

# Rapport final par Bilel Mezrani et Marco Cayuela

```
library(tidyverse)
library(openxlsx)
library(psych)
library(DataExplorer)
library(factoextra)
library(cluster)
library(gridExtra)
library(forecast)
library(fpp2)
library(cowplot)
```

## 1. Présentation du projet

Le F.B.I met à disposition un ensemble de données sur la criminalité, l'effectif policier et les personnes arrêtées chaque année de 1995 à 2019. Ces données peuvent être intéressantes pour visualiser la criminalité aux Etats-Unis et essayer d'expliquer les variations des crimes

Dans notre travail, nous nous sommes focalisés sur deux types de tableaux : l'évolution de la criminalité entre 2000 et 2019, ainsi que la répartition des crimes et des effectifs policiers selon les Etats.

## 2. Nettoyage des données

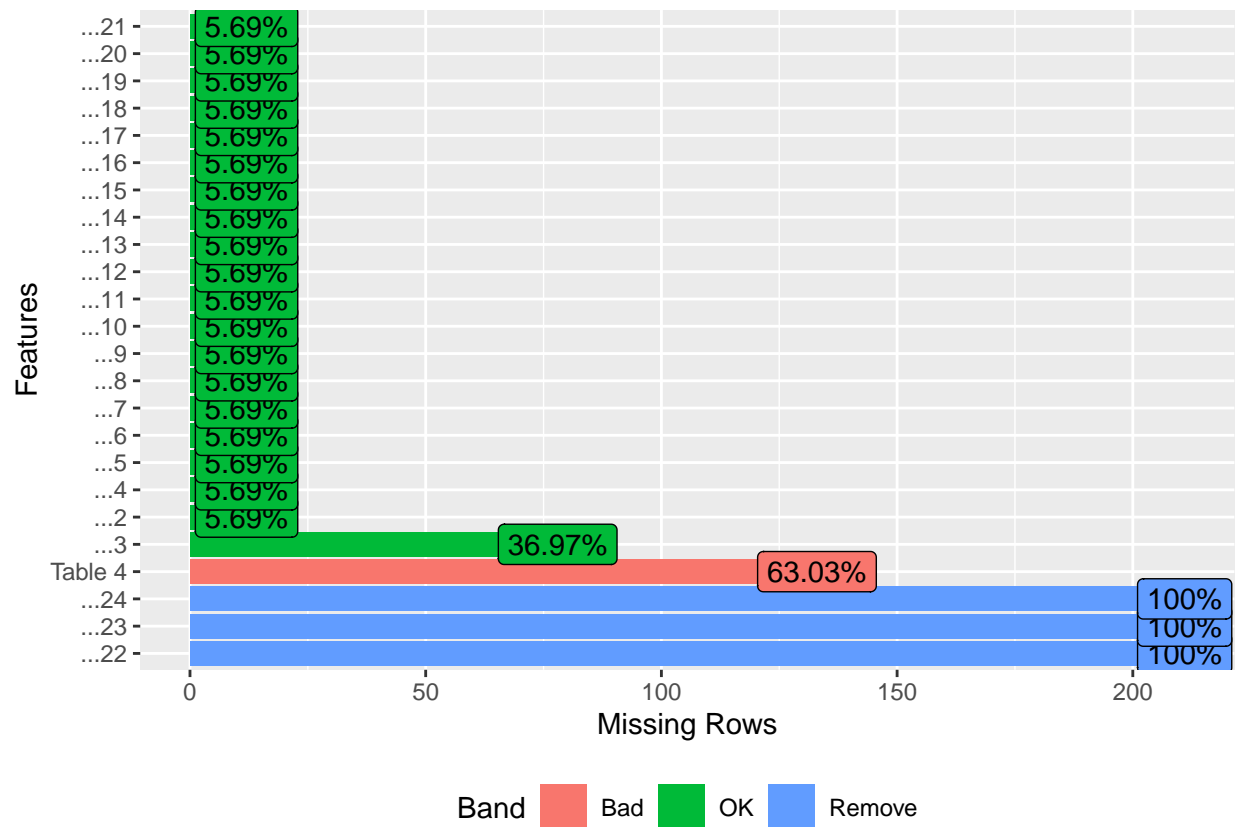
Les données téléchargées sur le site du F.B.I doivent être nettoyées car elles ne sont pas manipulables directement. En effet, il y a de nombreuses valeurs manquantes et les données ne sont pas en format tidy. Par conséquent, pour pouvoir visualiser les données et en tirer des conclusions, il faut effectuer quelques manipulations.

```
fbi_df <- readxl::read_xls("04-Crime by Region, Geographic Division, 2018-2019.xls")
head(fbi_df, n = 6)
```

```
## # A tibble: 6 x 24
##   'Table 4' ...2 ...3 ...4 ...5 ...6 ...7 ...8 ...9 ...10 ...11 ...12
##   <chr>      <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Crime in~ <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 2 by Regio~ <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA> <NA>
## 3 Area      Year Popu~ Viol~ <NA> "Mur~ <NA> "Rap~ <NA> Robb~ <NA> Aggr~
## 4 <NA>      <NA> <NA> <NA> "Rat~ <NA> "Rat~ <NA> "Rat~ <NA> "Rat~ <NA>
## 5 United S~ 2018 3266~ 1252~ "383~ "163~ "5"  "143~ "44" 2812~ "86.~ 8109~
## 6 <NA>      2019 3282~ 1245~ "379~ "164~ "5"  "139~ "42.~ 2679~ "81.~ 8211~
## # ... with 12 more variables: ...13 <chr>, ...14 <chr>, ...15 <chr>,
## #   ...16 <chr>, ...17 <chr>, ...18 <chr>, ...19 <chr>, ...20 <chr>,
## #   ...21 <chr>, ...22 <lgl>, ...23 <lgl>, ...24 <lgl>
```

```
options(repr.plot.width = 18, repr.plot.height = 8)

plot_missing(fbi_df)
```



On remarque que certaines colonnes et lignes sont inutiles. On les supprime. Par ailleurs, on assigne des noms aux colonnes.

On va aussi supprimer la criminalité pour 100000 habitants car cela crée des variables NA. On peut d'ailleurs faire réapparaître ces variables si on le souhaite donc ce n'est pas un problème.

```
crime_per_hundred_thousand <- which(names(fbi) %in% NA)

fbi <- fbi[-1, -crime_per_hundred_thousand]
fbi <- as_tibble(fbi, .name_repair = "unique")
fbi <- fbi %>% fill(Area)

head(fbi, n=6)
```

```
## # A tibble: 6 x 12
##   Area Year Population2 'Violent crime3' 'Murder and \nn~ 'Rape\n(revised~
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Unit~ 2018 326687501 1252399 16374 143765
## 2 Unit~ 2019 328239523 1245410 16425 139815
## 3 Unit~ Perc~ <NA> -0.599999999999999~ 0.299999999999999~ -2.70000000000000~
## 4 Nort~ 2018 56046620 164441 1926 17674
## 5 Nort~ 2019 55982803 163717 1834 17315
```

```
## 6 Nort~ Perc~ <NA> -0.400000000000~ -4.799999999999~ -2
## # ... with 6 more variables: Robbery <chr>, 'Aggravated assault' <chr>,
## # 'Property crime' <chr>, Burglary <chr>, 'Larceny-theft' <chr>, 'Motor
## # vehicle theft' <chr>
```

On crée des nouvelles variables représentant les régions et les divisions géographiques.

```
fbi <- fbi %>% mutate(Region = NA, Geographic_Division = NA )
fbi <- fbi[c(13:14, 1:12)]

fbi <- fbi %>% mutate(Area = str_trim(Area))

head(fbi, n=6)
```

```
## # A tibble: 6 x 14
##   Region Geographic_Divi~ Area Year Population2 'Violent crime3'
##   <lgl> <lgl> <chr> <chr> <chr> <chr>
## 1 NA NA Unit~ 2018 326687501 1252399
## 2 NA NA Unit~ 2019 328239523 1245410
## 3 NA NA Unit~ Perc~ <NA> -0.599999999999~
## 4 NA NA Nort~ 2018 56046620 164441
## 5 NA NA Nort~ 2019 55982803 163717
## 6 NA NA Nort~ Perc~ <NA> -0.400000000000~
## # ... with 8 more variables: 'Murder and \nnonnegligent \nmanslaughter' <chr>,
## # 'Rape\n(revised definition)4' <chr>, Robbery <chr>, 'Aggravated
## # assault' <chr>, 'Property crime' <chr>, Burglary <chr>,
## # 'Larceny-theft' <chr>, 'Motor vehicle theft' <chr>
```

