Université Cheikh Anta Diop

Faculté des Sciences et Techniques

Département de Mathématiques et Informatiques

Master 1 Modélisation Statistique et Informatique

Projet à rendre :

Analyse des composantes principales

Présenté par :

MAYENE Bienvenue Schékina

Sommaire :

Introduction :

La place des données dans la vie d'un pays, d'une entreprise ou d'un individu est incontestable. En effet, toute activité ou toutes décisions requièrent un certain nombre d'information. Pour tirer des informations adéquates, il nous faut employer des bonnes méthodes d'où l'intérêt d'utiliser des méthodes d'analyses multidimensionnelles. Parmi celles-ci nous avons la méthode d'analyse des composantes principales.

L'analyse des composantes principales est une méthode d'analyse multidimensionnelle qui s'applique à un tableau de données croisant des individus en ligne et des variables qualitatives ou quantitatives en colonnes.

Cette technique statistique nous permet de résumer l'information contenue dans un vaste tableau de données quantitatives à partir des représentations graphiques.

Notre travail portera sur la base de données nommée ***« Emission de CO2 des voitures »***. Il consistera tout d'abord à décrire cette base de données, ensuite réaliser l'Analyse des Composantes Principales (ACP) sur celle-ci et enfin interpréter les résultats obtenus. Notons que nous nous préoccuperons des données allant de la ligne 279 à 309.

I-/ Présentation des données

a) - Description de la base :

La base de données sur laquelle reposera notre travail est intitulé ***« Emission de CO2 des voitures ».*** Cet ensemble de données relève les détails sur la façon dont les émissions de CO2 d'un véhicule peuvent varier selon différentes caractéristiques. L'ensemble de données a été extrait du site Web officiel de données ouvertes du gouvernement du Canada. Ceci est une version compilée qui contient des données sur une période de 7 ans.

Il y a un total de 2053 observations sur 10 variables.

Signification des variables :

L'ensemble des variables de la base ont été codées en anglais. Notons que la marque et le modèle de voiture dans notre cas sont combinés et représentent ici le nom de nos individus c'est à dire le nom des voitures observées.

- **Engine size** : La taille du moteur
- **Cylinders** : Nombre de cylindres
- **Vehicule class :** Les modalités de cette variable sont au nombre de 16 et sont les suivantes : "COMPACT", "FULL-SIZE" "MID-SIZE", "MINICOMPACT", "MINIVAN", "PICKUP TRUCK - SMALL","PICKUP TRUCK -STANDARD", "SPECIAL PURPOSE VEHICLE", "STATION WAGON - MID-SIZE" "STATION WAGON - SMALL", "SUBCOMPACT", "SUV – SMALL ", "SUV - STANDARD" , "TWO-SEATER", "VAN - CARGO" et "VAN - PASSENGER".

- **Transmission** : Type de transmission des véhicules

Les modalités de cette variable sont au nombre de 5 et sont les suivantes :Automatic, Automated manual (Manuel automatisé), Automatic with select shift (automatic avec sélecteur de vitesse), Continuously variable et Manual.

- **Fuel type :** Type de carburant

Les modalités de cette variable sont : Regular gasoline, Premium gasoline et Diesel.

Fuel Consumption **: Consommation de carburant**

- **Fuel.Cons.City :** Quantités de consommation de carburant en ville
- **Fuel.Cons.Hwy :** Quantités de consommation de carburant sur autoroute

Fuel.Cons.City et ***Fuel.Cons.Hwy*** sont indiquées en litres aux 100 kilomètres (L/100 km).

- ***Fuel.Cons.Comb*** : Quantité de consommation de carburant combinée (55 % en ville, 45 % sur autoroute) est indiquée en L/100 km. Elle est aussi indiquée en milles par gallon « **Fuel.Cons.Comb.mpg** » (**mpg** ou mi/gal).

Avec 1 **Miles per gallon** (US) [**mpg**] = 235,2 Litre par 100 kilomètres [l/100km]

- **CO2.Emissions :** Emission en CO2 des voitures

b) - Statistiques descriptives sur la base de données

Avant de pouvoir faire des statistiques descriptives sur l'ensemble de la base de données, nous allons procéder au recodage de nos variables qualitatives. En effet, en visualisant nos variables à travers la commande *str(co2_emission)*, nous remarquons que nos variables qualitatives Vehicule.class, Transmission et Fuel.type sont de type character.

Si nous laissons ce type inchangé, nous ne pourrions pas mener correctement notre travail. Nous allons donc modifier le type de ces variables en type factor.

Ensuite, nous allons ordonner nos variables de sorte à mettre en première position nos variables quantitatives. Ceci nous aidera plus tard dans l'analyse des composantes principales que nous allons réaliser.

Les statistiques descriptives que nous allons réaliser vont se faire en deux phases : l'analyse univariée et l'analyse bivarié.

- ***Analyse univariée***

Nous allons étudions la distribution des variables quantitatives

Nous allons réaliser cette analyse à l'aide de la commande *summary(co2_emission)*. Nous pouvons aussi utiliser la commande *describe* pour pouvoir avoir des statistiques sur notre base de données.

Nous remarquons que R nous fournit pour les variables qualitatives les effectifs des modalités et pour les variables quantitatives les statistiques de base (moyenne, médiane, min, max et etc…).

Au risque de tenir compte des valeurs extrêmes, nous allons analyser pour les variables quantitatives la médiane, le minimum et le maximum.

Des résultats obtenus, nous pouvons avoir les commentaires suivants :

- 50% des voitures de notre base ont une taille de moteur inférieure à 3 tandis que les 50 autres pourcents ont en une supérieure à 3. La plus petite taille des moteurs que nous avons dans notre base est 1 et la plus grande est 8,4.

- 50% des voitures de notre base ont un nombre de cylindre inférieur à 6 tandis que les 50 autres pourcents ont un nombre de cylindre supérieure à 6. Le plus petit nombre de cylindre que nous observons dans notre base est 3, le maximum est 16

- 50% des voitures de notre base ont une Quantité de consommation de carburant combinée inférieure à 27 mi/gal tandis que les 50 autres pourcents ont en une

supérieure à cette valeur. La plus grande consommation de carburant combinée est de 69 mi/gal tandis que ka plus petite est de 13 mi/gal

- 50% des voitures de notre base ont une émission de co2 inférieure à 248 tandis que les 50 autres pourcents ont en une supérieure à 248. Nous pouvons voir que la plus grande émission de co2 sur l'ensemble des voitures de notre base est de 522, la plus petite étant de 96
- 50% des voitures de notre base ont une Quantité de consommation de carburant combinée inférieure à 10,6 L/100 km tandis que les 50 autres pourcents ont en une supérieure à cette valeur. Nous observons La plus grande consommation de carburant combinée est de 22,2 L/100 km tandis que ka plus petite est de 4,1 L/100 km
- 50% des voitures de notre base ont une Quantités de consommation de carburant en ville inférieure à 12,10 L/100 km tandis que les 50 autres pourcents ont en une supérieure à cette valeur. Nous observons La plus grande consommation de carburant combinée est de 26,8 L/100 km tandis que ka plus petite est de 4,2 L/100 km

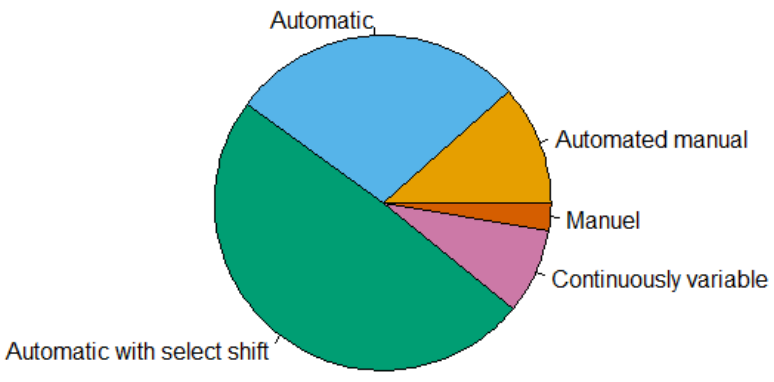
- 50% des voitures de notre base ont une Quantités de consommation de carburant sur autoroute inférieure à 8,7 L/100 km tandis que les 50 autres pourcents ont en une supérieure à cette valeur. Nous observons La plus grande consommation de carburant combinée est de 17,9 L/100 km tandis que ka plus petite est de 4,0 L/100 km

Pour les variables qualitatives nous allons étudier la Répartition des modalités à l'aide de la commande table.

