

Applied Problem Set 2

Qianyu Shao & Sandra Dai

01/11/2021

```
library(tidyverse)
```

Front matter

Add your collaborators: **Sandro**

Late coins used this pset: 2. Late coins left: 8.

Problems

Part 1

Read in one percent sample

```
start.time <- Sys.time()
pt <- read.csv("parking_tickets_one_percent.csv")
end.time <- Sys.time()
end.time-start.time
```

```
## Time difference of 11.21443 secs
```

```
library(testthat)
```

```
## Warning: 'testthat' R 4.1.1
```

```
##
## 'testthat'
```

```
## The following object is masked from 'package:dplyr':
##
## matches
```

```
## The following object is masked from 'package:purrr':
##
## is_null
```

```
## The following objects are masked from 'package:readr':
##
##   edition_get, local_edition
```

```
## The following object is masked from 'package:tidyr':
##
##   matches
```

```
test_that('row numbers correct', {
  expect_equal(nrow(pt), 287458)
})
```

```
## Test passed
```

2.

```
file.info("parking_tickets_one_percent.csv")
```

```
##                                size isdir mode                mtime
## parking_tickets_one_percent.csv 83942807 FALSE  666 2021-11-01 20:33:04
##                                ctime                atime exe
## parking_tickets_one_percent.csv 2021-11-01 20:33:04 2021-11-05 23:20:30 no
```

```
83942807/1048576
```

```
## [1] 80.0541
```

The file is 80.0541 megabytes. Given that this is only the sample containing ticket numbers that end in 01, which is generally 1/100 of full data set. The size of full data set might be 8005.41M.

3. It's ordered by issue date.

4.

```
pt %>% summarise(across(everything(), ~ sum(is.na(.))))
```

```
##   X ticket_number issue_date violation_location license_plate_number
## 1 0              0          0              0              0
##   license_plate_state license_plate_type zipcode violation_code
## 1          97          2054    54115          0
##   violation_description unit_description vehicle_make fine_level1_amount
## 1              0    29              0              0              0
##   fine_level2_amount current_amount_due total_payments ticket_queue
## 1              0              0              0              0
##   ticket_queue_date notice_level hearing_disposition notice_number officer
## 1              0          84068          259899              0          0
##   address
## 1      0
```

#from <https://stackoverflow.com/questions/24027605/determine-the-number-of-na-values-in-a-column>

5. The top three variables with most missing values are zipcode, hearing disposition and notice level. Some of them having “Define” under variable ticket_queue, which means city cannot identify the vehicle owner, thus zipcode, hearing disposition and notice level are all can’t be obtained in this circumstance. Also, variable hearing disposition is blank unless the ticket was contested.

Cleaning the data and benchmarking

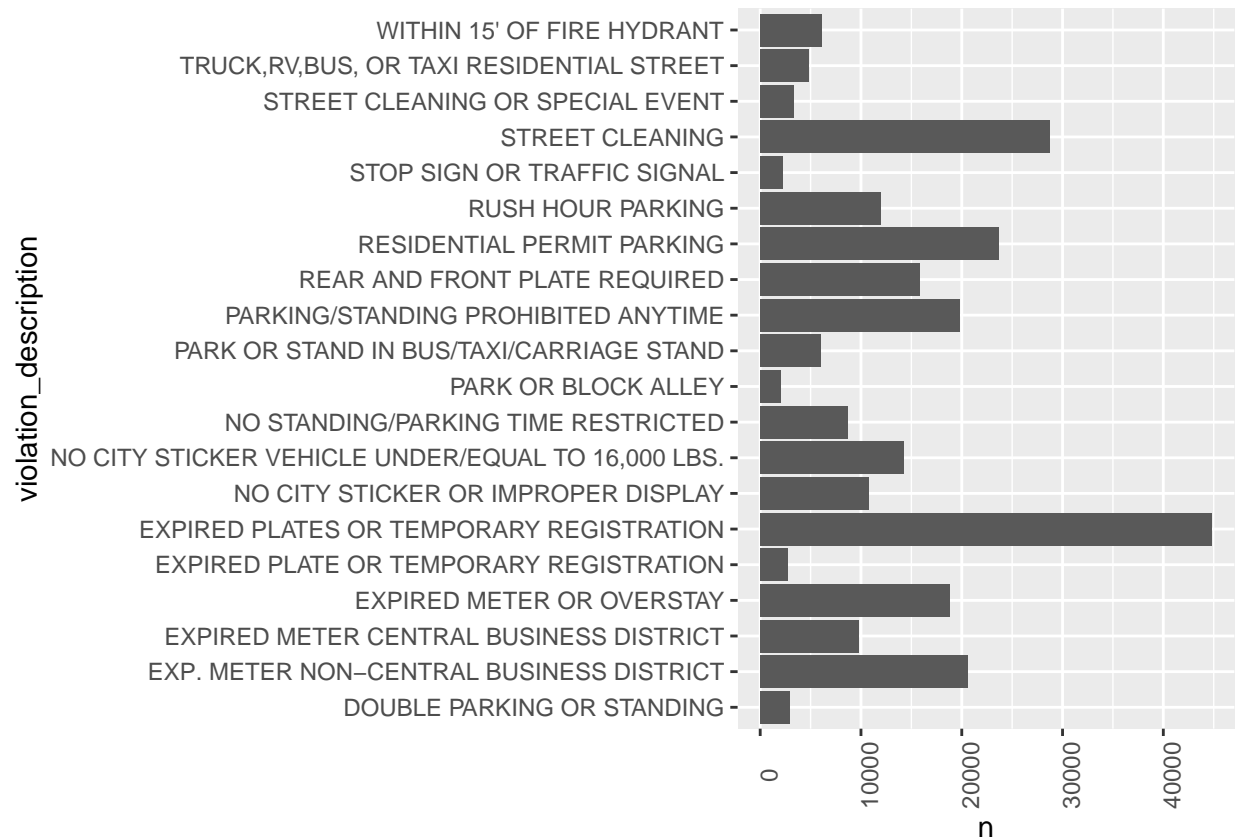
1. 22364 tickets were issued in 2017, which implies around 2236400 tickets were issued in total. According to ProPublica, the City of Chicago issues more than 3 million tickets each year. I think this is meaningful as being interpreted as that picking tickets ending with 01 is not so representative for full data set.

```
pt <- pt %>%
  mutate(year = str_extract(issue_date, "\\d{1,4}"))
pt %>%
  filter(year == "2017") %>%
  nrow()
```

```
## [1] 22364
```

