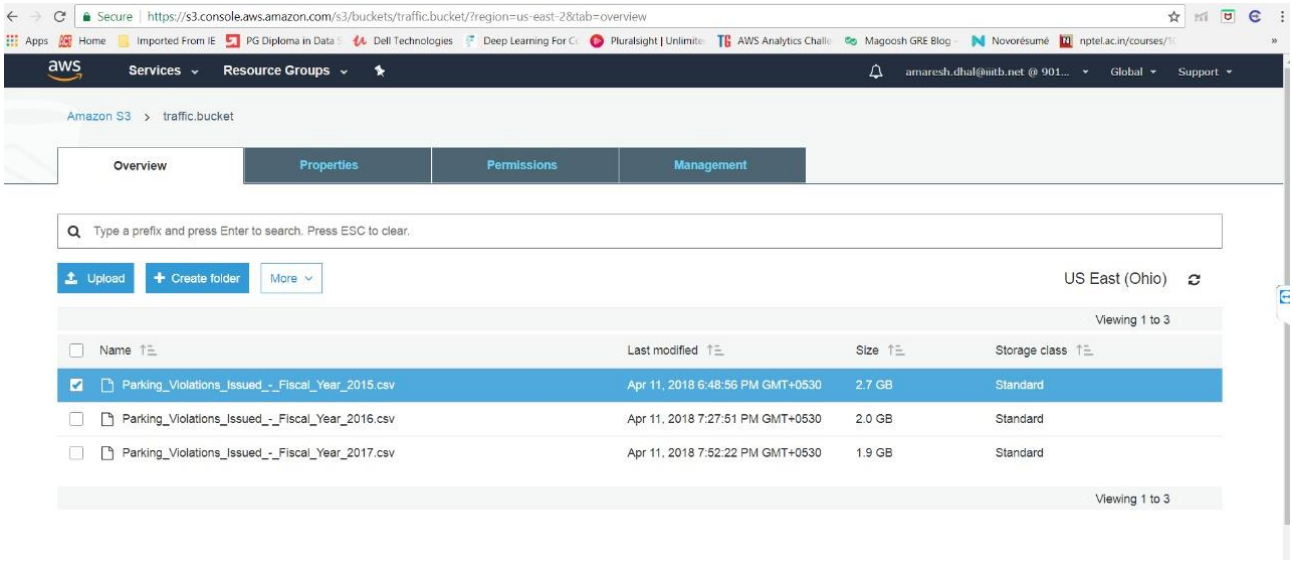
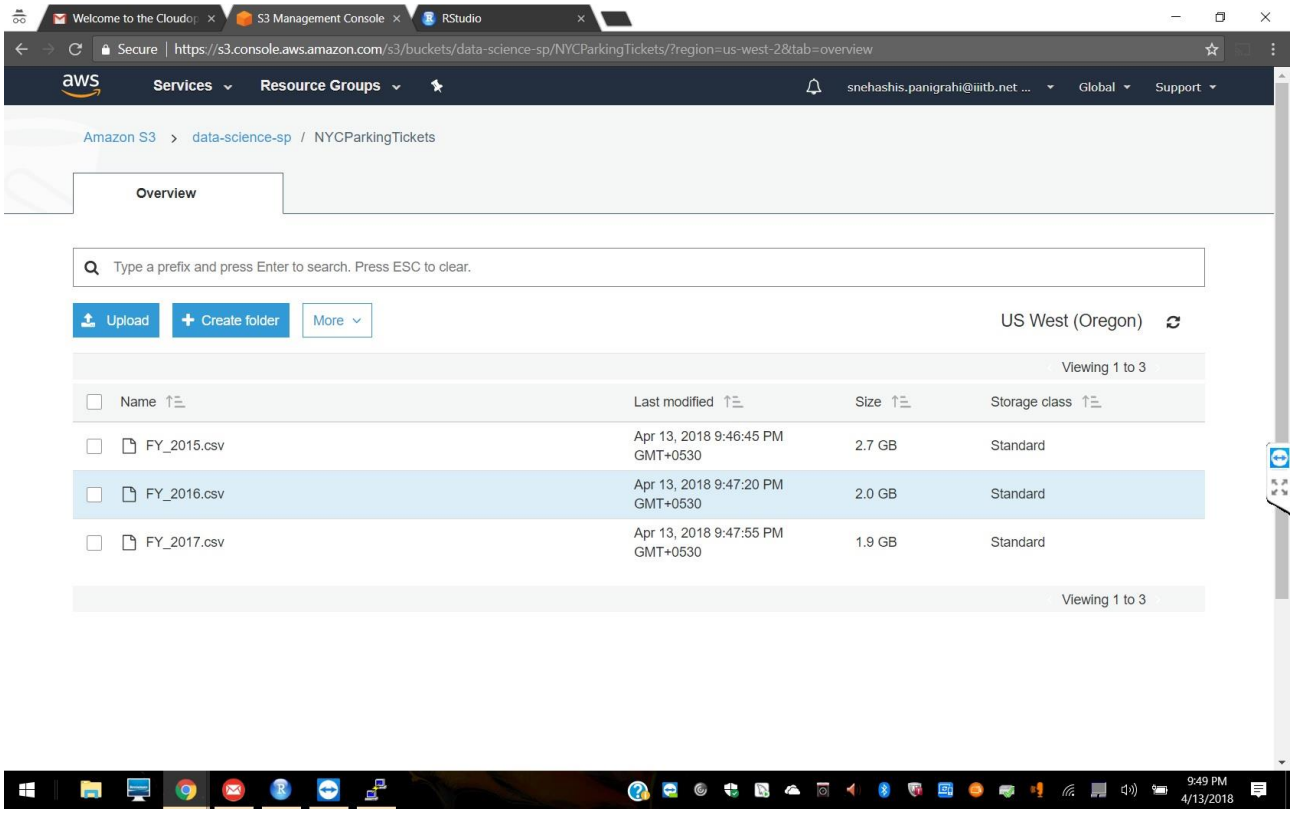
S3 Bucket screen shot:

