U.S. National Oceanic and Atmospheric Administration's Storm Data Analysis

OVERVIEW

Weather events cause public health and economic problems for communities and municipalities. Severe events result in fatalities, injuries, and damage. Predicting and/or preventing these outcomes is a primary objective.

This analysis examines the damaging effects of severe weather conditions (e.g. hurricanes, tornadoes, thunderstorms, floods, etc.) on human populations and the econonomy in the U.S. from 1950 to 2011.

As a result, the analysis will highlight the severe weather events associated with the greatest impact on the economy and population health.

SYNOPSIS

This is an exploration of the 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, including when and where they occur, which type of event, as well as the estimates of relevant fatalities, injuries, and various forms of damage. The dataset used in this project is provided by the U.S. National Oceanic and Atmospheric Administration (NOAA). This analysis discovered that tornados are responsible for a maximum number of fatalities and injuries. This analysis also discoered that floods are responsbile for maximum property damage, while Droughts cause maximum crop damage. Objective: Explore the NOAA Storm Database to help answer important questions about severe weather events.

DATA PROCESSING

```
# Load Libraries
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
# Load Data
storm_data <- read.csv('repdata_data_StormData.csv')
head(storm_data)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE  EVTYPE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL TORNADO
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL TORNADO
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL TORNADO
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL TORNADO
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL TORNADO
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL TORNADO
##   BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1         0                                               0         NA
## 2         0                                               0         NA
## 3         0                                               0         NA
## 4         0                                               0         NA
## 5         0                                               0         NA
## 6         0                                               0         NA
##   END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1         0                      14.0   100 3   0          0       15    25.0
## 2         0                       2.0   150 2   0          0        0     2.5
## 3         0                       0.1   123 2   0          0        2    25.0
## 4         0                       0.0   100 2   0          0        2     2.5
## 5         0                       0.0   150 2   0          0        2     2.5
## 6         0                       1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         3450      8748
##   LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1       3051       8806              1
## 2          0          0              2
## 3          0          0              3
## 4          0          0              4
## 5          0          0              5
## 6          0          0              6
```

```
summary(storm_data)
```

```
