



**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**

## A. Dataset URL

<https://data.lacity.org/A-Well-Run-City/Parking-Citations/wjz9-h9np>

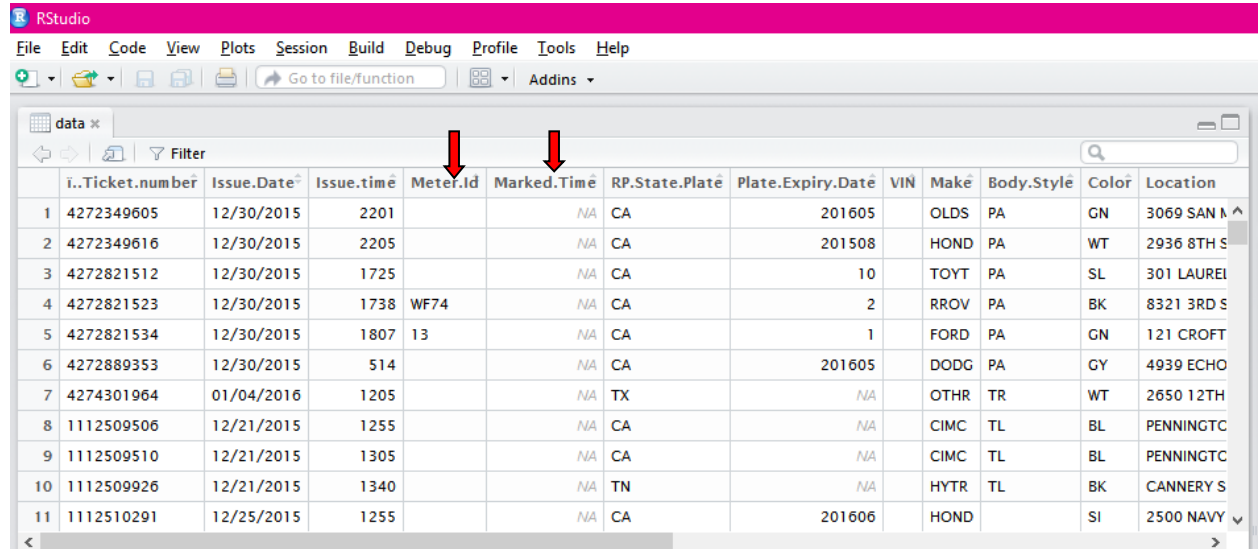
Above URL contains entire data regarding Parking Citation in City of Los Angeles from year 2010 onwards. Parking tickets or citations are issued by parking enforcement officials for violating state (California Vehicle Code) or city (Los Angeles Municipal Code) parking laws. (Angeles,2016) A parking violation 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.

## B. Data Cleaning:

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

Before cleaning:



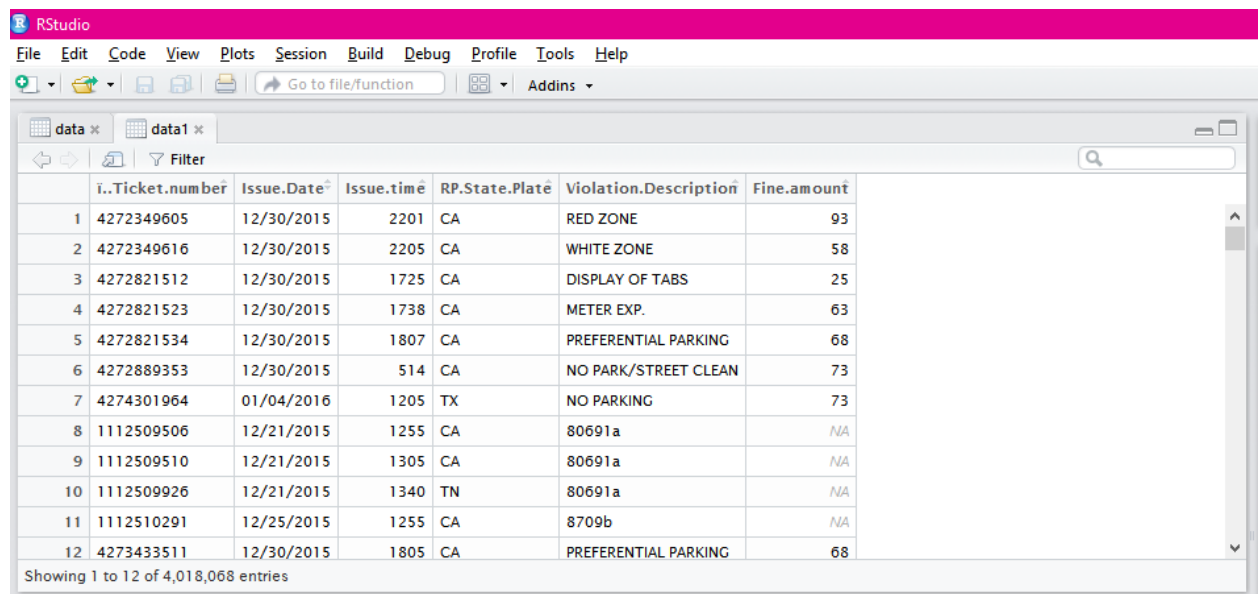
	i..Ticket.number	Issue.Date	Issue.time	Meter.Id	Marked.Time	RP.State.Plate	Plate.Expiry.Date	VIN	Make	Body.Style	Color	Location
1	4272349605	12/30/2015	2201		NA	CA	201605		OLDS	PA	GN	3069 SAN M
2	4272349616	12/30/2015	2205		NA	CA	201508		HOND	PA	WT	2936 8TH S
3	4272821512	12/30/2015	1725		NA	CA	10		TOYT	PA	SL	301 LAUREI
4	4272821523	12/30/2015	1738	WF74	NA	CA	2		RROV	PA	BK	8321 3RD S
5	4272821534	12/30/2015	1807	13	NA	CA	1		FORD	PA	GN	121 CROFT
6	4272889353	12/30/2015	514		NA	CA	201605		DODG	PA	GY	4939 ECHO
7	4274301964	01/04/2016	1205		NA	TX	NA		OTHR	TR	WT	2650 12TH
8	1112509506	12/21/2015	1255		NA	CA	NA		CIMC	TL	BL	PENNINGTC
9	1112509510	12/21/2015	1305		NA	CA	NA		CIMC	TL	BL	PENNINGTC
10	1112509926	12/21/2015	1340		NA	TN	NA		HYTR	TL	BK	CANNERY S
11	1112510291	12/25/2015	1255		NA	CA	201606		HOND		SI	2500 NAVY

