Reproducible Research: Peer Assessment 2

The Impact of Severe Weather Events on Public Health and Economy in the United States

Synopsis

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. Based on this data we found that tornados and excessive heat generate the highest number of fatalities and injuries while tornados, floods and powerful winds incur the highest economic damages.

Data Processing

```
suppressMessages(library(R.utils))
suppressMessages(library(ggplot2))
```

Here we download and unzip the data set.

```
if (!"StormData.csv.bz2" %in% dir("./data/")) {
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile =
"data/StormData.csv.bz2")
  bunzip2("data/StormData.csv.bz2", overwrite=TRUE, remove=FALSE)
}
```

Then we read the data in R.

```
if (!"StormData" %in% ls()) {
    StormData <- read.csv("data/StormData.csv", sep = ",")
}
dim(StormData)
```

```
## [1] 902297     37
```

We extract the variables related to fatalities, injuries, property and crop damages along with the event type.

```
StormDataVar <- c("EVTYPE","FATALITIES","INJURIES","PROPDMG","CROPDMG")
newStormData <- subset(StormData, select=StormDataVar)
```

We now aggregate the damages on event types.

```
StormDataVar1 <- c("FATALITIES","INJURIES","PROPDMG","CROPDMG")
StormDamage <- aggregate(newStormData[StormDataVar1], list(newStormData$EVTYPE), sum)
summary(StormDamage)
```

```
##                   Group.1      FATALITIES         INJURIES
##     HIGH SURF ADVISORY:  1   Min.   :   0.00   Min.   :    0.0
##   COASTAL FLOOD       :  1   1st Qu.:   0.00   1st Qu.:    0.0
##   FLASH FLOOD         :  1   Median :   0.00   Median :    0.0
##   LIGHTNING           :  1   Mean   :  15.38   Mean   :  142.7
##   TSTM WIND           :  1   3rd Qu.:   0.00   3rd Qu.:    0.0
##   TSTM WIND (G45)     :  1   Max.   :5633.00   Max.   :91346.0
##  (Other)              :979
##     PROPDMG           CROPDMG
##  Min.   :      0   Min.   :     0
##  1st Qu.:      0   1st Qu.:     0
##  Median :      0   Median :     0
##  Mean   :  11050   Mean   :  1399
##  3rd Qu.:     35   3rd Qu.:     0
##  Max.   :3212258   Max.   :579596
##
```

It is also a good idea to delete the rows with only zeros in the damages columns.

```
StormDamage <- StormDamage[rowSums(StormDamage[, -1]>0) != 0, ]
```

