# The Multifaceted Impact of Fire Damage: Beyond the Flames

Zekai Li

2024-01-19

**Abstract**

This study examines the relationship between financial losses from fire incidents and the number of responding personnel in Toronto, using data from opendatatoronto and advanced visualization with ggplot2. Our analysis reveals a pronounced right-skewness in both financial losses and personnel count, with significant outliers. By applying a logarithmic transformation to both variables, we transform the nonlinear relationship into an almost linear one, indicating a proportional increase in financial loss with the number of personnel. Code and data supporting this analysis is available at this link https://github.com/ZEKAILI032/STA302/tree/main

## Introduction

Fires unleash a spectrum of complex and intricate damages, often wreaking havoc on homes and businesses in ways that are not immediately visible. Beyond the obvious destruction of personal belongings and structural components, the intense heat from fires is capable of inducing a range of distortions, including swelling and buckling. Furthermore, smoke from fire damage insidiously infiltrates furniture, upholstery, drapes, wooden structural elements, and other porous materials. If left untreated, the pungent odor of smoke can stubbornly persist for years, embedding itself into the very fabric of the affected space.

In this project, `R` (R Core Team 2021) is used as our primary programming language. Our aim is to explore the correlation between the estimated financial impact of fire incidents and the number of responding personnel in Toronto. The dataset 'Fire Incidents', sourced from `opendatatoronto`, serves as the foundation for our analysis (Gelfand 2020). We have processed the data using the dplyr package (Wickham et al. 2020), and the presentation of tables is refined using the `kable` function from the knitr package (Xie 2014). All resources are in github[1]

[1] https://github.com/ZEKAILI032/STA302/tree/main

# Data

Given the presence of numerous missing values in the dataset, which could potentially skew the findings, we implemented a simulation approach. All instances with missing data were meticulously excluded to ensure the integrity of the analysis. While acknowledging that this approach may influence the outcomes, it was deemed essential at this stage of the research to proceed with the most accurate and reliable data available.

Table 1: Statistical Summary

| loss | personnel |
|---|---|
| Min. : 0 | Min. : 0.0 |
| 1st Qu.: 500 | 1st Qu.: 13.0 |
| Median : 3000 | Median : 21.0 |
| Mean : 37494 | Mean : 25.6 |
| 3rd Qu.: 17000 | 3rd Qu.: 31.0 |
| Max. :50000000 | Max. :1275.0 |

Table 1 presents a comprehensive statistical summary of fire-related financial losses and the number of personnel involved. The data reveals that the median fire loss is $3,000, while the average loss significantly escalates to $37,494. Notably, the losses extend into the upper echelons, with the third quartile at $17,000 and an extreme maximum loss reaching a staggering $50 million. On the personnel front, the median number stands at 21, with an average slightly higher at 25.6. The scale of personnel involvement can vary widely, as evidenced by the maximum number recorded at 1,275. The distribution of both variables exhibits a pronounced right skewness, indicating a concentration of lower values with sporadic but substantial high-value outliers.

