An Analyis of Weather Data

Matthew Dunne

August 6, 2017

Synopsis

Using the NOAA Storm Database we are able to establish that, in the United States, Tornadoes are the most harmful with respect to population health and have the most pronounced economic consequences. The data analysis below lays out how tornados cause the most fatalities, injuries, and economic damage (both in property and crops).

Data Processing

The code for processing the data is as follows:

```
##set working directory to where data file is stored
setwd("C:/Users/mjdun/Desktop/Coursera/Data Science Specialization/Course 5 Reproducible Research/Cours
##load data.table package to use fread() function
library(data.table)
##read data into R
data<-fread("storm data.csv")
```

First create a subset of fatality data

```
harm_data<-data.frame(EVTYPE=data$EVTYPE, FATALITIES=data$FATALITIES, INJURIES=data$INJURIES)
fatal<-tapply(harm_data$FATALITIES, harm_data$EVTYPE, sum)
fatal<-data.frame(Event=names(fatal), Fatalities=fatal)
fatal<-subset(fatal, fatal$Fatalities>=1)
```

and then create a subset of injury data.

```
injuries<-tapply(harm_data$INJURIES, harm_data$EVTYPE, sum)
injuries<-data.frame(Event=names(injuries), Injuries=injuries)
injuries<-subset(injuries, injuries$Injuries>=1)
```

The event type entries in the original are written in a similar but not identical way. Therefore, we need to consolidate them so that event types are denoted the same way from entry to entry.

First do this with the fatality data.

```
fatal$Event<-as.character(fatal$Event)
fatal_events<-unique(fatal$Event)
avalanche<-"AVALANCE|AVALANCHE"
fatal$Event<-ifelse(grepl(avalanche, fatal$Event), "AVALANCHE", fatal$Event)
coastal_flood<-"Coastal Flooding|COASTAL FLOODING|COASTAL FLOOD"
fatal$Event<-ifelse(grepl(coastal_flood, fatal$Event), "COASTAL FLOOD", fatal$Event)
snow<-"blowing snow|BLOWING SNOW|COLD AND SNOW|FALLING SNOW/ICE|FREEZING RAIN/SNOW|HEAVY SNOW|HEAVY SNO
fatal$Event<-ifelse(grepl(snow, fatal$Event), "SNOW", fatal$Event)
coast_storm<-"COASTAL STORM|COASTALSTORM"
fatal$Event<-ifelse(grepl(coast_storm, fatal$Event), "COASTAL STORM", fatal$Event)
cold<-"Cold|COLD|Cold Temperature|COLD WAVECOLD WEATHER|COLD/WIND CHILL|COLD/WINDS|Extended Cold|Extrem
fatal$Event<-ifelse(grepl(cold, fatal$Event), "COLD", fatal$Event)
```

```
heat<-"DROUGHT/EXCESSIVE HEAT|EXCESSIVE HEAT|EXTREME HEAT|HEAT|HEAT WAVE|HEAT WAVE DROUGHT|HEAT WAVES|H
fatal$Event<-ifelse(grepl(heat, fatal$Event), "HEAT", fatal$Event)
flood<-"EXCESSIVE RAINFALL|FLASH FLOOD|FLASH FLOOD/FLOOD|FLASH FLOODING|FLASH FLOODING/FLOOD|FLASH FLOO
fatal$Event<-ifelse(grepl(flood, fatal$Event), "FLOOD", fatal$Event)
hurricane<-"HURRICANE|HURRICANE ERIN|HURRICANE FELIX|HURRICANE OPAL|HURRICANE OPAL/HIGH WINDS|HURRICANE/
fatal$Event<-ifelse(grepl(hurricane, fatal$Event), "HURRICANE", fatal$Event)
ice<-"FREEZING DRIZZLE|FREEZING RAIN|Freezing Spray|FROST|GLAZE|ICE|ICE ON ROAD|ICE STORM|ICY ROADS|WIN
fatal$Event<-ifelse(grepl(ice, fatal$Event), "ICE", fatal$Event)
thunderstorm<-"THUNDERSTORM|THUNDERSTORM WIND|THUNDERSTORM WIND (G40)|THUNDERSTORM WIND G52|THUNDERSTORM
fatal$Event<-ifelse(grepl(thunderstorm, fatal$Event), "THUNDERSTORM", fatal$Event)
surf<-"Heavy Surf|HEAVY SURF|Heavy surf and wind|HEAVY SURF/HIGH SURF|High Surf|HIGH SURF|ROUGH SURF|STR
fatal$Event<-ifelse(grepl(surf, fatal$Event), "HEAVY SURF", fatal$Event)
wind<-"GUSTY WIND|GUSTY WINDS|HIGH WINDS|RAIN/WIND|STRONG WIND|Whirlwind|WIND|WINDS|WIND STORM|Strong Wi
fatal$Event<-ifelse(grepl(wind, fatal$Event), "WIND", fatal$Event)
landslides<-"LANDSLIDES|LANDSLIDE|Mudslide|Mudslides"
fatal$Event<-ifelse(grepl(landslides, fatal$Event), "LANDSLIDE", fatal$Event)
wildfire<-"WILD FIRES|WILD/FOREST FIRE|WILDFIRE"
fatal$Event<-ifelse(grepl(wildfire, fatal$Event), "WILD FIRES", fatal$Event)
tornado<-"TORNADO|WATERSPOUT|WATERSPOUT/TORNADO|WATERSPOUT|DUST DEVIL"
fatal$Event<-ifelse(grepl(tornado, fatal$Event), "TORNADO", fatal$Event)
tropical_storm<-"TROPICAL STORM|TROPICAL STORM GORDON"
fatal$Event<-ifelse(grepl(tropical_storm, fatal$Event), "TROPICAL STORM", fatal$Event)
heavy_seas<-"HEAVY SEAS|HIGH SEAS|HIGH SWELLS|ROUGH SEAS"
fatal$Event<-ifelse(grepl(heavy_seas, fatal$Event), "HEAVY SEAS", fatal$Event)
```

