Reproducible Research - Peer graded assignment 2

The report consists of an analysis of weather data in the US in terms of Helth and economy impact. Specifically, the goal is to inspect which weather events cause more injuries, fatalities, and economic loss.The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

**library**(knitr)

opts\_chunk$set(echo = TRUE)

The next libraries are needed for the code to run.

**library**(ggplot2)

**library**(grid)

**library**(gridExtra)

Reading data into workspace.

rm(list=ls())

**if**(!file.exists("./week4")){dir.create("./week4")}

fileUrl = "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"

fileDest = "./week4/repdata-data-StormData.csv.bz2"

**if**(!file.exists(fileDest))

{

download.file(fileUrl, fileDest)

}

**if**(!exists("projectData")){projectData <- read.csv(bzfile(fileDest))}

Data Processing

Prepping health impact data for presentation.

1. Taking in only events where there was at least 1 injury or fatality.
2. Collecting the data into two different data frames, corresponding to the consequence.
3. Aggregating to get the sum of impacts for each event.
4. Re-ordering the data.

fatalitiesData <- subset(projectData, FATALITIES > 0)

injuriesData <- subset(projectData, INJURIES > 0)

fatalitiesSum <- aggregate(FATALITIES ~ EVTYPE, fatalitiesData, sum, na.rm = TRUE)

injuriesSum <- aggregate(INJURIES ~ EVTYPE, injuriesData, sum, na.rm = TRUE)

fatalitiesSum <- fatalitiesSum[order(fatalitiesSum$FATALITIES, decreasing = TRUE),]

injuriesSum <- injuriesSum[order(injuriesSum$INJURIES, decreasing = TRUE),]

Prepping economic impact data.

1. As before, taking in events where the damage was higher than zero.
2. Replacing charecter data with numerical.
3. Adding a separate column with the total damage value.
4. Aggregating and re-ordering.

economyData <- subset(projectData, PROPDMG > 0 | CROPDMG > 0)

economyData$PROPDMGEXP <- sub("h|H", "2", economyData$PROPDMGEXP)

economyData$PROPDMGEXP <- sub("k|K", "3", economyData$PROPDMGEXP)

economyData$PROPDMGEXP <- sub("m|M", "6", economyData$PROPDMGEXP)

economyData$PROPDMGEXP <- sub("b|B", "9", economyData$PROPDMGEXP)

economyData$CROPDMGEXP <- sub("h|H", "2", economyData$CROPDMGEXP)

economyData$CROPDMGEXP <- sub("k|K", "3", economyData$CROPDMGEXP)

economyData$CROPDMGEXP <- sub("m|M", "6", economyData$CROPDMGEXP)

economyData$CROPDMGEXP <- sub("b|B", "9", economyData$CROPDMGEXP)

economyData$propDMGQuant <- as.numeric(economyData$PROPDMG) \* 10 ^ as.numeric(economyData$PROPDMGEXP)

## Warning: NAs introduced by coercion

economyData$cropDMGQuant <- as.numeric(economyData$CROPDMG) \* 10 ^ as.numeric(economyData$CROPDMGEXP)

## Warning: NAs introduced by coercion

economyData$totalLoss <- economyData$propDMGQuant + economyData$cropDMGQuant

economySum <- aggregate(totalLoss ~ EVTYPE, economyData, sum, na.rm = TRUE)

economySum <- economySum[order(economySum$totalLoss, decreasing = TRUE),]

economySum$totalLoss <- economySum$totalLoss/10e6

Plotting

There is an exhaustive list of weather events that had an impact in the US. For presentation purposes, the ten most influential were chosen.

fatalPlot <- ggplot(data=head(fatalitiesSum,10), aes(x=reorder(EVTYPE, -FATALITIES), y=FATALITIES)) +

geom\_bar(fill="goldenrod",stat="identity") +

ylab("Total number of fatalities") + xlab("Event type") +

theme(legend.position="none") + theme(axis.text.x = element\_text(size = 5, angle = 45, hjust = 1))

injPlot <- ggplot(data=head(injuriesSum,10), aes(x=reorder(EVTYPE, -INJURIES), y=INJURIES)) +

geom\_bar(fill="darkblue",stat="identity") +

ylab("Total number of injuries") + xlab("Event type") +

theme(legend.position="none") + theme(axis.text.x = element\_text(size = 5, angle = 45, hjust = 1))

pushViewport(viewport(layout = grid.layout(2, 2, heights = unit(c(1, 4), "null"))))

grid.text("Ten Most Health-Influening Weather Events in the US", vp = viewport(layout.pos.row = 1,

layout.pos.col = 1:2))

print(fatalPlot, vp = viewport(layout.pos.row = 2, layout.pos.col = 1))

print(injPlot, vp = viewport(layout.pos.row = 2, layout.pos.col = 2))

We see that tornadoes cause more fatalities and injuries in a rate that is significantly greater than the rest of events. However, what folloes tornadoes in both parameters vary.

economyPlot <- ggplot(data=head(economySum,10), aes(x=reorder(EVTYPE, -totalLoss), y=totalLoss)) +

geom\_bar(fill="darkblue",stat="identity") +

ylab("Loss (1M USD)") + xlab("Event type") +

theme(legend.position="none") + theme(axis.text.x = element\_text(size = 5, angle = 45, hjust = 1)) +

ggtitle("Ten Most Economically Influential Weather Events in the US")

grid.arrange(economyPlot, ncol = 1, nrow = 1)

From the analysis it is obvoius that flood is the most destructive weather events, as documented in the data.