Exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) SStorm database

Analysis of U.S. NOAA SStorm database

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, crop and property damage, and preventing such outcomes to the extent possible is a key concern.

This project tracks the type of events that are most harmful with respect to population health in terms of FATALITIES(Death of human beings) and Injuires. It also tracks the type of events that have the greatest economic consequences in dollar terms of crop and property damage.

Loading and Processing the raw data

Down the SStorm Data file at https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 and copy it in the R Studio working directory Download the SStorm Data file and Read in the data

```
# download.file(url="http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
# destfile="repdata-data-StormData.csv - Copy.bz2")

# Reading in the data using read.csv
SStorm <- read.csv("repdata-data-StormData.csv - Copy.bz2",stringsAsFactors=FALSE)
```

This data analysis addresses two questions about the dataset. 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? Subset the original dataset to satisfy population health(FATALITIES and INJURIES) and economic consequences(CROPDMG,CROPDMGEXP, PROPDMG, PROPDMGEXP) along with a few more self explanatory variables.

Variables subset: BGN DATE, BGN TIME, TIME_ZONE, COUNTYNAME, STATE, EVTYPE, LENGTH , WIDTH, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP

```
SStormSubset<- SStorm[ ,c("BGN_DATE","BGN_TIME","TIME_ZONE","COUNTYNAME","STATE","EVTYPE",
"LENGTH", "WIDTH", "FATALITIES","INJURIES", "PROPDMGEXP", "PROPDMG",
"CROPDMGEXP", "CROPDMG", "REFNUM")]
# Fix the outlier. Million(M) instead of Billion(B)
SStormSubset[SStormSubset$REFNUM == 605943, "PROPDMGEXP" ] <- "M"
```

Get the year component of the beginning date of the event(BGN_DATE) to enable us to display the popluation health and economic consequences on a yearly basis

```
# Get the Year of the Event in variable BGN_YEAR
SStormSubset$BGN_DATE1<-as.Date(as.character(SStormSubset$BGN_DATE), format="%m/%d/%Y",tz="")
SStormSubset$BGN_YEAR <- as.integer(substr(as.character(SStormSubset$BGN_DATE1), 1,4))
```

Determine the event types(variable EVTYPE) in the original dataset

```
length(unique(SStormSubset$EVTYPE))
```

[1] 985

From 1996 to present, 48 event types are recorded as defined in NWS Directive 10-1605. We must Consolidate the **985** unique events in the original dataset to one of the 48 possible values. However consolidating **985** event types to 48 is a very tedious task. Therefore we filter the data on the basis of non zero values in FATALITIES, INJURIES, CROPDMG or PROPDMG. This significantly reduces the unique EVENT types.

Create a new variable, EVTYPE1 to hold the consolidated EVENT type after converting to upper case

```
SStormSubset$EVTYPE <- toupper(SStormSubset$EVTYPE)
SStormSubset$EVTYPE1 <- SStormSubset$EVTYPE
```

Process to answer question 1: Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

Eliminate the observations in the dataset with zero(0) FATALITIES and INJURIES

```
SStormSubsetHealth <- SStormSubset[SStormSubset$FATALITIES != 0 & SStormSubset$INJURIES != 0,]
```

Determine the unique EVENT types(variable EVTYPE) in the health subset dataset, SStormSubsetHealth.

```
length(unique(SStormSubsetHealth$EVTYPE))
```

[1] 84

We must approximately consolidate the **84** unique events in the data subset to one of the 48 possible values.

Consolidation of EVENT types for the population health dataset Write the unique types of EVENTS to a csv file,events_health.csv for easy readability

```
EVENT_HEALTH <- data.frame(table(SStormSubsetHealth$EVTYPE1))
write.csv(EVENT_HEALTH, file="events_health.csv", row.names=FALSE)
```

Consolidation Methodology The data in terms of types of EVENT(EVTYPE) is not normalized in the TRUEST sense of the word. There are also multiple types of EVENTS that do not seem to have a match. However these EVENTS are very small in number and therefore can be ignored. No single routine can fix The consolidation of the types of EVENT. It is a manual process using the following methodology: 1. Search for a particular pattern(eg EXTREME) in EVTYPE1 for the EVENT "EXTREME COLD/WIND CHILL" using:

```
# pattern=".*EXTREME*." to match the EVENT "EXTREME COLD/WIND CHILL"
# Use the command below without the #.
# unique(grep( pattern, SStormSubsetProp$EVTYPE1, ignore.case = FALSE, value=TRUE))
```

This gives you the result of unique types of EVENT for that pattern. 2. If all the EVENTs in the result look like they match the event, "EXTREME COLD/WIND CHILL" then replace the EVTYPE1 values in the health dataset, SStormSubsetHealth using:

```
# pattern=".*EXTREME*." to match the EVENT "EXTREME COLD/WIND CHILL"
# Use the command below without the #.
# SStormSubsetHealth[grep(pattern, SStormSubsetHealth$EVTYPE1, ignore.case = FALSE), "EVTYPE1"]
#<- "EXTREME COLD/WIND CHILL"
```

