```
#Load requirement packages
library(R.utils)
library(data.table)
library(dplyr)
library(ggplot2)
library(tidyr)
install.packages("knitr")
library(knitr)

#Data Processing
data <- read.csv("/Users/asudeberber/Downloads/repdata_data_StormData.csv", header =
TRUE, sep = ",")
head(data)
dim(data)

#Based on the information above, the data table now has 902,297 rows and 7 columns.
Below is a brief description of each variable.

#1. evtype    : storm event type
#2. fatalities: amount of fatalities per event
#3. injuries  : amount of injuries per event
#4. propdmg   : property damage amount
#5. propdmgexp: property damage in exponents
#6. cropdmg   : crop damage amount
#7. cropdmgexp: crop damage in exponents

View(data)

#There are a lot of columns and I don't need all so I subset them
data2 <- data %>%

select(c("EVTYPE","FATALITIES","INJURIES","PROPDMG","PROPDMGEXP","CROPDMG","CROPDMGEXP"))
%>%
  rename_all(tolower)
str(data2)

length(unique(data2$EVTYPE))

#I select columns I need for the bar plot, group it by event type and calculate sum of
both fatalities and injuries. Then, arrange it in descending order and slice the first
10 rows, then gather it and turning it into categorical variables for creating a grouped
bar plot.
pop_health <-
  data2 %>% 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)

#the variable PROPDMGEXP is regarding property damage expenses, use for economic
consequensequences

unique(data2$propdmgexp)
unique(data2$cropdmgexp)

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
```

```
}

economic <-
  data2 %>% select("evtype", "propdmg", "propdmgexp", "cropdmg", "cropdmgexp") %>%
  mutate(prop_dmg = propdmg*sapply(propdmgexp, FUN = cost), crop_dmg =
cropdmg*sapply(cropdmgexp, FUN = cost), .keep="unused") %>%
  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
### 1. Across the United States, which types of events (as indicated in the EVTYPE
variable) are most harmful with respect to population health?

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("red", "pink"))

#Based on the bar plot, it's evident that tornadoes have the highest impact on the
popoulation health, since it causes the most fatalities and injuries.

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

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

# Conclusion

#On the basis of this analysis, funds should be allocated to strengthening
infrastructure or early warning systems so as to deal with tornadoes in order to protect
people's safety and health. In order to safeguard these assets and crops as much as
possible, there should be more funding for innovation in developing better protection
systems and infrastructure with a view to preventing damage from hurricanes and
typhoons.
```