Reproducible-research-project-2-Markdown

Ing. Greg Bennett

August 30, 2017

Load Data

This is an **R Markdown** document contains the Peer-graded Assignment: Course Project 2. Step 1 You will need to download & save in a specific folder, see: https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2.

```
library(readr)

## Warning: package 'readr' was built under R version 3.3.3

df <- read_csv("C:/Users/gbennett/Dropbox/Data
Scientists/5.Reproducible_Research/project/repdata_data_StormData.csv")

## Parsed with column specification:
## cols(
##   .default = col_character(),
##   STATE__ = col_double(),
##   COUNTY = col_double(),
##   BGN_RANGE = col_double(),
##   COUNTY_END = col_double(),
##   END_RANGE = col_double(),
##   LENGTH = col_double(),
##   WIDTH = col_double(),
##   F = col_integer(),
##   MAG = col_double(),
##   FATALITIES = col_double(),
##   INJURIES = col_double(),
##   PROPDMG = col_double(),
##   CROPDMG = col_double(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double(),
##   LATITUDE_E = col_double(),
##   LONGITUDE_ = col_double(),
##   REFNUM = col_double()
## )

## See spec(...) for full column specifications.
```

Process the data

To evaluate the health impact, the total fatalities and the total injuries for each event type (EVTYPE) are calculated. The codes for this calculation are:

```
## Warning: package 'dplyr' was built under R version 3.3.3

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Warning: package 'bindrcpp' was built under R version 3.3.3

## # A tibble: 10 x 2
##            EVTYPE total.fatalities
##             <chr>            <dbl>
##  1        TORNADO             5633
##  2 EXCESSIVE HEAT             1903
##  3    FLASH FLOOD              978
##  4           HEAT              937
##  5      LIGHTNING              816
##  6      TSTM WIND              504
##  7          FLOOD              470
##  8    RIP CURRENT              368
##  9      HIGH WIND              248
## 10      AVALANCHE              224

## # A tibble: 10 x 2
##               EVTYPE total.injuries
##                <chr>          <dbl>
##  1           TORNADO          91346
##  2         TSTM WIND           6957
##  3             FLOOD           6789
##  4    EXCESSIVE HEAT           6525
##  5         LIGHTNING           5230
##  6              HEAT           2100
##  7         ICE STORM           1975
##  8       FLASH FLOOD           1777
##  9 THUNDERSTORM WIND           1488
## 10              HAIL           1361
```

Economic impact:

```
## Warning in cbind(Multiplier, Symbol): number of rows of result is not a
## multiple of vector length (arg 2)
```

