Analysis of The Most Dangerous Weather Events in the United States Since 1950

Synopsis

This report analyzes data from U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database to identify the most dangerous types of weather events in terms of damage to people and property from 1950-2011. Tornados, thunderstorms, and excessive heat were the most dangeous to people. While floods, hurricanes, and tornados caused the most damage to property and crops.

	INJURIES	DAMAGES
1	TORNADO	FLOOD
2	THUNDERSTORM	HURRICANE
3	HEAT	TORNADO
4	FLOOD	STORM SURGE
5	LIGHTNING	HAIL
6	WINTER STORM	DROUGHT
7	STRONG WIND	THUNDERSTORM
8	ICE STORM	ICE STORM
9	FIRE	FIRE
10	HAIL	TROPICAL STORM

The top 10 weather events cause 99% of the damages so that is where safety efforts should focused. There is a long tail of other types events including "dust devils." Although they sound fierce, they have only caused $2,300 in total damages and no injuries.

Since weather patterns are signifincatly different in different regions of the united states. It is also useful to look at the data on a state by state basis, which I have done at the bottom of the report.

Data Processing

The data was downloaded from https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 and contains 902,297 records.

```
library(dplyr)
library(stringr)
library(choroplethr)
library(ggplot2)

stormdata <- read.csv("repdata-data-StormData.csv.bz2")
stormdata <- select(stormdata, REFNUM, BGN_DATE, STATE, EVTYPE, FATALITIES,INJURIES, PROPDMG, PROPDMGEX
```

The raw data contained several anomolies that required cleanup and processing before the analysis could begin. In particular, assessment of the damages stored a numeric figure in the PROPDMG collumn and then a factor in the PROPDMGEXP collumn that represented the exponent (K, M, or B) that should be applied to the numeric value to get the total damages.

```
ApplyExponents <- function (DMGCol, EXPCol) {
  #This will multiply the damages by the right factor to get the total damages

  for (i in 1:length(EXPCol)) {
    if (is.na(EXPCol[i])) { next()
    } else if (EXPCol[i] == "M" | EXPCol[i] == "m") {
      DMGCol[i] <- DMGCol[i] * 1000000
    } else if (EXPCol[i] == "B") {
      DMGCol[i] <- DMGCol[i] * 1000000000
    } else if (DMGCol[i] > 0) {
      # I assume the default scale for all other values (K,k, 2,0,) is thousands
      # Since K by far the median value
      DMGCol[i] <- DMGCol[i] * 1000
    }
  }
  return(DMGCol)
}

stormdata$PROPDMG <- ApplyExponents (stormdata$PROPDMG, stormdata$PROPDMGEXP)
stormdata$CROPDMG <- ApplyExponents (stormdata$CROPDMG, stormdata$CROPDMGEXP)
```

The worst event in terms of damages was a flood in California on 1/1/2006 with total damages listed as $115,000,000,000. This was an outlier so I checked other sources and there indeed was a flood on that day, but the damages were estimated at only 300,000,000. I manually changed that value because it was skewing the data.Then I summed the the property damages and the crop damages to get the total damages for each event. To estimate the impact to people I used the Injuries collumn because it seemed to also include any fatalities.

```
#According to this report the damage from the 1/1/2006 flood in california was $300M USD
maxdamage <- max(stormdata$PROPDMG)
stormdata[stormdata$PROPDMG==maxdamage,]$TOTALDMG <- 300000000
stormdata <- mutate(stormdata, TOTALDMG = PROPDMG+CROPDMG)
```

Another issue with the data was that there was no fixed taxonomy to classify the events. For example Hurricanes were listed as “Hurricanes”, “HURRICANE”, “Hurricane Florence”, etc. In order to create meaningful totals these needed to consolidated. I converted all names to upper case, removed items that had no damages or injuries, and combined similiar looking event types into a standard form such as “HURRICANE”. I cleaned up the top 25 types of events, which represented 99% of the total damage. There was a long tail of other names but I did not clean them all because they had low damages and would not have made an impact on the analysis.