Else note the EVENTS that are not a match and repeat step 1 again for each one of these unmatched EVENTS. Resolve the all unmatched EVENTS before coming back to the original EVENT(eg EXTREME COLD/WIND CHILL) in this example for replacement.

```
SStormSubsetHealth[grep(".*EXTREME*.", SStormSubsetHealth$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"
SStormSubsetHealth[grep(".*WIND CHILL*.", SStormSubsetHealth$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"

SStormSubsetHealth[grep("MARINE THUNDERSTORM.*WIND*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MTW"
SStormSubsetHealth[grep("THUNDER", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "THUNDERSTORM WIND"

SStormSubsetHealth[grep(".*MARINE TSTM WIND*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MTW"
SStormSubsetHealth[grep(".*TSTM*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "THUNDERSTORM WIND"
SStormSubsetHealth[grep("MTW", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MARINE THUNDERSTORM WIND"

SStormSubsetHealth[grep(".*SURF*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HIGH SURF"
SStormSubsetHealth[grep(".*HIGH SEAS*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HIGH SURF"

SStormSubsetHealth[grep(".*FLASH FLOOD*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "FF"
SStormSubsetHealth[grep(".*FLOOD*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "FLOOD"
SStormSubsetHealth[grep(".*FF*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "FLASH FLOOD"

SStormSubsetHealth[grep(".*MARINE STRONG WIND*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MSW"
SStormSubsetHealth[grep(".*STRONG*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "STRONG WIND"
SStormSubsetHealth[grep(".*GUSTY*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "STRONG WIND"
SStormSubsetHealth[grep("MSW", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MARINE STRONG WIND"

SStormSubsetHealth[grep(".*WINTER SStorm*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "WINTER SStorm"

SStormSubsetHealth[grep(".*MARINE HIGH WIND*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MHW"
SStormSubsetHealth[grep(".*HIGH WIND*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HIGH WIND"
SStormSubsetHealth[grep("MHW", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "MARINE HIGH WIND"

SStormSubsetHealth[grep(".*WINTER WEATHER*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "WINTER WEATHER"

SStormSubsetHealth[grep(".*HURRICANE*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HURRICANE/TYPHOON"

SStormSubsetHealth[grep(".*TORNADO*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "TORNADO"

SStormSubsetHealth[grep(".*TROPICAL*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "TROPICAL SStorm"

SStormSubsetHealth[grep(".*RIP CURRENT*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "RIP CURRENT"

SStormSubsetHealth[grep(".*SNOW*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HEAVY SNOW"

SStormSubsetHealth[grep(".*HEAT WAVE*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "HEAT"

SStormSubsetHealth[grep(".*ICY ROADS*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "ICE SStorm"
SStormSubsetHealth[grep(".*BLACK ICE*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "ICE SStorm"
SStormSubsetHealth[grep("ICE", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "ICE SStorm"

SStormSubsetHealth[grep(".*LANDSLIDE*.", SStormSubsetHealth$EVTYPE1,
ignore.case = FALSE), "EVTYPE1"] <- "DEBRIS FLOW"
```

Install and load the dplyr, knitr package to utilize grouping of data

```
#if (!require("dplyr")) {
#suppressWarnings(suppressMessages(install.packages("dplyr")))
#}
if (!require("knitr")) {
suppressWarnings(suppressMessages(install.packages("knitr")))
}
```

```
## Loading required package: knitr
```

```
## Warning: package 'knitr' was built under R version 3.1.3
```

Grouping and Summarizing the population health data Group the SStorm data set by EVENT type(EVTYPE1), summarize the values for totals of FATALITIES and INJURIES and then sort by highest FATALITIES first followed by highest INJURIES.

```r
SStormSubsetHealthFbyE_1 <- with(SStormSubsetHealth, aggregate(FATALITIES,by=list(EVENT=EVTYPE1), sum))
colnames(SStormSubsetHealthFbyE_1) <- c("EVENT", "FATALITIES")

SStormSubsetHealthIbyE_1 <- with(SStormSubsetHealth, aggregate(INJURIES,by=list(EVENT=EVTYPE1), sum))
colnames(SStormSubsetHealthIbyE_1) <- c("EVENT", "INJURIES")

SStormSubsetHealth3 <- merge(SStormSubsetHealthFbyE_1, SStormSubsetHealthIbyE_1, by.x = "EVENT", by.y =
SStormSubsetHealth4 <- SStormSubsetHealth3[order(-SStormSubsetHealth3$FATALITIES),]
SStormSubsetHealth5 <- SStormSubsetHealth3[order(-SStormSubsetHealth3$INJURIES),]

colnames(SStormSubsetHealth4) <- c("EVENT", "FATALITIES", "INJURIES")
colnames(SStormSubsetHealth5) <- c("EVENT", "FATALITIES", "INJURIES")
```

Group the SStorm data set by STATE and EVENT type(EVTYPE1), summarize the values for totals of FATALITIES and INJURIES and then sort by highest FATALITIES first followed by highest INJURIES.

