Severe Weather Events Impact Synopsis

This report looks at the impact of a number of weather events on both health and the economy in the US. A big part of the analysis is spent on cleaning up the classification for the different events. The report will show that in terms of health, tornados are clearly the most impactful, while in financial terms floods and droughts have the biggest impact.

Data Processing

Load & Subset

First the data needs to be loaded into R.

```
setwd("C:/Chris/cds/repdata/cp2")
data <- read.csv("storm_data.csv")
```

Also the required packages need to be loaded.

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Afterwards I subset the data to only include the relevant columns and only those rows which will have an impact on the result. This will exclude rows/events in which no human was harmed and no economic damage was caused.

```
data_sub <- data[, c(8, 23, 24, 25, 26, 27, 28)]
data_sub2 <- data_sub[(data_sub$PROPDMG > 0 | data_sub$CROPDMG > 0 |
                           data_sub$FATALITIES > 0 | data_sub$INJURIES > 0),]
head(data_sub2)
```

```
##    EVTYPE FATALITIES INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO          0       15    25.0          K       0
## 2 TORNADO          0        0     2.5          K       0
## 3 TORNADO          0        2    25.0          K       0
## 4 TORNADO          0        2     2.5          K       0
## 5 TORNADO          0        2     2.5          K       0
## 6 TORNADO          0        6     2.5          K       0
```

Factor Colum Cleanup

Event Type

After removing not needed rows there are still 488 different event types, many of them actually being the same.

As a first step I convert the EVTYPE column to character and back to factor in order to get rid of factor levels that are not relevant anymore (there were 985 EVTYPEs before subsetting the data). Afterwards I check my data frame and generate a table with all remaning event type.

```
data_sub2$EVTYPE <- as.character(data_sub2$EVTYPE)
data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
str(data_sub2)
```

```
## 'data.frame':    254633 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 488 levels "   HIGH SURF ADVISORY",..: 407 407 407 407 407 407 407 407 40
7 407 ...
##  $ 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 ...
```

First I apply general rules to clean up obvious mistakes. This includes extra whitespace and converting everything to lowercase.

```
data_sub2$EVTYPE <- as.character(data_sub2$EVTYPE)
data_sub2$EVTYPE <- str_trim(data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("   ", " ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("  ", " ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- tolower(data_sub2$EVTYPE)
data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
```

Next I get rid of plural forms of words, abbreviations, and words that are nearly similar in meaning. While this might lead to worse awkward names it will improve the quality of the data.

```
data_sub2$EVTYPE <- as.character(data_sub2$EVTYPE)

data_sub2$EVTYPE <- gsub("tstm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("winds", "wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("flooding", "flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("floods", "flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("cstl", "coastal", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("\\\\", "/", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("waves", "wave", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("rains", "rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("snow-squalls", "snow squalls", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("avalance", "avalanche", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("squalls", "squall", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("hvy", "heavy", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("landslides", "landslide", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("snowfall", "snow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("lighting", "lightning", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("precipitationitation", "rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("precip", "rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("slides", "slide", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("currents", "current", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("storms", "storm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thuderstorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thundeerstorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderestorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("trees", "tree", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("wins", "wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstormw", "thunderstorm wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstormwinds", "thunderstorm wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstormwind", "thunderstorm wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thundertorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunerstorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("tornadoes", "tornado", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("torndao", "tornado", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("tunderstorm", "thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("unseasonable", "unseasonably", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("fld", "flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("floodin", "flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("sml", "small", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("windchill", "wind chill", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("coastalstorm", "coastal storm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("excessive", "heavy", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("extreme", "heavy", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("ice on road", "icy roads", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("ice roads", "icy roads", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("lake effect", "lake-effect", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("lakeshore", "lake", data_sub2$EVTYPE)
```

```
data_sub2$EVTYPE <- gsub("ligntning", "lightning", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("mudslide", "mud slide", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("non thunderstorm", "non-thunderstorm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("severe", "heavy", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("small", "light", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("rainfall", "rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("shower", "rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("fires", "fire", data_sub2$EVTYPE)

data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
```

Afterwards I replace/remove special characters. The most important aspect here is how to link different event types. E.g. “rain and flood” vs. “rain/flood”. I have decided to use " / " to link multiple event types. E.g. “rain / flood”.

Also I have decided to treat event types with multiple events as separate events types, instead of attributing the damage to each single one or splitting it. The reason is that splitting would be arbitrary and counting damage twice will skew the data.

```
data_sub2$EVTYPE <- as.character(data_sub2$EVTYPE)

data_sub2$EVTYPE <- gsub("\\?", "other", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" and ", " / ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("[./-]$", "", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("/", " / ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("   ", " ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("  ", " ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" and ", " / ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" and$", "", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("\\&", "/", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" - ", " / ", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(", ", " / ", data_sub2$EVTYPE)

data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
```

Next I’m removing all names and indicators of intensity.

```
data_sub2$EVTYPE <- as.character(data_sub2$EVTYPE)

data_sub2$EVTYPE <- gsub(" 0.75| 075| 100| 125| 150| 175| 200| 275| 450| 75| damage| \\(g40\\)| 48| e
douard| emily| erin| felix| gordon| opal| accident| mishap| \\(41\\)| \\(g35\\)| \\(g40\\)| \\(g4
5\\)| / awning| / tree| 13| 40| 45| 55$| 60 mph| 63 mph| 65 mph| 65)| 65mph| 98 mph| g45| g50| g52| g
55| g58| g60| tree| windind| f0| f1| f2| f3| alberto| dean| gordon| jerry| \\(minor",
                        "", data_sub2$EVTYPE)

data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
```

As a last step I am combining different events that should be in the same group according to the data documentation, which specifies 48 unique event types.

