This is for Assignment 2 of Reproducible Research

Introduction

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.The section below loads the data and checks the initial length of the unique number of EVTYPE(Eent types)

```
library("data.table")
```

```
## Warning: package 'data.table' was built under R version 3.1.1
```

```
stData <- data.table(read.csv("data/repdata_data_StormData.csv", header = T,
    stringsAsFactors = F))
initial <- length(unique(stData$EVTYPE))
```

In the section below, the event type data is scanned and grouped. This will inturn reduce the number of EVTYPE

Data Processing

```
stData[grep("Summary", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Ignore"
stData[grep("Low Tide", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Astronomical Low Tide"
stData[grep("Avalanche", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Avalanche"
stData[grep("Blizzard", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Blizzard"
stData[grep("Coastal Flood", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Coastal Flood"
stData[grep("Debris Flow", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Debris Flow"
stData[grep("Dense Fog", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Dense Fog"
stData[grep("Dense Smoke", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Dense Smoke"
stData[grep("Drought", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Drought"
stData[grep("Dust Devil", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Dust Devil"
stData[grep("Dust Storm", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Dust Storm"
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
stData[grep("Excessive Heat", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Excessive Heat"
stData[grep("(Extreme|Unusually|Record|Bitter|Excess|Unseas|Abnor).*(Cold|Chill|Cool)",
    stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Extreme cold/Wind Chill"
stData[grep("(Cold|Wind).*chill", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Cold/Wind Chill"
stData[grep("^Flood", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Flash Flood"
stData[grep("Freezing Fog", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Freezing Fog"
stData[grep("Frost/Freeze", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Frost/Freeze"
stData[grep("Funnel Cloud", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Funnel Cloud"
stData[grep("Hail", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Hail"
stData[grep("Heat", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Heat"
stData[grep("Heavy Rain", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Heavy Rain"
stData[grep("Heavy Snow", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Heavy Snow"
stData[grep("High Surf", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "High Surf"
stData[grep("High Wind", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "High Wind"
stData[grep("Hurricane|Typhoon", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Hurricane/Typhoon"
stData[grep("Ice Storm", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Ice Storm"
stData[grep("Lakeshore Flood", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Lakeshore Flood"
stData[grep("Lake-Effect Snow", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Lake-Effect Snow"
stData[grep("Lightning", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Lightning"
stData[grep("Marine Hail", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Marine Hail"
stData[grep("Marine High Wind", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Marine High Wind"
stData[grep("Marine Strong Wind", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Marine Strong Wind"
stData[grep("Marine.*(Thunderstorm|tstm).*Wind ", stData$EVTYPE, ignore.case = TRUE),
    ]$EVTYPE = "Marine Thunderstorm Wind"
stData[grep("Rip Current", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Rip Current"
stData[grep("Seiche", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Seiche"
stData[grep("Sleet", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Sleet"
stData[grep("Storm Surge/Tide", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Storm Surge/Tide"
stData[grep("Strong Wind", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Strong Wind"
stData[grep("^Thund|^ *tstm", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Thunderstorm Wind"
stData[grep("Tornado", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Tornado"
stData[grep("Tropical Depression", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Tropical Depression"
stData[grep("Tropical Storm", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Tropical Storm"
stData[grep("Storm Surge/Tide", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Storm Surge/Tide"
stData[grep("Tsunami", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Tsunami"
stData[grep("Volcanic ASh", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Volcanic ASh"
stData[grep("Waterspout", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Waterspout"
stData[grep("Wildfire", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Wildfire"
stData[grep("Winter Storm", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Winter Storm"
stData[grep("Winter Weather", stData$EVTYPE, ignore.case = TRUE), ]$EVTYPE = "Winter Weather"
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
final <- length(unique(stData$EVTYPE))
```

The initial value of unique EVTYPE was more than 1000 and odd and the final value is 526. This easies the analysis of the data.

