Reproducible Research Project 2

James Hancock

July 3, 2016

Synopsis:

The basic goal of this assignment is to explore the NOAA Storm Database and answer some basic questions about severe weather events. You must use the database to answer the questions below and show the code for your entire analysis. Your analysis can consist of tables, figures, or other summaries. You may use any R package you want to support your analysis.

Storms and other severe weather events can cause both public health and economic problems for communities and municipalities. Many severe events can result in fatalities, injuries, and property damage, and preventing such outcomes to the extent possible is a key concern.

This project involves exploring the U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database. This database tracks characteristics of major storms and weather events in the United States, including when and where they occur, as well as estimates of any fatalities, injuries, and property damage.

Data analysis must address the following questions: 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? I answer this on both fatalities and injuries. 2. Across the United States, which types of events have the greatest economic consequences? I measure this in USD

Findings: 1. Tornados cause the most fatalities in North America. 2. Tornados also cause the most injuries in North America. 3. Flooding causes the most property damage in $

Data Processing:

load data set into R.

```
stormdata <- read.csv(bzfile("data/StormData.csv.bz2"))
}
```

Compile data on fatalities. Pare dataset to Event Type and Fatalities columns, and only rows with fatalities. Group by event type and take the 5 events with the most fatalities

```
{
fatalityData <- (subset(stormdata, FATALITIES > 0))[c(8, 23)]
fatalityGroupedData <- aggregate(FATALITIES ~ EVTYPE, data = fatalityData, FUN = "sum", na.rm = TRUE)
fatalityGroupedData <- fatalityGroupedData[order(fatalityGroupedData$FATALITIES, decreasing=TRUE), ]
fatalityGroupedData <- fatalityGroupedData[1:5, ]
fatalityGroupedData$EVTYPE <- factor(fatalityGroupedData$EVTYPE, levels=fatalityGroupedData$EVTYPE)
}
```

Compile data on injuries. Pare dataset to Event Type and Injuries columns, and only rows with injuries. Group by event type and take the 5 events with the most injuries.

```
{
damageAmount <- function(amount, magnitude)
}
```

```
{
returnAmount <- 0
if (toupper(magnitude)[1]=="K")
  {
    returnAmount <- (amount * 1000)
  }
  if (toupper(magnitude)[1]=="M")
  {
    returnAmount <- (amount * 1000000)
  }
  if (toupper(magnitude)[1]=="B")
  {
    returnAmount <- (amount * 1000000000)
  }
  return(returnAmount)
}

{
damageData <- (subset(stormdata, PROPDMG > 0 | CROPDMG > 0))[c(8, 25, 26, 27, 28)]
damageData$DamageAmount <- ((mapply(damageAmount, damageData$PROPDMG, damageData$PROPDMGEXP)) + (mapply

damageGroupedData <- aggregate(DamageAmount ~ EVTYPE, data = damageData, FUN = "sum", na.rm = TRUE)
damageGroupedData <- damageGroupedData[order(damageGroupedData$DamageAmount, decreasing=TRUE), ]
damageGroupedData <- damageGroupedData[1:5, ]
damageGroupedData$EVTYPE <- factor(damageGroupedData$EVTYPE, levels=damageGroupedData$EVTYPE)
}
```

Results:

Plot fatalities

```
{
library(ggplot2)
ggplot(fatalityGroupedData, aes(x=EVTYPE, y=FATALITIES)) +
geom_bar(stat="identity") +
xlab("Types") +
ylab("Fatalities") +
ggtitle("Total Fatalities By Weather Events")
}
```

Plot Injuries

```
{
library(ggplot2)
ggplot(injuryGroupedData, aes(x=EVTYPE, y=INJURIES)) +
geom_bar(stat="identity") +
xlab("Types") +
ylab("Injuries") +
ggtitle("Total Injuries By Weather Events")
}
```

Plot Damages Caused in $

```
{
library(ggplot2)
ggplot(damageGroupedData, aes(x=EVTYPE, y=DamageAmount)) +
geom_bar(stat="identity") +
xlab("Types") + ylab("Damage Amount (In Dollars)") +
ggtitle("Total Damage Amount By Weather Events")
}
```