Exploration of National Climatic Data Center Storm Events Data

JL

02/09/2021

Synopsis

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

Here we take the combined data from 1950 through to November 2011 and explore the most destructive weather events in terms of both public harm and economic harm. Also, we explore the same considerations on a State-by-state basis. Three visuals are produced to convey these findings. The data and code to produce the same visuals and conclusions is included in this report.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(ggthemes)
library(stringi)
```

Loading and Processing the Raw Data

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This 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.

```
# Download data
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2", "rawdata.csv.

# Download documentation for data
```

```r
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf", "rawdata

# import data into dataframe
raw <- read.csv("rawdata.csv.bz2")
```

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

The Data

The dataframe `raw` has 37 columns and 900,000+ observations across the united states. Here is a brief overview.

```r
str(raw)
```

```
## '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 ...
```

Information that we will explore are

- Location: STATE__, COUNTY, COUNTYNAME, STATE
- Time: BGN_DATE, BGN_TIME
- Event Information: EVTYPE, LENGTH, WIDTH, F, REMARKS
- Damage: FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP

For LENGTH and WIDTH, this is the path length (in miles and tenths of miles) and maximum path width (in yards) for all tornadoes.

For F, The "Saffir-Simpson Hurricane and Tropical Cyclone Scale" is used

1. Windspeed 64-82 kts (74-95 mph), storm tide: 4-5 FT, Damage: Minor

2. Windspeed 83-95 kts (96-110 mph), storm tide: 6-8 FT, Damage: Moderate

3. Windspeed 96-113 kts (111-130 mph), storm tide: 9-12 FT, Damage: Major

4. Windspeed 114-135 kts (131-155 mph), storm tide: 13-18 FT, Damage: Severe

5. Windspeed >135 kts (>155 mph), storm tide: >18 FT, Damage: Catastrophic

For REMARKS, this is a description of the event.

For PROPDMGEXP and CROPDMGEXP, characters are used to signify cost of damage and include "K" for thousands, "M" for millions, and "B" for billions.

Processing

To process the data for analysis, we will select only the variables that we have outlined above. Further, we will rename these variables to more accessible versions and ensure they are cast into the correct type for calculations.

Secondly, we will create a dataframe for locations, containing both state and county ids with their respective state and county names, for reference.

```
# collect location IDs for reference and remove duplicate rows
counties <- select(raw, "STATE__", "STATE", "COUNTY", "COUNTYNAME") %>%
  distinct(.keep_all = TRUE)

colnames(counties) <- c("state_id", "state_name", "county_id", "county_name")

states <- select(raw, "STATE__", "STATE") %>%
  distinct(.keep_all = TRUE)

colnames(states) <- c("state_id", "state_name")

# states contains duplicates, over separate IDs.
# Manually remove against a list of abbreviations
states <- states[-c(79, 59, 63, 57, 56, 55, 95, 94, 93, 65, 70,
```

```
                     60, 68, 61,  66, 73, 71, 72, 76, 74, 62,
                     67, 78, 77, 51, 69, 75, 53, 64, 58, 20, 8), ]


# trim raw to chosen variables
clean <- select(raw, c("STATE__", "COUNTY", "BGN_DATE", "BGN_TIME", "EVTYPE", "LENGTH", "WIDTH", "F", "

# rename columns
colnames(clean) <- c("state_id", "county_id", "start_date", "start_time", "event", "length", "width", "

# cast raw into correct types and formats
clean$state_id <- as.integer(clean$state_id)
clean$county_id <- as.integer(clean$county_id)

clean$event <- stri_trans_totitle(as.factor(clean$event))

clean$F <- as.factor(clean$F)

clean$start_date <- as.Date(as.character(
  strptime(clean$start_date, format = "%m/%d/%Y")))

# We want to include damage_category into damage counts.
# first swap K, M, B with 1e3, 1e6 and 1e9 respectively

clean$property_damage_cat[clean$property_damage_cat == ""] <- 1
clean$property_damage_cat[clean$property_damage_cat == "K"] <- 1e3
clean$property_damage_cat[clean$property_damage_cat == "M"] <- 1e6
clean$property_damage_cat[clean$property_damage_cat == "B"] <- 1e9

clean$crop_damage_cat[clean$crop_damage_cat == ""] <- 1
clean$crop_damage_cat[clean$crop_damage_cat == "K"] <- 1e3
clean$crop_damage_cat[clean$crop_damage_cat == "M"] <- 1e6
clean$crop_damage_cat[clean$crop_damage_cat == "B"] <- 1e9

# now multiply to get correct damage values

clean$property_damage <- as.numeric(clean$property_damage) * as.numeric(clean$property_damage_cat)
```

```
## Warning: NAs introduced by coercion
```

```
clean$crop_damage <- as.numeric(clean$crop_damage) * as.numeric(clean$crop_damage_cat)
```

```
## Warning: NAs introduced by coercion
```

Results

Most harmful events to people's health

This first question we want to investigate is the following

- Across the United States, which types of events are most harmful with respect to population health?

There are a few ways to interpret this. Direct harm, is easily fatalities and injuries, however indirectly we could have the economic impact such as property damage, or crop damage. So we will limit this to the direct harm.

We will group by event and summarise the data

