---
title: “Peer-graded Assignment: Course Project 2”
output: html_document
---

## Peer-graded Assignment: Course Project 2

By - Manan Vohra

December 7, 2020

## 1: Synopsis

The goal of this assignment is to exploration and analysis of the NOAA Storm Database to highlight the key insights regarding the severe weather events aon both the population and economy. The time line for this data is from 1950 till November 2011.

# The following analysis investigates which types of severe weather events are most harmful on:

1. Health (injuries and fatalities)
2. Property and crops (economic consequences)

Information on the Data: DocumentNational Weather Serviceation

## 2: Data Processing

### 2.1: Data Loading

```
library(dplyr)
library(ggplot2)
```

```
storm_data <- read.csv("StromData.csv.bz2")
colnames(storm_data)
```

[1] “STATE__” *"BGN_DATE" "BGN_TIME" "TIME_ZONE" "COUNTY"*
*[6] "COUNTYNAME" "STATE" "EVTYPE" "BGN_RANGE" "BGN_AZI"*
*[11] "BGN_LOCATI" "END_DATE" "END_TIME" "COUNTY_END" "COUNTYENDN" [16] "END_RANGE" "END_AZI" "END_LOCATI" "LENGTH" "WIDTH"*
*[21] "F" "MAG" "FATALITIES" "INJURIES" "PROPDMG"*
*[26] "PROPDMGEXP" "CROPDMG" "CROPDMGEXP" "WFO" "STATEOFFIC" [31] "ZONENAMES" "LATITUDE" "LONGITUDE" "LATITUDE_E" "LONGITUDE”* [36] “REMARKS” “REFNUM”
### 2.2: Selecting Data

we will be selecting only those columns which we need for the analysis from the dataset. AS we are only interested in the correlation between the Nation - wide event to Health and Economic consequences, we will be using the following columns only:

-EVTYPE: event type

-FATALITIES: number of fatalities

-INJURIES: number of injuries

-PROPDMG: property damage (dollars)

-PROPDMGEXP: magnitude of property damage (K = thousands, M = millions, B = billions)

-CROPDMG: crop damage (dollars)

-CROPDMGEXP: magnitude of crop damage (H = hundreds, K = thousands, M = millions, B = billions)

```
storm_data <- storm_data[, c('EVTYPE', 'FATALITIES', 'INJURIES', 'PROPDMG', 'PROPDMGEXP', 'CROPDMG', 'C
head(storm_data)
```

EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP 1 TOR-
NADO 0 15 25.0 K 0
2 TORNADO 0 0 2.5 K 0
3 TORNADO 0 2 25.0 K 0
4 TORNADO 0 2 2.5 K 0
5 TORNADO 0 2 2.5 K 0
6 TORNADO 0 6 2.5 K 0

```
summary(storm_data)
```

```
EVTYPE            FATALITIES          INJURIES            PROPDMG
```

Length:902297 Min. : 0.0000 Min. : 0.0000 Min. : 0.00
Class :character 1st Qu.: 0.0000 1st Qu.: 0.0000 1st Qu.: 0.00
Mode :character Median : 0.0000 Median : 0.0000 Median : 0.00
Mean : 0.0168 Mean : 0.1557 Mean : 12.06
3rd Qu.: 0.0000 3rd Qu.: 0.0000 3rd Qu.: 0.50
Max. :583.0000 Max. :1700.0000 Max. :5000.00
PROPDMGEXP CROPDMG CROPDMGEXP
Length:902297 Min. : 0.000 Length:902297
Class :character 1st Qu.: 0.000 Class :character
Mode :character Median : 0.000 Mode :character
Mean : 1.527
3rd Qu.: 0.000
Max. :990.000

## 3: Data Analysis

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

```
Storm_data_q1 <- storm_data[, 1:3] %>%
                 group_by(EVTYPE) %>%
                 summarise_all(sum)
summary(Storm_data_q1)
```

```
EVTYPE            FATALITIES         INJURIES
```

