PA2_Analysis

Robert Jones

Health and Economical of Natural Disasters

Synopsis

The ai'm of this report to is study the impact of natural disasters in the US with a view to understand how to manage resourcing and budgets to enable mitigation strategies. There shall be a particular focus on determining the even types which lead to the greatest harm to the poplation and those which have the greatest economic impact. The report shall also attempt to capture changes to trends over time to assess whether a contributor is persistant.

Data Processing

The storm data has been downloaded from the National Weather Service between 19520 and 2011, further information on the data can be reviewed in the following link

Only data relating to the onjects will be passed on for analysis.

```
if (!dir.exists("data")) {
  dir.create("data")
}

url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download.file(url, "data/storm data.bz2")


storm.data <- read.csv("data/storm data.bz2", as.is = T)

storm.data$PROP.MULTI[storm.data$PROPDMGEXP == "K"] <- 1000
storm.data$PROP.MULTI[storm.data$PROPDMGEXP == "M"] <- 1000000
storm.data$PROP.MULTI[storm.data$PROPDMGEXP == "B"] <- 1000000000

storm.data$CROP.MULTI[storm.data$CROPDMGEXP == "K"] <- 1000
storm.data$CROP.MULTI[storm.data$CROPDMGEXP == "M"] <- 1000000
storm.data$CROP.MULTI[storm.data$CROPDMGEXP == "B"] <- 1000000000


#translate dates, aggregate damage and select columns
storm.data<- mutate(storm.data, start.date = mdy_hms(storm.data$BGN_DATE),end.date = mdy_hms(storm.data$
  select(year, EVTYPE, STATE, damage, INJURIES, FATALITIES)

head(storm.data)
```

```
##   year  EVTYPE STATE damage INJURIES FATALITIES
## 1 1950 TORNADO    AL     NA       15          0
## 2 1950 TORNADO    AL     NA        0          0
## 3 1951 TORNADO    AL     NA        2          0
## 4 1951 TORNADO    AL     NA        2          0
```

```
## 5 1951 TORNADO    AL     NA        2          0
## 6 1951 TORNADO    AL     NA        6          0
```

The study shall focus on the top 40 event types arrange with highest fatalities and also damage. The analysis steps are as follows: 1. Group data and summarise damage and fatalities 2. Get top 40 event types for each (fatalities and damage) 3. Subset data based on top 40 3. Create yearly summary for top 10 events

```
#Get top 40 event types
storm.type.grouped <- group_by(storm.data, EVTYPE)

storm.type.summary <- summarise(storm.type.grouped, total.injuries = sum(INJURIES, na.rm = T), mean.inju

top40.fatality.event <- top_n(storm.type.summary,40,total.fatalities)
top40.damage.event <- top_n(storm.type.summary,40,total.damage)

#Subset Data
top10.fatality.year.summary <- filter(storm.data, EVTYPE %in% head(top40.fatality.event$EVTYPE,10)) %>%
                                    mean.injuries = mean(INJURIES, na.rm = T),
                                    total.fatalities = sum(FATALITIES, na.rm = T),
                                    mean.fatalities = mean(FATALITIES, na.rm = T),
                                    total.damage = sum(damage, na.rm = T))

head(top10.fatality.year.summary)
```

```
## Source: local data frame [6 x 7]
## Groups: year [6]
##
##    year  EVTYPE total.injuries mean.injuries total.fatalities
##   <dbl>   <chr>          <dbl>         <dbl>            <dbl>
## 1  1950 TORNADO            659      2.955157               70
## 2  1951 TORNADO            524      1.947955               34
## 3  1952 TORNADO           1915      7.040441              230
## 4  1953 TORNADO           5131     10.428862              519
## 5  1954 TORNADO            715      1.174056               36
## 6  1955 TORNADO            926      1.465190              129
## # ... with 2 more variables: mean.fatalities <dbl>, total.damage <dbl>
```

```
top10.damage.year.summary <- filter(storm.data, EVTYPE %in% head(top40.damage.event$EVTYPE,10)) %>%   g
                                    mean.injuries = mean(INJURIES, na.rm = T),
                                    total.fatalities = sum(FATALITIES, na.rm = T),
                                    mean.fatalities = mean(FATALITIES, na.rm = T),
                                    total.damage = sum(damage, na.rm = T)) %>% arrange(-total.damage)

head(top10.damage.year.summary)
```

```
## Source: local data frame [6 x 7]
## Groups: year [4]
##
##    year    EVTYPE total.injuries mean.injuries total.fatalities
##   <dbl>     <chr>          <dbl>         <dbl>            <dbl>
## 1  2006     FLOOD              5   0.004135649               19
## 2  2011   TORNADO           6163   2.811587591              587
## 3  2011     FLOOD             10   0.002764722               58
```

```
## 4  1994 ICE STORM           1614  36.681818182               2
## 5  2010     FLOOD            127   0.047942620              41
## 6  2010      HAIL             42   0.003845450               0
## # ... with 2 more variables: mean.fatalities <dbl>, total.damage <dbl>
```

