

PROJET_PLATA_CORP

Mona OSMAN,Alireza SEYF, Sami MANSOUR, Lola TOSETTO

2024-03-16

##Présentation de notre Base et Étude #Titre DCJS Adult Arrests by County

#Source Les données utilisées dans notre projet proviennent du Département des services de justice pénale de l'État de New York (DCJS), qui a pour mission d'améliorer la sécurité publique. Le DCJS recueille des données à partir de son système d'historique criminel informatisé (CCH), qui est le référentiel central des informations sur l'historique criminel dans l'État de New York. Le CCH contient les dossiers d'histoire criminelle de toutes les personnes arrêtées et poursuivies depuis 1970.

Les données sur les arrestations d'adultes pour des infractions pouvant faire l'objet de relevés d'empreintes digitales sont collectées à partir de diverses agences tout au long du traitement d'une affaire, depuis l'arrestation initiale jusqu'aux décisions judiciaires. Les infractions pouvant faire l'objet de relevés d'empreintes digitales comprennent les délits graves et les délits mineurs définis dans la loi pénale de l'État de New York, ainsi que les violations spécifiques. Les données sont mises à jour annuellement et peuvent être sujettes à des modifications au fil du temps en raison de l'ajout d'informations manquantes antérieurement.

Il convient de noter que ces données sont limitées aux arrestations d'adultes pour des infractions majeures et mineures pouvant faire l'objet de relevés d'empreintes digitales, et que les demandes d'informations supplémentaires peuvent être adressées au DCJS ou aux agences d'arrestation de l'État ou locales. Ces données sont destinées à permettre au public un accès rapide et facile à des informations publiques, bien que des erreurs humaines ou mécaniques soient possibles.

#Description colonnes Un document Data dictionary est fournie avec cette base de donnée, pour expliquer chaque colonne en plus de la recolte de donnée, nous nous focalisons surtout sur les variables Drug Felony (Délits liés aux drogues) qui donne le nombre d'arrestations d'adultes pour des délits liés aux drogues, y compris les infractions relatives aux substances contrôlées et à la marijuana et sur Violent Felony (Délits violents) qui indique le nombre d'arrestations d'adultes pour des délits violents, tels que définis dans la loi pénale de l'État de New York.

#Problématique Existe-t-il une corrélation entre le taux de criminalité liée à la drogue (drug felony) et le taux liée à la violence(violent felony) ? Dans quelle mesure la criminalité liée à la drogue influence-t-elle la criminalité liée à la violence ?

##Chargement des packages

```
library(dplyr)
```

```

## Warning: package 'dplyr' was built under R version 4.2.3
##
## 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(readr)

## Warning: package 'readr' was built under R version 4.2.3

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.2.3
## Warning: package 'ggplot2' was built under R version 4.2.3
## Warning: package 'tibble' was built under R version 4.2.3
## Warning: package 'tidyr' was built under R version 4.2.3
## Warning: package 'purrr' was built under R version 4.2.3
## Warning: package 'stringr' was built under R version 4.2.3
## Warning: package 'forcats' was built under R version 4.2.3
## Warning: package 'lubridate' was built under R version 4.2.3

## — Attaching core tidyverse packages ————— tidyverse
2.0.0 —
## ✓ forcats    1.0.0      ✓ stringr    1.5.1
## ✓ ggplot2    3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.3      ✓ tidyr     1.3.1
## ✓ purrr      1.0.2

## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all
conflicts to become errors

library(ggplot2)
library(tidyr) # Pour pivot_longer

##Chargement des Données

```

```
Data <- read_delim("adult-arrests-by-county-beginning-1970.csv",delim=",")

## Rows: 3181 Columns: 13
## — Column specification

```

```
## Delimiter: ","
## chr (1): County
## dbl (12): Year, Total, Felony Total, Drug Felony, Violent Felony, DWI
Felony...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

NB: On s'assure bien du typage des données.

On vérifie les types de données et les valeurs manquantes. On crée deux nouvelles colonnes, pour avoir un bon nommage, en s'assurant du typage tidy de nos variables : drug felony et violent felony.

