An Analysis of a Rising Trend in Dollar Loss in Fire Incidents With and Without Sprinkler Systems in Toronto over a decade (2011-2022)*

Aryaman Sharma

January 22, 2024

A Fire Incidents dataset was used to assess the general trend in dollar loss in fire incidents and the impact of existence of sprinkler systems installed across the city of Toronto. Based on the analysis, a general rise over the years in the dollar loss in fire incidents as well as a massive amount of loss to property was found wherever sprinkler systems were not installed. This paper suggests that an installation of sprinkler systems will help reduce the impact of fire incidents on loss to property in Toronto.

Table of contents

*Code and data are available at: https://github.com/Ary4m3n/fire-incidents.git

1 Introduction

The Great Fire of Toronto in 1904 caused mass destruction. The fire demolished around 20 acres of the industrial area in the city of Toronto, demolishing at least 98 buildings, causing a dollar loss of $10 million and leaving numerous thousands unemployed (Bradburn 2020). Since then we have seen an improvement in city codes and policies, better civil planning and extensive training of firefighters (Pandas 2022). Till date, the Toronto Fire Services report approximately 10,000 fires each year which is of immense concern to the city of Toronto and its residents (Ohrn 2019).

The Fire Incidents data of Toronto includes incident reports of fire incidents as defined by the Ontario Fire Marshal from 2011-2022 (Data 2024). This paper delves deeper into the trend of estimated dollar loss in fire incidents in the city of Toronto from 2011-2022 and also analyses the impact of an installation of sprinkler systems on the dollar loss. The aim of the paper is to analyse the impact of better planning, specifically an installation of sprinkler systems, on the extent of loss faced by the residents of Toronto.

In this paper, Fire Incidents data (Data 2024) was used to first explore if there is a trend in the estimated dollar loss over the years, and then was used to find any correlation between better civil planning, i.e. the installation of sprinkler systems and a lesser loss of property. It was found that there is a general rising trend in the dollar loss between 2011 and 2022. The paper also found that the dollar loss was drastically higher for fire incidents where there was no sprinkler system installed. This leads us to make an important judgement that better civil planning helps reduce the impact of such unfortunate events.

This paper is structured using the following sections: Data, Results and Discussion In the Data (Section 2) section, the data source of the dataset from Open Data Toronto (Data 2024) is discussed and the data cleaning process is outlined. In the Results (Section 3) section, the paper summarizes the data findings and relevant graphs of the trends observed. The paper ends with the Discussion (Section 5) section, where the findings of the paper have been analysed and delved deeper into, and a further scope for the paper has been discussed.

2 Data

The data analysed in this paper was from Open Data Toronto (Data 2024). The data was cleaned and analysed using the open source `R` programming language (R Core Team 2022). R libraries and packages such as `opendatatoronto` (Gelfand 2022), `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `ggplot2` (Wickham 2016), `knitr` (Xie 2023), `readr` (Wickham, Hester, and Bryan 2023) and `dplyr` (Wickham et al. 2023). In the following sections, we will discuss the raw data (Section 2.1) and then move on to discussing the data cleaning process (Section 2.2).

2.1 Raw Fire Incidents Data

The raw data for this paper was obtained from Open Data Toronto (Data 2024) and we specifically looked at the Fire Incidents Data provided by the Ontario Fire Marshal from 2011-2022. We also looked at another data set related to this topic, Fire Services Incident Data (Data 2019) but that data set firstly was last refreshed in 2019 and had been retired and replaced with the data set we used.

The raw data had loads of information (that was mainly unnecessary to us for this paper), which mainly included the area of origin, casualities, persons rescued, estimated dollar loss, fire incident type, fire timing, intersection (location), cause, sprinkler system presence etc. For this paper specifically, to observe and analyse the trend in dollar loss and the effect of better civil planning (i.e. existence of sprinkler systems), we did not require such a plethora of variables. The raw data set contains 2,357,639 data entries and 43 variables. Due to the size of the dataset, we could not include a table outlining the structure of the dataset.

