RepData_PeerAssessment2

Janus

Thursday, September 17, 2015

An Analysis Report on the Impact of Severe Weather Events on Public Health and Economy in the United States

```
setwd("C:/Users/ITSUPPORT/Desktop/R/A2")
library(R.utils)
```

Data processing

```
## Warning: package 'R.utils' was built under R version 3.2.2

## Loading required package: R.oo

## Warning: package 'R.oo' was built under R version 3.2.2

## Loading required package: R.methodsS3

## Warning: package 'R.methodsS3' was built under R version 3.2.2

## R.methodsS3 v1.7.0 (2015-02-19) successfully loaded. See ?R.methodsS3 for help.
## R.oo v1.19.0 (2015-02-27) successfully loaded. See ?R.oo for help.
##
## Attaching package: 'R.oo'
##
## The following objects are masked from 'package:methods':
##
##     getClasses, getMethods
##
## The following objects are masked from 'package:base':
##
##     attach, detach, gc, load, save
##
## R.utils v2.1.0 (2015-05-27) successfully loaded. See ?R.utils for help.
##
## Attaching package: 'R.utils'
##
## The following object is masked from 'package:utils':
##
##     timestamp
##
## The following objects are masked from 'package:base':
##
##     cat, commandArgs, getOption, inherits, isOpen, parse, warnings
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.2
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.2
```

```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
## Warning: package 'gridExtra' was built under R version 3.2.2
```

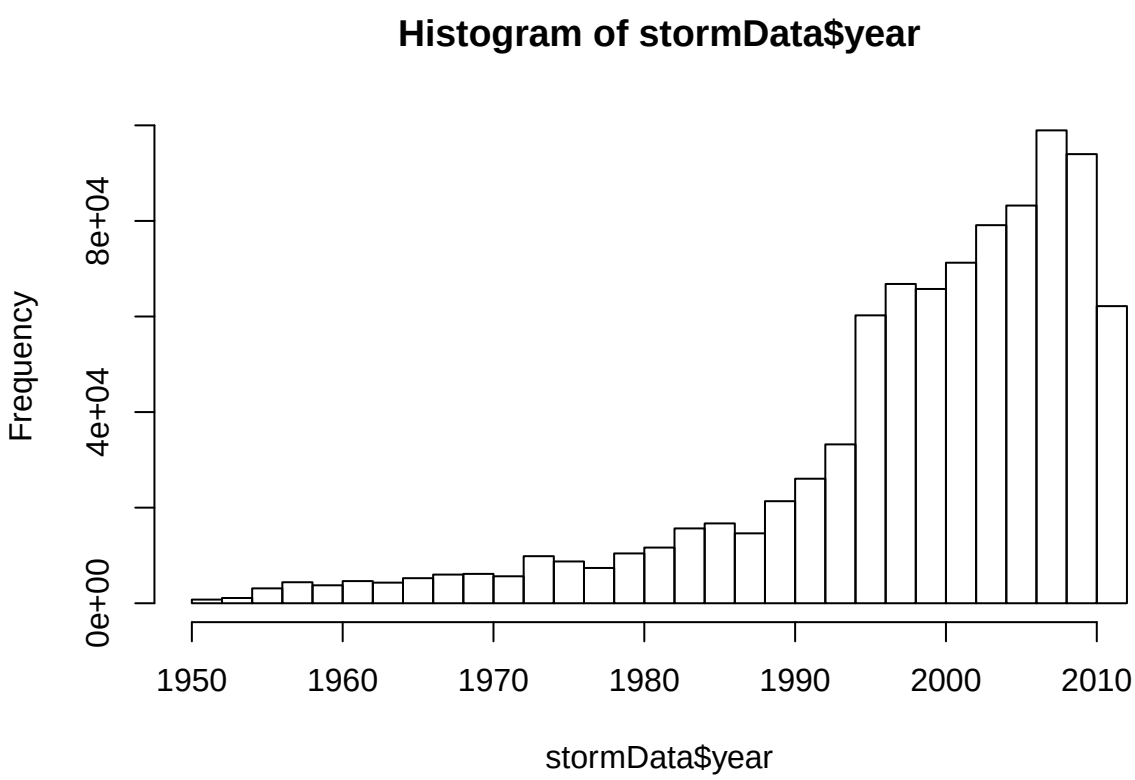
```
#####Read the data.
```

```
stormData <- read.csv("repdata-data-StormData.csv", sep = ",")
dim(stormData)
```

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

```
if (dim(stormData)[2] == 37) {
    stormData$year <- as.numeric(format(as.Date(stormData$BGN_DATE, format = "%m/%d/%Y %H:%M:%S"), "%Y"))
}
hist(stormData$year, breaks = 30)
```

The time period of the events begins in the year 1950 and ends in November 2011.