```
str(Data)

## spc_tbl_ [3,181 × 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ County : chr [1:3181] "Albany" "Allegany" "Bronx" "Broome"
...
## $ Year : num [1:3181] 2019 2019 2019 2019 2019 ...
## $ Total : num [1:3181] 6195 751 40450 4795 1520 ...
## $ Felony Total : num [1:3181] 2274 209 14954 1420 454 ...
## $ Drug Felony : num [1:3181] 416 33 2798 291 130 ...
## $ Violent Felony : num [1:3181] 460 44 5786 310 65 ...
## $ DWI Felony : num [1:3181] 98 15 117 60 55 25 58 55 17 56 ...
## $ Other Felony : num [1:3181] 1300 117 6253 759 204 ...
## $ Misdemeanor Total : num [1:3181] 3921 542 25496 3375 1066 ...
## $ Drug Misdemeanor : num [1:3181] 673 60 4105 657 162 ...
## $ DWI Misdemeanor : num [1:3181] 586 159 756 360 282 135 419 207 81
273 ...
## $ Property Misdemeanor: num [1:3181] 1463 106 6190 1289 237 ...
## $ Other Misdemeanor : num [1:3181] 1199 217 14445 1069 385 ...
## - attr(*, "spec")=
## .. cols(
## .. County = col_character(),
## .. Year = col_double(),
## .. Total = col_double(),
## .. `Felony Total` = col_double(),
## .. `Drug Felony` = col_double(),
## .. `Violent Felony` = col_double(),
## .. `DWI Felony` = col_double(),
## .. `Other Felony` = col_double(),
## .. `Misdemeanor Total` = col_double(),
## .. `Drug Misdemeanor` = col_double(),
## .. `DWI Misdemeanor` = col_double(),
```

```
## .. `Property Misdemeanor` = col_double(),
## .. `Other Misdemeanor` = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

```
summary(Data)
```

```
##      County      Year      Total      Felony Total
## Length:3181    Min.   :1970    Min.   :      1    Min.   :      0
## Class :character 1st Qu.:1982    1st Qu.:     817    1st Qu.:    181
## Mode  :character Median :1994    Median :    1478    Median :    349
##      Mean   :1994    Mean   :    6531    Mean   :   2224
##      3rd Qu.:2007    3rd Qu.:    4027    3rd Qu.:   1069
##      Max.   :2019    Max.   :   107786    Max.   :  44632
##      Drug Felony    Violent Felony    DWI Felony    Other Felony
## Min.   :      0.0    Min.   :      0.0    Min.   :      0.00    Min.   :      0.0
## 1st Qu.:    17.0    1st Qu.:    37.0    1st Qu.:   15.00    1st Qu.:    98.0
## Median :    47.0    Median :    74.0    Median :   35.00    Median :   187.0
## Mean   :   498.3    Mean   :   677.5    Mean   :   65.56    Mean   :   982.2
## 3rd Qu.:   179.0    3rd Qu.:   255.0    3rd Qu.:   74.00    3rd Qu.:   554.0
## Max.   : 17442.0    Max.   : 16217.0    Max.   :  613.00    Max.   : 15467.0
## Misdemeanor Total Drug Misdemeanor DWI Misdemeanor Property Misdemeanor
## Min.   :      1    Min.   :      0.0    Min.   :      0.0    Min.   :      0
## 1st Qu.:    628    1st Qu.:    22.0    1st Qu.:   173.0    1st Qu.:   148
## Median :   1157    Median :    65.0    Median :   328.0    Median :   313
## Mean   :   4307    Mean   :   870.7    Mean   :   617.2    Mean   :  1313
## 3rd Qu.:   2958    3rd Qu.:   242.0    3rd Qu.:   666.0    3rd Qu.:   813
## Max.   : 73366    Max.   : 29471.0    Max.   : 8954.0    Max.   : 33334
## Other Misdemeanor
## Min.   :      0
## 1st Qu.:   234
## Median :   444
## Mean   :  1506
## 3rd Qu.:  1141
## Max.   : 24875
```

```
donnees <- na.omit(Data)
```

```
donnees$drug_felony <- as.numeric(Data$"Drug Felony")
donnees$violent_felony <- as.numeric(Data$"Violent Felony")
```

##GRAPHES On souhaite visualiser l'évolution de nos variables, drug felony et violent felony, au cours du temps sur un même graphique. On a choisi de réaliser un graphe simple ayant en ordonnée le nombre de délits de chacune de nos varibales et en abscisse les années (1970-2019).

