Weather Impact on Fire Incidents in Toronto

Shijun Yu

GitHub Link: https://github.com/TonyYu123456/JSC370-Project

1 Introduction

1.1 Background

Fire incidents pose significant risks to public safety and cause substantial economic damage. Understanding the factors contributing to the occurrence and severity of these incidents is crucial for effective prevention and management. This study examines fire incidents within the city of Toronto, using publicly available data from the City of Toronto's open data portal. This dataset includes essential information for each reported fire incident, such as the exact date and time the alarm was raised, geographic location (latitude and longitude), civilian casualties, and estimated financial loss resulting from the fire.

Existing research indicates that environmental conditions, particularly weather, can influence the likelihood and severity of fire incidents. To further explore this potential relationship, weather data was obtained via the Open-Meteo Historical Weather API. This API provides detailed historical weather observations on an hourly basis, including important meteorological variables such as temperature (in °C), wind speed (in meters per second), and precipitation (in millimeters).

Due to API call restrictions under Open-Meteo's free plan (maximum of 5000 calls per hour and 10000 calls per day), data extraction was carried out in batches. Two batches of 5000 fire incidents each were processed separately and subsequently merged together. For each fire incident, weather conditions from the specific hour of the incident (e.g., if an incident occurred at 5:10 AM, then weather data from 5:00 AM was matched) were extracted and merged into a comprehensive dataset. The final dataset thus includes aligned hourly weather conditions corresponding precisely to each incident's reported alarm hour, enabling the accuracy and consistency of the analysis.

1.2 Research Question

With the motivation to explore external influences on fire incidents, several important considerations were made before clearly defining the study population and research question:

First, this study focuses specifically on incidents that have valid geographic coordinates and recorded alarm times, as accurate location and timing information are crucial for precise matching with hourly weather data.

Second, the severity of fire incidents can be measured using various metrics, such as civilian casualties and estimated dollar loss. Therefore, these two variables were selected as indicators of fire severity in this analysis.

Finally, after a preliminary exploration of the dataset, it was recognized that some of the variables seem to have a skewed distribution. Thus, appropriate transformations may be required to handle these distributions effectively during the analysis.

Considering these factors, the formulated research question for this study is:

"Do weather conditions (specifically precipitation, temperature, and wind speed) significantly influence the frequency and severity of fire incidents in Toronto?"

2 Methods

2.1 Data Acquisition

Two primary datasets were used in this analysis:

- Fire incident dataset was downloaded from the City of Toronto's Open Data Portal (https://open.toronto.ca/dataset/fire-incidents/). This dataset contains records for fire incidents within Toronto, providing comprehensive details such as the exact date and time of incident, geographical coordinates, the number of civilian casualties, and the estimated dollar loss incurred from each incident.
- Weather dataset was obtained using the Open-Meteo Historical Weather API (https://open-meteo.com/en/docs/historical-weather-api). The Open-Meteo API provides detailed meteorological information including temperature (°C), wind speed (m/s), and precipitation (mm) at an hourly granularity. For each incident, the exact hourly weather conditions matching the incident's alarm hour were fetched using latitude, longitude, and incident date as query parameters.

The datasets were initially stored as separate tibbles and later merged into a single comprehensive dataset. This final merged dataset provides a robust foundation for analyzing the relationship between weather conditions and fire incidents.

2.2 Data Cleaning and Wrangling

Following data acquisition, the merged dataset was first inspected to verify variable structure and data types. This initial assessment confirmed the presence of key variables including incident alarm time (TFS_Alarm_Time), date of the incident (Incident_Date), geographic coordinates (Latitude and Longitude), civilian casualties (Civilian_Casualties), estimated dollar loss (Estimated_Dollar_Loss), and hourly weather conditions (Temperature, Wind_Speed, and Precipitation).

The dataset was then cleaned for analysis. Several data type conversions and renaming procedures were conducted for consistency and clarity. Missing values were checked during the inspection, with only one observation containing NA. This incomplete record was removed from the dataset, resulting in a new dataset ready for further transformations and analysis.

Exploratory checks of variable distributions were conducted next, using histograms to assess potential skewness. Temperature was approximately normally distributed and required no transformation. However, wind speed exhibited significant right-skewness, prompting a log transformation (Wind_Speed_Log) to achieve a more symmetric distribution suitable for subsequent analysis. Due to the highly skewed nature of precipitation data (predominantly composed of zero values), this variable was converted into a binary categorical variable (Precipitation_Binary), indicating the presence or absence of rainfall at the time of the incident. Additionally, given the highly skewed and zero-inflated distribution of civilian casualties, this variable was also converted into a binary categorical variable (Civilian_Casualties_Binary), representing whether an incident resulted in civilian casualties or not. Similarly, estimated dollar loss exhibited significant right-skewness, driven by a small number of incidents with extremely high losses. To address this, a log transformation (Estimated_Dollar_Loss_Log) was applied, reducing skewness, improving interpretability, and maintaining valuable information regarding incident severity.

Outlier detection and removal were carried out for Temperature and Wind_Speed_Log using the 1.5*IQR method. These extreme values were safely removed. Conversely, outliers in the Estimated_Dollar_Loss_Log were intentionally retained, recognizing that extreme values in this variable reflect genuinely severe fire incidents critical to the study's objectives.

After completing these data cleaning and wrangling procedures, the final dataset comprised 9931 rows and 14 columns, each with properly transformed, cleaned, and aligned variables. This thoroughly prepared dataset provided a robust foundation for later analysis.

2.3 Data Exploration Tools

Data exploration was conducted using the cleaned dataset to uncover patterns, assess distributions, and identify relationships relevant to the research question.

Histograms were used to visualize the distribution of continuous predictor variables, including temperature and log-transformed wind speed. A bar chart was used to examine the frequency of incidents according to precipitation condition ("Rain" vs "No Rain").