Tableau des effectifs /pourcentages de la variable type de transmission du moteur

Modalités de la variable	Effectif	Pourcentage
Automated manual	238	11,59%
Automatic	582	28,35%
Automatic with select shift	1010	49,20%
Continously variable	168	8,18%
Manuel	55	2,68%

Répartition du type de transmission des moteurs

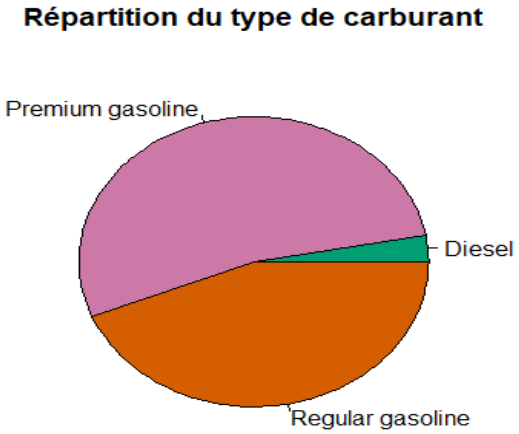


D’après le tableau des effectifs et fréquences des modalités et d’après le diagramme circulaire ci-dessus, nous pouvons voir que le type de transmission du moteur le plus courant est le type « *Automatic with select shift* » avec près de 49,20%.

Le type de transmission le moins présent dans la base est le type « *Manuel* » avec 2,68%

Tableau des pourcentages de la variable type de carburant 'Fuel.Type'

Modalités de la variable	Effectifs	Pourcentage
Diesel	61	2,97%
Premium Gasoline	1093	53,24%
Regular gasoline	899	43,79%



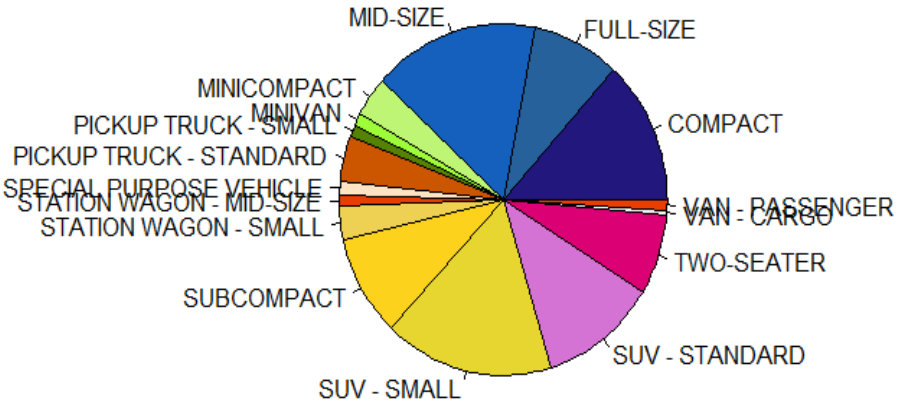
D’après le tableau des effectifs et fréquences des modalités et d’après le diagramme circulaire ci-dessus, nous pouvons voir que le type de carburant le plus courant est le type « *Premium gasoline*» avec près de 53,24%. Le type de carburant le moins présent dans la base est le type « Diesel » avec 2,97%

Tableau des pourcentages de la variable 'Vehicule.Class'

Modalités de la variable	Effectifs	Pourcentage
Compact	272	13,25%
Full-size	178	8 ,67%
Mid-size	335	16,32%
Minicompact	74	3,60%
Minivan	25	1,22%
Pickup Truck-small	21	1,02%
Pickup Truck-standard	86	4,19%
Special Purpose Vehicle	27	1,32%
Station Wagon-Mid-size	21	1,02%
Station Wagon-Small	64	3,12%
Subcompact	190	9,25%

Suv-Small	344	16,76%
SUV-Standard	235	11,45%
Two-Seater	152	7,40%
Van-Cargo	9	0,44%
Van-Passenger	20	0,97%

Répartition des classes de véhicules



D'après le tableau des effectifs et fréquences des modalités et d'après le diagramme circulaire ci-dessus, nous pouvons voir que le type de véhicule le plus courant est le type « *SUV-SMALL*» avec près de 16,76%

Le type de vehicule le moins présent dans la base est le type « Van-Cargo» avec 0,44%

- ***Analyse bivariée***

Dans cette première étude, nous allons chercher à voir l'effet du type de transmission du moteur sur le type de consommation en carburant.

Nous aurons donc comme variable indépendante (VI) = 'Transmission' et comme variable dépendante (VD) = 'Fuel Type'. Nous obtenons les résultats suivants :

Tableau des pourcentages colonnes

	Automated manual	Automatic	Automatic with select shift	Conctinuously variable	Manuel	Ensemble
Diesel	1,3	4,6	3,1	0,0	0,0	3,0
Premium gasoline	84,5	39	60,7	14,9	49,1	53,2
Regular gasoline	14,3	56,4	36,2	85,1	50,9	43,8
Total	100	100	100	100	100	100

Interprétation :

Nous pouvons voir sur le tableau que pour la modalité 'premium gasoline', nous remarquons qu'il y'a plus de voiture ayant pour type de transmission du moteur le type « Automated Manual » qui consomme ce genre de carburant. Les voitures qui consomment le moins le premium gasoline sont de type de transmission « manuel ».

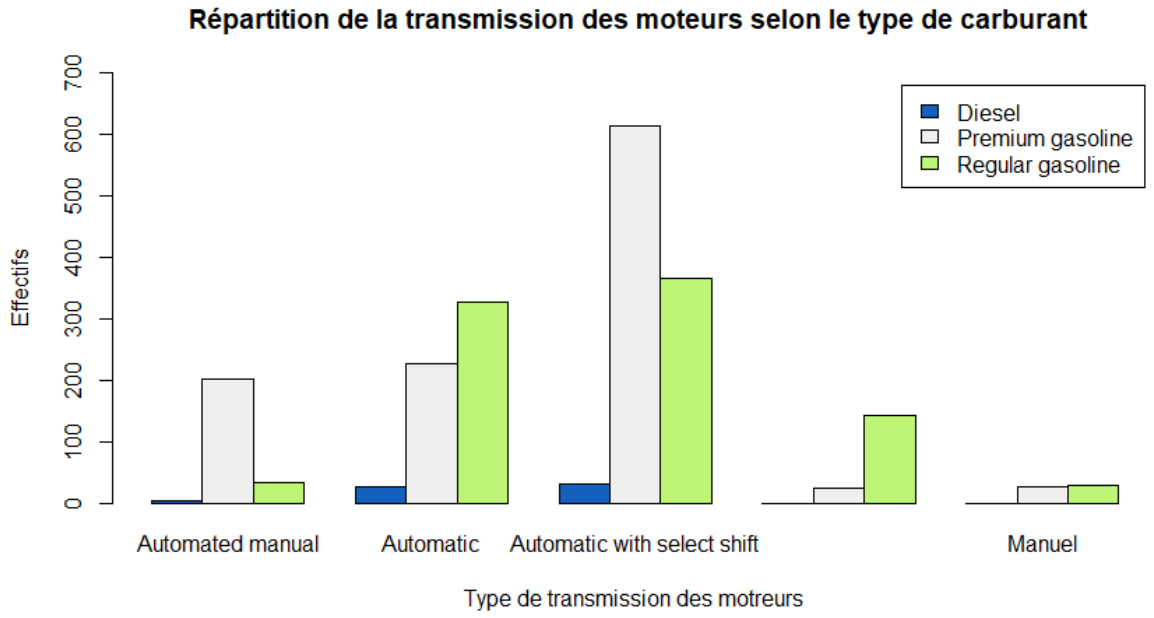
Tableau des pourcentages lignes

	Automated manual	Automatic	Automatic with select shift	Conctinuously variable	Manuel	Ensemble
Diesel	4,9	44,3	50,8	0,0	0,0	100
Premium gasoline	18,4	20,8	56,1	2,3	2,5	100
Regular gasoline	3,8	36,5	40,7	15,9	3,1	100
Total	11,6	28,3	49,2	8,2	2,7	100

