

# Assignment2

Di Y.

9/2/2020

## Severe weather events and their consequences

### Synopsis

In the USA, tornadoes have killed and hurt the largest number of American people. From 1950 to 2011, 5633 people have lost their lives due to tornadoes and more than 91000 people got injured. Moreover, tornadoes also caused the largest damage of properties. In the same time period, the loss of property was estimated as \$3.2 trillion, while hail caused estimatedly total crop damage of \$58 billion.

### 1. Data Reading and Processing

```
file <- "repdata_data_StormData.csv.bz2"
if (!file.exists(file)) {
  download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", file, method="curl")
  unzip("repdata_data_StormData.csv.bz2", exdir="data")
}
```

```
raw <- read.csv(file, header=T, sep=",")
dim(raw)
```

```
## [1] 902297 37
```

```
str(raw)
```

```
## 'data.frame': 902297 obs. of 37 variables:
## $ STATE__ : num 1 1 1 1 1 1 1 1 1 1 ...
## $ BGN_DATE : Factor w/ 16335 levels "1/1/1966 0:00:00",...: 6523 6523 4242 11116 2224 2224 2260 383 3980 3980 ...
## $ BGN_TIME : Factor w/ 3608 levels "00:00:00 AM",...: 272 287 2705 1683 2584 3186 242 1683 3186 3186 ...
## $ TIME_ZONE : Factor w/ 22 levels "ADT","AKS","AST",...: 7 7 7 7 7 7 7 7 7 7 ...
## $ COUNTY : num 97 3 57 89 43 77 9 123 125 57 ...
## $ COUNTYNAME: Factor w/ 29601 levels "", "5NM E OF MACKINAC BRIDGE TO PRESQUE ISLE LT MI",...: 13513 1873 4598 10592 4372 10094 197
3 23873 24418 4598 ...
## $ STATE : Factor w/ 72 levels "AK","AL","AM",...: 2 2 2 2 2 2 2 2 2 2 ...
## $ EVTYPE : Factor w/ 985 levels " HIGH SURF ADVISORY",...: 834 834 834 834 834 834 834 834 834 834 ...
## $ BGN_RANGE : num 0 0 0 0 0 0 0 0 0 0 ...
## $ BGN_AZI : Factor w/ 35 levels "", " N", " NW",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ BGN_LOCATI: Factor w/ 54429 levels "", " Christiansburg",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ END_DATE : Factor w/ 6663 levels "", "1/1/1993 0:00:00",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ END_TIME : Factor w/ 3647 levels "", " 0900CST",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ 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 : Factor w/ 24 levels "", "E", "ENE", "ESE",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ END_LOCATI: Factor w/ 34506 levels "", " CANTON", " TULIA",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ 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: Factor w/ 19 levels "", "-", "?", "+",...: 17 17 17 17 17 17 17 17 17 17 ...
## $ CROPDMG : num 0 0 0 0 0 0 0 0 0 0 ...
## $ CROPDMGEXP: Factor w/ 9 levels "", "?", "0", "2",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ WFO : Factor w/ 542 levels "", " CI", "%SD",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ STATEOFFIC: Factor w/ 250 levels "", "ALABAMA, Central",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ ZONENAMES : Factor w/ 25112 levels "", "
1 1 1 1 1 ...
## $ 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 : Factor w/ 436781 levels "", "\t", "\t\t",...: 1 1 1 1 1 1 1 1 1 1 ...
## $ REFNUM : num 1 2 3 4 5 6 7 8 9 10 ...
```

```
"|__truncated__,...: 1 1 1 1 1
```

