Most Harmful Storm Events Across the States for People and Economy

Sertinell

12 November 2015

Sipnosys

In this study we explore the impact of storm events in the life of the United States of America inhabitants. We explore both, the impact in health and in economy. To identify the most harmful events for health we compare the total number of fatalities and injuries caused by each kind of event. To identify the most harmful eventsin terms of economic impact we compare tha amount of property damage and crop damage. We found that only four type events account for almost the 80% of the fatalities and more than 85% of the injuries. Simlarly, the five most harmful type of events account for more than 97% of the property damage and more than 84% of the crop damage.

Data Processing

```
require(data.table)
```

```
## Loading required package: data.table
```

```
require(dplyr)
```

```
## Loading required package: dplyr
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:data.table':
##
##     between, last
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
require(scales)
```

```
## Loading required package: scales
```

Load data

Download the file, read it using fread through a bzcat to uncompress it into a *data.frame*.

```
filename <- "data/storms.csv.bz2"
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", destfile = fil
stormdata <-fread(sprintf("bzcat %s | tr -d '\\000'", filename))
```

```
##
Read 0.0% of 967216 rows
Read 22.7% of 967216 rows
Read 38.3% of 967216 rows
Read 50.7% of 967216 rows
Read 55.8% of 967216 rows
Read 70.3% of 967216 rows
Read 78.6% of 967216 rows
Read 84.8% of 967216 rows
Read 902297 rows and 37 (of 37) columns from 0.523 GB file in 00:00:12
```

```
## Warning in fread(sprintf("bzcat %s | tr -d '\\000'", filename)): Read less
## rows (902297) than were allocated (967216). Run again with verbose=TRUE and
## please report.
```

Clean up data

Since all the questions refer to data across the states, I have dropped out all the columns related to location.

```
notwanted <- c(1,4:7,9:11,14:20, 30:37)
stormdata <- select(stormdata, -notwanted)
```

Change the columns names to lower case and remove "_"

```
names(stormdata) <- tolower(names(stormdata))
names(stormdata) <- sub("_", "", names(stormdata))
names(stormdata)
```

```
##  [1] "bgndate"    "bgntime"    "evtype"     "enddate"    "endtime"
##  [6] "f"          "mag"        "fatalities" "injuries"   "propdmg"
## [11] "propdmgexp" "cropdmg"    "cropdmgexp" "wfo"
```

Calculate the actual damage

It seems that the *propdmgexp* and the *cropdmgexp* are the exponents of the *propdmg* and *cropdmg* respectively. We are using those colums to generate an extra column with the absolute value.

```r
table(stormdata$cropdmgexp)
```

```
## 
##                ?      0      2      B      k      K      m      M
## 618413         7     19      1      9     21 281832      1   1994
```

```r
table(stormdata$propdmgexp)
```

```
## 
##                -      ?      +      0      1      2      3      4      5
## 465934         1      8      5    216     25     13      4      4     28
##         6      7      8      B      h      H      K      m      M
##         4      5      1     40      1      6 424665      7  11330
```

Numbers are taken as exponent of scientific notation ($10*x$) and letters are taken as follows:

- B: Billions 10^12
- k|K : Kilo 10^3
- M|m : Millions 10^6
- H|h : Hecto 10^2

Other values and blanks are assigned 0.

```r
values <- data.frame(
  letters = c("B", "k", "K", "m", "M", "H", "h"),
  value =   c( 12,   3,   3,  6,   6 ,   2,  2)
  );
for(letter in values$letters){
  stormdata$cropdmgexp[stormdata$cropdmgexp == letter] <-
    values$value[values$letters == letter]
  stormdata$propdmgexp[stormdata$propdmgexp == letter] <-
    values$value[values$letters == letter]
}

stormdata$cropdmgexp <- as.integer(stormdata$cropdmgexp)
```

```
## Warning: NAs introduced by coercion
```

```r
stormdata$propdmgexp <- as.integer(stormdata$propdmgexp)
```

```
## Warning: NAs introduced by coercion
```

```r
stormdata$cropdmgexp[is.na(stormdata$cropdmgexp)] <- 0
stormdata$propdmgexp[is.na(stormdata$propdmgexp)] <- 0

stormdata <- mutate(stormdata, propertydamage = propdmg * 10^propdmgexp, cropdamage = cropdmg * 10^crop
```

Drop old columns

```
stormdata <- select(stormdata, -c(cropdmg, propdmg, cropdmgexp, propdmgexp))
```

Clean up the *evtype* column

Extract the evtype column and inspect it to decide how to clean up this messy column. Get all the unique
values and short it alphabetycally.

```
evtype <- unique(stormdata$evtype)
evtype <- evtype[order(evtype)]
evtype <- tolower(evtype)
evtype
```

```
##   [1] "?"                                "abnormally dry"
##   [3] "abnormally wet"                   "abnormal warmth"
##   [5] "accumulated snowfall"             "agricultural freeze"
##   [7] "apache county"                    "astronomical high tide"
##   [9] "astronomical low tide"            "avalance"
##  [11] "avalanche"                        "beach erosin"
##  [13] "beach erosion"                    "beach erosion"
##  [15] "beach erosion/coastal flood"      "beach flood"
##  [17] "below normal precipitation"       "bitter wind chill"
##  [19] "bitter wind chill temperatures"   "black ice"
##  [21] "black ice"                        "blizzard"
##  [23] "blizzard and extreme wind chil"   "blizzard and heavy snow"
##  [25] "blizzard/freezing rain"           "blizzard/heavy snow"
##  [27] "blizzard/high wind"               "blizzard summary"
##  [29] "blizzard weather"                 "blizzard/winter storm"
##  [31] "blowing dust"                     "blowing snow"
##  [33] "blowing snow"                     "blowing snow"
##  [35] "blowing snow & extreme wind ch"   "blowing snow- extreme wind chi"
##  [37] "blowing snow/extreme wind chil"   "blow-out tide"
##  [39] "blow-out tides"                   "breakup flooding"
##  [41] "brush fire"                       "brush fires"
##  [43] "coastal erosion"                  "coastal flood"
##  [45] "coastalflood"                     " coastal flood"
##  [47] "coastal flood"                    "coastal flooding"
##  [49] "coastal flooding"                 "coastal flooding"
##  [51] "coastal  flooding/erosion"        "coastal flooding/erosion"
##  [53] "coastal storm"                    "coastalstorm"
##  [55] "coastal storm"                    "coastal surge"
##  [57] "coastal/tidal flood"              "cold"
##  [59] "cold"                             "cold air funnel"
##  [61] "cold air funnels"                 "cold air tornado"
##  [63] "cold and frost"                   "cold and frost"
##  [65] "cold and snow"                    "cold and wet conditions"
##  [67] "cold temperature"                 "cold temperatures"
##  [69] "cold wave"                        "cold weather"
##  [71] "cold/wind chill"                  "cold wind chill temperatures"
##  [73] "cold/winds"                       "cool and wet"
##  [75] "cool spell"                       "cstl flooding/erosion"
##  [77] "damaging freeze"                  "damaging freeze"
##  [79] "dam break"                        "dam failure"
```

