Comparison of the number of six crimes in Toronto from 2014 to 2019

Abstract

This blog mainly uses the R data visualization package ggplot2 to compare crime in the urban area around Toronto.

Introduction

The data sets used in this blog are from the TORONTO OPEN DATA PORTAL. A brief introduction to the data set: This data set includes regional crime data from 2014 to 2019. This dataset can be used to count assaults, car thefts, break-ins, robberies, burglaries and killings. The data also include five-year average crime rates by community and crime per 100,000 people, based on 2016 census data.

Discussion

First, we import the data into R:

```
library(opendatatoronto)
library(dplyr)

# get package
package <- show_package("fc4d95a6-591f-411f-af17-327e6c5d03c7")
package

# get all resources for this package
resources <- list_package_resources("fc4d95a6-591f-411f-af17-327e6c5d03c7")

# identify datastore resources; by default, Toronto Open Data sets datastore
#resource format to CSV for non-geospatial and GeoJSON for geospatial resources
datastore_resources <- filter(resources, tolower(format) %in% c('csv', 'geojson'))

# load the first datastore resource as a sample
data <- filter(datastore_resources, row_number()==1) %>% get_resource()
```

Then, we can have a look at the data.

Due to the limited scope of the screen, I only captured part of the data as shown in the figure below.

_id	OBJECTID	Neighbourhood	Hood_ID	Population	Assault_2014	Assault_2015	Assault_2016	Assault_2017	Assault_2018	Assault_2019	Assault_AVG
1	16	South Parkdale	085	21849	202	226	231	229	220	251	226.5
2	17	South Riverdale	070	27876	215	207	236	243	304	261	244.3
3	18	St.Andrew-Windfields	040	17812	53	41	48	45	55	58	50.0
4	19	Taylor-Massey	061	15683	127	92	97	107	123	122	111.3
5	20	Humber Summit	021	12416	76	89	118	116	109	118	104.3
6	21	Humbermede	022	15545	117	132	114	157	119	161	133.3
7	22	Centennial Scarborough	133	13362	50	39	48	48	46	48	46.5
8	23	Church-Yonge Corridor	075	31340	480	495	543	694	766	879	642.8
9	24	Clairlea-Birchmount	120	26984	247	259	244	243	288	282	260.5
10	25	Cliffcrest	123	15935	79	97	106	114	132	133	110.2
11	26	Flemingdon Park	044	21933	128	147	153	122	145	152	141.2

Then, we use the package ggplot2 to visualize the data. We can compare the number of the six types of crime in different regions to get a more intuitive understanding of the data.

First, select the column we are analyzing, codes as follows:

```
data2 <- data %>%
  select(Neighbourhood,Assault_AVG,
         AutoTheft_AVG,
         BreakandEnter_AVG,
         Homicide_AVG,
         Robbery_AVG,
         TheftOver_AVG)

data.frame(data2)
```

This is the data look like after processing:

	Neighbourhood	Assault_AVG	AutoTheft_AVG	BreakandEnter_AVG	Homicide_AVG	Robbery_AVG	TheftOver_AVG
1	South Parkdale	226.5	18.7	65.3	0.3	33.0	10.0
2	South Riverdale	244.3	30.8	108.8	1.8	49.0	21.3
3	St.Andrew-Windfields	50.0	20.2	78.7	0.5	7.2	8.5
4	Taylor-Massey	111.3	9.3	46.0	0.7	21.0	3.5
5	Humber Summit	104.3	82.3	48.7	1.2	25.7	17.3
6	Humbermede	133.3	49.3	36.7	0.3	28.2	6.0
7	Centennial Scarborough	46.5	9.5	19.2	0.3	7.5	2.0
8	Church-Yonge Corridor	642.8	37.8	188.5	2.0	135.7	33.8
9	Clairlea-Birchmount	260.5	44.5	94.2	1.2	41.0	11.7
10	Cliffcrest	110.2	16.5	48.2	0.3	23.8	4.2
11	Flemingdon Park	141.2	7.8	21.0	0.2	20.0	3.8

Second, we load the ggplot2 package and then call the histogram function.

```
ggplot(data2, aes(x=Neighbourhood, y=Assault_AVG)) +
  geom_bar(stat='identity')
```

The results are as follows, we have got a picture:
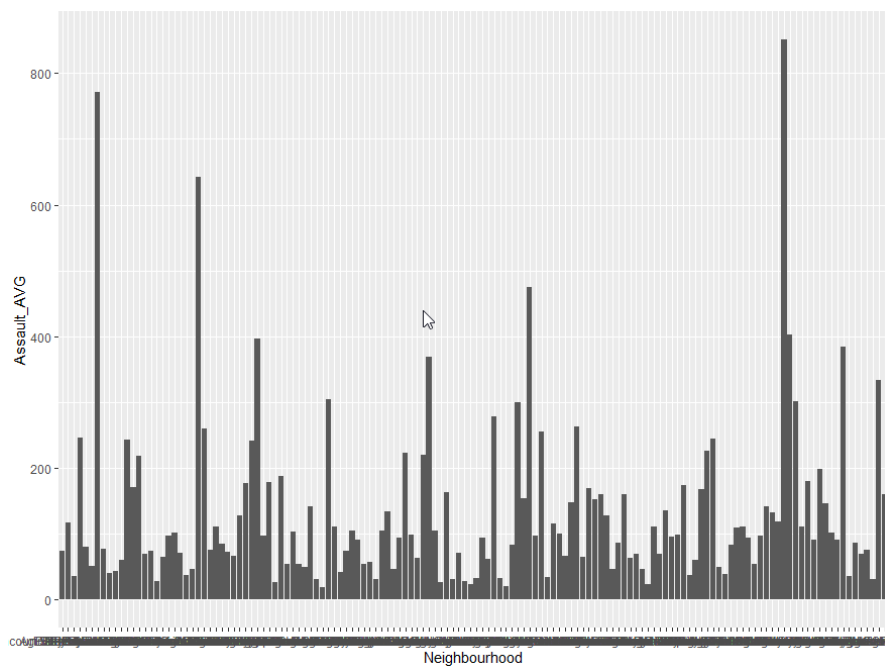
The treatment of other types of crime is similar to the above treatment and will not be repeated here.

Reference

Grolemund, G., & Wickham, H. (2018). R for data science.

Kabacoff, R. I. (2010). *R in Action*. manning.