Enfin, on complète en fonction des Etats.

```
fbi$Region[
  which(fbi$Area %in% c(
    "Connecticut", "Maine", "Massachusetts", "New Hampshire", "Rhode Island",
    "Vermont", "New Jersey", "New York", "Pennsylvania"
  ))
] <- "Northeast"

#### Northeast - New England
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Connecticut", "Maine", "Massachusetts", "New Hampshire", "Rhode Island", "Vermont"
  ))
] <- "New England"

#### Northeast - Middle Atlantic
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "New Jersey", "New York", "Pennsylvania"
  ))
] <- "Middle Atlantic"
```

```
##### Midwest #####
```

```
fbi$Region[
  which(fbi$Area %in% c(
    "Illinois", "Indiana", "Michigan", "Ohio", "Wisconsin",
    "Iowa5", "Kansas", "Minnesota", "Missouri", "Nebraska", "North Dakota", "South Dakota"
  )
)
] <- "Midwest"
```

```
#### Midwest - East North Central
```

```
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Illinois", "Indiana", "Michigan", "Ohio", "Wisconsin"
  )
)
] <- "East North Central"
```

```
#### Midwest - West North Central
```

```
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Iowa5", "Kansas", "Minnesota", "Missouri", "Nebraska", "North Dakota", "South Dakota"
  )
)
] <- "West North Central"
```

```
##### South #####
```

```
fbi$Region[
  which(fbi$Area %in% c(
    "Delaware", "District of Columbia6", "Florida",
    "Georgia", "Maryland", "North Carolina7", "South Carolina",
    "Virginia", "West Virginia", "Alabama", "Kentucky", "Mississippi",
    "Tennessee", "Arkansas", "Louisiana", "Oklahoma", "Texas"
  )
)
] <- "South"
```

```
#### South - South Atlantic
```

```
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Delaware", "District of Columbia6", "Florida", "Georgia",
    "Maryland", "North Carolina7", "South Carolina", "Virginia", "West Virginia"
  )
)
] <- "South Atlantic"
```

```
#### South - East South Central
```

```

fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Alabama", "Kentucky", "Mississippi", "Tennessee"
  )
)
] <- "East South Central"

#### South - West South Central
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Arkansas", "Louisiana", "Oklahoma", "Texas"
  )
)
] <- "West South Central"

##### West #####

fbi$Region[
  which(fbi$Area %in% c(
    "Arizona", "Colorado", "Idaho", "Montana", "Nevada",
    "New Mexico", "Utah", "Wyoming", "Alaska", "California",
    "Hawaii", "Oregon", "Washington", "Puerto Rico"
  )
)
] <- "West"

#### West - Mountain
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Arizona", "Colorado", "Idaho", "Montana", "Nevada", "New Mexico", "Utah", "Wyoming"
  )
)
] <- "Mountain"

#### West - Pacific
fbi$Geographic_Division[
  which(fbi$Area %in% c(
    "Alaska", "California", "Hawaii", "Oregon", "Washington", "Puerto Rico"
  )
)
] <- "Pacific"

##### TOTAL #####

# ve Amerika toplamını da Total olarak gösterelim.
fbi$Region[which(fbi$Area == "United States Total5, 6, 7")] <- "Total"

fbi$Geographic_Division[which(fbi$Area == "United States Total5, 6, 7")] <- "Total"

```

```
head(fbi, 10)
```

```
## # A tibble: 10 x 14
##   Region Geographic_Div~ Area Year Population2 'Violent crime3'
##   <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Total Total Unit~ 2018 326687501 1252399
## 2 Total Total Unit~ 2019 328239523 1245410
## 3 Total Total Unit~ Perc~ <NA> -0.5999999999999~
## 4 <NA> <NA> Nort~ 2018 56046620 164441
## 5 <NA> <NA> Nort~ 2019 55982803 163717
## 6 <NA> <NA> Nort~ Perc~ <NA> -0.4000000000000~
## 7 <NA> <NA> New ~ 2018 14829322 38294
## 8 <NA> <NA> New ~ 2019 14845063 36350
## 9 <NA> <NA> New ~ Perc~ <NA> -5.0999999999999~
## 10 North~ New England Conn~ 2018 3571520 7485
## # ... with 8 more variables: 'Murder and \nnonnegligent \nmanslaughter' <chr>,
## # 'Rape\n(revised definition)4' <chr>, Robbery <chr>, 'Aggravated
## # assault' <chr>, 'Property crime' <chr>, Burglary <chr>,
## # 'Larceny-theft' <chr>, 'Motor vehicle theft' <chr>
```

On s'occupe des derniers problèmes liés aux données et aux valeurs manquantes.

```
# Total
fbi_total <- fbi %>% filter(Region == "Total")

# Region
fbi_region <- fbi %>% filter(Area %in% c("Northeast", "Midwest5", "South6, 7", "West"))

# Geographic Division
fbi_geo_div <- fbi %>% filter(Geographic_Division %in%
                             c("New England",
                               "Middle Atlantic",
                               "East North Central",
                               "West North Central5",
                               "South Atlantic6, 7",
                               "East South Central",
                               "West South Central",
                               "Mountain",
                               "Pacific") )

# State
fbi_state <- fbi %>%
  filter(Region != "Total", !is.na(Region))

# Percent Change
fbi_percent_change <- fbi_state %>%
  filter(Year %in% "Percent change")

# States Data
fbi_state <- fbi_state %>%
  filter(Year != "Percent change",
         Area != c("New England", "Middle Atlantic", "East North Central",
                   "West North Central5",
```

```

        "South Atlantic6, 7",
        "East South Central",
        "West South Central",
        "Mountain",
        "Pacific")
    )

# Delete digits at the end of some Areas
for(i in c("","6","7","8","9", "5")){
  fbi_state$Area <- str_remove_all(fbi_state$Area, i)
}

head(fbi_state)

## # A tibble: 6 x 14
##   Region Geographic_Div~ Area   Year  Population2 'Violent crime3'
##   <chr>   <chr>          <chr> <chr> <chr>         <chr>
## 1 North~ New England    Conn~ 2018  3571520      7485
## 2 North~ New England    Conn~ 2019  3565287      6546
## 3 North~ New England    Maine 2018  1339057      1500
## 4 North~ New England    Maine 2019  1344212      1548
## 5 North~ New England    Mass~ 2018  6882635     23424
## 6 North~ New England    Mass~ 2019  6892503     22578
## # ... with 8 more variables: 'Murder and \nnonnegligent \nmanslaughter' <chr>,
## #   'Rape\n(revised definition)4' <chr>, Robbery <chr>, 'Aggravated
## #   assault' <chr>, 'Property crime' <chr>, Burglary <chr>,
## #   'Larceny-theft' <chr>, 'Motor vehicle theft' <chr>

```

On convertit certaines colonnes dans leurs types naturels.

```

fbi_state <- fbi_state %>%
  mutate_at(vars(Population2:`Motor vehicle theft` ),funs(as.numeric)) %>%
  rename(State = Area) %>%
  mutate_at(vars(c(State, Region:Year)), funs(as.factor))
names(fbi_state) <- c("Region",
  "Geographic_Division",
  "State",
  "Year",
  "Population",
  "Violent_crime",
  "Murder_and_nonnegligent_manslaughter",
  "Rape(revised definition)",
  "Robbery",
  "Aggravated_assault",
  "Property_crime",
  "Burglary",
  "Larceny_theft",
  "Motor_vehicle_theft")

```

```
head(fbi_state)
```

```
## # A tibble: 6 x 14
```

```
##   Region Geographic_Divi~ State Year   Population Violent_crime Murder_and_nonn~
##   <fct>   <fct>           <fct> <fct>         <dbl>         <dbl>         <dbl>
## 1 North~ New England     Conn~ 2018       3571520         7485           86
## 2 North~ New England     Conn~ 2019       3565287         6546          104
## 3 North~ New England     Maine 2018       1339057         1500           23
## 4 North~ New England     Maine 2019       1344212         1548           20
## 5 North~ New England     Mass~ 2018       6882635        23424          138
## 6 North~ New England     Mass~ 2019       6892503        22578          152
## # ... with 7 more variables: 'Rape(revised definition)' <dbl>, Robbery <dbl>,
## #   Aggravated_assault <dbl>, Property_crime <dbl>, Burglary <dbl>,
## #   Larceny_theft <dbl>, Motor_vehicle_theft <dbl>
```

Ceci était un exemple de nettoyage de données que l'on a effectué. Pour les autres tableaux, la méthode est sensiblement la même et nous obtenons alors des données au format tidy que nous pouvons plus facilement manipuler avec R.

Par ailleurs, il est important de noter que “Burglary”, “Larceny\_theft” et “Motor\_vehicle\_theft” sont des sous-catégories de “Property\_crime”, tandis que “Murder\_and\_nonnegligent\_manslaughter”, “Rape(revised definition)”, “Robbery” et “Aggravated\_assault” sont des sous-catégories de “Violent\_crime”.

### 3. Visualisation des données

**3.1 Analyse des données pour la criminalité en fonction de l'année** Voici à quoi ressemble le tableau correspondant :

```
fbi_cleaned_names <- read_csv("fbi_cleaned_names.csv")
head(fbi_cleaned_names, n=6)
```

```
## # A tibble: 6 x 11
##   year population1 violent_crime2 murder_and_nonn~ rape_legacy_def~ robbery
##   <dbl>         <dbl>         <dbl>         <dbl>         <dbl>     <dbl>
## 1 2000    281421906    1425486    15586    90178  408016
## 2 2001    285317559    1439480    16037    90863  423557
## 3 2002    287973924    1423677    16229    95235  420806
## 4 2003    290788976    1383676    16528    93883  414235
## 5 2004    293656842    1360088    16148    95089  401470
## 6 2005    296507061    1390745    16740    94347  417438
## # ... with 5 more variables: aggravated_assault <dbl>, property_crime <dbl>,
## #   burglary <dbl>, larceny_theft <dbl>, motor_vehicle_theft <dbl>
```

