Take_home_assignment

Vicky (Ke) Xu

15/02/2021

City of Toronto

PARKS, FORESTRY & RECREATION DIVISION

Data Analyst & Integrator (Data Science)

Job ID# 13025(X)

Technical Assignment Result

Load all requried packages

The very first step is to load all parking ticket data. I took it from year 2016 to 2018.

```r
all_file_names <- list.files('parking-ticket/parking-tickets-2016/',full.name=TRUE)
listDFs <- lapply(all_file_names, read.csv, quote='', stringsAsFactors=FALSE)
combDFs_2016 <- do.call("rbind", listDFs)

all_file_names <- list.files('parking-ticket/parking-tickets-2017/',full.name=TRUE)
listDFs <- lapply(all_file_names, read.csv, quote='', stringsAsFactors=FALSE)
combDFs_2017 <- do.call("rbind", listDFs)

all_file_names <- list.files('parking-ticket/parking-tickets-2018/',full.name=TRUE)
listDFs <- lapply(all_file_names, read.csv, quote='', stringsAsFactors=FALSE)
combDFs_2018 <- do.call("rbind", listDFs)


parking_ticket <- rbind(combDFs_2016,combDFs_2017)
parking_ticket <- rbind(parking_ticket,combDFs_2018)
```

Preliminary examinations on missing values

```r
# Check the number of missing values within each column

parking_ticket_missing_summary <- apply(parking_ticket,2,function(x) sum(!complete.cases(x)))
print(parking_ticket_missing_summary)
```

```
##      tag_number_masked     date_of_infraction        infraction_code
##                      0                     25                     41
## infraction_description        set_fine_amount     time_of_infraction
##                      0                     39                   4276
##              location1              location2              location3
##                      0                      0                      1
##              location4               province
##                      0                      0
```

```
# Remove missing values and $0 set_fine_amount (there are 93 of them)
parking_ticket <- parking_ticket[complete.cases(parking_ticket$date_of_infraction),]
parking_ticket <- parking_ticket[complete.cases(parking_ticket$infraction_code),]
parking_ticket <- parking_ticket[complete.cases(parking_ticket$set_fine_amount),]
parking_ticket <- parking_ticket[parking_ticket$set_fine_amount != 0,]

# Decided to leave all missing time_of_infraction as is because
# it is not within the objectives of this assignment.
nrow(parking_ticket[!complete.cases(parking_ticket$time_of_infraction),])
```

```
## [1] 4217
```

3.1.1 Top 20 ticket infractions (frequency)

```
top_20_freq <- parking_ticket %>%
  group_by(infraction_code,infraction_description) %>%
  summarise(freq=n()) %>%
  arrange(desc(freq)) %>%
  head(20) %>%
  as.data.frame()
top_20_freq
```

```
##    infraction_code         infraction_description    freq
## 1                3       PARK ON PRIVATE PROPERTY 1129942
## 2                5 PARK-SIGNED HWY-PROHIBIT DY/TM 1118241
## 3               29 PARK PROHIBITED TIME NO PERMIT  971362
## 4              207 PARK MACHINE-REQD FEE NOT PAID  615430
## 5                9 STOP-SIGNED HWY-PROHIBIT TM/DY  319581
## 6                2     PARK - LONGER THAN 3 HOURS  236371
## 7                8   STAND VEH.-PROHIBIT TIME/DAY  233899
## 8              406  PARK-VEH. W/O VALID ONT PLATE  231938
## 9              403  STOP-SIGNED HIGHWAY-RUSH HOUR  214804
## 10             210   PARK FAIL TO DISPLAY RECEIPT  213395
## 11               6 PARK-SIGNED HWY-EXC PERMT TIME  178102
## 12               2    PARK - LONGER THAN 3 HOURS   116497
## 13              15 PARK-WITHIN 3M OF FIRE HYDRANT   96584
## 14              28 PARK-N.YORK 2AM-6AM DEC1-MAR31   78585
## 15             192      STAND SIGNED TRANSIT STOP   73677
## 16             415 PARK COMMERC LOAD ZONE NOT LDG   49283
## 17             312   PARKING MACH-NOT USED/NO FEE   48892
## 18              30 STOP-(ON/OVER) (SIDEWK/FTPATH)   48291
## 19             134    PARK-SIGNED HWY-PUBLIC LANE   43298
## 20             347           PARK IN A FIRE ROUTE   40553
```

3.1.2 Top 20 ticket infractions (revenue)

```r
top_20_revenue <- parking_ticket %>%
  group_by(infraction_code,infraction_description) %>%
  summarise(revenue=sum(set_fine_amount)) %>%
  arrange(desc(revenue)) %>%
  head(20) %>%
  as.data.frame()
top_20_revenue
```

```
##    infraction_code         infraction_description  revenue
## 1                5 PARK-SIGNED HWY-PROHIBIT DY/TM 55022530
## 2                3       PARK ON PRIVATE PROPERTY 33898260
## 3              403  STOP-SIGNED HIGHWAY-RUSH HOUR 32220600
## 4               29 PARK PROHIBITED TIME NO PERMIT 29140860
## 5                9 STOP-SIGNED HWY-PROHIBIT TM/DY 19174860
## 6              207 PARK MACHINE-REQD FEE NOT PAID 18462900
## 7                8   STAND VEH.-PROHIBIT TIME/DAY 14033940
## 8              192      STAND SIGNED TRANSIT STOP 10440900
## 9              347           PARK IN A FIRE ROUTE 10138250
## 10              15 PARK-WITHIN 3M OF FIRE HYDRANT  9658400
## 11             406  PARK-VEH. W/O VALID ONT PLATE  9277520
## 12             367 STND ONSTRT ACCESSIBLE NO PRMT  8601300
## 13               6 PARK-SIGNED HWY-EXC PERMT TIME  7124080
## 14              30 STOP-(ON/OVER) (SIDEWK/FTPATH)  6761790
## 15             210   PARK FAIL TO DISPLAY RECEIPT  6401850
## 16             355   PARK IN ACCESSIBLE NO PERMIT  4631850
## 17             363 PARK ONSTRT ACCESSIBLE NO PRMT  3968100
## 18               2     PARK - LONGER THAN 3 HOURS  3545565
## 19              28 PARK-N.YORK 2AM-6AM DEC1-MAR31  3143400
## 20             384 STOP VEH OTR THN BCYCL-BYCL LN  2617950
```

3.1.4a How far (as the crow flies) is the closest parking lot (Green P) to each of the top 20 infraction locations?

```r
top_20_locations <- parking_ticket %>%
  group_by(location2) %>%
  summarise(freq=n()) %>%
  arrange(desc(freq)) %>%
  head(20) %>%
  as.data.frame()

# Paste all information to make it ready for geocoding
top_20_locations$location_new <- paste0(top_20_locations$location2,',TORONTO,ON,CANADA')

# Geocode all 20 addresses using the geocode function from tidygeocoder with method='osm'
top_20_locations_geocoded <- top_20_locations %>%
  tidygeocoder::geocode(location_new,method='osm')

# Load green-p-parking data
green_p <- fromJSON('green-p-parking/green-p-parking-2019.json') %>% as.data.frame()
```

```
# Convert carparks.lat and carparks.lng to numeric
green_p$carparks.lat <- as.numeric(green_p$carparks.lat)
green_p$carparks.lng <- as.numeric(green_p$carparks.lng)

# Make sure that there is no missing locations
nrow(green_p[!complete.cases(green_p$carparks.lat),])
```

