

New York City Parking Ticket Analysis

Submitted by: The Code Ninjas [1. Alvin Mark Windsor, 2. Uday Balerao, 3. Anand Agarwal]

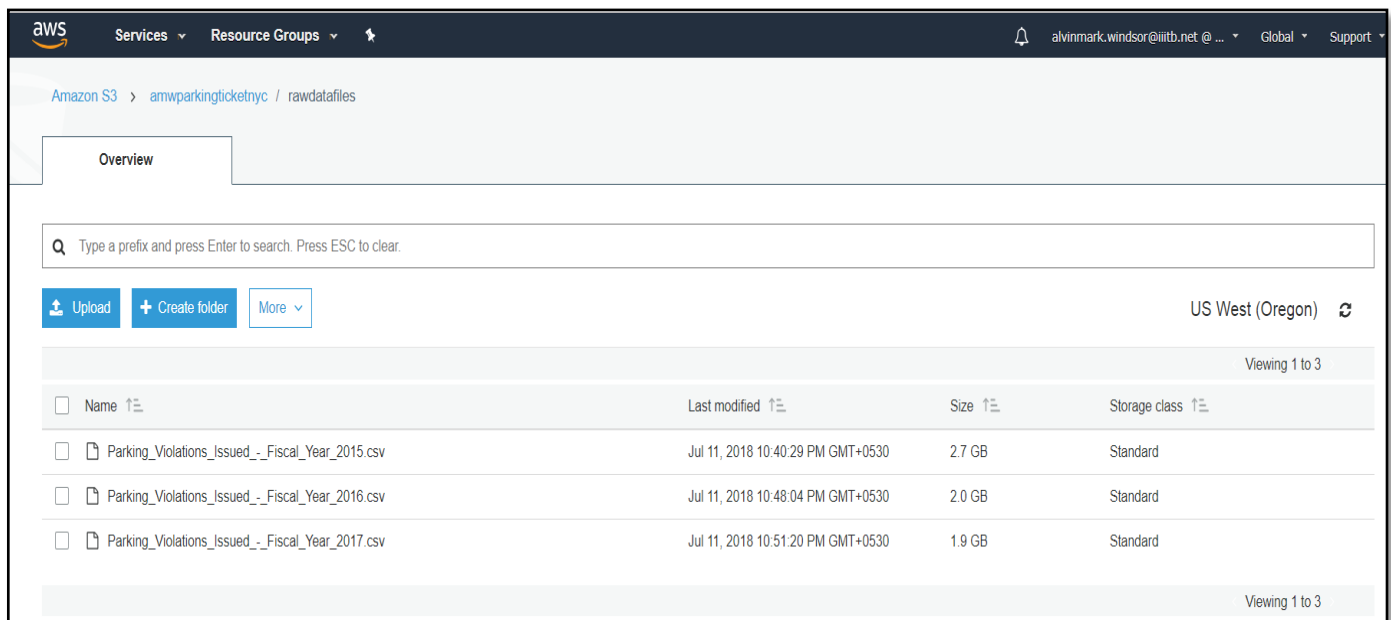
Brief: This Case Study is centered on the concepts of Ingesting and Analyzing big data on the Spark-R platform. Spark will allow us to analyze the full files at high speeds, as opposed to taking a series of random samples that will approximate the population.

Dataset: The dataset used for analysis consists of the Parking Violation Ticket details issued by the NYC Police Department for the fiscal year of 2015, 2016 and 2017.

Our Analysis is focused on performing exploratory analysis to understand the parking violation dataset. We will compare metrics/phenomenon related to Parking Tickets over three different years-2015, 2016 and 2017. The Analysis will be performed on the R-studio interface using SparkR package running on an AWS cluster.

[Step A]: Reading Data into S3 Buckets:

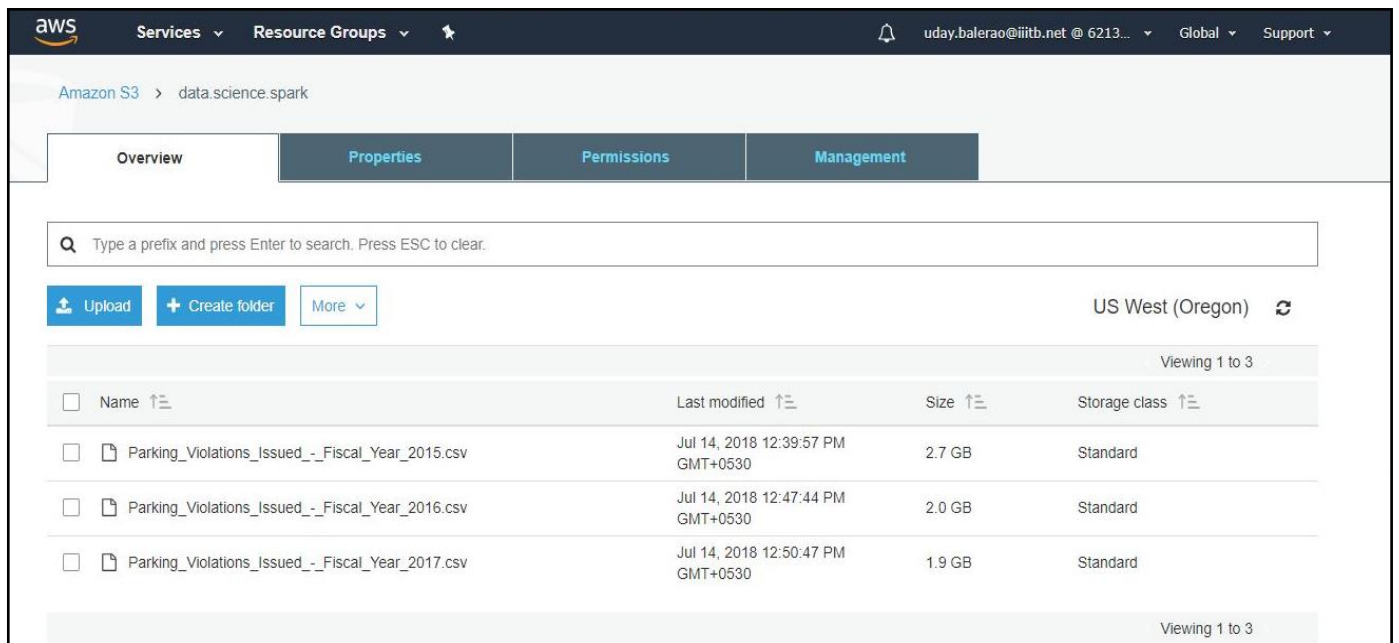
[1] Alvin Mark Windsor



The screenshot shows the AWS S3 console interface for the bucket 'amwparkingticketnyc' in the 'rawdatafiles' folder. The 'Overview' tab is selected. A search bar is present with the placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are buttons for 'Upload', 'Create folder', and 'More'. The region is set to 'US West (Oregon)'. A table lists the contents of the bucket:

Name	Last modified	Size	Storage class
Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Jul 11, 2018 10:40:29 PM GMT+0530	2.7 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Jul 11, 2018 10:48:04 PM GMT+0530	2.0 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Jul 11, 2018 10:51:20 PM GMT+0530	1.9 GB	Standard

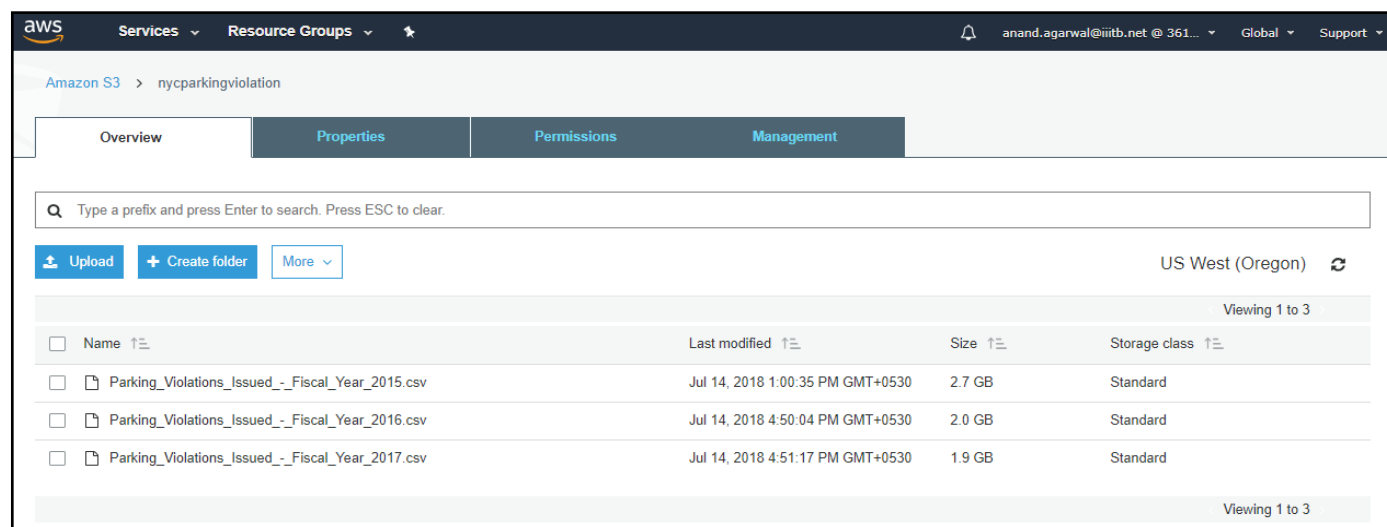
[2] Uday Balerao



The screenshot shows the AWS S3 console interface for the bucket 'data.science.spark'. The 'Overview' tab is selected. A search bar is present with the placeholder text 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are buttons for 'Upload', 'Create folder', and 'More'. The region is set to 'US West (Oregon)'. A table lists the contents of the bucket:

Name	Last modified	Size	Storage class
Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Jul 14, 2018 12:39:57 PM GMT+0530	2.7 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Jul 14, 2018 12:47:44 PM GMT+0530	2.0 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Jul 14, 2018 12:50:47 PM GMT+0530	1.9 GB	Standard

[3] Anand Agarwal



Name	Last modified	Size	Storage class
Parking_Violations_Issued_-_Fiscal_Year_2015.csv	Jul 14, 2018 1:00:35 PM GMT+0530	2.7 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2016.csv	Jul 14, 2018 4:50:04 PM GMT+0530	2.0 GB	Standard
Parking_Violations_Issued_-_Fiscal_Year_2017.csv	Jul 14, 2018 4:51:17 PM GMT+0530	1.9 GB	Standard

[Step B]: Prelude to Analysis

[A.] Methodology: We will conduct the analysis in 3 stages.

Stage 1: Data Quality Verification and Cleaning

Stage 2: Overview and examining the dataset

Stage 3: Deriving and Comparing Metrics through Aggregation Tasks.

[B.] Assumptions and Data Handling

1. We have three datasets available for analysis [Refer. Screenshot attached above]. We will split the dataset into three financial years each comprising of 12 months and conduct the analysis. In addition to the individual year analysis we will conduct a comparison study between the years whenever applicable.

2. The attribute names in the dataset contain white spaces and special characters. To facilitate ease of analysis we have replaced the white spaces with “_” and removed special characters. The date and time formats are not in a standard format therefore we have conducted data manipulation to extract the required information as per the requirement.

3. We used the source: <https://newyorkparkingticket.com/nyc-parking-ticket-required-elements-redux/> to gain an understanding of the components of NYC Parking Ticket. As each dataset has an excess of 10 million records and 43+ attributes, we have defined Summons Number and Issue Date as critical parameters that are required to consider a record/ticket as valid.

4. As Summons Number is a 10 digit unique identifier for a parking ticket we have used this attribute to remove duplicate entries in each dataset.

5. When observing the Issue Date range for each individual dataset we found the following non-conforming results.

a. 2015 Dataset: 1985-07-16 to 2015-06-30 [YYYY-MM-DD format]

b. 2016 Dataset: 1970-04-13 to 2069-10-02 [YYYY-MM-DD format]

c. 2017 Dataset: 1972-03-30 to 2069-11-19 [YYYY-MM-DD format]

Clearly the dataset needs to be cleaned before analysis. If we filter the data based on the Gregorian Calendar/Regular calendar it leads to drastic loss in data. For example in the 2015 dataset if we remove all records not belonging to the year 2015 the resulting dataset is only 46.34% of the original dataset.

Therefore, we considered USA federal government fiscal year. Refer Source: The most important fiscal year for the economy is the federal government's fiscal year. It defines the U.S. government's budget. It runs from October 1 of the budget's prior year through September 30 of the year being described. <https://www.thebalance.com/fiscal-year-definition-federal-budget-examples-3305794> . However, the 2015 dataset contains records only till 30th June 2015. Therefore in the interest of retaining majority of the valuable data we adapted the fiscal year to contain records from 1st July of the Previous Year to 30th June of the year being described. We have used the above concept to remove non conforming records.

6. There are differing number of attributes in the 2015, 2016 and the 2017 dataset. We have removed columns like "No_Standing_or_Stopping_Violation", "Hydrant_Violation", "Double_Parking_Violation", "Latitude", "Longitude", "Community_Board", "Community_Council", "Census_Tract", "BIN", "BBL" and "NTA" as they predominantly contained null/missing values. We've standardized the columns between the 3 datasets.

7. We checked the following Source: <https://www1.nyc.gov/site/nypd/bureaus/patrol/precincts-landing.page> and identified that there is no Precinct with the Code 0. However, while performing analysis for Violation_Precinct and Issuer_Precinct we have retained the records for which the Precinct Code was 0 as it contributed to the highest frequency of tickets. Therefore these records must be treated as tickets for which the Issuer/Violation Precinct has not been recorded clearly.

8. For analysis of parking tickets with missing address we have set the constraint that any ticket with either a missing House_Number or Street_Name will be treated as a ticket with missing violation address.

9. **The Violation Time** attribute is stored in a nonconforming string format. We have performed some data manipulation to convert it into a timestamp. While inspecting the records with missing Violation Time values:

Fiscal Year	Frequency of Missing Violation Time	Total Number of Records	Percentage Missing
2015	61603	10598035	0.5812%
2016	63567	10396894	0.6114%
2017	58000	10539563	0.5503%

From this table it is evident that less than 0.6% of records in each year have missing Violation Time and is negligible. Therefore, we will drop records with missing violation time while conducting the analysis on VT.

10. The Seasonality vs. Ticket Frequency analysis was performed by creating buckets for the season with reference to the ticket's Issue_Date [i.e, Issue_Month]. The seasons were defined as follows: [1] Spring runs from March 1 to May 31; [2] Summer runs from June 1 to August 31; [3] Fall (autumn) runs from September 1 to November 30; and [4] Winter runs from December 1 to February 28.

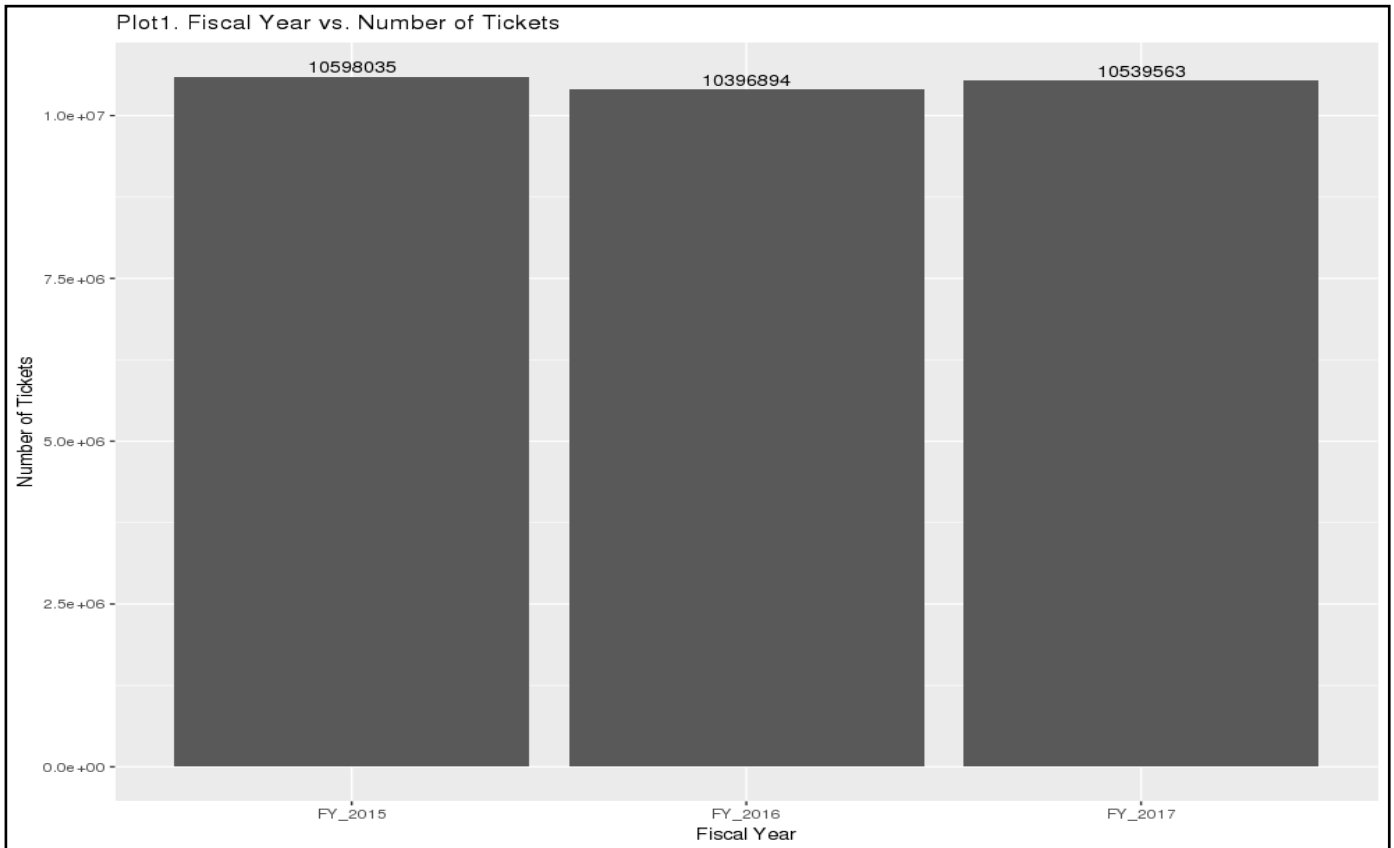
11. The Ticket Fine revenue estimation was carried out with the help of the following Source: <https://www1.nyc.gov/site/finance/vehicles/services-violation-codes.page> to extract the average fare for a ticket based on the violation code.

[C.] Dimensions of Cleaned Dataset used for Analysis

Fiscal Year	Number of Rows	Number of Columns
2015	10,598,035	45
2016	10,396,894	45
2017	10,539,563	45

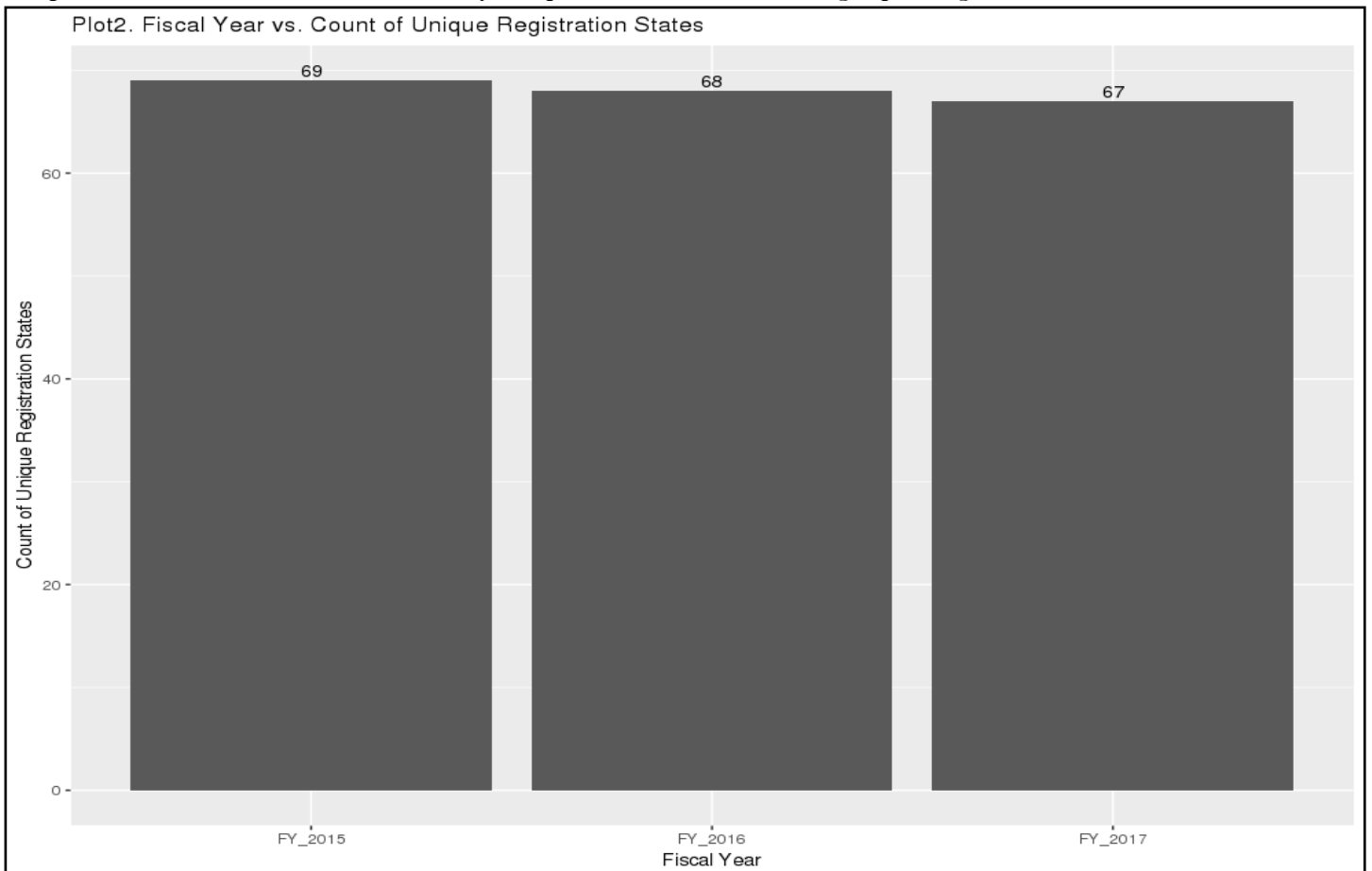
Stage 2: Overview and examining the dataset

Stage 2: Question 1: Find total number of tickets for each year.



Insights: The number of parking tickets issued annually for the past three years range between 10.39-10.59 Million tickets.

Stage 2: Question 2: Find out how many unique states the cars which got parking tickets came from.



