

# **NYC Parking Tickets - An Exploratory Analysis**

**(Case Study)**

Team,

Hariharan Subramaniam  
Thriphthi Raj  
Sivaiah Aelosrolu  
Abhishek Pandey

## **Problem Statement**

New York City is a thriving metropolis. Just like most other metros that size, one of the biggest problems its citizens face is parking. The classic combination of a huge number of cars and a cramped geography is the exact recipe that leads to a huge number of parking tickets.

The objective of this case study is to perform exploratory data analysis on the 'NYC Parking Tickets' dataset for the fiscal year 2014-2015, 2015-2016 and 2016-2017.

## **Approach**

1. Setup the Spark R environment.
2. Setup the SQL environment to run SQL like queries.
3. Import the data for analysis that is kept in the Hadoop cluster.
4. Perform Exploratory data analysis.
5. Check data quality issues and fix where possible.
6. Derive new variables where needed to answer the questions asked in the case study.
7. Subset the data so the data reflects the current fiscal years.
8. Answer the questions asked in case study.
9. Add plots where possible.
10. Submit the assignment with a presentation and coding file.

## **Exploratory data Analysis**

### **1. Explore the data**

There were 3 datasets that were imported from the following locations. Each dataset had data for fiscal years 2014-2015, 2015-2016 and 2016-2017 respectively,

/common\_folder/nyc\_parking/Parking\_Violations\_Issued\_-\_Fiscal\_Year\_2015.csv  
/common\_folder/nyc\_parking/Parking\_Violations\_Issued\_-\_Fiscal\_Year\_2016.csv  
/common\_folder/nyc\_parking/Parking\_Violations\_Issued\_-\_Fiscal\_Year\_2017.csv

### **Dimensions :**

#### **Dataset for fiscal year 2014-2015**

1. There were 11809233 rows and 51 columns

#### **Dataset for fiscal year 2015-2016**

1. There were 10626899 rows and 51 columns

#### **Dataset for fiscal year 2016-2017**

1. There were 10803028 rows and 43 columns

Note : The number of columns in this dataset is relatively lesser than the dataset of 2014-2015 and 2015-2016.

**Column name of the three datasets**

| <u>fiscal year 2014-2015</u>  | <u>fiscal year 2015-2016</u>  | <u>fiscal year 2016-2017</u>  |
|---|---|---|
| "Summons Number"<br>"Plate ID"<br>"Registration State"                      | "Summons Number"<br>"Plate ID"<br>"Registration State"                      | "Summons Number"<br>"Plate ID"<br>"Registration State"                      |
| "Plate Type"<br>"Issue Date"<br>"Violation Code"                            | "Plate Type"<br>"Issue Date"<br>"Violation Code"                            | "Plate Type"<br>"Issue Date"<br>"Violation Code"                            |
| "Vehicle Body Type"<br>"Vehicle Make"<br>"Issuing Agency"                   | "Vehicle Body Type"<br>"Vehicle Make"<br>"Issuing Agency"                   | "Vehicle Body Type"<br>"Vehicle Make"<br>"Issuing Agency"                   |
| "Street Code1"<br>"Street Code2"<br>"Street Code3"                          | "Street Code1"<br>"Street Code2"<br>"Street Code3"                          | "Street Code1"<br>"Street Code2"<br>"Street Code3"                          |
| "Vehicle Expiration Date"<br>"Violation Location"<br>"Violation Precinct"   | "Vehicle Expiration Date"<br>"Violation Location"<br>"Violation Precinct"   | "Vehicle Expiration Date"<br>"Violation Location"<br>"Violation Precinct"   |
| "Issuer Precinct"<br>"Issuer Code"<br>"Issuer Command"                      | "Issuer Precinct"<br>"Issuer Code"<br>"Issuer Command"                      | "Issuer Precinct"<br>"Issuer Code"<br>"Issuer Command"                      |
| "Issuer Squad"<br>"Violation Time"<br>"Time First Observed"                 | "Issuer Squad"<br>"Violation Time"<br>"Time First Observed"                 | "Issuer Squad"<br>"Violation Time"<br>"Time First Observed"                 |
| "Violation County"<br>"Violation In Front Of Or<br>Opposite" "House Number" | "Violation County"<br>"Violation In Front Of Or<br>Opposite" "House Number" | "Violation County"<br>"Violation In Front Of Or<br>Opposite" "House Number" |
| "Street Name"<br>"Intersecting Street"<br>"Date First Observed"             | "Street Name"<br>"Intersecting Street"<br>"Date First Observed"             | "Street Name"<br>"Intersecting Street"<br>"Date First Observed"             |
| "Law Section"<br>"Sub Division"<br>"Violation Legal Code"                   | "Law Section"<br>"Sub Division"<br>"Violation Legal Code"                   | "Law Section"<br>"Sub Division"<br>"Violation Legal Code"                   |

|   |   |   |
|---|---|---|
| "Days Parking In Effect"<br>"From Hours In Effect"<br>"To Hours In Effect"            | "Days Parking In Effect"<br>"From Hours In Effect"<br>"To Hours In Effect"            | "Days Parking In Effect"<br>"From Hours In Effect"<br>"To Hours In Effect"            |
| "Vehicle Color"<br>"Unregistered Vehicle?"<br>"Vehicle Year"                          | "Vehicle Color"<br>"Unregistered Vehicle?"<br>"Vehicle Year"                          | "Vehicle Color"<br>"Unregistered Vehicle?"<br>"Vehicle Year"                          |
| "Meter Number"<br>"Feet From Curb"<br>"Violation Post Code"                           | "Meter Number"<br>"Feet From Curb"<br>"Violation Post Code"                           | "Meter Number"<br>"Feet From Curb"<br>"Violation Post Code"                           |
| "Violation Description"<br>"No Standing or Stopping<br>Violation" "Hydrant Violation" | "Violation Description"<br>"No Standing or Stopping<br>Violation" "Hydrant Violation" | "Violation Description"<br>"No Standing or Stopping<br>Violation" "Hydrant Violation" |
| "Double Parking Violation"<br>"Latitude"<br>"Longitude"                               | "Double Parking Violation"<br>"Latitude"<br>"Longitude"                               | "Double Parking Violation"  |
| "Community Board"<br>"Community Council "<br>"Census Tract"                           | "Community Board"<br>"Community Council "<br>"Census Tract"                           |   |
| "BIN"<br>"BBL"<br>"NTA"   | "BIN"<br>"BBL"<br>"NTA"   |   |

## 2.Check data quality issues

- 1.The column names have spaces and unwanted characters such as ?.This can be cleaned
- 2.The columns "Latitude","Longitude","Community Board","Community Council ","Census Tract", "BIN", "BBL" and "NTA" are not present in 2017 dataset.
- 3.Summons\_numbers seems to be the unique id that represents a tickets .However the dataset for 2014-2015 seems to have duplicate entries of the Summons\_Numbers.Datasets 2015-2016 and 2016-2017 do not have duplicate entries
- 4.The plate\_ID is a combination of Alphanumeric value,However there is one value that is fully alphabetic.This is just an observation from the first five rows.This may not be an issue for the analysis.
- 5.The Registration state column seems to have values that are in numeric form as well which seems to be erroneous.This needs to be fixed as the question in the case study also demands the fix.

- 6.The street code columns seems to have values that are zeroes.
- 7.The Vehicle expiration date seems to have dates in erroneous manner as some numeric values are concatenated after year part of the timestamp.This will be fixed if necessary.
- 8.The violation time column seems to have values such as 0953A,0520P etc and the A and P seems to meant to be entered as AM and PM.This will be fixed.
- 9.Following rows of the dataset 2014-2015 and 2015-2016 appear to have NAs -Violation\_Post\_Code, Violation\_Description ,No\_Standing\_or\_Stopping\_Violation Hydrant\_Violation, Double\_Parking\_Violation Latitude Longitude Community\_Board Community\_Council Census\_Tract BIN BBL NTA.
- 10.Columns From\_Hours\_In\_Effect and To\_Hours\_In\_Effect seem to have time value entered erroneously
- 11.Date\_First\_Observed column seems to have year part of the date erroneous.

### **3.Fixing data quality issues**

The data quality issues stated below are observed across all columns.Where it is specific to a dataset it is mentioned.

- 1.*Column Names*: For all the imported data-frames fix the column names ,Remove starting and trailing spaces and add under scores between texts.
- 2.*Remove duplicates* : the dataset 2014-2015 has duplicate entries in Summons\_numbers ,The duplicate entries were removed for the dataset 2014-2015.  
Whereas the dataset 2015-2016 and 2016-2017 do not have duplicate entries in Summons\_numbers.
- 3.*Date format* : Convert The issue\_date column into date format.There are few observations across all datasets that have issue\_date entered out of range i.e not within the fiscal year and appear to be erroneous.
- 4.*Null values* : The columns of dataset 2014-2015 and 2015-2016 "Latitude","Longitude","Community Board","Community Council ","Census Tract","BIN","BBL" and "NTA" are entirely NAs and they can be removed.
- 5.*Fix time formats* : The columns Violation\_Time,Time\_First\_Observed,From\_Hours\_In\_Effect and To\_Hours\_In\_Effect do not have time entered in the full form of AM/PM.The 'M' is missing in AM/PM.This is was fixed.
- 6.*Timestamp columns* : The columns Violation\_Time,Time\_First\_Observed,From\_Hours\_In\_Effect and To\_Hours\_In\_Effect need to be converted into timestamps.

### **4.Derived Variables**

- 1.Columnn names Violation\_Time across all data frames will used to derive new column names such as hours and minutes that can be used to answer questions asked in the case study.

### **4.Subset data**

#### **1.Dataset 2014-2015**

As there are rows with incorrect dates which are out of range of the fiscal year 2014-2015.The dataset was subsetted for 1st July 2014 - 30 June 2015.

## 2.Dataset 2015-2016

As there are rows with incorrect dates which are out of range of the fiscal year 2014-2015.The dataset was subsetted for 1st July 2015 - 30 June 2016.

## 3.Dataset 2016-2017

As there are rows with incorrect dates which are out of range of the fiscal year 2014-2015.The dataset was subsetted for 1st July 2016 - 30 June 2017.

### Dimensions of the Dataframe for Analysis

Dimensions of the cleaned Dataframe were as below,

Fiscal year 2014-2015: Rows : 10598035 Cols : 48

Fiscal year 2015-2016: Rows : 10598035 Cols : 48

Fiscal year 2016-2017: Rows : 10598035 Cols : 48

### Answering the Case study questions

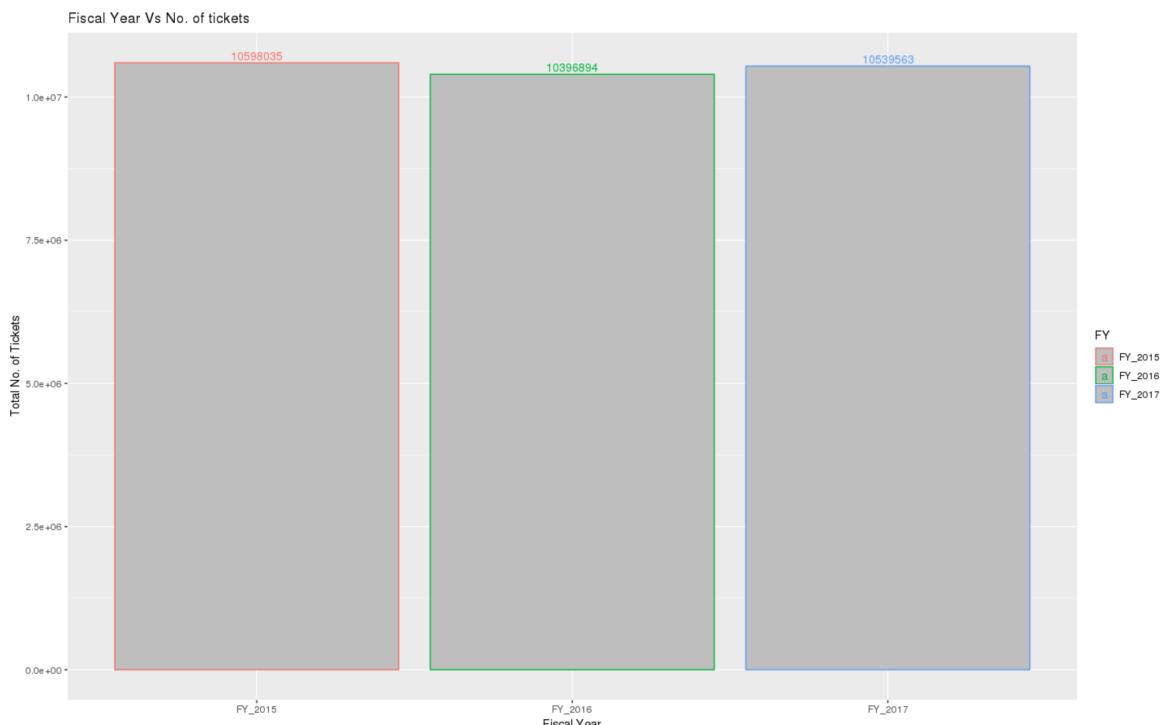
#### **Examine the data**

- Find the total number of tickets for each year.

