Economic and population health impact of storm events in the USA

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.
This data analysis address the following questions:
1. Across the United States, which types of storm events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?
Answers to these questions are based on the analysis of U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database, which is available by link https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2 (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2).

Data Processing

Data processing begins in assumption than data file **StormData.csv.bz2** is in current R working directory.

```
sd <- read.csv(bzfile("StormData.csv.bz2"))
```

According to National Weather Service Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf) (2.1.1 Storm Data Event Table) there are 48 event types. Variable *EVTYPE* in data file contains more unique variables.

```
length(unique(sd$EVTYPE))
```

```
## [1] 985
```

Below is the code, which reduce event type to documentation list and save the redused list to variable *ClrType*. Values, which can't be recognized, had been assigned to "Other" type.

```
sd$ClrType <- NA
sd[grepl("HURRICANE", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Hurricane (Typhoon)"
sd[grepl("TYPHOON", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Hurricane (Typhoon)"
sd[grepl("Astronomical Low Tide", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Astronomical Low Tide"
sd[grepl("Avalanche|AVALANCE", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Avalanche"
sd[grepl("Blizzard", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Blizzard"
sd[grepl("Wildfire", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Wildfire"
sd[grepl("Dense Fog", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dense Fog"
sd[grepl("^Fog$", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dense Fog"
sd[grepl("Smoke", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dense Smoke"
sd[grepl("Fog", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Freezing Fog"
sd[grepl("Drought|Dry", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Drought"
sd[grepl("Dust Storm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dust Storm"
sd[grepl("DustStorm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dust Storm"
sd[grepl("Dust", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Dust Devil"
sd[grepl("EXTREME HEAT|EXCESSIVE HEAT|RECORD HEAT", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Excessive Heat"
sd[grepl("HEAT", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Heat"
sd[grepl("TORNADO", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Tornado"
sd[grepl("COLD", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Extreme Cold/Wind Chill"
sd[grepl("Frost|Freeze", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Frost/Freeze"
sd[grepl("Cloud|FUNNEL", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Funnel Cloud"
sd[grepl("Seiche", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Seiche"
sd[grepl("Tide", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Storm Surge/Tide"
sd[grepl("Lakeshore Flood", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Lakeshore Flood"
sd[grepl("Coast.*Flood", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Coastal Flood"
sd[grepl("Flash.*Floo.*", sd$EVTYPE, ignore.case = TRUE) &
```

```
     is.na(sd$ClrType),"ClrType"] <- "Flash Flood"
sd[grepl("Flood", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Flood"
sd[grepl("Lightning|LIGHTING", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Lightning"
sd[grepl("Tropical Depression", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Tropical Depression"
sd[grepl("Tropical Storm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Tropical Storm"
sd[grepl("Tsunami", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Tsunami"
sd[grepl("Volcanic", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Volcanic Ash"
sd[grepl("Waterspout", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Waterspout"
sd[grepl("Lake.*Snow", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Lake-Effect Snow"
sd[grepl("Snow", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Heavy Snow"
sd[grepl("Winter Storm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Winter Storm"
sd[grepl("Winter", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Winter Weather"
sd[grepl("Marine Hail", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Marine Hail"
sd[grepl("Marine Thunderstorm Wind", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Marine Thunderstorm Wind"
sd[grepl("Marine High Wind", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Marine High Wind"
sd[grepl("Marine", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Marine Strong Wind"
sd[grepl("Thunderstorm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Thunderstorm Wind"
sd[grepl("Rip Current", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Rip Current"
sd[grepl("Surf", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "High Surf"
sd[grepl("Ice Storm", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Ice Storm"
sd[grepl("Ice|ICY", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Winter Weather"
sd[grepl("Sleet", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Sleet"
sd[grepl("Hail", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Hail"
sd[grepl("Rain", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Heavy Rain"
sd[grepl("High.*Wind", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "High Wind"
sd[grepl("Wind.*Chill", sd$EVTYPE, ignore.case = TRUE) &
```

```
     is.na(sd$ClrType),"ClrType"] <- "Cold/Wind Chill"
sd[grepl("^ ?TSTM|^TH[A-Za-z]*M|TUNDERSTORM", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Thunderstorm Wind"
sd[grepl("WIND", sd$EVTYPE, ignore.case = TRUE) &
     is.na(sd$ClrType),"ClrType"] <- "Strong Wind"
sd[is.na(sd$ClrType),"ClrType"] <- "Other"
```

Property damage and crop damage is stored in two pairs of varible. One variable stores value and another stores metric prefix. To convenience of this analysis damage amount is calculated to one variable.