# Let the unit of the damage value in columns "PROPDMGEXP" and "CROPDMGEXP" just be \$  
summary(raw)

```
## STATE__ BGN_DATE BGN_TIME
## Min. :1.0 5/25/2011 0:00:00: 1202 12:00:00 AM: 10163
## 1st Qu.:19.0 4/27/2011 0:00:00: 1193 06:00:00 PM: 7350
## Median :30.0 6/9/2011 0:00:00: 1030 04:00:00 PM: 7261
## Mean :31.2 5/30/2004 0:00:00: 1016 05:00:00 PM: 6891
## 3rd Qu.:45.0 4/4/2011 0:00:00: 1009 12:00:00 PM: 6703
## Max. :95.0 4/2/2006 0:00:00: 981 03:00:00 PM: 6700
## (Other) :895866 (Other) :857229
## TIME_ZONE COUNTY COUNTYNAME STATE
## CST :547493 Min. : 0.0 JEFFERSON : 7840 TX :83728
## EST :245558 1st Qu.:31.0 WASHINGTON: 7603 KS :53440
## MST :68390 Median :75.0 JACKSON : 6660 OK :46802
## PST :28302 Mean :100.6 FRANKLIN : 6256 MO :35648
## AST : 6360 3rd Qu.:131.0 LINCOLN : 5937 IA :31069
## HST : 2563 Max. :873.0 MADISON : 5632 NE :30271
## (Other): 3631 (Other) :862369 (Other):621339
## EVTYPE BGN_RANGE BGN_AZI
## HAIL :288661 Min. : 0.000 :547332
## TSTM WIND :219940 1st Qu.: 0.000 N :86752
## THUNDERSTORM WIND:82563 Median : 0.000 W :38446
## TORNADO :60652 Mean : 1.484 S :37558
## FLASH FLOOD :54277 3rd Qu.: 1.000 E :33178
## FLOOD :25326 Max. :3749.000 NW :24041
## (Other) :170878 (Other):134990
## BGN_LOCATI END_DATE END_TIME
## :287743 :243411 :238978
## COUNTYWIDE :19680 4/27/2011 0:00:00: 1214 06:00:00 PM: 9802
## Countywide : 993 5/25/2011 0:00:00: 1196 05:00:00 PM: 8314
## SPRINGFIELD : 843 6/9/2011 0:00:00: 1021 04:00:00 PM: 8104
## SOUTH PORTION: 810 4/4/2011 0:00:00: 1007 12:00:00 PM: 7483
## NORTH PORTION: 784 5/30/2004 0:00:00: 998 11:59:00 PM: 7184
## (Other) :591444 (Other) :653450 (Other) :622432
## COUNTY_END COUNTYENDN END_RANGE END_AZI
## Min. :0 Mode:logical Min. : 0.0000 :724837
## 1st Qu.:0 NA's:902297 1st Qu.: 0.0000 N :28082
## Median :0 Median : 0.0000 S :22510
## Mean :0 Mean : 0.9862 W :20119
## 3rd Qu.:0 3rd Qu.: 0.0000 E :20047
## Max. :0 Max. :925.0000 NE :14606
## (Other): 72096
## END_LOCATI LENGTH WIDTH
## :499225 Min. : 0.0000 Min. : 0.000
## COUNTYWIDE :19731 1st Qu.: 0.0000 1st Qu.: 0.000
## SOUTH PORTION : 833 Median : 0.0000 Median : 0.000
## NORTH PORTION : 780 Mean : 0.2301 Mean : 7.503
## CENTRAL PORTION: 617 3rd Qu.: 0.0000 3rd Qu.: 0.000
## SPRINGFIELD : 575 Max. :2315.0000 Max. :4400.000
## (Other) :380536
## F MAG FATALITIES INJURIES
## Min. :0.0 Min. : 0.0 Min. : 0.0000 Min. : 0.0000
## 1st Qu.:0.0 1st Qu.: 0.0 1st Qu.: 0.0000 1st Qu.: 0.0000
## Median :1.0 Median : 50.0 Median : 0.0000 Median : 0.0000
## Mean :0.9 Mean : 46.9 Mean : 0.0168 Mean : 0.1557
## 3rd Qu.:1.0 3rd Qu.: 75.0 3rd Qu.: 0.0000 3rd Qu.: 0.0000
## Max. :5.0 Max. :22000.0 Max. :583.0000 Max. :1700.0000
## NA's :843563
## PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP
## Min. : 0.00 :465934 Min. : 0.000 :618413
## 1st Qu.: 0.00 K :424665 1st Qu.: 0.000 K :281832
## Median : 0.00 M :11330 Median : 0.000 M :1994
## Mean :12.06 0 :216 Mean :1.527 k :21
## 3rd Qu.: 0.50 B :40 3rd Qu.: 0.000 0 :19
## Max. :5000.00 5 :28 Max. :990.000 B :9
## (Other): 84 (Other): 9
## WFO STATEOFFIC
## :142069 :248769
## OUN :17393 TEXAS, North :12193
## JAN :13889 ARKANSAS, Central and North Central:11738
## LWX :13174 IOWA, Central :11345
## PHI :12551 KANSAS, Southwest :11212
## TSA :12483 GEORGIA, North and Central :11120
## (Other):690738 (Other) :595920
##
```

