# NOAA Storm Database and severe weather events.

*Jun Li jun.li3@bms.com*

## Contents

# 1 Introduction

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.

# 2 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. It is downloaded from the course web site:

# 3 Assignment

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. You must use the database to answer the questions below and show the code for your entire analysis. Your analysis can consist of tables, figures, or other summaries. You may use any R package you want to support your analysis.

Two Questions to address

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?

# 4 Load Libraries

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(printr) # Pretty printing in knitr
library(gplots)
library(plyr)
library(gridExtra)
```

# 5 Read Data

```
storm = read.csv("repdata-data-StormData.csv", header = TRUE, sep = ",")
storm <- mutate_each(storm, funs(toupper))
```

# 6 Events most harmful to population health: fatalities and Injuries

```
event = as.data.frame(storm[, c("EVTYPE" , "FATALITIES", "INJURIES")])
event[,2]<- as.numeric(event[,2])
event[,3]<- as.numeric(event[,3])
colclass <- lapply(event, class)
colclass
```

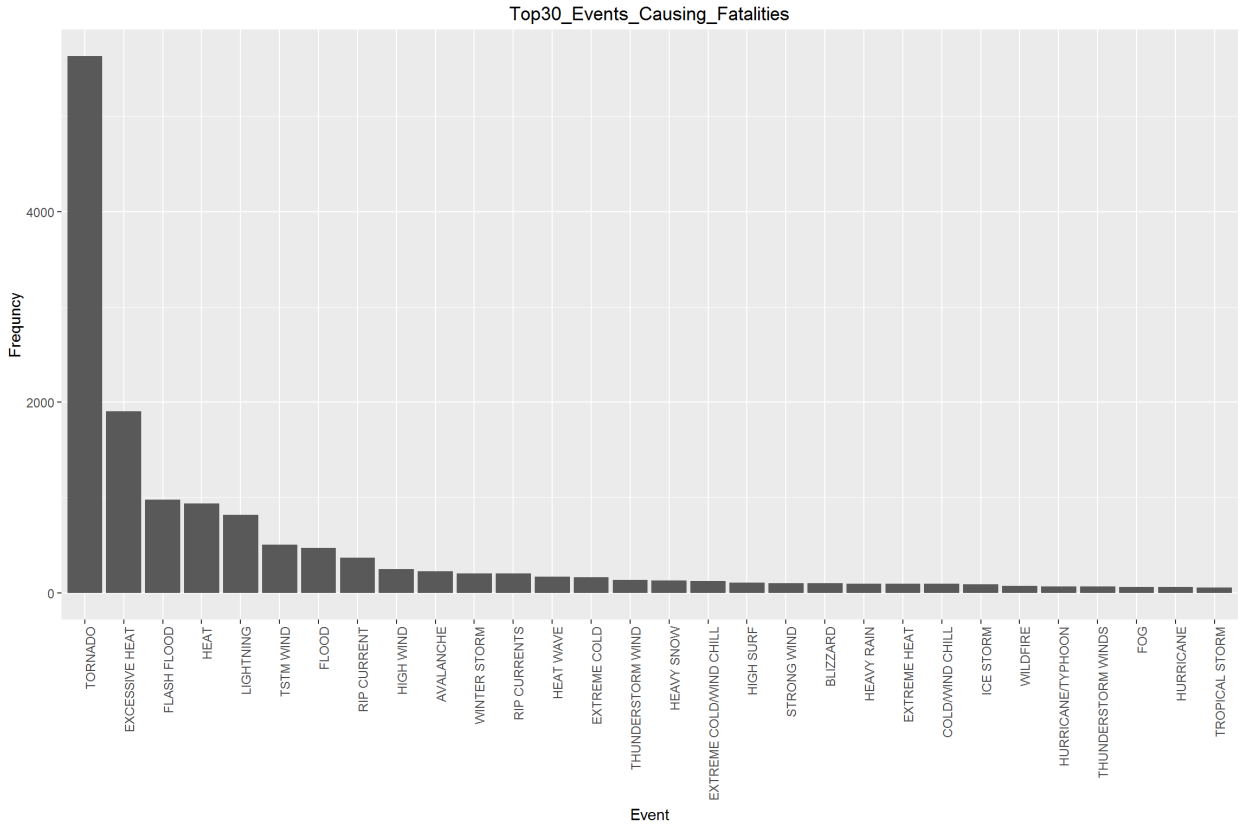
```
## $EVTYPE
## [1] "character"
##
## $FATALITIES
## [1] "numeric"
##
## $INJURIES
## [1] "numeric"
```

## 6.1 Events Causing Fatalities

