Analysis of NOAA Storm Database Synopsis

This assignment analyzes the US NOAA Storm Database to address the following two questions:

1. Across the United States, which types of events are most harmful with respect to population health?
2. Across the United States, which types of events have the greatest economic consequences?

Environment

- Windows 8.1 64 bit
- R v3.1.1
- Rstudio v0.98.1062.0

Data Processing

```
library(stringr)
library(plyr)
library(dplyr)
library(tidyr)
library(ggplot2)
```

```
fileurl <- "https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2"
fileurl <- sub("^https", "http", fileurl)

filename <- "StormData.csv.bz2"
if (!file.exists(filename)) {download.file(fileurl, filename)}

stormdata <- read.csv(bzfile(filename), stringsAsFactors=FALSE)
dim(stormdata)
```

```
## [1] 902297     37
```

```
stormdf <- tbl_df(stormdata) %>%
    filter(FATALITIES > 0 | INJURIES > 0 | PROPDMG > 0 | CROPDMG > 0) %>%
    mutate(YEAR = as.integer(str_extract(BGN_DATE, "\\d{4}"))) %>%
    filter(YEAR >= 1996) %>%
    select(EVTYPE, FATALITIES, INJURIES, PROPDMG, PROPDMGEXP, CROPDMG, CROPDMGEXP)

length(unique(stormdf$EVTYPE))
```

```
## [1] 222
```

```
clean_evtype <- function(x) {
    x <- toupper(str_trim(x))

    if (grepl("ACCIDENT|DROWNING", x)) "OTHER"
    else if (grepl("HIGH TIDE", x)) "COASTAL FLOOD"
    else if (grepl("COASTAL ?STORM", x)) "STORM SURGE/TIDE"
    else if (grepl("COASTAL|CSTL|BEACH", x)) "COASTAL FLOOD"
    else if (grepl("EXTREME", x)) "EXTREME COLD/WIND CHILL"
    else if (grepl("COLD", x)) "COLD/WIND CHILL"
    else if (grepl("FREEZING", x)) "FREEZING FOG"
    else if (grepl("FOG", x)) "DENSE FOG"
    else if (grepl("BLOWING DUST", x)) "DUST STORM"
    else if (grepl("^HEAT$", x)) "HEAT"
    else if (grepl("HEAT", x)) "EXCESSIVE HEAT"
    else if (grepl("THERMIA|WARM", x)) "HEAT"
    else if (grepl("FLASH|DAM BREAK|HIGH WATER", x)) "FLASH FLOOD"
    else if (grepl("LAKESHORE FLOOD", x)) "LAKESHORE FLOOD"
    else if (grepl("FLOOD|FLD", x)) "FLOOD"
    else if (grepl("MARINE T[^ ]*M\\b", x)) "MARINE THUNDERSTORM WIND"
    else if (grepl("MARINE", x)) x
    else if (grepl("ICE STORM", x)) "ICE STORM"
    else if (grepl("FROST|FREEZE|ICE|ICY", x)) "FROST/FREEZE"
    else if (grepl("LAKE.*SNOW", x)) "LAKE-EFFECT SNOW"
    else if (grepl("SNOW", x)) "HEAVY SNOW"
    else if (grepl("RAIN", x)) "HEAVY RAIN"
    else if (grepl("HAIL", x)) "HAIL"
    else if (grepl("SURF", x)) "HIGH SURF"
    else if (grepl("WINTER STORM", x)) "WINTER STORM"
    else if (grepl("WINTER|WINTRY|GLAZE|PRECIP", x)) "WINTER WEATHER"
    else if (grepl("SLIDE|SLUMP", x)) "DEBRIS FLOW"
    else if (grepl("RIVER", x)) "FLOOD"
    else if (grepl("GUSTY", x)) "STRONG WIND"
    else if (grepl("NON.?TSTM", x)) "OTHER"
    else if (grepl("THUNDERSTORM|TSTM|BURST", x)) "THUNDERSTORM WIND"
    else if (grepl("STRONG WIND", x)) "STRONG WIND"
    else if (grepl("WIND", x)) "HIGH WIND"
    else if (grepl("HURRICANE|TYPHOON", x)) "HURRICANE(TYPHOON)"
    else if (grepl("RIP", x)) "RIP CURRENT"
    else if (grepl("TORNADO|LANDSPOUT", x)) "TORNADO"
    else if (grepl("TROPICAL", x)) x
    else if (grepl("STORM|TIDE", x)) "STORM SURGE/TIDE"
    else if (grepl("FIRE", x)) "WILDFIRE"
    else if (grepl("WAVE|SWELL| SEAS$", x)) "HIGH SURF"
    else x
}
```

```
stormdf$EVTYPE <- sapply(stormdf$EVTYPE, clean_evtype)

length(unique(stormdf$EVTYPE))
```

```
## [1] 47
```

```
symbol2value <- function(x) {
    if (x == "K") 1e3
    else if (x == "M") 1e6
    else if (x == "B") 1e9
    else 1
}
```

```
stormdf$PROPDMG <- stormdf$PROPDMG * sapply(stormdf$PROPDMGEXP, symbol2value)
stormdf$CROPDMG <- stormdf$CROPDMG * sapply(stormdf$CROPDMGEXP, symbol2value)

stormdf <- stormdf %>%
    select(-PROPDMGEXP, -CROPDMGEXP) %>%
    group_by(EVTYPE) %>%
    summarize(FATALITIES = sum(FATALITIES), INJURIES = sum(INJURIES),
              PROPDMG = sum(PROPDMG), CROPDMG = sum(CROPDMG)) %>%
    mutate(TOTHEALTHDMG = FATALITIES + INJURIES, TOTECONOMICDMG = PROPDMG + CROPDMG)
```

