

# Étude de l'accidentologie en France en 2022

Iruomachi IRUOMAH, Emmanuel PAGUIEL, Yann BROCHET

## Sommaire

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Présentation du sujet . . . . .	2
1.2	Présentation de la base de données . . . . .	2
1.3	Quelques ajustements . . . . .	4
1.4	Statistiques . . . . .	5
1.5	Décisions . . . . .	6
<b>2</b>	<b>ACM</b>	<b>7</b>
2.1	Mise en place de l'ACM . . . . .	7
2.2	Étude de l'inertie . . . . .	7
2.3	Réalisation de l'ACM . . . . .	10
2.4	Points importants de cette ACM . . . . .	11
2.5	Variables supplémentaires . . . . .	12
<b>3</b>	<b>Classification</b>	<b>14</b>
3.1	Groupe 1 . . . . .	14
3.2	Groupe 2 . . . . .	15
3.3	Groupe 3 . . . . .	15
<b>4</b>	<b>Conclusion</b>	<b>16</b>
4.1	Perspectives . . . . .	16



# 1 Introduction

## 1.1 Présentation du sujet

L'objectif de cette étude sera plutôt global. Nous allons étudier **l'accidentologie en France** (en nous concentrant sur la ville de Tours). Le but est de repérer certaines **corrélations** entre les types d'individus et les accidents provoqués. Nous allons aussi essayer de voir à travers nos données si certains stéréotypes sont réellement fondés ou non (par exemple, sur la capacité de conduite en fonction du sexe).

Pour ce faire, nous allons commencer par présenter nos données, les filtrer pour sélectionner celles qui nous paraissent pertinentes, afin de nous lancer dans la réalisation d'une **ACM** (Analyse des Correspondances Multiples). Nous terminerons notre étude par une **classification**.

## 1.2 Présentation de la base de données

Pour la réalisation de cette étude, nous avons utilisé **4 bases de données** concernant les accidents en France en 2022, toutes liées entre elles. Nous avons tout d'abord des données sur les **caractéristiques de l'accident** (15 variables) :

Accident_Id	jour	mois	an	hrmn	lum	dep	com	agg	int	atm	col
202200000001	19	10	2022	16:15	1	26	26198	2	3	1	3
202200000002	20	10	2022	08:34	1	25	25204	2	3	1	3
202200000003	20	10	2022	17:15	1	22	22360	2	6	1	2
202200000004	20	10	2022	18:00	1	16	16102	2	3	8	6
202200000005	19	10	2022	11:45	1	13	13103	1	1	1	2
202200000006	6	10	2022	14:00	1	13	13056	2	2	1	3

Nous avons comme première variable **l'identifiant de l'accident**. Les 4 bases de données possèdent cette variable.

Nous avons ensuite des variables sur l'adresse, le moment de l'accident, les conditions météorologiques ainsi que le lieu précis de l'accident donné grâce aux coordonnées (longitudes et latitudes), qui nous serviront par la suite.

Comme on peut l'observer, les variables prennent des chiffres comme modalités ; elles correspondent toutes à une situation (dont nous avons leur signification sur le site data.gouv). Pour rendre la lecture et la compréhension de nos données **plus lisibles**, nous avons donc décidé de **transformer** les modalités afin qu'elles soient plus **explicites**.

Num_Acc	jour	mois	an	hrmn	lum	dep	com
202200000001	Milieu du mois	Octobre	2022	16:15	Plein jour	26	26198
202200000002	Milieu du mois	Octobre	2022	08:34	Plein jour	25	25204
202200000003	Milieu du mois	Octobre	2022	17:15	Plein jour	22	22360
202200000004	Milieu du mois	Octobre	2022	18:00	Plein jour	16	16102
202200000005	Milieu du mois	Octobre	2022	11:45	Plein jour	13	13103
202200000006	Début du mois	Octobre	2022	14:00	Plein jour	13	13056

Voici notre deuxième base de données concernant le **lieu de l'accident** (18 variables) :

Num_Acc	catr	voie	v1	v2
202200000001	Voie Communales	TEIL(vieille route du)	0	N/A
202200000002	Voie Communales	N/A	0	N/A
202200000003	Route Départementale	ROND POINT DE BREZILLET	0	N/A
202200000004	Voie Communales	QUATORZE JUILLET (RUE DU)	0	N/A
202200000005	Route Départementale	ROUTE DE JEAN MOULIN-RN 538	0	N/A
202200000006	Route Départementale	TURCAN FRANCIS	0	N/A

On peut observer ici des variables typiquement liées au lieu de l'accident. Certaines colonnes restent inutilisables car nous n'avons pas la signification de certaines variables ou bien il y a des valeurs manquantes. Nous allons donc procéder à un **nettoyage de ces données**.

Voici notre troisième base de données concernant **les usagers**, acteurs de l'accident (16 variables) :

Num_Acc	id_usager	id_vehicule	num_veh	place	catu	grav	sexe
202200000001	1 099 700	813 952	A01	1	Conducteur	Blessé hospitalisé	Masculin
202200000001	1 099 701	813 953	B01	1	Conducteur	Indemne	Masculin
202200000002	1 099 698	813 950	B01	1	Conducteur	Blessé léger	Masculin
202200000002	1 099 699	813 951	A01	1	Conducteur	Indemne	Masculin
202200000003	1 099 696	813 948	A01	1	Conducteur	Indemne	Masculin
202200000003	1 099 697	813 949	B01	1	Conducteur	Blessé léger	Féminin

Ici, nous avons principalement des données sur le type d'usagers (année de naissance, sexe). Nous allons principalement les utiliser pour notre étude.