To examine relationships between the transformed continuous predictor variables and the response variables, scatterplots with fitted linear regression lines were generated to explore the associations between Temperature, Wind Speed Log, and log-transformed Estimated Dollar Loss. Additionally, a boxplot was used to visualize differences in log-transformed Estimated Dollar Loss across rainy and non-rainy conditions, highlighting potential variations in fire severity based on precipitation.

Further, to assess how weather conditions may influence the occurrence of civilian casualties, boxplots were employed to compare the distributions of Temperature and Wind Speed Log between incidents with and without casualties. Additionally, a stacked bar plot was used to examine the proportion of fire incidents that resulted in civilian casualties across rainy and non-rainy conditions.

To quantify linear associations between numeric variables, a correlation heatmap was conducted. The heatmap provided initial insights into the strength and direction (positive or negative) of linear relationships among these numeric variables.

This exploratory approach provided valuable insights into potential weather-related influences on fire incidents, guiding the subsequent modeling and hypothesis testing.

3 Preliminary Results

3.1 Summary Statistics

Summary statistics are presented in Table 1 and Table 2 below. These statistics reflect the cleaned and transformed dataset after outlier removal and transformation steps described in the Methods section.

Table 1: Table 1: Summary Statistics of Numeric Variables

Temperature	Wind_Speed_Log	Estimated_Dollar_Loss_Log
Min. :-24.600	Min. :0.4055	Min. : 0.000
1st Qu.: 0.700	1st Qu.:1.2326	1st Qu.: 5.525
Median : 8.900	Median :1.5217	Median : 7.824
Mean : 8.748	Mean :1.5074	Mean : 7.029
3rd Qu.: 17.900	3rd Qu.:1.7951	3rd Qu.: 9.210
Max. : 34.800	Max. :2.6115	Max. :16.380

Table 2: Table 2: Summary Statistics of Binary Variables

Casualties_Proportion	No_Casualties_Proportion	Rain_Proportion	No_Rain_Proportion
0.0600141	0.9399859	0.1164032	0.8835968

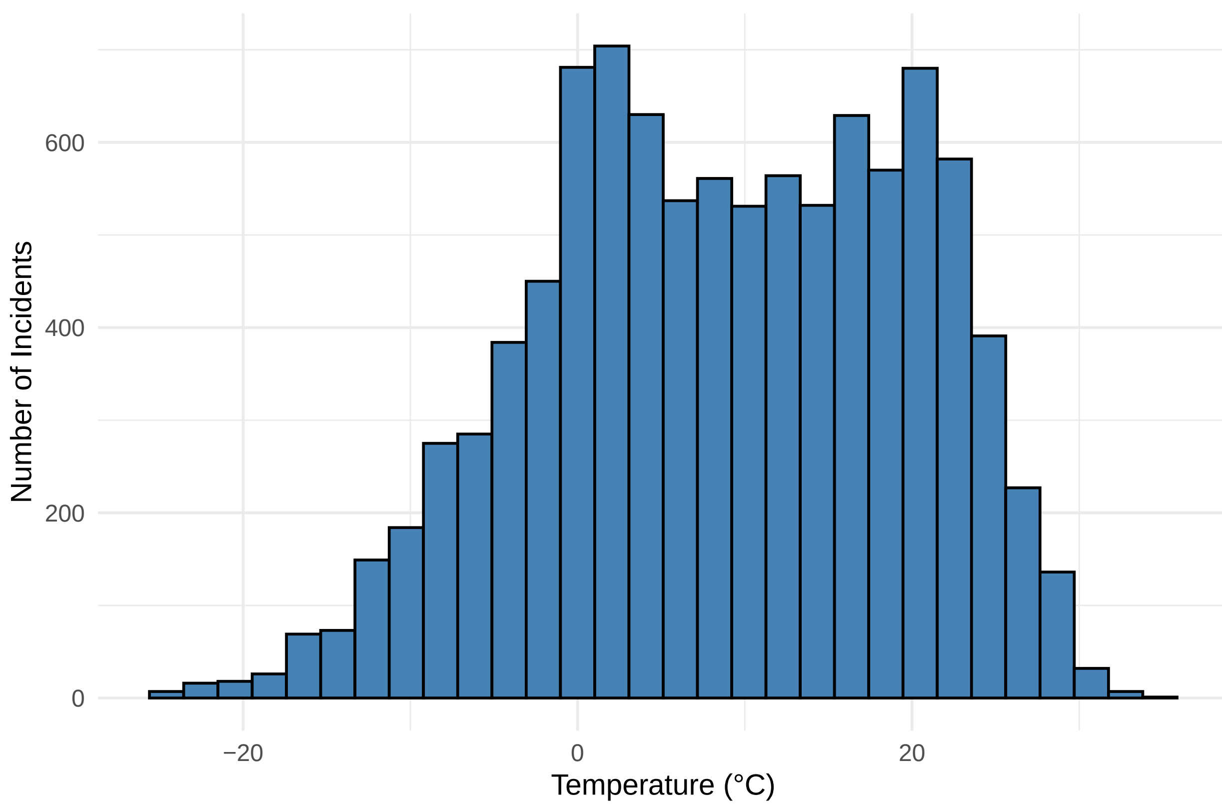
Table 1 reveals that fire incidents occurred across a wide range of temperatures, from -24.6°C to 34.8°C, with a median of 8.9°C. Wind speed, after log transformation, ranged from approximately 0.4m/s to 2.6m/s, with a median of 1.5m/s. The log-transformed estimated dollar loss varied considerably, with a minimum value of 0 and a maximum value of 16.38, indicating substantial variability in fire-related financial damages.

Table 2 shows that only 6% of fire incidents involved civilian casualties, while 94% resulted in no reported casualties. Additionally, only 11.6% of incidents occurred during rainy conditions, whereas the majority (88.4%) took place in dry conditions. This suggests that most fires happened in the absence of precipitation, warranting further investigation into whether rain reduces fire frequency or severity.

