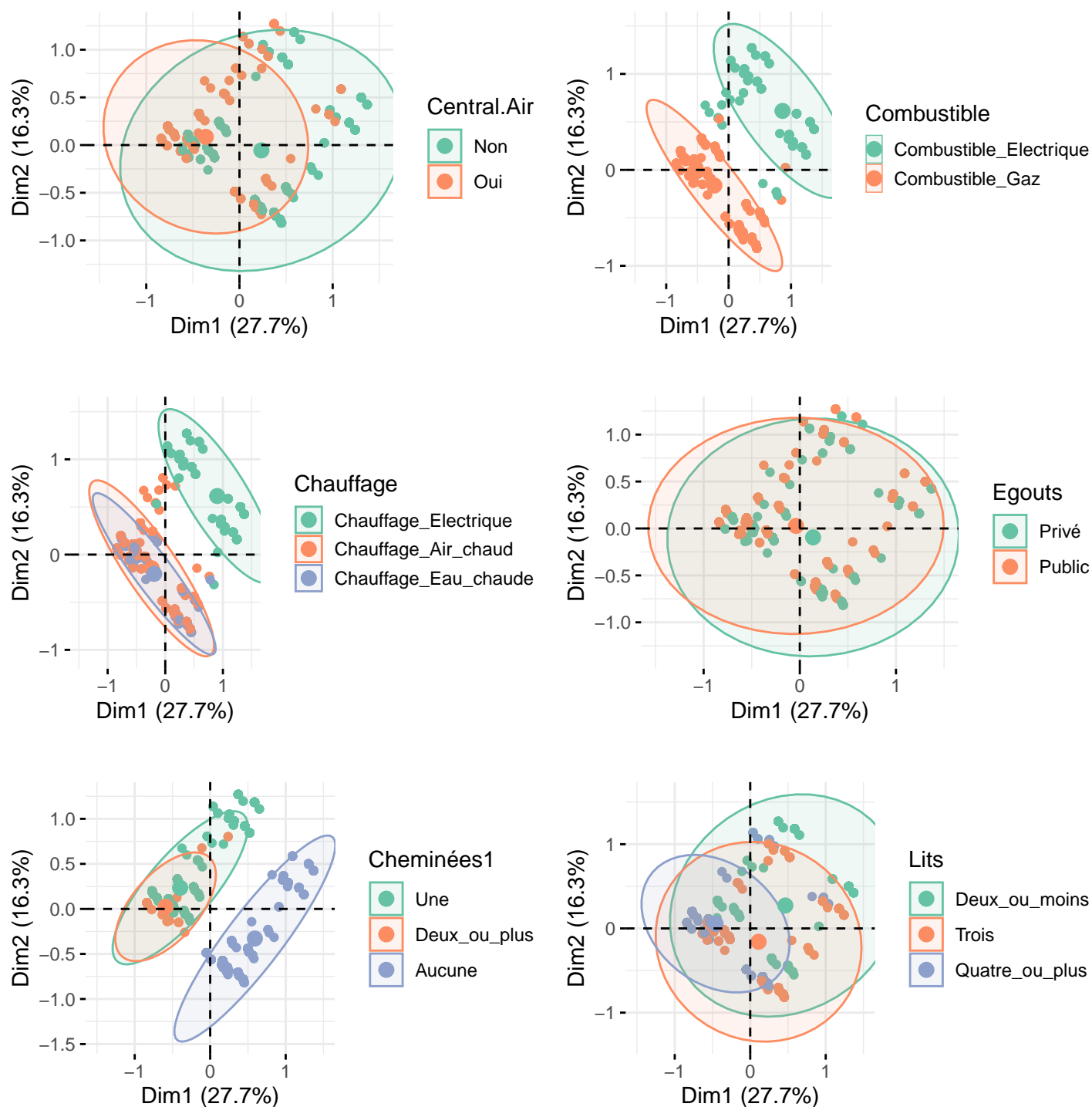
4.3 Représentation simultanée des individus et des variables



Cette représentation nous permet de mieux visualiser ce qui se passe dans nos données et ainsi conforter notre analyse précédente.

4.4 Approfondissement

Représentation des modalités de chaque variable sur le plan F(1,2)



Ces représentations d'aides à l'interprétation viennent conforter nos conclusions tirées lors de l'observation du nuage des individus et des modalités des variables. En effet, on observe que la structure du nuage de points pour chaque variable suggère encore une interprétation en fonction de la première et de la seconde bissectrice.

Ainsi, nous voyons que la première bissectrice marque effectivement une opposition entre les types de combustibles et le type de chauffage utilisé dans les biens immobiliers et que la seconde bissectrice marque une opposition entre les biens qui possèdent une cheminée et ceux n'en possèdent pas, ceux qui ont 4 lits ou plus et ceux qui en ont moins.

5 Classification des biens immobiliers

Cette méthode consiste en un découpage de la population en classes d'individus ayant les mêmes caractéristiques. Pour ce faire, nous avons utilisé les résultats de l'analyse des correspondances multiples effectuée plus tôt. Ainsi, nous avons décidé de conserver les 7 premières dimensions car elles cumulent 93.8% de l'information. L'objectif ici est de débruiter et non de résumer au mieux l'information.

5.1 Répartitions des biens en classes

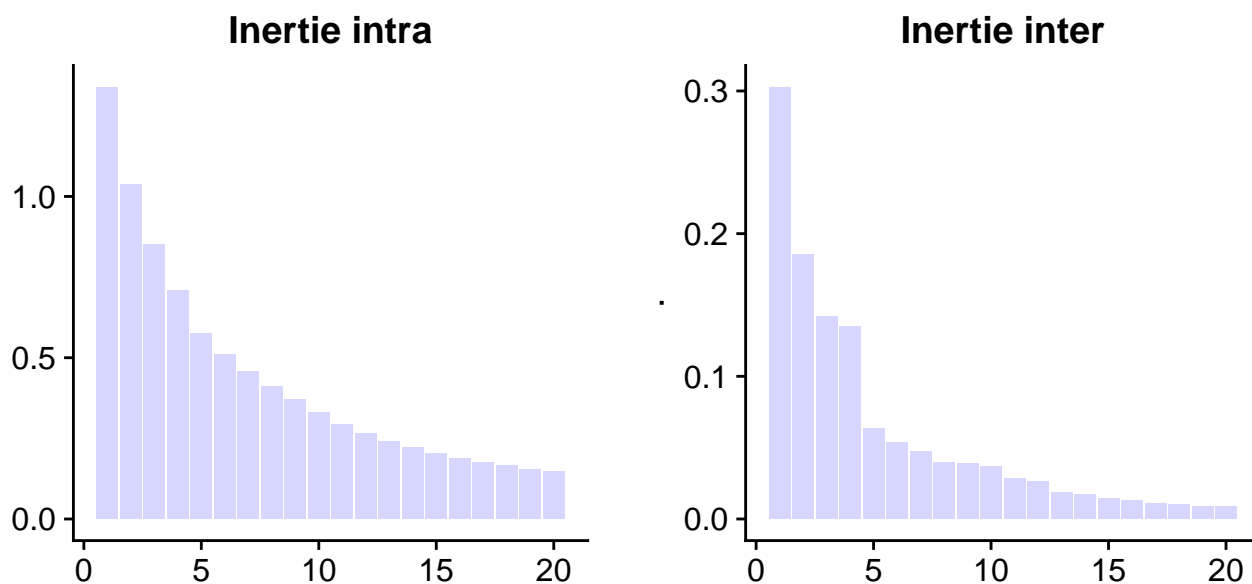
5.1.1 Arbre hiérarchique

Classification – Méthode de Ward



Au vu de la forme de l'arbre hiérarchique, il serait judicieux d'opter pour une partition en 4 classes.

