

PROJE STA101 : ANALYSE DE DONNEES DESCRIPTIVES

Accident corporel de la circulation routière en Ile de France

Auditeur : **Jean Marie ABENG ZOA**
Année : 2022 / 2023
Cnam Paris

Table des matières

1	Introduction	3
2	Objectif	3
3	Jeu de données	3
3.1	Prétraitement des données	4
4	Analyse univarié	6
4.1	Variables qualitatives	6
4.2	Variables quantitatives	9
5	Analyse bivariée	10
6	Analyse des correspondances multiples	13
6.1	Préparation des données pour l'ACM	13
6.2	Choix des valeurs propres	14
7	Classification ascendante hiérarchique	20
7.1	Description des classes	22
8	Conclusion	24

1 Introduction

Un accident de la circulation est bien entendu comme une collision non voulue, non prévue et mal anticipée entraînant au moins un engin roulant avec une chose, un animal ou une personne sur une voie publique ou privée ouverte à la circulation. En France, plus de 50 000 accidents corporels sont recensés chaque année par l'Observatoire Nationale Interministériel de la sécurité routière ONISR et les compagnies d'assurances indiquent qu'il y a 1,8 millions de constats amiables durant la même période.

2 Objectif

Cette étude traite les accidents corporels de la circulation routière dans la région Ile de France (IDF) en 2021.

L'objectif est de réaliser une analyse descriptive tout en posant un certain nombre de questions. Les conditions atmosphériques influent-elles sur la gravité des usagers accidentés ? Le type de collision a-t-elle une influence sur la gravité des accidents ? Les conditions d'éclairages sont-elles liées avec la localisation ? La vitesse et la catégorie de la route sont-elles liées ? Y'a-t-il un lien entre la gravité de l'accident et la circulation ?

Pour répondre à toutes ces questions, nous allons commencer par décrire le jeu de données, puis nous allons mettre en application les techniques d'analyse descriptive vu en cours tout en exploitant les outils d'analyse comme (ACP, ACM, la classification, AFC, AFD...).

3 Jeu de données

Le jeu de données est issu du site open data du gouvernement. Le lien de téléchargement se trouve [ici](#). Ce jeu de données décrit les accidents corporels routiers en France en 2021. Pour chaque accident corporel, des saisies d'informations décrivant l'accident sont effectuées par une unité de force de l'ordre qui est intervenu sur le lieu d'accident. Ces saisies sont rassemblées dans une fiche intitulée bulletin d'analyse des accidents corporels. L'ensemble de ces fiches constitue le fichier national des accidents corporels de la circulation dite « Fichier BAAC » administré par l'Observatoire National interministériel de la sécurité routière 'ONISR'.

Le jeu de données se compose de quatre fichiers :

- Fichier **Véhicules**, qui va rassembler les véhicules impliqués pendant l'accident,
- Fichier **Lieux**, décrit le lieu principal de l'accident,
- Fichier **Usagers**, rassemble les informations sur les personnes impliqués,
- Fichier **Caractéristiques** va décrire les circonstances générales de l'accident.

3.1 Prétraitement des données

Nous avons décidé de nous concentrer sur l'étude des accidents en Région Parisienne. Pour ce faire, nous avons procédé au nettoyage de nos données sous Jupyter Notebook (en python).

Dans les fichiers d'origines, nous avons remarqué que dans certaines variables il y'avait des informations qui étaient manquantes d'une part et d'autre part les informations dans certaines variables étaient vide. Nous avons concaténé les quatre fichiers puis supprimer les valeurs nulles et manquantes. Nous avons également créé une variable âge puisque qu'on avait seulement l'année de naissance des usages. Après ce traitement, nous nous sommes retrouvés avec un jeu de données de 10798 observations et 21 variables avec 5 variables quantitatives et 16 variables qualitatives. Le tableau de description des variables est ce dessous.

Trajet	Promenade loisirs	Qualitative
	Domicile travail	
	Utilisation professionnelle	
	Autres	
	Courses achats	
	Domicile école	
Catégorie route (catr)	Route départementale	Qualitative
	Voie comunale	
	Autoroute	
	Route nationale	
	Hors réseau public	
	Autres	
Circulation (circ)	Parc de stationnement ouvert public	
	Bidirectionnelle	Qualitative
	A sens unique	
	A chaussées séparées	
	Avec voie d'affection variable	
Nombre de voies de cir	Voie de circulation	Quantitative
Signal l'existence voie	Piste cyclable	Qualitative
	Voie réservée	
	Bande cyclable	
	Sans objet	
Catégorie usagers (catu)	Conducteur	Qualitative
	Passagers	
	Pieton	

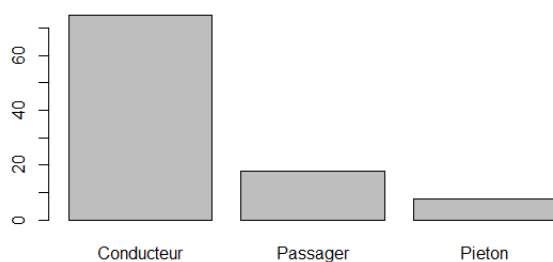
Surface (surf)	Normale Mouillée Enneigée Verglacée Flaques Autres Corps gras huile Inondée Boue	Qualitative
Infrastructures (infra)	Carrefour aménagé Autres Souterrain tunnel Pont autopont Bretelle de raccordement Zonne piétonne Chantier Voie ferrée Zone de péage Aucun	Qualitative
Situation accident	Sur chaussée Sur piste cyclable Sur trottoir Autres Sur autres voie spéciale Sur accotement Sur bande d'arrêt d'urgence	Qualitative
Lumière	Plein jour Nuit avec éclairage public allumé Nuit sans éclairage public Aube Nuit avec éclairage public éteint	Qualitative
Localisation (agg)	En agglomération Hors agglomération	Qualitative
Intersection (int)	Hors intersection Intersection en X Intersection en T Intersection en Y Autres intersection Place Giratoire Intersection à plus de 4 branches Passage à niveau	Qualitative

Condition atmosphérique	Normal Pluie légère Temps couvert Pluie forte Tempes éblouissant Neige grêle Brouillard fummée Vent fort tempête	Qualitative
Type de Collision	Deux véhicules par le coté Autres collision Deux véhicules par l'arrière Sans collision Deux véhucules frontales Trois véhicules et plus en chaine Trois véhicules et plus collision mult	Qualitative
Age	Age	Quantitative
Sexe	Sexe masculin	Quantitative
	Sexe féminin	Quantitative
Victime	Indemne Décéder Blésser hospitalisé Blésser legé	Quantitative
Mois accident	Mois	Quantitative
Vitesse max	vitesse	Quantitative
Lieux acc Dep	Departement	Quantitative

4 Analyse univariée

Afin d'avoir une première idée de la distribution des variables ainsi qu'identifier s'il y'a des anomalies dans les données, nous allons effectuer une analyse univariée.

4.1 Variables qualitatives



Concernant la Figure 1 : Variable catégorie usager, nous pouvons voir que les individus les plus impliqués aux accidents sont les conducteurs.

Figure 1 : Variable catégorie usager

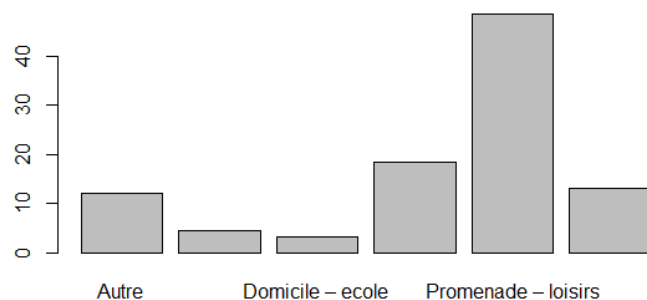


Figure 2 : Variable trajet

En explorant la *Figure 2 : Variable trajet*, nous remarquons qu'il y'a plus d'accident chez des personnes qui font des promenades-loisirs. Les individus faisant le trajet Domicile-école sont moins accidentés.

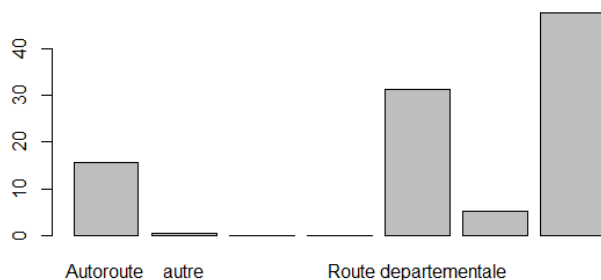


Figure 3 : Variable catégorie route

Nous pouvons remarquer sur la *Figure 3 : Variable catégorie route* que les routes ou il y'a le plus d'accidents sont les Routes départementales et les voies communales.

