

Cars

Jean Clark, Lilian Boissé et Théo De Sa Morais

Contents

1 I- Présentation du document	1
1.1 Les données	1
2 L'analyse	7
2.1 L'analyse par composante principale	7
2.2 L'analyse factorielle des composantes	11

1 I- Présentation du document

1.1 Les données

Ce document présente les données sur la vente de voitures aux États-Unis, entre ventes aux enchères et ventes au détaillant. Chaque ligne correspond à un véhicule vendu aux enchères entre 2009 et 2010.

Structure de notre document :

Notre tableau de données contient 28 variables détaillées très brièvement :

- **Mauvais_Achat** : facteur indiquant si le véhicule était un achat évitable ou non.
- **Date_d_achat** : date d'achat du véhicule.
- **Encheres** : dans quelle enchère le véhicule a été vendu.
- **Annee_du_vehicule** : année où le véhicule a été vendu pour la première fois.
- **Age_du_vehicule** : âge du véhicule.
- **Fabrication** : marque de fabrication du véhicule.
- **Modele** : modèle du véhicule.
- **Trim** : aménagements et/ou options du véhicule.
- **Sous_Marque** :
- **Couleur** : couleur du véhicule.
- **Type_de_transmission** : type de transmission du véhicule, deux facteurs : automatique ou manuelle.
- **Type_de_roues** : le type de roues sur le véhicule.
- **Kilometrage** : kilométrage du véhicule au moment de la dernière vente.
- **Nationalite** : nationalité du constructeur du véhicule.
- **Taille** : taille du véhicule.
- **Top_trois_cons_Americains** : facteur déterminant si le véhicule a été construit par l'un des géants américains suivant : Ford, Chrysler ou GM(General Motors).
- **CVh_M_E_1** : cote du véhicule aux enchères au moment de l'achat si il est en moyen état.
- **CVh_B_E_1** : cote du véhicule aux enchères au moment de l'achat si il est en bon état.
- **CVh_M_D_1** : cote du véhicule au détaillant au moment de l'achat si il est en moyen état.

- **CVh_B_D_1** : cote du véhicule au détaillant au moment de l'achat si il est en bon état.
- **CVh_M_E_2** : cote du véhicule actuelle aux enchères si il est en moyen état.
- **CVh_B_E_2** : cote du véhicule actuelle aux enchères si il est en bon état.
- **CVh_M_D_2** : cote du véhicule actuelle au détaillant si il est en moyen état.
- **CVh_B_D_2** : cote du véhicule actuelle au détaillant si il est en bon état.
- **Numero_de_l_acheteur**
- **Etat** : État d'origine de l'acheteur.
- **Ct_vhcl_au_momt_de_l_achat** : coût du véhicule au moment de l'achat
- **Est_une_vente_en_ligne** : facteur oui ou non.
- **Cout_de_la_garantie**

Nous avons décidé, par souci de redondance de l'information, de supprimer trois variables : l'ID du type de roues qui donnait la même information que le type de roues, l'ID du véhicule et enfin le code postal de l'acheteur qui est redondant avec l'État d'origine de l'acheteur.

Les variables PRIMEUNIT et AUCGART ont été supprimées car trop d'informations sur ces variables sont manquantes. Les premières observations en regardant la structure de nos données seront sur l'âge des véhicules vendus qui ont tous moins de 9 ans ou 9 ans ce qui en font des véhicules relativement récents. Ce marché représenté par nos données est donc un marché de l'occasion mais récente, pas de véhicule de collection, pas de véhicule en mauvais état et/ou obsolète. Nous garderons cette observation en tête pour le restant de l'analyse.

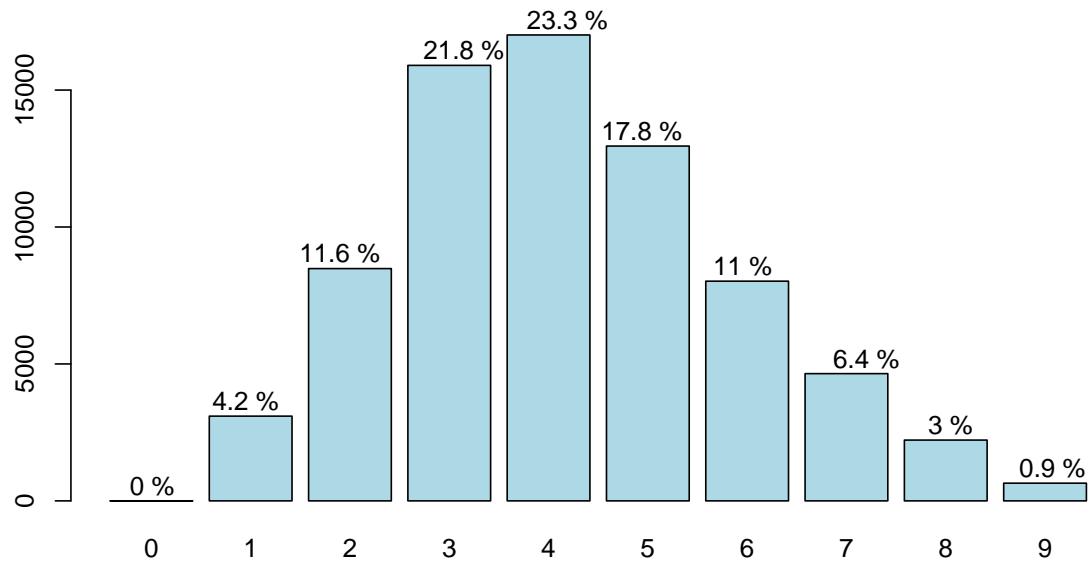
1.1.1 Qualitatives

Commençons par analyser nos données qualitatives.

Les véhicules les plus vendus sont de fabrication **Chevrolet, Dodge, Ford et Chrysler**. Ces quatres marques totalisent à elles seules plus de **68% des ventes aux enchères aux États-Unis**. Quand aux modèles, ce sont encore une fois les modèles de ces marques dominantes qui sont les plus vendus avec tout de même une petite surprise, la **PT Cruiser** qui est construite par Chrysler est la voiture la plus vendue alors que Chrysler est seulement la quatrième marque la plus vendue aux États-Unis (voir Annexes Tables 1 et 2).

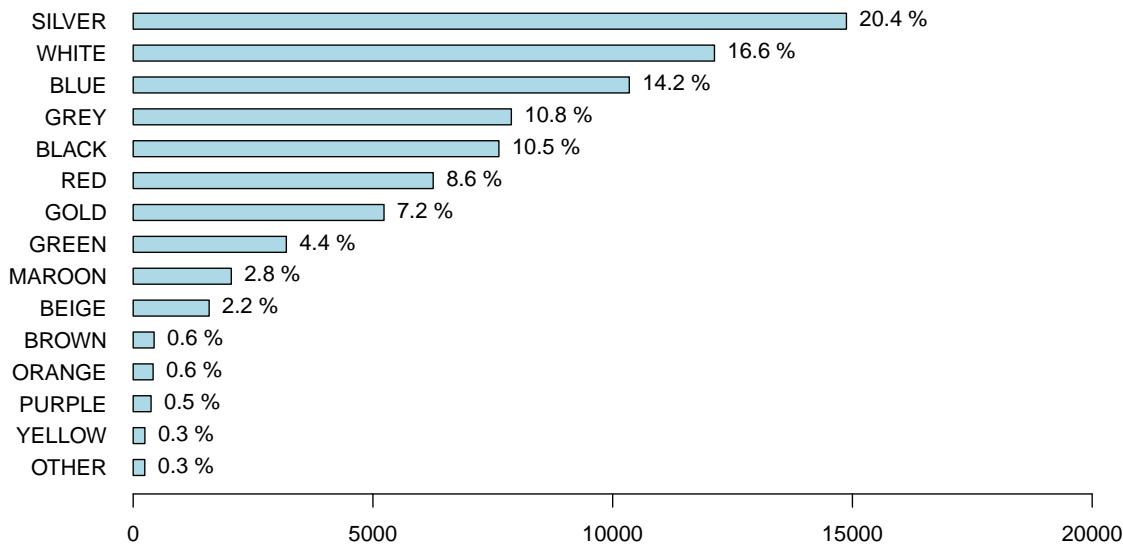
On observe logiquement que **83,6% des véhicules sont d'origine Américaine**, mais il existe aussi un marché des voitures asiatiques. Les voitures européennes sont très rares sur le marché de la vente aux enchères aux États-Unis (voir Annexes Table 3).

Répartition des véhicules vendus selon leur âge



On voit ici que la plupart des véhicules ont été mis en circulation autour de l'année 2006 (voir Annexes Table 4).

