Accident Analysis using the apply family and sqldf

Surabhi Chouhan

December 20, 2016

```
library("jsonlite")

library("curl")

library("sqldf")

## Loading required package: gsubfn

## Loading required package: proto

## Loading required package: RSQLite

library(plyr)

acc_data <- fromJSON("https://data.maryland.gov/api/views/pdvh-
tf2u/rows.json?accessType=DOWNLOAD") #this command gets the json from given
url



accidents <- data.frame(acc_data[['data']]) #as the json contains metadata as
well as data, so this command extracts the data from json and stores it as a
dataframe
accident_analysis <- accidents[,-(1:8)] #removing the first 8 columns and
storng it into another variable

#this command is creating a vector for the column names of the data frame
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_WI
TH_1","COLLISION_WITH_2")
names(accident_analysis) <- namesOfColumns #naming the columns of dataframe



#removing extra space from the day of week using gsub function, and as it
removes the space, the factor gets converted to character, so next step makes
it a factor again
accident_analysis$DAY_OF_WEEK <-
gsub('\\s+','',accident_analysis$DAY_OF_WEEK)
accident_analysis$DAY_OF_WEEK<-as.factor(accident_analysis$DAY_OF_WEEK)
```

```
#this command counts the number of accidents that took place on Sunday
sqldf("select count(*) from accident_analysis where DAY_OF_WEEK == 'SUNDAY'")
```

```
##   count(*)
## 1     2373
```

```
#this command counts the number of accidents that had injuries
sqldf("select count(*) from accident_analysis where INJURY == 'YES' AND
INJURY IS NOT NULL")
```

```
##   count(*)
## 1     6433
```

```
#this command gives the count of injuries according to the day of week
sqldf("select DAY_OF_WEEK, count(*) from accident_analysis where INJURY ==
'YES'  group by 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
```

```
#this command counts the number of accidents that happened on Sunday. The
difference in output is that is shows the count for other days under FALSE
and for Sunday under TRUE as I have given the group by condition specifying
the value of DAY_OF_WEEK. Rest the count in both sql and tapply remains the
same.
acc_on_sunday <-
tapply(accident_analysis$DAY_OF_WEEK,accident_analysis$DAY_OF_WEEK
=="SUNDAY",count)
acc_on_sunday
```

```
## $`FALSE`
##           x freq
## 1    FRIDAY 3014
## 2    MONDAY 2554
## 3  SATURDAY 2732
## 4  THURSDAY 2671
## 5   TUESDAY 2676
## 6 WEDNESDAY 2618
##
## $`TRUE`
##        x freq
## 1 SUNDAY 2373
```

```
#this command gives the count of accidents which had injuries. The difference
in output is that it gives the count of INJURY=NO under FALSE and for
INJURY=YES under TRUE, as I have given the group by condition specifying the
```

```
value of INJURY. Rest the count in both sql and tapply remains the same.
acc_with_injury <- tapply(accident_analysis$INJURY,accident_analysis$INJURY
== "YES",count)
acc_with_injury
```

```
## $`FALSE`
##    x  freq
## 1 NO 12204
##
## $`TRUE`
##     x freq
## 1 YES 6433
```

```
#this command gives the count of injuries according to the days, with output
showing count of INJURY=NO under FALSE and count of INJURY=YES under TRUE as
my group by condition contains a specific value of INJURY. Rest the count in
both sql and tapply remains the same.
injury_list_day <-
tapply(accident_analysis$DAY_OF_WEEK,accident_analysis$INJURY=="YES",count)
injury_list_day
```

```
## $`FALSE`
##           x freq
## 1    FRIDAY 1971
## 2    MONDAY 1639
## 3  SATURDAY 1781
## 4    SUNDAY 1555
## 5  THURSDAY 1703
## 6   TUESDAY 1833
## 7 WEDNESDAY 1722
##
## $`TRUE`
##           x freq
## 1    FRIDAY 1043
## 2    MONDAY  915
## 3  SATURDAY  950
## 4    SUNDAY  818
## 5  THURSDAY  968
## 6   TUESDAY  843
## 7 WEDNESDAY  896
```