#####According to the histogram, the number of events significantly increases around 1995. By utilizing the subset of the data from 1990 to 2011.

```
storm <- stormData[stormData$year >= 1995, ]
dim(storm)
```

```
## [1] 681500     38
```

After subsetting, there are 681500 rows and 38 columns.

Impact on Public Health

```
sortHelper <- function(fieldName, top = 15, dataset = stormData) {
    index <- which(colnames(dataset) == fieldName)
    field <- aggregate(dataset[, index], by = list(dataset$EVTYPE), FUN = "sum")
    names(field) <- c("EVTYPE", fieldName)
    field <- arrange(field, field[, 2], decreasing = T)
    field <- head(field, n = top)
    field <- within(field, EVTYPE <- factor(x = EVTYPE, levels = field$EVTYPE))
    return(field)
}
```

```
fatalities <- sortHelper("FATALITIES", dataset = storm)
```

```
injuries <- sortHelper("INJURIES", dataset = storm)
```

We notices at the number of fatalities and injuries that are caused by several weather events.

Impact on Economy

```
convertHelper <- function(dataset = storm, fieldName, newFieldName) {
    totalLen <- dim(dataset)[2]
    index <- which(colnames(dataset) == fieldName)
    dataset[, index] <- as.character(dataset[, index])
    logic <- !is.na(toupper(dataset[, index]))
    dataset[logic & toupper(dataset[, index]) == "B", index] <- "9"
    dataset[logic & toupper(dataset[, index]) == "M", index] <- "6"
    dataset[logic & toupper(dataset[, index]) == "K", index] <- "3"
    dataset[logic & toupper(dataset[, index]) == "H", index] <- "2"
    dataset[logic & toupper(dataset[, index]) == "", index] <- "0"
    dataset[, index] <- as.numeric(dataset[, index])
    dataset[is.na(dataset[, index]), index] <- 0
    dataset <- cbind(dataset, dataset[, index - 1] * 10^dataset[, index])
    names(dataset)[totalLen + 1] <- newFieldName
    return(dataset)
}

storm <- convertHelper(storm, "PROPDMGEXP", "propertyDamage")
```

We will convert the property damage and crop damage data into comparable numerical forms according to the meaning of units described in the code book.

```
## Warning in convertHelper(storm, "PROPDMGEXP", "propertyDamage"): NAs
## introduced by coercion
```

```
storm <- convertHelper(storm, "CROPDMGEXP", "cropDamage")
```

```
## Warning in convertHelper(storm, "CROPDMGEXP", "cropDamage"): NAs introduced
## by coercion
```

```
names(storm)
```

```
##  [1] "STATE__"        "BGN_DATE"       "BGN_TIME"       "TIME_ZONE"
##  [5] "COUNTY"         "COUNTYNAME"     "STATE"          "EVTYPE"
##  [9] "BGN_RANGE"      "BGN_AZI"        "BGN_LOCATI"     "END_DATE"
## [13] "END_TIME"       "COUNTY_END"     "COUNTYENDN"     "END_RANGE"
```

```
## [17] "END_AZI"        "END_LOCATI"     "LENGTH"         "WIDTH"
## [21] "F"              "MAG"            "FATALITIES"     "INJURIES"
## [25] "PROPDMG"        "PROPDMGEXP"     "CROPDMG"        "CROPDMGEXP"
## [29] "WFO"            "STATEOFFIC"     "ZONENAMES"      "LATITUDE"
## [33] "LONGITUDE"      "LATITUDE_E"     "LONGITUDE_"     "REMARKS"
## [37] "REFNUM"         "year"           "propertyDamage" "cropDamage"
```

```
options(scipen=999)
property <- sortHelper("propertyDamage", dataset = storm)
crop <- sortHelper("cropDamage", dataset = storm)
```

Results

```
fatalities
```

Looking at the impact on public health, there are two sorted lists of severe weather events shown below by the number of people badly affected.

```
##               EVTYPE FATALITIES
## 1     EXCESSIVE HEAT       1903
## 2            TORNADO       1545
## 3        FLASH FLOOD        934
## 4               HEAT        924
## 5          LIGHTNING        729
## 6              FLOOD        423
## 7        RIP CURRENT        360
## 8          HIGH WIND        241
## 9          TSTM WIND        241
## 10         AVALANCHE        223
## 11      RIP CURRENTS        204
## 12      WINTER STORM        195
## 13         HEAT WAVE        161
## 14 THUNDERSTORM WIND        131
## 15      EXTREME COLD        126
```