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

```
nrow(green_p[!complete.cases(green_p$carparks.lng),])
```

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

```
# Create spatial objects for top_20_locations_geocoded and green_p
sp_top_20_locations_geocoded <- top_20_locations_geocoded
coordinates(sp_top_20_locations_geocoded) <- ~long+lat

sp_green_p <- green_p
coordinates(sp_green_p) <- ~carparks.lng+carparks.lat

# Create a new data frame with two addtional columns to store results
top_20_locations_w_min_parking_dist <- top_20_locations_geocoded
top_20_locations_w_min_parking_dist['min_dist_ind'] <- 0
top_20_locations_w_min_parking_dist['min_dist'] <- 0

# Loop through all 20 locaitons and find the closest parking location
for (i in 1:20)
{
  dist_mat <- distm(coordinates(sp_top_20_locations_geocoded[i,]),coordinates(sp_green_p))
  top_20_locations_w_min_parking_dist$min_dist_ind[i] <- which.min(dist_mat)
  top_20_locations_w_min_parking_dist$min_dist[i]  <- min(dist_mat)
}

green_p_new <- green_p %>%
  mutate(index=1:nrow(green_p))

top_20_locations_w_min_parking_dist<- left_join(top_20_locations_w_min_parking_dist,
                                                green_p_new,by=c('min_dist_ind' = 'index')) %>%
                                                select(-min_dist_ind)
                                                as.data.frame()
```

```
## Error in as.data.frame(): argument "x" is missing, with no default
```

