

Analyse des accidents de la circulation : tendances et facteurs déterminants (2019-2023)

LAURIE BEHLOUL - RIEULLE BRUSQ -
BAPTISTE DELAUNAY - JOANNA GOSSE



Plan



I Contexte et objectifs

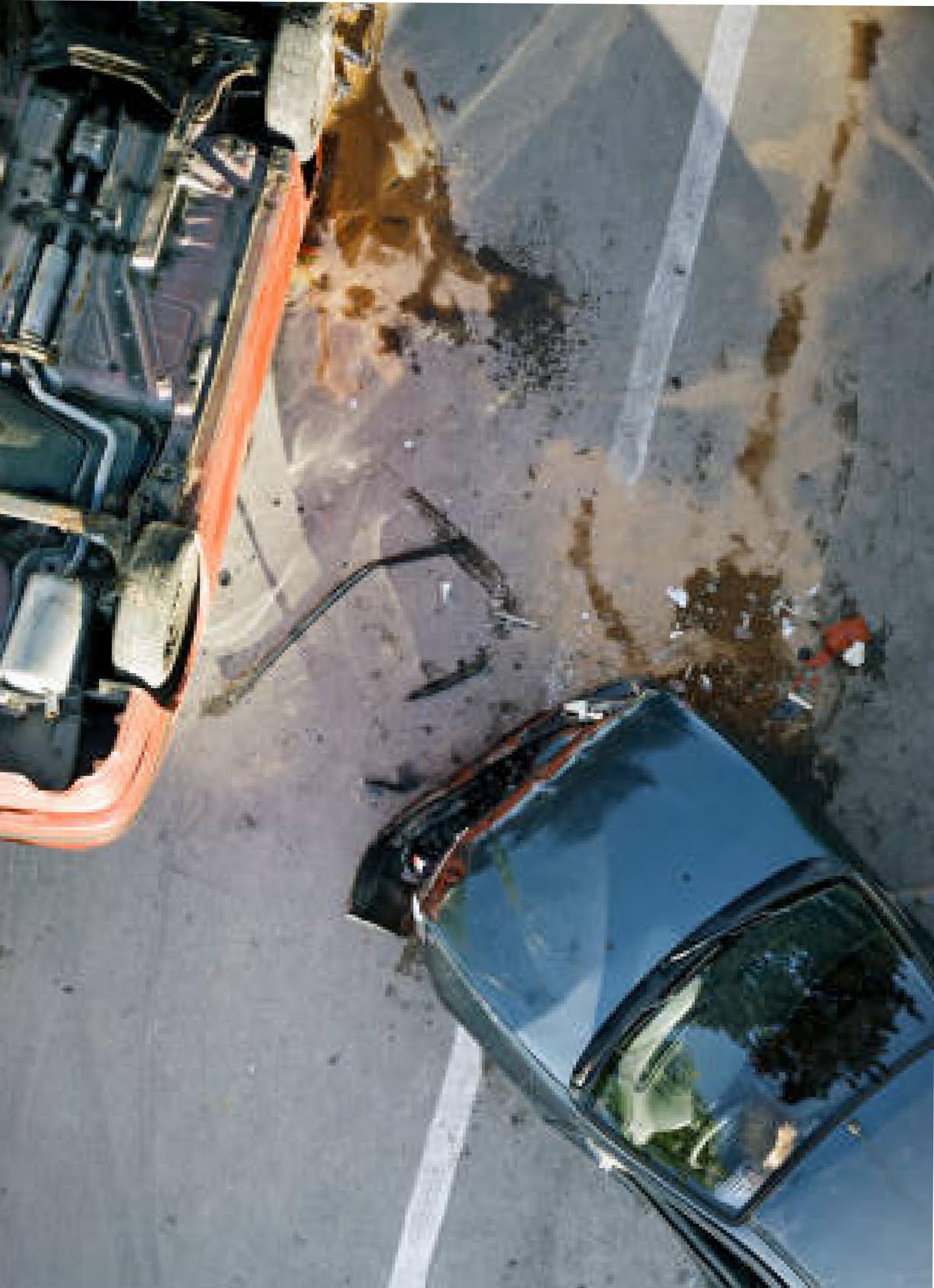
IV Prédiction avec Random Forest

II Présentation des données
et analyse descriptive

V Recommandations et
perspectives

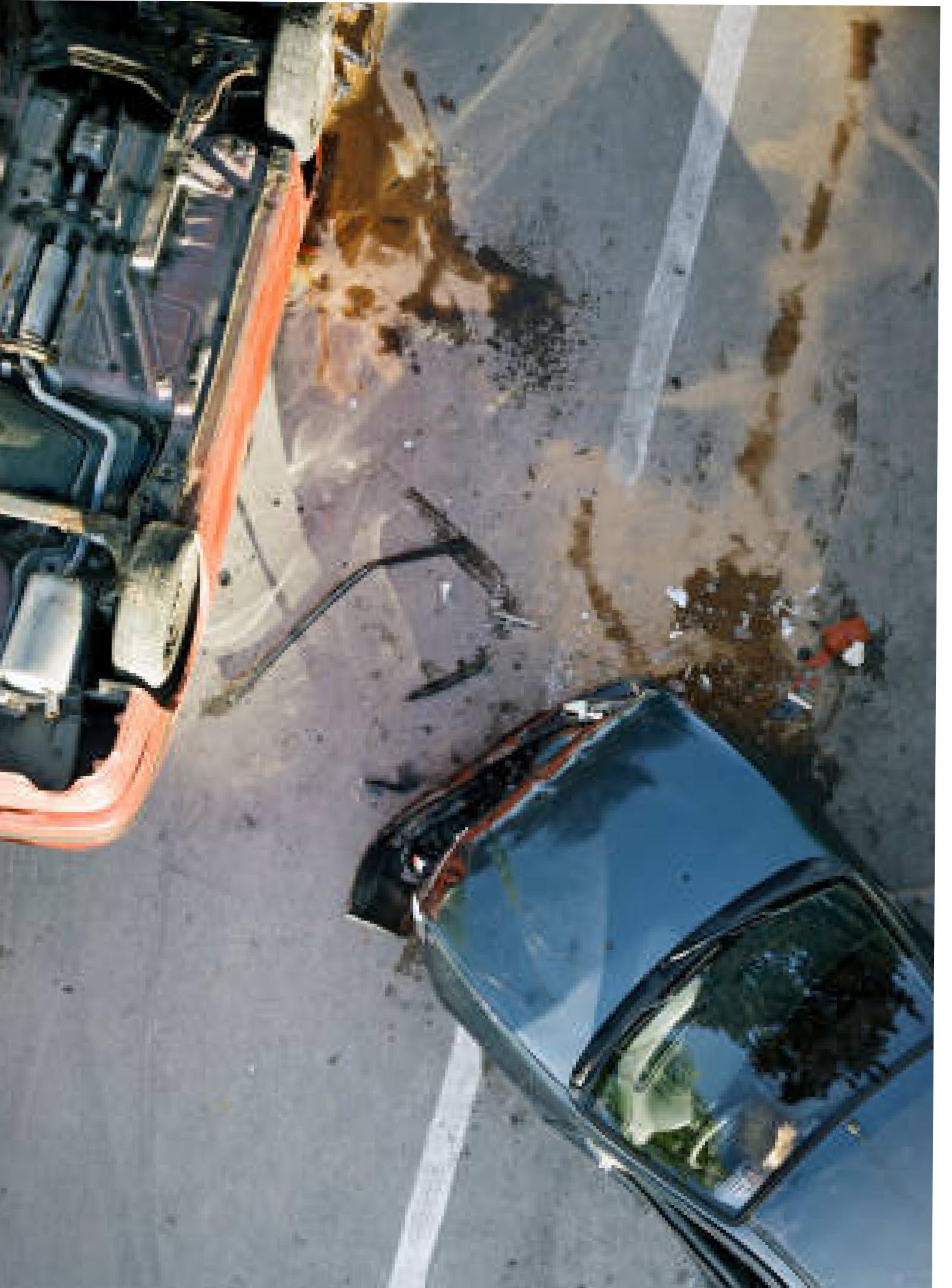
III Analyse statistique et
modélisation

VI Conclusion



Contexte

- ▶ La sécurité routière : un enjeu majeur
- ▶ 3 398 personnes décédées en 2023 pour 54 822 accidents recensés
- ▶ Exploitation du fichier BAAC
(Bulletin d'Analyse des Accidents Corporels)



Objectifs

- ▶ Décrire et visualiser les tendances temporelles et spatiales
- ▶ Identifier les corrélations entre les variables explicatives
- ▶ Utiliser des techniques statistiques et cartographiques pour proposer des analyses précises
- ▶ Déterminer la gravité d'un accident selon les caractéristiques observées

