Analysis of NOAA Storm Database

Nikhila Arkalgud

February 22, 2015

Synopsis:

We will use the NOAA Storm Database to exmaine the effects these disaster events have on lives. We will specifically try to analyze the effects on population health and economic consequences. From the results obtained we found that tornadoes had the highest effect on the population health, it had both the highest number of fatalities and injuries. The impact of tornadoes on population health were almost 2 to 3 times higher than the second highest event. Floods have resulted in the highest impact on the economy. Although property damages due to floods are more than twice the second highest event(hurricane), it has not contributed to a high number of fatalities or injuries compared to tornadoes. The report below provides the steps taken to process this dataset and then provides the results.

Data Processing:

Download the Storm Data from https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) Unzip the file and read the csv file into R.

```
stdata <- read.csv("repdata-data-StormData.csv")
```

We will check the dimentionality of this dataset

```
dim(stdata)
```

```
## [1] 902297     37
```

```
head(stdata)
```

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

We will find the total fatalities and injuries resulted from each disaster event type which is represented in this dataset under EVTYPE

```
# aggregate the total fatalities for each evtype
ph_fatality <- aggregate(stdata$FATALITIES, by=list(stdata$EVTYPE), FUN=sum)
names(ph_fatality) <- make.names(c("evtype","fatality"))

# aggregate the total injuries for each evtype
ph_injury <- aggregate(stdata$INJURIES, by=list(stdata$EVTYPE), FUN=sum)
names(ph_injury) <- make.names(c("evtype","injury"))
```

We will now process these subsets to find the cummulative sum and cummulative fraction for both fatalities and injuries. This will help us isolate the events that cummulatively contribute to 95% of the total fatality and the events that cummulatively contribute to 95% of the total injuries.

```
# we will calculate the cummulative sum and find the cummulative fraction for each event
# this will help us identify the events that
# cummulatively contribute to 95% of the total fatality

# sort the ph_fatality in desending order
phfo <- ph_fatality[order(-ph_fatality[,2]),]

#append a col cufrac to phfo
phfo$cufrac <- cumsum(phfo$fatality)/sum(phfo$fatality)
```

```
# we will calculate the cummulative sum and find the cummulative fraction for each event
# this will help us identify the events that
# cummulatively contribute to 95% of the total injuries

# sort the ph_injury in desending order
phio <- ph_injury[order(-ph_injury[,2]),]

#append a col cufrac to phio
phio$cufrac <- cumsum(phio$injury)/sum(phio$injury)
```

We will also extract the total property damage resulted from each disaster event type. Economic disaster for each event type can be measured using PROPDMG and PROPDMGEXP columns. The PROPDMGEXP column procides the magnitude, with ‘K’ representing thousand dollars, ‘M’ for millon dollars and ‘B’ for billion dollars. We will use these two rows to obtain the raw dollar amount of economic damage for each event type.

```
#append a column to replace 'K', 'M' and 'B' provided under PROPDMGEXP
#with 1000,1000000,1000000000, empty rows are replaced by 1
#then multiply this value with the value provided under PROPDMG

# start by making all rows 1
stdata$economicDamage <- 1

#extract the K, M and B logical indices
kindices <- stdata$PROPDMGEXP == 'K'
mindices <- stdata$PROPDMGEXP == 'M'
bindices <- stdata$PROPDMGEXP == 'B'

# assign 1000, 1000000 and 1000000000 to replace the K, M and B
stdata$economicDamage[kindices] <- 1000
stdata$economicDamage[mindices] <- 1000000
stdata$economicDamage[bindices] <- 1000000000

#multiply with PROPDMG column to get the actual economic damage for this event
stdata$economicDamage <- stdata$PROPDMG * stdata$economicDamage

#aggregate the economic damage by each event type
ed <- aggregate(stdata$economicDamage, by=list(stdata$EVTYPE), FUN=sum)
names(ed) <- make.names(c("evtype","economicDamage"))

# sort the data by economicDamage in desending order
ed_ordered <- ed[order(-ed[,2]),]
```

```
# we will calculate the cummulative sum and find the cummulative fraction for each event
# this will help us identify the events that
# cummulatively contribute to 95% of the total economic damage

#append a col cufrac to ed_ordered
ed_ordered$cufrac <- cumsum(ed_ordered$economicDamage)/sum(ed_ordered$economicDamage)
```

Results:

We will examine two keys effects that result due to the occurance of these events:

1. How does this affect the population health?
2. What are the economic consequences?

For each of these cases, examining the events that contribute to the 95% of the cummulative damage is a good measure to identify the most damaging events. We will display these events and their respective values. We will then plot the top 10 events for each of these categories.