then consolidate the injury data

```
injuries<-tapply(harm_data$INJURIES, harm_data$EVTYPE, sum)
injuries<-data.frame(Event=names(injuries), Injuries=injuries)
injuries<-subset(injuries, injuries$Injuries>=1)
injuries$Event<-as.character(injuries$Event)
injuries$Event<-ifelse(grepl(avalanche, injuries$Event), "AVALANCHE", injuries$Event)
injuries$Event<-ifelse(grepl(snow, injuries$Event), "SNOW", injuries$Event)
injuries$Event<-ifelse(grepl(cold, injuries$Event), "COLD", injuries$Event)
injuries$Event<-ifelse(grepl(heat, injuries$Event), "HEAT", injuries$Event)
injuries$Event<-ifelse(grepl(flood, injuries$Event), "FLOOD", injuries$Event)
injuries$Event<-ifelse(grepl(hurricane, injuries$Event), "HURRICANE", injuries$Event)
injuries$Event<-ifelse(grepl(ice, injuries$Event), "ICE", injuries$Event)
injuries$Event<-ifelse(grepl(thunderstorm, injuries$Event), "THUNDERSTORM", injuries$Event)
injuries$Event<-ifelse(grepl(surf, injuries$Event), "HEAVY SURF", injuries$Event)
injuries$Event<-ifelse(grepl(wind, injuries$Event), "WIND", injuries$Event)
injuries$Event<-ifelse(grepl(landslides, injuries$Event), "LANDSLIDE", injuries$Event)
injuries$Event<-ifelse(grepl(wildfire, injuries$Event), "WILD FIRES", injuries$Event)
injuries$Event<-ifelse(grepl(tornado, injuries$Event), "TORNADO", injuries$Event)
injuries$Event<-ifelse(grepl(tropical_storm, injuries$Event), "TROPICAL STORM", injuries$Event)
injuries$Event<-ifelse(grepl(heavy_seas, injuries$Event), "HEAVY SEAS", injuries$Event)
```

From this, we can calculate the number of fatalities by event type.

```
fatalities_by_type<-tapply(fatal$Fatalities, fatal$Event, sum)
fatalities_by_type<-data.frame(Event=names(fatalities_by_type), Fatalities=fatalities_by_type)
fatalities_by_type<-fatalities_by_type[with(fatalities_by_type, order(-Fatalities)), ]
```

as well as the number of injuries by event type.

```
injuries_by_type<-tapply(injuries$Injuries, injuries$Event, sum)
injuries_by_type<-data.frame(Event=names(injuries_by_type), Injuries=injuries_by_type)
injuries_by_type<-injuries_by_type[with(injuries_by_type, order(-Injuries)), ]
```