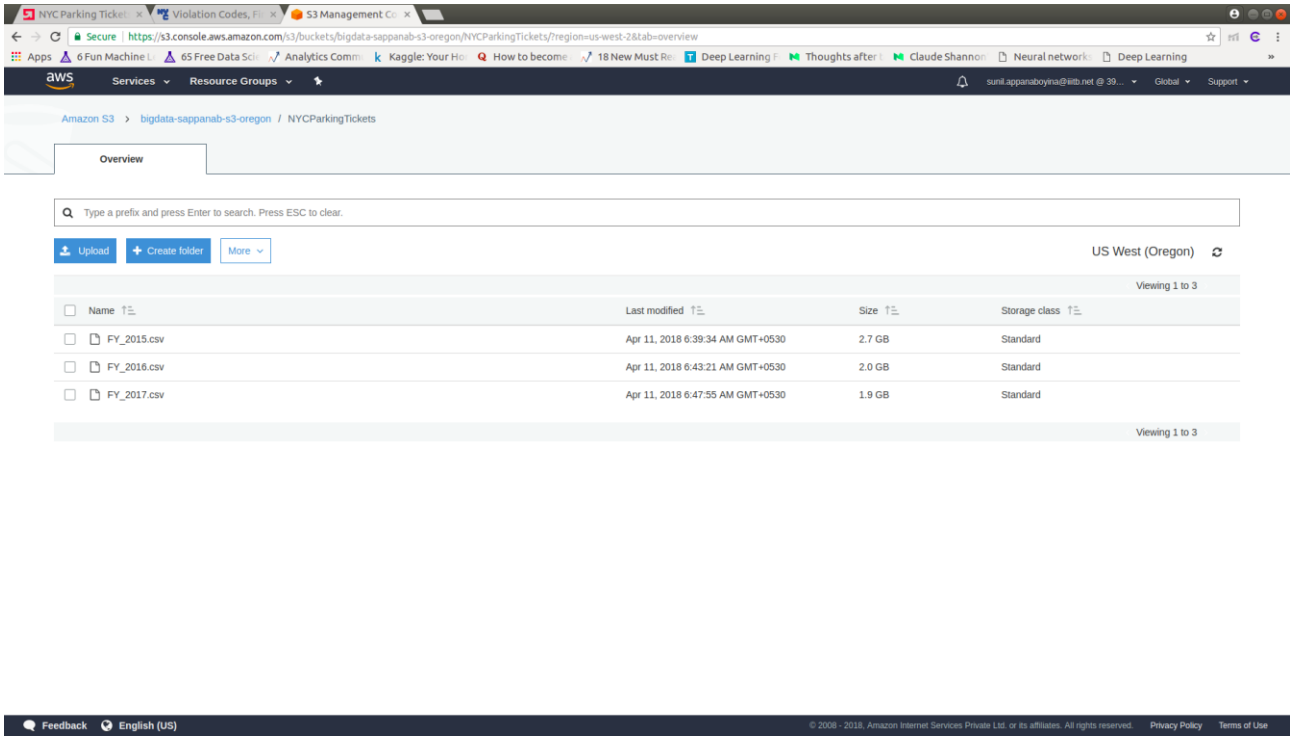
Amaresh Dhal:



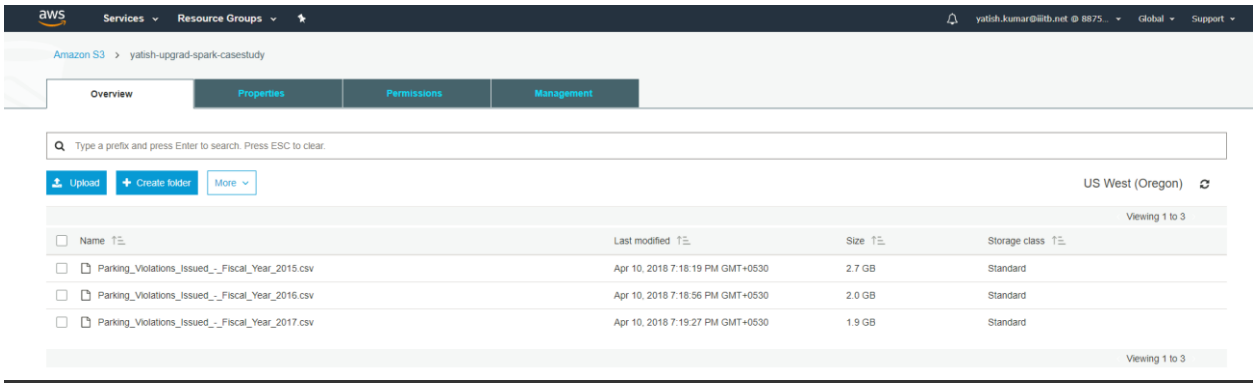
Snehashish Panigrahi:



Sunil Appanaboyina:



Yatish Kumar



Assumptions:

1) Each CSV file has data from other years as well, we have taken only the records with respective year provided as per file name.

2) We have considered all states apart from US alone, Canadian states where also reported the analysis.

3) Missing Violation location was assumed as location where the violation took place. So where ever the violation location missing we considered it as address missing.

4) The 'Violation Time' column has error data such as "8023P". So considering only records with proper time between 0-2400 hrs.

Data Dimensions:

2015 data dimensions: 11809233 51
2016 data dimensions: 10626899 51
2017 data dimensions: 10803028 43

observation: the number of rows in the data are different for the three years and 2017 data has different number of columns

2015 data has records from 1985 to 2015
2016 data has records from 1970 to 2069
2017 data has records from 1972 to 2069

Selected data only from 2015, 2016 and 2017 for further analysis

Examine the data.