**Ans:**

| # | FY      | Total_tickets |
|---|---------|---------------|
| # | FY_2015 | 10598035      |
| # | FY_2016 | 10396894      |
| # | FY_2017 | 10539563      |

#### **Plot :**



2. Find out the number of unique states from where the cars that got parking tickets came from. (Hint: Use the column 'Registration State') There is a numeric entry in the column which should be corrected. Replace it with the state having maximum entries. Give the number of unique states for each year again.

There was a numeric value called '99' in Registration State and that was replaced with 'NY' as NY had maximum entries in all data frames of fiscal year 2014-2015, 2015-2016 and 2016-2017

#### **For fiscal year 2014-2015**

There were 68 unique states in the data frame of 2014-2015

Snapshot of top5 as below

|   | <b>Registration_State.</b> | <b>No_tickets_state</b> |
|---|----------------------------|-------------------------|
| 1 | NY                         | 8294768                 |
| 2 | NJ                         | 969222                  |
| 3 | PA                         | 261192                  |
| 4 | CT                         | 144719                  |
| 5 | FL                         | 133980                  |

#### **For fiscal year 2015-2016**

There were 67 unique states in the data frame of 2014-2015

*Snapshot of top5 as below*

|   | <b>Registration_State</b> | <b>No_tickets_state</b> |
|---|---------------------------|-------------------------|
| 1 | NY                        | 8123559                 |
| 2 | NJ                        | 949163                  |
| 3 | PA                        | 252681                  |
| 4 | CT                        | 142006                  |
| 5 | FL                        | 135273                  |

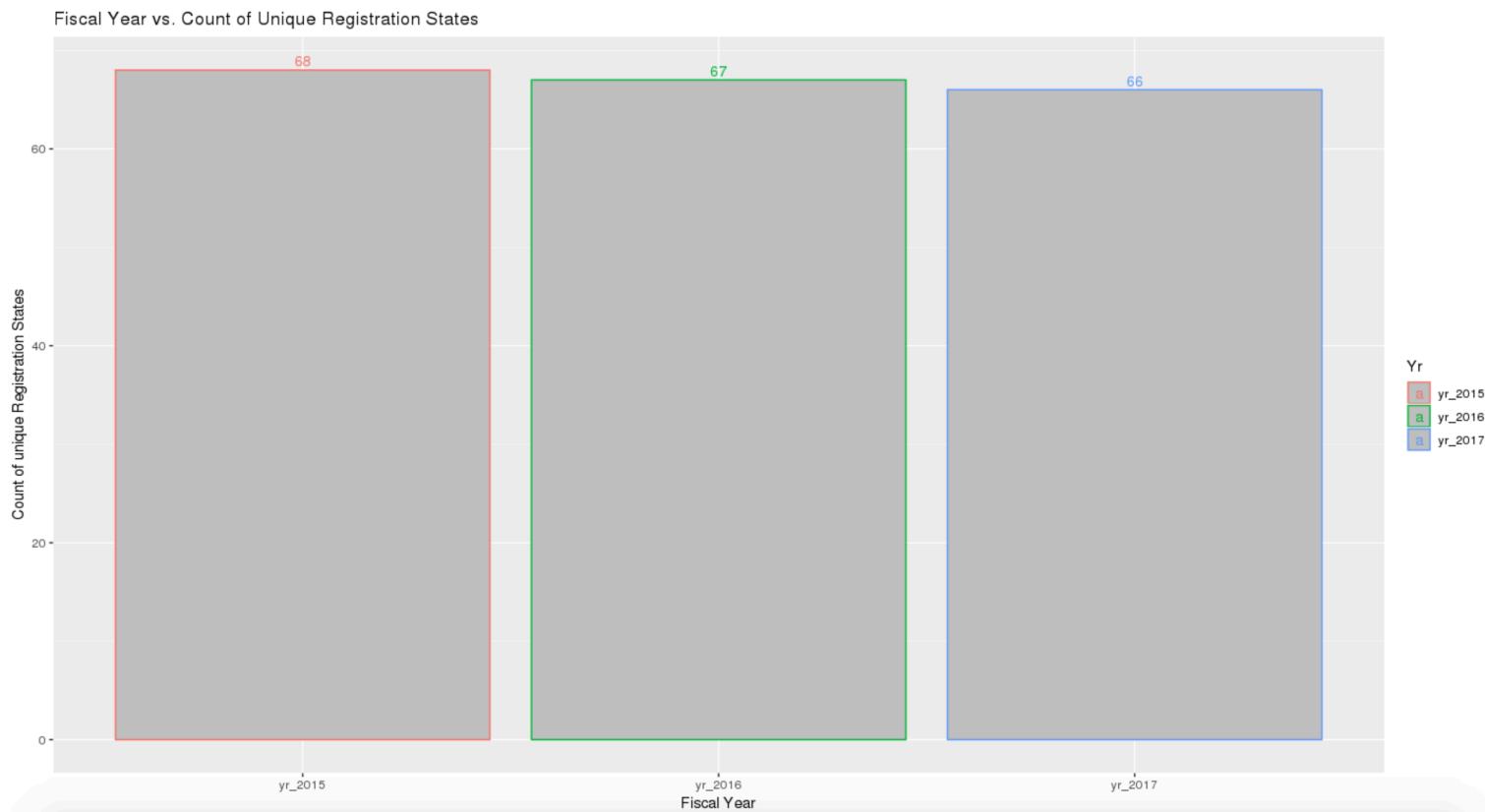
#### **For fiscal year 2016-2017**

There were 66 unique states in the data frame of 2016-2017

*Snapshot of top5 as below*

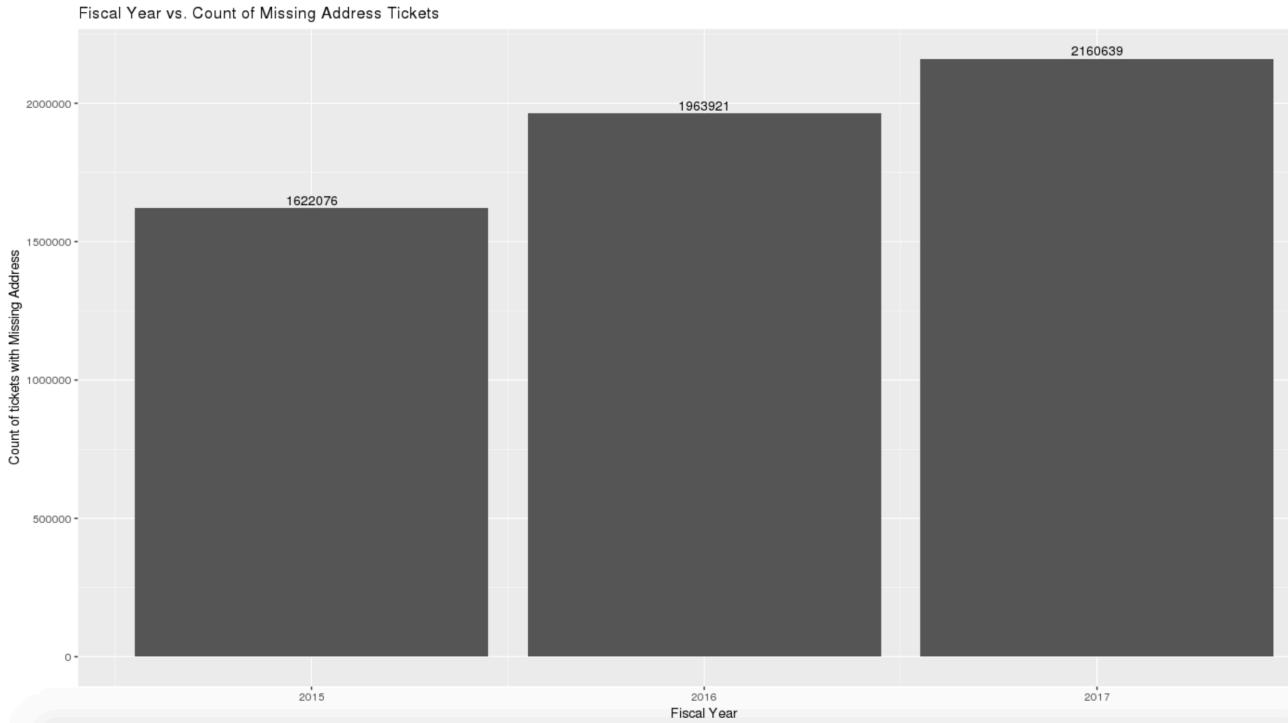
|   | <b>Registration_State</b> | <b>No_tickets_state</b> |
|---|---------------------------|-------------------------|
| 1 | NY                        | 8307594                 |
| 2 | NJ                        | 905942                  |
| 3 | PA                        | 278005                  |
| 4 | FL                        | 140624                  |
| 5 | CT                        | 137963                  |

**Plot:**



3. Some parking tickets don't have the address for violation location on them, which is a cause for concern. Write a query to check the number of such tickets.

1. About 1622076 records(15% of the total record) of the data frame 2014\_2015 have missing address for violation location
2. About 1963921 records(19% of the total record) of the data frame 2015\_2016 have missing address for violation location
3. About 2160639 records(20% of the total record) of the data frame 2016\_2017 have missing address for violation location



### Aggregation tasks

1. How often does each violation code occur? Display the frequency of the top five violation codes.

**Ans:**

#### For fiscal year 2014 - 2015

|   | <b>Violation_Code</b> | <b>Frequency_of_Tickets</b> |
|---|-----------------------|-----------------------------|
| 1 | 21                    | 1469228                     |
| 2 | 38                    | 1305007                     |
| 3 | 14                    | 908418                      |
| 4 | 36                    | 747098                      |
| 5 | 37                    | 735600                      |