- 2.

```
pt %>%
  group_by(violation_description) %>%
  count() %>%
  arrange(-n) %>%
  head(20) %>%
  ungroup() %>%
  ggplot() +
  geom_col(aes(x = violation_description, y = n)) +
  theme(axis.text.x = element_text(angle = 90)) +
  coord_flip()
```



Joins - unit

1. 29 rows have missing data for Unit.
2. There are 385 units in unit_key.

```
uk <- read.csv("unit_key.csv",encoding = "UTF-8")
uk <- uk %>%
  rename(unit = X.U.FEFF.Reporting.District)
uk %>%
  count(unit) %>%
  nrow()
```

```
## [1] 385
```

3. 287429 rows in tickets are matched, 29 rows in tickets are unmatched. 128 rows in unit_key are matched, 257 rows are not.

```
pt <- pt %>%
  mutate(unit=as.character(unit)) %>%
  left_join(uk, by = "unit")

pt1 <- pt %>%
  mutate(unit=as.character(unit)) %>%
```

```
full_join(uk, by = "unit")
```

```
287715-287458
```

```
## [1] 257
```

```
sum(is.na(uk$Department.Name))
```

```
## [1] 0
```

4.

DOF issued more tickets. 1160 N. Larrabee, 6464 N. Clark, OEMC, 3315 W. Ogden, 5555 W. Grand are the top 5 departments under CPO issued the most tickets.

```
pt %>% count(unit_description) %>% arrange(-n)
```

```
##   unit_description      n
## 1                DOF 143909
## 2                CPD 120712
## 3   Miscellaneous  16442
## 4        CPD-Other   3750
## 5        CPD-Airport  2617
## 6        Unidentified    28
```

```
pt %>%
  count(unit_description, Department.Description, sort = TRUE) %>%
  head(8)
```

```
##   unit_description Department.Description      n
## 1                DOF Department of Finance 106482
## 2                DOF                SERCO   37426
## 3                CPD      1160 N. Larrabee   9478
## 4   Miscellaneous                CPM      8064
## 5                CPD      6464 N. Clark   7946
## 6                CPD                OEMC   7374
## 7                CPD      3315 W. Ogden   5469
## 8                CPD      5555 W. Grand   5464
```

Joins - ZIP code

1.

```
zip <- read.csv("chi_zips.csv", encoding="UTF-8")
zip <- zip %>% rename(zipcode=X.U.FEFFF.ZIP) %>%
  mutate(zipcode=as.character(zipcode))

library(tidycensus)
```

```
## Warning: 'tidycensus' R 4.1.1
```

```

ACS_VARS_18<-load_variables(2018,"acs5",cache=TRUE)

acs_data<-get_acs(
  geography = "zcta",
  variables=c("B01003_001","B02001_003","B06011_001"),
  year=2018,
  state = "IL"
)

```

Getting data from the 2014-2018 5-year ACS

```

acs_data <- acs_data %>%
  select(-moe) %>%
  pivot_wider(names_from = variable,
              values_from = estimate) %>%
  rename(population=B01003_001,
         race=B02001_003,
         median_income=B06011_001) %>%
  filter(!is.na(population)) %>%
  filter(!is.na(race)) %>%
  filter(!is.na(median_income))

pt <- pt %>%
  mutate(zipcode = str_extract(zipcode, "\\d{1,5}")) %>%
  filter(!is.na(zipcode)) %>%
  left_join(acs_data, by = c("zipcode"="GEOID"))

pt %>%
  filter(ticket_queue!="paid") %>%
  filter(!is.na(population)) %>%
  count(population, zipcode) %>%
  mutate(unpaid = n/population) %>%
  arrange(-unpaid) %>%
  head(3)

```

```

##   population zipcode    n   unpaid
## 1      52793   60622 4755 0.09006876
## 2      32203   60636 2741 0.08511629
## 3       782   60604   65 0.08312020