3.1.4b How far (as the crow flies) is the closest TTC stop to the top 20 infraction locations?

```
# Load TTC data
ttc_stops <- read.table('TTC Routes and Schedules/opendata_ttc_schedules/stops.txt',
                        header=TRUE,
                        sep=',',
                        stringsAsFactors = FALSE,
                        quote = "")
head(ttc_stops)
```

```
##   stop_id stop_code                      stop_name stop_desc stop_lat  stop_lon
## 1     262       662      DANFORTH RD AT KENNEDY RD        NA 43.71438 -79.26094
## 2     263       929     DAVENPORT RD AT BEDFORD RD        NA 43.67445 -79.39966
## 3     264       940      DAVENPORT RD AT DUPONT ST        NA 43.67551 -79.40194
## 4     265      1871 DAVISVILLE AVE AT CLEVELAND ST        NA 43.70209 -79.37811
## 5     266     11700         DISCO RD AT ATTWELL DR        NA 43.70136 -79.59484
## 6     267      3478         DISCO RD AT ATTWELL DR        NA 43.70104 -79.59581
##   zone_id stop_url location_type parent_station stop_timezone
## 1      NA       NA            NA             NA            NA
## 2      NA       NA            NA             NA            NA
## 3      NA       NA            NA             NA            NA
## 4      NA       NA            NA             NA            NA
## 5      NA       NA            NA             NA            NA
## 6      NA       NA            NA             NA            NA
##   wheelchair_boarding
## 1                   2
## 2                   1
## 3                   2
## 4                   1
## 5                   1
## 6                   1
```

```
dim(ttc_stops)
```

```
## [1] 9503   12
```

```
ttc_stops$stop_lat <- as.numeric(ttc_stops$stop_lat)
ttc_stops$stop_lon <- as.numeric(ttc_stops$stop_lon)

sp_ttc_stops <- ttc_stops
coordinates(sp_ttc_stops) <- ~stop_lon+stop_lat

# Create two columns to store results
top_20_locations_w_min_ttc_stops <- top_20_locations_geocoded
top_20_locations_w_min_ttc_stops['min_dist_ind'] <- 0
top_20_locations_w_min_ttc_stops['min_dist'] <- 0

# Loop through all 20 locations and find the closest parking location
for (i in 1:20)
{
  dist_mat <- distm(coordinates(sp_top_20_locations_geocoded[i,]),coordinates(sp_ttc_stops))
  top_20_locations_w_min_ttc_stops$min_dist_ind[i] <- which.min(dist_mat)
  top_20_locations_w_min_ttc_stops$min_dist[i]  <- min(dist_mat)
}

ttc_stops_new <- ttc_stops %>%
  mutate(index=1:nrow(ttc_stops))

top_20_locations_w_min_ttc_stops<- left_join(top_20_locations_w_min_ttc_stops,
                                             ttc_stops_new,by=c('min_dist_ind' = 'index')) %>%
                                             select(-min_dist_ind) %>%
                                             as.data.frame()
top_20_locations_w_min_ttc_stops
```

```
##               location2  freq                           location_new      lat
## 1          20 EDWARD ST 16612         20 EDWARD ST,TORONTO,ON,CANADA 43.65694
## 2      2075 BAYVIEW AVE 16422     2075 BAYVIEW AVE,TORONTO,ON,CANADA 43.72286
## 3     1265 MILITARY TRL  9731    1265 MILITARY TRL,TORONTO,ON,CANADA 43.78336
## 4         LA PLANTE AVE  8680        LA PLANTE AVE,TORONTO,ON,CANADA 43.65919
## 5     199 RICHMOND ST W  8664    199 RICHMOND ST W,TORONTO,ON,CANADA 43.64952
## 6     103 THE QUEENSWAY  8503    103 THE QUEENSWAY,TORONTO,ON,CANADA 43.63629
## 7              JAMES ST  8066             JAMES ST,TORONTO,ON,CANADA 43.65334
## 8   15 MARINE PARADE DR  7626  15 MARINE PARADE DR,TORONTO,ON,CANADA 43.62911
## 9        1 BRIMLEY RD S  7193       1 BRIMLEY RD S,TORONTO,ON,CANADA 43.70596
## 10     3401 DUFFERIN ST  7176     3401 DUFFERIN ST,TORONTO,ON,CANADA 43.72562
## 11     1000 FINCH AVE W  6968     1000 FINCH AVE W,TORONTO,ON,CANADA 43.76885
## 12      2075 BAYVIEW AV  6782      2075 BAYVIEW AV,TORONTO,ON,CANADA 43.72286
## 13            EDWARD ST  6563            EDWARD ST,TORONTO,ON,CANADA 43.65643
## 14     150 GERRARD ST W  6503     150 GERRARD ST W,TORONTO,ON,CANADA 43.65810
## 15   150 DAN LECKIE WAY  6325   150 DAN LECKIE WAY,TORONTO,ON,CANADA 43.63979
## 16       273 BLOOR ST W  6046       273 BLOOR ST W,TORONTO,ON,CANADA 43.66793
## 17       250 FRONT ST E  5792       250 FRONT ST E,TORONTO,ON,CANADA 43.65179
## 18  19 GRAND TRUNK CRES  5678  19 GRAND TRUNK CRES,TORONTO,ON,CANADA 43.64168
## 19      WELLINGTON ST W  5652      WELLINGTON ST W,TORONTO,ON,CANADA 43.64706
## 20 40 ORCHARD VIEW BLVD  5540 40 ORCHARD VIEW BLVD,TORONTO,ON,CANADA 43.70838
##         long   min_dist stop_id stop_code
## 1  -79.38207   74.59047   14454     13809
## 2  -79.37566   33.95817     851     11248
## 3  -79.18708  183.14252   14921     15114
## 4  -79.38602  104.07835    3450       248
## 5  -79.38841  106.54775   15131     15340
## 6  -79.47037   88.29467    5383     14279
## 7  -79.38120  116.09544    8008       233
## 8  -79.47541  125.74072   14053     14544
## 9  -79.23549 1071.07623    5336      3307
## 10 -79.45231  287.15598    4099      1677
## 11 -79.46910  147.43462     623      3587
## 12 -79.37566   33.95817     851     11248
## 13 -79.38399   68.31765    7505       242
## 14 -79.38794   89.82043    7730      1111
## 15 -79.39855   49.38852   15323     15497
## 16 -79.39625  105.34994    8736       495
## 17 -79.36362   44.50113     640      9536
## 18 -79.38304  251.78222   15123     15332
## 19 -79.38094  112.84821    5077       270
## 20 -79.40007  102.58733   24067     15709
##                                              stop_name stop_desc stop_lat
## 1                 DUNDAS STATION - NORTHBOUND PLATFORM        NA 43.65715
## 2     