# Peer Assessment 2 - Analysis of Storm Data

*Kevin Gardner*

*02/14/2016*

## Synopsis

This document contains R code and documentation to prepare and analyze the following dataset `https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2`.

Our analysis answers two main questions:

- Which types of events are most harmful to population health?
- Which types of events have the greatest economic consequences?

The report contains three plots which illustrate the results of the analysis.

## Data processing

Setting up environment

```
rm(list = ls())
library("data.table")
library("dplyr")
library("ggplot2")
setwd("~/GitHub/05_ReproducableResearch/RepData_PeerAssessment2")
```

Loading data

```
f <- file.path(getwd(), "2FStormData.csv.bz2")
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url, f, mode = "wb")
DT_storm <- data.table(read.csv(f))
```

Selecting relevant subset

```
dt <- DT_storm %>% select(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP,
    CROPDMG, CROPDMGEXP)
```

Cleaning the data

```
# Replace empty values with 0's
dt[FATALITIES == "", `:=`(FATALITIES, "0")]
dt[INJURIES == "", `:=`(INJURIES, "0")]
dt[PROPDMG == "", `:=`(PROPDMG, "0")]
dt[CROPDMG == "", `:=`(CROPDMG, "0")]
# Normalize damage values
dt[PROPDMGEXP == "", `:=`(PROPDMGEXP, "0")]
```

```
dt[PROPDMGEXP == "+" | PROPDMGEXP == "-" | PROPDMGEXP == "?", `:=`(PROPDMGEXP,
    "1")]
dt[PROPDMGEXP == "h" | PROPDMGEXP == "H", `:=`(PROPDMGEXP, "2")]
dt[PROPDMGEXP == "k" | PROPDMGEXP == "K", `:=`(PROPDMGEXP, "3")]
dt[PROPDMGEXP == "m" | PROPDMGEXP == "M", `:=`(PROPDMGEXP, "6")]
dt[PROPDMGEXP == "B", `:=`(PROPDMGEXP, "9")]
dt[CROPDMGEXP == "", `:=`(CROPDMGEXP, "0")]
dt[CROPDMGEXP == "+" | CROPDMGEXP == "-" | CROPDMGEXP == "?", `:=`(CROPDMGEXP,
    "1")]
dt[CROPDMGEXP == "h" | CROPDMGEXP == "H", `:=`(CROPDMGEXP, "2")]
dt[CROPDMGEXP == "k" | CROPDMGEXP == "K", `:=`(CROPDMGEXP, "3")]
dt[CROPDMGEXP == "m" | CROPDMGEXP == "M", `:=`(CROPDMGEXP, "6")]
dt[CROPDMGEXP == "B", `:=`(CROPDMGEXP, "9")]
# Verify normalized values look ok
dt %>% select(PROPDMGEXP) %>% distinct() %>% setorder(PROPDMGEXP) %>% data.table() %>%
    invisible()
dt %>% select(CROPDMGEXP) %>% distinct() %>% setorder(CROPDMGEXP) %>% data.table() %>%
    invisible()
# Recode values as numeric for computation of next step
dt[, `:=`(PROPDMGEXP, as.integer(PROPDMGEXP))]
dt[, `:=`(CROPDMGEXP, as.integer(CROPDMGEXP))]
```

## Data manipulations

Calculating Total Damages for each weather event

```
dt[, `:=`(damageUSD, PROPDMG * 10^(PROPDMGEXP) + CROPDMG * 10^(CROPDMGEXP))]
```

## Data aggregations

Calculating total `FATALITIES` by `EVTYPE`

```
dt <- group_by(dt, EVTYPE)  #group data
fatalities <- dt %>% summarize(FATALITIES = sum(FATALITIES))
summarise(ungroup(dt), sum(FATALITIES)) %>% invisible()  #ungroup data
```

Calculating total `INJURIES` by `EVTYPE`

```
dt <- group_by(dt, EVTYPE)  #group data
injuries <- dt %>% summarize(INJURIES = sum(INJURIES))
summarise(ungroup(dt), sum(INJURIES)) %>% invisible()  #ungroup data
```

Calculating total `damageUSD` by `EVTYPE`

```
dt <- group_by(dt, EVTYPE)  #group data
damages <- dt %>% summarize(DAMAGES.USD = sum(damageUSD))
summarise(ungroup(dt), sum(damageUSD)) %>% invisible()  #ungroup data
```

