- ➢ Analysis (Assumptions, Data Cleaning)
- ➢ Solutions for Questions

## Analysis:

### Assumptions:

- Considered only records having issue date relevant to Fiscal year(July 1 to June 30)
- Interpreted State Code 99 as Invalid

### Data Cleaning:

- Remove duplicate rows in each data set.
- Parsed Date Columns
- Discarded the records with incorrect Fiscal year

Examine the data.

1. *Find total number of tickets for each year.*

   **ASSUMPTION:**
   - Considered "Summons Number" as primary key.
   - Counting Number of Tickets based on number of unique records for column 'Summons Number'
   - Interpreted "Year" in the question as "Fiscal Year"

   | Year | Total Number of Tickets |
   | --- | --- |
   | 2015 | 10558797 |
   | 2016 | 10357238 |
   | 2017 | 10504843 |

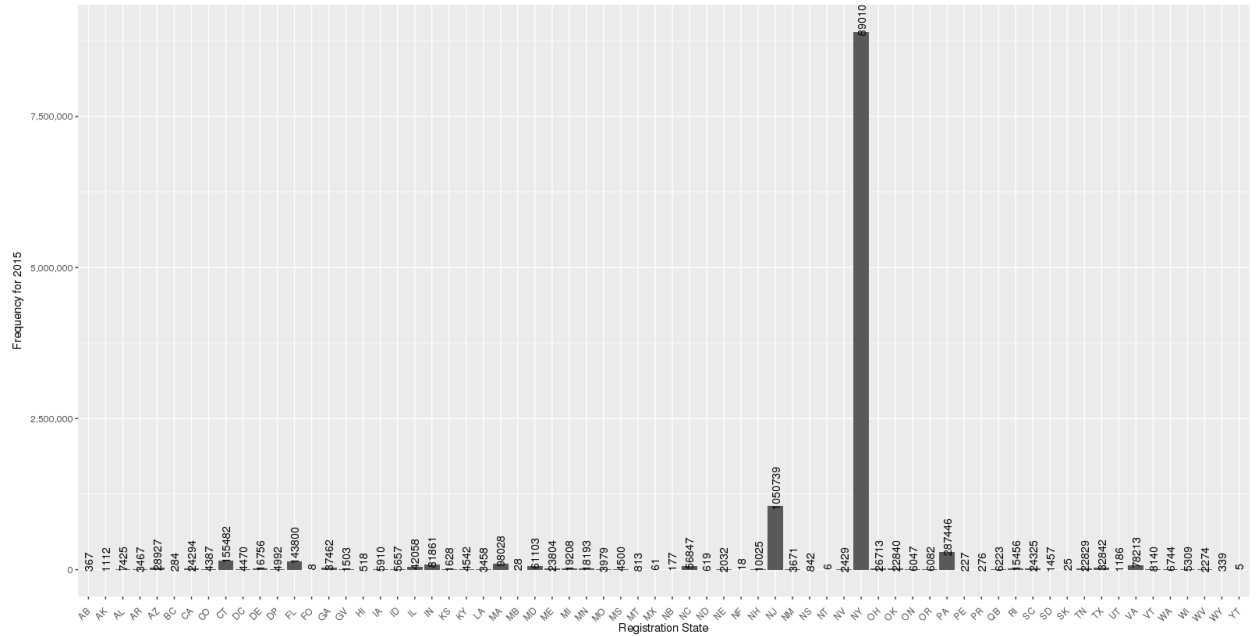   **OBSERVATION:**
   Almost all the 3 years have same total number of tickets. Looks like tickets count has not increased year on year.
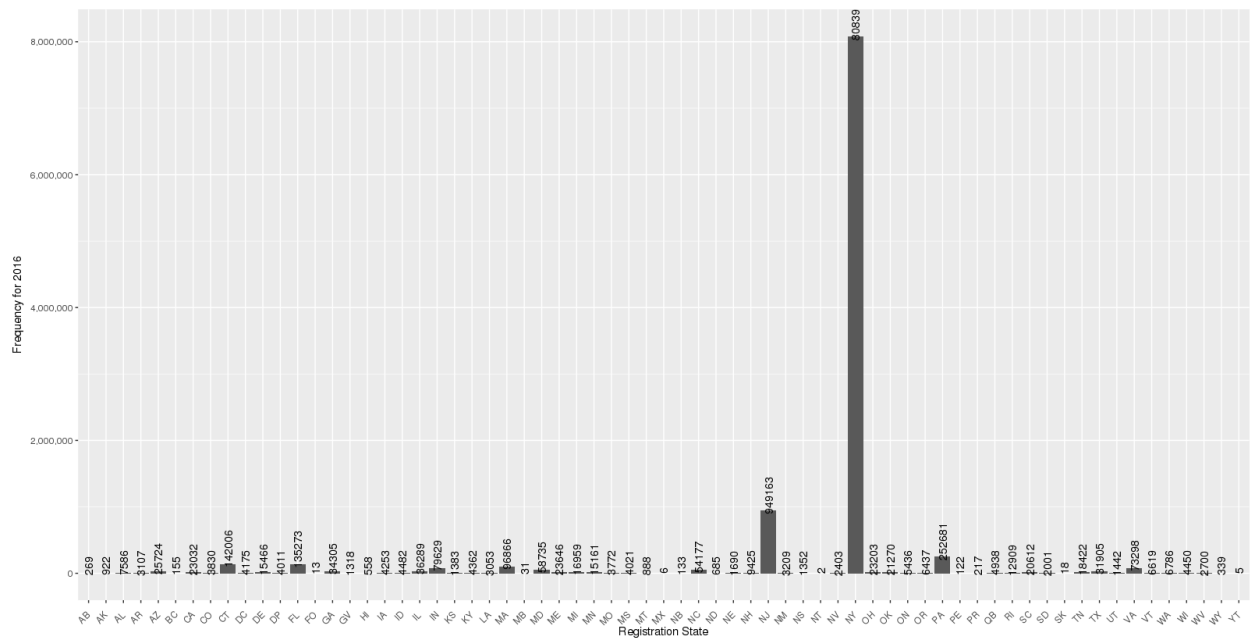
*2. Find out how many unique states the cars which got parking tickets came from.*

