Fire incident in Toronto

Yilin Lan

04/06/2022

Abstract

In this report, I introduce the Toronto public safety background and the major concerns about public safety hazard. The dataset I used collects fire incidents data in Toronto. Then, I introduce the observations, variables of the dataset as well as how the data is collected. The pros and cons are also mentioned in the report. I give some possible improvement on future data collection. Then, I analyze the relationship between estimated dollar loss and number of exposure fires, civilian casualties, building status. Finally, I give my data analyzation result based on the previse work.

Keywords— Toronto, fire incidents

Introduction

With the development of technology and society, citizens and government concern more about the public safety. Among different types of public safety hazard, fire accident one of the most common thing in people's daily life. In January 19, 2022, in the city of Toronto, a fire broke out at a home on Gladstone Avenue, east of Dufferin Street at 7 p.m. A woman is dead and a man is in critical condition in this fire accident [@d]. Fire accidents happened almost every day, people get injury from the accidents often. They even cause death. Statistically, there are about 24,000 house fires each year in Canada, resulting in an average of 377 deaths and 3,048 injuries per year. Fire causes huge casualties and property loss. Since then, due to the global warming and technology improvement, people use more electrical goods during their life so that there is an increasing trend in number of fire incidents globally every year.

Base on the previse background of public safety. Government needs to collect the fire incidents information in order to improve the city's safe infrastructures as well as promote fire prevention. The dataset I used is provided by Ontario Fire Marshal (OFM) and share to the public on City of Toronto Open Data Portal. The OFM provides leadership and expertise on fire safety and it promote changes in laws, fire equipment in order to minimize the impact of fire and other public safety hazards on people, property and the environment in Ontario [@a]. According to Fire Protection and Prevention Act (FPPA), the Fire Marshal is responsible for the investigation of the cause, origin and circumstances of any fire, explosion, etc [@a]. Every year, the OFM investigates approximately 600 fire cases, including 80-100 death and some injury of the firefighters [@a]. These investigations help experts to determine why fire accidents happened and how to prevent them in the future. They also help identify trends in fire emergencies and ensure fire protection services are maintained for every Ontarian [@a].

Given the importance of fire incidents statistics in characterizing the most common cause of fire incident and making improvement about fire protection infrastructure, it is important to understand how the fire incident dataset is reported and how it may be interpreted. For this report, I will use fire incident dataset from City of Toronto Open Data Portal to analyze how the estimated dollar loss in a fire incident related to civilian casualties, number of exposure fires and different building status. Also, I summarize the major possible cause and the common origin of fire. Moreover, I will discuss the potential bias in how these data are collected and the impact on the final data analyzation. Some possible improvement when collect data in the future has also been given. The dataset will be processed and analyzed in R (R Core Team 2020) primarily using the tidyverse [@c] and dplyr [@f] packages. Figures and tables will be created with ggplot (Wickham 2016) and kableextra (Zhu 2020). The packages knitr (Xie 2021) are used to generate the R markdown report.

Data

Data Source

This dataset provides data about all the cases that sent to the Ontario Fire Marshal (OFM) relating to only fire incidents. The Dataset includes only Fire incidents as defined by the Ontario Fire Marshal and some incidents have been excluded pursuant to exemptions under Section 8 of Municipal Freedom of Information and Protection of Privacy Act (MFIPPA)[@e]. The dataset was published and updated on City of Toronto Open Data Portal annually by Fire Services. The fire incidents analyzed in this report was obtained in csv format from the City of Toronto Open Data Portal using the R package opendatatoronto. It was last updated on Jan 31, 2022.

Methodology and Data Collection

The dataset contains all the fire incidents cases that have been reported to the OFM from year 2011 to 2022. Fire incidents are reported to the OFM through all Ontario fire departments. They are responsible for recording data on every emergency response they make [@a]. Every fire incident is detailed reported, including the start and end time, location, estimated dollar loss, origin of fire, etc. Based on the background, it is always hard for experts to figure what happened in a fire incident. Crews usually need to be on scene for two to five days before conducting further analysis for weeks after [@b]. This situation has been reflected on the dataset. There are many missing values on number of exposure fires, Alarm Failure code, Alarm Type code and building status, etc. In fact, it takes a long time to investigate the origin of a fire incident. Manny Garcia, a training specialist with the OFM said, "Say the fire is on the interior of the building, then we commence our interior analysis so that we can then determine what's our room of origin and where that fire started and then once we've determined the room of origin, we try to determine the area of origin within that room and see what ignition sources were there." [@b]. Once the origin has been determined, people can start to investigate the building status and Alarm Type code. These are all hard to get an exact information because fire has already destroyed a lot of things.

