Analysis of Weather Events Exploring Most Detrimental Events On Human Population's Health and Economic Impact

Synopsis

The NOAA Storm Events Database is processed to provide for meaningful features. Mostly processing involves checking missing values and filtering out where applicable. Then creating meaningful multipliers to get the actual number of economic or health impact respectively. Evaluating the top 10 events per category being total health and total economic damage. Note that the top 10 events includes states but is not top 10 per state, as this would create a very large uninterpretable plot. In addition health category will include a plot for deaths by EVTYPE and STATE, as a death is very absolute whereas injuries is very broad - usually a death should be considered more pertinent than an injury.

Processing

```r
# Determining Empty or NA Values
colSums(is.na(df))
```

```
##    BGN_DATE      EVTYPE FATALITIES    INJURIES       STATE     PROPDMG PROPDMGEXP
##           0           0          0           0           0           0          0
##    CROPDMG CROPDMGEXP
##          0          0
```

```r
sapply(df, function(col) sum(col == "", na.rm = TRUE))
```

```
##    BGN_DATE      EVTYPE FATALITIES    INJURIES       STATE     PROPDMG PROPDMGEXP
##           0           0          0           0           0           0     465934
##    CROPDMG CROPDMGEXP
##          0     618413
```

```r
# Check where dmg is greater than 0 but multiplier is missing, should be a red flag here
sum(df$PROPDMG > 0 & df$PROPDMGEXP == "")
```

```
## [1] 76
```

```r
# Similarly with cropdmg
sum(df$CROPDMG > 0 & df$CROPDMGEXP == "")
```

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

```r
# Filter out these rows
df <- df[!(df$PROPDMG > 0 & df$PROPDMGEXP == ""), ]
df <- df[!(df$CROPDMG > 0 & df$CROPDMGEXP == ""), ]

# Change date column to datetime
df$BGN_DATE <- as.Date(df$BGN_DATE, format = "%m/%d/%Y")

# Feature Engineering
# Total Injuries and Fatalities
df$total_health_impact <- df$INJURIES + df$FATALITIES
```

```r
# Total for property and crop damage
# First Need to Map These Exponents: Looking at unique values we see:
unique(df$CROPDMGEXP)
```

```
## [1] ""  "M" "K" "m" "B" "0" "k" "?" "2"
```

```r
unique(df$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```r
# Make mapping
exp_map <- c(
  'H' = 1e2, 'K' = 1e3, 'M' = 1e6, 'B' = 1e9,
  'h' = 1e2, 'k' = 1e3, 'm' = 1e6, 'b' = 1e9,
  '0' = 1, '1' = 10, '2' = 100, '3' = 1000,
  '4' = 10000, '5' = 1e5, '6' = 1e6, '7' = 1e7,
  '8' = 1e8, '+' = 1, '-' = 1, '?' = 1
)


df$prop_multiplier <- exp_map[as.character(df$PROPDMGEXP)]
df$crop_multiplier <- exp_map[as.character(df$CROPDMGEXP)]
# Make NAs zero as these were events that didnt cause dmg or have economic impact
df$prop_multiplier[is.na(df$prop_multiplier)] <- 0
df$crop_multiplier[is.na(df$crop_multiplier)] <- 0
# Also add actual values for crop and prop dmg
df$actual_property_damage <- df$PROPDMG * df$prop_multiplier
df$actual_crop_damage     <- df$CROPDMG * df$crop_multiplier
```

```r
# Actually Empty Going to Make total economic based on Either both 0 evaluates to 0 otherwise multiplie
df_empty_exp <- df[
  (df$PROPDMG > 0 & df$PROPDMGEXP == "") |
  (df$CROPDMG > 0 & df$CROPDMGEXP == ""),
]

# Create total_econ_impact safely
df$total_econ_impact <- ifelse(
  df$PROPDMG == 0 & df$CROPDMG == 0,
  0,
  (df$PROPDMG * df$prop_multiplier) + (df$CROPDMG * df$crop_multiplier)
)
```

