Reproducible Research: Peer Assessment 2

Impact of severe weather events on Public Health and Economy in the United States between 1950 and 2011. An Analysis based on NOAA Database.

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities in the United States. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern. Impacts of major storms and weather events were observed in the United States between 1950 and 2011. Analysis of the weather events and the impacts was based on the information contained in the NOAA database at the time. The data analysis addressed 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?

Analysis showed that tornado had the highest impact on population health since they caused most of the fatalities and injuries. While, flood and drought caused the highest damages on property and crops respectively.

Data Processing

The file containing the data for the analysis can be downloaded from the web site: "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2" The file was then unzipped and loaded into R.

```
# set working directory
setwd("C:/VION/Emmanuel/R_WD_Coursera/repdata_StormData")
```

Download data

```
# download file from URL
download.file("http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
              destfile = "stormData.csv.bz2")
```

```
# unzip file
# install.packages("R.utils")
library(R.utils)
```

```
bunzip2("stormData.csv.bz2", overwrite=T, remove=F)
```

Load data into R & explore data

```
stormData <- read.csv("stormData.csv")
# exploring the data contents
dim(stormData)
```

```
## [1] 902297     37
```

```
head(stormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
```

```
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
## 2         NA         0                       2.0   150 2   0          0
## 3         NA         0                       0.1   123 2   0          0
## 4         NA         0                       0.0   100 2   0          0
## 5         NA         0                       0.0   150 2   0          0
## 6         NA         0                       1.5   177 2   0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

```r
names(stormData)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Obtaining the most harmful weather events with respect to population health

Dataset required from stormData are EVTYPE, FATALITIES and INJURIES attributes

Geting the Top 10 events with highest fatalities. Aggregating the data by event:

```r
dataset_fatalities <- aggregate(FATALITIES ~ EVTYPE, data = stormData, sum, na.rm = TRUE)
dataset_fatalities <- dataset_fatalities[order(-dataset_fatalities$FATALITIES), ]
Top10Fatalities <- dataset_fatalities[1:10, ]
Top10Fatalities
```

```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
```

```
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

Geting the Top 10 events with highest injuries. Aggregating the data by event:

```r
dataset_injuries <- aggregate(INJURIES ~ EVTYPE, data = stormData, sum, na.rm = TRUE)
dataset_injuries = dataset_injuries[order(-dataset_injuries$INJURIES), ]
Top10Injuries <- dataset_injuries[1:10, ]
Top10Injuries
```

```
##                EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

Obtaining the events with the greatest economic consequences

Data transformation: The damage value is represented in two parts "-DMG" (numeric) and "-DMGEXP" (alphanumeric). The records in both PROPDMGEXP and CROPDMGEXP columns (Hundred (H) = 2, Thousand (K) = 3, Million (M) = 6, Billion (B) = 9) are converted into exponent values (numeric) and then used to calcualte the damage costs for property and crops.

```r
# Property damage values
stormData$PROPDMGEXP <- as.character(stormData$PROPDMGEXP)
stormData$PROPDMGEXP[toupper(stormData$PROPDMGEXP) == 'H'] <- "2"
stormData$PROPDMGEXP[toupper(stormData$PROPDMGEXP) == 'K'] <- "3"
stormData$PROPDMGEXP[toupper(stormData$PROPDMGEXP) == 'M'] <- "6"
stormData$PROPDMGEXP[toupper(stormData$PROPDMGEXP) == 'B'] <- "9"
stormData$PROPDMGEXP <- as.numeric(stormData$PROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

```r
stormData$PROPDMGEXP[is.na(stormData$PROPDMGEXP)] <- 0
stormData$TotalCost_PROPDMG <- stormData$PROPDMG * 10^stormData$PROPDMGEXP
```

```r
# Crop damage values
stormData$CROPDMGEXP <- as.character(stormData$CROPDMGEXP)
stormData$CROPDMGEXP[toupper(stormData$CROPDMGEXP) == 'H'] <- "2"
stormData$CROPDMGEXP[toupper(stormData$CROPDMGEXP) == 'K'] <- "3"
stormData$CROPDMGEXP[toupper(stormData$CROPDMGEXP) == 'M'] <- "6"
stormData$CROPDMGEXP[toupper(stormData$CROPDMGEXP) == 'B'] <- "9"
stormData$CROPDMGEXP <- as.numeric(stormData$CROPDMGEXP)
```