In the next section (Section 2.2), we will outline the data-cleaning process and also show the first few rows of the cleaned data.

2.2 Cleaning Fire Incidents Data

As stated above in (Section 2.1), we cleaned the data to cater to our needs of analyzing the trend in dollar loss and the effects of better civil planning. Specifically, we only kept 3 variables, the estimated dollar loss, the year in which the incident took place and sprinkler system presence which outlines whether a sprinkler system had been installed in the place where the incident took place. Table 1 shows the structure of the cleaned data by presenting the first 6 datapoints or rows.

Here, in order to obtain this cleaned data table, I got the year of the incident from the `alarm_time` variable in the raw data and cleaned the sprinkler system presence column to only show 4 unique statuses, “No sprinkler system”, “Full sprinkler system present”, “Undetermined” and “NA” as shown clearly in Table 1 using `R` (R Core Team 2022). Table 1 shows the head of the cleaned data.

Table 1: Cleaned Data showing the first 6 rows

Dollar Loss	Year	Sprinkler System Presence
15000	2018	NA
50	2018	NA
0	2018	Undetermined
1500	2018	NA
2000	2018	No sprinkler system
100000	2018	No sprinkler system

Now that we have cleaned our raw fire incidents data and have generated a clean table. We will go on to the Results section (Section 3) and look at the trend generated for the dollar loss over the years, 2011-2022 (Section 3.1), and then look at the effect of the presence of sprinkler systems on the property or dollar loss (Section 3.2).

3 Results

3.1 Trend in Dollar Loss over 2011-2022 in Fire Incidents

Table 2 shows the table of the total estimated dollar loss over the years 2011-2022.

Table 2: Trend

Year	Dollar Loss	Dollar Loss in Millions
2011	50014115	50.01411
2012	42482142	42.48214
2013	52232801	52.23280
2014	61145851	61.14585
2015	42223795	42.22380
2016	60803825	60.80383
2017	77320995	77.32099
2018	77291443	77.29144

Year	Dollar Loss	Dollar Loss in Millions
2019	119116686	119.11669
2020	70524276	70.52428
2021	84039521	84.03952
2022	88676002	88.67600

Figure 1 shows the bar plot of the Table 2 where the Dollar Loss in millions is plotted against the years 2011-2022.

