Untitled

September 30, 2020

Analysis the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database

This project explores the NOAA storm database, which tracks major storms and weather events, to address the most severe types of weather events in the USA, which caused greatest damage to human population in terms of fatalities/injuries and economic loss during the years 1950 - 2011.

There are two goals of this analysis:

- identify the weather events that are most harmful with respect to population health
- identify the weather events that have the greatest economic consequences.

Based on our analysis, we conclude that TORNADOS and FLOODS are most harmful weather events in the USA in terms of the risk to human health and economic impact.

Data processing

[10]:
```
Url_data <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.
 ↪csv.bz2"
```

[11]:
```
File_data <- "StormData.csv.bz2"
if (!file.exists(File_data)) {
 download.file(Url_data, File_data, mode = "wb") }
```

[12]:
```
data <-read.csv(file = File_data, header=TRUE, sep=",")
```

[13]:
```
head(data)
```

A data.frame: 6 × 37

	STATE___ <dbl>	BGN_DATE <fct>	BGN_TIME <fct>	TIME_ZONE <fct>	COUNTY <dbl>	COUN <fct>
1	1	4/18/1950 0:00:00	0130	CST	97	MOBII
2	1	4/18/1950 0:00:00	0145	CST	3	BALDV
3	1	2/20/1951 0:00:00	1600	CST	57	FAYET
4	1	6/8/1951 0:00:00	0900	CST	89	MADIS
5	1	11/15/1951 0:00:00	1500	CST	43	CULLM
6	1	11/15/1951 0:00:00	2000	CST	77	LAUDI