Interprétation :

Nous voyons d'après les données que les modèles de transmission 'automatic' représentent 28% des voitures en général, mais seulement 21% des premium gasoline contre 44% des diesel. Les modèles de transmission 'Manuel' représentent seulement 2,7% des voitures dans l'ensemble, mais 2,5% des premium gasoline. Notons qu'il n'y a aucun modèle de transmission de type Manuel qui consomme du Diesel

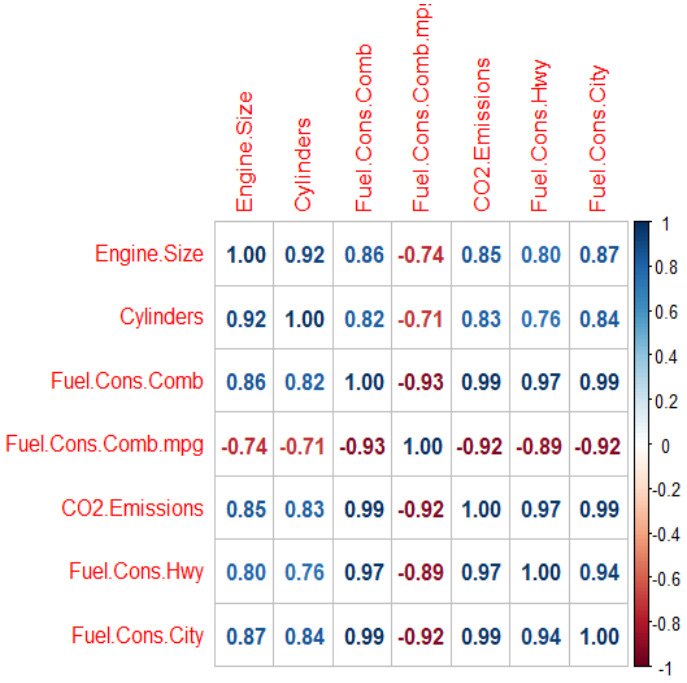
Les résultats obtenus plus haut peuvent s'illustrer également à travers le graphique suivant :



Dans la pratique pour s'assurer qu'il existe réellement une association 'significative' entre deux variables qualitatives on utilise un test statistique. Ainsi, dans le cas de figure ou les 2 variables sont qualitatives nous pouvons utiliser le test de chi-2 de comparaison de deux pourcentages ou proportion. Dans ce test on cherche à rejeter l'hypothèse d'indépendance (l'hypothèse nulle H0) au seuil 5% Entre autres, on cherche à comparer deux pourcentages ou proportions"

Commentaire : La 'p-value' est très significative (p-value < 2.2e-16 c'est à dire inférieur au seuil). Ainsi, on peut affirmer avec une haute certitude qu'il existe ***une association statistique significative*** entre le type de carbutrant (fuel.type) et le type de transmission du moteur (Transmission)

Dans une deuxième étude on va s'intéresser aux potentielle liaisons entre nos variables quantitatives au travers de la matrice de corrélation. La figure ci-dessous illustre les coefficients de corrélations entre chaque variable



Plus la couleur est bleu foncé, plus il existe une corrélation linéaire positive entre les variables. Plus la couleur est rouge foncé, plus il existe une corrélation linéaire négative entre les variables.

Nous voyons par exemple que le coefficient de corrélation entre les variables Engine.Size et CO2.Emission est de 0,85 (proche de 1). On peut donc conclure que plus la taille des moteurs est grande, plus la voiture contenant ce moteur émet du CO2

II-/ Analyse des composantes principales

Avant de commencer l'analyse des composantes principales, nous devons standardiser nos variables c'est-à-dire les centrer et de les réduire. En effet, notre tableau de données regroupe des variables de nature très différentes exprimées dans plusieurs unités. Afin de s'affranchir de l'influence de ces unités dans les calculs, il est préférable de travailler sur des données indépendantes des unités de mesures et des échelles de grandeur. Notons que la fonction PCA issue du package FactomineR normalise automatiquement les données pendant l'ACP.

Le fait d'avoir ordonner plus tôt nos variables nous a permis de mettre en premier les variables quantitatives et ensuite les variables qualitatives. En effet, l'analyse des composantes principales est utilisée lorsque **les variables dites d'intérêts sont quantitatives**. On les appelle encore variables actives

Les variables d'intérêts seront les seuls à participer à la construction des axes et elles seront utilisées pour comparer et décrire les individus Les variables supplémentaires quant à elles caractériserons les groupes d'individus et les relations entre les variables actives.

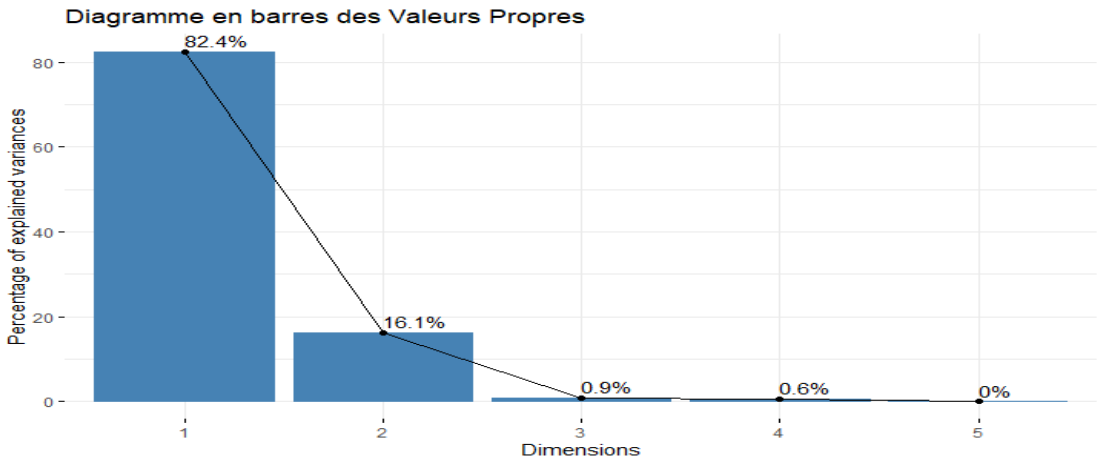
Dans notre étude nous aurons donc les variables quantitatives actives indexées de la position 1 à la position 5, les variables quantitatives supplémentaire indexées de la position 6 à 7 et toutes les variables qualitatives en variables supplémentaires

a-) Choix des axes à retenir

Chaque axe restitue une part de l'information (inertie) contenue dans le nuage des points. Cette part est proportionnelle à la valeur propre associée (eigenvalues en anglais) qui mesurent la quantité de variance expliquée par chaque axe principal. Les valeurs propres sont grandes pour les premiers axes et petits pour les axes suivants.

