ThulasiRam_RuppaKrishnan_HW5

JSON & tapply Homework: Accident Analysis

```
#Step   1:  Load    the data

#Read   in  the following   JSON    dataset
#http://data.maryland.gov/api/views/pdvh-tf2u/rows.json?accessType=DOWNLOAD

# load required libraries
library(bitops)
library(RCurl)
library(jsonlite)
library(RJSONIO)
```

```
##
## Attaching package: 'RJSONIO'
```

```
## The following objects are masked from 'package:jsonlite':
##
##     fromJSON, toJSON
```

```
library(proto)
library(gsubfn)
library(RSQLite)
library(sqldf)

# load data
mv_URL <- "http://data.maryland.gov/api/views/pdvh-tf2u/rows.json?accessType=DOWNLOAD"
mvApiResult <- getURL(mv_URL)
mvResults <- RJSONIO::fromJSON(mvApiResult)
summary(mvResults)
```

```
##      Length Class  Mode
## meta     1  -none- list
## data 18638  -none- list
```

```
#summary(mvResults$data)
mvData <-mvResults$data
```

```
nullToNA <- function(x) {
  x[sapply(x, is.null)] <- NA
  return(x)
}

namesOfColumns <-
  c("CASE_NUMBER","BARRACK","ACC_DATE","ACC_TIME","ACC_TIME_CODE","DAY_OF_WEEK","ROAD","INTERSEC
T_ROAD","DIST_FROM_INTERSECT","DIST_DIRECTION","CITY_NAME","COUNTY_CODE","COUNTY_NAME","VEHICLE_
COUNT","PROP_DET","INJURY","COLLISION_WITH_1","COLLISION_WITH_2")

mv_df <- data.frame(matrix(unlist(lapply(mvData,nullToNA)),nrow=length(mvResults$data),ncol = le
ngth(mvResults$data[[1]]),byrow = T), stringsAsFactors = FALSE)
mv_df <- mv_df[,-c(1:8)]
colnames(mv_df) <- namesOfColumns
mv_df$DAY_OF_WEEK <-sapply(mv_df$DAY_OF_WEEK,trimws,which='right')
View(mv_df)
```

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

# How   many    accidents   happen  on  SUNDAY
sqldf('select count(case_number) accidents_cnt from mv_df where (day_of_week) ="SUNDAY"')
```

```
##   accidents_cnt
## 1          2373
```

```
# How   many    accidents   had injuries
sqldf('select count(1) accidents_with_injury from mv_df where injury="YES" ')
```

```
##   accidents_with_injury
## 1                  6433
```

```
# List  the injuries    by  day
sqldf('select (day_of_week) day_of_week,count(1) injuries_cnt from mv_df where injury="YES" grou
p by (day_of_week) order by case (day_of_week) when "SUNDAY" then 1 when "MONDAY" then 2 when "T
UESDAY" then 3 when "WEDNESDAY" then 4 when "THURSDAY" then 5 when "FRIDAY" then 6 when "SATURDA
Y" then 7 end')
```

```
##   day_of_week injuries_cnt
## 1      SUNDAY          818
## 2      MONDAY          915
## 3     TUESDAY          843
## 4   WEDNESDAY          896
## 5    THURSDAY          968
## 6      FRIDAY         1043
## 7    SATURDAY          950
```

```
#Step   4:  Understand  the data    using   tapply

# How   many    accidents   happen  on  SUNDAY
data.frame(`colnames<-`(matrix(tapply(mv_df$DAY_OF_WEEK, mv_df$DAY_OF_WEEK=='SUNDAY', length)[2
]),"accidents_cnt"))
```

```
##   accidents_cnt
## 1          2373
```

```
# How   many    accidents   had injuries
data.frame(`colnames<-`(matrix(tapply(mv_df$CASE_NUMBER, mv_df$INJURY=='YES', length)[2]),"accid
ents_with_injury"))
```

```
##   accidents_with_injury
## 1                  6433
```

```
# List  the injuries    by  day
`colnames<-`(data.frame(tapply(mv_df[which(mv_df$INJURY=='YES'),][,1], mv_df[which(mv_df$INJURY=
='YES'),][,which(colnames(mv_df)=="DAY_OF_WEEK")]  , length)),"injuries_cnt")
```

```
##           injuries_cnt
## FRIDAY            1043
## MONDAY             915
## SATURDAY           950
## SUNDAY             818
## THURSDAY           968
## TUESDAY            843
## WEDNESDAY          896
```