

Reproducible Research: Peer Assessment 2

Introduction

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

Data

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site: Storm Data [47Mb] There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

National Weather Service Storm Data Documentation

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

Library

```
library(plyr)
library(knitr)
library(ggplot2)
library(grid)
```

Loading Data

```
setwd("/testx")
NOAA <- read.csv("FstormData.csv", stringsAsFactors = FALSE, strip.w
```



Data Cleaning and Processing

```
# Filter of principal columns
FilteredNOAA <- NOAA[ , c("EVTTYPE", "BGN_DATE", "FATALITIES", "INJURIE
                        "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMG

# Ajust the format of Date
FilteredNOAA$BGN_DATE <- as.POSIXct(FilteredNOAA$BGN_DATE, format="%m/

# PROP and CROP Damage countains index of your numbers (Kilo, Hecto,
# that we go transform to decimals (10^2, 10^3, ..)
### Create vector of mapvalues
TransformPropDamage <- mapvalues(FilteredNOAA$PROPDMGEXP,
                                c("K", "M", "", "B", "m", "+", "0", "5", "6",
                                c(10^3, 10^6, 1, 10^9, 10^6, 1, 1, 10^5, 1

TransformCropDamage <- mapvalues(FilteredNOAA$CROPDMGEXP,
                                c("", "M", "K", "m", "B", "?", "0", "k", "2"),
                                c( 1, 10^6, 10^3, 10^6, 10^9, 1, 1, 10^3, 10^2

# Create 2 new columns with exponential applied in to the numbers
FilteredNOAA$PROPTOTALDMG <- as.numeric(TransformPropDamage) * Filter

FilteredNOAA$CROPTOTALDMG <- as.numeric(TransformCropDamage) * Filter

# Sum PROP+CROP Damage (Group total by Event Type )
FilteredNOAA$TOTALDMG <- FilteredNOAA$PROPTOTALDMG + FilteredNOAA$CRO
```

Which types of events are most harmful with respect to population health?

Preparing for plots...

First, compose the data frame, labels and FAIPlot (Fatalities and Injuries Plot)

```

dataf <- data.frame(tolower(gsub(" ","", NOAA$EVTYPE, fixed=TRUE)),
                    NOAA$FATALITIES, NOAA$INJURIES,
                    NOAA$PROPDGMG, tolower(NOAA$PROPDGMGEXP),
                    NOAA$CROPDMG, tolower(NOAA$CROPDMGEXP)
                    )

names(dataf) <- c("EVTYPE", "FATALITIES", "INJURIES", "PROPDGMG", "PROPDGMGEXP", "CROPDMG", "CROPDMGEXP")

# Fatalities and Injures (Harmful events)
fai = dataf[dataf$FATALITIES > 0 | dataf$INJURIES > 0 , ]

FAI <- aggregate ( x = list ( fai$FATALITIES, fai$INJURIES ),
                  by= list ( fai$EVTYPE), FUN=sum, na.rm = TRUE)

names(FAI) <- c("EVTYPE", "FATALITIES", "INJURIES")

FAIplot <- head(FAI[order(-FAI$FATALITIES, -FAI$INJURIES) , ])