Results

Types of events that are most harmful with respect to population health:

```
top_health_dmg <- stormdf %>%
    select(-c(PROPDMG, CROPDMG, TOTECONOMICDMG)) %>%
    filter(FATALITIES >= 275 | INJURIES >= 875) %>%
    arrange(desc(FATALITIES), desc(TOTHEALTHDMG))

top_health_dmg
```

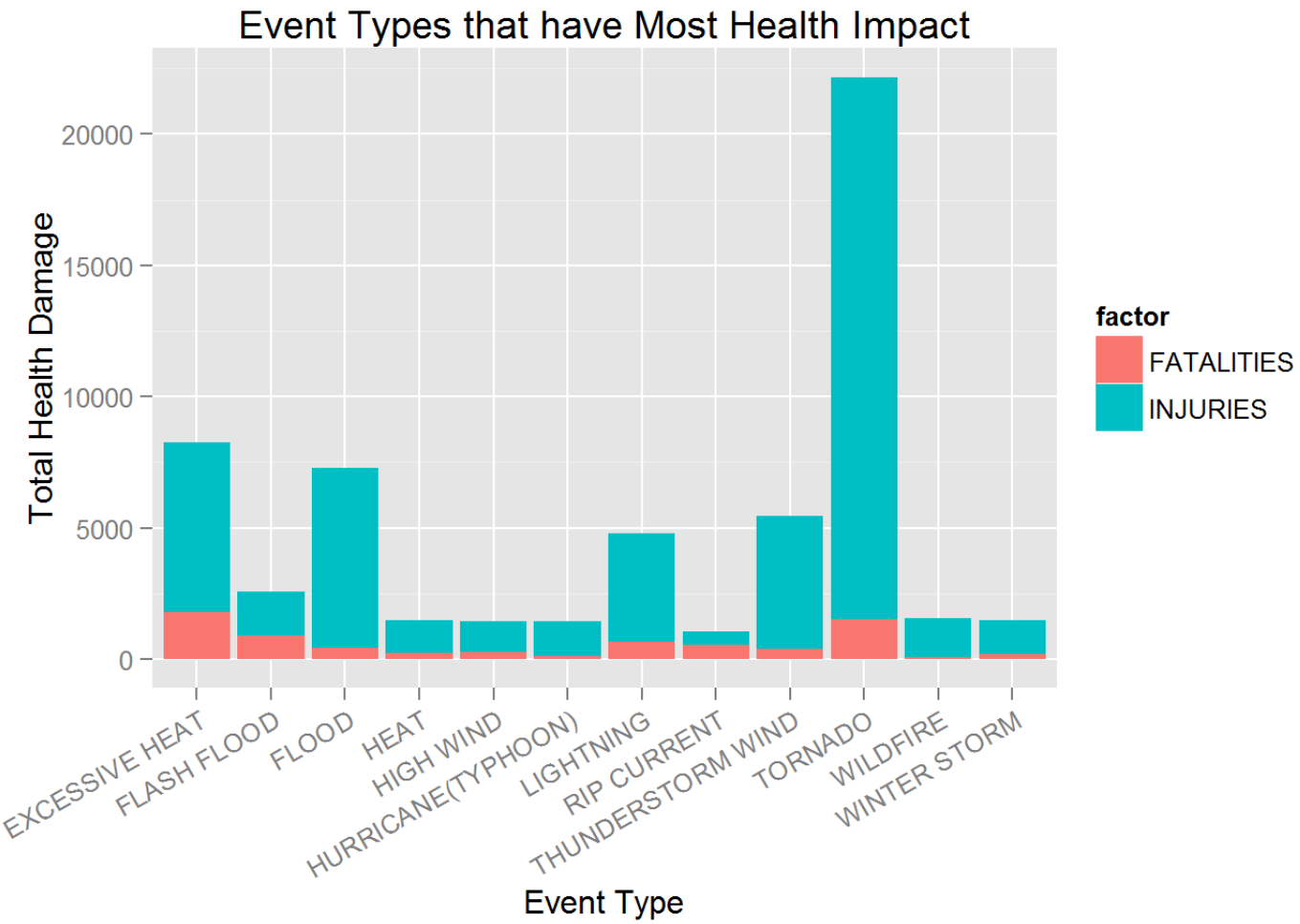
```
## Source: local data frame [12 x 4]
##
##                EVTYPE FATALITIES INJURIES TOTHEALTHDMG
## 1      EXCESSIVE HEAT       1799     6461         8260
## 2             TORNADO       1511    20667        22178
## 3         FLASH FLOOD        890     1674         2564
## 4           LIGHTNING        651     4141         4792
## 5         RIP CURRENT        542      503         1045
## 6               FLOOD        444     6839         7283
## 7   THUNDERSTORM WIND        377     5059         5436
## 8           HIGH WIND        255     1174         1429
## 9                HEAT        245     1241         1486
## 