```
## Warning: NAs introduced by coercion
```

```r
stormData$CROPDMGEXP[is.na(stormData$CROPDMGEXP)] <- 0
stormData$TotalCost_CROPDMG <- stormData$CROPDMG * 10^stormData$CROPDMGEXP
```

Results

1. Across the United States, which types of events are most harmful with respect to population health?

Impact on Public Health

```r
# Get Top 10 weather events with highest fatalities
Top10Fatalities <- dataset_fatalities[1:10, ]
Top10Fatalities
```

```
##             EVTYPE FATALITIES
## 834        TORNADO       5633
## 130 EXCESSIVE HEAT       1903
## 153    FLASH FLOOD        978
## 275           HEAT        937
## 464      LIGHTNING        816
## 856      TSTM WIND        504
## 170          FLOOD        470
## 585    RIP CURRENT        368
## 359      HIGH WIND        248
## 19       AVALANCHE        224
```

```r
# Get Top 10 weather events with highest injuries
Top10Injuries <- dataset_injuries[1:10, ]
Top10Injuries
```

```
##                EVTYPE INJURIES
## 834           TORNADO    91346
## 856         TSTM WIND     6957
## 170             FLOOD     6789
## 130    EXCESSIVE HEAT     6525
## 464         LIGHTNING     5230
## 275              HEAT     2100
## 427         ICE STORM     1975
## 153       FLASH FLOOD     1777
## 760 THUNDERSTORM WIND     1488
## 244              HAIL     1361
```

```r
par(mfrow = c(1, 2), mar = c(12, 5, 3, 2), mgp = c(3, 1, 0), cex = 0.8, las = 3)
barplot(Top10Fatalities$FATALITIES, names.arg = Top10Fatalities$EVTYPE, col = 'blue',
        main = 'Fatalities Top 10 Weather Events', ylab = 'Number of Fatalities')
barplot(Top10Injuries$INJURIES, names.arg = Top10Injuries$EVTYPE, col = 'red',
        main = 'Injuries Top 10 Weather Events', ylab = 'Number of Injuries')
```

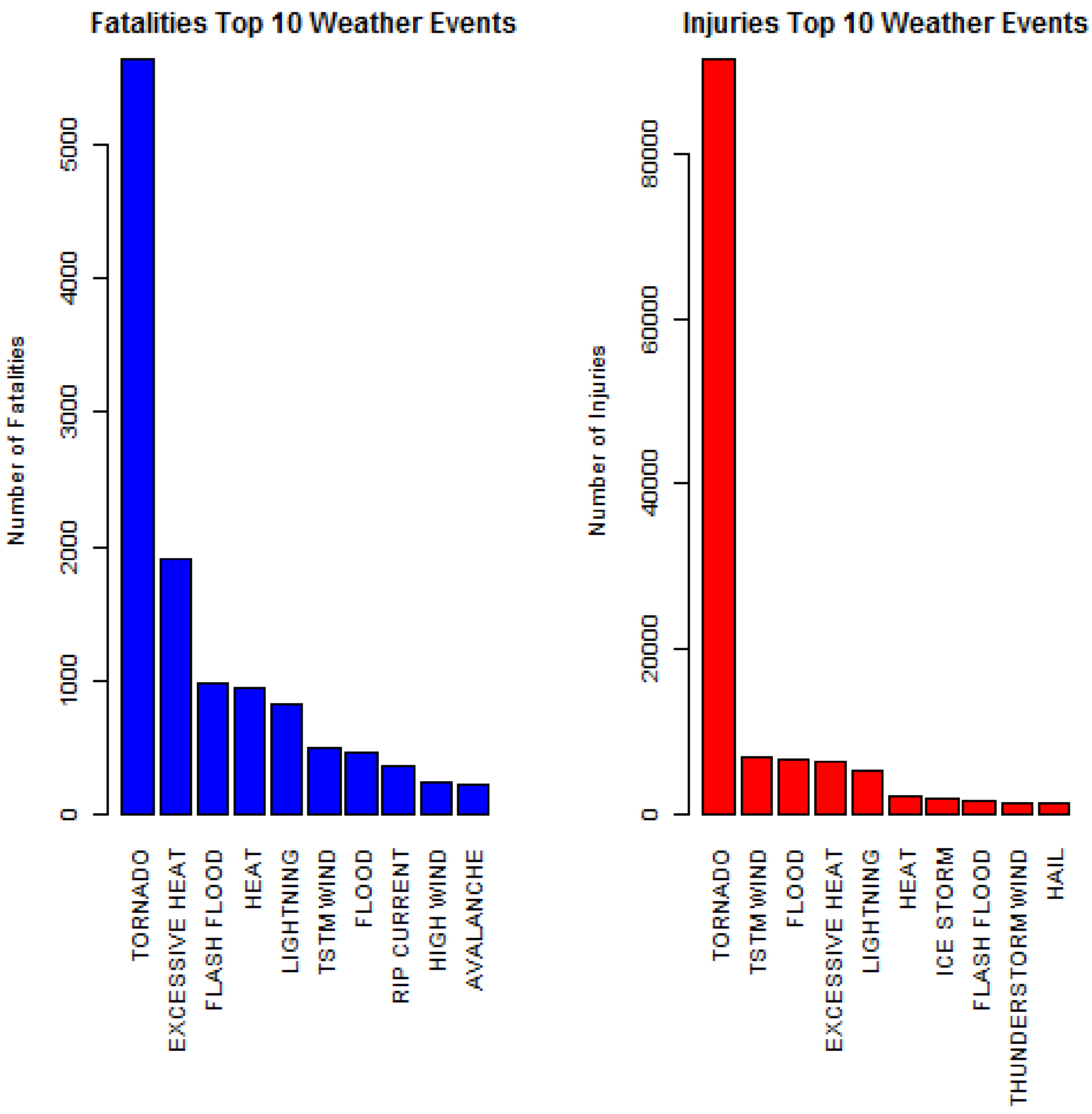

Figure 1: plot of chunk PHealth_Plot

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

Impact on Economy

