Title: “PA2_Template”

Author: “Michael Malin Jr”

Date: “Tuesday, October 21, 2014”

Title:

Reproducible Research: Peer Assessment 2

Synopsis:

Tornadoes have been the most harmful weather related event since 1950. Producing over 90,000 injuries and over 5,000 fatalities. The second most harmful event does not come close to the number of injuries or fatalities cased by tornadoes. Thunderstorms had the second most injuries at over 6,000 injuries. Excessive Heat had the 2nd most fatalities at over 1,900 deaths. It might be interesting to see in future analysis to amount of fatalities due to excessive heat over the years. I would assume that the amount of excessive heat deaths have increased over the years due to record amount of heat temperatures recorded in the past decade. Tornadoes top the list of most damaging weather related event with over 48 million of total damage. A not so close second is Flood with over 34 million. Presented below is a high level view the most damaging and harmful weather related events. Additional analysis is still needed into studying the impact of these events beyond what I have presented. Property damage totals have not been adjusted for inflation.

Data Processing:

First, to unzip a bz2 file, intall the “R.utils” package into R. Unzip the storm data file using the bunzip2() function from “R.utils”. Read the storm data into R by using read.csv(). Load the “dplyr” library into R and use tbl_df() function to view your data in R:

```
# install.packages("R.utils")
suppressWarnings(library("R.utils"))
bunzip2("repdata-data-StormData.csv.bz2")
```

```
## Error: File already exists: repdata-data-StormData.csv
```

```
data = read.csv("repdata-data-StormData.csv")

suppressWarnings(library("dplyr"))
suppressWarnings(library(ggplot2))
data_df = tbl_df(data); data_df
```

```
## Source: local data frame [902,297 x 37]
##
##    STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME
## 1        1  4/18/1950 0:00:00     0130       CST     97     MOBILE
## 2        1  4/18/1950 0:00:00     0145       CST      3    BALDWIN
## 3        1  2/20/1951 0:00:00     1600       CST     57    FAYETTE
## 4        1   6/8/1951 0:00:00     0900       CST     89    MADISON
## 5        1 11/15/1951 0:00:00     1500       CST     43    CULLMAN
## 6        1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE
## 7        1 11/16/1951 0:00:00     0100       CST      9     BLOUNT
## 8        1  1/22/1952 0:00:00     0900       CST    123 TALLAPOOSA
## 9        1  2/13/1952 0:00:00     2000       CST    125 TUSCALOOSA
## 10       1  2/13/1952 0:00:00     2000       CST     57    FAYETTE
## ..     ...                 ...      ...       ...    ...        ...
## Variables not shown: EVTYPE (fctr), BGN_RANGE (dbl), BGN_AZI (fctr
##   BGN_LOCATI (fctr), END_DATE (fctr), END_TIME (fctr), COUNTY_END
##   COUNTYENDN (lgl), END_RANGE (dbl), END_AZI (fctr), END_LOCATI (f
##   LENGTH (dbl), WIDTH (dbl), F (int), MAG (dbl), FATALITIES (dbl),
##   INJURIES (dbl), PROPDMG (dbl), PROPDMGEXP (fctr), CROPDMG (dbl),
##   CROPDMGEXP (fctr), WFO (fctr), STATEOFFIC (fctr), ZONENAMES (fct
##   LATITUDE (dbl), LONGITUDE (dbl), LATITUDE_E (dbl), LONGITUDE_ (d
##   REMARKS (fctr), REFNUM (dbl)
```

Question 1: Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

First, compare the top 5 events to the total amount of injuries and fatalities that occurred.

```
#Top 5 Injury Events: ##########################################
data %>%
        group_by(EVTYPE) %>%
        select(EVTYPE, INJURIES) %>%
        summarise(
                SUM_INJURIES = sum(INJURIES, na.rm = T)
                ) %>%
        filter(SUM_INJURIES > 0) %>%
        arrange(desc(SUM_INJURIES))%.%
        head(n=5)
```

```
## Source: local data frame [5 x 2]
##
##           EVTYPE SUM_INJURIES
## 1        TORNADO        91346
## 2      TSTM WIND         6957
## 3          FLOOD         6789
## 4 EXCESSIVE HEAT         6525
## 5      LIGHTNING         5230
```

```
## Top 5 Fatality Events: ########################################
data %>%
        group_by(EVTYPE) %>%
        select(EVTYPE, FATALITIES) %>%
        summarise(
                SUM_FATALITIES = sum(FATALITIES, na.rm = T)
                ) %>%
        filter(SUM_FATALITIES > 0) %>%
        arrange(desc(SUM_FATALITIES))%.%
        head(n=5)
```

