

# Accidents corporels de la circulation routière 2018 (Work in Progress)

2020-11-23

## Contents

<b>Introduction</b>	<b>2</b>
<b>Préparation des données</b>	<b>2</b>
Table Caractéristiques : . . . . .	2
Table Lieux : . . . . .	5
Table Véhicules : . . . . .	7
Table Usagers : . . . . .	11
<b>Data Cleaning</b>	<b>13</b>
Table Caractéristiques . . . . .	13
Table Lieux . . . . .	15
<b>Axiomes</b>	<b>16</b>
<b>Quelques chiffres</b>	<b>16</b>
<b>Analyse Spatiale</b>	<b>16</b>
Carte . . . . .	16
Analyse par départements . . . . .	18
<b>Analyse temporelle</b>	<b>20</b>
Analyse sur l'année . . . . .	20
Chart 3 : Répartitions des accidents par mois . . . . .	21
Table 1 : Récapitulatif de la répartition des accidents sur l'année. . . . .	21
Analyse selon les jours de la semaine. . . . .	21
Chart 4 : Répartitions des accidents par jours . . . . .	23
Chart 5 : Répartitions des accidents mortels par jours . . . . .	24
Analyse selon l'heure de la journée . . . . .	24
Chart 6 : Répartitions des accidents selon l'heure de la journée. . . . .	25
Chart 7 : Répartitions des accidents mortels selon l'heure de la journée . . . . .	26
Chart 8 : Nombre d'accidents mortels vs Nombre d'accidents totaux selon l'heure de la journée	27
<b>Analyse des conditions routières</b>	<b>27</b>
Analyse Selon le type de route . . . . .	27
Chart 9 : Répartitions des accidents selon le type de route . . . . .	27
Type d'intersection si applicable . . . . .	28
Chart 10 : Répartitions des accidents selon le type d'intersection . . . . .	28
Le régime circulaire . . . . .	29
Chart 11 : Répartitions des accidents selon le régime circulaire . . . . .	29
Le type de tracé . . . . .	30
Chart 12 : Répartitions des accidents selon le type de tracé . . . . .	30
Le type de surface . . . . .	31

Chart 13 : Répartitions des accidents selon le type de surface . . . . .	31
Le profil de la route . . . . .	32
Chart 14 : Répartitions des accidents selon le profil de la route . . . . .	32
Analyse des conditions météorologiques . . . . .	33
Chart 15 : Répartitions des accidents selon les conditions météorologiques . . . . .	33
Selon la visibilité . . . . .	34
Chart 16 : Répartitions des accidents selon la visibilité . . . . .	34
<b>Prediction</b>	<b>35</b>
Regression logistique . . . . .	35
Model 1 . . . . .	35
Model 2 . . . . .	41
Neural Network . . . . .	43
Arbre de décision . . . . .	43

## Introduction

Ce rapport a pour but d’analyser les données concernant les accidents de la route sur l’année 2018. J’espère pouvoir . . .

La base de données est disponible sur le site [data.gouv.fr](http://data.gouv.fr)

Dans un premier temps j’analyserai globalement les données, ensuite il serait peut-être intéressant d’essayer de prédire par exemple à quel moment et à quel endroit mettre en place des mesures pour prévenir les accidents.

## Préparation des données

### Table Caractéristiques :

La table Caractéristiques décrit les circonstances générales de l’accident, chaque ligne correspond à un unique accident (il n’y pas de doublon).

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>Num_Acc</b>	Numéro d’identifiant de l’accident	Categorical	Nominal
<b>an</b>	Année de l’accident	Numerical Discrete	Interval
<b>mois</b>	Mois de l’accident	Categorical	Ordinal
<b>jour</b>	Jour de l’accident	Categorical	Ordinal
<b>hrmn</b>	Heure et minutes de l’accident	Numerical Continuous	Interval
<b>lum</b>	“Lumière” = conditions d’éclairage dans lesquelles l’accident s’est produit 1 - Plein jour 2 - Crépuscule ou aube 3 - Nuit sans éclairage public (Nuit S EP) 4 - Nuit avec éclairage public non allumé (Nuit A EP NA) 5 - Nuit avec éclairage public allumé (Nuit A EP A)	Categorical	Nominal

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>agg</b>	Localisation : 1 - Hors agglomération 2- En agglomération	Categorical	Nominal
<b>int</b>	Intersection : 1 - Hors intersection 2 - Intersection en X 3 - Intersection en T 4 - Intersection en Y 5 - Intersection à plus de 4 branches 6 - Giratoire 7 - Place 8 - Passage à niveau 9 - Autres intersection	Categorical	Nominal
<b>atm</b>	Conditions atmosphériques : 1 - Normale 2 - Pluie légère 3 - Pluie forte 4 - Neige, grêle 5 - Brouillard - fumée 6 - Vent fort, tempête 7 - Temps éblouissant 8 - Temps couvert 9 - Autre	Categorical	Nominal
<b>col</b>	Type de collision : 1 - Deux véhicules frontale 2 - Deux véhicules par l'arrière 3 - Deux véhicules par le côté 4 - Trois véhicules et plus en chaîne 5 - Trois véhicules et plus collisions multiples 6 - Autres collision 7 - Sans collision	Categorical	Nominal
<b>com</b>	Numéro de commune, code donné par l'INSEE comportant 3 chiffres calés à droite	Categorical	Nominal
<b>adr</b>	Adresse postale : variables renseignée par les accidents survenus en agglomeration	Text	
<b>gps</b>	Codage GPS (1 caractère indicateur de provenance) : M = Métropole A = Antilles (Martinique ou Guadeloupe) G = Guyane R = Réunion Y = Mayotte	Categorical	Nominal
<b>lat</b>	Latitude (degrés décimaux)	Numerical Continuous	Interval

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>long</b>	Longitude (degrés décimaux)	Numerical Continuous	Interval
<b>dep</b>	Code INSEE du département d'un 0 (201 : Corse du Sud, 202 : Haute-Corse)	Categorical	Nominal

En voyant le descriptif des données, certaines ont un type “integer” mais devraient être de type “factor”. Ensuite j’aimerais modifier la manière dont est répertorié la date; je vais créer une nouvelle donnée sous le format ISO YYYY-MM-DD hh:mm. Pour finir je vais modifier le nom des colonnes pour les rendre plus explicite.

```
## [1] "Sommaire du data"
```

```
##          Num_Acc          mois          condlumiere
## 2018000000001:    1  Octobre   : 5698  Plein_Jour      :38701
## 2018000000002:    1   Juin    : 5579  Crepuscule_Aube: 3839
## 2018000000003:    1  Septembre: 5532  Nuit_S_EP       : 5574
## 2018000000004:    1   Juillet : 5241  Nuit_A_EP_NA    :  575
## 2018000000005:    1    Mai     : 5055  Nuit_A_EP_A     : 9094
## 2018000000006:    1  Novembre : 4883
## (Other)       :57777 (Other)   :25795
##          agglo          typeintersect          condatmos
## Hors_agglomeration:20723  Hors_Intersection:38313  Normale          :45824
## En_agglomeration   :37060  Intersection_X   : 7024  Pluie_Legere     : 6350
##                   Intersection_T   : 5836  Temps_Couvert   : 2039
##                   Autres           : 2266  Pluie_Forte     : 1333
##                   Giratoire        : 2099  Temps_Eblouissant: 1038
##                   Intersection_Y   : 1186  (Other)         : 1194
##                   (Other)         : 1059  NA's           :    5
##          typecollision  numcommune  adresse          codagegps
## Autre_Collision:19522  55      : 2331  Length:57783    Metropole:55716
## 2V_Cote         :16284  7        : 694   Class :character Antilles :  597
## 2V_Arriere      : 7393  116      : 644   Mode  :character  Guyane  :  519
## 2V_Frontale     : 5211  112      : 643                   Reunion :  621
## Sans_Collision : 5185  117      : 545                   Mayotte :  280
## (Other)        : 4186  118      : 538                   NA's    :   50
## NA's           :    2  (Other):52388
##          latitude      longitude      numdept
## Min.      : 364247  Min.    :-508705  75      : 5591
## 1st Qu.:4482685  1st Qu.: 171017  13      : 3413
## Median :4780120  Median   : 242505  93      : 2735
## Mean     :4587661  Mean     : 457331  94      : 2548
## 3rd Qu.:4885140  3rd Qu.: 490700  92      : 2520
## Max.     :5542440  Max.     :6545675  69      : 2435
## NA's     :1823    NA's     :1823    (Other):38541
##          dateacci          joursem          season
## Min.      :2018-01-01 00:25:00  Fri     : 9908  Automne :15247
## 1st Qu.:2018-04-17 13:15:00  Thu     : 8511  Été     :15109
## Median :2018-07-09 00:35:00  Tue     : 8378  Hiver   :11960
## Mean     :2018-07-09 04:15:20  Wed     : 8195  Printemps:15467
## 3rd Qu.:2018-10-05 08:15:00  Sat     : 8164
```

## Max. :2018-12-31 23:50:00 (Other):14623  
 ## NA's :4 NA's : 4

### Table Lieux :

La table Lieux décrit le lieu principal de l'accident même si celui-ci s'est déroulé à une intersection, chaque ligne correspond à un unique lieu d'accident (il n'y pas de doublon).

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>Num_Acc</b>	Numéro d'identifiant de l'accident	Categorical	Nominal
<b>catr</b>	Catégorie de route : 1 - Autoroute 2 - Route Nationale 3 - Route Départementale 4 - Voie Communale 5 - Hors réseau public 6 - Parc de stationnement ouvert à la circulation publique 9 - Autre	Categorical	Nominal
<b>voie</b>	Numero de la route	Categorical	Nominal
<b>V1</b>	Indice numérique du numéro de route (ex : 2 bis, 3 ter etc...)	Text	
<b>V2</b>	Lettre indice alphanumérique de la route	Text	
<b>circ</b>	Régime de circulation : 1 - A sens unique 2 - Bidirectionnelle 3 - A chaussées séparées 4 - Avec voies d'affectation variable	Categorical	Nominal
<b>nbv</b>	Nombre total de voies de circulation	Numerical Discrete	Ratio
<b>vosp</b>	Signale l'existence d'une voie réservée, indépendamment du fait que l'accident ait lieu ou non sur cette voie : 1 - Piste cyclable 2 - Banque cyclable 3 - Voie réservée	Categorical	Nominal
<b>prof</b>	Profil décrivant la déclivité de la route à l'endroit de l'accident : 1 - Plat 2 - Pente 3 - Sommet de côte Bas de côte	Categorical	Nominal
<b>pr</b>	Numéro du PR de rattachement (numéro de la borne amont)	Categorical	Nominal

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>pr1</b>	Distance en mètres au PR (par rapport à la borne amont)	Numerical Continuous	Ratio
<b>plan</b>	Tracé en plan : 1 - Partie rectiligne 2 - En courbe à gauche 3 - En courbe à droite 4 - En "S"	Categorical	Nominal
<b>lartpc</b>	Largeur du terre plein central (TPC) s'il existe	Numerical Continuous	Ratio
<b>larrou</b>	Largeur de la chaussée affectée à la circulation des véhicules (ne sont pas compris les bandes d'arrêt d'urgence, les TPC et les places de stationnement)	Numerical Continuous	Ratio
<b>surf</b>	Etat de la surface : 1 - normale 2 - mouillée 3 - flaques 4 - inondée 5 - enneigée 6 - boue 7 - verglacée 8 - corps gras, huile 9 - autre	Categorical	Nominal
<b>infra</b>	Aménagement - Infrastructure : 1 - Souterrain, tunnel 2 - Pont, autopont 3 - Bretelle d'échangeur ou de raccordement 4 - Voie ferrée 5 - Carrefour aménagé 6 - Zone piétonne 7 - Zone de péage	Categorical	Nominal
<b>situ</b>	Situation de l'accident : 1 - Sur chaussée 2 - Sur bande d'arrêt d'urgence 3 - Sur accotement 4 - Sur trottoir 5- Sur piste cyclable	Categorical	Nominal
<b>env1</b>	point école : proximité d'une école		

J'aimerais vérifier cette dernière variable (env1) car je ne suis pas sûr de la comprendre, est-ce une valeur binaire ou une valeur continue ?

```
## [1] 0 NA 99 3
```

En regardant les données, cette variable peut prendre 4 valeurs : 0, 99, 3, NA, Je ne comprends pas à quoi elles correspondent. Par conséquent, je n'inclurai pas cette variable dans les analyses tant et aussi longtemps que je ne l'a comprendrai pas.

Comme précédemment, certaines variables devraient avoir un type factor

```
## [1] "Sommaire du data"
```

```

##          Num_Acc          Cat_Route          Num_Voie
## 201800000001:      1 Voie_Communale :26728          :21893
## 201800000002:      1 Route_Dept      :20087  0          : 6369
## 201800000003:      1 Autoroute       : 5572  1          : 849
## 201800000004:      1 Route_Nationale : 3857  86          : 834
## 201800000005:      1 Non_Renseigné   : 622   6          : 819
## 201800000006:      1 Parc_Stationnement: 423   7          : 795
## (Other)      :57777 (Other)          : 494 (Other):26224
##          Regime_Circ          Nb_Voie          Num_PR
## A_Sens_Unique      :10061 Min.      : 0.000  0          :17452
## Bidirectionnelle    :35397 1st Qu.: 2.000  1          : 2676
## Chaussees_Separees : 8781 Median : 2.000  2          : 1367
## Voies_Affect_Variable: 361 Mean   : 2.345  3          : 1077
## Non_Renseigné      :    0 3rd Qu.: 2.000  4          : 956
## NA's               : 3183 Max.    :13.000 (Other):18480
##                   NA's      :507   NA's    :15775
## Distance_PR          Exist_Voie_Reservee          Profil_Route
## Min.      : 0.0 Piste_Cyclable : 1615 Plat          :45934
## 1st Qu.: 0.0 Banque_Cyclable : 860 Pente          : 9110
## Median : 40.0 Voie_Reservee : 1774 Sommet_de_Cote: 992
## Mean   : 253.9 Non_Voie_Réservee: 0 Bas_de_Cote   : 909
## 3rd Qu.: 500.0 NA's          :53534 NA's          : 838
## Max.    :8902.0
## NA's    :16048
##          Trace_Plan          Largeur_TPC          Largeur_Chause          Etat_Surface
## Partie_Rectiligne:45638 Min.      : 0.00 Min.      : 0.00 Normale :45726
## En_Courbe_Gauche : 5220 1st Qu.: 0.00 1st Qu.: 0.00 Mouillee :10247
## En_Courbe_Droite : 4909 Median : 0.00 Median : 0.00 Autres  : 312
## En_S              : 731 Mean   : 3.59 Mean   : 39.05 Verglacee: 244
## NA's              : 1285 3rd Qu.: 0.00 3rd Qu.: 67.00 Enneigee : 158
##                   Max.    :714.00 Max.    :800.00 (Other) : 296
##                   NA's    :44490 NA's    :44074 NA's    : 800
##          Infrastrucutre          Situation_Acc
## Carrefour_Amenage : 3489 Chaussee          :31995
## Echangeur_Raccordement: 1083 Bande_Arret_Urgence: 863
## Pont_Autopont     : 898 Accotement          :22680
## Souterrain_Tunnel : 623 Trottoir           : 686
## Zone_Pietonne     : 546 Piste_Cyclable      : 317
## (Other)           : 261 NA's                : 1242
## NA's              :50883

```

## Table Véhicules :

La table Véhicules décrit les véhicules impliqués. Plusieurs véhicules peuvent être impliqué par accident.

