Impact of the severe weather events on human population and property based on the U.S. National Oceanic and Atmospheric Administration's storm database.

Synopsis

Goal of this report is to idenify severe weather events of the highest impact on both US population and US economy based on the information retrieved from U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. According to official documentation NOAA storm database sources include

> county, state and federal emergency management officials,local law enforcement officials, skywarn spotters, NWS damage surveys, newspaper clipping services, the insurance industry and the general public.

Database version used for this report contain entries for years 1950-2011.

Data processing

R libraries

```
library(data.table)
library(ggplot2)
library(gridExtra)
```

```
## Loading required package: grid
```

```
library(reshape)
```

```
## Loading required package: plyr
##
## Attaching package: 'reshape'
##
## The following objects are masked from 'package:plyr':
##
##     rename, round_any
```

Loading dataset and initital transofrmations

Dataset has been decompressed and loaded into `data.frame` and after that convered into `data.table`. Resulting `data.table` contained 436108 rows and 38.

```
setwd
("~/Workspace/coursera/reproducible_research/RepData_PeerAssessment2
dt <- data.table(read.csv(bzfile("StormData.csv.bz2")))
str(dt)
```

```
## Classes 'data.table' and 'data.frame':   902297 obs. of  37
variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : Factor w/ 16335 levels "10/10/1954
0:00:00",..: 6523 6523 4213 11116 1426 1426 1462 2873 3980
3980 ...
##  $ BGN_TIME  : Factor w/ 3608 levels "000","0000","00:00:00
AM",..: 212 257 2645 1563 2524 3126 122 1563 3126 3126 ...
##  $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7
7 7 7 7 7 7 7 7 7 ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC
BRIDGE TO PRESQUE ISLE LT MI",..: 13513 1873 4598 10592 4372
10094 1973 23873 24418 4598 ...
##  $ STATE     : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2
2 2 2 2 2 2 2 ...
##  $ EVTYPE    : Factor w/ 985 levels "?","ABNORMALLY
DRY",..: 830 830 830 830 830 830 830 830 830 830 ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : Factor w/ 35 levels "","E","Eas","EE",..: 1
1 1 1 1 1 1 1 1 1 ...
##  $ BGN_LOCATI: Factor w/ 54429 levels "","?","(01R)AFB GNRY
RNG AL",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_DATE  : Factor w/ 6663 levels "","10/10/1993
0:00:00",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ END_TIME  : Factor w/ 3647 levels "","?","0000",..: 1 1
1 1 1 1 1 1 1 1 ...
##  $ 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   : Factor w/ 24 levels "","E","ENE","ESE",..: 1
1 1 1 1 1 1 1 1 1 ...
##  $ END_LOCATI: Factor w/ 34506 levels "","(0E4)PAYSON
ARPT",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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: Factor w/ 19 levels "","-","?","+",..: 17 17
17 17 17 17 17 17 17 17 ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1
1 1 1 1 1 1 1 ...
##  $ WFO       : Factor w/ 542 levels "","2","43","9V9",..: 1
1 1 1 1 1 1 1 1 1 ...
##  $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA,
Central",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ ZONENAMES : Factor w/ 25112 levels
"","
"|__truncated__,..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ 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   : Factor w/ 436781 levels "","  ","  ","
",..: 1 1 1 1 1 1 1 1 1 1 ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
##  - attr(*, ".internal.selfref")=<externalptr>
```

To simplify further aggregations dataset has been extended with following column:

- COUNT column with value 1
- YEAR year of the event extract from event BGN_DATE

```
dt$COUNT <- 1
dt$YEAR <- year(strptime(dt$BGN_DATE, "%m/%d/%Y %H:%M:%S"))
```

Processing PROPDMGEXP and CROPDMGEXP.

Based on the provided documentation and a relevant discussion on the Coursera forum I decided to interpret provided data as an expotent for PROPDMG and CROPDMG respectively.

```
levels(dt$PROPDMGEXP)
```

```
##  [1] ""  "-" "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8"
"B" "h" "H" "K"
## [18] "m" "M"
```

```
levels(dt$CROPDMGEXP)
```

```
## [1] ""  "?" "0" "2" "B" "k" "K" "m" "M"
```

```
fixexp <- function(x) {
    return(as.numeric(ifelse(x == "", 0, ifelse(x == "K",
"3", ifelse(x == "M",
        "6", ifelse(x == "B", "9", ifelse(x %in% c("", "-",
"?", "+", "0", "H"),
            NA, as.character(x)))))))
}

dt$PROPDMGEXP <- fixexp(toupper(dt$PROPDMGEXP))
dt$CROPDMGEXP <- fixexp(toupper(dt$CROPDMGEXP))

dt$PROPDMG.TOTAL <- (dt$PROPDMG * 10^dt$PROPDMGEXP)/1e+06
dt$CROPDMG.TOTAL <- (dt$CROPDMG * 10^dt$CROPDMGEXP)/1e+06
```

Discrepancy between events description and damage valu

There is large number of entries where result of above computation does not much value which can be found in the recodrd description. For the purpose of this I dedcided to leave these intacted with one exception

```
dt[which.max(dt$PROPDMG.TOTAL)]
```

```
##    STATE__          BGN_DATE    BGN_TIME TIME_ZONE COUNTY
COUNTYNAME STATE
## 1:       6 1/1/2006 0:00:00 12:00:00 AM       PST     55
NAPA    CA
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI          END_DATE
END_TIME
## 1:  FLOOD         0          COUNTYWIDE 1/1/2006 0:00:00
07:00:00 AM
##    COUNTY_END COUNTYENDN END_RANGE END_AZI END_LOCATI
LENGTH WIDTH  F MAG
## 1:          0          NA         0          COUNTYWIDE
0     0 NA   0
##    FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG
CROPDMGEXP WFO
## 1:          0        0     115          9    32.5
6 MTR
##              STATEOFFIC ZONENAMES LATITUDE LONGITUDE
LATITUDE_E LONGITUDE_
## 1: CALIFORNIA, Western               3828     12218
3828      12218
##
REMARKS
## 1: Major flooding continued into the early hours of January
1st, before the Napa River finally fell below flood stage and
the water receeded. Flooding was severe in Downtown Napa from
the Napa Creek and the City and Parks Department was hit with
$6 million in damage alone. The City of Napa had 600 homes
with moderate damage, 150 damaged businesses with costs of at
least $70 million.
##    REFNUM COUNT YEAR PROPDMG.TOTAL CROPDMG.TOTAL
## 1: 605943     1 2006        115000          32.5
```

```
dt[YEAR == 2006 & REFNUM == 605943]$PROPDMG.TOTAL = 70
```

Cleaning and categorizing event types

Due to lack of standarized vocabulary and as result of muliple typographical errors dataset of interest contains large number of unique envent labels. To obtain meaningful results I decided to reduce number of events labels by fixing obvious errors and introducing classification based on National Weather Service Instruction

```
length(unique(dt$EVTYPE))
```

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

Whitespace normalization

At the beginning I removed leading, trailing and repeated whitspaces from events labels.

```
dt$EVTYPE <- toupper(gsub("(^\\s*)|(\\s*$)|((?<=\\s)\\s+)",
"", dt$EVTYPE, perl = T))
```

Excluding long-term events, summary rows and entries which cannot be easly attributed to the weather events.

Events corresponding to summary fields and isolated events or ambiguous have been removed.

```
length(unique(dt$EVTYPE))
```

```
## [1] 883
```

```
# Remove summary rows
exclude.list <- c("NONE", "?", "OTHER", "MARINE ACCIDENT",
"WET MONTH", "WET YEAR",
    "APACHE COUNTY", "NO SEVERE WEATHER", "MONTHLY
PRECIPITATION", "UNSEASONABLY WARM YEAR",
    "DROWNING", "SOUTHEAST", "EXCESSIVE", "HIGH", "MILD
PATTERN", "NORTHERN LIGHTS",
    "RECORD TEMPERATURES", "RECORD TEMPERATURE", "SEVERE
TURBULENCE", "MONTHLY TEMPERATURE",
    "TEMPERATURE RECORD")
dt <- dt[!grepl("(^SUMMARY)|(SUMMARY$)", EVTYPE) & !EVTYPE %
in% exclude.list,
    ]
length(unique(dt$EVTYPE))
```

```
## [1] 795
```

Combining labels based on term similarity

Reamaining labels have been replaced with general terms inspired by National Weather Service Instruction classification using set of regular expressions. Detailed mapping and number of removed entries at each step can be found below.