RAAB BLVD AT HOSPITAL RD WEST SIDE (K & L WINGS)        NA 43.72264
## 3                    UNIVERSITY OF TORONTO SCARBOROUGH        NA 43.78453
## 4                 BAY ST AT GERRARD ST WEST NORTH SIDE        NA 43.65878
## 5                       QUEEN ST WEST AT ST PATRICK ST        NA 43.65048
## 6                      THE QUEENSWAY AT WINDERMERE AVE        NA 43.63704
## 7                                  BAY ST AT ALBERT ST        NA 43.65289
## 8      MARINE PARADE DR (EAST) AT LAKE SHORE BLVD WEST        NA 43.62998
## 9                        BARKDENE HILL AT LARWOOD BLVD        NA 43.71544
## 10       YORKDALE RD AT GO TERMINAL (YORKDALE STATION)        NA 43.72589
## 11                       FINCH AVE WEST AT DUFFERIN ST        NA 43.76862
## 12    RAAB BLVD AT HOSPITAL RD WEST SIDE (K & L WINGS)        NA 43.72264
## 13 BAY ST AT DUNDAS ST WEST (TORONTO COACH TERMINAL)        NA 43.65583
## 14                   GERRARD ST EAST AT UNIVERSITY AVE        NA 43.65773
## 15        FORT YORK BLVD AT DAN LECKIE WAY WEST SIDE        NA 43.63935
## 16                       BLOOR ST WEST AT BEDFORD RD        NA 43.66790
## 17                      FRONT ST EAST AT BERKELEY ST        NA 43.65140
## 18           QUEENS QUAY WEST AT HARBOURFRONT CENTRE        NA 43.63961
## 19                      BAY ST AT WELLINGTON ST WEST        NA 43.64722
## 20                     YONGE ST AT ORCHARD VIEW BLVD        NA 43.70830
##     stop_lon zone_id stop_url location_type parent_station stop_timezone
## 1  -79.38119      NA       NA            NA             NA            NA
## 2  -79.37596      NA       NA            NA             NA            NA
## 3  -79.18548      NA       NA            NA             NA            NA
## 4  -79.38487      NA       NA            NA             NA            NA
## 5  -79.38848      NA       NA            NA             NA            NA
## 6  -79.47001      NA       NA            NA             NA            NA
## 7  -79.38249      NA       NA            NA             NA            NA
## 8  -79.47640      NA       NA            NA             NA            NA
## 9  -79.23791      NA       NA            NA             NA            NA
## 10 -79.44876      NA       NA            NA             NA            NA
## 11 -79.46730      NA       NA            NA             NA            NA
## 12 -79.37596      NA       NA            NA             NA            NA
## 13 -79.38388      NA       NA            NA             NA            NA
## 14 -79.38893      NA       NA            NA             NA            NA
## 15 -79.39847      NA       NA            NA             NA            NA
## 16 -79.39756      NA       NA            NA             NA            NA
## 17 -79.36374      NA       NA            NA             NA            NA
## 18 -79.38177      NA       NA            NA             NA            NA
## 19 -79.37955      NA       NA            NA             NA            NA
## 20 -79.39880      NA       NA            NA             NA            NA
##    wheelchair_boarding
## 1                    1
## 2                    1
## 3                    1
## 4                    1
## 5                    1
## 6                    1
## 7                    1
## 8                    1
## 9                    1
## 10                   1
## 11                   1
## 12                   1
## 13                   1
## 14                   1
## 15                   1
## 16                   1
## 17                   1
## 18                   1
## 19                   1
## 20                   1
```

3.1.5a Impact of day of week in all infractions

```
# First create columns for weekday, month, season and year
```

```
parking_ticket$month <- as.factor(month(ymd(parking_ticket$date_of_infraction)))
parking_ticket$year <- as.factor(year(ymd(parking_ticket$date_of_infraction)))
parking_ticket$weekday <- weekdays(as.Date(parking_ticket$date_of_infraction,
                 tryFormats = c('%Y%m%d')))

getSeason <- function(input.date){
  numeric.date <- 100*month(input.date)+day(input.date)
  ## input Seasons upper limits in the form MMDD in the "break =" option:
  cuts <- base::cut(numeric.date, breaks = c(0,319,0620,0921,1220,1231))
  # rename the resulting groups (could've been done within cut(...levels=) if "Winter" wasn't double
  levels(cuts) <- c("Winter","Spring","Summer","Fall","Winter")
  return(cuts)
}
parking_ticket$season <- getSeason(ymd(parking_ticket$date_of_infraction))

parking_ticket %>%
  group_by(weekday) %>%
  summarise(number_of_infractions = n())
```

```
## # A tibble: 7 x 2
##   weekday   number_of_infractions
##   <chr>                     <int>
## 1 Friday                  1005609
## 2 Monday                   889083
## 3 Saturday                 832535
## 4 Sunday                   651118
## 5 Thursday                1015880
## 6 Tuesday                 1018032
## 7 Wednesday               1030527
```

3.1.5b Impact of month of week in all infractions

```
parking_ticket %>%
  group_by(month) %>%
  summarise(number_of_infractions = n())
```

```
## # A tibble: 12 x 2
##    month number_of_infractions
##    <fct>                 <int>
##  1 1                    533714
##  2 2                    490695
##  3 3                    564517
##  4 4                    552767
##  5 5                    575895
##  6 6                    554569
##  7 7                    533534
##  8 8                    548016
##  9 9                    533590
## 10 10                   561249
## 11 11                   529474
## 12 12                   464764
```

