

Analyse des accidents de la route dans le département 37 : Une exploration statistique

MELLOT Rachel, DUMONT-ROTY Pierr, MARTIN Chloé

10 janvier 2025

Abstract

Ce document présente une analyse des données issues de l'Observatoire National Inter-ministériel de la Sécurité Routière (ONISR) concernant les accidents de la route dans le département 37. À travers une analyse des correspondances multiples (ACM) et un clustering, nous explorons les profils types des accidents afin de proposer des pistes pour la prévention routière.

Contents

1	Accidents de la route en Indre-et-Loire : enjeux et questions clés	2
2	Description des bases de données : méthodologie et contenu	2
2.1	Origine et contenu des données	2
2.2	Filtrage et nettoyage des données	3
3	Analyse des Correspondances Multiples : exploration des relations entre les facteurs d'accidents	4
3.1	Méthodologie de l'Analyse des Correspondances Multiples (ACM)	4
3.2	Analyse de la part d'inertie expliquée	5
3.3	Représentation des individus dans le plan factoriel	6
3.4	Relations entre Variables et Axes Factoriels	7
4	Clustering des accidents : des profils pour mieux comprendre les risques	10
4.1	Classification Hiérarchique consolidée : Méthodologie	10
4.2	Choix du nombre de clusters : Critères et Méthodes	11
4.3	Exploration visuelle des profils types d'accidents	11
4.4	Compréhension des Groupes : Interprétation des Clusters	13
5	Synthèse des résultats et recommandations en prévention routière	15
6	Annexe	16

1 Accidents de la route en Indre-et-Loire : enjeux et questions clés

Chaque année, les accidents de la route causent des pertes humaines et matérielles considérables, posant des enjeux majeurs dans différents secteurs, comme la santé ou les dépenses publiques. En effet, en France, on dénombre en 2022 environ 3 260 décès et plus de 57 000 blessés graves nécessitant une hospitalisation. Les coûts associés (soins médicaux, pertes de productivité, dommages matériels) dépassent 40 milliards d’euros par an, illustrant le poids économique pour la société.

Malgré des progrès notables en matière de sécurité routière, ces chiffres restent préoccupants, soulignant la nécessité d’identifier et de comprendre les facteurs à l’origine des accidents afin de proposer des solutions adaptées. Réduire ces accidents est une priorité qui dépasse les simples statistiques : il s’agit d’une démarche essentielle pour sauver des vies et alléger les souffrances des victimes et de leurs proches.

En mobilisant les données issues de l’Observatoire National Interministériel de la Sécurité Routière (ONISR), ce travail cherche à répondre à une question centrale : quels sont les principaux facteurs et configurations qui favorisent les accidents, et comment peut-on agir pour les prévenir ?

Notre étude porte spécifiquement sur les accidents survenus dans le département de l’Indre-et-Loire. Le département de l’Indre-et-Loire (37) constitue un terrain d’étude particulièrement pertinent. Situé en région Centre-Val de Loire, il se caractérise par une diversité de contextes routiers, entre zones urbaines denses (comme Tours) et territoires ruraux plus isolés. Cette dualité offre l’opportunité d’explorer une large gamme de configurations accidentogènes, allant des grands axes routiers aux petites routes départementales.

Cette analyse a pour ambition de répondre à des enjeux majeurs liés à la sécurité routière. Notamment, identifier des typologies (profils types) d’accidents dans le but de proposer des recommandations pour des futures campagnes de prévention. En combinant une analyse des correspondances multiples (ACM) et une classification hiérarchique, nous espérons offrir une vision claire des problématiques locales et contribuer à l’amélioration des dispositifs de prévention dans le département de l’Indre-et-Loire.

2 Description des bases de données : méthodologie et contenu

2.1 Origine et contenu des données

Pour cette étude nous avons choisi de travailler sur des bases de données fournies par l’Observatoire national interministériel de la sécurité routière (ONISR) répertoriant des accidents corporels de la circulation routière.

Les données ont été collectées par les forces de l’ordre qui sont intervenues sur le lieu de l’accident. Elles concernent l’ensemble des accidents survenus en France métropolitaine et dans les territoires d’Outre-mer, sur une année précise, ici l’année 2023.

Les informations enregistrées sont réparties en quatre bases de données distinctes, couvrant des aspects variés des accidents :

- Rubrique Caractéristiques : contenant les informations correspondants aux circonstances générales de l’accident (date, heure, type de collision, etc.).

- Rubrique Lieux : référençant la localisation précise de l'accident (type de route, conditions météorologiques, intersection, etc.).
- Rubrique Véhicules : renseignant les données relatives aux véhicules impliqués (type de véhicule, manœuvre effectuée, etc.).
- Rubrique Usagers : décrivant les caractéristiques des personnes impliquées (âge, sexe, gravité des blessures, etc.).

Ces bases sont reliées entre elles grâce à un identifiant unique, le numéro de l'accident (Num_Acc), permettant de combiner les informations pour chaque événement. Afin de faciliter l'analyse statistique, nous avons fusionné ces quatre bases en une seule base de données consolidée, en nous appuyant sur cet identifiant.

Les tableaux suivants illustrent un échantillon des données mobilisées dans cette étude, offrant un aperçu des variables extraites de nos différentes bases de données.

Table 1: Extrait des données de la rubrique 'Caractéristiques'

Num_Acc	jour	mois	an	hrmn	lum	dep	com	agg	atm	col	adr
2.023e+11	7	5	2023	06:00	1	75	75101	2	2	7	RUE DE RIVOLI
2.023e+11	7	5	2023	05:30	5	94	94080	2	3	6	Avenue de Paris

Table 2: Extrait des données de la rubrique 'Lieux'

Num_Acc	catr	voie	vosp	plan	larout	surf	infra	situ	vma
2.023e+11	4	RUE DE RIVOLI	0	1	-1	2	0	1	30
2.023e+11	4	RUE SAINT FLORENTIN	0	1	-1	2	0	1	30

Table 3: Extrait des données de la rubrique 'Usagers'

Num_Acc	id_usager			id_vehicule			num_veh	place	catu	grav	sexe	an_nais	trajet
2.023e+11	203	851	184	155	680	557	A01	1	1	4	1	1978	5
2.023e+11	203	851	182	155	680	556	A01	1	1	1	2	1997	9

Table 4: Extrait des données de la rubrique 'Véhicules'

Num_Acc	id_vehicule			num_veh	senc	catv	obs	obsm	choc	manv	motor
2.023e+11	155	680	557	A01	1	30	0	0	5	1	1
2.023e+11	155	680	556	A01	2	7	0	1	1	1	1

2.2 Filtrage et nettoyage des données

Pour cette étude, nous avons limité l'analyse aux accidents survenus dans le département de l'Indre-et-Loire (37), ce qui a impliqué un filtrage géographique des données.

Par ailleurs, un nettoyage des données a été effectué afin d'assurer leur qualité et leur cohérence. Les valeurs manquantes (NA) ou non renseignées ont été exclues, car elles ne permettaient pas une interprétation fiable. De plus, certaines modalités imprécises ou redondantes ont été regroupées ou transformées pour équilibrer leur nombre dans chaque variable et ainsi améliorer la robustesse des résultats lors de l'Analyse des Correspondances Multiples (ACM) et de la classification hiérarchique. Par exemple, les types de routes ont été regroupés en catégories plus

larges, et les conditions météorologiques ont été réorganisées en fonction de leur impact sur la visibilité et l’adhérence.