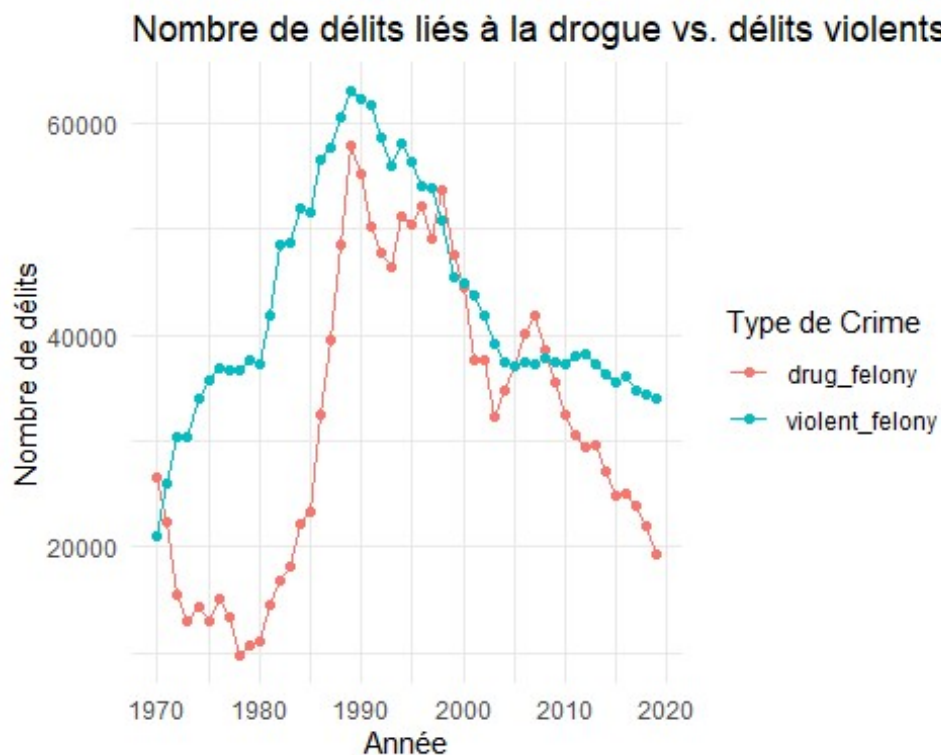
```
# Transformer Les données au format Long
donnees_long <- pivot_longer(donnees,
                             cols = c("drug_felony", "violent_felony"),
                             names_to = "Type_of_Crime",
                             values_to = "Count")
```

```

donnees_long_sum <- aggregate(Count ~ Year + Type_of_Crime, data =
donnees_long, sum) #aggregation sur les années qui fait la somme de
drug_felony et violent_felony pour chaque année.

# Créer le graphique
ggplot(donnees_long_sum, aes(x = Year, y = Count, color = Type_of_Crime)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Nombre de délits liés à la drogue vs. délits violents par
année",
       x = "Année",
       y = "Nombre de délits",
       color = "Type de Crime")

```



On peut remarquer que les deux variables suivent des tendances similaires après 1990. Avant cela, il y avait peu d'observations de la variable drug felony, contrairement à celle de violent felony. Donc nous faisons notre étude sur ces deux intervalles de temps différents. Pour expliquer ce changement de tendances, on a établie une timeline:

Années 1960-70 : Les États-Unis constatent une augmentation de l'usage des drogues au milieu du mouvement de contre-culture. Le président Nixon déclare que l'abus de drogues est "l'ennemi public numéro un" en 1971, ce qui mène à un accent sur des politiques antidrogues strictes. 1973 : L'État de New York promulgue les lois Rockefeller sur la drogue, imposant de lourdes peines pour possession et vente de narcotiques, ayant un impact significatif sur les arrestations liées à la drogue et sur les populations carcérales.

