Fire Incident Analysis*

Exploring Toronto Fire Incident Data

Qisheng Yu

September 27, 2024

In this paper, we analyze fire incident data in Toronto to identify patterns and relationships between response times, estimated losses, and various incident characteristics. Using visualization techniques, we provide insights into key metrics such as incident type, response time, and resource allocation.

1 Introduction

Fire safety is a concern in urban environments, where the density of buildings and huge population increase risks of fire-related incidents. In Toronto, understanding the dynamics of fire incidents, response times, and the associated economic losses is important for policymakers and emergency responders to optimize their strategies. Although fire department interventions and fire alarm systems are in place to mitigate risks, there is a need for data-driven analyses that shed light on how these factors interact and impact overall fire safety outcomes.(This analysis follows path in the course material *Telling Stories with Data* by Rohan Alexander(Alexander 2023).)

In this paper, we analyze fire incidents in Toronto using data obtained from the City of Toronto (Toronto Open Data Initiative 2024). Our analysis was conducted using R (R Core Team 2024), a robust tool for statistical computing. This paper focuses on the analysis of fire incidents in Toronto, drawing from a dataset provided by the City of Toronto's Open Data initiative. The dataset contains detailed records of fire incidents, response times, estimated financial losses, and fire alarm system statuses. Previous studies have largely focused on the importance of timely responses and fire alarm systems in minimizing fire damage. However, there remains a gap in quantifying the direct relationship between response times, fire alarm presence, and the financial impact of fire incidents. Additionally, little attention has been given to spatial disparities in fire incidents across different areas of the city.

*Code and data are available at: LINK.

To address these gaps, we conducted a comprehensive analysis of fire incidents in Toronto, examining how response times and the presence of fire alarm systems influence the financial losses incurred from fires. We also explored the distribution of fire incidents across different station areas, seeking to identify regions with higher fire risks. Using statistical methods, we visualized the distributions and relationships between key variables such as response times, estimated financial losses, and fire alarm system efficacy.

Our findings suggest that, while shorter response times generally correlate with lower estimated losses, there are notable exceptions, especially in incidents where fire alarms were either absent or ineffective. Certain areas also exhibited higher frequencies of fire incidents, indicating possible regional risk factors that warrant further investigation.

2 Data

The dataset used in this analysis contains over 32,000 records from Toronto fire incidents. It includes various features such as response times, estimated losses, number of responding units, presence of fire alarms, and more. Each fire incident is grouped by incident type, and there are records for both civilian and firefighter casualties. We used the R (R Core Team 2024) to process the data and create visualizations using the package called ggplot2 (Wickham 2016).

2.1 Variables

Each row in the dataset represents an individual fire incident, and the following variables are of particular importance in the analysis:

Incident_ID: A unique identifier assigned to each fire incident.

Incident_Type: The classification of the fire incident, for a structural fire/vehicle fire/another type of incident.

Response_Time (minutes): The time taken from the moment the alarm was raised until the first fire response unit arrived on the scene. This is a key measure of the efficiency of emergency services.

Estimated_Loss ($): The estimated monetary loss resulting from the fire.

Civilian_Casualties: The number of civilian casualties associated with the fire.

Firefighter_Casualties: The number of firefighter casualties associated with the fire.

Persons_Rescued: The number of individuals rescued during the incident.

Fire_Alarm_Operation: A categorical variable indicating whether the fire alarm system functioned during the incident (e.g., “Yes”, “No”, “Unknown”).

Fire_Alarm_Presence: Indicates whether a fire alarm system was present in the building or area where the fire occurred.

Area_of_Origin: The location within the building where the fire started (e.g., “Kitchen”, “Living Room”, “Garage”).

Extent_of_Fire: A categorical variable representing the extent to which the fire spread (e.g., “Confined to object of origin”, “Spread beyond room of origin”).

Incident_Station_Area: The fire station area responsible for responding to the fire.

Incident_Ward: The Toronto ward where the incident occurred, providing geographic context.

Latitude and Longitude: The exact location of the incident.

Responding_Units: The number of firefighting units (e.g., trucks or apparatus) that responded to the incident.

Responding_Personnel: The number of firefighting personnel who responded to the incident.