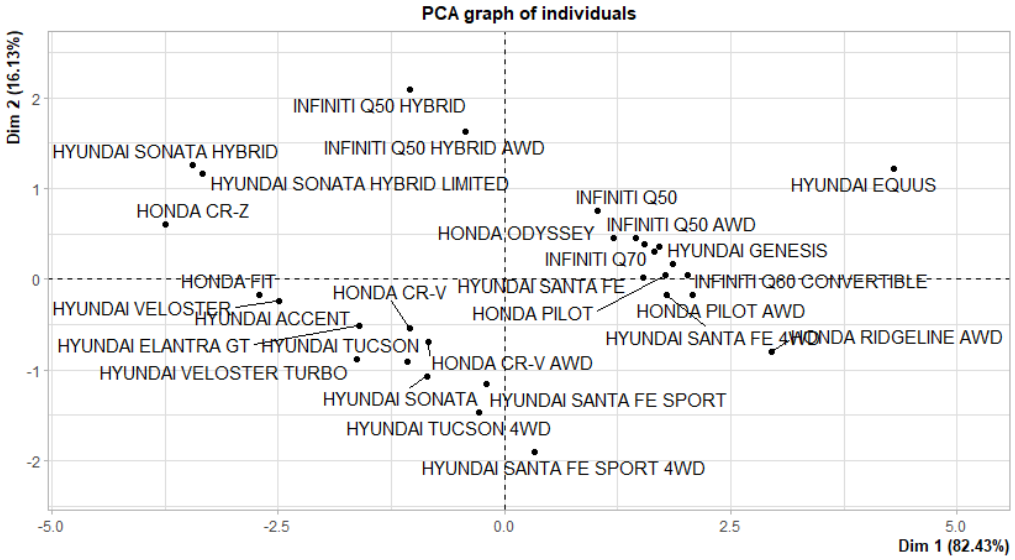
Nous examinons les valeurs propres pour déterminer le nombre de composantes principales à prendre en considération. Les valeurs propres et la proportion de variances (information) retenues par les composantes principales peuvent être extraites à l'aide de la fonction *get_eigenvalue()* du package factoextra.

Plusieurs solutions existent pour déterminer le nombre d'axes à analyser en ACP.
Celle que nous allons utiliser est de regarder le graphique des valeurs propres (appelé scree plot). Le nombre d'axes est déterminé par le point, au-delà duquel les valeurs propres restantes sont toutes relativement petites et de tailles comparables
.



Nous observons sur le graphique ci-dessus qu'à partir de la 3e dimension le pourcentage d'inertie décroît considérablement et devient très insignifiant. Au-delà de la 2e dimension, il y'a quasiment aucune d'information restituée par les autres dimensions. Par conséquent, nous retiendrons les deux premiers axes qui résument à eux deux près de **98,56%** de l'information de l'ensemble.

b-) Description du nuage des points des individus



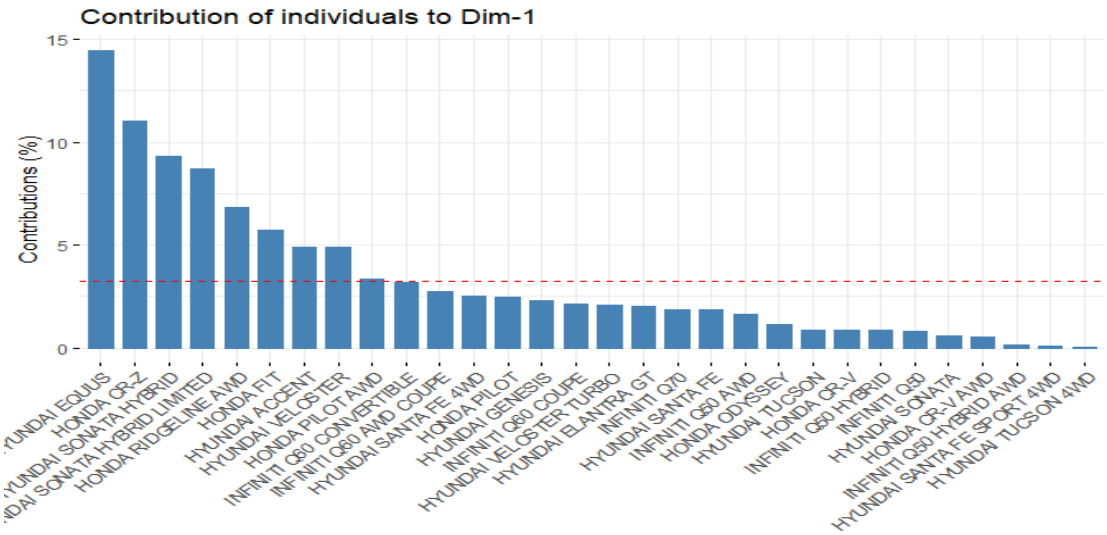
Pour décrire le nuage des points individu, nous allons dans un premier temps nous focaliser sur les individus les plus originaux c'est-à-dire qui ont une grande disto et qui dans un deuxième temps contribuent fortement à la construction de l'axe.

- Commentaires pour l'axe 1

Sur le graphique, nous pouvons voir qu'il y'a des individus qui se démarquent des autres individus et qui sont assez éloignés du centre. Nous remarquons que les véhicules HYUNDAI SONATA HYBRID et HYUNDAI SONATA HYBRID LIMITED sont proches, c'est-à-dire que ces voitures prennent à peu près les mêmes valeurs par rapport aux variables. Ces individus ont

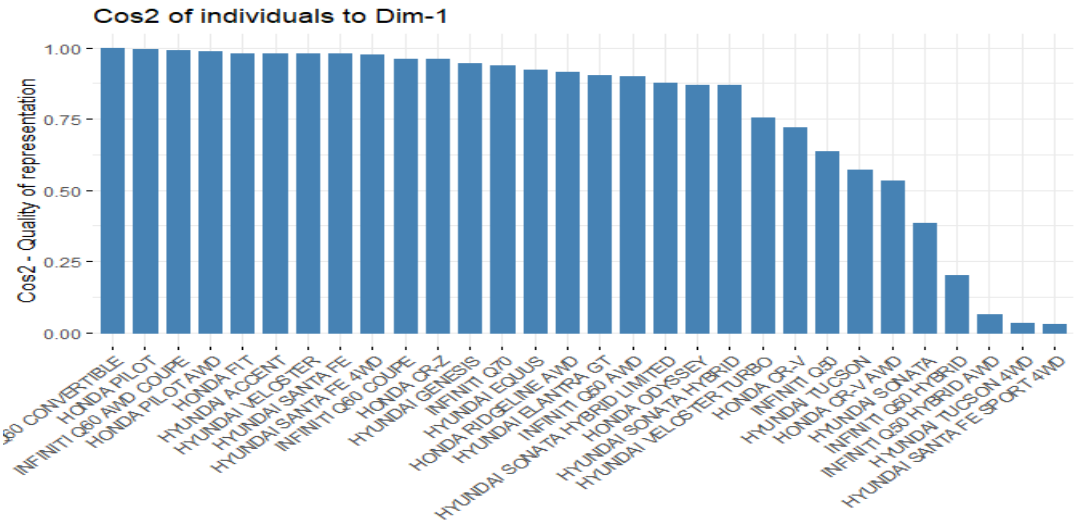
chacun une grande disto et sont tous de coordonnées négatives. Nous dirons que ces individus prennent des fortes valeurs sur l'axe 1.

Comme autres individus atypiques nous remarquons de suite les véhicules de marque HYUNDAI EQUUS et HONDA CR-Z qui sont aussi très éloignés du centre et par conséquent ont une grande disto. Les véhicules de marque HYUNDAI EQUUS sont de coordonnée positive et ceux de marque HONDA CR-Z sont de coordonnés négatives. Nous dirons également que ces individus prennent des fortes valeurs sur l'axe 1



Selon le graph ci-dessus et d'après les résultats obtenus, nous pouvons voir qu'il s'agit de deux groupes qui contribuent très fortement à la construction de l'axe 1 (contribution supérieure à 8).