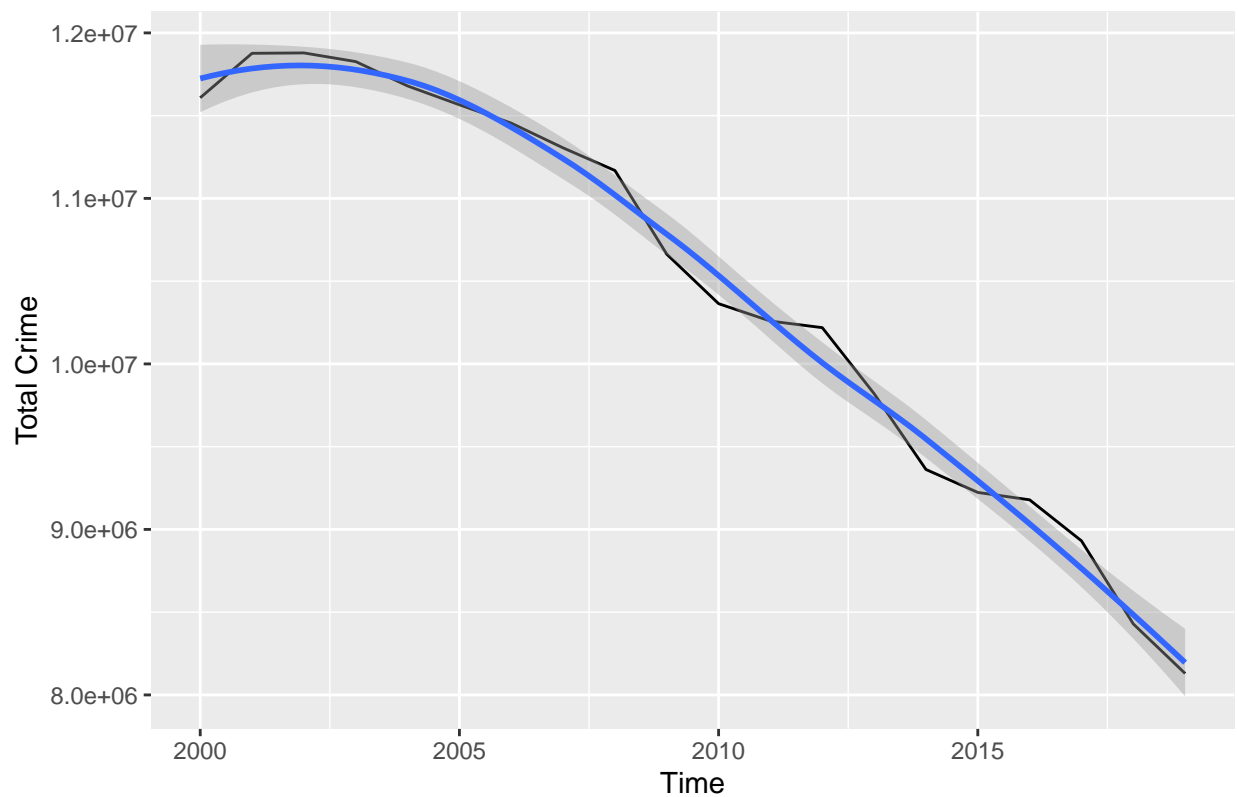
On va commencer par changer le type et transformer le Data Frame en Time Series.

```
pass.ts <- ts(as.matrix(fbi_cleaned_names[,2:11]), start = 2000)
```

Dans un premier temps, on veut avoir une vision d'ensemble de l'évolution de la criminalité en fonction des années.

```
fbi_total <- fbi_cleaned_names %>% dplyr::summarise(year = year, sum_crime = rowSums(fbi_cleaned_names[,
total.ts <- ts(as.matrix(fbi_total[,2]), start = 2000)
autoplot(total.ts, ) + geom_smooth() + labs(y = 'Total Crime')
```

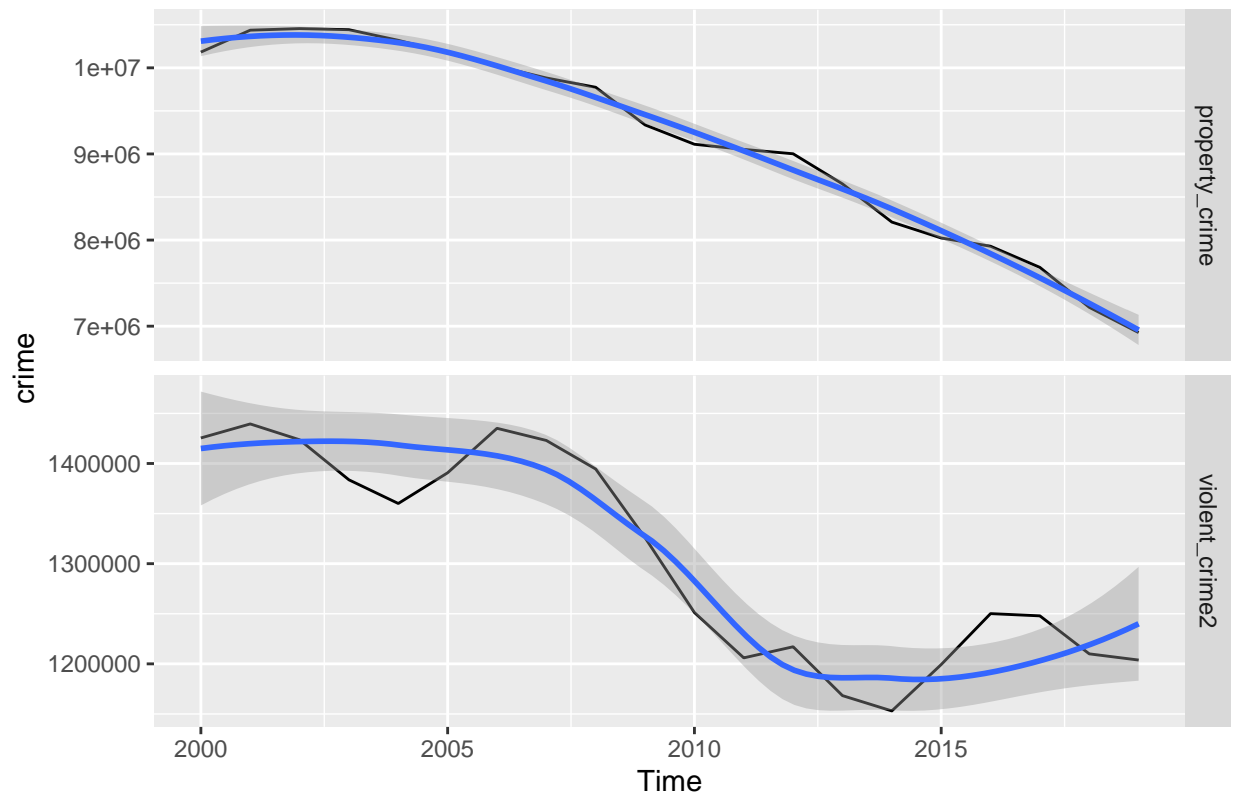




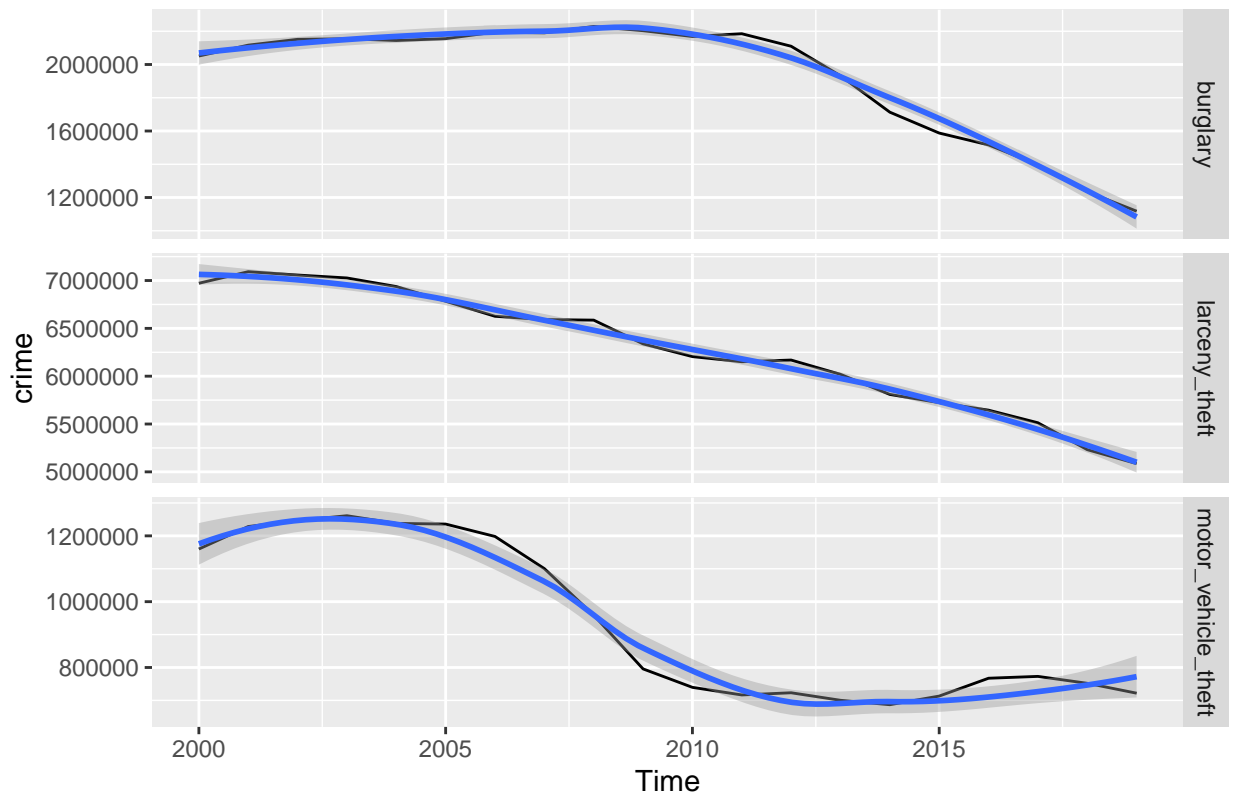
On remarque que la criminalité globale diminue au fil des années. Cependant, en jetant un coup d'œil au tableau, on remarque directement que le nombre de 'Property\_crime' est beaucoup plus élevé que le nombre de 'Violent\_crime'. Par conséquent, la courbe tracée ci-dessus est représentative de l'évolution de "Property\_crime".

On va maintenant tracer l'évolution temporelle de certains crimes dont les nombres enregistrés sont nombreux : property\_crime, burglary, larceny\_theft, violent\_crime, motor\_vehicle\_theft.

```
autoplot(facets =TRUE, pass.ts[,c('property_crime',  
                                  'violent_crime2')]) +  
  geom_smooth()+labs(y='crime')
```



```
autoplot(facets =TRUE, pass.ts[,c('burglary',
                                   'larceny_theft',
                                   'motor_vehicle_theft'))+
geom_smooth()+labs(y='crime')
```



Deux phénomènes sont visibles. On remarque que `property_crime` est en constante décroissance alors que `violent_crime` décroît à partir de 2007 mais commence à recroître à partir de 2015.

