

# Analyse des accidents de la route pendant l'année 2021 en fonction du lieu, de la période, de l'âge et du sexe

Groupe FLAL

## Contents

Etude globale sur les accidents en 2021 en France . . . . .	2
<b>Est ce que les usagers influencent la gravité des accidents ?</b>	<b>4</b>
Commençons par étudier si l'âge du conducteur est en corrélation avec la gravité de l'accident . . .	4
<b>Est-ce que le sexe influence la gravité des accidents ?</b>	<b>8</b>
Est ce que le lieu influence la gravité des accidents ? . . . . .	10
Est ce que l'environnement influence la gravité des accidents ? . . . . .	15

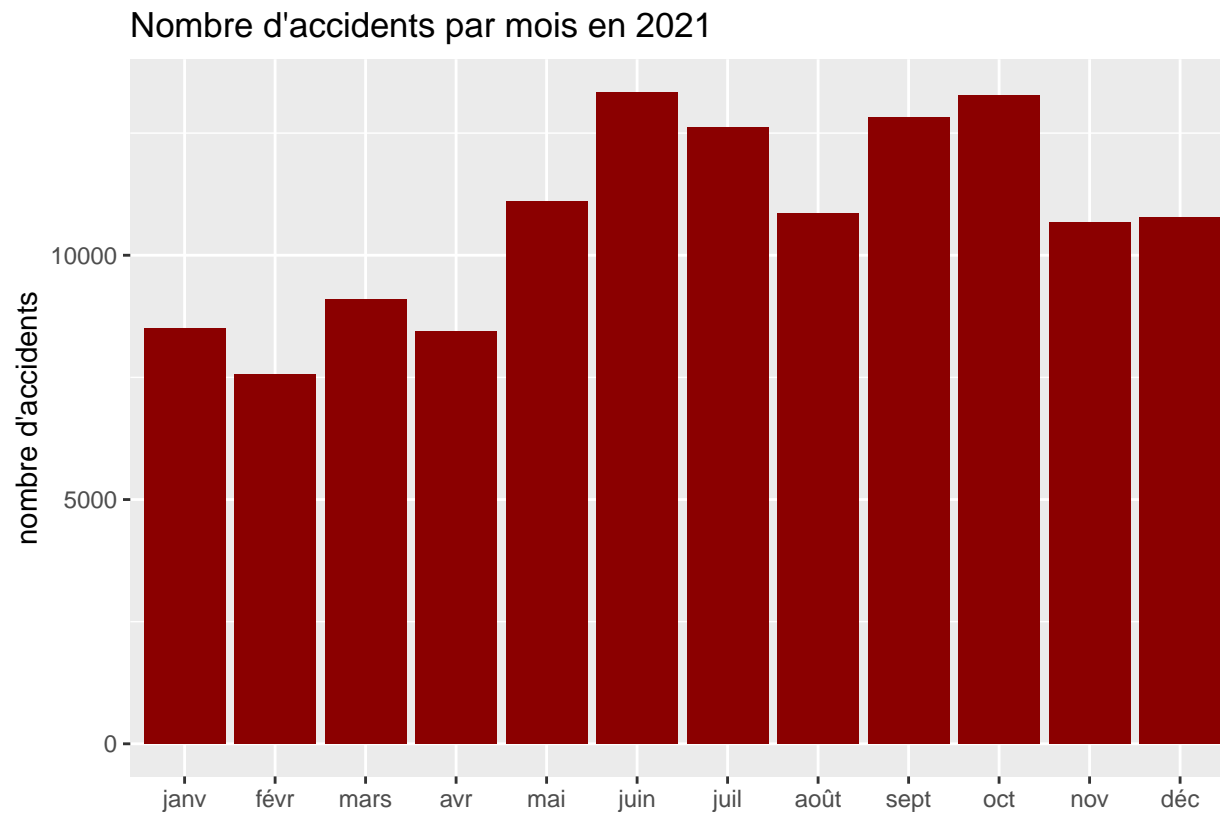
**Alex Delagrangé, Léo Bouvier, Lucas Giry, Farah Seifeddine** Dans ce rapport, nous cherchons à expliquer la gravité des accidents en France en 2021. Le jeu de données vient du site [data.gouv.fr](https://data.gouv.fr). La question dans ce dossier est :

#quels sont les facteurs importants impactant la gravité des accidents en France en 2021 ?

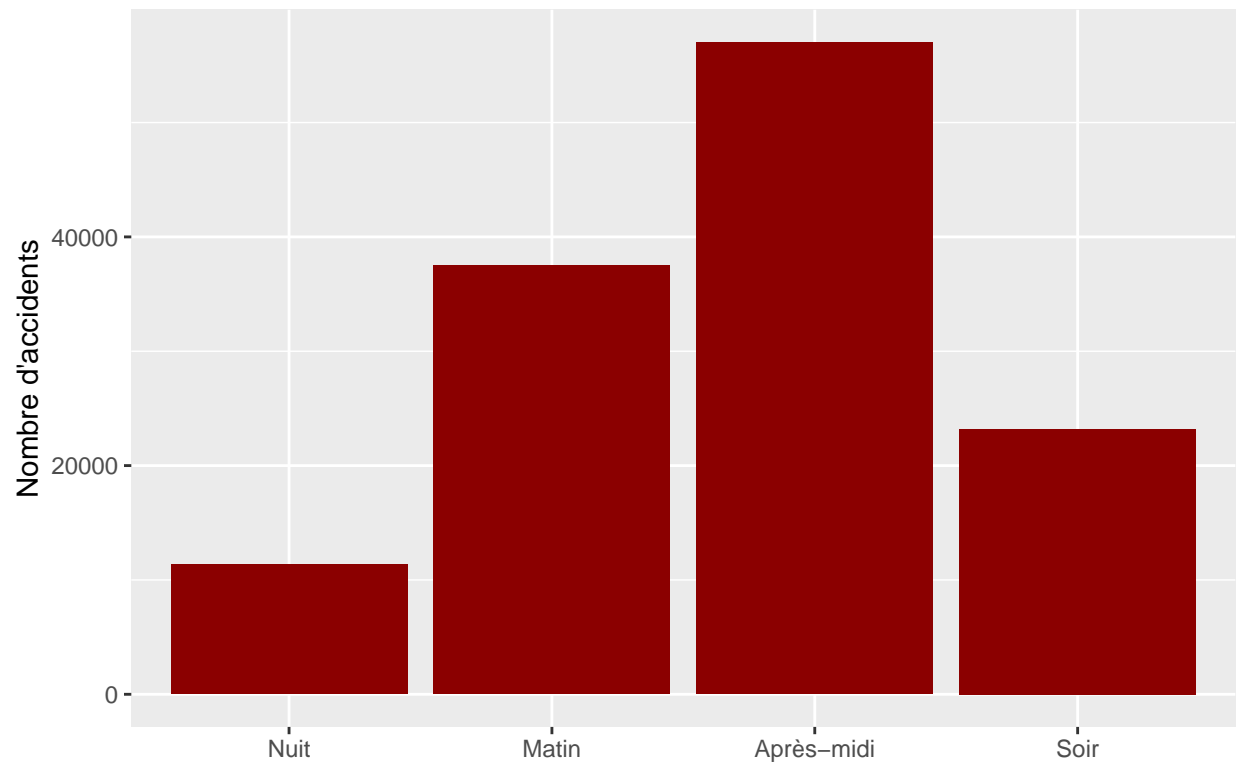
Pour trier le jeu de données, nous avons d'abord commencé par enlever les lignes de la dataframe sans données dans lesquelles les colonnes étaient remplies de -1. On a ensuite trié les colonnes pour supprimer les colonnes que nous n'allons pas regarder dans notre études comme par exemple le type de moteur du véhicule. Pour finir il fallait enlever les caractères spéciaux des régions. Après tout cela nous avons un jeu de données propre et exploitable.

