Reproducible Research: Peer Assessment 2

Synopsis

After loading the data, we only subset the data we need like the data related to population health and property damage with crop damage. Then we sum up the result corresponding to the specific event type. In the damage data set, we need to take the exponent part into consideration. In order to better view the result, we sort the data descending and plot the first 10 result. From the data, we can see that across the United States, TORNADO is the most harmful with respect to population health, while FLOOD causes the most property damage and DRAUGHT causes the most crop damage.

Data Loading

```
# read csv
stormData <- read.csv("StormData/repdata-data-StormData.csv")
# show names to get a basic understanding
names(stormData)
```

```
##  [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"
```

Population health

```
# subset data for population health
populationHealthData <- stormData[, c("EVTYPE", "FATALITIES", "INJURIES")]
# convert factor to integer for later aggregation
populationHealthData$FATALITIES <- as.integer(populationHealthData$FATALITIES)
populationHealthData$INJURIES <- as.integer(populationHealthData$INJURIES)
# sum up the result and show
populationHealthResult <- aggregate(cbind(INJURIES, FATALITIES) ~ EVTYPE, data = populationHealthData, s
head(populationHealthResult)
```

```
##                 EVTYPE INJURIES FATALITIES
## 1    HIGH SURF ADVISORY        0          0
## 2         COASTAL FLOOD        0          0
## 3           FLASH FLOOD        0          0
## 4             LIGHTNING        0          0
## 5             TSTM WIND        0          0
## 6       TSTM WIND (G45)        0          0
```

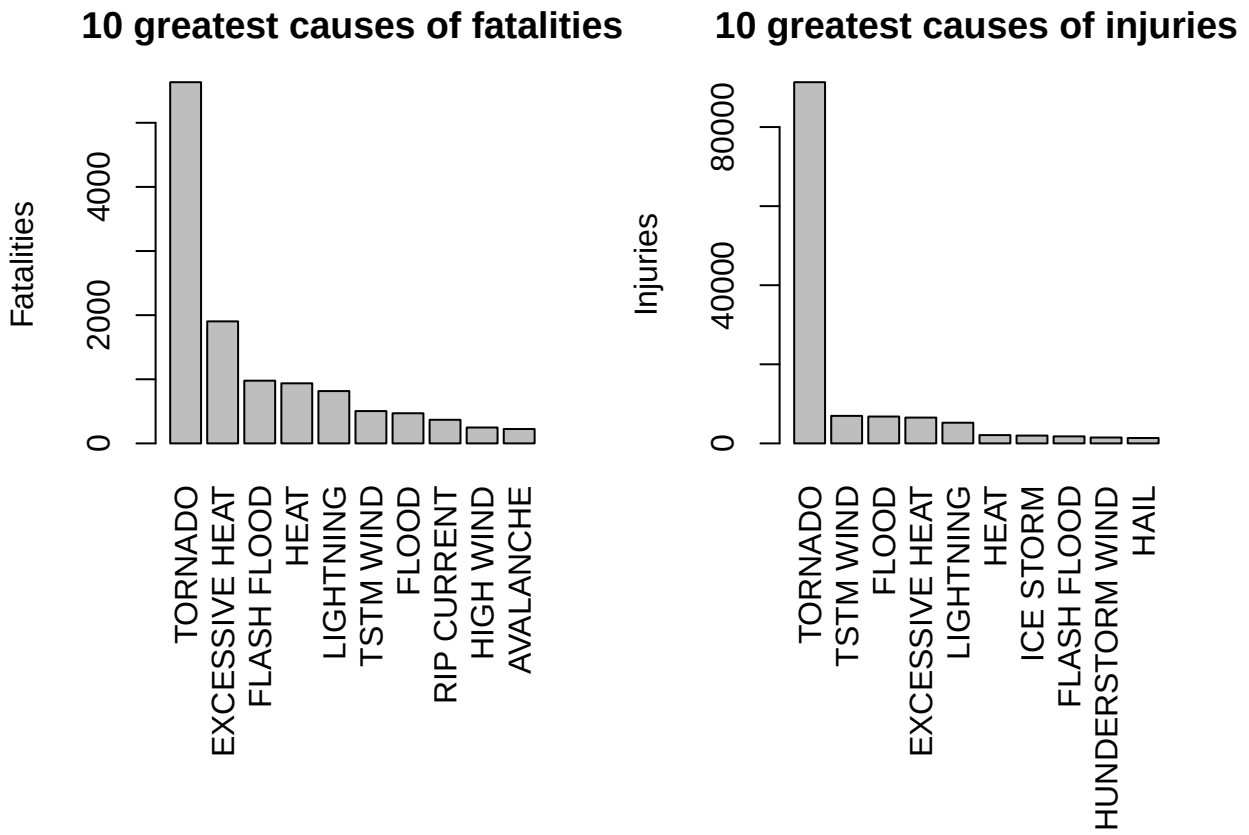
```r
# sort the data and show
sortedFatalityResult <- populationHealthResult[order(-populationHealthResult$FATALITIES), ]
head(sortedFatalityResult)
```

```
##             EVTYPE INJURIES FATALITIES
## 834        TORNADO    91346       5633
## 130 EXCESSIVE HEAT     6525       1903
## 153    FLASH FLOOD     1777        978
## 275           HEAT     2100        937
## 464      LIGHTNING     5230        816
## 856      TSTM WIND     6957        504
```

```r
sortedInjuryResult <- populationHealthResult[order(-populationHealthResult$INJURIES), ]
head(sortedInjuryResult)
```

```
##             EVTYPE INJURIES FATALITIES
## 834        TORNADO    91346       5633
## 856      TSTM WIND     6957        504
## 170          FLOOD     6789        470
## 130 EXCESSIVE HEAT     6525       1903
## 464      LIGHTNING     5230        816
## 275           HEAT     2100        937
```

```r
# plot the sorted result
par(mfrow = c(1,2), mar = c(10, 4, 3, 2))
barplot(sortedFatalityResult$FATALITIES[1:10], names.arg = sortedFatalityResult$EVTYPE[1:10], las = 3, y
barplot(sortedInjuryResult$INJURIES[1:10], names.arg = sortedInjuryResult$EVTYPE[1:10], las = 3, ylab =
```