4

```
##  [81] "deep hail"                    "dense fog"
##  [83] "dense smoke"                   "downburst"
##  [85] "downburst winds"              "driest month"
##  [87] "drifting snow"                "drought"
##  [89] "drought/excessive heat"       "drowning"
##  [91] "dry"                          "dry conditions"
##  [93] "dry hot weather"              "dry microburst"
##  [95] "dry microburst 50"            "dry microburst 53"
##  [97] "dry microburst 58"            "dry microburst 61"
##  [99] "dry microburst 84"            "dry microburst winds"
## [101] "dry mircoburst winds"         "dryness"
## [103] "dry pattern"                  "dry spell"
## [105] "dry weather"                  "dust devel"
## [107] "dust devil"                   "dust devil"
## [109] "dust devil waterspout"        "duststorm"
## [111] "dust storm"                   "dust storm/high winds"
## [113] "early freeze"                 "early frost"
## [115] "early frost"                  "early rain"
## [117] "early snow"                   "early snowfall"
## [119] "early snowfall"               "erosion/cstl flood"
## [121] "excessive"                    "excessive cold"
## [123] "excessive heat"               "excessive heat/drought"
## [125] "excessively dry"              "excessive precipitation"
## [127] "excessive rain"               "excessive rainfall"
## [129] "excessive snow"               "excessive wetness"
## [131] "extended cold"                "extreme cold"
## [133] "extreme cold"                 "extreme cold/wind chill"
## [135] "extreme heat"                 "extremely wet"
## [137] "extreme/record cold"          "extreme windchill"
## [139] "extreme wind chill"           "extreme wind chill/blowing sno"
## [141] "extreme wind chills"          "extreme windchill temperatures"
## [143] "falling snow/ice"             "first frost"
## [145] "first snow"                   " flash flood"
## [147] "flash flood"                  "flash flood/"
## [149] "flash flood/flood"            "flash flood/ flood"
## [151] "flash flood from ice jams"    "flash flood - heavy rain"
## [153] "flash flood/heavy rain"       "flash flooding"
## [155] "flash flooding/flood"         "flash flooding/thunderstorm wi"
## [157] "flash flood/landslide"        "flash flood landslides"
## [159] "flash floods"                 "flash flood/ street"
## [161] "flash flood winds"            "flash floooding"
## [163] "flood"                        "flood"
## [165] "flood flash"                  "flood/flash"
## [167] "flood/flash flood"            "flood/flashflood"
## [169] "flood/flash flood"            "flood/flash/flood"
## [171] "flood/flash flooding"         "flood flood/flash"
## [173] "flood & heavy rain"           "flooding"
## [175] "flooding/heavy rain"          "flood/rain/wind"
## [177] "flood/rain/winds"             "flood/river flood"
## [179] "floods"                       "flood/strong wind"
## [181] "flood watch/"                 "fog"
## [183] "fog and cold temperatures"    "forest fires"
## [185] "freeze"                       "freeze"
## [187] "freezing drizzle"             "freezing drizzle"
```

```
## [189] "freezing drizzle"            "freezing drizzle and freezing"
## [191] "freezing fog"                "freezing fog"
## [193] "freezing rain"               "freezing rain"
## [195] "freezing rain"               "freezing rain and sleet"
## [197] "freezing rain and snow"      "freezing rain/sleet"
## [199] "freezing rain sleet and"     "freezing rain sleet and light"
## [201] "freezing rain/snow"          "freezing spray"
## [203] "frost"                       "frost"
## [205] "frost/freeze"                "frost/freeze"
## [207] "frost\\freeze"               "funnel"
## [209] "funnel cloud"                "funnel cloud"
## [211] "funnel cloud."               "funnel cloud/hail"
## [213] "funnel clouds"               "funnels"
## [215] "glaze"                       "glaze"
## [217] "glaze ice"                   "glaze/ice storm"
## [219] "gradient wind"               "gradient wind"
## [221] "gradient wind"               "gradient winds"
## [223] "grass fires"                 "ground blizzard"
## [225] "gustnado"                    "gustnado and"
## [227] "gusty lake wind"             "gusty thunderstorm wind"
## [229] "gusty thunderstorm winds"    "gusty wind"
## [231] "gusty wind"                  "gusty wind/hail"
## [233] "gusty wind/hvy rain"         "gusty wind/rain"
## [235] "gusty winds"                 "gusty winds"
## [237] "gusty winds"                 "hail"
## [239] "hail(0.75)"                  "hail 075"
## [241] "hail 0.75"                   "hail 088"
## [243] "hail 0.88"                   "hail 100"
## [245] "hail 1.00"                   "hail 125"
## [247] "hail 150"                    "hail 175"
## [249] "hail 1.75"                   "hail 1.75)"
## [251] "hail 200"                    "hail 225"
## [253] "hail 275"                    "hail 450"
## [255] "hail 75"                     "hail 80"
## [257] "hail 88"                     "hail aloft"
## [259] "hail damage"                 "hail flooding"
## [261] "hail/icy roads"              "hailstorm"
## [263] "hail storm"                  "hailstorms"
## [265] "hail/wind"                   "hail/winds"
## [267] "hard freeze"                 "hazardous surf"
## [269] "heat"                        "heatburst"
## [271] "heat drought"                "heat/drought"
## [273] "heat wave"                   "heat wave"
## [275] "heat wave drought"           "heat waves"
## [277] "heavy lake snow"             "heavy mix"
## [279] "heavy precipatation"         "heavy precipitation"
## [281] "heavy precipitation"         "heavy rain"
## [283] "heavy rain"                  "heavy rain"
## [285] "heavy rain and flood"        "heavy rain and wind"
## [287] "heavy rain effects"          "heavy rainfall"
## [289] "heavy rain/flooding"         "heavy rain/high surf"
## [291] "heavy rain/lightning"        "heavy rain/mudslides/flood"
## [293] "heavy rains"                 "heavy rain/severe weather"
## [295] "heavy rains/flooding"        "heavy rain/small stream urban"
```

```