Présentation des données

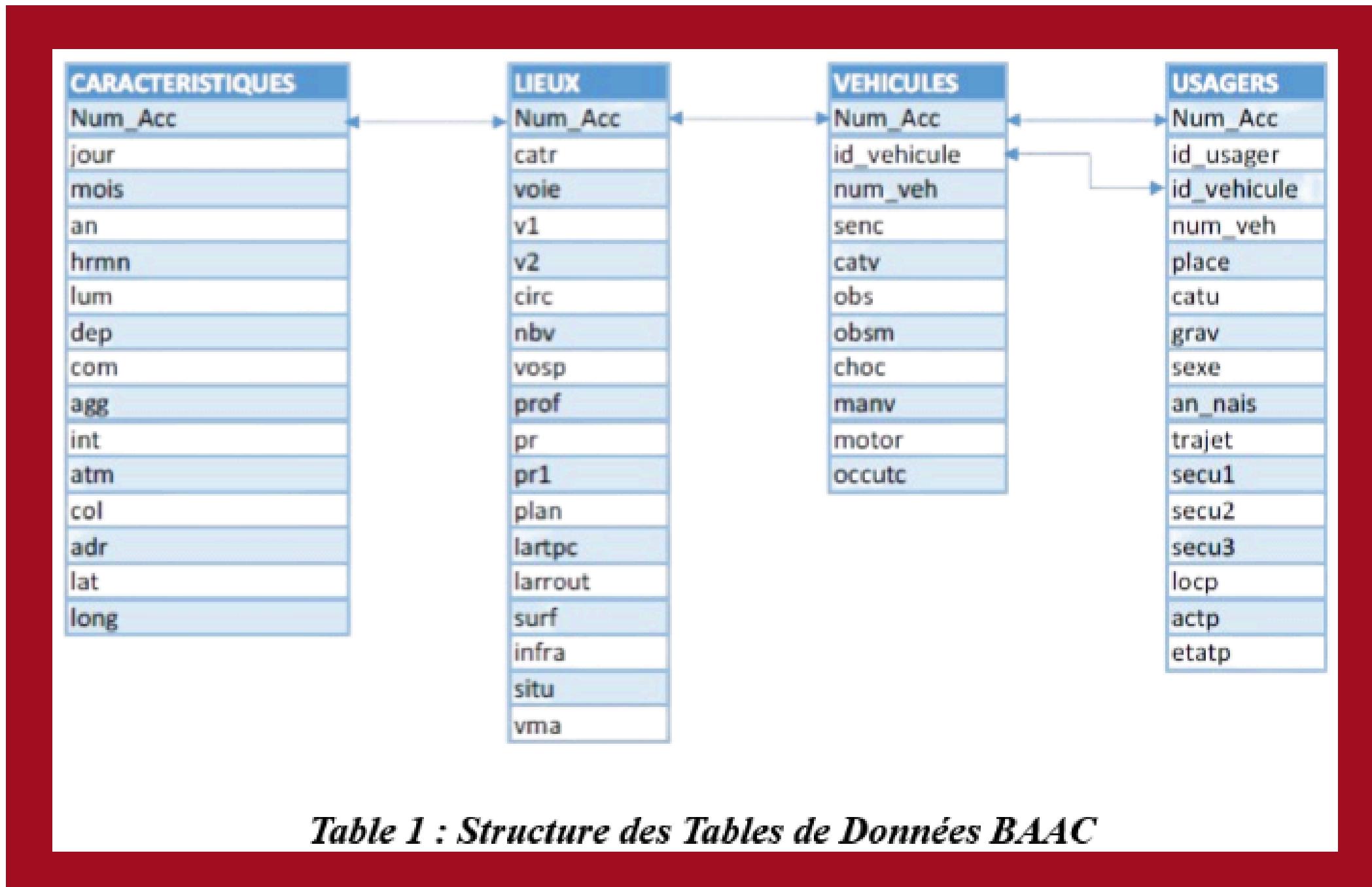
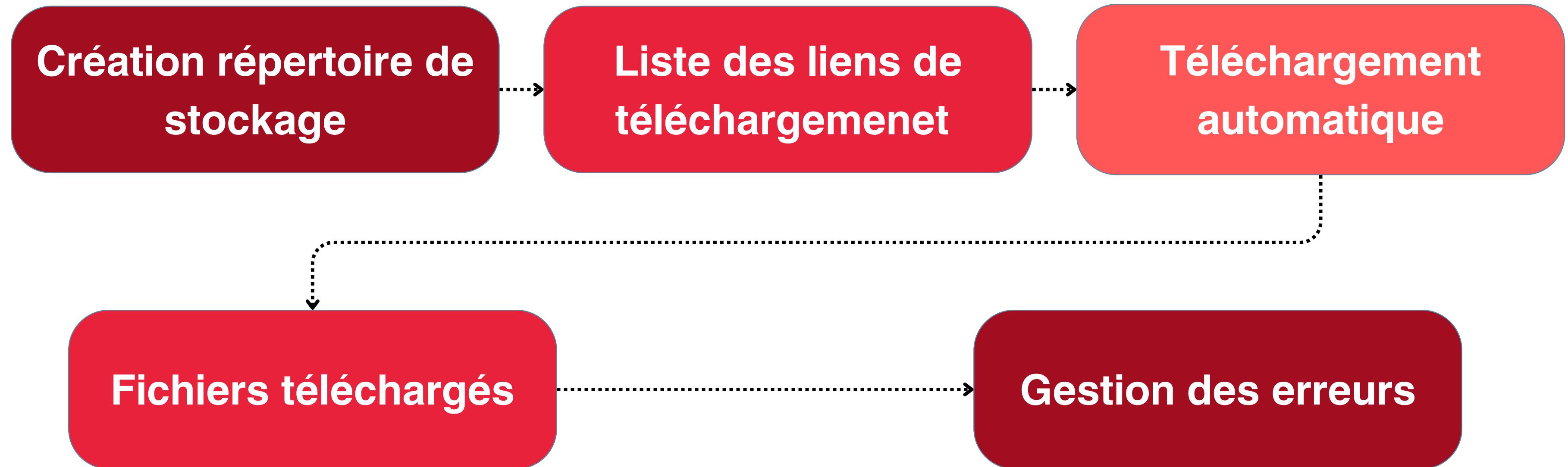


Table 1 : Structure des Tables de Données BAAC

Num_Acc : identifiant unique pour chaque accident

Collecte des Données

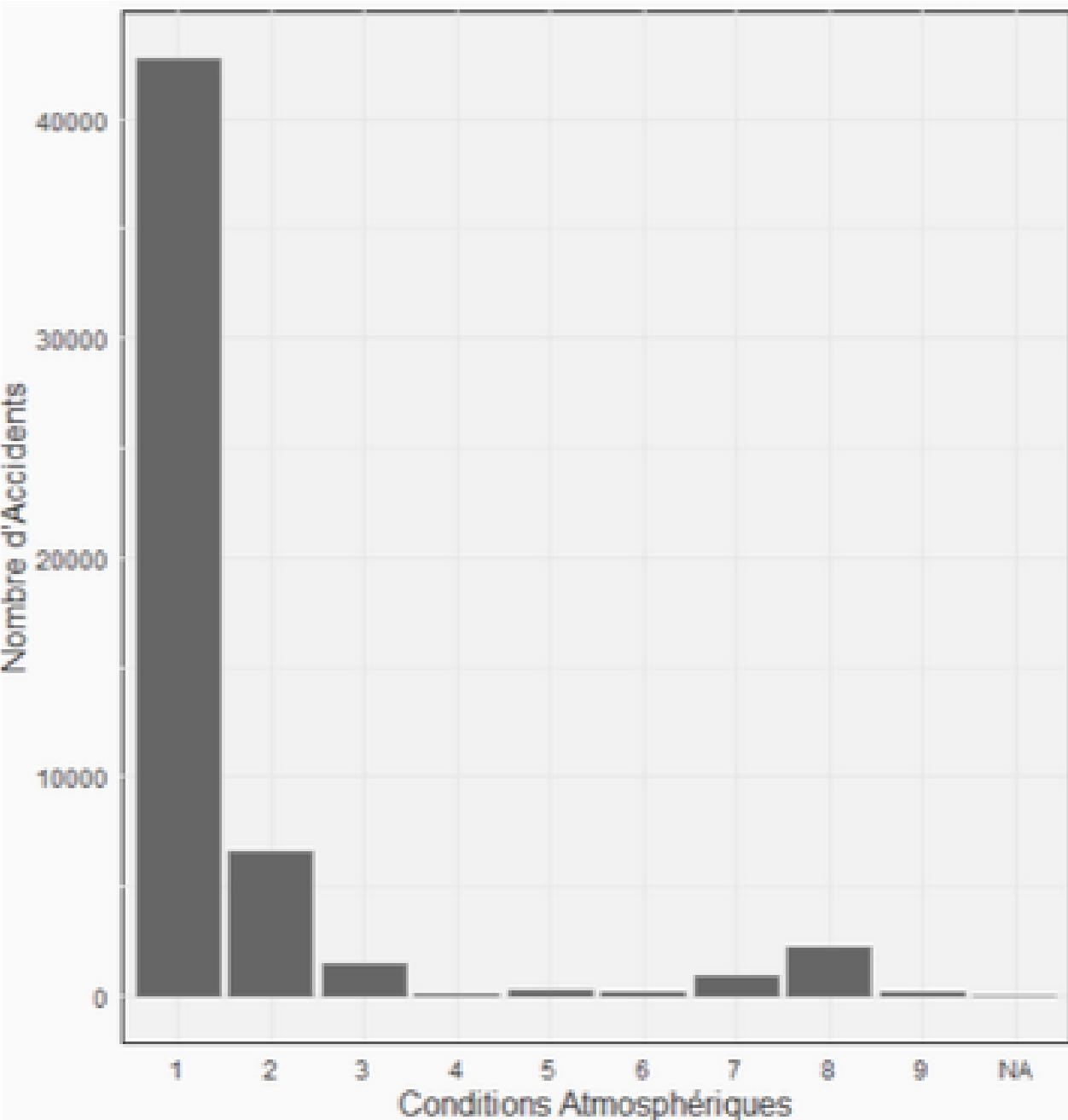
Fichiers issus du site data.gouv.fr : plateforme ouverte des données publiques françaises



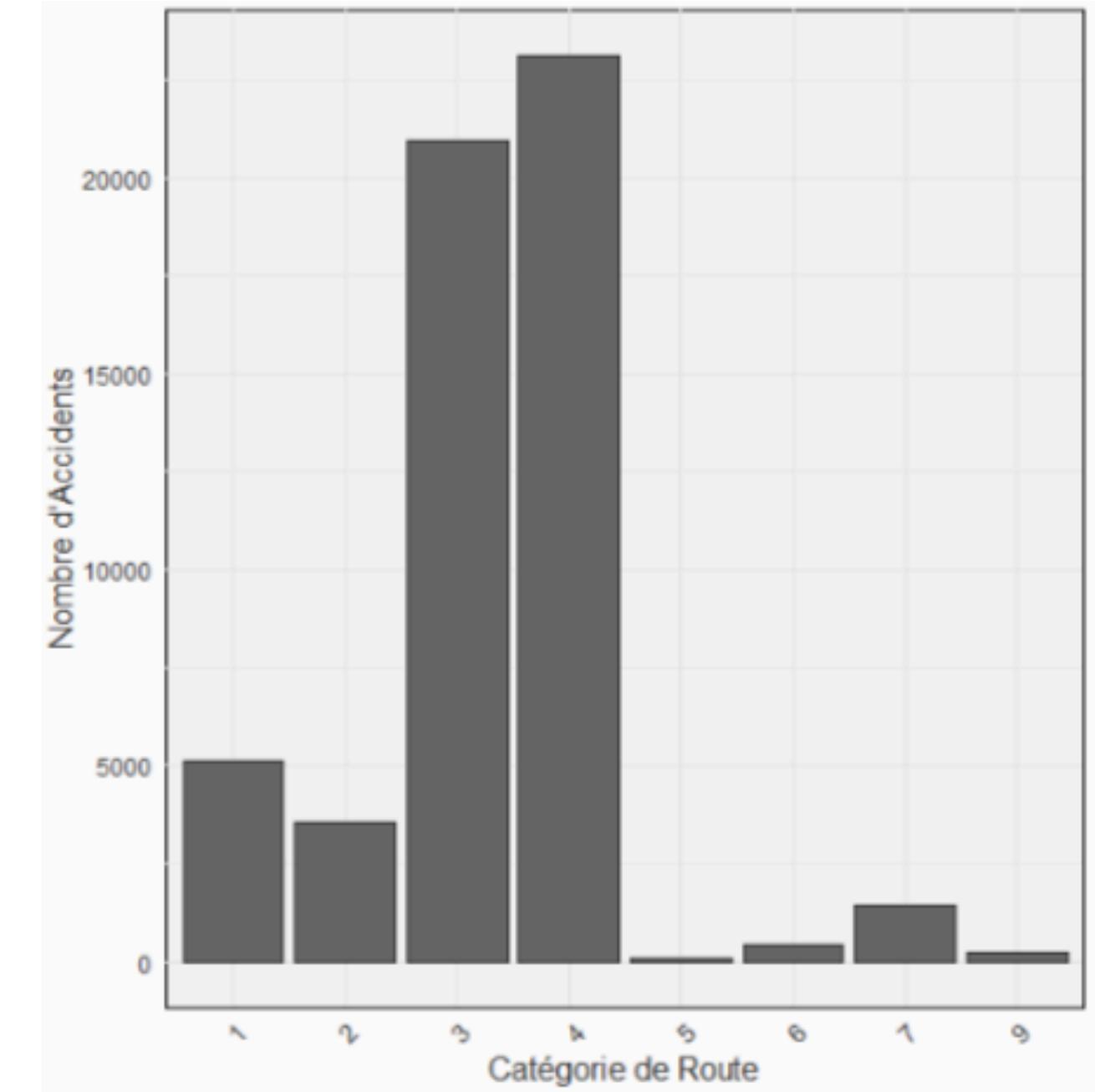
Analyse descriptive

Effectué sur
les données de
2023

fichier recensant tous
les accidents, avec
nombre
de mort, nombre de
véhicules et
d'usagers
impliqués



