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

Analysis of U.S. NOAA Storm 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 Storm Data file at https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 and copy it in the R Studio working directory Down the Storm Data file and Read in the data

```
# setInternet2(use = TRUE)
download.file(url="http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
destfile="repdata-data-StormData.csv.bz2")

# Reading in the data using read.csv
Storm <- read.csv("repdata-data-StormData.csv.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

```
StormSubset<- Storm[ ,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)
StormSubset[StormSubset$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
StormSubset$BGN_DATE1<-as.Date(as.character(StormSubset$BGN_DATE), format="%m/%d/%Y",tz="")
StormSubset$BGN_YEAR <- as.integer(substr(as.character(StormSubset$BGN_DATE1), 1,4))
```

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

```
length(unique(StormSubset$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

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

```
StormSubsetHealth <- StormSubset[StormSubset$FATALITIES != 0 & StormSubset$INJURIES != 0,]
```

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

```
length(unique(StormSubsetHealth$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(StormSubsetHealth$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, StormSubsetProp$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, StormSubsetHealth using:

```
# pattern=".*EXTREME*." to match the EVENT "EXTREME COLD/WIND CHILL"
# Use the command below without the #.
# StormSubsetHealth[grep(pattern, StormSubsetHealth$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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

StormSubsetHealth[grep(".*LANDSLIDE*.", StormSubsetHealth$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")))
}
```

```
## Loading required package: dplyr
```

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

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
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 storm 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.

```
StormSubsetHealth <- group_by(StormSubsetHealth, EVTYPE1)
StormSubsetHealth1 <- summarise(StormSubsetHealth, FATALITIES=sum(FATALITIES),
INJURIES=sum(INJURIES))
StormSubsetHealth1 <- arrange(StormSubsetHealth1, desc(FATALITIES), desc(INJURIES))
colnames(StormSubsetHealth1) <- c("EVENT", "FATALITIES", "INJURIES")
```

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

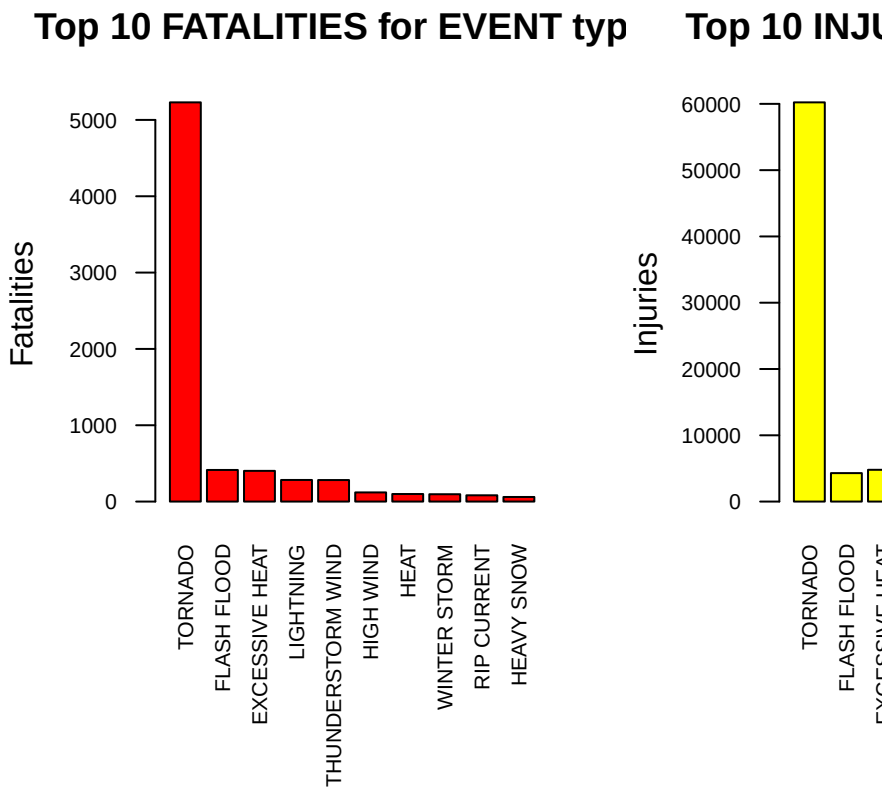
```
kable(head(StormSubsetHealth1,10))
```

EVENT	FATALITIES	INJURIES
TORNADO	5230	60226
FLASH FLOOD	414	4293
EXCESSIVE HEAT	402	4791
LIGHTNING	283	649
THUNDERSTORM WIND	282	833
HIGH WIND	120	375
HEAT	99	1704
WINTER STORM	96	631
RIP CURRENT	82	154
HEAVY SNOW	60	250