## [297] "heavy rain/snow"                    "heavy rain/urban flood"
## [299] "heavy rain; urban flood winds;"     "heavy rain/wind"
## [301] "heavy seas"                          "heavy shower"
## [303] "heavy showers"                       "heavy snow"
## [305] "heavy snow and"                      "heavy snow andblowing snow"
## [307] "heavy snow and high winds"           "heavy snow and ice"
## [309] "heavy snow and ice storm"            "heavy snow and strong winds"
## [311] "heavy snow/blizzard"                 "heavy snow/blizzard/avalanche"
## [313] "heavy snow/blowing snow"             "heavy snow   freezing rain"
## [315] "heavy snow/freezing rain"            "heavy snow/high"
## [317] "heavy snow/high wind"                "heavy snow/high winds"
## [319] "heavy snow/high winds & flood"       "heavy snow/high winds/freezing"
## [321] "heavy snow & ice"                    "heavy snow/ice"
## [323] "heavy snow/ice storm"                "heavy snowpack"
## [325] "heavy snow shower"                   "heavy snow/sleet"
## [327] "heavy snow squalls"                  "heavy snow-squalls"
## [329] "heavy snow/squalls"                  "heavy snow/wind"
## [331] "heavy snow/winter storm"            "heavy surf"
## [333] "heavy surf"                          "heavy surf and wind"
## [335] "heavy surf coastal flooding"         "heavy surf/high surf"
## [337] "heavy swells"                        "heavy wet snow"
## [339] "high"                                "high seas"
## [341] "high surf"                           "high surf"
## [343] "high surf advisories"               "   high surf advisory"
## [345] "high surf advisory"                  "high swells"
## [347] "high  swells"                        "high temperature record"
## [349] "high tides"                          "high water"
## [351] "high waves"                          "highway flooding"
## [353] "high wind"                           "high wind"
## [355] "high wind 48"                        "high wind 63"
## [357] "high wind 70"                        "high wind and heavy snow"
## [359] "high wind and high tides"            "high wind and seas"
## [361] "high wind/blizzard"                  "high wind/ blizzard"
## [363] "high wind/blizzard/freezing ra"     "high wind damage"
## [365] "high wind (g40)"                     "high wind/heavy snow"
## [367] "high wind/low wind chill"            "high winds"
## [369] "high  winds"                         "high winds/"
## [371] "high winds 55"                       "high winds 57"
## [373] "high winds 58"                       "high winds 63"
## [375] "high winds 66"                       "high winds 67"
## [377] "high winds 73"                       "high winds 76"
## [379] "high winds 80"                       "high winds 82"
## [381] "high winds and wind chill"           "high winds/coastal flood"
## [383] "high winds/cold"                     "high winds dust storm"
## [385] "high wind/seas"                      "high winds/flooding"
## [387] "high winds/heavy rain"               "high winds heavy rains"
## [389] "high winds/snow"                     "high wind/wind chill"
## [391] "high wind/wind chill/blizzard"       "hot and dry"
## [393] "hot/dry pattern"                     "hot pattern"
## [395] "hot spell"                           "hot weather"
## [397] "hurricane"                           "hurricane edouard"
## [399] "hurricane emily"                     "hurricane erin"
## [401] "hurricane felix"                     "hurricane-generated swells"
## [403] "hurricane gordon"                    "hurricane opal"
```

```
## [405] "hurricane opal/high winds"       "hurricane/typhoon"
## [407] "hvy rain"                          "hyperthermia/exposure"
## [409] "hypothermia"                       "hypothermia/exposure"
## [411] "hypothermia/exposure"             "ice"
## [413] "ice and snow"                      "ice floes"
## [415] "ice fog"                           "ice jam"
## [417] "ice jam flooding"                  "ice jam flood (minor"
## [419] "ice on road"                       "ice pellets"
## [421] "ice roads"                         "ice/snow"
## [423] "ice/snow"                          "ice storm"
## [425] "ice storm and snow"                "icestorm/blizzard"
## [427] "ice storm/flash flood"            "ice/strong winds"
## [429] "icy roads"                         "icy roads"
## [431] "lack of snow"                      "lake effect snow"
## [433] "lake effect snow"                  "lake-effect snow"
## [435] "lake flood"                        "lakeshore flood"
## [437] "landslide"                         "landslides"
## [439] "landslide/urban flood"            "landslump"
## [441] "landslump"                         "landspout"
## [443] "large wall cloud"                  "late freeze"
## [445] "late season hail"                  "late season snow"
## [447] "late-season snowfall"             "late season snowfall"
## [449] "late snow"                         "light freezing rain"
## [451] "lighting"                          "lightning"
## [453] " lightning"                        "lightning."
## [455] "lightning and heavy rain"          "lightning and thunderstorm win"
## [457] "lightning and winds"               "lightning damage"
## [459] "lightning fire"                    "lightning/heavy rain"
## [461] "lightning injury"                  "lightning thunderstorm winds"
## [463] "lightning thunderstorm windss"     "lightning  wauseon"
## [465] "light snow"                        "light snow"
## [467] "light snow"                        "light snow and sleet"
## [469] "light snowfall"                    "light snow/flurries"
## [471] "light snow/freezing precip"        "ligntning"
## [473] "local flash flood"                 "local flood"
## [475] "locally heavy rain"                "low temperature"
## [477] "low temperature record"            "low wind chill"
## [479] "major flood"                       "marine accident"
## [481] "marine hail"                       "marine high wind"
## [483] "marine mishap"                     "marine strong wind"
## [485] "marine thunderstorm wind"          "marine tstm wind"
## [487] "metro storm, may 26"               "microburst"
## [489] "microburst"                        "microburst winds"
## [491] "mild and dry pattern"              "mild/dry pattern"
## [493] "mild pattern"                      "minor flood"
## [495] "minor flooding"                    "minor flooding"
## [497] "mixed precip"                      "mixed precipitation"
## [499] "mixed precipitation"               "moderate snow"
## [501] "moderate snowfall"                 "monthly precipitation"
## [503] "monthly rainfall"                  "monthly rainfall"
## [505] "monthly snowfall"                  "monthly snowfall"
## [507] "monthly temperature"               "mountain snows"
## [509] "mud/rock slide"                    "mudslide"
## [511] "mudslide"                          "mud slide"
```

```