En ce qui concerne la représentation des données on a décidé de représenter toutes les données dans une grande dataframe dont on récupérera les colonnes nécessaires pour les différents facteurs.

## Etude globale sur les accidents en 2021 en France

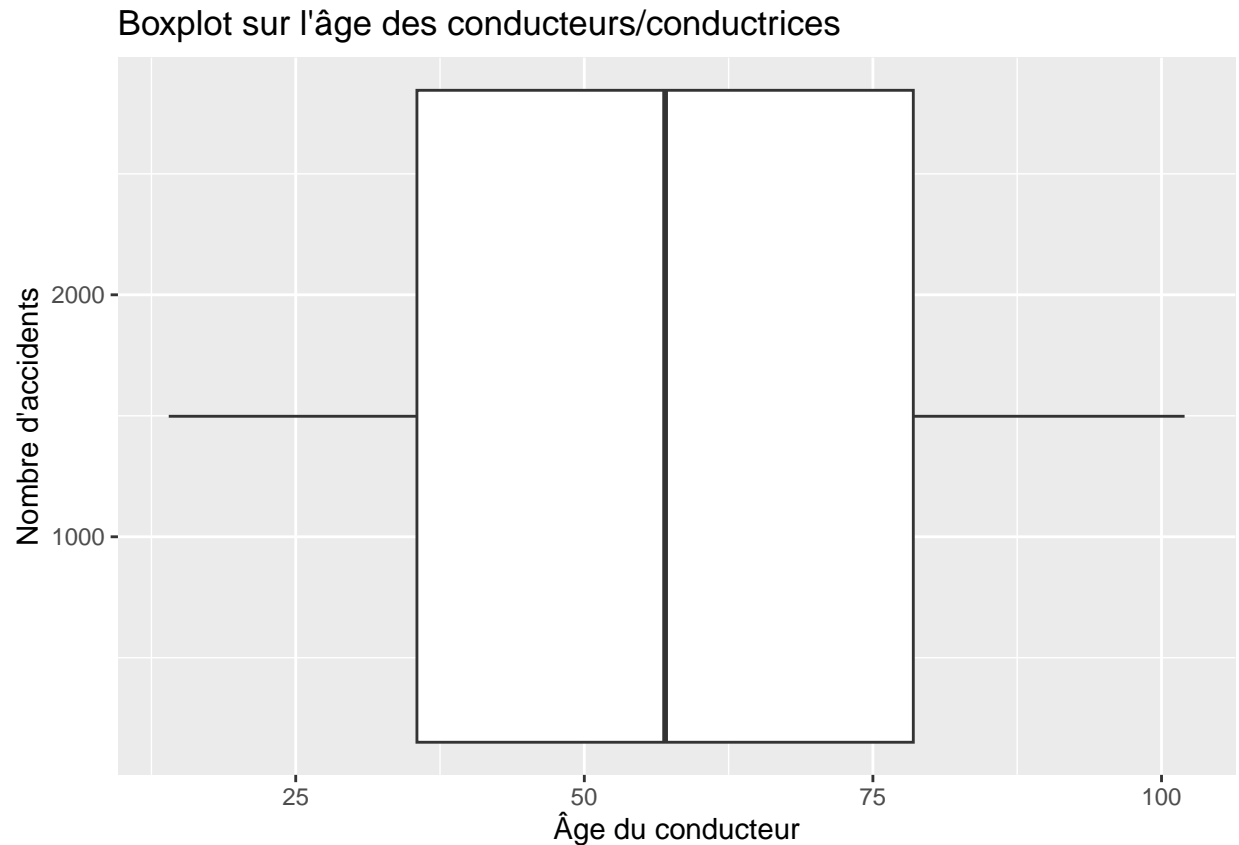


Nombre d'accidents par période de la journée en 2021



## Est ce que les usagers influencent la gravité des accidents ?

Commençons par étudier si l'âge du conducteur est en corrélation avec la gravité de l'accident



```
## $stats
## [1] 14.0 35.5 57.0 78.5 102.0
##
## $n
## [1] 87
##
## $conf
## [1] 49.71607 64.28393
##
## $out
## numeric(0)
```

