Examine the data:

1. Find the total number of tickets for each year. (we are assuming fiscal year)

2015	10,598,035
2016	10,396,894
2017	10,539,563

2. Find out the number of unique states from where the cars that got parking tickets came from. *Give the number of unique states for each year again.*

2015	69
2016	68
2017	67

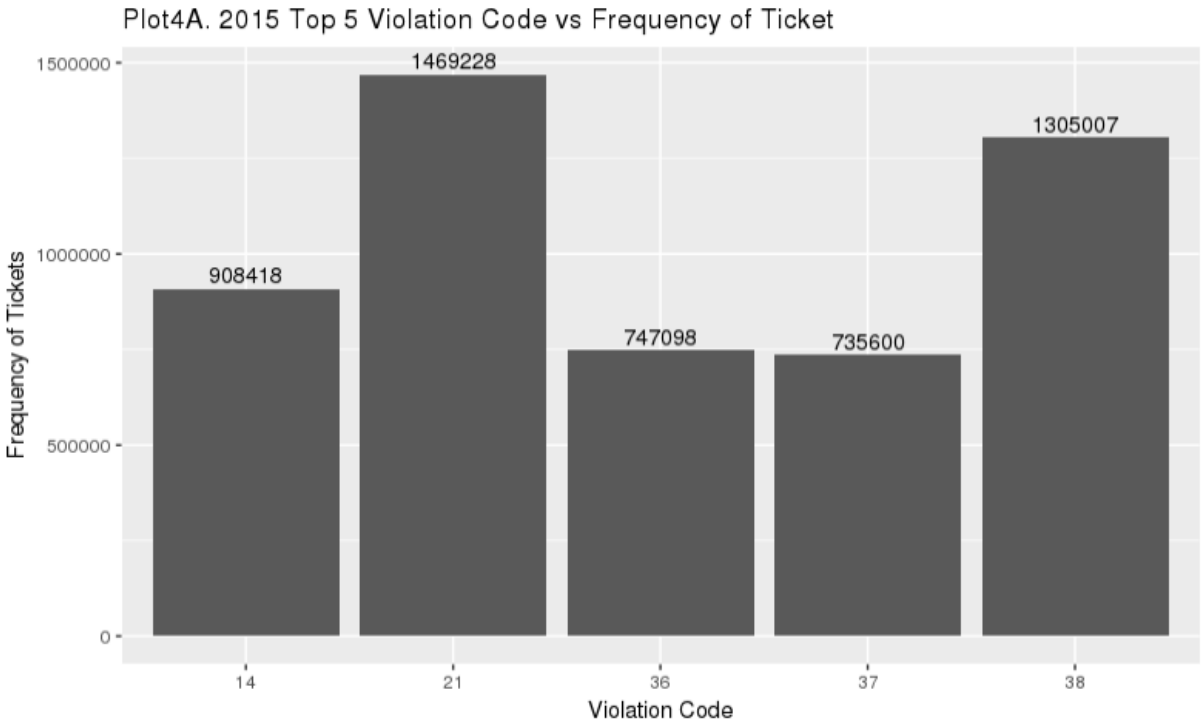
3. Some parking tickets don't have the address for violation location on them, which is a cause for concern. Write a query to check the number of such tickets.

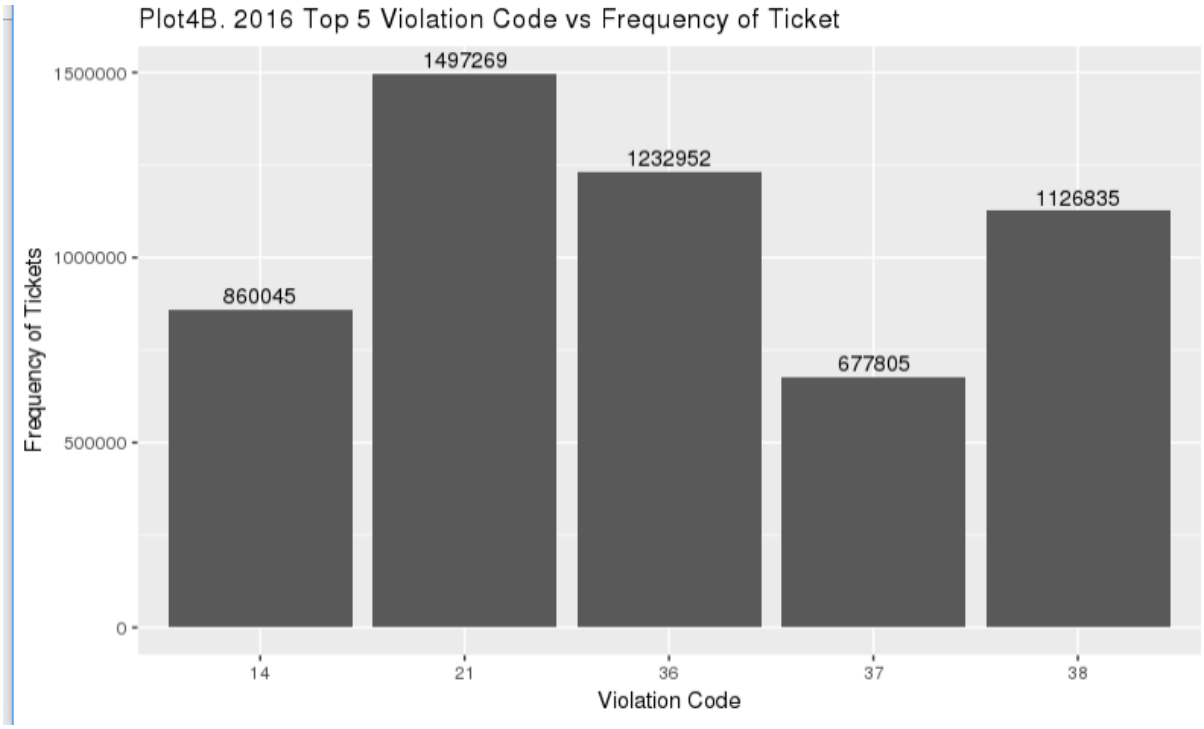
2015	1,622,076
2016	1,963,921
2017	2,160,639