2.2 Data Analysis

2.2.1 Distribution of Response Time

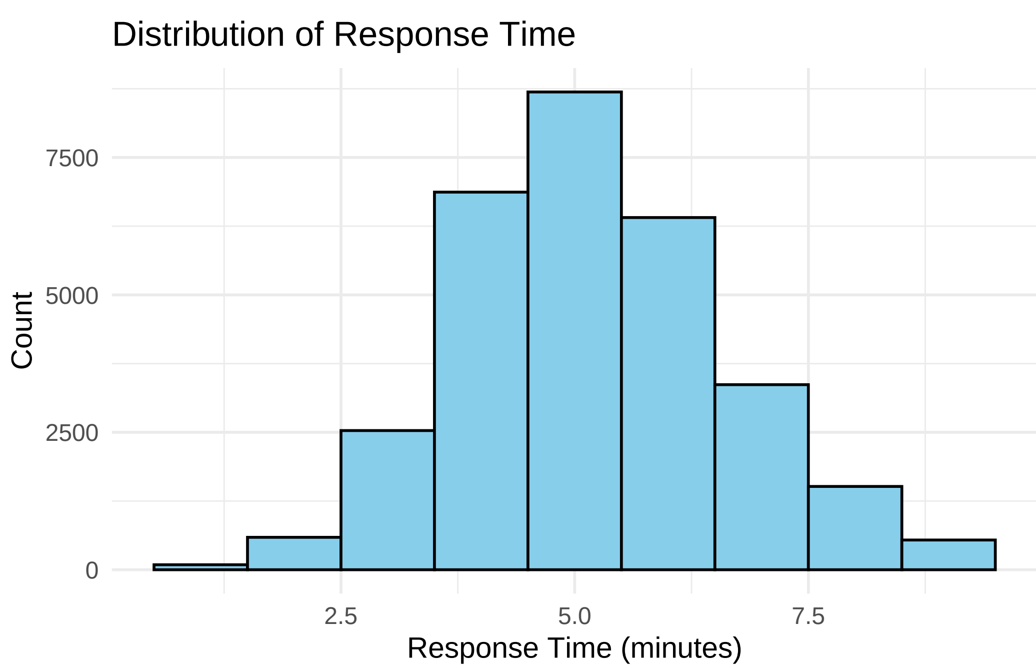

Figure 1: Distribution of Response Time

The distribution of response times for fire incidents is shown in Figure 1. The majority of incidents have response times between 4 and 6 minutes, with a few incidents having much longer response times.

2.2.2 Distribution of Estimated Loss

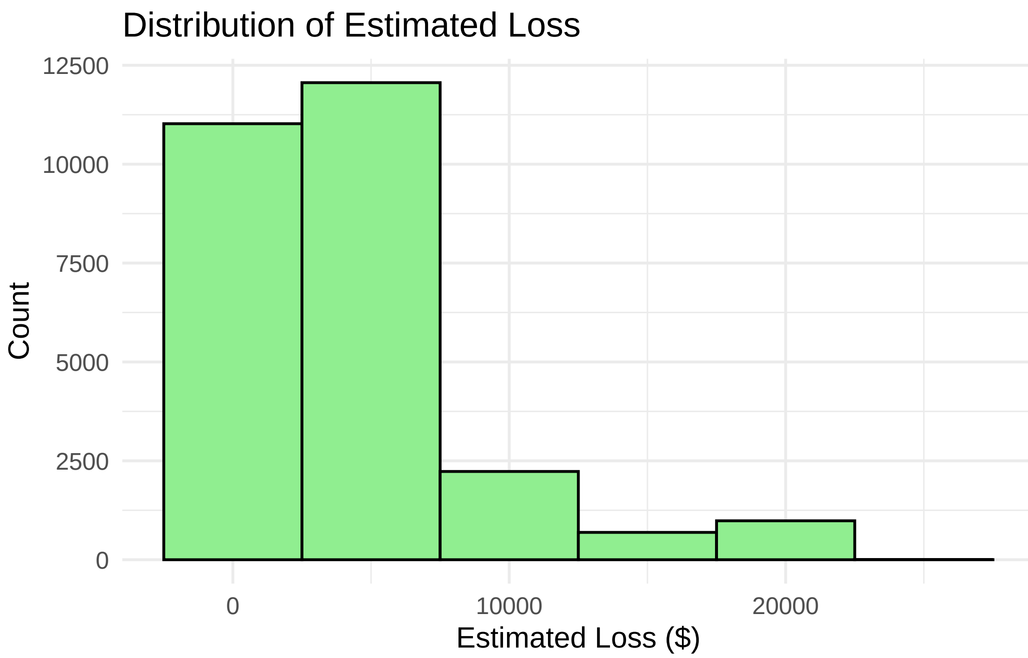

Figure 2: Distribution of Estimated Loss

The distribution of estimated loss for fire incidents is highly skewed, as shown in Figure 2. Most incidents report estimated losses below $5,000, though a small number of incidents have significantly higher losses, up to $50,000 or more.

2.3 Response Time by Incident Type

To examine how response times vary across different types of incidents, we plotted a boxplot of response times grouped by incident type in Figure 3. It shows that the median response time is relatively similar across most incident types, with a few outliers where response times are much longer.

