

Analyse exploratoire

Les activités terroristes mondiales depuis 1970

5 SDBD A2 - CAZAUBON Thomas, FEUILLET Laure, PLANTEC Maël

Filtrage des données

Avant toute chose, on commence par supprimer les données du dataset qui ne nous intéressent pas, on crée ensuite un nouveau dataset, bien plus léger, qui nous servira de base de travail pour la suite.

```
#####
### PREMIER ECREMAGE DES DONNEES #####
#####

#Lecture du dataset original
gterror = read.csv("terrorism.csv", sep=',', header=TRUE)

#Filtrage pour ne garder que les colonnes pertinentes
filtered <- subset(gterror, select=c(iyear, imonth, extended, country_txt, region_txt, provstate, city, region_locale))

#Renommage des colonnes
names(filtered) <- c("année", "mois", "étendu", "pays", "région", "région_locale", "ville", "latitude", "longitude")

#Sauvegarde sur csv
write.csv(filtered, "./filtered_terrorism.csv")
```

Analyse des données

C'est parti !

```
#####
##### COMMENCER EN DOUCEUR #####
#####

gterror = read.csv("filtered_terrorism.csv", sep=',', header=TRUE)
require("ggplot2")

## Loading required package: ggplot2

require("plyr")

## Loading required package: plyr

require("tidyverse")
```

Loading required package: tidyverse

Se familiariser avec les données :

```

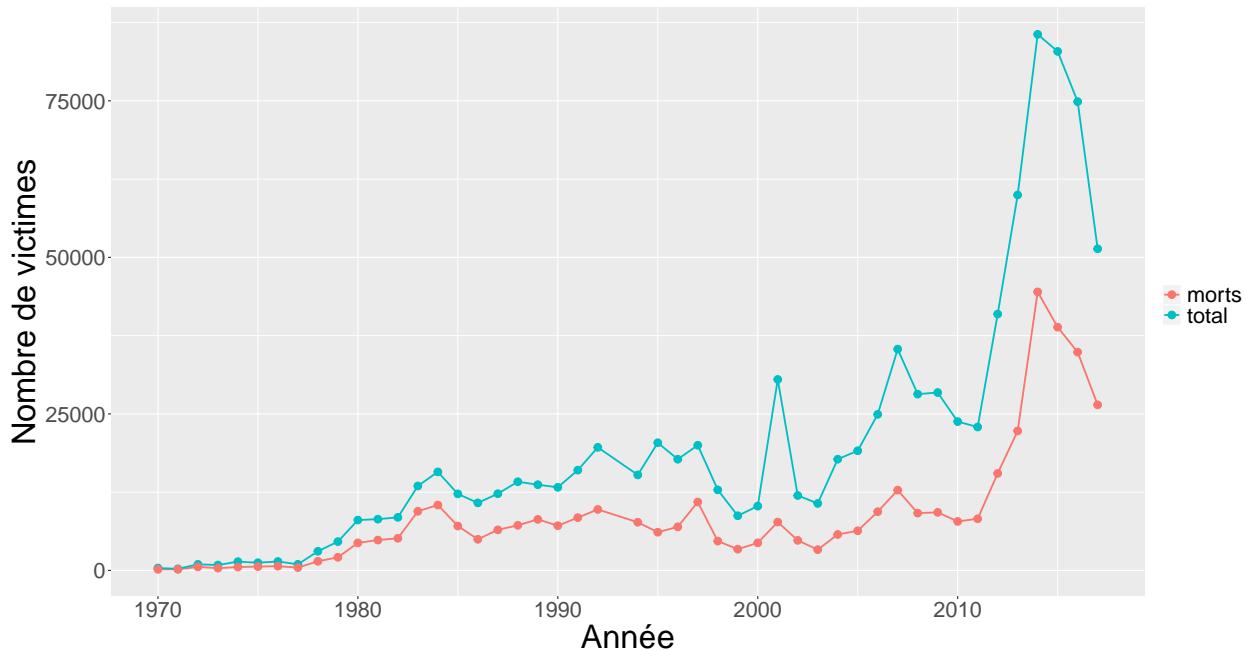
#Nombre de victimes (décès) du terrorisme depuis 1970
nbkills = sum(na.omit(gterror$nombre_morts))
nbkills

## [1] 411868

#Nombre de victimes du terrorisme par année : 2014 est l'année la plus meurtrière avec quasiment 45000
killsPerYear = ddply(gterror, .(année), summarize,
                      total=sum(na.omit(nombre_blessés))+sum(na.omit(nombre_morts)),
                      morts=sum(na.omit(nombre_morts)))
killsPerYear = gather(data=killsPerYear, key=type, value=victimes, total, morts)
plot = ggplot(killsPerYear, aes(x=année, y=victimes, color=type)) +
  geom_line(size=0.8) + geom_point(size=3) +
  labs(title="Evolution du nombre de victimes du terrorisme dans le monde", x ="Année", y="Nombre de victimes",
       theme(title=element_text(size=30),
             axis.text.x=element_text(size=20), axis.text.y=element_text(size=20),
             legend.title=element_blank(), legend.text=element_text(size=20)))
ggsave(plot, filename="graphiques/killsPerYear.pdf", width=16, height=9)
plot

