Reproducible Research StormData Assignment 2

Neel9jain

13/11/2021

Reproducible research Assignment

This Report summarizes the Effects of different Natural events on the Human health as well as the Economic impacts of these events.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

This Involves use of the Comma separated CSV file "Storm data" .The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records

Data Processing

The data for the project is downloaded from the Coursera website .The downloaded data comes in the form of a zipped file in bz2 algorithm.

The file is first downloaded and then comma separated file is converted to human readable format using the read.table command.

```
download.file(url="https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
              destfile = "StormData.csv.bz2")


Storm_Data <- read.table("StormData.csv.bz2",header=T, quote="\"", sep=",")
```

This assignment answers two questions 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?

So we will get the results in two parts

1. Effect of different events on Human Health

First we read the data for the Fatalities cased by variuos events which will evnetually be plotted along with the other occurances.

```
#1# Fatalities
Event_Fatalities <- tapply(Storm_Data$FATALITIES, Storm_Data$EVTYPE, sum)
##Top 5 Fatal
```

```
Top_5_Fatal <- Event_Fatalities[order(Event_Fatalities,decreasing = TRUE)][1:5]
Df_Top_5_Fatal <- data.frame(names(Top_5_Fatal),Top_5_Fatal)
names(Df_Top_5_Fatal) <- c("Event","Fatalities")
## Max Fatalities
Max_Fatalities <- Event_Fatalities[Event_Fatalities == max(Event_Fatalities)]
```

The Max number of fatalities have occurred from the following event:

```
print (Max_Fatalities)
```

```
## TORNADO
##    5633
```

These events have been harmful to the human health by the way of injuries also :

```
#2# Injuries
Event_Injuries <- tapply(Storm_Data$INJURIES, Storm_Data$EVTYPE, sum)
##TOP 5 Injuries
Top_5_Injuries <- Event_Injuries[order(Event_Injuries,decreasing = TRUE)][1:5]
Df_Top_5_Injuries <- data.frame(names(Top_5_Injuries),Top_5_Injuries)
names(Df_Top_5_Injuries) <- c("Event","Injuries")
## Max Injuries
Max_Injuries <- Event_Injuries[Event_Injuries == max(Event_Injuries)]
```

The Max number of Injuries have occurred from the following event:

```
print (Max_Injuries)
```

```
## TORNADO
##   91346
```