```
event_agg <- aggregate(x=event[,2], by=list(event[,1]), FUN=sum, na.rm=TRUE)
names(event_agg)[1] <- colnames(event[1])
names(event_agg)[2] <- colnames(event[2])

event_order<- event_agg[order(event_agg[,2], decreasing = TRUE),]
Top30 <- event_order[1:30,]
rownames(Top30) <- NULL
eve_counts <- Top30[,2]
eve_list <- c(Top30[,1])
eve_name <- Top30[,1]

g <- ggplot(Top30, aes(x = reorder(eve_list, -eve_counts), y = eve_counts)) +
        geom_bar(stat = "identity") +
        ggtitle("Top30_Events_Causing_Fatalities") +
        labs(x="Event",y="Frequncy") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(g)
```




```
png(file = "Top_30_Events_Causing_Fatalities.png", width=8,height=6, units = 'in', res = 300)

tb <- knitr::kable(Top30, digits = 5, caption = "Top_30_Events_Causing_Fatalities", col.names = c("Even
print(tb)
```

```
##
##
## Table: Top_30_Events_Causing_Fatalities
##
## Events                      Fatalities
## ----------------------  -----------
## TORNADO                         5633
## EXCESSIVE HEAT                  1903
## FLASH FLOOD                      978
## HEAT                             937
## LIGHTNING                        816
## TSTM WIND                        504
## FLOOD                            470
## RIP CURRENT                      368
## HIGH WIND                        248
## AVALANCHE                        224
## WINTER STORM                     206
## RIP CURRENTS                     204
## HEAT WAVE                        172
## EXTREME COLD                     162
## THUNDERSTORM WIND                133
```

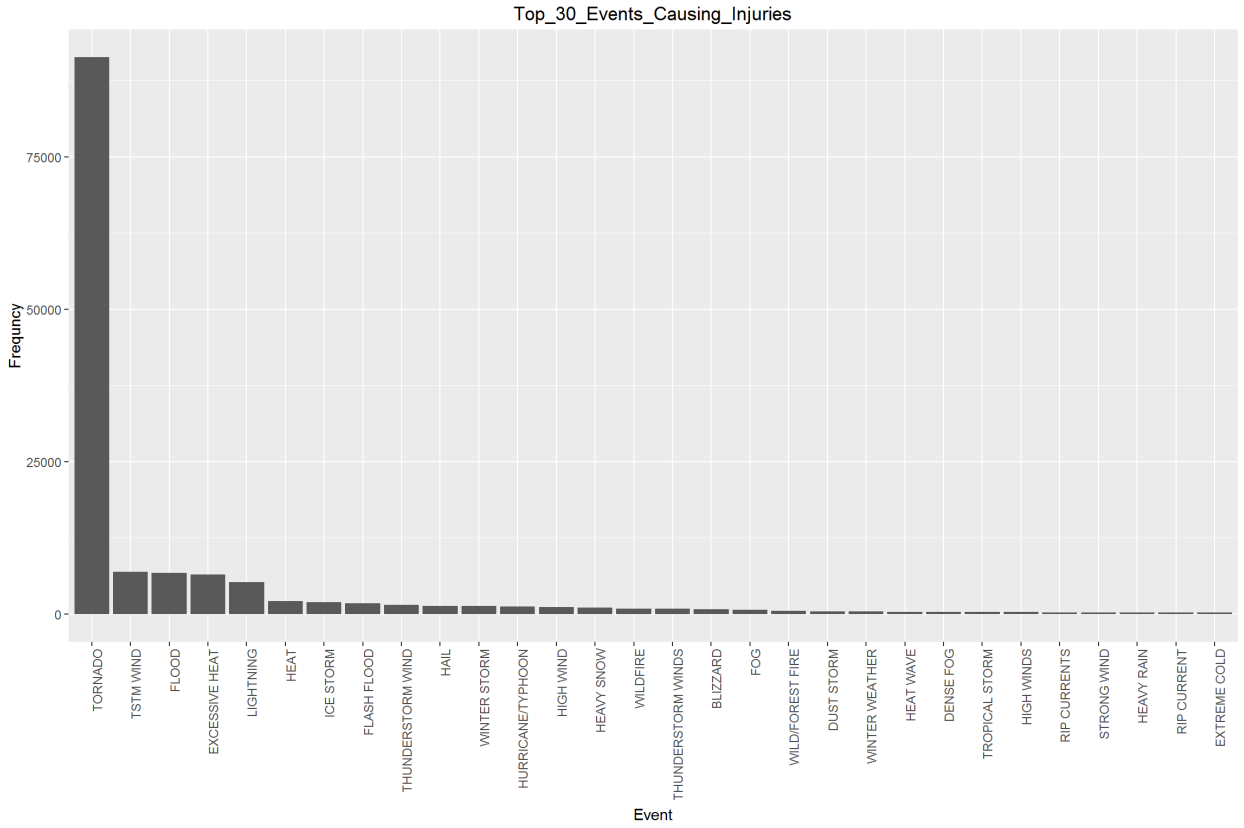
```
## HEAVY SNOW                     127
## EXTREME COLD/WIND CHILL        125
## HIGH SURF                      104
## STRONG WIND                    103
## BLIZZARD                       101
## HEAVY RAIN                      98
## EXTREME HEAT                    96
## COLD/WIND CHILL                 95
## ICE STORM                       89
## WILDFIRE                        75
## HURRICANE/TYPHOON               64
## THUNDERSTORM WINDS              64
## FOG                             62
## HURRICANE                       61
## TROPICAL STORM                  58
```

## 6.2 Events Causing Injuries