From the plot, we can see that across the United States, TORNADO is the most harmful with respect to population health.

Economic consequences

```
# show names
names(stormData)
```

```
##  [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"
```

```
# subset data for economic consequences
economicData <- stormData[, c("EVTYPE", "PROPDMG", "PROPDMGEXP","CROPDMG", "CROPDMGEXP")]
# convert factor to integer for later aggregation
economicData$PROPDMG <- as.numeric(economicData$PROPDMG)
economicData$CROPDMG <- as.numeric(economicData$CROPDMG)

head(economicData)
```

```
##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## 1 TORNADO    25.0          K       0
## 2 TORNADO     2.5          K       0
## 3 TORNADO    25.0          K       0
## 4 TORNADO     2.5          K       0
## 5 TORNADO     2.5          K       0
## 6 TORNADO     2.5          K       0
```

```
# deal with exponent
unique(economicData$PROPDMGEXP)
```

```
##  [1] K M   B m + 0 5 6 ? 4 2 3 h 7 H - 1 8
## Levels:  - ? + 0 1 2 3 4 5 6 7 8 B h H K m M
```

```
economicData$PROPEXP[economicData$PROPDMGEXP == "K"] <- 1000
economicData$PROPEXP[economicData$PROPDMGEXP == "M"] <- 1e+06
economicData$PROPEXP[economicData$PROPDMGEXP == ""] <- 1
economicData$PROPEXP[economicData$PROPDMGEXP == "B"] <- 1e+09
economicData$PROPEXP[economicData$PROPDMGEXP == "m"] <- 1e+06
economicData$PROPEXP[economicData$PROPDMGEXP == "+"] <- 0
economicData$PROPEXP[economicData$PROPDMGEXP == "0"] <- 1
economicData$PROPEXP[economicData$PROPDMGEXP == "5"] <- 1e+05
economicData$PROPEXP[economicData$PROPDMGEXP == "6"] <- 1e+06
economicData$PROPEXP[economicData$PROPDMGEXP == "?"] <- 0
economicData$PROPEXP[economicData$PROPDMGEXP == "4"] <- 1e+04
economicData$PROPEXP[economicData$PROPDMGEXP == "2"] <- 1e+02
economicData$PROPEXP[economicData$PROPDMGEXP == "3"] <- 1e+03
economicData$PROPEXP[economicData$PROPDMGEXP == "h"] <- 1e+02
economicData$PROPEXP[economicData$PROPDMGEXP == "7"] <- 1e+07
economicData$PROPEXP[economicData$PROPDMGEXP == "H"] <- 1e+02
economicData$PROPEXP[economicData$PROPDMGEXP == "-"] <- 0
economicData$PROPEXP[economicData$PROPDMGEXP == "1"] <- 1e+01
economicData$PROPEXP[economicData$PROPDMGEXP == "8"] <- 1e+08

unique(economicData$CROPDMGEXP)
```

```
## [1]   M K m B ? 0 k 2
## Levels:  ? 0 2 B k K m M
```

```
economicData$CROPEXP[economicData$CROPDMGEXP == ""] <- 1
economicData$CROPEXP[economicData$CROPDMGEXP == "M"] <- 1e+06
economicData$CROPEXP[economicData$CROPDMGEXP == "k"] <- 1e+03
economicData$CROPEXP[economicData$CROPDMGEXP == "m"] <- 1e+06
economicData$CROPEXP[economicData$CROPDMGEXP == "B"] <- 1e+09
economicData$CROPEXP[economicData$CROPDMGEXP == "?"] <- 0
economicData$CROPEXP[economicData$CROPDMGEXP == "0"] <- 1
economicData$CROPEXP[economicData$CROPDMGEXP == "k"] <- 1e+03
economicData$CROPEXP[economicData$CROPDMGEXP == "2"] <- 1e+02

# Compute the value
economicData$PROPDMGVALUE <- economicData$PROPDMG * economicData$PROPEXP
economicData$CROPDMGVALUE <- economicData$CROPDMG * economicData$CROPEXP
```

```r
head(economicData)
```

```
##    EVTYPE PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP PROPEXP CROPEXP
## 1 TORNADO    25.0          K       0                1000       1
## 2 TORNADO     2.5          K       0                1000       1
## 3 TORNADO    25.0          K       0                1000       1
## 4 TORNADO     2.5          K       0                1000       1
## 5 TORNADO     2.5          K       0                1000       1
## 6 TORNADO     2.5          K       0                1000       1
##   PROPDMGVALUE CROPDMGVALUE
## 1        25000            0
## 2         2500            0
## 3        25000            0
## 4         2500            0
## 5         2500            0
## 6         2500            0
```

```r
# sum up the result and show
economicResult <- aggregate(cbind(PROPDMGVALUE, CROPDMGVALUE) ~ EVTYPE, data = economicData, sum)
head(economicResult)
```

```
##                  EVTYPE PROPDMGVALUE CROPDMGVALUE
## 1    HIGH SURF ADVISORY       200000            0
## 2         COASTAL FLOOD            0            0
## 3           FLASH FLOOD        50000            0
## 4             LIGHTNING            0            0
## 5             TSTM WIND      8100000            0
## 6       TSTM WIND (G45)         8000            0
```

