My title*

An analysis of solved and unsolved homicides from 2010 to 2017 in the United States's 2 largest cities, New York and Los Angeles

Emily Su

November 30, 2024

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

Table of contents

---

*Code and data are available at: https://github.com/moonsdust/unsolved-murders.

1

1 Introduction

Overview paragraph

This led to us investigate the following question in our paper: what are the differences in homicide case information like the year and city the homicide took place and the victims' perceived characteristics (age, sex, and race) between solved and unsolved homicides in 2 of the largest cities in the United States, New York and Los Angeles, from 2010 to 2017?

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2024).... Our data (The Washington Post 2018).... 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.

Limitation of dataset. - There's only data available from 2010 onwards for both New York and Los Angeles - Not all victims were able to be identified and they were removed from the dataset during data cleaning

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (**?@fig-bills**), from (**palmerpenguins?**).

Talk more about it.

And also planes (**?@fig-planes**). (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.

2.4 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 A.7.

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.

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

We run the model in R (R Core Team 2024) using the `rstanarm` package of (**rstanarm?**). We use the default priors from `rstanarm`.

3.2 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

4.1 Differences in Homicide Case Information Between Solved and Unsolved Cases in New York and Los Angeles (2010 to 2017)

4.1.1 Date (Month and Year)
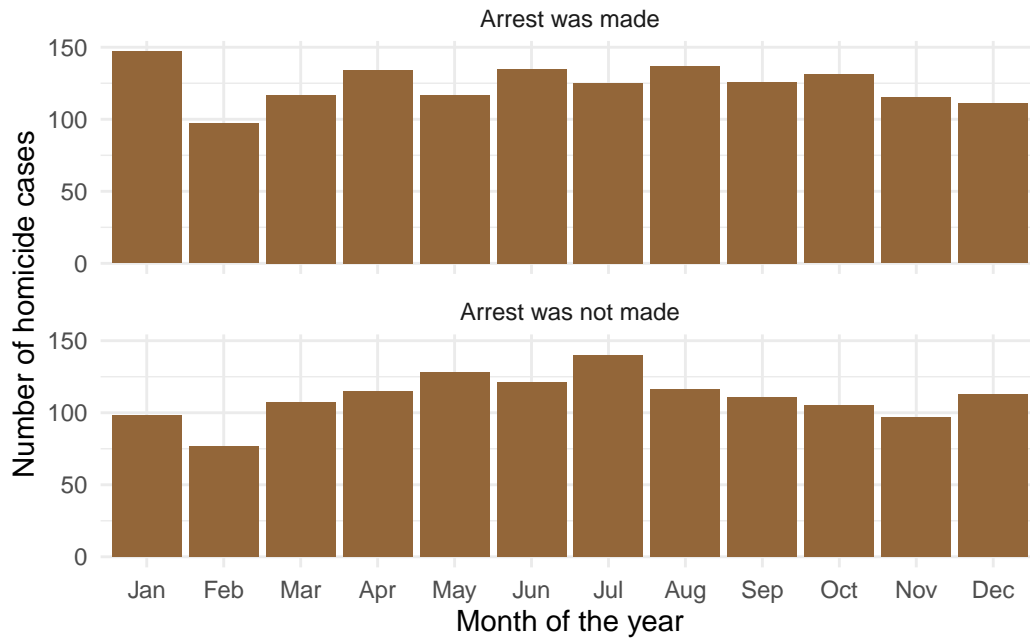
Figure 1: Number of solved and unsolved homicides across the 12 months of a year in Los Angeles and New York (2010 to 2017)

4.1.2 City

Table 1: Proportion and number of solved and unsolved homicides in Los Angeles and New York (2010 to 2017)