```
direct_harm <- clean %>%
  group_by(event) %>%
  summarise(number_of_events = n(),
            total_fatalities = sum(fatalities),
            total_injuries = sum(injuries),
            total_sum = sum(fatalities) + sum(injuries)) %>%
  arrange(desc(total_sum))

head(direct_harm)
```

```
## # A tibble: 6 x 5
##   event          number_of_events total_fatalities total_injuries total_sum
##   <chr>                     <int>            <dbl>          <dbl>     <dbl>
## 1 Tornado                   60652             5633          91346     96979
## 2 Excessive Heat             1678             1903           6525      8428
## 3 Tstm Wind                219942              504           6957      7461
## 4 Flood                     25327              470           6789      7259
## 5 Lightning                 15754              816           5230      6046
## 6 Heat                        767              937           2100      3037
```

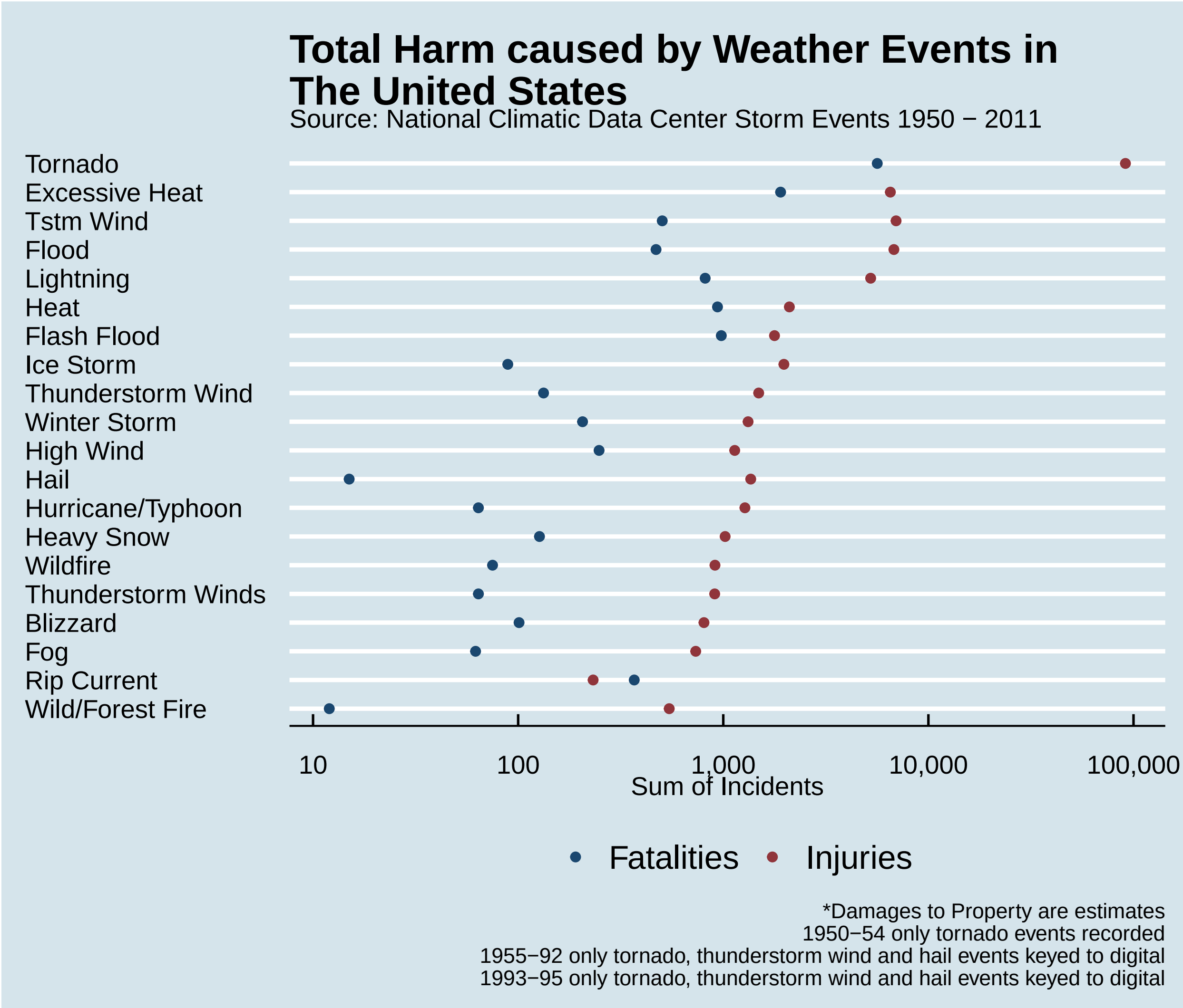
From the summary above, we can see that Tornados have the most fatalities and injuries associated with them. We can also see that the most frequent event is Thunderstorm Wind (Tstm Wind).

To build a visual, we will adjust the dataframe. Trim to just the top 20 contributing events, sorted by sum of fatalities and injuries. Then collapse fatalities and injuries into one column, but associating a label with each, for ggplot to interpret.

