Exploring Destructive Weather using NOAA Data

Matt LeVasseur

Saturday, March 21, 2015

### Synopsis

This brief paper is the second Peer Assessment for the Coursera *Reproducible* *Research* course.

Presented is an exploratory data analysis of destructive weather since 1995, defined in terms of: -Impact on life -Economic cost

The analysis suggests that tornados and heat are the most fatal factors and surf / flooding cause the most total cost of property and cost damage.

The raw data file has data since 1950 and through November, 2011. Since earlier data files are considered incomplete, the data was filtered to 1970.

## Data Processing

Note: *No preprocessing was done outside what is documented here*

**Downloading Data and creating raw data file**

Here is how the NOAA data was downloaded and the raw file prepared for analysis, i.e., produce the raw datafile.

library(knitr)

## Warning: package 'knitr' was built under R version 3.1.2

rm(list=ls())  
setwd("C:/Users/Space Jockey 19/Desktop/DeskWD") #my working directory  
  
URL = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"  
## This URL comes from hovering over the link in the assignment  
  
## If get this error:   
## "Error in download.file(zipURL, Dataset.bz2) : unsupported URL scheme"  
  
## Then remove the comments to set the following command before download.file:  
setInternet2(use = TRUE)   
download.file(URL, destfile="NOAAdata.bz2", cacheOK = TRUE)  
  
NOAAdata.csv <- read.csv(bzfile("NOAAdata.bz2")) ## The raw .csv file

Analysis is ready to start with the raw CSV file, named "NOAAdata.csv".

However, the NOAA data set has data in it since 1950, and data collected early on is considered incomplete. Indeed, over the years since then data collection has increasingly become more powerful and therefore the data is not uniform.

My approach is to use the beginning date (column 2) as a factor to determine a year after which it appears we get consistent, modern data.

### Additional pre-processing to justify and subset to data since 1995:

First, extract the year, then summarize data by year. Then, filter the dataset after a year which appears data are consistently collected.

library(lubridate)

## Warning: package 'lubridate' was built under R version 3.1.2

NOAAdates <- strptime(NOAAdata.csv$BGN\_DATE, "%m/%d/%Y %H:%M:%S")   
# BGN\_DATE to date format (time zone irrelevant for this, only need year)  
  
## Display number of rows (observations) for each year  
## To determine uniformity of data collection  
Yearsum <- year(NOAAdates)  
  
#Output for decision to filter data  
table(Yearsum)

## Yearsum  
## 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961   
## 223 269 272 492 609 1413 1703 2184 2213 1813 1945 2246   
## 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973   
## 2389 1968 2348 2855 2388 2688 3312 2926 3215 3471 2168 4463   
## 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985   
## 5386 4975 3768 3728 3657 4279 6146 4517 7132 8322 7335 7979   
## 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997   
## 8726 7367 7257 10410 10946 12522 13534 12607 20631 27970 32270 28680   
## 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009   
## 38128 31289 34471 34962 36293 39752 39363 39184 44034 43289 55663 45817   
## 2010 2011   
## 48161 62174

summary(Yearsum) #shows Year 1995 and above = 75% of data

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 1950 1995 2002 1999 2007 2011

From this data table, see that the number of observations per year increases over time, doubling at several dates like 1952, 1955, 1973, 1982, over 10,000 in 1989, and 30,000 consistently in 1998.

This analysis uses data just since 1996 as more indictive of modern data collection and recent destructive weather patterns, when the breadth of the obs reached over 30,000.

Questions relate to personal health risks and economic damage, so only variables that relate to Event Type, Injuries, Fatalities, Property and Crop Damage will be selected.

## Select columns for Event Type, Injuries, Fatalities, Property, Crop Damage  
## I always include remarks for interpretation and year to filter out older obs  
  
NOAAdata.csv <- cbind(Yearsum, NOAAdata.csv)  
  
library(dplyr, warn=FALSE)

## Warning: package 'dplyr' was built under R version 3.1.2

NOAAdata\_tbl <- tbl\_df(NOAAdata.csv)  
NOAA96 <- select(NOAAdata\_tbl, 1,9,24:29)  
  
# Subset the data for years since 1995 inclusive  
NOAA96 <- filter(NOAA96, Yearsum > 1995)  
  
## Filter out events with no impact  
NOAA96 <- filter(NOAA96, !(FATALITIES==0 & INJURIES==0 & PROPDMG==0 & CROPDMG==0))  
  
## NOAA96 is now the data table to be used for analysis, 201,318 obs  
dim(NOAA96)

## [1] 201318 8

## Further processing for first question

The EVTYPE variable will need to be cleaned up to combine like categories that in the dataset originally sit in 985 separate levels. For this first question, we can also ignore all observations where FATALITIES and INJURIES = 0.