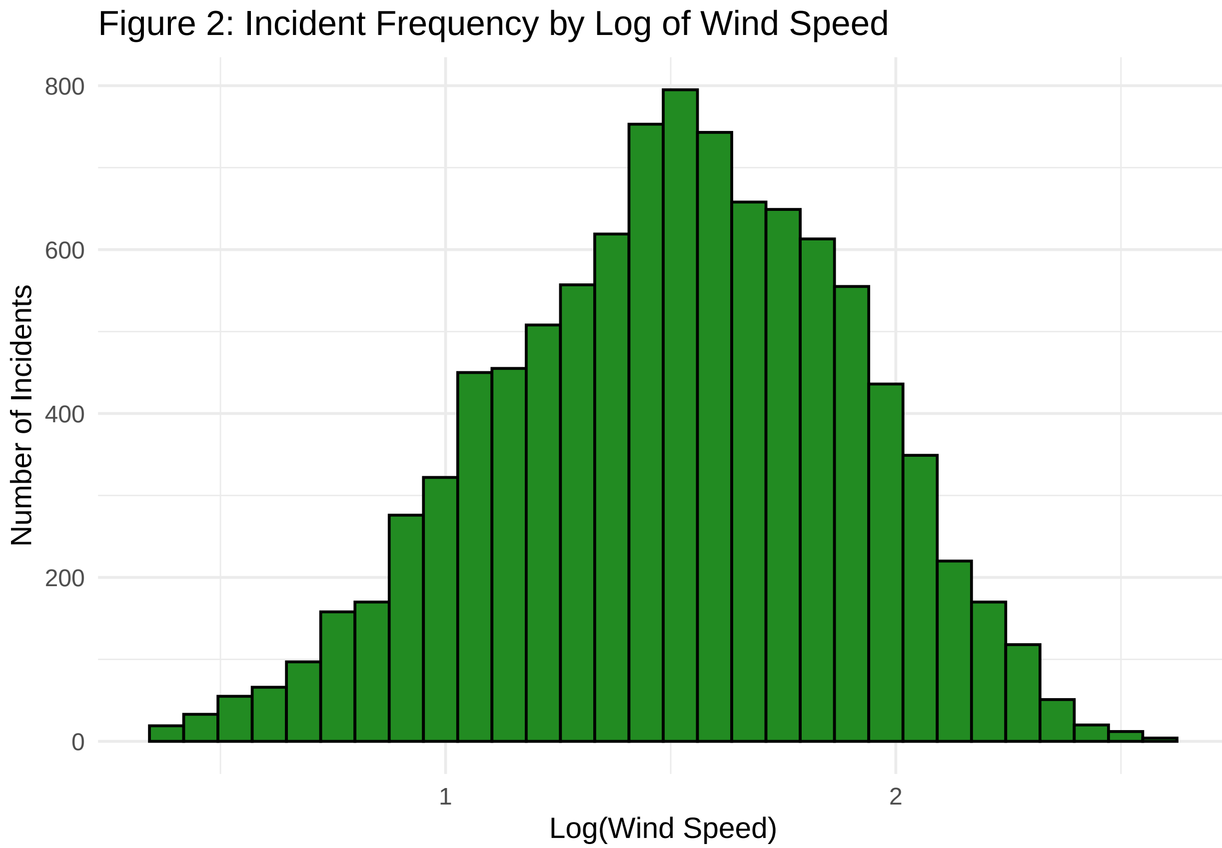
3.2 Frequency of Fire Incidents by Weather Conditions

The frequency of fire incidents varied notably across different weather conditions. Figure 1 below shows the distribution of fire incidents by temperature, indicating that fire incidents occurred across a broad temperature range, though incidents were relatively more frequent during moderate temperatures.

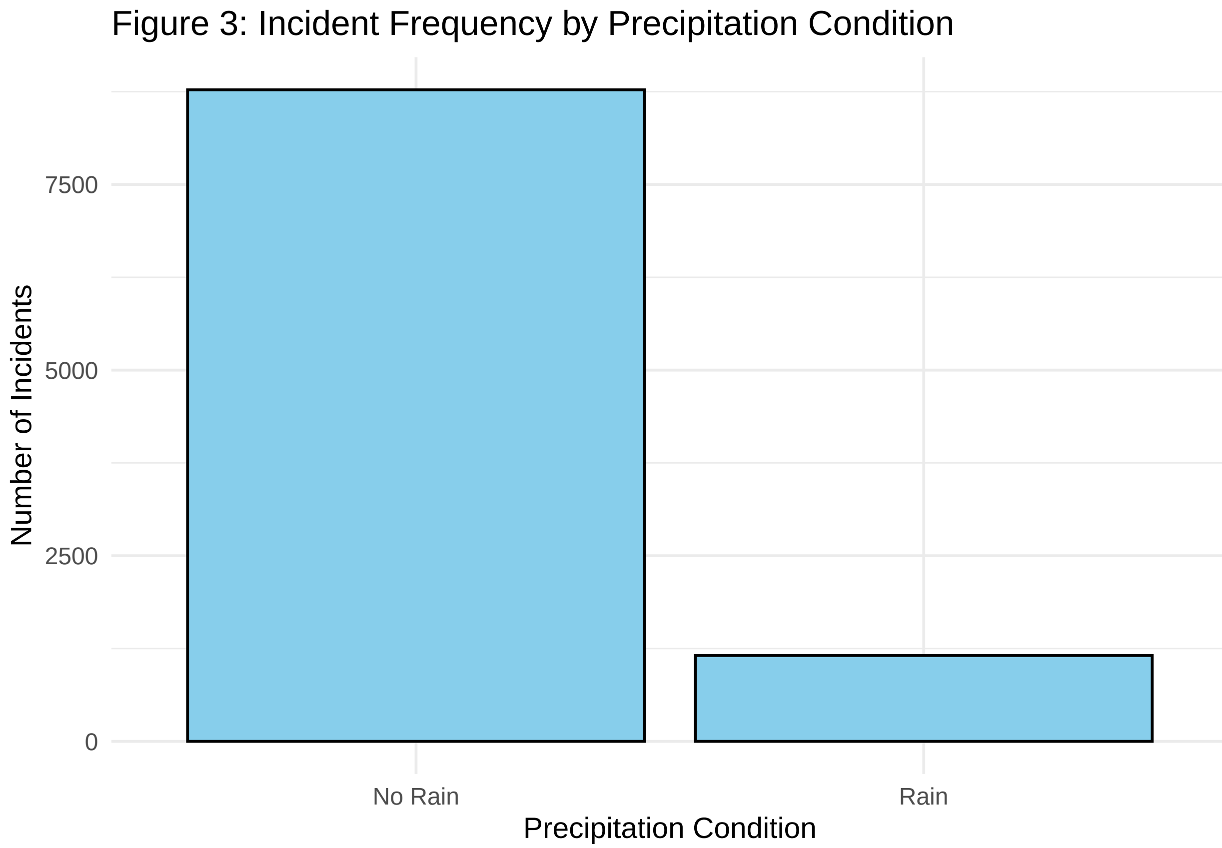
Figure 1: Incident Frequency by Temperature