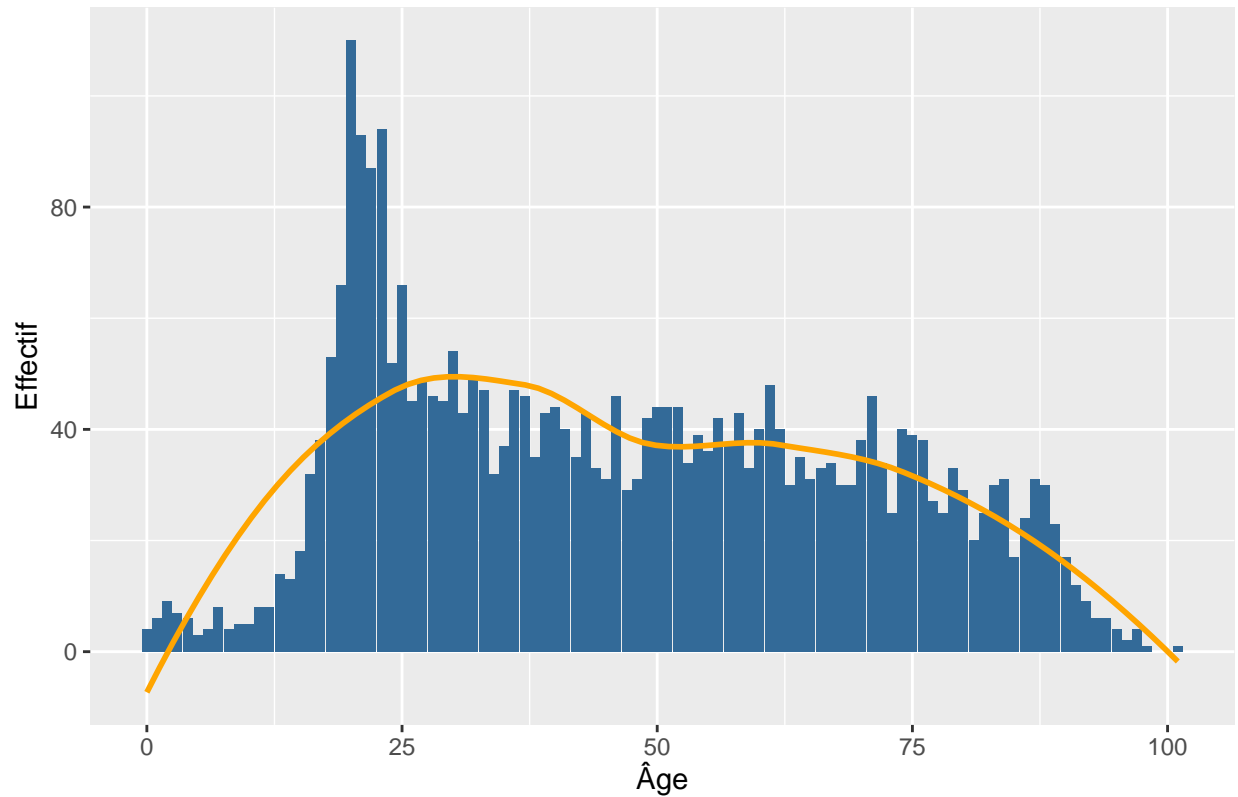
Sur 126086 accidents de la route en 2021, 50 % des conducteurs ont moins de 57 ans. 50 % d'entre eux ont entre 35.5 et 78.5 ans.

Dans ce cas, l'intervalle de confiance est [49.7,64.3], ce qui signifie que l'on peut être raisonnablement sûr que la moyenne de l'âge dans la population dont l'échantillon a été prélevé se trouve dans cette plage avec une probabilité de 95%.

Effectif des accidents mortels par âge du conducteurs

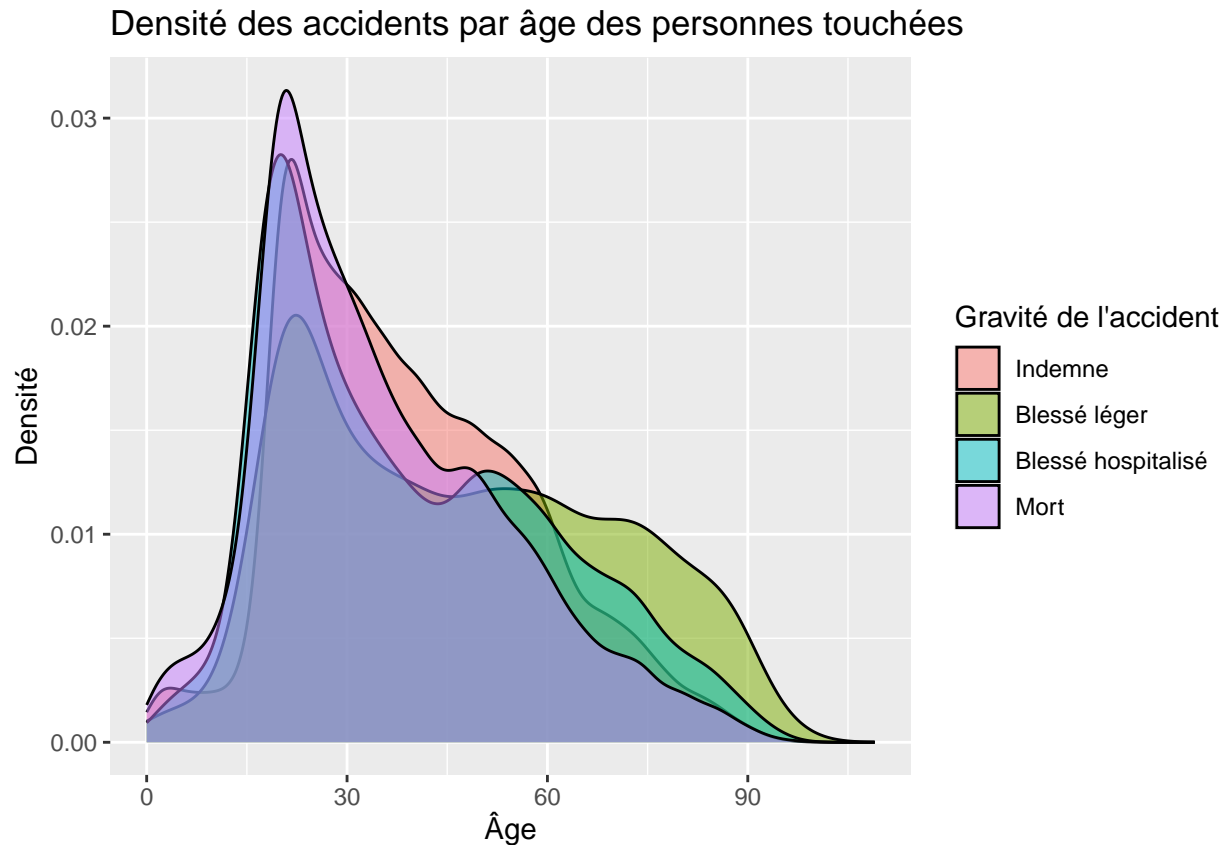


## Effectif des accidents mortels par âge des personnes touchées



On constate la même croissance du nombre d'accidents mortels (conducteur ou non) de 0 à 23 ans, les personnes touchées sont plus nombreuses, ce qui est normal il y a les passagers ajoutés. La tendance décroît fortement de 23 à 40 ans avant de rester plus ou moins au même niveau jusqu'à environ 70 ans. Pour les conducteurs, on constate une décroissance nette plus tôt que chez les personnes touchées, une hypothèse serait que plus l'âge augmente, moins il y a de conducteurs.

Le fort pic autour de 20 ans pourrait s'expliquer par l'âge de l'obtention du permis de conduire qui se traduit par un manque d'expérience en tant que conducteur.