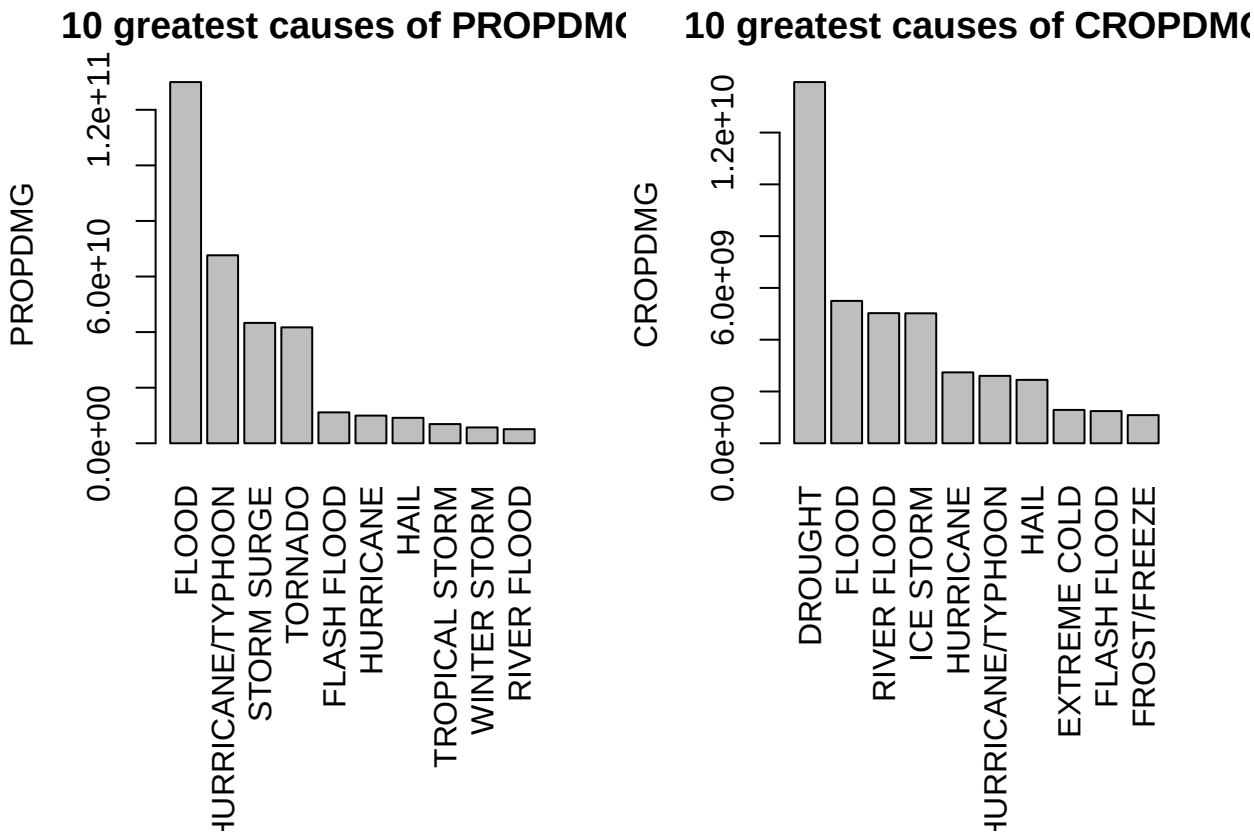
```r
# sort the data and show
sortedPropResult <- economicResult[order(-economicResult$PROPDMGVALUE), ]
head(sortedPropResult)
```

```
##                EVTYPE PROPDMGVALUE CROPDMGVALUE
## 166             FLOOD 129983590857   5499430000
## 397 HURRICANE/TYPHOON  67657180000   2604170000
## 654       STORM SURGE  43320621000            0
## 815           TORNADO  41720310877    315410160
## 150       FLASH FLOOD  11124311527   1243360000
## 388         HURRICANE   9958241010   2739310000
```

```r
sortedCropResult <- economicResult[order(-economicResult$CROPDMGVALUE), ]
head(sortedCropResult)
```

```
##                EVTYPE PROPDMGVALUE CROPDMGVALUE
## 93            DROUGHT   1012937000  13951120000
## 166             FLOOD 129983590857   5499430000
## 574       RIVER FLOOD   5063310500   5026000000
## 413         ICE STORM   3122365510   5020450000
## 388         HURRICANE   9958241010   2739310000
## 397 HURRICANE/TYPHOON  67657180000   2604170000
```

```r
# plot the sorted result
par(mfrow = c(1,2), mar = c(10, 4, 3, 2))
barplot(sortedPropResult$PROPDMGVALUE[1:10], names.arg = sortedPropResult$EVTYPE[1:10], las = 3, ylab =
barplot(sortedCropResult$CROPDMGVALUE[1:10], names.arg = sortedCropResult$EVTYPE[1:10], las = 3, ylab =
```



From the plot, we can see FLOOD causes the most property damage and DRAUGHT causes the most crop damage.