1975 : New York fait face à une grave crise fiscale conduisant à des coupes dans les services, y compris la police, pendant une période d'augmentation des taux de criminalité et d'instabilité économique. Les années 1980 :

Début des années 1980 : La prévalence du crack commence à augmenter, conduisant à une hausse de la consommation de drogues et des crimes associés. Milieu des années 1980 : Une augmentation notable des décès liés à la drogue indique une consommation plus élevée, avec une forte augmentation des arrestations pour délits liés à la drogue à New York. Cette période voit une hausse des crimes liés à la propriété, tels que les cambriolages et les vols de véhicules, qui sont corrélés à l'augmentation de l'usage des drogues. 1986-1989 : Les taux d'homicides et les crimes violents augmentent, avec New York faisant face à une grave épidémie de crack et à la criminalité violente associée.

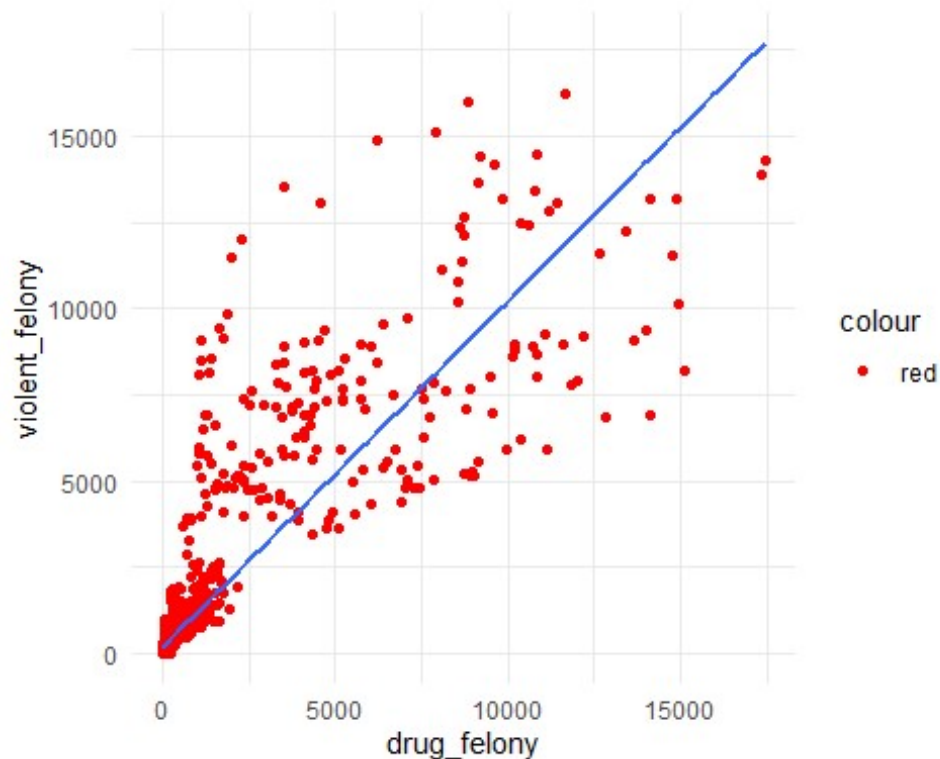
Nos sources: Public Enemy Number One: A Pragmatic Approach to America's Drug Problem-publié le 29 juin 2016 sur le site Web de Richard Nixon Foundation.

Ainsi, on peut émettre l'hypothèse que le peu d'observation de drug felony avant 1990 est dû au fait qu'il n'y avait pas autant de pression politique sur les arrestations liées aux drogues, et qu'après 1990, l'événement politique War On Drugs explique l'augmentation de la variable Drug Felony.

