

Gun violence data

Auteurs :

Marc Kratz

Maxime Heron

Maxime Moineau

Romain Cavallo

Descriptif du sujet et de ses finalités

Sujet :

Il s'agira de mener une série de statistiques sur l'archive du même intitulée qui se trouve ici. Cet archive (au format CSV) contient des entrée relative aux faits divers causées par des armes à feu qui ont eu lieu aux Etats Unis.

Entre le 1er janvier 2013 et le 31 mars 2018. Il s'agira de montrer l'évolution d'un certain nombre de variables pendant la période concernée, d'en calculer la moyenne et l'écart type (quand possible). Les variable qui nous intéressent sont :

- nombre de blessés;*
- nombre de morts;*
- nombre de malfaiteurs;*
- âge des malfaiteurs;*
- état dans lequel le fait divers a eu lieu. ´*

Il s'agira enfin de voir si vous trouvez des corrélations entre ces variables et s'il y aussi des corrélations avec la période de l'année.

Le sujet demandait la récupération de données de faits divers concernant les armes à feu, une observation statistique des différentes données à disposition et l'établissement de corrélations entre ces données.

Hypothèse :

Un descriptif des données utilisées

Le fichier CSV ressource contient pour chaque fait divers :

"incident_id" : un identifiant, non utilisé,
"date" : la date, au format YYYY-MM-DD, parfois absente ou incohérente, utilisée,
"state" : l'Etat où a eu lieu le fait divers, utilisé,
"city_or_county" : la ville a eu lieu le fait divers, non utilisé,
"address" : une adresse, non utilisée,
"n_killed" : le nombre de victimes décédées, utilisé,
"n_injured" : le nombre de blessés, utilisé,
"incident_url" : une url, parfois absente ou incohérente, non utilisée,
"source_url": parfois absente ou incohérente, non utilisée,
"incident_url_fields_missing" : non utilisé,
"congressional_district" : non utilisé,
"gun_stolen" : non utilisé,
"gun_type" : non utilisé
"incident_characteristics" : non utilisé,

"latitude" : non utilisée,
"location_description":parfois abstente ou incohérente, non utilisée
"longitude" : non utilisé
"n_guns_involved": parfois abstent ou incohérent, non utilisé
"notes": non utilisé,
"participant_age": parfois abstent ou incohérent, non utilisé
"participant_age_group": parfois abstent ou incohérent, non utilisé
"participant_gender":parfois abstent ou incohérent, non utilisé
"participant_name":non utilisé
"participant_relationship":non utilisé
"participant_status": non utilisé
"participant_type":non utilisé
"sources": non utilisé
"state_house_district": non utilisé
"state_senate_district": non utilisé

La méthodologie suivie pour répondre aux questions

Nous avons tenté d'une part de recueillir les données, puis de les mettre en opposition afin de pouvoir les représenter graphiquement.

Le but était par la suite de pouvoir représenter les propriétés statistiques de ces objets afin de les comparer.

L'objectif était de représenter les variations de ces données sous la forme de lois de probabilité dont nous aurions déterminé les paramètres.

Les résultats obtenus et leur visualisation graphique autant que possible

Nous avons réussi à récupérer certaines données, en revanche pas toutes celles que nous souhaitions, notamment celles de l'âge et de la période de l'année.