De manière générale, le graphique de densité montre comment les distributions de l'âge des conducteurs diffèrent selon la gravité de l'accident. On remarque un fort pic de densité autour de 20 ans ce qui laisse penser à une forte corrélation entre l'âge et la gravité de l'accident.

```
##
## Call:
## glm(formula = grav_bin ~ age, family = binomial(), data = sub_df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.4426  -1.3180   0.9822   1.0296   1.2298
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.6048416  0.0129956  46.54  <2e-16 ***
## age         -0.0066719  0.0003048 -21.89  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 170975  on 126085  degrees of freedom
## Residual deviance: 170495  on 126084  degrees of freedom
## AIC: 170499
##
## Number of Fisher Scoring iterations: 4
```

Les résultats de la régression logistique indiquent que l'âge est significativement associé à la gravité des accidents. Plus précisément, pour une unité d'augmentation de l'âge, la log-odds d'avoir une gravité plus élevée diminue de 0,00667. Cela peut être interprété comme une diminution de la probabilité d'avoir une gravité plus élevée pour chaque année supplémentaire.

Le rapport des deviances (null deviance et residual deviance) montre que le modèle ajuste bien les données, car il y a une réduction significative de la deviance résiduelle par rapport à la deviance nulle. En outre, l'AIC est relativement faible, ce qui indique que le modèle est un ajustement approprié pour les données.

Le test de significativité indique que la relation entre l'âge et la gravité de l'accident est très significative ( $p\text{-value} < 2e-16$ ), ce qui renforce l'idée que l'âge est un facteur important à prendre en compte dans l'évaluation de la gravité des accidents de la route.

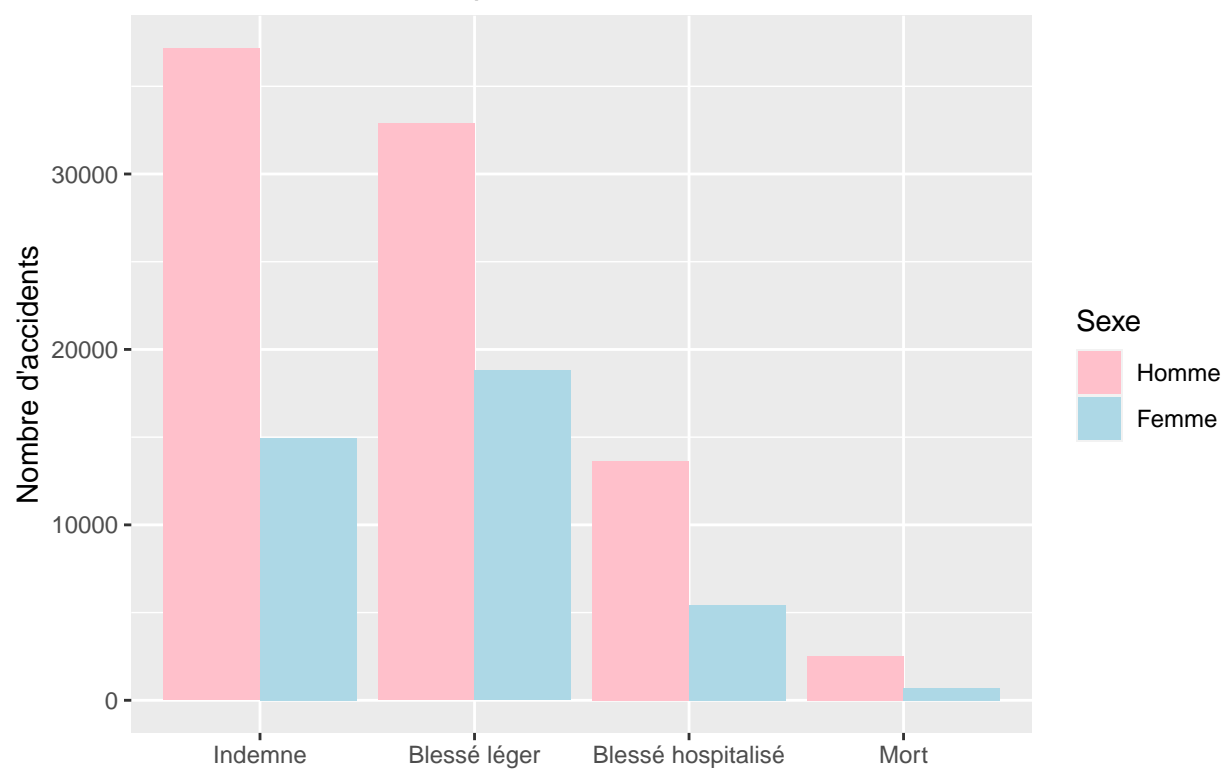
En résumé, les résultats de la régression logistique soutiennent l'idée que l'âge est significativement associé à la gravité des accidents de la route et que le modèle est un ajustement approprié pour les données.