```

4.

```

zip_unpaid <- pt %>%
  filter(ticket_queue!="paid") %>%
  filter(!is.na(population)) %>%
  count(population, zipcode) %>%
  mutate(unpaid = n/population)

map<-get_acs(
  geography = "zcta",
  variables=c("B01003_001","B02001_003","B06011_001"),
  year=2018,

```

```

state = "IL",
geometry=TRUE
) %>%
inner_join(zip_unpaid, by = c("GEOID"="zipcode")) %>%
inner_join(zip, by = c("GEOID"="zipcode")) %>%
filter(!is.na(unpaid)) %>%
ggplot(aes(fill=unpaid, color=unpaid))+
geom_sf()+
labs(title = "",
      caption="")

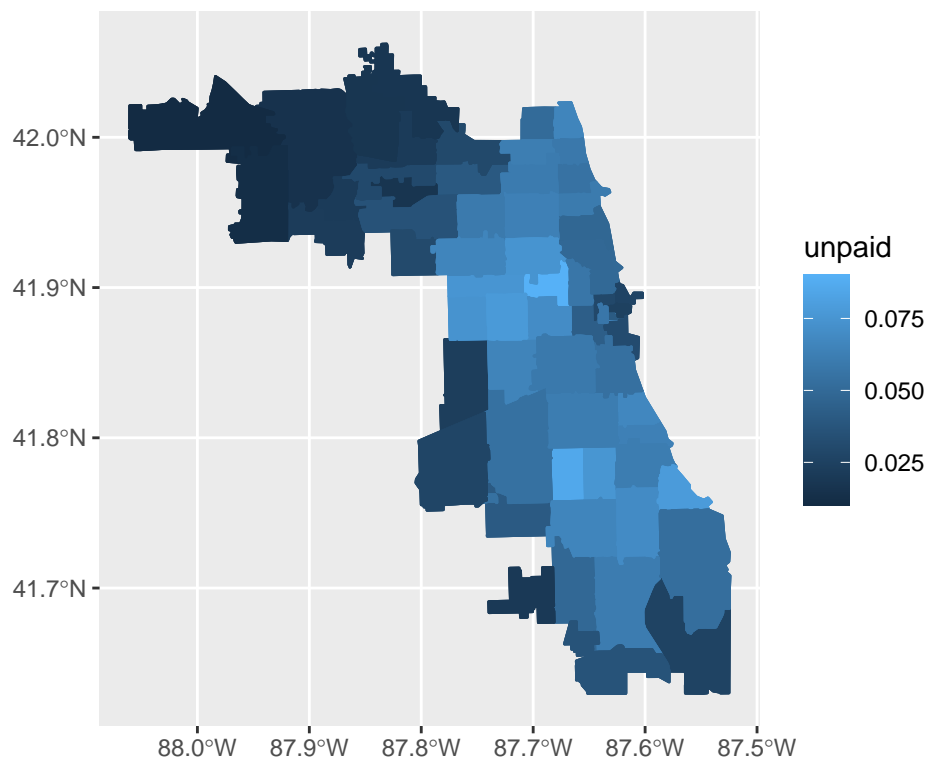
```

```
## Getting data from the 2014-2018 5-year ACS
```

```
## Downloading feature geometry from the Census website. To cache shapefiles for use in future session
```

```
## |
```

```
map
```



Part II

Understanding the structure of the data

1.

```
fine <- pt %>%
  filter(fine_level2_amount<2*fine_level1_amount) %>%
  count(violation_code) %>%
  filter(n>100)
fine
```

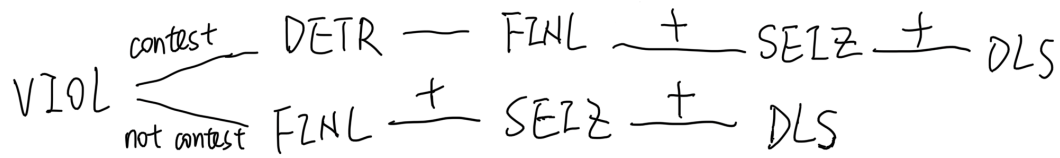
```
## violation_code    n
## 1      0964050J  857
## 2      0964100C  423
## 3      0964130 1347
## 4      0976220B  570
```

```
pt %>%
  filter(violation_code %in% fine$violation_code) %>%
  mutate(unpaid_increase = fine_level2_amount-fine_level1_amount) %>%
  count(violation_code,violation_description,unpaid_increase)
```