Combining thunderstorm realated entries

```
thuderstorm.pattern <- "^(THUN?D?EE?RE?STORM)|(TSTMW)|(TSTM)|
(THUNDERSTORMW)|(THUNDERTORM)|(THUNDERTSORM)|(THUNDERSTROM)|
(TUNDERSTORM)|(SEVERE THUNDERSTORM)|(GUSTNADO)|(GUSTY
THUNDERSTORM WINDS)|(^GUSTY THUNDERSTORM WIND$)|(^THUNDESTORM
WINDS$)|(MICR?OBURST)|(^DOWNBURST)"
grep(thuderstorm.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##   [1] "DOWNBURST"                      "DOWNBURST
WINDS"
##   [3] "DRY MICROBURST"                 "DRY MICROBURST
50"
##   [5] "DRY MICROBURST 53"              "DRY MICROBURST
58"
##   [7] "DRY MICROBURST 61"              "DRY MICROBURST
84"
##   [9] "DRY MICROBURST WINDS"
"GUSTNADO"
##  [11] "GUSTNADO AND"                   "GUSTY THUNDERSTORM
WIND"
##  [13] "GUSTY THUNDERSTORM WINDS"       "MARINE TSTM
WIND"
##  [15] "MICROBURST"                     "MICROBURST
WINDS"
##  [17] "NON TSTM WIND"                  "NON-TSTM
WIND"
##  [19] "SEVERE THUNDERSTORM"            "SEVERE
THUNDERSTORMS"
##  [21] "SEVERE THUNDERSTORM WINDS"      "THUDERSTORM
WINDS"
##  [23] "THUNDEERSTORM WINDS"            "THUNDERESTORM
WINDS"
##  [25] "THUNDERSTORM"                   "THUNDERSTORM
DAMAGE"
##  [27] "THUNDERSTORM DAMAGE TO"         "THUNDERSTORM
HAIL"
##  [29] "THUNDERSTORMS"                  "THUNDERSTORMS
WIND"
##  [31] "THUNDERSTORMS WINDS"
"THUNDERSTORMW"
##  [33] "THUNDERSTORMW 50"               "THUNDERSTORM
WIND"
##  [35] "THUNDERSTORM WIND."             "THUNDERSTORM WIND
50"
##  [37] "THUNDERSTORM WIND 52"           "THUNDERSTORM WIND
56"
##  [39] "THUNDERSTORM WIND 59"           "THUNDERSTORM WIND
59 MPH"
##  [41] "THUNDERSTORM WIND 59 MPH."      "THUNDERSTORM WIND
60 MPH"
##  [43] "THUNDERSTORM WIND 65MPH"        "THUNDERSTORM WIND
65 MPH"
##  [45] "THUNDERSTORM WIND 69"           "THUNDERSTORM WIND
98 MPH"
##  [47] "THUNDERSTORM WIND/AWNING"       "THUNDERSTORM WIND
(G40)"
##  [49] "THUNDERSTORM WIND G50"          "THUNDERSTORM WIND
G51"
##  [51] "THUNDERSTORM WIND G52"          "THUNDERSTORM WIND
G55"
##  [53] "THUNDERSTORM WIND G60"          "THUNDERSTORM WIND
G61"
##  [55] "THUNDERSTORM WIND/HAIL"         "THUNDERSTORM
WIND/LIGHTNING"
##  [57] "THUNDERSTORMWINDS"              "THUNDERSTORM
WINDS"
##  [59] "THUNDERSTORM W INDS"            "THUNDERSTORM
WINDS."
##  [61] "THUNDERSTORM WINDS 13"          "THUNDERSTORM WINDS
2"
```

```
##  [63] "THUNDERSTORM WINDS 50"          "THUNDERSTORM WINDS
52"
##  [65] "THUNDERSTORM WINDS53"           "THUNDERSTORM WINDS
53"
##  [67] "THUNDERSTORM WINDS 60"          "THUNDERSTORM WINDS
61"
##  [69] "THUNDERSTORM WINDS 62"          "THUNDERSTORM WINDS
63 MPH"
##  [71] "THUNDERSTORM WINDS AND"         "THUNDERSTORM
WINDS/FLASH FLOOD"
##  [73] "THUNDERSTORM WINDS/ FLOOD"      "THUNDERSTORM
WINDS/FLOODING"
##  [75] "THUNDERSTORM WINDS FUNNEL CLOU" "THUNDERSTORM
WINDS/FUNNEL CLOU"
##  [77] "THUNDERSTORM WINDS G"           "THUNDERSTORM WINDS
G60"
##  [79] "THUNDERSTORM WINDSHAIL"         "THUNDERSTORM WINDS
HAIL"
##  [81] "THUNDERSTORM WINDS/HAIL"        "THUNDERSTORM WINDS/
HAIL"
##  [83] "THUNDERSTORM WINDS HEAVY RAIN"  "THUNDERSTORM
WINDS/HEAVY RAIN"
##  [85] "THUNDERSTORM WINDS LE CEN"      "THUNDERSTORM WINDS
LIGHTNING"
##  [87] "THUNDERSTORM WINDSS"            "THUNDERSTORM WINDS
SMALL STREA"
##  [89] "THUNDERSTORM WINDS URBAN FLOOD" "THUNDERSTORM WIND/
TREE"
##  [91] "THUNDERSTORM WIND TREES"        "THUNDERSTORM WIND/
TREES"
##  [93] "THUNDERSTORM WINS"              "THUNDERSTORMW
WINDS"
##  [95] "THUNDERSTROM WIND"              "THUNDERSTROM
WINDS"
##  [97] "THUNDERTORM WINDS"              "THUNDERTSORM
WIND"
##  [99] "THUNDESTORM WINDS"              "THUNERSTORM
WINDS"
## [101] "TORNADOES, TSTM WIND, HAIL"
"TSTM"
## [103] "TSTM HEAVY RAIN"
"TSTMW"
## [105] "TSTM WIND"                     "TSTM WIND
40"
## [107] "TSTM WIND (41)"                "TSTM WIND
45"
## [109] "TSTM WIND 50"                  "TSTM WIND
51"
## [111] "TSTM WIND 52"                  "TSTM WIND
55"
## [113] "TSTM WIND 65)"                 "TSTM WIND AND
LIGHTNING"
## [115] "TSTM WIND DAMAGE"              "TSTM WIND
(G35)"
## [117] "TSTM WIND (G40)"               "TSTM WIND
G45"
## [119] "TSTM WIND (G45)"               "TSTM WIND
G58"
## [121] "TSTM WIND/HAIL"                "TSTM
WINDS"
## [123] "TSTM WND"                      "TUNDERSTORM
WIND"
## [125] "WET MICOBURST"                 "WET MICROBURST"
```

```
dt[grepl(thuderstorm.pattern, EVTYPE), ]$EVTYPE <-
"THUNDERSTORM / THUNDERSTORM WIND"
length(unique(dt$EVTYPE))
```

```
## [1] 670
```

Combining hail realated entries

```
hail.pattern <- "HAIL"
grep(hail.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "DEEP HAIL"          "FUNNEL CLOUD/HAIL" "GUSTY
WIND/HAIL"
##  [4] "HAIL"               "HAIL 075"          "HAIL
0.75"
##  [7] "HAIL(0.75)"         "HAIL 088"          "HAIL
0.88"
## [10] "HAIL 100"           "HAIL 1.00"         "HAIL
125"
## [13] "HAIL 150"           "HAIL 175"          "HAIL
1.75"
## [16] "HAIL 1.75)"         "HAIL 200"          "HAIL
225"
## [19] "HAIL 275"           "HAIL 450"          "HAIL
75"
## [22] "HAIL 80"            "HAIL 88"           "HAIL
ALOFT"
## [25] "HAIL DAMAGE"        "HAIL FLOODING"     "HAIL/ICY
ROADS"
## [28] "HAILSTORM"          "HAIL STORM"
"HAILSTORMS"
## [31] "HAIL/WIND"          "HAIL/WINDS"        "LATE SEASON
HAIL"
## [34] "MARINE HAIL"        "NON SEVERE HAIL"   "SMALL
HAIL"
## [37] "WIND/HAIL"
```

```
dt[grepl(hail.pattern, EVTYPE), ]$EVTYPE <- "HAIL"
length(unique(dt$EVTYPE))
```

```
## [1] 634
```

Combining ornado realated entries

```
tornado.pattern <- "(TORNADO)|(TORNDAO)|(WAY?TER\\s?SPOUT)|
(FUNNEL)|(LANDSPOUT)|(WHIRLWIND)"
grep(tornado.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "COLD AIR FUNNEL"          "COLD AIR FUNNELS"
##  [3] "COLD AIR TORNADO"         "DUST DEVIL WATERSPOUT"
##  [5] "FUNNEL"                   "FUNNEL CLOUD"
##  [7] "FUNNEL CLOUD."            "FUNNEL CLOUDS"
##  [9] "FUNNELS"                  "LANDSPOUT"
## [11] "TORNADO"                  "TORNADO DEBRIS"
## [13] "TORNADOES"                "TORNADO F0"
## [15] "TORNADO F1"               "TORNADO F2"
## [17] "TORNADO F3"               "TORNADOS"
## [19] "TORNADO/WATERSPOUT"       "TORNDAO"
## [21] "WALL CLOUD/FUNNEL CLOUD" "WATERSPOUT"
## [23] "WATER SPOUT"              "WATERSPOUT-"
## [25] "WATERSPOUT/"              "WATERSPOUT FUNNEL CLOUD"
## [27] "WATERSPOUTS"              "WATERSPOUT TORNADO"
## [29] "WATERSPOUT-TORNADO"       "WATERSPOUT/TORNADO"
## [31] "WATERSPOUT/ TORNADO"      "WAYTERSPOUT"
## [33] "WHIRLWIND"
```

```
dt[grepl(tornado.pattern, EVTYPE), ]$EVTYPE <- "TORNADO"
length(unique(dt$EVTYPE))
```

```
## [1] 602
```

Combining wildfire realated entries

```
wildfire.pattern <- "(^((BRUSH)|(WILD)|(FOREST)|
(GRASS)).*FIRES?$)|(^RED FLAG CRITERIA)|(^RED FLAG FIRE WX)"
grep(wildfire.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "BRUSH FIRE"        "BRUSH FIRES"       "FOREST
FIRES"
##  [4] "GRASS FIRES"       "RED FLAG CRITERIA" "RED FLAG FIRE
WX"
##  [7] "WILDFIRE"          "WILDFIRES"         "WILD
FIRES"
## [10] "WILD/FOREST FIRE"  "WILD/FOREST FIRES"
```

```
dt[grepl(wildfire.pattern, EVTYPE), ]$EVTYPE <- "WILDFIRE"
length(unique(dt$EVTYPE))
```

```
## [1] 592
```

Combining surge realated entries

```
surge.pattern <- "SURGE"
grep(surge.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "COASTAL SURGE"    "STORM SURGE"      "STORM
SURGE/TIDE"
```

```
dt[grepl(surge.pattern, EVTYPE), ]$EVTYPE <- "SURGE"
length(unique(dt$EVTYPE))
```

```
## [1] 590
```