To analyze the impact of different event types I aggregated the data by the sums of each event type by state–since the individual records for each event were not needed.

```
stormdata.agg <- aggregate(stormdata[c("TOTALDMG", "INJURIES")], by=list(EVENT=stormdata$EVTYPE, STATE=s
stormdata.agg <- filter(stormdata.agg, stormdata.agg$TOTALDMG > 0 | stormdata.agg$INJURIES > 0)
stormdata.agg$EVENT <- str_trim(toupper(as.character(stormdata.agg$EVENT)))

for (i in 1:nrow(stormdata.agg)) {
  #clean up the event names
  if (stormdata.agg[i,]$EVENT=="?") {stormdata.agg[i,]$EVENT <- "OTHER"}
  if (grepl("HURR", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HURRICANE"}
  if (grepl("HURRICANE/TYPHOON", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HURRICANE"}
```

```
  if (grepl("THUNDER", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "THUNDERSTORM"}
  if (grepl("TSTM", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "THUNDERSTORM"}
  if (grepl("TORN", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "TORNADO"}
  if (grepl("FLOOD", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "FLOOD"}
  if (grepl("HEAT", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HEAT"}
  if (grepl("HAIL", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HAIL"}
  if (grepl("FREEZE", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "FREEZE"}
  if (grepl("FIRE", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "FIRE"}
  if (grepl("STORM SURGE", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "STORM SURGE"}
  if (grepl("WIND", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "STRONG WIND"}
  if (grepl("HEAVY RAIN", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HEAVY RAIN"}
  if (grepl("HEAVY SNOW", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "HEAVY SNOW"}
  if (grepl("COLD", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "EXTREME COLD"}
  if (grepl("BLIZZARD", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "WINTER STORM"}
  if (grepl("FLD", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "FLOOD"}
  if (grepl("FROST", stormdata.agg[i,]$EVENT)) {stormdata.agg[i,]$EVENT <- "FROST"}
}

#aggregate one more time to flatten the new duplicate event names
stormdata.agg <- aggregate(stormdata.agg[c("TOTALDMG", "INJURIES")], by=list(EVENT=stormdata.agg$EVENT,
```

