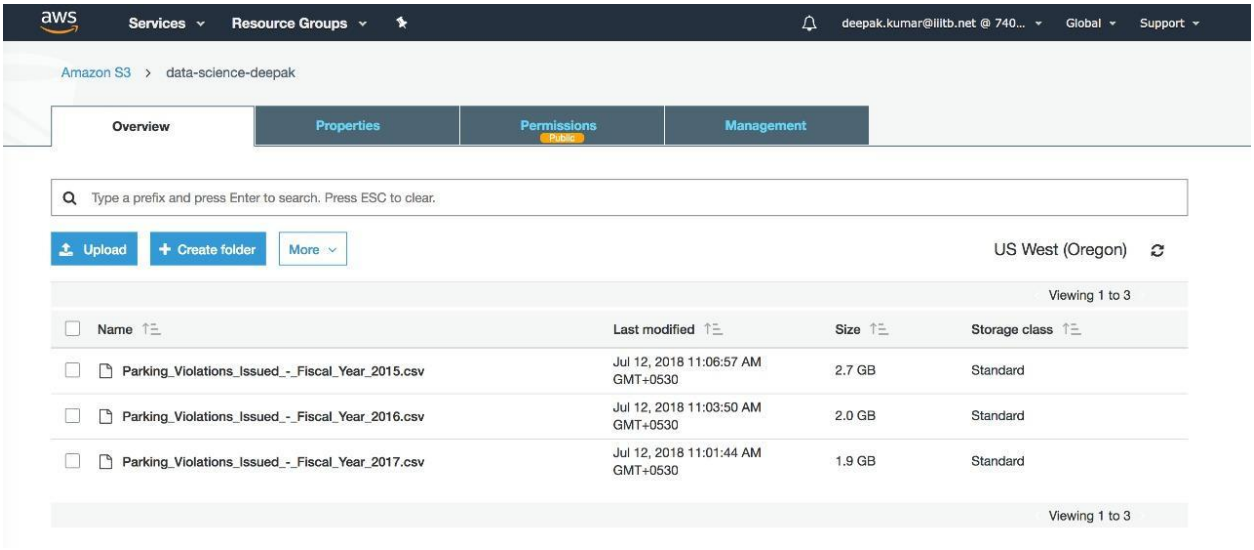
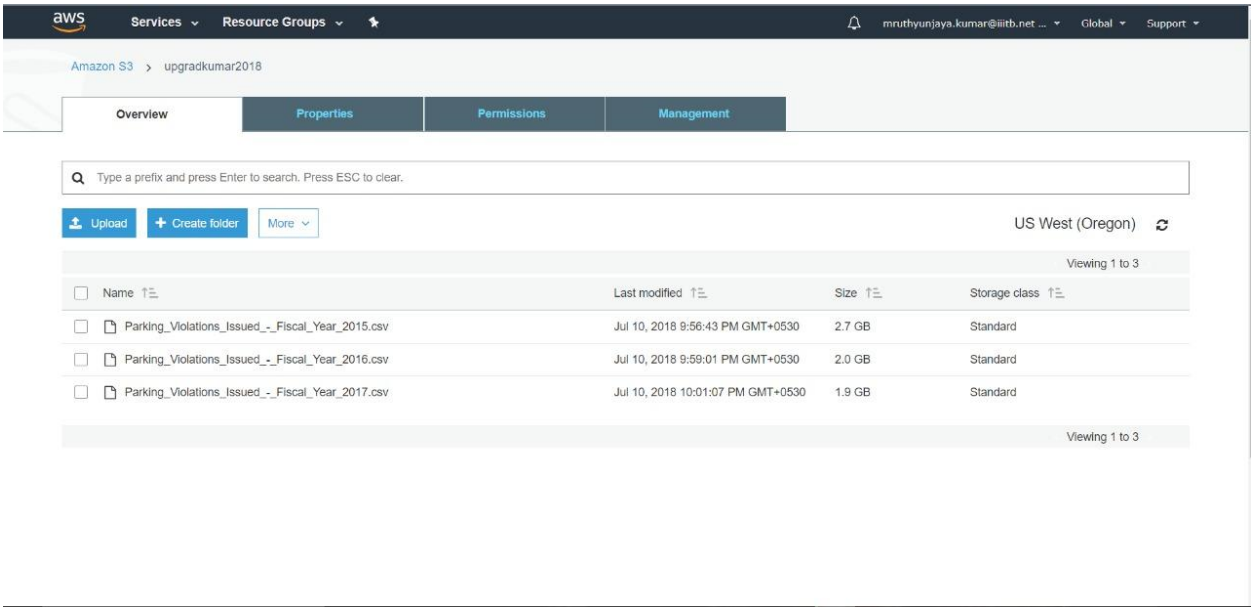
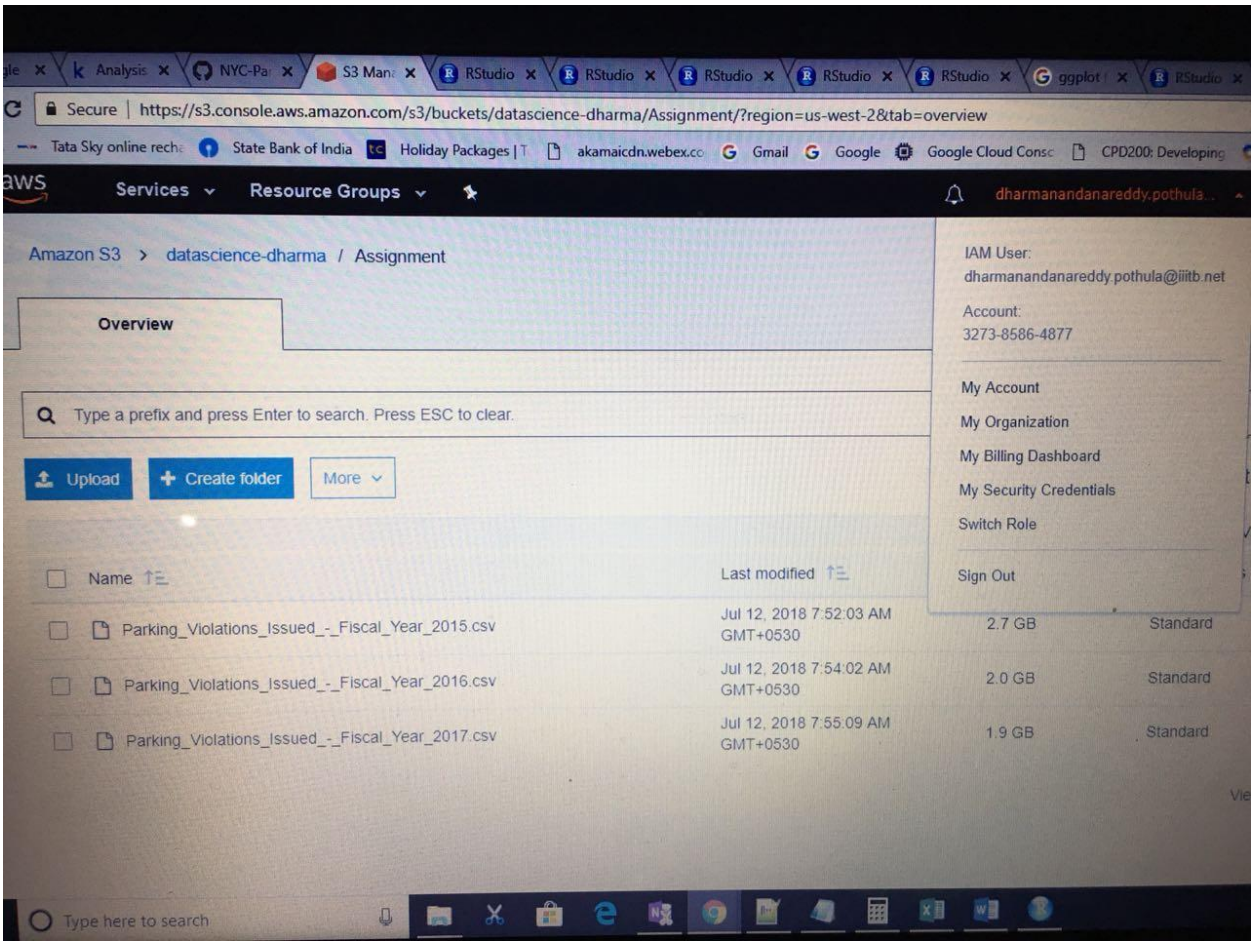
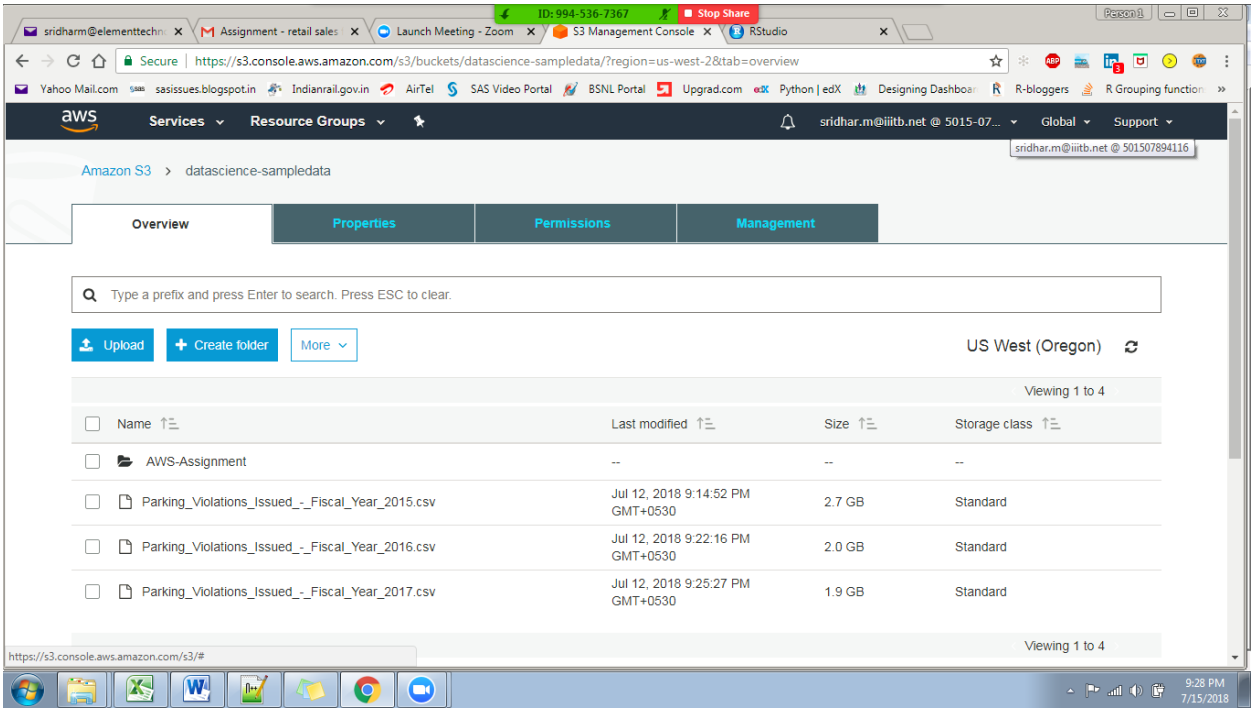
Screenshots showing files and user IDs are not present for all 4 members of the team






