Impact of major weather events in the United States between 2001 and 2011

Synopsis

In this report we analyse the public health and economic impact of storms and other major weather events in United States. Many of such severe events can result in fatalities, injuries, crop and property damage, and preventing such outcomes is a key concern for federal and state government and for municipalities.

We explored the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database and, using the data from 2001 to 2011, we found which type of event had most health and economic impact in those 10 years span.

This report could be of interest to government or municipal managers who might be responsible for preparing for severe weather events and will need to prioritize resources for different types of events.

Note: this report was made as an assignment for the Reproducible Research Coursera course. The report is made with RMarkdown and Knitr.

Data Processing

Reading the data

The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage. The events in the database start in the year 1950 and end in November 2011.

We obtain a compressed file from the Coursera course site: Storm data [47Mb], and we read the raw text file included.

```
storm_data_raw <- read.csv(bzfile("StormData.csv.bz2"))
```

We check the dimension and the first few rows of the data set.

```
storm_data_dim <- dim(storm_data_raw)
storm_data_dim
## [1] 902297     37
head(storm_data_raw)
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
## 1         NA         0                      14.0   100 3   0          0
```

```
## 2         NA         0                         2.0    150 2    0          0
## 3         NA         0                         0.1    123 2    0          0
## 4         NA         0                         0.0    100 2    0          0
## 5         NA         0                         0.0    150 2    0          0
## 6         NA         0                         1.5    177 2    0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

The data set contains 902297 events and 37 variables.

We are interested in the following variables.

BGN_DATE

Begin date of the event. The events could span many days, but for our purposes we date them with the begin date. We extract the year and we will use it to filter the events registered between 2001 and 2011.

There are two reasons for this filter:

- in the earlier years of the database (1950s-1960s-1970s-1980s) there are generally fewer events recorded, most likely due to a lack of good records, and so data from those periods could not be used to judge the relative impact of type of events without a much accurate analysis of the data
- we think that that the last 10 years of data is enough to evaluate which are the most impactful type of events

EVTYPE

The NOAA storm database code book reports 48 event type. The event types in the data set are more than 9 hundred.

```
length(unique(storm_data_raw$EVTYPE))
## [1] 985
```

Some difference are caused by upper and lower cases, but many of them are caused by incorrect imputation, especially in the early years. There is not an easy way to correct them so we looked for event types with big values for fatalities and damage and use regular expression substitution to correct them.

```
storm_data_corrected <- storm_data_raw
storm_data_corrected$EVTYPE <- toupper(storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^(SMALL )?HAIL.*", "HAIL", 
storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("TSTM|THUNDERSTORMS?", "THUNDERSTORM", 
storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("STORMS?", "STORM", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("WINDS?|WINDS?/HAIL", "WIND", 
storm_data_corrected$EVTYPE)
```

