

# project2

2023-05-04

```
library(ggplot2)
library(plyr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:plyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
```

```
## The following objects are masked from 'package:stats':
##
##   filter, lag
```

```
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
if(!exists("storm.data")) {
  storm.data <- read.csv(bzfile("repdata-data-StormData.csv.bz2"),header = TRUE)
}
dim(storm.data)
```

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

```
str(storm.data)
```

```
## '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 ...

```

```
tail(storm.data)
```

```

##          STATE__      BGN_DATE    BGN_TIME TIME_ZONE COUNTY
## 902292      47 11/28/2011 0:00:00 03:00:00 PM      CST      21
## 902293      56 11/30/2011 0:00:00 10:30:00 PM      MST       7
## 902294      30 11/10/2011 0:00:00 02:48:00 PM      MST       9
## 902295       2 11/8/2011 0:00:00 02:58:00 PM      AKS      213
## 902296       2 11/9/2011 0:00:00 10:21:00 AM      AKS      202
## 902297       1 11/28/2011 0:00:00 08:00:00 PM      CST       6
##
##          COUNTYNAME STATE      EVTYPE BGN_RANGE
## 902292 TNZ001>004 - 019>021 - 048>055 - 088    TN WINTER WEATHER      0
## 902293          WYZ007 - 017     WY      HIGH WIND      0
## 902294          MTZ009 - 010     MT      HIGH WIND      0
## 902295          AKZ213     AK      HIGH WIND      0
## 902296          AKZ202     AK      BLIZZARD      0
## 902297          ALZ006     AL      HEAVY SNOW      0
##
##          BGN_AZI BGN_LOCATI      END_DATE    END_TIME COUNTY_END COUNTYENDN
## 902292          11/29/2011 0:00:00 12:00:00 PM          0          NA
## 902293          11/30/2011 0:00:00 10:30:00 PM          0          NA
## 902294          11/10/2011 0:00:00 02:48:00 PM          0          NA
## 902295          11/9/2011 0:00:00 01:15:00 PM          0          NA
## 902296          11/9/2011 0:00:00 05:00:00 PM          0          NA
## 902297          11/29/2011 0:00:00 04:00:00 AM          0          NA
##
##          END_RANGE END_AZI END_LOCATI LENGTH WIDTH  F MAG FATALITIES INJURIES
## 902292          0          0          0 0 NA  0          0          0
## 902293          0          0          0 0 NA 66          0          0
## 902294          0          0          0 0 NA 52          0          0
## 902295          0          0          0 0 NA 81          0          0

```

```

## 902296      0      0      0 NA      0      0
## 902297      0      0      0 NA      0      0
##      PROPDMG PROPDMGEXP CROPDGMG CROPDMGEXP WFO      STATEOFFIC
## 902292      0      K      0      K MEG      TENNESSEE, West
## 902293      0      K      0      K RIW WYOMING, Central and West
## 902294      0      K      0      K TFX      MONTANA, Central
## 902295      0      K      0      K AFG      ALASKA, Northern
## 902296      0      K      0      K AFG      ALASKA, Northern
## 902297      0      K      0      K HUN      ALABAMA, North
##
## 902292 LAKE - LAKE - OBION - WEAKLEY - HENRY - DYER - GIBSON - CARROLL - LAUDERDALE - TIPTON - HAYWOOD
## 902293      OWL CREEK & BRIDGE
## 902294      NORTH ROCK
## 902295
## 902296
## 902297
##      LATITUDE LONGITUDE LATITUDE_E LONGITUDE_
## 902292      0      0      0      0
## 902293      0      0      0      0
## 902294      0      0      0      0
## 902295      0      0      0      0
## 902296      0      0      0      0
## 902297      0      0      0      0
##
## 902292
## 902293
## 902294
## 902295 EPISODE NARRATIVE: A 960 mb low over the southern Aleutians at 0300AKST on the 8th intensified
## 902296 EPISODE NARRATIVE: A 960 mb low over the southern Aleutians at 0300AKST on the 8th intensified
## 902297      EPISODE NARRATIVE: An intense upper level low developed on the 28th
##      REFNUM
## 902292 902292
## 902293 902293
## 902294 902294
## 902295 902295
## 902296 902296
## 902297 902297

