Statistical Analysis Based on the Police Annual Statistical Report of Victims of Crime

Qianyuan_Huang

02/05/2022

Abstract

The data of victims of crime are important resource in analyzing the funding distribution of police department and allocation of police resources. This report analyzes the Police Annual Statistical Report (ASR) to investigate: (1)the crime trend from 2014 to 2020 through the number of crime victims (2) the bias of the dataset and the possible influence to the result (3)the distribution of victims in different age group in each categories of crime. As a result, the crime rate from 2014 to 2020 increases around 8% in GTA area, where assault crime contributes largest percentage of increase in all categories of crime.

Introduction

Toronto used to be considered as a relatively safe city in the world. In 2017, Toronto was ranked as the 24th safest city in the world by Economics. Although the rank dropped to 95 in 2018, it was stilled one of the top 10 safest city in North America. After 2005 the "year of guns" and a long stable period from 2009 to 2015 (Statistics Canada, 2021)Canada (2021), the public started not to be satisfied with the crime rate in Toronto once again. People even started to protest against the over-funding for the police department. Most of the protestors were once victims or families of victims in crime. They conveyed their dissatisfaction to the government including the compensation for crime victims, penalty to the criminals, and the increasing number of crime victims in the police report every year.

The purpose of this analysis is to investigate the crime in Toronto through various age groups of victims within different categories of crimes. Statistical reports are essential tools for the government or police department to relocate the funding. By analyzing the relationship between the age group of victims in every categories of crime, the funding could be used more efficiently. In this report I will use the Police Annual Statistical Report of crime victims to acquire the data I need, analyze the reliability of the report and process the data with R markdown.

The data is included in The Toronto's Police Service Annual Statistical Report (ASR), which is a comprehensive overview of police related statistics including reported crimes, victims of crime, search of persons, firearms, traffic collisions, personnel, budget, communications, public complaints, regulated interactions and other administrative information. The data I used in this report includes all identified victims of crimes against the person from 2014 to 2020. The crime could be roughly categorized as assault, robbery, sexual violation, and others. Most of the data was collected from official police report where the victims were found or unfound inside or outside of GTA.

However, from this particular resource, some of the data was not collected directly by respondent but by third-party police report and the count of the victim might not be as accurate as anticipated. Since the victims were also included when they were not found, the data could not be perfectly accurate. There might be situations such as no response, where the only source of the data was the description from witness or criminal. However, when describing crime or victims witness of violent crime seem to be more cautious and inaccurate due to the fear towards the criminal (Tobolowsky 2001)Tobolowsky (2001). The age, category, or even gender as subjective factors could also influence the description. Duplicate persons were also possible if the same person was victimized for the same offense during the same period. Therefore the actual count could be smaller than the number in the data. At the same time, as this data only include three main categories and no more than ten subtypes of crime, it might not be detailed enough for all the victims and

crime between 2014 and 2020. Owing to the inaccuracy of the data, I believe it is necessary to avoid using the data with unknown gender or age group. The reliability of the unknown data could be far worse than the others. Mixing all the data together could cause profound influence to the whole dataset inevitably.

The rest of the data could be collected from internet where people could report themselves as victims of crime through official website such as canada.ca, Canadian Resource Center, Department of Justice or local victim service. As the terminology "crime" could be misunderstood by people due to subjective understanding, the credibility of self-report victims could be far from that collected from police report. Sometimes people including criminal themselves could considered themselves as victim and this will surely bring huge inaccuracy in ASR. Under a more radical circumstance people could pretend to be the victims in order to get access to the government funding and gain compensation for the actual crime victims. The victims might also repeatedly report themselves due to serious mental damage as the aftermath of victimization (Greenberg 2012)Martin S. Greenberg (2012). Because of these two aspects the dataset could be overestimated, the true data could be less than the statistical number in the reports.

Data

The dataset of crime victim contains 12 variables with 854 observation from 2014 to 2020. The 12 variables were ID, Index, reported year, category, subtype, assault subtype, gender, age group, age cohort, count, object ID, and geometry. ID and Index were the unique identifier for Open Data database. Reported year was the year the crime victims were reported. Category and subtype were the type of crime including assault, robbery, and sexual assault. Assault subtype was broken into a more specific categorization of crime. Gender, age group and age cohort were the sex and age identification of the victims. Count was the count of identified victims, although duplicate persons are possible if the same person was victimized for the same offense during the same period.

