NewYork City Parking – Spark Case Study

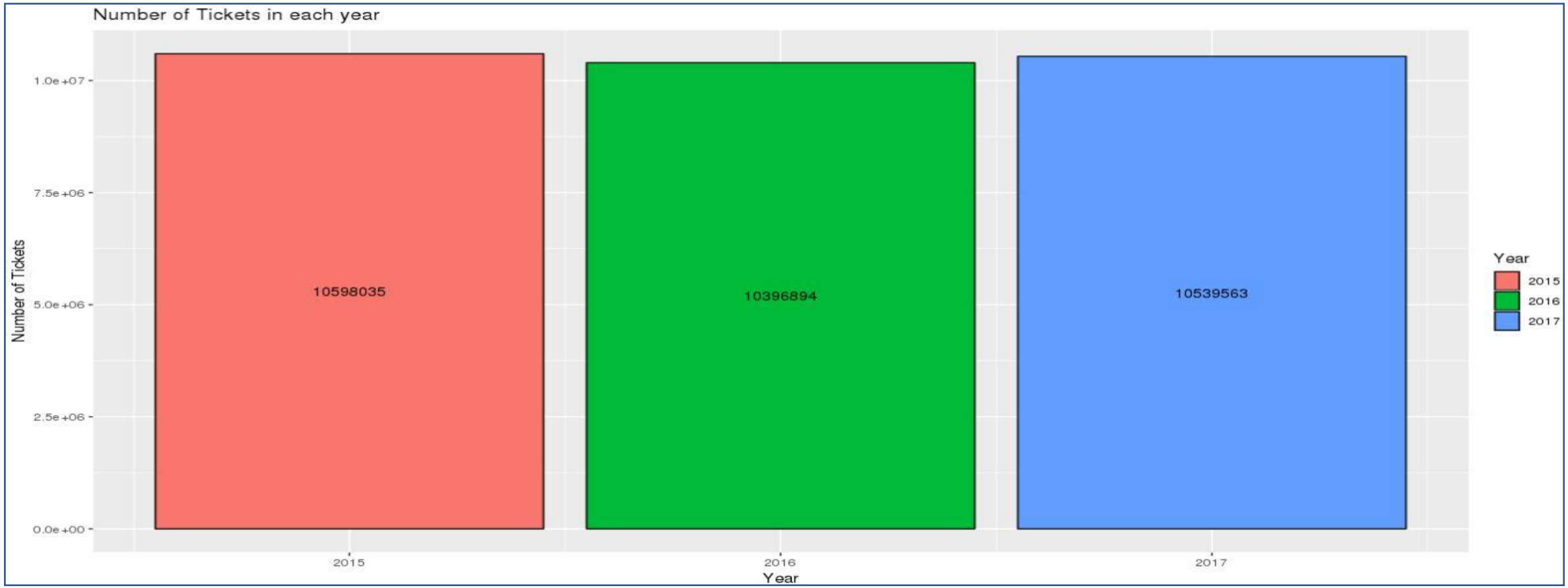
Payel Das
Shalini
Sridar Ganesh
Ravishankar DS

Assumptions

- We are using Fiscal year time frame for our analysis.
- We are using JULY14 to JUNE15 as year 2015 ,JULY15 to JUNE16 as year 2016 and JULY16 to JUNE 2017 as YEAR 2017
- Duplicate Summons number are removed.
- Dataset attribute names has blank special characters/ we have replaced with “_” to do the analysis
- Since "Census_Tract", "BIN", "BBL","Community_Board", "Community_Council", "Latitude", "Longitude", and "NTA" are captured in 2017 but it is there in 2015/2016 and hence the same will be removed.
- Since we have 00:00AM and 12:00AM we will combine . We have replaced all 00XXAM with 12XXAM
- We have standardized timestamp
- We have considered four major seasons: Summer(Jun-Aug), Winter(Dec-Jan), Fall(Sep-Nov) & Spring(Mar-May)

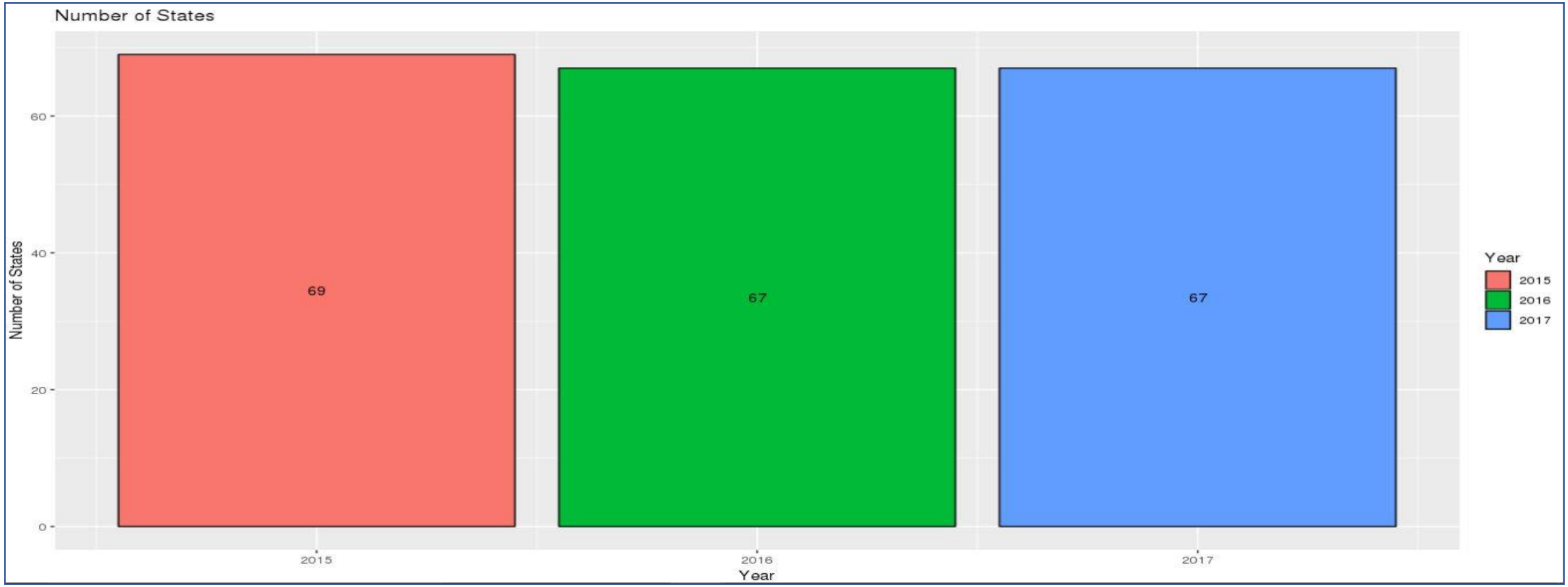
Examine the data

- **Question 1. Find total number of tickets for each year**
  - Total number of tickets issued for year 2015 is 10,598,035
  - Total number of tickets issued for year 2016 is 10,396,894
  - Total number of tickets issued for year 2017 is 10,539,563