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>Num_Acc</b>	Numéro d'identifiant de l'accident	Categorical	Nominal

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>Num_Veh</b>	Identifiant du véhicule repris pour chacun des usagers occupant ce véhicules (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés)	Categorical	Nominal
<b>senc</b>	Sens de circulation : 1 - PK ou PR ou numéro adresse postale croissant 2 - PK ou PR ou numéro adresse postale décroissant	Categorical	Nominal
<b>catv</b>	Catégorie du véhicule : 01 - Bicyclette 02 - Cyclomoteur < 50 cm3 03 - Voiturette (Quadricycle à moteur carrossé) 07 - VL seul 10 - VU seul 1.5T <= PTAC <= 3.5T 13 - PL seul 3.5T < PTCA <= 7.5T 14 - PL seul > 7,5T 15 - PL > 3.5T + remorque 16 - Tracteur routier seul 17 - Tracteur routier + semi-remorque 20 - Engin spécial 21 - Tracteur agricole 30 - Scooter < 50 cm3 31 - Motocyclette > 50 cm3 et <= 125 cm3 32 - Scooter > 50 cm3 et <= 125 cm3 33 - Motocyclette > 125 cm3 34 - Scooter > 125 cm3 35 - Quad léger <= 50 cm3 36 - Quad lourd > 50 cm3 37 - Autobus 38 - Autocar 39 - Train 40 - Tramway 99 - Autre véhicule (dont piéton en roller ou en trottinette)	Categorical	Nominal



<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>obs</b>	Obstacle fixe heurté : 1 - Véhicule en stationnement 2 - Arbre 3 - Glissière métallique 4 - Glissière béton Autre glissière 6 - Bâtiment, mur, pile de pont 7 - Support signalisation verticale ou poste d'appel d'urgence 8 - Poteau 9 - Mobilier urbain 10 - Parapet 11 - Ilot, refuge, borne haute 12 - Bordure de trottoir 13 - Fossé, talus, paroi rocheuse 14 - Autre obstacle fixe sur chaussée 15 - Autre obstacle fixe sur trottoir ou accotement 16 - Sortie de chaussée sans obstacle	Categorical	Nominal
<b>obsm</b>	Obstacle mobile heurté : 1 - Piéton 2 - Véhicule 4 - Véhicule sur rail 5 - Animal domestique 6 - Animal sauvage 9 - Autre	Categorical	Nominal
<b>choc</b>	Point de choc initial : 1 - Avant 2 - Avant droit 3 - Avant gauche 4 - Arrière 5 - Arrière droit 6 - Arrière gauche 7 - Côté droit 8 - Côté gauche 9 - Chocs multiples (tonneaux)	Categorical	Nominal

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>manv</b>	Manoeuvre principal avant l'accident (MPAA) : 1 - Sans changement de direction 2 - Même sens, même file 3 - Entre 2 files 4 - En marche arrière 5 - A contresens 6 - En franchissant le terre-plein central 7 - Dans le couloir bus, dans le même sens 8 - Dans le couloir de bus, dans le sens inverse 9 - En s'insérant 10 - En faisant demi-tour sur la chaussée Changeant de file : 11 - À gauche 12 - À droite Déporté : 13 - À gauche 14 - À droite Tournant : 15 - À gauche 16 - À droite Dépassant : 17 - À gauche 18 - À droite Divers : 19 - Traversant la chaussée 20 - Manoeuvre de stationnement 21 - Manoeuvre d'évitement 22 - Ouverture de porte 23 - Arrêté (hors stationnement) 24 - En stationnement (avec occupants)	Categorical	Nominal
<b>occutc</b>	Nombre d'occupants dans le transport en commun	Numerical Continuous	Ratio

Comme pour les autres rebriques, je vais transformer quelques variables en factor.

```
## [1] "Sommaire du data"
```

```
##           Num_Acc           Sens_Circ           Cat_Veh
## 201800005731: 13   Croissant :47358   VL seul           :60384
## 201800007042: 12   Decroissant:33040   Motocyclette > 125 cm3 : 8277
## 201800020085: 12   NA's           :18478   VU seul 1.5T <= PTAC <= 3.5T: 6390
## 201800006070: 10                                     Bicyclette           : 4967
## 201800021773: 10                                     Cyclomoteur < 50 cm3  : 3967
## 201800006038: 9                                     Scooter < 50 cm3     : 2740
## (Other)      :98810                                (Other)              :12151
##   Nb_Occupant                                Obstacle_Fixe                                Obstacle_Mobile
## Min.      : 0.0000   Vehicule_en_stationnement: 2255   Pion           :10261
## 1st Qu.: 0.0000   Fosse_talus           : 1884   Vehicule       :69438
## Median : 0.0000   Arbre                 : 1502   Vehicule_sur_rail: 101
## Mean    : 0.1481   Glissiere_metallique  : 1317   Animal_domestique: 85
```

```

## 3rd Qu.: 0.0000  Glissiere_beton      : 1298  Animal_sauvage : 662
## Max. :900.0000  (Other)              : 5860  Autre          : 1007
##              NA's                :84760  NA's          :17322
##              Point_Choc              Manoeuvre      Num_Veh
## Avant      :35421  MPAA_ss_chgmt_direction:39560  A01      :56736
## Avant_gauche:14766  MPAA_mm_sens_mm_file   :12425  B01      :34577
## Avant_droit :12117  Tournant_AG             : 8089  C01      : 4317
## Arriere     :10136  Deporte_AG              : 4652  Z01      : 1700
## Cote_gauche : 7103  Depassant_AG            : 3374  D01      : 950
## (Other)     :13939  (Other)                 :22842  E01      : 239
## NA's        : 5394  NA's                    : 7934  (Other): 357

```

## Table Usagers :

La table Usagers décrit les usagers impliqués. Plusieurs usagers peuvent avoir été impliqué dans l'accident.

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>Num_Acc</b>	Numéro d'identifiant de l'accident	Categorical	Nominal
<b>Num_Veh</b>	Identifiant du véhicule repris pour chacun des usagers occupant ce véhicules (y compris les piétons qui sont rattachés aux véhicules qui les ont heurtés)	Categorical	Nominal
<b>place</b>	Permet de situer la place occupée dans le véhicule par l'utilisateur au moment de l'accident (valeur de 1 à 9)*	Categorical	Nominal
<b>catu</b>	Catégorie d'utilisateur : 1 - Conducteur 2 - Passager 3 - Piéton 4 - Piéton en roller ou en trottinette	Categorical	Nominal
<b>grav</b>	Gravité de l'accident : 1 - Indemne 2 - Tué 3 - Blessé hospitalisé 4 - Blessé léger	Categorical	Nominal
<b>sexe</b>	Sexe de l'utilisateur : 1 - Masculin 2 - Féminin	Categorical	Nominal
<b>An_nais</b>	Année de naissance de l'utilisateur	Numerical discrete	Ratio
<b>trajet</b>	Motif du déplacement au moment de l'accident : 1 - Domicile_Travail 2 - Domicile_École 3 - Courses_Achats 4 - Utilisation_Professionnelle 5 - Promenade_Loisirs 9 - Autre	Categorical	Nominal

<i>Variables</i>	<i>Description</i>	<i>Type of Data</i>	<i>Level of Measurement</i>
<b>secu</b>	sur 2 caractères : Le premier concerne l'existence d'un équipement de sécurité : 1 - Ceinture 2 - Casque 3 - Dispositif enfants 4 - Equipement réfléchissant 9 - Autre Le second concerne l'utilisation de l'équipement de sécurité : 1 - Oui 2 - Non 3 - Non déterminable	Categorical	Nominal
<b>locp</b>	Localisation du piéton : Sur chaussée (SC) : 1 - A > 50m du passage piéton (PP) 2 - A < 50m du passage piéton (PP) Sur passage piéton (SPP) : 3 - Sans signalisation lumineuse (SSL) 4 - Avec signalisation lumineuse (ASL) Divers (D) : 5 - Sur trottoir 6 - Sur accotement 7 - Sur refuge ou BAU 8 - Sur contre allée	Categorical	Nominal
<b>actp</b>	Action du piéton : 0 - Déplacement non renseigné 1 - Déplacement dans le sens du véhicule heurtant 2 - Déplacement dans le sens inverse du véhicule heurtant 3 - Traversant 4 - Masqué 5 - Jouant_Courant 6 - Avec animal 9 - Autre	Categorical	Nominal
<b>etatp</b>	Cette variable permet de préciser si le piéton accidenté était seul ou non : 1 - Seul 2 - Accompagné 3 - En groupe	Categorical	Nominal

\*Mettre une photo des dispositions

Comme pour les autres rebriques, je vais transformer quelques variables en factor. Aussi, je vais séparer la variable **secu** en deux. La première variable concernera l'existence ou non d'équipements de sécurité. La deuxième concernera l'utilisation de ces équipements.

```
## [1] "Sommaire du data"
```

```
##          Num_Acc          Num_Place          Cat_Usager
```

```

## 201800017680:    51    1      :96011  Conducteur:96025
## 201800019383:    46    2      :14465  Passenger  :23184
## 201800013008:    44    3      : 2475  Piéton     :10960
## 201800029284:    39    4      : 2049
## 201800006589:    38    9      : 1435
## 201800006560:    37    (Other): 2774
## (Other)         :129914  NA's    :10960
##               Gravite      Sexe
## Indemne         :54248  Masculin:88218
## Tué             : 3392  Féminin :41951
## Blessé_Hospitalisé:22169
## Blessé_Léger     :50360
##
##
##
##               Type_Trajet      secu      Localisation_Pieton
## Domicile_Travail      :18015  Min.    :11.00  PP_SSL      : 3225
## Domicile_École         : 3174  1st Qu.:11.00  Chaussée_<50m_PP : 2483
## Courses_Achats         : 3758  Median  :11.00  PP_AS_L      : 1739
## Utilisation_Professionnelle:12723  Mean    :15.29  Chaussée_>=50m_PP: 1452
## Promenade_Loisirs      :52671  3rd Qu.:21.00  Sur_Trottoir    : 694
## Autre                  :11372  Max.    :93.00  (Other)         : 450
## NA's                   :28456  NA's    :14094  NA's           :120126
##               Action_Pieton      Etat_Pieton      An_Naiss
## Traversant             : 7928  Seul     : 7851  Min.          :1911
## Autre                   : 760  Accompagné: 2172  1st Qu.:1966
## Depl_Sens_Vehicule      : 629  En_Groupe : 480  Median        :1982
## Jouant_Courant          : 484  NA's      :119666  Mean          :1979
## Depl_Sens_Inverse_Vehicule: 293  3rd Qu.:1994
## (Other)                 : 209  Max.      :2018
## NA's                   :119866  NA's      :30
##   Num_Veh      Equipement_Seurite Utilisation_Equipement
## Length:130169  Length:130169      Length:130169
## Class :character  Class :character  Class :character
## Mode  :character  Mode  :character  Mode  :character
##
##
##
##

```

## Data Cleaning

### Table Caractéristiques

```
## [1] "Sommaire Table Caractéristiques"
```

```

##               Num_Acc      mois      condlumiere
## 201800000001:    1  Octobre : 5698  Plein_Jour      :38701
## 201800000002:    1   Juin   : 5579  Crepuscule_Aube: 3839
## 201800000003:    1  Septembre: 5532  Nuit_S_EP       : 5574
## 201800000004:    1   Juillet : 5241  Nuit_A_EP_NA    : 575
## 201800000005:    1    Mai    : 5055  Nuit_A_EP_A     : 9094
## 201800000006:    1  Novembre : 4883
## (Other)       :57777  (Other) :25795

```

```

##                agglo                typeintersect                condatmos
## Hors_agglomeration:20723 Hors_Intersection:38313 Normale :45824
## En_agglomeration :37060 Intersection_X : 7024 Pluie_Legere : 6350
## Intersection_T : 5836 Temps_Couvert : 2039
## Autres : 2266 Pluie_Forte : 1333
## Giratoire : 2099 Temps_Eblouissant: 1038
## Intersection_Y : 1186 (Other) : 1194
## (Other) : 1059 NA's : 5
##                typecollision numcommune adresse                codagegps
## Autre_Collision:19522 55 : 2331 Length:57783 Metropole:55716
## 2V_Cote :16284 7 : 694 Class :character Antilles : 597
## 2V_Arriere : 7393 116 : 644 Mode :character Guyane : 519
## 2V_Frontale : 5211 112 : 643 Reunion : 621
## Sans_Collision : 5185 117 : 545 Mayotte : 280
## (Other) : 4186 118 : 538 NA's : 50
## NA's : 2 (Other):52388
## latitude longitude numdept
## Min. : 364247 Min. : -508705 75 : 5591
## 1st Qu.:4482685 1st Qu.: 171017 13 : 3413
## Median :4780120 Median : 242505 93 : 2735
## Mean :4587661 Mean : 457331 94 : 2548
## 3rd Qu.:4885140 3rd Qu.: 490700 92 : 2520
## Max. :5542440 Max. :6545675 69 : 2435
## NA's :1823 NA's :1823 (Other):38541
## dateacci joursem season
## Min. :2018-01-01 00:25:00 Fri : 9908 Automne :15247
## 1st Qu.:2018-04-17 13:15:00 Thu : 8511 Été :15109
## Median :2018-07-09 00:35:00 Tue : 8378 Hiver :11960
## Mean :2018-07-09 04:15:20 Wed : 8195 Printemps:15467
## 3rd Qu.:2018-10-05 08:15:00 Sat : 8164
## Max. :2018-12-31 23:50:00 (Other):14623
## NA's :4 NA's : 4

```

En regardant le sommaire des données de cette table, il y a 4 lignes sans date d'accidents qui doivent être inclus dans l'étude car ils comportent des blessés hospitalisés.

Concernant les valeurs manquantes de la colonne "typecollision", les accidents avec les mêmes caractéristiques sont catégorisés en "Autre\_Collision".

Ensuite, les valeurs des deux colonnes "latitude" et "longitude" pourront être estimées par des valeurs de latitude et longitude associées au département, si le besoin est justifié.

Les NA's associées à la colonne "codagegps" pourront être comblées grâce au numéro de département associé.

```

## [1] 2V_Frontale Sans_Collision 2V_Cote Autre_Collision
## [5] 2V_Arriere 3V&pl_Multiple 3V&pl_Chaine <NA>
## 7 Levels: 2V_Frontale 2V_Arriere 2V_Cote 3V&pl_Chaine ... Sans_Collision
##
## 2V_Frontale 2V_Arriere 2V_Cote 3V&pl_Chaine 3V&pl_Multiple
## 21 35 29 4 11
## Autre_Collision Sans_Collision
## 9964 158
## [1] "Normale" "Pluie_Legere" "Pluie_Forte"
## [4] "Neige_Grele" "Brouillard_Fumee" "Vent_Fort_Tempete"
## [7] "Temps_Eblouissant" "Temps_Couvert" "Autres"

```

## Table Lieux

```

##          Num_Acc          Cat_Route          Num_Voie
## 201800000001:      1 Voie_Communale :26728          :21893
## 201800000002:      1 Route_Dept      :20087      0      : 6369
## 201800000003:      1 Autoroute       : 5572      1      : 849
## 201800000004:      1 Route_Nationale : 3857      86      : 834
## 201800000005:      1 Non_Renseigné   : 622       6      : 819
## 201800000006:      1 Parc_Stationnement: 423       7      : 795
## (Other)      :57777 (Other)         : 494 (Other):26224
##          Regime_Circ          Nb_Voie          Num_PR
## A_Sens_Unique :10061 Min. : 0.000 0 :17452
## Bidirectionnelle :35397 1st Qu.: 2.000 1 : 2676
## Chaussees_Separees : 8781 Median : 2.000 2 : 1367
## Voies_Affect_Variable: 361 Mean : 2.345 3 : 1077
## Non_Renseigné : 0 3rd Qu.: 2.000 4 : 956
## NA's : 3183 Max. :13.000 (Other):18480
##          NA's :507 NA's :15775
## Distance_PR          Exist_Voie_Reservee          Profil_Route
## Min. : 0.0 Piste_Cyclable : 1615 Plat :45934
## 1st Qu.: 0.0 Banque_Cyclable : 860 Pente : 9110
## Median : 40.0 Voie_Reservee : 1774 Sommet_de_Cote: 992
## Mean : 253.9 Non_Voie_Réservee: 0 Bas_de_Cote : 909
## 3rd Qu.: 500.0 NA's :53534 NA's : 838
## Max. :8902.0
## NA's :16048
##          Trace_Plan          Largeur_TPC          Largeur_Chausee          Etat_Surface
## Partie_Rectiligne:45638 Min. : 0.00 Min. : 0.00 Normale :45726
## En_Courbe_Gauche : 5220 1st Qu.: 0.00 1st Qu.: 0.00 Mouillee :10247
## En_Courbe_Droite : 4909 Median : 0.00 Median : 0.00 Autres : 312
## En_S : 731 Mean : 3.59 Mean : 39.05 Verglacee: 244
## NA's : 1285 3rd Qu.: 0.00 3rd Qu.: 67.00 Enneigee : 158
##          Max. :714.00 Max. :800.00 (Other) : 296
##          NA's :44490 NA's :44074 NA's : 800
##          Infrastrucutre          Situation_Acc
## Carrefour_Amenage : 3489 Chaussee :31995
## Echangeur_Raccordement: 1083 Bande_Arret_Urgence: 863
## Pont_Autopont : 898 Accotement :22680
## Souterrain_Tunnel : 623 Trottoir : 686
## Zone_Pietonne : 546 Piste_Cyclable : 317
## (Other) : 261 NA's : 1242
## NA's :50883
## [1] "Autoroute" "Route_Nationale" "Route_Dept"
## [4] "Voie_Communale" "Hors_Res_Public" "Parc_Stationnement"
## [7] "Autres" "Non_Renseigné"
## [1] "A_Sens_Unique" "Bidirectionnelle" "Chaussees_Separees"
## [4] "Voies_Affect_Variable" "Non_Renseigné"
## [1] "Plat" "Pente" "Sommet_de_Cote" "Bas_de_Cote"
## [1] "Partie_Rectiligne" "En_Courbe_Gauche" "En_Courbe_Droite"
## [4] "En_S"
## [1] "Normale" "Mouillee" "Flaques" "Innondee"
## [5] "Enneigee" "Boue" "Verglacee" "Corps_Gras_Huile"

```

## [9] "Autres"

## Axiomes

- Pour rendre les résultats plus compréhensible, j'ai pris la liberté d'arrondir les nombres décimaux à l'entier supérieur.  
Par exemple, si la moyenne est égale à 158,4 le résultats affiché sera de 159.  
Ce choix a été motivé par la gravité même d'un accident.
- Ce rapport reste à l'échelle de la France. Toutes les analyses réalisées ci-dessous peuvent être appliquées à l'échelle départementale. - Chaque analyse est structurée ainsi: il y a une première analyse sur la globalité des accidents suivie directement par une deuxième concernant spécifiquement les accidents mortels.