Combining flash flood realated entries

```
flash.flood.pattern <- "(FLASH.*FLOOD)|(FLOOD.*FLASH)|(^FLASH
FLOOODING)"
grep(flash.flood.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "FLASH FLOOD"                      "FLASH
FLOOD/"
##  [3] "FLASH FLOOD/FLOOD"                "FLASH FLOOD/
FLOOD"
##  [5] "FLASH FLOOD FROM ICE JAMS"        "FLASH FLOOD - HEAVY
RAIN"
##  [7] "FLASH FLOOD/HEAVY RAIN"           "FLASH
FLOODING"
##  [9] "FLASH FLOODING/FLOOD"             "FLASH
FLOODING/THUNDERSTORM WI"
## [11] "FLASH FLOOD/LANDSLIDE"            "FLASH FLOOD
LANDSLIDES"
## [13] "FLASH FLOODS"                     "FLASH FLOOD/
STREET"
## [15] "FLASH FLOOD WINDS"                "FLASH
FLOOODING"
## [17] "FLOOD FLASH"
"FLOOD/FLASH"
## [19] "FLOOD/FLASHFLOOD"                 "FLOOD/FLASH
FLOOD"
## [21] "FLOOD/FLASH/FLOOD"                "FLOOD/FLASH
FLOODING"
## [23] "FLOOD FLOOD/FLASH"                "ICE STORM/FLASH
FLOOD"
## [25] "LOCAL FLASH FLOOD"
```

```
dt[grepl(flash.flood.pattern, EVTYPE), ]$EVTYPE <- "FLASH
FLOOD"
length(unique(dt$EVTYPE))
```

```
## [1] 566
```

Combining coatal flood, beach erosion and tide realated entries

```
coastal.flood.pattern <- "((COASTAL)|(CSTL)|(TIDAL)|(BEACH)).*
((EROSION)|(FLOOD))|(ASTRONOMICAL.*TIDE)|(^HIGH TIDE)|(^BEACH
EROSIN)|(^RAPIDLY RISING WATER)|(^RAPIDLY RISING WATER)|
(^BLOW-OUT TIDE)"
grep(coastal.flood.pattern, sort(unique(dt$EVTYPE)), value =
T)
```

```
##  [1] "ASTRONOMICAL HIGH TIDE"      "ASTRONOMICAL LOW
TIDE"
##  [3] "BEACH EROSIN"                "BEACH
EROSION"
##  [5] "BEACH EROSION/COASTAL FLOOD" "BEACH
FLOOD"
##  [7] "BLOW-OUT TIDE"               "BLOW-OUT
TIDES"
##  [9] "COASTAL EROSION"
"COASTALFLOOD"
## [11] "COASTAL FLOOD"               "COASTAL
FLOODING"
## [13] "COASTAL FLOODING/EROSION"    "COASTAL/TIDAL
FLOOD"
## [15] "CSTL FLOODING/EROSION"       "EROSION/CSTL
FLOOD"
## [17] "HEAVY SURF COASTAL FLOODING" "HIGH
TIDES"
## [19] "HIGH WINDS/COASTAL FLOOD"    "RAPIDLY RISING
WATER"
## [21] "TIDAL FLOOD"                 "TIDAL FLOODING"
```

```
dt[grepl(coastal.flood.pattern, EVTYPE), ]$EVTYPE <- "COASTAL
FLOOD OR EROSION / TIDE"
length(unique(dt$EVTYPE))
```

```
## [1] 545
```

Combining remaining flood realated entries

```
other.flood.pattern <- "(^FLOOD((S)|(ING))?$)|(^LAKE(SHORE)?
FLOOD$)|(^(ICE JAM)|(SNOWMELT)|(MAJOR)|(MINOR)|(RIVER( AND
STREAM )?)|(RURAL) FLOOD(ING)?$)|(^FLOOD & HEAVY RAIN$)|
(^HIGH WATER$)|(^SMA?L?L STREAM)|(^LOCAL FLOOD)|(^HIGHWAY
FLOODING)|(^FLOOD WATCH)|(^BREAKUP FLOODING)|(^STREAM
FLOODING)|(URBAN)|(^STREET FLOOD)"
grep(other.flood.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "BREAKUP FLOODING"
"FLOOD"
##  [3] "FLOOD & HEAVY RAIN"
"FLOODING"
##  [5] "FLOOD/RIVER FLOOD"
"FLOODS"
##  [7] "FLOOD WATCH/"                      "HEAVY RAIN/SMALL
STREAM URBAN"
##  [9] "HEAVY RAIN/URBAN FLOOD"            "HEAVY RAIN; URBAN
FLOOD WINDS;"
## [11] "HIGH WATER"                        "HIGHWAY
FLOODING"
## [13] "ICE JAM"                           "ICE JAM
FLOODING"
## [15] "ICE JAM FLOOD (MINOR"              "LAKE
FLOOD"
## [17] "LAKESHORE FLOOD"                   "LANDSLIDE/URBAN
FLOOD"
## [19] "LOCAL FLOOD"                       "MAJOR
FLOOD"
## [21] "MINOR FLOOD"                       "MINOR
FLOODING"
## [23] "MUD SLIDES URBAN FLOODING"         "RIVER AND STREAM
FLOOD"
## [25] "RIVER FLOOD"                       "RIVER
FLOODING"
## [27] "RURAL FLOOD"                       "SMALL
STREAM"
## [29] "SMALL STREAM AND"                  "SMALL STREAM AND
URBAN FLOOD"
## [31] "SMALL STREAM AND URBAN FLOODIN"    "SMALL STREAM
FLOOD"
## [33] "SMALL STREAM FLOODING"             "SMALL STREAM URBAN
FLOOD"
## [35] "SMALL STREAM/URBAN FLOOD"          "SML STREAM
FLD"
## [37] "SNOWMELT FLOODING"                 "STREAM
FLOODING"
## [39] "STREET FLOOD"                      "STREET
FLOODING"
## [41] "URBAN AND SMALL"                   "URBAN AND SMALL
STREAM"
## [43] "URBAN AND SMALL STREAM FLOOD"      "URBAN AND SMALL
STREAM FLOODIN"
## [45] "URBAN FLOOD"                       "URBAN
FLOODING"
## [47] "URBAN FLOOD LANDSLIDE"             "URBAN
FLOODS"
## [49] "URBAN SMALL"
"URBAN/SMALL"
## [51] "URBAN/SMALL FLOODING"              "URBAN/SMALL
STREAM"
## [53] "URBAN SMALL STREAM FLOOD"          "URBAN/SMALL STREAM
FLOOD"
## [55] "URBAN/SMALL STREAM FLOODING"       "URBAN/SMALL STRM
FLDG"
## [57] "URBAN/SML STREAM FLD"              "URBAN/SML STREAM
FLDG"
## [59] "URBAN/STREET FLOODING"
```

```
dt[grepl(other.flood.pattern, EVTYPE), ]$EVTYPE <- "FLOOD"
length(unique(dt$EVTYPE))
```

```
## [1] 487
```

Combining tropical storm related entries

```
tropical.strom.pattern <- "(TROPICAL STORM)|(HURRICANE)|
(TYPHOON)|(TROPICAL DEPRESSION)|(^REMNANTS OF FLOYD)"
grep(tropical.strom.pattern, sort(unique(dt$EVTYPE)), value =
T)
```

```
##  [1] "HURRICANE"                    "HURRICANE
EDOUARD"
##  [3] "HURRICANE EMILY"              "HURRICANE
ERIN"
##  [5] "HURRICANE FELIX"              "HURRICANE-GENERATED
SWELLS"
##  [7] "HURRICANE GORDON"             "HURRICANE
OPAL"
##  [9] "HURRICANE OPAL/HIGH WINDS"
"HURRICANE/TYPHOON"
## [11] "REMNANTS OF FLOYD"            "TROPICAL
DEPRESSION"
## [13] "TROPICAL STORM"               "TROPICAL STORM
ALBERTO"
## [15] "TROPICAL STORM DEAN"          "TROPICAL STORM
GORDON"
## [17] "TROPICAL STORM JERRY"         "TYPHOON"
```

```
dt[grepl(tropical.strom.pattern, EVTYPE), ]$EVTYPE <-
"TROPICAL STORM / HURRICANE / TYPHOON"
length(unique(dt$EVTYPE))
```

```
## [1] 470
```