Length:985 Min. : 0.00 Min. : 0.0
Class :character 1st Qu.: 0.00 1st Qu.: 0.0
Mode :character Median : 0.00 Median : 0.0
Mean : 15.38 Mean : 142.7
3rd Qu.: 0.00 3rd Qu.: 0.0
Max. :5633.00 Max. :91346.0

```
# Top 10 events by Fatalities
top_fatalities <- Storm_data_q1[order(Storm_data_q1$FATALITIES, decreasing = TRUE), ]
# Top 10 events by Injuries
top_injuries <- Storm_data_q1[order(Storm_data_q1$INJURIES, decreasing = TRUE), ]
# Top 10 events that causes both Fatalities and Injuries
Storm_data_q1$total <- rowSums(Storm_data_q1[, 2:3])
top_health <- Storm_data_q1[order(Storm_data_q1$total, decreasing = "TRUE"), ]
```

**3.2: Across the United States, which types of events have the greatest economic consequences?**

Function for converting the 'EXP' column code to power of 10's

```
value_conversion <- function(expType) {
  if (expType %in% c('h', 'H')) {
    return(2)
  } else if (expType %in% c('k', 'K')) {
    return(3)
  } else if (expType %in% c('m', 'M')) {
    return(6)
  } else if (expType %in% c('b', 'B')) {
    return(9)
  } else if (suppressWarnings(!is.na(as.numeric(expType)))) {
    #won't show: NAs introduced by coercion when given a non-number
    return(as.numeric(expType))
  } else {
    return(0)
  }
}
```

```
Storm_data_q2 <- storm_data[, c(1, 4:7)] %>% rowwise() %>% mutate(prop_dmg = PROPDMG*10**value_conversi

head(Storm_data_q2)
```

# A tibble: 6 x 7

# Rowwise:

EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP prop_dmg crop_dmg 1 TORNADO 25 K 0 "" 25000 0 2 TORNADO 2.5 K 0 "" 2500 0 3 TORNADO 25 K 0 "" 25000 0 4 TORNADO 2.5 K 0 "" 2500 0 5 TORNADO 2.5 K 0 "" 2500 0 6 TORNADO 2.5 K 0 "" 2500 0

```r
Storm_data_q2 <- Storm_data_q2[, c(1, 6, 7)] %>%
  group_by(EVTYPE) %>%
  summarise_all(sum)
summary(Storm_data_q2)
```