## Quelques chiffres

Selon les données disponibles, il y a eu :

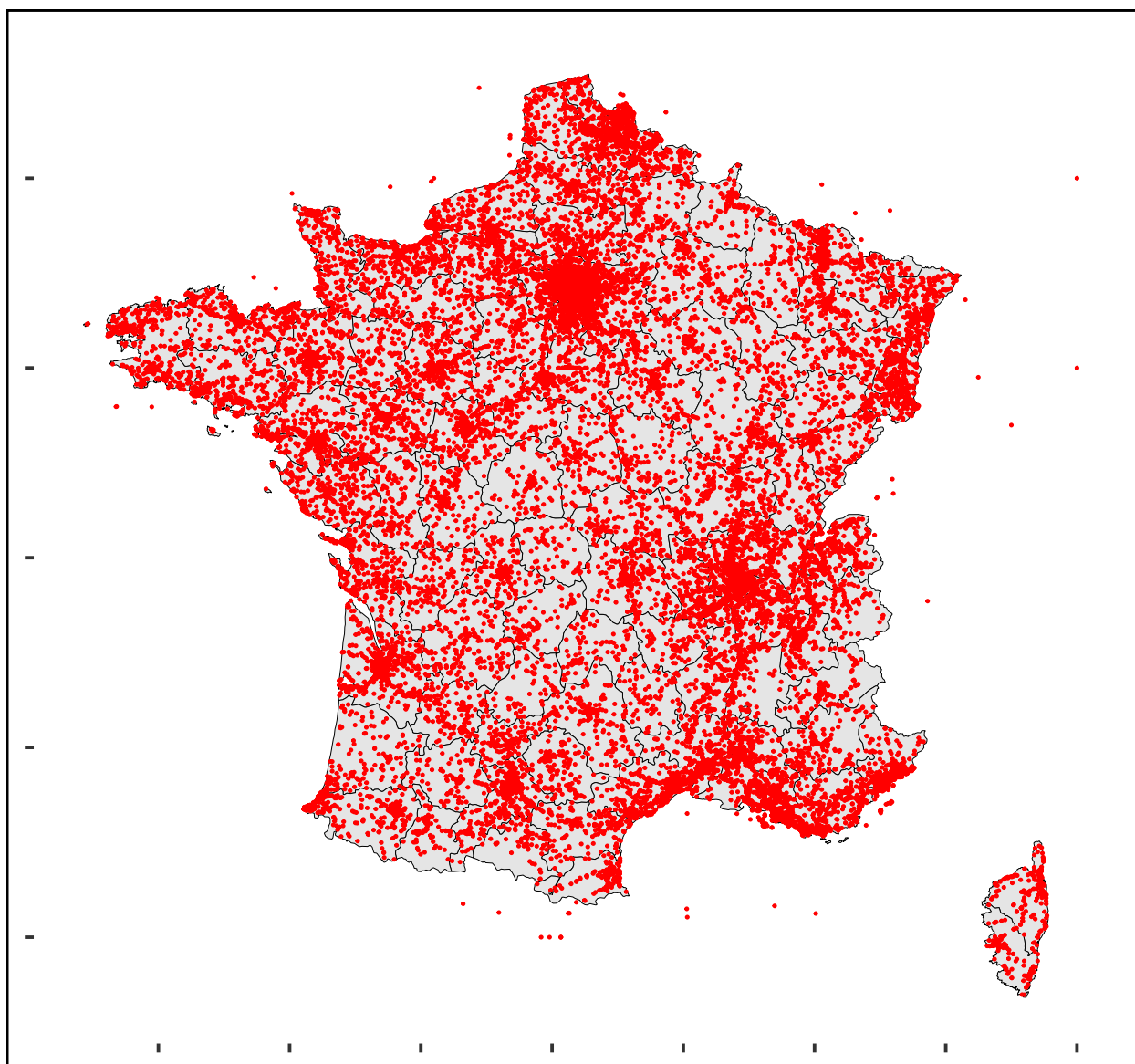
- 57783 accidents impliquant 130169 usagers, soit plus de 2 usagers impliqués par accidents en moyenne
- 54248 usagers sont sortis indemne (soit environ 41,7%)
- 72529 usagers ont été blessés (soit environ 55.7%)
- 3392 usagers sont décédés (soit environ 2.6%)
- 3186 accidents mortels sur les 57783 répertoriés (correspond à 5.5%)

## Analyse Spatiale

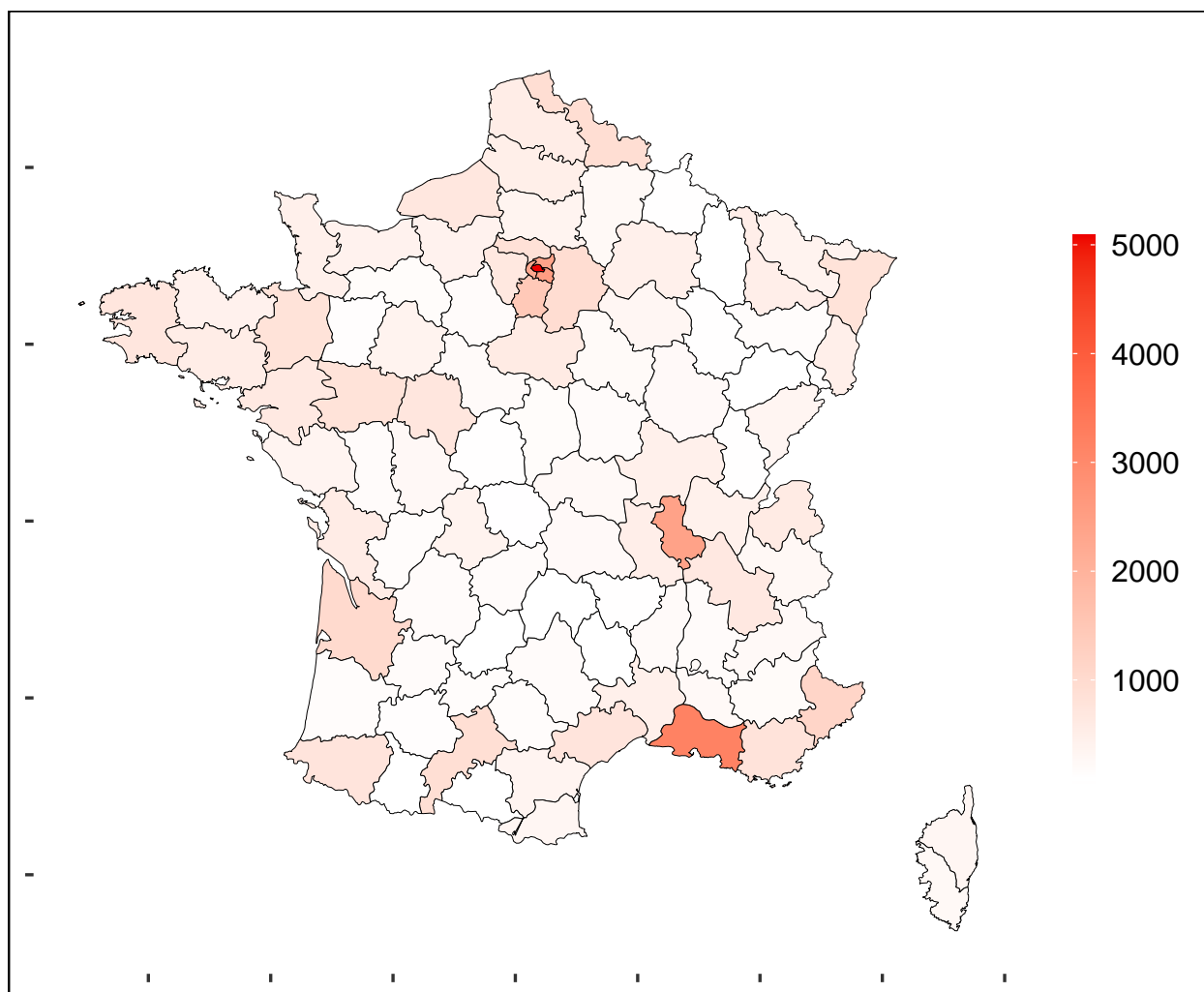
### Carte

La carte ci-dessous nous présente les coordonnées GPS des accidents sur le territoire français. Quelques points sont en dehors des frontières, il s'agit probablement d'erreurs de prises de données. La densité d'accidents dans les grandes villes et les alentours est bien visible sur la carte.





Puis, celle ci-dessous, nous permet de visualiser rapidement la répartition des accidents par départements. Il y a 3 régions critiques qui sont concernées par un grand nombre d'accidents: la région de Paris, du Rhône, et des Bouches du Rhône.



## Analyse par départements

En regardant plus précisément les départements, les régions Parisiennes et les Bouche-du-Rhône font partis du top 5 des régions comptant le plus d'accidents

Mais quand on analyse les accidents mortels, le département Bouches-du-Rhône (13) est celui qui compte le plus décès et fait partie aussi des départements comptant le plus d'accident au total. À contrario, les départements parisiens ne sont pas dans le top 5. Par conséquent, à Marseille les accidents semblent plutôt grave.

Les départements en bleu ciel sur les graphiques ci-dessous représentent les départements comptabilisant à eux seuls 50% des accidents.

**Chart 1 : Répartitions des accidents par départements**

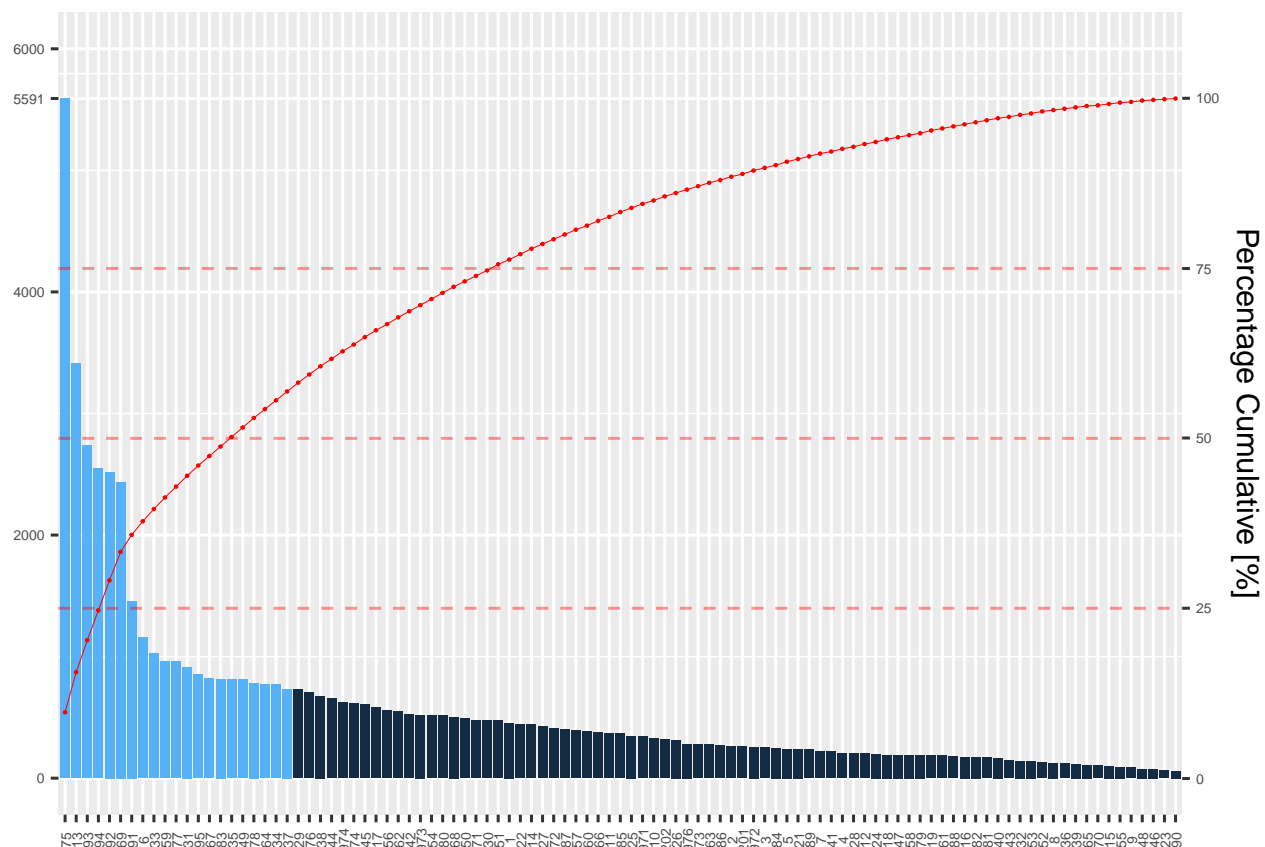
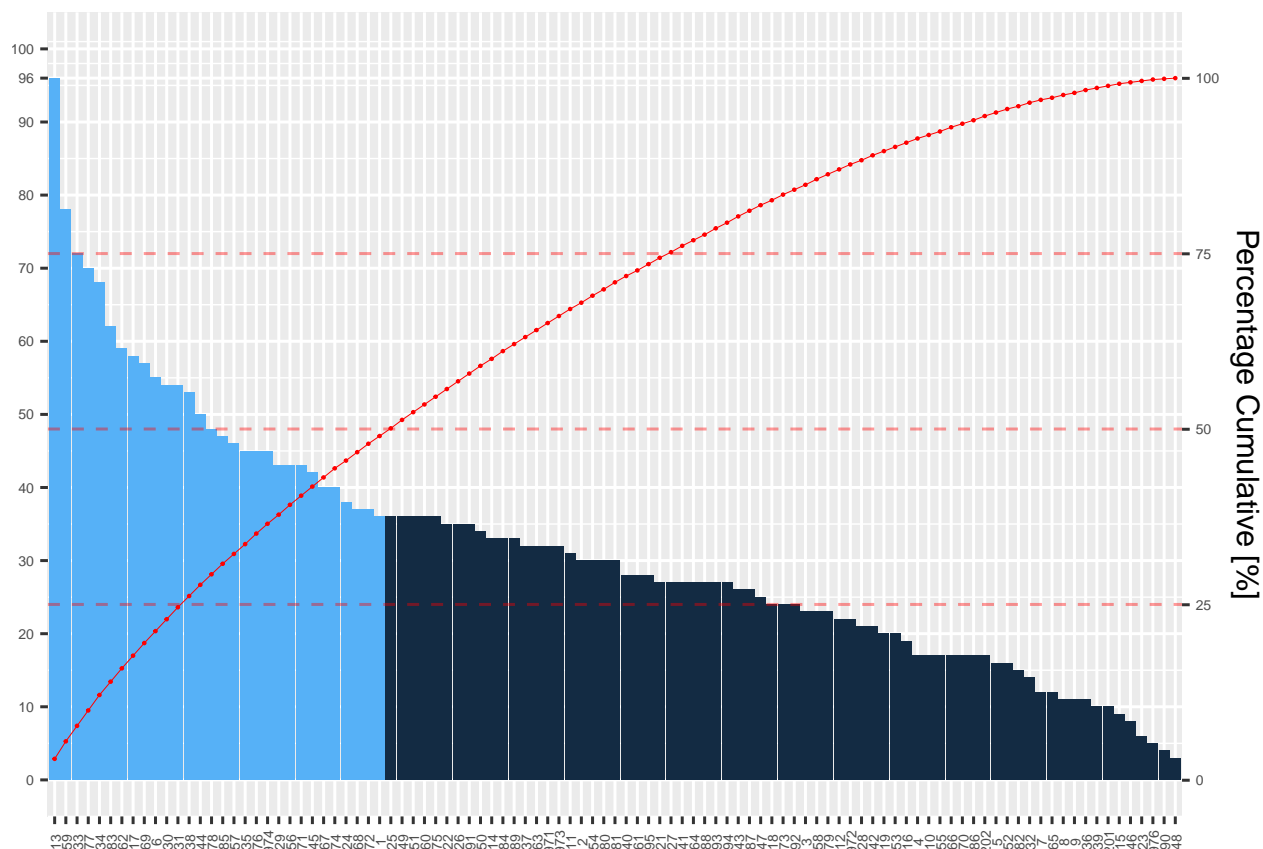


