USA NOAA Storm Database Event Impacts Upon Health and the Economy

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

This project involves exploring the U.S. National Oceanic and Atmospheric Administration’s (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size.

Storm Data [47Mb] Unzipped Size (397MB)

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

Questions answered

The data analysis addresses the following questions:

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?

Data Processing

The code for loading and processing the storm data for analysis follows this brief outline of what it achieves:-

- Load the data (this can be time consuming first time round so cache=TRUE is set)
- We are answering 2 basic questions as described above so it is my opinion that we need to get the fatalities and injuries as an indicator of the health effects (this is crude I know but it is difficult to put a value on the psychological effects that these fatalities/injuries subsequently may also have had). We are also trying to determine the economic effects of the storm analysis so totalling both the property and crop related damages whilst also crude (and does not take account of inflation) is the best I can do with the provided dataset in the time given.
- There is an inherent problem in the provided dataset which relates to the uniqueness of the event types provided, this is a very large problem since the storm FAQ sites 48 event types but the dataset contains hundreds so I've devised a means to converting with pure best guessing based upon the wording used for each event that has occurred and tried to map it one to the agreed 48 event types as per the storm FAQ documentation.
- I've then decided to provide very simply insights into the top 10 event types for each of the health (fatalities/injuries) and the economic effects with simple charts.

The code section below shows all of the above mentioned stage from start to obtaining the top10 datasets.

- Load the data (this can be time consuming first time round so cache=TRUE is set)

```
## Load the storm data into memory - time consuming (397.8 MB - 902297 obs. of 37 vars)
setwd("~/GitHub/RepData_PeerAssessment2")
stormdata <- read.table("repdata-data-StormData.csv.bz2", sep=",", header=TRUE)
```

- We are answering 2 basic questions as described above so it is my opinion that we need to get the fatalities and injuries as an indicator of the health effects (this is crude I know but it is difficult to put a value on the phychological effects and subsequents these fatalaties/injuries may also have had). We are also trying to determine the economic effects of the storm analysis so totaling both the property and crop related damages whilst also crude (and does not take account of inflation) is the best I can do with the provided dataset.

```
## reduce the dataset to what interests us
## 1. health effects (fatalities/injuries)
## 2. economic effects (property/crop damage)
sd_health <- subset(stormdata, FATALITIES>0 | INJURIES>0)
sd_health$EVTYPE <- toupper(sd_health$EVTYPE)
sd_econ <- subset(stormdata, PROPDMG>0 | CROPDMG>0)
sd_econ$EVTYPE <- toupper(sd_econ$EVTYPE)



## aggregate the health numbers
## I've decided the FATALITIES be made to stand out
sdhf <- aggregate(FATALITIES ~ EVTYPE, data=sd_health, sum)
## I've decided the INJURIES be made to stand out
sdhi <- aggregate(INJURIES ~ EVTYPE, data=sd_health, sum)

## aggregate the economic numbers (note PROPDMGEXP/CROPDMGEXP adjustments)
## need to put a value on the economic values first
## need to multiply by (100 for H, 1000 for K, 1000000 for M, 1000000000 for B)
## if K/k, M/m or B/b don't exist then leave values as they are
sd_econ$PROPDMGEXP <- toupper(sd_econ$PROPDMGEXP)
sd_econ$PROPDMGVAL <- sd_econ$PROPDMG*ifelse(sd_econ$PROPDMGEXP=="K", 1000, ifelse(sd_econ$PROPDMGEXP=="M",
1000000, ifelse(sd_econ$PROPDMGEXP=="B", 1000000000, ifelse(sd_econ$PROPDMGEXP=="H", 100, sd_econ$PROPDMG))
))
## for +/-/?/K/M/B simply leave the values as they are, where it is a number
## 1 - 8 multiply the value by 10^(the number)
```

```
sd_econ$PROPDMGVAL <- sd_econ$PROPDMGVAL * ifelse(grepl("12345678",sd_econ$PROPDMGEXP), as.numeric(sd_econ$
PROPDMGEXP), 1 )
## Crop Damage calculation
sd_econ$CROPDMGVAL <- sd_econ$CROPDMG*ifelse(sd_econ$CROPDMGEXP=="K", 1000, ifelse(sd_econ$CROPDMGEXP=="M",
1000000, ifelse(sd_econ$CROPDMGEXP=="B", 1000000000, ifelse(sd_econ$CROPDMGEXP=="H", 100, sd_econ$CROPDMGEX
P))))
## for +/-/?/K/M/B simply leave the values as they are, where it is a number
## 1 - 8 multiply the value by 10^(the number)
sd_econ$CROPDMGVAL <- sd_econ$CROPDMGVAL * ifelse(grepl("12345678",sd_econ$CROPDMGEXP), as.numeric(sd_econ$
CROPDMGEXP), 1 )

## aggregate the economic numbers
sde <- aggregate(PROPDMGVAL + CROPDMGVAL ~ EVTYPE, data=sd_econ, sum)
## change the column name
colnames(sde)[2] <- "TOTALDMGVAL"
```

