RepData_PeerAssessment2

by Fabio Bianchini

The following timelines show the different time spans for each period of unique data collection and processing procedures. Select below for detailed decriptions of each data collection type. https://www.ncdc.noaa.gov/stormevents/details.jsp

Loading Raw Data

```
url <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
download(url, "Storm_Data.bz2", mode = "wb") #Download dataset from specific URL
bunzip2("Storm_Data.bz2", "Storm_data.csv", remove = FALSE, skip = TRUE) # unzip data file
```

```
## [1] "Storm_data.csv"
## attr(,"temporary")
## [1] FALSE
```

```
Storm_Data <- read.csv("Storm_data.csv") #
```

```
dim(Storm_Data) # Original dataset dimension
```

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

```
names(Storm_Data) # Variables name in the orginal 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"
```

Process/transform the data into a format suitable for the analysis

```
ds1 <- as_tibble(Storm_Data)
# variable must have a unique name in the dataset
names(ds1)[names(ds1)=="STATE__"] <- "STATE_NUM"
names(ds1)[names(ds1)=="LONGITUDE_"] <- "LONGITUDE_E"
names(ds1) <- str_to_lower(names(ds1)) # Force lowercase dataset columb names
names(ds1) <-str_replace(names(ds1), "_+$","") # Remove final underscore from columb names
names(ds1) <- str_replace(names(ds1), "_",".") #
names(ds1)
```

```
##  [1] "state.num"   "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.e"
## [36] "remarks"     "refnum"
```

```r
# Remove the observation with no interest for answer the question for this analysis
ds2 <- ds1[ds1$fatalities > 0 | ds1$injuries > 0 | ds1$cropdmg > 0 | ds1$propdmg > 0,]
dim(ds2)
```

```
## [1] 254633     37
```

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

The variables of interest, for analazing the impact on population healt are `fatalites` and `injuries`so we create a subset from the original dataset with only the variable of interest.

```r
# Create a dataset with only the columb/variable of interest to answer this question
ds3 <- select(ds2, fatalities, injuries, evtype)
# Force all `evtypes` to uppercase
ds3$evtype <- str_to_upper(ds3$evtype)
# replace multiple spaces with single space
ds3$evtype <- gsub(" +", " ", ds3$evtype)
# Summarize fatalities and injuries valure grouped by `evtype`
ds4 <- ds3 %>% group_by(evtype) %>%
        summarise(tot.fatalities = sum(fatalities), tot.injuries = sum(injuries))

# Dimension for summarized dataset
dim(ds4) #
```

```
## [1] 443   3
```

```r
# Re-organize the dataset
fatalities <- arrange(ds4, desc(tot.fatalities))
head(fatalities)
```

