****

**CIS 5270: Business Intelligence**

**Spring 2017**

**Parking Citations in Los Angeles using RStudio**

Submitted to: Dr. Shilpa Balan

**Team Members: Bhagyashree Jadhav CIN:305849354**

**Narendra Mali CIN:305848717**

1. **Dataset URL**

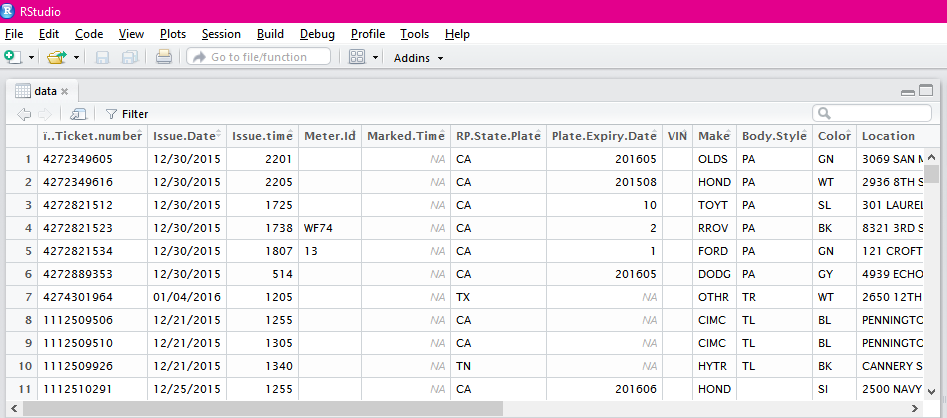
[**https://data.lacity.org/A-Well-Run-City/Parking-Citations/wjz9-h9np**](https://data.lacity.org/A-Well-Run-City/Parking-Citations/wjz9-h9np)

Above URL contains entire data regarding Parking Citation in City of Los Angeles from year 2010 onwards. Parking tickets or citations are issued by parking enforcement officials for violating state (California Vehicle Code) or city (Los Angeles Municipal Code) parking laws. (Angeles,2016) A parking violation is made when a motor vehicle is parked in a restricted place or for parking in an unauthorized manner. Analysis in Citation data every year gives us insights in approximating the revenue government will make in coming years and allow them to take measures necessary for improving parking scenario in the city. (Romero,2016)

Citations made in and around university campuses, commercial complexes and private properties are separately handled by parking enforcement of respective officials. LADOT operates 118 public parking facilities throughout the City of Los Angeles with more than 11,500 parking spaces.(LACity,2016) Traffic officers issue about 2.5 million citations each year. The dataset has parking citations listed from 2010 to 2016 and the size is 480 MB. There are more than four million entries and every citation has big amount of data such as person’s vehicle VIN number, Violation Description, date and time, RP state plate, location, latitude and longitude etc. which is recorded every day. This recorded data goes on increasing day by day, and hence analyzing this dataset with R will provide clear picture of parking citations in the City of Los Angeles. This analysis will help us to understand and recognize extra parking required in certain areas with less or reserved/permit parking.

1. **Data Cleaning:**
2. Data set contained more than 20 columns so, we removed unwanted columns such as “Meter ID”, “Marked Time” etc.

Before cleaning:

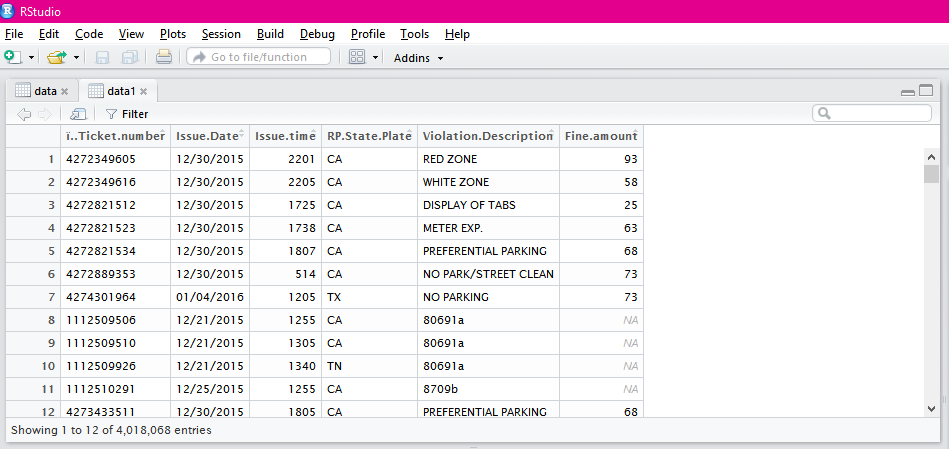


Code:

data1<- data[,c(1,2,3,6,16,17)]