Chart 2 : Répartitions des accidents mortels par départements



## Analyse temporelle

### Analyse sur l'année

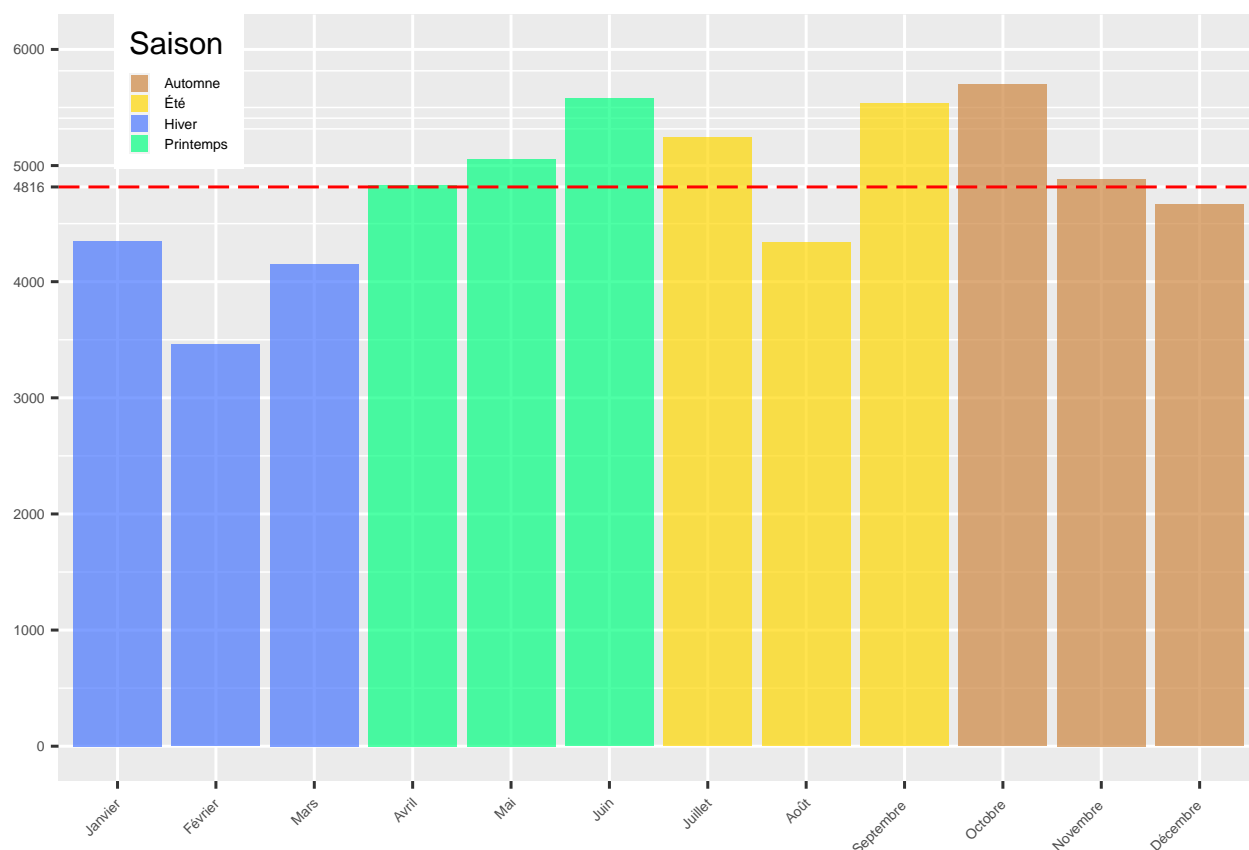
Durant l'année 2018, il y a eu 57783 accidents recensés, soit environ 4,816 accidents par mois. En visualisant le graphe Chart 3, nous pouvons ressortir quelques remarques :

- La période d'hiver compte le moins d'accidents ;
- Chacun des mois du printemps (Avril, Mai, Juin) est au-dessus de la moyenne d'accidents, avec une augmentation de mois en mois;
- En été, le mois d'Aout est un mois en baisse pour augmenter brusquement en Septembre;
- Octobre correspond au mois comptant le plus d'accident, mais avec 4.7% d'accidents mortels;
- le mois de Février et Juillet ont un pourcentage d'accidents mortels au dessus de 6%;
- Le mois de Juillet est aussi le mois avec le plus d'accidents mortels dans l'année;
- Pendant l'automne les accidents baissent progressivement.

Aussi, en se concentrant sur les accidents mortels (voir Barchart 2), nous pouvons constater que l'été est une période de l'année très meurtrière sur les routes françaises.

Un indicateur du nombre d'accidents/estimation du nombre de voitures sur les routes pourrait nous aider à nuancer ces observations. En effet, la période estivale correspond aux vacances d'été et par conséquent il y a une forte affluence de voitures, les accidents devraient augmenter en conséquence. Néanmoins c'est à vérifier car n'ayant pas les données pour le moment je ne peux pas confirmer la véracité de cette affirmation.

**Chart 3 : Répartitions des accidents par mois**



Le tableau ci-dessous résume les indicateurs disponibles par mois, groupé en saison.

**Table 1 : Récapitulatif de la répartition des accidents sur l'année.**

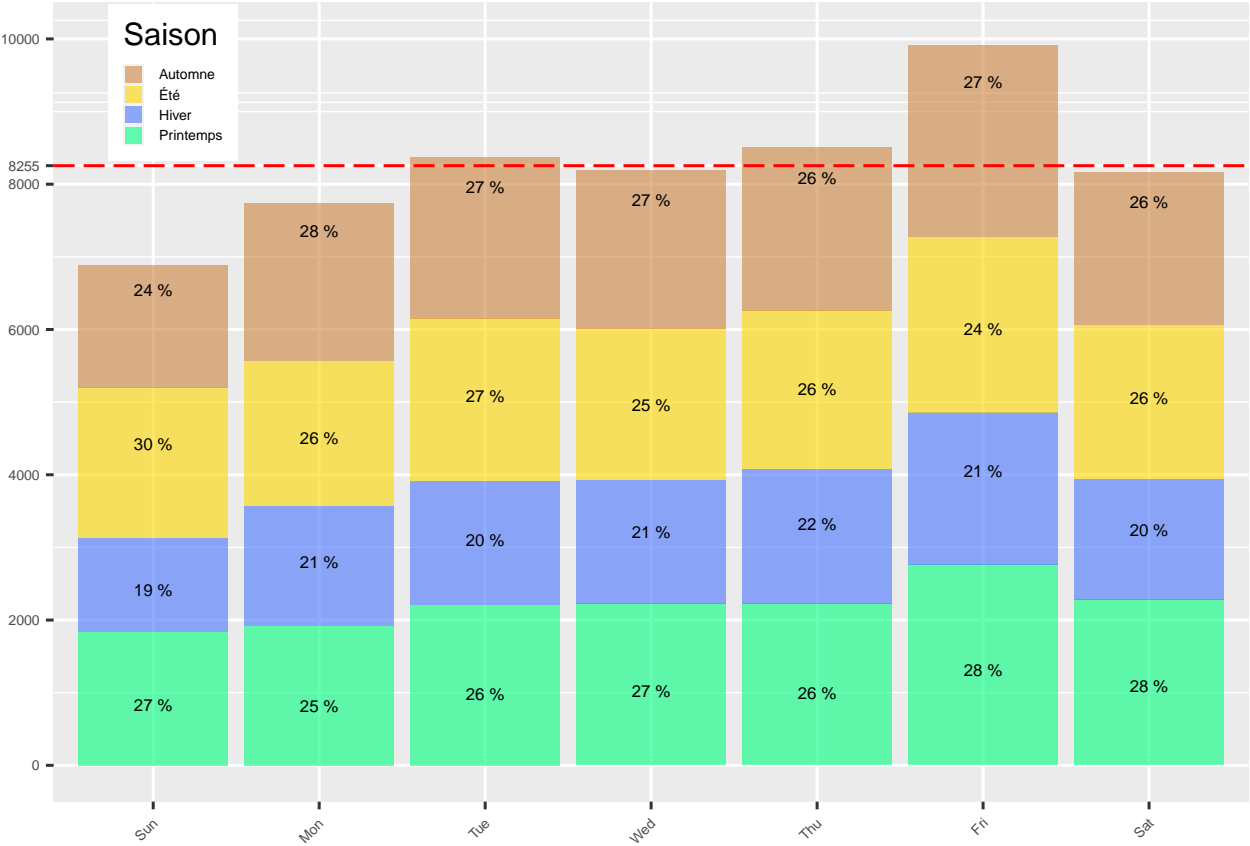
### Analyse selon les jours de la semaine.

Il y a eu en moyenne 8255 accidents au total et 456 accidents mortels par jours. D'un point de vue global le nombre d'accidents est assez régulé, sauf pour le vendredi où le nombre d'accidents augmente fortement, et le dimanche où ils sont en baisse. Mais en se concentrant sur les accidents mortels le constat est différent: en fin de semaine le nombre d'accidents mortels est supérieur à la semaine. Il semblerait donc qu'en semaine les accidents sont élevés mais sont de gravité moindre. Or, pour la fin de semaine, c'est le contraire, spécialement pour le dimanche, où il y a moins d'accidents mais ils sont plus mortels. Le samedi et dimanche ont autour de 7% d'accidents mortels, tandis que les autres jours on se situe plus autour de 5%.

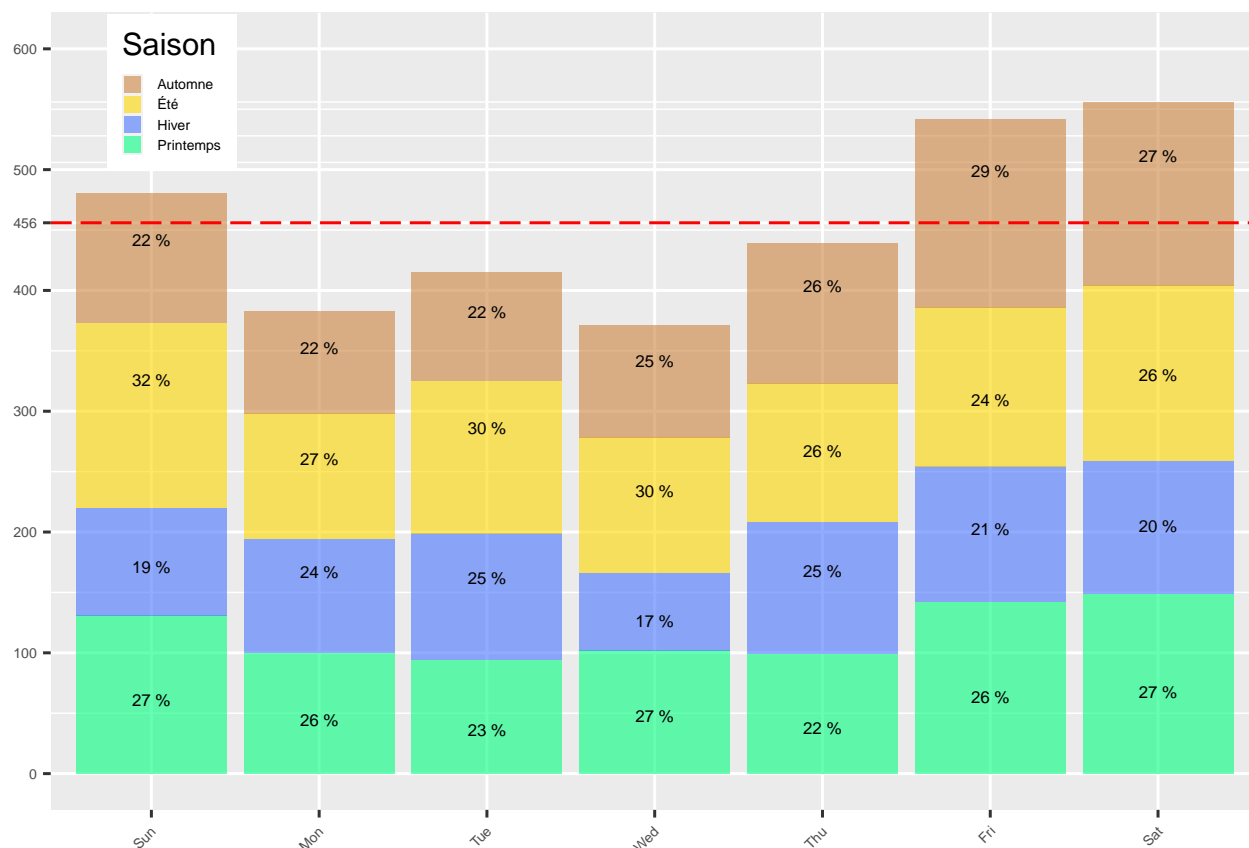
Mois	Ensemble des accidents		Accidents Mortels		% Accidents Mortels / Total Accidents du mois
	Total Accidents	% Accidents	Total Accidents Mortels	% Accidents Mortels	
Hiver					
Janvier	4350	7.53%	226	7.09%	5.20%
Février	3458	5.98%	221	6.94%	6.39%
Mars	4152	7.19%	236	7.41%	5.68%
Printemps					
Avril	4833	8.36%	275	8.63%	5.69%
Mai	5055	8.75%	254	7.97%	5.02%
Juin	5579	9.66%	288	9.04%	5.16%
Été					
Juillet	5241	9.07%	324	10.17%	6.18%
Août	4336	7.50%	246	7.72%	5.67%
Septembre	5532	9.57%	317	9.95%	5.73%
Automne					
Octobre	5698	9.86%	268	8.41%	4.70%
Novembre	4883	8.45%	252	7.91%	5.16%
Décembre	4666	8.08%	279	8.76%	5.98%

Figure 1: Tab

Chart 4 : Répartitions des accidents par jours



**Chart 5 : Répartitions des accidents mortels par jours**



### Analyse selon l'heure de la journée

Tout les jours de la semaine, il y a un premier pic le matin entre 8h et 9h et surtout un pic le soir entre 16h et 19h, correspondant aux heures de début et de fin de travail. Ce pic en fin de journée est aussi visible en fin de semaine. Le constat est pratiquement le même quant on regarde les accidents mortels, tout les jours de la semaine, il y a un pic le soir entre 16h et 19h.

On peut aussi constater que dans les nuits de vendredi à samedi et samedi à dimanche, le nombre d'accidents et le nombre d'accidents mortels doublent par rapport à la semaine.



Chart 6 : Répartitions des accidents selon l'heure de la journée.

