

Weather Data Analysis for Course

Comments (—)

Share

Hide Toolbars

Research course

Synopsis

This assignment is to show a few dimensions to the weather data. The data contains records for many weather events, their types, # of fatalities, # of injuries, and the dollar amounts associated with each event. This report shows some analysis of that data.

Data Processing

Loading the data. It is just a csv file.

```
weatherData <- read.csv("repdata_data_StormData.csv")
```

The data is really dirty. Some basic manipulations are done here, such as trimming leading/trailing spaces and converting all to upper case. More cleanup could be done, but that isn't the focus of this assignment.

```
library(reshape)

weatherData$evType_clean <- gsub("^\\s+|\\s+$", "", toupper(weatherData$EVTYPE))
evTypeMelt <- melt(weatherData, id = c("evType_clean"), measure.vars = c("FATALITIES",
  "INJURIES"))

evTypeSummary <- cast(evTypeMelt, evType_clean ~ variable, sum)
```

Fatalities

For the fatalities data, we'll pick the top 12 most fatal types of events. The #12 is chosen because graphing more than that gets tricky and the numbers drop off quickly enough that we can answer the questions we need to without any more data.

```
fatalities <- head(evTypeSummary[order(evTypeSummary$FATALITIES, decreasing = TRUE),
  c("evType_clean", "FATALITIES")], n = 12)

fatalities$eventOrdered <- reorder(fatalities$evType_clean, -fatalities$FATALITIES)
```

Injuries

For the injuries data, we'll pick the top 12 most fatal types of events. The #12 is chosen because graphing more than that gets tricky and the numbers drop off quickly enough that we can answer the questions we need to without any more data.

```
injuries <- head(evTypeSummary[order(evTypeSummary$FATALITIES, decreasing = TRUE),
  c("evType_clean", "INJURIES")], n = 12)
```