Then reset the factor to the reduced levels that remain, and print table as a reference to be explicit on the interpretations made for final categories:

# To get only events with injuries or fatalities:  
NOAA96health <- filter(NOAA96, !(FATALITIES==0 & INJURIES==0))  
  
## This brings the dataset down to 12,764obs  
## Originally 985 levels in EVTYPE --> now reduced to 138   
## This table shows the remaining events as recorded in the raw data

levelset <- as.character(NOAA96health$EVTYPE)  
  
#Combine all Wind not attached to other events, reduces factors to 125  
levelset1 <- gsub("WIND|STRONG WIND|GUSTY WIND|GUSTY WINDS|STRONG WINDS|HIGH WIND|MARINE STRONG WIND|Gusty winds|MARINE HIGH WIND|NON TSTM WIND|WINDS|MARINE WIND|NON-SEVERE WIND DAMAGE", "WIND", levelset, ignore.case = TRUE)  
  
#Combine all Thunderstorm related, reduces factors to 119  
levelset2 <- gsub("TSTM WIND|THUNDERSTORM WIND|TSTM WIND/HAIL|MARINE THUNDERSTORM WIND|MARINE TSTM WIND|TSTM|TSTM WIND (G40)|THUNDERSTORM WIND (G40)|TSTM WIND (G35)|TSTM WIND (G45)|THUNDERSTORMS (G35)|THUNDERSTORMS (G40)|THUNDERSTORMS (G45)|THUNDERSTORM" , "THUNDERSTORMS", levelset1, ignore.case = TRUE)  
  
#Combine all Flooding, reduces to 108 factors  
levelset3 <- gsub("FLASH FLOOD|FLOOD|COASTAL FLOOD|Coastal Flooding|COASTAL FLOODING/EROSION|RIVER FLOOD|River Flooding|HIGH WATER|TIDAL FLOODING|URBAN/SML STREAM FLD" , "FLOODING", levelset2, ignore.case = TRUE)  
   
#Combine Rip Currents and Surf related, reduces to 89 factors  
levelset4 <- gsub("RIP CURRENT|RIP CURRENTS|HIGH SURF|HEAVY SURF/HIGH SURF|HEAVY SURF|ROUGH SURF|STORM SURGE/TIDE|STORM SURGE|Heavy surf and wind|COASTAL STORM|Coastal Storm|Heavy Seas|HIGH SWELLS|ROUGH SEAS|HIGH SEAS|DROWNING|HAZARDOUS SURF|TSUNAMI|ROUGE WAVE" , "SURF", levelset3, ignore.case = TRUE)  
  
#Combine all Cold related, reduces to 76 factors  
levelset5 <- gsub("EXTREME COLD|EXTREME COLD/WIND CHILL|EXTREME WINDCHILL|Hypothermia/Exposure|Cold|COLD WEATHER|Extended Cold|Frost|HYPERTHERMIA/EXPOSURE|Cold Termperature|COLD/WIND CHILL|COLD Temperature" , "COLD", levelset4, ignore.case = TRUE)  
  
#Combine all Heat related, reduces to 71 factors  
levelset6 <- gsub("EXCESSIVE HEAT|UNSEASONABLY WARM|Heat Wave|WARM WEATHER|RECORD HEAT" , "HEAT", levelset5, ignore.case = TRUE)  
  
#Combine all Winter Weather related, reduces to 47 factors  
levelset7 <- gsub("WINTER STORM|HEAVY SNOW|ICE STORM|WINTER WEATHER/MIX|ICY ROADS|SNOW|SNOW SQUALL|WINTRY MIX|FREEZING DRIZZLE|EXCESSIVE SNOW|BLACK ICE|blowing snow|WINTER WEATHER MIX|FREEZING RAIN|Heavy snow shower|LIGHT SNOW|ICE ON ROAD|SNOW AND ICE|ICE ROADS|COLD AND SNOW|FALLING SNOW/ICE|Freezing Spray|BLIZZARD|MIXED PRECIP" , "SNOW\_ICE", levelset6, ignore.case = TRUE)  
  