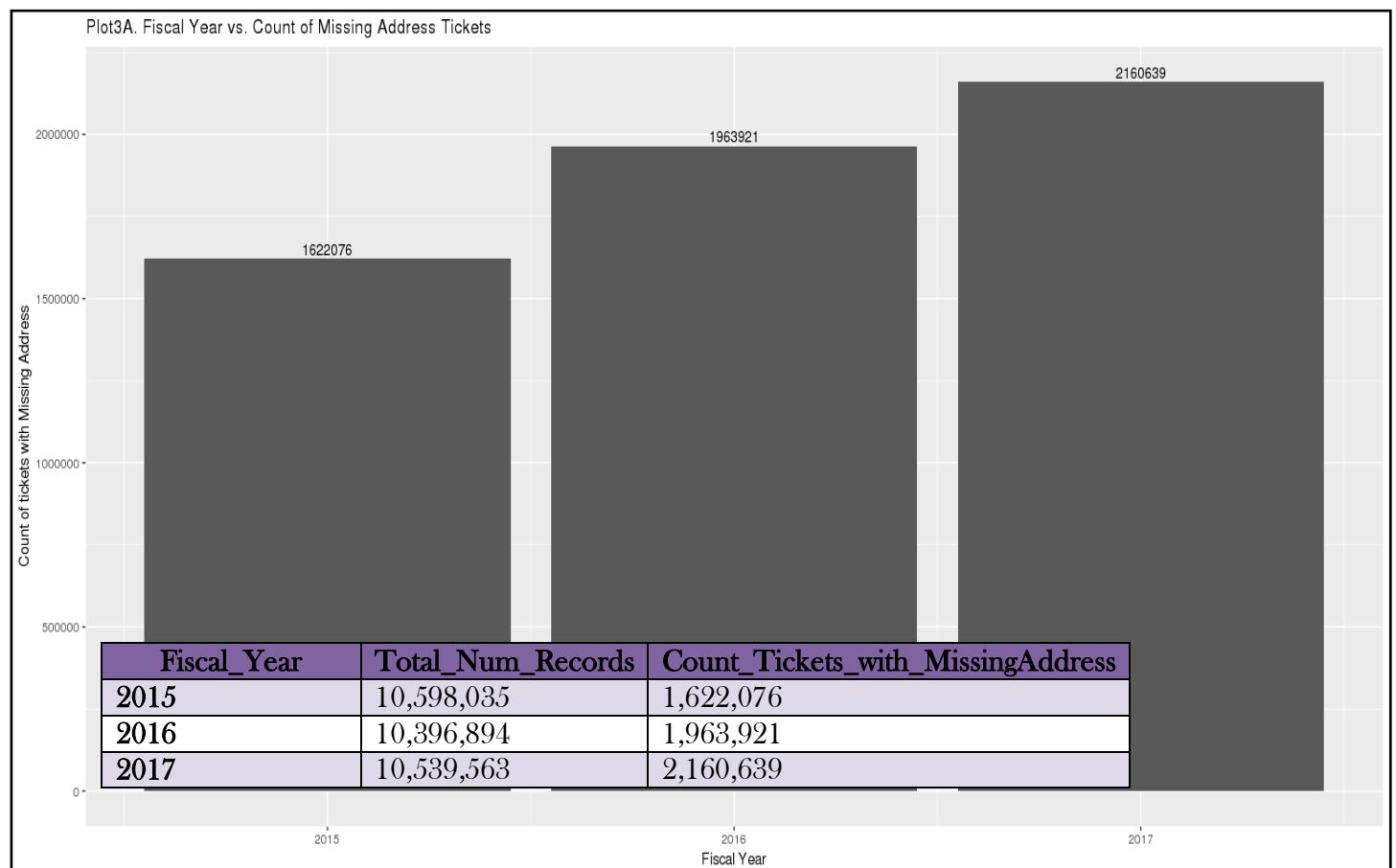
Sl. No.	2015			2016			2017		
	State Code	Number_of_Tickets	State Name	State Code	Number_of_Tickets	State Name	State Code	Number_of_Tickets	State Name
1	NY	8533145	New York	NY	8260189	New York	NY	8481061	New York
2	NJ	996951	New Jersey	NJ	968839	New Jersey	NJ	925965	New Jersey
3	PA	271388	Pennsylvania	PA	259177	Pennsylvania	PA	285419	Pennsylvania
4	CT	149462	Connecticut	CT	145153	Connecticut	FL	144556	Florida
5	FL	138892	Florida	FL	138647	Florida	CT	141088	Connecticut
6	MA	94262	Massachusetts	MA	99115	Massachusetts	MA	85547	Massachusetts
7	IN	78789	Indiana	IN	81141	Indiana	IN	80749	Indiana
8	VA	73361	Virginia	VA	75093	Virginia	VA	72626	Virginia
9	MD	59185	Maryland	MD	60240	Maryland	MD	61800	Maryland
10	NC	55062	North Carolina	NC	55629	North Carolina	NC	55806	North Carolina
11	99	41294	#N/A	99	41585	#N/A	IL	37329	Illinois
12	IL	39827	Illinois	IL	37222	Illinois	GA	36852	Georgia
13	GA	36084	Georgia	GA	35300	Georgia	99	36625	#N/A
14	TX	31697	Texas	TX	32635	Texas	TX	36516	Texas
15	AZ	28760	Arizona	AZ	26603	Arizona	AZ	26426	Arizona

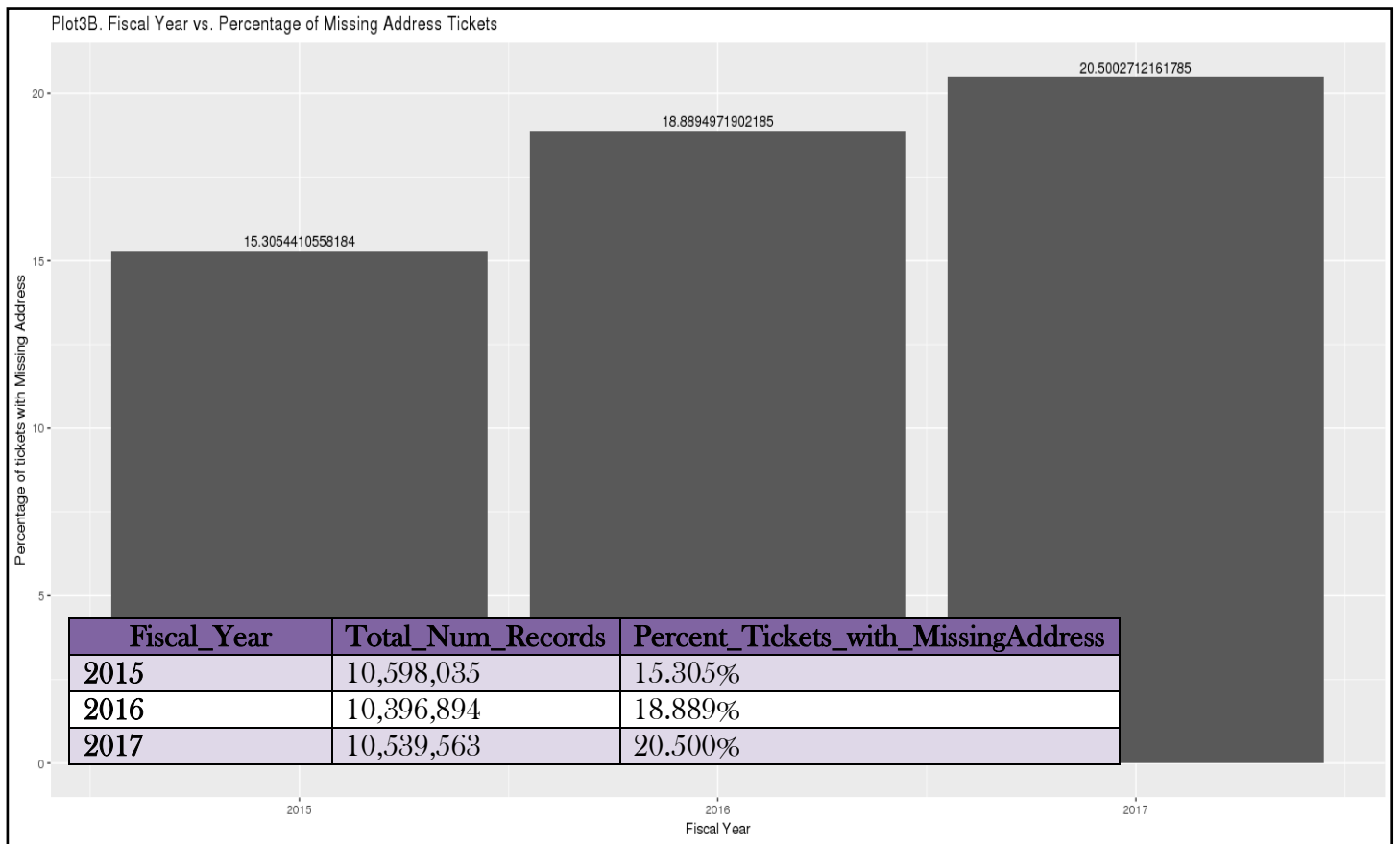
Insight: From the graph it is clear that the number of unique states on the registration plate of cars receiving parking tickets gradually decrease by one from 2015 to 2017. However, while looking at the top 15 registration state wise distribution of parking tickets. There are records with the state code as 99. This implies that there is several parking tickets issued to vehicles with an undefined or unknown registration state.

The state codes present in 2015 dataset that are not present in 2016 and 2017 are show below.

State Code	2015 Dataset	2016 Dataset	2017 Dataset
NF	✓	✗	✗
YT	✓	✓	✗

Stage 2: Question 3: Some parking tickets don't have addresses on them, which is cause for concern. Find out how many such tickets there are.





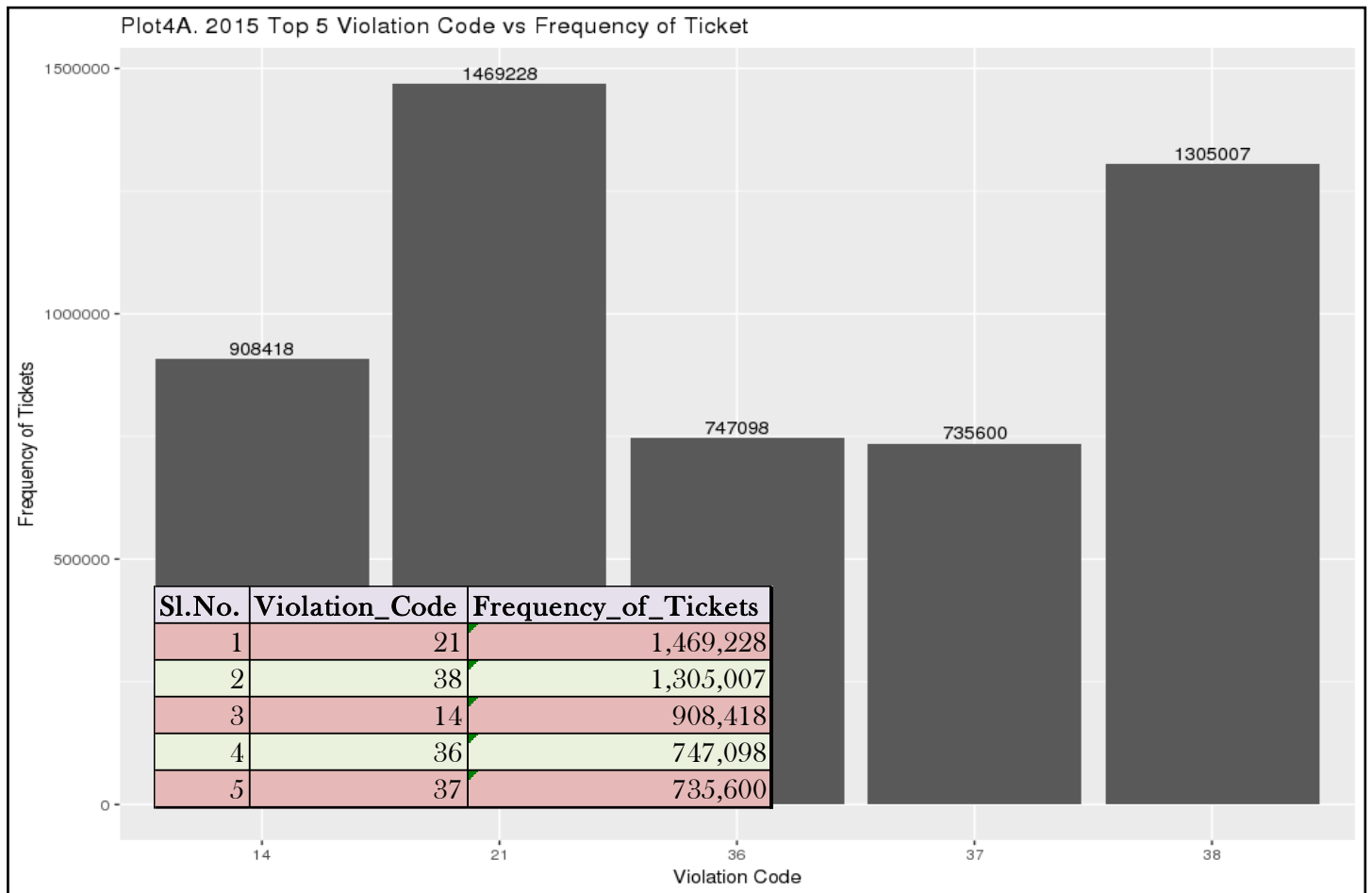
Insight: Tickets with either a missing violation location house number or a missing violation location street name are considered as tickets issued with a missing violation address. From graph 3A there is a clear increasing trend in the count of tickets issued with a missing violation location. It ranges from 1.6 million in 2015 to 2.1 million tickets in 2017. However, since the count is relative to the total number of tickets issued, in graph 3B we study the percentage of tickets issued with a missing violation location. From this graph also it is evident that there is an increasing trend of tickets issued with missing violation address from 15.3% in 2015 to 20.5% in 2017. This needs to be looked into as tickets with missing violation locations can either be misattributed or faulty tickets or it can be contested by the charged vehicle owners in court.

Total_Num_Records	Count_Tickets_with_MissingAddress	Percent_Tickets_with_MissingAddress	Fiscal_Year
10,598,035	1,622,076	15.30544	2015
10,396,894	1,963,921	18.8895	2016
10,539,563	2,160,639	20.50027	2017

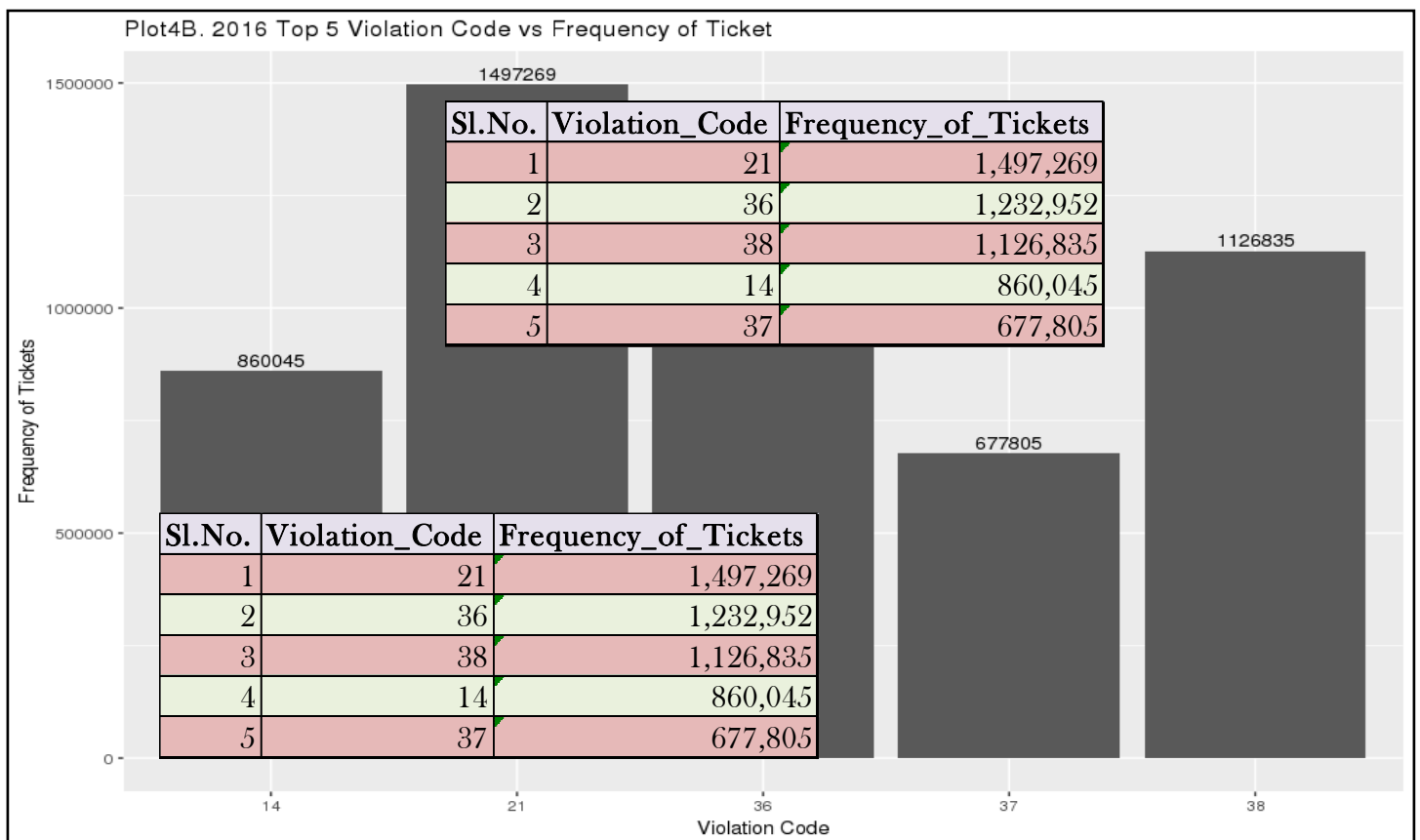
Stage 3: Deriving and Comparing Metrics through Aggregation Tasks.

Stage 3: Question 1: How often does each violation code occur? (Frequency of violation codes - find the top 5)

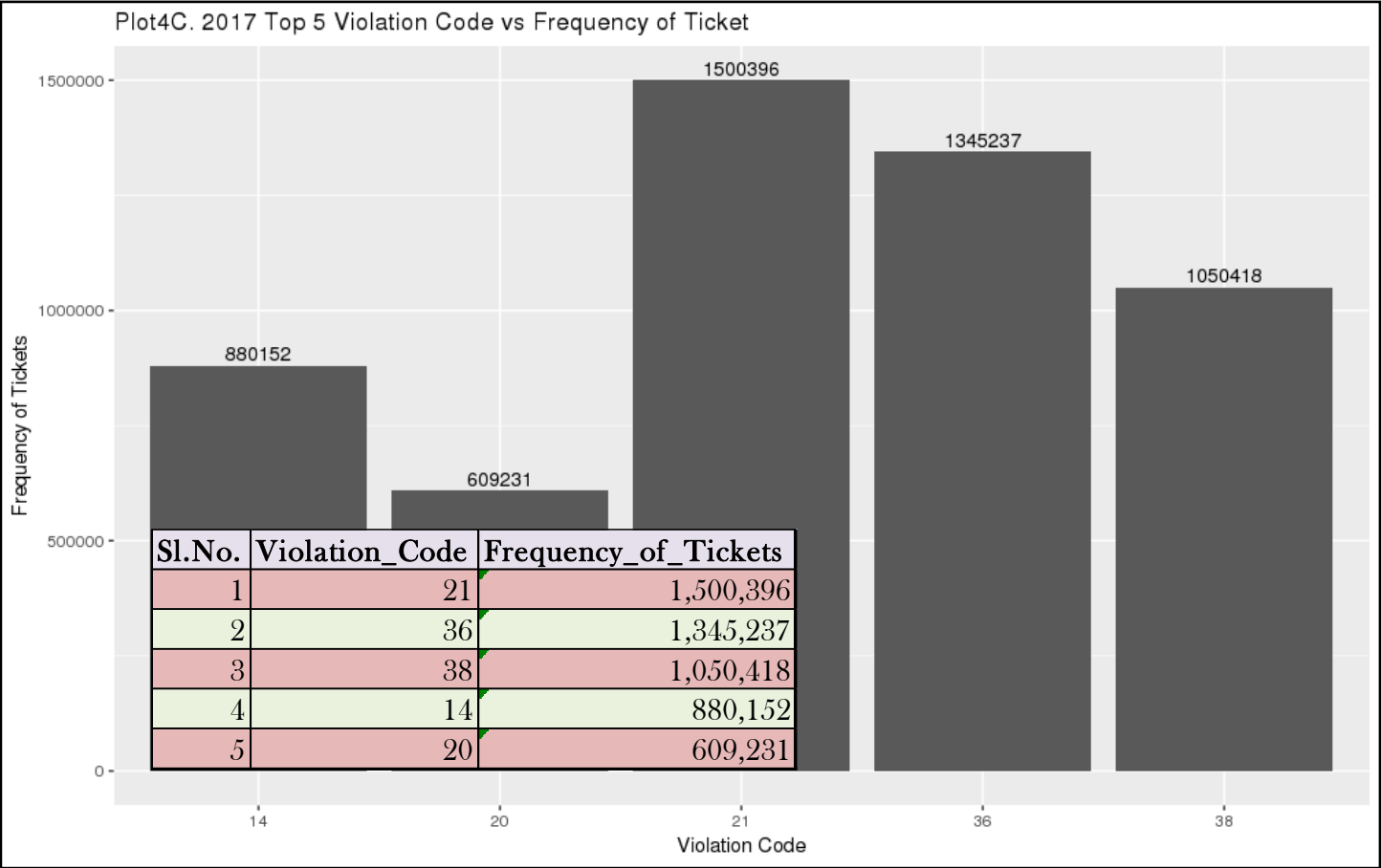
[3.1.1] 2015 Dataset



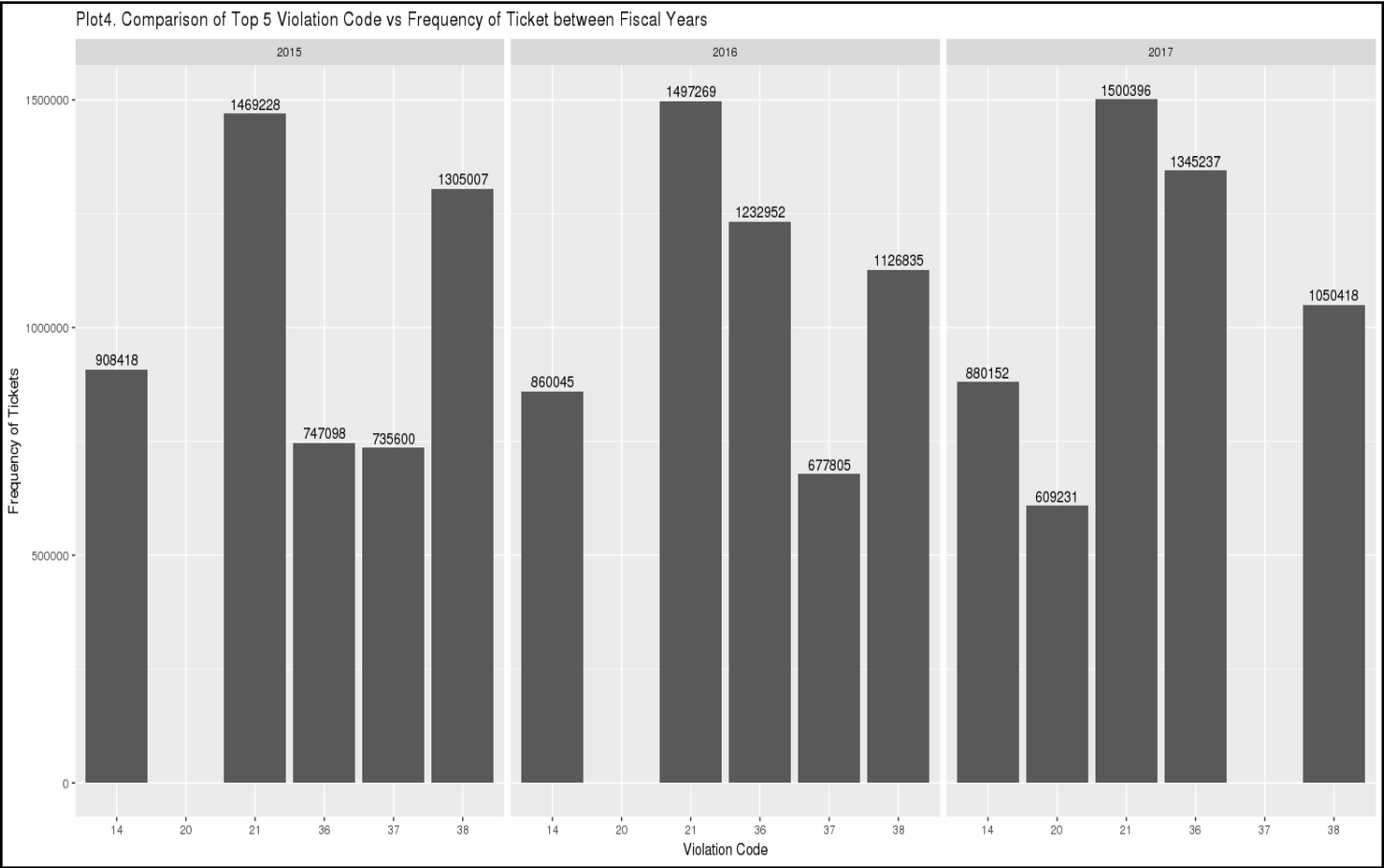
[3.1.2] 2016 Dataset



[3.1.3] 2017 Dataset



[3.1.4] Combined Frequency Distribution of Violation Codes



Insight: From the Analysis of Violation Code frequency between the years it is clear that the most frequent violation for which a parking ticket is issued if for:

[1] Violation Code 21: Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.