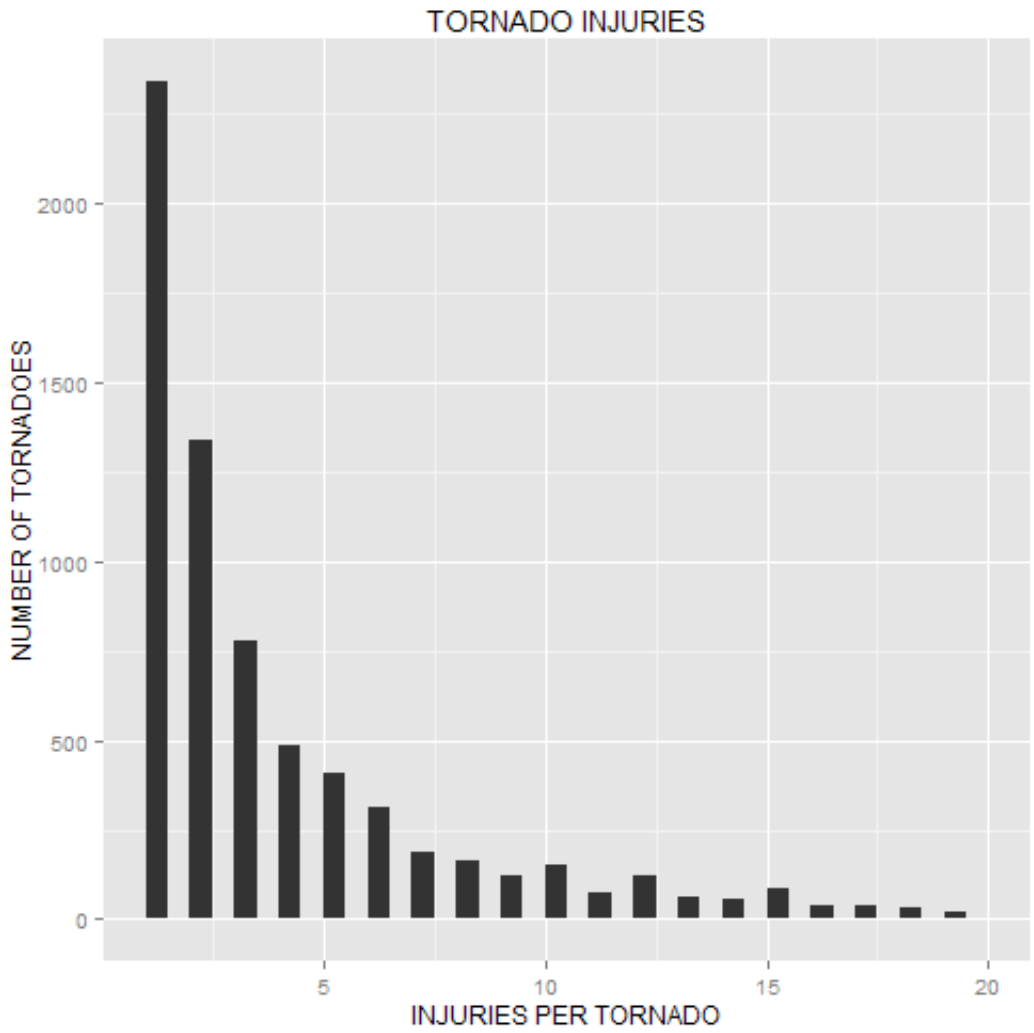
```
## Source: local data frame [5 x 2]
##
##           EVTYPE SUM_FATALITIES
## 1        TORNADO           5633
## 2 EXCESSIVE HEAT           1903
## 3    FLASH FLOOD            978
## 4           HEAT            937
## 5      LIGHTNING            816
```