Effectifs par couleur de la voiture



Les couleurs froides comme le gris, le blanc et le bleu semblent se vendre le plus, tandis que les véhicules de couleurs chaude et vive sont les moins représentés. On note aussi que les véhicules de moyenne et grande taille utilisant une boîte automatique sont les plus présents sur le marché des enchères de voitures (voir Annexes Tables 5 et 6).

Ce ne sont pas des particuliers qui achètent les voitures aux enchères, **les concessionnaires et revendeurs professionnels** sont les premiers consommateurs de véhicules aux enchères. On observe que les états où il y a le plus d'enchères sont les États dont la densité urbaine est moins importante comme **le Texas, l'Arizona ou la Floride** (voir Annexes Tables 6 et 7).

Les entreprises *MANHEIM* et *ADESA* dominent le marché de la vente de véhicules aux enchères aux États-Unis.

1.1.2 I.1.B- Quantitatives

La variable la plus corrélée avec le prix d'achat du véhicule est la cote d'un véhicule en moyen état vendu aux enchères au moment de l'achat. À l'inverse Le prix de vente du véhicule ne semble pas corrélé avec le kilométrage et le coup de la garantie.

	CVh_M_E_1	CVh_B_E_1	CVh_M_D_1	CVh_B_D_1
Moyenne	6199.44	7459.79	8599.31	9970.23
Ecart-type	2386.62	2621.30	3039.01	3235.37
Minimum	1475.00	2122.18	2187.18	2908.00
Q1	4339.00	5480.00	6373.00	7571.00
Médiane	6139.00	7355.00	8484.00	9848.00
Q3	7782.00	9044.00	10682.00	12123.00
Maximum	35722.00	36859.00	39080.00	41482.00

En moyenne la côte d'un véhicule pendant la période d'achat est plus haute chez un détaillant qu'aux enchères. Il semble donc plus intéressant pour un acheteur de s'orienter vers une enchère pour acquérir une nouvelle voiture.

	CVh_M_E_2	CVh_B_E_2	CVh_M_D_2	CVh_B_D_2
Moyenne	6174.24	7441.95	8832.75	10212.49
Ecart-type	2389.28	2624.50	3014.56	3212.84
Minimum	1406.00	1973.00	2134.80	2812.33
Q1	4311.00	5464.00	6585.00	7833.00
Médiane	6085.00	7339.00	8756.00	10122.00
Q3	7746.00	9024.00	10921.00	12322.00
Maximum	35722.00	36859.00	39080.00	41062.00

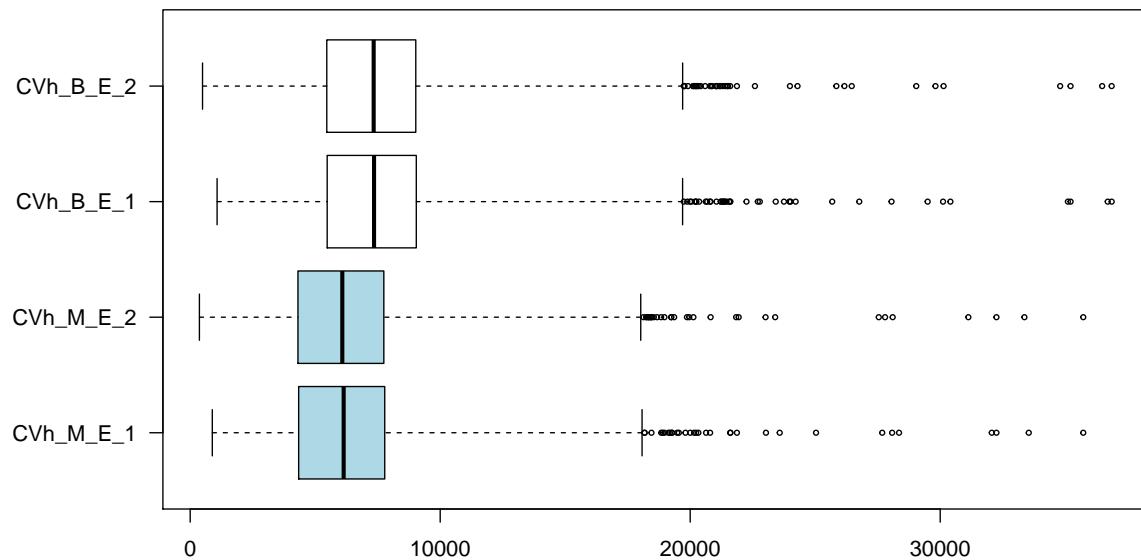
Il semblerait d'après les moyennes que les côtes des détaillants, pour une même voiture, gagnent plus de valeur avec le temps que celles des enchères.

	Kilometrage	Ct_vhcl_au_momt_de_l_achat	Cout_de_la_garantie
Moyenne	71524.78	6729.58	1276.9
Ecart-type	14565.41	1770.28	598.1
Minimum	22917.33	2620.00	462.0
Q1	61869.00	5430.00	837.0
Médiane	73376.00	6700.00	1169.0
Q3	82441.00	7900.00	1623.0
Maximum	115717.00	45469.00	7498.0

En moyenne le prix d'achat d'une voiture est plus grand que sa côte aux enchères. Il faut vérifier si cette différence est significative.

Testons la côte du véhicule aux enchères au moment de l'achat contre la côte du véhicule aux enchères actuellement pour le même état de la voiture.

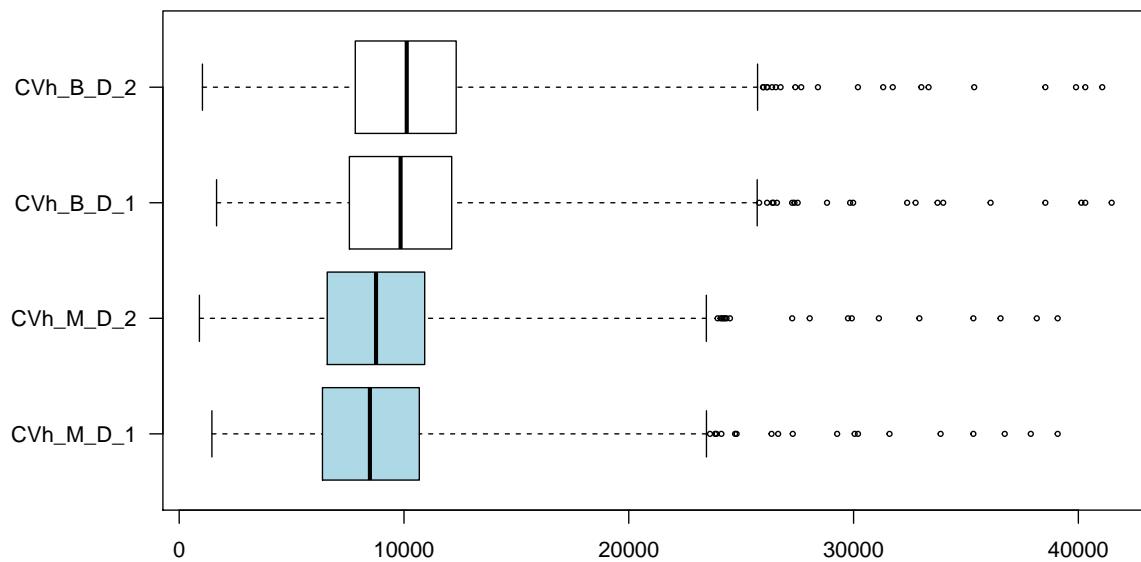
Cote aux enchères selon la période



commentaires à refaire

Testons la cote du véhicule chez un détaillant au moment de l'achat contre la cote du véhicule chez un détaillant actuellement pour le même état de la voiture.

Cote au détaillant selon la période



Test de comparaison Kilometrage et Cout de la garantie d'un Mauvais/Bon achat

	Kilometrage	Cout_de_la_garantie
mean in group Non	71073.22	1265.225
mean in group Oui	74746.93	1360.231
p-value	0.00	0.000

La p-value de chacun de ces tests de comparaison des moyennes est proche de 0, ce qui indiquent que l'on rejette l'hypothèse d'égalité des moyennes. En moyenne, les voitures identifiées comme bon achat ont un kilometrage et un cout de garantie plus faible que les voitures identifiées comme mauvais achat.

Test de comparaison Age et Cout du véhicule d'un Mauvais/Bon achat

	Age_du_vehicule	Ct_vhcl_au_momt_de_l_achat
mean in group Non	5.073112	6795.637
mean in group Oui	5.946062	6258.213
p-value	1.000000	0.000

La p-value du test comparant la moyenne des ages des véhicules dans le deux groupes est égale à 1. L'age du véhicule du groupe bon achat est donc en moyenne plus petit que celui du groupe mauvais achat. La p-value du test comparant le cout du véhicules est égale à 0. En moyenne, le cout d'achat d'un véhicule identifié comme bon achat est plus élevée que celu d'un véhicule identifié comme mauvais achat.