Exploring Health Impact - Deaths and Injuries

```r
library(dplyr)
```

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

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

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

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

```r
# Creating Summary Table Should be a lot less observations because of grouping
# Note n_events denoting the count of given event, may look at given summary per
# event to get an idea of severity on avg of an event
event_summary <- df %>%
  group_by(EVTYPE) %>%
  summarise(
    total_fatalities   = sum(FATALITIES, na.rm = TRUE),
    total_injuries     = sum(INJURIES, na.rm = TRUE),
    total_health_impact = sum(FATALITIES + INJURIES, na.rm = TRUE),

    total_prop_damage  = sum(actual_property_damage, na.rm = TRUE),
    total_crop_damage  = sum(actual_crop_damage, na.rm = TRUE),
    total_econ_impact  = sum(total_econ_impact, na.rm = TRUE),

    n_events           = n()
  ) %>%
  arrange(desc(total_health_impact))  # Just the one of many sorts might do
```

```r
str(event_summary)
```

```
## tibble [984 x 8] (S3: tbl_df/tbl/data.frame)
##  $ EVTYPE             : chr [1:984] "TORNADO" "EXCESSIVE HEAT" "TSTM WIND" "FLOOD" ...
##  $ total_fatalities   : num [1:984] 5633 1903 504 470 816 ...
##  $ total_injuries     : num [1:984] 91346 6525 6957 6789 5228 ...
##  $ total_health_impact: num [1:984] 96979 8428 7461 7259 6044 ...
##  $ total_prop_damage  : num [1:984] 5.69e+10 7.75e+06 4.48e+09 1.45e+11 9.30e+08 ...
##  $ total_crop_damage  : num [1:984] 4.15e+08 4.92e+08 5.54e+08 5.66e+09 1.21e+07 ...
##  $ total_econ_impact  : num [1:984] 5.74e+10 5.00e+08 5.04e+09 1.50e+11 9.42e+08 ...
##  $ n_events           : int [1:984] 60651 1678 219940 25324 15753 767 54257 2006 82562 11433 ...
```

```r
# Going to Get a Per Event Average
event_summary <- event_summary %>%
  mutate(
    econ_per_event = total_econ_impact / n_events,
    health_per_event = total_health_impact / n_events
  )
```

```r
# Get summary but by state also
event_summary_state <- df %>%
  group_by(EVTYPE, STATE) %>%
  summarise(
    total_fatalities   = sum(FATALITIES, na.rm = TRUE),
    total_injuries     = sum(INJURIES, na.rm = TRUE),
    total_health_impact = sum(FATALITIES + INJURIES, na.rm = TRUE),

    total_prop_damage  = sum(actual_property_damage, na.rm = TRUE),
    total_crop_damage  = sum(actual_crop_damage, na.rm = TRUE),
    total_econ_impact  = sum(total_econ_impact, na.rm = TRUE),

    n_events           = n(),
    .groups = "drop"
  ) %>%
```

```r
    arrange(desc(total_health_impact))  # Just the one of many sorts might do

event_summary_state <- event_summary_state %>%
  mutate(
    econ_per_event = total_econ_impact / n_events,
    health_per_event = total_health_impact / n_events
  )
```
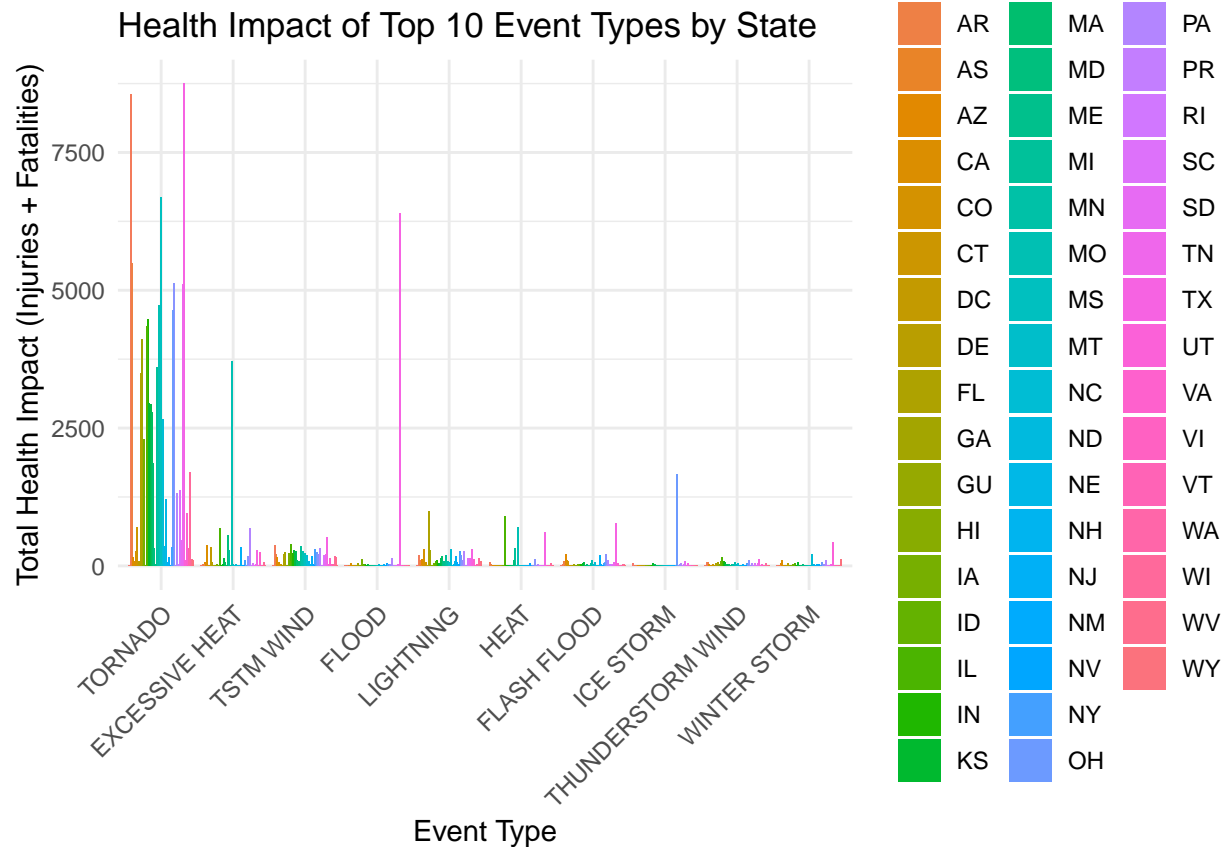
```r
top10_events <- event_summary_state %>%
  group_by(EVTYPE) %>%
  summarise(total_health_impact = sum(total_health_impact, na.rm = TRUE)) %>%
  arrange(desc(total_health_impact)) %>%
  slice_head(n = 10)

filtered_data <- event_summary_state %>%
  filter(EVTYPE %in% top10_events$EVTYPE)

library(ggplot2)

# This looks a bit too cluttered
ggplot(filtered_data, aes(x = reorder(EVTYPE, -total_health_impact),
                          y = total_health_impact, fill = STATE)) +
  geom_col(position = "dodge") +
  labs(title = "Health Impact of Top 10 Event Types by State",
       x = "Event Type", y = "Total Health Impact (Injuries + Fatalities)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Health Impact of Top 10 Event Types by State


