Economical consequences of storms in the United States

Synopsis

The weather event data originated from the National Environmental Satellite, Data, and Information Service (NESDIS). The weather events were classified into valid events using a string pattern matching algorithm. Further information is given in section **Data Processing**. With the algorithm the total number of weather events was reduced from originally 977 to 20. The total number of deaths, injuries and costs is used as metric to incorporate both, the number of events and their respective magnitude. Thus mild weather events which appear on a common basis are weighted similar to severe events which only happen rarely. The most dangeraous wether events are tornados, heat and floods accounting to 66% of all observed deaths in the United States. Furthermore, 63% of all injuries are caused by tornados. Wind, tornados, floods, thunderstorms amd hail cause the most economic damage accounting to almost 90% of total costs.

Data Processing

```
library("readr")
library("dplyr")
```

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

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

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

```
library("ggplot2")
library("stringr")
library("purrr")
```

```
##
## Attaching package: 'purrr'
```

```
## The following objects are masked from 'package:dplyr':
##
##     contains, order_by
```

```
library("tidyr")

storm <- read_csv("data/repdata_data_StormData.csv") %>%
  select(EVTYPE, FATALITIES, INJURIES, PROPDMG, CROPDMG)
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   STATE__ = col_double(),
##   COUNTY = col_double(),
##   BGN_RANGE = col_double(),
##   COUNTY_END = col_double(),
##   END_RANGE = col_double(),
```

```
##   LENGTH = col_double(),
##   WIDTH = col_double(),
##   F = col_integer(),
##   MAG = col_double(),
##   FATALITIES = col_double(),
##   INJURIES = col_double(),
##   PROPDMG = col_double(),
##   CROPDMG = col_double(),
##   LATITUDE = col_double(),
##   LONGITUDE = col_double(),
##   LATITUDE_E = col_double(),
##   LONGITUDE_ = col_double(),
##   REFNUM = col_double()
## )

## See spec(...) for full column specifications.
names(storm) <- c("event", "deaths", "injuries", "propdmg", "cropdmg")

nrevents <- length(unique(storm$event))
```

In total there are 977 unique weather events present in the data. However, due to different capitalisation e.g. **Freezing Fog** and **FREEZING FOG** the number of duplicated events can be reduced using only lowercase letters.

```
storm$event <- tolower(storm$event)
nrevents <- length(unique(storm$event))
```

Thus lowering the number of events to 890. In addition an event classifier is applied to the complete list of events to further simplify complex events like for example **flash flood - heavy wind**. The raw event string is split at each ___ ___ (space) and / character. A pattern match between each single event string and the following valid events:

- **blizzard**
- **cold**
- **fire**
- **flood**
- **hail**
- **heat**
- **hurricane**
- **lightning**
- **microburst**
- **rain**
- **snow**
- **storm**
- **thunderstorm**
- **tornado**
- **tsunami**
- **typhoon**
- **wind**
- **waterspout**

is applied. Thus complex events like **flash flood - heavy wind** are assigned to multiple valid events e.g. **flood** and **wind**. If none of the event strings match with any of the valid events the valid event **other** was assigned.

```
classify_events <- function(data) {
  events <- str_split(string = data$event, pattern = " |/")

  valid_events <- c("blizzard", "cold", "fire", "flood", "fog", "hail", "heat",
                    "hurricane", "lightning", "microburst", "rain", "snow", "thunderstorm",
                    "tornado", "tsunami", "typhoon", "wind","waterspout")

  find_ids <- function(strings) {
    ids <- sapply(valid_events, function(x) str_detect(string = strings, pattern = x))
    if (is.array(ids)) {
      ids <- unlist(apply(ids, MARGIN = 1, which))
    } else {
      ids <- which(ids)
    }
    # Add storm here, because it is part of 'thunderstorm'
    if (any(strings %in% "storm")) ids <- c(ids, length(valid_events) + 1)
    return(ids)
  }

  ids <- vector(mode = "list", length = length(events))
  for (i in seq_along(ids)) {
    ids[[i]] <- find_ids(strings = events[[i]])
    if (i %% 10000 == 0) print(i)
  }

  # Add final event category and replicate data for events with multiple entries
  data$id <- 1:nrow(data)
  no_id <- map_int(ids, length) == 0
  ids[no_id] <- length(valid_events) + 2
  select_events <- data.frame(valid_event = c(valid_events, "storm", "other"))
  select_events$valid_id <- 1:nrow(select_events)
  ids_df <- data.frame(id = rep(1:nrow(data), times = map_int(ids, length)),
                       valid_id = unlist(ids))

  final_df <- left_join(ids_df, data, by = "id") %>%
    left_join(select_events, by = "valid_id")

  return(final_df)
}
```

```
# Classification is very time consuming. The classified data is stored in storm_clean.Rda
storm_clean <- classify_events(storm)
save(storm_clean, file = "storm_clean.rda")
```

```
load(file = "storm_clean.rda", verbose = TRUE)
```

```
## Loading objects:
##   storm_clean
```