Secure | https://s3.console.aws.amazon.com/s3/buckets/datascience-dharma/Assignment/?region=us-west-2&tab=overview
Services
Resource Groups
dharmanandanareddy.pothula...
Amazon S3 > datascience-dharma / Assignment
Overview
Type a prefix and press Enter to search. Press ESC to clear.
Upload
Create folder
More
IAM User:
dharmanandanareddy.pothula@iiitb.net
Account:
3273-8586-4877
My Account
My Organization
My Billing Dashboard
My Security Credentials
Switch Role
Sign Out
Name
Last modified
Parking_Violations_Issued_-_Fiscal_Year_2015.csv
Jul 12, 2018 7:52:03 AM GMT+0530
2.7 GB
Standard
Parking_Violations_Issued_-_Fiscal_Year_2016.csv
Jul 12, 2018 7:54:02 AM GMT+0530
2.0 GB
Standard
Parking_Violations_Issued_-_Fiscal_Year_2017.csv
Jul 12, 2018 7:55:09 AM GMT+0530
1.9 GB
Standard
Type here to search



Examine the data.

1. Find total number of tickets for each year.

2015: 10951257 (51 columns)

2016: 10626899 (51 Columns)

2017: 10803028

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

2015: 68 (99 excluded)

2016: 68 (99 excluded)