- There is an inherent problem in the provided dataset which relates to the uniqueness of the event types provided, this is a very large problem since the storm FAQ sites 48 event types but the dataset contains hundreds so I've devised a means to converting with pure best guessing based upon the wording used for each event that has occurred and tried to map it one to the agreed 48 event types as per the storm FAQ documentation.

```
## the EVTYPE aggregation for sdhf, sdhi and sde is very difficult to deal with based on the
## number of spelling mistakes and the fact that there is a recommended set of 48 naming conventions
## in use when logging these types of meteorological events.
## So. I put together the following function to help decipher the entered information as an attempt
## to coerse the EVTYPE to conform more against the accepted 48 types of events expected.
## As an example when run against the aggregate economic set of data which resulted in 397 unique EVTYPEs
## we get down to under 48 EVTYPEs, the function is here for review purposes too and essentially it takes
```

```
## each word of the logged EVTYPE (removes S and ING from the end of each word) and then filters the agreed 
48
## event types to see if the word is included.  This results in a set of agreed 48 events of which an event

## could appear more than one, I favour the larger frequency of occurrence of an event name and simply pick
## the topmost event (even if it does clash with another event of the same frequency).  If no event matches 
then
## I return OTHER as the event.
getBestEVTYPE <- function( sEVTYPE )  {

  ## create set of event type groups we are interested in reporting
  ## the breakdown of both the health and economic effects of the storm data
  ## NOTE: Taken from Storm Events-FAQ
  ##       http://www.ncdc.noaa.gov/stormevents/pd01016005curr.pdf
  ##
  ## 48 agreed event types for analysis
  evt_groups <- c("Astronomical Low Tide", "Avalanche", "Blizzard",
                  "Coastal Flood", "Cold/Wind Chill", "Debris Flow",
                  "Dense Fog", "Dense Smoke", "Drought",
                  "Dust Devil", "Dust Storm", "Excessive Heat",
                  "Extreme Cold/Wind Chill", "Flash Flood", "Flood",
                  "Frost/Freeze", "Funnel Cloud", "Freezing Fog",
                  "Hail", "Heat", "Heavy Rain", "Heavy Snow",
                  "High Surf", "High Wind", "Hurricane",
                  "Ice Storm", "Lake-Effect Snow", "Lakeshore Flood",
                  "Lightning", "Marine Hail", "Marine High Wind",
                  "Marine Strong Wind", "Marine Thunderstorm Wind",
                  "Rip Current", "Seiche", "Sleet",
```

```r
                "Storm Surge/Tide", "Strong Wind", "Thunderstorm Wind",
                "Tornado Typhoon", "Tropical Depression", "Tropical Storm",
                "Tsunami", "Volcanic Ash", "Waterspout",
                "Wildfire", "Winter Storm", "Winter Weather" )
evt_groups <- toupper(evt_groups)

##sEVTYPE <- "HIGH SURF ADVISORY"    ## set manually here for testing

sEVTYPE <- as.character(sEVTYPE)
## tidy up the logged EVTYPE so we process each word - still can fail if words don't match afterwards
sEVTYPE <- gsub("/", " ", sEVTYPE)
sEVTYPE <- gsub("\\)", " ", sEVTYPE)
sEVTYPE <- gsub("\\(", " ", sEVTYPE)
sEVTYPE <- gsub("\\.", " ", sEVTYPE)
sEVTYPE <- gsub("\\\\", " ", sEVTYPE)
sEVTYPE <- gsub("-", " ", sEVTYPE)

## split into words, get a vector of words back
vStr <- unlist(strsplit(sEVTYPE, " "))

bestMatch <- ""
iMatches <- 0
matched_evtypes <- c()
## loop through EVTYPE words
for(word in vStr) {
  ## some words may be empty due to double/treble SPACing
  if (nchar(word)>0) {
    ## remove ING or S at the end of the word otherwise it won't match (eg. STORMS, RAINING)
```

```
      if (substr(word, nchar(word)-3+1, nchar(word))=="ING") {
          word <- substr(word, 1, nchar(word)-3)
      } else if (substr(word, nchar(word)-1+1, nchar(word))=="S") {
          word <- substr(word, 1, nchar(word)-1)
      }
      ## is the word in any of the event groups
      ## go through groups and build matching word groups
      matched_evtypes <- c(matched_evtypes, evt_groups[grepl(word,evt_groups)])
    }
  }
  best_evtype <- "OTHER" ## default return value where words didn't match anything

  if (length(matched_evtypes)>0) {
    ## create a data frame containing a table of frequencies of matching words
    ## from the event type
    mevt <- as.data.frame(table(matched_evtypes))
    ## sort by frequency - descending - to give top matching event type
    mevt <- mevt[order(mevt$Freq, decreasing=TRUE),]
    ## return best matching event type name - it could still be wrong but better to have 48 than 205/397 ev
ents
    best_evtype <- as.character(mevt[1,c(1)])

  }

  best_evtype
}
```