#Combine all pure rain related, reduces to 44 factors  
levelset8 <- gsub("HEAVY RAIN|Torrential Rainfall|HAIL|SMALL HAIL|RAIN/SNOW|RAIN\_HAIL\_ICE", "RAIN\_HAIL" , levelset7, ignore.case = TRUE)  
  
#Combine all fire related, reduces to 42 factors  
levelset9 <- gsub("WILD/FOREST FIRE|WILDFIRE|BRUSH FIRE", "FIRE" , levelset8, ignore.case = TRUE)  
  
#Combine all hurricane related, reduces to 38 factors  
levelset10 <- gsub("HURRICANE/TYPHOON|TROPICAL STORM|TYPHOON|Hurricane Edouard", "HURRICANE" , levelset9, ignore.case = TRUE)  
  
#Combine all landslide related, reduces to 35 factors  
levelset11 <- gsub("Mudslides|LANDSLIDES|Mudslide", "LANDSLIDE" , levelset10, ignore.case = TRUE)  
  
#Combine all tornado related, reduces to 31 factors, a manageable number  
levelset12 <- gsub("Dust Devil|WATERSPOUT|FUNNEL CLOUD|Whirlwind", "TORNADO" , levelset11, ignore.case = TRUE)  
  
NOAA96health$EVTYPE <- levelset12  
NOAA96health$EVTYPE <- as.factor(NOAA96health$EVTYPE)  
  
NOAA96health <- filter(NOAA96health, FATALITIES> 1 & INJURIES>1)  
  
sort(table(NOAA96health$EVTYPE), decreasing = TRUE)

##   
## TORNADO HEAT SNOW\_ICE   
## 183 36 35   
## FLOODING WIND THUNDERSTORMS   
## 29 18 17   
## SURF FIRE FOG   
## 10 9 8   
## HURRICANE LIGHTNING COLD   
## 8 7 6   
## AVALANCHE LANDSLIDE RAIN\_HAIL   
## 3 3 3   
## DUST STORM WINTER WEATHER DENSE FOG   
## 2 2 1   
## RAIN\_HAIL\_ICE COASTALSTORM DROUGHT   
## 1 0 0   
## DRY MICROBURST GLAZE Marine Accident   
## 0 0 0   
## OTHER ROGUE WAVE SNOW\_ICEs   
## 0 0 0   
## THUNDERSTORMS (G35) THUNDERSTORMS (G40) THUNDERSTORMS (G45)   
## 0 0 0

str(NOAA96health$EVTYPE)

## Factor w/ 30 levels "AVALANCHE","COASTALSTORM",..: 28 28 28 28 28 28 29 21 29 15 ...

## I played with a clustering method, interesting not yet perfected  
## hcluster <- hclust(dist(log10(table(NOAA96health$EVTYPE))+1))  
## plot(hcluster, cex=0.8, main=NULL, xlab=NULL, ylab=NULL)

Close enough for our purposes. Now to prepare the dataset for Q2, economic impact, in a similar way.

NOAA96econ <- filter(NOAA96, !(PROPDMG==0 & CROPDMG ==0))  
# leaves 194,525

econset <- as.character(NOAA96econ$EVTYPE) #starts with 186 unique values  
  
#Combine all Wind not attached to other events, reduces factors to 157  
econset1 <- gsub("WIND|STRONG WIND|GUSTY WIND|GUSTY WINDS|STRONG WINDS|HIGH WIND|MARINE STRONG WIND|Gusty winds|MARINE HIGH WIND|NON TSTM WIND|WINDS|MARINE WIND|NON-SEVERE WIND DAMAGE|WIND Damage|WIND.\*|NON-TSTM WIND|WINDS|GRADIENT WIND|EXTREME WIND", "WINDS", econset, ignore.case = TRUE)  
  
