Characteristics of major storms and weather events in the United States across the period 1996-2001

Characteristics of major storms and weather events in the United States across the period 1996-2001

Synopsis

This report aims to present the analysis of the NOAA Storm Database and offer answers to the questions proposed in the assignment instructions. After a preliminary analysis of the data structure and NCDC documentation, we established a valid time period for the weather events and limited the fields to be used. Other additional tasks related to cleaning data were performed, but our focus in this work has been the exploration and analysis of the provided dataset. We have found that extreme hot temperatures and tornados lead the negative effect on human lifes and economic damage, but it seems that the importance of hurricane Katrina in 2005 distorts the results.

Loading and Processing the Raw Data

The dataset analysed in this report was obtained from the link made available at the instructions provided for this assignment.According to the National Climatic Data Center (NCDC), this dataset records 48 event types from 1996 to present; before that period a reduced number of events has been incorporated. The assignment instructions stablished that the events in the database start in the year 1950 and end in November 2011. Given those constraints, this report analyses the data corresponding to the 1996-2011 period.

Downloading and reading the NCDC file

We first download the raw file and read the data including the header.

```
library(data.table)
library(ggplot2)
library(xtable)

# The palette
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73", "#F0E442", "#0072B2",
    "#D55E00", "#CC79A7", "gold", "green3")

# To use for fills
v <- scale_fill_manual(values = cbPalette)
# To use for line and point colors
v <- scale_colour_manual(values = cbPalette)


## setInternet2(TRUE) f <- tempfile()
## download.file('https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2',
## f) datZero <- data.table(read.csv(bzfile(f)))

datZero <- data.table(read.csv("repdata_data_StormData.csv", sep = ",", header = TRUE))
```

In order to process only the 1996-2011 period, we filter the data produced in other years.

```
# cutting years until 1996 and comments, adding formatted date
dat1 <- datZero[year(strptime(datZero[["BGN_DATE"]], "%m/%d/%Y %H:%M:%S")) >=
    1996]
```

The number of fields ignored by our analysis increases the size of the dataset to be manipulated; with the purpose of work with a smaller data object,we are using only the columns with values significative to our analysis.

```
dat1 <- dat1[, c("BGN_DATE", "STATE", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG",
    "PROPDMGEXP", "CROPDMG", "CROPDMGEXP"), with = FALSE]
```

The number of records with zero value for all the analysis-significative columns is very important.

```
nrow(dat1[FATALITIES == 0 & INJURIES == 0 & PROPDMG == 0 & CROPDMG == 0])

## [1] 452212
```

Therefore, we cut off the rows without values different from zero.

```
dat1 <- dat1[FATALITIES > 0 | INJURIES > 0 | PROPDMG > 0 | CROPDMG > 0, ]
```

Finally, a valid Datetime-formatted column is incorporated in order to facilitate further manipulation of data.

```
dat1 <- dat1[, `:=`(bgn_date2, as.Date(strptime(dat1$BGN_DATE, "%m/%d/%Y %H:%M:%S")))]
dat1 <- dat1[, `:=`(bgn_year, year(bgn_date2))]
```

Obtaining totals

According to the NCDC, the column PROPDMGEXP is related to the value of PROPDMG and it stores alphabetical characters used to represent magnitude ("K" for thousands, "M" for millions, and "B" for billions). We use this characteristic in order to obtain a homogeneous representation for the value of damage to properties.

```
dat1 <- dat1[, `:=`(propdmg0, 0)]
dat1 <- dat1[, `:=`(propdmgK, 0)]
dat1 <- dat1[, `:=`(propdmgM, 0)]
dat1 <- dat1[, `:=`(propdmgB, 0)]

dat1 <- dat1[as.character(dat1$PROPDMGEXP) == "0" & "PROPDMG" > 0, `:=`(propdmg0,
    PROPDMG * 1)]
dat1 <- dat1[as.character(dat1$PROPDMGEXP) == "K" & "PROPDMG" > 0, `:=`(propdmgK,
    PROPDMG * 1000)]
dat1 <- dat1[as.character(dat1$PROPDMGEXP) == "M" & "PROPDMG" > 0, `:=`(propdmgM,
    PROPDMG * 1e+06)]
dat1 <- dat1[as.character(dat1$PROPDMGEXP) == "B" & "PROPDMG" > 0, `:=`(propdmgB,
    PROPDMG * 1e+09)]

dat1 <- dat1[is.na(dat1$propdmg0), `:=`(propdmg0, 0)]
dat1 <- dat1[is.na(dat1$propdmgK), `:=`(propdmgK, 0)]
dat1 <- dat1[is.na(dat1$propdmgM), `:=`(propdmgM, 0)]
dat1 <- dat1[is.na(dat1$propdmgB), `:=`(propdmgB, 0)]
```