```

Evolution du nombre de victimes du terrorisme dans le monde

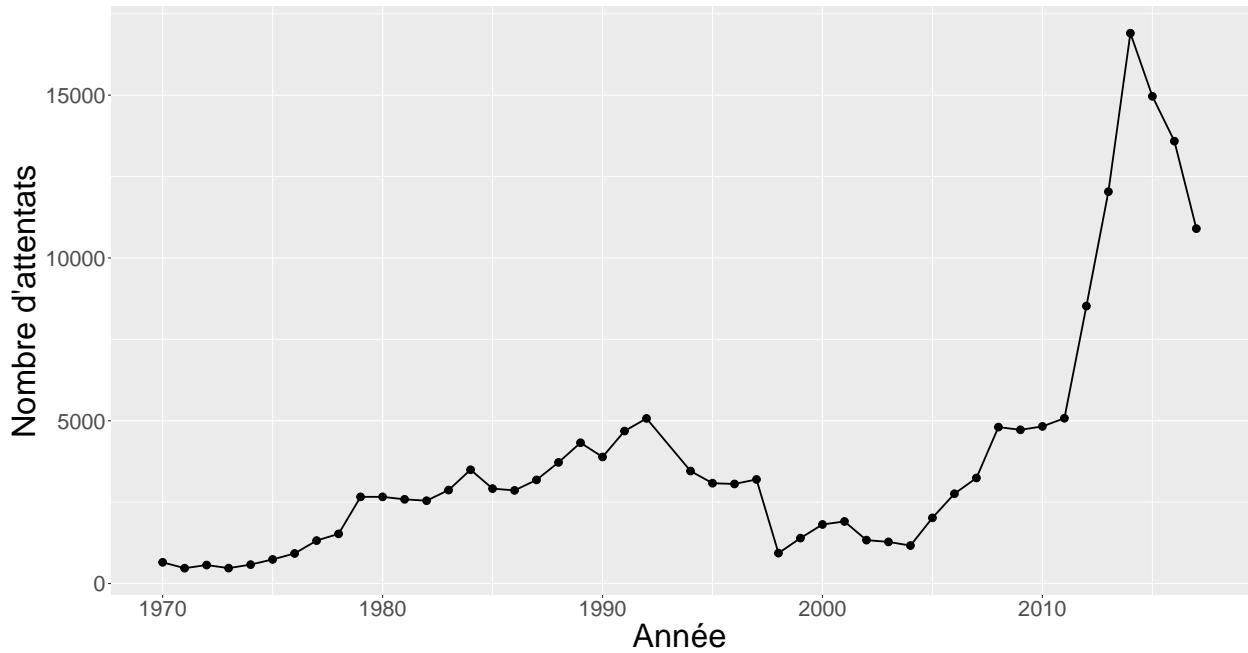


```

#Nombre d'attentats par année
attacksPerYear = ddply(gterror, .(année), summarize,
                        attentats=length(pays))
plot = ggplot(attacksPerYear, aes(x=année, y=attentats)) +
  geom_line(size=0.8) + geom_point(size=3) +
  labs(title="Evolution du nombre d'attentats dans le monde", x ="Année", y="Nombre d'attentats") +
  theme(title=element_text(size=30),
        axis.text.x=element_text(size=20), axis.text.y=element_text(size=20))
ggsave(plot, filename="graphiques/attacksPerYear.pdf", width=16, height=9)
plot

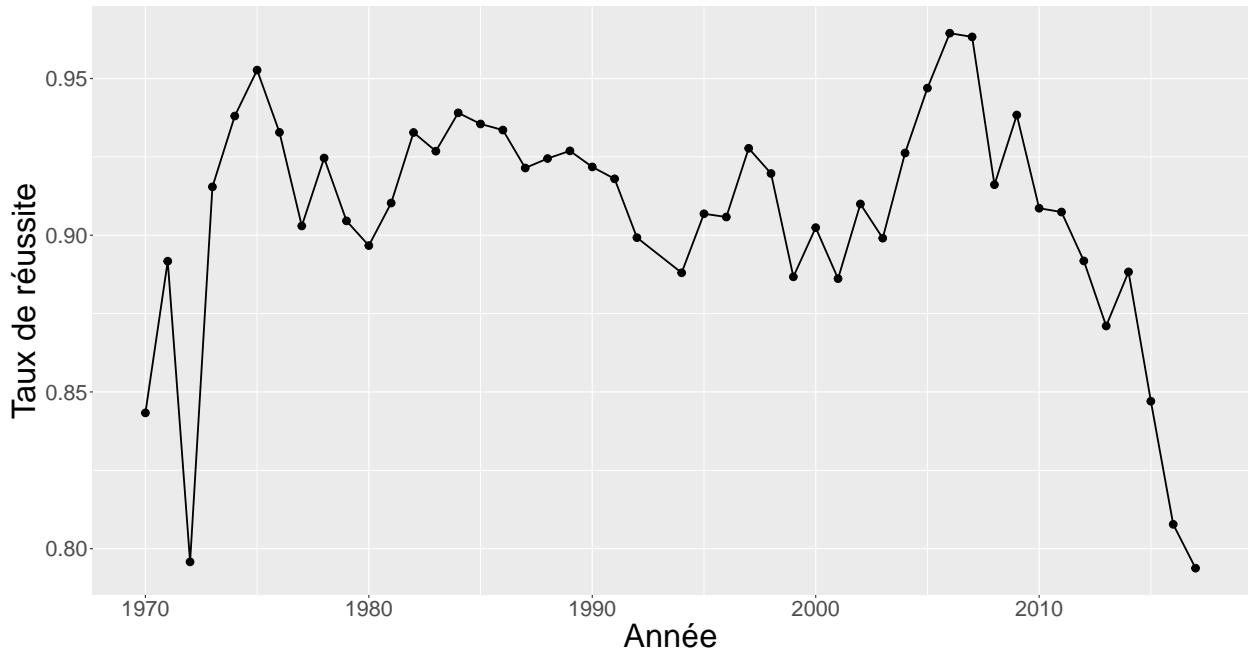
```