Voici notre dernière base de données concernant les **véhicules** utilisés lors d'un accident (11 variables) :

Num_Acc	id_vehicule	num_veh	senc
202200000001	813 952	A01	PK ou PR ou numéro d'adresse postale croissant
202200000001	813 953	B01	PK ou PR ou numéro d'adresse postale croissant
202200000002	813 950	B01	PK ou PR ou numéro d'adresse postale décroissant
202200000002	813 951	A01	PK ou PR ou numéro d'adresse postale décroissant
202200000003	813 948	A01	PK ou PR ou numéro d'adresse postale décroissant
202200000003	813 949	B01	PK ou PR ou numéro d'adresse postale croissant

Nous utiliserons moins ces données, sauf pour une variable. Cependant, nous avons ici des informations sur les types de véhicules, le carburant, la manœuvre lors de l'accident, etc.

Ensuite, nous avons décidé de **concaténer nos 4 bases de données** par un inner join en prenant en référence la variable ID\_accident (Num\_Acc) :

```
df2 <- dfcarac |>
  inner_join(df lieux, by = "Num_Acc") |>
  inner_join(df vehicules, by = "Num_Acc") |>
  inner_join(df usagers, by = c("Num_Acc", "id_vehicule"))
```

Voici donc notre base de données comportant **48 variables** et **114 076 observations** avant le nettoyage des lignes avec des valeurs manquantes et des variables inutilisables :

Num_Acc	jour	mois	an	hrmn	lum	dep	com	agg
202200000001	Milieu du mois	Octobre	2022	16:15	Plein jour	26	26198	En agglomération
202200000001	Milieu du mois	Octobre	2022	16:15	Plein jour	26	26198	En agglomération
202200000003	Milieu du mois	Octobre	2022	17:15	Plein jour	22	22360	En agglomération
202200000003	Milieu du mois	Octobre	2022	17:15	Plein jour	22	22360	En agglomération

Ces 114 076 observations correspondent au **nombre d'usagers** impliqués dans les accidents en 2022.

### 1.3 Quelques ajustements

Pour notre étude, une information cruciale sur nos individus nous manque, **leur âge**.

Mais nous avons repéré une variable qui va nous permettre d'obtenir l'âge de nos individus. Comme nous avons l'année de naissance de nos individus, nous allons faire **2024 – l'année de naissance** pour chacun d'eux.

```
mutate(age = 2024 - as.numeric(df2$an_nais))
```

Par la suite, nous avons créé **5 classes d'âge** :

- **Les mineurs**
- **De 18 à 25 ans**, des jeunes adultes
- **De 26 à 38 ans**, des adultes avec un peu plus d'expérience et des responsabilités
- **De 39 à 55 ans**, des adultes avec beaucoup d'expérience
- **Au-dessus de 55 ans**, les séniors.

```
mutate(age_class = case_when(
  age >= 0 & age <= 17 ~ "Mineurs",
  age >= 18 & age <= 25 ~ "18-25",
  age >= 26 & age <= 38 ~ "26-38",
  age >= 39 & age <= 55 ~ "39-55",
  age > 55 ~ "Séniors"))
```

Voici un exemple du rendu :

age	age_class
16	Mineurs
76	Séniors

## 1.4 Statistiques

Avant de commencer notre **ACM**, place à quelques petites **statistiques** intéressantes.

Tout d'abord, l'une de nos bases était structurée de manière à ce que chaque ligne corresponde à un accident unique. Pour récupérer le nombre d'accidents, nous avons donc fait :

```
sum(nrow(dfcarac))
```

Nombre d'accidents en 2022
55 302

Par la suite, nous avons trouvé intéressant de calculer le **taux de mortalité**. Nous avons pu le faire grâce à la variable **grav**, qui prenait comme modalité « tué ».

```
nb_morts <- sum(dfusagers$grav == "Tué")  
round((nb_morts/nrow(dfusagers)*100),2)
```

taux de mortalité
2,8%

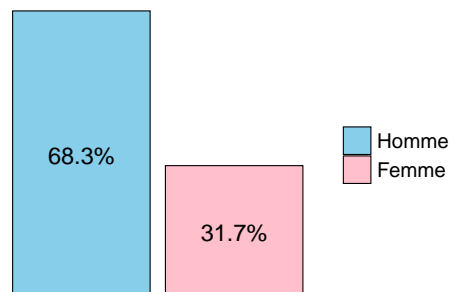
Puis, nous avons trouvé pertinent de calculer l'**âge moyen** de nos individus pour le comparer avec celui de la France, qui est de 42 ans et demi.

```
mean(df5$age)
```

Âge moyen de nos individus
40 ans et demi

Nos données se rapprochent donc de la **réalité**.

Pour finir, regardons la **répartition homme/femme** de notre étude.



On peut déjà observer que, statistiquement, dans cette étude, les hommes sont environ **deux fois** plus impliqués dans des accidents que les femmes, ce qui peut contredire le préjugé selon lequel les femmes conduisent moins bien que les hommes.

## 1.5 Décisions

Maintenant, comme nous étudions à l'**université de Tours**, nous trouvons judicieux d'organiser notre **ACM** sur les accidents de Tours et **sa périphérie**.

Pour ce faire, on a récupéré les données de **longitude** et de **latitude** de Tours, et on a travaillé avec les données de nos observations pour limiter notre dataframe à des observations provenant uniquement de chez nous.

```
lat_min <- 47
lat_max <- 47.7
long_min <- 0.3
long_max <- 1.3

tours_data <- df5[df5$lat >= lat_min & df5$lat <= lat_max &
                  df5$long >= long_min & df5$long <= long_max, ]
```

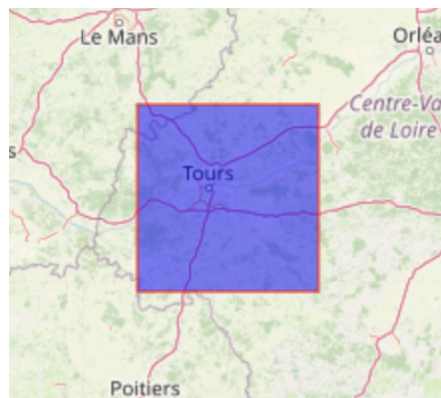


Figure 1: Périmètre de l'étude

Au minimum nous avons **943 observations** (si l'on réduit le périmètre, on ne capte plus aucune observation).