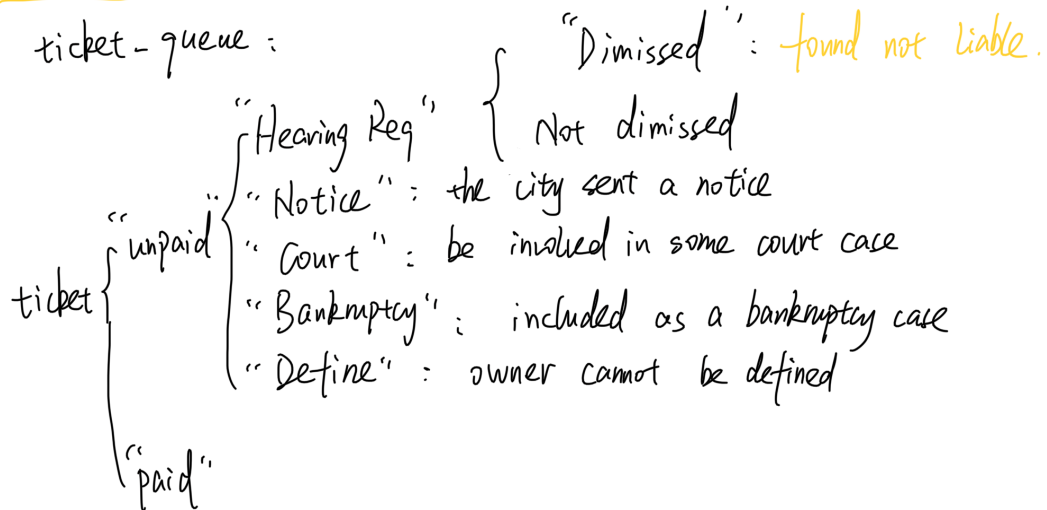
```
## violation_code violation_description unpaid_increase    n
## 1      0964050J      DISABLED PARKING ZONE           50  857
## 2      0964050J      DISABLED PARKING ZONE          200  276
## 3      0964050J      DISABLED PARKING ZONE          250  591
## 4      0964100C BLOCK ACCESS/ALLEY/DRIVEWAY/FIRELANE    75  138
## 5      0964100C BLOCK ACCESS/ALLEY/DRIVEWAY/FIRELANE   100  423
## 6      0964100C BLOCK ACCESS/ALLEY/DRIVEWAY/FIRELANE   150  723
## 7      0964130      PARK OR BLOCK ALLEY             100 1347
## 8      0964130      PARK OR BLOCK ALLEY             150  338
## 9      0976220B SMOKED/TINTED WINDOWS PARKED/STANDING    0  570
## 10     0976220B SMOKED/TINTED WINDOWS PARKED/STANDING   25  621
## 11     0976220B SMOKED/TINTED WINDOWS PARKED/STANDING  250  290
```

2.

```
knitr::include_graphics("graph.png")
```

ticket - queue :



3.

```

vd <- pt %>%
  count(violation_description, violation_code) %>%
  count(violation_description) %>%
  filter(n>1)

pt %>%
  filter(violation_description %in% vd$violation_description) %>%
  count(violation_description, violation_code)

```

##	violation_description	violation_code	n
## 1	3-7 AM SNOW ROUTE	0964060	658
## 2	3-7 AM SNOW ROUTE	0964060B	8
## 3	CURB LOADING ZONE	0964160A	1
## 4	CURB LOADING ZONE	0964160B	913
## 5	INDUSTRIAL PERMIT PARKING	0964091	95
## 6	INDUSTRIAL PERMIT PARKING	0964091B	2
## 7	NO CITY STICKER OR IMPROPER DISPLAY	0964125	9809
## 8	NO CITY STICKER OR IMPROPER DISPLAY	0976170	12
## 9	SPECIAL EVENTS RESTRICTION	0964041	213
## 10	SPECIAL EVENTS RESTRICTION	0964041B	164

4.

```

vc <- pt %>%
  count(violation_description, violation_code) %>%
  count(violation_code) %>%

```

```

filter(n>1)
pt %>%
  filter(violation_code %in% vc$violation_code) %>%
  count(violation_code, violation_description)

```

##	violation_code	violation_description	n
## 1	0964040B	STREET CLEANING	22816
## 2	0964040B	STREET CLEANING OR SPECIAL EVENT	2657
## 3	0964041B	Special Events	11
## 4	0964041B	SPECIAL EVENTS RESTRICTION	164
## 5	0964070	SNOW ROUTE: 2' OF SNOW OR MORE	115
## 6	0964070	SNOW ROUTE: 2' OF SNOW OR MORE	16
## 7	0964170D	TRUCK OR SEMI-TRAILER PROHIBITED	92
## 8	0964170D	TRUCK TRAILOR/SEMI/TRAILER PROHIBITED	47
## 9	0964200B	OUTSIDE METERED SPACE	54
## 10	0964200B	PARK OUTSIDE METERED SPACE	245
## 11	0976160A	MISSING/NONCOMPLIANT FRONT AND/OR REAR PLATE	645
## 12	0976160A	REAR AND FRONT PLATE REQUIRED	11099
## 13	0976160B	EXPIRED PLATE OR TEMPORARY REGISTRATION	2283
## 14	0976160B	REAR PLATE REQUIRED MOTORCYCLE/TRAILER	281
## 15	0980110B	HAZARDOUS DILAPIDATED VEHICLE	104
## 16	0980110B	HAZARDOUS DILAPITATED VEHICLE	200

5. For example, “EXPIRED PLATES OR TEMPORARY REGISTRATION” and “EXPIRED PLATE OR TEMPORARY REGISTRATION” have different violation_code.

```

vd50 <- pt %>%
  count(violation_description, sort = TRUE) %>%
  head(50)

pt %>%
  filter(violation_description %in% vd50$violation_description) %>%
  count(violation_description, violation_code)

