A brief report on population and economic impacts of natural disasters in USA between 1950-2011

Kamran A.

Synopsis:

The this exploratory report deals with the data on population and economic impact of natural disasters in USA, from 1950 to 2011.

Each recorded disaster in that period was assigned to one of 898 categories (such as flood, tornado, typhoon, etc) and the aim of this report is to highlight the categories with highest population and economic impacts.

The population impact is measured by the a total number of injuries and fatalities resulting from a given disaster. Likewise, economic impact is measured by combined estimated cost of property and agricultural damages resulting from a given disaster.

This report may be of use to government officials in charge of allocating funds towards programs, aiming to predict the onset and mitigate the impact of natural disasters.

Data Processing:

This report makes use of the following packages:

-data.table -R.utils -ggplot2 -reshape2 -gridExtra.

```
library(data.table)
library(R.utils)
library(ggplot2)
library(reshape2)
library(gridExtra)
```

Data is loaded from the Coursera website, unzipped and read into R.

```
URL<- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

download.file(url = URL, destfile = "StormData.bz2")

bunzip2(filename = "StormData.bz2", destname = "StormData.csv", overwrite=TRUE)

raw<-fread(input = "StormData.csv", sep = ",", header = TRUE, stringsAsFactors = TRUE, showPr
ogress = FALSE)

head(raw)
```

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

Since this report aims to analyze the aggregate impact of various disaster types, we will not be needing most of the variables.

```
raw1<- subset(x = raw, select = c(EVTYPE, FATALITIES, INJURIES, PROPDMG, CROPDMG))
```

Finally, lets rename our variables for convenience and set natural disaster types to lowercase.

```
names(raw1)<-c("event", "fatalities", "injuries", "property_damage", "crop_damage")

raw1$event<-tolower(raw1$event)

head(raw1)
```

```
##      event fatalities injuries property_damage crop_damage
## 1: tornado          0       15            25.0           0
## 2: tornado          0        0             2.5           0
## 3: tornado          0        2            25.0           0
## 4: tornado          0        2             2.5           0
## 5: tornado          0        2             2.5           0
## 6: tornado          0        6             2.5           0
```

Now that our data is tidy, lets find the total population/economic damage of each disaster type, from 1950 to 2011.

```
raw2<-raw1[, list(injuries=sum(injuries), fatalities=sum(fatalities), totalhd=sum(injuries, f
atalities), property_damage=sum(property_damage), crop_damage=sum(crop_damage), totaled=sum(p
roperty_damage, crop_damage)), by= event]

head(raw2)
```

```
##                    event injuries fatalities totalhd property_damage
## 1:               tornado    91346       5633   96979      3212258.16
## 2:             tstm wind     6957        504    7461      1335995.61
## 3:                  hail     1361         15    1376       688693.38
## 4:         freezing rain       23          7      30         2951.70
## 5:                  snow       31          5      36         3069.32
## 6: ice storm/flash flood        2          0       2            0.00
##    crop_damage    totaled
## 1:    100018.5 3312276.68
## 2:    109202.6 1445198.21
## 3:    579596.3 1268289.66
## 4:         0.0    2951.70
## 5:        10.0    3079.32
## 6:         0.0       0.00
```

Data Analysis

Selecting Disaster Types

To find out which disaster types have the highest population impact, lets order the data by “totalhd” variable

```
data1<- raw2[,1:4][order(totalhd, decreasing = TRUE),]

head(data1)
```

```
##             event injuries fatalities totalhd
## 1:        tornado    91346       5633   96979
## 2: excessive heat     6525       1903    8428
## 3:      tstm wind     6957        504    7461
## 4:          flood     6789        470    7259
## 5:      lightning     5230        816    6046
## 6:           heat     2100        937    3037
```

