Exploring the NOAA Storm Database

R Markdown

```
## Introduction
# Severe storms and other weather events can cause significant health and economic impacts to communiti

# The dataset used for this project is the Storm Database of the National Oceanic and Atmospheric Admin

# The purpose of this project is to analyze the NOAA Storm Database to identify trends and patterns in 

# The analysis will focus on the following questions:
# - What are the most common types of severe weather events in the United States?
# - Where do severe weather events occur most frequently?
# - What are the most significant health and economic impacts of severe weather events?
# - The analysis will use a variety of statistical methods to answer these questions.

# The results of the analysis will be presented in a report that includes a discussion of the findings,
# ---------------------------------------------------------------------------------------------------
```

how the data were loaded into R and processed for analysis

```
#dataset<-download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2","E:/

dataset <- read.csv(bzfile("E:/Data Science Foundations using R/5 Reproducible Research/Woche 4/Projekt
```

some content about the data

```
head(dataset)
```

```
##   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
```

```r
str(dataset)
```

```
## '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" "2/20/1951 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 ...
```

```
names(dataset)
```

```
##  [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"
```

Injuries sorting, top 10

```
total_injuries <- aggregate(INJURIES~EVTYPE, dataset, sum)
total_injuries <- arrange(total_injuries, desc(INJURIES))
total_injuries <- total_injuries[1:20, ]
total_injuries
```

```
##                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
## 11       WINTER STORM     1321
## 12  HURRICANE/TYPHOON     1275
## 13          HIGH WIND     1137
## 14         HEAVY SNOW     1021
## 15           WILDFIRE      911
## 16 THUNDERSTORM WINDS      908
## 17           BLIZZARD      805
## 18                FOG      734
## 19   WILD/FOREST FIRE      545
## 20         DUST STORM      440
```

Total fatalities

```r
total_fatalities <- aggregate(FATALITIES~EVTYPE,dataset, sum)
total_fatalities <- arrange(total_fatalities, desc(FATALITIES))
total_fatalities <- total_fatalities[1:20, ]
total_fatalities
```

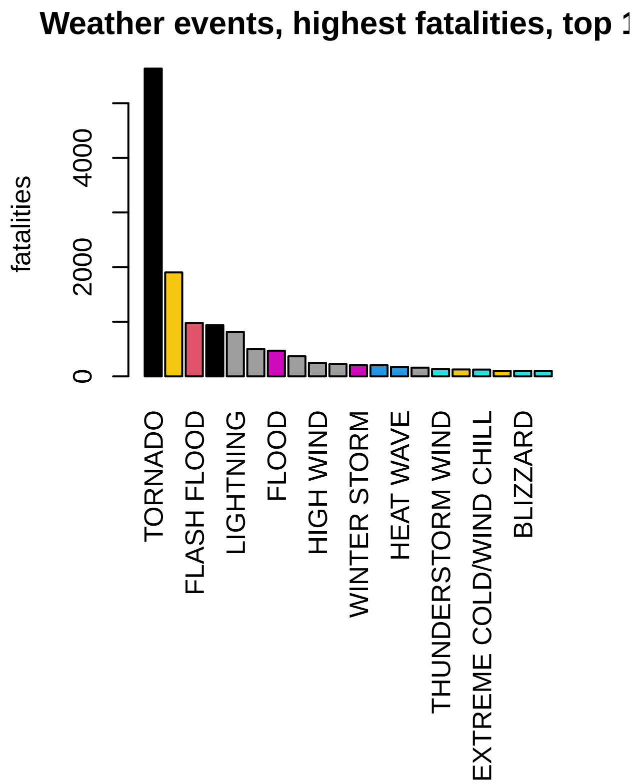
```
##                    EVTYPE FATALITIES
## 1                 TORNADO       5633
## 2          EXCESSIVE HEAT       1903
## 3             FLASH FLOOD        978
## 4                    HEAT        937
## 5               LIGHTNING        816
## 6               TSTM WIND        504
## 7                   FLOOD        470
## 8             RIP CURRENT        368
## 9               HIGH WIND        248
## 10              AVALANCHE        224
## 11           WINTER STORM        206
## 12           RIP CURRENTS        204
## 13              HEAT WAVE        172
## 14           EXTREME COLD        160
## 15      THUNDERSTORM WIND        133
## 16             HEAVY SNOW        127
## 17 EXTREME COLD/WIND CHILL        125
## 18            STRONG WIND        103
## 19               BLIZZARD        101
## 20              HIGH SURF        101
```

```r
str(total_fatalities)
```

```
## 'data.frame':    20 obs. of  2 variables:
##  $ EVTYPE    : chr  "TORNADO" "EXCESSIVE HEAT" "FLASH FLOOD" "HEAT" ...
##  $ FATALITIES: num  5633 1903 978 937 816 ...
```

