Reproducible Research: Peer Assessment 2

Introduction

Reproducible Research - Peer Assignment 2
Author: Tim Westran
Purpose: Answer questions about the NOAA Storm Database dataset
Data set used: https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
Github repo for the Course: Reproducible Research

Synopsis

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. Various questions have been posed, and are answered below using the dataset. Code is provided for the analysis performed.

Questions answered: 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

Load libraries and set options

```
## Ensure we have loaded the libraries we need.
library("data.table")
library("ggplot2")

## disable scientific notation
options(scipen = 999)
```

Download data

Download the bz2 file and read it into a dataframe.

```
## Set the URL of the file to download
fileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

## Download the file to the working directory, save with filename:  repdata-data-StormData.csv
.bz2
download.file(fileUrl, destfile = paste0("./", "repdata-data-StormData.csv.bz2"))

## Read in .bz2 file and store in a dataframe called stormData
stormData <- read.csv(bzfile("repdata-data-StormData.csv.bz2"))

## Convert StormData dataframe to datatable called stormDataTable
stormDataTable <- as.data.table(stormData)
```

Extract columns we care about.

We only need to keep some of the columns, so we remove everything except: EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP

```
## Remove everything except EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDM
GEXP
columnsToDrop <- colnames(stormDataTable[, !c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "
PROPDMGEXP", "CROPDMG", "CROPDMGEXP")])

## Set all other columns to NULL.
stormDataTable[, c(columnsToDrop) := NULL]

## Grab all rows where a fatality or injury occurred, looking only at the columns we care abou
t.
stormDataTable <- stormDataTable[(EVTYPE != "?" & (INJURIES > 0 | FATALITIES > 0 | PROPDMG > 0
 | CROPDMG > 0)), c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "C
ROPDMGEXP") ]
```

convert dollars and exponents to a single value

To determine the impact of weather on the economy, we need to example property and crop damage figures. Property damage is represented by PROPDMG, which contains the raw dollars, and PROPDMGEXP, which contains the exponents. We need to transform these two variables into a single value for each row.

Similarly, crop damage is represented by CROPDMG, which contains the raw dollars, and CROPDMGEXP, which contains the exponents. We need to transform these two variables into a single value for each row.

```
## First, we modify all column names so they are all uppercase.
colNamesUpper <- c("PROPDMGEXP", "CROPDMGEXP")
stormDataTable[,  (colNamesUpper) := c(lapply(.SD, toupper)), .SDcols = colNamesUpper]

## Now, convert all property and crop damage exponent symbols to actual values.  The data set
has different
## symbols mapped to different exponents.
## First, property damage
propertyDmgConv <-  c("\"\"" = 1, "-" = 1, "+" = 1, "0" = 1, "1" = 10, "2" = 100,
               "H" = 100,  "3" = 1000,  "K" = 1000, "4" = 10000, "5" = 10^5, "6" = 10^6,
               "M" = 10^6, "7" = 10^7, "8" = 10^8, "9" = 10^9, "B" = 10^9)

## Then crop damage exponent symbol mapping.
cropDmgConv <-  c("\"\"" = 1, "?" = 1, "0" = 1, "K" = 1000, "M" = 10^6, "B" = 10^9)

## Now, insert our converted exponents back into the data table.
## First, property damage.
stormDataTable[, PROPDMGEXP := propertyDmgConv[as.character(stormDataTable[,PROPDMGEXP])]]

## Setting all exponents with a value of NA to 1 (no exponent).
stormDataTable[is.na(PROPDMGEXP), PROPDMGEXP := 1 ]

## Then, property damage.
stormDataTable[, CROPDMGEXP := cropDmgConv[as.character(stormDataTable[,CROPDMGEXP])] ]
```

```
## Again, setting all exponents with a value Of NA to 1 (no exponent).
stormDataTable[is.na(CROPDMGEXP), CROPDMGEXP := 1 ]

## Add two new columns, propertyCost and cropCost
stormDataTable <- stormDataTable[, .(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, proper
tyCost = PROPDMG * PROPDMGEXP, CROPDMG, CROPDMGEXP, cropCost = CROPDMG * CROPDMGEXP)]
```

