# Reproducible Research

Saurabh Ghadge

14/12/2021

# Introduction

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

The data for this assignment come in the form of a comma-separated-value file compressed via the bzip2 algorithm to reduce its size. You can download the file from the course web site:

Storm Data (https://d396qusza40orc.cloudfront.net/repdata%2Fdata%2FStormData.csv.bz2) [47Mb]

There is also some documentation of the database available. Here you will find how some of the variables are constructed/defined.

National Weather Service Storm Data Documentation (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2Fpd01016005curr.pdf)

National Climatic Data Center Storm Events FAQ (https://d396qusza40orc.cloudfront.net/repdata%2Fpeer2_doc%2FNCDC%20Storm%20Events-FAQ%20Page.pdf)

The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete. #### About Data Transformation
Data is first downloaded in working directory and then it is read in R studio.
Data is in CSV form so we read it using *read.csv()* command,and then it is transform in tibble table using *tibble()* command. All Data transformation is carried out in dplyr of tidyverse.

## Reading data

first few rows and column of this storm data is as follows.

```
data <- read.csv("repdata_data_stormdata.csv")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data <- tibble(data)
head(data,5)
```

```
## # A tibble: 5 x 37
##   STATE__ BGN_DATE   BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE EVTYPE BGN_RANGE
##     <dbl> <chr>      <chr>    <chr>      <dbl> <chr>      <chr> <chr>      <dbl>
## 1       1 4/18/1950~ 0130     CST           97 MOBILE     AL    TORNA~         0
## 2       1 4/18/1950~ 0145     CST            3 BALDWIN    AL    TORNA~         0
## 3       1 2/20/1951~ 1600     CST           57 FAYETTE    AL    TORNA~         0
## 4       1 6/8/1951 ~ 0900     CST           89 MADISON    AL    TORNA~         0
## 5       1 11/15/195~ 1500     CST           43 CULLMAN    AL    TORNA~         0
## # ... with 28 more variables: BGN_AZI <chr>, BGN_LOCATI <chr>, END_DATE <chr>,
## #   END_TIME <chr>, COUNTY_END <dbl>, COUNTYENDN <lgl>, END_RANGE <dbl>,
## #   END_AZI <chr>, END_LOCATI <chr>, LENGTH <dbl>, WIDTH <dbl>, F <int>,
## #   MAG <dbl>, FATALITIES <dbl>, INJURIES <dbl>, PROPDMG <dbl>,
## #   PROPDMGEXP <chr>, CROPDMG <dbl>, CROPDMGEXP <chr>, WFO <chr>,
## #   STATEOFFIC <chr>, ZONENAMES <chr>, LATITUDE <dbl>, LONGITUDE <dbl>,
## #   LATITUDE_E <dbl>, LONGITUDE_ <dbl>, REMARKS <chr>, REFNUM <dbl>
```

## Dive in

To find Across the United States, which types of events are most harmful with respect to population health I just group the data as events type and then calculate the deaths and injuries because of that events.
Variables FATALATIES and INJURIES represents damage to population.

```
harm_event <- data %>% group_by(EVTYPE) %>%
        summarise(death = sum(FATALITIES),Injury = sum(INJURIES)) %>%
        arrange(desc(death),desc(Injury))
harm_event
```

```
## # A tibble: 985 x 3
##    EVTYPE         death Injury
##    <chr>          <dbl>  <dbl>
##  1 TORNADO         5633  91346
##  2 EXCESSIVE HEAT  1903   6525
##  3 FLASH FLOOD      978   1777
##  4 HEAT             937   2100
##  5 LIGHTNING        816   5230
##  6 TSTM WIND        504   6957
##  7 FLOOD            470   6789
##  8 RIP CURRENT      368    232
##  9 HIGH WIND        248   1137
## 10 AVALANCHE        224    170
## # ... with 975 more rows
```

Here we see that death and injuries are double type clearly It should be Integer type and for some of the events death and Injuries are zero so we neglect such events.

