PA2

Peggy Fan

September 1, 2014

Title:

Impact of storms on population health and the economy: Data from the U.S. National Oceanic and Atmospheric Administration (NOAA)

Synopsis:

This analysis uses the storm data from NOAA dated from 1950 to November 2011 answer two questions: what are the types of storm events are most harmful to population health, and what types of storm events caused the highest levels of damages in the economy? After sorting the event types into categories listed on page. 6 of the NOAA Storm Documentation File, the total number of fatalities and juries, as well as the property and crop damage in dollar amount, were calculated for each event type. The top three event types that were the most harmful were tornado, heat, and flood. The top three event types that caused the most property damage were hurricane, high wind, and wild fire, and for crop damage, they were flood, hurricane, and high wind.

Data Processing

Load file

```
setwd("/Users/peggyfan/Downloads/R_data/Reproducible_research")
url <- "http://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
temp <- tempfile()
download.file(url, destfile = "./repdata-data-StormData.csv.bz2", method = "curl", mode = "w
b")
bunzip2(filename = "repdata-data-StormData.csv.bz2", destname = "stormData.csv")
```

storm <- read.csv("./repdata-data-StormData.csv")

Subset data to only relevant columns. Population health is represented by fatalities and injuries. Impact on the economy is presernted by property damage and crop damage.

```
names(storm)
```

```
##  [1] "STATE__"    "BGN_DATE"   "BGN_TIME"   "TIME_ZONE"  "COUNTY"
##  [6] "COUNTYNAME" "STATE"      "EVTYPE"     "BGN_RANGE"  "BGN_AZI"
## [11] "BGN_LOCATI" "END_DATE"   "END_TIME"   "COUNTY_END" "COUNTYENDN"
## [16] "END_RANGE"  "END_AZI"    "END_LOCATI" "LENGTH"     "WIDTH"
## [21] "F"          "MAG"        "FATALITIES" "INJURIES"   "PROPDMG"
## [26] "PROPDMGEXP" "CROPDMG"    "CROPDMGEXP" "WFO"        "STATEOFFIC"
## [31] "ZONENAMES"  "LATITUDE"   "LONGITUDE"  "LATITUDE_E" "LONGITUDE_"
## [36] "REMARKS"    "REFNUM"
```

```
keeps <- c("STATE","COUNTY", "EVTYPE", "FATALITIES", "INJURIES", "PROPDMG", "PROPDMGEXP", "CR
OPDMG", "CROPDMGEXP")
storm1 <-storm[keeps]
storm1$EVTYPE <- as.factor(toupper(storm1$EVTYPE))
storm1$EVTYPE <- gsub("/", "", storm1$EVTYPE)
storm1$EVTYPE <- as.factor(storm1$EVTYPE)
```

Question 1:

Subset data that contains non-zero values of pop_health

```
health_data <- subset(storm1, storm1$FATALITIES!=0 & storm1$INJURIES!=0)
```

Calculate the sum for fatalities and injuries by event type, and create a new table of those sums.

```
library(plyr)
Fatalities_sum <-ddply(health_data, c("EVTYPE"),summarize,sum=sum(FATALITIES), na.rm=TRUE)
Fatalities_sum <- Fatalities_sum[, -3]
Fatalities_sum$Category <- "Fatalities"

Injuries_sum <-ddply(health_data, c("EVTYPE"),summarize,sum=sum(INJURIES), na.rm=TRUE)
Injuries_sum <- Injuries_sum[, -3]
Injuries_sum$Category <- "Injuries"

health_graph <- rbind(Fatalities_sum, Injuries_sum)
```

