

Impacts of Weather Events on Health and Economy in the US

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities. This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database which tracks characteristics of major storms and weather events in the United States.

In this report, the impacts of weather events on both population health and economic are analyzed. Regarding population health, the results reveal that tornadoes and excessive heat are the two most harmful weather events. In addition, the largest costs (property damages) to US economy between 1950 and 2011 are caused by flash floods and thunderstorm winds.

Data Processing

- Data source: The original data are from Storm Events Database (<http://www.ncdc.noaa.gov/stormevents/ftp.jsp>), provided by National Climatic Data Center (<http://www.ncdc.noaa.gov/>).
- Data format: The data are stored in a comma-separated-value file available here (<https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>).
- Data documentation: There is documentation of the data available here (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf).

First of all, the original data are loaded into a data frame.

```
data <- read.csv(bzfile("StormData.csv.bz2"))
```

Then, the data need to be cleaned. Specifically, the event types are converted to lowercase and all punctuation characters are replaced with a space.

```
# convert event types to lowercase
eventTypes <- tolower(data$EVTTYPE)
# replace punctuation characters with a space
eventTypes <- gsub("[[:blank:][:punct:]]+", " ", eventTypes)
# update the original data
data$EVTTYPE <- eventTypes
```

The above cleaned data are used in the following analysis.

Analysis of Harmful Events to Population Health

The number of fatalities are aggregated by the weather event type.

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.1.2
```

```
casualties <- ddply(data, .(EVTYPE), summarize,  
                    fatalities = sum(FATALITIES))  
# Find top 5 weather events causing most death  
fatalEvents <- head(casualties[order(casualties$fatalities, decreasing = T), ], 5)
```

Top 5 weather events causing the largest number of deaths include:

```
fatalEvents[, c("EVTYPE", "fatalities")]
```

```
##           EVTYPE fatalities  
## 741         tornado      5633  
## 116 excessive heat      1903  
## 138   flash flood       978  
## 240         heat       937  
## 410   lightning       816
```

Analysis of Economic Impacts of Weather Events

To find events that have the greatest economic consequences, property damages are aggregated by the weather event type.

```
computeExponent <- function(e) {  
  if (e %in% c('h', 'H')) # h -> hundred  
    return(2)  
  else if (e %in% c('k', 'K')) # k -> thousand  
    return(3)  
  else if (e %in% c('m', 'M')) # m -> million  
    return(6)  
  else if (e %in% c('b', 'B')) # b -> billion  
    return(9)  
  else if (!is.na(as.numeric(e)))  
    return(as.numeric(e))  
  else if (e %in% c('', '-', '?', '+'))  
    return(0)  
  else {  
    stop("Invalid exponent value")  
  }  
}
```

```
exponent <- sapply(data$PROPDGMGEXP, FUN=computeExponent)  
data$prop_dmg <- data$PROPDMG * (10 ** exponent)
```

```
library(plyr)
```

```
# Calculate the economy loss by event type
```

```
economyLoss <- ddply(data, .(EVTYPE), summarize, prop_dmg = sum(prop_dmg))
```

```
economyLoss <- economyLoss[(economyLoss$prop_dmg > 0), ]
```

```
# Find top 5 wether events causing largest economy loss
```

```
propertyDamageEvents <- head(economyLoss[order(economyLoss$prop_dmg, decreasing = T), ],  
5)
```

Top 5 costly weather events causing most property damages include:

