Weather Events Damage Analysis

Synopsis

This is a report of damages caused by various weather events in the U.S. For this report we used data provided by U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, it stores data of when and where they happen, and how damaging those events were in terms of fatalities, injuries and property and crops damage resulting from these events.

The present document analyses which of those weather events caused the most damage in population health as well as economic cost. This damages caused by weather events are exponentially distributed. We found that most injuries and fatalities are caused by tornados, while in the economic damage sector, the most damage in property historically comes from floods, and the most damage done to crops is caused by droughts.

Data Processing

- Unzip the data and place it in the current R working directory (change working directory if needed)
- Read data into R using read.csv()
- Remove unwanted symbols in the variable names and convert variable names to lower cases to be able to cover all data when looking for matches on a specific character string.
- Clean the values for the "evtype" variable by doing these steps:

    1. Remove all irregular and unwanted symbols in the observations.
    2. Remove all values shorter than 3 characters, this ensures that values that are too short to actually be an event type get excluded.
    3. Remove all values that start with the word "summary", this ensures we will only work with individual observations.
    4. Group all similar events under a common variable name for each weather event, this ensures that we won't repeat events only because of a small difference in the name of the variables (for example grouping all types of wind under "wind").
    5. Group all events that account for too few observations under a new variable called "others".

4.

- Calculate damage values for property damage and crop damage and store them under new variables.

Load Data

```
# Read data file
data <- read.csv("repdata-data-StormData.csv", header = T)

# Remove unwanted symbols in the variable names and convert them to lower cases

names(data) <- tolower(names(data))
names(data) <- gsub("_","",names(data))
```

Clean Values for the “evtype” variable

According to the dataset [documentation]() the event names should be included in the fifty event types listed.

```
# Remove all irregular and unwanted symbols in the observations
data$evtype <- tolower(as.character(data$evtype))
data$evtype <- gsub("^(([^:]+)://)?([^:/]+)(:([0-9]+))?(/.*)","",data$evtype)
#  Remove all values shorter than 3 characters
data <- subset(data,nchar(data$evtype)>=2)
#  Remove all values that start with the word"summary"
data$evtype[grep("summary", data$evtype)] <- "tbm"
data <- subset(data,data$evtype != "tbm")

#  Group all similar events
data$evtype[grep("hail", data$evtype)] <- "hail"
data$evtype[grep("wind", data$evtype)] <- "wind"
data$evtype[grep("tornado", data$evtype)] <- "tornado"
data$evtype[grep("flood", data$evtype)] <- "flood"
data$evtype[grep("lightning", data$evtype)] <- "lightning"
data$evtype[grep("snow", data$evtype)] <- "snow"
```

```
data$evtype[grep("rain", data$evtype)] <- "rain"
data$evtype[grep("winter", data$evtype)] <- "winter"
data$evtype[grep("heat", data$evtype)] <- "heat"
data$evtype[grep("fog", data$evtype)] <- "fog"
data$evtype[grep("surf", data$evtype)] <- "surf"
data$evtype[grep("ice storm", data$evtype)] <- "ice storm"
data$evtype[grep("fire", data$evtype)] <- "wild fire"
data$evtype[grep("storm surge", data$evtype)] <- "strom surge"
data$evtype[grep("hurricane", data$evtype)] <- "hurricane"
data$evtype[grep("drought", data$evtype)] <- "drought"
data$evtype[grep("thunderstorm", data$evtype)] <- "thunderstorm"
```

We can calculate that these events account for 97.8% of all observations, so we will group all other events and label this new category as “others”.

```
sum(data$evtype %in%
c("flood","wind","snow","tornado","hail","rain","lightning","winter","fog","heat","surf","ice storm","wild
fire","storm surge","hurricane ","drought","thunderstorm"))/nrow(data)

tbc <- data$evtype %in%
c("flood","wind","snow","tornado","hail","rain","lightn ing","winter","heat","surf","fog","ice storm","wild
fire","storm surge","hurric ane","drought","thunderstorm") == F

data$evtype[tbc == T] <- "others"
```

By looking at this data ordered by event type, we can clearly see how many observations of each type are in the data.

```
sort(table(data$evtype))
```