## Est-ce que le sexe influence la gravité des accidents ?

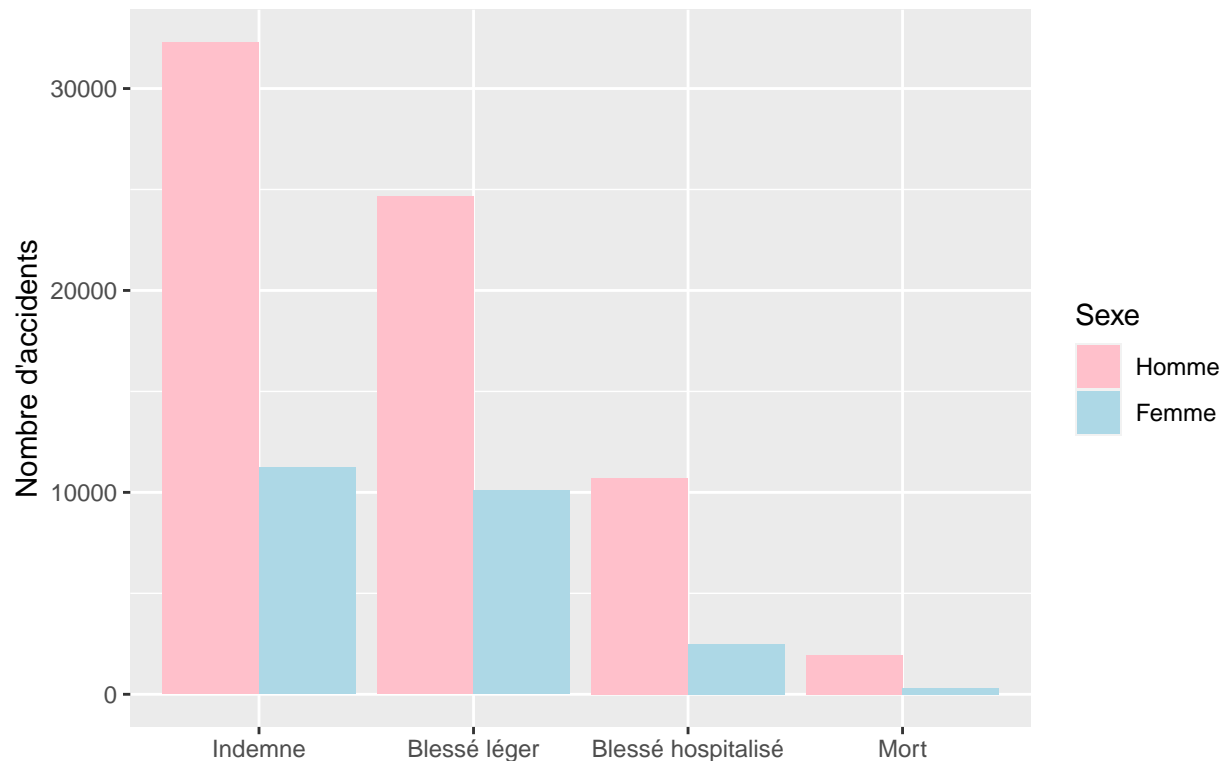
```
sub_df <- df[, c("grav", "sexe")]
# enleve les na
sub_df <- na.omit(sub_df) %>% filter(sexe != -1)
tab <- table(sub_df$sexe, sub_df$grav)
# Conversion de la table croisée en un data frame
df_s <- as.data.frame.matrix(tab)
df_s$sexe <- rownames(df_s)
# Mise en forme des données pour un diagramme en barres empilées
df_long <- tidyr::gather(df_s, key = "grav", value = "count", -sexe)
df_long$grav <- factor(df_long$grav, levels = c("1", "2", "3", "4"))
df_long$sexe <- factor(df_long$sexe, levels = c("1", "2"))
# Convertir la variable grav en factor avec les niveaux correspondants
df_long$grav <- factor(df_long$grav, levels = 1:4,
  labels = c("Indemne", "Mort", "Blessé hospitalisé", "Blessé léger"))
df_long$sexe <- factor(df_long$sexe, levels = 1:2,
  labels = c("Homme", "Femme"))
# Définir les couleurs personnalisées
colors <- c("Homme" = "pink",
  "Femme" = "lightblue")
# Réorganiser les niveaux de la variable "grav" en fonction de la fréquence
df_long$grav <- reorder(df_long$grav, desc(df_long$count))
# Créer le graphique
ggplot(df_long, aes(x = grav, fill = sexe)) +
  geom_col(position = "dodge", aes(y = count)) +
  scale_fill_manual(values = colors, name = "Sexe",
    labels = c("Homme", "Femme")) +
  labs(title = "Gravité des accidents par sexe", x = "", y = "Nombre d'accidents") + theme_gray()
```



Gravité des accidents par sexe



## Gravité des accidents par sexe du conducteur



On remarque que, conducteur/trice ou non, la tendance reste la même au niveau de la gravité. La quantité d'accidents reste plus importante chez les hommes que chez les femmes.

```
##
## Pearson's Chi-squared test
##
## data: table(df_conducteurs$sexe, df_conducteurs$grav)
## X-squared = 686.69, df = 6, p-value < 2.2e-16
```

Puisque la p-value est inférieure au seuil de significativité communément utilisé de 0,05, on peut rejeter l'hypothèse nulle selon laquelle il n'y a pas d'association entre le sexe et la gravité des accidents de la route. On peut donc conclure qu'il y a une association significative entre le sexe et la gravité des accidents de la route en France en 2021.

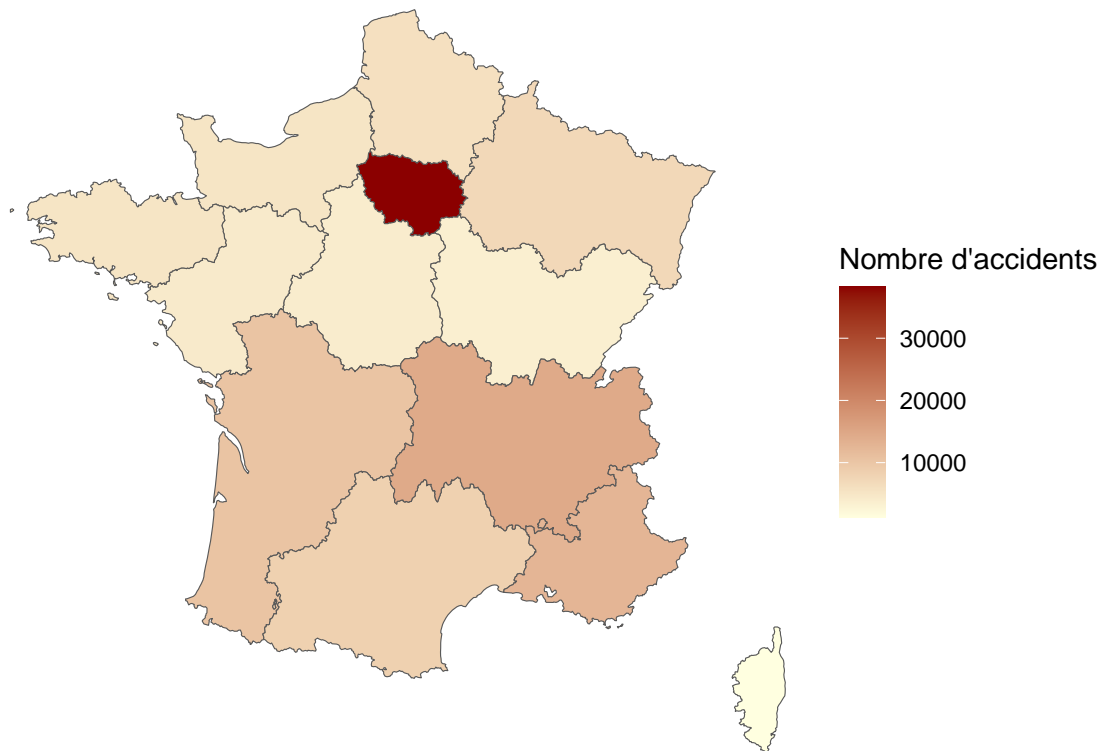
Cela signifie que le sexe des conducteurs est statistiquement significatif pour prédire la gravité des accidents de la route en France en 2021. Cependant, il est important de noter que les résultats de ce test ne permettent pas de déterminer la direction de l'association (c'est-à-dire, si les accidents graves sont plus fréquents chez les hommes ou chez les femmes).