- Pour le premier phénomène, on peut interpréter cela par une meilleure prospérité (au moins jusqu'à la crise de 2007 et après 2011) . Les gens ont moins ressenti l'obligation de faire des vols ou des cambriolages au fil des années. Aujourd'hui, les jeunes, majoritairement responsables de ce type de crime, passent plus de temps sur les réseaux sociaux que dehors, ce qui peut expliquer cette diminution. Cependant, il ne faut pas se tromper : les crimes liés aux cambriolages sont beaucoup plus nombreux que les crimes violents.
- On peut fournir plusieurs explications au deuxième phénomène. Premièrement, on observe rapidement une constante croissance de la population depuis 2000. Souvent, une telle augmentation est due à de nouvelles naissances et un vieillissement global de la population. Par conséquent, la population vieillissante est moins encline à faire des crimes. Par ailleurs, la décroissance coïncide aussi avec l'élection d'Obama qui représentait un espoir pour l'Amérique, et notamment les minorités qui sont victimes et acteurs de la criminalité. On parle souvent de "l'effet Obama". Cet effet a été renouvelé en 2012 lors de sa seconde élection. Cependant, cet effet s'essouffle vers 2014 et avec l'élection de Trump en 2016, `violent_crime` réaugmente. Cela est sûrement due à la défiance des minorités envers la politique de Trump. Par ailleurs, il faut aussi prendre en compte la récession après la crise financière de 2007 qui a sûrement joué un rôle dans cette diminution de `violent_crime` à cette période., les gens cherchent plutôt des moyens 'malins' d'arnaquer et non à commettre des crimes.

**3.2 Visualisation de la criminalité en 2019** Nous allons maintenant nous concentrer sur les données de 2019, et notamment la répartition par Etats.

Dans un premier temps, on peut observer quels sont les crimes les plus commis en 2019.

```

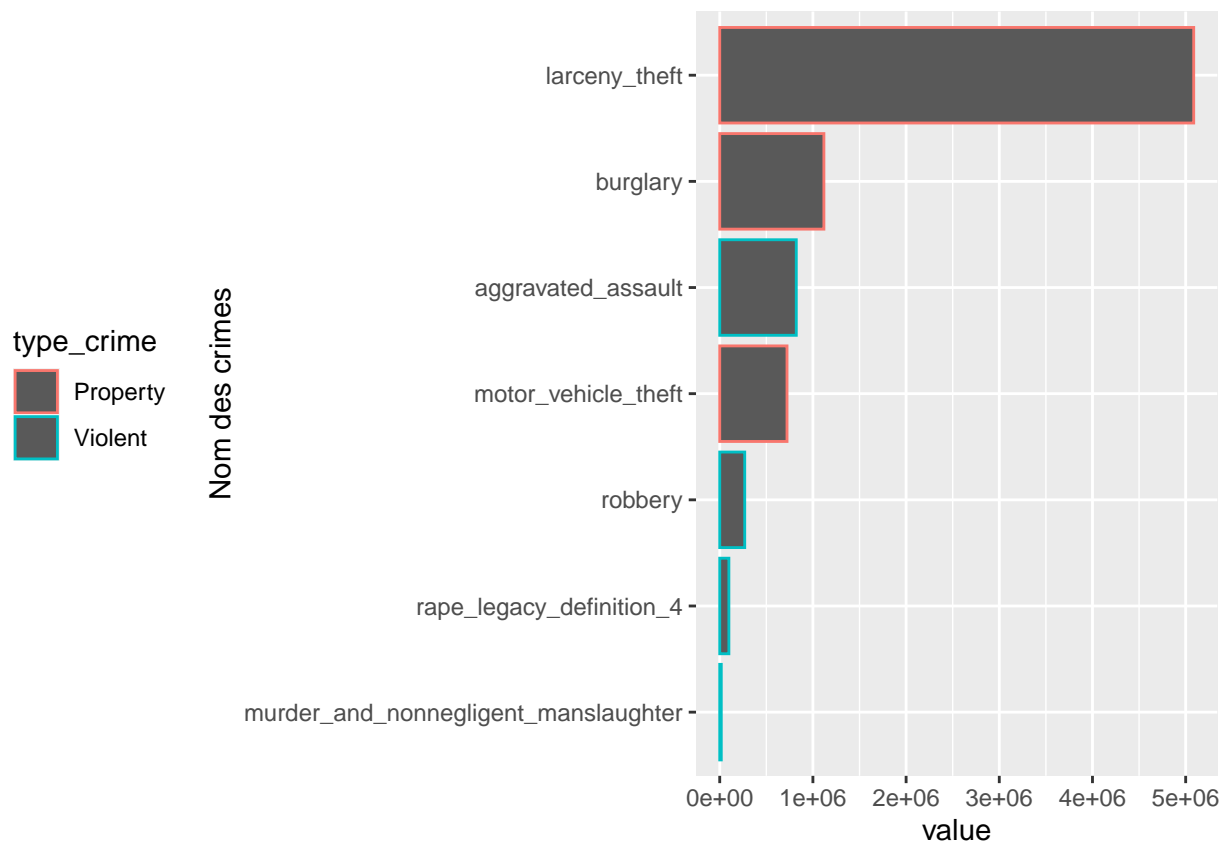
x1<- c("murder_and_nonnegligent_manslaughter",
      "rape_legacy_definition_4",
      "robbery",
      "aggravated_assault",
      "burglary",
      "larceny_theft",
      "motor_vehicle_theft")

y1<-as.numeric(fbi_cleaned_names[20,x1])

z1 <- c("Violent", "Violent", "Violent", 'Violent', 'Property','Property', 'Property')

data <- data.frame(names = x1, value = y1, type_crime = z1)
# Barplot
ggplot(data, aes(x=reorder(names, value), y=value, color = type_crime)) +
  geom_bar(stat = "identity") +
  theme(legend.position="left")+
  coord_flip()+
  labs(x = 'Nom des crimes')

```



On remarque, comme à la partie précédente, que les crimes de propriétés sont beaucoup plus nombreux, ce qui est normal car ces crimes sont réalisés pour de l'argent et non par seul but d'agression.

Nous allons maintenant nous concentrer sur les Etats.

```
fbi_state <- read.csv('fbi_state.csv')
fbi_state <- fbi_state %>% filter(Year %in% c('2019'))
head(fbi_state, n=10)
```

##	X	Region	Geographic_Division	State	Year	Population2
## 1	2	Northeast	New England	Connecticut	2019	3565287
## 2	4	Northeast	New England	Maine	2019	1344212
## 3	6	Northeast	New England	Massachusetts	2019	6892503
## 4	8	Northeast	New England	New Hampshire	2019	1359711
## 5	10	Northeast	New England	Rhode Island	2019	1059361
## 6	12	Northeast	New England	Vermont	2019	623989
## 7	14	Northeast	Middle Atlantic	New Jersey	2019	8882190
## 8	16	Northeast	Middle Atlantic	New York	2019	19453561
## 9	18	Northeast	Middle Atlantic	Pennsylvania	2019	12801989
## 10	20	Midwest	East North Central	Illinois	2019	12671821
##		Violent.crime3	Murder.and..nonnegligent..manslaughter			
## 1		6546				104
## 2		1548				20
## 3		22578				152
## 4		2074				33
## 5		2342				25
## 6		1262				11
## 7		18375				262
## 8		69764				558
## 9		39228				669
## 10		51561				832
##		Rape..revised.definition.4	Robbery	Aggravated.assault	Property.crime	
## 1		771	1929	3742	50862	
## 2		516	188	824	16743	
## 3		2204	3613	16609	81317	
## 4		590	313	1138	16442	
## 5		491	418	1408	16259	
## 6		278	71	902	8888	
## 7		1531	5730	10852	118637	
## 8		6583	18068	44555	267155	
## 9		4351	9743	24465	179665	
## 10		6078	12464	32187	233984	
##		Burglary	Larceny.theft	Motor.vehicle.theft		
## 1		6441	38457	5964		
## 2		2350	13667	726		
## 3		12341	62844	6132		
## 4		1717	13832	893		
## 5		2321	12580	1358		
## 6		1275	7315	298		
## 7		16399	91902	10336		
## 8		27600	226851	12704		
## 9		23354	143921	12390		
## 10		34433	180776	18775		

On va maintenant visualiser la criminalité par Etat sur une carte des Etats-Unis.

```
library(maps)
us_states <- map_data("state")
fbi_state$region <- tolower(fbi_state$State)
us_states_crime <- left_join(us_states, fbi_state)
head(us_states_crime)
```