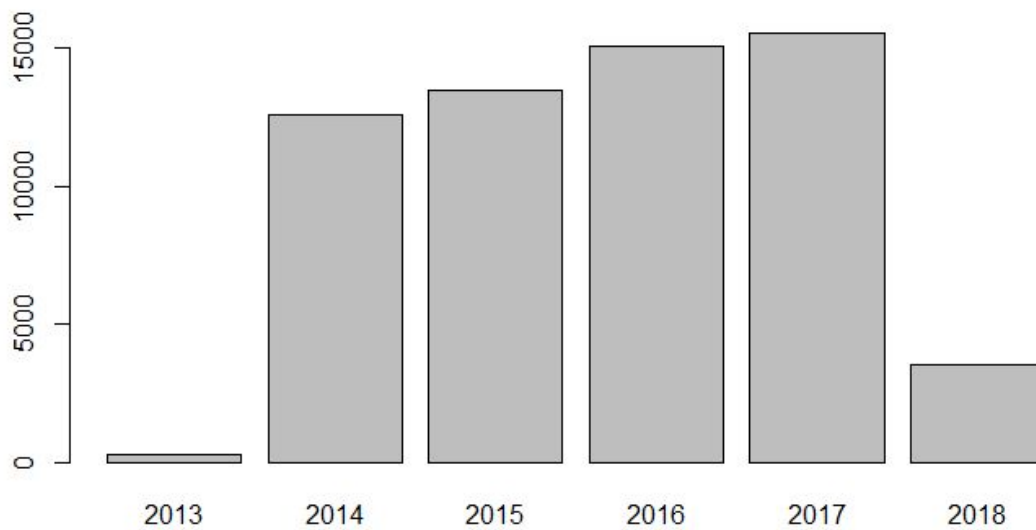
La difficulté d'abord du dataset à travers le langage R a été une barrière pour nous qui nous a retenu de progresser vers des conclusions poussées sur ces données.

Une courte présentation des membres du groupe et de leur rôle dans le projet.

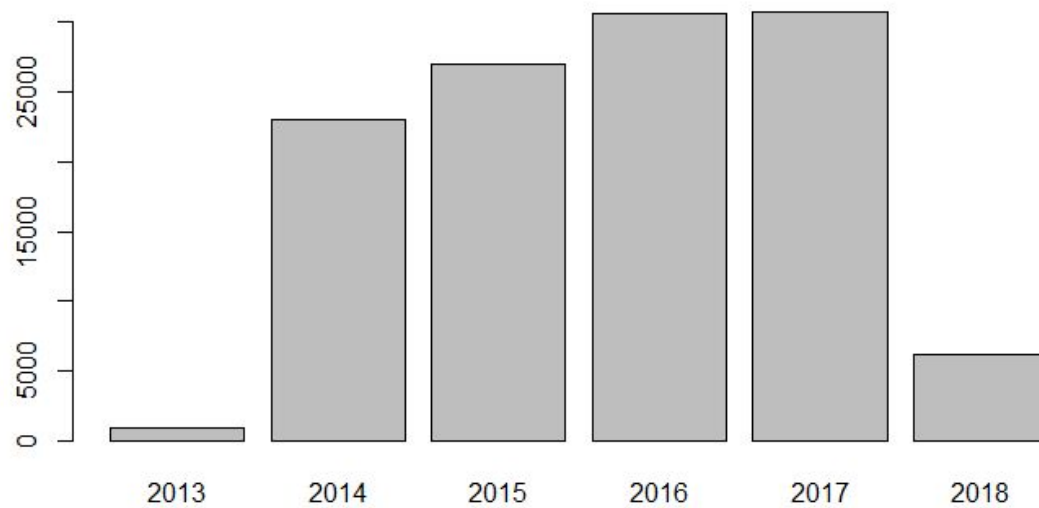
Maxime M : J'ai tenté de réunir les données nécessaire à notre étude, afin que nous puissions appliquer des traitements statistiques sur celles-ci pour espérer dégager des corrélations et en comparer la pertinence. Il a été difficile malgré tout de passer la barrière des types et des contraintes du langage pour parvenir une collection de données exploitable et complète pour répondre au sujet.

