Analysis of Public Health and Economy Impacts of Weather Events

Synopsis

Severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events result in fatalities, injuries, and property damage.

The goal of this assignment is to explore the U.S. National Oceanic and Atmospheric Administration's (NOAA) Storm Database and determine, across the United States, which types of events are most harmful with respect to population health, and which types of events have the greatest economic consequences.

Using the Storm Events Database dataset, we were able to determine that the type of storm event with the highest impact to human health are:

- Tornadoes, thunderstorm wind, and hail and Hurricanes and tropical storms.
- The former occurs with the greatest frequency and still his the highest average impact to health per storm event.

- In terms of economical impact, our model determined that flooding and heavy rain along with the previous two have the highest negative impact.
- Overall, tornadoes appear to be the worst threat to both health and economies of all storm events.

Data Processing

Initialization

We clean-up our workspace, set the seed for reproducibility, remember processing date and time and R version for reference, and show the parameters of the session.

```
# cleanup
rm(list=ls())

#libraries
library (lattice)
library (xtable)

# for reproductibility
set.seed(590607)

# some usefule variables
dt = Sys.time()
date <- format(dt,"%d-%b-%Y")
time <- format(dt,"%H:%M:%S")
Rversion <- version$version.string
```

Downloading data

```r
# baseDir will be prefixing all data accesses
baseDir <- "C:/Users/Prasanna/Desktop/Coursera/Reproducible_Research"

# create data sub-directory if necessary
dataDir <- file.path(baseDir, "data")
if(!file.exists(dataDir)) { dir.create(dataDir) }

zipFilePath <- file.path(dataDir, "repdata_data_StormData.csv.bz2")
dateFilePath <- file.path(dataDir, "date_time_downloaded.txt")
# download original data if necessary (skip if exists already as it takes time and bandwith)
# just delete the bz2 file in data subdir to trigger a fresh download
if (!file.exists(zipFilePath)) {
  zipFileUrl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
  download.file (zipFileUrl, zipFilePath, method="curl")
  DTDownloaded <- format(Sys.time(), "%Y-%b-%d %H:%M:%S")
  cat (DTDownloaded, file=dateFilePath)
} else {
  DTDownloaded <- scan(file=dateFilePath, what="character", sep="\n")
}
```

Reading data

We load data in R using read.csv and bzfile to uncompress it.

```r
# read dataset and load data in R
dataset.raw <- read.csv(bzfile(zipFilePath), header = TRUE)
```

Selecting variables

Looking at the raw dataset and its documentation, only some variables will be interesting for this work, mainly:

- EVTYPE: event type.
- FATALITIES: number of dead people resulting from the event.
- INJURIES: number of non fatal casualties.
- PROPDMG and PROPDMGEXP: property damage in USD, PROPDMG being the mantissa (significand) and PROPDMGEXP the exponent.
- CROPDMG and CROPDMGEXP: crop damage in USD, CROPDMG being the mantissa and CROPDMGEXP the exponent.
- BNG_DATE, END_DATE: start/end date, useful if we want to calculate the duration of the event.
- STATE: state where the event occurred, if we want to show the geographic variations.

Let's start with selecting only these variables to start our tidy dataset:

```r
dataset.tidy <- dataset.raw[, c("EVTYPE", "FATALITIES", "INJURIES",
                     "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP",
                     "BGN_DATE", "END_DATE", "STATE") ]
```

Health Impacts

Health impacts are recorded through two variables: FATALITIES and INJURIES. Let's calculate the total number of casualties in the variable CASUALTIES:

```
dataset.tidy$CASUALTIES <- dataset.tidy$FATALITIES + dataset.tidy$INJURIES
```

The number of casualties ranges from 0 to 1742.