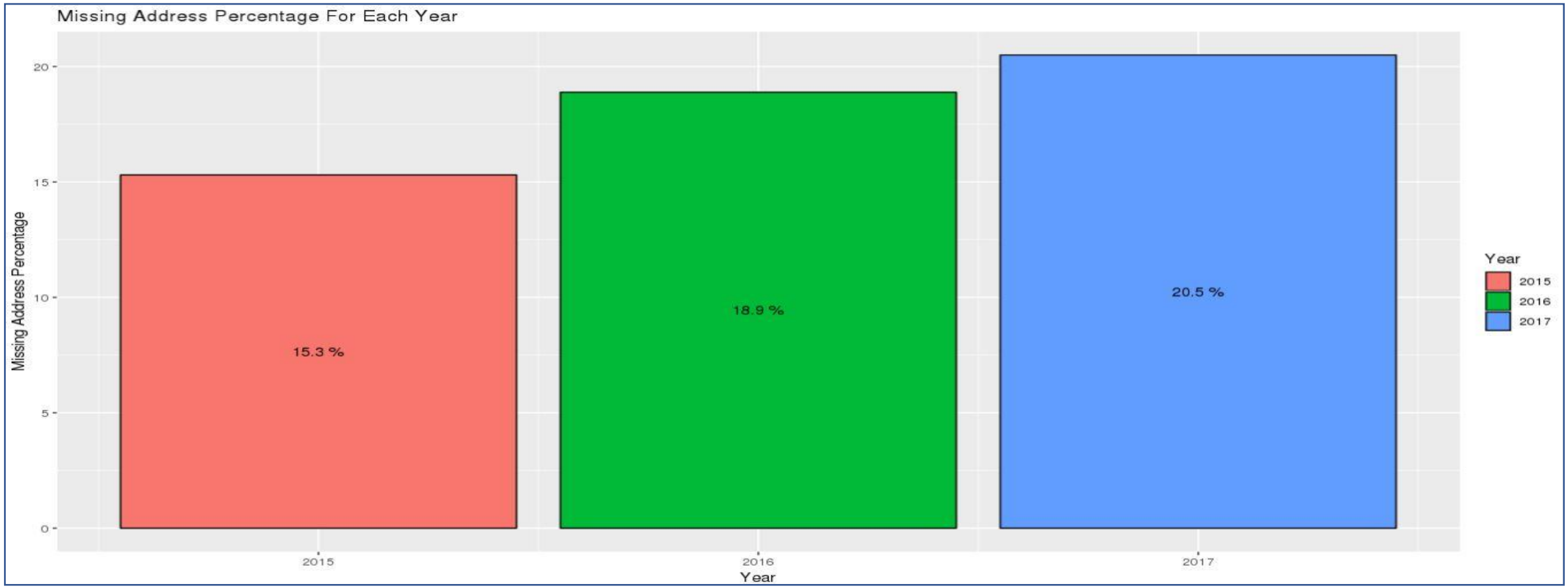
Examine the data

- **<u>Question 2: Find out how many unique states the cars which got parking tickets came from.</u>**
  - There is a numeric entry(“99”) in the registrationState column in 2016 data. We have replaced it with the most occurring state name i.e. ‘NY’ and got the below numbers:
  - 2015 data has data for **69** unique states, 2016 data has data for **67** unique states, 2017 data has data for **67** unique states



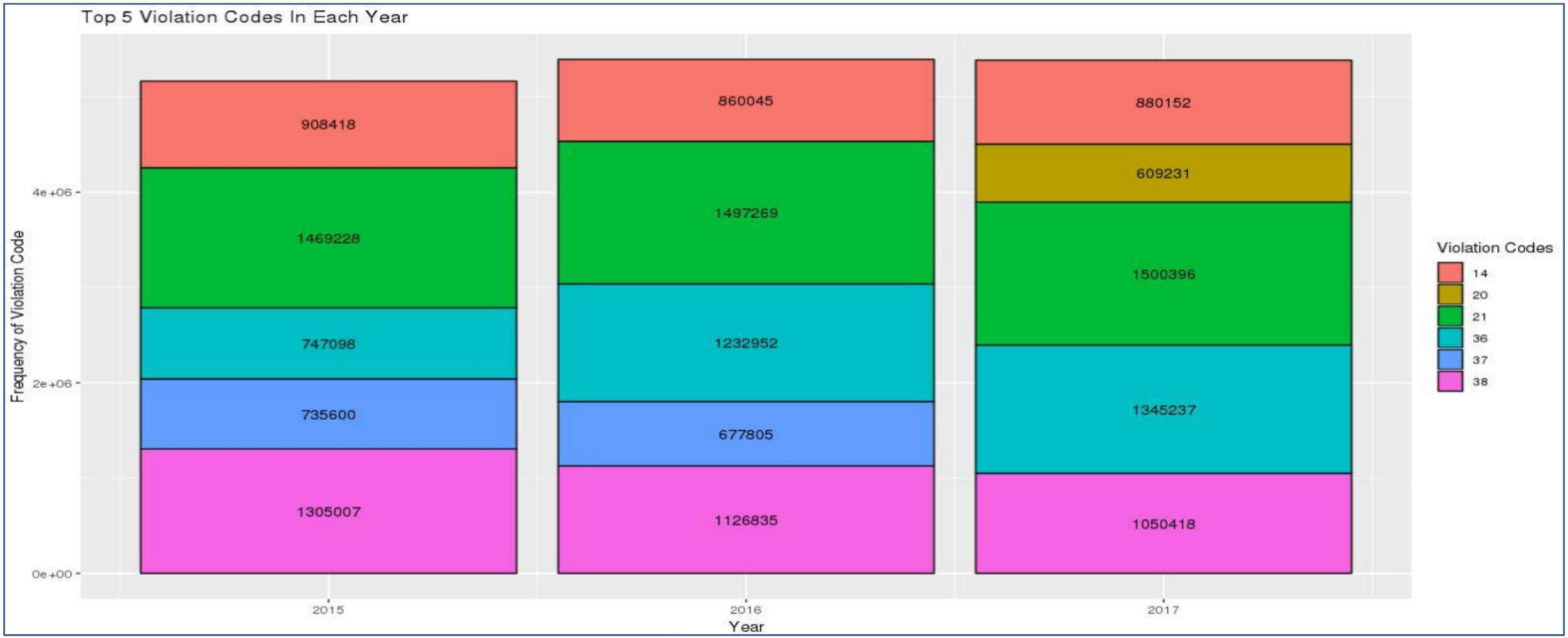
Examine the data

- **Question 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**
  - For 2015 - Total of 1622076 Missing Address are there in 2015 which is 15.3% of overall data.
  - For 2016 - Total of 1963921 Missing Address are there in 2015 which is 18.8% of overall data.
  - For 2017 - Total of 2160639 Missing Address are there in 2015 which is 20.5% of overall data.



Aggregation tasks

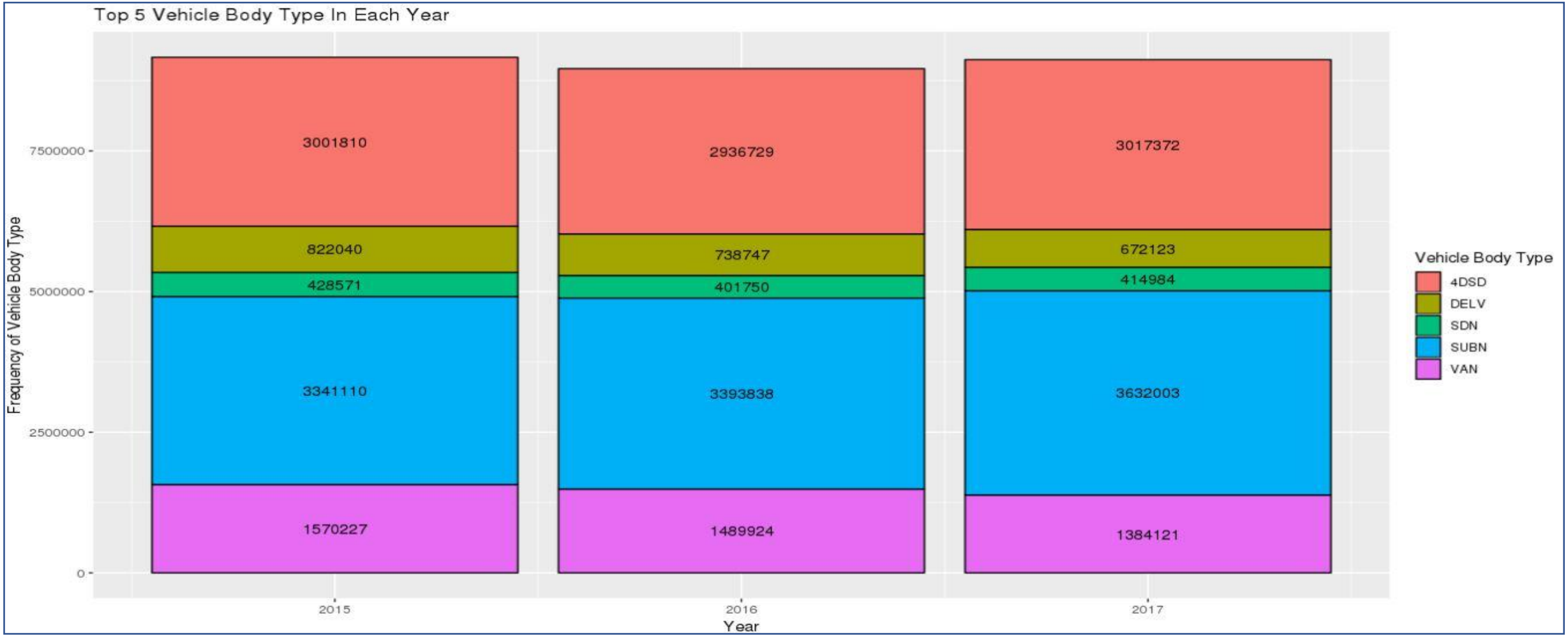
- **Question 1. How often does each violation code occur? Display the frequency of the top five violation codes**



