January 2020 Parking Violations

Parking Tickets from Washington D.C.’s Open Data Site

justjensen.co

16 Oct, 2020

Contents

This notebook provides information about parking tickets in Washington D.C.

Aquiring the Data

First, we'll need to grab data directly in R using the [Parking Ticket dataset]'s API.

```
# url <- 'https://opendata.arcgis.com/datasets/009dedf-
 ↪  baf364905a8e25181b3490cd9_0.geojson'
# destination_file <- 'january_2020_dc_parking_tickets.geojson'
# download.file(url, destination_file, 'curl')
# library(rgdal)
# sp_parking_tickets <- readOGR('january_2020_dc_parking_tickets.geojson')

library(tidyverse)
url <- 'https://opendata.arcgis.com/datasets/009dedf-
 ↪  baf364905a8e25181b3490cd9_0.csv'
destination_file <- 'january_2020_dc_parking_tickets.csv'
download.file(url, destination_file, 'curl')
df_violations <- read_csv(destination_file)
colnames(df_violations) <- tolower(colnames(df_violations))
```

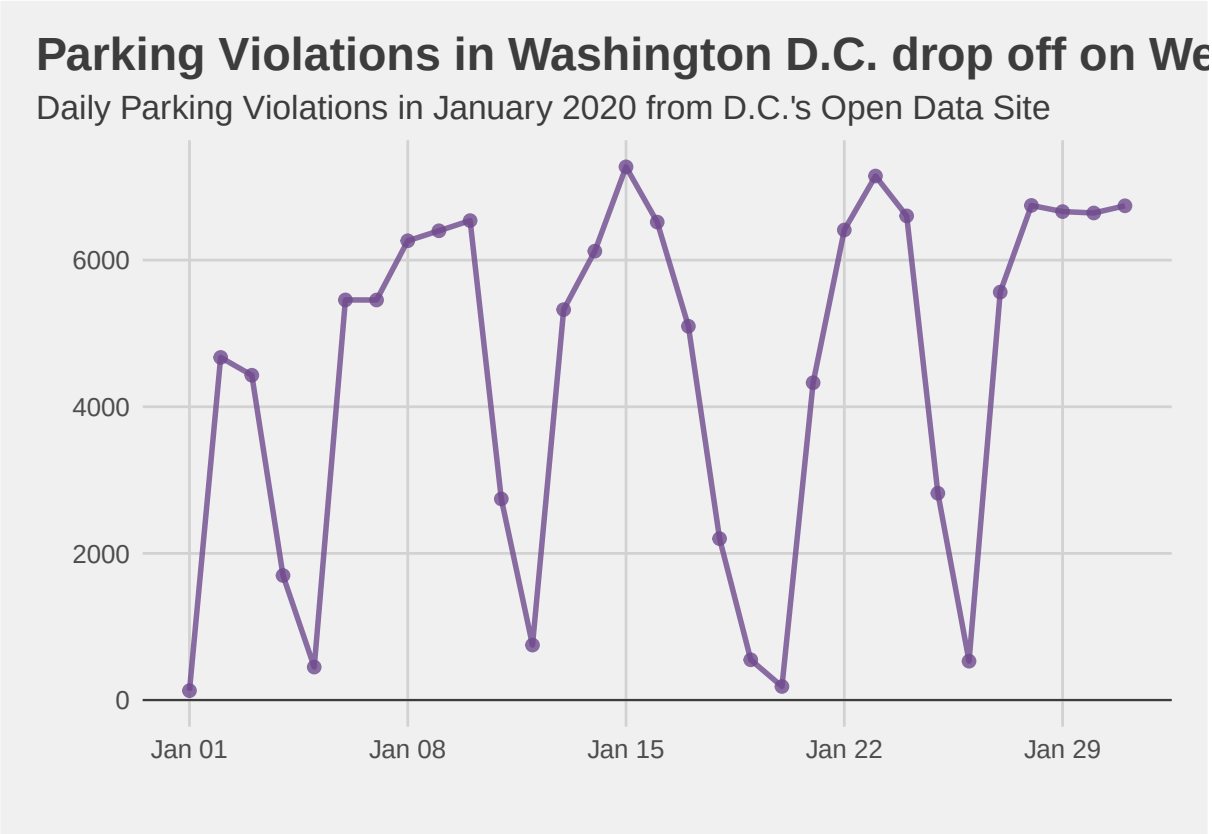
The first violation in our data set is a 'PARK IN OFFICIAL PARKING PERMIT ONLY SPACE' violation.