Clean up event type variables for health data. The strategy is to use the event type table on page.6 of the Storm Data Documentation File, where keywards related to an event type are used to recode that variable into the particular event type.

```
vars <- as.character(health_graph$EVTYPE)
vars <- gsub("ICE STORM.*", "TEST", vars)
vars <- gsub("EXTREME COLDWIND CHILL", "TEST2", vars)
vars <- gsub("AVALAN.*","AVALANCHE", vars)
vars <- gsub("BLIZZARD.*|BLOWING SNOW.*|.+SQUALL.*|SNOWHIGH.*", "BLIZZARD", vars)
vars <- gsub("COASTAL.*", "COASTAL FLOOD", vars)
vars <- gsub("HIGH WINDSEAS", "MARINE HIGH WIND", vars)
vars <- gsub("HIGH.*|\\WIND\\>|\\<WINDS\\>|RAINWIND", "HIGH WIND", vars)
vars <- gsub(".+COLD|COLD.*", "COLD/WIND CHILL", vars)
vars <- gsub("DROUGHT.*", "DROUGHT", vars)
vars <- gsub("EXTREME HEAT", "EXCESSIVE HEAT", vars)
vars <- gsub("EXTREME COLD.*|.+WINDCHILL|EXTREME COLDWIND CHILL", "EXTREME COLD/WIND CHILL",
vars)
vars <- gsub("LOW TEMPERATURE|HYPOTHERMIA.*", "EXTREME COLD/WIND CHILL", vars)
vars <- gsub("FLASH FLOOD.*", "FLASH FLOOD", vars)
vars <- gsub(".+FLOOD.*|FLOOD.*", "FLOOD", vars)
vars <- gsub("FREEZ.*|FROST|\\<ICE\\>|.+ROAD.*|BLACK ICE|GLAZE.*", "FROST/FREEZE", vars)
vars <- gsub("FOG.*", "FOG", vars)
vars <- gsub(".+HAIL.*", "HAIL", vars)
vars <- gsub(".+WARM|WARM.*|.+HEAT.*|HYPERTHERMIA.*", "HEAT", vars)
vars <- gsub(".+RAIN.*|.+SLIDE.*", "HEAVY RAIN", vars)
vars <- gsub(".+SNOW|SNOW.*", "HEAVY SNOW", vars)
vars <- gsub(".+SURF.*", "HEAVY SURF", vars)
vars <- gsub("HURRICANE.*|TYPHOON", "HURRICANE", vars)
vars <- gsub("TSTM WIND.*", "MARINE STRONG WIND", vars)
vars <- gsub("RIP CURRENT.*", "RIP CURRENT", vars)
vars <- gsub("STORM SURGE.*", "STORM SURGE/TIDE", vars)
vars <- gsub("THUNDERS.*|DRY MICROBUST.*|LIGHTNING.*|\\<WIND STORM\\>", "THUNDERSTORM WIND",
vars)
vars <- gsub("TROPICAL.*", "TROPICAL STORM", vars)
vars <- gsub("TORNADO.*", "TORNADO", vars)
vars <- gsub("WATERSPOUT.*", "WATERSPOUT", vars)
vars <- gsub(".+FIRE.*", "WILD FIRE", vars)
vars <- gsub("WINT.*|.+MIX.*", "WINTER STORM", vars)
vars <- gsub("TEST", "ICE STORM", vars)
vars <- gsub("TEST2", "EXTREME COLD/WIND CHILL", vars)
vars <- gsub("STRONG WINDS", "STRONG WIND", vars)

health_graph$EVTYPE <- (vars)
```

Recode event type as factor variable

```
health_graph$EVTYPE <- as.factor(health_graph$EVTYPE)
levels(health_graph$EVTYPE)
```

```
##  [1] "AVALANCHE"                    "BLIZZARD"
##  [3] "COLD/WIND CHILL"              "COLD/WIND CHILL TEMPERATURES"
##  [5] "DENSE FOG"                    "DUST STORM"
##  [7] "EXTREME COLD/WIND CHILL"      "FLOOD"
##  [9] "FOG"                          "FROST/FREEZE"
## [11] "GUSTY HIGH WIND"              "HAIL"
## [13] "HEAT"                         "HEAT WAVE"
## [15] "HEAT WAVE DROUGHT"            "HEAVY RAIN"
## [17] "HEAVY SNOW"                   "HEAVY SURF"
## [19] "HIGH WIND"                    "HURRICANE"
## [21] "ICE STORM"                    "ICE STORM2"
## [23] "MARINE ACCIDENT"              "MARINE HIGH WIND"
## [25] "MARINE MARINE STRONG WIND"    "MARINE MISHAP"
## [27] "MARINE STRONG WIND"           "MARINE THUNDERSTORM WIND"
## [29] "MIXED PRECIP"                 "RIP CURRENT"
## [31] "ROUGH SEAS"                   "STORM SURGE/TIDE"
## [33] "STRONG HIGH WIND"             "STRONG WIND"
## [35] "THUNDERSTORM WIND"            "TORNADO"
## [37] "TROPICAL STORM"               "TSUNAMI"
## [39] "URBANSML STREAM FLD"          "WATERSPOUT"
## [41] "WILD FIRE"                    "WIND"
## [43] "WINTER STORM"
```