Results: Weather events, highest fatalities, top 10

```r
totals <- total_fatalities
par(mfrow = c(1, 2), mar = c(15, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)
barplot(totals$FATALITIES, las = 3, names.arg = totals$EVTYPE, main = "Weather events, highest fataliti
```



Sad stuff - top 5

```
totals<- merge(total_fatalities, total_injuries, by.x = "EVTYPE", by.y = "EVTYPE")
totals<-arrange(totals,desc(FATALITIES+INJURIES))

sad_stuff <- melt(totals, id.vars="EVTYPE", variable.name = "bad_thing")
tail(sad_stuff, 5)
```

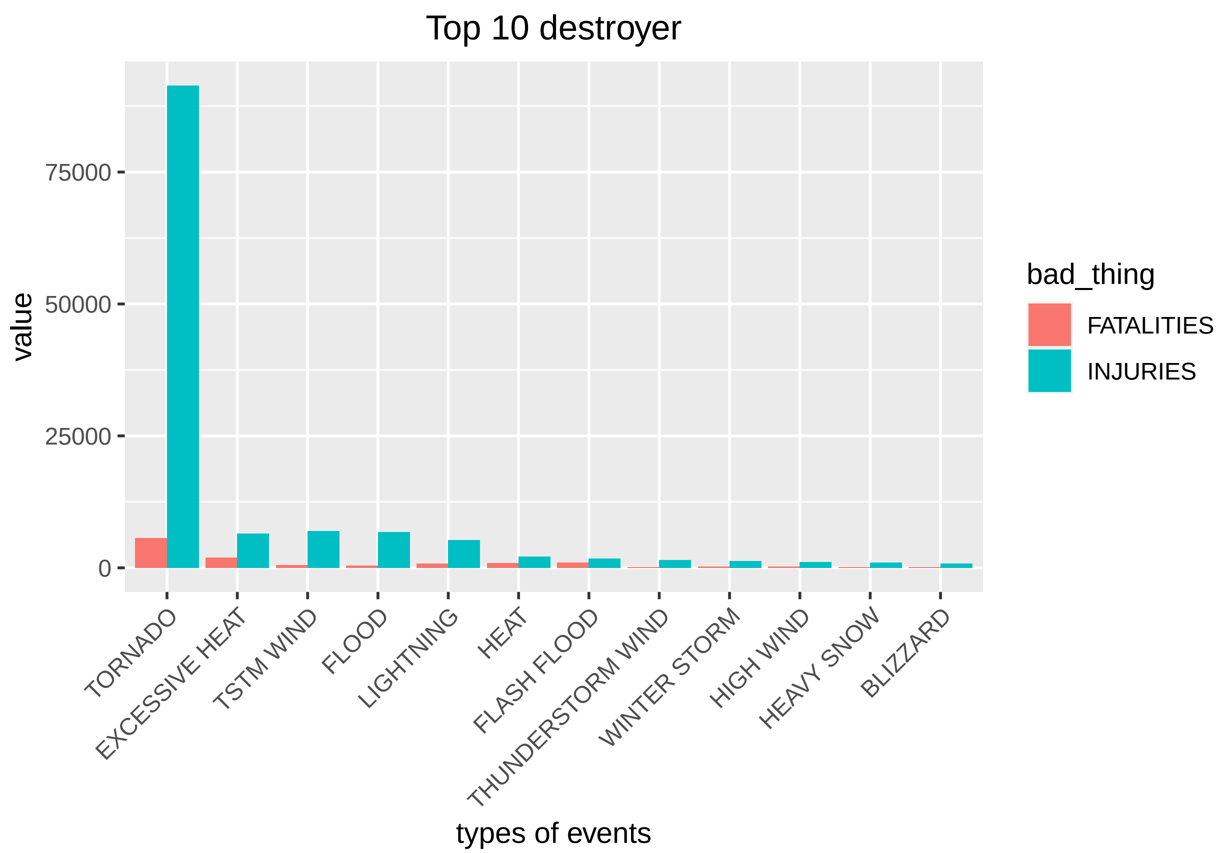
```
##                EVTYPE bad_thing value
## 20 THUNDERSTORM WIND  INJURIES  1488
## 21      WINTER STORM  INJURIES  1321
## 22         HIGH WIND  INJURIES  1137
## 23        HEAVY SNOW  INJURIES  1021
## 24          BLIZZARD  INJURIES   805
```

