Impact of Weather in USA to Economy and Population Health

Synopsis

In this report we aim to find out which weather event in the United States between years 1950-2011 has been most harmful respect to population health and had the greatest economic consequences. This research involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. The database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. NOAA has 50 different event categories in their database. From these data, we found that tornado is by far most harmful respect to population health accounting for 91346 injuries and 5633 fatalities recorded. Most property damage has caused floods (122.5 billion USD), runner up is hurricanes/typhoons (74.40 billion USD) and third most damaging is storm surges (46.56 billion USD). Other events do not have closely significant impact as top 3. The biggest contributor to crop damage is drought (13.95 billion USD).

Loading and Processing the Raw Data

```
SD <- read.csv(file = "repdata_data_StormData.csv.bz2", sep = ",")
```

Having a look of the column names and variables

```
names(SD)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

Subsetting only relevant columns which are "STATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG", "CROPDMGEXP"

```
SD <- SD[,c(2,7,8,23:28)]
```

Renaming some columns for easier programming

```
colnames(SD)[1:5] <- c("Date","State","EvType","Fat","Inj")
```

Converting date to date format

```
SD$Date <- as.Date(SD$Date, format = "%m/%d/%Y")
```

Converting "EvType" and "State" form character vectors to factors

```
SD$EvType <- as.factor(SD$EvType)
SD$State <- as.factor(SD$State)
```

Copying event variables from "Storm Data Documentation" and saving as variable "vars"

```
vars <- c("Astronomical Low Tide", "Avalanche", "Blizzard", "Coastal Flood" , "Cold/Wind Chill", "Debris
vars <- toupper(vars) # As our dataset is in uppercase letters we need to convert "vars" to uppercase l
```

Weather Impact on Fatalities

Calculating the sum of fatalities for each event

```
library(tidyverse)
SumFat <- group_by(SD, EvType) %>%
    summarise(Fat = sum(Fat)) %>%
    arrange(desc(Fat)) # sorting data frame to descending order
```

Removing events with low frequency (30 or less), because a lot of those entries are difficult to classify and it has insignificant impact on analysis:

```
SumFat <- filter(SumFat, Fat > 30)
```

Checking if the variables match with documentation after renaming "EvTypes"

```
which(!SumFat$EvType %in% vars) #showing which do not match
```

```
##  [1]  6 12 13 14 17 22 27 28 29 31 32 33 34
```

Now we have to take a look why does the mismatches occur. For example the entry "TSTM WIND" is abbrevation of"Thunderstorm Wind". Another examples include but not exhaustive "HEAT" = HEAT WAVE","EXTREME COLD/WINDCHILL" = "COLD" etc.

Collapsing duplicate factor levels:

```
library(forcats)
SumFat$EvType <- fct_collapse(SumFat$EvType,
           `RIP CURRENTS` = c("RIP CURRENTS", "RIP CURRENT"),
           `THUNDERSTORM WIND` = c("TSTM WIND", "THUNDERSTORM WINDS"),
           `EXTREME COLD/WINDCHILL`= c("EXTREME COLD", "COLD", "EXTREME COLD/WIND CHILL")
         )
SumFat$EvType <- fct_recode(SumFat$EvType,
                            "HEAT" = "HEAT WAVE",
                            "EXCESSIVE HEAT" = "EXTREME HEAT",
                            "DENSE FOG" = "FOG",
                            "DEBRIS FLOW" = "LANDSLIDE",
                            "HIGH WIND" = "HIGH WINDS")
```

Finding the sum of duplicate entries of events

```
SumFat <- SumFat %>%
  group_by(EvType) %>%
  summarise(Fat, Fat = sum(Fat))
```

Removing duplicate entries

```
SumFat <- distinct(SumFat) %>%
  arrange(desc(Fat))
```

What is sum of top 10 weather events?

```
sum(SumFat$Fat[1:10])
```

```
## [1] 12881
```

That is 79,9 % of total fatalities which means 10 events is sufficient for this analysis.

Leaving only 10 weather events with most fatalities

```
SumFat <- SumFat[1:10,]
```

Creating "Type" column to show type of impact

```
SumFat$Type <- "fatalities"
SumFat <- rename(SumFat, count = Fat) #renaming "Fat" column to count
```

Weather Impact on Injuries

Making a table of total Injuries

```
  SumInj <- SD %>% group_by(EvType) %>%
    summarise(Inj = sum(Inj)) %>%
    arrange(desc(Inj))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

How many injuries in total?

```
sum(SumInj$Inj)
```

```
## [1] 140528
```

How many injuries in top 10 events with most injuries?

```
sum(SumInj$Inj[1:10])
```

```
## [1] 125548
```

Top 10 events account for

```
sum(SumInj$Inj[1:10])/sum(SumInj$Inj)*100
```

```
## [1] 89.3402
```