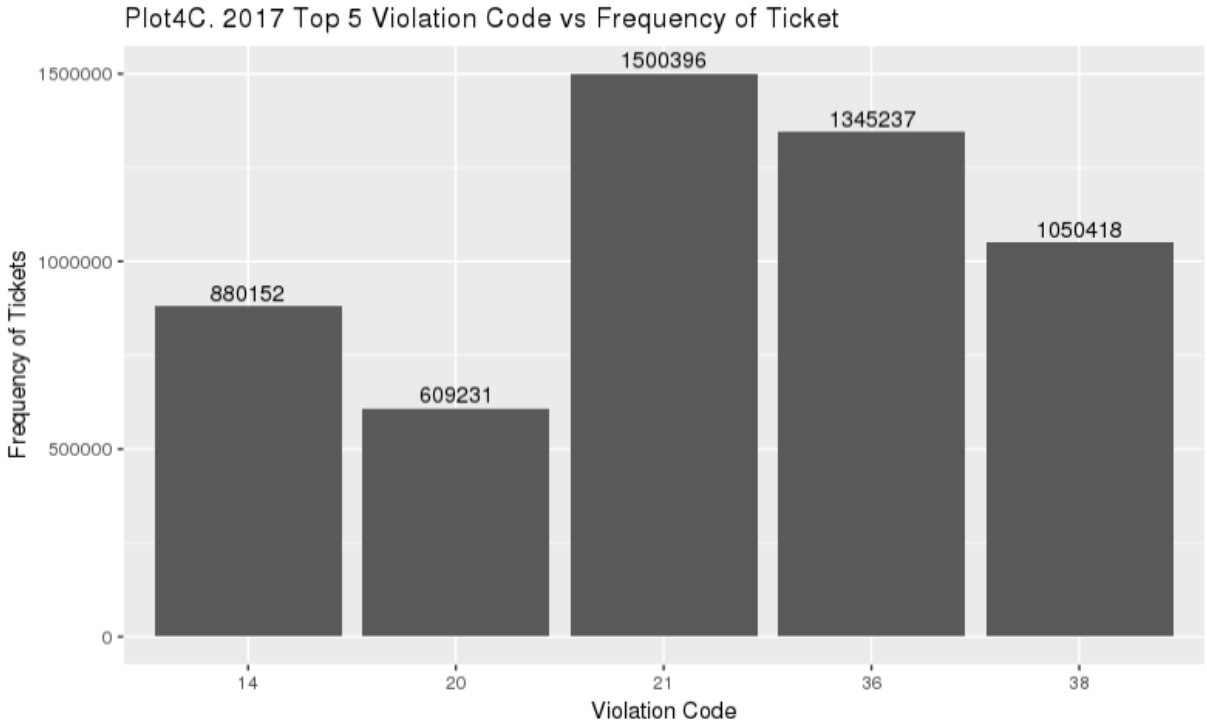
Aggregation tasks

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

Violation code	***Frequency of tickets for 2015***	***Frequency of tickets for 2016***	***Frequency of tickets for 2017***
21	1469228	1497269	1500396
38	1305007	1126835	1050418
14	908418	860045	880152
36	747098	1232952	1345237
37	735600	677805	N/A
20	N/A	N/A	609231


Plot4A. 2015 Top 5 Violation Code vs Frequency of Ticket
Frequency of Tickets
1500000
1000000
500000
0
908418
1469228
747098
735600
1305007
14
21
36
37
38
Violation Code

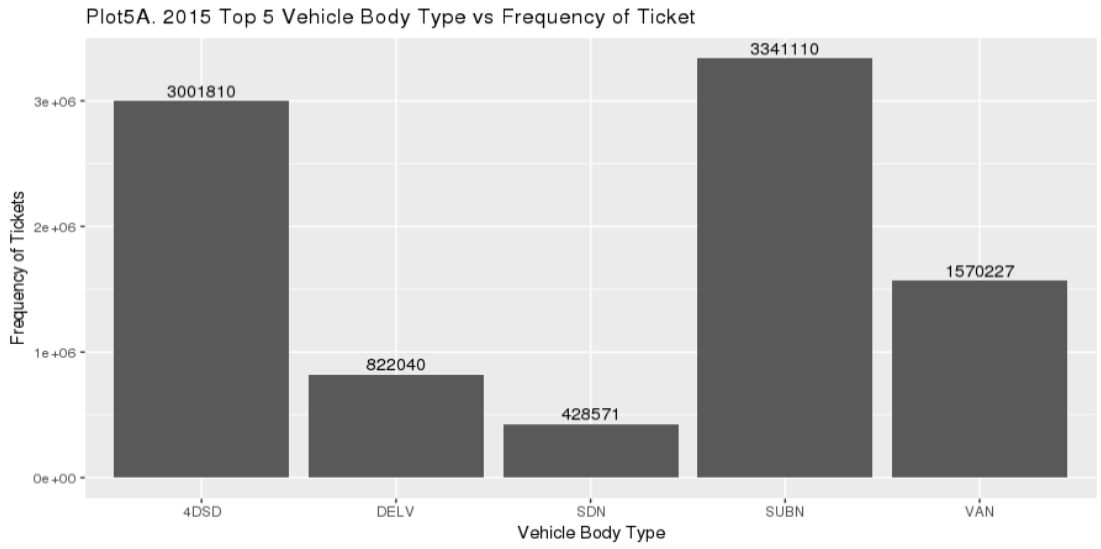

Plot4B. 2016 Top 5 Violation Code vs Frequency of Ticket
Frequency of Tickets
1500000
1000000
500000
0
860045
1497269
1232952
677805
1126835
14
21
36
37
38
Violation Code

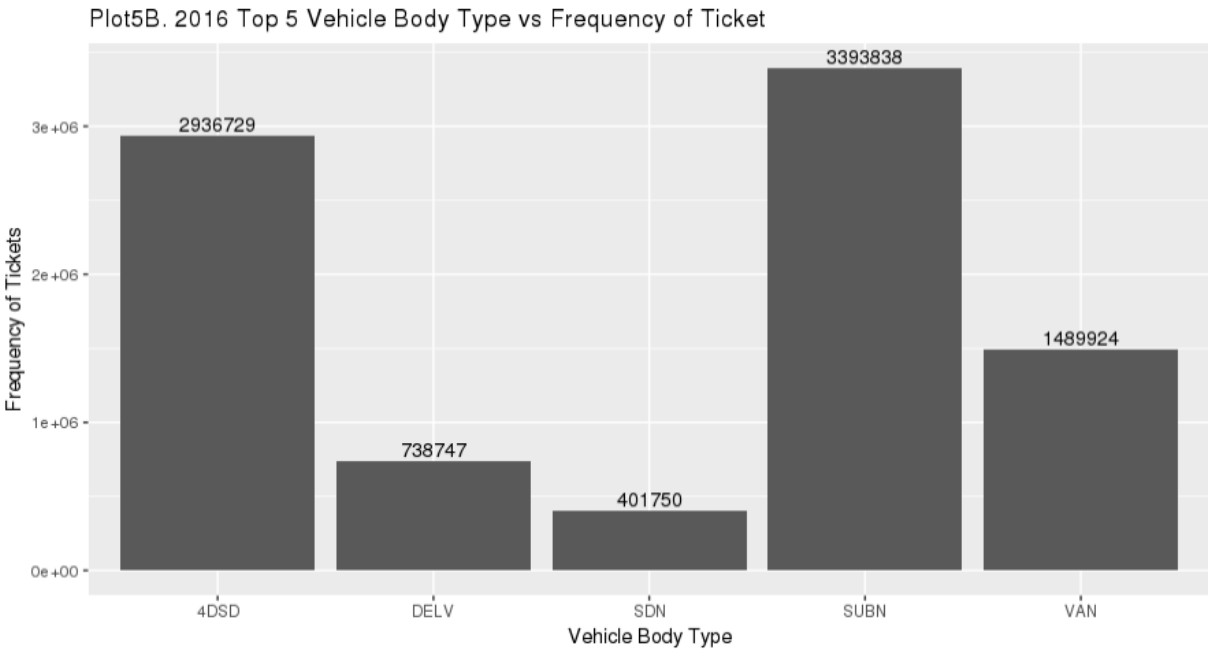


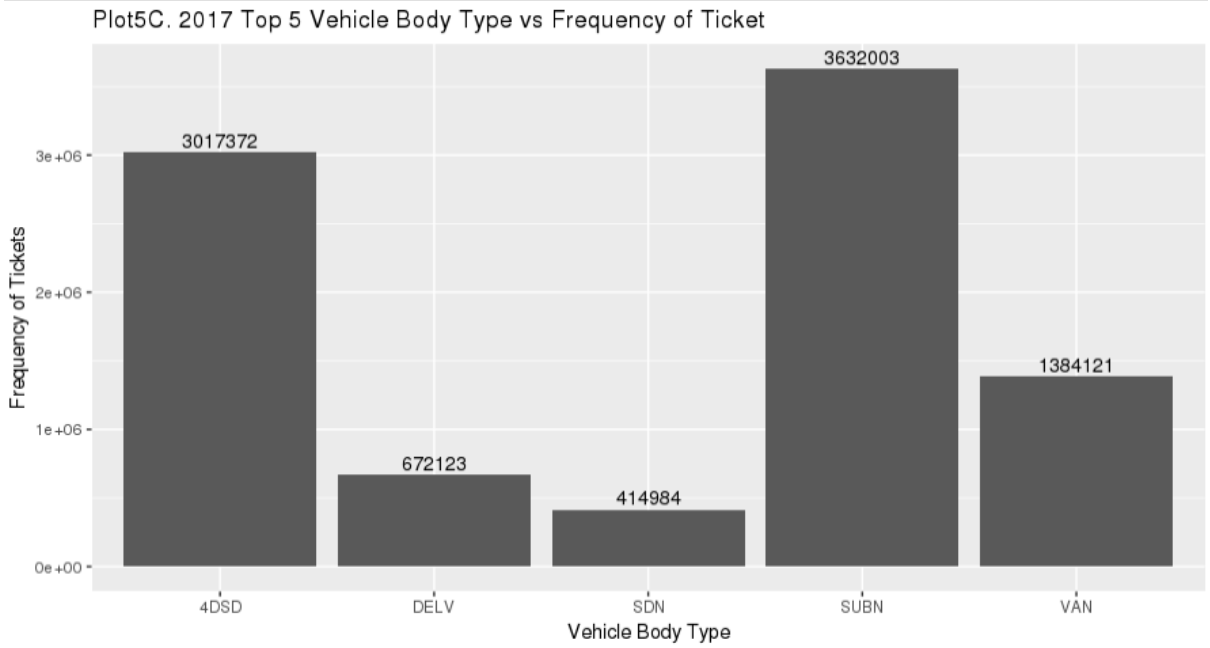
2. How often does each 'vehicle body type' get a parking ticket? How about the 'vehicle Make'? (***Hint***: *find the top 5 for both)*