```

##	violation_description	violation_code	n
## 1	20' OF CROSSWALK	0964100F	330
## 2	3-7 AM SNOW ROUTE	0964060	658
## 3	3-7 AM SNOW ROUTE	0964060B	8
## 4	ABANDONED VEH. FOR 7 DAYS OR INOPERABLE	0980110A	707
## 5	BLOCK ACCESS/ALLEY/DRIVEWAY/FIRELANE	0964100C	1284
## 6	CURB LOADING ZONE	0964160A	1
## 7	CURB LOADING ZONE	0964160B	913
## 8	DISABLED CURB CUT	0964100D	360
## 9	DISABLED PARKING ZONE	0964050J	1724
## 10	DOUBLE PARKING OR STANDING	0964110A	2518
## 11	EXP. METER NON-CENTRAL BUSINESS DISTRICT	0964190A	16735
## 12	EXPIRED METER CENTRAL BUSINESS DISTRICT	0964190B	7837
## 13	EXPIRED METER OR OVERSTAY	0964190	15061
## 14	EXPIRED PLATE OR TEMPORARY REGISTRATION	0976160B	2283
## 15	EXPIRED PLATES OR TEMPORARY REGISTRATION	0976160F	38016
## 16	HAZARDOUS DILAPITATED VEHICLE	0980110B	200
## 17	IMPROPER DISPLAY OF CITY STICKER	0964125D	352

## 18	MISSING/NONCOMPLIANT FRONT AND/OR REAR PLATE	0976160A	645
## 19	NO CITY STICKER OR IMPROPER DISPLAY	0964125	9809
## 20	NO CITY STICKER OR IMPROPER DISPLAY	0976170	12
## 21	NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.	0964125B	13002
## 22	NO PARK IN PRIVATE LOT	0980120B	298
## 23	NO STANDING/PARKING TIME RESTRICTED	0964080B	6747
## 24	NONCOMPLIANT PLATE(S)	0976160D	1625
## 25	OBSTRUCT ROADWAY	0964020B	1331
## 26	OBSTRUCTED OR IMPROPERLY TINTED WINDOWS	0976220A	228
## 27	PARK ALLEY	0964130A	828
## 28	PARK OR BLOCK ALLEY	0964130	1685
## 29	PARK OR STAND IN BUS/TAXI/CARRIAGE STAND	0964140B	5030
## 30	PARK OR STAND IN VIADUCT/UNDERPASS	0964110G	199
## 31	PARK OR STAND ON CROSSWALK	0964110C	1543
## 32	PARK OR STAND ON PARKWAY	0964110E	393
## 33	PARK OR STAND ON SIDEWALK	0964110D	825
## 34	PARK OUTSIDE METERED SPACE	0964200B	245
## 35	PARK VEHICLE SOLE PURPOSE OF DISPLAYING FOR SALE	0980080A	208
## 36	PARKING/STANDING PROHIBITED ANYTIME	0964150B	15856
## 37	REAR AND FRONT PLATE REQUIRED	0976160A	11099
## 38	REAR PLATE REQUIRED MOTORCYCLE/TRAILER	0976160B	281
## 39	RESIDENTIAL PERMIT PARKING	0964090E	18989
## 40	RUSH HOUR PARKING	0964080A	9187
## 41	SAFETY BELTS REQUIRED	0976180A	768
## 42	SMOKED/TINTED WINDOWS PARKED/STANDING	0976220B	1481
## 43	SPECIAL EVENTS RESTRICTION	0964041	213
## 44	SPECIAL EVENTS RESTRICTION	0964041B	164
## 45	STAND, PARK, OR OTHER USE OF BUS LANE	0912060	1063
## 46	STOP SIGN OR TRAFFIC SIGNAL	0964100G	1820
## 47	STREET CLEANING	0964040B	22816
## 48	STREET CLEANING OR SPECIAL EVENT	0964040B	2657
## 49	TRUCK,MOTOR HOME, BUS BUSINESS STREET	0964170B	354
## 50	TRUCK,RV,BUS, OR TAXI RESIDENTIAL STREET	0964170A	3600
## 51	TWO HEAD LAMPS REQUIRED VISIBLE 1000'	0976050B	403
## 52	WINDOWS MISSING OR CRACKED BEYOND 6	0976210B	453
## 53	WITHIN 15' OF FIRE HYDRANT	0964100A	5016
## 54	WRONG DIRECTION OR 12'' FROM CURB	0964020A	848

Revenue increase from “missing city sticker” tickets

1. The new code is “0964125B”, the old is “0964125” and “0976170”. The initial cost is 200 and 120.

```
pt %>%
  mutate(sticker=str_extract(pt$violation_description, ".STICKER.")) %>%
  filter(!is.na(sticker)) %>%
  select(violation_description, violation_code, fine_level1_amount,year) %>%
  count(violation_description, violation_code, fine_level1_amount,year)
```