As you can see from the table, tornadoes have by far the largest number of casualties. Visual inspection of the data shows that the numbers of fatalities and injuries decline rapidly for all other event types, which is confirmed by the quantile function:

```
final1<-subset(data1, totalhd>=quantile(totalhd, .985))

final1
```

```
##                event injuries fatalities totalhd
##  1:          tornado    91346       5633   96979
##  2:   excessive heat     6525       1903    8428
##  3:        tstm wind     6957        504    7461
##  4:            flood     6789        470    7259
##  5:        lightning     5230        816    6046
##  6:             heat     2100        937    3037
##  7:      flash flood     1777        978    2755
##  8:        ice storm     1975         89    2064
##  9: thunderstorm wind     1488        133    1621
## 10:     winter storm     1321        206    1527
## 11:        high wind     1137        248    1385
## 12:             hail     1361         15    1376
## 13: hurricane/typhoon     1275         64    1339
## 14:       heavy snow     1021        127    1148
```

It seems that 14 event types out of 898 account for 98.5% of all casualties caused by natural disasters. This is very useful for policy makers, as they can focus their efforts on mitigating the effects a much smaller number of disasters.

Next, let us perform the same analysis to highlight disaster types with the highest economic impact.

```
data2<- raw2[,c(1, 5:7)][order(totaled, decreasing = TRUE),]

final2<-subset(data2, totaled>=quantile(totaled, .985))

final2
```

```
##                 event property_damage crop_damage    totaled
##  1:           tornado      3212258.16   100018.52 3312276.68
##  2:       flash flood      1420124.59   179200.46 1599325.05
##  3:         tstm wind      1335995.61   109202.60 1445198.21
##  4:              hail       688693.38   579596.28 1268289.66
##  5:             flood       899938.48   168037.88 1067976.36
##  6:  thunderstorm wind       876844.17    66791.45  943635.62
##  7:         lightning       603351.78     3580.61  606932.39
##  8: thunderstorm winds       446293.18    18684.93  464978.11
##  9:         high wind       324731.56    17283.21  342014.77
## 10:      winter storm       132720.59     1978.99  134699.58
## 11:        heavy snow       122251.99     2165.72  124417.71
## 12:          wildfire        84459.34     4364.20   88823.54
## 13:         ice storm        66000.67     1688.95   67689.62
## 14:       strong wind        63011.81     1616.90   64628.71
```

Once again, we have 14 disaster types, which account for 98.5% of economic damage from natural disasters. Not surprisingly, the property damage of tornadoes greatly exceeds that of the other types.

Preparing theCcharts

Since the impact of each event has 2 components (injuries/fatalities and property/agricultural damages), it would be suitable to represent our findings in the shape of a stacked bar chart.

The following code prepares our first set of results, pertaining to population damages.

We begin by subsetting the data and converting disaster event types to an ordered factor variable.

```
final1<-final1[,1:3]

final1$event<- ordered(final1$event, levels=unique(final1$event))

head(final1)
```

```
##              event injuries fatalities
## 1:         tornado    91346       5633
## 2: excessive heat     6525       1903
## 3:       tstm wind     6957        504
## 4:           flood     6789        470
## 5:       lightning     5230        816
## 6:            heat     2100        937
```

Next, we melt the injuries and fatalities variables, as well as sort the data table by ascending event and descending type.

```
final1<-melt(final1, id.vars ="event", value.name = "casualties", variable.name = "type")

final1<-final1[order(event, -type)]

head(final1)
```

```
##              event       type casualties
## 1:         tornado fatalities       5633
## 2:         tornado   injuries      91346
## 3: excessive heat fatalities       1903
## 4: excessive heat   injuries       6525
## 5:       tstm wind fatalities        504
## 6:       tstm wind   injuries       6957
```

Lastly, we add a fourth variable, which is a cumulative sum of fatalities and injuries, per disaster event type. We will use this variable to indicate label positions in the bar chart later on.

```
final1<-final1[, list(type, casualties, y_pos=cumsum(casualties)), by=event]

head(final1)
```

```
##              event       type casualties y_pos
## 1:         tornado fatalities       5633  5633
## 2:         tornado   injuries      91346 96979
## 3: excessive heat fatalities       1903  1903
## 4: excessive heat   injuries       6525  8428
## 5:       tstm wind fatalities        504   504
## 6:       tstm wind   injuries       6957  7461
```