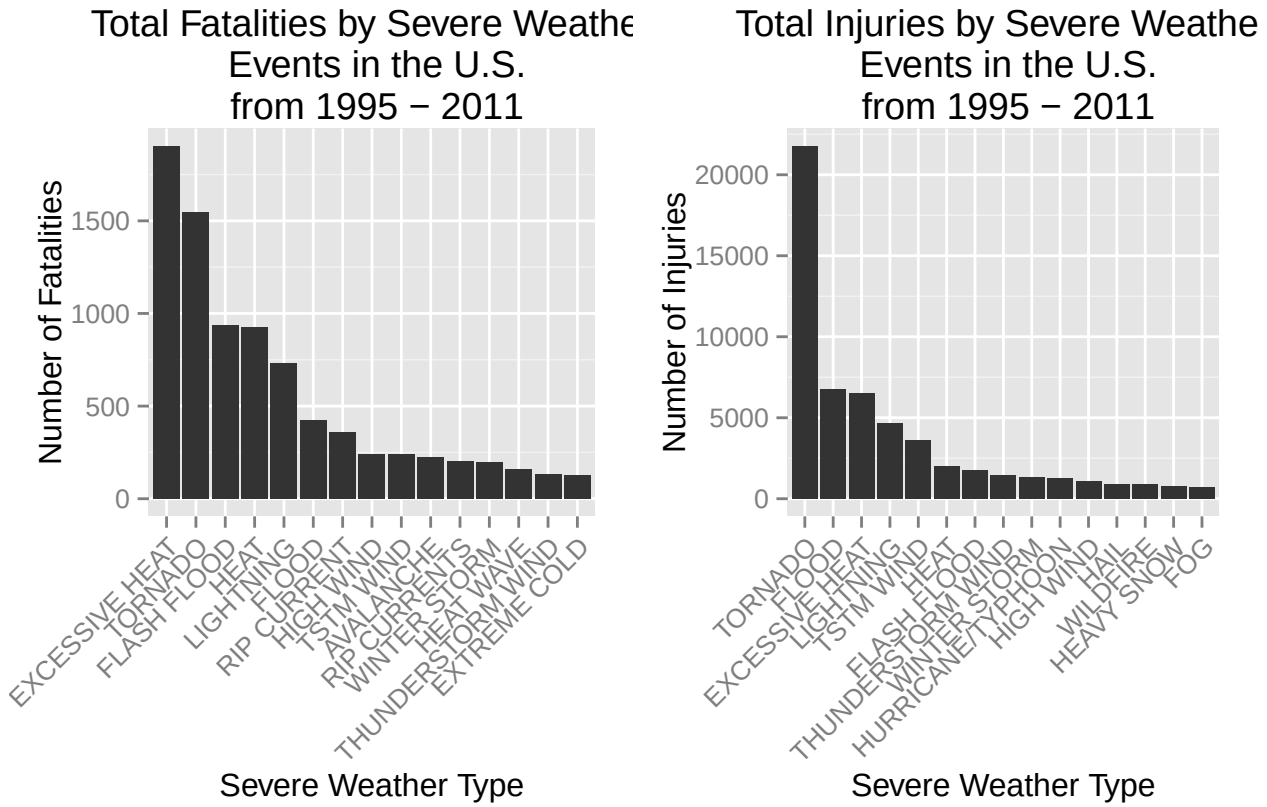
```
injuries
```

```
##               EVTYPE INJURIES
## 1            TORNADO    21765
## 2              FLOOD     6769
## 3     EXCESSIVE HEAT     6525
## 4          LIGHTNING     4631
## 5          TSTM WIND     3630
## 6               HEAT     2030
## 7        FLASH FLOOD     1734
## 8  THUNDERSTORM WIND     1426
## 9       WINTER STORM     1298
## 10 HURRICANE/TYPHOON     1275
## 11         HIGH WIND     1093
```

```
## 12              HAIL      916
## 13         WILDFIRE      911
## 14       HEAVY SNOW      751
## 15              FOG      718
```

```
fatalitiesPlot <- qplot(EVTYPE, data = fatalities, weight = FATALITIES, geom = "bar", binwidth = 1) +
    scale_y_continuous("Number of Fatalities") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Fatalities by Severe Weather\n Events in the U.S.\n from 1995 - 2011")
injuriesPlot <- qplot(EVTYPE, data = injuries, weight = INJURIES, geom = "bar", binwidth = 1) +
    scale_y_continuous("Number of Injuries") +
    theme(axis.text.x = element_text(angle = 45,
    hjust = 1)) + xlab("Severe Weather Type") +
    ggtitle("Total Injuries by Severe Weather\n Events in the U.S.\n from 1995 - 2011")
grid.arrange(fatalitiesPlot, injuriesPlot, ncol = 2)
```

According to the graphs shows the total fatalities and injuries resulting from serveral weather



events.

#####According to the histograms, flood and hurricane/typhoon result in the most property damage, while drought and flood result in the most crop damage in the United States from 1995 to 2011.

Conclusion

From the analysis, we conclude that excessive heat and tornado are most harmful with respect to population health, while flood, drought, and hurricane/typhoon have the greatest economic consequences.