Code:

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

View(data1)

After Cleaning:



	i..Ticket.number	Issue.Date	Issue.time	RP.State.Plate	Violation.Description	Fine.amount
1	4272349605	12/30/2015	2201	CA	RED ZONE	93
2	4272349616	12/30/2015	2205	CA	WHITE ZONE	58
3	4272821512	12/30/2015	1725	CA	DISPLAY OF TABS	25
4	4272821523	12/30/2015	1738	CA	METER EXP.	63
5	4272821534	12/30/2015	1807	CA	PREFERENTIAL PARKING	68
6	4272889353	12/30/2015	514	CA	NO PARK/STREET CLEAN	73
7	4274301964	01/04/2016	1205	TX	NO PARKING	73
8	1112509506	12/21/2015	1255	CA	80691a	NA
9	1112509510	12/21/2015	1305	CA	80691a	NA
10	1112509926	12/21/2015	1340	TN	80691a	NA
11	1112510291	12/25/2015	1255	CA	8709b	NA
12	4273433511	12/30/2015	1805	CA	PREFERENTIAL PARKING	68

Showing 1 to 12 of 4,018,068 entries

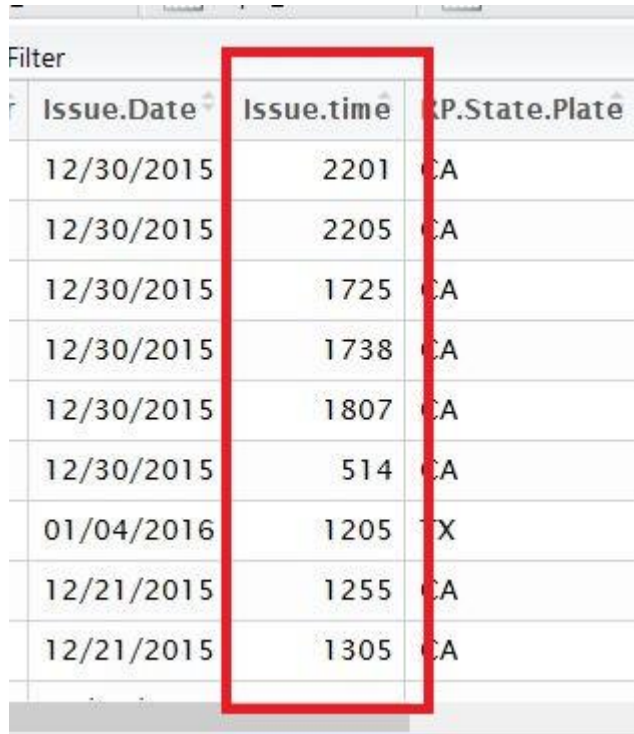
2. Issued time column has entry in numbered format instead of AM/PM format. We changed the Issue time column with the help of following code.

