

My title\*

My subtitle if needed

Luxuan Zhu

30 January 2022

#### **Abstract**

First sentence. Second sentence. Third sentence. Fourth sentence.

**Introduction (talk about dataset, findings, and conclusion)**

**1st paragraph: motivation and broad**

**2nd paragraph: what I did and findings**

**3 paragraph: implications**

**precisely discuss the source of data and the bias it brings (ethical, statistical, etc.)**

Fire incidents occur everyday around the world. In most societies, these incidents are a source of threat to safety, capital, buildings, and societal development. They occur in different locations, with different causes, and result in damages to different extends. To further understand the causes of unintentional house fires, Turner and her peers conducted literature research and found that demographic factors such as age, gender, and socio-economic status can contribute to fire incidents. Moreover, social factors such as household composition, property characteristics, and fire safety practices can potentially cause house fires. An important observation from the study is that older and disabled residents are more likely to suffer an injury or fatality in the event of a fire, while they are less likely to ignite the fire in the first place. (Turner et al. 2016) Therefore, It is important for a society to look at the characteristics across fire incidents and further prevent future events, especially to protect those who are more in need for help and support.

An inevitable consideration while working with a dataset of real incidents is ethics. One needs to consider who are represented by the dataset and who are excluded. Under ethical considerations, what does it mean to make conclusions and thus influences based on the dataset? Looking at the source of the dataset, one can assume that the dataset does not record all fire incidents in Toronto. For instance, some incidents were excluded due to privacy concerns, while some may simply not be reported. Then, it is important to consider the groups of people influenced by the fire incidents. The dataset does not reflect the demographic of the persons involved, thus each incident may associate with any one person or group, with any demographic information, under any circumstance. Thus, for instance, when a statistician concludes that one type of fire incident is more significant than another type, the same affect may apply to the individuals associated behind the two types of incidents. Does that mean the group of individuals associated with the more severe incident type “deserve” more support? These questions are crucial to people such as statisticians and policy

---

\*Code and data are available at: [LINK](#).

setters who might rely on results derived from the dataset to make certain decisions that would, directly or indirectly, influence the society.

This paper provides an in-depth analysis on the fire incidents occurred in cooking areas or kitchens located in Toronto, ON. Specifically, the author is interested in exploring the differences in estimated money loss as a result of the fire incidents, categorized by whether the fire alarm system was operated. The author's hypothesis is that fire incidents where the fire alarm system did not operate or did not exist would result in higher estimated dollar loss compared to those where the fire alarm system operated. Such a hypothesis

The author found that...

The remainder of the paper will follow the structure of: Section @ref(data) explains the data; Section @ref(Model) showcases the models to support the analysis; Section @ref(Results) covers the results of the models; Section...

## Data (generate a table and talk about the variables)

The dataset is obtained from Open Data Toronto (Gelfand 2020) using R (R Core Team 2020). The dataset provides information on 17536 fire incidents to which Toronto Fire responds in more detail, as displayed by the 43 variables including Area of Origin, Extend of Fire, and Estimated Dollar Loss. The dataset includes only fire incidents defined by the Ontario Fire Marshal, which means it does not capture all fire incidents in Toronto. Since the fire incidents recorded are actual events related to citizens of Toronto and the society as a whole, personal information is not provided and the exact address have been approximated to the nearest intersections for privacy purposes. Moreover, the dataset follows exemptions under Section 8 of Municipal Freedom of Information and Protection of Privacy Act (MFIPPA) and excludes certain incidents.

For the purpose of the paper, the analysis will focus on the fire incidents occurred in cooking areas or kitchens. Moreover, the author is interested in exploring the relationship between the estimated dollar loss and the fire alarm system operation among the selected kitchen fire incidents. To do so, the author uses Tidyverse (Wickham et al. 2019) to select the variables of "Area of Origin," "Fire Alarm System Operation," "Latitude," "Longitude," and "Estimated Dollar Loss." The author also removes all entrants with missing values for any of the variables. The following shows an extract of the cleaned dataset (Table @ref(tab:dataextract)).