```

```

vars <- c( "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CROPDGMG", "CROPDMGEXP")
mydata <- storm.data[, vars]

```

```

sum(is.na(mydata$FATALITIES))

```

```

## [1] 0

```

```

sum(is.na(mydata$INJURIES))

```

```

## [1] 0

```

```

sum(is.na(mydata$PROPDMG))

```

```

## [1] 0

```

```
sum(is.na(mydata$CROPDMG))
```

```
## [1] 0
```

```
sum(is.na(mydata$PROPDMGEXP))
```

```
## [1] 0
```

```
sum(is.na(mydata$CROPDMGEXP))
```

```
## [1] 0
```

```
sort(table(mydata$EVTYPE), decreasing = TRUE)[1:10]
```

```
##
##          HAIL          TSTM WIND  THUNDERSTORM WIND          TORNADO
##          288661          219940          82563          60652
##    FLASH FLOOD          FLOOD THUNDERSTORM WINDS    HIGH WIND
##          54277          25326          20843          20212
##    LIGHTNING    HEAVY SNOW
##          15754          15708
```

```
mydata$EVENT<- "OTHER"
mydata$EVENT[grepl("HAIL", mydata$EVTYPE, ignore.case = TRUE)] <- "HAIL"
mydata$EVENT[grepl("HEAT", mydata$EVTYPE, ignore.case = TRUE)] <- "HEAT"
mydata$EVENT[grepl("FLOOD", mydata$EVTYPE, ignore.case = TRUE)] <- "FLOOD"
mydata$EVENT[grepl("WIND", mydata$EVTYPE, ignore.case = TRUE)] <- "WIND"
mydata$EVENT[grepl("STORM", mydata$EVTYPE, ignore.case = TRUE)] <- "STORM"
mydata$EVENT[grepl("SNOW", mydata$EVTYPE, ignore.case = TRUE)] <- "SNOW"
mydata$EVENT[grepl("TORNADO", mydata$EVTYPE, ignore.case = TRUE)] <- "TORNADO"
mydata$EVENT[grepl("WINTER", mydata$EVTYPE, ignore.case = TRUE)] <- "WINTER"
mydata$EVENT[grepl("RAIN", mydata$EVTYPE, ignore.case = TRUE)] <- "RAIN"
sort(table(mydata$EVENT), decreasing = TRUE)
```

```
##
##    HAIL    WIND    STORM    FLOOD TORNADO    OTHER    WINTER    SNOW    RAIN    HEAT
## 289270 255362 113156  82686  60700  48970  19604  17660  12241  2648
```

```
sort(table(mydata$PROPDMGEXP), decreasing = TRUE)[1:10]
```

```
##
##          K          M          0          B          5          1          2          ?          m
## 465934 424665 11330  216  40  28  25  13  8  7
```

```
sort(table(mydata$CROPDMGEXP), decreasing = TRUE)[1:10]
```