Romain

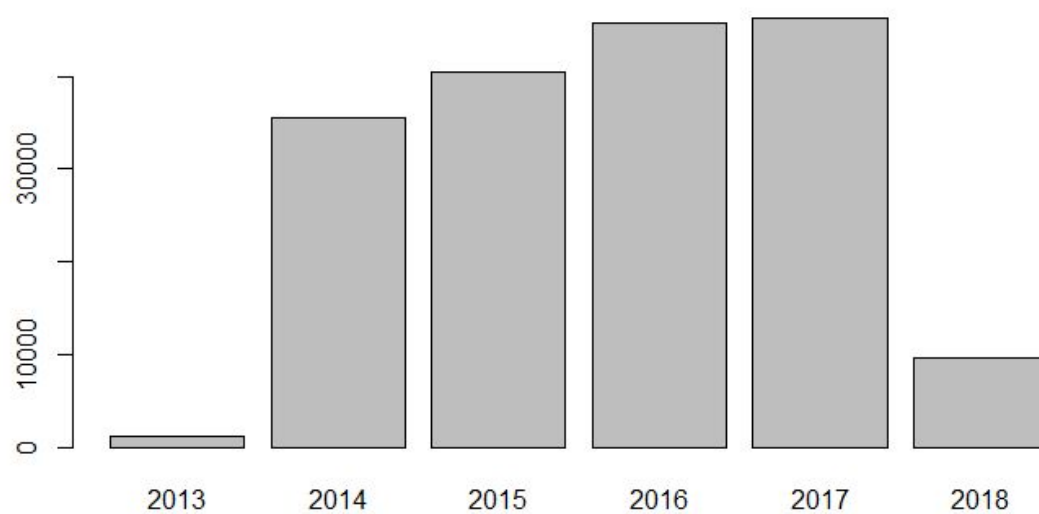
#Mort par années

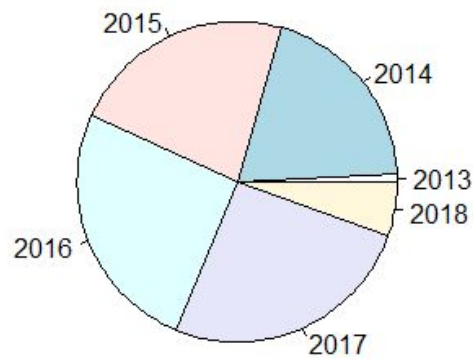


#Blésés par années



#NB Tot blésés mort + blésés par années



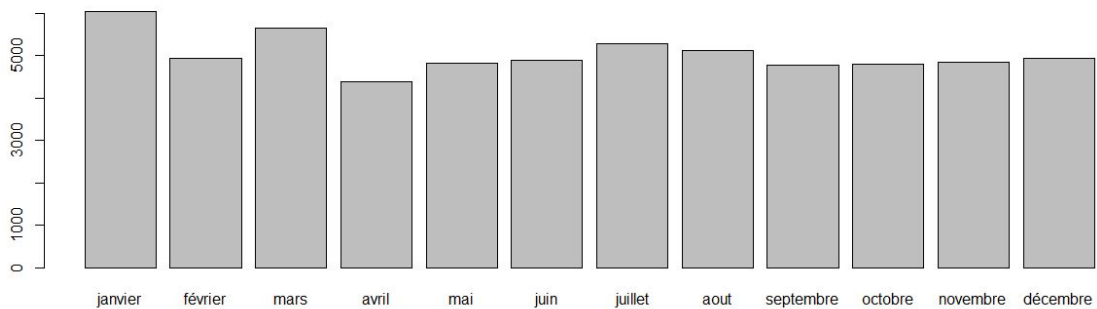


2013	2014	2015	2016	2017	2018
1296	35559	40451	45646	46214	9704

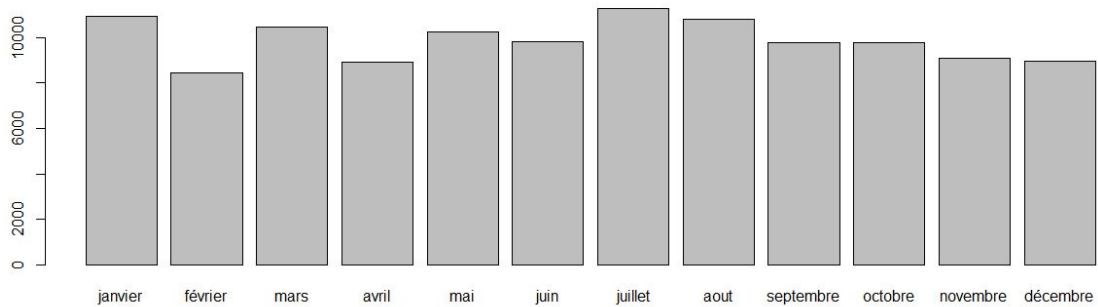
Les chiffres correspondent entre le nombre de morts et de blessés

On y voit toujours une enorme augmentation de 2013 à 2014 puis en continuelle expansion jusqu'a 2017 ou après les chiffres chutent énormément