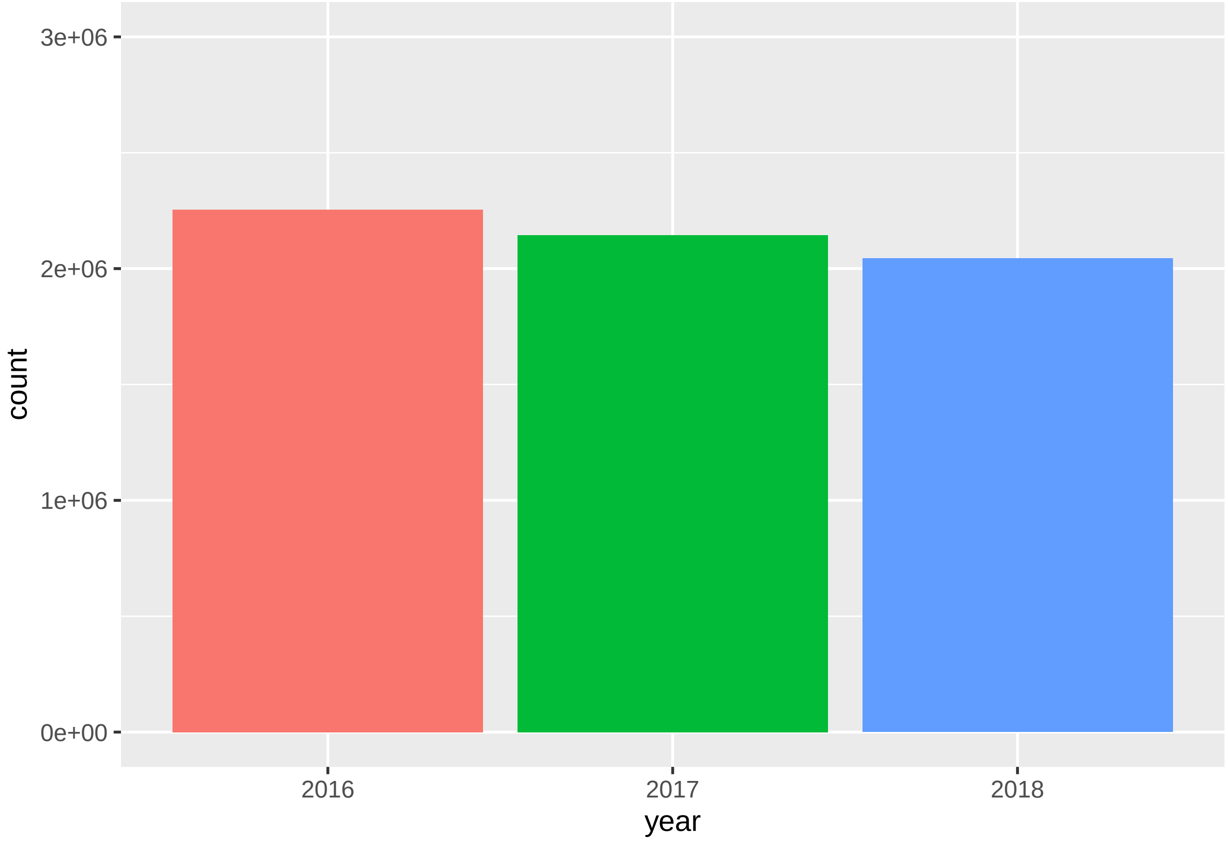
3.1.5c Impact of season of week in all infractions

```
parking_ticket %>%
  group_by(season) %>%
  summarise(number_of_infractions = n())
```

```
## # A tibble: 4 x 2
##   season number_of_infractions
##   <fct>                  <int>
## 1 Winter               1497414
## 2 Spring               1717610
## 3 Summer               1634753
## 4 Fall                 1593007
```

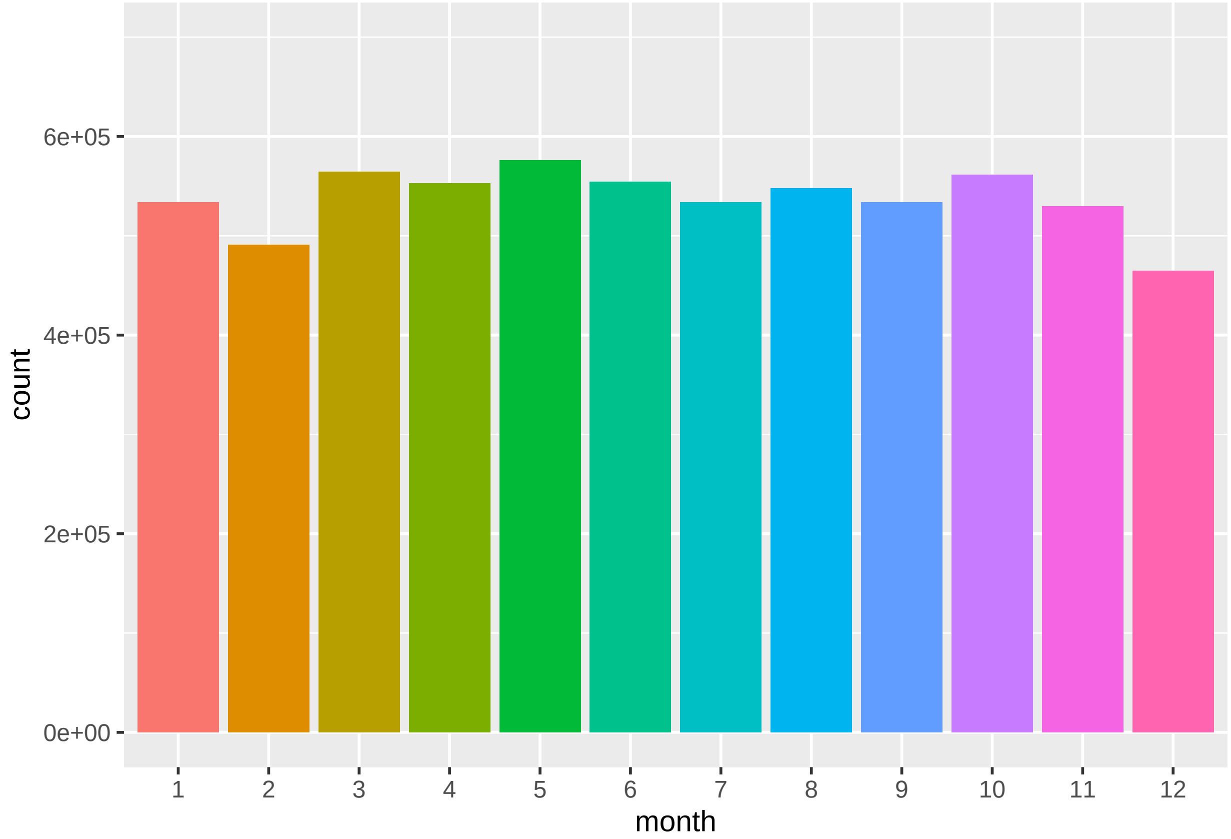
3.3.1.a.i Distribution of infractions by year

```
parking_ticket %>%
  group_by(year) %>%
  summarise(count=n()) %>%
  ggplot(aes(x=year,y=count)) +
    geom_bar(aes(fill = year), position = "dodge", stat="identity")+
    # geom_point(aes(x = year, y = count), size = 1.5, color="black", group = 2) +
    # geom_line(aes(x = year, y = count), size = 1.5, color="red", group = 1) +
    ylim(0,3000000) +
    theme(legend.position = "none")
```