```
event_agg2 <- aggregate(x=event[,3], by=list(event[,1]), FUN=sum, na.rm=TRUE)
names(event_agg2)[1] <- colnames(event[1])
names(event_agg2)[2] <- colnames(event[3])

event_order2<- event_agg2[order(event_agg2[,2], decreasing = TRUE),]
Top30_2 <- event_order2[1:30,]
rownames(Top30_2) <- NULL
eve_counts2 <- Top30_2[,2]
eve_list2 <- c(Top30_2[,1])
eve_name2 <- Top30_2[,1]

g2 <- ggplot(Top30_2, aes(x = reorder(eve_list2, -eve_counts2), y = eve_counts2)) +
        geom_bar(stat = "identity") +
        ggtitle("Top_30_Events_Causing_Injuries") +
        labs(x="Event",y="Frequncy") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(g2)
```




```
png(file = "Top_30_Events_Causing_Injuries.png", width=8,height=6, units = 'in', res = 300)

tb2 <- knitr::kable(Top30_2, digits = 5, caption = "Top_30_Events_Causing_Injuries", col.names = c("Eve
print(tb2)
```

```
##
##
## Table: Top_30_Events_Causing_Injuries
##
## Events                Injuries
## -------------------  ---------
## TORNADO                  91346
## TSTM WIND                 6957
## FLOOD                     6789
## EXCESSIVE HEAT            6525
## LIGHTNING                 5230
## HEAT                      2100
## ICE STORM                 1975
## FLASH FLOOD               1777
## THUNDERSTORM WIND         1488
## HAIL                      1361
## WINTER STORM              1321
## HURRICANE/TYPHOON         1275
## HIGH WIND                 1137
## HEAVY SNOW                1021
## WILDFIRE                   911
```

```
## THUNDERSTORM WINDS         908
## BLIZZARD                   805
## FOG                        734
## WILD/FOREST FIRE           545
## DUST STORM                 440
## WINTER WEATHER             398
## HEAT WAVE                  379
## DENSE FOG                  342
## TROPICAL STORM             340
## HIGH WINDS                 302
## RIP CURRENTS               297
## STRONG WIND                280
## HEAVY RAIN                 251
## RIP CURRENT                232
## EXTREME COLD               231
```

# 7 Events have the greatest economic consequences

```
##  [1] ""  "-" "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8" "B" "H" "K" "M"
```