Question 2:

Subset data to only retain cases where both property damage and crop damage are non-zero

```
damage_data <- subset(storm1, storm1$PROPDMG!=0 & storm1$CROPDMG!=0 )
```

Convert the property damages into numeric values, using the variable PROPDMGEXP and CROPDMGEXP. Values outside of H, K, M, and B are ignored due to the confusion of what those values might mean.

First convert all letters to upper case

```
damage_data$PROPDMGEXP <- toupper(damage_data$PROPDMGEXP)
damage_data$CROPDMGEXP <- toupper(damage_data$CROPDMGEXP)
```

For property damage, recode the numeric values of H, K, M, B, then calculate the numeric values of PROPDMG.

```
damage_data$pexpH <- ifelse(damage_data[ , "PROPDMGEXP"] == "H", 100, NA )
damage_data$pexpK <- ifelse(damage_data[ , "PROPDMGEXP"] == "K", 1000, NA )
damage_data$pexpM <- ifelse(damage_data[ , "PROPDMGEXP"] == "M", 1000000, NA )
damage_data$pexpB <- ifelse(damage_data[ , "PROPDMGEXP"] == "B", 1000000000, NA )

damage_data$pEXP <- rowMeans(damage_data[, c("pexpH", "pexpK", "pexpM", "pexpB")], na.rm=TRU
E)
damage_data$prop_d <- damage_data$PROPDMG*damage_data$pEXP
```

Repeat the same recoding steps for crop damage.

```
damage_data$cexpH <- ifelse(damage_data[ , "CROPDMGEXP"] == "H", 100, NA )
damage_data$cexpK <- ifelse(damage_data[ , "CROPDMGEXP"] == "K", 1000, NA )
damage_data$cexpM <- ifelse(damage_data[ , "CROPDMGEXP"] == "M", 1000000, NA )
damage_data$cexpB <- ifelse(damage_data[ , "CROPDMGEXP"] == "B", 1000000000, NA )

damage_data$cEXP <- rowMeans(damage_data[, c("cexpH", "cexpK", "cexpM", "cexpB")], na.rm=TRU
E)
damage_data$crop_d <- damage_data$CROPDMG*damage_data$cEXP
```

Calculate the sums for property damage and crop damage by event type and create a new table for both types of sum.

```
Prop_d_sum <-ddply(damage_data, c("EVTYPE"),summarize,sum=sum(prop_d), na.rm=TRUE)
Prop_d_sum$Category <- "Property Damage"

Crop_d_sum <-ddply(damage_data, c("EVTYPE"),summarize,sum=sum(crop_d), na.rm=TRUE)
Crop_d_sum$Category <- "Crop Damage"

damage_graph <- rbind(Prop_d_sum, Crop_d_sum)
```