Results: Top 10 destroyer

```
healthChart <- ggplot(sad_stuff, aes(x=reorder(EVTYPE, -value), y=value))

healthChart = healthChart + geom_bar(stat="identity", aes(fill=bad_thing), position="dodge")
```

```
healthChart = healthChart + xlab("types of events")
healthChart = healthChart + theme(axis.text.x = element_text(angle=45, hjust=1))
healthChart = healthChart + ggtitle("Top 10 destroyer") + theme(plot.title = element_text(hjust = 0.5))
healthChart
```



Event type - sorting PROPDMG

```
propdmg <- aggregate(PROPDMG ~ EVTYPE, data = dataset, FUN = sum)
propdmg <- propdmg[order(propdmg$PROPDMG, decreasing = TRUE), ]
propdmgMax <- propdmg[1:10, ]
print(propdmgMax)
```

```
##                EVTYPE   PROPDMG
## 834           TORNADO 3212258.2
## 153       FLASH FLOOD 1420124.6
## 856         TSTM WIND 1335965.6
## 170             FLOOD  899938.5
## 760 THUNDERSTORM WIND  876844.2
## 244              HAIL  688693.4
## 464         LIGHTNING  603351.8
## 786 THUNDERSTORM WINDS  446293.2
## 359         HIGH WIND  324731.6
## 972      WINTER STORM  132720.6
```

Event type - sorting CROPDMG

```r
cropdmg <- aggregate(CROPDMG ~ EVTYPE, data = dataset, FUN = sum)
cropdmg <- cropdmg[order(cropdmg$CROPDMG, decreasing = TRUE), ]
cropdmgMax <- cropdmg[1:10, ]
print(cropdmgMax)
```

```
##                  EVTYPE   CROPDMG
## 244                HAIL 579596.28
## 153         FLASH FLOOD 179200.46
## 170               FLOOD 168037.88
## 856           TSTM WIND 109202.60
## 834             TORNADO 100018.52
## 760   THUNDERSTORM WIND  66791.45
## 95              DROUGHT  33898.62
## 786  THUNDERSTORM WINDS  18684.93
## 359           HIGH WIND  17283.21
## 290          HEAVY RAIN  11122.80
```