- Now you will see the application of the above function the datasets of interest

```
## reduces 205 EVTYPEs to 39 N_EVTYPEs
sdhf$N_EVTYPE <- lapply(as.character(sdhf$EVTYPE), getBestEVTYPE)
## reduces 205 EVTYPEs to 39 N_EVTYPEs
sdhi$N_EVTYPE <- lapply(as.character(sdhi$EVTYPE), getBestEVTYPE)
## reduces 397 EVTYPEs to 43 N_EVTYPEs
sde$N_EVTYPE <- lapply(as.character(sde$EVTYPE), getBestEVTYPE)
```

- After application of the event type reduction we must redo the aggregation part as the event type has changed

```
## now we have the results (sdhf, sdhi and sde) and it is time to present some analysis
## but first we must re-aggregate the data again to reduce the datasets to within the 48 accepted event typ
es
sde$N_EVTYPE <- as.factor(as.character(sde$N_EVTYPE))      ## make factor to aggregate
sde <- aggregate(TOTALDMGVAL ~ N_EVTYPE, data=sde, sum)    ## aggregate
sde <- sde[order(sde$TOTALDMGVAL, decreasing=TRUE),]       ## sort by damage value
sde$Position <- rep(1:nrow(sde))

sdetop10 <- sde[1:10,]                                     ## get top10 for analysis
## divide damage value by 1000000 (million) to view the table results better
sdemill <- sde
sdemill$TOTALDMGVAL <- sdemill$TOTALDMGVAL/1000000
colnames(sdemill)[2] <- "DAMAGE-millions"

sdhf$N_EVTYPE <- as.factor(as.character(sdhf$N_EVTYPE))
sdhf <- aggregate(FATALITIES ~ N_EVTYPE, data=sdhf, sum)
sdhf <- sdhf[order(sdhf$FATALITIES, decreasing=TRUE),]       ## sort by number of fatalities
sdhf$Position <- rep(1:nrow(sdhf))
sdhftop10 <- sdhf[1:10,]                                      ## get top10 for analysis
```

```
sdhi$N_EVTYPE <- as.factor(as.character(sdhi$N_EVTYPE))
sdhi <- aggregate(INJURIES ~ N_EVTYPE, data=sdhi, sum)
sdhi <- sdhi[order(sdhi$INJURIES, decreasing=TRUE),]       ## sort by number of injuries
sdhi$Position <- rep(1:nrow(sdhi))
sdhitop10 <- sdhi[1:10,]                                     ## get top10 for analysis
```