```
# trim a copy of dataframe to top20, select key columns
dh_short <- direct_harm[1:20, ]
dh_short2 <- dh_short

# rename two columns to merge on
colnames(dh_short)[3] <- "Incidents"
colnames(dh_short2)[4] <- "Incidents"

# select only the columns we need for a merge
dh_short <- select(dh_short, c("event", "total_sum", "Incidents"))
dh_short2 <- select(dh_short2, c("event", "total_sum", "Incidents"))

# create a label to split data on plotting
dh_short$label <- "Fatalities"
dh_short2$label <- "Injuries"

# merge dataframes, remove the temprorary second
dh_short <- rbind(dh_short, dh_short2)

rm(dh_short2)
rm(direct_harm)
```

Now we have the data we need for plotting an informational visual in a way that allows us to create it easily.

```
# plot for the top 20 by total of fatalities and injuries
ggplot(dh_short, aes(x = Incidents,
                     y = reorder(event, total_sum),
                     colour = label),
                     size = 3) +

  theme_economist() +

  scale_colour_stata() +

  geom_point() +

  labs(x = "Sum of Incidents",
       y = NULL,
       colour = NULL,
       title = "Total Harm caused by Weather Events in\nThe United States",
       subtitle = "Source: National Climatic Data Center Storm Events 1950 - 2011",
       caption = "*Damages to Property are estimates\n1950-54 only tornado events recorded\n 1955-92 onl

  guides(fill = guide_legend(title = NULL)) +

  theme(
    legend.position = "bottom") +

  scale_x_log10(breaks = c(1e1, 1e2, 1e3, 1e4, 1e5),
                labels = c("10", "100", "1,000", "10,000", "100,000"))
```



As you can see from the graph above, for the top 20 events, the number of injuries surpasses the number of fatalities. Which is to expected. The exception is Rip Current, which understandably is very dangerous once one has been dragged into one. The most destructive event by far in the U.S. is Tornado, with both regards to total fatalities and total injuries. The latter by a stunning degree. Tornadoes are highly destructive events and they frequent the country. At the bottom of the top 20 events is Wild/Forest Fire, which, alike Hail, show that incidents are much less likely to be fatal, however could still be quite serious.

Most harmful events with respect to the economy

This second question we want to investigate is the following

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

In this regard, we will investigate the effects of these weather events on damage to Property and damage to Crops. To do so we will group by events and take the sum of property damage and the sum of crop damage, then sort by the total of the two.

```r
economic_harm <- clean %>%
  group_by(event) %>%
  summarise(number_of_events = n(),
            property_damage = sum(property_damage),
            crop_damage = sum(crop_damage),
            total_damage = sum(property_damage) + sum(crop_damage)) %>%
  arrange(desc(total_damage))

