Most Harmful Weather Events

Shivanand R Koppalkar

April 3, 2016

Synopsis

Across the United States, tornadoes, excessive heat, and flash floods are most harmful with respect to population health.

Across the United States, tornadoes, thunderstorm winds, and flash floods have the greatest economic consequences.

Our raw data are taken from National Weather Service Instruction 10-1605. The events in the database start in the year 1950 and end in November 2011. Fatalities, injuries, and property damage (in dollars) are totalled over that time.

Data Processing

We load our data.

```
storm.data = read.csv(bzfile("repdata-data-StormData.csv.bz2"), header =
TRUE)
```

We don't need all the columns.

```
reduced.storm.data <-
        storm.data[,c("EVTYPE", "FATALITIES", "INJURIES", "PROPDMG")]
```

Normalize event names.

```
reduced.storm.data$EVTYPE <-
    gsub("^HEAT$", "EXCESSIVE HEAT", reduced.storm.data$EVTYPE)
reduced.storm.data$EVTYPE <-
    gsub("^TSTM WIND$", "THUNDERSTORM WIND", reduced.storm.data$EVTYPE)
reduced.storm.data$EVTYPE <-
    gsub("^THUNDERSTORM WIND$", "THUNDERSTORM WINDS",
reduced.storm.data$EVTYPE)
```

First we aggregate data on fatalities and find which events are the top 10 causes of fatalities.

```
agg.fatalities.data <-
        aggregate(
```

```
                reduced.storm.data$FATALITIES,
                by=list(reduced.storm.data$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(agg.fatalities.data) = c("event.type", "fatality.total")
fatalities.sorted <-
    agg.fatalities.data[order(-agg.fatalities.data$fatality.total),]
top.fatalities <- fatalities.sorted[1:10,]
top.fatalities$event.type <-
    factor(
        top.fatalities$event.type, levels=top.fatalities$event.type,
        ordered=TRUE)
```

We next do the same for injuries.

```
agg.injuries.data <-
        aggregate(
                reduced.storm.data$INJURIES,
                by=list(reduced.storm.data$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(agg.injuries.data) = c("event.type", "injury.total")
injuries.sorted <- agg.injuries.data[order(-
agg.injuries.data$injury.total),]
top.injuries <- injuries.sorted[1:10,]
top.injuries$event.type <-
    factor(
        top.injuries$event.type, levels=top.injuries$event.type,
        ordered=TRUE)
```

Finally we do the same for property damage.

```
agg.prop.dmg.data <-
        aggregate(
                reduced.storm.data$PROPDMG,
                by=list(reduced.storm.data$EVTYPE), FUN=sum, na.rm=TRUE)
colnames(agg.prop.dmg.data) = c("event.type", "prop.dmg.total")
prop.dmg.sorted <- agg.prop.dmg.data[order(-
agg.prop.dmg.data$prop.dmg.total),]
top.prop.dmg <- prop.dmg.sorted[1:10,]
top.prop.dmg$event.type <-
    factor(
        top.prop.dmg$event.type, levels=top.prop.dmg$event.type,
        ordered=TRUE)
```

Results

We graph the top 10 causes of fatalities.

```
library(ggplot2)
ggplot(data=top.fatalities, aes(x=event.type, y=fatality.total)) +
    geom_bar(stat="identity") + xlab("Event type") + ylab("Total
fatalities") +
    ggtitle("Fatalities By Event Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

We do the same for injuries.

```
ggplot(data=top.injuries, aes(x=event.type, y=injury.total)) +
    geom_bar(stat="identity") + xlab("Event type") + ylab("Total injuries")
+
    ggtitle("Injuries By Event Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Finally we do so for property damage.

```
ggplot(data=top.prop.dmg, aes(x=event.type, y=prop.dmg.total)) +
    geom_bar(stat="identity") + xlab("Event type") +
    ylab("Total property damage") +  ggtitle("Property Damage By Event
Type") +
    theme(axis.text.x = element_text(angle = 45, hjust = 1))
```