## [513] "mudslide/landslide"           "mudslides"
## [515] "mudslides"                    "mud slides"
## [517] "mud slides urban flooding"    "near record snow"
## [519] "none"                         "non severe hail"
## [521] "non-severe wind damage"       "non tstm wind"
## [523] "non-tstm wind"                "normal precipitation"
## [525] "northern lights"              "no severe weather"
## [527] "other"                        "other"
## [529] "patchy dense fog"             "patchy ice"
## [531] "prolong cold"                 "prolong cold"
## [533] "prolong cold/snow"            "prolonged rain"
## [535] "prolong warmth"               "rain"
## [537] "rain and wind"                "rain damage"
## [539] "rain (heavy)"                 "rain/snow"
## [541] "rainstorm"                    "rain/wind"
## [543] "rapidly rising water"         "record cold"
## [545] "record cold"                  "record  cold"
## [547] "record cold and high wind"    "record cold/frost"
## [549] "record cool"                  "record dry month"
## [551] "record dryness"               "record/excessive heat"
## [553] "record/excessive rainfall"    "record heat"
## [555] "record heat"                  "record heat wave"
## [557] "record high"                  "record high"
## [559] "record high temperature"      "record high temperatures"
## [561] "record low"                   "record low rainfall"
## [563] "record may snow"              "record precipitation"
## [565] "record rainfall"              "record snow"
## [567] "record snow/cold"             "record snowfall"
## [569] "record temperature"           "record temperature"
## [571] "record temperatures"          "record temperatures"
## [573] "record warm"                  "record warm temps."
## [575] "record warmth"                "record warmth"
## [577] "record winter snow"           "red flag criteria"
## [579] "red flag fire wx"             "remnants of floyd"
## [581] "rip current"                  "rip currents"
## [583] "rip currents heavy surf"      "rip currents/heavy surf"
## [585] "river and stream flood"       "river flood"
## [587] "river flooding"               "river flooding"
## [589] "rock slide"                   "rogue wave"
## [591] "rotating wall cloud"          "rough seas"
## [593] "rough surf"                   "rural flood"
## [595] "saharan dust"                 "saharan dust"
## [597] "seasonal snowfall"            "seiche"
## [599] "severe cold"                  "severe thunderstorm"
## [601] "severe thunderstorms"         "severe thunderstorm winds"
## [603] "severe turbulence"            "sleet"
## [605] "sleet & freezing rain"        "sleet/freezing rain"
## [607] "sleet/ice storm"              "sleet/rain/snow"
## [609] "sleet/snow"                   "sleet storm"
## [611] "small hail"                   "small hail"
## [613] "small hail"                   "small stream"
## [615] "small stream and"             "small stream and urban flood"
## [617] "small stream and urban floodin" "small stream flood"
## [619] "small stream flooding"        "small stream urban flood"
```

```
## [621] "small stream/urban flood"        "sml stream fld"
## [623] "smoke"                            "snow"
## [625] "snow"                             "snow accumulation"
## [627] "snow accumulation"                "snow advisory"
## [629] "snow and cold"                    "snow and heavy snow"
## [631] "snow and ice"                     "snow and ice"
## [633] "snow and ice storm"               "snow and sleet"
## [635] "snow and sleet"                   "snow and wind"
## [637] "snow/ bitter cold"                "snow/blowing snow"
## [639] "snow/cold"                        "snow\\cold"
## [641] "snow drought"                     "snowfall record"
## [643] "snow freezing rain"               "snow/freezing rain"
## [645] "snow/heavy snow"                  "snow/high winds"
## [647] "snow- high wind- wind chill"      "snow/ice"
## [649] "snow/ ice"                        "snow/ice storm"
## [651] "snowmelt flooding"                "snow/rain"
## [653] "snow/rain/sleet"                  "snow showers"
## [655] "snow sleet"                       "snow/sleet"
## [657] "snow/sleet/freezing rain"         "snow/sleet/rain"
## [659] "snow squall"                      "snow squalls"
## [661] "snow squalls"                     "snow squalls"
## [663] "snowstorm"                        "southeast"
## [665] "storm force winds"                "storm surge"
## [667] "storm surge/tide"                 "stream flooding"
## [669] "street flood"                     "street flooding"
## [671] "strong wind"                       "strong wind"
## [673] "strong wind gust"                 "strong winds"
## [675] "strong winds"                     "strong winds"
## [677] "summary august 10"                "summary august 11"
## [679] "summary august 17"                "summary august 21"
## [681] "summary august 2-3"               "summary august 28"
## [683] "summary august 4"                 "summary august 7"
## [685] "summary august 9"                 "summary jan 17"
## [687] "summary july 23-24"               "summary june 18-19"
## [689] "summary june 5-6"                 "summary june 6"
## [691] "summary: nov. 16"                 "summary: nov. 6-7"
## [693] "summary: oct. 20-21"              "summary: october 31"
## [695] "summary of april 12"              "summary of april 13"
## [697] "summary of april 21"              "summary of april 27"
## [699] "summary of april 3rd"             "summary of august 1"
## [701] "summary of july 11"               "summary of july 2"
## [703] "summary of july 22"               "summary of july 26"
## [705] "summary of july 29"               "summary of july 3"
## [707] "summary of june 10"               "summary of june 11"
## [709] "summary of june 12"               "summary of june 13"
## [711] "summary of june 15"               "summary of june 16"
## [713] "summary of june 18"               "summary of june 23"
## [715] "summary of june 24"               "summary of june 3"
## [717] "summary of june 30"               "summary of june 4"
## [719] "summary of june 6"                "summary of march 14"
## [721] "summary of march 23"              "summary of march 24"
## [723] "summary of march 24-25"           "summary of march 27"
## [725] "summary of march 29"              "summary of may 10"
## [727] "summary of may 13"                "summary of may 14"
```

```