View(data1)

After Cleaning:



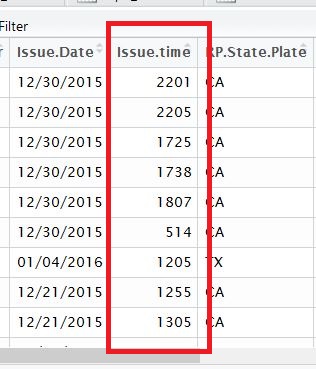
1. Issued time column has entry in numbered format instead of AM/PM format. We changed the Issue time column with the help of following code.

Code:

data1$Issue.time<- data1$Issue.time/100

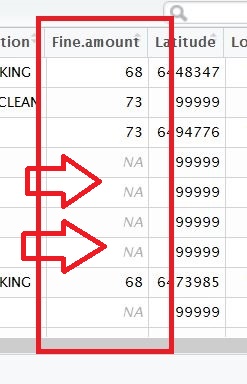
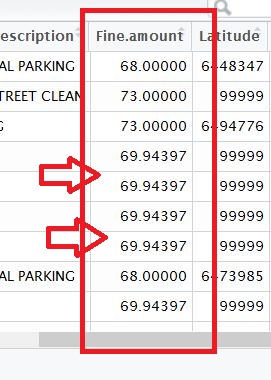
data1$Issue.time <- format(strptime(data1$Issue.time, '%H.%M'), '%I:%M %p')

Before cleaning After cleaning

1. Fine amount column had many NA values. We replaced NA values with mean of the fine amount column values.

Before Cleaning After Cleaning

Code:

fine\_amt<-data1$Fine.amount

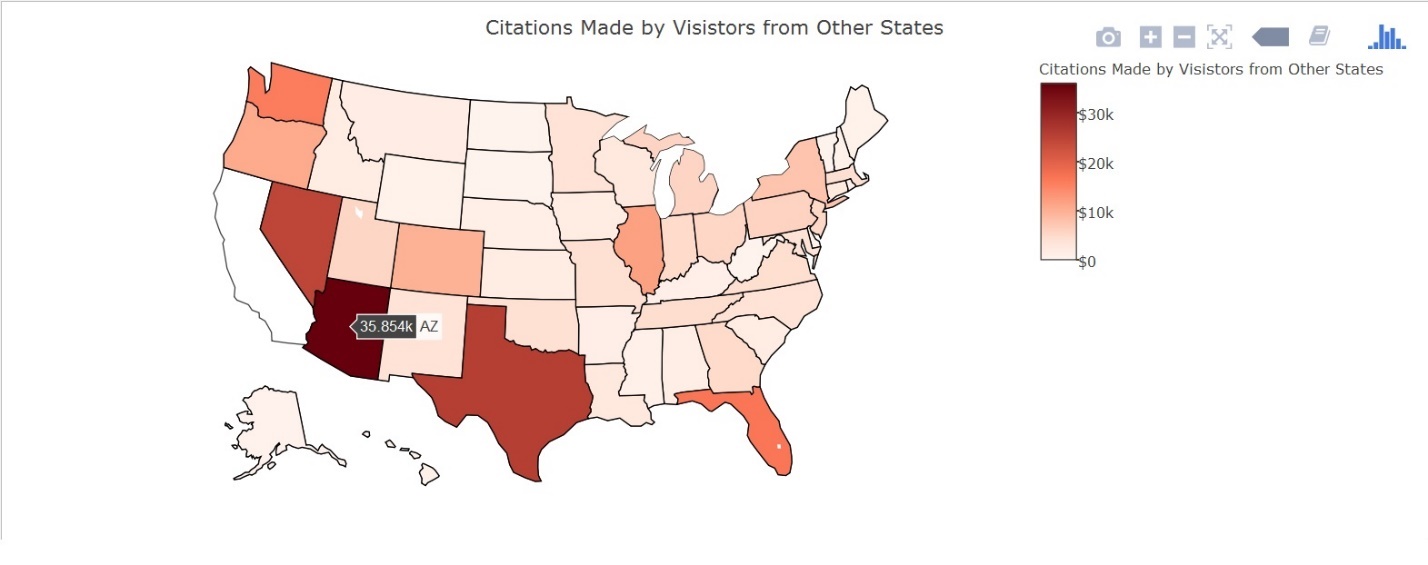
mean(fine\_amt,na.rm=TRUE)