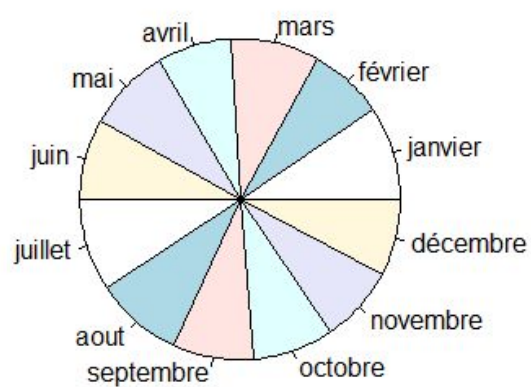
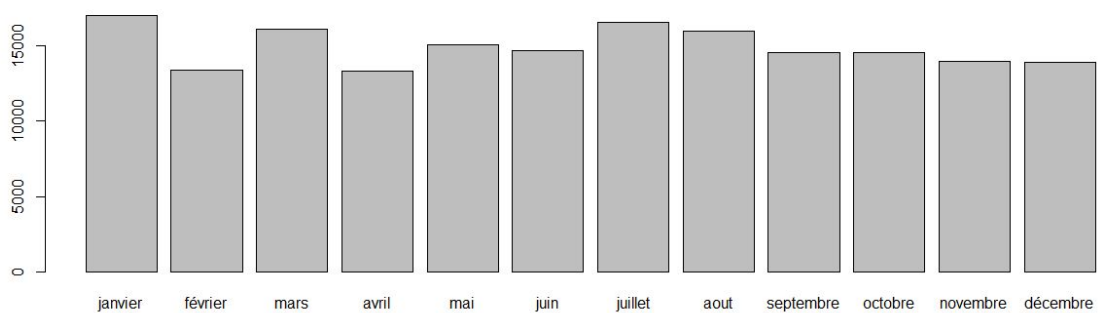
#NB mort par mois