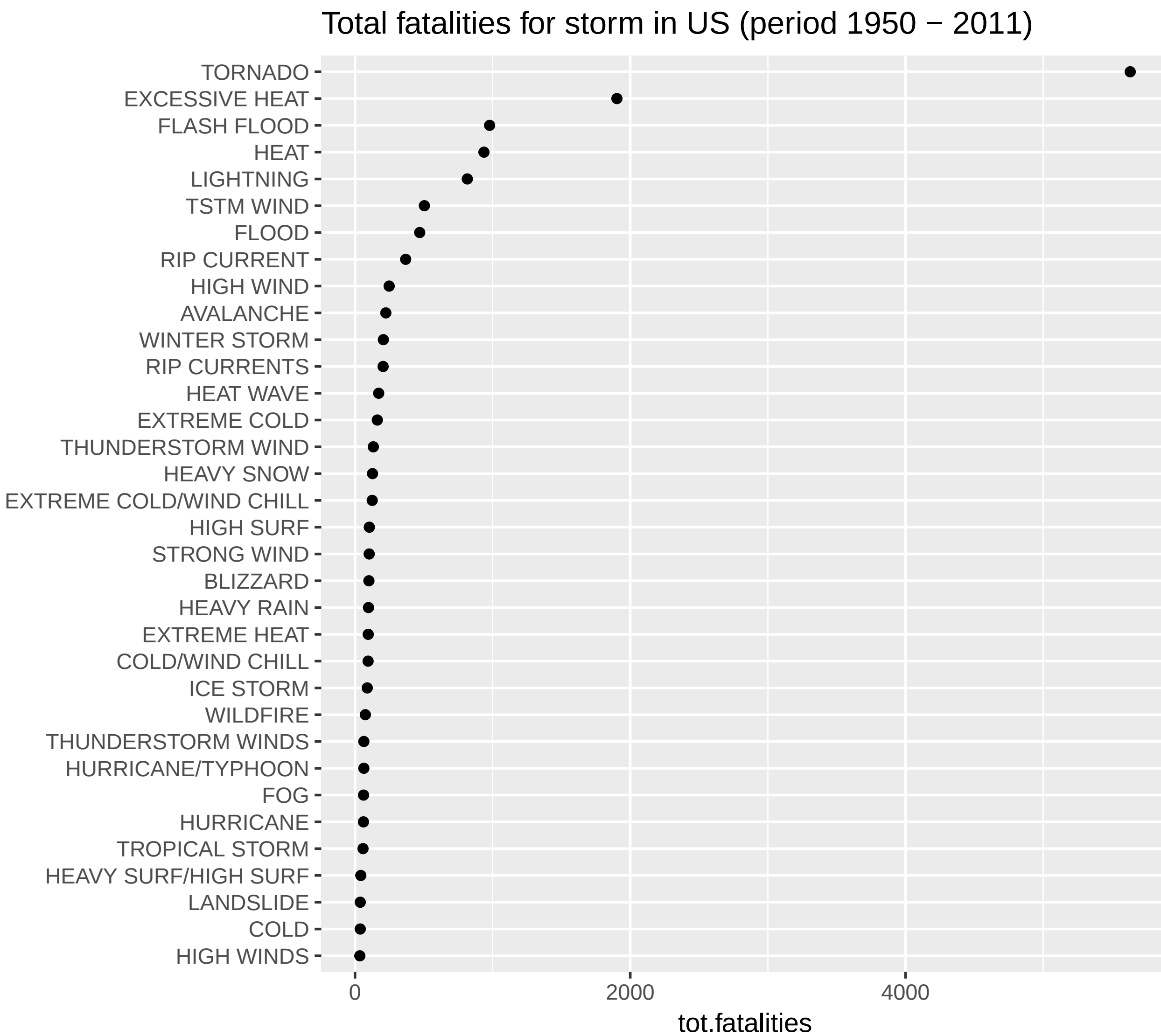
```
## # A tibble: 6 × 3
##           evtype tot.fatalities tot.injuries
##            <chr>          <dbl>        <dbl>
## 1        TORNADO           5633        91346
## 2 EXCESSIVE HEAT           1903         6525
## 3    FLASH FLOOD            978         1777
## 4           HEAT            937         2100
## 5      LIGHTNING            816         5230
## 6      TSTM WIND            504         6957
```

Fatalitis analysis For this analysis we will consider only the events with n. of fatalities greater that the mean

```r
plot_fatalities <- fatalities[fatalities$tot.fatalities > mean(fatalities$tot.fatalities), ]
nrow(plot_fatalities) # Events with n. of fatalities greater that the mean
```

```
## [1] 34
```

```r
ggplot(plot_fatalities, aes(tot.fatalities, fct_reorder(evtype, tot.fatalities))) + geom_point() + labs
```



The TORNADO event has most harmful impact on public health with n. **5633** total fatalities.