[60]:
```
# subsetting by data
main_data<-data
```

[61]:
```
main_data$BGN_DATE <- strptime(data$BGN_DATE, "%m/%d/%Y %H:%M:%S")
```

[62]:
```
main_data <- subset(data,main_data$BGN_DATE>1996-01-01, )
```

[63]:
```
str(main_data)
```

```
'data.frame':   866041 obs. of  37 variables:
 $ STATE__    : num  1 1 1 1 1 1 1 1 1 1 …
 $ BGN_DATE   : Factor w/ 16335 levels "1/1/1966 0:00:00",..: 3852 3852 5704 4834
4834 5092 5185 5185 5185 5185 …
 $ BGN_TIME   : Factor w/ 3608 levels "00:00:00 AM",..: 2623 2865 2584 2732 2735
3090 1143 1555 1683 3555 …
 $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",..: 7 7 7 7 7 7 7 7 7 7 …
 $ COUNTY     : num  97 63 129 67 5 57 65 73 125 17 …
 $ COUNTYNAME: Factor w/ 29601 levels "","5NM E OF MACKINAC BRIDGE TO PRESQUE
ISLE LT MI",..: 13513 5653 27256 5778 1887 4598 5714 8485 24418 2808 …
 $ STATE      : Factor w/ 72 levels "AK","AL","AM",..: 2 2 2 2 2 2 2 2 2 2 …
 $ EVTYPE     : Factor w/ 985 levels "   HIGH SURF ADVISORY",..: 856 856 834 856
856 244 834 244 856 834 …
 $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 …
 $ BGN_AZI    : Factor w/ 35 levels "","  N"," NW",..: 1 1 1 1 1 1 1 1 1 1 …
 $ BGN_LOCATI: Factor w/ 54429 levels "","  Christiansburg",..: 1 1 1 1 1 1 1 1 1
1 …
 $ END_DATE   : Factor w/ 6663 levels "","1/1/1993 0:00:00",..: 1 1 1 1 1 1 1 1 1
1 …
 $ END_TIME   : Factor w/ 3647 levels "","  0900CST",..: 1 1 1 1 1 1 1 1 1 1 …
 $ COUNTY_END: num  0 0 0 0 0 0 0 0 0 0 …
 $ COUNTYENDN: logi  NA NA NA NA NA NA …
 $ END_RANGE : num  0 0 0 0 0 0 0 0 0 0 …
 $ END_AZI    : Factor w/ 24 levels "","E","ENE","ESE",..: 1 1 1 1 1 1 1 1 1 1
…
 $ END_LOCATI: Factor w/ 34506 levels "","  CANTON","  TULIA",..: 1 1 1 1 1 1 1 1
1 1 …
 $ LENGTH     : num  0 0 0 0 0 0 1 0 0 2 …
 $ WIDTH      : num  0 0 33 0 0 0 33 0 0 200 …
 $ F          : int  NA NA 2 NA NA NA 2 NA NA 3 …
 $ MAG        : num  51 0 0 0 0 75 0 100 0 0 …
 $ FATALITIES: num  0 0 0 0 0 0 0 0 0 2 …
 $ INJURIES   : num  0 0 0 0 0 0 0 0 0 14 …
 $ PROPDMG    : num  0 0 25 0 0 0 25 0 0 250 …
 $ PROPDMGEXP: Factor w/ 19 levels "","-","?","+",..: 1 1 17 1 1 1 17 1 1 17 …
 $ CROPDMG    : num  0 0 0 0 0 0 0 0 0 0 …
 $ CROPDMGEXP: Factor w/ 9 levels "","?","0","2",..: 1 1 1 1 1 1 1 1 1 1 …
 $ WFO        : Factor w/ 542 levels "","  CI","%SD",..: 1 1 1 1 1 1 1 1 1 1 …
 $ STATEOFFIC: Factor w/ 250 levels "","ALABAMA, Central",..: 1 1 1 1 1 1 1 1 1
1 …
 $ ZONENAMES : Factor w/ 25112 levels "","
"| __truncated__,..: 1 1 1 1 1 1 1 1 1 1 …
 $ LATITUDE   : num  3042 3247 3128 3136 3142 …
```

```
 $ LONGITUDE : num  8817 8754 8815 8517 8524 …
 $ LATITUDE_E: num  0 0 0 0 0 0 0 0 0 0 …
 $ LONGITUDE_: num  0 0 0 0 0 0 0 0 0 0 …
 $ REMARKS   : Factor w/ 436781 levels "","\t","\t\t",..: 1 1 1 1 1 1 1 1 1 1
…
 $ REFNUM    : num  799 800 801 802 803 804 805 806 807 808 …
```

[64]:
```
main_data <- subset(main_data, select = c(EVTYPE, FATALITIES, INJURIES, PROPDMG,
PROPDMGEXP, CROPDMG, CROPDMGEXP))
```

[65]:
```
#cleaning event types names
main_data$EVTYPE <- toupper(main_data$EVTYPE)
```

[66]:
```
# eliminating zero data
#main_data <- main_data[main_data$FATALITIES!=0 |main_data$INJURIES !=0 |␣
 ↪main_data$PROPDMG !=0 | main_data$CROPDMG !=0, ]
```

Population health data processing

[68]:
```
health_data <- aggregate(cbind(FATALITIES, INJURIES) ~ EVTYPE, data =␣
 ↪main_data, FUN=sum)
```

[70]:
```
health_data$PEOPLE_LOSS <-health_data$FATALITIES + health_data$INJURIES
```

[71]:
```
health_data <- health_data[order(health_data$PEOPLE_LOSS, decreasing =
TRUE), ]
```

[72]:
```
Top10_events_people <- health_data[1:10,]
```

[73]:
```
Top10_events_people
```

A data.frame: 10 × 4

	EVTYPE <chr>	FATALITIES <dbl>	INJURIES <dbl>	PEOPLE_LOSS <dbl>
758	TORNADO	3272	59611	62883
116	EXCESSIVE HEAT	1903	6525	8428
779	TSTM WIND	504	6957	7461
154	FLOOD	470	6789	7259
418	LIGHTNING	816	5230	6046
243	HEAT	937	2100	3037
138	FLASH FLOOD	978	1777	2755
387	ICE STORM	89	1975	2064
685	THUNDERSTORM WIND	133	1488	1621
888	WINTER STORM	206	1321	1527