## 2 ACM

### 2.1 Mise en place de l'ACM

Tout d'abord, nous avons dû faire un choix de variables pour une **ACM claire** qui nous donne des informations **pertinentes**.

Nous sommes donc partis sur des variables qui identifient le **type d'individu**, ainsi que quelques **caractéristiques de l'accident**.

#### 2.1.1 Variables prises

On a sélectionné :

- la variable **choc** qui nous dit de quel côté a eu le choc qui a causé l'accident
- la variable **grav** qui nous dit la gravité de l'état de santé de l'individu (blessé, indemne, tué)
- la variable **catu** qui nous donne l'information sur la place de l'individu au moment de l'accident (conducteur, passager, piéton)
- la variable **sexe** qui nous dit si l'individu était une femme ou un homme
- la variable **age\_class** que l'on a créée pour avoir l'information de l'âge de l'individu au moment de l'accident

#### 2.1.2 Variables supplémentaires

- la variable **vma** qui nous donne la vitesse maximale autorisée au moment de l'accident
- la variable **trajet** qui nous donne le type de sortie de l'individu
- la variable **situ**, qui nous donne la situation de l'accident (sur un trottoir, la chaussée, etc). On aurait pu logiquement mettre cette variable dans notre ACM, c'est ce que nous avons fait au départ, mais nous avons eu des résultats peu convaincants, c'est pourquoi nous l'avons mise en supplémentaire pour garder l'information.

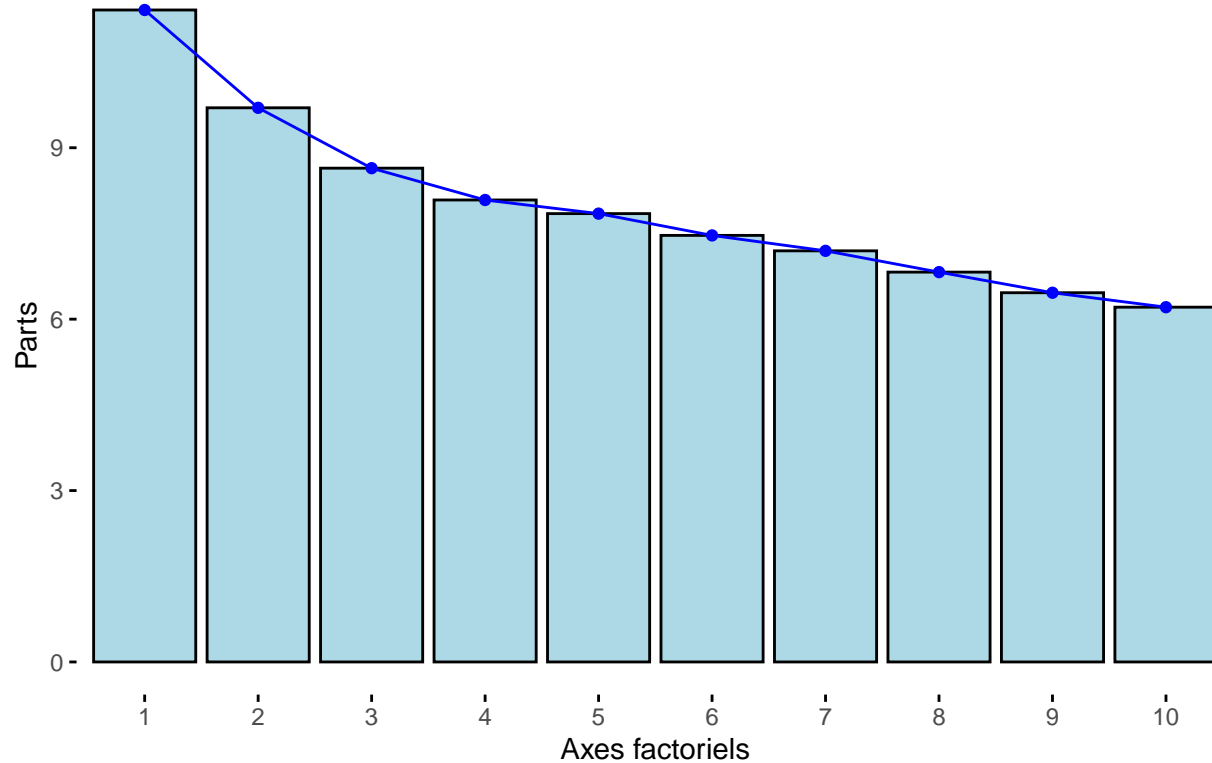
```
resMCA<-MCA(tours_data, quali.sup = c(1,2,7), graph=FALSE)
```

## 2.2 Étude de l'inertie

	Valeur propre	Part d'inertie	Part cumulée
dim 1	0.32	11.41	11.41
dim 2	0.27	9.70	21.11
dim 3	0.24	8.64	29.75
dim 4	0.23	8.09	37.84
dim 5	0.22	7.85	45.68
dim 6	0.21	7.47	53.15
dim 7	0.20	7.20	60.34
dim 8	0.19	6.82	67.17

Il serait intéressant de regarder jusqu'à la **4ème dimension**, qui explique **37,84 %** de la variance, mais nous nous concentrerons principalement sur les **deux premiers axes** pour notre étude.