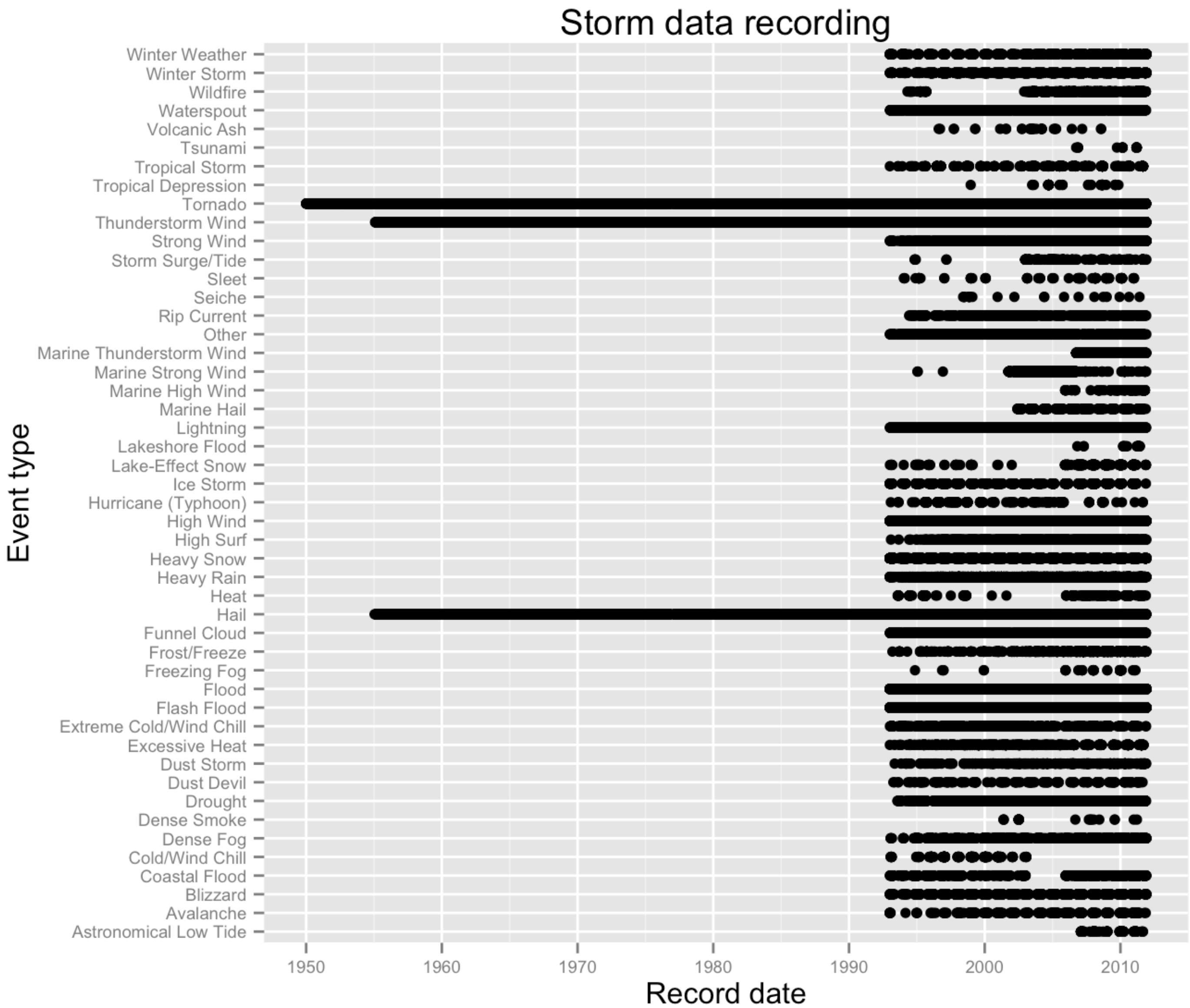
```
sd$ClrPROPDMG <- NA
suppressWarnings(power <- !is.na(as.numeric(as.character(sd$PROPDMGEXP))))
sd[power,"ClrPROPDMG"] <-
  sd[power,"PROPDMG"]*10^as.numeric(as.character(sd[power,"PROPDMGEXP"]))
power <- sd$PROPDMGEXP == "B"
sd[power,"ClrPROPDMG"] <- sd[power,"PROPDMG"]*10^9
power <- sd$PROPDMGEXP %in% c("M","m")
sd[power,"ClrPROPDMG"] <- sd[power,"PROPDMG"]*10^6
power <- sd$PROPDMGEXP == "K"
sd[power,"ClrPROPDMG"] <- sd[power,"PROPDMG"]*10^3
power <- sd$PROPDMGEXP %in% c("H","h")
sd[power,"ClrPROPDMG"] <- sd[power,"PROPDMG"]*10^2
power <- is.na(sd$ClrPROPDMG)
sd[power,"ClrPROPDMG"] <- sd[power,"PROPDMG"]

sd$ClrCROPDMG <- NA
suppressWarnings(power <- !is.na(as.numeric(as.character(sd$CROPDMGEXP))))
sd[power,"ClrCROPDMG"] <-
  sd[power,"CROPDMG"]*10^as.numeric(as.character(sd[power,"CROPDMGEXP"]))
power <- sd$CROPDMGEXP == "B"
sd[power,"ClrCROPDMG"] <- sd[power,"CROPDMG"]*10^9
power <- sd$CROPDMGEXP %in% c("M","m")
sd[power,"ClrCROPDMG"] <- sd[power,"CROPDMG"]*10^6
power <- sd$CROPDMGEXP == "K"
sd[power,"ClrCROPDMG"] <- sd[power,"CROPDMG"]*10^3
power <- is.na(sd$ClrCROPDMG)
sd[power,"ClrCROPDMG"] <- sd[power,"CROPDMG"]
```

