Storm Data Analysis using NOAA provided data(1976-2011)

DainaE

Tuesday, March 17, 2015

Title:

This research presents a study into the Storm Data which is an official publication of the National Oceanic and Atmospheric Administration (NOAA). There is a study done on occurrence of storms and other significant weather phenomena and their intensity to cause loss of life, injuries, significant property damage, and/or disruption to commerce

Synopsis:

The research involves study into the Storm Data published by National Oceanic and Atmospheric Administration. The objective of the study is to clasify the data to correspond to key events classification , to identify the events which are major causes of impact on Human and Economic loss. The significance of the study is to identify areas of how to improve the event reporting to gather more accurate data and also to propose areas of focus to the municipal regulatories. Scope of the study is limited to the NOAA Storm Database. The limitations in the current analysis of the events would be : 1. The event mapping is done based on the judgement and ability if the author which can be refined and also corrected in certain cases. 2. The time constraint of one week without domain knowledge in the field of weather events can lead to biased judgement

Data Processing

Read the Storm Data for analysis Data is read from the CSV file

```
setwd("C://Personal//Coursera//Assignment//Data Storm")
sd_data <- read.csv("repdata-data-StormData.csv.bz2")
library(dplyr)

## Warning: package 'dplyr' was built under R version 3.1.2

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

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

```
sd_tbl <- tbl_df(sd_data)
```

Cleaning Data

Steps involved to get a clean Data

1. We are interested in evaluation the loss as part of Question 1 and Question 2 of this analysis. Hence we filter and extract only those rows where one of the loss is >0 ie Fatalities, Injuries, Cropdmg,Propdmg This brings us to a 447 row dataset.

2. From the codebook section 2.1.1 we understand the number of valid event types is 48. We have to clean the data to bring it to that form. Valid types from codebook is saved into valid_events flat file.

Based on manual analysis of data we understand the kind of errors in the naming of EVTYPE variables. We will replace them to uniformly be one of the 48 valid event types. There are 48 replacements which will be done to clean up the data. We introduce a new column EventMap. We extract all the rows which should correspond to a given valid EVTYPE and save the value in EventMap The file lookupTable (https://github.com/demmanuel123/DataStormAnalysis/blob/master/lookupTable.csv) is available in github , which gives the strategy for cleanup and the keys used to segregate teh 447 line items based on relevance into the 48 events.

```
library(splus2R)
```

```
## Warning: package 'splus2R' was built under R version 3.1.3
```

```
events_tbl <- filter(sd_tbl,FATALITIES >0 | INJURIES >0 | CROPDMG >0 |
PROPDMG>0 )
events_tbl$EVTYPE <- lowerCase(events_tbl$EVTYPE)
events_tbl <- select(events_tbl,EVTYPE,BGN_DATE,
FATALITIES,INJURIES,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP) %>%
group_by(EVTYPE) %>% summarise_each(funs(sum))
#Event list from code book section 2.1.1 copied to the valid_events file
events_validlist <- read.fwf("valid_events",widths = 24)
```

```
## Warning in readLines(file, n = thisblock): incomplete final line found on
## 'valid_events'
```

```
events_validlist <- validSlotNames(events_validlist)
events_validlist$V1 <- gsub(c(" "),"",events_validlist$V1)
events_tbl <- mutate(events_tbl,EventMap=NA)

#447*8 matrix
dim(events_tbl)
```

```
## [1] 447   9

pick.event <- function(x)
{as.character(events_validlist$V1[grepl(x,events_validlist$V1)])}
events_tbl$EventMap<- NA
#Clean up data by associating to the right 48 event types

nrow(events_tbl[is.na(events_tbl$EventMap),])