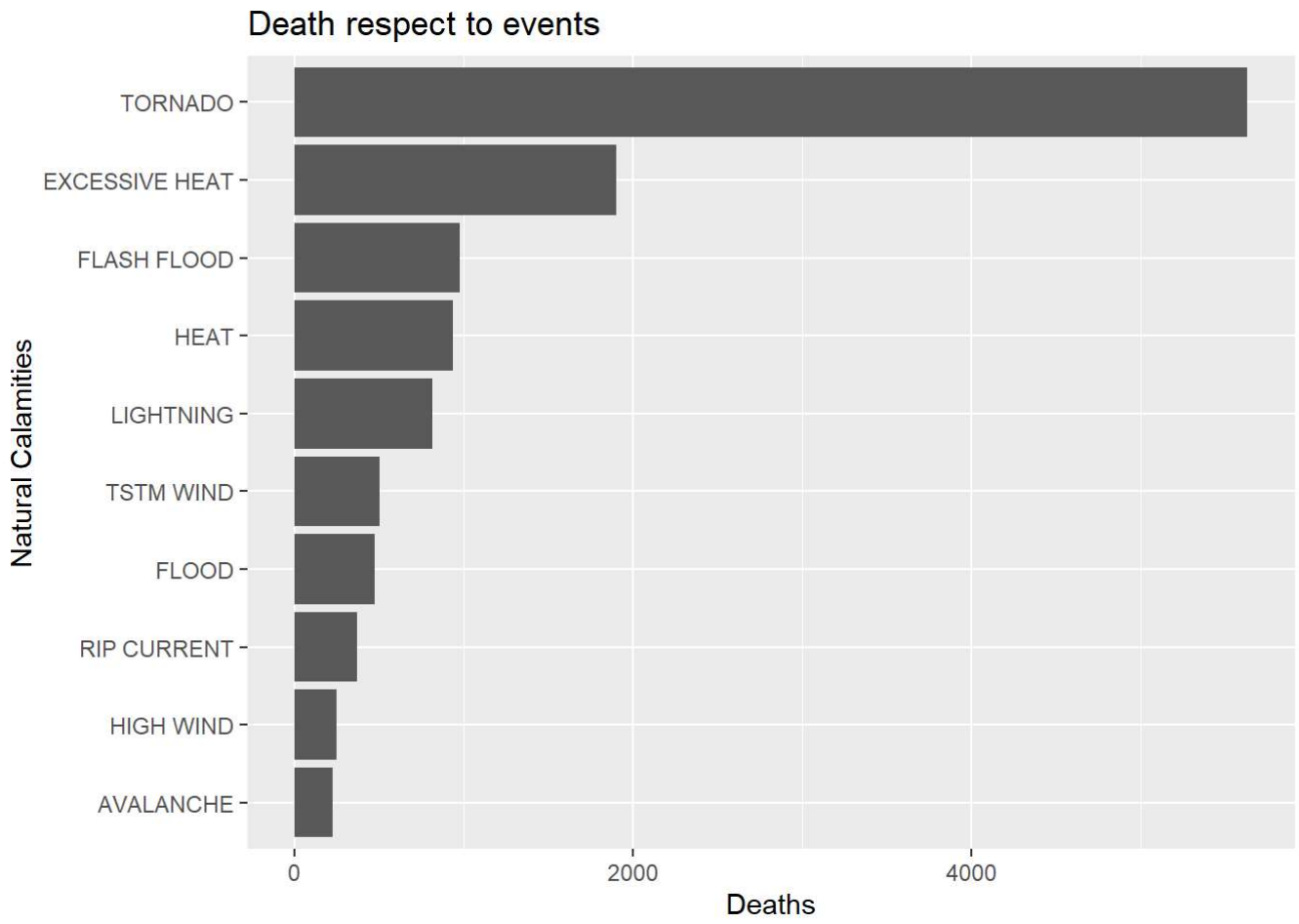
```
harm_event$death <- as.integer(harm_event$death)
harm_event$Injury <- as.integer(harm_event$Injury)
```