## Results

Identifying top event types by fatalities

```
topFatalities <- fatalities %>% setorder(-FATALITIES) %>% head(10)
```

Identifying top event types by injuries

```
topInjuries <- injuries %>% setorder(-INJURIES) %>% head(10)
```

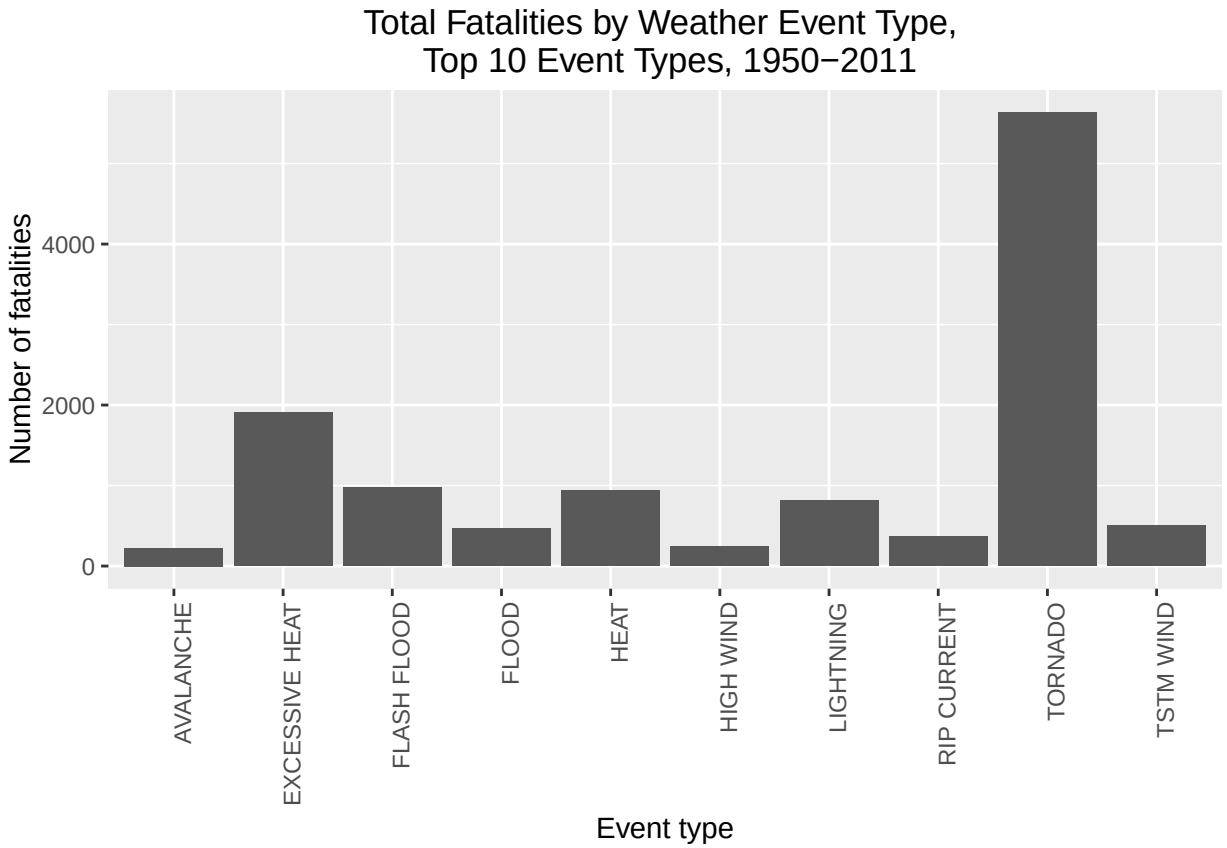
Identifying top event types by damages

```
topDamages <- damages %>% setorder(-DAMAGES.USD) %>% head(10)
```