I group everything that does not fall into one of these 48 event types into “other”.

Blizzard

```
data_sub2$EVTYPE <- gsub("blizzard / winter storm|ground blizzard",
                         "blizzard", data_sub2$EVTYPE)
```

Coastal Flood

```
data_sub2$EVTYPE <- gsub("coastal erosion|astronomical high tide|coastal flood / erosion|coastal surg
e|erosion / coastal flood|high tides|high water|tidal flood",
                         "coastal flood", data_sub2$EVTYPE)
```

Cold / Wind Chill

```
data_sub2$EVTYPE <- gsub("cold / snow|cold / wet conditions|cold temperature|cold wave|cold weather|c
ool / wet|hypothermia|hypothermia / exposure|low temperature|bitter cold|unseasonably cold|cold / win
d$|cold$",
                         "cold / wind chill", data_sub2$EVTYPE)
```

Debris Flow

```
data_sub2$EVTYPE <- gsub("landslide|mud slide urban flood|mud slide|rock slide",
                         "debris flow", data_sub2$EVTYPE)
```

Dense Fog

```
data_sub2$EVTYPE <- gsub("^fog", "dense fog", data_sub2$EVTYPE)
```

Excessive Heat

```
data_sub2$EVTYPE <- gsub("heavy heat|record / heavy heat|record heat|hyperthermia / exposure",
                         "excessive heat", data_sub2$EVTYPE)
```

Extreme Cold / Wind Chill

```
data_sub2$EVTYPE <- gsub("extended cold|heavy wind chill|record cold",
                         "extreme cold / wind chill", data_sub2$EVTYPE)
```

Flash Flood

```
data_sub2$EVTYPE <- gsub("dam break|flash flood / street|flash flood from ice jams|flash flood wind|f
lood flash|flood / flash|ice jam flood|ice jam|ice floes",
                         "flash flood", data_sub2$EVTYPE)
```

Flood

```
data_sub2$EVTYPE <- gsub("breakup flood|flood / river flood|light stream flood|major flood|minor floo
d|river / stream flood|river flood|rural flood|snowmelt flood|urban / light stream flood|urban floo
d|urban light|urban / light stream|urban / light",
                         "flood", data_sub2$EVTYPE)
```

Freezing Fog

```
data_sub2$EVTYPE <- gsub("glaze ice|glaze / ice storm|glaze|freezing dense fog",
                         "freezing fog", data_sub2$EVTYPE)
```

Frost / Freeze

```
data_sub2$EVTYPE <- gsub("black ice|damaging freeze|early frost|hard freeze|icy roads",
                         "frost / freeze", data_sub2$EVTYPE)
```

Funnel Cloud

```
data_sub2$EVTYPE <- gsub("funnel clou$",
                         "funnel cloud", data_sub2$EVTYPE)
```

Hail

```
data_sub2$EVTYPE <- gsub("hailstorm|light hail",
                         "hail", data_sub2$EVTYPE)
```

Heat

```
data_sub2$EVTYPE <- gsub("heat wave|unseasonably warm / dry|unseasonably warm|warm weather",
                         "heat", data_sub2$EVTYPE)
```

Heavy Rain

```
data_sub2$EVTYPE <- gsub("raintorm|record rainfall|torrential rainfall|unseasonal rain|torrential rai
n|record rain|^rain",
                         "heavy rain", data_sub2$EVTYPE)
```

Heavy Snow

```
data_sub2$EVTYPE <- gsub("falling snow / ice|heavy snow shower|heavy snow squall|heavy snowpack|recor
d snow|snow / blowing snow|snow accumulation|snow squall|thundersnow|^snow",
                         "heavy snow", data_sub2$EVTYPE)
```

High Surf

```
data_sub2$EVTYPE <- gsub("beach erosion|hazardous surf|heavy surf|heavy swells|high surf advisory|hig
h swells|high wave|rough surf",
                         "high surf", data_sub2$EVTYPE)
```

Hurricane / Typhoon

```
data_sub2$EVTYPE <- gsub("^typhoon",
                         "hurricane / typhoon", data_sub2$EVTYPE)
```

Ice Storm

```
data_sub2$EVTYPE <- gsub("ice$",
                         "ice storm", data_sub2$EVTYPE)
```

Hurricane / Typhoon

```
data_sub2$EVTYPE <- gsub("hurricane|hurricane-generated swells",
                         "hurricane / typhoon", data_sub2$EVTYPE)
```

Lakeshore Flood

```
data_sub2$EVTYPE <- gsub("lake flood",
                         "lakeshore flood", data_sub2$EVTYPE)
```

Lake-Effect Snow

```
data_sub2$EVTYPE <- gsub("heavy lake snow",
                         "lake-effect snow", data_sub2$EVTYPE)
```

Lightning

```
data_sub2$EVTYPE <- gsub("lightning fire|lightning wauseon|lightning injury",
                         "lightning", data_sub2$EVTYPE)
```

Storm Tide

```
data_sub2$EVTYPE <- gsub("storm surge|storm surge / tide",
                         "storm tide", data_sub2$EVTYPE)
```

Strong Wind

```
data_sub2$EVTYPE <- gsub("wind / wave|gusty wind|wind storm|^wind",
                         "strong wind", data_sub2$EVTYPE)
```

Thunderstorm Wind

```
data_sub2$EVTYPE <- gsub("downburst|dry microburst|dry mircoburst wind|gustnado|microburst|microburst
wind|wet microburst|whirlwind|thunderstorm win$|thunderstorm wi$|heavy thunderstorm wind|heavy thunde
rstorm",
                         "thunderstorm wind", data_sub2$EVTYPE)
```