The first variable, Area of Origin, shows the initial cause of the fire incidents. In the context of this paper, all observations should contain the Area of Origin of "24 - Cooking Area or Kitchen." The second variable of Fire Alarm System Operation shows the status of the fire alarm system during the fire incident. The four statuses include: Fire alarm system operated; Fire alarm system did not operate; not applicable (no system); and Fire alarm system operation undetermined. The next two variables, longitude and latitude, are used in mapping the locations of the fire incidents. Lastly, the Estimated Dollar Loss showcases the financial consequences of each fire incident.

```
fire_incidents <-  
  read_csv("~/Session1/STA304_Paper_1/inputs/data/fire_incidents.csv")  
  
## Rows: 17536 Columns: 43  
  
## -- Column specification -----  
## Delimiter: ","  
## chr   (25): Area_of_Origin, Building_Status, Business_Impact, Extent_Of_Fire,...  
## dbl   (13): _id, Civilian_Casualties, Count_of_Persons_Rescued, Estimated_Dol...  
## dtm   (5): Ext_agent_app_or_defer_time, Fire_Under_Control_Time, Last_TFS_Un...  
  
##  
## i Use `spec()` to retrieve the full column specification for this data.  
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.  
  
cleaned_data <-  
  fire_incidents %>%
```

```

select(Area_of_Origin, Fire_Alarm_System_Operation, Extent_Of_Fire, Building_Status, Latitude, Longitude)
filter(Area_of_Origin == "24 - Cooking Area or Kitchen") %>%
filter(!is.na(Fire_Alarm_System_Operation)) %>%
filter(!is.na(Latitude)) %>%
filter(!is.na(Longitude)) %>%
filter(!is.na(Extent_Of_Fire)) %>%
filter(!is.na(Estimated_Dollar_Loss))

data_extract <-
  cleaned_data %>%
  slice(1:10) %>%
  kable(
    caption = "First ten rows of a dataset of fire incidents in Toronto",
    col.names = c("Area of Origin", "Fire Alarm System Operation", "Extent of Fire", "Building Status", "Latitude", "Longitude", "Estimated Dollar Loss"),
    digits = 1,
    booktabs = TRUE,
    linesep = ""
  )
data_extract

```

Table 1: First ten rows of a dataset of fire incidents in Toronto

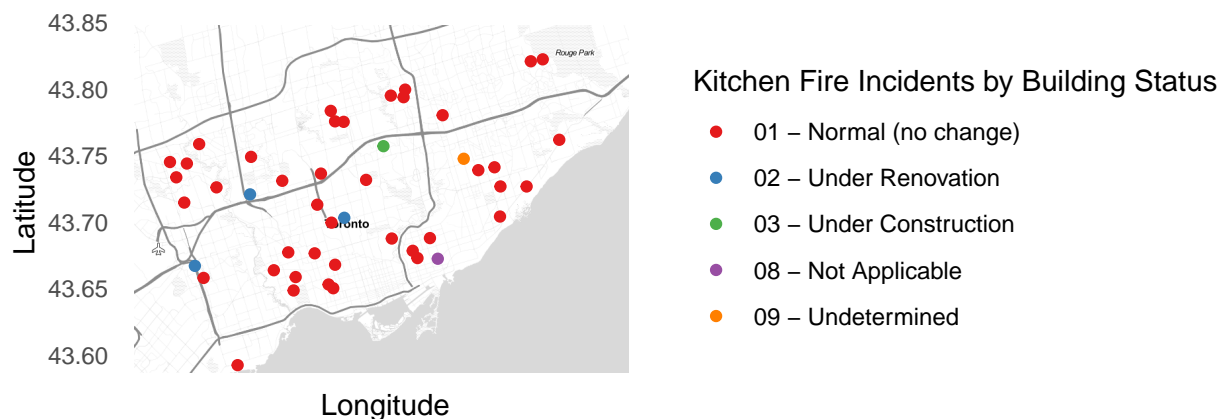
Area of Origin	Fire Alarm System Operation	Extent of Fire	Building Status	Latitude	Longitude	Estimated Dollar Loss
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	1 - Confined to object of origin	01 - Normal (no change)	43.7	-79.4	500
24 - Cooking Area or Kitchen	8 - Not applicable (no system)	1 - Confined to object of origin	01 - Normal (no change)	43.7	-79.4	0
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	1 - Confined to object of origin	01 - Normal (no change)	43.8	-79.4	0
24 - Cooking Area or Kitchen	8 - Not applicable (no system)	2 - Confined to part of room/area of origin	01 - Normal (no change)	43.7	-79.4	60000
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	1 - Confined to object of origin	01 - Normal (no change)	43.7	-79.6	3000
24 - Cooking Area or Kitchen	8 - Not applicable (no system)	1 - Confined to object of origin	01 - Normal (no change)	43.7	-79.4	200
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	2 - Confined to part of room/area of origin	03 - Under Construction	43.8	-79.3	15000
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	2 - Confined to part of room/area of origin	01 - Normal (no change)	43.8	-79.4	10000
24 - Cooking Area or Kitchen	1 - Fire alarm system operated	1 - Confined to object of origin	01 - Normal (no change)	43.7	-79.4	5000