In Figure 2, fire incidents are shown by log-transformed wind speed, illustrating a relatively normal distribution with a concentration of incidents at moderate wind speeds.

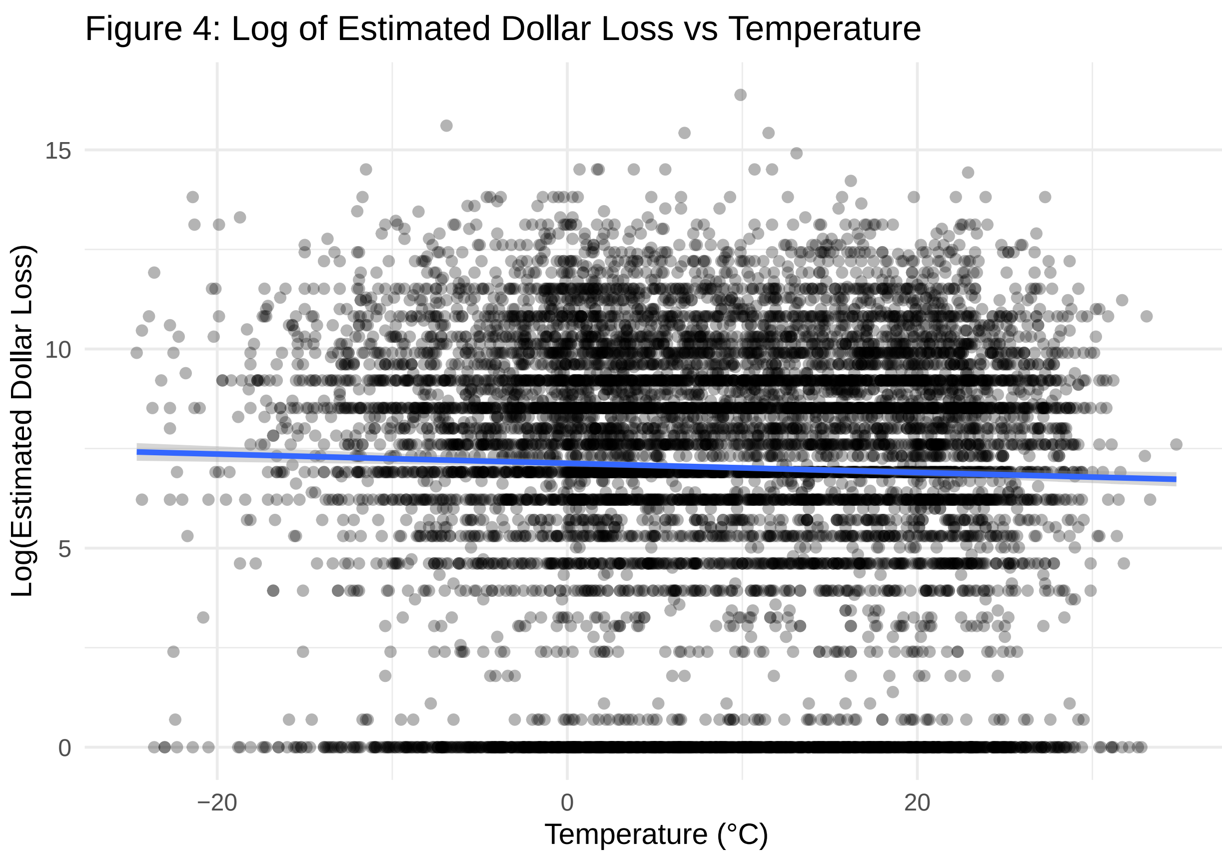


Regarding precipitation, Figure 3 below highlights that the vast majority of fire incidents occurred during dry conditions (no rain), suggesting potential influences of precipitation on the frequency of incidents.



3.3 Relationships Between Weather and Severity of Incidents

Scatterplots provided initial visual insights into how temperature and wind speed relate to the severity of fire incidents measured by estimated dollar loss. Figure 4 illustrates a scatterplot of temperature versus log-transformed estimated dollar loss, showing a weak negative linear relationship.