Since the damage done by tornadoes is much greater than then other 13 disaster types, plotting all on the same scale will make the bar chart hard to read. Hence, we will make a two panel plot, whereby the first panel will show all 14 disaster types and the second panel will show the same types, with tornadoes excluded.

```
plot1<- ggplot(final1, aes(event, casualties, fill=type)) + geom_bar(stat = "identity") +

        geom_text(aes(y=y_pos, label=casualties), vjust=-0.3, color="black", size=3.5) +

        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
c(0.9,0.95)) +

        xlab("Disaster Type") + ylab("Number of Casualties") +

        scale_fill_discrete(name="Legend",labels=c("Injuries", "Fatalities"))

plot2<- ggplot(final1[c(-1, -2),], aes(event, casualties, fill=type)) + geom_bar(stat = "iden
tity") +

        geom_text(aes(y=y_pos, label=casualties), vjust=-0.3, color="black", size=3.5) +

        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
c(0.9,0.95)) +

        xlab("Disaster Type") + ylab("Number of Casualties") +

        scale_fill_discrete(name="Legend",labels=c("Injuries", "Fatalities"))
```

Now, let us perform the same set of steps on the data table for economic effects, with an extra step towards the end to round up the cost to the nearest thousand, mostly to have shorter labels on the bar chart

```
data2<- raw2[,c(1, 5:7)][order(totaled, decreasing = TRUE),]

final2<-subset(data2, totaled>=quantile(totaled, .985))

final2<-final2[,1:3]

final2$event<- ordered(final2$event, levels=unique(final2$event))

final2<-melt(final2, id.vars ="event", value.name = "cost", variable.name = "type")

final2<-final2[order(event, -type)]

final2$cost<- sapply(final2$cost, function(x){ round(x/1000, digits = 1)})

final2<-final2[, list(type, cost, y_pos=cumsum(cost)), by=event]

final2
```

```
##                 event            type   cost  y_pos
##  1:           tornado     crop_damage  100.0  100.0
##  2:           tornado property_damage 3212.3 3312.3
##  3:       flash flood     crop_damage  179.2  179.2
##  4:       flash flood property_damage 1420.1 1599.3
##  5:         tstm wind     crop_damage  109.2  109.2
##  6:         tstm wind property_damage 1336.0 1445.2
##  7:              hail     crop_damage  579.6  579.6
##  8:              hail property_damage  688.7 1268.3
##  9:             flood     crop_damage  168.0  168.0
## 10:             flood property_damage  899.9 1067.9
## 11:  thunderstorm wind     crop_damage   66.8   66.8
## 12:  thunderstorm wind property_damage  876.8  943.6
## 13:         lightning     crop_damage    3.6    3.6
## 14:         lightning property_damage  603.4  607.0
## 15: thunderstorm winds     crop_damage   18.7   18.7
## 16: thunderstorm winds property_damage  446.3  465.0
## 17:         high wind     crop_damage   17.3   17.3
## 18:         high wind property_damage  324.7  342.0
## 19:      winter storm     crop_damage    2.0    2.0
## 20:      winter storm property_damage  132.7  134.7
## 21:        heavy snow     crop_damage    2.2    2.2
## 22:        heavy snow property_damage  122.3  124.5
## 23:          wildfire     crop_damage    4.4    4.4
## 24:          wildfire property_damage   84.5   88.9
## 25:         ice storm     crop_damage    1.7    1.7
## 26:         ice storm property_damage   66.0   67.7
## 27:       strong wind     crop_damage    1.6    1.6
## 28:       strong wind property_damage   63.0   64.6
##                 event            type   cost  y_pos
```