Code:

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

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

Before cleaning



Filter	Issue.Date	Issue.time	P.State.Plate
	12/30/2015	2201	CA
	12/30/2015	2205	CA
	12/30/2015	1725	CA
	12/30/2015	1738	CA
	12/30/2015	1807	CA
	12/30/2015	514	CA
	01/04/2016	1205	TX
	12/21/2015	1255	CA
	12/21/2015	1305	CA

After cleaning



Filter	Issue.Date	Issue.time
	12/30/2015	10:01 PM
	12/30/2015	10:05 PM
	12/30/2015	05:25 PM
	12/30/2015	05:38 PM
	12/30/2015	06:07 PM
	12/30/2015	05:14 AM
	01/04/2016	12:05 PM
	12/21/2015	12:55 PM
	12/21/2015	01:05 PM

18,068 entries

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

Before Cleaning

After Cleaning

tion	Fine.amount	Latitude	Lo
KING	68	64.48347	
CLEAN	73	99999	
	73	64.94776	
	NA	99999	
	NA	99999	
	NA	99999	
	NA	99999	
KING	68	64.73985	
	NA	99999	

escription	Fine.amount	Latitude
AL PARKING	68.00000	64.48347
REET CLEAN	73.00000	99999
	73.00000	64.94776
	69.94397	99999
	69.94397	99999
	69.94397	99999
	69.94397	99999
AL PARKING	68.00000	64.73985
	69.94397	99999

Code:

```
fine_amt<-data1$Fine.amount
```

```
mean(fine_amt,na.rm=TRUE)
```

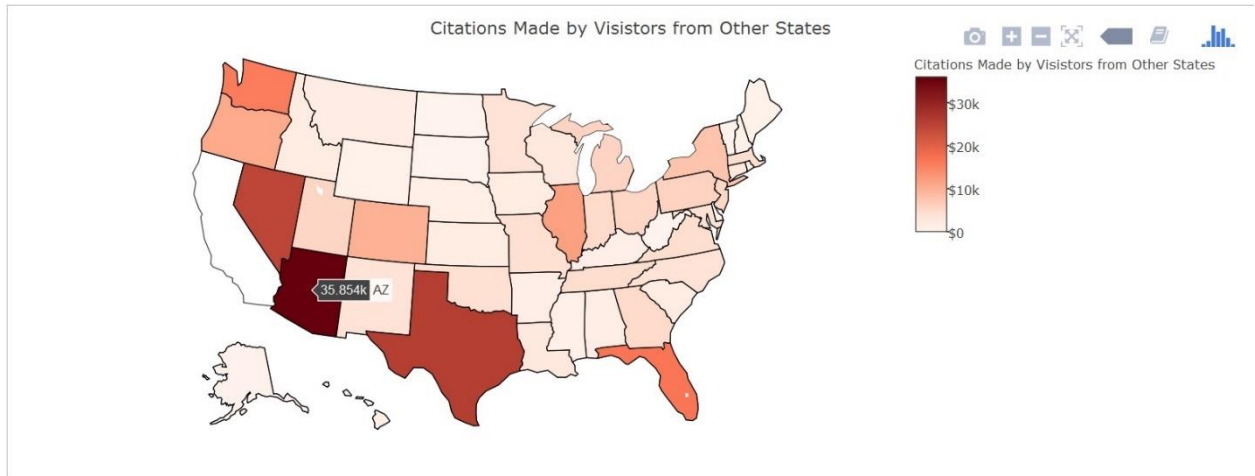
```
fine_amt[is.na(fine_amt)]=mean(fine_amt, na.rm = TRUE)
```

```
data1$Fine.amount<-fine_amt
```

```
View(data1)
```

## C. Data Visualization:

### 1. What is the number of citations made by visitors from other state?

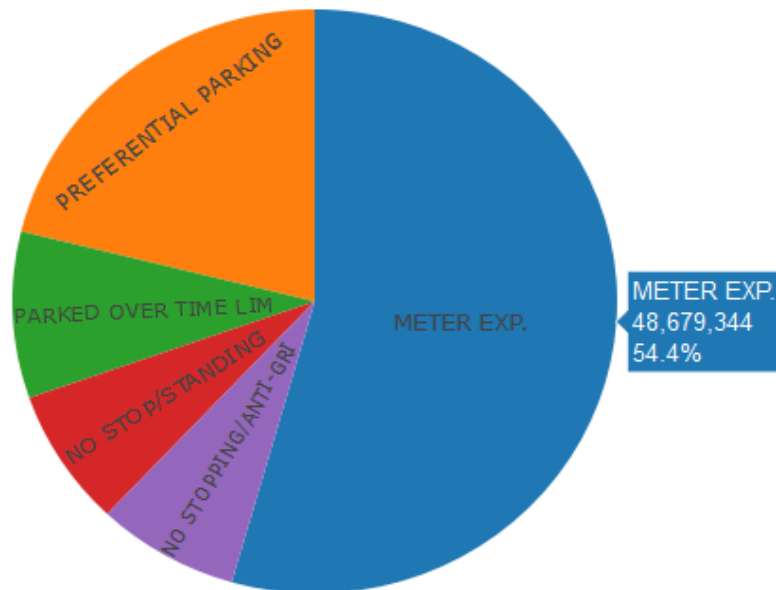


(Highlights from R script – plotly, dplyr packages, filter, geographic map)

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.