1.1.3 I.1.C- Axes d'analyses

Du point de vue d'un acheteur :

- Repérer les véhicules à éviter d'acheter
- Repérer les types de véhicule selon les différentes enchères.
- Au contraire repérer les véhicules qui prennent le plus de valeur par rapport à leur cote aux enchères.

Du point de vue des enchères :

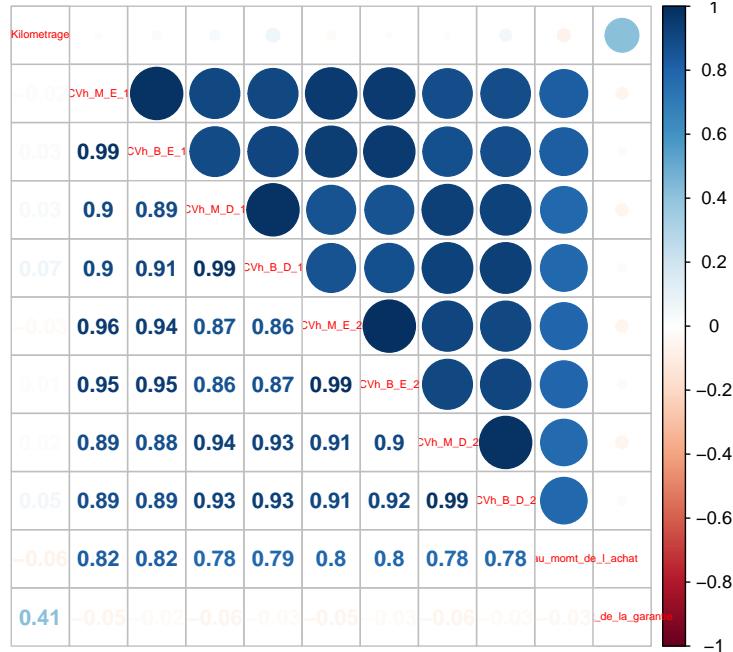
- Identifier les acheteurs
- Contrôler le niveau des cotes
- Repérer les différents groupes de véhicules

Du point de vue d'un vendeur :

- Étudier la différence entre les cotes du véhicule entre les détaillants et les enchères selon les périodes et les différentes caractéristiques du véhicule.
- Étudier la différence entre la valeur du véhicule et les cotes selon les caractéristiques du véhicule(marque, modèle, type de roues...).

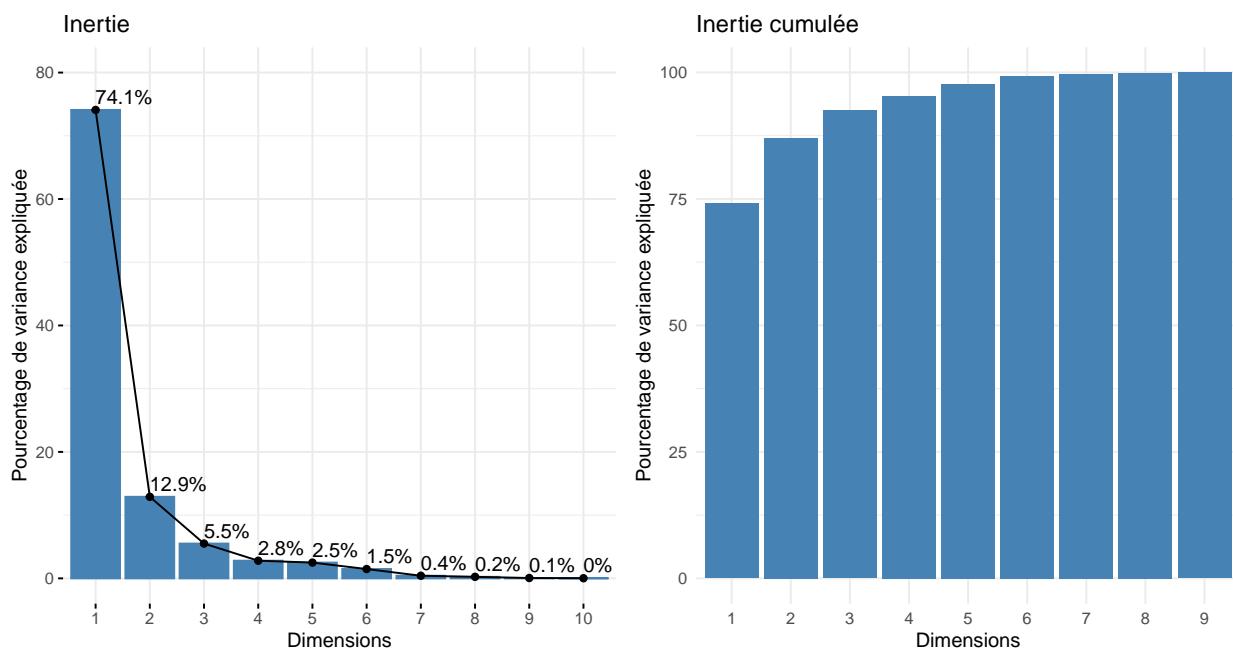
2 L'analyse

2.1 L'analyse par composante principale



2.1.1 Inerties

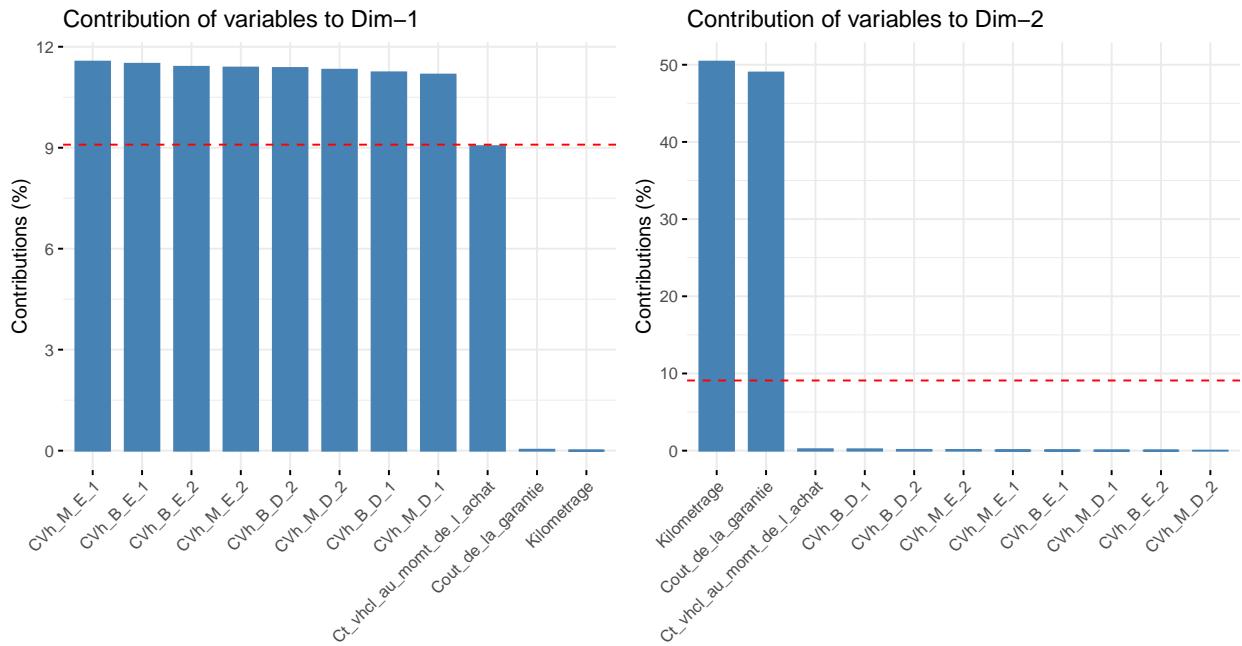
	1	2	3	4	5	6	7	8	9	10	11
Inerties	8.1	1.4	0.6	0.3	0.3	0.2	0.0	0.0	0.0	0	0
Inerties relatives %	74.1	12.9	5.5	2.8	2.5	1.5	0.4	0.2	0.1	0	0
Inerties relatives cumulées	74.1	87.0	92.5	95.3	97.8	99.3	99.7	99.9	100.0	100	100