3.3.1.a.ii Distribution of infractions by month

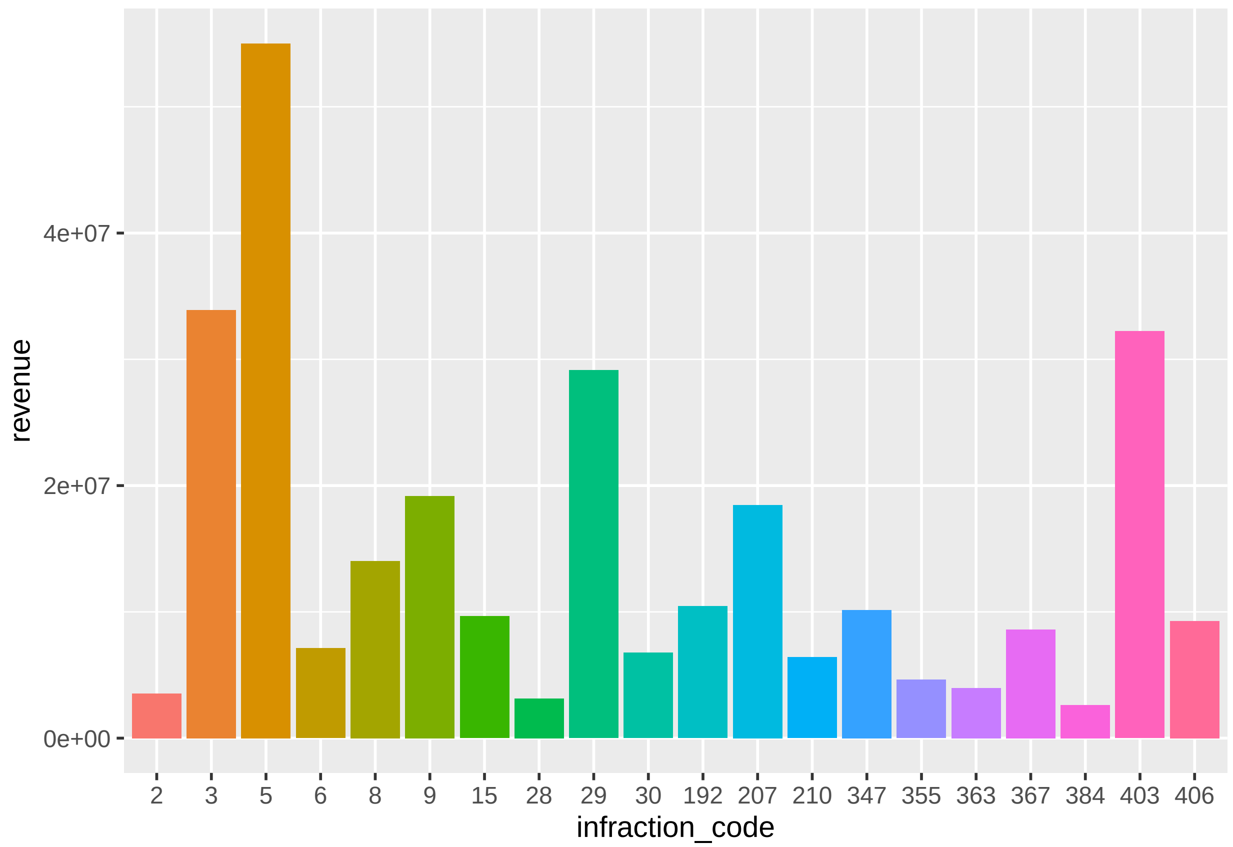
```
parking_ticket %>%
  group_by(month) %>%
  summarise(count=n()) %>%
  ggplot(aes(x=month,y=count)) +
  geom_bar(aes(fill = month), position = "dodge", stat="identity")+
  ylim(0,700000)+
  theme(legend.position = "none")
```



```
  # geom_point(aes(x = month, y = count), size = 1.5, color="black", group = 1) +
  # geom_line(aes(x = month, y = count), size = 1.5, color="blue", group = 1)
```

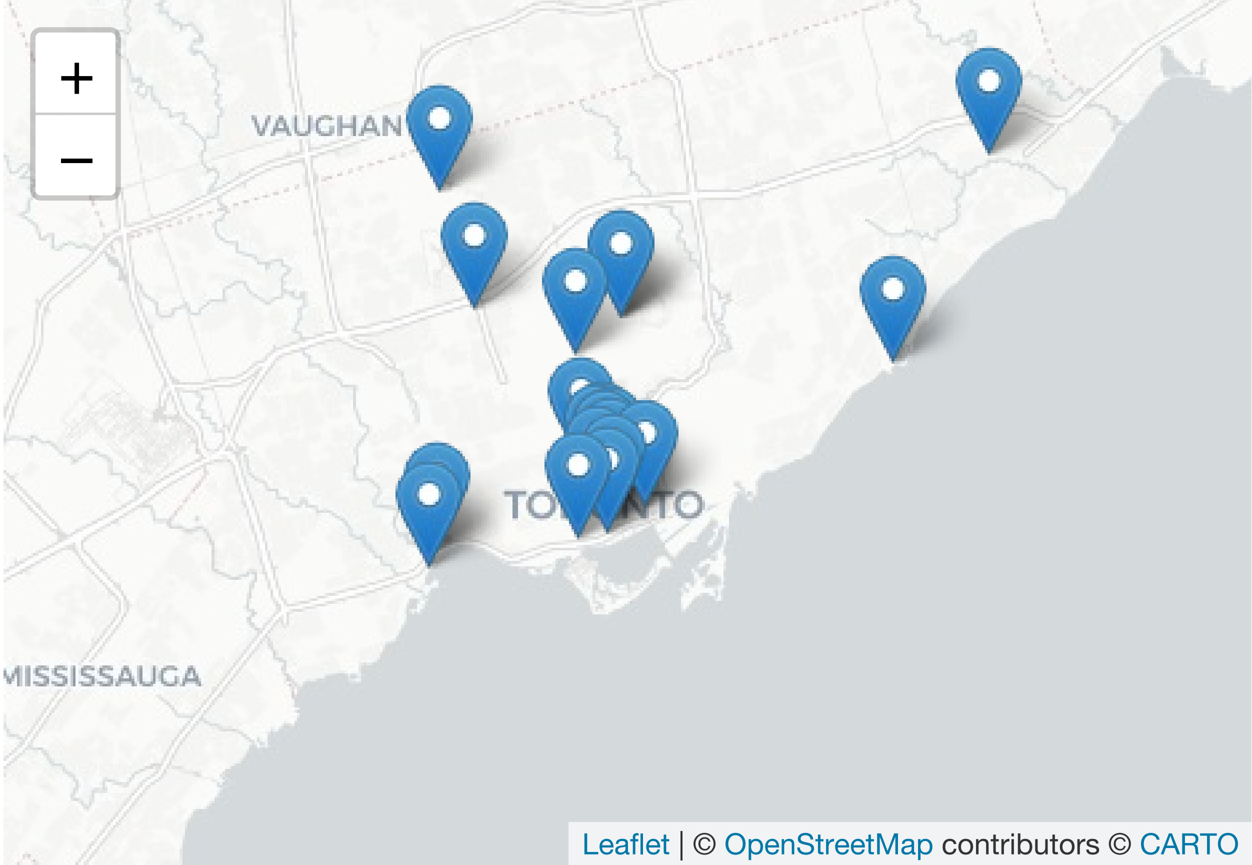
3.3.1.b Distribution of top 20 infractions by fines

```
top_20_revenue <- top_20_revenue[order(top_20_revenue$infraction_code),]
top_20_revenue$infraction_code <- as.factor(top_20_revenue$infraction_code)
top_20_revenue %>%
  ggplot(aes(x=infraction_code,y=revenue)) +
  geom_bar(aes(fill = infraction_code), position = "dodge", stat="identity")+
  theme(legend.position = "none")
```