Figure 3: Response Time by Incident Type

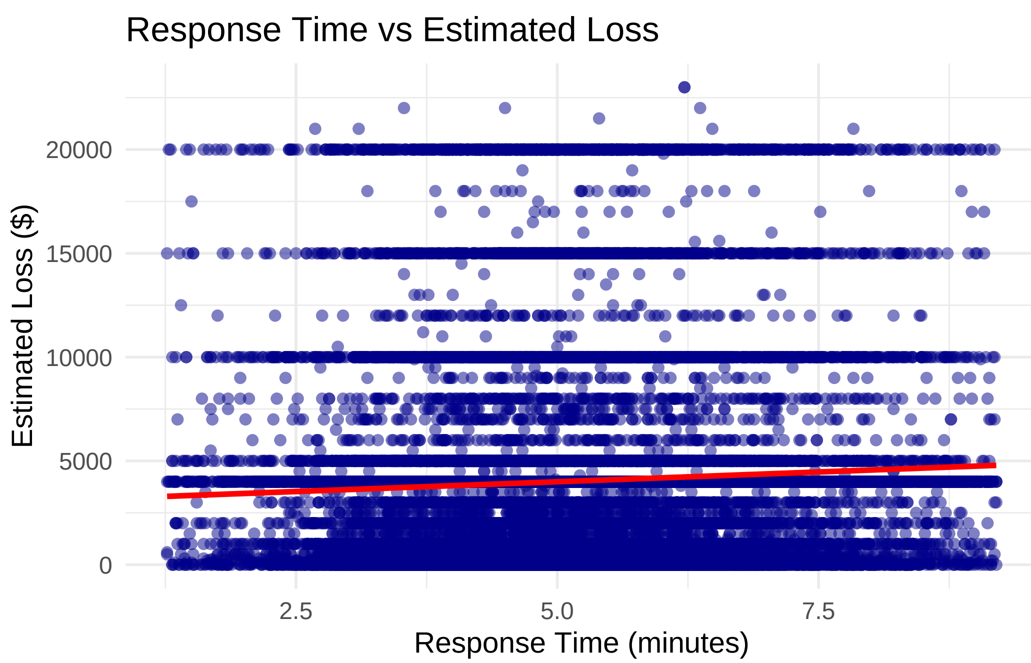

Figure 4: Response Time vs Estimated Loss

2.4 Response Time vs Estimated Loss

We also explored the relationship between response time and estimated loss. The scatter plot in Figure 4 suggests that there is a weak positive correlation between response time and estimated loss. However, the trend is not strong, and there are a large number of incidents with zero or low losses regardless of response time.

2.5 Number of Fire Incidents by Station Area

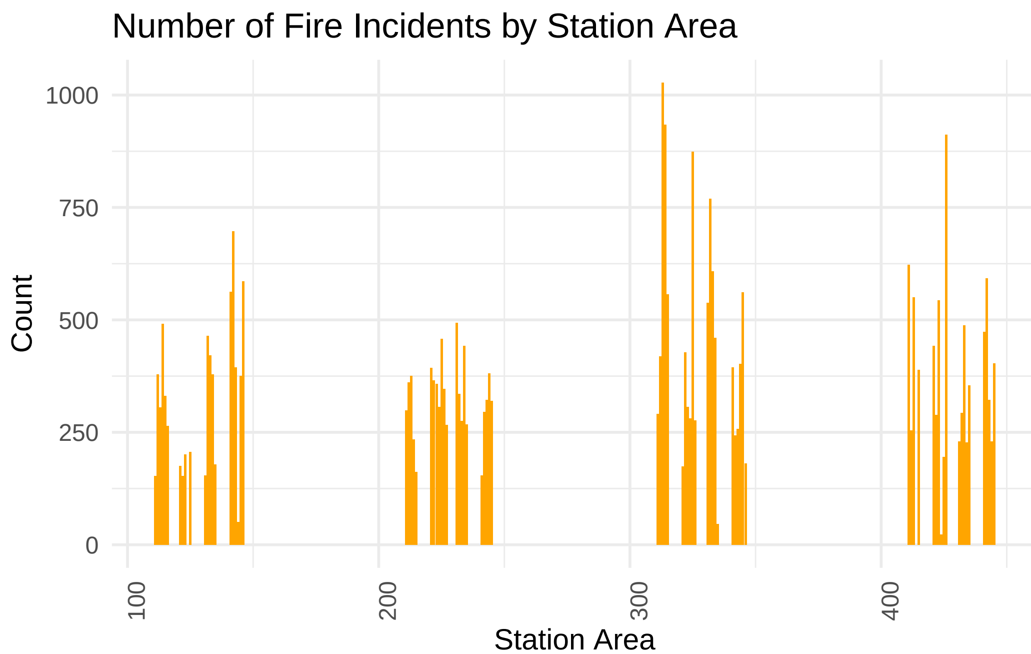

Figure 5: Number of Fire Incidents by Station Area

The number of fire incidents by station area is visualized in Figure 5. The bar plot shows the number of fire incidents by station area, with the station areas segmented into four distinct groups on the x-axis (around 100, 200, 300, and 400). Each group represents different geographic regions or station numbers in the dataset.

