Reproducible Research II

Nishant Sisingi

October 6, 2016

Coursera : Reproducible Research - Harmful and Economic Impact of Severe Weather Events

Synopsis

Severe weather events can cause both public health and economic problems. In this report we aim to address two questions:

Across the United States, which types of events are most harmful with respect to population health? Across the United States, which types of events have the greatest economic consequences? To answer these questions, we explored the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, property, and crop damage.

From these data, it was established that across the United States, tornado caused the largest numbers of fatalities and injuries and was most harmful to population health. Tornado also caused the highest property damange, while hail caused the highest crop damange. Therefore, tornado and hail had the greatest economic consequences

Data Processing

The data for this project was downloaded from Reproducible Research Coursera course website. It came in the form of a comma-separated-value file compressed via the bzip2 algorithm. National Weather Service Storm Data Documentation, made available through Coursera course website as well, were used to identify variable for the data analysis.

```
library(dplyr)
library(ggplot2)
library(lubridate)
library(tibble)
library(readr)
```

Read the data and select appropriate variables for analysis ie. variables for fatalities, injuries, crop-damage and prop-damage.

```
download.file("https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2",destfile = "repdata-data-StormData.
csv.bz2", method = "curl")
weather_data <- read_csv(bzfile("repdata-data-StormData.csv.bz2"))
weather_data <- as_tibble(weather_data)
weather <- weather_data %>% select(EVTYPE,FATALITIES,INJURIES,PROPDMG,PROPDMGEXP,CROPDMG,CROPDMGEXP)
```

Converting the follwing variables to character, since realtional databases to be used later works better with character variables.

```
weather$PROPDMGEXP <- as.character(weather$PROPDMGEXP)
weather$CROPDMGEXP <- as.character(weather$CROPDMGEXP)
```

Creating proxy tables to map the codes with their multipliers.

```
crop <- tibble(CROPDMGEXP = c("0","1","2","3","4","5","6","7","8","B","h","H","k","K","m","M"), c = c(0:8,1000000000,100,100,1
00,1000,1000000,1000000))

prop <- tibble(PROPDMGEXP = c("0","1","2","3","4","5","6","7","8","B","h","H","K","m","M"), p = c(0:8,1000000000,100,100,100
0,1000000,1000000))
```

Joining both the tables with the parent, mapping the multiplier values

```
weather_new <- left_join(weather,prop,by = "PROPDMGEXP")
weather_new <- left_join(weather_new,crop,by = "CROPDMGEXP")
```

Obtaining the damage dollar values.

```
weather_new <- weather_new %>% mutate(prop_dmg = PROPDMG*p, crop_dmg = CROPDMG*c)
```

Results

In this section, we consider separetely two questions about these data.

Across the United States, which types of events are most harmful with respect to population health?

