STA304 Paper1

Suofeiya Guo 1005857208

2/5/2022

Abstract

In all parts of the world, the most important thing for people is their personal safety. Today, we are going to write a report about the crime rate in Toronto in recent years. This article will let people know about the crime situation in Toronto in recent years by recording various information of victims, so as to better protect themselves. The data for this article was obtained from the Open Toronto website. The data is authentic and valid and protects the privacy of the victims to a large extent. Data have been recorded since 2014 and were last collected in August 2021. The report found the proportion and pattern of victims in terms of sex, age and so on, as detailed in the report below.

Introduction

The most important thing for people is their personal safety. Today, we are going to write a report about the crime rate in Toronto in recent years, hoping to better protect ourselves. The data for this article was obtained from the Open Toronto website. The data is authentic and valid and protects the privacy of the victims to a large extent. The report found patterns and proportions of victims in terms of sex, age and so on. I presented the results of the report through the analysis of bar charts, pie charts and tables, including age, crime type and gender.

DATA

The background of this data

No matter what time, their life safety is one of the most concerned things. Whether it's robbery, assault, or sexual assault, it can be very damaging to the victim. Toronto also has a large number of crimes every year. The crime rate in Toronto has been on the rise for the past three years, and more and more people are worried about their safety. This set of statistics is about Police Annual Statistical Report - Victims of Crime records the number of crimes recorded by Toronto police in the past seven years, which contains a lot of information about crimes but does not involve the privacy of victims. This dataset was found in open Toronto data website, as an annually updated data, the data last refreshed in 2021.08.12.

The population of this data should be all victims in Toronto, however, the sample of this data be focus on these victims who reported their situations to the police. The frame of this should be the police. This is because not all victims choose to report their situations to the police. For example, some victims did not report to the police in time when the accident happened and did not pursue further investigation after the accident. Some victims were reluctant to report to the police for privacy reasons, or some private reasons were not handled by the police. These are all cases that are not counted. So there is a slight bias between the data and the actual situation. It is true that the situation not reported to the police cannot be changed, but it should not affect our subsequent analysis.
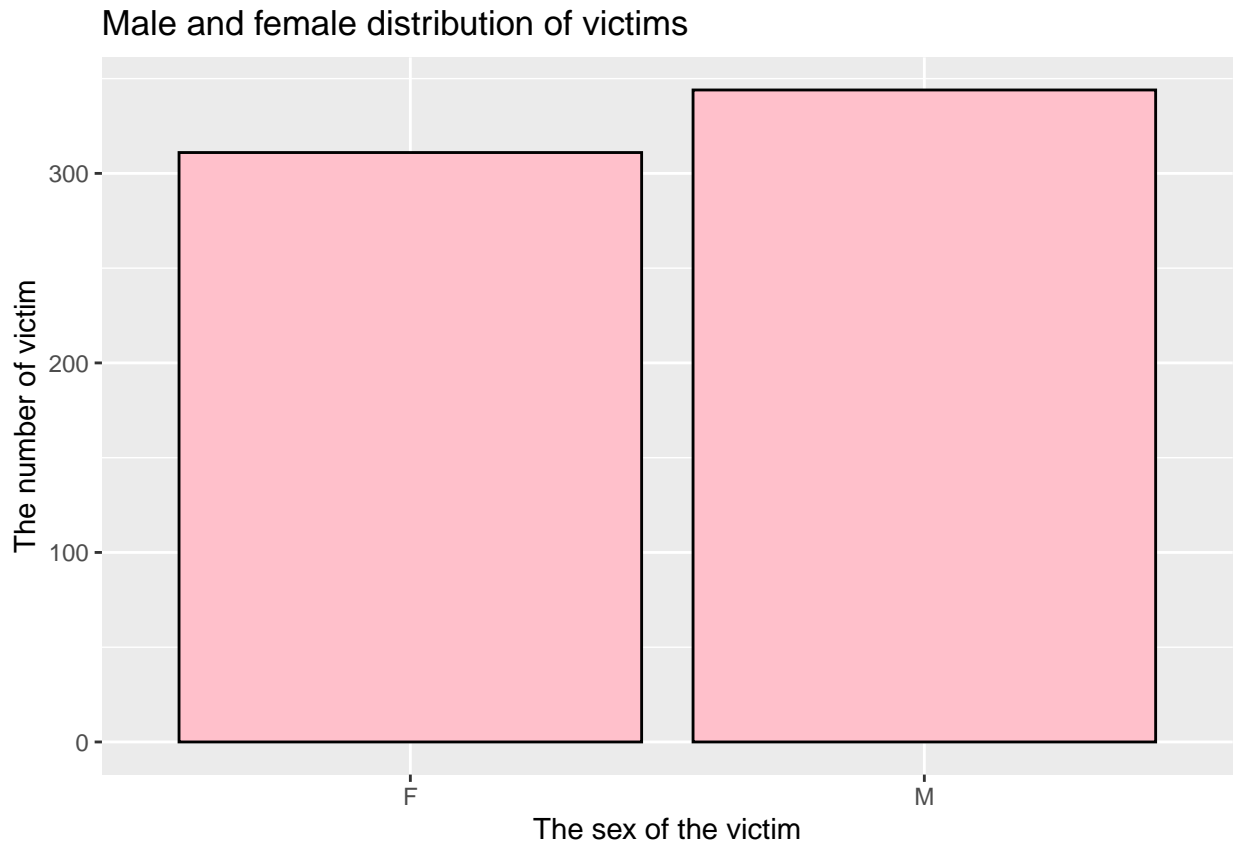
The data cleaning process