```
# PLEASE DO IGNORE THIS PLOT IN FIG COUNT IT IS NOT USED IN RESULTS JUST FOR INTUITION!
```

```
top10_event_state <- event_summary_state %>%
  arrange(desc(total_health_impact)) %>%
  slice_head(n = 10)



health_plot <- ggplot(top10_event_state, aes(x = reorder(paste(EVTYPE, STATE, sep = " - "),
                                 -total_health_impact),
                        y = total_health_impact, fill = EVTYPE)) +
  geom_col() +
  labs(title = "Top 10 Most Harmful Event-State Combinations",
       x = "Event Type - State", y = "Total Health Impact (Injuries + Fatalities)") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(health_plot)
```
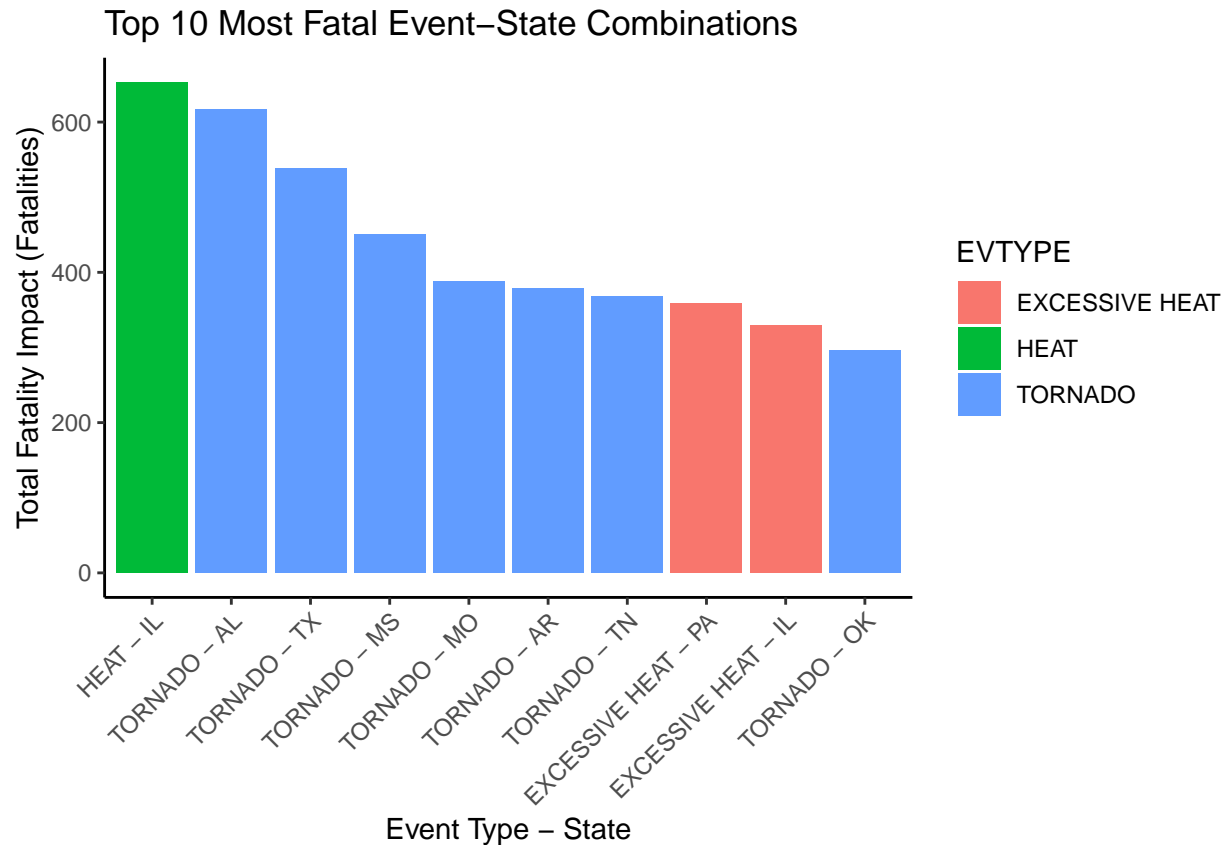
Top 10 Most Harmful Event–State Combinations
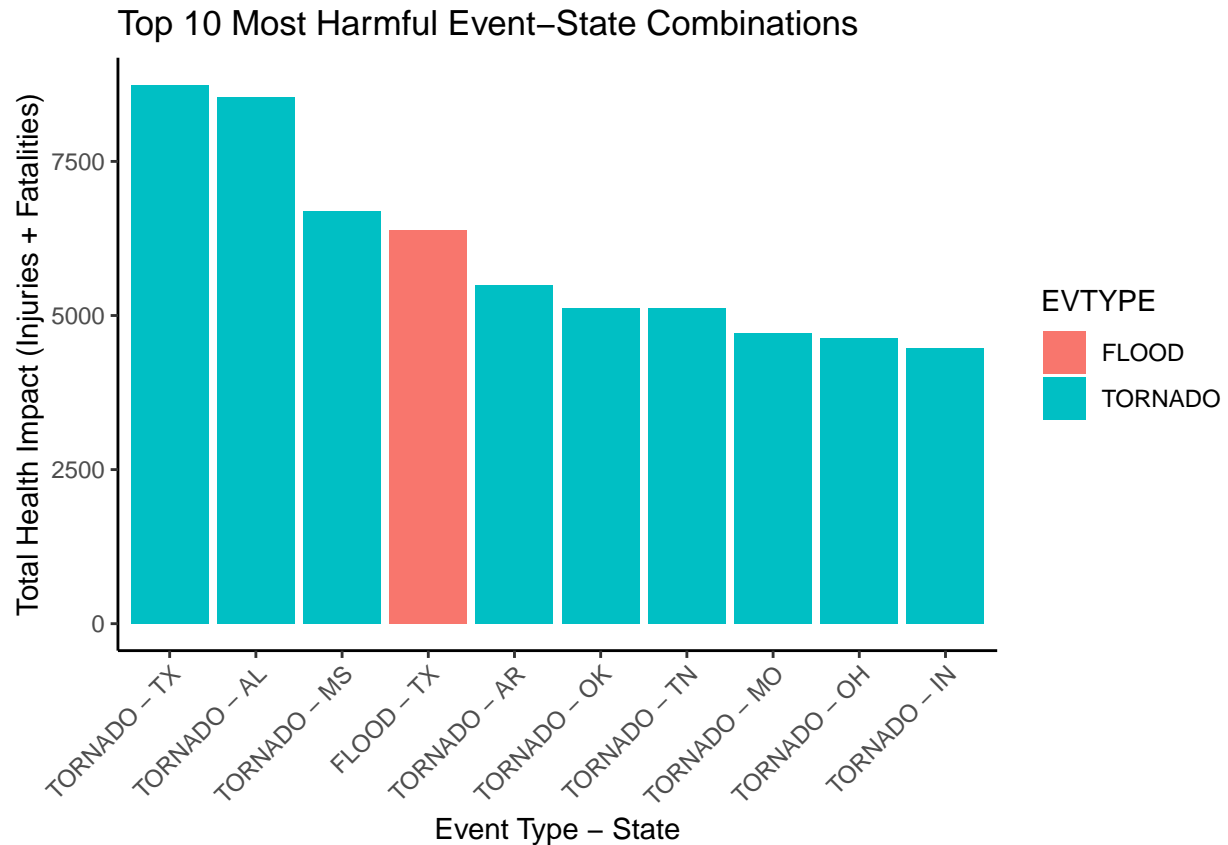