## [1] 447

#25
events_tbl$EventMap[grepl("low tide",events_tbl$EVTYPE)] <-
pick.event("Astronomical*")
events_tbl$EventMap[grepl("avalan",events_tbl$EVTYPE)] <-
pick.event("Avalanche*")
events_tbl$EventMap[grepl("blizzard",events_tbl$EVTYPE)] <-
pick.event("Blizzard*")
events_tbl$EventMap[grepl("dense fog",events_tbl$EVTYPE)] <-
pick.event("DenseFog*")
events_tbl$EventMap[grepl("dense smoke",events_tbl$EVTYPE)] <-
pick.event("Smoke*")
events_tbl$EventMap[grepl("dust devil",events_tbl$EVTYPE)] <-
pick.event("DustDevil*")
events_tbl$EventMap[grepl("dust storm",events_tbl$EVTYPE)] <-
pick.event("DustStorm*")
events_tbl$EventMap[grepl("hail",events_tbl$EVTYPE)] <- pick.event("Hail*")
events_tbl$EventMap[grepl("hurricane",events_tbl$EVTYPE)] <-
pick.event("Hurricane*")
events_tbl$EventMap[grepl("tropical depression",events_tbl$EVTYPE)] <-
pick.event("TropicalDep*")
events_tbl$EventMap[grepl("tropical storm",events_tbl$EVTYPE)] <-
pick.event("TropicalStorm*")
events_tbl$EventMap[grepl("tsunami",events_tbl$EVTYPE)] <-
pick.event("Tsunami*")
events_tbl$EventMap[grepl("volcanic",events_tbl$EVTYPE)] <-
pick.event("Volcanic*")
events_tbl$EventMap[grepl("lakeshore",events_tbl$EVTYPE)] <-
pick.event("Lakeshore*")
events_tbl$EventMap[grepl("marine hail",events_tbl$EVTYPE)] <-
pick.event("MarineHail*")
events_tbl$EventMap[grepl("marine high wind",events_tbl$EVTYPE)] <-
pick.event("MarineHigh*")
events_tbl$EventMap[grepl("marine strong wind",events_tbl$EVTYPE) |
grepl("marine mishap",events_tbl$EVTYPE) ] <- pick.event("MarineStrong*")
events_tbl$EventMap[grepl("funnel",events_tbl$EVTYPE)] <-
pick.event("FunnelCloud*")
events_tbl$EventMap[grepl("drought",events_tbl$EVTYPE)] <-
pick.event("Drought*")
events_tbl$EventMap[grepl("rip ",events_tbl$EVTYPE) ] <- pick.event("Rip*")
events_tbl$EventMap[grepl("seiche",events_tbl$EVTYPE)] <-
```

```
pick.event("Seiche*")
events_tbl$EventMap[grepl("lake-effect",events_tbl$EVTYPE)] <-
pick.event("Lake-Effect*")
events_tbl$EventMap[grepl("storm surge",events_tbl$EVTYPE)] <-
pick.event("StormSurge*")

#8
events_tbl$EventMap[grepl("^winter*",events_tbl$EVTYPE)] <-
pick.event("WinterWeather*")

events_tbl$EventMap[grepl("wild",events_tbl$EVTYPE) |
grepl("forest",events_tbl$EVTYPE) | grepl("grass",events_tbl$EVTYPE)] <-
pick.event("Wildfire*")
events_tbl$EventMap[grepl("coastal",events_tbl$EVTYPE) |
grepl("*cstl*",events_tbl$EVTYPE) | grepl("beach*",events_tbl$EVTYPE)] <-
pick.event("Coastal*")
events_tbl$EventMap[grepl("flash",events_tbl$EVTYPE) &
grepl("flood",events_tbl$EVTYPE) | grepl("dam break",events_tbl$EVTYPE)] <-
pick.event("FlashFlood*")
events_tbl$EventMap[grepl("flood",events_tbl$EVTYPE) |
grepl("fld",events_tbl$EVTYPE)& !grepl("*flash",events_tbl$EVTYPE) &
!grepl("coastal",events_tbl$EVTYPE) & !grepl("*cstl",events_tbl$EVTYPE) &
!grepl("lakeshore",events_tbl$EVTYPE)]<- pick.event("^Flood*")
events_tbl$EventMap[(grepl("tstm",events_tbl$EVTYPE) |
grepl("micro",events_tbl$EVTYPE) | grepl("downburst",events_tbl$EVTYPE) |
grepl("*rstorm winds*",events_tbl$EVTYPE) |
grepl("thunder",events_tbl$EVTYPE)) & !grepl("marine*",events_tbl$EVTYPE) &
!grepl("hail",events_tbl$EVTYPE) & !grepl("flood",events_tbl$EVTYPE)] <-
pick.event("^Thunderstorm*")
events_tbl$EventMap[(grepl("^cold",events_tbl$EVTYPE) ) & !grepl("extreme
cold/wind",events_tbl$EVTYPE) & !grepl("tornado",events_tbl$EVTYPE)] <-
pick.event("^Cold/WindChill*")
events_tbl$EventMap[grepl("extreme cold",events_tbl$EVTYPE) |
grepl("hypo",events_tbl$EVTYPE) | grepl("unseasonabl.
cold",events_tbl$EVTYPE) | grepl("extended cold",events_tbl$EVTYPE) |
grepl("record cold",events_tbl$EVTYPE) | grepl("extreme
wind*",events_tbl$EVTYPE)] <- pick.event("ExtremeCold/WindChill")
events_tbl$EventMap[grepl("lightning",events_tbl$EVTYPE) |
grepl("lighting",events_tbl$EVTYPE) | grepl("ligntnin*",events_tbl$EVTYPE)&
!grepl("tstm",events_tbl$EVTYPE) & !grepl("thunderstorm",events_tbl$EVTYPE) ]
<- pick.event("Lightning*")