Vehicle Body Type:

Vehicle Body Type	***Frequency of tickets for 2015***	***Frequency of tickets for 2016***	***Frequency of tickets for 2017***
SUBN	3341110	3393838	3632003
4DSD	3001810	2936729	3017372
VAN	1570227	1489924	1384121
DELV	822040	738747	672123
SDN	428571	401750	414984

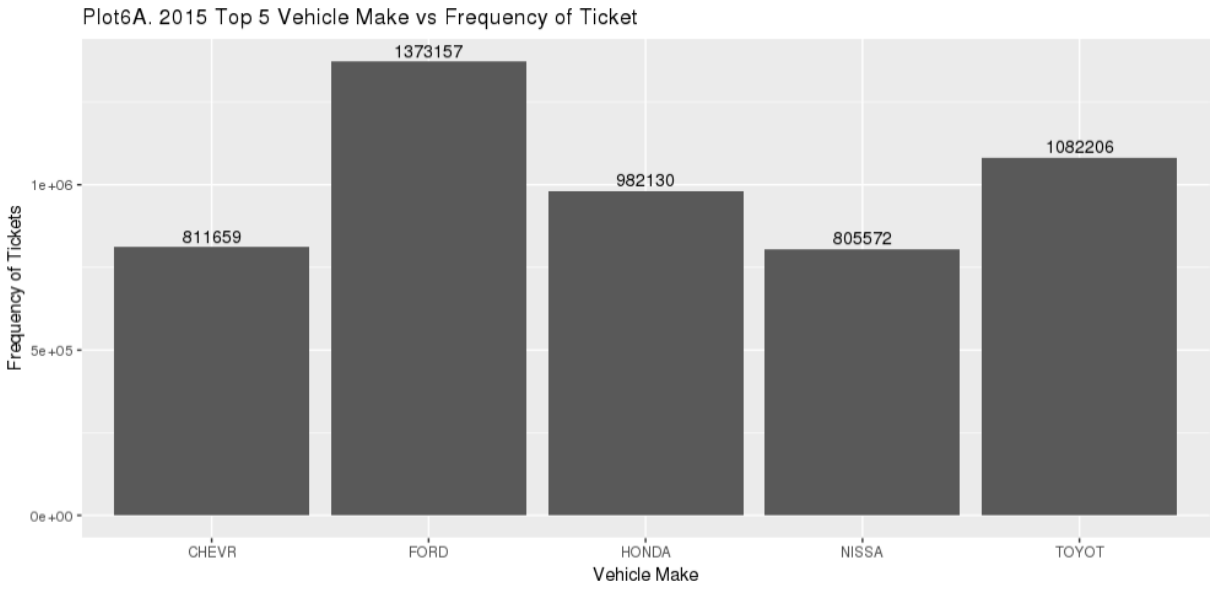


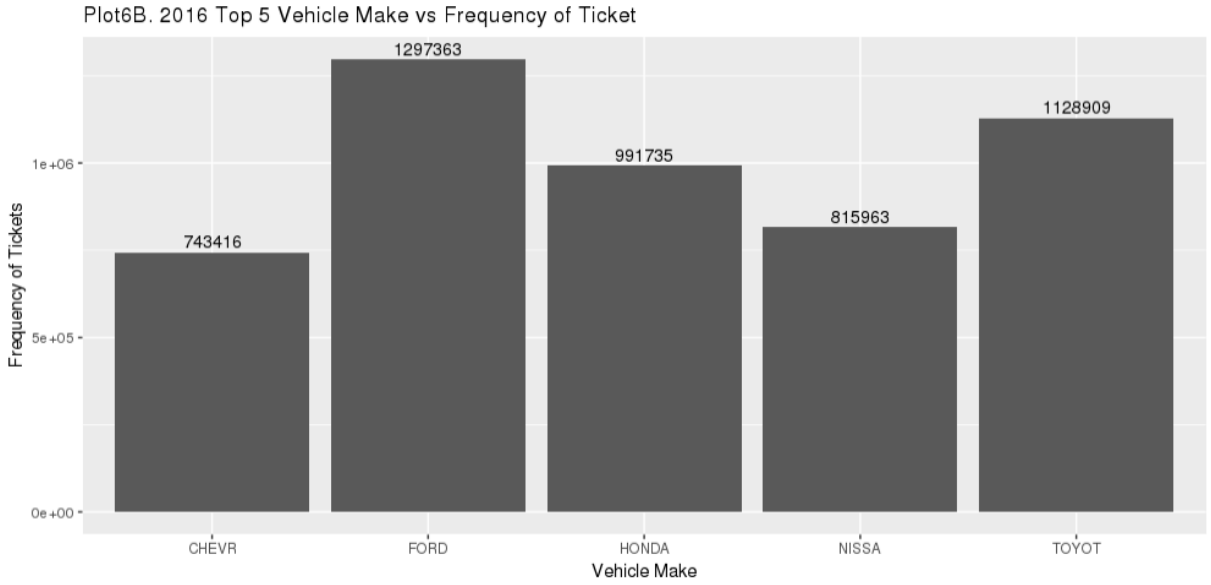


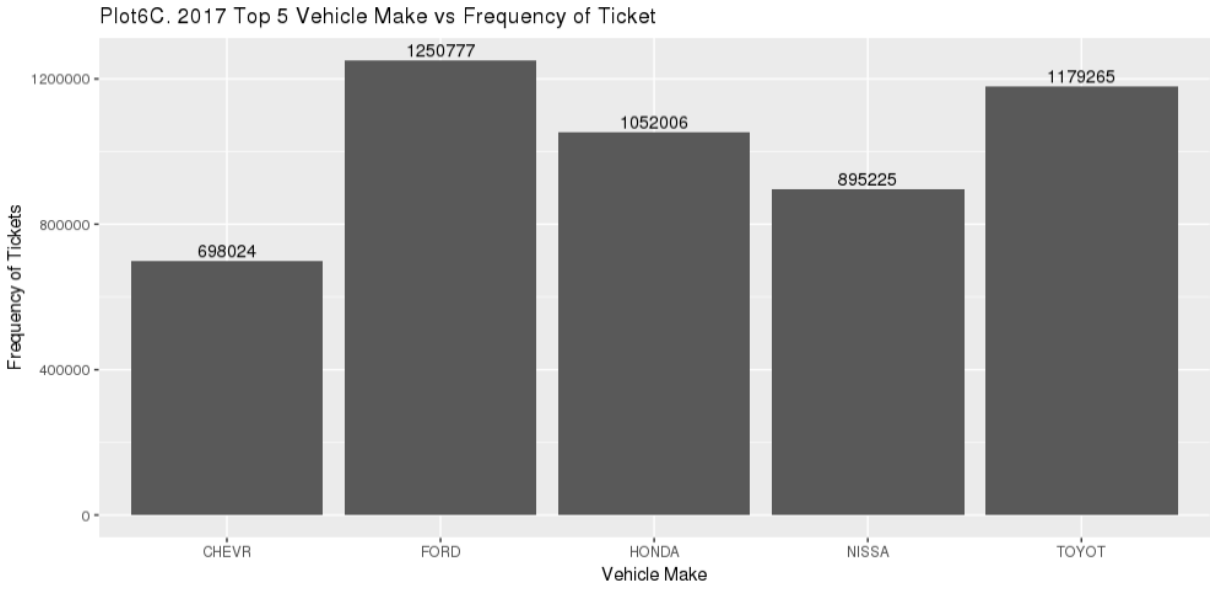


Vehicle Make:

Vehicle Make	*Frequency of tickets for 2015*	*Frequency of tickets for 2016*	*Frequency of tickets for 2017*
FORD	1373157	1297363	1250777
TOYOTA	1082206	1128909	1179265
HONDA	982130	991735	1052006
CHEVR	811659	815963	895225
NISSAN	805572	743416	698024

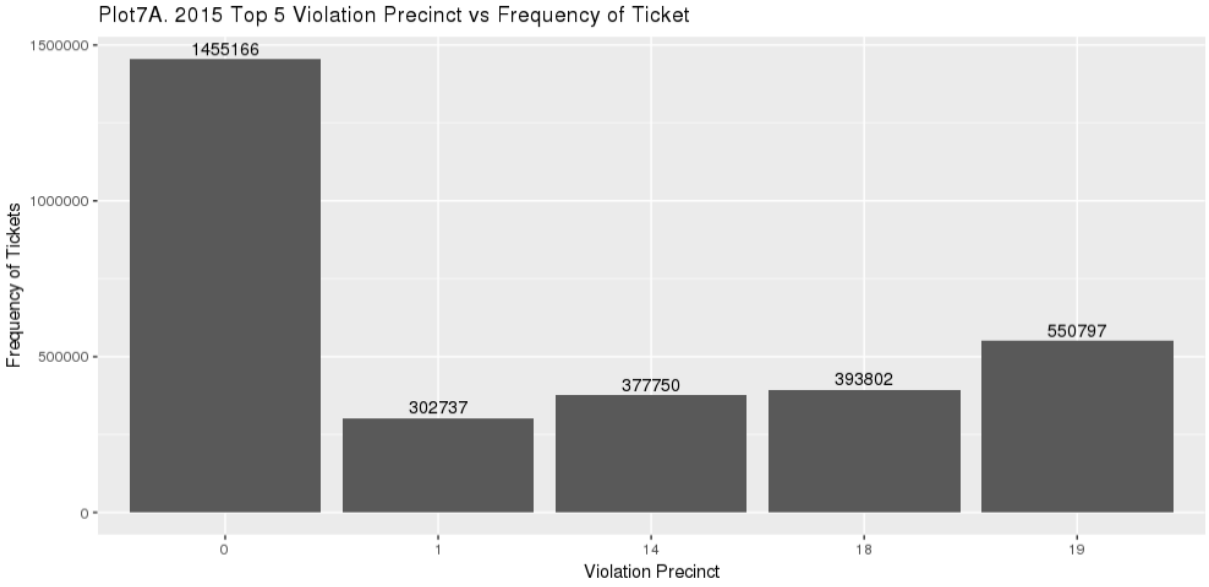

Plot6A. 2015 Top 5 Vehicle Make vs Frequency of Ticket
1373157
1082206
982130
811659
805572
1e+06
5e+05
0e+00
Frequency of Tickets
CHEVR
FORD
HONDA
NISSA
TOYOT
Vehicle Make

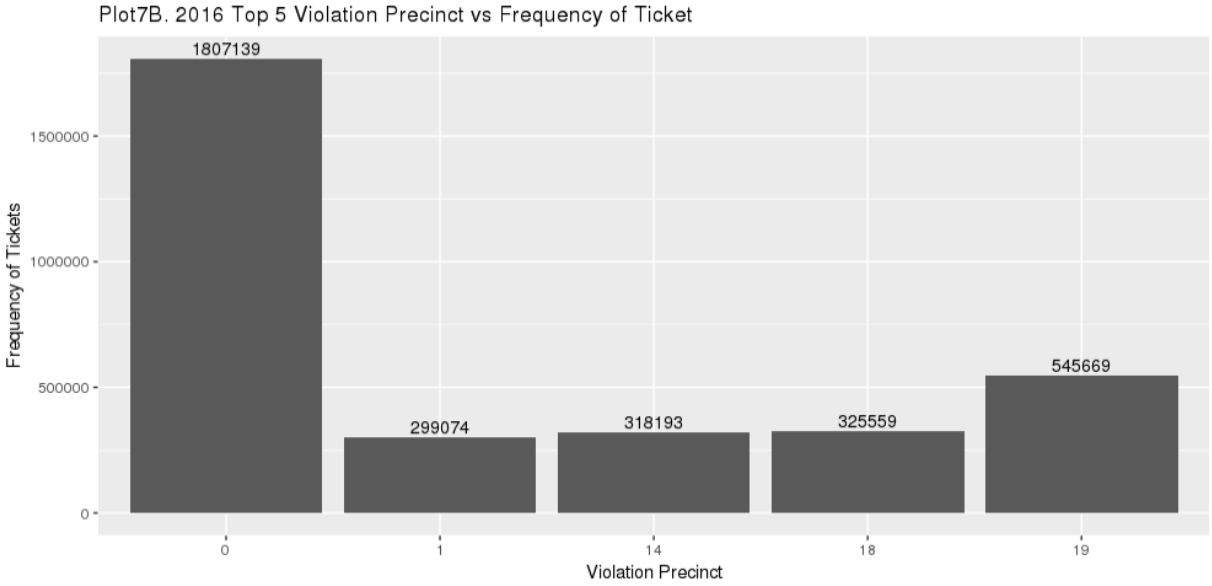

Plot6B. 2016 Top 5 Vehicle Make vs Frequency of Ticket
1297363
1128909
991735
815963
743416
1e+06
5e+05
0e+00
Frequency of Tickets
CHEVR
FORD
HONDA
NISSA
TOYOT
Vehicle Make

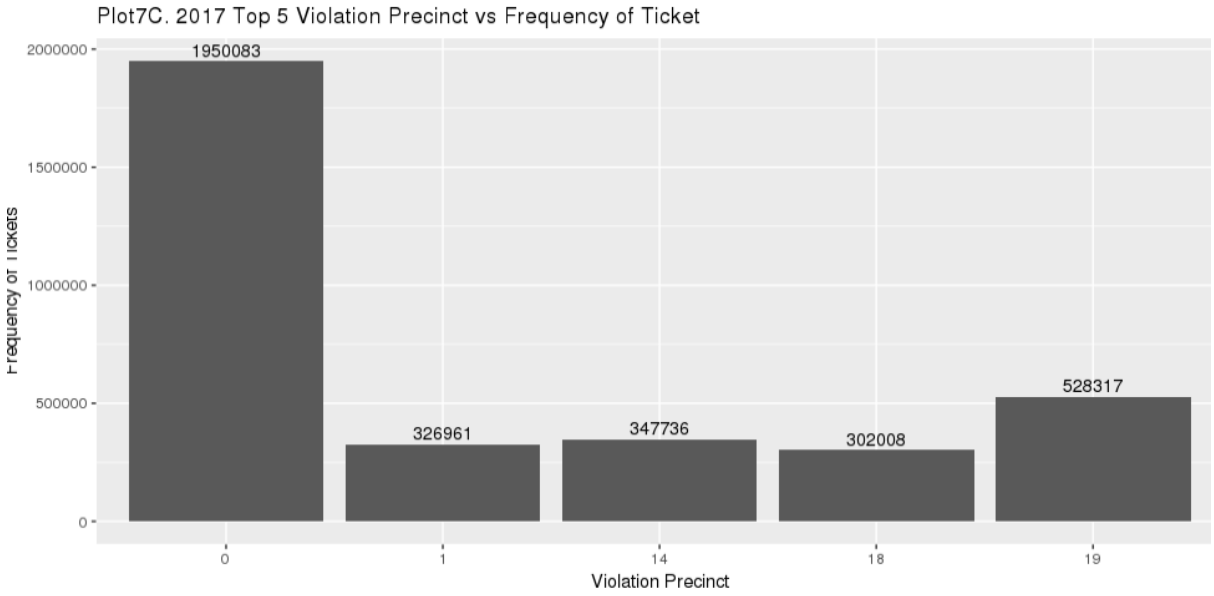

Plot6C. 2017 Top 5 Vehicle Make vs Frequency of Ticket
1250777
1179265
1052006
895225
698024
1200000
800000
400000
0
Frequency of Tickets
CHEVR
FORD
HONDA
NISSA
TOYOT
Vehicle Make

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:
    a. '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?