I used R to clean the data set with tidyverse@R1 and Knitr@R2. Because of the strong correlation between some of the variables I deleted the columns in the data that would not have contribution to my analysis. Then I mutated the variable and categorized the crime count into five categories named "A" to "E". Due to the unreliability of the Unknown data, I filtered all the observations with unknown gender or unknown age group and named the final cleaned dataset as data4.

According to the ASR crime category, all crime was divided into 4 categories. As the dataset of crime victims only gives the information of each observation respectively. It is necessary to investigate the relationship between the number of victims in every category according to the reported years. Figure 1 shows the distribution of crime category in every year from 2014 to 2020.

```r
library(opendatatoronto)
library(dplyr)

# get package
package <- show_package("9cd09dd7-4453-43bd-af7a-caf42e565103")
package
```

```
## # A tibble: 1 x 11
##   title       id        topics civic_issues publisher excerpt dataset_category
##   <chr>       <chr>     <chr>  <chr>        <chr>     <chr>   <chr>
## 1 Police Annua~ 9cd09dd7~ <NA>   <NA>         <NA>      <NA>    <NA>
## # ... with 4 more variables: num_resources <int>, formats <chr>,
## #   refresh_rate <chr>, last_refreshed <date>
```

```r
# get all resources for this package
resources <- list_package_resources("9cd09dd7-4453-43bd-af7a-caf42e565103")

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

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

```
## # A tibble: 854 x 12
##    `_id` Index_ ReportedYear Category     Subtype AssaultSubtype   Sex   AgeGroup
##    <int> <lgl>         <int> <chr>        <chr>   <chr>            <chr> <chr>
##  1     1 NA             2014 Crimes Agai~ Assault Aggravated Pea~  M     Adult
##  2     2 NA             2014 Crimes Agai~ Assault Aggravated Pea~  M     Adult
##  3     3 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Adult
##  4     4 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Adult
##  5     5 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Adult
##  6     6 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Adult
##  7     7 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Adult
##  8     8 NA             2014 Crimes Agai~ Assault Assault Peace ~  F     Unknown
##  9     9 NA             2014 Crimes Agai~ Assault Assault Peace ~  M     Adult
## 10    10 NA             2014 Crimes Agai~ Assault Assault Peace ~  M     Adult
## # ... with 844 more rows, and 4 more variables: AgeCohort <chr>, Count_ <int>,
## #   ObjectId <int>, geometry <chr>
```
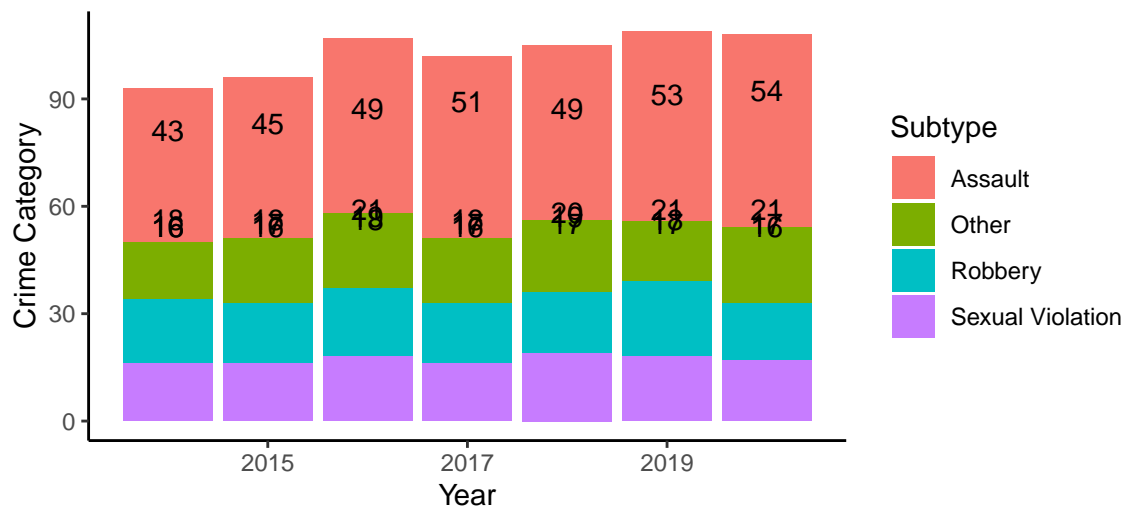
```r
data1 <-
  data %>%
  select(-geometry)
data2 <-
  data1 %>%
  select(-ObjectId)
```