#5
events_tbl$EventMap[grepl("freezing",events_tbl$EVTYPE) |
grepl("freez*",events_tbl$EVTYPE) & !grepl("sleet",events_tbl$EVTYPE) &
!grepl("dense",events_tbl$EVTYPE)] <- pick.event("Freezing*")
events_tbl$EventMap[ grepl("frost",events_tbl$EVTYPE)|
grepl("freez*",events_tbl$EVTYPE) | grepl("fog",events_tbl$EVTYPE)&
!grepl("sleet",events_tbl$EVTYPE) & !grepl("dense",events_tbl$EVTYPE)] <-
pick.event("Frost*")
```

```
events_tbl$EventMap[grepl("winter storm",events_tbl$EVTYPE) &
!grepl("blizzard",events_tbl$EVTYPE) & !grepl("heavy
snow*",events_tbl$EVTYPE)] <- pick.event("WinterStorm*")
events_tbl$EventMap[grepl("snow",events_tbl$EVTYPE) & !grepl("lake-
effect",events_tbl$EVTYPE) & !grepl("winter storm",events_tbl$EVTYPE)] <-
pick.event("HeavySnow*")
events_tbl$EventMap[grepl("ice storm",events_tbl$EVTYPE) |
grepl("^ice*",events_tbl$EVTYPE) | grepl("^glaze",events_tbl$EVTYPE) &
!grepl("black$",events_tbl$EVTYPE) & !grepl("^snow",events_tbl$EVTYPE)&
!grepl("sleet",events_tbl$EVTYPE) & !grepl("flash flood",events_tbl$EVTYPE)]
<- pick.event("IceStorm*")

events_tbl$EventMap[grepl("marine thunder*",events_tbl$EVTYPE) |
grepl("marine tst*",events_tbl$EVTYPE)] <- pick.event("MarineThunderstorm*")