Combining snow and ice related entries

```
ice.and.snow.pattern <- "(SNOW)|(ICE)|(GLAZE)|(FREEZING
DRIZZLE)|(FREEZE)|(FROST)|(HEAVY MIX)|(SLEET)|(ICY ROADS)|
(MIXED PRECIPITATION)|(^MIXED PRECIP)|(^FREEZING SPRAY)"
print(grep(ice.and.snow.pattern, sort(unique(dt$EVTYPE)),
value = T))
```

```
##   [1] "ACCUMULATED SNOWFALL"            "AGRICULTURAL
FREEZE"
##   [3] "BLACK ICE"                       "BLIZZARD AND HEAVY
SNOW"
##   [5] "BLIZZARD/HEAVY SNOW"             "BLOWING
SNOW"
##   [7] "BLOWING SNOW & EXTREME WIND CH" "BLOWING SNOW-
EXTREME WIND CHI"
##   [9] "BLOWING SNOW/EXTREME WIND CHIL" "COLD AND
FROST"
##  [11] "COLD AND SNOW"                   "DAMAGING
FREEZE"
##  [13] "DRIFTING SNOW"                   "EARLY
FREEZE"
##  [15] "EARLY FROST"                     "EARLY
SNOW"
##  [17] "EARLY SNOWFALL"                  "EXCESSIVE
SNOW"
##  [19] "FALLING SNOW/ICE"                "FIRST
FROST"
##  [21] "FIRST SNOW"
"FREEZE"
##  [23] "FREEZING DRIZZLE"                "FREEZING DRIZZLE
AND FREEZING"
##  [25] "FREEZING RAIN AND SLEET"         "FREEZING RAIN AND
SNOW"
##  [27] "FREEZING RAIN/SLEET"             "FREEZING RAIN SLEET
AND"
##  [29] "FREEZING RAIN SLEET AND LIGHT"  "FREEZING
RAIN/SNOW"
##  [31] "FREEZING SPRAY"
"FROST"
##  [33] "FROST/FREEZE"
"FROST\\FREEZE"
##  [35] "GLAZE"                           "GLAZE
ICE"
##  [37] "GLAZE/ICE STORM"                 "HARD
FREEZE"
##  [39] "HEAVY LAKE SNOW"                 "HEAVY
MIX"
##  [41] "HEAVY RAIN/SNOW"                 "HEAVY
SNOW"
##  [43] "HEAVY SNOW AND"                  "HEAVY SNOW
ANDBLOWING SNOW"
##  [45] "HEAVY SNOW AND HIGH WINDS"       "HEAVY SNOW AND
ICE"
##  [47] "HEAVY SNOW AND ICE STORM"        "HEAVY SNOW AND
STRONG WINDS"
##  [49] "HEAVY SNOW/BLIZZARD"             "HEAVY
SNOW/BLIZZARD/AVALANCHE"
##  [51] "HEAVY SNOW/BLOWING SNOW"         "HEAVY SNOW FREEZING
RAIN"
##  [53] "HEAVY SNOW/FREEZING RAIN"        "HEAVY
SNOW/HIGH"
##  [55] "HEAVY SNOW/HIGH WIND"            "HEAVY SNOW/HIGH
WINDS"
##  [57] "HEAVY SNOW/HIGH WINDS & FLOOD"  "HEAVY SNOW/HIGH
WINDS/FREEZING"
##  [59] "HEAVY SNOW & ICE"                "HEAVY
SNOW/ICE"
##  [61] "HEAVY SNOW/ICE STORM"            "HEAVY
SNOWPACK"
```

```
##  [63] "HEAVY SNOW SHOWER"               "HEAVY
SNOW/SLEET"
##  [65] "HEAVY SNOW SQUALLS"              "HEAVY SNOW-
SQUALLS"
##  [67] "HEAVY SNOW/SQUALLS"              "HEAVY
SNOW/WIND"
##  [69] "HEAVY SNOW/WINTER STORM"         "HEAVY WET
SNOW"
##  [71] "HIGH WIND AND HEAVY SNOW"        "HIGH WIND/HEAVY
SNOW"
##  [73] "HIGH WINDS/SNOW"
"ICE"
##  [75] "ICE AND SNOW"                    "ICE
FLOES"
##  [77] "ICE FOG"                         "ICE ON
ROAD"
##  [79] "ICE PELLETS"                     "ICE
ROADS"
##  [81] "ICE/SNOW"                        "ICE
STORM"
##  [83] "ICE STORM AND SNOW"
"ICESTORM/BLIZZARD"
##  [85] "ICE/STRONG WINDS"                "ICY
ROADS"
##  [87] "LACK OF SNOW"                    "LAKE EFFECT
SNOW"
##  [89] "LAKE-EFFECT SNOW"                "LATE
FREEZE"
##  [91] "LATE SEASON SNOW"                "LATE SEASON
SNOWFALL"
##  [93] "LATE-SEASON SNOWFALL"            "LATE
SNOW"
##  [95] "LIGHT SNOW"                      "LIGHT SNOW AND
SLEET"
##  [97] "LIGHT SNOWFALL"                  "LIGHT
SNOW/FLURRIES"
##  [99] "LIGHT SNOW/FREEZING PRECIP"      "MIXED
PRECIP"
## [101] "MIXED PRECIPITATION"             "MODERATE
SNOW"
## [103] "MODERATE SNOWFALL"               "MONTHLY
SNOWFALL"
## [105] "MOUNTAIN SNOWS"                  "NEAR RECORD
SNOW"
## [107] "PATCHY ICE"                      "PROLONG
COLD/SNOW"
## [109] "RAIN/SNOW"                       "RECORD
COLD/FROST"
## [111] "RECORD MAY SNOW"                 "RECORD
SNOW"
## [113] "RECORD SNOW/COLD"                "RECORD
SNOWFALL"
## [115] "RECORD WINTER SNOW"              "SEASONAL
SNOWFALL"
## [117] "SLEET"                           "SLEET & FREEZING
RAIN"
## [119] "SLEET/FREEZING RAIN"             "SLEET/ICE
STORM"
## [121] "SLEET/RAIN/SNOW"
"SLEET/SNOW"
## [123] "SLEET STORM"
"SNOW"
## [125] "SNOW ACCUMULATION"               "SNOW
```

```
ADVISORY"
## [127] "SNOW AND COLD"                    "SNOW AND HEAVY
SNOW"
## [129] "SNOW AND ICE"                     "SNOW AND ICE
STORM"
## [131] "SNOW AND SLEET"                   "SNOW AND
WIND"
## [133] "SNOW/ BITTER COLD"                "SNOW/BLOWING
SNOW"
## [135] "SNOW/COLD"
"SNOW\\COLD"
## [137] "SNOW DROUGHT"                     "SNOWFALL
RECORD"
## [139] "SNOW FREEZING RAIN"               "SNOW/FREEZING
RAIN"
## [141] "SNOW/HEAVY SNOW"                  "SNOW/HIGH
WINDS"
## [143] "SNOW- HIGH WIND- WIND CHILL"
"SNOW/ICE"
## [145] "SNOW/ ICE"                        "SNOW/ICE
STORM"
## [147] "SNOW/RAIN"
"SNOW/RAIN/SLEET"
## [149] "SNOW SHOWERS"                     "SNOW
SLEET"
## [151] "SNOW/SLEET"                       "SNOW/SLEET/FREEZING
RAIN"
## [153] "SNOW/SLEET/RAIN"                  "SNOW
SQUALL"
## [155] "SNOW SQUALLS"
"SNOWSTORM"
## [157] "THUNDERSNOW"                      "THUNDERSNOW
SHOWER"
## [159] "UNUSUALLY LATE SNOW"              "WET SNOW"
```

```
dt[grepl(ice.and.snow.pattern, EVTYPE), ]$EVTYPE <- "ICE /
SNOW / FROST"
length(unique(dt$EVTYPE))
```

```
## [1] 311
```

Combining dust realated entries

```
dust.storm.pattern <- "DUST"
grep(dust.storm.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "BLOWING DUST"          "DUST DEVEL"            "DUST
DEVIL"
## [4] "DUSTSTORM"             "DUST STORM"            "DUST
STORM/HIGH WINDS"
## [7] "HIGH WINDS DUST STORM" "SAHARAN DUST"
```

```
dt[grepl(dust.storm.pattern, EVTYPE), ]$EVTYPE <- "DUST
STORM / DEVIL"
length(unique(dt$EVTYPE))
```

```
## [1] 304
```

Combining cold realated entries

```
cold.pattern <- "(COLD)|(CHILL)|(COOL)|(LOW TEMPERATURE)|
(HYPOTHERMIA)|(^UNSEASONAL LOW TEMP)|(^RECORD LOW$)"
grep(cold.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "BITTER WIND CHILL"              "BITTER WIND CHILL
TEMPERATURES"
##  [3] "COLD"                           "COLD AND WET
CONDITIONS"
##  [5] "COLD TEMPERATURE"               "COLD
TEMPERATURES"
##  [7] "COLD WAVE"                      "COLD
WEATHER"
##  [9] "COLD/WIND CHILL"                "COLD WIND CHILL
TEMPERATURES"
## [11] "COLD/WINDS"                     "COOL AND
WET"
## [13] "COOL SPELL"                     "EXCESSIVE
COLD"
## [15] "EXTENDED COLD"                  "EXTREME
COLD"
## [17] "EXTREME COLD/WIND CHILL"        "EXTREME/RECORD
COLD"
## [19] "EXTREME WINDCHILL"              "EXTREME WIND
CHILL"
## [21] "EXTREME WIND CHILL/BLOWING SNO" "EXTREME WIND
CHILLS"
## [23] "EXTREME WINDCHILL TEMPERATURES" "FOG AND COLD
TEMPERATURES"
## [25] "HIGH WIND/LOW WIND CHILL"       "HIGH WINDS AND WIND
CHILL"
## [27] "HIGH WINDS/COLD"                "HIGH WIND/WIND
CHILL"
## [29] "HIGH WIND/WIND CHILL/BLIZZARD"
"HYPOTHERMIA"
## [31] "HYPOTHERMIA/EXPOSURE"           "LOW
TEMPERATURE"
## [33] "LOW TEMPERATURE RECORD"         "LOW WIND
CHILL"
## [35] "PROLONG COLD"                   "RECORD
COLD"
## [37] "RECORD COLD AND HIGH WIND"      "RECORD
COOL"
## [39] "RECORD LOW"                     "SEVERE
COLD"
## [41] "UNSEASONABLE COLD"              "UNSEASONABLY
COLD"
## [43] "UNSEASONABLY COOL"              "UNSEASONABLY COOL &
WET"
## [45] "UNSEASONAL LOW TEMP"            "UNUSUALLY
COLD"
## [47] "WIND CHILL"                     "WIND CHILL/HIGH
WIND"
```

```
dt[grepl(cold.pattern, EVTYPE), ]$EVTYPE <- "COLD AND
WINDCHILL"
length(unique(dt$EVTYPE))
```

```
## [1] 257
```