Now it is time to process the data on property and crop damage. Create data frames that subset to events that caused $50 million or more of property damage and crop damage respectively.

```
prop_dmg<-data.frame(EVTYPE=data$EVTYPE, PROPDMG=data$PROPDMG, PROPDMGEXP=data$PROPDMGEXP)
prop_dmg<-subset(prop_dmg, (PROPDMGEXP=="K"|PROPDMGEXP=="M"|PROPDMGEXP=="m"|PROPDMGEXP=="B"))
prop_dmg$PROPDMG<-with(prop_dmg, ifelse(PROPDMGEXP=="K", PROPDMG*1000, ifelse((PROPDMGEXP=="M"|PROPDMGEX
prop<-tapply(prop_dmg$PROPDMG, prop_dmg$EVTYPE, sum, na.rm=TRUE)
property_dmg<-data.frame(Event=names(prop), Damage=prop)
property_dmg[with(property_dmg, order(-Damage)), ]
property_dmg<-subset(property_dmg, property_dmg$Damage>=50000000)
crop_dmg<-data.frame(EVTYPE=data$EVTYPE, CROPDMG=data$CROPDMG, CROPDMGEXP=data$CROPDMGEXP)
crop_dmg<-subset(crop_dmg, (CROPDMGEXP=="K"|CROPDMGEXP=="k"|CROPDMGEXP=="M"|CROPDMGEXP=="m"|CROPDMGEXP==
crop_dmg$CROPDMG<-with(crop_dmg, ifelse((CROPDMGEXP=="K"|CROPDMGEXP=="k"), CROPDMG*1000, ifelse((CROPDM
crop<-tapply(crop_dmg$CROPDMG, crop_dmg$EVTYPE, sum, na.rm=TRUE)
crops<-data.frame(Event=names(crop), Damage=crop)
crops<-crops[with(crops, order(-Damage)), ]
crops<-subset(crops, crops$Damage>=50000000)
```

Then consolidate the data based on event type as before.

```
property_dmg$Event<-as.character(property_dmg$Event)
property_dmg$Event<-ifelse(grepl("FLOOD", property_dmg$Event), "FLOOD/FLASH FLOOD", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("FLD", property_dmg$Event), "FLOOD/FLASH FLOOD", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("Flood", property_dmg$Event), "FLOOD/FLASH FLOOD", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("HURRICANE", property_dmg$Event), "HURRICANE", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("TYPHOON", property_dmg$Event), "HURRICANE", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("THUNDERSTORM", property_dmg$Event), "THUNDERSTORM", property_dmg$Even
property_dmg$Event<-ifelse(grepl("TSTM", property_dmg$Event), "THUNDERSTORM", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("COLD", property_dmg$Event), "COLD", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("TORNADO", property_dmg$Event), "TORNADO", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("WIND", property_dmg$Event), "HIGH WIND", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("FIRE", property_dmg$Event), "WILD/FOREST FIRE", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("RAIN", property_dmg$Event), "HEAVY RAINS", property_dmg$Event)
property_dmg$Event<-ifelse(grepl("SURGE", property_dmg$Event), "STORM SURGE", property_dmg$Event)
crops$Event<-as.character(crops$Event)
crops$Event<-ifelse(grepl("HURRICANE", crops$Event), "HURRICANE", crops$Event)
crops$Event<-ifelse(grepl("FREEZE", crops$Event), "FREEZE", crops$Event)
crops$Event<-ifelse(grepl("THUNDERSTORM", crops$Event), "THUNDERSTORM", crops$Event)
crops$Event<-ifelse(grepl("TSTM", crops$Event), "THUNDERSTORM", crops$Event)
crops$Event<-ifelse(grepl("WIND", crops$Event), "HIGH WIND", crops$Event)
crops$Event<-ifelse(grepl("FIRE", crops$Event), "WILD/FOREST FIRE", crops$Event)
crops$Event<-ifelse(grepl("FLOOD", crops$Event), "FLOOD/FLASH FLOOD", crops$Event)
crops$Event<-ifelse(grepl("RAIN", crops$Event), "HEAVY RAINS", crops$Event)
crops$Event<-ifelse(grepl("COLD", crops$Event), "COLD", crops$Event)
```

Combine the property damage and crop damage into one data frame.