#NB blésés par mois



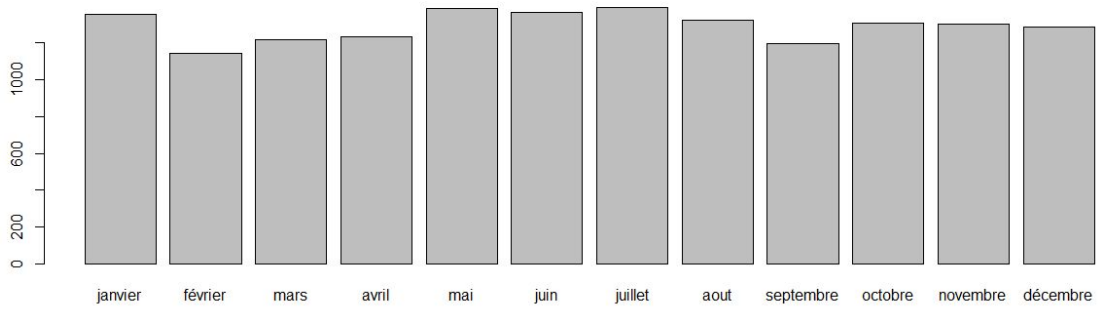
#NB Tot blésés mort + blésés par mois



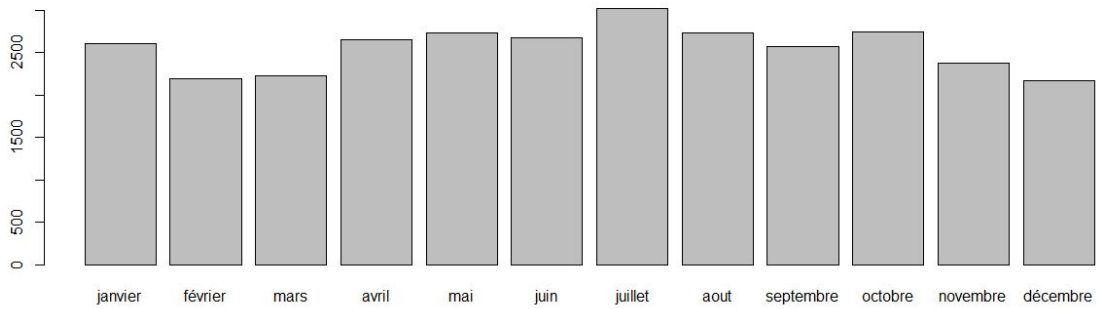
On constate que le nombre de mort

#Analyse année avec le plus de crime 2017

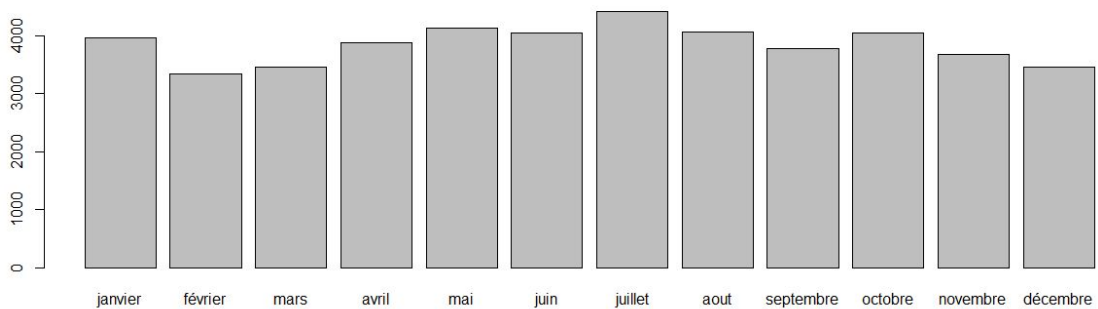
#Nb de mort par mois 2017



