

Title: Storm Data Analysis using Data from NOAA

Synopsis : This R markdown file analyzes ‘Storm Data’ from “U.S. National Oceanic and Atmospheric Administration’s (NOAA) storm database” to see,

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

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

Data Processing Part

Step 1 Downloading: At this analysis, it will download the file compressed with bzip2 algorithm named, “StormData.csv.bz2”. This file will be decompressed to get the StormData.csv file.

Step 2 Loading: This extracted file will be used as the raw data for analysis. This file will be loaded into ‘data’ variable.

Step 3 Data Processing: The data will be examined to be used for the analysis to answer above two questions.

Analysis Part

Step 4 Analysis:

1) ‘EVTYPE’ variable and time variable will be used to see which event is the most harmful respect to population health.

2) Again, ‘EVTYPE’ variable and variables related to economic consequences will be used to see which event is the influential to economic of USA.

Step 5 Report: Results from the analysis will be posted in here and to Rpubs.

For question 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

Data Processing

Step 1 Downloading

From the URL “<http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>”

(<http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2>), the raw data will be downloaded as “StormData.csv.bz2”

```
setwd("D:/Jung Soo Lee@JSLEE/Documents/MOOC/Coursera/[Coursera]Data Science Specialization/Reproducible Research/Course Project 2")

if(!file.exists("StormData.csv.bz2")) {
  Original_Data_URL <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file(Original_Data_URL, destfile="StormData.csv.bz2")
}
```

Step 2 Loading

The downloaded zipfile will be loaded.

File failed to load: /extensions/MathZoom.js

```
## No need to decompress this downloaded file.
## StormData = bzfile("StormData.csv.bz2")
data <- read.csv("StormData.csv.bz2", stringsAsFactors=F)
```

Step 3 Data Processing

The loaded data will be explored at first and processed for further analysis.

```
# Check the data
str(data)
# ==> 902297 obs. of  37 variables
head(data)
tail(data)
summary(data)
summary(data$EVTYPE)
# EVTYPE
# Length:902297
# Class :character
# Mode  :character
table(data$EVTYPE)
eventtype_unique <- unique(data$EVTYPE)
summary(eventtype_unique)
# Length      Class      Mode
# 985         character character
data$EVTYPE <- tolower(data$EVTYPE)
eventtype_unique <- unique(data$EVTYPE)
summary(eventtype_unique)
# Length      Class      Mode
# 898         character character
head(data$INJURIES)
summary(data$INJURIES)
head(data$FATALITIES)
summary(data$FATALITIES)
```

Analysis

Step 4 Analysis:

```
# install.packages("ggplot2", dependencies = T)
library(ggplot2)
library(lattice)

# For "1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?" question,
# I will calculate how many injuries+fatalities an event caused using aggregate function.
casualties <- with(data, aggregate(INJURIES + FATALITIES ~ EVTYPE, data=data, FUN = "sum"))

# Change the name
names(casualties)[2] <- "Totalcasualties"

# Order the number of casualties by decending method
ordered_casualties <- casualties[order(-casualties$Totalcasualties),]

# Just see top6 using head()
Top6 <- head(ordered_casualties)
```