## [729] "summary of may 22"             "summary of may 22 am"
## [731] "summary of may 22 pm"          "summary of may 26 am"
## [733] "summary of may 26 pm"          "summary of may 31 am"
## [735] "summary of may 31 pm"          "summary of may 9-10"
## [737] "summary: sept. 18"             "summary sept. 25-26"
## [739] "summary september 20"          "summary september 23"
## [741] "summary september 3"           "summary september 4"
## [743] "temperature record"            "thuderstorm winds"
## [745] "thundeerstorm winds"           "thunderestorm winds"
## [747] "thundersnow"                   "thundersnow shower"
## [749] "thunderstorm"                  "thunderstorm damage"
## [751] "thunderstorm damage to"        "thunderstorm hail"
## [753] "thunderstorms"                 "thunderstorms wind"
## [755] "thunderstorms winds"           "thunderstormw"
## [757] "thunderstormw 50"              "thunderstorm wind"
## [759] "thunderstorm wind"             "thunderstorm wind."
## [761] "thunderstorm wind 50"          "thunderstorm wind 52"
## [763] "thunderstorm wind 56"          "thunderstorm wind 59"
## [765] "thunderstorm wind 59 mph"      "thunderstorm wind 59 mph."
## [767] "thunderstorm wind 60 mph"      "thunderstorm wind 65mph"
## [769] "thunderstorm wind 65 mph"      "thunderstorm wind 69"
## [771] "thunderstorm wind 98 mph"      "thunderstorm wind/awning"
## [773] "thunderstorm wind (g40)"       "thunderstorm wind g50"
## [775] "thunderstorm wind g51"         "thunderstorm wind g52"
## [777] "thunderstorm wind g55"         "thunderstorm wind g60"
## [779] "thunderstorm wind g61"         "thunderstorm wind/hail"
## [781] "thunderstorm wind/lightning"   "thunderstormwinds"
## [783] "thunderstorm winds"            "thunderstorm  winds"
## [785] "thunderstorm w inds"           "thunderstorm winds."
## [787] "thunderstorm winds 13"         "thunderstorm winds 2"
## [789] "thunderstorm winds 50"         "thunderstorm winds 52"
## [791] "thunderstorm winds53"          "thunderstorm winds 53"
## [793] "thunderstorm winds 60"         "thunderstorm winds 61"
## [795] "thunderstorm winds 62"         "thunderstorm winds 63 mph"
## [797] "thunderstorm winds and"        "thunderstorm winds/flash flood"
## [799] "thunderstorm winds/ flood"     "thunderstorm winds/flooding"
## [801] "thunderstorm winds funnel clou" "thunderstorm winds/funnel clou"
## [803] "thunderstorm winds g"          "thunderstorm winds g60"
## [805] "thunderstorm windshail"        "thunderstorm winds hail"
## [807] "thunderstorm winds/hail"       "thunderstorm winds/ hail"
## [809] "thunderstorm winds heavy rain" "thunderstorm winds/heavy rain"
## [811] "thunderstorm winds     le cen" "thunderstorm winds lightning"
## [813] "thunderstorm windss"           "thunderstorm winds small strea"
## [815] "thunderstorm winds urban flood" "thunderstorm wind/ tree"
## [817] "thunderstorm wind trees"       "thunderstorm wind/ trees"
## [819] "thunderstorm wins"             "thunderstormw winds"
## [821] "thunderstrom wind"             "thunderstrom winds"
## [823] "thundertorm winds"             "thundertsorm wind"
## [825] "thundestorm winds"             "thunerstorm winds"
## [827] "tidal flood"                   "tidal flooding"
## [829] "tidal flooding"                "tornado"
## [831] "tornado debris"                "tornadoes"
## [833] "tornadoes, tstm wind, hail"    "tornado f0"
## [835] "tornado f1"                    "tornado f2"
```

```
## [837] "tornado f3"                      "tornados"
## [839] "tornado/waterspout"              "torndao"
## [841] "torrential rain"                 "torrential rainfall"
## [843] "tropical depression"            "tropical storm"
## [845] "tropical storm alberto"         "tropical storm dean"
## [847] "tropical storm gordon"          "tropical storm jerry"
## [849] "tstm"                            "tstm heavy rain"
## [851] "tstmw"                           "tstm wind"
## [853] " tstm wind"                      "tstm wind"
## [855] "tstm wind 40"                    "tstm wind (41)"
## [857] "tstm wind 45"                    "tstm wind 50"
## [859] "tstm wind 51"                    "tstm wind 52"
## [861] "tstm wind 55"                    "tstm wind 65)"
## [863] "tstm wind and lightning"        "tstm wind damage"
## [865] "tstm wind (g35)"                 "tstm wind (g40)"
## [867] " tstm wind (g45)"               "tstm wind g45"
## [869] "tstm wind  (g45)"               "tstm wind (g45)"
## [871] "tstm wind g58"                   "tstm wind/hail"
## [873] "tstm winds"                      "tstm wnd"
## [875] "tsunami"                         "tunderstorm wind"
## [877] "typhoon"                         "unseasonable cold"
## [879] "unseasonably cold"              "unseasonably cool"
## [881] "unseasonably cool & wet"        "unseasonably dry"
## [883] "unseasonably hot"               "unseasonably warm"
## [885] "unseasonably warm and dry"      "unseasonably warm & wet"
## [887] "unseasonably warm/wet"          "unseasonably warm year"
## [889] "unseasonably wet"               "unseasonal low temp"
## [891] "unseasonal rain"                "unusually cold"
## [893] "unusually late snow"            "unusually warm"
## [895] "unusual/record warmth"          "unusual warmth"
## [897] "urban and small"                "urban and small stream"
## [899] "urban and small stream flood"   "urban and small stream floodin"
## [901] "urban flood"                     "urban flood"
## [903] "urban flood"                     "urban flooding"
## [905] "urban flooding"                  "urban flood landslide"
## [907] "urban floods"                    "urban small"
## [909] "urban/small"                     "urban/small flooding"
## [911] "urban/small stream"              "urban small stream flood"
## [913] "urban/small stream flood"        "urban/small stream  flood"
## [915] "urban/small stream flooding"     "urban/small strm fldg"
## [917] "urban/sml stream fld"            "urban/sml stream fldg"
## [919] "urban/street flooding"           "very dry"
## [921] "very warm"                       "vog"
## [923] "volcanic ash"                    "volcanic ash"
## [925] "volcanic ashfall"               "volcanic ash plume"
## [927] "volcanic eruption"              "wake low wind"
## [929] "wall cloud"                      "wall cloud/funnel cloud"
## [931] "warm dry conditions"            "warm weather"
## [933] "waterspout"                      " waterspout"
## [935] "water spout"                     "waterspout-"
## [937] "waterspout/"                     "waterspout funnel cloud"
## [939] "waterspouts"                     "waterspout tornado"
## [941] "waterspout-tornado"             "waterspout/tornado"
## [943] "waterspout/ tornado"            "wayterspout"
```

```
## [945] "wet micoburst"            "wet microburst"
## [947] "wet month"                "wet snow"
## [949] "wet weather"              "wet year"
## [951] "whirlwind"                "whirlwind"
## [953] "wildfire"                 "wildfires"
## [955] "wild fires"               "wild/forest fire"
## [957] "wild/forest fires"        "wind"
## [959] "wind"                     " wind"
## [961] "wind advisory"            "wind and wave"
## [963] "wind chill"               "wind chill/high wind"
## [965] "wind damage"              "wind damage"
## [967] "wind gusts"               "wind/hail"
## [969] "winds"                    "wind storm"
## [971] "winter mix"               "winter storm"
## [973] "winter storm/high wind"   "winter storm high winds"
## [975] "winter storm/high winds"  "winter storms"
## [977] "winter weather"           "winter weather"
## [979] "winter weather mix"       "winter weather/mix"
## [981] "wintery mix"              "wintry mix"
## [983] "wintry mix"               "wintry mix"
## [985] "wnd"
```