##     STATE__       BGN_DATE           BGN_TIME          TIME_ZONE
##  Min.   : 1.0   Length:902297      Length:902297      Length:902297
##  1st Qu.:19.0   Class :character   Class :character   Class :character
##  Median :30.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :31.2
##  3rd Qu.:45.0
##  Max.   :95.0
##
##      COUNTY       COUNTYNAME           STATE              EVTYPE
##  Min.   :  0.0   Length:902297      Length:902297      Length:902297
##  1st Qu.: 31.0   Class :character   Class :character   Class :character
##  Median : 75.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :100.6
##  3rd Qu.:131.0
##  Max.   :873.0
##
##    BGN_RANGE          BGN_AZI           BGN_LOCATI          END_DATE
##  Min.   :   0.000   Length:902297      Length:902297      Length:902297
##  1st Qu.:   0.000   Class :character   Class :character   Class :character
##  Median :   0.000   Mode  :character   Mode  :character   Mode  :character
##  Mean   :   1.484
##  3rd Qu.:   1.000
##  Max.   :3749.000
##
##    END_TIME           COUNTY_END COUNTYENDN       END_RANGE
##  Length:902297      Min.   :0    Mode:logical   Min.   :  0.0000
##  Class :character   1st Qu.:0    NA's:902297    1st Qu.:  0.0000
##  Mode  :character   Median :0                   Median :  0.0000
##                     Mean   :0                   Mean   :  0.9862
##                     3rd Qu.:0                   3rd Qu.:  0.0000
##                     Max.   :0                   Max.   :925.0000
##
##    END_AZI           END_LOCATI            LENGTH              WIDTH
##  Length:902297      Length:902297      Min.   :   0.0000   Min.   :   0.000
##  Class :character   Class :character   1st Qu.:   0.0000   1st Qu.:   0.000
##  Mode  :character   Mode  :character   Median :   0.0000   Median :   0.000
##                                        Mean   :   0.2301   Mean   :   7.503
##                                        3rd Qu.:   0.0000   3rd Qu.:   0.000
##                                        Max.   :2315.0000   Max.   :4400.000
##
##        F               MAG            FATALITIES          INJURIES
##  Min.   :0.0      Min.   :    0.0   Min.   :  0.0000   Min.   :   0.0000
##  1st Qu.:0.0      1st Qu.:    0.0   1st Qu.:  0.0000   1st Qu.:   0.0000
##  Median :1.0      Median :   50.0   Median :  0.0000   Median :   0.0000
##  Mean   :0.9      Mean   :   46.9   Mean   :  0.0168   Mean   :   0.1557
##  3rd Qu.:1.0      3rd Qu.:   75.0   3rd Qu.:  0.0000   3rd Qu.:   0.0000
##  Max.   :5.0      Max.   :22000.0   Max.   :583.0000   Max.   :1700.0000
##  NA's   :843563
##     PROPDMG          PROPDMGEXP           CROPDMG         CROPDMGEXP
##  Min.   :   0.00   Length:902297      Min.   :  0.000   Length:902297
##  1st Qu.:   0.00   Class :character   1st Qu.:  0.000   Class :character
##  Median :   0.00   Mode  :character   Median :  0.000   Mode  :character
##  Mean   :  12.06                      Mean   :  1.527
##  3rd Qu.:   0.50                      3rd Qu.:  0.000
##  Max.   :5000.00                      Max.   :990.000
```

```
##
##      WFO             STATEOFFIC         ZONENAMES            LATITUDE
##  Length:902297      Length:902297      Length:902297      Min.   :   0
##  Class :character   Class :character   Class :character   1st Qu.:2802
##  Mode  :character   Mode  :character   Mode  :character   Median :3540
##                                                           Mean   :2875
##                                                           3rd Qu.:4019
##                                                           Max.   :9706
##                                                           NA's   :47
##    LONGITUDE        LATITUDE_E     LONGITUDE_       REMARKS
##  Min.   :-14451   Min.   :   0   Min.   :-14455   Length:902297
##  1st Qu.:  7247   1st Qu.:   0   1st Qu.:     0   Class :character
##  Median :  8707   Median :   0   Median :     0   Mode  :character
##  Mean   :  6940   Mean   :1452   Mean   :  3509
##  3rd Qu.:  9605   3rd Qu.:3549   3rd Qu.:  8735
##  Max.   : 17124   Max.   :9706   Max.   :106220
##                   NA's   :40
##      REFNUM
##  Min.   :     1
##  1st Qu.:225575
##  Median :451149
##  Mean   :451149
##  3rd Qu.:676723
##  Max.   :902297
##
```

