Storm

```
setwd("C:/Users/John/Desktop/ReprdResearch-Week4-STORM-data")
#Synopsis
# Storms and other severe weather events can cause both public health and economic problems for communi
# This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) sto
```

```
#Data proccessing

tinytex::install_tinytex()
library("data.table")
library("ggplot2")
library(tinytex)

raw_data <- read.csv("/Users/John/Desktop/ReprdResearch-Week4-STORM-data/repdata_data_StormData.csv.bz2
#Questions #1

# For Questions #1, events will be evaluated by which are most harmful with respect to population helth
# Total number of casualties (TOTAL_CAS)
# is estimated as the sum of INJURIES and FATALITIES for all events.
```

```
raw_data$TOTAL_CAS <- raw_data$FATALITIES + raw_data$INJURIES
```

```
# Questions #2
# For Question #2, events which have the greatest economic consequences will be evaluated.
# The assumption where the economic consequences is the sum of crop damage and property damage TOTAL_DM
# To construct the TOTAL_DMG variable, PROPDMGEXP and CROPDMGEXP will be transformed in 1000, 1e+06 and

value <- function(x) {
  x <- tolower(x)
  if (x == "k") res <- 1000
  if (x == "m") res <- 1e+06
  if (x == "b") res <- 1e+09
  else res <- 1
  res
}

raw_data$PROP_DMG <- raw_data$PROPDMG * sapply(raw_data$PROPDMGEXP, value) /1000000
raw_data$CROP_DMG <- raw_data$CROPDMG * sapply(raw_data$CROPDMGEXP, value) /1000000
raw_data$TOTAL_DMG <- raw_data$PROP_DMG + raw_data$CROP_DMG

#Taking only relevant variables, a new data set is constructed. The new data set is aggregated by EVTYP

proc_data <- raw_data[,c("EVTYPE", "FATALITIES", "INJURIES", "TOTAL_CAS", "PROP_DMG", "CROP_DMG", "TOTA
proc_data <- aggregate(proc_data[,2:7], by=list(proc_data$EVTYPE),FUN=sum, na.rm=TRUE)