Evolution du nombre d'attentats dans le monde



```
#Evolution du nombre de succès
successesPerYear = ddply(gterror, .(année), summarize, ratio=(sum(na.omit(succès))/length(année)))
plot = ggplot(successesPerYear, aes(x=année, y=ratio)) +
  geom_line(size=0.8) + geom_point(size=3) +
  labs(title="Evolution du taux de réussite des attentats", x="Année", y="Taux de réussite") +
  theme(title=element_text(size=30), axis.text.x=element_text(size=20), axis.text.y=element_text(size=20))
ggsave(plot, filename="graphiques/successRatioPerYear.pdf", width=16, height=9)
plot
```

Evolution du taux de réussite des attentats



```

# Je garde uniquement certaines colonnes qui m'interessent.
tw = subset(gterror, select=c(année, pays, type_attaque, responsable, type_arre, nombre_morts, nombre_ba))

'%ni%' <- Negate('%in%')

# Nombre de morts par type d'arme utilisé.
morts_par_type_arre = ddply(tw, .(type_arre), summarize, nombre_morts=sum(na.omit(nombre_morts)))

# Evolution de ce top 5 et last 5 au fil du temps
morts_par_type_arre_par_année = ddply(tw, .(type_arre, année), summarize,
                                       nombre_morts=sum(na.omit(nombre_morts)))
top5_armes = arrange(morts_par_type_arre, -morts_par_type_arre$nombre_morts)[1:5,]

top5_morts_par_type_arre_par_année = morts_par_type_arre_par_année[morts_par_type_arre_par_année$type_arre %in% top5_armes,]

bottom5_armes = arrange(morts_par_type_arre, -morts_par_type_arre$nombre_morts)[7:12,]