5.2 Etude des inerties

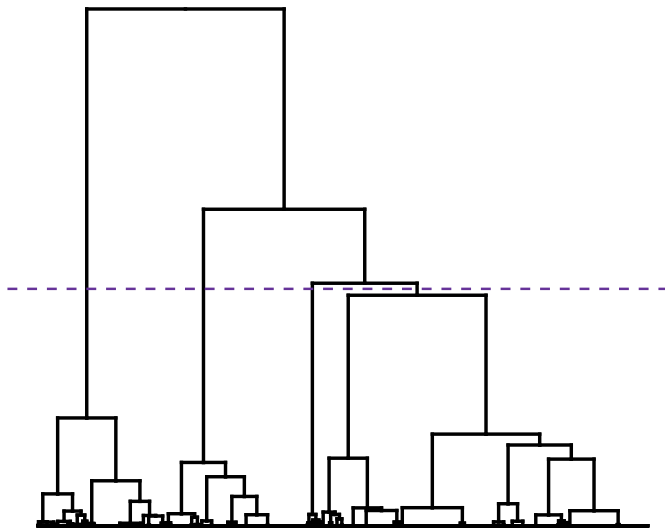


Lorsque l'on passe du 4ème au 5ème groupe, on observe une forte baisse de l'inertie inter puis une stabilisation lors du passage du 5ème au 6ème groupe. Par ailleurs, lors du passage du 4ème au 5ème groupe, on remarque que la perte d'inertie intra est faible. L'arbitrage est tel qu'un découpage en 4 groupes serait donc préférable.

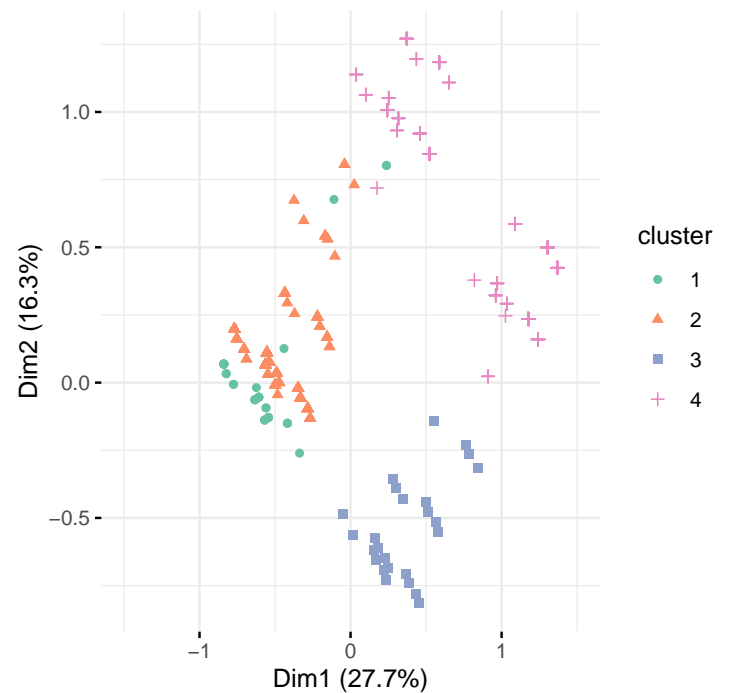
5.3 Visualisation des groupes

Classification

avec la méthode de Ward



Visualisation des groupes



5.4 Description des classes

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Cheminées1=Deux_ou_plus	100.0000000	100.0000000	2.5299601	0.0000000	18.513004
Cheminées2=Yes	4.2648709	100.0000000	59.3209055	0.0000000	6.020850
Pièces.cut=(10,12]	12.2137405	42.105263	8.7217044	0.0000000	5.597792
Prix.cut=(555.000,665.000]	41.6666667	13.157895	0.7989348	0.0000057	4.538634
Lits=Quatre_ou_plus	4.7422680	60.526316	32.2902796	0.0003426	3.580791
Prix.cut=(225.000,335.000]	5.0131926	50.0000000	25.2330226	0.0009813	3.295839
Central.Air=Oui	4.0472175	63.157895	39.4806924	0.0032122	2.946662
Combustible=Combustible_Gaz	2.9461279	92.105263	79.0945406	0.0351820	2.106257
Prix.cut=(335.000,445.000]	5.8333333	18.421053	7.9893475	0.0356498	2.100900
Egouts=Privé	4.1791045	36.842105	22.3035952	0.0400762	2.052963
Egouts=Public	2.0565553	63.157895	77.6964048	0.0400762	-2.052963
Combustible=Combustible_Electrique	0.9554140	7.894737	20.9054594	0.0351820	-2.106257
Prix.cut=[5000,115.000]	0.0000000	0.0000000	9.8535286	0.0184262	-2.356941
Chauffage=Chauffage_Electrique	0.6600660	5.263158	20.1731025	0.0117706	-2.518949
Lits=Deux_ou_moins	0.6430868	5.263158	20.7057257	0.0096024	-2.589826
Central.Air=Non	1.5401540	36.842105	60.5193076	0.0032122	-2.946662
Prix.cut=(115.000,225.000]	0.4901961	10.526316	54.3275632	0.0000000	-5.669922
Cheminées2=No	0.0000000	0.0000000	40.6790945	0.0000000	-6.020850
Cheminées1=Aucune	0.0000000	0.0000000	40.6790945	0.0000000	-6.020850
Cheminées1=Une	0.0000000	0.0000000	56.7909454	0.0000000	-7.775170

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Cheminées1=Une	86.518171	100.0000000	56.790945	0.0000000	37.391351
Cheminées2=Yes	82.828283	100.0000000	59.320906	0.0000000	35.438221
Combustible=Combustible_Gaz	61.195286	98.5094851	79.094541	0.0000000	19.924055
Lits=Quatre_ou_plus	73.195876	48.1029810	32.290280	0.0000000	13.041648
Chauffage=Chauffage_Air_chaud	61.260331	80.3523035	64.447403	0.0000000	12.817759
Central.Air=Oui	66.779089	53.6585366	39.480692	0.0000000	11.115917
Pièces.cut=(8,10]	74.103586	25.2032520	16.711052	0.0000000	8.790562
Prix.cut=(225.000,335.000]	65.963061	33.8753388	25.233023	0.0000000	7.612095
Prix.cut=(335.000,445.000]	78.333333	12.7371274	7.989348	0.0000000	6.808164
Pièces.cut=(10,12]	75.572519	13.4146341	8.721704	0.0000000	6.433766
Chauffage=Chauffage_Eau_chaude	61.904762	19.3766938	15.379494	0.0000241	4.223530
Egouts=Public	51.670951	81.7073171	77.696405	0.0002401	3.672583
Prix.cut=(445.000,555.000]	81.818182	2.4390244	1.464714	0.0018714	3.109919
Egouts=Privé	40.298507	18.2926829	22.303595	0.0002401	-3.672583
Pièces.cut=(4,6]	40.050378	21.5447154	26.431425	0.0000239	-4.225116
Lits=Trois	41.926346	40.1084011	47.003995	0.0000001	-5.264022
Prix.cut=(115.000,225.000]	41.911765	46.3414634	54.327563	0.0000000	-6.110476
Cheminées1=Deux_ou_plus	0.000000	0.0000000	2.529960	0.0000000	-6.923929
Pièces.cut=[2,4]	25.974026	8.1300813	15.379494	0.0000000	-7.776856
Prix.cut=[5000,115.000]	17.567568	3.5230352	9.853529	0.0000000	-8.371495
Lits=Deux_ou_moins	27.974276	11.7886179	20.705726	0.0000000	-8.488126
Central.Air=Non	37.623762	46.3414634	60.519308	0.0000000	-11.115917
Combustible=Combustible_Electrique	3.503185	1.4905149	20.905459	0.0000000	-19.924055
Chauffage=Chauffage_Electrique	0.660066	0.2710027	20.173102	0.0000000	-21.306670
Cheminées2=No	0.000000	0.0000000	40.679094	0.0000000	-35.438221
Cheminées1=Aucune	0.000000	0.0000000	40.679094	0.0000000	-35.438221

