**Analysis of Natural Disasters Impact on Human Lives and the Economy in the United States**

**Synopsis**

The U.S. National Oceanic and Atmospheric Administration’s (NOAA) has been collecting and annually reporting the events related to any natural disaster occurring across the country and the impact of each on the lives of citizens and the economy in the effected regions. These data are archived in the NOAA database and are available to the public.

For this report we have used some of the data from NOAA database to study and analysis the data to identify most costly events that have been reported from across the country between 1993 and 2011. We will focus on two aspects of the aftermath of these events: the human casualties and the setback on the economy of the affected region as a whole.

**Loading and Processing the Raw Data**

We read the data from [Dataset:U.S. National Oceanic and Atmospheric Administration’s (NOAA) storm database](https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) location and used the bzip2 decompression algorithm to decompress the file before reading it into a data frame. Free version of the bzip2 package can be downloaded from the Internet. The data file is stored in directory ~/Storm\_Data/repdata\_data\_StormData.csv of the local hard drive.

We use read.csv2 to read in the data as character class and store them in a data frame. To see if we need to do some clean up and formatting of the data we take a look at the structure of the data.

data <- read.csv2("Storm\_Data/repdata\_data\_StormData.csv", header=TRUE, sep = ",", colClasses = "character")

dim(data)

## [1] 902297 37

str(data)

## 'data.frame': 902297 obs. of 37 variables:

## $ STATE\_\_ : chr "1.00" "1.00" "1.00" "1.00" ...

## $ BGN\_DATE : chr "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" ...

## $ BGN\_TIME : chr "0130" "0145" "1600" "0900" ...

## $ TIME\_ZONE : chr "CST" "CST" "CST" "CST" ...

## $ COUNTY : chr "97.00" "3.00" "57.00" "89.00" ...

## $ COUNTYNAME: chr "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...

## $ STATE : chr "AL" "AL" "AL" "AL" ...

## $ EVTYPE : chr "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...

## $ BGN\_RANGE : chr "0.00" "0.00" "0.00" "0.00" ...

## $ BGN\_AZI : chr "" "" "" "" ...

## $ BGN\_LOCATI: chr "" "" "" "" ...

## $ END\_DATE : chr "" "" "" "" ...

## $ END\_TIME : chr "" "" "" "" ...

## $ COUNTY\_END: chr "0.00" "0.00" "0.00" "0.00" ...

## $ COUNTYENDN: chr "" "" "" "" ...

## $ END\_RANGE : chr "0.00" "0.00" "0.00" "0.00" ...

## $ END\_AZI : chr "" "" "" "" ...

## $ END\_LOCATI: chr "" "" "" "" ...

## $ LENGTH : chr "14.00" "2.00" "0.10" "0.00" ...

## $ WIDTH : chr "100.00" "150.00" "123.00" "100.00" ...

## $ F : chr "3" "2" "2" "2" ...

## $ MAG : chr "0.00" "0.00" "0.00" "0.00" ...

## $ FATALITIES: chr "0.00" "0.00" "0.00" "0.00" ...

## $ INJURIES : chr "15.00" "0.00" "2.00" "2.00" ...

## $ PROPDMG : chr "25.00" "2.50" "25.00" "2.50" ...

## $ PROPDMGEXP: chr "K" "K" "K" "K" ...

## $ CROPDMG : chr "0.00" "0.00" "0.00" "0.00" ...

## $ CROPDMGEXP: chr "" "" "" "" ...

## $ WFO : chr "" "" "" "" ...

## $ STATEOFFIC: chr "" "" "" "" ...

## $ ZONENAMES : chr "" "" "" "" ...

## $ LATITUDE : chr "3040.00" "3042.00" "3340.00" "3458.00" ...

## $ LONGITUDE : chr "8812.00" "8755.00" "8742.00" "8626.00" ...

## $ LATITUDE\_E: chr "3051.00" "0.00" "0.00" "0.00" ...

## $ LONGITUDE\_: chr "8806.00" "0.00" "0.00" "0.00" ...

## $ REMARKS : chr "" "" "" "" ...

## $ REFNUM : chr "1.00" "2.00" "3.00" "4.00" ...

In this study we are interested in the impact of natural events on human lives and property damages, hence need to keep the few columns that have the data for fatalities/injuries, property/crop damages, and their respective exponent (EXP). We also keep few of factored columns such as type of event, beginning date of event, state, etc. We also checked for any NA’s in the remaining data, which none existed, and formatted appropriate columns to numeric.

data1 <- data[,-c(1,5,9:22,29:37)]

sum(is.na(data1))

## [1] 0

data1[,7] <- as.numeric(data1[,7]); data1[,8] <- as.numeric(data1[,8]); data1[,9] <- as.numeric(data1[,9]); data1[,11] <- as.numeric(data1[,11])

We noticed there are some undefined characters in PROPDMGEXP and CROPDMGEXP. Rows with these value have to be removed from the data before we replace the defined values (H,K,M,B) with their numeric values (100, 1000, etc.), so the total cost of damages can be calculated by adding the dollar amount from properties and crops.

unique(data1$PROPDMGEXP)

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

## [18] "1" "8"

unique(data1$CROPDMGEXP)