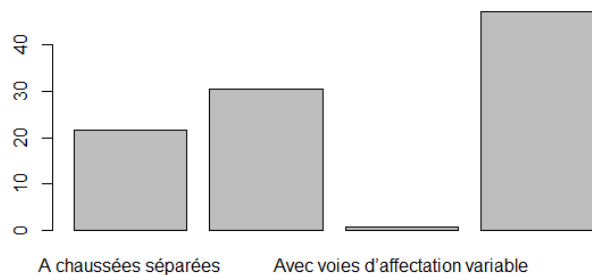


Figure 4 : Régime circulation

Nous pouvons voir sur la *Figure 4 : Régime circulation* que les individus font le plus d'accident sur les voies bidirectionnelle puis les voies à sens unique.

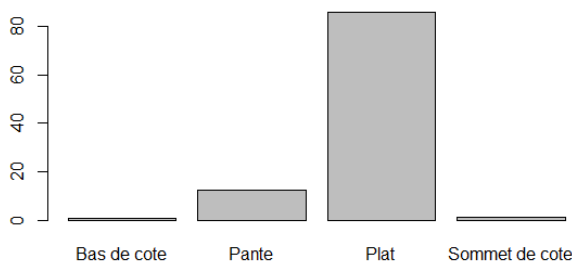


Figure 5 : Profile route

La *Figure 5 : Profile route* nous montre que la plupart des accidents ont eu lieu sur les routes de profile plat.

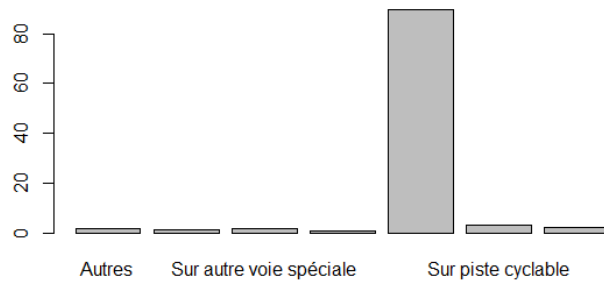


Figure 6 : Variable situation accident

La Figure 6 : Variable situation accident nous montre que la plupart des accidents ont eu lieu sur la chaussée.

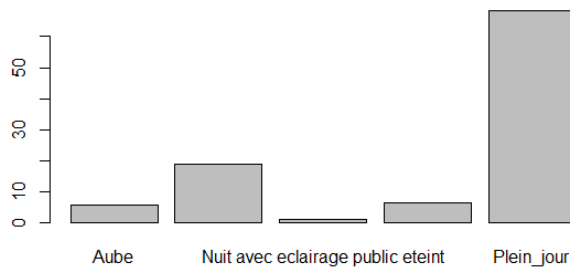


Figure 8 : Variable lumière

Nous voyons sur la Figure 8 : Variable lumière que la plupart des individus ayant fait l'accident c'était en plein jour.

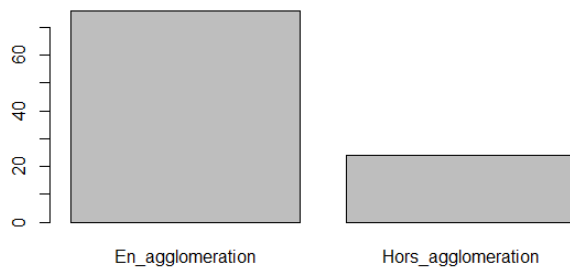


Figure 7 : Variable Localisation

Nous pouvons voir ici que la plupart des individus ont fait l'accident en agglomération.

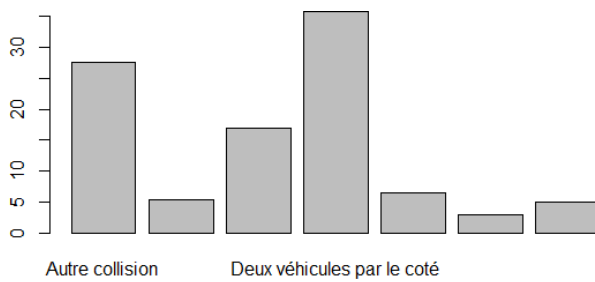
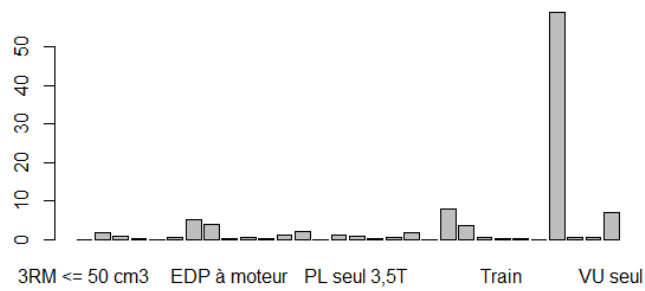


Figure 9 : Type de collision

La Figure 9 : Type de collision montre que le plus des collisions ayant lieu lors des accidents sont des collisions impliquant deux véhicules par l'arrière, par le coté et autres collisions.



Dans cette variable catégorie de véhicule les individus les plus impliqués aux accidents sont les véhicules seuls.

Figure 10 : Variable Catégorie de véhicule

4.2 Variables quantitatives

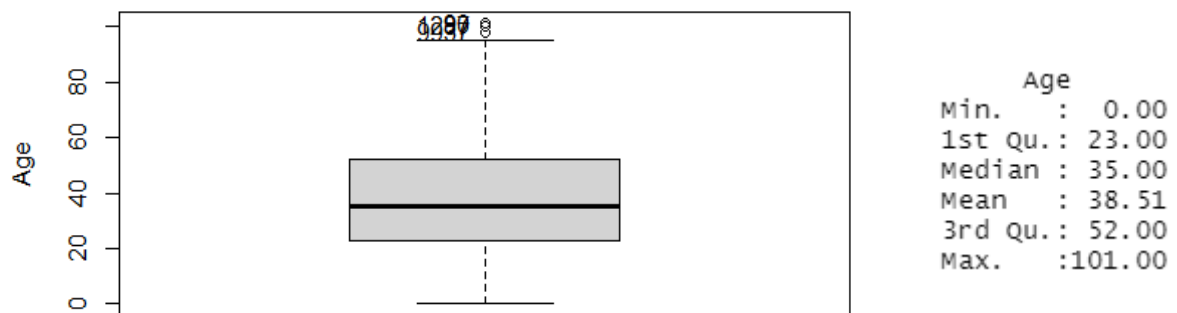


Figure 11 : Variable Age

Concernant la variable **Age**, la moyenne d'âge des individus ayant été impliqué aux accidents est de 38, 5 ans. Nous pouvons voir que 25% de ces individus ont moins de 23 ans et 75% plus de 23 ans.

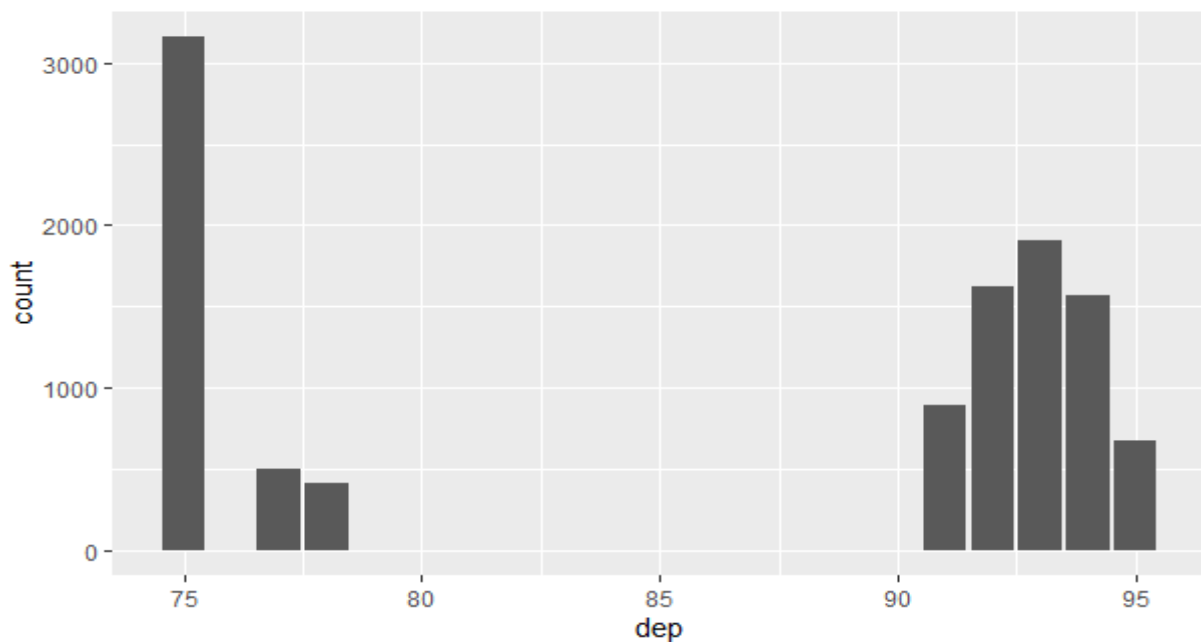
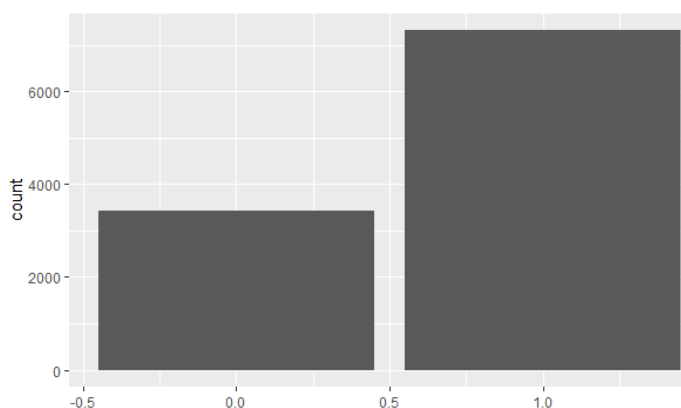