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Cheminées2=No	69.885434	100.0000000	40.679094	0.0000000	32.207737
Cheminées1=Aucune	69.885434	100.0000000	40.679094	0.0000000	32.207737
Combustible=Combustible_Gaz	35.606061	99.0632319	79.094541	0.0000000	14.164494
Chauffage=Chauffage_Air_chaud	35.950413	81.4988290	64.447403	0.0000000	8.987870
Lits=Trois	36.827196	60.8899297	47.003995	0.0000000	6.795117
Central.Air=Non	33.663366	71.6627635	60.519308	0.0000000	5.634673
Prix.cut=[5000,115.000]	42.567568	14.7540984	9.853529	0.0001071	3.873818
Prix.cut=(115.000,225.000]	31.985294	61.1241218	54.327563	0.0008434	3.338159
Pièces.cut=(4,6]	33.753149	31.3817330	26.431425	0.0066971	2.711543
Pièces.cut=(6,8]	32.723577	37.7049180	32.756325	0.0105881	2.556009
Pièces.cut=[2,4]	35.497836	19.2037471	15.379494	0.0109858	2.543150
Prix.cut=(445.000,555.000]	4.545454	0.2341920	1.464714	0.0065365	-2.719581
Prix.cut=(225.000,335.000]	22.427441	19.9063232	25.233023	0.0024275	-3.032239
Prix.cut=(335.000,445.000]	11.666667	3.2786885	7.989348	0.0000056	-4.541013
Cheminées1=Deux_ou_plus	0.000000	0.0000000	2.529960	0.0000025	-4.708254
Pièces.cut=(10,12]	9.923664	3.0444965	8.721704	0.0000001	-5.327369
Pièces.cut=(8,10]	14.741036	8.6651054	16.711052	0.0000000	-5.524925
Central.Air=Oui	20.404722	28.3372365	39.480692	0.0000000	-5.634673
Lits=Quatre_ou_plus	15.876289	18.0327869	32.290280	0.0000000	-7.685777
Combustible=Combustible_Electrique	1.273885	0.9367681	20.905459	0.0000000	-14.164494
Chauffage=Chauffage_Electrique	0.330033	0.2341920	20.173102	0.0000000	-14.682522
Cheminées1=Une	0.000000	0.0000000	56.790945	0.0000000	-30.844466
Cheminées2=Yes	0.000000	0.0000000	59.320906	0.0000000	-32.207737

	Cla/Mod	Mod/Cla	Global	p.value	v.test
Chauffage=Chauffage_Electrique	98.3498350	99.6655518	20.173102	0.0000000	37.717561
Combustible=Combustible_Electrique	94.2675159	98.9966555	20.905459	0.0000000	36.225450
Lits=Deux_ou_moins	42.4437299	44.1471572	20.705726	0.0000000	10.457220
Central.Air=Non	27.1727173	82.6086957	60.519308	0.0000000	9.104808
Cheminées2=No	30.1145663	61.5384615	40.679094	0.0000000	8.125618
Cheminées1=Aucune	30.1145663	61.5384615	40.679094	0.0000000	8.125618
Pièces.cut=[2,4]	37.6623377	29.0969900	15.379494	0.0000000	6.875471
Prix.cut=(115.000,225.000]	25.6127451	69.8996656	54.327563	0.0000000	6.110657
Prix.cut=[5000,115.000]	39.8648649	19.7324415	9.853529	0.0000000	5.911359
Pièces.cut=(4,6]	24.6851385	32.7759197	26.431425	0.0062822	2.732676
Cheminées1=Deux_ou_plus	0.0000000	0.0000000	2.529960	0.0001928	-3.728244
Prix.cut=(335.000,445.000]	4.1666667	1.6722408	7.989348	0.0000003	-5.120312
Pièces.cut=(8,10]	8.7649402	7.3578595	16.711052	0.0000002	-5.191074
Pièces.cut=(10,12]	2.2900763	1.0033445	8.721704	0.0000000	-6.225783
Cheminées1=Une	13.4818288	38.4615385	56.790945	0.0000000	-7.111104
Cheminées2=Yes	12.9068462	38.4615385	59.320906	0.0000000	-8.125618
Prix.cut=(225.000,335.000]	6.5963061	8.3612040	25.233023	0.0000000	-8.149735
Central.Air=Oui	8.7689713	17.3913043	39.480692	0.0000000	-9.104808
Lits=Quatre_ou_plus	6.1855670	10.0334448	32.290280	0.0000000	-9.917877
Chauffage=Chauffage_Eau_chaude	0.0000000	0.0000000	15.379494	0.0000000	-10.361666
Chauffage=Chauffage_Air_chaud	0.1033058	0.3344482	64.447403	0.0000000	-27.291646
Combustible=Combustible_Gaz	0.2525253	1.0033445	79.094541	0.0000000	-36.225450

Nous voyons que 100 % des biens immobiliers ayant deux cheminées ou plus sont dans la classe 1. De même, 60 % des biens qui sont dans la classe 1 possèdent 4 lits ou plus. De plus, 86.5 % des biens immobiliers ayant 1 cheminées sont dans la classe 2. Par ailleurs, 70 % des biens immobiliers n'ayant pas de cheminées sont dans la classe 3. 100 % des biens de qui n'ont pas de cheminées sont dans la classe 3. Pour finir, 98 % des biens immobiliers chauffés à l'électrique sont dans la classe 4.

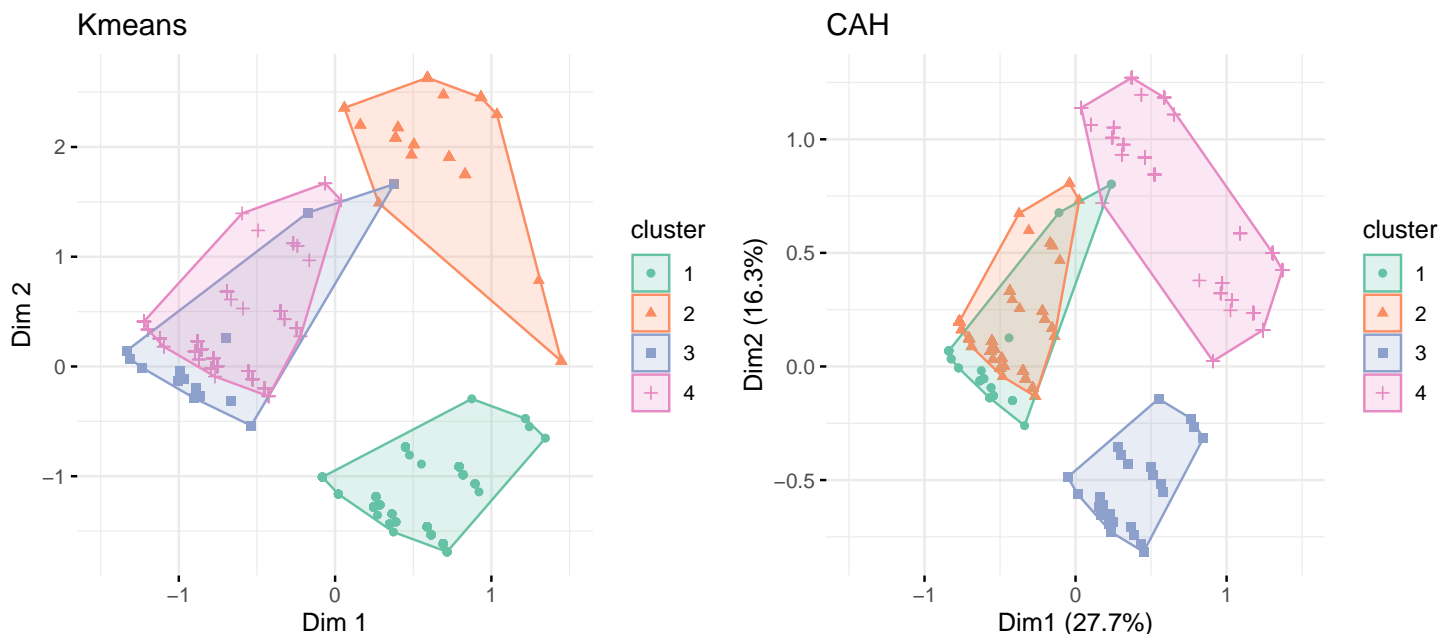
Ces 4 classes peuvent représenter 4 types différents de biens allant du plus grand (classe 1) au plus petit (classe 4).