Nous avons d'une part les véhicules de marque HYUNDAI SONATA HYBRID, HYUNDAI SONATA HYBRID LIMITED et HONDA CR-Z et d'autres part les véhicules de marque HYUNDAI EQUUS. Le fait que ces voitures soient opposées signifie que ce sont des voitures très différentes puisque le 1er axe est celui qui sépare au mieux les points. En effet le premier axe représente près de 82,4% de l'inertie. Donc ces deux groupes voitures ont des comportements très différents et ce sur l'ensemble de leurs caractéristique observables

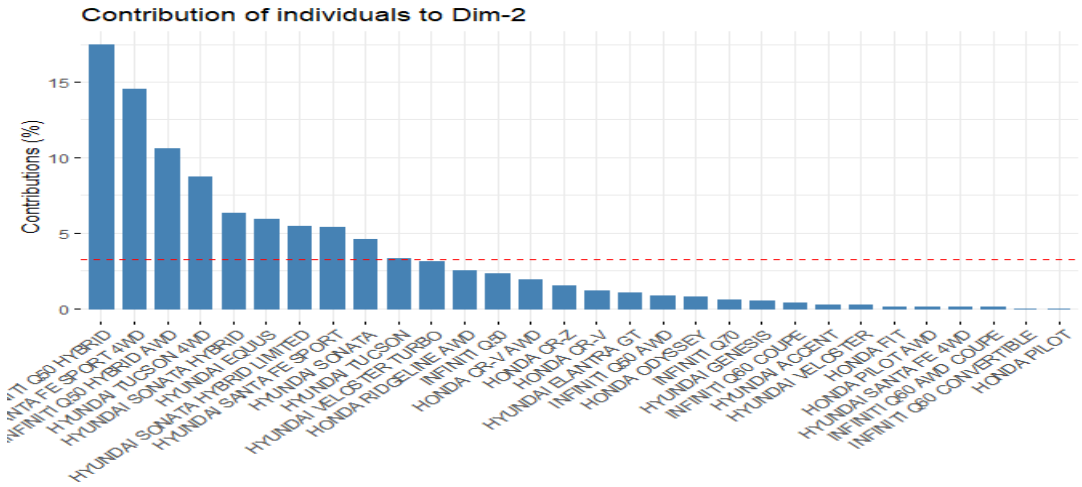


Par ailleurs nous pouvons aussi voir sur le graphique ci-dessus que les individus qui contribuent fortement à la construction de l'axe 1 c'est-à-dire les véhicules de marque HYUNDAI SONATA HYBRID, HYUNDAI SONATA HYBRID LIMITED, HONDA CR-Z et HYUNDAI EQUUS sont tous bien représentés sur cet axe. En effet le cosinus carré de ces individus est très proche de 1.

- Commentaires pour l'axe 2 :

Sur le graphique du nuage des points individus, nous pouvons voir qu'il y'a des individus qui se démarquent des autres individus et qui sont assez éloignés du centre. Nous remarquons que les véhicules HYUNDAI SANTA FE SPORT 4WD et HYUNDAI TUCSON 4WD sont à peu près proches. Ces individus ont chacun une grande disto et sont tous de coordonnées négatives. Nous dirons que ces individus prennent des fortes valeurs sur l'axe 2.

Comme autres individus atypiques nous remarquons de suite les véhicules de marque INFINITI Q50 HYBRID et INFINITI Q50 HYBRID AWD qui sont aussi très éloignés du centre mais ont des coordonnées positives. Ces véhicules sont proches c'est-à-dire qu'elles prennent à peu près les valeurs par rapport aux variables. Nous dirons également que ces individus prennent des fortes valeurs sur l'axe 2.

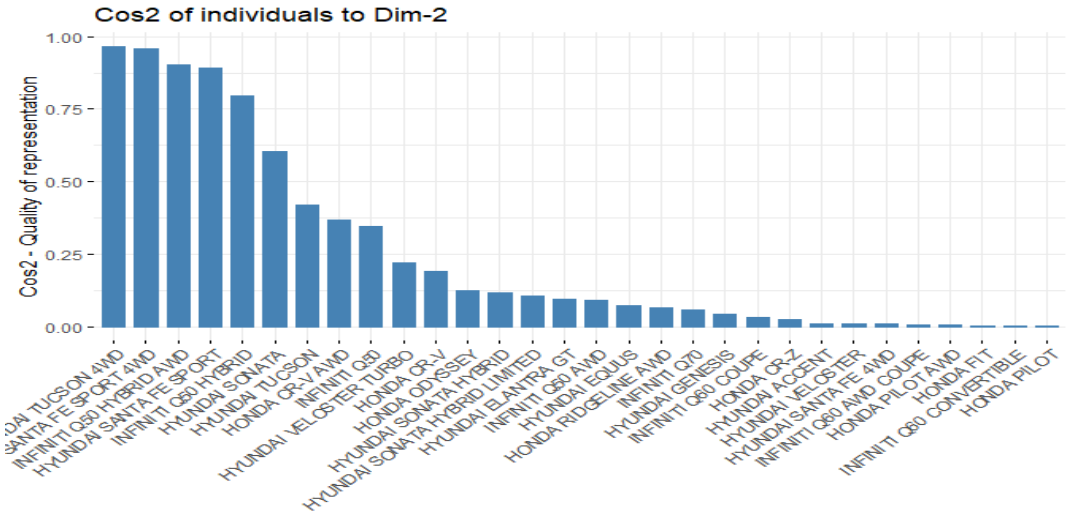


Selon le graph ci-dessus et d'après les résultats obtenus, nous pouvons voir qu'il s'agit de deux groupes qui contribuent très fortement à la construction de l'axe 2 (contribution supérieure à 8).

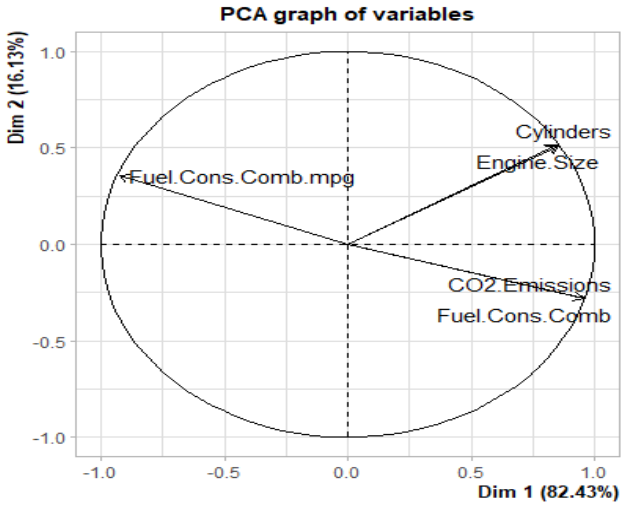
Nous avons d'une part les véhicules de marque HYUNDAI SANTA FE SPORT 4WD, HYUNDAI TUCSON 4WD et d'autres part les véhicules de marque INFINITI Q50 HYBRID et INFINITI Q50 HYBRID AWD. Le fait que ces voitures soient opposées signifie que ces deux groupes voitures ont des comportements très différents sur l'ensemble de leurs caractéristique observables.

Par ailleurs nous pouvons aussi voir sur le graphique ci-dessous que les individus qui contribuent fortement à la construction de l'axe 2 c'est-à-dire les véhicules de marque HYUNDAI SANTA FE SPORT 4WD, HYUNDAI TUCSON 4WD, INFINITI Q50 HYBRID et INFINITI

Q50 HYBRID AWD sont tous bien représentés sur cet axe. En effet le cosinus carré de ces individus est très proche de 1.



c-) Description du nuage des points des variables



La corrélation entre une variable et une composante principale (PC) est utilisée comme coordonnées de la variable sur la composante principale. La représentation des variables diffère de celle des observations : les observations sont représentées par leurs projections, mais les variables sont représentées par leurs corrélations

Sur le graphique ci-dessus nous pouvons voir que les variables Engine.size et Cylinders sont très proches. Le coefficient de corrélation entre ces deux variables est très proche de 1. C'est le cas aussi pour les variables CO2.Emission et Fuel.Cons.comb qui semblent même indistinguable à l'œil nu. Ces variables sont très proches et le coefficient de corrélation entre celles-ci est quasiment proche égal à 1.