Figure 12 : Variable Département

Nous pouvons voir ici que la plupart des accidents ont eu lieu dans 4 départements. Dans Paris, on peut voir plus de 3000 accidents. Dans la Seine Saint Denis (93), les Hauts de Seine (92) et Val de Marne on peut compter près de 2000 accidents par départements.



En regardant la variable sexe, on voit que les individus les plus accidentés sont de sexe masculin, on peut compter un peu plus de 7000 accidents.

Figure 13 : Variable sexe

5 Analyse bivariable

Afin d'avoir plus ample information nous allons regarder si les variables sont liées entre elles (deux à deux). Ce qui peut nous permettre de répondre à certaines questions que nous nous sommes posés au début.

- Liaison entre **type de collision** et **gravité d'accident** (grav ou victime) :

En effectuant le calcul sur le rapport de corrélation et en vérifiant le test de l'association, nous avons trouvé que le rapport de corrélation est très proche de 0 (**rapport de corrélation = 0.0005793012**) et que la P-value est proche de 5% (**P-value = 3.99%**) ce qui veut dire que les deux variables ne sont pas liées.

- Liaison entre **condition d'éclairage (lum)** et **la localisation (agg)**:

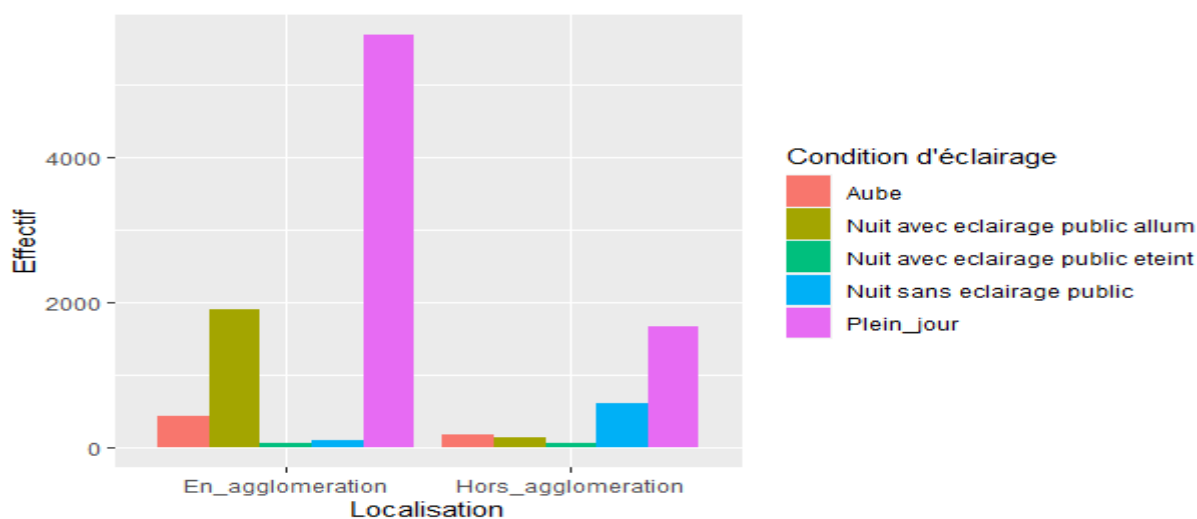


Figure 14 : Localisation-Lumière

Lorsqu'on observe la Figure 14 : Localisation-Lumière , on voit que les conditions d'éclairages influent sur la localisation. En faisant un zoom **en agglomération**, on remarque que le nombre d'accident (près de 6000 individus) est plus important en **plein jour** et cela peut s'expliquer par le fait qu'il y'a un nombre important d'utilisateurs sur les voies. **Hors agglomération** on observe également un nombre assez important (près de 2000 individus) des accidents. Le test de Khi deux nous a donné une p-value extrêmement faible (**p-value < 2.2e-16**), l'hypothèse d'indépendance est rejetée. Les deux variables sont donc liées statistiquement.

X-squared = 1849.6, df = 4, p-value < 2.2e-16

- Liaison la **gravité de l'accident** (victime ou grav) et **Circulation (circ)** :

Après calcul, nous avons une **p-value = 0.09855** est faible. L'hypothèse d'indépendance est rejetée, les deux variables gravités de l'accident et circulation sont statistiquement liées. Nous pouvons voir que les individus ayant été accidenté soit en circulation à sens unique, à chaussée séparée ou bidirectionnelle ont des blessures légères.



Figure 15 : Gravité accident – Circulation

- Liaison entre **trajet** et **gravité de l'accident** (victime ou grav) :

Cette liaison nous donne une p-value < 2.2e-16, valeur extrêmement inférieure à 5%. L'hypothèse d'indépendance est rejetée. Les deux variables **trajet** et **gravité de l'accident** (grav) sont donc liées statistiquement. On peut voir sur Figure 16 : Gravité accident - Trajet que les individus accidentés qui sont sortis indemnes près de 2000 sortaient des promenades loisirs (1). On observe également que la plupart de ceux qui ont été blessé hospitaliser (3) et blessé léger (4) sortaient aussi des promenades loisirs.

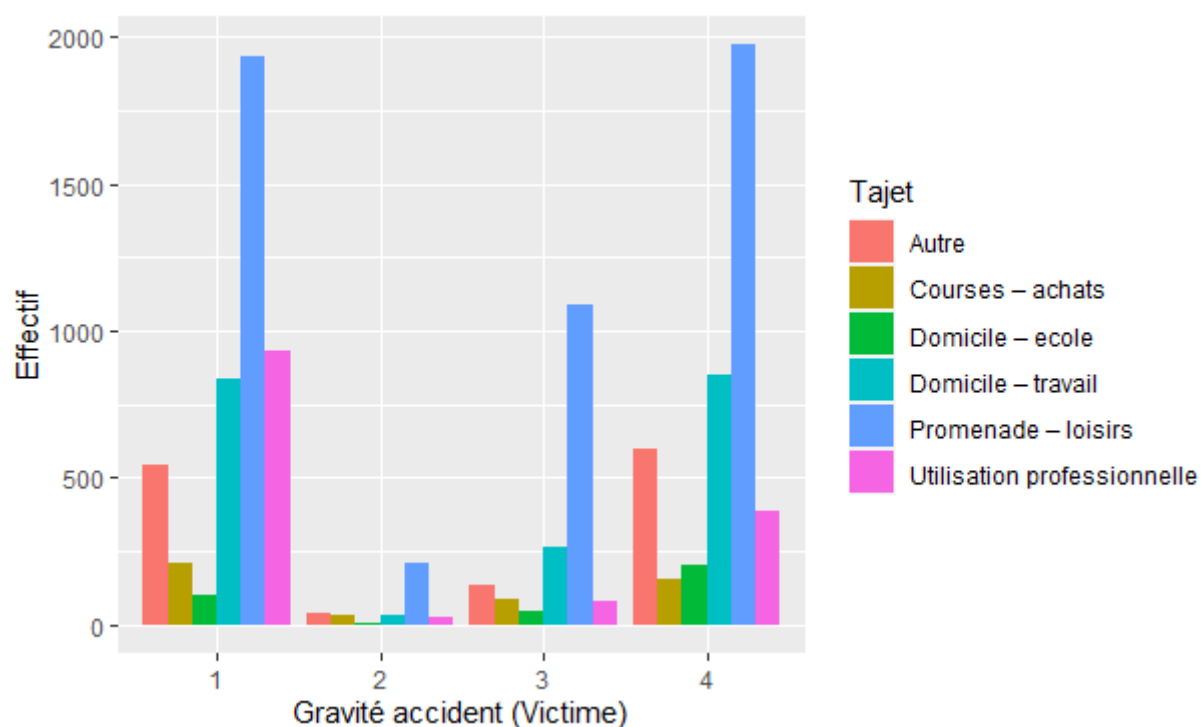


Figure 16 : Gravité accident - Trajet

Pour voir si les autres variables sont liées, nous avons réalisé le test d'anova (variable qualitative et variable quantitative) le test de Khi deux pour les variables qualitatives.