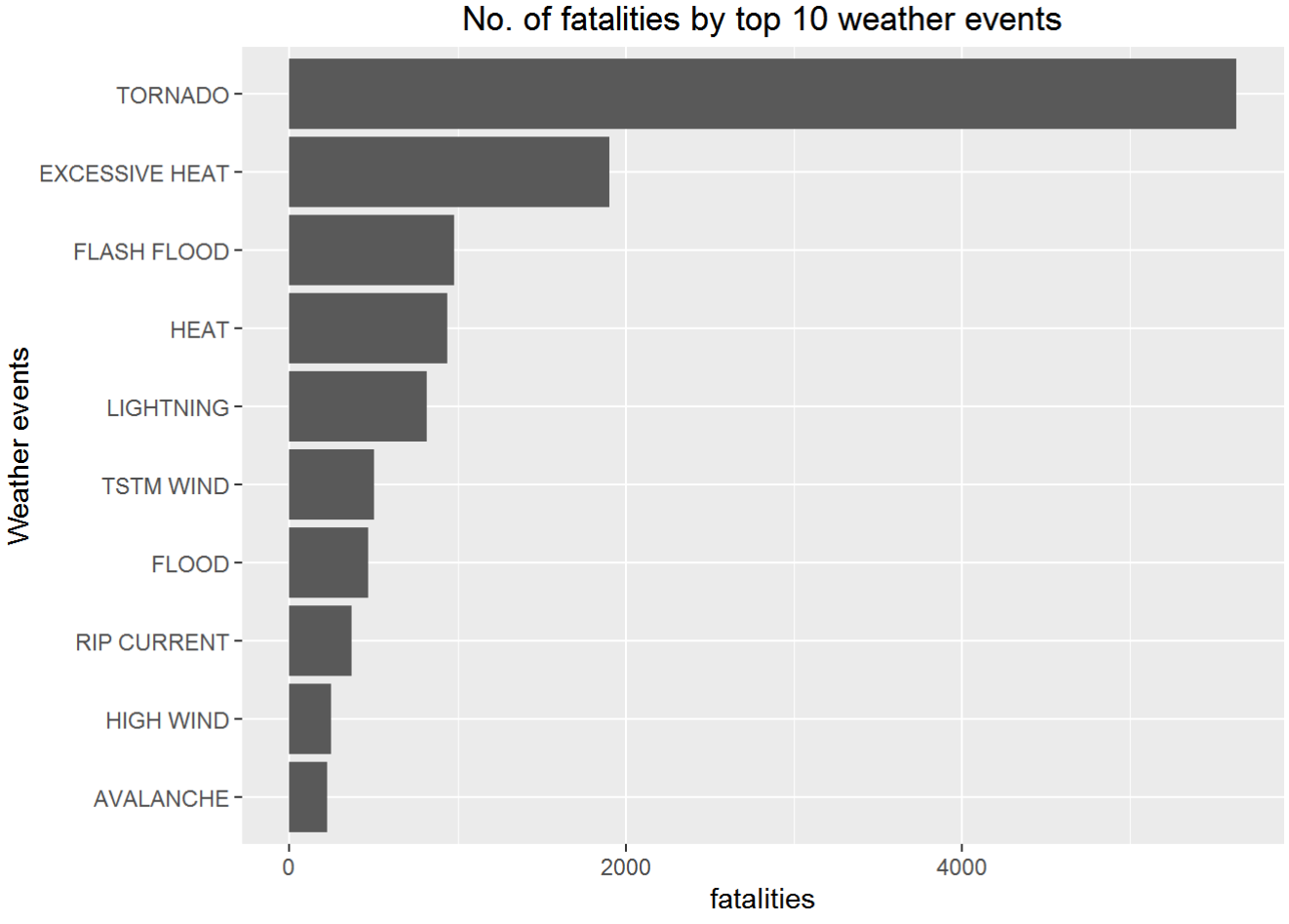
There are 2 variables, FATALITIES and INJURIES, that indicate how harmlful the severe weather events were with respect to population heatlh.

Let's examing the FATALITIES first. All fatalities are summarized by the event type into new table, which is labeled appropriately and sorted in decending order, so that the events with the highest fatality total are at the top of the table.

```
fatalities <- weather_new %>% select(EVTYPE,FATALITIES) %>% group_by(EVTYPE) %>% summarise(fatalities_tot = sum(FATALITIES))
f <- arrange(fatalities,desc(fatalities_tot))
f
```

```
## # A tibble: 977 × 2
##            EVTYPE fatalities_tot
##             <chr>          <dbl>
## 1         TORNADO           5633
## 2  EXCESSIVE HEAT           1903
## 3     FLASH FLOOD            978
## 4            HEAT            937
## 5       LIGHTNING            816
## 6       TSTM WIND            504
## 7           FLOOD            470
## 8     RIP CURRENT            368
## 9       HIGH WIND            248
## 10      AVALANCHE            224
## # ... with 967 more rows
```

```
ggplot(head(f,10)) + geom_bar(aes(x=reorder(EVTYPE,fatalities_tot),y=fatalities_tot), stat = "identity") + coord_flip() + gg
title("No. of fatalities by top 10 weather events") + xlab("Weather events") + ylab("fatalities")
```