Tornado

```
data_sub2$EVTYPE <- gsub("cold air tornado|landspout",
                         "tornado", data_sub2$EVTYPE)
```

Tropical Depression

```
data_sub2$EVTYPE <- gsub("gradient wind",
                         "tropical depression", data_sub2$EVTYPE)
```

Tropical Storm

```
data_sub2$EVTYPE <- gsub("coastal storm",
                         "tropical storm", data_sub2$EVTYPE)
```

Wildfire

```
data_sub2$EVTYPE <- gsub("brush fire|forest fire|grass fire|wild / forest fire|wild fire",
                         "wildfire", data_sub2$EVTYPE)
```

Winter Storm

```
data_sub2$EVTYPE <- gsub("heavy snow / strong wind|heavy snow / wind",
                         "winter storm", data_sub2$EVTYPE)
```

Winter Weather

```
data_sub2$EVTYPE <- gsub("freezing drizzle|freezing rain / sleet| freezing rain / snow|freezing rai
n|freezing spray|light snow|winter weather / mix|winter weather mix|wintry mix|light winter weather",
                         "winter weather", data_sub2$EVTYPE)
```

Other & Multi Events

```
data_sub2$EVTYPE <- gsub("agricultural freeze|apache county|blowing dust|blowing snow|heavy mix| / he
avy weather|/ light stream urban|heavy seas| / squall|heavy turbulence|heavy wetness|high seas|landsl
ump|late season snow|non-heavy wind|non-thunderstorm wind|rapidly rising water|rogue wave|rough sea
s|storm force wind|drowning|marine$|high$",
                         "other", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("dust devil waterspout",
                         "dust devil / waterspout", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("flash flood debris flow",
```

```
                        "flash flood / debris flow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("flood / flash flood|flood / flashflood",
                        "flash flood / flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("flood / rain / wind|flood and heavy rain",
                        "flood / heavy rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("lightning thunderstorm wind",
                        "lightning / thunderstorm wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("rain / snow",
                        "heavy rain / heavy snow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("rain / wind",
                        "heavy rain / strong wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("waterspout tornado|waterspout-tornado",
                        "waterspout / tornado", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("winter storm high wind",
                        "high wind / winter storm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("hail / wind",
                        "hail / strong wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heat drought",
                        "drought / heat", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("high wind heavy rain",
                        "heavy rain / high wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy snow winter weather",
                        "heavy snow / winter weather", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy snow rain",
                        "heavy snow / heavy rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("high surf coastal flood",
                        "high surf / coastal flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("high surf / wind",
                        "high surf / strong wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("ice / strong wind",
                        "ice storm / strong wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("ice / snow",
                        "heavy snow / ice storm", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("strong wind / rain",
                        "heavy rain / strong wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstorm hail|thunderstorm wind hail|thunderstorm windhail",
                        "thunderstorm wind / hail", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstorm wind lightning",
                        "thunderstorm wind / lightning", data_sub2$EVTYPE)

data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)
```

Lastly some of the repitions need to be removed.

```
data_sub2$EVTYPE <- gsub("dense dense","dense", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("wind chill / wind chill","wind chill", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("flood flood|floodflood","flood", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("hail $","hail", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy heavy","heavy", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy rain other|heavy rainitation|heavy rainother|mixed rain|mixed rainita
tion",
                         "heavy rain", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy snow / heavy snow","heavy snow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("heavy snowother","heavy snow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("high surf / high surf","high surf", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("high wind $","high wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("typhoon / typhoon","typhoon", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("  to","wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("wind $","wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("wind53","wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("thunderstrom wind|thunderstorm winds|thunderstorm$|thunderstorm to",
                         "thunderstorm wind", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" / snow"," / heavy snow", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub("^freeze|frost$","frost / freeze", data_sub2$EVTYPE)
data_sub2$EVTYPE <- gsub(" / seas","", data_sub2$EVTYPE)

data_sub2$EVTYPE <- as.factor(data_sub2$EVTYPE)

str(data_sub2)
```

```
## 'data.frame':    254633 obs. of  7 variables:
##  $ EVTYPE    : Factor w/ 103 levels "astronomical low tide",..: 92 92 92 92 92 92 92 92 92 92 ...
##  $ 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 ...
```

Results

The results are split in two parts, one looking at the impact of severe weather events on human health, the other part looking at economic consequences.

Health Impact

Impact on human health is measured by two variables in the data: One for fatalities and one for injuries. I will assess which event is more harmful based on the total number of fatalities/injuries. This gives different result than looking at the mean per event type because some events are less frequent but more harmful when they occur. However the assignment asks for impact on population health, so the total seems more fitting.