head(economic_harm)
```

```
## # A tibble: 6 x 5
##   event             number_of_events property_damage crop_damage total_damage
##   <chr>                        <int>           <dbl>       <dbl>        <dbl>
## 1 Flood                        25327    144657709807  5661968450 150319678257
## 2 Hurricane/Typhoon               88     69305840000  2607872800  71913712800
## 3 Storm Surge                    261     43323536000        5000  43323541000
## 4 Drought                       2488      1046106000 13972566000  15018672000
## 5 Hurricane                      174     11868319010  2741910000  14610229010
## 6 River Flood                    173      5118945500  5029459000  10148404500
```

From the summary above we can see that Flood causes the most expensive damage, over twice the next event, Hurricane/Typhoon. The event most damaging to crops is Drought. Especially noteworthy as it causes relatively low property damage. Floods are second in crop damage, but as already mentioned, deal significant property damage also.

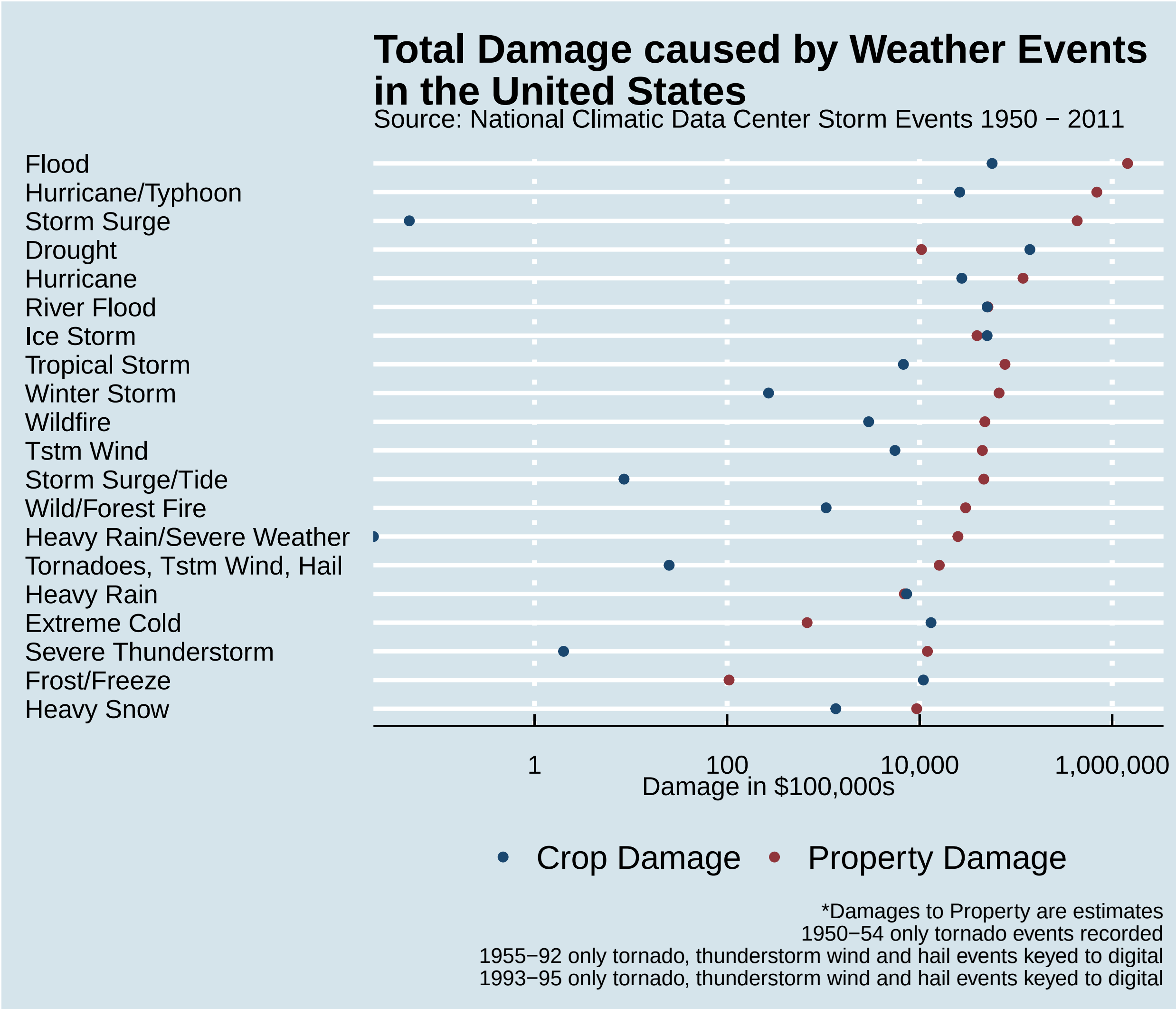
To build a visual, we will adjust the dataframe. Trim to just the top 20 contributing events, sorted by sum of damages. Then collapse property damages and crop damages into one column, but associating a label with each, for ggplot to interpret.

```r
# trim a copy of dataframe to top20, select key columns
eco_short <- economic_harm[1:20, ]
eco_short2 <- eco_short

# rename two columns to merge on
colnames(eco_short)[3] <- "Incidents"
colnames(eco_short2)[4] <- "Incidents"

# select only the columns we need for a merge
eco_short <- select(eco_short, c("event", "total_damage", "Incidents"))
eco_short2 <- select(eco_short2, c("event", "total_damage", "Incidents"))

# create a label to split data on plotting
eco_short$label <- "Property Damage"
eco_short2$label <- "Crop Damage"

# merge dataframes, remove the temprorary second
eco_short <- rbind(eco_short, eco_short2)

# rescale damage to thousands of dollars
eco_short$total_damage <- eco_short$total_damage

rm(eco_short2)
rm(economic_harm)
```

Now we have the data we need for plotting an informational visual in a way that allows us to create it easily.

```
# plot for the top 20 by total of fatalities and injuries
ggplot(eco_short, aes(x = Incidents,
                 y = reorder(event, total_damage),
                 colour = label),
                 size = 3) +

  theme_economist() +

  scale_colour_stata() +

  geom_point() +

  labs(x = "Damage in $100,000s",
       y = NULL,
       colour = NULL,
       title = "Total Damage caused by Weather Events\nin the United States",
       subtitle = "Source: National Climatic Data Center Storm Events 1950 - 2011",
       caption = "*Damages to Property are estimates\n1950-54 only tornado events recorded\n 1955-92 onl

  guides(fill = guide_legend(title = NULL)) +

  theme(legend.position = "bottom",
        panel.grid.major.x = element_line(colour = "white",
                                          size = 1,
                                          linetype = "dotted")) +

  scale_x_log10(breaks = c(1e5, 1e7, 1e9, 1e11),
                labels = c("1", "100", "10,000", "1,000,000"))
```

```
## Warning: Transformation introduced infinite values in continuous x-axis
```



As mentioned, Floods cause the most damage to property, but summed with River Flood, these together overtake Drought as most damaging for crops, and both are likely to occur in tandem. Storm Surge is another event to take note of. It is third in property damage yet causes relatively small crop damage. Understandably so as it would almost solely affects electronics. Events that cause more crop damage than property damage include Frost/Freeze, Extreme Cold, Heavy Rain, Ice Storm, and Drought. Concluding that temperature affects primarily crops over property, though still result in many damages.

State that suffers the most

As we have the data that allows us to investigate down to the state level, we will. We will use the same steps as before, yet grouping by state and aggregating for each of what we have explored already. We'll sort by the number of events in this case.