The other common violation codes shared between the years are as follows:

[2] Violation Code 14: General No Standing: Standing or parking where standing is not allowed by sign, street marking or; traffic control device.

[3] Violation Code 36: Exceeding the posted speed limit in or near a designated school zone.

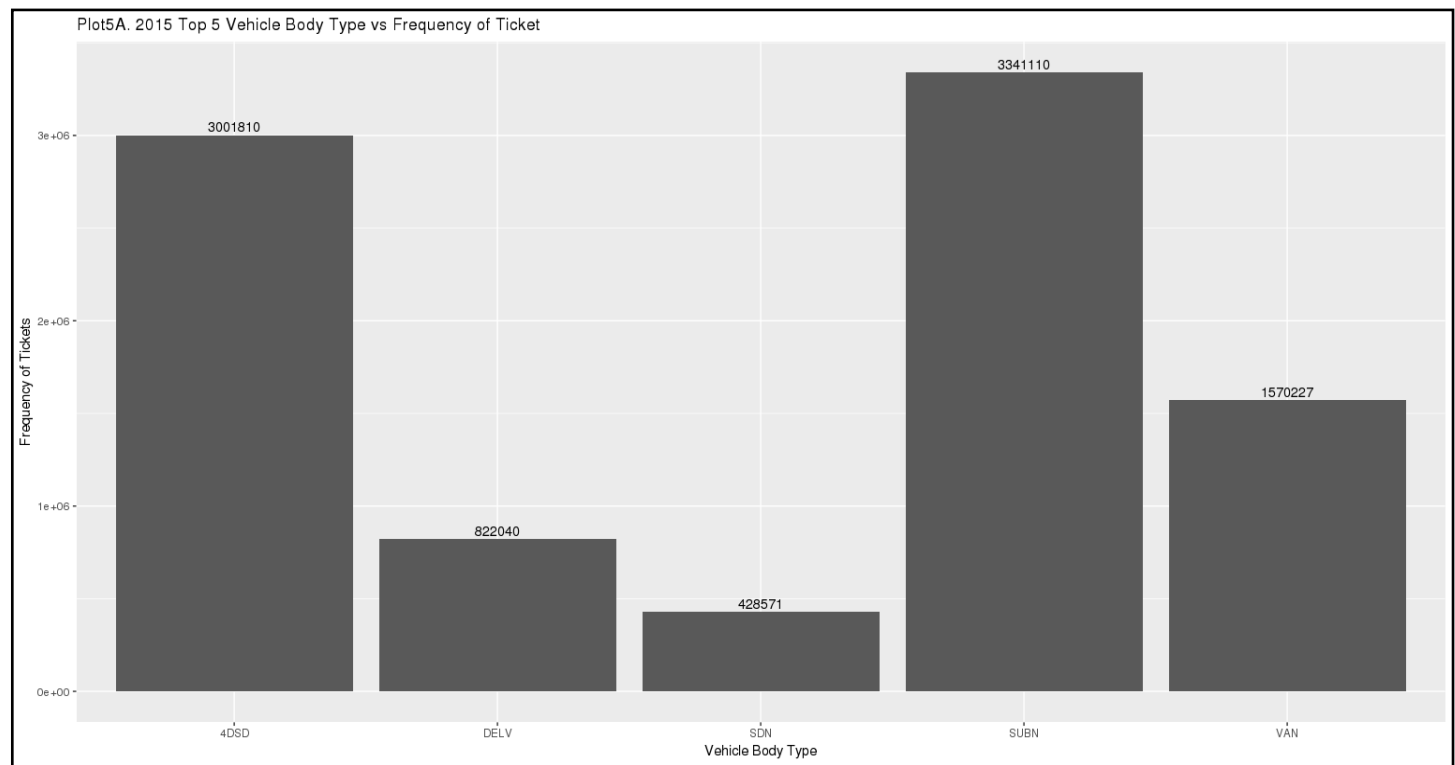
[4] Violation Code 37-38: Muni Meter – (37) Parking in excess of the allowed time. (38) Failing to show a receipt or tag in the windshield. Drivers get a 5-minute grace period past the expired time on Muni-Meter receipts.

Therefore, the highest frequency of tickets through the years is issued for No-parking violations, Exceeding Speed limit near school zones and Expired Muni Meter for parked vehicle.

Stage 3: Question 2: How often does each vehicle body type get a parking ticket? How about the vehicle make? (Find the top 5 for both)

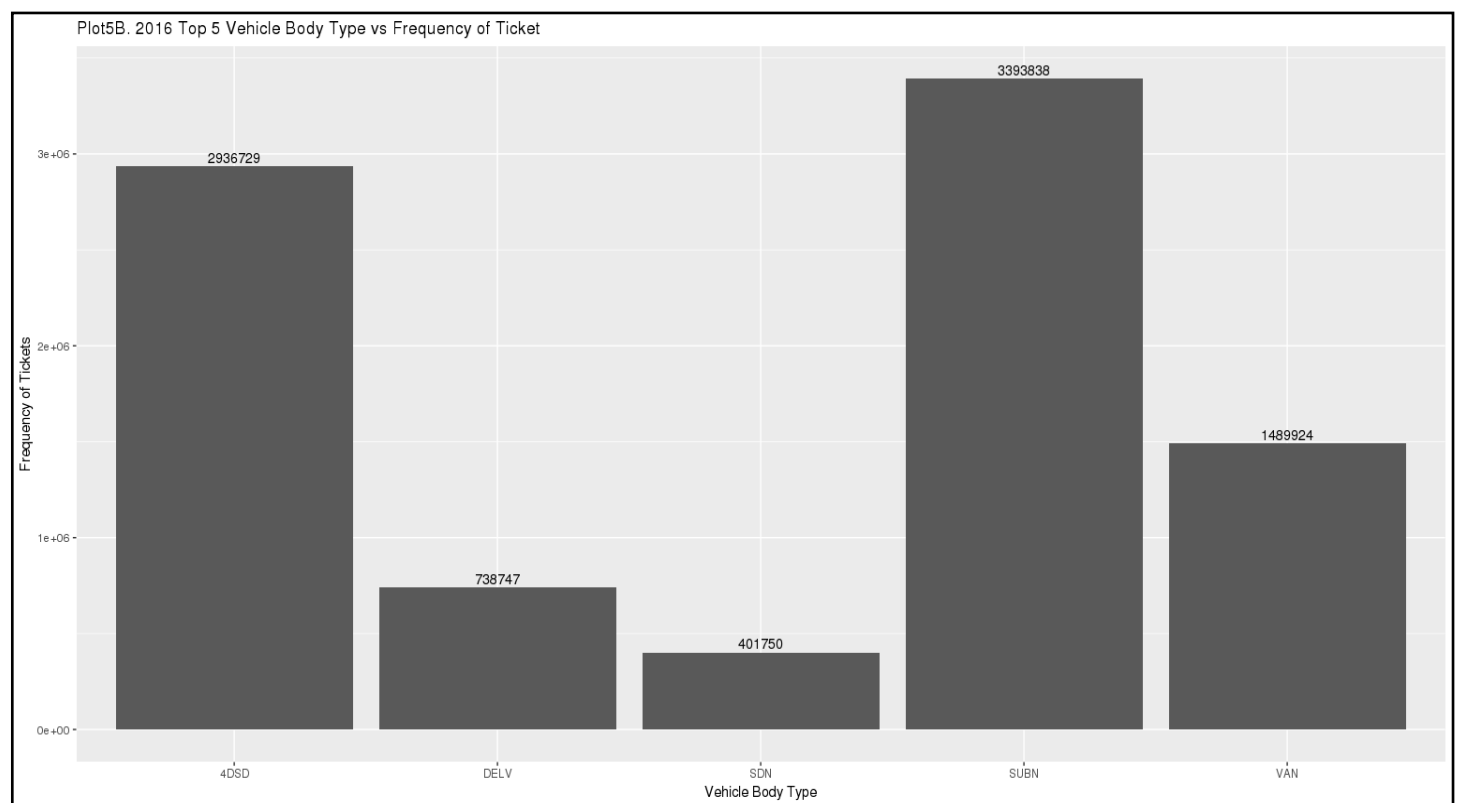
[3.2.1.] Vehicle Body Type vs. Frequency of Ticket Analysis

[3.2.1.1] 2015 Dataset



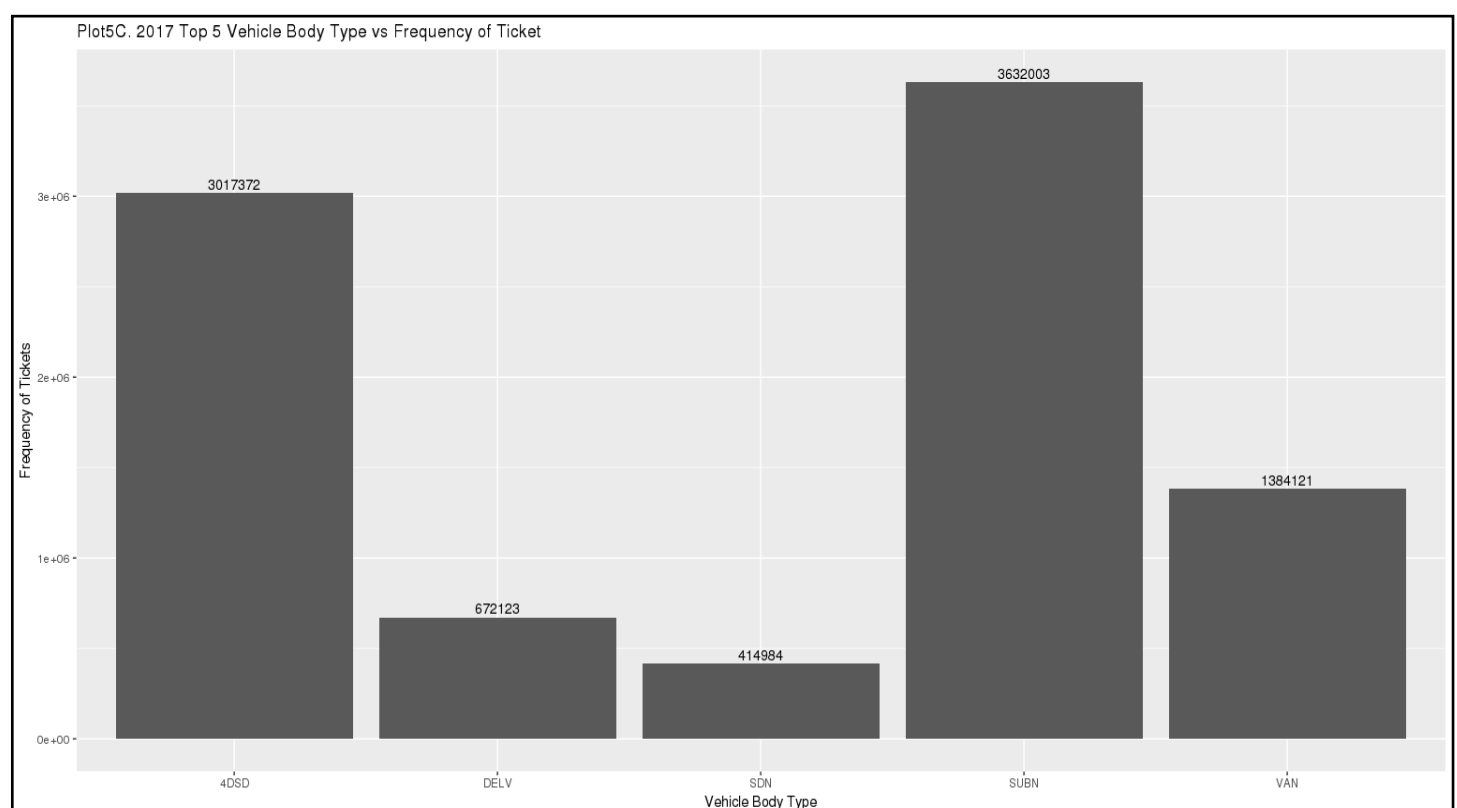
Sl.No.	Vehicle_Body_Type	Frequency_of_Tickets
1	SUBN	3,341,110
2	4DSD	3,001,810
3	VAN	1,570,227
4	DELV	822,040
5	SDN	428,571

[3.2.1.2] 2016 Dataset



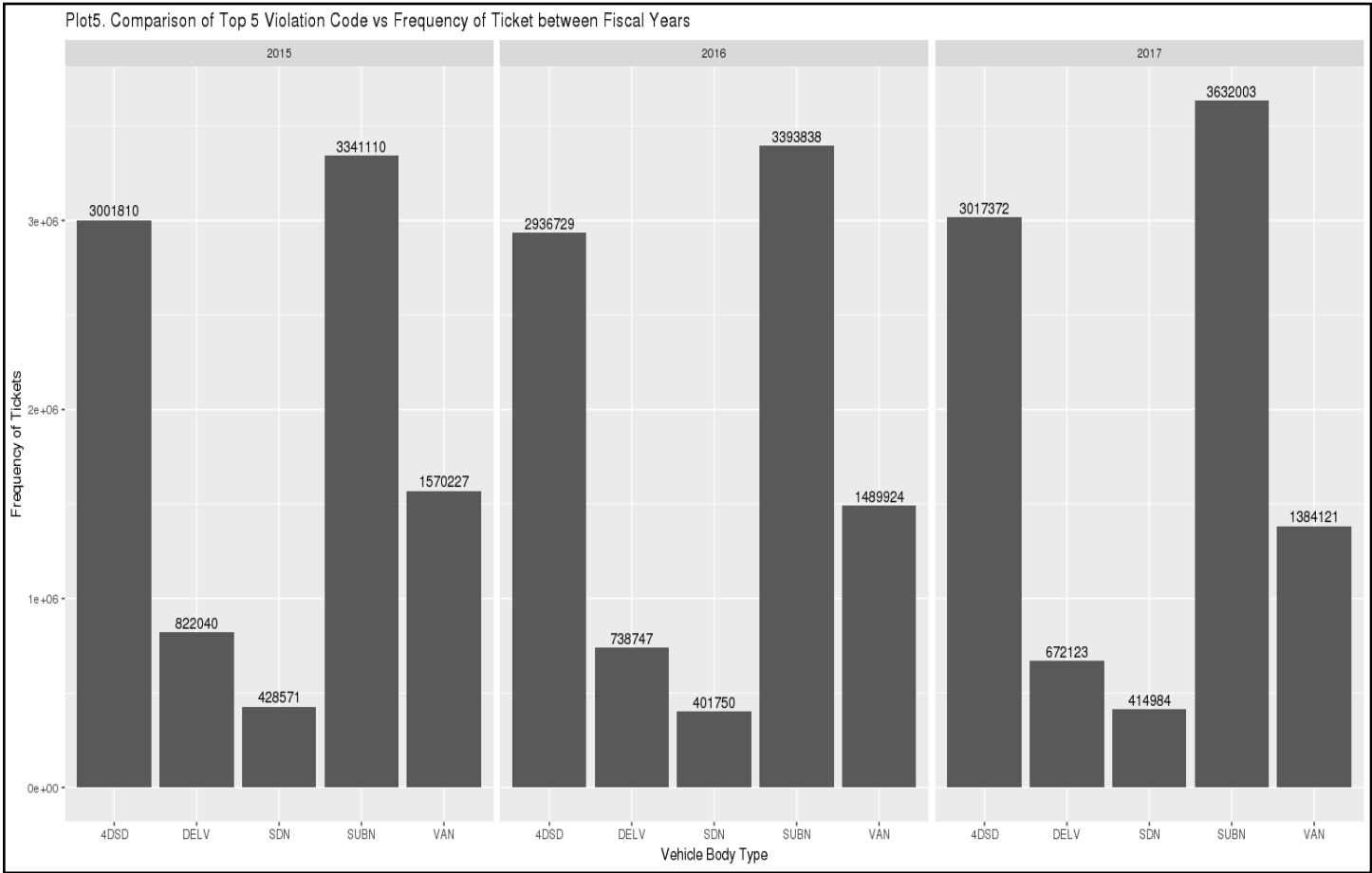
Sl.No.	Vehicle_Body_Type	Frequency_of_Tickets
1	SUBN	3,393,838
2	4DSD	2,936,729
3	VAN	1,489,924
4	DELV	738,747
5	SDN	401,750

[3.2.1.3] 2017 Dataset



Sl.No.	Vehicle_Body_Type	Frequency_of_Tickets
1	SUBN	3,632,003
2	4DSD	3,017,372
3	VAN	1,384,121
4	DELV	672,123
5	SDN	414,984

[3.2.1.4] Combined Vehicle Body Type vs. Frequency of Parking Tickets



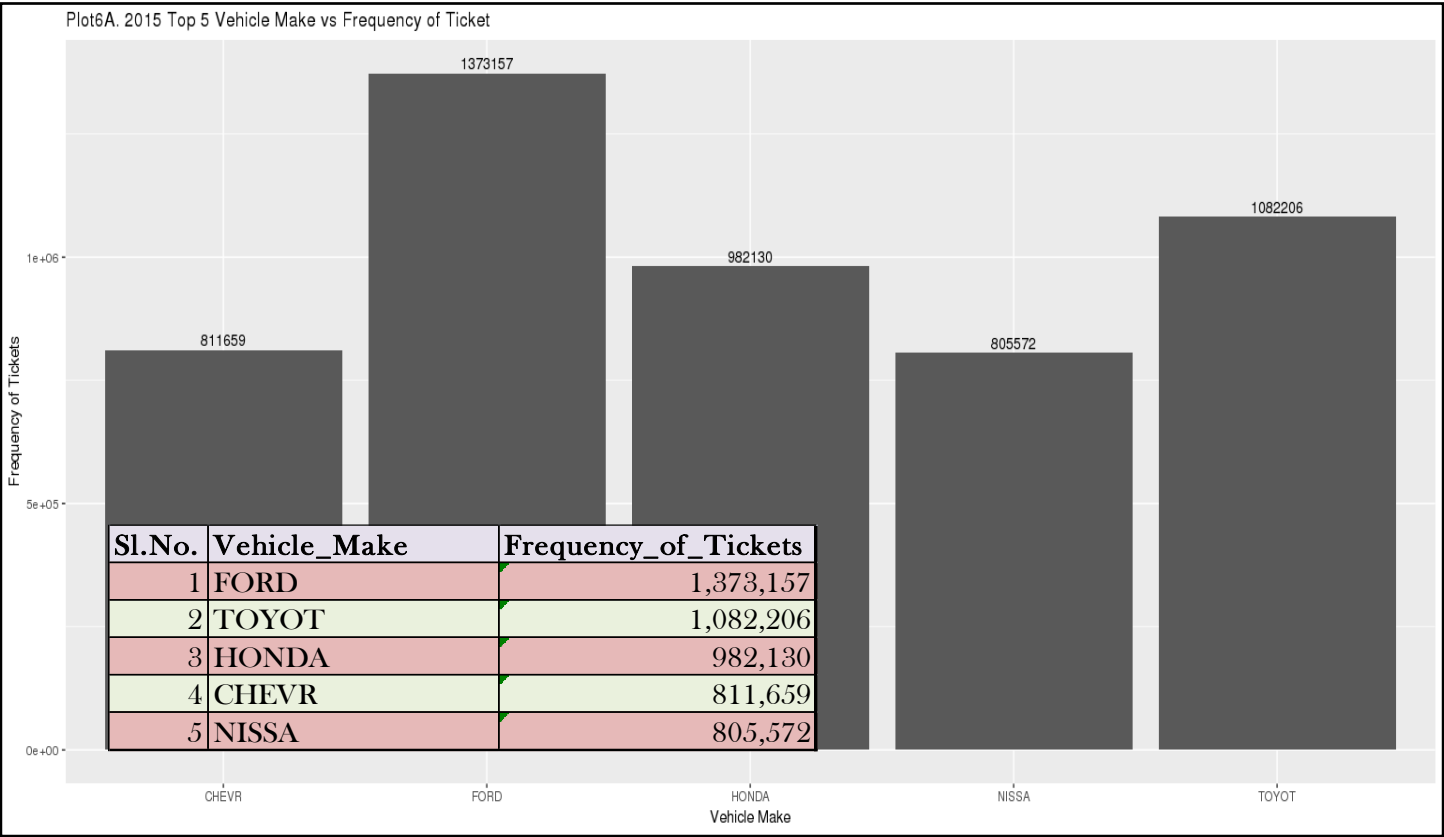
Insights: The top 5 Vehicle Body Types issued with parking tickets remain the same throughout the years. The Vehicle Body Type Codes have the following interpretation:

- [1] SUBN-The law defines a suburban as a vehicle that can be used to carry passengers and cargo. Vehicles that can be registered with the suburban body type include station wagons, sport utility vehicles, hearses and ambulances.
- [2] 4DSD-Four-door sedan
- [3] VAN-Type of road vehicle used for transporting goods or people.
- [4] DELV-Delivery Truck
- [5] SDN-Civilian Sedan

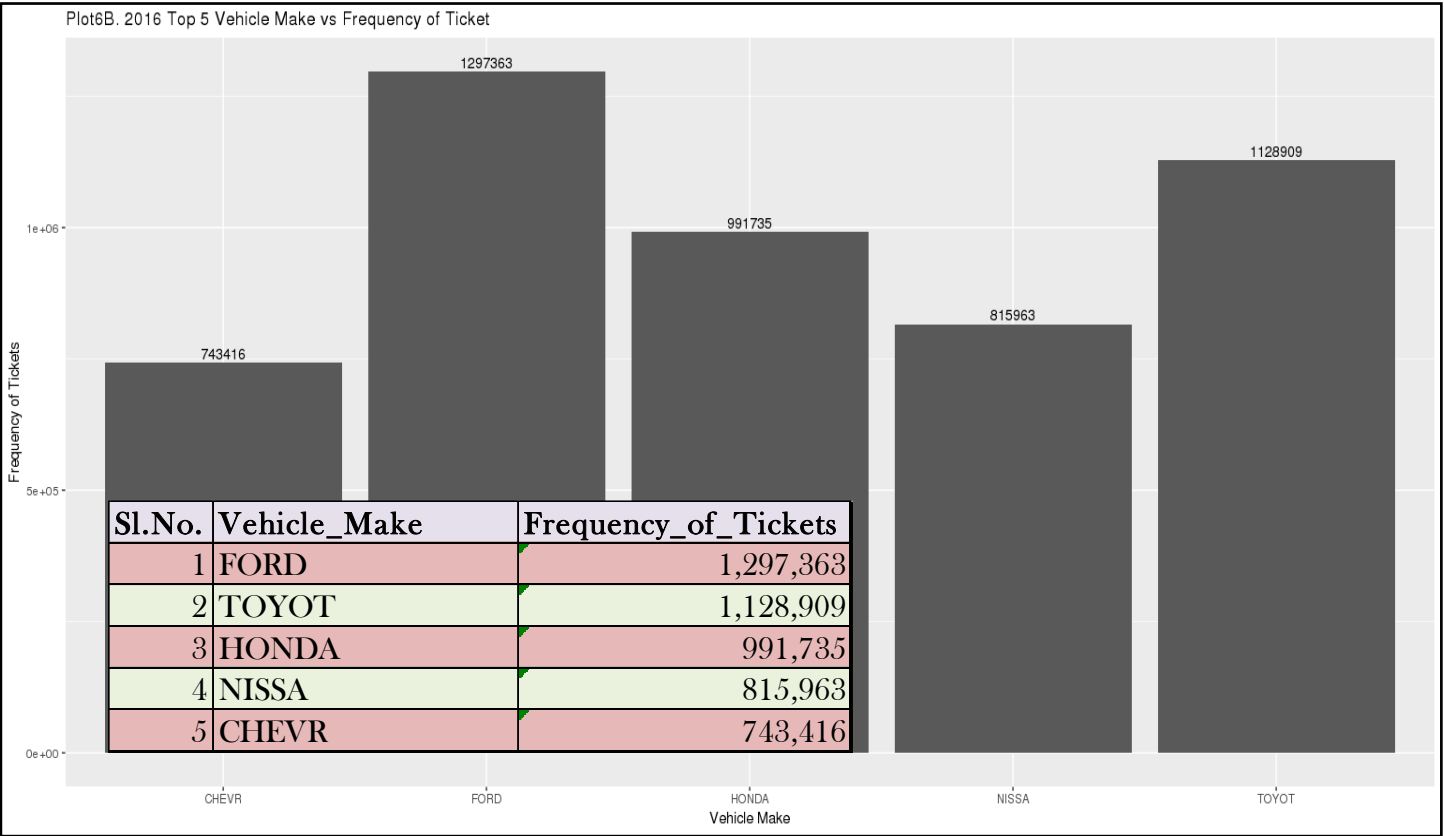
The SUBN body type results in the highest number of tickets issued throughout the years and the top 3 vehicle body types that are issued parking tickets [i.e., SUBN, 4DSN and VAN] contribute towards 74.6% of the total number of tickets issued in 2015. 75.2% of total tickets issued in 2016 and 76.2% of total tickets issued in 2017.

3.2.2.] Vehicle Make vs. Frequency of Ticket Analysis

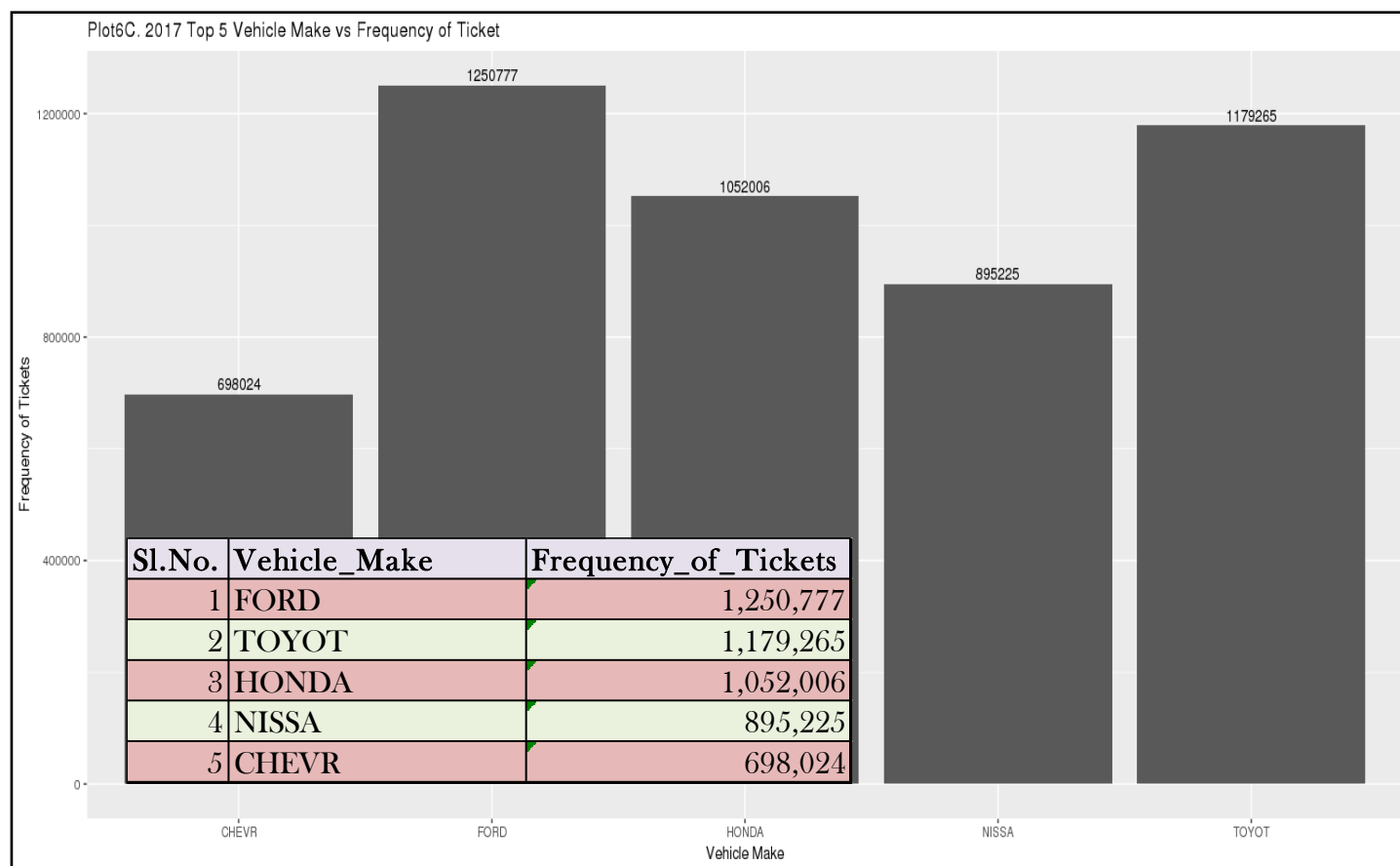
3.2.2.1] 2015 Dataset



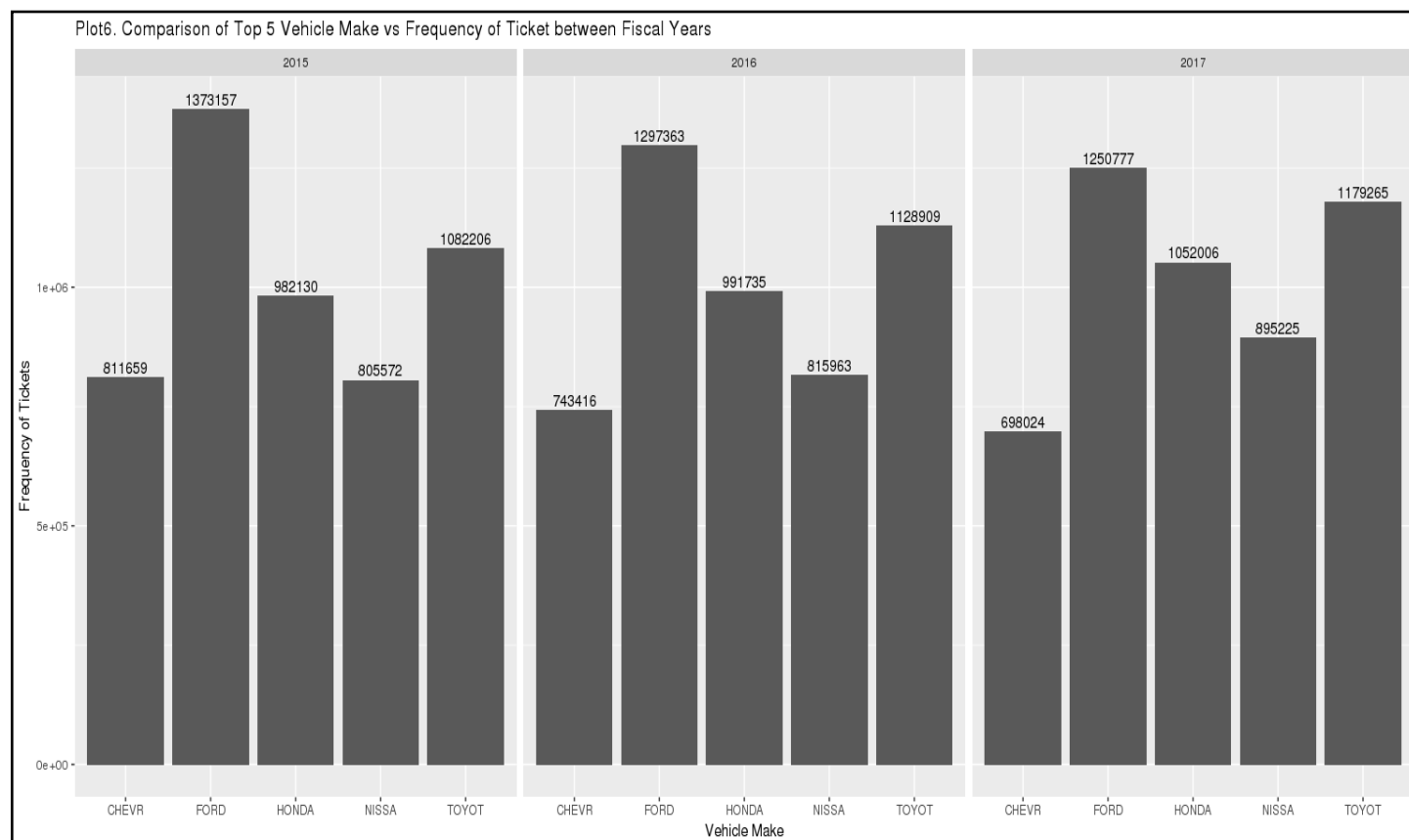
3.2.2.2] 2016 Dataset



[3.2.2.3] 2017 Dataset



[3.2.2.4] Combined Vehicle Body Make vs. Frequency of Parking Tickets



Insight: From the Vehicle Make vs. Frequency of Parking Tickets analysis it is clear that the same 5 vehicle makers feature in the top violations throughout the years. Ford vehicles contribute to the highest frequency of

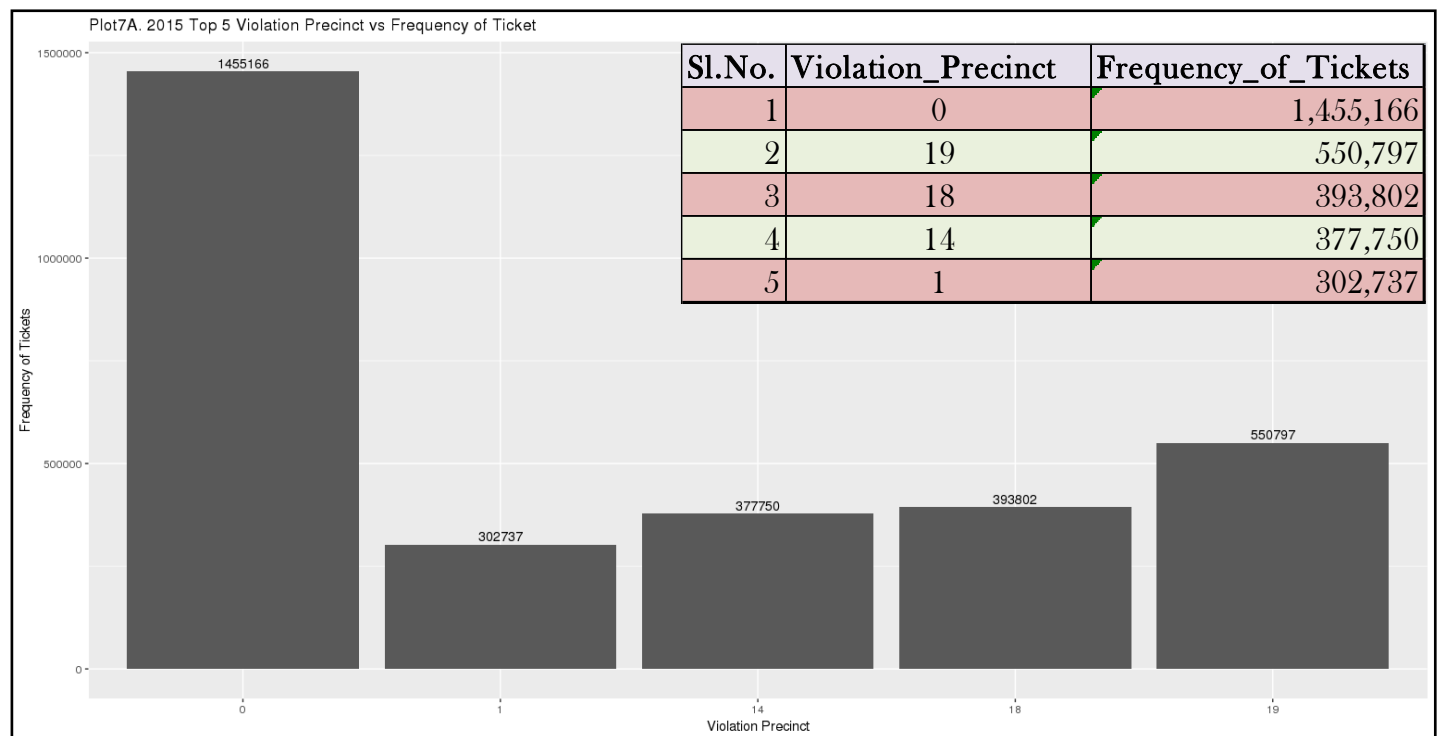
Parking Tickets. Whereas Ford, Toyota and Honda contribute towards 32.43% of total parking violations in 2015, 32.87% of total parking violations in 2016 and 33.04% of total parking violations in 2017. This may due to a higher number of vehicles from these makers owned by the public in NYC.

Stage 3: Question 3: A precinct is a police station that has a certain zone of the city under its command. Find the (5 highest) frequencies of:

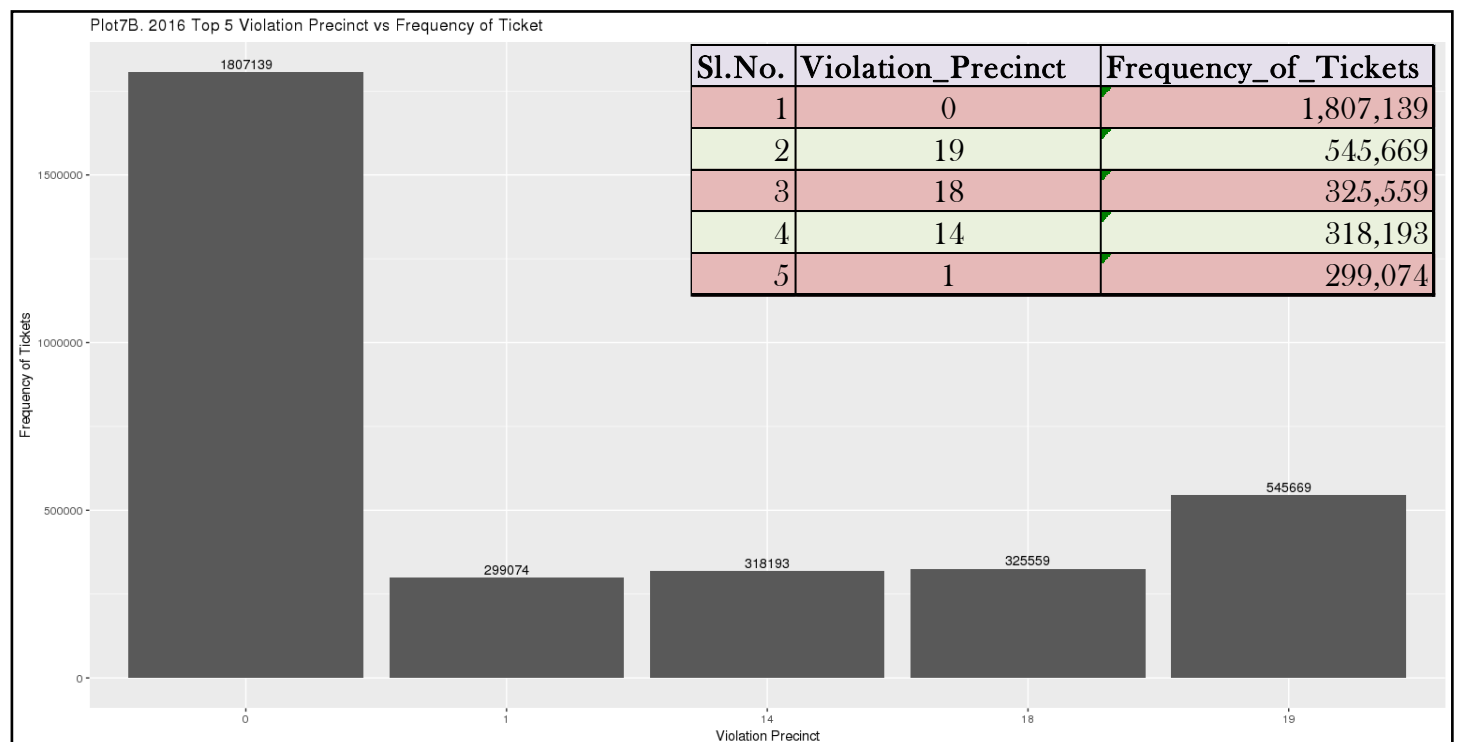
1. Violating Precincts (this is the precinct of the zone where the violation occurred)
2. Issuing Precincts (this is the precinct that issued the ticket)

[3.3.1.] Violating Precincts [Zone where Violation occurred] vs. Frequency of Ticket Analysis

