

Storm_data_assignment

Believe Ayodele

3 May 2016

```
## Importing the dplyr library

library (dplyr)

## Warning: package 'dplyr' was built under R version 3.2.5

##
## 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

dat <-
download.file('https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2', 'C:/Apps/python/R/data.csv.bz2')

dat <- read.csv('C:/Apps/python/R/data.csv.bz2', stringsAsFactors = F)

dat2 <- dat

names (dat2) <- tolower(names(dat2))

dat3 <- select (tbl_df(dat2), evtype, fatalities:cropdmgexp)
sum(is.na(dat3))

## [1] 0

names (dat3)

## [1] "evtype"      "fatalities" "injuries"    "propdmg"     "propdmgexp"
## [6] "cropdmg"     "cropdmgexp"

##Let's look at variables, there are many garbage, need manipulating and
cleaning

unique(dat3$propdmgexp)
```

```
## [1] "K" "M" "" "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-"
## [18] "1" "8"
```

```
unique(dat3$cropdmgexp)
```

```
## [1] "" "M" "K" "m" "B" "?" "0" "k" "2"
```

```
dat3$propdmgexp [
(dat3$propdmgexp=='')|(dat3$propdmgexp=='+')|(dat3$propdmgexp=='?')|
(dat3$propdmgexp=='-')| (dat3$propdmgexp=='0')| (dat3$propdmgexp=='h') |
(dat3$propdmgexp=='H')] <-0
dat3$propdmgexp[(dat3$propdmgexp=='K')] <- 3
dat3$propdmgexp[(dat3$propdmgexp=='M')|(dat3$propdmgexp=='m')] <- 6
dat3$propdmgexp[(dat3$propdmgexp=='B')] <- 9
dat3$cropdmgexp[(dat3$cropdmgexp=='')|(dat3$cropdmgexp=='?')|(dat3$cropdmgexp
=='0')] <- 1
dat3$cropdmgexp[(dat3$cropdmgexp=='K')|(dat3$cropdmgexp=='k')] <- 3
dat3$cropdmgexp[(dat3$cropdmgexp=='M')|(dat3$cropdmgexp=='m')] <- 6
dat3$cropdmgexp[(dat3$cropdmgexp=='B')] <- 9
dat3$propdmgexp <- as.numeric(dat3$propdmgexp)
dat3$cropdmgexp <- as.numeric(dat3$cropdmgexp)
```

```
dat3$propdmg <- dat3$propdmg*(10^dat3$propdmgexp)
dat3$cropdmg <- dat3$cropdmg*(10^dat3$cropdmgexp)
```

```
head(unique(dat3$evtype), 30)
```

```