Before going further, we'll need to do a little more work to get our dataframe set up properly!

```
df_violations <- df_violations[c('objectid', 'issue_date',
 ↪  'issuing_agency_code',
                                 'issuing_agency_name',
                                 ↪  'issuing_agency_short',
                                 ↪  'violation_code',
                                 'violation_proc_desc', 'location',
                                 ↪  'disposition_type',
                                 'disposition_date', 'fine_amount',
                                 ↪  'total_paid', 'latitude',
                                 'longitude', 'mar_id',
                                 ↪  'gis_last_mod_dttm')]
cat(paste('#', 'This is a heading for Object', df_violations$objectid[1], '
 ↪  \n'))
```

Creating a Chart of Violations

```
df_violations$issue_date <- as.Date(df_violations$issue_date,
 ↪  format='%Y/%m/%d')
df_violations_per_day <- df_violations %>% count(issue_date)
plt <- ggplot(df_violations_per_day) +
  geom_hline(yintercept=0, size=0.4, color='#3C3C3C')+
  geom_line(aes(x=issue_date, y=n), color='#6f4a8e', alpha=0.8, size=1) +
  geom_point(aes(x=issue_date, y=n), color='#6f4a8e', alpha=0.8, size=2) +
  labs(x='', y='', title='Parking Violations in Washington D.C. drop off on
 ↪  Weekends',
       subtitle="Daily Parking Violations in January 2020 from D.C.'s Open
        ↪  Data Site") +
  scale_x_date(date_labels='%b %d', breaks=seq(as.Date('2020-01-01'),
 ↪  as.Date('2020-01-31'), by='weeks')) +
  theme(text=element_text(size=12, color='#3C3C3C'),
        plot.title=element_text(hjust=0, size=rel(1.5), face='bold'),
        plot.subtitle = element_text(hjust=0, size=rel(1.1)),
        plot.caption=element_text(hjust=0),
        plot.title.position = 'plot',
        plot.background = element_rect(fill='#F0F0F0'),
         ↪  axis.ticks=element_blank(),
        panel.background = element_rect(fill='#F0F0F0'),
         ↪  panel.grid=element_line(color=NULL),
        panel.grid.major=element_line(color='#d2d2d2'),
         ↪  panel.grid.minor=element_blank(),
        strip.background=element_blank(),
         ↪  strip.text=element_text(face='bold'),
        plot.margin=unit(c(1,1,1,1), 'lines'))
print(plt)
```




```
ggsave('Parking Violations in Washington DC.png', plot=plt, type='cairo',
       height=5,width=8,units='in')
```

Creating a Table of Violations

```
library(kableExtra)
df_violations$fine_paid <-
 ↪  ifelse(df_violations$fine_amount==df_violations$total_paid,
                                  1, 0)

df_violations$fine_bin <- case_when(
  df_violations$fine_amount < 50 ~ '<$50',
  df_violations$fine_amount < 100 ~ '$50 - $99',
  df_violations$fine_amount < 200 ~ '$100 - $199',
  df_violations$fine_amount >= 200 ~ '$200+'
)
```

```
# Preparing the final dataframe for table generation
df_violations_fines <- df_violations %>%
  drop_na(fine_amount) %>%
  group_by(fine_bin) %>%
  summarise('Tickets (thousands)'=round(length(fine_paid)/1000,1),
            'Percent
             ↪  Paid'=paste0('%',round(sum(fine_paid)/length(fine_paid),3)*100))
             ↪  %>%
  gather('Tickets (thousands)', 'Percent Paid', key='value_type',
 ↪  value='value') %>%
  spread('fine_bin', 'value') %>%
  select(value_type, '<$50', '$50 - $99', '$100 - $199', '$200+')

colnames(df_violations_fines)[1] <- 'Fine Amount'

kb <- kbl(df_violations_fines, format = 'latex',
          booktabs=T, digits=1, linesep='', align=c('lrrrr')) %>%
  kable_styling(latex_options = 'striped') %>%
  add_header_above(c(' ', 'Cheaper'=2, 'More Expensive'=2))

# kb <- kbl(df_violations_fines,
#             'latex', booktabs=T, digits=2, linesep='',
#             col.names = c('', 'Hours', strftime(recent_weeks[1],
 ↪  '%Y-%m-%d'),
#                           strftime(recent_weeks[2], '%Y-%m-%d'),'Perc.
 ↪  Diff'),
#             align=c('lrrrr')) %>%
#   kable_styling(full_width = F, latex_options = 'striped')
print(kb)
```

	Cheaper		More Expensive	
Fine Amount	<$50	$50 - $99	$100 - $199	$200+
Percent Paid	%30.7	%22.7	%19.3	%10.3
Tickets (thousands)	62.3	40.6	27.2	3.3