RESULTS

###QUESTION 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

```
#a) aggregating EVTYPE wrt injuries : aggregate the top 10 injuries by the event type and sor
t the output in descending order
total_injuries <- aggregate(INJURIES~EVTYPE, storm_data, sum)
total_injuries <- arrange(total_injuries, desc(INJURIES))
total_injuries <- total_injuries[1:20, ]
total_injuries
```

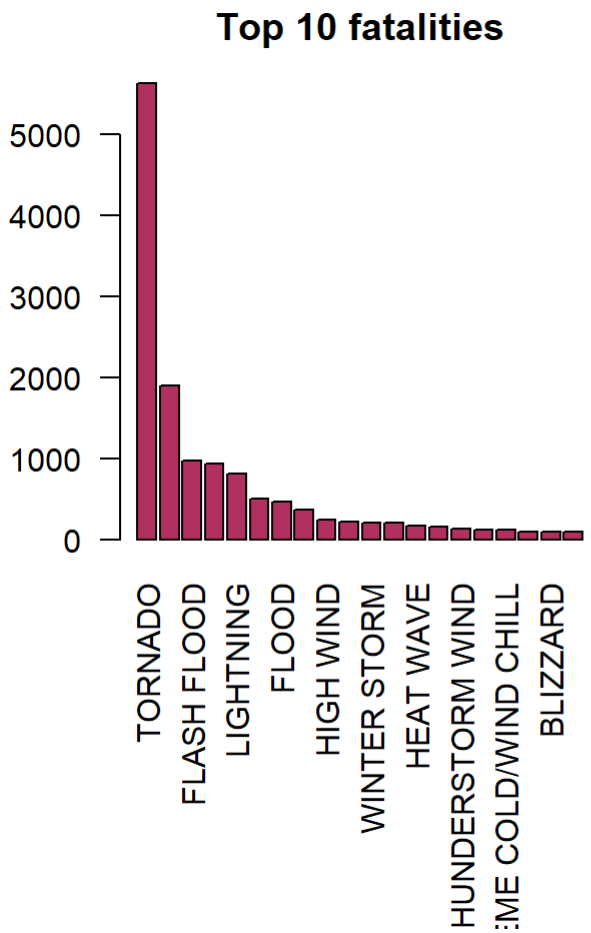
```
##                EVTYPE INJURIES
## 1             TORNADO    91346
## 2           TSTM WIND     6957
## 3               FLOOD     6789
## 4      EXCESSIVE HEAT     6525
## 5           LIGHTNING     5230
## 6                HEAT     2100
## 7           ICE STORM     1975
## 8         FLASH FLOOD     1777
## 9   THUNDERSTORM WIND     1488
## 10               HAIL     1361
## 11       WINTER STORM     1321
## 12  HURRICANE/TYPHOON     1275
## 13          HIGH WIND     1137
## 14         HEAVY SNOW     1021
## 15           WILDFIRE      911
## 16 THUNDERSTORM WINDS      908
## 17           BLIZZARD      805
## 18                FOG      734
## 19   WILD/FOREST FIRE      545
## 20         DUST STORM      440
```

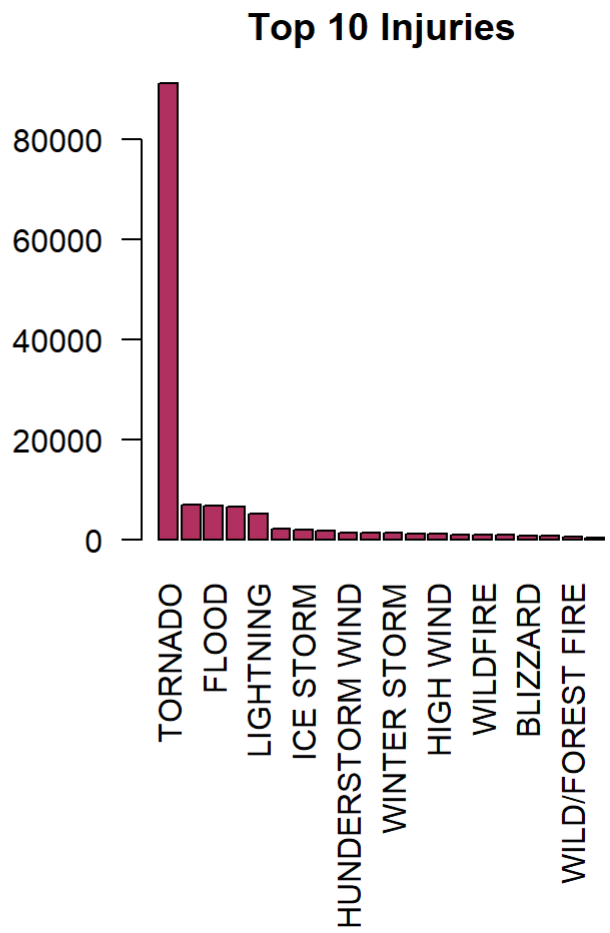
```
#b) aggregating EVTYPE wrt fatalities : aggregate the top 10 fatalities by the event type and sort the output in descending order
total_fatalities <- aggregate(FATALITIES~EVTYPE,storm_data, sum)
total_fatalities <- arrange(total_fatalities, desc(FATALITIES))
total_fatalities <- total_fatalities[1:20, ]
total_fatalities
```

```
##                   EVTYPE FATALITIES
## 1                 TORNADO       5633
## 2           EXCESSIVE HEAT       1903
## 3              FLASH FLOOD        978
## 4                    HEAT        937
## 5               LIGHTNING        816
## 6               TSTM WIND        504
## 7                   FLOOD        470
## 8             RIP CURRENT        368
## 9               HIGH WIND        248
## 10              AVALANCHE        224
## 11           WINTER STORM        206
## 12           RIP CURRENTS        204
## 13              HEAT WAVE        172
## 14           EXTREME COLD        160
## 15       THUNDERSTORM WIND        133
## 16             HEAVY SNOW        127
## 17 EXTREME COLD/WIND CHILL        125
## 18             STRONG WIND        103
## 19                BLIZZARD        101
## 20              HIGH SURF        101
```