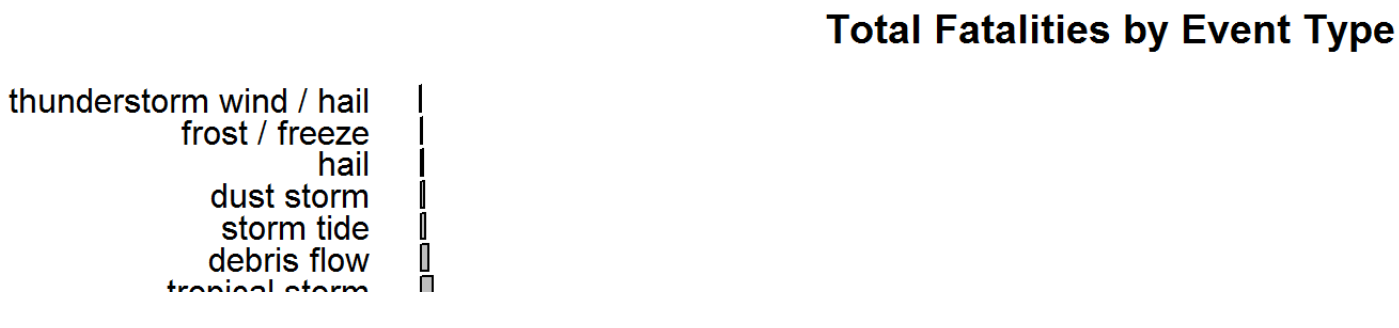
```
data_grouped <- group_by(data_sub2, EVTYPE)
summary <- summarize(data_grouped, sum_fat = sum(FATALITIES),
                     mean_fat = round(mean(FATALITIES, na.rm=T), digits = 2),
                     sum_inj = sum(INJURIES),
                     mean_inj = round(mean(INJURIES, na.rm=T), digits = 2),
                     count = n())
df <- data.frame(summary)
df_filt <- filter(df, count > 100 & (sum_inj > 40 | sum_fat > 40))
df_fat <- arrange(df_filt, desc(sum_fat))
df_inj <- arrange(df_filt, desc(sum_inj))
df_fat
```

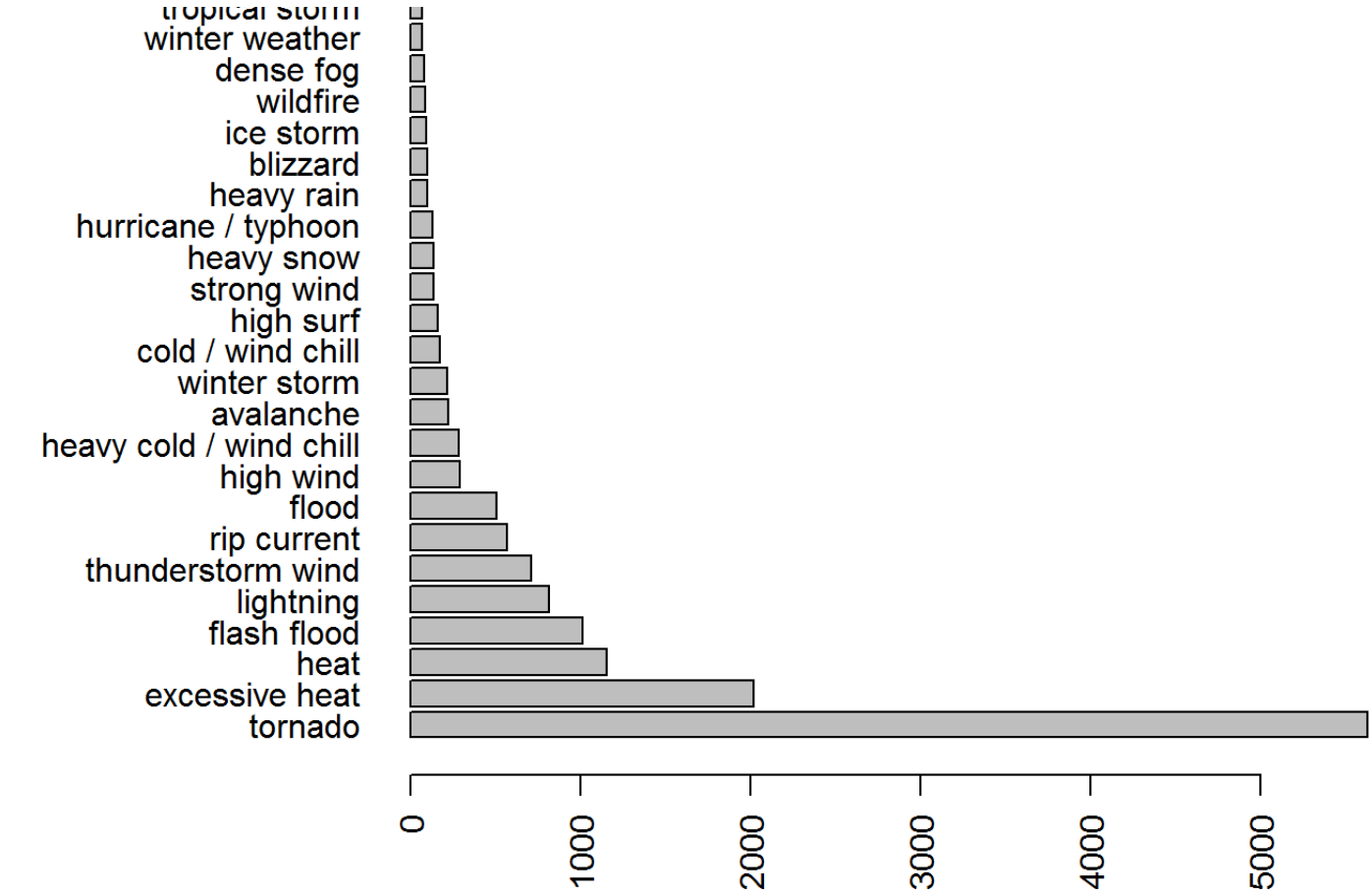
```
##                       EVTYPE sum_fat mean_fat sum_inj mean_inj  count
## 1                    tornado    5633     0.14   91364     2.29  39963
## 2             excessive heat    2019     2.81    6730     9.36    719
## 3                       heat    1154     4.44    2498     9.61    260
## 4                flash flood    1016     0.05    1800     0.08  21584
## 5                  lightning     817     0.06    5231     0.39  13299
## 6          thunderstorm wind     710     0.01    9441     0.08 119252
## 7                rip current     572     0.89     529     0.83    641
## 8                      flood     511     0.05    6873     0.61  11310
## 9                  high wind     290     0.05    1460     0.24   6188
## 10  heavy cold / wind chill     287     0.93     255     0.82    310
## 11                 avalanche     225     0.84     170     0.63    269
## 12              winter storm     216     0.14    1338     0.89   1511
## 13         cold / wind chill     175     1.11      60     0.38    158
## 14                 high surf     160     0.70     246     1.08    227
## 15               strong wind     141     0.04     400     0.11   3582
## 16                heavy snow     137     0.09    1091     0.73   1498
## 17      hurricane / typhoon     133     0.57    1333     5.75    232
## 18                heavy rain     102     0.09     306     0.26   1166
## 19                  blizzard     101     0.39     805     3.14    256
## 20                 ice storm      95     0.13    2112     2.89    732
## 21                  wildfire      90     0.07    1608     1.28   1258
## 22                 dense fog      80     0.44    1076     5.94    181
## 23            winter weather      73     0.10     655     0.89    739
## 24            tropical storm      70     0.16     385     0.90    426
## 25               debris flow      44     0.21      55     0.26    210
## 26                storm tide      24     0.11      43     0.19    224
## 27                dust storm      22     0.21     440     4.27    103
## 28                      hail      15     0.00    1371     0.05  26158
## 29            frost / freeze       9     0.05      59     0.33    177
## 30 thunderstorm wind / hail       5     0.01      96     0.19    506
```

```
df_inj
```