```
## $EVTYPE
## [1] "character"
##
## $PROPDMG
## [1] "numeric"
##
## $PROPDMGEXP
## [1] "character"
```

## 7.1 Events causing Property Damage

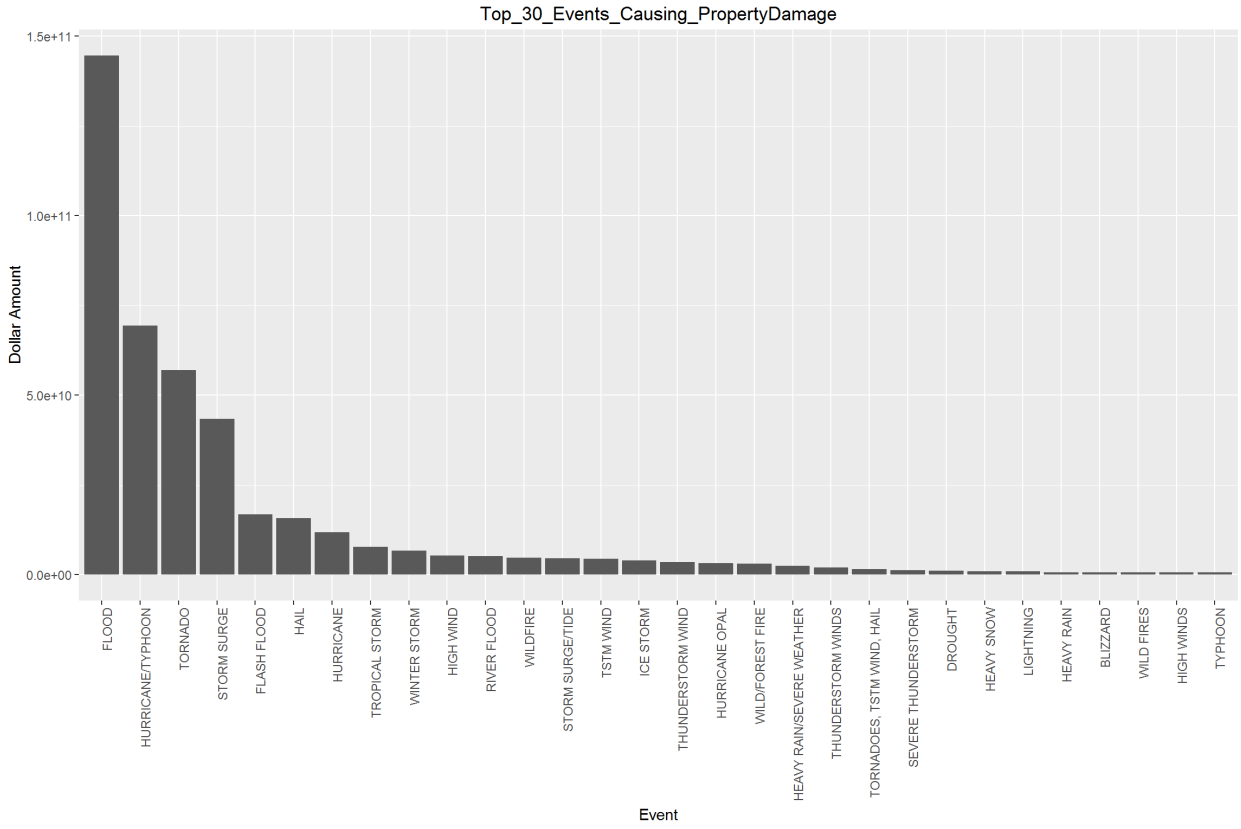
```
## $EVTYPE
## [1] "character"
##
## $PROPDMG
## [1] "numeric"
##
## $PROPDMGEXP
## [1] "character"
```

```
propnew_agg <- aggregate(x=propnew[,2], by=list(propnew[,1]), FUN=sum, na.rm=TRUE)
names(propnew_agg)[1] <- colnames(propnew[1])
names(propnew_agg)[2] <- colnames(propnew[2])

propnew_order<- propnew_agg[order(propnew_agg[,2], decreasing = TRUE),]
Top30_ProDmg <- propnew_order[1:30,]
eve_counts <- Top30_ProDmg[,2]
eve_list <- c(Top30_ProDmg[,1])
eve_name <- Top30_ProDmg[,1]

g <- ggplot(Top30_ProDmg, aes(x = reorder(eve_list, -eve_counts), y = eve_counts)) +
```

```
        geom_bar(stat = "identity") +
        ggtitle("Top_30_Events_Causing_PropertyDamage") +
        labs(x="Event",y="Dollar Amount") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(g)
```




```
png(file = "Top_30_Events_Causing_PropertyDamage.png", width=8,height=6, units = 'in', res = 300)

tb3 <- knitr::kable(Top30_ProDmg, digits = 5, caption = "Top_30_Events_Causing_PropertyDamage", col.name
print(tb3)
```

```
##
##
## Table: Top_30_Events_Causing_PropertyDamage
##
##       Events                          PropDmg
## ----  ---------------------------  ------------
## 154   FLOOD                        144657709807
## 372   HURRICANE/TYPHOON             69305840000
## 758   TORNADO                       56947380677
## 599   STORM SURGE                   43323536000
## 138   FLASH FLOOD                   16822673979
## 212   HAIL                          15735267513
## 363   HURRICANE                     11868319010
```