14 is the common violation code that occurs most frequently every year

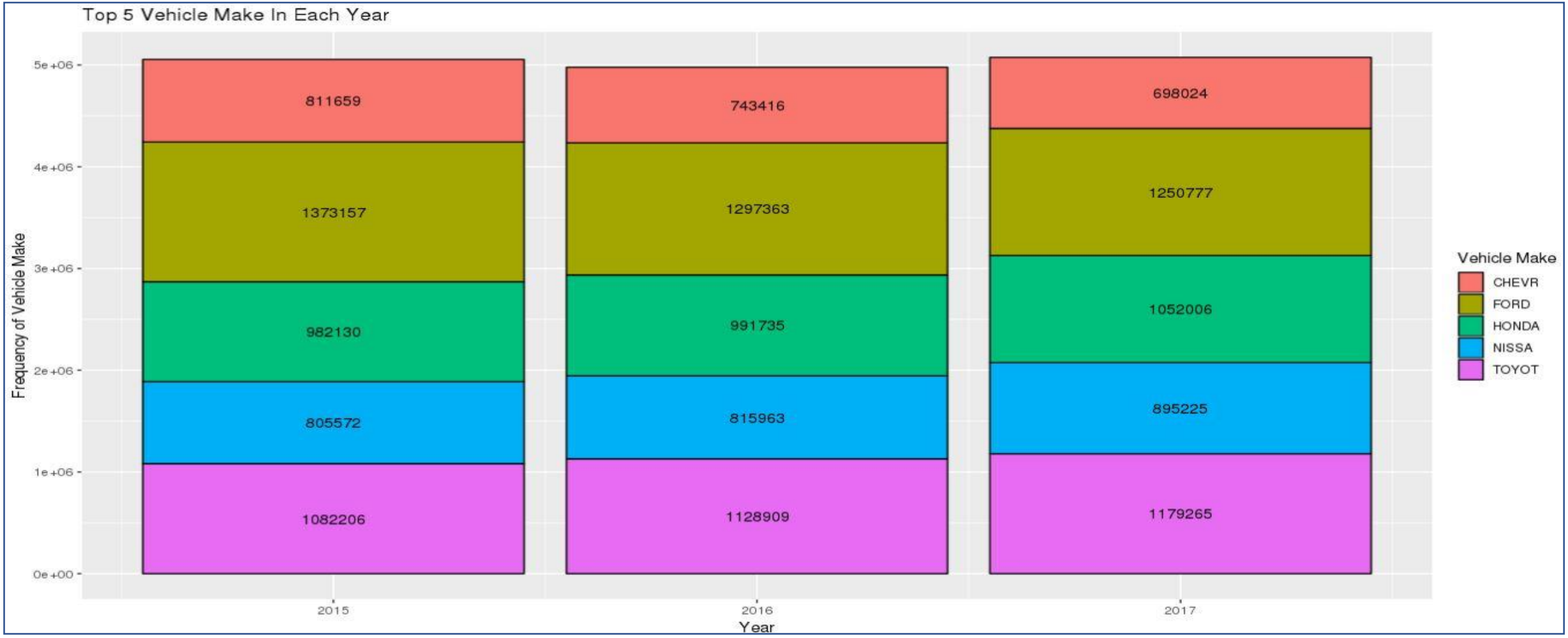
Aggregation tasks

- **Question 2. How often does each 'vehicle body type' get a parking ticket?**



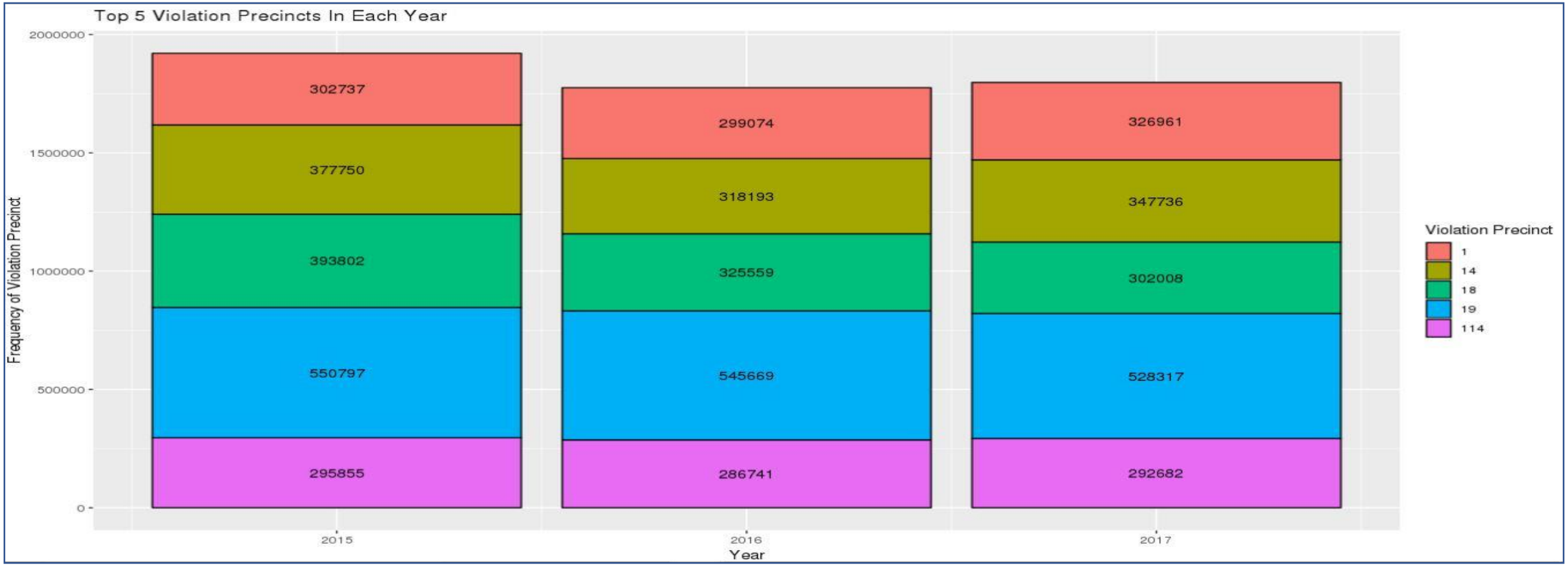
Aggregation tasks

- **Question 2(continued). How about the 'vehicle make'?**



Aggregation tasks

- **Question 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:**
- **'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?**
- **Insight: In Precinct 18 & 19 the frequency of violation has reduced year over year.**

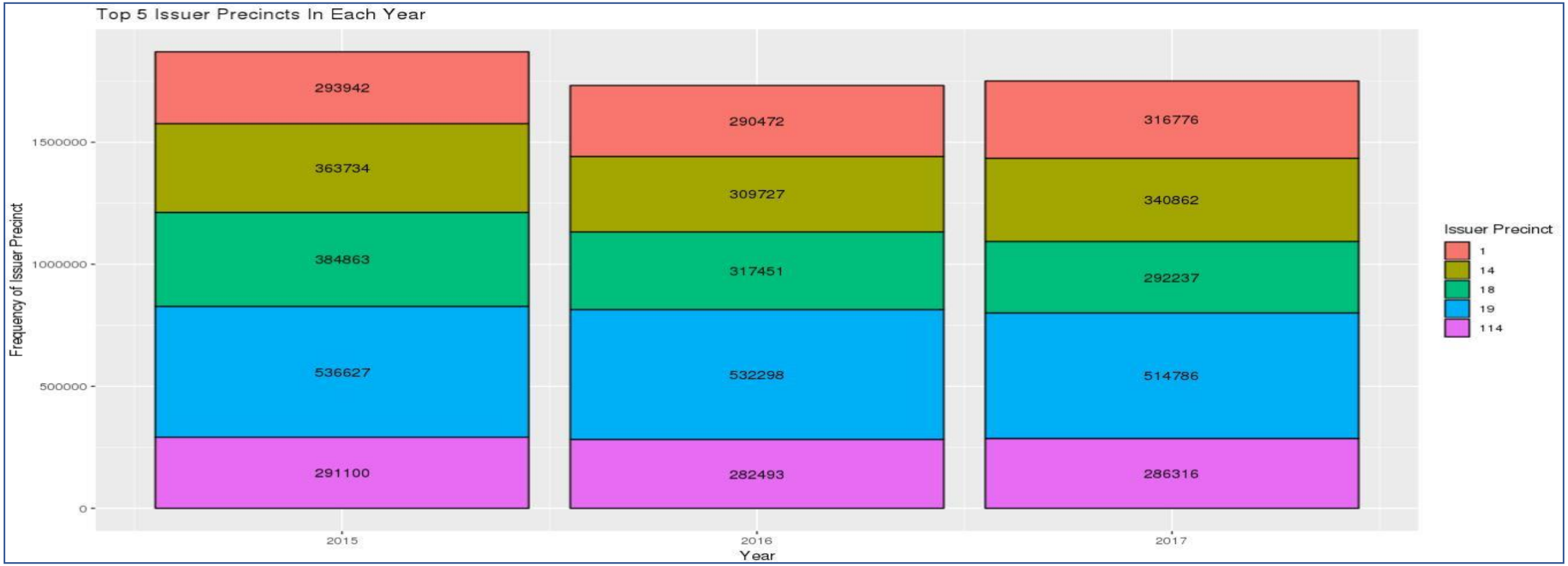


Aggregation tasks

- **Question 3 (continued): 'Issuer Precinct' (this is the precinct that issued the ticket)**