Calculating and accounting for each damage type

```
# Change quantity markers for actual quantities
data$propdmgexp <- as.character(data$propdmgexp)
data$propdmgexp[grep("K", data$propdmgexp)] <- "1000"
data$propdmgexp[grep("M", data$propdmgexp)] <- "1000000"
data$propdmgexp[grep("m", data$propdmgexp)] <- "1000000"
data$propdmgexp[grep("B", data$propdmgexp)] <- "1000000000"
tbc <- data$propdmgexp %in% c("1000","1000000","1000000000") == F
data$propdmgexp[tbc == T] <- "1"
data$propdmgexp <- as.numeric(data$propdmgexp)

# Change quantity markers for actual quantities
data$cropdmgexp <- as.character(data$cropdmgexp)
data$cropdmgexp[grep("K", data$cropdmgexp)] <- "1000"
data$cropdmgexp[grep("M", data$cropdmgexp)] <- "1000000"
data$cropdmgexp[grep("m", data$cropdmgexp)] <- "1000000"
data$cropdmgexp[grep("B", data$cropdmgexp)] <- "1000000000"
tbc <- data$cropdmgexp %in% c("1000","1000000","1000000000") == F
data$cropdmgexp[tbc == T] <- "1"
data$cropdmgexp <- as.numeric(data$cropdmgexp)
```

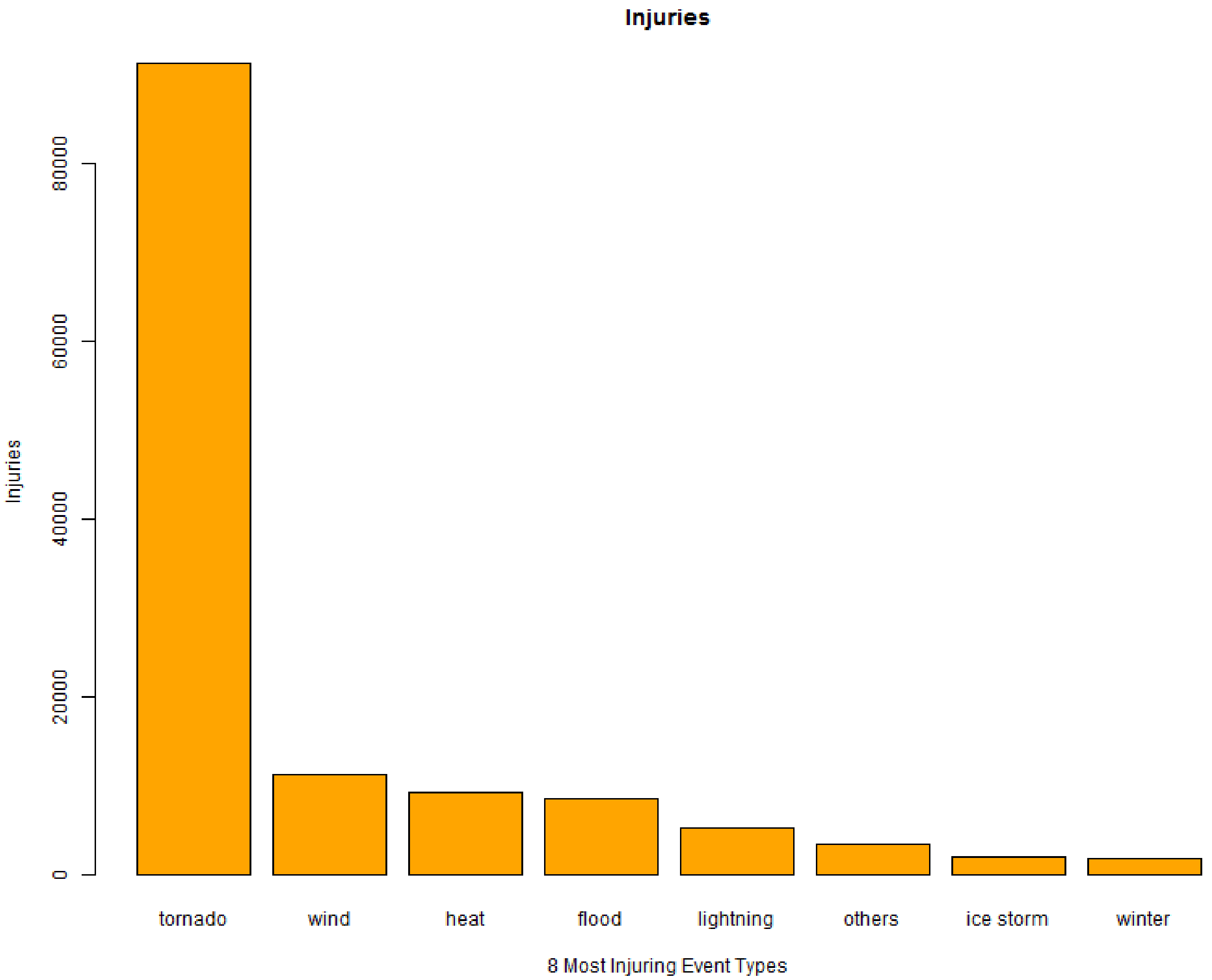
Calculate property damage and crop damage values and store them in new variables labeled "propdamage" for property damage expenses and "cropdamage" for crop damage expenses.

```
data$propdamage <- data$propdmg * data$propdmgexp
data$cropdamage <- data$cropdmg * data$cropdmgexp
```

