RepData_PeerAssessment2

Chuwei Zhong

10/14/2019

This report presented the data analysis for U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. We extract the most harmful weather events to people and economics.

Download the data from website: https://www.coursera.org/learn/reproducible-research/peer/OMZ37/course-project-2 This is a csv.bz2 file and unzipped to working directory.

load library

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(gridExtra)
```

analysis start from the raw data file

```
data <- read.csv("repdata_data_StormData.csv")

#get useful columns
data <- data[ , c(8, 23:28)]
```

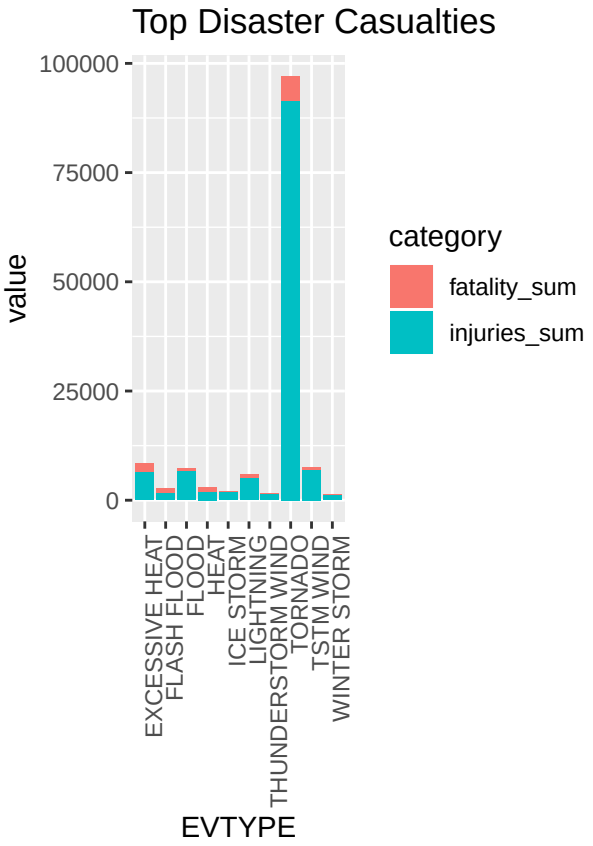
which types of events are most harmful to population health?

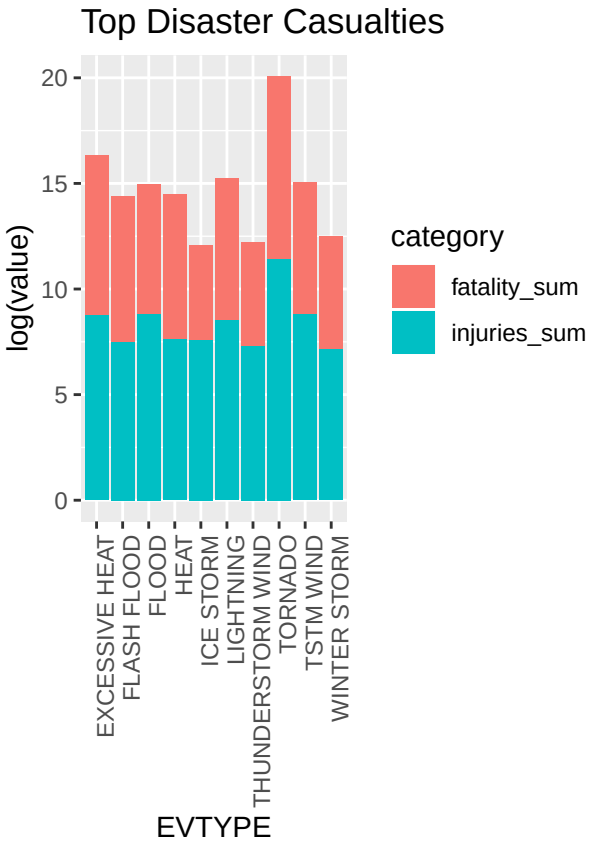
In this question, we will get the result from "FATALITIES" and "INJURIES". A bar graph is used to present and only top 10 factors will show. To make y-axis easier to show, I use log function.