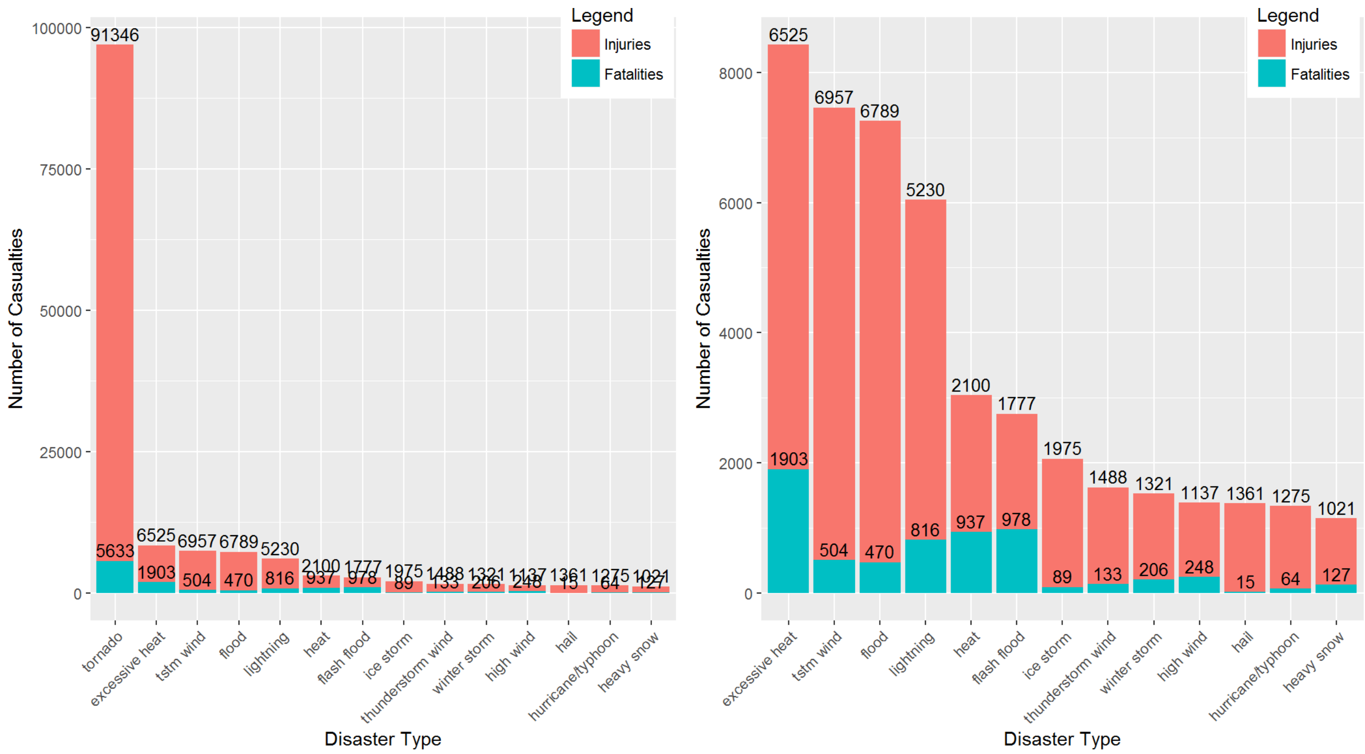
Here is the corresponding bar chart code for economic damage.

```
plot3<- ggplot(final2, aes(event, cost, fill=type)) + geom_bar(stat = "identity") +

        geom_text(aes(y=y_pos, label=cost), vjust=-0.3, color="black", size=3.5) +

        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
c(0.9,0.95)) +

        xlab("Disaster Type") + ylab("Cost Estimate (thousands of USD)") +

        scale_fill_discrete(name="Legend",labels=c("Property Damage", "Crop Damage"))+
guides(fill=FALSE)


plot4<- ggplot(final2[c(-1, -2),], aes(event, cost, fill=type)) + geom_bar(stat = "identity")
 +

        geom_text(aes(y=y_pos, label=cost), vjust=-0.3, color="black", size=3.5) +

        theme(axis.text.x = element_text(angle = 45, hjust = 1), legend.position =
c(0.9,0.95)) +

        xlab("Disaster Type") + ylab("Cost Estimate (thousands of USD)") +

        scale_fill_discrete(name="Legend",labels=c("Property Damage", "Crop Damage"))
```

Results

Here is the first plot, where we can clearly see that tornadoes cause the most population damage.

