

RepReschProject2 using the NOAA Storm Database to answer some questions about severe weather events

Christopher Stonell

2023-10-01

Synopsis

In this analysis project I am using the data from the National Oceanic and Atmospheric Administration's Storm Database from 1950 to 2011. I will address two questions in particular:

1. What weather events have the greatest impact on public health 2. What weather events have the greatest impact in economic indicators.

The analysis will be presented as tables and figures.

Read data from .bz2 file

Data read from the provided .bz2 file. read.csv supports directly reading from compressed .bz2 files.

The data is read as a data frame into the variable "data".

```
# read.csv allows for reading directly from compressed .bz2  
data <- as.data.frame(read.csv("/home/christopher/Documents/Coursera_Data_Science/datasciencecoursera/r/
```

The dataset consists of 902297 rows and 37 variables.

Pre-processing & data transformations

```
# set colClasses for the columns of interest to ensure computations correct.  
data$EVTYPE <- as.character(data$EVTYPE)  
data$FATALITIES <- as.numeric(data$FATALITIES)  
data$INJURIES <- as.numeric(data$INJURIES)  
data$PROPDGM <- as.numeric(data$PROPDGM)  
data$CROPDGM <- as.numeric(data$CROPDGM)
```

Data Processing

```
# do computations for results in this block  
  
# measures of harm to health are columns FATALITIES & INJUTRIES.  
topFatalities <- sort(tapply(data$FATALITIES, data$EVTYPE, FUN=sum), decreasing = TRUE)  
topInjuries <- sort(tapply(data$INJURIES, data$EVTYPE, FUN=sum), decreasing = TRUE)
```

```

topCombined <- sort(tapply((data$FATALITIES+data$INJURIES), list(data$EVTYPE), FUN=sum), decreasing = TRUE)
d_comb <- data.frame(x = topCombined[1:10], group=names(topCombined[1:10]))

# measures of economic impact
topEconomic <- sort(tapply((data$PROPDMG+data$CROPDMG), list(data$EVTYPE), FUN=sum), decreasing = TRUE)

d_econ <- data.frame(x = topEconomic[1:10], group=names(topEconomic[1:10]))

```

results

Questions 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health?

The events are captured in the variable called EVTYPE. There are two variables for health impacts: FATALITIES and INJURIES. The harm to public health can be considered measured by the sum of fatalities and injuries.

The top 10 events with the greatest number of fatalities and injuries were:

Table 1: Table of number of Fatalities and Injuries by Severe Weather Event type.

Number		
TORNADO	96979	TORNADO
EXCESSIVE HEAT	8428	EXCESSIVE HEAT
TSTM WIND	7461	TSTM WIND
FLOOD	7259	FLOOD
LIGHTNING	6046	LIGHTNING
HEAT	3037	HEAT
FLASH FLOOD	2755	FLASH FLOOD
ICE STORM	2064	ICE STORM
THUNDERSTORM WIND	1621	THUNDERSTORM WIND
WINTER STORM	1527	WINTER STORM

Question 2. Across the United States, which types of events have the greatest economic consequences?

The economic consequences are captured as dollar amounts in variables PROPDMG for cost of property damage and CROPDMG for cost of crop damage.

The events with the top 10 greatest economic impacts were:

Table 2: Table of economic cost of Severe Weather Events by Types.

\$		
TORNADO	3312276.7	TORNADO
FLASH FLOOD	1599325.1	FLASH FLOOD
TSTM WIND	1445168.2	TSTM WIND
HAIL	1268289.7	HAIL
FLOOD	1067976.4	FLOOD
THUNDERSTORM WIND	943635.6	THUNDERSTORM WIND

	\$	
LIGHTNING	606932.4	LIGHTNING
THUNDERSTORM WINDS	464978.1	THUNDERSTORM WINDS
HIGH WIND	342014.8	HIGH WIND
WINTER STORM	134699.6	WINTER STORM

Bar Graph of economic impact by year

```
## use tapply again to generate the data for barplot by state
barplot(tapply((data$PROPDMG+data$CROPDMG), list(data$STATE), FUN=sum), xlab="State", ylab = "Total $ economic damage")
```

Barplot of Total economic impact of severe weather events by State