Area of Origin	Fire Alarm System Operation	Extent of Fire	Building Status	Latitude	Longitude	Estimated Dollar Loss
24 - Cooking Area or Kitchen	8 - Not applicable (no system)	2 - Confined to part of room/area of origin	01 - Normal (no change)	43.7	-79.6	5000

## Model

## Map

To gain a visual representation of the distribution of fire incidents within the city of Toronto, the author uses ggmap (Kahle and Wickham 2013), mapproj (R by Ray Brownrigg, Minka, and Plan 9 codebase by Roger Bivand. 2022), and maps (Richard A. Becker, Ray Brownrigg. Enhancements by Thomas P Minka, and Deckmyn. 2021) to construct a map pinpointing the locations of selected data points. Due to the high number of observations, the author chooses to focus on fire incidents that either spread to other floors within the building, spread to the entire structure, or spread to the roof or the exterior of the building. All fire incidents are still initialized in the kitchen area. The justification for such a niche selection is to have the map display the fire incidents that extended beyond its area of origin and caused larger areas of damage. Then, the author plots the data points using colours sorted by the building status of each incident. The possible values are: normal, under rennovation, under construction, and unknown. The final resulting map is shown below.



## Scatterplot

```
final_graph1 <-
  cleaned_data %>%
  filter(Fire_Alarm_System_Operation == "1 - Fire alarm system operated") %>%
  ggplot(mapping = aes(x = Estimated_Dollar_Loss)) +
  geom_histogram(bins = 100) +
  xlim(0, 250000) +
  ylim(0, 250) +
  labs(x = "Estimated Dollar Loss",
       y = "Count")

final_graph2 <-
  cleaned_data %>%
```

```

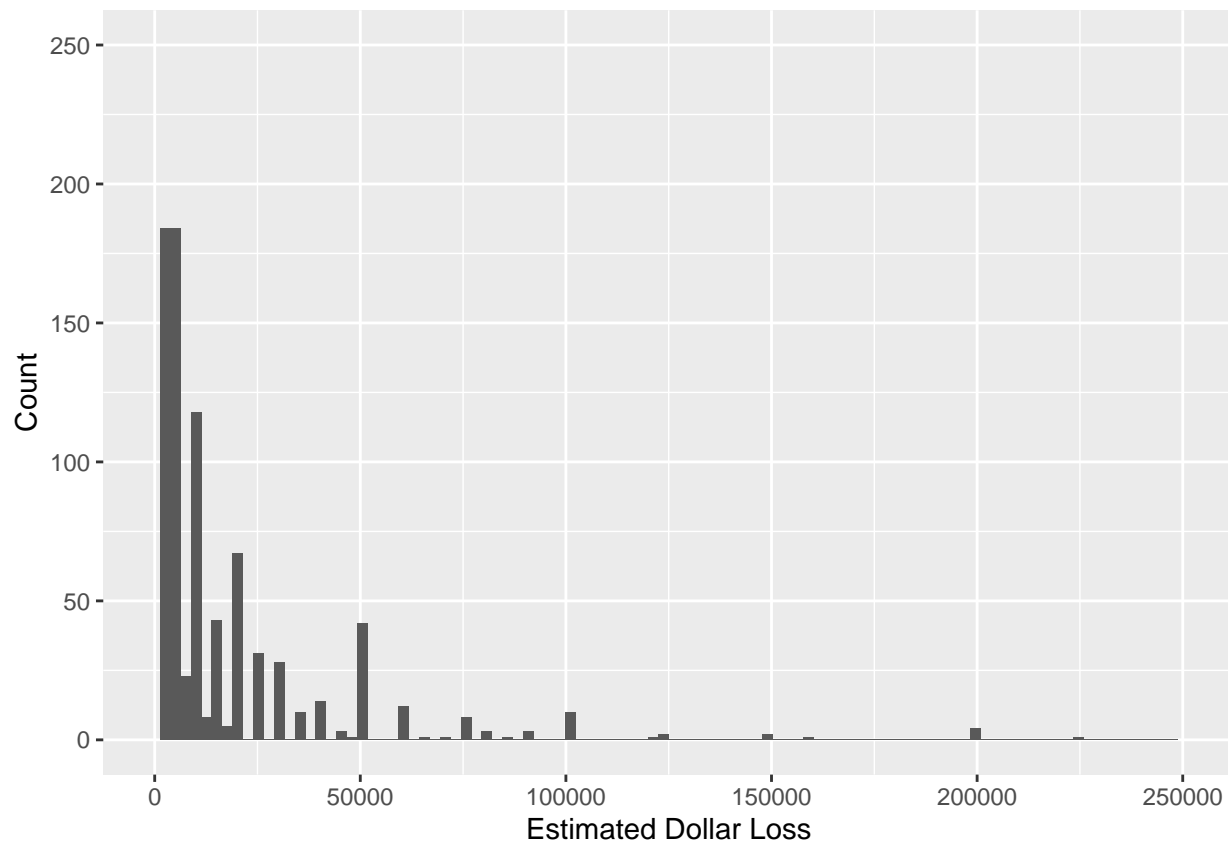
filter(Fire_Alarm_System_Operation != "1 - Fire alarm system operated") %>%
ggplot(mapping = aes(x = Estimated_Dollar_Loss)) +
geom_histogram(bins = 100) +
xlim(0, 250000) +
ylim(0, 250) +
labs(x = "Estimated Dollar Loss",
     y = "Count")

par(mfrow=c(1,2))
final_graph1