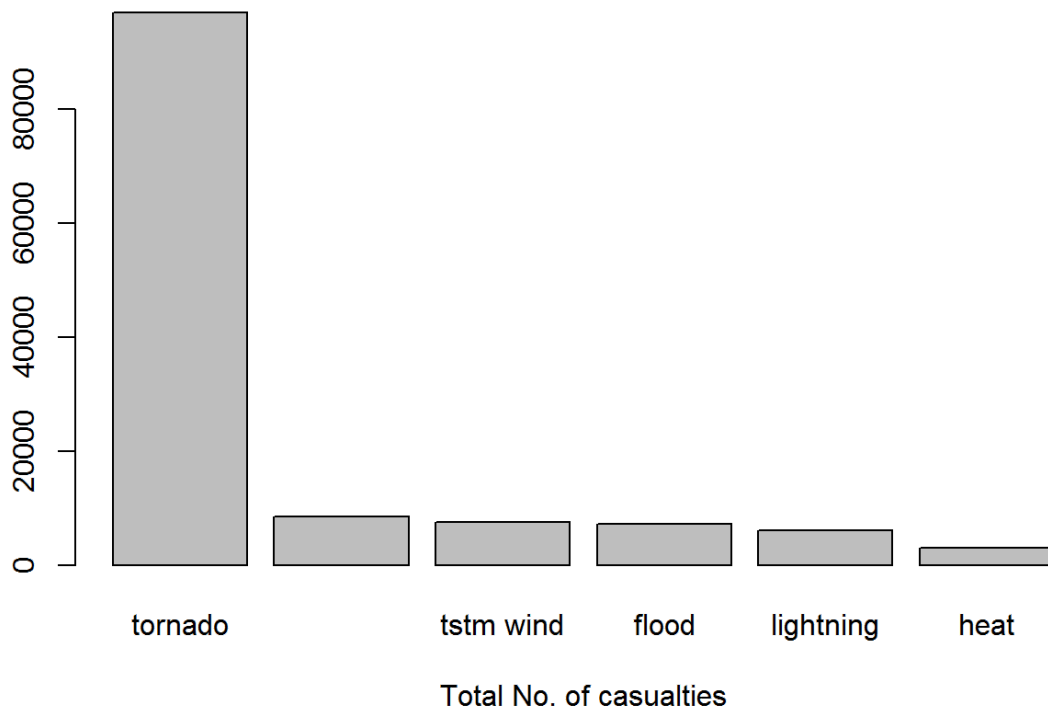
Step 5 Report:

File failed to load: /extensions/MathZoom.js

```
# Draw a barplot
```

```
barplot(Top6$Totalcasualties, main = "Which event caused the most harmful with respect to population health", xlab = "Total No. of casualties", names.arg=Top6$EVTYPE)
```

Which event caused the most harmful with respect to population health?



```
# The answer
```

```
"The most harmful event is"
```

```
## [1] "The most harmful event is"
```

```
Top6$EVTYPE[1]
```

```
## [1] "tornado"
```

For the “2. Across the United States, which types of events have the greatest economic consequences?” question

Data Processing

Step 1 Downloading

Data to download is the same as the one downloaded in question 1.

Step 2 Loading(same as above in question1)

Loading is the same as that of question 1. “Data” variable will be used.

Step 3 Data Processing**

```

# The loaded data will be explored for further analysis of question2
str(data)
str(data$CROPDMGEXP)
unique(data$CROPDMGEXP)
unique(data$PROPDGMGEXP)
str(data$PROPDGMGEXP)
table(data$CROPDMGEXP)
table(data$PROPDGMGEXP)

# Variable 'CROPDMGEXP' and 'PROPDGMGEXP' will be subsetted values with "M", "K", "B" since these are values mention
ed in pdf "NATIONAL WEATHER SERVICE INSTRUCTION" p12.
# First, 'CROPDMGEXP' subsetting
data2 <- subset(data, (data$CROPDMGEXP == "M" | data$CROPDMGEXP == "K" | data$CROPDMGEXP == "B") | (data$PROPDGMGEXP
=="M" | data$PROPDGMGEXP == "K" | data$PROPDGMGEXP == "B") )

str(data2)
# CROPDMG value will be multiplied accordingly with,
# "M", "K", "B" which mean 1000000, 1000, 1000000000 respectively.
summary(data2$CROPDMG)
table(data2$CROPDMGEXP)
# Character with 'B', 'K', 'M' are assigned values accordingly as follows. Other than these three character, the m
ultiplier will be 1.

for(i in 1:length(data2$CROPDMGEXP)) {
  ifelse(data2$CROPDMGEXP[i] == "M", data2$CROPDMG[i] <- data2$CROPDMG[i] * 1000000,
    ifelse(data2$CROPDMGEXP[i] == "K", data2$CROPDMG[i] <- data2$CROPDMG[i] * 1000,
      ifelse(data2$CROPDMGEXP[i] == "B", data2$CROPDMG[i] <- data2$CROPDMG[i] * 1000000000, data2$CROPDMG
[i] <- data2$CROPDMG[i] * 1)))
}
summary(data2$CROPDMG)

# PROPDGMG value will be multiplied accordingly with,
# "M", "K", "B" which mean 1000000, 1000, 1000000000 respectively.
summary(data2$PROPDGMG)
table(data2$PROPDGMGEXP)
# Character with 'B', 'K', 'M' are assigned values accordingly as follows. Other than these three character, the m
ultiplier will be 1.

for(i in 1:length(data2$PROPDGMGEXP)) {
  ifelse(data2$PROPDGMGEXP[i] == "M", data2$PROPDGMG[i] <- data2$PROPDGMG[i] * 1000000,
    ifelse(data2$PROPDGMGEXP[i] == "K", data2$PROPDGMG[i] <- data2$PROPDGMG[i] * 1000,
      ifelse(data2$PROPDGMGEXP[i] == "B", data2$PROPDGMG[i] <- data2$PROPDGMG[i] * 1000000000, data2$PROPDGMG
[i] <- data2$PROPDGMG[i] * 1)))
}
summary(data2$PROPDGMG)

