Health and Economic Risks of US Weather

K. Huisamen

5 June 2016

Synopsis

Good weather observations feedback can support future decisons in allocating critical resources to minimise injuries and fatalities. For example droughts have significant impact but is not a major health risk, but heat is a major health risk with a much lower economic impact. By having this information available proper resources can be allocated to ensure the health and safety of the inhabitants.

Data Processing

This weather data from 1950 to November 2011, contains historic observations on the health and economic impact of the various weather events. In order to use this information some processing is required. The events in the data was cleaned by checking for common mistakes and categorisation errors to align it with the standard weather circular.[1] First the costing was derived by making some assumptions about the missing information. Where cost unit of measure was missing, or non-standard reference used it is assumed to be in the thousands. Numbers in the unit of measure column was assumed that it indicated the number of zero's and allocated as such to the various applicable unit of measure. Where a observation referenced multiple event types additional observation was registered for each event type.

Download raw data from the internet.

```
# Analysis start from the raw CSV file containing the data
# if exist then
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
zipfile <- "~/40 L&G/Coursera/ReproducibleResearch/Wk4_Project2/repdata-data-StormData.csv.bz2"

if (!exists("dat", envir = .GlobalEnv)){
        download.file(url, zipfile)
        datedownload <- date()
        dat <- read.csv(zipfile, strip.white = TRUE,
                        na.strings=c("","  ","NA"))#Blanks are real sods to catch
}
```

Tidy Economic Cost

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
```

[1] NATIONAL WEATHER SERVICE INSTRUCTION 10-1605

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
dat2 <- dat %>% select(EVTYPE, PROPDMGEXP, CROPDMGEXP, PROPDMG, CROPDMG, FATALITIES, INJURIES) %>%
        mutate(ID = seq_along(EVTYPE))

# Substitute H
DMG <- c("H", "h")
datHprop <- dat2 %>%
        filter(PROPDMGEXP %in% DMG) %>%
                mutate(PROPDMGEXPtidy = "H", DMGUOM = 100, PropCost = PROPDMG*100)
# table(datHprop$PROPDMGEXPtidy)

# Substitute K
DMG <- c("K", "k","-", "?", "+", "0", "1", "2", "3", "4", "5")
datKprop <- dat2 %>%
        filter(PROPDMGEXP %in% DMG|is.na(PROPDMGEXP)) %>%
                mutate(PROPDMGEXPtidy = "K", DMGUOM = 1000, PropCost = PROPDMG*1000)
# table(datKprop$PROPDMGEXPtidy)

datKcrop <- dat2 %>%
        filter(CROPDMGEXP %in% DMG|is.na(CROPDMGEXP)) %>%
                mutate(CROPDMGEXPtidy = "K", DMGUOM = 1000, CropCost = CROPDMG*1000)
# table(datKcrop$CROPDMGEXPtidy)

# Substitute M
DMG <- c("M", "m", "6", "7", "8")
datMprop <- dat2 %>%
        filter(PROPDMGEXP %in% DMG) %>%
                mutate(PROPDMGEXPtidy = "M", DMGUOM = 1000000, PropCost = PROPDMG*1000000)
# table(datMprop$PROPDMGEXPtidy)

datMcrop <- dat2 %>%
        filter(CROPDMGEXP %in% DMG) %>%
                mutate(CROPDMGEXPtidy = "M", DMGUOM = 1000000, CropCost = CROPDMG*1000000)
# table(datMcrop$CROPDMGEXPtidy)

# Substitute B
DMG <- c("B", "b")
datBprop <- dat2 %>%
        filter(PROPDMGEXP %in% DMG) %>%
                mutate(PROPDMGEXPtidy = "B", DMGUOM = 1000000000, PropCost = PROPDMG*1000000000)
# table(datBprop$PROPDMGEXPtidy)

datBcrop <- dat2 %>%
        filter(CROPDMGEXP %in% DMG) %>%
                mutate(CROPDMGEXPtidy = "B", DMGUOM = 1000000000, CropCost = CROPDMG*1000000000)
# table(datBcrop$CROPDMGEXPtidy)
```

```
# Combine rows
datcrop <- rbind(datKcrop, datMcrop, datBcrop)
# table(datcrop$CROPDMGEXPtidy)

datprop <- rbind(datHprop, datKprop, datMprop, datBprop)
# table(datprop$PROPDMGEXPtidy)

# Combine Columns
dat3 <- merge(datprop, datcrop, by.x = "ID", by.y = "ID")

dat4 <- dat3 %>% mutate(ID, EVENT = toupper(EVTYPE.x),
                        COST = PropCost+CropCost,
                        FATALITIES = FATALITIES.x,
                        INJURIES = INJURIES.x) %>%
        select(ID, EVENT, COST, FATALITIES,INJURIES)
```