5.5 Méthode de partitionnement

Tout comme la classification hiérarchique ascendante, l'algorithme de partitionnement appelé "k-means" permet de découper la population en classes.

5.6 Comparaison des groupes réalisés avec k-means et classification hiérarchique

Avec les 2 méthodes de répartition de la population en classe, nous observons que les partitionnements sont assez équivalents.



6 Annexes

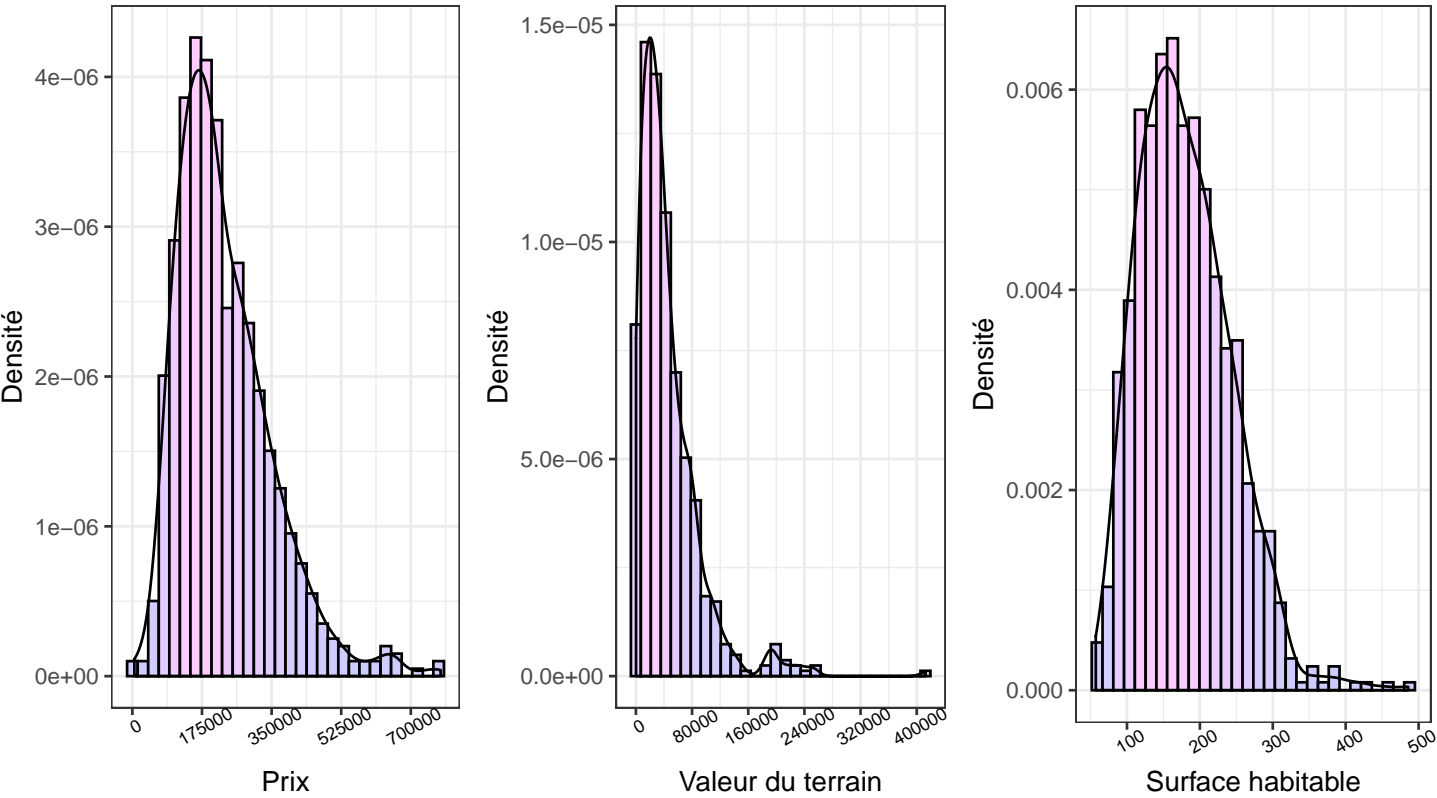
6.1 Analyse descriptive des variables

6.1.1 Résumé des variables quantitatives

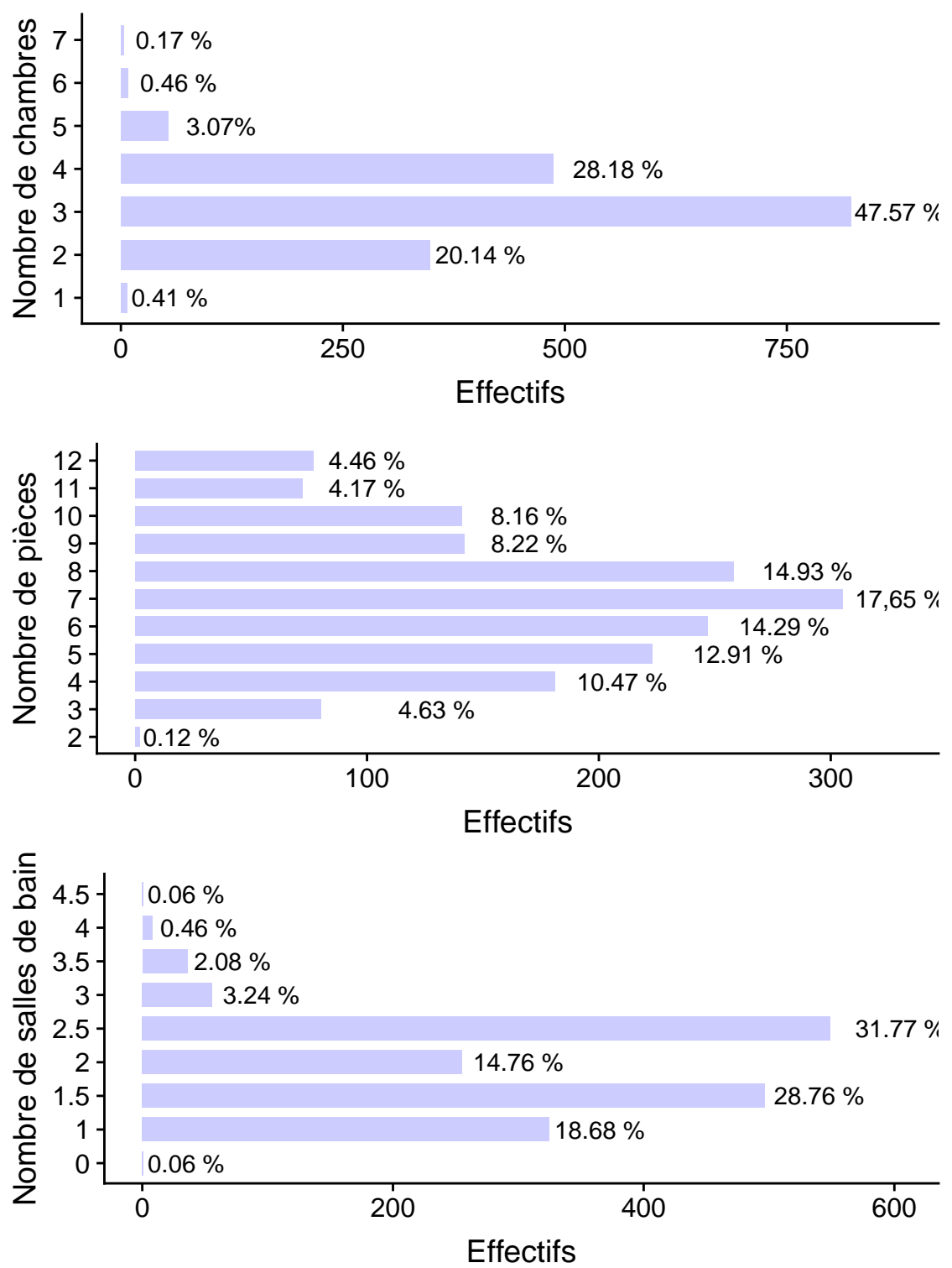
	Prix	Taille	Age	Valeur_du_terrain	Surface_habitable	Pct.College	Chambres	Salles_de_bains	Pièces
Moyenne	211966.71	0.50	27.92	34557.19	163.040	55.57	3.15	1.90	7.04
Écart type	98441.39	0.70	29.21	35021.17	57.590	10.33	0.82	0.66	2.32
Minimum	5000.00	0.00	0.00	200.00	57.200	20.00	1.00	0.00	2.00
Q1	145000.00	0.17	13.00	15100.00	120.800	52.00	3.00	1.50	5.00
Médiane	189900.00	0.37	19.00	25000.00	151.850	57.00	3.00	2.00	7.00
Q3	259000.00	0.54	34.00	40200.00	198.575	64.00	4.00	2.50	8.25
Maximum	775000.00	12.20	225.00	412600.00	485.700	82.00	7.00	4.50	12.00