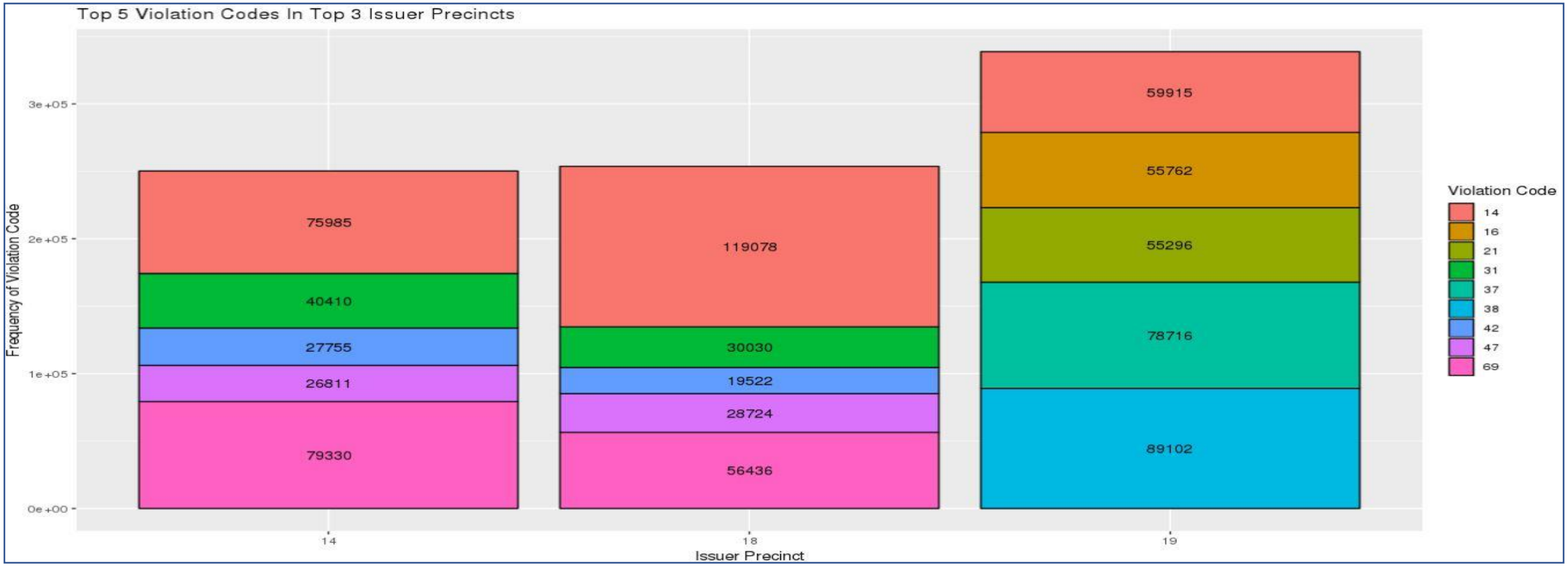
**We have excluded precinct 0 from our analysis as in the question it was mentioned as an erroneous code. We have looked at other valid precinct values.*

- **Insight: In Precinct 18 & 19 the frequency of Issuing ticket has reduced year over year, because violation of rules has also reduced there over the years.**



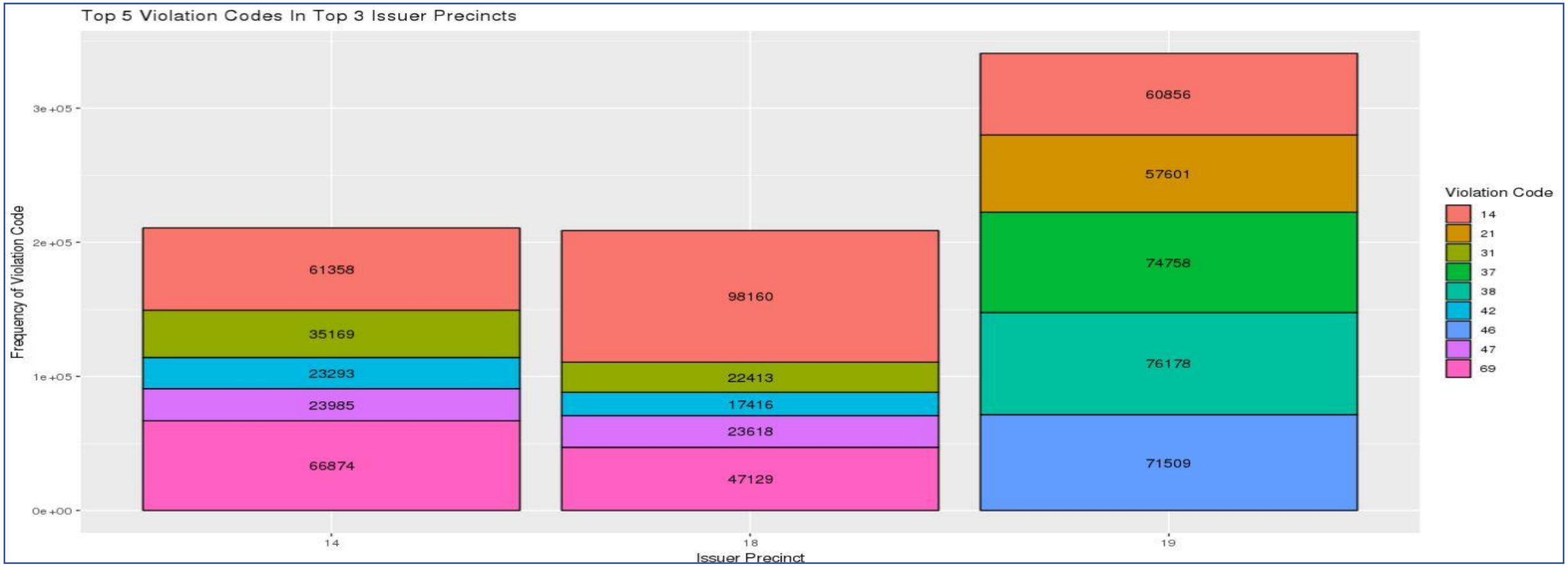
Aggregation tasks

- **Question 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?**
- **Year 2015:**



Aggregation tasks

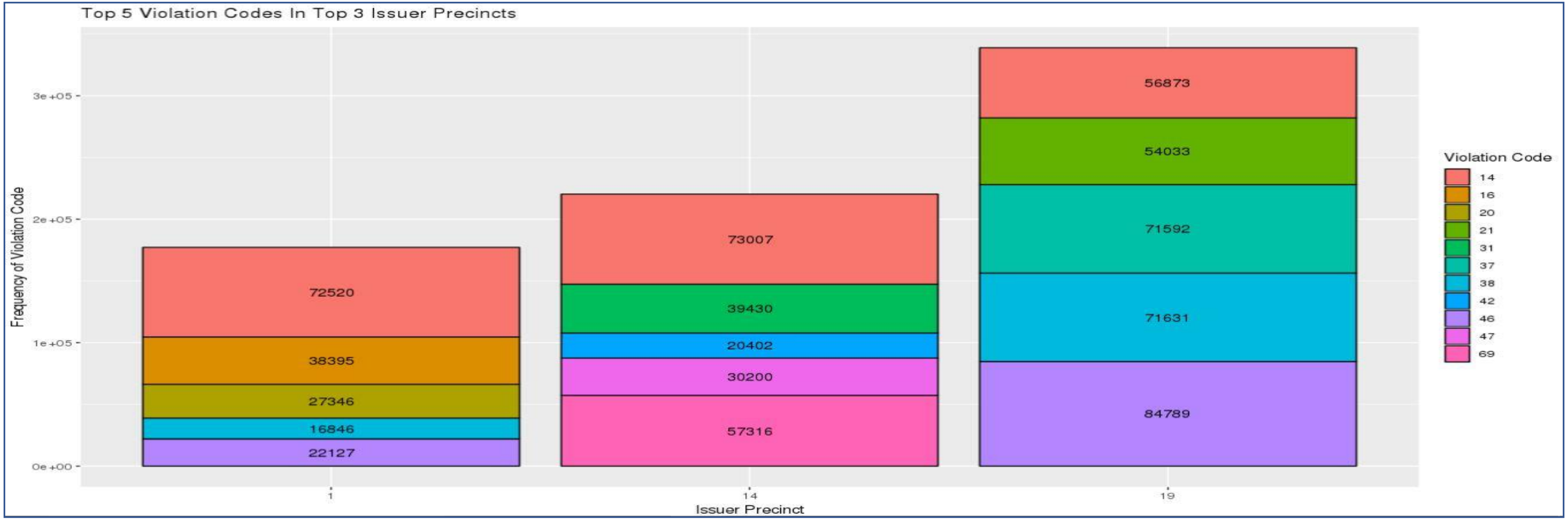
- **Question 4(cont.): Find the violation code frequency across three precincts which have issued the most number of tickets - do these precinct zones have an exceptionally high frequency of certain violation codes? Are these codes common across precincts?**
- **Year 2016:**