### Which types of events are most harmful to population health?

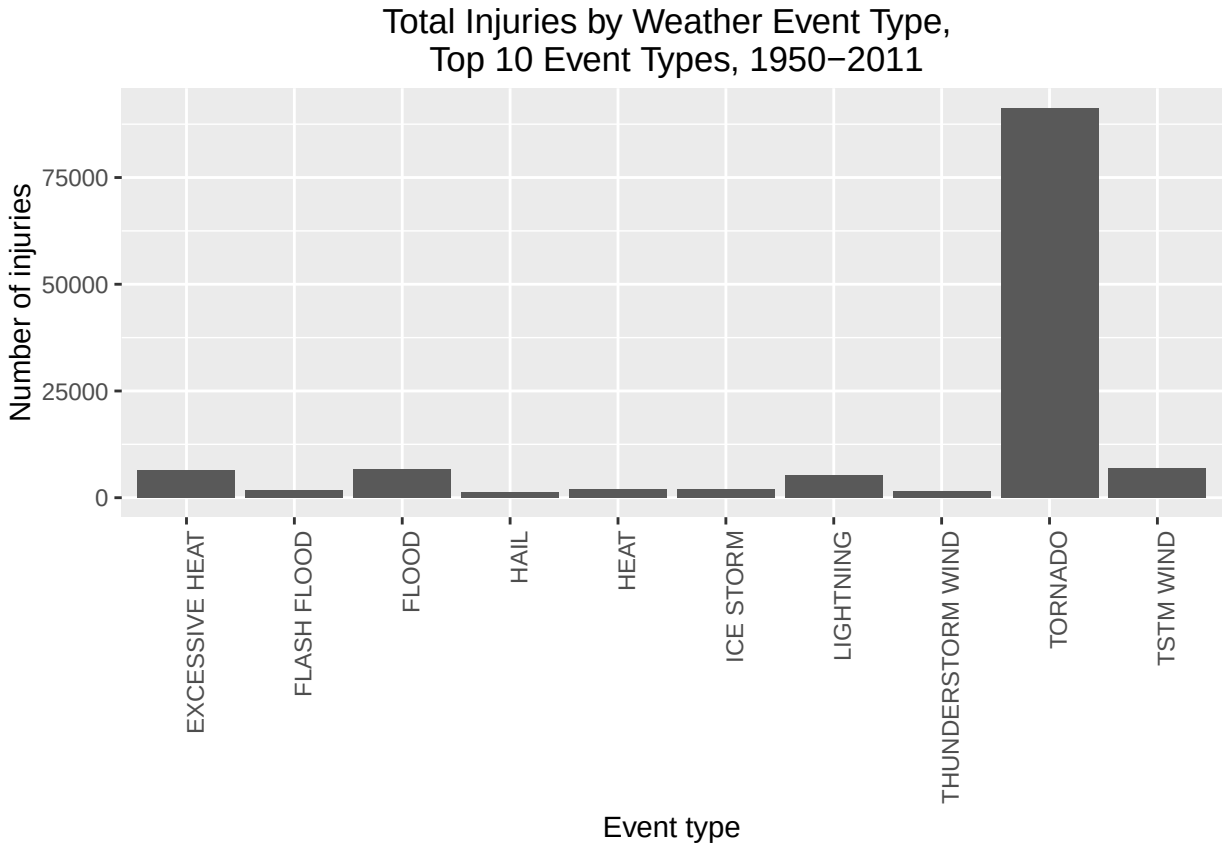
Plotting total `FATALITIES` by `EVTYPE` for top 10 event types

```
ggplot(data = topFatalities, aes(x = topFatalities$EVTYPE, y = topFatalities$FATALITIES)) +
    geom_bar(stat = "identity") + theme(axis.text.x = element_text(angle = 90,
    hjust = 1)) + xlab("Event type") + ylab("Number of fatalities") + ggtitle("Total Fatalities by Weatl
```




Plotting total `INJURIES` by `EVTYPE` for top 10 event types

```
ggplot(data = topInjuries, aes(x = topInjuries$EVTYPE, y = topInjuries$INJURIES)) +
    geom_bar(stat = "identity") + theme(axis.text.x = element_text(angle = 90,
    hjust = 1)) + xlab("Event type") + ylab("Number of injuries") + ggtitle("Total Injuries by Weather
```




Above we observe that, as measured by fatalities and injuries, Tornados are most harmful to population health.