The top 10 SStorm Data Events that are most harmful with respect to population health in terms of FATALITIES is are as shown below.

```r
kable(head(SStormSubsetHealth4,10))
```

	EVENT	FATALITIES	INJURIES
30	TORNADO	5230	60226
9	FLASH FLOOD	414	4293
7	EXCESSIVE HEAT	402	4791
19	LIGHTNING	283	649
29	THUNDERSTORM WIND	282	833
16	HIGH WIND	121	390
12	HEAT	99	1704
34	WINTER STORM	85	599
26	RIP CURRENT	82	154
14	HEAVY SNOW	60	250

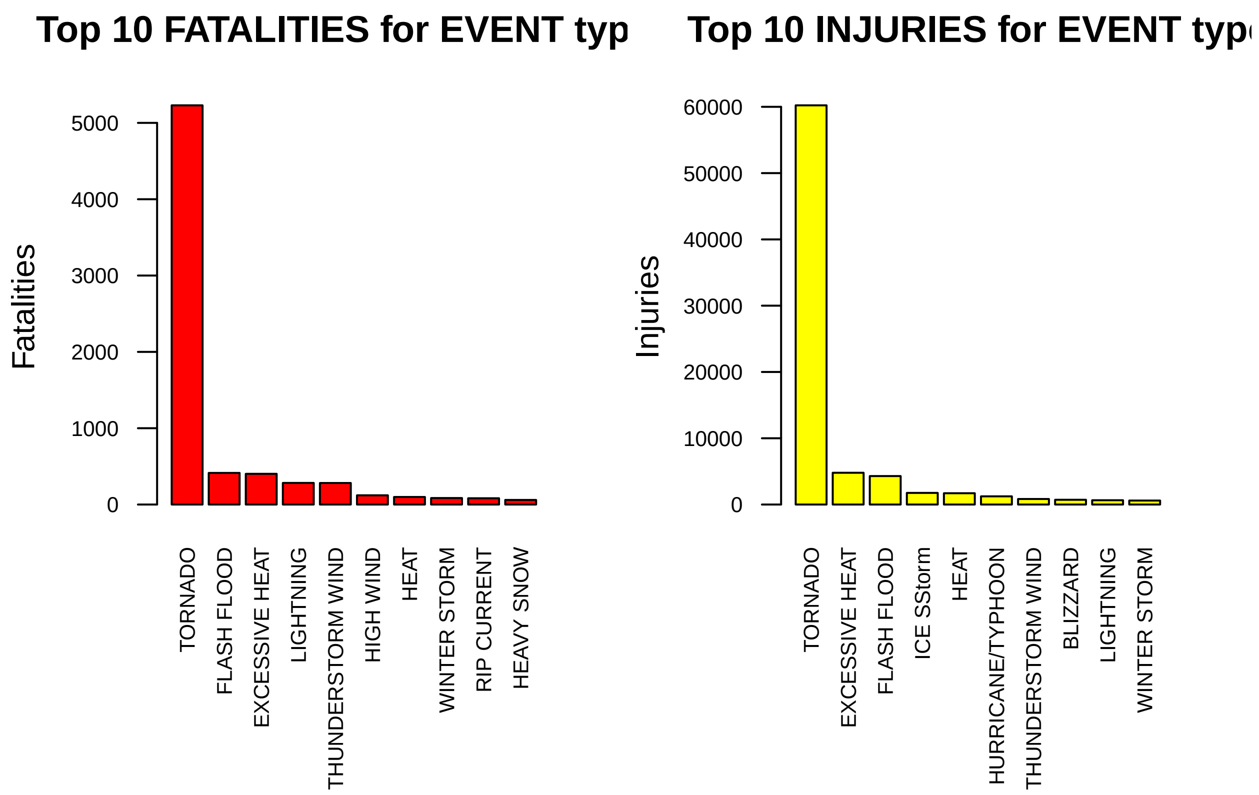
The top 10 SStorm Data Events that are most harmful with respect to population health in terms of INJURIES is are as shown below.

```r
kable(head(SStormSubsetHealth5,10))
```

	EVENT	FATALITIES	INJURIES
30	TORNADO	5230	60226
7	EXCESSIVE HEAT	402	4791
9	FLASH FLOOD	414	4293
18	ICE SStorm	44	1755
12	HEAT	99	1704
17	HURRICANE/TYPHOON	46	1237
29	THUNDERSTORM WIND	282	833
2	BLIZZARD	48	718
19	LIGHTNING	283	649
34	WINTER STORM	85	599

```
par(mfrow = c(1, 2), mar=c(8, 4, 4, 2), las = 2, cex.axis=0.65)
with(SStormSubsetHealth4,
{
    fat <- as.table(head(SStormSubsetHealth4[,c("FATALITIES")],10))
    dimnames(fat) <- list(head(SStormSubsetHealth4[,c("EVENT")],10))
    barplot(fat, main="Top 10 FATALITIES for EVENT types", col="red", ylab="Fatalities")

    inj <- as.table(head(SStormSubsetHealth5[,c("INJURIES")],10))
    dimnames(inj) <- list(head(SStormSubsetHealth5[,c("EVENT")],10))
    barplot(inj, main="Top 10 INJURIES for EVENT type", col="yellow", , ylab="Injuries")
})
```



Process to answer question 2: Across the United States, which types of events have the greatest economic consequences? This will be broken down into two parts. The first is Crop damage and second is Property damage.