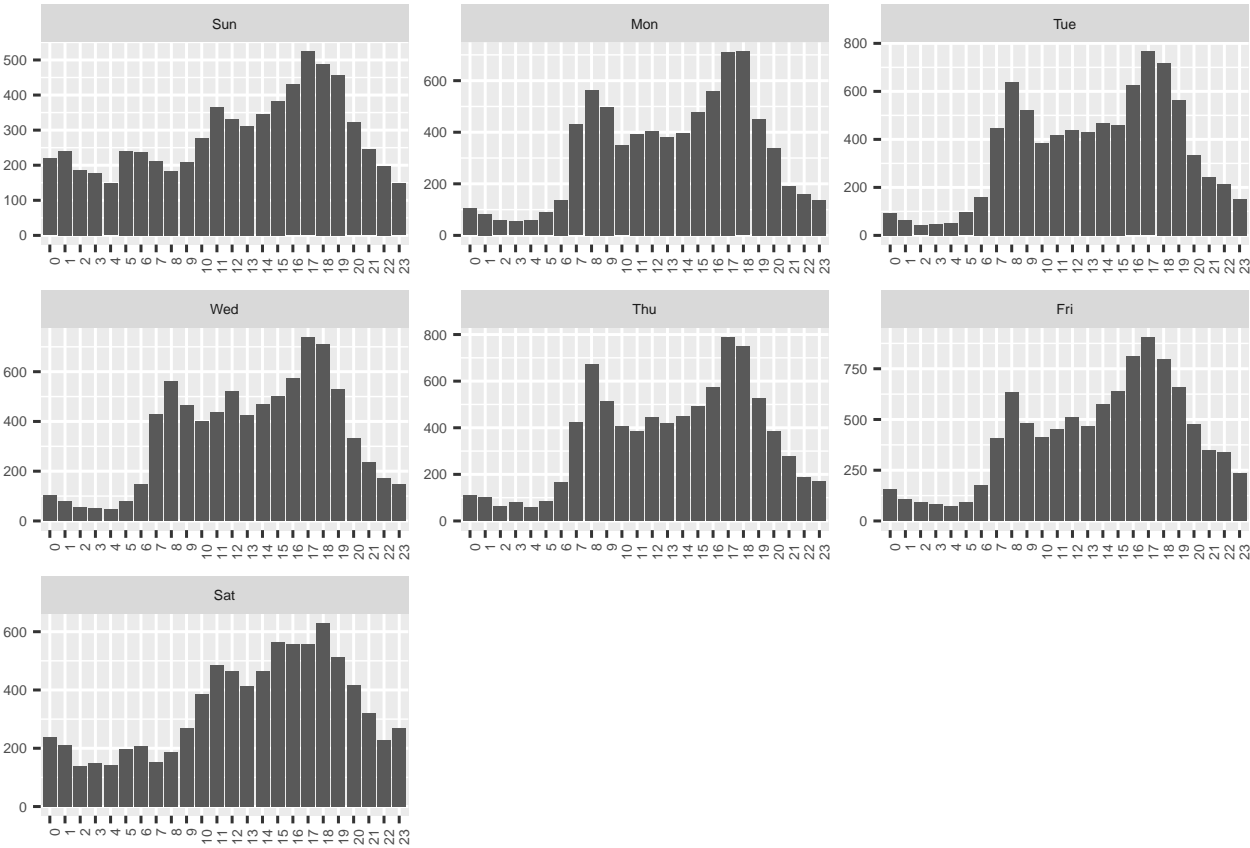
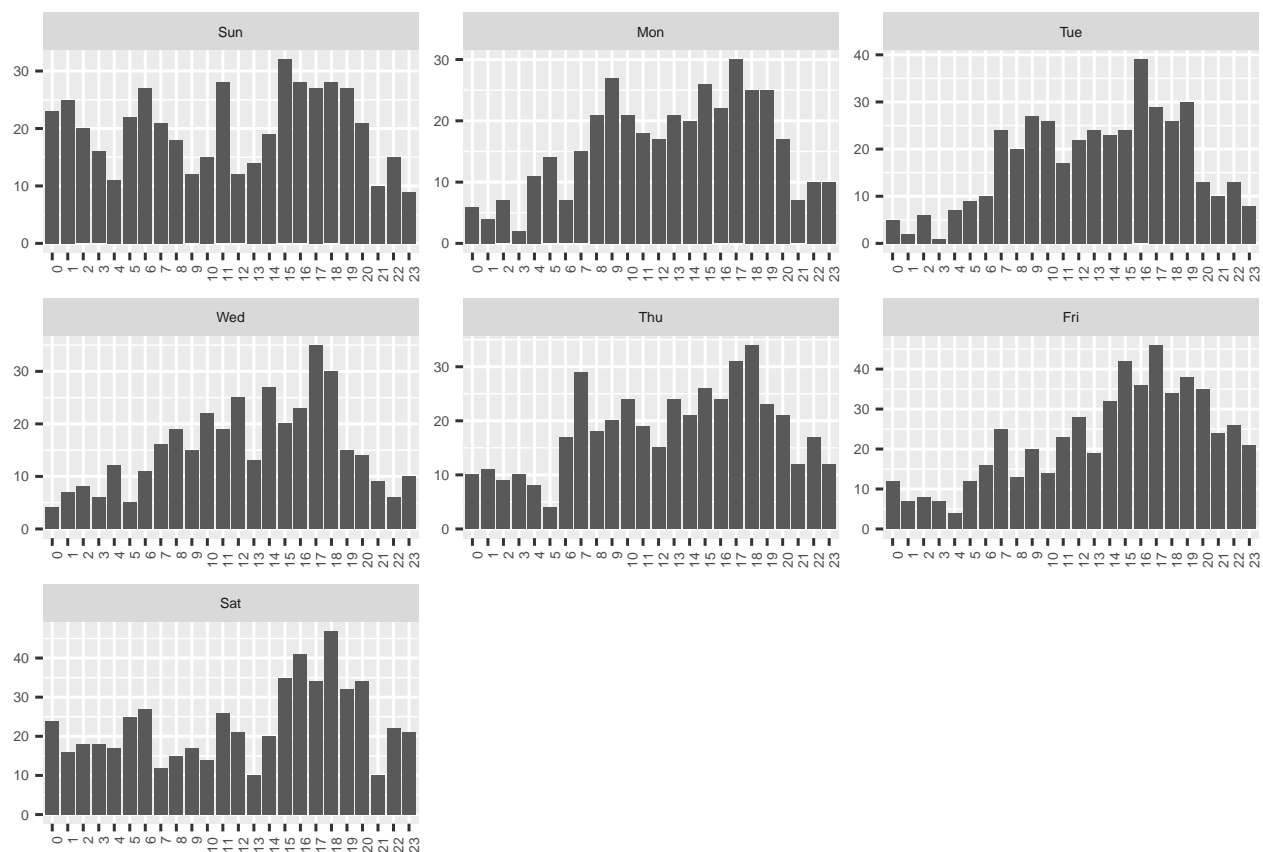
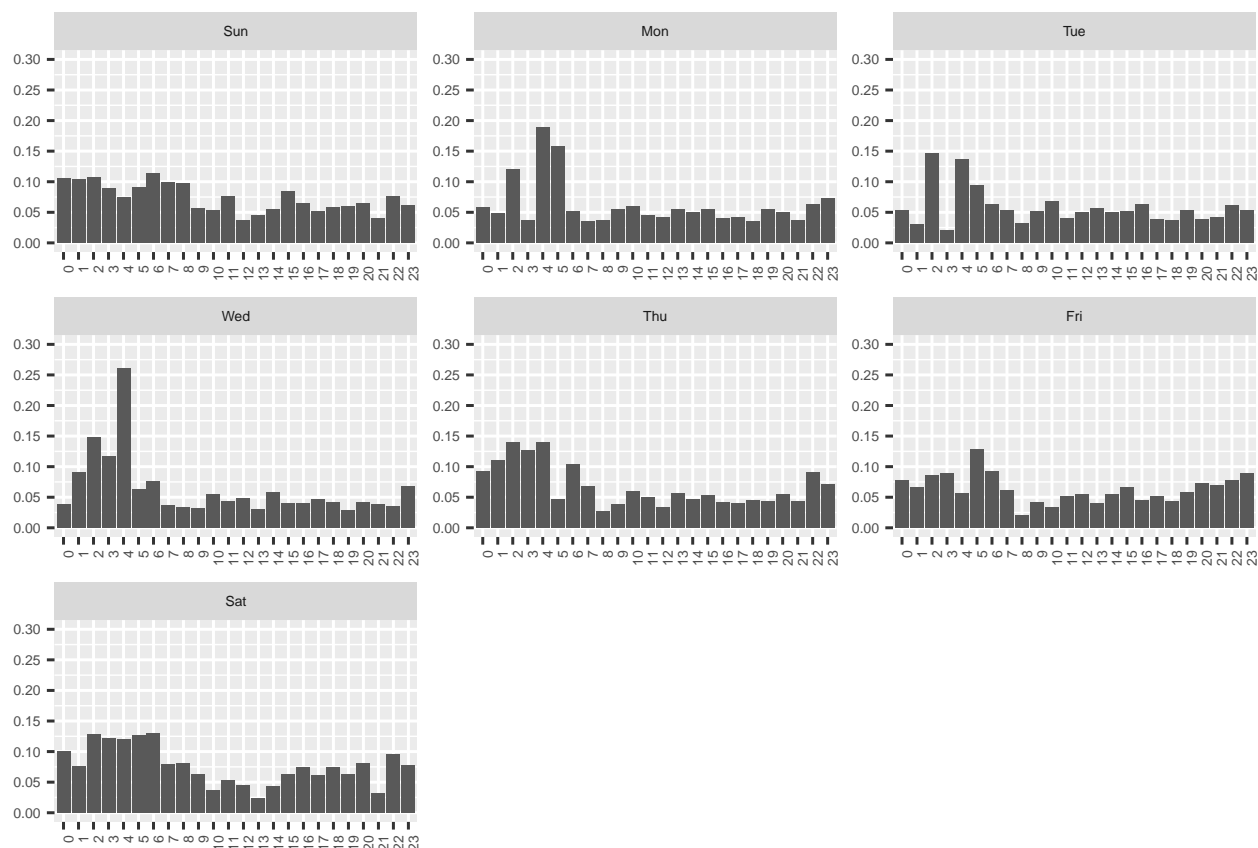


Chart 7 : Répartitions des accidents mortels selon l'heure de la journée



de plus, grâce au graphique ci-dessous, nous pouvons observer que les accidents survenant la nuit ont 10% et plus de mortalité, spécialement la fin de semaine

**Chart 8 : Nombre d'accidents mortels vs Nombre d'accidents totaux selon l'heure de la journée**



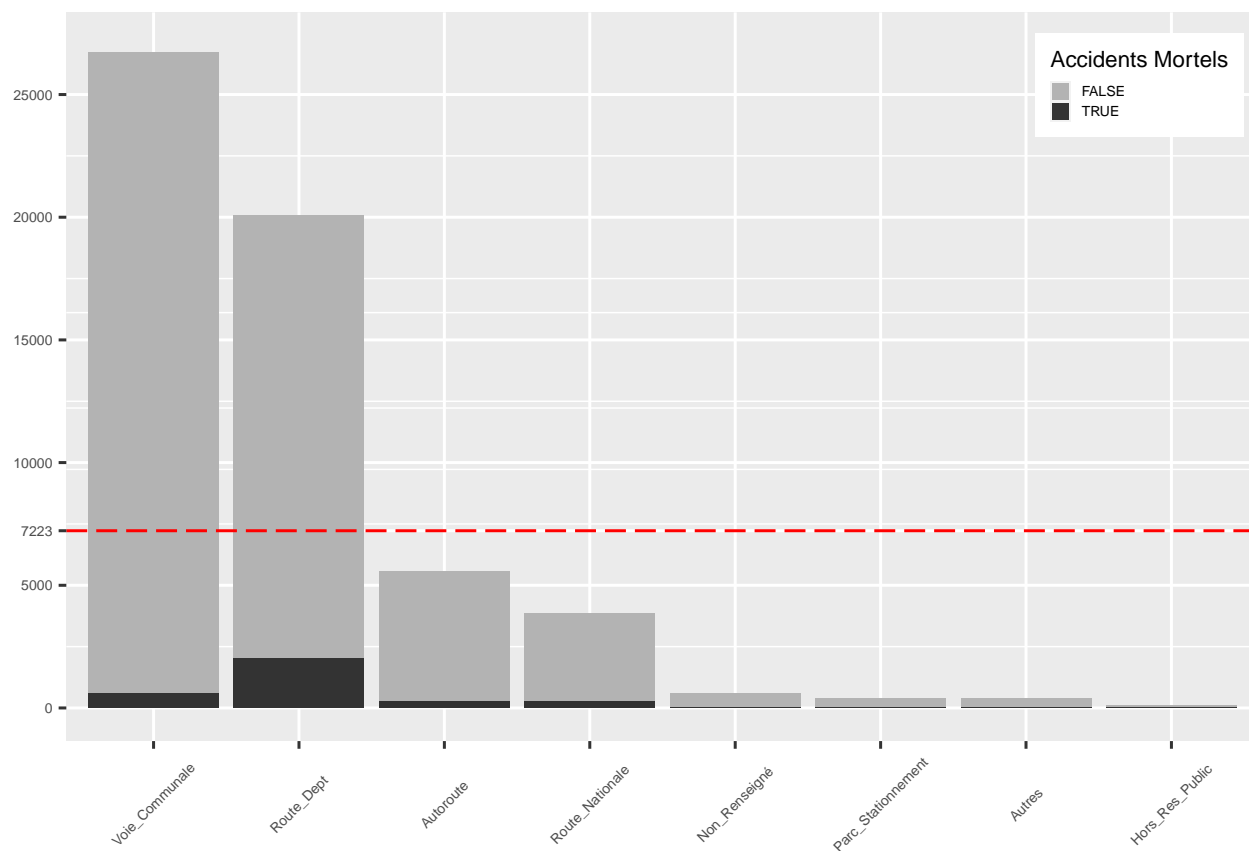
## Analyse des conditions routières

Cette section sera consacré à analyser les données liées aux conditions routières.

### Analyse Selon le type de route

La grande majorité des accidents se déroulent sur les voies communales et départementales. Mais la plupart des accidents mortels surviennent sur les routes départementales. Donc sur les voies communales, les accidents semblent être la plupart du temps juste des accrochages, alors que sur les voies départementales ils sont plus grave, la vitesse étant peut-être un facteur aggravant.

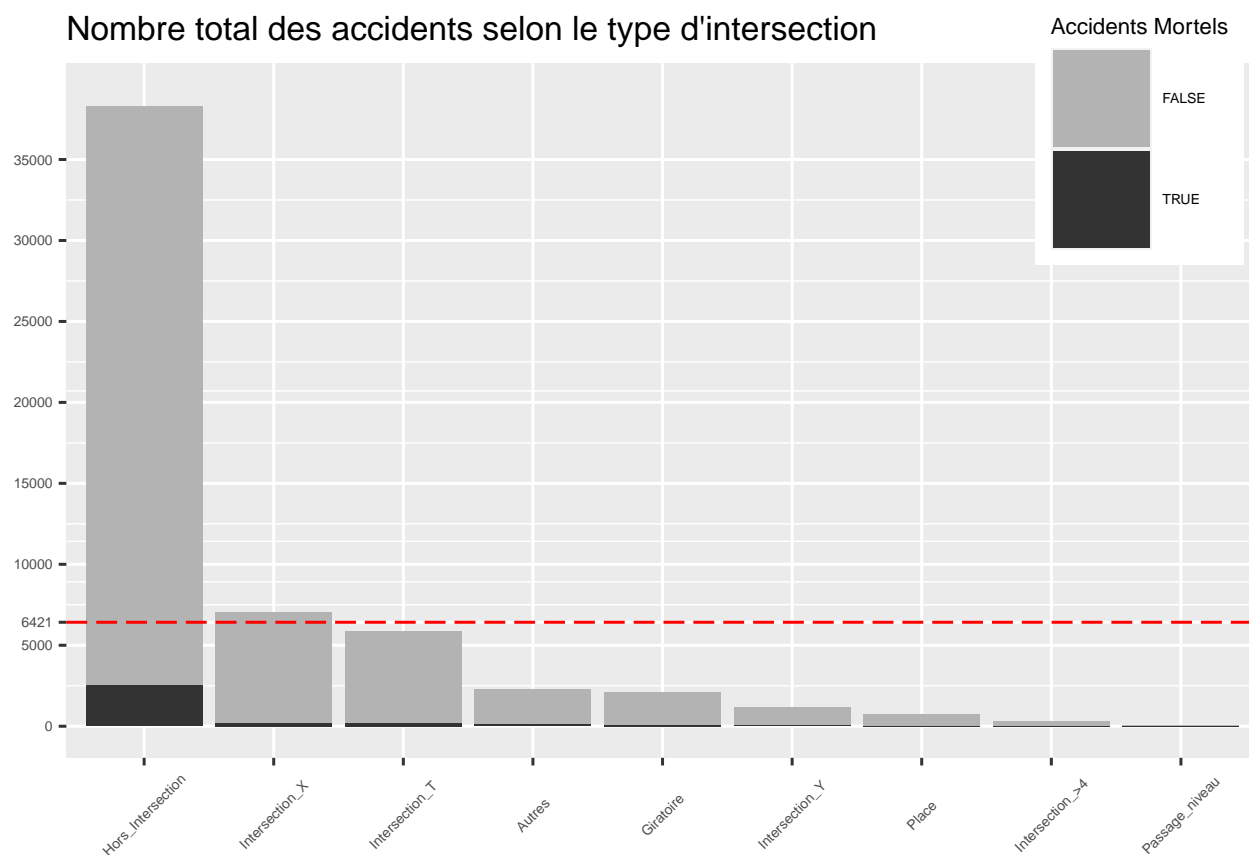
**Chart 9 : Répartitions des accidents selon le type de route**



### Type d'intersection si applicable

La grande majorité des accidents surviennent hors intersection avec 6.5% de mortalité. Les autres types d'intersection ont une mortalité de 3.5% environ.