A similar criteria is applied to the value of damage to crops.

```
dat1 <- dat1[, `:=`(cropdmg0, 0)]
dat1 <- dat1[, `:=`(cropdmgK, 0)]
dat1 <- dat1[, `:=`(cropdmgM, 0)]
dat1 <- dat1[, `:=`(cropdmgB, 0)]

dat1 <- dat1[as.character(dat1$CROPDMGEXP) == "0" & "CROPDMG" > 0, `:=`(cropdmg0,
    CROPDMG * 1)]
dat1 <- dat1[as.character(dat1$CROPDMGEXP) == "K" & "CROPDMG" > 0, `:=`(cropdmgK,
    CROPDMG * 1000)]
```

```
dat1 <- dat1[as.character(dat1$CROPDMGEXP) == "M" & "CROPDMG" > 0, `:=`(cropdmgM,
    CROPDMG * 1e+06)]
dat1 <- dat1[as.character(dat1$CROPDMGEXP) == "B" & "CROPDMG" > 0, `:=`(cropdmgB,
    CROPDMG * 1e+09)]

dat1 <- dat1[is.na(dat1$cropdmg0), `:=`(cropdmg0, 0)]
dat1 <- dat1[is.na(dat1$cropdmgK), `:=`(cropdmgK, 0)]
dat1 <- dat1[is.na(dat1$cropdmgM), `:=`(cropdmgM, 0)]
dat1 <- dat1[is.na(dat1$cropdmgB), `:=`(cropdmgB, 0)]
```

Results

The total effect of Weather events in the U.S.

In order to establish the importance of these events in terms of economic losses and human lifes, we have totalized by event both the economic damage and fatalities occurred. To avoid big-numbers display issues, we are converting those totals to thousand-dollars units. Finally, we obtain a unique representation for each event.

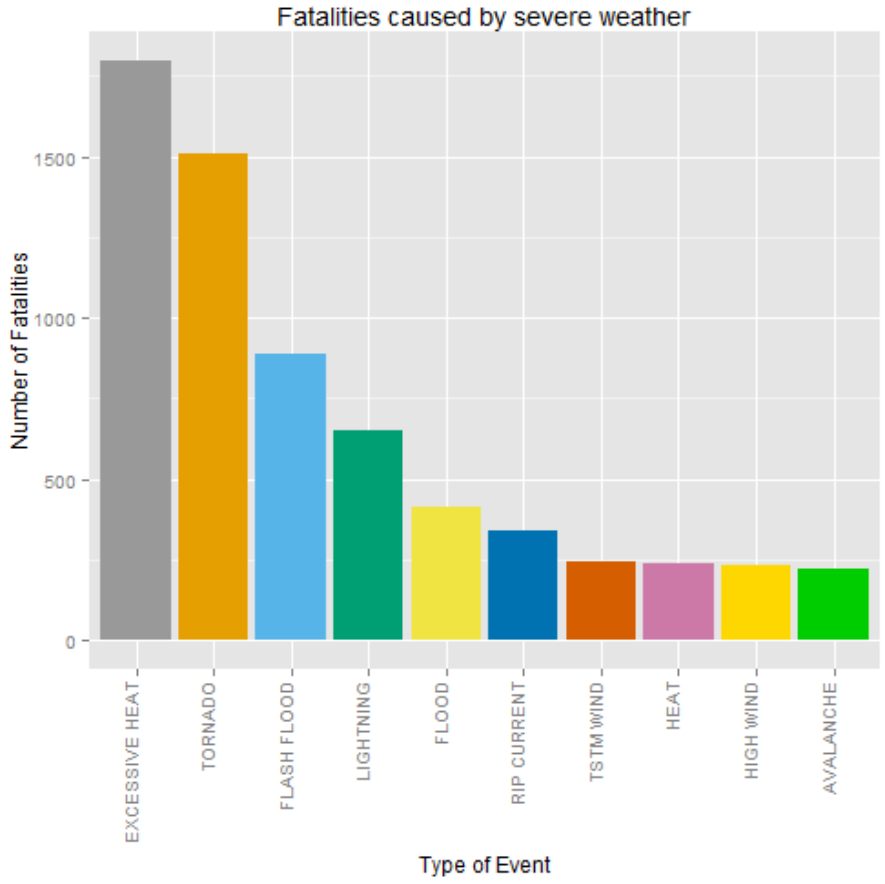
```
datTotEvent <- dat1[order(EVTYPE, bgn_year)]
datTotEvent <- datTotEvent[, by = "EVTYPE", `:=`(totFATALITIES = sum(FATALITIES),
    totINJURIES = sum(INJURIES), totPROPDMG = sum(propdmg0/1000) + sum(propdmgK/1000) +
        sum(propdmgM/1000) + sum(propdmgB/1000), totCROPDMG = sum(cropdmg0/1000) +
        sum(cropdmgK/1000) + sum(cropdmgM/1000) + sum(cropdmgB/1000))]
datTotEvent <- unique(datTotEvent, by = "EVTYPE")
```