Aggregation tasks

- **Question 4(cont.): 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?**

- **Year 2017:**



14 is the most commonly occurring violation code in all three years

Aggregation tasks

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

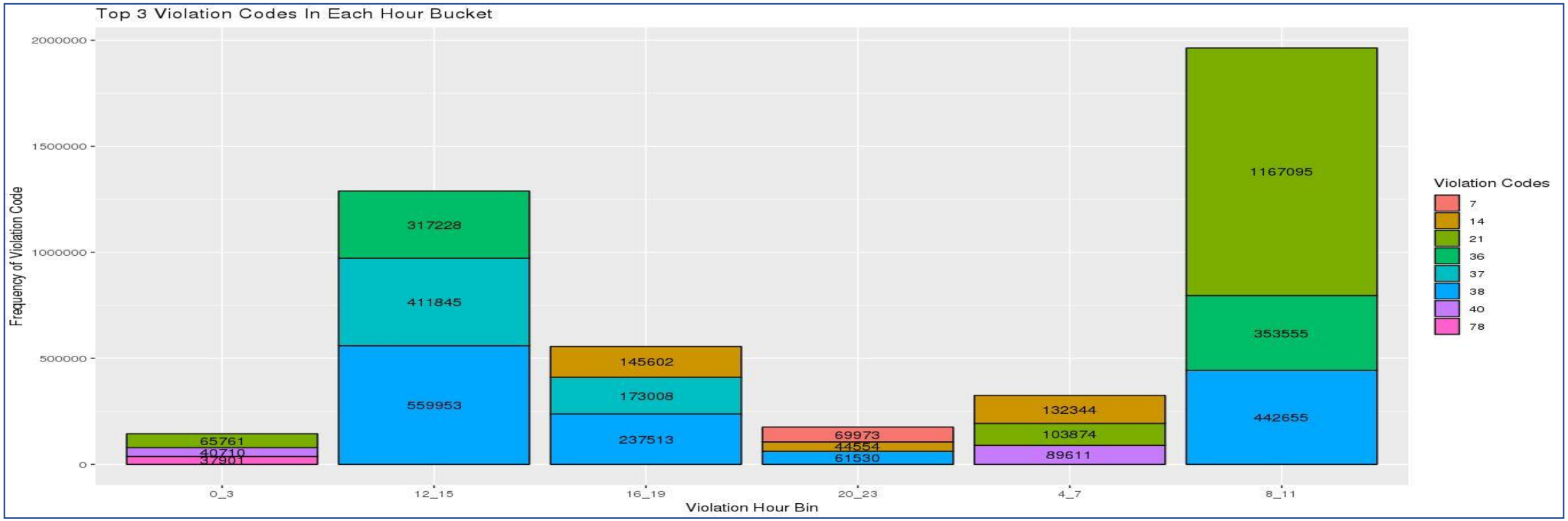
- **Find a way to deal with missing values, if any.:**
  - We have removed Null values since there were very few null violation times.

- **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.:**
  - Violation time was in string format having some special characters like '+' & '.' which we had replaced with 0s & we have converted 00XXAM into 12XXAM. Violation time had only 'A' or 'P' at the end of the string & we just added one 'M' next to it so that we can convert it into a timestamp format.

- **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.**
  - We have divided 24 hours into 6 bins namely: 0-3,4-7,8-11,12-15,16-19,20-23
  - Year wise graphs are shown in next slides

Aggregation tasks

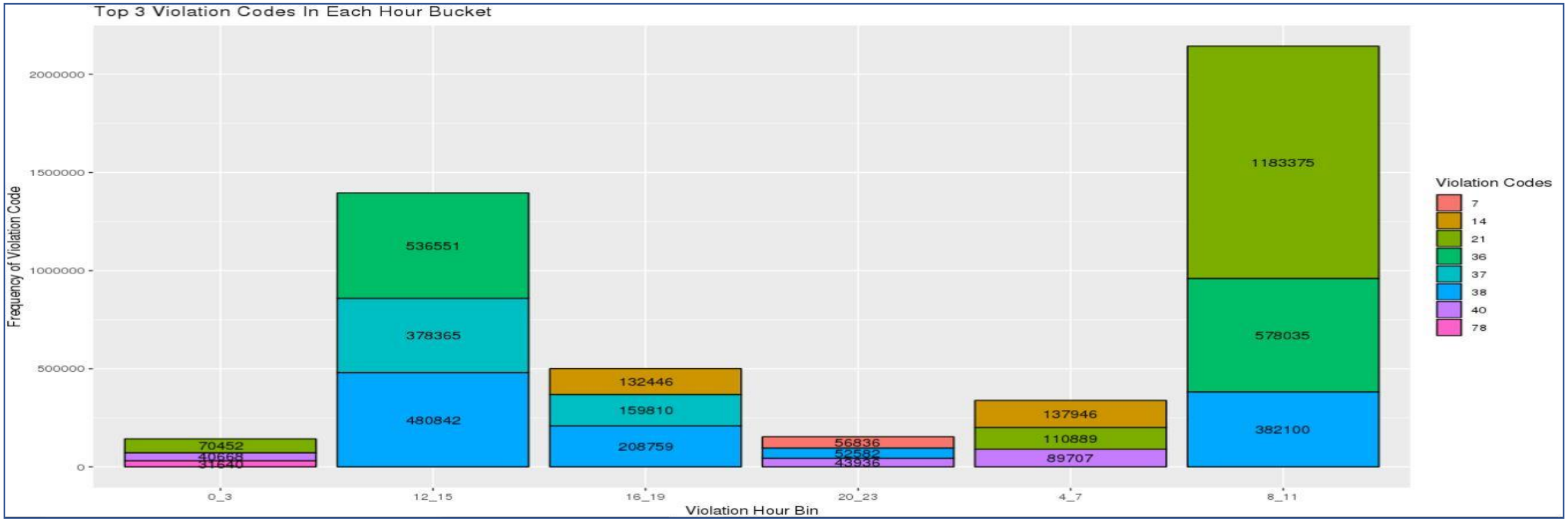
- **Question 5(part 1): Divide 24 hours into six equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the three most commonly occurring violations.**
- **Year 2015:**