Calculating costs

We now want to calculate some actual dollars from PROPDMG, PROPDMGEXP, CROPDMG and CROPDMGEXP in newly created variables: PROP_COST, CROP_COST, and TOTAL_COST. According to the documentation, xROPDMGEXP variables are supposed to contain K, M, or B, for Kilo, Million, Billion. But a quick table scan shows that we can also find several other characters.

```
t1 <- as.data.frame(table(dataset.tidy$PROPDMGEXP))
t2 <- as.data.frame(table(dataset.tidy$CROPDMGEXP))
tt <- merge(t1,t2,by="Var1",all.x=T,all.y=T)
colnames( tt ) <- c("Value","Count Prop","Count Crop")
print (xtable(tt,caption="-ROPDMGEXP variable"), type="html", include.rownames=FALSE)
```

-ROPDMGEXP variable

Value

Count Prop

Count Crop

465934

618413

- 1

  ?

  8

  7

  - 5

    0

    216

    19

    1

    25

    2

    13

    1

    3

    4

    4

    4

5
28
6
4
7
5
8
1
B
40
9
h
1
H
6
K
424665
281832
m
7
1
M
11330
1994
k
21

Thus, we generate NA in the xxx_COST variables for all records whose variable PROPDMGEXP or CROPDMGEXP doesn't show empty, K, M, or B (being insensitive to case). For correct exponent, we adjust the numbers accordingly.

```
f <- function(x) {
  x <- toupper(x);
  if (x == "") return (1);
  if (x == "K") return (1000);
  if (x == "M") return (1000000);
  if (x == "B") return (1000000000);
  return (NA);
}
dataset.tidy$PROP_COST <- with(dataset.tidy, as.numeric(PROPDMG)  * sapply(PROPDMGEXP,f))
dataset.tidy$CROP_COST <- with(dataset.tidy, as.numeric(CROPDMG) * sapply(CROPDMGEXP,f))
dataset.tidy$TOTAL_COST <- dataset.tidy$PROP_COST + dataset.tidy$CROP_COST
```

Evaluating dates and duration

Then, let's create additional variables: DATE_START, DATE_END, YEAR, and DURATION. In the raw dataset, BNG_START is recorded as a factor but want to have a start date available as an actual date that could be manipulated. DURATION will be created as a number of hours, neglecting the time VARIABLES for now.

```
dataset.tidy$DATE_START <- strptime(as.character(dataset.tidy$BGN_DATE), "%m/%d/%Y")
dataset.tidy$DATE_END <- strptime(as.character(dataset.tidy$END_DATE), "%m/%d/%Y")
dataset.tidy$YEAR <- as.integer(format(dataset.tidy$DATE_START, "%Y"))
dataset.tidy$DURATION <- as.numeric(dataset.tidy$DATE_END - dataset.tidy$DATE_START)/3600
```

We verify that the records range from 1950 to 2011.

Working on event types

The type of event is determined by the variable EVTYPE. Looking at the raw dataset, we can see that EVTYPE is not cleanly recorded. Some same events are recorded with different names (e.g "THUNDERTORM WINDS", "THUNDERSTORMWINDS", "THUNDERSTROM WIND"), different events are grouped within the same records (e.g. "GUSTY WIND/HVY RAIN"), and some records show summary instead of individual events (e.g. "Summary of June 3").

```
nev0 <- length(sort(unique(dataset.tidy$EVTYPE)))
```

There are 985 different event types in the initial dataset.

We first delete the daily summaries which don't correspond to an event:

```
dataset.tidy <- subset(dataset.tidy, substr(EVTYPE,1,7) != "Summary")
```

Then, in order to normalize EVTYPE, we perform the following transformations:

- Normalize the letter-case.
- Suppress plural marks.
- Correct spelling issues.
- Regroup similar events.

```
dataset.tidy$EVTYPE <- toupper(dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("/"," ",dataset.tidy$EVTYPE) # might select one instead
dataset.tidy$EVTYPE <- gsub("  "," ",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^ ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^SEVERE ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^RECORD ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^UNSEASONABLY ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^UNSEASONABLY ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^UNSEASONABLE ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- sub("^UNSEASONAL ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("SML ","SMALL ",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("FLD","FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WINTERY","WINTER",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WINTRY","WINTER",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("TSTM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("HVY","HEAVY",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("FOREST FIRE","FIRE",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("FLOODING","FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("FLDG","FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("LIGHTING","LIGHTNING",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("LIGNTNING","LIGHTNING",dataset.tidy$EVTYPE)
```

```
dataset.tidy$EVTYPE <- gsub("FLOOD FLASH","FLASH FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^FLASH FLOOD.+$","FLASH FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^FLOOD.+$","FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDERTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDESTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUDERSTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDERSTROM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDERTSROM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNERSTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDERESTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDEERSTORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("THUNDERTSORM","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^THUNDERSTORM? WIND.+$","THUNDERSTORM WINDS",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^THUNDERSTORM.+$","THUNDERSTORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("TORNADOE","TORNADO",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("TORNADOES","TORNADO",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("TORNDAO","TORNADO",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^TORNADO.+$","TORNADO",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^HURRICANE.+$","HURRICANE",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^TROPICAL STORM.+$","TROPICAL STORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^SNOW.+$","SNOW",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^HIGH WIND.+$","HIGH WIND",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^LIGHTNING.+$","LIGHTNING",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^ICE .+$","ICE",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^HAIL .+$","HAIL",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^HEAVY RAIN.+$","HEAVY RAIN",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^HEAVY SNOW.+$","HEAVY SNOW",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^FREEZING RAIN.+$","FREEZING RAIN",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^BLIZZARD.+$","BLIZZARD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^RIP CURRENT.+$","RIP CURRENT",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^SLEET.+$","SLEET",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WIND HAIL","HAIL",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^DRY MICROBURST.+$","DRY MICROBURST",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^SMALL STREAM.+$","SMALL STREAM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^FUNNEL CLOUD.+$","FUNNEL CLOUD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^RAIN.+$","RAIN",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^WINTER WEATHER.+$","WINTER WEATHER",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^WINTER STORM.+$","WINTER STORM",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^WIND .+$","WIND",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WND","WIND",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("HEAT WAVE","HEAT",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WILDFIRE","WILD FIRE",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("HIGH ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("EXCESSIVE ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("HEAVY ","",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("WINTER SNOW","SNOW",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("RIVER FLOOD","FLOOD",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^FROST.+$","FROST",dataset.tidy$EVTYPE)
dataset.tidy$EVTYPE <- gsub("^LANDSLIDE.+$","LANDSLIDE",dataset.tidy$EVTYPE)
#sort(unique(dataset.tidy$EVTYPE))
nev1 <- length(sort(unique(dataset.tidy$EVTYPE)))
```

We have now 376 distinct event types. Note that these transformations are heuristic and that further work should be done to properly regroup events, including selecting the most relevant event when several are reported in the same record.

Grouping for preparing the analysis

We can now create analysis datasets by grouping the different outcomes by event type, state, and year.

```
byEvType <- subset(dataset.tidy, select=c(EVTYPE, INJURIES, FATALITIES, CASUALTIES, PROP_COST, CROP_COS
byEvType <- aggregate (. ~ EVTYPE, data = byEvType, sum)
# keep event with some impact
byEvType <- subset(byEvType,  CASUALTIES+TOTAL_COST > 0)
# in order to avoid decimals
byEvType$CASUALTIES <- as.integer(byEvType$CASUALTIES)
byEvType$FATALITIES <- as.integer(byEvType$FATALITIES)
byEvType$INJURIES <- as.integer(byEvType$INJURIES)

byState <- subset(dataset.tidy, select=c(STATE, INJURIES, FATALITIES, CASUALTIES, PROP_COST, CROP_COST,
byState <- aggregate(. ~ STATE, data = byState, FUN="sum")

byYear <- subset(dataset.tidy, select=c(YEAR, INJURIES, FATALITIES, CASUALTIES, PROP_COST, CROP_COST, T
byYear <- aggregate(. ~ YEAR, data = byYear, FUN="sum")
```