3.3.2.a Geographic distribution (location) of top 20 infractions (count)

```
top_20_infraction_location_map <- leaflet(sp_top_20_locations_geocoded) %>%
  addProviderTiles("CartoDB.Positron") %>%
  setView(lng = -79.372573, lat = 43.679434, zoom = 10) %>%
  addMarkers(popup = paste("Address: ", sp_top_20_locations_geocoded$location2, "<br>",
                           "Total number of infractions: ", sp_top_20_locations_geocoded$freq, "<br>"))
top_20_infraction_location_map
```



3.3.2.b Geographic distribution by ward for top 20 infractions (count)

Interpretation: find the total number of infractions within each ward for all top 20 infraction locations.

```
# Import City of Toronto ward

sf_city_ward <- sf::st_read('City Wards/City Wards Data/City Wards Data.shp',
                            stringsAsFactors=F, options = "ENCODING=UTF8")
```

```
## options:        ENCODING=UTF8
## Reading layer `City Wards Data' from data source `/Users/jasonjiang/Desktop/CityOfToronto/City Wards
## Simple feature collection with 25 features and 25 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: -79.63926 ymin: 43.581 xmax: -79.11545 ymax: 43.85546
## CRS:            4326
```

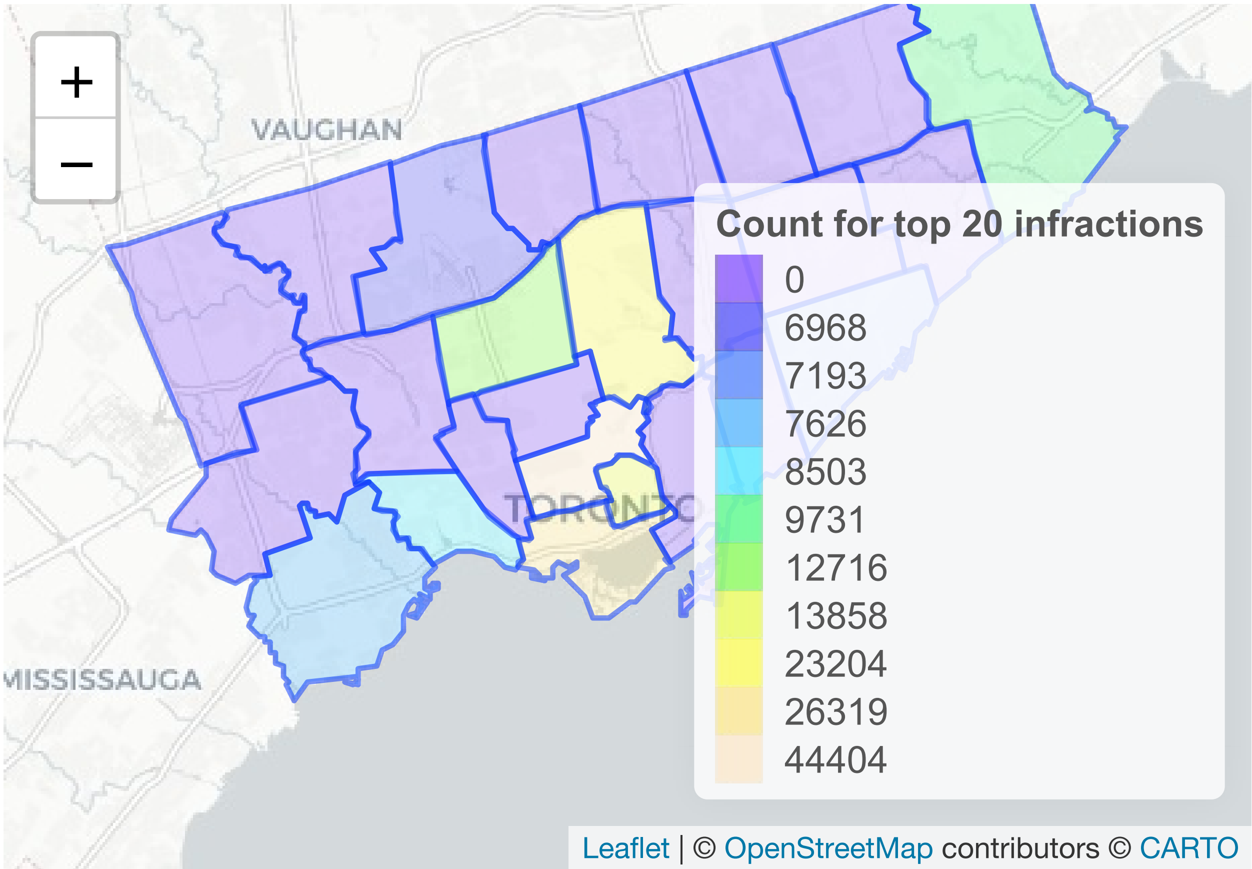
```
city_ward_col_names <- read.csv('City Wards/City Wards Data/City Wards Data_fields.csv',
                                stringsAsFactors = F)
colnames(sf_city_ward) <- city_ward_col_names$name
head(sf_city_ward)
```

```
## Simple feature collection with 6 features and 25 fields
## geometry type:  POLYGON
## dimension:      XY
## bbox:           xmin: -79.58521 ymin: 43.64933 xmax: -79.21276 ymax: 43.80327
## CRS:            4326
##    _id AREA_ID      DATE_EFFECTIVE         DATE_EXPIRY AREA_ATTR_ID
## 1 2001 2457740 2018-08-07T18:11:06 3000-01-01T05:00:00     25993196
## 2 2002 2457739 2018-08-07T18:11:06 3000-01-01T05:00:00     25993195
## 3 2003 2457738 2018-08-07T18:11:06 3000-01-01T05:00:00     25993194
## 4 2004 2457737 2018-08-07T18:11:06 3000-01-01T05:00:00     25993193
## 5 2005 2457736 2018-08-07T18:11:06 3000-01-01T05:00:00     25993192
## 6 2006 2457735 2018-08-07T18:11:06 3000-01-01T05:00:00     25993191
##   AREA_TYPE_ID PARENT_AREA_ID AREA_TYPE AREA_CLASS_ID AREA_CLASS
## 1          528             NA      CITW            NA         NA
## 2          528             NA      CITW            NA         NA
## 3          528             NA      CITW            NA         NA
## 4          528             NA      CITW            NA         NA
## 5          528             NA      CITW            NA         NA
## 6          528             NA      CITW            NA         NA
##   AREA_SHORT_CODE AREA_LONG_CODE                AREA_NAME
## 1               7              7 Humber River-Black Creek
## 2               6              6              York Centre
## 3              18             18               Willowdale
## 4              11             11      University-Rosedale
## 5              19             19        Beaches-East York
## 6              20             20    Scarborough Southwest
##                      AREA_DESC FEATURE_CODE FEATURE_CODE_DESC TRANS_ID_CREATE
## 1 Humber River-Black Creek (7)           NA                NA          279754
## 2              York Centre (6)           NA                NA          279754
## 3              Willowdale (18)           NA                NA          279754
## 4     University-Rosedale (11)           NA                NA          279754
## 5       Beaches-East York (19)           NA                NA          279754
## 6   Scarborough Southwest (20)           NA                NA          279754
##   TRANS_ID_EXPIRE  X  Y LONGITUDE LATITUDE OBJECTID Shape__Area Shape__Length
## 1              -1 NA NA        NA       NA 17344785    58868733      43438.94
## 2              -1 NA NA        NA       NA 17344801    67805389      40910.17
## 3              -1 NA NA        NA       NA 17344817    37926494      24767.00
## 4              -1 NA NA        NA       NA 17344833    26002991      29861.63
## 5              -1 NA NA        NA       NA 17344849    32155033      30975.88
## 6              -1 NA NA        NA       NA 17344865    53987873      45132.35
##                         geometry
## 1 POLYGON ((-79.49105 43.7635...
## 2 POLYGON ((-79.44043 43.7634...
## 3 POLYGON ((-79.39449 43.7615...
## 4 POLYGON ((-79.39004 43.6905...
## 5 POLYGON ((-79.29864 43.7151...
## 6 POLYGON ((-79.27903 43.6716...
```

```r
# First determine the polygon each location belongs to

top_20_locations_geocoded$AREA_NAME <- apply(top_20_locations_geocoded[,c('long','lat')], 1, function(r
  # Transformation to palnar (epsg:2163) is required, since sf library assumes planar projection
  sf_city_ward_pl <- st_transform(sf_city_ward$geometry, 2163)
  coords <- c(as.numeric(row[1]),as.numeric(row[2]))
  pnt_sf <- st_transform(st_sfc(st_point(coords),crs = 4326), 2163)
  # Use st_intersects to see which polygon does the point intersects with
  area <- sf_city_ward[which(st_intersects(pnt_sf, sf_city_ward_pl, sparse = FALSE)), ]$AREA_NAME
  return(area)
})

top_20_locations_geocoded
```

```
## # A tibble: 20 x 6
##    location2        freq location_new                 lat  long AREA_NAME
##    <chr>           <int> <chr>                      <dbl> <dbl> <chr>
##  1 20 EDWARD ST    16612 20 EDWARD ST,TORONTO,ON,C~  43.7 -79.4 University-Rose~
##  2 2075 BAYVIEW A~ 16422 2075 BAYVIEW AVE,TORONTO,~  43.7 -79.4 Don Valley West
##  3 1265 MILITARY ~  9731 1265 MILITARY TRL,TORONTO~  43.8 -79.2 Scarborough-Rou~
##  4 LA PLANTE AVE    8680 LA PLANTE AVE,TORONTO,ON,~  43.7 -79.4 University-Rose~
##  5 199 RICHMOND S~  8664 199 RICHMOND ST W,TORONTO~  43.6 -79.4 Spadina-Fort Yo~
##  6 103 THE QUEENS~  8503 103 THE QUEENSWAY,TORONTO~  43.6 -79.5 Parkdale-High P~
##  7 JAMES ST         8066 JAMES ST,TORONTO,ON,CANADA  43.7 -79.4 Toronto Centre
##  8 15 MARINE PARA~  7626 15 MARINE PARADE DR,TORON~  43.6 -79.5 Etobicoke-Lakes~
##  9 1 BRIMLEY RD S   7193 1 BRIMLEY RD S,TORONTO,ON~  43.7 -79.2 Scarborough Sou~
## 10 3401 DUFFERIN ~  7176 3401 DUFFERIN ST,TORONTO,~  43.7 -79.5 Eglinton-Lawren~
## 11 1000 FINCH AVE~  6968 1000 FINCH AVE W,TORONTO,~  43.8 -79.5 York Centre
## 12 2075 BAYVIEW AV  6782 2075 BAYVIEW AV,TORONTO,O~  43.7 -79.4 Don Valley West
## 13 EDWARD ST        6563 EDWARD ST,TORONTO,ON,CANA~  43.7 -79.4 University-Rose~
## 14 150 GERRARD ST~  6503 150 GERRARD ST W,TORONTO,~  43.7 -79.4 University-Rose~
## 15 150 DAN LECKIE~  6325 150 DAN LECKIE WAY,TORONT~  43.6 -79.4 Spadina-Fort Yo~
## 16 273 BLOOR ST W   6046 273 BLOOR ST W,TORONTO,ON~  43.7 -79.4 University-Rose~
## 17 250 FRONT ST E   5792 250 FRONT ST E,TORONTO,ON~  43.7 -79.4 Toronto Centre
## 18 19 GRAND TRUNK~  5678 19 GRAND TRUNK CRES,TORON~  43.6 -79.4 Spadina-Fort Yo~
## 19 WELLINGTON ST W  5652 WELLINGTON ST W,TORONTO,O~  43.6 -79.4 Spadina-Fort Yo~
## 20 40 ORCHARD VIE~  5540 40 ORCHARD VIEW BLVD,TORO~  43.7 -79.4 Eglinton-Lawren~
```