Plotting the graph depicting the top 10 causes for Fatalities and Injuries

```
# plot graphs showing the top 10 fatalities and injuries
par(mfrow=c(1,2),mar=c(10,3,3,2))
barplot(total_fatalities$FATALITIES,names.arg=total_fatalities$EVTYPE,las=2,col="maroon",ylab
="fatalities",main="Top 10 fatalities")
barplot(total_injuries$INJURIES,names.arg=total_injuries$EVTYPE,las=2,col="maroon",ylab="inju
ries",main="Top 10 Injuries")
```





###QUESTION 2. Across the United States, which types of events have the greatest economic consequences? An analysis of the weather events responsible for the greatest economic consequences

Hypothesis: Economic consequences means damages. The two significant types of damage typically caused by weather events include ‘properties and crops’

We have property Damage and crop damage

```
# Aggregate Data for Property Damage
property_damage <- aggregate(PROPDMG ~ EVTYPE, data = storm_data, FUN = sum)
property_damage <- property_damage[order(property_damage$PROPDMG, decreasing = TRUE), ]
# 10 most harmful causes of injuries
property_damageMax <- property_damage[1:10, ]
print(property_damageMax)
```

```
##                EVTYPE   PROPDMG
## 834           TORNADO 3212258.2
## 153       FLASH FLOOD 1420124.6
## 856         TSTM WIND 1335965.6
## 170             FLOOD  899938.5
## 760 THUNDERSTORM WIND  876844.2
## 244              HAIL  688693.4
## 464         LIGHTNING  603351.8
## 786 THUNDERSTORM WINDS  446293.2
## 359         HIGH WIND  324731.6
## 972       WINTER STORM  132720.6
```

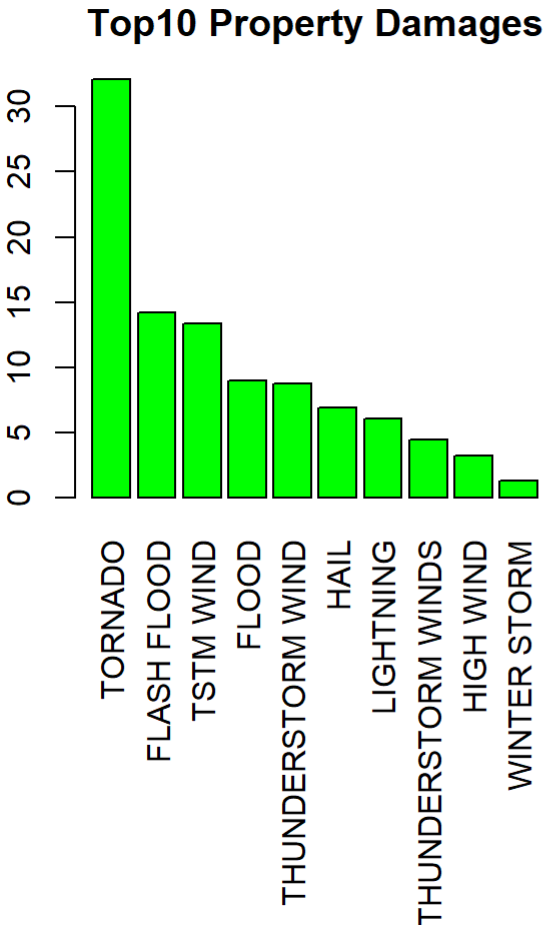
```
#Aggregate Data for Crop Damage
crop_damage <- aggregate(CROPDMG ~ EVTYPE, data = storm_data, FUN = sum)
crop_damage <- crop_damage[order(crop_damage$CROPDMG, decreasing = TRUE), ]
# 10 most harmful causes of injuries
crop_damageMax <- crop_damage[1:10, ]
print(crop_damageMax)
```