```

Analysis

Step 4 Analysis:

```
# install.packages("ggplot2", dependencies = T)
library(ggplot2)
library(lattice)

# I will calculate how much economic values are lost using aggregate function.
Eco_dmg <- with(data2, aggregate(CROPDMG + PROPDMG ~ EVTYPE, data=data, FUN = "sum"))

# Change the name
names(Eco_dmg)[2] <- "Total_Economic_Damage"

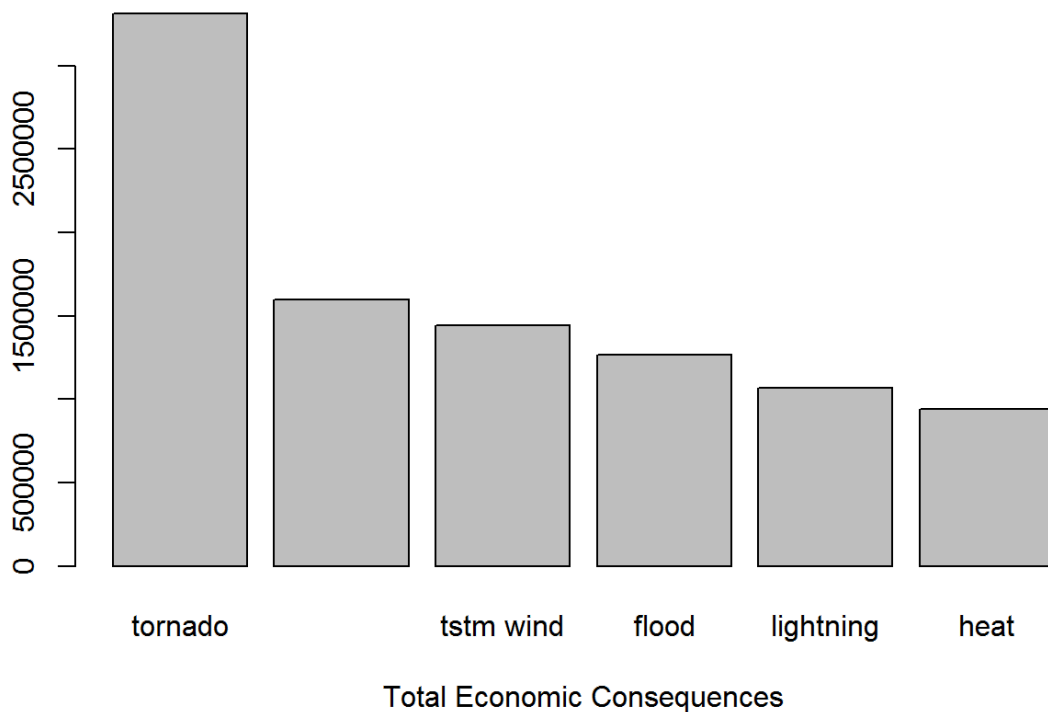
# Order the economic values by decending method
ordered_eco_dmg <- Eco_dmg[order(-Eco_dmg$Total_Economic_Damage),]

# Just see top6 using head()
Top6_Eco <- head(ordered_eco_dmg)
```

Step 5 Report:

```
# Draw a barplot
barplot(Top6_Eco$Total_Economic_Damage, main = "Which event caused the greatest economic consequence", xlab = "Total Economic Consequences", names.arg=Top6_Eco$EVTYPE)
```

Which event caused the greatest economic consequence



```
# The answer
"The greatest economic consequence event is"
```

```
## [1] "The greatest economic consequence event is"
```

```
Top6_Eco$EVTYPE[1]
```

```
## [1] "tornado"
```

File failed to load: /extensions/MathZoom.js