1. How does this affect the population health?

We will analyze the number of fatalities and injuries that have occurred due to these events. This will help us analyze the effect these events have on the population health.

We will look at the events that contribute to 95% of the cummulative fatalites that results from these events.

```
# display the events that contribute to 90% of the total fatality
top95_f <- phfo[phfo$cufrac < 0.95,]

#total number of events
nrow(phfo)
```

```
## [1] 985
```

```
#total number of events contributing to the 95% of total fatality
nrow(top95_f)
```

```
## [1] 32
```

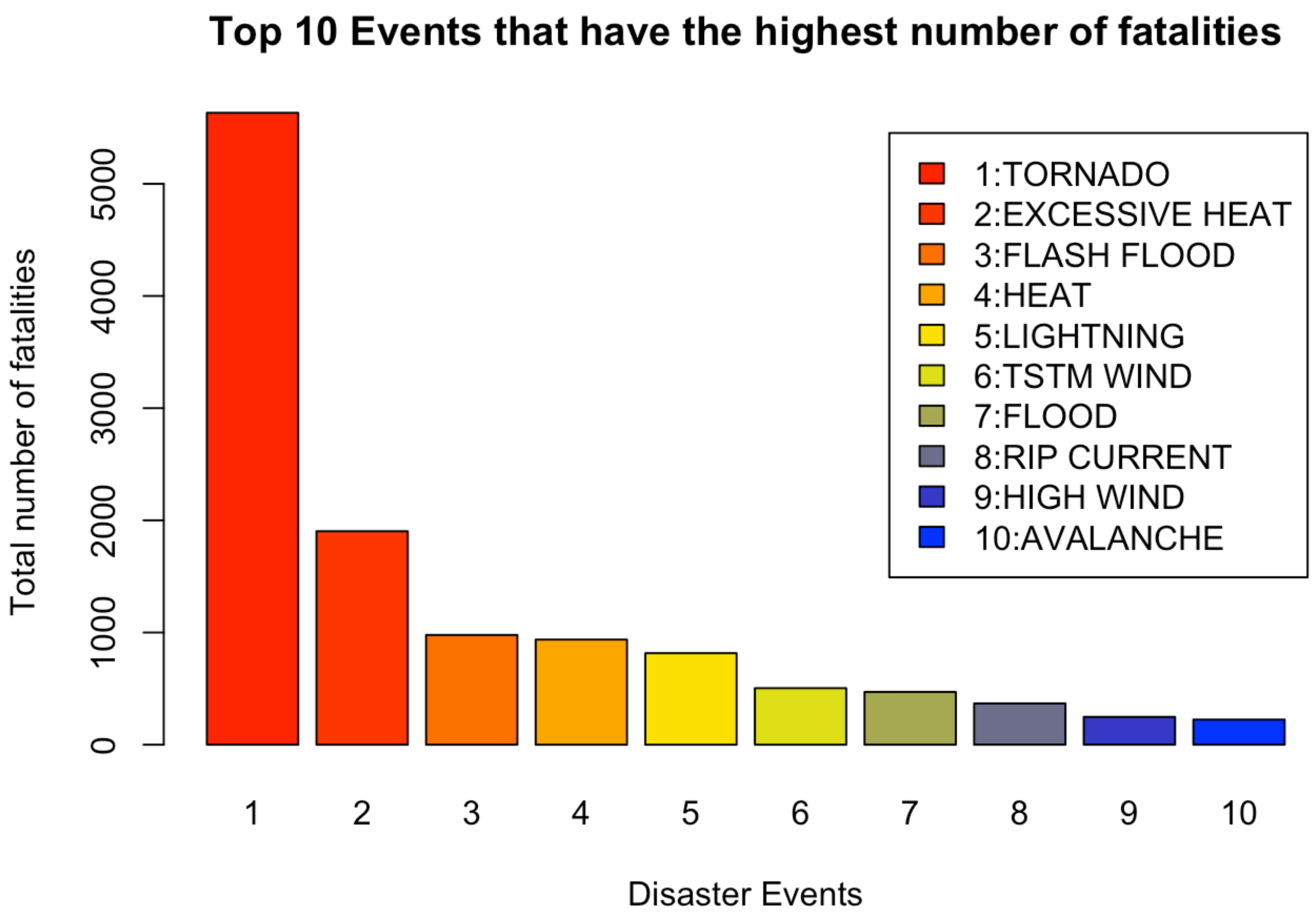
```
#list of these events contributing to the 95% of total fatality
top95_f
```

```
##                      evtype fatality    cufrac
## 834                 TORNADO     5633 0.3719379
## 130          EXCESSIVE HEAT     1903 0.4975900
## 153             FLASH FLOOD      978 0.5621657
## 275                    HEAT      937 0.6240343
## 464               LIGHTNING      816 0.6779135
## 856               TSTM WIND      504 0.7111918
## 170                   FLOOD      470 0.7422252
## 585             RIP CURRENT      368 0.7665236
## 