## Reproducible Research Course Project 2 - Storm Data Analysis

Loading libraries

```
library(readr)
library(data.table)
library(dplyr)
library(ggplot2)
library(lubridate)
library(magrittr)
```

### Synopsis

This R-analysis explores the NOAA Strom Database to evaluate the impact of natural disasters (e.g. tornados, hurricanes, . . . ) on public health and the economy (property damage) in the US.

### Get the data

1. Data is downloaded from the following URL: https://www.coursera.org/learn/reproducible-research/peer/OMZ37/course-project-2 and the Link: Storm Data [47 Mb] is used.
2. For understanding the data set the mentioned documentation is used (ref. National Weather Service-Link: Storm Data Documentation)

### Load the data

```
storm <- read.table("./repdata_data_StormData/repdata_data_StormData.csv",
                    na.strings = c("NA", ""), sep = ",", header=TRUE)
```

### Data Processing I

1. Across the United States, which types of events (as indicated in the EVTYPE) are most harmful with respect to population health (most fatalities and injuries)?

Step 1: Select relevant data variables (columns)

```
storm <- select(storm, STATE, EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP)
```

Step 2: Sum the fatalities per event type (e.g. tornados, hurricanes, . . . ) & arrange the result in decreasing order to find the events with most deaths and injuries:

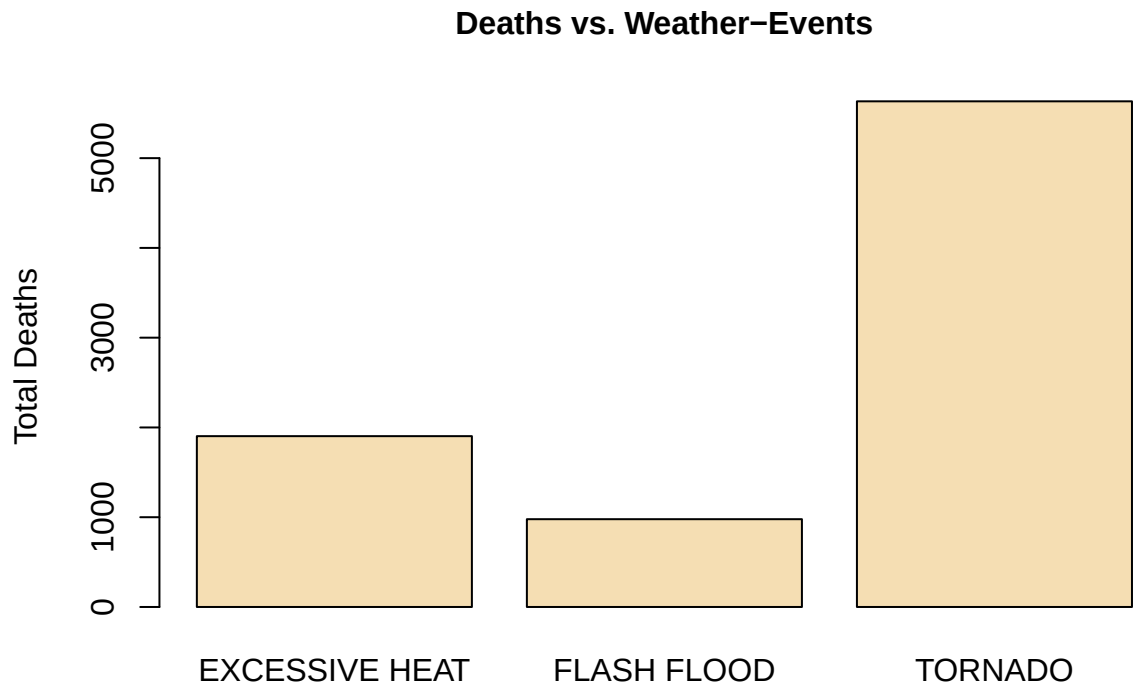
```
fatal_evtype <- aggregate(storm$FATALITIES, by=list(storm$EVTYPE), sum, na.rm = TRUE)
fatal_evtype <- setNames(fatal_evtype, c("EVTYPE","FATALITIES"))
fatal_evtype <- arrange(fatal_evtype, desc(FATALITIES))
injur_evtype <- aggregate(storm$INJURIES, by=list(storm$EVTYPE), sum, na.rm = TRUE)
injur_evtype <- setNames(injur_evtype, c("EVTYPE","INJURIES"))
injur_evtype <- arrange(injur_evtype, desc(INJURIES))
```