In the original dataset, there are 854 observations and 12 variables. Here are all variables:

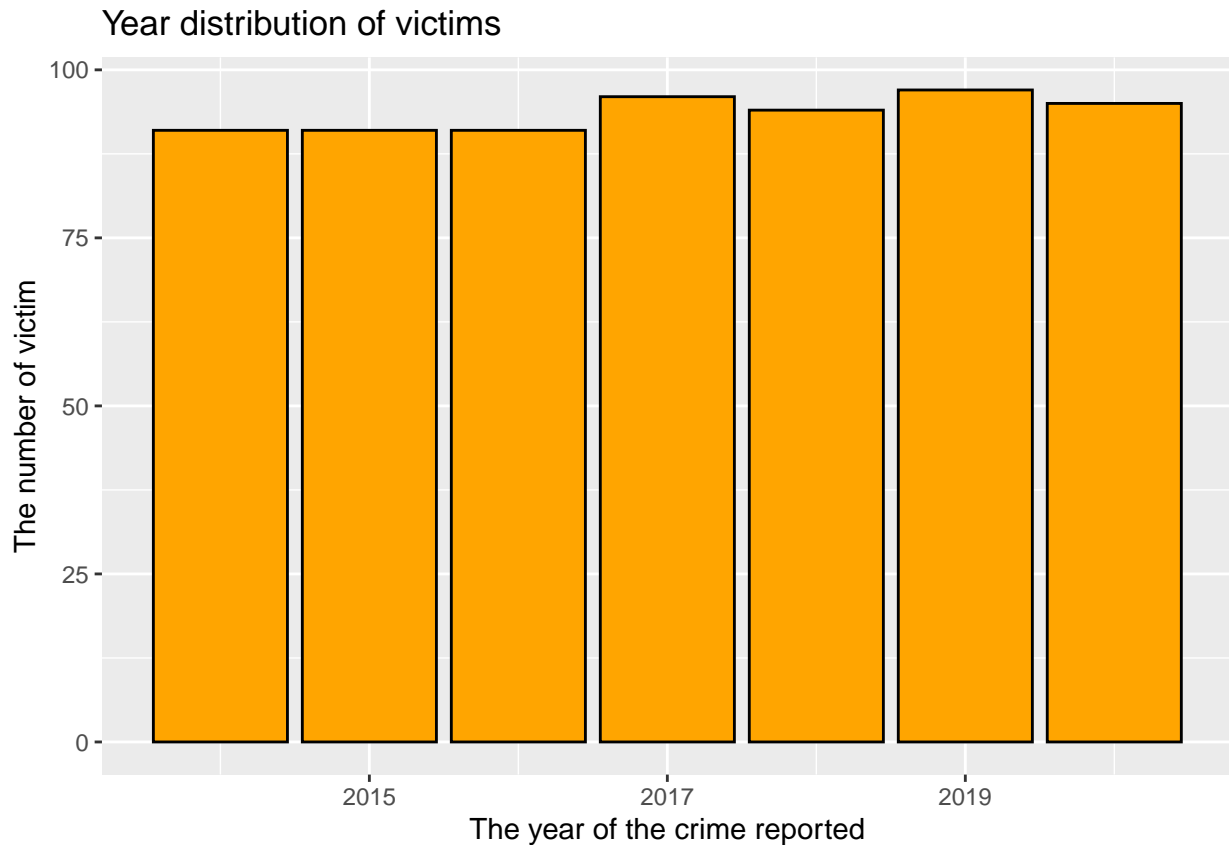| Variable Name	*Meaning*
Index__	Unique identifier
ReportedYear	Year crime was reported
Category	Crime category
Subtype	Crime category subtype
AssaultSubtype	Breakdown of assault subtypes
Sex	Sex of identified victim
AgeGroup	Age group of identified victim, adult or youth
AgeCohort	Age cohort of identified victim
Count__	Count of identified victims; duplicate persons are possible if the same person was victimized for the same offence during the same period
ObjectId	geometry