Calculate the total cost and fatalities / injuries

Here, we total up both the property and crop costs, then add all costs together to generate a total cost.

```
## Create a new data table called totalCost that holds all summed cost figures.
totalCost <- stormDataTable[, .(propertyCost = sum(propertyCost), cropCost = sum(cropCost), To
tal_Cost = sum(propertyCost) + sum(cropCost)), by = .(EVTYPE)]

## Order from highest to lowest.
totalCost <- totalCost[order(-Total_Cost), ]

## Grab the top 10 costs and store them in the totalCost data table.
totalCost <- totalCost[1:10, ]

## Now, create a new data table called totalInjuriesAndFatailities that holds all of the summe
d injury
## and fatality statistics.
totalInjuriesAndFatailities <- stormDataTable[, .(FATALITIES = sum(FATALITIES), INJURIES = sum
(INJURIES), totals = sum(FATALITIES) + sum(INJURIES)), by = .(EVTYPE)]

## Order from highest to lowest.
totalInjuriesAndFatailities <- totalInjuriesAndFatailities[order(-FATALITIES), ]

## Grab the top 10 injuries and fatalities, and store them in the data table.
totalInjuriesAndFatailities <- totalInjuriesAndFatailities[1:10, ]
```

Results

Health Inpact of Weather Events

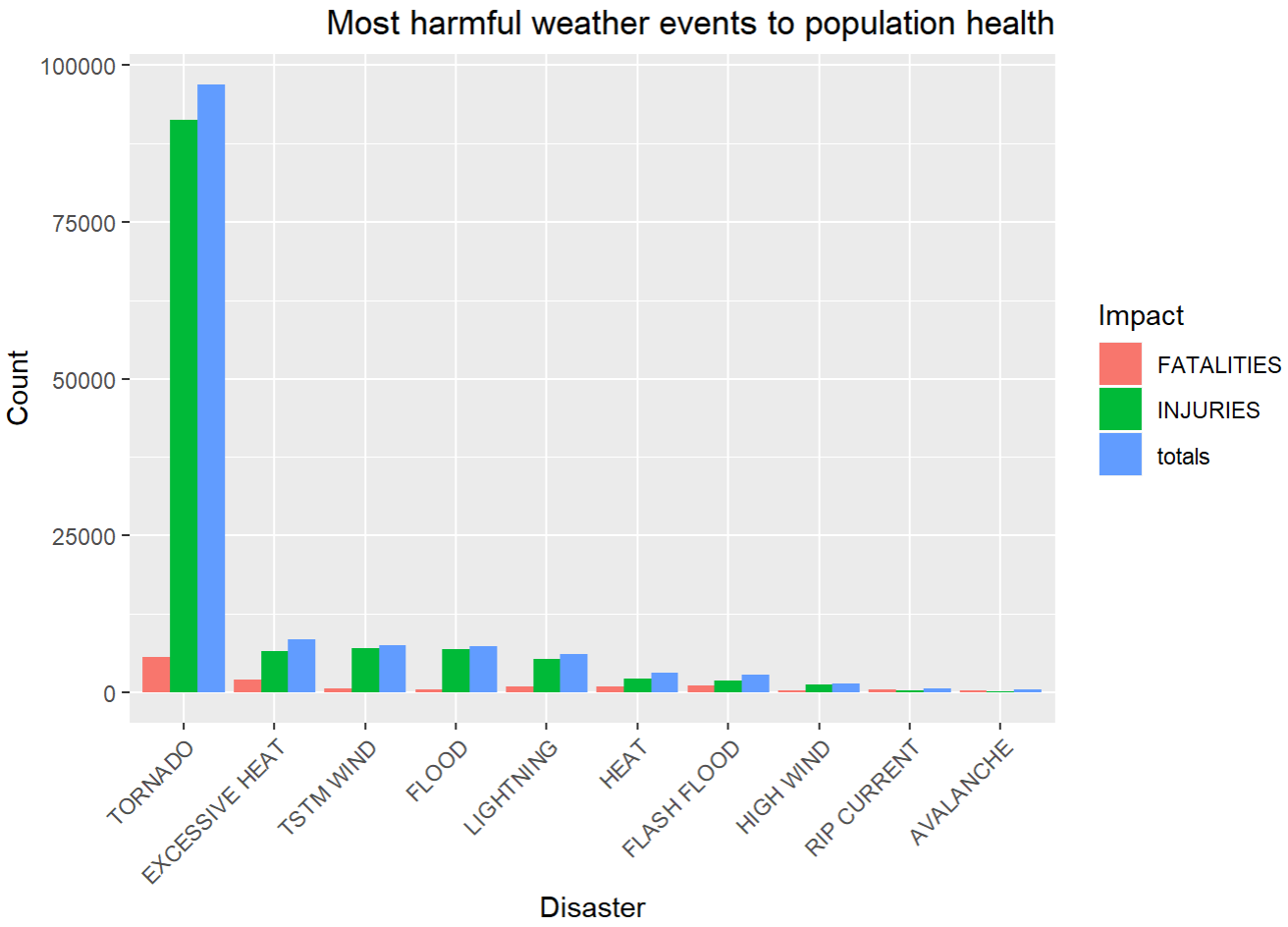
Health impact of weather events - Weather events which are most harmful.

```
## First, create a new variable named harmfulEvents.
harmfulData <- melt(totalInjuriesAndFatailities, id.vars="EVTYPE", variable.name = "Impact")

## Create the chart to answer the question:  Which types of events are most harmful to populat
ion health?
healthChart <- ggplot(harmfulData, aes(x=reorder(EVTYPE, -value), y=value)) +
    geom_bar(stat="identity", aes(fill=Impact), position="dodge") +
    ylab("Count") +
    xlab("Disaster") +
    theme(axis.text.x = element_text(angle=45, hjust=1))+
    ggtitle("Most harmful weather events to population health") +
    theme(plot.title = element text(hjust = 1))
```

