Proyecto2

Nosequiensoy007

12/11/2020

set libraries

```
library(dplyr)
```

```
##
## 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
```

```
library(ggplot2)
library(summarytools)
```

```
## Registered S3 method overwritten by 'pryr':
##   method      from
##   print.bytes Rcpp
## For best results, restart R session and update pander using devtool
s:: or remotes::install_github('rapporter/pander')
```

```
library(cowplot)
```

#After downloadzip, file, unzip and then change wd where folder and file are in my case

```
setwd("C:/Users/juanfidel18/Desktop/Coursera/ReproducibleResearch")
```

reading and Procesing Data

```
stormdata <- read.csv("repdata_data_StormData.csv")
```

Check dimensions and structure

```
str(stormdata)
```

```
## 'data.frame':    902297 obs. of  37 variables:
##  $ STATE__   : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ BGN_DATE  : chr  "4/18/1950 0:00:00" "4/18/1950 0:00:00" "2/20/1
951 0:00:00" "6/8/1951 0:00:00" ...
##  $ BGN_TIME  : chr  "0130" "0145" "1600" "0900" ...
##  $ TIME_ZONE : chr  "CST" "CST" "CST" "CST" ...
##  $ COUNTY    : num  97 3 57 89 43 77 9 123 125 57 ...
##  $ COUNTYNAME: chr  "MOBILE" "BALDWIN" "FAYETTE" "MADISON" ...
##  $ STATE     : chr  "AL" "AL" "AL" "AL" ...
##  $ EVTYPE    : chr  "TORNADO" "TORNADO" "TORNADO" "TORNADO" ...
##  $ BGN_RANGE : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ BGN_AZI   : chr  "" "" "" "" ...
##  $ BGN_LOCATI: chr  "" "" "" "" ...
##  $ END_DATE  : chr  "" "" "" "" ...
##  $ END_TIME  : chr  "" "" "" "" ...
##  $ 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   : chr  "" "" "" "" ...
##  $ END_LOCATI: chr  "" "" "" "" ...
##  $ LENGTH    : num  14 2 0.1 0 0 1.5 1.5 0 3.3 2.3 ...
##  $ WIDTH     : num  100 150 123 100 150 177 33 33 100 100 ...
##  $ F         : int  3 2 2 2 2 2 2 1 3 3 ...
##  $ MAG       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ FATALITIES: num  0 0 0 0 0 0 0 0 1 0 ...
##  $ INJURIES  : num  15 0 2 2 2 6 1 0 14 0 ...
##  $ PROPDMG   : num  25 2.5 25 2.5 2.5 2.5 2.5 2.5 25 25 ...
##  $ PROPDMGEXP: chr  "K" "K" "K" "K" ...
##  $ CROPDMG   : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ CROPDMGEXP: chr  "" "" "" "" ...
##  $ WFO       : chr  "" "" "" "" ...
##  $ STATEOFFIC: chr  "" "" "" "" ...
##  $ ZONENAMES : chr  "" "" "" "" ...
##  $ LATITUDE  : num  3040 3042 3340 3458 3412 ...
##  $ LONGITUDE : num  8812 8755 8742 8626 8642 ...
```

```
##  $ LATITUDE_E: num  3051 0 0 0 0 ...
##  $ LONGITUDE_: num  8806 0 0 0 0 ...
##  $ REMARKS   : chr  "" "" "" "" ...
##  $ REFNUM    : num  1 2 3 4 5 6 7 8 9 10 ...
```

```
freq(stormdata$EVTYPE, order = "freq", rows = 1:10)
```

```
## Frequencies
## stormdata$EVTYPE
## Type: Character
##
##                              Freq   % Valid   % Valid Cum.   % Tota
l   % Total Cum.
## ------------------------ -------- --------- -------------- --------
- --------------
##                      HAIL   288661     31.99          31.99     31.9
9          31.99
##                 TSTM WIND   219940     24.38          56.37     24.3
8          56.37
##         THUNDERSTORM WIND    82563      9.15          65.52      9.1
5          65.52
##                   TORNADO    60652      6.72          72.24      6.7
2          72.24
##               FLASH FLOOD    54277      6.02          78.26      6.0
2          78.26
##                     FLOOD    25326      2.81          81.06      2.8
1          81.06
##        THUNDERSTORM WINDS    20843      2.31          83.37      2.3
1          83.37
##                 HIGH WIND    20212      2.24          85.61      2.2
4          85.61
##                 LIGHTNING    15754      1.75          87.36      1.7
5          87.36
##               HEAVY SNOW    15708      1.74          89.10      1.7
4          89.10