## [1] "TORNADO" "TSTM WIND"
## [3] "HAIL" "FREEZING RAIN"
## [5] "SNOW" "ICE STORM/FLASH FLOOD"
## [7] "SNOW/ICE" "WINTER STORM"
## [9] "HURRICANE OPAL/HIGH WINDS" "THUNDERSTORM WINDS"
## [11] "RECORD COLD" "HURRICANE ERIN"
## [13] "HURRICANE OPAL" "HEAVY RAIN"
## [15] "LIGHTNING" "THUNDERSTORM WIND"
## [17] "DENSE FOG" "RIP CURRENT"
## [19] "THUNDERSTORM WINS" "FLASH FLOOD"
## [21] "FLASH FLOODING" "HIGH WINDS"
## [23] "FUNNEL CLOUD" "TORNADO F0"
## [25] "THUNDERSTORM WINDS LIGHTNING" "THUNDERSTORM WINDS/HAIL"
## [27] "HEAT" "WIND"
## [29] "LIGHTING" "HEAVY RAINS"
```

```
length(unique(dat3$evtype))
```

```
## [1] 985
```

```
dat3$evtype <- gsub('.*STORM.*', 'STORM', dat3$evtype)
dat3$evtype <- gsub('.*FLOOD.*', 'FLOOD', dat3$evtype)
dat3$evtype <- gsub('.*WIND.*', 'WIND', dat3$evtype)
```

```

dat3$evtype <- gsub('.*TORN.*', 'TORNADO', dat3$evtype)
dat3$evtype <- gsub('.*HAIL.*', 'HAIL', dat3$evtype)
dat3$evtype <- gsub('.*HURRICANE.*', 'HURRICANE', dat3$evtype)
dat3$evtype <- gsub('.*RAIN.*', 'RAIN', dat3$evtype)
dat3$evtype <- gsub('.*SNOW.*', 'SNOW', dat3$evtype)
dat3$evtype <- gsub('.*COLD.*', 'COLD', dat3$evtype)
dat3$evtype <- gsub('.*LOW.*TEMPER.*', 'COLD', dat3$evtype)
dat3$evtype <- gsub('.*FROST.*', 'COLD', dat3$evtype)
dat3$evtype <- gsub('.*HIGH.*TEMPER.*', 'HEAT', dat3$evtype)
dat3$evtype <- gsub('.*HEAT.*', 'HEAT', dat3$evtype)
dat3$evtype <- gsub('.*FIRE.*', 'FIRE', dat3$evtype)

length(unique(dat3$evtype))

## [1] 433

## Grouping data by event type and looking at fatalities, injuries and
economic damage

dat3 <- group_by(dat3, evtype)
dat4 <- summarise(dat3, all_fatalities=sum(fatalities),
all_injuries=sum(injuries),
all_propdmg=sum(propdmg), all_cropdmg=sum(cropdmg))

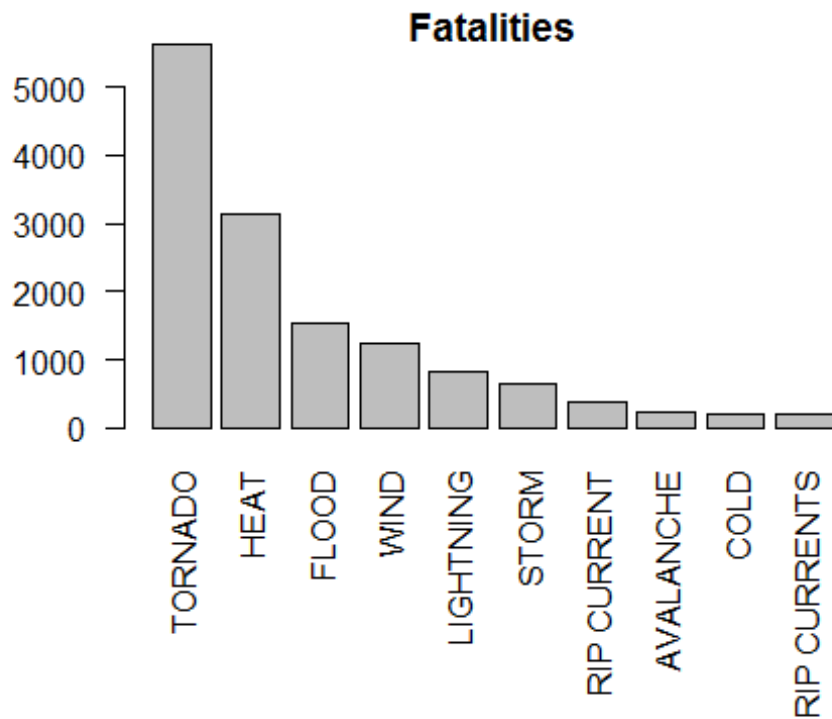
table_fatalities <- arrange(select(dat4, evtype, all_fatalities),
desc(all_fatalities))[1:10,]
table_fatalities

## Source: local data frame [10 x 2]
##
##      evtype all_fatalities
##      (chr)      (dbl)
## 1  TORNADO      5636
## 2   HEAT       3138
## 3  FLOOD       1523
## 4   WIND       1235
## 5 LIGHTNING      816
## 6   STORM       633
## 7 RIP CURRENT   368
## 8 AVALANCHE     224
## 9    COLD       215
## 10 RIP CURRENTS  204

## Drawing the various charts

par(mar=c(9,5,1,1))
barplot(height = table_fatalities$all_fatalities, names.arg =
table_fatalities$evtype, main = 'Fatalities', las=2)