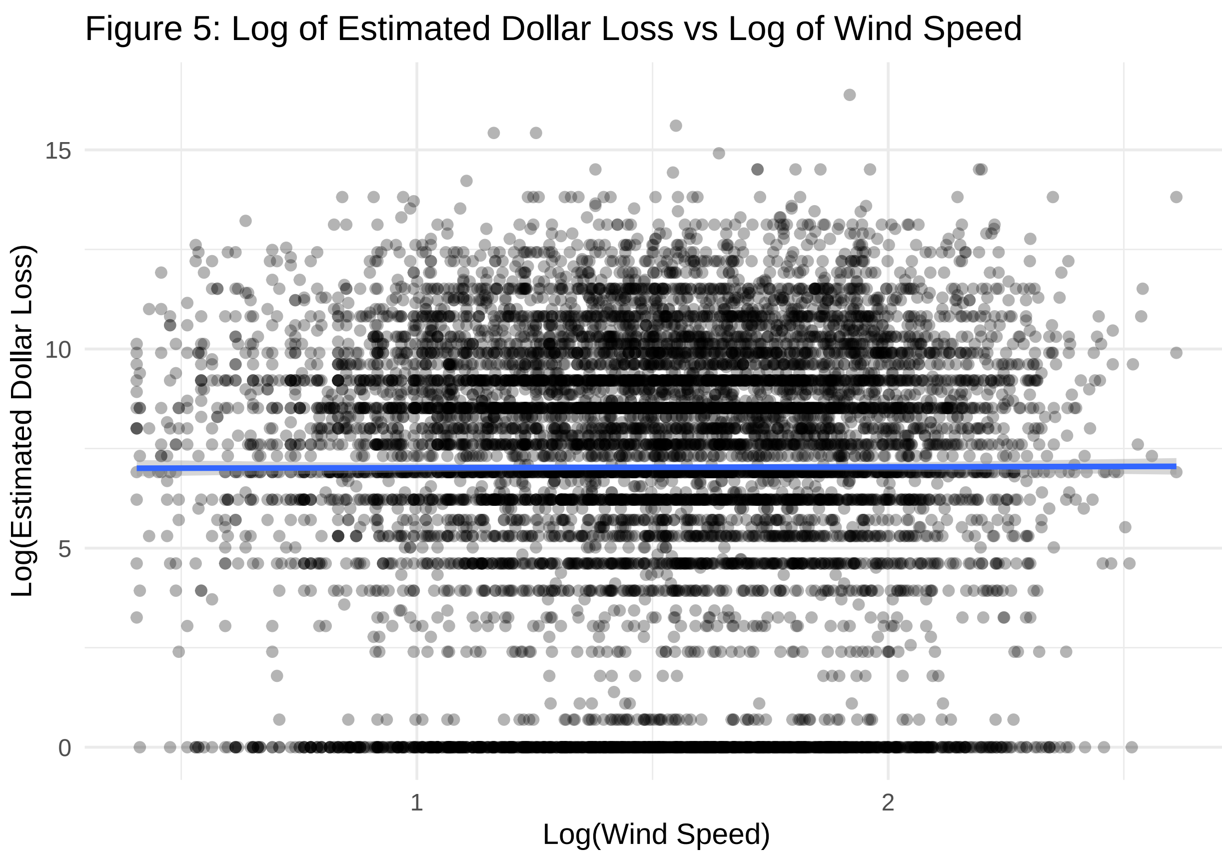


The summary of this linear model is shown in Table 3. It confirms the previous indication of a weak but statistically significant negative relationship between temperature and the log of estimated dollar loss. Specifically, the estimated coefficient for Temperature is approximately -0.01 (p-value < 0.05), suggesting that for each 1°C increase in temperature, the log-transformed dollar loss decreases by about 0.01 units. Interpreted on the original scale, this translates to roughly a 1% decrease in estimated dollar loss per degree Celsius increase.

Table 3: Table 3: Summary Table of Linear Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	7.130457	0.0452463	157.591888	0.0000000
Temperature	-0.011555	0.0032256	-3.582259	0.0003423

Similarly, Figure 5 depicts the relationship between log-transformed wind speed and log-transformed estimated dollar loss, suggesting no strong linear relationship but potentially indicating a subtle pattern worth further investigation.