```
par(las = 2, mar=c(8, 4, 4, 2), cex.axis=0.65, mfrow = c(1, 2))
with(StormSubsetHealth1,
{
    fat <- data.frame(head(StormSubsetHealth1[,c("FATALITIES")],10))
    fat1 <- as.table(fat$FATALITIES)
    dimnames(fat1) <- as.vector(head(StormSubsetHealth1[,c("EVENT")],10))
    barplot(fat1, main="Top 10 FATALITIES for EVENT types", col="red", ylab="Fatalities")
```

```
    inj <- data.frame(head(StormSubsetHealth1[,c("INJURIES")],10))
    inj1 <- as.table(inj$INJURIES)
    dimnames(inj1) <- as.vector(head(StormSubsetHealth1[,c("EVENT")],10))
    barplot(inj1, main="Top 10 INJURIES for EVENT type", col="yellow", , ylab="Injuries")
})
```

Barplot for top 10 Storm Data Events that are most harmful with respect to population health



in terms of FATALITIES and 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(StormSubsetEconomic) of data with non zero values. This subset, StormSubsetEconomic will be processed separately for Crop damage and Property damage.

```
StormSubsetEconomic <- StormSubset[(StormSubset$CROPDMG != 0 | StormSubset$PROPDMG != 0),]
```

The Crop damage cost is estimated by multiplying the variable ending "exp"(CROPDMGEXP and PROPDMG-EXP) 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

```
StormSubsetCrop <- StormSubsetEconomic
```

Get the distinct values for CROPDMGEXP in the dataset

```
cexp <- data.frame(table(StormSubsetCrop$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)

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

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

```
StormSubsetCrop <- StormSubsetCrop[(StormSubsetCrop$CROPDMGEXP == "K") |
(StormSubsetCrop$CROPDMGEXP == "M") | (StormSubsetCrop$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

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

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

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

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

```
EVENT_CROP <- data.frame(table(StormSubsetCrop$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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

StormSubsetCrop[grep("CF", StormSubsetCrop$EVTYPE1, ignore.case = FALSE),
```

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

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

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

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

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

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

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

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

StormSubsetCrop[grep(".*LANDSLIDE*.", StormSubsetCrop$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.

```
StormSubsetCrop <- group_by(StormSubsetCrop, EVTYPE1)
StormSubsetCrop1 <- summarise(StormSubsetCrop, CROPDMGCOST=sum(CROPDMGCOST))
StormSubsetCrop1 <- arrange(StormSubsetCrop1, desc(CROPDMGCOST))
StormSubsetCrop1$CROP_DAMAGE_COST <- ifelse(StormSubsetCrop1$CROPDMGCOST > 1000000000,
paste(signif(StormSubsetCrop1$CROPDMGCOST/1000000000, digits=5),"B", sep=" "),
StormSubsetCrop1$CROPDMGCOST)
colnames(StormSubsetCrop1) <- c("EVENT","CROPDMGCOST", "CROP DAMAGE COST")
```

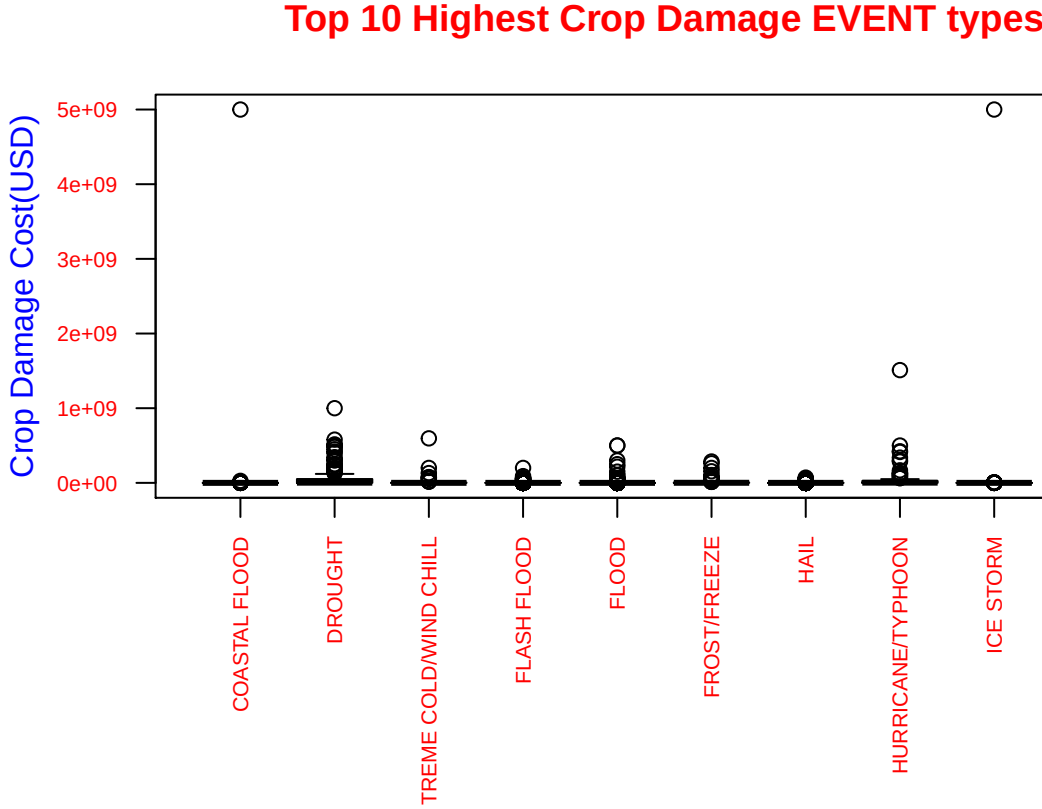
The Top 10 Storm Data Events with greatest economic consequences being Crop damage are as shown below in tabular form.