```
#aggregate the casualties data, reorder and extract top 10 weather events
health_data <- data %>%
  group_by(EVTYPE) %>%
  summarize(fatality_sum = sum(FATALITIES,na.rm = TRUE),injuries_sum = sum(INJURIES,na.rm = TRUE))

health_data <- arrange(health_data, desc(fatality_sum+injuries_sum))
health_data <- health_data[1:10,]

#show the result with ggplot2
health_data <- gather(health_data, category, value, fatality_sum, injuries_sum)
plot1 <- ggplot(data=health_data, aes(x=EVTYPE, y=value, fill=category)) +
  geom_bar(stat="identity")+
  ggtitle("Top Disaster Casualties") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
plot2 <- ggplot(data=health_data, aes(x=EVTYPE, y=log(value), fill=category)) +
  geom_bar(stat="identity")+
  ggtitle("Top Disaster Casualties") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
grid.arrange(plot1, plot2, ncol=2)
```





So, Tornado is the weather event which causes the most casualty.

which types of events have the greatest economic consequences?

In this question, we will get the result from "PROPDMG" "PROPDMGEXP" "CROPDMG" and "CROPDMGEXP".

```
table(data$PROPDMGEXP)
```

```
##
##             -      ?      +      0      1      2      3      4      5
## 465934      1      8      5    216     25     13      4      4     28
##      6      7      8      B      h      H      K      m      M
##      4      5      1     40      1      6 424665      7  11330
```

```
table(data$CROPDMGEXP)
```

```
##
##             ?      0      2      B      k      K      m      M
## 618413      7     19      1      9     21 281832      1   1994
```

First, the damages need to be calculated. We only concern the relative large damage amount.

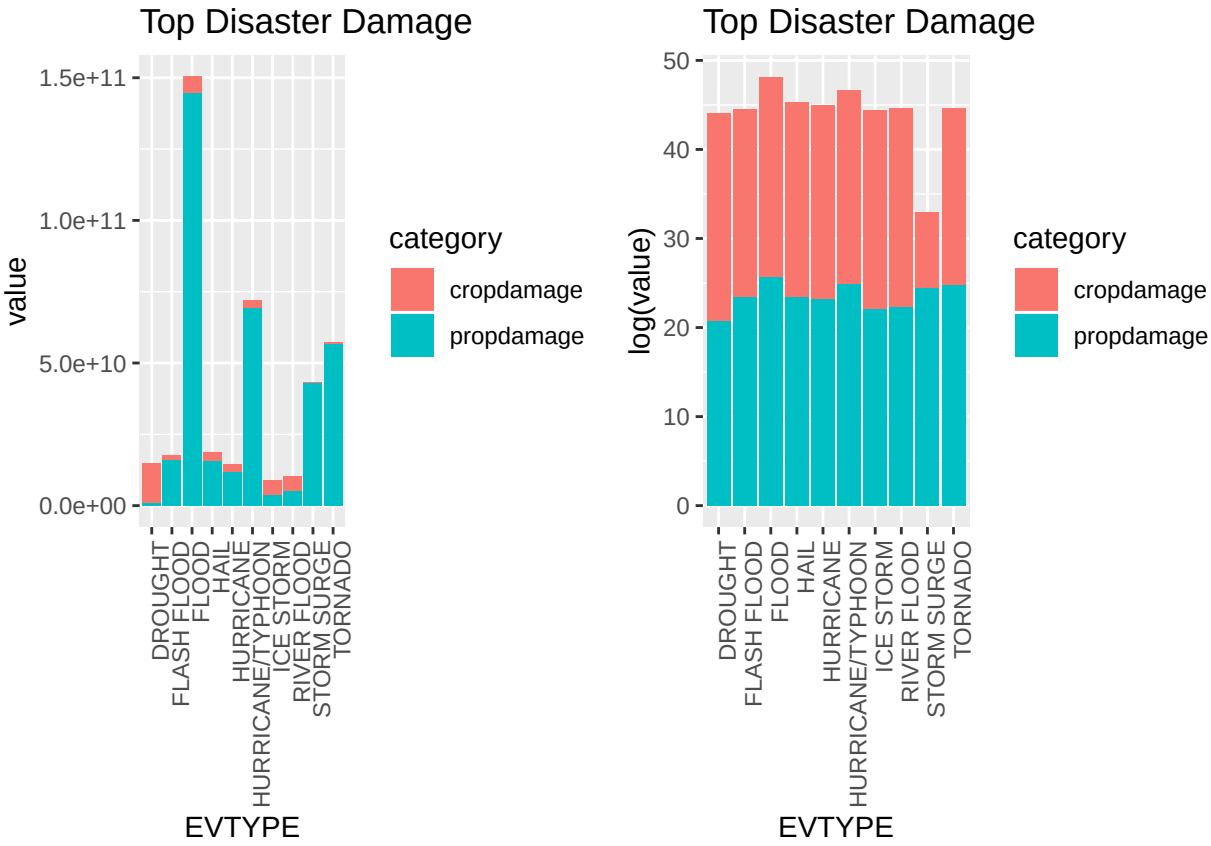
```
data$propdamage <- 0
data[data$PROPDMGEXP=="H",]$propdamage <- data[data$PROPDMGEXP=="H",]$PROPDMG * 10^2
data[data$PROPDMGEXP=="h",]$propdamage <- data[data$PROPDMGEXP=="h",]$PROPDMG * 10^2
data[data$PROPDMGEXP=="K",]$propdamage <- data[data$PROPDMGEXP=="K",]$PROPDMG * 10^3
data[data$PROPDMGEXP=="M",]$propdamage <- data[data$PROPDMGEXP=="M",]$PROPDMG * 10^6
data[data$PROPDMGEXP=="m",]$propdamage <- data[data$PROPDMGEXP=="m",]$PROPDMG * 10^6
data[data$PROPDMGEXP=="B",]$propdamage <- data[data$PROPDMGEXP=="B",]$PROPDMG * 10^9
```