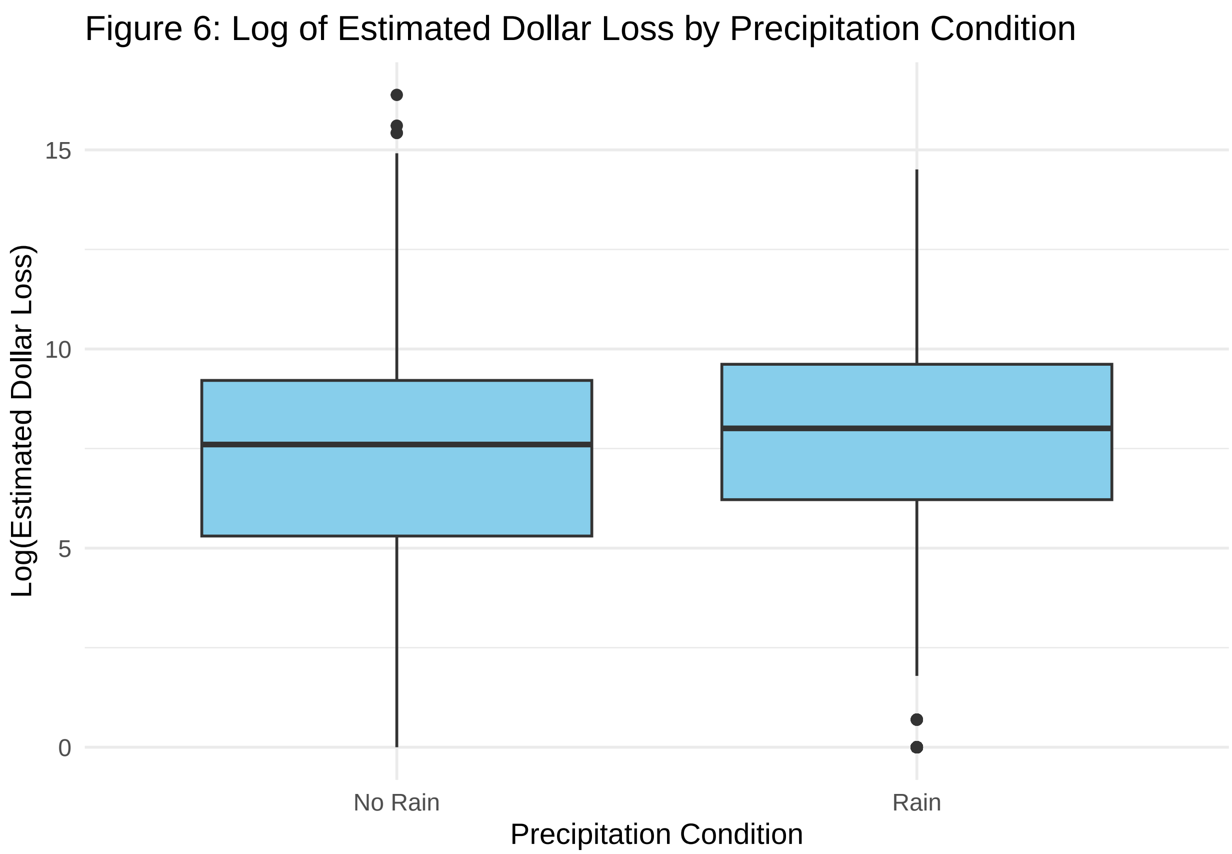


The summary of this linear model is shown in Table 4. In contrast to the previous model, Wind_Speed_Log exhibits a small positive coefficient (approximately 0.02) with a high p-value (0.81), indicating no statistically significant linear relationship between the log of wind speed and the log of estimated dollar loss.

Table 4: Table 4: Summary Table of Linear Model

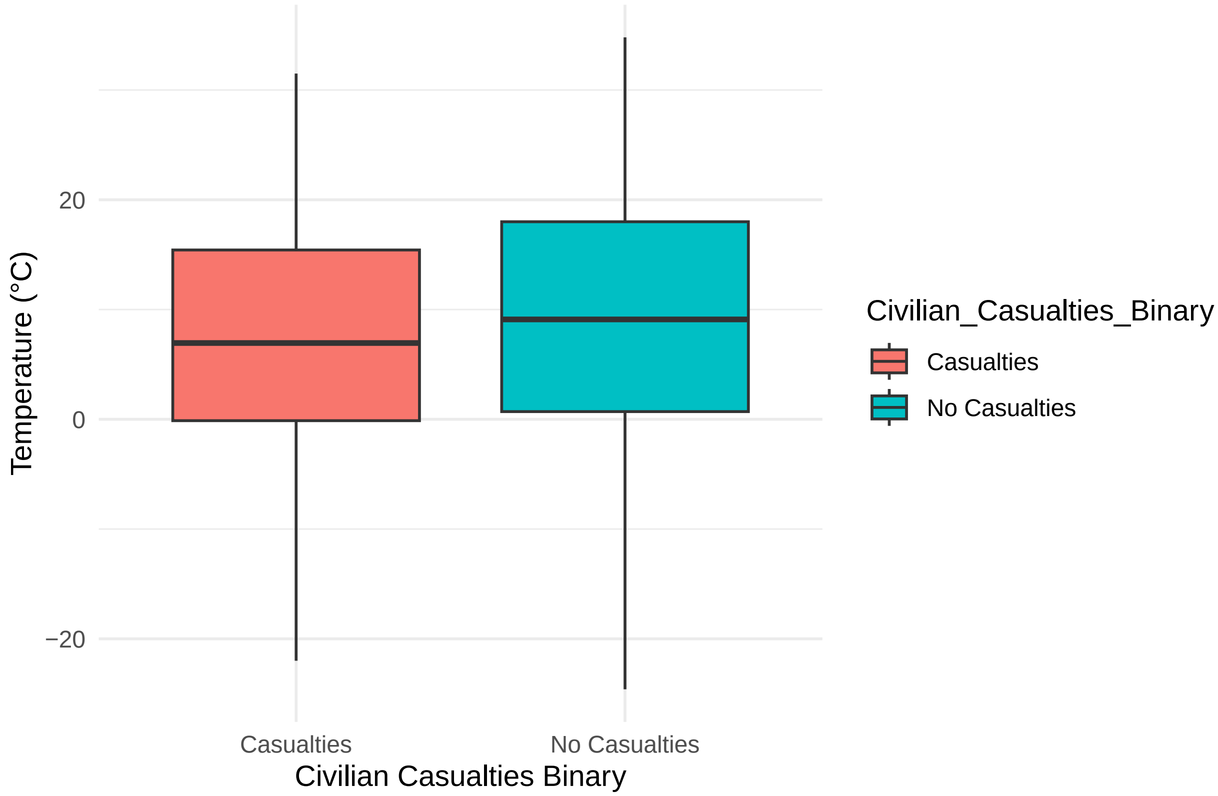
	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	6.9972629	0.1404805	49.8094791	0.0000000
Wind_Speed_Log	0.0213014	0.0901894	0.2361852	0.8132938

