Reproducible Research project 2.Rmd

Jiangpeng Pan

6/15/2020

Title: Analysis of Weather Data

Synopsis:

In this project, the data frame and histogram were used to analyze the relationship between weather event type and faltality, injury, property damage

Data Processing

```
if (!file.exists("stormData.csv.bz2"))
    {
     url <- 'https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2
'
     download.file(url,destfile='StormData.csv.bz2',mode='wb')

    }
if (!exists("stormData")) {
  stormDataFile <- bzfile(description = "StormData.csv.bz2", open = "r")
  stormData <- read.csv(stormDataFile, fill = TRUE, header = TRUE)
  close(stormDataFile)
}
```

import necessary packages

```
library(plyr)
require(plyr)
library(ggplot2)
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
head(stormData)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME 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    25.0
## 2         0                       2.0   150 2   0          0        0     2.5
## 3         0                       0.1   123 2   0          0        2    25.0
## 4         0                       0.0   100 2   0          0        2     2.5
## 5         0                       0.0   150 2   0          0        2     2.5
## 6         0                       1.5   177 2   0          0        6     2.5
##   PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES LATITUDE LONGITUDE
## 1          K       0                                         3040      8812
## 2          K       0                                         3042      8755
## 3          K       0                                         3340      8742
## 4          K       0                                         3458      8626
## 5          K       0                                         3412      8642
## 6          K       0                                         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
```

```
summary(stormData)
```

```
##     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
##
:594029
##
:205988
##  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)
:100444
##     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
```

```
names(stormData)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

sort data: Injury with weather type, fatality with weather type, property damage with weather type

```
injury <- aggregate(INJURIES ~ EVTYPE, data = stormData, sum)

fatality <- aggregate(FATALITIES ~ EVTYPE, stormData, sum)

propdmg <- aggregate(PROPDMG ~ EVTYPE, stormData, sum)
```

list the top 10 of injury and fatality with related weather type

```
rankinjury <- arrange(injury, desc(injury[, 2]))
rankfatality <- arrange(fatality, desc(fatality[, 2]))
rankpropdmg <- arrange(propdmg, desc(propdmg[, 2]))
top10injury <- head(rankinjury, 10)
top10fatality <- head(rankfatality, 10)
top10propdmg <- head(rankpropdmg, 10)
```

the top 10 weather type injury is

```
top10injury
```

```
##               EVTYPE INJURIES
## 1            TORNADO    91346
## 2          TSTM WIND     6957
## 3              FLOOD     6789
## 4     EXCESSIVE HEAT     6525
## 5          LIGHTNING     5230
## 6               HEAT     2100
## 7          ICE STORM     1975
## 8        FLASH FLOOD     1777
## 9  THUNDERSTORM WIND     1488
## 10             HAIL     1361
```

the top 10 weather type fatality is

```
top10fatality
```

```
##            EVTYPE FATALITIES
## 1         TORNADO       5633
## 2  EXCESSIVE HEAT       1903
## 3     FLASH FLOOD        978
## 4            HEAT        937
## 5       LIGHTNING        816
## 6       TSTM WIND        504
## 7           FLOOD        470
## 8     RIP CURRENT        368
## 9       HIGH WIND        248
## 10      AVALANCHE        224
```

the top 10 weather type property damage is

