Reproducible Research - Peer Assessment 2

Joe Cannon

October 12, 2015

Impact of Severe Weather Events on Public Health and Economy in the United States Synopsis In this report, the goal is to analyze the impact of different weather events on public health and economy based on the storm database collected from the U.S. National Oceanic and Atmospheric Administration's (NOAA) from 1950 - 2011. The data used will be estimates of:

-Fatalities, and injuries to caclulate damage with regards to population health -Property and crop damage to caclulate damage with regards to economic impact

From these data, I found that tornado are the most damaging with respect to both population health, as well as economic impacts (followed closely by Flash Floods). I also noted that the most damage occurs between the hours of 10pm and 5am.

Load Libraries

```
suppressWarnings(suppressMessages(library(ggplot2)))
suppressWarnings(suppressMessages(library(dplyr)))
suppressWarnings(suppressMessages(library(lubridate)))
suppressWarnings(suppressMessages(library(gridExtra)))
```

Data Processing First, we check to make sure the data is present, if not download it. Then we decompress it so we can proess it.

```
if (!"stormdata.csv.bz2" %in% tolower(dir("./"))) {
    print("hhhh")
    download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile =
}
stormdata <- read.table("stormdata.csv.bz2",
                        header = T,quote="\"", sep=",",na.strings = NA)
```

```
stormdata$BGN_DATE <- as.Date(stormdata$BGN_DATE,format="%m/%d/%Y")
stormdata$BGN_TIME <- as.character(stormdata$BGN_TIME)
stormdata$EVTYPE   <- as.character(stormdata$EVTYPE)
bn12hr <- length(stormdata[grep("AM|PM",stormdata$BGN_TIME),]$BGN_TIME)
```

Now we have to clean up the begin dates. Even though they are suppose to enter the time in 24hr format; we noticed that 653530 timeshave entered in 12hr format. So we are going to convert those to 24hr format.

NOTE: The Ending times are so sparse and in such bad shape that it does not make sense to work with that data at this point.

Now we pad the time column.

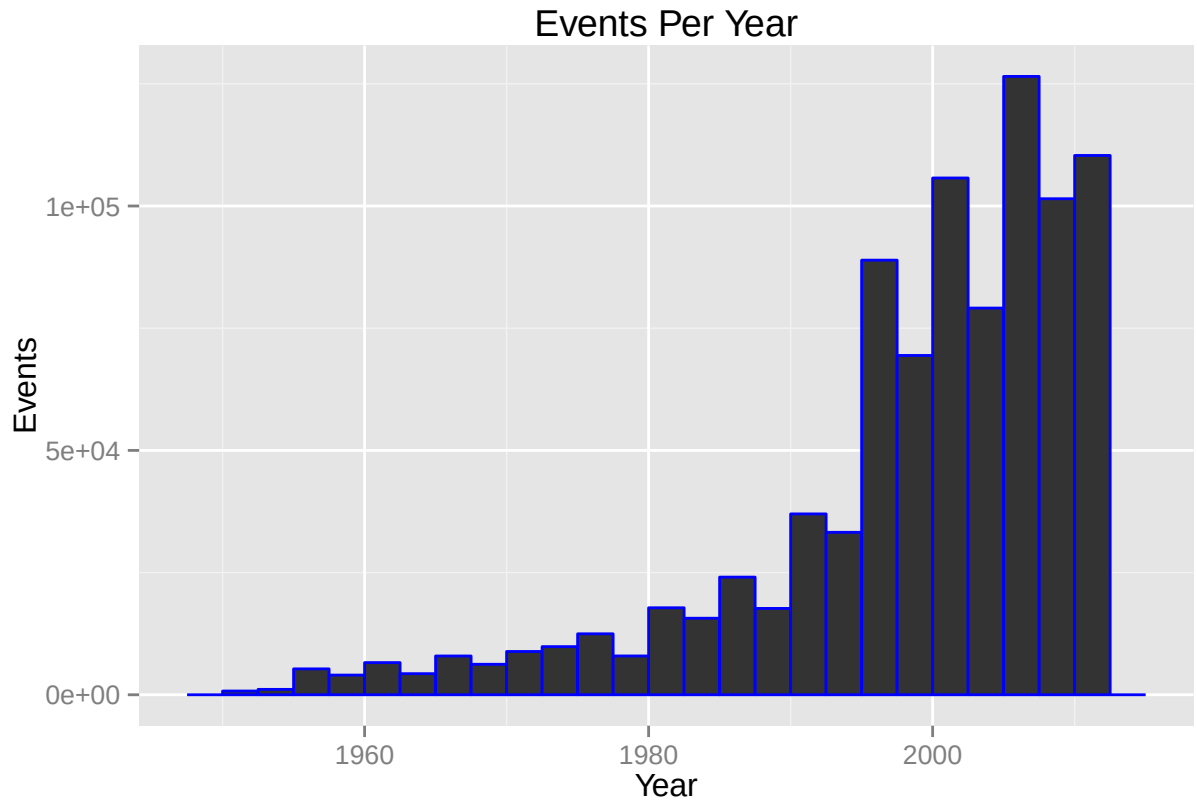
```
AMPMData <-stormdata[grepl("[AM]|[PM]",stormdata$BGN_TIME),]
t2 <-stormdata[!(grepl("[AM]|[PM]",stormdata$BGN_TIME)),]
AMPMData$NBGN_TIME <- paste0(substr(strptime(AMPMData$BGN_TIME, "%I:%M:%S %p" ),12,13),
AMPMData[grepl("NA",AMPMData$NBGN_TIME),]$NBGN_TIME<- paste0(substr(AMPMData[grepl("NA",AMPMData$NBGN_T
AMPMData$BGN_TIME <-AMPMData$NBGN_TIME
AMPMData <- select(AMPMData,-(NBGN_TIME))
stormdata <- rbind(t2,AMPMData)
stormdata$BGN_TIME <-  sprintf("%04s", stormdata$BGN_TIME)
```

