

# Mini Tp, Etude du taux de résolution des homicides par état des Etats unis depuis 1980

*MugnierJu*

*21 mars 2017*

## Librairies à inclure

```
doInstall <- TRUE
toInstall <- c("maps", "ggplot2")

library(ggplot2)
library(maps)
library(dplyr);

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(magrittr);
```

## Introduction

Nous avons choisi comme jeu de test un fichier contenant chaque homicide commis aux Etat Unis d'Amérique depuis 1980. Nous nous sommes posés la question suivante : y a-t-il des zones/groupes d'états dans lesquels le taux de résolution des homicides est plus fort ou plus faible ? Hypothèse : Nous pourrions distinguer des groupes d'états proches ayant un taux de résolution similaire. De plus nous ne pensons pas que le taux de résolution sera homogène.

Nous considérons donc que les moyens alloués à la police dans l'ensemble du pays sont équivalents. De plus, pour des raisons de limitation technique, nous ne pouvons pas présenter les données concernant les états d'Hawaii et de l'Alaska. En effet le jeu de données contenant les coordonnées que nous avons utilisés afin de construire nos cartes ne contient pas ces deux états.

Notre analyse va donc s'articuler en deux parties, une première permettant de présenter les données et une deuxième répondant plus en profondeur à notre question.

Le jeu de données utilisé : <https://www.kaggle.com/murderaccountability/homicide-reports>

```
data <- read.csv("/home/julien/Documents/ModelStat/proglitt/MiniTp/database.csv")
```

## Première partie : présentation des données

Nous commençons, dans un premier temps, par compter le nombre d'homicides commis dans chaque état depuis 1980.

```
data %>% group_by(State) %>% count(State) ->data2
```

Notre but est de créer une carte affichant le nombre d'homicides par état, nous utilisons donc la fonction `map_data()` de la librairie `maps` afin de charger un jeu de données extérieur contenant les coordonnées nécessaires à la création d'une carte.

Comme dit précédemment, le jeu de données contenant les informations de la carte ne prenant pas en compte les états de Hawaii et de l'Alaska, nous devons les retirer de notre jeu de données.

```
all_states <- map_data("state")
data2$region <- tolower(data2$State)

data2 %>% filter(region != "alaska") -> data2
data2 %>% filter(region != "hawaii") -> data2
```

Les actions précédentes ont été mises en place dans le but de recouper les données de nos deux jeux de test. Nous effectuons cette fusion sur le champ `region` qui leur est commun et que nous avons harmonisé.

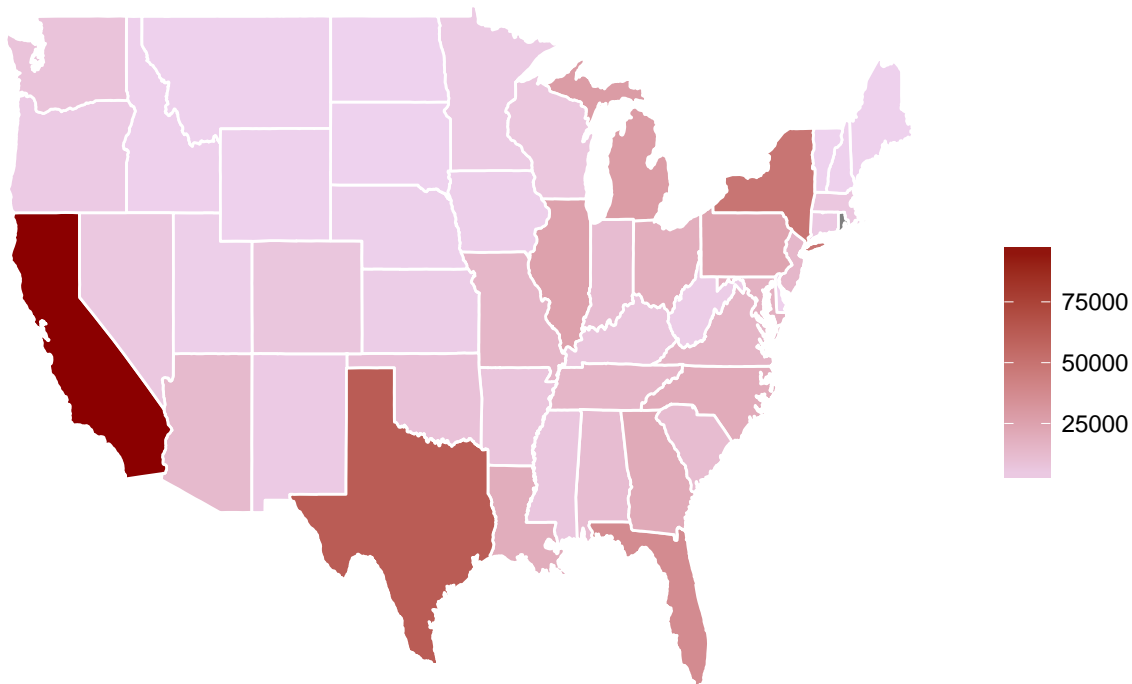
```
Total <- left_join(all_states, data2, by="region")
```