Combining generic wind realated entries

```
wind.pattern <- "(((GRADIENT)|(HIGH)|(GUSTY)|(STRONG)) WIND)|
(^WINDS?$)|(^NON-SEVERE WIND DAMAGE$)|(^WIND DAMAGE$)|(^STORM
FORCE WINDS$)|(^WND$)|(^WIND GUSTS)|(^WIND ADVISORY)|(^WIND
STORM)|(^GUSTY LAKE WIND)|(^WAKE LOW WIND)"
grep(wind.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "BLIZZARD/HIGH WIND"              "FLOOD/STRONG
WIND"
##  [3] "GRADIENT WIND"                   "GRADIENT
WINDS"
##  [5] "GUSTY LAKE WIND"                 "GUSTY
WIND"
##  [7] "GUSTY WIND/HVY RAIN"             "GUSTY
WIND/RAIN"
##  [9] "GUSTY WINDS"                     "HIGH
WIND"
## [11] "HIGH WIND 48"                    "HIGH WIND
63"
## [13] "HIGH WIND 70"                    "HIGH WIND AND HIGH
TIDES"
## [15] "HIGH WIND AND SEAS"              "HIGH
WIND/BLIZZARD"
## [17] "HIGH WIND/ BLIZZARD"             "HIGH
WIND/BLIZZARD/FREEZING RA"
## [19] "HIGH WIND DAMAGE"                "HIGH WIND
(G40)"
## [21] "HIGH WINDS"                      "HIGH
WINDS/"
## [23] "HIGH WINDS 55"                   "HIGH WINDS
57"
## [25] "HIGH WINDS 58"                   "HIGH WINDS
63"
## [27] "HIGH WINDS 66"                   "HIGH WINDS
67"
## [29] "HIGH WINDS 73"                   "HIGH WINDS
76"
## [31] "HIGH WINDS 80"                   "HIGH WINDS
82"
## [33] "HIGH WIND/SEAS"                  "HIGH
WINDS/FLOODING"
## [35] "HIGH WINDS/HEAVY RAIN"           "HIGH WINDS HEAVY
RAINS"
## [37] "MARINE HIGH WIND"                "MARINE STRONG
WIND"
## [39] "NON-SEVERE WIND DAMAGE"          "STORM FORCE
WINDS"
## [41] "STRONG WIND"                     "STRONG WIND
GUST"
## [43] "STRONG WINDS"                    "WAKE LOW
WIND"
## [45] "WIND"                            "WIND
ADVISORY"
## [47] "WIND DAMAGE"                     "WIND
GUSTS"
## [49] "WINDS"                           "WIND
STORM"
## [51] "WINTER STORM/HIGH WIND"          "WINTER STORM HIGH
WINDS"
## [53] "WINTER STORM/HIGH WINDS"         "WND"
```

```
dt[grepl(wind.pattern, EVTYPE), ]$EVTYPE <- "WIND"
length(unique(dt$EVTYPE))
```

```
## [1] 204
```

Combining rain and wet weather realated entries

```
rain.pattern <- "(RAIN)|(HEAVY PRECIPITATION)|(HEAVY SHOWER)|
(^METRO STORM, MAY 26)|(^HEAVY PRECIPATATION)|(^UNSEASONABLY
WET)|(^EXCESSIVE PRECIPITATION)|(^NORMAL PRECIPITATION)|(^WET
WEATHER)|(^EXCESSIVE WETNESS)|(^EXTREMELY WET)|(^RECORD
PRECIPITATION)|(ABNORMALLY WET)"
grep(rain.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "ABNORMALLY WET"              "BLIZZARD/FREEZING
RAIN"
##  [3] "EARLY RAIN"                  "EXCESSIVE
PRECIPITATION"
##  [5] "EXCESSIVE RAIN"              "EXCESSIVE
RAINFALL"
##  [7] "EXCESSIVE WETNESS"           "EXTREMELY
WET"
##  [9] "FLOODING/HEAVY RAIN"
"FLOOD/RAIN/WIND"
## [11] "FLOOD/RAIN/WINDS"            "FREEZING
RAIN"
## [13] "HEAVY PRECIPATATION"         "HEAVY
PRECIPITATION"
## [15] "HEAVY RAIN"                  "HEAVY RAIN AND
FLOOD"
## [17] "HEAVY RAIN AND WIND"         "HEAVY RAIN
EFFECTS"
## [19] "HEAVY RAINFALL"              "HEAVY
RAIN/FLOODING"
## [21] "HEAVY RAIN/HIGH SURF"        "HEAVY
RAIN/LIGHTNING"
## [23] "HEAVY RAIN/MUDSLIDES/FLOOD" "HEAVY
RAINS"
## [25] "HEAVY RAIN/SEVERE WEATHER"  "HEAVY
RAINS/FLOODING"
## [27] "HEAVY RAIN/WIND"             "HEAVY
SHOWER"
## [29] "HEAVY SHOWERS"               "HVY
RAIN"
## [31] "LIGHT FREEZING RAIN"         "LIGHTNING AND HEAVY
RAIN"
## [33] "LIGHTNING/HEAVY RAIN"        "LOCALLY HEAVY
RAIN"
## [35] "METRO STORM, MAY 26"         "MONTHLY
RAINFALL"
## [37] "NORMAL PRECIPITATION"        "PROLONGED
RAIN"
## [39] "RAIN"                        "RAIN AND
WIND"
## [41] "RAIN DAMAGE"                 "RAIN
(HEAVY)"
## [43] "RAINSTORM"
"RAIN/WIND"
## [45] "RECORD/EXCESSIVE RAINFALL"  "RECORD LOW
RAINFALL"
## [47] "RECORD PRECIPITATION"        "RECORD
RAINFALL"
## [49] "TORRENTIAL RAIN"             "TORRENTIAL
RAINFALL"
## [51] "UNSEASONABLY WET"            "UNSEASONAL
RAIN"
## [53] "WET WEATHER"
```

```
dt[grepl(rain.pattern, EVTYPE), ]$EVTYPE <- "RAIN OR WET"
length(unique(dt$EVTYPE))
```

```
## [1] 152
```

Combining mudslide, rock slide and landslide realated entries

```
land.slide.pattern <- "(MUD\\s?SLIDE)|(ROCK SLIDE)|(LANDSLUMP)
|(LANDSLIDES?)"
grep(land.slide.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "LANDSLIDE"          "LANDSLIDES"
"LANDSLUMP"
##  [4] "MUD/ROCK SLIDE"     "MUDSLIDE"           "MUD
SLIDE"
##  [7] "MUDSLIDE/LANDSLIDE" "MUDSLIDES"          "MUD
SLIDES"
## [10] "ROCK SLIDE"
```

```
dt[grepl(land.slide.pattern, EVTYPE), ]$EVTYPE <- "LANDSLIDE /
MUDSLIDE / ROCK SLIDE"
length(unique(dt$EVTYPE))
```

```
## [1] 143
```

Combining lightning realated entries

```
lightning.pattern <- "(LIGHTNING)|(LIGHTING)|(LIGNTNING)"
grep(lightning.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "LIGHTING"
"LIGHTNING"
##  [3] "LIGHTNING."                    "LIGHTNING AND
THUNDERSTORM WIN"
##  [5] "LIGHTNING AND WINDS"           "LIGHTNING
DAMAGE"
##  [7] "LIGHTNING FIRE"                "LIGHTNING
INJURY"
##  [9] "LIGHTNING THUNDERSTORM WINDS"  "LIGHTNING
THUNDERSTORM WINDSS"
## [11] "LIGHTNING WAUSEON"             "LIGNTNING"
```

```
dt[grepl(lightning.pattern, EVTYPE), ]$EVTYPE <- "LIGHTNING"
length(unique(dt$EVTYPE))
```

```
## [1] 132
```

Combining dust realated entries

```
fog.pattern <- "FOG"
grep(fog.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "DENSE FOG"        "FOG"              "FREEZING
FOG"
## [4] "PATCHY DENSE FOG"
```

```
dt[grepl(fog.pattern, EVTYPE), ]$EVTYPE <- "FOG"
length(unique(dt$EVTYPE))
```

```
## [1] 129
```

Combining winter storm and blizzard realated entries

```
blizzard.pattern <- "(BLIZZARD)|(WINTER STORM)"
grep(blizzard.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "BLIZZARD"                          "BLIZZARD AND EXTREME
WIND CHIL"
## [3] "BLIZZARD WEATHER"                  "BLIZZARD/WINTER
STORM"
## [5] "GROUND BLIZZARD"                   "WINTER
STORM"
## [7] "WINTER STORMS"
```

```
dt[grepl(blizzard.pattern, EVTYPE), ]$EVTYPE <- "BLIZZARD /
WINTER STORM"
length(unique(dt$EVTYPE))
```

```
## [1] 123
```

Combining winter weather realated entries

```
winter.pattern <- "(WINTER WEATHER)|(WINTRY MIX)|(^WINTER MIX)
|(^WINTERY MIX)"
grep(winter.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "WINTER MIX"         "WINTER WEATHER"     "WINTER
WEATHER MIX"
## [4] "WINTER WEATHER/MIX" "WINTERY MIX"        "WINTRY MIX"
```

```
dt[grepl(winter.pattern, EVTYPE), ]$EVTYPE <- "WINTER WEATHER"
length(unique(dt$EVTYPE))
```

```
## [1] 118
```

Combining heat realated entries

```
heat.pattern <- "(HEAT)|(RECORD WARMTH)|(HOT)|(WARM)|(^RECORD
HIGH)|(HYPERTHERMIA)|(HIGH TEMPERATURE RECORD)"
grep(heat.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "ABNORMAL WARMTH"            "DROUGHT/EXCESSIVE
HEAT"
##  [3] "DRY HOT WEATHER"            "EXCESSIVE
HEAT"
##  [5] "EXCESSIVE HEAT/DROUGHT"     "EXTREME
HEAT"
##  [7] "HEAT"
"HEATBURST"
##  [9] "HEAT DROUGHT"
"HEAT/DROUGHT"
## [11] "HEAT WAVE"                  "HEAT WAVE
DROUGHT"
## [13] "HEAT WAVES"                 "HIGH TEMPERATURE
RECORD"
## [15] "HOT AND DRY"                "HOT/DRY
PATTERN"
## [17] "HOT PATTERN"                "HOT
SPELL"
## [19] "HOT WEATHER"
"HYPERTHERMIA/EXPOSURE"
## [21] "PROLONG WARMTH"             "RECORD/EXCESSIVE
HEAT"
## [23] "RECORD HEAT"                "RECORD HEAT
WAVE"
## [25] "RECORD HIGH"                "RECORD HIGH
TEMPERATURE"
## [27] "RECORD HIGH TEMPERATURES"   "RECORD
WARM"
## [29] "RECORD WARM TEMPS."         "RECORD
WARMTH"
## [31] "UNSEASONABLY HOT"           "UNSEASONABLY
WARM"
## [33] "UNSEASONABLY WARM AND DRY" "UNSEASONABLY WARM &
WET"
## [35] "UNSEASONABLY WARM/WET"      "UNUSUALLY
WARM"
## [37] "UNUSUAL/RECORD WARMTH"      "UNUSUAL
WARMTH"
## [39] "VERY WARM"                  "WARM DRY
CONDITIONS"
## [41] "WARM WEATHER"
```

```
dt[grepl(heat.pattern, EVTYPE), ]$EVTYPE <- "HEAT"
length(unique(dt$EVTYPE))
```

```
## [1] 78
```