#### For fiscal year 2015 - 2016

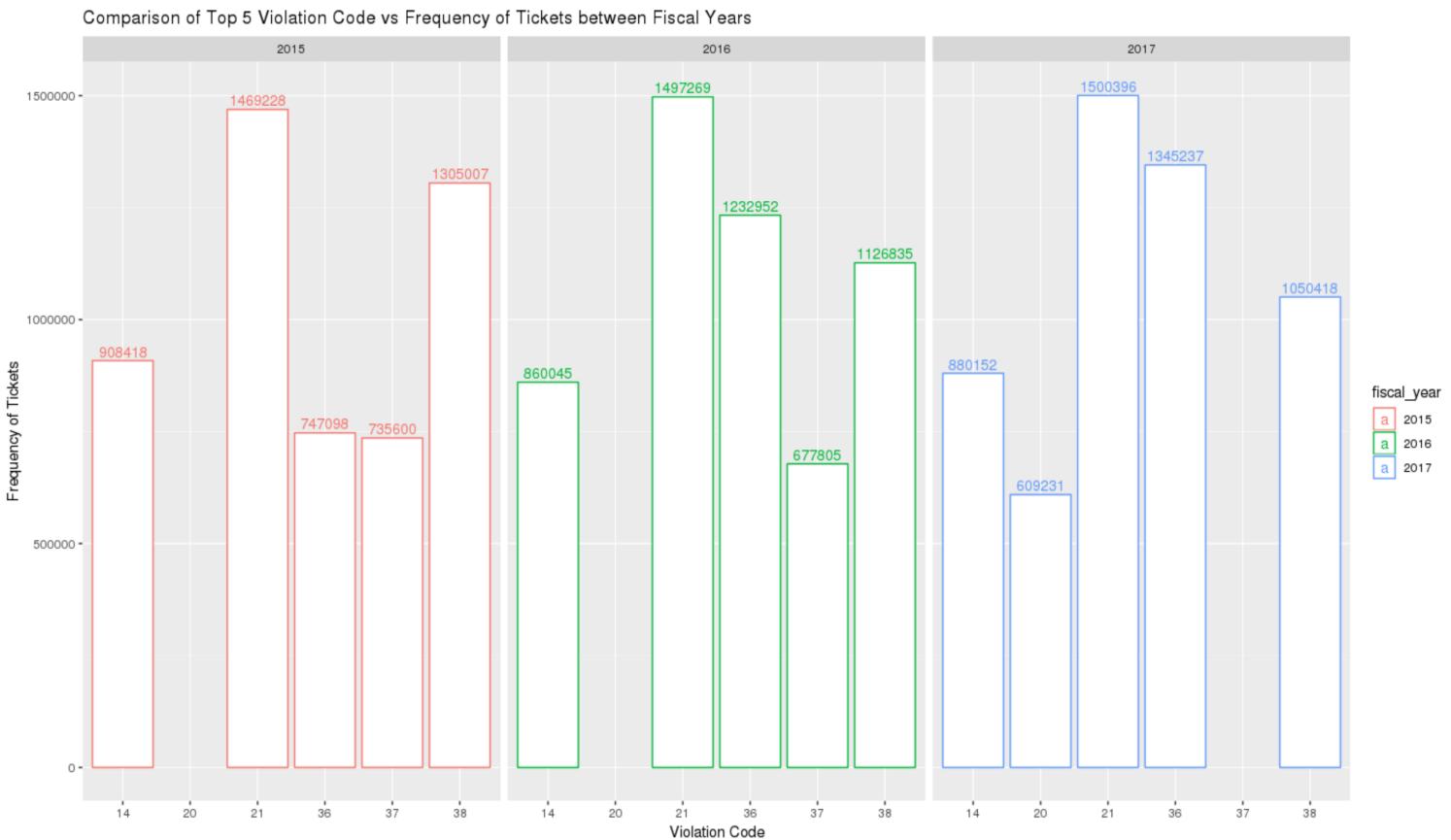
|   | <b>Violation_Code</b> | <b>Frequency_of_Tickets</b> |
|---|-----------------------|-----------------------------|
| 1 | 21                    | 1497269                     |
| 2 | 36                    | 1232952                     |
| 3 | 38                    | 1126835                     |
| 4 | 14                    | 860045                      |
| 5 | 37                    | 677805                      |

#### For fiscal year 2016 - 2017

|   | <b>Violation_Code</b> | <b>Frequency_of_Tickets</b> |
|---|-----------------------|-----------------------------|
| 1 | 21                    | 1500396                     |
| 2 | 36                    | 1345237                     |
| 3 | 38                    | 1050418                     |
| 4 | 14                    | 880152                      |

5            20            609231

### Combined Chart



2. How often does each 'vehicle body type' get a parking ticket? (**Hint:** find the top 5 for both)

### For fiscal year 2014 - 2015

|   | Vehicle_Body_Type | Frequency_of_Tickets |
|---|-------------------|----------------------|
| 1 | SUBN              | 3341110              |
| 2 | 4DSD              | 3001810              |
| 3 | VAN               | 1570227              |
| 4 | DELV              | 822040               |
| 5 | SDN               | 428571"              |

### For fiscal year 2015 - 2016

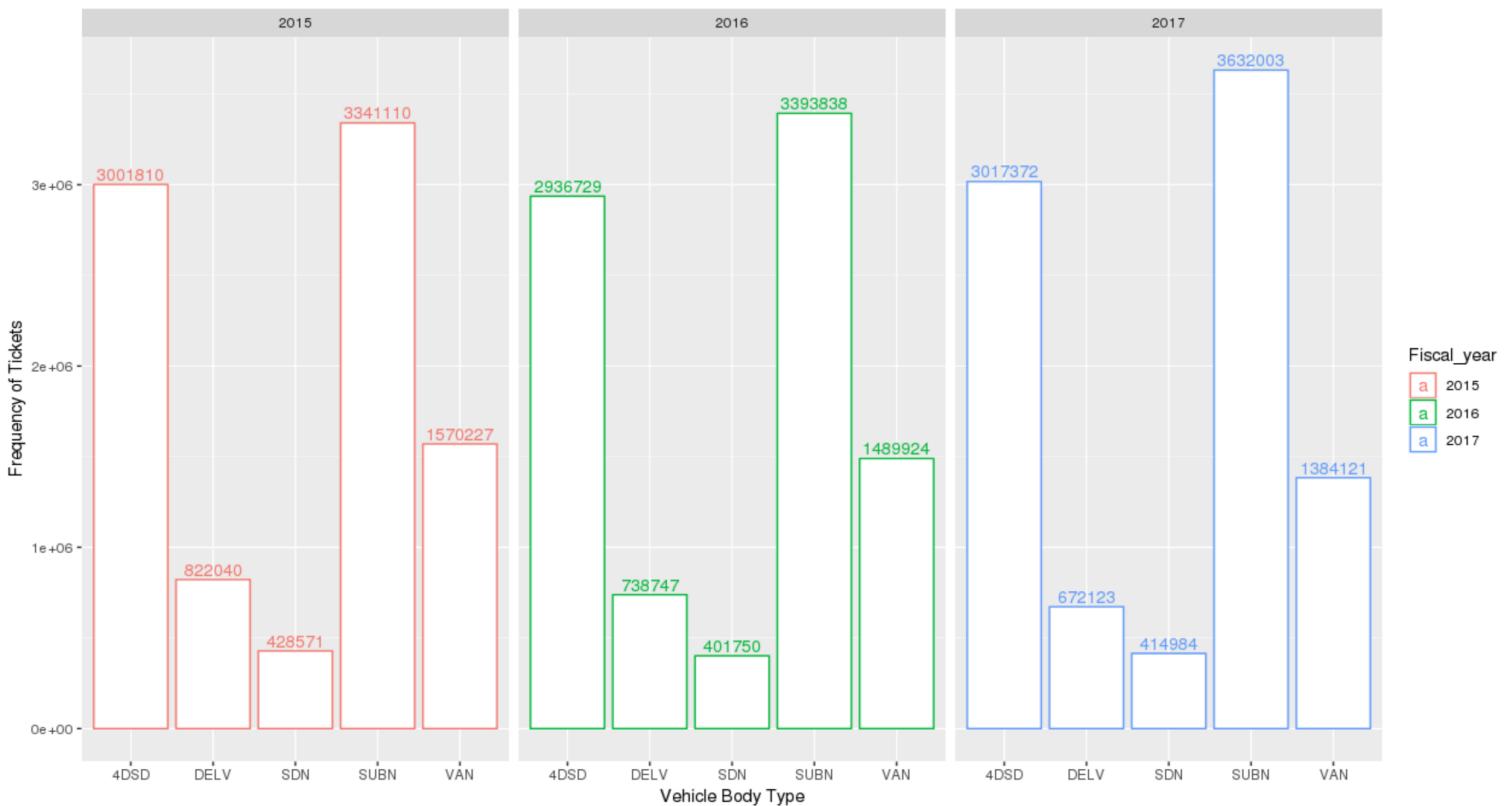
|   | Vehicle_Body_Type | Frequency_of_Tickets |
|---|-------------------|----------------------|
| 1 | SUBN              | 3393838              |
| 2 | 4DSD              | 2936729              |
| 3 | VAN               | 1489924              |
| 4 | DELV              | 738747               |
| 5 | SDN               | 401750"              |

### For fiscal year 2016 - 2017

| Vehicle_Body_Type | Frequency_of_Tickets |
|-------------------|----------------------|
| 1 SUBN            | 3632003              |
| 2 4DSD            | 3017372              |
| 3 VAN             | 1384121              |
| 4 DELV            | 672123               |
| 5 SDN             | 414984               |

### Combined Chart:

Comparison of Top 5 Violation Bodytype vs Frequency of Ticket between Fiscal Years



### 2.2 How about the 'vehicle make'?

### For fiscal year 2014 - 2015

| Vehicle_Make | Frequency_of_Tickets |
|--------------|----------------------|
| 1 FORD       | 1373157              |
| 2 TOYOT      | 1082206              |
| 3 HONDA      | 982130               |
| 4 CHEVR      | 811659               |
| 5 NISSA      | 805572               |

### For fiscal year 2015 - 2016

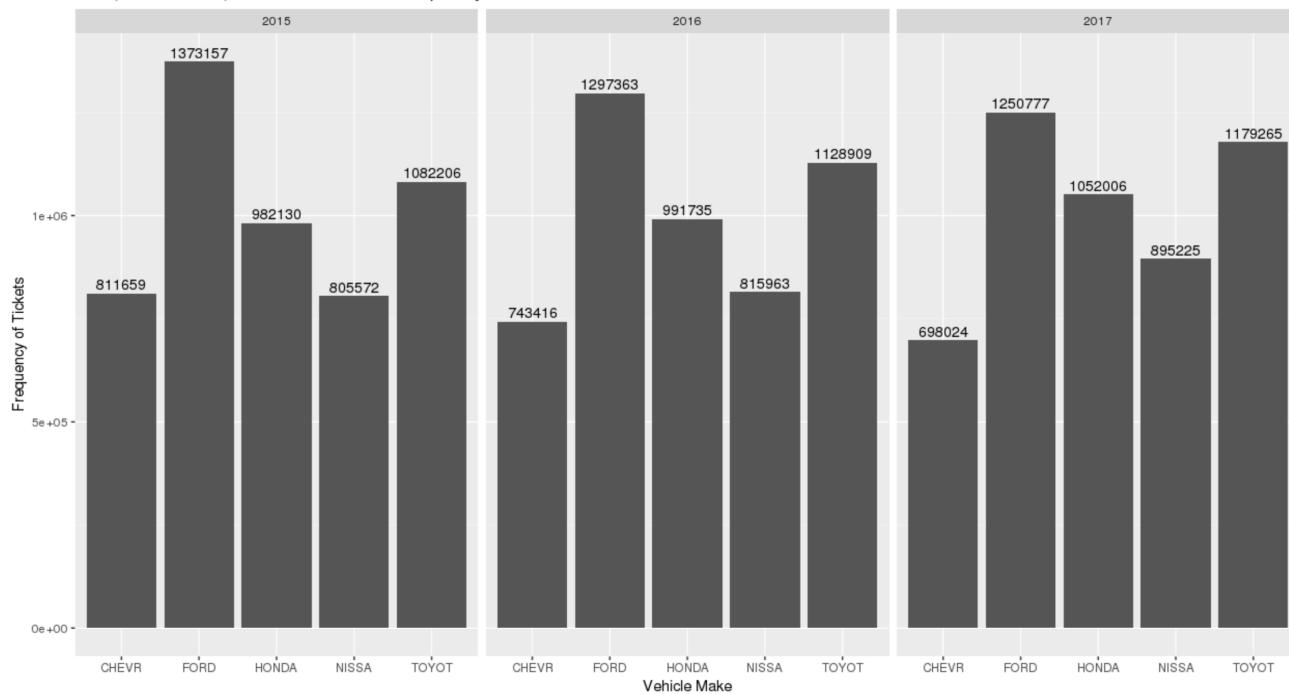
| Vehicle_Make Frequency_of_Tickets |       |         |
|-----------------------------------|-------|---------|
| 1                                 | FORD  | 1297363 |
| 2                                 | TOYOT | 1128909 |
| 3                                 | HONDA | 991735  |
| 4                                 | NISSA | 815963  |
| 5                                 | CHEVR | 743416  |

### For fiscal year 2016 - 2017

| Vehicle_Make Frequency_of_Tickets |       |         |
|-----------------------------------|-------|---------|
| 1                                 | FORD  | 1250777 |
| 2                                 | TOYOT | 1179265 |
| 3                                 | HONDA | 1052006 |
| 4                                 | NISSA | 895225  |
| 5                                 | CHEVR | 698024  |

### Combined Chart

Comparison of Top 5 Vehicle Make vs Frequency of Ticket between Fiscal Years



3. A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequency of tickets for each of the following:
1. 'Violation Precinct' (this is the precinct of the zone where the violation occurred). Using this, can you make any insights for parking violations in any specific areas of the city?