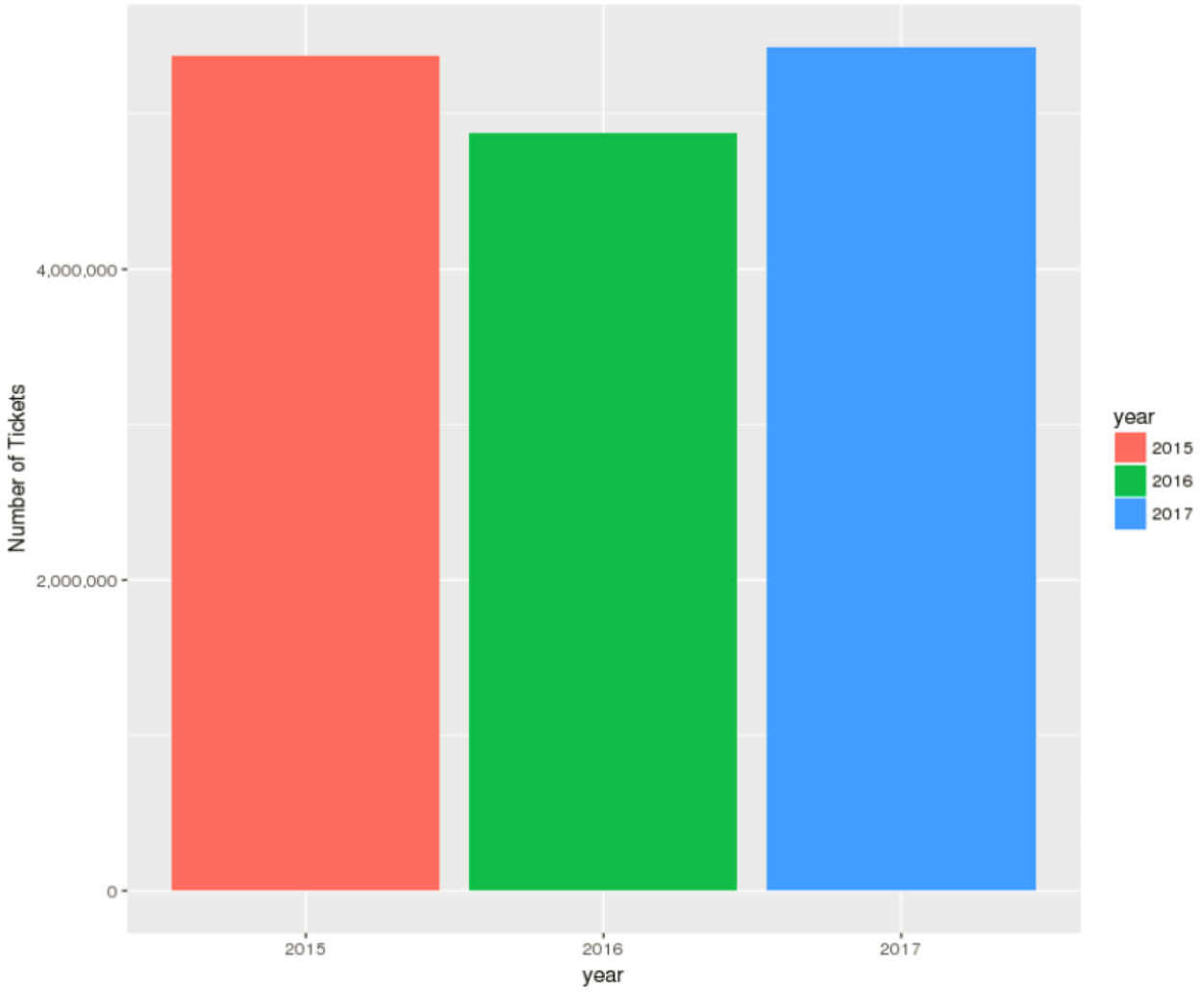
1.Find total number of tickets for each year.

2015 has 5373971 tickets

2016 has 4872621 tickets

2017 has 5431918 tickets

Graph:



Looks like we have more violation happing in 2017 compared to previous years.

Approach taken:

There are no missing or null values in the 'Summons Number` column in the selected data for the three years

`Summons Number` column in 2015 has 586770 duplicates
`Summons Number` column in 2016 has 0 duplicates
`Summons Number` column in 2017 has 0 duplicates

new dimensions

Year	Tickets	Dimensions
2015	5373971	51
2016	4872621	51
2017	5431918	43

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

2015 - 68 (considering 99 also as a state. It has states from Canada)
2016 - 67 (considering 99 also as a state. It has states from Canada)
2017 - 65 (considering 99 also as a state. It has states from Canada)

QUESTION 3. SOME PARKING TICKETS DONT HAVE ADDRESSES ON THEM. FIND OUT HOW MANY TICKETS SUCH THERE ARE.

Assuming address here means the location where the violation took place, finding the missing values in `Violation Location`

Year	Count of Tickets with No Location
2015	721275
2016	828348
2017	925596

AGGREGATION TASKS

QUESTION 1. HOW OFTEN DOES EACH VIOLATION CODE OCCUR? (FREQUENCY OF VIOLATION CODES - FIND THE TOP 5)

2015

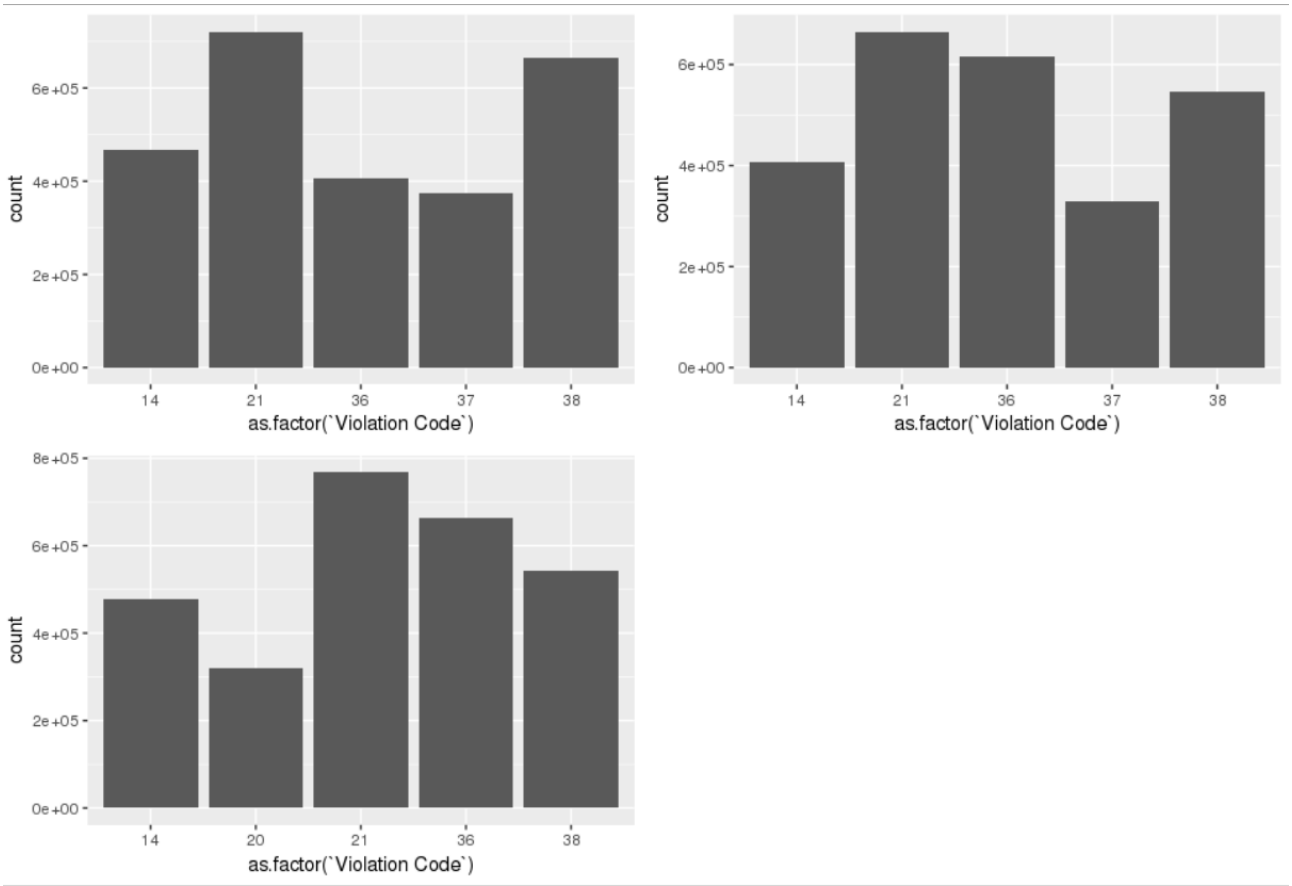
Top 5	Violation Code	count
1	21	720902
2	38	663904
3	14	466488
4	36	406249
5	37	373229