Results

```
bar.plot1 <- ggplot() + geom_bar(data = top40.fatality.event, aes(EVTYPE, total.injuries), stat = "iden
  theme(axis.text.x=element_text(angle=90,hjust=1)) + ggtitle("Fig 1. Injuries caused by top 40 events"

print(bar.plot1)
```

```
bar.plot2 <- ggplot() + geom_bar(data = top40.fatality.event, aes(EVTYPE, total.damage), stat = "identi
  theme(axis.text.x=element_text(angle=90,hjust=1)) + ggtitle("Fig 2. Damage caused by top 40 events")

print(bar.plot2)
```

- From figure 1 it is clear tornados are the largest cause of injuries out of the dataset between 1950 and 2011.
- From figure 1 it is clear floods are the largest cause of damage out of the dataset between 1950 and 2011.

```
line.plot1 <- ggplot() + geom_line(data = top10.fatality.year.summary, aes(year, total.injuries, colour

print(line.plot1)
```

- From figure 3 it is clear tornados are persistantly the largest contributor to injury.

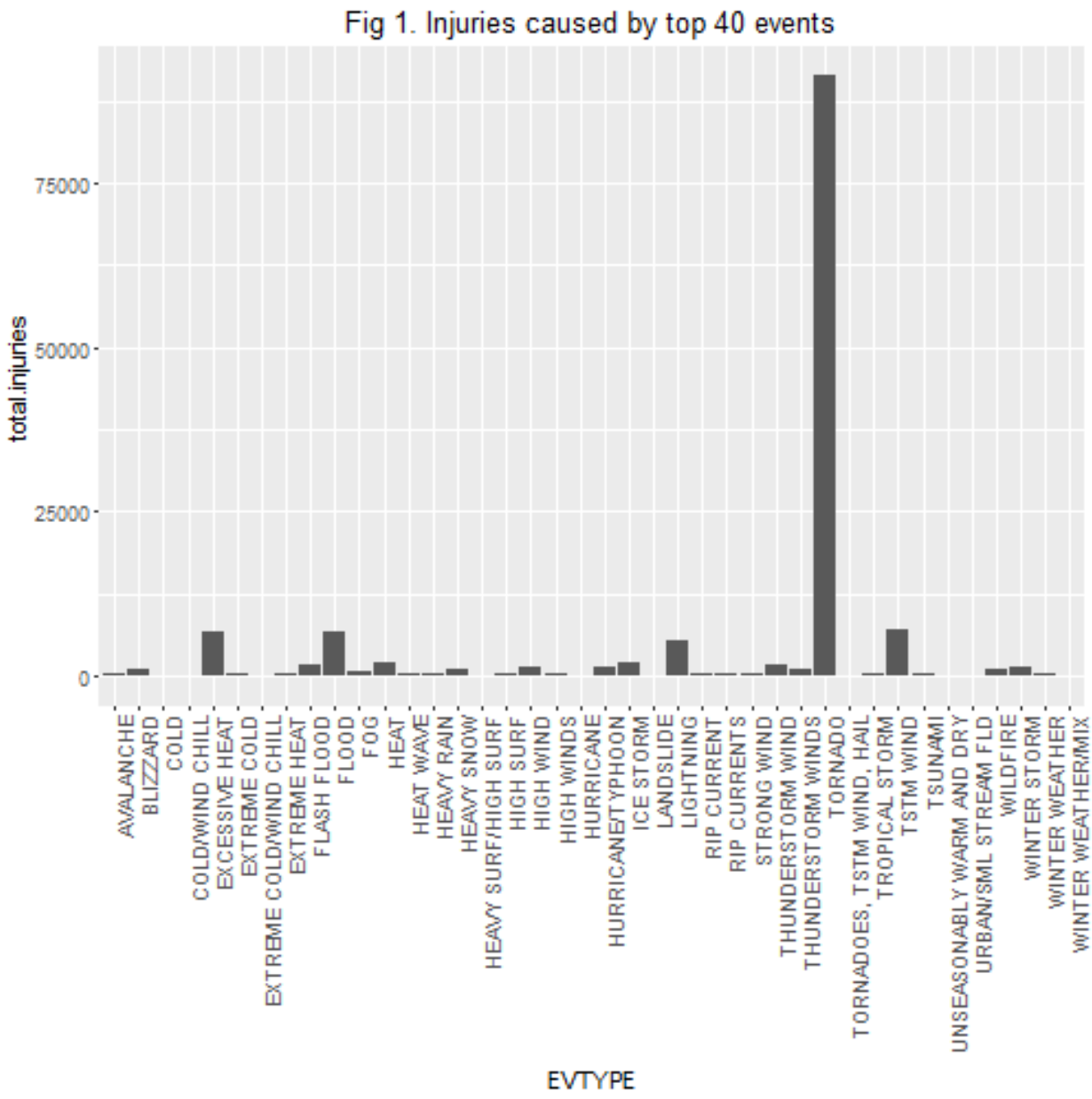

Figure 1: plot of chunk results1

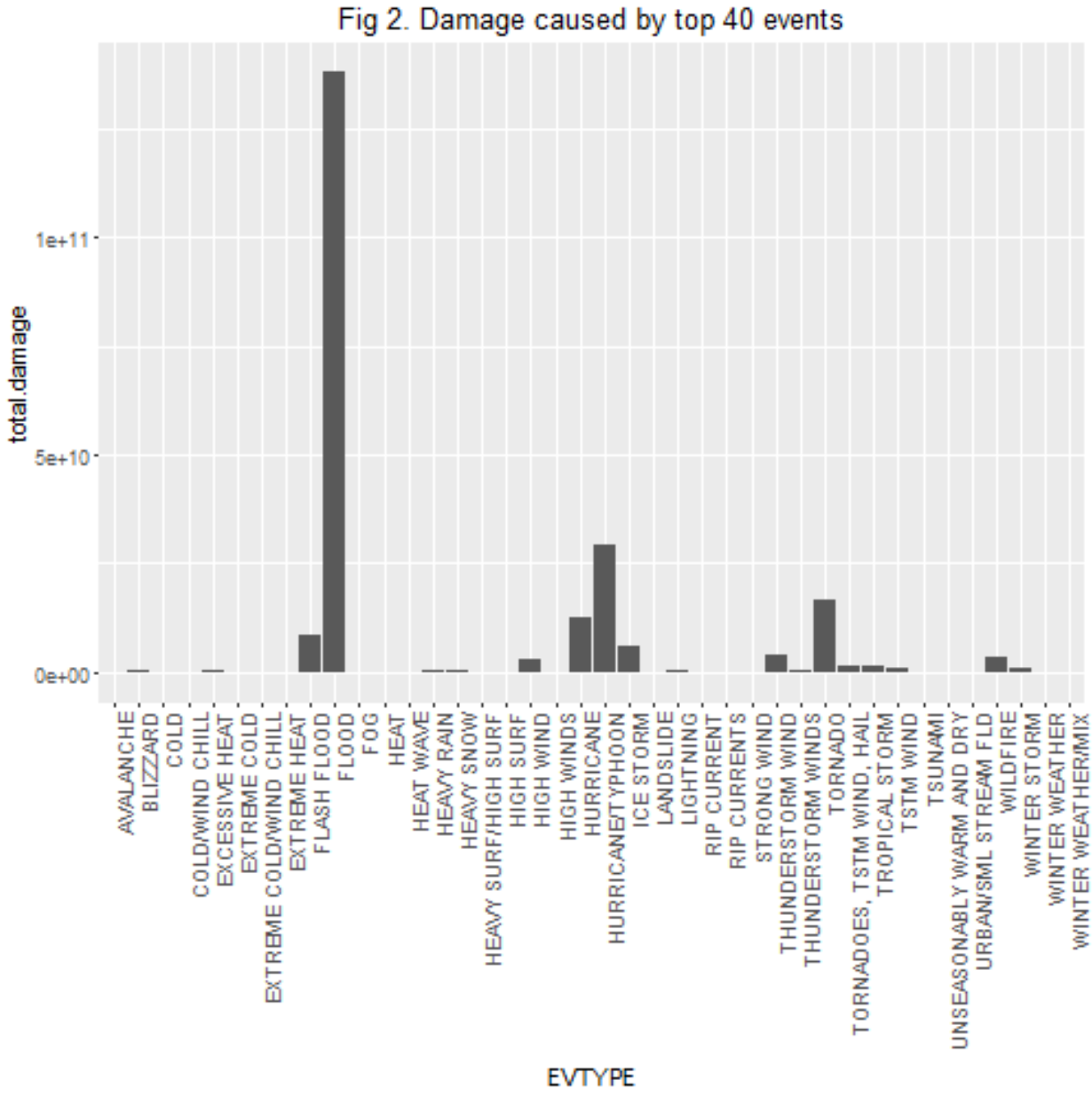

Figure 2: plot of chunk results1

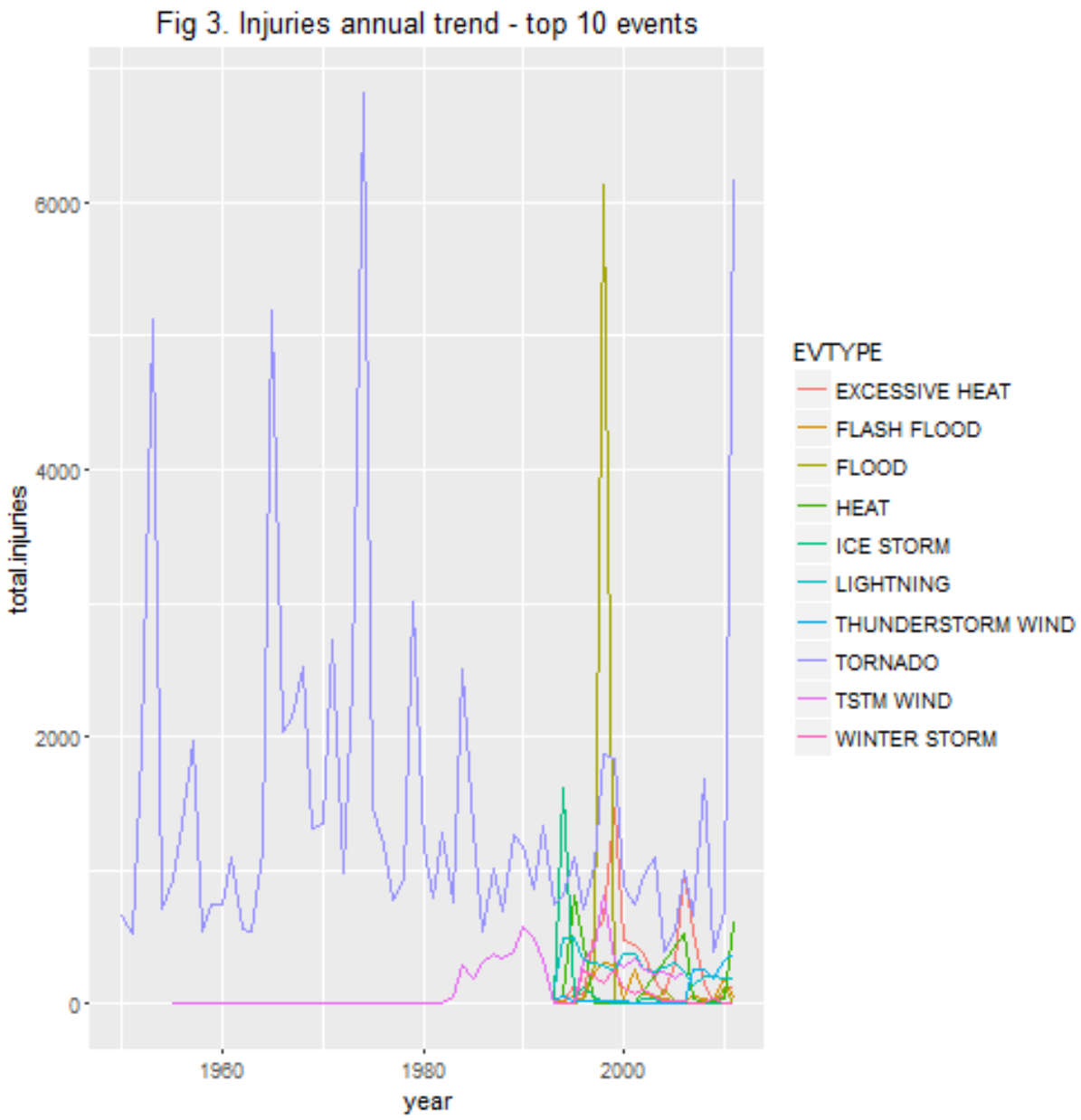

Figure 3: plot of chunk results2