### For fiscal year 2014 - 2015

| Violation_Precinct | Frequency_of_Tickets |
|--------------------|----------------------|
| 1                  | 0                    |
| 2                  | 19                   |
| 3                  | 18                   |
| 4                  | 14                   |
|                    | 1455166              |
|                    | 550797               |
|                    | 393802               |
|                    | 377750               |

|   |     |        |
|---|-----|--------|
| 5 | 1   | 302737 |
| 6 | 114 | 295855 |

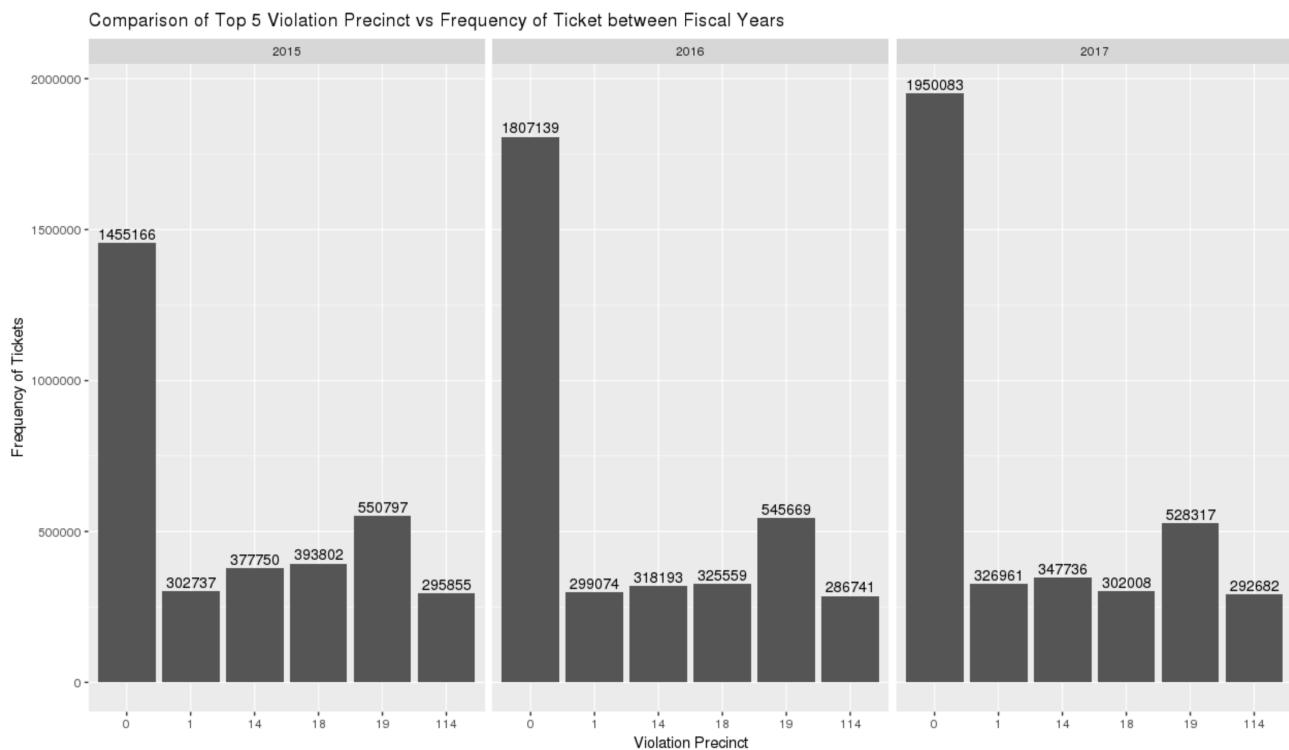
#### For fiscal year 2015 - 2016

| Violation_Precinct | Frequency_of_Tickets |
|--------------------|----------------------|
| 1                  | 0                    |
| 2                  | 19                   |
| 3                  | 18                   |
| 4                  | 14                   |
| 5                  | 1                    |
| 6                  | 114                  |

#### For fiscal year 2016 - 2017

| Violation_Precinct. | Frequency_of_Tickets |
|---------------------|----------------------|
| 1                   | 0                    |
| 2                   | 19                   |
| 3                   | 14                   |
| 4                   | 1                    |
| 5                   | 18                   |
| 6                   | 114                  |

#### Combined plot :



- 'Issuer Precinct' (this is the precinct that issued the ticket)

#### For fiscal year 2014 - 2015

| Issuer_Precinct | Frequency_of_Tickets |
|-----------------|----------------------|
| 1               | 0                    |
| 2               | 19                   |

|   |     |        |
|---|-----|--------|
| 3 | 18  | 384863 |
| 4 | 14  | 363734 |
| 5 | 1   | 293942 |
| 6 | 114 | 291100 |

#### **For fiscal year 2015 - 2016**

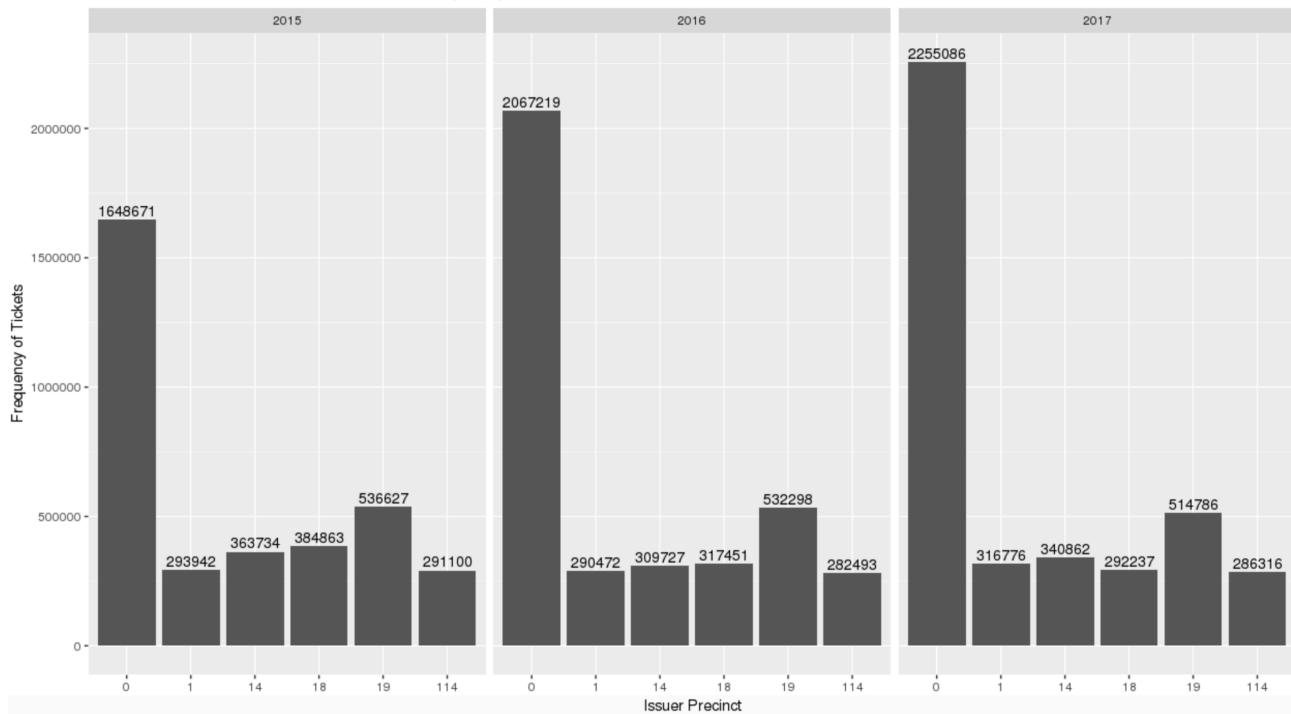
| Issuer_Precinct | Frequency_of_Tickets |         |
|-----------------|----------------------|---------|
| 1               | 0                    | 2067219 |
| 2               | 19                   | 532298  |
| 3               | 18                   | 317451  |
| 4               | 14                   | 309727  |
| 5               | 1                    | 290472  |
| 6               | 114                  | 282493  |

#### **For fiscal year 2016 - 2017**

| Issuer_Precinct | Frequency_of_Tickets |         |
|-----------------|----------------------|---------|
| 1               | 0                    | 2255086 |
| 2               | 19                   | 514786  |
| 3               | 14                   | 340862  |
| 4               | 1                    | 316776  |
| 5               | 18                   | 292237  |
| 6               | 114                  | 286316  |

#### **Combined Plot**

Comparison of Top 5 Issuer Precinct vs Frequency of Ticket between Fiscal Years



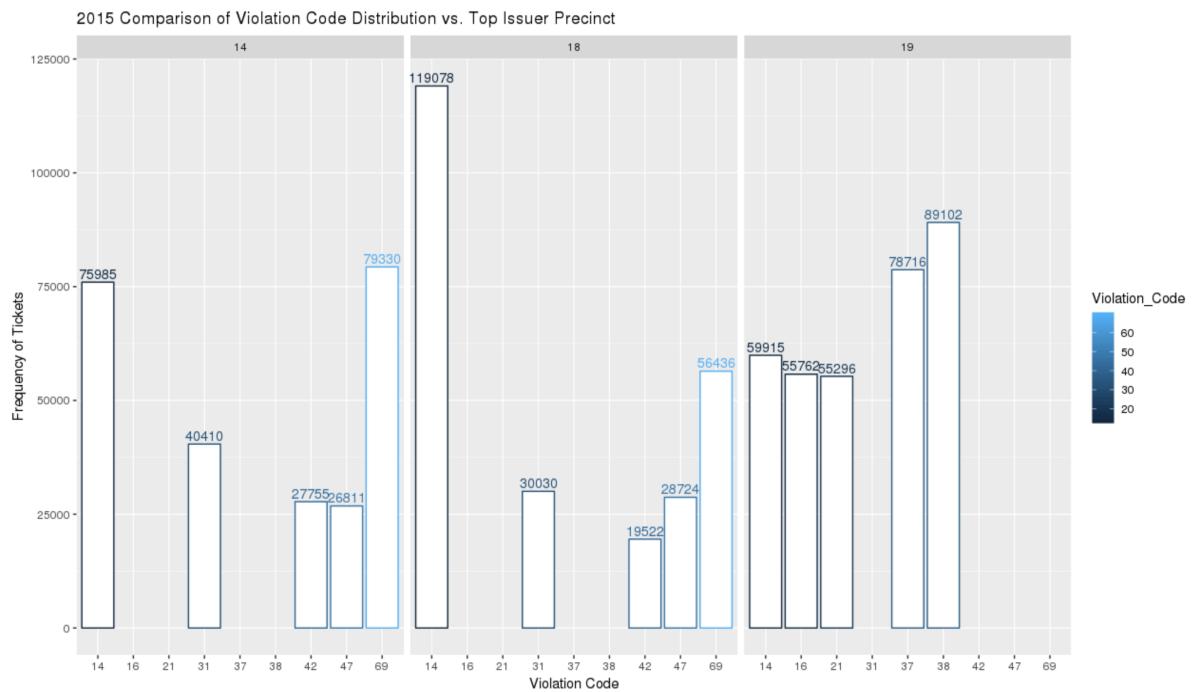
- Find the violation code frequency across three precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?