```
## 772   TROPICAL STORM                 7703890550
## 888   WINTER STORM                   6688497251
## 320   HIGH WIND                      5270046295
## 529   RIVER FLOOD                    5118945500
## 875   WILDFIRE                       4765114000
## 600   STORM SURGE/TIDE               4641188000
## 779   TSTM WIND                      4484958495
## 387   ICE STORM                      3944927860
## 685   THUNDERSTORM WIND              3483122472
## 370   HURRICANE OPAL                 3172846000
## 873   WILD/FOREST FIRE               3001829500
## 262   HEAVY RAIN/SEVERE WEATHER      2500000000
## 711   THUNDERSTORM WINDS             1944590859
## 766   TORNADOES, TSTM WIND, HAIL     1600000000
## 541   SEVERE THUNDERSTORM            1205360000
## 84    DROUGHT                       1046106000
## 274   HEAVY SNOW                     932759140
## 418   LIGHTNING                      930379430
## 254   HEAVY RAIN                     694248090
## 28    BLIZZARD                      659213950
## 872   WILD FIRES                     624100000
## 337   HIGH WINDS                    608323748
## 802   TYPHOON                       600230000
```

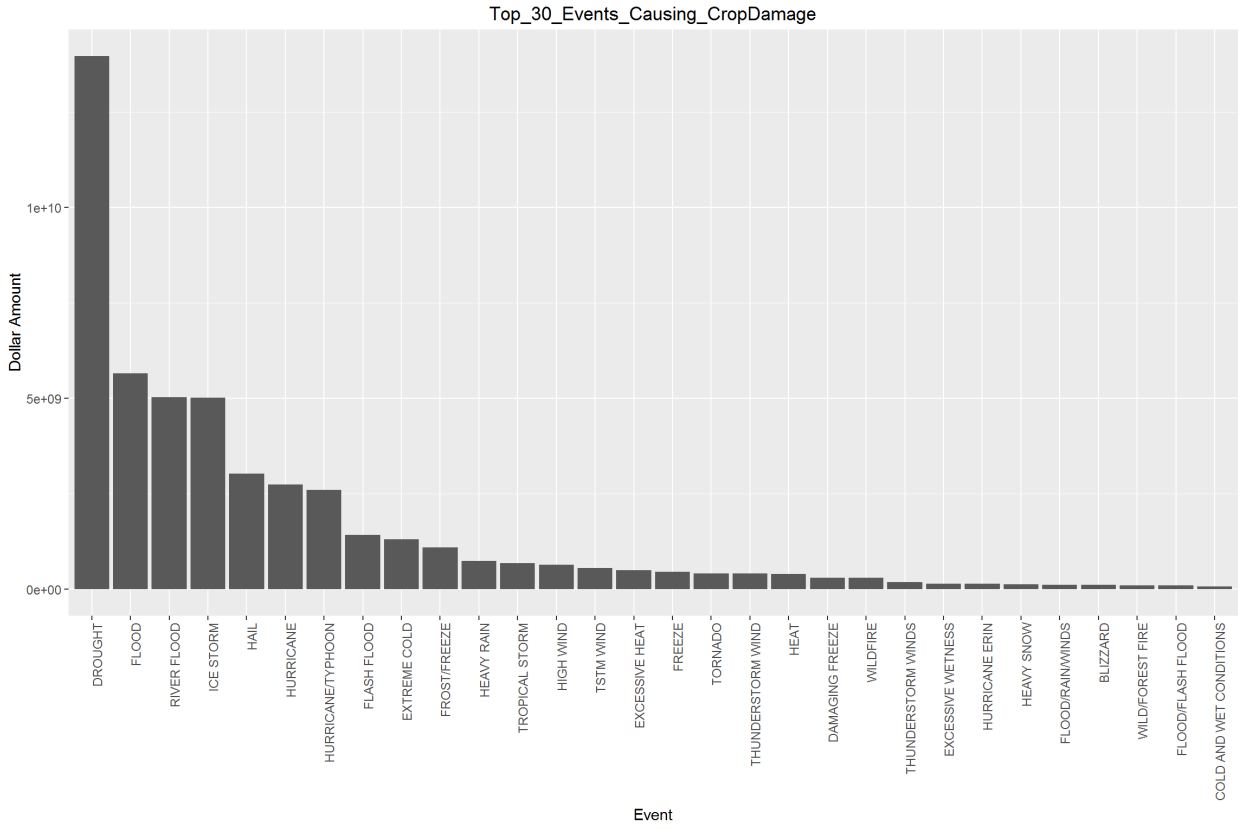
## 7.2 Events Causing CROP DAMAGE

