RRPA2

Synopsis

The purpose of this report is to investigate the types of event are most harmful to the population health and economy. Based on the analysis of the data, it is observed that the top aggregate impact to population health and property to be Tornados and floods have caused the most significant economic damage.

Data Processing

Environment Initialization

```
library(ggplot2)
library(xtable)
```

```
## Warning: package 'xtable' was built under R version 3.1.1
```

```
setwd("C:/Users/Xiaoyu Liu/Dropbox/Onlinecourse/DataAnalysisonline/reproductive")
raw.data.file.name <- "storm-raw-data.csv.bz2"

if (!file.exists(raw.data.file.name)) {
    download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",raw.data.fi
}
```

Load the data

The following fields were used for this analysis:

Field Description BGN_DATE Date the storm event began EVTYPE Type of storm event (not abbreviated) FATALITIES The number of deaths directly related to the weather event INJURIES The number of injuries directly related to the weather event PROPDMG The estimated amount of damage to property incurred by the weather event (whole numbers and hundredths) PROPDMGEXP A multiplier where “H” denotes hundreds, ’K" denotes thousands, “M” denotes millions, and “B” denotes billions CROPDMG The estimated amount of damage to crops incurred by the weather event (whole numbers and hundredths) CROPDMGEXP A multiplier where “H” denotes hundreds, ’K" denotes thousands, “M” denotes millions, and “B” denotes billions

```
data.file.name <- "storm-data.csv"

if (!file.exists(data.file.name)) {
    raw.data <- read.csv(bzfile(raw.data.file.name), header = TRUE, stringsAsFactors = FALSE)
    raw.data$BGN_DATE <- as.Date(raw.data$BGN_DATE, format = "%m/%d/%Y %H:%M:%S")
    # Horizontal subset
    data <- raw.data[raw.data$BGN_DATE >= as.Date("1993-01-01"), ]
    data <- data[data$INJURIES > 0 | data$FATALITIES > 0 | data$PROPDMG > 0 | data$CROPDMG > 0, ]
    # Vertical subset - limit the data to the required fields.
    data <- data[, c("BGN_DATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDMG"
    # Transform PROPDMG multiplier
```

```
    data$PROPDMGEXP[is.na(data$PROPDMGEXP)] <- 0
    data$PROPDMGEXP[data$PROPDMGEXP == ""] <- 1
    data$PROPDMGEXP[grep("[-+?]", data$PROPDMGEXP)] <- 1
    data$PROPDMGEXP[grep("[Hh]", data$PROPDMGEXP)] <- 100
    data$PROPDMGEXP[grep("[Kk]", data$PROPDMGEXP)] <- 1000
    data$PROPDMGEXP[grep("[Mm]", data$PROPDMGEXP)] <- 1e+06
    data$PROPDMGEXP[data$PROPDMGEXP == "B"] <- 1e+09
    data$PROPDMGEXP <- as.numeric(data$PROPDMGEXP)
    # Transform CROPDMG multiplier
    data$CROPDMGEXP[is.na(data$CROPDMGEXP)] <- 0
    data$CROPDMGEXP[data$CROPDMGEXP == ""] <- 1
    data$CROPDMGEXP[grep("[-+?]", data$CROPDMGEXP)] <- 1
    data$CROPDMGEXP[grep("[Hh]", data$CROPDMGEXP)] <- 100
    data$CROPDMGEXP[grep("[Kk]", data$CROPDMGEXP)] <- 1000
    data$CROPDMGEXP[grep("[Mm]", data$CROPDMGEXP)] <- 1e+06
    data$CROPDMGEXP[data$CROPDMGEXP == "B"] <- 1e+09
    data$CROPDMGEXP <- as.numeric(data$CROPDMGEXP)
    # Adjust property and crop damages
    data$PROPDMG <- data$PROPDMG * data$PROPDMGEXP
    data$CROPDMG <- data$CROPDMG * data$CROPDMGEXP

    # Clean Event Type
    data$EVTYPE <- toupper(data$EVTYPE)
    # Trim multiple spaces and leading, trailing spaces
    trim <- function(x) gsub("[[:space:]]{2,}", " ", gsub("(^[[:space:]]+|[[:space:]]+$)", "", x))
    data$EVTYPE <- trim(data$EVTYPE)
    write.csv(data, file = data.file.name, row.names = FALSE)
} else {
    data <- read.csv(data.file.name, header = TRUE, stringsAsFactors = FALSE)
}
```