```
data$cropdamage <- 0
data[data$CROPDMGEXP=="k",]$cropdamage <- data[data$CROPDMGEXP=="k",]$CROPDMG * 10^3
data[data$CROPDMGEXP=="K",]$cropdamage <- data[data$CROPDMGEXP=="K",]$CROPDMG * 10^3
data[data$CROPDMGEXP=="M",]$cropdamage <- data[data$CROPDMGEXP=="M",]$CROPDMG * 10^6
data[data$CROPDMGEXP=="m",]$cropdamage <- data[data$CROPDMGEXP=="m",]$CROPDMG * 10^6
data[data$CROPDMGEXP=="B",]$cropdamage <- data[data$CROPDMGEXP=="B",]$CROPDMG * 10^9

#damage_data <- aggregate(c(propdamage,cropdamage)~EVTYPE, data = data, sum, na.rm=TRUE)
#aggregate the casualties data, reorder and extract top 10 weather events
damage_data <- data %>%
  group_by(EVTYPE) %>%
  summarize(propdamage = sum(propdamage,na.rm = TRUE),cropdamage = sum(cropdamage,na.rm = TRUE))
damage_data <- arrange(damage_data, desc(propdamage+cropdamage))
damage_data <- damage_data[1:10,]

#show the result with ggplot2
damage_data <- gather(damage_data, category, value, propdamage, cropdamage)

plot1 <- ggplot(data=damage_data, aes(x=EVTYPE, y=value, fill=category)) +
  geom_bar(stat="identity")+
  ggtitle("Top Disaster Damage") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
plot2 <- ggplot(data=damage_data, aes(x=EVTYPE, y=log(value), fill=category)) +
  geom_bar(stat="identity")+
  ggtitle("Top Disaster Damage") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
grid.arrange(plot1, plot2, ncol=2)
```



So, Flood is the weather event which causes the most economic damage.