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

As the question requests for information about damage inflicted to population health, we are using the number of fatalities caused by a event to find the most harmful meteorological phenomenon.

```
datQ1 <- datTotEvent[order(totFATALITIES, totINJURIES, decreasing = TRUE)]
datQ1 <- datQ1[totFATALITIES > 0, ]
datQ1 <- datQ1[order(totFATALITIES, decreasing = TRUE)]

levelsQ1 <- datQ1[, c("EVTYPE", "totFATALITIES"), with = FALSE]
levelsQ1 <- levelsQ1[order(totFATALITIES, decreasing = TRUE)]
levelsQ1 <- levelsQ1[["EVTYPE"]]

datQ1 <- transform(datQ1, EVTYPE = factor(EVTYPE, levels = levelsQ1))
```

```
datQ1 <- head(datQ1, 10)

## Question 1

qplot(x = EVTYPE, y = totFATALITIES, data = datQ1, fill = EVTYPE, geom = "bar",
    stat = "identity", xlab = "Type of Event", ylab = "Number of Fatalities") +
    labs(title = "Fatalities caused by severe weather") + opts(legend.position = "none") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust = 0)) + scale_fill_manual(

## 'opts' is deprecated. Use 'theme' instead. (Deprecated; last used in version 0.9.1)
```



Answering Question 2: Across the United States, which types of events have the greatest economic consequences?

The answer to this question relates to property and crop damage; therefore we are aggregating these values into a total for economic losses. After that, we order the data by the economic totals and obtain the 10 most significant events in terms of damage to properties and crops.

```
datQ2 <- datTotEvent[, `:=`(totECONDMG, totPROPDMG + totCROPDMG)]
datQ2 <- datTotEvent[totECONDMG > 0, ]
datQ2 <- datTotEvent[order(totECONDMG, decreasing = TRUE)]

levelsQ2 <- datQ2[, c("EVTYPE", "totECONDMG"), with = FALSE]
levelsQ2 <- levelsQ2[order(totECONDMG, decreasing = TRUE)]
levelsQ2 <- levelsQ2[["EVTYPE"]]

datQ2 <- transform(datQ2, EVTYPE = factor(EVTYPE, levels = levelsQ2))

datQ2 <- datQ2[, list(EVTYPE, totFATALITIES, totINJURIES, totPROPDMG, totCROPDMG,
    totECONDMG)]
setnames(datQ2, c("Event", "Fatalities", "Injuries", "PropertyDamage", "CropsDamage",
    "EconomicDamage"))
datQ2 <- head(datQ2, 10)

## Question 2

## d.table <- xtable(datQ2)

d.table <- xtable(prettyNum(datQ2, align = "rllllll", format = "fg"), caption = "Economic dam


print(d.table, type = "html", caption.placement = "top")
```