We have to find most destructive events with respective population health,So simply Here we are going to select top 10 events that cause more casualty for population.

```
most_harm <- harm_event[1:10,]
```

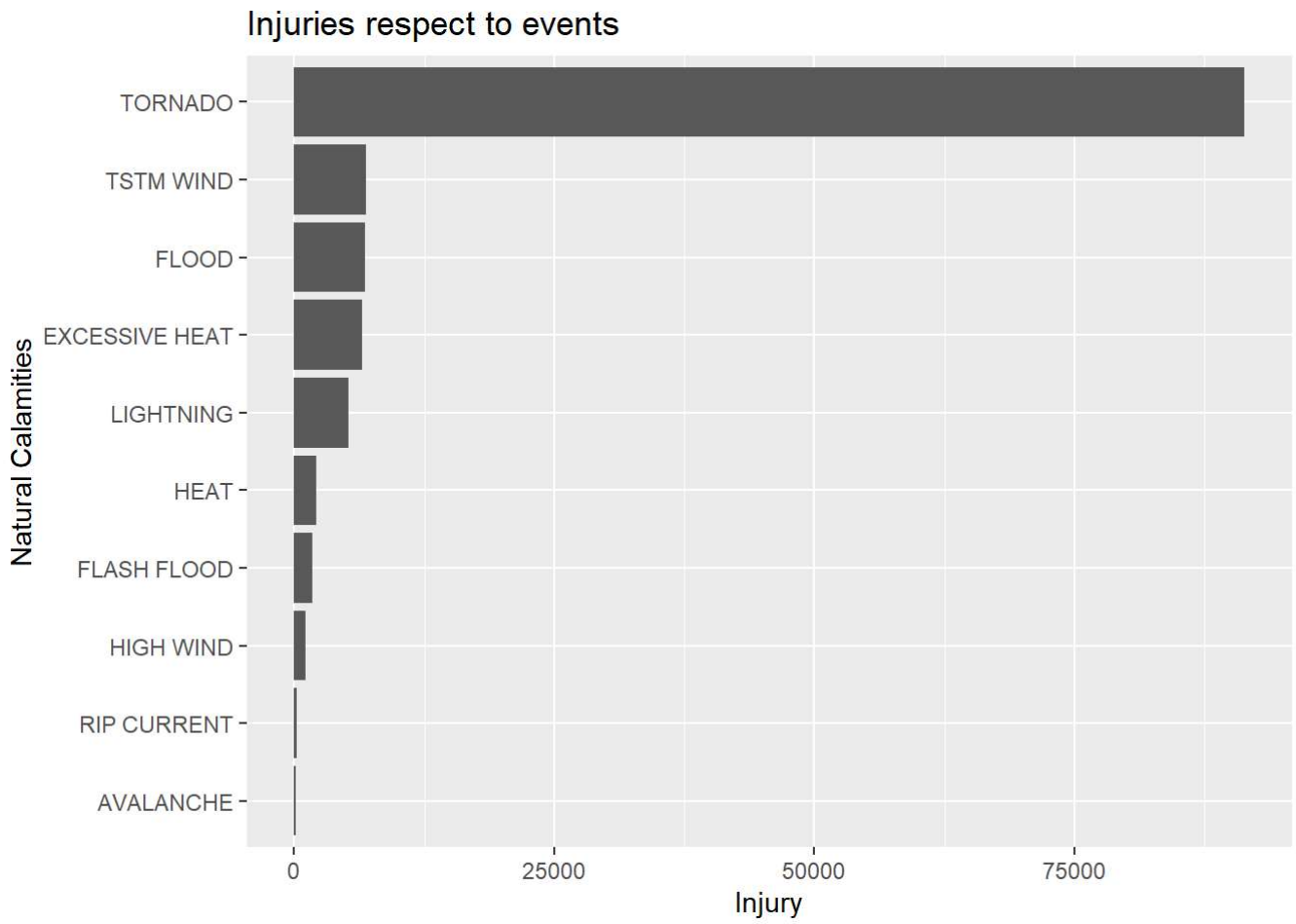
## Bar plot showing deaths Due Natural calamities(most harmuful/Which causes more deaths)

```
ggplot(data = most_harm) + geom_bar(mapping = aes( x = reorder(EVTYPE,death), y = death),
                                    stat = "identity") + coord_flip() +
        ggtitle("Death respect to events") + labs(x = "Natural Calamities",y="Deaths")
```