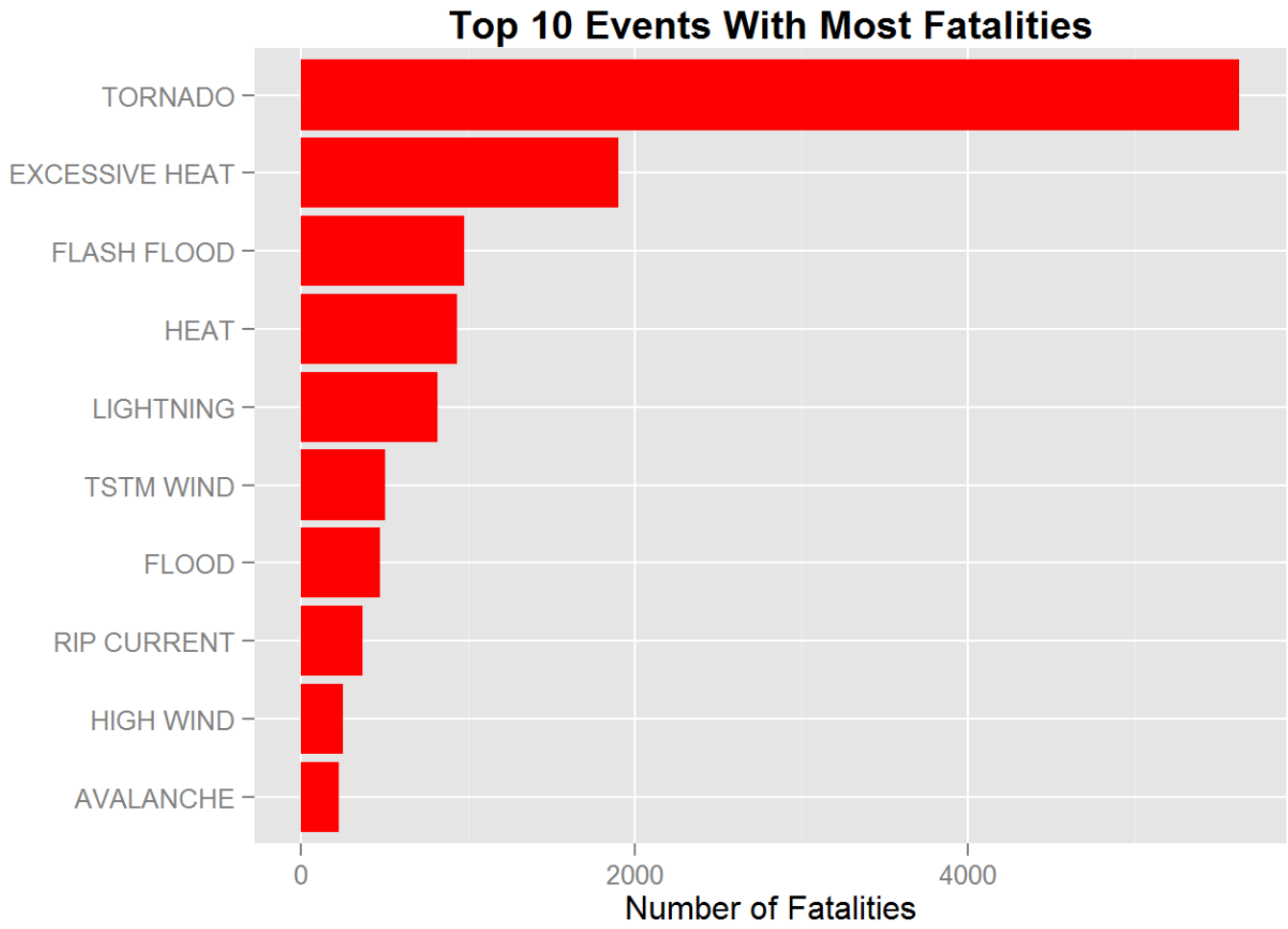
Then we should combine the damages from property and crops into one variable since they are monetary based.

```
StormDamage$ECONOMIC <- StormDamage$PROPDMG + StormDamage$CROPDMG
StormDamage$CROPDMG = NULL
StormDamage$PROPDMG = NULL
```