```
state_harm <- clean %>%
  group_by(state_id) %>%
  summarise(number_of_events = n(),
            damage = (sum(property_damage, na.rm=TRUE) + sum(crop_damage, na.rm=TRUE))/1e4,
            fatalities = sum(fatalities),
            injuries = sum(injuries)) %>%
```

```r
  arrange(desc(number_of_events))

state_harm <- merge(state_harm, states)

head(state_harm)
```

```
##   state_id number_of_events      damage fatalities injuries state_name
## 1        1            22739  1781823.11        784     8742         AL
## 2        2             4390    29790.63         74      112         AK
## 3        4             6156   396053.07        208      968         AZ
## 4        5            27102   455709.64        530     5550         AR
## 5        6            10780 12711585.94        550     3278         CA
## 6        8            20473   288994.99        163     1004         CO
```

From the summary above we can see that there is a great range of values across the United States. Undoubtedly, size of state and population are strong factors in here, However, it is not unreasonable to assume that both location and government management of prevention and preparation play a significant role also.

To build a visual, we will adjust the dataframe. Collapse number of events, damages, fatalities and injuries into one column, but associating a label with each, for ggplot to interpret.

```r
# copies of dataframe to merge
state_n <- state_harm
state_p <- state_harm
state_f <- state_harm
state_i <- state_harm

# rename two columns to merge on
state_n$values <- state_n$number_of_events
colnames(state_p)[3] <- "values"
colnames(state_f)[4] <- "values"
colnames(state_i)[5] <- "values"

# create a label to split data on plotting
state_n$label <- "Number of Events"
state_p$label <- "Damages $10,000s"
state_f$label <- "Fatalities"
state_i$label <- "Injuries"

# select only the columns we need for a merge
state_n <- state_n %>% select(c("state_name", "values", "label", "number_of_events"))
state_p <- state_p %>% select(c("state_name", "values", "label", "number_of_events"))
state_f <- state_f %>% select(c("state_name", "values", "label", "number_of_events"))
state_i <- state_i %>% select(c("state_name", "values", "label", "number_of_events"))


# merge dataframes, remove the temprorary dataframes
state_n <- rbind(state_n, state_p)
state_n <- rbind(state_n, state_f)
state_n <- rbind(state_n, state_i)

state_harm <- state_n
```

```r
# sorting
arrange(state_harm, desc(number_of_events));
```

```
##      state_name     values            label number_of_events
## 1            TX   83728.00 Number of Events            83728
## 2            TX 3394243.80 Damages $10,000s            83728
## 3            TX    1366.00       Fatalities            83728
## 4            TX   17667.00         Injuries            83728
## 5            KS   53441.00 Number of Events            53441
## 6            KS  505472.03 Damages $10,000s            53441
## 7            KS     356.00       Fatalities            53441
## 8            KS    3449.00         Injuries            53441
## 9            OK   46802.00 Number of Events            46802
## 10           OK  671968.71 Damages $10,000s            46802
## 11           OK     458.00       Fatalities            46802
## 12           OK    5710.00         Injuries            46802
## 13           MO   35648.00 Number of Events            35648
## 14           MO  793182.72 Damages $10,000s            35648
## 15           MO     754.00       Fatalities            35648
## 16           MO    8998.00         Injuries            35648
## 17           IA   31069.00 Number of Events            31069
## 18           IA 1018659.20 Damages $10,000s            31069
## 19           IA     140.00       Fatalities            31069
## 20           IA    2892.00         Injuries            31069
## 21           NE   30271.00 Number of Events            30271
## 22           NE  529499.04 Damages $10,000s            30271
## 23           NE     102.00       Fatalities            30271
## 24           NE    1471.00         Injuries            30271
## 25           IL   28488.00 Number of Events            28488
## 26           IL 1408744.26 Damages $10,000s            28488
## 27           IL    1421.00       Fatalities            28488
## 28           IL    5563.00         Injuries            28488
## 29           AR   27102.00 Number of Events            27102
## 30           AR  455709.64 Damages $10,000s            27102
## 31           AR     530.00       Fatalities            27102
## 32           AR    5550.00         Injuries            27102
## 33           NC   25351.00 Number of Events            25351
## 34           NC 1028355.78 Damages $10,000s            25351
## 35           NC     398.00       Fatalities            25351
## 36           NC    3415.00         Injuries            25351
## 37           GA   25259.00 Number of Events            25259
## 38           GA  601749.74 Damages $10,000s            25259
## 39           GA     327.00       Fatalities            25259
## 40           GA    5061.00         Injuries            25259
## 41           OH   24923.00 Number of Events            24923
## 42           OH  725030.57 Damages $10,000s            24923
## 43           OH     403.00       Fatalities            24923
## 44           OH    7112.00         Injuries            24923
## 45           MN   23609.00 Number of Events            23609
## 46           MN  570154.59 Damages $10,000s            23609
## 47           MN     168.00       Fatalities            23609
## 48           MN    2282.00         Injuries            23609
## 49           AL   22739.00 Number of Events            22739
```