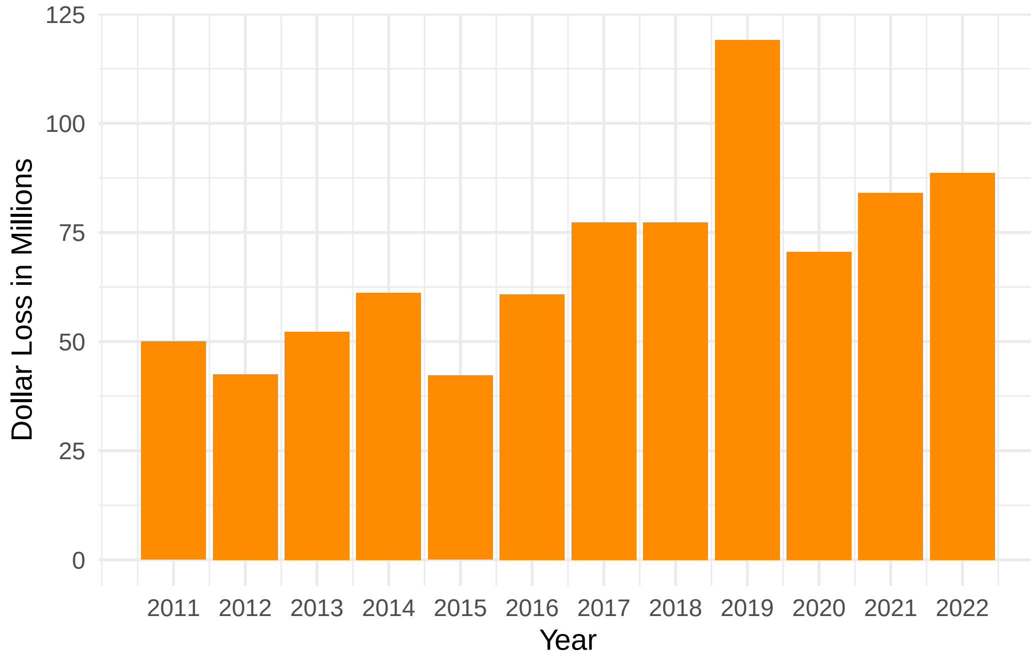


Figure 1: Year vs Dollar Loss ($million) in Fire Incidents in Toronto (2011-2022)

3.2 Effect of presence of Sprinkler Systems on Dollar Loss

Table 3 shows the table of Dollar Loss over the years in the presence of proper sprinkler systems.

Table 3: Full Sprinkler System Present

Year	Dollar Loss	Dollar Loss in Millions
2011	3826290	3.826290
2012	8488092	8.488092

Year	Dollar Loss	Dollar Loss in Millions
2013	3872366	3.872366
2014	6040676	6.040676
2015	3054784	3.054784
2016	4082601	4.082601
2017	2949971	2.949971
2018	4864365	4.864365
2019	5949001	5.949001
2020	5466891	5.466891
2021	7583691	7.583691
2022	5888177	5.888177

Table 4 shows the table of Dollar Loss over the years when there were no sprinkler systems present.

Table 4: No Sprinkler System Present

Year	Dollar Loss	Dollar Loss in Millions
2011	37276365	37.27636
2012	22779879	22.77988
2013	35350140	35.35014
2014	31118999	31.11900
2015	25615889	25.61589
2016	38632389	38.63239
2017	37853068	37.85307
2018	43650217	43.65022
2019	94093947	94.09395
2020	41548697	41.54870
2021	39780346	39.78035
2022	53476753	53.47675

Some of our data is of penguins (Figure 2), from Horst, Hill, and Gorman (2020).

Talk more about it.