I have decided to keep only the most common *words* and consider every evtype that contains a certain word
as an event of this word type.

Split all the evtype values by words to find the most common words longer than 2 characters:

```
words <- unlist(strsplit(evtype,'[^a-zA-Z]', perl = TRUE))
words <- unique(words)
words <- words[order(words)]
words <- words[nchar(words) > 2]
```

Count the times every type appears:

```
stormdata$evtype <- tolower(stormdata$evtype)
wordtable <- lapply(words, FUN=function(x){
              sum(grepl(x, stormdata$evtype))
            })
wordtable <- data.frame(word = words, count = unlist(wordtable))
wordtable <- wordtable[order(-wordtable$count),]
head(wordtable, 20)
```

```
##               word  count
## 329            win 384628
## 330           wind 364901
## 125           hail 290401
## 295           tstm 227236
## 256          storm 124624
## 275   thunderstorm 109572
## 99           flood  82731
## 287        tornado  60700
## 16             ash  55704
## 94           flash  55677
```

13

```
## 132        heavy  27984
## 133         high  23068
## 143         inds  22989
## 332        winds  22988
## 336       winter  19604
## 243          sno  17706
## 244         snow  17705
## 160        light  15987
## 162     lightning  15776
## 170       marine  12615
```

We can see that words like *heavy*, *high* and others appear on the top of our list. Drop them because they are too vague.

```
dropwords <- c("heavy", "high", "strong", "extreme", "record", "light")
wordtable <- wordtable[-which(wordtable$word %in% dropwords),]
```

Now we need to run a loop and assign the *word* value to a new column called *cleanevent* to every event that matches each of the *words*. Every event is assigned a *cleanevent* value only once.

```
stormdata$cleanevent <- as.character("")
stormdata$cleanevent <- NA
for(event in wordtable$word)
  {
    stormdata$cleanevent[is.na(stormdata$cleanevent) &
                         grepl(event, stormdata$evtype) ] <- event
  }
```

We got *only* 85 levels and there is only 9 NA event. The NA values don't seem to be important.

```
length(table(stormdata$cleanevent))
```

```
## [1] 85
```

```
sum(is.na(stormdata$cleanevent))
```

```
## [1] 9
```

```
stormdata[is.na(stormdata$cleanevent),]
```

```
##              bgndate       bgntime      evtype           enddate
## 1:  6/11/1995 0:00:00        0000 record high
## 2:  6/11/1995 0:00:00        0000 record high
## 3:  6/12/1995 0:00:00        0000 record high
## 4:  6/13/1995 0:00:00        0000 record high
## 5:  6/11/1995 0:00:00        0000 record high
## 6: 11/16/1994 0:00:00        0800        high
## 7:   2/9/1994 0:00:00        0600           ?  2/9/1994 0:00:00
## 8:  2/22/1997 0:00:00 12:28:00 PM record high 2/22/1997 0:00:00
## 9:  2/27/1997 0:00:00 05:24:00 PM record high 2/27/1997 0:00:00
##         endtime f mag fatalities injuries wfo propertydamage cropdamage
```

```
## 1:                      0          0        0                  0        0
## 2:                      0          0        0                  0        0
## 3:                      0          0        0                  0        0
## 4:                      0          0        0                  0        0
## 5:                      0          0        0                  0        0
## 6:      1600PST          0          0        1                  0        0
## 7:      1500EST          0          0        0               5000        0
## 8: 12:28:00 PM          0          0        0 ALY               0        0
## 9: 05:24:00 PM          0          0        0 ALY               0        0
##    cleanevent
## 1:          NA
## 2:          NA
## 3:          NA
## 4:          NA
## 5:          NA
## 6:          NA
## 7:          NA
## 8:          NA
## 9:          NA
```

Which types of events are most harmful with respect to population health?

To show the impact of events in population health we need to consider the columns *injuries* and *fatalities*. Short the data according to those columns. It seems the worst recorded events were tornadoes and heatwaves.

```
stormdata <- arrange(stormdata, desc(fatalities),
                     desc(injuries))
head(stormdata)
```

```
##              bgndate    bgntime      evtype          enddate
## 1: 7/12/1995 0:00:00       1100        heat 7/16/1995 0:00:00
## 2: 5/22/2011 0:00:00 04:40:00 PM     tornado 5/22/2011 0:00:00
## 3:  6/8/1953 0:00:00       1930     tornado
## 4: 5/11/1953 0:00:00       1610     tornado
## 5: 7/28/1999 0:00:00 11:00:00 AM excessive heat 7/31/1999 0:00:00
## 6:  6/9/1953 0:00:00       1425     tornado
##        endtime f mag fatalities injuries wfo propertydamage cropdamage
## 1:     1700CST   0        583        0            0.0e+00          0
## 2: 05:00:00 PM 5   0        158     1150 SGF       2.8e+12          0
## 3:             5   0        116      785          2.5e+07          0
## 4:             5   0        114      597          2.5e+07          0
## 5: 07:00:00 AM   0         99        0 LOT         0.0e+00          0
## 6:             4   0         90     1228          2.5e+08          0
##    cleanevent
## 1:       heat
## 2:    tornado
## 3:    tornado
## 4:    tornado
## 5:       heat
## 6:    tornado
```

Now we can summarize the data to check the totals:

15

```
databyev <- group_by(stormdata, cleanevent)
databyev <- summarise_each(databyev, funs(sum), fatalities, injuries)
databyev <- arrange(databyev, desc(fatalities), desc(injuries))
head(databyev)
```

```
##   cleanevent fatalities injuries
## 1    tornado       5636    91407
## 2       heat       3138     9224
## 3        win       1731    13466
## 4      flood       1525     8602
## 5  lightning        817     5231
## 6    current        572      529
```