Variables	P-value
Grav / Situa	0.6373
Grav / Catr	0.2828
Grav / Catu	< 2.2e-16
Grav / Infra	0.5578
Grav / Sexe	< 2.2e-16
Grav / Age	< 2.2e-16
Vma / Catr	< 2.2e-16
Grav / Vma	0.6355
Col / Lum	< 2.2e-16

Au vu des résultats du tableau, les variables gravité accident (grav) et situation (situa), grav et infrastructure (infra), grav et vitesse maximale (vma) ont une P-value supérieur à 5%. On ne peut donc pas mettre en évidence un lien entre les variables ces variables.

Par contre on observe bien que les autres variables ont une P-value extrêmement faible, l'hypothèse d'indépendance est donc rejetée. On peut donc conclure que les variables ayant une P-value faible comme Grav et Catu, Grav et Sexe, Grav et Age, Vma et Catr, Col et Lum sont liées statistiquement.

L'analyse bivariée nous a permis de voir que la variable gravité de l'accident (victime) est liée à d'autres variables. Nous avons répondu à certaines problématiques que nous avons posé au départ au moyens des tests statistiques et des graphiques.

6 Analyse des correspondances multiples

Afin de pouvoir répondre aux problématiques posées c'est-à-dire savoir si les conditions d'éclairages (Lumière) influent sur la localisation (agg) ou l'accident a eu lieu, si la vitesse est liée à la catégorie de la route (catr), la catégorie de la route (catr) est – elle liée avec la localisation de l'accident (agg) ? Nous avons choisi l'ACM parce que notre jeu de données contient en majorité les variables qualitatives donc cette solution semble appropriée pour notre analyse.

6.1 Préparation des données pour l'ACM

Après avoir calcul comme le montre les distributions des variables **annexe 1**, nous avons vu qu'il y'a plusieurs modalités avec de faibles effectifs. Ces modalités pourraient très influencer sur la construction des axes. Pour essayer de remédier à cela, nous allons regrouper les effectifs de certaines modalités.

Pour la variable circulation, les effectifs des modalités « Avec voie d'affectation variable » et « A chaussée séparée » sont regroupés. Pour la variable trajet, les effectifs des modalités « courses – achats » et « domiciles – écoles » sont regroupés avec « autres ». Pour la variable signal l'existence voie les effectifs des modalités « voie réservée » et « bande cyclable » sont regroupés avec « piste cyclable ». Pour la variable catégorie route les effectifs des modalités « parc de stationnement public », « hors réseau publique », « autres » et « routes nationales » sont regroupés avec « auto route ». Pour la variable profil les effectifs des modalités « Bas de côte » et « Sommet de côte » sont regroupés avec « Pante ». Pour la variable plan les effectifs des modalités « en S » sont regroupés « En courbe à droite ». Pour la variable surface, les modalités avec les effectifs faibles sont regroupées avec les affectifs de la modalité « mouillée ». Pour la variable lumière (éclairage) les effectifs de la modalité « nuit avec éclairage public éteint » sont regroupés avec « nuit sans éclairage public ». Le même est réalisé sur la variable condition atmosphérique, intersection et infrastructure.

I

Choix des variables :

Pour poursuivre notre travail, nous allons choisir comme variables actives celles qui étaient liées statiquement car elles contribueront le plus à la construction des axes factoriels. Nous allons écarter de notre analyse les variables (nombre de voies, mois accident) qui ne semblent pertinentes. Comme variable illustrative nous utiliserons la vitesse maximale (vma), le département (dep), l'âge (Age), et sexe car elles vont beaucoup plus nous aider lors de l'interprétation des résultats.

La variable gravitée a été transformé en qualitative.

6.2 Choix des valeurs propres

Pour effectuer notre analyse, nous avons besoin de trouver le nombre d'axes en fonction des valeurs propres. Selon la règle de Kaiser, on pourrait potentiellement décider de conserver les axes dont les valeurs propres sont supérieures à $1/P$ ($1/16 = 0,0625$) avec P le nombre des variables actives conservées. En appliquant donc cette règle, on pourrait restreindre l'analyse aux 14 premiers axes. Ce qui rendra trop lourd notre analyse. En appliquant la règle du coude, on conservera 6 axes qui portent près de 30% de l'inertie.

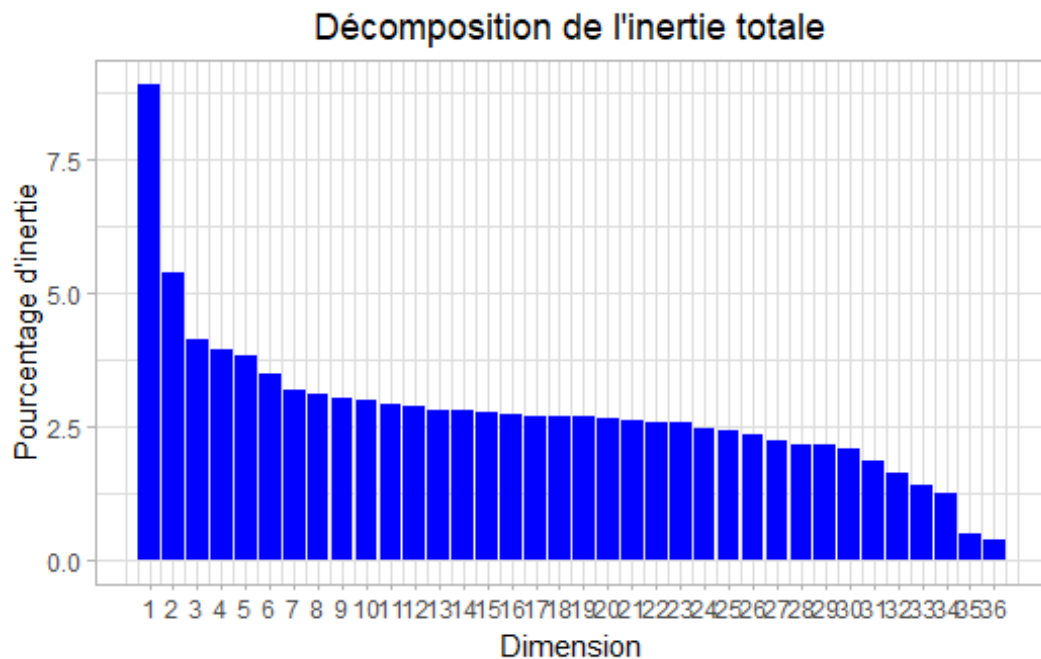


Figure 17 : Décomposition de l'inertie

Graphe des individus Dim 1 et Dim 2 :

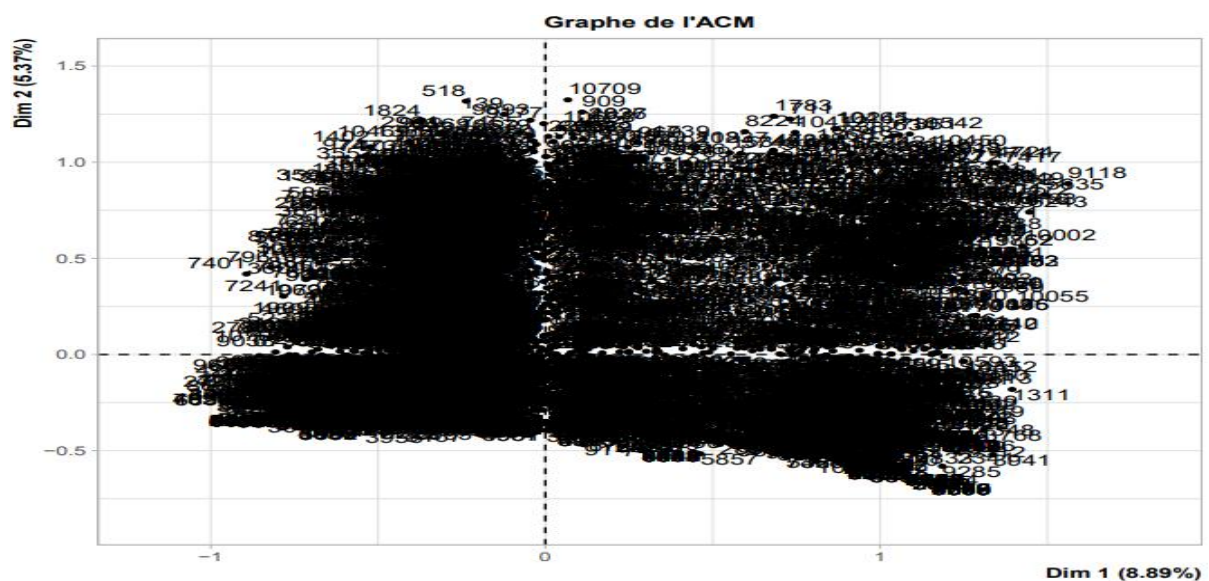


Figure 18 : Graphe des individus Dim 1 et Dim 2

Grphe des individus Dim 5 et Dim 6 :