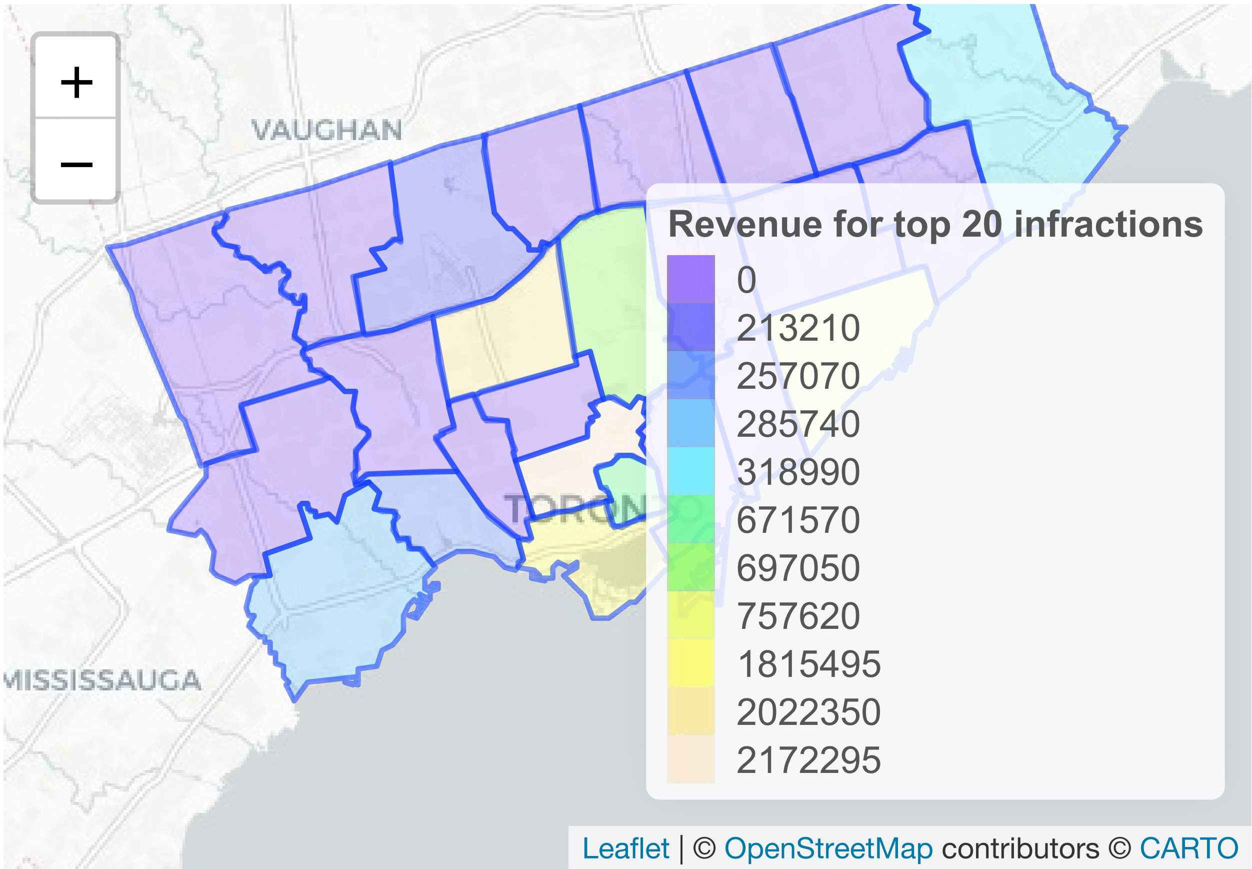
```
# Then create a frequency table

count_20_locations_by_ward <- top_20_locations_geocoded %>%
  group_by(AREA_NAME) %>%
  summarise(count=sum(freq)) %>%
  select(AREA_NAME,count) %>%
  right_join(sf_city_ward,by='AREA_NAME') %>%
  select(AREA_NAME,count,geometry) %>%
  replace(is.na(.), 0)

# Finally, plot it out

unique_color_df <- data.frame(count = sort(unique(count_20_locations_by_ward$count)),
                              color = gsub(".{2}$","",topo.colors(length(unique(count_20_locations_by_wa

colors <- unique_color_df %>%
  right_join(count_20_locations_by_ward,by='count') %>%
  select(color)
colors <-  as.vector(as.character(colors[,1]))

leaflet(sf_city_ward) %>%
  addProviderTiles("CartoDB.Positron") %>%
  setView(lng = -79.372573, lat = 43.679434, zoom = 10) %>%
  addPolygons(data = count_20_locations_by_ward$geometry, weight = 2,
              popup = paste('Area:',count_20_locations_by_ward$AREA_NAME, "<br>",
                            'Count:',count_20_locations_by_ward$count, "<br>"),
              fillColor = colors) %>%
  addLegend(position = 'bottomright',
            colors=unique_color_df$color,
            labels=unique_color_df$count,
            title='Count for top 20 infractions')
```


+
−
VAUGHAN
TORONTO
MISSISSAUGA
Count for top 20 infractions
0
6968
7193
7626
8503
9731
12716
13858
23204
26319
44404
Leaflet | © OpenStreetMap contributors © CARTO

3.3.2.c Geographic distribution by ward for top 20 infractions (revenue)

Interpretation: find the revenue of infractions within each ward and find the top 20 wards

```
# First retrieve aggregated sum revenues per location
revenue_df <- parking_ticket[(parking_ticket$location2 %in% top_20_locations_geocoded$location2),] %>%
  group_by(location2) %>%
  summarise(revenue=sum(set_fine_amount)) %>%
  as.data.frame() %>%
  right_join(top_20_locations_geocoded,by='location2')

# Then create a table with summed revenue

revenue_20_locations_by_ward <- revenue_df %>%
  group_by(AREA_NAME) %>%
  summarise(sum=sum(revenue)) %>%
  select(AREA_NAME,sum) %>%
  right_join(sf_city_ward,by='AREA_NAME') %>%
  select(AREA_NAME,sum,geometry) %>%
  replace(is.na(.), 0)

# Finally, plot it out

unique_color_df <- data.frame(sum = sort(unique(revenue_20_locations_by_ward$sum)),
                              color = gsub(".{2}$","",topo.colors(length(unique(revenue_20_locations_by_

colors <- unique_color_df %>%
  right_join(revenue_20_locations_by_ward,by=c('sum')) %>%
  select(color)
colors <-  as.vector(as.character(colors[,1]))

leaflet(sf_city_ward) %>%
  addProviderTiles("CartoDB.Positron") %>%
  setView(lng = -79.372573, lat = 43.679434, zoom = 10) %>%
  addPolygons(data = revenue_20_locations_by_ward$geometry, weight = 2,
              popup = paste('Area:',revenue_20_locations_by_ward$AREA_NAME, "<br>",
                            'Count:',revenue_20_locations_by_ward$sum, "<br>"),
              fillColor = colors) %>%
  addLegend(position = 'bottomright',
            colors=unique_color_df$color,
            labels=unique_color_df$sum,
            title='Revenue for top 20 infractions')
```


+
−
VAUGHAN
MISSISSAUGA
TORONTO
Revenue for top 20 infractions
0
213210
257070
285740
318990
671570
697050
757620
1815495
2022350
2172295
Leaflet | © OpenStreetMap contributors © CARTO