```
propertyDamageEvents[, c("EVTYPE", "prop_dmg")]
```

```
##              EVTYPE  prop_dmg  
## 138      flash flood 6.820e+13  
## 697 thunderstorm winds 2.087e+13  
## 741          tornado 1.079e+12  
## 209             hail 3.158e+11  
## 410      lightning 1.729e+11
```

Results

Health impacts of weather events

The top deadly weather event types are shown in the following plot.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

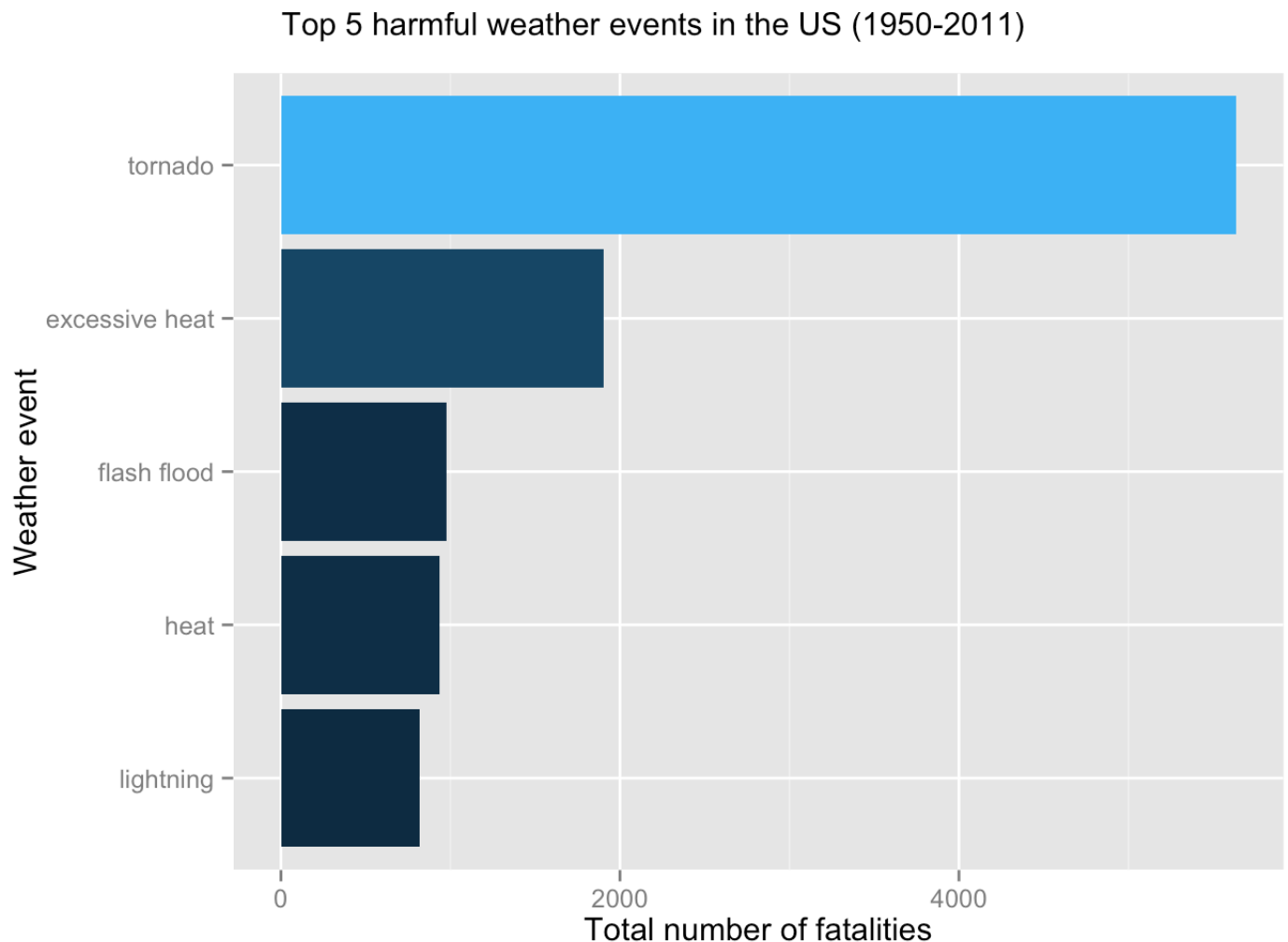
```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.1.2
```

```
## Loading required package: grid
```

```
p1 <- ggplot(data=fatalEvents,
             aes(x=reorder(EVTYPE, fatalities), y=fatalities, fill=fatalities)) +
  geom_bar(stat="identity") +
  coord_flip() +
  ylab("Total number of fatalities") +
  xlab("Weather event") +
  theme(legend.position="none")