Eliminate the observations in the dataset with zero(0) CROP DAMAGE(CROPDMG) or PROPERTY DAMAGE (PROPDMG) to create a broad economic subset(SStormSubsetEconomic) of data with non zero values. This subset, SStormSubsetEconomic will be processed separately for Crop damage and Property damage.

```
SStormSubsetEconomic <- SStormSubset[(SStormSubset$CROPDMG != 0 | SStormSubset$PROPDMG != 0),]
```

The Crop damage cost is estimated by multiplying the variable ending "exp"(CROPDMGEXP and PROPDMGEXP) by their respective damage count variables(CROPDMG and PROPDMG). The valid values for the exp variables K(thousand = 1000), M(million = 1000000) and B(billion = 1000000000).

Crop Damage data consolidation Create a new dataset for Crop damage clean up from the economic dataset

```
SStormSubsetCrop <- SStormSubsetEconomic
```

Get the distinct values for CROPDMGEXP in the dataset

```
cexp <- data.frame(table(SStormSubsetCrop$CROPDMGEXP))
colnames(cexp) <- c("CROPDMGEXP", "Freq")
kable(cexp)
```

CROPDMGEXP	Freq
	145037
?	6
0	17
B	7
k	21
K	97960
m	1
M	1982

There are CROPDMGEXP values of k and m. Convert these to K and M respectively. Eliminate observations with values 0, ? and “”(blank)

```
SStormSubsetCrop[SStormSubsetCrop$CROPDMGEXP=="k", "CROPDMGEXP"] <- "K"
SStormSubsetCrop[SStormSubsetCrop$CROPDMGEXP=="m", "CROPDMGEXP"] <- "M"
```

Subset only observations with CROPDMGEXP values of K, M and B

```
SStormSubsetCrop <- SStormSubsetCrop[(SStormSubsetCrop$CROPDMGEXP == "K") |
(SStormSubsetCrop$CROPDMGEXP == "M") | (SStormSubsetCrop$CROPDMGEXP == "B"),]
```

Calculate actual Crop damage cost by multiplying CROPDMG by CROPDMGEXP. Valid values for CROPDMGEXP are K(thousand = 1000), M(million = 1000000) and B(billion = 1000000000). Store the Crop damage cost in variable, CROPDMGCOST

```
SStormSubsetCrop <- cbind(SStormSubsetCrop,CROPDMGCOST =
ifelse((SStormSubsetCrop$CROPDMGEXP=="B"),1000000000 * SStormSubsetCrop$CROPDMG,0))

SStormSubsetCrop$CROPDMGCOST <- ifelse((SStormSubsetCrop$CROPDMGEXP == "M"),
1000000 * SStormSubsetCrop$CROPDMG,  SStormSubsetCrop$CROPDMGCOST)

SStormSubsetCrop$CROPDMGCOST <- ifelse((SStormSubsetCrop$CROPDMGEXP == "K"),
1000 * SStormSubsetCrop$CROPDMG, SStormSubsetCrop$CROPDMGCOST)
```

Write the unique types of EVENTS to a csv file,events_crop.csv for easy readability

```
EVENT_CROP <- data.frame(table(SStormSubsetCrop$EVTYPE1))
write.csv(EVENT_CROP, file="events_crop.csv", row.names=FALSE)
```

Consolidate the **152** EVENT type to 48 possible values in variable EVTYPE1. Look above for Consolidation Methodology in this document.

```
SStormSubsetCrop[grep(".*EXTREME*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"

SStormSubsetCrop[grep(".*THUNDER*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"
SStormSubsetCrop[grep("THUDERSTORM WINDS", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"

SStormSubsetCrop[grep(".*TORNADO*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "TORNADO"

SStormSubsetCrop[grep(".*TSTM*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"

SStormSubsetCrop[grep(".*SStorm SURGE*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "SStorm SURGE/TIDE"

SStormSubsetCrop[grep(".*HEAVY RAIN/HIGH SURF*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH SURF"

SStormSubsetCrop[grep(".*HEAVY SNOW*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAVY SNOW"

SStormSubsetCrop[grep(".*HURRICANE*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HURRICANE/TYPHOON"
SStormSubsetCrop[grep(".*TYPHOON*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HURRICANE/TYPHOON"

SStormSubsetCrop[grep(".*WINTER*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WINTER SStorm"

SStormSubsetCrop[grep(".*DUST*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "DUST SStorm"

SStormSubsetCrop[grep(".*HIGH WIND*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH WIND"

SStormSubsetCrop[grep(".*HEAVY RAIN*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAVY RAIN"

SStormSubsetCrop[grep(".*RIVER FLOOD*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetCrop[grep(".*COASTAL FLOODING*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetCrop[grep(".*COASTAL FLOOD*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "CF"
```

```
SStormSubsetCrop[grep(".*FLASH FLOOD*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FF"
SStormSubsetCrop[grep(".*FLOOD*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLOOD"

SStormSubsetCrop[grep("CF", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetCrop[grep("FF", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLASH FLOOD"

SStormSubsetCrop[grep(".*HAIL*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HAIL"

SStormSubsetCrop[grep(".*TROPICAL SStorm*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "TROPICAL SStorm"

SStormSubsetCrop[grep(".*MARINE STRONG WIND*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MSW"
SStormSubsetCrop[grep(".*STRONG WIND*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "STRONG WIND"
SStormSubsetCrop[grep(".*MSW*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE STRONG WIND"

SStormSubsetCrop[grep(".*GUST*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "STRONG WIND"

SStormSubsetCrop[grep(".*HEAT WAVE*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAT"

SStormSubsetCrop[grep(".*WILD*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WF"
SStormSubsetCrop[grep(".*FOREST*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WILDFIRE"
SStormSubsetCrop[grep("WF", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WILDFIRE"

SStormSubsetCrop[grep(".*FREEZE*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FROST/FREEZE"

SStormSubsetCrop[grep(".*LANDSLIDE*.", SStormSubsetCrop$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "DEBRIS FLOW"
```