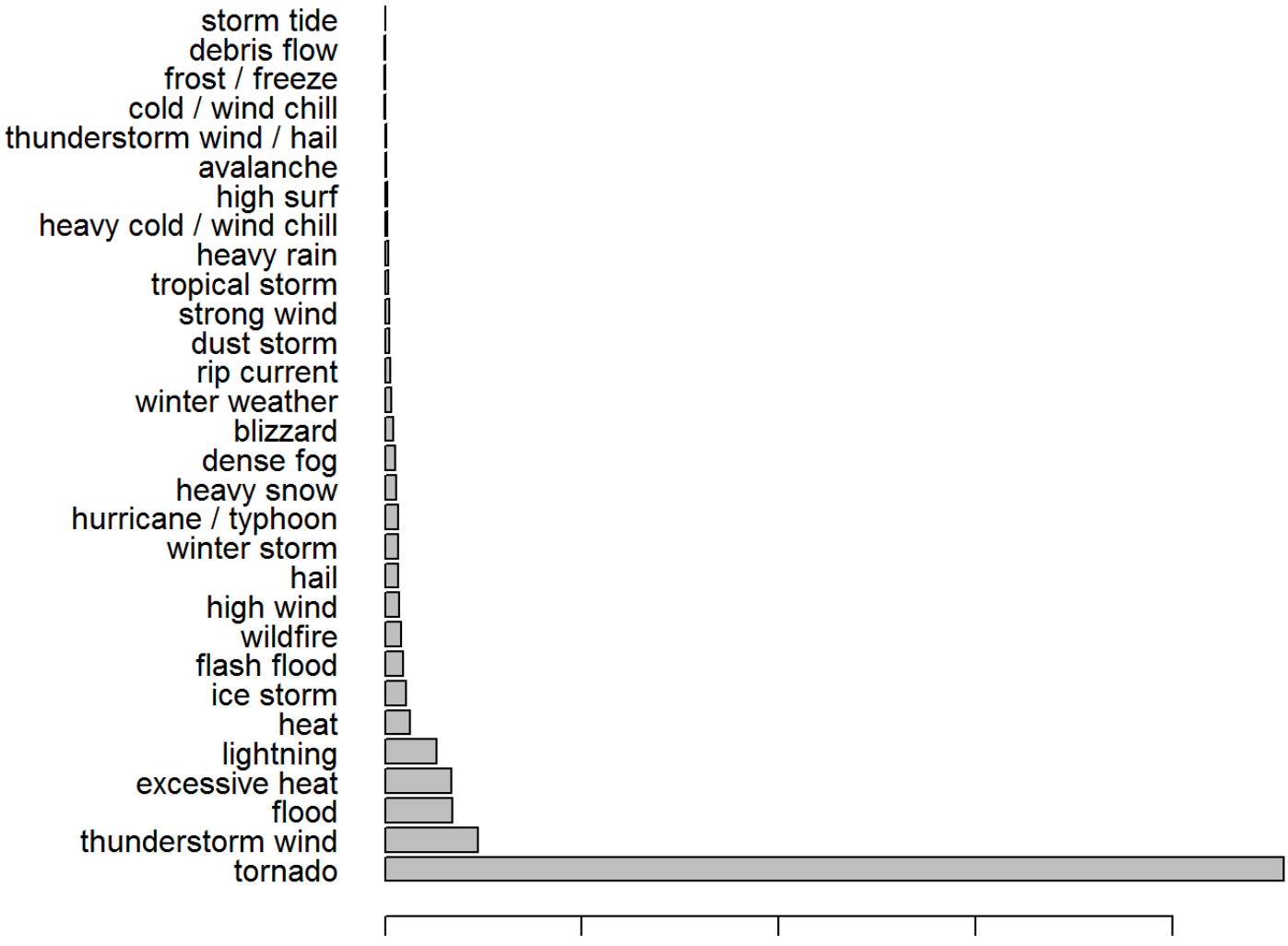
```
##                    EVTYPE sum_fat mean_fat sum_inj mean_inj  count
## 1                 tornado    5633     0.14   91364     2.29  39963
## 2       thunderstorm wind     710     0.01    9441     0.08 119252
## 3                   flood     511     0.05    6873     0.61  11310
## 4          excessive heat    2019     2.81    6730     9.36    719
## 5               lightning     817     0.06    5231     0.39  13299
## 6                    heat    1154     4.44    2498     9.61    260
## 7              ice storm      95     0.13    2112     2.89    732
## 8            flash flood    1016     0.05    1800     0.08  21584
## 9              wildfire      90     0.07    1608     1.28   1258
## 10             high wind     290     0.05    1460     0.24   6188
## 11                  hail      15     0.00    1371     0.05  26158
## 12           winter storm     216     0.14    1338     0.89   1511
## 13    hurricane / typhoon     133     0.57    1333     5.75    232
## 14             heavy snow     137     0.09    1091     0.73   1498
## 15              dense fog      80     0.44    1076     5.94    181
## 16               blizzard     101     0.39     805     3.14    256
## 17         winter weather      73     0.10     655     0.89    739
## 18            rip current     572     0.89     529     0.83    641
## 19             dust storm      22     0.21     440     4.27    103
## 20             strong wind     141     0.04     400     0.11   3582
## 21         tropical storm      70     0.16     385     0.90    426
## 22             heavy rain     102     0.09     306     0.26   1166
## 23  heavy cold / wind chill     287     0.93     255     0.82    310
## 24              high surf     160     0.70     246     1.08    227
## 25              avalanche     225     0.84     170     0.63    269
## 26 thunderstorm wind / hail       5     0.01      96     0.19    506
## 27        cold / wind chill     175     1.11      60     0.38    158
## 28         frost / freeze       9     0.05      59     0.33    177
## 29            debris flow      44     0.21      55     0.26    210
## 30             storm tide      24     0.11      43     0.19    224
```

```
par(mar = c(4,10,1,1), mfrow = c(2,1))
barplot(height = df_fat$sum_fat, names.arg = df_fat$EVTYPE, horiz = T,
        las = 2)
title("Total Fatalities by Event Type", adj=1)
barplot(height = df_inj$sum_inj, names.arg = df_inj$EVTYPE, horiz = T,
        las = 2)
title("Total Injuries by Event Type", adj=1)
```