Check the percentage of total victims per event:

```
databyev <- mutate(databyev,
                   fatalitiespercent  = fatalities/sum(fatalities),
                   injuriespercent = injuries/sum(injuries) )
databyev <- mutate(databyev,
                   cumsumfat  = cumsum(fatalitiespercent),
                   cumsuminj = cumsum(injuriespercent) )
head(databyev, 10)
```

```
##     cleanevent fatalities injuries fatalitiespercent injuriespercent
## 1      tornado       5636    91407        0.37213602     0.650454002
## 2         heat       3138     9224        0.20719709     0.065638165
## 3          win       1731    13466        0.11429515     0.095824320
## 4        flood       1525     8602        0.10069330     0.061212000
## 5    lightning        817     5231        0.05394520     0.037223898
## 6      current        572      529        0.03776824     0.003764374
## 7    avalanche        224      170        0.01479036     0.001209723
## 8         cold        210      280        0.01386596     0.001992485
## 9        storm        206     2899        0.01360185     0.020629341
## 10         sno        162     1109        0.01069660     0.007891666
##     cumsumfat cumsuminj
## 1   0.3721360 0.6504540
## 2   0.5793331 0.7160922
## 3   0.6936283 0.8119165
## 4   0.7943216 0.8731285
## 5   0.8482668 0.9103524
## 6   0.8860350 0.9141168
## 7   0.9008254 0.9153265
## 8   0.9146913 0.9173190
## 9   0.9282932 0.9379483
## 10  0.9389898 0.9458400
```

Which types of events have the greatest economic consequences?

To show the impact on economy we inspect the columns *propertydanage* and *cropdamage*.

```
stormdata <- arrange(stormdata, desc(propertydamage),
                     desc(cropdamage))
head(stormdata)
```

```
##                bgndate     bgntime         evtype             enddate
## 1:    1/1/2006 0:00:00 12:00:00 AM          flood    1/1/2006 0:00:00
## 2:   8/29/2005 0:00:00 02:00:00 AM    storm surge   8/29/2005 0:00:00
## 3:   8/28/2005 0:00:00 11:00:00 AM hurricane/typhoon 8/29/2005 0:00:00
## 4:   8/29/2005 0:00:00 02:00:00 AM    storm surge   8/29/2005 0:00:00
## 5:  10/24/2005 0:00:00 04:00:00 AM hurricane/typhoon 10/24/2005 0:00:00
## 6:   8/28/2005 0:00:00 11:00:00 AM hurricane/typhoon 8/29/2005 0:00:00
##        endtime f mag fatalities injuries wfo propertydamage cropdamage
## 1: 07:00:00 AM   0        0        0 MTR     1.150e+14   32500000
## 2: 05:00:00 PM   0        0        0 LIX     3.130e+13          0
## 3: 05:00:00 PM   0        0        0 LIX     1.693e+13          0
## 4: 05:00:00 PM   0        0        0 LIX     1.126e+13          0
## 5: 12:00:00 PM   0        5        0 MFL     1.000e+13          0
## 6: 05:00:00 PM   0        0        0 LIX     7.350e+12          0
##    cleanevent
## 1:      flood
## 2:      storm
## 3:  hurricane
## 4:      storm
## 5:  hurricane
## 6:  hurricane
```

It seems like floods, storms and hurricanes are the most harmful events in terms of economical impact. Summarize the data to get the totals.

```
databyeveco <- group_by(stormdata, cleanevent)
databyeveco <- summarise_each(databyeveco, funs(sum), propertydamage, cropdamage)
databyeveco <- arrange(databyeveco, desc(propertydamage), desc(cropdamage))
head(databyeveco)
```

```
##   cleanevent propertydamage   cropdamage
## 1      flood   1.285397e+14 5.007267e+12
## 2  hurricane   7.431036e+13 1.513995e+12
## 3      storm   5.291793e+13 5.000739e+12
## 4        win   8.016669e+12 2.201749e+09
## 5    tornado   5.351703e+12 4.149615e+08
## 6       fire   2.545957e+12 4.032816e+08
```

Check the percentage of damage for every event:

```
databyeveco <- mutate(databyeveco,
               propertydamagepercent  = propertydamage/sum(propertydamage),
               cropdamagepercent = cropdamage/sum(cropdamage) )
databyeveco <- mutate(databyeveco,
               cumsumprop  = cumsum(propertydamagepercent),
               cumsumcrop = cumsum(cropdamagepercent) )
head(databyeveco)
```

17

```
##   cleanevent propertydamage    cropdamage propertydamagepercent
## 1      flood   1.285397e+14 5.007267e+12           0.465719562
## 2  hurricane   7.431036e+13 1.513995e+12           0.269238104
## 3      storm   5.291793e+13 5.000739e+12           0.191729980
## 4        win   8.016669e+12 2.201749e+09           0.029045651
## 5    tornado   5.351703e+12 4.149615e+08           0.019390063
## 6       fire   2.545957e+12 4.032816e+08           0.009224401
##   cropdamagepercent cumsumprop cumsumcrop
## 1      3.669539e-01  0.4657196  0.3669539
## 2      1.109520e-01  0.7349577  0.4779059
## 3      3.664755e-01  0.9266876  0.8443814
## 4      1.613535e-04  0.9557333  0.8445428
## 5      3.041015e-05  0.9751234  0.8445732
## 6      2.955420e-05  0.9843478  0.8446027
```