359               HIGH WIND      248 0.7828986
## 19                AVALANCHE      224 0.7976890
## 972            WINTER STORM      206 0.8112909
## 586            RIP CURRENTS      204 0.8247606
## 278               HEAT WAVE      172 0.8361175
## 140            EXTREME COLD      160 0.8466821
## 760       THUNDERSTORM WIND      133 0.8554638
## 310              HEAVY SNOW      127 0.8638495
## 141 EXTREME COLD/WIND CHILL      125 0.8721030
## 676             STRONG WIND      103 0.8789039
## 30                 BLIZZARD      101 0.8855728
## 350               HIGH SURF      101 0.8922417
## 290              HEAVY RAIN       98 0.8987124
## 142            EXTREME HEAT       96 0.9050512
## 79          COLD/WIND CHILL       95 0.9113239
## 427               ICE STORM       89 0.9172004
## 957                WILDFIRE       75 0.9221525
## 411       HURRICANE/TYPHOON       64 0.9263783
## 786      THUNDERSTORM WINDS       64 0.9306042
## 188                     FOG       62 0.9346979
## 402               HURRICANE       61 0.9387257
## 848          TROPICAL STORM       58 0.9425553
## 342    HEAVY SURF/HIGH SURF       42 0.9453285
## 442               LANDSLIDE       38 0.9478376
```