The first 10th Fatalities events

```r
library(knitr)
kable(plot_fatalities[1:10,])
```

evtype	tot.fatalities	tot.injuries
TORNADO	5633	91346
EXCESSIVE HEAT	1903	6525
FLASH FLOOD	978	1777
HEAT	937	2100
LIGHTNING	816	5230
TSTM WIND	504	6957
FLOOD	470	6789
RIP CURRENT	368	232
HIGH WIND	248	1137

evtype	tot.fatalities	tot.injuries
AVALANCHE	224	170
Injuries analys	is**	
For this analysis	we will consider	only the events with n. of injuries greater that the mean

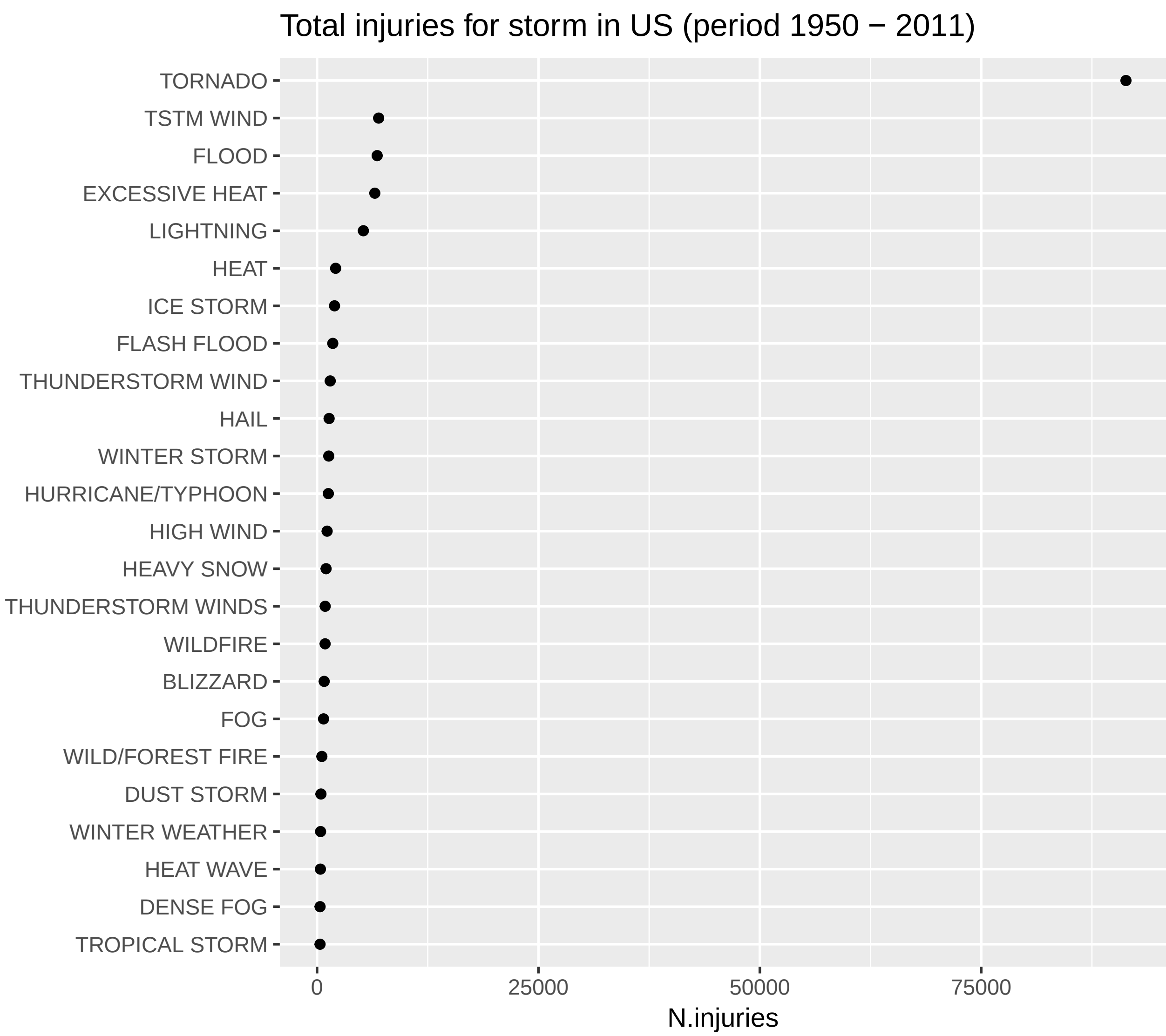
```
injuries <- arrange(ds4, desc(tot.injuries))
mean(injuries$tot.injuries) # Mean value for injuries
```

```
## [1] 317.219
```

```
plot_injuries <- injuries[injuries$tot.injuries > mean(injuries$tot.injuries), ]
nrow(plot_injuries) # Events with n. of injuries greater that the mean
```

```
## [1] 24
```

```
ggplot(plot_injuries, aes(tot.injuries, fct_reorder(evtype, tot.injuries))) + geom_point() + labs(title=
```



