Reproducible Research : Peer Assessment 2

synopsis

This data analysis address the following questions:
1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

After plotting the MOST HARMFUUL EVENTS graph, it is clear TORNADO is most harmful event , AVALANCE and RIP CURRENT are the least harmful.

Costliest events graphs shows FLOOD is costliest event,

Pre-req:

The following packages are installed 1. R.utils 2. ggplot2 3. plyr 4. reshape2

Mannually unzip the Storm Data file

Data Processing:

We load the needed libraries.

```
library(R.utils)
library(ggplot2)
library(plyr)
library(reshape2)
```

Then load the data.

```
storm_data <- read.csv("repdata-data-StormData.csv", stringsAsFactors=FALSE)

storm_data <- data.frame(as.Date(storm_data$BGN_DATE, "%m/%d/%Y %H:%M:%S"),
                      storm_data$EVTYPE,
                      storm_data$FATALITIES,
                      storm_data$INJURIES,
                      storm_data$PROPDMG,
                      as.character(storm_data$PROPDMGEXP),
                      storm_data$CROPDMG,
                      as.character(storm_data$CROPDMGEXP),
                      storm_data$REFNUM)
colnames(storm_data) <- c("BGN_DATE", "EVTYPE", "FATALITIES", "INJURIES",
                          "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP","REFNUM")
```

And, finally we calculate new numeric fields as PROPCASH (combines the PROPDMG and PROPDMGEXP fields to create a numeric value), CROPCASH (combines the CROPDMG and CROPDMGEXP fields to create a numeric value) and TOTCASH (combines the PROPCASH and CROPCASH fields to create a numeric value).

```
# Do not use scientific notation
options(scipen=999)

# Mapping exponents
text.values <- c("h","H","k","K","m","M","b","B")
exp.values <- c(10^2,10^2,10^3,10^3,10^6,10^6,10^9,10^9)
map.exponents <- data.frame(text.values, exp.values)

#Calculating cash values
storm_data <- merge(map.exponents, storm_data,
                    by.x="text.values", by.y="PROPDMGEXP", all.y=TRUE)
names(storm_data)[2] <- "prop.exponents"
storm_data$PROPCASH <- storm_data$PROPDMG * storm_data$prop.exponents
storm_data$PROPCASH[is.na(storm_data$PROPCASH)] <- 0

storm_data <- merge(map.exponents, storm_data[,2:11],
                    by.x="text.values", by.y="CROPDMGEXP", all.y=TRUE)
names(storm_data)[2] <- "crop.exponents"
storm_data$CROPCASH <- storm_data$CROPDMG * storm_data$crop.exponents
storm_data$CROPCASH[is.na(storm_data$CROPCASH)] <- 0

storm_data$TOTCASH <- storm_data$PROPCASH + storm_data$CROPCASH

#cleaning data frame
storm_data <- storm_data[,c(4:7,10:13)]
```

Results

Most harmful types of events with respect to population health across the United States