In this case, since not all of the variable is useful, we deleted some of the useless ones to keep our data brief. Below is the one we keep:

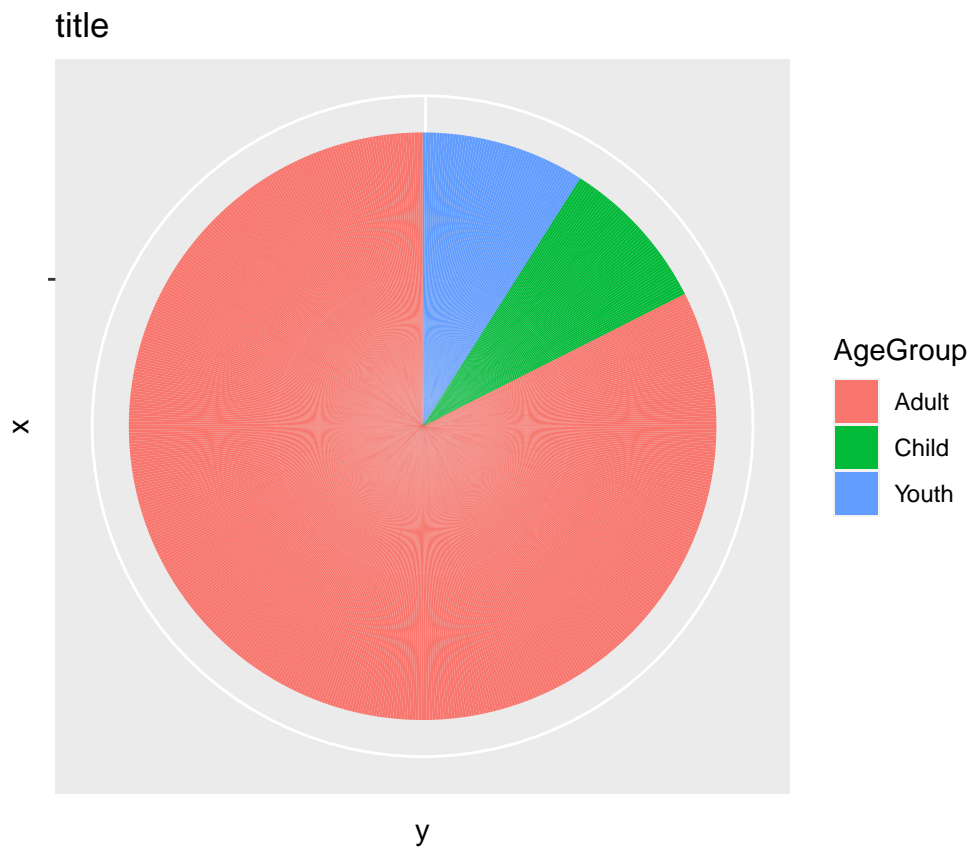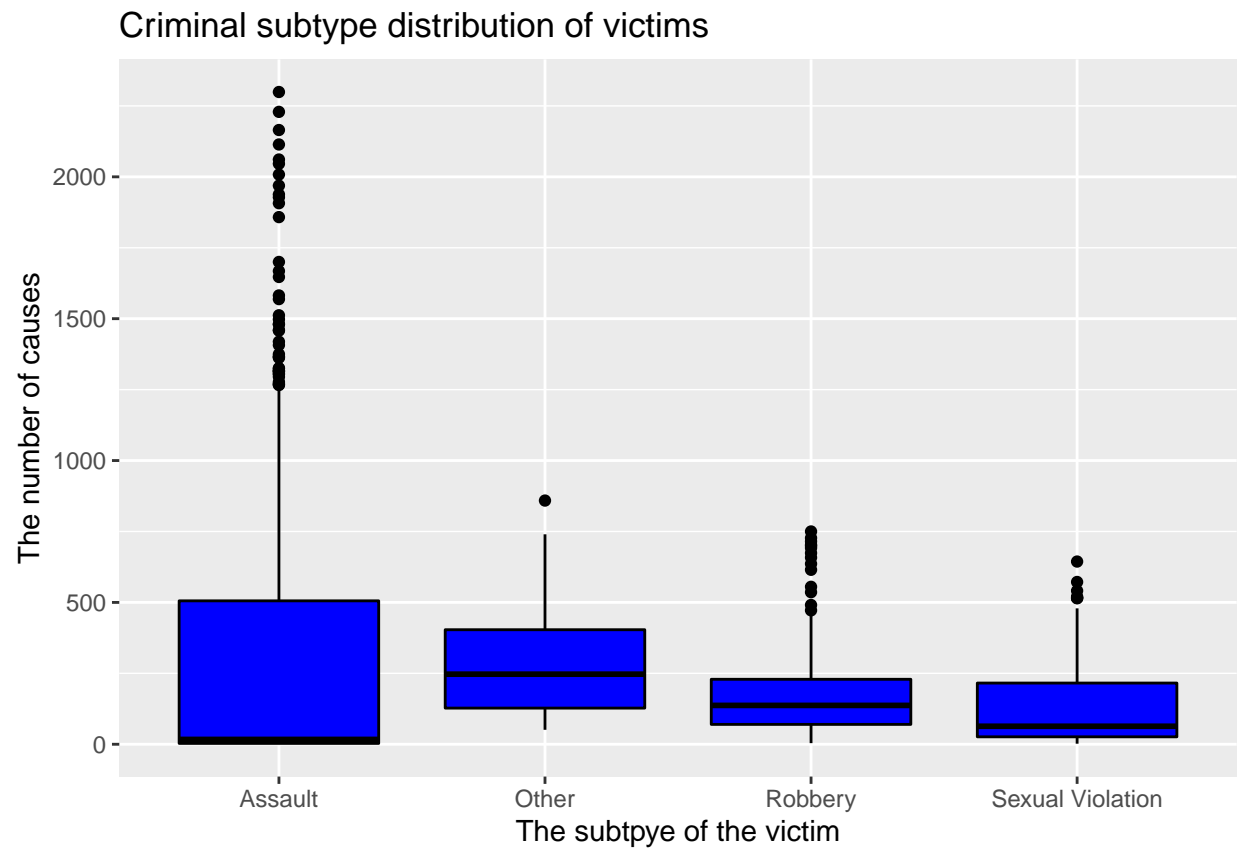| Variable Name	*Meaning*
Index__	Unique identifier
ReportedYear	Year crime was reported
Subtype	Crime category subtype
Sex	Sex of identified victim
AgeGroup	Age group of identified victim, adult or youth
AgeCohort	Age cohort of identified victim
Count__	Count of identified victims; duplicate persons are possible if the same person was victimized for the same offence during the same period