events_tbl$EventMap[grepl("mud",events_tbl$EVTYPE) |
grepl("rock",events_tbl$EVTYPE) | grepl("landslide",events_tbl$EVTYPE) &
!grepl("flash flood",events_tbl$EVTYPE)] <- pick.event("Debris*")
events_tbl$EventMap[grepl("^tor*",events_tbl$EVTYPE) |
grepl("*tornado*",events_tbl$EVTYPE)] <- pick.event("Tornado*")
events_tbl$EventMap[grepl("waterspout",events_tbl$EVTYPE) &
!grepl("*tornado*",events_tbl$EVTYPE)] <- pick.event("Waterspout*")
events_tbl$EventMap[grepl("excessive heat",events_tbl$EVTYPE) |
grepl("extreme heat",events_tbl$EVTYPE) &
!grepl("*drought*",events_tbl$EVTYPE)] <- pick.event("ExcessiveHeat*")
events_tbl$EventMap[grepl("*heat",events_tbl$EVTYPE) |
grepl("*warm",events_tbl$EVTYPE) & !grepl("*excessive*",events_tbl$EVTYPE) &
!grepl("*extreme*",events_tbl$EVTYPE) &
!grepl("*drought*",events_tbl$EVTYPE)] <- pick.event("Heat*")
events_tbl$EventMap[grepl("*heavy rain*",events_tbl$EVTYPE) |
grepl("excessive rain*",events_tbl$EVTYPE) &
!grepl("*flash*",events_tbl$EVTYPE) & !grepl("*flood*",events_tbl$EVTYPE) &
!grepl("*high surf*",events_tbl$EVTYPE) &
!grepl("lightning*",events_tbl$EVTYPE) ] <- pick.event("HeavyRain*")
events_tbl$EventMap[grepl("high surf",events_tbl$EVTYPE) |
grepl("^high*",events_tbl$EVTYPE) | grepl("heavy seas",events_tbl$EVTYPE)  |
grepl("heavy surf",events_tbl$EVTYPE) | grepl("high tide",events_tbl$EVTYPE)
| grepl("rough ",events_tbl$EVTYPE)] <- pick.event("HighSurf*")
events_tbl$EventMap[grepl("high wind",events_tbl$EVTYPE) |
grepl("^wind$",events_tbl$EVTYPE) | grepl("gusty wind",events_tbl$EVTYPE)&
!grepl("flood",events_tbl$EVTYPE) & !grepl("hail",events_tbl$EVTYPE)&
!grepl("*blizzard*",events_tbl$EVTYPE)] <- pick.event("^HighWind*")
events_tbl$EventMap[grepl("strong wind*",events_tbl$EVTYPE) |
grepl("typhoon*",events_tbl$EVTYPE) & !grepl("marine*",events_tbl$EVTYPE)] <-
pick.event("^StrongWind*")
events_tbl$EventMap[grepl("sleet",events_tbl$EVTYPE) | grepl("black
ice",events_tbl$EVTYPE)] <- pick.event("Sleet*")
```

```
events_new_tbl <- events_tbl[!is.na(events_tbl$EventMap),] %>%
group_by(EventMap) %>% summarise_each(funs(sum),-EVTYPE) %>%
arrange(desc(FATALITIES,INJURIES))
nrow(events_tbl[is.na(events_tbl$EventMap),])

## [1] 44

dim(events_new_tbl)

## [1] 45  8
```

Question 1

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

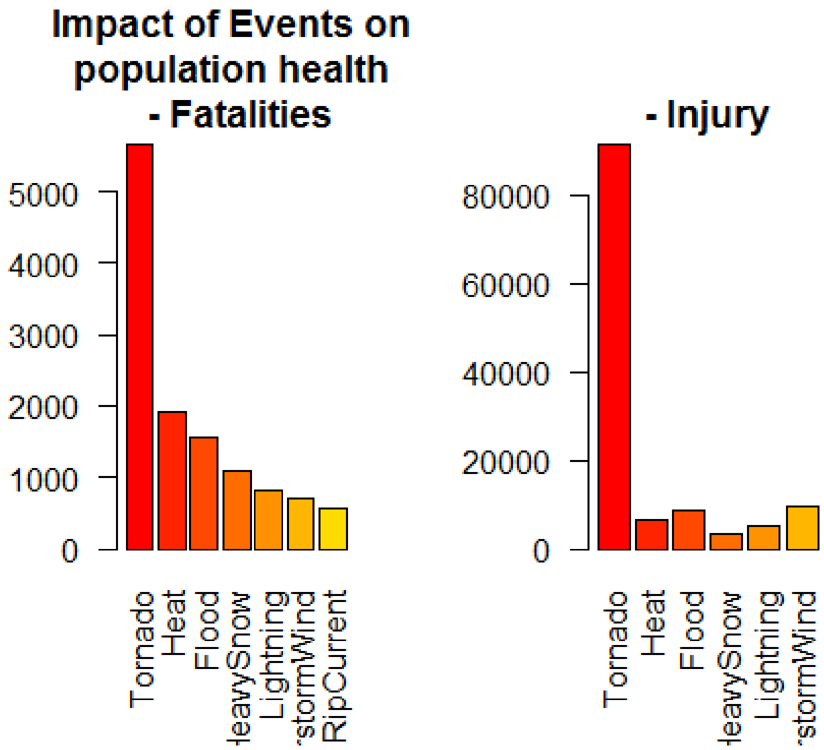
```
# EVents covered by Fatalities > mean and Injury > mean
FatMeanCoverage<-
100*sum(events_new_tbl$FATALITIES[which(events_new_tbl$FATALITIES
>mean(events_new_tbl$FATALITIES))])/sum(events_new_tbl$FATALITIES)
InjMeanCoverage<-
100*sum(events_new_tbl$INJURIES[which(events_new_tbl$INJURIES
>mean(events_new_tbl$INJURIES))])/sum(events_new_tbl$INJURIES)