2017: 67

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

2015: 5463

2016: 8274

2017: 4009

Aggregation tasks

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

2015

#ViolationCode count

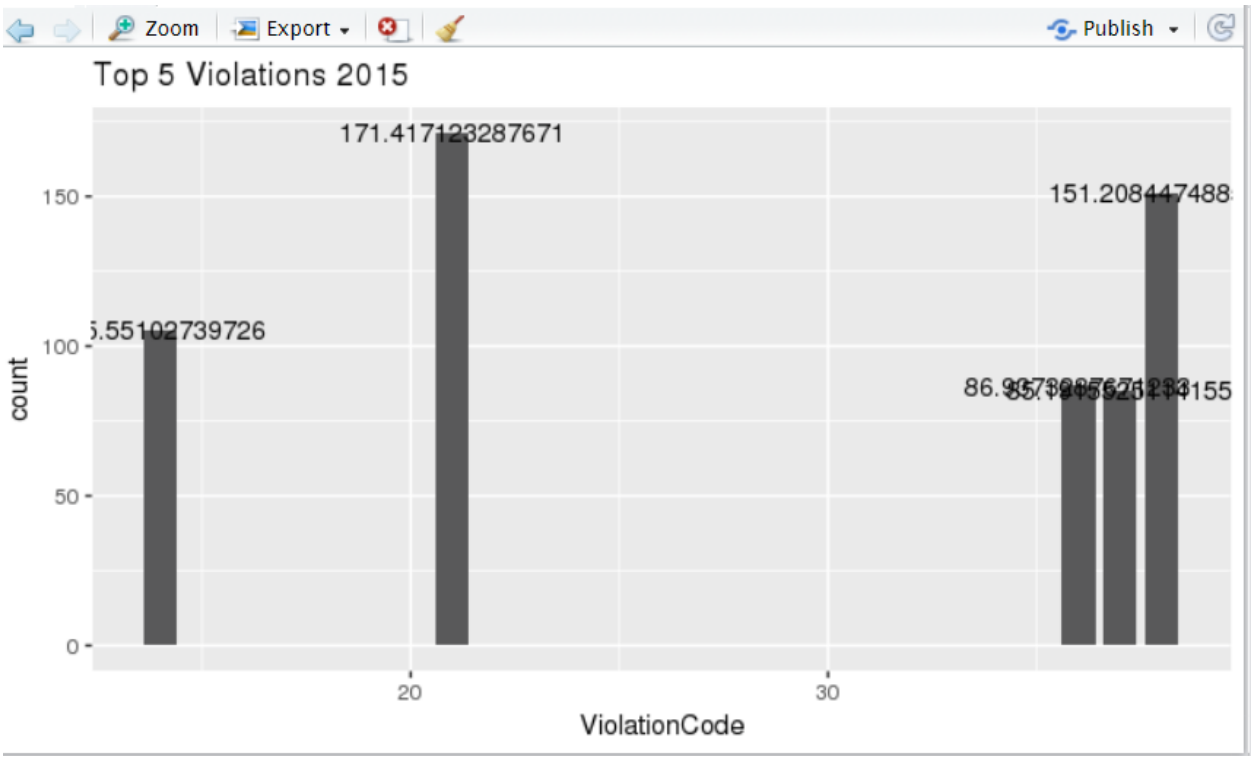
#1 21 171.41712

#2 38 151.20845

#3 14 105.55103

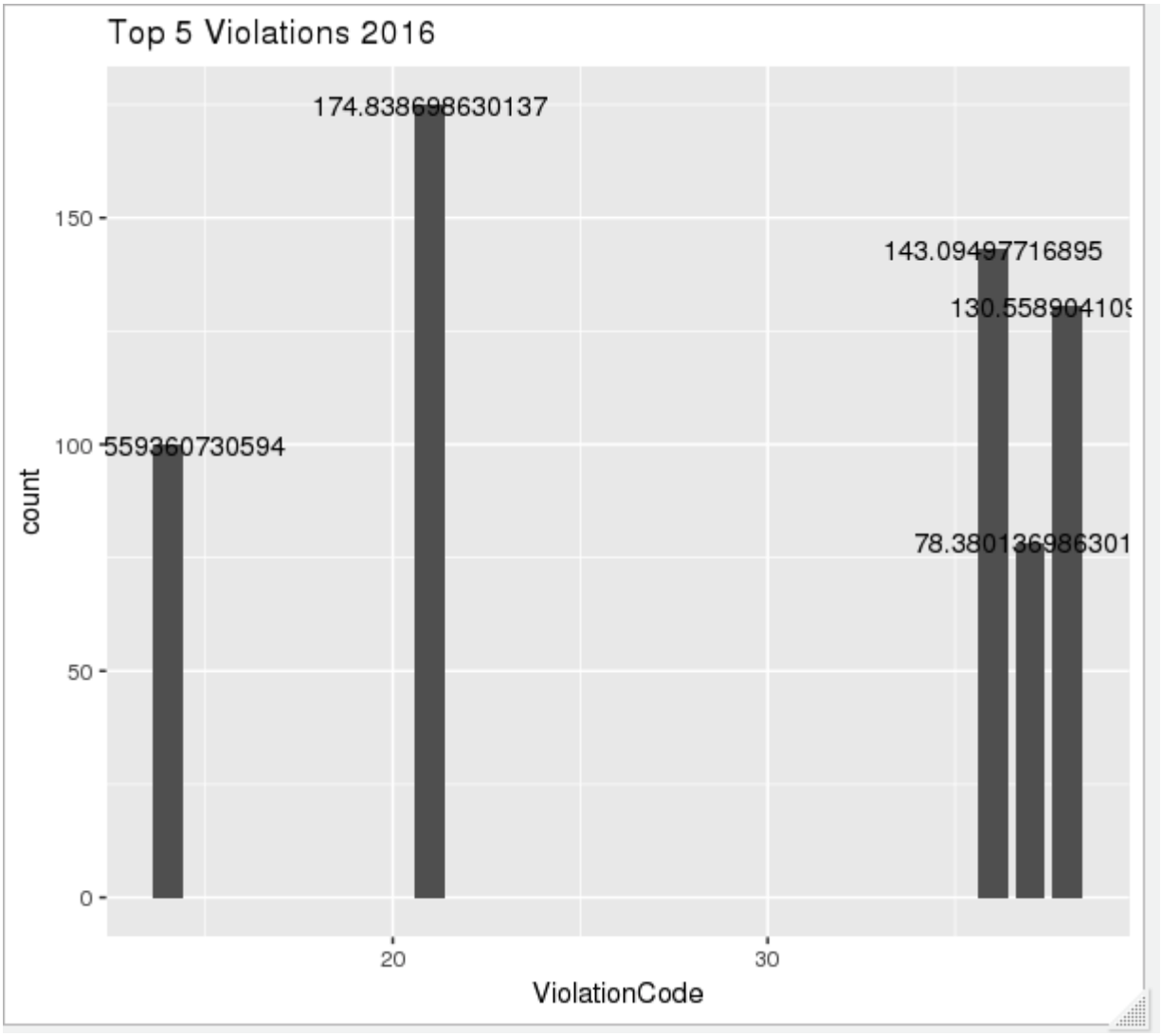
#4 36 86.93733

#5 37 85.19155



2016

```
#ViolationCode     count
#             21 174.83870
#             38 130.55890
#             14  99.95594
#             36 143.09498
#             37  78.38014
```