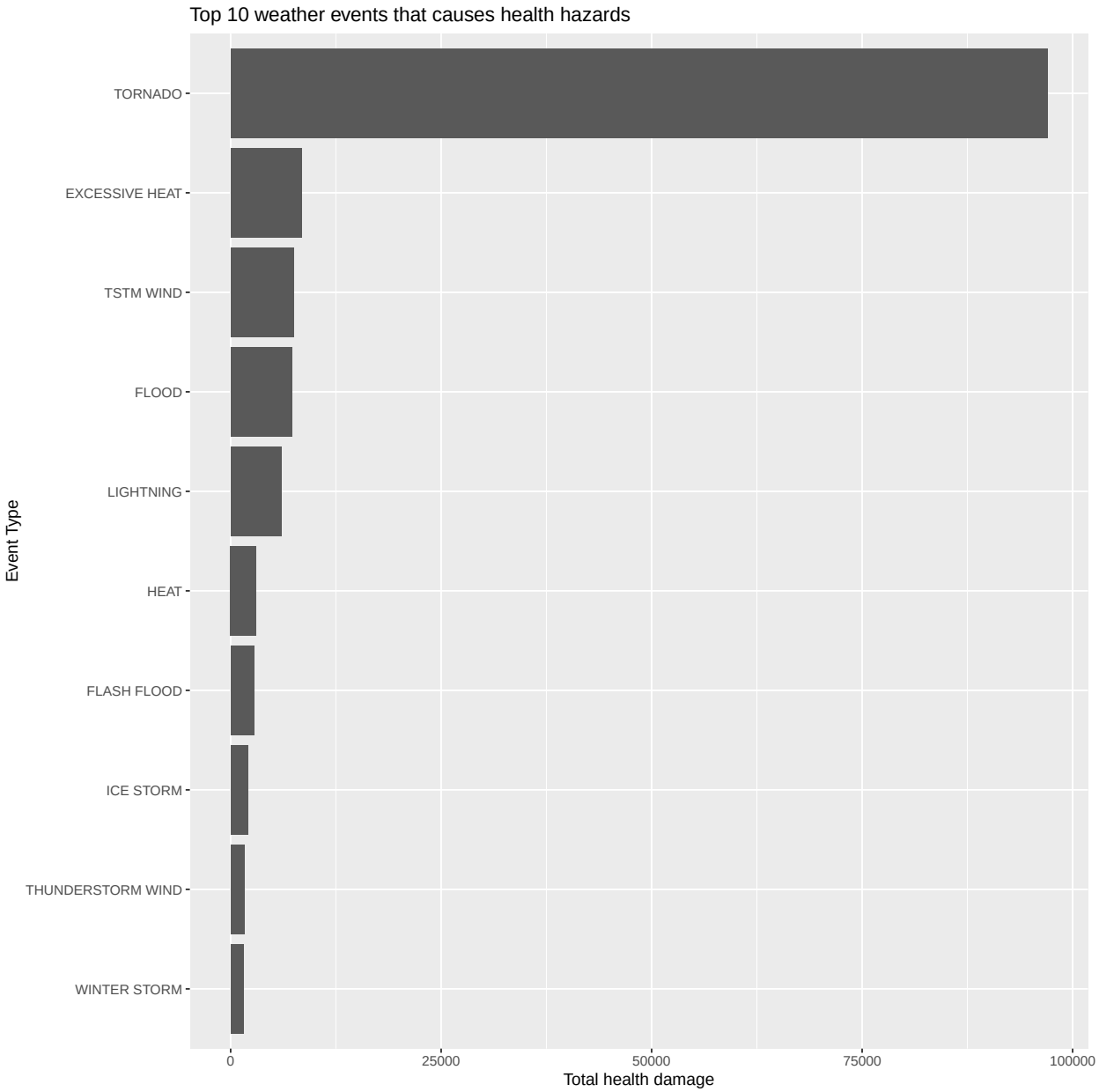
```
EVTYPE             prop_dmg            crop_dmg

Length:985 Min. :0.000e+00 Min. :0.000e+00
Class :character 1st Qu.:0.000e+00 1st Qu.:0.000e+00
Mode :character Median :0.000e+00 Median :0.000e+00
Mean :4.347e+08 Mean :4.985e+07
3rd Qu.:5.105e+04 3rd Qu.:0.000e+00
Max. :1.447e+11 Max. :1.397e+10
```

```r
# Top 10 events by Property dmage
top_prop_dmg <- Storm_data_q2[order(Storm_data_q2$prop_dmg, decreasing = TRUE), ]
# Top 10 events by Crop Damage
top_crop_dmg <- Storm_data_q2[order(Storm_data_q2$crop_dmg, decreasing = TRUE), ]
# Top 10 events that causes both Property and Crop Damage
Storm_data_q2$total <- rowSums(Storm_data_q2[, 2:3])
top_economic <- Storm_data_q2[order(Storm_data_q2$total, decreasing = "TRUE"), ]
```