#Nb blésé par mois 2017

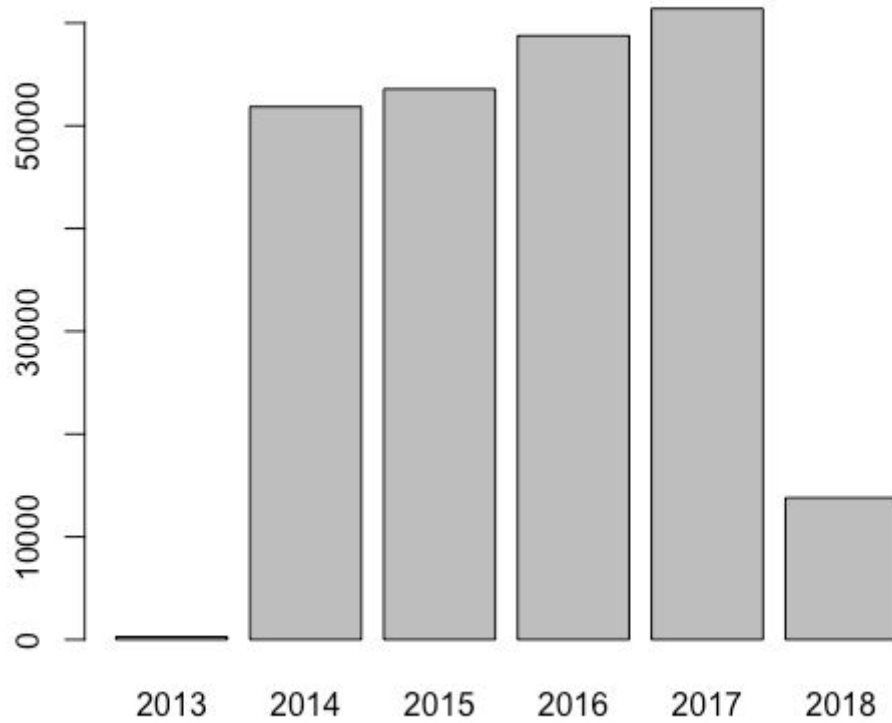


#Nb blésé total : mort + blésés 2017

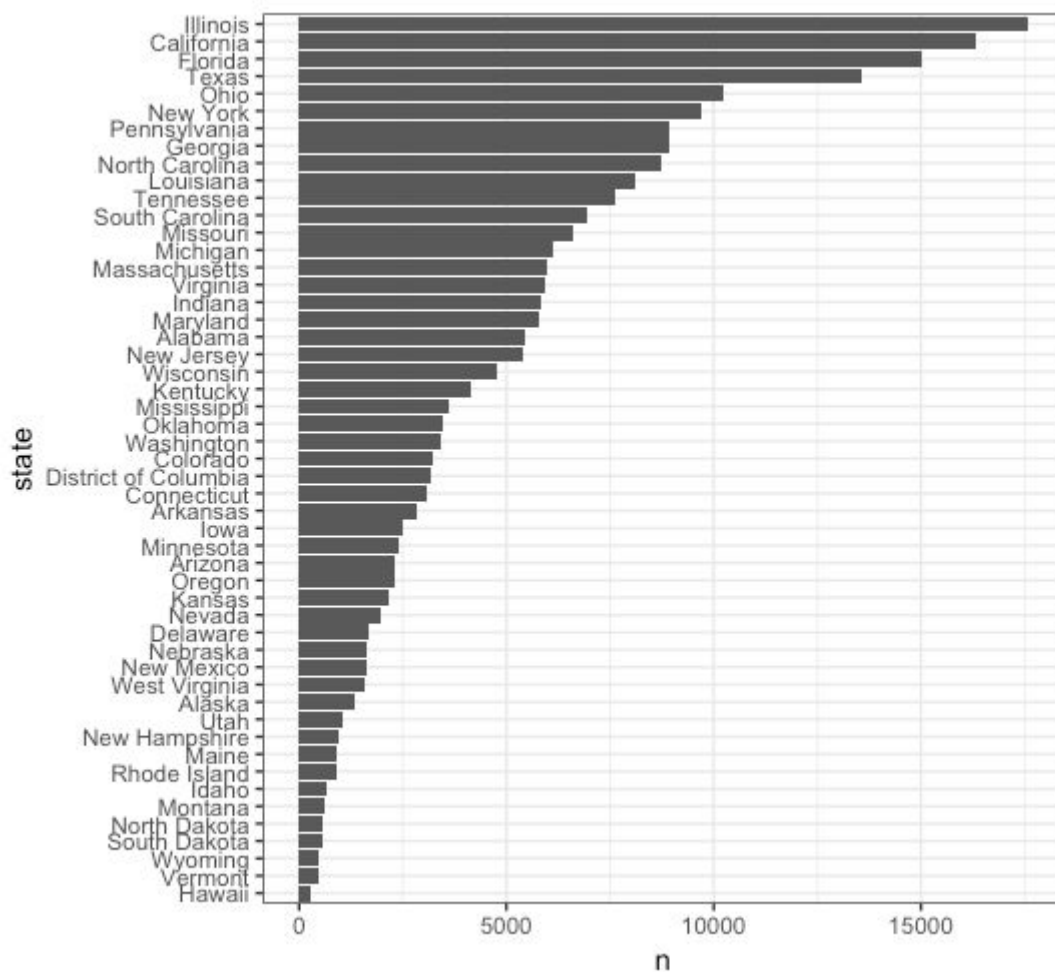


#Pays avec le plus d'armes impliquées

Marc

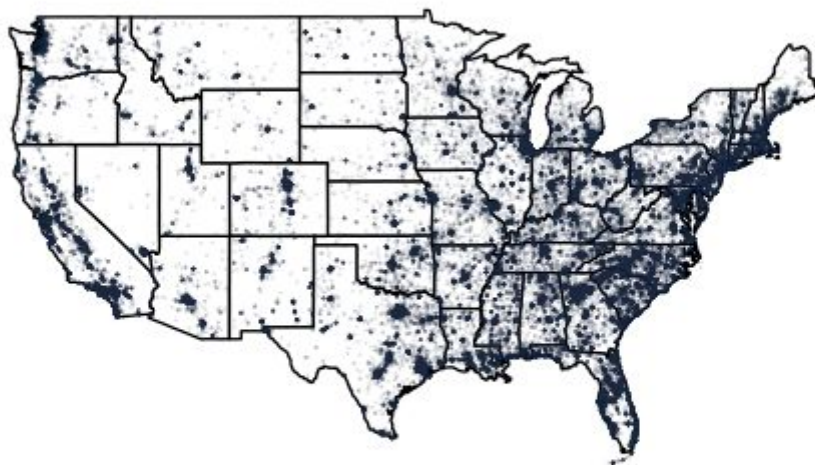


Nombre d'incident sur chaque année
Ici, on note