## Est ce que le lieu influence la gravité des accidents ?

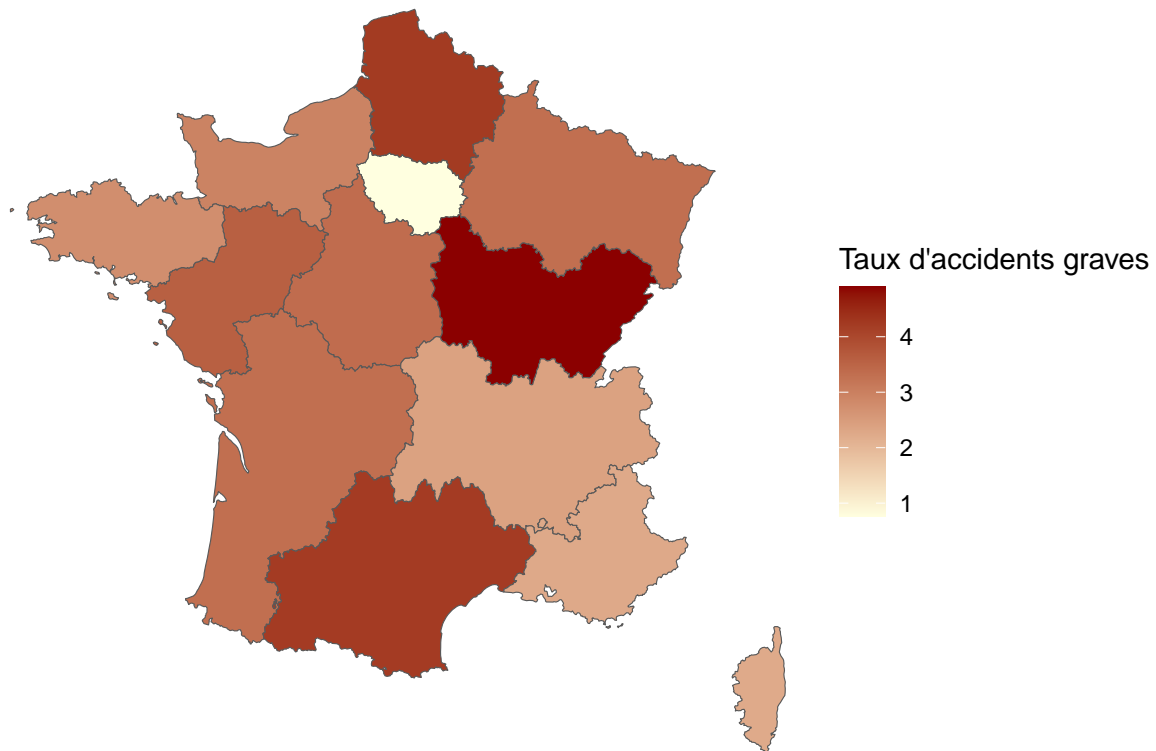
```
## Reading layer 'regions-version-simplifiee' from data source
## 'https://raw.githubusercontent.com/gregoireddavid/france-geojson/master/regions-version-simplifiee.'
## using driver 'GeoJSON'
## Simple feature collection with 13 features and 2 fields
```

```
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -5.103601 ymin: 41.36705 xmax: 9.559721 ymax: 51.0884
## Geodetic CRS:   WGS 84
```

## Nombre d'accidents par région en France en 2021



## Rapport du nombre d'accidents mortels sur le nombre d'accidents par région



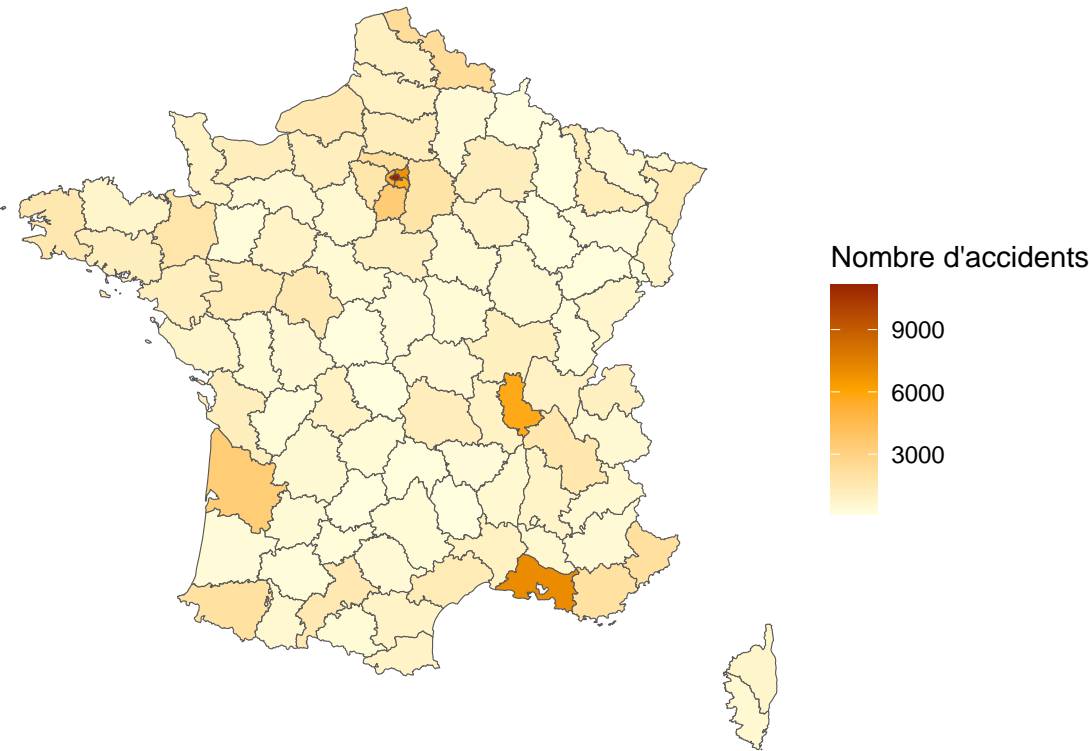
```
##  
## Pearson's Chi-squared test  
##  
## data:  cont_table  
## X-squared = 6065.5, df = 39, p-value < 2.2e-16
```

Le résultat du test montre une statistique de test de 6065.5 et un degré de liberté de 39, ce qui donne une p-value inférieure à  $2.2e-16$ .