## Bar plot showing Injuries Due to Natural calamities

```
ggplot(data = most_harm) + geom_bar(mapping = aes( x = reorder(EVTYPE,Injury),
                                                   y = Injury),stat = "identity") +
        coord_flip() + ggtitle("Injuries respect to events") + labs(x = "Natural Calamities")
```




## To find which types of events have the greatest economic consequences

Variable PROPDMG, PROPDMGEXP, CROPDMG AND CROPDMGEXP REPRESENTS Economic consequences on property and crop.

```
unique(data$PROPDMGEXP)
```

```
##  [1] "K" "M" ""  "B" "m" "+" "0" "5" "6" "?" "4" "2" "3" "h" "7" "H" "-" "1" "8"
```

```
unique(data$CROPDMGEXP)
```

```
## [1] ""  "M" "K" "m" "B" "?" "0" "k" "2"
```

These are damage exponents(millions,billions…)
we will convert them into a same unit for analysing which events takes more money or which event has greatest economic consequences.

```
data$PROPDMGEXP <- toupper(data$PROPDMGEXP)
old <- c("B","M","K","H")
new <- c(9,6,3,2)
data$PROPDMGEXP[data$PROPDMGEXP %in% old] <- new[match(data$PROPDMGEXP,old,nomatch = 0)]
data$PROPDMGEXP[data$PROPDMGEXP %in% c("","+","-","?")] <- "0"
data$CROPDMGEXP <- toupper(data$CROPDMGEXP)
data$CROPDMGEXP[data$CROPDMGEXP %in% old] <- new[match(data$CROPDMGEXP,old,nomatch = 0)]
data$CROPDMGEXP[data$CROPDMGEXP %in% c("","+","-","?")] <- "0"
```

We will now convert property damage and crop damage In same meassurable unit.