‘‘‘

```
lastyear <- max(year(stormdata$BGN_DATE))
firstyear <- min(year(stormdata$BGN_DATE))
```

The events in the database start in the year 1950 and end in 2011. The spread is as follows.

```
ggplot(stormdata, aes(x = year(stormdata$BGN_DATE))) +
         geom_histogram(color="blue",binwidth=2.5) +
        labs(title="Events Per Year",
             x = "Year", y = "Events")
```



Trim down to the last 20 years. Events occurring before then are not as relevant as technologies such as storm warnings, architectural integrity have dramatically improved. Also data entered in prior to 1996 is much more sparse and suspect in coding. It should be noted that during my experiments I noted changing the cutoff date above 1991 alters the top economic event as shown

- All Years -> Tornado
- 1991(last 20 years) -> Tornado
- 1995(Big jump in avail data) -> Flash Flood
- 1996(last 15 years) -> TSTM Wind
- 2001(last 10 years) -> Flash Flood

Personal cost remained consistant with Tornados being the #1 cause of injury/death by a hefty margin. This point came up in the discussion forums.

```
cstormdata <- stormdata[which(year(stormdata$BGN_DATE)>=1991),]%>%
              select(BGN_DATE,BGN_TIME,EVTYPE,INJURIES,FATALITIES,PROPDMG,CROPDMG)
cstormdata$EVTYPE <- as.character(cstormdata$EVTYPE)
```

```
personalcost_event <- aggregate(cbind(INJURIES,FATALITIES,INJURIES+FATALITIES)~EVTYPE , cstormdata, sum

colnames(personalcost_event) <- c("EVTYPE","INJURIES","FATALITIES","TOTALPC")
personalcost_event$EVTYPE = as.character(personalcost_event$EVTYPE)

TPI <- head(personalcost_event[order(-personalcost_event$TOTALPC),],5)

p1<- ggplot(TPI, aes(x=EVTYPE, y=TOTALPC)) +
          geom_histogram(color="blue",stat = "identity") +
         labs(title="Top 10 Cost(in population health ) Per Event",
              x = "Event Type", y = "Cost")
colnames(TPI) <- c("Event","Injuries","Fatalities","Total Death/Fatalities")
grid.table(TPI, rows=NULL)
```