**4: Results**

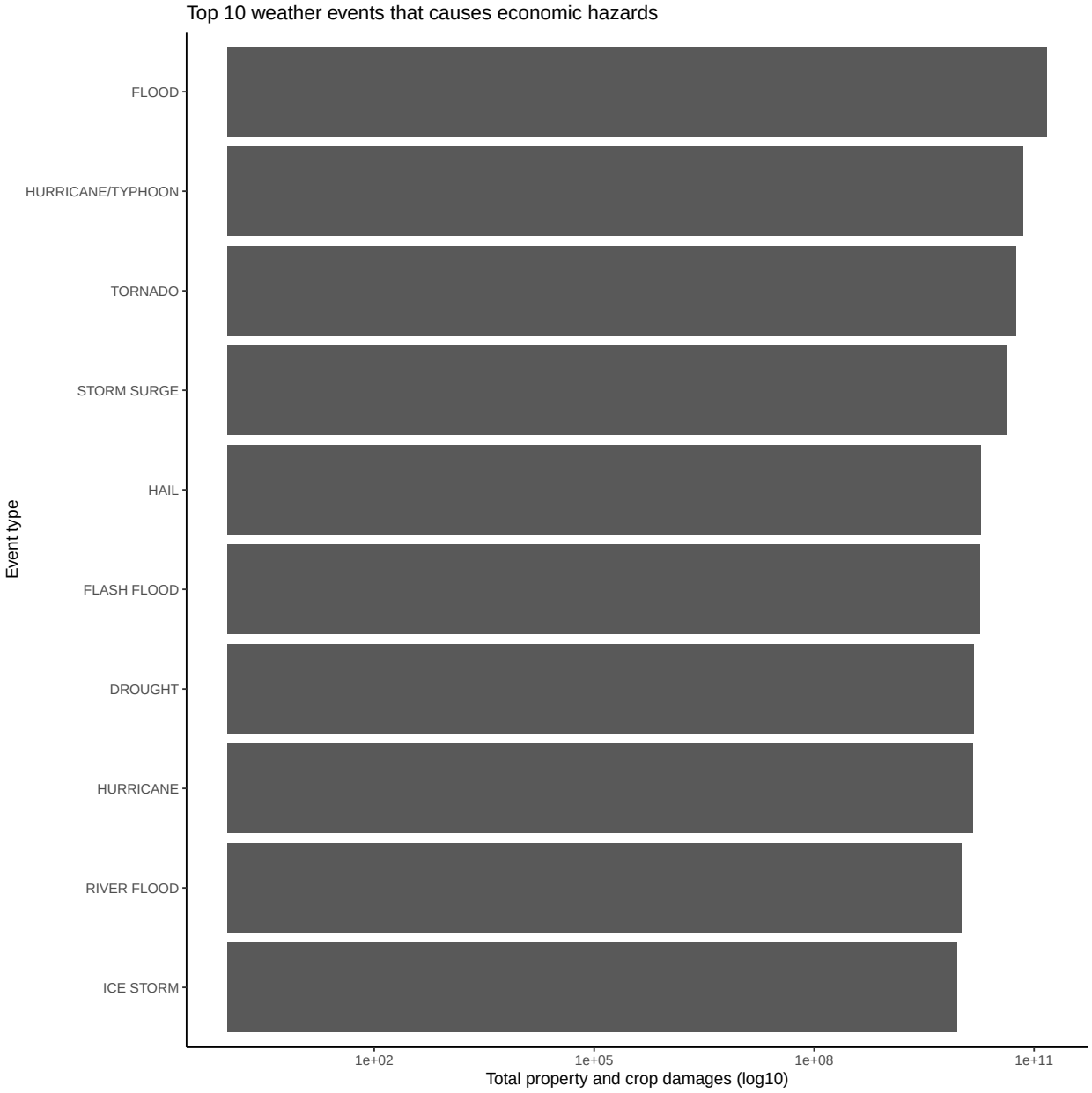
```r
options(repr.plot.width = 14, repr.plot.height = 8)
ggplot(data = top_health[1:10,], aes(x=reorder(EVTYPE, total), y=total))+
  geom_bar(stat = 'identity') +
  coord_flip() +
  xlab('Event Type') +
  ylab('Total health damage') +
  ggtitle('Top 10 weather events that causes health hazards')
```

**Find types of events that are most harmful with respect to population health**




```
ggplot(data = top_economic[1:10, ], aes(x = reorder(EVTYPE, total), y = total)) +
  #need to use reorder to prevent the categorical data from reordering
  geom_bar(stat = 'identity') +
  coord_flip() +
  scale_y_continuous(trans = 'log10') +
  xlab('Event type') +
  ylab('Total property and crop damages (log10)') +
  ggtitle('Top 10 weather events that causes economic hazards') +
  theme_classic()
```

**Find types of events that have the greatest economic consequences**


Top 10 weather events that causes economic hazards
Event type
FLOOD
HURRICANE/TYPHOON
TORNADO
STORM SURGE
HAIL
FLASH FLOOD
DROUGHT
HURRICANE
RIVER FLOOD
ICE STORM
1e+02
1e+05
1e+08
1e+11
Total property and crop damages (log10)