- Commentaires pour l'axe 1 :

D'après les résultats de nos données et les observations faites sur le graphe ci-dessus, nous voyons que toutes les variables ont des fortes coordonnées sur l'axe 1. Ces variables sont

donc fortement corrélées à cet axe. Nous voyons aussi que les variables CO2.Emission, Fuel.Cons.comb et Fuel.Cons.Comb.mpg contribuent le plus à la construction de l'axe 1. Les variables CO2.Emission et Fuel.Cons.comb sont de coordonnées positives tandis que la variable Fuel.Cons.Comb.mpg est de coordonnées négatives

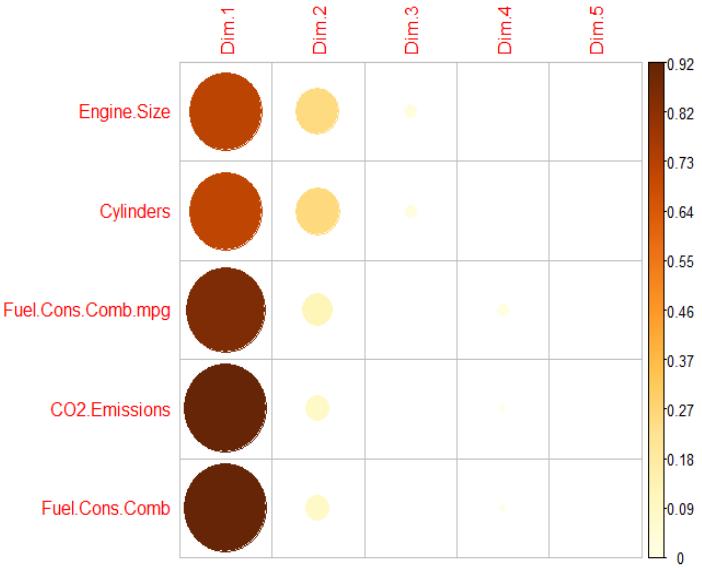
Nous pouvons voir également sur le graphique NLB ci-dessus que ces variables : CO2.Emission, Fuel.Cons.comb et Fuel.Cons.comb.mpg ont des cosinus carré proche de 1 sur l'axe 1 par conséquent ces variables sont donc bien représentées sur cet axe.

- Commentaires pour l'axe 2 :

D'après les données et selon les résultats obtenus, nous voyons que seulement les variables Engine.Size et Cylinders prennent des coordonnées positives tout juste supérieures à 0,5. Nous pouvons dire que ces variables sont tout de même corrélées à l'axe 2 mais pas fortement. Malgré cela, nous voyons que celles-ci contribuent le plus à la construction de cet axe. Les autres variables sont corrélées à l'axe 2 également mais très faiblement.

Nous pouvons aussi voir sur le graphique NLB ci-dessus que les variables Engine.size et Cylinders ont des cosinus carrés très faibles sur l'axe 2 par conséquent ces variables ne sont pas bien représentées sur cet axe. Par la suite, nous avons essayé de regarder la qualité de représentation de ces variables sur le plan et il s'avère que celles-ci sont bien représentées sur le plan factorielle (voir *figure MGS* juste en bas).

Etant donné que ces variables sont les mieux représentées sur l'axe 2 et qu'elles sont bien représentées sur le plan nous avons décidé de les retenir dans le cadre de l'analyse de cet axe.



Graphique NLB

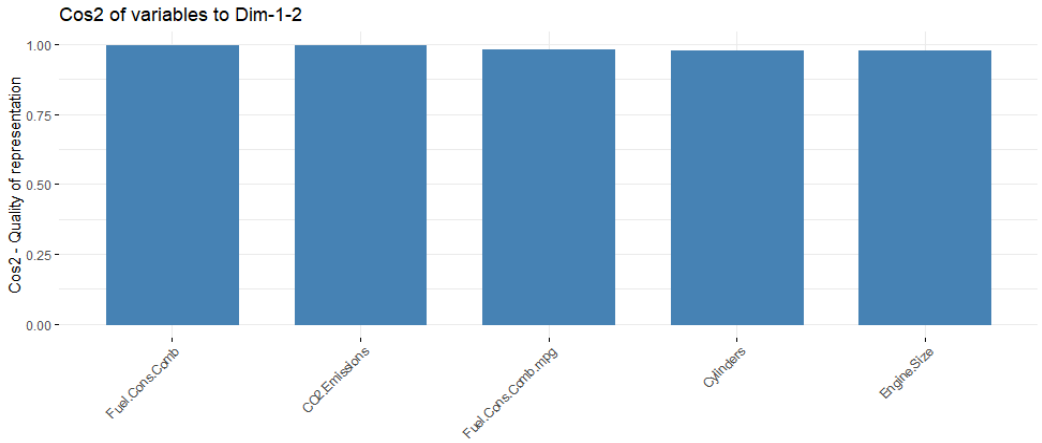


Figure MGS

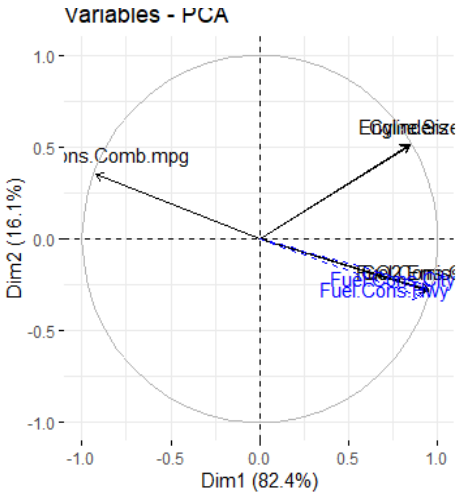
e) Variables supplémentaires

Comme mentionnée dans le script, notre base de données contient des *variables continues supplémentaires* (quanti.sup, colonnes 6 et 7) et des *variables qualitatives supplémentaires* (quali.sup, colonne 8 à 10).

Les variables supplémentaires ne sont pas utilisées pour la détermination des composantes principales. Leurs coordonnées sont prédites en utilisant uniquement les informations fournies par l'analyse en composantes principales effectuée sur les variables actives.

- **Variables continues supplémentaire**

Les variables continues supplémentaires que nous avons dans notre base de données sont : *Fuel.Cons.City* et *Fuel.Cons.Hwy.*



Sur le cercle de corrélation ci-dessus, les variables continues supplémentaires de notre base sont en pointillées et de couleur bleue. D'après les résultats obtenus, nous voyons que ces deux variables ont des fortes coordonnées sur l'axe 1, par conséquent elles sont fortement

corrélées à celui-ci. Ces variables sont également bien représentées sur l'axe 1 du fait de leurs cosinus carré proche de 1 sur cet axe.

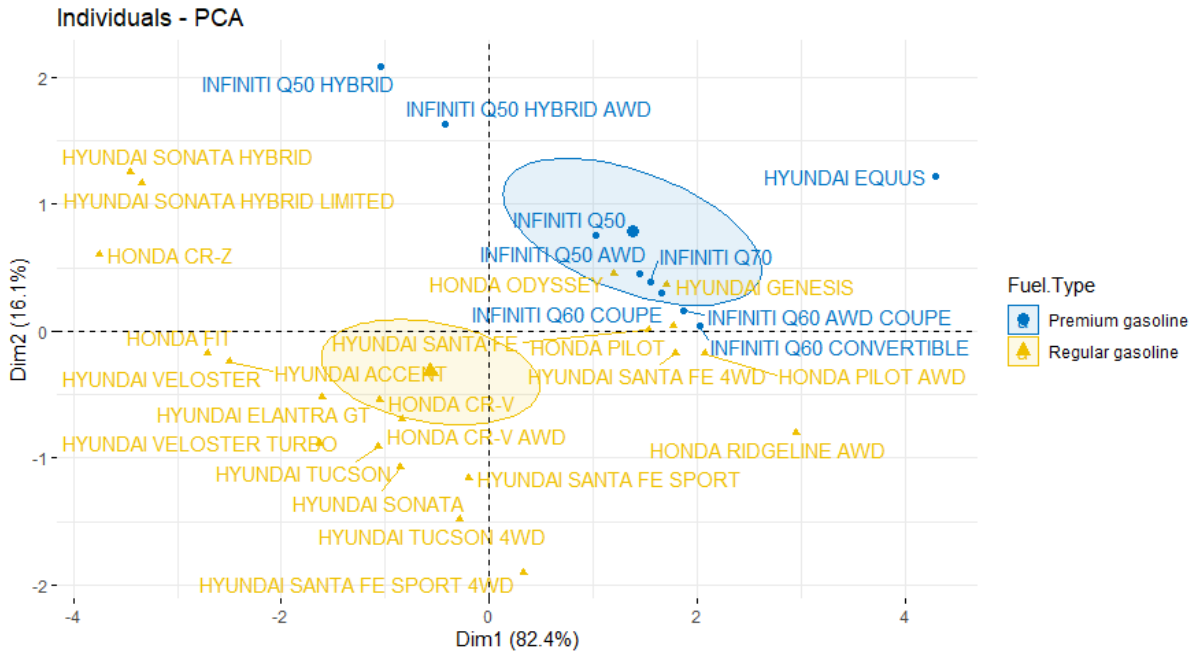
Sur l'axe 2, les variables *Fuel.Cons.City* et *Fuel.Cons.Hwy* ont des coordonnées négatives peu significatif par conséquent ces variables ne sont pas fortement corrélées à l'axe. Le cosinus carré de ces variables sur cet axe n'étant pas assez significatifs (proche de 1) fait en sorte qu'elles ne sont pas bien représentées sur l'axe 2.