Nombre d'accidents par
catégorie de conditions
atmosphériques

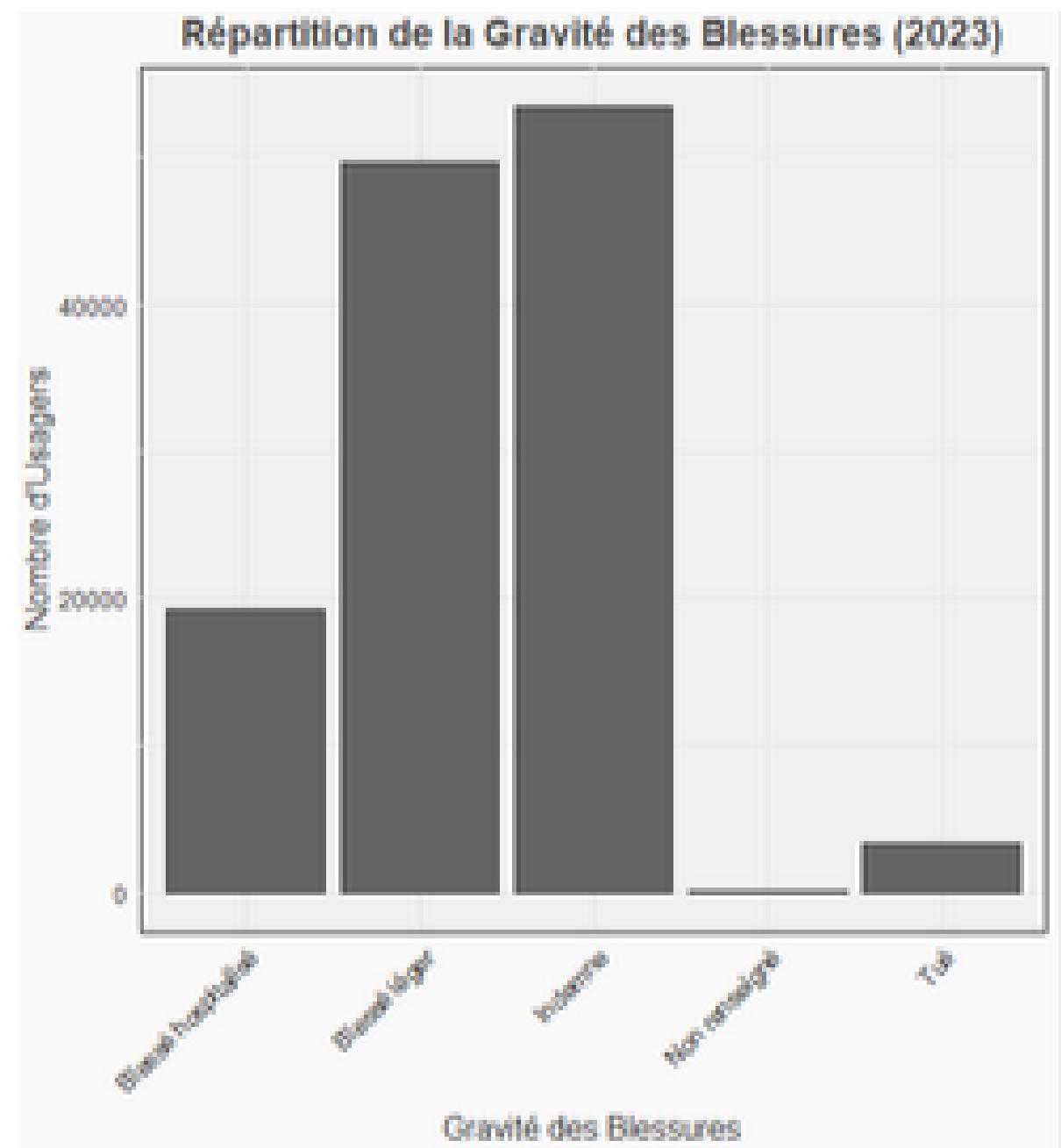


Nombre d'accidents par
type de route

En 2023,

- majorité des accidents ont lieu sous de bonnes conditions atmosphériques
- majorité des accidents ont lieu sur des routes départementales et communales
- majorité des accidentés ressortent légèrement blessés ou indemne de l'accident

