How weather related events impacts the US

Mario Azevedo

Sunday, May 24, 2015

### Synopsis

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

### Data Processing

#### Loading the libraries that will be used

require(data.table)  
require(dplyr)  
require(tidyr)  
require(readr)

#### Downloading and reading the data

#download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",  
# "repdata-data-StormData.csv.bz2")  
data <- read.csv('repdata-data-StormData.csv.bz2')  
#Reading a data file prepared from the "Storm Data Event Table"  
type <- readLines('storm.txt')  
data <- data.table(data)

#### Cleaning the data

data[data$EVTYPE == "TSTM WIND"]$EVTYPE <- "THUNDERSTORM WIND"  
data[data$EVTYPE == "THUNDERSTORM WINDS"]$EVTYPE <- "THUNDERSTORM WIND"  
data[data$EVTYPE == "MARINE TSTM WIND"]$EVTYPE <- "MARINE THUNDERSTORM WIND"  
data[data$EVTYPE == "MARINE THUNDERSTORM WINDS"]$EVTYPE <- "MARINE THUNDERSTORM WIND"  
data <- filter(data,EVTYPE %in% type)

#### Preparing data for first question

harm <- group\_by(data,EVTYPE) %>%   
 summarise(TofEV=n(),Fatalities = sum(FATALITIES),  
 Injuries=sum(INJURIES)) %>%  
 arrange(desc(Fatalities))  
harm

## Source: local data table [46 x 4]  
##   
## EVTYPE TofEV Fatalities Injuries  
## 1 TORNADO 60652 5633 91346  
## 2 EXCESSIVE HEAT 1678 1903 6525  
## 3 FLASH FLOOD 54277 978 1777  
## 4 HEAT 767 937 2100  
## 5 LIGHTNING 15754 816 5230  
## 6 THUNDERSTORM WIND 323346 701 9353  
## 7 FLOOD 25326 470 6789  
## 8 RIP CURRENT 470 368 232  
## 9 HIGH WIND 20212 248 1137  
## 10 AVALANCHE 386 224 170  
## .. ... ... ... ...

* 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.

### Results

* 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).