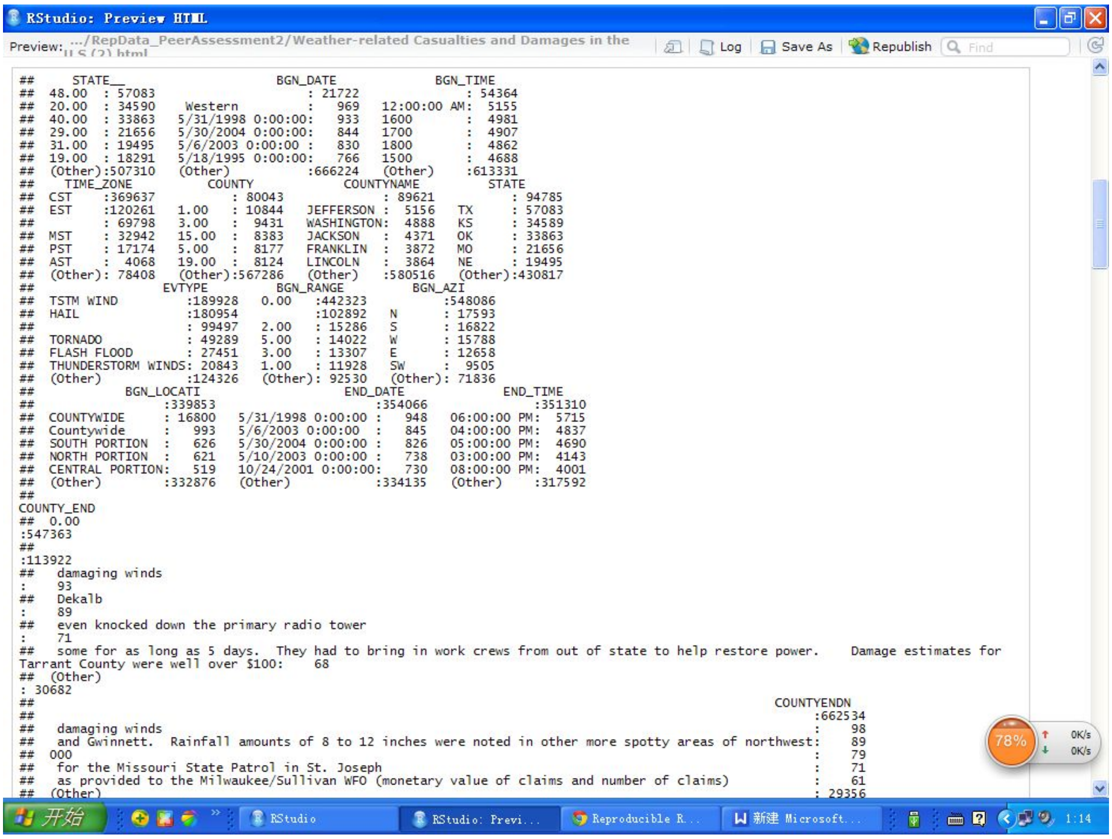
~~~
# Weather-related Casualties and Damages in the U.S.
###==========================================================================
## Reading data
```{r,results='hide'}
setwd("H:\\许晟\\Rprogramming\\Github\\RepData_PeerAssessment2")
library(stringr)
library(dplyr)
library(reshape2)
library(ggplot2)
if (!file.exists("stormData.csv.bz2")){
    setInternet2(use = TRUE)
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
"stormData.csv.bz2")
}
rawStorm = read.csv("StormData.csv.bz2",header = TRUE, sep = ",")
```
### let's take a peek at the data:
```{r}
summary(rawStorm)
```
~~~