Next, we will focus our attention on tornadoes. We will begin by filtering on all the cases of tornadoes that had an injury or fatality.

```
## tornado_injuries ##############################################
data_tornado =   filter(data, EVTYPE == 'TORNADO' |
                                EVTYPE == 'TORNADOS' |
                                EVTYPE == 'TORNADO F0' |
                                EVTYPE == 'TORNADO F1' |
                                EVTYPE == 'TORNADO F2' |
                                EVTYPE == 'TORNADO F3' |
                                EVTYPE == 'TORNADOES' |
                                EVTYPE == 'TORNADO DEBRIS' |
                                EVTYPE == 'COLD AIR TORNADO' |
                                EVTYPE == 'TORNADOES, TSTM WIND, HAI
                        INJURIES > 0)

tornado_group = group_by(data_tornado, INJURIES)
tornado_summarise = summarise(tornado_group, TCOUNT = n())

## tornado_fatalities ############################################
data_tornado2 =   filter(data, EVTYPE == 'TORNADO' |
                                EVTYPE == 'TORNADOS' |
                                EVTYPE == 'TORNADO F0' |
                                EVTYPE == 'TORNADO F1' |
                                EVTYPE == 'TORNADO F2' |
                                EVTYPE == 'TORNADO F3' |
                                EVTYPE == 'TORNADOES' |
                                EVTYPE == 'TORNADO DEBRIS' |
                                EVTYPE == 'COLD AIR TORNADO' |
                                EVTYPE == 'TORNADOES, TSTM WIND, HAIL
                        FATALITIES > 0)

tornado_group2 = group_by(data_tornado2, FATALITIES)
tornado_summarise2 = summarise(tornado_group2, TCOUNT = n())
```

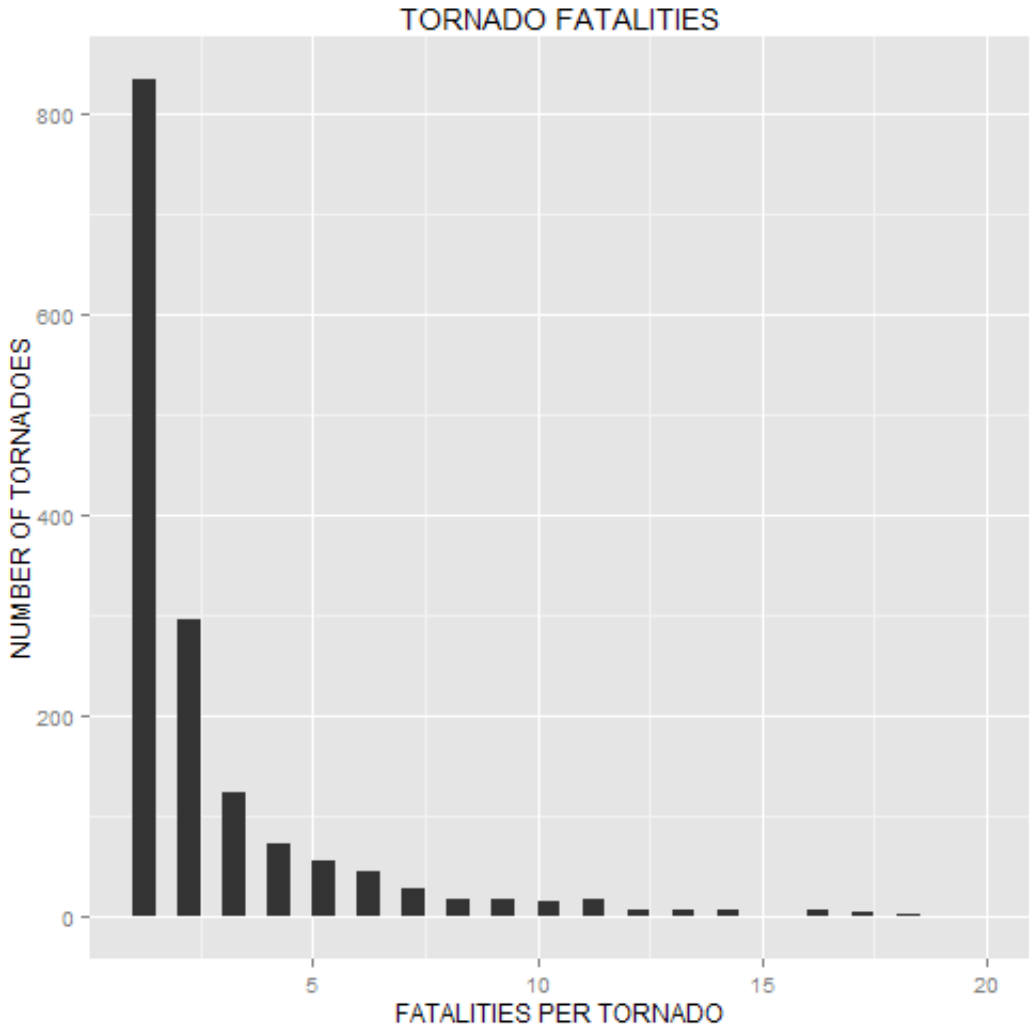
Below we will compare the number of tornadoes with the number of injuries per tornado:

```
# windows()
qplot(INJURIES,
      data=tornado_group,
      xlim=c(1,20),
      binwidth = 0.5,
      main = "TORNADO INJURIES",
      xlab="INJURIES PER TORNADO",
      ylab="NUMBER OF TORNADOES")
```