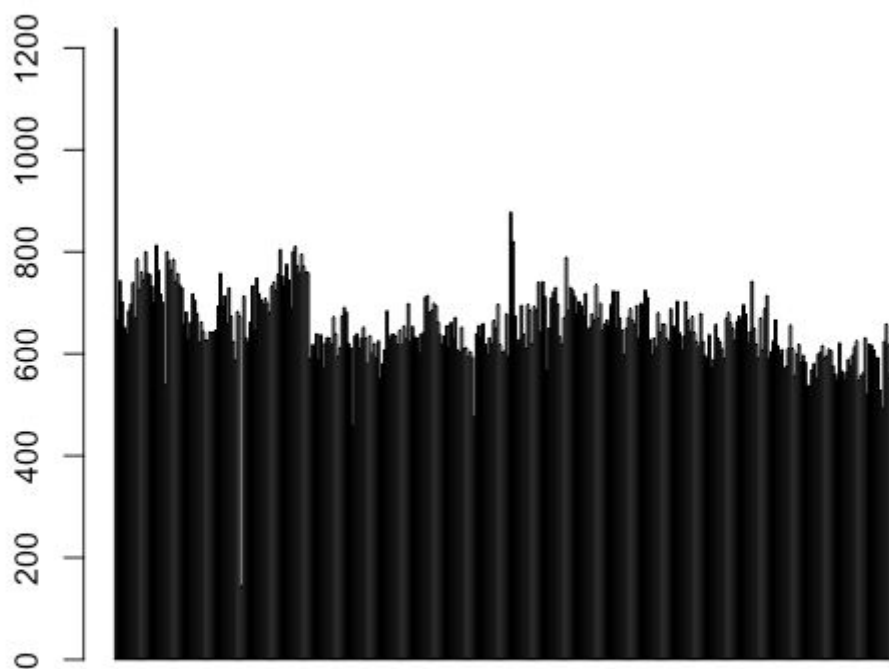
Nombre d'incident par état sur les 5 ans Ici on observe que Illinois est en tête

On peut voir que ces chiffres coïncide avec les etat les plus habitées des etats unis



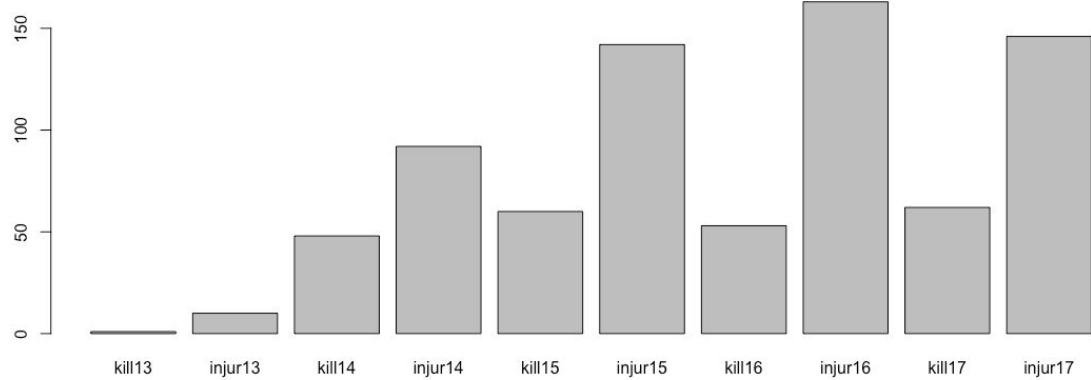
Répartition des incidents sur le territoire