Analyse descriptive

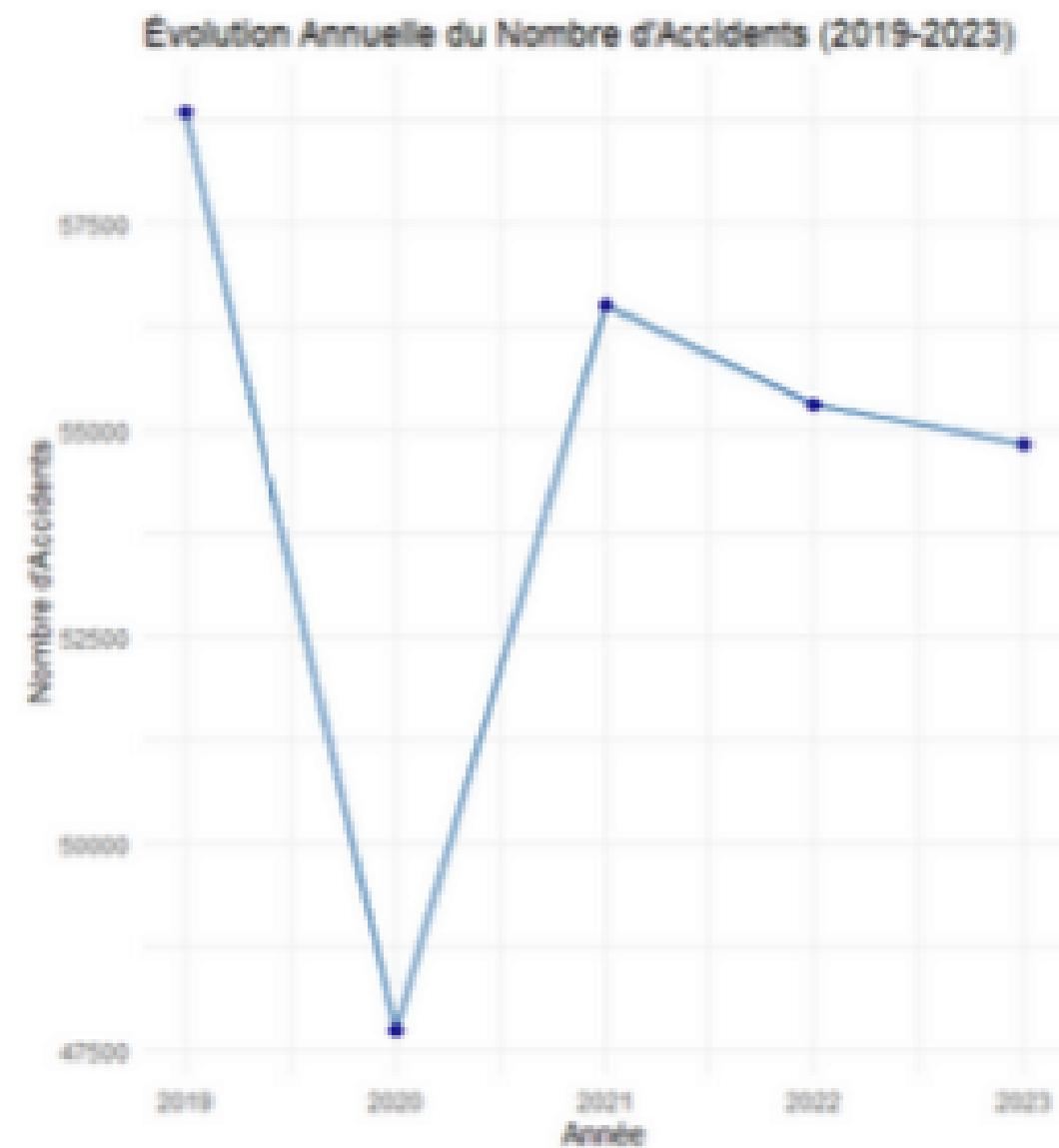


Gravité des blessures

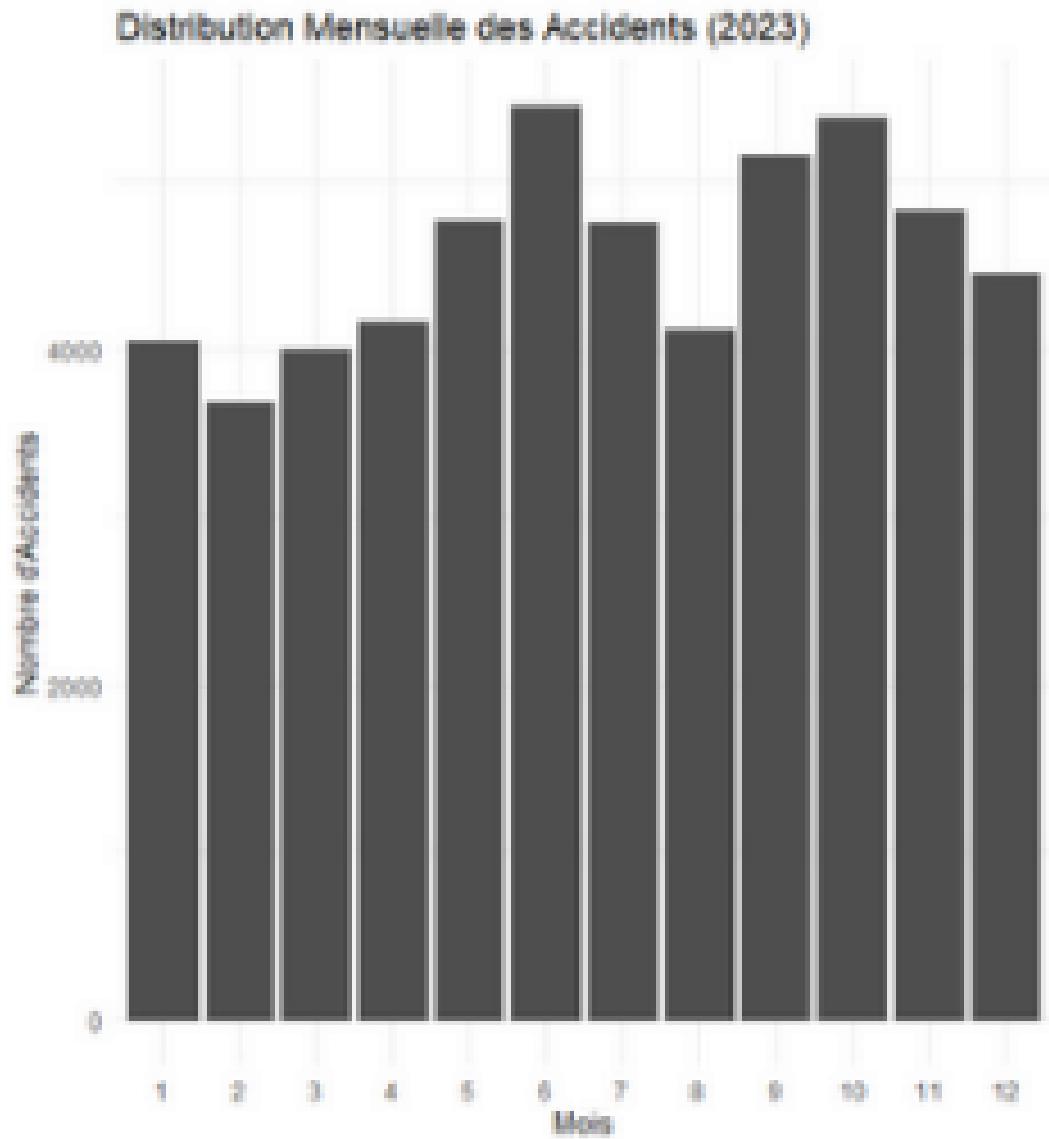
Analyse descriptive

Tendances annuelles entre 2019 et 2023

Distribution mensuelle sur l'année 2023



Nombre d'accident entre 2019 et 2023

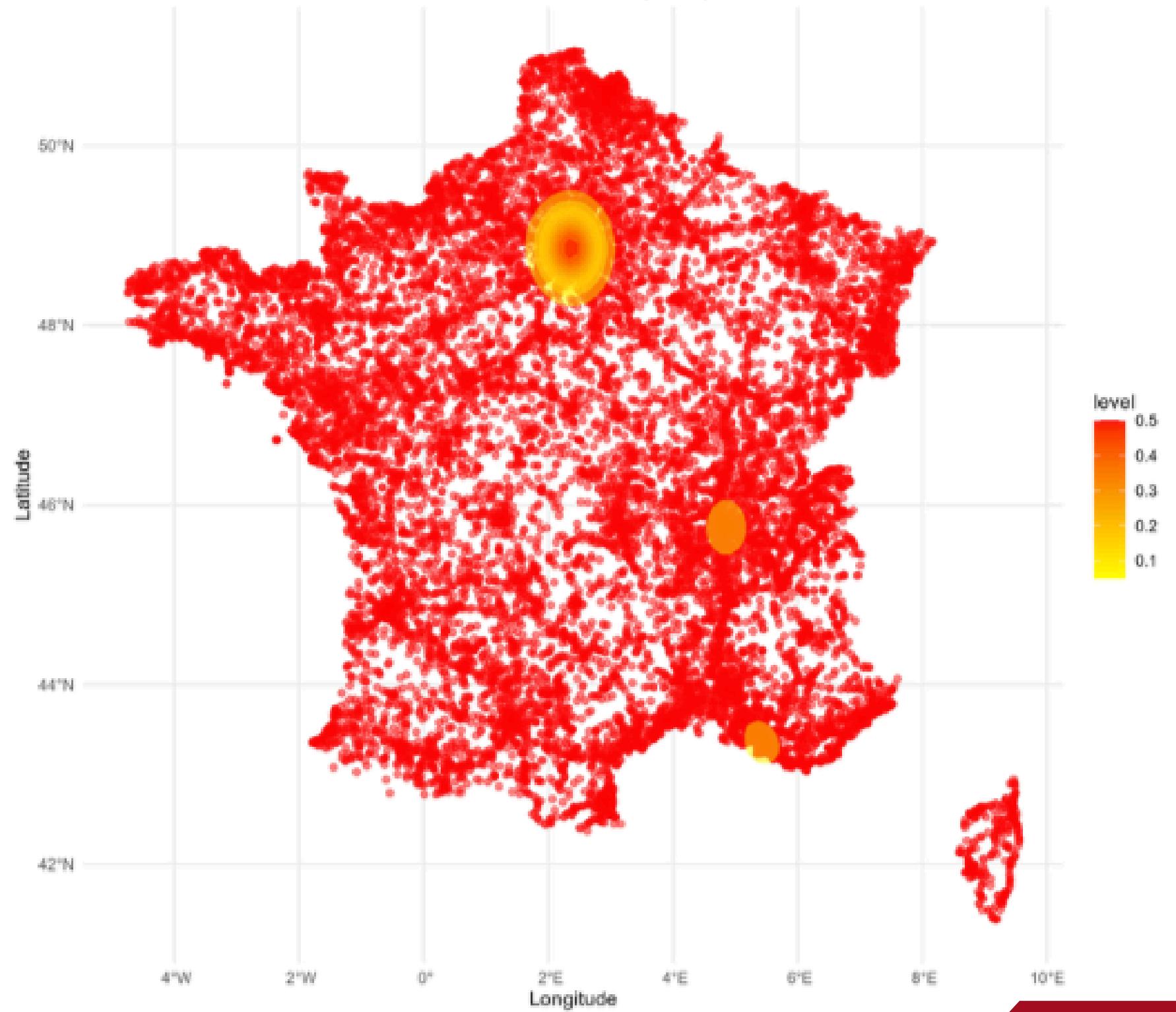


Nombre d'accidents par mois en 2023

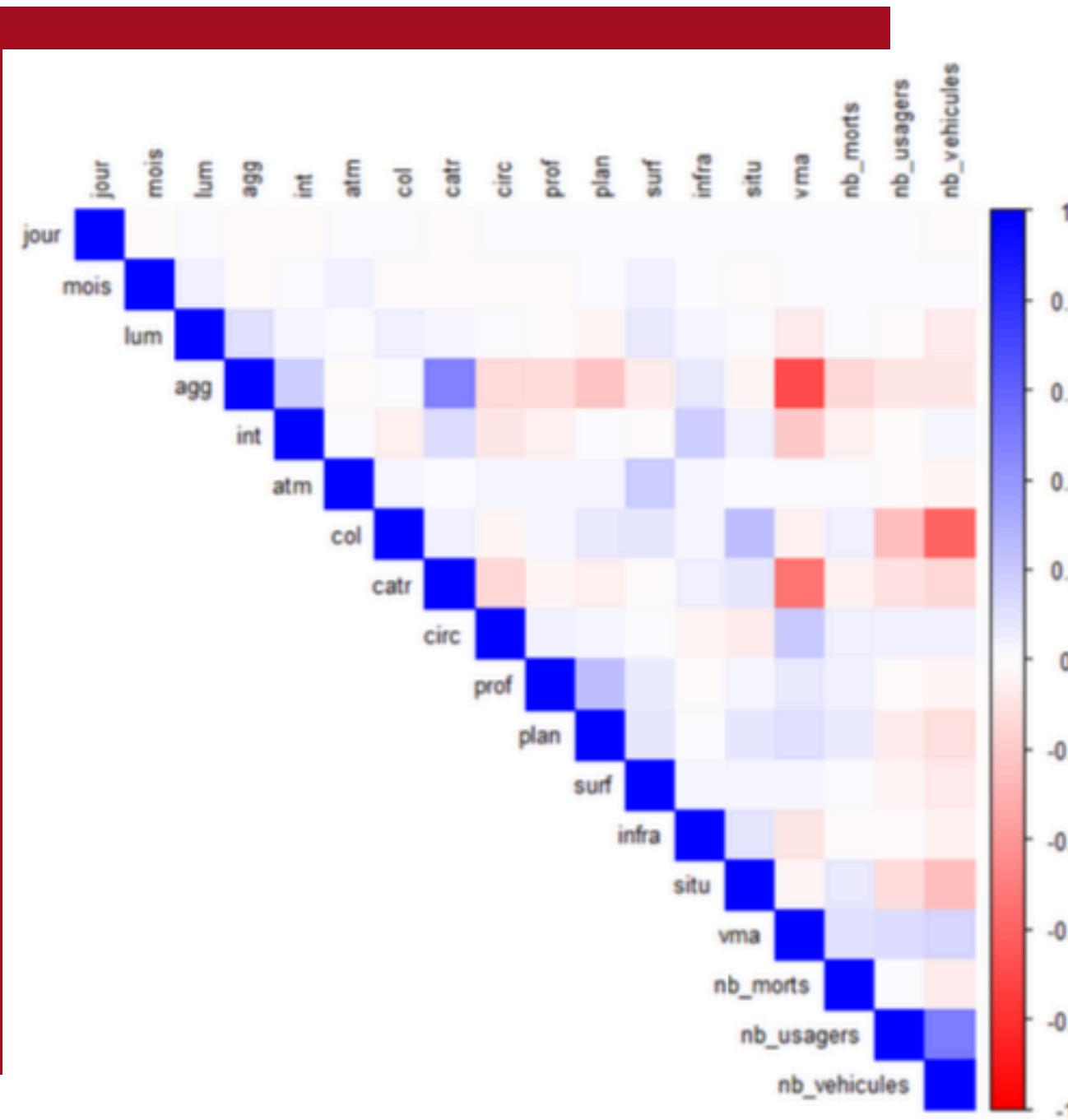
Analyse descriptive

Nombre d'accidents en France (biaisé par la densité de population)

Carte de Densité des Accidents Routiers en France (2023)



Analyse statistique et modélisation



Corrélations significatives identifiées

- Forte corrélation entre catr (type de route) et contexte urbain/non urbain(agg) avec nb_vehicules.
- Relation constante entre vma et gravité des accidents.

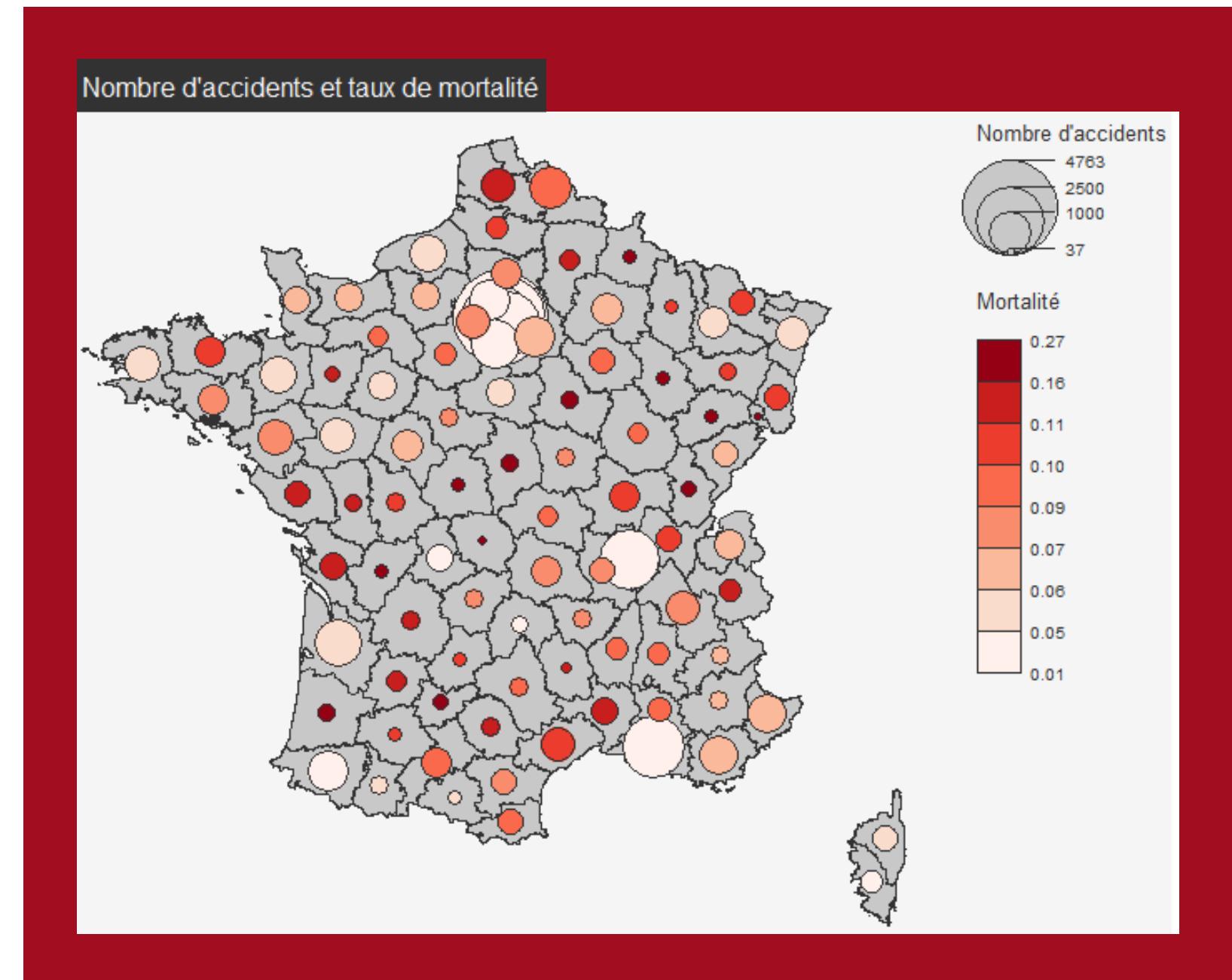
Stabilité et variations dans le temps

- Stabilité vma & gravité.
- Fluctuations entre atm & infra.