Economic consequences data processing

[74]:
```
main_data$PROPDMGEXP <- gsub("[Hh]", "2", main_data$PROPDMGEXP)
main_data$PROPDMGEXP <- gsub("[Kk]", "3", main_data$PROPDMGEXP)
main_data$PROPDMGEXP <- gsub("[Mm]", "6", main_data$PROPDMGEXP)
```

```
main_data$PROPDMGEXP <- gsub("[Bb]", "9", main_data$PROPDMGEXP)
main_data$PROPDMGEXP <- gsub("\\+", "1", main_data$PROPDMGEXP)
main_data$PROPDMGEXP <- gsub("\\?|\\-|\\ ", "0", main_data$PROPDMGEXP)
main_data$PROPDMGEXP <- as.numeric(main_data$PROPDMGEXP)
```

[76]:
```
main_data$CROPDMGEXP <- gsub("[Hh]", "2", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- gsub("[Kk]", "3", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- gsub("[Mm]", "6", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- gsub("[Bb]", "9", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- gsub("\\+", "1", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- gsub("\\-|\\?|\\ ", "0", main_data$CROPDMGEXP)
main_data$CROPDMGEXP <- as.numeric(main_data$CROPDMGEXP)
main_data$PROPDMGEXP[is.na(main_data$PROPDMGEXP)] <- 0
main_data$CROPDMGEXP[is.na(main_data$CROPDMGEXP)] <- 0
```

[78]:
```
#creating total damage values
library(dplyr)
```

[80]:
```
main_data <- mutate(main_data, PROPDMGTOTAL = PROPDMG * (10 ^ PROPDMGEXP),␣
↪CROPDMGTOTAL = CROPDMG * (10 ^ CROPDMGEXP))
```

[81]:
```
Economic <- aggregate(cbind(PROPDMGTOTAL, CROPDMGTOTAL) ~ EVTYPE, data =␣
↪main_data, FUN=sum)
```

[82]:
```
Economic $ECONOMIC_LOSS <- Economic$PROPDMGTOTAL + Economic $CROPDMGTOTAL
Economic <- Economic[order(Economic$ECONOMIC_LOSS, decreasing =TRUE), ]
```