% of the all the injuries ever recorded

Events with injuries less than 100 are statistically irrelevant for our analysis so we remove them

```
SumInj <- SumInj %>% filter(Inj > 100)
```

Checking if the variables match with documentation after renaming "EvTypes"

```
which(!SumInj$EvType %in% vars) #showing which do not match
```

```
##  [1]  2 16 18 19 24 25 26 30 31 33 35 36
```

Now we have to take a look why does the mismatches occur. There are lot of typos, but we will fix the isse by renaming the factors.

```
SumInj$EvType <- fct_recode(SumInj$EvType,
                            "THUNDERSTORM WIND" = "TSTM WIND",
                            "THUNDERSTORM WIND" = "THUNDERSTORM WINDS",
                            "DENSE FOG" = "FOG",
                            "WILDFIRE" = "WILD/FOREST FIRE",
                            "WILDFIRE" = "WILD FIRES",
                            "HEAT" = "HEAT WAVE",
                            "HIGH WIND" = "HIGH WINDS",
                            "EXTREME COLD/WINDCHILL" = "EXTREME COLD",
                            "FREEZING FOG" = "GLAZE",
                            "EXCESSIVE HEAT" = "EXTREME HEAT"
  )
```

Finding the sum of duplicate entries of events

```
  SumInj <- SumInj %>%
    group_by(EvType) %>%
    summarise(Inj, Inj = sum(Inj))
```

Removing duplicate entries

```
  SumInj <- distinct(SumInj) %>%
    arrange(desc(Inj))
```

Leaving only 10 weather events with most crop damage

```
  SumInj <- SumInj[1:10,]
```

Creating "Type" column to show type of impact

```
SumInj$Type <- "injuries"
```

Renaming “Inj” to “count” to match the same variable in “SumFat”

```
SumInj <- rename(SumInj, count = Inj)
```

Results Of Weather Impact on Population Health

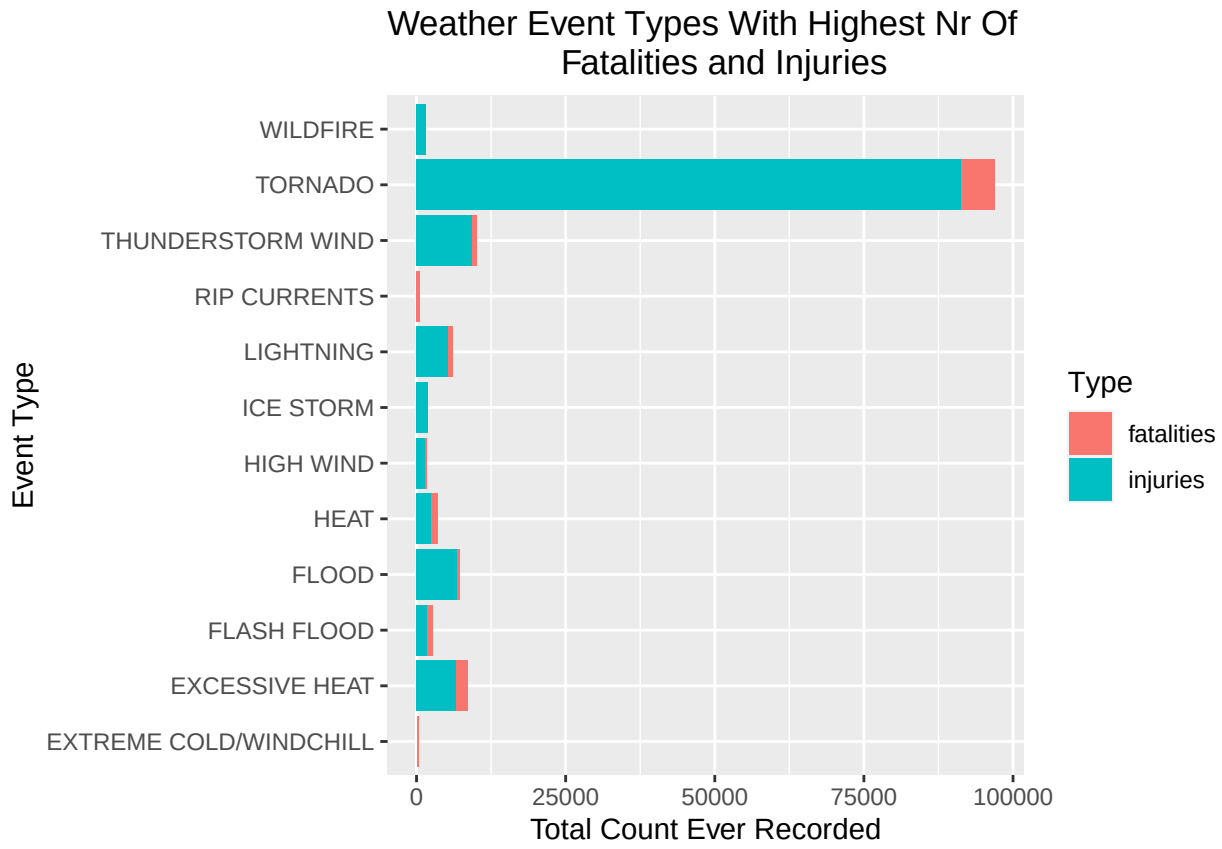
Combining injuries and fatalities into one table

```
Fat.Inj.combined <-rbind(SumFat, SumInj)
```