And also planes (Figure 3). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

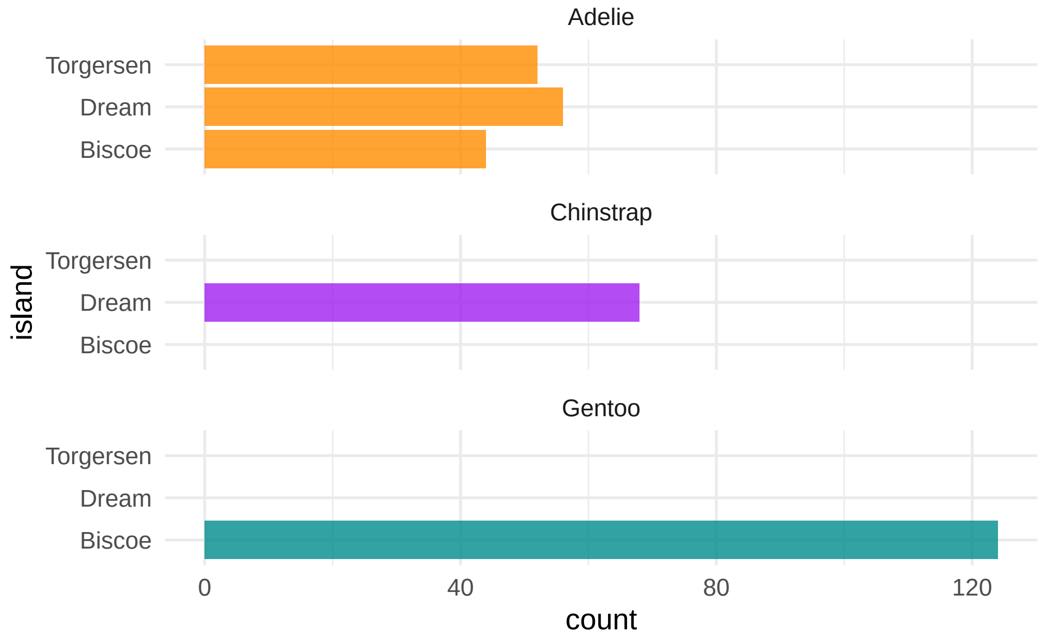


Figure 2: Bills of penguins

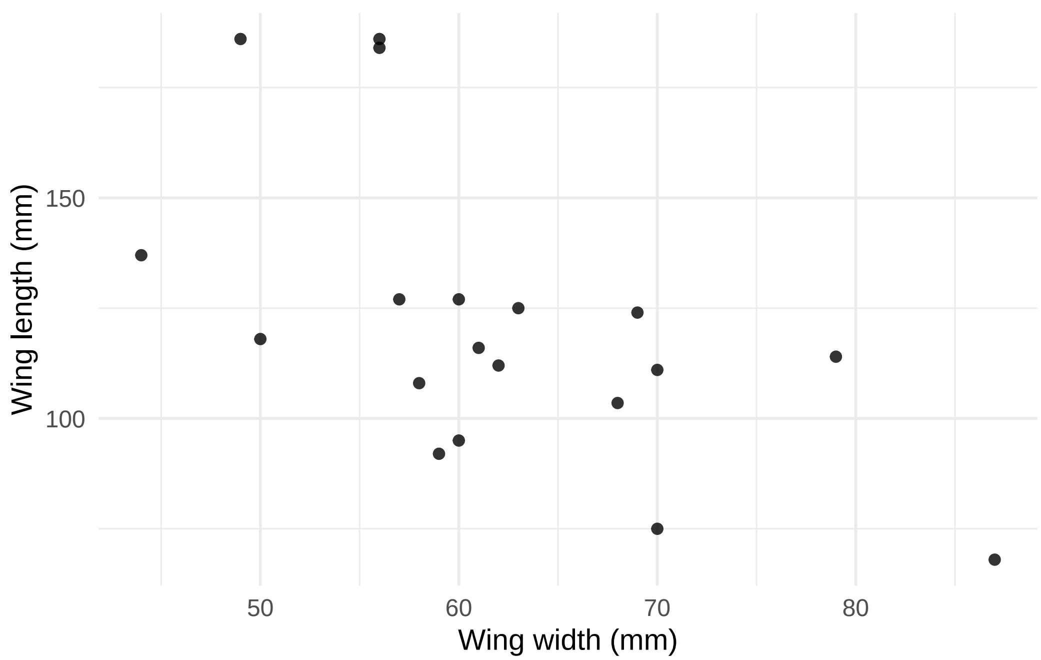


Figure 3: Relationship between wing length and width

4 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

4.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$
\begin{aligned}
y_i|\mu_i, \sigma &\sim \text{Normal}(\mu_i, \sigma) && (1)\\
\mu_i &= \alpha + \beta_i + \gamma_i && (2)\\
\alpha &\sim \text{Normal}(0, 2.5) && (3)\\
\beta &\sim \text{Normal}(0, 2.5) && (4)\\
\gamma &\sim \text{Normal}(0, 2.5) && (5)\\
\sigma &\sim \text{Exponential}(1) && (6)
\end{aligned}
$$

We run the model in R (R Core Team 2022) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

4.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

Our results are summarized in Table 5.

5 Discussion

5.1 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.