UniqueCauses<- unique(events_new_tbl$EventMap[which(events_new_tbl$FATALITIES
>mean(events_new_tbl$FATALITIES))],events_new_tbl$EventMap[which(events_new_t
bl$INJURIES >mean(events_new_tbl$INJURIES))])
UniqueCauses <- gsub(", "," ",toString(UniqueCauses))

TITLE<- paste("The graph shows the impact of natural events on population
health, causing", round(FatMeanCoverage,0),"% of Fatalities and"
,round(InjMeanCoverage,0) ,"% of Injuries, Events responsible are ",
UniqueCauses )

par(mfrow = c(1,2))
barplot(height = events_new_tbl$FATALITIES[which(events_new_tbl$FATALITIES
>mean(events_new_tbl$FATALITIES))],names.arg =
(events_new_tbl$EventMap[which(events_new_tbl$FATALITIES
>mean(events_new_tbl$FATALITIES))]),col = heat.colors(10), las=2 ,border =
par("fg"), legend.text = TRUE,main = "Impact of Events on \npopulation health
\n - Fatalities")

barplot(height = events_new_tbl$INJURIES[which(events_new_tbl$INJURIES
>mean(events_new_tbl$INJURIES))],names.arg =
(events_new_tbl$EventMap[which(events_new_tbl$INJURIES
>mean(events_new_tbl$INJURIES))]),col = heat.colors(10),las=2, border =
par("fg"), legend.text = TRUE,main ="\n\n- Injury")
```



Analysis 1

The graph shows the impact of natural events on population health, causing 82 % of Fatalities and 89 % of Injuries, Events responsible are Tornado Heat Flood HeavySnow Lightning ThunderstormWind RipCurrent

Question 2

Across the United States, which types of events have the greatest economic consequences?

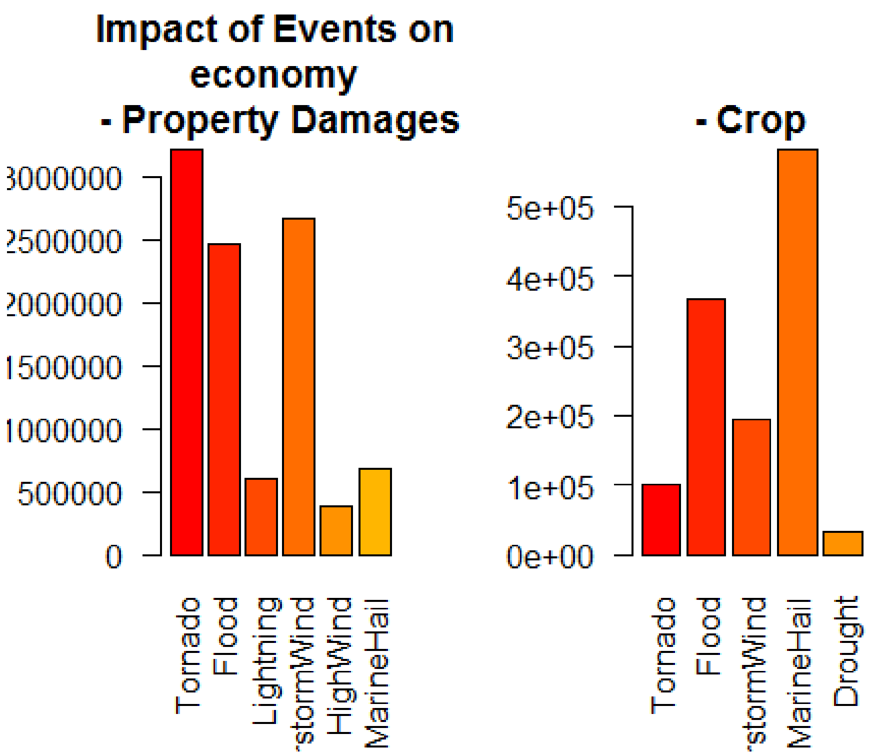
```
# EVents covered by Fatalities > mean and Injury > mean
PropMeanCoverage<-
100*sum(events_new_tbl$PROPDMG[which(events_new_tbl$PROPDMG
>mean(events_new_tbl$PROPDMG))])/sum(events_new_tbl$PROPDMG)
CropMeanCoverage<-
100*sum(events_new_tbl$CROPDMG[which(events_new_tbl$CROPDMG
>mean(events_new_tbl$CROPDMG))])/sum(events_new_tbl$CROPDMG)
UniqueCauses2<- unique(events_new_tbl$EventMap[which(events_new_tbl$PROPDMG
>mean(events_new_tbl$PROPDMG))],events_new_tbl$CROPDMG[which(events_new_tbl$C
ROPDMG >mean(events_new_tbl$CROPDMG))])