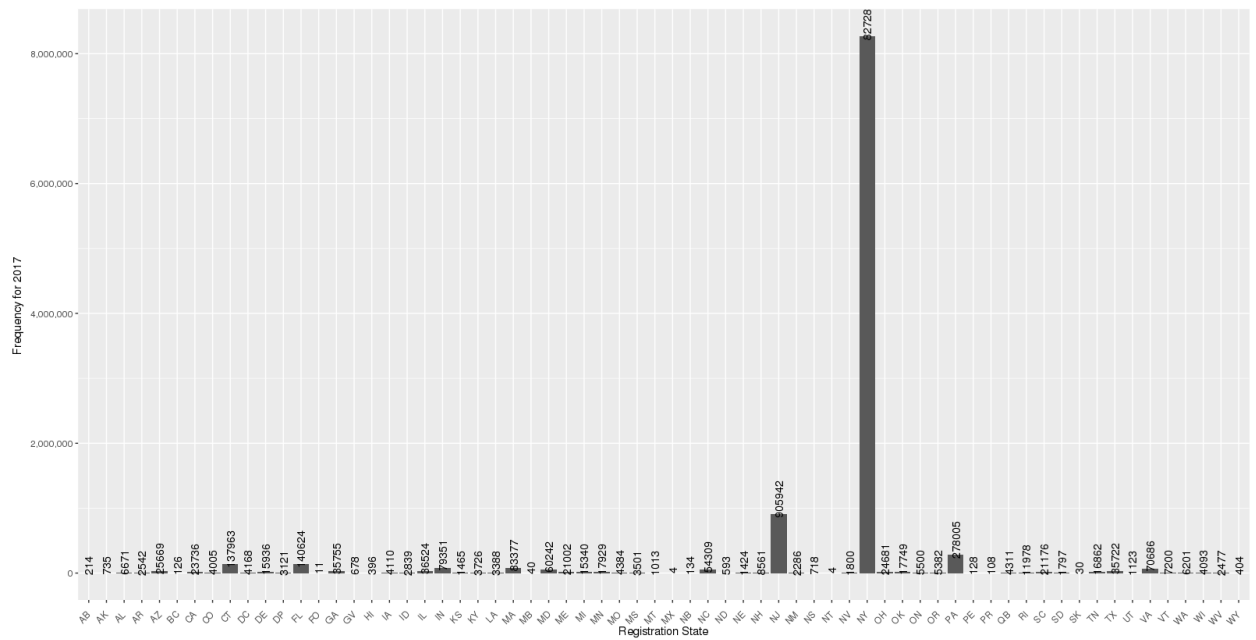| Year | Total Number of Unique States |
|------|-------------------------------|
| 2015 | 68 |
| 2016 | 67 |
| 2017 | 66 |