Table 5: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

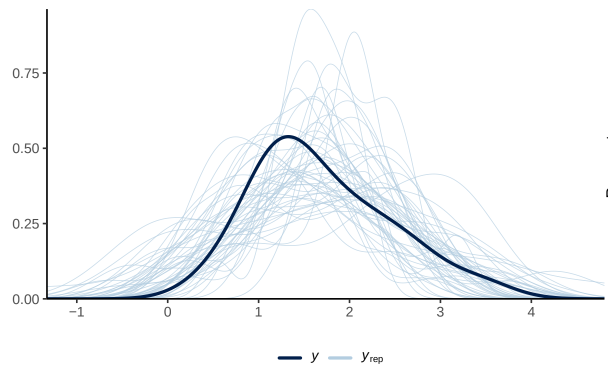
A Additional data details

B Model details

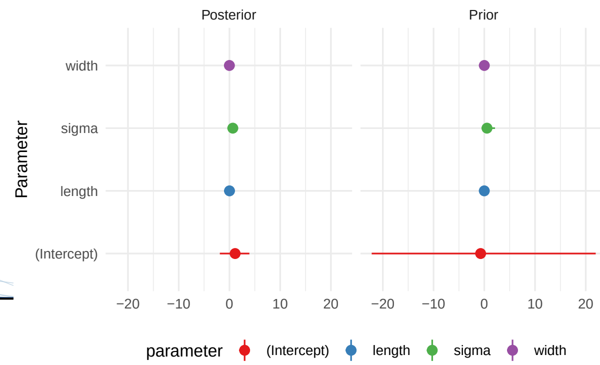
B.1 Posterior predictive check

In Figure 4a we implement a posterior predictive check. This shows...

In Figure 4b we compare the posterior with the prior. This shows...



(a) Posterior prediction check



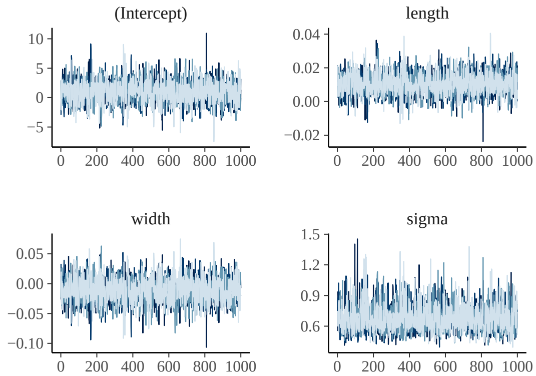
(b) Comparing the posterior with the prior

Figure 4: Examining how the model fits, and is affected by, the data

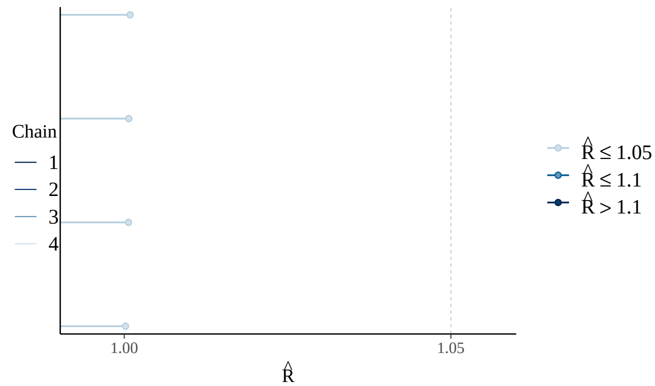
B.2 Diagnostics

Figure 5a is a trace plot. It shows... This suggests...

Figure 5b is a Rhat plot. It shows... This suggests...



(a) Trace plot



(b) Rhat plot

Figure 5: Checking the convergence of the MCMC algorithm

References

Bradburn, Jamie. 2020. *Great Fire of Toronto (1904)*. *The Canadian Encyclopedia*. www.thecanadianencyclopedia.ca/en/article/great-fire-of-toronto-1904.

Data, Open. 2019. *Fire Service Incident Data*. https://open.toronto.ca/dataset/fire-services-incident-data/.

———. 2024. *Fire Incidents*. https://open.toronto.ca/dataset/fire-incidents/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *Palmerpenguins: Palmer Archipelago (Antarctica) Penguin Data*. https://doi.org/10.5281/zenodo.3960218.

Ohrn, Anders. 2019. *Toronto on Fire in Data, Part 1*. *Towards Data Science*. https://towardsdatascience.com/toronto-on-fire-in-data-part-1-484435eca880.

Pandas, Toronto Department of. 2022. *An Analysis of the City of Toronto's Fire Response Between 2011–2019*. *Medium*. https://medium.com/@cityoftorontopandas/an-analysis-of-the-city-of-torontos-fire-response-between-2011-2019-6f03f08b89d4.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.