Results

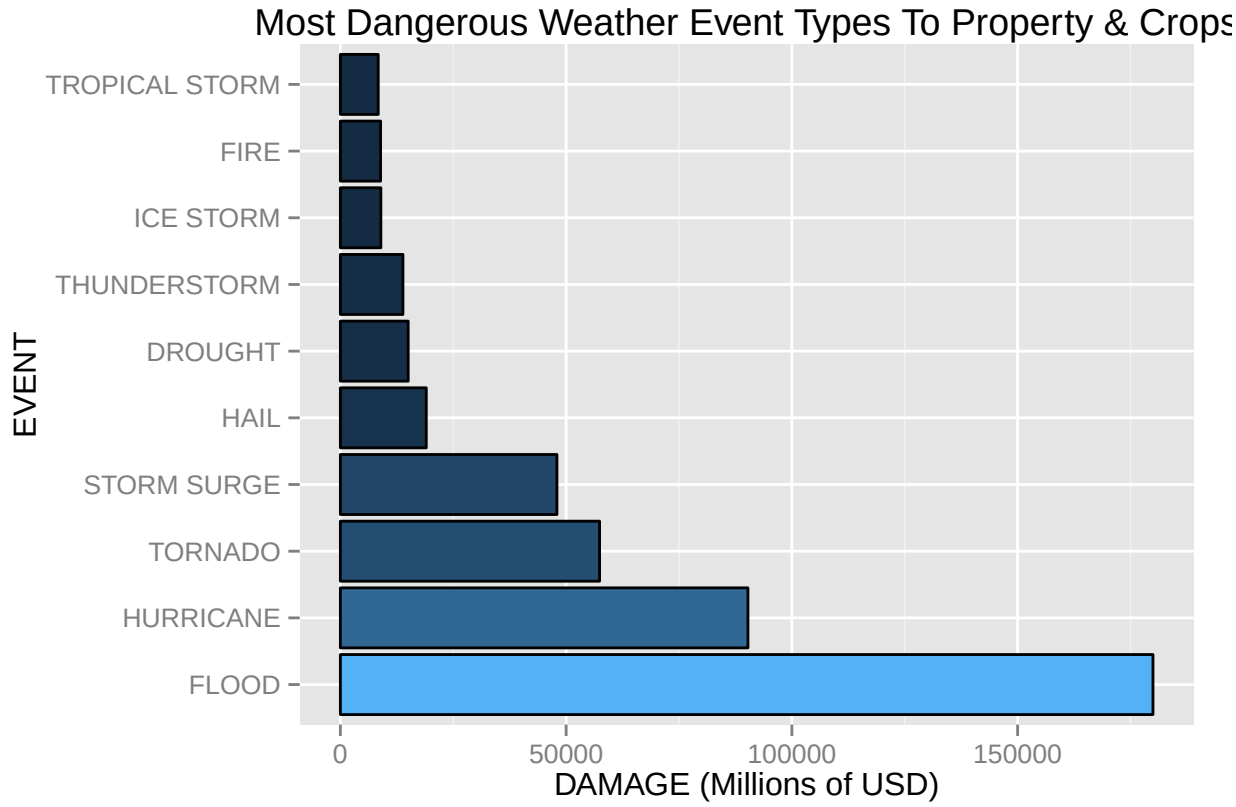
Damage to property and crops

Floods by far caused the most damage to property and crops. Hurricanes were number 2, though flooding and storm surge are caused by hurricanes so in fact hurricanes may be the most dangerous. Further, row by row anaysis is recomended to check the relationship between hurricanes, storm surge, and floods.

```
#chart the top weather events by total damage
stormdata.top <- group_by(stormdata.agg, EVENT) %>%
  summarize(DAMAGE=round(sum(TOTALDMG)/1000000)) %>%
  arrange(desc(DAMAGE)) %>% top_n(10)

stormdata.top$DAMAGE <- as.numeric(as.character(stormdata.top$DAMAGE))
stormdata.top$EVENT <- as.factor(stormdata.top$EVENT)
stormdata.top$EVENT <- factor(stormdata.top$EVENT, stormdata.top$EVENT)

ggplot(data=stormdata.top, aes(x=EVENT, y=DAMAGE, fill=DAMAGE)) +
  geom_bar(colour="black", stat="identity")  + coord_flip() +
  theme(legend.position="none") +
  ggtitle("Most Dangerous Weather Event Types To Property & Crops") +
  ylab("DAMAGE (Millions of USD)")
```