```
top10propdmg
```

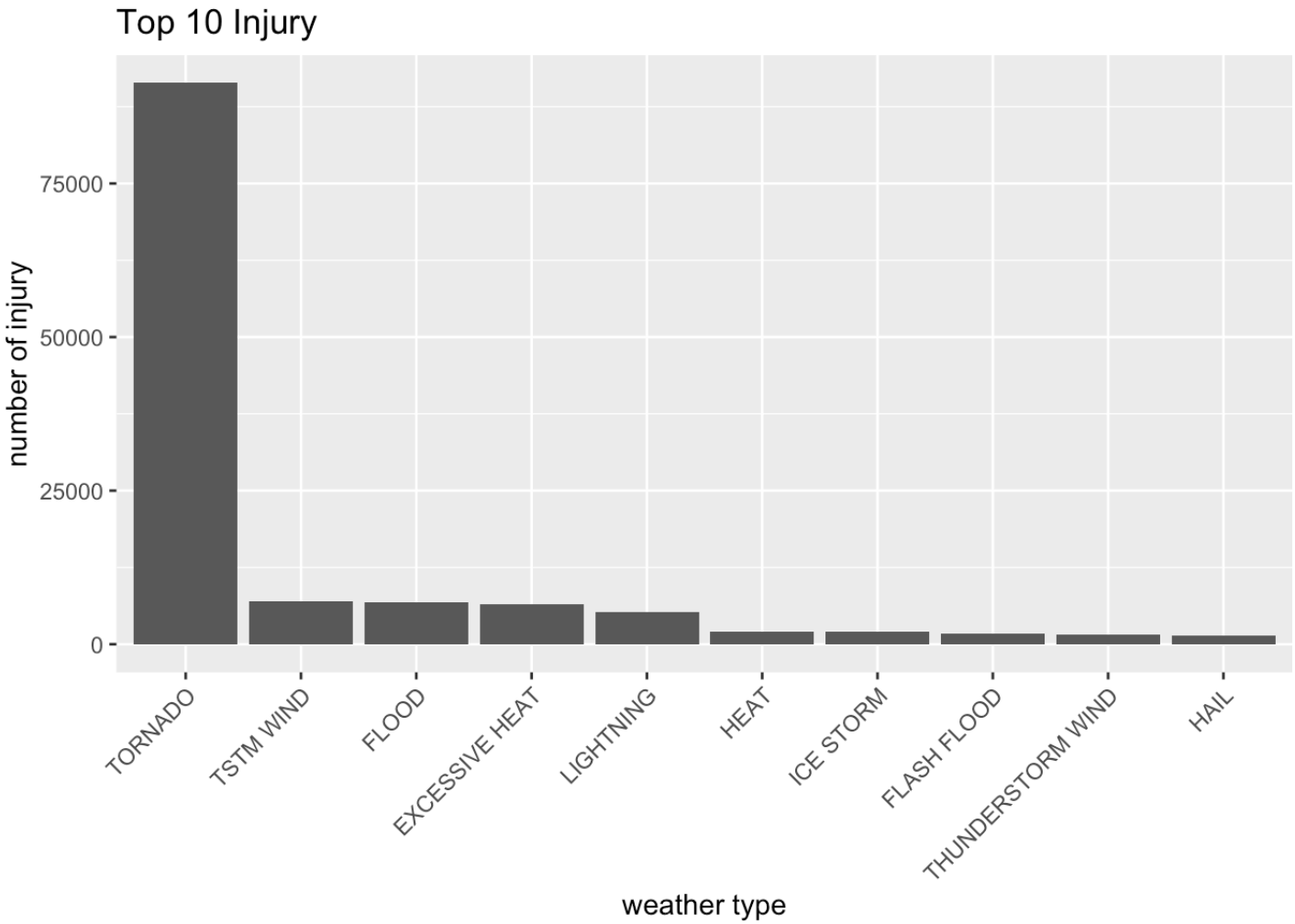
```
##                EVTYPE   PROPDMG
## 1             TORNADO 3212258.2
## 2         FLASH FLOOD 1420124.6
## 3           TSTM WIND 1335965.6
## 4               FLOOD  899938.5
## 5   THUNDERSTORM WIND  876844.2
## 