Combining drought realated entries

```
dry.pattern <- "(DROUGHT)|(DRY)|(DRIEST)|(^BELOW NORMAL
PRECIPITATION)"
grep(dry.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "ABNORMALLY DRY"              "BELOW NORMAL
PRECIPITATION"
##  [3] "DRIEST MONTH"
"DROUGHT"
##  [5] "DRY"                         "DRY
CONDITIONS"
##  [7] "DRY MIRCOBURST WINDS"
"DRYNESS"
##  [9] "DRY PATTERN"                 "DRY
SPELL"
## [11] "DRY WEATHER"                 "EXCESSIVELY
DRY"
## [13] "MILD AND DRY PATTERN"        "MILD/DRY
PATTERN"
## [15] "RECORD DRY MONTH"            "RECORD
DRYNESS"
## [17] "UNSEASONABLY DRY"            "VERY DRY"
```

```
dt[grepl(dry.pattern, EVTYPE), ]$EVTYPE <- "DROUGHT"
length(unique(dt$EVTYPE))
```

```
## [1] 61
```

Combining volcanic activity realated entries

```
volcanic.pattern <- "(VOLCANIC ((ASH)|(ERUPTION)))"
grep(volcanic.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "VOLCANIC ASH"       "VOLCANIC ASHFALL"   "VOLCANIC ASH
PLUME"
## [4] "VOLCANIC ERUPTION"
```

```
dt[grepl(volcanic.pattern, EVTYPE), ]$EVTYPE <- "VOLCANIC
ACTIVITY"
length(unique(dt$EVTYPE))
```

```
## [1] 58
```

Combining various marine events

```
marine.pattern <- "(HIGH SEAS)|(SURF)|(RIP CURRENT)|(SWELLS)|
(MARINE)|(HEAVY SEAS)|(^ROUGH SEAS)|(^HIGH WAVES)|(^WIND AND
WAVE)|(^ROGUE WAVE)|(HIGH SURF)|(SEICHE)"
grep(marine.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
##  [1] "HAZARDOUS SURF"            "HEAVY SEAS"
##  [3] "HEAVY SURF"                "HEAVY SURF AND WIND"
##  [5] "HEAVY SURF/HIGH SURF"      "HEAVY SWELLS"
##  [7] "HIGH SEAS"                 "HIGH SURF"
##  [9] "HIGH SURF ADVISORIES"      "HIGH SURF ADVISORY"
## [11] "HIGH SWELLS"               "HIGH WAVES"
## [13] "MARINE MISHAP"             "MARINE THUNDERSTORM WIND"
## [15] "RIP CURRENT"               "RIP CURRENTS"
## [17] "RIP CURRENTS HEAVY SURF"   "RIP CURRENTS/HEAVY SURF"
## [19] "ROGUE WAVE"                "ROUGH SEAS"
## [21] "ROUGH SURF"                "SEICHE"
## [23] "WIND AND WAVE"
```

```
dt[grepl(marine.pattern, EVTYPE), ]$EVTYPE <- "MARINE WIND /
SWELL / SURF"
length(unique(dt$EVTYPE))
```

```
## [1] 36
```

Combining vog and smoke realated entries

```
vog.and.smoke.pattern <- "(VOG)|(SMOKE)"
grep(vog.and.smoke.pattern, sort(unique(dt$EVTYPE)), value =
T)
```

```
## [1] "DENSE SMOKE" "SMOKE"       "VOG"
```

```
dt[grepl(vog.and.smoke.pattern, EVTYPE), ]$EVTYPE <- "VOG /
SMOKE"
length(unique(dt$EVTYPE))
```

```
## [1] 34
```

Combinine avalanche realated entries

```
avalanche.pattern <- "(AVALANCH?E)"
grep(avalanche.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "AVALANCE"  "AVALANCHE"
```

```
dt[grepl(avalanche.pattern, EVTYPE), ]$EVTYPE <- "AVALANCHE"
length(unique(dt$EVTYPE))
```

```
## [1] 33
```

Combining dam failure realated entries

```
dam.pattern <- "(DAM)"
grep(dam.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "DAM BREAK"   "DAM FAILURE"
```

```
dt[grepl(dam.pattern, EVTYPE), ]$EVTYPE <- "DAM FAILURE"
length(unique(dt$EVTYPE))
```

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

Combining coastal storm realated entries

```
coastalstorm.pattern <- "(COASTAL\\s?STORM)"
grep(coastalstorm.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "COASTALSTORM"  "COASTAL STORM"
```

```
dt[grepl(coastalstorm.pattern, EVTYPE), ]$EVTYPE <- "COASTAL
STORM"
length(unique(dt$EVTYPE))
```

```
## [1] 31
```

Combining wall cloud realated entries

```
wallcloud.pattern <- "(WALL CLOUD)"
grep(wallcloud.pattern, sort(unique(dt$EVTYPE)), value = T)
```

```
## [1] "LARGE WALL CLOUD"    "ROTATING WALL CLOUD" "WALL
CLOUD"
```

```
dt[grepl(wallcloud.pattern, EVTYPE), ]$EVTYPE <- "WALL CLOUD"
length(unique(dt$EVTYPE))
```

```
## [1] 29
```

```
print(paste("After initial data processing number of records
have been reduced to",
    dim(dt)[1], "assigned to", length(unique(dt$EVTYPE)),
"categories"))
```

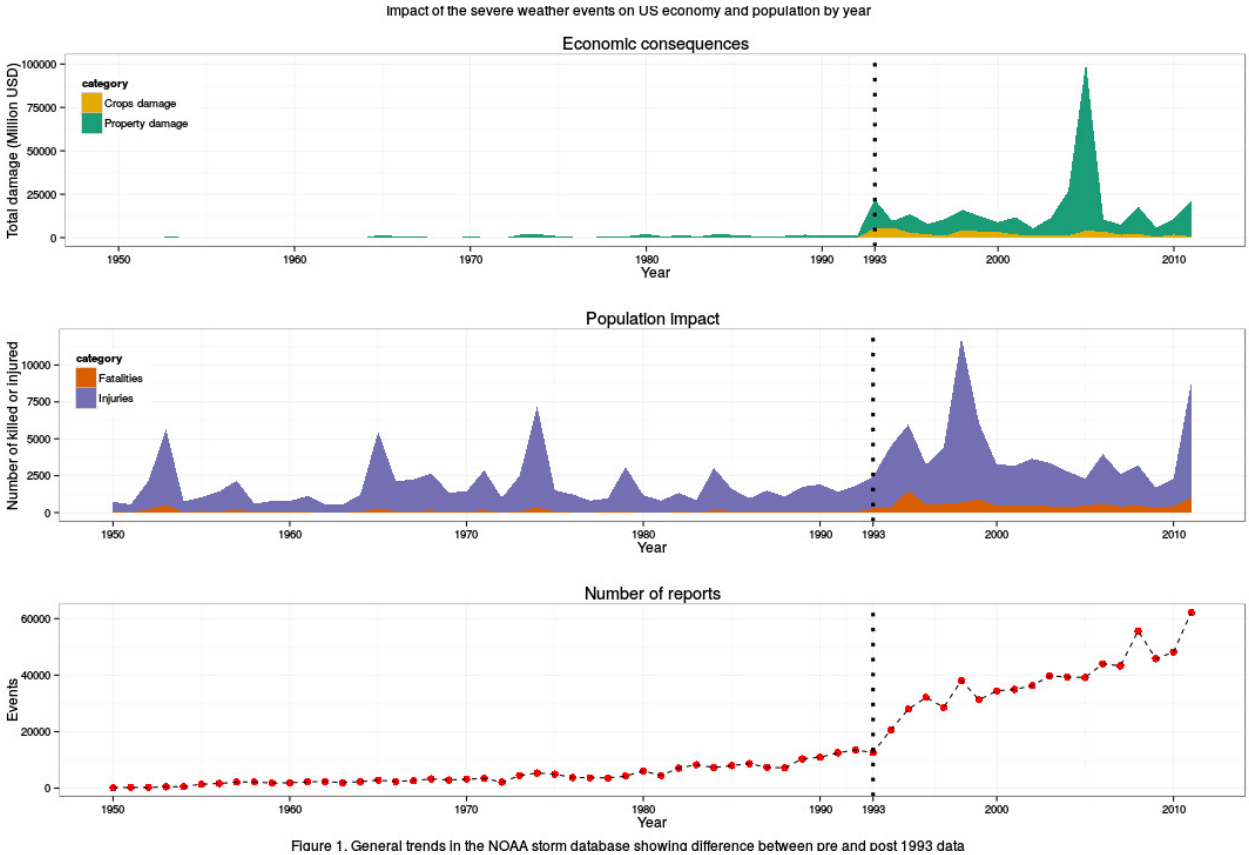
```
## [1] "After initial data processing number of records have
been reduced to 902042 assigned to 29 categories"
```

