IST687 – 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

Step 2: Clean the data

After you load the data, remove the first 8 columns, and then, to make it easier to work with, name the rest of the columns as follows:

Note, not surprisingly, it is in JSON format. You should be able to see that the first result is the metadata (information about the data) and the second is the actual data.

```
namesOfColumns <-
c("CASE_NUMBER","BARRACK","ACC_DATE","ACC_TIME","ACC_TIME_CODE","DAY_OF_WE
EK","ROAD","INTERSECT_ROAD","DIST_FROM_INTERSECT","DIST_DIRECTION","CITY_NA
ME","COUNTY_CODE","COUNTY_NAME","VEHICLE_COUNT","PROP_DEST","INJURY","COLLI
SION_WITH_1","COLLISION_WITH_2")
```

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

Answer the following questions:

- How many accidents happen on SUNDAY
- How many accidents had injuries (might need to remove NAs from the data)
- List the injuries by day

Step 4: Understand the data using tapply

Answer the following questions (same as before) – compare results:

- How many accidents happen on Sunday
- How many accidents had injuries (might need to remove NAs from the data)
- List the injuries by day