Étant donné le grand nombre de variables (68 au total) contenues dans la base fusionnée, nous présentons ici un extrait limité aux variables les plus pertinentes pour notre analyse :

Table 5: Extrait des données de notre base d’étude

Num_Acc	catv_grouped	manv_grouped	surf_grouped	lum_grouped	atm_grouped
2.023e+11	Poids lourds	Trajectoire rectiligne	Mouillée	Jour	atm-Normal
2.023e+11	Poids lourds	Trajectoire rectiligne	Mouillée	Jour	atm-Normal
2.023e+11	Voitures	Changement de file	Mouillée	Jour	atm-Normal

3 Analyse des Correspondances Multiples : exploration des relations entre les facteurs d’accidents

3.1 Méthodologie de l’Analyse des Correspondances Multiples (ACM)

L’Analyse des Correspondances Multiples (ACM) est une méthode d’analyse exploratoire utilisée pour étudier les relations entre des variables qualitatives. Cette méthode réduit la dimensionnalité des données tout en conservant l’essentiel de l’information, facilitant ainsi l’interprétation des liens entre les variables et la visualisation des résultats.

Dans notre étude, nous avons inclus plusieurs groupes de variables dans l’ACM, distinguées selon leur nature et leur rôle dans l’analyse.

Les variables actives qualitatives choisies pour cette étude sont :

- “agg” : Caractère urbain ou rural de la zone.
- “catv” : Catégorie des véhicules impliqués.
- “manv” : Manœuvres des véhicules.
- “surf” : État de la surface de la route.
- “lum” : Conditions de luminosité.
- “atm” : Conditions météorologiques.
- “catr” : Catégorie de la route.
- “situ” : Situation de l’accident sur la voie.

Les variables qualitatives supplémentaires comprennent :

- “mois” : Mois de l’accident.
- “catu” : Catégorie des usagers impliqués (conducteurs, passagers, piétons).
- “sexe” : Sexe des individus impliqués.
- “trajet” : Nature du trajet (domicile-travail, loisirs, etc.).
- “grav” : Gravité de l’accident.
- “vosp” : Présence d’une voie réservée aux vélos.

Les variables quantitatives supplémentaires incluent :

- “an_nais” : Année de naissance des usagers.
- “vma” : Vitesse maximale autorisée sur le lieu de l’accident.

En combinant ces différentes variables, l'ACM permet d'explorer les interactions complexes entre les conditions, les acteurs et les circonstances des accidents de la route dans l'Indre-et-Loire, tout en offrant une représentation synthétique et informative de ces données.

3.2 Analyse de la part d'inertie expliquée

Afin de déterminer quels axes factoriels seront retenus pour l'interprétation, nous avons examiné la part d'inertie expliquée par chaque axe, c'est-à-dire la proportion de variance totale capturée.

Le graphique et le tableau ci-dessous, illustrant la part d'inertie des axes, nous guideront dans le choix du nombre d'axes à retenir pour notre analyse.

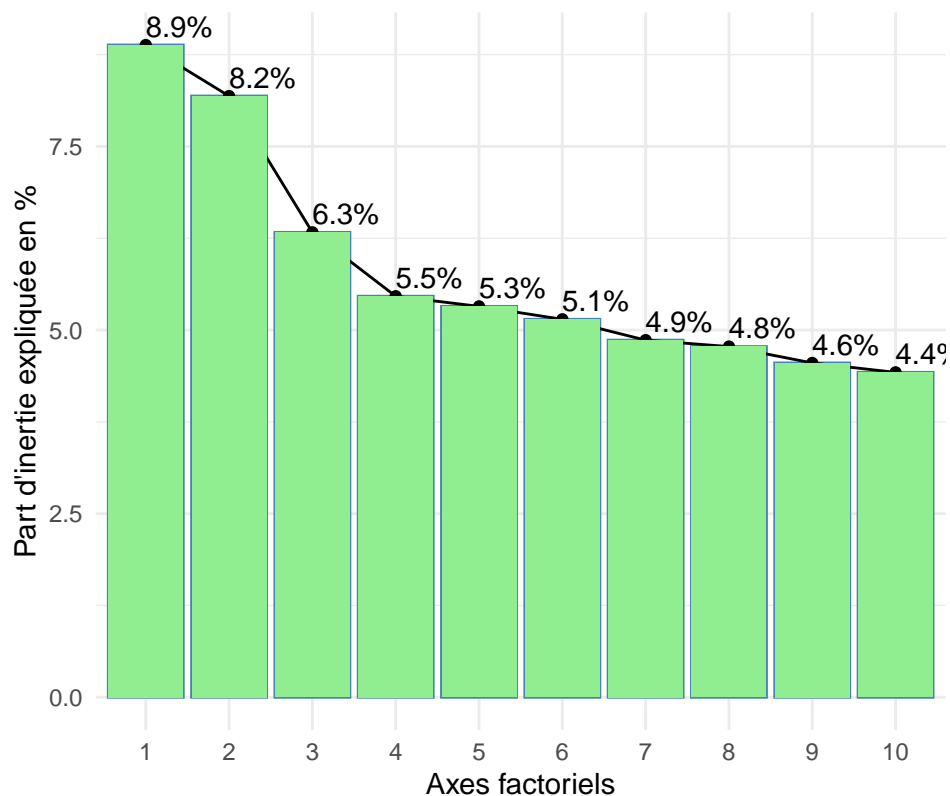


Figure 1: Graphique des Valeurs Propres

Table 6: Tableau des inerties

	dim 1	dim 2	dim 3	dim 4	dim 5	dim 6	dim 7	dim 8	dim 9	dim 10
Valeurs Propre	0.3	0.2	0.2	0.2	0.2	0.1	0.1	0.1	0.1	0.1
Parts d'inertie %	8.9	8.2	6.3	5.5	5.3	5.1	4.9	4.8	4.6	4.4
Parts cumulées	8.9	17.1	23.4	28.9	34.2	39.3	44.2	49.0	53.5	57.9

L'analyse de la part d'inertie montre que les deux premiers axes expliquent ensemble 17.1% de la variance totale :

- L'axe 1 explique 8.9 % de la variance.
- L'axe 2 explique 8.2 % de la variance.

Ces deux axes, bien qu'individuellement modestes, concentrent une proportion significative de l'information globale et sont donc retenus pour l'analyse.

Le troisième axe explique 6.3 % de la variance. Bien que moins important, il sera occasionnellement pris en compte pour approfondir certains aspects de l'interprétation.

En revanche, les axes suivants expliquent chacun une part d'inertie oscillant entre 5 % et 4 %, sans qu'une chute marquée ne soit observée dans le graphique. Ces axes ne seront pas retenus dans notre étude principale, car leur contribution est relativement faible et ils n'apporteraient qu'une complexité supplémentaire à l'analyse.

Ainsi, nous concentrons notre étude principalement sur les deux premiers axes, tout en considérant ponctuellement l'axe 3 pour enrichir l'interprétation.

3.3 Représentation des individus dans le plan factoriel

Le nuage des individus est une visualisation clé de l'analyse en composantes multiples (ACM), où chaque point représente un individu, c'est-à-dire un accident dans notre étude. Ce graphique permet de visualiser la distribution des 2718 accidents selon les deux premiers axes factoriels retenus.