6.1.1.1 Variables quantitatives continues

Dans cette base de données, nous avons deux variables quantitatives continues qui sont : “Prix” et “Valeur du terrain”.

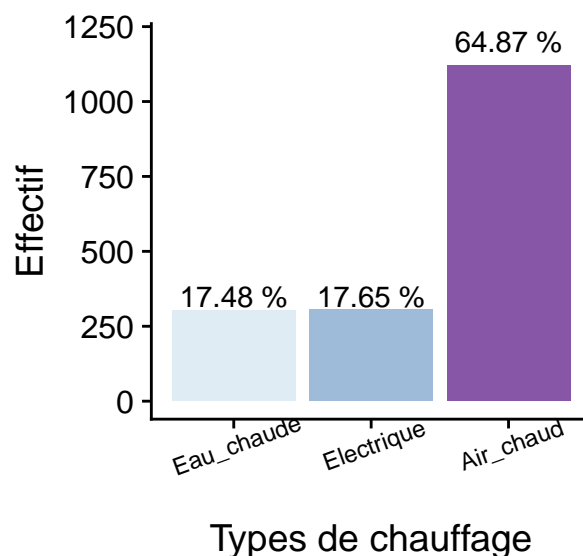
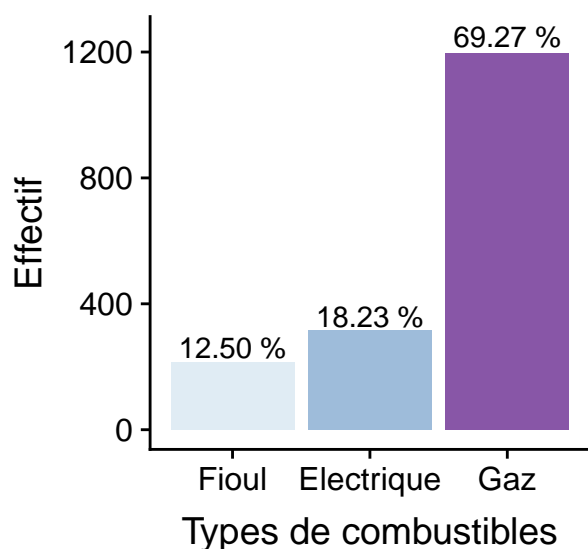
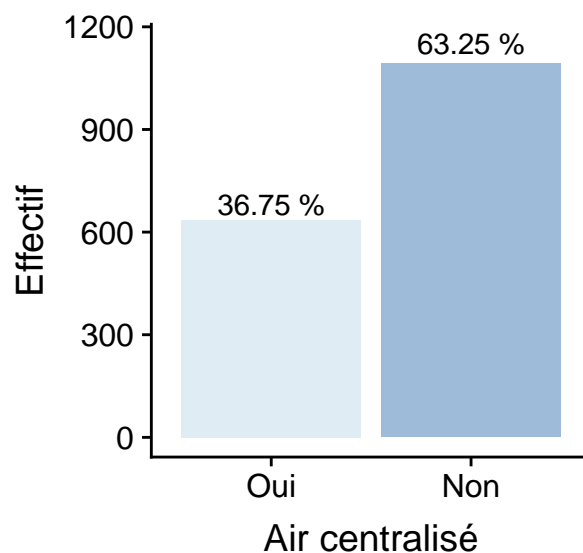
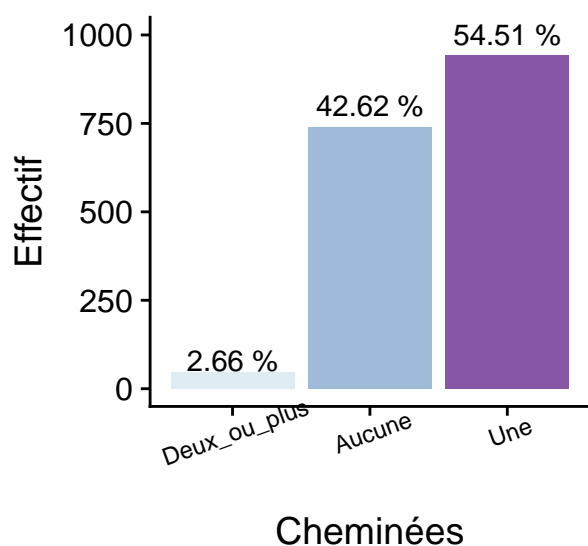
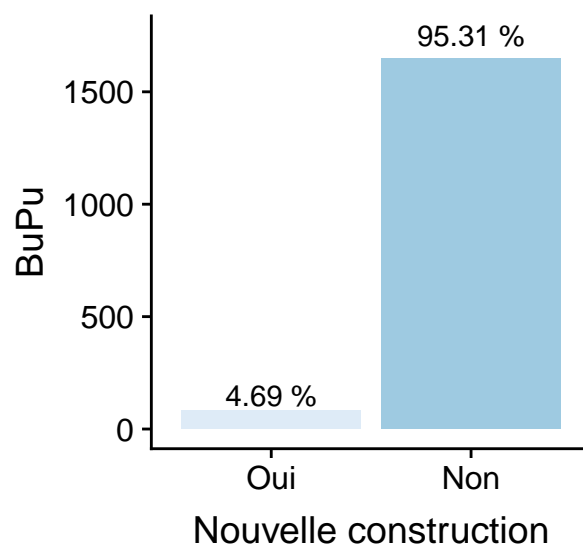
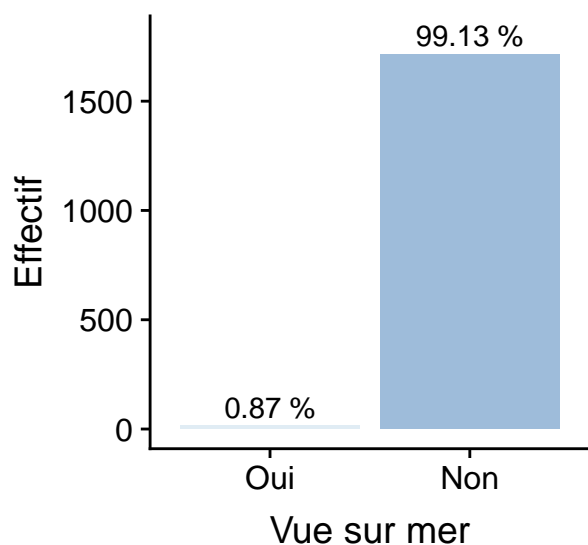