```

## Warning: Removed 4 rows containing non-finite values (stat\_bin).

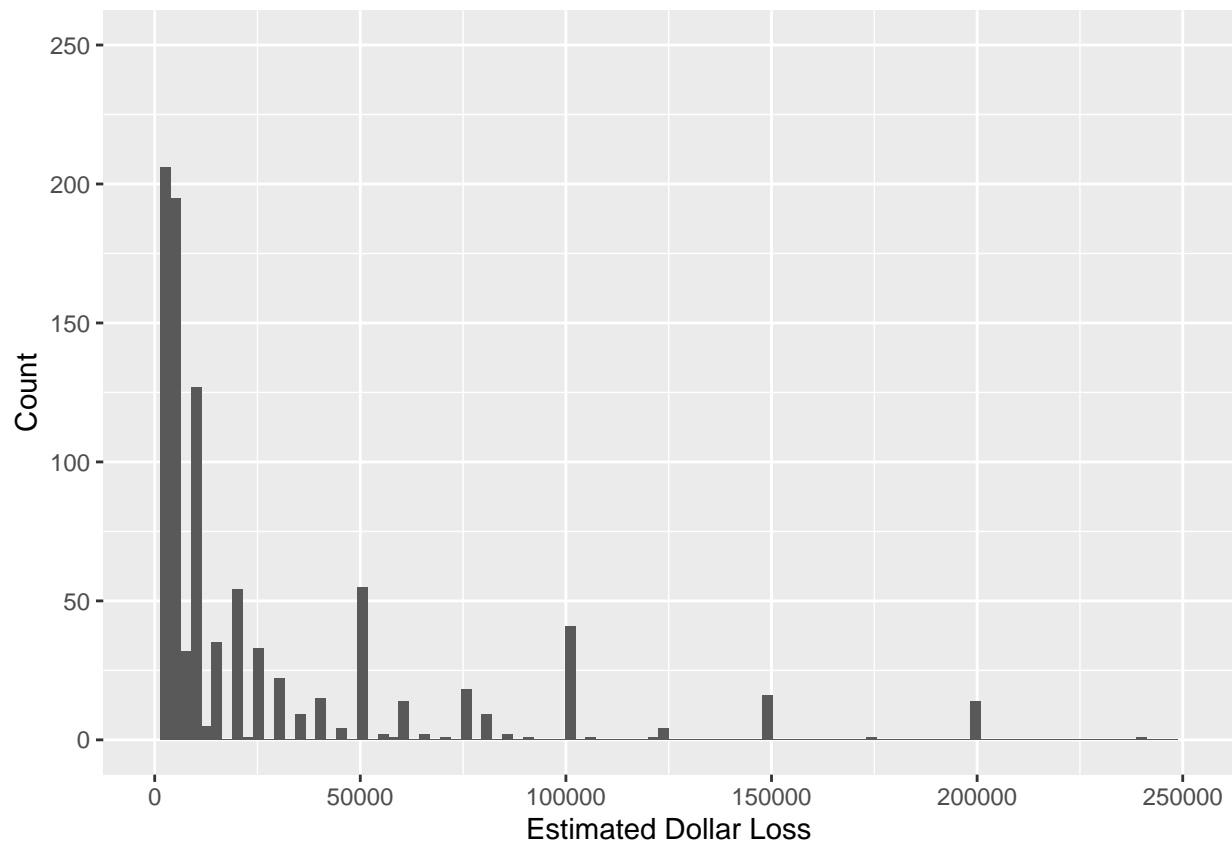
## Warning: Removed 2 rows containing missing values (geom\_bar).



```
final_graph2
```

## Warning: Removed 26 rows containing non-finite values (stat\_bin).

## Warning: Removed 2 rows containing missing values (geom\_bar).



## 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

- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Kahle, David, and Hadley Wickham. 2013. “Ggmap: Spatial Visualization with Ggplot2.” *The R Journal* 5 (1): 144–61. <https://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>.
- R by Ray Brownrigg, Doug McIlroy. Packaged for, Thomas P Minka, and transition to Plan 9 codebase by Roger Bivand. 2022. *Mapproj: Map Projections*. <https://CRAN.R-project.org/package=mapproj>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Richard A. Becker, Original S code by, Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka, and Alex Deckmyn. 2021. *Maps: Draw Geographical Maps*. <https://CRAN.R-project.org/package=maps>.
- Turner, Samantha L, Rhodri Johnson, Alison L Weightman, Sarah E Rodgers, Geri Arthur, Rowena Bailey, and Ronan A Lyons. 2016. “RISK FACTORS FOR UNINTENTIONAL HOUSE FIRE INCIDENTS, INJURIES AND DEATHS: A SYSTEMATIC REVIEW.” *Injury Prevention* 22: A301.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.