City	Status of the homicide case	Number of cases	Proportion of cases
Los Angeles	Arrest was made	1109	0.51
Los Angeles	Arrest was not made	1087	0.49
New York	Arrest was made	383	0.61
New York	Arrest was not made	241	0.39
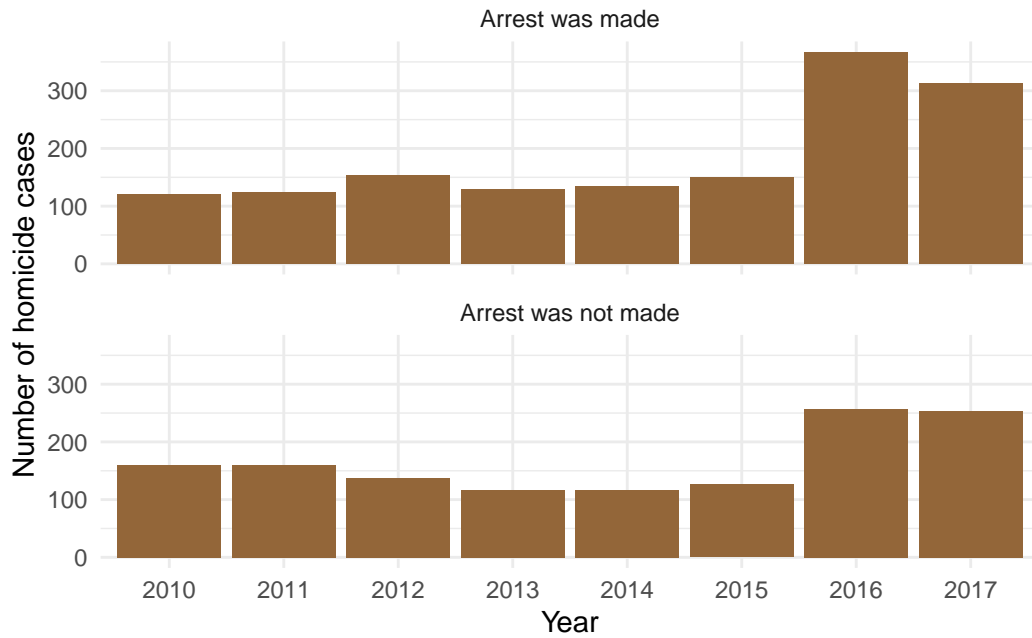
Figure 2: Number of solved and unsolved homicides from 2010 to 2017 in Los Angeles and New York

4.1.3 Disposition

Table 2: Disposition of homicide cases in New York and Los Angeles (2010 to 2017)

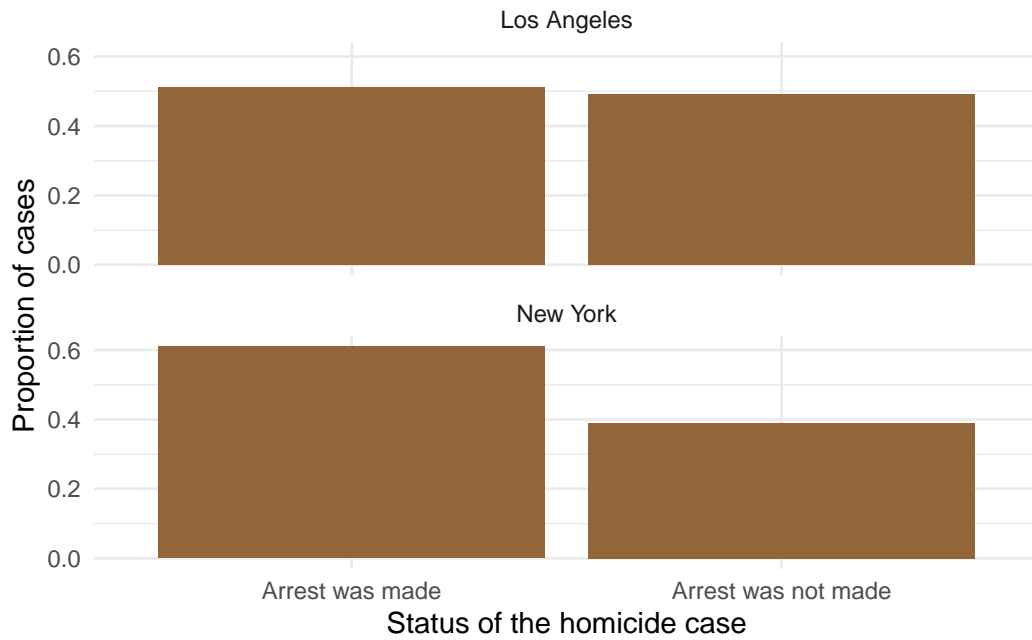| City	Disposition of the homicide case	Number of cases
Los Angeles	Closed by arrest	1109
Los Angeles	Open/No arrest	1087
New York	Closed by arrest	383
New York	Closed without arrest	17
New York	Open/No arrest	224