Now we can look at the data in the context of the types of events and dates on which they were recorded.

```
library("ggplot2")
library(scales)

ggplot(sd, aes(x=ClrType, y= as.Date(BGN_DATE, format = "%m/%d/%Y"))) +
  geom_point() +
  ylab("Record date") +
  xlab("Event type") +
  ggtitle("Storm data recording") +
  theme(axis.text.x=element_text(size=7),
        axis.text.y=element_text(size=7)) +
  coord_flip()
```



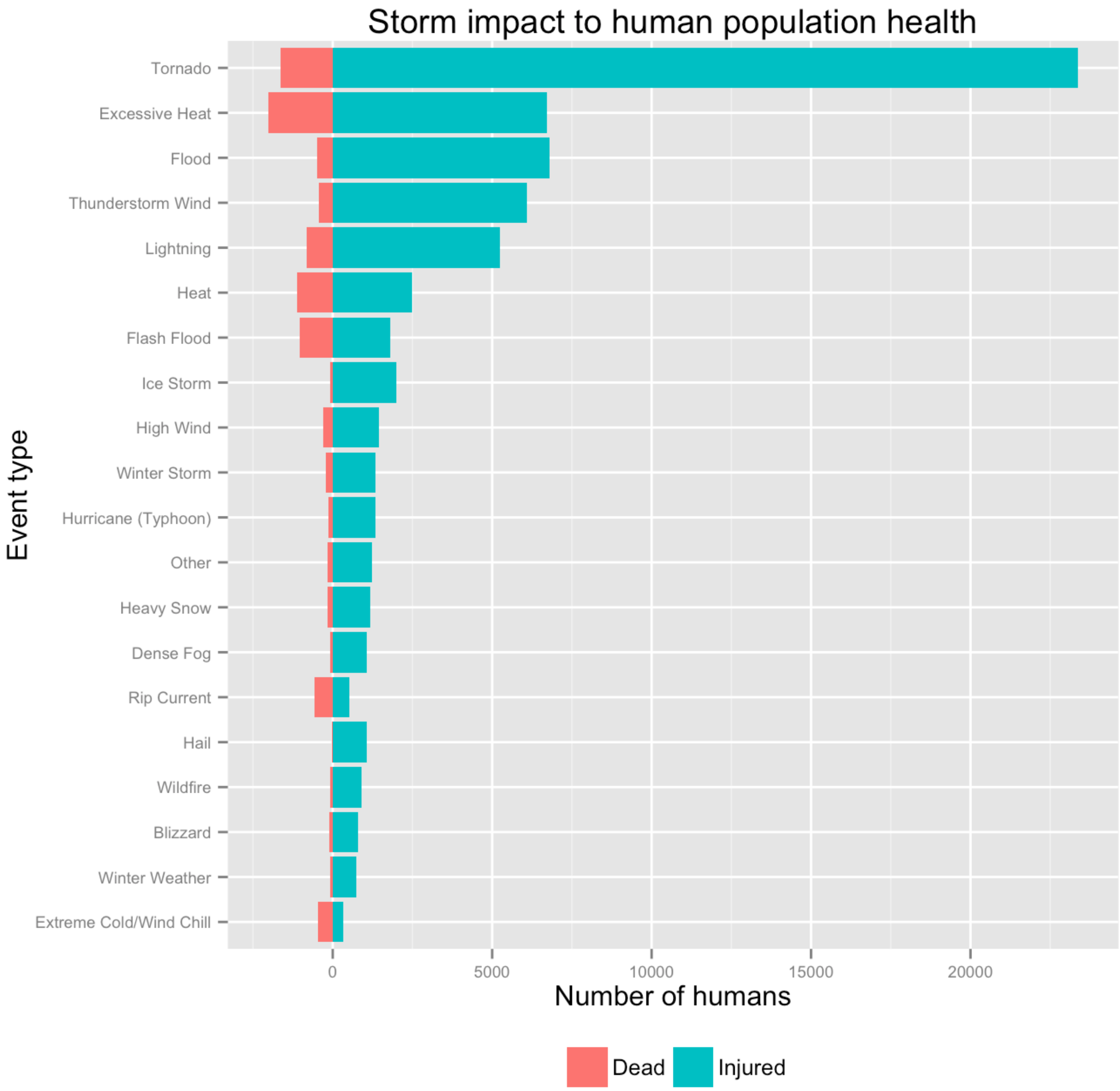
The plot shows that data of only few event types had been recorded till year of 1993. So for the purpose of this analysis we'll consider only data, which was recorded from Jan 1, 1993.