[3.3.1.1] 2015 Dataset



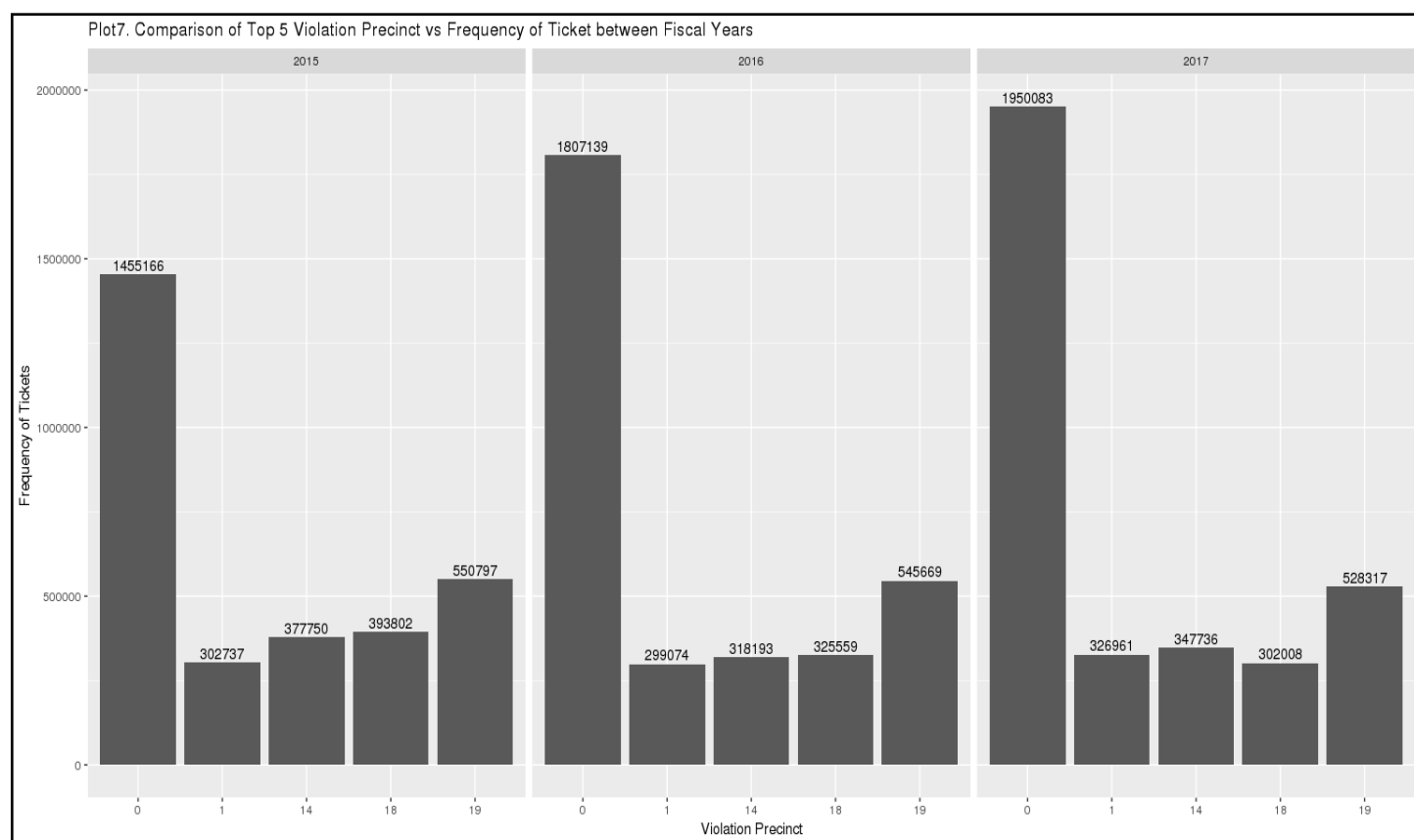
[3.3.1.2] 2016 Dataset



[3.3.1.3] 2017 Dataset



[3.3.1.4] Combined Violation Precinct vs. Frequency of Parking Tickets

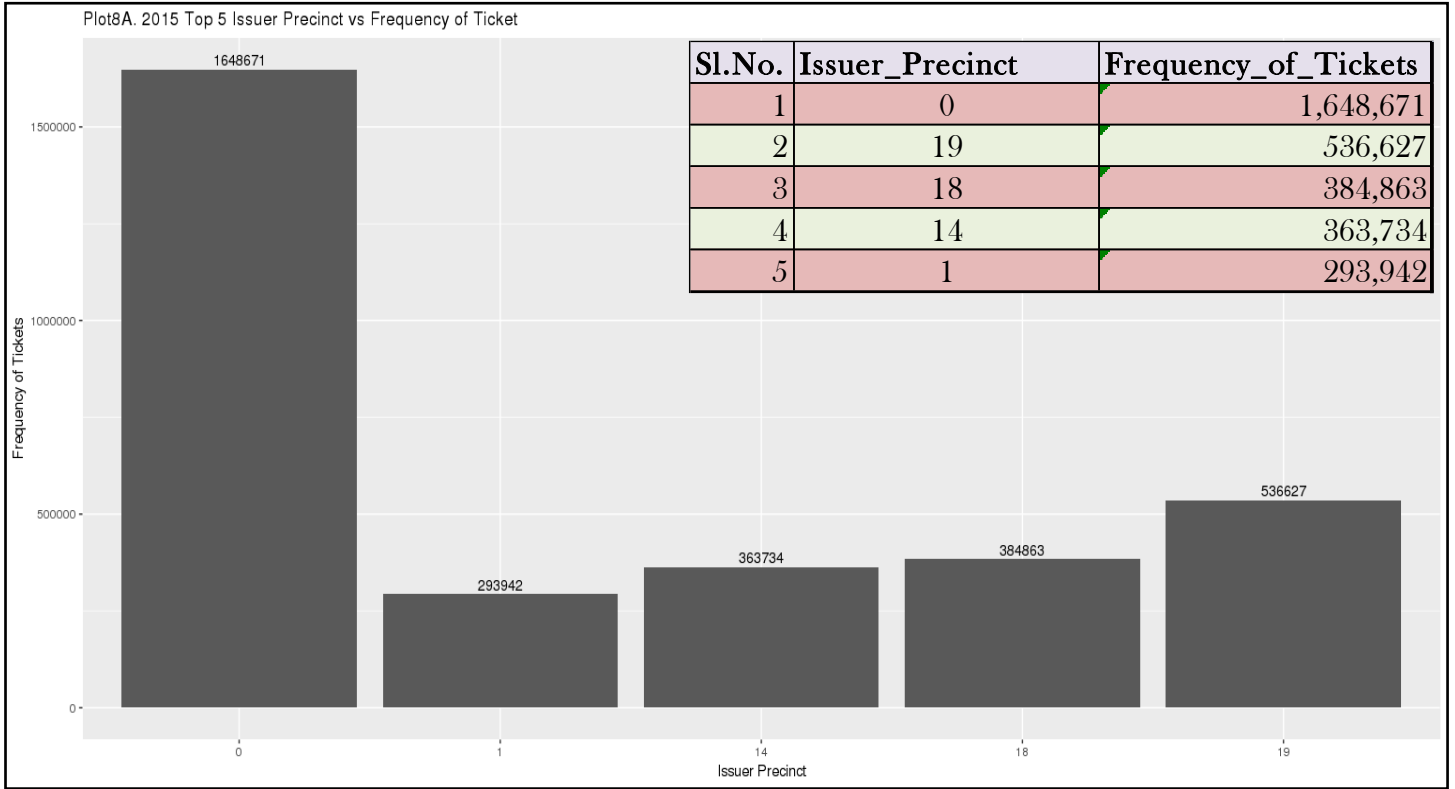


Insight: While analyzing the Violation Precinct vs. Frequency of Parking Ticket it is clear that the highest number of tickets in each year were issued with the Violation Precinct as 0. From the list of Precinct Codes and Zones it is evident that Precinct Code 0 does not exist, therefore this raises a red flag. These tickets may be issued with missing Violating Precinct Code or might be faulty tickets. Therefore, this is one area that must be inspected and

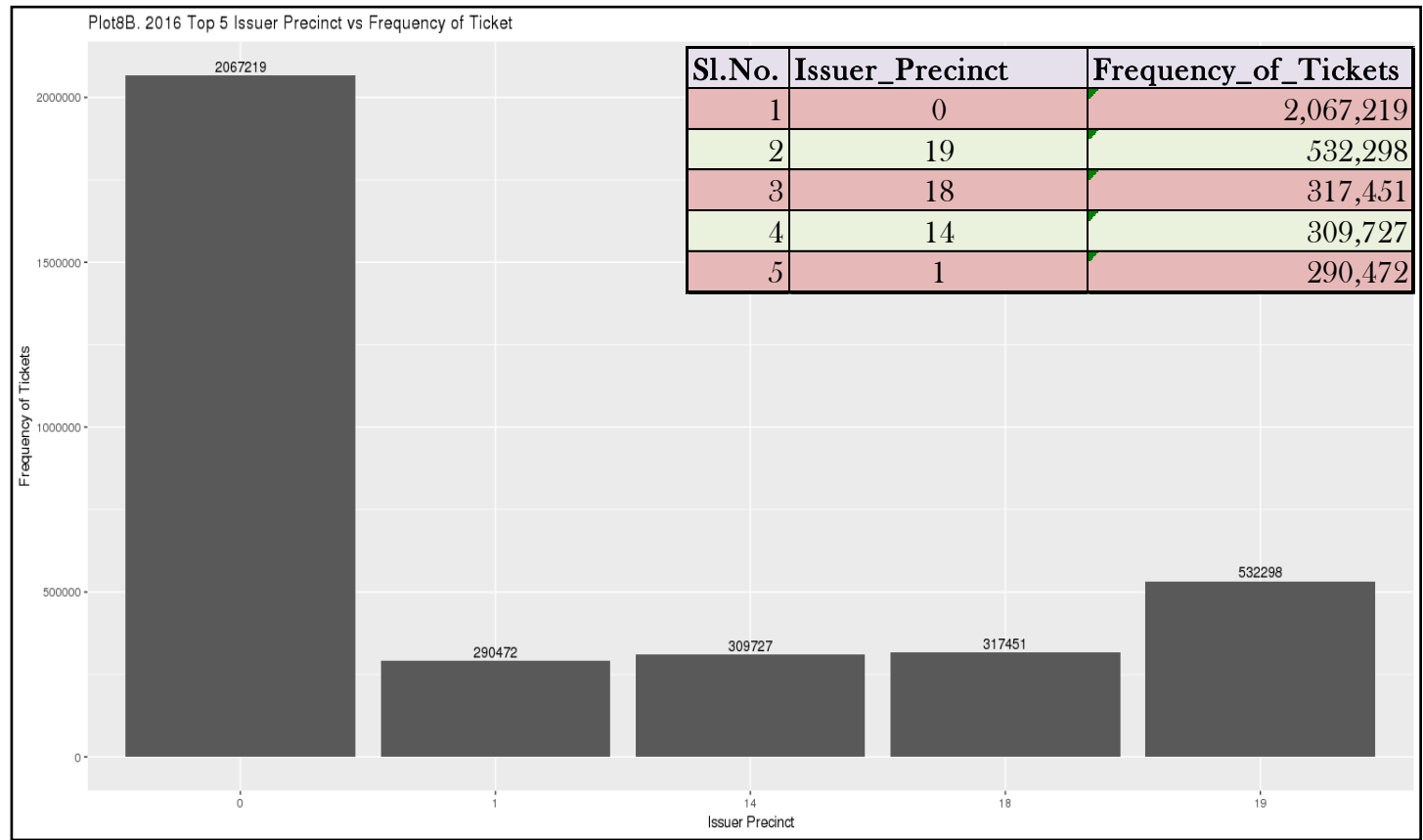
rectified. The other violation precincts 19, 18, 14 and 1 feature in all three years and contribute to 29.06% of total tickets issued in 2015, 31.69% of all tickets issued in 2016 and 32.78% of all tickets issued in 2017.

[3.3.2.] Issuing Precincts [Precinct issuing the Ticket] vs. Frequency of Ticket Analysis

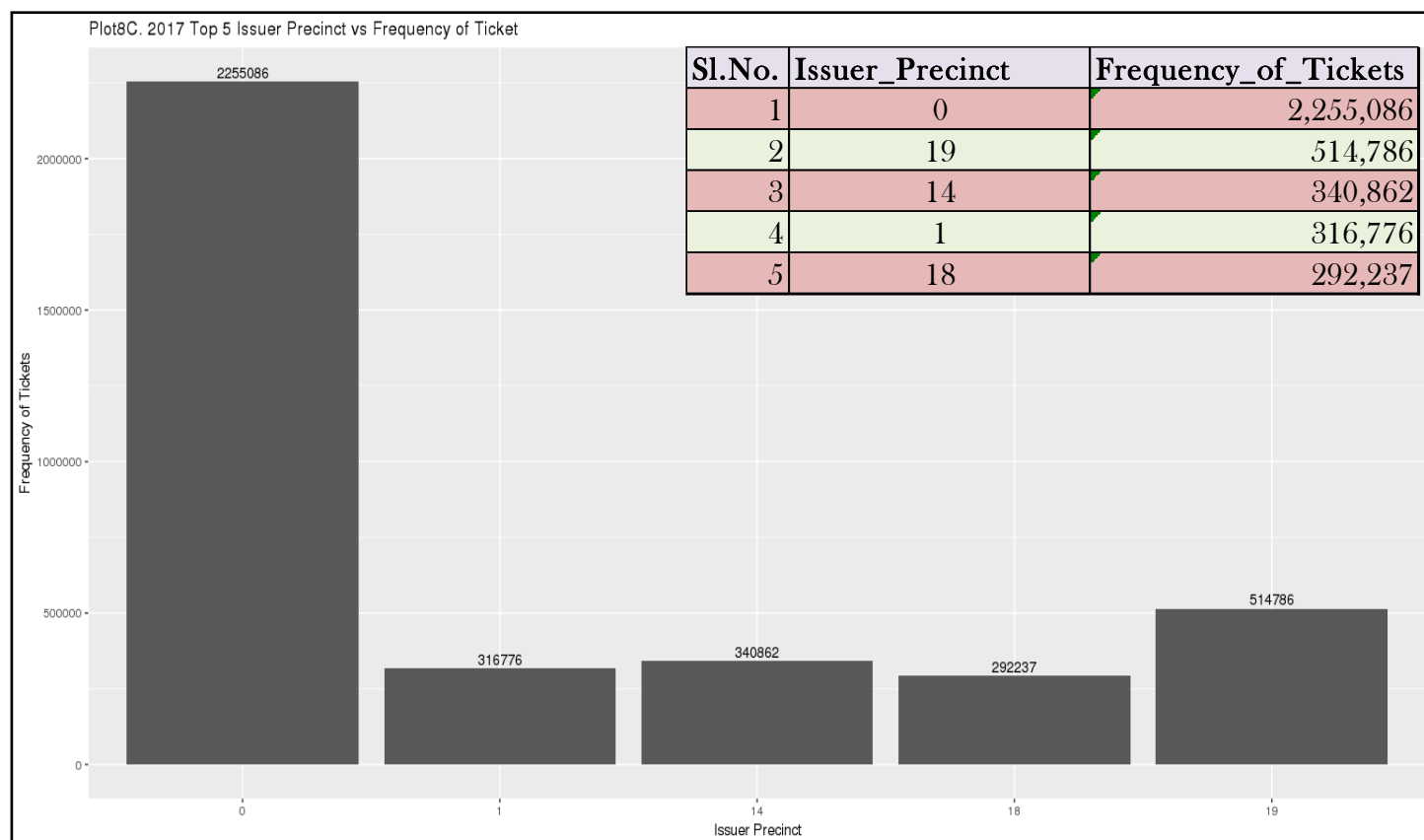
[3.3.2.1] 2015 Dataset



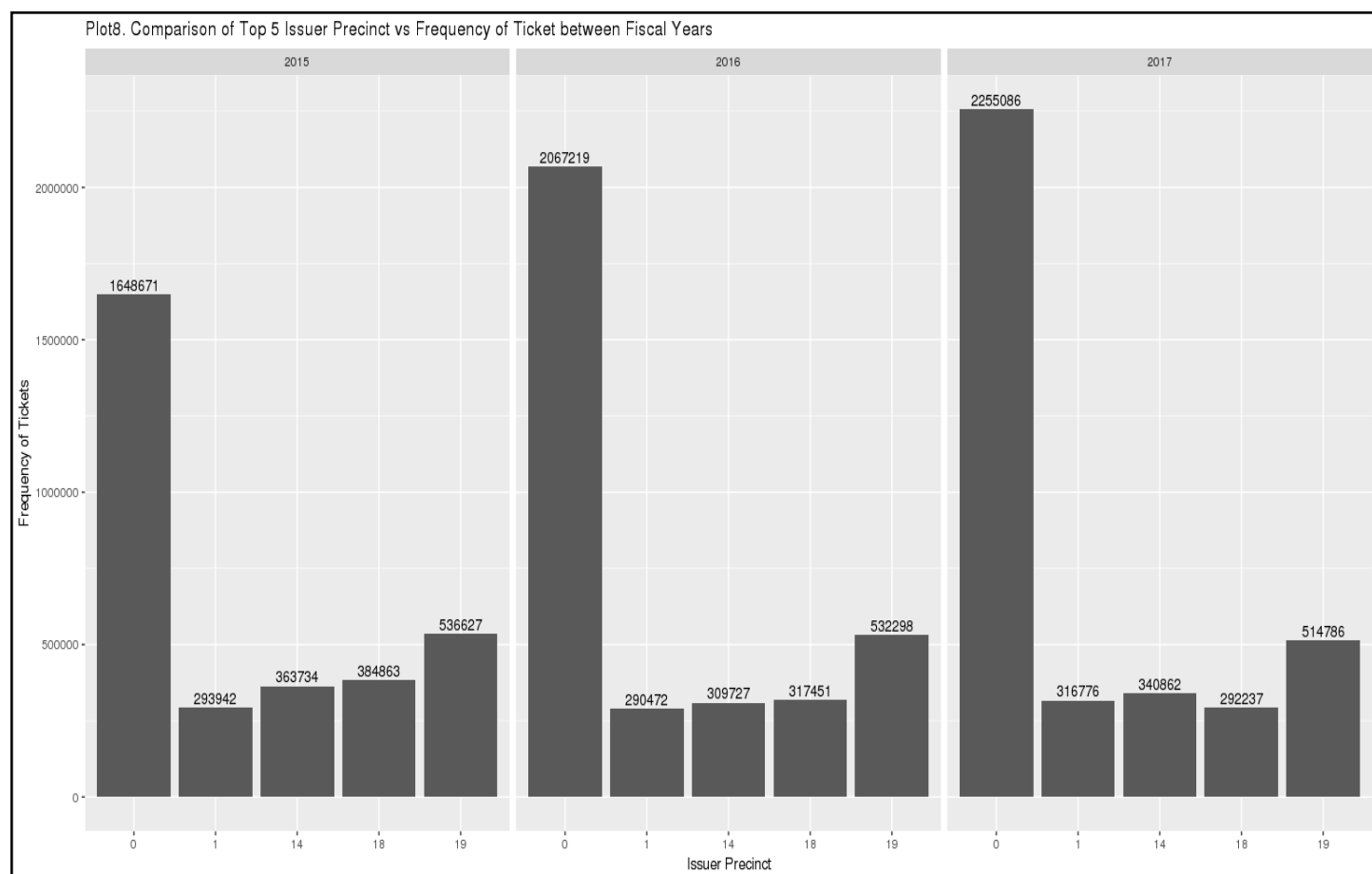
[3.3.2.2] 2016 Dataset



[3.3.2.3] 2017 Dataset



[3.3.2.4] Combined Issuer Precinct vs. Frequency of Parking Tickets



Insight: The Issuer Precinct vs. Frequency of Ticket Analysis revealed that the highest number of tickets from each year was issued with Issuer Precinct Code 0. Clearly this is peculiar as Precinct Code 0 does not exist and

therefore raises doubts regarding the validity of these tickets. The NYC Parking Ticket Authority needs to look into the root cause for this unusually high number of tickets with Issuer Precinct code 0. Another insight to be considered is that the number of tickets with Issuer Code 0 is almost equivalent to the next 4 highest Precinct codes combined. The other top Issuer Precinct codes that feature throughout the years are 19, 18, 14 and 1.

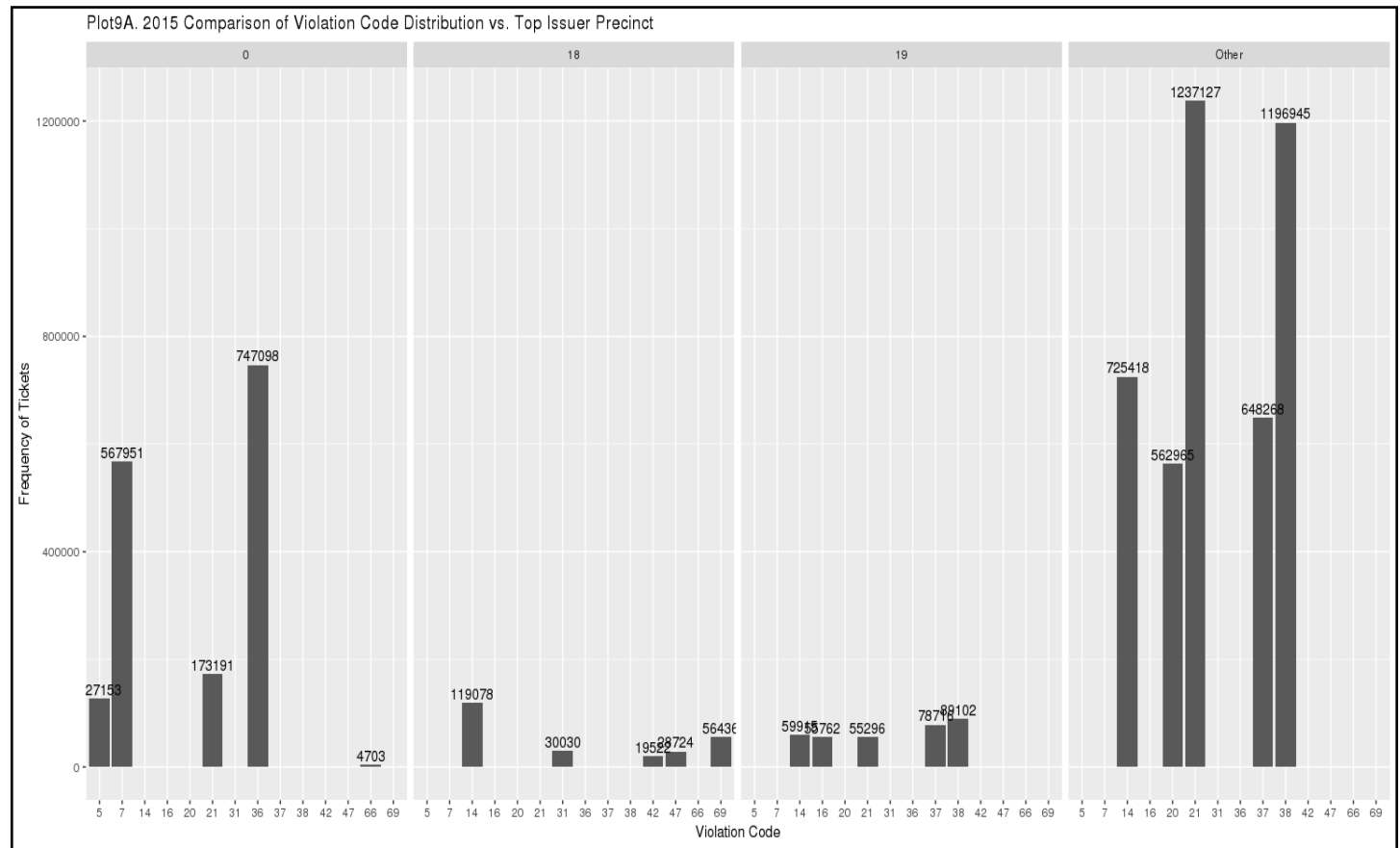
Stage 3: Question 4: Find the violation code frequency across 3 precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?