2015:



2016:

2017:



**OBSERVATION**:
Following three states consistently shown more number of tickets for all the 3years.
- o NY
- o NJ
- o PA

3. *Some parking tickets don't have addresses on them, which is cause for concern. Find out how many such tickets there are.*

### ASSUMPTION:

- We've interpreted "address" in the question as "Violation Location" in the given data set.
- Considered records having "Violation Location" as NULL for analysis

| Year | Total Number of Tickets having Violation Location as NULL |
|---|---|
| 2015 | 1602550 |
| 2016 | 1806516 |
| 2017 | 1949347 |

Aggregation tasks

1. *How often does each violation code occur? (frequency of violation codes - find the top 5)*

################ **2015** ################

#Top 5 violation codes for year 2015

| Violation Code | Frequency of violation codes |
|---|---|
| 21 | 1592519 |
| 38 | 1397978 |
| 14 | 969177 |
| 36 | 824610 |
| 37 | 784747 |

################ **2016** ################

#Top 5 violation codes for year 2016

| Violation Code | Frequency of violation codes |
|---|---|
| 21 | 1490775 |
| 36 | 1232910 |
| 38 | 1125950 |
| 14 | 857521 |
| 37 | 677448 |

################ **2017** ################

#Top 5 violation codes for year 2017

| Violation Code | Frequency of violation codes |
|---|---|
| 21 | 1494775 |
| 36 | 1345192 |
| 38 | 1049457 |
| 14 | 877730 |
| 20 | 607510 |

**OBSERVATION**:
Violation Code "21" identified as top 1 violation code for all the 3years based on frequency.

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

################ **2015** ################

#Top 5 vehicle body types got a parking ticket for year 2015

| Vehicle Body Type | Frequency of parking ticket |
|---|---|
| SUBN | 3604356 |
| 4DSD | 3231197 |
| VAN | 1668301 |
| DELV | 869720 |
| SDN | 486665 |

#Top 5 vehicle make got a parking ticket for year 2015

| Vehicle Make | Frequency of parking ticket |
|---|---|
| FORD | 1471180 |
| TOYOT | 1170476 |
| HONDA | 1061237 |
| NISSA | 872262 |
| CHEVR | 869751 |

################ **2016** ################

#Top 5 vehicle body types got a parking ticket for year 2016

| Vehicle Body Type | Frequency of parking ticket |
|---|---|
| SUBN | 3384675 |
| 4DSD | 2933883 |
| VAN | 1485753 |
| DELV | 734962 |
| SDN | 392262 |

#Top 5 vehicle make got a parking ticket for year 2016

| Vehicle Make | Frequency of parking ticket |
|---|---|
| FORD | 1293656 |
| TOYOT | 1126329 |
| HONDA | 987863 |
| NISSA | 813695 |
| CHEVR | 741641 |

#Top 5 vehicle body types got a parking ticket for year 2017

| Vehicle Body Type | Frequency of parking ticket |
|---|---|
| SUBN | 3624913 |
| 4DSD | 3016847 |
| VAN | 1380288 |
| DELV | 668485 |
| SDN | 405170 |

#Top 5 vehicle make got a parking ticket for year 2017

| Vehicle Make | Frequency of parking ticket |
|---|---|
| FORD | 1247606 |
| TOYOT | 1176814 |
| HONDA | 1049257 |
| NISSA | 893226 |
| CHEVR | 696596 |

**OBSERVATION**:
For the all three years, following Vehicle Body Types got parking ticket.
- SUBN
- 4DSD
- VAN
- DELV
- SDN

For the all three years, following Vehicle Make got parking ticket.
- FORD
- TOYOT
- HONDA
- NISSA
- CHEVR