Results

Below is the barplot which answers the question of which types of events are most harmful to population health.

```
sd93 <- sd[as.Date(sd$BGN_DATE, format="%m/%d/%Y") >= as.Date("1993-01-01"),]
EventType <- aggregate(cbind(sd93[,c("FATALITIES", "INJURIES",
                                     "ClrPROPDMG", "ClrCROPDMG")],1),
                       by=list(ClrType=sd93$ClrType), sum)
EventType$ClrType <- factor(EventType$ClrType,
                            levels = EventType$ClrType[order(EventType$FATALITIES +
                                                               EventType$INJURIES)])
 tst<- EventType
EventType <- EventType[order(EventType$FATALITIES + EventType$INJURIES, decreasing = TRU
E),]
 EventType <- EventType[1:20,]

ggplot(EventType, aes(x=ClrType)) +
  geom_bar(aes(y=-FATALITIES,fill="Dead"), stat="identity", position = "identity") +
  geom_bar(aes(y=INJURIES,fill="Injured"), stat="identity", position = "identity") +
  guides(fill = guide_legend(title = "")) +
  ylab("Number of humans") +
  xlab("Event type") +
  ggtitle("Storm impact to human population health") +
  theme(axis.text.x=element_text(size=7),
        axis.text.y=element_text(size=7)) +
  theme(legend.position="bottom") +
  coord_flip()
```