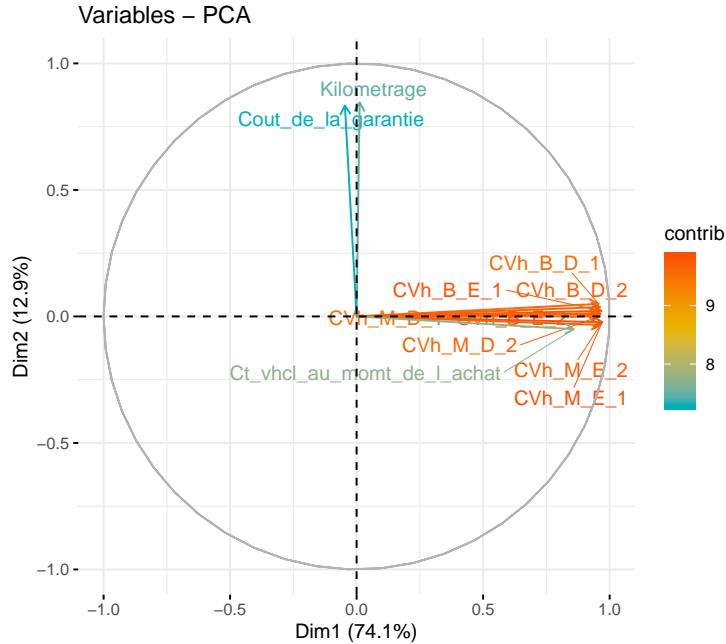


L'étude des inerties nous conduit à sélectionner deux axes pour notre étude, qui à eux deux enregistrent 86% de l'inertie totale (74% pour l'axe 1 et 13% pour l'axe 2).

2.1.2 Etudes des variables

	Coord. principales			Contribution %			Qualité de représentation			Qualité.axes double	
	F1	F2	F3	F1	F2	F3	F1	F2	F3	Qual(F1,F2)	Qual(F2,F3)
Kilometrage	0.01	0.85	-0.52	0.00	50.40	44.75	0.00	0.72	0.27	0.72	0.99
CVh_M_E_1	0.97	-0.02	0.04	11.56	0.04	0.22	0.94	0.00	0.00	0.94	0.00
CVh_B_E_1	0.97	0.02	0.03	11.49	0.03	0.12	0.94	0.00	0.00	0.94	0.00
CVh_M_D_1	0.95	0.01	-0.07	11.17	0.01	0.76	0.91	0.00	0.00	0.91	0.00
CVh_B_D_1	0.96	0.05	-0.07	11.24	0.17	0.80	0.92	0.00	0.00	0.92	0.01
CVh_M_E_2	0.96	-0.03	0.04	11.38	0.08	0.33	0.93	0.00	0.00	0.93	0.00
CVh_B_E_2	0.96	0.01	0.04	11.40	0.01	0.20	0.93	0.00	0.00	0.93	0.00
CVh_M_D_2	0.96	0.00	-0.05	11.32	0.00	0.35	0.92	0.00	0.00	0.92	0.00
CVh_B_D_2	0.96	0.04	-0.05	11.37	0.10	0.38	0.93	0.00	0.00	0.93	0.00
Ct_vhcl_au_momt_de_l_achat	0.86	-0.05	0.14	9.05	0.18	3.00	0.74	0.00	0.02	0.74	0.02
Cout_de_la_garantie	-0.05	0.83	0.55	0.03	48.99	49.08	0.00	0.70	0.30	0.70	0.99



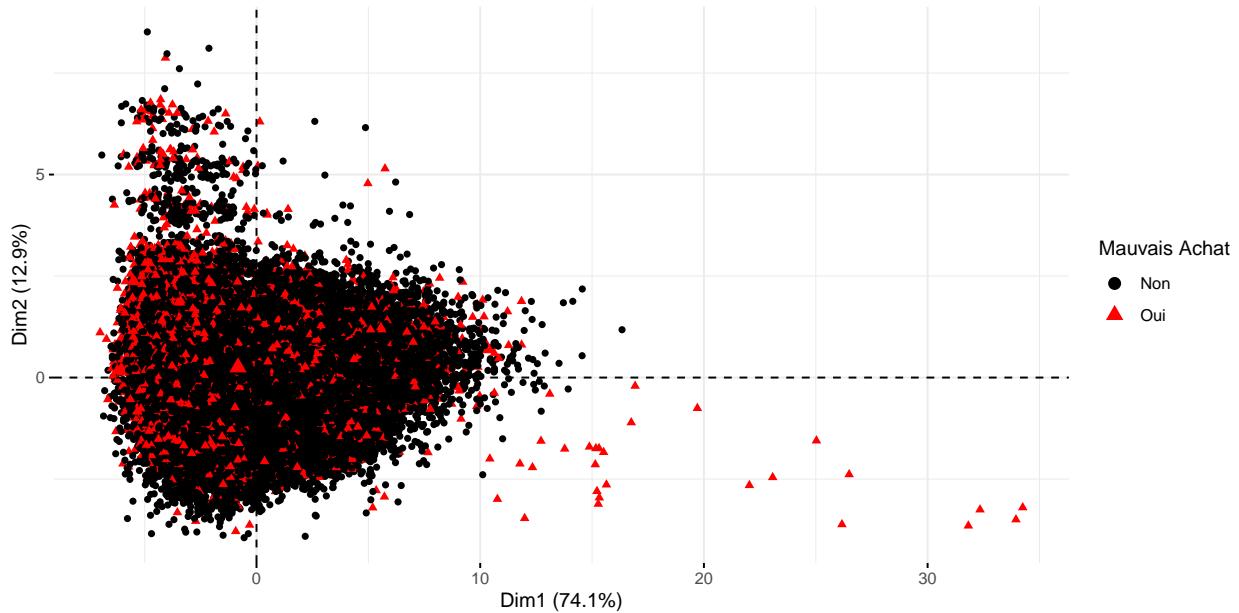


Axe 1 : Les variables qui contribuent le plus à la formation de cet axe sont les variables correspondant aux différentes côtes des véhicules. En effet, chacune de ces variables contribue positivement (environ 11%) à la formation de cet axe. De plus, la variable *Cout Vehicule Au Moment de L Achat* (8,5% de contribution à l'axe 1) est très proche de la contribution moyenne qui est d'environ 9%. Cette variable, étant donnée qu'elle représente un prix comme les côtes, peut être intégrée à l'interprétation de cet axe. De plus, toutes ces variables sont très bien représentées (supérieur à 0,9 pour toutes les côtes et environ 0,7 pour le *Cout Vehicule Au Moment de L Achat*). La corrélation entre toutes ces variables nous permettait d'avoir cette intuition. Cet axe représente le prix (via les côtes principalement) des voitures.

Axe 2 : Cet axe représente beaucoup moins le nuage que l'axe 1 mais il reste tout de même intéressant à étudier. Les variables *kilométrages* et *Cout de la garantie* contribuent très majoritairement à la construction de cet axe (respectivement 45% et 49%). La qualité de représentation n'est pas excellente mais reste correcte (environ 0,7 pour les deux variables). Cet axe représente donc le kilométrage et la garantie des véhicules.

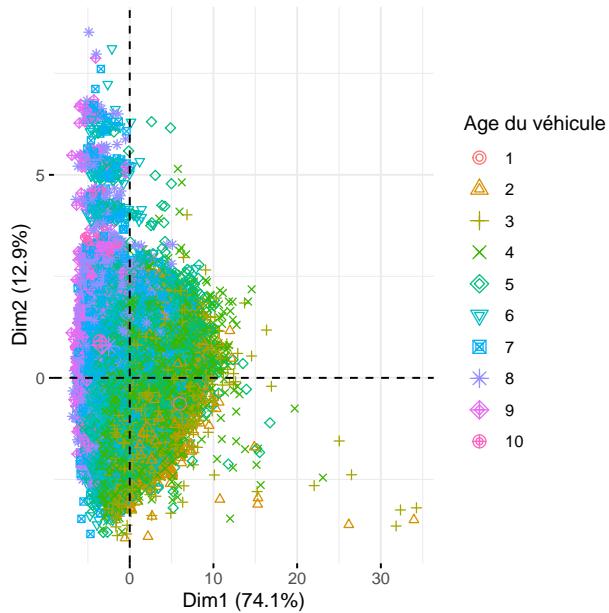
2.1.3 Etudes des individus

Individuals – PCA

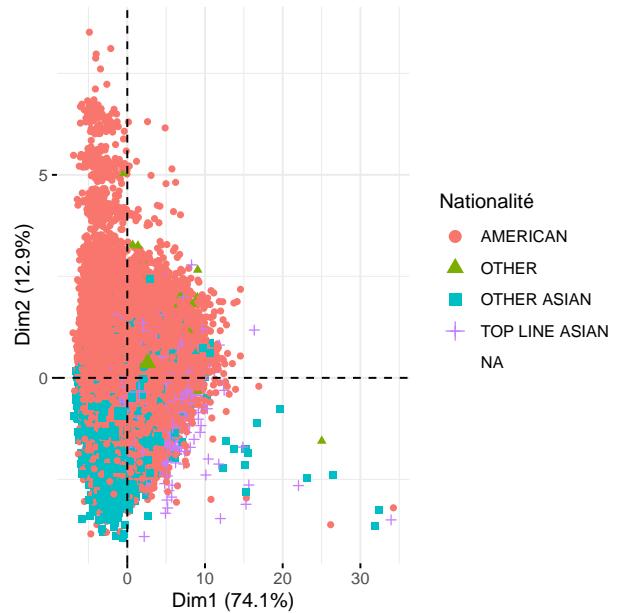


Cette représentation du nuage de points des individus met en évidence les véhicules identifiés comme “lemon” ou non. Ce graphique ne permet pas d’observer une vraie corrélation entre les variables quantitatives et la variable *Mauvais Achat*. Cependant, la partie inférieure droite du graphique laisse penser que les voitures ayant une côte trop élevée, par rapport à la moyenne, conjugué à un coût de garantie et un kilométrage faible pourraient être de mauvais achat. Cette analyse doit toutefois être modérée étant donné le peu d’individus concernés (une trentaine au maximum)