There are 32 events out of the total 985 events that contribute to 95% of the total number of fatalities.

The below bar chart shows top 10 events that contribute to the highest fatality

```
events <- top95_f$evtype[1:10]
labels <- paste(seq(1:10),events,sep=":")
ramp <- colorRamp(c("red","yellow","blue"))
barplot(top95_f$fatality[1:10], main = "Top 10 Events that have the highest number of fatal
ities", xlab = "Disaster Events", ylab = "Total number of fatalities", names.arg = seq(1:10
), legend = labels, col = rgb(ramp(seq(0,1,length=10)),max=255))
```



From the barchart and also the table above we see that tornado has the highest number of fatalities.

```
top95_f[1,1:2]
```

```
##      evtype fatality
## 834 TORNADO     5633
```

Now we will examine the injuries that have resulted due to these disaster events.

We will look at the events that contribute to 95% of the cummulative injuries that results from these events.

```
# display the events that contribute to 95% of the total injuries
top95_i <- phio[phio$cufrac < 0.95,]

#total number of events
nrow(phio)
```

```
## [1] 985
```

```
#total number of events contributing to the 95% of total injuries
nrow(top95_i)
```

```
## [1] 17
```

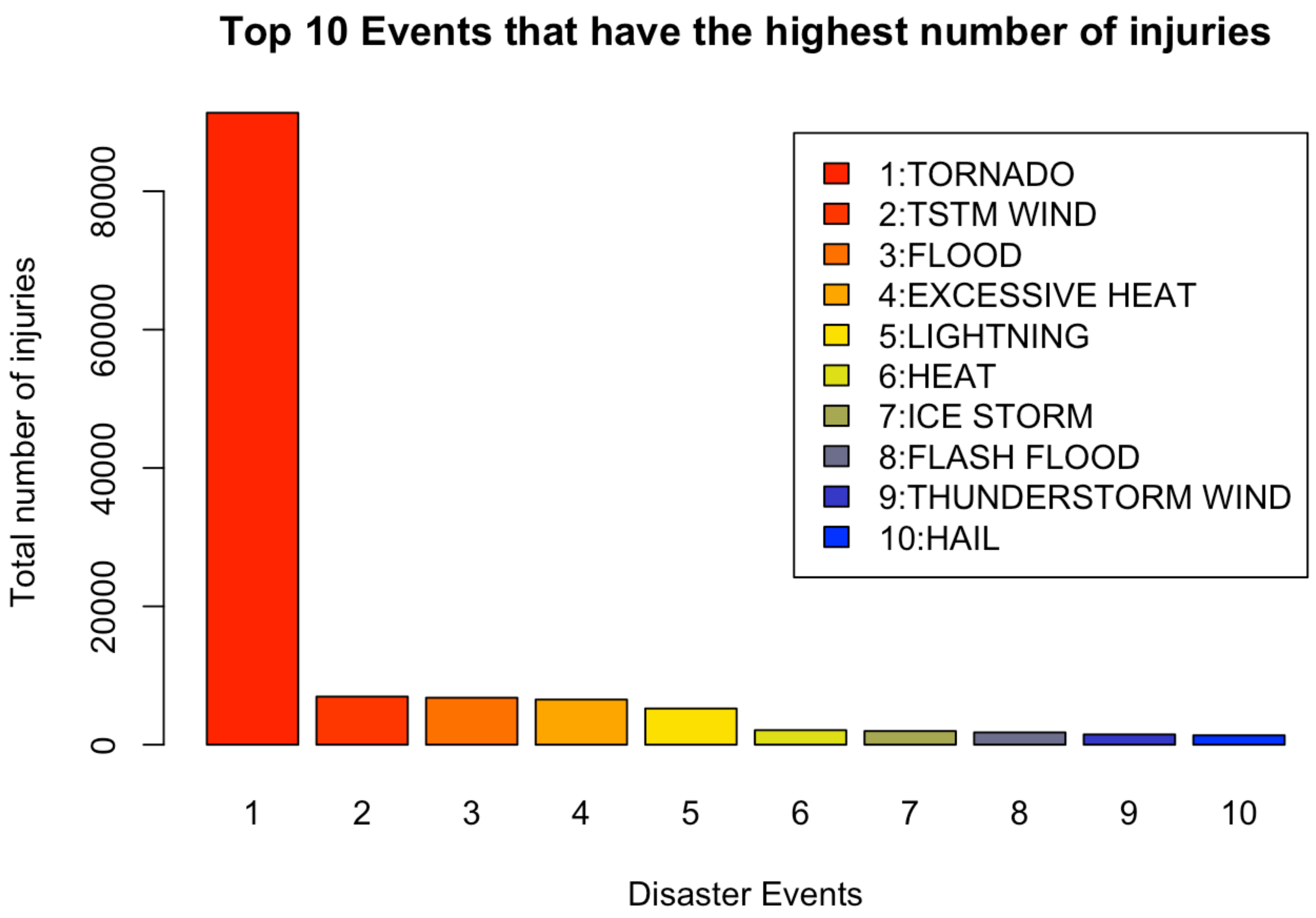
```
#list of these events contributing to the 95% of total injuries
top95_i
```

```
##                  evtype injury    cufrac
## 834             TORNADO  91346 0.6500199
## 856           TSTM WIND   6957 0.6995261
## 170               FLOOD   6789 0.7478367
## 130      EXCESSIVE HEAT   6525 0.7942688
## 464           LIGHTNING   5230 0.8314855
## 275                HEAT   2100 0.8464292
## 427           ICE STORM   1975 0.8604833
## 153         FLASH FLOOD   1777 0.8731285
## 760   THUNDERSTORM WIND   1488 0.8837171
## 244                HAIL   1361 0.8934020
## 972        WINTER STORM   1321 0.9028023
## 411   HURRICANE/TYPHOON   1275 0.9118752
## 359           HIGH WIND   1137 0.9199661
## 310          HEAVY SNOW   1021 0.9272316
## 957            WILDFIRE    911 0.9337143
## 786  THUNDERSTORM WINDS    908 0.9401756
## 30             BLIZZARD    805 0.9459040
```