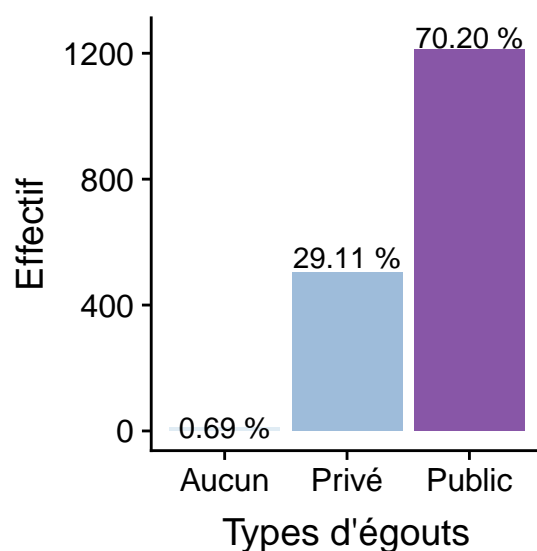
6.1.1.2 Variables quantitatives discrètes



6.1.2 Variables qualitatives

Dès maintenant nous allons présenter une étude illustrative des caractéristiques des variables qualitatives de notre jeu de données.





Au vu de ces graphiques, nous pouvons dire que les biens qui sont de nouvelles constructions et ceux qui ont une vue sur mer sont très minoritaire dans notre base de données. Le type de combustible le plus prédominant est le gaz suivi de loin par l'électrique et le pétrole. Le type de chauffage le plus prédominant est l'air chaud suivi de loin par l'électrique et l'eau chaude. En ce qui concerne l'air centralisé, 63% des biens en sont équipés contre 36% non. Pour finir, 54% des biens possèdent une ou deux cheminées contre 42% qui n'en ont pas. Pour le type d'égouts, pour 70% des biens, ils sont publics.

6.2 Mise en relation des variables - Statistiques inférentielles

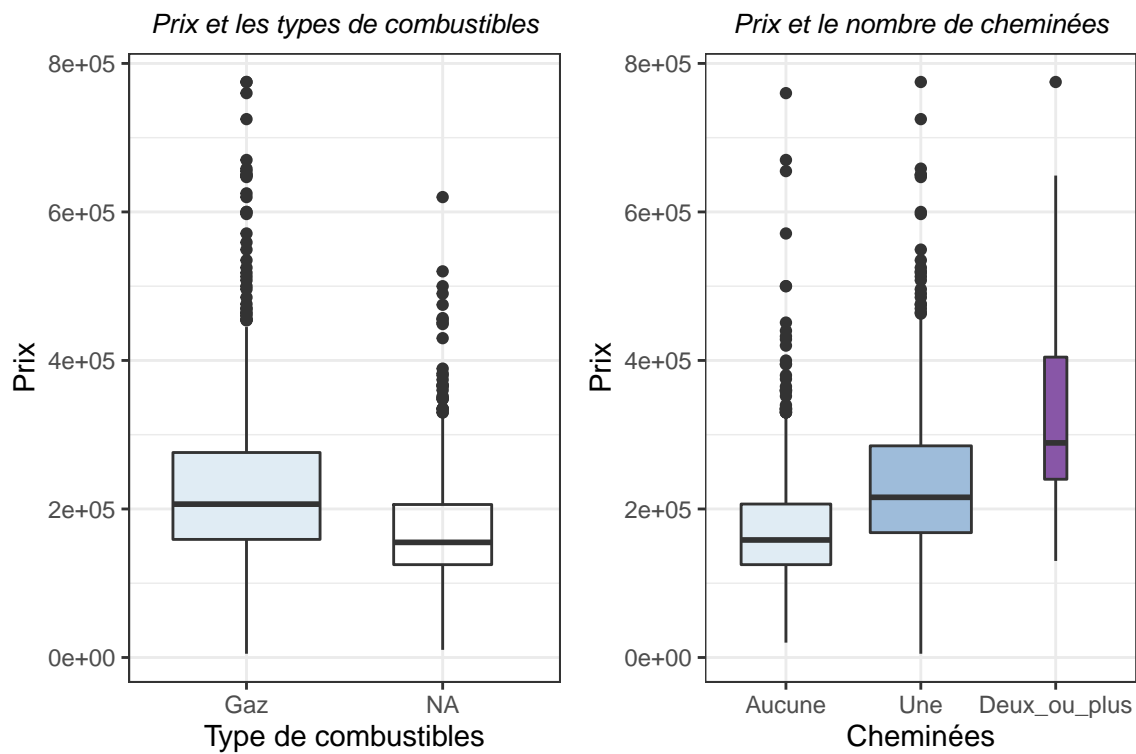
Tout d'abord, nous avons décidé de mettre en relation les variables qui nous semblaient les plus pertinentes avec le Prix. Nous allons tenter de trouver des variables qui aurait un lien et éventuellement un impact sur le prix.

6.2.1 Test d'indépendance avec la variable "Prix"

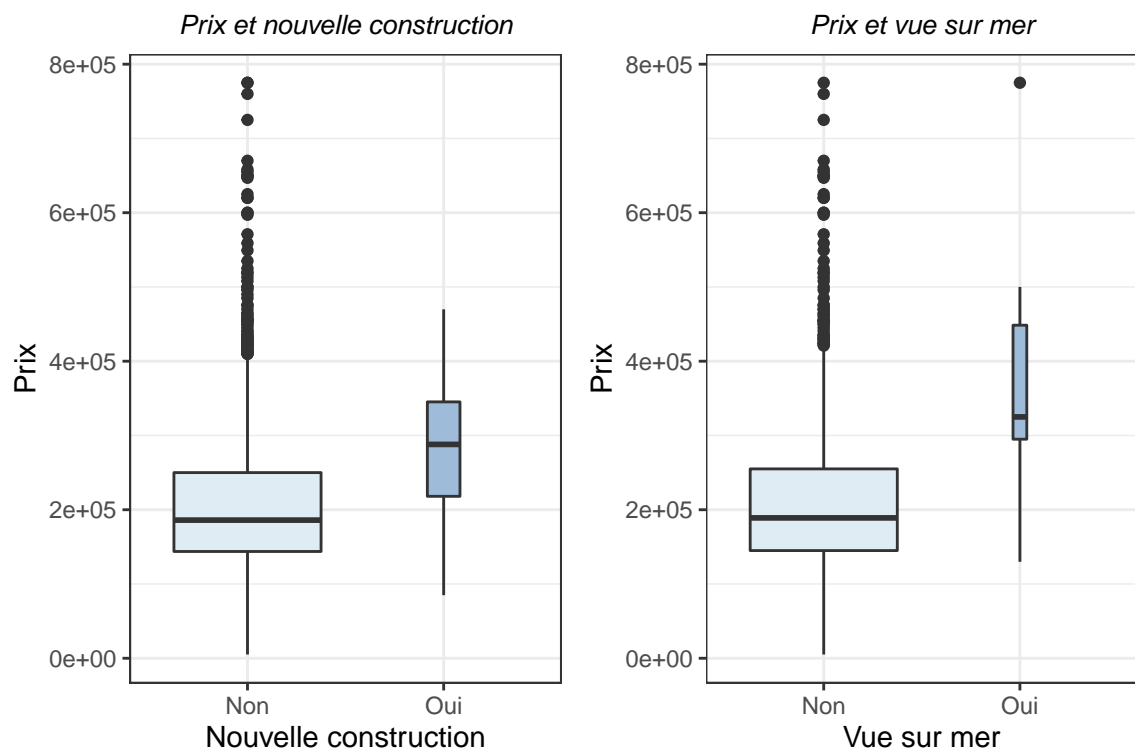
	Vue_sur_mer	Nouvelle_construction	Central.Air	Combustible	Chauffage	Egouts
Eff. théo. min.	0.0086806	0.046875	0.3674769	0.1250000	0.1747685	0.0069444
p-value	0.0182894	0.000000	0.0005208	0.3585736	0.3254314	0.0000000

Au vu du test anova réalisé entre les prix et les différents types de combustible, au risque de 5%, nous pouvons dire qu'il existe bien une différence de prix entre les biens n'utilisant pas les mêmes combustibles. Le test anova réalisé entre le prix et le nombre de cheminées au risque de 5% nous permet de rejeter l'hypothèse nulle telle que les prix des biens immobiliers sont égaux selon le nombre de cheminées

6.2.2 Illustrations



Globalement, on remarque que les biens immobiliers dont le type de combustible est le gaz ont des prix plus élevés que les autres. On remarque que globalement les prix des maisons qui ont deux cheminées ou plus sont plus chères que les autres. Puis viennent celles avec une seule cheminée et pour finir celles avec aucunes.