Individuals – PCA



Individuals – PCA



Ces deux nouvelles représentations graphiques du nuage de points des individus permettent une analyse pour les compagnies d’enchère. En effet, l’habillage permet d’observer un rapport entre prix et âge du véhicule et, dans une moindre mesure, entre prix et nationalité du véhicule. Les véhicules les plus agés sont souvent

les moins chers tandis que les plus récents ont un prix plus élevé. Par conséquent, pour gagner plus, les entreprises d'enchère doivent vendre des véhicules plutôt récent. Le deuxième graphique met en évidence un lien entre prix et nationalité du véhicule. Les voitures des top constructeurs asiatiques sont plus chères que celle des autres constructeurs asiatiques. Les voitures américaines sont en revanche dispersées dans les gammes de prix. Afc correspondante dans l'annexe.

2.2 L'analyse factorielle des composantes

Nous essayons ici d'observer les caractéristiques classiques des ventes et ainsi de trouver celle qui s'éloigne de ces profils moyens.

2.2.1 L'acheteur

La localisation de l'acheteur a t-il un lien avec l'achat de tel ou tel véhicule?

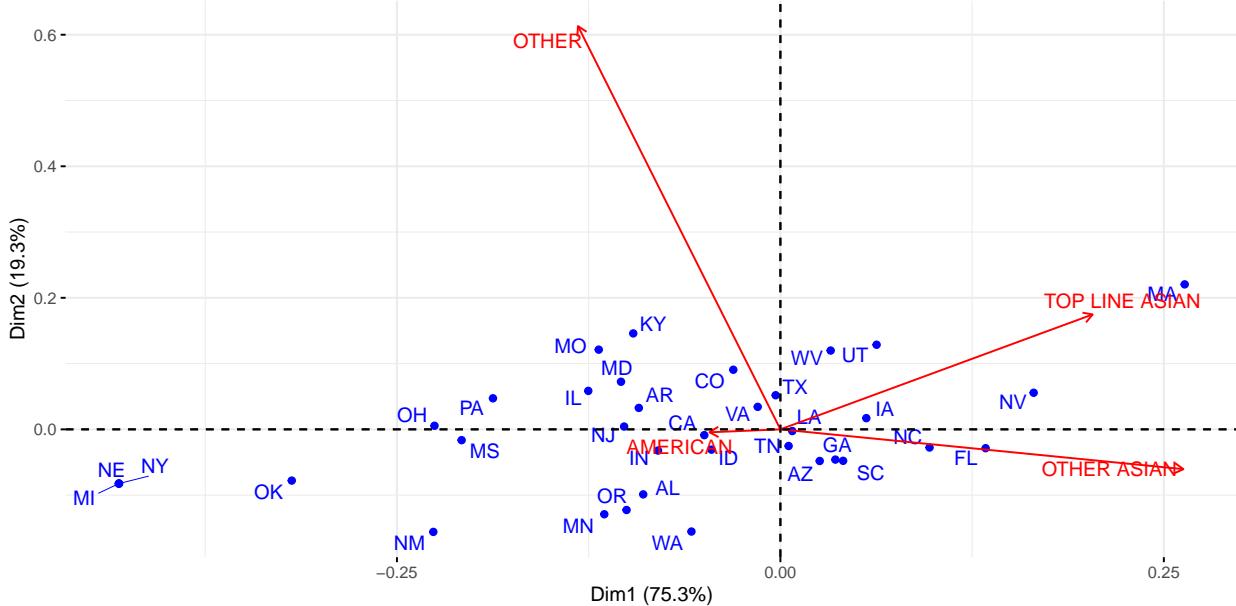
Les tableaux de contingence entre l'État d'origine de l'acheteur, la nationalité du véhicule, sa taille et son âge ayant été réalisés, des différences importantes ressortent. Les données suivantes seront analysées sans les ventes de véhicules neufs et sans les acheteurs du New Hampshire car leur inertie est bien trop importante et gène la visualisation. Ces données seront affichées dans les annexes.

Table 1: Profils colonnes

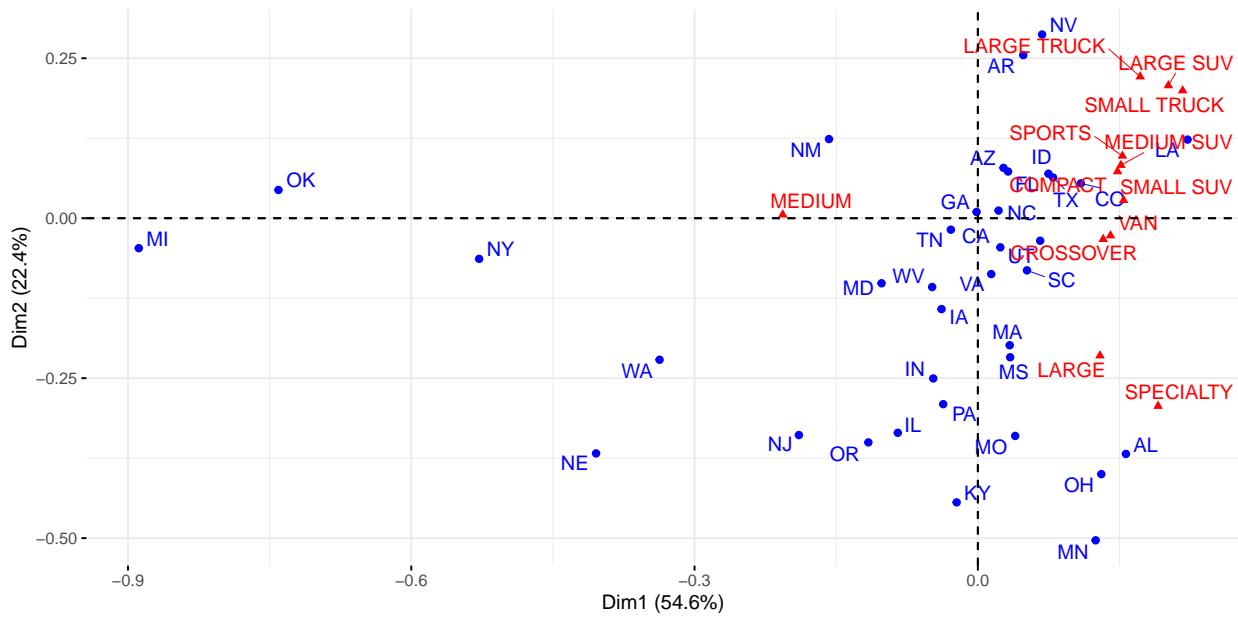
	Poids	Coord. principales			Inertie		Contribution %			Qualité de rep		
		F1	F2	F3	Inertie	Inertie relative	F1	F2	F3	F1	F2	F3
AMERICAN	0.84	-0.05	0.00	0.00	0.00	11.84	15.57	0.57	0.12	0.99	0.01	0.00
OTHER	0.00	-0.13	0.61	0.46	0.00	10.35	0.40	33.67	65.67	0.03	0.63	0.34
OTHER ASIAN	0.11	0.26	-0.06	0.03	0.01	52.80	65.94	13.66	9.42	0.94	0.05	0.01
TOP LINE ASIAN	0.05	0.20	0.18	-0.06	0.00	25.00	18.09	52.10	24.79	0.55	0.40	0.05

Les tableaux pour les analyses suivantes seront mis en annexes.