Figure 3: Proportion of solved and unsolved homicides in Los Angeles and New York (2010 to 2017)
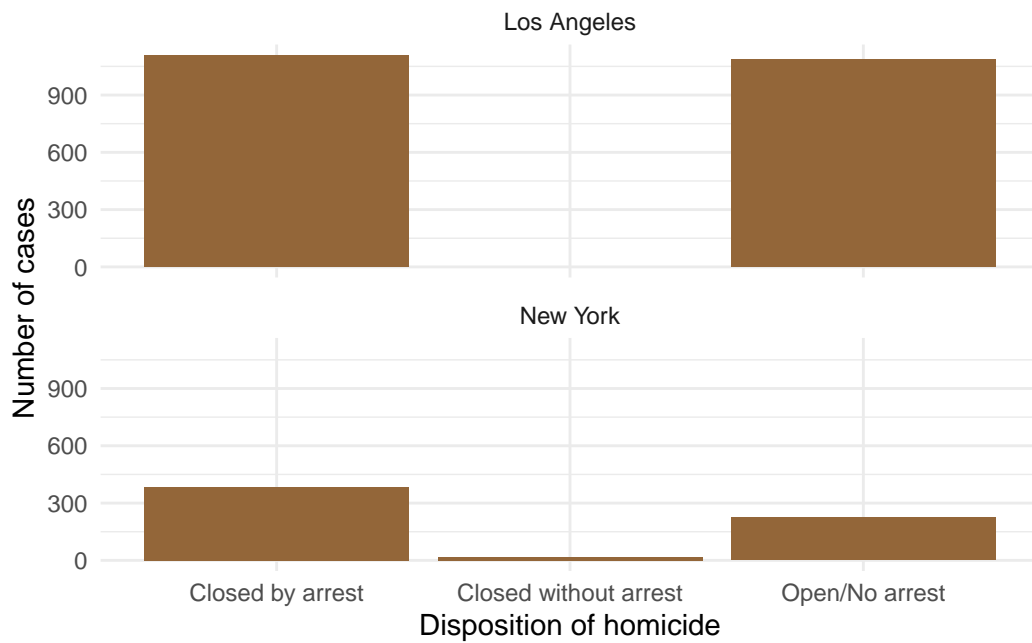
Figure 4: Disposition of homicide cases in New York and Los Angeles (2010 to 2017)

4.1.4 Victim's Age

Figure 5: Distribution of victim's age in solved and unsolved homicides in New York and Los Angeles (2010 to 2017)

4.1.5 Victim's Sex

Table 3: Proportion and number of homicide cases per sex in New York and Los Angeles (2010 to 2017)

Victim's sex	Status of the homicide case	Number of cases	Proportion of cases
Female	Arrest was made	269	0.67
Female	Arrest was not made	135	0.33
Male	Arrest was made	1223	0.51
Male	Arrest was not made	1193	0.49

4.1.6 Victim's Race

Figure 6: Proportion of homicide cases per sex in New York and Los Angeles (2010 to 2017)
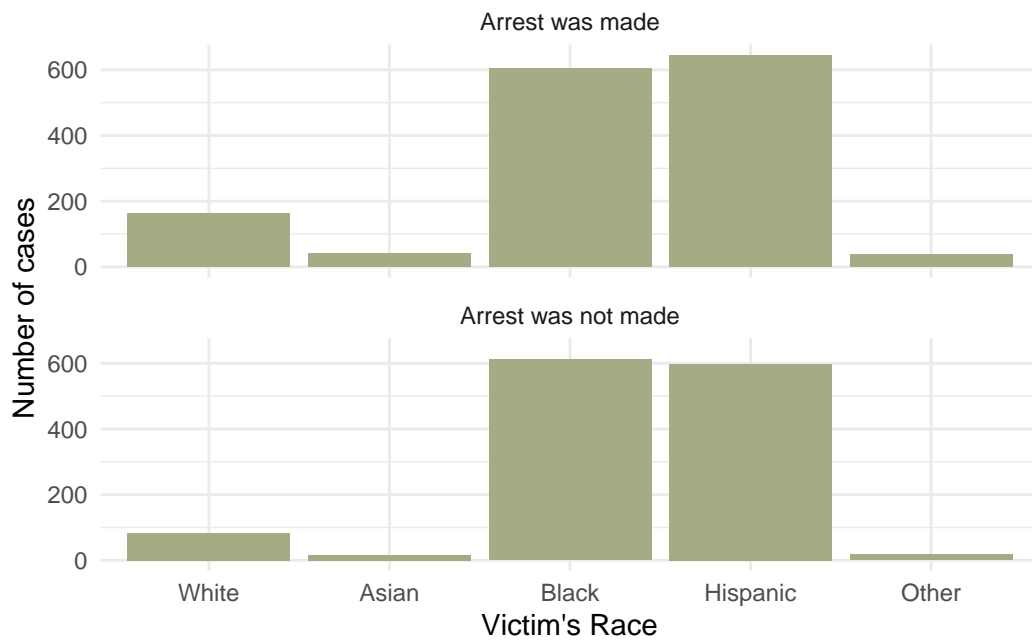
Figure 7: Number of homicide cases per race in New York and Los Angeles (2010 to 2017)