bottom5_morts_par_type_arre_par_année = morts_par_type_arre_par_année[morts_par_type_arre_par_année$type_arre %in% bottom5_armes,]

# Plots
gtop5 = ggplot(top5_morts_par_type_arre_par_année, aes(x=année, y=nombre_morts, color=type_arre)) +
  geom_line(aes(group=type_arre, linetype=type_arre), size= 1) +
  labs(title="Top 5 des armes : Evolution des morts par année", x="Année", y="Nombre de morts", color="Type d'arme") +
  theme(title=element_text(size=30), axis.text.x=element_text(size=20), axis.text.y=element_text(size=20),
        scale_colour_discrete("") +
        scale_linetype_manual("", values=c(1,2,3,4,5,6)) +
        scale_shape_manual("", values=c(17,18, 19,20,21,22)))

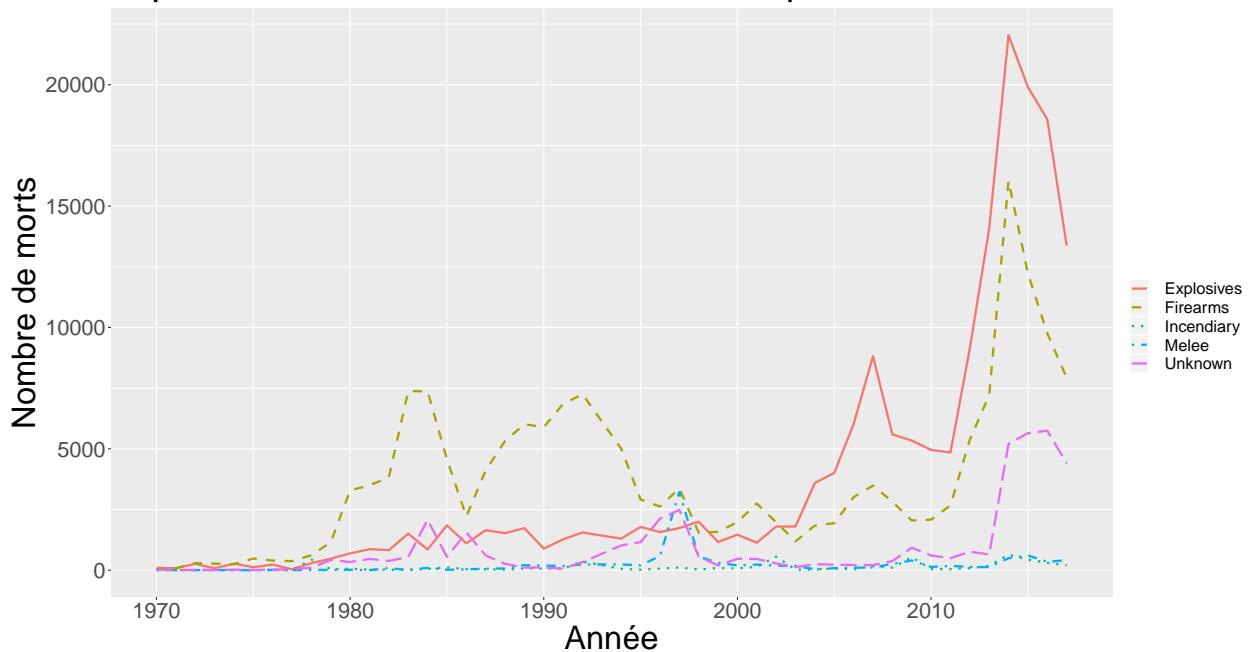
gbottom5 = ggplot(bottom5_morts_par_type_arre_par_année, aes(x=année, y=nombre_morts, color=type_arre)) +
  geom_line(aes(group=type_arre, linetype=type_arre), size= 1) +
  labs(title="Bottom 5 des armes : Evolution des morts par année", x="Année", y="Nombre de morts", color="Type d'arme") +
  theme(title=element_text(size=30), axis.text.x=element_text(size=20), axis.text.y=element_text(size=20),
        scale_colour_discrete("") +
        scale_linetype_manual("", values=c(1,2,3,4,5,6)) +
        scale_shape_manual("", values=c(17,18, 19,20,21,22)))

ggsave(gtop5, filename="graphiques/top5weaponsEvolutionOfKills.pdf", width=16, height=9)
ggsave(gbottom5, filename="graphiques/bottom5weaponsEvolutionOfKills.pdf", width=16, height=9)

gtop5

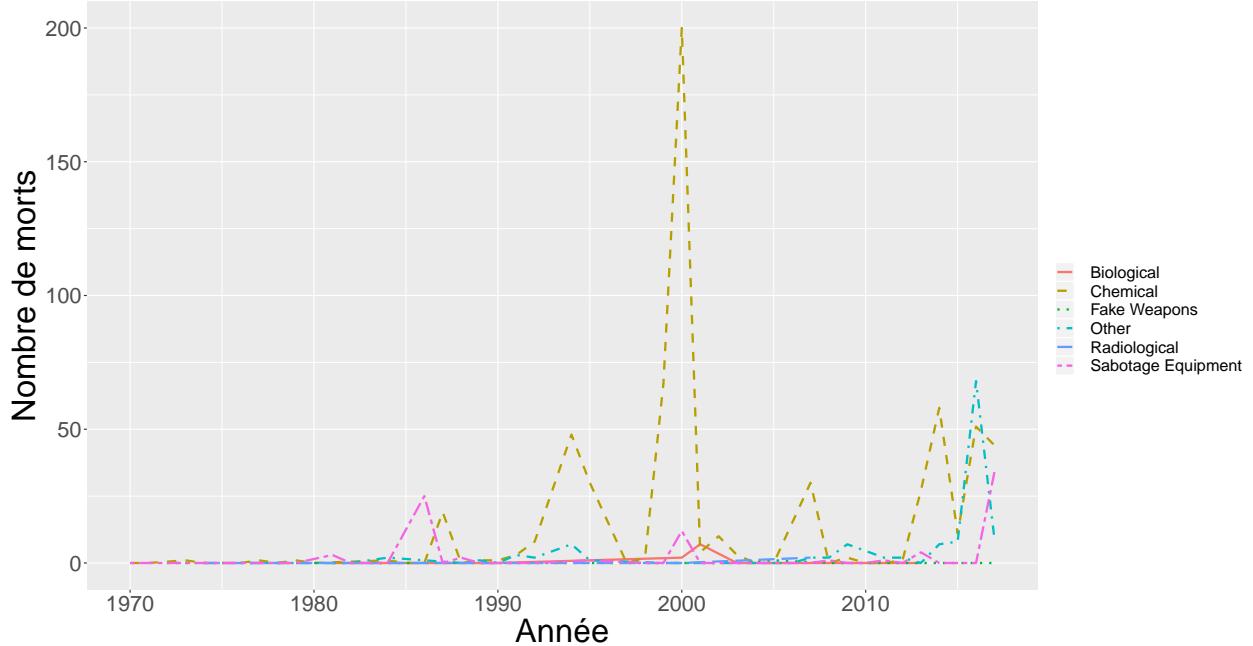
```