Step 3: Transform Event-Types into factor-variable & make barplots for the top 3 of most hazard weather events:

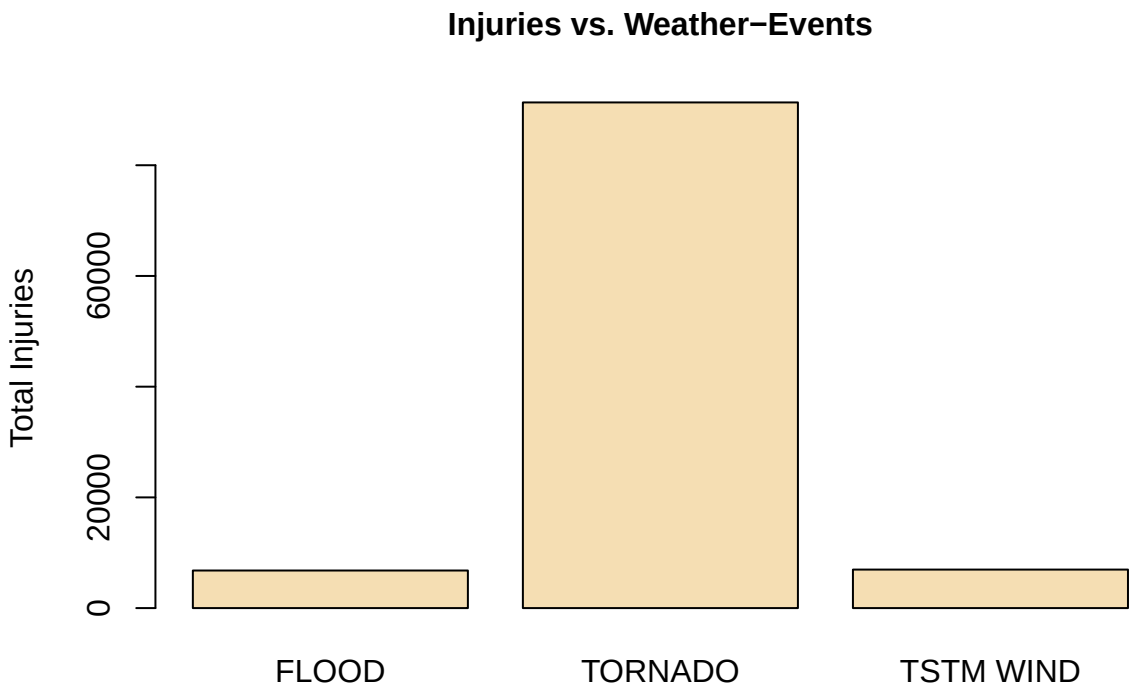
```
fatal_evtype <- fatal_evtype[1:3,]
injur_evtype <- injur_evtype[1:3,]
fatal_evtype <- transform(fatal_evtype, EVTYPE   = factor(EVTYPE))
injur_evtype <- transform(injur_evtype, EVTYPE   = factor(EVTYPE))
```

**Results I:**

```
barplot(FATALITIES ~ EVTYPE, fatal_evtype, main="Deaths vs. Weather-Events",
        xlab = "", ylab="Total Deaths", cex.lab=1, cex.main=1, col="wheat")
```




```
barplot(INJURIES ~ EVTYPE, injur_evtype, main="Injuries vs. Weather-Events",
        xlab = "", ylab="Total Injuries", cex.lab=1, cex.main=1, col="wheat")
```




**Data Processing II**

2. Across the United States, which types of events have the greatest economic consequences?

Step 1: Calculate absolute PROPDMG in US-Dollars (USD) by taking into account PROPDMGEXP (K = kilo, M = million, B = billion)

```
storm$PROPDMGEXP <- gsub("K", "1E3", storm$PROPDMGEXP)
storm$PROPDMGEXP <- gsub("M", "1E6", storm$PROPDMGEXP)
storm$PROPDMGEXP <- gsub("B", "1E9", storm$PROPDMGEXP)
storm$PROPDMGEXP <- as.numeric(storm$PROPDMGEXP)
```

```
## Warning: NAs durch Umwandlung erzeugt
```

```
storm <- mutate(storm, PROPDMG = PROPDMG * PROPDMGEXP)
storm <- select(storm, -PROPDMGEXP)
```

Step 2: Sum the property damage per event type & order the result to find the top 3 most costly events.