ZONENAMES

```
##
##
## GREATER RENO / CARSON CITY / M - GREATER RENO / CARSON CITY / M
: 639
## GREATER LAKE TAHOE AREA - GREATER LAKE TAHOE AREA
: 592
## JEFFERSON - JEFFERSON
: 303
## MADISON - MADISON
302
## (Other)
## LATITUDE LONGITUDE LATITUDE_E LONGITUDE_
## Min. : 0 Min. : -14451 Min. : 0 Min. : -14455
## 1st Qu.:2802 1st Qu.: 7247 1st Qu.: 0 1st Qu.: 0
## Median :3540 Median : 8707 Median : 0 Median : 0
## Mean :2875 Mean : 6940 Mean :1452 Mean : 3509
## 3rd Qu.:4019 3rd Qu.: 9605 3rd Qu.:3549 3rd Qu.: 8735
## Max. :9706 Max. :17124 Max. :9706 Max. :106220
## NA's :47 NA's :40
## REMARKS REFNUM
## :287433 Min. : 1
## :24013 1st Qu.:225575
## Trees down.\n : 1110 Median :451149
## Several trees were blown down.\n : 568 Mean :451149
## Trees were downed.\n : 446 3rd Qu.:676723
## Large trees and power lines were blown down.\n: 432 Max. :902297
## (Other) :588295
```

```
table(raw$PROPDMGEXP)
```

```
##
## - ? + 0 1 2 3 4 5 6
## 465934 1 8 5 216 25 13 4 4 28 4
## 7 8 B h H K m M
## 5 1 40 1 6 424665 7 11330
```

```
table(raw$CROPDMGEXP)
```

```
##
## ? 0 2 B k K m M
## 618413 7 19 1 9 21 281832 1 1994
```

```
unit_fun <- function(unit){
  if (unit=="") {
    return(1)
  } else if(unit %in% c("?", "-", "+")){
    return(NA)
  } else if(as.numeric(unit) %in% c(1:10)){
    return(as.numeric(unit))
  } else if(unit %in% c("k", "K")){
    return(10^3)
  } else if(unit %in% c("m", "M")){
    return(10^6)
  } else if(unit %in% c("h", "H")){
    return(10^2)
  } else if(unit %in% c("B", "b")){
    return(10^12)
  }
}

raw$PROPDMG <- raw$PROPDMG*unit_fun(raw$PROPDMGEXP)
```

```
## Warning in if (unit == "") {: Bedingung hat Länge > 1 und nur das erste Element
## wird benutzt
```

```
## Warning in if (unit %in% c("?", "-", "+")) {: Bedingung hat Länge > 1 und nur
## das erste Element wird benutzt
```

```
## Warning in if (as.numeric(unit) %in% c(1:10)) {: Bedingung hat Länge > 1 und nur
## das erste Element wird benutzt
```

```
## Warning in if (unit %in% c("k", "K")) {: Bedingung hat Länge > 1 und nur das
## erste Element wird benutzt
```

```
raw$CROPDMG <- raw$CROPDMG*unit_fun(raw$CROPDMGEXP)
```

```
## Warning in if (unit == "") {: Bedingung hat Länge > 1 und nur das erste Element
## wird benutzt
```

```
raw$PROPDGMGEXP <- c(1)
raw$CROPDMGEXP <- c(1)
head(raw)
```

```
## STATE__      BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAMES STATE  EVTYPE
## 1      1 4/18/1950 0:00:00 0130   CST   97  MOBILE   AL TORNADO
## 2      1 4/18/1950 0:00:00 0145   CST    3  BALDWIN  AL TORNADO
## 3      1 2/20/1951 0:00:00 1600   CST   57  FAYETTE  AL TORNADO
## 4      1 6/8/1951 0:00:00 0900   CST   89  MADISON  AL TORNADO
## 5      1 11/15/1951 0:00:00 1500   CST   43  CULLMAN  AL TORNADO
## 6      1 11/15/1951 0:00:00 2000   CST   77  LAUDERDALE  AL TORNADO
## BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END COUNTYENDN
## 1      0              0      NA
## 2      0              0      NA
## 3      0              0      NA
## 4      0              0      NA
## 5      0              0      NA
## 6      0              0      NA
## END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES INJURIES PROPDMG
## 1      0          14.0 100 3 0      0 15 25000
## 2      0           2.0 150 2 0      0 0 2500
## 3      0           0.1 123 2 0      0 2 25000
## 4      0           0.0 100 2 0      0 2 2500
## 5      0           0.0 150 2 0      0 2 2500
## 6      0           1.5 177 2 0      0 6 2500
## PROPDGMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1      1      0      1          3040 8812
## 2      1      0      1          3042 8755
## 3      1      0      1          3340 8742
## 4      1      0      1          3458 8626
## 5      1      0      1          3412 8642
## 6      1      0      1          3450 8748
## LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1      3051      8806      1
## 2      0      0      2
## 3      0      0      3
## 4      0      0      4
## 5      0      0      5
## 6      0      0      6
```

```
dim(raw)
```

```
## [1] 902297 37
```