9

Table 4: Number of homicide cases per sex in New York and Los Angeles (2010 to 2017)

Victim's race	Status of the homicide case	Number of cases
White	Arrest was made	164
White	Arrest was not made	84
Asian	Arrest was made	41
Asian	Arrest was not made	17
Black	Arrest was made	604
Black	Arrest was not made	612
Hispanic	Arrest was made	645
Hispanic	Arrest was not made	597
Other	Arrest was made	38
Other	Arrest was not made	18

4.2 Model Results

Figure 8: The credible intervals (line) for coefficient estimates (dot) of predictor variables for homicides that go unsolved from 2010 to 2017.

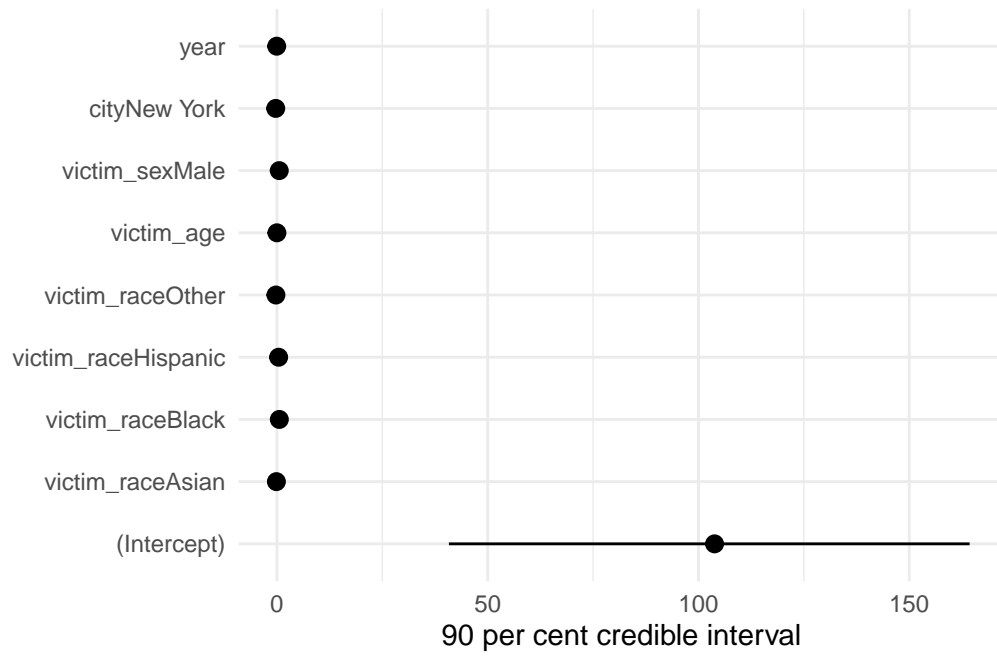Table 5: Relationship between a homicide being unsolved from 2010 to 2017 with the city and year a victim is found in/on and the race, age, and sex of a victim. Mean absolute deviation (MAD) values are in parenthesis.

	Unsolved homicides (2010 to 2017)
(Intercept)	103.817
	(36.659)
victim_raceAsian	−0.125
	(0.325)
victim_raceBlack	0.570
	(0.154)
victim_raceHispanic	0.406
	(0.154)
victim_raceOther	−0.239
	(0.317)
victim_age	−0.004
	(0.003)
victim_sexMale	0.548
	(0.116)
cityNew York	−0.282
	(0.111)
year	−0.052
	(0.018)
Num.Obs.	2820
R2	0.035
Log.Lik.	−1902.792
ELPD	−1912.0
ELPD s.e.	10.1
LOOIC	3824.0
LOOIC s.e.	20.1
WAIC	3824.0
RMSE	0.49

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.

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 Areas of improvement and next steps

Weaknesses and next steps should also be included.

A Appendix

A.1 Note on Reproducing

In order to reproduce the results in the paper, first run the 00-install_packages.R in the scripts folder located in this paper's GitHub repository. Then run the other scripts based on the number at the beginning of the script name.

A.2 Acknowledgments

We would like to thank Alexander (2023) for providing assistance with the R code used to produce the tables and graphs in this paper.

A.3 Code styling

Code written in the scripts was checked and styled with lintr (Hester et al. 2024) and styler (Müller and Walthert 2024).