At this point the costing is complete and events are addressed.

Tidy Event Types

Event types are consolidated into the standard event list and a total risk value is assigned based on the number of fatalities and injuries. This is also where the event summary is compiled in terms of risk and cost. Five risk levels have been incorporated if goverment agencies wish to implement a standard operating procedure system.

```
# 48 Standard Events
ET <- list("Astronomical Low Tide" = "Astronomical Low Tide",
        "Avalanche" = "Avalanche|Avalance",
        "Blizzard" = "Blizzard",
        "Coastal Flood" = "Coastal Flood|Tidal Flood",
        "Cold/Wind Chill"= "Cold/Wind Chill|Cold",
        "Debris Flow" = "Debris Flow",
        "Dense Fog" = "Dense Fog|Fog",
        "Dense Smoke" = "Dense Smoke",
        "Drought" = "Drought",
        "Dust Devil" = "Dust Devil",
        "Dust Storm" = "Dust Storm",
        "Excessive Heat" = "Excessive Heat|Extreme Heat|Heat Wave|Record Heat",
        "Extreme Cold/Wind Chill" = "Extreme Cold/Wind Chill|Cold|Wind Chill",
        "Flash Flood" = "Flash Flood|Flash",
        "Flood" = "^Flood|Flooding|River Flood|Erosion",
        "Freezing Fog" = "Freezing Fog",
        "Frost/Freeze" = "Frost|Freez",
        "Funnel Cloud" = "Funnel Cloud",
        "Hail" = "Hail",
        "Heat" = "^Heat",
        "Heavy Rain" = "Heavy Rain",
        "Heavy Snow" = "Heavy Snow|Snow",
        "High Surf" = "High Surf|Surf",
        "High Wind" = "High Wind",
        "Hurricane/Typhoon" = "Hurricane|Typhoon",
        "Ice Storm" = "Ice Storm",
```