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

```

mydata$PROPDMGEXP <- as.character(mydata$PROPDMGEXP)
mydata$PROPDMGEXP[is.na(mydata$PROPDMGEXP)] <- 0
mydata$PROPDMGEXP[!grepl("K|M|B", mydata$PROPDMGEXP, ignore.case = TRUE)] <- 0
mydata$PROPDMGEXP[grepl("K", mydata$PROPDMGEXP, ignore.case = TRUE)] <- "3"
mydata$PROPDMGEXP[grepl("M", mydata$PROPDMGEXP, ignore.case = TRUE)] <- "6"
mydata$PROPDMGEXP[grepl("B", mydata$PROPDMGEXP, ignore.case = TRUE)] <- "9"
mydata$PROPDMGEXP <- as.numeric(as.character(mydata$PROPDMGEXP))
mydata$property.damage <- mydata$PROPDMG * 10^mydata$PROPDMGEXP

mydata$CROPDMGEXP <- as.character(mydata$CROPDMGEXP)
mydata$CROPDMGEXP[is.na(mydata$CROPDMGEXP)] <- 0
mydata$CROPDMGEXP[!grepl("K|M|B", mydata$CROPDMGEXP, ignore.case = TRUE)] <- 0
mydata$CROPDMGEXP[grepl("K", mydata$CROPDMGEXP, ignore.case = TRUE)] <- "3"
mydata$CROPDMGEXP[grepl("M", mydata$CROPDMGEXP, ignore.case = TRUE)] <- "6"
mydata$CROPDMGEXP[grepl("B", mydata$CROPDMGEXP, ignore.case = TRUE)] <- "9"
mydata$CROPDMGEXP <- as.numeric(as.character(mydata$CROPDMGEXP))
mydata$crop.damage <- mydata$CROPDMG * 10^mydata$CROPDMGEXP

```

## aggregate

```

agg.fatalities.and.injuries <- ddply(mydata, .(EVENT), summarize, Total = sum(FATALITIES + INJURIES, na.rm = TRUE))
agg.fatalities.and.injuries$type <- "fatalities and injuries"
agg.fatalities <- ddply(mydata, .(EVENT), summarize, Total = sum(FATALITIES, na.rm = TRUE))
agg.fatalities$type <- "fatalities"
agg.injuries <- ddply(mydata, .(EVENT), summarize, Total = sum(INJURIES, na.rm = TRUE))
agg.injuries$type <- "injuries"
agg.health <- rbind(agg.fatalities, agg.injuries)

health.by.event <- join(agg.fatalities, agg.injuries, by="EVENT", type="inner")
health.by.event