Data aggregation

Cleaned dataset has been aggregated by year to obtain insight about general trends in number of reported events, number of affected people and amount of property damage.

```
number.of.events.by.year <- dt[, list(COUNT = sum(COUNT)), by
= list(YEAR)]

population.damage.by.year <- rbind(dt[, list(total = sum
(FATALITIES, na.rm = TRUE),
    category = "fatalities"), by = list(YEAR)], dt[, list
(total = sum(INJURIES,
    na.rm = TRUE), category = "injuries"), by = list(YEAR)])

population.damage.by.year$category <- factor
(population.damage.by.year$category,
    labels = c("Fatalities", "Injuries"))

property.and.crop.damage.by.year <- rbind(dt[, list(total =
sum(CROPDMG.TOTAL,
    na.rm = TRUE), category = "cropdmg"), by = list(YEAR)], dt
[, list(total = sum(PROPDMG.TOTAL,
    na.rm = TRUE), category = "propdmg"), by = list(YEAR)])

property.and.crop.damage.by.year$category <- factor
(property.and.crop.damage.by.year$category,
    labels = c("Crops damage", "Property damage"))


plot.number.of.events.by.year <- ggplot
(number.of.events.by.year, aes(x = YEAR,
    y = COUNT)) + geom_point(size = 3, colour = "red") +
geom_line(linetype = "dashed")
plot.number.of.events.by.year <-
plot.number.of.events.by.year + geom_vline(xintercept = 1993,
    linetype = "dotted", colour = "black", size = 1.5)
plot.number.of.events.by.year <-
plot.number.of.events.by.year + scale_x_continuous(breaks = c
(seq(1950,
    1990, 10), 1993, 2000, 2010)) + theme_bw()
plot.number.of.events.by.year <-
plot.number.of.events.by.year + labs(x = "Year",
    y = "Events", title = "Number of reports")

plot.property.and.crop.damage.by.year <- ggplot
(property.and.crop.damage.by.year,
    aes(x = YEAR, y = total, fill = category)) + geom_area
(position = "stack")
plot.property.and.crop.damage.by.year <-
plot.property.and.crop.damage.by.year +
    geom_vline(xintercept = 1993, linetype = "dotted", colour
= "black", size = 1.5)
plot.property.and.crop.damage.by.year <-
plot.property.and.crop.damage.by.year +
    scale_x_continuous(breaks = c(seq(1950, 1990, 10), 1993,
2000, 2010))
plot.property.and.crop.damage.by.year <-
plot.property.and.crop.damage.by.year +
    scale_fill_manual(values = c("#e6ab02", "#1b9e77")) +
theme_bw() + theme(legend.justification = c(0,
    0), legend.position = c(0, 0.5))
plot.property.and.crop.damage.by.year <-
plot.property.and.crop.damage.by.year +
    labs(x = "Year", y = "Total damage (Million USD)", title
= "Economic consequences")
```

```
plot.population.damage.by.year <- ggplot
(population.damage.by.year, aes(x = YEAR,
    y = total, fill = category)) + geom_area(position =
"stack")
plot.population.damage.by.year <-
plot.population.damage.by.year + geom_vline(xintercept =
1993,
    linetype = "dotted", colour = "black", size = 1.5)
plot.population.damage.by.year <-
plot.population.damage.by.year + scale_x_continuous(breaks = c
(seq(1950,
    1990, 10), 1993, 2000, 2010))
plot.population.damage.by.year <-
plot.population.damage.by.year + scale_fill_manual(values = c
("#d95f02",
    "#7570b3")) + theme_bw() + theme(legend.justification = c
(0, 0), legend.position = c(0,
    0.5))
plot.population.damage.by.year <-
plot.population.damage.by.year + labs(x = "Year",
    y = "Number of killed or injured", title = "Population
impact")

grid.arrange(plot.property.and.crop.damage.by.year,
plot.population.damage.by.year,
    plot.number.of.events.by.year, main = "Impact of the
severe weather events on US economy and population by year",
    sub = textGrob("Figure 1. General trends in the NOAA storm
database showing difference between pre and post 1993 data",
        gp = gpar(font = 0.7)))
```



As we can see there is substantial difference between data collected before and after 1993. We can suspect that main reason is a change in methods of collecting data. Since we are interested in the results which can applied to the current situation I decided to keep only the data collected after 1992.

```
print(paste("It contains", dim(dt[YEAR >= 1993])[1], "records
assigned to",
    length(unique(dt[YEAR >= 1993]$EVTYPE)), "categories"))
```

```
## [1] "It contains 714483 records assigned to 29 categories"
```

To analyse impact of the individaual classes of events I decided to use two metrics:

- population impact - sum of the number of injured and killed
- economic impact - sum of amount of crops damage and property damage

Since we have no detailed knowledge about long-term effects of events this approach seems to be justified.

```
dt.agg.by.year.and.evtype.gte.1993 <- dt[YEAR >= 1993,
    list(
        n_events=sum(COUNT),

        total_population_impact=sum(FATALITIES + INJURIES,
na.rm=TRUE),
        avg_population_impact=mean(FATALITIES + INJURIES,
na.rm=TRUE),
        median_population_impact=median(FATALITIES +
INJURIES, na.rm=TRUE),
        sd_population_impact=sd(FATALITIES + INJURIES,
na.rm=TRUE),
        max_population_impact=max(FATALITIES + INJURIES,
na.rm=TRUE),

        avg_economic_impact=mean(PROPDMG.TOTAL +
CROPDMG.TOTAL, na.rm=TRUE),
        median_economic_impact=median(PROPDMG.TOTAL +
CROPDMG.TOTAL, na.rm=TRUE),
        sd_economic_impact=sd(PROPDMG.TOTAL + CROPDMG.TOTAL,
na.rm=TRUE),
        max_economic_impact=sd(PROPDMG.TOTAL + CROPDMG.TOTAL,
na.rm=TRUE),
        total_economic_impact=sum(PROPDMG.TOTAL +
CROPDMG.TOTAL, na.rm=TRUE)

    ),
    by=list(EVTYPE)
]
```

Majority of events recorded between 1993 and 2011 had only marginal impact on the population and property. For the most common classes of events both means and meadians are low or close to 0 as shown below.

```
grid.arrange(
    tableGrob(dt.agg.by.year.and.evtype.gte.1993[
        order(-n_events),
        list(
            Event=EVTYPE,
            N=n_events,
            Total=total_population_impact,
            Mean=avg_population_impact,
            Median=median_population_impact,
            SD=sd_population_impact,
            Max=max_population_impact
        )
    ]),
    sub = textGrob("Table 1. Population impact statistics
1993-2011", gp = gpar(font = 0.7))
)
```

	Event	N	Total	Mean	Median	SD	Max
1	THUNDERSTORM / THUNDERSTORM WIND	240274	6704	0.027901	0	0.53167	70
2	HAIL	227451	980	0.004309	0	0.38663	109
3	FLASH FLOOD	55676	2837	0.050956	0	1.29285	159
4	TORNADO	36776	25031	0.680634	0	11.06392	1308
5	FLOOD	29567	7388	0.249873	0	10.22343	802
6	WIND	26291	2352	0.089460	0	1.01704	89
7	ICE / SNOW / FROST	21557	3896	0.180730	0	10.86275	1569
8	LIGHTNING	15768	6049	0.383625	0	1.31827	51
9	BLIZZARD / WINTER STORM	14161	2460	0.173717	0	4.12213	390
10	RAIN OR WET	12199	411	0.033691	0	0.59505	32
11	WINTER WEATHER	8254	677	0.082021	0	2.14108	138
12	MARINE WIND / SWELL / SURF	7698	1594	0.207067	0	1.14596	55
13	WILDFIRE	4241	1698	0.400377	0	3.63533	153
14	HEAT	3006	12422	4.132402	0	25.87434	583
15	COLD AND WINDCHILL	2739	793	0.289522	0	2.80492	132
16	DROUGHT	2583	5	0.001936	0	0.08112	4
17	FOG	1880	1156	0.614894	0	4.06828	80
18	COASTAL FLOOD OR EROSION / TIDE	1172	15	0.012799	0	0.18205	5
19	TROPICAL STORM / HURRICANE / TYPHOON	1058	1917	1.811909	0	27.19709	787
20	LANDSLIDE / MUDSLIDE / ROCK SLIDE	648	99	0.152778	0	1.26165	24
21	DUST STORM / DEVIL	588	507	0.862245	0	3.89554	40
22	SURGE	411	67	0.163017	0	1.55711	27
23	AVALANCHE	387	395	1.020672	1	1.28949	12
24	VOLCANIC ACTIVITY	29	0	0.000000	0	0.00000	0
25	VOG / SMOKE	22	0	0.000000	0	0.00000	0
26	TSUNAMI	20	162	8.100000	0	35.98962	161
27	WALL CLOUD	11	0	0.000000	0	0.00000	0
28	COASTAL STORM	11	6	0.545455	0	0.68755	2
29	DAM FAILURE	5	0	0.000000	0	0.00000	0

Table 1. Population impact statistics 1993-2011

```
grid.arrange(
    tableGrob(dt.agg.by.year.and.evtype.gte.1993[
        order(-n_events),
        list(
            Event=EVTYPE,
            N=n_events,
            Total=total_economic_impact,
            Mean=avg_economic_impact,
            Median=median_economic_impact,
            SD=sd_economic_impact,
            Max=max_economic_impact
        )
    ]),
    sub = textGrob("Table 2. Economic impact statistics 1993-
2011", gp = gpar(font = 0.7))

)
```