3. *A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequencies of:*
   3.1. *Violating Precincts (this is the precinct of the zone where the violation occurred)*
   3.2. *Issuing Precincts (this is the precinct that issued the ticket)*

**ASSUMPTION:**
 - Considered Precint '0' as valid Precint

############### **2015** ###############

#5 highest frequencies of Violating Precincts for year 2015

| Violation Precinct | Frequency |
| --- | --- |
| 0 | 1602550 |
| 19 | 588293 |
| 18 | 419173 |
| 14 | 400926 |
| 1 | 322696 |

#5 highest frequencies of Issuing Precincts for year 2015

| Issuer Precinct | Frequency |
| --- | --- |
| 0 | 1828089 |
| 19 | 570884 |
| 18 | 409740 |
| 14 | 386171 |
| 1 | 313469 |

############### **2016** ###############

#5 highest frequencies of Violating Precincts for year 2016

| Violation Precinct | Frequency |
| --- | --- |
| 0 | 1806515 |
| 19 | 544847 |
| 18 | 324632 |
| 14 | 316995 |
| 1 | 298075 |

#5 highest frequencies of Issuing Precincts for year 2016

| Issuer Precinct | Frequency |
|---|---|
| 0 | 2061872 |
| 19 | 531651 |
| 18 | 316730 |
| 14 | 308893 |
| 1 | 289897 |

################ **2017** ################

#5 highest frequencies of Violating Precincts for year 2017

| Violation Precinct | Frequency |
|---|---|
| 0 | 1949347 |
| 19 | 527366 |
| 18 | 346580 |
| 14 | 325871 |
| 1 | 301093 |

#5 highest frequencies of Issuing Precincts for year 2017

| Issuer Precinct | Frequency |
|---|---|
| 0 | 2248357 |
| 19 | 514061 |
| 14 | 340003 |
| 1 | 316090 |
| 18 | 291563 |

OBSERVATION:
For the all three years, following Precincts have highest frequencies of Violating Precincts & Issuing Precincts.
- 0
- 19
- 14
- 18
- 1

4. *Find the violation code frequency across 3 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?*

**ASSUMPTION**:
- Considered Top 5 violation codes for comparison across precincts.

################ **2015** ################

#Top 5 violation codes across 3 precincts which have issued the most number of tickets for year 2015

| Violation Code | Frequency |
|---|---|
| 36 | 824610 |
| 7 | 619543 |
| 21 | 263242 |
| 14 | 195009 |
| 5 | 144352 |

################ **2016** ################

#Top 5 violation codes across 3 precincts which have issued the most number of tickets for year 2016

| Violation Code | Frequency |
|---|---|
| 36 | 1232909 |
| 7 | 457855 |
| 21 | 283745 |
| 14 | 164411 |
| 5 | 106607 |

################ **2017** ################

#Top 5 violation codes across 3 precincts which have issued the most number of tickets for year 2017

| Violation Code | Frequency |
|---|---|
| 36 | 1345192 |
| 7 | 464676 |
| 21 | 310394 |
| 14 | 136277 |
| 5 | 130958 |

**OBSERVATION**:
For the all three years, following violation codes have highest frequencies of Parking Violations across 3 precincts, which have issued the most number of tickets.

- 36
- 7
- 21
- 14
- 5

5. **You'd want to find out the properties of parking violations across different times of the day:**
   5.1. *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.*
   5.2. *Find a way to deal with missing values, if any.*
   5.3. *Divide 24 hours into 6 equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the 3 most commonly occurring violations*
   5.4. *Now, try another direction. For the 3 most commonly occurring violation codes, find the most common times of day (in terms of the bins from the previous part)*

ASSUMPTION:
- Ignored Missing Values, as the percentage of missing values compared to overall data set is very less.
- Considered Hour bins as below
  - Bin 1 -- 0 to 3 hrs
  - Bin 2 -- 4 to 7 hrs
  - Bin 3 -- 8 to 11 hrs
  - Bin 4 -- 12 to 15 hrs
  - Bin 5 -- 16 to 19 hrs
  - Bin 6 -- 20 to 23 hrs