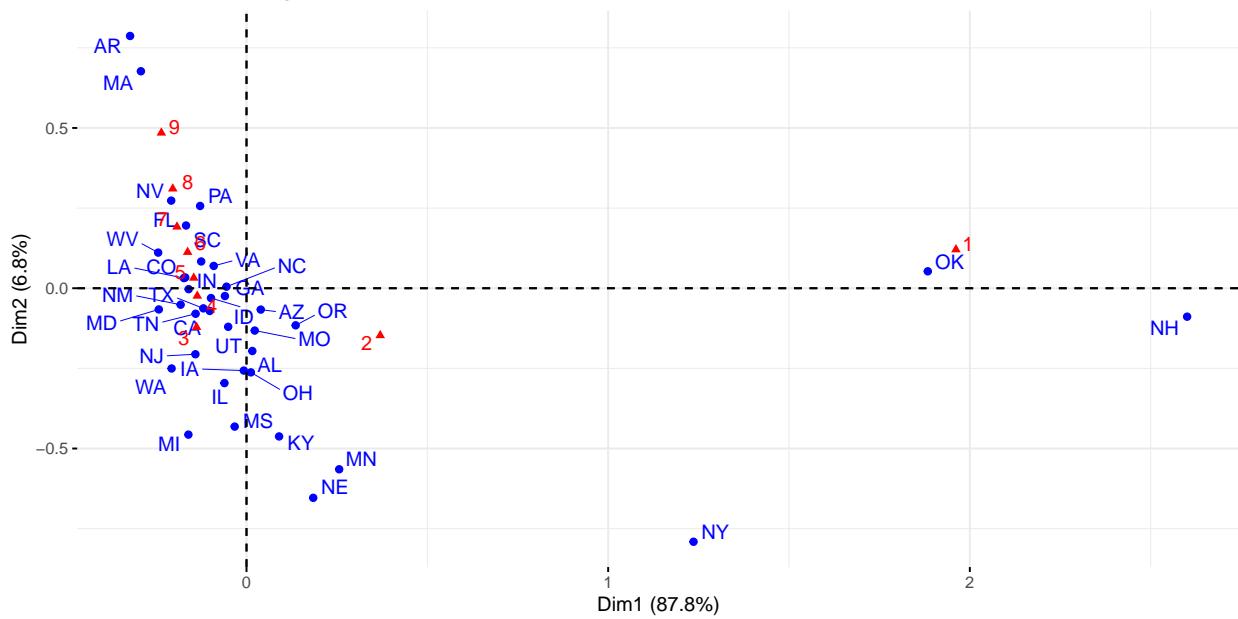
État de l'acheteur/nationalité du véhicule



État d'origine de l'acheteur/Taille du véhicule



État de l'acheteur/âge du véhicule



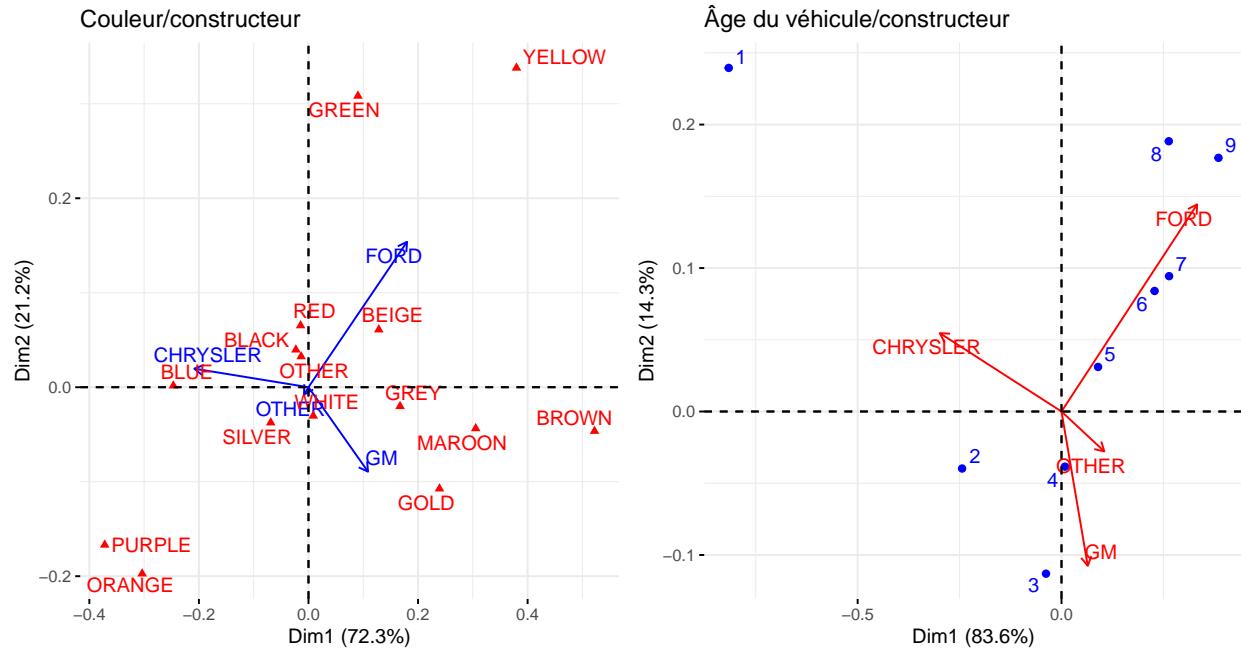
Cette première analyse nous montre que les véhicules achetés par des acheteurs des États du Massachusetts, Nevada et Utah s'éloignent de la même façon du profil moyen que les véhicules Top Line Asian. Il semblerait que s'éloigner vers la droite de l'axe c'est retrouver des profils de véhicules asiatiques.

Dans le second graphique, les profils qui s'éloigne le plus de l'axe tels que Miami et Oklahoma sont des profils avec très peu d'effectifs. Mais ils sont très largement associés aux tailles de véhicules médium.

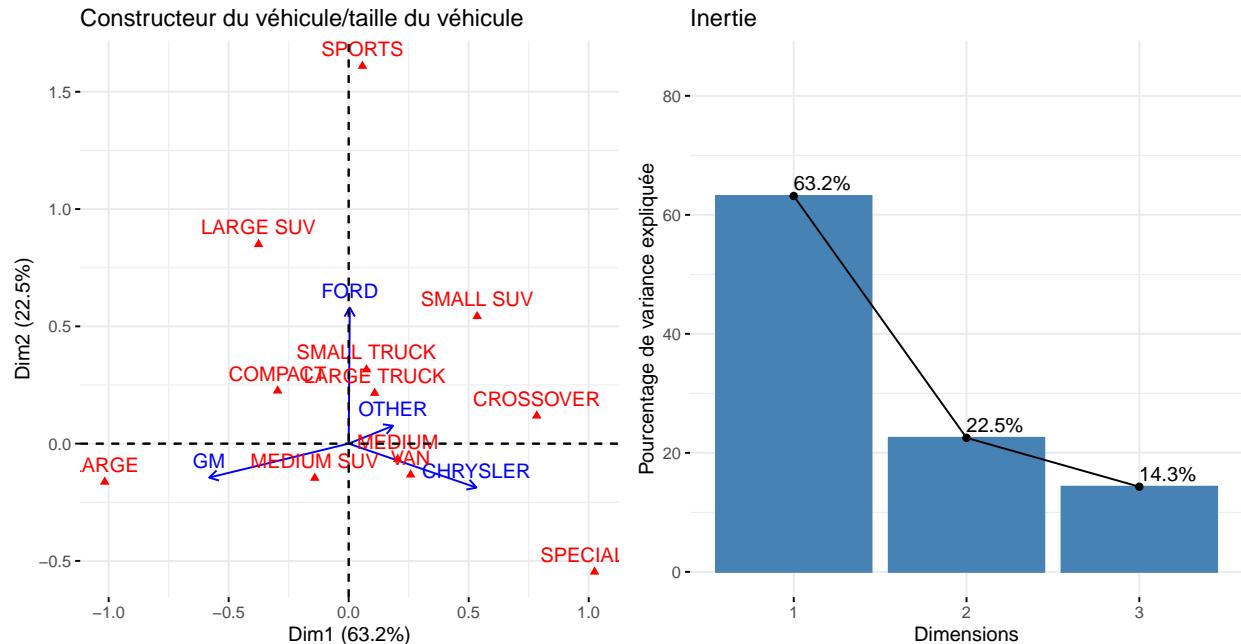
Dans le troisième graphique, les acheteurs venant d'Oklahoma et New Hampshire s'éloignent de la même façon que les véhicules neufs. Le reste n'est pas clair, mais on peut discerner une dynamique vers le haut, à gauche de l'éloignement du profil moyen selon l'âge du véhicule.