Next, we will compare the number of tornadoes with the count of fatalities per tornado:

```
# windows()
qplot(FATALITIES,
      data=tornado_group2,
      xlim=c(1,20),
      binwidth = 0.5,
      main = "TORNADO FATALITIES",
      xlab="FATALITIES PER TORNADO",
      ylab="NUMBER OF TORNADOES")
```



Question 2: Across the United States, which types of events have the greatest economic consequences?

First, filter your data to only include the damage in the millions (as indicated in the PROPDMGEXP variable). Next, we will display the top 5 economically damaging events since 1950.

```
#Most expensive events
dataM = filter(data, PROPDMGEXP == 'M')

dataM %>%
        group_by(EVTYPE) %>%
        select(EVTYPE, PROPDMG) %>%
        summarise(
                MILLIONS = sum(PROPDMG, na.rm = T)
        ) %>%
        filter(MILLIONS > 0) %>%
        arrange(desc(MILLIONS)) %.%
        head(n=5)
```

```
## Source: local data frame [5 x 2]
##
##         EVTYPE MILLIONS
## 1      TORNADO    48462
## 2        FLOOD    21279
## 3 FLASH FLOOD    13735
## 4         HAIL    13252
## 5    HURRICANE     6159
```

Similar to above, we will focus our attention to only the damage done by tornadoes. We will show the overall number of tornadoes with the amount of damage done per tornado.

We begin by filtering on every tornado with over a million dollars of damage. (Totals are not adjusted for inflation):