Results

First we will plot the results for the first question i.e

Across the United States, which types of events are most harmful with respect to population health?

```
## The plot is made via ggplot package  and arranged via gridExtra package
library(gridExtra)
library(ggplot2)

## Plotting the Data for Fatalities
Plot1 <- qplot(x=Df_Top_5_Fatal$Fatalities,y=Df_Top_5_Fatal$Event,
              geom = "col",xlab = "Fatalities" ,ylab ="Event")

## Plotting the Data for Injuries
Plot2 <- qplot(x=Df_Top_5_Injuries$Injuries,y=Df_Top_5_Injuries$Event,
              geom = "col",xlab = "Injuries" ,ylab ="Event")

grid.arrange(Plot1,Plot2 ,ncol=2)
```
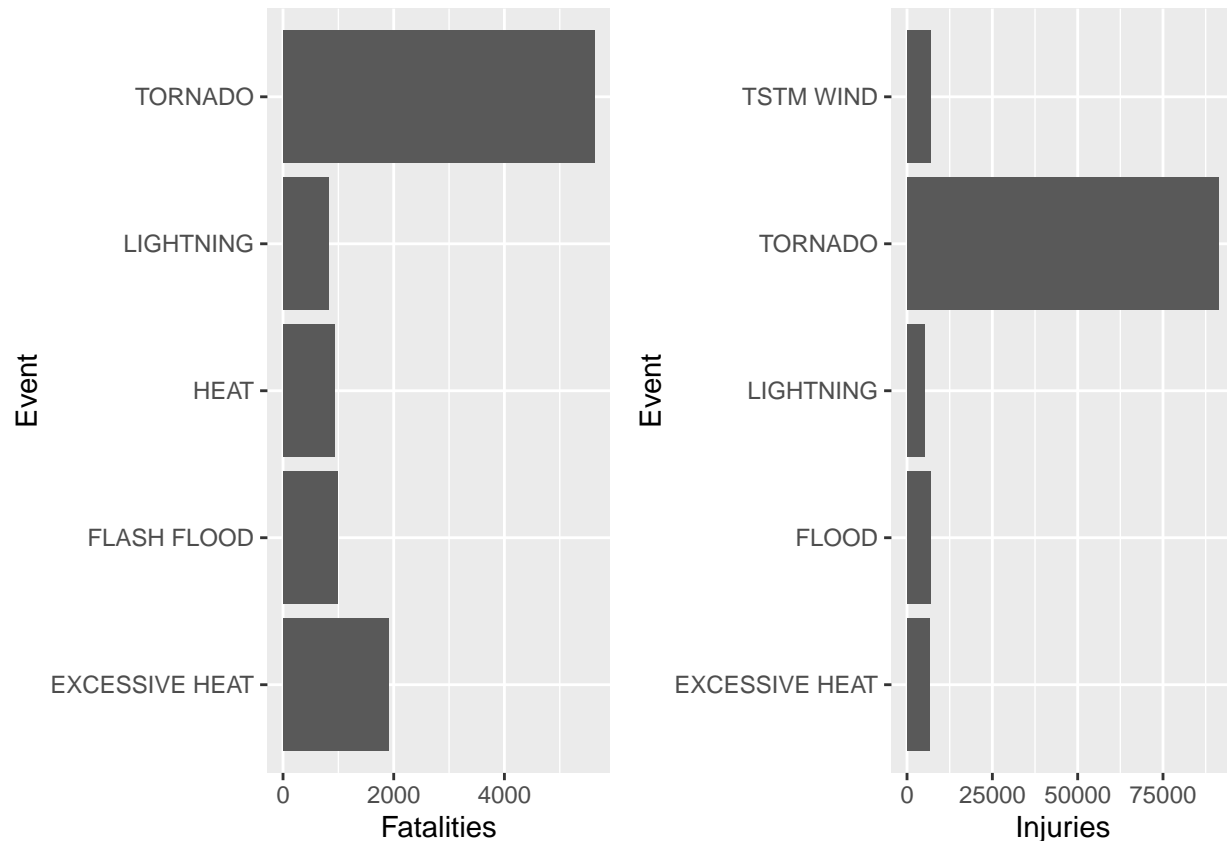
2. Effect of different events on Economics

Since we have to plot the economics impact of the various events by the way of Property Damage and Crop Damages hence analysing the Data

```r
#1# Property Damage
Prop_Damage <- tapply(Storm_Data$PROPDMG, Storm_Data$EVTYPE, sum)
##Top 5 property Damage
Top_5_Property_DM <- Prop_Damage[order(Prop_Damage,decreasing = TRUE)][1:5]
Df_Top_5_Property_DM <- data.frame(names(Top_5_Property_DM),Top_5_Property_DM)
names(Df_Top_5_Property_DM) <- c("Event","Property Damage")
## Max Property Damage
Max_Prop_Damage <- Prop_Damage[Prop_Damage == max(Prop_Damage)]
```

The Max Property Damage has occurred from the following event:

```r
print (Max_Prop_Damage)
```

```
## TORNADO
## 3212258
```

```r
#2# Crop Damage
Crop_Damage <- tapply(Storm_Data$CROPDMG, Storm_Data$EVTYPE, sum)
## Top 5 crop Damage
```

```
Top_5_Crop_DM <- Crop_Damage[order(Crop_Damage,decreasing = TRUE)][1:5]
Df_Top_5_Crop_DM <- data.frame(names(Top_5_Crop_DM),Top_5_Crop_DM)
names(Df_Top_5_Crop_DM) <- c("Event","Crop Damage")
## Max Crop_Damage
Max_Crop_Damage <- Crop_Damage[Crop_Damage == max(Crop_Damage)]
```

The Max Crop Damage has occurred from the following event:

```
print (Max_Crop_Damage)
```

```
##       HAIL
## 579596.3
```