Méthodologie et objectifs

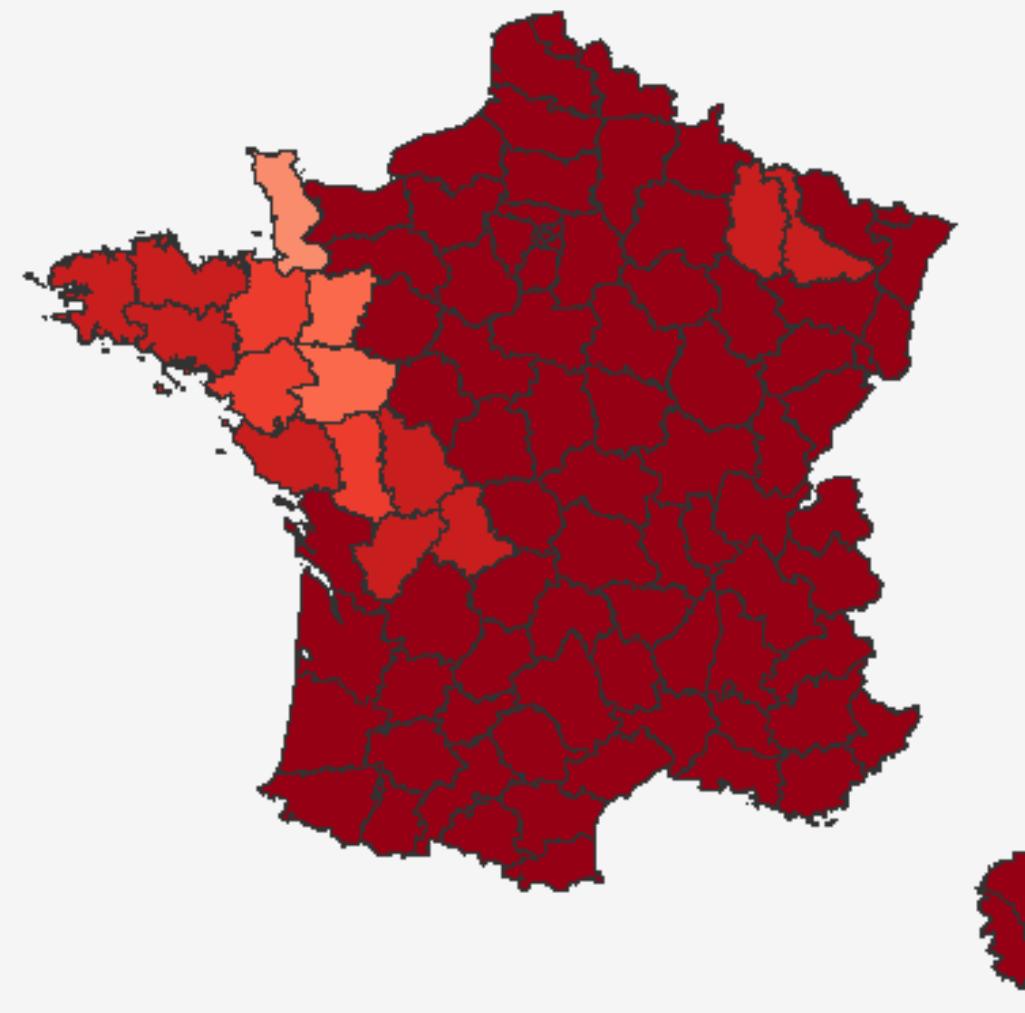
- Matrice de corrélation: détection des associations significatives et variables redondante.
- Comprendre les relations structurelles des accidents et leur évolution temporelle.

Représentation géographique



Régression géographique pondérée

Régression géographique départementale en 2023



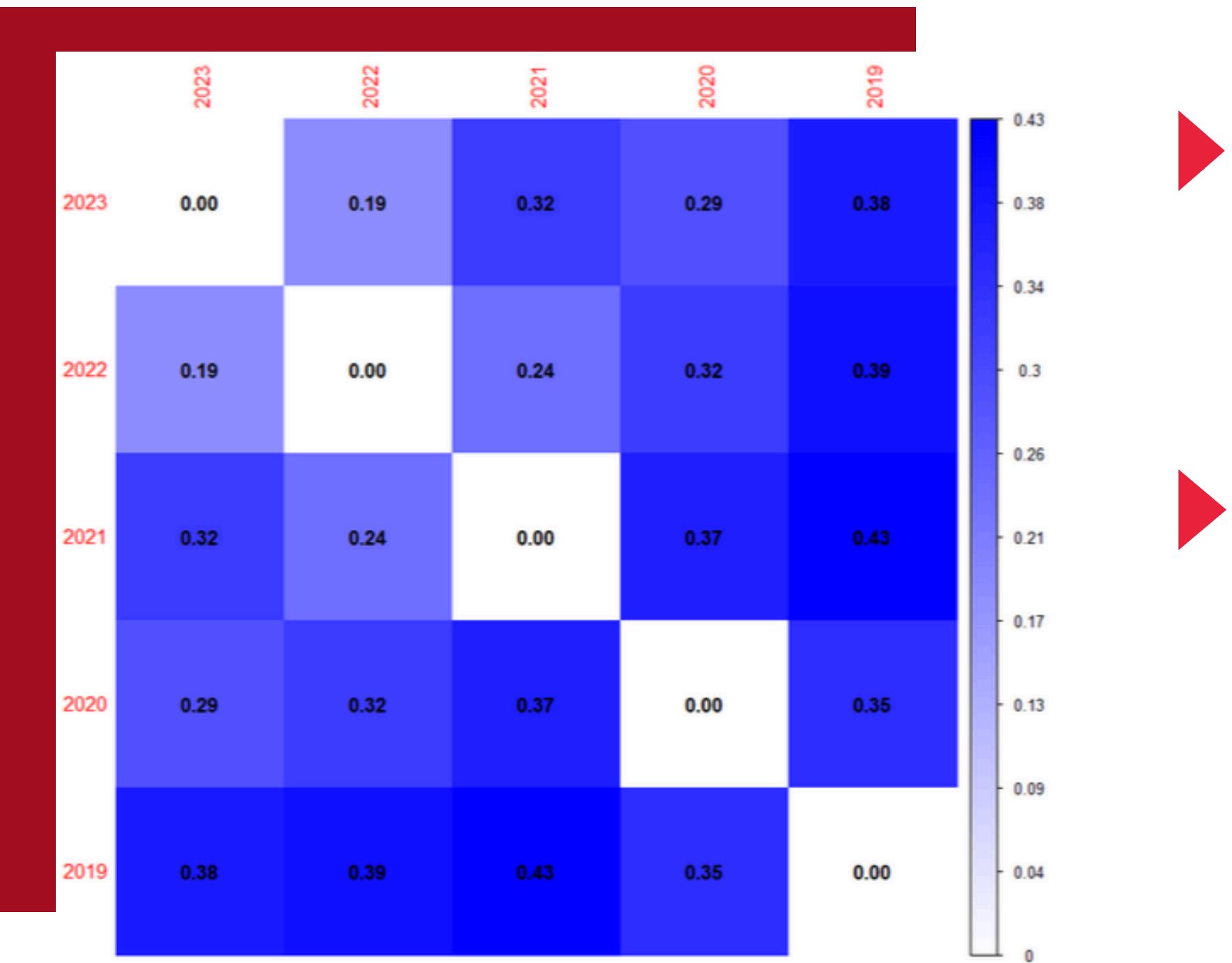
➤ Comparaison de différents paramètres à la densité :

- Nombres d'accidents, mortalité, taux d'accidents en agglomérations et nombres de véhicules impliqués

➤ Librairie GWmodel

- Utilisation de la méthode de "forward selection"

Analyse statistique et modélisation



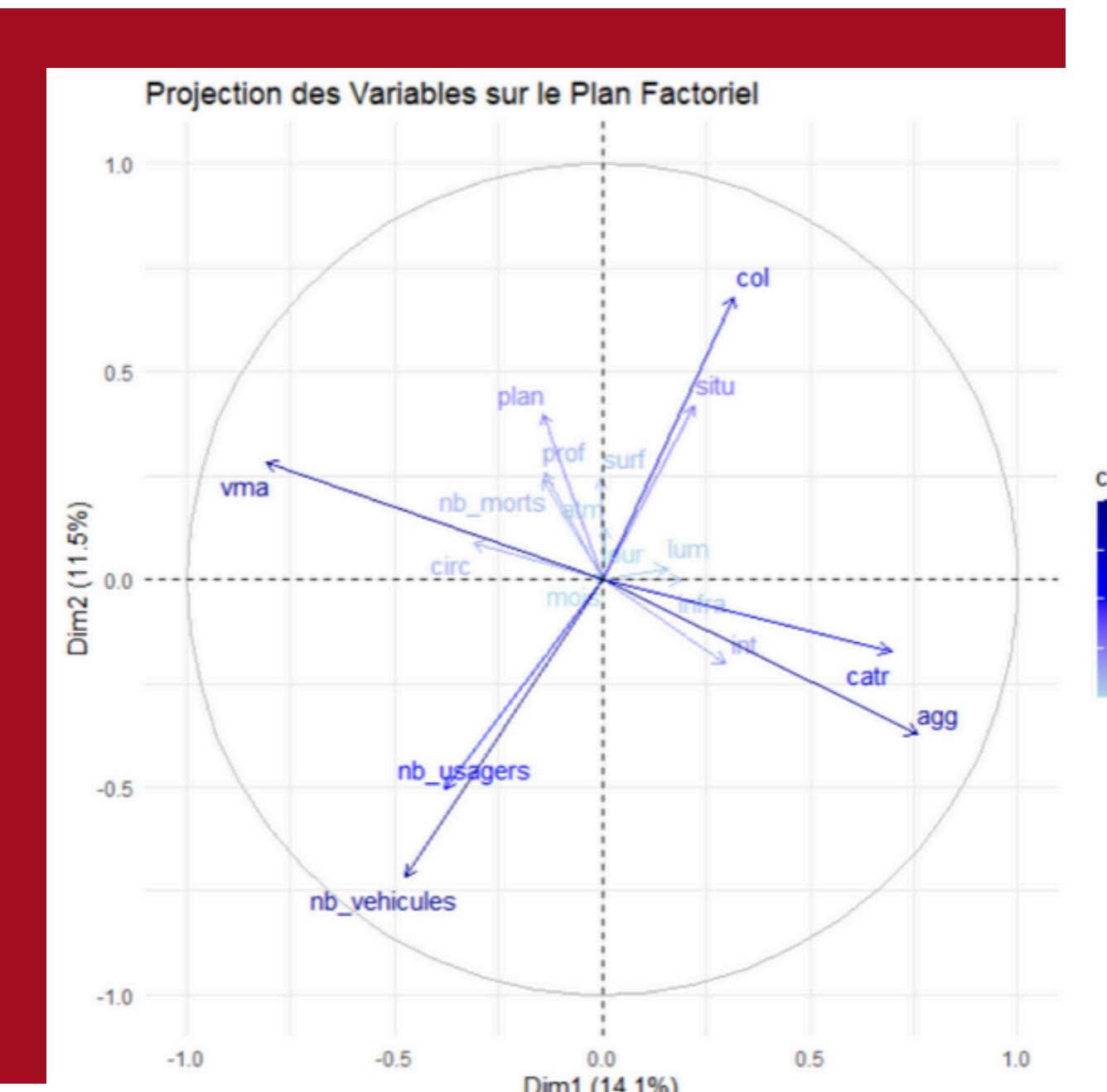
Distance de Frobenius

- Quantifie la différence entre les matrices de corrélation de chaque années
- Mesure la stabilité ou les changements significatifs dans les relations entre variables.