2017

ViolationCode count

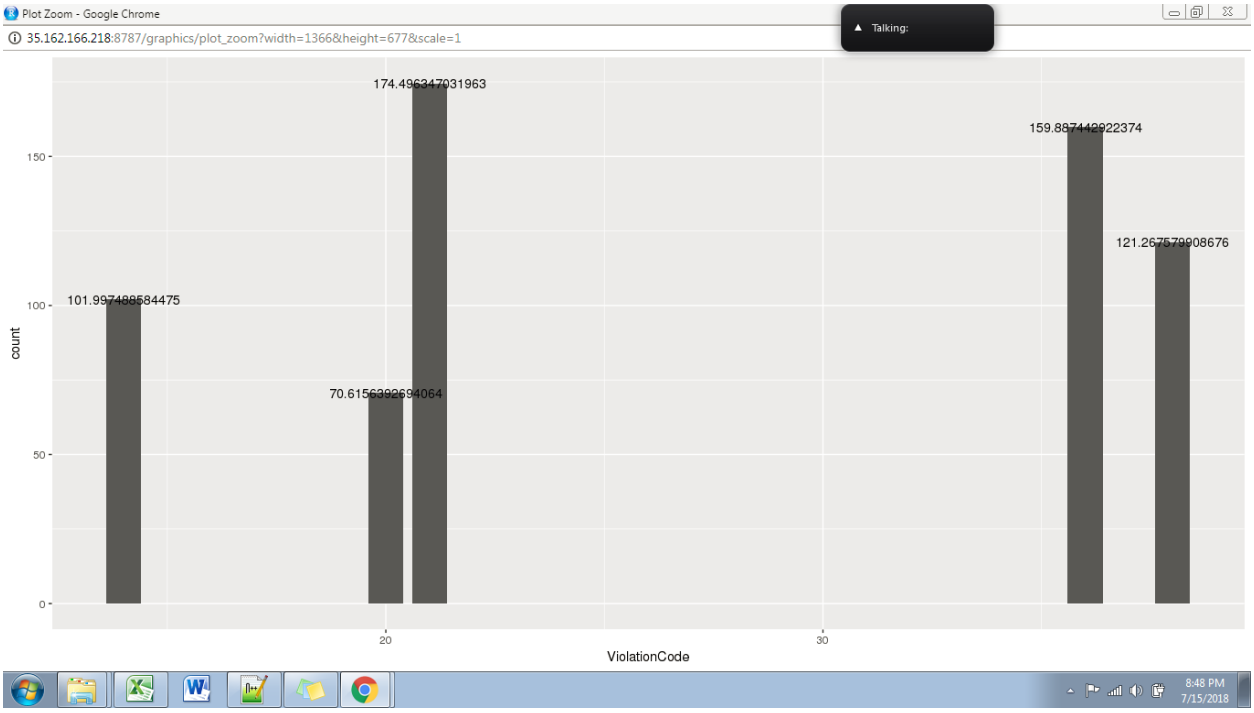
#1 21 174.49635

#2 36 159.88744

#3 38 121.26758

#4 14 101.99749

#5 20 70.61564



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

2015

VehicleBodyType count

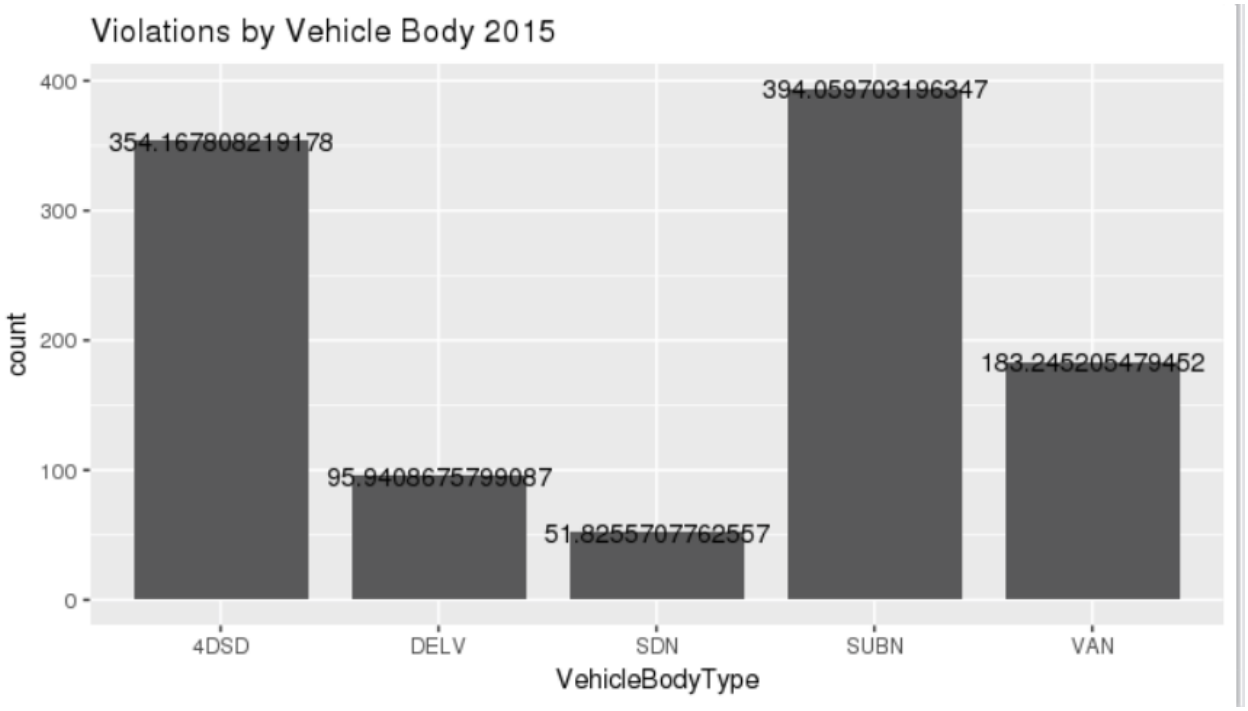
1 SUBN 394.05970

2 4DSD 354.16781

3 VAN 183.24521

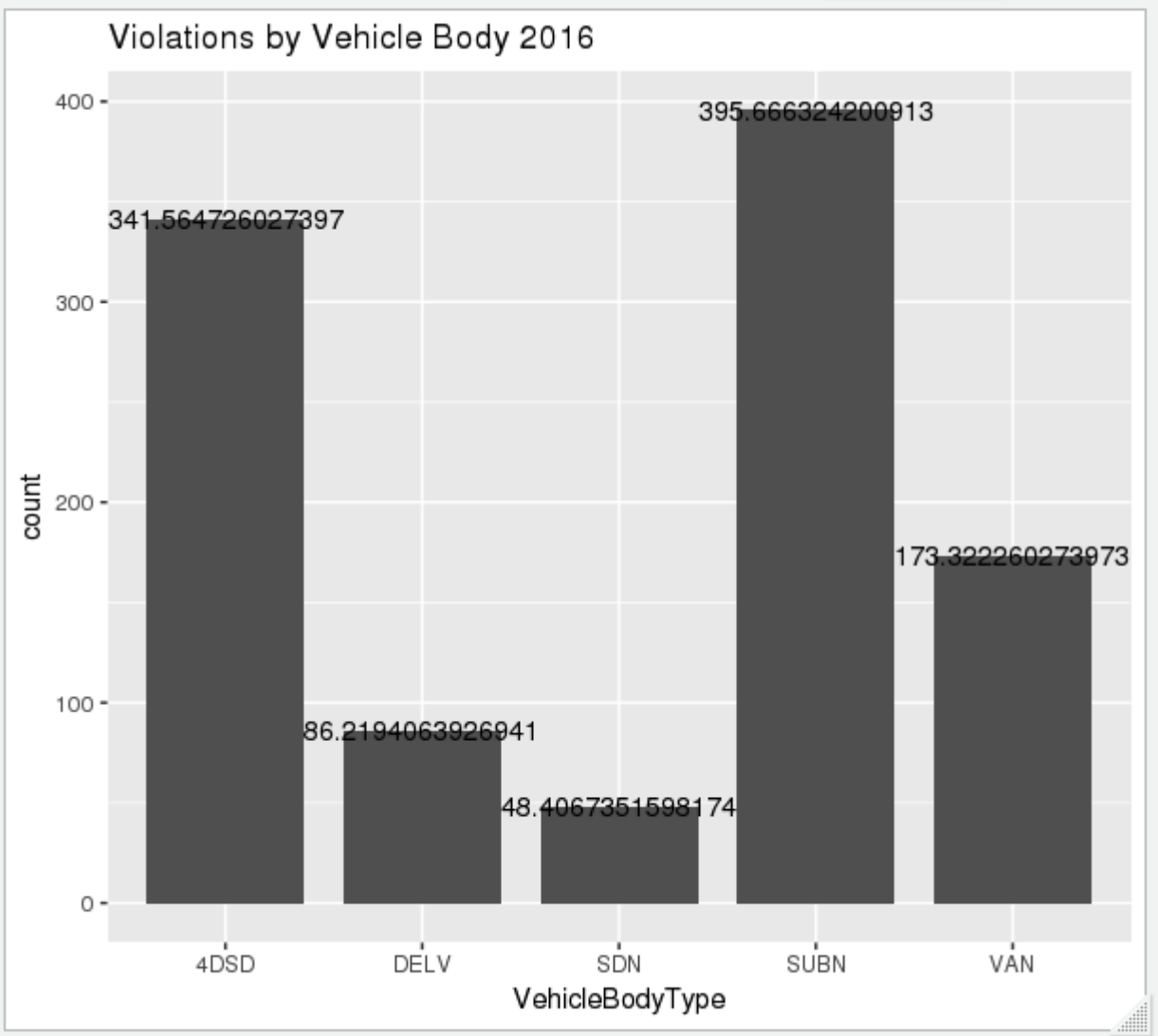
4 DELV 95.94087

5 SDN 51.82557



2016

```
#VehicleBodyType     count
#1             SUBN 395.66632
#2             4DSD 341.56473
#3              VAN 173.32226
#4             DELV  86.21941
#5              SDN  48.40674
```



2017

VehicleBodyType count

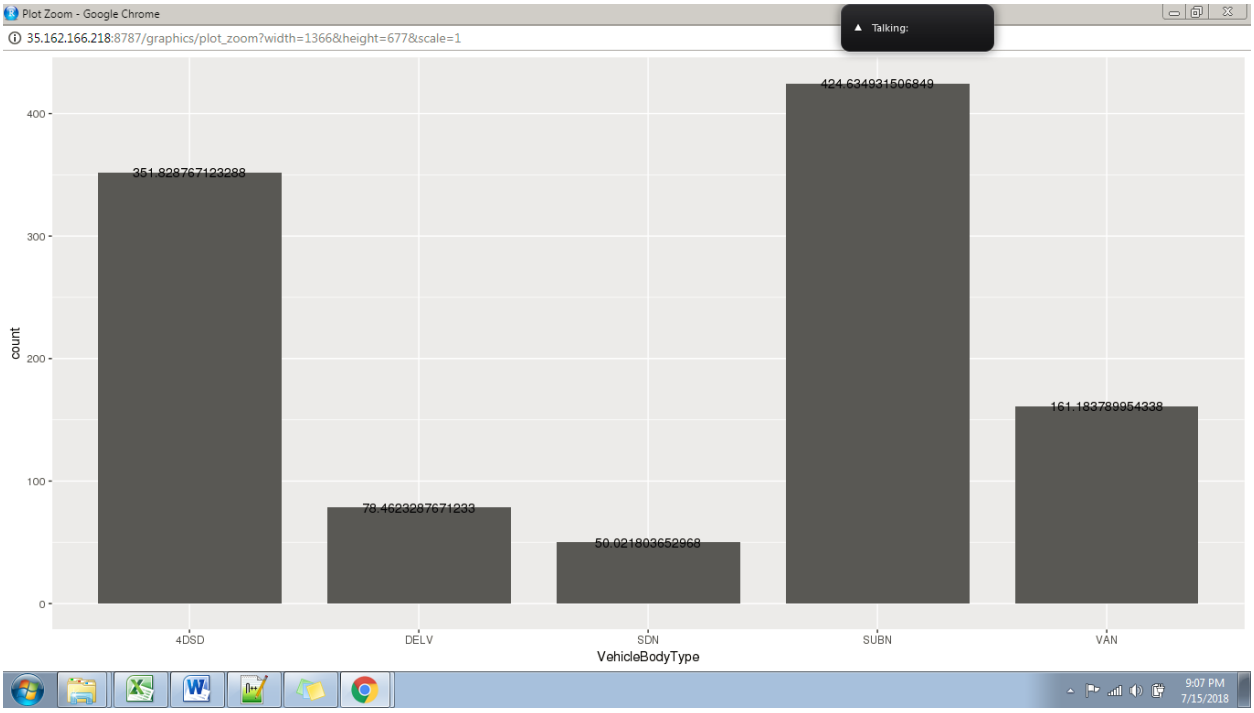
#1 SUBN 424.63493

#2 4DSD 351.82877

#3 VAN 161.18379

#4 DELV 78.46233

#5 SDN 50.02180



Vehicle Make Violations

2015 (Ford)

VehicleMake count

1 FORD 161.7926

2 TOYOT 128.2561

3 HONDA 116.2156

4 NISSA 95.6129

5 CHEVR 95.4782

2016

```
#VehicleMake    count
#1         FORD 151.22991
#2        TOYOT 131.82534
#3        HONDA 115.76187
#4        NISSA  95.30057
#5        CHEVR  86.71952
```

2017

VehicleMake count

#1 FORD 146.22808

#2 TOYOT 138.29349

#3 HONDA 123.20068

#4 NISSA 104.86187

#5 CHEVR 81.58162

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)