## Inerties relatives de chaque axe



Nous voyons ici qu'il serait en réalité intéressant de regarder jusqu'au **3ème axe**, car au-delà, l'inertie cumulée décroît de moins en moins.

	Dim 1	Dim 2
Arrière	0.74	18.60
Aucun	0.57	0.00
Avant	0.06	1.22
Chocs multiples (tonneaux)	0.11	7.68
Côté	1.04	1.38
Conducteur	10.50	0.07
Passager	19.64	3.76
Piéton	7.79	14.23
Blessé hospitalisé	0.71	20.30
Blessé léger	7.94	0.88
Indemne	3.40	17.02
Tué	0.85	2.42
Féminin	12.07	0.24
Masculin	6.41	0.13
18-25	0.00	2.33
26-38	0.78	1.71
39-55	2.85	3.14
Mineurs	24.49	0.88
Séniors	0.06	4.04



### 2.2.1 Dimension 1

Cet axe est principalement construit par les modalités “**Conducteur/Passager**”, mais aussi par les modalités “**Masculin/Féminin**” ainsi que “**Mineurs**”.

### 2.2.2 Dimension 2

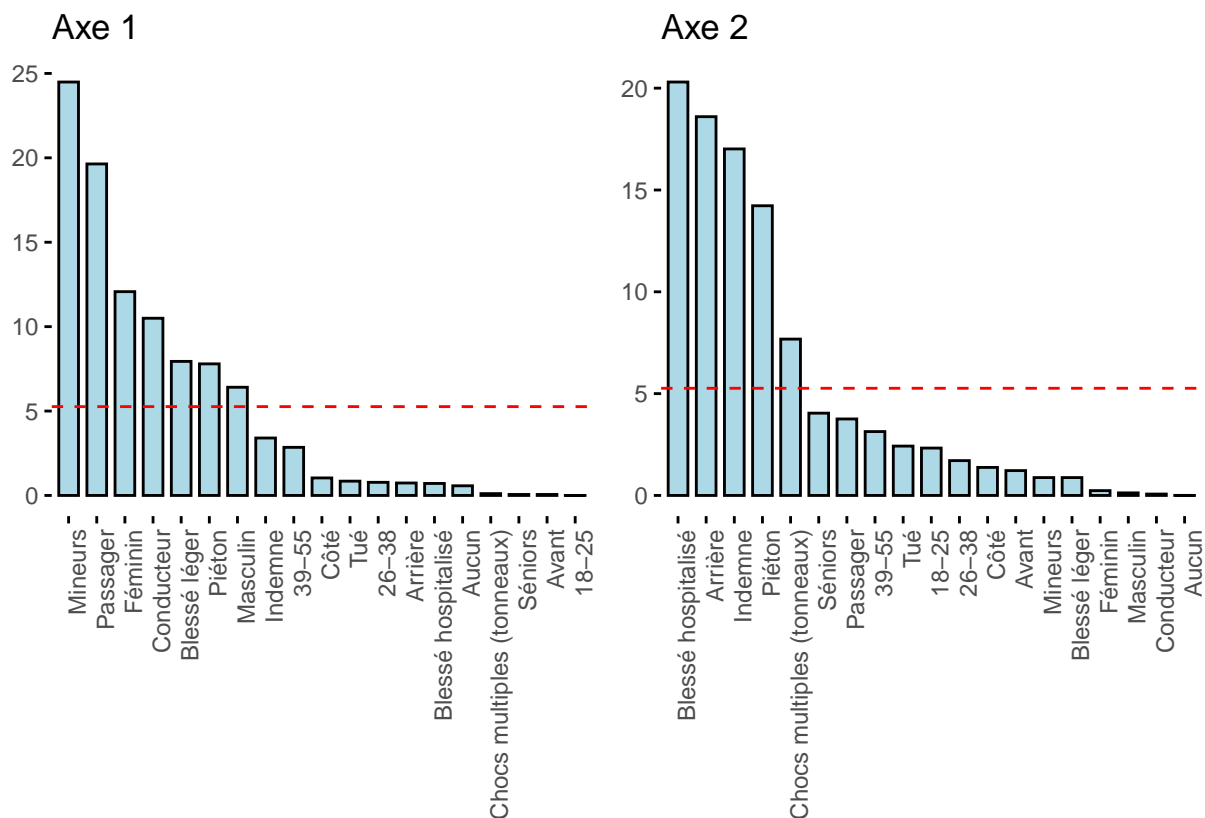
Cet axe est plutôt construit par les modalités “**Arrière**”, “**Blessé hospitalisé**”, “**Indemne**”, “**Chocs multiples**” (tonneaux) et “**Piéton**”.

### 2.2.3 Interprétation

Nous comprenons donc que l’axe 1 est en partie construit par **l’identité de l’individu** au moment de l’accident, c’est-à-dire son sexe et sa place lors de l’accident.