Results

Question 1. Across the United States, which types of events (as indicated in the [EVTYPE] variable) are most harmful with respect to human health?

```
population.health.data <- data[, c("EVTYPE", "INJURIES", "FATALITIES")]
population.health.data <- aggregate.data.frame(x = population.health.data[, c(2, 3)], by = list(populat
colnames(population.health.data)[1] <- "EVTYPE"
population.health.data$TOTAL <- population.health.data$FATALITIES + population.health.data$INJURIES
population.health.data <- population.health.data[order(population.health.data$TOTAL, decreasing = TRUE)
row.names(population.health.data) <- population.health.data$EVTYPE
population.health.data <- population.health.data[population.health.data$TOTAL != 0, ]
```

```
population.health.table <- xtable(population.health.data[1:10, c("INJURIES", "FATALITIES")], caption =
names(population.health.table) <- c("Injuries", "Fatalities")
print(population.health.table, type = "html")
```

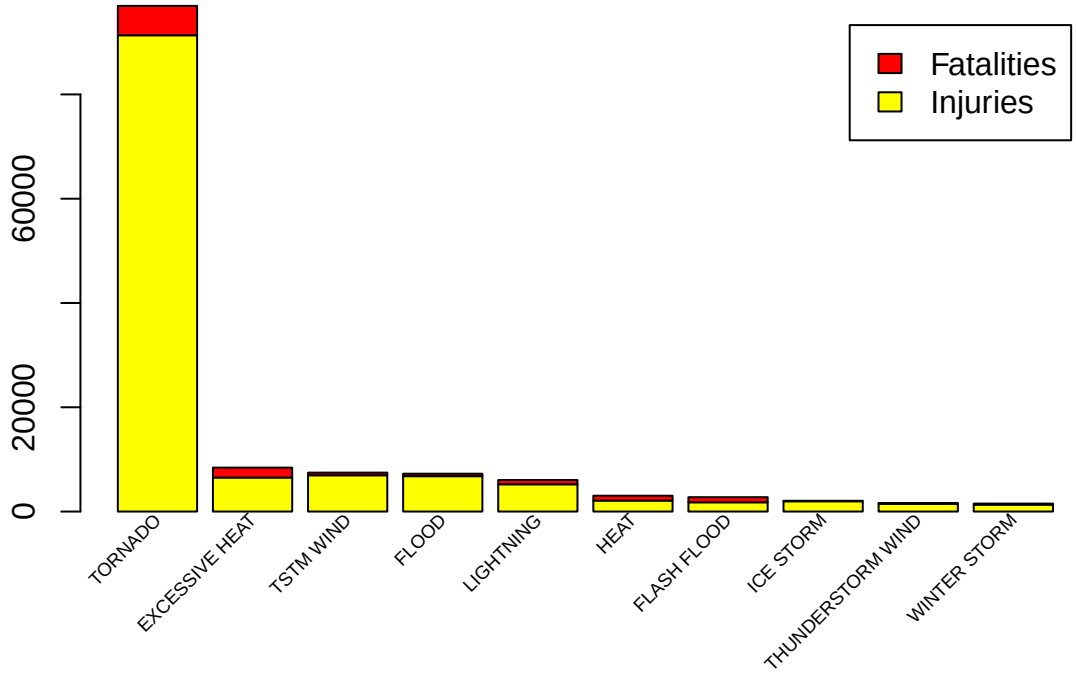
```
## <!-- html table generated in R 3.1.0 by xtable 1.7-3 package -->
## <!-- Sun Jul 27 20:49:40 2014 -->
## <TABLE border=1>
## <CAPTION ALIGN="bottom"> Table 1. Event types that accounted for the highest cumulative injuries and
## <TR> <TH>  </TH> <TH> Injuries </TH> <TH> Fatalities </TH>  </TR>
##   <TR> <TD align="right"> TORNADO </TD> <TD align="right"> 91346.00 </TD> <TD align="right"> 5633.00
##   <TR> <TD align="right"> EXCESSIVE HEAT </TD> <TD align="right"> 6525.00 </TD> <TD align="right"> 1
##   <TR> <TD align="right"> TSTM WIND </TD> <TD align="right"> 6957.00 </TD> <TD align="right"> 504.00
##   <TR> <TD align="right"> FLOOD </TD> <TD align="right"> 6789.00 </TD> <TD align="right"> 470.00 </T
##   <TR> <TD align="right"> LIGHTNING </TD> <TD align="right"> 5230.00 </TD> <TD align="right"> 816.00
##   <TR> <TD align="right"> HEAT </TD> <TD align="right"> 2100.00 </TD> <TD align="right"> 937.00 </TD
##   <TR> <TD align="right"> FLASH FLOOD </TD> <TD align="right"> 1777.00 </TD> <TD align="right"> 978.
##   <TR> <TD align="right"> ICE STORM </TD> <TD align="right"> 1975.00 </TD> <TD align="right"> 89.00
##   <TR> <TD align="right"> THUNDERSTORM WIND </TD> <TD align="right"> 1488.00 </TD> <TD align="right"
##   <TR> <TD align="right"> WINTER STORM </TD> <TD align="right"> 1321.00 </TD> <TD align="right"> 206
##    </TABLE>
```