```
top10_event_state_deaths <- event_summary_state %>%
  arrange(desc(total_fatalities)) %>%
  slice_head(n = 10)


health_plot_deaths <- ggplot(top10_event_state_deaths, aes(x = reorder(paste(EVTYPE, STATE, sep = " - "
                                    -total_fatalities),
                          y = total_fatalities, fill = EVTYPE)) +
  geom_col() +
  labs(title = "Top 10 Most Fatal Event-State Combinations",
       x = "Event Type - State", y = "Total Fatality Impact (Fatalities)") +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(health_plot_deaths)
```

Top 10 Most Fatal Event–State Combinations

Exploring Econ Impact - Property damage and Crop damage

```r
# Note in will have to rewrap results chunk with caption too
# Similarly doing this for total_econ_impact
top10_event_state_econ <- event_summary_state %>%
  arrange(desc(total_econ_impact)) %>%
  slice_head(n = 10)

econ_plot <- ggplot(top10_event_state_econ, aes(x = reorder(paste(EVTYPE, STATE, sep = " - "),
                                      -total_econ_impact),
                              y = total_econ_impact,
                              fill = EVTYPE)) +
  geom_col() +
  labs(title = "Top 10 Most Harmful Event-State Combinations For Economy",
       x = "Event Type - State", y = "Total Economic Impact (Property + Crop)",) +
  theme_classic() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(econ_plot)
```

Top 10 Most Harmful Event–State Combinations For Economy
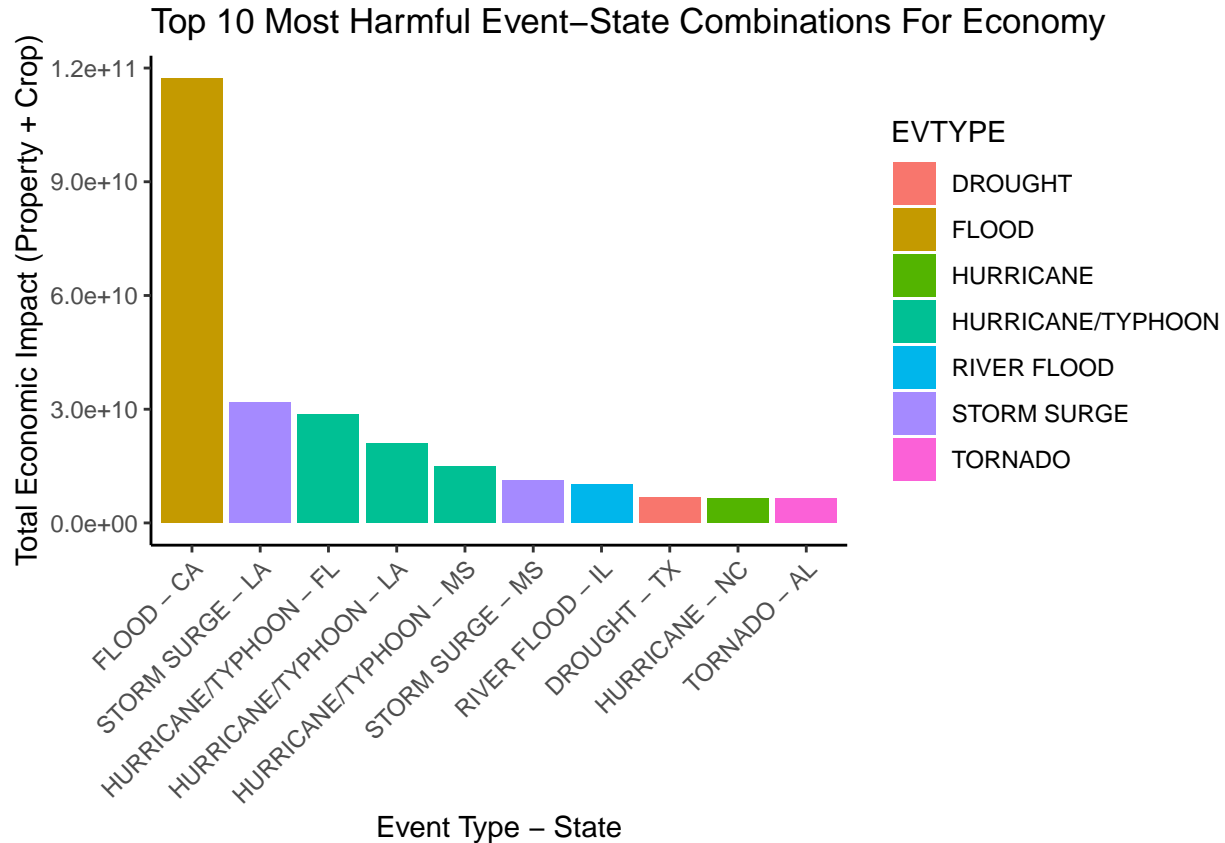
Results

Introduction

The weather events can cause two distinct types of damage that of property and crops which falls under economic impact and the other category is human health which relates to deaths and injuries. This fundamentally seems like it may cause a conflict of interest with policy makers as the main incentives to be re-elected could be from measures of economic growth, whereas human life should be viewed as the most important thing to uphold from an ethical viewpoint.