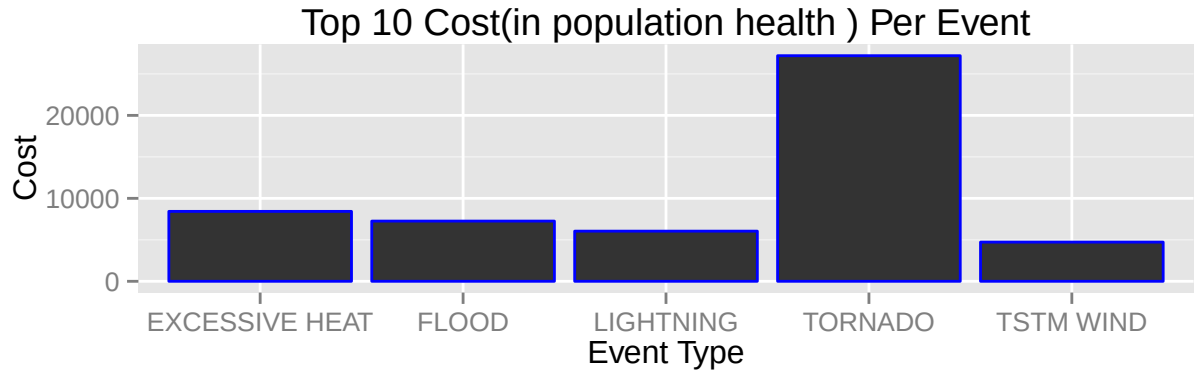
Event	Injuries	Fatalities	Total Death/Fatalities
TORNADO	25497	1699	27196
EXCESSIVE HEAT	6525	1903	8428
FLOOD	6789	470	7259
LIGHTNING	5230	816	6046
TSTM WIND	4441	285	4726

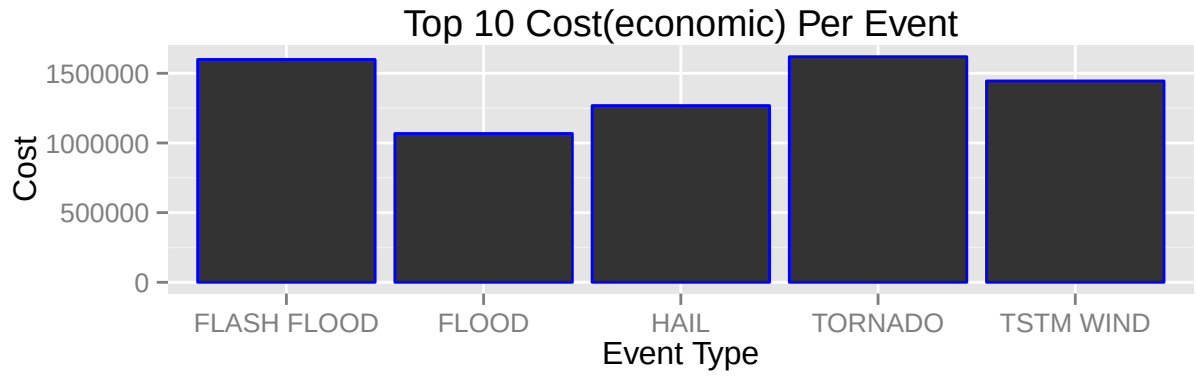
```
economiccost_event <- aggregate(cbind(PROPDMG,CROPDMG,PROPDMG+CROPDMG)~EVTYPE , cstormdata, sum, na.acti
colnames(economiccost_event) <- c("EVTYPE","PROPDMG","CROPDMG","TOTALEC")
economiccost_event$EVTYPE = as.character(economiccost_event$EVTYPE)
TEC <- head(economiccost_event[order(-economiccost_event$TOTALEC),],5)
p3<- ggplot(TEC, aes(x=EVTYPE, y=TOTALEC)) +
         geom_histogram(color="blue",stat = "identity") +
        labs(title="Top 10 Cost(economic) Per Event",
             x = "Event Type", y = "Cost")

colnames(TEC) <- c("Event","Prop Damage","Crop Damage","Total Damage")
grid.table(TEC, rows=NULL)
```