Finalement nous créons une carte des Etats Unis avec le nombre de meurtres par état en valeur absolue.

```
p <- ggplot()
p <- p + geom_polygon(data=Total, aes(x=long, y=lat, group = group, fill=Total$n), colour="white"
) + scale_fill_continuous(low = "thistle2", high = "darkred", guide="colorbar")
P1 <- p + theme_bw() + labs(fill = ""
, title = "Nombre de meurtres aux USA par états", x="", y="")

P1 + scale_y_continuous(breaks=c()) + scale_x_continuous(breaks=c()) + theme(panel.border = element_bla
```

## Nombre de meurtres aux USA par états



Comme nous pouvons le voir, cette carte ne fait que résumer notre jeu de données mais n'a pas beaucoup de sens. En effet, les états les plus peuplés, comme la Californie et le Texas, sont sur-représentés. Si notre but était de montrer le nombre de meurtres par tranche de mille habitants, nous aurions pu le faire en fusionnant notre jeu de données actuel avec un fichier contenant le nombre d'habitants par états et ainsi dégager un taux grâce à cela. Néanmoins, cette analyse, bien que plus pertinente, manque encore de sens, notamment car nous n'avons pas séparé les données par années.

## Seconde partie : étude du taux de résolution des crimes

Nous utilisons globalement la même méthode que dans la partie précédente afin de créer notre carte.