Station areas in the 300 range show the highest density of fire incidents, with several exceeding 750 incidents. This might indicate that these regions are more populated or prone to higher fire risks. Other station areas (around 100, 200, and 400) also see fire incidents, but their distribution is less dense compared to the 300 series. The clustering of incidents around certain station areas could be driven by multiple factors, such as population density, industrial areas, or geographical layouts prone to fire hazards. Further investigation is needed to determine whether these station areas correspond to specific urban or suburban regions.

2.6 Responding Units vs Estimated Loss

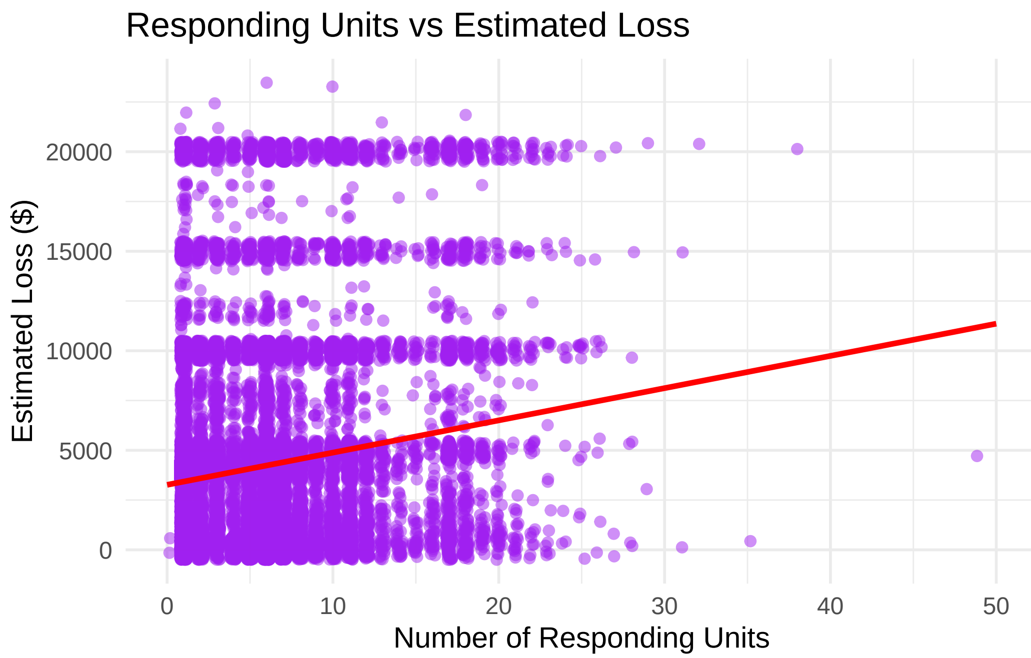

Figure 6: Responding Units vs Estimated Loss

The relationship between the number of responding units and estimated loss is shown in Figure 6. There appears to be a positive correlation between the number of units and the estimated loss, which may indicate that larger incidents with higher estimated losses require more resources.

3 Appendix

3.1 Data Cleaning Process

Handling Missing Values: Several variables, such as Response_Time and Estimated_Loss, contained missing or NA values. For numerical variables, missing values were imputed with the median value of the variable where appropriate. For categorical variables, missing data were handled by categorizing them as "Unknown" or "Not Available".

Outlier Detection and Treatment: In variables such as Response_Time and Estimated_Loss, extreme values were identified as potential outliers. For example, some incidents reported extremely high response times or losses that could distort the analysis. These outliers were treated by limiting the scale of visualizations to focus on the central tendency of the data, though no data points were removed during the cleaning process.

Data Transformation: The Response_Time variable was converted into numeric format where necessary to allow for proper visualization and statistical analysis. Similarly, Incident_Ward and Incident_Station_Area were transformed into factors to support categorical analysis.

The following sections explore different aspects of this dataset using various visualization techniques.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. https://tellingstorieswithdata.com/.

R Core Team. 2024. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Open Data Initiative, City of. 2024. "Toronto Fire Incidents Data." https://open.toronto.ca/dataset/fire-incidents/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org/.