

# Projet Modele Statistique

---

ETUDES DE LA MORTALITE DUE A L'ABUS DE DROGUES AUX  
USA

HADJ ARAB ANIS – BODIANG ASSANE  
| L3 MIAGE

## **Sommaire:**

### **I) Etude du Connecticut**

- a) Carte des USA
- b) Quelques chiffres
- c) Graphes des morts par rapport à la durée (2012-2016)
- d) Graphe des morts selon la drogue utilisée
- e) Graphe par groupe ethnique :

### **II) Etude générale des USA**

- a) Comparaison Population/Mort
- b) Graphe selon le Sexe
- c) Graphe selon l'ethnicité
- d) Graphe selon l'âge

## **Introduction:**

Notre objectif est d'utiliser nos connaissances en RStudio pour récupérer des graphes qui nous montrent l'évolution du nombre de morts, par rapport au Sexe, l'Age et peut être même l'ethnicité pour étudier en profondeur les conditions de ces décès (il est regrettable de ne pas pouvoir faire en parallèle une étude des conditions de vie selon les différentes régions aux USA pour pouvoir émettre des hypothèses concrètes sur les raisons de cet abus des drogues).

Pour faire cette étude, nous avons choisi deux fichiers obtenus du site du gouvernement Américain :

US Data.gov: <https://catalog.data.gov/dataset>

Les fichiers obtenus sont les suivants :

<https://catalog.data.gov/dataset/accidental-drug-related-deaths-january-2012-sept-2015>  
<https://catalog.data.gov/dataset/drug-poisoning-mortality-u-s-and-state-trends-united-states-20022014>

La question que nous nous sommes posés et à laquelle nous voulons essayer de répondre est la suivante :

Y'a-t-il un rapport entre l'âge, le sexe ou l'ethnicité avec le nombre de décès dus à la drogue ?  
Et l'évolution du taux de mortalité suis-t-il la même évolution que la croissance de la population ?

Nous nous sommes posé ces questions tout d'abord car l'âge est un facteur important dans tout domaine et est la première variable à prendre en compte. Le sexe en est un aussi et cela car les habitudes et mode de vies des Hommes et des Femmes sont quelque peu différents. Quant à l'ethnie, ayant été présente dans notre DataBase nous avons jugé important de comparer pour voir s'il y a vraiment des différences et peut-être donner des hypothèses pour les expliquer.

Voir le RMarkdown pour plus de détails sur chaque exécution et les commentaires de nos choix à tel moment de l'avancée du projet.

## **Méthodologie :**

Tout d'abord, nous avons décidé, en lisant le fichier d'enlever les informations que nous jugées inutiles pour notre travail (Etat de décès sachant que tout se passe au Connecticut). Ensuite, montrer le Connecticut sur une carte grâce à ggmap et maptools. Nous avons utilisé gsub pour nettoyer la colonne DeathLoc pour récupérer la longitude et la Latitude. Puis avons décidé de donner quelques chiffres clés de notre étude (moyenne d'âge des morts (Par sexe également), un sommaire des âges ainsi que le nombre de morts par sexe) et cela pour avoir une idée sur ces informations sur lesquelles nous sommes en train de travailler.

Ensuite, viens le moment où il faut représenter nos chiffres par des graphes. Il nous a fallu transformer nos dates de type factor en type Date et donc pour cela nous avons utilisé la commande as.Date. Une fois cela fait, nous avons voulu voir un diagramme des décès par jour pour avoir une base sur quoi concentrer nos recherches.

Une fois le graphe des morts par jour fait, nous nous sommes pu voir que le pic des morts par jour est de 8 et cela dans un seul Etat ce qui est énorme !

Cependant, il aurait été préférable d'avoir une étude en parallèle de la signification de chacune de ces jours pour voir si par exemple les pics sont causés lors de jours fériés, weekends ou une fête quelconque, ou même par exemple d'une vague de licenciement ou autre (problèmes économiques).

Ce graphe obtenu ne montre pas bien l'évolution que nous voulons constater. Nous avons donc par la suite fait une étude mensuelle

Nous avons essayé de récupérer juste les Hommes qui consomment de l'Héroïne et nous avons essayé de les arranger par date. Cela nous a permis de nous rendre compte que la date est du type factor et donc nos valeurs ne sont pas arrangées par ordre de date (nous ne pouvons donc pas faire de graphe par rapport à la date) :

Grace à la librairie as et à la commande as.Date, nous avons réussi à transformer notre facteur Date en un type Date qui peut être plus facilement manipulé.

Ce qui nous permettra de faire plus facilement nos futurs graphes. Par exemple le graphe Du nombre des morts par jour suivant :

Enfin grâce à la commande : `mois <- strptime(df4$Date, "%m%Y")` et gsub :

`Mutate (Annee = gsub("-", "$", "", Date))`, nous avons récupéré seulement l'année et nous avons fait le graphe avec le nombre de morts annuel au Connecticut. Ce graphe nous montre bien mieux l'évolution qui a eu lieu entre 2012 à 2016. Il montre une augmentation de 3 fois du nombre de mort en 4 ans !!