## 2. What is the total fine amount for each violation?

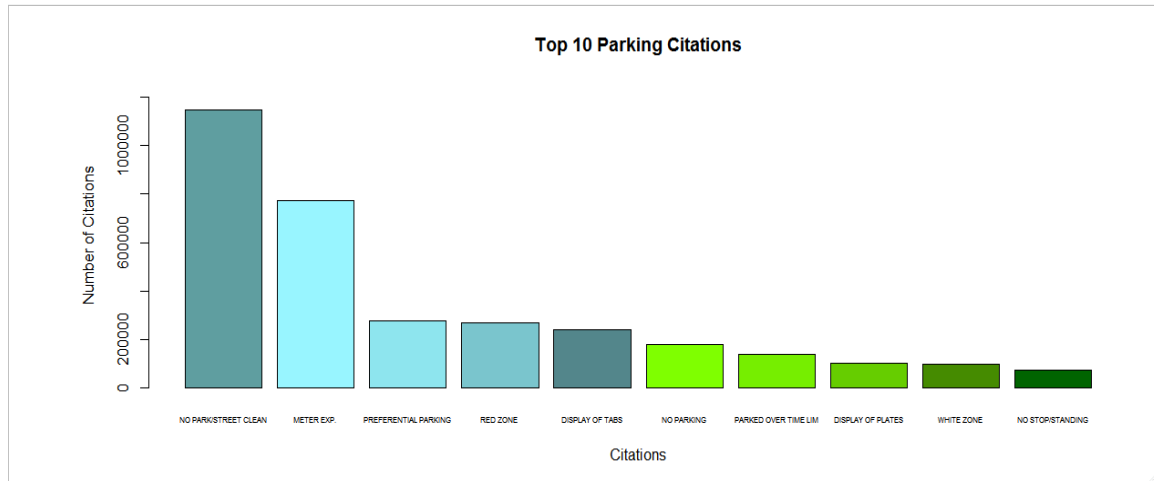
Five Highest Number of Parking Citations



(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.

### 3. How does count of different types of parking citations differ?

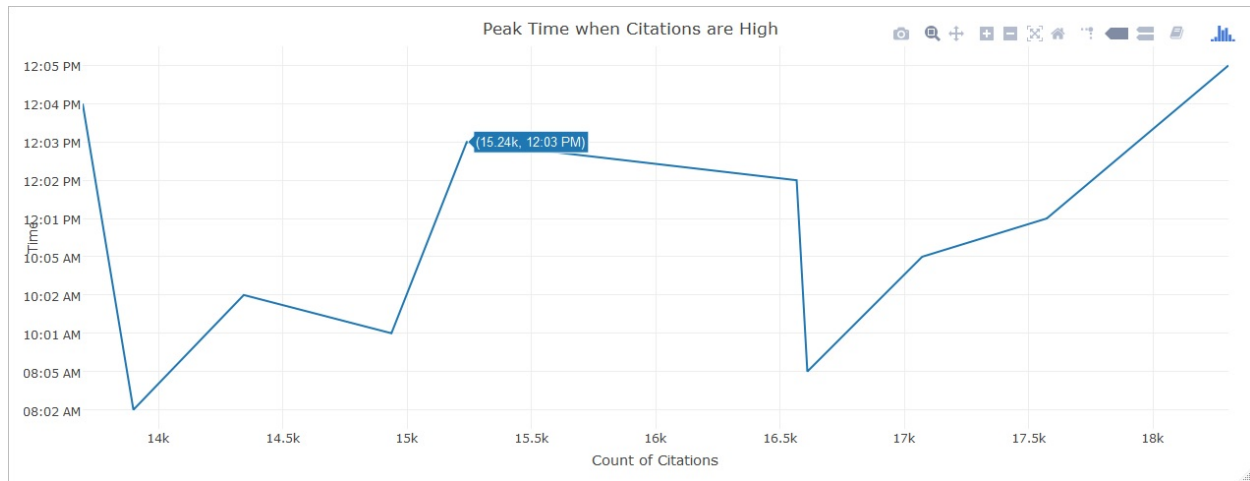


(Highlights from R script – dplyr, barplot, bar chart, different colors)

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.



#### 4. 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.