The data is first collected by all Ontario fire departments and sent to the OFM. Then, the OFM is responsible for analyzes and transforms the collected data into valuable information [@a]. After a fire incident, the specialists survey the scene in order to get the information about the fire incident. Data has been collected through the process. One of the advantages of the dataset is that it contains lots of information about each fire incidents, all possible information is detailed recorded in the dataset. When people need to review a fire incident, the dataset can provide enough information such as the origin of the fire, building status, start and end time, etc. It is convenient for the OFM and experts to summary the best way to prevent the fire incident in order to minimize the impact of public safety hazards.

However, due to some privacy concern, some data have been excluded in the dataset and that will lead to the incompleteness of the dataset. The data analyze result will not be completely accurate. There exist some people that don't want to share the fire incident information to the public. This problem cannot be fixed because the law requires that local government institutions protect the privacy of an individual's personal information existing in government records. Moreover, there are over 90% of observations contain missing value. Due to the limited technology, it is hard to get complete and detailed information in the fire incident. Although it is hard to collect all the information after the fire incidents, the missing value will increase the uncertainty of the data analyze result. This problem can be ameliorated by improving the survey methods. Overall, there will be some potential bias occurred during the data analyzation. Some fire incidents have not been recorded and there are some missing values in the dataset. People will get smaller than actual number of fire incidents and the mean of total dollar loss will be lower than average. Moreover, the result of the most common origin of fire will not be accurate.

```
##
##    'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
```

```
##
##     group_rows
```