[84]:
```
Top10_events_economy <- Economic[1:10,]
```

[85]:
```
Top10_events_economy
```

A data.frame: 10 × 4

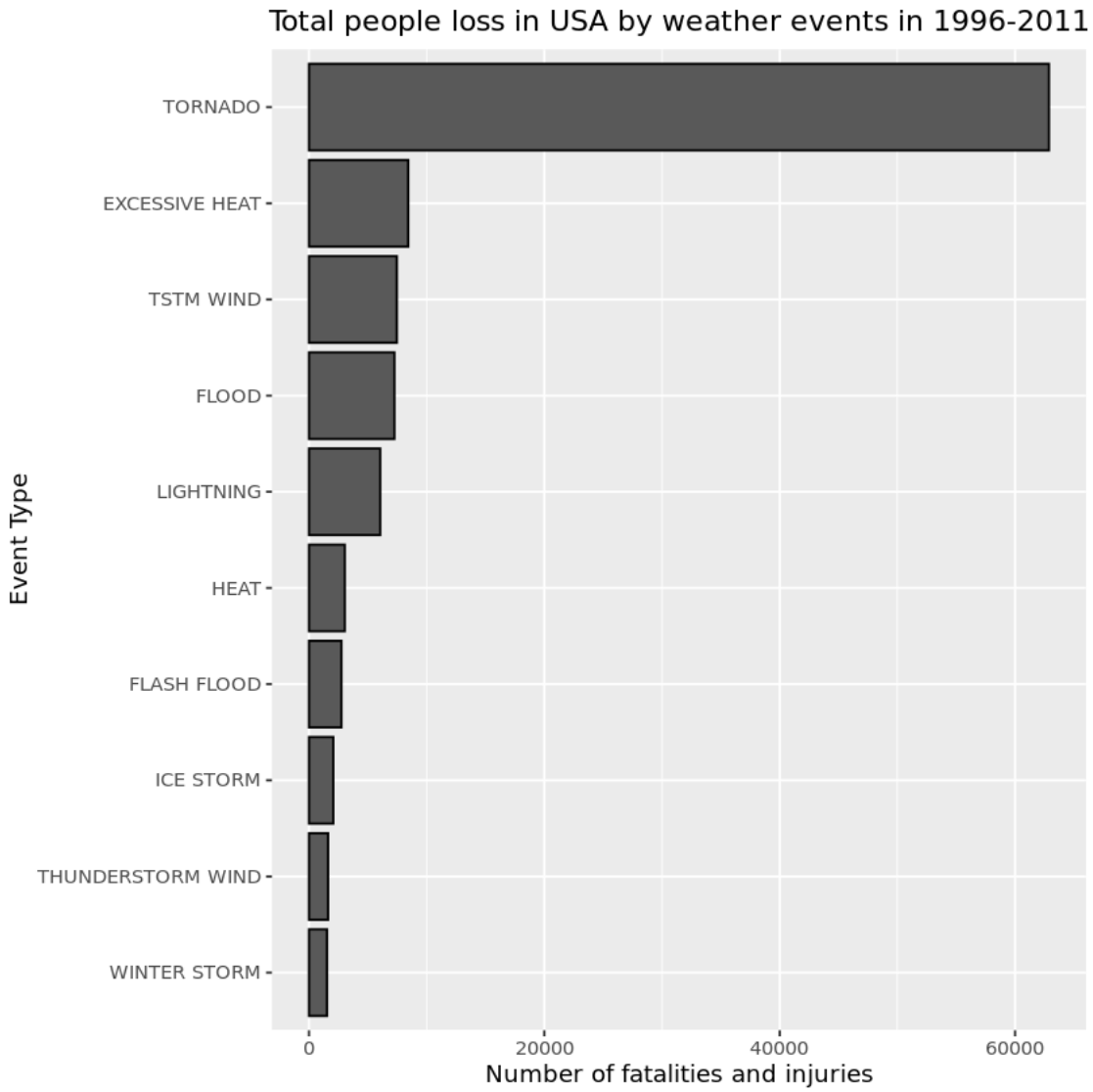
	EVTYPE <chr>	PROPDMGTOTAL <dbl>	CROPDMGTOTAL <dbl>	ECONOM <dbl>
154	FLOOD	144657709807	5661968450	15031967
372	HURRICANE/TYPHOON	69305840000	2607872800	71913712
758	TORNADO	51673434936	414953270	52088388
599	STORM SURGE	43323536000	5000	43323541
212	HAIL	15735267513	3025954473	18761221
138	FLASH FLOOD	16822673978	1421317100	18243991
84	DROUGHT	1046106000	13972566000	15018672
363	HURRICANE	11868319010	2741910000	14610229
529	RIVER FLOOD	5118945500	5029459000	10148404
387	ICE STORM	3944927860	5022113500	89670413

Results

[86]:
```
#plotting health loss
library(ggplot2)
```

```
g <- ggplot(data = Top10_events_people, aes(x = reorder(EVTYPE, PEOPLE_LOSS),␣
 ↪y= PEOPLE_LOSS))
g <- g + geom_bar(stat = "identity", colour = "black") + labs(title = "Total␣
 ↪people loss in USA by weather events in 1996-2011") + theme(plot.title =␣
 ↪element_text(hjust = 0.5))  + labs(y = "Number of fatalities and injuries",␣
 ↪x = "Event Type") + coord_flip()
```