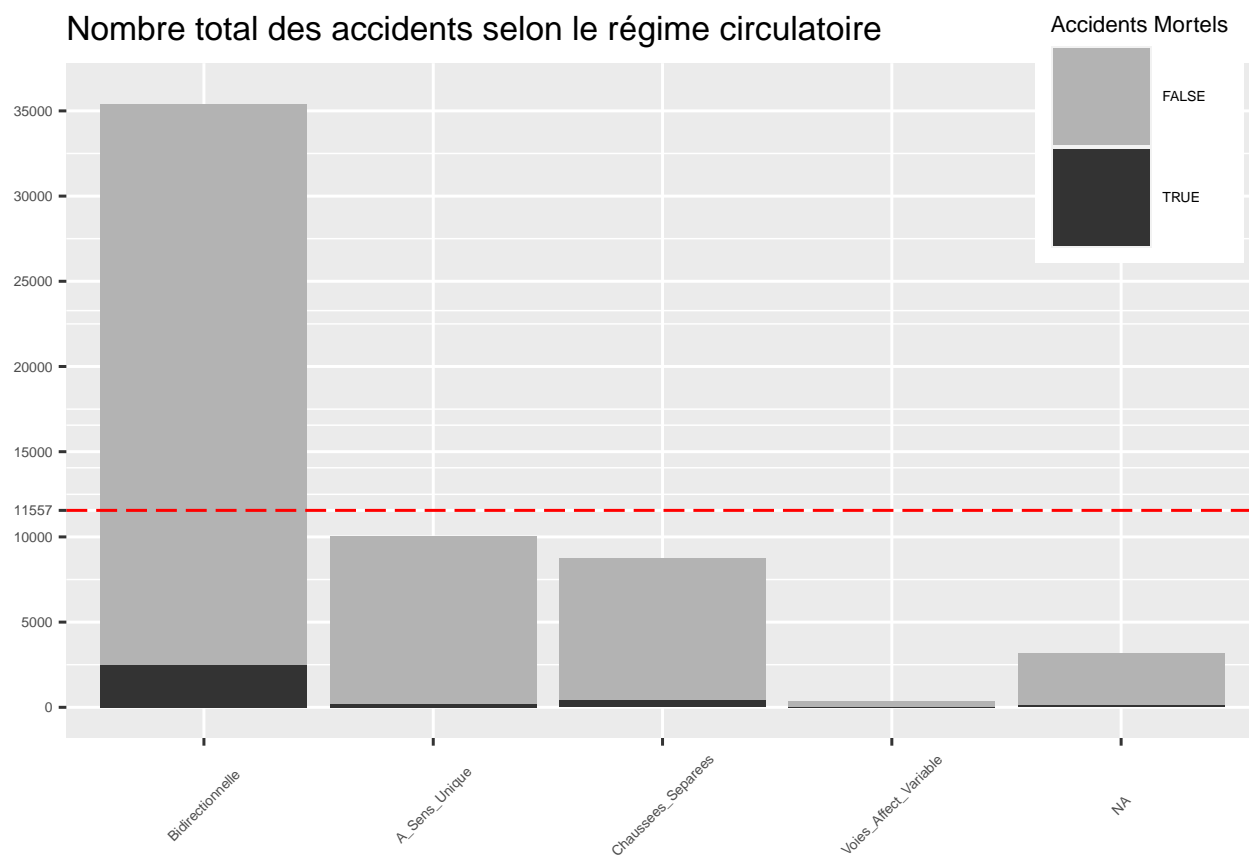
**Chart 10 : Répartitions des accidents selon le type d'intersection**



### Le régime circulaire

Les accidents surviennent en major partie sur des routes à double sens avec près de 7% de mortalité.

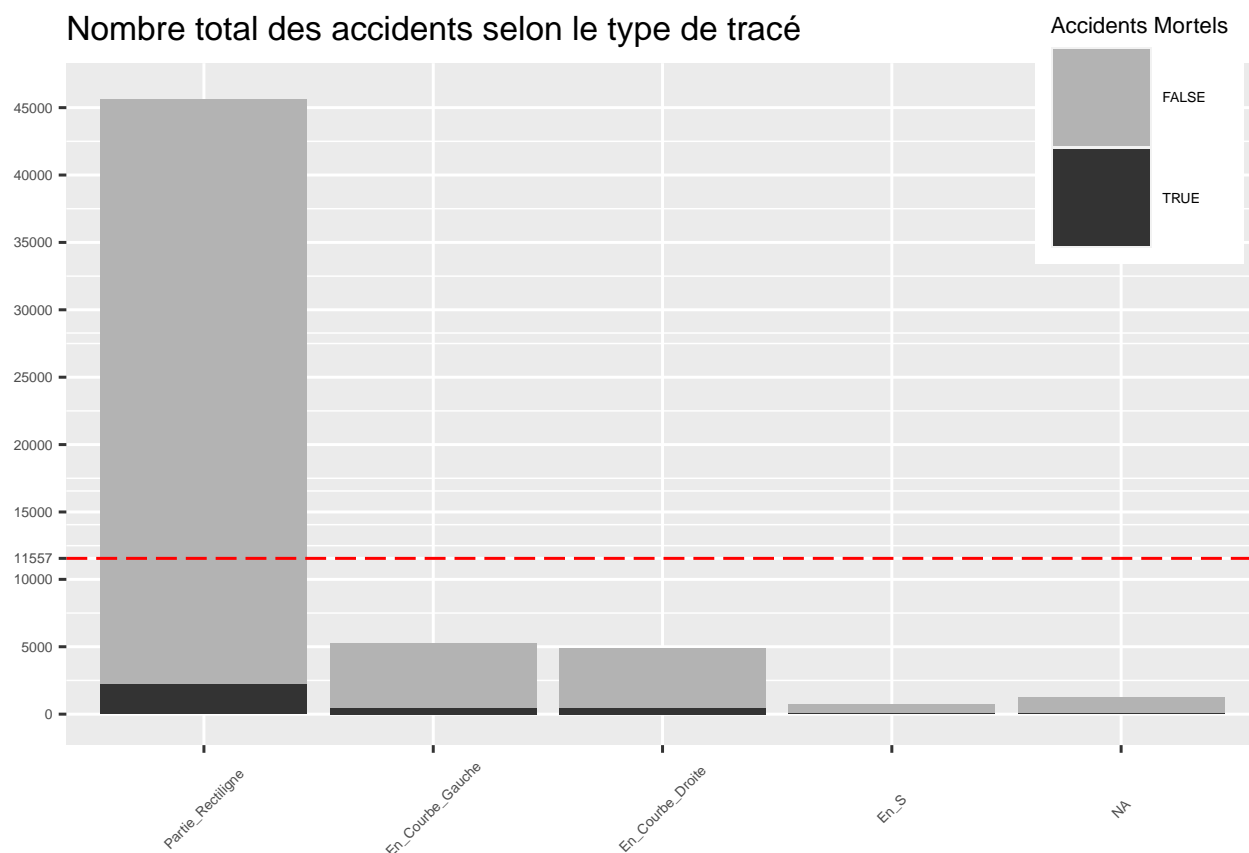
Chart 11 : Répartitions des accidents selon le régime circulaire



### Le type de tracé

Même situation que précédemment, la grande majorité des accidents sont sur une route rectiligne avec environ de 5% de mortels. Mais lorsque les accidents surviennent dans un virage ou une courbe en S avec une proportion d'accidents mortels autour de 8.7%. Par conséquent, les accidents en courbe sont plus dangereux.

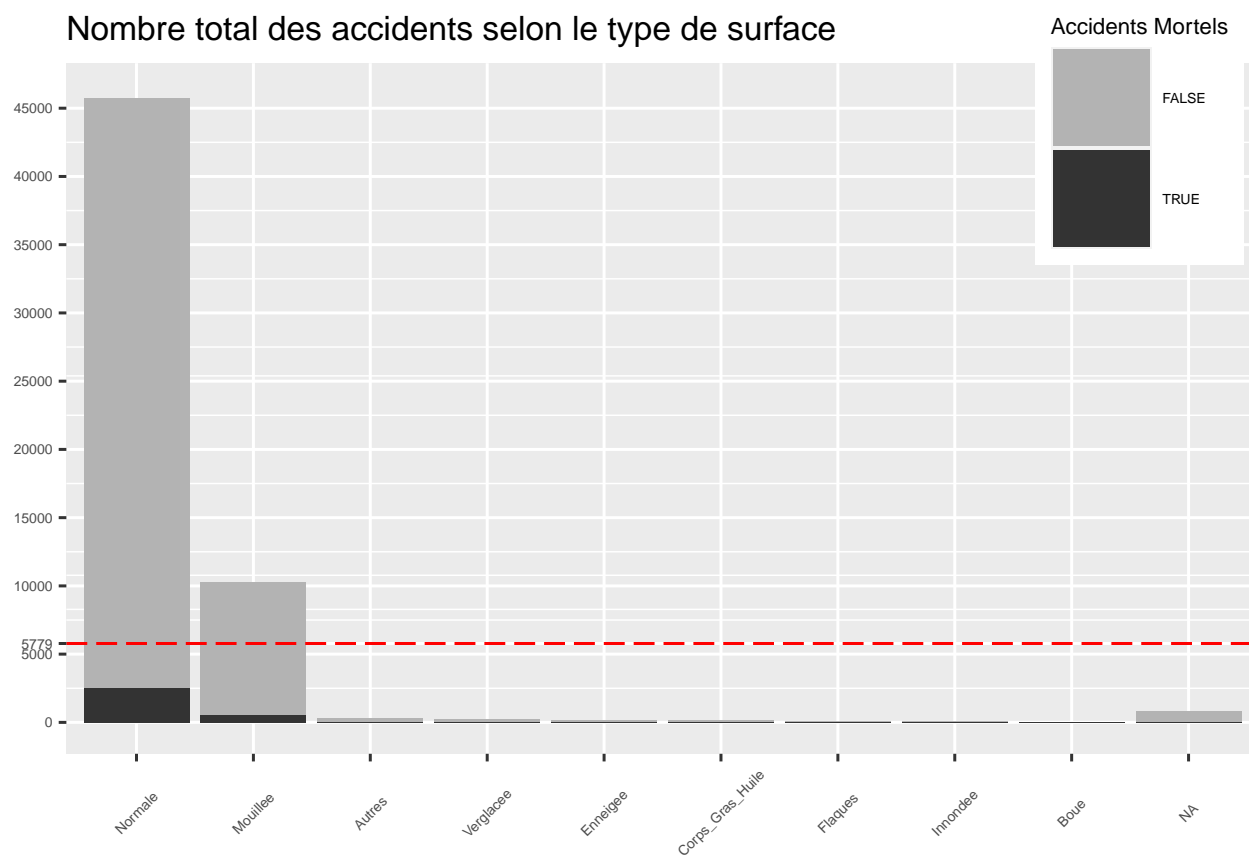
**Chart 12 : Répartitions des accidents selon le type de tracé**



### Le type de surface

79% des accidents surviennent sur une surface dite normale et 18% sur une surface mouillée avec un taux d'accidents mortels pour les deux surfaces de 5.5%. Dans 15% des cas un accident sur une route inondée a été mortel et dans 9% des cas sur les routes verglacées. Si les conditions météorologiques sont favorables pour entrainer des surfaces comme celle-ci, des mesures devraient être prise pour éviter tout accident.

**Chart 13 : Répartitions des accidents selon le type de surface**

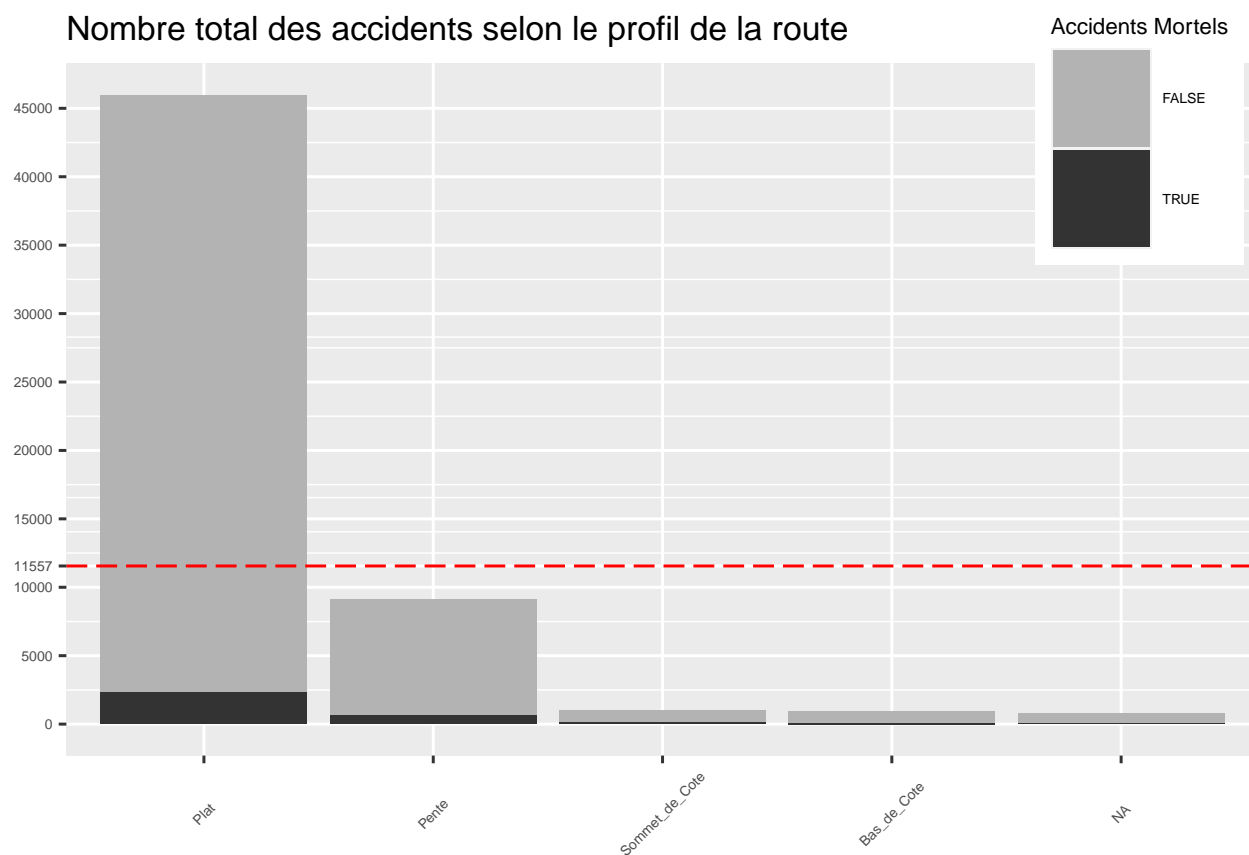


### Le profil de la route

Là encore, 80% des accidents surviennent sur une route plate avec un taux d'accident mortel d'environ de 5%. Par contre, si un accident survient au sommet d'une cote, le taux augmente à 10%.

Chart 14 : Répartitions des accidents selon le profil de la route

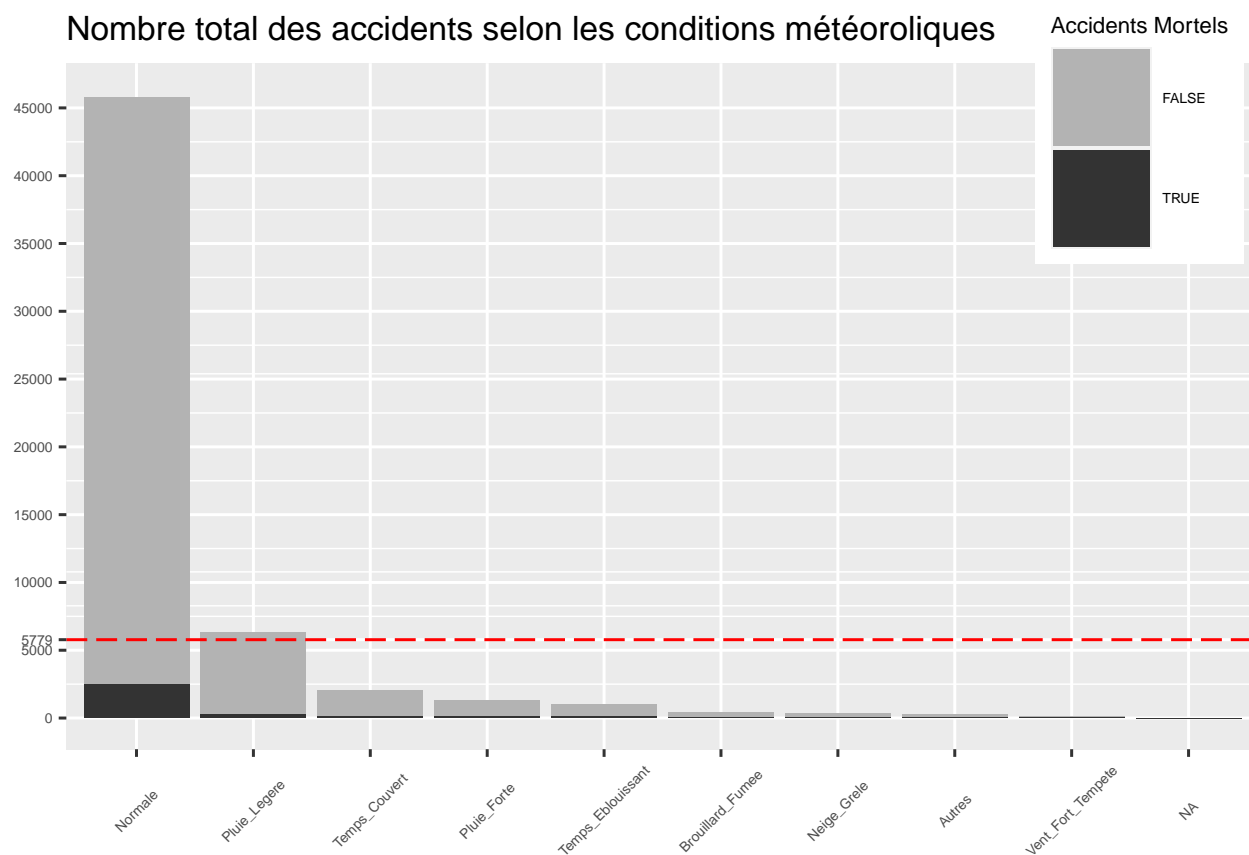




### Analyse des conditions météorologiques

Hors mis le constat que 80% des accidents surviennent dans des conditions météorologiques normales, les situations où il y a un manque de visibilité augmentent le risque d'accidents mortels. En effet, par temps de brouillards le risque d'avoir un accident mortel est de 14% environ et par vent fort pendant une tempête ou par temps éblouissant, le risque est de 9%. Ces situations doivent accroître la vigilance des automobilistes.