Let's add some percent for relevant columns.

```
byEvType$pctCasualties <- with(byEvType, round(CASUALTIES/sum(CASUALTIES) * 100, 2))
byEvType$pctInjuries <- with(byEvType, round(INJURIES/sum(INJURIES) * 100, 2))
byEvType$pctFatalities <- with(byEvType, round(FATALITIES/sum(FATALITIES) * 100, 2))
byEvType$pctTotalCost <- with(byEvType, round(TOTAL_COST/sum(TOTAL_COST) * 100, 2))
byEvType$pctPropCost <- with(byEvType, round(PROP_COST/sum(PROP_COST) * 100, 2))
byEvType$pctCropCost <- with(byEvType, round(CROP_COST/sum(CROP_COST) * 100, 2))
```

Results

Which types of events are most harmful to population health?

Let's create a table showing the 20 most important event types in term of casualties.

```
tmp <- byEvType[with(byEvType, order(-CASUALTIES)),c("EVTYPE","INJURIES","pctInjuries","FATALITIES","pc
tmp <- tmp[1:20,]
tmp2 <- tmp
colnames(tmp2) <- c("Type of Event", "Injuries", "%", "Fatalities", "%", "Total", "%")
print(xtable(tmp2, caption="US Casualties From Weather Event, 1950-2011"),type="html",include.rownames=
```