[3.4.1] 2015 Dataset

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

Let’s observe the Top-5 violation codes for which tickets were issued in each of the above 3 Issuing Precincts and compare the results with all other Issuing Precincts for the year 2015.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	747,098	0	1	38	89,102	19
2	7	567,951	0	2	37	78,716	19
3	21	173,191	0	3	14	59,915	19
4	5	127,153	0	4	16	55,762	19
5	66	4,703	0	5	21	55,296	19
Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	119,078	18	1	21	1,237,127	Other
2	69	56,436	18	2	38	1,196,945	Other
3	31	30,030	18	3	14	725,418	Other
4	47	28,724	18	4	37	648,268	Other
5	42	19,522	18	5	20	562,965	Other



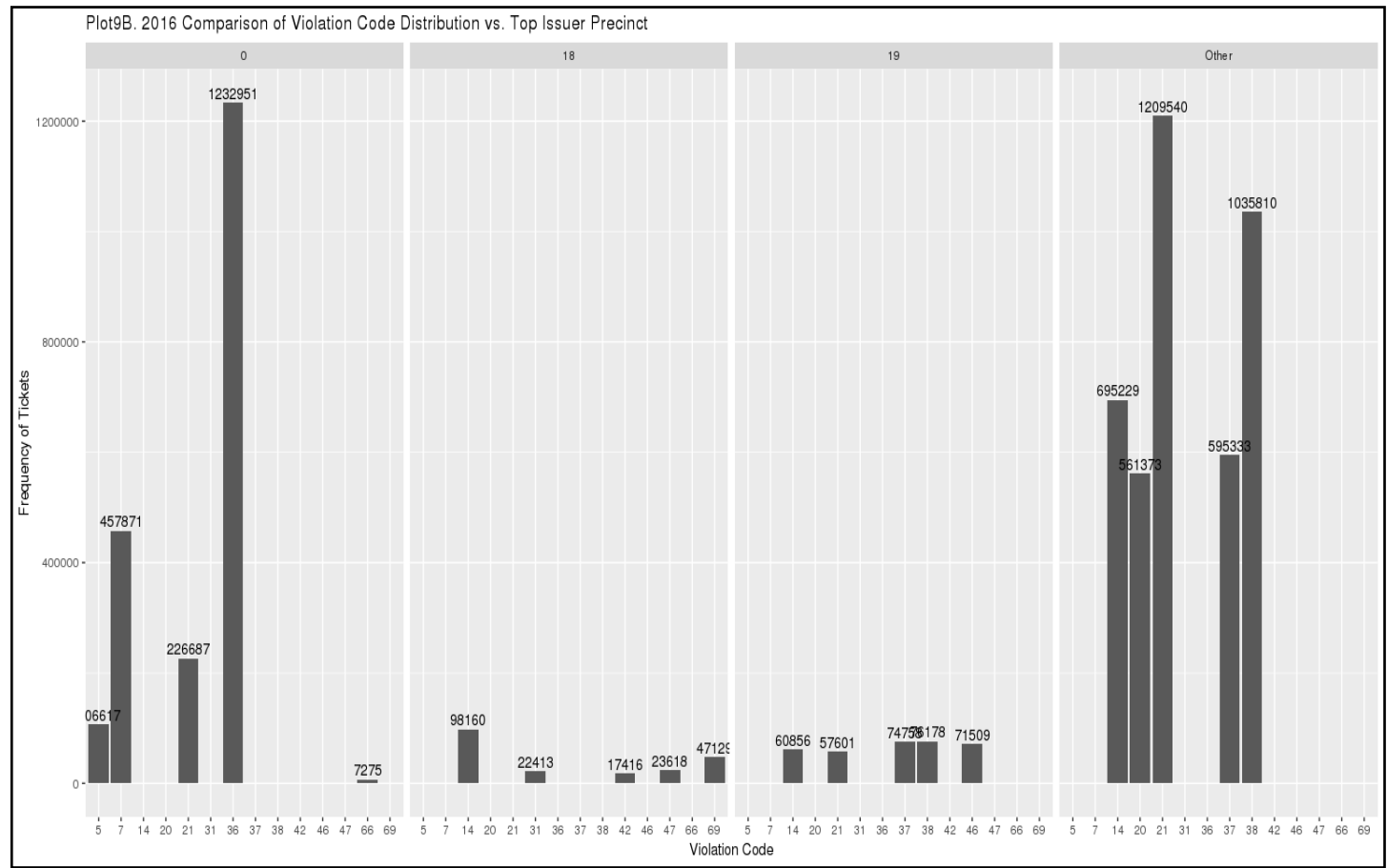
Insights: While observing the violation code distribution across Issuer Precincts for 2015 it is clear that an exceptionally high number of Tickets with Violation Code 36 and 7 are issued from Precinct Code 0 which is not consistent with the other Precincts. There is also a relatively low count [Inconsistent] of Violation Code 38 in the Top-3 Precincts when compared to the others. However, Violation Codes 14 and 21 are consistent throughout precincts.

[3.4.2] 2016 Dataset

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

Let’s observe the Top-5 violation codes for which tickets were issued in each of the above 3 Issuing Precincts and compare the results with all other Issuing Precincts for the year 2016.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	1,232,951	0	1	38	76,178	19
2	7	457,871	0	2	37	74,758	19
3	21	226,687	0	3	46	71,509	19
4	5	106,617	0	4	14	60,856	19
5	66	7,275	0	5	21	57,601	19
Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	98,160	18	1	21	1,209,540	Other
2	69	47,129	18	2	38	1,035,810	Other
3	47	23,618	18	3	14	695,229	Other
4	31	22,413	18	4	37	595,333	Other
5	42	17,416	18	5	20	561,373	Other



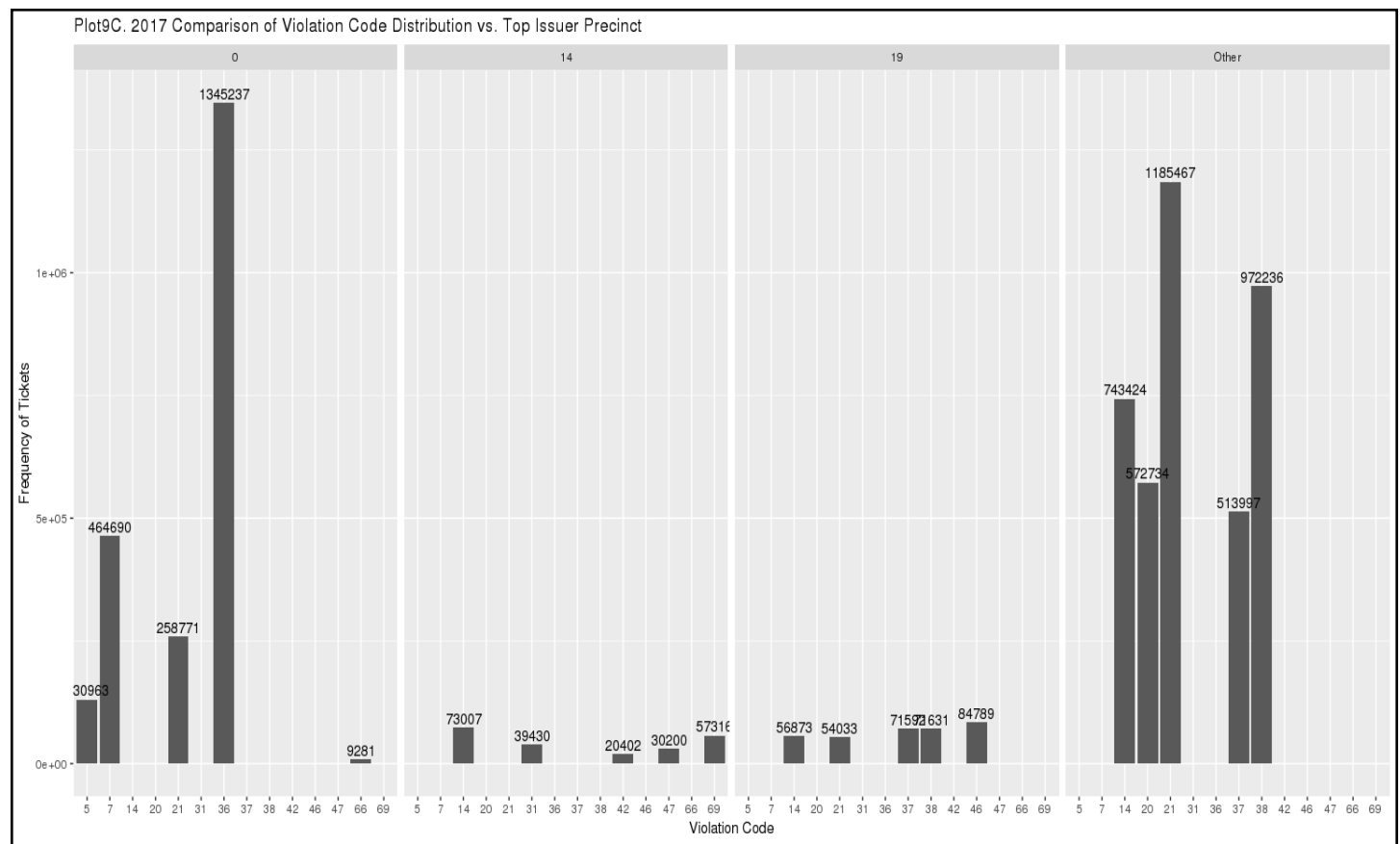
Insights: While observing the violation code distribution across Issuer Precincts for 2016 there is a clear inconsistency of an exceptionally high count of Tickets with Violation Code 36 and 7 that are issued from Precinct Code 0 which are not consistent in other Precincts. There is also a relatively low count [Inconsistent] of Violation Code 38 in the Top-3 Precincts when compared to the others.

[3.4.3] 2017 Dataset

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

Let’s observe the Top-5 violation codes for which tickets were issued in each of the above 3 Issuing Precincts and compare the results with all other Issuing Precincts for the year 2017.

Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	36	1,345,237	0	1	46	84,789	19
2	7	464,690	0	2	38	71,631	19
3	21	258,771	0	3	37	71,592	19
4	5	130,963	0	4	14	56,873	19
5	66	9,281	0	5	21	54,033	19
Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct	Sl. No.	Violation_Code	Frequency_of_Tickets	Issuer_Precinct
1	14	73,007	14	1	21	1,185,467	Other
2	69	57,316	14	2	38	972,236	Other
3	31	39,430	14	3	14	743,424	Other
4	47	30,200	14	4	20	572,734	Other
5	42	20,402	14	5	37	513,997	Other



Insight: 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 others.

Stage 3: Question 5: [a] You'd want to find out the properties of parking violations across different times of the day: 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.

The Violation Time has been transformed into a suitable timestamp format prior to Analysis.

[b] Find a way to deal with missing values, if any.

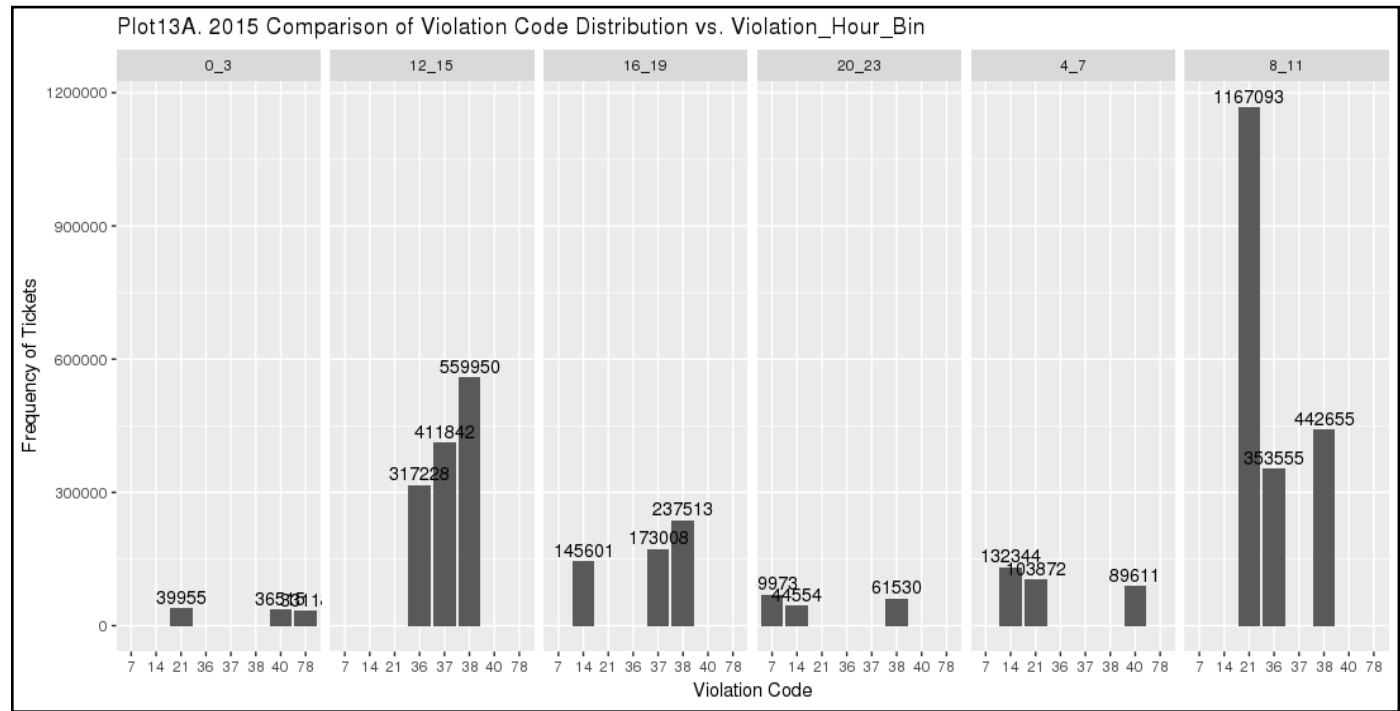
Fiscal Year	Frequency of Missing Violation Time	Total Number of Records	Percentage Missing
2015	61603	10598035	0.5812%
2016	63567	10396894	0.6114%
2017	58000	10539563	0.5503%

From this table it is evident that less than 0.6% of records in each year have missing Violation Time and is negligible. Therefore, we will drop records with missing violation time while conducting the analysis on VT.

[c - Part 1] 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.

[3.5.1.1] 2015 Dataset

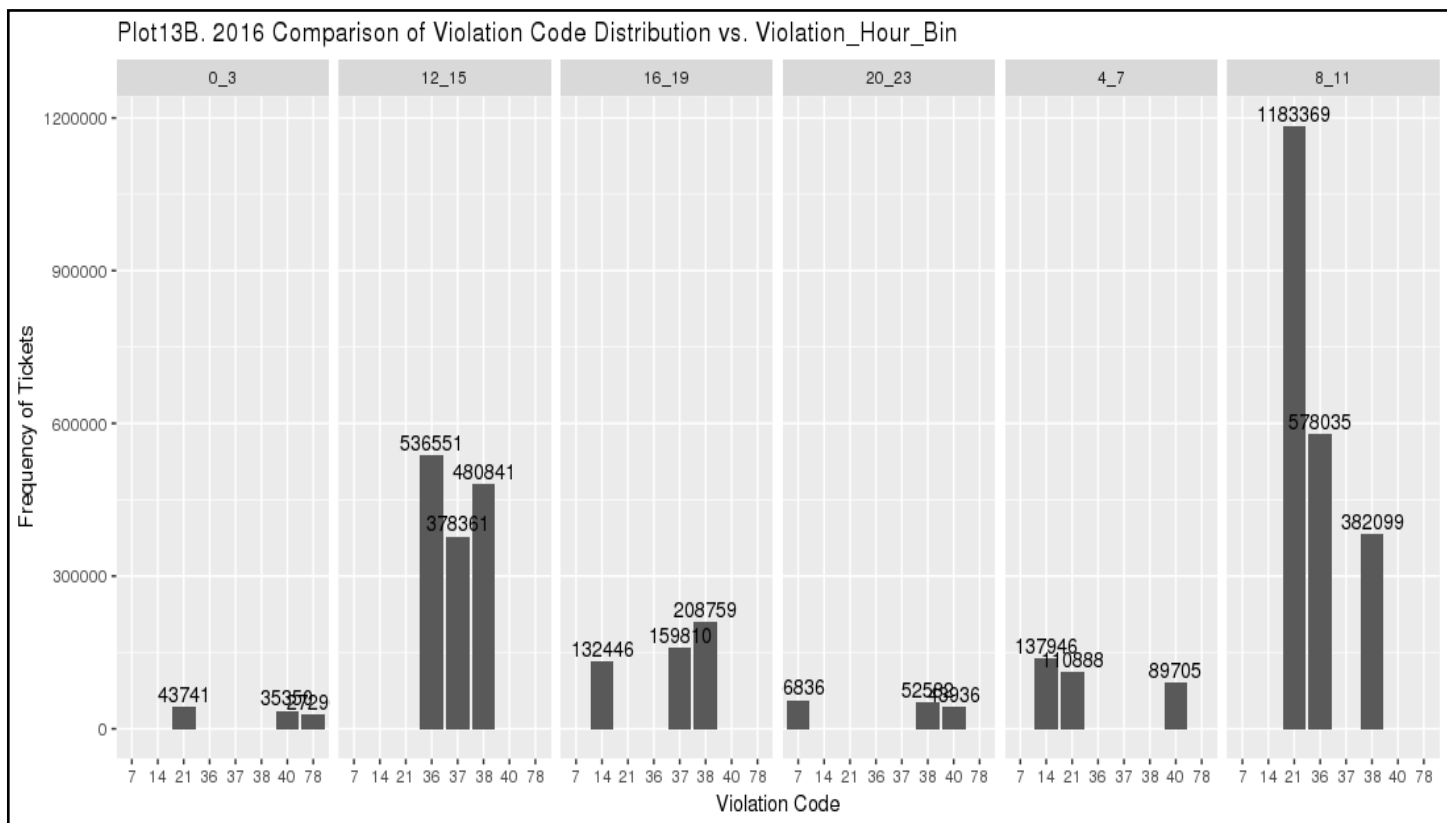
Sl. No.	Violation_Hour_Bin	Violation_Code	Frequency_of_Tickets
1	16_19	38	237513
2	16_19	37	173008
3	16_19	14	145601
4	8_11	21	1167093
5	8_11	38	442655
6	8_11	36	353555
7	4_7	14	132344
8	4_7	21	103872
9	4_7	40	89611
10	12_15	38	559950
11	12_15	37	411842
12	12_15	36	317228
13	0_3	21	39955
14	0_3	40	36515
15	0_3	78	33118
16	20_23	7	69973
17	20_23	38	61530
18	20_23	14	44554



Insight: While observing the Frequency of Top-3 Violation Codes in each Violation Time Bin it is clear that the majority of the Tickets are issued between 0800-1100 Hrs and 1200-1500 Hrs. It is also important to note the exceptionally high number of tickets for Violation Code 21 [This is expected as Code 21 stands for No-Parking Zone Tickets. Majority of the public might park inappropriately during the morning rush] issued between 0800-1100 Hrs. There is also a high frequency of tickets for Violation Code 36 and 38 between 0800-1500 Hrs. [This is also expected as Code 36 is due to exceeding speed limit near school zones, there are spikes between 0800-1500 Hrs during start and end of school day while Codes 37&38 are due to an expired parking meter between 1200-1500 Hrs. and 1600-1900 Hrs.]

[3.5.1.2] 2016 Dataset

Sl. No.	Violation_Hour_Bin	Violation_Code	Frequency_of_Tickets
1	16_19	38	208759
2	16_19	37	159810
3	16_19	14	132446
4	8_11	21	1183369
5	8_11	36	578035
6	8_11	38	382099
7	4_7	14	137946
8	4_7	21	110888
9	4_7	40	89705
10	12_15	36	536551
11	12_15	38	480841
12	12_15	37	378361
13	0_3	21	43741
14	0_3	40	35350
15	0_3	78	27296
16	20_23	7	56836
17	20_23	38	52582
18	20_23	40	43936

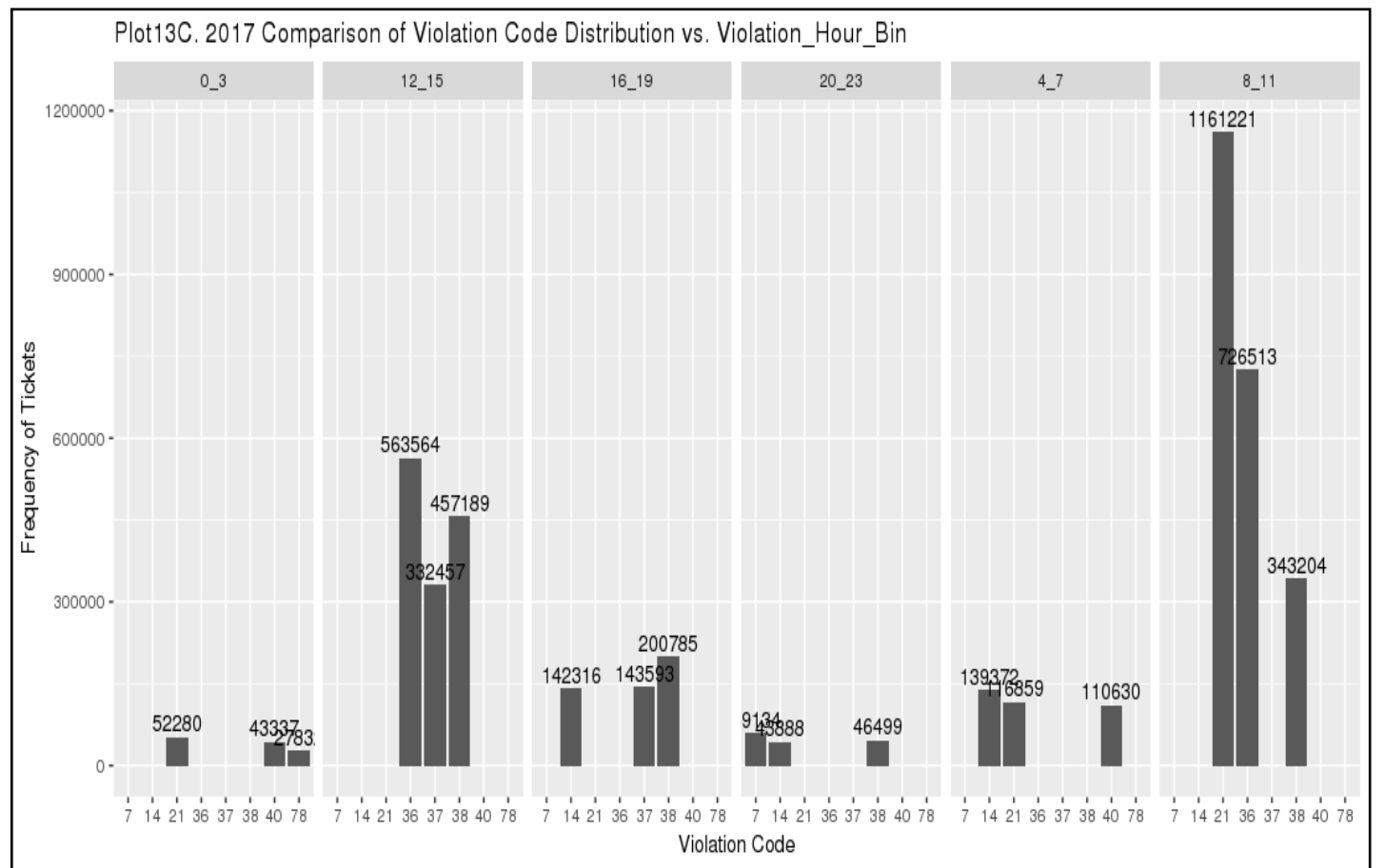


Insight: While observing the Frequency of Top-3 Violation Codes in each Violation Time Bin it is clear that the majority of the Tickets are issued between 0800-1100 Hrs and 1200-1500 Hrs. It is also important to note

the exceptionally high number of tickets for Violation Code 21 [This is expected as Code 21 stands for No-Parking Zone Tickets. Majority of the public might park inappropriately during the morning rush] issued between 0800-1100 Hrs. There is also a high frequency of tickets for Violation Code 36 and 38 between 0800-1500 Hrs. [This is also expected as Code 36 is due to exceeding speed limit near school zones, there are spikes between 0800-1500 Hrs during start and end of school day while Codes 37&38 are due to an expired parking meter between 1200-1500 Hrs. and 1600-1900 Hrs.]