The TORNADO event has most harmful impact on public health with n. **91346** total injuries.

The first 10th injuries events

```r
library(knitr)
kable(plot_injuries[1:10,])
```

evtype	tot.fatalities	tot.injuries
TORNADO	5633	91346
TSTM WIND	504	6957
FLOOD	470	6789
EXCESSIVE HEAT	1903	6525
LIGHTNING	816	5230
HEAT	937	2100
ICE STORM	89	1975
FLASH FLOOD	978	1777
THUNDERSTORM WIND	133	1488
HAIL	15	1361

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

The variables of interest for analazing the **greatest economic consequences of a Storm event** are `Property damage` and `Crop damage`, so we create a subset from the original dataset with only the variables of interest

```r
damage <- select(ds2, evtype, propdmg, propdmgexp, cropdmg, cropdmgexp)
```

Due to the particulary form for storm data damage in the original dataset, we need to convert this variables in a form suitable per the correct analysis and rappresentation.

```r
# Convert cropdmgexp and propdmgexp variables
damage$propdmgexp <- as.character(damage$propdmgexp)
damage$cropdmgexp <- as.character(damage$cropdmgexp)
damage$propdmgexp <- str_to_upper(damage$propdmgexp)
damage$cropdmgexp <- str_to_upper(damage$cropdmgexp)
#
damage$propdmg.value <- 0 # New dataset columb for property damage value
damage[damage$propdmgexp == "K", ]$propdmg.value <- 3
damage[damage$propdmgexp == "M", ]$propdmg.value <- 6
damage[damage$propdmgexp == "B", ]$propdmg.value <- 7
#
damage$cropdmg.value <- 0 # New dataset columb for crop damage value
damage[damage$cropdmgexp == "K", ]$cropdmg.value <- 3
damage[damage$cropdmgexp == "M", ]$cropdmg.value <- 6
damage[damage$cropdmgexp == "B", ]$cropdmg.value <- 7
#
damage$totdmg.value <- 0 # New dataset columb for total damage value
names(damage)
```

```
## [1] "evtype"        "propdmg"       "propdmgexp"    "cropdmg"
## [5] "cropdmgexp"    "propdmg.value" "cropdmg.value" "totdmg.value"
```