Figure 19 : Nuage individus Dim 5 et 6

15

Graphe des modalités Dim 1 et 2 :

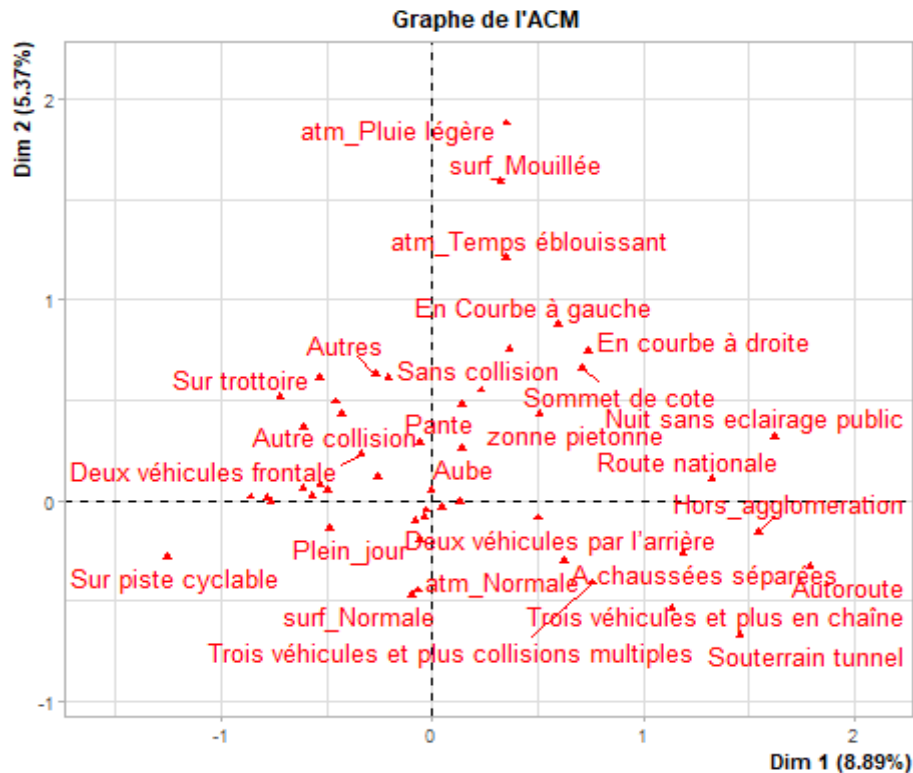


Figure 21 : Graphe des modalités Dim 1 et 2

Sur le graphe des modalités, nous observons que les modalités de certaines variables ressortent beaucoup par rapports au autres comme par exemple la variable condition atmosphérique (atm), condition d'éclairage ou lumière (lum), type de collision (col). Nous allons améliorer la qualité de représentation afin de pouvoir réaliser une bonne analyse.

Graphe des contribution des modalités Dim 1 et 2:

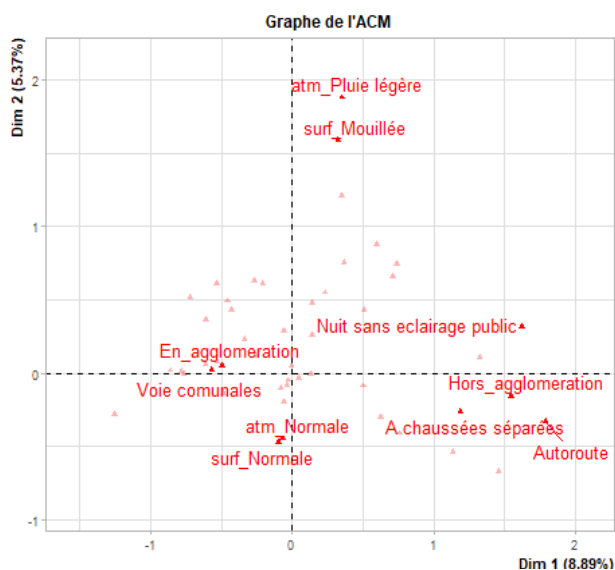


Figure 22 : Modalités de contribution

La Figure 22 : Modalités de contribution montre les 10 modalités ayant le plus contribuer à la construction du plan factoriel (Dim 1 et 2). On peut voir que la modalité autoroute contribue significativement sur l'axe des abscisses. Nous ferons une interprétation détaillée par la suite.

En analysant les axes du plan factoriel (Dim 1 et Dim 2) qui résume 14,26% de l'information, sur l'axe Dim 1 du premier plan factoriel nous observons que cette dimension sépare les modalités en deux groupes. D'une part Hors_agglomération, Autoroute, A chaussées séparées, deux véhicules par l'arrière, atm_normal, nuit_sans_eclairage_public, route nationale qui sont caractérisées par les coordonnées positives. D'autre part les modalités plein_jour, voie communale, bidirectionnelle, en agglomération, sur piste cyclable, autre collision, sur trottoir qui ont des coordonnées négatives. Sur cette même dimension, nous observons bien les modalités en agglomération et voie communales des variables (agg « localisation » et catr « catégorie de route ») qui sont très proches et donc très corrélés. On voit que les individus les plus accidentés étaient en agglomération et ces les plus accidentés sur l'autoroute étaient hors agglomération. On peut également dire que les modalités nuit sans éclairage public et route nationale sont dans le même groupe et donc les individus accidentés lorsque l'éclairage public était éteint ces individus se trouvaient dans les routes nationales. L'Annex 2 montre la corrélation des modalités les plus liées au moins liées.

Sur l'axe Dim 2, nous observons également deux groupes. D'une part les modalités atm_pluie légère, sur_mouillée, atm_temps éblouissant, nuit sans éclairage public qui sont caractérisé par les coordonnées positives. D'autre part les modalités plein_jour, atm_normal, surf_normal, trois_voitures_plus_collision_multiples qui sont caractérisés négativement. Les modalités qui contribuent le plus sont celles ayant une forte coordonnée. On peut voir que les modalités atm pluie légère et surface mouillée des variables (conditions atmosphériques et surfaces) sont très liées donc corrélés. Cette même dimension oppose les individus 1788 et 10098 provenant de la zone des accidents ayant eu lieu hors agglomération sur les autoroutes aux individus 10709 et 909 ayant fait l'accident la nuit sans éclairage public sur les routes nationales.

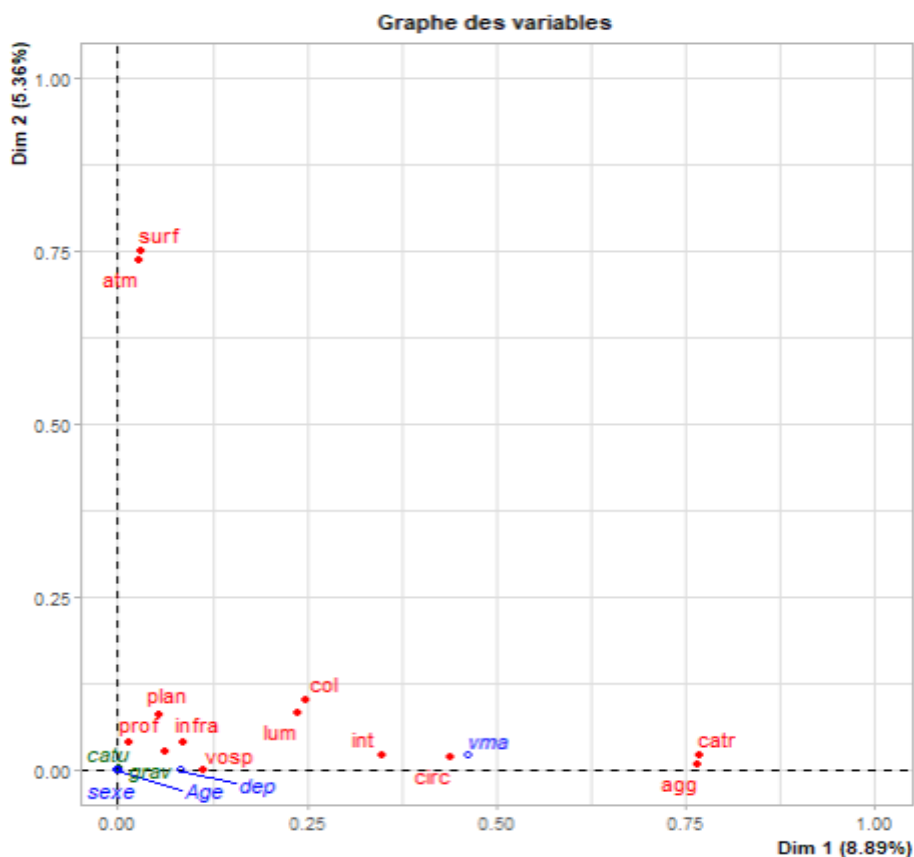


Figure 23 : Graphe des variables

Globalement, les accidents ayant eu lieu hors agglomération étaient sur des routes nationales et sur les autoroutes généralement la nuit sans éclairage public. Les individus étaient plus impliqués aux accidents en agglomération sur routes départementales et sur des voies communales et la plupart de ces accidents étaient en plein jour.