```
kable(head(StormSubsetCrop1[, c("EVENT","CROP DAMAGE COST")],10))
```

EVENT	CROP DAMAGE COST
DROUGHT	13.973 B
FLOOD	5.79 B
HURRICANE/TYPHOON	5.5161 B
COASTAL FLOOD	5.0575 B
ICE STORM	5.0221 B
HAIL	3.0469 B
FROST/FREEZE	1.8891 B
FLASH FLOOD	1.5322 B
EXTREME COLD/WIND CHILL	1.335 B
THUNDERSTORM WIND	1.2717 B

```
crop10 <- data.frame(head(StormSubsetCrop1[, "EVENT"],10))
par(mfrow=c(1,1), las = 2, mar=c(8, 4, 4, 2), cex.axis=0.65)
boxplot(CROPDMGCOST ~ EVTYPE1, data=subset(StormSubsetCrop,
subset=StormSubsetCrop$EVTYPE1 %in% crop10$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")
```

Boxplot for the range of values for Top 10 Storm Data Events with greatest economic consequences being Crop damage.



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

```
StormSubsetProp <- StormSubsetEconomic
```

Get the distinct values for CROPDMGEXP in the dataset

```
pexp <- data.frame(table(StormSubsetProp$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
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)

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

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

```
StormSubsetProp <- StormSubsetProp[(StormSubsetProp$PROPDMGEXP == "K") |
(StormSubsetProp$PROPDMGEXP == "M") | (StormSubsetProp$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

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

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

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

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

```
EVENT_PROP <- data.frame(table(StormSubsetProp$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

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

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

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

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

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

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

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

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

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

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

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

StormSubsetProp[grep(".*WINTER STORM*.", StormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "WINTER STORM"
```

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

```
"EVTYPE1"] <- "COASTAL FLOOD"
StormSubsetProp[grep("RIVER AND STREAM FLOOD", StormSubsetProp$EVTYPE1, ignore.case = FALSE),
"EVTYPE1"] <- "COASTAL FLOOD"

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

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

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

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

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

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

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

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

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

```
StormSubsetProp <- group_by(StormSubsetProp, EVTYPE1)
StormSubsetProp1 <- summarise(StormSubsetProp, PROPDMGCOST=sum(PROPDMGCOST))
StormSubsetProp1 <- arrange(StormSubsetProp1, desc(PROPDMGCOST))
StormSubsetProp1$PROPERTY_DAMAGE_COST <- ifelse(StormSubsetProp1$PROPDMGCOST > 1000000000,
paste(signif(StormSubsetProp1$PROPDMGCOST/1000000000, digits=5),"B", sep=" "), StormSubsetProp1$PROPDMG
colnames(StormSubsetProp1) <- c("EVENT","PROPDMGCOST","PROPERTY DAMAGE COST")
```

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

```
kable(head(StormSubsetProp1[, c("EVENT","PROPERTY DAMAGE COST")],10))
```

EVENT	PROPERTY DAMAGE COST
HURRICANE/TYPHOON	85.356 B
TORNADO	58.542 B
STORM SURGE/TIDE	47.965 B
FLOOD	30.033 B
FLASH FLOOD	16.947 B
HAIL	15.975 B
THUNDERSTORM WIND	10.971 B
WILDFIRE	8.4966 B
TROPICAL STORM	7.7144 B
WINTER STORM	6.749 B

```
prop10 <- data.frame(head(StormSubsetProp1[, "EVENT"],10))
par(mfrow=c(1,1), las = 2, mar=c(8, 4, 4, 2), cex.axis=0.65)
boxplot(PROPDMGCOST ~ EVTYPE1, data=subset(StormSubsetProp,
subset=StormSubsetProp$EVTYPE1 %in% prop10$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")
```

Boxplot for the range of values for Top 10 Storm Data Events with greatest economic consequences being Property damage.