Pour évaluer la pertinence de la projection des individus sur ces axes, nous avons utilisé le cosinus carré (\cos^2), une mesure indiquant la qualité de la représentation des individus. Plus la valeur du \cos^2 est élevée, plus la projection d'un individu sur le plan factoriel est fidèle à sa position dans l'espace initial multidimensionnel. Les individus sont colorisés en fonction de leur \cos^2 , ce qui permet d'identifier visuellement les points bien représentés (en rouge ou orange) et ceux dont la qualité de représentation est plus faible (en bleu ou vert).

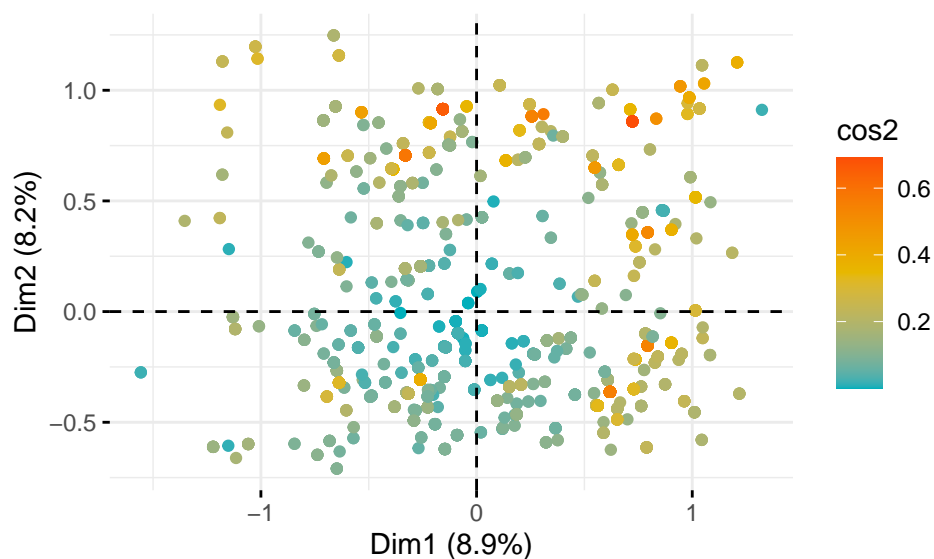


Figure 2: Nuage des individus

Le nuage des individus révèle une distribution dense et homogène des accidents, suggérant que ces derniers forment un seul groupe principal. Cette absence de séparation marquée entre sous-groupes d'accidents reflète un certain manque de différenciation dans la distribution des individus sur les deux premiers axes.

Il est important de noter que nos individus sont des numéros d'accidents, ce qui rend difficile une observation intuitive de leurs caractéristiques uniquement à partir de leurs identifiants numériques. Malgré cette limitation, nous pouvons pousser plus loin notre étude des individus.

Après étude des contributions des individus, il ressort qu'une majeure partie des individus participe significativement à la construction des axes. Cependant, nous pouvons remarquer

que certains individus contribuent plus fortement à la formation des axes. Ces individus vont jouer un rôle clé dans l'interprétation des dimensions, car ils capturent des profils spécifiques ou extrêmes dans les données.

3.4 Relations entre Variables et Axes Factoriels

L'Analyse des Correspondances Multiples (ACM) met en lumière les variables et modalités qui structurent nos données, révélant des dimensions pertinentes pour comprendre les profils d'accidents de la route. Les contributions des variables aux axes factoriels, tout comme leur visualisation dans les nuages de variables et de modalités, permettent d'identifier les éléments clés influençant la construction de chaque axe.

Le premier axe est principalement structuré par des variables liées au lieu de l'accident et aux infrastructures routières. Les modalités les plus contributives concernent la localisation (en agglomération ou hors agglomération), le type de route (autoroute, départementale, nationale, urbaine) et le type de véhicule impliqué, en particulier les engins personnels (vélos, trottinettes, fauteuils roulants). Par ailleurs, les accidents sur des voies cyclables se distinguent fortement. Ces observations soulignent l'importance de l'environnement infrastructurel dans la structuration de cette première dimension.

Le deuxième axe se concentre sur les variables environnementales et météorologiques. Les conditions atmosphériques (pluie, temps sec, etc.), l'état de la chaussée (mouillée ou sèche) et la luminosité (jour ou nuit) jouent un rôle déterminant. Cet axe traduit une opposition entre des conditions météorologiques difficiles, souvent rencontrées durant les mois d'hiver comme novembre ou décembre, et des conditions normales, plus fréquentes pendant les périodes plus clémentes. Ainsi, ce second axe reflète une dimension environnementale essentielle, associant les caractéristiques saisonnières et météorologiques aux profils d'accidents.

Quant au troisième axe, il met en lumière des situations plus spécifiques, particulièrement liées aux cyclistes et utilisateurs d'engins personnels. Les modalités les plus représentatives de cet axe incluent les accidents sur des voies cyclables, les manœuvres inhabituelles telles que traverser une chaussée, ainsi que les véhicules particuliers comme les engins personnels. Ce troisième axe révèle ainsi une dimension centrée sur les spécificités des accidents en milieu urbain impliquant des usagers vulnérables, notamment les cyclistes.

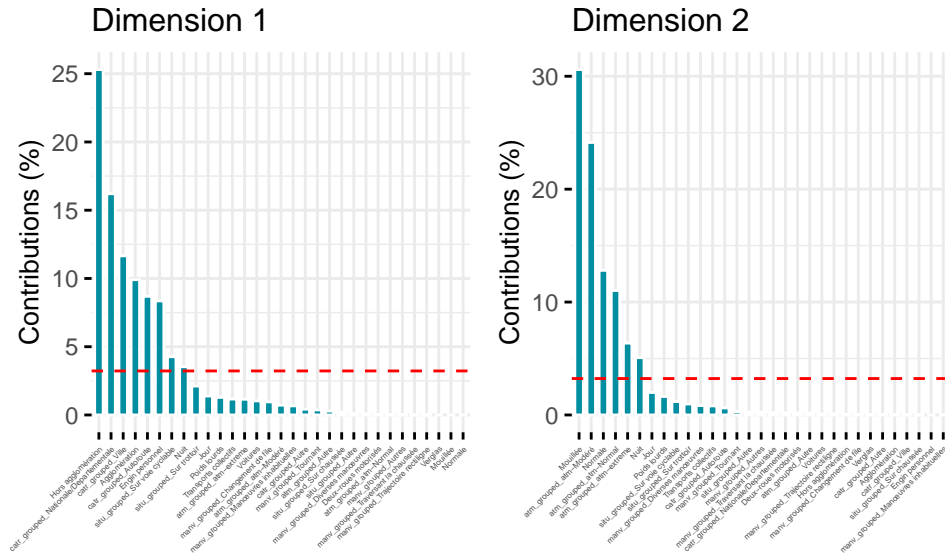


Figure 3: Histogramme des contributions des variables aux axes

Le nuage des variables offre une vue synthétique des relations entre les variables qualitatives et les axes factoriels. Les variables associées à l'infrastructure et aux véhicules se regroupent le long de l'axe 1, reflétant leur importance dans cette dimension. À l'inverse, les variables décrivant les conditions météorologiques et saisonnières se rassemblent autour de l'axe 2, confirmant leur rôle structurant dans ce contexte. Cette représentation facilite la compréhension globale des données et met en évidence les groupes de variables corrélées, renforçant les interprétations issues des contributions.