```

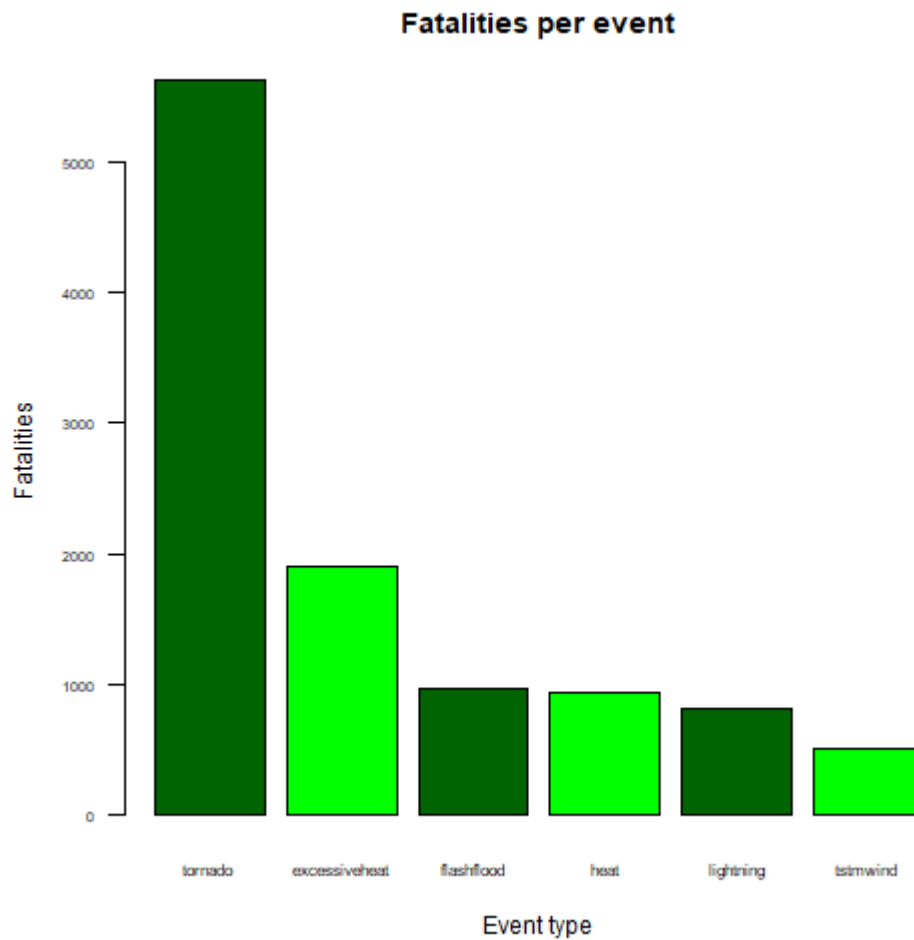
Graph 1) Fatalities

With function barplot, view the Number of Fatalities ty event type.

```

barplot(FAIplot$FATALITIES,
        main="Fatalities per event",
        col=c("dark green","green"),
        xlab="Event type", ylab="Fatalities",
        cex.axis=0.6, cex.names=0.7,
        names.arg = FAIplot$EVTYPE,
        las= 1)

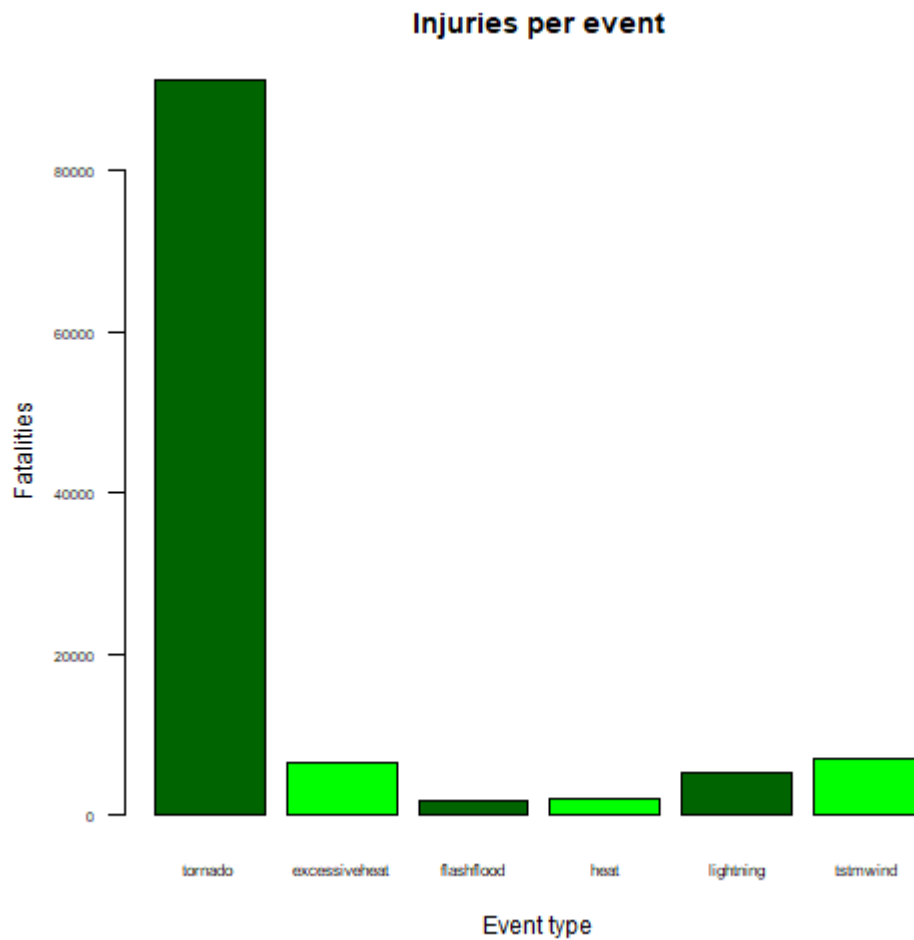
```



Graph 2) Injuries

Now, view the Number of Injuries ty event type.

```
barplot(FAIplot$INJURIES,
        main="Injuries per event",
        col=c("dark green","green"),
        xlab="Event type", ylab="Fatalities",
        cex.axis=0.6, cex.names=0.7,
        names.arg = FAIplot$EVTYPE,
        las= 1)
```



Which types of events have the greatest economic consequences?

Let us now see, through the same collection of dies, which disaster will be considered the greatest impact economically.

Preparing for plot...