Une fois ayant comparés les sexes, nous nous sommes concentrés sur les différentes drogues consommées et le nombre de décès qu'elles ont causé au cours du temps :

Ne sachant pas comment transformer mes colonnes de chaque drogue en une ligne par rapport à chaque mort, nous avons décidé de créer un data frame dans lequel on met dans la première colonne le nom des drogues et dans la seconde le nombre de morts pour chacune de ces drogues.

Pour cela, nous avons tout d'abord récupéré le nombre de mort pour chaque drogue séparément (sachant la nature du data frame de base nous ne pouvions pas les récupérer en même temps). Par la suite, nous avons utilisé data.frame pour créer le data frame voulu. Enfin, nous avons plot nos résultats dans le graphe suivant en bar (car nous pensons que cela est plus pertinent).

Finalement, pour avoir une étude plus concrète et complète, nous avons décidé d'étudier l'ethnicité aussi (ce que nous ne voulions pas faire au départ) et donc nous avons créé un graphe qui représente le nombre de morts durant cette période de 4 ans par groupe ethnique. Ce que nous avons obtenus montre très clairement que le nombre de morts aux USA est largement plus élevé pour les Blancs (plus de 5 fois plus élevé que le second groupe : Hispaniques). Plusieurs raisons peuvent expliquer ça, cependant sachant que les chiffres récupérés sont au Connecticut, nous ne pouvons faire de conclusions hâtives. Nous concluons dessus à la fin.

Les mêmes étapes ont été prises mais cette fois pour faire l'étude sur la globalité des USA et cette fois sans la difficulté de devoir arranger la date. Il aurait été préférable de pouvoir faire une Map sur laquelle on plot, selon l'importance des populations ou du nombre de décès, des cercles plus ou moins larges. Mais n'ayant pas de Latitude et de Longitude, nous avons trouvé difficile cette tâche malgré le fait que la librairie map permette de récupérer les zones selon le nom de l'état Américain. Je pense qu'ici une carte aurait été plus pertinente qu'un graphe ou un histogramme.

Enfin, nous avons également fait un graphe qui se concentre sur les fourchettes d'âge : Nous avons remarqué d'abord qu'il y a des morts de moins de 14 ans ! De plus, les groupes d'âges les plus touchés sont entre 25 et 54 ans et ne cessent d'augmenter et les plus touchés sont les personnes entre 45 et 55 ans.

#### Choix de représentation :

Tout au long de ce projet, nous nous sommes concentrée sur les graphes, parfois des lignes ou des points quand il s'agit de durées, et des diagrammes en barres pour comparer des quantités selon des variables différentes (par exemple les différentes drogue). Nous avons choisi ceux-là qu'ils sont bien plus facilement lisibles, nos chiffres n'étant pas très complexes, ces graphes sont parfaits pour les étudier. Nous avons aussi fait des tableaux pour montrer en détail les valeurs que nous avons mis dans nos graphes. Finalement, nous aurions bien aimé nous concentrer sur présenter nos chiffres avec des map mais ceci s'est avéré assez difficile.

## **Analysis in Literate Programming :**

```
i=0
While != EOF
    Get the element we're looking for (ex : get Date)
    j=0
    While != EOF
        If current.element = element
            Then i++
            Delete row
        End if
    End while
i++
end while
```

⇒ Comparable au groupe\_by(element) %>% summarize(N=n())

```
While != EOF
    If current.element = element
        Keep/Delete row
    Else Delete/Keep
    End if
End while
```

⇒ Comparable au filter (..=.. / ..!=..)

## **Conclusion :**

Au cours de ce projet, nous avons appris à mieux manipuler les informations que nous avons et à souterir que les informations dont nous avons besoin et les présenter d'une façon lisible par tous et compréhensible. Les statistiques nous donnent des valeurs réelles, lesquelles nous pouvons analyser et dont on peut sortir les informations nécessaires pour prendre des décisions ou émettre des hypothèses sur certains sujets.

Par exemple dans notre sujet ici par rapport aux décès causes par la drogue, nous avons pu émettre des hypothèses tel que le comportement des femmes et hommes est différent et donc leur consommation de drogue diffère d'où les chiffres très espacés que nous avons trouvés. Une autre serait que l'âge est un autre facteur (ceux de 70 ans sont moins aptes à en consommer que les personnes de 40 ans).

Cependant cette étude a montré certainement une chose importante : elle nous montre que les drogues dures sont de plus en plus présentes et consommées de nos jours (effet de mode peut être ; par ex : des jeunes qui font la fête ou des d'autres voulant diminuer leur stress), et il faudrait donc renforcer les mesures prises face à ce problème qui continue d'augmenter !!