```
## 50          AL  1781823.11 Damages $10,000s            22739
## 51          AL      784.00       Fatalities            22739
## 52          AL     8742.00         Injuries            22739
## 53          PA    22226.00 Number of Events            22226
## 54          PA   542343.75 Damages $10,000s            22226
## 55          PA      846.00       Fatalities            22226
## 56          PA     3223.00         Injuries            22226
## 57          MS    22192.00 Number of Events            22192
## 58          MS  3641893.98 Damages $10,000s            22192
## 59          MS      555.00       Fatalities            22192
## 60          MS     6675.00         Injuries            22192
## 61          FL    22124.00 Number of Events            22124
## 62          FL  4541296.97 Damages $10,000s            22124
## 63          FL      746.00       Fatalities            22124
## 64          FL     5918.00         Injuries            22124
## 65          KY    22092.00 Number of Events            22092
## 66          KY   303838.66 Damages $10,000s            22092
## 67          KY      239.00       Fatalities            22092
## 68          KY     3480.00         Injuries            22092
## 69          SD    21728.00 Number of Events            21728
## 70          SD    85211.99 Damages $10,000s            21728
## 71          SD       61.00       Fatalities            21728
## 72          SD      868.00         Injuries            21728
## 73          TN    21721.00 Number of Events            21721
## 74          TN   658304.47 Damages $10,000s            21721
## 75          TN      521.00       Fatalities            21721
## 76          TN     5202.00         Injuries            21721
## 77          IN    21506.00 Number of Events            21506
## 78          IN   489051.50 Damages $10,000s            21506
## 79          IN      391.00       Fatalities            21506
## 80          IN     4720.00         Injuries            21506
## 81          VA    21189.00 Number of Events            21189
## 82          VA   253184.77 Damages $10,000s            21189
## 83          VA      174.00       Fatalities            21189
## 84          VA     1703.00         Injuries            21189
## 85          NY    21058.00 Number of Events            21058
## 86          NY   497183.12 Damages $10,000s            21058
## 87          NY      342.00       Fatalities            21058
## 88          NY     1340.00         Injuries            21058
## 89          CO    20473.00 Number of Events            20473
## 90          CO   288994.99 Damages $10,000s            20473
## 91          CO      163.00       Fatalities            20473
## 92          CO     1004.00         Injuries            20473
## 93          WI    19781.00 Number of Events            19781
## 94          WI   420268.59 Damages $10,000s            19781
## 95          WI      279.00       Fatalities            19781
## 96          WI     2309.00         Injuries            19781
## 97          MI    17911.00 Number of Events            17911
## 98          MI   268553.20 Damages $10,000s            17911
## 99          MI      398.00       Fatalities            17911
## 100         MI     4586.00         Injuries            17911
## 101         LA    17323.00 Number of Events            17323
## 102         LA  6130171.17 Damages $10,000s            17323
## 103         LA      310.00       Fatalities            17323
```

```
## 104          LA     3215.00         Injuries            17323
## 105          SC    17125.00 Number of Events            17125
## 106          SC   125843.89 Damages $10,000s            17125
## 107          SC      221.00       Fatalities            17125
## 108          SC     1786.00         Injuries            17125
## 109          MT    14695.00 Number of Events            14695
## 110          MT    40526.99 Damages $10,000s            14695
## 111          MT       58.00       Fatalities            14695
## 112          MT      181.00         Injuries            14695
## 113          ND    14630.00 Number of Events            14630
## 114          ND   586589.07 Damages $10,000s            14630
## 115          ND       69.00       Fatalities            14630
## 116          ND      608.00         Injuries            14630
## 117          CA    10780.00 Number of Events            10780
## 118          CA 12711585.94 Damages $10,000s            10780
## 119          CA      550.00       Fatalities            10780
## 120          CA     3278.00         Injuries            10780
## 121          WV     9099.00 Number of Events             9099
## 122          WV   102659.00 Damages $10,000s             9099
## 123          WV       92.00       Fatalities             9099
## 124          WV      363.00         Injuries             9099
## 125          NJ     8074.00 Number of Events             8074
## 126          NJ   329519.12 Damages $10,000s             8074
## 127          NJ      180.00       Fatalities             8074
## 128          NJ     1152.00         Injuries             8074
## 129          WY     7332.00 Number of Events             7332
## 130          WY    21212.06 Damages $10,000s             7332
## 131          WY       56.00       Fatalities             7332
## 132          WY      432.00         Injuries             7332
## 133          NM     7130.00 Number of Events             7130
## 134          NM   198061.24 Damages $10,000s             7130
## 135          NM       72.00       Fatalities             7130
## 136          NM      385.00         Injuries             7130
## 137          AZ     6156.00 Number of Events             6156
## 138          AZ   396053.07 Damages $10,000s             6156
## 139          AZ      208.00       Fatalities             6156
## 140          AZ      968.00         Injuries             6156
## 141          MA     5651.00 Number of Events             5651
## 142          MA   128015.54 Damages $10,000s             5651
## 143          MA      140.00       Fatalities             5651
## 144          MA     2121.00         Injuries             5651
## 145          OR     4821.00 Number of Events             4821
## 146          OR   106515.39 Damages $10,000s             4821
## 147          OR       75.00       Fatalities             4821
## 148          OR      225.00         Injuries             4821
## 149          ID     4767.00 Number of Events             4767
## 150          ID    26745.21 Damages $10,000s             4767
## 151          ID       58.00       Fatalities             4767
## 152          ID      273.00         Injuries             4767
## 153          ME     4524.00 Number of Events             4524
## 154          ME    56225.95 Damages $10,000s             4524
## 155          ME       25.00       Fatalities             4524
## 156          ME      177.00         Injuries             4524
## 157          AK     4390.00 Number of Events             4390
```

