library("jsonlite")

library("curl")

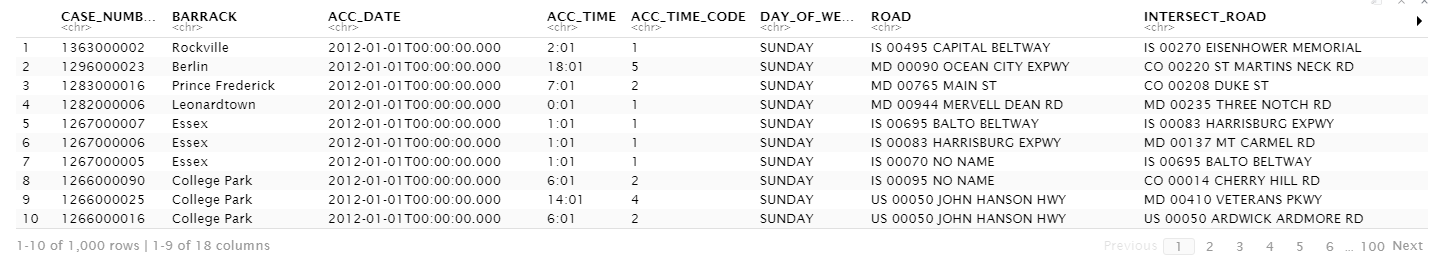
### step 1: load the data

```{r}

df <- fromJSON("https://opendata.maryland.gov/resource/pdvh-tf2u.json")

print(df)

```



### Step 2: Clean the data - these steps appear to already be compelted when printing the DF

#df <- df[,-(1:8)] didnt need to do this with new dataset as instructed

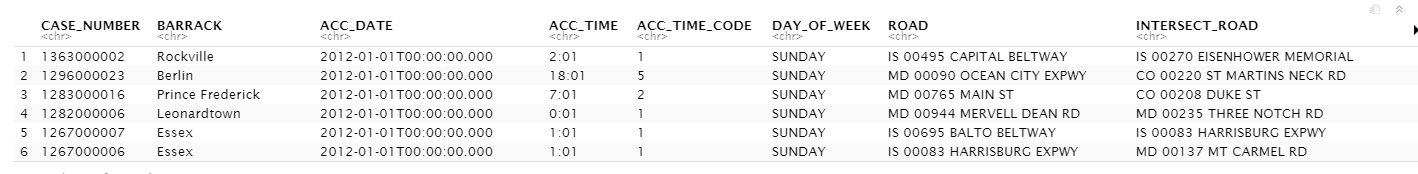
```{r}

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")

names(df) <- namesOfColumns

head(df)

```



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

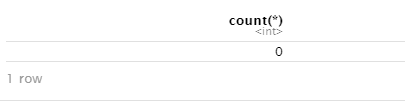
install.packages("sqldf")

library("sqldf")

```{r}

sqldf("select count(\*) from df where DAY\_OF\_WEEK == 'SUNDAY'")

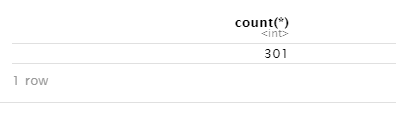
```



```{r}

sqldf("select count(\*) from df where INJURY == 'YES' AND INJURY IS NOT NULL")

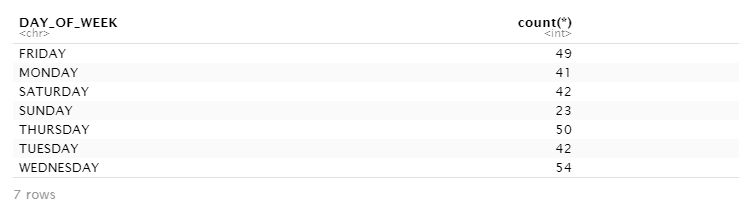
```



```{r}

sqldf("select DAY\_OF\_WEEK, count(\*) from df where INJURY == 'YES' group by DAY\_OF\_WEEK")

```



###step 4: Understand the data using tapply

install.packages("plyr")

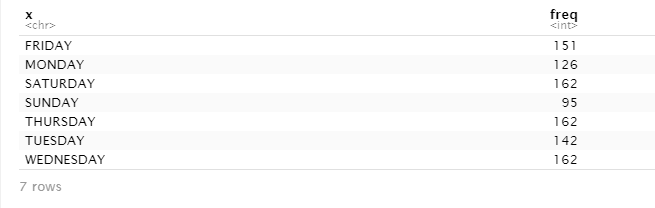
library(plyr)

```{r}

acc\_on\_sunday <- tapply(df$DAY\_OF\_WEEK,df$DAY\_OF\_WEEK =="SUNDAY",count)

acc\_on\_sunday

```

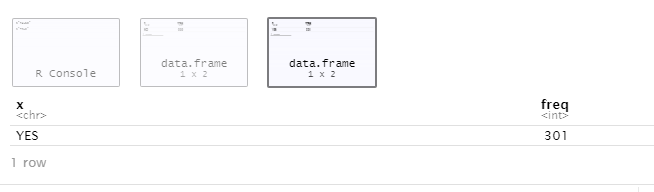


```{r}

acc\_with\_injury <- tapply(df$INJURY,df$INJURY == "YES",count)

acc\_with\_injury

```



```{r}

injury\_list\_day <- tapply(df$DAY\_OF\_WEEK,df$INJURY=="YES",count)

injury\_list\_day

```