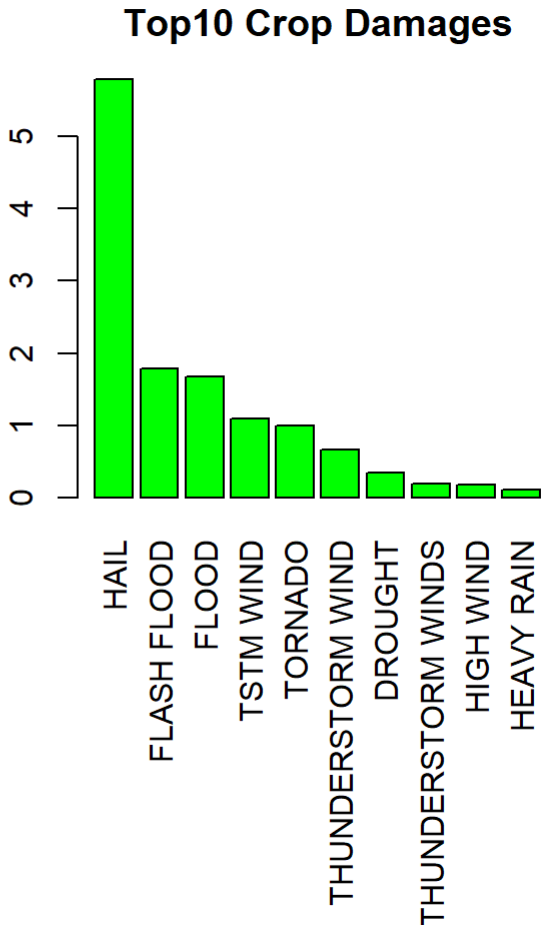
```
##                EVTYPE   CROPDMG
## 244              HAIL 579596.28
## 153       FLASH FLOOD 179200.46
## 170             FLOOD 168037.88
## 856         TSTM WIND 109202.60
## 834           TORNADO 100018.52
## 760 THUNDERSTORM WIND  66791.45
## 95            DROUGHT  33898.62
## 786 THUNDERSTORM WINDS  18684.93
## 359         HIGH WIND  17283.21
## 290        HEAVY RAIN  11122.80
```

Plotting the graph depicting the top 10 causes for Property and Crop Damage

```
##plot the graph showing the top 10 property and crop damages

par(mfrow=c(1,2),mar=c(11,3,3,2))
barplot(property_damageMax$PROPDMG/(10^5),names.arg=property_damageMax$EVTYPE,las=3,col="gree
n",ylab="Property damage(billions)",main="Top10 Property Damages")
barplot(crop_damageMax$CROPDMG/(10^5),names.arg=crop_damageMax$EVTYPE,las=3,col="green",ylab
="Crop damage(billions)",main="Top10 Crop Damages")
```





SUMMARY

Tornados are responsible for the maximum number of fatalities and injuries, followed by Excessive Heat for fatalities and Thunderstorm wind for injuries.

Floods are responsbile for maximum property damage, while Droughts cause maximum crop damage. Second major events that caused the maximum damage was Hurricanes/Typhoos for property damage and Floods for crop damage.