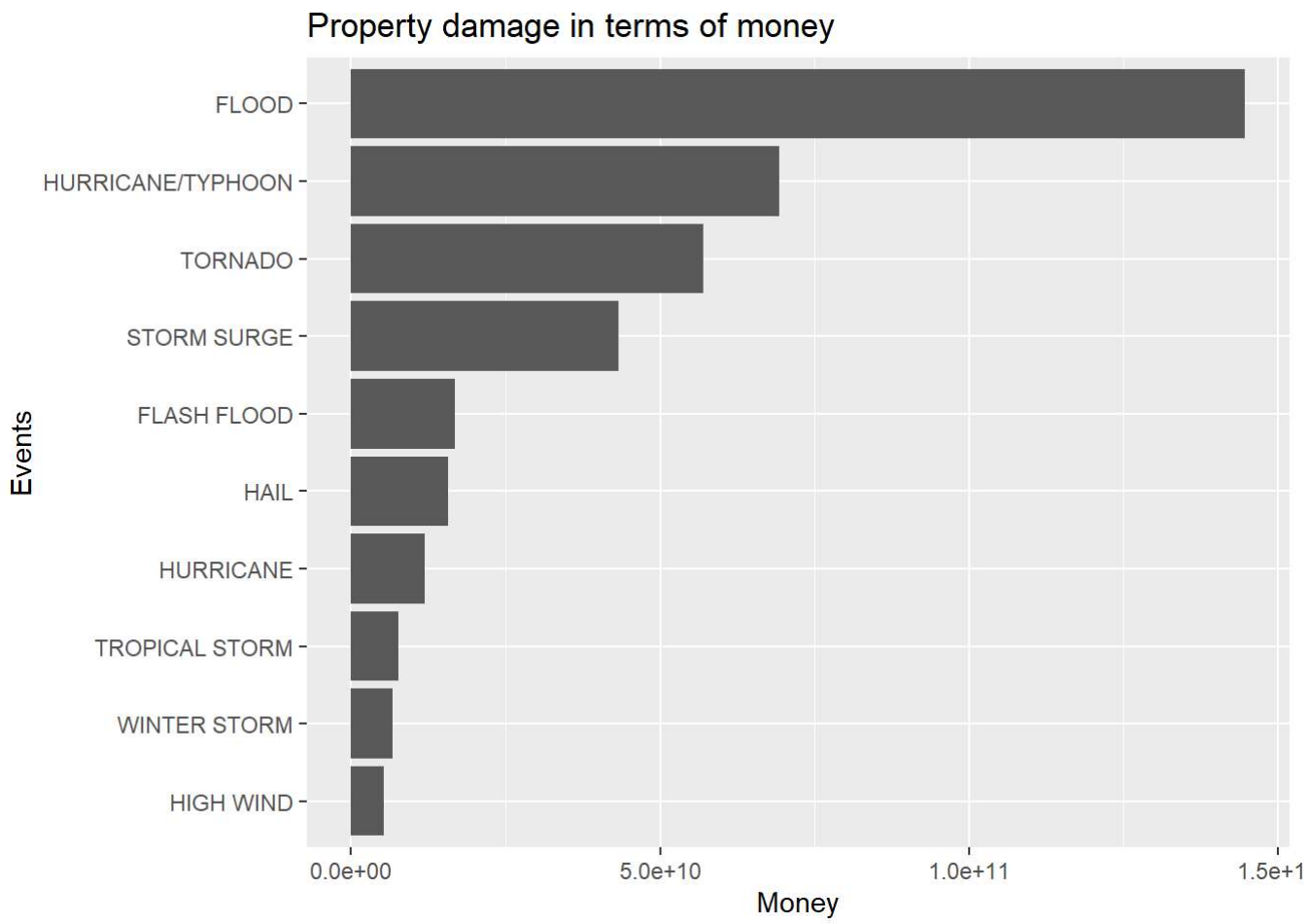
```
data$crop_dmg <- data$CROPDMG * (10^as.numeric(data$CROPDMGEXP))
data$prop_dmg <-  data$PROPDMG * (10^as.numeric(data$PROPDMGEXP))
```

now we will group the events in order to get which event produces more economic consequences.
we only select first 10 events which produces most economic consequences comare to other.