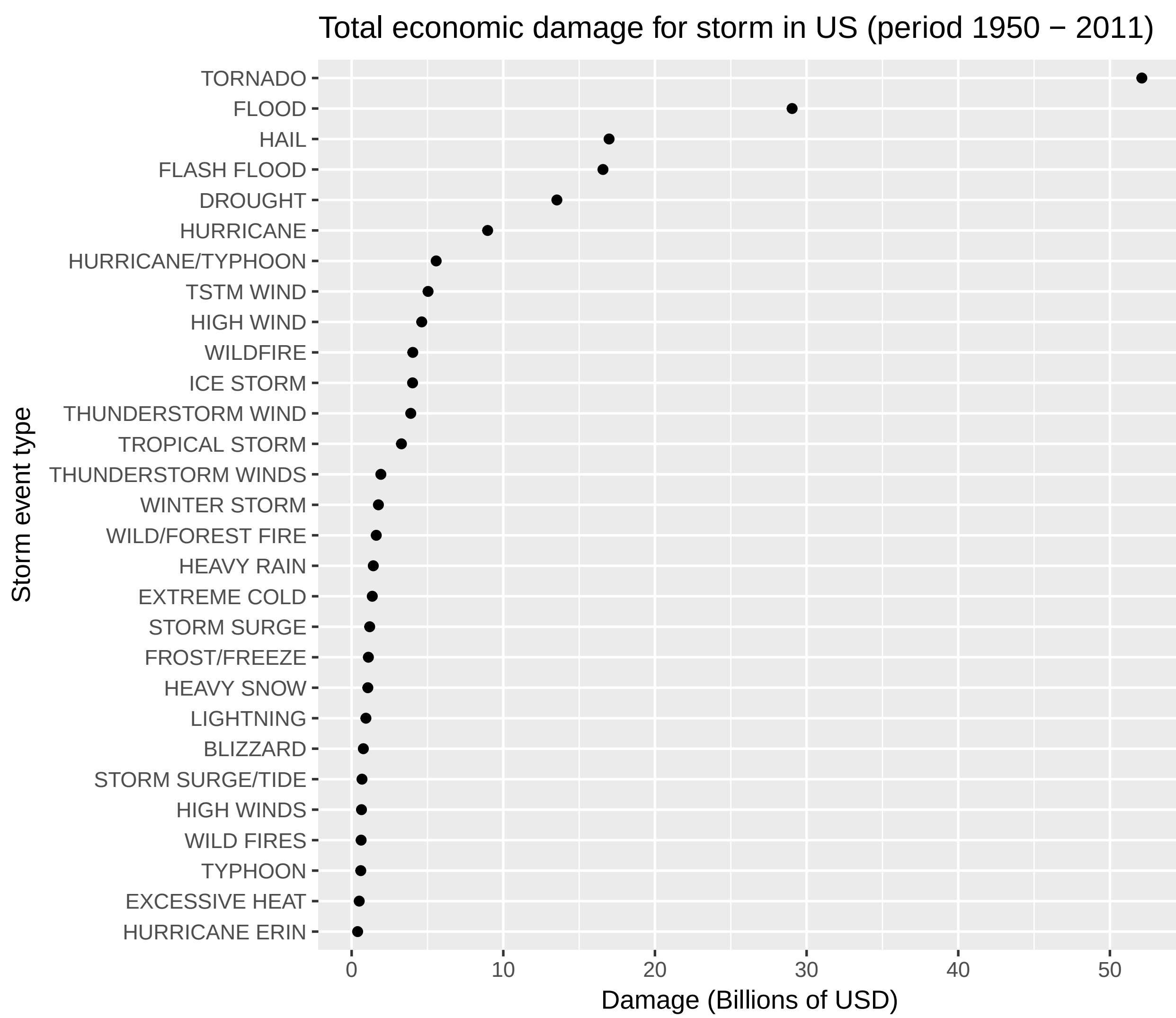
Now valorize the new `total damage value` columb as a `total property damage` and `total crop damage` value summ

```r
damage$totdmg.value <- damage$propdmg*(10^damage$propdmg.value) + damage$cropdmg*(10^damage$cropdmg.valu
```

```r
# Summarize property damage and crop damage valure grouped by `evtype`
ds5 <- damage %>% group_by(evtype) %>% summarise(total = sum(totdmg.value))
plot_damage <- arrange(ds5, desc(total))
# For the plot porpuose we consider only events with total damage value greater that the mean
plot_damage <- plot_damage[plot_damage$total > mean(plot_damage$total), ]
nrow(plot_damage) # Events with total damage amount greater that the mean
```

```
## [1] 29
```

```r
ggplot(plot_damage, aes(total/10^9, fct_reorder(evtype, total))) + geom_point() + labs(title="Total econ
```



The TORNADO event has the greatest economic consequences with **52 Billions of USD** total damage value.

The first 10th great economic events

```r
library(knitr)
kable(plot_damage[1:10,])
```

evtype	total
TORNADO	52105114049

evtype	total
FLOOD	29044678257
HAIL	16976221521
FLASH FLOOD	16572129167
DROUGHT	13533672000
HURRICANE	8967229010
HURRICANE/TYPHOON	5573812800
TSTM WIND	5038935845
HIGH WIND	4621617595
WILDFIRE	4030986800