#### **For fiscal year 2014 - 2015**

#Violation code 14 is common across Precinct 14,18,19

#69,14,31 are the top three in precinct 14  
#69,14,31 are again top 3 in precinct 18  
#38,37,14 are top 3 in precinct 19

### Combined Plot

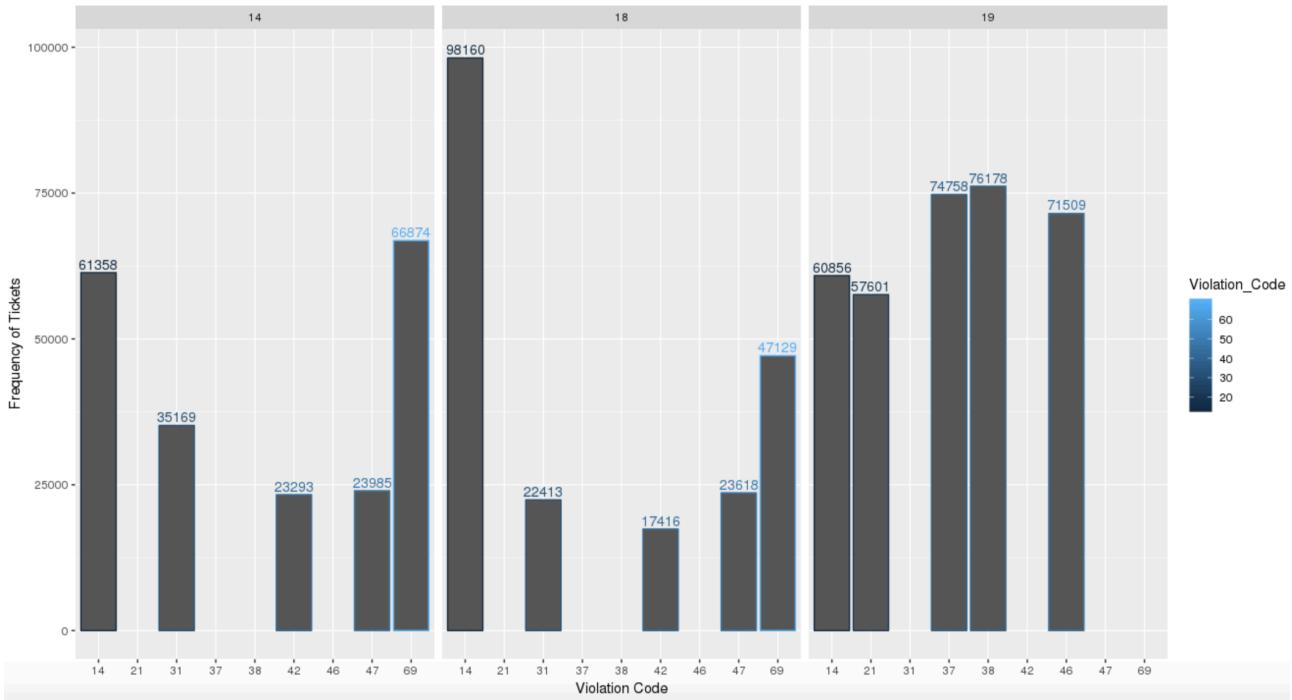


## For fiscal year 2015 - 2016

#69,14,31 are top 3 in precinct 14  
#14,69,47 are top 3 in precinct 18  
#38,37,46 are top 3 in precinct 19

## Combined Plot

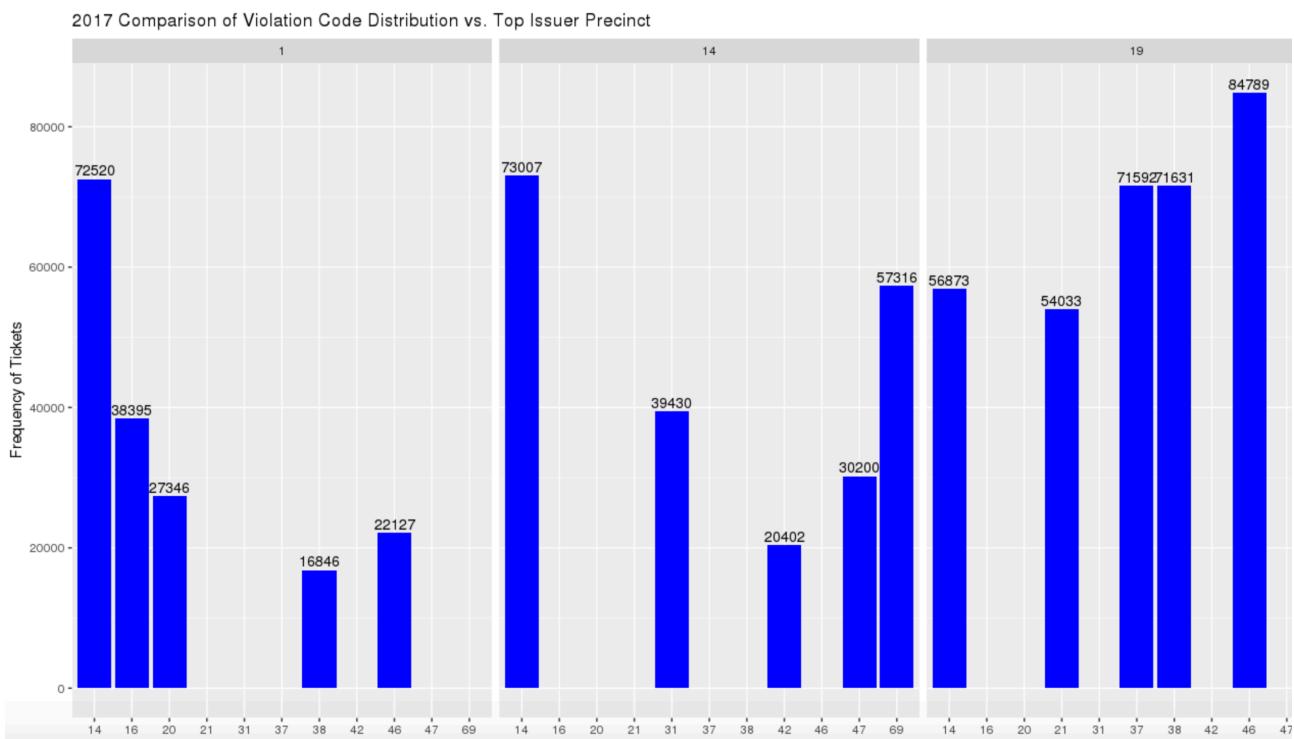
2016 Comparison of Violation Code Distribution vs. Top Issuer Precinct



## For fiscal year 2016 - 2017

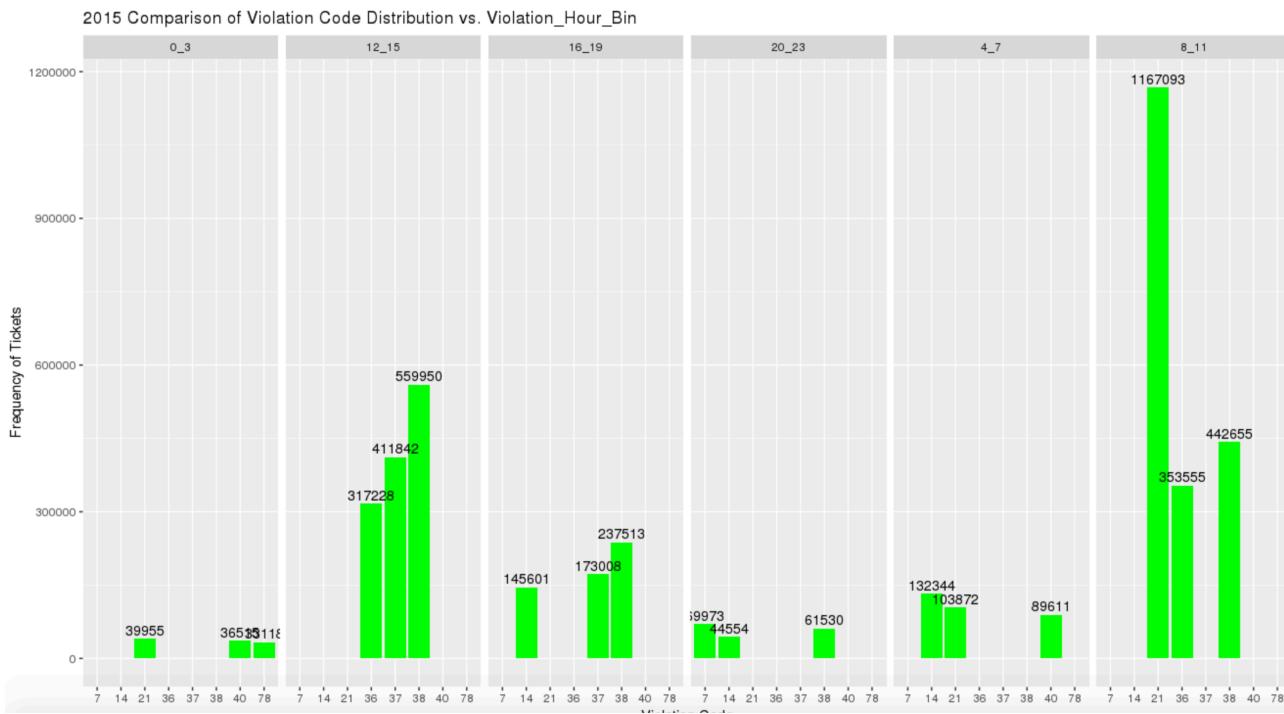
```
#top 3 violation codes in precinct 1 are 14,16,20  
#top 3 violation codes in precinct 14 are 14,69,31  
#top 3 violation codes in precinct 19 are 46,37,38 .Also 14 is seen in this precinct as 4th  
#violation code 14 is common across all 3 precincts
```

## Combined Plot



5. You'd want to find out the properties of parking violations across different times of the day:
  1. Find a way to deal with missing values, if any.  
**Hint:** Check for the null values using 'isNull' under the SQL. Also, to remove the null values, check the 'dropna' command in the API documentation.  
**Ans.** All the NA's were dropped from violation\_time column using *is not null* filter in *subset* function
  2. The Violation Time field is specified in a strange format. Find a way to make this into a time attribute that you can use to divide into groups.  
**Ans.** casted into data type 'string'
    3. Divide 24 hours into six equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the three most commonly occurring violations.  
**Hint:** Use the CASE-WHEN in SQL view to segregate into bins. For finding the most commonly occurring violations, a similar approach can be used as mention in the hint for question 4.  
**Ans.** The data was binned into 0\_3,4\_7,8\_11,12\_15,16\_19,20\_23

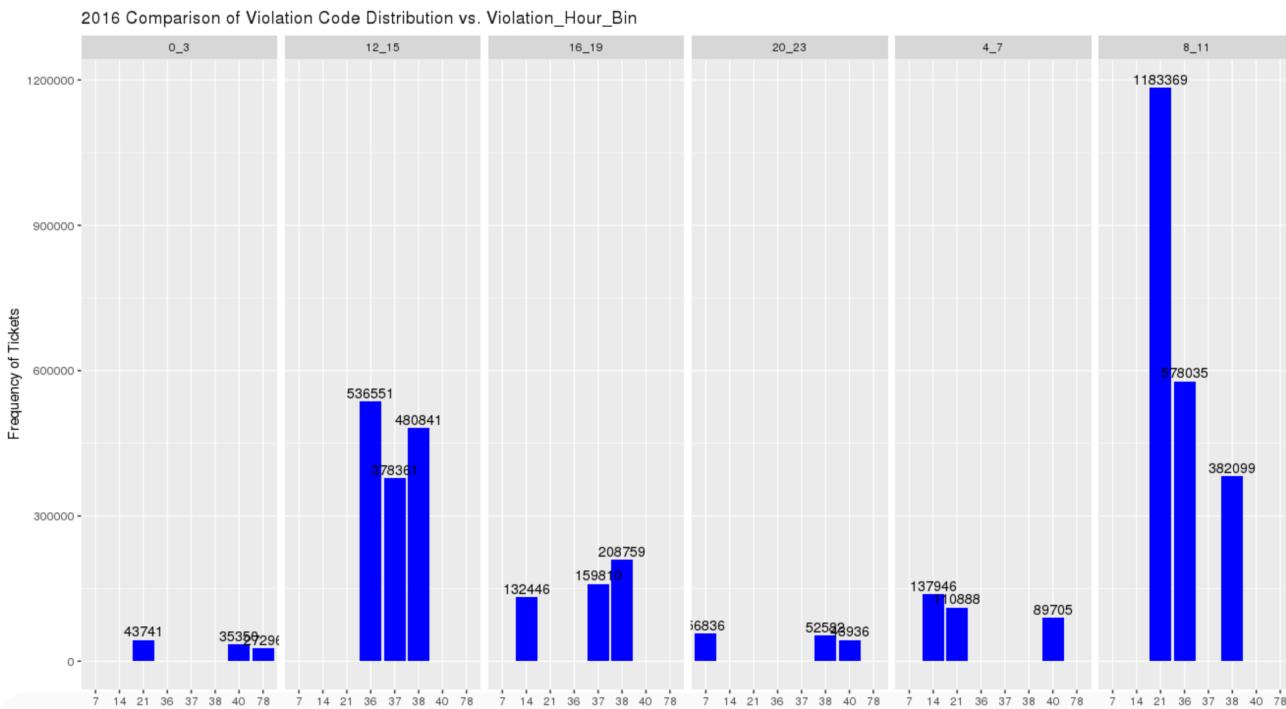
## For fiscal year 2014 - 2015



- 1.Between noon to 15 hrs - There are relatively higher violations seen such as for violation code 36,37,38
- 2.Between 8 to 11 Am - There are relatively highest violations seen for a day.Highest violation code being 21 followed by 38 and 36