Group the Crop damage data set by EVENT type(EVTYPE1), summarize the values for the totals of the Crop Damage cost(CROPDMGCOST) and then sort by the highest Crop Damage cost(CROPDMGCOST) first.

```
SStormSubsetCrop1 <- with(SStormSubsetCrop, aggregate(CROPDMGCOST,by=list(EVENT=EVTYPE1), sum))
colnames(SStormSubsetCrop1) <- c("EVENT", "CROPDMGCOST")

SStormSubsetCrop2 <- SStormSubsetCrop1[order(-SStormSubsetCrop1$CROPDMGCOST),]

SStormSubsetCrop2$CROP_DAMAGE_COST <- ifelse(SStormSubsetCrop2$CROPDMGCOST > 1000000000,
paste(signif(SStormSubsetCrop2$CROPDMGCOST/1000000000, digits=5),"B", sep=" "),
SStormSubsetCrop2$CROPDMGCOST)
colnames(SStormSubsetCrop2) <- c("EVENT","CROPDMGCOST", "CROP DAMAGE COST")
SStormSubsetCrop2$EVENT <- factor(SStormSubsetCrop2$EVENT)
```

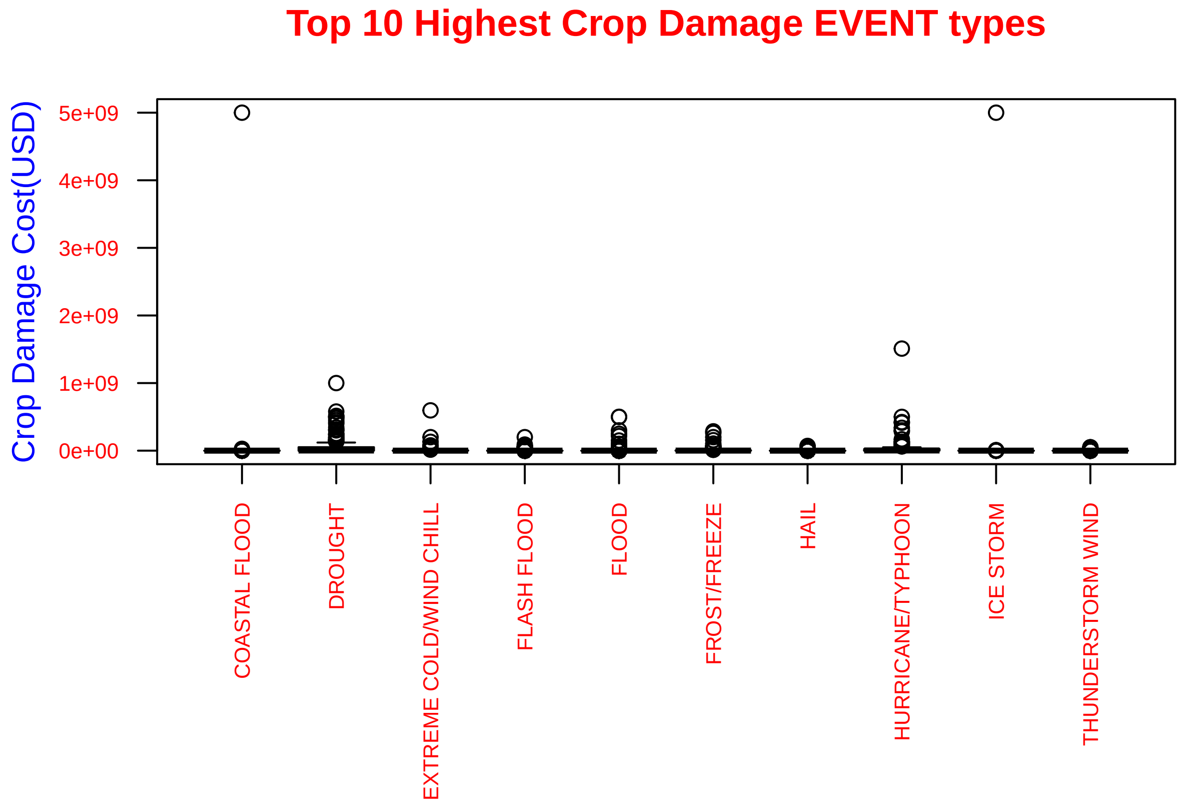
The Top 10 SStorm Data Events with greatest economic consequences being Crop damage are as shown below in tabular form.

```
kable(head(SStormSubsetCrop2[, c("EVENT","CROP DAMAGE COST")],10))
```