```r
storm_data_corrected$EVTYPE <- gsub("RAINS?", "RAIN", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^TH?UN?DEE?RS?TO?RO?M ?WIND.*|^(SEVERE )?
THUNDERSTORM$|^WIND STORM$|^(DRY )?MI[CR][CR]OBURST.*|^THUNDERSTORMW$", "THUNDERSTORM
WIND", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^COASTAL ?STORM$|^MARINE ACCIDENT$", "MARINE
THUNDERSTORM WIND", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^FLOODS?.*|^URBAN/SML STREAM FLD$|^(RIVER|TIDAL|
MAJOR|URBAN|MINOR|ICE JAM|RIVER AND STREAM|URBAN/SMALL STREAM)? FLOOD(ING)?S?$|^HIGH
WATER$|^URBAN AND SMALL STREAM FLOODIN$|^DROWNING$|^DAM BREAK$", "FLOOD",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^FLASH FLOOD.*|^RAPIDLY RISING WATER$", "FLASH
FLOOD", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("WATERSPOUTS?", "WATERSPOUT",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("WEATHER/MIX", "WEATHER",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("CURRENTS?", "CURRENT", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^WINDCHILL$|^COLD.*|^LOW TEMPERATURE$|^UNSEASONABLY
COLD$", "COLD/WIND CHILL", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^EXTREME WIND ?CHILL$|^(EXTENDED|EXTREME|RECORD)?
COLDS?$", "EXTREME COLD/WIND CHILL", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^WILD/FOREST FIRE$|^(WILD|BRUSH|FOREST)? ?FIRES?$",
"WILDFIRE", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^RAIN/SNOW$|^(BLOWING|HEAVY|EXCESSIVE|BLOWING|ICE
AND|RECORD)? ?SNOWS?.*", "HEAVY SNOW", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^FOG$", "DENSE FOG", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^(GUSTY|NON-SEVERE|NON ?-?THUNDERSTORM)? ?WIND.*|
^ICE/STRONG WIND$", "STRONG WIND", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("SURGE$", "SURGE/TIDE", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("CLOUDS?", "CLOUD", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^FROST[/\\]FREEZE$|^FROST$|^(DAMAGING)? ?FREEZE$|
^HYP[OE]R?THERMIA.*|^ICE$|^(ICY|ICE) ROADS$|^BLACK ICE$|^ICE ON ROAD$", "FROST/FREEZE",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^GLAZE.*|^FREEZING (RAIN|DRIZZLE|RAIN/SNOW|SPRAY$)
$|^WINTRY MIX$|^MIXED PRECIP(ITATION)?$|^WINTER WEATHER MIX$|^LIGHT SNOW$|^FALLING
SNOW/ICE$|^SLEET.*", "SLEET", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^HURRICANE.*", "HURRICANE/TYPHOON",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^HEAT WAVES?$|^UNSEASONABLY WARM$|^WARM WEATHER$",
"HEAT", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^(EXTREME|RECORD/EXCESSIVE|RECORD) HEAT$",
"EXCESSIVE HEAT", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^HEAVY SURF(/HIGH SURF)?.*$|^(ROUGH|HEAVY) SEAS?.*|
^(ROUGH|ROGUE|HAZARDOUS) SURF.*|^HIGH WIND AND SEAS$|^HIGH SURF.*", "HIGH SURF",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^LAND(SLUMP|SLIDE)?S?$|^MUD ?SLIDES?$|^AVALANCH?
E$", "AVALANCHE", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^UNSEASONABLY WARM AND DRY$|^DROUGHT.*|^HEAT WAVE
DROUGHT$", "DROUGHT", storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^TORNADO.*", "TORNADO",
storm_data_corrected$EVTYPE)

storm_data_corrected$EVTYPE <- gsub("^TROPICAL STORM.*", "TROPICAL STORM",
storm_data_corrected$EVTYPE)
```

```
storm_data_corrected$EVTYPE <- gsub("^MARINE MISHAP$|^HIGH WIND/SEAS$", "MARINE HIGH WIND", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^HIGH WIND.*", "HIGH WIND", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^HIGH SEAS$", "MARINE STRONG WIND", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^RIP CURRENT.*", "RIP CURRENT", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^WATERSPOUT.*", "WATERSPOUT", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^EXCESSIVE RAINFALL$|^RAIN.*|^TORRENTIAL RAINFALL$|^(HEAVY|HVY)? (RAIN|MIX|PRECIPITATION).*", "HEAVY RAIN", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^FOG.*", "FREEZING FOG", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^WINTER STORM.*", "WINTER STORM", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^THUNDERSNOW$|^ICE STORM.*", "ICE STORM", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("WAVES?|SWELLS?", "SURF", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^LIGHTNING.*", "LIGHTNING", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^WHIRLWIND$|^GUSTNADO$|^TORNDAO$", "TORNADO", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^COASTAL FLOOD.*", "COASTAL FLOOD", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^TYPHOON", "HURRICANE/TYPHOON", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^EROSION/CSTL FLOOD$|^COASTAL FLOOD/EROSION$|^COASTAL SURGE/TIDE$", "COASTAL FLOOD", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^ASTRONOMICAL HIGH TIDE$", "STORM SURGE/TIDE", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^(GROUND)? ?BLIZZARD.*$", "BLIZZARD", storm_data_corrected$EVTYPE)
storm_data_corrected$EVTYPE <- gsub("^DUST STORM.*$", "DUST STORM", storm_data_corrected$EVTYPE)
```

FATALITIES and INJURIES

Fatalities and injuries estimated for the event. These values are used to estimate the public health impact for type of events.

PROPDMG and CROPDMG

Property and crop damage estimated for the event. These values are used to estimate the economic impact for type of events.

Exponent

These variables are associated with PROPDMGEXP and CROPDMGEXP which are used as exponents to interpret the numeric values for the damage. There are not much information in the data code book about these variables.

The only symbols with a clear meaning are:

- H or h: for hundredth of dollars
- K or k: for thousands of dollars
- M or m: for million of dollars
- B or b: for billion of dollars

We sum the damage values grouped by the symbols.