	Event	N	Total	Mean	Median	SD	Max
1	THUNDERSTORM / THUNDERSTORM WIND	240274	14065.991	5.859e-02	0.000	4.731e+00	4.731e+00
2	HAIL	227451	19024.446	8.365e-02	0.000	5.121e+00	5.121e+00
3	FLASH FLOOD	55676	19120.949	3.435e-01	0.000	6.883e+00	6.883e+00
4	TORNADO	36776	26779.681	7.285e-01	0.000	1.914e+01	1.914e+01
5	FLOOD	29567	46038.565	1.557e+00	0.000	6.419e+01	6.419e+01
6	WIND	26291	6893.935	2.623e-01	0.000	1.176e+01	1.176e+01
7	ICE / SNOW / FROST	21557	12166.655	5.645e-01	0.000	3.461e+01	3.461e+01
8	LIGHTNING	15768	947.521	6.014e-02	0.005	3.204e-01	3.204e-01
9	BLIZZARD / WINTER STORM	14161	7488.315	5.288e-01	0.000	4.253e+01	4.253e+01
10	RAIN OR WET	12199	4306.280	3.530e-01	0.000	2.287e+01	2.287e+01
11	WINTER WEATHER	8254	42.310	5.126e-03	0.000	1.732e-01	1.732e-01
12	MARINE WIND / SWELL / SURF	7698	104.140	1.353e-02	0.000	4.434e-01	4.434e-01
13	WILDFIRE	4241	8899.910	2.099e+00	0.000	3.485e+01	3.485e+01
14	HEAT	3006	924.805	3.077e-01	0.000	1.157e+01	1.157e+01
15	COLD AND WINDCHILL	2739	1685.100	6.152e-01	0.000	1.276e+01	1.276e+01
16	DROUGHT	2583	15018.672	5.817e+00	0.000	4.317e+01	4.317e+01
17	FOG	1880	25.011	1.330e-02	0.000	9.927e-02	9.927e-02
18	COASTAL FLOOD OR EROSION / TIDE	1172	459.496	3.924e-01	0.000	3.261e+00	3.261e+00
19	TROPICAL STORM / HURRICANE / TYPHOON	1058	99283.551	9.384e+01	0.030	7.805e+02	7.805e+02
20	LANDSLIDE / MUDSLIDE / ROCK SLIDE	648	347.413	5.361e-01	0.000	3.736e+00	3.736e+00
21	DUST STORM / DEVIL	588	9.938	1.690e-02	0.000	1.065e-01	1.065e-01
22	SURGE	411	47966.079	1.167e+02	0.005	1.651e+03	1.651e+03
23	AVALANCHE	387	3.722	9.617e-03	0.000	8.130e-02	8.130e-02
24	VOLCANIC ACTIVITY	29	0.500	1.724e-02	0.000	6.584e-02	6.584e-02
25	VOG / SMOKE	22	0.100	4.545e-03	0.000	2.132e-02	2.132e-02
26	TSUNAMI	20	144.082	7.204e+00	0.115	1.862e+01	1.862e+01
27	WALL CLOUD	11	0.000	0.000e+00	0.000	0.000e+00	0.000e+00
28	COASTAL STORM	11	0.050	4.545e-03	0.000	1.508e-02	1.508e-02
29	DAM FAILURE	5	1.002	2.004e-01	0.000	4.470e-01	4.470e-01

Table 2. Economic impact statistics 1993-2011

Since distribution of the number of events by class is highly skewed and measures of centrality can be in our case somewhat misleading I decided to use maximum value as a statistic describing general severity of the group. While this approach can be disputed I think it is justified by the fact that most of the time we try avoid worst case scenario.

Finally data has been splitted into four categories:

- Low impact,
- High population, low economic impact
- Low population, high economic impact
- High population, high economic impact

Population impact equal to 50 and economic impact equal to 10 million USD have been used as the spliting lines. This values are arbitrary and where adjusted to fit logscale distribution of the data.

As a auxiliary metric sum of the normalized averages will be used.

```
attach(dt.agg.by.year.and.evtype.gte.1993)
dt.agg.by.year.and.evtype.gte.1993$group <- factor(ifelse
(max_economic_impact <
    10 & max_population_impact < 50, 1, ifelse
(max_economic_impact < 10 & max_population_impact >=
    50, 2, ifelse(max_economic_impact >= 10 &
max_population_impact < 50, 3,
    4))), labels = c("Low impact", "High population, low
economic impact", "Low population, high economic impact",
    "High population, high economic impact"))

detach(dt.agg.by.year.and.evtype.gte.1993)
dt.agg.by.year.and.evtype.gte.1993$EVTYPE <- factor
(dt.agg.by.year.and.evtype.gte.1993$EVTYPE)
dt.agg.by.year.and.evtype.gte.1993$Normalized.average.impact
<- scale(dt.agg.by.year.and.evtype.gte.1993
$avg_economic_impact) +
    scale(dt.agg.by.year.and.evtype.gte.1993
$avg_population_impact)
```

Results

Most of the observed differences can be attributed to the differences in the size of the samples. Nevertheless I was able to identift four types of events which had both high normalized average and maximum impact and should be considered highly harmful. These are the following:

- Surge
- Heat / Heat Wave
- Tropical storm
- Tsunami

Remaining types of events should especially those assigned to non low-impact groups.

```
plot.high.impact.events <- ggplot
(dt.agg.by.year.and.evtype.gte.1993, aes(x =
max_population_impact,
    y = max_economic_impact, label = EVTYPE, colour = group))
plot.high.impact.events <- plot.high.impact.events + geom_text
(hjust = -0.1,
    vjust = -0.1, size = 3, alpha = 0.7)
plot.high.impact.events <- plot.high.impact.events +
geom_point(aes(size = Normalized.average.impact),
    shape = 21) + theme_bw()
plot.high.impact.events <- plot.high.impact.events + labs
(title = "Categories of the severe weather events\n U.S. 2003-
2011",
    y = "Economic impact (million USD)", x = "Population
impact (number of injured and killed)") +
    scale_x_log10(expand = c(0.7, 0)) + scale_y_log10()
plot.high.impact.events <- plot.high.impact.events +
scale_color_manual(values = c("#66c2a5",
    "#fdae61", "#4393c3", "#d53e4f"))

grid.arrange(plot.high.impact.events, sub = textGrob("Figure
2. Categories of the severe weather events. \n Variables had
been plotted  using a logarithmic scale. Size of the points
is proportional to the logarithm of the number of recorded
events",
    gp = gpar(font = 0.7)))
```

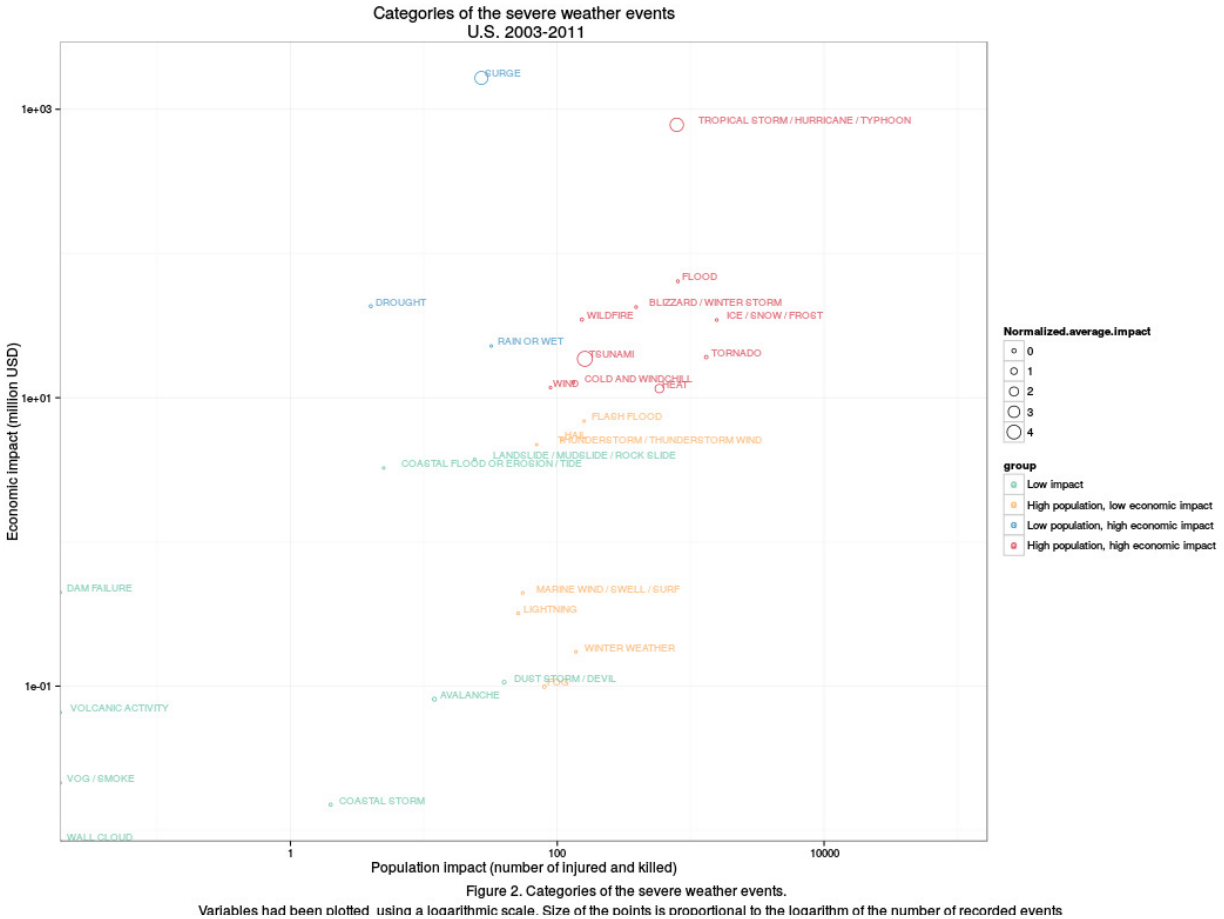


Figure 2. Categories of the severe weather events.
Variables had been plotted using a logarithmic scale. Size of the points is proportional to the logarithm of the number of recorded events