Pour commencer, on utilise un graphique pour visualiser la ligne de régression linéaire, pour l'intervalle d'années de bases (1970-1990)

```
# Nuage de points pour visualiser la relation
ggplot() +
  geom_point(data = donnees, aes(x = drug_felony, y = violent_felony, color =
"red")) +
  geom_smooth(data = donnees, aes(x = drug_felony, y = violent_felony),
method = lm, se = FALSE) +
  scale_color_manual(values = c("red")) +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



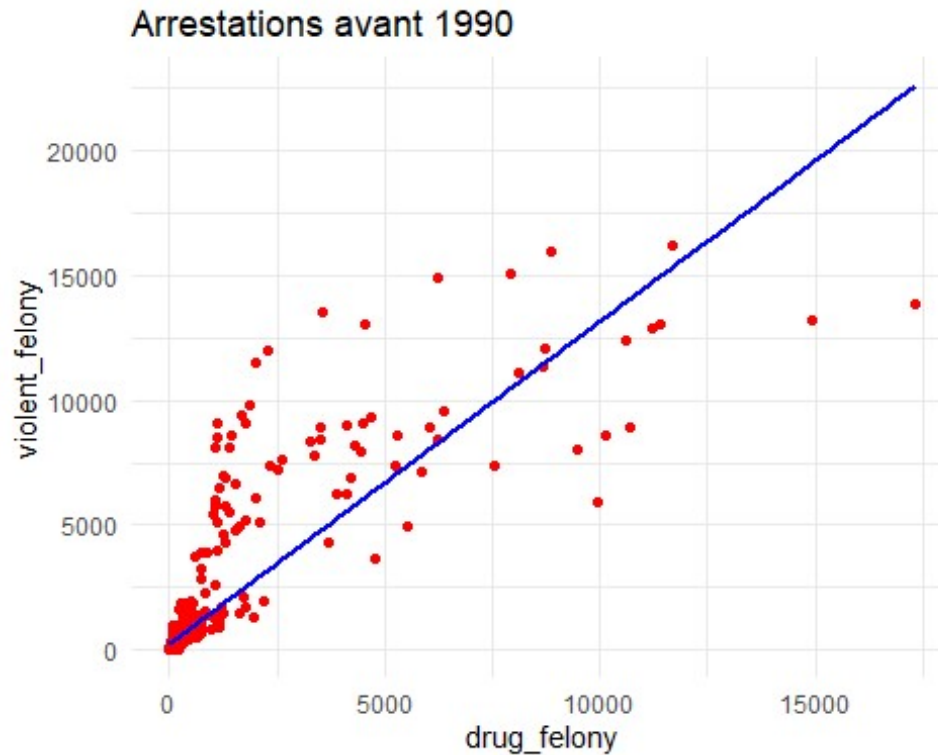
Ligne de tendance

(La ligne bleue) représente la ligne de régression linéaire, qui est le meilleur ajustement linéaire pour les points de données. Elle indique la tendance générale des données : dans ce cas, une tendance positive indique que, généralement, à mesure que le nombre de délits liés à la drogue augmente, le nombre de délits violents tend également à augmenter. Ainsi, on décide de faire deux cas d'études différents sur les intervalles [1970,1990] et [1990,2019], et cela commence par refaire le graphe précédents avec ces nouvelles contraintes pour observer le nuage de points.

```
# Filtrer les données pour les arrestations avant 1990
donnees_avant_1990 <- donnees %>% filter(Year < 1990)

# Graphique des arrestations avant 1990
ggplot(donnees_avant_1990, aes(x = drug_felony, y = violent_felony)) +
  geom_point(color = "red") +
  geom_smooth(aes(x = drug_felony, y = violent_felony), method = lm, se =
FALSE, color = "blue") +
  scale_color_manual(values = c("red")) +
  theme_minimal() +
  ggtitle("Arrestations avant 1990")