Most frequent Violation Codes 21,36,38

Aggregation tasks

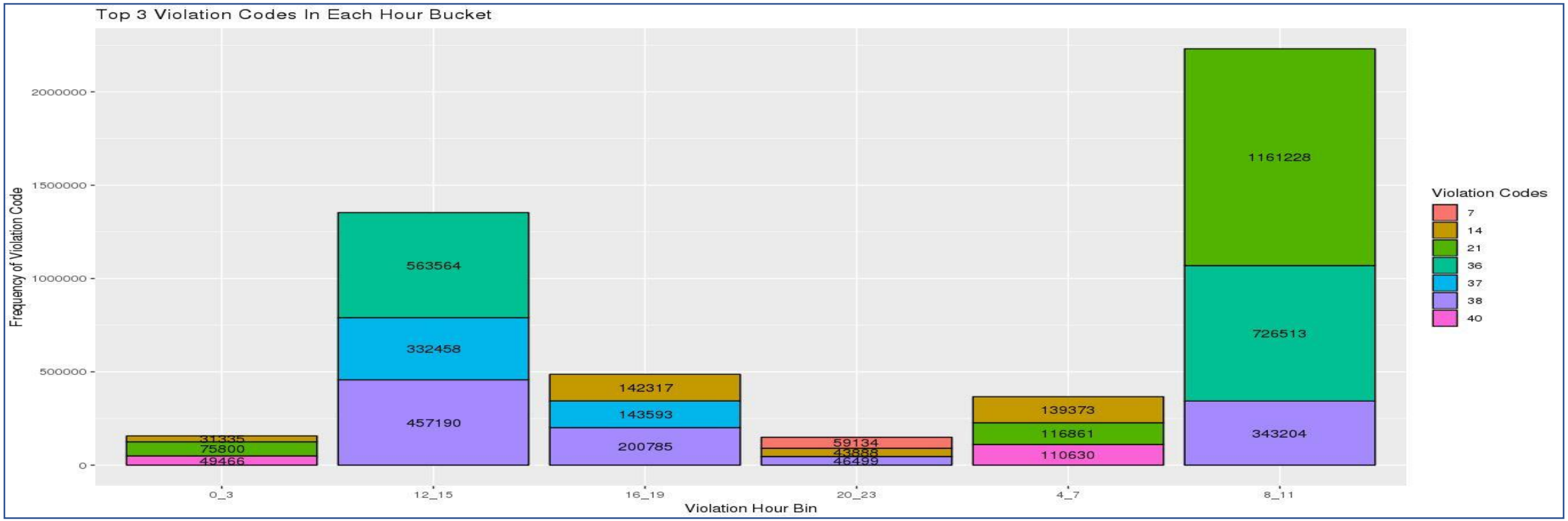
- **Question 5 (part 1): Divide 24 hours into six equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the three most commonly occurring violations.**
- **Year 2016:**



Most frequent Violation Codes 21,36,38

Aggregation tasks

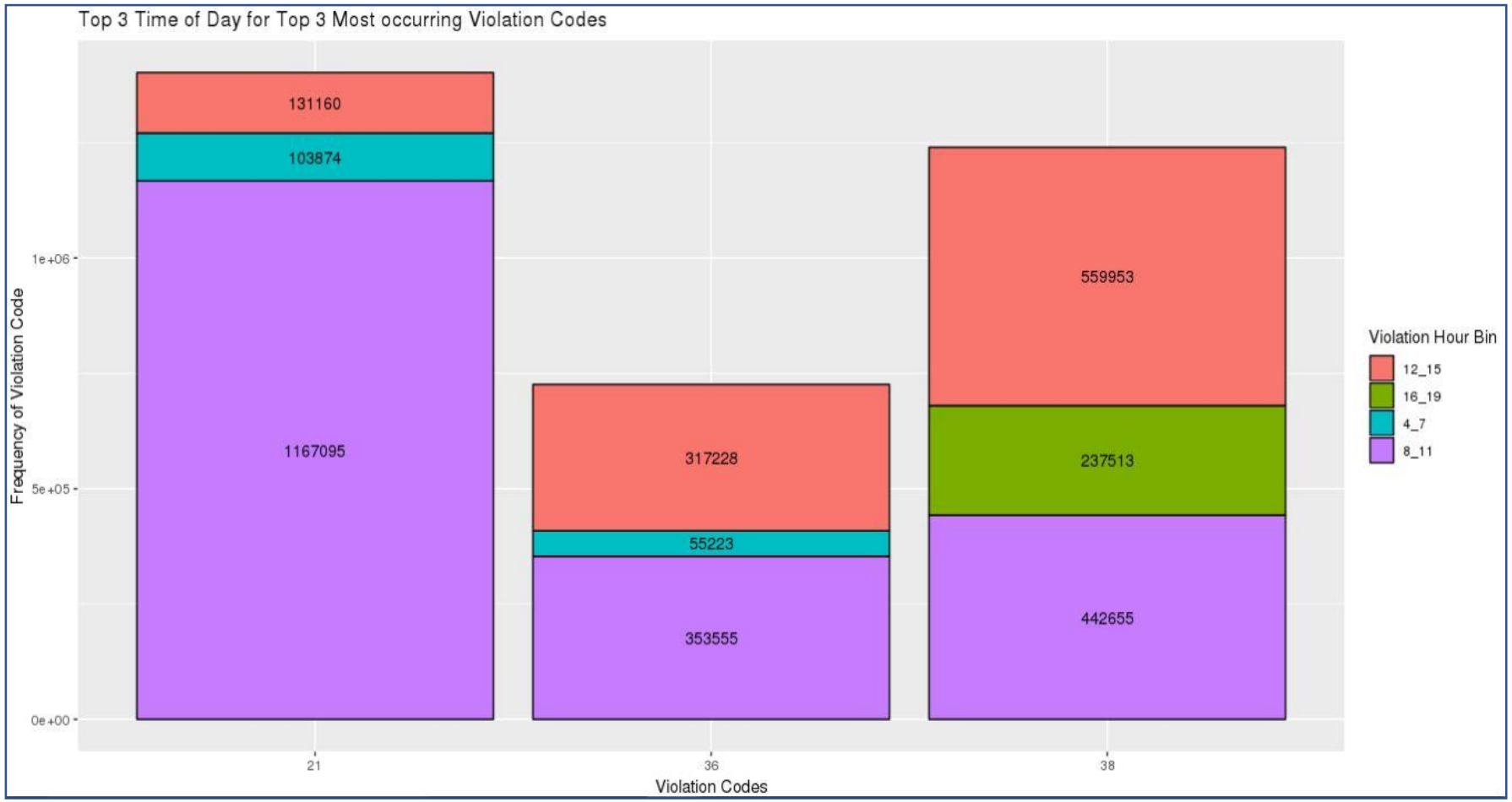
- **Question 5 (part 1): Divide 24 hours into six equal discrete bins of time. The intervals you choose are at your discretion. For each of these groups, find the three most commonly occurring violations.**
- **Year 2017:**



Most frequent Violation Codes 21,36,38

Aggregation tasks

- **Question 5 (part 2): 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)**
- **Year 2015:**

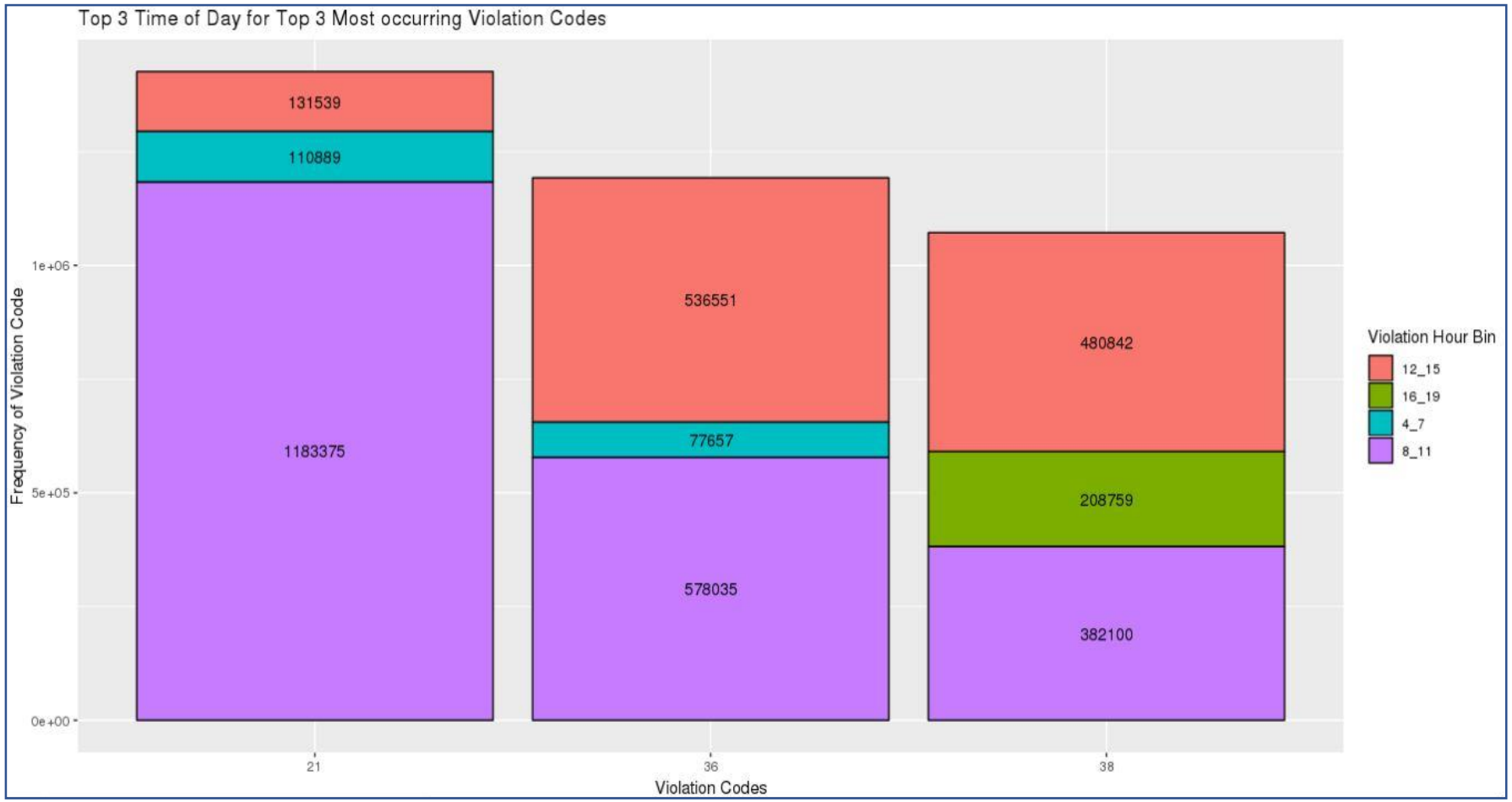


Top 3 Time of day for each of the top 3 violation codes

21: 8-11,4-7,12-15
36: 8-11,4-7,12-15
38: 8-11,16-19,12-15

Aggregation tasks

- **Question 5 (part 2): 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)**
- **Year 2016:**