Nous avons dans ce cas besoin d'un taux d'homicide que nous avons mis en pourcentage. Pour arriver à cela, nous avons compté le nombre de meurtres par état, comme précédemment, puis nous avons compté le nombre de crimes résolus par état. Par la suite nous en avons déduit un taux que nous avons ramené en pourcent.

```
data %>% group_by(State) %>% count(State) -> dataPrimaire
data %>% filter(Crime.Solved == "Yes") %>% group_by(State) %>% count(State) -> dataSecondaires

dataMeurtresElucides <- merge(dataSecondaires, dataPrimaire, by="State")
dataMeurtresElucides %>% mutate(proportionElucide = (n.x/n.y)*100) -> dataMeurtresElucides

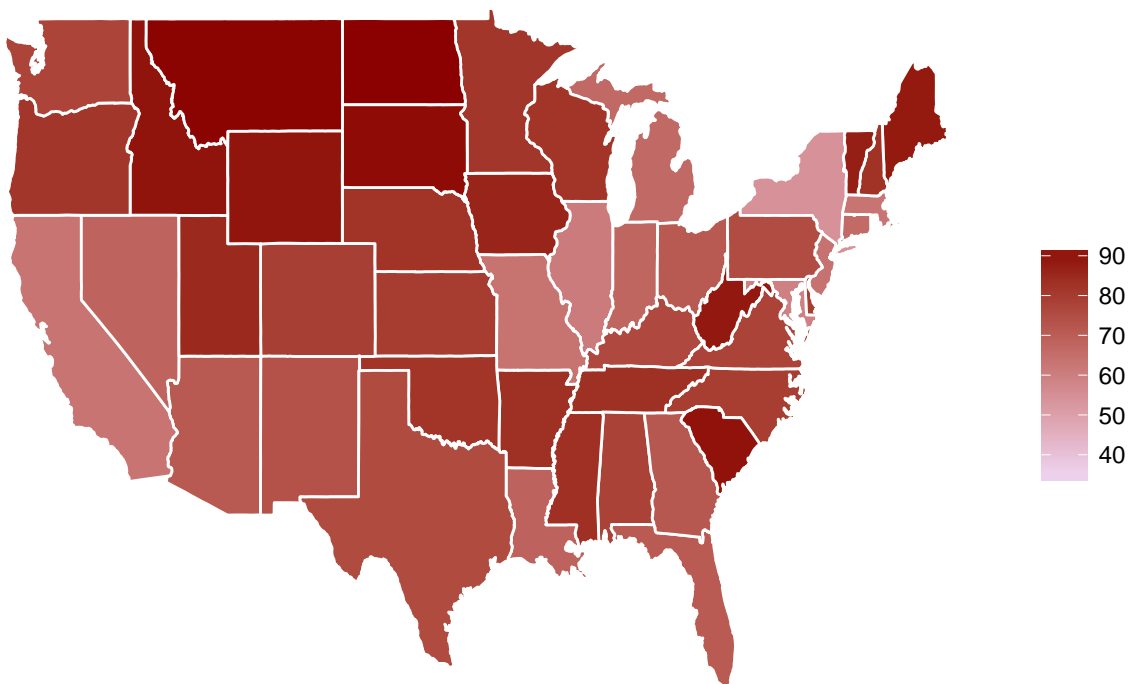
all_states <- map_data("state")
dataMeurtresElucides$region <- tolower(dataMeurtresElucides$State)

dataMeurtresElucides %>% filter(region != "alaska") -> dataMeurtresElucides
dataMeurtresElucides %>% filter(region != "hawaii") -> dataMeurtresElucides
```

```
Total <- left_join(dataMeurtresElucides,all_states, by="region")
```

```
p <- ggplot()
p <- p + geom_polygon(data=Total, aes(x=long, y=lat, group = group, fill=Total$proportionElucide), colour = "black") + scale_fill_continuous(low = "thistle2", high = "darkred", guide="colorbar")
P1 <- p + theme_bw() + labs(fill = "Proportion d'élucidation", title = "Nombre de meurtres aux USA par états", x="", y="")
P1 + scale_y_continuous(breaks=c()) + scale_x_continuous(breaks=c()) + theme(panel.border = element_blank)
```

## Nombre de meurtres aux USA par états



Taux de résolution moyen :

```
head(mean(dataMeurtresElucides$proportionElucide))
```

```
## [1] 76.68543
```

Ecart Type:

```
head(sd(dataMeurtresElucides$proportionElucide))
```

```
## [1] 11.68627
```

Nous avons donc obtenu une seconde carte représentant le taux de résolution par état. Comme nous en avons fait l'hypothèse, le taux n'est pas homogène et l'on peut distinguer plusieurs groupes d'états par ce taux. En effet nous observons un groupe d'état au nord ouest (Idaho, Montana, Utah, Dakota Nord/Sud...) dont le taux de résolution est très élevé par rapport au reste du pays. A l'inverse on peut distinguer des

groupes d'états tel que Illinois, missouri, indiana, michigan qui eux aussi sont proches géographiquement et qui possèdent un taux de résolution bien plus faible que le reste du pays. On remarque donc grâce à cette carte de grosses disparités entre les états. On distingue également le fait que les états sont souvent groupés avec d'autres états ayant un taux proche.

Pour conclure, nous avons validé notre hypothèse de départ en trouvant une corrélation entre l'espace géographique et le taux de résolution des meurtres au USA. Pour pousser notre étude, nous pourrions faire cette même étude avec une dimension supplémentaire : le temps. Aussi, il serait intéressant de découvrir les réelles causes entraînant nos résultats.