Interprétation

- 2023 et 2022 montrent une forte similarité (distance de 0.19).
- Les matrices de 2019 sont les plus éloignées des années récentes (distance de 0.43 avec 2021).

Analyse statistique et modélisation



Objectifs de l'ACP:

- Identifier les variables les plus contributrices à la variabilité des données.
- Analyser les évolutions temporelles des relations entre variables.

Variance expliquée :

- CP1 explique 13.7% à 14.1% de la variance selon l'année.
- Les deux premières composantes cumulées expliquent environ 25% de la variance totale.

Contributions des variables :

- CP1 : Variables principales – vma, agg, et nb_vehicules.
- CP2 : Variables clés – catr, col, et nb_usagers.

Analyse statistique et modélisation

Test du Chi² (relation entre luminosité et accidents) :

- Valeur de Chi² : **101171**.
- p-value : < **2.2e-16**.
- Interprétation : Distribution non uniforme de la luminosité, certaines catégories étant plus associées aux accidents.

Test de Fisher (luminosité vs type d'agglomération) :

- p-value : **0.01307**.
- Odds ratio : **1.05 (faible effet)**.
- Interprétation : Association significative mais faible entre jour/nuit et le type d'agglomération.

Analyse statistique et modélisation

Exemple pour le paramètre “int” donnant le type d’intersection

Table de contingence observée

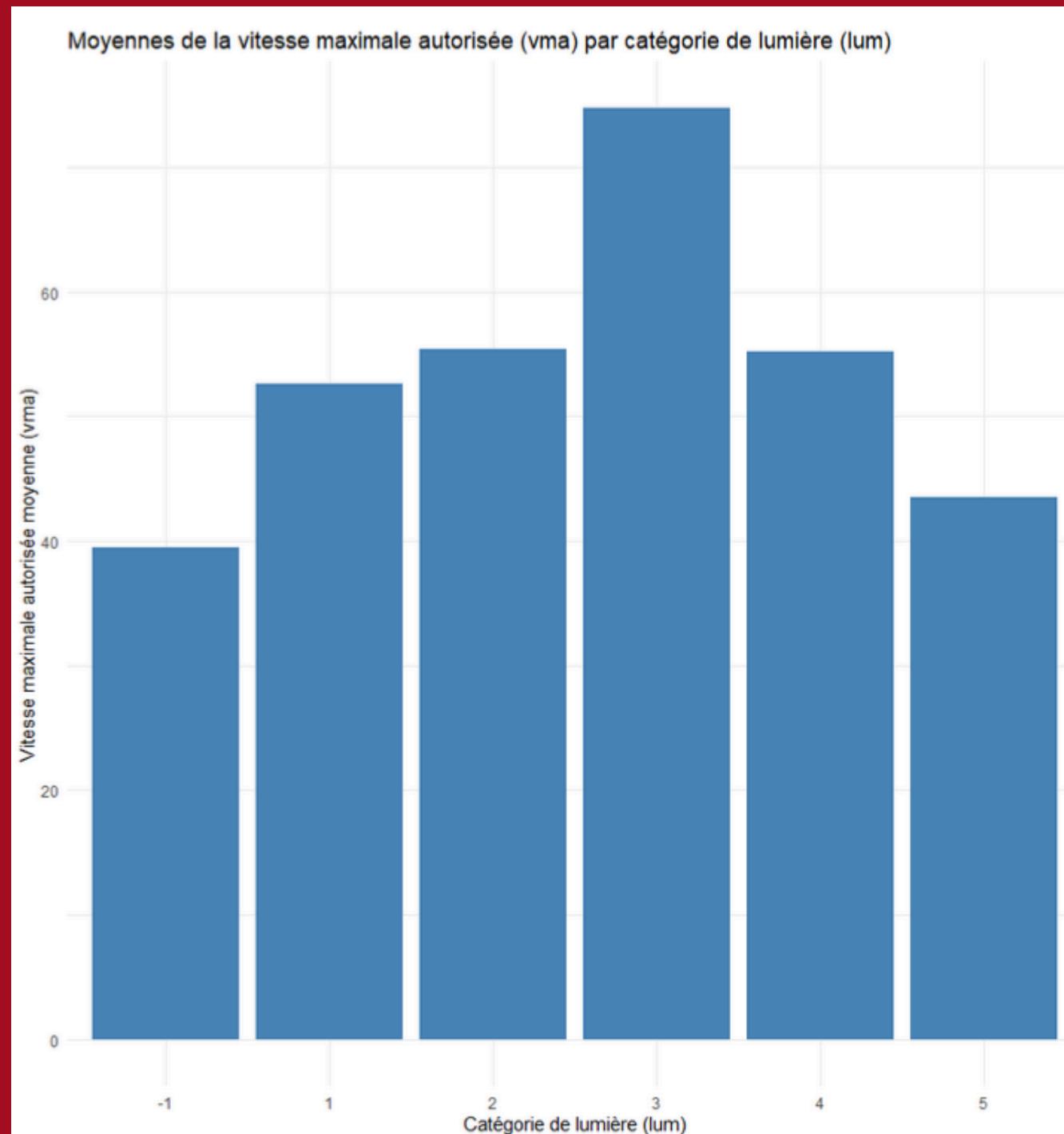
INT	Mortel	Non mortel
-1	0	6
1	2556	32286
2	187	6556
3	188	6171
4	37	1058
5	6	286
6	83	2357
7	6	409
8	32	109
9	108	2381

- Statistique du test : $X^2 = 497.11$
- Degrés de liberté (df) : 9
- p-value : $p < 2.2 \times 10^{-16}$

Table de contingence théorique

INT	Mortel	Non mortel
-1	0.148	5.852
1	1200.420	32241.580
2	179.258	6563.742
3	176.412	6182.588
4	31.499	1063.501
5	7.518	284.482
6	76.464	2363.536
7	11.891	403.109
8	3.623	137.377
9	113.768	2375.232

Analyse statistique et modélisation



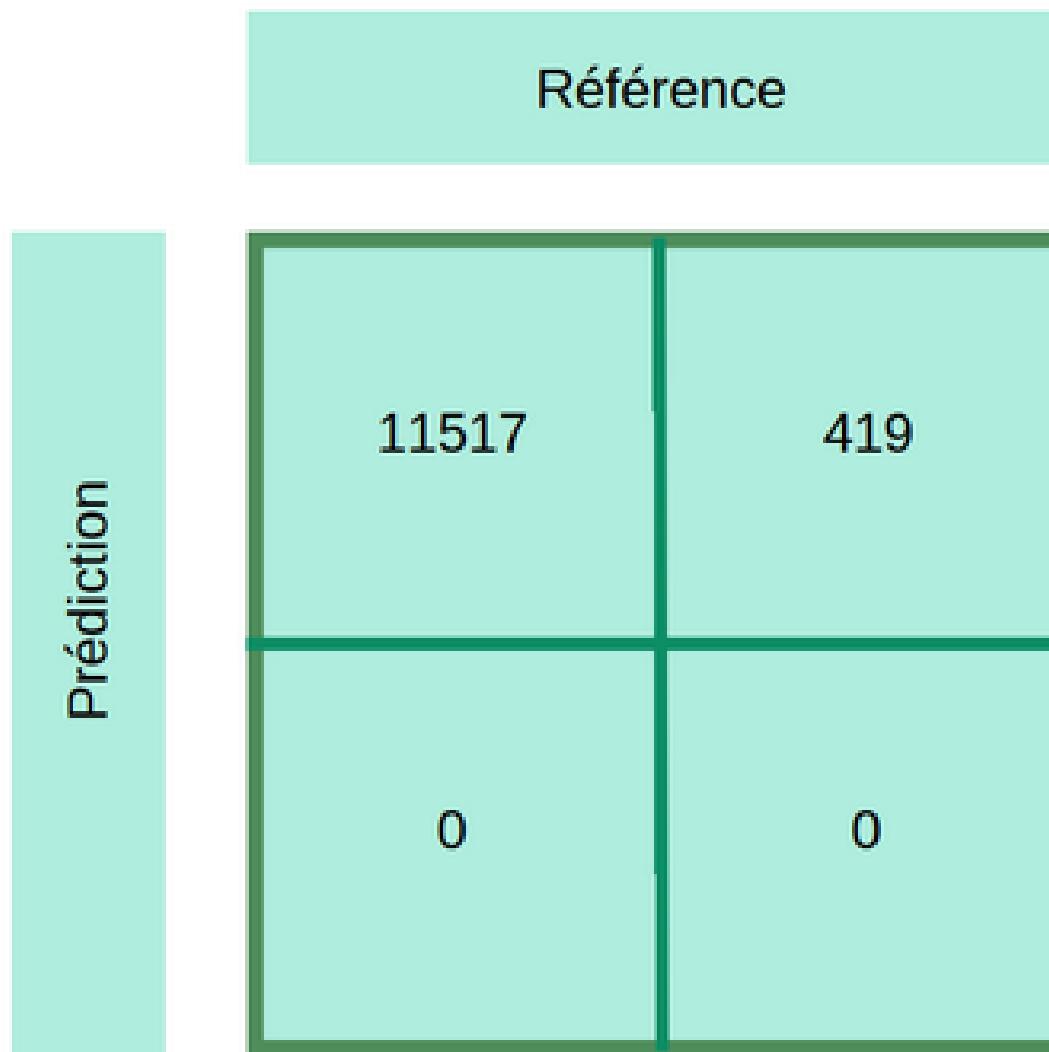
Analyse ANOVA (luminosité et vitesse maximale autorisée) :

- p-value : < 2e-16.
- Moyennes de vitesse (vma) par catégorie de luminosité, par exemples: Niveau 1 : 52.6 km/h, Niveau 3 : 74.7 km/h.
- Interprétation : La luminosité impacte significativement la vitesse maximale autorisée.