Top 5 des armes : Evolution des morts par année



gbottom5

Bottom 5 des armes : Evolution des morts par année



#Nationalité la plus visée par année

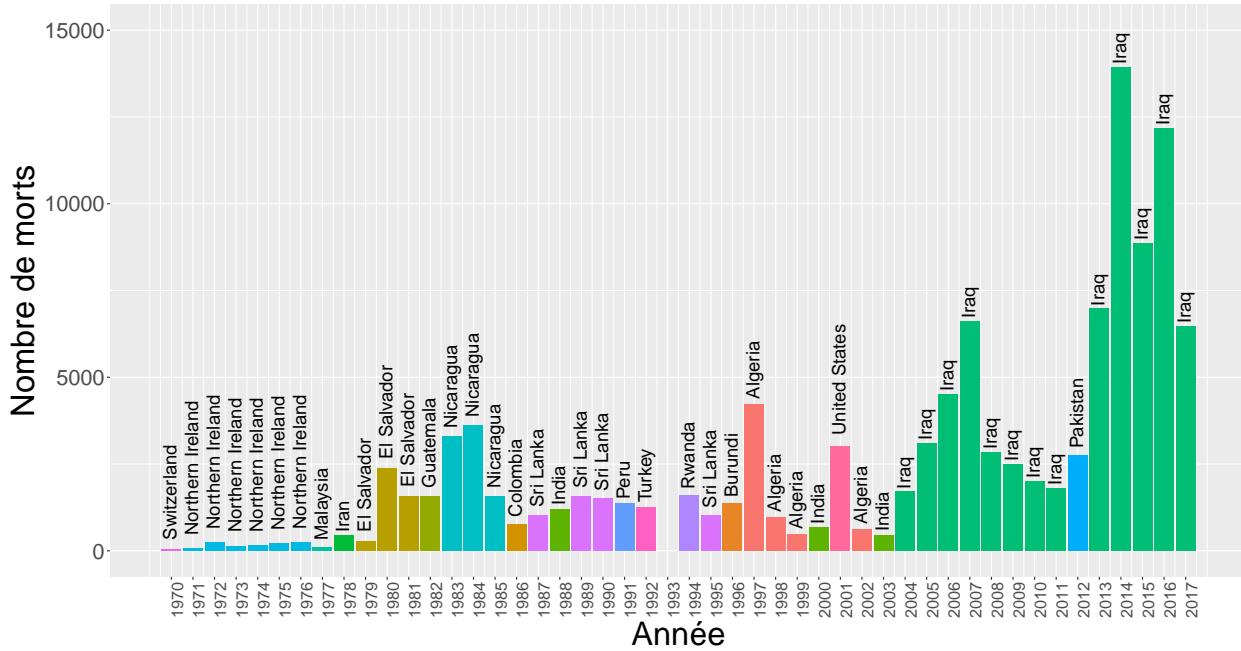
```
mostTargetedNationality = ddply(gterror, .(année, nationalité_ciblée), summarize, morts=sum(na.omit(nom))
mostTargetedNationality = mostTargetedNationality[order(mostTargetedNationality$année, -mostTargetedNationalité
mostTargetedNationality = ddply(mostTargetedNationality, .(année), summarize, nationalité=nationalité_ciblée)
plot = ggplot(mostTargetedNationality, aes(x=année,y=morts, fill = nationalité)) +
```

```

geom_bar(stat="identity") +
  labs(title="La nationalité la plus visée par le terrorisme selon l'année", x="Année", y="Nombre de morts"),
  geom_text(aes(label=nationalite), position=position_dodge(width=0.9), size=6, vjust=0.4, hjust=-0.1, color="black"),
  theme(title=element_text(size=30),
        axis.text.x=element_text(size=15, angle=90),
        axis.text.y=element_text(size=20),
        legend.position="none")+
  scale_x_continuous(breaks = c(1970:2017))+ylim(min=0,max=15000)
  ggsave(plot, filename="graphiques/mostTargetedNationalityPerYear.pdf", width=16, height=9)
  plot