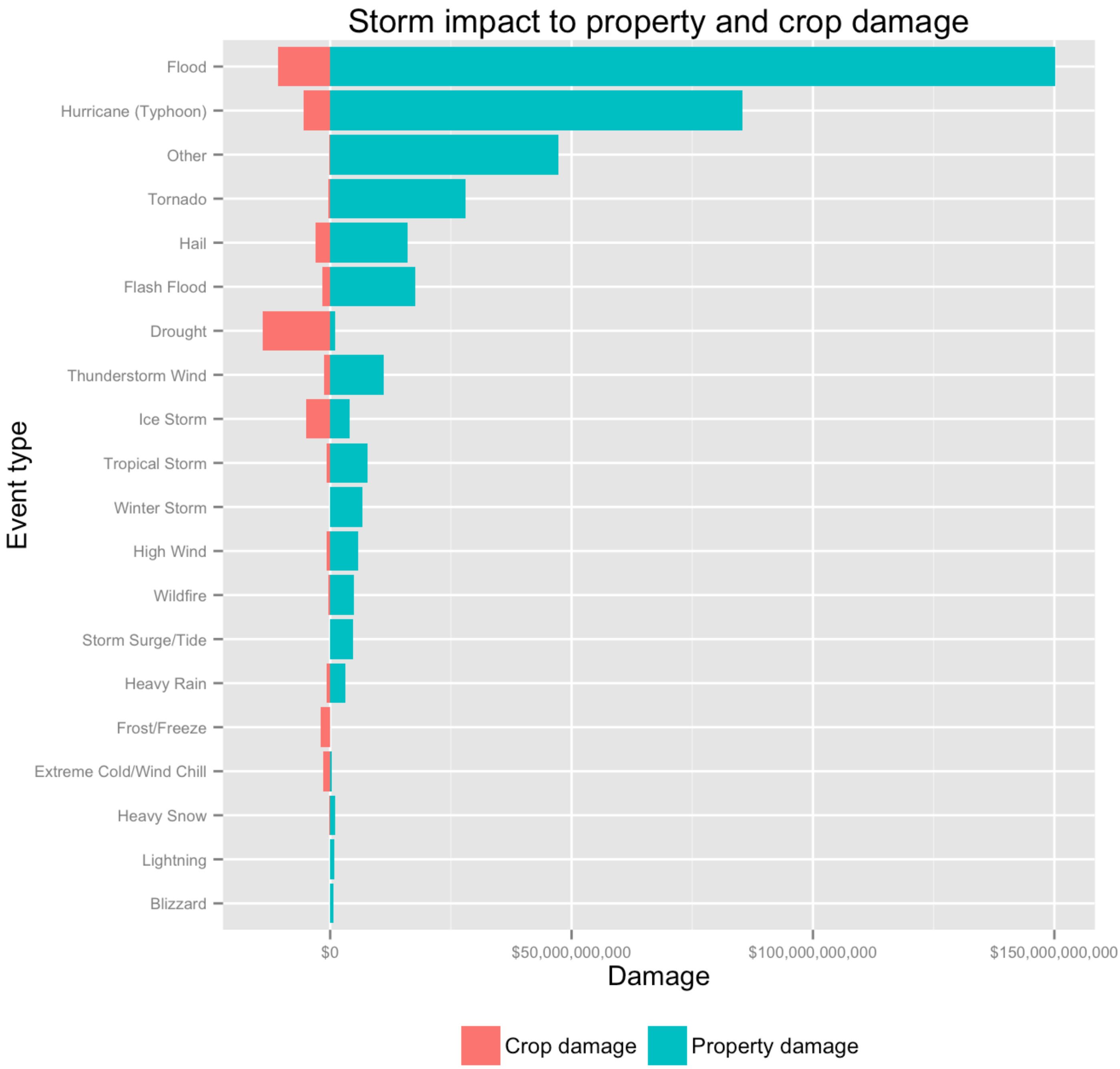
Below is the barplot which answers the question of which types of events have the greatest economic consequences.

```
 EventType<- tst

EventType$ClrType <- factor(EventType$ClrType,
                            levels = EventType$ClrType[order(EventType$ClrPROPDMG +
                                                                  EventType$ClrCROPDMG)])

EventType <- EventType[order(EventType$ClrPROPDMG + EventType$ClrCROPDMG, decreasing = TRU
E),]
EventType <- EventType[1:20,]

ggplot(EventType, aes(x=ClrType)) +
  geom_bar(aes(y=-ClrCROPDMG,fill="Crop damage"), stat="identity", position = "identity")
+
  geom_bar(aes(y=ClrPROPDMG,fill="Property damage"), stat="identity", position = "identit
y") +
  guides(fill = guide_legend(title = "")) +
  ylab("Damage") +
  xlab("Event type") +
  scale_y_continuous(labels = dollar) +
  ggtitle("Storm impact to property and crop damage") +
  theme(axis.text.x=element_text(size=7),
        axis.text.y=element_text(size=7)) +
  theme(legend.position="bottom") +
  coord_flip()
```


Storm impact to property and crop damage
Flood
Hurricane (Typhoon)
Other
Tornado
Hail
Flash Flood
Drought
Thunderstorm Wind
Ice Storm
Tropical Storm
Winter Storm
High Wind
Wildfire
Storm Surge/Tide
Heavy Rain
Frost/Freeze
Extreme Cold/Wind Chill
Heavy Snow
Lightning
Blizzard
Event type
$0
$50,000,000,000
$100,000,000,000
$150,000,000,000
Damage
Crop damage
Property damage