US Casualties From Weather Event, 1950-2011

Type of Event

Injuries

%

Fatalities

%

Total

%

TORNADO

91339

65.02

5655

37.36

96994

62.33

HEAT

9054

6.45

3031

20.03

12085

7.77

THUNDERSTORM

9507

6.77

711

4.70

10218

6.57

FLOOD

6794

4.84

482

3.18

7276

4.68

LIGHTNING

5232

3.72

817

5.40

6049

3.89

FLASH FLOOD

1800

1.28

1035

6.84

2835

1.82

ICE

2115

1.51

96

0.63

2211

1.42

WIND

1535

1.09

312

2.06

1847

1.19

WINTER STORM

1353

0.96

217

1.43

1570

1.01

WILD FIRE

1456

1.04

87

0.57

1543

0.99

HURRICANE

1328

0.95

135

0.89

1463

0.94

HAIL

1358

0.97

15

0.10

1373

0.88

SNOW

1134

0.81

140

0.93

1274

0.82

RIP CURRENT

529

0.38

577

3.81

1106

0.71

BLIZZARD

805

0.57

101

0.67

906

0.58

FOG

734

0.52

62

0.41

796

0.51

WINTER WEATHER

538

0.38

61

0.40

599

0.38

DUST STORM

440

0.31

22

0.15

462

0.30

TROPICAL STORM

383

0.27

66

0.44

449

0.29

AVALANCHE

170

0.12

224

1.48

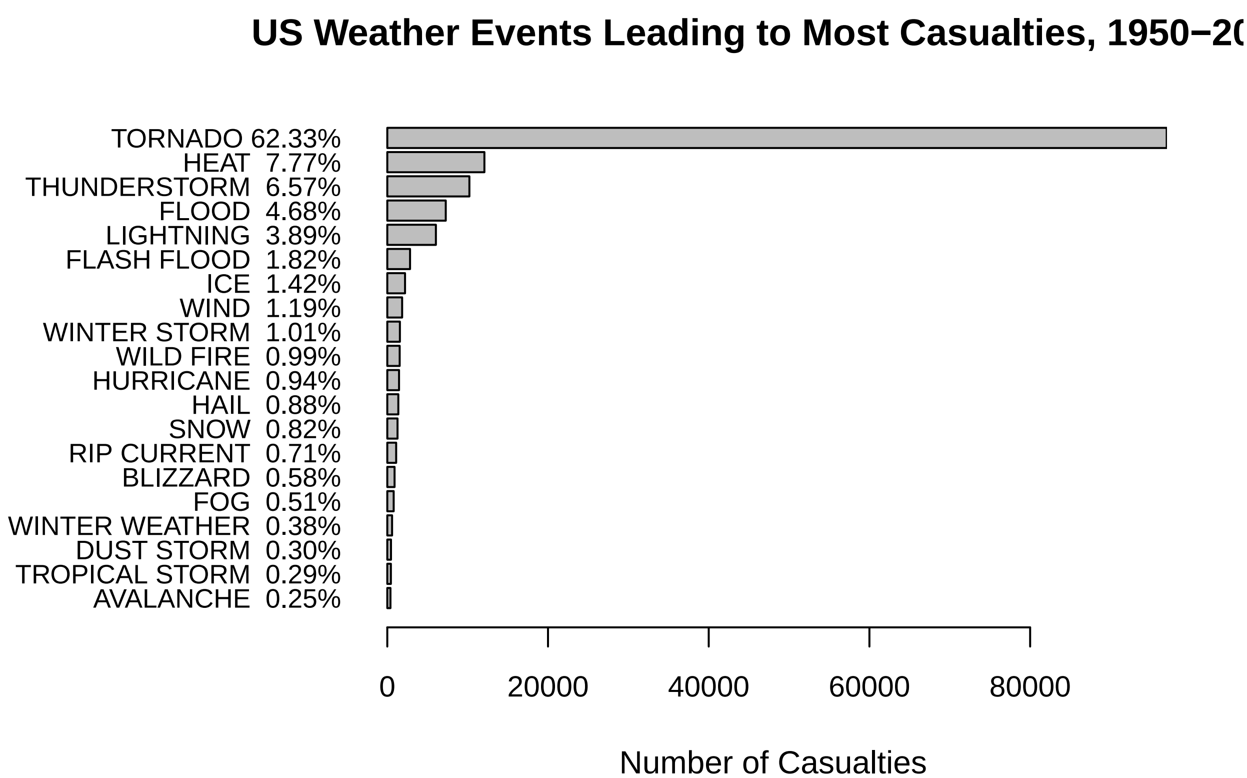
394

0.25

And put the total number of casualities per event type on a chart.

```
tmp <- tmp[order(tmp$CASUALTIES),]
labels <- paste(tmp$EVTYPE, format(tmp$pctCasualties,digits=2))  # add percents to labels
labels <- paste(labels, "%", sep = "")  # ad % to labels

par(las=1, mar=c(5,10,4,2))
barplot(tmp$CASUALTIES,
        horiz=T, las=1, space=.2,
        names.arg = labels, # tmp$EVTYPE,
        cex.axis = .9,
        cex.names = .8, xpd=F,
        main="US Weather Events Leading to Most Casualties, 1950-2011",
        xlab="Number of Casualties")
```




```
tot <- sum(dataset.tidy$CASUALTIES,na.rm=T)
inj <- sum(dataset.tidy$INJURIES,na.rm=T)
fat <- sum(dataset.tidy$FATALITIES,na.rm=T)
pctInj <- round(100 * inj / tot,2)
pctFat <- round(100 * fat / tot,2)
```

The total number of casualties over the 61 years period analysed is 155,673, with 140,528 injuries (90.27%) and 15,145 deaths (9.73%).

Which types of events have the greatest economic consequences?