## For fiscal year 2015 - 2016

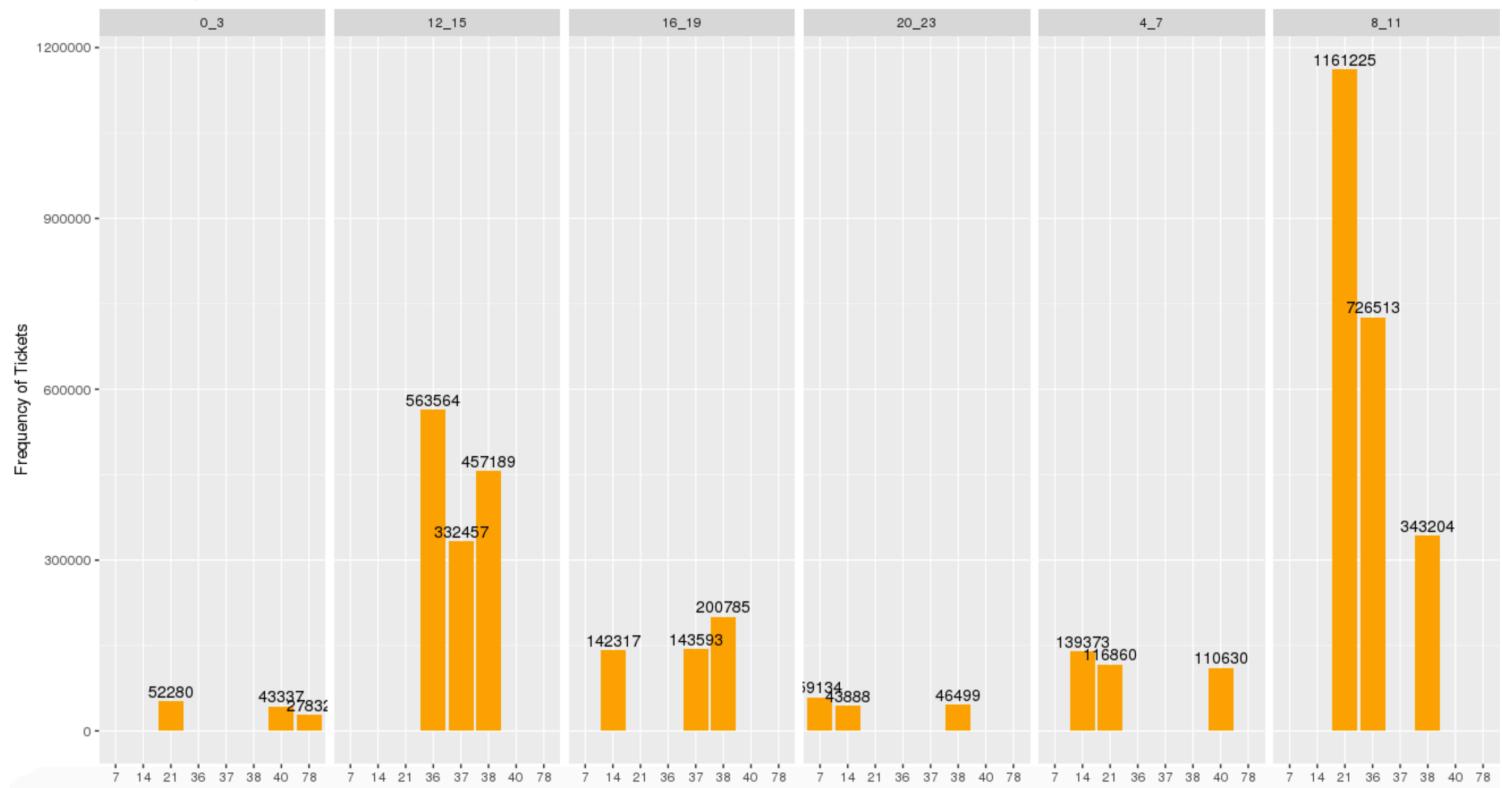
- 1.Between noon to 15 hrs - There are relatively higher violations seen such as for violation code 36,37,38
- 2.Between 8 to 11 Am - There are relatively highest violations seen for a day.Highest violation code being 21 followed by 38 and 36



## For fiscal year 2016 - 2017

- 1.Between noon to 15 hrs - There are relatively higher violations seen such as for violation code 36,37,38
- 2.Between 8 to 11 Am - There are relatively highest violations seen for a day.Highest violation code being 21 followed by 38 and 36

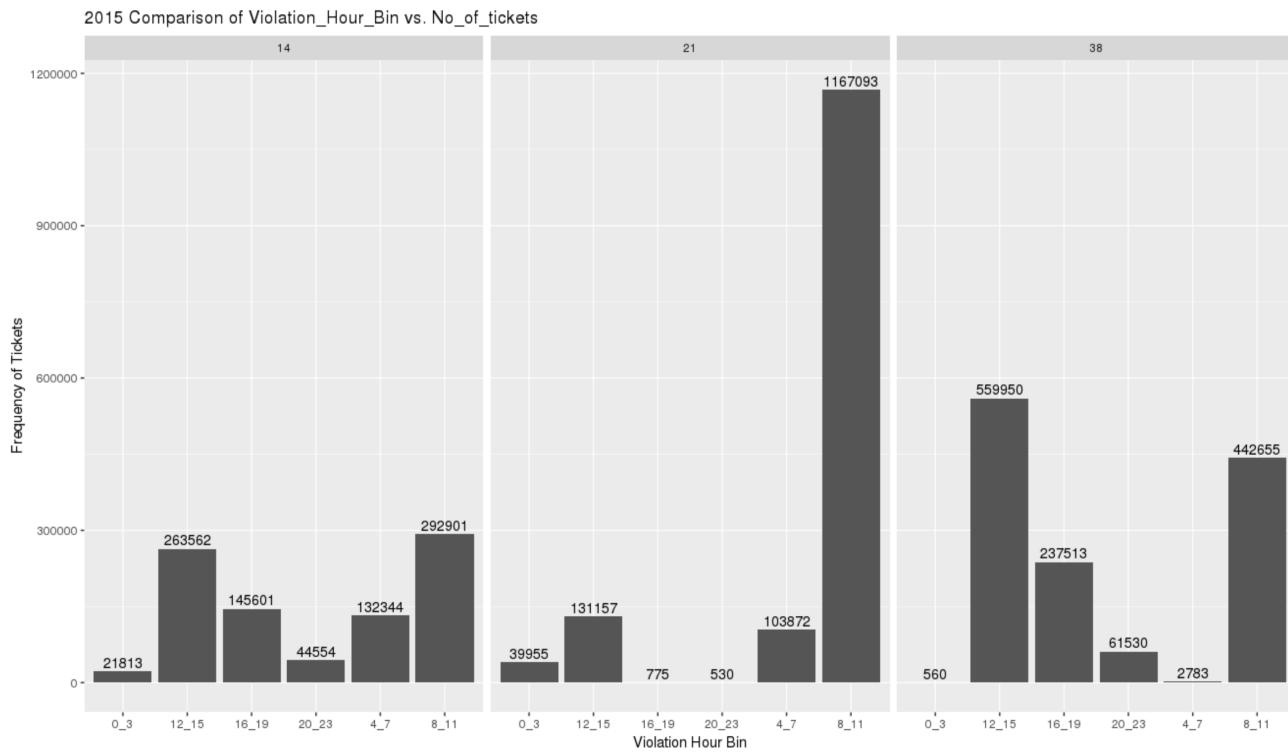
2017 Comparison of Violation Code Distribution vs. Violation\_Hour\_Bin



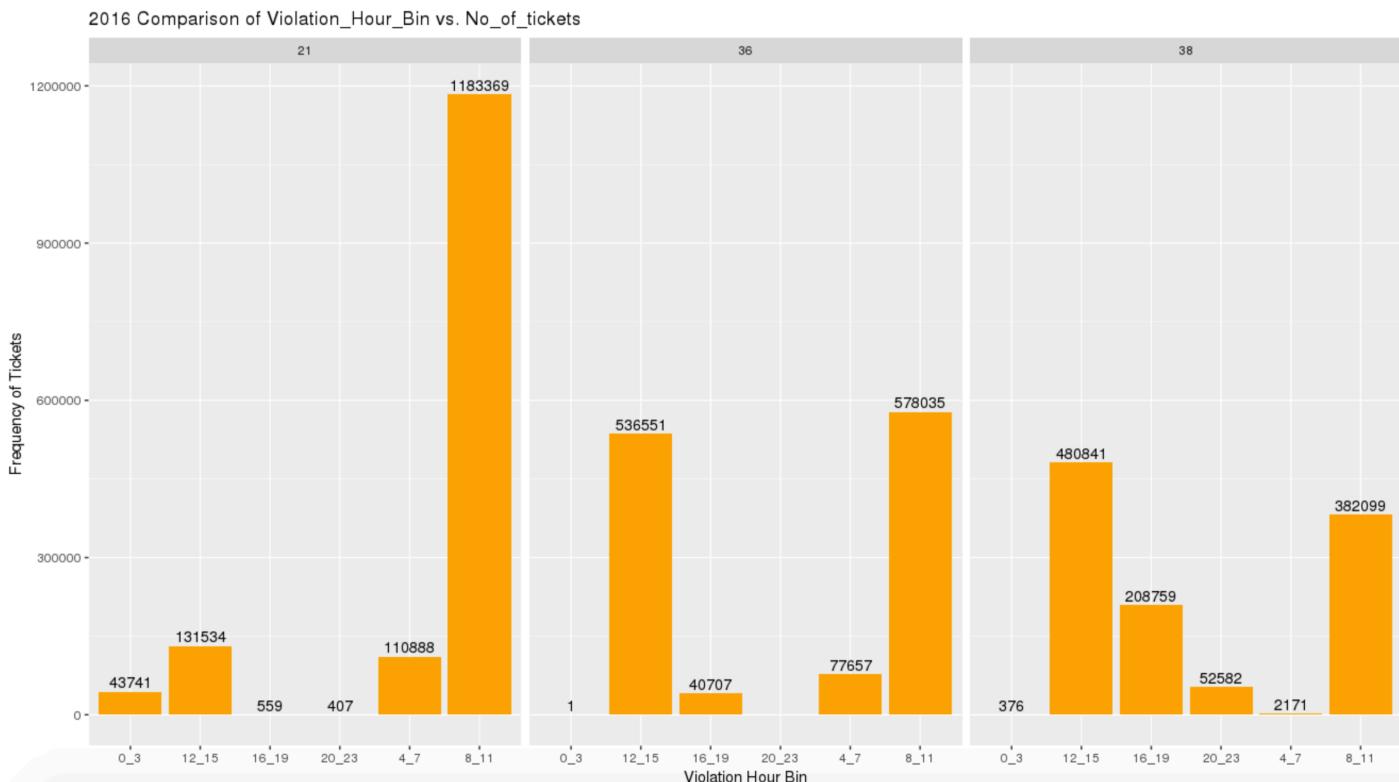
All fiscal years show similar pattern and show similar violations and high number of parking related fines during noon to 15 hr and 8AM -11 AM

4. Now, try another direction. For the 3 most commonly occurring violation codes, find the most common time of the day (in terms of the bins from the previous part)