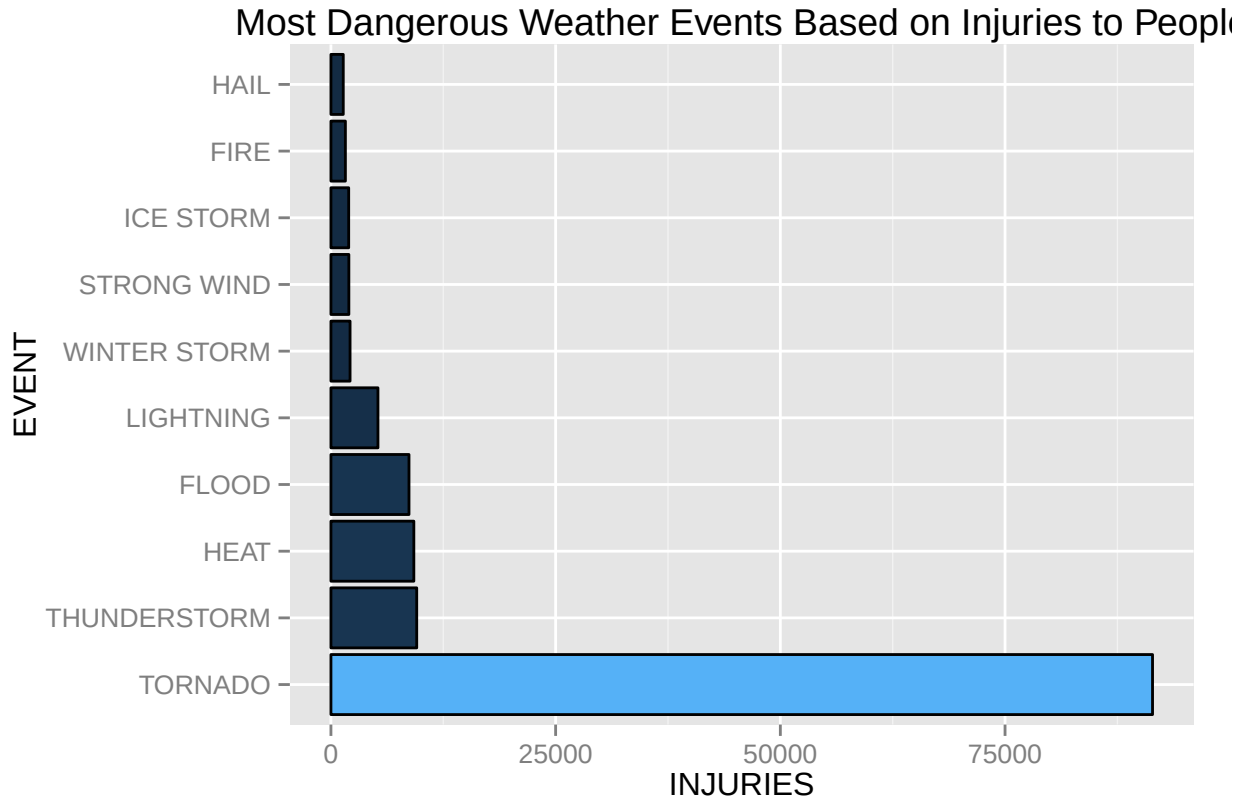
Injuries to people

Tornados cause more injuries than all other types of events combined.

```
#chart the top weather events by injuries
stormdata.top <- group_by(stormdata.agg, EVENT) %>%
  summarize(INJURIES=sum(INJURIES)) %>%
  arrange(desc(INJURIES)) %>% top_n(10)

stormdata.top$INJURIES <- as.numeric(as.character(stormdata.top$INJURIES))
stormdata.top$EVENT <- as.factor(stormdata.top$EVENT)
stormdata.top$EVENT <- factor(stormdata.top$EVENT, stormdata.top$EVENT)

ggplot(data=stormdata.top, aes(x=EVENT, y=INJURIES, fill=INJURIES)) +
  geom_bar(colour="black", stat="identity")  + coord_flip() +
  theme(legend.position="none") +
  ggtitle("Most Dangerous Weather Events Based on Injuries to People")
```



Analyze by state

Analyzing this data by country total may not be helpful for individual regions since weather patterns are different in different regions. For example the Northwest has very few tornados. As a final analysis I plotted the top causes of damages in each state to reveal the regional differences.

```
#filter by the event with the max damage in each state
agg.gb <- group_by(stormdata.agg, STATE) %>%
  filter(TOTALDMG == max(TOTALDMG)) %>%
  filter(TOTALDMG > 0)  %>%  filter(TOTALDMG > 0)  %>%
  filter(STATE %in% state.abb ) %>% select(region=STATE, value=EVENT)

#plot the map with most dangerous event in each state by total damages
agg.gb$region <- as.character(agg.gb$region)
agg.gb$region <- tolower(state.name[match(agg.gb$region,state.abb)])

choro = StateChoropleth$new(agg.gb)
choro$title = "Most Damaging Weather In Each State"
choro$ggplot_scale = scale_fill_manual(name="Weather Event",values=c("#d73027", "#f46d43", "#fdae61","#
choro$render()
```

Most Damaging Weather In Each State