```



```
table_injuries <- arrange(select(dat4, evtype, all_injuries),
desc(all_injuries))[1:10,]
table_injuries
```

```
## Source: local data frame [10 x 2]
```

```
##
##      evtype all_injuries
##      (chr)      (dbl)
## 1  TORNADO      9140
## 2    HEAT       915
## 3    WIND       904
## 4   FLOOD       860
## 5   STORM       669
## 6 LIGHTNING     523
## 7    FIRE       160
## 8    HAIL       137
## 9  HURRICANE     132
## 10   SNOW       111
```

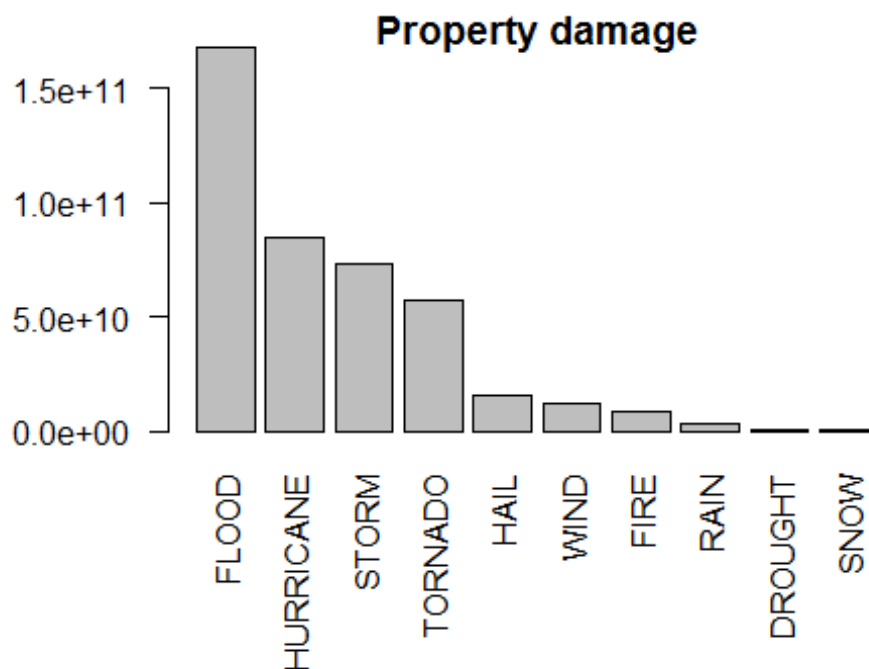
```
table_propdmg <- arrange(select(dat4, evtype, all_propdmg),
desc(all_propdmg))[1:10,]
table_propdmg
```

```
## Source: local data frame [10 x 2]
```

```
##
##      evtype all_propdmg
##      (chr)      (dbl)
```

```
## 1      FLOOD 168061094835
## 2  HURRICANE 84656180010
## 3      STORM 73263643688
## 4    TORNADO 57003318427
## 5      HAIL 15736043018
## 6      WIND 12450581618
## 7      FIRE 8501628500
## 8      RAIN 3233664190
## 9    DROUGHT 1046106000
## 10     SNOW 1014264750
```

```
par(mar=c(9,5,1,1))
barplot(height = table_propdmg$all_propdmg, names.arg = table_propdmg$evtype,
main = 'Property damage', las=2)
```