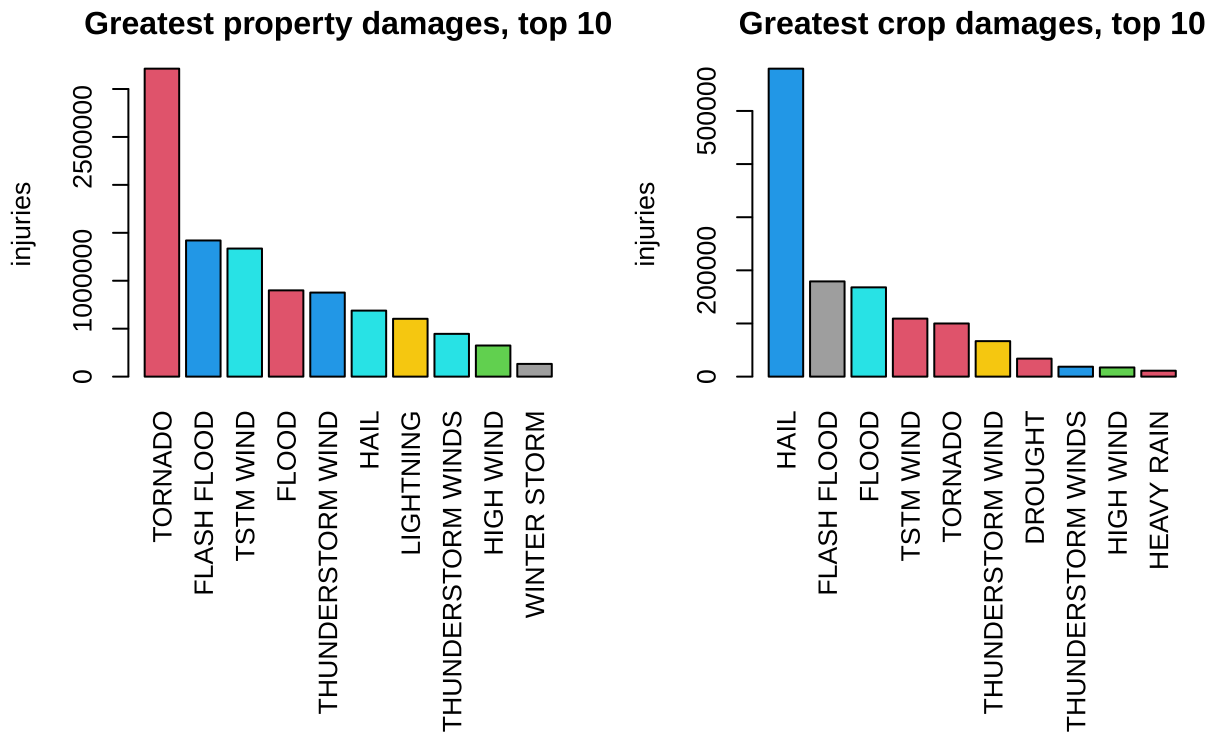
Greatest property damages & greatest crop damages, top10

```r
par(mfrow = c(1, 2), mar = c(15, 4, 3, 2), mgp = c(3, 1, 0), cex = 0.8)

options(scipen = 999)

barplot(propdmgMax$PROPDMG, las = 3, names.arg = propdmgMax$EVTYPE,
        main = "Greatest property damages, top 10",
        ylab = "injuries", col = propdmgMax$PROPDMG)

barplot(cropdmgMax$CROPDMG, las = 3, names.arg = cropdmgMax$EVTYPE,
        main = "Greatest crop damages, top 10",
        ylab = "injuries", col = cropdmgMax$CROPDMG)
```