	EVENT	CROP DAMAGE COST
12	DROUGHT	13.973 B
21	FLOOD	5.79 B
33	HURRICANE/TYPHOON	5.5161 B
5	COASTAL FLOOD	5.0575 B
35	ICE STORM	5.0221 B
27	HAIL	3.0469 B
24	FROST/FREEZE	1.8891 B
20	FLASH FLOOD	1.5322 B
19	EXTREME COLD/WIND CHILL	1.335 B
47	THUNDERSTORM WIND	1.2717 B

Boxplot for Top 10 Storm Data Events with greatest economic consequences being Crop damage.

```
crop10 <- data.frame(head(SStormSubsetCrop2[, "EVENT"],10))
colnames(crop10)<- "EVENT"
par(mfrow=c(1,1), mar=c(9, 4, 4, 2),las = 2, cex.axis=0.65)
boxplot(CROPDMGCOST ~ EVTYPE1, data=subset(SStormSubsetCrop,
subset=SStormSubsetCrop$EVTYPE1 %in% crop10$EVENT), plot=TRUE,
main="Top 10 Highest Crop Damage EVENT types", ylab="Crop Damage Cost(USD)",
col.axis = "red", col.main="red", col.lab = "blue")
```



Property Damage data consolidation Create a new dataset for Property damage clean up from the economic dataset

```
SStormSubsetProp <- SStormSubsetEconomic
```

Get the distinct values for CROPDMGEXP in the dataset

```
pexp <- data.frame(table(SStormSubsetProp$PROPDMGEXP))
colnames(pexp) <- c("PROPDMGEXP", "Freq")
kable(pexp)
```

PROPDMGEXP	Freq
	4357
-	1
+	5
0	209
2	1
3	1
4	4
5	18
6	3
7	2
B	39

PROPDMGEXP	Freq
h	1
H	6
K	229057
m	7
M	11320

There are PROPDMGEXP values of m. Convert m to M respectively. Eliminate observations with values 0, ? and “”(blank)

```
SStormSubsetProp[SStormSubsetProp$PROPDMGEXP=="m", "PROPDMGEXP"] <- "M"
```

Subset only observations with PROPDMGEXP values of K, M and B

```
SStormSubsetProp <- SStormSubsetProp[(SStormSubsetProp$PROPDMGEXP == "K") |
(SStormSubsetProp$PROPDMGEXP == "M") | (SStormSubsetProp$PROPDMGEXP == "B"),]
```

Calculate actual Property damage cost by multiplying PROPDMG by PROPDMGEXP. Values for PROPDMGEXP are K(thousand = 1000), M(million = 1000000) and B(billion = 1000000000). Store the Property damage cost in variable, PROPDMGCOST

```
SStormSubsetProp <- cbind(SStormSubsetProp,PROPDMGCOST =
ifelse((SStormSubsetProp$PROPDMGEXP == "B"), 1000000000 * SStormSubsetProp$PROPDMG, 0))

SStormSubsetProp$PROPDMGCOST <- ifelse((SStormSubsetProp$PROPDMGEXP == "M"),
1000000 * SStormSubsetProp$PROPDMG,  SStormSubsetProp$PROPDMGCOST)

SStormSubsetProp$PROPDMGCOST <- ifelse((SStormSubsetProp$PROPDMGEXP == "K"),
1000 * SStormSubsetProp$PROPDMG, SStormSubsetProp$PROPDMGCOST)
```

Write the unique types of EVENTS to a csv file,events_prop.csv for easy readability

```
EVENT_PROP <- data.frame(table(SStormSubsetProp$EVTYPE1))
write.csv(EVENT_PROP, file="events_prop.csv", row.names=FALSE)
```