##                   (Other)    98361     10.90         100.00     10.9
0         100.00
##                      <NA>        0                               0.0
0         100.00
##                     Total   902297    100.00         100.00    100.0
0         100.00
```

```
summary(stormdata$INJURIES)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
##    0.0000    0.0000    0.0000    0.1557    0.0000 1700.0000
```

```
summary(stormdata$FATALITIES)
```

```
##      Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
##   0.0000   0.0000   0.0000   0.0168   0.0000 583.0000
```

```
summary(stormdata$PROPDMG)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00    0.00    0.00   12.06    0.50 5000.00
```

```
freq(stormdata$PROPDMGEXP)
```

```
## Frequencies
## stormdata$PROPDMGEXP
## Type: Character
##
##                          Freq   % Valid   % Valid Cum.   % Total
% Total Cum.
## -------------------- -------- --------- -------------- --------- --
------------
##       (Empty string)   465934     51.64          51.64     51.64
51.64
##                    -        1      0.00          51.64      0.00
51.64
##                    ?        8      0.00          51.64      0.00
51.64
##                    +        5      0.00          51.64      0.00
51.64
##                    0      216      0.02          51.66      0.02
51.66
##                    1       25      0.00          51.67      0.00
51.67
##                    2       13      0.00          51.67      0.00
51.67
##                    3        4      0.00          51.67      0.00
51.67
##                    4        4      0.00          51.67      0.00
51.67
##                    5       28      0.00          51.67      0.00
51.67


##                    6        4      0.00          51.67      0.00
51.67
##                    7        5      0.00          51.67      0.00
51.67
```

```
##                     8        1      0.00          51.67      0.00
51.67
##                     B       40      0.00          51.68      0.00
51.68
##                     h        1      0.00          51.68      0.00
51.68
##                     H        6      0.00          51.68      0.00
51.68
```

```
##                     K   424665     47.06          98.74     47.06
98.74
##                     m        7      0.00          98.74      0.00
98.74
##                     M    11330      1.26         100.00      1.26
100.00
##                  <NA>        0                                0.00
100.00
##                 Total   902297    100.00         100.00    100.00
100.00
```

```
summary(stormdata$CROPDMG)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   1.527   0.000 990.000
```

```
freq(stormdata$CROPDMGEXP)
```

```
## Frequencies
## stormdata$CROPDMGEXP
## Type: Character
##
##                           Freq   % Valid   % Valid Cum.   % Total
% Total Cum.
## -------------------- -------- --------- -------------- --------- --
------------
##       (Empty string)   618413     68.54          68.54     68.54
68.54
##                     ?        7      0.00          68.54      0.00
68.54
##                     0       19      0.00          68.54      0.00
68.54
##                     2        1      0.00          68.54      0.00
68.54
##                     B        9      0.00          68.54      0.00
68.54
##                     k       21      0.00          68.54      0.00
68.54
```

```
##                      K    281832      31.23           99.78      31.23
99.78
##                      m         1       0.00           99.78       0.00
99.78
##                      M      1994       0.22          100.00       0.22
100.00
##                   <NA>         0                                   0.00
100.00
```

```
##                  Total    902297     100.00          100.00     100.00
100.00
```