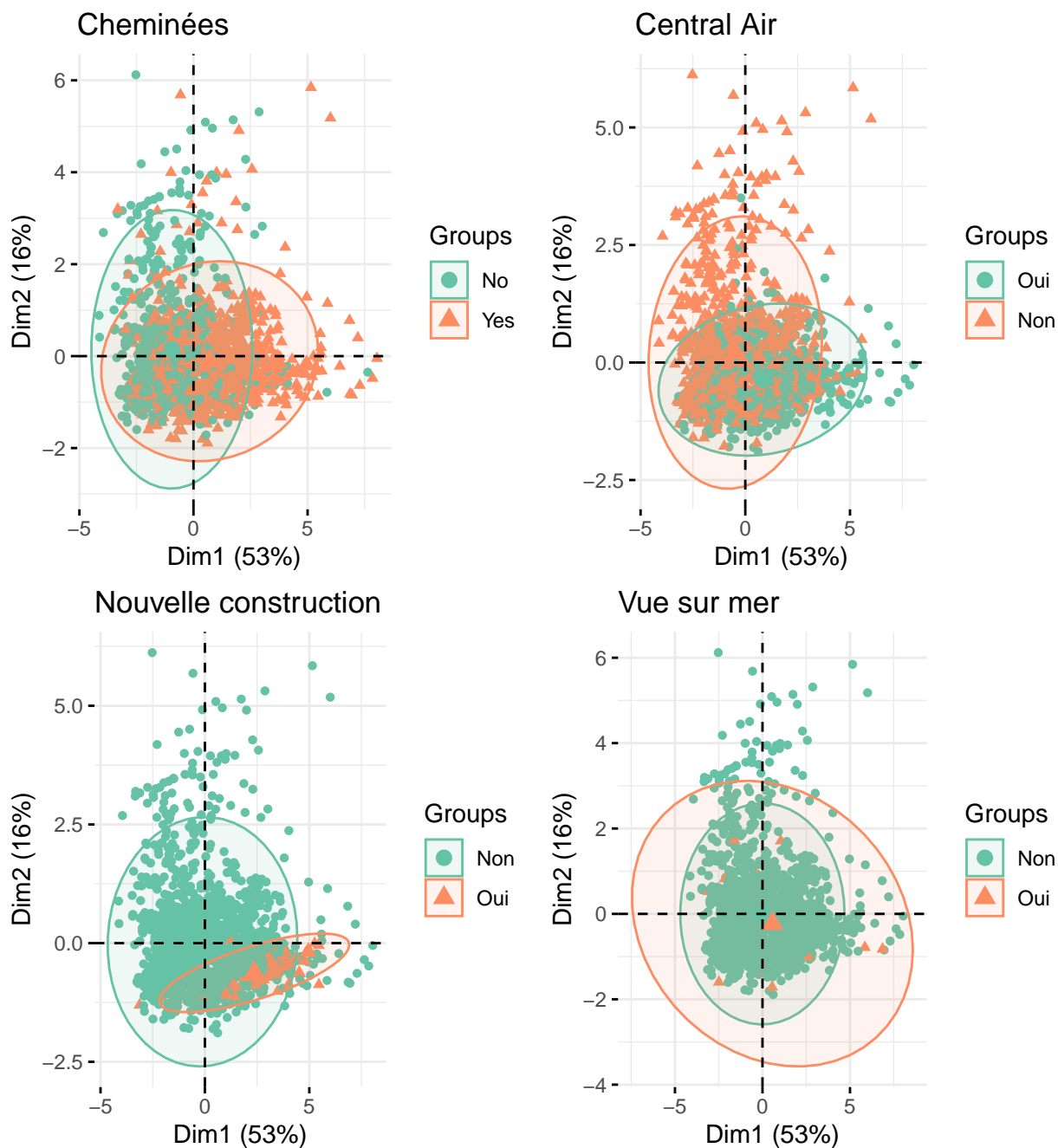


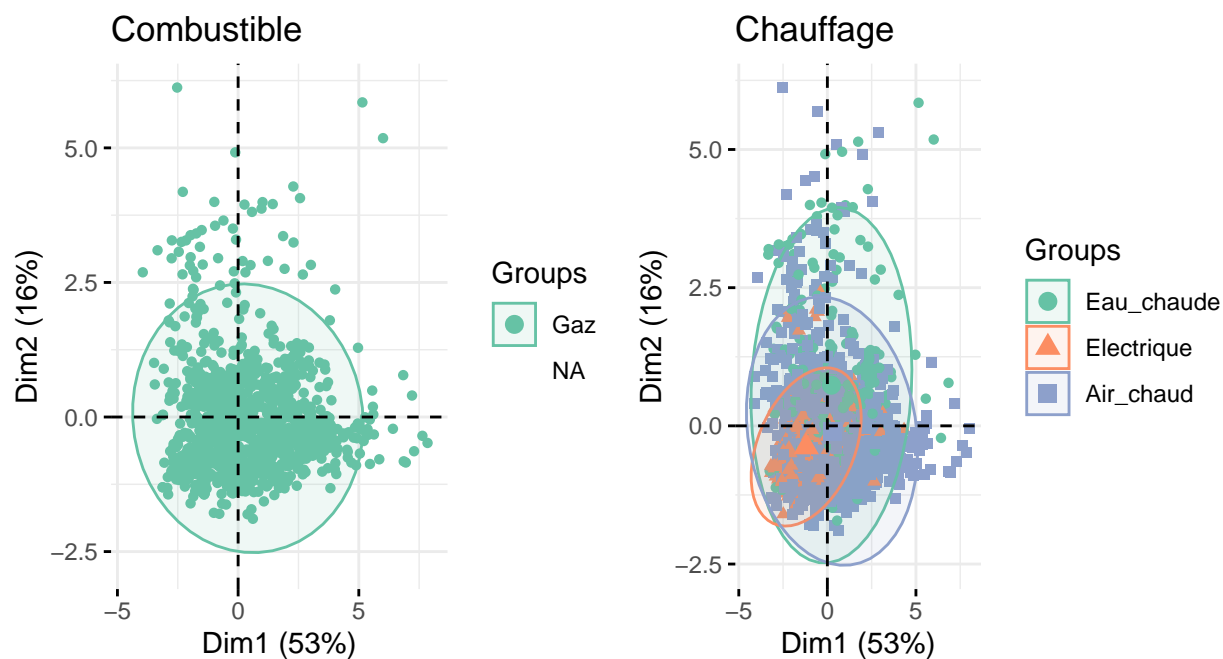
Grâce à cette représentation de la variable Nouvelle Construction, on peut constater effectivement que les nouvelles constructions valent plus chères que les anciennes.

Nous voyons clairement que les maisons qui ont vue sur mer ont des prix plus élevés que celles qui n'en ont pas.