################ **2015** ################

#3 most commonly occurring violations for each discrete bins of time for year 2015

| Hour Bin Number | Violation Code | Total Tickets |
| --- | --- | --- |
| 1 | 21 | 67638 |
| 1 | 40 | 38350 |
| 1 | 78 | 35674 |
| | | |
| 2 | 40 | 23190 |
| 2 | 14 | 18095 |
| 2 | 7 | 14933 |
| | | |
| 3 | 21 | 1372141 |
| 3 | 38 | 476141 |
| 3 | 36 | 450539 |
| | | |
| 4 | 38 | 600861 |
| 4 | 37 | 440471 |
| 4 | 36 | 350967 |
| | | |
| 5 | 38 | 254773 |
| 5 | 37 | 184227 |
| 5 | 14 | 156727 |
| | | |

| | | |
|---|---|---|
| 6 | 7 | 76972 |
| 6 | 38 | 65109 |
| 6 | 40 | 48087 |

#Most common times of day, when 3 most commonly violation codes are occurring for year 2015

| Hour Bin Number | Total Tickets |
|---|---|
| 3 | 2282575 |
| 4 | 1031802 |
| 5 | 412328 |
| 6 | 113602 |
| 1 | 96052 |
| 2 | 23266 |

################  **2016**  ################

#3 most commonly occurring violations for each discrete bins of time for year 2016

| Hour Bin Number | Violation Code | Total Tickets |
|---|---|---|
| 1 | 21 | 65759 |
| 1 | 40 | 35611 |
| 1 | 78 | 28096 |
| | | |
| 2 | 40 | 20961 |
| 2 | 14 | 18901 |
| 2 | 20 | 11308 |
| | | |
| 3 | 21 | 1284358 |
| 3 | 36 | 655669 |
| 3 | 14 | 389865 |
| | | |
| 4 | 36 | 536532 |
| 4 | 38 | 480483 |
| 4 | 37 | 378180 |
| | | |
| 5 | 38 | 208553 |
| 5 | 37 | 159722 |
| 5 | 14 | 131900 |
| | | |
| 6 | 7 | 56835 |
| 6 | 38 | 52566 |
| 6 | 40 | 43565 |

#Most common times of day, when 3 most commonly violation codes are occurring for year 2016

| Hour Bin Number | Total Tickets |
|---|---|
| 3 | 2323764 |
| 4 | 1152077 |
| 5 | 249800 |
| 1 | 66076 |
| 6 | 52964 |
| 2 | 4907 |

################  2017  ################

#3 most commonly occurring violations for each discrete bins of time for year 2017

| Hour Bin Number | Violation Code | Total Tickets |
|---|---|---|
| 1 | 21 | 71306 |
| 1 | 40 | 44161 |
| 1 | 14 | 28470 |
|  |  |  |
| 2 | 40 | 25365 |
| 2 | 14 | 15176 |
| 2 | 20 | 13379 |
|  |  |  |
| 3 | 21 | 1268233 |
| 3 | 36 | 756556 |
| 3 | 14 | 393574 |
|  |  |  |
| 4 | 36 | 563548 |
| 4 | 38 | 456919 |
| 4 | 37 | 332335 |
|  |  |  |
| 5 | 38 | 200613 |
| 5 | 37 | 143521 |
| 5 | 14 | 141816 |
|  |  |  |
| 6 | 7 | 59133 |
| 6 | 38 | 46473 |
| 6 | 40 | 43635 |

#Most common times of day, when 3 most commonly violation codes are occurring for year 2017

| Hour Bin Number | Total Tickets |
|---|---|

| 3 | 2369651 |
| 4 | 1169562 |
| 5 | 226193 |
| 1 | 71660 |
| 6 | 46812 |
| 2 | 5505 |

**OBSERVATION**:
For the all three years, majorly top 3 violations occurred in between 8 AM to 11 AM (bin 3), followed by in between 12 PM to 3 PM (bin 4).