#Combine all Thunderstorm related, reduces factors to 153  
econset2 <- gsub("TSTM WIND|THUNDERSTORM WIND|TSTM WIND/HAIL|MARINE THUNDERSTORM WIND|MARINE TSTM WIND|TSTM|.THUNDERSTORMS|TSTM WIND (G40)|THUNDERSTORM WIND (G40)|TSTM WIND (G35)|TSTM WIND (G45)|THUNDERSTORMS (G35)|THUNDERSTORMS (G40)|THUNDERSTORMS (G45)|THUNDERSTORM" , "THUNDERSTORMS", econset1, ignore.case = TRUE)  
  
#Combine all Flooding, reduces to 142 factors  
econset3 <- gsub("FLASH FLOOD|.FLOODING|FLOOD|COASTAL FLOOD|Coastal Flooding|COASTAL FLOODING/EROSION|RIVER FLOOD|River Flooding|HIGH WATER|TIDAL FLOODING|URBAN/SML STREAM FLD" , "FLOODING", econset2, ignore.case = TRUE)  
   
#Combine Rip Currents and Surf related, reduces to 129 factors  
econset4 <- gsub("RIP CURRENT|RIP CURRENTS|HIGH SURF|HEAVY SURF/HIGH SURF|HEAVY SURF|ROUGH SURF|STORM SURGE/TIDE|STORM SURGE|Heavy surf and wind|COASTAL STORM|Coastal Storm|Heavy Seas|HIGH SWELLS|ROUGH SEAS|HIGH SEAS|DROWNING|HAZARDOUS SURF|TSUNAMI|ROUGE WAVE" , "SURF", econset3, ignore.case = TRUE)  
  
#Combine all Cold related, reduces to 115 factors  
econset5 <- gsub("\*.FREEZE|HARD FREEZE|FREEZE|FREEZING FOG|DAMAGING FREEZE|EXTREME COLD|EXTREME COLD.\*|EXTREME WINDCHILL|Hypothermia/Exposure|Cold|COLD WEATHER|Extended Cold|Frost|HYPERTHERMIA/EXPOSURE|Cold Termperature|COLD/WIND CHILL|COLD Temperature|Early COLD|Unseasonable COLD" , "COLD", econset4, ignore.case = TRUE)  
  
#Combine all Heat related, reduces to 113 factors  
econset6 <- gsub("EXCESSIVE HEAT|UNSEASONABLY WARM|Heat Wave|WARM WEATHER|RECORD HEAT" , "HEAT", econset5, ignore.case = TRUE)  
  
#Combine all Winter Weather related, reduces to 91 factors  
econset7 <- gsub("LATE SEASON SNOW\_ICE|SNOW\_ICEfall|SNOW\_ICEitation|SNOW\_ICES|SNOW\_ICEs|LIGHT SNOW\_ICE|\*.SNOW\_ICE|SNOWINTER STORM|HEAVY SNOW|ICE STORM|WINTER WEATHER/MIX|ICY ROADS|SNOW|SNOW SQUALL|WINTRY MIX|FREEZING DRIZZLE|EXCESSIVE SNOW|BLACK ICE|blowing snow|WINTER WEATHER MIX|FREEZING RAIN|Heavy snow shower|LIGHT SNOW|ICE ON ROAD|SNOW AND ICE|ICE ROADS|COLD AND SNOW|FALLING SNOW/ICE|Freezing Spray|BLIZZARD|MIXED PRECIP" , "SNOW\_ICE", econset6, ignore.case = TRUE)  
  
#Combine all pure rain related, reduces to 88 factors  
econset8 <- gsub("HEAVY RAIN|Torrential Rainfall|HAIL|SMALL HAIL|RAIN/SNOW|RAIN\_HAIL\_ICE|UNSEASONAL RAIN", "RAIN\_HAIL" , econset7, ignore.case = TRUE)  
  
#Combine all fire related, reduces to 87 factors  
econset9 <- gsub("WILD/FOREST FIRE|WILDFIRE|BRUSH FIRE", "FIRE" , econset8, ignore.case = TRUE)  
  
#Combine all hurricane related, reduces to 83 factors  
econset10 <- gsub("HURRICANE/TYPHOON|TROPICAL STORM|TYPHOON|Hurricane Edouard|TROPICAL DEPRESSION", "HURRICANE" , econset9, ignore.case = TRUE)  
  
#Combine all landslide related, reduces to 81 factors  
econset11 <- gsub("Mudslides|LANDSLIDES|Mudslide", "LANDSLIDE" , econset10, ignore.case = TRUE)  
  