Results

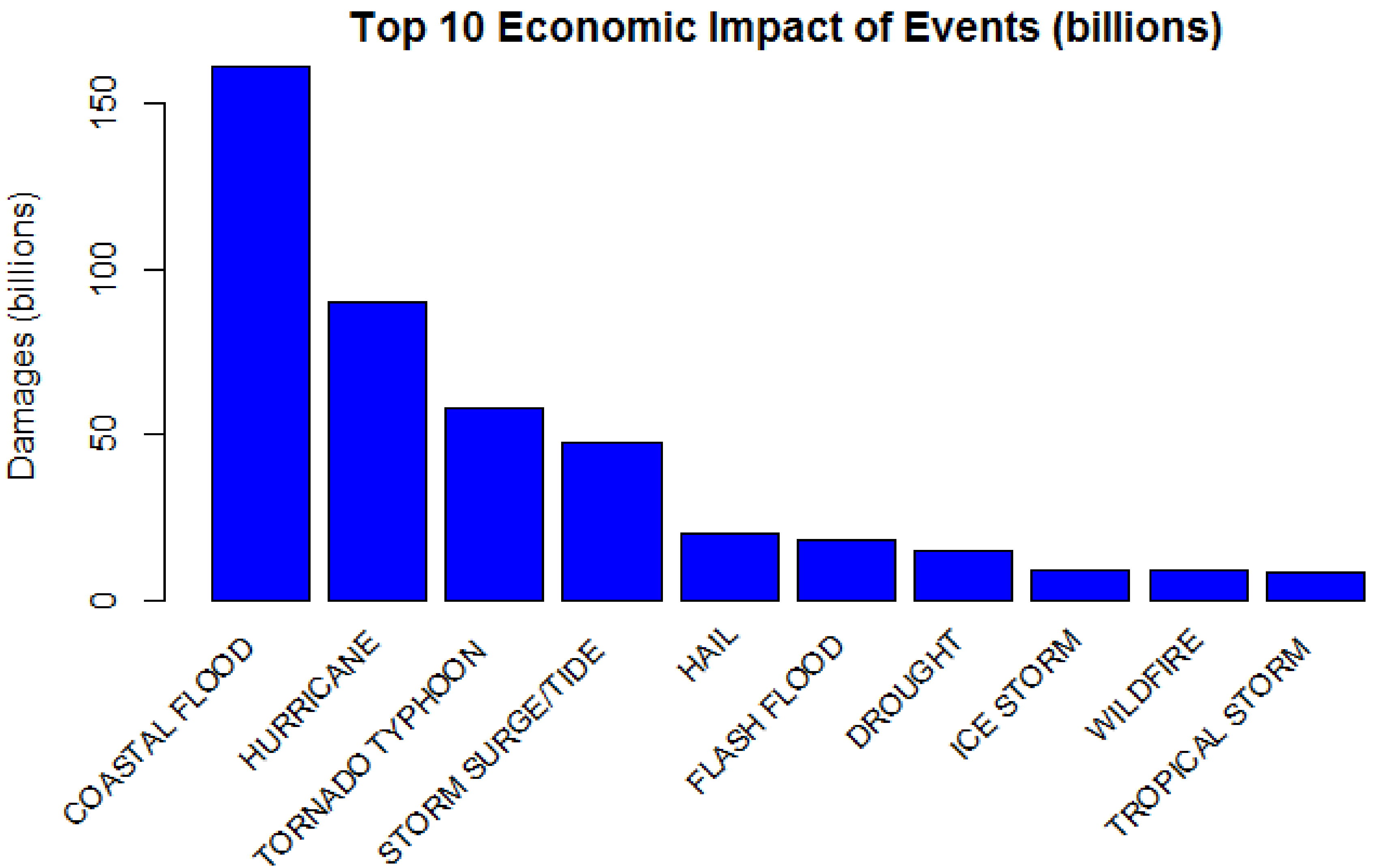
I've then decided to provide very simply insights into the top 10 event types for each of the health (fatalities/injuries) and the economic effects with simple charts.