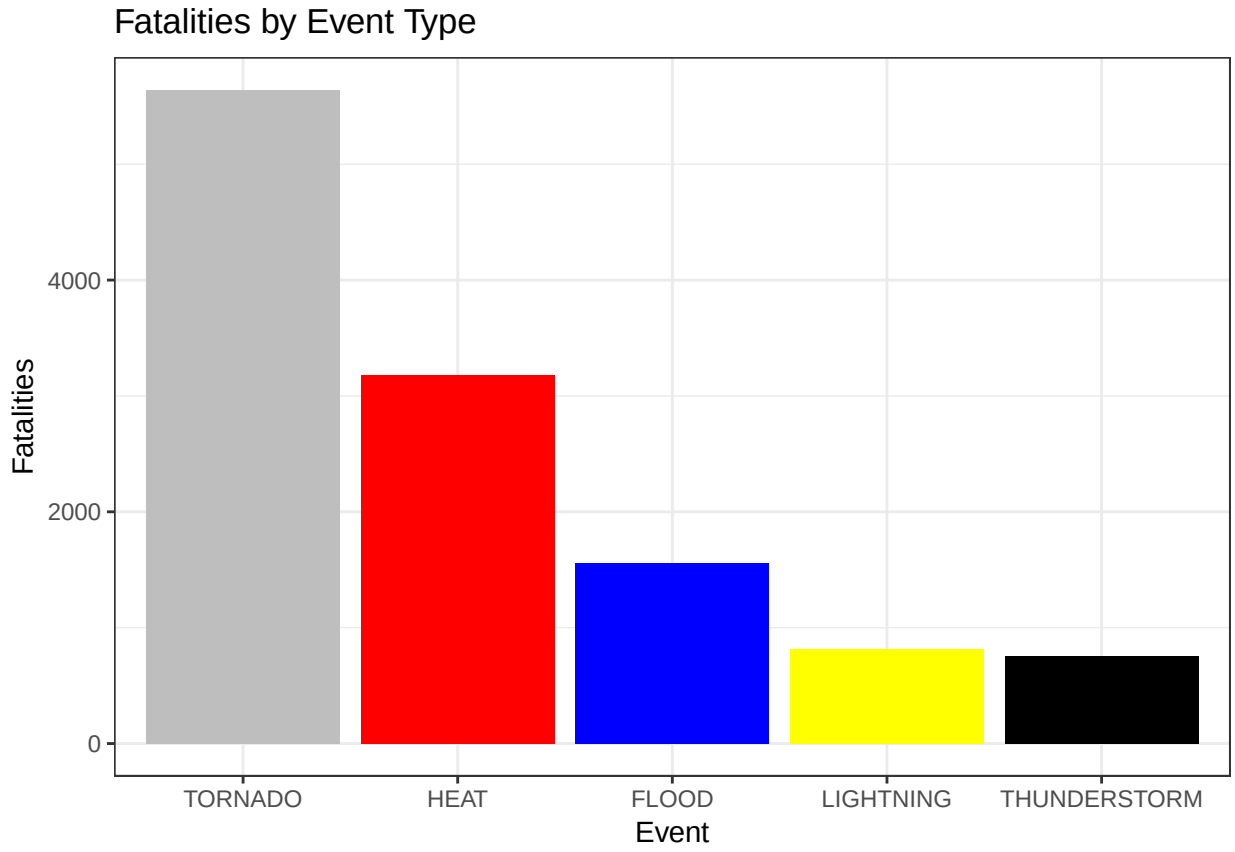
```
economic<-rbind(property_dmg, crops)
economic<-tapply(economic$Damage, economic$Event, sum)
economic<-data.frame(Event=names(economic), Damage=economic)
economic<-economic[with(economic, order(-Damage)), ]
```

```
row.names(economic)<-NULL
```

Results:

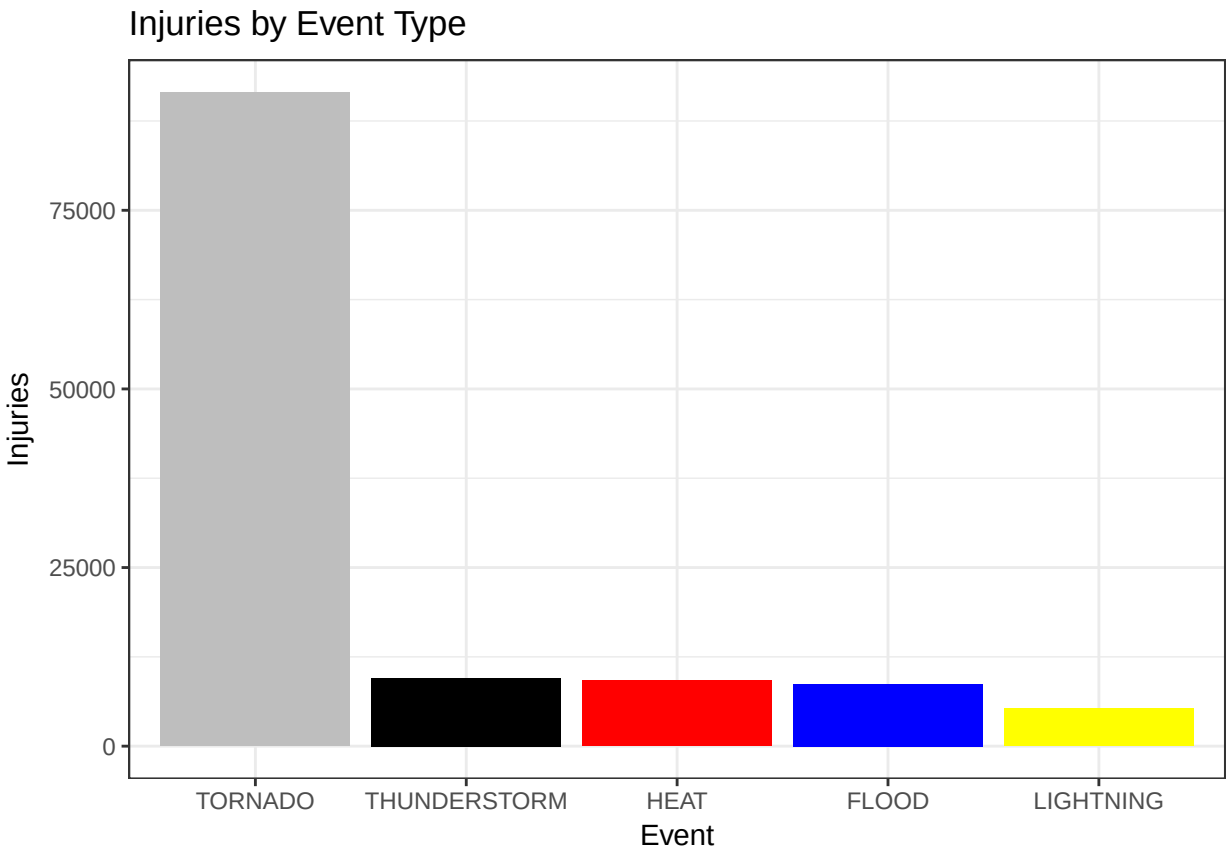
From this data we can derive the following graphs. First, a bar graph of the fatalities from the five most deadly weather events.

```
top_fat<-head(fatalities_by_type, 5)
library(ggplot2)
top_fat$Event<-reorder(top_fat$Event, -top_fat$Fatalities)
f<-ggplot(top_fat, aes(Event, Fatalities))
f+geom_col(aes(fill=Event))+scale_fill_manual(values=c("gray", "red", "blue", "yellow", "black"))+labs(
```