UniqueCauses2 <- gsub(", "," ",toString(UniqueCauses2))
TITLE2<- paste("The graph shows the impact of natural events on economic
damage, causing", round(PropMeanCoverage,0),"% of Property Damage and"
```

```
,round(CropMeanCoverage,0) ,"% of Crop Damage, Events responsible are ",
UniqueCauses2 )

par(mfrow = c(1,2))
barplot(height = events_new_tbl$PROPDMG[which(events_new_tbl$PROPDMG
>mean(events_new_tbl$PROPDMG))],names.arg =
(events_new_tbl$EventMap[which(events_new_tbl$PROPDMG
>mean(events_new_tbl$PROPDMG))]),col = heat.colors(10),las=2, border =
par("fg"), legend.text = TRUE,main = "Impact of Events on \neconomy \n -
Property Damages ")

barplot(height = events_new_tbl$CROPDMG[which(events_new_tbl$CROPDMG
>mean(events_new_tbl$CROPDMG))],names.arg =
(events_new_tbl$EventMap[which(events_new_tbl$CROPDMG
>mean(events_new_tbl$CROPDMG))]),col = heat.colors(10),las=2, border =
par("fg"), legend.text = TRUE,main ="\n\n- Crop")
```



Analysis 2

The graph shows the impact of natural events on economic damage, causing 92 % of Property Damage and 93 % of Crop Damage, Events responsible are Tornado Flood Lightning ThunderstormWind HighWind MarineHail

Question 3

Consider writing your report as if it were to be read by a government or municipal manager who might be responsible for preparing for severe weather events and will need to prioritize resources for different types of events. However, there is no need to make any specific recommendations in your report.

```
actionableEvents <- unique(UniqueCauses,UniqueCauses2)
```

Analysis 3

The research on the events / natural calamities leading to people and economic loss leads us to give highest attention to the event Tornado Heat Flood HeavySnow Lightning ThunderstormWind RipCurrent , they being mostly responsible for 92 %property damage , 93 % crop damage, 82 %Fatlities, and 89 % Injuries of all the natural events.

We need focus for measures for Pre Event, During the event and Post Event restruction. In revival, the focus should be on the erection of facilities of greater competence than those built in pre-disaster stage.

For Tornados which is a heavy impact Natural Calamity , If possible to determine, it is very desirable to include in the event narrative the type of thunderstorm that was associated with the tornado, such as high-precipitation supercell, low precipitation supercell, non-supercell thunderstorm, line thunderstorm, bookend vortex, etc.

There can also be a proposal made for improvement of event reporting format specially for the major loss causing events. This can lead to more accurate reporting of data and better analysis.

Results

The research on the events / natural calamities leading to people and economic loss leads us to give highest attention to the event Tornado Heat Flood HeavySnow Lightning ThunderstormWind RipCurrent , they being mostly responsible for the 92 %property damage , 93 % crop damage, 82 %Fatlities, and 89 % Injuries of all the natural events. We need focus for measures for Pre Event, During the event and Post Event restruction. In revival, the focus should be on the erection of facilities of greater competence than those built in pre-disaster stage.