grid.arrange(p1, main="Top 5 harmful weather events in the US (1950-2011)")
```



As can be seen from the above figure, tornadoes cause the most number of fatalities among all event types. Specifically, in the last 60 years, more than 5000 fatalities are caused by tornadoes in the US. The other event types such as excessive heat and flash floods are also most harmful to the population health.

Economic impacts of weather events

The top 5 most costly weather event types since 1950s are reported in the following plot.

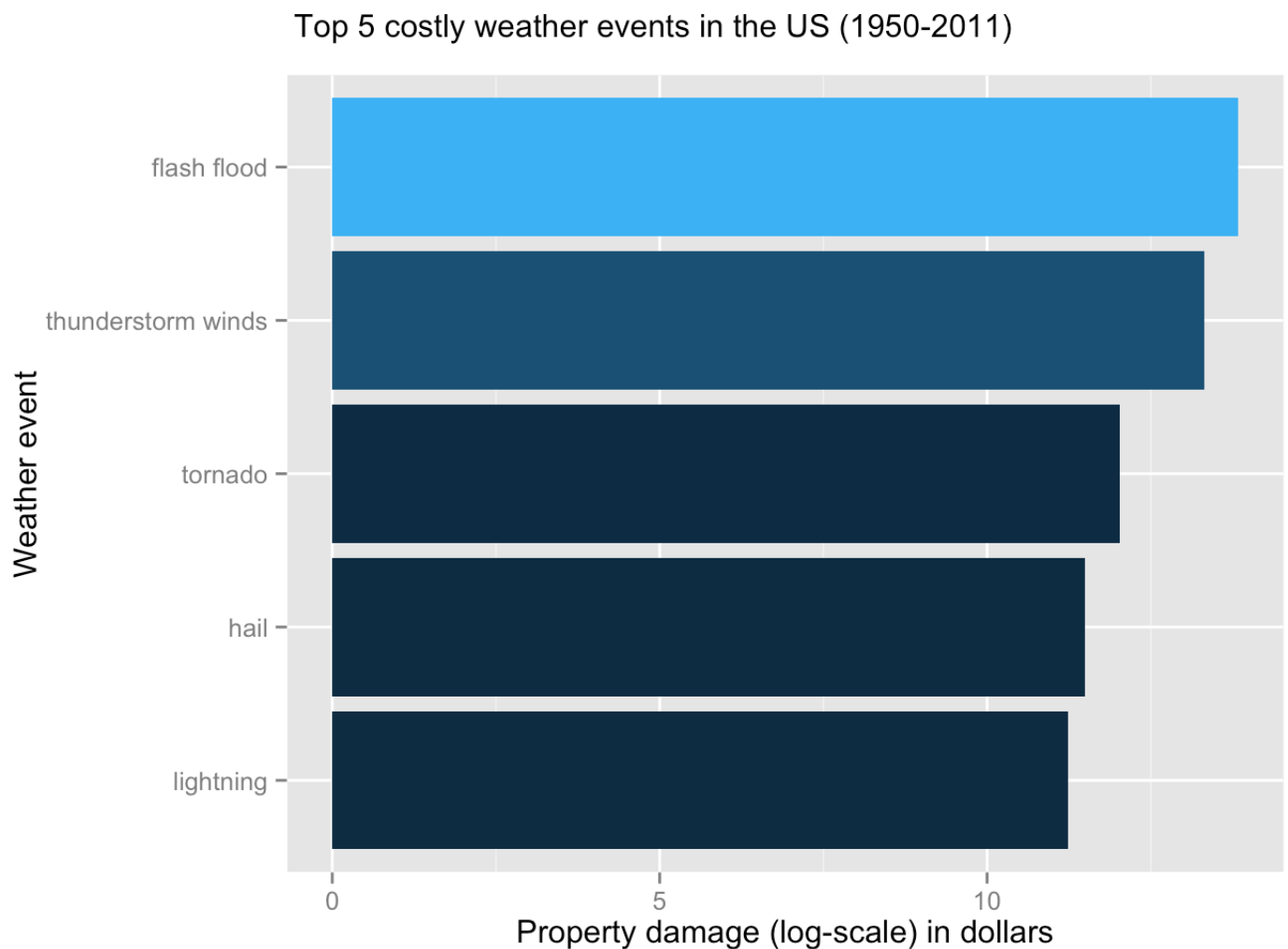
```

library(ggplot2)
library(gridExtra)

p1 <- ggplot(data=propertyDamageEvents,
             aes(x=reorder(EVTYPE, prop_dmg), y=log10(prop_dmg), fill=prop_dmg )) +
  geom_bar(stat="identity") +
  coord_flip() +
  xlab("Weather event") +
  ylab("Property damage (log-scale) in dollars") +
  theme(legend.position="none")

grid.arrange(p1, main="Top 5 costly weather events in the US (1950-2011)")

```



In the above figure, the property damages are shown in logarithmic scale given the large range of their values. The results reveal that the largest costs (property damages) to US economy between 1950 and 2011 are caused by flash floods and thunderstorm winds.