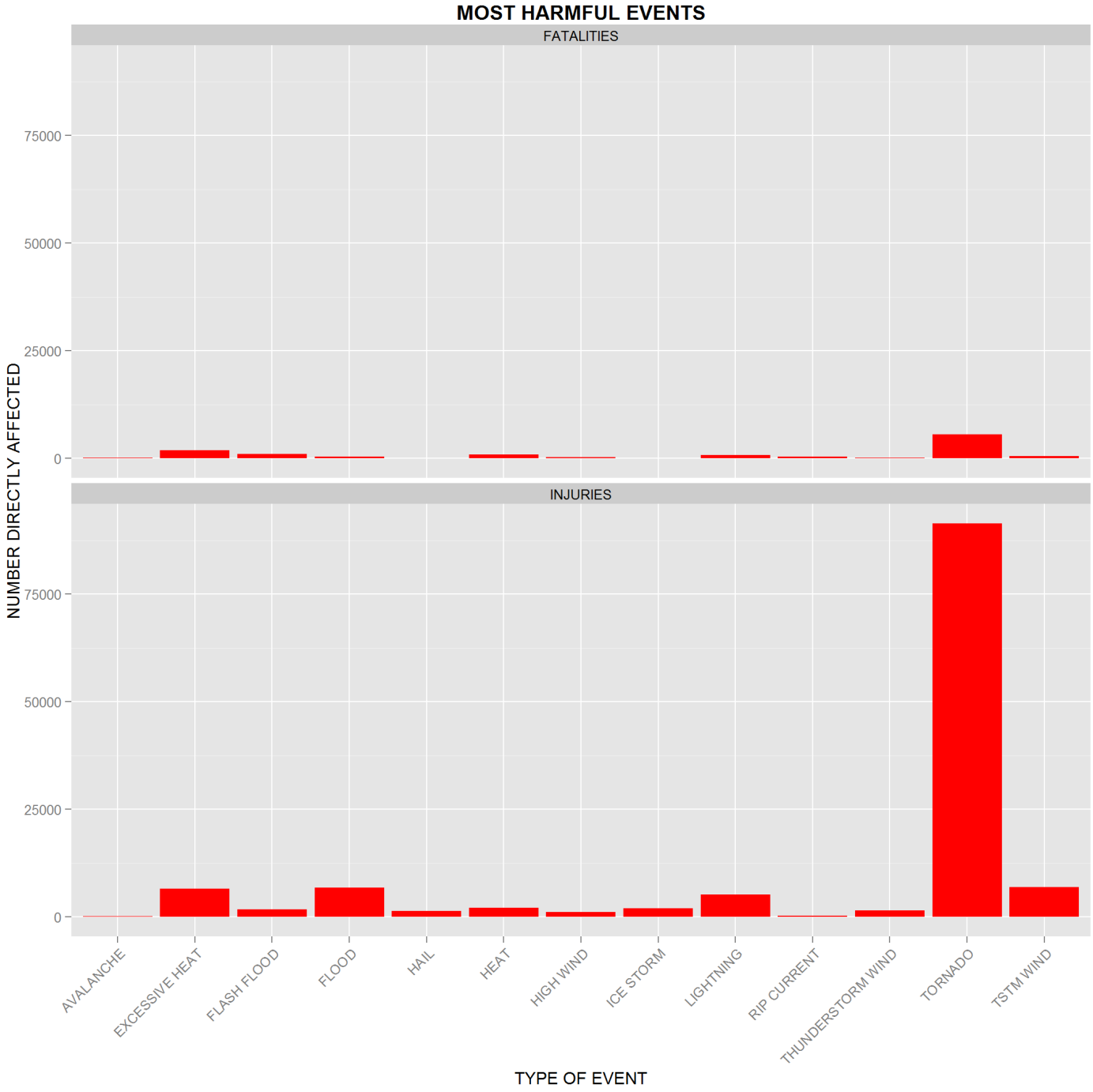
Firstly, we summarize data about fatalities and injuries by type of event. And we create a total data frame.

```
fatalities.total <- ddply(storm_data,.(EVTYPE),summarize,FATALITIES=sum(FATALITIES, na.r
m=TRUE))
injuries.total <- ddply(storm_data,.(EVTYPE),summarize,INJURIES=sum(INJURIES, na.rm=TRU
E))

total <- merge(fatalities.total, injuries.total,
                    by.x="EVTYPE", by.y="EVTYPE", all=TRUE)
```

Now, as we need only the most harmful types of events, we take only those that are greater than 99th percentile. We reshape the data, and draw the graph.

```
total <- total[total$FATALITIES > quantile(total$FATALITIES, probs=0.99) |
                   total$INJURIES > quantile(total$INJURIES, probs=0.99),]

summary <- melt(total, id=c("EVTYPE"), measure.vars=c("FATALITIES","INJURIES"))
```

```
g <- ggplot(summary,
            aes(x=EVTYPE,
                y=value))
g <- g + geom_bar(fill="red", stat="identity")
g <- g + labs(x = "TYPE OF EVENT")
g <- g + labs(y = "NUMBER DIRECTLY AFFECTED")
g <- g + labs(title="MOST HARMFUL EVENTS")
g <- g + facet_wrap( ~ variable, ncol=1)
g <- g + theme(plot.title = element_text(lineheight=.8, face="bold"),
               axis.text.x=element_text(angle=45,vjust=1,hjust=1))
print(g)
```



Costliers types of events across the United States

Now, as we need only the costliers types of events, we take only those that are greater than 99th percentile.

```
economic.total <- ddply(storm_data,.(EVTYPE),summarize,TOTCASH=sum(TOTCASH, na.rm=TRUE))

g <- ggplot(economic.total[economic.total$TOTCASH > quantile(economic.total$TOTCASH, prob
s=0.99),],
            aes(x=EVTYPE,
                y=TOTCASH/10^9))
#g <- g + geom_bar(fill="#00BFC4", stat="identity")
g <- g + geom_bar(fill="red", stat="identity")

g <- g + labs(x = "TYPES OF EVENT")
g <- g + labs(y = "BILLION US $ ")
g <- g + labs(title="COSTLIEST EVENTS")
g <- g + theme(plot.title = element_text(lineheight=.8, face="bold"),
               axis.text.x=element_text(angle=45,vjust=1,hjust=1))
print(g)
```