6. *Let's try and find some seasonality in this data*
    6.1. *First, divide the year into some number of seasons, and find frequencies of tickets for each season.*
    6.2. *Then, find the 3 most common violations for each of these season*

## ASSUMPTION:
- Considered Seasons as below
  - Season 1 -- January to March
  - Season 2 -- April to June
  - Season 3 -- July to September
  - Season 4 -- October to December

############### **2015** ###############

# Frequencies of tickets for each season for year 2015

| Season Bin Number | Total Tickets |
|---|---|
| 1 | 2475936 |
| 2 | 3250418 |
| 3 | 2789425 |
| 4 | 2435478 |

#3 Most common violations for each season for Year 2016

| Season Bin Number | Violation Code | Total Tickets |
|---|---|---|
| 1 | 38 | 336762 |
| 1 | 21 | 281600 |
| 1 | 14 | 220029 |
| | | |
| 2 | 21 | 471580 |
| 2 | 38 | 346719 |
| 2 | 14 | 262595 |
| | | |
| 3 | 21 | 397871 |
| 3 | 38 | 348466 |
| 3 | 14 | 234606 |
| | | |
| 4 | 21 | 350563 |
| 4 | 38 | 292639 |
| 4 | 14 | 207397 |

############### **2016** ###############

# Frequencies of tickets for each season for year 2016

| Season Bin Number | Total Tickets |
|---|---|
| 1 | 2671331 |
| 2 | 2425877 |
| 3 | 2728663 |
| 4 | 2801028 |

#3 Most common violations for each season for Year 2016

| Season Bin Number | Violation Code | Total Tickets |
|---|---|---|
| 1 | 21 | 349644 |
| 1 | 36 | 341787 |
| 1 | 38 | 308999 |
| | | |
| 2 | 21 | 348473 |
| 2 | 36 | 294015 |
| 2 | 38 | 254909 |
| | | |
| 3 | 21 | 403720 |
| 3 | 38 | 305360 |
| 3 | 14 | 234943 |
| | | |
| 4 | 36 | 433966 |
| 4 | 21 | 429750 |
| 4 | 28 | 274428 |

################ **2017** ################

# Frequencies of tickets for each season for year 2017

| Season Bin Number | Total Tickets |
|---|---|
| 1 | 2671332 |
| 2 | 3018840 |
| 3 | 2463936 |
| 4 | 2648920 |

#3 Most common violations for each season for Year 2017

| Season Bin Number | Violation Code | Total Tickets |
|---|---|---|
| 1 | 21 | 374202 |
| 1 | 36 | 348240 |

| | | |
|---|---|---|
| 1 | 38 | 287017 |
| | | |
| 2 | 21 | 421184 |
| 2 | 36 | 369902 |
| 2 | 38 | 266909 |
| | | |
| 3 | 21 | 385774 |
| 3 | 38 | 244985 |
| 3 | 36 | 239879 |
| | | |
| 4 | 36 | 442593 |
| 4 | 21 | 347428 |
| 4 | 38 | 263393 |

**OBSERVATION**:

For the all three years, majorly following violations occurred highly compared to others in all seasons.

