Reproducible Research:Storm data Analysis (Course Project 2)

AbdulAziz Ascandari

2024-06-02

#synopsis

Severe weather events, such as storms, can significantly impact public health and the economy, often resulting in fatalities, injuries, and extensive property damage. Mitigating these outcomes is a primary concern for many communities and municipalities. This analysis examines the effects of severe weather events on public health and economic damage in the United States using NOAA's storm database. The results show that floods and hurricanes/typhoons cause the highest property and crop damage costs, leading to the most substantial economic consequences. Additionally, tornadoes have the greatest impact on population health, causing the highest number of fatalities and injuries.

#Introduction

Severe weather events, such as storms, can have significant public health and economic impacts on communities and municipalities. These events often lead to fatalities, injuries, and extensive property damage. Mitigating these outcomes is a primary concern for many.

This project aims to analyze the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database provides detailed information on major storms and weather events across the United States, including their occurrence, locations, and estimates of associated fatalities, injuries, and property damage.

Data Processing

Software Environment information

First, we have to ascertain some information about software environment where we are going to conduct the analysis. We can get this by using the code below:

"'{r, echo=TRUE}

sessionInfo()

```
## Loading packages

We can now load the required packages for the analysis.
```

```
'''{r,echo=TRUE}

library(R.utils) # load bz2 file
library(data.table)
```

```
library(dplyr)
library(ggplot2)
library(tidyr)
```

Load the dataset

Now we can download and load the bz2 file to run the analysis. Afterwards, we can visualize the first few lines of the file using the *head* function

“‘{r, echo =TRUE}

if (!file.exists(“stormdata.csv.bz2”)) { url <- “https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2” download.file(url, “stormdata.csv.bz2”) bunzip2(“stormdata.csv.bz2”, “stormdata.csv”, remove=FALSE) }

storm <- data.table::fread(“stormdata.csv”, fill=TRUE, header=TRUE) head(storm)

We then look at the column names of the data using the code below:

‘‘‘{r, echo=TRUE}

```
names(storm)
```

The results shows a number of variables, however we shall subset for the variables that we would need for our analysis and convert them to lower case. Thereafter we shall look at the structure using *str* command.

“‘{r, echo = TRUE}

storm2 <- storm %>% select(c(“EVTYPE”,“FATALITIES”,“INJURIES”,“PROPDMG”,“PROPDMGEXP”,“CROPDMG”,“
%>% rename_all(tolower) str(storm2)

Now,the data table shows that we have 902,297 rows and 7 columns. The information on these variables a

-evtype : storm event type
-fatalities: amount of fatalities per event
-injuries : amount of injuries per event
-propdmg : property damage amount
-propdmgexp: property damage in exponents
-cropdmg : crop damage amount
-cropdmgexp: crop damage in exponents

Processing data for population health analysis

‘‘‘{r, echo = TRUE}

```
length(unique(storm$EVTYPE))
```

First, we will select the columns needed for the bar plot and group the data by event type. We will then calculate the sum of both fatalities and injuries for each event type. Next, we will arrange the results in descending order and select the top 10 rows. Finally, we will gather the data and convert it into categorical variables to create a grouped bar plot.

“‘{r, echo = TRUE}

pop_health <- storm2 %>% select(evtype, fatalities, injuries) %>% group_by(evtype) %>% summarize(fatalities = sum(fatalities), injuries = sum(injuries), .groups='drop') %>% arrange(desc(fatalities), desc(injuries)) %>% slice(1:10) %>% gather(key = type, value = value, fatalities, injuries)

```
## Processing data for economic consequences analysis

Here, since the variable PROPDMGEXP is regarding property damage expenses,it can be utilized to denote t

'''{r, echo =TRUE}

unique(storm2$propdmgexp)

unique(storm2$cropdmgexp)
```

Given the messy values for the exponents of property and crop damage costs, we created a function to standardize these values and calculate the costs in millions accordingly.

"'{r, echo =TRUE}

create function to calculate cost

cost <- function(x) { if (x == "H") 1E-4 else if (x == "K") 1E-3 else if (x == "M") 1 else if (x == "B") 1E3 else 1-6 }

```
Aside from the function to calculate cost, the methods is basically much the same for the rest of the a

'''{r, echo = TRUE}

economic <-
    storm2 %>% select("evtype", "propdmg", "propdmgexp", "cropdmg", "cropdmgexp") %>%
    mutate(prop_dmg = propdmg*sapply(propdmgexp, FUN = cost), crop_dmg = cropdmg*sapply(cropdmgexp, FUN
    group_by(evtype) %>%
    summarize(property = sum(prop_dmg), crop = sum(crop_dmg), .groups='drop') %>%
    arrange(desc(property), desc(crop)) %>%
    slice(1:10) %>%
    gather(key = type, value = value, property, crop)
```

Results

This section shall look at the results to the questions that the analysis seeks to answer after the data processing steps.

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

"'{r, echo = TRUE} ggplot(data=pop_health, aes(reorder(evtype, -value), value, fill=type)) + geom_bar(position = "dodge", stat="identity") + labs(x="Event Type", y="Count") + theme_bw()

+ theme(axis.text.x = element_text(angle = 20, vjust=0.7)) + ggtitle("Total Number of Fatalities and Injuries of top 10 storm event types") + scale_fill_manual(values=c("blue", "yellow"))

```
Interpretation : Based on the bar plot, it is evident that tornadoes have the highest impact on the popu

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

```{r, echo = TRUE}

ggplot(data=economic, aes(reorder(evtype, -value), value, fill=type)) +
  geom_bar(position = "dodge", stat="identity") +
  labs(x="Event Type", y="Count (millions)") +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 25, vjust=0.5)) +
  ggtitle("Total Cost of Property and Crop Damage by top 10 storm event types") +
  scale_fill_manual(values=c("darkgreen", "red"))
```

Interpretation : From the bar plot, Floods and Hurricanes/Typhoons have highest property and crop damage costs, thus resulting in the biggest economic consequences.

Conclusion

Based on the analysis, resources should be prioritized towards addressing tornadoes to enhance public safety and health. This can be achieved by investing in better infrastructure and early warning systems. Additionally, to mitigate the impacts of hurricanes and typhoons, there should be increased funding for innovative solutions aimed at developing robust systems and infrastructure. These improvements are essential to protect properties and crops, thereby minimizing potential damages.
```