Results

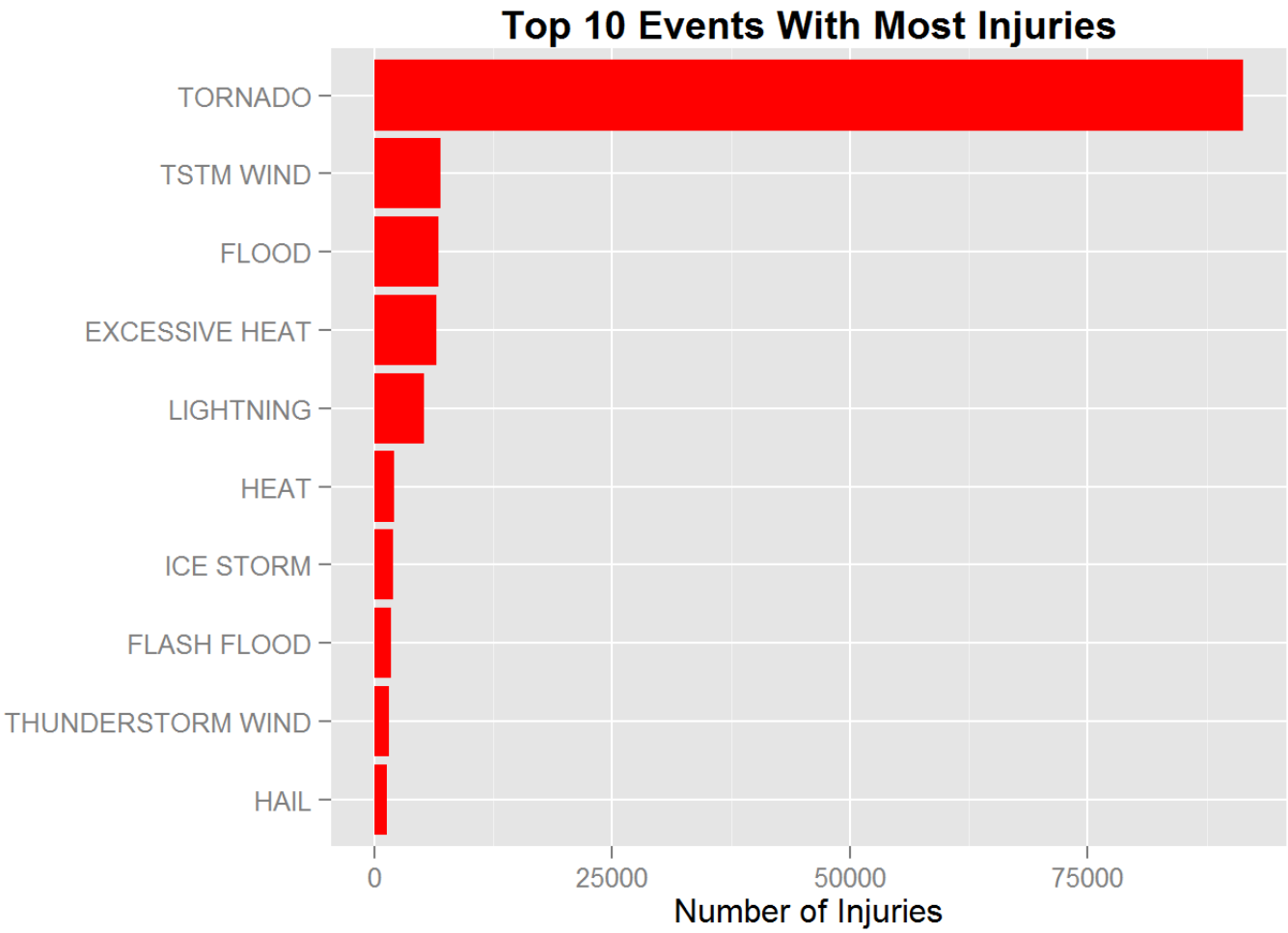
Most Harmful Events Based on the Number of Fatalities

```
fatalities <- data.frame(StormDamage$Group.1,StormDamage$FATALITIES)
names(fatalities) <- c("EVTYPE","FATALITIES")
fatalities <- fatalities[order(-fatalities$FATALITIES),]
fatalities <- head(fatalities,10)
ggplot(data=fatalities, aes(x = reorder(EVTYPE, FATALITIES), y = FATALITIES)) +
    geom_bar(stat="identity", fill="red") + coord_flip() +
    xlab("") + ylab("Number of Fatalities") +
    ggtitle("Top 10 Events With Most Fatalities") +
    theme(plot.title = element_text(lineheight=10, face="bold"))
```