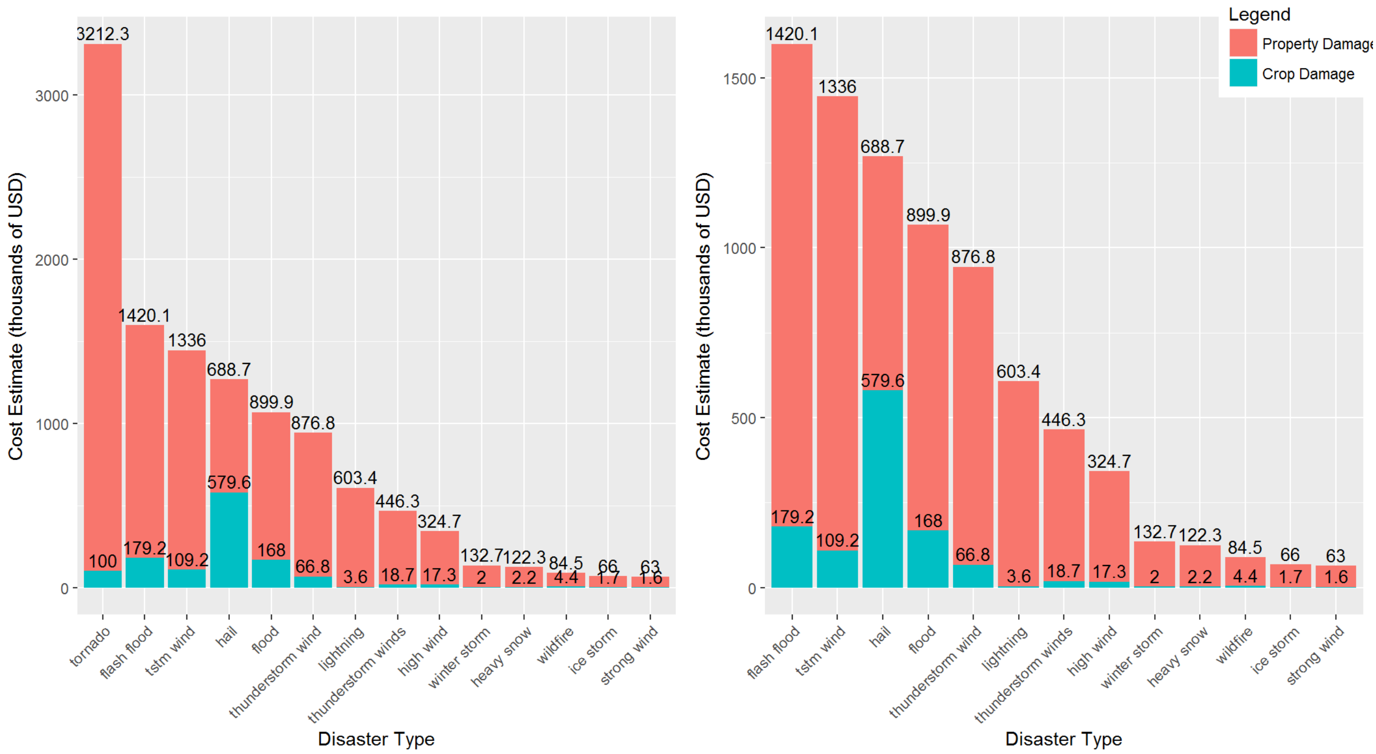
```
grid.arrange(plot1, plot2, ncol=2)
```



Fourteen natural disaster types accounting for 98.5% of population damage from 1950 to 2011 (with and without tornadoes)

And here is the second plot, which once more shows that tornadoes cause the most economic damage.

```
grid.arrange(plot3, plot4, ncol=2)
```



Fourteen natural disaster types accounting for 98.5% of economic damage from 1950 to 2011 (with and without tornadoes)