Creating the plot

```
g <- ggplot(Fat.Inj.combined, aes(x = EvType, y = count, fill = Type)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("Total Count Ever Recorded") +
  ggtitle("Weather Event Types With Highest Nr Of
               Fatalities and Injuries")

g
```



Weather Impact on Property

The highest exponent in the dataset is "B" which is billion. First we need to filter out only data where "PROPDMG" is equal to "B". Then we sort it by event and calculate the sum.

```
prop.dmg <- filter(SD, PROPDMGEXP == "B") %>%
    group_by(EvType) %>%
     summarise(PROPDMG = sum(PROPDMG))
```

Checking if the variables match with documentation after renaming "EvTypes"

```
which(!prop.dmg$EvType %in% vars) #showing which do not match
```

```
## [1]  4  6  7  8 10 11 13 15 17
```

We can ignore FALSE == "TORNADOES, TSTM WIND, HAIL" as it is not possible to classify

Renaming factors according to documentation

```
prop.dmg$EvType <- fct_recode(prop.dmg$EvType,
                              "FLASH FLOOD" = "RIVER FLOOD",
                              "HEAVY RAIN" = "HEAVY RAIN/SEVERE WEATHER",
                              "WILDFIRE" = "WILD/FOREST FIRE",
                              "THUNDERSTORM WIND" ="SEVERE THUNDERSTORM",
                              "STORM SURGE" = "STORM SURGE/TIDE"
                             )
prop.dmg$EvType <- fct_collapse(prop.dmg$EvType,
                                "HURRICANE/TYPHOON" = c("HURRICANE OPAL", "HURRICANE OPAL/HIGH WINDS",
                                )
```

Finding the sum by each weather event

```
prop.dmg <- prop.dmg %>% group_by(EvType) %>%
   summarise(total = sum(PROPDMG)) %>%
   arrange(desc(total)) %>%
   slice(1:10) # we need only 10 events for our plot
```

creating "Type" column for later comparison with crop damage

```
prop.dmg$Type <- "property"
```

Weather Impact on Crops

Creating the data frame

```
  crop.dmg <- group_by(SD, EvType)
```

Taking a look of the count of the exponents

```
table(crop.dmg$CROPDMGEXP)
```

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

We see that there the count of "B" is 9 and M is "1994" which means we weed to combine them to get enough data.

Creating data frames for exponent "B" and "M"

```
crop.dmg.B <- filter(crop.dmg, CROPDMGEXP == "B") %>%
  summarise(total = sum(CROPDMG))
crop.dmg.M <- filter(crop.dmg, CROPDMGEXP == "M") %>%
  summarise(total = sum(CROPDMG)) %>%
  mutate(total = total/1000)                    #converting exponent "M" to "B"
```

Creating single data frame for the exponents

```
crop.dmg <- rbind(crop.dmg.B, crop.dmg.M) %>%
  arrange(desc(total))
```

Removing all events with insignificant impact (less than 0.1 billion USD)

```
crop.dmg <- filter(crop.dmg, total > 0.1)
```

Do the "EvTypes" match with events described in the documentation?

```
which(!crop.dmg$EvType %in% vars) #showing which do not match
```

```
##  [1]  4  5  9 17 22 23 24 25 26 27 29 31
```

Renaming factors according to documentation

```
crop.dmg$EvType <- fct_recode(crop.dmg$EvType,
                              "FLASH FLOOD" ="RIVER FLOOD",
                              "WILDFIRE" ="WILD/FOREST FIRE",
                              "HURRICANE/TYPHOON" = "HURRICANE",
                              "HURRICANE/TYPHOON" = "HURRICANE ERIN",
                              "FROST/FREEZE" ="DAMAGING FREEZE",
                              "FROST/FREEZE" = "FREEZE"
                              )
```

Checking if the variables match with documentation after renaming "EvTypes"

```
which(!crop.dmg$EvType %in% vars)
```

```
## [1]  9 17 25 26 29
```

Only rows nr 24 and 27 do not match, but they are not in documentation so we can ignore them