##Data Processing

```
stormpro <- select(stormdata, EVTYPE, FATALITIES, INJURIES, PROPDMG, P
ROPDMGEXP, CROPDMG, CROPDMGEXP)
```

#now we can transform

```
stormpro$EVTYPE <- as.character(stormpro$EVTYPE)
stormpro$EVTYPE <- toupper(stormpro$EVTYPE)
stormpro$PROPDMGEXP <- toupper(stormpro$PROPDMGEXP)
stormpro <- mutate(stormpro, propdmgexp2 = recode(PROPDMGEXP,
                                                  "1" = 10,
                                                  "2" = 100,
                                                  "3" = 1000,
                                                  "4" = 10000,
                                                  "5" = 100000,
                                                  "6" = 1000000,
                                                  "7" = 10000000,
                                                  "8" = 100000000,
                                                  "H" = 100,
                                                  "K" = 1000,
                                                  "M" = 1000000,
                                                  "B" = 1000000000,
                                                  .default = 1))
stormpro <- mutate(stormpro, propdmgvalue = PROPDMG*propdmgexp2)
```

#calculate damage

```
stormpro$CROPDMGEXP <- toupper(stormpro$CROPDMGEXP)
```

```
stormpro <- mutate(stormpro, cropdmgexp2 = recode(CROPDMGEXP,
                                                  "1" = 10,
                                                  "2" = 100,
                                                  "3" = 1000,
                                                  "4" = 10000,
                                                  "5" = 100000,
                                                  "6" = 1000000,
                                                  "7" = 10000000,
                                                  "8" = 100000000,
                                                  "H" = 100,
                                                  "K" = 1000,
                                                  "M" = 1000000,
                                                  "B" = 1000000000,
                                                 .default = 1))
stormpro <- mutate(stormpro, cropdmgvalue = CROPDMG*cropdmgexp2)
```

#subset injuries info

```
injuries <- aggregate(INJURIES~EVTYPE, data = stormpro, FUN = sum)
injuries <- injuries[order(injuries$INJURIES, decreasing = TRUE), ]
injuries <- injuries[1:5, ]
```

#fatality data by event

```
fatalities <- aggregate(FATALITIES~EVTYPE, data = stormpro, FUN = sum)
fatalities <- fatalities[order(fatalities$FATALITIES, decreasing = TRU
E), ]
fatalities <- fatalities[1:5, ]
```

#Properly damage by type

```
propdamage <- aggregate(propdmgvalue~EVTYPE, data = stormpro, FUN = su
m)
propdamage <- propdamage[order(propdamage$propdmgvalue, decreasing = T
RUE), ]
propdamage <- propdamage[1:5, ]
cropdamage <- aggregate(cropdmgvalue~EVTYPE, data = stormpro, FUN = su
m)
cropdamage <- cropdamage[order(cropdamage$cropdmgvalue, decreasing = T
RUE), ]
cropdamage <- cropdamage[1:5, ]
```

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

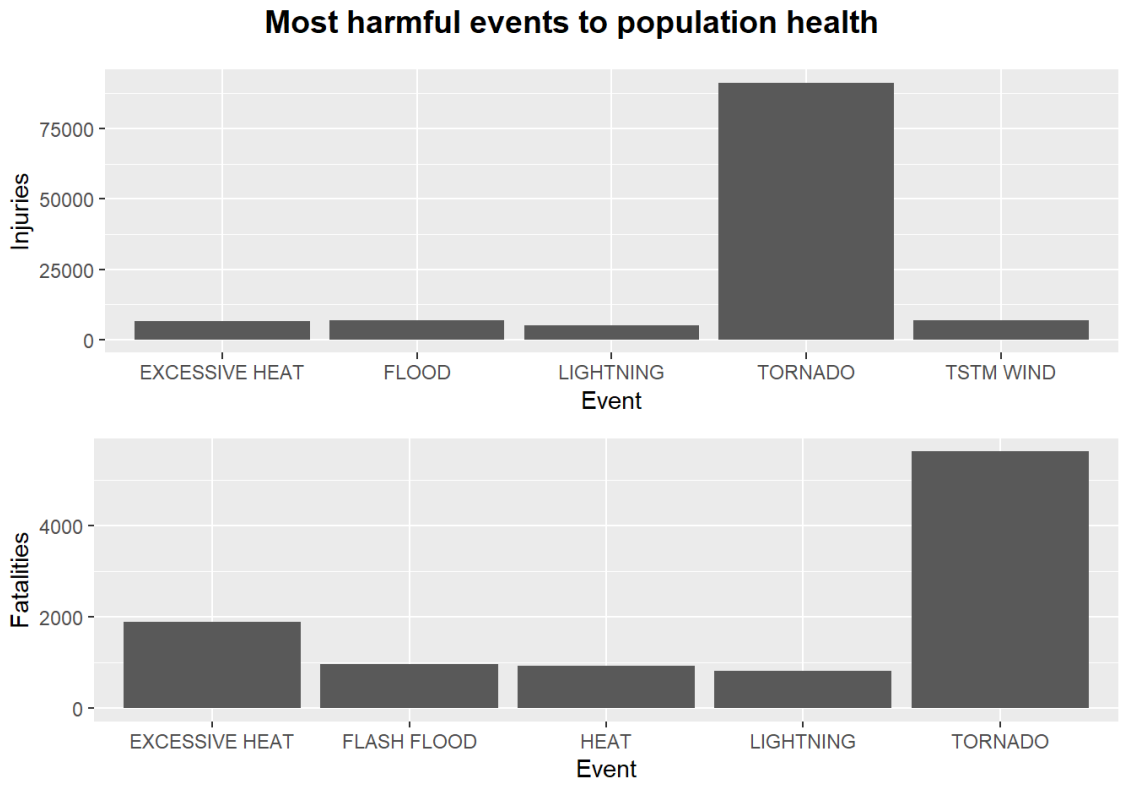
```
injgraph <- ggplot(data = injuries, aes(x = EVTYPE, y = INJURIES)) + l
abs(x="Event", y="Injuries") + geom_col()

fatgraph <- ggplot(data = fatalities, aes(x = EVTYPE, y = FATALITIES))
+ labs(x="Event", y="Fatalities") + geom_col()

title <- ggdraw() + draw_label("Most harmful events to population heal
th", fontface = 'bold') + theme(plot.margin = margin(0,0,0,7))

plot_col <- plot_grid(injgraph,fatgraph, ncol = 1)
```