[3.5.1.3] 2017 Dataset

Sl. No.	Violation_Hour_Bin	Violation_Code	Frequency_of_Tickets
1	16_19	38	200785
2	16_19	37	143593
3	16_19	14	142316
4	8_11	21	1161221
5	8_11	36	726513
6	8_11	38	343204
7	4_7	14	139372
8	4_7	21	116859
9	4_7	40	110630
10	12_15	36	563564
11	12_15	38	457189
12	12_15	37	332457
13	0_3	21	52280
14	0_3	40	43337
15	0_3	78	27832
16	20_23	7	59134
17	20_23	38	46499
18	20_23	14	43888

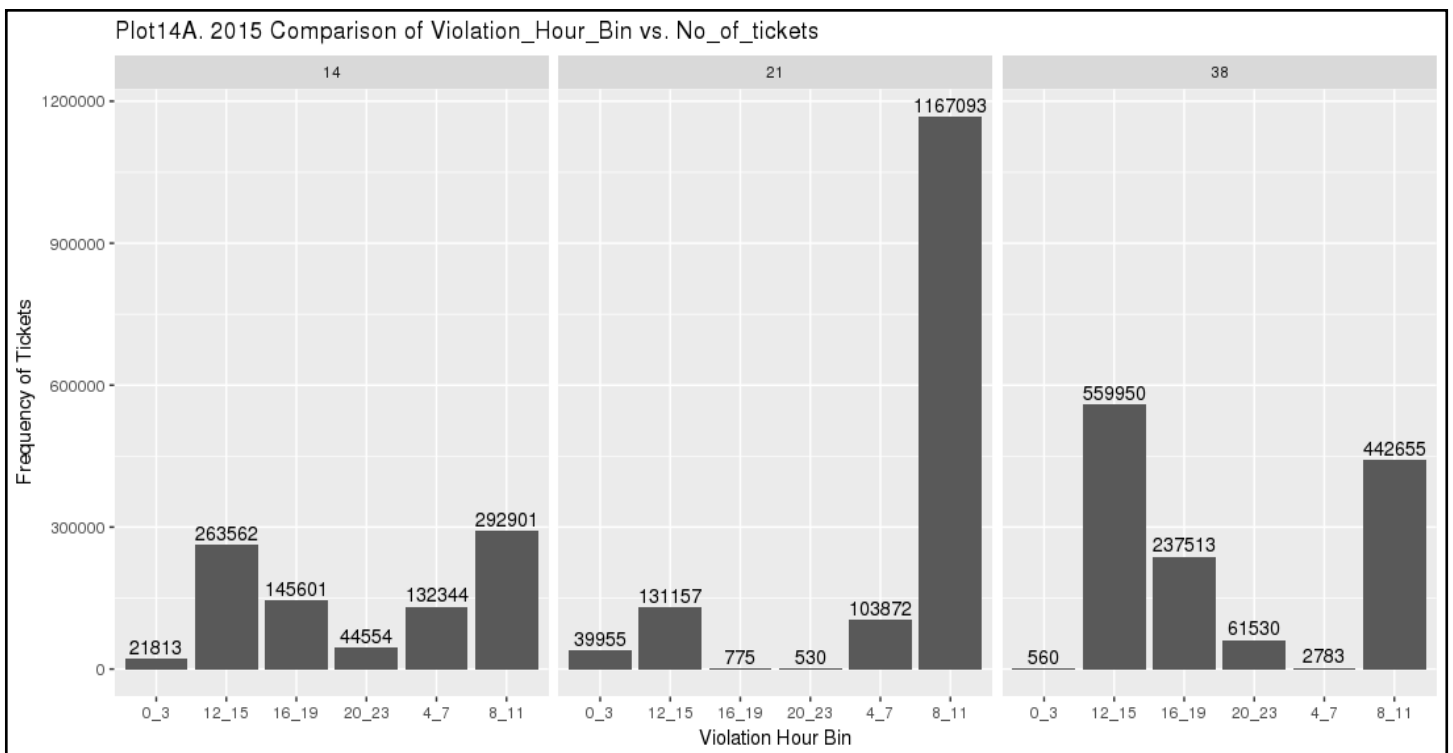


Insight: Insight: While observing the Frequency of Top-3 Violation Codes in each Violation Time Bin it is clear that the majority of the Tickets are issued between 0800-1100 Hrs and 1200-1500 Hrs. It is also important to note the exceptionally high number of tickets for Violation Code 21 [This is expected as Code 21 stands for No-Parking Zone Tickets. Majority of the public might park inappropriately during the morning rush] issued between 0800-1100 Hrs. There is also a high frequency of tickets for Violation Code 36 and 38 between 0800-1500 Hrs. [This is also expected as Code 36 is due to exceeding speed limit near school zones, there are spikes between 0800-1500 Hrs during start and end of school day while Codes 37&38 are due to an expired parking meter between 1200-1500 Hrs. and 1600-1900 Hrs.]

[d - Part 2] 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)

3.5.2.1| 2015 Dataset

Sl. No.	Violation_Code	Violation_Hour_Bin	No_of_tickets
1	14	0_3	21813
2	14	12_15	263562
3	14	16_19	145601
4	14	20_23	44554
5	14	4_7	132344
6	14	8_11	292901
7	21	0_3	39955
8	21	12_15	131157
9	21	16_19	775
10	21	20_23	530
11	21	4_7	103872
12	21	8_11	1167093
13	38	0_3	560
14	38	12_15	559950
15	38	16_19	237513
16	38	20_23	61530
17	38	4_7	2783
18	38	8_11	442655

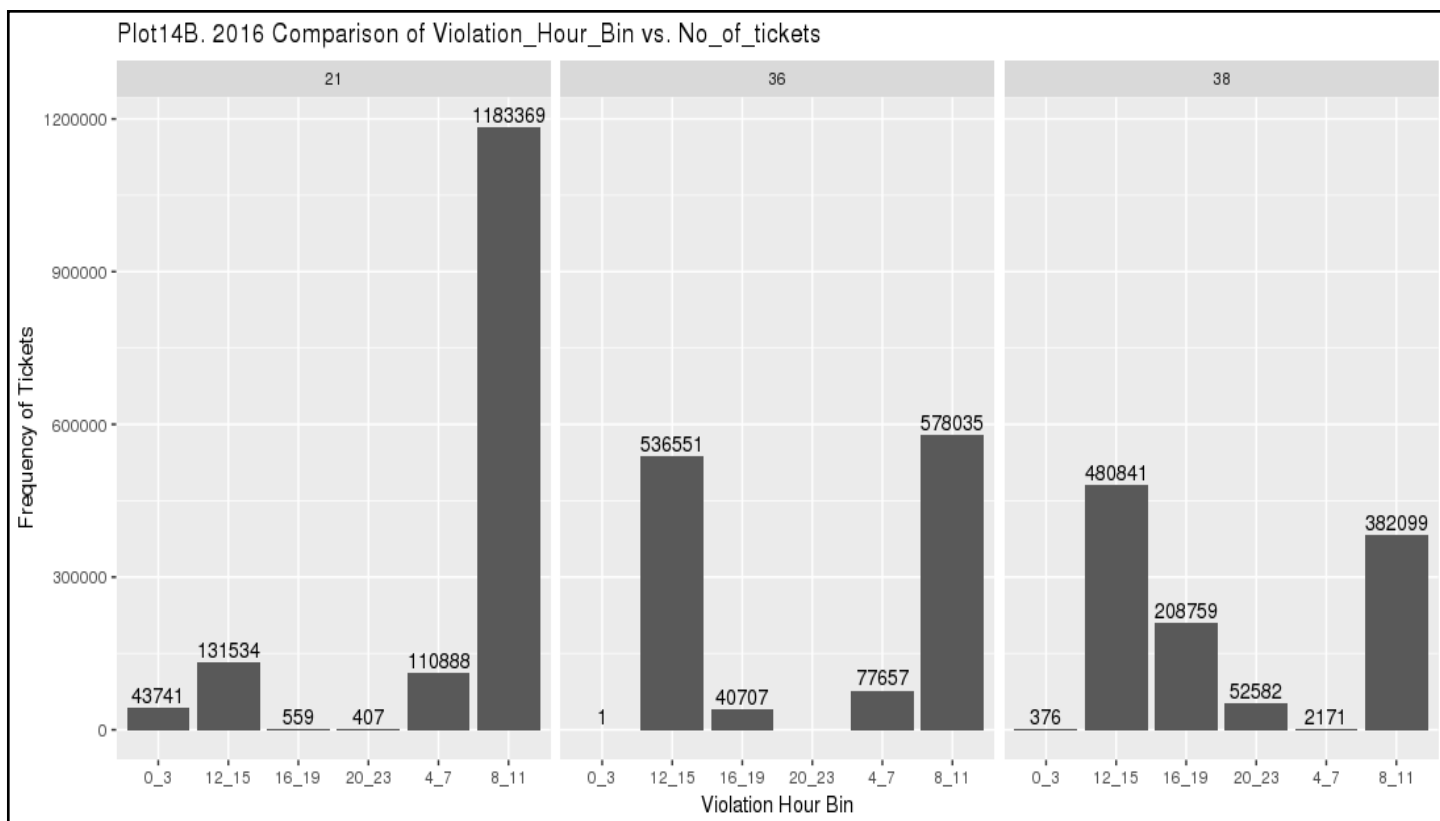


Insight: The Top-3 Violation Codes for 2015 are 14: No Standing Violation, 21: No Parking Violation, 38: Parking Muni Meter Expired.

While looking at the distribution of Violation Code 14, it looks fairly even with peaks between 0800-1100 Hrs and 1200-1500 Hrs and dips between 2000-2300 Hrs and 00-0300 Hrs. For Violation Code 21, there is noticeably a significant peak between 0800-1100 Hrs and dips between 1600-2300 Hrs. For Violation Code 38 there are peaks between 0800-1500 Hrs and dips between 00-0700 Hrs.

[3.5.2.2] 2016 Dataset

Sl. No.	Violation_Code	Violation_Hour_Bin	no_of_tickets
1	21	0_3	43741
2	21	12_15	131534
3	21	16_19	559
4	21	20_23	407
5	21	4_7	110888
6	21	8_11	1183369
7	36	0_3	1
8	36	12_15	536551
9	36	16_19	40707
10	36	4_7	77657
11	36	8_11	578035
12	38	0_3	376
13	38	12_15	480841
14	38	16_19	208759
15	38	20_23	52582
16	38	4_7	2171
17	38	8_11	382099



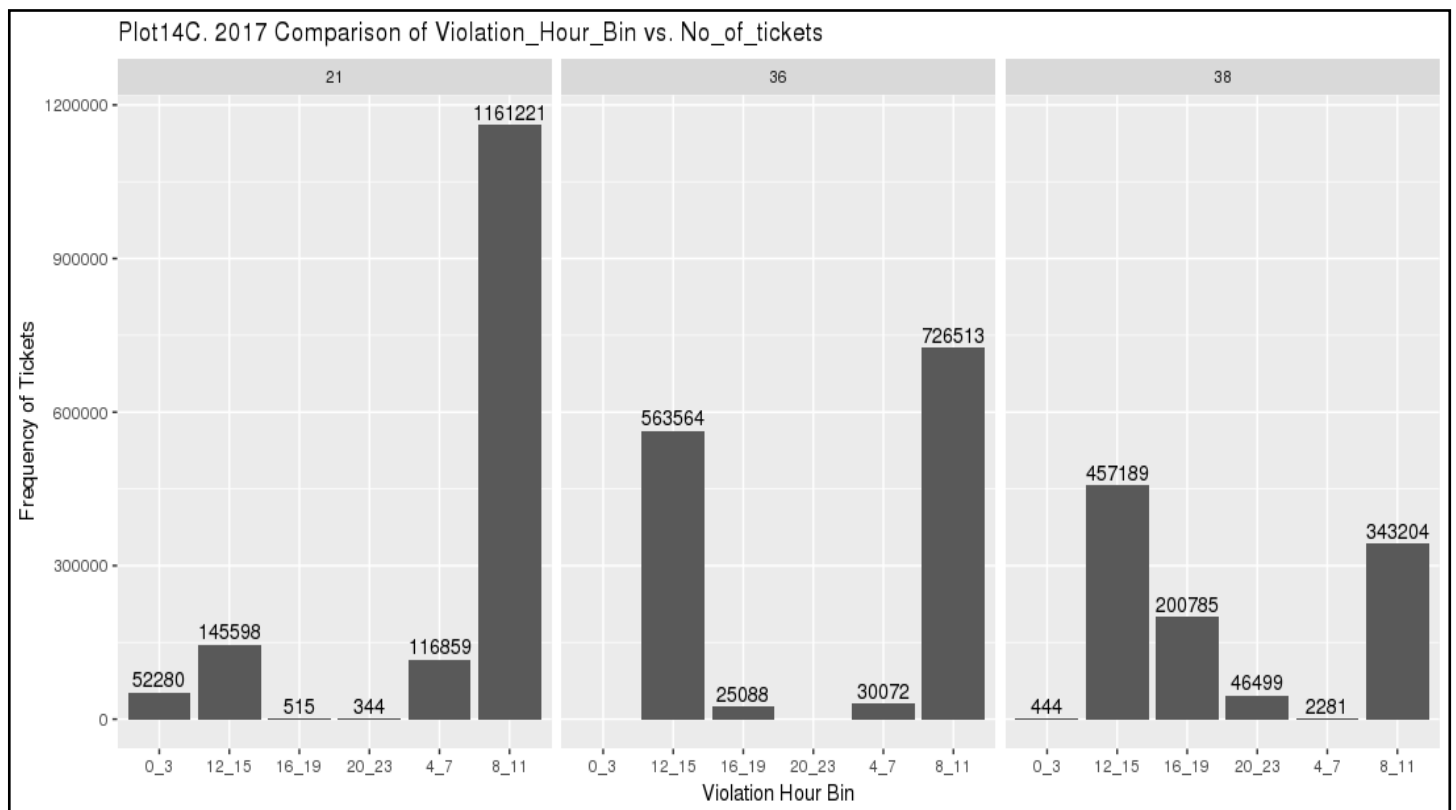
Insight: The Top-3 Violation Codes for 2016 are 21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired.

While looking at the distribution of Violation Code 21, there is noticeably a significant peak between 0800-1100 Hrs and dips between 1600-2300 Hrs. For Violation Code 36 there are almost equal peaks between 0800-1100

Hrs and 1200-1500 Hrs this may be as a result of people rushing to drop and pickup children to and from school. Note: There are almost no instances of Violation Code 36 between 2000-0300 Hrs. For Violation Code 38 there are peaks between 0800-1500 Hrs and dips between 00-0700 Hrs.

[3.5.2.3] 2017 Dataset

Sl. No.	Violation_Code	Violation_Hour_Bin	No_of_tickets
1	21	0_3	52280
2	21	12_15	145598
3	21	16_19	515
4	21	20_23	344
5	21	4_7	116859
6	21	8_11	1161221
7	36	12_15	563564
8	36	16_19	25088
9	36	4_7	30072
10	36	8_11	726513
11	38	0_3	444
12	38	12_15	457189
13	38	16_19	200785
14	38	20_23	46499
15	38	4_7	2281
16	38	8_11	343204



Insight: The Top-3 Violation Codes for 2016 are 21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired.

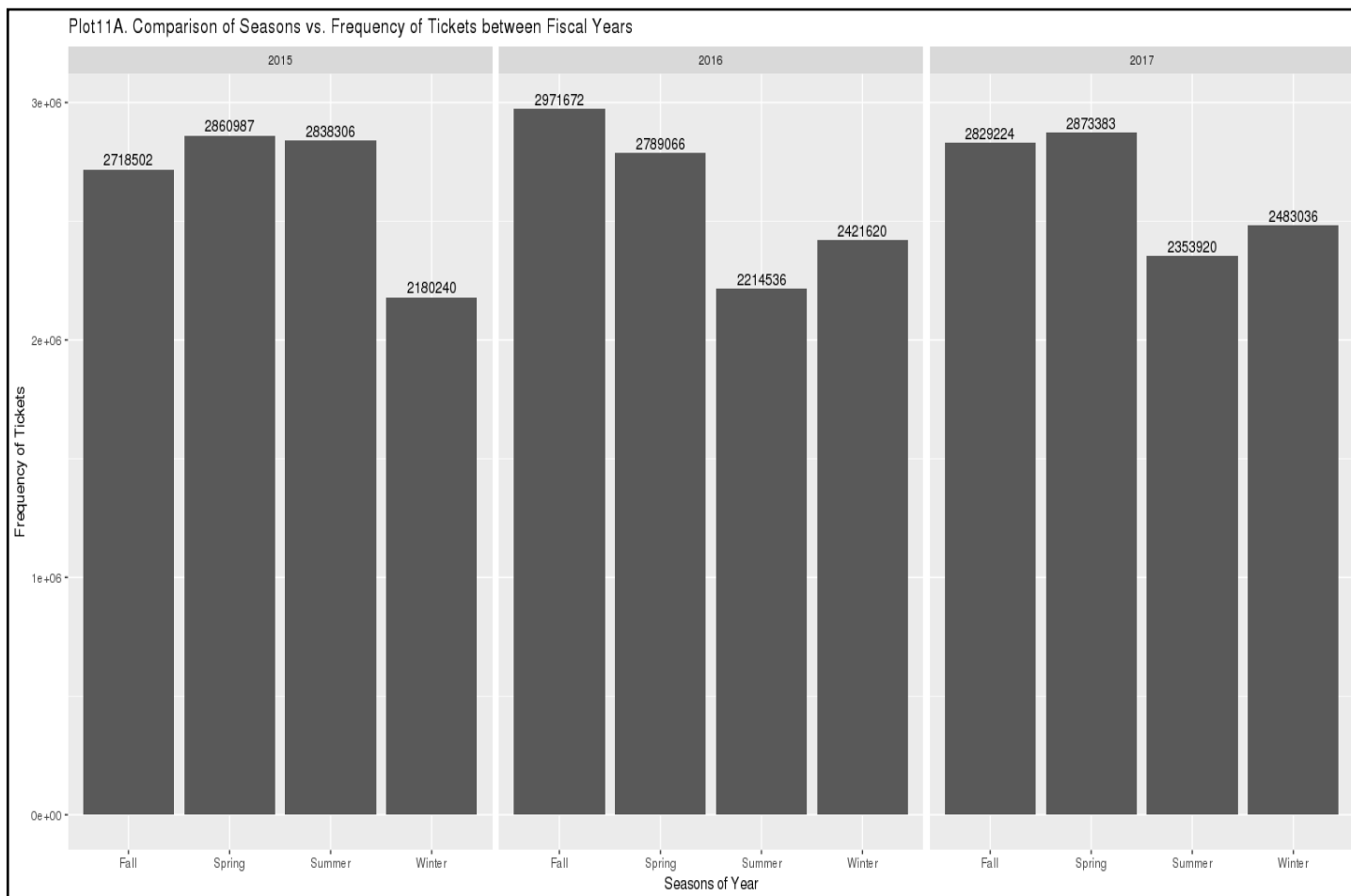
While looking at the distribution of Violation Code 21, there is noticeably a significant peak between 0800-1100 Hrs and dips between 1600-2300 Hrs. For Violation Code 36 there are almost equal peaks between 0800-1100 Hrs and 1200-1500 Hrs this may be as a result of people rushing to drop and pickup children to and from

school. Note: There are no instances of Violation Code 36 between 2000-0300 Hrs. For Violation Code 38 there are peaks between 0800-1500 Hrs and dips between 00-0700 Hrs.

Stage 3: Question 6: Let's try and find some seasonality in this data

[Part-1] First, divide the year into some number of seasons, and find frequencies of tickets for each season.