Clean up event variable for damage data using the same strategy as the recoding of event type of quesiton 1.

```
vars1 <- as.character(damage_graph$EVTYPE)
vars1 <- gsub("ICE STORM.*", "TEST", vars1)
vars1 <- gsub("EXTREME COLDWIND CHILL", "TEST2", vars1)
vars1 <- gsub("AVALAN.*","AVALANCHE", vars1)
vars1 <- gsub("BLIZZARD.*|BLOWING SNOW.*|.+SQUALL.*|SNOWHIGH.*", "BLIZZARD", vars1)
vars1 <- gsub("COASTAL.*", "COASTAL FLOOD", vars1)
vars1 <- gsub("HIGH WINDSEAS", "MARINE HIGH WIND", vars1)
vars1 <- gsub("HIGH.*|\\WIND\\>|\\<WINDS\\>|RAINWIND", "HIGH WIND", vars1)
vars1 <- gsub(".+COLD|COLD.*", "COLD/WIND CHILL", vars1)
vars1 <- gsub("DROUGHT.*", "DROUGHT", vars1)
vars1 <- gsub("EXTREME HEAT", "EXCESSIVE HEAT", vars1)
vars1 <- gsub("EXTREME COLD.*|.+WINDCHILL|EXTREME COLDWIND CHILL", "EXTREME COLD/WIND CHILL",
vars1)
vars1 <- gsub("LOW TEMPERATURE|HYPOTHERMIA.*", "EXTREME COLD/WIND CHILL", vars1)
vars1 <- gsub("FLASH FLOOD.*", "FLASH FLOOD", vars1)
vars1 <- gsub(".+FLOOD.*|FLOOD.*", "FLOOD", vars1)
vars1 <- gsub("FREEZ.*|FROST.*|\\<ICE\\>|.+ROAD.*|BLACK ICE|GLAZE.*", "FROST/FREEZE", vars1)
vars1 <- gsub("FOG.*", "FOG", vars1)
vars1 <- gsub(".+HAIL.*", "HAIL", vars1)
vars1 <- gsub(".+WARM|WARM.*|.+HEAT.*|HYPERTHERMIA.*", "HEAT", vars1)
vars1 <- gsub(".+RAIN.*|.+SLIDE.*", "HEAVY RAIN", vars1)
vars1 <- gsub(".+SNOW|SNOW.*", "HEAVY SNOW", vars1)
vars1 <- gsub(".+SURF.*", "HEAVY SURF", vars1)
vars1 <- gsub("HURRICANE.*|TYPHOON", "HURRICANE", vars1)
vars1 <- gsub("TSTM WIND.*", "MARINE STRONG WIND", vars1)
vars1 <- gsub("RIP CURRENT.*", "RIP CURRENT", vars1)
vars1 <- gsub("STORM SURGE.*", "STORM SURGE/TIDE", vars1)
vars1 <- gsub("THUNDERS.*|DRY MICROBUST.*|LIGHTNING.*|\\<WIND STORM\\>|SEVERE.*", "THUNDERSTO
RM WIND", vars1)
vars1 <- gsub("TROPICAL.*", "TROPICAL STORM", vars1)
vars1 <- gsub("TORNADO.*", "TORNADO", vars1)
vars1 <- gsub("WATERSPOUT.*", "WATERSPOUT", vars1)
vars1 <- gsub(".+FIRE.*", "WILD FIRE", vars1)
vars1 <- gsub("WINT.*|.+MIX.*", "WINTER STORM", vars1)
vars1 <- gsub("TEST", "ICE STORM", vars1)
vars1 <- gsub("TEST2", "EXTREME COLD/WIND CHILL", vars1)
vars1 <- gsub("STRONG WINDS", "STRONG WIND", vars1)

damage_graph$EVTYPE <- (vars1)
```

Recode event type as factor variable

```
damage_graph$EVTYPE <- as.factor(damage_graph$EVTYPE)
```

Results

Question 1:

Calculating new sums by the recoded event type variable. Remove duplicates.

