IST 687
Homework 5
Due Date: 11/9

Code requires the following packages to run

```
library(RCurl) #install.packages('RCurl')
library(RJSONIO) #install.packages('RJSONIO')
library(jsonlite) #install.packages('jsonlite')
library(DescTools) #install.packages('DescTools')
library(na.tools) #install.packages('na.tools')
library(RSQLite) #install.packages('RSQLite')
library(sqldf) #install.packages('sqldf')
```

Step 1: Load the data

```
#Read in the following JSON dataset
maryland_URL <- "http://opendata.maryland.gov/api/views/pdvh-tf2u/rows.json?,%E2%86%92accessType=DOWNLOAD"
maryland_retrieve_URL <- getURL(maryland_URL)
maryland_results <- fromJSON(maryland_retrieve_URL)
```

Step 2: Clean the data

```
#remove the first eight columns
maryland_cleaned <- maryland_results[["data"]][,-1:-8]

#define the column names
namesOfColumns <- c("CASE_NUMBER", "BARRACK", "ACC_DATE", "ACC_TIME", "ACC_TIME_CODE", "DAY_OF_WEEK", "ROAD", "INTERSECT_ROAD", "DIST_FROM_INTERSECT", "DIST_DIRECTION", "CITY_NAME", "COUNTY_CODE", "COUNTY_NAME", "VEHICLE_COUNT", "PROP_DEST", "INJURY", "COLLISION_WITH_1", "COLLISION_WITH_2")

#add the column names to the dataset
colnames(maryland_cleaned) <- namesOfColumns
```

Step 3: Understand the data using SQL (via SQLDF)

```
#Turn into data frame to enable SQL query use
maryland_cleaned_SQL <- as.data.frame(maryland_cleaned)

#How many accidents happen on Sunday?
sqldf('select count(maryland_cleaned_SQL.CASE_NUMBER) from maryland_cleaned_SQL where trim(DAY
_OF_WEEK) = "SUNDAY"')
##   count(maryland_cleaned_SQL.CASE_NUMBER)
## 1                                    2373

#How many accidents had injuries?
sqldf('select count(maryland_cleaned_SQL.CASE_NUMBER) from maryland_cleaned_SQL where INJURY =
"YES"')
##   count(maryland_cleaned_SQL.CASE_NUMBER)
## 1                                    6433

#List the injuries by day
sqldf('select maryland_cleaned_SQL.DAY_OF_WEEK, count(*) from maryland_cleaned_SQL where INJURY
= "YES" group by maryland_cleaned_SQL.DAY_OF_WEEK')
##   DAY_OF_WEEK count(*)
## 1   FRIDAY        1043
## 2   MONDAY         915
## 3   SATURDAY       950
## 4   SUNDAY         818
## 5   THURSDAY       968
## 6   TUESDAY        843
## 7   WEDNESDAY      896
```

Step 4: Understand the data using tapply

```
#Clean out spaces from DAY_OF_WEEK
maryland_cleaned[,6] <- gsub(" ", "", maryland_cleaned[,6])

#How many accidents happen on Sunday?
sum(as.numeric(maryland_cleaned[,6]=="SUNDAY"))
## [1] 2373

#How many accidents had injuries?
sum(as.numeric(maryland_cleaned[,16]=="YES"),na.rm = TRUE)
## [1] 6433

#List the injuries by day
tapply(as.numeric(na.replace(maryland_cleaned[,16], "NO")=="YES"),maryland_cleaned[,6],sum)
##    FRIDAY    MONDAY  SATURDAY    SUNDAY  THURSDAY   TUESDAY WEDNESDAY
##      1043       915       950       818       968       843       896
```