```

La nationalité la plus visée par le terrorisme selon l'année

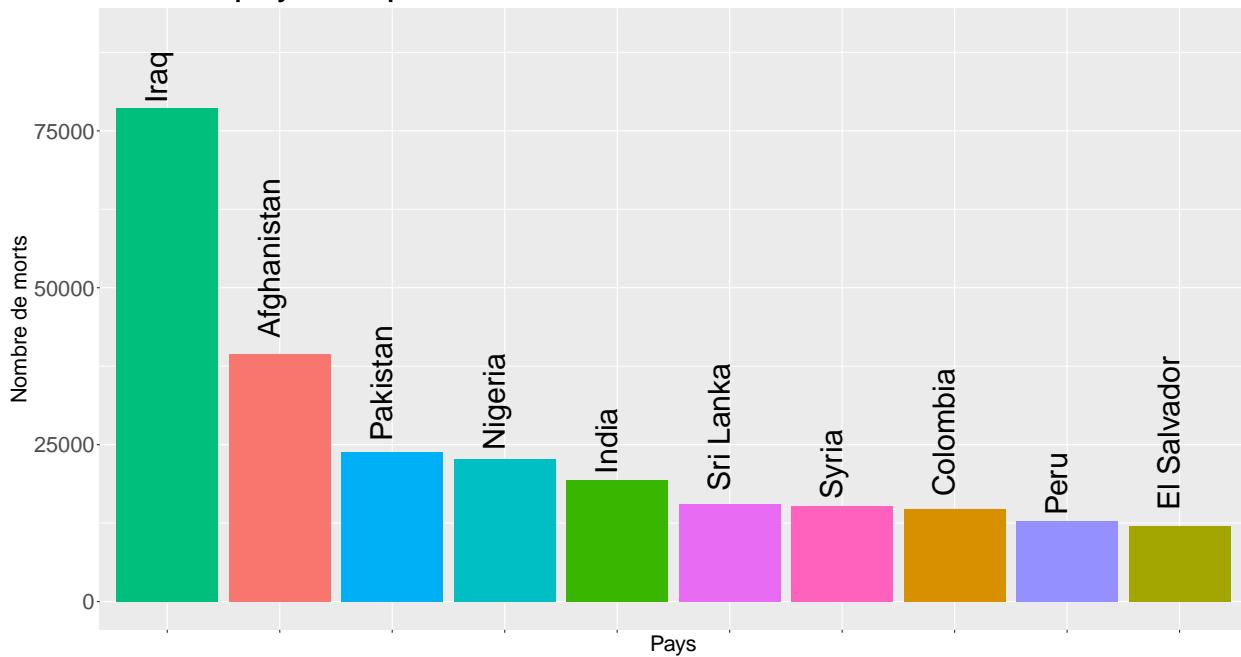


```

#Les 10 pays les plus touchés
killsPerCountry = dplyr::ggroup(gterror, .by=pays, summarize, morts=sum(na.omit(nombre_morts)))
killsPerCountry = killsPerCountry[order(-killsPerCountry$morts),][0:10,]
plot = ggplot(killsPerCountry, aes(x=reorder(pays,-morts),y=morts, fill=pays)) +
  geom_bar(stat="identity") +
  labs(title="Les dix pays les plus touchés", x="Pays", y="Nombre de morts") +