```

| ##    | EVENT   | Total | type       | Total | type     |
|-------|---------|-------|------------|-------|----------|
| ## 1  | FLOOD   | 1524  | fatalities | 8602  | injuries |
| ## 2  | HAIL    | 15    | fatalities | 1371  | injuries |
| ## 3  | HEAT    | 3138  | fatalities | 9224  | injuries |
| ## 4  | OTHER   | 2626  | fatalities | 12224 | injuries |
| ## 5  | RAIN    | 114   | fatalities | 305   | injuries |
| ## 6  | SNOW    | 164   | fatalities | 1164  | injuries |
| ## 7  | STORM   | 416   | fatalities | 5339  | injuries |
| ## 8  | TORNADO | 5661  | fatalities | 91407 | injuries |
| ## 9  | WIND    | 1209  | fatalities | 9001  | injuries |
| ## 10 | WINTER  | 278   | fatalities | 1891  | injuries |

```

agg.propdmg.and.cropdmg <- ddply(mydata, .(EVENT), summarize, Total = sum(property.damage + crop.damage, na.rm = TRUE))
agg.propdmg.and.cropdmg$type <- "property and crop damage"

agg.prop <- ddply(mydata, .(EVENT), summarize, Total = sum(property.damage, na.rm = TRUE))
agg.prop$type <- "property"

agg.crop <- ddply(mydata, .(EVENT), summarize, Total = sum(crop.damage, na.rm = TRUE))

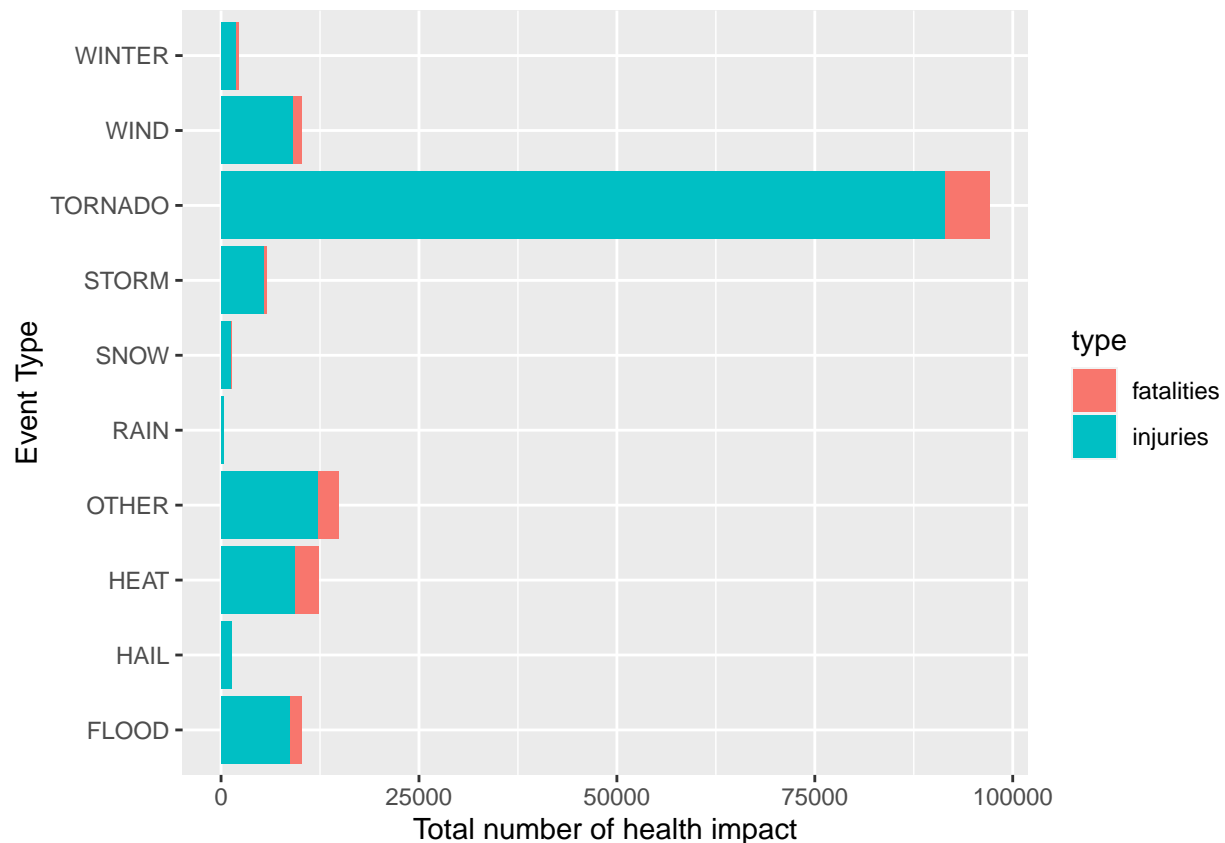
```

```
agg.crop$type <- "crop"
agg.economic <- rbind(agg.prop, agg.crop)

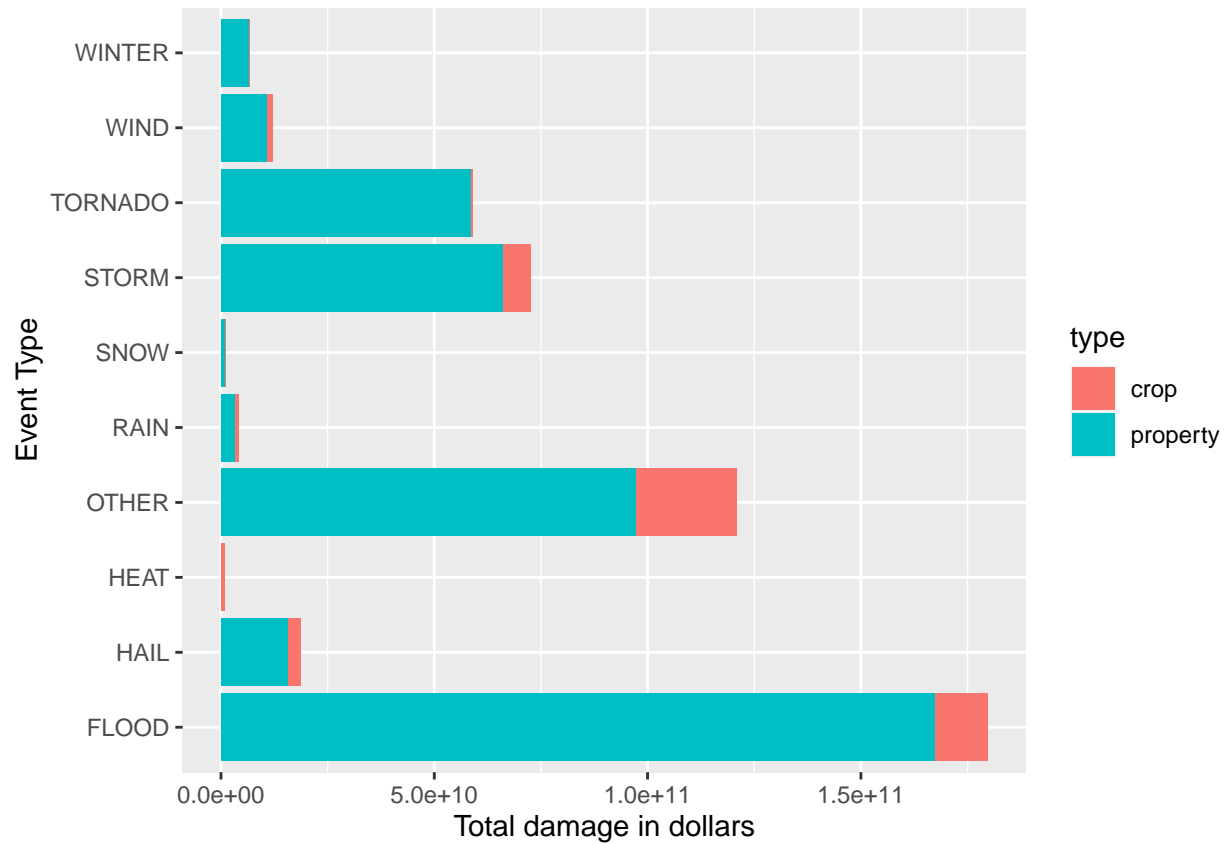