Economic - Property + Crop Damage

By far the most impactful event for the economy is floods and this effects CA (California), this is more than double the next which is storm surge for LA and then closely is hurricane/typhoon in FL(Florida), LA(Louisiana) and MS(Mississippi), then river flood in IL(Illinois), lastly of about equal total economic impact is Drought in Texas, Hurricane in North Carolina and then Tornado in Alabama.

This comes as a surprise and it might be overly emphasized depending on how weather events are classified, i.e, are floods considered as a separate event after a hurricane/typhoon, because then this economic impact would be overstated in a sense and the measures taken to protect against a flood might not work with the aftermath of a hurricane. I believe it's very important to understand what the data might be conveying in the real world context because this context is vital in understanding how to address such issues. In general it seems like hurricanes are the most devastating event across the entire country, but it seems that California in particular has had massive devastation in the past with floods. Although, I bet more recent data would show a much worse impact for wildfires in California as per recent events, being the LA wildfires. Looking at the latest begin date for the data set hows that we are only analyzing data up until 2011-11-30. So this analysis
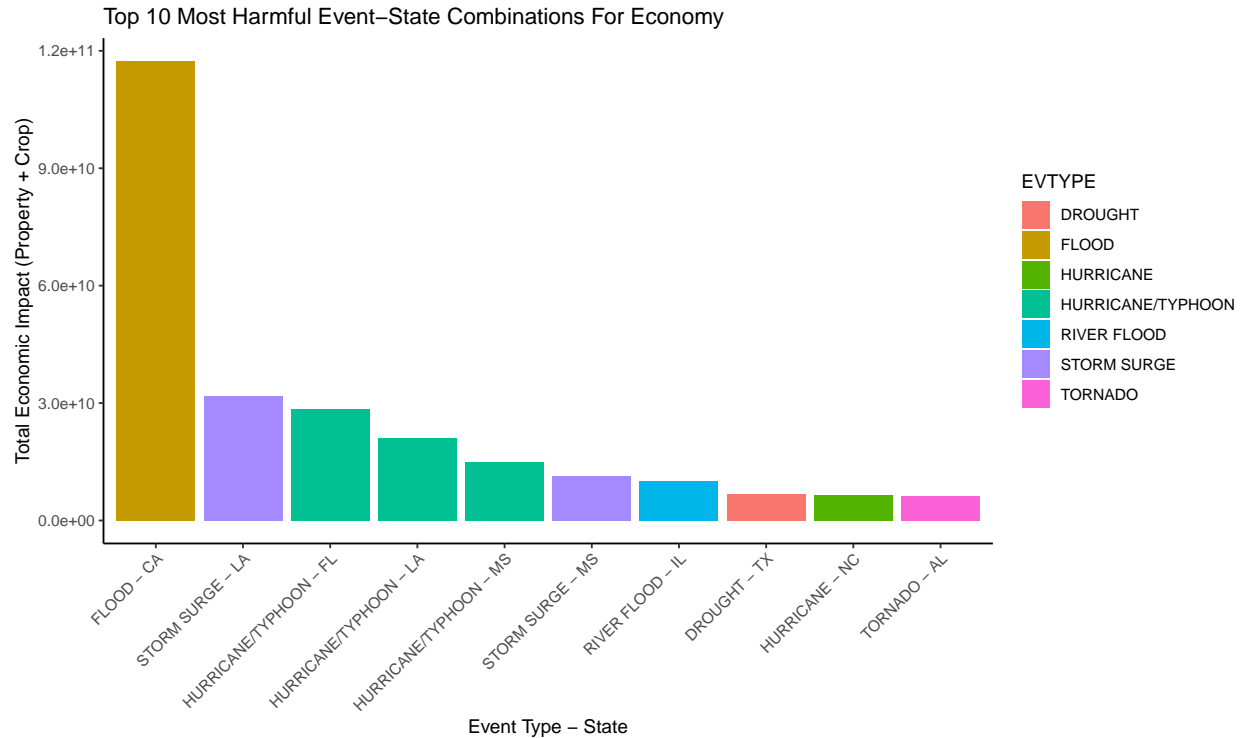
Figure 1: Similarly top 10 events capped providing state for which events toom place (this is summarized data).

may not be as relevant to today's most pertinent issues surrounding weather events.