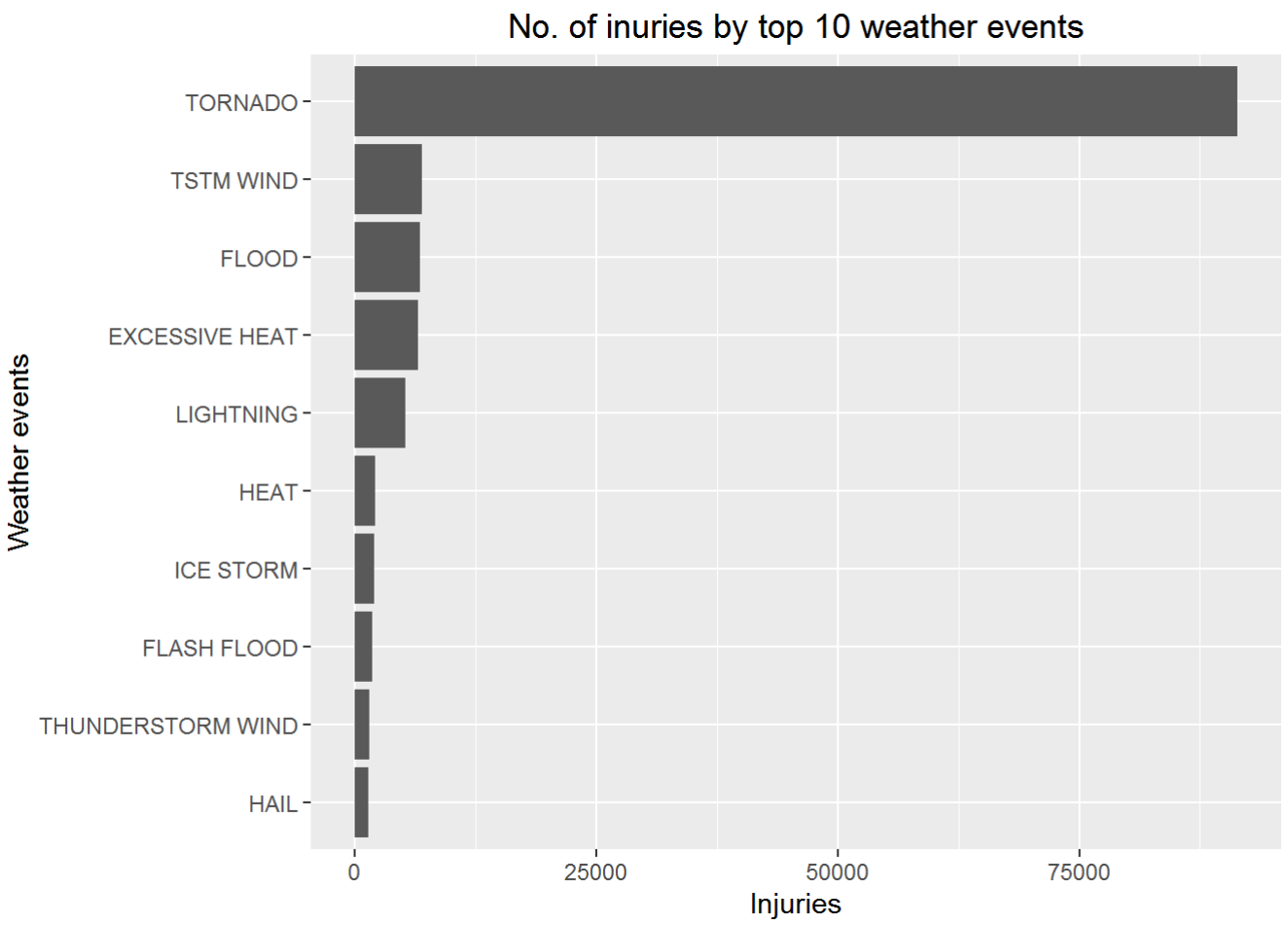
Therefore, across the United States, tornado and excessive heat caused the largest numbers of fatalities and were most harmful with respect to population health.