Chart - Which weather events are most harmful?

```
## Display the chart of harmful effects
healthChart
```



From the above chart, we can see that Tornados are the most harmful events, with 5,633 fatalities, and 91,346 injuries.

This is followed by excessive heat, with 1,903 fatalities, and 6,525 injuries.

With flash floods coming in 3rd, with 978 fatalities, and 1,777 injuries.

Economic Inpact of Weather Events

Economic impact of weather events - Weather events which have the greatest economic impact

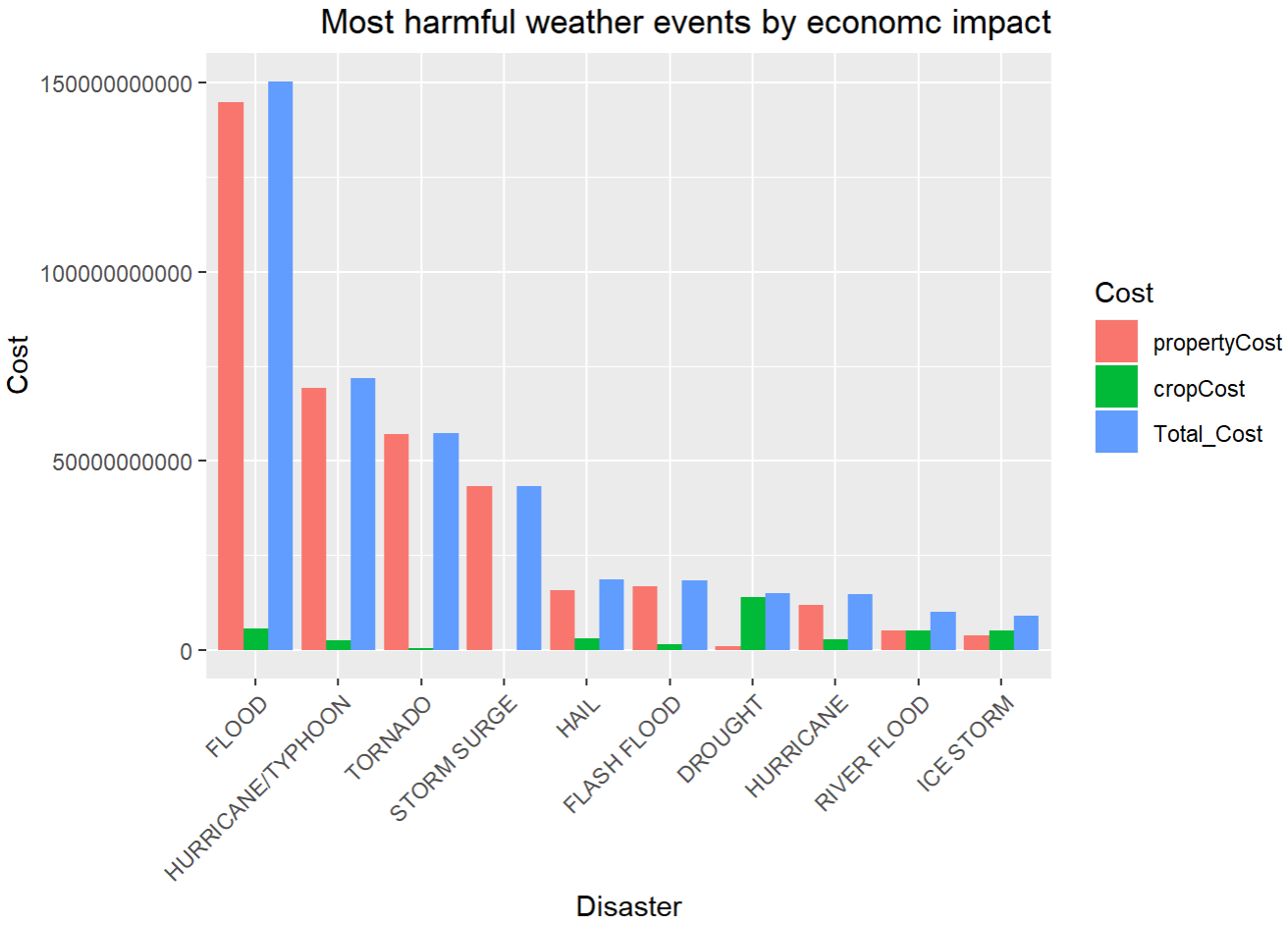
```
## First, create a new variable named Cost.
economicImpact <- melt(totalCost, id.vars="EVTYPE", variable.name = "Cost")

## Create the chart to answer the question:  Which types of events are most harmful to populat
ion health?
economicImpactChart <- ggplot(economicImpact, aes(x=reorder(EVTYPE, -value), y=value)) +
    geom_bar(stat="identity", aes(fill=Cost), position="dodge") +
    ylab("Cost") +
    xlab("Disaster") +
```

```
    theme(axis.text.x = element_text(angle=45, hjust=1)) +
    ggtitle("Most harmful weather events by economc impact") +
    theme(plot.title = element_text(hjust = 1))
```