Economic damage occassioned by weather events

Event

Fatalities

Injuries

PropertyDamage

CropsDamage

EconomicDamage

1

FLOOD

414

6758

143944834

4974778

148919612

2

HURRICANE/TYPHOON

64

1275

69305840

2607873

71913713

3

STORM SURGE

2

37

43193536

5

43193541

4

TORNADO

1511

20667

24616946

283425

24900371

5

HAIL

7

713

14595143

2476029

17071173

6

FLASH FLOOD

887

1674

15222204

1334902

16557106

7

HURRICANE

61

46

11812819

2741410

14554229

8

DROUGHT

0

4

1046101

13367566

14413667

9

TROPICAL STORM

57

338

7642476

677711

8320187

10

HIGH WIND

235

1083

5247860

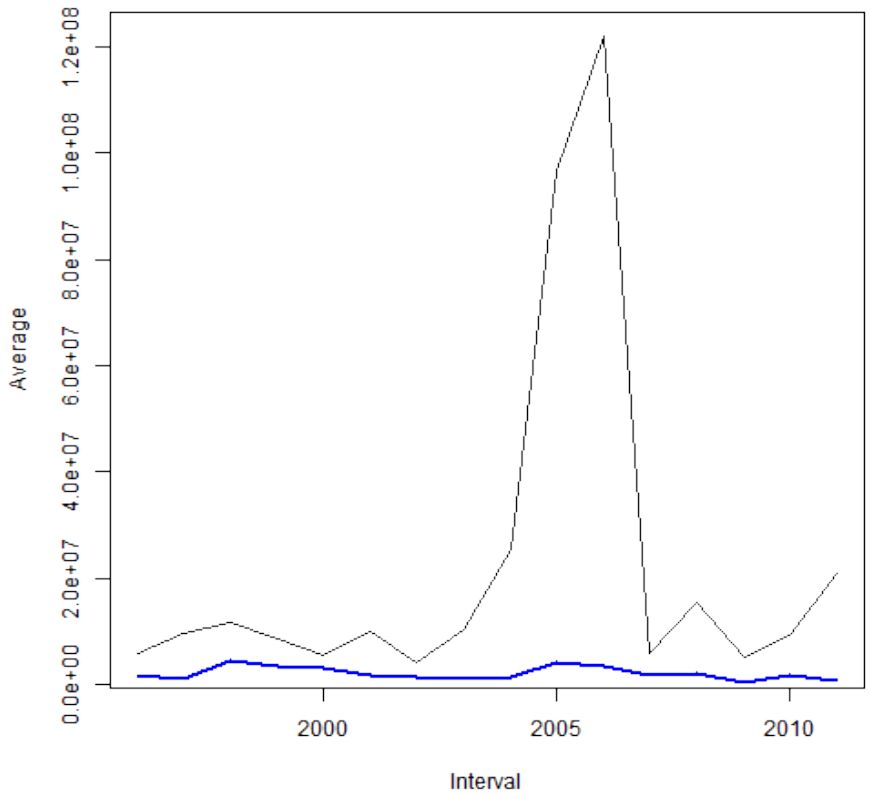
633561

5881422

Additional

Along the period

```
datTotYear <- dat1[order(bgn_year, EVTYPE)]
datTotYear <- datTotYear[, by = year(datTotYear$bgn_date2), `:=`(totFATALITIES = sum(FATALIT
    totINJURIES = sum(INJURIES), totPROPDMG = sum(propdmg0/1000) + sum(propdmgK/1000) +
        sum(propdmgM/1000) + sum(propdmgB/1000), totCROPDMG = sum(cropdmg0/1000) +
        sum(cropdmgK/1000) + sum(cropdmgM/1000) + sum(cropdmgB/1000))]
datTotYear <- unique(datTotYear, by = "bgn_year")
datTotYear <- datTotYear[, list(bgn_year, EVTYPE, totFATALITIES, totINJURIES,
    totPROPDMG, totCROPDMG)]


with(datTotYear, plot(bgn_year, totPROPDMG, pch = NA, type = "l", ylab = "Average",
    xlab = "Interval"))
lines(datTotYear$bgn_year, datTotYear$totCROPDMG, col = "blue", lwd = 2)
```



Across the US

```
datTotGeo <- dat1[order(STATE, EVTYPE)]
datTotGeo <- datTotGeo[, by = STATE, `:=`(totFATALITIES = sum(FATALITIES), totINJURIES = sum
    totPROPDMG = sum(propdmg0/1e+06) + sum(propdmgK/1e+06) + sum(propdmgM/1e+06) +
        sum(propdmgB/1e+06), totCROPDMG = sum(cropdmg0/1e+06) + sum(cropdmgK/1e+06) +
        sum(cropdmgM/1e+06) + sum(cropdmgB/1e+06))]

datTotGeo <- unique(datTotGeo, by = "STATE")
datTotGeo <- datTotGeo[, list(STATE, totFATALITIES, totINJURIES, totPROPDMG,
    totCROPDMG)]

datTotGeo <- datTotGeo[order(totPROPDMG, decreasing = TRUE)]
datTotGeo <- head(datTotGeo, 10)
```

```
datTotGeo$STATE <- as.character(datTotGeo$STATE)
datTotGeo$STATE <- as.factor(datTotGeo$STATE)

ggplot(datTotGeo, aes(STATE, totPROPDMG, fill = STATE)) + geom_bar(stat = "identity",
    binwidth = 0.7) + labs(title = "Property damage from 1996-2011 for USA") +
    labs(x = "State", y = "Property damage (Thousand Millions)") + scale_fill_manual(values
```