2015

#ViolationPrecinct count

#1 0 1633006

#2 19 559716

#3 18 400887

#4 14 384596

#5 1 307808

#IssuerPrecinct count

#1 0 1834343

#2 19 544946

#3 18 391501

#4 14 369725

#5 1 298594

2016

#ViolationPrecinct count

#1 0 1868655

#2 19 554465

#3 18 331704

#4 14 324467

#5 1 303850

#IssuerPrecinct count

#1 0 1834343

#2 19 544946

#3 18 391501

#4 14 369725

#5 1 298594

2017

ViolationPrecinct count

#1 0 2072400

#2 19 535671

#3 14 352450

#4 1 331810

#5 18 306920

IssuerPrecinct count

#1 0 2388479

#2 19 521513

#3 14 344977

#4 1 321170

#5 18 296553

---

[1] 3206545

#ViolationCode count

#1 36 1400614

#2 7 516389

#3 21 326255

---

[1] 3206545

---

[1] 2914991

ViolationCode count

#1 36 1400614

#2 7 516388

#3 14 154780

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

#ViolationPrecinct count

#1 0 1619126

#2 25 9

#3 18 3

#4 104 3

#5 7 2

#6 106 2

#7 14 2

#8 26 1

#9 44 1

#10 122 1

#11 47 1

#12 1 1

#13 13 1

#14 48 1

#15 41 1

#16 33 1

#17 10 1

#18 77 1

#19 102 1

#20 50 1

#21 113 1

#22 121 1

#23 75 1

#24 71 1

#25 123 1

#26 901 1

#27 66 1

2016

#ViolationPrecinct	count
#1	0 1619126
#2	25 9
#3	18 3
#4	104 3
#5	7 2
#6	106 2
#7	14 2
#8	26 1
#9	44 1
#10	122 1
#11	47 1
#12	1 1

```
#13                 13       1
#14                 48       1
#15                 41       1
#16                 33       1
#17                 10       1
#18                 77       1
#19                102       1
#20                 50       1
#21                113       1
#22                121       1
#23                 75       1
#24                 71       1
#25                123       1
#26                901       1
#27                 66       1
```

2017

[1] 2062664

ViolationPrecinct count

#1 0 2062644

#2 13 5

#3 7 3

#4 1 2

#5 78 2

#6 14 2

#7 34 1

#8 6 1

#9 17 1

#10 84 1

#11 70 1

#12 10 1

---

5. 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.
- Find a way to deal with missing values, if any.
- 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
- 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)