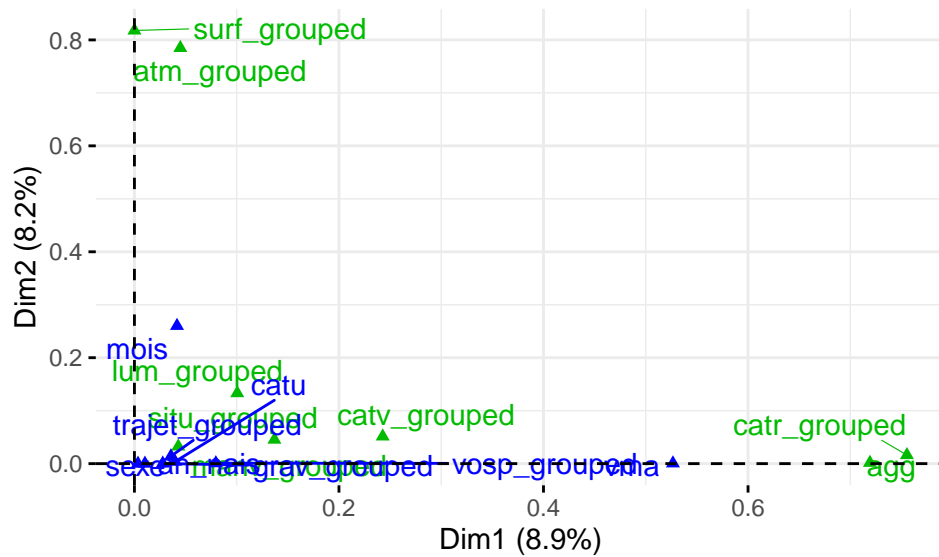


Figure 4: Nuage des variables

Sur l'axe 2, la partie supérieure regroupe des modalités associées à des conditions environnementales défavorables. Les accidents survenus dans des conditions météorologiques extrêmes, telles que des pluies fortes ou des tempêtes, sont particulièrement représentés, de même que ceux ayant eu lieu sur des sols mouillés. Ces circonstances sont souvent accompagnées de périodes hivernales, notamment en décembre, janvier et novembre, et de situations survenant principalement la nuit, accentuant ainsi le caractère risqué des conditions dans lesquelles ces accidents se produisent. À l'inverse, la partie inférieure de l'axe est dominée par des modalités correspondant à des conditions météorologiques plus favorables. Les accidents survenus sur un sol sec ou dans un état normal y sont largement représentés, tout comme ceux qui ont eu lieu durant les mois d'été, notamment en juin, juillet et septembre. Ces événements se produisent principalement en journée, sous des conditions lumineuses classiques, contribuant à des situations moins risquées sur le plan environnemental.

9

En somme, l'ACM dévoile trois dimensions principales : une première axée sur les infrastructures et les véhicules, une seconde liée aux conditions environnementales et météorologiques, et une troisième reflétant les contextes spécifiques aux cyclistes et utilisateurs d'engins personnels.

Les variables quantitatives supplémentaires, bien qu'elles n'interviennent pas directement dans la construction des axes factoriels, offrent des informations complémentaires pour affiner l'interprétation des dimensions principales. La vitesse maximale autorisée (VMA) est corrélée positivement à l'axe 1 de façon importante, reliant les zones à haute vitesse (routes hors agglomération) aux modalités à droite de l'axe. L'année de naissance, quant à elle, est négativement corrélée à cet axe, montrant que les jeunes conducteurs se situent à gauche, souvent impliqués dans des accidents urbains. Tandis que les conducteurs plus âgés se trouvent à droite. Ces variables confirment les différences entre accidents urbains et hors agglomération.

4 Clustering des accidents : des profils pour mieux comprendre les risques

4.1 Classification Hiérarchique consolidée : Méthodologie

Dans le prolongement de l'Analyse des Correspondances Multiples (ACM) réalisée précédemment, une classification hiérarchique consolidée a été effectuée afin d'identifier des groupes homogènes d'accidents de la route. Cette méthode permet de regrouper les individus en fonction de leurs similitudes dans l'espace factoriel défini par l'ACM, maximisant ainsi la cohérence interne des clusters tout en différenciant au mieux les groupes entre eux. La classification hiérarchique offre un cadre méthodologique rigoureux pour explorer les structures latentes dans les données, en tenant compte des dimensions principales révélées par l'ACM. En consolidant cette classification par un reclassement des individus, nous renforçons la stabilité et la pertinence des résultats. Cette étape constitue une avancée cruciale dans l'interprétation des données, permettant de mettre en lumière des profils types d'accidents et d'enrichir la compréhension des facteurs sous-jacents.

Le dendrogramme ci-dessous illustre les résultats de notre classification hiérarchique. Ce graphique représente le processus de regroupement progressif des individus selon le critère appliqué, ici le critère de Ward. Cette méthode cherche à maximiser la variance inter tout en minimisant la variance intra. Les branches du dendrogramme indiquent les distances entre les groupes, permettant d'identifier des seuils pertinents pour la formation des clusters.

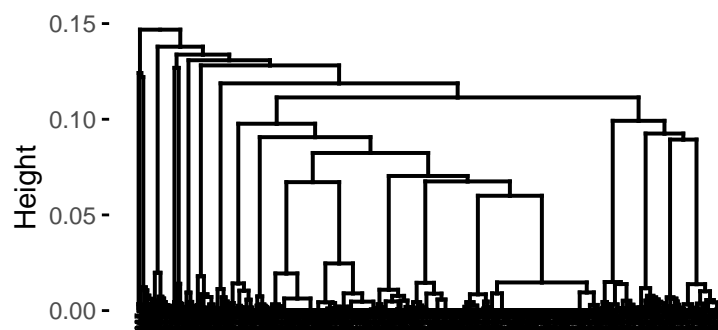


Figure 6: Dendrogramme : Clustering with Ward method

L'analyse du dendrogramme ne nous permet pas de déterminer visuellement le découpage optimal. Ainsi une analyse supplémentaire des résultats de notre classification hiérarchique est nécessaire pour permettre une interprétation approfondie des données.

4.2 Choix du nombre de clusters : Critères et Méthodes

Pour déterminer le nombre optimal de clusters, nous nous sommes appuyés sur deux graphiques complémentaires présentés ci-dessous. Le premier illustre l'évolution de l'inertie intra-cluster en fonction du nombre de clusters, tandis que le second met en évidence le gain d'inertie inter-cluster. Ces deux indicateurs statistiques sont des outils fiables pour guider notre choix.