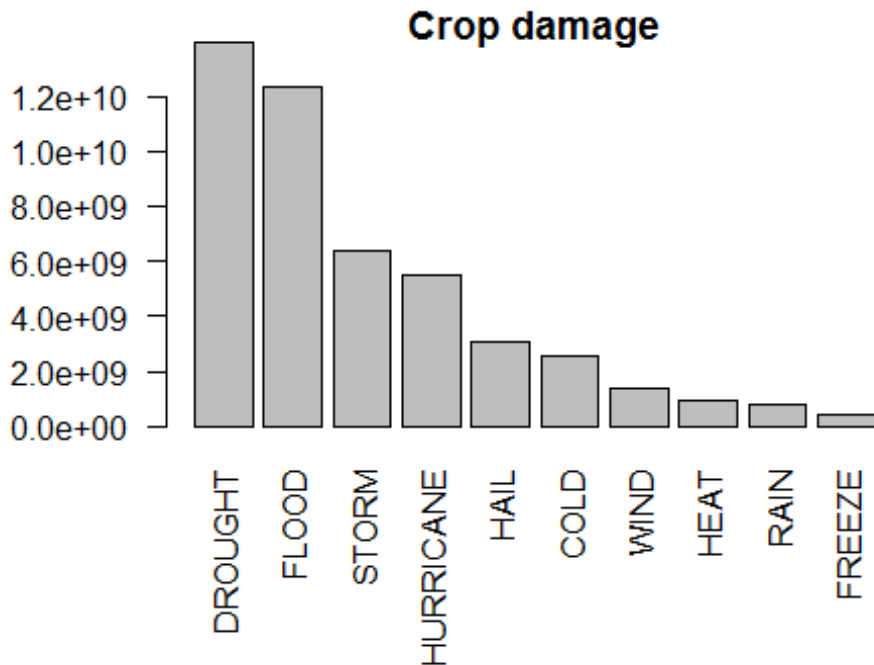


```
table_cropdmg <- arrange(select(dat4, evtype, all_cropdmg),
desc(all_cropdmg))[1:10,]
table_cropdmg
```

```
## Source: local data frame [10 x 2]
##
##      evtype all_cropdmg
##      (chr)      (dbl)
## 1  DROUGHT 13972566000
## 2   FLOOD 12352059100
## 3   STORM 6406919680
## 4 HURRICANE 5505292800
```

```
## 5      HAIL  3046837680
## 6      COLD  2544101500
## 7      WIND  1406229150
## 8      HEAT   904469280
## 9      RAIN   804652800
## 10     FREEZE  446225000
```

```
par(mar=c(9,5,1,1))
barplot(height = table_cropdmg$all_cropdmg, names.arg = table_cropdmg$evtype,
main = 'Crop damage', las=2)
```



```
compare_table <- matrix(nrow = 10, ncol = 4)
compare_table[,1] <- arrange(dat4, desc(all_fatalities))$evtype[1:10]
compare_table[,2] <- arrange(dat4, desc(all_injuries))$evtype[1:10]
compare_table[,3] <- arrange(dat4, desc(all_propdmg))$evtype[1:10]
compare_table[,4] <- arrange(dat4, desc(all_cropdmg))$evtype[1:10]
colnames(compare_table) <- c('all_fatalities', 'all_injuries', 'all_propdmg',
'all_cropdmg')
compare_table
```

```
##      all_fatalities all_injuries all_propdmg all_cropdmg
## [1,] "TORNADO"      "TORNADO"      "FLOOD"       "DROUGHT"
## [2,] "HEAT"         "HEAT"         "HURRICANE"   "FLOOD"
## [3,] "FLOOD"        "WIND"         "STORM"       "STORM"
## [4,] "WIND"         "FLOOD"        "TORNADO"     "HURRICANE"
## [5,] "LIGHTNING"    "STORM"        "HAIL"        "HAIL"
## [6,] "STORM"        "LIGHTNING"    "WIND"        "COLD"
```

##	[7,]	"RIP CURRENT"	"FIRE"	"FIRE"	"WIND"
##	[8,]	"AVALANCHE"	"HAIL"	"RAIN"	"HEAT"
##	[9,]	"COLD"	"HURRICANE"	"DROUGHT"	"RAIN"
##	[10,]	"RIP CURRENTS"	"SNOW"	"SNOW"	"FREEZE"