Finding the sum of duplicate entries of events

```
crop.dmg <- crop.dmg %>% group_by(EvType) %>%
  summarise(total, total = sum(total))
```

```
## `summarise()` regrouping output by 'EvType' (override with `.groups` argument)
```

Removing duplicate entries and leaving only 10 weather events with most crop damage

```
crop.dmg <- distinct(crop.dmg) %>%
  arrange(desc(total))
crop.dmg <- crop.dmg[1:10,]
```

Creating "Type" column to match with "SumFat" data frame

```
crop.dmg$Type <- "crops"
```

Results Of Weather Impact on Economy

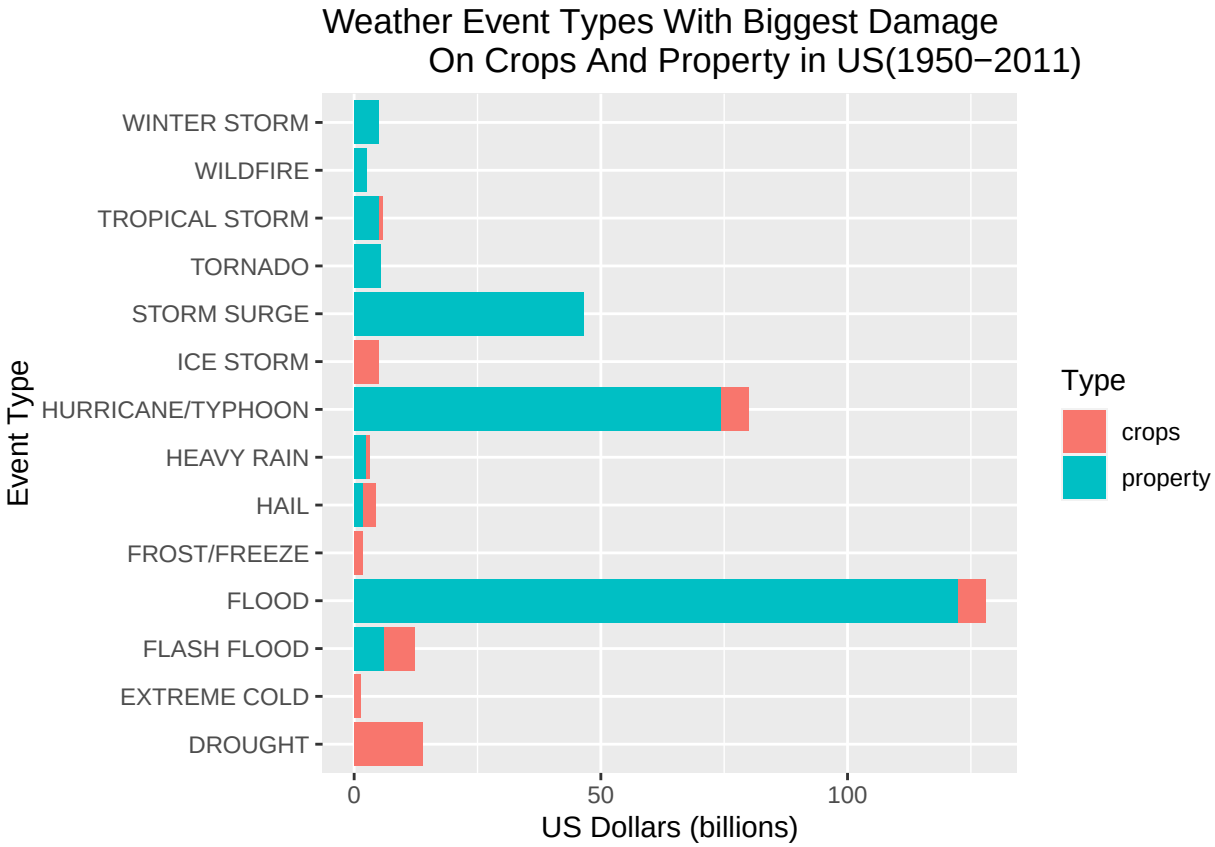
Combing property damage and crop damage into single table

```
prop.crop.combined <- rbind(prop.dmg, crop.dmg)
```

Creating the plot

```
g2 <- ggplot(prop.crop.combined, aes(x = EvType, y = total, fill = Type)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("US Dollars (billions)") +
  ggtitle("Weather Event Types With Biggest Damage
          On Crops And Property in US(1950-2011)")

g2
```


Weather Event Types With Biggest Damage
On Crops And Property in US(1950−2011)
Event Type
WINTER STORM
WILDFIRE
TROPICAL STORM
TORNADO
STORM SURGE
ICE STORM
HURRICANE/TYPHOON
HEAVY RAIN
HAIL
FROST/FREEZE
FLOOD
FLASH FLOOD
EXTREME COLD
DROUGHT
0
50
100
US Dollars (billions)
Type
crops
property