My title*

My subtitle if needed

Andy Jiang

November 26, 2024

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

1 Introduction

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

2.3 Traffic Speed Dataset

The traffic speed dataset contains information collected from multiple locations and months, detailing percentile speeds, total traffic volume, and counts of vehicles traveling within specific speed ranges. The dataset aggregates these metrics monthly, offering insights into traffic behavior and variations over time.

*Code and data are available at: [https://github.com/AndyYanxunJiang/toronto-vehicles-speed-analysis).

This dataset includes: - Percentile speeds (5th, 10th, … 95th) representing speed thresholds exceeded by corresponding proportions of vehicles. - Traffic volume as the total number of vehicles observed monthly. - Speed bins (e.g., 0-4 km/h, 5-9 km/h) for detailed breakdowns of vehicle counts across speed ranges. - The table below, Table 1, showcases a preview of the cleaned dataset with selected columns to highlight key features.

Table 1: Preview of cleaned traffic speed dataset.

Month	5th Percentile	Median (50th Percentile)	95th Percentile	Vehicles (30-34 km/h)	Vehicles (50-54 km/h)	Vehicles (>100 km/h)	Total Volume
2024-09-01	7	21	32	2636	5	0	57758
2024-09-01	5	18	29	1463	1	0	48365
2024-09-01	0	0	28	162	0	0	2116
2024-09-01	21	40	55	16492	11560	4	163591
2024-09-01	0	28	46	3524	444	0	30606
2024-09-01	17	38	49	8954	1957	1	64690
2024-09-01	0	29	40	8312	72	0	27790
2024-09-01	0	11	26	535	0	0	59467
2024-09-01	0	49	65	1472	15730	7	62706
2024-09-01	6	33	44	8899	257	0	34309

2.4 Random Sample from the Dataset

Table 2 provides a random sample of 10 observations, offering an unbiased snapshot of key metrics such as speed percentiles, total traffic volume, and counts of vehicles exceeding 100 km/h. This selection illustrates the dataset's diversity without emphasizing specific patterns.

Table 2: Random sample of traffic speed dataset.

Month	5th Percentile	Median (50th Percentile)	95th Percentile	Total Volume	Vehicles (>100 km/h)
2024-03-01	15	39	51	33333	0
2020-02-01	11	28	42	40020	0
2024-09-01	9	24	38	117573	0
2022-12-01	0	20	34	31121	4
2021-07-01	14	37	51	160808	1
2023-09-01	30	50	65	182810	23
2022-06-01	8	22	37	58684	0
2022-02-01	15	25	47	195854	0
2020-11-01	14	36	47	87896	0
2024-02-01	0	0	28	3472	0

2.5 Extreme Speeds Dataset

Vehicles traveling at extreme speeds (over 100 km/h) represent a critical factor in understanding traffic safety and violations. The table below, Table 3, highlights the top 5 months with the highest counts of vehicles exceeding 100 km/h. This provides insights into when extreme speeds are most prevalent.

Table 3: Top 5 months with the highest counts of vehicles exceeding 100 km/h.

Month	Total Volume	Vehicles (>100 km/h)	95th Percentile Speed
2021-11-01	45735	3423	122
2023-01-01	62476	1640	85
2023-01-01	87523	1475	77
2020-06-01	273454	1391	87
2020-07-01	224624	1304	52

2.6 Proportion of Vehicles at Moderate Speeds

Another key insight from this dataset is the proportion of vehicles traveling within a moderate speed range (50–70 km/h). This measure helps understand how typical driving speeds align with safe and efficient traffic flow.

The table below, Table 4, highlights monthly proportions of vehicles in this speed range.

Table 4: Monthly proportions of vehicles traveling at moderate speeds (50–70 km/h).

Month	Moderate Speed Vehicles	Total Volume	Proportion (%)
Jan	32248526	208619163	15.46
Feb	31376308	213960232	14.66
Mar	35727340	235271611	15.19
Apr	31383270	194534832	16.13
May	32859985	207592245	15.83
Jun	32981330	201948797	16.33
Jul	33826108	196693112	17.20
Aug	33837333	195652825	17.29
Sep	36352915	240764265	15.10
Oct	32154660	203667881	15.79
Nov	33186997	217011827	15.29
Dec	32651478	211458699	15.44

Talk way more about it.

2.7 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

3 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.