Top 5 Event-Type - Damage Type ="PROPDMG"

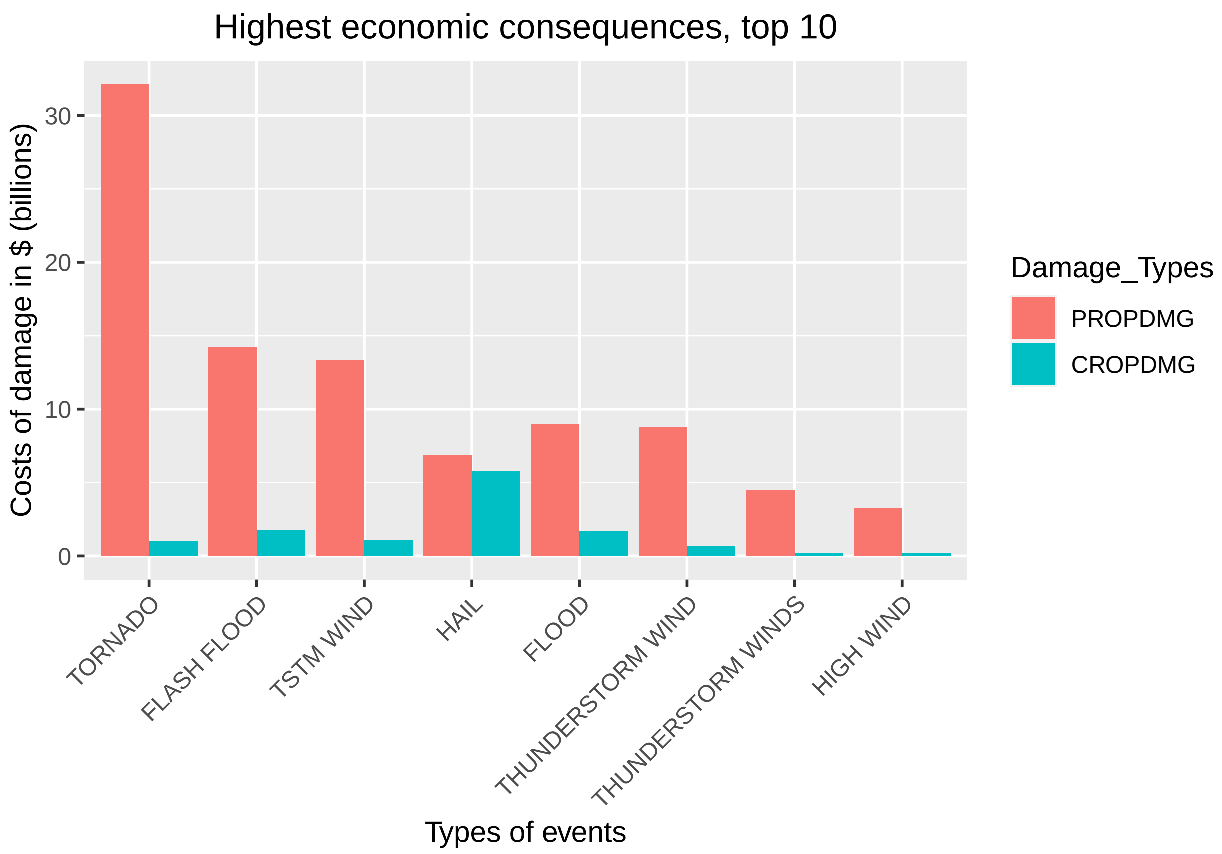
```
totalDamage<- merge(propdmgMax,cropdmgMax,by.x = "EVTYPE", by.y = "EVTYPE")
totalDamage<-arrange(totalDamage,desc(PROPDMG + CROPDMG))
top10damages <- melt(totalDamage, id.vars="EVTYPE", variable.name = "Damage_Types")
head(top10damages, 5)
```

```
##        EVTYPE Damage_Types     value
## 1     TORNADO      PROPDMG 3212258.2
## 2 FLASH FLOOD      PROPDMG 1420124.6
## 3   TSTM WIND      PROPDMG 1335965.6
## 4        HAIL      PROPDMG  688693.4
## 5       FLOOD      PROPDMG  899938.5
```