#Combine all tornado related, reduces to 76 factors  
econset12 <- gsub("Dust Devil|WATERSPOUT|FUNNEL CLOUD|Whirlwind", "TORNADO" , econset11, ignore.case = TRUE)  
  
NOAA96econ$EVTYPE <- econset12  
NOAA96econ$EVTYPE <- as.factor(NOAA96econ$EVTYPE)  
  
  
  
sort(table(NOAA96econ$EVTYPE), decreasing = TRUE)

##   
## THUNDERSTORMSS FLOODING RAIN\_HAIL   
## 104841 29084 23600   
## TORNADO LIGHTNING WINDS   
## 12256 8794 8689   
## SNOW\_ICE WINTER STORM FIRE   
## 2074 1347 1084   
## HURRICANE WINTER WEATHER SURF   
## 635 374 326   
## DROUGHT LAKE-EFFECT SNOW\_ICE LANDSLIDE   
## 258 194 190   
## COLD COLDCOLD DUST STORM   
## 133 117 76   
## DRY MICROBURST FOG DENSE FOG   
## 68 64 56   
## AVALANCHE OTHER HEAT   
## 52 32 30   
## LIGHT SNOW\_ICE COLD/WINDS SNOW\_ICEITATION   
## 22 16 15   
## SEICHE ASTRONOMICAL HIGH TIDE LAKESHORE FLOODING   
## 9 8 5   
## WINDSCHILL THUNDERSTORMSS AGRICULTURALCOLD   
## 4 3 3   
## LAKE EFFECT SNOW\_ICE RAIN SNOW\_ICEitation   
## 3 3 3   
## SNOW\_ICES UNSEASONABLY COLD WET MICROBURST   
## 3 3 3   
## ASTRONOMICAL LOW TIDE DAM BREAK Erosion/Cstl FLOODING   
## 2 2 2   
## Glaze LANDSPOUT MARINE RAIN\_HAIL   
## 2 2 2   
## MUD SLIDE SNOW\_ICEs TORNADOS   
## 2 2 2   
## VOLCANIC ASH WINDS (G40) SURF ADVISORY   
## 2 2 1   
## FLOODING Beach Erosion BLOWING DUST   
## 1 1 1   
## COASTAL EROSION COASTAL FLOODING/EROSION DENSE SMOKE   
## 1 1 1   
## DOWNBURST Early COLD FLOODING/FLASH/FLOODING   
## 1 1 1   
## FLOODING/FLOODING GLAZE Ice jam FLOODING (minor   
## 1 1 1   
## Lake Effect SNOW\_ICE Landslump LATE SEASON SNOW\_ICE   
## 1 1 1   
## Marine Accident Microburst Other   
## 1 1 1   
## RAIN\_HAIL/SURF ROCK SLIDE SNOW\_ICEfall   
## 1 1 1   
## WINDS/HVY RAIN WINDS/rain WINDS/RAIN\_HAIL   
## 1 1 1   
## WINDSS   
## 1

str(NOAA96econ$EVTYPE)

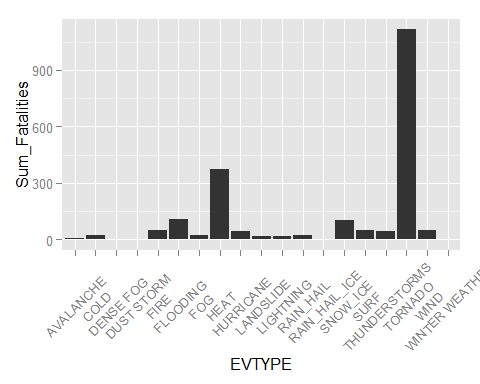
## Factor w/ 76 levels " SURF ADVISORY",..: 75 63 62 62 62 68 62 62 62 25 ...

# Results

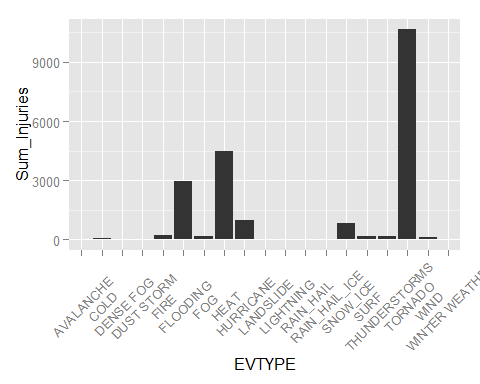
Creating graphs for analysis with respect to "Fatalities" and "Injuries," by event type.