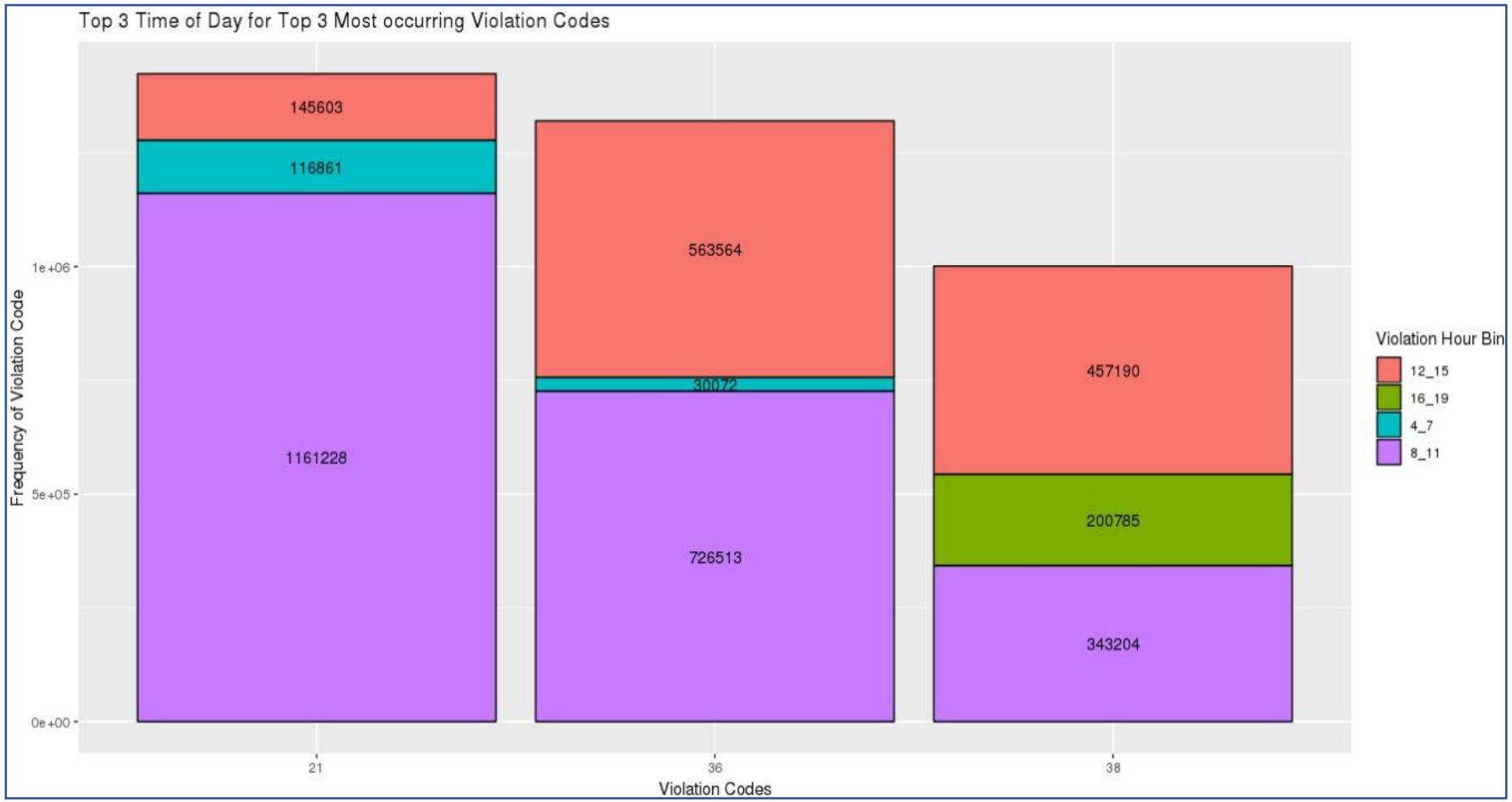


Top 3 Time of day for each of the top 3 violation codes

21: 8-11,4-7,12-15
36: 8-11,4-7,12-15
38: 8-11,16-19,12-15

Aggregation tasks

- **Question 5 (part 2): 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)**
- **Year 2017:**

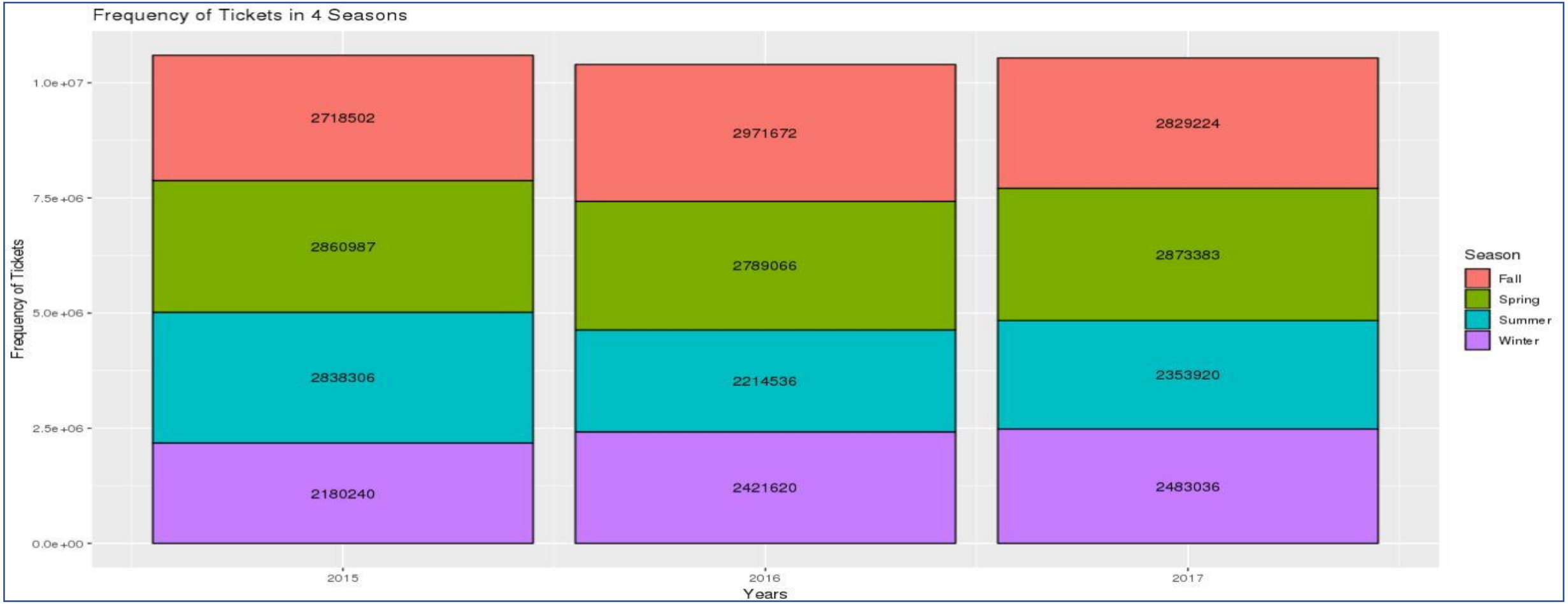


Top 3 Time of day for each of the top 3 violation codes

21: 8-11,4-7,12-15
36: 8-11,4-7,12-15
38: 8-11,16-19,12-15

Aggregation tasks

- **Question 6: Let's try and find some seasonality in this data**
  - We have considered four major seasons: Summer(Jun-Aug), Winter(Dec-Jan), Fall(Sep-Nov) & Spring(Mar-May)
- **First, divide the year into some number of seasons, and find frequencies of tickets for each season.**