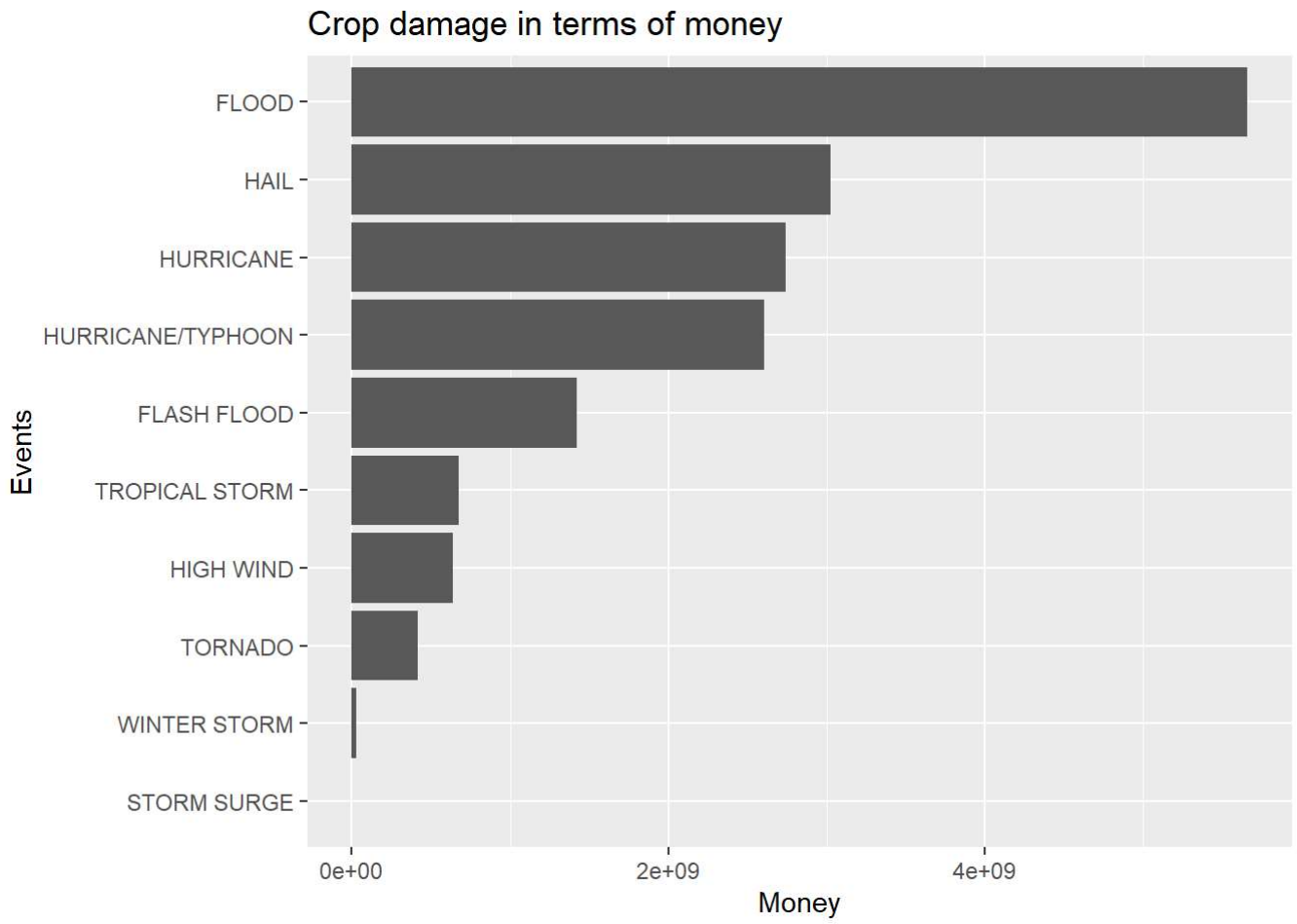
```
eco_even <- data %>% group_by(EVTYPE) %>% summarise( prop_dmg_money = sum(prop_dmg),crop_dmg_mon
ey = sum(crop_dmg)) %>% arrange(desc(prop_dmg_money),desc(crop_dmg_money))
most_eco_even <- eco_even[1:10,]
```

following bar plot shows how much economic consequences produces by events for property damage and crop damage.

```
ggplot(data = most_eco_even,mapping = aes(x = reorder(EVTYPE,prop_dmg_money),y = prop_dmg_mone
y)) + geom_bar(stat = "identity") + coord_flip() + ggtitle("Property damage in terms of money")
 + labs(x = "Events",y = "Money")
```




```
ggplot(data = most_eco_even,mapping = aes(x = reorder(EVTYPE,crop_dmg_money),y = crop_dmg_mone
y)) + geom_bar(stat = "identity") + coord_flip() + ggtitle("Crop damage in terms of money") + la
bs(x = "Events",y = "Money")
```




# Results

1. Tornado and Excessive heat events causes more deaths, and Tornado and TSTM Wind are responsible for most of injuries in US.
2. Flood and Hurricane/typhoon are responsible for most of property damage while flood and hail are responsible most of crop damage.