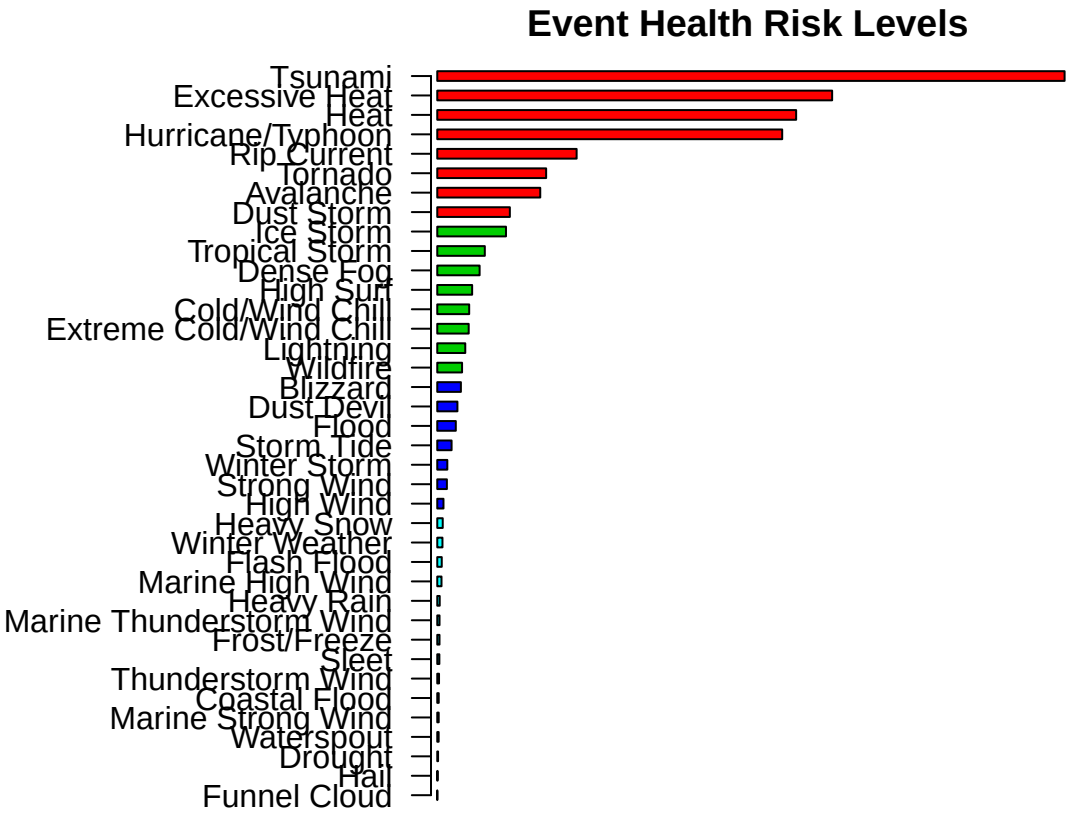
```
        "Lakeshore Flood" = "Lakeshore Flood",
        "Lake-Effect Snow" = "Lake-Effect Snow",
        "Lightning" = "Lightning",
        "Marine Hail" = "Marine Hail",
        "Marine High Wind" = "Marine High Wind|Heavy Seas",
        "Marine Strong Wind" = "Marine High Wind",
        "Marine Thunderstorm Wind" = "Marine Thunderstorm Wind|TSTM|MARINE MISHAP|Coastalstorm|Coastal S
        "Rip Current" = "Rip Current",
        "Seiche" = "Seiche",
        "Sleet" = "Sleet",
        "Storm Tide" = "Storm Tide|Storm Surge",
        "Strong Wind" = "Strong Wind",
        "Thunderstorm Wind" = "^Thunder",
        "Tornado" = "Tornado",
        "Tropical Depression" = "Tropical Depression",
        "Tropical Storm" = "Tropical Storm",
        "Tsunami" = "Tsunami",
        "Volcanic Ash" = "Volcanic Ash",
        "Waterspout" = "Waterspout",
        "Wildfire" = "Wildfire|Forest",
        "Winter Storm" = "Winter Storm",
        "Winter Weather" = "Winter Weather")

dat2 <- data.frame(ID = NULL, EVENT = NULL, COST = NULL, FATALITIES = NULL, INJURIES = NULL)
for (i in 1:length(ET)){
        searchstring <- toupper(ET[[i]])
        tidyeventlist <- dat4 %>%
                filter(grepl(searchstring, EVENT)) %>%
                mutate(EVENT = names(ET[i]))
        dat2 <- rbind(dat2, tidyeventlist)
}

rm(tidyeventlist)
Event.Types <- unique(dat2$EVENT)

# Quantify Health Risk
FatalityRisk <- 2
InjuryRisk <- 1
RiskLevels <- data.frame(Category = 5:1, Label = c("Disaster", "Significant", "Moderate", "Minor", "Mini
# change here for the new list using different column for cost????
HarmfulEvents <- dat2 %>%
        group_by(EVENT) %>%
        summarise(Fatalities = mean(FATALITIES), Injuries = mean(INJURIES), Cost = mean(COST)) %>%
        mutate(HealthRisk = (FatalityRisk*Fatalities)+(InjuryRisk*Injuries)) %>%
        filter(HealthRisk > 0) %>%
        arrange(desc(HealthRisk)) %>%
        mutate(Category = ntile(HealthRisk, 5)) #Split into 5 categories

row.names(HarmfulEvents) <- HarmfulEvents$EVENT
HarmfulEvents <- merge(x = HarmfulEvents, y = RiskLevels, sort = FALSE)
```

Results

Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

The bar plot below is arranged by priority, indicating the top item is the most signifcant health risk. The plot is color coded for a 5 level standard operating procedure system.