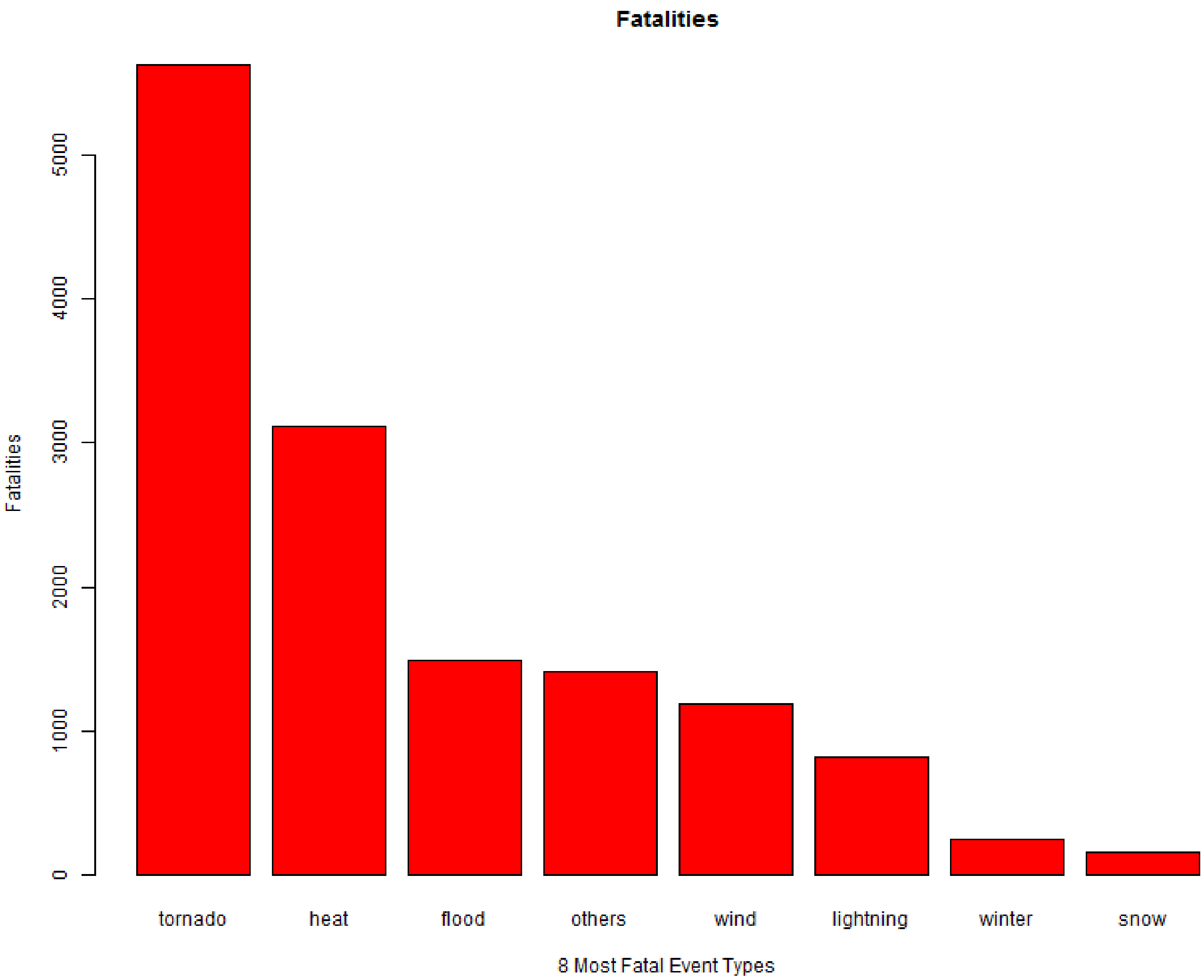
Results

Damages to Population Health

- Injuries: Tornados are the cause for the majority of injuries(66.1%) among all weather events. If we look at the five weather events that cause more injuries to the population, which are tornados, wind, heat, floods and lightnings, in that order, they account for over 90% of all injuries.



Fatalities: Tornados are too the main cause for fatalities (38.4%) in the population among all weather events. The most damaging weather events not labeled as “others” are responsible for over 78% of fatalities recorded, these are: tornados, heat, floods and wind, in that order.



It is clear that the three most damaging weather events to the health of the general population in the U.S. are tornados, heat and flood in that order, however, tornados remain well above floods and heat as the most dangerous weather event.

```
#  Get the event that caused the most injuries
#  totalInjuries <- tapply(data$injuries, data$evtype, sum)
sort(totalInjuries, decreasing = T)[1]
#  Calculate the percentaje of total injuries caused by the top event in this area
sum(sort(totalInjuries, decreasing = T)[1])/sum(totalInjuries)
#  Get the 5 events that cause the most injuries
sort(totalInjuries, decreasing = T)[1:5]
#  Calculate the percentage of injuries that the top 5 events represent
sum(sort(totalInjuries, decreasing = T)[1:5])/sum(totalInjuries)
#  Get the event that caused the most fatalities
totalFatal <- tapply(data$fatalities, data$evtype, sum)
sort(totalFatal, decreasing = T)[1]
#  Calculate the percentaje of total fatalities caused by the top event in this area
sum(sort(totalFatal, decreasing = T)[1])/sum(totalFatal)
#  Get the 5 events that cause the most fatalities
sort(totalFatal, decreasing = T)[1:5]
# Calculate the percentage of fatalities that the top events represent
sum(sort(totalFatal, decreasing = T)[1:3])/sum(totalFatal) + sum(sort(totalFatal, decreasing =
T)[5])/sum(totalFatal)
```

Figure 1: Injuries

Plotting Injuries

```
barplot(sort(totalInjuries, decreasing = T)[1:8], main = "Injuries", col="orang e",xlab="8 Most Injuring
Event Types", ylab="Injuries")
```