According to this data it might be reasonable to suggest that water based events like hurricanes and floods cause the most economic damage this may be two-fold, because they are the most common events that cannot be controlled as easily - a wildfire under some circumstances can be contained by firefighters, but the most horrific hurricanes will simply destroy most places that are within its path as well as after effects left in its wake. Technology is not advanced enough to be able to contain a the most devastating of hurricanes but early warning systems and preventative measures for after effects may help reduce the economic impact - albeit it makes more sense that human loss of life and injuries are more easily avoided through the use of early warning systems as crops cannot be relocated nor can buildings be dynamically moved.

Health - Injuries + Fatalities

Firstly, look at the most fatal events which show loss of human life.

Looking at the most fatal events, these events are heat in Illinois and Tornadoes in Alabama, Texas, Mississippi, Arkansas, then with excessive heat in Pennsylvania and Illinois with CK(Not US State Maybe Cook Islands?) being 10th with TOrnadoes. Overall this shows Tornadoes and heat as causing the most fatalities. Unexpectedly a colder state being Pensylvannia ranked in top 10 with a heat related event which may suggest that a lot of these fatalities could have been incurred from a lack of preparedness on the part of the state. I.e, these had unavoidable deaths.

The most harmful events to human health in absolute totality summing up both injuries and fatalities are mostly from tornadoes, with the exception of Texas being ranked fourth by floods. Texas is also ranked 1st most with tornadoes dealing most devastation to human life. Then the rest of the Tornadoes occur in Mississippi, Arkansas, CK? etc. I wonder if these are just because they share a geographic locatedness to a predisposition to tornadoes forming, if there are certain characteristics such as high population densities, lack