colnames(proc_data) <- c("EVTYPE", colnames(proc_data[2:7]))
#top_data() function takes data frame (df), column number (col) and returns the top results.
```

```r
top_data <- function(df, col, top) {
  df <- df[,c(1, col)]
  df <- df[order(df[,2], decreasing = T),]
  df <- df[1:top,]
  rownames(df) <- NULL
  df
}
```

```r
#Results
# Question 1

# Across the United States, which types of events are most harmful with respect to population health?
#   The top 3 events with most FATALITIES are:
top_data(proc_data, 2,3)
```

```
##           EVTYPE FATALITIES
## 1        TORNADO       5633
## 2 EXCESSIVE HEAT       1903
## 3    FLASH FLOOD        978
```

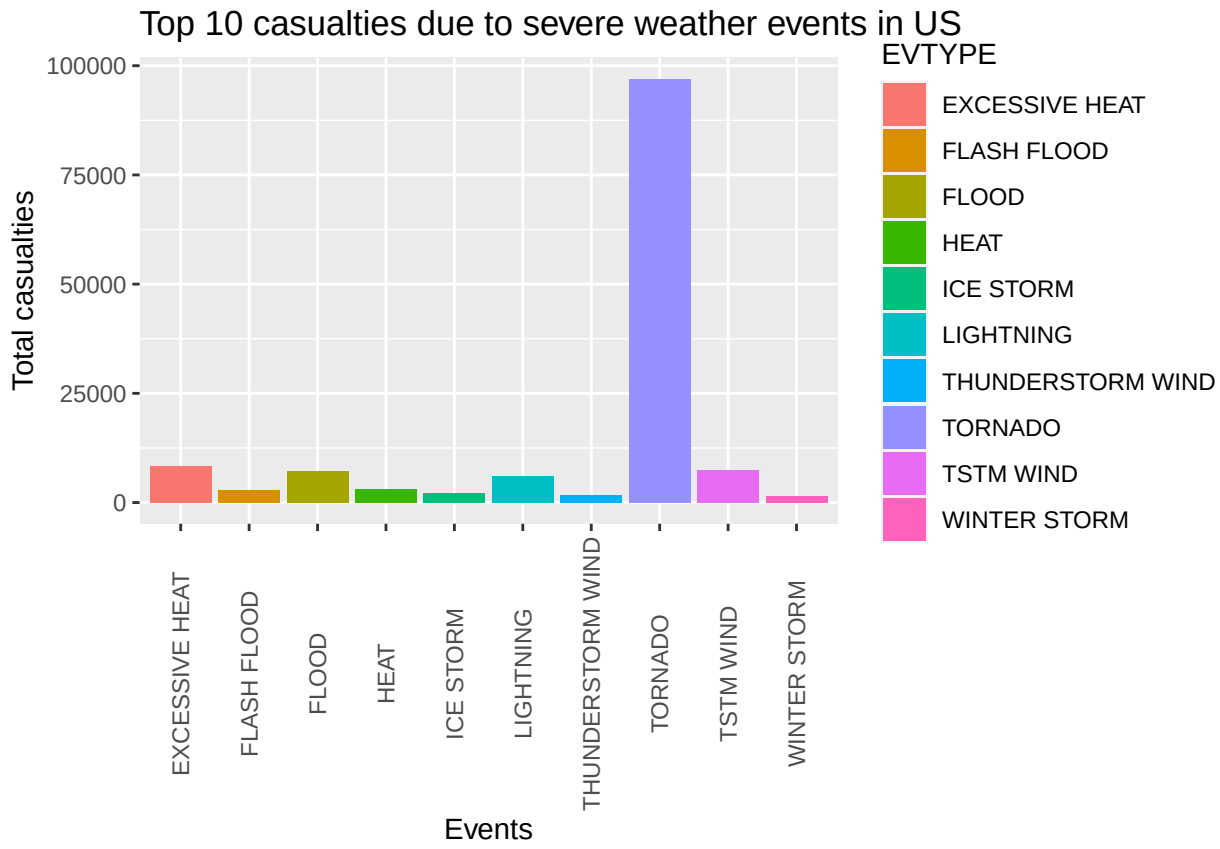
```r
# The top 3 events with most INJURIES are:
top_data(proc_data, 3,3)
```

```
##      EVTYPE INJURIES
## 1   TORNADO    91346
## 2 TSTM WIND     6957
## 3     FLOOD     6789
```

```r
# The top 10 events with most Total casualties are:
cas_data <- top_data(proc_data, 4,10)
print(cas_data)
```

```
##               EVTYPE TOTAL_CAS
## 1            TORNADO     96979
## 2     EXCESSIVE HEAT      8428
## 3          TSTM WIND      7461
## 4              FLOOD      7259
## 5          LIGHTNING      6046
## 6               HEAT      3037
## 7        FLASH FLOOD      2755
## 8          ICE STORM      2064
## 9  THUNDERSTORM WIND      1621
## 10      WINTER STORM      1527
```

```r
# Plot top 10 events with most Total casualties:
ggplot(cas_data, aes(x=EVTYPE, y=TOTAL_CAS, fill=EVTYPE)) +
  geom_bar(stat="identity") +
  ggtitle("Top 10 casualties due to severe weather events in US") +
  xlab("Events") +
  ylab("Total casualties") +
  theme(axis.text.x = element_text(angle = 90, vjust=0.5))
```