economic.by.event <- join (agg.prop, agg.crop, by="EVENT", type="inner")
economic.by.event
```

```
##      EVENT      Total      type      Total type
## 1  FLOOD 167502193929 property 12266906100 crop
## 2   HAIL 15733043048 property  3046837473 crop
## 3   HEAT  20325750 property   904469280 crop
## 4  OTHER 97246712337 property 23588880870 crop
## 5   RAIN  3270230192 property   919315800 crop
## 6   SNOW 1024169752 property   134683100 crop
## 7  STORM 66304415393 property  6374474888 crop
## 8 TORNADO 58593098029 property   417461520 crop
## 9   WIND 10847166618 property  1403719150 crop
## 10 WINTER  6777295251 property   47444000 crop
```

```
agg.health$EVENT <- as.factor(agg.health$EVENT)
ggplot(agg.health, aes(x = EVENT, y = Total, fill = type)) + geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("Total number of health impact") +
  theme(plot.title = element_text(hjust = 0.5))
```



```
ggplot(agg.economic, aes(x = EVENT, y = Total, fill = type)) + geom_bar(stat = "identity") +
  coord_flip() +
  xlab("Event Type") +
  ylab("Total damage in dollars") +
  theme(plot.title = element_text(hjust = 0.5))
```



Tornado is the most harmful weather event for health. Flood has the greatest economic consequences.