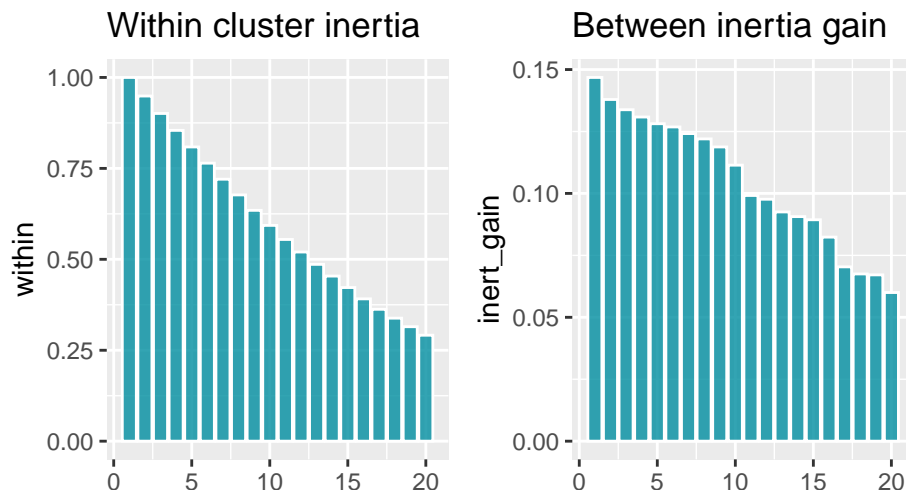


Figure 7: Représentation de l'inertie inter et intra cluster

Le graphique de l'inertie intra-cluster montre une diminution progressive à mesure que le nombre de clusters augmente, ce qui est attendu : un nombre plus élevé de clusters permet une meilleure homogénéité au sein de chaque groupe.

Quant au graphique du gain d'inertie inter-cluster, il révèle une marche significative entre 10 et 11 clusters. Ce seuil marque une rupture claire, justifiant le choix de 10 clusters comme solution optimale.

Il convient de noter que l'option d'un 11 cluster aurait pu être envisagée. Toutefois, une analyse plus approfondie a montré qu'un cluster supplémentaire n'apportait pas de nouvelles informations interprétables et rendait l'analyse plus complexe. Ainsi, le choix de 10 clusters représente un équilibre entre la qualité statistique et la simplicité interprétative.

4.3 Exploration visuelle des profils types d'accidents

Après avoir déterminé le nombre optimal de clusters, il est essentiel de les visualiser afin de mieux comprendre leur formation et leur structure. Pour ce faire, nous avons utilisé deux représentations graphiques complémentaires :

- Le dendrogramme colorisé, qui illustre de manière hiérarchique les relations entre les individus et met en avant les clusters définis.
- Le nuage des individus colorisé, enrichi par des ellipses représentant chaque cluster. Ce graphique situe les groupes dans l'espace factoriel défini par l'ACM, offrant une perspective globale sur leur positionnement et leur proximité relative.

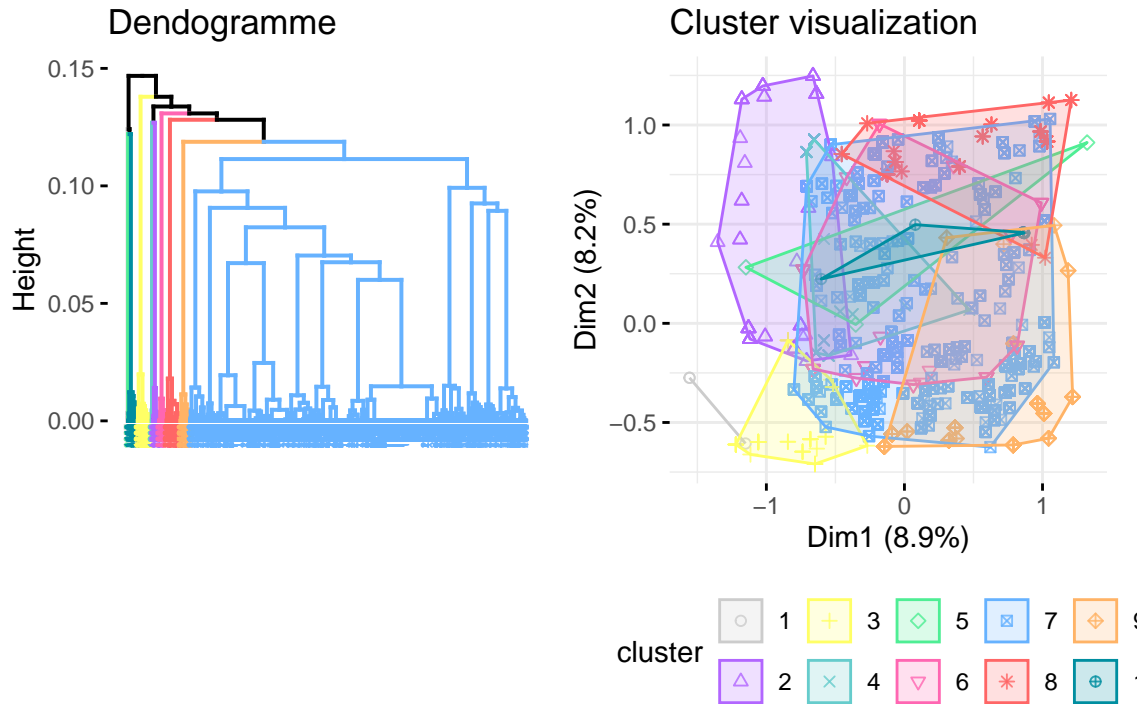


Figure 8: Visualisation des clusters

Cependant, la visualisation de ces groupes n'est pas toujours évidente sur ces graphiques. Cela s'explique notamment par la disparité des tailles des clusters : certains groupes sont très petits, regroupant à peine 3 individus, tandis que d'autres sont beaucoup plus grands, comme le cluster 7, qui rassemble la quasi-totalité des individus (2287 individus).

Ainsi, les clusters de plus petite taille se distinguent par des modalités très spécifiques, ce qui les rend particulièrement intéressants d'un point de vue interprétatif. En revanche, le cluster 7 représente un groupe plus généraliste, regroupant des individus partageant des caractéristiques moins distinctives. Ces observations soulignent la diversité et le niveau de détail des groupes identifiés.

Afin de compléter l'analyse de la visualisation des clusters, le tableau ci-dessous présente la répartition des individus au sein de chaque cluster. Cette représentation permet de mieux appréhender l'importance relative de chaque groupe, en mettant en évidence la variabilité de leur taille, allant de clusters très restreints à des regroupements beaucoup plus larges.

Table 7: Nombre d'individus par cluster

Cluster	1	2	3	4	5	6	7	8	9	10
Nombre d'individus	3	64	43	50	3	67	2287	95	96	10

Grâce à ce test, il est possible d'identifier les variables catégorielles les plus discriminantes pour la segmentation des clusters. Une p-valeur faible (inférieure au seuil conventionnel de 0,05) indique que la variable est significativement associée aux clusters, ce qui signifie qu'elle joue un rôle déterminant dans la différenciation des groupes.

Parmi les variables analysées, celles liées à la catégorie du véhicule, la manœuvre effectuée, l'état de la surface de la route, les conditions atmosphériques, et la catégorie de route se sont révélées

particulièrement pertinentes pour expliquer et segmenter les clusters. Ces résultats confirment l'importance de ces dimensions dans la structuration des profils d'accidents.