Top dates sur toutes les années



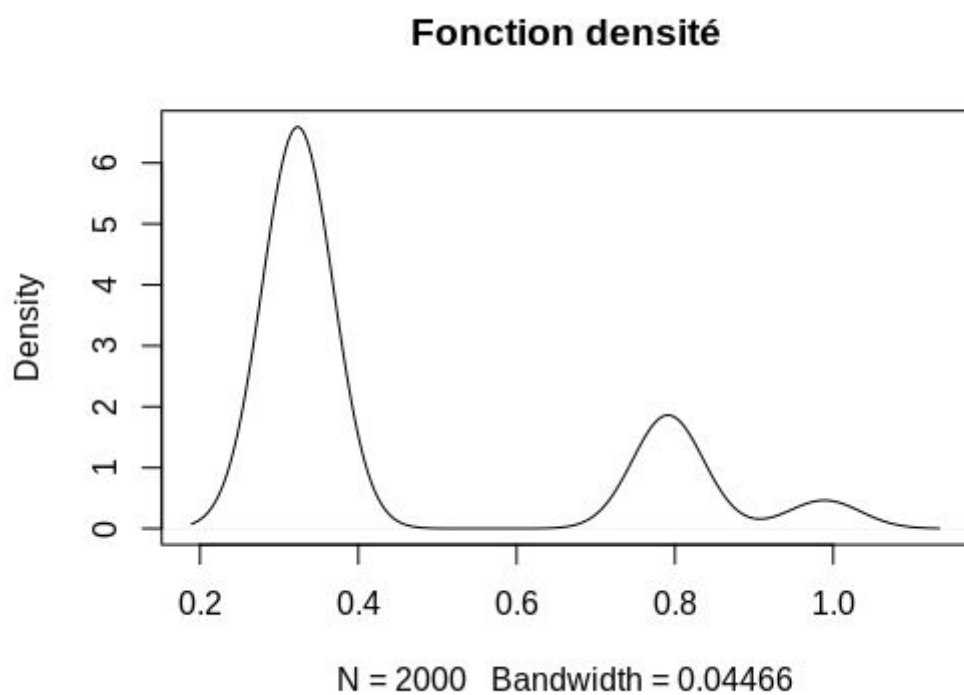
Date	01/01	07/04	07/05	01/20	03/25	03/18	01/15	01/25	03/24	03/28
Nombre d'incidents	1238	877	820	813	810	804	800	800	800	795

Répartition des incidents sur le long d'une "année classique"



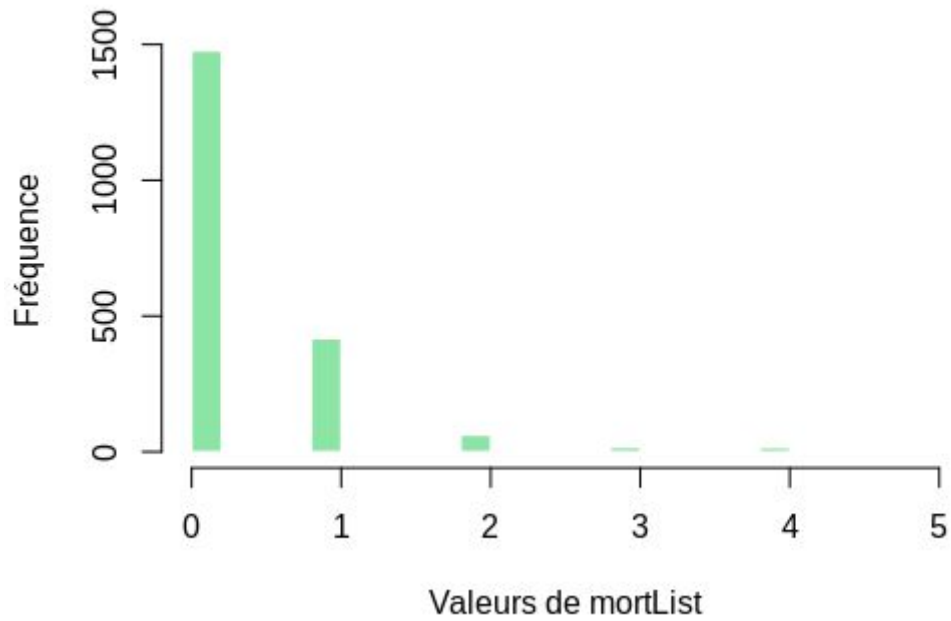
répartition des mort blessé le 4 juillet de chaque année

Maxime



Densité des morts

Distribution de la variable mortList



Distribution de la variable blessesList