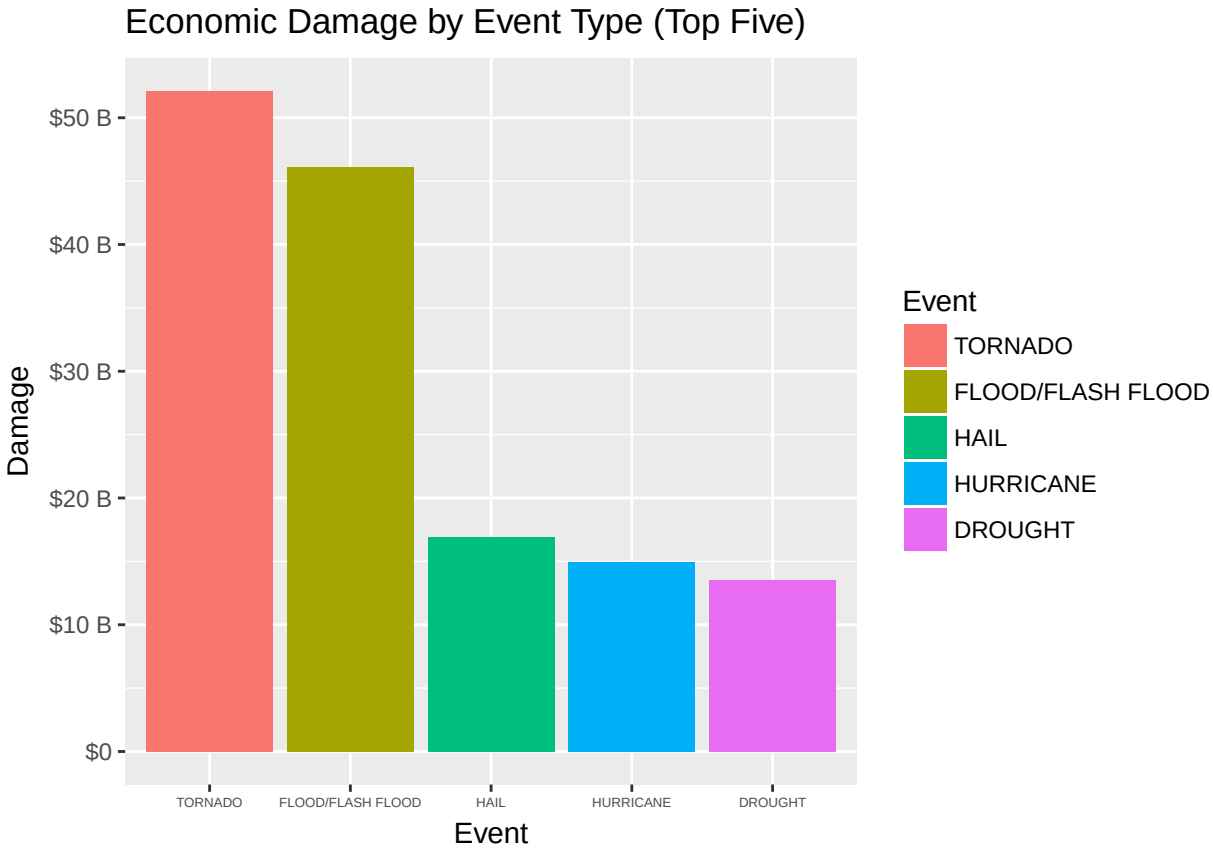
We can see that tornadoes are by far the most deadly weather events. We also see that tornadoes also cause the most injuries in the following graph.

```
top_inj<-head(injuries_by_type, 5)
top_inj$Event<-reorder(top_inj$Event, -top_inj$Injuries)
i<-ggplot(top_inj, aes(Event, Injuries))
i+geom_col(aes(fill=Event))+scale_fill_manual(values=c("gray", "black", "red", "blue", "yellow"))+labs(
```



Looking at economic costs (both property and crops) it is clear that tornadoes also cause the most damage, though not by the same pronounced margin as in the case of injuries and fatalities.

```
top_damage<-head(economic, 5)
top_damage$Event<-reorder(top_damage$Event, -top_damage$Damage)
e<-ggplot(top_damage, aes(Event, Damage))
e+geom_col(aes(fill=Event))+scale_y_continuous(label=c("$0", "$10 B", "$20 B", "$30 B", "$40 B", "$50 B
```


Economic Damage by Event Type (Top Five)
Damage
$50 B
$40 B
$30 B
$20 B
$10 B
$0
TORNADO
FLOOD/FLASH FLOOD
HAIL
HURRICANE
DROUGHT
Event
Event
TORNADO
FLOOD/FLASH FLOOD
HAIL
HURRICANE
DROUGHT