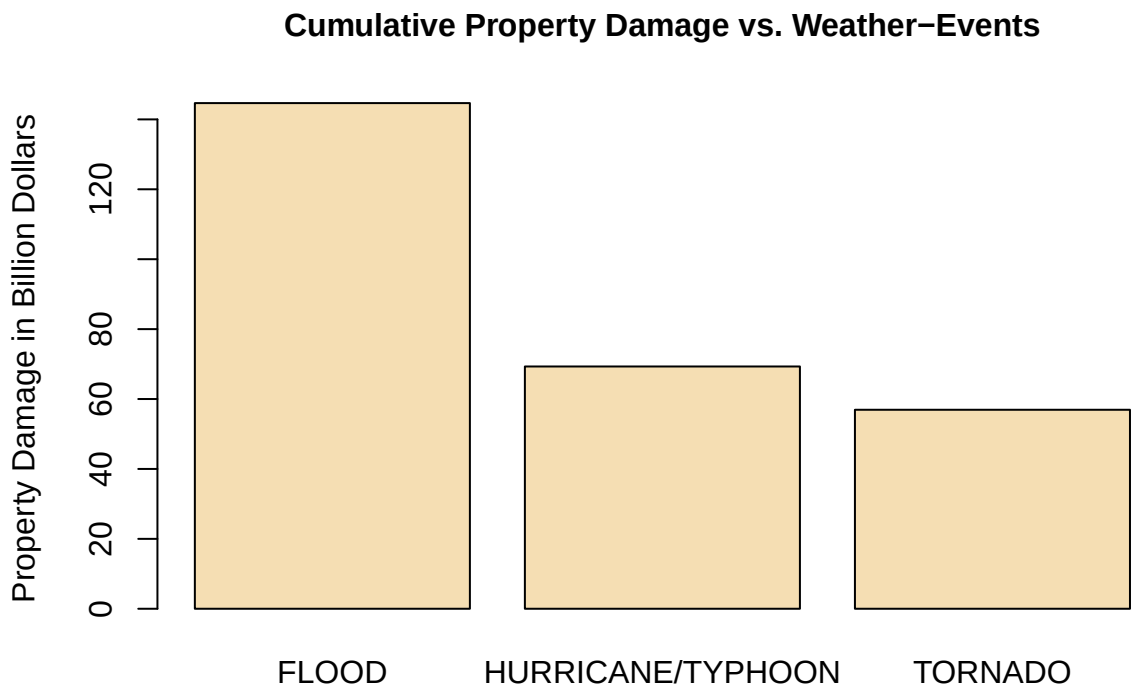
```
propdmg_evtype <- aggregate(storm$PROPDMG, by=list(storm$EVTYPE), sum, na.rm = TRUE)
propdmg_evtype <- setNames(propdmg_evtype, c("EVTYPE","PROPDMG"))
propdmg_evtype <- arrange(propdmg_evtype, desc(PROPDMG))
```

Step 3: Transform Event-Types into factor-variable & make barplot

```r
propdmg_evtype <- propdmg_evtype[1:3,]
propdmg_evtype <- transform(propdmg_evtype, EVTYPE = factor(EVTYPE))
```

**Results II:**

```r
barplot(PROPDMG/1e9 ~ EVTYPE, propdmg_evtype,
        main="Cumulative Property Damage vs. Weather-Events",
        xlab = "", ylab="Property Damage in Billion Dollars", cex.lab=1,
        cex.main=1,   col="wheat")
```




**Conclusion**

In absolute terms of all recorded weather events **tornados** caused the most fatalities (*5633*)/ injuries (*91346*). The most property damage (approximately *140 billion USD*) was caused by **floods**.