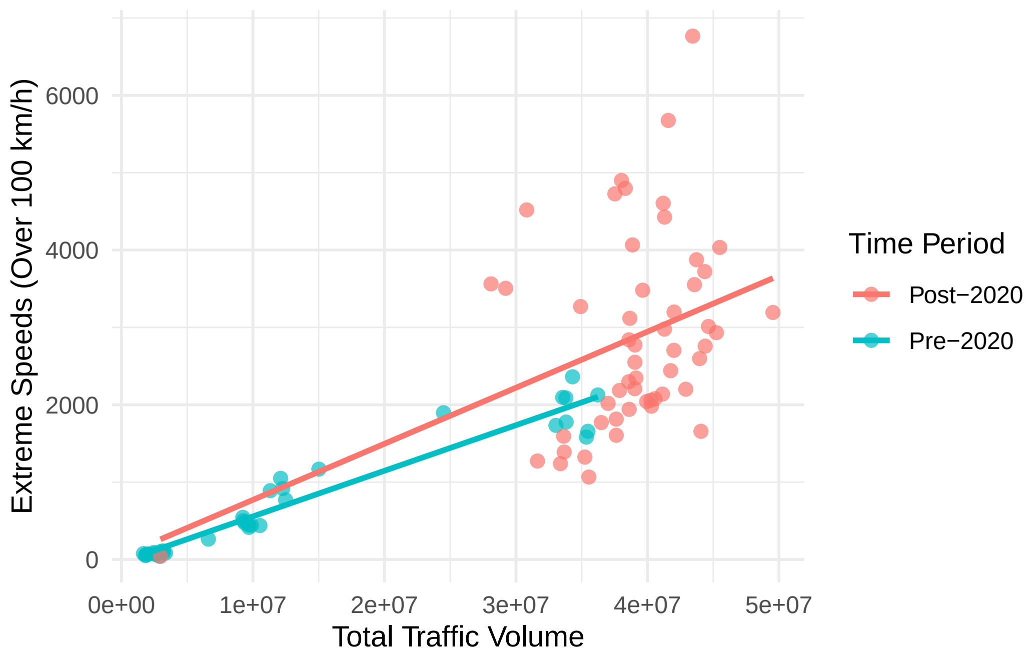


Figure 1: Relationship between extreme speeds (over 100 km/h) and total traffic volume, categorized by pre-2020 and post-2020 time periods.

3.1 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 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

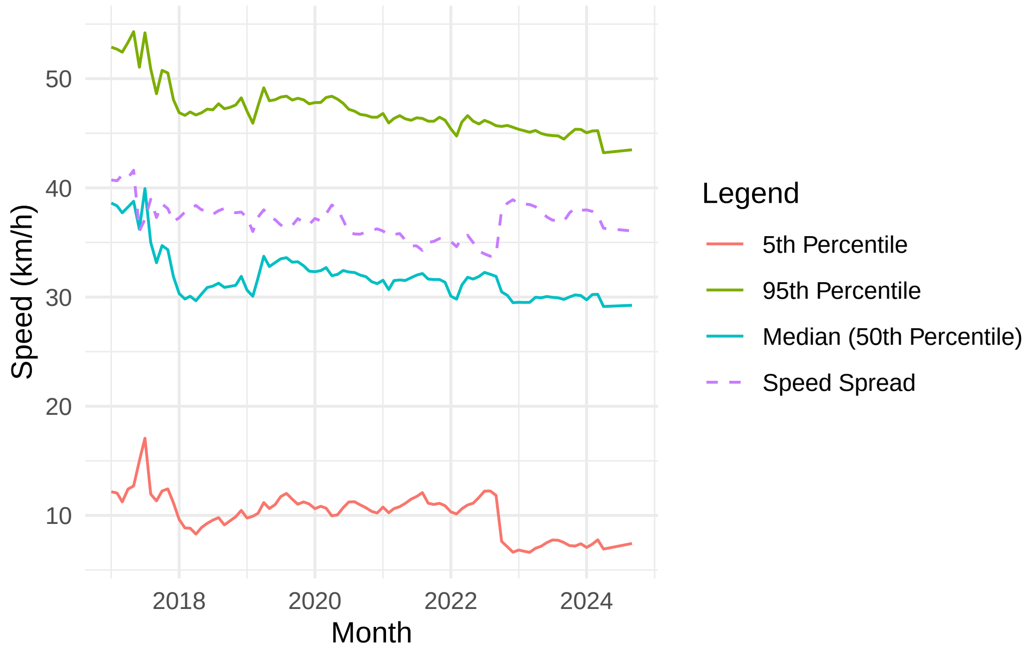


Figure 2: Trends in key speed percentiles (5th, 50th, and 95th) and the spread between the 95th and 5th percentiles over time.