## Results

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

```
library(dplyr)
injury <- raw %>% group_by(EVTYPE) %>%
  summarise(sum_injury=sum(INJURIES,na.rm=T)) %>%
  arrange(desc(sum_injury)) %>%
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# Top 10 severe weathers, which brought the most injuries
injury_top10 <- injury %>% top_n(10)
```

```
## Selecting by sum_injury
```

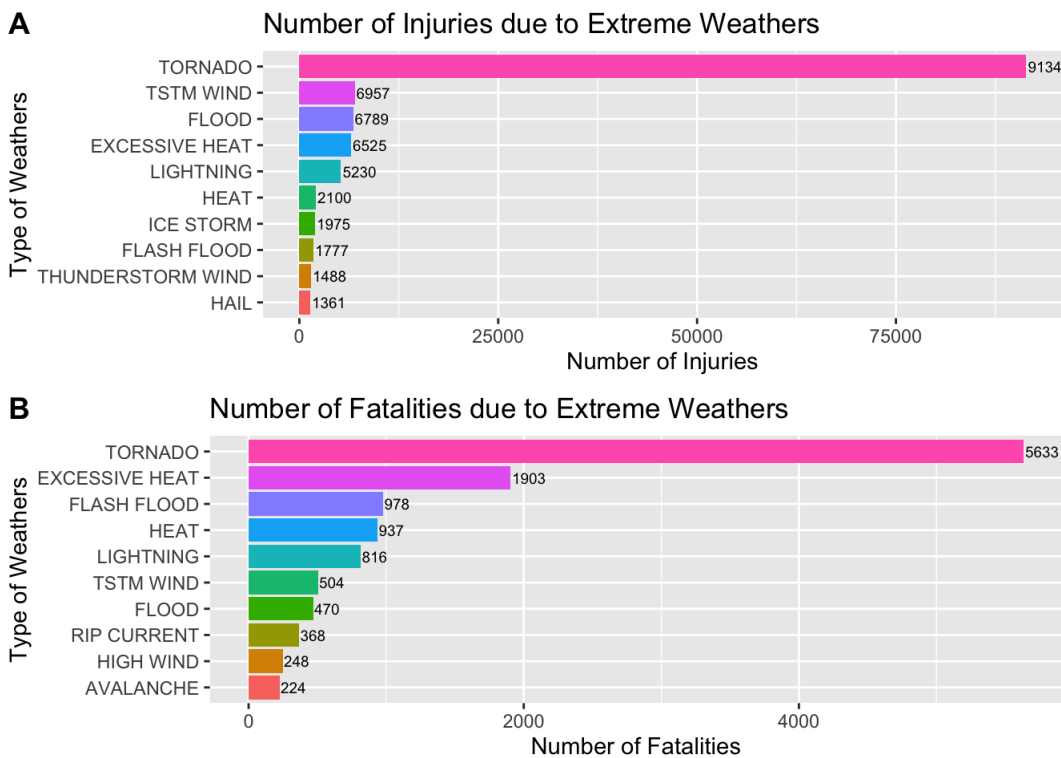
```
fatality <- raw %>% group_by(EVTYPE) %>%  
  summarise(sum_fatality=sum(FATALITIES,na.rm=T)) %>%  
  arrange(desc(sum_fatality)) %>%  
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# Top 10 severe weathers, which led to the most fatalities  
fatality_top10 <- fatality %>% top_n(10)
```

```
## Selecting by sum_fatality
```

```