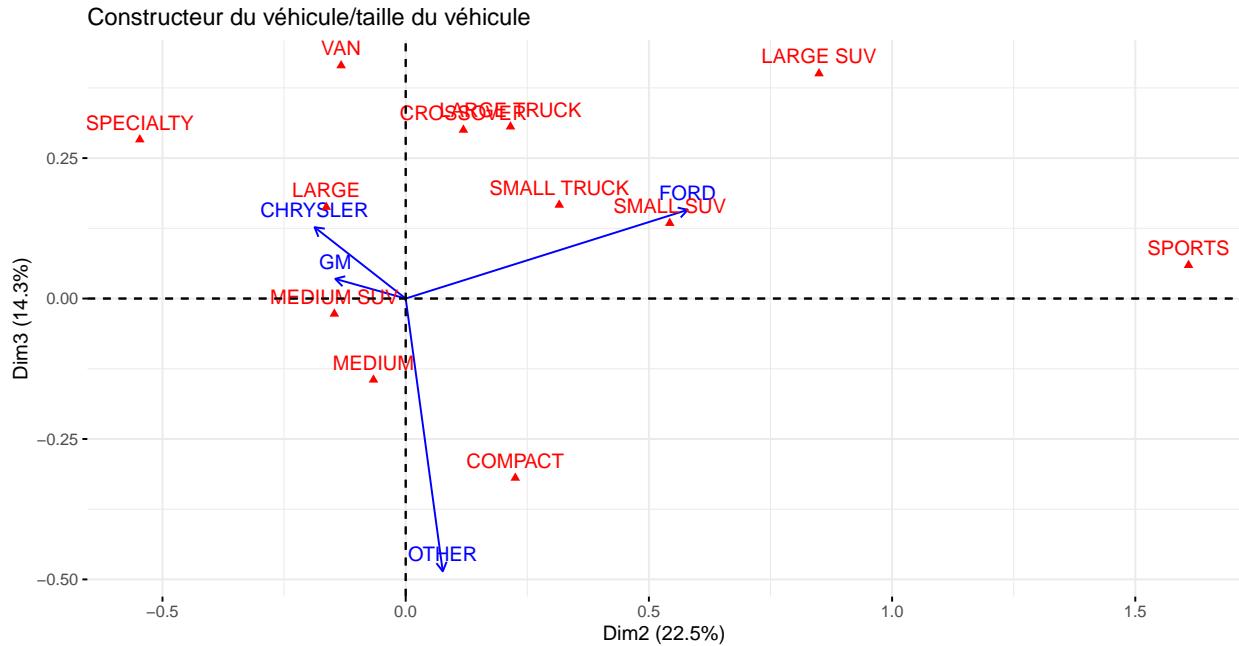
2.2.2 Le vendeur

Quel type de véhicule peut-on trouver selon tel ou tel constructeur?



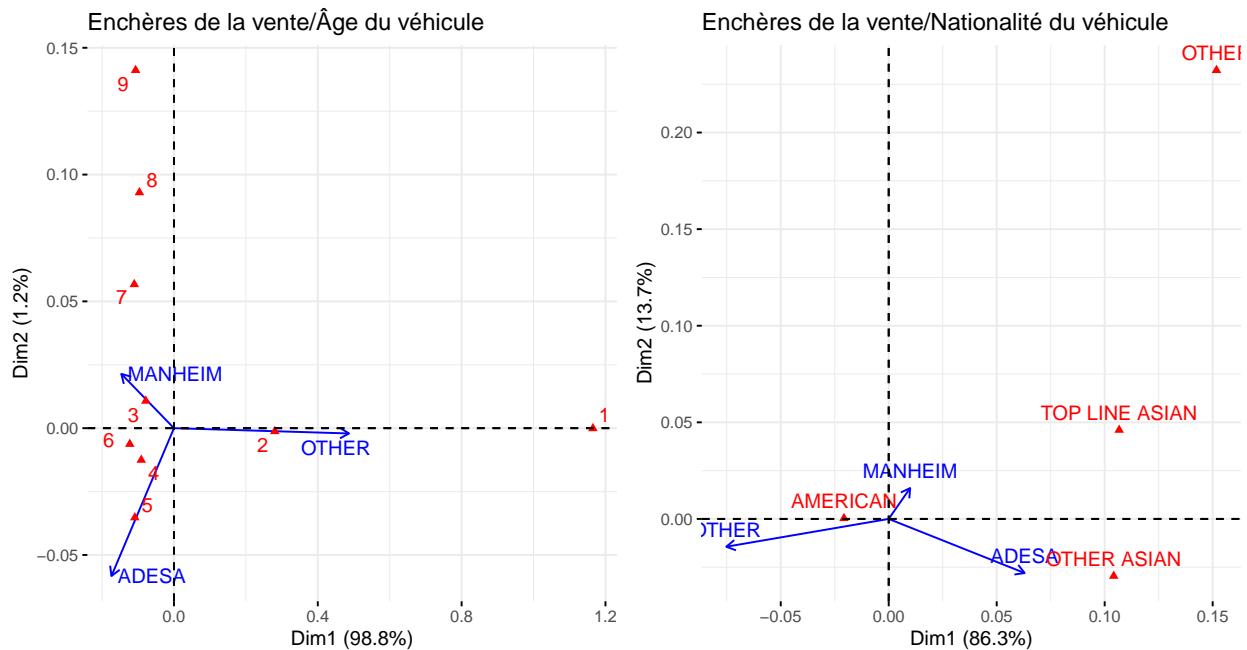
Le constructeur Ford s'éloigne de la même façon que les véhicules les plus vieux du profil moyen, les occurrences de véhicules plus âgés sont plus importantes chez le constructeur Ford que chez le profil moyen. Il semblerait de la même façon que véhicules Chrysler et véhicules âgés s'éloignent de la même façon.



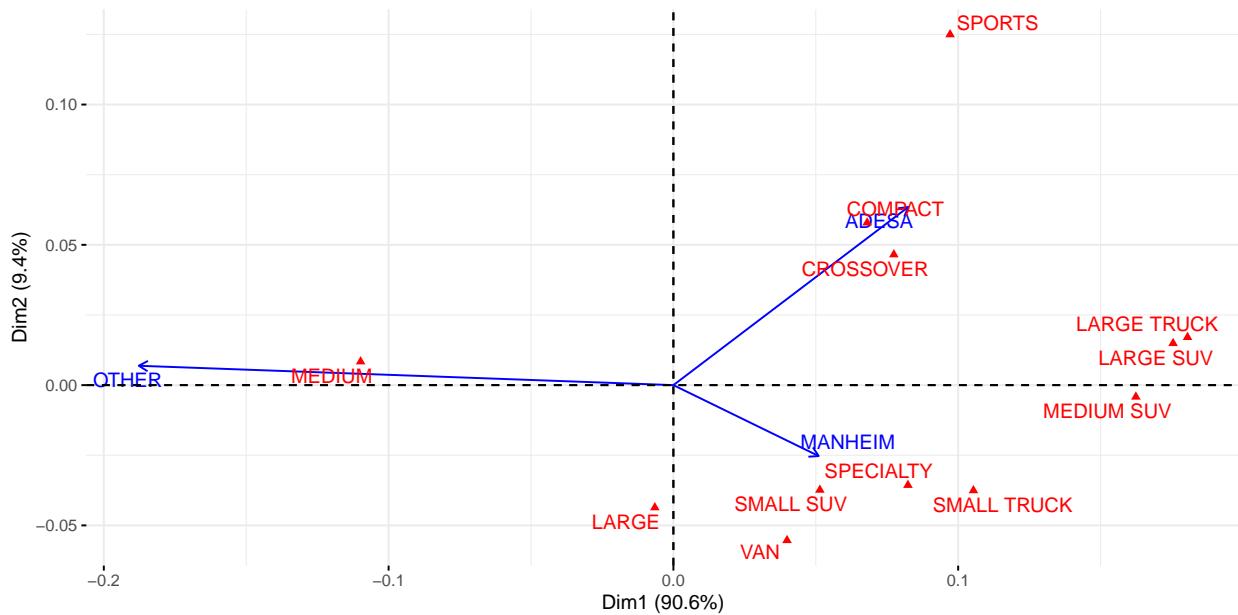


2.2.3 Les enchères

Quel type de véhicule se trouve plus fréquemment chez tel ou tel enchère?