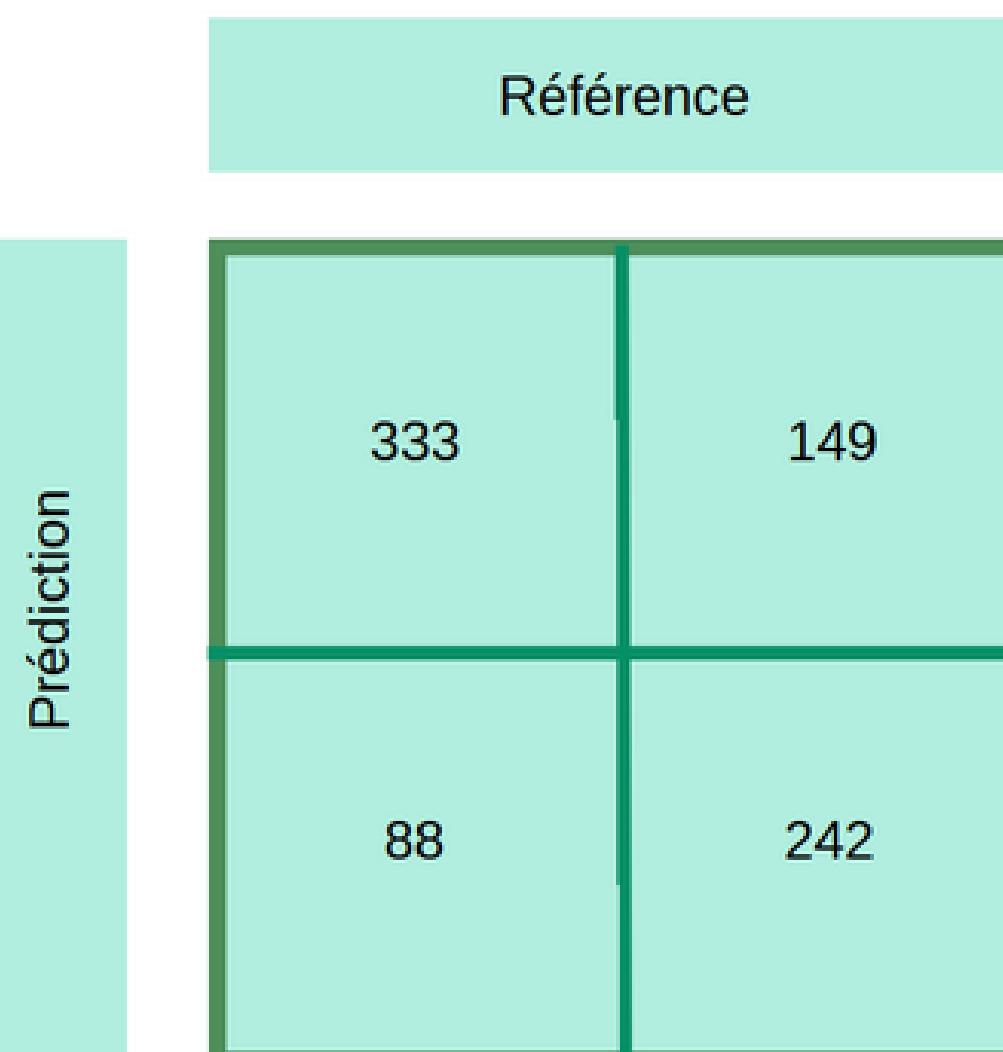
Prédiction - Random Forest

Année 2020



Matrice de confusion sans équilibrage des données

F1 score = 98%
p-value = 0.513

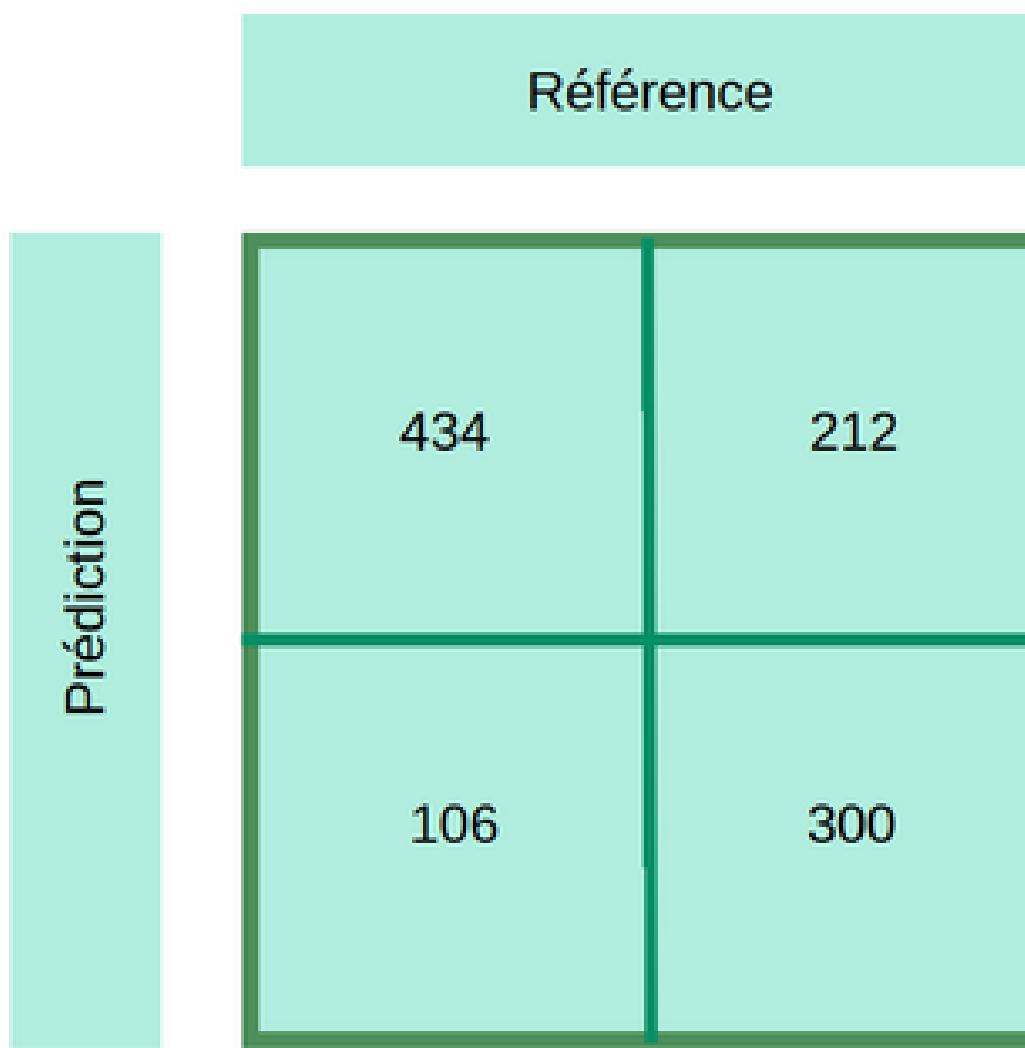


Matrice de confusion avec équilibrage des données

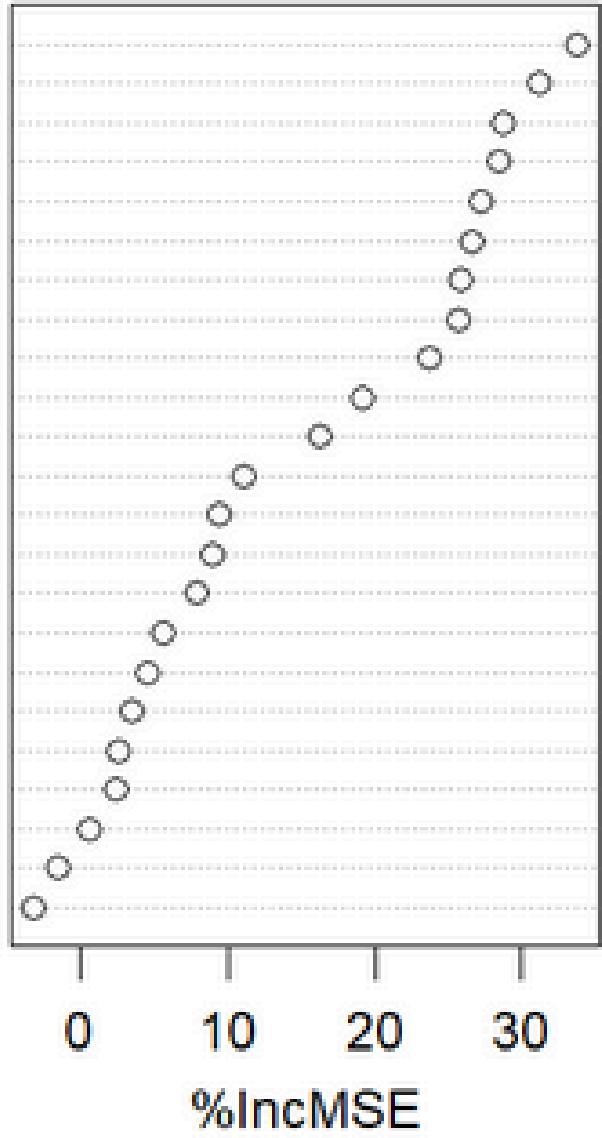
F1 score = 74%
p-value < 2.2e-16

Année 2022

Pour évaluer les différences entre les années



nb_vehicules
col
com
vma
agg
lat
situ
dep
catr
long
circ
hrmn
int
plan
lum
adr
surf
atm
prof
Infra
an
mois
jour



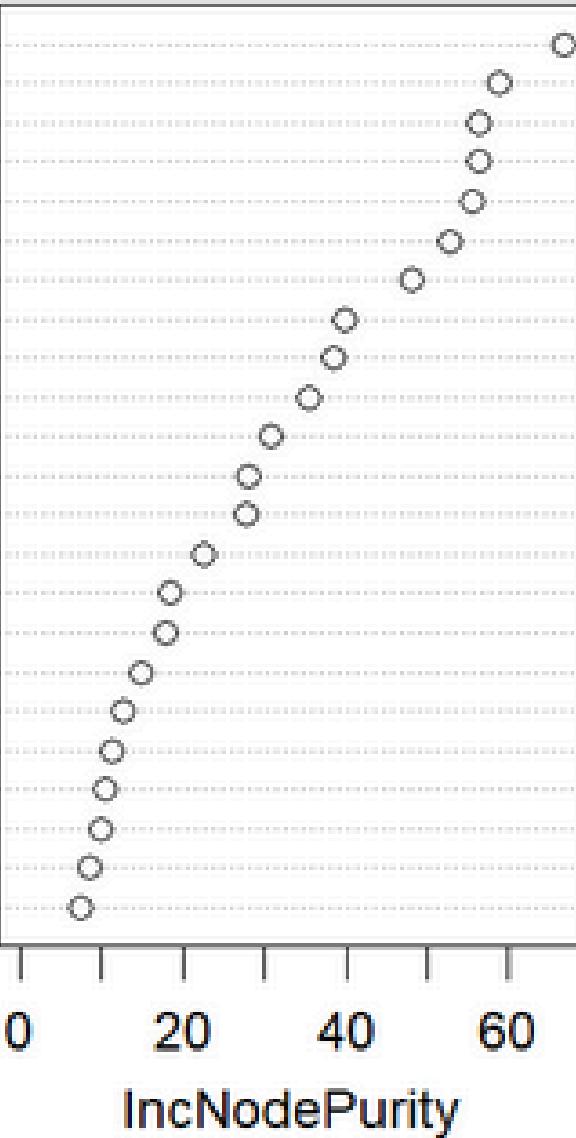
Matrice de confusion avec
équilibrage des données