The following plot demonstrates that tornados are most the most harmful weather event to population health.

```
population.health.plot <- barplot(height = t(as.matrix(population.health.data[1:10,  c("INJURIES", "FATA

text(population.health.plot, par("usr")[3], labels = row.names(population.health.data)[1:10],srt = 45, a
axis(2)
```

Figure 1. Top 10 severe weather events caused injuries and fatalities
U.S. 1993 − 2011



Question 2. Across the United States, which types of events (as indicated in the [EVTYPE] variable) have the greatest economic consequences?

```
economy.data <- data[, c("EVTYPE", "PROPDMG", "CROPDMG")]
economy.data <- aggregate.data.frame(x = economy.data[, c(2, 3)], by = list(economy.data$EVTYPE),FUN = s
colnames(economy.data)[1] <- "EVTYPE"
economy.data$TOTAL <- economy.data$PROPDMG + economy.data$CROPDMG
economy.data <- economy.data[order(economy.data$TOTAL, decreasing = TRUE), ]
row.names(economy.data) <- economy.data$EVTYPE
economy.data <- economy.data[economy.data$TOTAL != 0, ]
```

```
economy.table <- xtable(economy.data[1:10, c("PROPDMG", "CROPDMG")], caption = "Table 2. Event types tha
names(economy.table) <- c("Property Damage", "Crop Damage")
print(economy.table, type = "html")
```

```
## <!-- html table generated in R 3.1.0 by xtable 1.7-3 package -->
## <!-- Sun Jul 27 20:49:56 2014 -->
## <TABLE border=1>
## <CAPTION ALIGN="bottom"> Table 2. Event types that accounted for the highest property damage and crop
## <TR> <TH>  </TH> <TH> Property Damage </TH> <TH> Crop Damage </TH>  </TR>
##   <TR> <TD align="right"> TORNADO </TD> <TD align="right"> 3212258.16 </TD> <TD align="right"> 100018
##   <TR> <TD align="right"> FLASH FLOOD </TD> <TD align="right"> 1420124.59 </TD> <TD align="right"> 17
##   <TR> <TD align="right"> TSTM WIND </TD> <TD align="right"> 1335965.61 </TD> <TD align="right"> 109
##   <TR> <TD align="right"> HAIL </TD> <TD align="right"> 688693.38 </TD> <TD align="right"> 579596.28
##   <TR> <TD align="right"> FLOOD </TD> <TD align="right"> 899938.48 </TD> <TD align="right"> 168037.88
##   <TR> <TD align="right"> THUNDERSTORM WIND </TD> <TD align="right"> 876844.17 </TD> <TD align="right
##   <TR> <TD align="right"> LIGHTNING </TD> <TD align="right"> 603351.78 </TD> <TD align="right"> 3580
##   <TR> <TD align="right"> THUNDERSTORM WINDS </TD> <TD align="right"> 446293.18 </TD> <TD align="righ
##   <TR> <TD align="right"> HIGH WIND </TD> <TD align="right"> 324731.56 </TD> <TD align="right"> 17283
##   <TR> <TD align="right"> WINTER STORM </TD> <TD align="right"> 132720.59 </TD> <TD align="right"> 19
##    </TABLE>
```