4.4 Compréhension des Groupes : Interprétation des Clusters

4.4.1 cluster 1

Le cluster 1 se distingue par des accidents impliquant uniquement des engins personnels (vélos, trottinettes, skateboards, etc.), sans la présence d'autres types de véhicules. Tous ces accidents ont eu lieu à proximité d'une voie cyclable ou directement sur celle-ci. En comparaison avec l'ensemble des accidents impliquant des engins personnels (9.3%), ce cluster ne représente que 1.2% des accidents où un engin personnel est en cause.

Ce cluster semble donc correspondre à des accidents principalement liés aux vélos sur des voies cyclables urbaines, dans des conditions météorologiques normales.

4.4.2 cluster 2

Le cluster 2 partage des caractéristiques similaires avec le cluster 1, notamment en termes de localisation des accidents en milieu urbain, sur des pistes cyclables. Cependant, à la différence du cluster 1, ce groupe inclut plusieurs types de véhicules. En effet, 58% des accidents impliquent des engins personnels, tandis que 39% concernent des voitures.

Ce cluster reflète des accidents sur des pistes cyclables en ville, mais avec une plus grande diversité de véhicules impliqués.

4.4.3 cluster 3

Le cluster 3 regroupe des accidents survenus en ville, spécifiquement sur des trottoirs. En effet, tous les individus de ce groupe ont été impliqués dans des accidents sur trottoir, une modalité relativement rare, puisque seulement 1.6% des accidents dans l'ensemble de la base de données concernent cette situation. Cette rareté a probablement contribué à la formation distincte de ce cluster. De plus, 88.4% des accidents de ce cluster ont eu lieu en plein jour, ce qui est un autre facteur marquant de cette catégorie. En termes de gravité, 70% des accidents ont entraîné des blessés, ce qui est cohérent avec le fait que ce cluster implique principalement des piétons, des usagers particulièrement vulnérables.

Le cluster 3 représente des accidents urbains sur des trottoirs, souvent survenus en journée, et impliquant principalement des piétons. La haute proportion d'accidents avec blessés souligne la vulnérabilité de ces usagers.

4.4.4 cluster 4

Le cluster 4 regroupe tous les accidents impliquant des transports en commun, soit 50 accidents en 2023. Ces accidents se produisent presque exclusivement en milieu urbain, sur la chaussée. En ce qui concerne les autres modalités associées à ce cluster, leur interprétation est moins évidente, car elles sont relativement partagées. Par exemple, 50% des accidents ont eu lieu dans des conditions atmosphériques normales (avec un sol sec), tandis que l'autre moitié s'est produite dans des conditions météorologiques modérées (avec un sol mouillé). Ainsi, la modalité la plus discriminante de ce cluster reste le type de véhicule, à savoir les transports en commun.

Le cluster 4 représente principalement des accidents en milieu urbain impliquant des transports en commun, la variable "type de véhicule" étant le critère le plus déterminant dans la constitution de ce groupe. Les autres facteurs, bien que présents, ne semblent pas aussi discriminants.

4.4.5 cluster 5

Le cluster 5 est un groupe très spécifique et relativement petit, composé de seulement 3 individus. Ce cluster se distingue par un facteur rare : la présence de verglas sur la route le jour de l'accident. Cette modalité est peu fréquente dans l'ensemble des données, ce qui est cohérent avec le contexte local, puisque le département de l'Indre-et-Loire n'est pas particulièrement connu pour ses périodes de gel sévères.

Le cluster 5 représente des accidents isolés, survenus dans des conditions météorologiques extrêmes liées au verglas, un phénomène rare mais potentiellement très dangereux.

4.4.6 cluster 6

Le cluster 6 semble principalement être défini par la modalité "Autre" de la variable "type de manœuvre effectuée", qui est relativement rare. Ce groupe présente des accidents survenus sur la chaussée, impliquant principalement des voitures, mais aussi dans une moindre mesure des vélos. Une caractéristique notable de ce cluster est que près de deux tiers des accidents ont causé des blessures, tandis qu'un tiers des accidents a concerné des individus indemnes.

4.4.7 cluster 7

Le cluster 7 est le cluster le plus peuplé, représentant presque la totalité des accidents. Il est principalement constitué d'accidents impliquant des voitures (81%), ce qui reflète bien la proportion générale observée dans la base de données. Ces accidents se produisent majoritairement sur la chaussée. Les facteurs environnementaux, tels que les conditions météorologiques, ne semblent pas jouer un rôle déterminant dans ce cluster, la majorité des accidents ayant eu lieu dans des conditions atmosphériques dites "normales". En conclusion, le cluster 7 représente de manière générale les accidents de notre base de données, agissant ainsi comme un cluster "moyen" ou de référence.

4.4.8 cluster 8

Le cluster 8 rassemble les accidents survenus dans des conditions météorologiques extrêmes, telles que des tempêtes, de fortes pluies ou des vents violents, accompagnés d'une surface de route mouillée. Il ressort que la majorité de ces accidents ont eu lieu pendant les mois de mai et novembre, des périodes caractérisées par des variations climatiques importantes et des intempéries imprévisibles. De plus, il est à noter que les accidents du cluster 8 se sont principalement produits sur des trajectoires rectilignes, ce qui suggère que la manœuvre du véhicule n'a pas été un facteur déterminant dans la survenue de l'accident.

4.4.9 cluster 9

Le cluster 9 regroupe les accidents impliquant des poids lourds. Une caractéristique notable de ce groupe est la prédominance des trajets effectués dans un cadre professionnel, ce qui est en accord avec le type de véhicule concerné. En effet, 85% des accidents surviennent pendant la journée, ce qui peut être expliqué par les restrictions horaires concernant les chauffeurs routiers.

4.4.10 cluster 10

Le cluster 10 est difficile à interpréter en raison de sa petite taille (seulement 10 individus). Il regroupe les accidents pour lesquels les forces de l'ordre ont codifié la condition atmosphérique comme « autre », une modalité relativement rare dans notre échantillon. Cette rareté semble avoir joué un rôle dans la formation de ce cluster. De plus, on peut supposer que la modalité « sol mouillé » associée à ces accidents indique que la condition « autre » pourrait correspondre à

des intempéries spécifiques, probablement des conditions météorologiques particulières qui ont été mal renseigné lors du rapport des forces de l'ordre.

5 Synthèse des résultats et recommandations en prévention routière

L'analyse des accidents de la route en Indre-et-Loire, à travers l'Analyse des Correspondances Multiples (ACM) et la Classification Hiérarchique (clustering), a permis de dégager des enseignements précieux sur les facteurs structurant ces accidents. L'ACM a révélé que les accidents peuvent être principalement expliqués par deux grands axes factoriels : les conditions environnementales et le lieu de l'accident. Le premier axe met en évidence la distinction entre accidents survenant en milieu urbain ou hors agglomération, et les types de véhicules impliqués (engins personnels, véhicules légers, poids lourds, etc.), tandis que le second axe met en lumière l'importance des facteurs météorologiques et de l'état de la surface de la route (par exemple, verglas ou sol mouillé).

Le processus de clustering a ensuite permis d'identifier des groupes homogènes d'accidents, révélant des profils spécifiques et des typologies d'incidents ayant des caractéristiques communes. Ces groupes, issus de la classification hiérarchique, se distinguent notamment par des critères tels que la présence de vélos et trottinettes sur les voies cyclables, les accidents impliquant des poids lourds, ou encore les incidents en conditions météorologiques extrêmes. En combinant ces deux méthodes, nous avons pu cerner les principaux facteurs contributifs aux accidents dans la région.