F1 score = 73%

p-value < 2.2e-16

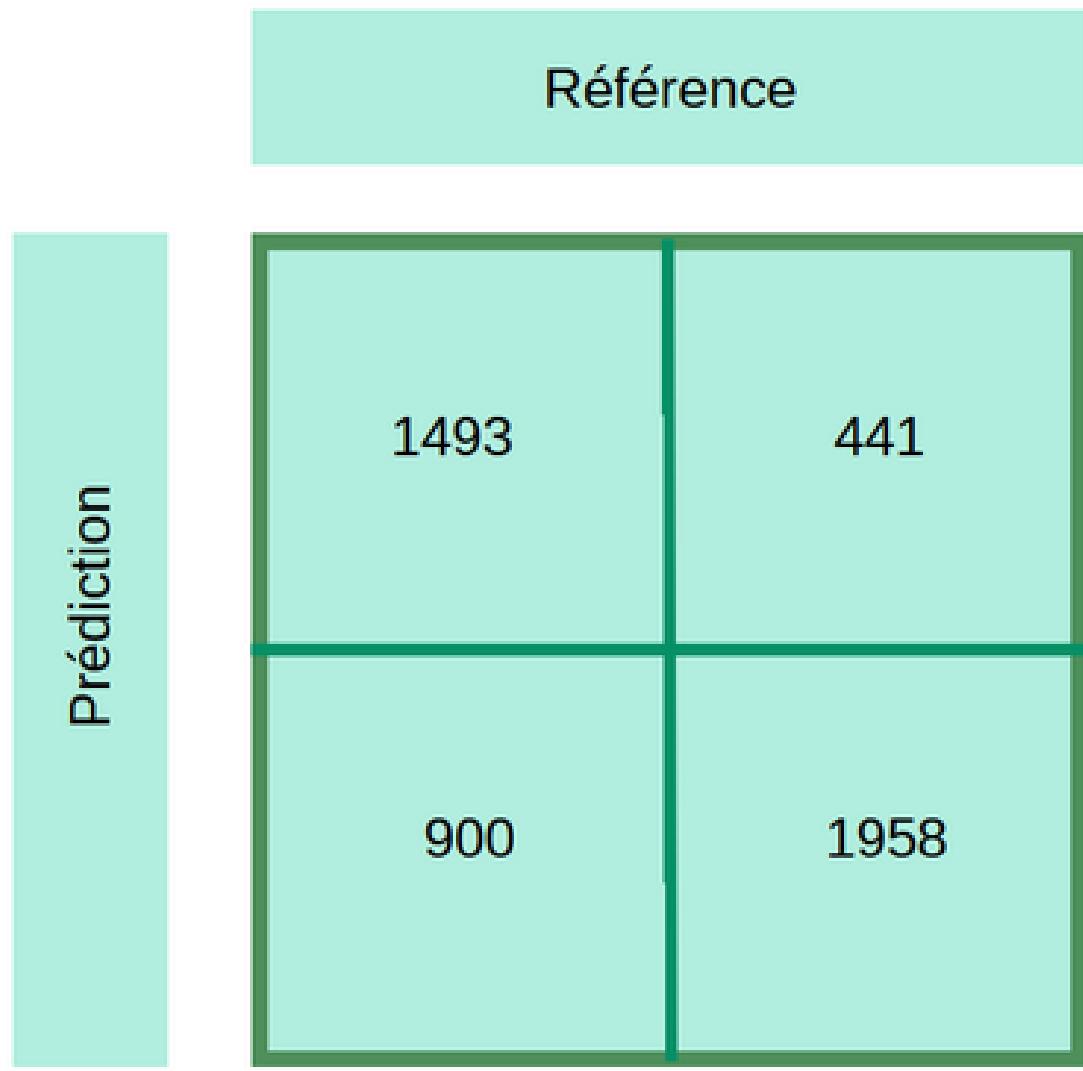
%IncMSE (Pourcentage
d'augmentation de l'erreur
quadratique moyenne)

lat
long
col
hrmn
com
adr
nb_vehicules
dep
jour
vma
situ
mois
agg
catr
an
circ
int
plan
lum
atm
prof
Infra
surf



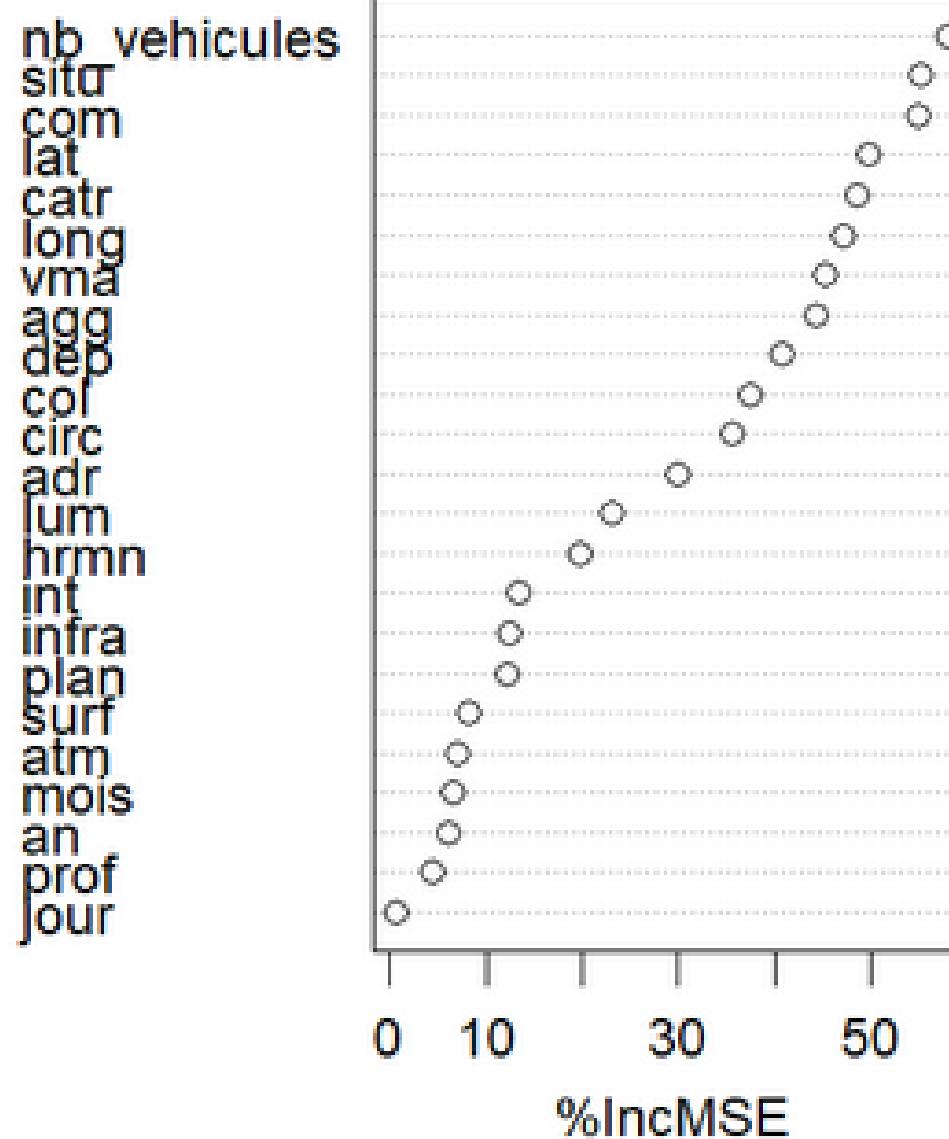
IncNodePurity (Importance en
fonction de l'amélioration de la pureté
des nœuds des arbres de décision)

Années 2019-2023

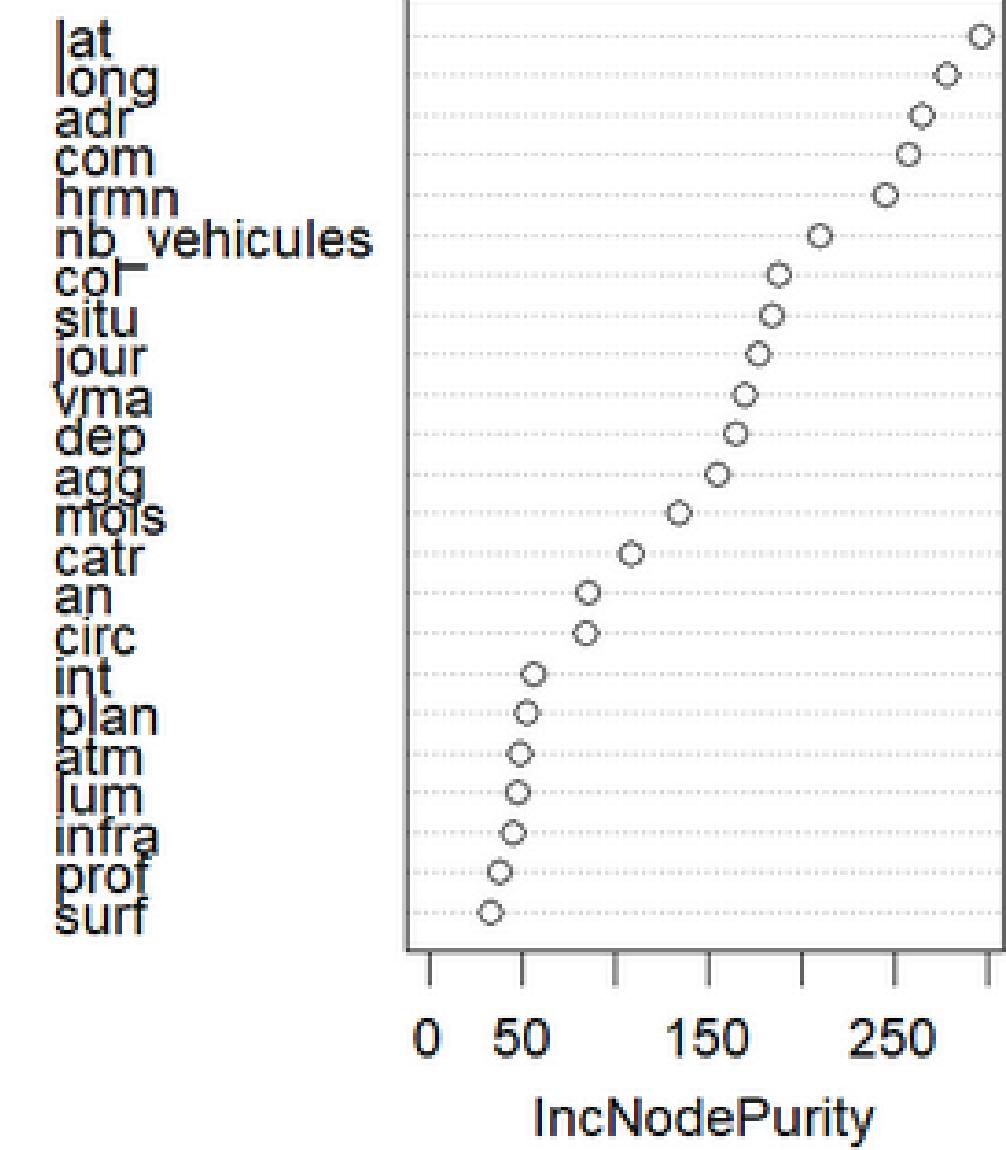


Matrice de confusion avec équilibrage des données

F1 score = 69%
p-value < 2.2e-16



%IncMSE (Pourcentage d'augmentation de l'erreur quadratique moyenne)



IncNodePurity (Importance en fonction de l'amélioration de la pureté des nœuds des arbres de décision)

Validation croisée et Bootstrap

Résultats avec validation croisée :

précision moyenne = 0.704

écart-type de la précision = 0.0098

F1 score = 65.4%

écart-type du F1 score = 0.0133

→ Résultats cohérents avec résultats précédents : stabilité du modèle validée

Résultats avec bootstrap avec 25 répétitions :

Moyenne F1 score = 0.668
écart-type F1 score = 0.0081

Quantile des F1 score:

0%	25%	50%	75%	100%
0.66	0.66	0.67	0.67	0.69

→ bonne variabilité des statistiques



**Merci pour votre
attention !**

Bibliographie

[1] Al-Mamlook, R., Abouchabaka, J., & Al-Gadhib, A. (2020).

Comparaison des algorithmes d'apprentissage automatique pour la prédition de la gravité des accidents de la route.

Disponible sur [ResearchGate](#).

[2] Bulut, A. (2020).

Prédiction de la gravité des accidents de la route à l'aide d'algorithmes d'apprentissage automatique.

DergiPark Journal. Disponible sur [DergiPark](#).

[3] Cybergeo: Revue européenne de géographie (2018).

Statistiques spatiales des accidents de la route.

Disponible sur [Cybergeo](#).

Bibliographie

- ▶ [4] **IEEE Xplore (2019).**
Comparaison des algorithmes d'apprentissage automatique pour la prédition de la gravité des accidents de la route.
Disponible sur [IEEE Xplore](#).

- ▶ [5] **Université Gustave Eiffel, LASTIG. (2025).**
Projet d'analyse des données – ENSG Geo Data Science UE2.
Dépôt disponible sur [GitHub](#).