2015

#bin_number ViolationCode count

#1 1 21 63574

#2 1 40 36490

#3 1 78 34842

#1 2 14 134458

#2 2 21 106858

#3 2 40 91344

#bin_number ViolationCode count

#1 3 21 1192163

#2 3 38 449070

#3 3 36 360365

#bin_number ViolationCode count

#1 4 38 432287

#2 4 37 324905

#3 4 36 220663

#1 5 38 241327

#2 5 37 175802

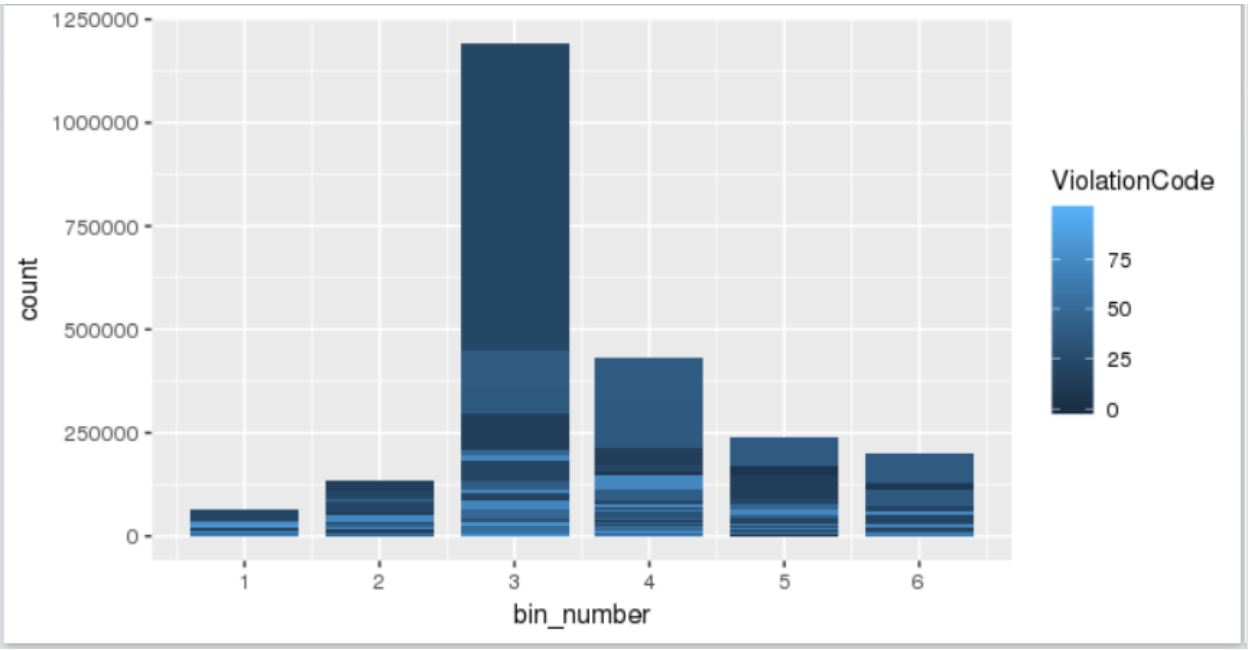
#3 5 7 168888

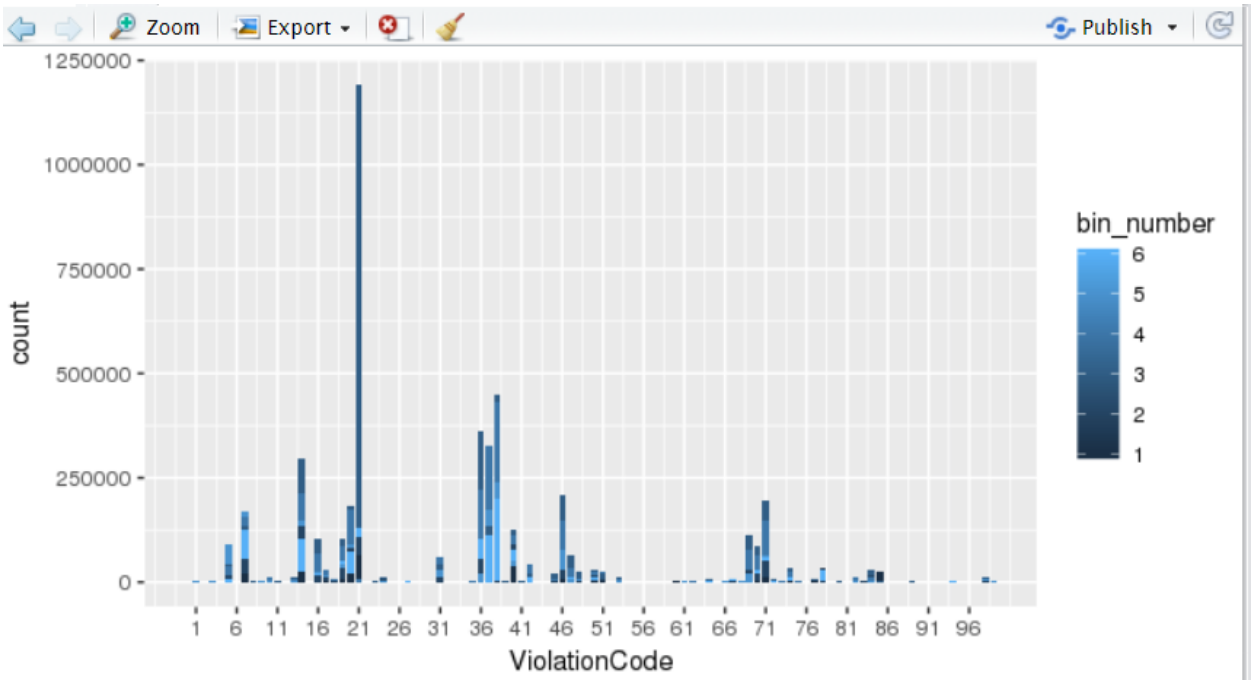
#bin_number ViolationCode count

#1 6 38 198472

#2 6 21 130164

#3 6 7 124456





2016

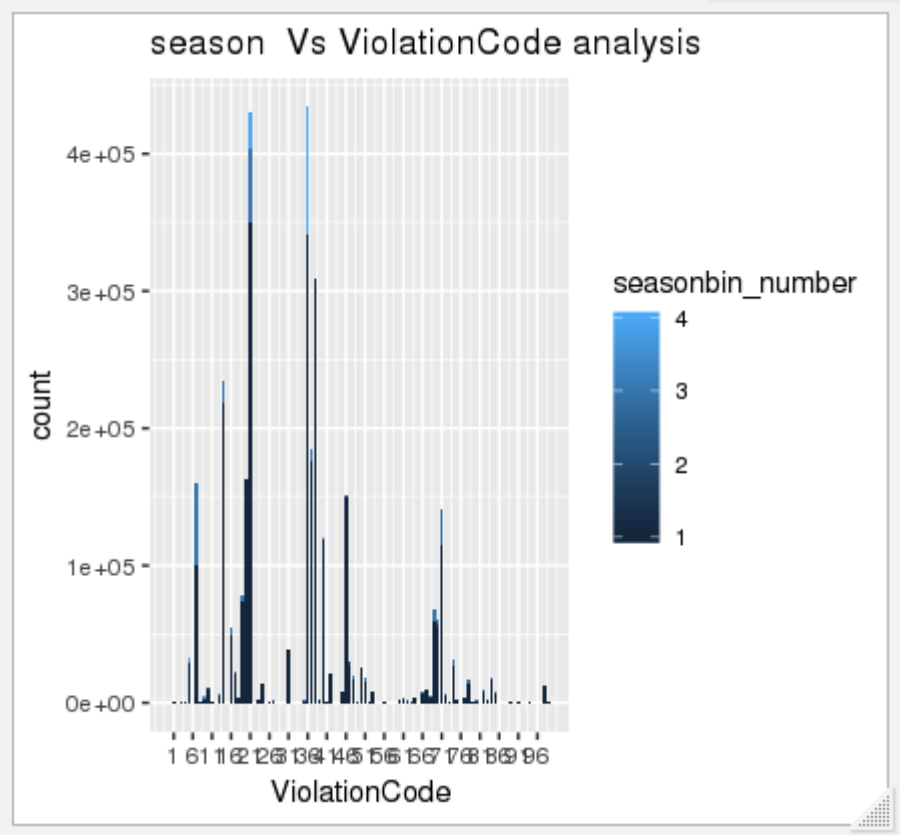
bin_number	ViolationCode	count
1	21	63574
1	40	36490
1	78	34842
1	14	26545
1	85	24865
2	14	134458
2	21	106858
2	40	91344
2	20	81103
2	36	56550
3	21	1192163
3	38	449070
3	36	360365
3	14	297711
3	46	210978
4	38	432287
4	37	324905
4	36	220663
4	14	211833
4	20	171943
5	38	241327
5	37	175802
5	7	168888
5	14	148538

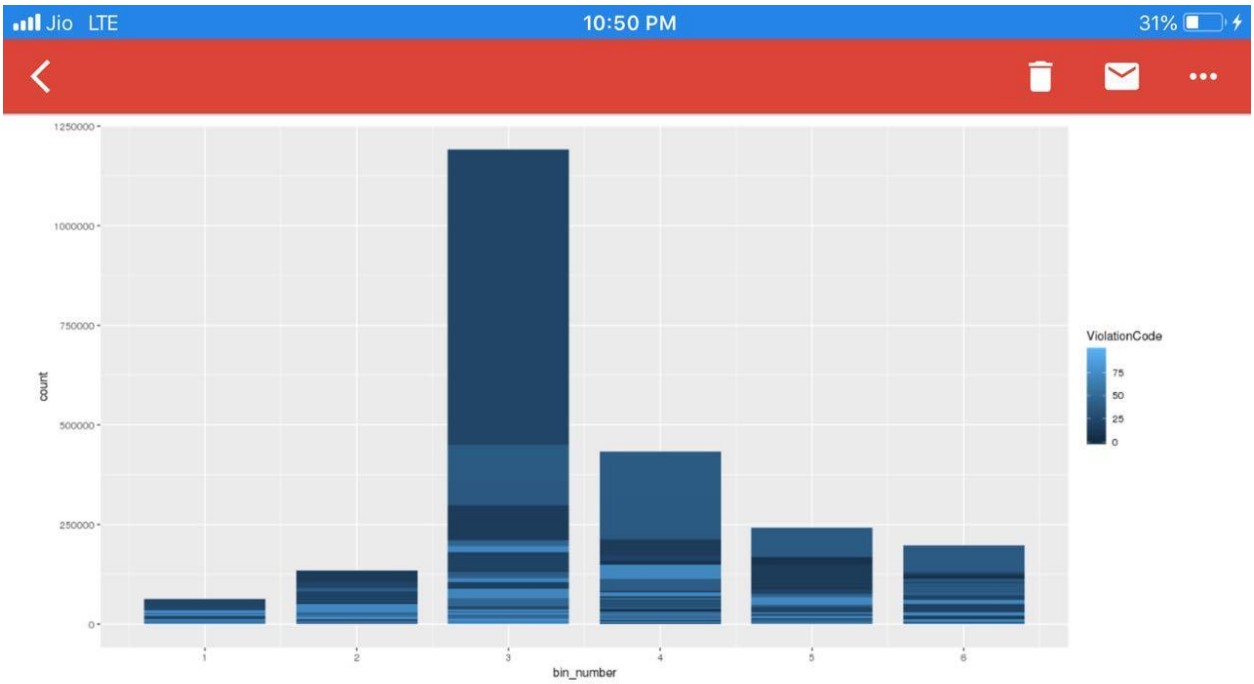
5	5	89709
6	38	198472
6	21	130164
6	7	124456
6	37	112638
6	14	105542