##	violation_description	violation_code
## 1	NO CITY STICKER OR IMPROPER DISPLAY	0964125
## 2	NO CITY STICKER OR IMPROPER DISPLAY	0964125
## 3	NO CITY STICKER OR IMPROPER DISPLAY	0964125
## 4	NO CITY STICKER OR IMPROPER DISPLAY	0964125

```

## 5          NO CITY STICKER OR IMPROPER DISPLAY          0964125
## 6          NO CITY STICKER OR IMPROPER DISPLAY          0964125
## 7          NO CITY STICKER OR IMPROPER DISPLAY          0976170
## 8          NO CITY STICKER OR IMPROPER DISPLAY          0976170
## 9          NO CITY STICKER OR IMPROPER DISPLAY          0976170
## 10         NO CITY STICKER OR IMPROPER DISPLAY          0976170
## 11         NO CITY STICKER OR IMPROPER DISPLAY          0976170
## 12         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 13         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 14         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 15         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 16         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 17         NO CITY STICKER VEHICLE OVER 16,000 LBS.    0964125C
## 18 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 19 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 20 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 21 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 22 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 23 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
## 24 NO CITY STICKER VEHICLE UNDER/EQUAL TO 16,000 LBS.  0964125B
##   fine_level1_amount year      n
## 1          120 2007 2026
## 2          120 2008 1987
## 3          120 2009 1963
## 4          120 2010 1842
## 5          120 2011 1801
## 6          120 2012  190
## 7          120 2007    6
## 8          120 2008    1
## 9          120 2009    2
## 10         120 2010    1
## 11         120 2011    2
## 12         500 2012   19
## 13         500 2013   32
## 14         500 2014   17
## 15         500 2015   15
## 16         500 2016   12
## 17         500 2017    7
## 18         200 2012 1847
## 19         200 2013 2318
## 20         200 2014 1856
## 21         200 2015 2250
## 22         200 2016 2057
## 23         200 2017 2047
## 24         200 2018  627

```

2. There is no obvious trend for number of missing sticker tickets over years. They keep around 2000 per year.

```

pt %>%
  mutate(sticker=str_extract(pt$violation_description, ".STICKER.")) %>%
  filter(!is.na(sticker)) %>%
  select(violation_description, violation_code, fine_level1_amount,year) %>%
  count(year)

```

```
##   year    n
## 1  2007 2032
## 2  2008 1988
## 3  2009 1965
## 4  2010 1843
## 5  2011 1803
## 6  2012 2056
## 7  2013 2350
## 8  2014 1873
## 9  2015 2265
## 10 2016 2069
## 11 2017 2054
## 12 2018  627
```

3. Maybe 2012-02-25.

```
library(lubridate)
```

```
## Warning:   'lubridate' R 4.1.1
```

```
##
##   'lubridate'
```

```
## The following objects are masked from 'package:base':
```

```
##
##   date, intersect, setdiff, union
```

```
pt %>%
  filter(violation_code == "0964125B") %>%
  select(issue_date) %>%
  arrange() %>%
  head(1)
```

```
##           issue_date
## 1 2012-02-25 02:00:00
```

4. It should have been 10 million.

```
revenue_2011 <- pt %>%
  filter(year=="2011"&ticket_queue=="Paid") %>%
  filter(violation_code=="0964125"|violation_code=="0976170") %>%
  mutate(revenue=sum(total_payments)) %>%
  select(revenue) %>%
  head(1)

revenue_2012 <- pt %>%
  filter(year=="2012"&ticket_queue=="Paid") %>%
  filter(violation_code=="0964125C"|violation_code=="0964125B") %>%
  mutate(revenue=sum(total_payments)) %>%
  select(revenue) %>%
  head(1)

(revenue_2012-revenue_2011)*100
```

```
## revenue
## 1 10581470
```

5. After the increase, the repayment rate is 0.49. Before the increase, the repayment rate is 0.54. If the number of issued tickets stays constant, the revenue would increase by just 14.94%.

#After the fine increase

```
pt %>%
  filter(year=="2012") %>%
  filter(violation_code=="0964125C"|violation_code=="0964125B") %>%
  count(ticket_queue) %>%
  mutate(all_ticket_issued = sum(n), repayment_rate = n/all_ticket_issued)
```

```
## ticket_queue n all_ticket_issued repayment_rate
## 1 Bankruptcy 70 1866 0.0375133976
## 2 Court 1 1866 0.0005359057
## 3 Define 52 1866 0.0278670954
## 4 Dismissed 278 1866 0.1489817792
## 5 Notice 550 1866 0.2947481243
## 6 Paid 915 1866 0.4903536977
```

#Before the fine increase

```
pt %>%
  filter(year=="2011") %>%
  filter(violation_code=="0976170"|violation_code=="0964125") %>%
  count(ticket_queue) %>%
  mutate(all_ticket_issued = sum(n), repayment_rate = n/all_ticket_issued)
```

```
## ticket_queue n all_ticket_issued repayment_rate
## 1 Bankruptcy 55 1803 0.03050471
## 2 Define 58 1803 0.03216861
## 3 Dismissed 237 1803 0.13144759
## 4 Notice 465 1803 0.25790349
## 5 Paid 988 1803 0.54797560
```

#revenue change after increasing ticket cost from 120 to 200, if not started issuing more of these tickets
 $(200 \times 0.49 - 120 \times 0.54) / (120 \times 0.54)$