Chart - Which weather events have the greatest economic impact?

```
## Display the chart of harmful effects
economicImpactChart
```



From the above chart, we can see that floods are the most harmful events economically, with 144,657,709,807 in property damage, 5,661,968,450 in crop damage, and 150,319,678,257 in total damage.

This is followed by hurricanes / typhoons, with 69,305,840,000 in property damage, 2,607,872,800 in crop damage, and 71,913,712,800 in total damage.

With tornados coming in 3rd, with 56,947,380,677 in property damage, 414,953,270 in crop damage, and 57,362,333,947 in total damage.

```
## Display top 10 most harmful events economically.
head(totalCost,10)
```

```
##                EVTYPE propertyCost    cropCost   Total_Cost
##  1:             FLOOD 144657709807  5661968450 150319678257
##  2: HURRICANE/TYPHOON  69305840000  2607872800  71913712800
##  3:           TORNADO  56947380677   414953270  57362333947
```

```
##  4:       STORM SURGE  43323536000        5000  43323541000
##  5:              HAIL  15735267513  3025954473  18761221986
##  6:       FLASH FLOOD  16822673979  1421317100  18243991079
##  7:           DROUGHT   1046106000 13972566000  15018672000
##  8:         HURRICANE  11868319010  2741910000  14610229010
##  9:       RIVER FLOOD   5118945500  5029459000  10148404500
## 10:         ICE STORM   3944927860  5022113500   8967041360
```

```
## Display top 10 most harmful events health-wise.
head(totalInjuriesAndFatailities,10)
```

```
##             EVTYPE FATALITIES INJURIES totals
##  1:        TORNADO       5633    91346  96979
##  2: EXCESSIVE HEAT       1903     6525   8428
##  3:    FLASH FLOOD        978     1777   2755
##  4:           HEAT        937     2100   3037
##  5:      LIGHTNING        816     5230   6046
##  6:      TSTM WIND        504     6957   7461
##  7:          FLOOD        470     6789   7259
##  8:    RIP CURRENT        368      232    600
##  9:      HIGH WIND        248     1137   1385
## 10:      AVALANCHE        224      170    394
```