bin_number	ViolationCode	count
3	36	360365
4	36	220663
6	36	102881

bin_number	ViolationCode	count
3	38	449070
4	38	432287
5	38	241327

bin_number	ViolationCode	count
3	21	1192163
6	21	130164
2	21	106858





2017

bin_number ViolationCode count

#1 3 21 1182689

#2 3 36 751422

#3 4 36 376961

#4 4 38 356354

#5 3 38 346518

#6 3 14 274288

---

[1] 572 3

#bin_number ViolationCode count

#1 1 21 73160

#2 1 40 45960

#3 1 14 29311

---

bin_number ViolationCode count

#1 2 14 141276

#2 2 21 119469

#3 2 40 112186

#bin_number ViolationCode count

#1 3 21 1182689

#2 3 36 751422

#3 3 38 346518

#bin_number ViolationCode count

#1 4 36 376961

#2 4 38 356354

#3 4 37 265869

#bin_number ViolationCode count

#1 5 38 203232

#2 5 37 145784

#3 5 14 144749

#bin_number ViolationCode count

#1 6 36 211434

#2 6 38 153537

#3 6 21 144082

---

bin_number ViolationCode count

#1 4 38 356354

#2 3 38 346518

#3 5 38 203232

#bin_number ViolationCode count

#1 3 21 1182689

#2 6 21 144082

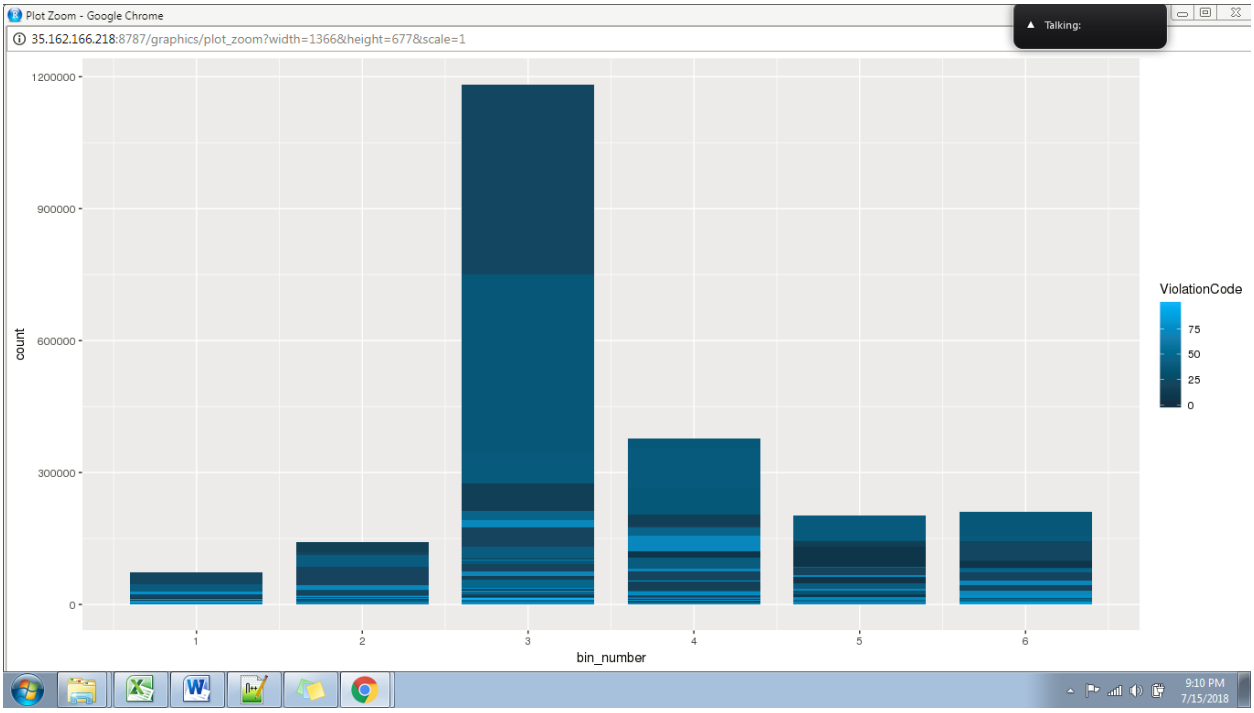
#3 2 21 119469

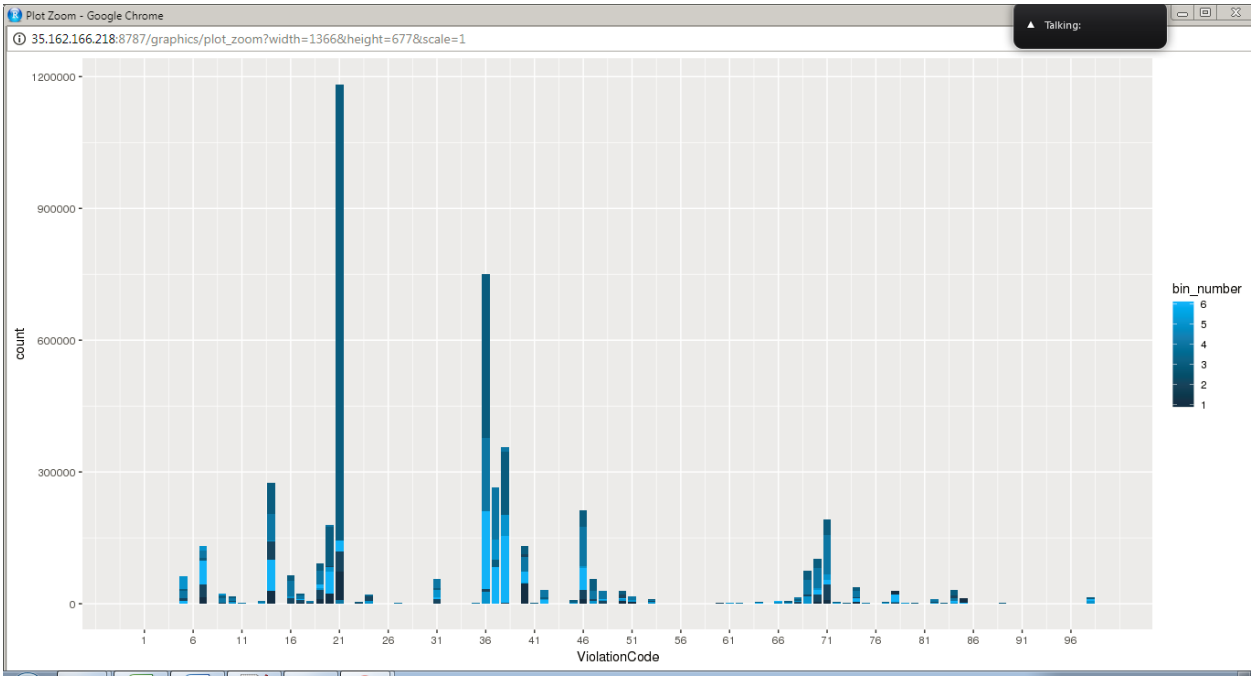
#bin_number ViolationCode count

#1 3 36 751422

#2 4 36 376961

#3 6 36 211434





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

- First, divide the year into some number of seasons, and find frequencies of tickets for each season.
- Then, find the 3 most common violations for each of these season