## `geom_smooth()` using formula = 'y ~ x'
```



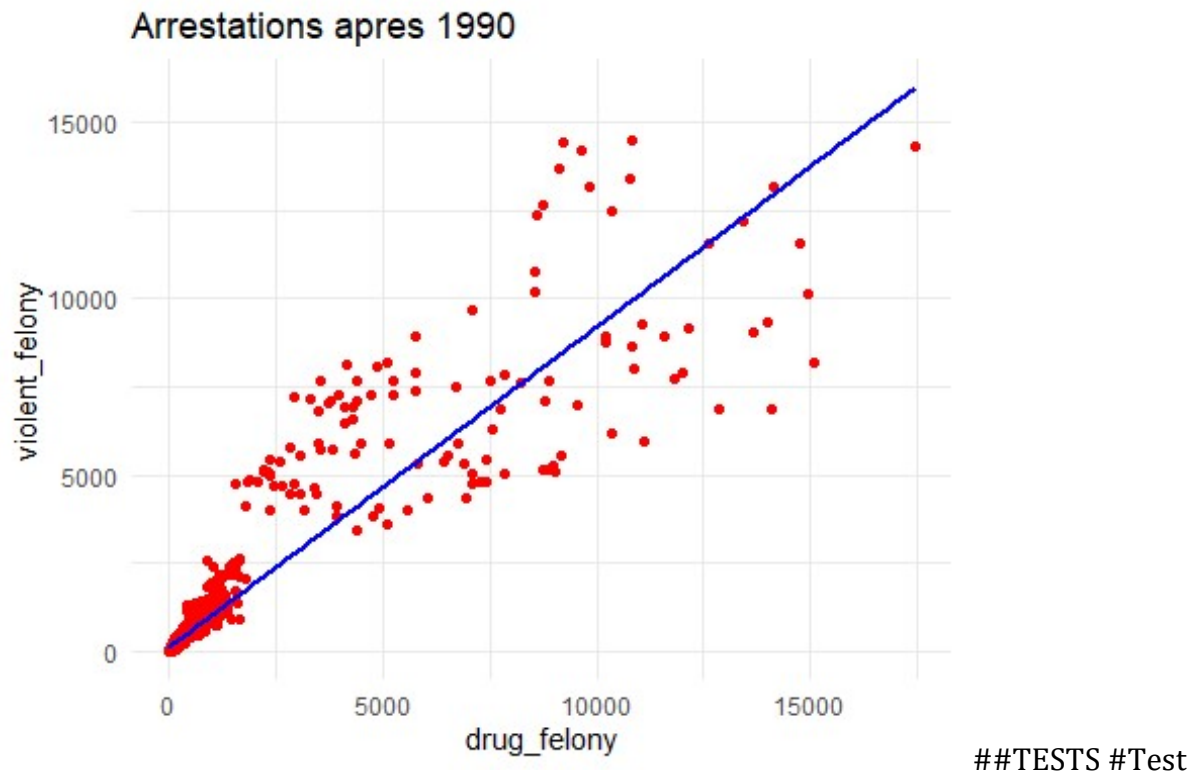
```
# Filtrer Les données pour Les arrestations apres 1990
```

```
donnees_apres_1990 <- donnees %>% filter(Year >= 1990)
```

```
# Graphique des arrestations apres 1990
```

```
ggplot(donnees_apres_1990, aes(x = drug_felony, y = violent_felony)) +  
  geom_point(color = "red") +  
  geom_smooth(aes(x = drug_felony, y = violent_felony), method = lm, se =  
FALSE, color = "blue") +  
  scale_color_manual(values = c("red")) +  
  theme_minimal() +  
  ggtitle("Arrestations apres 1990")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

correlation pour [1970,1990]

```
cor.test(donnees_avant_1990$drug_felony, donnees_avant_1990$violent_felony)

##
## Pearson's product-moment correlation
##
## data:  donnees_avant_1990$drug_felony and
##        donnees_avant_1990$violent_felony
## t = 58.054, df = 1278, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8357006 0.8658957
## sample estimates:
##      cor
## 0.8515025
```

#Test correlation pour [1990,2019]

```
cor.test(donnees_apres_1990$drug_felony, donnees_apres_1990$violent_felony)

##
## Pearson's product-moment correlation
##
## data:  donnees_apres_1990$drug_felony and
##        donnees_apres_1990$violent_felony
## t = 102.84, df = 1899, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
```

```
## 95 percent confidence interval:
##  0.9136155 0.9273253
## sample estimates:
##          cor
## 0.9207542
```