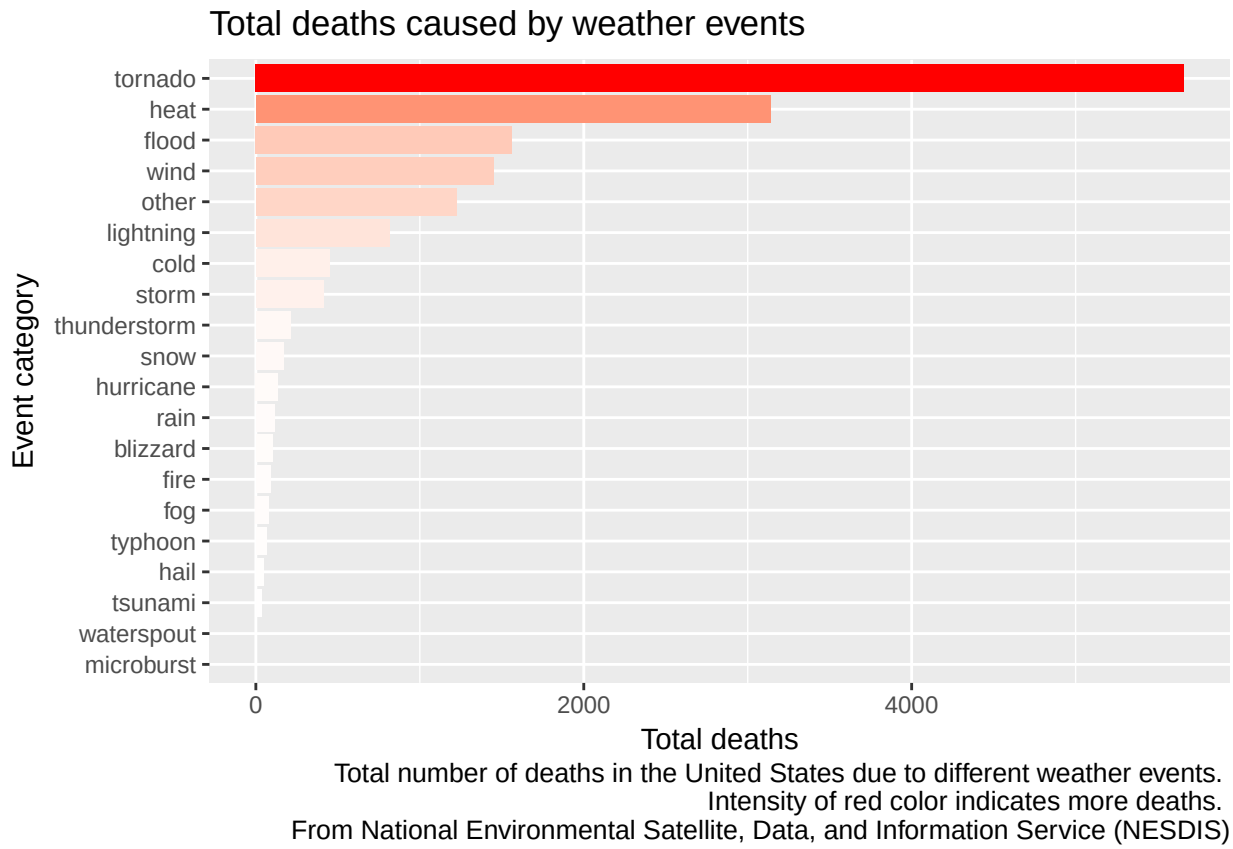
```
nrevents <- length(unique(storm_clean$valid_event))
```

The number of events has been reduced to 20.

Results

The total number of deaths and injuries is used as metric to incorporate both, the number of deaths per event and the total number of events. Thus events with a small number of deaths which happen on a regular basis e.g. **heat** are weighted similar to severe events which only happen in rare cases.