Aggregation tasks

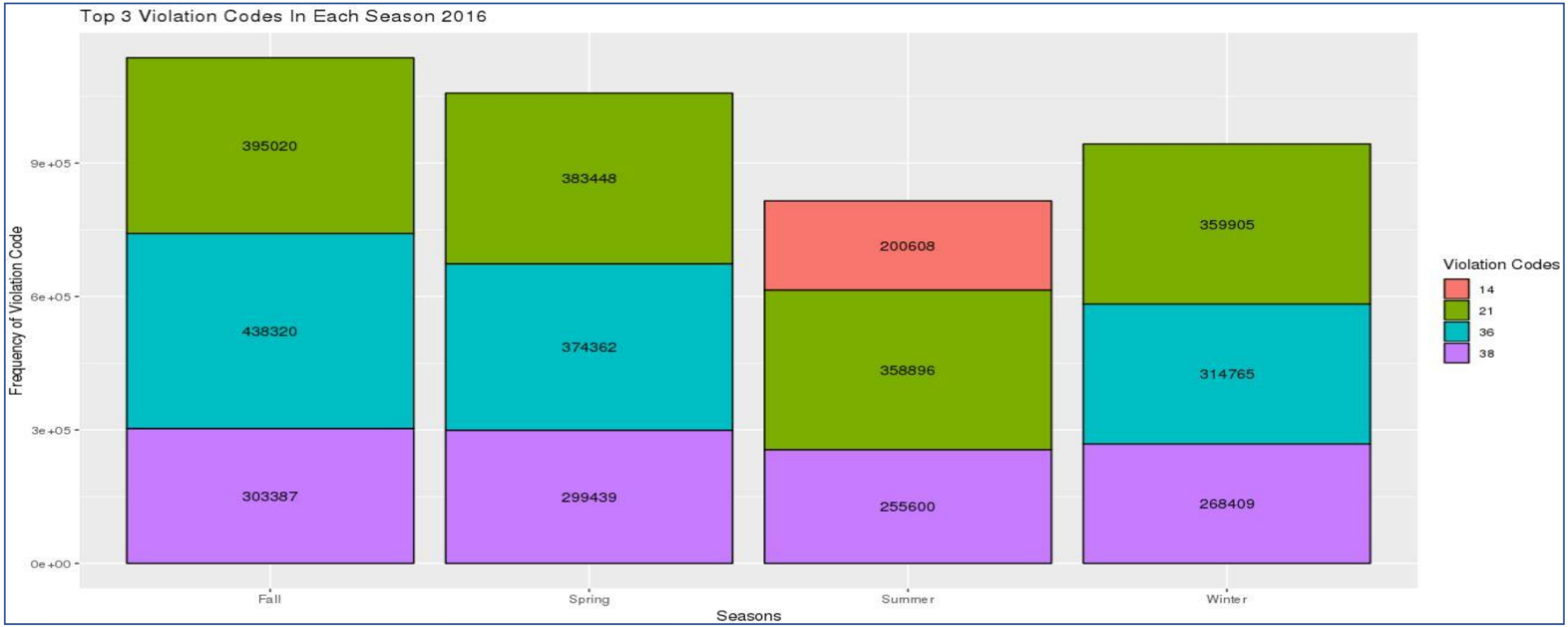
- **<u>Question 6: Then, find the three most common violations for each of these seasons.</u>**

- In the year 2015, most common violations for all these seasons were 14,21,38

Aggregation tasks

- **Question 6: Then, find the three most common violations for each of these seasons.**

- In the year 2016, most common violations were 21,26,38 for Fall, Spring & Winter. Only for Summer it was 14,21,38



Aggregation tasks

- **<u>Question 6: Then, find the three most common violations for each of these seasons.</u>**

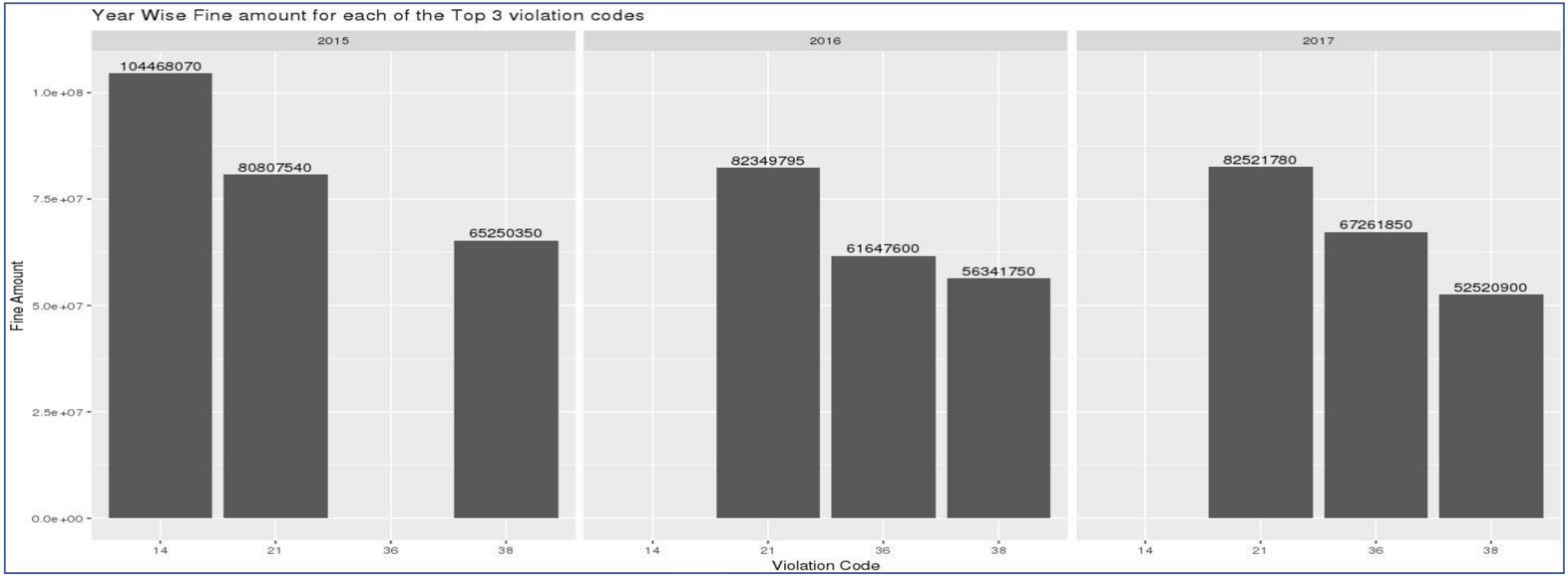
- In the year 2017, most common violations were 21,26,38 for Fall, Spring & Winter. Only for Summer it was 14,21,38

Aggregation tasks

- **Question 7: Find total occurrences of the three most common violation codes.**
  - In 2015 the most common violation codes were 14,21,38
  - In 2016 the most common violation codes were 21,36,38
  - In 2017 the most common violation codes were 21,36,38

Aggregation tasks

- **Question 7: 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..**
  - In 2015 the most common violation codes were 14,21,38, Violation Code 14 had the highest fine amount: $104.4M
  - In 2016 the most common violation codes were 21,36,38, Violation Code 21 had the highest fine amount: $82.3M
  - In 2017 the most common violation codes were 21,36,38, Violation Code 21 had the highest fine amount: $82.5M



Aggregation tasks

- **Question 7: What can you intuitively infer from these findings?**

  - Except 2015, Violation Code 21(Street Cleaning: No parking where parking is not allowed by sign, street marking or traffic control device.) was charged the highest annual fine amount and stays in the range of $82.3-$82.5 million.
  - In 2015, Violation Code 14(General No Standing: Standing or parking where standing is not allowed by sign, street marking or; traffic control device.) brings the highest fine amount $104.4 million. However, Violation Code 14 does appear in the top-3 violation codes of the year 2016 and 2017.
  - Violation Code 38(Failing to show a receipt or tag in the windshield.) shows a gradual decrease in the total fine amount collected from $65.25 million in 2015 to $52.52 million in 2017. Which means violation of rules in this category has reduced over the years.
  - Violation Code 36(Exceeding the posted speed limit in or near a designated school zone.) only appears in the top-3 for 2016 and 2017 with a mild increasing trend of $61.64 million to $67.26 million between 2016 and 2017 respectively. Which means in 2017, people tend to break the speed limit more as compared to 2016 & hence, violated the traffic rules.