Let's examing the INJURIES next. All injuries are summarized by the event type into new table, which is labeled appropriately and sorted in decending order, so that the events with the highest injury total are at the top of the table.

```
injuries <- weather_new %>% select(EVTYPE,INJURIES) %>% group_by(EVTYPE) %>% summarise(injuries_tot = sum(INJURIES))
i <- arrange(injuries,desc(injuries_tot))
i
```

```
## # A tibble: 977 × 2
##               EVTYPE injuries_tot
##                <chr>        <dbl>
## 1            TORNADO        91346
## 2          TSTM WIND         6957
## 3              FLOOD         6789
## 4     EXCESSIVE HEAT         6525
## 5           LIGHTNING         5230
## 6               HEAT         2100
## 7          ICE STORM         1975
## 8        FLASH FLOOD         1777
## 9  THUNDERSTORM WIND         1488
## 10             HAIL         1361
## # ... with 967 more rows
```

```
ggplot(head(i,10)) + geom_bar(aes(x=reorder(EVTYPE,injuries_tot),y=injuries_tot), stat = "identity") + coord_flip() + ggtitl
e("No. of inuries by top 10 weather events") + xlab("Weather events") + ylab("Injuries")
```



Therefore, across the United States, tornado and thunderstorm caused the largest numbers of injuries and were most harmful with respect to population health.

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

There are 2 variables, “prop_dmg” (property damage) and “crop_dmg” (crop damage), that indicate which types of severe weather events have the greatest economic consequences.

Let’s examing the prop_dmg first. All property damanges are summarized by the event type into new table, which is labeled appropriately and sorted in decending order, so that the events with the highest property damange total are at the top of the table.

```
prop_dmg <- weather_new %>% select(EVTYPE,prop_dmg) %>% group_by(EVTYPE) %>% summarise(prop_tot = sum(prop_dmg))
arrange(prop_dmg,desc(prop_tot))
```

```
## # A tibble: 977 × 2
##                         EVTYPE   prop_tot
##                          <chr>      <dbl>
## 1  TORNADOES, TSTM WIND, HAIL 1600000000
## 2                   WILD FIRES  624100000
## 3                    HAILSTORM  241000000
## 4              HIGH WINDS/COLD  110500000
## 5               River Flooding  106155000
## 6                  MAJOR FLOOD  105000000
## 7   HURRICANE OPAL/HIGH WINDS  100000000
## 8     WINTER STORM HIGH WINDS   60000000
## 9              HURRICANE EMILY   50000000
## 10         Erosion/Cstl Flood   16200000
## # ... with 967 more rows
```

Therefore, across the United States, tornado,tstm wind,hail caused the highest property damange and had the greatest economic consequences.

Let’s examing the crop_dmg next. All crop damages are summarized by the event type into new table, which is labeled appropriately and sorted in decending order, so that the events with the highest crop damange total are at the top of the table.

```
crop_dmg <- weather_new %>% select(EVTYPE,crop_dmg) %>% group_by(EVTYPE) %>% summarise(crop_tot = sum(crop_dmg))
arrange(crop_dmg,desc(crop_tot))
```

```
## # A tibble: 977 × 2
##                       EVTYPE  crop_tot
##                        <chr>     <dbl>
## 1          EXCESSIVE WETNESS 142000000
## 2    COLD AND WET CONDITIONS  66000000
## 3                Early Frost  42000000
## 4            Damaging Freeze  34130000
## 5                     Freeze  10500000
## 6  HURRICANE OPAL/HIGH WINDS  10000000
## 7             UNSEASONAL RAIN  10000000
## 8             HIGH WINDS/COLD   7000000
## 9          Unseasonable Cold   5100000
## 10              COOL AND WET   5000000
## # ... with 967 more rows
```

Therefore, across the United States, excessive wetness and wet conditions caused the highest crop damange and had the greatest economic consequences.

Conlclusion

From these data, it was established that across the United States, tornado caused the largest numbers of fatalities and injuries and was most harmful to population health. Tornadoes,tstm wind,hail caused the highest property damage, while excessive wetness and wet conditions caused the highest crop damange.