Violation Precinct	***Frequency of tickets for 2015***	***Frequency of tickets for 2016***	***Frequency of tickets for 2017***
0	1455166	1807139	1950083
19	550797	545669	528317
18	393802	325559	302008
14	377750	318193	347736
1	302737	299074	326961

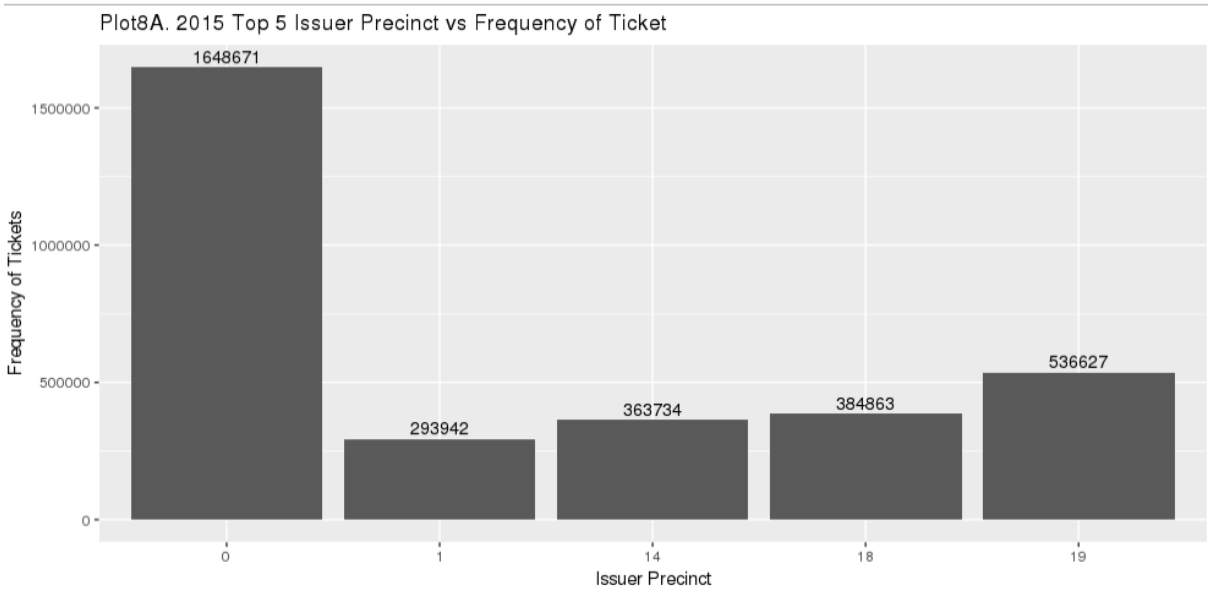


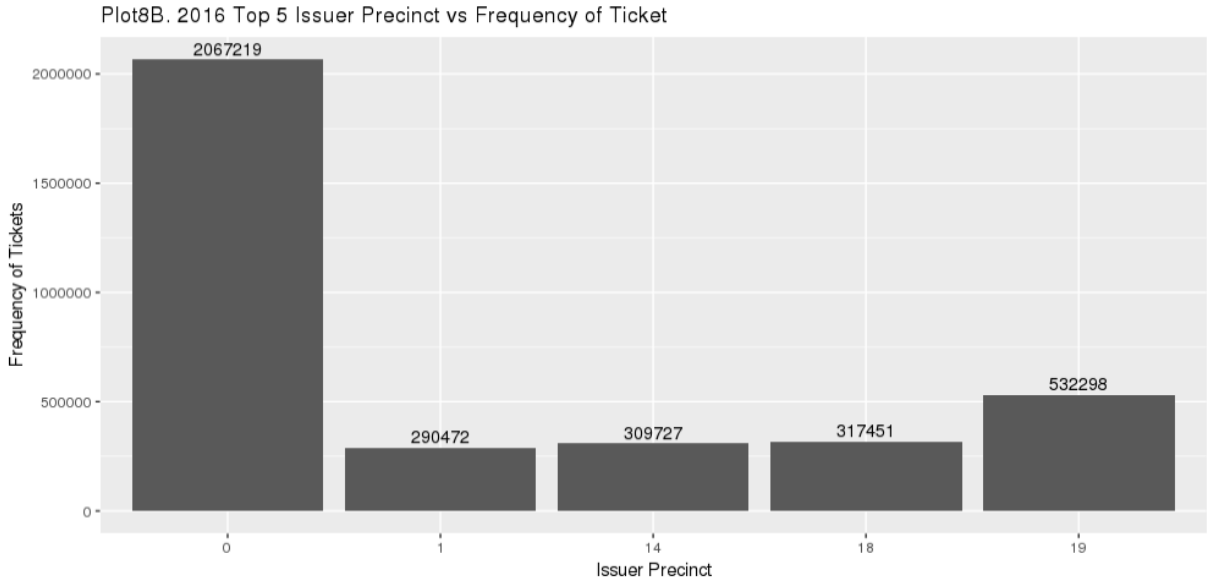


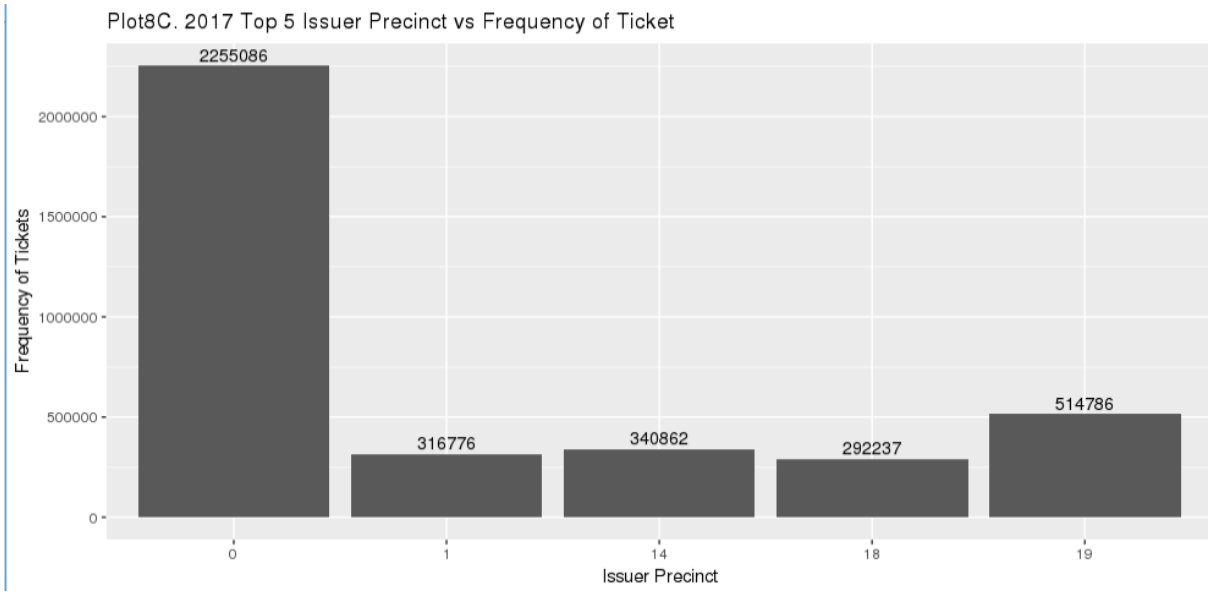


b. 'Issuer Precinct' (this is the precinct that issued the ticket)
*Here you would have noticed that the dataframe has 'Violating Precinct' or 'Issuing Precinct' as '0'. These are the erroneous entries. Hence, provide the record for five correct precincts. (**Hint**: print top six entries after sorting)*

Issuer Precinct	***Frequency of tickets for 2015***	***Frequency of tickets for 2016***	***Frequency of tickets for 2017***
0	1648671	2067219	2255086
19	536627	532298	514786
18	384863	317451	292237
14	363734	309727	340862
1	293942	290472	316776







4. 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?