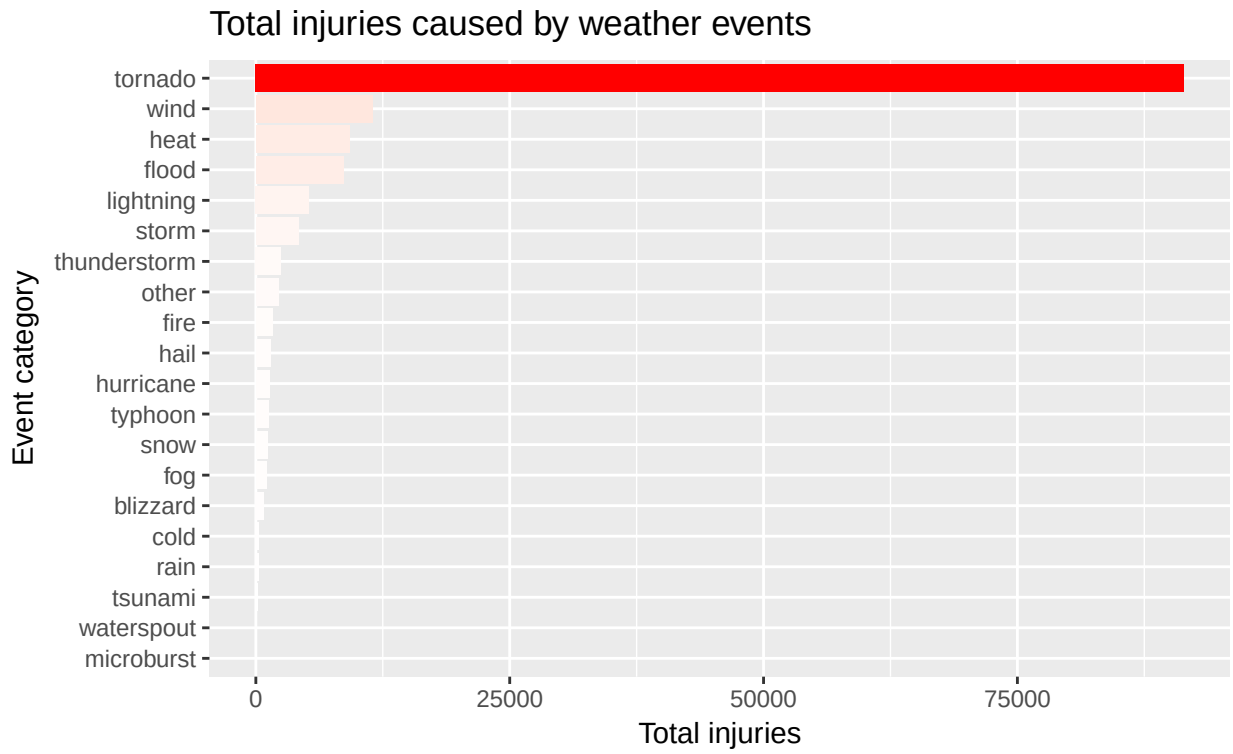
```
agg <- group_by(storm_clean, valid_event) %>%
  summarise(deaths_tot = sum(deaths),
            injuries_tot = sum(injuries))

deaths <- arrange(agg, deaths_tot)
deaths$valid_event <- factor(deaths$valid_event,
                             levels = deaths$valid_event,
                             labels = deaths$valid_event)

capt <- "Total number of deaths in the United States due to different weather events.
Intensity of red color indicates more deaths.
From National Environmental Satellite, Data, and Information Service (NESDIS)"

ggplot(deaths, aes(x = valid_event, y = deaths_tot, fill = deaths_tot)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(high = "red", low = "white") +
  coord_flip() +
  labs(x = "Event category",
       y = "Total deaths",
       title = "Total deaths caused by weather events",
       caption = capt) +
  theme(legend.position = "none")
```




```
top3 <- (nrow(deaths) - 2):nrow(deaths)
top3_death <- rev(as.character(deaths$valid_event[top3]))
top3_perc <- round(sum(deaths$deaths_tot[top3]) /
                     sum(deaths$deaths_tot), digits = 2) * 100
```

The 3 most dangerous weather events are tornado, heat, flood accounting to 66% of all observed deaths in the United States.

```
injuries <- arrange(agg, injuries_tot)
injuries$valid_event <- factor(injuries$valid_event,
                               levels = injuries$valid_event,
                               labels = injuries$valid_event)

capt <- "Total number of injuries in the United States due to different weather events.
Intensity of red color indicates more injuries.
From National Environmental Satellite, Data, and Information Service (NESDIS)"

ggplot(injuries, aes(x = valid_event, y = injuries_tot, fill = injuries_tot)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(high = "red", low = "white") +
  coord_flip() +
  labs(x = "Event category",
       y = "Total injuries",
       title = "Total injuries caused by weather events",
       caption = capt) +
  theme(legend.position = "none")
```