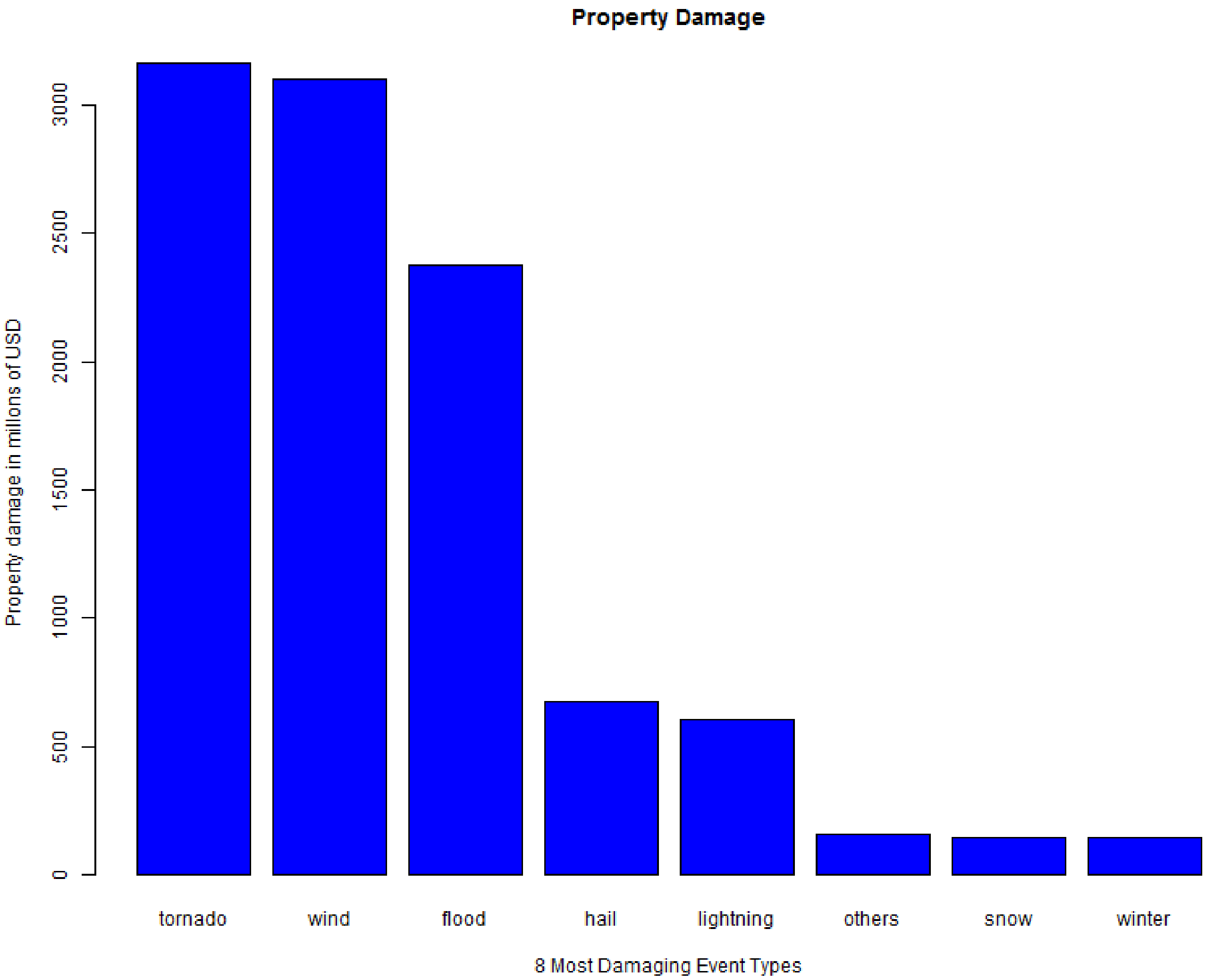
Figure 2: Fatalities

Plotting Fatalities

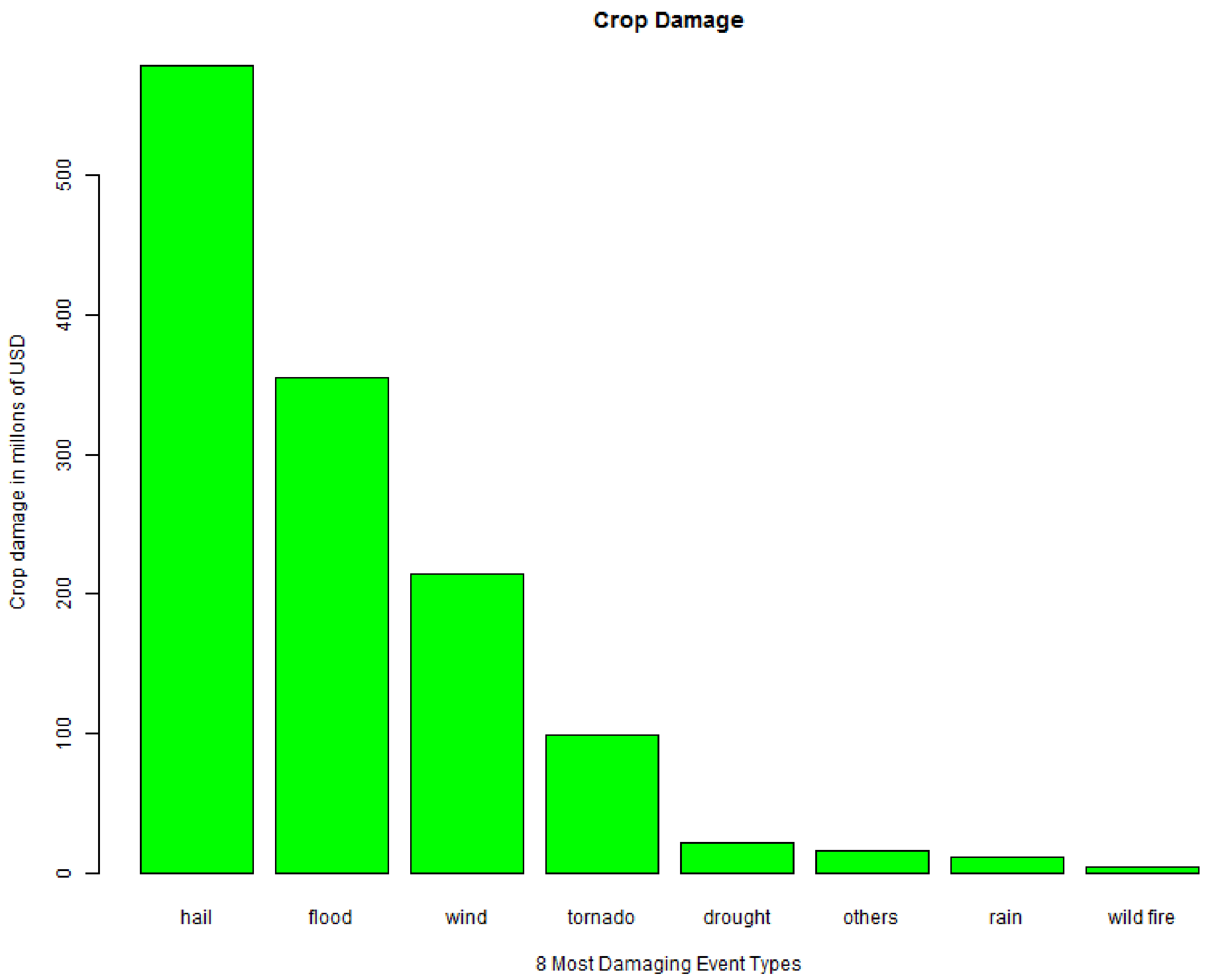
```
barplot(sort(totalFatal, decreasing = T)[1:8], main = "Fatalities", col="red",x lab="8 Most Fatal Event
Types", ylab="Fatalities")
```

