United States weather events analysis report

Zhao Mingli

October 29, 2016

Introduction of this data analysis.

This report explores the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database and answers two questions below:

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

Data processing

The data processing are stressed in following steps:

1. Get the raw csv and load the data into dataframe named *stormData*

```
library(dplyr)
```

```
##
## 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
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```
#download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",destfile = "s
# Read data as a dataframe
stormData <- read.csv(file = "storm_data.csv.bz2",stringsAsFactors = FALSE)
```

2. Compute the injuries and fatalities

```
# Compute the injuries
stormDataByEVType <- stormData %>% group_by(EVTYPE)
topInjuries <- stormDataByEVType %>% summarise(mean_injuries=mean(INJURIES)) %>% arrange(desc(mean_inju
# Compute the fatalities
topFatalities <- stormDataByEVType %>% summarise(mean_fatalities=mean(FATALITIES)) %>% arrange(desc(mea
```
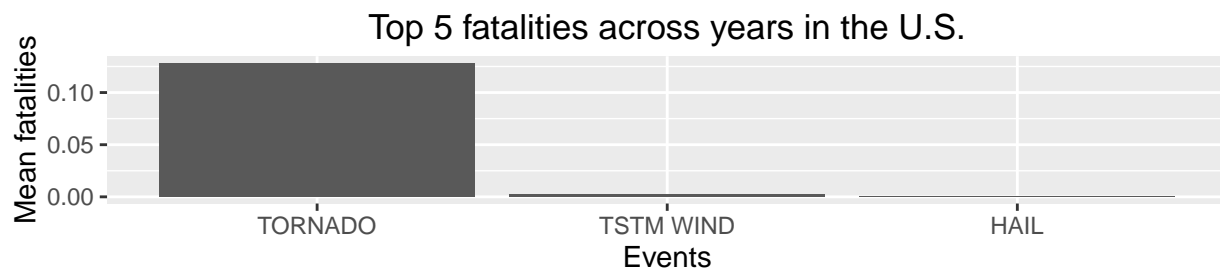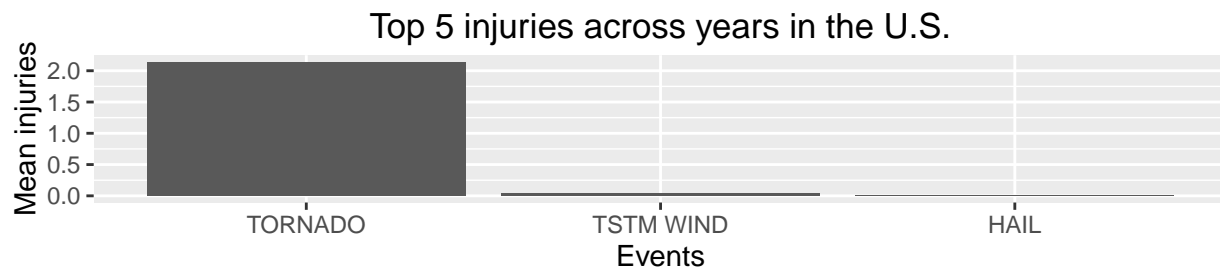
3. Compute the economic consequences

```
# summarize property damages
topPropDmg <- stormDataByEVType %>% summarise(sum_propDmg=sum(PROPDMG)) %>% arrange(desc(sum_propDmg))
# summarize crop damages
topCropDmg <- stormDataByEVType %>% summarise(sum_cropDmg=sum(CROPDMG)) %>% arrange(desc(sum_cropDmg))
```

Results

Tornado is the most harmful event to human's health. More events that causes injuries or fatalness are
listed below:

```
p1 <- ggplot(head(topInjuries,n=5),aes(x=reorder(EVTYPE,-mean_injuries),y=mean_injuries)) + geom_bar(sta
p2 <- ggplot(head(topFatalities,n=5),aes(x=reorder(EVTYPE,-mean_fatalities),y=mean_fatalities)) + geom_
multiplot(p1, p2,rows=2)
```
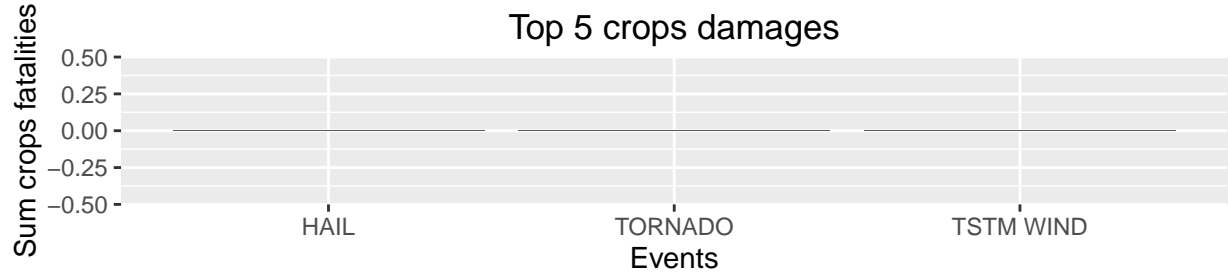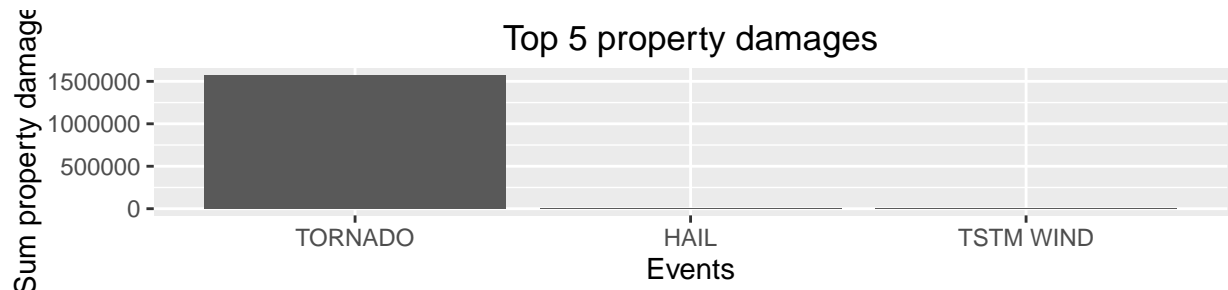

[1] 2

Regarding economic consequences, **Tornado** causes property lost most among all events whereas **Heat**
damages crops most.
The top 3 events causes economic losses are stressed as following:

```
p3 <- ggplot(head(topPropDmg,n=5),aes(x=reorder(EVTYPE,-sum_propDmg),y=sum_propDmg)) + geom_bar(stat =
p4 <- ggplot(head(topCropDmg,n=5),aes(x=reorder(EVTYPE,-sum_cropDmg),y=sum_cropDmg)) + geom_bar(stat =
multiplot(p3, p4,rows=2)
```

Top 5 property damages

Top 5 crops damages

```
## [1] 2
```