  geom_text(aes(label=pays), vjust=0, hjust=-0.1, color="black", size=10, angle=90) +
  theme(title=element_text(size=30),
        axis.title.x=element_text(size=20), axis.title.y=element_text(size=20),
        axis.text.x=element_blank(), axis.text.y=element_text(size=20),
        legend.position="none") +
  ylim(min=0, max=90000)
  ggsave(plot, filename="graphiques/tenTargetedCountries.pdf", width=16, height=9)
  plot

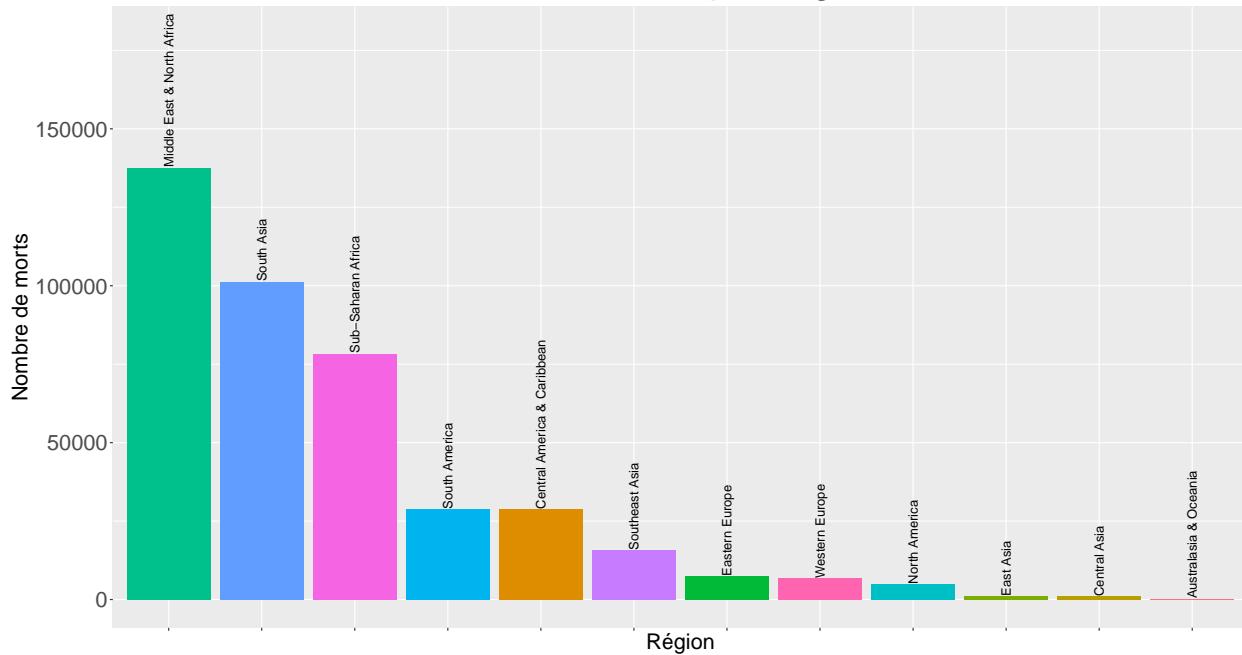
```