- 21
- 38

7. *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 3 most commonly occurring codes.*
   7.1. *Find total occurrences of the 3 most common violation codes*
   7.2. *Then, search the internet for NYC parking violation code fines. You will find a website (on the nyc.gov URL) that lists these fines. They're divided into two categories, one for the highest-density locations of the city, the other for the rest of the city. For simplicity, take an average of the two.*
   7.3. *Using this information, find the total amount collected for all of the fines. State the code which has the highest total collection.*
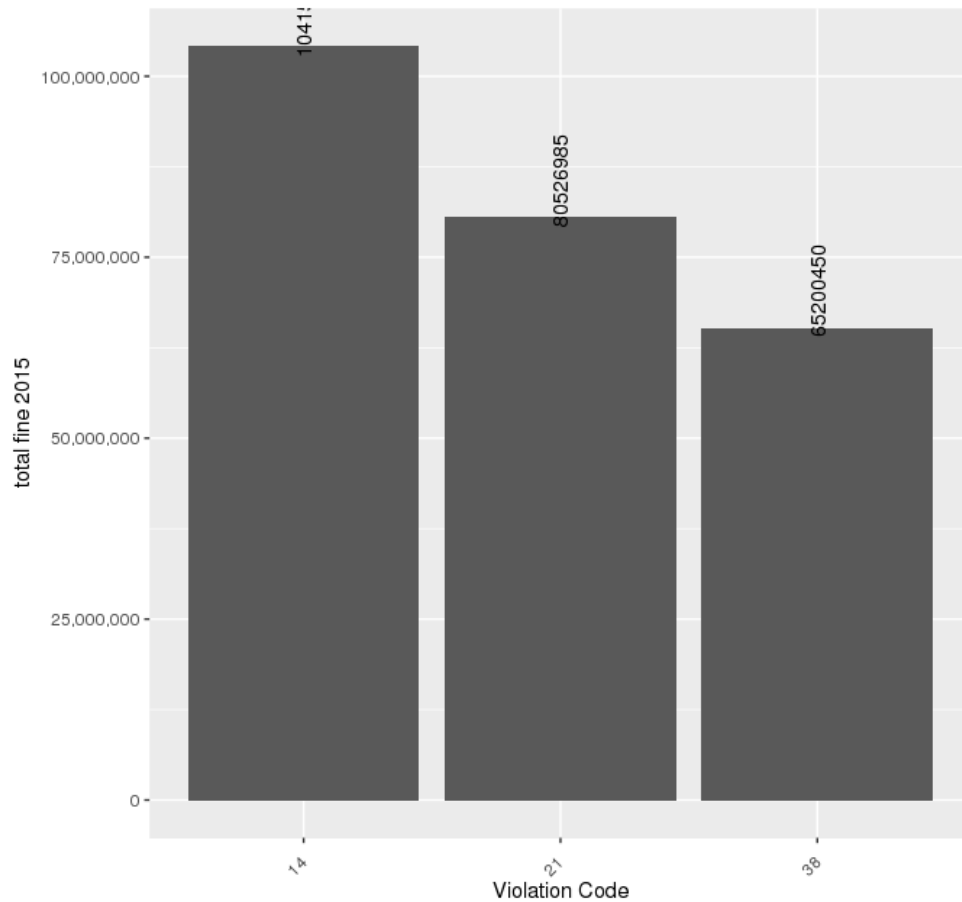   7.4. *What can you intuitively infer from these findings?*

################ **2015** ################

#Top 3 violation codes for year 2015 based on count

| Violation Code | Frequency |
|----------------|-----------|
| 21 | 1592519 |
| 38 | 1397978 |
| 14 | 969177 |

#Top 3 violation codes for year 2015 based on total fine collection

| Violation Code | Total Fine |
|----------------|------------|
| 14 | 111455355 |
| 21 | 87588545 |
| 38 | 69898900 |

Bar chart showing total fine 2015 by Violation Code (14: 1041!, 21: 80526985, 38: 65200450)

#Top 3 violation codes for year 2016 based on count

| Violation Code | Frequency |
|:---:|:---:|
| 21 | 1490775 |
| 36 | 1232910 |
| 38 | 1125950 |

#Top 3 violation codes for year 2016 based on total fine collection

| Violation Code | Total Fine |
|:---:|:---:|
| 14 | 98614915 |
| 21 | 81992625 |
| 36 | 61645500 |
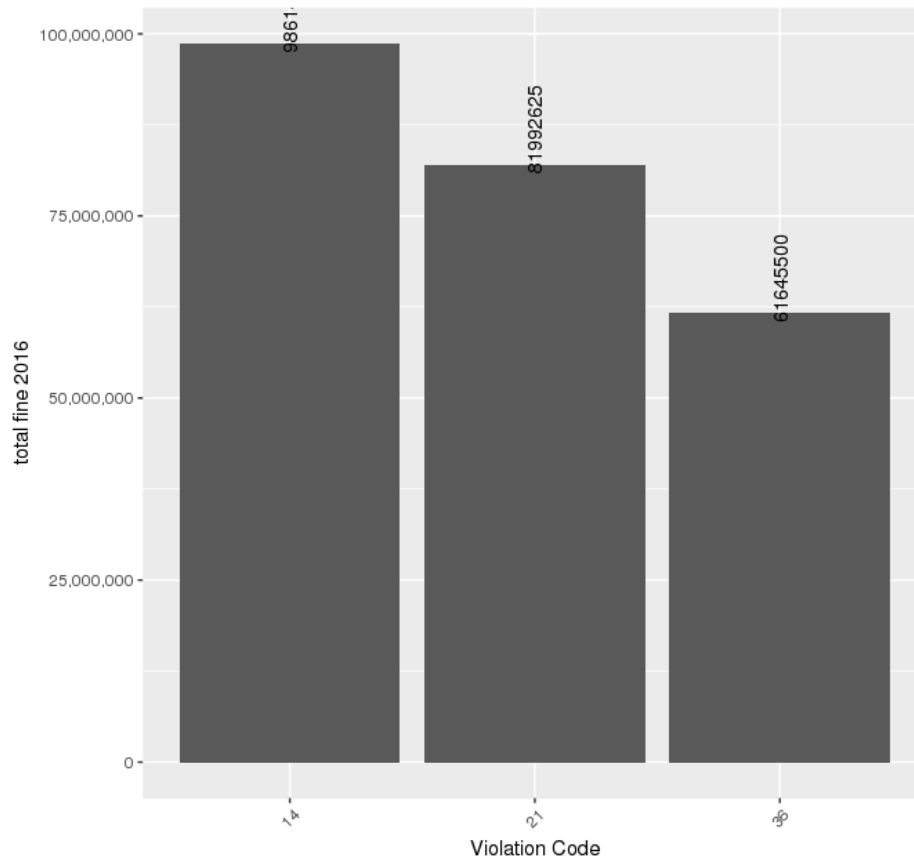
#Top 3 violation codes for year 2017 based on count