```
# Get Top 10 events with highest property damage:
Top10Cost_PROPDMG <- aggregate(stormData$TotalCost_PROPDMG, by = list(stormData$EVTYPE), "sum")
names(Top10Cost_PROPDMG) <- c("Event", "Cost")
Top10Cost_PROPDMG <- Top10Cost_PROPDMG[order(-Top10Cost_PROPDMG$Cost), ][1:10, ]
Top10Cost_PROPDMG
```

```
##                 Event         Cost
## 170             FLOOD 144657709807
## 411 HURRICANE/TYPHOON  69305840000
## 834           TORNADO  56947380677
## 670       STORM SURGE  43323536000
## 153       FLASH FLOOD  16822673979
## 244              HAIL  15735267513
## 402         HURRICANE  11868319010
## 848    TROPICAL STORM   7703890550
## 972      WINTER STORM   6688497251
## 359         HIGH WIND   5270046295
```

```
# Get Top 10 events with highest crop damage:
Top10Cost_CROPDMG <- aggregate(stormData$TotalCost_CROPDMG, by = list(stormData$EVTYPE), "sum")
names(Top10Cost_CROPDMG) <- c("Event", "Cost")
Top10Cost_CROPDMG <- Top10Cost_CROPDMG[order(-Top10Cost_CROPDMG$Cost), ][1:10, ]
Top10Cost_CROPDMG
```

```
##                 Event        Cost
## 95            DROUGHT 13972566000
## 170             FLOOD  5661968450
## 590       RIVER FLOOD  5029459000
## 427         ICE STORM  5022113500
## 244              HAIL  3025954473
## 402         HURRICANE  2741910000
## 411 HURRICANE/TYPHOON  2607872800
## 153       FLASH FLOOD  1421317100
## 140      EXTREME COLD  1292973000
## 212      FROST/FREEZE  1094086000
```