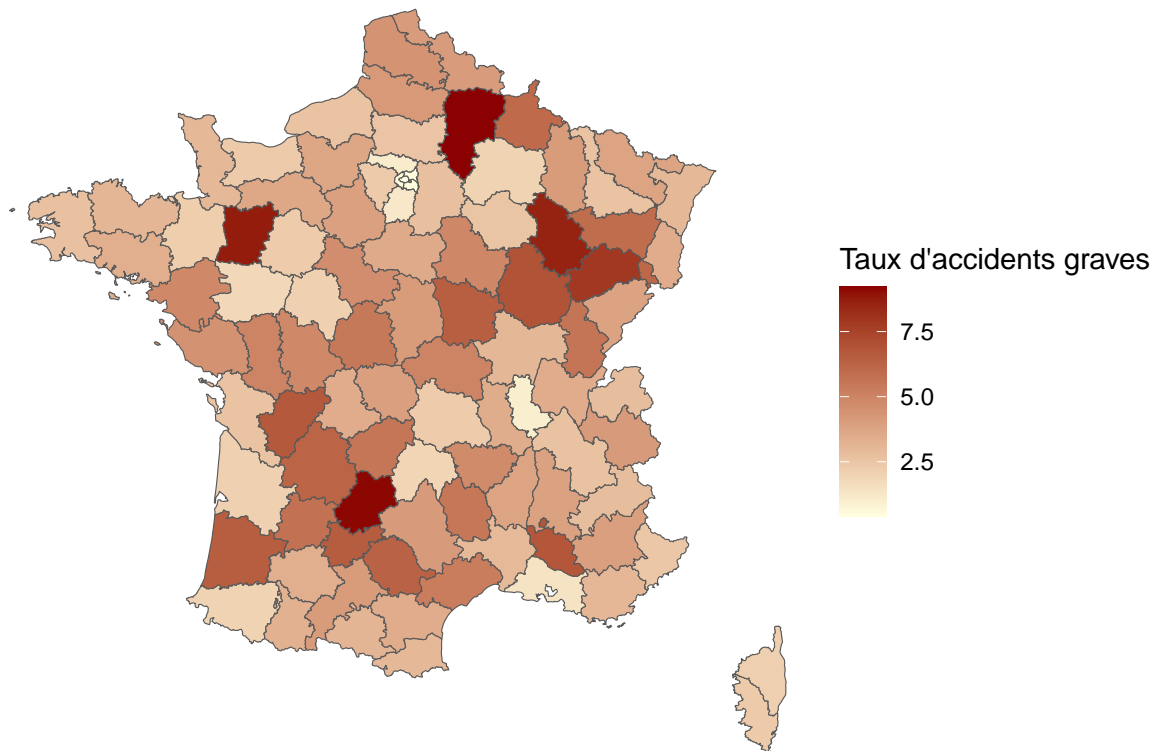
La p-value est très faible, ce qui suggère qu'il y a une forte association entre la région et la gravité de l'accident. Autrement dit, la gravité des accidents semble varier significativement selon la région où ils se produisent.

```
## Reading layer 'departements' from data source  
## 'https://france-geojson.gregoireddavid.fr/repo/departements.geojson'  
## using driver 'GeoJSON'  
## Simple feature collection with 96 features and 2 fields  
## Geometry type: MULTIPOLYGON  
## Dimension: XY  
## Bounding box: xmin: -5.138001 ymin: 41.36216 xmax: 9.559226 ymax: 51.08854  
## Geodetic CRS: WGS 84
```

Nombre d'accidents par département en France en 2021



## Rapport du nombre d'accidents mortels sur le nombre d'accidents par départen

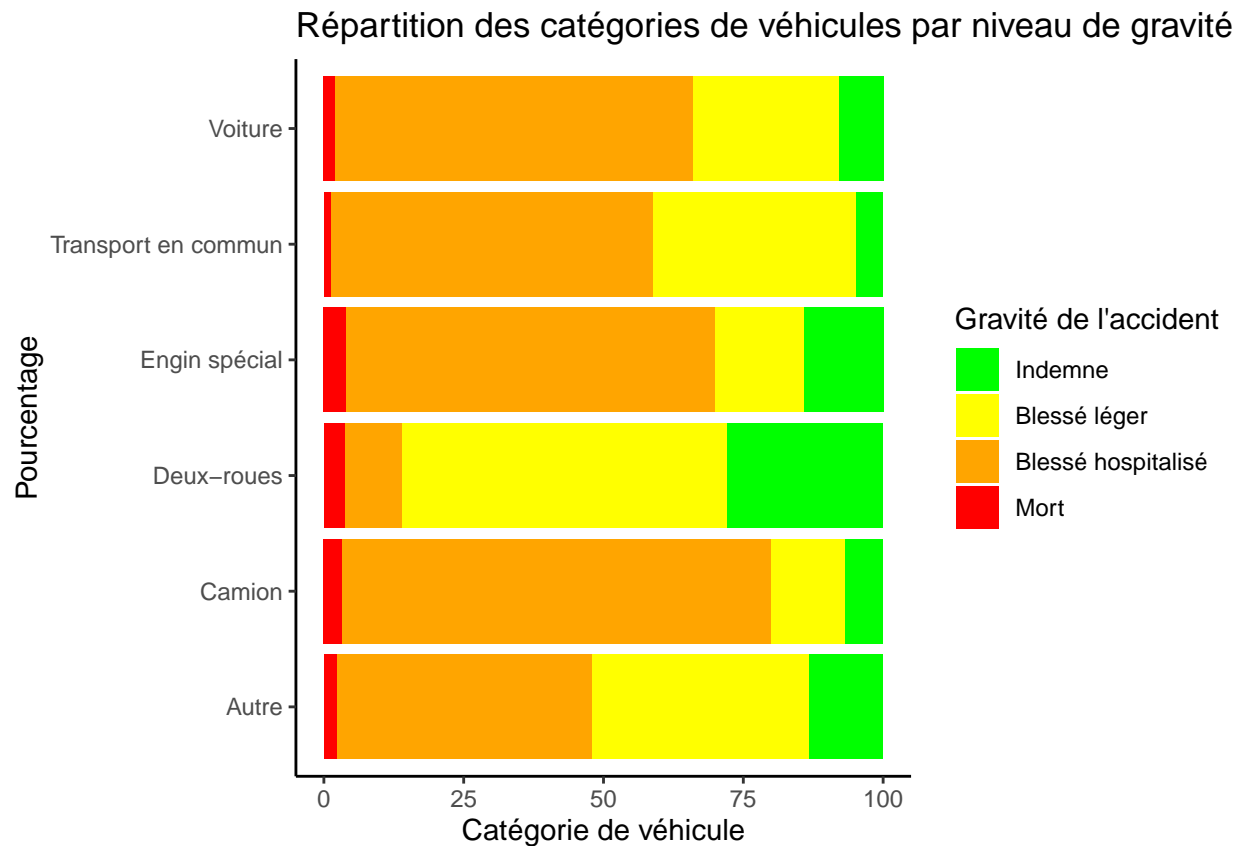


```
##  
## Pearson's Chi-squared test  
##  
## data:  cont_table  
## X-squared = 10960, df = 318, p-value < 2.2e-16
```

Dans les deux résultats, la statistique de test est très élevée (6065.5 pour les régions et 10960 pour les départements), ce qui suggère qu'il existe un lien significatif entre la région/département et la gravité de l'accident.