2016

Top 5	Violation Code	count
1	21	664947
2	36	615242
3	38	547080
4	14	405885
5	37	330489

2017

Top 5	Violation Code	count
1	21	768087
2	36	662765
3	38	542079
4	14	476664
5	20	319646



QUESTION 2. HOW OFTEN DOES EACH VEHICLE BODY GET PARKING TICKET ? (FIND THE TOP 5)

2015

Top 5	Vehicle Body Type	Count
1	SUBN	1715517
2	4DSD	1514580
3	VAN	795457
4	DELV	419548
5	SDN	209381

2016

Top 5	Vehicle Body Type	Count
1	SUBN	1596326
2	4DSD	1354001
3	VAN	722234
4	DELV	354388
5	SDN	178954

2017

Top 5	Vehicle Body Type	Count
1	SUBN	1883954
2	4DSD	1547312
3	VAN	724029
4	DELV	358984

5	SDN	194197

SUBN model has been booked for parking ticket consistently in last 3 years.

HOW ABOUT VEHICLE MAKE ? (FIND THE TOP 5)

2015

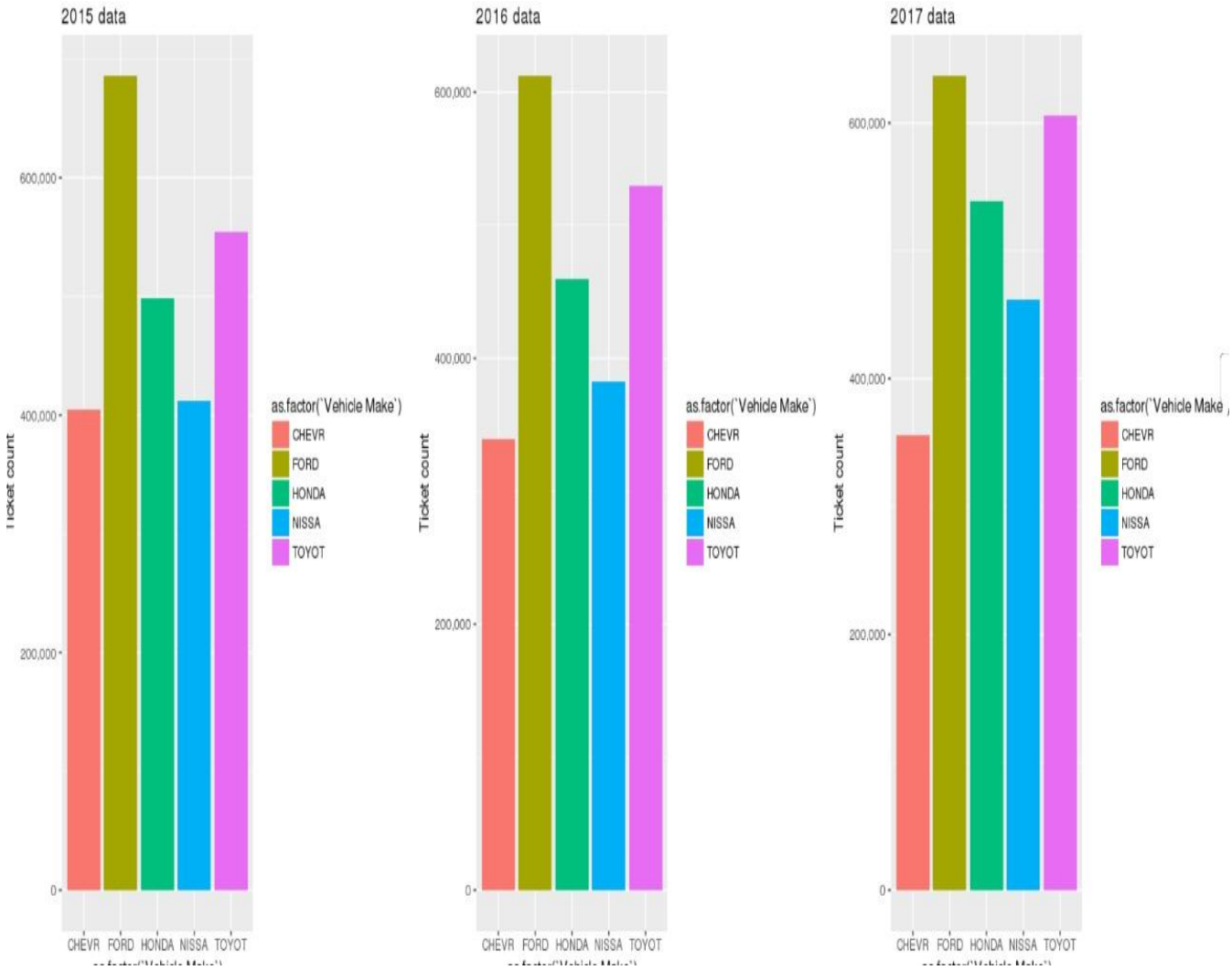
Top 5	Vehicle Make	Count
1	FORD	685900
2	TOYOT	554392
3	HONDA	498858
4	NISSA	411857
5	CHEVR	404841

2016

Top 5	Vehicle Make	Count
1	FORD	612276
2	TOYOT	529115
3	HONDA	459469
4	NISSA	382082
5	CHEVR	339466

2017

Top 5	Vehicle Make	Count
1	FORD	636844
2	TOYOT	605291
3	HONDA	538884
4	NISSA	462017
5	CHEVR	356032



QUESTION 3. FIND THE (5 HIGHEST) FREQUENCIES OF:
 # VIOLATING PRECINCTS

2015

Top 5	Violation Precinct	**Count**
1	0	721275
2	19	287403
3	14	197011
4	18	193593
5	1	152040

Looks like there is no precinct 0. Not sure if this is an error in the data. Since it has highest count not removing it.