To my untrained eye based on each of the charts presented I would conclude that TORNADO TYPHOON, HURRICAN, STORM SURGES, COASTAL AND FLASH FLOODing have major impact upon both the health and economic well-being of people in general as they are typically associated events (the association is guessing without further looking at the data in more detail). This is followed by EXCESSIVE HEAT events where health impacts are felt most.

Top 10 Events having most Economic Impact

See the Appendix section for full dataset details.

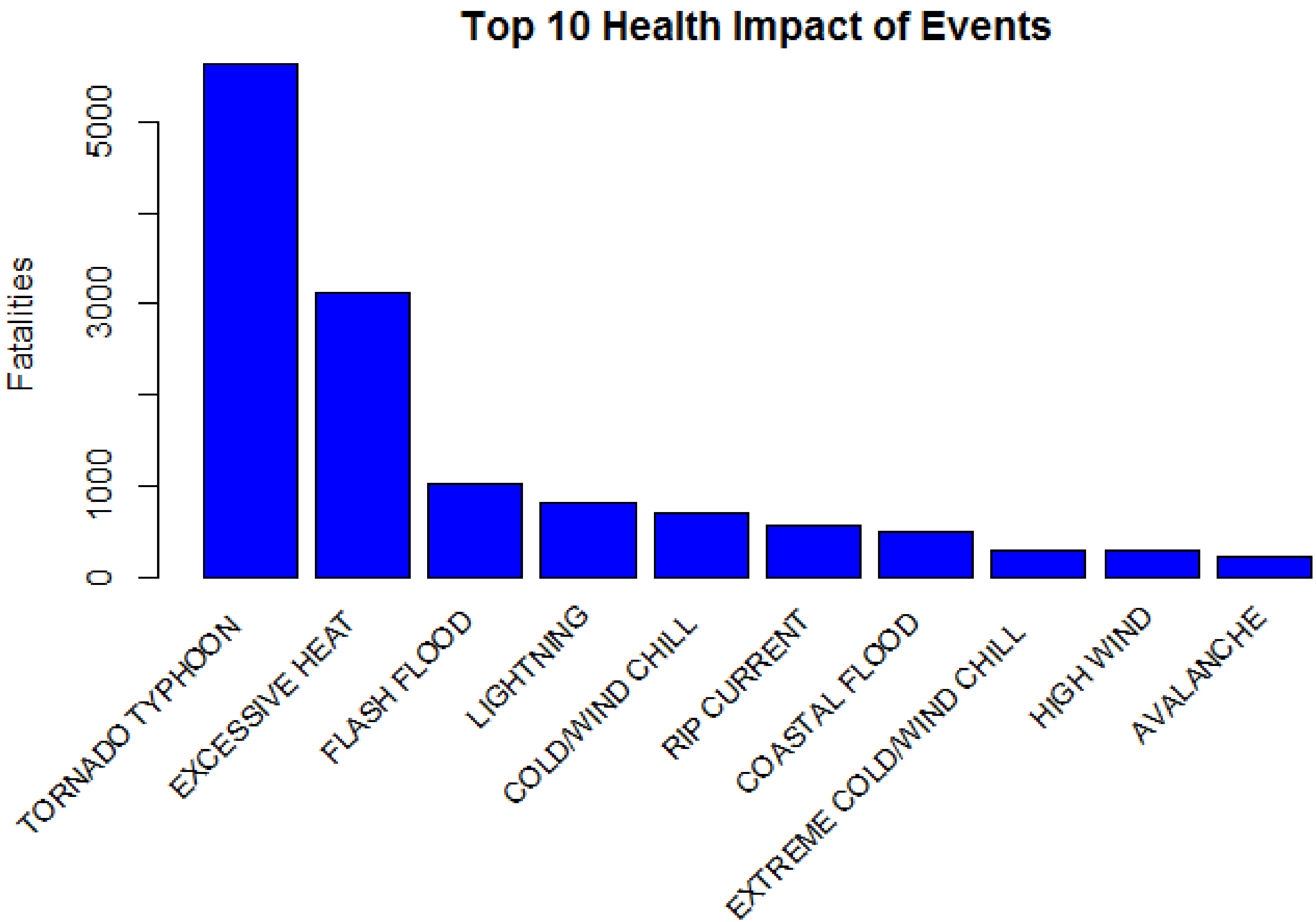
```
par(mar=c(10,4,2,1))
bp <- barplot(sdetop10$TOTALDMGVAL/1000000000,axes=FALSE,col="blue",main="Top 10 Economic Impact of Events
(billions)", ylab="Damages (billions)")
labels <- paste(sdetop10$N_EVTYPE,"")
text(bp,par("usr")[3],labels=labels,srt=45,adj=c(1.1,1.1),xpd=TRUE,cex=.9)
axis(2)
```