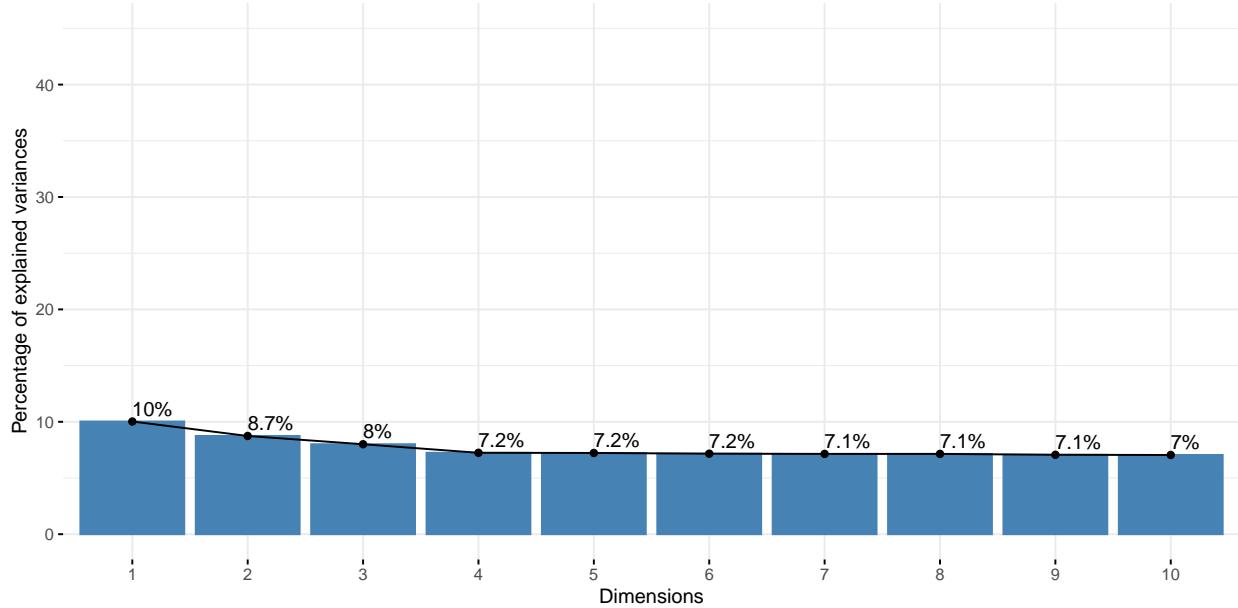


Enchères de la vente/Taille du véhicule



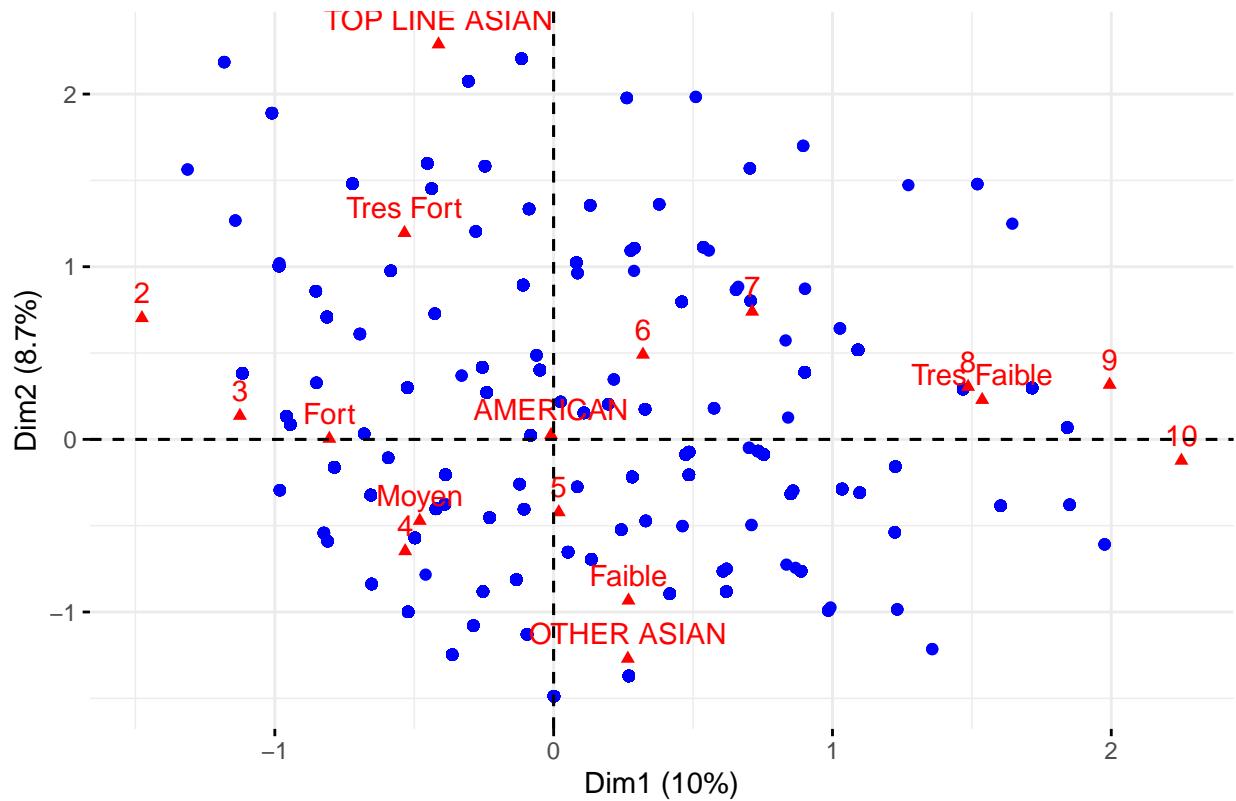
2.2.4 Analyse des correspondances multiples

Scree plot



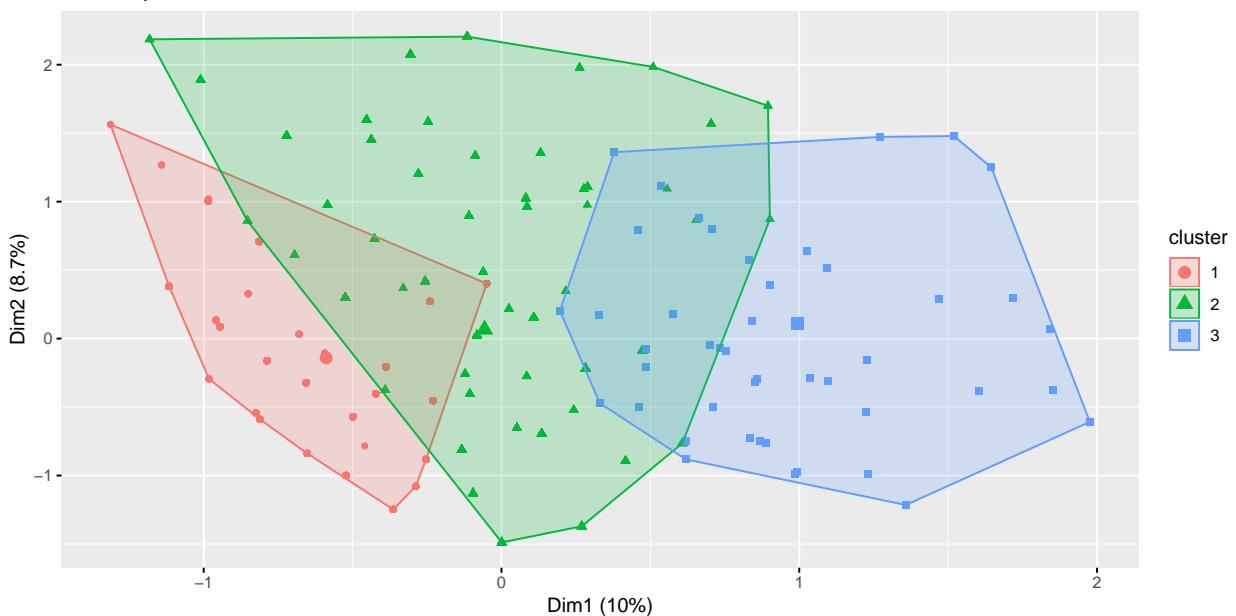
Nous avons réalisé une ACM sur les variables coûts du véhicule, âge du véhicule et nationalité du véhicule. Malheureusement nos ordinateurs ne nous permettant pas de réaliser unclustering sur l'ensemble des individus, nous avons donc pris un échantillon de 10 000 individus aléatoirement tiré au sein de notre base de données.

MCA – Biplot



En observant le nuage des variables on constate que le prix semble corrélé avec l'âge du véhicule en effet plus on se trouve à droite de l'axe 1 plus l'âge des véhicules augmente et plus son prix diminue. À l'inverse plus on se trouve à gauche de l'axe 2 plus le prix augmente et plus l'âge diminue. De plus, les voitures appartenant à la classe "Top line asian" semblent corrélées avec des prix très fort alors que la classe de voitures "Other asian" a l'air d'être corrélé aux prix faibles.

Cluster plot



Bien que nous n'ayons pas réalisé le clustering sur l'ensemble des individus il semblerait que trois groupes d'individus se dégagent.