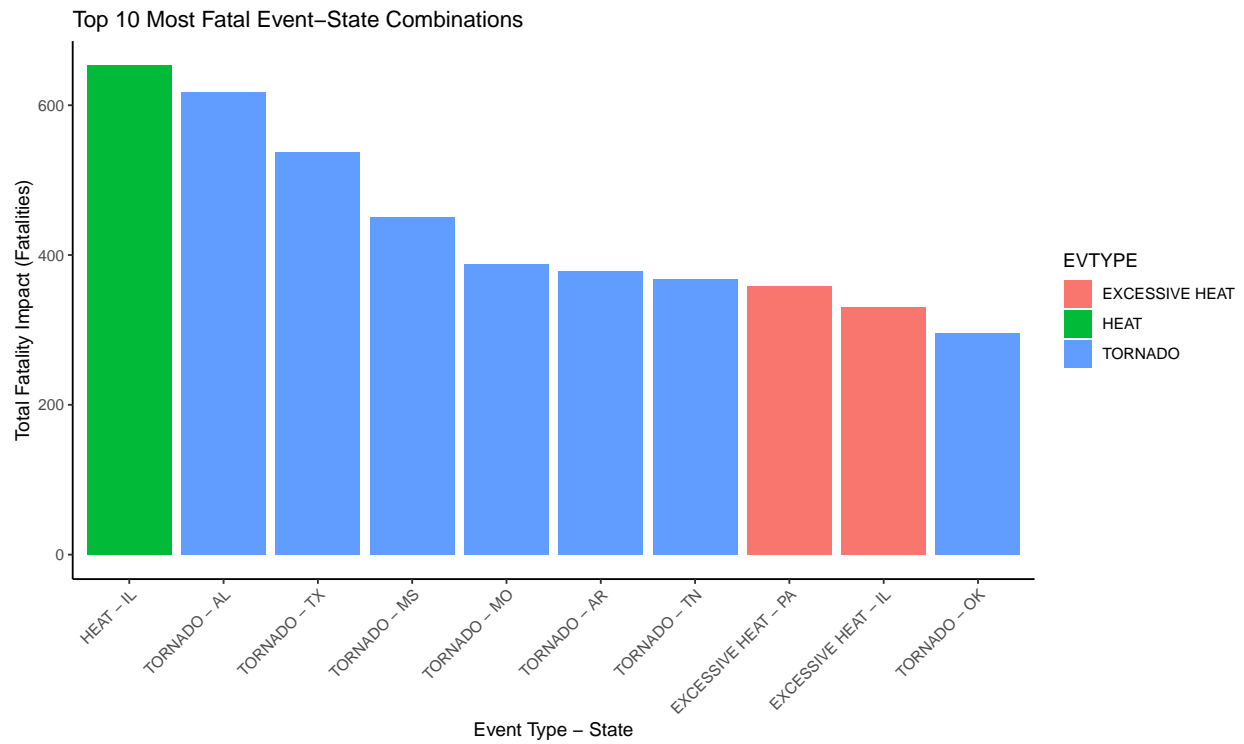
Figure 2: Alarmingly shows which weather events are deadliest including which state it occurs within, this is because including state may shed light on actions to take, for example developing measures to combat heatstroke in a colder climate up north in the United States wouldn't make too much sense. It is a very large country, hence including the state is pertinent information.
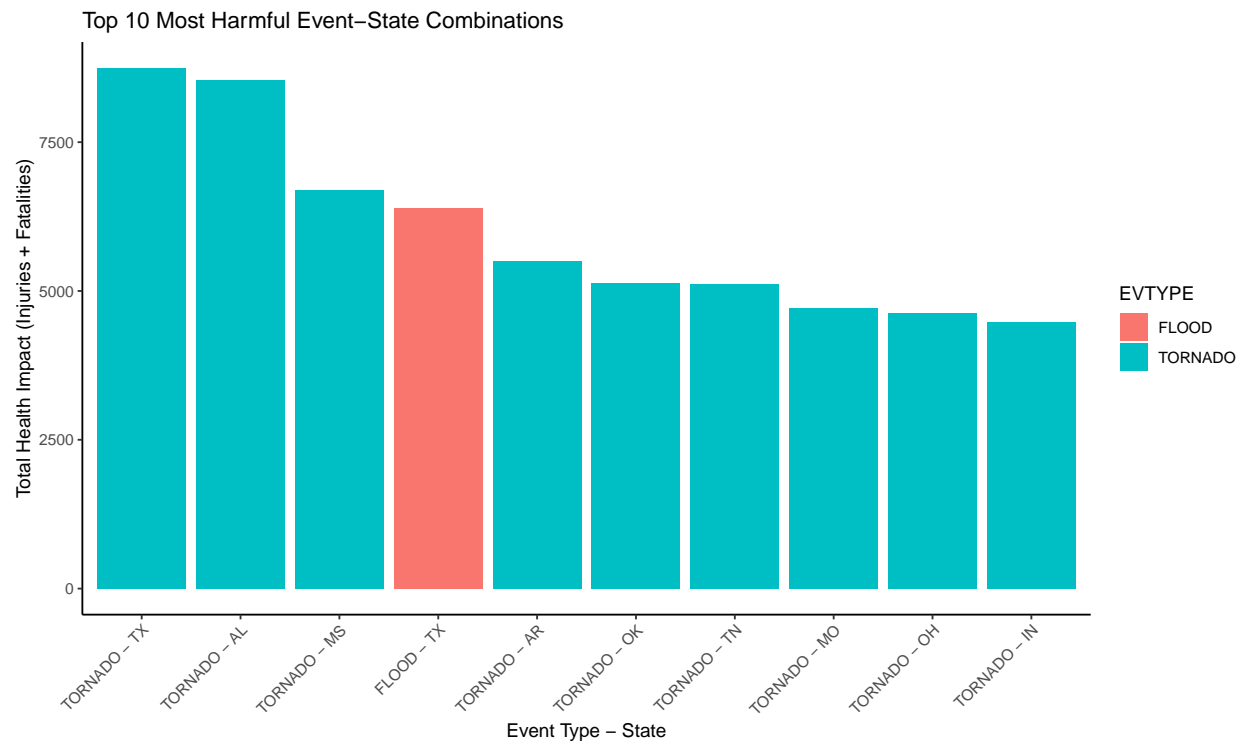
Figure 3: This takes the top 10 most harmful events but includes states to see how weather effects United States across the country, note that this is hard capped to only 10 instead of having 10 for each as this is more interpretable.

of early warning systems, or if these tornadoes are more likely to form during the night which would cause less alertness from the population etc. A deeper investigation into these issues should be performed to see why so many people are being effected by this event.

Conclusion

Overall there is a clear separation of which weather events impact the economy the most and which weather events are more detrimental to human life. This may cause a conflict of interest as policy makers now potentially have to balance measures to either prepare for economically impactful events or prioritize human life. Recommendations for future research should be to see which measures can protect both human life and defend against economic hardships if possible. This is no easy task, but innovative solutions may possibly come forth which may help against both. One such measure to me would be to have a very advanced warning system, such a system should be able to determine the formation or likely formation of an amalgamation of events such as excessive heat or tornadoes, floods as these events may have common properties which can be measured early on to prepare civilians and businesses against such events.