Cependant, cette étude présente certaines limites. Tout d'abord, elle se limite au département d'Indre-et-Loire (37), ce qui restreint la généralisation des résultats à d'autres régions, la France étant un pays avec une grande diversité de climat. De plus, certaines variables potentielles, telles que la consommation d'alcool, l'utilisation du téléphone au volant, n'ont pas été intégrées dans cette analyse, ce qui pourrait constituer un axe d'amélioration pour des études futures.

Enfin, à partir de ces résultats, plusieurs recommandations pour la sécurité routière peuvent être proposées. D'une part, des campagnes de prévention ciblées sur les zones identifiées comme à risque pourraient être mises en place. Par exemple, au sein des zones urbaines avec une forte concentration de personnes vulnérables, une amélioration des infrastructures cyclables pourrait contribuer à réduire le nombre d'accidents. D'autre part, des initiatives visant à améliorer la circulation sous conditions météorologiques extrêmes, notamment lors de périodes de verglas ou de tempêtes, devraient être renforcées.

En conclusion, cette étude ouvre la voie à une meilleure compréhension des facteurs contribuant aux accidents de la route en Indre-et-Loire, tout en fournissant des pistes concrètes pour améliorer la sécurité routière et réduire les risques d'accidents sur le long terme.

6 Annexe

Contributions des individus aux axes 1 et 2 :

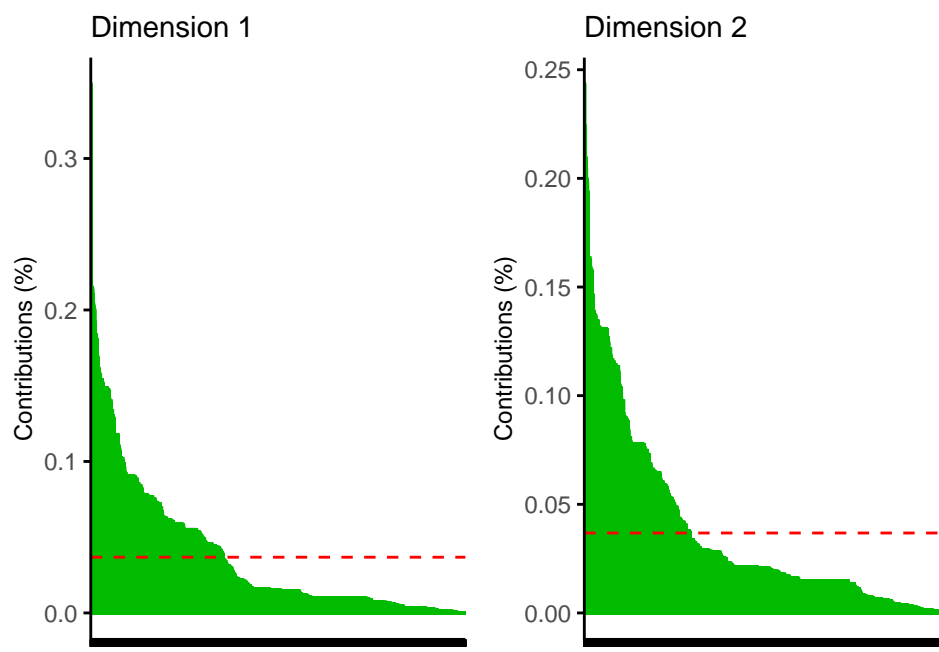


Figure 9: Histogramme des contributions des individus aux axes

Contributions des variables à l'axe 3 :

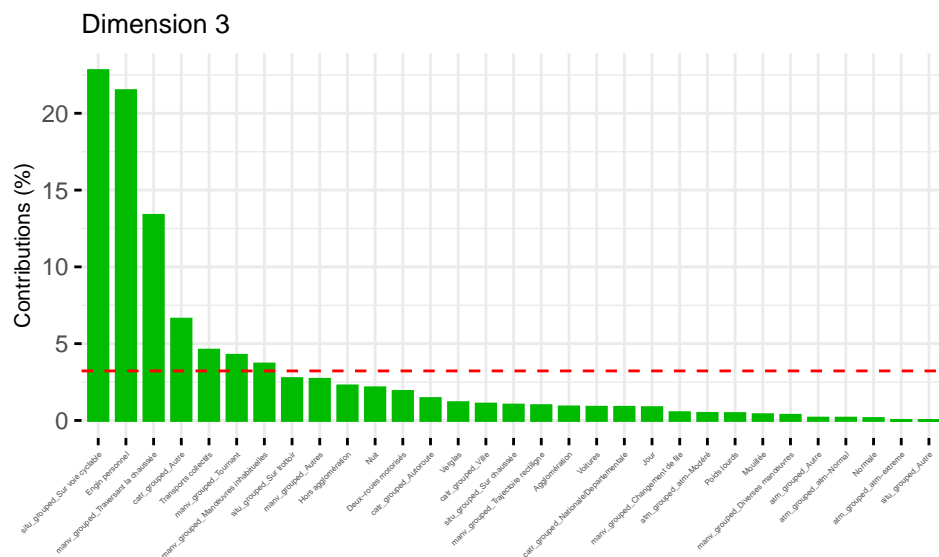


Figure 10: Histogramme des contributions des variables à l'axe 3

Tableau des coefficients corrélation :

Table 8: Dim 1

	R2	p-value
agg	0.72	0.000000000
catr_grouped	0.76	0.000000000
catv_grouped	0.24	0.000000000
situ_grouped	0.14	0.000000000
lum_grouped	0.10	0.000000000
vosp_grouped	0.08	0.000000000
atm_grouped	0.04	0.000000000
catu	0.04	0.000000000
manv_grouped	0.04	0.000000000
trajet_grouped	0.04	0.000000000
mois	0.04	0.000000000
grav_grouped	0.03	0.000000000
sexe	0.00	0.002091381

Table 9: Dim 2

	R2	p-value
surf_grouped	0.82	0.000000e+00
atm_grouped	0.78	0.000000e+00
mois	0.26	0.000000e+00
lum_grouped	0.13	0.000000e+00
catv_grouped	0.05	0.000000e+00
situ_grouped	0.05	0.000000e+00
manv_grouped	0.03	0.000000e+00
catr_grouped	0.02	1.298329e-09
trajet_grouped	0.01	2.243101e-08
catu	0.00	6.344184e-03
agg	0.00	3.696365e-02