[illegible]

La dimension 3 oppose les modalités des individus qui sont positifs sur l'axe des abscisses aux modalités des individus caractérisés négativement. A droite de cette dimension, on a les modalités route départementale, deux véhicules frontale, atmosphère normale qui forment un groupe. Ce qui nous amènes à dire que ces modalités sont liées et donc les variables catégorie de routes, conditions atmosphériques et types de collision sont corrélés l'Annexe 2 des corrélations

des modalités le confirme. On a donc tendance à voir que les individus ayant fait l'accident sur les routes départementales les conditions atmosphériques étaient normales et le type de collision était des collisions impliquant deux véhicules frontaux. A gauche on trouve l'atmosphère pluie légère, deux véhicules frontaux. La dimension 4 oppose les individus ayant fait l'accident sur les voies réservées à ceux ayant fait l'accident sur les routes normales (route départementale par exemple).

Graphe des modalités Dim 5 et 6 :

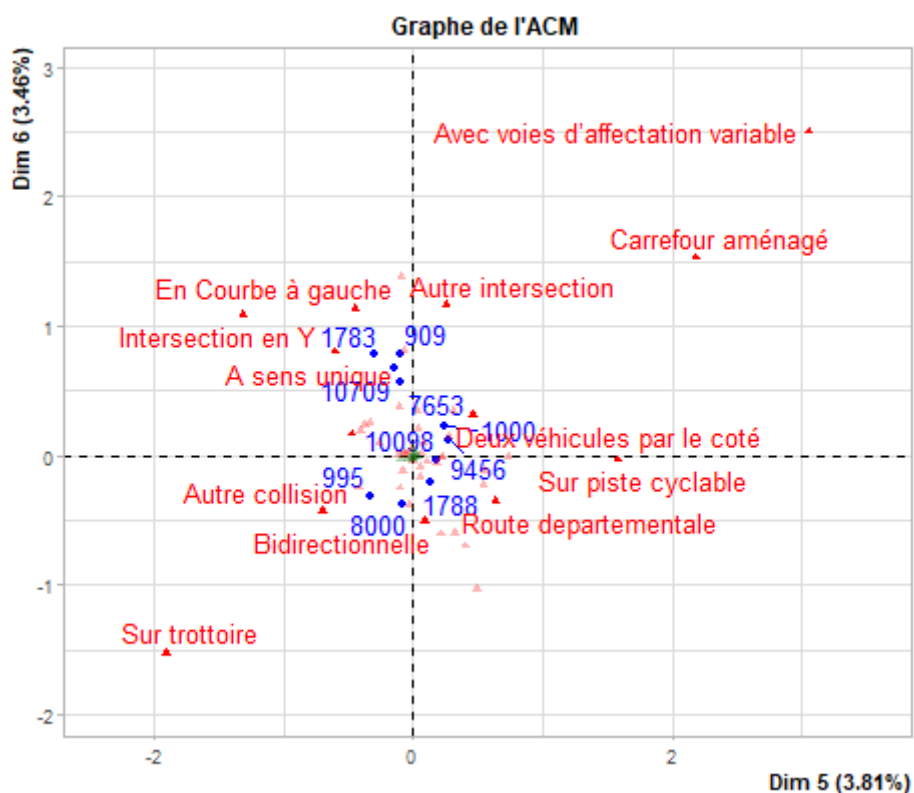


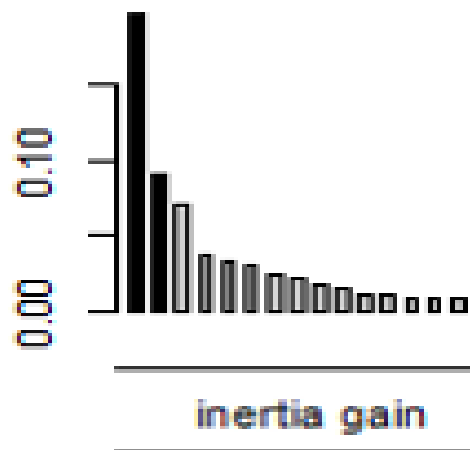
Figure 26 : Graphe des modalités Dim 5 et 6

La dimension 6 oppose des individus ayant fait l'accident dans les voies de circulation à sens unique, à chaussée séparée dont les catégories des routes étaient des voies communales et donc les conditions d'éclairage étaient la nuit avec éclairage public allumé. Par exemple si on prend les individus 10709 et 7653, ils sont opposés aux individus ayant fait l'accident dans les catégories des routes départementales dont la circulation est bidirectionnelle (individus 8000 et 1788). Lorsqu'on repart sur le jeu de données regarder ces individus pris, on voit que les individus 10709 et 7653 ont été accidentés sur des voies communales à sens unique la nuit avec éclairage public allumé et que les individus 8000 et 1788 ont été accidentés sur les routes départementales bidirectionnelles la nuit avec éclairage public allumé. Le fait que certaines de ses modalités soient proches et en se référant sur le tableau des corrélations des modalités Annexe 2, on va dire ici que les individus ayant été accidentés la nuit avec éclairage public allumé étaient sur des routes communales à sens unique et départementale bidirectionnelle.

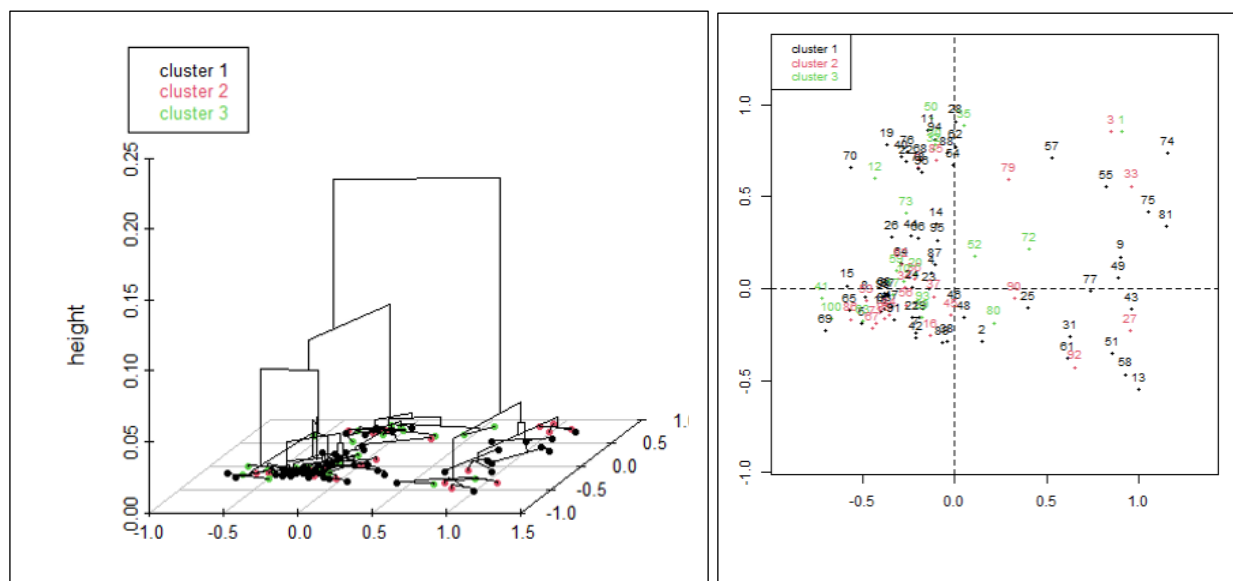
Globalement, nous allons dire que les conditions d'éclairages ou lumière (lum) influent sur la localisation de l'accident (agg). En effet, on a découvert que les individus accidentés hors agglomération généralement sur les routes nationales, les autoroutes la nuit sans éclairage public. En agglomération la plupart des accidentés étaient sur des voies communales et c'était en plein jour. Ce qui nous amènes à dire que les conditions d'éclairage lumière (lum) influent sur la localisation de l'accident et que la catégorie des routes à aussi une influence sur la localisation ces deux variables sont même d'ailleurs très liées et ont contribuent beaucoup à la construction des plans factoriels.