```
## 158         AK    29790.63 Damages $10,000s             4390
## 159         AK       74.00       Fatalities             4390
## 160         AK      112.00         Injuries             4390
## 161         UT     4135.00 Number of Events             4135
## 162         UT    79815.05 Damages $10,000s             4135
## 163         UT      136.00       Fatalities             4135
## 164         UT     1070.00         Injuries             4135
## 165         VT     3871.00 Number of Events             3871
## 166         VT   153810.83 Damages $10,000s             3871
## 167         VT       23.00       Fatalities             3871
## 168         VT       71.00         Injuries             3871
## 169         WA     3312.00 Number of Events             3312
## 170         WA   141276.83 Damages $10,000s             3312
## 171         WA      146.00       Fatalities             3312
## 172         WA      753.00         Injuries             3312
## 173         CT     3294.00 Number of Events             3294
## 174         CT    76157.12 Damages $10,000s             3294
## 175         CT       41.00       Fatalities             3294
## 176         CT      897.00         Injuries             3294
## 177         NV     3139.00 Number of Events             3139
## 178         NV    84076.91 Damages $10,000s             3139
## 179         NV      105.00       Fatalities             3139
## 180         NV      232.00         Injuries             3139
## 181         NH     3022.00 Number of Events             3022
## 182         NH    22041.88 Damages $10,000s             3022
## 183         NH       32.00       Fatalities             3022
## 184         NH      195.00         Injuries             3022
## 185         HI     2547.00 Number of Events             2547
## 186         HI    22016.44 Damages $10,000s             2547
## 187         HI       44.00       Fatalities             2547
## 188         HI       95.00         Injuries             2547
## 189         DE     1913.00 Number of Events             1913
## 190         DE    17902.89 Damages $10,000s             1913
## 191         DE       30.00       Fatalities             1913
## 192         DE      338.00         Injuries             1913
## 193         RI      839.00 Number of Events              839
## 194         RI    12064.60 Damages $10,000s              839
## 195         RI        7.00       Fatalities              839
## 196         RI       48.00         Injuries              839
## 197         MD      450.00 Number of Events              450
## 198         MD    15800.56 Damages $10,000s              450
## 199         MD       31.00       Fatalities              450
## 200         MD      392.00         Injuries              450
```

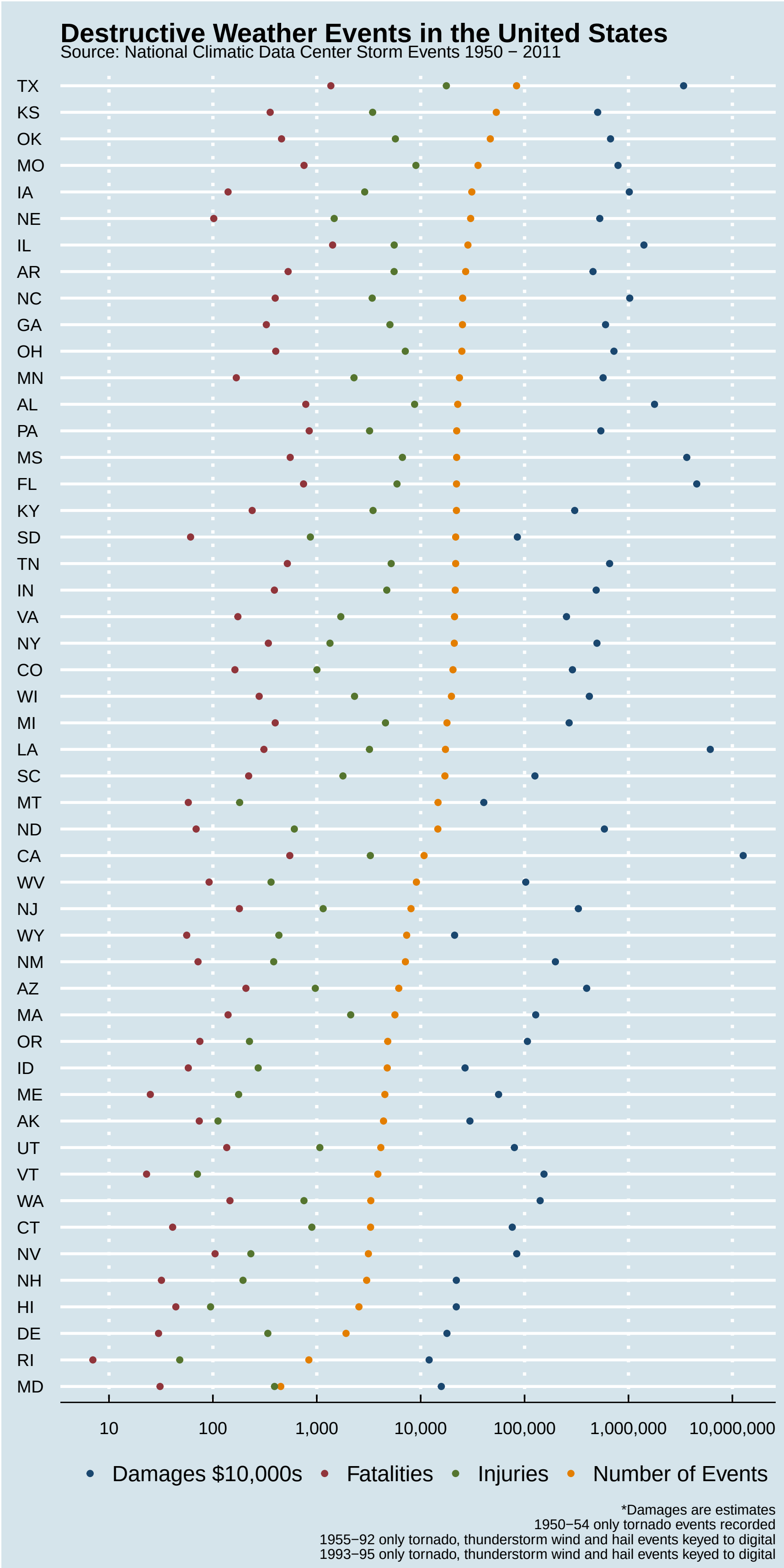
```
# remove temporary dataframes
rm(state_n, state_p, state_f, state_i)