Let's see the table first (in billions of US dollars).

```r
# sorted by total cost
tmp <- byEvType[with(byEvType, order(-TOTAL_COST)),c("EVTYPE","PROP_COST","pctPropCost","CROP_COST","pc
tmp <- tmp[1:20,]
# transform to billions
tmp$TOTAL_COST <- tmp$TOTAL_COST / 1000000000
tmp$PROP_COST <- tmp$PROP_COST / 1000000000
tmp$CROP_COST <- tmp$CROP_COST / 1000000000
tmp2 <- tmp
colnames(tmp2) <- c("Type of Event", "Property", "%", "Crop", "%", "Total", "%")
print (xtable(tmp2, caption="US Weather Events Costs (billions USD), 1950-2011"), type="html", include.r
```

US Weather Events Costs (billions USD), 1950-2011

Type of Event

Property

%

Crop

%

Total

%

FLOOD

150.02

35.11

10.84

22.11

160.86

33.77

HURRICANE

84.76

19.83

5.52

11.25

90.27

18.95

TORNADO

58.54

13.70

0.37

0.75

58.91

12.37

STORM SURGE

43.32

10.14

0.00

0.00

43.32

9.09

HAIL

15.73

3.68

3.00

6.12

18.73

3.93

FLASH FLOOD

16.91

3.96

1.53

3.12

18.44

3.87

DROUGHT

1.05

0.24

13.97

28.50

15.02

3.15

THUNDERSTORM

10.97

2.57

1.27

2.59

12.24

2.57

ICE

3.97

0.93

5.03

10.25

9.00

1.89

TROPICAL STORM

7.71

1.81

0.69

1.42

8.41

1.77

WILD FIRE

7.77

1.82

0.40

0.82

8.17

1.71

WINTER STORM

6.75

1.58

0.03

0.07

6.78

1.42

WIND

6.01

1.41

0.69

1.40

6.70

1.41

STORM SURGE TIDE

4.64

1.09

0.00

0.00

4.64

0.97

RAIN

3.24

0.76

0.81

1.64

4.04

0.85

EXTREME COLD

0.07

0.02

1.31

2.68

1.38

0.29

FROST

0.01

0.00

1.16

2.37

1.17

0.25

SNOW

0.98

0.23

0.13

0.27

1.12

0.23

LIGHTNING

0.93

0.22

0.01

0.02

0.95

0.20

HEAT

0.02

0.00

0.90

1.83

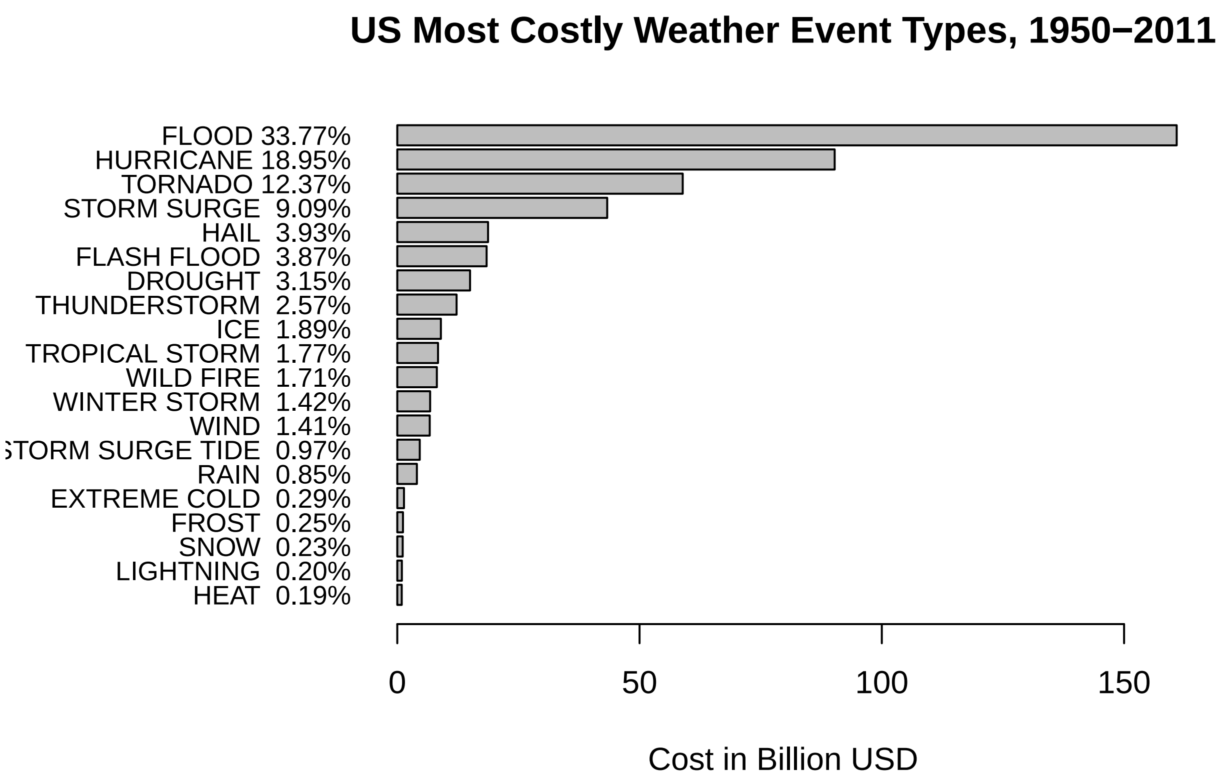
0.92

0.19

Then, the plot of the total cost:

```
tmp <- tmp[order(tmp$TOTAL_COST),]
labels <- paste(tmp$EVTYPE, format(tmp$pctTotalCost,digits=2))  # add percents to labels
labels <- paste(labels, "%", sep = "")  # ad % to labels

par(las=1, mar=c(5,10,4,2))
p <- barplot(tmp$TOTAL_COST,
        horiz=T, las=1, space=.2,
        axisnames=T, names.arg = labels, # tmp$EVTYPE,
        cex.axis = 1, offset=0,
        cex.names = .8, xpd=T,
        main="US Most Costly Weather Event Types, 1950-2011",