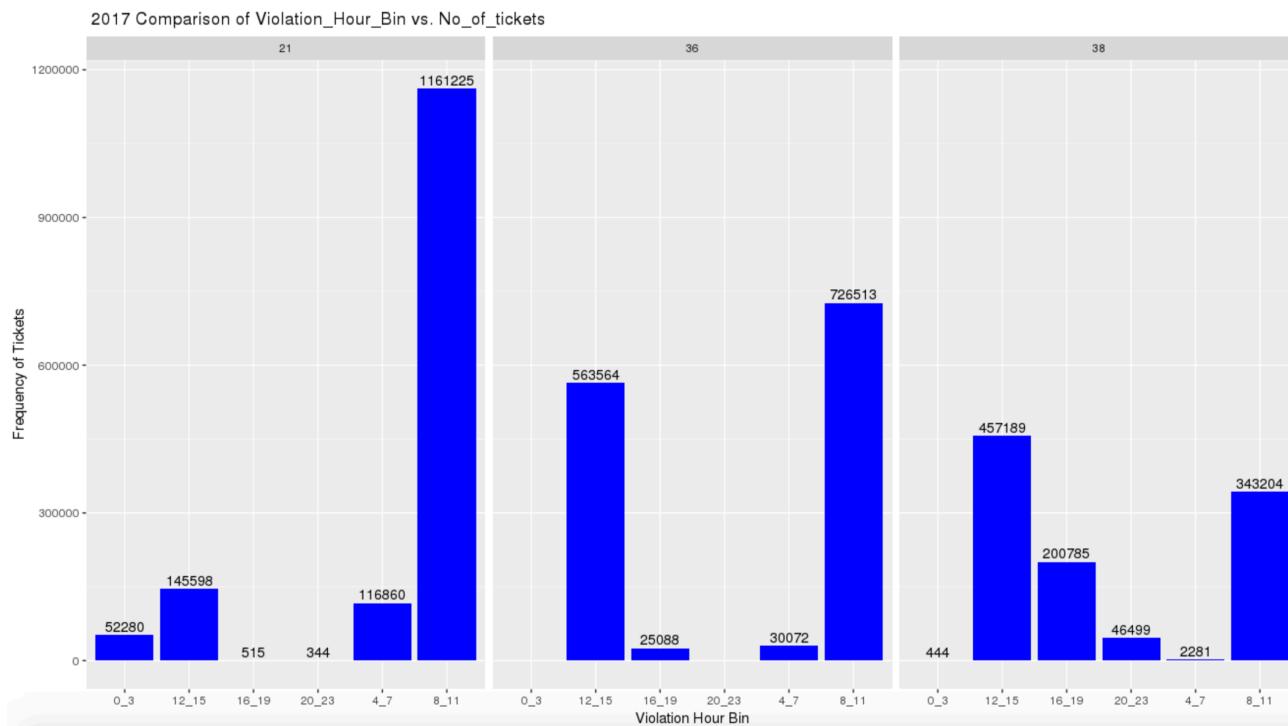
#### For fiscal year 2014 - 2015



#### For fiscal year 2015 - 2016



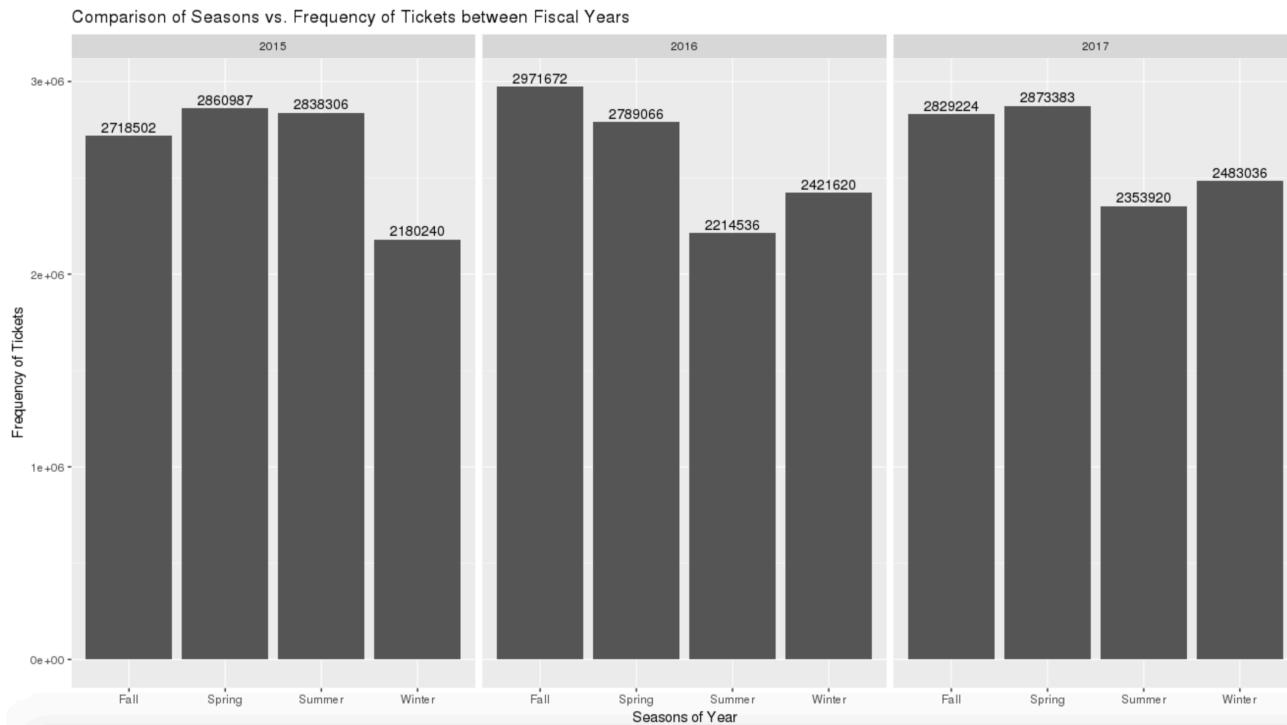
## For fiscal year 2016 - 2017



All fiscal years show similar pattern and show similar violations and high number of parking related fines during noon to 15 hr and 8AM -11 AM

6. Let's try and find some seasonality in this data

1. First, divide the year into some number of seasons, and find frequencies of tickets for each season. (**Hint:** Use Issue Date to segregate into seasons)



#All seasons have shown fines related to parking

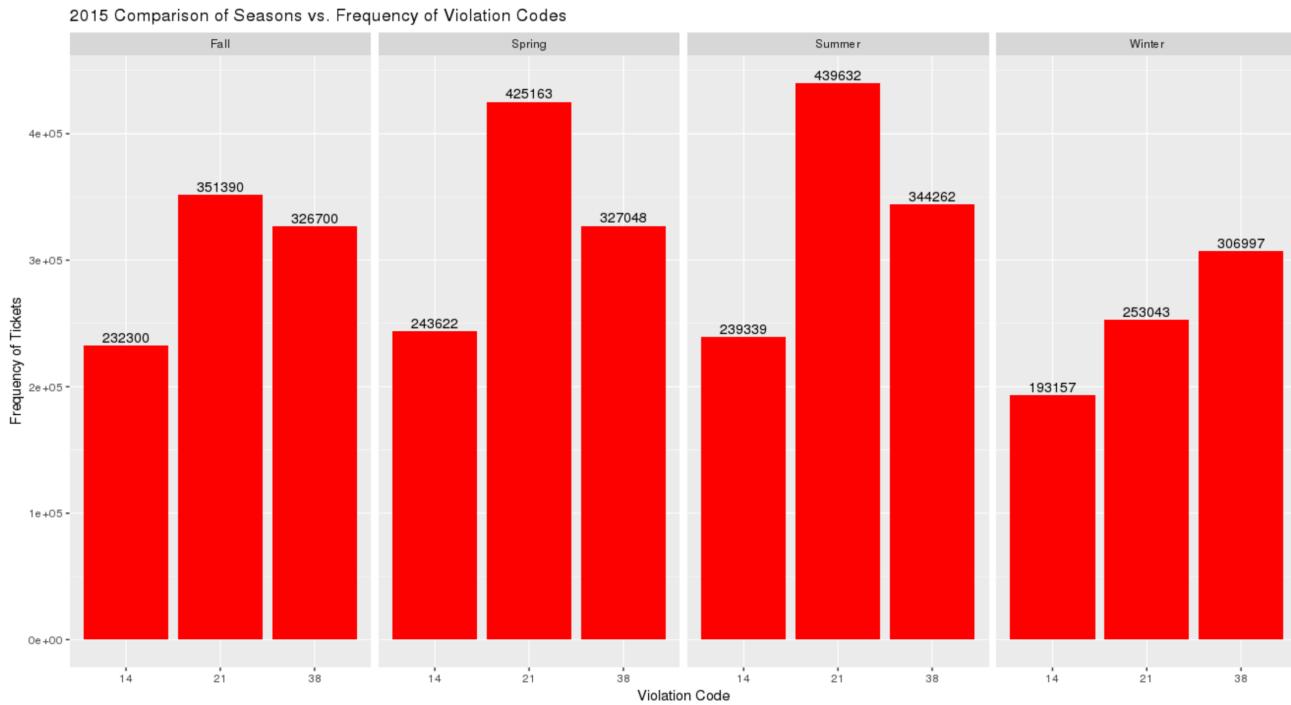
#During the fiscal year 2014-2015 Spring shows the highest number of tickets followed by Summer, Fall and Winter

#During the fiscal year 2015-2016 Fall shows the highest number of tickets followed by Spring, Winter and Summer

#During the fiscal year 2015-2016 Spring shows the highest number of tickets followed by Fall, Winter and Summer

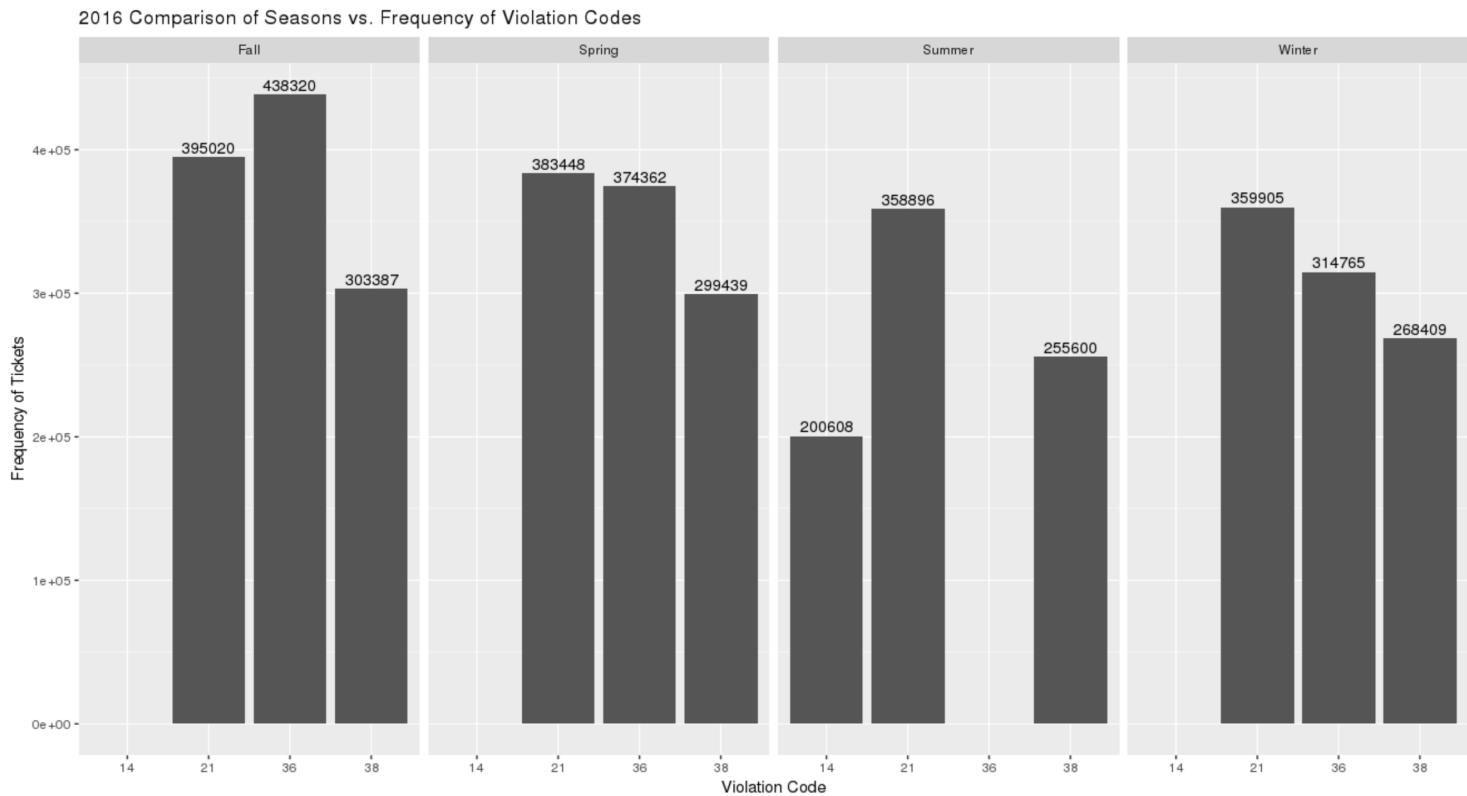
2. Then, find the three most common violations for each of these seasons.  
**(Hint:** A similar approach can be used as mention in the hint for question 4.)

