Human and Economic Impact of Major Weather Events

Alex R Abraham

Sunday, July 20, 2014

Environment Details

```
sessionInfo()
```

```
## R version 3.1.0 (2014-04-10)

## Platform: x86_64-w64-mingw32/x64 (64-bit)

##

## locale:

## [1] LC_COLLATE=English_Australia.1252  LC_CTYPE=English_Australia.1252

## [3] LC_MONETARY=English_Australia.1252 LC_NUMERIC=C

## [5] LC_TIME=English_Australia.1252

##

## attached base packages:

## [1] stats     graphics  grDevices utils     datasets  methods   base

##

## loaded via a namespace (and not attached):

## [1] digest_0.6.4     evaluate_0.5.5   formatR_0.10     htmltools_0.2.4

## [5] knitr_1.6        rmarkdown_0.2.49 stringr_0.6.2    tools_3.1.0

## [9] yaml_2.1.13
```

Assignment Summary

Document Layout

Language: Your document should be written in English.

Title: Your document should have a title that briefly summarizes your data analysis

Synopsis: Immediately after the title, there should be a synopsis which describes and summarizes your analysis in at most 10 complete sentences.

There should be a section titled Data Processing which describes (in words and code) how the data were loaded into R and processed for analysis. In particular, your analysis must start from the raw CSV file containing the data. You cannot do any preprocessing outside the document. If

preprocessing is time-consuming you may consider using the cache = TRUE option for certain code chunks.

There should be a section titled Results in which your results are presented.

You may have other sections in your analysis, but Data Processing and Results are required.

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).

Loading the data.

Reading the Storm Data file
(Assumes that the file is present in the local data folder)

```
stormData <- read.table("./data/repdata-data-StormData.csv.bz2", header=T, sep = ",")
```

Questions

Your data analysis must address the following questions:

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

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

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.

Affect to health - Fatality and Injuries

Code chunk to identify the top 10 Fatality and the top 10 Injury data
This is done by totaling the values in FATALITIES and INJURIES columns by each EVTYPE category.
This is then sorted in descending order or value of total columns
The top ten by the value are picked to determine the top 10 categories

```
fa<-aggregate( FATALITIES~EVTYPE, stormData, sum, na.action=na.pass )

 topFa <- head(fa[with(fa, order(-FATALITIES)),],10)



inj<-aggregate( INJURIES~EVTYPE, stormData, sum, na.action=na.pass )

 topInj <- head(inj[with(inj, order(-INJURIES)),],10)
```

Loading ggplot2 library

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.1
```

Setting up and printing the plot for Fatalities

```
p1 <- ggplot(topFa , aes(x = EVTYPE, y = FATALITIES)) +
    geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle = 60, hjust = 1))+
    labs(title = "Top 10 climatic causes of mortality",
        x = "Event", y = "Total Deaths")


print(p1)
```

Setting up and printing the plot for Injuries

```
p2 <- ggplot(topInj , aes(x = EVTYPE, y = INJURIES)) +
    geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle = 60, hjust = 1))+
    labs(title = "Top 10 climatic causes of injury",
    x = "Event", y = "Total Injuries")


print(p2)
```

Converting to character

```
stormData$PROPDMGEXP <- as.character(stormData$PROPDMGEXP)
stormData$CROPDMGEXP <- as.character(stormData$CROPDMGEXP)
```

Assigning a numeric value to be used as Exponential factor for Property Damage

```
stormData[stormData$PROPDMGEXP %in% c("-" , "+" , "" , "?","0" , "1", NA), c("PROPDMGEXP")] <-1
stormData[stormData$PROPDMGEXP %in% c("2","h", "H"), c("PROPDMGEXP")] <-2
stormData[stormData$PROPDMGEXP %in% c("3","k", "K"), c("PROPDMGEXP")] <-3
stormData[stormData$PROPDMGEXP %in% c("4"), c("PROPDMGEXP")] <-4
stormData[stormData$PROPDMGEXP %in% c("5"), c("PROPDMGEXP")] <-5
stormData[stormData$PROPDMGEXP %in% c("6","m", "M"), c("PROPDMGEXP")] <-6
```

```
stormData[stormData$PROPDMGEXP %in% c("7"), c("PROPDMGEXP")] <-7

stormData[stormData$PROPDMGEXP %in% c("8"), c("PROPDMGEXP")] <-8

stormData[stormData$PROPDMGEXP %in% c("9","b", "B"), c("PROPDMGEXP")] <-9
```

Assigning a numeric value to be used as Exponential factor for Crop Damage

```
stormData[stormData$CROPDMGEXP %in% c("-" , "+" , "" , "?","0" , "1",NA), c("CROPDM
GEXP")] <-1

stormData[stormData$CROPDMGEXP %in% c("2","h", "H"), c("CROPDMGEXP")] <-2

stormData[stormData$CROPDMGEXP %in% c("3","k", "K"), c("CROPDMGEXP")] <-3

stormData[stormData$CROPDMGEXP %in% c("4"), c("CROPDMGEXP")] <-4

stormData[stormData$CROPDMGEXP %in% c("5"), c("CROPDMGEXP")] <-5

stormData[stormData$CROPDMGEXP %in% c("6","m", "M"), c("CROPDMGEXP")] <-6

stormData[stormData$CROPDMGEXP %in% c("7"), c("CROPDMGEXP")] <-7

stormData[stormData$CROPDMGEXP %in% c("8"), c("CROPDMGEXP")] <-8

stormData[stormData$CROPDMGEXP %in% c("9","b", "B"), c("CROPDMGEXP")] <-9
```

Converting Property and Crop exponentials to numeric to enable computations

```
stormData$PROPDMGEXP <- as.numeric(stormData$PROPDMGEXP)

stormData$CROPDMGEXP <- as.numeric(stormData$CROPDMGEXP)
```

Calculating the total of property and crop damage for each event

```
stormData$TOTDMG <- stormData$PROPDMG *10^stormData$PROPDMGEXP +  stormData$CROPDMG
*10^stormData$CROPDMGEXP
```

Aggregating by EVTYPE and picking top 10 only

```
td<-aggregate( TOTDMG~EVTYPE, stormData, sum, na.action=na.pass )

topTd <- head(td[with(td, order(-TOTDMG)),],10)
```

Setting up the ggplot and printing top categories of damage in $ Billions

```
p3 <- ggplot(topTd , aes(x = EVTYPE, y = TOTDMG/10^9)) +

    geom_bar(stat = "identity") +

    theme(axis.text.x = element_text(angle = 60, hjust = 1))+

    labs(title = "Top 10 climate events by Total $ Billion in damage",

         x = "Event", y = "Total Damage")


print(p3)
```