2016

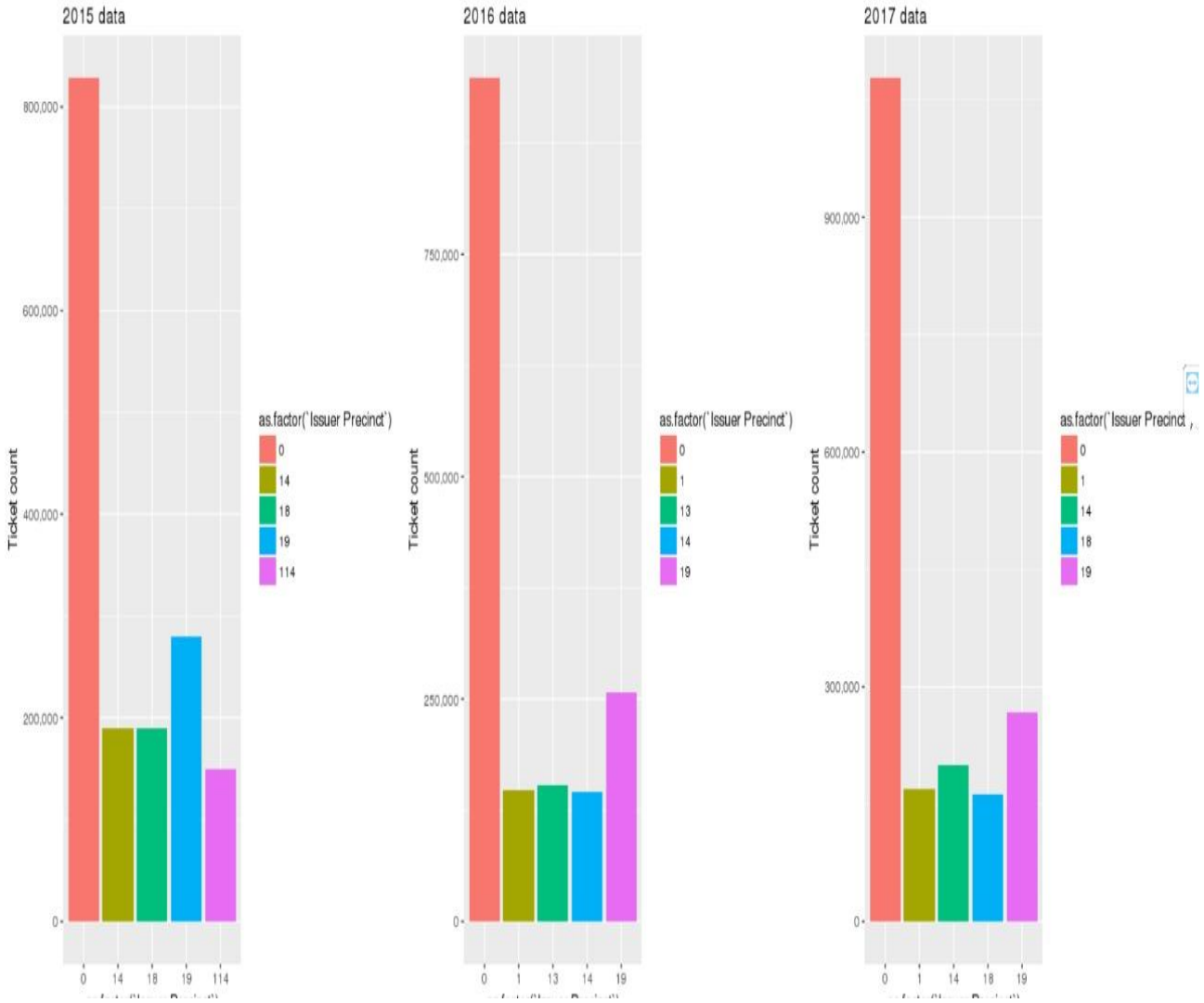
Top 5	Violation Precinct	**Count**
1	0	828348
2	19	264299
3	13	156144
4	1	152231
5	14	150637

Looks like there is no precinct 0. Not sure if this is an error in the data. Since it has highest count not removing it.

2017

Top 5	Violation Precinct	**Count**
1	0	925596
2	19	274445
3	14	203553
4	1	174702
5	18	169131

Looks like there is no precinct 0. Not sure if this is an error in the data. Since it has highest count not removing it.



All years has zero vehicle precincts counts highest, so we have assumed zero has valid vehicle precincts.

ISSUING PRECINCTS

2015

Top 5	Issuing Precinct	**Count**
1	0	828570
2	19	279931
3	14	190403
4	18	190337
5	114	149532

2016

Top 5	Issuing Precinct	**Count**
1	0	948438
2	19	258049
3	13	153478
4	1	146987
5	14	146165

2017

Top 5	Issuing Precinct	Count
1	0	1078406
2	19	266961
3	14	200495
4	1	168740
5	18	162994

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?

2015
Issuer Precincts which issued most are 0, 19 and 14

Issue Precinct	Top 5 Violation code	Count
0	36	406249
	7	253730
	21	96218
	5	55192
	66	2343
19	38	45647
	37	40665
	14	31295
	16	29738
	46	27049
14	69	41004
	14	38696
	31	20676
	47	14480
	42	14446

2016
Issuer Precincts which issued most are 0, 19 and 13

Issue Precinct	Top 5 Violation code	Count
0	36	615242
	7	165111
	21	104351
	5	43467
	66	3821
19	37	38052
	38	37855
	46	36442
	14	28772
	21	25588
13	69	23356
	47	17532
	38	16447
	14	15812
	37	13589

2017
Issuer Precincts which issued most are 0, 19 and 14

Issue Precinct	Top 5 Violation code	Count
0	36	662765
	7	210175
	21	126053
	5	48076
	66	5258
19	46	48445
	38	36386
	37	36056
	14	29797
	21	28415
14	14	45036
	69	30464
	31	22555
	47	18364
	42	10027

QUESTION 5. FIND PROPERTIES OF PARKING VIOLATIONS ACROSS DIFFERENT TIMES OF THE DAY. FIND WAY TO DEAL WITH MISSING VALUES.
 # DIVIDE 24 HRS INTO EQUAL BINS AND FIND VIOLATIONS OCCURRING. ALSO FOR A PARTICULAR VIOLATION FIND COMMONLY OCCURRING TIMES