Nuage des modalités : axe 3 et axe 4

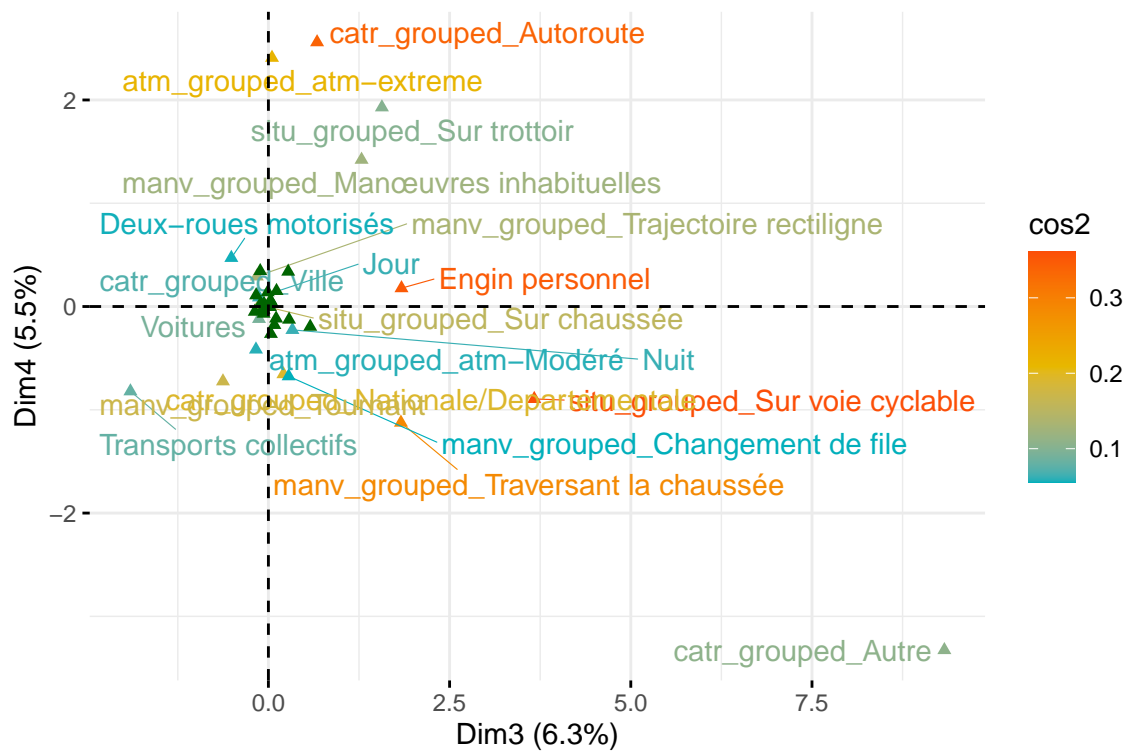


Figure 11: Nuage des modalités : axe 3 et 4

Représentation des variables supplémentaire quantitative (axe 1 et axe 2) :

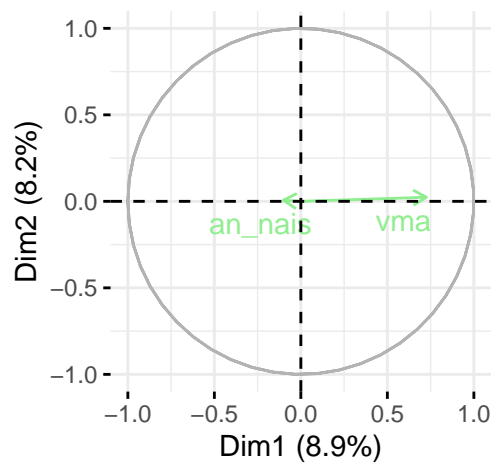


Figure 12: Représentation des variables supplémentaires quantitatives

Tableau interprétation clusters :

Table 10: Cluster 1

	Cla/Mod	Mod/Cla	Global	p.value
catr_grouped=catr_grouped_Autre	100.0	100.0	0.1	2.991e-10
catv_grouped=Engin personnel	1.2	100.0	9.3	7.884e-04
situ_grouped=situ_grouped_Sur voie cyclable	3.0	66.7	2.5	1.782e-03
vosp_grouped=Voie cyclable	0.6	100.0	17.4	5.276e-03

Table 11: Cluster 2

	Cla/Mod	Mod/Cla	Global	p.value
situ_grouped=situ_grouped_Sur voie cyclable	95.5	100	2.5	2.1e-126
vosp_grouped=Voie cyclable	10.5	78	17.4	6.0e-27
catv_grouped=Engin personnel	14.7	58	9.3	4.8e-23
agg=Agglomération	3.3	100	71.9	4.8e-10

Table 12: Cluster 3

	Cla/Mod	Mod/Cla	Global	p.value
situ_grouped=situ_grouped_Sur trottoir	97.7	100.0	1.6	7.887e-94
manv_grouped=manv_grouped_Manœuvres inhabituelles	14.8	30.2	3.2	3.084e-10
catr_grouped=catr_grouped_Ville	2.3	97.7	66.3	4.242e-07
agg=Agglomération	2.2	100.0	71.9	5.892e-07

Table 13: Cluster 4

	Cla/Mod	Mod/Cla	Global	p.value
catv_grouped=Transports collectifs	100.0	100	1.8	9.280e-108
trajet_grouped=trajet_grouped_Utilisation professionnelle	8.5	52	11.2	8.053e-13
agg=Agglomération	2.5	98	71.9	1.227e-06
catr_grouped=catr_grouped_Ville	2.6	94	66.3	3.226e-06

Table 14: Cluster 5

	Cla/Mod	Mod/Cla	Global	p.value
surf_grouped=Verglas	100.0	100.0	0.1	2.991e-10
mois=mois_1	1.3	66.7	5.7	9.515e-03
situ_grouped=situ_grouped_Sur chaussée	0.0	33.3	90.4	2.669e-02
surf_grouped=Normale	0.0	0.0	70.5	2.562e-02

Table 15: Cluster 6

	Cla/Mod	Mod/Cla	Global	p.value
manv_grouped=manv_grouped_Autres	93.1	100.0	2.6	9.318e-129
catv_grouped=Engin personnel	7.9	29.9	9.3	1.362e-06
mois=mois_1	7.1	16.4	5.7	1.458e-03
situ_grouped=situ_grouped_Sur chaussée	2.7	98.5	90.4	9.848e-03

Table 16: Cluster 7

	Cla/Mod	Mod/Cla	Global	p.value
catv_grouped=Voitures	91.5	81.3	74.8	6.518e-64
situ_grouped=situ_grouped_Sur chaussée	87.6	94.1	90.4	2.763e-39
atm_grouped=atm_grouped_atm-Normal	87.7	76.4	73.3	5.287e-16
surf_grouped=Normale	87.7	73.5	70.5	3.877e-14

Table 17: Cluster 8

	Cla/Mod	Mod/Cla	Global	p.value
atm_grouped=atm_grouped_atm-extreme	97.9	100.0	3.6	1.413e-174
surf_grouped=Mouillée	11.6	97.9	29.4	1.610e-48
mois=mois_5	16.0	33.7	7.4	1.992e-14
catr_grouped=catr_grouped_Autoroute	16.0	21.1	4.6	4.335e-09

Table 18: Cluster 9

	Cla/Mod	Mod/Cla	Global	p.value
catv_grouped=Poids lourds	98.0	100.0	3.6	5.281e-176
trajet_grouped=trajet_grouped_Utilisation professionnelle	11.5	36.5	11.2	3.701e-11
mois=mois_2	14.6	26.0	6.3	3.083e-10
agg=Hors agglomération	7.3	58.3	28.1	3.565e-10

Table 19: Cluster 10

	Cla/Mod	Mod/Cla	Global	p.value
atm_grouped=atm_grouped_Autre	100.0	100	0.4	1.677e-28
mois=mois_8	4.4	80	6.7	1.483e-08
surf_grouped=Mouillée	1.3	100	29.4	4.630e-06
trajet_grouped=trajet_grouped_Domicile-travail/école	1.1	90	30.9	1.864e-04