Top 10 Events having most Health Impact due to Fatalities

See the Appendix section for full dataset details.

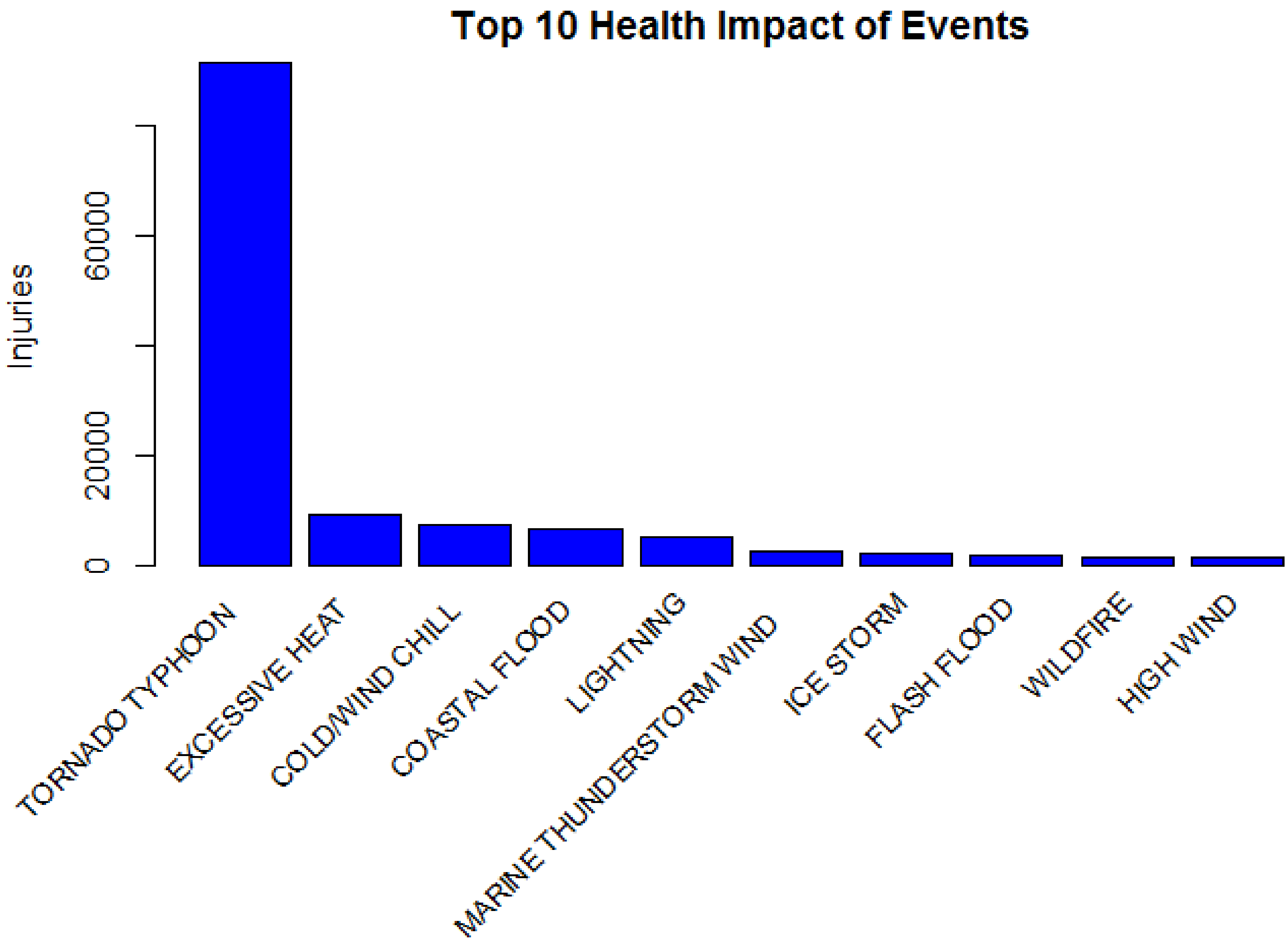
```
par(mar=c(10,4,2,1))
bp <- barplot(sdhftop10$FATALITIES,axes=FALSE,col="blue",main="Top 10 Health Impact of Events", ylab="Fatal
ities")
labels <- paste(sdhftop10$N_EVTYPE,"")
text(bp,par("usr")[3],labels=labels,srt=45,adj=c(1.1,1.1),xpd=TRUE,cex=.9)
axis(2)
```