### 1. Across the United States, which types of events (EVTYPE variable) are most harmful with respect to population health?

library(ggplot2, warn = FALSE)  
# making a table of EVTYPE vs. FATALITIES  
event\_grp <- group\_by(NOAA96health, EVTYPE)  
Health\_tbl <- summarize(event\_grp, sum(FATALITIES), sum(INJURIES))  
names(Health\_tbl) <- c("EVTYPE", "Sum\_Fatalities", "Sum\_Injuries")  
  
#plots  
  
p <- qplot(x=EVTYPE, y=Sum\_Fatalities, data=Health\_tbl, geom="bar", stat="identity") + theme (axis.text.x = element\_text(angle = 45, vjust = .5))  
  
q <- qplot(x=EVTYPE, y=Sum\_Injuries, data=Health\_tbl, geom="bar", stat="identity") + theme (axis.text.x = element\_text(angle = 45, vjust = .5))  
  
print(p)



print(q)



Comparing these graphs, Tornados are the most lethal followed by Heat, and then closely by Flooding and Snow/Ice related fatalities.

Tornados also lead the way for injuries, Heat and Flooding as well, but then rain/hail/ice jumps into the picture.

### 2. Across the United States, which types of events have the greatest economic consequences?

Creating graphs for analysis with respect to "PROPDMG" and "CROPDMG" by event type also requires accounting for the units in the PROPDMGEXP and CROPDMGEXP columns

Create a new column with the actual number multiplying in by the factor in the final analysis table.

library(dplyr)  
  
## K is thousand (1000 or 10 ^ 3)  
## M is million (1 000 000 or 10 ^ 6)  
## B is billion (1 000 000 000 or 10 ^ 9)  
  
  
# Adding a factor column to multiply property costs by  
NOAA96econ$Propfactor <- gsub("K" , "1000" , NOAA96econ$PROPDMGEXP, ignore.case=TRUE)  
NOAA96econ$Propfactor <- gsub("M" , "1000000" , NOAA96econ$Propfactor, ignore.case=TRUE)  
NOAA96econ$Propfactor <- gsub("B" , "1000000000" , NOAA96econ$Propfactor, ignore.case=TRUE)  
NOAA96econ$Propfactor <- as.numeric(NOAA96econ$Propfactor)  
  
#multiply the Property Damage by its factor  
NOAA96econ <- mutate(NOAA96econ, PropCost = PROPDMG \* Propfactor)  
  
NOAA96econ$Cropfactor <- gsub("K" , "1000" , NOAA96econ$PROPDMGEXP, ignore.case=TRUE)  
NOAA96econ$Cropfactor <- gsub("M" , "1000000" , NOAA96econ$Cropfactor, ignore.case=TRUE)  
NOAA96econ$Cropfactor <- gsub("B" , "1000000000" , NOAA96econ$Cropfactor, ignore.case=TRUE)  
NOAA96econ$Cropfactor <- as.numeric(NOAA96econ$Cropfactor)  
  
NOAA96econ <- mutate(NOAA96econ, CropCost = CROPDMG \* Cropfactor)  
  
  
  
econ\_grp <- group\_by(NOAA96econ, EVTYPE)  
Econ\_tbl <- summarize(econ\_grp, sum(PropCost), sum(CropCost))  
  
names(Econ\_tbl) <- c("EVTYPE", "Sum\_Property", "Sum\_Crop")  
  