Consolidate the **370** EVENT types to 48 possible values as defined in NWS Directive 10-1605 in variable EVTYPE1

```
SStormSubsetProp[grep("MARINE THUNDERSTORM WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MTW"
SStormSubsetProp[grep(".*THUN*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"
SStormSubsetProp[grep("THUDERSTORM WINDS", SStormSubsetProp$EVTYPE1, ignore.case = FALSE), "EVTYPE1"] <-
SStormSubsetProp[grep("TUNDERSTORM WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"

SStormSubsetProp[grep("MARINE TSTM WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MTW"

SStormSubsetProp[grep(".*DUST DEVIL*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "DUST DEVIL"
SStormSubsetProp[grep(".*WATERSPOUT*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WATERSPOUT"

SStormSubsetProp[grep(".*TORNADO*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "TORNADO"
SStormSubsetProp[grep("TORNDAO", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "TORNADO"

SStormSubsetProp[grep("NON-TSTM WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH WIND"

SStormSubsetProp[grep(".*TSTM*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "THUNDERSTORM WIND"

SStormSubsetProp[grep("MTW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE THUNDERSTORM WIND"

SStormSubsetProp[grep(".*HURRICANE*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HURRICANE/TYPHOON"
SStormSubsetProp[grep(".*TYPHOON*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HURRICANE/TYPHOON"

SStormSubsetProp[grep(".*HEAVY SNOW*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAVY SNOW"

SStormSubsetProp[grep(".*TROPICAL SStorm*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "TROPICAL SStorm"

SStormSubsetProp[grep("BLIZZARD/WINTER SStorm", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "BLIZZARD"

SStormSubsetProp[grep(".*WINTER SStorm*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WINTER SStorm"

SStormSubsetProp[grep(".*WINTER WEATHER*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WINTER WEATHER"

SStormSubsetProp[grep(".*WILD*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WILDFIRE"
SStormSubsetProp[grep(".*FOREST*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WILDFIRE"

SStormSubsetProp[grep("HEAVY SURF", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH SURF"

SStormSubsetProp[grep(".*DUST SStorm*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "DUST SStorm"

SStormSubsetProp[grep("SNOW/HIGH WINDS", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH WIND"

SStormSubsetProp[grep("MARINE HIGH WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MHW"

SStormSubsetProp[grep(".*HIGH WIND*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HIGH WIND"

SStormSubsetProp[grep("MHW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE HIGH WIND"

SStormSubsetProp[grep(".*EFFECT*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LAKE-EFFECT SNOW"

SStormSubsetProp[grep("FLASH FLOOD - HEAVY RAIN", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLASH FLOOD"

SStormSubsetProp[grep("FLOOD & HEAVY RAIN", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLOOD"

SStormSubsetProp[grep(".*LIGHTNING*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LIGHTNING"
SStormSubsetProp[grep("LIGNTNING", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LIGHTNING"
SStormSubsetProp[grep("LIGHTING", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LIGHTNING"

SStormSubsetProp[grep(".*HEAVY RAIN*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAVY RAIN"

SStormSubsetProp[grep("LAKE-EFFECT SNOW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LES"

SStormSubsetProp[grep("SNOW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HEAVY SNOW"

SStormSubsetProp[grep("LES", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LAKE-EFFECT SNOW"

SStormSubsetProp[grep("MARINE HAIL", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MH"
SStormSubsetProp[grep("MARINE STRONG WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MSW"
SStormSubsetProp[grep("MARINE THUNDERSTORM WIND", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MTW"

SStormSubsetProp[grep(".*GUSTY*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "STRONG WIND"

SStormSubsetProp[grep(".*HAIL*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "HAIL"

SStormSubsetProp[grep(".*STRONG WIND*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "STRONG WIND"

SStormSubsetProp[grep("MH", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE HAIL"
SStormSubsetProp[grep("MSW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE STRONG WIND"
SStormSubsetProp[grep("MTW", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "MARINE THUNDERSTORM WIND"

SStormSubsetProp[grep(".*COASTAL*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"

SStormSubsetProp[grep(".*COASTAL*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"

SStormSubsetProp[grep(".*EXTREME HEAT*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXCESSIVE HEAT"

SStormSubsetProp[grep(".*EXTREME*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"

SStormSubsetProp[grep("EXTENDED COLD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"
SStormSubsetProp[grep("RECORD COLD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"

SStormSubsetProp[grep("FLASH FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FF"
SStormSubsetProp[grep("FLOOD FLASH", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FF"

SStormSubsetProp[grep("FLOOD/RIVER FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLOOD"

SStormSubsetProp[grep(".*RIVER FLOOD*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetProp[grep("EROSION/CSTL FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetProp[grep("TIDAL FLOODING", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetProp[grep("RIVER AND STREAM FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"

SStormSubsetProp[grep("COASTAL FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "CF"
SStormSubsetProp[grep("EROSION/CSTL FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"

SStormSubsetProp[grep("LAKE FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LAKESHORE FLOOD"
SStormSubsetProp[grep("LAKESHORE FLOOD", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LF"

SStormSubsetProp[grep(".*FLOOD*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLOOD"
```

```
SStormSubsetProp[grep("CF", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"
SStormSubsetProp[grep("LF", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "LAKESHORE FLOOD"
SStormSubsetProp[grep("FF", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "FLASH FLOOD"

SStormSubsetProp[grep("EXTREME COLD/WIND CHILL", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "ECWC"

SStormSubsetProp[grep(".*COLD*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COLD/WIND CHILL"
SStormSubsetProp[grep("ECWC", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "EXTREME COLD/WIND CHILL"

SStormSubsetProp[grep(".*SStorm SURGE*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "SStorm SURGE/TIDE"

SStormSubsetProp[grep(".*LANDSLIDE*.", SStormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "DEBRIS FLOW"
```

Group the Property damage data set by EVENT type(EVTYPE1), summarize the values for Property Damage cost(PROPDMGCOST) and then sort by highest PROPDMGCOST first.

```
SStormSubsetProp1 <- with(SStormSubsetProp, aggregate(PROPDMGCOST,by=list(EVENT=EVTYPE1), sum))
colnames(SStormSubsetProp1) <- c("EVENT", "PROPDMGCOST")

SStormSubsetProp2 <- SStormSubsetProp1[order(-SStormSubsetProp1$PROPDMGCOST),]

SStormSubsetProp2$PROPERTY_DAMAGE_COST <- ifelse(SStormSubsetProp2$PROPDMGCOST > 1000000000,
paste(signif(SStormSubsetProp2$PROPDMGCOST/1000000000, digits=5),"B", sep=" "), SStormSubsetProp2$PROPD
colnames(SStormSubsetProp2) <- c("EVENT","PROPDMGCOST","PROPERTY DAMAGE COST")
```