Est ce que l'environnement influence la gravité des accidents ?

Est ce que la catégorie de véhicule influence la gravité ?



```
##
## Call:
## glm(formula = grav_binary ~ categorie, family = binomial, data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -0.8727  -0.5807  -0.5807  -0.4624   2.3695
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -1.694609   0.008903  -190.339 < 2e-16 ***
## categorieCamion    -0.487282   0.060194   -8.095 5.72e-16 ***
## categorieDeux-roues  0.925468   0.017959   51.533 < 2e-16 ***
## categorieEngin spécial  0.184850   0.125430    1.474  0.141
## categorieTransport en commun -1.050340   0.106775   -9.837 < 2e-16 ***
## categorieVoiture   -0.513298   0.036831  -13.937 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 118832  on 129092  degrees of freedom
```

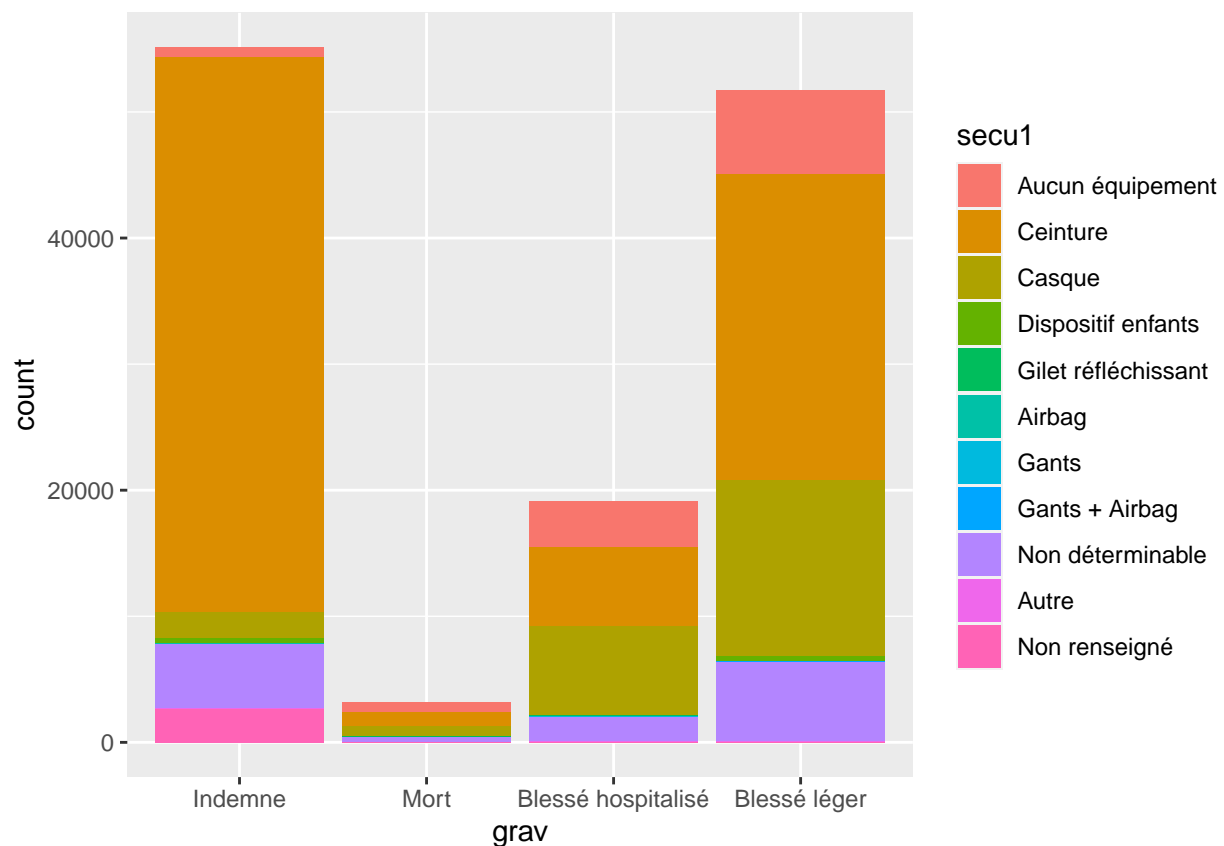
```
## Residual deviance: 115606 on 129087 degrees of freedom
## AIC: 115618
##
## Number of Fisher Scoring iterations: 5
```

La première partie de la sortie donne quelques informations sur la qualité de l'ajustement du modèle, notamment les résidus de deviance. Dans l'ensemble, les résidus semblent assez faibles, ce qui suggère que le modèle s'adapte bien aux données.

La deuxième partie de la sortie présente les coefficients estimés pour chaque niveau de la variable explicative (les différentes catégories de véhicules). Les coefficients indiquent l'effet de chaque niveau de la variable explicative sur la probabilité de gravité de l'accident. Par exemple, la variable `categorieTransport` en commun a un coefficient négatif (-1.050340), ce qui signifie qu'être impliqué dans un accident avec un transport en commun diminue la probabilité de gravité de l'accident par rapport aux autres catégories de véhicules.

En somme, ces résultats indiquent que la catégorie de véhicule est un facteur significatif pour prédire la gravité de l'accident, et que certains types de véhicules ont une probabilité de gravité plus élevée ou plus faible que d'autres.

### Est ce que l'équipement de sécurité influence la gravité ?



- Pour les véhicules :

- On regarde la gravité en fonction de la sécurité installée pour l'utilisateur (ceinture airbag pour le



- Pour l'environnement : - On regarde la gravité en fonction de la vitesse autorisée (en aggro (50km/h)n etc...) - On regarde le nb et la gravité en fonction de l'état de la route (mouillée, enneigée, sèche) - On regarde le nb et la gravité en fonction de l'éclairage de la route (nuit noire, jour, éclairage urbain etc...)