Econ\_tbl

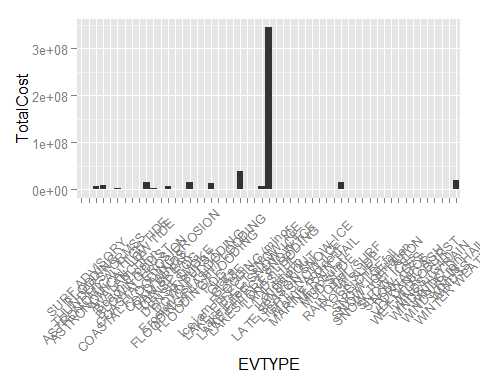
## Source: local data frame [76 x 3]  
##   
## EVTYPE Sum\_Property Sum\_Crop  
## 1 SURF ADVISORY 200000 0  
## 2 FLOODING 50000 0  
## 3 THUNDERSTORMSS 8108000 0  
## 4 AGRICULTURALCOLD NA NA  
## 5 ASTRONOMICAL HIGH TIDE 9425000 0  
## 6 ASTRONOMICAL LOW TIDE 320000 0  
## 7 AVALANCHE 3711800 0  
## 8 Beach Erosion 100000 0  
## 9 BLOWING DUST 20000 0  
## 10 COASTAL EROSION 766000 0  
## 11 COASTAL FLOODING/EROSION 15000000 0  
## 12 COLD NA NA  
## 13 COLD/WINDS 1990000 600000  
## 14 COLDCOLD NA NA  
## 15 DAM BREAK 1002000 0  
## 16 DENSE FOG 7319000 0  
## 17 DENSE SMOKE 100000 0  
## 18 DOWNBURST 2000 0  
## 19 DROUGHT NA NA  
## 20 DRY MICROBURST NA NA  
## 21 DUST STORM NA NA  
## 22 Early COLD NA NA  
## 23 Erosion/Cstl FLOODING 16200000 0  
## 24 FIRE NA NA  
## 25 FLOODING NA NA  
## 26 FLOODING/FLASH/FLOODING 10000 0  
## 27 FLOODING/FLOODING 5000 0  
## 28 FOG 13145500 0  
## 29 Glaze 90000 0  
## 30 GLAZE 60000 0  
## 31 HEAT NA NA  
## 32 HURRICANE NA NA  
## 33 Ice jam FLOODING (minor 1000 0  
## 34 LAKE-EFFECT SNOW\_ICE 40115000 0  
## 35 Lake Effect SNOW\_ICE 50000 0  
## 36 LAKE EFFECT SNOW\_ICE 17000 0  
## 37 LAKESHORE FLOODING 7540000 0  
## 38 LANDSLIDE 325808000 20017000  
## 39 Landslump 570000 0  
## 40 LANDSPOUT 7000 0  
## 41 LATE SEASON SNOW\_ICE 180000 0  
## 42 LIGHT SNOW\_ICE 451000 0  
## 43 LIGHTNING NA NA  
## 44 Marine Accident 50000 0  
## 45 MARINE RAIN\_HAIL 4000 0  
## 46 Microburst 20000 0  
## 47 MUD SLIDE 100100 0  
## 48 Other 50000 0  
## 49 OTHER NA NA  
## 50 RAIN NA NA  
## 51 RAIN\_HAIL NA NA  
## 52 RAIN\_HAIL/SURF 13500000 1500000  
## 53 ROCK SLIDE 150000 0  
## 54 SEICHE 980000 0  
## 55 SNOW\_ICE NA NA  
## 56 SNOW\_ICEfall 85000 0  
## 57 SNOW\_ICEitation 235000 0  
## 58 SNOW\_ICEITATION 555000 0  
## 59 SNOW\_ICEs 55000 0  
## 60 SNOW\_ICES 15000 0  
## 61 SURF 48074193500 25850000  
## 62 THUNDERSTORMSS NA NA  
## 63 TORNADO NA NA  
## 64 TORNADOS 12000 0  
## 65 UNSEASONABLY COLD NA NA  
## 66 VOLCANIC ASH 500000 0  
## 67 WET MICROBURST 35000 0  
## 68 WINDS NA NA  
## 69 WINDS (G40) 18000 0  
## 70 WINDS/HVY RAIN 2000 0  
## 71 WINDS/rain 2000 0  
## 72 WINDS/RAIN\_HAIL 20000 0  
## 73 WINDSCHILL NA NA  
## 74 WINDSS 500000 0  
## 75 WINTER STORM NA NA  
## 76 WINTER WEATHER 20866000 15000

NOAA96econ <- mutate(NOAA96econ, TotalCost = CropCost+PropCost)  
  
total\_grp <- group\_by(NOAA96econ, EVTYPE)  
Total\_tbl <- summarize(total\_grp, sum(TotalCost))  
  
names(Total\_tbl) <- c("EVTYPE", "TotalCost")  
  