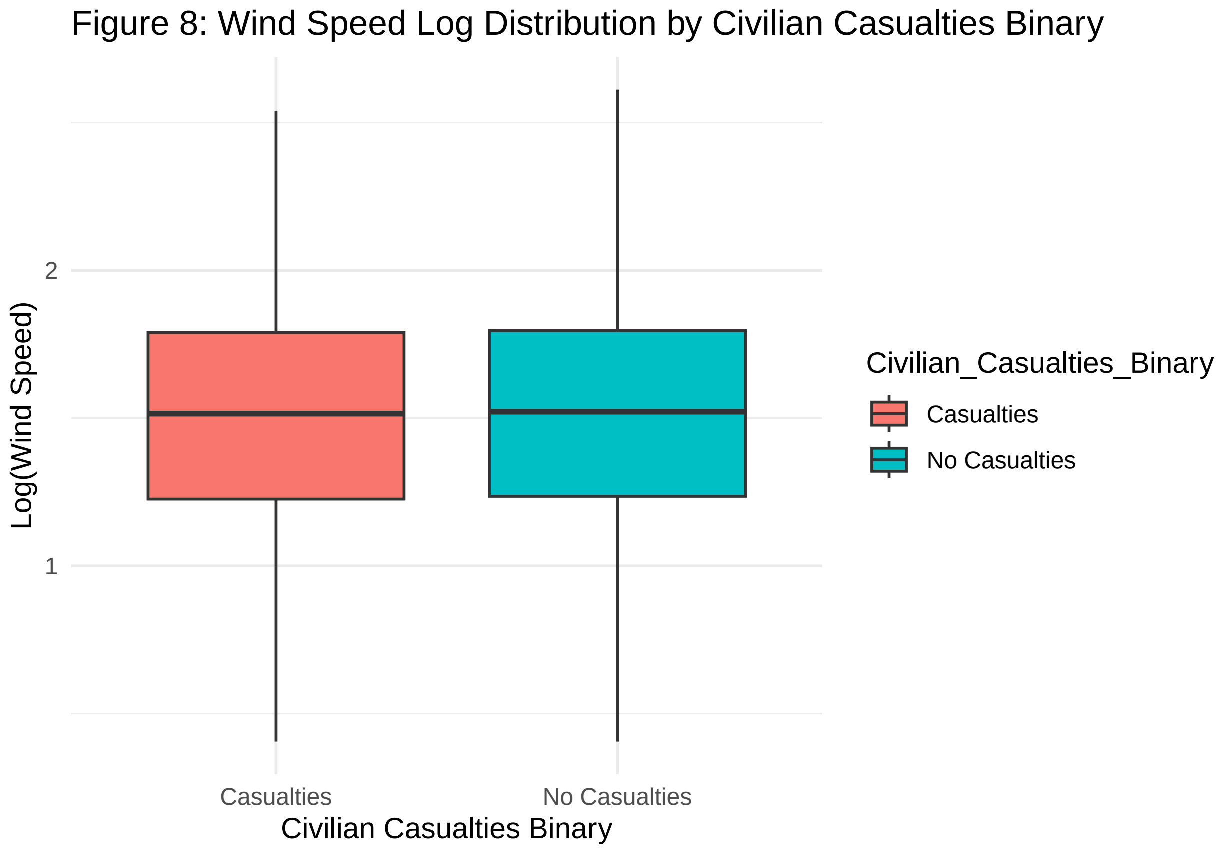
The boxplot in Figure 6 further highlights a comparison of the log-transformed estimated dollar loss between rainy and non-rainy conditions, suggesting slight differences in financial severity based on precipitation.



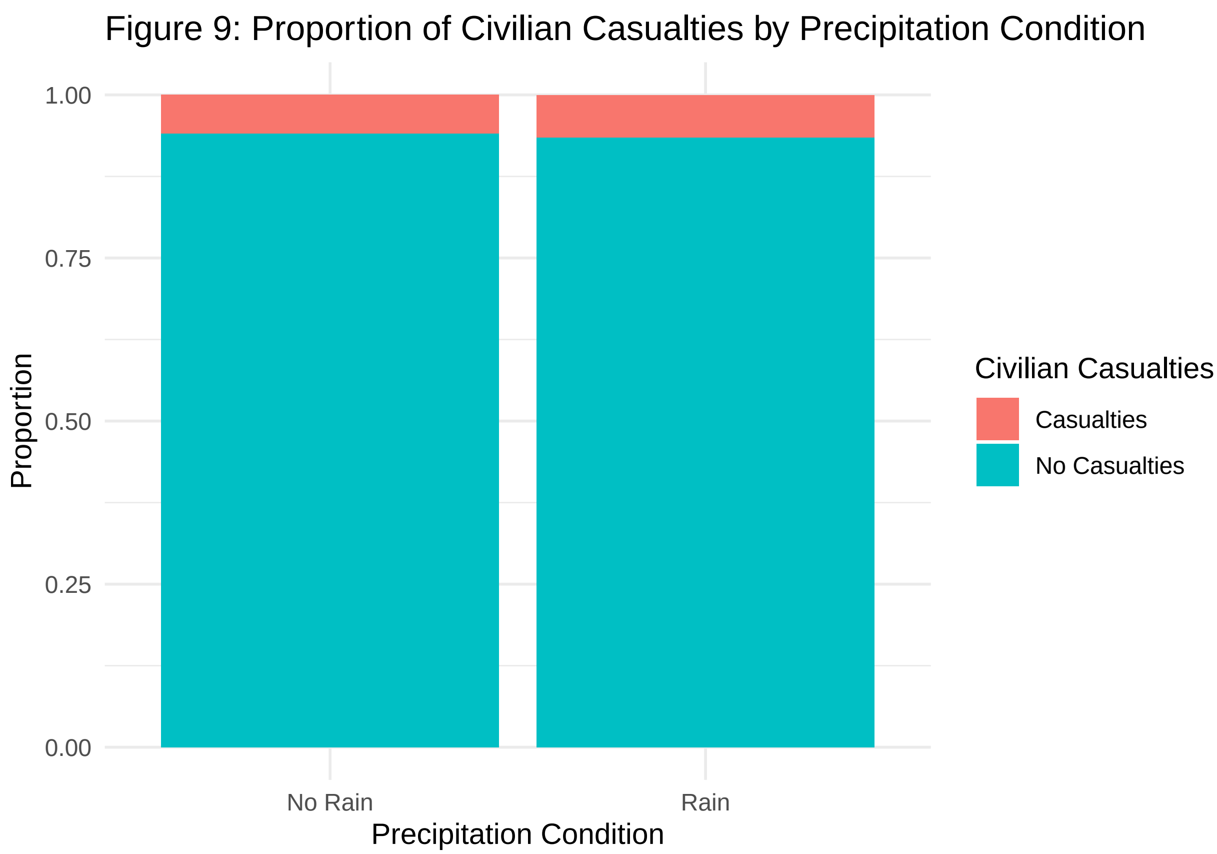
Additionally, to explore how weather conditions may influence the occurrence of civilian casualties, Figure 7 (Temperature vs Casualties) and Figure 8 (Log of Wind Speed vs Casualties) were created, comparing the distributions of temperature and log-transformed wind speed between incidents with and without casualties. Though initial differences appear minimal, these visuals justify further statistical investigation.