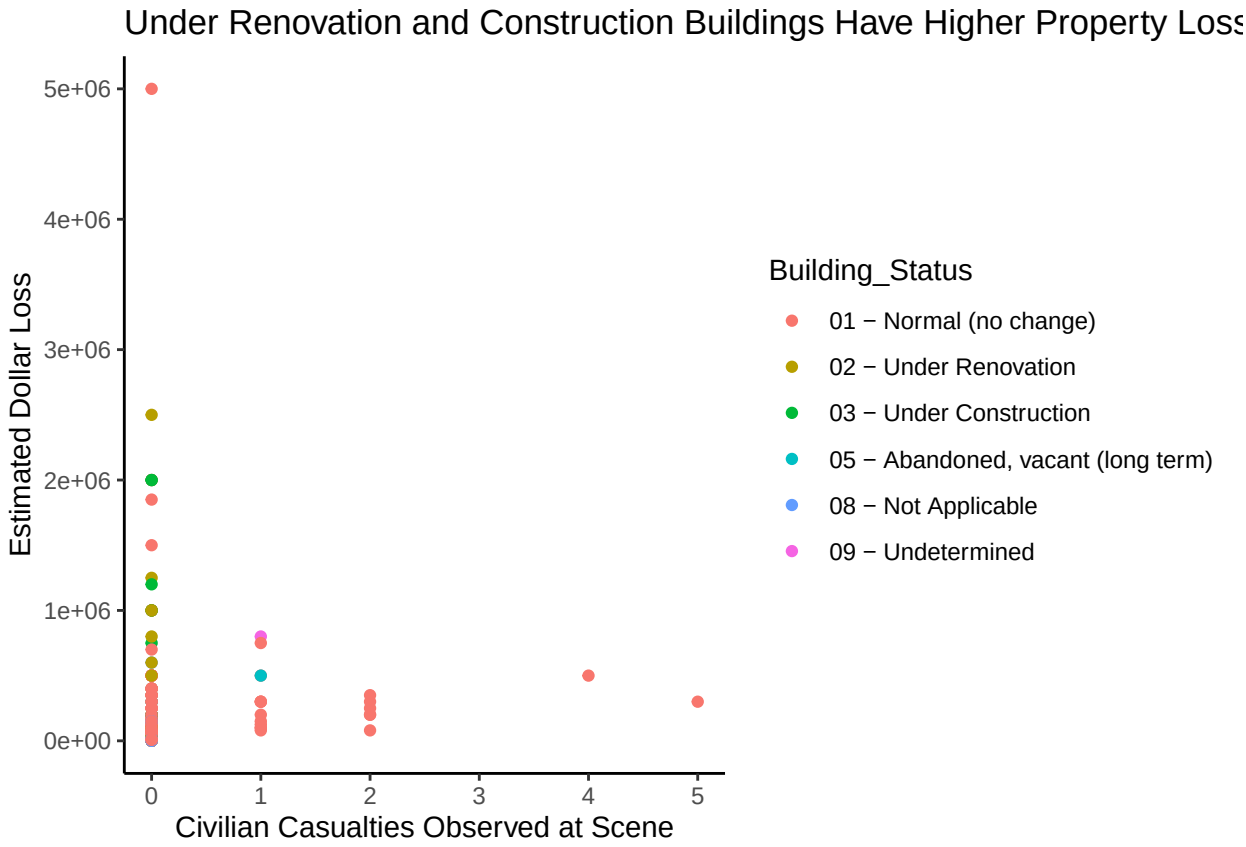
```
## # A tibble: 1 x 11
##   title          id       topics civic_issues publisher excerpt dataset_category
##   <chr>          <chr>    <chr>  <chr>        <chr>     <chr>   <chr>
## 1 Fire Incidents 64a2669~ Publi~ <NA>         Fire Ser~ "This ~ Table
## # ... with 4 more variables: num_resources <int>, formats <chr>,
## #   refresh_rate <chr>, last_refreshed <date>
```

```
## # A tibble: 17,536 x 43
##    `_id` Area_of_Origin         Building_Status Business_Impact Civilian_Casual~
##    <int> <chr>                  <chr>           <chr>                      <int>
##  1 52609 81 - Engine Area       <NA>            <NA>                           0
##  2 52610 75 - Trash, rubbish a~ <NA>            <NA>                           0
##  3 52611 <NA>                   <NA>            <NA>                           0
##  4 52612 75 - Trash, rubbish a~ 01 - Normal (n~ 1 - No busines~               0
##  5 52613 <NA>                   <NA>            <NA>                           0
##  6 52614 81 - Engine Area       <NA>            <NA>                           0
##  7 52615 22 - Sleeping Area or~ 01 - Normal (n~ 1 - No busines~               0
##  8 52616 55 - Mechanical/Elect~ 01 - Normal (n~ 2 - May resume~               0
##  9 52617 28 - Office            01 - Normal (n~ 1 - No busines~               0
## 10 52618 <NA>                   <NA>            <NA>                           0
## # ... with 17,526 more rows, and 38 more variables:
## #   Count_of_Persons_Rescued <int>, Estimated_Dollar_Loss <int>,
## #   Estimated_Number_Of_Persons_Displaced <chr>, Exposures <int>,
## #   Ext_agent_app_or_defer_time <chr>, Extent_Of_Fire <chr>,
## #   Final_Incident_Type <chr>, Fire_Alarm_System_Impact_on_Evacuation <chr>,
## #   Fire_Alarm_System_Operation <chr>, Fire_Alarm_System_Presence <chr>,
## #   Fire_Under_Control_Time <chr>, Ignition_Source <chr>, ...
```

Data Characteristic

The fire incident data set contains aggregate data of all fire incidents reported to the OFM between year of 2011 and 2022 [@e]. There are 17536 observations in the dataset and 43 attributes. Since the dataset includes comprehensive information about every fire incident, I will mainly focus on some important variables that closely related to fire incident and people's safety. The selected variable includes area of origin, building status, possible cause, civilian casualties, estimated Dollar loss and number of exposure fires. The variable types are either integer or character. As I mentioned in the previous part, the dataset contains lot of missing values and that will impact on the data analyzation process so that I remove all the missing data. The impact on the final result interpretation will be mentioned again below. A sample view of the dataset is displayed below.

```
data0= na.omit(data)
data0%>%ggplot(aes(y=Estimated_Dollar_Loss,x=Civilian_Casualties,color=Building_Status))+geom_point()+tl
```



According to the news about a fire incident, people care more about the property loss and civilian casualties observed at scene. Figure 1 illustrates the relationship between them.