- **Variables qualitatives supplémentaire**

Notre base de données contient trois variables qualitatives supplémentaires à savoir : *Fuel.type, Vehicule.Class* et *Transmission*. Ces variables seront utilisées pour caractériser ou illustrer les différents groupes d'individus existant dans notre base.

Notons que les individus que nous allons caractériser sont ceux que nous avons retenu lors de la description du nuage de point des individus.

Caractérisation des individus avec la variable qualitative sur le type de carburant (Fuel.Type)



Sur le graphique ci-dessus nous pouvons voir une opposition entre les véhicules atypiques que nous avions remarqué lors de la description du nuage des points des individus. Il s'agissait des véhicules HYUNDAI SONATA HYBRID, HYUNDAI SONATA HYBRID LIMITED, HONDA CR-Z, HYUNDAI EQUUS sur l'axe 1 et HYUNDAI TUCSON 4WD, HYUNDAI SANTA FE SPORT 4WD, INFINI Q50 HYBRID, INFINI Q50 HYBRID AWD sur l'axe 2.

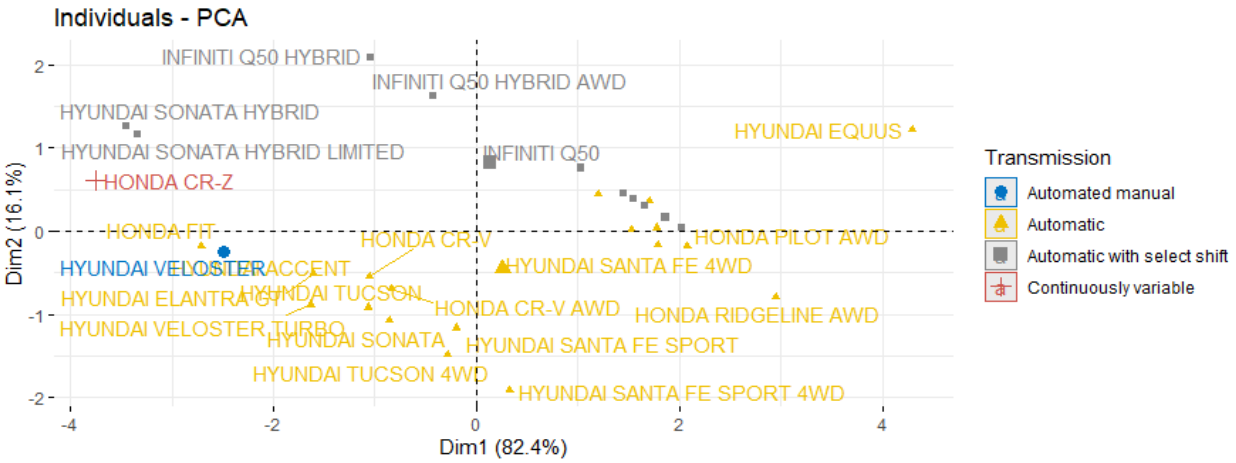
En effet nous avons comme information par rapport à ce graphe que les véhicules de marques HYNDAI SONATA HYBRID, HYNDAI SONATA HYBRID LIMITED et HONDA CR-Z consomment du « Regular gasoline » tandis que ceux de marque HYNDAI EQUUS consomment du « Premium gasoline ».

Les véhicules de marques HYUNDAI TUCSON 4WD et HYUNDAI SANTA FE SPORT 4WD consomment du « Regular gasoline » tandis que les véhicules de marque INFINI Q50 HYBRID et INFINI Q50 HYBRID AWD consomment du « Premium gasoline ».

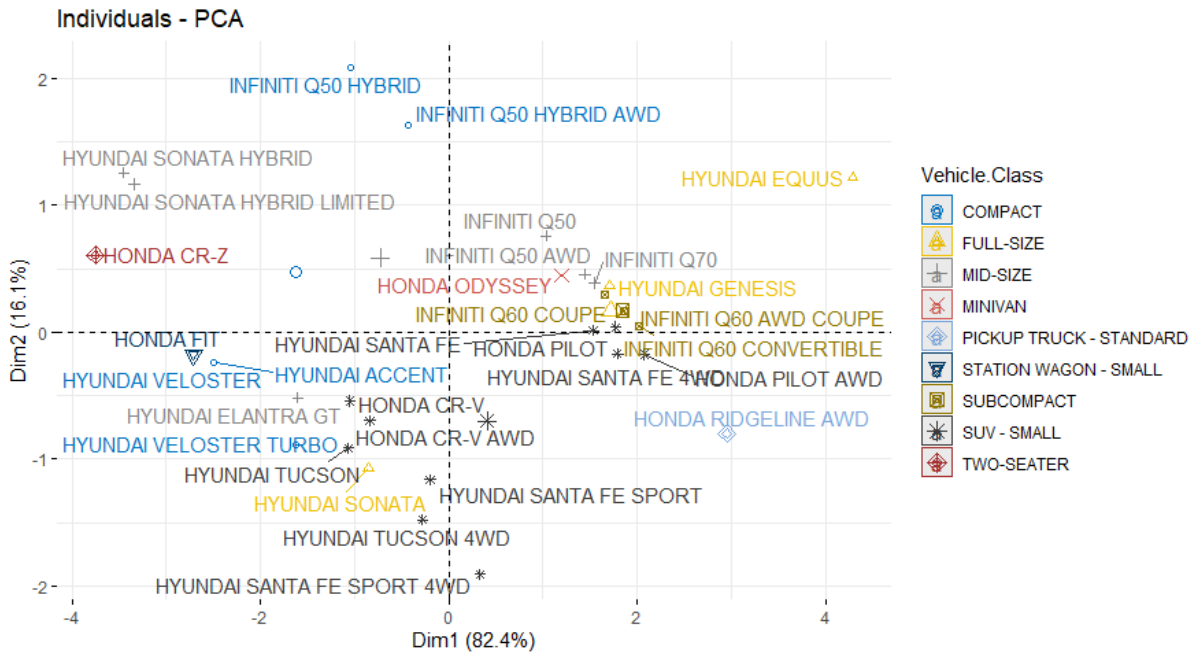
<u>Caractérisation des individus avec la variable qualitative sur le type de transmission du moteur</u> (Transmission) :

L'information que nous pouvons avoir sur le graphique ci-dessous est que les véhicules de marques HYNDAI SONATA HYBRID, HYNDAI SONATA HYBRID LIMITED ont comme type de transmission du moteur le type « Automatic with select shift », les vehicules de marque HONDA CR-Z ont comme type de transmission le type « continuously variable » et les voitures de marque HYUNDAI EQUUS ont comme type de transmision le type « Automatic ».

