

Reproducible Research: Course Project 2

Samvedna

19/06/2021

```
library(ggplot2)
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
setwd("C:/Users/peign/Desktop/Last project")
```

```
data <- read.table("repdata_data_StormData.csv.bz2",
```

```
header=TRUE,
```

```
sep=";",
```

```
stringsAsFactors = FALSE)
```

```
mini_data <- data[,c("EVTYPE", "FATALITIES", "INJURIES", "PROPDGMG",  
"PROPDGMGEXP", "CROPDGMG", "CROPDGMGEXP")]
```

```
mini_data$PROPDGMGEXP <- as.character(mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP = gsub("\\-|\\+|\\?|0", "0", mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP = gsub("h|H", "100", mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP = gsub("k|K", "1000", mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP = gsub("m|M", "1000000", mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP = gsub("b|B", "1000000000", mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP <- as.numeric(mini_data$PROPDGMGEXP)
```

```
mini_data$PROPDGMGEXP[is.na(mini_data$PROPDGMGEXP)] = 0
```

```
mini_data$CROPDGMGEXP <- as.character(mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP = gsub("\\-|\\+|\\?|0", "0", mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP = gsub("h|H", "100", mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP = gsub("k|K", "1000", mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP = gsub("m|M", "1000000", mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP = gsub("b|B", "1000000000", mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP <- as.numeric(mini_data$CROPDGMGEXP)
```

```
mini_data$CROPDGMGEXP[is.na(mini_data$CROPDGMGEXP)] = 0
```

```
1
```

First question : Across the United States, which types of events are most harmful with respect to population health?

```
library(grid)
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 4.0.5
```

```
## -----
```

```
## You have loaded plyr after dplyr - this is likely to cause problems.
```

```
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
```

```
## library(plyr); library(dplyr)
```

```
## -----
```

```
##
## Attaching package: 'plyr'
## The following objects are masked from 'package:dplyr':
##
## arrange, count, desc, failwith, id, mutate, rename, summarise,
## summarize
mini_data$HEALTH_TOTAL <- mini_data$FATALITIES + mini_data$INJURIES
mini <- mini_data
graph1 <- aggregate(x=list(Total=mini$HEALTH_TOTAL), by=list(Event=mini$EVTYPE),
FUN=sum)
graph1_bis <- order(graph1$Total, decreasing = TRUE)
graph1 <- graph1[graph1_bis,]
plot <- ggplot(graph1[1:10,], aes(x = reorder(Event, -Total), y = Total)) +
geom_bar(color="BLUE", stat = "identity") +
theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position = "none") +
labs(x = "Event Type", y = "Total Casualties", title = "Casualties")
```

```
plot
2
0
25000
50000
75000
100000
TORNADO
EXCESSIVE HEAT
TSTM WIND
FLOOD
LIGHTNING
HEAT
FLASH FLOOD
ICE STORM
THUNDERSTORM WIND
WINTER STORM
```

```
Event Type
Total Casualties
Casualties
```

Thus we understand that the most dangerous natural phenomenon is Tornado.

Second Question: Across the United States, which types of events have the greatest economic consequences?

```
mini$ECONOMIC_TOTAL <- (mini$PROPDMG * mini$PROPDMGEXP) + (mini$CROPDMG
* mini$CROPDMGEXP)
graph2 <- aggregate(x=list(Total=mini$ECONOMIC_TOTAL), by=list(Event=mini$EVTYPE),
FUN=sum)
graph2_bis <- order(graph2$Total, decreasing = TRUE)
graph2 <- graph2[graph2_bis,]
plot_bis <- ggplot(graph2[1:10,], aes(x = reorder(Event, -Total), y = Total / 1000000000)) +
geom_bar(color="BLUE",stat = "identity") +
theme(axis.text.x = element_text(angle = 90, hjust = 1), legend.position = "none") +
```

```
labs(x = "Event Type", y = "Economic loses, $", title = "Economic loses")
```

```
plot_bis
```

```
3
```

```
0
```

```
50
```

```
100
```

```
150
```

```
FLOOD
```

```
HURRICANE/TYPHOON
```

```
TORNADO
```

```
STORM SURGE
```

```
HAIL
```

```
FLASH FLOOD
```

```
DROUGHT
```

```
HURRICANE
```

```
RIVER FLOOD
```

```
ICE STORM
```

```
Event Type
```

```
Economic loses, $
```

```
Economic loses
```

```
### Thus, flood causes the most economics loses
```

```
4
```