Total Injuries by Event Type



I filtered out the least relevant event types, based on their frequency, number of fatalities, and number of injuries.

It can be seen that for both variables, tornado is the most harmful event. However when only looking at fatalities, heat and excessive heat take the second and third place, while they fall further behind when only looking at injuries. Based on injuries thunderstorm wind moves to second place.

Economical Impact

The two relevant variables for economical impact are crop damage and property damage. They both need to first be manipulated based on the PROPDMGEXP and CROPDMGEXP columns.

```
econ <- data_sub2
econ$PROPDMGEXP <- gsub("^$","1",econ$PROPDMGEXP)
econ$PROPDMGEXP <- gsub("\\-|\\+|0|2|3|4|5|6|7|8","1",econ$PROPDMGEXP)
econ$PROPDMGEXP <- gsub("h|H","100",econ$PROPDMGEXP)
econ$PROPDMGEXP <- gsub("K","1000",econ$PROPDMGEXP)
econ$PROPDMGEXP <- gsub("m|M","1000000",econ$PROPDMGEXP)
econ$PROPDMGEXP <- gsub("b|B","1000000000",econ$PROPDMGEXP)

econ$CROPDMGEXP <- gsub("^$","1",econ$CROPDMGEXP)
econ$CROPDMGEXP <- gsub("\\?|0","1",econ$CROPDMGEXP)
econ$CROPDMGEXP <- gsub("k|K","1000",econ$CROPDMGEXP)
econ$CROPDMGEXP <- gsub("m|M","1000000",econ$CROPDMGEXP)
econ$CROPDMGEXP <- gsub("b|B","1000000000",econ$CROPDMGEXP)

econ$CROPDMGEXP <- as.numeric(econ$CROPDMGEXP)
econ$PROPDMGEXP <- as.numeric(econ$PROPDMGEXP)

econ <- mutate(econ, prop = PROPDMG*PROPDMGEXP, crop = CROPDMG*CROPDMGEXP)
```

The resulting new variables prop and crop, which indicate the damage caused by each event, can be analyzed in the same fashion.