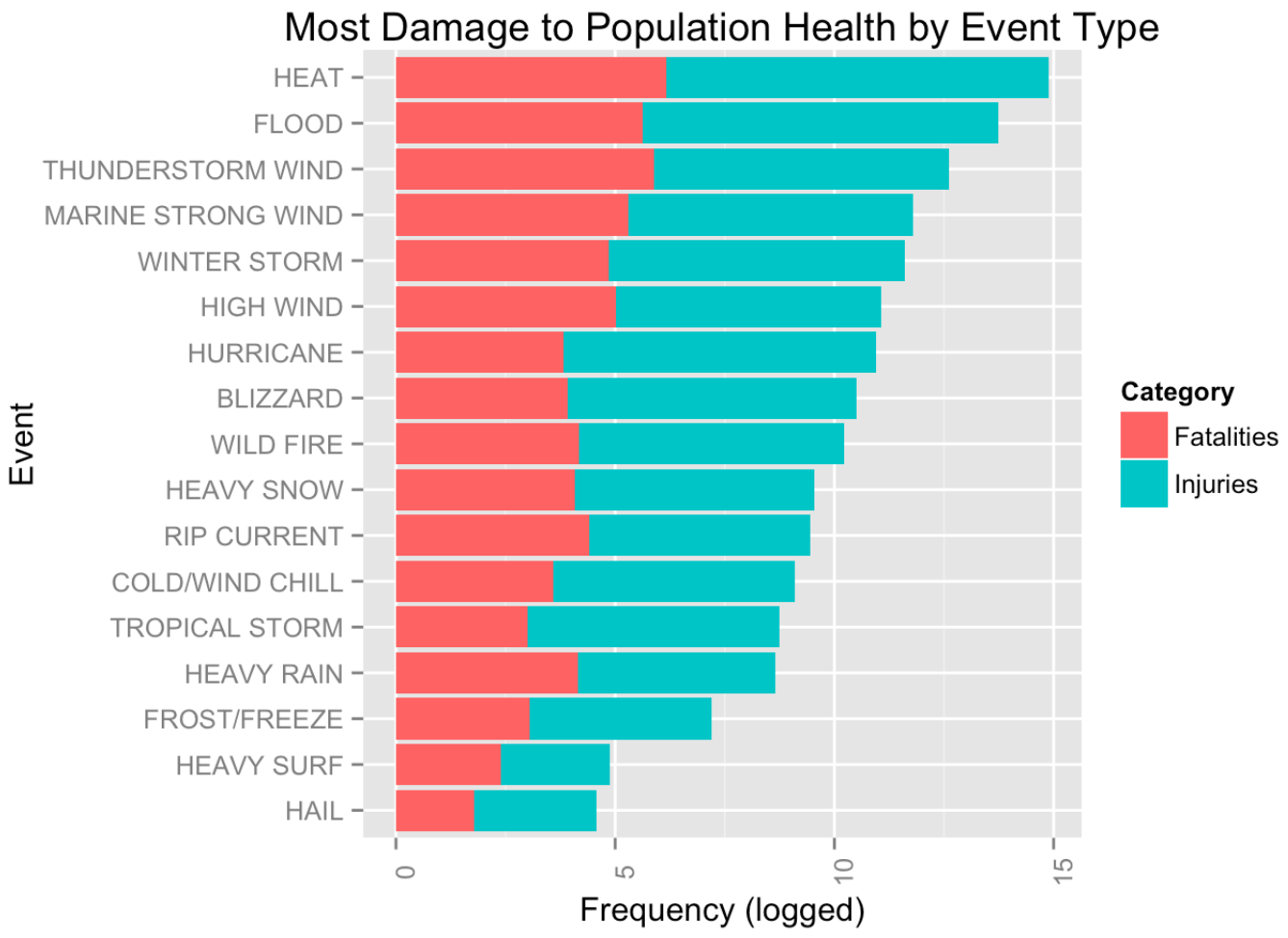
```
fn <- function(x) sum(x)
test <- ddply(health_graph, c("EVTYPE", "Category"),transform, sum=fn(sum), na.rm=TRUE)

health_graph1 <- unique(test[duplicated(test),])
```

Log the sums to highlight lower frequencies due to the huge discrepancy between the sums. For instance, tornado's number was near 100,000, and the second highest event "Heat" has only about 8800.

```
health_graph1$sum <- log(health_graph1$sum)
```

Create graph for damage of population health

```
library(ggplot2)
ggplot(data=health_graph1, aes(reorder(EVTYPE, sum), y=sum, fill = Category)) +
  geom_bar(stat ="identity") + coord_flip() +
  theme(axis.text.x=element_text(angle=90)) +
  ggtitle("Most Damage to Population Health by Event Type") +
  labs(y = 'Frequency (logged)', x = 'Event')
```



Question 2:

Calculating new sums by the recoded event type variable.

```
fn <- function(x) sum(x)
test1 <- ddply(damage_graph, c("EVTYPE", "Category"),transform, sum=fn(sum), na.rm=TRUE)
```