3.1.1 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.

4 Results

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)	-0.88
	(0.03)
volume	0.00
	(0.00)
pct_05	0.21
	(0.00)
pct_50	0.65
	(0.00)
pct_95	0.18
	(0.00)
Num.Obs.	31 990
R2	0.992
R2 Adj.	0.992
Log.Lik.	$-41\,414.100$
ELPD	$-41\,436.5$
ELPD s.e.	385.2
LOOIC	82 872.9
LOOIC s.e.	770.4
WAIC	82 872.9
RMSE	0.88

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

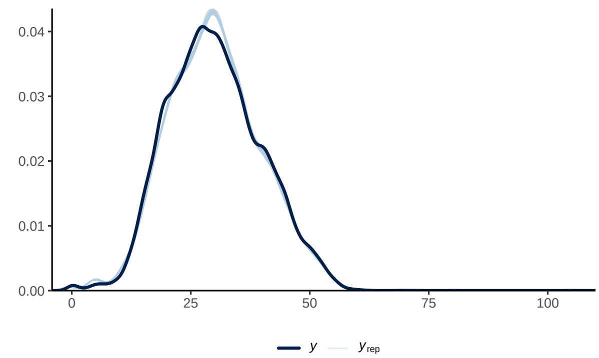
A Additional data details

B Model details

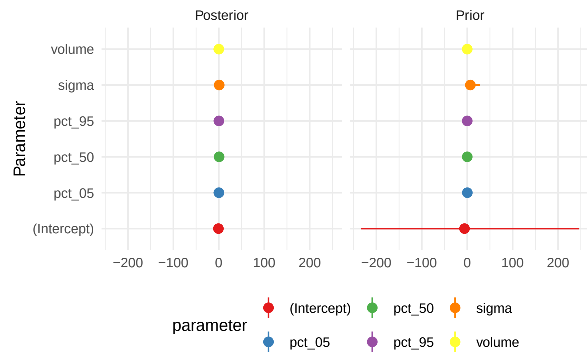
B.1 Posterior predictive check

In Figure 3a we implement a posterior predictive check. This shows...

In Figure 3b we compare the posterior with the prior. This shows...



(a) Posterior prediction check



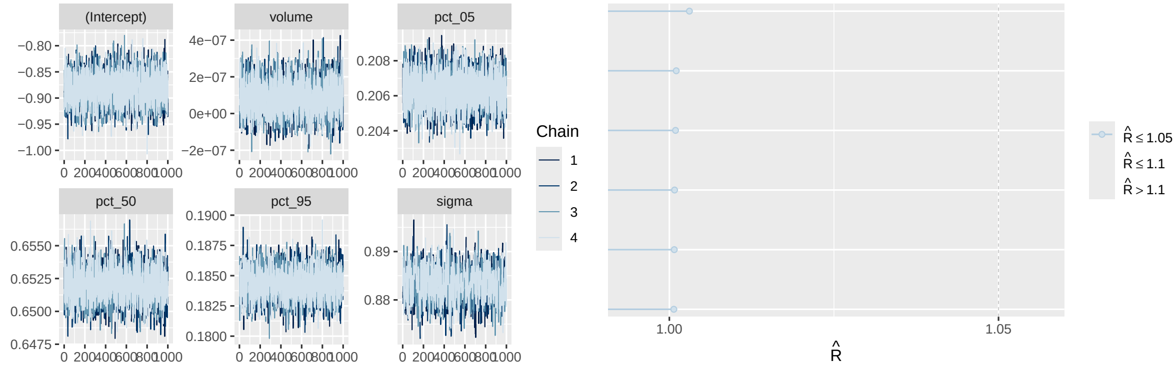
(b) Comparing the posterior with the prior

Figure 3: Examining how the model fits, and is affected by, the data

B.2 Diagnostics

Figure 4a is a trace plot. It shows... This suggests...

Figure 4b is a Rhat plot. It shows... This suggests...



(a) Trace plot

(b) Rhat plot

Figure 4: Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.