Event	Prop Damage	Crop Damage	Total Damage
TORNADO	1519172.4	100018.5	1619191
FLASH FLOOD	1420124.6	179200.5	1599325
TSTM WIND	1335965.6	109202.6	1445168
HAIL	688693.4	579596.3	1268290
FLOOD	899938.5	168037.9	1067976

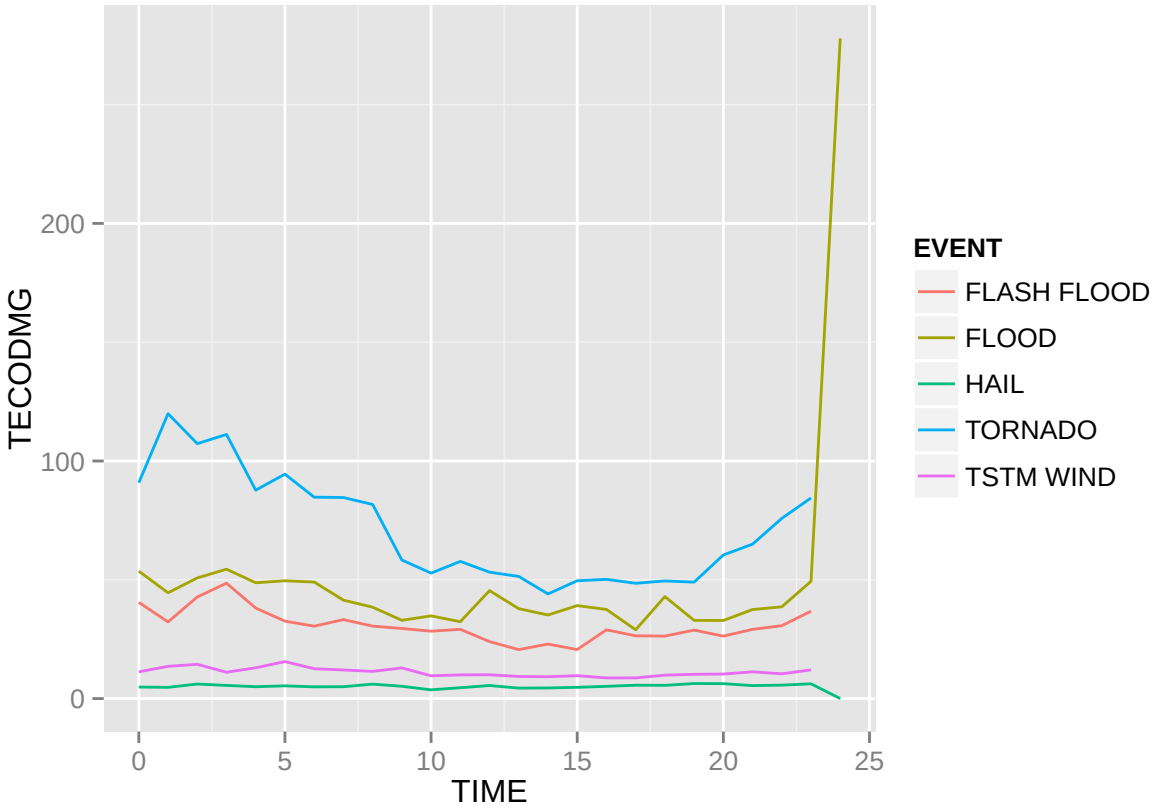
```
grid.arrange(p1,p3)
```





For economic damage time seems to be the most sensitive to Tornados, Floods and Flash Floods

```
Tornados <- cstormdata[cstormdata$EVTYPE %in% TEC$Event,]
Tornado_TimeTrend <- aggregate(cbind(PROPDMG+CROPDMG,INJURIES+FATALITIES)~as.integer(substr(BGN_TIME,1,
colnames(Tornado_TimeTrend) <- c("TIME","EVENT","TECODMG","TPERDMG")
 ggplot(Tornado_TimeTrend, aes(x=TIME, y=TECODMG,colour = EVENT)) +
  geom_line()
```


TECODMG
200
100
0
0
5
10
15
20
25
TIME
EVENT
FLASH FLOOD
FLOOD
HAIL
TORNADO
TSTM WIND