```r
data4<-
  data2%>%
  filter(AgeGroup != "Unknown")
```

```r
data5 <-
  data4 %>%
  mutate(Count_=
           case_when(
             Count_ <= 50 ~ "A",
             Count_ <= 200 ~ "B",
             Count_ <= 500 ~ "C",
             Count_ <= 1000 ~ "D",
             Count_ <= 1000000 ~ "E",
             TRUE ~ "Unknown or error"
             )
  )
```

Figure 1: Year and Crime Relationship

Through figure 1 we could tell the total number of crime victims were gradually increasing from 2014 to 2020 except 2016. However, the year in this dataset only represent the year when the crime victims was reported and it also collected the victims who might be conducted crime outside of Toronto. Therefore, the data did not actually give reliable information about the relationship between the number of crime victims and year. But the graph still showed the upward trend of crime in the last six years, especially the assault crime. As the substantial decrease of crime after the 2005 "year of the gun" and a long stable period between 2009-2014(Statistics Canada, 2021)Canada (2021), the assault crime rise from 12.5% in 2014 to 15.6% in 2020. Comparing to assault events, the other category of crime seems to be at a stable rate. It is worth noting that the robbery crime In 2016 was abnormally higher than 2015 and 2014. As robbery crime were not as serious as assault or sexual violation relatively, most of the people would report crime by themselves. 2016 happened to be the first year when some of the local victims of crime services station put into use, which could be one of the reason the robbery data was higher than the other years. In the other hand, this abnormal data could also be the proof of the underestimation of the dataset, some victims might not be bothered to report themselves as victims when the crime was not serious.