Top 10 Events having most Health Impact due to Injuries

See the Appendix section for full dataset details.

```
par(mar=c(10,4,2,1))
bp <- barplot(sdhitop10$INJURIES,axes=FALSE,col="blue",main="Top 10 Health Impact of Events", ylab="Injuries")
labels <- paste(sdhitop10$N_EVTYPE,"")
text(bp,par("usr")[3],labels=labels,srt=45,adj=c(1.1,1.1),xpd=TRUE,cex=.9)
axis(2)
```



- The analysis document must have at least one figure containing a plot.
- Your analyis must have no more than three figures. Figures may have multiple plots in them (i.e. panel plots), but there cannot be more than three figures total.

- You must show all your code for the work in your analysis document. This may make the document a bit verbose, but that is okay. In general, you should ensure that echo = TRUE for every code chunk (this is the default setting in knitr).

Appendix - Full Event Summary Datasets

Events having most Economic Impact

```
sde
```

```
##                     N_EVTYPE TOTALDMGVAL Position
## 4              COASTAL FLOOD   1.614e+11        1
## 22                 HURRICANE   9.015e+10        2
## 35           TORNADO TYPHOON   5.801e+10        3
## 34          STORM SURGE/TIDE   4.797e+10        4
## 17                      HAIL   2.038e+10        5
## 13               FLASH FLOOD   1.844e+10        6
## 8                    DROUGHT   1.502e+10        7
## 23                 ICE STORM   8.981e+09        8
## 41                  WILDFIRE   8.900e+09        9
## 37            TROPICAL STORM   8.409e+09       10
## 30 MARINE THUNDERSTORM WIND   7.086e+09       11
## 21                 HIGH WIND   6.735e+09       12
## 42              WINTER STORM   6.717e+09       13
## 5            COLD/WIND CHILL   5.446e+09       14
## 18                HEAVY RAIN   4.074e+09       15
## 15              FROST/FREEZE   2.016e+09       16
## 12  EXTREME COLD/WIND CHILL   1.407e+09       17
```

```
## 19              HEAVY SNOW  1.105e+09      18
## 11          EXCESSIVE HEAT  1.068e+09      19
## 26               LIGHTNING  9.459e+08      20
## 3                 BLIZZARD  7.714e+08      21
## 31                   OTHER  6.719e+08      22
## 29       MARINE STRONG WIND  2.515e+08      23
## 38                 TSUNAMI  1.441e+08      24
## 20               HIGH SURF  1.006e+08      25
## 43          WINTER WEATHER  4.230e+07      26
## 24        LAKE-EFFECT SNOW  4.018e+07      27
## 6                DENSE FOG  2.283e+07      28
## 14            FREEZING FOG  1.365e+07      29
## 1     ASTRONOMICAL LOW TIDE  9.772e+06      30
## 40              WATERSPOUT  9.564e+06      31
## 10              DUST STORM  9.199e+06      32
## 25         LAKESHORE FLOOD  7.570e+06      33
## 28        MARINE HIGH WIND  6.718e+06      34
## 2                AVALANCHE  3.722e+06      35
## 36      TROPICAL DEPRESSION  1.737e+06      36
## 33                  SEICHE  9.800e+05      37
## 9               DUST DEVIL  7.391e+05      38
## 39            VOLCANIC ASH  5.000e+05      39
## 16            FUNNEL CLOUD  1.996e+05      40
## 32             RIP CURRENT  1.630e+05      41
## 7              DENSE SMOKE  1.000e+05      42
## 27             MARINE HAIL  5.400e+04      43
```