Les véhicules de marques HYUNDAI TUCSON 4WD et HYUNDAI SANTA FE SPORT 4WD ont comme type de transmission du moteur le type « Automatic » tandis que les véhicules de marque INFINI Q50 HYBRID et INFINI Q50 HYBRID AWD ont comme type de transmission le type « Automatic with select shift ».



<u>Caractérisation des individus avec la variable qualitative sur le type de véhicule</u> (Vehicule.Class) :

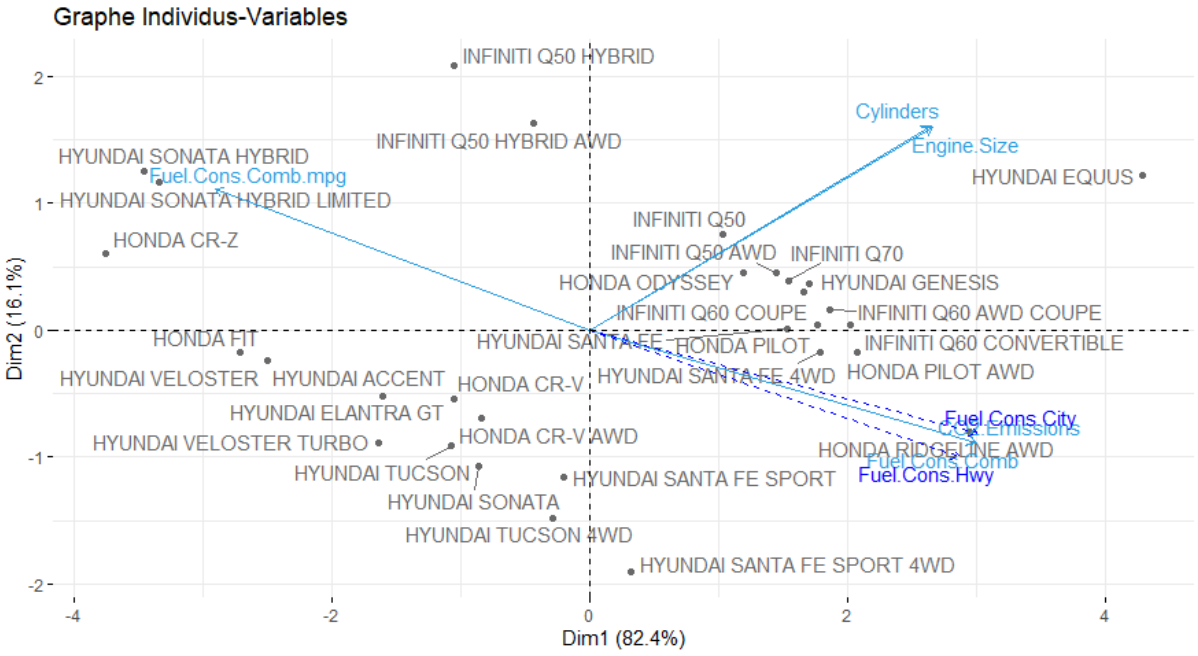


L'information que nous pouvons avoir sur le graphique ci-dessus est que les véhicules de marques HYNDAI SONATA HYBRID, HYNDAI SONATA HYBRID LIMITED sont des véhicules de classe « MID-SIZE » c'est-à-dire des véhicules de taille moyenne, les véhicules de marque HONDA CR-Z sont de type « TWO-SEATER » c'est-à-dire des véhicules deux places et les voitures de marque HYUNDAI EQUUS sont de type « FULL-SIZE » c'est-à-dire des véhicules de grande taille.

Les véhicules de marques HYUNDAI TUCSON 4WD et HYUNDAI SANTA FE SPORT 4WD sont de type « SUV-SMALL » tandis que les véhicules de marque INFINI Q50 HYBRID et INFINI Q50 HYBRID AWD sont de type « COMPACT » c'est-à-dire les voitures familiales de taille moyenne.

III-/ Interprétations des résultats

a-) Description du nuage des points individus-variables



Sur ce graphe nous pouvons voir que la variable Fuel.Cons.Comb.mpg les individus HYUNDAI SONATA HYBRID, HYUNDAI SONATA HYBRID et HONDA CR-Z sont du coté positif de l'axe 1 tandis que les variables CO2.Emissions, Fuel.Cons.Comb et l'individu HYUNDAI EQUUS sont du coté positif de l'axe.

Nous voyons aussi que les variables Cylinders et Engize.Size sont du coté positif de l'axe 2 avec les individus INFINI Q50 HYBRID ET INFINI Q50 HYBRID AWD.

b) - Résumé des informations obtenues et Identification des groupes d'individus à l'aide des variables

- Résumé des informations obtenues

Axe 1 :

-	+
Variables	Variables
Fuel.Cons.Comb.mpg	CO2.Emissions Fuel.cons.comb
Individus	Individus
Hyundai Sonata Hybrid Hyundai Sonata Hybrid limited Honda CR-Z	Hyundai EQUUS

Axe 2 :

-	+
Variables	Variables
	Engine.Size Cylinders
Individus	Individus
Hyundai Santa Fe Sport 4WD Hyundai Tucson 4WD	Infini Q50 Hybrid Infini Q50 Hybrid AWD

- Identification des groupes d'individus à l'aide des variables

Nous pouvons dire que l'axe 1 oppose deux groupes de véhicules :

Les véhicules de marque HYUNDAI EQUUS qui émettent beaucoup de CO2, qui ont une grande consommation de carburant combiné et qui en revanche ont une faible consommation de carburant combiné en mpg aux véhicules de marque HYUNDAI SONATA HYBRYD, HYUNDAI SONATA HYBRYD LIMITED et HONDA CR-Z qui ont une grande consommation de carburant combiné en mpg et qui en revanche ont une faible consommation de carburant combiné et une faible émission de CO2.

Les voitures de marque HYUNDAI EQUUS consomment du « premium gasoline », sont de type de transmission « Automatic » et sont de voiture de grande taille. Elles sont donc opposées aux voitures de marque HYUNDAI SONATA HYBRYD, HYUNDAI SONATA HYBRYD LIMITED et HONDA CR-Z qui consomment du « Regular gasoline », qui ont comme type de transmission « Automatic with select shift » et qui sont de taille moyenne (excepté les voitures de marque HONDA CR-Z qui ont comme type de transmission du moteur « continuously variable » et qui sont des véhicules de classe « TWO-SEATER ».

Nous pouvons dire que l'axe 2 oppose deux groupes de véhicules :

Les véhicules de marque INFINI Q50 HYBRID et INFINI Q50 HYBRID AWD qui ont un bon nombre de cylindre et un grand moteur aux voitures de marque HYUNDAI TUCSON 4WD et HYUNDAI SANTA FE HYBREID 4WD qui ont un petit nombre de cylindre et un petit moteur.

Les véhicules de marque INFINI Q50 HYBRID et INFINI Q50 HYBRID AWD consomment du « premium gasoline », sont de type de transmission « Automatic with select shift » et sont des voitures compactes c'est-à-dire familiale mais de taille moyenne. Elles sont donc opposées aux voitures de marque HYUNDAI TUCSON 4WD et HYUNDAI SANTA FE HYBREID 4WD qui consomment du « regular gasoline », sont de type de transmission « Automatic » et sont des voitures de type « SUV-SMALL ».

Conclusion :

L'analyse en composantes principales (ACP) est un outil extrêmement puissant de synthèse de l'information, très utile lorsque l'on est en présence d'une somme importante de données quantitatives à traiter et interpréter. Le principe de cette analyse est de chercher le plan qui résume le mieux l'information contenue dans le tableau de dimension k. On obtient ainsi une représentation approchée du nuage dans un espace de faible dimension

Malheureusement cette analyse ne se limite qu'aux variables quantitatives. Heureusement, il existe d'autres méthodes factorielles permettant de remédier à cela, comme l'**Analyse des Correspondances Multiples** pour des variables qualitatives, ou l'**Analyse Factorielle des Données Mixtes.**