```
data5 %>% filter(!is.na(Count_)) %>%
  ggplot(aes(x= Subtype, fill = AgeGroup)) + geom_bar() + theme_classic()+
  geom_text(stat = 'count',aes(label =..count.., vjust =0))+
  labs(x = "Crime Category", y = "Age Group", title = "Figure 2: Age Group and Crime Category")
```

Figure 2: Age Group and Crime Category
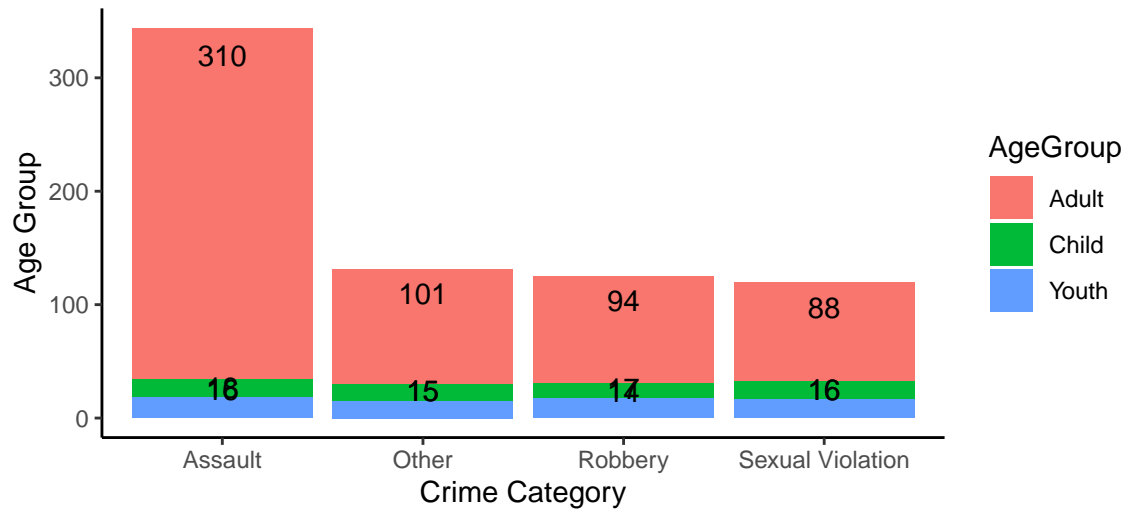
The next graph shows the age group distribution of the victims in crime and we can see that over 80% of the victims are adults. Assault crime was defined as when someone either intends to, commits, or threatens to inflict serious bodily harm on another person. The motivation for this kind of crime was always social disorganization(Du Toit, 2017)Du Toit (2017). Far away from the society and well-protected by the parents, the child and teenager were less likely to become the target of aggravated assault, thus the child and teenager victims only took 9.8% of total victims.

However, in robbery the criminals were always rational and tried to conduct the crime to fulfill their own monetary desire. Thus, teenagers and child with relatively lower chance to resist become the best target of crime and the number of these young victims took 24.8% of the total victims in robbery.

Sexual violation is defined as the act in which one intentionally sexually touches another person without that person's consent. In this specific kind of crime, the criminal was often psychologically distorted and did not have clear recognition of the consequences of their criminal behavior. At the same time due to the existence of pedophilia, the child and teenager victims in sexual violation reached 26.7%. Yet this result might still be underestimated. Firstly, some of the child victims are too young to understand the definition of sexual violation(Katz, 2021)Katz (2021). They would not report the crime to their parents who could be the only way of data collection for them. Secondly, children could be were threatened or bribed by the abusers so that they are too afraid to report the abuse to parents. The sexual violation might even cause mental health problems and as a result hide the fact that they are actually victims of sexual crime. Last but not least, parents or custodian of child victims might ignore their word and considering it as stories made up by children. They might also refuse to report this to official website or police department out of guilty, embarrassment, and fear.

```r
data4 %>% ggplot(aes(x = AgeCohort, y=Count_ ,color = Subtype)) + geom_point() + theme_classic()+
  labs(x = "Age Cohort", y = "Count", title = "Figure 3: Age Cohort in Different Crimes")
```

Figure 3: Age Cohort in Different Crimes
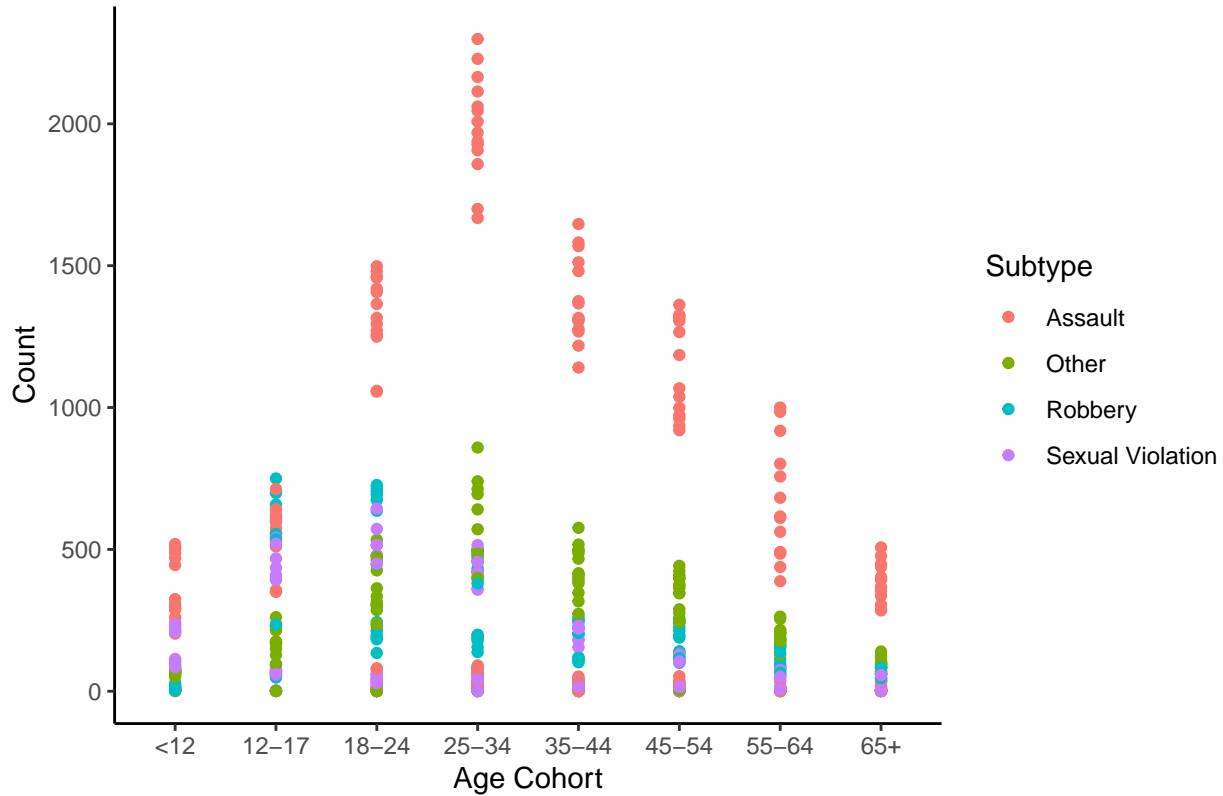