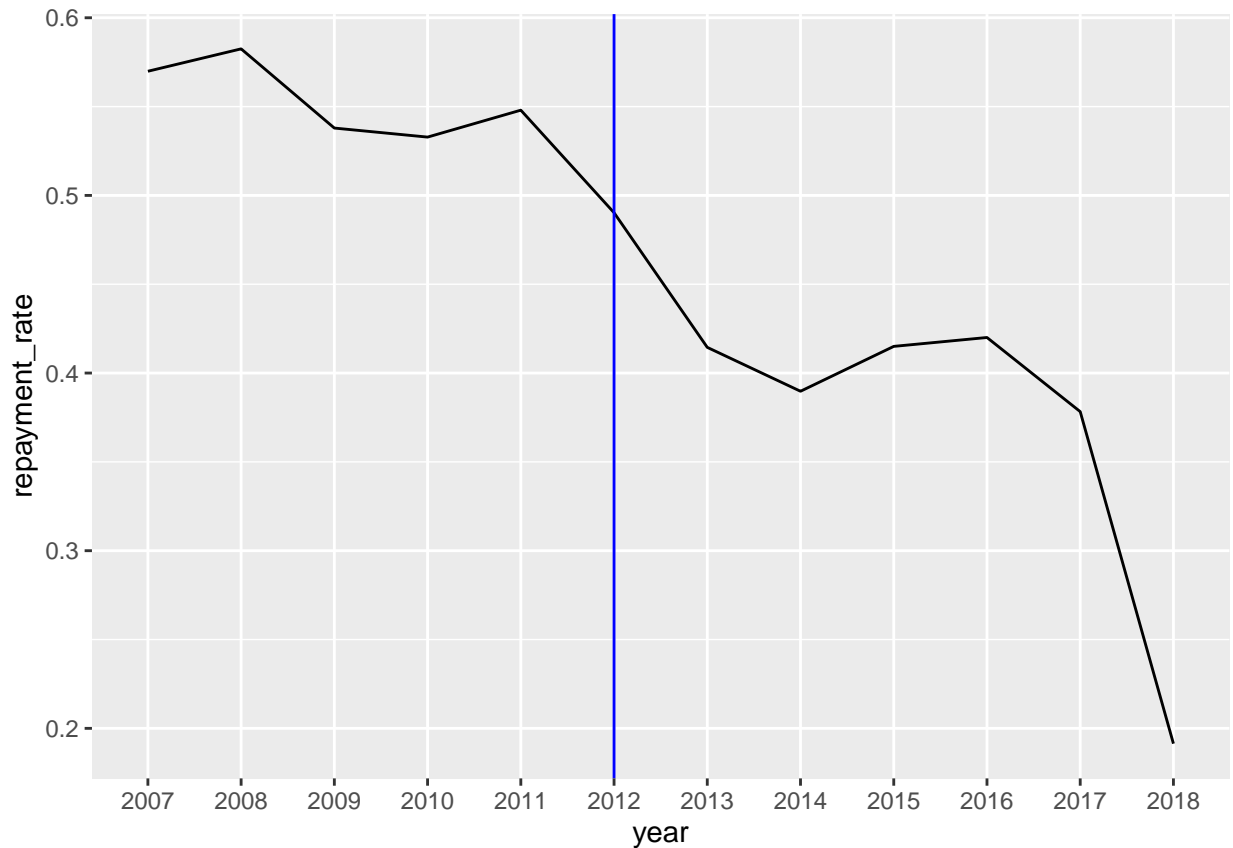
```
## [1] 0.1494
```

- 6.

```
map <- pt %>%
  mutate(sticker=str_extract(pt$violation_description, ".STICKER.")) %>%
  filter(!is.na(sticker)) %>%
  group_by(year) %>%
  count(ticket_queue) %>%
  mutate(all_ticket_issued = sum(n), repayment_rate = n/all_ticket_issued) %>%
  filter(ticket_queue == "Paid") %>%
  select(ticket_queue, repayment_rate) %>%
  ungroup() %>%
  ungroup()
```

```
## Adding missing grouping variables: `year`
```

```
map %>% ggplot(aes(x=year,y=repayment_rate))+
  geom_line(group = 1)+
  geom_vline(aes(xintercept = "2012"), color="blue")
```



7. For 2010 and 2011, “EXPIRED METER OR OVERSTAY”, “EXPIRED PLATES OR TEMPORARY REGISTRATION” and “STREET CLEANING” have the most paid tickets, so I suggest we should increase the amount of fine on these tickets.

```
pt %>%
  filter(year == "2011") %>%
  filter(ticket_queue == "Paid") %>%
  count(violation_description) %>%
  arrange(-n) %>%
  head(5)
```

```
##           violation_description      n
## 1      EXPIRED METER OR OVERSTAY 3399
## 2 EXPIRED PLATES OR TEMPORARY REGISTRATION 2492
## 3           STREET CLEANING 1838
## 4      RESIDENTIAL PERMIT PARKING 1207
## 5    NO CITY STICKER OR IMPROPER DISPLAY  988
```

```
pt %>%
  filter(year == "2010") %>%
  filter(ticket_queue == "Paid") %>%
  count(violation_description) %>%
  arrange(-n) %>%
  head(5)
```

```
##              violation_description      n
## 1          EXPIRED METER OR OVERSTAY 3250
## 2 EXPIRED PLATES OR TEMPORARY REGISTRATION 2318
## 3              STREET CLEANING 1753
## 4          RESIDENTIAL PERMIT PARKING 1287
## 5      NO CITY STICKER OR IMPROPER DISPLAY  982
```

8.