AS per the documentation there are seven variables related to these questions in the Assignment. They are:

EVTYPE –> representing the event type FATALITIES –> measure of harm to human health INJURIES –> measure of harm to human health PROPDMG –> magnitude of measure inturn ecominc damage in USD PROPDMGEXP –> measure of magnitude of property damage CROPDMG –> measure of crop damage inturn econmic damage in USD CROPDMGEXP –> measure of magnitude of crop damage

A new data frame is created with these 7 variables .

```
data2 <- data.frame(stData$EVTYPE, stData$FATALITIES, stData$INJURIES, stData$PROPDMG,
    stData$PROPDMGEXP, stData$CROPDMG, stData$CROPDMGEXP, stringsAsFactors = FALSE)
# replace missing values with 0's
data2$stData.FATALITIES[(data2$stData.FATALITIES == "")] <- 0
data2$stData.INJURIES[(data2$stData.INJURIES == "")] <- 0
data2$stData.PROPDMG[(data2$stData.PROPDMG == "")] <- 0
data2$stData.CROPDMG[(data2$stData.CROPDMG == "")] <- 0

# use a uniform numbers-only damage magnitude representation
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "")] <- 0
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "+") | (data2$stData.PROPDMGEXP ==
    "-") | (data2$stData.PROPDMGEXP == "?")] <- 1
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "H") | (data2$stData.PROPDMGEXP ==
    "h")] <- 2
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "K") | (data2$stData.PROPDMGEXP ==
    "k")] <- 3
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "M") | (data2$stData.PROPDMGEXP ==
    "m")] <- 6
data2$stData.PROPDMGEXP[(data2$stData.PROPDMGEXP == "B") | (data2$stData.PROPDMGEXP ==
    "b")] <- 9

data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "")] <- 0
data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "+") | (data2$CROPDMGEXP ==
    "-") | (data2$CROPDMGEXP == "?")] <- 1
data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "H") | (data2$CROPDMGEXP ==
    "h")] <- 2
data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "K") | (data2$CROPDMGEXP ==
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
    "k")] <- 3
data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "M") | (data2$CROPDMGEXP ==
    "m")] <- 6
data2$stData.CROPDMGEXP[(data2$stData.CROPDMGEXP == "B") | (data2$CROPDMGEXP ==
    "b")] <- 9

# convert magnitude variables into integers
data2$stData.PROPDMGEXP <- as.integer(as.character(data2$stData.PROPDMGEXP))
```

```
## Warning: NAs introduced by coercion
```

```
data2$stData.CROPDMGEXP <- as.integer(as.character(data2$stData.CROPDMGEXP))
```

```
## Warning: NAs introduced by coercion
```

We are using the multiplot function from Cookbook from R for plotting two graphs in a single plot.

```
# Multiple plot function ggplot objects can be passed in ..., or to plotlist
# (as a list of ggplot objects) - cols: Number of columns in layout -
# layout: A matrix specifying the layout. If present, 'cols' is ignored.  If
# the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE), then
# plot 1 will go in the upper left, 2 will go in the upper right, and 3 will
# go all the way across the bottom.
multiplot <- function(..., plotlist = NULL, file, cols = 1, layout = NULL) {
    require(grid)

    # Make a list from the ... arguments and plotlist
    plots <- c(list(...), plotlist)

    numPlots = length(plots)

    # If layout is NULL, then use 'cols' to determine layout
    if (is.null(layout)) {
        # Make the panel ncol: Number of columns of plots nrow: Number of rows
        # needed, calculated from # of cols
        layout <- matrix(seq(1, cols * ceiling(numPlots/cols)), ncol = cols,
            nrow = ceiling(numPlots/cols))
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
    }

    if (numPlots == 1) {
        print(plots[[1]])

    } else {
        # Set up the page
        grid.newpage()
        pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

        # Make each plot, in the correct location
        for (i in 1:numPlots) {
            # Get the i,j matrix positions of the regions that contain this subplot
            matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

            print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row, layout.pos.col = matchidx$col))
        }
    }
}
```

Results There are two questions are addressed in this project: 1. Across the United States, which types of events (as indicated in the EVTYPE variable) are most harmful with respect to population health? 2. Across the United States, which types of events have the greatest economic consequences?