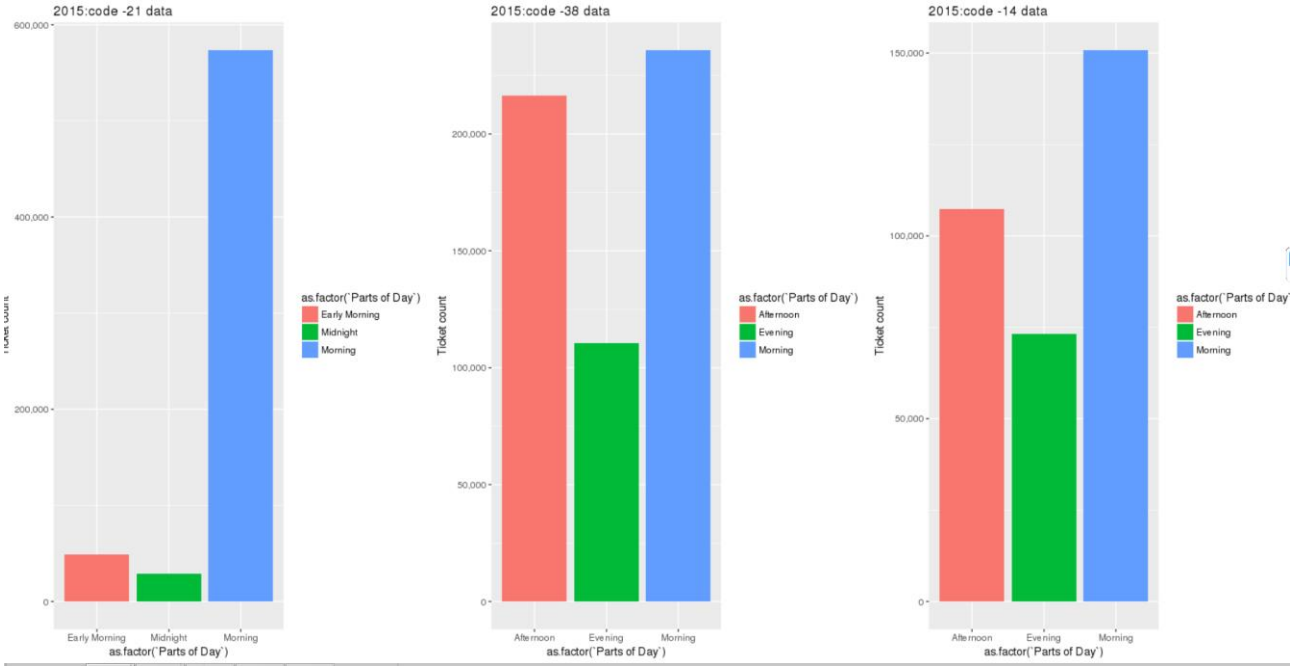
2015

Top 3 violation code in time of day

Time of Day	Top 3 Violation Code	Count
Midnight	21	29139
	40	18032
	78	15342
For Early Morning	14	69606
	21	49094
	40	46967
For Morning	21	573751
	38	235926
	36	188843
For Afternoon	38	216312
	37	163607
	36	122363
For Evening	38	110512
	37	83159
	14	73131
For Night	7	30013
	38	28448
	40	22661

Violation code in times of day:

Violation code	Time of day	Count
21	Morning	573751
	Early Morning	49094
	Midnight	29139
38	Morning	235926
	Afternoon	216312
	Evening	110512
14	Morning	150832
	Afternoon	107317
	Evening	73131



2016

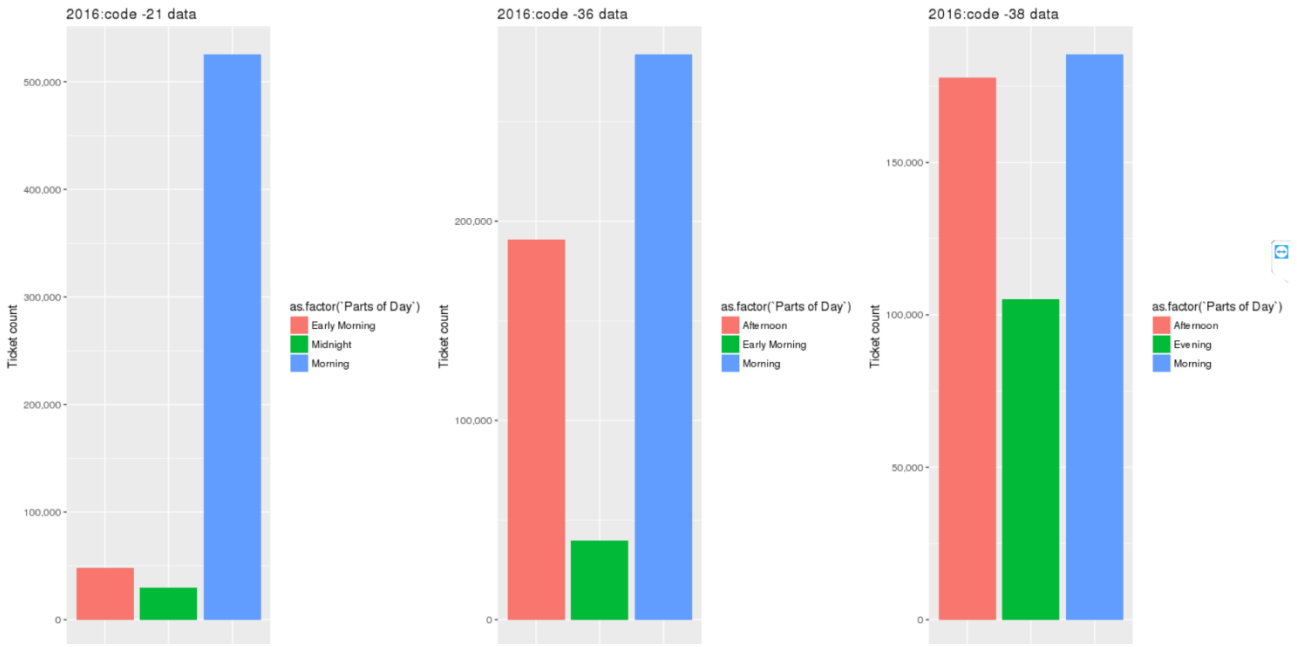
Top 3 violation code in time of day

Time of Day	Top 3 Violation Code	Count
Midnight	21	29927
	40	17105
	78	13354
For Early Morning	14	65793
	21	48224
	40	42437
For Morning	21	525293
	36	283605
	38	185376
For Afternoon	36	190892
	38	177881
	37	144199
	38	105053

For Evening	37	79460
	14	63559
For Night	38	21589
	7	20297
	40	20184

Violation code in times of day:

Violation code	Time of day	Count
21	Morning	525293
	Early Morning	48224
	Midnight	29927
36	Morning	283605
	Afternoon	190892
	Early Morning	39803
38	Morning	185376
	Afternoon	177881
	Evening	105053