Events having most Health Impact due to Fatalities

```
sdhf
```

```
##                     N_EVTYPE FATALITIES Position
## 33           TORNADO TYPHOON       5636        1
## 11            EXCESSIVE HEAT       3136        2
## 13               FLASH FLOOD       1035        3
## 24                 LIGHTNING        817        4
## 5            COLD/WIND CHILL        708        5
## 30               RIP CURRENT        577        6
## 4              COASTAL FLOOD        491        7
## 12  EXTREME COLD/WIND CHILL        304        8
## 21                 HIGH WIND        296        9
## 2                  AVALANCHE        224       10
## 38              WINTER STORM        216       11
## 28 MARINE THUNDERSTORM WIND        211       12
## 20                 HIGH SURF        156       13
## 19                HEAVY SNOW        145       14
## 29                     OTHER        145       15
## 22                 HURRICANE        133       16
## 27        MARINE STRONG WIND        125       17
## 18                HEAVY RAIN        114       18
## 3                   BLIZZARD        101       19
## 23                 ICE STORM         96       20
## 37                  WILDFIRE         90       21
## 7                  DENSE FOG         80       22
```

```
## 34          TROPICAL STORM         66       23
## 39          WINTER WEATHER         61       24
## 17                    HAIL         40       25
## 35                 TSUNAMI         33       26
## 32        STORM SURGE/TIDE         24       27
## 10              DUST STORM         22       28
## 8                  DROUGHT         16       29
## 14            FREEZING FOG         11       30
## 26         MARINE HIGH WIND         10       31
## 1     ASTRONOMICAL LOW TIDE          8       32
## 25             MARINE HAIL          8       33
## 36              WATERSPOUT          3       34
## 9               DUST DEVIL          2       35
## 15            FROST/FREEZE          2       36
## 31                   SLEET          2       37
## 6              DEBRIS FLOW          1       38
## 16            FUNNEL CLOUD          0       39
```

Events having most Health Impact due to Injuries

```
sdhi
```

```
##                     N_EVTYPE INJURIES Position
## 33           TORNADO TYPHOON    91412        1
## 11            EXCESSIVE HEAT     9232        2
## 5            COLD/WIND CHILL     7229        3
## 4              COASTAL FLOOD     6804        4
```

```
## 24                LIGHTNING     5231        5
## 28 MARINE THUNDERSTORM WIND     2452        6
## 23                ICE STORM     2152        7
## 13              FLASH FLOOD     1802        8
## 37                 WILDFIRE     1608        9
## 21                HIGH WIND     1517       10
## 17                     HAIL     1371       11
## 38             WINTER STORM     1338       12
## 22                HURRICANE     1328       13
## 19               HEAVY SNOW     1119       14
## 7                 DENSE FOG     1076       15
## 3                  BLIZZARD      805       16
## 29                    OTHER      567       17
## 39           WINTER WEATHER      540       18
## 30              RIP CURRENT      529       19
## 10               DUST STORM      440       20
## 34           TROPICAL STORM      383       21
## 27        MARINE STRONG WIND      323      22
## 18               HEAVY RAIN      297       23
## 12  EXTREME COLD/WIND CHILL      260       24
## 20                HIGH SURF      214       25
## 2                 AVALANCHE      170       26
## 35                  TSUNAMI      129       27
## 9                DUST DEVIL       43       28
## 32         STORM SURGE/TIDE       43       29
## 14             FREEZING FOG       38       30
## 36               WATERSPOUT       29       31
## 8                   DROUGHT       25       32
```

```
## 26         MARINE HIGH WIND      9       33
## 25              MARINE HAIL      7       34
## 15             FROST/FREEZE      3       35
## 16             FUNNEL CLOUD      3       36
## 1     ASTRONOMICAL LOW TIDE      0       37
## 6               DEBRIS FLOW      0       38
## 31                    SLEET      0       39
```