Events most harmful with respect to population health

Take the aggregate the fatalities, injuries and economic damage by event type and plot.

Facilities

```
library(ggplot2)

tFatalities <- sort(tapply(as.numeric(data2$stData.FATALITIES), data2$stData.EVTYPE,
    sum), decreasing = T)
```

```
## Warning: NAs introduced by coercion
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
tInjuries <- sort(tapply(as.numeric(data2$stData.INJURIES), data2$stData.EVTYPE,
    sum), decreasing = T)
```

```
## Warning: NAs introduced by coercion
```

```
names <- as.vector(names(head(tFatalities)))
vals <- as.vector(head(tFatalities))
df <- data.frame(names, vals)

# ggplot(data=df, aes(x=df$names,y=head(tFatalities))) + geom_bar(stat =
# 'identity') +xlab('Event Type') + ylab('# of Fatalities') + ggtitle('NOAA
# Top 6 Fatality Counts, 1950-2011') +theme(axis.text.x = element_text(angle
# = 90, hjust = 1))

# (df$names, head(tFatalities))
```

Injuries

```
tInjuries <- sort(tapply(as.numeric(data2$stData.INJURIES), data2$stData.EVTYPE,
    sum), decreasing = T)
```

```
## Warning: NAs introduced by coercion
```

```
head(tInjuries)
```

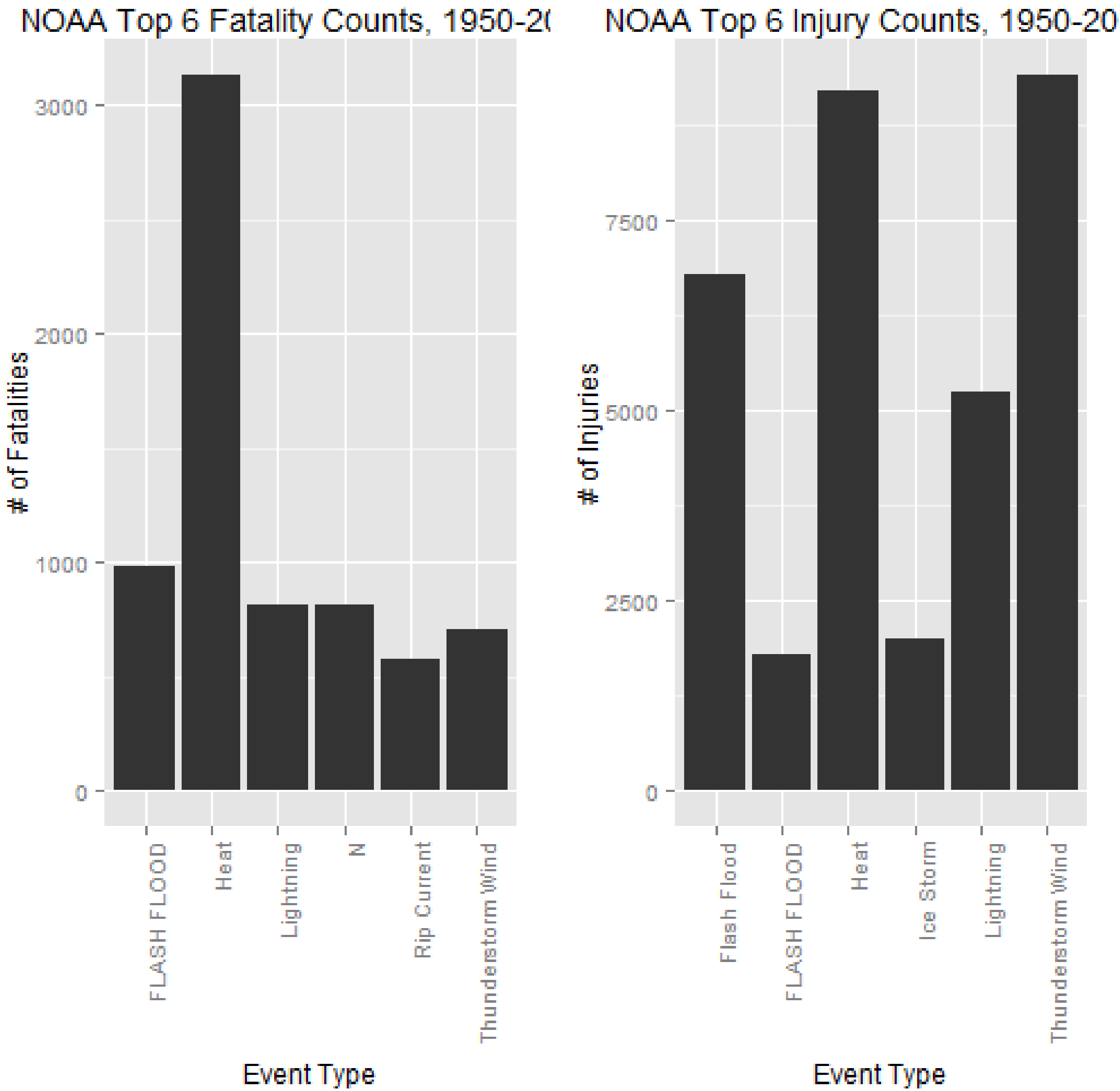
```
## Thunderstorm Wind              Heat       Flash Flood         Lightning
##              9413              9209              6806              5232
##         Ice Storm       FLASH FLOOD
##              1992              1777
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