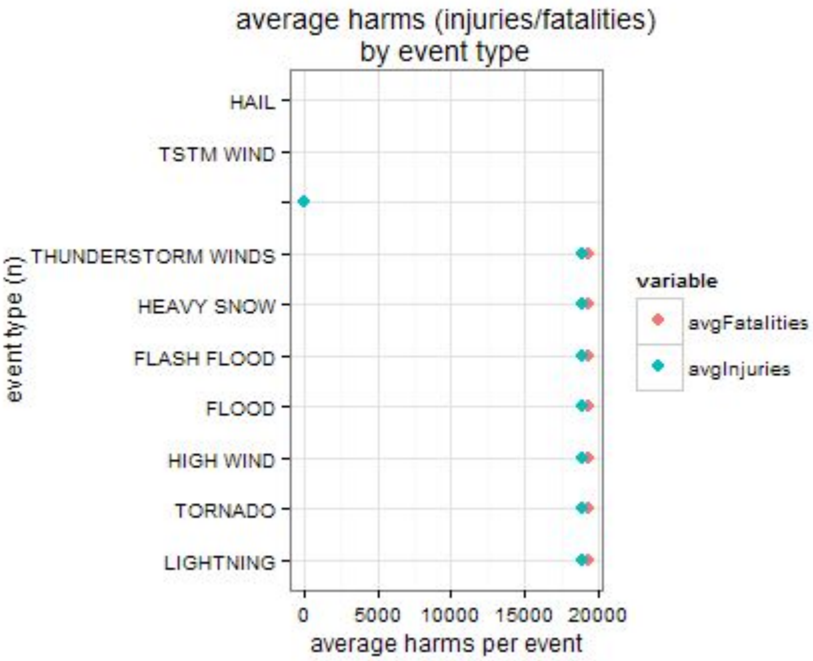


```
##     STATE__                BGN_DATE             BGN_TIME
##  48.00   : 57083                    : 21722              : 54364
##  20.00   : 34590    Western         :   969   12:00:00 AM:  5155
##  40.00   : 33863   5/31/1998 0:00:00:   933   1600       :  4981
##  29.00   : 21656   5/30/2004 0:00:00:   844   1700       :  4907
##  31.00   : 19495   5/6/2003 0:00:00 :   830   1800       :  4862
##  19.00   : 18291   5/18/1995 0:00:00:   766   1500       :  4688
##  (Other):507310   (Other)          :666224   (Other)    :613331
##    TIME_ZONE         COUNTY             COUNTYNAME          STATE
##  CST     :369637            : 80043              : 89621             : 94785
##  EST     :120261   1.00    : 10844   JEFFERSON :  5156   TX      : 57083
##          : 69798   3.00    :  9431   WASHINGTON:  4888   KS      : 34589
##  MST     : 32942   15.00   :  8383   JACKSON   :  4371   OK      : 33863
##  PST     : 17174   5.00    :  8177   FRANKLIN  :  3872   MO      : 21656
##  AST     :  4068   19.00   :  8124   LINCOLN   :  3864   NE      : 19495
##  (Other): 78408   (Other):567286   (Other)   :580516   (Other):430817
##                EVTYPE          BGN_RANGE          BGN_AZI
##  TSTM WIND          :189928   0.00    :442323           :548086
##  HAIL               :180954           :102892   N       : 17593
##                     : 99497   2.00    : 15286   S       : 16822
##  TORNADO            : 49289   5.00    : 14022   W       : 15788
##  FLASH FLOOD        : 27451   3.00    : 13307   E       : 12658
##  THUNDERSTORM WINDS: 20843   1.00    : 11928   SW      :  9505
##  (Other)            :124326   (Other): 92530   (Other): 71836
##            BGN_LOCATI                    END_DATE               END_TIME
##                 :339853                       :354066                 :351310
##  COUNTYWIDE     : 16800   5/31/1998 0:00:00 :   948   06:00:00 PM:  5715
##  Countywide     :   993   5/6/2003 0:00:00  :   845   04:00:00 PM:  4837
##  SOUTH PORTION  :   626   5/30/2004 0:00:00 :   826   05:00:00 PM:  4690
##  NORTH PORTION  :   621   5/10/2003 0:00:00 :   738   03:00:00 PM:  4143
##  CENTRAL PORTION:   519   10/24/2001 0:00:00:   730   08:00:00 PM:  4001
##  (Other)        :332876   (Other)             :334135   (Other)    :317592
##
COUNTY_END
##  0.00
:547363
##
:113922
##   damaging winds
:    93
##   Dekalb
:    89
##   even knocked down the primary radio tower
:    71
##   some for as long as 5 days.  They had to bring in work crews from out of state to help restore power.    Damage estimates for
Tarrant County were well over $100:    68
##  (Other)
: 30682
##                                                                                                          COUNTYENDN
##                                                                                                         :662534
##   damaging winds                                                                                        :     98
##   and Gwinnett.  Rainfall amounts of 8 to 12 inches were noted in other more spotty areas of northwest:     89
##  000                                                                                                    :     79
##   for the Missouri State Patrol in St. Joseph                                                           :     71
##   as provided to the Milwaukee/Sullivan WFO (monetary value of claims and number of claims)             :     61
##  (Other)                                                                                                : 29356
```

~~~
```{r}
# convert crop and property amounts into total dollars based on EXP fields
# first change all char values to lower case
rawStorm$PROPDMGEXP = tolower(rawStorm$PROPDMGEXP)
~~~

~~~
rawStorm$PROPDMGEXP = gsub('-', 0, rawStorm$PROPDMGEXP)
rawStorm$PROPDMGEXP = gsub('h', 2, rawStorm$PROPDMGEXP)
rawStorm$PROPDMGEXP = gsub('k', 3, rawStorm$PROPDMGEXP)
rawStorm$PROPDMGEXP = gsub('m', 6, rawStorm$PROPDMGEXP)
rawStorm$PROPDMGEXP = gsub('b', 9, rawStorm$PROPDMGEXP)

rawStorm$PROPDMGEXP = as.numeric(rawStorm$PROPDMGEXP)
```
```{r}
rawStorm$PROPDMGEXP[is.na(rawStorm$PROPDMGEXP) == TRUE] = 0