```
## [1] ""  "?" "0" "2" "B" "K" "M"
```

```
## $EVTYPE
## [1] "character"
##
## $CROPDMG
## [1] "numeric"
##
## $CROPDMGEXP
## [1] "character"
```

```
cropnew_order<- cropnew_agg[order(cropnew_agg[,2], decreasing = TRUE),]
Top30_CropDmg <- cropnew_order[1:30,]
eve_counts <- Top30_CropDmg[,2]
eve_list <- c(Top30_CropDmg[,1])
eve_name <- Top30_CropDmg[,1]

g <- ggplot(Top30_CropDmg, aes(x = reorder(eve_list, -eve_counts), y = eve_counts)) +
        geom_bar(stat = "identity") +
        ggtitle("Top_30_Events_Causing_CropDamage") +
        labs(x="Event",y="Dollar Amount") +
        theme(axis.text.x = element_text(angle = 90, hjust = 1))
print(g)
```




```
png(file = "Top_30_Events_Causing_CropDamage.png", width=8,height=6, units = 'in', res = 300)

tb4 <- knitr::kable(Top30_CropDmg, digits = 5, caption = "Top_30_Events_Causing_CropDamage", col.names =
print(tb4)
```

```
##
##
## Table: Top_30_Events_Causing_CropDamage
##
##       Events                       CropDmg
## ----  ------------------------  ------------
## 84    DROUGHT                    13972566000
## 154   FLOOD                       5661968450
## 529   RIVER FLOOD                 5029459000
## 387   ICE STORM                   5022113500
## 212   HAIL                        3025954473
## 363   HURRICANE                   2741910000
## 372   HURRICANE/TYPHOON           2607872800
## 138   FLASH FLOOD                 1421317100
## 125   EXTREME COLD                1312973000
## 187   FROST/FREEZE                1094186000
## 254   HEAVY RAIN                   733399800
## 772   TROPICAL STORM               678346000
## 320   HIGH WIND                    638571300
## 779   TSTM WIND                    554007350
## 116   EXCESSIVE HEAT               492402000
```

```
## 174   FREEZE                     456725000
## 758   TORNADO                    414953270
## 685   THUNDERSTORM WIND          414843050
## 243   HEAT                       401461500
## 76    DAMAGING FREEZE            296230000
## 875   WILDFIRE                   295472800
## 711   THUNDERSTORM WINDS         190654788
## 122   EXCESSIVE WETNESS          142000000
## 367   HURRICANE ERIN             136010000
## 274   HEAVY SNOW                 134653100
## 165   FLOOD/RAIN/WINDS           112800000
## 28    BLIZZARD                   112060000
## 873   WILD/FOREST FIRE           106796830
## 160   FLOOD/FLASH FLOOD           95034000
## 63    COLD AND WET CONDITIONS     66000000
```

## 7.3 System Information

***Time required to process this report:*** *2.68012 mins*

***R session information:***

```
## R version 3.2.3 (2015-12-10)
## Platform: i386-w64-mingw32/i386 (32-bit)
## Running under: Windows 7 (build 7601) Service Pack 1
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] gridExtra_2.0.0 plyr_1.8.3      gplots_2.17.0   printr_0.0.5
## [5] reshape2_1.4.1  ggplot2_2.0.0   dplyr_0.4.3
##
## loaded via a namespace (and not attached):
##  [1] Rcpp_0.12.3        knitr_1.12.3       magrittr_1.5
##  [4] munsell_0.4.3      colorspace_1.2-6   R6_2.1.2
##  [7] highr_0.5.1        stringr_1.0.0      caTools_1.17.1
## [10] tools_3.2.3        parallel_3.2.3     grid_3.2.3
## [13] gtable_0.1.2       KernSmooth_2.23-15 DBI_0.3.1
## [16] htmltools_0.3      gtools_3.5.0       lazyeval_0.1.10
## [19] yaml_2.1.13        assertthat_0.1     digest_0.6.9
## [22] formatR_1.2.1      bitops_1.0-6       evaluate_0.8
## [25] rmarkdown_0.9.2    labeling_0.3       gdata_2.17.0
## [28] stringi_1.0-1      scales_0.3.0
```