## [1] "" "M" "K" "m" "B" "?" "0" "k" "2"

data1 <- data1[data1$CROPDMGEXP %in% c("K", "M", "B", "b", "m", "k"),]

data1 <- data1[data1$PROPDMGEXP %in% c("K", "M", "B", "b", "m", "k", "H", "h"),]

data1$CROPDMGEXP <- gsub("B", "10e+9", data1$CROPDMGEXP, ignore.case=T )

data1$CROPDMGEXP <- gsub("M", "10e+6", data1$CROPDMGEXP, ignore.case=T )

data1$CROPDMGEXP <- gsub("k", "1000", data1$CROPDMGEXP, ignore.case=T )

data1$PROPDMGEXP <- gsub("B", "10e+9", data1$PROPDMGEXP, ignore.case=T )

data1$PROPDMGEXP <- gsub("M", "10e+6", data1$PROPDMGEXP, ignore.case=T )

data1$PROPDMGEXP <- gsub("k", "1000", data1$PROPDMGEXP, ignore.case=T )

data1$PROPDMGEXP <- gsub("H", "100", data1$PROPDMGEXP, ignore.case=T )

data1$YEAR <- substr(as.Date(data1$BGN\_DATE, format="%m/%d/%Y"),1,4)

Now we are ready to identify events that have the greatest economic consequences.

library(ggplot2)

library(plyr)

# Calculate damages to cROps and properties and add them to get the total damage (TOTALDMG)

data1$CROP <- data1$CROPDMG \* as.numeric(data1$CROPDMGEXP)

data1$PROP <- data1$PROPDMG \* as.numeric(data1$PROPDMGEXP)

data1$TOTALDMG <- data1$CROP + data1$PROP

# Get the sum of reported damages for each event

GEC <- ddply(data1, "EVTYPE", summarise, Total\_Damage = sum(TOTALDMG))

GEC <- GEC[order(GEC$Total\_Damage, decreasing = T),]

# Just look at the top events, those with higher than $1B

topGEC <- GEC[GEC$Total\_Damage >= 10e+9,]

names(topGEC) <- c("EVTYPE","Total\_Damage\_Millions")

topGEC[,2] <- topGEC[,2]/1000

ggplot(topGEC, aes(x=reorder(EVTYPE,Total\_Damage\_Millions), y=Total\_Damage\_Millions)) +

geom\_bar(stat="identity", fill="blue", colr="black") +

xlab("Event Type") + ylab("Total Damage in Millions") +

ggtitle("Types of events with highest impact on the the economy") +

theme(axis.text.x = element\_text(angle=60, hjust=1))

topGEC[1,]

## EVTYPE Total\_Damage\_Millions

## 23 FLOOD 1.374e+09

Damages to properties and crops across the country from floods is by far the most costly event reported as of 2011, a total of 1.3739 × 109 million dollars.

Similar calculation for causalities (fatality + injuries) caused by natural disasters is done below in order to identify which type of event has caused the most damage to human health across the country.

data1$FAT\_INJ <- data1$FATALITIES + data1$INJURIES

health <- ddply(data1, "EVTYPE", summarise, Tot\_Fatality = sum(FATALITIES), Tot\_INJURIES = sum(INJURIES), Total = sum(FAT\_INJ))

health <- health[order(health$Total, decreasing = T),]

ggplot(health[1:20,], aes(x=reorder(EVTYPE,Total), y=Total)) +

geom\_bar(stat="identity", fill="blue", colr="black") +

xlab("Event Type") + ylab("Total Casualties") +

ggtitle("Top 20 types of events with largest number of Fatalities + Injuries") +

theme(axis.text.x = element\_text(angle=60, hjust=1))

head(health,10)

## EVTYPE Tot\_Fatality Tot\_INJURIES Total

## 98 TORNADO 1064 11960 13024

## 23 FLOOD 261 6495 6756

## 63 ICE STORM 13 1616 1629

## 88 THUNDERSTORM WIND 130 1412 1542

## 41 HEAT 204 1172 1376

## 68 LIGHTNING 173 1010 1183

## 16 EXCESSIVE HEAT 171 899 1070

## 19 FLASH FLOOD 387 667 1054

## 61 HURRICANE/TYPHOON 40 909 949

## 117 WILDFIRE 59 555 614

Data shows that Tornadoes have taken the lives of 1064 people and caused injuries to 1.196 × 104 people.

Which state has experienced the highest cost due to flood or any other natural disasters in a year?

data1 <- data1[order(data1$TOTALDMG, decreasing=T),]

head(data1[,c(5:6,13:16)])

## STATE EVTYPE YEAR CROP PROP TOTALDMG

## 605953 CA FLOOD 2006 3.25e+08 1.15e+12 1.150e+12

## 198389 IL RIVER FLOOD 1993 5.00e+10 5.00e+10 1.000e+11

## 581537 MS HURRICANE/TYPHOON 2005 1.51e+10 5.88e+10 7.390e+10

## 529351 FL HURRICANE/TYPHOON 2004 2.85e+09 5.42e+10 5.705e+10

## 211900 MS ICE STORM 1994 5.00e+10 5.00e+05 5.000e+10

## 529436 FL HURRICANE/TYPHOON 2004 9.32e+08 4.83e+10 4.923e+10

The flood in the state of California in 2006 caused the most costly damage to properties and crops over the course of 20 years since 1993.Interestingly, large portion of the cost was related to property damages, a nearly 99.97 percentage.

**Results**

During the course of nearly 20 years (1993-2011) US was hit with 900K natural events. Among them all, Tornadoes have been taking more lives and causing injuries to people of effected regions followed by flood as the second most life threatening event. With respect to damages to properties and crops, floods have caused the most impact to the economy. Just in one event in 2006 the State of California experienced the most devastating flood from a long lasting rainfall in the winter causing major damages to properties and wineries in NAPA. In that year a record high of $115B lose in property and crop damage was reported.