7 Classification ascendante hiérarchique

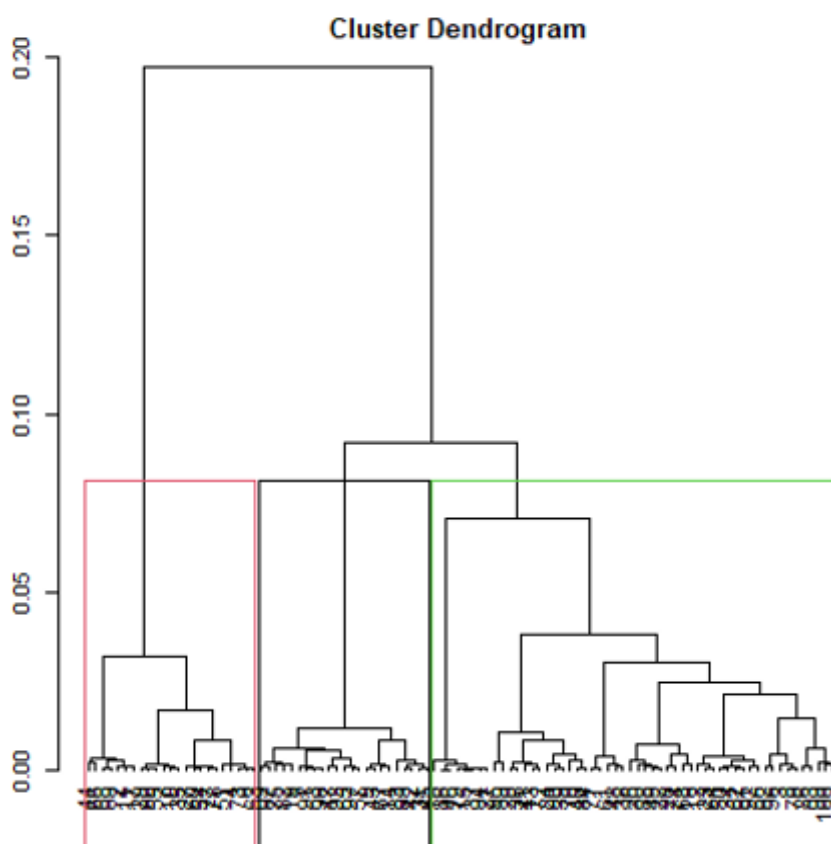
Pour faire suite à ces observations, nous allons réaliser une classification en utilisant la méthode de Ward. Elle consiste donc à regrouper les individus pour que l'augmentation de l'inertie interclasse soit maximale. Nous allons utiliser la classification ascendante hiérarchique, l'ACM effectuer avant nous a permis de réduire la dimension de nos données et de conserver les axes pertinents.



Ci-dessus, nous avons un diagramme de gain d'inertie qui regroupe 15 classes au total. On remarque que lorsqu'on quitte de la classe 15 à la classe 14, il y'a une perte considérable d'inertie et lorsqu'on passe de la classe 14 à la classe 13 on perd également l'inertie. On peut donc voir qu'il y'a une perte importe d'inertie lorsqu'on passe de 3 à 2 classes et lorsqu'on passe de 2 à 1 classe. Nous resterons donc sur les 3 classes suggérer par le diagramme.



Dans la figure à gauche, on peut voir que les classes noires sont moins séparés que les autres. Dans les axes 3 et 4, les classes noires sont mieux séparés.



7.1 Description des classes

Classe 1 :

	Cla/Mod	Mod/Cla	Global	P.value	V.test
atm=atm_Normale	78.36	92.20	77.17	0.00	Inf
agg=En_agglomeration	81.42	94.05	75.76	0.00	Inf
catr=Voie comunales	80.79	58.74	47.68	0.00	Inf
circ=Bidirectionnelle	79.61	57.30	47.20	4.99e - 22	32.23
catr=Route departementale	81.13	38.67	31.26	1.30e - 19	23.40
col=Deux véhicules par le coté	77.01	42.06	35.82	1.68e - 80	19
lum=Plein_jour	69.45	72.225	68.19	8.18e - 35	12.30
col=Deux véhicules frontale	79.93	6.46	5.30	1.03e - 14	7.73
lum=Nuit avec eclairage public allum	70.17	20.13	18.81	1.08e - 06	4.87
circ=A sens unique	67.99	31.58	30.46	4.72e - 04	3.49
col=Autre collision	67.94	28.55	27.56	1.42e - 03	3.18
atm=atm_Temps couvert	58.97	4.22	4.69	1.52e - 03	-3.17

Le tableau ci-dessus nous montre la composition de la classe 1 de notre classification. Cla/Mod va nous donner le pourcentage de présence de cette modalité dans cette classe par rapport aux autres classes. Mod/Cla indique le pourcentage des individus qui ont cette modalité dans cette classe. Vtest montre la comparaison par rapport à la fréquence moyenne de cette modalité. Cet indicateur permettra de mettre en valeur la tendance de classe.

Pour la classe 1, les individus cette classe prennent des valeurs significativement différentes de 0, la plupart des Vtest sont positives. Les individus regroupés accidentés dans cette classe ont été majoritairement accidenté sur les voies communales, les routes départementales en agglomération. Ces accidents ont eu lieu sur des routes à circulation bidirectionnelle et à sens unique, en majoritairement en plein jour et impliquant des collisions au moins deux véhicules frontaux. Les conditions atmosphériques étaient normales.

Ci-dessous les individus les plus proches du centre de la classe. En allant vérifier le jeu de données, tous ces individus ont les mêmes modalités. Ils ont été accidentés en agglomération (agg) sur route départementale (catr) à sens unique (circ) en plein jour avec des conditions atmosphériques normales (atm) et sur un même trajet (promenade – loisir).

```
Cluster: 1
      6642      7371      8319      8614      9436
0.1021999 0.1720836 0.1720836 0.1720836 0.1720836
```

Classe 2 :

	Cla/Mod	Mod/Clas	Global	P.value	V.test
atm=atm_Pluie légère	65.71	64.17	14.28	0.00	Inf
surf=surf_Mouillée	62.85	98.10	22.83	0.00	Inf
atm=atm_Temps éblouissant	45.78	12.02	3.84	1.40e - 54	15.55
lum=Aube	19.30	7.40	5.30	1.21e - 03	3.23
circ=A chaussées séparées	12.08	56.32	68.19	3.48e - 4	- 13.26

Les individus de cette classe prennent des valeurs Vtest significativement élevées et différentes de 0. Ces individus regroupés de cette classe ont été accidentés sur les voies à chaussées séparées (très peu d'individus) à lors des conditions atmosphériques pas agréables (pluie légère, temps éblouissant) et c'était plutôt à l'aube.

Nous avons ici les individus les plus proches du centre de classe 2. Après avoir vérifier sur le jeu de données, ces individus ont les mêmes caractéristiques. Ils ont été accidentés sur des voies à chaussées séparées à l'aube et surtout lorsqu'il pleuvait.

Cluster: 2					
120	10284	2793	3510	688	
0.1433507	0.2017878	0.2058747	0.2118446	0.2217945	

Classe 3 :

	Cla/Mod	Mod/Cla	Global	P.value	V.test
agg=Hors_agglomeration	80.28	98.36	24.23	0.00	Inf
catr=Autoroute	96.84	76.17	15.55	0.00	Inf
lum=Nuit sans éclairage public	67.83	25.28	7.37	1.26e - 21	30.97
catr=Route nationale	73.35	20.36	5.49	3.90e - 18	29.25
col=Trois véhicules et plus en chaîne	56.93	14.41	5.01	5.19e - 86	19.65
col=Trois véhicules et plus collisions multiples	40.76	5.99	2.90	5.101e - 18	8.65
col=Deux véhicules par l'arrière	38.61	33.19	17.00	2.37e - 96	20.82

Lorsqu'on regarde l'indicateur Vtest, les individus regroupés dans la classe 3 ont été accidentés en majoritairement hors agglomération (agg) sur les autoroutes et route nationales (catr). Ces accidents ont eu lieu en partie la nuit sans éclairage public impliquaient trois ou deux véhicules (col). Les conditions atmosphériques étaient parfois mauvaises (légère pluie, temps éblouissant) et parfois bonnes.

Ci-dessous les individus les plus proches du centre de la classe 3. Lorsqu'on regarde sur le jeu de données, on a bien la confirmation que 4 de ces individus ont été accidentés sur les autoroutes et l'autre sur une route nationale. Les conditions atmosphériques sont les mêmes et surtout tous ont été accidentés la nuit sans éclairage public.

```
Cluster: 3
      132      707      7454      929      169
0.1718305 0.1718305 0.1718305 0.1905932 0.1926423
```

8 Conclusion

L'étude avait pour objectif d'utiliser les méthodes descriptives vu en cours pour analyser un jeu de données de notre choix. Le jeu de données utilisé était celui des accidents corporels en France en 2021 et nous avons limité l'étude en région Parisienne.