Highest economic consequences, top 10

```
DamageChart <- ggplot(top10damages, aes(x=reorder(EVTYPE, -value/100000), y=value/100000))
DamageChart = DamageChart + geom_bar(stat="identity", aes(fill=Damage_Types), position="dodge")
DamageChart = DamageChart + xlab("Types of events") +ylab("Costs of damage in $ (billions)")
DamageChart = DamageChart + theme(axis.text.x = element_text(angle=45, hjust=1))
```

```r
DamageChart = DamageChart + ggtitle("Highest economic consequences, top 10") + theme(plot.title = elemen
DamageChart
```



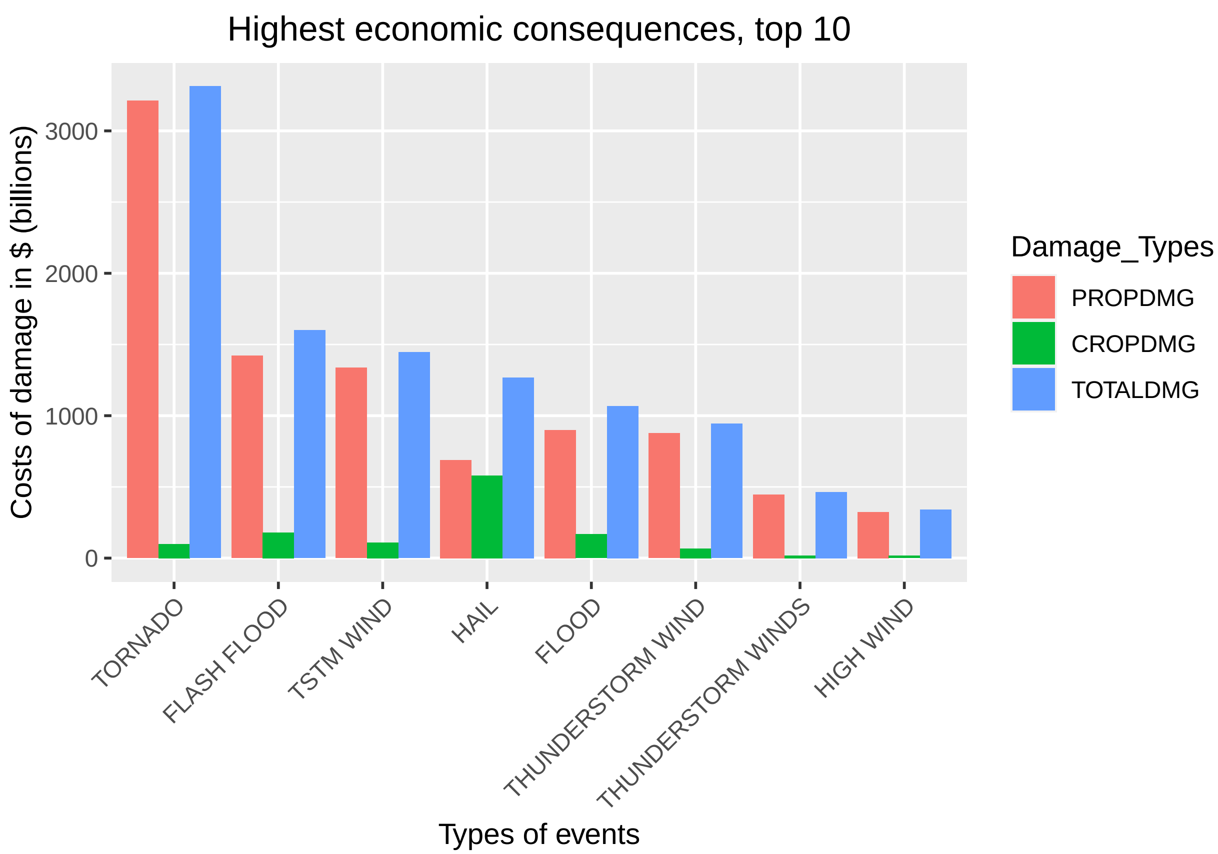
Top 5 Event-Type - Damage Type ="TOTALDMG"

```r
totalDamage<- merge(propdmgMax,cropdmgMax,by.x = "EVTYPE", by.y = "EVTYPE")
totalDamage$TOTALDMG <- totalDamage$PROPDMG + totalDamage$CROPDMG
totalDamage<-arrange(totalDamage,desc(TOTALDMG))
top10damages <- melt(totalDamage, id.vars="EVTYPE", variable.name = "Damage_Types")
tail(top10damages, 5)
```

```
##                EVTYPE Damage_Types     value
## 20               HAIL     TOTALDMG 1268289.7
## 21              FLOOD     TOTALDMG 1067976.4
## 22  THUNDERSTORM WIND     TOTALDMG  943635.6
## 23 THUNDERSTORM WINDS     TOTALDMG  464978.1
## 24          HIGH WIND     TOTALDMG  342014.8
```

Highest economic consequences, top 10

```
DamageChart <- ggplot(top10damages, aes(x=reorder(EVTYPE, -value/1000), y=value/1000),fill=Damage_Types
DamageChart = DamageChart + geom_bar(stat="identity", aes(fill=Damage_Types), position="dodge")
DamageChart = DamageChart + xlab("Types of events") + ylab("Costs of damage in $ (billions)")
DamageChart = DamageChart + theme(axis.text.x = element_text(angle=45, hjust=1))
DamageChart = DamageChart + ggtitle("Highest economic consequences, top 10") + theme(plot.title = elemen
DamageChart
```



Summary: The analysis of the NOAA Storm Database has shown that the most common types of severe weather events in the United States are tornadoes, hurricanes, and storms. These events occur most frequently in the South and Southeast United States. The most serious health and economic impacts of severe weather events are fatalities, injuries, and property damage.

Discussion of Implications: The findings of the analysis have important implications for public policy and decision-making. The government should take action to protect the public from the consequences of severe weather events. This includes improved early warning systems, disaster preparedness plans, and recovery efforts. Businesses and individuals should also take action to protect themselves from the consequences of severe weather events. This includes insurance, emergency plans, and supplies.

Recommendations for Further Research: The findings of the analysis can be supported or expanded by further research. Areas that should be further investigated include:

- The impacts of severe weather events on different population groups
- The impacts of severe weather events on the environment
- The costs of severe weather events
- The effectiveness of disaster mitigation measures

Conclusion: The analysis of the NOAA Storm Database has provided important insights into the severity and extent of severe weather events in the United States. The findings of the analysis have important implications for public policy and decision-making. By implementing disaster mitigation measures, the public can be protected from the consequences of these events.