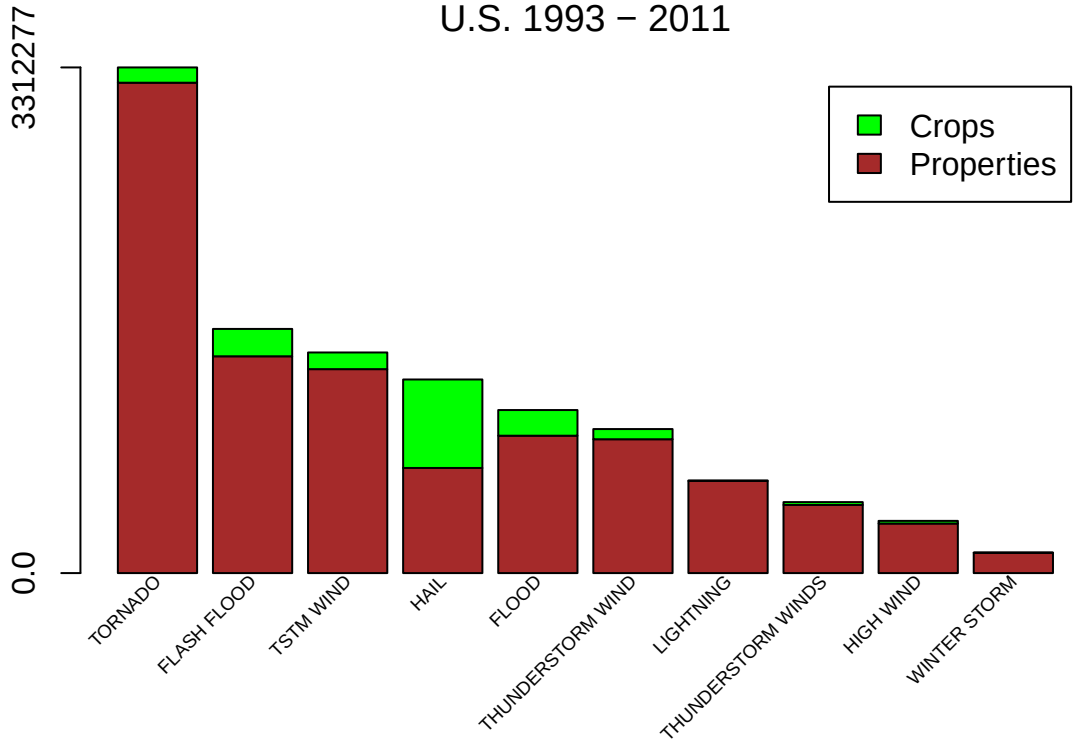
The following plot demonstrates that floods have the greatest economic damages impact on property and crops.

```
economy.plot <- barplot(height = t(as.matrix(economy.data[1:10, c("PROPDMG", "CROPDMG")])), main = bquo

text(economy.plot, par("usr")[3], labels = row.names(economy.data)[1:10], srt = 45, adj = 1, cex = 0.6,

axis(2, at = c(0, economy.data[1, "TOTAL"]), labels = c("0.0", format(economy.data[1,"TOTAL"], digits =
```

Figure 2. Top 10 weather events caused property damage
U.S. 1993 − 2011



Conclusions

Tornados have the greatest impact on the population health. Flood has the largest impact on the property and second largest impact on crops. Drought has the largest impact on the crops. Overall, flood brings the largest economic consequence.