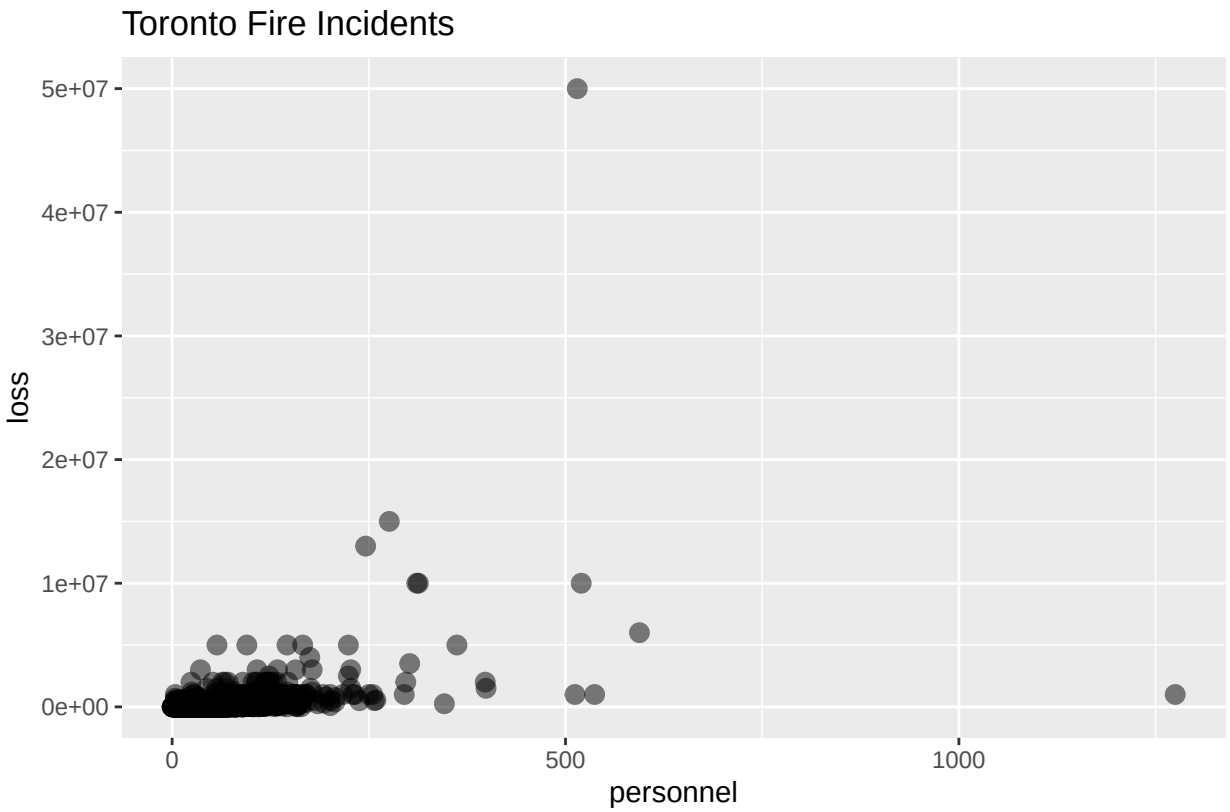


Figure 1

Figure 1 displays a scatter plot illustrating the relationship between the number of responding personnel and the estimated dollar loss, generated using the `ggplot2` package (Wickham 2011). The plot vividly highlights outliers in both the loss and personnel data, corroborating the observations made from the statistical summary in Table 1. The raw data suggest a nonlinear relationship between the number of personnel and the dollar loss. To address this and potentially reveal a clearer pattern, applying a logarithmic transformation to both the loss and personnel data could be beneficial, thereby normalizing the scale and stabilizing the variance of these variables.

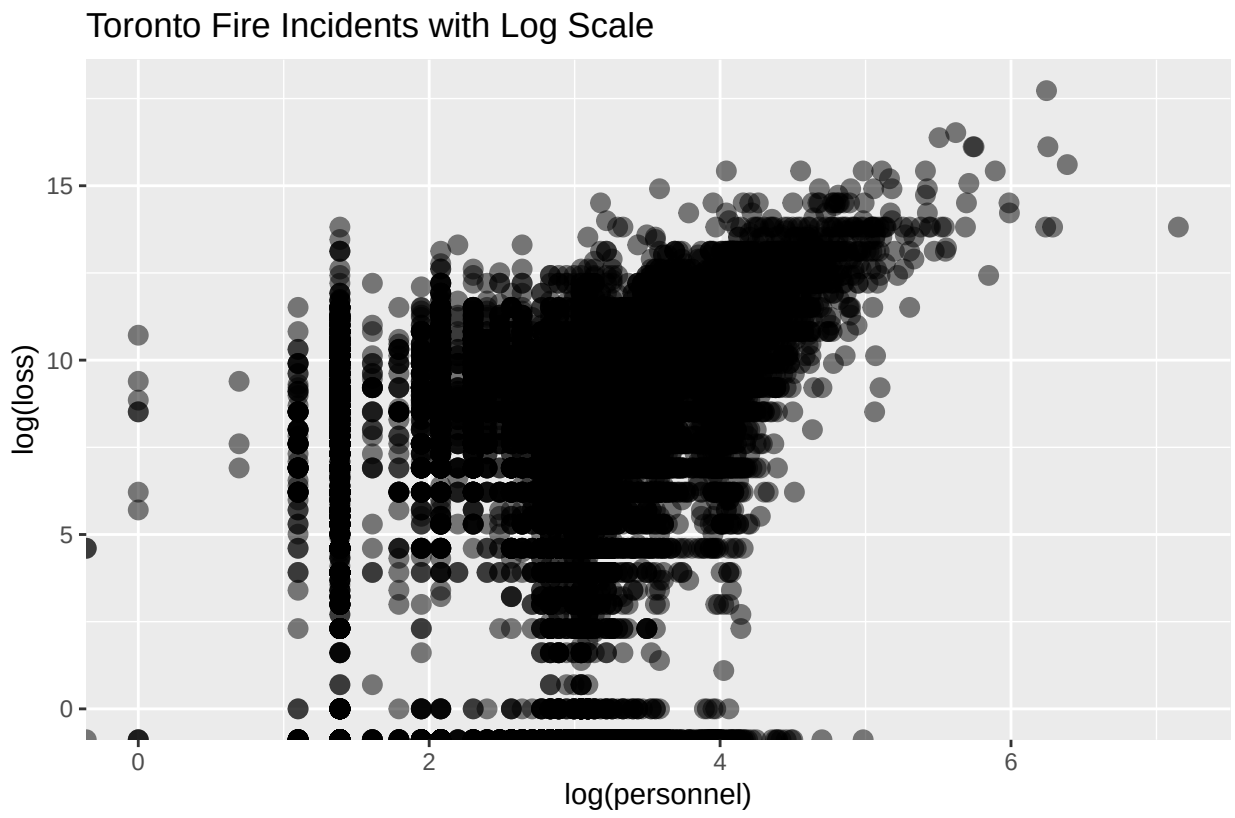


Figure 2

Figure 2 depicts the relationship between personnel and loss due to fire, both plotted on a logarithmic scale. Following the log transformation, the relationship between these two variables becomes markedly linear. The graph demonstrates a consistent trend: as the log-transformed value of personnel `log(personnel)` increases, there is a corresponding increase in the log-transformed value of the estimated loss `log(loss)`. This linear relationship highlights a proportional linkage between the number of personnel involved and the financial impact of the fire.

# Reference


Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2011. "Ggplot2." *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (2): 180–85.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.