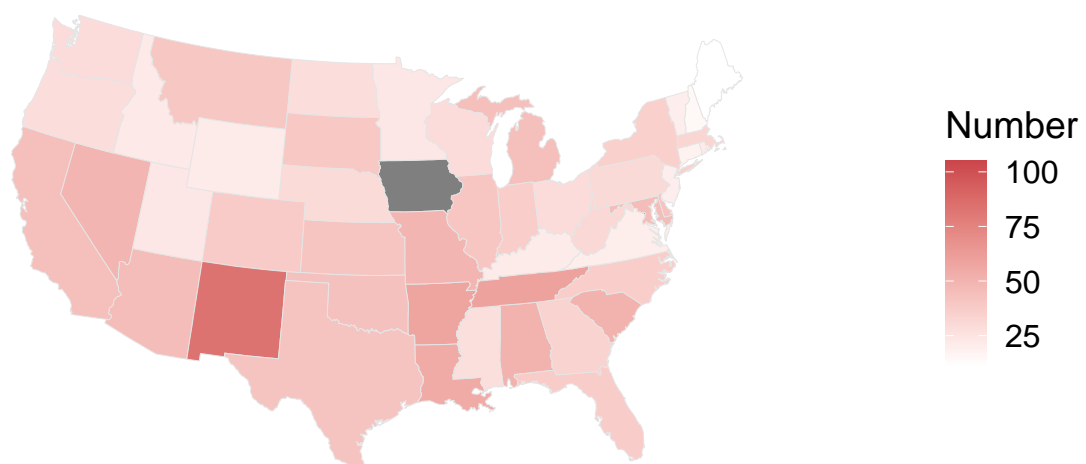
```
##      long      lat group order  region subregion  X Region
## 1 -87.46201 30.38968     1     1 alabama      <NA> 62  South
## 2 -87.48493 30.37249     1     2 alabama      <NA> 62  South
## 3 -87.52503 30.37249     1     3 alabama      <NA> 62  South
## 4 -87.53076 30.33239     1     4 alabama      <NA> 62  South
## 5 -87.57087 30.32665     1     5 alabama      <NA> 62  South
## 6 -87.58806 30.32665     1     6 alabama      <NA> 62  South
## Geographic_Division State Year Population2 Violent.crime3
## 1 East South Central Alabama 2019      4903185          25046
## 2 East South Central Alabama 2019      4903185          25046
## 3 East South Central Alabama 2019      4903185          25046
## 4 East South Central Alabama 2019      4903185          25046
## 5 East South Central Alabama 2019      4903185          25046
## 6 East South Central Alabama 2019      4903185          25046
## Murder.and..nonnegligent..manslaughter Rape..revised.definition.4 Robbery
## 1                                358                    2068      3941
## 2                                358                    2068      3941
## 3                                358                    2068      3941
## 4                                358                    2068      3941
## 5                                358                    2068      3941
## 6                                358                    2068      3941
## Aggravated.assault Property.crime Burglary Larceny.theft Motor.vehicle.theft
## 1              18679              131133      26079          92477          12577
## 2              18679              131133      26079          92477          12577
## 3              18679              131133      26079          92477          12577
## 4              18679              131133      26079          92477          12577
## 5              18679              131133      26079          92477          12577
## 6              18679              131133      26079          92477          12577
```

```
p0 <- ggplot(data = us_states_crime,
             mapping = aes(x = long, y = lat,
                           group = group,
                           fill = (Violent.crime3/Population2)*10000))

p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +
  coord_map(projection = "albers", lat0 = 39, lat1 = 45)

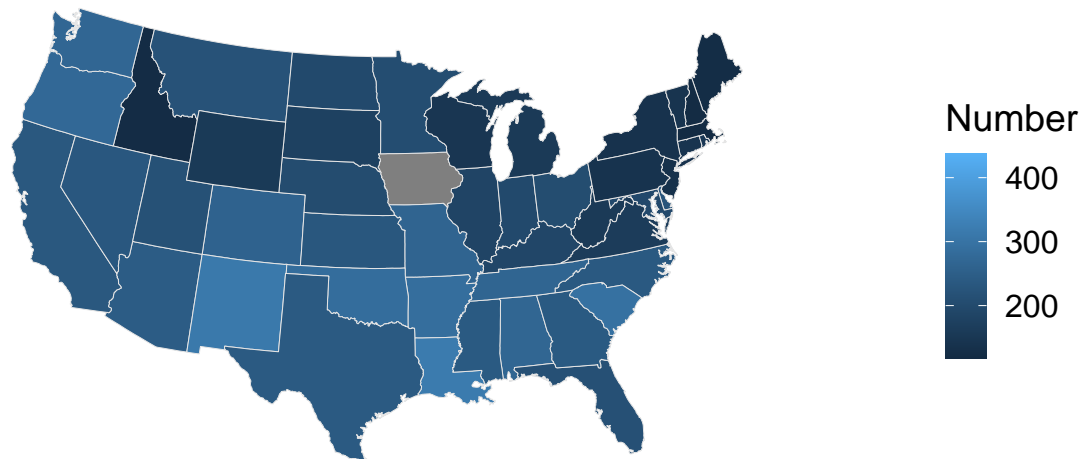
p2 <- p1 + scale_fill_gradient(low = "white", high = "#CB454A") +
  labs(title = "Violent Crime per 10000 inhabitants in each State")
p2 + theme_map() + labs(fill = "Number")
```

## Violent Crime per 10000 inhabitants in each State



```
p0 <- ggplot(data = us_states_crime,  
             mapping = aes(x = long, y = lat,  
                           group = group,  
                           fill = (Property.crime/Population2)*10000))  
  
p1 <- p0 + geom_polygon(color = "gray90", size = 0.1) +  
  coord_map(projection = "albers", lat0 = 39, lat1 = 45)  
  
p1 + labs(title = "Property Crime per 10000 inhabitants in each State") +  
  theme_map() +  
  labs(fill = "Number")
```

## Property Crime per 10000 inhabitants in each State



On remarque que le Sud des Etats-Unis et la Floride sont les régions où le nombre de crimes est le plus grand pour 10,000 habitants. En effet, cette échelle est représentative de la criminalité car le centre des Etats-Unis étant moins peuplé, il est sûr que le nombre de crimes sera plus faible dans cette région. Dans la suite, nous allons aussi étudier l'influence de la police sur la criminalité et voir si les théories économiques, indiquant une baisse de criminalité avec un plus grand nombre de policiers, sont vérifiées en pratique.

```
police_state <- readr::read_csv('police_state.csv')
fbi_state2019 <- readr::read_csv('fbi_state.csv') %>% filter(Year==2019)

police_state <- select(police_state,
  c("State", "Total_law_enforcement_employess"))
police_state <- police_state[-c(1,2),]
police_state <- rename(police_state,
  Total=Total_law_enforcement_employess)

fbi_state2019 <- mutate(fbi_state2019,
  Total_Crime = rowSums(fbi_state2019[c(7,12)]))
fbi_state2019[, "State"] <- apply(X=fbi_state2019[, "State"],
  1, FUN=toupper)

data_joined <- dplyr::left_join(police_state, fbi_state2019)
data_joined[, "Total"] <- apply(data_joined[, "Total"],
  1, FUN=as.integer)
```



```

data_joined <- mutate(data_joined,
  Police_rate= Total/Population2*10000)
data_joined <- mutate(data_joined,
  Violent_Crime_Rate= `Violent crime3`/Population2*10000)
data_joined <- mutate(data_joined,
  Property_Crime_rate= `Property crime`/Population2*10000)
data_joined <- mutate(data_joined,
  Crime_Rate = Property_Crime_rate+Violent_Crime_Rate)
mycols <- c("#0073C2FF", "#EFC000FF", "#868686FF", "#CD534CFF")

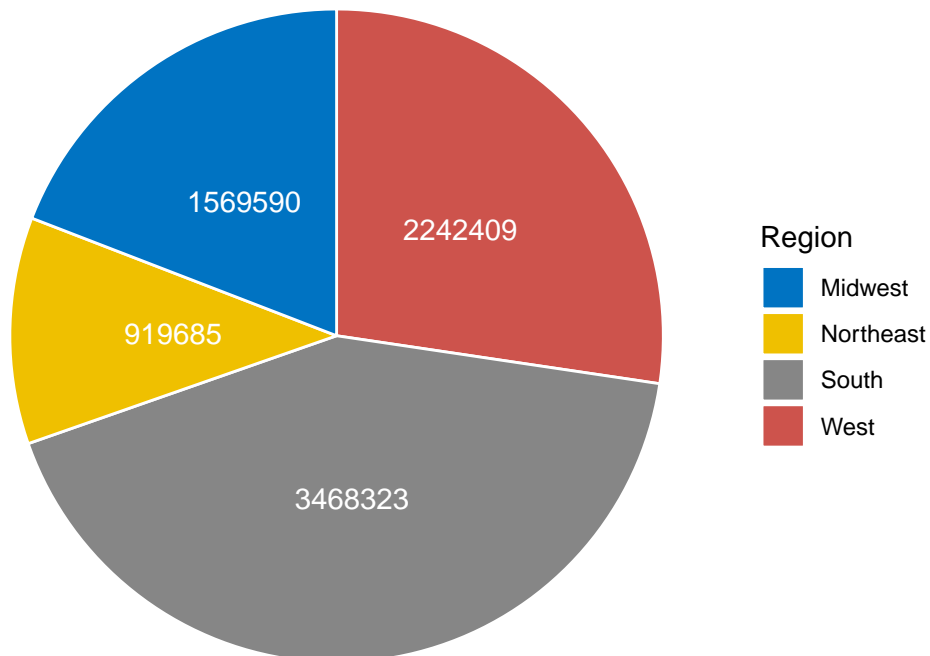
fbi_region <- fbi_state2019[c(2,6,7,12,16)] %>% group_by(Region)%>%
  summarise_if(is.numeric,sum)

fbi_region <- fbi_region %>%
  arrange(desc(Region)) %>%
  mutate(lab.ypos = cumsum(Total_Crime) - 0.5*Total_Crime)

ggplot(fbi_region, aes(x = "", y = Total_Crime, fill = Region)) +
  geom_bar(width = 1, stat = "identity", color = "white") +
  coord_polar("y", start = 0)+
  geom_text(aes(y=lab.ypos, label = Total_Crime), color = "white")+
  scale_fill_manual(values = mycols) +
  theme_void()+
  labs( title="Number of crimes for each region")