fine\_amt[is.na(fine\_amt)]=mean(fine\_amt, na.rm = TRUE)

data1$Fine.amount<-fine\_amt

View(data1)

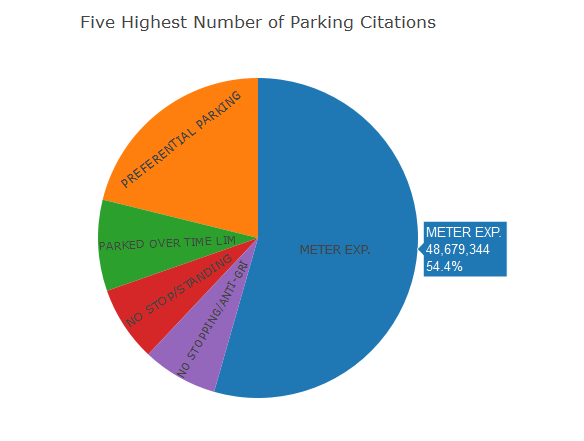
1. **Data Visualization:**
2. **What is the number of citations made by visitors from other state?**



(Highlights from R script – plotly, dplyr packages, filter, geographic map)

Tourists from all over the United States come to Los Angeles for vacation or for a quick trip. Parking citations for tourists are quite high since they don’t know where to park or safe areas for parking. Among all states, people whose cars are registered in the state of Arizona have most number of parking citations in Los Angeles, followed by Texas and Nevada. Total of 35,000+ citations record is found from the Arizona state vehicles alone. Geographic map has been used to display out of state citations. Hence, we suggest that tourist from Arizona, Texas and Nevada should take extra care before parking their vehicle in Los Angeles.

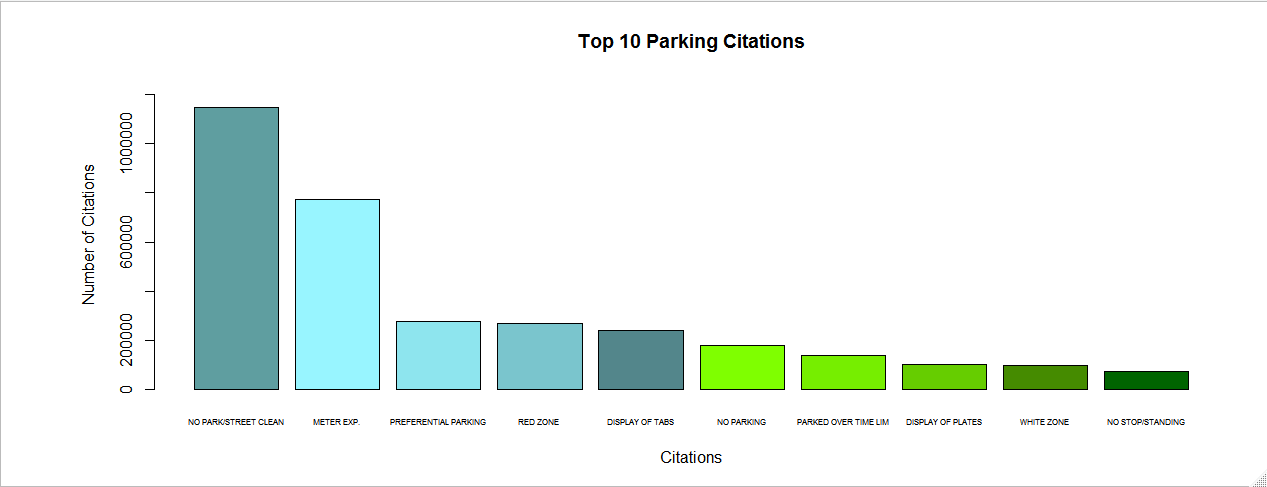
1. **What is the total fine amount for each violation?**