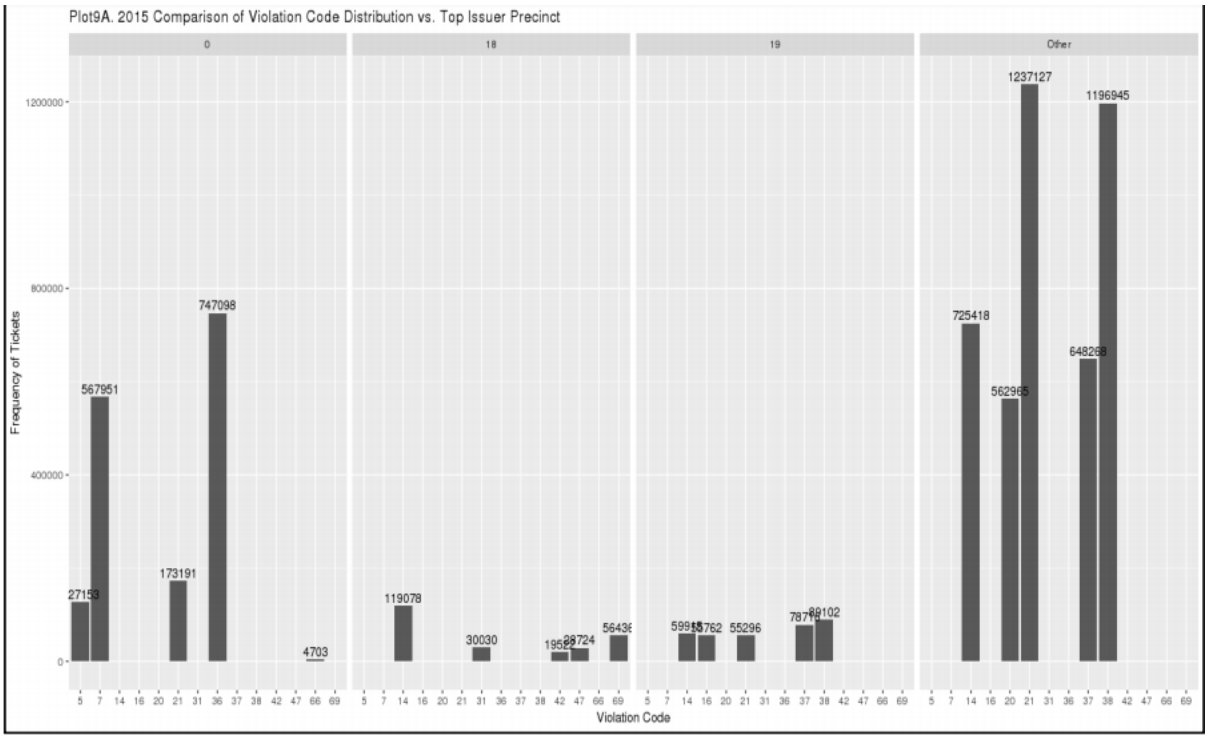
The top 3 ticket Issuing Precincts Codes for the year 2015 are 0, 19 and 18.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	747,098	0
2	7	567,951	0
3	21	173,191	0
4	5	127,153	0
5	66	4,703	0

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	38	89,102	19
2	37	78,716	19
3	14	59,915	19
4	16	55,762	19
5	21	55,296	19

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	119,078	18
2	69	56,436	18
3	31	30,030	18
4	47	28,724	18
5	42	19,522	18

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	21	1,237,127	Other
2	38	1,196,945	Other
3	14	725,418	Other
4	37	648,268	Other
5	20	562,965	Other



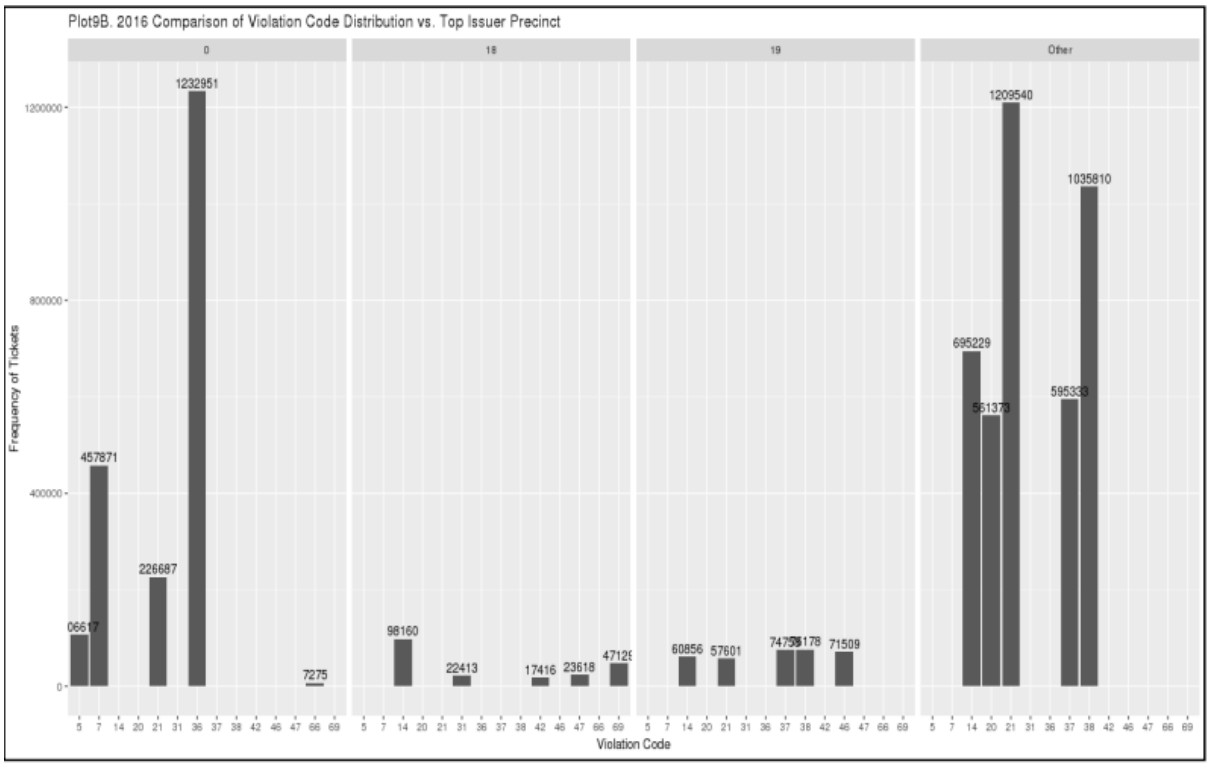
The top 3 ticket Issuing Precincts Codes for the year 2016 are 0, 19 and 18.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	1,232,951	0
2	7	457,871	0
3	21	226,687	0
4	5	106,617	0
5	66	7,275	0