```

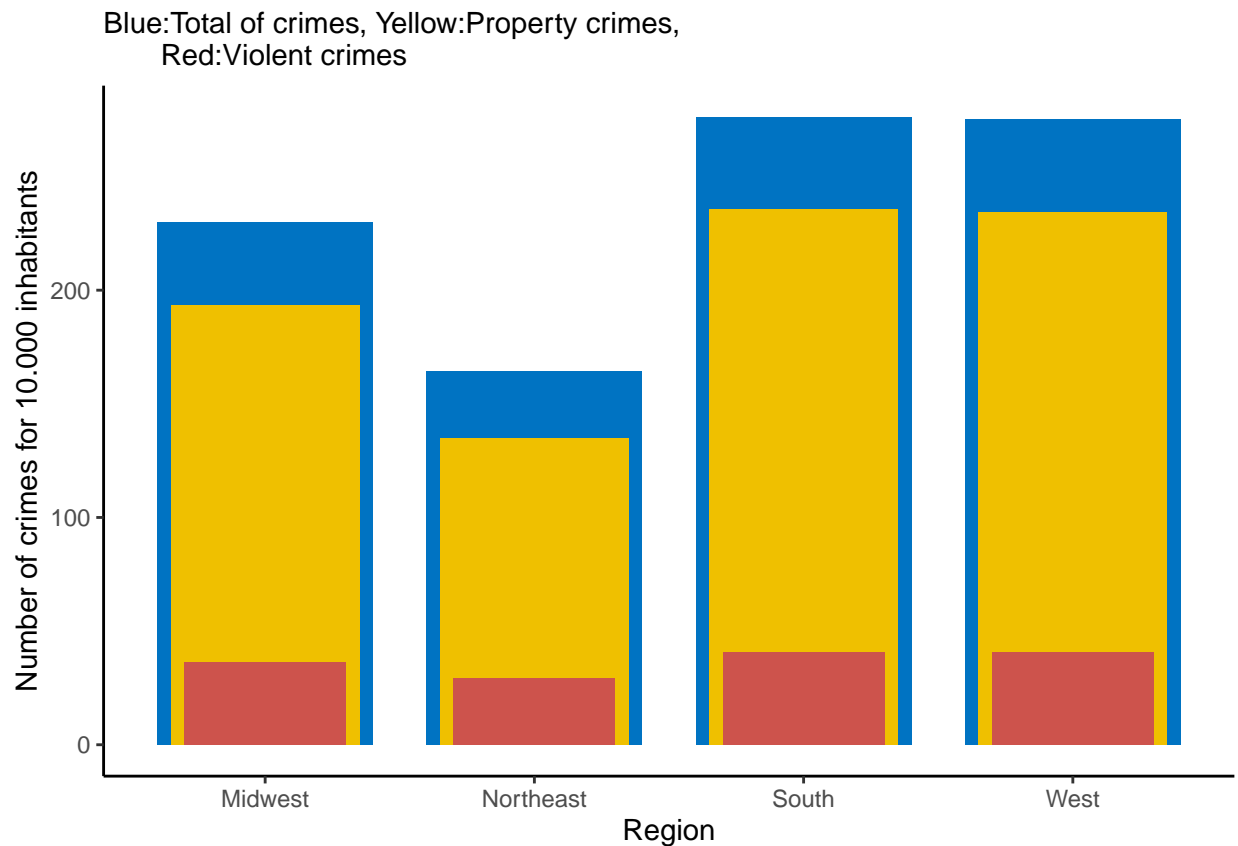
Number of crimes for each region



```

ggplot(data=fbi_region) +
  geom_bar(data = fbi_region, aes(x = Region,
                                y = Total_Crime/Population2*10000),
          width = 0.8,
          stat = 'identity',
          fill="#0073C2FF") +
  geom_bar(data = fbi_region,
          aes(x = Region,
              y = `Property crime`/Population2*10000),
          width = 0.7,
          stat = 'identity',
          fill = "#EFC000FF") +
  geom_bar(data = fbi_region ,
          aes(x = Region, y = `Violent crime3`/Population2*10000),
          width = 0.6,
          stat = 'identity',
          fill = "#CD534CFF") +
  theme_classic()+
  labs(y="Number of crimes for 10.000 inhabitants",
       subtitle = "Blue:Total of crimes, Yellow:Property crimes,
       Red:Violent crimes")

```

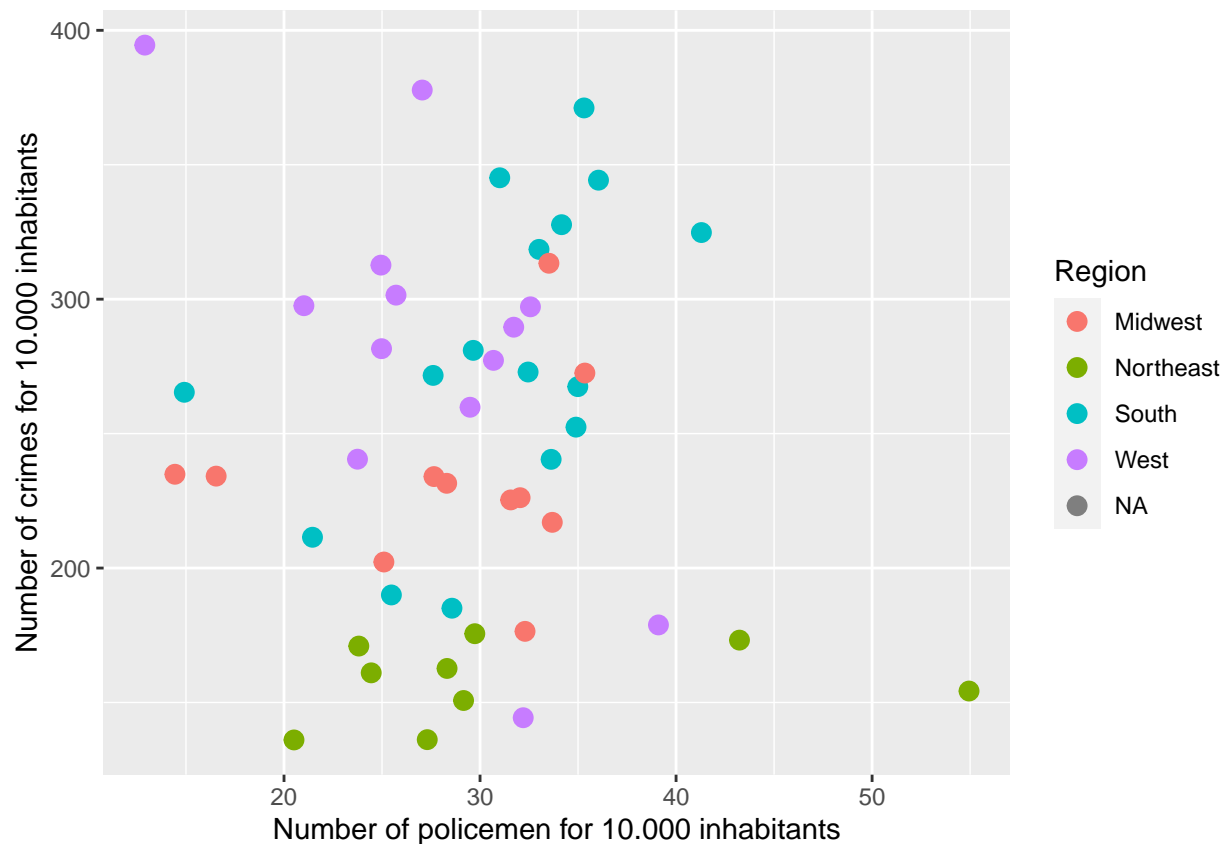


On a choisi de représenter dans un premier temps ces deux visualisations car on observe que le nombre de crimes est plus élevé dans le Sud et l'Ouest des Etats-Unis. En mettant en parallèle avec la criminalité pour 10,000 habitants, on en déduit que le nombre de crime est élevé à l'Ouest parce qu'il y a beaucoup de population. On en déduit aussi que la criminalité dans le Sud est la plus élevée des Etats-Unis.

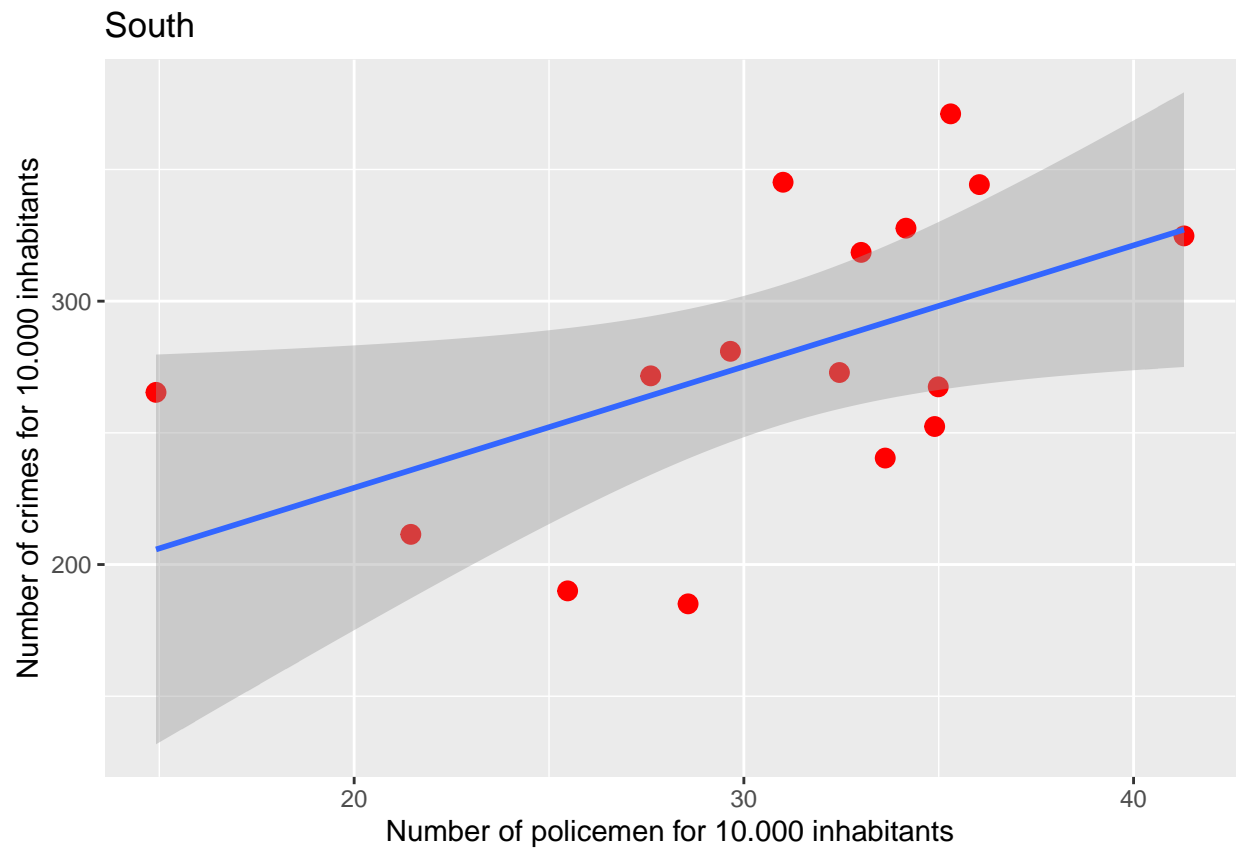
On va maintenant étudier l'influence de la Police sur la criminalité. La question logique qu'on peut se poser est la suivante : Un plus grand nombre de policiers implique-t-il nécessairement un moins grand nombre de crimes ?