Le coefficient de corrélation de Pearson est de 0.8515025 pour les données avant 1990, et celui d'après 1990 est 0.92. Ces valeurs sont très proches de 1, ce qui indique une forte corrélation positive entre les deux variables. La valeur p est inférieure à $2.2e-16$, ce qui est largement en dessous du seuil habituel de 0.05, cela signifie qu'on peut rejeter l'hypothèse nulle qu'il n'y a pas de corrélation entre les deux variables. En plus, le coefficient de corrélation a augmenté entre ces deux intervalles, ce qui montre l'évolution de leur relation. Ces données montrent une corrélation positive et statistiquement significative entre le nombre d'arrestations pour criminalité liée à la drogue et le nombre d'arrestations pour criminalité liée à la violence. Cela suggère que ces deux types de criminalité sont liés d'une certaine manière.

On va passer à la causalité de ces deux variables. Ce n'est pas quelque chose de tranché mais on peut prouver la dépendance de la variable `violent_felony` à `drug_felony` pour ces deux intervalles.

#Test de regression linéaire avant 1990

```
lm(donnees_avant_1990$violent_felony ~ donnees_avant_1990$drug_felony)

##
## Call:
## lm(formula = donnees_avant_1990$violent_felony ~
## donnees_avant_1990$drug_felony)
##
## Coefficients:
##                (Intercept)  donnees_avant_1990$drug_felony
##                215.997                                1.294
```

Le test de régression linéaire montre que le modèle est statistiquement significatif ($p\text{-value} < 2.2e-16$), ce qui confirme qu'il y a une relation linéaire entre les délits liés aux drogues et les crimes violents avant 1990.

#Test de regression linéaire après 1990

```
lm(donnees_apres_1990$violent_felony ~ donnees_apres_1990$drug_felony)

##
## Call:
## lm(formula = donnees_apres_1990$violent_felony ~
## donnees_apres_1990$drug_felony)
##
## Coefficients:
##                (Intercept)  donnees_apres_1990$drug_felony
##                142.307                                0.908
```

Le test de régression linéaire montre également une forte significativité statistique ($p\text{-value} < 2.2e-16$), confirmant ainsi la présence d'une relation linéaire entre les délits liés aux drogues et les crimes violents après 1990. Cependant le coefficient du taux de regression linéaire a diminué au fil de temps, on peut expliquer cette diminution par plusieurs facteur possibles: il se peut qu'après 1990 des facteurs de confusion supplémentaires soient entrés en jeu, affectant la relation entre drug_felony et violent_felony. Ces facteurs de confusion pourraient inclure des changements dans les politiques de sécurité, l'accès aux soins de santé mentale, ou d'autres facteurs socio-économiques qui n'ont pas été pris en compte dans le modèle de régression linéaire.

##CONCLUSION Les résultats des tests de corrélation et des modèles de régression linéaire indiquent une relation positive significative entre les drug felony et violent felony, tant avant qu'après 1990 dans l'État de New York. Cependant, il est important de noter que la corrélation entre ces deux types de crimes semble être devenue encore plus forte après 1990.

Cette corrélation ne doit pas être interprétée comme une preuve de causalité directe. Cependant, elle suggère fortement qu'il existe un lien entre drug_felony et violent_felony, qui peut être influencé par divers facteurs socio-économiques, politiques et culturels. Par exemple, la disponibilité accrue de drogues illicites peut conduire à une augmentation des activités criminelles, y compris les crimes violents, tandis que les politiques de lutte contre la drogue et les programmes de prévention peuvent avoir un impact sur ces tendances. En conclusion, nous ne pouvons pas réfuter l'hypothèse posé au début de cette étude, mais on ne pourra pas l'accepter non plus même si les résultats obtenus soulignent l'importance de politiques publiques qui abordent à la fois les problèmes de drogue et de criminalité violente, tout en reconnaissant les dynamiques complexes qui sous-tendent leur relation.