First, adjusting data frames.

```

EconImpact = dataf[dataf$PROPDMGEXP %in% c("h","k","m","b") &
  dataf$PROPDMG > 0 &
  dataf$CROPDMGEXP %in% c("h","k","m","b") &
  dataf$CROPDMG > 0, ]

EconImpact$CROPDMGEXP <- as.numeric(
  lapply(EconImpact$CROPDMGEXP,
    function(x) if (x=="h") 100
                  else if (x=="k") 1000
                  else if (x=="m") 1e+06
                  else if (x=="b") 1e+09
                  else 0 ))

EconImpact$PROPDMGEXP <- as.numeric(
  lapply(EconImpact$PROPDMGEXP,
    function(x) if (x=="h") 100
                  else if (x=="k") 1000
                  else if (x=="m") 1e+06
                  else if (x=="b") 1e+09
                  else 0 ))

```

Graph 3) Making plot of economic impact.

```

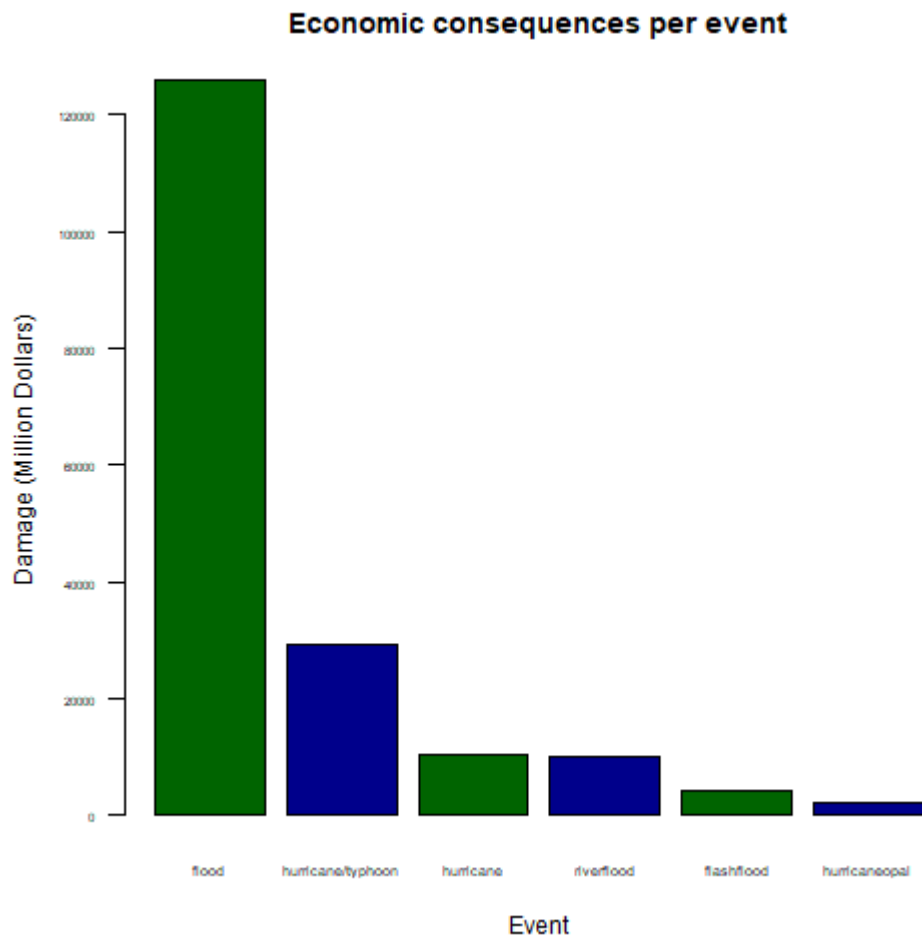
ECONIMPACT <- aggregate(x = list(EconImpact$PROPDMG*EconImpact$PROPDMGEXP,
  by=list(EconImpact$EVTYPE), FUN=sum, na.rm = TRUE)

names(ECONIMPACT) <- c("EVTYPE","PRODMGs","CROPDMGs")

ECONIMPACTplot <- head(ECONIMPACT[order(-ECONIMPACT$PRODMGs, -ECONIMPACT$CROPDMGs), ], 10)

barplot( ( ECONIMPACTplot$PRODMGs+ECONIMPACTplot$CROPDMGs) / 1000000,
  main="Economic consequences per event",
  col=c("dark green","dark blue"),
  xlab="Event",
  ylab="Damage (Million Dollars)",
  cex.axis=0.5, cex.names=0.6,
  names.arg = ECONIMPACTplot$EVTYPE, las= 1, beside = TRUE)

```



Results

Through a quick analysis, we can see that tornadoes are leading the list of personal disasters, in both cases, according to the data worked.

We can clearly see that the floods caused an appreciably greater economic impact compared to other types of damage.

We consider that this historical series brings a kind of reality where, although we had previously had a great impact on people's lives in relation to tornadoes, the economic consequences of the floods were much greater.

Perhaps this can be explained in more detail by better working the data, or even if this analysis promotes a new search, and with more data can be inferred and explain the causes of these relationships of personal and economic damages.