6                HAIL  688693.4
## 7           LIGHTNING  603351.8
## 8  THUNDERSTORM WINDS  446293.2
## 9           HIGH WIND  324731.6
## 10       WINTER STORM  132720.6
```

Results

plot analysis

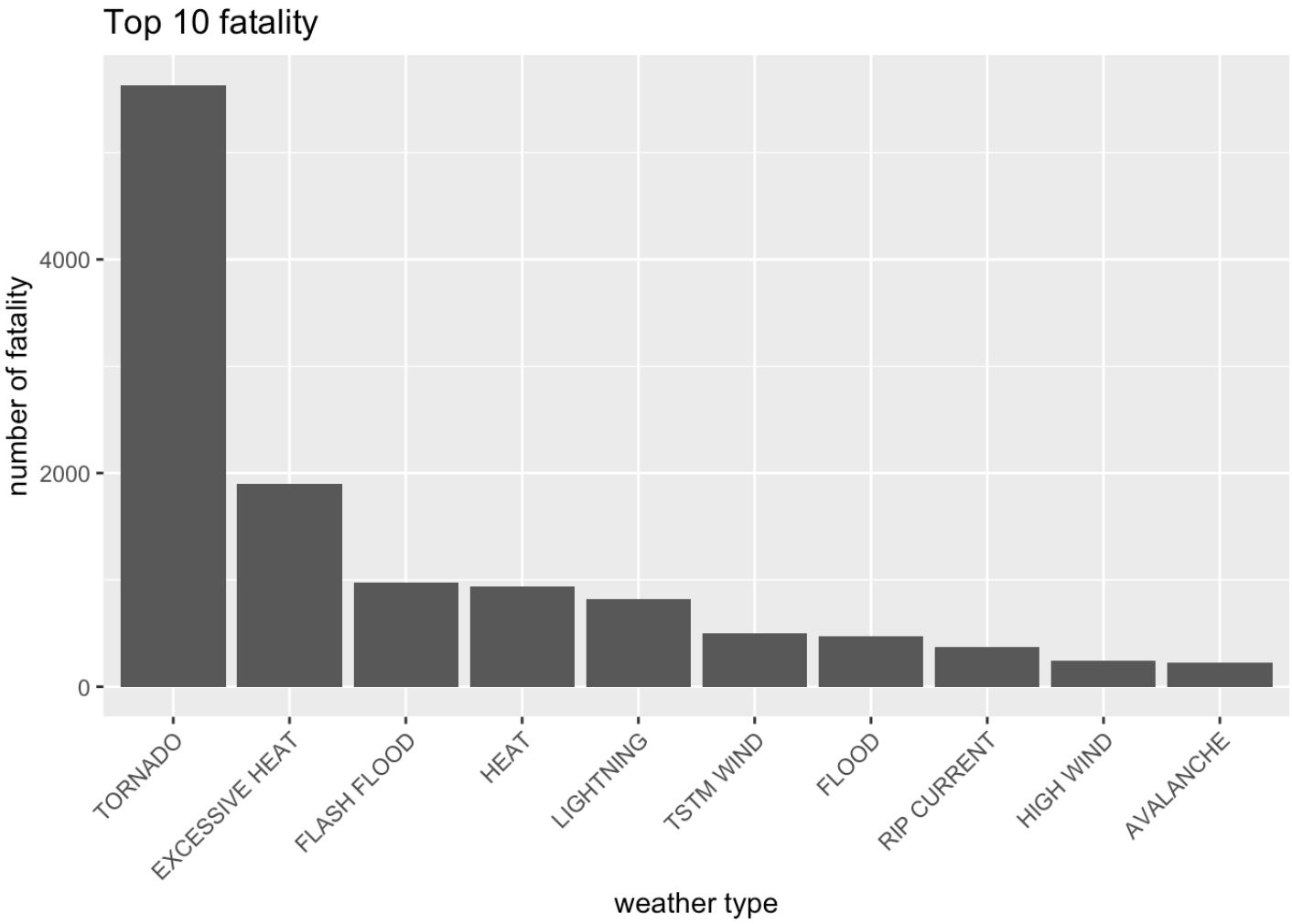
the histogram of weather type and injury number