        xlab="Cost in Billion USD")
```




```
billionTotal <- round(sum(dataset.tidy$TOTAL_COST,na.rm=T)/1000000000)
billionProp <- round(sum(dataset.tidy$PROP_COST,na.rm=T)/1000000000)
billionCrop <- round(sum(dataset.tidy$CROP_COST,na.rm=T)/1000000000)
pctProp <- round(100 * billionProp / billionTotal,2)
pctCrop <- round(100 * billionCrop / billionTotal,2)
```

The total cost over the 61 years period analyzed (without inflation nor cost of life correction) is 476 billions of US dollards, 427 (89.71%) for properties damages and 49 (10.29%) for crop damages.

```
c1 <- round(byEvType[order(-byEvType$TOTAL_COST),"pctTotalCost"][1])
c2 <- round(byEvType[order(-byEvType$TOTAL_COST),"pctTotalCost"][2])
c3 <- round(byEvType[order(-byEvType$TOTAL_COST),"pctTotalCost"][3])
h1 <- round(byEvType[order(-byEvType$CASUALTIES),"pctCasualties"][1])
```

Summary

Flood has the greatest economic impact (34%), followed by Hurricanes (19%) and Tornadoes(12%). Preventive measures should focus on these events, specially floods, for example in improving the management of water flow.

Concerning population health, tornadoes create 62% of the casualties, far above the rest of the events. Preventive actions could for example be focused on detecting early signs, alerting populations, and better shelters.

It would be interesting to further group events by types relating to the same prevention measures.

Furthermore, an analysis of the progression of the main event types over time could determine trends and help better selecting the most relevant measures to address these trends.

Finally, the calculated costs should be adjusted to reflect current costs, considering inflation over time and the comparative cost of life.

Conclusion

This brief analysis prepared for the second assignment of Coursera "Reproducible Research" aims at determining which weather events are most harmful to the population health and to the economy, from a US survey of weather events between 1950 and 2011.

Tornadoes, floods, and hurricanes show the greatest impact on economy and on health. Prevention efforts should therefore be focused on these events.