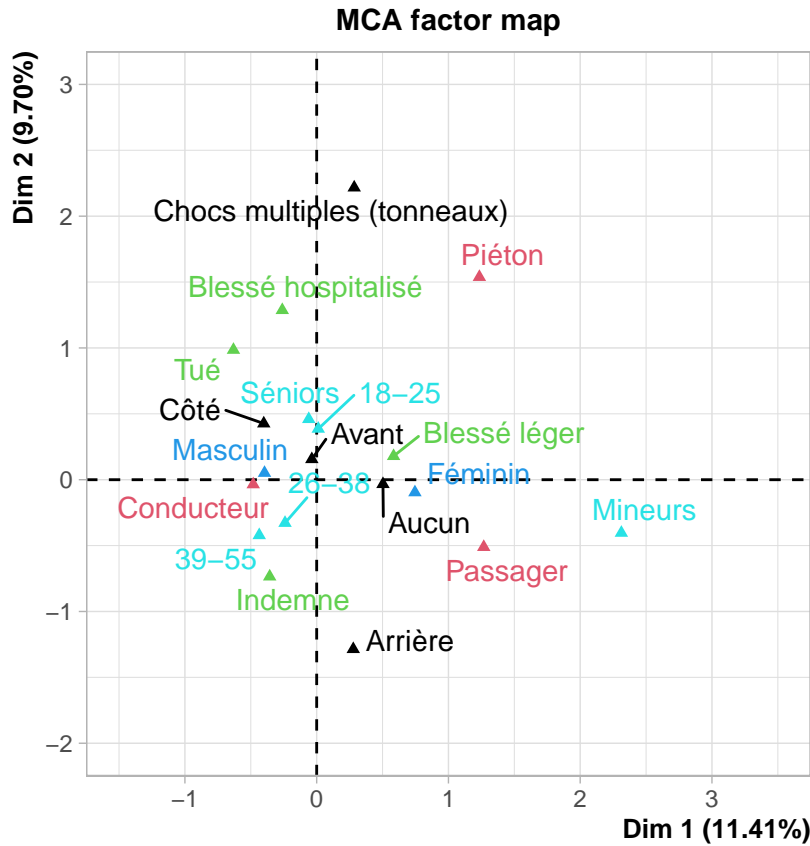
En revanche, l’axe 2 est plutôt construit par **les caractéristiques de l’accident**, telles que le type de choc et la gravité de l’accident (état des individus). Cependant, on retrouve la variable Piéton. On peut se demander pourquoi cette variable n’apparaît pas dans la construction de l’axe 1. Piéton contribue à l’axe 2 sans doute parce que cet axe capture les caractéristiques de l’accident, telles que la gravité des blessures, qui varient de manière significative pour les piétons par rapport aux conducteurs et passagers.

### 2.2.4 Tableau des contributions



Ces tableaux viennent confirmer notre interprétation.

## 2.3 Réalisation de l'ACM



### 2.3.1 Dimension 1

Comme nous l'avons remarqué auparavant, cette dimension semble être liée à l'**identité de l'individu** et à sa **position** au moment de l'accident.

- À gauche, nous trouvons les modalités “**Conducteur**” et “**Masculin**”. Cela signifie que cette partie de l'axe est associée aux conducteurs, souvent des hommes adultes.
- À droite, nous trouvons les modalités “**Passager**”, “**Mineurs**” et “**Féminin**”, ce qui pourrait indiquer une association avec les passagers plutôt jeunes ou féminins.

### 2.3.2 Dimension 2

Cet axe là reflète plutôt les **caractéristiques** liées à la **gravité des accidents**.

- En haut, nous retrouvons les modalités “**Blessé hospitalisé**”, “**Tué**”, “**Piéton**” et “**Chocs multiples (tonneaux)**”. Cela semble traduire une gravité élevée des accidents concernant les piétons et les accidents impliquant des tonneaux (chocs multiples).
- En bas, nous retrouvons des modalités comme “**Indemne**” et “**Arrière**”, traduisant des accidents moins graves. Les chocs à l'arrière d'un véhicule sont souvent moins impressionnants que ceux sur les côtés.

### 2.3.3 En haut à droite

Cette zone est marquée par les modalités **“Piéton”** et **“Blessé hospitalisé”**, indiquant une association entre les piétons et les accidents graves.

### 2.3.4 En bas à droite

Nous retrouvons ici un regroupement des modalités **“Passager”**, **“Mineurs”** et **“Arrière”**, ce qui montre que les passagers, souvent jeunes, sont moins exposés à des chocs graves.

### 2.3.5 En haut à gauche

Cette zone est associée aux accidents graves impliquant probablement des conducteurs, car nous retrouvons les modalités **“Tué”**, **“Blessé hospitalisé”** et **“Chocs multiples”**.

### 2.3.6 En bas à gauche

On trouve des modalités telles que **“Conducteur”**, **“Masculin”**, **“Indemne”** et les classes d’âge **“26-38”** et **“39-55”**. Cela pourrait indiquer que les conducteurs adultes masculins sont souvent indemnes ou moins touchés par des chocs graves (probablement car ils ont des responsabilités, de l’expérience et ils sont plus prudents).