10       WINTER STORM        191     1292         1483
## 11 HURRICANE(TYPHOON)        125     1328         1453
## 12           WILDFIRE         87     1458         1545
```

```
fatalities <- select(top_health_dmg, EVTYPE, FATALITIES)
colnames(fatalities) <- c("EVTYPE", "TOTHEALTHDMG")

injuries <- select(top_health_dmg, EVTYPE, INJURIES)
colnames(injuries) <- c("EVTYPE", "TOTHEALTHDMG")

# health_dmg <- rbind(fatalities, injuries)
health_dmg = rbind(mutate(fatalities, factor = rep("FATALITIES",12)),
                   mutate(injuries, factor = rep("INJURIES",12)))

ggplot(health_dmg, aes(x = EVTYPE, y = TOTHEALTHDMG, fill = factor)) +
    geom_bar(stat = "identity") +
    labs(title = "Event Types that have Most Health Impact",
         x = "Event Type", y = "Total Health Damage") +
    theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



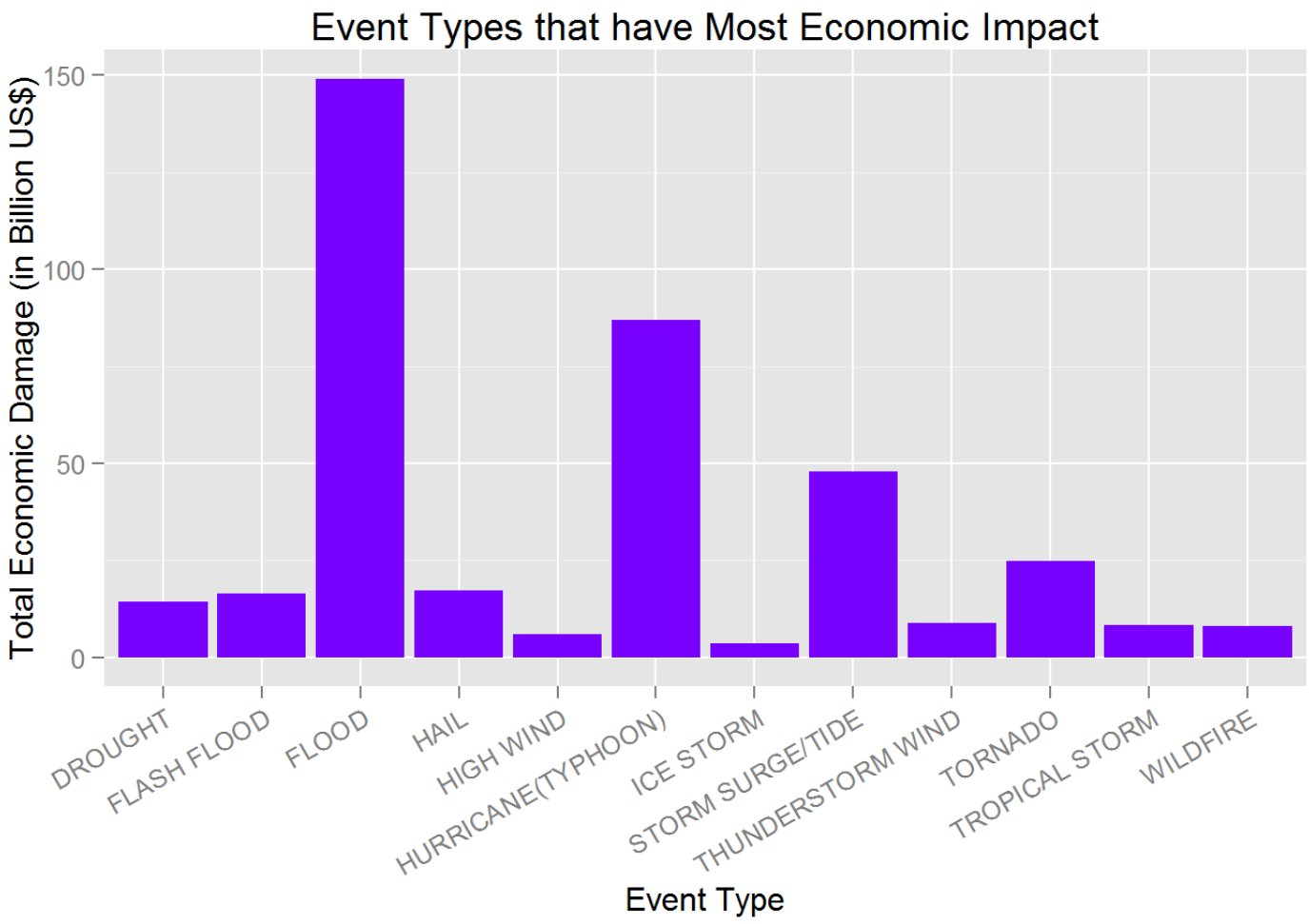
Types of events that have the greatest economic consequences:

```
top_economic_dmg <- stormdf %>%
    select(-c(FATALITIES, INJURIES, TOTHEALTHDMG)) %>%
    mutate(TOTECONOMICDMG = TOTECONOMICDMG / 1e9) %>% # amount to billion
    filter(TOTECONOMICDMG >= 2) %>%
    arrange(desc(TOTECONOMICDMG))

top_economic_dmg
```

```
## Source: local data frame [12 x 4]
##
##                EVTYPE   PROPDMG   CROPDMG TOTECONOMICDMG
## 1               FLOOD 1.441e+11 5.013e+09        149.143
## 2   HURRICANE(TYPHOON) 8.172e+10 5.350e+09         87.069
## 3     STORM SURGE/TIDE 4.784e+10 3.955e+06         47.845
## 4             TORNADO 2.462e+10 2.834e+08         24.900
## 5                HAIL 1.464e+10 2.562e+09         17.201
## 6         FLASH FLOOD 1.522e+10 1.335e+09         16.558
## 7             DROUGHT 1.046e+09 1.337e+10         14.414
## 8    THUNDERSTORM WIND 7.871e+09 9.523e+08          8.823
## 9      TROPICAL STORM 7.642e+09 6.777e+08          8.320
## 10           WILDFIRE 7.760e+09 4.023e+08          8.163
## 11          HIGH WIND 5.252e+09 6.339e+08          5.886
## 12          ICE STORM 3.642e+09 1.566e+07          3.658
```

```
ggplot(top_economic_dmg, aes(x = EVTYPE, y = TOTECONOMICDMG)) +
    geom_bar(stat = "identity", fill = "#7700FF") +
    labs(title = "Event Types that have Most Economic Impact",
         x = "Event Type", y = "Total Economic Damage (in Billion US$)") +
    theme(axis.text.x = element_text(angle = 30, hjust = 1))
```



Note

This analysis was done under Peer Assesment 2 of Coursera course *Reproducible Research*.