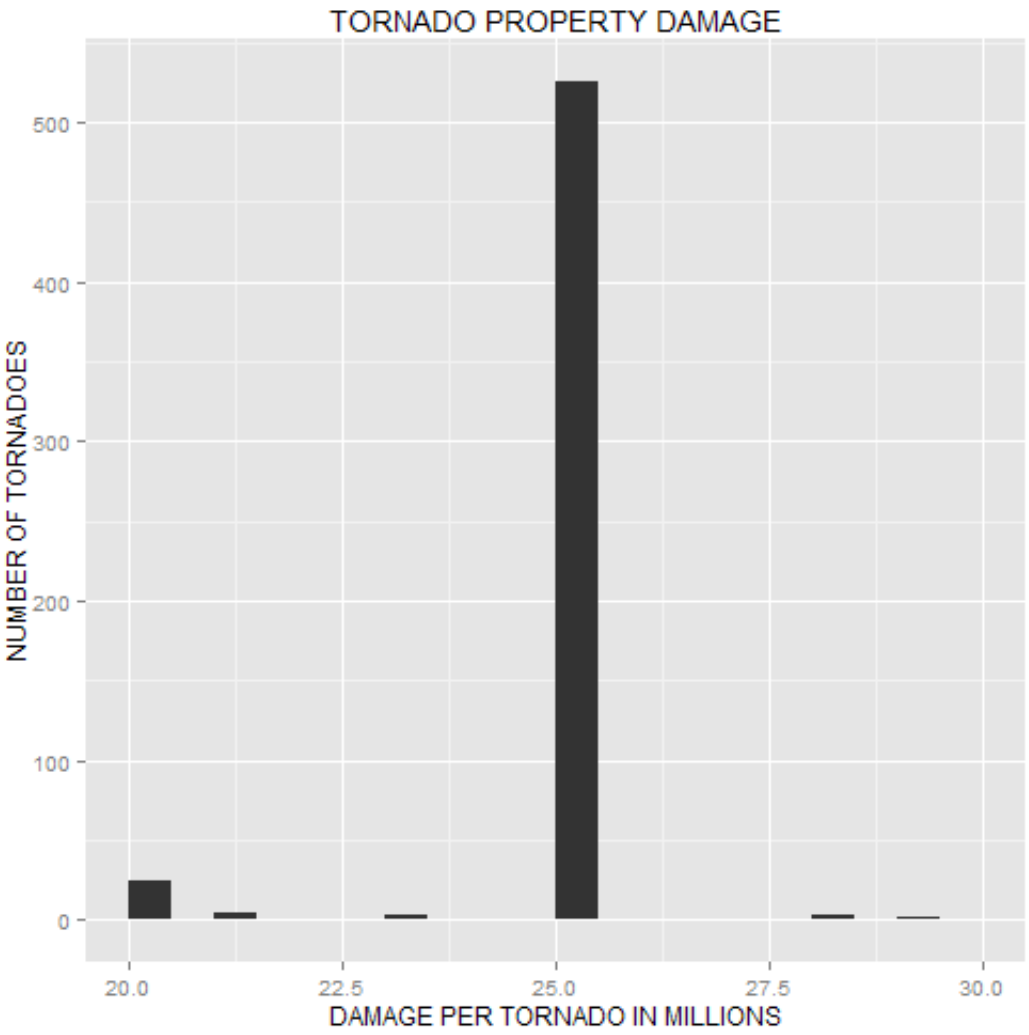
```
data_tornado3 =   filter(data, EVTYPE == 'TORNADO' |
                                 EVTYPE == 'TORNADOS' |
                                 EVTYPE == 'TORNADO F0' |
                                 EVTYPE == 'TORNADO F1' |
                                 EVTYPE == 'TORNADO F2' |
                                 EVTYPE == 'TORNADO F3' |
                                 EVTYPE == 'TORNADOES' |
                                 EVTYPE == 'TORNADO DEBRIS' |
                                 EVTYPE == 'COLD AIR TORNADO' |
                                 EVTYPE == 'TORNADOES, TSTM WIND, HAI
                          PROPDMGEXP == 'M')

tornado_group3 = group_by(data_tornado3, PROPDMG)
tornado_summarise3 = summarise(tornado_group3, TCOUNT = n())
```

Below is the code to view overall plot with tornado damage between 1 and 30 million:

```
# windows()
overall = qplot(PROPDMG,
               data=tornado_group3,
               xlim=c(1,30),
               binwidth = 0.5,
               main = "TORNADOE DAMAGE",
               xlab="DAMAGE PER TORNADO IN MILLIONS",
               ylab="NUMBER OF TORNADOES")
```

Plot for tornado damage between 20 and 30 million:

```
# windows()
qplot(PROPDMG,
      data=tornado_group3,
      xlim=c(20,30),
      binwidth = 0.5,
      main = "TORNADO PROPERTY DAMAGE",
      xlab="DAMAGE PER TORNADO IN MILLIONS",
      ylab="NUMBER OF TORNADOES")
```



Below prints your markdown file to html:

```
# library(knitr)
# knit2html("PA2_template.Rmd")
```