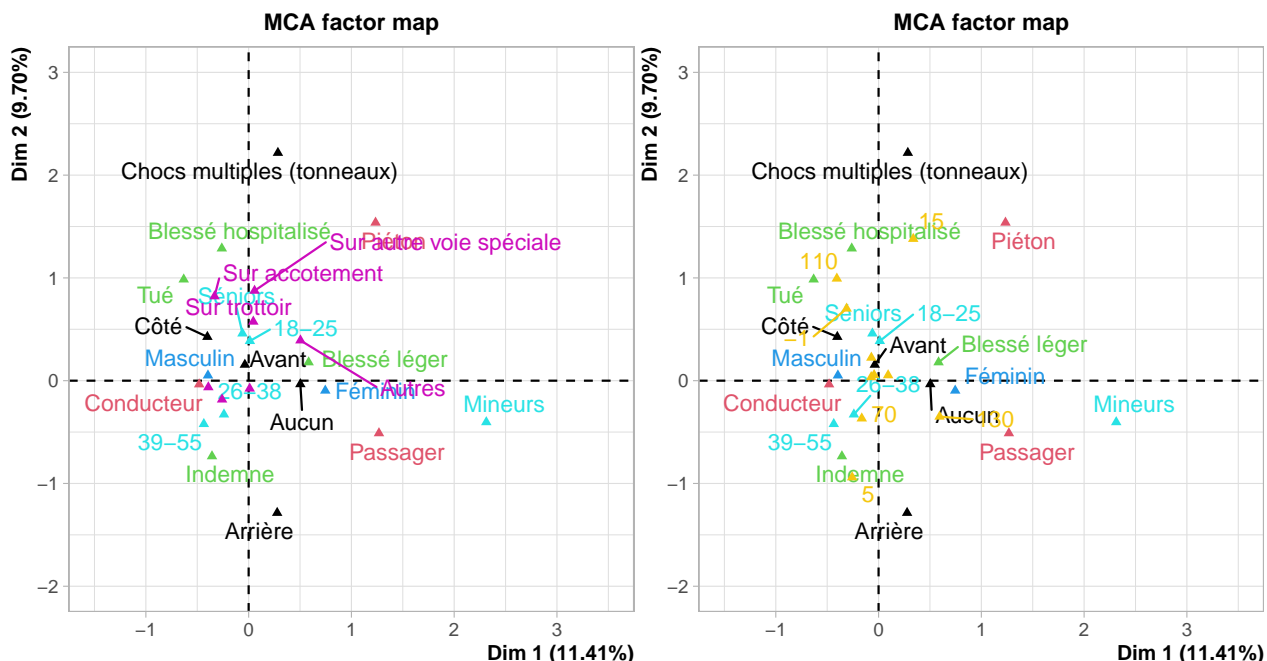
## 2.4 Points importants de cette ACM

- **L’axe 1** oppose principalement les conducteurs masculins adultes aux passagers, notamment jeunes ou féminins.
- **L’axe 2** met en opposition les caractéristiques de gravité des accidents : en haut les accidents graves (blessures, décès, piétons), en bas les accidents moins graves (indemnes).
- La modalité **“Piéton”** contribue fortement à **l’axe 2**, probablement en raison de la gravité des blessures spécifiques aux piétons.

Notre analyse nous permet ainsi de mieux comprendre la **répartition des individus** en fonction des **caractéristiques** des accidents.

## 2.5 Variables supplémentaires

Notre analyse des **variables supplémentaires** s'est principalement concentrée sur leur positionnement par rapport aux axes et sur les **informations complémentaires** qu'elles pouvaient apporter à nos variables principales.



### 2.5.1 situ

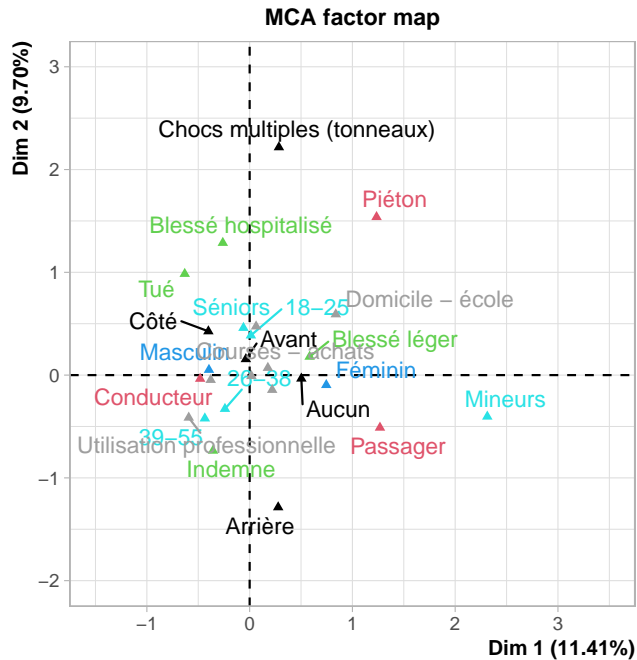
- Les modalités supplémentaires “**Sur accotement**”, “**Sur autre voie spéciale**”, et “**Sur trottoir**” renforcent l’interprétation précédente selon laquelle l’**axe 2** est associé à la gravité des accidents. Leur présence dans les zones impliquant des piétons et des blessures graves montre que les accidents en dehors des voies principales ou sur des trottoirs sont particulièrement dangereux pour les piétons.

### 2.5.2 vma

Les valeurs 5, 15, 70, 110, 130 représentent la **vitesse maximale autorisée** (en km/h) au moment de l’accident.

- VMA = 5**, située en bas près de l’axe central, correspond à des accidents survenus à très faible vitesse, comme en zone de rencontre. Cette faible vitesse est logiquement associée à la modalité « Indemne », car les chances de sortir sans blessures dans ce type de situations sont très élevées.
- VMA = 70**, Les routes à 70 km/h, souvent secondaires, sont associées à des accidents frontaux impliquant des conducteurs masculins adultes (26-38 ans), ce qui pourrait refléter des comportements de conduite plus risqués.
- VMA = 110**, positionnées en haut à gauche proches des modalités « Blessé hospitalisé » et « Tué ». Cela indique que des accidents à des vitesses élevées sont fortement associés à des conséquences graves, comme des blessures sévères ou des décès.
- VMA = 130**, bien que l’on pourrait s’attendre à la voir proche des modalités « Blessé hospitalisé » et « Tué », cela peut s’expliquer par des facteurs tels que la sécurité accrue et la meilleure visibilité sur autoroute, réduisant la fréquence des accidents graves. Ce qui est intéressant, c’est sa proximité avec la modalité « Aucun », suggérant que ces accidents pourraient être liés à un problème technique du véhicule ou une perte de contrôle plutôt qu’à un choc violent.