Les dix pays les plus touchés



```
#Nombre de victimes du terrorisme par région du monde
killsPerRegion = ddply(gterror, .(région), summarize, morts=sum(na.omit(nombre_morts)))
killsPerRegion = killsPerRegion[order(-killsPerRegion$morts),]
plot = ggplot(killsPerRegion, aes(x=reorder(région,-morts), y=morts, fill=région)) +
  geom_bar(stat="identity") +
  labs(title="Nombre de victimes du terrorisme par région", x="Région", y="Nombre de morts") +
  geom_text(aes(label=région), color="black", hjust=-0.005, size=4.3, angle=90) +
  theme(title=element_text(size=30),
        axis.title.x=element_text(size=20), axis.title.y=element_text(size=20),
        axis.text.x=element_blank(), axis.text.y=element_text(size=20),
        legend.position="none") +
  ylim(min=0, max=180000)
ggsave(plot, filename="graphiques/killsPerRegion.pdf", width=16, height=9)
plot
```

Nombre de victimes du terrorisme par région



```
# Carte des activités terroristes depuis 1970
data = subset(gterror, select=c(latitude, longitude))
data <- subset(data, data$longitude >= -10000)
#map = plot(data$longitude, data$latitude, main="Map ?", xlab="Longitude", ylab="Latitude", pch=20)
map = ggplot( data , aes( data$longitude , data$latitude )) +
  labs(title="Le terrorisme dans le monde - 1970 à nos jours", x="", y="") +
  geom_point(shape=19, colour="black", size=0.6, fill = NA, alpha=0.1, stroke=1) +
  theme(title=element_text(size=30),
        axis.text.x=element_blank(), axis.text.y=element_blank(),
        axis.ticks=element_blank(), axis.line=element_blank(),
        axis.title.x=element_blank(), axis.title.y=element_blank(),
        legend.position="none",
        panel.background=element_blank(), panel.border=element_blank(),
        panel.grid.major=element_blank(), panel.grid.minor=element_blank(),
        plot.background=element_blank())
ggsave(map, filename="graphiques/map.pdf", width=16, height=9)
map
```

Le terrorisme dans le monde – 1970 à nos jours



Graphe annexe (que nous n'avons pas analysé dans le rapport) :

```
#Evolution du nombre de victimes en France
gterrorfr = gterror[gterror$pays=="France",]
gterrorfr = ddply(gterrorfr, .(année), summarize, morts = sum(na.omit(nombre_morts)))
plot = ggplot(gterrorfr, aes(x=année,y=morts)) +
  geom_line(size=0.8) + geom_point(size=3) +
  labs(title="Evolution du nombre de victimes en France", x="Année", y="Nombre de morts") +
  theme(title=element_text(size=30),
        axis.text.x=element_text(size=20), axis.text.y=element_text(size=20))
ggsave(plot, filename = "graphiques/killsInFrancePerYear.pdf", width=16, height=9)
plot
```

Evolution du nombre de victimes en France