I choose variables Estimated_Dollar_Loss, Civilian_Casualties to draw a scatterplot and categorize the plot using variable building status. From the graph, the estimated property loss spreads around 5000-500000 dollars and there is no obvious relationship between civilian casualties. The estimated property loss tends to be lower when there are more civilian casualties. Moreover, there are three outliers in the plot. One has extremely high estimated dollar loss and the other two have high civilian casualties. The three outliers will generally rise the average of both two variables. The dark yellow and green dots are located at the upper side of the scatterplot. That means when the building status is under renovation or under construction, there will be more property loss. Besides, when the property loss is high, there likely to be no civilian casualties.

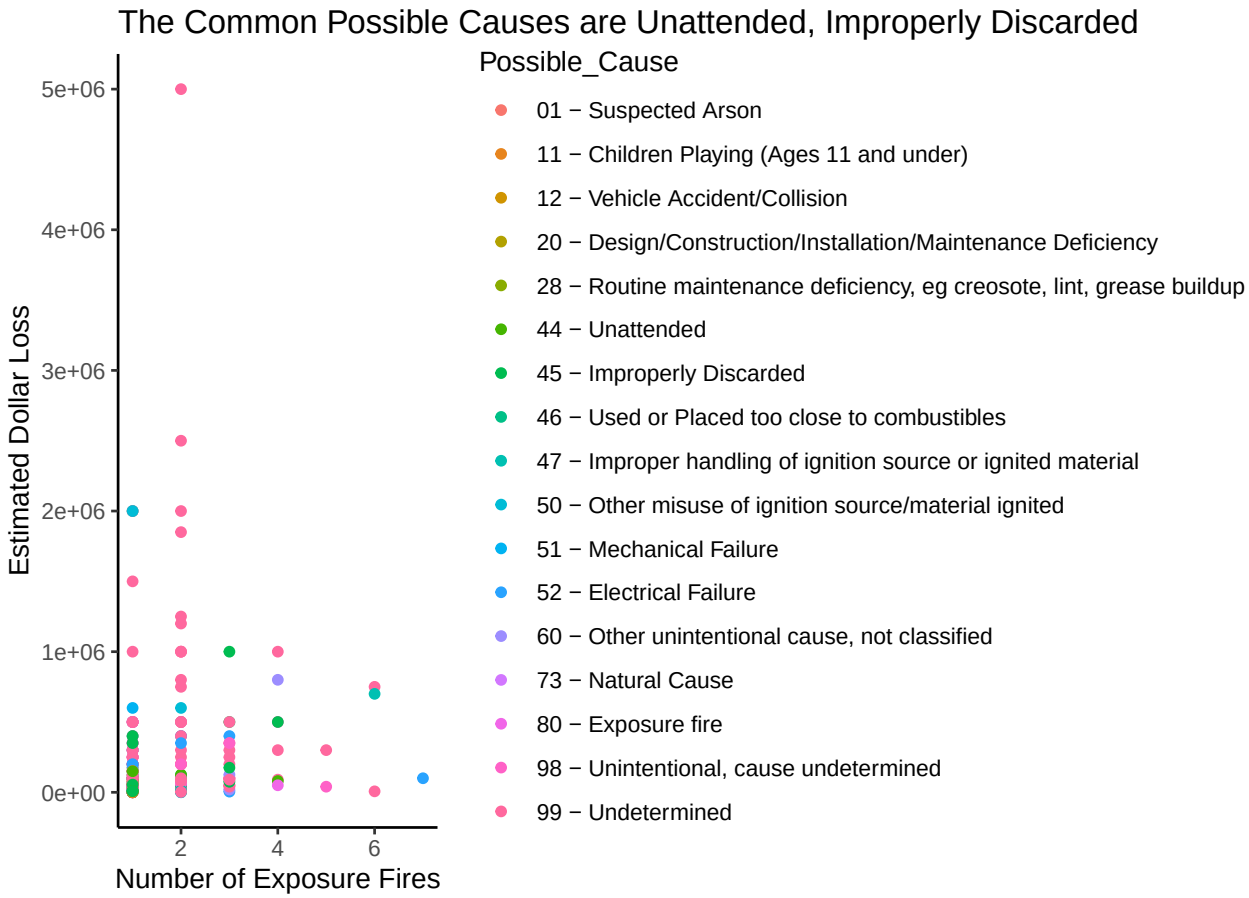


Figure 1: Estimated Dollar Loss by Number of Exposure Fires and Possible Cause

Model

Results

Discussion

First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

Second discussion point

Third discussion point

Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

Additional details

References