(Highlights from R script - dplyr, plotly package, sum ,mean, pie chart)

Each Violation has a particular fine assigned with it. It is important to find out that which type of citation gives large amount of fine to LA City. Above pie chart states the same thing, it provides top categories in which maximum fine is collected. It also shows percentage of each citation and different colors make it easier to read. For creating this insight mean of the fine amount has been replaced with N\A to avoid the incorrect output. Sum is also used to calculate total fine for each citation. Highest fine collected is for meter expiry citation. Preferential parking, parked over time limit, no stop\standing, no stopping\anti GRI are the subsequent violation description which has ranked amongst top 5 for parking citation with respect to amount of fine collected.

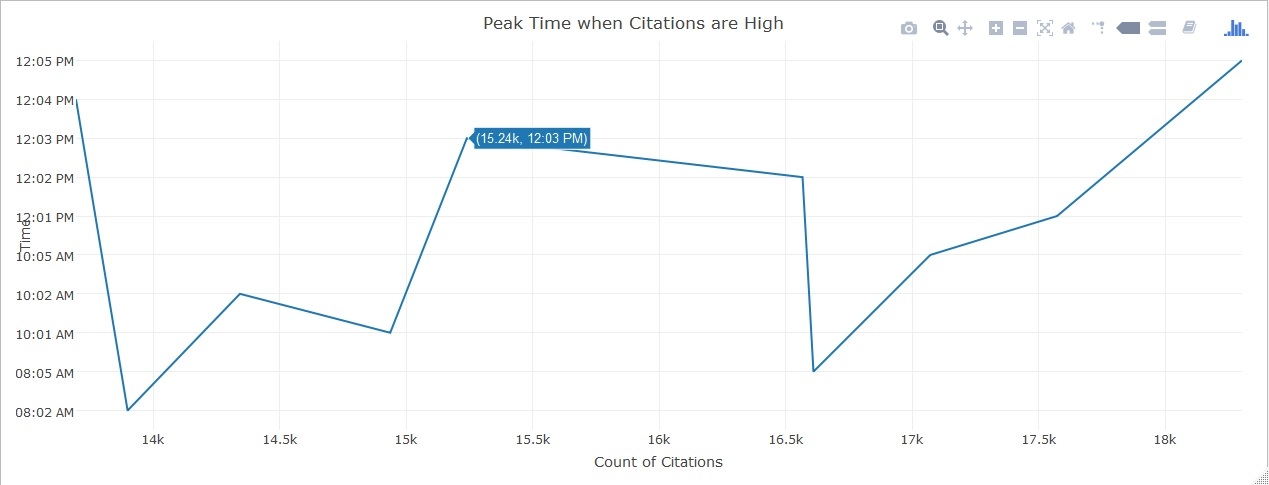
1. **How does count of different types of parking citations differ?**



(Highlights from R script – dplyr, barplot, bar chart, different colors)

From the above Bar chart, it can be seen that No Parking on Street Cleaning category leads citation category with total of 1,148,218 numbers of citations. Other categories are Preferential parking, Display of tabs, No Parking and No stopping. These categories have received high number of citations such as Preferential parking has received more than 700,000 citations and Display of tabs has received more than 275,000 citations. We recommend drivers to give special attentions to street cleaning boards because street cleaning information is always boarded on one of the street signs and with the help of this they can avoid parking citations.

1. What is the peak time when citations are high?



(Highlights from R script – dplyr, plotly package, sqldf package, User defined function, Line chart)

As seen from the above time series chart, citations are very high during noon time. User defined function Count\_time has been created to calculate the count of citations at a particular time. Data frame and issue time is passed in this function to calculate the count of citations. There are more than eighteen thousand citations found in the data for the duration 12:00 and 12:05 PM. Also, the visualization shows top 10 timings in which citations are high. More number of citations occurs in the first half of the day. Tourists as well as Los Angeles citizen should make a note of this and park carefully since the citations at this time are not only for the tourist area. These citations are found in entire LA City. Street sweeping and No parking signs should be read carefully before parking since noon is the time on most streets goes through cleaning.