### 2.5.3 trajet

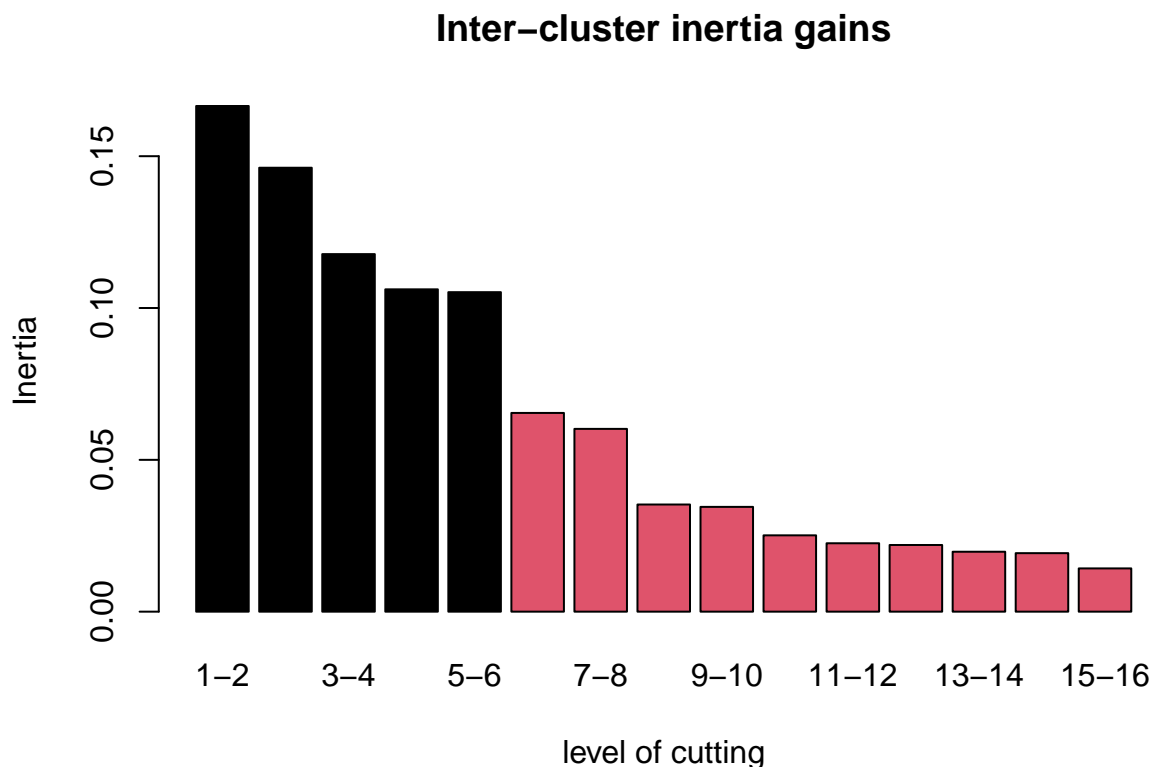


- **Domicile-école** : sa position centrale sur la carte factorielle montre qu'elle n'est pas fortement associée à des conséquences graves ou à un type d'accident particulier. Cela signifie que les accidents ayant lieu lors de **trajets domicile-école** sont souvent **dispersés** dans toutes les situations : ils peuvent être bénins ou graves, sans tendance marquée. Ce type de trajet concerne souvent des **déplacements courts**, souvent en zone urbaine, ce qui peut expliquer l'absence de spécificité.
- **Utilisation professionnelle** : cette modalité est proche des modalités qui décrivent une personne en âge moyen de travail comme on peut le voir avec les modalités « Conducteur », « 26-38 » et « 39-55 », en raison d'un **trajets régulier** et **maîtrisé** du fait que le chemin emprunté doit être le même on va retrouver une absence en générale d'accident grave. Des blessures légères plutôt que des conséquences graves dont la raison qu'on y retrouve la modalité « **Indemne** ».
- **Courses-achats** : cette modalité se situe à proximité des modalités telles que « **Blessé léger** » et « **Sénior** », ce qui suggère que les accidents survenant lors de trajets pour courses et achats sont souvent moins graves dû à l'utilisation courte du véhicule, souvent en zone urbaine et non loin de « **Sénior** » du fait que la personne arrivée à cette âge-là non **plus grand-chose à faire** et vont se retrouver régulièrement à utiliser leur véhicule pour leur course.

### 3 Classification

```
mca.clus <- HCPC(resMCA, graph = FALSE, consol = TRUE)
```

Pour notre **classification**, regardons le graphique de gain d'inertie **inter-groupe** afin de déterminer le nombre de cluster optimal pour notre étude :



D'après le graphique, nous pourrions envisager **7 à 8 clusters**, car le gain d'inertie inter-groupe devient faible au-delà. Cependant, après plusieurs essais, les résultats étaient plus pertinents avec **6 clusters**. Nous avons donc choisi cette option.

Nous avons identifié **3 clusters** particulièrement intéressants :

#### 3.1 Groupe 1

	Cla/Mod	Mod/Cla	Global	p.value	v.test
age_class=26-38	76.35	47.55	21.53	0	13.88
catu=Conducteur	45.45	95.09	72.32	0	12.42
age_class=39-55	67.26	46.63	23.97	0	11.61
grav=Indemne	46.65	57.67	42.74	0	6.72
trajet=Utilisation professionnelle	57.29	16.87	10.18	0	4.78
trajet=Domicile – travail	52.63	21.47	14.10	0	4.60
choc=Avant	38.89	77.30	68.72	0	4.18
choc=Arrière	50.00	22.09	15.27	0	4.13

- Environ **2 tiers** des individus de **26 à 55 ans** se retrouvent dans ce groupe, et **95%** des individus de ce groupe sont des conducteurs.
- De plus, **45%** des individus indemnes après un accident sont aussi dans ce cluster (environ la moitié).

- **57%** des individus qui ont un accident lors de trajets professionnels se retrouvent aussi dans ce cluster.
- Pour finir, nous retrouvons les accidents provoqués par des chocs à l'avant ou à l'arrière (qui se retrouvent en grande partie dans ce cluster).