There are 17 events out of the total 985 events that contribute to 95% of the total number of injuries.

The below bar chart shows top 10 events that contribute to the highest number of injuries.

```
events <- top95_i$evtype[1:10]
labels <- paste(seq(1:10),events,sep=":")
ramp <- colorRamp(c("red","yellow","blue"))
barplot(top95_i$injury[1:10], main = "Top 10 Events that have the highest number of injurie
s", xlab = "Disaster Events", ylab = "Total number of injuries", names.arg = seq(1:10), leg
end = labels, col = rgb(ramp(seq(0,1,length=10)),max=255))
```



From both the barchart and the table above we see that tornados also have also resulted in the highest number of injuries.

```
top95_i[1,1:2]
```

```
##      evtype injury
## 834 TORNADO  91346
```

We can see that tornadoes have resulted in the highest number of fatalities and injuries, the numbers are almost 2-3 times higher than the second highest events leading it to have the highest impact on the population health.

2. What are the economic consequences?

We will now examine the events that have the greatest economic consequences. We will look at the property damage associated to each of these events.

We will look at the events that contribute to 95% of the cummulative property damage that results from these events.

```
# display the events that contribute to 95% of the total economic damage
top95_e <- ed_ordered[ed_ordered$cufrac < 0.95,]

#total number of events
nrow(ed_ordered)
```

```
## [1] 985
```

```
#total number of events contributing to the 95% of total economic damage
nrow(top95_e)
```

```
## [1] 16
```

