Extreme Weather Events

Synopsis

Data from the US NOAA National Climatic Data Center *Storm Events Database* [1] include statistics on loss of human life and damage to property. This study will briefly explore the relationship between different categories of weather events and thier impact.

NOAA's National Weather Service has been collecting data on storm events since 1950. [2] Over the history of the program varied and increasingly detailed information about the categories of weather events has been reported and compiled. In order to form a uniform data set of events this study uses data beginning in 1996 which follows NWS Instruction 10-1605 [3] outlaying 48 weather categories.

Data Processing

The data come in a bz2-compressed csv file with all events on record from 1950 through 2011.

```
need <- function(libraries) {
  # Load libraries, installing first if necessary
  for (i in base::setdiff(libraries, .packages(all = TRUE))) {install.packages(i)}
  for (i in base::setdiff(libraries, .packages(all = FALSE))) {
    library(package = i, character.only = TRUE)}
}
need(c("dplyr", "lubridate", "tidyr", "ggplot2", "magrittr"))
```

```
if (!file.exists("data/StormData.csv.bz2")) {
  download.file(url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",
                destfile = "data/StormData.csv.bz2", method = "curl")
}

load_storm_data_csv_bz <- function(bzf) {
  read.csv(file = bzfile(bzf), header = TRUE,
# read directly from bz2 file, wrap output dataframe in dplyr::tbl_df class
    colClasses = c(
      "NULL",
      "character", #date
      rep("NULL", times = 4), #skip columns
      "factor", #state
      "factor", #type
      rep("NULL", times = 14), #skip columns
      "numeric", #fatal
      "numeric", #injuries
      "numeric", #property
      "factor", #propexp
      "numeric", #crops
      "factor", #cropsexp
      rep("NULL", times = 9))) %>% tbl_df # put the data frame in dplyr df object
}
```

[1] http://www.ncdc.noaa.gov/stormevents/
[2] http://www.ncdc.noaa.gov/stormevents/details.jsp?type=eventtype
[3] http://www.ncdc.noaa.gov/stormevents/pd01016005curr.pdf

```
if(!exists("storm.data")) {
  if (file.exists("data/stormdata.rds")) storm.data <- readRDS(file = "data//stormdata.rds")
    # load saved object on disk if available, otherwise:
  else {
    if (!file.exists("data/StormData.csv.bz2")) download.file(
      destfile = "data/StormData.csv.bz2", method = "curl",
      url = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2")
    storm.data <- load_storm_data_csv_bz("data/StormData.csv.bz2")
    saveRDS(object = storm.data, ascii = FALSE, compress = "bzip2",
      file = "data//stormdata.rds") #cache data in only 1.2MB on disk
  }
}
```

```
storm.data2 <- storm.data %>%
  transmute( # transmute ony selects named columns
    date = lubridate::mdy(
      x = sub(x = BGN_DATE, pattern = "([0-9/]+).*", replacement = "\\1")),
    year = lubridate::year(date),
    EVTYPE, FATALITIES, INJURIES,
    PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP
  ) %>% dplyr::filter(year >= 1996)

# This filter would be part of prev op, but some unknown bug in knitr environment stops it from working

storm.data2 %<>% mutate( # mutate changes named columns, but does not affect the rest
  property = PROPDMG *
    ifelse(PROPDMGEXP == "1", 1, # treat 1 as multiplying by 10^0 or 1
       ifelse(PROPDMGEXP == "h" || PROPDMGEXP == "H", 100, #hundreds mult by 10^2
          ifelse(PROPDMGEXP == "K" || PROPDMGEXP == "k", 1e3, #thousands, times 10^3
             ifelse(PROPDMGEXP == "M" || PROPDMGEXP == "M", 1e6, #millions, times 10^6
                ifelse(PROPDMGEXP == "B", 1e9, #billions, times 10^9
    NA))))) # make everything else NA
  )

storm.data2 %<>% mutate(
  crops = CROPDMG *
    ifelse(CROPDMGEXP == 1, 1,
       ifelse(CROPDMGEXP == "h" || CROPDMGEXP == "H", 100, #hundreds
          ifelse(CROPDMGEXP == "K" || CROPDMGEXP == "k", 1e3, #thousands
             ifelse(CROPDMGEXP == "M" || CROPDMGEXP == "M", 1e6, #millions
                ifelse(CROPDMGEXP == "B", 1e9,
    NA))))) #billions, make everything else NA
  )

storm.data3 <- storm.data2 %>%
  select(year, type = EVTYPE, fatality = FATALITIES, property, crops)
```

Overview

This analysis focuses on extreme weather events from 1996 through 2011. While the documentation seems to say that there are 48 types of weather tracked corresponding with broad categories: tornadoes, hurricanes,

floods, and heat.