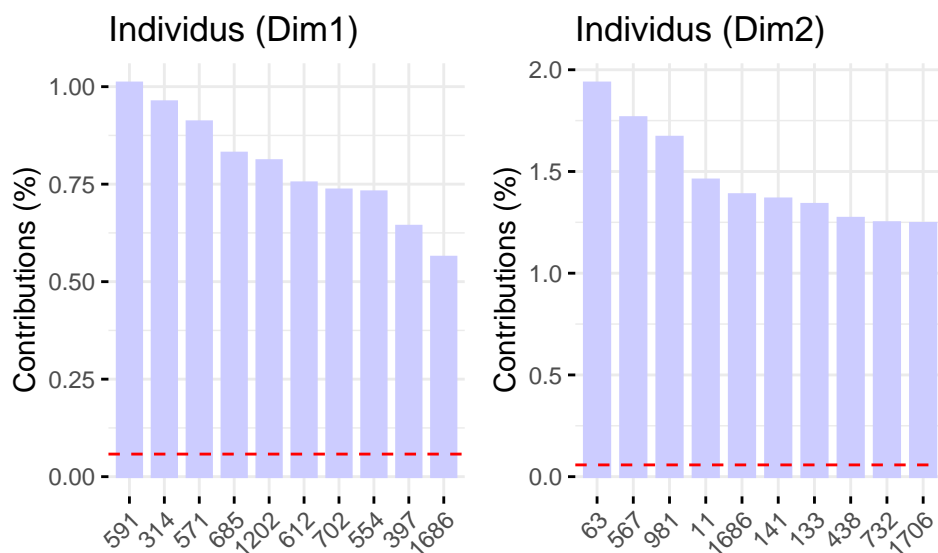
6.3 Analyse en composantes principales

Nous avons réalisé le nuage des individus en distinguant les différents groupes correspondant aux modalités de chaque variables et cela pour toutes les variables qualitatives. Les nuages sont représentés ci-dessous.





On remarque toutes ces variable ne sont pas pertinentes car elles ne permettent pas de séparer le nuage des individus. Seule la variable "Lits" permettait une séparation.



6.4 Analyse factorielle des correspondances

TABLE 8 – Tableau des profils ligne

	Type de chauffage		
	Electrique	Air_chaud	Eau_chaude
10	7.1	68.1	24.8
11	2.8	75.0	22.2
12	2.6	74.0	23.4
2	0.0	100.0	0.0
3	41.2	45.0	13.8
4	29.8	57.5	12.7
5	23.3	59.6	17.0
6	19.0	67.2	13.8
7	18.7	64.6	16.7
8	13.6	68.2	18.2
9	9.2	70.4	20.4
Profil moyen	167.3	749.6	183.0

TABLE 9 – Caractéristiques des profils ligne

	Poids	CoordF1	CoordF2	ContribF1	ContribF2	QualiteF1	QualitéF2	QualF1F2
Electrique	0.177	-0.516	0.015	81.198	1.151	0.999	0.001	1
Air_chaud	0.649	0.094	-0.036	9.973	25.154	0.875	0.125	1
Eau_chaude	0.175	0.171	0.118	8.828	73.695	0.679	0.321	1

TABLE 10 – Tableau des profils colonne

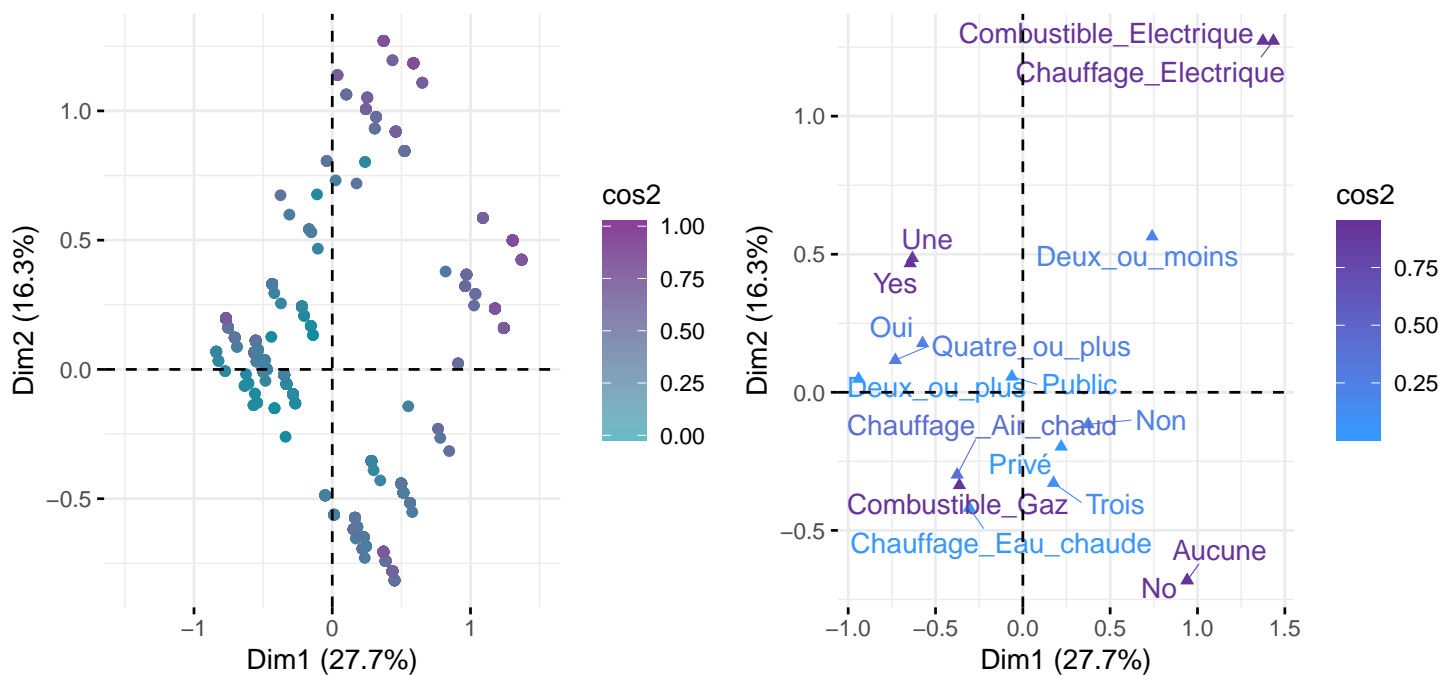
	Nombre de pièces										
	10	11	12	2	3	4	5	6	7	8	9
Electrique	3.3	0.7	0.7	0.0	10.8	17.7	17.0	15.4	18.7	11.5	4.3
Air_chaud	8.6	4.8	5.1	0.2	3.2	9.3	11.9	14.8	17.6	15.7	8.9
Eau_chaude	11.6	5.3	6.0	0.0	3.6	7.6	12.6	11.3	16.9	15.6	9.6
Profil moyen	23.5	10.8	11.8	0.2	17.6	34.6	41.5	41.5	53.2	42.8	22.8

TABLE 11 – Caractéristiques des profils colonne

	Poids	CoordF1	CoordF2	ContribF1	ContribF2	QualiteF1	QualitéF2	QualF1F2
10	0.082	0.291	0.104	11.943	26.881	0.887	0.113	1
11	0.042	0.392	-0.004	11.071	0.016	1.000	0.000	1
12	0.045	0.401	0.026	12.343	0.899	0.996	0.004	1
2	0.001	0.392	-0.623	0.307	13.701	0.284	0.716	1
3	0.046	-0.611	0.107	29.779	16.330	0.970	0.030	1
4	0.105	-0.324	-0.021	19.006	1.365	0.996	0.004	1
5	0.129	-0.145	0.038	4.693	5.716	0.936	0.064	1
6	0.143	-0.047	-0.087	0.540	33.207	0.223	0.777	1
7	0.177	-0.029	-0.011	0.252	0.664	0.870	0.130	1
8	0.149	0.106	-0.016	2.893	1.175	0.978	0.022	1
9	0.082	0.225	0.004	7.172	0.045	1.000	0.000	1

6.5 Analyse des correspondances multiples

6.5.1 Qualité de représentation



6.5.2 Contributions