Results

Effects on Population Health

Four events account for almost the 80% of the fatalities and more than 85% of the injuries. Those events are:
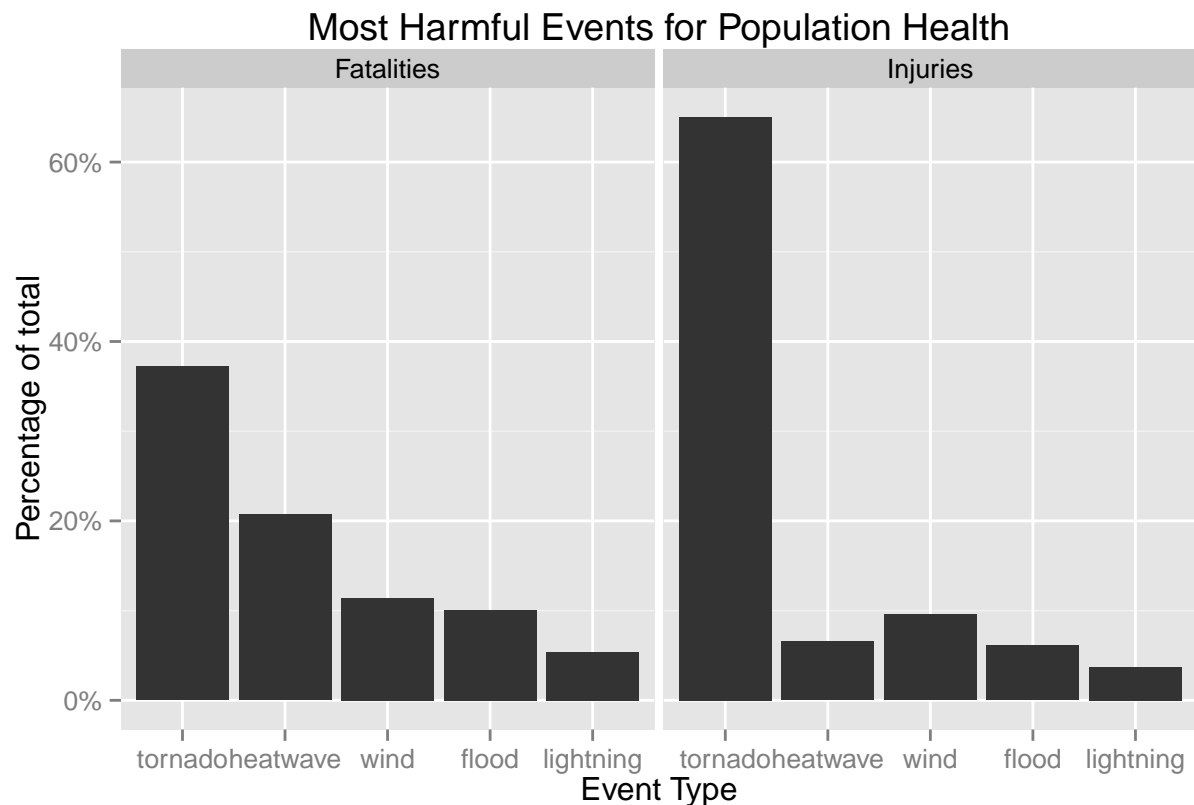
1. Tornadoes
2. Heatwaves
3. Wind
4. Flood
5. Lightning

```
head(databyev, 10)
```

```
##     cleanevent fatalities injuries fatalitiespercent injuriespercent
## 1      tornado       5636    91407        0.37213602     0.650454002
## 2         heat       3138     9224        0.20719709     0.065638165
## 3          win       1731    13466        0.11429515     0.095824320
## 4        flood       1525     8602        0.10069330     0.061212000
## 5     lightning        817     5231        0.05394520     0.037223898
## 6      current        572      529        0.03776824     0.003764374
## 7     avalanche        224      170        0.01479036     0.001209723
## 8         cold        210      280        0.01386596     0.001992485
## 9        storm        206     2899        0.01360185     0.020629341
## 10         sno        162     1109        0.01069660     0.007891666
##    cumsumfat cumsuminj
## 1  0.3721360 0.6504540
## 2  0.5793331 0.7160922
## 3  0.6936283 0.8119165
## 4  0.7943216 0.8731285
## 5  0.8482668 0.9103524
## 6  0.8860350 0.9141168
## 7  0.9008254 0.9153265
## 8  0.9146913 0.9173190
## 9  0.9282932 0.9379483
## 10 0.9389898 0.9458400
```

Plot a barplot:

```
temp <- head(databyev, 5)
temp <- select(temp, cleanevent, fatalitiespercent, injuriespercent)
temp$cleanevent <- as.factor(temp$cleanevent)
levels(temp$cleanevent)<- c("flood", "heatwave", "lightning",
                            "tornado", "wind")
temp$cleanevent <- reorder(temp$cleanevent, desc(temp$fatalitiespercent), min)
temp <- melt(temp, id.vars = "cleanevent", variable.name = "variable")
levels(temp$variable) <- c("Fatalities", "Injuries")
ggplot(temp, aes(x=cleanevent, y=value))+
  geom_bar(stat = "identity") +
  facet_grid( . ~ variable) +
  ggtitle("Most Harmful Events for Population Health")+
  xlab("Event Type") +
  ylab("Percentage of total")+
  scale_y_continuous(labels = percent)
```


Effects on Economy

The five most harmful events account for more than 97% of the property damage and more than 84% of the crop damage. Those events are:

1. Floods
2. Hurricanes
3. Storms

19

4. Wind
5. Tornadoes

```
head(databyeveco, 10)
```

```
##      cleanevent propertydamage   cropdamage propertydamagepercent
## 1        flood   1.285397e+14 5.007267e+12          4.657196e-01
## 2    hurricane   7.431036e+13 1.513995e+12          2.692381e-01
## 3        storm   5.291793e+13 5.000739e+12          1.917300e-01
## 4          win   8.016669e+12 2.201749e+09          2.904565e-02
## 5       tornado   5.351703e+12 4.149615e+08          1.939006e-02
## 6         fire   2.545957e+12 4.032816e+08          9.224401e-03
## 7         rain   2.500734e+12 8.061528e+08          9.060552e-03
## 8         hail   1.814177e+12 3.046887e+09          6.573049e-03
## 9        rough   1.046116e+09 1.512473e+12          3.790243e-06
## 10         sno   1.023255e+09 1.346631e+08          3.707413e-06
##      cropdamagepercent cumsumprop cumsumcrop
## 1        3.669539e-01  0.4657196  0.3669539
## 2        1.109520e-01  0.7349577  0.4779059
## 3        3.664755e-01  0.9266876  0.8443814
## 4        1.613535e-04  0.9557333  0.8445428
## 5        3.041015e-05  0.9751234  0.8445732
## 6        2.955420e-05  0.9843478  0.8446027
## 7        5.907831e-05  0.9934083  0.8446618
## 8        2.232889e-04  0.9999814  0.8448851
## 9        1.108404e-01  0.9999852  0.9557255
## 10       9.868686e-06  0.9999889  0.9557354
```

Plot a barplot:

```
temp <- head(databyeveco, 5)
temp <- select(temp, cleanevent, propertydamagepercent, cropdamagepercent)
temp$cleanevent <- as.factor(temp$cleanevent)
levels(temp$cleanevent)[5] <- "wind"
temp$cleanevent <- reorder(temp$cleanevent, desc(temp$propertydamagepercent), min)
temp <- melt(temp, id.vars = "cleanevent", variable.name = "variable")
levels(temp$variable) <- c("Property Damage", "Crop Damage")
ggplot(temp, aes(x=cleanevent, y=value))+
  geom_bar(stat = "identity") +
  facet_grid( . ~ variable) +
  ggtitle("Most Harmful Events for Economy")+
  xlab("Event Type") +
  ylab("Percentage of total")+
  scale_y_continuous(labels = percent)
```

Most Harmful Events for Economy