Comme précédemment, on va prendre le nombre de policiers pour 10,000 habitants et le nombre de crimes pour 10,000 habitants.

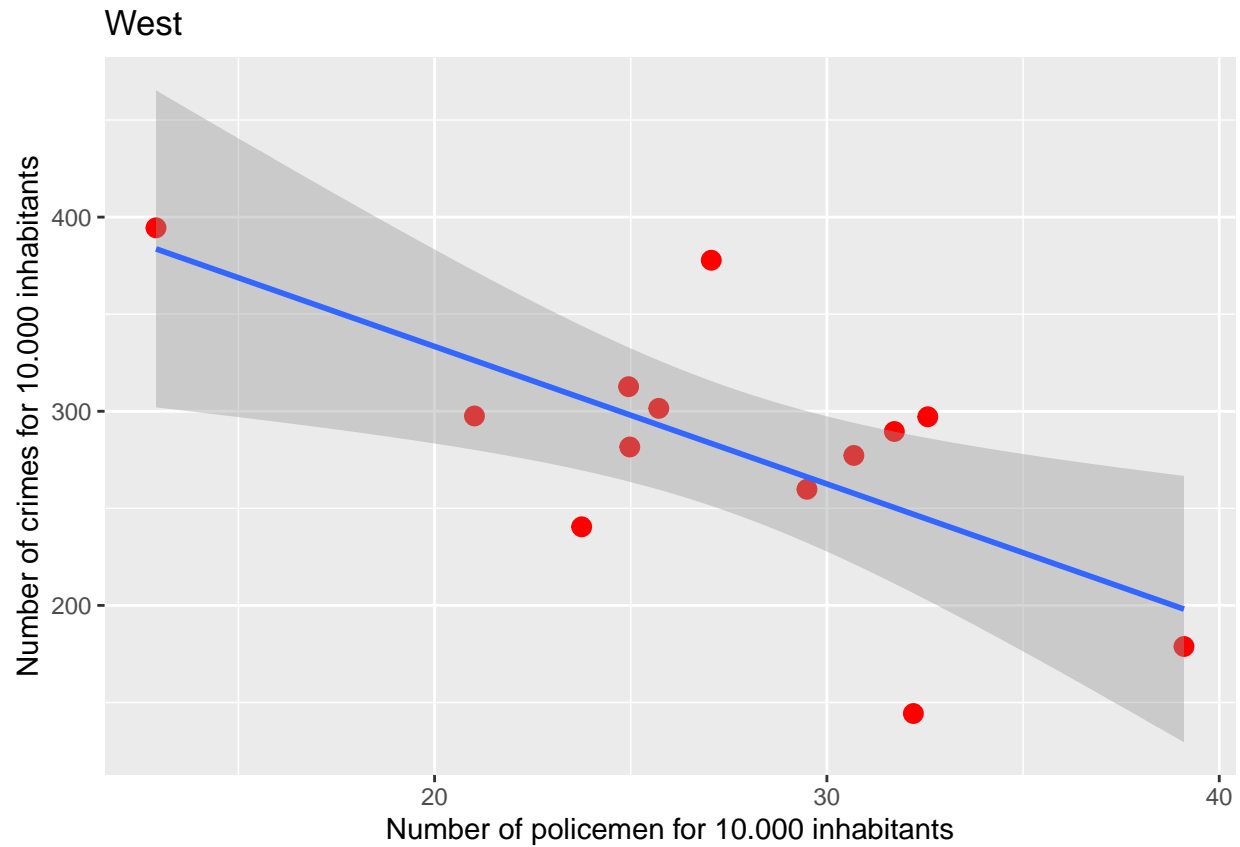
```
ggplot(data_joined,
       mapping=aes(x=Police_rate,
                   y=Crime_Rate, color=Region))+
  geom_point(size=3)+
  labs(x="Number of policemen for 10.000 inhabitants",
       y="Number of crimes for 10.000 inhabitants")
```



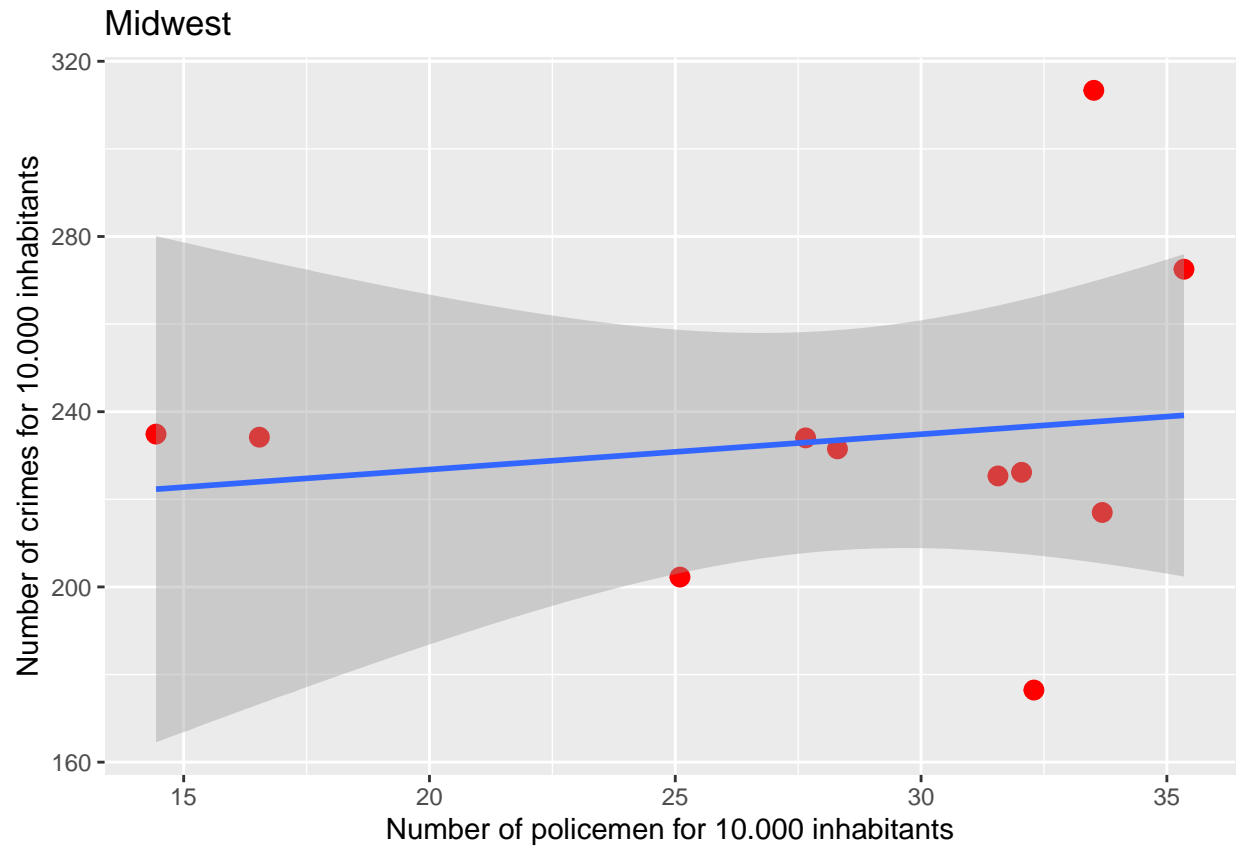
```
ggplot(data_joined %>% filter(Region=="South"),
       mapping=aes(x=Police_rate, y=Crime_Rate))+
  geom_point(size=3, color='red')+geom_smooth(method=lm)+
  labs(x="Number of policemen for 10.000 inhabitants",
       y="Number of crimes for 10.000 inhabitants")+
  labs(title = "South")
```