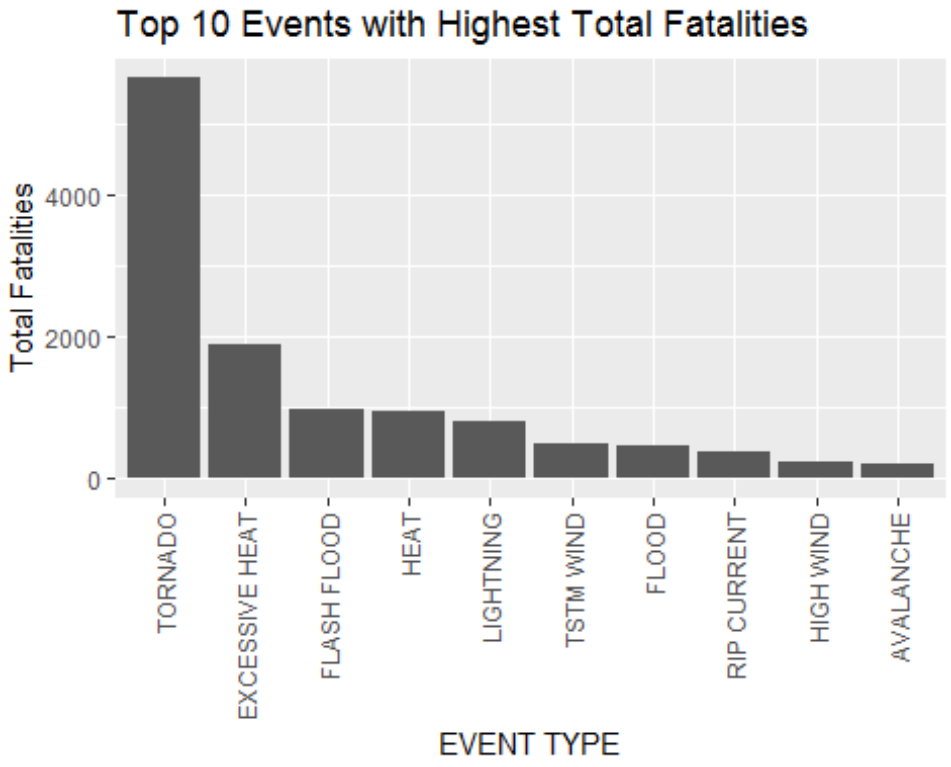
Results

Health Impact The **top 10 events** with the highest total fatalities and injuries are shown graphically.

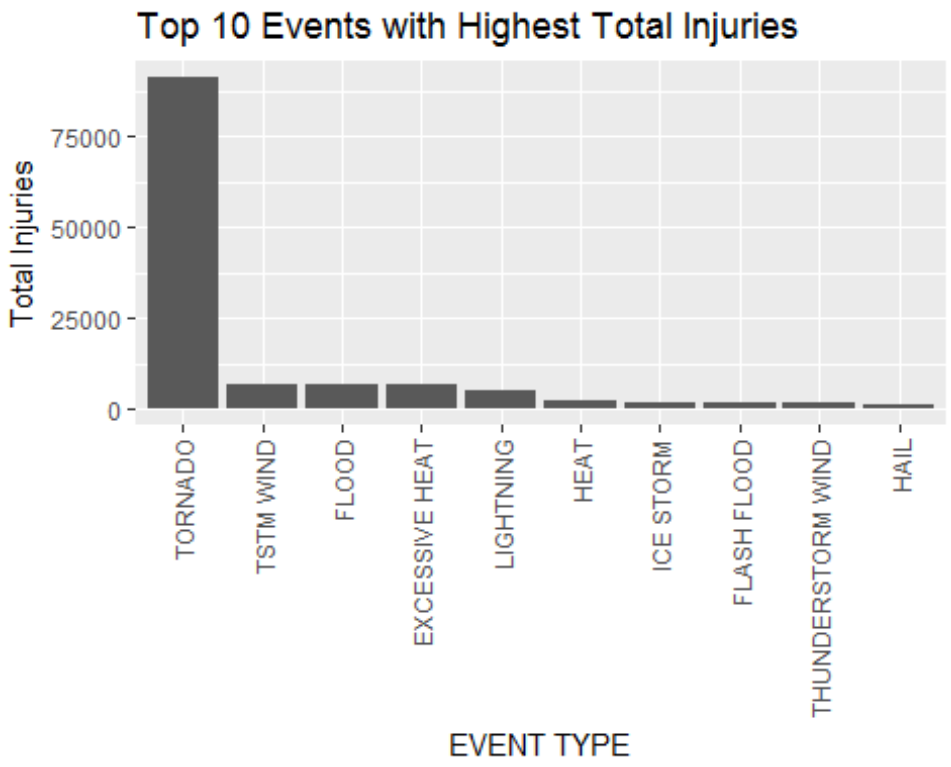
```
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.3

HTF <- ggplot(df.fatalities[1:10,], aes(x=reorder(EVTYPE, -total.fatalities),
y=total.fatalities))+geom_bar(stat="identity") + theme(axis.text.x =
element_text(angle=90, vjust=0.5, hjust=1))+ggtitle("Top 10 Events with
Highest Total Fatalities") +labs(x="EVENT TYPE", y="Total Fatalities")
HTI <- ggplot(df.injuries[1:10,], aes(x=reorder(EVTYPE, -total.injuries),
y=total.injuries))+geom_bar(stat="identity") + theme(axis.text.x =
element_text(angle=90, vjust=0.5, hjust=1))+ggtitle("Top 10 Events with
Highest Total Injuries") +labs(x="EVENT TYPE", y="Total Injuries")
HTF
```




```
HTI
```


Top 10 Events with Highest Total Injuries
Total Injuries
75000
50000
25000
0
TORNADO
TSTM WIND
FLOOD
EXCESSIVE HEAT
LIGHTNING
HEAT
ICE STORM
FLASH FLOOD
THUNDERSTORM WIND
HAIL
EVENT TYPE