Le **groupe 1** est donc composé principalement de conducteurs adultes âgés entre **26 et 55 ans**, engagés dans des trajets professionnels ou domicile-travail. Les accidents sont peu graves, avec une majorité d'individus indemnes après l'accident. Les types de chocs avant et arrière sont les plus fréquents, ce qui pourrait être lié à des contextes de **trafic professionnel ou d'embouteillages** (fréquent dans les trajets domicile-travail).

Ce groupe représente probablement des accidents de la vie courante, impliquant des conducteurs réguliers dans des situations de routine.

### 3.2 Groupe 2

	Cla/Mod	Mod/Cla	Global	p.value	v.test
<b>age_class=Séniors</b>	74.69	83.56	25.98	0.00	21.42
<b>catu=Conducteur</b>	30.50	94.98	72.32	0.00	9.53
<b>grav=Tué</b>	81.25	11.87	3.39	0.00	7.01
<b>grav=Blessé hospitalisé</b>	35.03	25.11	16.65	0.00	3.69
<b>trajet=Courses – achats</b>	42.42	6.39	3.50	0.01	2.47

- **74%** des seniors se retrouve dans ce groupe et **83%** des individus de ce groupe sont des seniors
- **95%** des individus de ce groupe sont des conducteurs
- **81%** des individus tués se retrouvent dans ce groupe mais peu d'individus de ce groupe sont réellement tués (12%), car cette modalité est plutôt rare. On retrouve une petite part aussi des individus blessés hospitalisés.
- **42%** des individus qui ont subi un accident sur leur trajet pour aller faire des courses se retrouvent dans ce groupe.

Ce groupe est fortement associé aux **seniors**, qui représentent une part très importante des individus. Les accidents dans ce groupe sont particulièrement graves, avec une **forte sur-représentation des décès** (tués) et des blessures hospitalisées. Cela suggère que lorsqu'un accident implique un senior, il est souvent plus grave en termes de conséquences.

Les trajets typiques dans ce groupe sont liés aux **courses et achats**, ce qui pourrait indiquer que ces accidents se produisent dans des contextes de déplacement de **proximité**, typiques des seniors.

### 3.3 Groupe 3

	Cla/Mod	Mod/Cla	Global	p.value	v.test
<b>catu=Passager</b>	67.93	93.98	19.51	0.00	21.07
<b>age_class=Mineurs</b>	72.46	37.59	7.32	0.00	11.66
<b>sexe=Féminin</b>	27.83	68.42	34.68	0.00	8.56
<b>vma=130</b>	43.16	30.83	10.07	0.00	7.39
<b>trajet=Promenade – loisirs</b>	21.66	51.13	33.30	0.00	4.57
<b>choc=Arrière</b>	27.08	29.32	15.27	0.00	4.49
<b>grav=Blessé léger</b>	19.66	51.88	37.22	0.00	3.70
<b>choc=Avant</b>	12.35	60.15	68.72	0.02	-2.26

- **2 tiers** des individus qui sont passagers lors d'un accident se retrouvent dans ce groupe, et **93%** des individus de ce groupe sont des passagers.
- De plus, **72%** des mineurs sont dans ce groupe et environ 30% des femmes aussi (cependant, environ **70%** des individus de ce cluster sont des femmes).
- Nous pouvons aussi observer que **43%** des accidents sur autoroute se retrouvent dans ce groupe.
- Nous retrouvons les modalités, "**Blessé léger**", "**choc avant**" et "**choc arrière**".

Nous pouvons donc en déduire que ce groupe est constitué de passagers **mineurs**, avec une **sur-représentation des femmes**. Les accidents surviennent souvent :

- Lors de trajets de loisirs ou de promenades,
- Sur des autoroutes (VMA 130 km/h), ce qui peut s'expliquer par des **trajets familiaux longs**. Les chocs arrières sont caractéristiques dans ce groupe, ce qui peut indiquer une conduite en file ou un contexte de ralentissement, typique des **autoroutes**. Les accidents dans ce groupe ont une gravité relativement modérée, avec une proportion notable de blessés légers.

## 4 Conclusion

L'objectif de notre étude était de mieux comprendre l'**accidentologie en France**, en particulier dans la région de Tours, tout en cherchant à vérifier si certains stéréotypes liés à la conduite sont fondés ou non. À travers l'analyse des données et la réalisation d'une **ACM** (Analyse des Correspondances Multiples) suivie d'une **classification**, nous avons pu identifier des **corrélations significatives** entre les caractéristiques des individus et les accidents dans lesquels ils sont impliqués.

Nos principaux résultats sont les suivants :

- Les stéréotypes concernant les femmes au volant sont **infondés** : Contrairement aux idées reçues, ce sont les hommes qui sont plus souvent impliqués dans des accidents, qu'ils soient conducteurs ou victimes.
- Le stéréotype des "personnes âgées" impliquées dans des accidents est **partiellement confirmé** : Les individus âgés sont en effet sur-représentés dans certains types d'accidents, et les conséquences pour ces derniers tendent à être plus graves (hospitalisations, décès).
- La répartition des rôles "**hommes conducteurs**" et "**femmes passagères**" est **avérée** : Les données montrent une présence marquée des hommes au volant tandis que les femmes apparaissent plus fréquemment en tant que passagères, notamment dans des contextes familiaux ou de trajets de loisirs.

En outre, nous avons relevé que :

- Les accidents graves (hospitalisations et décès) surviennent souvent chez des groupes bien identifiés comme les **seniors** ou lors de trajets à haute vitesse.
- Les **mineurs** sont sur-représentés dans des accidents impliquant des **passagers**, souvent sur des trajets de loisirs.

### 4.1 Perspectives

Cette étude a permis de mettre en lumière certaines dynamiques importantes de l'accidentologie. Elle ouvre également la voie à des réflexions sur les politiques de **sécurité routière**, en insistant sur des populations plus vulnérables comme les **seniors** et les **mineurs**.