```
property_data_exp <- storm_data_raw %.%
  group_by(PROPDMGEXP) %.%
  summarise(property_damage_per_exp = sum(PROPDMG)) %.%
  arrange(PROPDMGEXP)
print(xtable(property_data_exp), type="html")
```

	PROPDMGEXP	property_damage_per_exp
1		527.41
2	-	15.00
3	?	0.00
4	+	117.00
5	0	7108.30
6	1	0.00
7	2	12.00
8	3	20.00
9	4	14.50
10	5	210.50
11	6	65.00
12	7	82.00
13	8	0.00
14	B	275.85
15	h	2.00
16	H	25.00
17	K	10735292.10

18	m	38.90
19	M	140694.45

```
crop_data_exp <- storm_data_raw %.%
  group_by(CROPDMGEXP) %.%
  summarise(crop_damage_per_exp = sum(CROPDMG)) %.%
  arrange(CROPDMGEXP)
print(xtable(crop_data_exp), type="html")
```

	CROPDMGEXP	crop_damage_per_exp
1		11.00
2	?	0.00
3	0	260.00
4	2	0.00
5	B	13.61
6	k	436.00
7	K	1342955.91
8	m	10.00
9	M	34140.80

We see that the symbols without a clear meaning are associated with minimal values. So we use only the H, K, M, B symbols to interpret the damage amounts and we clear the amounts for the other symbols.

We create a function to decode the symbols and return a multiplier for the amounts. We use as base value the million of dollars, as we are interested only in the most impactful damage. This function will be used to prepare the final data set.

```
# Function to decode the EXP symbol
decode_exp <- function(exp_symbol) {
  # Normalize to millions of dollars
  if (toupper(exp_symbol) == "B") exp <-1000
  else if (toupper(exp_symbol) == "M") exp <-1
  else if (toupper(exp_symbol) == "K") exp <- 1/1000
  else if (toupper(exp_symbol) == "H") exp <- 1/10000
  # Don't know how to interpet other values
  else exp <-0
  return(exp)
}
decode_exp_v <- Vectorize(decode_exp)
```

Constant dollars

To add up and compare amounts of different years, we have to convert them in constant dollars, i.e. inflation adjusted amounts.

To compute a conversion factor for each year, we download the Consumer Price Index for All Urban Consumers: All Items from Federal Reserve Economic Data, we average the monthly CPI to obtain an annual value, and we compute a factor based at 2011 that could be multiplied with the amounts to adjust for the inflation.

```
# Get Consumer Price Index from Federal Reserve Economic Data
getSymbols("CPIAUCSL", src='FRED')
# CPI is monthly. Calculate an annual average.
annual_cpi <- apply.yearly(CPIAUCSL, mean)
# Calculate conversion factor using 2011 as the base year
conversion_factor <- 1 / annual_cpi * as.numeric(annual_cpi['2011'])
```

Final dataset

The final data set, used for the analysis, is obtained by:

- filtering the 2001-2011 events
- applying the exponent and the constant dollars factor to the property and crop amounts
- renaming the variables to make them clearer

```
storm_data <- storm_data_corrected %.%
  filter(year(mdy_hms(BGN_DATE)) >= 2001) %.%
  mutate(property_damage_exp = decode_exp_v(PROPDMGEXP)
         ,crop_damage_exp = decode_exp_v(CROPDMGEXP)
         ,event_year = year(mdy_hms(BGN_DATE))
         ,factor_cd = as.numeric(conversion_factor[as.character(event_year)])
         ,property_damage = PROPDMG * property_damage_exp * factor_cd
         ,crop_damage = CROPDMG * crop_damage_exp * factor_cd) %.%
  select(event_year
         ,event_type = EVTYPE
         ,fatalities = FATALITIES
         ,injuries = INJURIES
         ,property_damage
         ,crop_damage)
storm_data_final_dim <- dim(storm_data)
storm_data_final_dim
```

[1] 488692 6

```
print(xtable(head(storm_data), digits = c(0,0,0,0,0,6,6)), type = "html")
```

	event_year	event_type	fatalities	injuries	property_damage	crop_damage
1	2001	THUNDERSTORM WIND	0	0	0.002541	0.000000
2	2001	THUNDERSTORM WIND	0	0	0.019058	0.000000
3	2001	HAIL	0	0	0.000000	0.000000

4	2001	THUNDERSTORM WIND	0	0	0.006353	0.000000
5	2001	THUNDERSTORM WIND	0	0	0.006353	0.000000
6	2001	THUNDERSTORM WIND	0	0	0.003812	0.000000

The final data set contains 488692 events and 6 variables.

Results

The analysis aims to answer two question:

- •Across the United States, which types of events are most harmful with respect to population health?
- •Across the United States, which types of events have the greatest economic consequences?

Which types of events are most harmful?

We consider as "harmful to population health"" the events with registered fatalities and injuries.

We select only the events with fatalities and injuries.