Results

Secondly we will plot the results for the second question i.e

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

```
## The plot is made via ggplot package  and arranged via gridExtra package
library(gridExtra)
library(ggplot2)

## Plotting the Data for Property Damage
Plot3 <- qplot(x=Df_Top_5_Property_DM$`Property Damage`,y=Df_Top_5_Property_DM$Event,
               geom = "col",xlab = "Propety Damage (USD)" ,ylab ="Event")

## Plotting the Data for Crop Damage
Plot4 <- qplot(x=Df_Top_5_Crop_DM$`Crop Damage`,y=Df_Top_5_Crop_DM$Event,
               geom = "col",xlab = "Crop Damage (USD)" ,ylab ="Event")

grid.arrange(Plot3,Plot4 ,ncol=2)
```
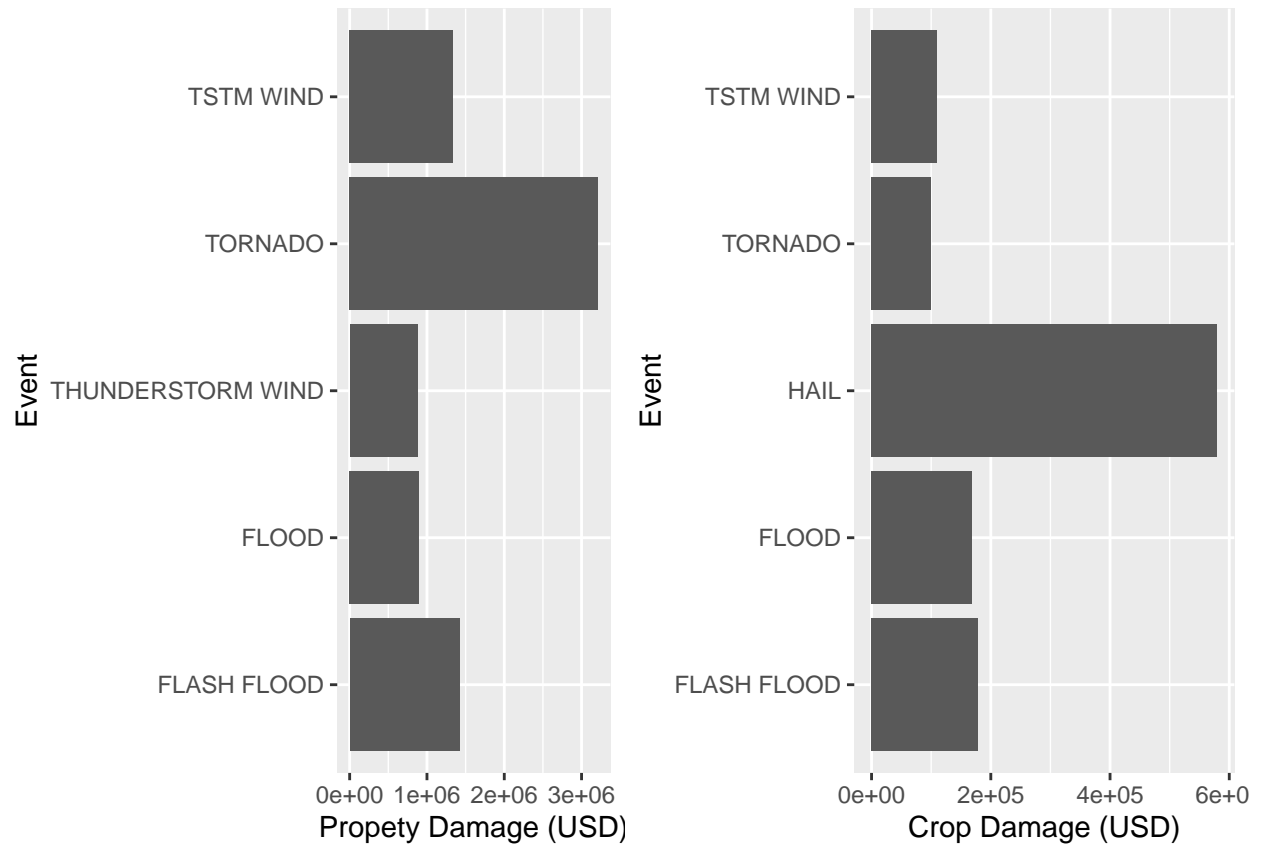
The document contains all the reproducible code and plot packages and may be used to reproduce the research .

PS : All the data has been taken from the public sources.