```
g <- ggplot(top10injury, aes(x = reorder(EVTYPE, -INJURIES), y = INJURIES))
g + geom_bar(stat = "identity") + xlab("weather type") + ylab("number of injury") + g
gtitle("Top 10 Injury") + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



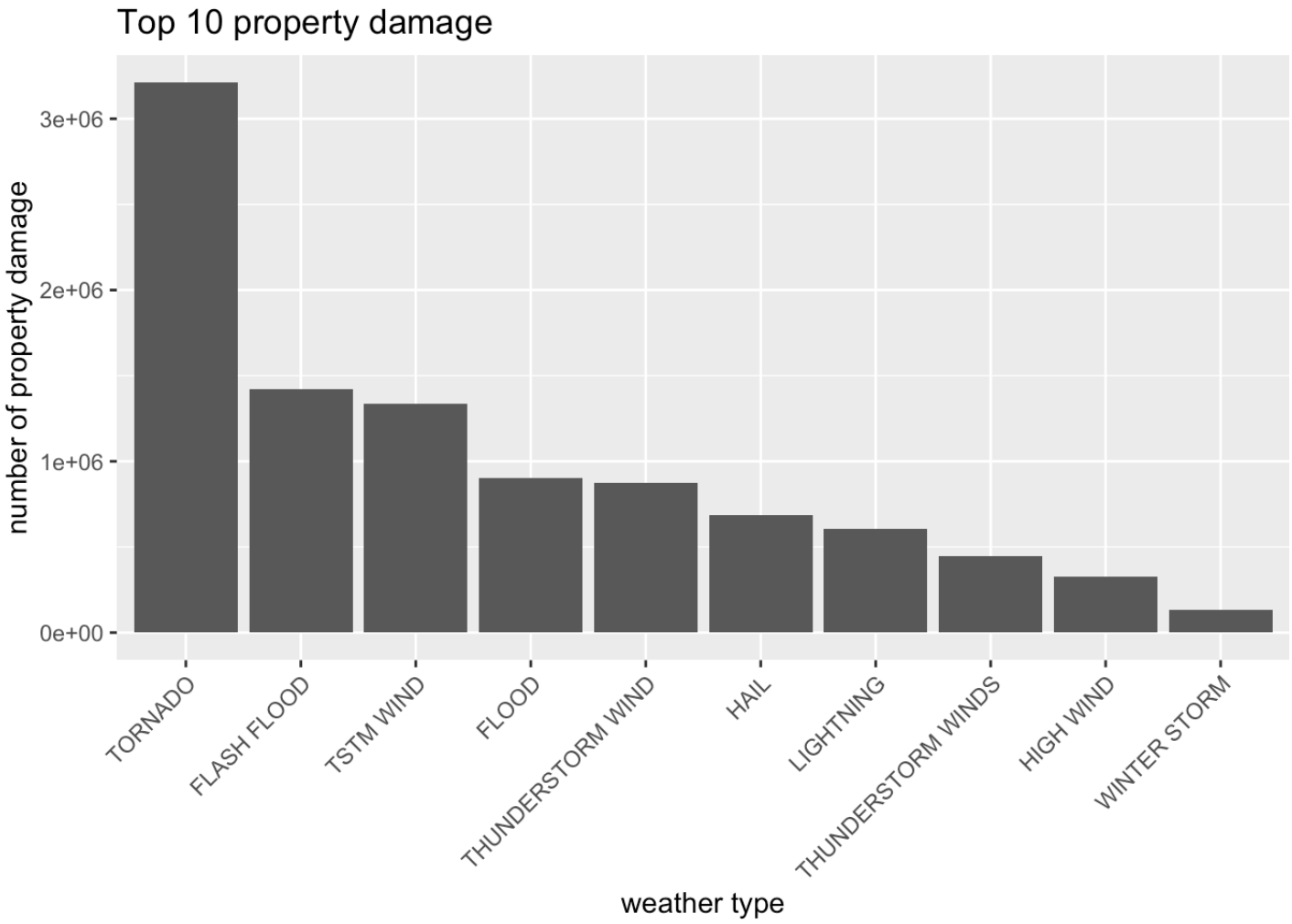
the histogram of weather type and fatality number

```
g <- ggplot(top10fatality, aes(x = reorder(EVTYPE, -FATALITIES), y = FATALITIES))
g + geom_bar(stat = "identity") + xlab("weather type") + ylab("number of fatality") +
ggtitle("Top 10 fatality") + theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



the histogram of weather type and property danage number

```
g <- ggplot(top10propdmg, aes(x = reorder(EVTYPE, -PROPDMG), y = PROPDMG))
g + geom_bar(stat = "identity") + xlab("weather type") + ylab("number of property dam
age") + ggtitle("Top 10 property damage") + theme(axis.text.x = element_text(angle =
45, hjust = 1))
```


Top 10 property damage
number of property damage
3e+06
2e+06
1e+06
0e+00
TORNADO
FLASH FLOOD
TSTM WIND
FLOOD
THUNDERSTORM WIND
HAIL
LIGHTNING
THUNDERSTORM WINDS
HIGH WIND
WINTER STORM
weather type