```
fatalities_registered <- storm_data[storm_data$fatalities != 0, 3]
injuries_registered <- storm_data[storm_data$injuries != 0, 4]
fatalities_registered_sum <- sum(fatalities_registered)
injuries_registered_sum <- sum(injuries_registered)
summary(fatalities_registered)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1.00    1.00    1.00    1.71    1.00  158.00
summary(injuries_registered)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     1.0     1.0     1.0     5.7     3.0  1150.0
```

We can see that in the 2001-2011 period were registered 5517 fatalities and 32330 injuries.

The distribution of fatalities has a median of 1, a third quartile of 1 and a max value of 158. This mean that more than the 75% of the events have 1 fatalities and there are few events with a large number of fatalities.

Injuries has a similar distribution with a median of 1, a third quartile of 3 and a max value of 1150.

We can visually confirm these distribution with a boxplot of fatalities and injuries (figure 1). As the distribution is highly skewed towards 1, we have plotted the y axis as the logarithm in base 10 of fatalities and injuries to make the distribution clearer.

```
par(mfrow=c(1,2), mar=c(5,2,2,2), oma=c(0,0,3,0))
boxplot(fatalities_registered, log = "y", xlab = "Fatalities", ylim = c(1,1200))
boxplot(injuries_registered, log = "y", xlab = "Injuries", ylim = c(1,1200))
mtext("Fig. 1 - Fatalities and Injuries (log10)", side=3, line=1, outer=TRUE)
```

We find out the top 10 event types by fatalities and we compute the average number of fatalities per event types. For all the event type the average is less than 1.

```
event_type_fatalities_summary <- storm_data %.%
  group_by(event_type) %.%
  summarise(fatalities_per_type = sum(fatalities), event_count = n()) %.%
```

```
  mutate(average_fatalities = fatalities_per_type / event_count) %.%
  arrange(desc(fatalities_per_type, event_count))
print(xtable(head(event_type_fatalities_summary, n = 10), digits = c(0,0,0,0,2)),
type="html")
```

	event_type	fatalities_per_type	event_count	average_fatalities
1	TORNADO	1152	16520	0.07
2	EXCESSIVE HEAT	856	1059	0.81
3	FLASH FLOOD	573	38412	0.01
4	RIP CURRENT	433	570	0.76
5	LIGHTNING	414	8779	0.05
6	FLOOD	270	19939	0.01
7	HEAT	230	735	0.31
8	THUNDERSTORM WIND	227	154645	0.00
9	AVALANCHE	200	895	0.22
10	EXTREME COLD/WIND CHILL	143	1152	0.12

We find out the top 10 event types by injuries and we compute the average number of injuries per event types. Here stand out HURRICAN/TYPHOON with 9 average injuries per event, and EXCESSIVE HEAT with 3.

```
event_type_injuries_summary <- storm_data %.%
  group_by(event_type) %.%
  summarise(injuries_per_type = sum(injuries), event_count = n()) %.%
  mutate(average_injuries = injuries_per_type / event_count) %.%
  arrange(desc(injuries_per_type, event_count))
print(xtable(head(event_type_injuries_summary, n = 10), digits = c(0,0,0,0,2)),
type="html")
```

	event_type	injuries_per_type	event_count	average_injuries
1	TORNADO	14331	16520	0.87
2	EXCESSIVE HEAT	3242	1059	3.06
3	THUNDERSTORM WIND	2913	154645	0.02
4	LIGHTNING	2622	8779	0.30
5	HURRICANE/TYPHOON	1291	133	9.71
6	HEAT	1222	735	1.66

7	WILDFIRE	1099	3294	0.33
8	FLASH FLOOD	780	38412	0.02
9	HIGH WIND	557	15569	0.04
10	HAIL	488	154472	0.00

Merging the two list we can see that the event types with the big impact on public health are: TORNADO, EXCESSIVE HEAT, FLASH FLOOD, LIGHTNING, HEAT, THUNDERSTORM WIND.

Which types of events have the greatest economic consequences?

We consider as "with economic consequences”“ the events with registered property and crop damage.

We select only the events with property and crop damage.