```
par(mfrow = c(1, 2), mar = c(12, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(Top10Cost_PROPDMG$Cost/(10^9), las = 3, names.arg = Top10Cost_PROPDMG$Event,
        main = "Top 10 Events with Greatest Property Damages", ylab = "Cost of damages ($billions)",
        col = "light blue")
barplot(Top10Cost_CROPDMG$Cost/(10^9), las = 3, names.arg = Top10Cost_CROPDMG$Event,
        main = "Top 10 Events With Greatest Crop Damages", ylab = "Cost of damages ($ billions)",
        col = "pink")
```

Conclusion

Across the United States in the time period between 1950 and 2011, flood, drought, and hurricane/typhoon were the weather events with the greatest economic impact. While tornados were the most harmful with respect to population health.

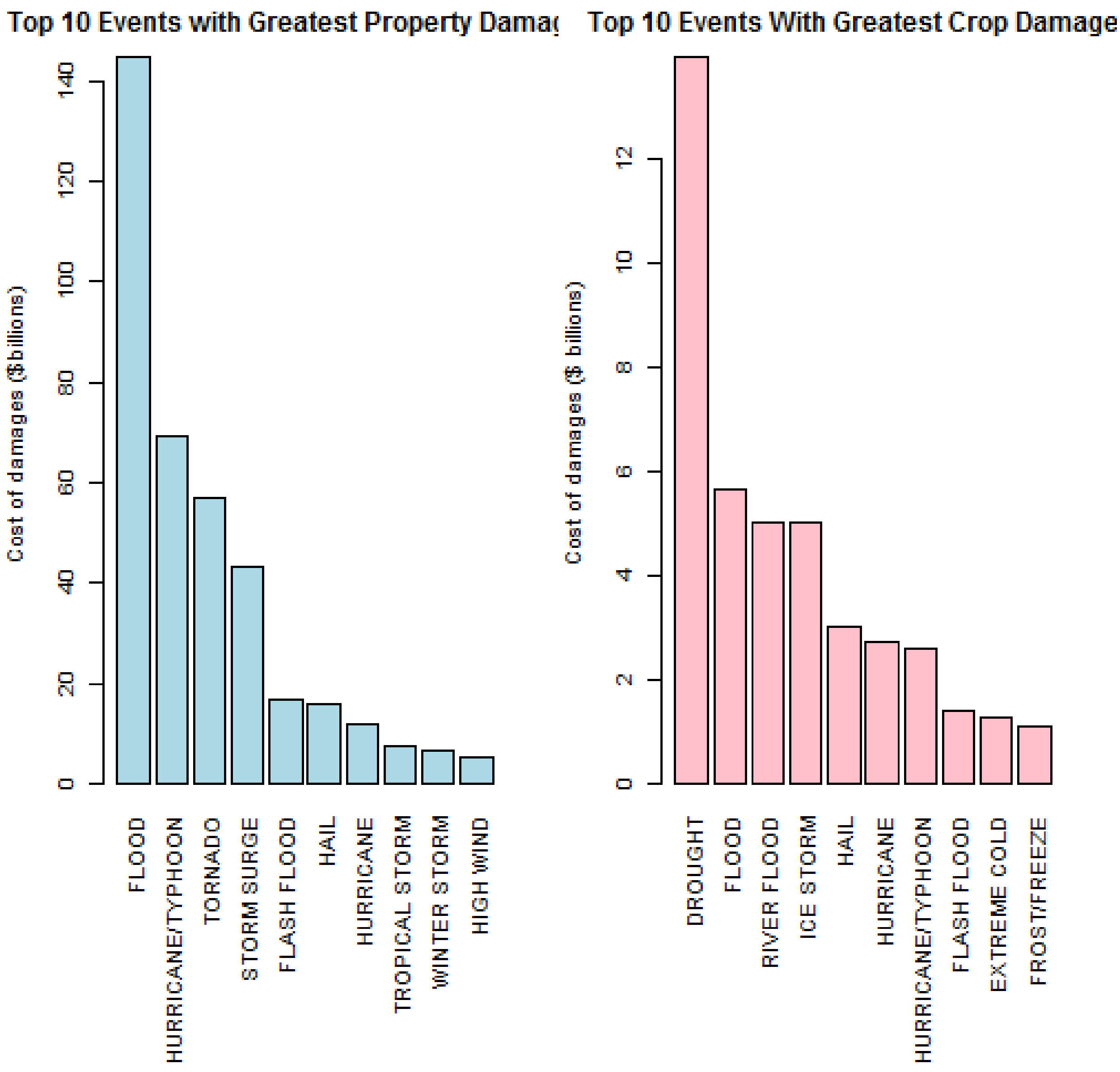

Figure 2: plot of chunk Economy_Plot