Sl.No.	Season	Frequency_of_Tickets	Fiscal_Year	Sl.No.	Season	Frequency_of_Tickets	Fiscal_Year
1	Spring	2,860,987	2015	1	Fall	2,971,672	2016
2	Summer	2,838,306	2015	2	Spring	2,789,066	2016
3	Fall	2,718,502	2015	3	Winter	2,421,620	2016
4	Winter	2,180,240	2015	4	Summer	2,214,536	2016
Sl.No.	Season	Frequency_of_Tickets	Fiscal_Year				
1	Spring	2,873,383	2017				
2	Fall	2,829,224	2017				
3	Winter	2,483,036	2017				
4	Summer	2,353,920	2017				

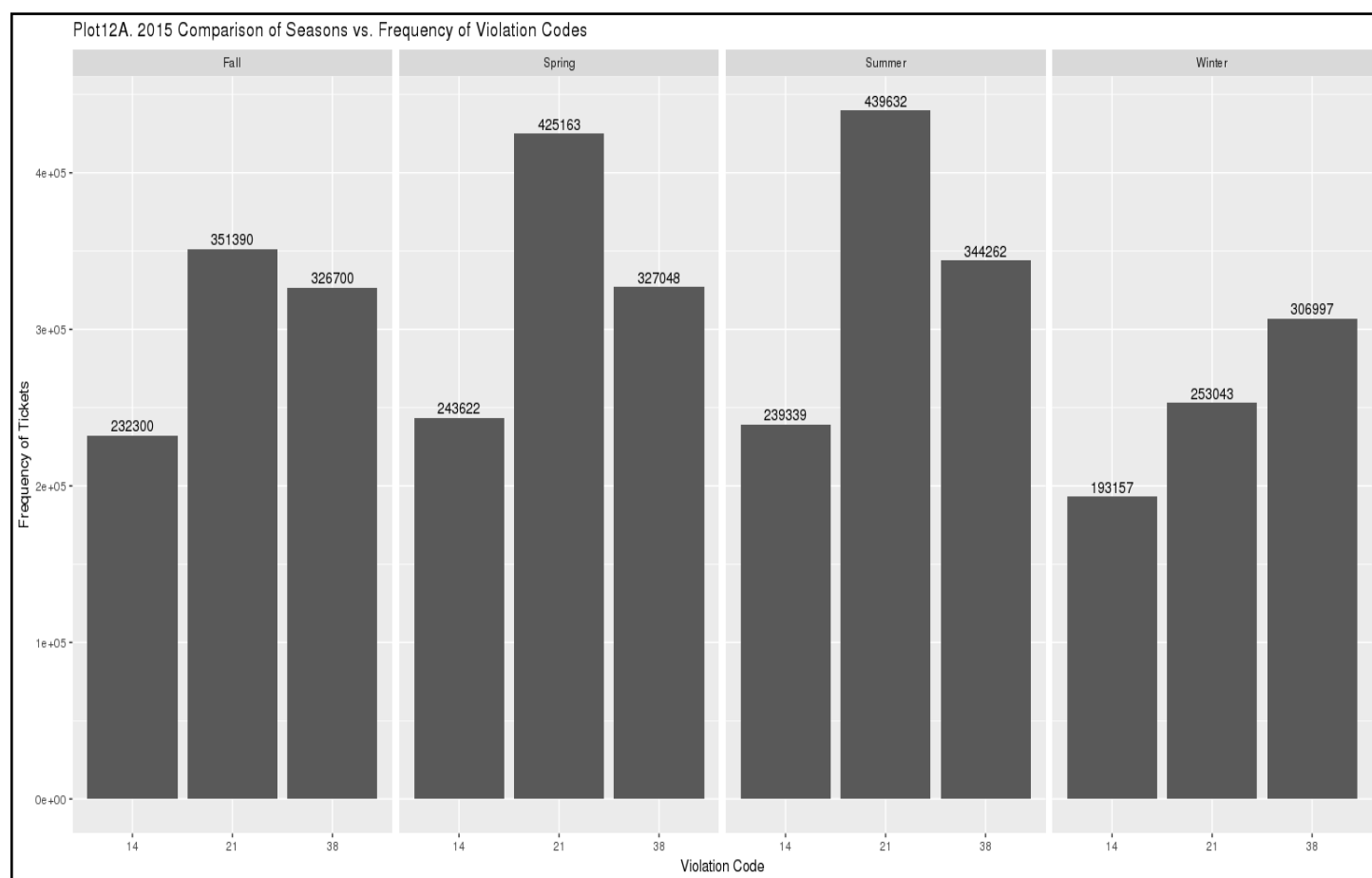


Insight: For the year 2015, Spring and Summer account for the highest number of parking tickets and comprises 53.77% of the total tickets issued in 2015. However, for the years 2016 and 2017, Fall and Spring account for the highest number of parking tickets and comprises of 55.41% of the total tickets issued in 2016 and 54.11% of all tickets issued in 2017.

[Part-2] Then, find the 3 most common violations for each of these seasons.

[3.6.1] 2015 Dataset [Top- 3 Violation Codes 21, 38 and 14]

Sl.No.	Season	Violation_Code	Frequency_of_Tickets
1	Fall	21	351390
2	Fall	38	326700
3	Fall	14	232300
4	Spring	21	425163
5	Spring	38	327048
6	Spring	14	243622
7	Summer	21	439632
8	Summer	38	344262
9	Summer	14	239339
10	Winter	38	306997
11	Winter	21	253043
12	Winter	14	193157

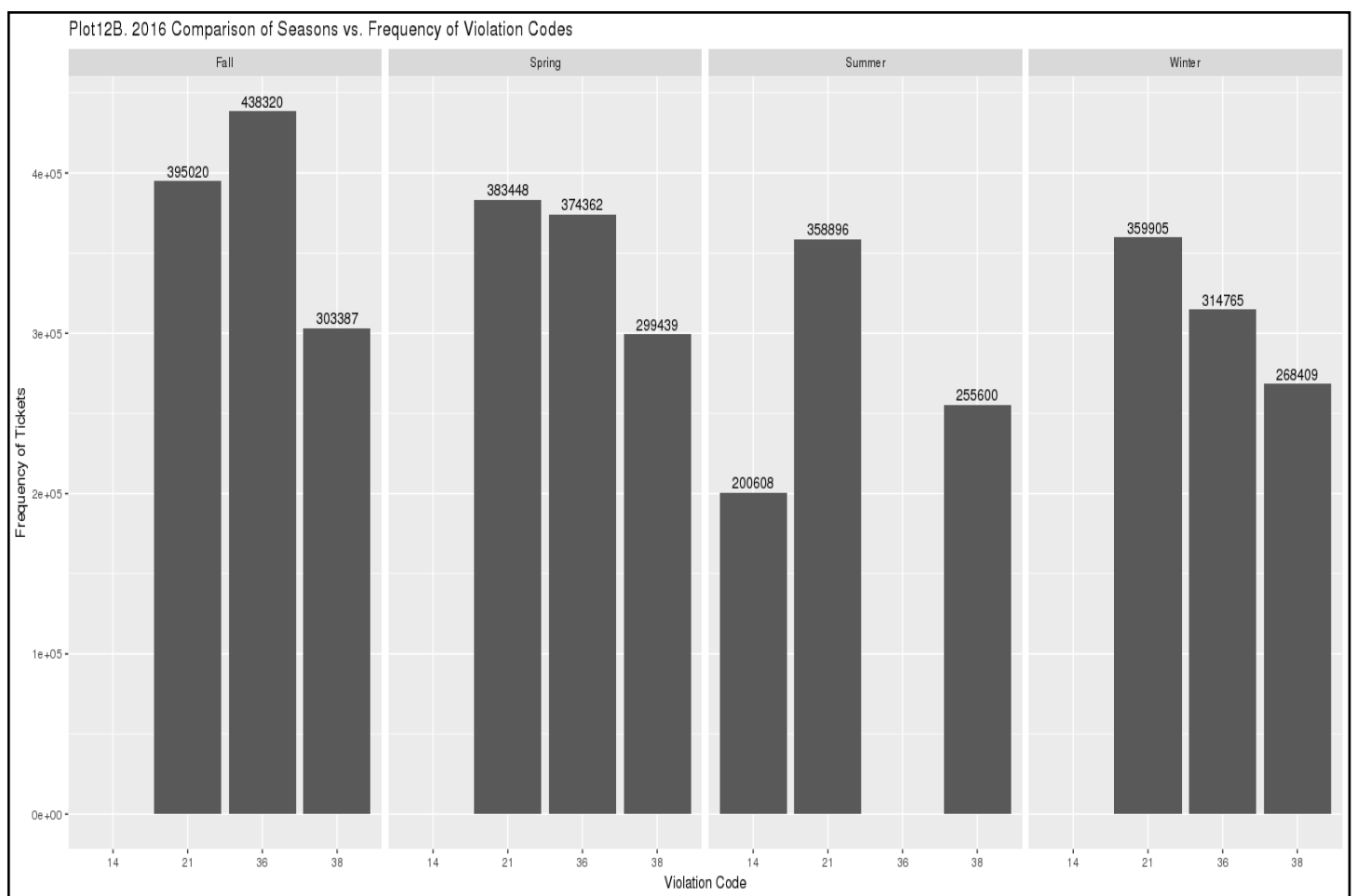


Insight: The Top-3 Violation Codes are 14: No Standing Violation, 21: No Parking Violation, 38: Parking Muni Meter Expired.

Violation Code 14 doesn't display any apparent pattern of seasonality and remains fairly constant. However, Violation Code 21 shows seasonal variation, with peaks in Summer and Spring and dips in Winter. Violation Code 38 doesn't show any significant seasonality it remains fairly constant with a dip in Winter. Overall, all three Violation codes show a dip in Winter when compared to the other seasons.

[3.6.2] 2016 Dataset [Top-3 Violation Codes 21, 36, 38 and 14]

Sl.No.	Season	Violation_Code	Frequency_of_Tickets
1	Fall	36	438320
2	Fall	21	395020
3	Fall	38	303387
4	Spring	21	383448
5	Spring	36	374362
6	Spring	38	299439
7	Summer	21	358896
8	Summer	38	255600
9	Summer	14	200608
10	Winter	21	359905
11	Winter	36	314765
12	Winter	38	268409

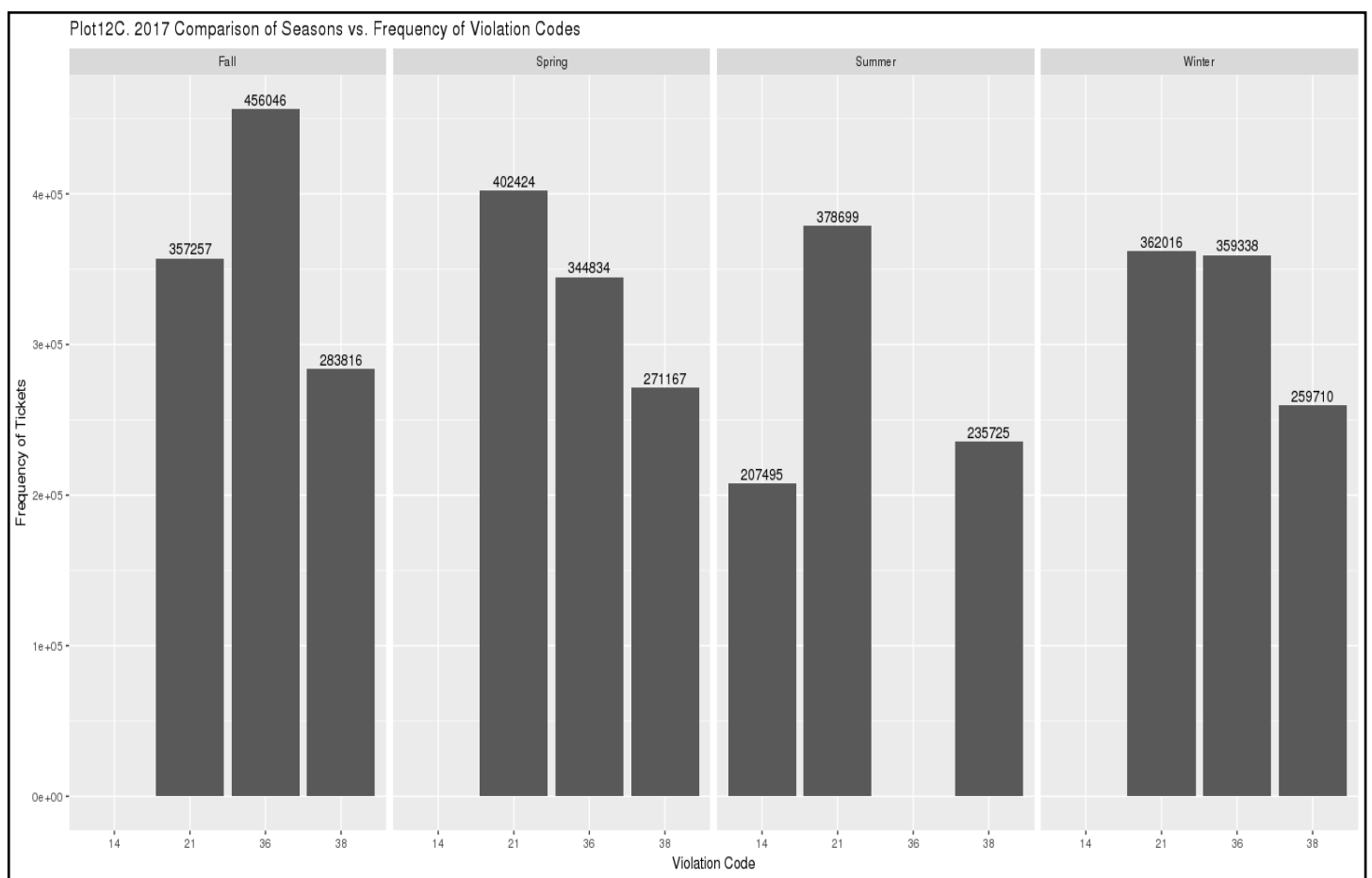


Insight: The Top Violation Codes are 21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired and 14: No Standing Violation

For Violation Code 21, it's fairly constant between season with a slight peak in Fall and Spring. For Violation Code 36 there is a significant peak in Fall. However, in Summer Violation Code 36 doesn't feature in the top-3 violation codes and is replaced with Violation Code 14. Violation Code 38 is fairly constant with peaks in Fall and Spring.

[3.6.3] 2017 Dataset [Top-3 Violation Codes 21, 36, 38 and 14]

Sl.No.	Season	Violation_Code	Frequency_of_Tickets
1	Fall	36	456046
2	Fall	21	357257
3	Fall	38	283816
4	Spring	21	402424
5	Spring	36	344834
6	Spring	38	271167
7	Summer	21	378699
8	Summer	38	235725
9	Summer	14	207495
10	Winter	21	362016
11	Winter	36	359338
12	Winter	38	259710



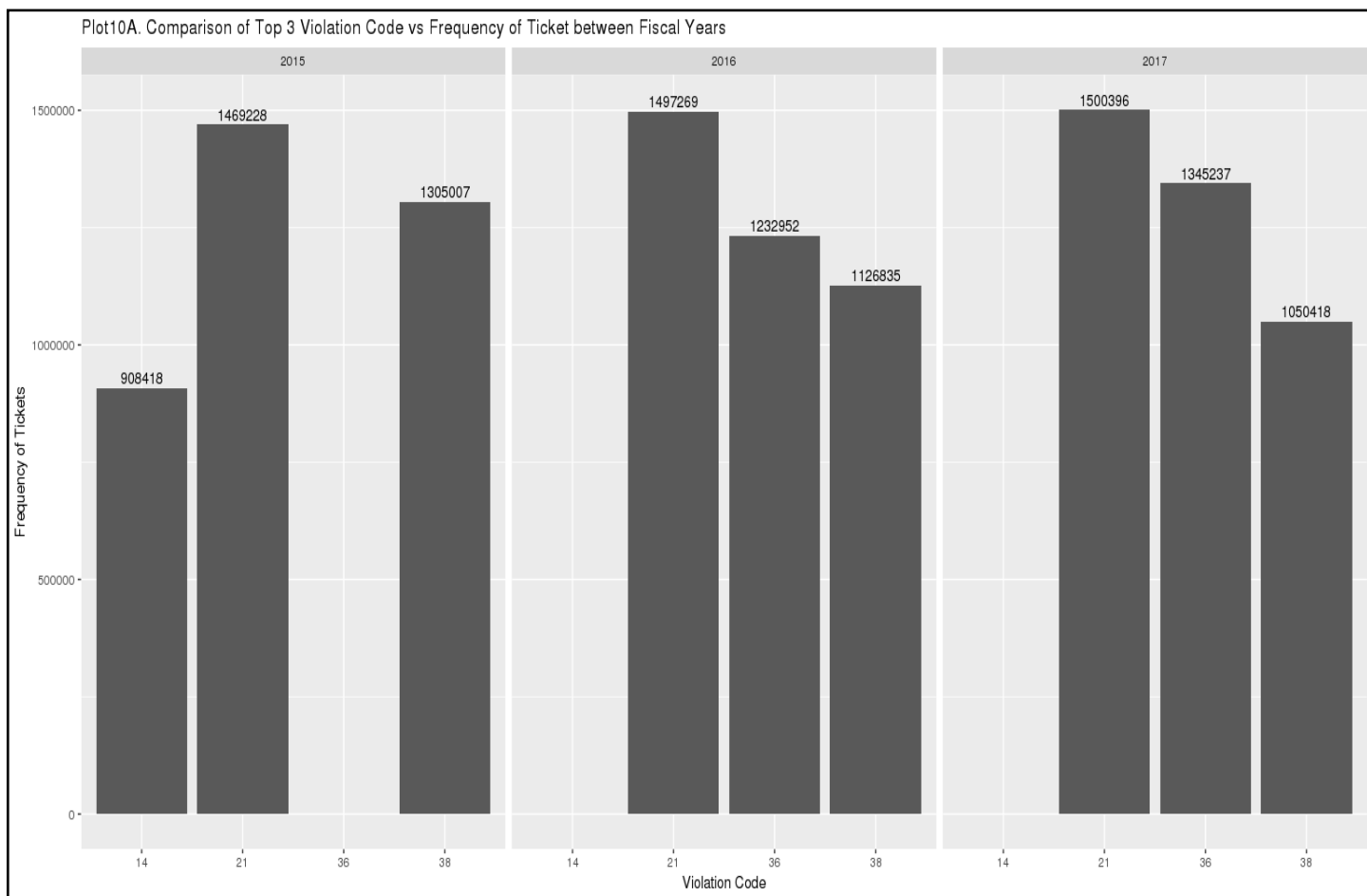
Insight: The Top Violation Codes are 21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired and 14: No Standing Violation

For Violation Code 21, there is a significant peak in Fall and remains fairly constant through the other seasons. For Violation Code 36 there is a significant peak in Fall. However, in Summer Violation Code 36 doesn't feature in the top-3 violation codes and is replaced with Violation Code 14. Violation Code 38 is fairly constant with peaks in Fall and Spring.

Stage 3: Question 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.

Sl.No.	Violation_Code	Frequency_of_Tickets	Fiscal_Year	Average_Fine_PerTicket	Total_Fine_Amount
1	21	1,469,228	2015	55	80,807,540
2	38	1,305,007	2015	50	65,250,350
3	14	908,418	2015	115	104,468,070
4	21	1,497,269	2016	55	82,349,795
5	36	1,232,952	2016	50	61,647,600
6	38	1,126,835	2016	50	56,341,750
7	21	1,500,396	2017	55	82,521,780
8	36	1,345,237	2017	50	67,261,850
9	38	1,050,418	2017	50	52,520,900

[Part-1] Find total occurrences of the 3 most common violation codes [Frequency vs. Violation Code]



Insight: The Top-3 most frequent violation codes for the year 2015 are 14, 21 and 38 [14: No Standing Violation, 21: No Parking Violation, 38: Parking Muni Meter Expired]. However, the frequency of Violation Codes 21 and 38 are significantly higher than Code 14 with the highest frequency of tickets being issued for Violation Code 21 [No-Parking Tickets].

The Top-3 most frequent Violation Codes for the year 2016 are 21,36 and 38 [21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired]. The highest frequency of tickets was issued for Violation Code 21 followed by 36 and 38 respectively.

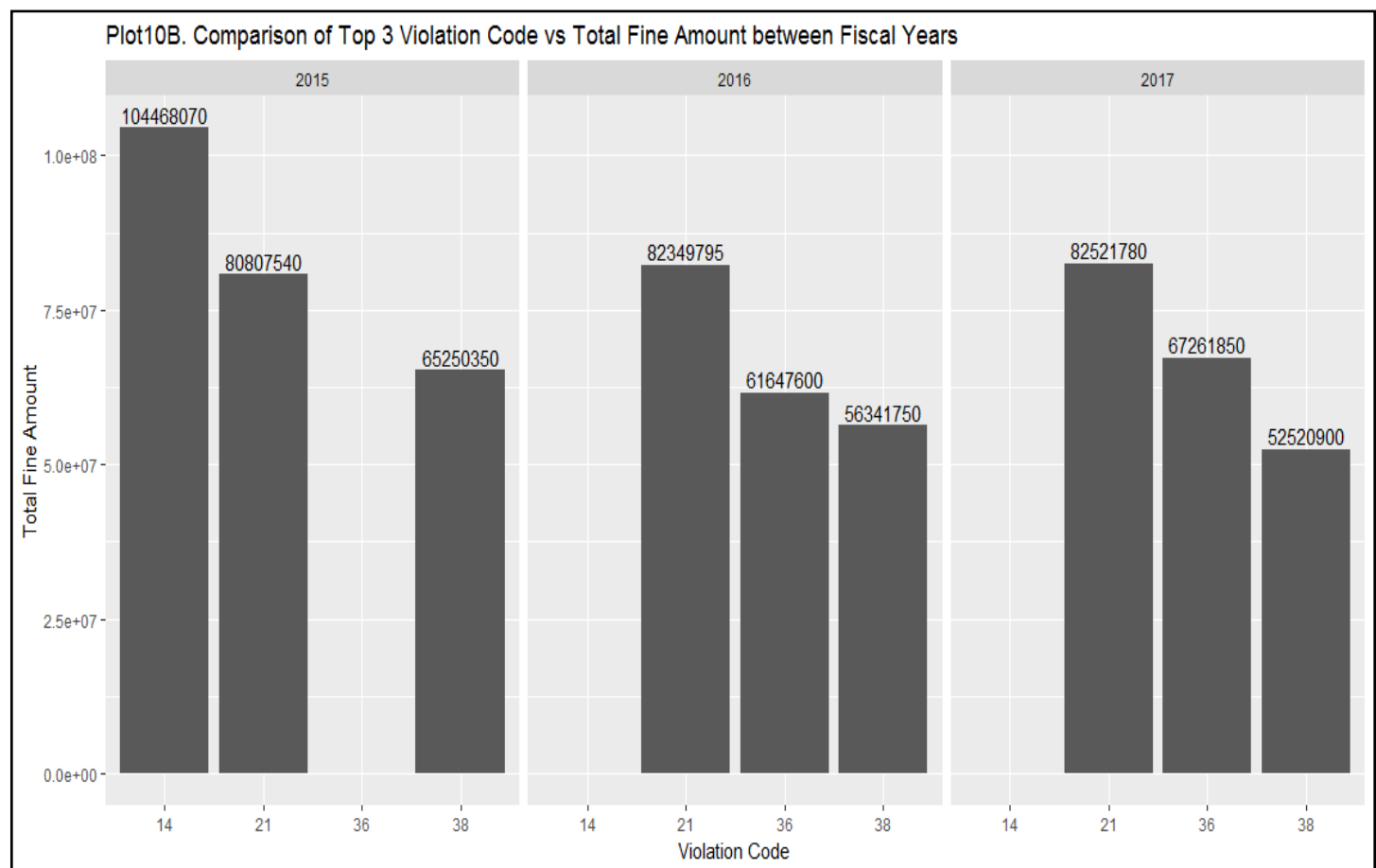
The Top-3 most frequent Violation Codes for the year 2017 are 21,36 and 38 [21: No Parking Violation, 36: Exceeding speed limit near School Zone, 38: Parking Muni Meter Expired]. The highest frequency of tickets was issued for Violation Code 21 followed by 36 and 38 respectively.

Overall Trends between Years: There is a gradual increasing trend in the frequency of tickets issued with Violation Code 21 from 1.46 million in the year 2015 to 1.50 million tickets in 2017. There is a significant drop in the frequency of tickets issued with Violation Code 38 from 1.30 million in 2015 to 1.05 million tickets in 2017. Whereas, Violation Code 14 only features in the top-3 for the year 2015 and Violation Code 36 features in top-3 for the year 2016 and 2017 with a gradual increasing trend of 1.23 million to 1.34 million tickets.

[Part-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.

Using this information, find the total amount collected for all of the fines. State the code which has the highest total collection.

What can you intuitively infer from these findings?



Insight: From the Total Collection vs. Violation Code Analysis graphs, 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. 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 Trends for Total Fine Amount between Years: With the exception of 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.