```
property_damage_registered <- storm_data[storm_data$property_damage != 0, 5]
crop_damage_registered <- storm_data[storm_data$crop_damage != 0, 6]
property_damage_registered_sum <- sum(property_damage_registered)
crop_damage_registered_sum <- sum(crop_damage_registered)
summary(property_damage_registered)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       0       0       3       0  128000
summary(crop_damage_registered)
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0     0.0     0.0     1.9     0.1  1740.0
```

We can see that in the 2001-2011 period were registered 366608 property damage and 22989 crop damage (millions of dollars).

The distribution of property damage has a median of 0, a third quartile of 0 and a max value of 128000 (million of dollars). This mean that more than the 75% of the events have sub-million property damage and there are few events with damage in the millions and even in the billions of dollars.

Crop damage has a similar distribution with a median of 0, a third quartile of 0.1 and a max value of 1740 (millions of dollars).

We can visually confirm these distribution with a boxplot of property and crop damage (figure 2). As the distribution is highly skewed towards 0, we have plotted the y axis as the logarithm in base 10 of property and crop damage to make the distribution clearer.

```
par(mfrow=c(1,2), mar=c(5,2,2,2), oma=c(0,0,3,0))
boxplot(property_damage_registered, log = "y", xlab = "Property damage", ylim =
c(.00001,100000))
boxplot(crop_damage_registered, log = "y", xlab = "Crop damage", ylim =
c(.00001,100000))
mtext("Fig. 2 - Property and Crop damage - Millions of dollars (log10)", side=3, line=1,
outer=TRUE)
```

We find out the top 10 event types by property damage and we compute the average number of property damage per event types. Here stand out HURRICAN/TYPHOON with 631 million of dollars average property damage per event, and STORM SURGE/TIDE with 147 million of dollars.

```
event_type_property_damage_summary <- storm_data %.%
  group_by(event_type) %.%
```

```
  summarise(property_damage_per_type = sum(property_damage), event_count = n()) %.%
  mutate(average_property_damage = property_damage_per_type / event_count) %.%
  arrange(desc(property_damage_per_type, event_count))
print(xtable(head(event_type_property_damage_summary, n = 10), digits = c(0,0,0,0,2)),
type="html")
```

	event_type	property_damage_per_type	event_count	average_property_damage
1	FLOOD	148450	19939	7.45
2	HURRICANE/TYPHOON	83926	133	631.02
3	STORM SURGE/TIDE	54595	371	147.16
4	TORNADO	20202	16520	1.22
5	HAIL	12917	154472	0.08
6	FLASH FLOOD	12711	38412	0.33
7	TROPICAL STORM	8965	605	14.82
8	THUNDERSTORM WIND	5834	154645	0.04
9	WILDFIRE	5722	3294	1.74
10	HIGH WIND	5668	15569	0.36

We find out the top 10 event types by crop damage and we compute the average number of crop damage per event types. Here stands out HURRICAN/TYPHOON with 26 million of dollars average crop damage per event.

```
event_type_crop_damage_summary <- storm_data %.%
  group_by(event_type) %.%
  summarise(crop_damage_per_type = sum(crop_damage), event_count = n()) %.%
  mutate(average_crop_damage = crop_damage_per_type / event_count) %.%
  arrange(desc(crop_damage_per_type, event_count))
print(xtable(head(event_type_crop_damage_summary, n = 10), digits = c(0,0,0,0,2)),
type="html")
```

	event_type	crop_damage_per_type	event_count	average_crop_damage
1	DROUGHT	7862	1933	4.07
2	FLOOD	3879	19939	0.19
3	HURRICANE/TYPHOON	3552	133	26.71
4	HAIL	1880	154472	0.01

5	FROST/FREEZE	1232	1407	0.88
6	FLASH FLOOD	888	38412	0.02
7	THUNDERSTORM WIND	712	154645	0.00
8	HIGH WIND	578	15569	0.04
9	EXCESSIVE HEAT	550	1059	0.52
10	HEAVY RAIN	465	9102	0.05

Merging the two list we can see that the event types with the big impact on public health are: FLOOD, HURRICANE/TYPHOON, HAIL, FLASH FLOOD, THUNDERSTORM WIND, HIGH WIND.