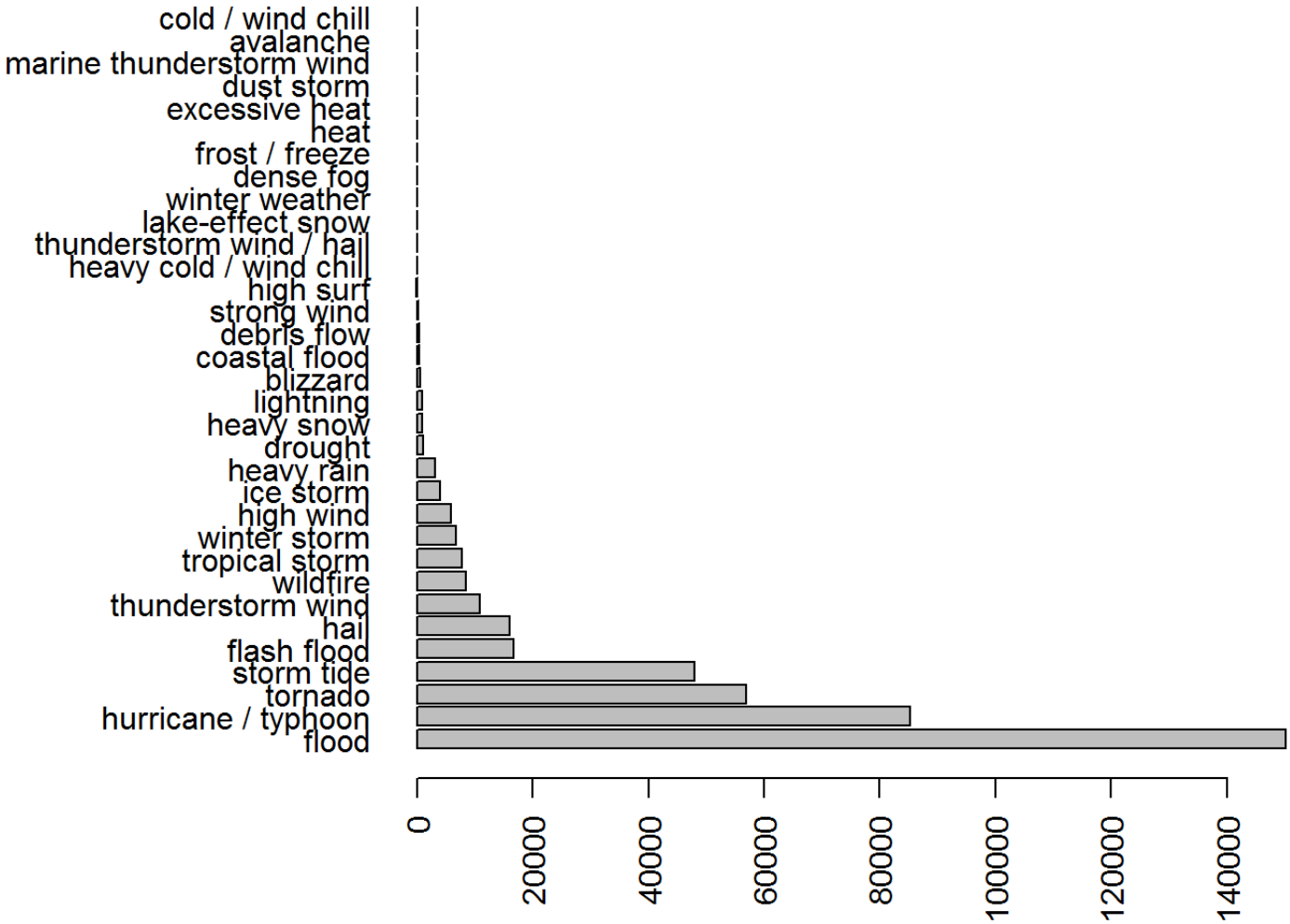
```
data_grouped <- group_by(econ, EVTYPE)
summary <- summarize(data_grouped,
                     sum_prop = round(sum(prop)/1000000, digits = 0),
                     mean_prop = round(mean(prop, na.rm=T)/1000000,digits = 2),
                     sum_crop = round(sum(crop)/1000000, digits = 0),
                     mean_crop = round(mean(crop, na.rm=T)/1000000, digits = 2),
                     count = n())
df <- data.frame(summary)
df_filt <- filter(df, count > 100 & (sum_prop > 0 | sum_crop > 0))
df_prop <- arrange(df_filt, desc(sum_prop))
df_crop <- arrange(df_filt, desc(sum_crop))
df_prop
```

```
##                     EVTYPE sum_prop mean_prop sum_crop mean_crop  count
## 1                    flood   150201     13.28    10738      0.95  11310
## 2       hurricane / typhoon    85256    367.48     5506     23.73    232
## 3                  tornado    56942      1.42      415      0.01  39963
## 4               storm tide    47965    214.13        1      0.00    224
## 5              flash flood    16640      0.77     1536      0.07  21584
## 6                     hail    15974      0.61     3047      0.12  26158
## 7        thunderstorm wind    10932      0.09     1207      0.01 119252
## 8                 wildfire     8497      6.75      403      0.32   1258
## 9           tropical storm     7714     18.11      695      1.63    426
## 10            winter storm     6691      4.43       27      0.02   1511
## 11               high wind     5880      0.95      679      0.11   6188
## 12               ice storm     3958      5.41     5022      6.86    732
## 13              heavy rain     3213      2.76      805      0.69   1166
## 14                 drought     1046      3.93    13973     52.53    266
## 15              heavy snow      958      0.64      135      0.09   1498
## 16               lightning      934      0.07       12      0.00  13299
## 17               blizzard      660      2.58      112      0.44    256
## 18           coastal flood      455      1.79        0      0.00    254
## 19             debris flow      327      1.56       20      0.10    210
## 20             strong wind      190      0.05       71      0.02   3582
## 21               high surf      102      0.45        0      0.00    227
## 22  heavy cold / wind chill       76      0.25     1313      4.24    310
## 23 thunderstorm wind / hail       45      0.09       65      0.13    506
## 24         lake-effect snow       41      0.20        0      0.00    199
## 25           winter weather       38      0.05       15      0.02    739
## 26                dense fog       23      0.13        0      0.00    181
## 27           frost / freeze       20      0.11     1968     11.12    177
## 28                     heat       12      0.05      407      1.57    260
## 29           excessive heat        8      0.01      497      0.69    719
## 30               dust storm        6      0.05        3      0.03    103
## 31 marine thunderstorm wind        6      0.04        0      0.00    142
## 32               avalanche        4      0.01        0      0.00    269
## 33        cold / wind chill        3      0.02      102      0.64    158
```