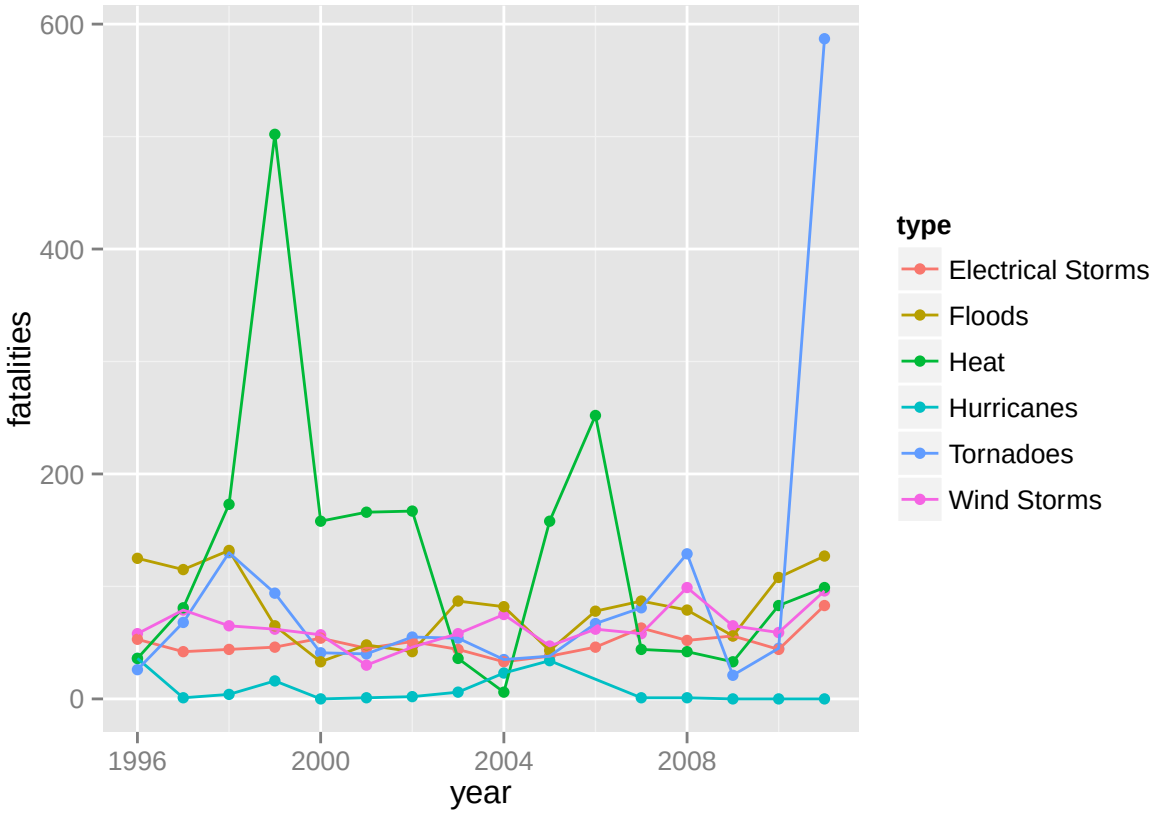
```
#Based on some rough exploration of only event counts:
# for (i in c("wind", "thun", "light", "floo", "torn", "heat", "hurr")) {
#   print(paste(i, length(grep(storm.data3$EVTYPE, pattern = i, ignore.case = TRUE))))}
# [1] "wind 364901"
# [1] "thun 109587"
# [1] "light 15987"
# [1] "floo 82732"
# [1] "torn 60701"
# [1] "heat 2648"
# [1] "hurr 288"

impact <- storm.data2 %>%
  group_by(EVTYPE, year) %>%
  summarise(
    fatalities = sum(FATALITIES),
    property = sum(property),
    count = n()
    ) %>%
  mutate(
    wind = grepl(pattern = "WIND", x = EVTYPE),
    elec = grepl(pattern = "THUN|LIGHT", x = EVTYPE),
    torn = grepl(pattern = "TORN", x = EVTYPE),
    hurr = grepl(pattern = "HURR|TYPH", x = EVTYPE),
    floo = grepl(pattern = "FLOO", x = EVTYPE),
    heat = grepl(pattern = "HEAT", x = EVTYPE)
)

impact.all <- rbind(
  impact %>%
    dplyr::filter(torn) %>%
    group_by(year) %>%
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Tornadoes")
  ,
  impact %>%
    dplyr::filter(wind) %>%
    group_by(year) %>%
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Wind Storms")
  ,
  impact %>%
    dplyr::filter(elec) %>%
    group_by(year) %>%
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Electrical Storms")
  ,
  impact %>%
    dplyr::filter(hurr) %>%
    group_by(year) %>%
```

```
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Hurricanes")
  ,
  impact %>%
    dplyr::filter(floo) %>%
    group_by(year) %>%
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Floods")
  ,
  impact %>%
    dplyr::filter(heat) %>%
    group_by(year) %>%
    summarise(fatalities = sum(fatalities),
              property = sum(property)) %>%
    mutate(type = "Heat")
  )
```