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	38	76,178	19
2	37	74,758	19
3	46	71,509	19
4	14	60,856	19
5	21	57,601	19

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	98,160	18
2	69	47,129	18
3	47	23,618	18
4	31	22,413	18
5	42	17,416	18

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	21	1,209,540	Other
2	38	1,035,810	Other
3	14	695,229	Other
4	37	595,333	Other
5	20	561,373	Other



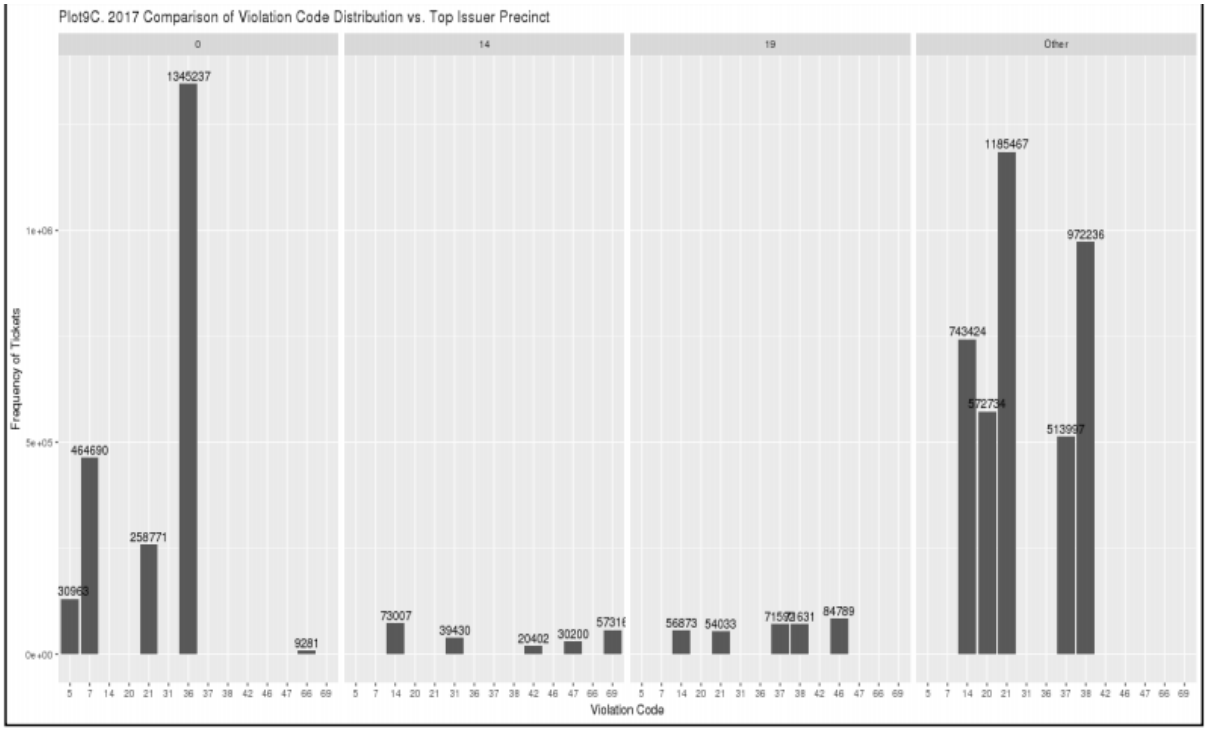
The top 3 ticket Issuing Precincts Codes for the year 2017 are 0, 19 and 14.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	1,345,237	0
2	7	464,690	0
3	21	258,771	0
4	5	130,963	0
5	66	9,281	0

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	46	84,789	19
2	38	71,631	19
3	37	71,592	19
4	14	56,873	19
5	21	54,033	19

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	73,007	14
2	69	57,316	14
3	31	39,430	14
4	47	30,200	14
5	42	20,402	14

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	21	1,185,467	Other
2	38	972,236	Other
3	14	743,424	Other
4	20	572,734	Other
5	37	513,997	Other



While observing the violation code distribution across Issuer Precincts for 2017 there is a clear inconsistency of an exceptionally high count of Tickets with Violation Code 36 and 7 that are issued from Precinct Code 0 which are not consistent in other Precincts. There is also a relatively low count [Inconsistent] of Violation Code 38 in the Top-3 Precincts when compared to the other

5. You'd want to find out the properties of parking violations across different times of the day:

    - Find a way to deal with missing values, if any: ***We will be dropping the missing value rows***.
    - 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.
    - 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.

2015:

Violation Hour (Bin)	***Violation code***	***Freq of tickets***
16-19	38	237513
16-19	37	173008
16-19	14	145601
8-11	21	1167093
8-11	38	442655

8-11	36	353555
4-7	14	132344
4-7	21	103872
4-7	40	89611
12-15	38	559950
12-15	37	411842
12-15	36	317228
0-3	21	39955
0-3	40	36515
0-3	78	33118
20-23	7	69973
20-23	38	61530
20-23	14	44554

2016:

Violation Hour (Bin)	***Violation code***	***Freq of tickets***
16-19	38	208759
16-19	37	159810
16-19	14	132446
8-11	21	1183369
8-11	36	578035
8-11	38	382099
4-7	14	137946
4-7	21	110888
4-7	40	89705
12-15	36	536551
12-15	38	480841
12-15	37	378361
0-3	21	43741
0-3	40	35350
0-3	78	27296
20-23	7	56836
20-23	38	52582
20-23	40	43936

2017:

Violation Hour (Bin)	***Violation code***	***Freq of tickets***
16-19	38	208759
16-19	37	159810
16-19	14	132446
8-11	21	1183369

8-11	36	578035
8-11	38	382099
4-7	14	137946
4-7	21	110888
4-7	40	89705
12-15	36	536551
12-15	38	480841
12-15	37	378361
0-3	21	43741
0-3	40	35350
0-3	78	27296
20-23	7	56836
20-23	38	52582
20-23	14	43936

- 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)

2015:

Violation code	***Violation Hour (Bin)***	***Freq of tickets***
14	0-3	21813
14	12-15	263562
14	16-19	145601
14	20-23	44554
14	4-7	132344
14	8-11	292901
21	0-3	39955
21	12-15	131157
21	16-19	775
21	20-23	530
21	4-7	103872
21	8-11	1167093
38	0-3	560
38	12-15	559950
38	16-19	237513
38	20-23	61530
38	4-7	2783
38	8-11	442655

2016:

Violation code	***Violation Hour (Bin)***	***Freq of tickets***
21	0-3	43741
21	12-15	131534
21	16-19	559

21	20-23	407
21	4-7	110888
21	8-11	1183369
36	0-3	1
36	12-15	536551
36	16-19	40707
36	4-7	77657
36	8-11	578035
38	0-3	376
38	12-15	480841
38	16-19	208759
38	20-23	52582
38	4-7	2171
38	8-11	382099

2017:

Violation code	***Violation Hour (Bin)***	***Freq of tickets***
21	0-3	52280
21	12-15	145598
21	16-19	515
21	20-23	344
21	4-7	116859
21	8-11	1161221
36	12-15	563564
36	16-19	25088
36	4-7	30072
36	8-11	726513
38	0-3	444
38	12-15	457189
38	16-19	200785
38	20-23	46499
38	4-7	2281
38	8-11	343204

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

    Frequency of tickets:

Season	***2015***	***2016***	***2017***

Spring	2860987	2789066	2873383
Summer	2838306	2214536	2353920
Fall	2718502	2971672	2829224
Winter	2180240	2421620	2483036

b. 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 question4).*
2015:

Season	***Violation Code***	***Freq of tickets***
Fall	21	351390
Fall	38	326700
Fall	14	232300
Spring	21	425163
Spring	38	327048
Spring	14	243622
Summer	21	439632
Summer	38	344262
Summer	14	239339
Winter	38	306997
Winter	21	253043
Winter	14	193157

2016:

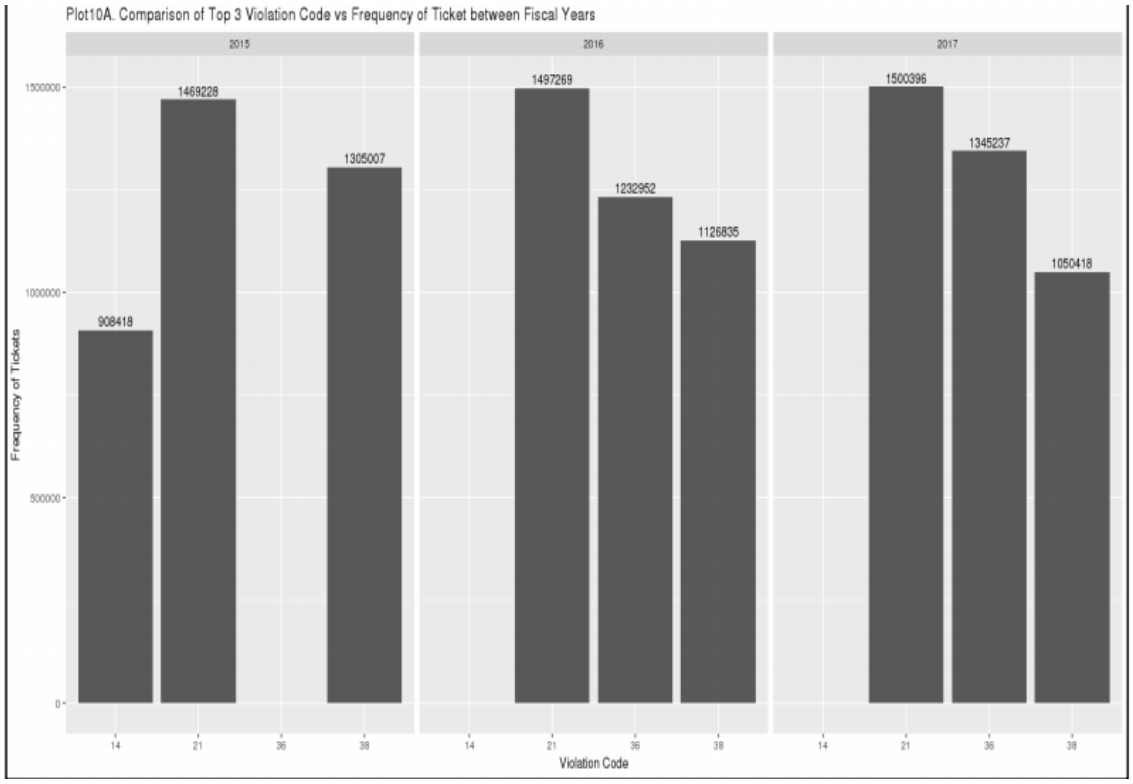
Season	***Violation Code***	***Freq of tickets***
Fall	36	438320
Fall	21	395020
Fall	38	303387
Spring	21	383448
Spring	36	374362
Spring	38	299439
Summer	21	358896
Summer	38	255600
Summer	14	200608
Winter	21	359905
Winter	36	314765
Winter	38	268409

2017:

Season	***Violation Code***	***Freq of tickets***
Fall	36	456046
Fall	21	357257

Fall	38	283216
Spring	21	402424
Spring	36	344834
Spring	38	271167
Summer	21	378699
Summer	38	235725
Summer	14	207495
Winter	21	362016
Winter	36	359338
Winter	38	259710

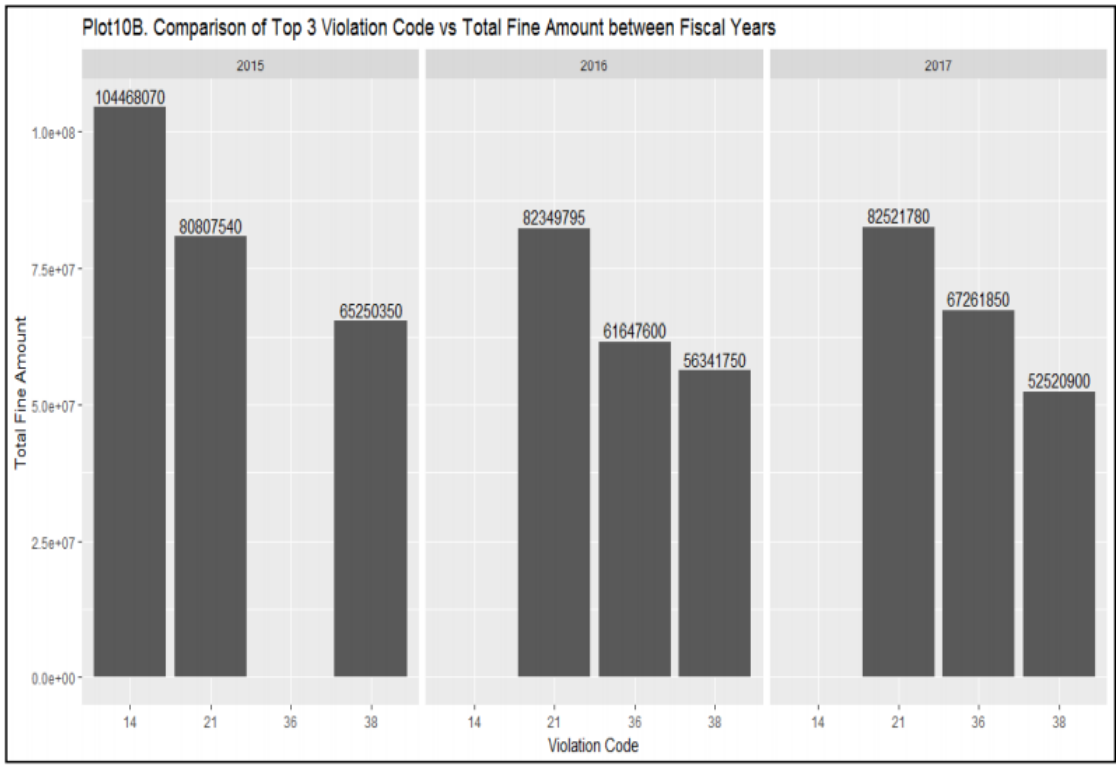
7. The fines collected from all the parking violation constitute a revenue source for the NYC police department. Let's take an example of estimating that for the three most commonly occurring codes.
    a. Find total occurrences of the three most common violation codes



    b. Then, visit the website: http://www1.nyc.gov/site/finance/vehicles/services-violation-codes.page
    It lists the fines associated with different violation codes. 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.

c. Using this information, find the total amount collected for the three violation codes with maximum tickets. State the code which has the highest total collection.
d. What can you intuitively infer from these findings?



From the Total Collection vs. Violation Code Analysis graphs:

a. For the year 2015 Violation Code 14 has the highest total collection at $104.46 million even though it had the lowest frequency out of the top-3 Violation Codes of 2015. This is due to the high average fine amount of $115 per ticket with Violation Code 14. This is followed up closely with Violation Code 21 bringing in $80.80 million in total fine amount for 2015.
b. In the year 2016 and 2017 Violation Code 21 brought the highest total fine amount at $82.34 million and $82.52 million in 2016 and 2017 respectively

Overall Collection Summary:

Except for the year 2015, Violation Code 21 brings in the highest total annual fine amount and remains constant around $80.8-$82.52 million.
Violation Code 14 is a peculiar event bringing in the highest fine amount in the year 2015 at $104.46 million. However, Violation Code 14 does not feature in the top-3 violation codes of the year 2016 and 2017. Violation Code 38 shows a gradual decrease in the total fine amount collected from $65.25 million in 2015 to $52.52 million in 2017. Violation Code 36 only features in the top-3 analysis for 2016 and

2017 with a mild increasing trend of $61.64 million to $67.26 million between 2016 and 2017 respectively.