```
summary_table <- data4 %>% filter(!is.na(Count_)) %>%
  group_by(AgeCohort) %>% summarise(min = min(Count_),
                          Q1 = quantile(Count_,0.25),
                          median = median(Count_),
                          Q3 = quantile(Count_,0.75),
                          max = max(Count_),
                          IQR = Q3 - Q1,
                          mean = mean(Count_),
                          sd = sd(Count_),
                          Small_Outliers = sum(Count_ < Q1 - 1.5*IQR),
                          Large_Outliers = sum(Count_ > Q3 + 1.5*IQR))
# change their orders, add or remove some of them
knitr::kable(summary_table)
```

AgeCohort	min	Q1	median	Q3	max	IQR	mean	sd	Small_Outliers	Large_Outliers
<12	1	16	76.0	233.00	519	217.00	140.40984	151.5441	0	0
12-17	1	67	215.5	530.25	750	463.25	280.57576	240.0102	0	0
18-24	1	3	108.0	506.00	1498	503.00	347.41837	458.3143	0	10
25-34	1	6	66.0	463.75	2299	457.75	396.86607	643.7698	0	14
35-44	1	7	37.0	317.00	1647	310.00	284.74286	460.0344	0	14
45-54	1	5	31.0	262.50	1362	257.50	237.36538	383.1395	0	14
55-64	1	4	56.0	185.00	1000	181.00	159.48315	242.1584	0	12
65+	1	3	46.0	95.00	507	92.00	95.27059	136.4379	0	14

In order to further investigate the relationship of age and crime, I specified the victims' age group into

age cohort and set up a plot graph with different subtype of crime. We could tell from the figure that the distribution of victims for assault crime is roughly a normal distribution. Victims aged between 25-34 are the group with largest risk of becoming the victims of assault crime. As for sexual violation crime, the largest group of the victims were aged between 18 to 24. Yet the age group of 12-17 and 25-34 were also experiencing a high risk of exposing to sexual crime. The robbery crime targeted the victims aged 12-24 much higher than other age cohort. For the teenager aged between 12-17 the risk of getting robbed was even higher than getting assault.

reference

Canada, Statistics. 2021. "Number and Rate of Homicide Victims, by Census Metropolitan Areas."

Du Toit, Pieter. 2017. "A Fresh Perspective on Historical Sexual Abuse : The Case of Hewitt."

Katz, D., Tener. 2021. "'We Took Turns': How Do Child Victims of Intrafamilial Child Sexual Abuse Perceive and Experience Their Siblings?"

Martin S. Greenberg, R. Barry Ruback. 2012. "After the Crime: Victim Decision Making."

Tobolowsky, Peggy M. 2001. "Crime Victim Rights and Remedies."