According to the reported data, how the cost to human life of the different

types of storms compare:

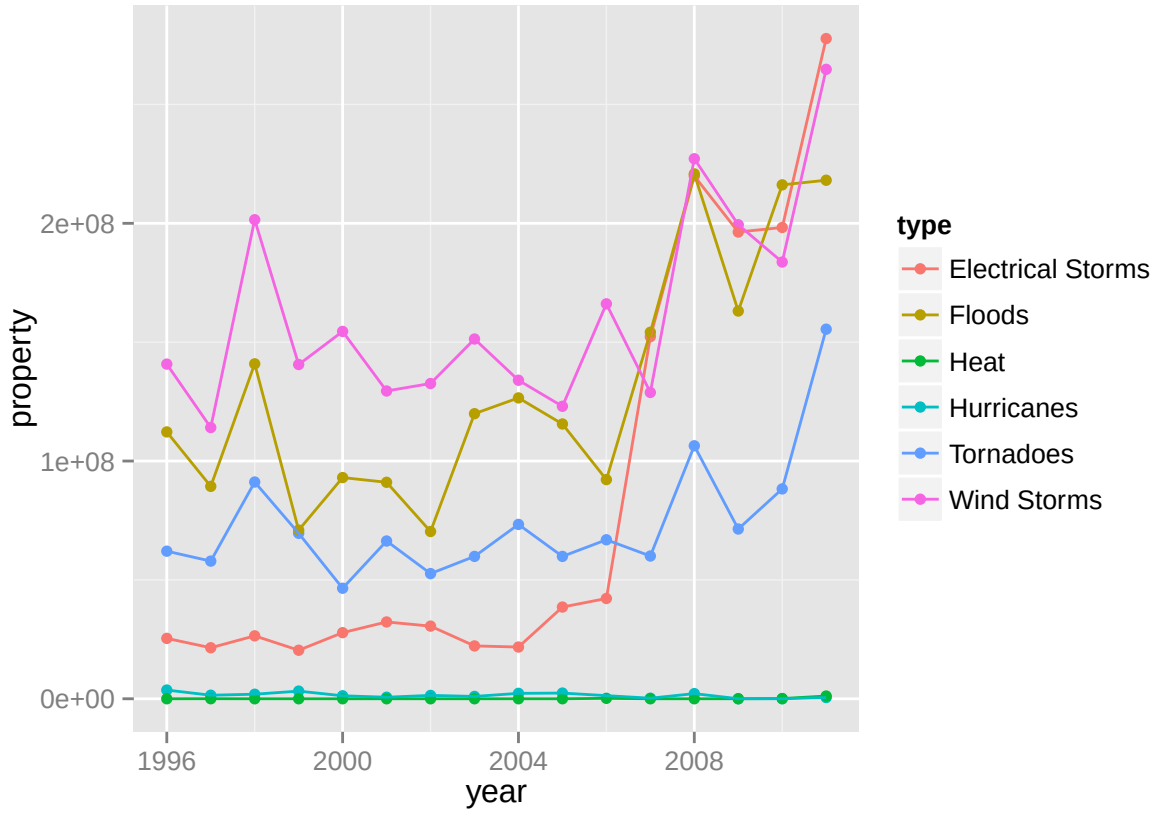
```
ggplot(data = impact.all) +
  aes(x = year, y = fatalities) +
  geom_point(mapping = aes(group = type, color = type)) +
  geom_line(aes(color = type))
```



According to the reported data, how the property damage of the different

types of storms compare:

```
ggplot(data = impact.all) +
  aes(x = year, y = property) +
  geom_point(mapping = aes(group = type, color = type)) +
  geom_line(aes(color = type))
```



Conclusion

According to the data hurricanes tend to result in fewer deaths than the other three types. Despite their prominence in the news they are rarer events than fatal heat waves, or deaths from flooding and tornadoes.

While heat events are notable for the number of deaths, they result in little property damage. However, floods cause significant property damage, as do tornadoes.