##  
## The downloaded binary packages are in  
## /var/folders/5x/5xkzjtpx77l47xk83kvx7fl00000gn/T//RtmpaG1z7O/downloaded_packages
```



2) Across the United States, which types of events have the greatest economic consequences?

```
prop_damage <- raw %>% group_by(EVTYPE) %>%  
  summarise(damage_p=sum(PROPDMG,na.rm=T)) %>%  
  arrange(desc(damage_p)) %>%  
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
prop_damage_top10 <- prop_damage %>% top_n(10)
```

```
## Selecting by damage_p
```

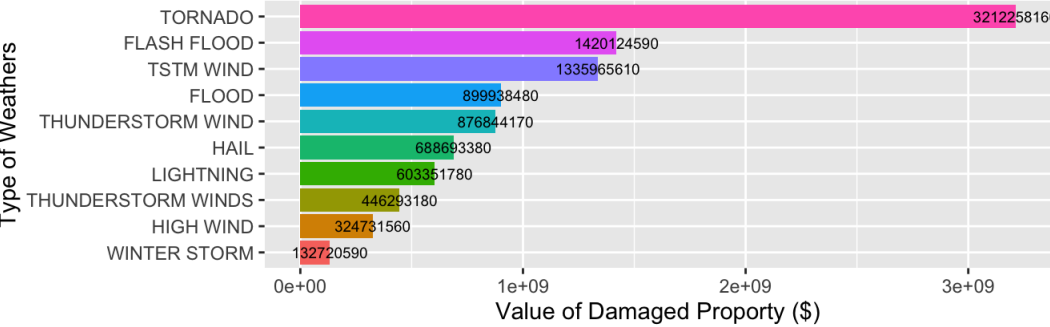
```
crop_damage <- raw %>% group_by(EVTYPE) %>%  
  summarise(damage_c=sum(CROPDMG,na.rm=T)) %>%  
  arrange(desc(damage_c)) %>%  
  ungroup()
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
crop_damage_top10 <- crop_damage %>% top_n(10)
```

C

Damage of Property due to Extreme Weathers



D

Damage of Crop Damage due to Extreme Weathers