#plot




```
plot_grid(title,plot_col, ncol = 1, rel_heights = c(0.1,1))
```

#results of question 1 #Tornados are the most harmful to population health, causing the most injuries and fatalities(33.2%).

#now go for question 2/ 2. Across the United States, which types of events have the greatest economic consequences?

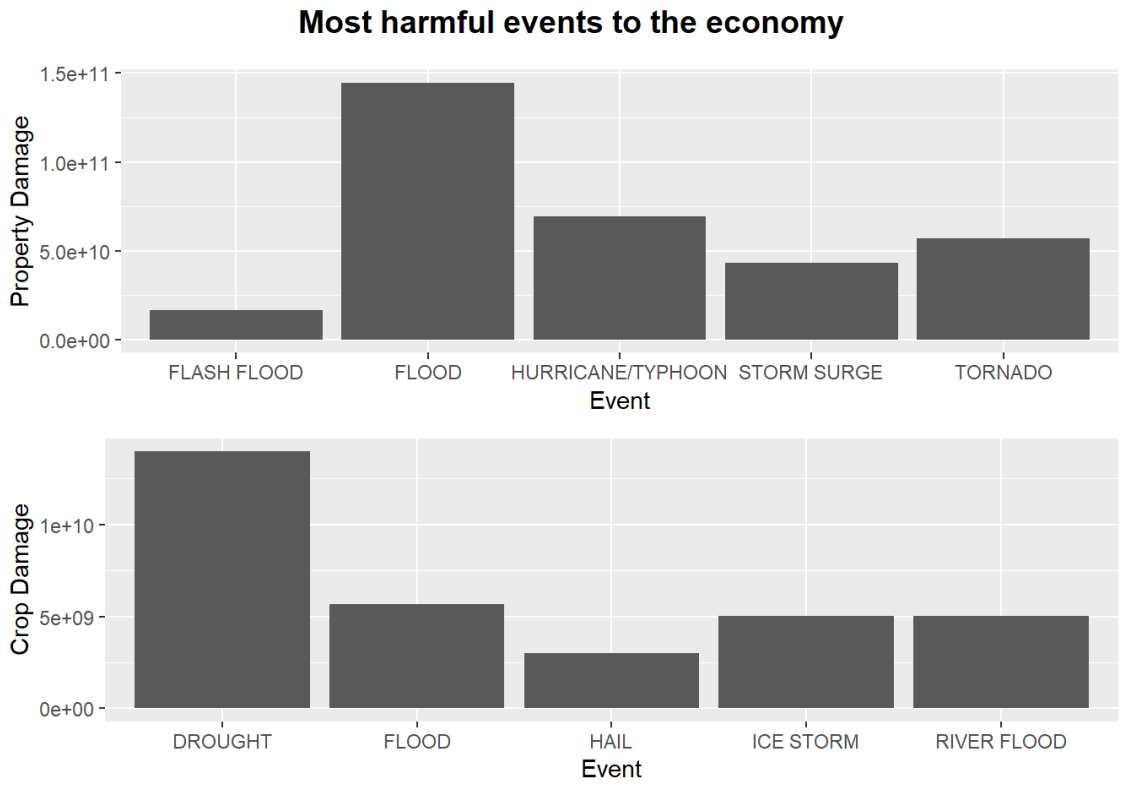
```
propdmg  <- ggplot(data = propdamage, aes(x = EVTYPE, y = propdmgvalue
)) + labs(x="Event", y="Property Damage") + geom_col()
```

```
cropdmg  <- ggplot(data = cropdamage, aes(x = EVTYPE, y = cropdmgvalue
)) + labs(x="Event", y="Crop Damage") + geom_col()

title2 <- ggdraw() + draw_label("Most harmful events to the economy",
fontface = 'bold') + theme(plot.margin = margin(0,0,0,7))

plot_col2 <- plot_grid(propdmg,cropdmg, ncol = 1)

plot_grid(title2,plot_col2, ncol = 1, rel_heights = c(0.1,1))
```



#results of question 2

#Floods cause the most property damage while droughts cause the most crop damage (37.1%)