### **Fiscal Year : 2014 - 2015**



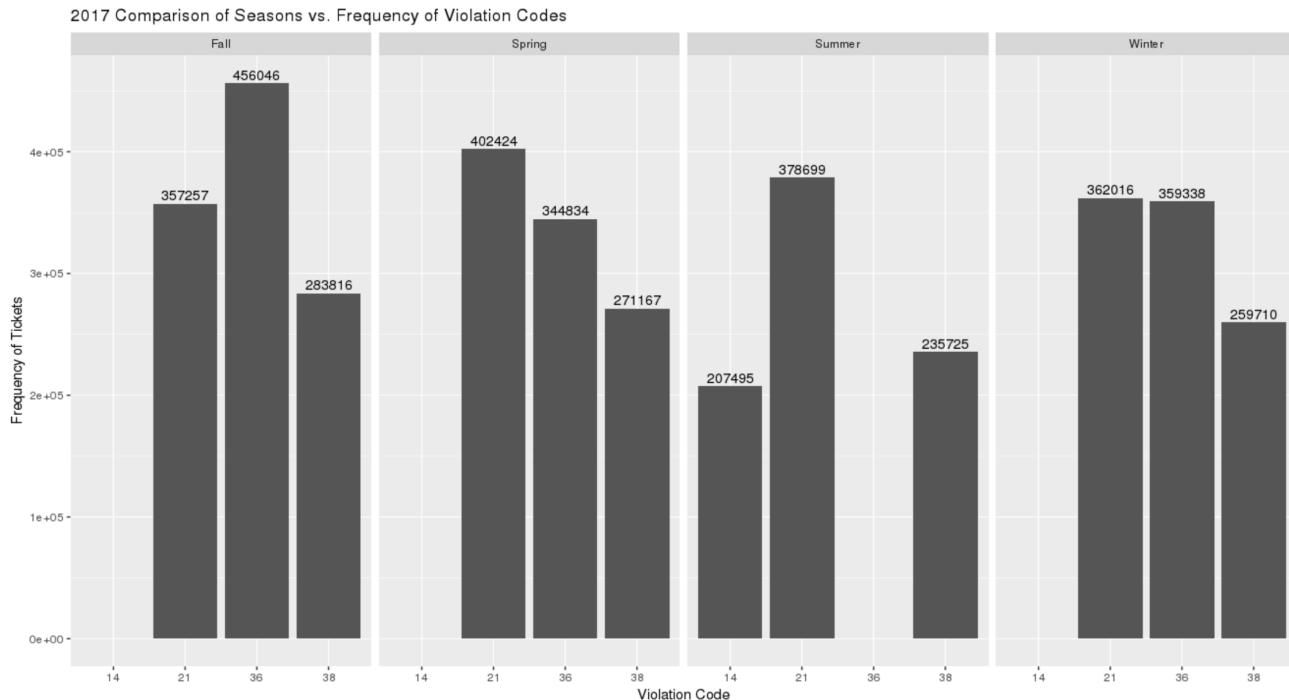
#Violations 14,21 and 28 are common during all seasons  
#Violation 21 is highest in the seasons Fall,Spring and Summer  
#Violation 38 is highest in the seasons Winter

### **Fiscal Year : 2015 - 2016**



#Fall - This fiscal year shows violations such as 21,36 and 38 during Fall  
 #Spring - This fiscal year shows violations such as 21,36 and 38 during Spring  
 #Summer - This fiscal year shows violations such as 14,21 and 38 during Summer  
 #Winter - This fiscal year shows violations such as 21,36 and 38 during Winter

### Fiscal Year : 2016 - 2017



#Fall - This fiscal year shows violations such as 21,36 and 38 during Fall  
 #Spring - This fiscal year shows violations such as 21,36 and 38 during Spring  
 #Summer - This fiscal year shows violations such as 14,21 and 38 during Summer  
 #Winter - This fiscal year shows violations such as 21,36 and 38 during Winter

**Q7.The fines collected from all the parking violation constitute a revenue source for the NYC police department. Let's take an example of estimating that for the three most commonly occurring codes.**

- Find total occurrences of the three most common violation code and respective fines earned by NYPD

Below tables depict the total occurrences and fines earned by departments

#### Fiscal Year : 2014 - 2015

| Violation_Code | Frequency_of_Tickets | Fiscal_Year | Average_Fine_PerTicket | Total_Fine_Amount |
|----------------|----------------------|-------------|------------------------|-------------------|
| 1              | 21                   | 1469228     | 2015                   | 55                |
| 2              | 38                   | 1305007     | 2015                   | 50                |
| 3              | 14                   | 908418      | 2015                   | 115               |

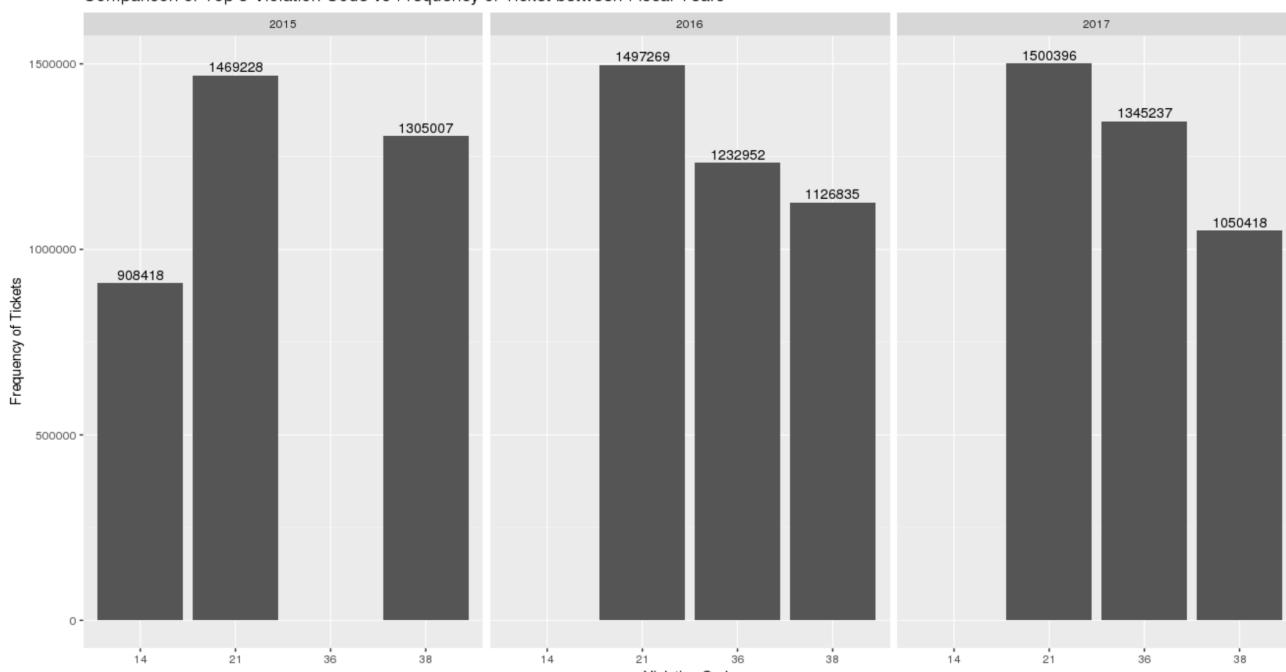
#### Fiscal Year : 2015 - 2016

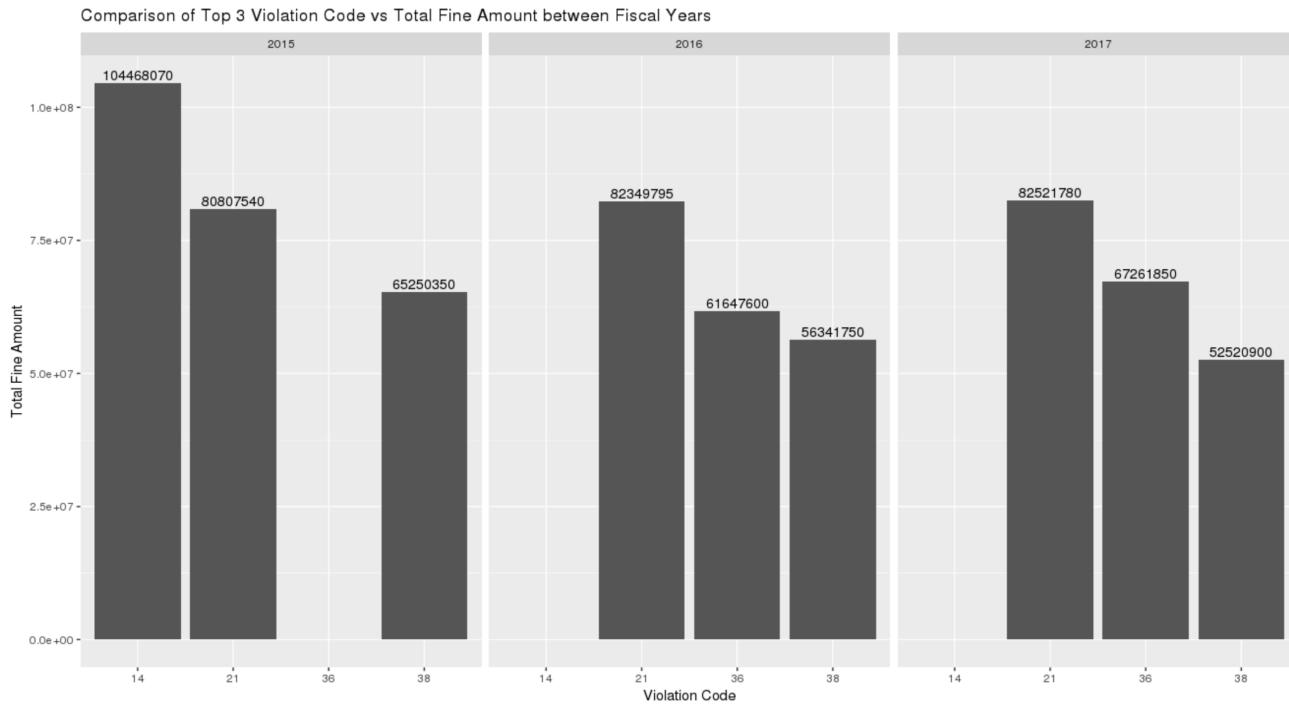
| Violation_Code | Frequency_of_Tickets | Fiscal_Year | Average_Fine_PerTicket | Total_Fine_Amount |
|----------------|----------------------|-------------|------------------------|-------------------|
| 1              | 21                   | 1497269     | 2016                   | 55                |
| 2              | 36                   | 1232952     | 2016                   | 50                |
| 3              | 38                   | 1126835     | 2016                   | 50                |

#### Fiscal Year : 2016 - 2017

| Violation_Code | Frequency_of_Tickets | Fiscal_Year | Average_Fine_PerTicket | Total_Fine_Amount |
|----------------|----------------------|-------------|------------------------|-------------------|
| 1              | 21                   | 1500396     | 2017                   | 55                |
| 2              | 36                   | 1345237     | 2017                   | 50                |
| 3              | 38                   | 1050418     | 2017                   | 50                |

Comparison of Top 3 Violation Code vs Frequency of Ticket between Fiscal Years





## 2. What can you intuitively infer from these findings?

1. **Violation code 14** - i.e *General No Standing: Standing or parking where standing is not allowed by sign, street marking or; traffic control device* was the highest fine collection in the fiscal year 2014 - 2015, However that diminished in the later years, it appears improved discipline in people
2. **Violation code 36** - i.e *Exceeding the posted speed limit in or near a designated school zone*. It seems to be introduced and in the fiscal year 2015-2016 however fiscal year 2014-2015 did not show top collections
3. **Violation code 21** - i.e *Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device*.
4. There is no improvement in the Violation 21,38 in all the fiscal years 2014 - 2015, 2015 - 2016 and 2016 - 2017