2017

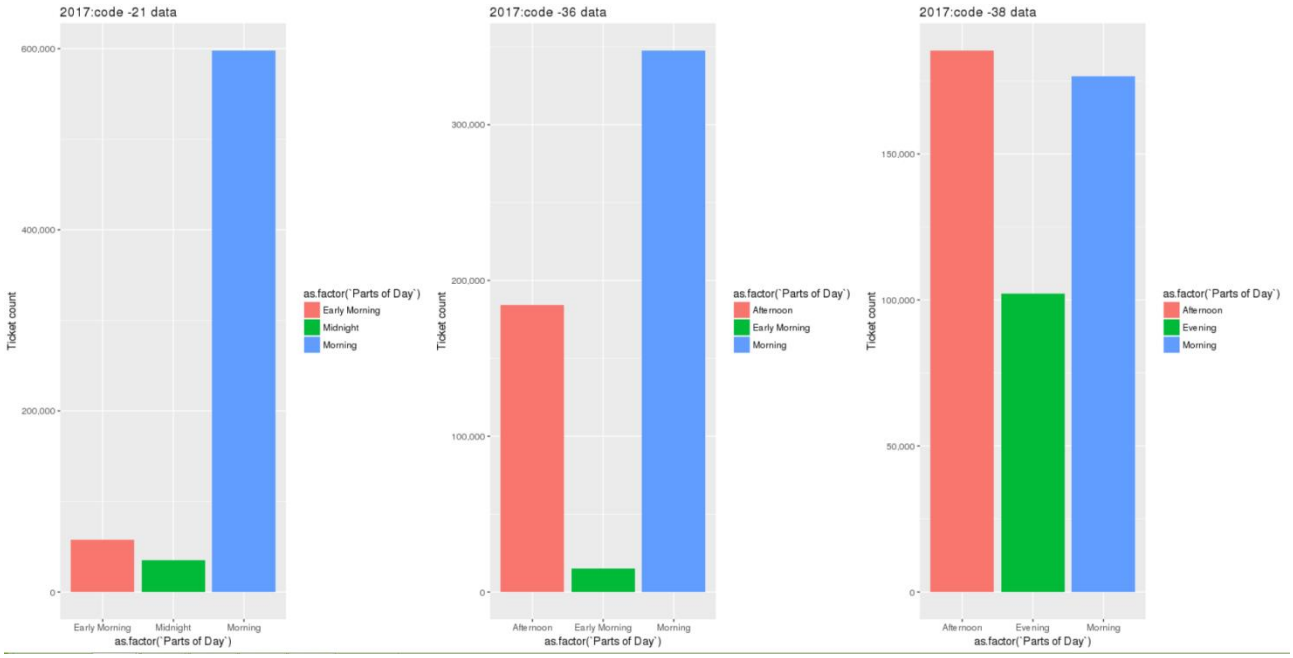
Top 3 violation code in time of day

Time of Day	Top 3 Violation Code	Count
Midnight	21	34751
	40	23751
	14	14208
For Early Morning	14	74685
	40	60817
	21	57889
For Morning	21	598063
	36	347672
	38	176558
	38	185447

For Afternoon	36	184476
	37	131190
For Evening	38	102293
	14	75529
	37	69862
For Night	7	26360
	40	22541
	14	21540

Violation code in times of day:

Violation code	Time of day	Count
21	Morning	598063
	Early Morning	57889
	Midnight	34751
36	Morning	347672
	Afternoon	184476
	Early Morning	15275
38	Afternoon	185447
	Morning	176558
	Evening	102293



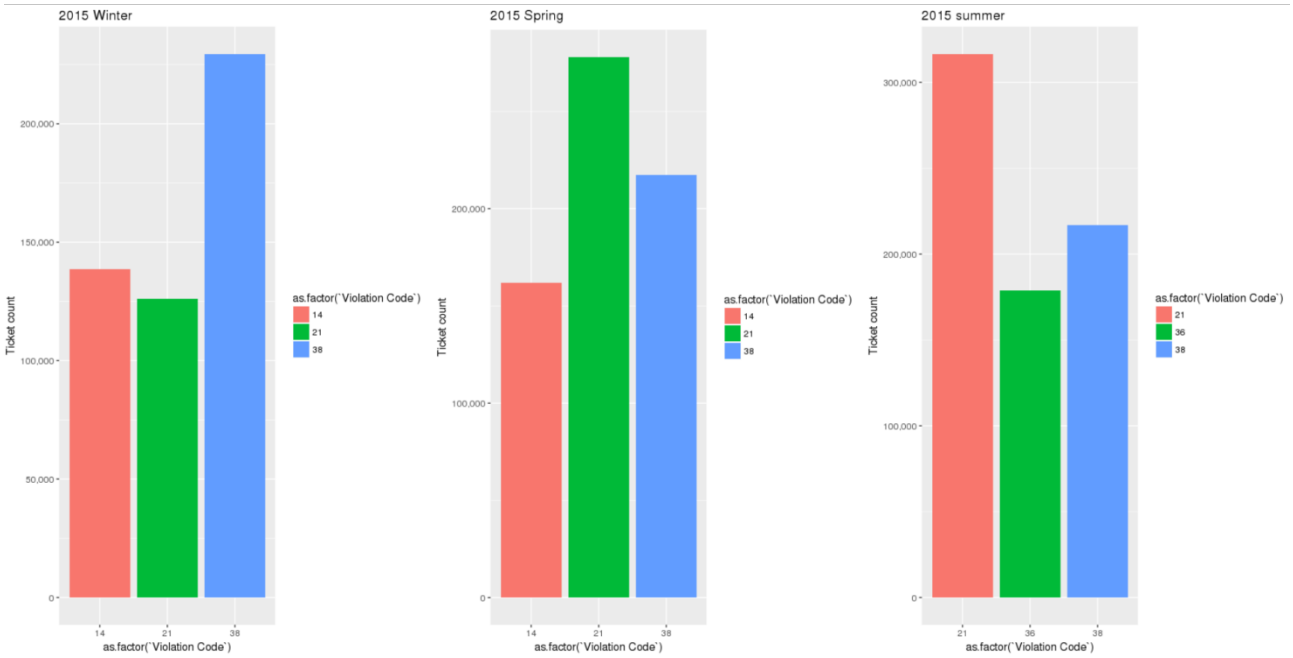
 # QUESTION 6. DIVIDE THE YEAR INTO SOME NUMBER OF SEASONS, AND FIND FREQUENCIES OF TICKETS FOR EACH SEASON. FIND 3 MOST COMMON VIOLATIONS FOR EACH SEASON