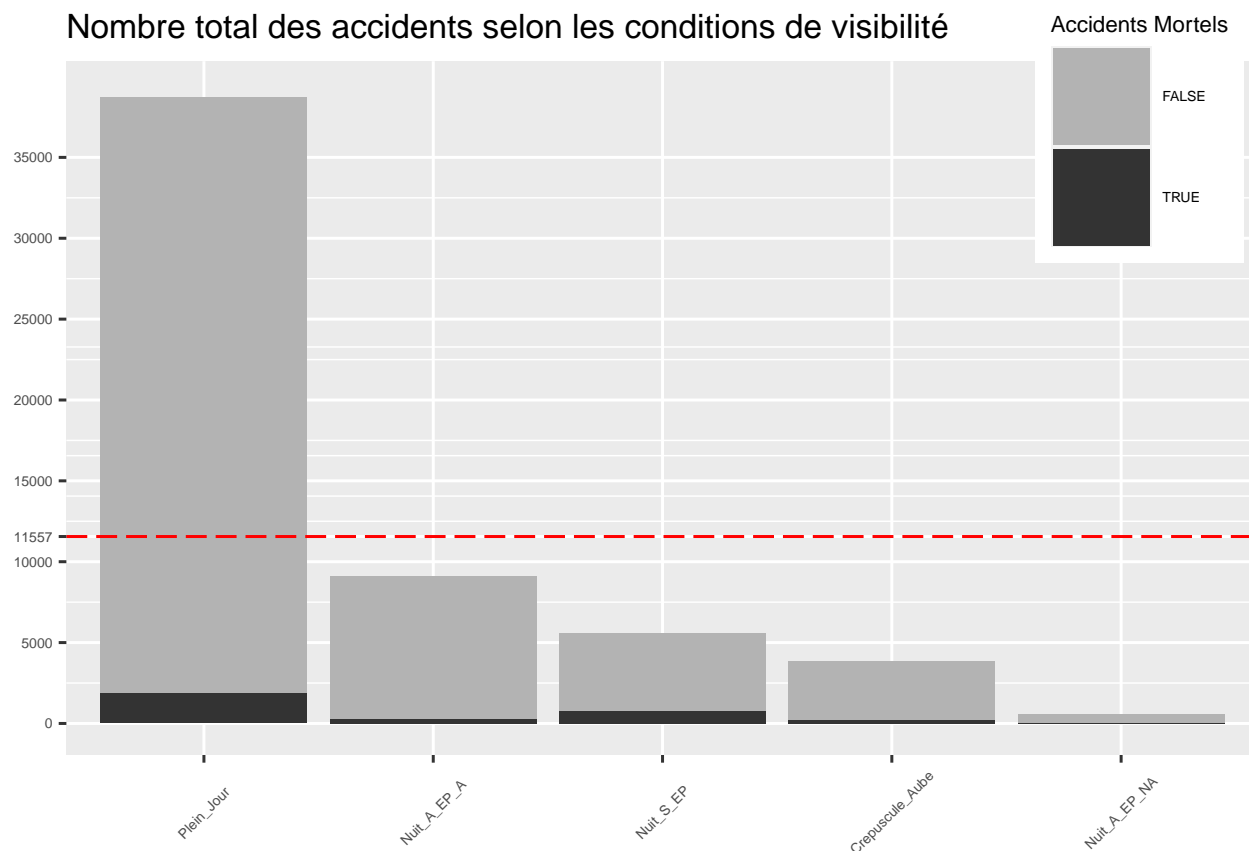
**Chart 15 : Répartitions des accidents selon les conditions météorologiques**



### Selon la visibilité

Les accidents surviennent dans 67% des cas en plein jour, et dans 10% des cas pendant la nuit sur une route sans éclairage donc des conditions de visibilité diminuée. C'est d'ailleurs dans cette dernière condition que le taux d'accidents mortels est le plus élevé, autour de 14%.

Chart 16 : Répartitions des accidents selon la visibilité



## Prediction

### Regression logistique

#### Model 1

```
##          Num_Acc          mois          condlumiere
## 201800000001: 1  Octobre : 5698  Plein_Jour      :38701
## 201800000002: 1   Juin   : 5579  Crepuscule_Aube: 3839
## 201800000003: 1  Septembre: 5532  Nuit_S_EP      : 5574
## 201800000004: 1   Juillet : 5241  Nuit_A_EP_NA   :  575
## 201800000005: 1    Mai    : 5055  Nuit_A_EP_A    : 9094
## 201800000006: 1  Novembre : 4883
## (Other)      :57777  (Other) :25795
##          agglo          typeintersect          condatmos
## Hors_agglomeration:20723  Hors_Intersection:38313  Normale          :45824
## En_agglomeration  :37060  Intersection_X   : 7024  Pluie_Legere     : 6350
##                               Intersection_T   : 5836  Temps_Couvert    : 2039
##                               Autres            : 2266  Pluie_Forte      : 1333
##                               Giratoire         : 2099  Temps_Eblouissant: 1038
##                               Intersection_Y    : 1186  (Other)          : 1194
##                               (Other)          : 1059  NA's             :    5
##          typecollision  numcommune  adresse          codagegps
## Autre_Collision:19522  55      : 2331  Length:57783     Metropole:55716
## 2V_Cote          :16284  7        :  694  Class :character  Antilles :  597
## 2V_Arriere       : 7393 116       :  644  Mode  :character  Guyane   :  519
## 2V_Frontale      : 5211 112       :  643                Reunion  :  621
```

```

## Sans_Collision : 5185 117 : 545 Mayotte : 280
## (Other) : 4186 118 : 538 NA's : 50
## NA's : 2 (Other):52388
## latitude longitude numdept
## Min. : 364247 Min. : -508705 75 : 5591
## 1st Qu.:4482685 1st Qu.: 171017 13 : 3413
## Median :4780120 Median : 242505 93 : 2735
## Mean :4587661 Mean : 457331 94 : 2548
## 3rd Qu.:4885140 3rd Qu.: 490700 92 : 2520
## Max. :5542440 Max. :6545675 69 : 2435
## NA's :1823 NA's :1823 (Other):38541
## dateacci joursem season
## Min. :2018-01-01 00:25:00 Fri : 9908 Automne :15247
## 1st Qu.:2018-04-17 13:15:00 Thu : 8511 Été :15109
## Median :2018-07-09 00:35:00 Tue : 8378 Hiver :11960
## Mean :2018-07-09 04:15:20 Wed : 8195 Printemps:15467
## 3rd Qu.:2018-10-05 08:15:00 Sat : 8164
## Max. :2018-12-31 23:50:00 (Other):14623
## NA's :4 NA's : 4
## Cat_Route Num_Voie Regime_Circ
## Voie_Communale :26728 :21893 A_Sens_Unique :10061
## Route_Dept :20087 0 : 6369 Bidirectionnelle :35397
## Autoroute : 5572 1 : 849 Chaussees_Separees : 8781
## Route_Nationale : 3857 86 : 834 Voies_Affect_Variable: 361
## Non_Renseigné : 622 6 : 819 Non_Renseigné : 0
## Parc_Stationnement: 423 7 : 795 NA's : 3183
## (Other) : 494 (Other):26224
## Nb_Voie Num_PR Distance_PR Exist_Voie_Reservee
## Min. : 0.000 0 :17452 Min. : 0.0 Piste_Cyclable : 1615
## 1st Qu.: 2.000 1 : 2676 1st Qu.: 0.0 Banque_Cyclable : 860
## Median : 2.000 2 : 1367 Median : 40.0 Voie_Reservee : 1774
## Mean : 2.345 3 : 1077 Mean : 253.9 Non_Voie_Reservee: 0
## 3rd Qu.: 2.000 4 : 956 3rd Qu.: 500.0 NA's :53534
## Max. :13.000 (Other):18480 Max. :8902.0
## NA's :507 NA's :15775 NA's :16048
## Profil_Route Trace_Plan Largeur_TPC
## Plat :45934 Partie_Rectiligne:45638 Min. : 0.00
## Pente : 9110 En_Courbe_Gauche : 5220 1st Qu.: 0.00
## Sommet_de_Cote: 992 En_Courbe_Droite : 4909 Median : 0.00
## Bas_de_Cote : 909 En_S : 731 Mean : 3.59
## NA's : 838 NA's : 1285 3rd Qu.: 0.00
## Max. :714.00
## NA's :44490
## Largeur_Chausee Etat_Surface Infrastrucutre
## Min. : 0.00 Normale :45726 Carrefour_Amenage : 3489
## 1st Qu.: 0.00 Mouillee :10247 Echangeur_Raccordement: 1083
## Median : 0.00 Autres : 312 Pont_Autopont : 898
## Mean : 39.05 Verglacee: 244 Souterrain_Tunnel : 623
## 3rd Qu.: 67.00 Enneigee : 158 Zone_Pietonne : 546
## Max. :800.00 (Other) : 296 (Other) : 261
## NA's :44074 NA's : 800 NA's :50883
## Situation_Acc Acc_Mortel Acc_Grave
## Chaussee :31995 Mode :logical Min. :0.0000
## Bande_Arret_Urgence: 863 FALSE:54597 1st Qu.:0.0000

```

```
## Accotement      :22680 TRUE :3186      Median :0.0000
## Trottoir        : 686              Mean    :0.3872
## Piste_Cyclable  : 317              3rd Qu.:1.0000
## NA's            : 1242             Max.     :1.0000
##
## hr_jour
## Min.           :0.000694
## 1st Qu.:0.409722
## Median :0.614583
## Mean      :0.580534
## 3rd Qu.:0.756944
## Max.      :0.999306
## NA's       :4
```

Nous pouvons constater que les colonnes “Largeur\_TPC”, “Largeur\_Chause”, “Exist\_Voie\_Reservee” et “Infrastrucutre” comporte beaucoup trop de valeurs NA, n’ayant pas d’information sur ces valeurs, je ne peux pas interpréter ces valeurs manquantes. Par conséquents, après avoir fait un premier essai de regression logistique avec ces colonnes, le model ne s’avérait pas significatif, alors je les ai supprimé du model.

```
## [1] "Sommaire du model_1"
##
## Call:
## glm(formula = Acc_Grave ~ ., family = binomial, data = train,
##      na.action = na.exclude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0017  -0.8351  -0.5740   0.9407   2.4423
##
## Coefficients:
##              Estimate Std. Error z value
## (Intercept)    -0.369158   0.097513  -3.786
## moisFévrier      0.082213   0.067819   1.212
## moisMars        -0.043187   0.064140  -0.673
## moisAvril       0.017306   0.063076   0.274
## moisMai         0.039751   0.062831   0.633
## moisJuin       0.077721   0.061873   1.256
## moisJuillet    0.049220   0.062914   0.782
## moisAoût       0.019328   0.065072   0.297
## moisSeptembre  -0.168040   0.062365  -2.694
## moisOctobre    -0.228285   0.061396  -3.718
## moisNovembre   -0.197859   0.062580  -3.162
## moisDécembre   -0.180380   0.063297  -2.850
## condlumiereCrepuscule_Aube  0.081076   0.049503   1.638
## condlumiereNuit_S_EP      0.278065   0.045785   6.073
## condlumiereNuit_A_EP_NA    0.079231   0.125889   0.629
## condlumiereNuit_A_EP_A     0.045453   0.036816   1.235
## agglomEn_agglomeration    -0.888181   0.034966 -25.401
## typeintersectIntersection_X -0.097791   0.042145  -2.320
## typeintersectIntersection_T  0.003308   0.043549   0.076
## typeintersectIntersection_Y -0.183094   0.090489  -2.023
## typeintersectIntersection_>4 -0.201183   0.203222  -0.990
## typeintersectGiratoire     0.078393   0.067184   1.167
## typeintersectPlace        -0.613841   0.142107  -4.320
```

## typeintersectPassage_niveau	0.027267	0.356318	0.077
## typeintersectAutres	0.097028	0.064080	1.514
## condatmosPluie_Legere	-0.065044	0.063481	-1.025
## condatmosPluie_Forte	0.036495	0.098656	0.370
## condatmosNeige_Grele	-0.246222	0.208463	-1.181
## condatmosBrouillard_Fumee	0.302940	0.143518	2.111
## condatmosVent_Fort_Tempete	0.480696	0.272574	1.764
## condatmosTemps_Eblouissant	0.505890	0.089886	5.628
## condatmosTemps_Couvert	0.009233	0.072170	0.128
## condatmosAutres	0.380999	0.180697	2.108
## typecollision2V_Arriere	-0.944153	0.056029	-16.851
## typecollision2V_Cote	-0.508071	0.047669	-10.658
## typecollision3V&pl_Chaine	-1.379770	0.087243	-15.815
## typecollision3V&pl_Multiple	-0.356089	0.077210	-4.612
## typecollisionAutre_Collision	-0.055235	0.046024	-1.200
## typecollisionSans_Collision	-0.144910	0.056832	-2.550
## joursem.L	-0.039912	0.033046	-1.208
## joursem.Q	0.201762	0.033485	6.025
## joursem.C	-0.063558	0.032509	-1.955
## joursem^4	0.074229	0.032151	2.309
## joursem^5	-0.016364	0.032045	-0.511
## joursem^6	0.004589	0.032339	0.142
## Cat_RouteRoute_Nationale	0.563686	0.066149	8.521
## Cat_RouteRoute_Dept	1.004493	0.061537	16.323
## Cat_RouteVoie_Communale	0.400179	0.065487	6.111
## Cat_RouteHors_Res_Public	1.128586	0.370467	3.046
## Cat_RouteParc_Stationnement	0.886835	0.312663	2.836
## Cat_RouteAutres	0.657635	0.155939	4.217
## Cat_RouteNon_Renseigné	0.881087	0.132126	6.669
## Regime_CircBidirectionnelle	0.651029	0.037666	17.284
## Regime_CircChaussees_Separees	0.310559	0.053572	5.797
## Regime_CircVoies_Affect_Variable	0.518897	0.151054	3.435
## Nb_Voie	-0.122191	0.012593	-9.703
## Profil_RoutePente	0.112672	0.033699	3.343
## Profil_RouteSommet_de_Cote	0.168796	0.093564	1.804
## Profil_RouteBas_de_Cote	0.037688	0.096931	0.389
## Trace_PlanEn_Courbe_Gauche	0.022251	0.044701	0.498
## Trace_PlanEn_Courbe_Droite	0.075070	0.044796	1.676
## Trace_PlanEn_S	-0.071267	0.107061	-0.666
## Etat_SurfaceMouillee	-0.266397	0.056069	-4.751
## Etat_SurfaceFlaques	-0.741271	0.332783	-2.227
## Etat_SurfaceInnondee	-0.450953	0.434824	-1.037
## Etat_SurfaceEnneigee	-0.443612	0.312691	-1.419
## Etat_SurfaceBoue	-0.732351	0.590857	-1.239
## Etat_SurfaceVerglacee	-0.317343	0.199787	-1.588
## Etat_SurfaceCorps_Gras_Huile	-0.387297	0.272581	-1.421
## Etat_SurfaceAutres	-0.088199	0.172260	-0.512
## Situation_AccBande_Arret_Urgence	-0.031316	0.096577	-0.324
## Situation_AccAccotement	-0.330398	0.027387	-12.064
## Situation_AccTrottoir	0.441657	0.108192	4.082
## Situation_AccPiste_Cyclable	-0.066605	0.167385	-0.398
## Acc_MortelTRUE	17.594447	78.057559	0.225
## hr_jour	0.072228	0.054366	1.329
##		Pr(> z )	

```

## (Intercept) 0.000153 ***
## moisFévrier 0.225421
## moisMars 0.500738
## moisAvril 0.783808
## moisMai 0.526957
## moisJuin 0.209064
## moisJuillet 0.434013
## moisAoût 0.766447
## moisSeptembre 0.007050 **
## moisOctobre 0.000201 ***
## moisNovembre 0.001569 **
## moisDécembre 0.004375 **
## condlumiereCrepuscule_Aube 0.101461
## condlumiereNuit_S_EP 0.0000000012530 ***
## condlumiereNuit_A_EP_NA 0.529106
## condlumiereNuit_A_EP_A 0.216979
## aggroEn_agglomeration < 0.0000000000000002 ***
## typeintersectIntersection_X 0.020322 *
## typeintersectIntersection_T 0.939456
## typeintersectIntersection_Y 0.043035 *
## typeintersectIntersection_>4 0.322191
## typeintersectGiratoire 0.243274
## typeintersectPlace 0.0000156338957 ***
## typeintersectPassage_niveau 0.939002
## typeintersectAutres 0.129986
## condatmosPluie_Legere 0.305538
## condatmosPluie_Forte 0.711440
## condatmosNeige_Grele 0.237550
## condatmosBrouillard_Fumee 0.034789 *
## condatmosVent_Fort_Tempete 0.077809 .
## condatmosTemps_Eblouissant 0.0000000182178 ***
## condatmosTemps_Couvert 0.898203
## condatmosAutres 0.034988 *
## typecollision2V_Arriere < 0.0000000000000002 ***
## typecollision2V_Cote < 0.0000000000000002 ***
## typecollision3V&pl_Chaine < 0.0000000000000002 ***
## typecollision3V&pl_Multiple 0.0000039887893 ***
## typecollisionAutre_Collision 0.230079
## typecollisionSans_Collision 0.010779 *
## joursem.L 0.227126
## joursem.Q 0.0000000016860 ***
## joursem.C 0.050572 .
## joursem^4 0.020956 *
## joursem^5 0.609588
## joursem^6 0.887147
## Cat_RouteRoute_Nationale < 0.0000000000000002 ***
## Cat_RouteRoute_Dept < 0.0000000000000002 ***
## Cat_RouteVoie_Communale 0.0000000009915 ***
## Cat_RouteHors_Res_Public 0.002316 **
## Cat_RouteParc_Stationnement 0.004563 **
## Cat_RouteAutres 0.0000247294859 ***
## Cat_RouteNon_Renseigné 0.0000000000258 ***
## Regime_CircBidirectionnelle < 0.0000000000000002 ***
## Regime_CircChaussees_Separees 0.0000000067519 ***