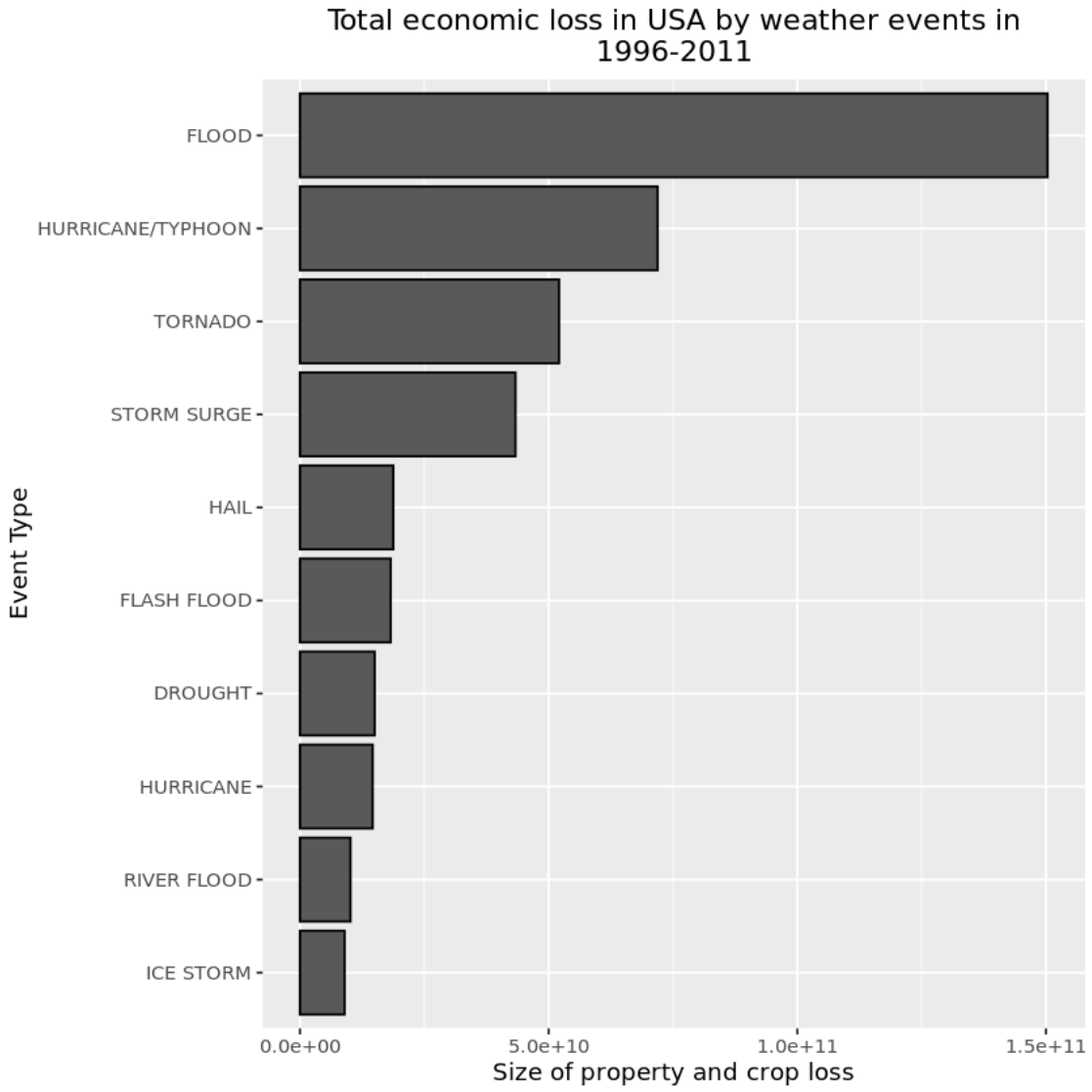
[87]:
```
print(g)
```



[88]:
```
#plotting economic loss
g <- ggplot(data = Top10_events_economy, aes(x = reorder(EVTYPE,␣
 ↪ECONOMIC_LOSS), y = ECONOMIC_LOSS))
```

```
g <- g + geom_bar(stat = "identity", colour = "black") + labs(title = "Total␣
 ↪economic loss in USA by weather events in
1996-2011") + theme(plot.title = element_text(hjust = 0.5)) + labs(y = "Size of␣
 ↪property and crop loss", x = "Event Type") + coord_flip()
```

[89]:
```
print(g)
```



[ ]:
```

```