```
names2 <- as.vector(names(head(tInjuries)))
vals2 <- as.vector(head(tInjuries))
df2 <- data.frame(names2, vals2)

p1 <- ggplot(data = df, aes(x = df$names, y = head(tFatalities))) + geom_bar(stat = "identity") +
    xlab("Event Type") + ylab("# of Fatalities") + ggtitle("NOAA Top 6 Fatality Counts, 1950-2011") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1))


p2 <- ggplot(data = df2, aes(x = df2$names2, y = head(tInjuries))) + geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1)) + xlab("Event Type") +
    ylab("# of Injuries") + ggtitle("NOAA Top 6 Injury Counts, 1950-2011")

multiplot(p1, p2, cols = 2)
```

```
## Loading required package: grid
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com



EVents that have the greatest economic consequences

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

Add the property and crop damage data to get the total economic impact per event record.

```
damage_total <- as.numeric(data2$stData.PROPDMG) * 10^data2$stData.PROPDMGEXP +
    as.numeric(data2$stData.CROPDMG) * 10^data2$stData.CROPDMGEXP
```

```
## Warning: NAs introduced by coercion
## Warning: NAs introduced by coercion
```

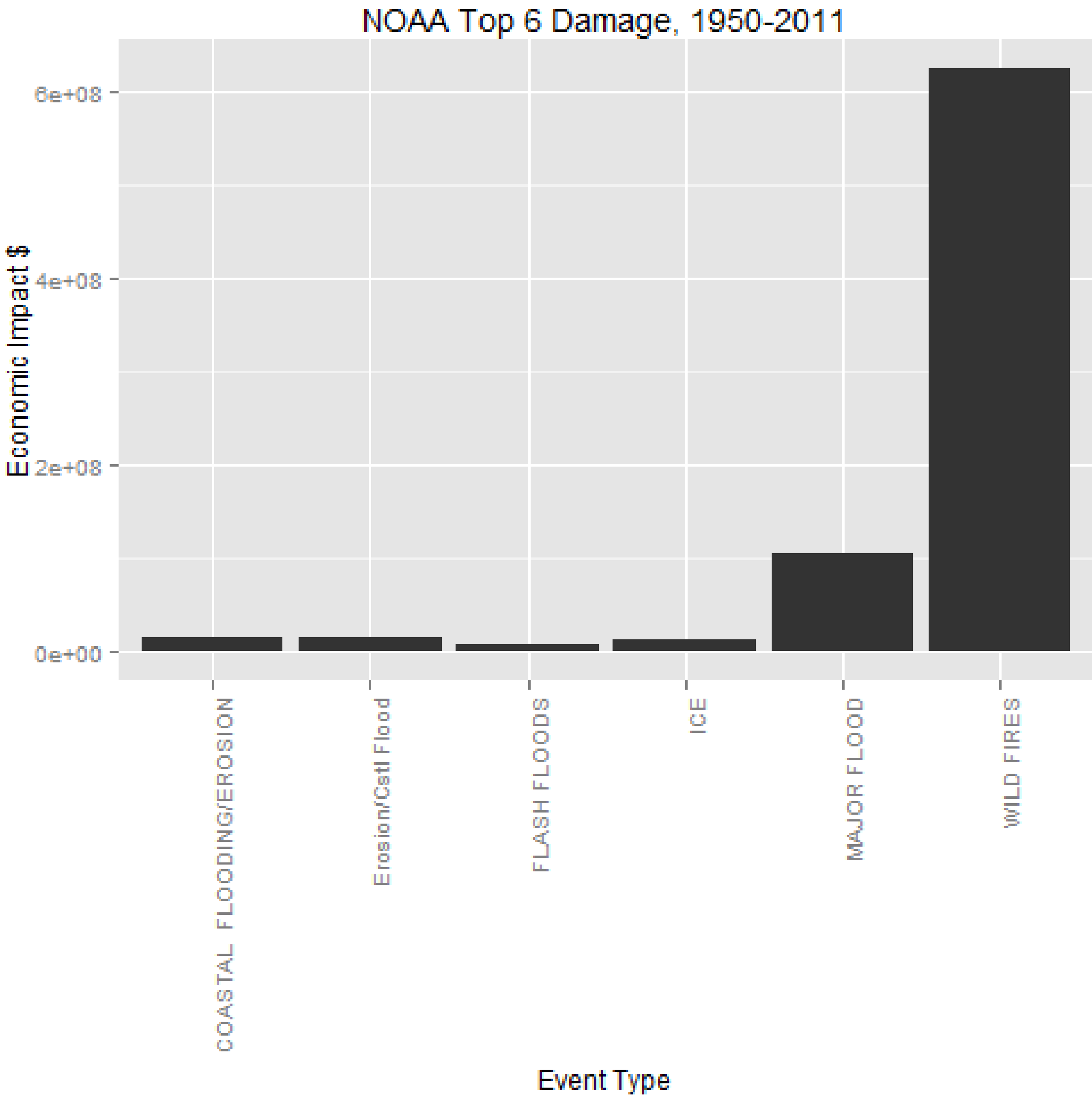
```
data2 <- cbind(data2, damage_total)

tEconomicImpact <- sort(tapply(data2$damage_total, data2$stData.EVTYPE, sum),
    decreasing = T)
head(tEconomicImpact)
```

```
##                WILD FIRES                MAJOR FLOOD
##                 624100000                  105000000
##        Erosion/Cstl Flood COASTAL  FLOODING/EROSION
##                  16200000                   15000000
##                       ICE               FLASH FLOODS
##                  12655000                    8728500
```

Now the graphical representation of this is as below:

```
names <- as.vector(names(head(tEconomicImpact)))
vals <- as.vector(head(tEconomicImpact))
df <- data.frame(names, vals)
ggplot(data = df, aes(x = df$names, y = df$vals)) + geom_bar(stat = "identity") +
    theme(axis.text.x = element_text(angle = 90, hjust = 1)) + xlab("Event Type") +
    ylab("Economic Impact $") + ggtitle("NOAA Top 6 Damage, 1950-2011")
```

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com



Synomsys

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com

EventTYpe Heat has caused the greatest number of Fatalities followed by FLASH FLOOD. Event Type Thunderstorm Wind has cuased the greatest number of Injuries followed by Heat

Event Type WILD FIRES has caused the greatest economic damage 6e+08 USD followed by MAJOR FLOOD.

PRO version Are you a developer? Try out the HTML to PDF API pdfcrowd.com