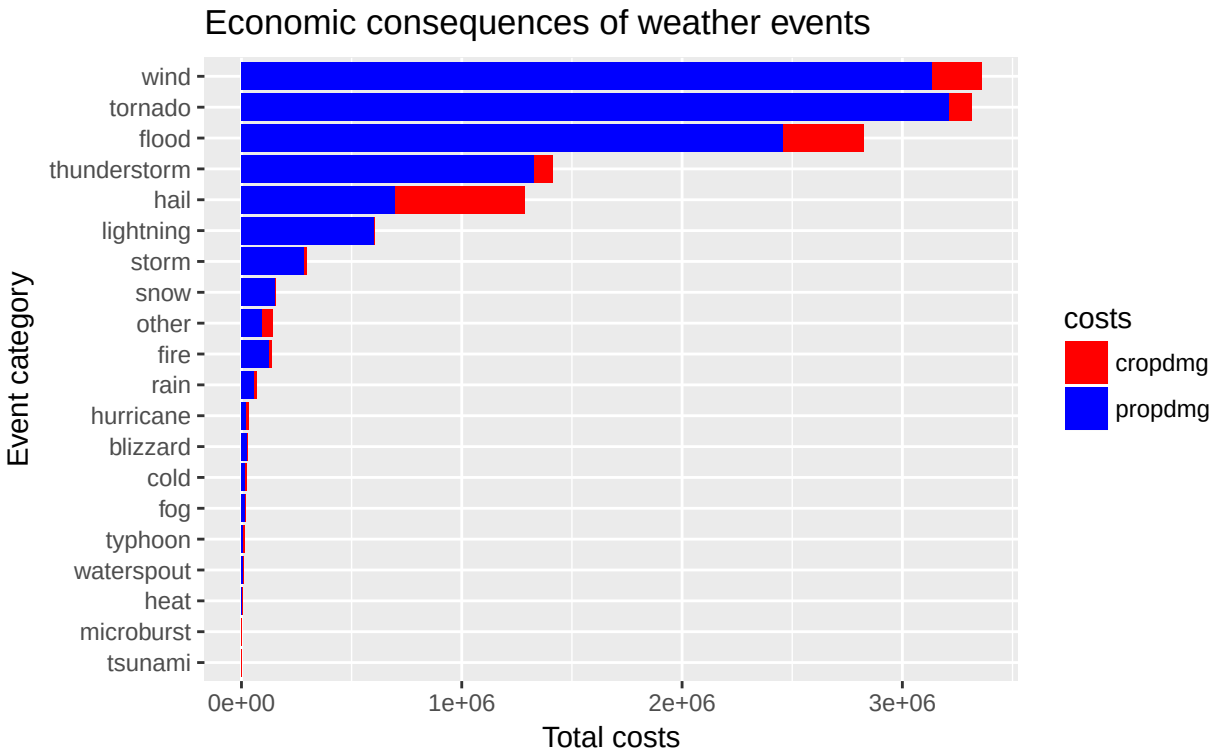
Total number of injuries in the United States due to different weather events.
Intensity of red color indicates more injuries.
From National Environmental Satellite, Data, and Information Service (NESDIS)

```
top1 <- nrow(injuries)
top1_injury <- as.character(injuries$valid_event[top1])
top1_perc <- round(sum(injuries$injuries_tot[top1]) /
                     sum(injuries$injuries_tot), digits = 2) * 100
```

63 % of all injuries due to weather events are caused by tornado.

```
cost <- select(storm_clean, event, valid_event, propdmg, cropdmg) %>%
  gather(propdmg, cropdmg, key = "type", value = "cost") %>%
  group_by(valid_event, type) %>%
  summarise(cost = sum(cost))

order <- group_by(cost, valid_event) %>%
  summarise(sum_cost = sum(cost)) %>%
  arrange(sum_cost)

cost$valid_event <- factor(cost$valid_event, levels = order$valid_event)

capt <- "Total number of costs in the United States due to different weather events.
Red color coding indicates costs related to crop damage and blue to prop damage.
From National Environmental Satellite, Data, and Information Service (NESDIS)"


ggplot(cost, aes(x = valid_event, y = cost, fill = type)) +
  geom_bar(stat = "identity") +
  coord_flip() +
  scale_fill_manual("costs", values = c("red", "blue")) +
```

```
  labs(x = "Event category",
       y = "Total costs",
       title = "Economic consequences of weather events",
       caption = capt)
```




```
top5 <- (nrow(order) - 4):nrow(order)
top5_cost <- rev(as.character(order$valid_event[top5]))
top5_perc <- round(sum(order$sum_cost[top5]) /
                     sum(order$sum_cost), digits = 2) * 100
```

The weather events causing the most economic consequences are wind, tornado, flood, thunderstorm, hail. Together they account to 89% of all costs. Prop related damage is much higher compared to crop related damage.