Remove duplicates and NAs

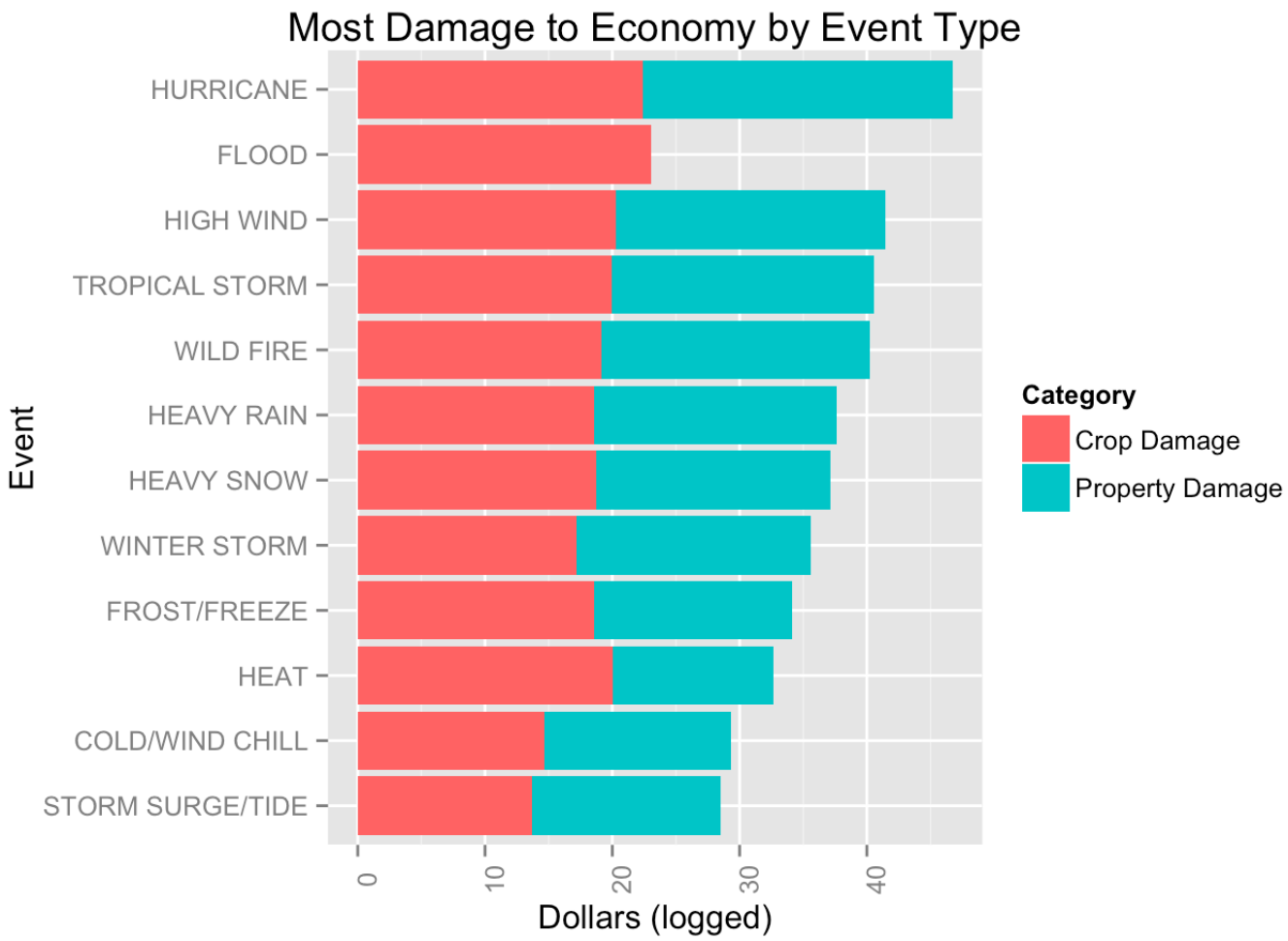
```
damage_graph1 <- unique(test1[duplicated(test1),])
damage_graph1 <-damage_graph1[complete.cases(damage_graph1), ]
```

Taking log of the sum because the sums for hurricane and flood are many times higher than the rest of events, causing a highly screwed distribtuion.

```
damage_graph1$sum1 <- log(damage_graph1$sum)
```

Create graph for property and crop damage.

```
library(ggplot2)
ggplot(data=damage_graph1, aes(reorder(EVTYPE, sum1), y=sum1, fill = Category)) +
  geom_bar(stat ="identity") + coord_flip() +
  theme(axis.text.x=element_text(angle=90)) +
  ggtitle("Most Damage to Economy by Event Type") +
  labs(y = 'Dollars (logged)', x = 'Event')
```


Most Damage to Economy by Event Type
HURRICANE
FLOOD
HIGH WIND
TROPICAL STORM
WILD FIRE
HEAVY RAIN
HEAVY SNOW
WINTER STORM
FROST/FREEZE
HEAT
COLD/WIND CHILL
STORM SURGE/TIDE
Event
0
10
20
30
40
Dollars (logged)
Category
Crop Damage
Property Damage