R code for Analysis and Visualization

#Data cleaning code in R

setwd("E:/RProject/data")

data<- read.csv('Parking\_Citations.csv', header=T, sep=',')

View(data)

install.packages('tidyr')

library(tidyr)

data1<- data[,c(1,2,3,6,16,17)]

View(data1)

data1$Issue.time<- data1$Issue.time/100

data1$Issue.time <- format(strptime(data1$Issue.time, '%H.%M'), '%I:%M %p')

names(data1)[names(data1)=="Issue.time"]<-"Issuetime"

View(data1)

fine\_amt<-data1$Fine.amount

mean(fine\_amt,na.rm=TRUE)

fine\_amt[is.na(fine\_amt)]=mean(fine\_amt, na.rm = TRUE)

data1$Fine.amount<-fine\_amt

View(data1)

Visualization#1 Geographic map for citations by the visitors from the other state

install.packages("dplyr")

library(dplyr)

install.packages("plotly")

library(plotly)

data\_f<-group\_by(data1,RP.State.Plate) %>% summarise(cnt = n()) %>% filter(RP.State.Plate != "CA")

data2 <-data.frame(data\_f)

View(data2)

map<- list(scope = 'usa',projection = list(type = 'albers usa'),showlakes = TRUE,lakecolor = toRGB('white'))

plot\_geo(data2,locationmode='USA-states') %>% add\_trace(z=~cnt,locations=~RP.State.Plate,color=~cnt,colors='Reds') %>% colorbar(title = "Citations Made by Visistors from Other States",tickprefix = '$') %>% layout(title = 'Citations Made by Visistors from Other States', geo = map)

Visualization #2: Pie chart for Total fine amount of each violation

Cnt\_Violation<-group\_by(data1,Violation.Description) %>% summarise(cnt = n())

Count\_Violation <-data.frame(Cnt\_Violation)

View(Count\_Violation)

sorted\_violation<- Count\_Violation[order(Count\_Violation$cnt,decreasing = T),]

top10\_citations<-head(sorted\_violation,10)

barplot(top10\_citations$cnt, ylim=c(0,1.1\*max(top10\_citations$cnt)), main="Top 10 Parking Citations", xlab="Citations",ylab = "Number of Citations", names.arg = top10\_citations$Citationname, cex.names=0.5,col=c("cadetblue","cadetblue1", "cadetblue2", "cadetblue3", " cadetblue4","chartreuse", "chartreuse2", "chartreuse3", "chartreuse4", "darkgreen"))

Visualization #3: Bar chart for count of different types of parking citations

**Script**

library(dplyr)

fine\_amt<-group\_by(data1,Violation.Description) %>% summarise(Total\_fine=sum(Fine.amount))

fine<-data.frame(fine\_amt)

sorted\_fine<-fine[order(fine$Total\_fine,decreasing = T),]

top5\_fine<-head(sorted\_fine,5)

View(top5\_fine)

plot\_ly(top5\_fine, labels = ~Violation.Description, values = ~Total\_fine, type = 'pie',textposition = 'inside',textinfo = 'label') %>% layout(title = 'Five Highest Number of Parking Citations', xaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE),yaxis = list(showgrid = FALSE, zeroline = FALSE, showticklabels = FALSE))

**Console**

source('barplot.R')

Visualization #4: Line chart for peak time when citations are high

**Script**

count\_time<-function(data1, Issuetime)

{

output<-sqldf("select Issuetime, count(Issuetime) as cnt from data1 group by Issuetime")

}

**Console**

source('userfunction.R')

myoutput<-count\_time(data1,Issuetime = "12:01 AM")

sort\_time<-myoutput[order(myoutput$cnt,decreasing = T),]

top10\_citations<-head(sort\_time,10)

plot\_ly(top10\_citations,x=~cnt,y=~Issuetime,type='scatter',mode='lines') %>% layout(title = 'Peak Time when Citations are High', yaxis = list(autotick = F,title= 'Time', dtick = 1), xaxis=list(title='Count of Citations'))

# Works Cited

Angeles, City of Los. *Parking Tickets*. 2016. . Online Article. 12 April 2017.

LACity. *PUBLIC PARKING*. 2016. Online Article. 13 April 2017.

Romero, Dennis. *Court Issues Huge Victory Over L.A. Parking Tickets*. 11 August 2016. Online Article. 17 April 2017.