Total\_tbl

## Source: local data frame [76 x 2]  
##   
## EVTYPE TotalCost  
## 1 SURF ADVISORY 200000  
## 2 FLOODING 50000  
## 3 THUNDERSTORMSS 8108000  
## 4 AGRICULTURALCOLD NA  
## 5 ASTRONOMICAL HIGH TIDE 9425000  
## 6 ASTRONOMICAL LOW TIDE 320000  
## 7 AVALANCHE 3711800  
## 8 Beach Erosion 100000  
## 9 BLOWING DUST 20000  
## 10 COASTAL EROSION 766000  
## 11 COASTAL FLOODING/EROSION 15000000  
## 12 COLD NA  
## 13 COLD/WINDS 2590000  
## 14 COLDCOLD NA  
## 15 DAM BREAK 1002000  
## 16 DENSE FOG 7319000  
## 17 DENSE SMOKE 100000  
## 18 DOWNBURST 2000  
## 19 DROUGHT NA  
## 20 DRY MICROBURST NA  
## 21 DUST STORM NA  
## 22 Early COLD NA  
## 23 Erosion/Cstl FLOODING 16200000  
## 24 FIRE NA  
## 25 FLOODING NA  
## 26 FLOODING/FLASH/FLOODING 10000  
## 27 FLOODING/FLOODING 5000  
## 28 FOG 13145500  
## 29 Glaze 90000  
## 30 GLAZE 60000  
## 31 HEAT NA  
## 32 HURRICANE NA  
## 33 Ice jam FLOODING (minor 1000  
## 34 LAKE-EFFECT SNOW\_ICE 40115000  
## 35 Lake Effect SNOW\_ICE 50000  
## 36 LAKE EFFECT SNOW\_ICE 17000  
## 37 LAKESHORE FLOODING 7540000  
## 38 LANDSLIDE 345825000  
## 39 Landslump 570000  
## 40 LANDSPOUT 7000  
## 41 LATE SEASON SNOW\_ICE 180000  
## 42 LIGHT SNOW\_ICE 451000  
## 43 LIGHTNING NA  
## 44 Marine Accident 50000  
## 45 MARINE RAIN\_HAIL 4000  
## 46 Microburst 20000  
## 47 MUD SLIDE 100100  
## 48 Other 50000  
## 49 OTHER NA  
## 50 RAIN NA  
## 51 RAIN\_HAIL NA  
## 52 RAIN\_HAIL/SURF 15000000  
## 53 ROCK SLIDE 150000  
## 54 SEICHE 980000  
## 55 SNOW\_ICE NA  
## 56 SNOW\_ICEfall 85000  
## 57 SNOW\_ICEitation 235000  
## 58 SNOW\_ICEITATION 555000  
## 59 SNOW\_ICEs 55000  
## 60 SNOW\_ICES 15000  
## 61 SURF 48100043500  
## 62 THUNDERSTORMSS NA  
## 63 TORNADO NA  
## 64 TORNADOS 12000  
## 65 UNSEASONABLY COLD NA  
## 66 VOLCANIC ASH 500000  
## 67 WET MICROBURST 35000  
## 68 WINDS NA  
## 69 WINDS (G40) 18000  
## 70 WINDS/HVY RAIN 2000  
## 71 WINDS/rain 2000  
## 72 WINDS/RAIN\_HAIL 20000  
## 73 WINDSCHILL NA  
## 74 WINDSS 500000  
## 75 WINTER STORM NA  
## 76 WINTER WEATHER 20881000

Notice that SURF contains data orders of magnitude above. From the discussion boards, we have discovered an error in the data. Conceding this is the largest one, removing it to get a better look at the others, and plot the total economic impact.

## Remove NA's  
  
badobs <- is.na(Total\_tbl$TotalCost)  
NoNa <- Total\_tbl[!badobs, ] #subset of df, with only not = NA  
  
NoSurf1 <- NoNa[1:44,]  
NoSurf2 <- NoNa[46:54,]  
  
NoSurf <- rbind(NoSurf1, NoSurf2)  
  
T <- qplot(x=EVTYPE, y=TotalCost, data=NoSurf, geom="bar", stat="identity") + theme (axis.text.x = element\_text(angle = 45, vjust = .5))  
  
print(T)



Still working on the axes, but flooding is a serious category along with related items like tides, but several significant other factors like landslides were unexpected.