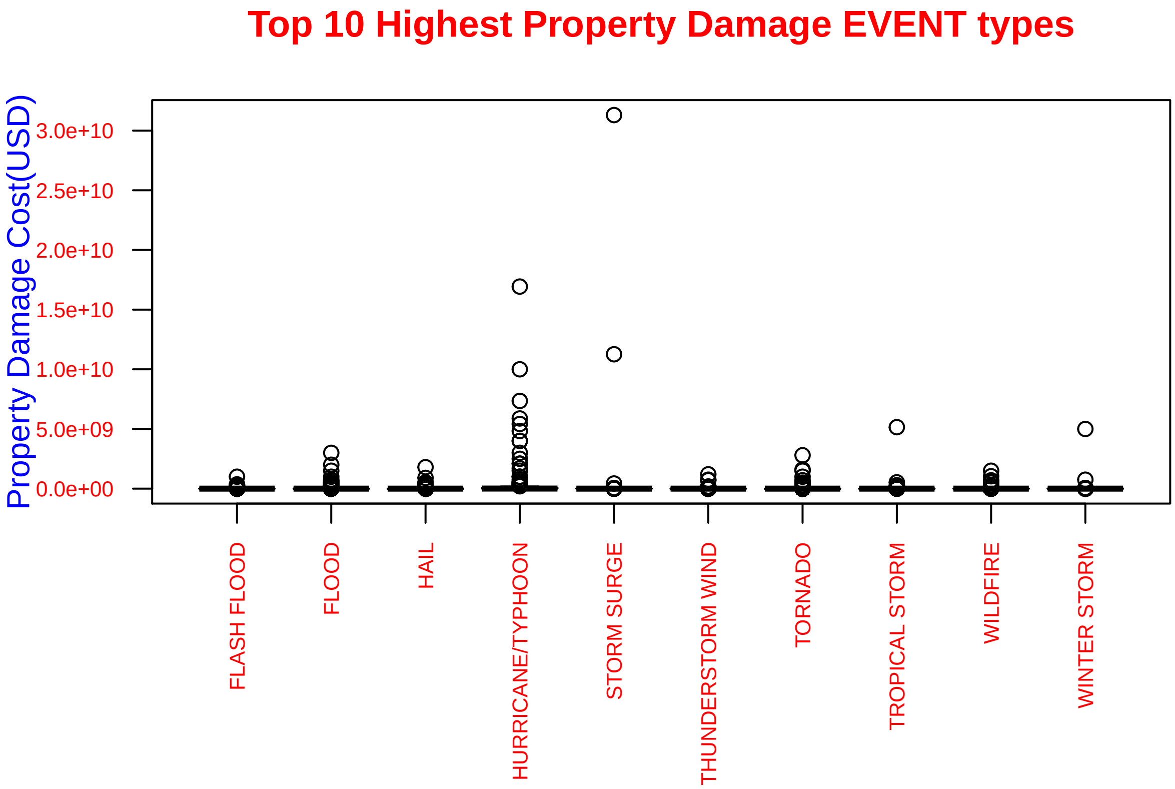
The top 10 SStorm Data Events with greatest economic consequences being Property damage are as shown below.

```
kable(head(SStormSubsetProp2[, c("EVENT","PROPERTY DAMAGE COST")],10))
```

	EVENT	PROPERTY DAMAGE COST
59	HURRICANE/TYPHOON	85.356 B
100	TORNADO	58.542 B
96	STORM SURGE	43.324 B
27	FLOOD	30.033 B
26	FLASH FLOOD	16.947 B
43	HAIL	15.975 B
99	THUNDERSTORM WIND	10.971 B
116	WILDFIRE	8.4966 B
102	TROPICAL STORM	7.7039 B
122	WINTER STORM	6.6885 B

Boxplot for Top 10 Storm Data Events with greatest economic consequences being Property damage.

```
prop10 <- data.frame(head(SStormSubsetProp2[, "EVENT"],10))
colnames(prop10)<- "EVENT"
par(mfrow=c(1,1), mar=c(8, 4, 4, 2), las = 2, cex.axis=0.65)
boxplot(PROPDMGCOST ~ EVTYPE1, data=subset(SStormSubsetProp,
subset=SStormSubsetProp$EVTYPE1 %in% prop10$EVENT), plot=TRUE,
main="Top 10 Highest Property Damage EVENT types", ylab="Property Damage Cost(USD)",
col.axis = "red", col.main="red", col.lab = "blue")
```


Top 10 Highest Property Damage EVENT types
Property Damage Cost(USD)
3.0e+10
2.5e+10
2.0e+10
1.5e+10
1.0e+10
5.0e+09
0.0e+00
FLASH FLOOD
FLOOD
HAIL
HURRICANE/TYPHOON
STORM SURGE
THUNDERSTORM WIND
TORNADO
TROPICAL STORM
WILDFIRE
WINTER STORM