- a. Using tickets data in 2011, among the top 10 mostly paid tickets, the total amount of tickets paid by high income neighborhoods is more than those by low income neighborhood for the following kinds of tickets:

EXPIRED METER OR OVERSTAY; NO STANDING/PARKING TIME RESTRICTED; REAR AND FRONT PLATE REQUIRED; RESIDENTIAL PERMIT PARKING; RUSH HOUR PARKING; STREET CLEANING

- b. About 62 million increase in revenue using number of tickets paid in 2011.

```
pt %>%
  filter(!is.na(median_income)) %>%
  distinct(zipcode, .keep_all = TRUE) %>%
  summarise(median=median(median_income))
```

```
##      median
## 1  30773
```

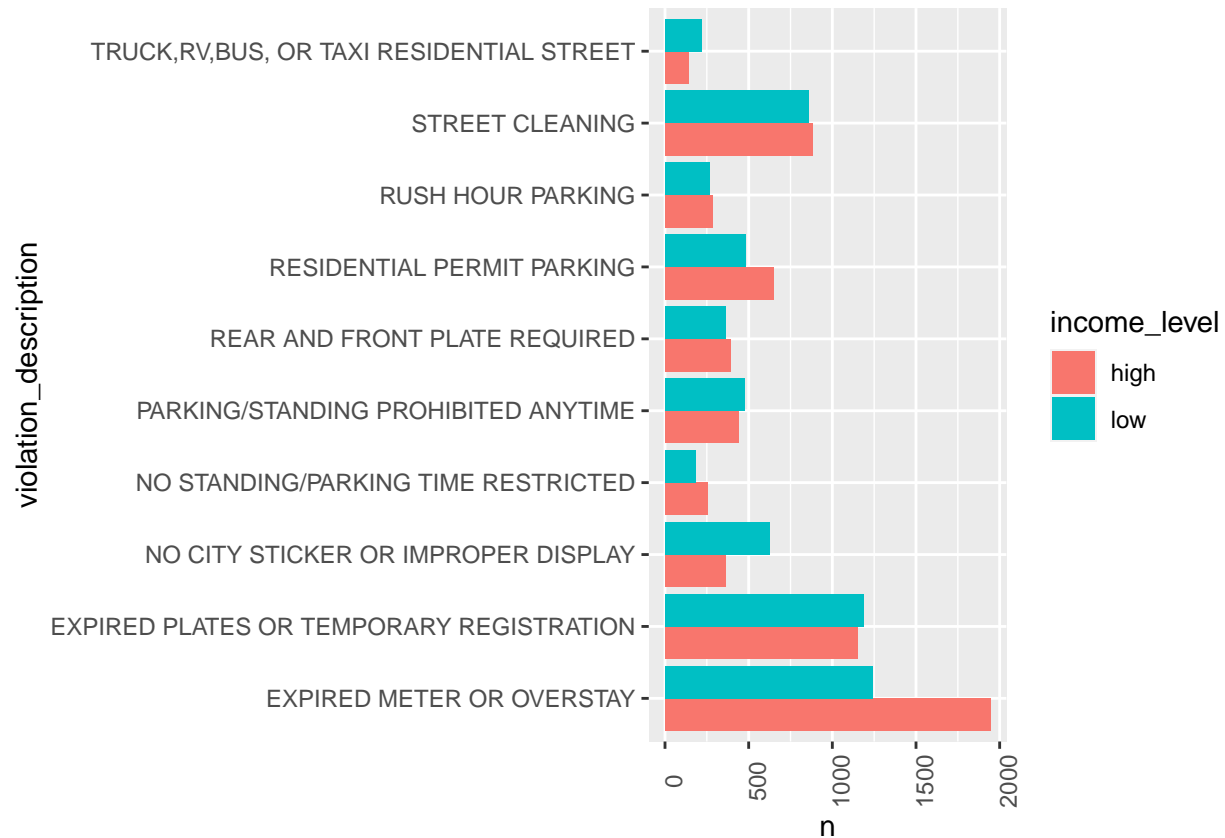
```
pt <- pt %>%
  filter(!is.na(median_income)) %>%
  mutate(income_level = if_else(median_income > 30773, "high", "low"))
```

```
high_paid_tickets <- pt %>%
  filter(year == "2011") %>%
  filter(ticket_queue == "Paid") %>%
  count(violation_description) %>%
  arrange(-n) %>%
  select(violation_description) %>%
  head(10)
```

```
pt %>%
  filter(year=="2011") %>%
  filter(violation_description %in% high_paid_tickets$violation_description) %>%
  filter(ticket_queue == "Paid") %>%
```



```
count(violation_description, income_level) %>%
ggplot(aes(x=violation_description, y=n, fill=income_level))+
geom_col(position = "dodge")+
coord_flip()+
theme(axis.text.x = element_text(angle = 90))
```



```
pt %>%
  filter(year=="2011") %>%
  filter(violation_description %in% high_paid_tickets$violation_description) %>%
  filter(ticket_queue == "Paid") %>%
  count(violation_description, income_level) %>%
  pivot_wider(names_from = income_level, values_from = n) %>%
  filter(high>low) %>%
  summarise(increase_revenue = sum(high+low)*80*100)
```

```
## # A tibble: 1 x 1
##   increase_revenue
##             <dbl>
## 1           62344000
```