Figure 7: Temperature Distribution by Civilian Casualties Binary





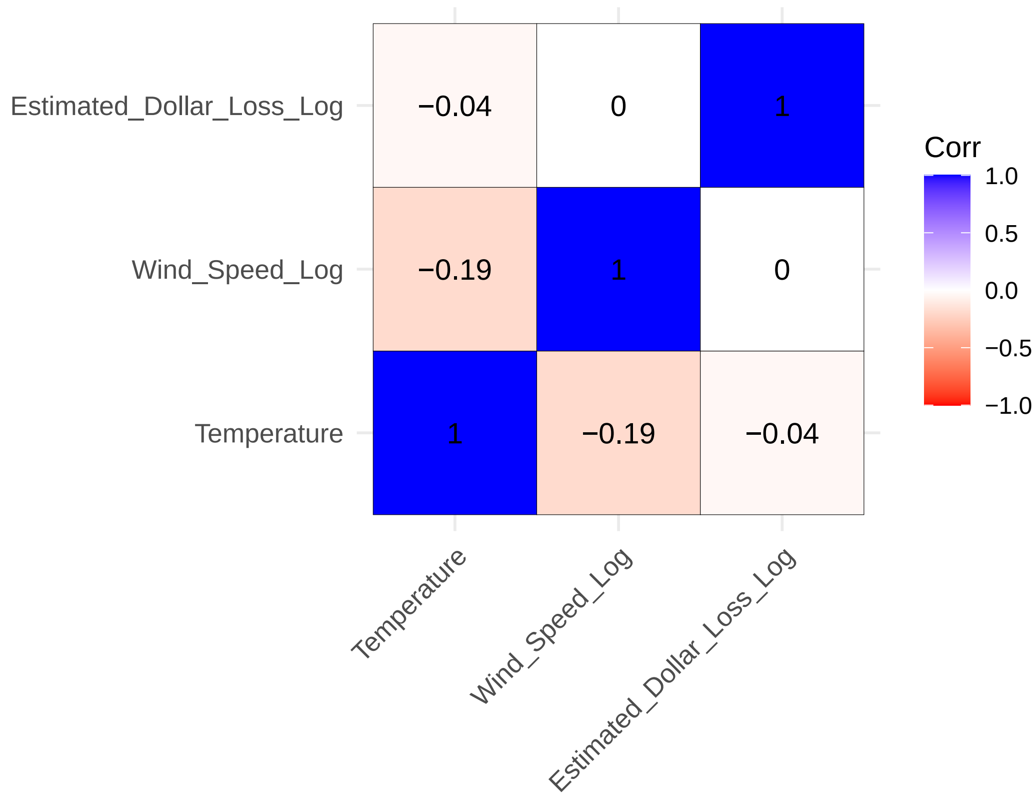
Lastly, a stacked bar plot (Figure 9) visually summarizes the proportion of incidents with civilian casualties across rainy and non-rainy conditions, providing initial insights into whether precipitation might influence fire-related injuries or fatalities.



3.4 Correlation Among Numeric Variables

The correlation heatmap (Figure 10) provides a concise summary of correlations among numerical variables. The weak correlations indicate potential non-linear relationships or interactions that may be explored further in statistical modeling.

Figure 10: Correlation Heatmap of Numeric Variables



4 Summary

4.1 Current Findings

The preliminary analysis provided insightful initial observations regarding the relationship between weather conditions and the frequency and severity of fire incidents in Toronto. Descriptive statistics and exploratory visualizations suggest that temperature, wind speed, and precipitation might have subtle, yet potentially meaningful effects.

Specifically, the data revealed that fire incidents occurred across a wide temperature range, indicating that temperature alone may not strongly limit the occurrence of fires but could still influence their severity. Wind speed displayed variability, but its relationship with estimated financial damage was weak, suggesting possible non-linear or complex interactions that require more advanced modeling to detect clearly. Precipitation showed a clearer pattern: majority of incidents occurred under dry conditions, raising the hypothesis that rainfall may reduce the likelihood of fire incidents or influence their severity.

Furthermore, the analysis highlighted substantial variation in estimated dollar loss, with a small proportion of incidents involving civilian casualties. This emphasizes the importance of separately analyzing incident severity, as measured by casualties and financial losses, to more thoroughly assess the impact of weather conditions.

4.2 Limitations

There are limitations in the current exploration. For example, the above analysis only involves a restricted amount of data and more data should be pulled using the Open-Meteo Historical Weather API for a more comprehensive analysis. Additionally, the above analysis only fits simple linear regression models, which does not have much expressive power to model a complicated correlation. Therefore, more complicated models should be fit to better capture the intrinsic patterns within the data.

4.3 Plan for Final Project

To complete the analysis and address the formulated research question comprehensively, the following steps are planned for the subsequent stages of the project:

1. Statistical Modeling

- Apply generalized linear models (GLM) or generalized additive models (GAM) to explore the relationships between weather conditions and both measures of severity: the occurrence of civilian casualties (binary response) and the magnitude of estimated dollar loss (continuous response).
- Evaluate the significance and predictive strength of temperature, wind speed, and precipitation through model diagnostics and statistical tests, such as likelihood-ratio tests and ANOVA tests.

2. Advanced Machine Learning Approaches

- Implement decision trees, random forests, and gradient boosting models to capture potential non-linear interactions or complex dependencies among variables that were not clearly revealed in preliminary linear analysis.
- Compare the performance of these models using appropriate metrics to identify the most reliable predictive factors.

3. Temporal and Interaction Effects

- Investigate whether the impact of weather conditions varies by season or time of day, given that temporal factors could potentially influence these relationships.
- Analyze potential interaction effects (e.g., how precipitation may modify the influence of wind speed or temperature).

4. Interactive Visualizations

- Create interactive visualizations to effectively communicate key findings.