| Violation Code | Frequency |
|:---:|:---:|
| 21 | 1494775 |
| 36 | 1345192 |
| 38 | 1049457 |

#Top 3 violation codes for year 2017 based on total fine collection

| Violation Code | Total Fine |
|:---:|:---:|
| 14 | 100938950 |
| 21 | 82212625 |
| 36 | 67259600 |
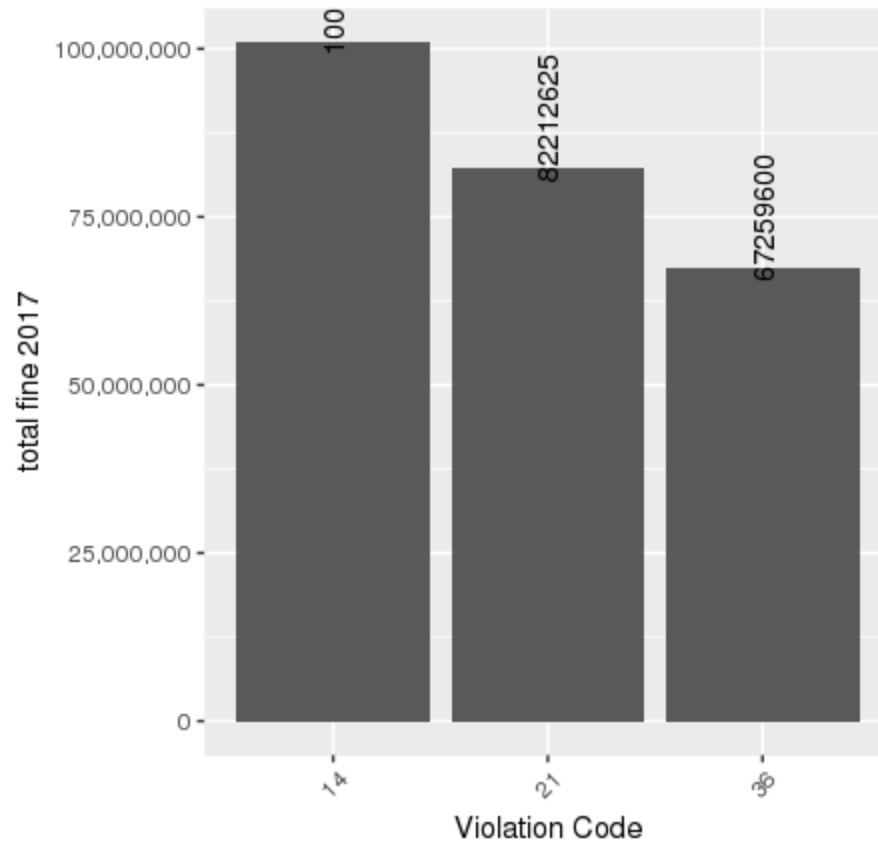
**OBSERVATION**:

- For the all three years, Violation code "14" has the highest total collection.
- Followed by Violation code "14", following violation codes have highest total collection.
  - 21
  - 36