2015

#seasonbin_number ViolationCode count

#1 1 38 336762

#2 1 21 281600

#3 1 14 220029

#seasonbin_number ViolationCode count

#1 2 21 471580

#2 2 38 346719

#3 2 14 262595

#seasonbin_number ViolationCode count

#1 3 21 397871

#2 3 38 348466

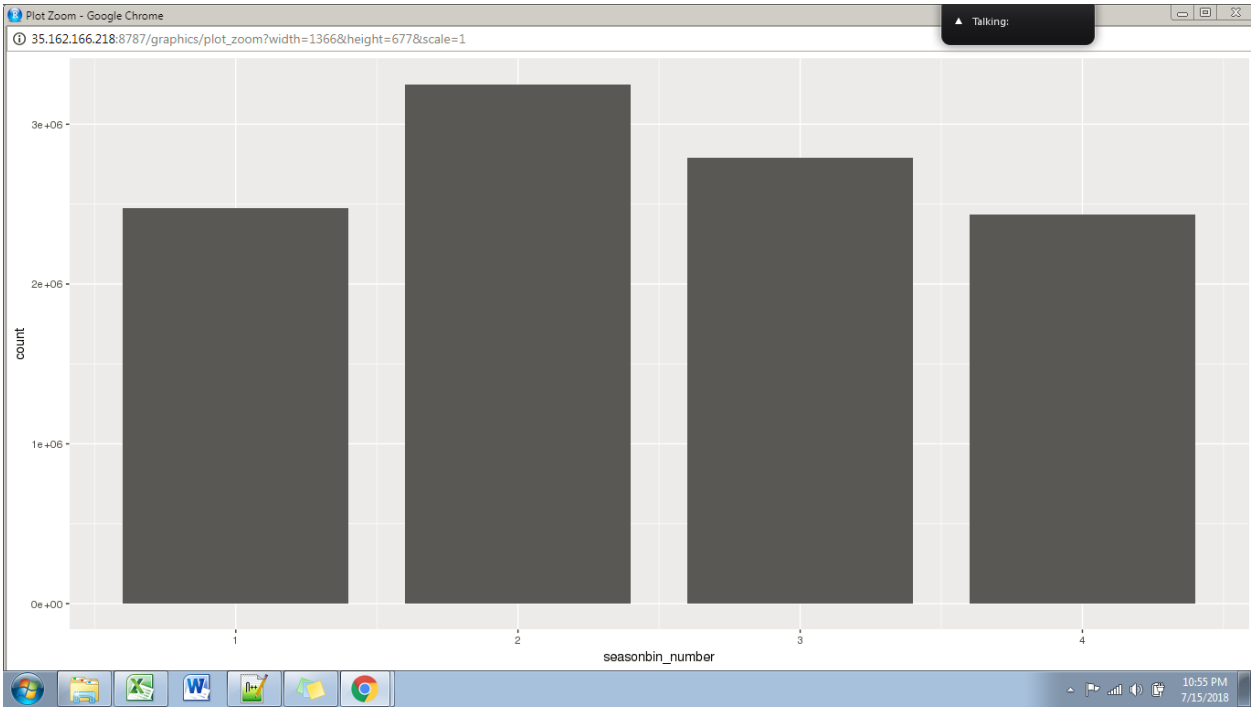
#3 3 14 234606

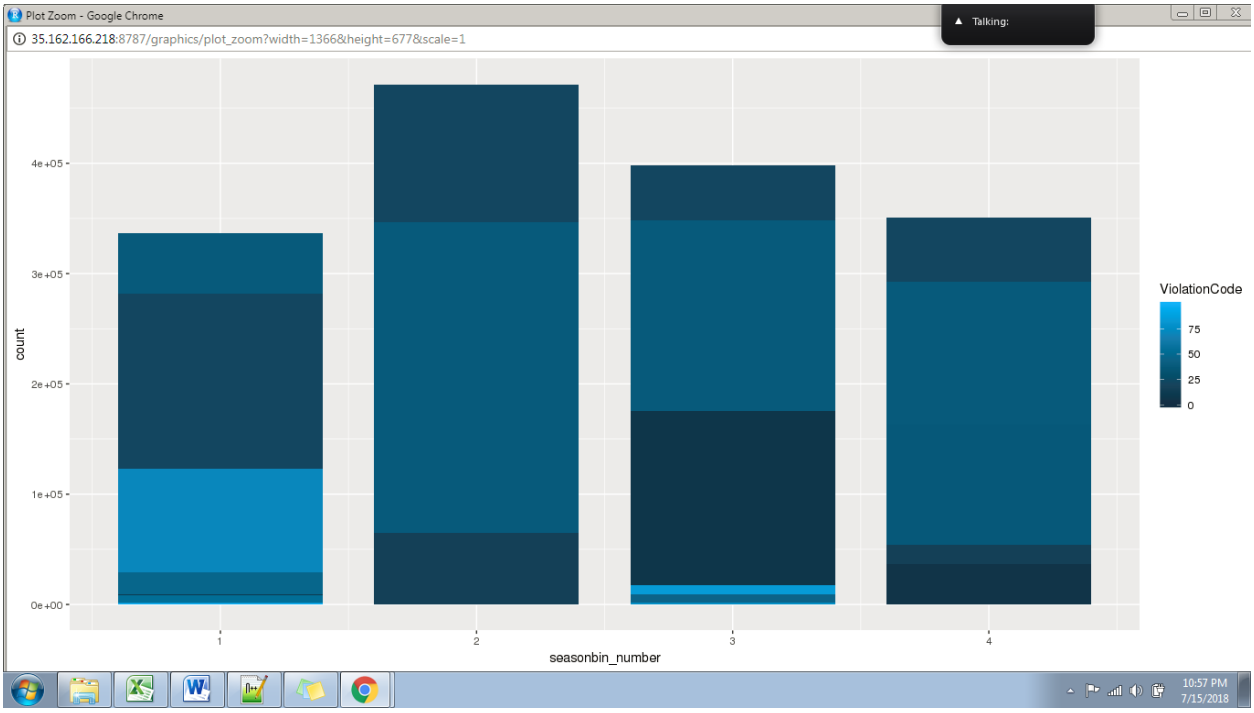
#seasonbin_number ViolationCode count

#1 4 21 350563

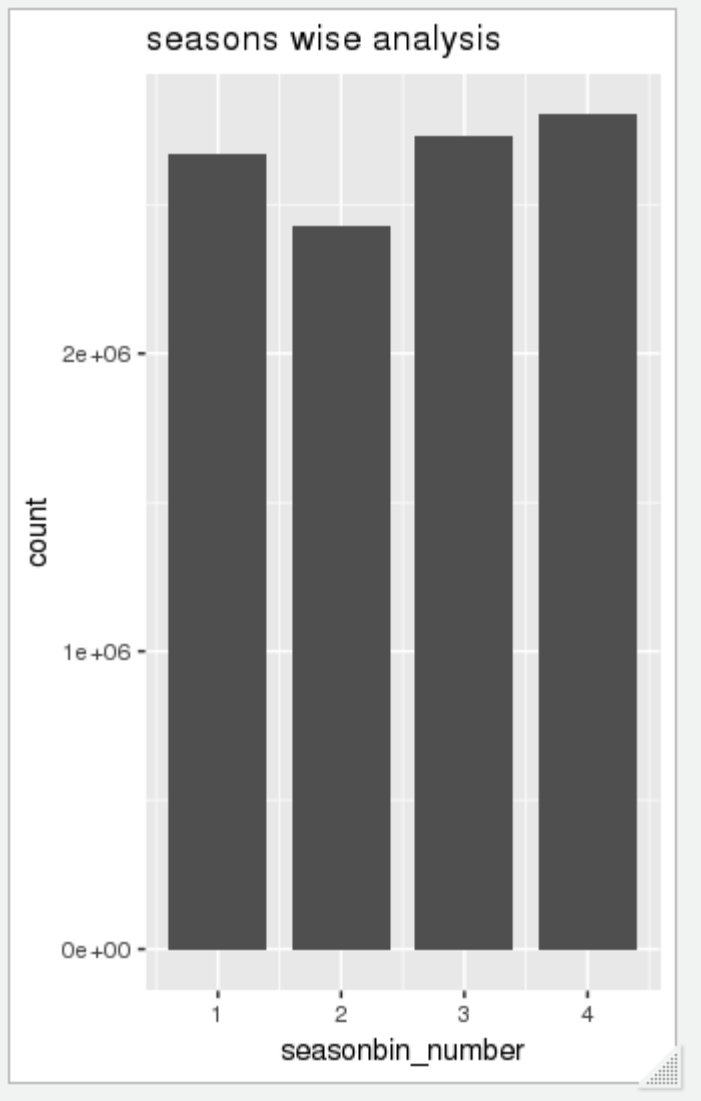
#2 4 38 292639

#3 4 14 207397

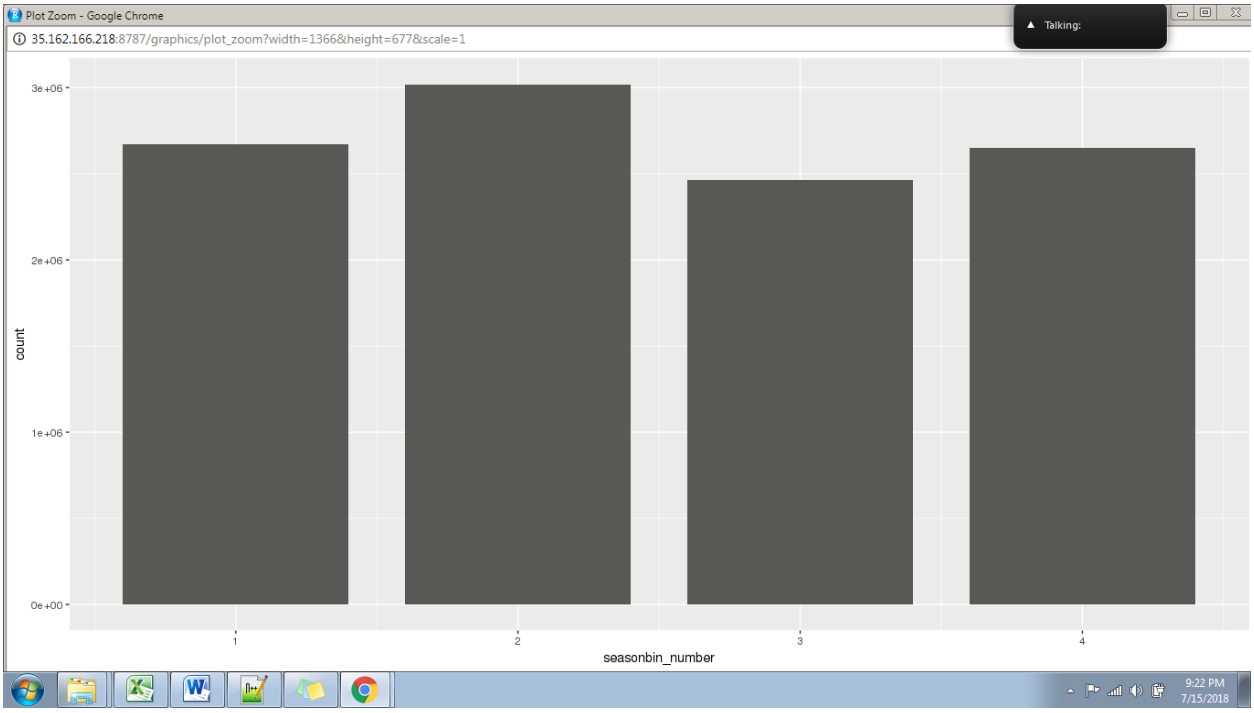



Plot Zoom - Google Chrome
35.162.166.218:8787/graphics/plot_zoom?width=1366&height=677&scale=1
Talking:
4e+05
3e+05
2e+05
1e+05
0e+00
count
1
2
3
4
seasonbin_number
ViolationCode
75
50
25
0
10:57 PM
7/15/2018

2016


seasons wise analysis
count
2e+06
1e+06
0e+00
1
2
3
4
seasonbin_number

2017



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.

- Find total occurrences of the 3 most common violation codes
- 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?

2015

#ViolationCode count fine

#1 21 1501614 75080700

#2 38 1324586 66229300

#3 14 924627 106332105

2016

#ViolationCode count fine

#1 21 1531587 76579350

#2 36 1253512 62675600

#3 38 1143696 131525040

2017

#ViolationCode	count	fine
#1	21 1528588	76429400
#2	36 1400614	70030700
#3	38 1062304	122164960