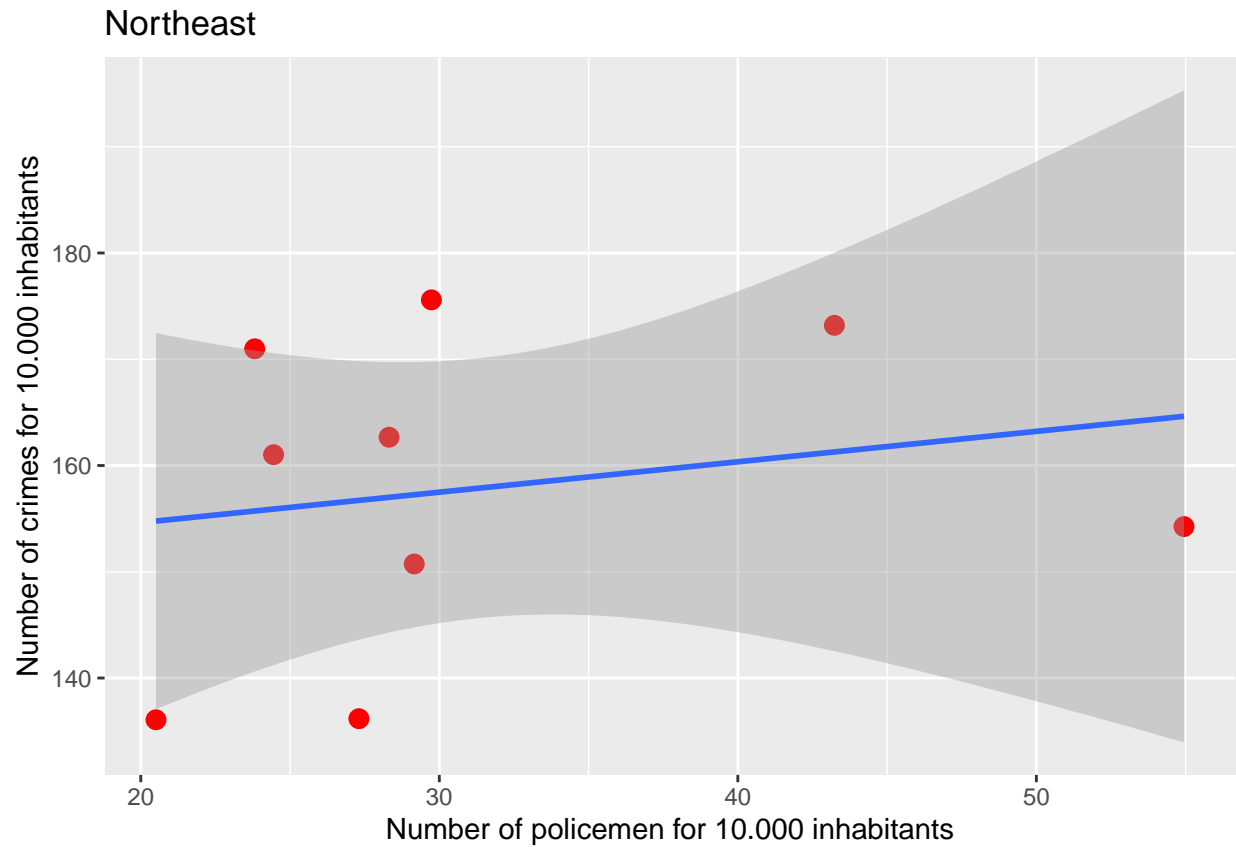
```
ggplot(data_joined %>% filter(Region=="West"),
  mapping=aes(x=Police_rate, y=Crime_Rate))+
  geom_point(size=3, color='red')+geom_smooth(method=lm)+
  labs(x="Number of policemen for 10.000 inhabitants",
    y="Number of crimes for 10.000 inhabitants")+
  labs(title = "West")
```



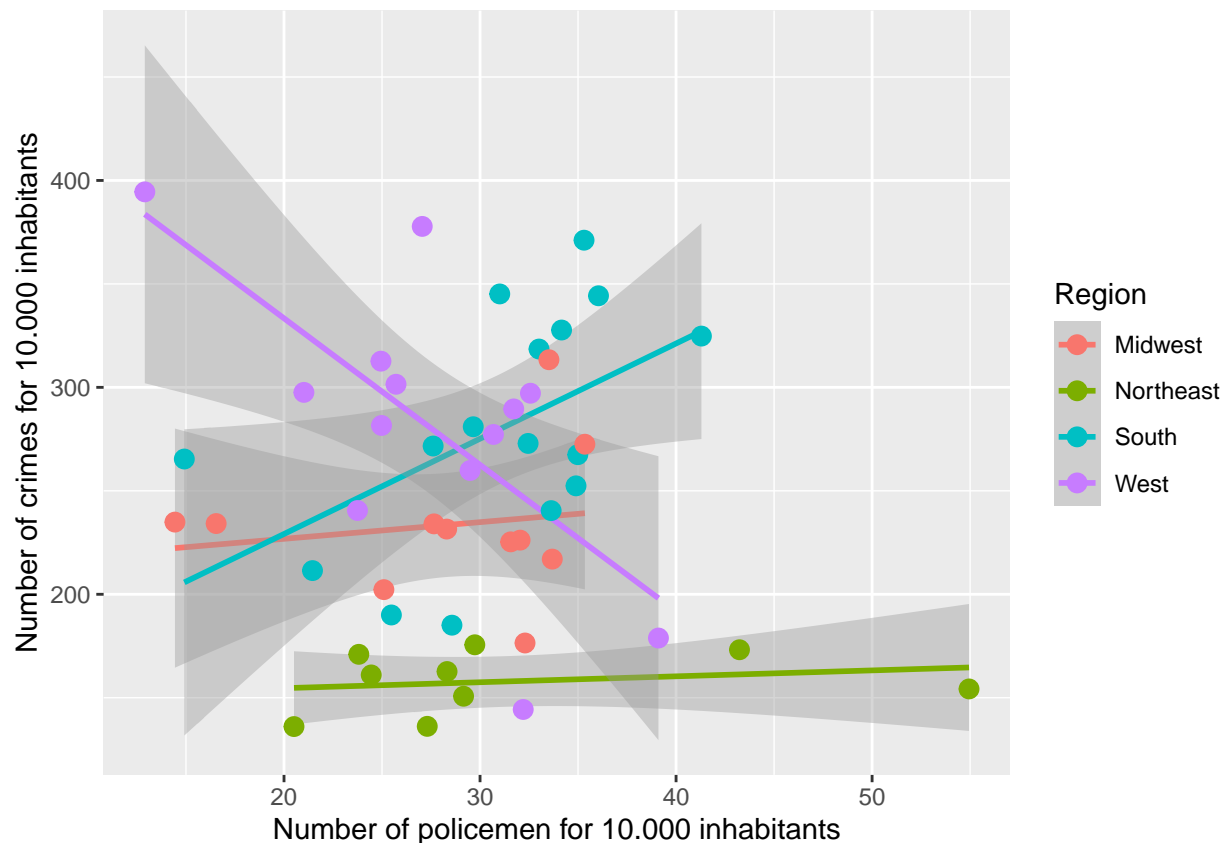
```
ggplot(data_joined %>% filter(Region=="Midwest"),
  mapping=aes(x=Police_rate, y=Crime_Rate))+
  geom_point(size=3, color='red')+geom_smooth(method=lm)+
  labs(x="Number of policemen for 10.000 inhabitants",
    y="Number of crimes for 10.000 inhabitants")+
  labs(title = "Midwest")
```



```
ggplot(data_joined %>% filter(Region=="Northeast"),
  mapping=aes(x=Police_rate, y=Crime_Rate))+
  geom_point(size=3, color='red')+geom_smooth(method=lm)+
  labs(x="Number of policemen for 10.000 inhabitants",
    y="Number of crimes for 10.000 inhabitants")+
  labs(title = "Northeast")
```



```
ggplot(data_joined, mapping=aes(x=Police_rate, y=Crime_Rate, color=Region))+geom_smooth(method=lm)+geom.  
  labs(x="Number of policemen for 10.000 inhabitants",  
        y="Number of crimes for 10.000 inhabitants")
```



- Le nuage de points indiquant taux de criminalité en fonction du nombre de policier pour 10.000 habitants dans la région “South” permet d’obtenir une bonne régression linéaire. Elle montre donc une proportionnalité entre ces deux mesures. Plus il y a de police déployée dans un état, plus le taux de criminalité y est important. On observe donc le phénomène inverse que dans la région “West”. Essayons d’expliquer cela en analysant d’abord le rapport à la violence de cette région.
- Le “South” est connu depuis longtemps pour être une région particulièrement violente. Les visualisations précédentes le prouvent d’ailleurs puisque c’est celle où il y a le plus de crimes et où le taux de criminalité est le plus élevé aux États-Unis. Ainsi, peu importe le nombre de policiers, cette partie des États-Unis serait encline au crime. Il y a plusieurs tentatives d’explications à cela.
- D’abord, les “Southern men” semblent avoir une “culture de l’honneur” importante. Une étude a été publiée par “the Journal of Personality and Social Psychology” lors de laquelle des gens ont été insulté aléatoirement dans les régions “South” and “North”. Les “Southern men” réagissent en moyenne beaucoup plus agressivement que les autres. De plus, on sait que pauvreté et criminalité sont étroitement liées. Or la région “South” est très pauvre. Effectivement, les états avec les “revenus par tête” les plus bas se trouvent dans le “South” (Mississippi, West Virginia, Arkansas,...). Enfin, un dernier facteur qui peut être pris en compte est le climat. Les fortes chaleurs dans le “South” pourraient favoriser la criminalité car les conditions de vie sont alors plus difficiles. Tous ces éléments pris en compte peuvent expliquer le pourquoi de cette criminalité dans cette région qui est ancrée culturellement depuis longtemps. La corrélation entre policiers et criminalité peut alors être expliquée par les efforts mis en place pour freiner cette criminalité. Plus d’employés de police sont déployés dans les états les plus touchés. Mais contrairement à la région “West”, cela ne semble pas avoir un réel impact.



## 4. Conclusion

Grâce à ce projet, on peut se rendre compte qu'une bonne visualisation des données facilite l'interprétation permet d'identifier certains comportements, que ce soit en fonction du temps ou de l'espace. D'autres tables auraient pu être jointes pour accentuer un peu plus l'analyse, notamment le lien entre le nombre de personnes issus de minorités dans un Etat et la criminalité. Les tables fournies par le F.B.I sont exhaustives, donc quiconque souhaitant prolonger l'analyse de ce projet pourra proposer de nombreux angles d'attaque.