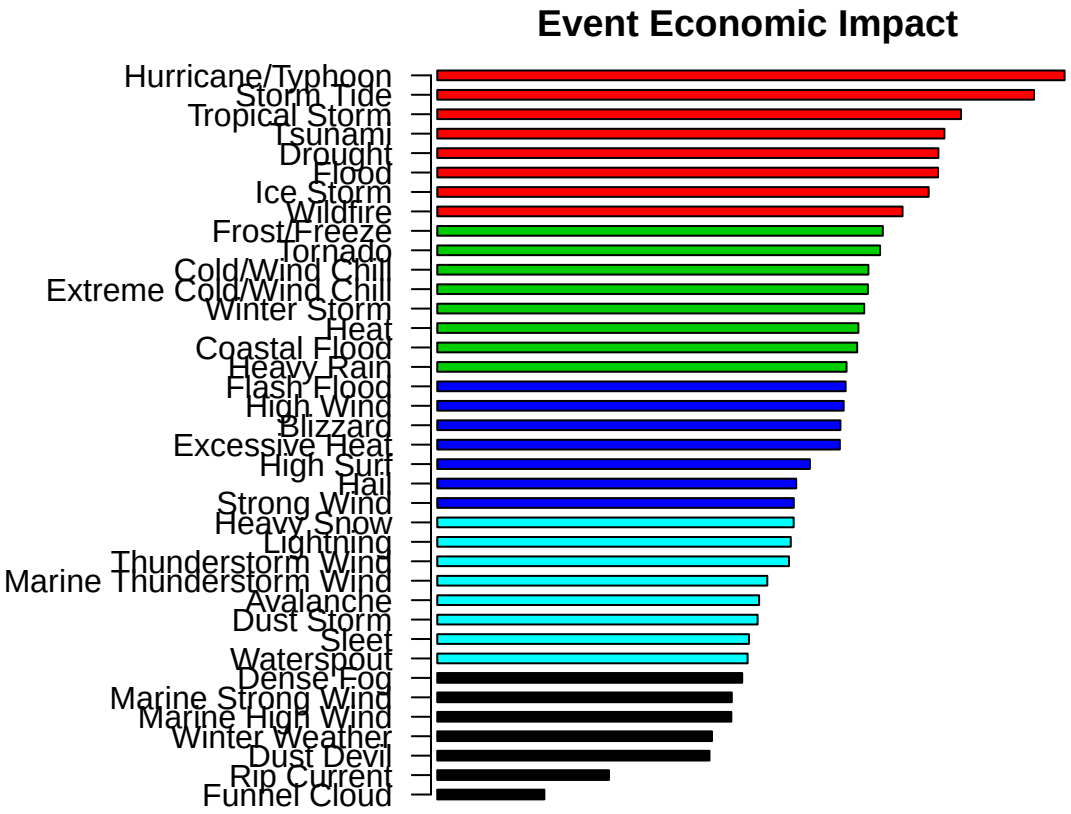
```
# Make a plot of the risk levels
old.par <- par(no.readonly=T) #backup margin settings
par(las=2) #Set the labels horizontal
par(mar = c(1,15,1,1)) #Give the labels some space
barplot(names.arg = rev(HarmfulEvents$EVENT),
        height = rev(HarmfulEvents$HealthRisk),
        horiz = TRUE, #Change to horizontal histogram
        axes = FALSE,
        col = HarmfulEvents$Label,
        space = 1,
        main = "Event Health Risk Levels",
        xlab = "Health Risk",
        axis.lty = 1
        )
```



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

The bar plot below is arranged by priority, indicating the top item is the most signifcant economic risk. The plot is color coded for a 5 level standard operating procedure system. A **log scale** is used due to the high variation in cost.

```
# EconomicEvents
EconomicEvents <- HarmfulEvents %>%
        arrange(desc(Cost))
par(las=2) #Set the labels horizontal
par(mar = c(1,15,1,1)) #Give the labels some space
barplot(names.arg = rev(EconomicEvents$EVENT),
        height = rev(log2(EconomicEvents$Cost)),
        horiz = TRUE, #Change to horizontal histogram
        axes = FALSE,
        col = HarmfulEvents$Label,
        space = 1,
        main = "Event Economic Impact",
        xlab = "Cost",
        axis.lty = 1
        )
```




```
par <- par(old.par) # reset
```