```

```

## Regime_CircVoies_Affect_Variable          0.000592 ***
## Nb_Voie                                < 0.0000000000000002 ***
## Profil_RoutePente                      0.000827 ***
## Profil_RouteSommet_de_Cote              0.071218 .
## Profil_RouteBas_de_Cote                 0.697410
## Trace_PlanEn_Courbe_Gauche              0.618639
## Trace_PlanEn_Courbe_Droite              0.093769 .
## Trace_PlanEn_S                          0.505625
## Etat_SurfaceMouillee                    0.0000020215844 ***
## Etat_SurfaceFlaques                     0.025914 *
## Etat_SurfaceInnondee                    0.299692
## Etat_SurfaceEnneigee                    0.155989
## Etat_SurfaceBoue                        0.215171
## Etat_SurfaceVerglacee                   0.112194
## Etat_SurfaceCorps_Gras_Huile             0.155360
## Etat_SurfaceAutres                      0.608645
## Situation_AccBande_Arret_Urgence         0.745745
## Situation_AccAccotement                  < 0.0000000000000002 ***
## Situation_AccTrottoir                   0.0000446195767 ***
## Situation_AccPiste_Cyclable              0.690694
## Acc_MortelTRUE                          0.821665
## hr_jour                                 0.183990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 49837 on 37224 degrees of freedom
## Residual deviance: 39715 on 37149 degrees of freedom
## (3223 observations deleted due to missingness)
## AIC: 39867
##
## Number of Fisher Scoring iterations: 16
## [1] "Variables significatives (avec une p-value <= 0.1)"
## [1] "(Intercept)" "moisSeptembre"
## [3] "moisOctobre" "moisNovembre"
## [5] "moisDecembre" "condlumiereNuit_S_EP"
## [7] "agglomEn_agglomeration" "typeintersectIntersection_X"
## [9] "typeintersectIntersection_Y" "typeintersectPlace"
## [11] "condatmosBrouillard_Fumee" "condatmosVent_Fort_Tempete"
## [13] "condatmosTemps_Eblouissant" "condatmosAutres"
## [15] "typecollision2V_Arriere" "typecollision2V_Cote"
## [17] "typecollision3V&pl_Chaine" "typecollision3V&pl_Multiple"
## [19] "typecollisionSans_Collision" "joursem.Q"
## [21] "joursem.C" "joursem^4"
## [23] "Cat_RouteRoute_Nationale" "Cat_RouteRoute_Dept"
## [25] "Cat_RouteVoie_Communale" "Cat_RouteHors_Res_Public"
## [27] "Cat_RouteParc_Stationnement" "Cat_RouteAutres"
## [29] "Cat_RouteNon_Renseigné" "Regime_CircBidirectionnelle"
## [31] "Regime_CircChaussees_Separees" "Regime_CircVoies_Affect_Variable"
## [33] "Nb_Voie" "Profil_RoutePente"
## [35] "Profil_RouteSommet_de_Cote" "Trace_PlanEn_Courbe_Droite"
## [37] "Etat_SurfaceMouillee" "Etat_SurfaceFlaques"

```



```
## [39] "Situation_AccAccotement"          "Situation_AccTrottoir"
##
##           0      1
## FALSE 5738 1420
##  TRUE 3988 4818
## [1] "Précision : 61%"
## [1] "Erreur Faux/negatif : 23%"
```

Avec ce premier model rapide nous arrivons à une précision de 58%. Il prédit qu'un accident n'est pas grave alors qu'en réalité il l'est dans 22% des cas.

## Model 2

Je vais modifier la database en rendant les variables de types facteurs en valeur numérique binaires (dummies variables). Ensuite, j'aimerais utiliser le package "mice" pour combler les valeurs manquantes. Pour finir, je vais normaliser les données de la variable "Nb\_Voie", le reste étant déjà normalisé.

```
##
## Call:
## glm(formula = Acc_Grave ~ ., family = binomial, data = train_2,
##      na.action = na.exclude)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1511  -0.8750  -0.6035   0.9762   2.4048
##
## Coefficients:
##                                Estimate Std. Error z value
## (Intercept)                   -0.01302    0.07157  -0.182
## Nb_Voie                       -0.11075    0.01132  -9.782
## mois_Septembre                 -0.19543    0.03967  -4.927
## mois_Octobre                   -0.25841    0.03939  -6.560
## mois_Novembre                  -0.25625    0.04289  -5.974
## mois_Décembre                  -0.20133    0.04374  -4.603
## condlumiere_Nuit_S_EP           0.34771    0.04127   8.426
## agglom_En_agglomeration        -0.96496    0.03119 -30.940
## typeintersect_Intersection_X    -0.11002    0.03712  -2.964
## typeintersect_Intersection_Y    -0.17165    0.07995  -2.147
## typeintersect_Place            -0.56467    0.12141  -4.651
## condatmos_Brouillard_Fumee      0.39087    0.13201   2.961
## condatmos_Vent_Fort_Tempete     0.60326    0.24126   2.500
## condatmos_Temps_Eblouissant     0.48464    0.08343   5.809
## condatmos_Autres                0.35824    0.15479   2.314
## typecollision_2V_Arriere        -0.96758    0.03951 -24.489
## typecollision_2V_Cote           -0.53311    0.02834 -18.814
## typecollision_3V.pl_Chaine      -1.37603    0.07315 -18.810
## typecollision_3V.pl_Multiple    -0.31106    0.06323  -4.920
## typecollision_Sans_Collision    -0.15549    0.04079  -3.812
## Cat_Route_Route_Nationale       0.58777    0.06094   9.645
## Cat_Route_Route_Dept            1.06783    0.05612  19.026
## Cat_Route_Voie_Communale        0.42525    0.06006   7.081
## Cat_Route_Hors_Res_Public       0.96466    0.26124   3.693
## Cat_Route_Parc_Stationnement    0.72216    0.13873   5.206
## Cat_Route_Autres                0.69613    0.14519   4.794
```

## Cat_Route_Non_Renseigné	0.89443	0.11598	7.712
## Regime_Circ_Bidirectionnelle	0.61168	0.03351	18.252
## Regime_Circ_Chausees_Separees	0.27586	0.04894	5.637
## Regime_Circ_Voies_Affect_Variable	0.38106	0.13643	2.793
## Profil_Route_Pente	0.12537	0.03101	4.043
## Profil_Route_Sommet_de_Cote	0.20574	0.08718	2.360
## Trace_Plan_En_Courbe_Droite	0.09170	0.04051	2.264
## Etat_Surface_Mouillee	-0.29711	0.03114	-9.541
## Etat_Surface_Flaques	-0.71575	0.30995	-2.309
## Situation_Acc_Accotement	-0.31330	0.02413	-12.984
## Situation_Acc_Trottoir	0.45551	0.09684	4.704
## joursem_Mon	-0.22277	0.04501	-4.949
## joursem_Tue	-0.23486	0.04432	-5.299
## joursem_Wed	-0.23064	0.04456	-5.176
## joursem_Thu	-0.19156	0.04389	-4.364
## joursem_Fri	-0.16212	0.04254	-3.811
## joursem_Sat	-0.06710	0.04392	-1.528
## joursem_NA	11.07854	68.79286	0.161
##		Pr(> z )	
## (Intercept)		0.855690	
## Nb_Voie	< 0.0000000000000002	***	
## mois_Septembre	0.0000008356696788	***	
## mois_Octobre	0.0000000000538026	***	
## mois_Novembre	0.0000000023141102	***	
## mois_Décembre	0.0000041726891331	***	
## condlumiere_Nuit_S_EP	< 0.0000000000000002	***	
## aggro_En_agglomeration	< 0.0000000000000002	***	
## typeintersect_Intersection_X	0.003036	**	
## typeintersect_Intersection_Y	0.031802	*	
## typeintersect_Place	0.0000033046652101	***	
## condatmos_Brouillard_Fumee	0.003067	**	
## condatmos_Vent_Fort_Tempete	0.012403	*	
## condatmos_Temps_Eblouissant	0.0000000062862873	***	
## condatmos_Autres	0.020648	*	
## typecollision_2V_Arriere	< 0.0000000000000002	***	
## typecollision_2V_Cote	< 0.0000000000000002	***	
## typecollision_3V.pl_Chaine	< 0.0000000000000002	***	
## typecollision_3V.pl_Multiple	0.0000008676109240	***	
## typecollision_Sans_Collision	0.000138	***	
## Cat_Route_Route_Nationale	< 0.0000000000000002	***	
## Cat_Route_Route_Dept	< 0.0000000000000002	***	
## Cat_Route_Voie_Communale	0.0000000000014325	***	
## Cat_Route_Hors_Res_Public	0.000222	***	
## Cat_Route_Parc_Stationnement	0.0000001933024867	***	
## Cat_Route_Autres	0.0000016308957101	***	
## Cat_Route_Non_Renseigné	0.00000000000000124	***	
## Regime_Circ_Bidirectionnelle	< 0.0000000000000002	***	
## Regime_Circ_Chausees_Separees	0.0000000173348830	***	
## Regime_Circ_Voies_Affect_Variable	0.005221	**	
## Profil_Route_Pente	0.0000526721989953	***	
## Profil_Route_Sommet_de_Cote	0.018282	*	
## Trace_Plan_En_Courbe_Droite	0.023590	*	
## Etat_Surface_Mouillee	< 0.0000000000000002	***	
## Etat_Surface_Flaques	0.020928	*	

```

## Situation_Acc_Accotement < 0.0000000000000002 ***
## Situation_Acc_Trottoir 0.0000025520945956 ***
## joursem_Mon 0.0000007452883417 ***
## joursem_Tue 0.0000001164134555 ***
## joursem_Wed 0.0000002263289852 ***
## joursem_Thu 0.0000127584275701 ***
## joursem_Fri 0.000138 ***
## joursem_Sat 0.126630
## joursem_NA 0.872060
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 53995 on 40447 degrees of freedom
## Residual deviance: 46359 on 40404 degrees of freedom
## AIC: 46447
##
## Number of Fisher Scoring iterations: 9
##
##           0      1
## FALSE 5741 1532
## TRUE  4883 5179
## [1] "Précision : 63%"
## [1] "Erreur Faux/negatif : 23%"

```

Avec les nouvelles modifications, la précision des prédictions est de 71% et une erreur de type faux/negatif (proportion à classer correctement ) de 15%. En terme de précision le model c'est amélioré.

## Neural Network

```

##
##           0      1
## FALSE 6076 1641
## TRUE  4548 5070
## [1] "Précision : 64%"
## [1] "Erreur Faux/Positif : 24%"

```

Avec cet algorithme, la précision du model est de 65%.

## Arbre de décision

```

## CART
##
## 57783 samples
## 43 predictor
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 52005, 52005, 52004, 52004, 52004, 52005, ...
## Resampling results across tuning parameters:
##
## cp      RMSE      Rsquared    MAE

```

```

## 0.0001 0.4412223 0.18028455 0.3838672
## 0.0021 0.4475009 0.15617338 0.4003354
## 0.0041 0.4522159 0.13834391 0.4089662
## 0.0061 0.4534140 0.13375074 0.4111556
## 0.0081 0.4534140 0.13375074 0.4111556
## 0.0101 0.4534140 0.13375074 0.4111556
## 0.0121 0.4534140 0.13375074 0.4111556
## 0.0141 0.4600718 0.10807464 0.4233250
## 0.0161 0.4600718 0.10807464 0.4233250
## 0.0181 0.4600718 0.10807464 0.4233250
## 0.0201 0.4600718 0.10807464 0.4233250
## 0.0221 0.4600718 0.10807464 0.4233250
## 0.0241 0.4600718 0.10807464 0.4233250
## 0.0261 0.4600718 0.10807464 0.4233250
## 0.0281 0.4600718 0.10807464 0.4233250
## 0.0301 0.4675067 0.07901399 0.4371211
## 0.0321 0.4675067 0.07901399 0.4371211
## 0.0341 0.4675067 0.07901399 0.4371211
## 0.0361 0.4675067 0.07901399 0.4371211
## 0.0381 0.4675067 0.07901399 0.4371211
## 0.0401 0.4675067 0.07901399 0.4371211
## 0.0421 0.4675067 0.07901399 0.4371211
## 0.0441 0.4675067 0.07901399 0.4371211
## 0.0461 0.4675067 0.07901399 0.4371211
## 0.0481 0.4675067 0.07901399 0.4371211
## 0.0501 0.4675067 0.07901399 0.4371211
## 0.0521 0.4675067 0.07901399 0.4371211
## 0.0541 0.4675067 0.07901399 0.4371211
## 0.0561 0.4675067 0.07901399 0.4371211
## 0.0581 0.4675067 0.07901399 0.4371211
## 0.0601 0.4675067 0.07901399 0.4371211
## 0.0621 0.4675067 0.07901399 0.4371211
## 0.0641 0.4675067 0.07901399 0.4371211
## 0.0661 0.4675067 0.07901399 0.4371211
## 0.0681 0.4675067 0.07901399 0.4371211
## 0.0701 0.4675067 0.07901399 0.4371211
## 0.0721 0.4675067 0.07901399 0.4371211
## 0.0741 0.4675067 0.07901399 0.4371211
## 0.0761 0.4675067 0.07901399 0.4371211
## 0.0781 0.4719425 0.07642283 0.4450211
## 0.0801 0.4855187 0.06499735 0.4711260
## 0.0821 0.4871050      NaN 0.4745385
## 0.0841 0.4871050      NaN 0.4745385
## 0.0861 0.4871050      NaN 0.4745385
## 0.0881 0.4871050      NaN 0.4745385
## 0.0901 0.4871050      NaN 0.4745385
## 0.0921 0.4871050      NaN 0.4745385
## 0.0941 0.4871050      NaN 0.4745385
## 0.0961 0.4871050      NaN 0.4745385
## 0.0981 0.4871050      NaN 0.4745385
##
## RMSE was used to select the optimal model using the smallest value.
## The final value used for the model was cp = 1e-04.

```

```
##
##           0      1
##  FALSE 6264 1701
##   TRUE  4360 5010

## [1] "Précision : 65%"
## [1] "Erreur Faux/Positif : 25%"
```

Ici le model nous donne 69% de précision avec une erreur de 8%.

Pour comparer les models entre eux, on peut se baser sur deux paramètres : la précision et la sensibilité. La sensibilité est le ratio du nombre de prédiction d'accidents grave correct (prédire 1 et réalité correspond à 1) / nombre réel d'accidents grave

Algorithm	Accuracy	False Negative Rate	Sensitivity	F1_score
Logistic Regression	61	23	77	68.072
Logistic Regression with dummy variables/NA estimation	63	23	77	69.300
Neural Network	64	24	76	69.486
CART Model	65	25	75	69.643

En comparant le F1-score de ces modèles, nous constatons qu'ils sont à peu près équivalent avec une précision supérieur à 60%. Néanmoins le modèle CART, et d'après le F1-score, serait un bon choix dans notre situation.

En incluant toutes les données disponibles, ces modèles peuvent être améliorés pour probablement gagner en précision. (coming soon)