# calculate new property damage
rawStorm$propDamage = ( as.numeric(rawStorm$PROPDMG) * ( 10 ^
rawStorm$PROPDMGEXP ) )
```


## Results
### Which events are most harmful to pop. health?
```{r}
# create data frame of total and average population harms by event type
harmByEvent = rawStorm %.%
  group_by(EVTYPE) %.%
  summarize(cntEvents = n(), totalFatalities = sum(FATALITIES), totalInjuries =
sum(INJURIES),
              avgFatalities = totalFatalities / cntEvents, avgInjuries = totalInjuries / cntEvents )

# order by occurence of event
harmByEvent = arrange(harmByEvent, -cntEvents)

# convert event types to lowercase format
harmByEvent$EVTYPE = paste( tolower(harmByEvent$EVTYPE), " (",
harmByEvent$cntEvents, ")" )

# plot top ten
# create plot data frame and melt (for easier plotting)
dfPlot = melt(harmByEvent[1:10,][c('EVTYPE', 'avgFatalities', 'avgInjuries')], id.vars='EVTYPE')

dfPlot = arrange(dfPlot, variable, -value)

# plot top ten events
ggplot(dfPlot, aes(x = EVTYPE, y = value, color = variable)) +
  geom_point(stat = 'identity', size = 3) +
  scale_x_discrete(limits = dfPlot$EVTYPE[1:10]) +
~~~

```
  coord_flip() +
  xlab('event type (n)') + ylab('average harms per event') +
  theme_bw() +
  ggtitle('average harms (injuries/fatalities)\nby event type')
```



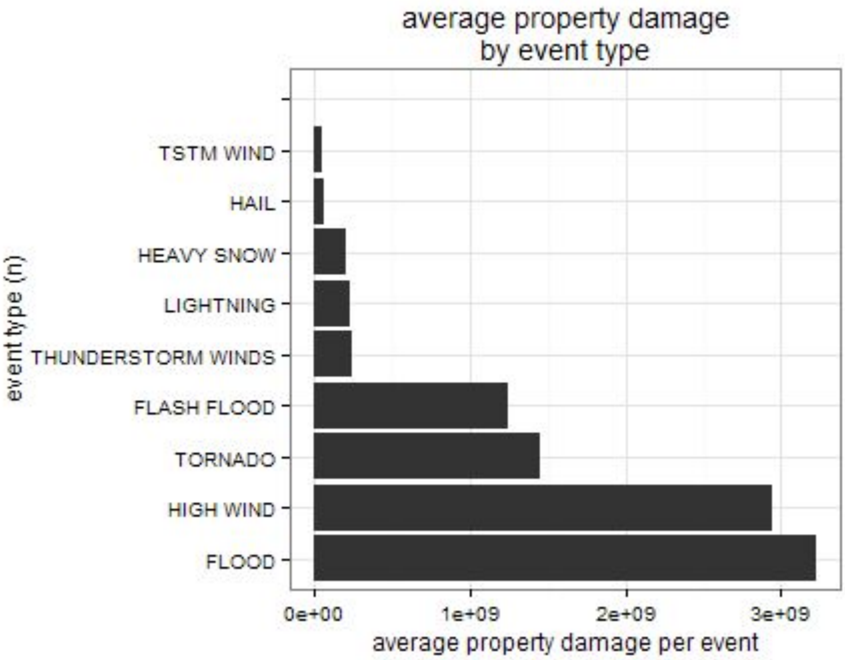
Of the most prevalent weather event types, tornadoes not only have the highest average deaths per event, but also the highest average injuries per event - and by a wide margin.

### Which events are most harmful to economy?

```{r}
# create data frame of property damage by event type
costByEvent = rawStorm %.%
  group_by(EVTYPE) %.%
  summarize(cntEvents = n(), totalPropDamage = sum(propDamage), avgPropDamage = totalPropDamage / cntEvents )

# order by occurence of event
costByEvent = arrange(costByEvent, -cntEvents)

# convert event types to lowercase format
costByEvent$EVTYPE = paste0( tolower(costByEvent$EVTYPE), " (", costByEvent$cntEvents, ")" )

# plot top ten
# create plot data frame
dfCostPlot = costByEvent[1:10, ]
dfCostPlot = arrange(dfCostPlot, -avgPropDamage)
```

```
# plot top ten events
ggplot(dfCostPlot, aes(x = EVTYPE, y = avgPropDamage) ) +
   geom_bar(stat = 'identity') +
   scale_x_discrete(limits = dfCostPlot$EVTYPE[1:10]) +
   coord_flip() +
   xlab('event type (n)') + ylab('average property damage per event') +
   theme_bw() + ggtitle('average property damage\nby event type')
```

Of the most prevalent weather event types, floods have the highest average property damage per event. Tornadoes come in a distance second, followed by flash floods


average property damage
by event type
TSTM WIND
HAIL
HEAVY SNOW
LIGHTNING
THUNDERSTORM WINDS
FLASH FLOOD
TORNADO
HIGH WIND
FLOOD
event type (n)
0e+00
1e+09
2e+09
3e+09
average property damage per event