```
# Question 2
# Across the United States, which types of events have the greatest economic consequences?
#   The top 3 events causing most property damages are:
top_data(proc_data, 5,3)
```

```
##              EVTYPE  PROP_DMG
## 1             FLOOD 122500.90
## 2 HURRICANE/TYPHOON  65500.01
## 3       STORM SURGE  42560.02
```

```
# The top 3 events causing most crop damages are:
top_data(proc_data, 6,3)
```

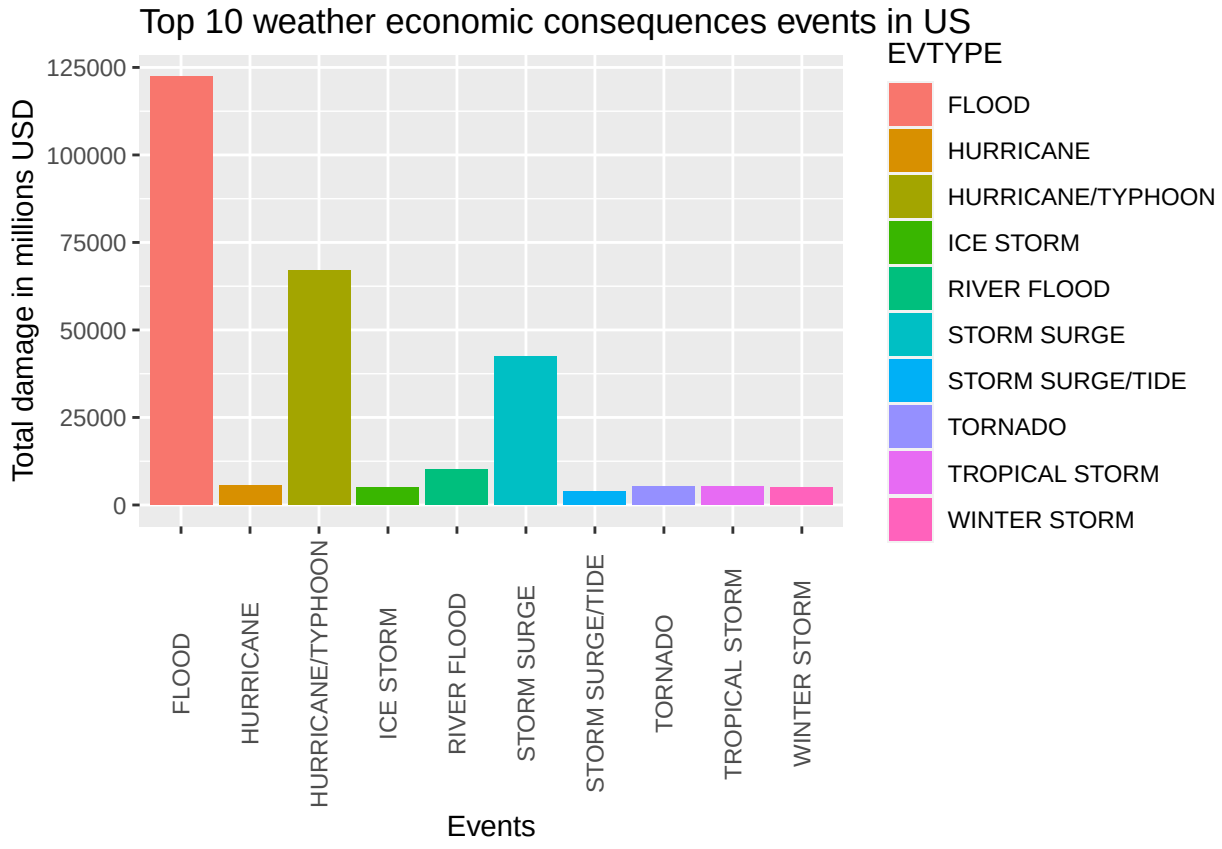
```
##              EVTYPE CROP_DMG
## 1       RIVER FLOOD 5000.003
## 2         ICE STORM 5000.002
## 3 HURRICANE/TYPHOON 1510.005
```

```
# The top 10 events with most Total casualties are:
dmg_data <- top_data(proc_data, 7,10)
print(dmg_data)
```

```
##               EVTYPE  TOTAL_DMG
## 1              FLOOD 122501.068
```

```
## 2  HURRICANE/TYPHOON  67010.011
## 3        STORM SURGE  42560.019
## 4        RIVER FLOOD  10000.017
## 5          HURRICANE   5700.021
## 6            TORNADO   5303.312
## 7     TROPICAL STORM   5150.054
## 8       WINTER STORM   5000.135
## 9          ICE STORM   5000.068
## 10  STORM SURGE/TIDE   4000.008
```

```r
# Plot top 10 events causing most damage:
ggplot(dmg_data, aes(x=EVTYPE, y=TOTAL_DMG, fill=EVTYPE)) +
  geom_bar(stat="identity") +
  ggtitle("Top 10 weather economic consequences events in US") +
  xlab("Events") +
  ylab("Total damage in millions USD") +
  theme(axis.text.x = element_text(angle = 90, vjust=0.5))
```




```r
# Results
#  Across the United States, which types of events are most harmful with respect to population health?
# Across the United States, which types of events have the greatest economic consequences? - The 5 most
```