Il était donc question de répondre à certaines problématiques liées à ces accidents. Durant cette étude, nous avons vu que la moyenne d'âge des individus accidentés était de 38,5 ans et que la plupart de ses accidents ces accidents avaient lieu sur le trajet promenade loisirs, généralement en journée et en majoritairement dans 4 départements (Paris, Seine Saint Denis, Haut Seine, Val de Marne).

Nous avons vu que les conditions d'éclairages influent sur la localisation de l'accident. Par exemple en agglomération le nombre d'individus accidentés étaient plus important en plein jour. Nous avons également vu que la gravité des accidents avaient une influence sur le type de circulation. Par exemple on a vu qu'il y'avait plus de victimes sur les voies bidirectionnelles et à sens unique, des victimes avec des blessures légères et blessures hospitalisées. On a vu aussi que la Vitesse avait une influence sur la catégorie des routes puisque ces deux variables étaient liées statistiquement. L'ACM et la classification nous ont permis de mieux comprendre nos données et de confirmer nos réponses.

Annexe1

Catégorie Usagers

	n	%	val%
Conducteur	8055	74.6	74.6
Passager	1914	17.7	17.7
Pieton	829	7.7	7.7

Circulation

	n	%	val%
A chaussées séparées	2333	21.6	21.6
A sens unique	3290	30.5	30.5
Avec voies d'affectation variable	78	0.7	0.7
Bidirectionnelle	5097	47.2	47.2

Trajet

	n	%	val%
Autre	1311	12.1	12.1
Courses – achats	487	4.5	4.5
Domicile – école	345	3.2	3.2
Domicile – travail	1992	18.4	18.4
Promenade – loisirs	5237	48.5	48.5
Utilisation professionnelle	1426	13.2	13.2

Voies de séparation (vosp)

	n	%	val%
Bande cyclable	171	1.6	1.6
Piste cyclable	810	7.5	7.5
Sans objet	9353	86.6	86.6
Voie réservée	464	4.3	4.3

Catégorie de route

	n	%	val%
Autoroute	1680	15.6	15.6
autre	29	0.3	0.3
Hors réseau public	5	0.0	0.0
Parc de stationnement ouvert à la circulation publique	1	0.0	0.0
Route départementale	3376	31.3	31.3
Route nationale	558	5.2	5.2
Voie communales	5149	47.7	47.7

Profil

	n	%	val%
Bas de cote	92	0.9	0.9
Pante	1310	12.1	12.1
Plat	9263	85.8	85.8
Sommet de cote	133	1.2	1.2

Plan

	n	%	val%
En courbe à droite	564	5.2	5.2
En Courbe à gauche	590	5.5	5.5
En S	35	0.3	0.3
Partie Rectiligne	9609	89.0	89.0

Surface

	n	%	val%
Autre	13	0.1	0.1
Boue	2	0.0	0.0
Corps gras huile	9	0.1	0.1
Enneigée	31	0.3	0.3
Flaques	17	0.2	0.2
Inondée	5	0.0	0.0
Mouillée	2364	21.9	21.9
Normale	8332	77.2	77.2
Verglacée	25	0.2	0.2

Situation

	n	%	val%
Autres	174	1.6	1.6
Sur accotement	141	1.3	1.3
Sur autre voie spéciale	171	1.6	1.6
Sur bande d'arrêt d'urgence	67	0.6	0.6
Sur chaussée	9667	89.5	89.5
Sur piste cyclable	360	3.3	3.3
Sur trottoir	218	2.0	2.0

Infrastructure

	n	%	val%
Aucun	9307	86.2	86.2
Autres	309	2.9	2.9
Bretelle d'échangeur ou de raccordement	117	1.1	1.1
Carrefour aménagé	429	4.0	4.0
Chantier	79	0.7	0.7
Pont autopont	169	1.6	1.6
Souterrain tunnel	262	2.4	2.4
Voie ferrée	20	0.2	0.2
Zone de péage	2	0.0	0.0
zone piétonne	104	1.0	1.0

Lumière (éclairage)

	n	%	val%
Aube	606	5.6	5.6
Nuit avec éclairage public allum	2032	18.8	18.8
Nuit avec éclairage public éteint	103	1.0	1.0
Nuit sans éclairage public	693	6.4	6.4
Plein jour	7364	68.2	68.2

Localisation (Agg)

	n	%	val%
En agglomération	8181	75.8	75.8
Hors agglomération	2617	24.2	24.2

Intersection

	n	%	val%
Autre intersection	474	4.4	4.4
Giratoire	203	1.9	1.9
Hors intersection	6196	57.4	57.4
Intersection à plus de 4 branches	146	1.4	1.4
Intersection en T	1321	12.2	12.2
Intersection en X	1773	16.4	16.4
Intersection en Y	366	3.4	3.4
Passage à niveau	10	0.1	0.1
Place	309	2.9	2.9

Condition atmosphérique

	n	%	val%
Brouillard_fumée	25	0.2	0.2
Neige_grêle	55	0.5	0.5
Normale	8333	77.2	77.2
Pluie forte	226	2.1	2.1
Pluie légère	1543	14.3	14.3
Temps couvert	507	4.7	4.7
Temps éblouissant	103	1.0	1.0
Vent fort_tempête	6	0.1	0.1

Collision

	n	%	val%
Autre collision	2976	27.6	27.6
Deux véhicules frontale	573	5.3	5.3
Deux véhicules par l'arrière	1836	17.0	17.0
Deux véhicules par le côté	3868	35.8	35.8
Sans collision	690	6.4	6.4
Trois véhicules et plus collisions multiples	314	2.9	2.9
Trois véhicules et plus en chaîne	541	5.0	5.0

ANNEXE 2

	Estimate	p.value
int=Hors intersection	4.568E-01	0.000E+00
agg=Hors_agglomeration	5.067E-01	0.000E+00
lum=Nuit sans éclairage public	6.499E-01	0.000E+00
circ=A chaussées séparées	5.840E-01	0.000E+00
catr=Autoroute	6.043E-01	0.000E+00
vosp=Sans objet	2.463E-01	1.785E-286
catr=Route nationale	3.731E-01	4.089E-254
col=Deux véhicules par l'arrière	1.682E-01	2.293E-197
col=Trois véhicules et plus en chaîne	4.206E-01	1.446E-166
infra=Souterrain tunnel	6.290E-01	1.803E-129
plan=En courbe à droite	1.597E-01	1.952E-78
surf=surf_Mouillée	1.043E-01	3.430E-76
atm=atm_Pluie légère	1.222E-01	4.538E-51
plan=En Courbe à gauche	8.805E-02	2.533E-50
situ=Sur chaussée	2.103E-01	7.689E-44
col=Trois véhicules et plus collisions multiples	2.347E-01	1.154E-42
infra=zonne piétonne	1.568E-01	1.117E-30
prof=Sommet de cote	2.184E-01	3.054E-27

	Estimate	p.value
col=Sans collision	3.960E-02	2.993E-23
atm=atm_Temps éblouissant	1.216E-01	2.757E-13
situ=Sur accotement	3.028E-01	2.106E-08
lum=Aube	-8.485E-02	4.355E-04
col=Autre collision	-1.699E-01	2.532E-04
infra=Autres	-2.267E-01	2.082E-06
atm=atm_Temps couvert	-1.562E-01	1.516E-06
prof=Pante	-6.533E-02	6.785E-08
circ=Avec voies d'affectation variable	-3.074E-01	6.679E-08
infra=Aucun	-1.072E-01	2.242E-11
situ=Sur trottoir	-7.771E-02	1.598E-15
col=Deux véhicules frontale	-3.097E-01	7.189E-17
lum=Plein_jour	-1.833E-01	1.026E-19
prof=Plat	-1.530E-01	2.739E-21
int=Intersection en Y	-9.650E-02	6.558E-33
atm=atm_Normale	-8.755E-02	3.331E-41
infra=Carrefour aménagé	-4.518E-01	7.672E-53
int=Autre intersection	-5.297E-03	2.624E-53

	Estimate	p.value
catr=Route departementale	-4.115E-01	1.134E-73
surf=surf_Normale	-1.043E-01	3.430E-76
lum=Nuit avec eclairage public allum	-3.818E-01	1.191E-119
situ=Sur piste cyclable	-4.354E-01	2.950E-133
plan=Partie Rectiligne	-2.478E-01	6.360E-135
int=Intersection en T	-1.726E-01	5.599E-202
vosp=Piste cyclable	-2.463E-01	1.785E-286
int=Intersection en X	-1.824E-01	7.048E-305
col=Deux véhicules par le coté	-3.835E-01	0.000E+00
agg=En_agglomeration	-5.067E-01	0.000E+00
circ=Bidirectionnelle	-2.687E-01	0.000E+00
catr=Voie comunales	-5.658E-01	0.000E+00