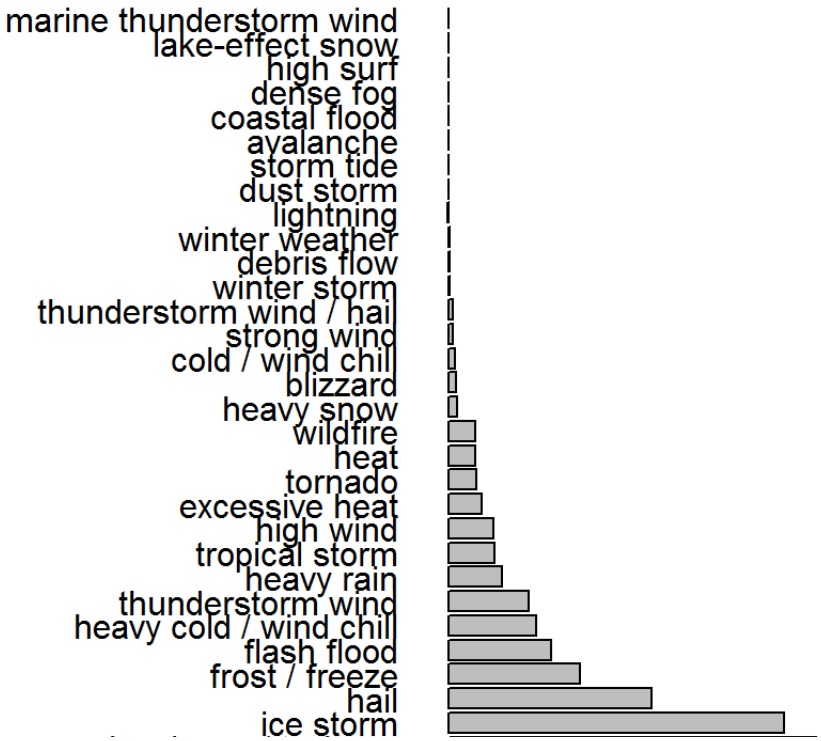
```
df_crop
```

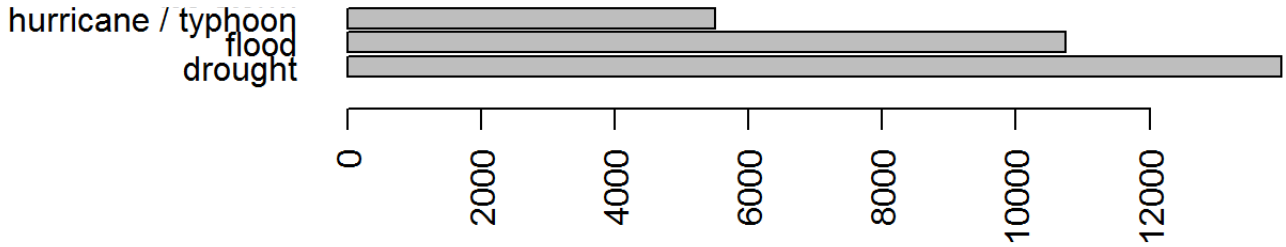
```
##                      EVTYPE sum_prop mean_prop sum_crop mean_crop  count
## 1                   drought     1046      3.93    13973     52.53    266
## 2                     flood   150201     13.28    10738      0.95  11310
## 3       hurricane / typhoon    85256    367.48     5506     23.73    232
## 4                 ice storm     3958      5.41     5022      6.86    732
## 5                      hail    15974      0.61     3047      0.12  26158
## 6            frost / freeze       20      0.11     1968     11.12    177
## 7               flash flood    16640      0.77     1536      0.07  21584
## 8   heavy cold / wind chill       76      0.25     1313      4.24    310
## 9         thunderstorm wind    10932      0.09     1207      0.01 119252
## 10               heavy rain     3213      2.76      805      0.69   1166
## 11           tropical storm     7714     18.11      695      1.63    426
## 12                high wind     5880      0.95      679      0.11   6188
## 13           excessive heat        8      0.01      497      0.69    719
## 14                  tornado    56942      1.42      415      0.01  39963
## 15                     heat       12      0.05      407      1.57    260
## 16                 wildfire     8497      6.75      403      0.32   1258
## 17               heavy snow      958      0.64      135      0.09   1498
## 18                 blizzard      660      2.58      112      0.44    256
## 19        cold / wind chill        3      0.02      102      0.64    158
## 20               strong wind      190      0.05       71      0.02   3582
## 21 thunderstorm wind / hail       45      0.09       65      0.13    506
## 22             winter storm     6691      4.43       27      0.02   1511
## 23              debris flow      327      1.56       20      0.10    210
## 24           winter weather       38      0.05       15      0.02    739
## 25                lightning      934      0.07       12      0.00  13299
## 26               dust storm        6      0.05        3      0.03    103
## 27               storm tide    47965    214.13        1      0.00    224
## 28                avalanche        4      0.01        0      0.00    269
## 29            coastal flood      455      1.79        0      0.00    254
## 30                dense fog       23      0.13        0      0.00    181
## 31                high surf      102      0.45        0      0.00    227
## 32         lake-effect snow       41      0.20        0      0.00    199
## 33 marine thunderstorm wind        6      0.04        0      0.00    142
```

```
par(mar = c(4,11,5,1), mfrow = c(2,1))
barplot(height = df_prop$sum_prop, names.arg = df_prop$EVTYPE, horiz = T,
        las = 2)
title("Total Property Damage (in m) by Event Type", adj=1)
barplot(height = df_crop$sum_crop, names.arg = df_crop$EVTYPE, horiz = T,
        las = 2)
title("Total Crop Damage (in m) by Event Type", adj=1)
```



Total Crop Damage (in m) by Event Type





Again, I filtered out the least relevant entries, based on frequency, property damage and crop damage.

When looking at property damage, it can be seen that floods have been the most expensive, followed by hurricanes/typhoons and tornados. However with regards to crop damage droughts clearly overtake floods as the most expensive event.