head(state_harm)
```

```
##   state_name values            label number_of_events
## 1         AL  22739 Number of Events            22739
## 2         AK   4390 Number of Events             4390
## 3         AZ   6156 Number of Events             6156
## 4         AR  27102 Number of Events            27102
```

```
## 5         CA  10780 Number of Events           10780
## 6         CO  20473 Number of Events           20473
```

Now we have the data we need for plotting an informational visual in a way that allows us to create it easily.

```r
# plot for the top 20 by total of fatalities and injuries
ggplot(state_harm, aes(x = values,
                 y = reorder(state_name, number_of_events),
                 colour = label),
                 size = 3, shape = 22) +

  theme_economist() +

  scale_colour_stata() +

  geom_point() +

  labs(x = NULL,
       y = NULL,
       colour = NULL,
       title = "Destructive Weather Events in the United States",
       subtitle = "Source: National Climatic Data Center Storm Events 1950 - 2011",
       caption = "*Damages are estimates\n1950-54 only tornado events recorded\n 1955-92 only tornado, t

  guides(fill = guide_legend(title = NULL)) +

  scale_fill_discrete(labels = c("Fatalities",
                                 "Injuries",
                                 "Events",
                                 "Damages $10,000s")) +

  theme(
    legend.position = "bottom",
    panel.grid.major.x = element_line(colour = "white",
                                      size = 1,
                                      linetype = "dotted")) +

  scale_x_log10(breaks = c(1e1, 1e2, 1e3, 1e4, 1e5, 1e6, 1e7),
                labels = c("10", "100", "1,000","10,000", "100,000", "1,000,000", "10,000,000"))
```


Destructive Weather Events in the United States
Source: National Climatic Data Center Storm Events 1950 − 2011
TX
KS
OK
MO
IA
NE
IL
AR
NC
GA
OH
MN
AL
PA
MS
FL
KY
SD
TN
IN
VA
NY
CO
WI
MI
LA
SC
MT
ND
CA
WV
NJ
WY
NM
AZ
MA
OR
ID
ME
AK
UT
VT
WA
CT
NV
NH
HI
DE
RI
MD
10
100
1,000
10,000
100,000
1,000,000
10,000,000
Damages $10,000s
Fatalities
Injuries
Number of Events
*Damages are estimates
1950−54 only tornado events recorded
1955−92 only tornado, thunderstorm wind and hail events keyed to digital
1993−95 only tornado, thunderstorm wind and hail events keyed to digital

From this visual, we can see that a strong association between the number of events and all other variables. Which is completely expected. California has had the most damages overall, we know that several droughts have been very costly. However it is Illinois and Texas (in that order) where most fatalities have occurred due to weather. Illinois has a history of Tornados, notably the Tri-State Torndao. Texas has a very variable climate and also suffers Hurricanes and Tornados and other storms. Naturally, this contributes to the high level of injuries in Texas. The "safest" states then are Rhode Island and Maryland. Interestingly, for Maryland, it suggests that with almost as many injuries as number of events, that makes it almost one person injured for each event. Though, there are fewer fatalities and injuries due to weather in Rhode Island, making it the "safest state for bad weather".