

Reproducible Research - Week 4 Peer Project

Vignesh C Iyer

7/12/2020

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

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.

Assignment

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. You must use the database to answer the questions below and show the code for your entire analysis. Your analysis can consist of tables, figures, or other summaries. You may use any R package you want to support your analysis.

Data

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site:

Storm Data

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

- National Weather Service Storm Data Documentation
- National Climatic Data Center Storm Events FAQ

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. More recent years should be considered more complete.

Data Pre-processing

The Storm Data is fetched, downloaded to the local system and then its contents are read based on the code given below

```
# This section deals with the downloading the compressed file and  
# extracting its contents.  
  
stormData <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
  
# The file is downloaded using the download.file function.  
download.file(stormData, destfile = "../StormData.csv.bz2")
```

```

# reading data from the file
readStormData <- read.csv("../StormData.csv.bz2")

# Fetching column names of Storm Data using the colNames function
colnames(readStormData)

## [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"

str(readStormData)

```

```

## 'data.frame': 902297 obs. of 37 variables:
## $ STATE_ : num 1 1 1 1 1 1 1 1 1 1 ...
## $ BGN_DATE : chr "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1951 0:00:00" "6/8/1951 0:00:00" .
## $ BGN_TIME : chr "0130" "0145" "1600" "0900" ...
## $ TIME_ZONE : chr "CST" "CST" "CST" "CST" ...
## $ COUNTY : num 97 3 57 89 43 77 9 123 125 57 ...
## $ COUNTYNAME: chr "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
## $ STATE : chr "AL" "AL" "AL" "AL" ...
## $ EVTYPE : chr "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
## $ BGN_RANGE : num 0 0 0 0 0 0 0 0 0 0 ...
## $ BGN_AZI : chr "" "" "" "" ...
## $ BGN_LOCATI: chr "" "" "" "" ...
## $ END_DATE : chr "" "" "" "" ...
## $ END_TIME : chr "" "" "" "" ...
## $ COUNTY_END: num 0 0 0 0 0 0 0 0 0 0 ...
## $ COUNTYENDN: logi NA NA NA NA NA NA ...
## $ END_RANGE : num 0 0 0 0 0 0 0 0 0 0 ...
## $ END_AZI : chr "" "" "" "" ...
## $ END_LOCATI: chr "" "" "" "" ...
## $ LENGTH : num 14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
## $ WIDTH : num 100 150 123 100 150 177 33 33 100 100 ...
## $ F : int 3 2 2 2 2 2 2 1 3 3 ...
## $ MAG : num 0 0 0 0 0 0 0 0 0 0 ...
## $ FATALITIES: num 0 0 0 0 0 0 0 0 1 0 ...
## $ INJURIES : num 15 0 2 2 2 6 1 0 14 0 ...
## $ PROPDMG : num 25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
## $ PROPDMGEXP: chr "K" "K" "K" "K" ...
## $ CROPDMG : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CROPDMGEXP: chr "" "" "" "" ...
## $ WFO : chr "" "" "" "" ...
## $ STATEOFFIC: chr "" "" "" "" ...
## $ ZONENAMES : chr "" "" "" "" ...
## $ LATITUDE : num 3040 3042 3340 3458 3412 ...
## $ LONGITUDE : num 8812 8755 8742 8626 8642 ...
## $ LATITUDE_E: num 3051 0 0 0 0 ...
## $ LONGITUDE_ : num 8806 0 0 0 0 ...
## $ REMARKS : chr "" "" "" "" ...

```

```
## $ REFNUM      : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
# Fetching first few rows of Storm Data
```

```
head(readStormData)
```

```
##   STATE__      BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1      1  4/18/1950 0:00:00    0130     CST    97    MOBILE    AL  TORNADO
## 2      1  4/18/1950 0:00:00    0145     CST     3    BALDWIN   AL  TORNADO
## 3      1  2/20/1951 0:00:00    1600     CST    57    FAYETTE    AL  TORNADO
## 4      1   6/8/1951 0:00:00    0900     CST    89    MADISON    AL  TORNADO
## 5      1 11/15/1951 0:00:00    1500     CST    43    CULLMAN    AL  TORNADO
## 6      1 11/15/1951 0:00:00    2000     CST    77 LAUDERDALE   AL  TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0              0              0      NA
## 2         0              0              0      NA
## 3         0              0              0      NA
## 4         0              0              0      NA
## 5         0              0              0      NA
## 6         0              0              0      NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0              14.0   100 3  0      0      15      25.0
## 2         0              2.0   150 2  0      0       0       2.5
## 3         0              0.1   123 2  0      0       2      25.0
## 4         0              0.0   100 2  0      0       2       2.5
## 5         0              0.0   150 2  0      0       2       2.5
## 6         0              1.5   177 2  0      0       6       2.5
##   PROPDMGEXP CROPDGM CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K        0          0      3040      8812
## 2          K        0          0      3042      8755
## 3          K        0          0      3340      8742
## 4          K        0          0      3458      8626
## 5          K        0          0      3412      8642
## 6          K        0          0      3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1      3051      8806         1
## 2         0         0         2
## 3         0         0         3
## 4         0         0         4
## 5         0         0         5
## 6         0         0         6
```

```
# fetching the unique event type in the Storm Data
```

```
head(unique(readStormData$EVTYPE))
```

```
## [1] "TORNADO"      "TSTM WIND"    "HAIL"
## [4] "FREEZING RAIN" "SNOW"         "ICE STORM/FLASH FLOOD"
```

We notice that the Date format is that of a Character from the below code

```
class(readStormData$BGN_DATE)
```

```
## [1] "character"
```

We will convert it to Date format using the as.Date function and assign it to a new variable stormDate

```
stormDate <- as.Date(readStormData$BGN_DATE, format = "%m%d%Y %H:%m:%s")
class(stormDate)
```

```
## [1] "Date"
```