In the data cleaning process, we choose a new name for our data after clean, it is named "data_clean" instead of "data". After this, we need to delete these useless variables "geometry, ObjectId, Index__, AssaultSubtype, Category" to keep above variables stay. In this case, we are using the function "select" here. At the same time, we are focus on research the female and male in this case, so we are keeping these observation which are female or male, using the function "filter". Then, within this data set, age is divided into three stages, adult, child, and youth. Some victims do not provide this in order to protect their privacy. So we are using the function "filter" to delete these observations who don't provide their age.

Male and female distribution of victims

Here we using the bar plot. It visually shows the ratio of male to female victims. We can see that, in this set of data, female victim is around 320. The male victims are around 400. Male victims were 1.25 times more likely than female victims.

Year distribution of victims

We can see from the bar chart that there were about 85 reported cases in year 2015, year 2016 and year 2014. From year 2017, the number of victims began to increase, nearly reaching 90cases, but in 2018, it showed a slight downward trend. However, in 2019 and 2020, the number of victims continued to increase, and even nearly exceeded 100 cases in 2019. Generally speaking, the crime rate in Toronto has been on the rise in recent years. Although it occasionally decreases, the overall trend is still on the rise.

Criminal subtype distribution of victims

title

As can be seen from the pie chart, adults account for the largest proportion of all victims, accounting for nearly 80 percent of the total

Table 3: Age of the victim

min	Q1	median	Q3	max	IQR	mean	sd
0	18	35	55	65	37	34.10076	19.63876

victims. Teenagers make up about 10 percent, toddlers make up about 10 percent. Toddlers are aged 0-12, teenagers are aged 12-19, and adults are over 19. Although the proportion of teenagers and young children is not very heavy compared with adults, but for victims of crime, the proportion of young children is not small.

Reference

https://open.toronto.ca/dataset/police-annual-statistical-report-victims-of-crime/

https://www.numbeo.com/crime/in/Toronto

https://www.journaldev.com/43863/head-and-tail-function-r

https://statisticsglobe.com/r-extract-first-or-last-n-characters-from-string

Git

https://github.com/Sofiaaabbbccc/Sofia-.git