Economic Cost

- Porperty Damage: Floods are the main cause for porperty damage(48%) among all weather events. If we account the property damage caused by floods, tornados, hail and wind, they account for over 74% of the property damage caused by all weather events.



- Crop Damage: The main cause for crop damages among all weather events is drought(31%). If we account for the crop damage caused by droughts, floods, ice storms, and hail, they account for over 76% of crop damages caused by all weather events.




```
# Get the event that caused the most property damage
totalPropDamage <- tapply(data$propdamage, data$evtype, sum)
sort(totalPropDamage, decreasing = T)[1]
# Calculate the percentaje of property damage caused by the top event
sort(totalPropDamage, decreasing = T)[1]/sum(totalPropDamage)
# Get the 5 events that cause the most property damage
sort(totalPropDamage, decreasing = T)[1:5]
# Calculate the percentage of damages caused by the top causes not labeled "others"
```

```
sum(sort(totalPropDamage, decreasing = T)[1:2])/sum(totalPropDamage) +
sum(sort(totalPropDamage, decreasing = T)[4:5])/sum(totalPropDamage)
#  Get the event that caused the most crop damage
totalCropDamage <- tapply(data$cropdamage, data$evtype, sum)
sort(totalCropDamage, decreasing = T)[1]
#  Calculate the percentaje of crop damage caused by the top event
sort(totalCropDamage, decreasing = T)[1]/sum(totalCropDamage)
#  Get the 5 events that cause the most crop damage
sort(totalCropDamage, decreasing = T)[1:5]
# Calculate the percentage of damages caused by the top causes not labeled "oth ers"
sum(sort(totalCropDamage, decreasing = T)[1:3])/sum(totalCropDamage) +
sum(sort(totalCropDamage, decreasing = T)[5])/sum(totalCropDamage)
```

Figure 3: Property Damage

Plotting Property Damage

```
reducedPropDamage <- totalPropDamage/1000000
barplot(sort(reducedPropDamage, decreasing = T)[1:8], main = "Property Damage", col="blue",xlab="8
Most Damaging Event Types", ylab="Property damage in millons of USD")
```

Figure 4: Crop Damage

Plotting Crop Damage

```
reducedCropDamage <- totalCropDamage/1000000
barplot(sort(reducedCropDamage, decreasing = T)[1:8], main = "Crop Damage", col ="green",xlab="8
Most Damaging Event Types", ylab="Crop damage in millons of US D")
```