Reporting the top event type for fatalities and injuries we get

```
fatalities %>% setorder(-FATALITIES) %>% head(1) %>% data.table()
```

```
##     EVTYPE FATALITIES
## 1: TORNADO       5633
```

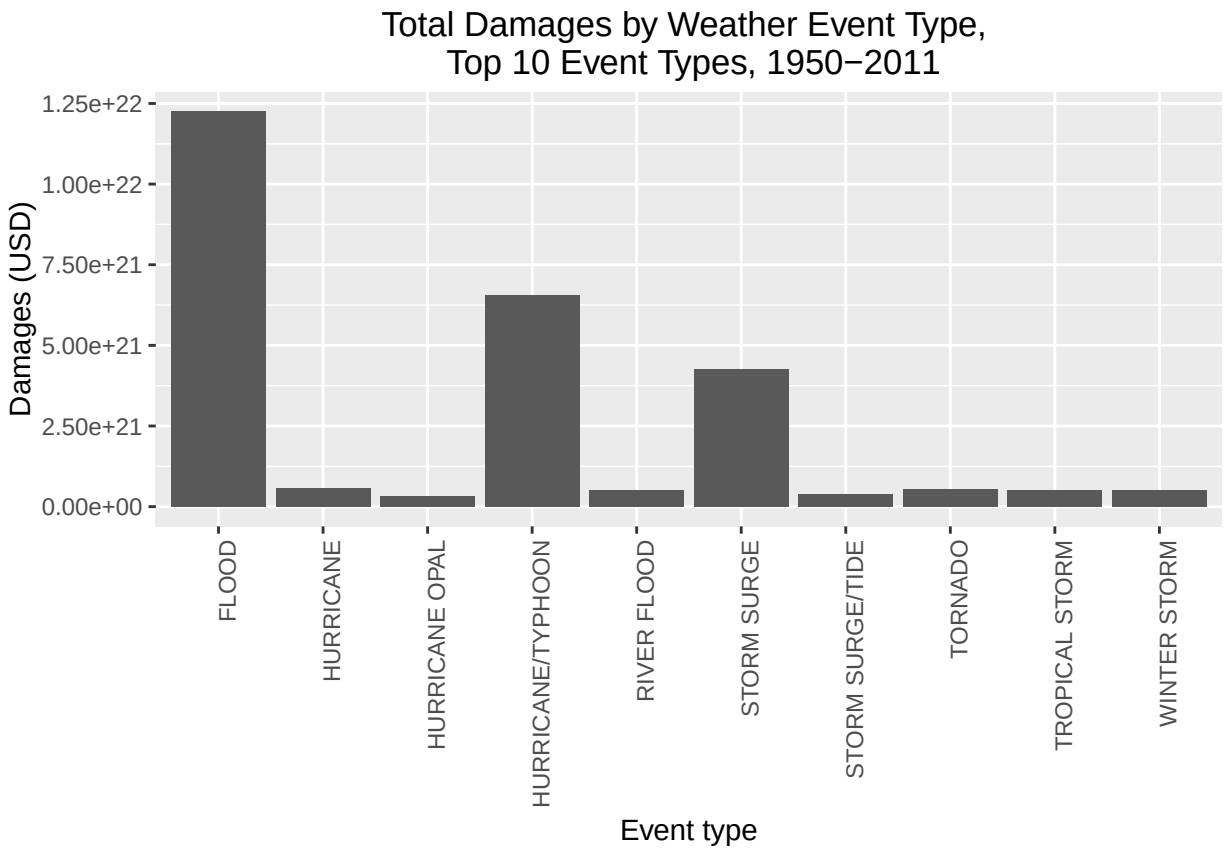
```
injuries %>% setorder(-INJURIES) %>% head(1) %>% data.table()
```

```
##     EVTYPE INJURIES
## 1: TORNADO    91346
```

**Which types of events have the greatest economic consequences?**

Plotting total `DAMAGES.USD` by `EVTYPE` for top 10 event types

```
ggplot(data = topDamages,
       aes(x = topDamages$EVTYPE, y = topDamages$DAMAGES.USD)) +
        geom_bar(stat = "identity") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1)) +
        xlab("Event type") +
        ylab("Damages (USD)") +
        ggtitle("Total Damages by Weather Event Type, \n Top 10 Event Types, 1950-2011")
```




Above we observe that, as measured by damages, Floods have the greatest economic consequences?

Reporting the top event type for damages we get

```
damages %>% setorder(-DAMAGES.USD) %>% head(1) %>% data.table()
```

```
##    EVTYPE  DAMAGES.USD
## 1:  FLOOD 1.225002e+22
```