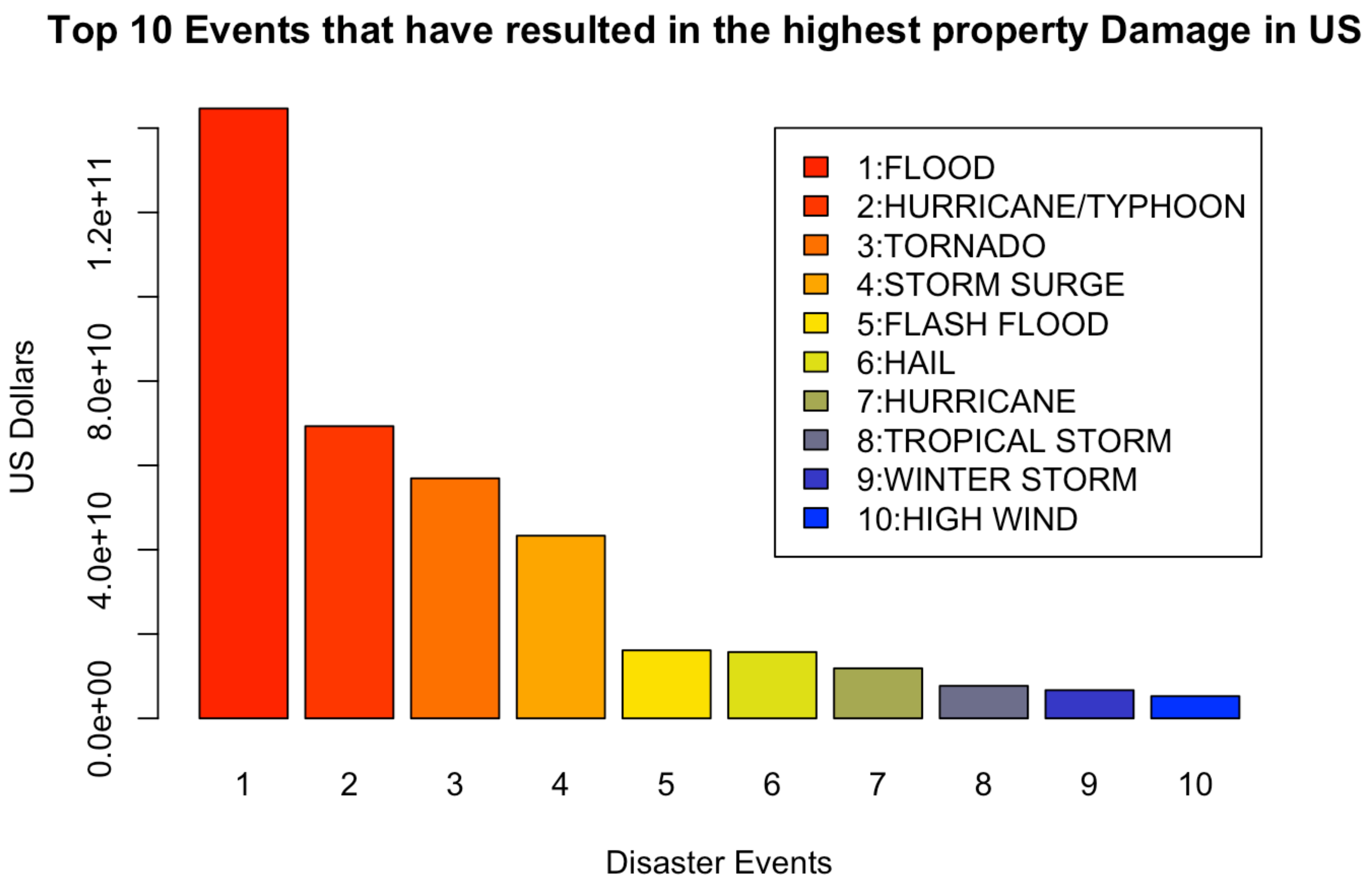
```
#list of these events contributing to the 95% of total economic damage
top95_e
```

```
##                evtype economicDamage    cufrac
## 170             FLOOD   144657709807 0.3385550
## 411 HURRICANE/TYPHOON    69305840000 0.5007575
## 834           TORNADO    56925660790 0.6339856
## 670       STORM SURGE    43323536000 0.7353794
## 153       FLASH FLOOD    16140812067 0.7731552
## 244              HAIL    15727367053 0.8099633
## 402         HURRICANE    11868319010 0.8377398
## 848    TROPICAL STORM     7703890550 0.8557699
## 972      WINTER STORM     6688497251 0.8714235
## 359         HIGH WIND     5270046295 0.8837575
## 590       RIVER FLOOD     5118945500 0.8957378
## 957          WILDFIRE     4765114000 0.9068900
## 671  STORM SURGE/TIDE     4641188000 0.9177522
## 856         TSTM WIND     4484928495 0.9282487
## 427         ICE STORM     3944927860 0.9374813
## 760 THUNDERSTORM WIND     3483121284 0.9456332
```

There are 16 events out of the total 985 events that contribute to 95% of the total economic damage.

The below bar chart shows top 10 events that contribute to the highest property damage

```
events <- top95_e$evtype[1:10]
labels <- paste(seq(1:10),events,sep=":")
ramp <- colorRamp(c("red","yellow","blue"))
barplot(top95_e$economicDamage[1:10], main = "Top 10 Events that have resulted in the highe
st property Damage in USD", xlab = "Disaster Events", ylab = "US Dollars", names.arg = seq(
1:10), legend = labels, col = rgb(ramp(seq(0,1,length=10)),max=255))
```



From the barchart and the table above we see that Floods have resulted in the maximum property damage leading to the highest economic consequence.

```
top95_e[1,1:2]
```

```
##     evtype economicDamage
## 170  FLOOD   144657709807
```

Floods have resulted in the highest property damage and resulted in high economic consequences, the US Dollars reported are more than twice the second highest event(hurricane).

Interestingly, although floods have high economic consequences, they seem to have less effect on population health. Tornados in contract are the highest in population health and have significant effects on economy as well, tornadoes are the third highest contributing event to property damages.