2015

Seasons	Count
Summer	1989078

Spring	1875996
Winter	1508897
Autumn	No Tickets

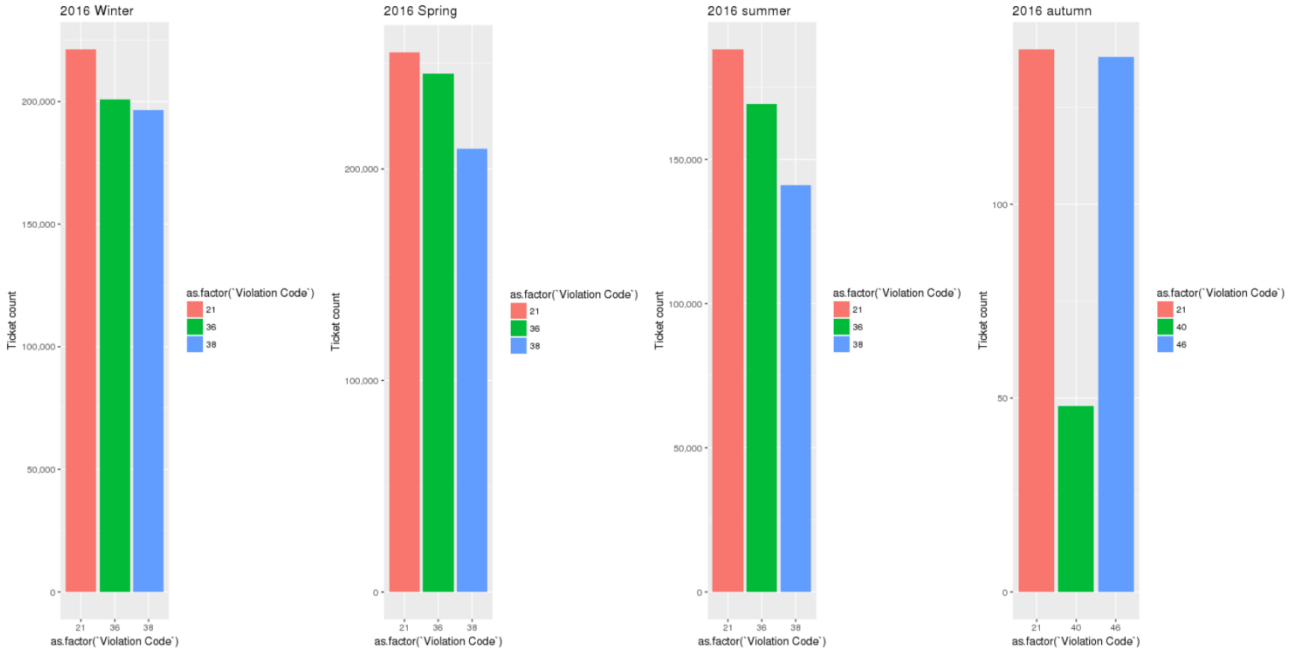
Season	**Violation Code**	**Count**
Winter	38	229493
	14	138696
	21	126269
Spring	21	278056
	38	217319
	14	161784
Summer	21	316577
	38	217092
	36	178682



2016

Seasons	Count
Sprint	1914597
Winter	1655128
Summer	1302265
Autmn	631

Season	Violation Code	Count
Winter	21	221352
	36	200971
	38	196560
Spring	21	255234
	36	245050
	38	209416
Summer	21	188221
	36	169221
	38	141100
Autmn	21	140
	46	138
	40	48

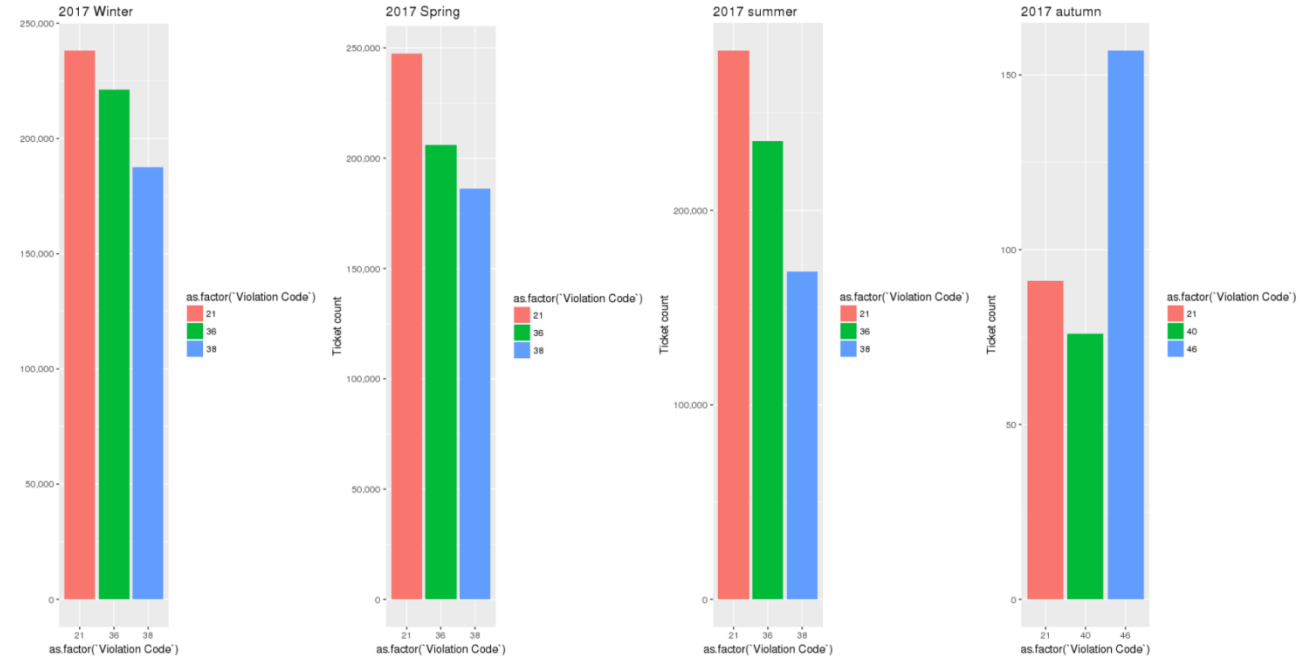


2017

Seasons	Count
Summer	1873110
Spring	1853139
Winter	1705028
Autmn	641

Season	Violation Code	Count
Autmn	46	157
	21	91
	40	76
Summer	21	282222
	36	235544
	38	168563
Spring	21	247554

	36	205953
	38	186122
Winter	21	238220
	36	221268
	38	187388



 # QUESTION 7. FIND TOTAL OCCURRENCES OF THE 3 MOST COMMON VIOLATION CODES. FIND THE TOTAL AMOUNT COLLECTED FOR ALL THE FINES.

#2015
From Question 1 we know that for 2015 the top Violation Codes are 21, 38 and 14. Their average fines are $55,$50 and $115 respectively.

Violation.Code	Count	Fine Amount
14	466488	53646120
21	720902	39649610
38	663904	33195200

14 highest collection

#2016

From Question 1 we know that for 2016 the top Violation Codes are 21, 36 and 38. Their average fines are $55, $50 and $50 respectively.

Violation Code	Count	Fine Amount
21	664947	36572085
36	615242	30762100
38	547080	27354000

21 highest collection

#2017

From Question 1 we know that for 2017 the top Violation Codes are 21, 36 and 38. Their average fines are $55, $50 and $50 respectively.

21 highest collection

Violation.Code	Count	Fine Amount
21	768087	42244785
36	662765	33138250
38	542079	27103950