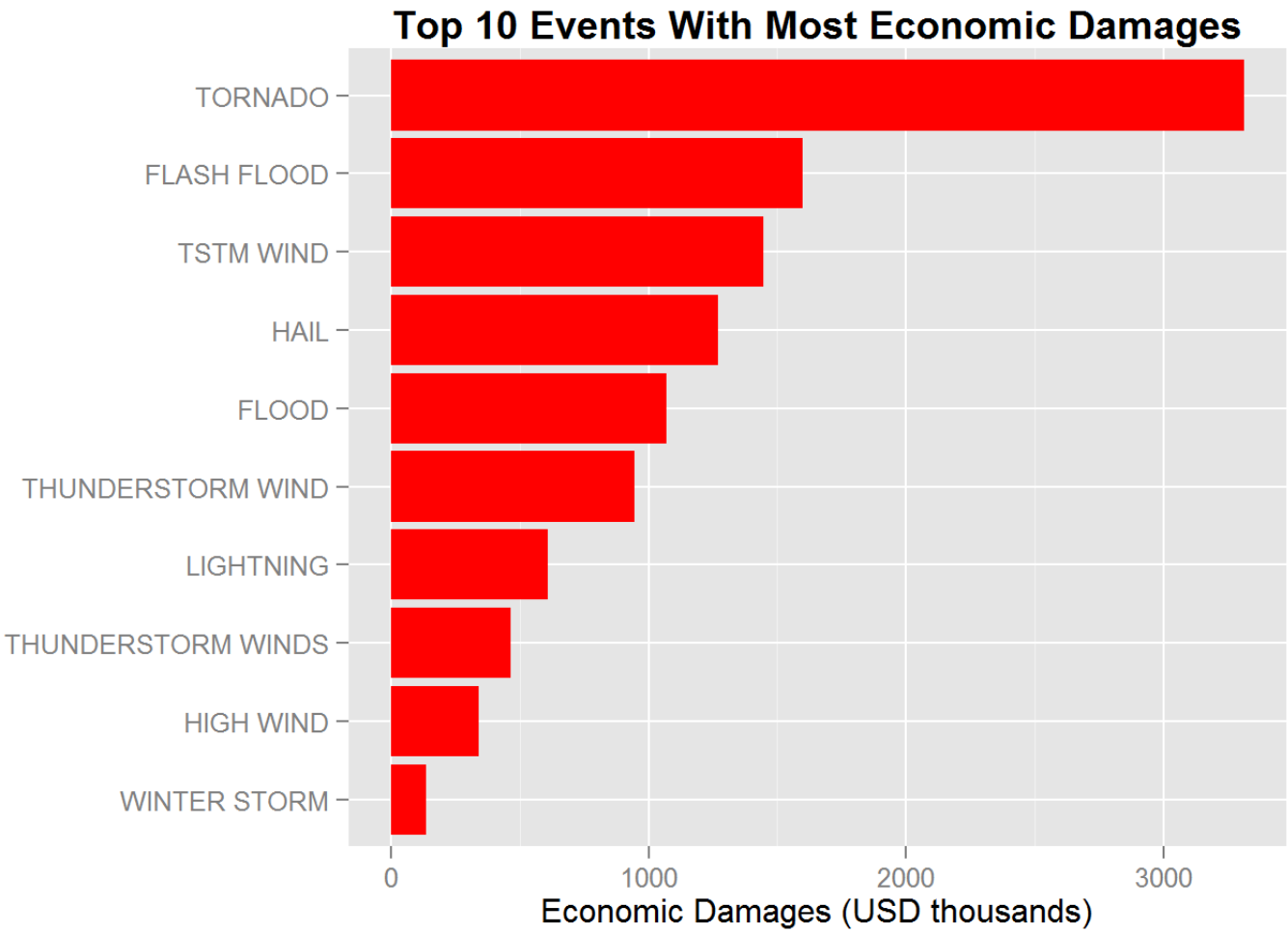
Most Harmful Events Based on the Number of Injuries

```
injuries <- data.frame(StormDamage$Group.1,StormDamage$INJURIES)
names(injuries) <- c("EVTYPE","INJURIES")
injuries <- injuries[order(-injuries$INJURIES),]
injuries <- head(injuries,10)
ggplot(data=injuries, aes(x = reorder(EVTYPE, INJURIES), y = INJURIES)) +
    geom_bar(stat="identity", fill="red") + coord_flip() +
    xlab("") + ylab("Number of Injuries") +
    ggtitle("Top 10 Events With Most Injuries") +
    theme(plot.title = element_text(lineheight=10, face="bold"))
```



Most Harmful Events Based on Economic Damages

```
economic <- data.frame(StormDamage$Group.1,StormDamage$ECONOMIC)
names(economic) <- c("EVTYPE","ECONOMIC")
economic <- economic[order(-economic$ECONOMIC), ]
economic <- head(economic,10)
economic$ECONOMIC <- economic$ECONOMIC/1000
ggplot(data=economic, aes(x = reorder(EVTYPE, ECONOMIC), y = ECONOMIC)) +
    geom_bar(stat="identity", fill="red") + coord_flip() +
    xlab("") + ylab("Economic Damages (USD thousands)") +
    ggtitle("Top 10 Events With Most Economic Damages") +
    theme(plot.title = element_text(lineheight=10, face="bold"))
```



Conclusion

Based on this analysis we found that **tornados** and **excessive heat** are the most harmful to the population health while **tornados**, **flash floods** and **thunderstorm winds** generate bigges economic damages.