A.4 Additional Tables

Table 6: Number of solved and unsolved homicides across the 12 months of a year in Los Angeles and New York (2010 to 2017)

Status of the homicide case	Month	Number of cases in the month
Arrest was made	Jan	147
Arrest was made	Feb	97
Arrest was made	Mar	117
Arrest was made	Apr	134
Arrest was made	May	117
Arrest was made	Jun	135
Arrest was made	Jul	125
Arrest was made	Aug	137
Arrest was made	Sep	126
Arrest was made	Oct	131
Arrest was made	Nov	115
Arrest was made	Dec	111
Arrest was not made	Jan	98
Arrest was not made	Feb	77
Arrest was not made	Mar	107
Arrest was not made	Apr	115

Table 6: Number of solved and unsolved homicides across the 12 months of a year in Los Angeles and New York (2010 to 2017)

Status of the homicide case	Month	Number of cases in the month
Arrest was not made	May	128
Arrest was not made	Jun	121
Arrest was not made	Jul	140
Arrest was not made	Aug	116
Arrest was not made	Sep	111
Arrest was not made	Oct	105
Arrest was not made	Nov	97
Arrest was not made	Dec	113

Table 7: Number of solved and unsolved homicides from 2010 to 2017 in Los Angeles and New York

Status of the homicide case	Year	Number of cases in the year
Arrest was made	2010	121
Arrest was made	2011	124
Arrest was made	2012	154
Arrest was made	2013	129
Arrest was made	2014	134
Arrest was made	2015	150
Arrest was made	2016	367
Arrest was made	2017	313
Arrest was not made	2010	160
Arrest was not made	2011	160
Arrest was not made	2012	137
Arrest was not made	2013	117
Arrest was not made	2014	117
Arrest was not made	2015	126
Arrest was not made	2016	257
Arrest was not made	2017	254

A.5 Idealized Survey and Methodology

- Link to literature

A.5.1 Idealized Survey Objectives

A.5.2 Sampling Approach

A.5.3 Respondent Recruitment

A.5.4 Data Validation

A.5.5 Idealized Survey Design

A.5.6 Link to Idealized Survey

- Using Google Forms

A.5.7 Limitations

A.5.8 Idealized Survey Questions

- Should have an introductory section and include details of a contact person
- Question type should be varied and appropriate.
- Have a final section that thank the respondents

A.6 Overview and Evaluation of The Washington Post's Dataset

- TODO: Make sure to link evaluation to literature

A.6.1 Overview

A.6.2 Sampling Approach

- what is the population, frame, and sample;
- how is the sample recruited;
- what sampling approach is taken, and what are some of the trade-offs of this;
- how is non-response handled;
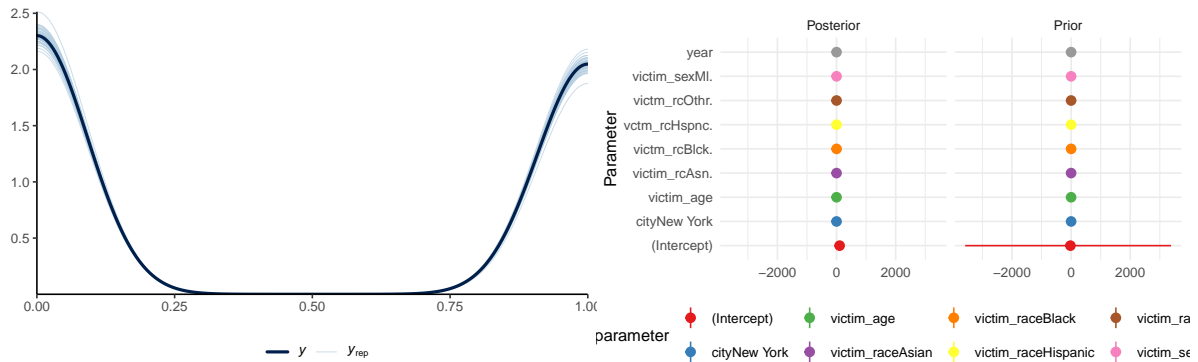
A.6.3 Strengths and limitations

- what is good and bad about the sampling.

A.7 Model details

A.7.1 Posterior predictive check

In Figure 9a we implement a posterior predictive check. This shows...

In Figure 9b we compare the posterior with the prior. This shows...

(a) Posterior prediction check   (b) Comparing the posterior with the prior

Figure 9: Examining how the model fits, and is affected by, the data

A.7.2 Diagnostics

Figure 10a is a trace plot. It shows... This suggests...

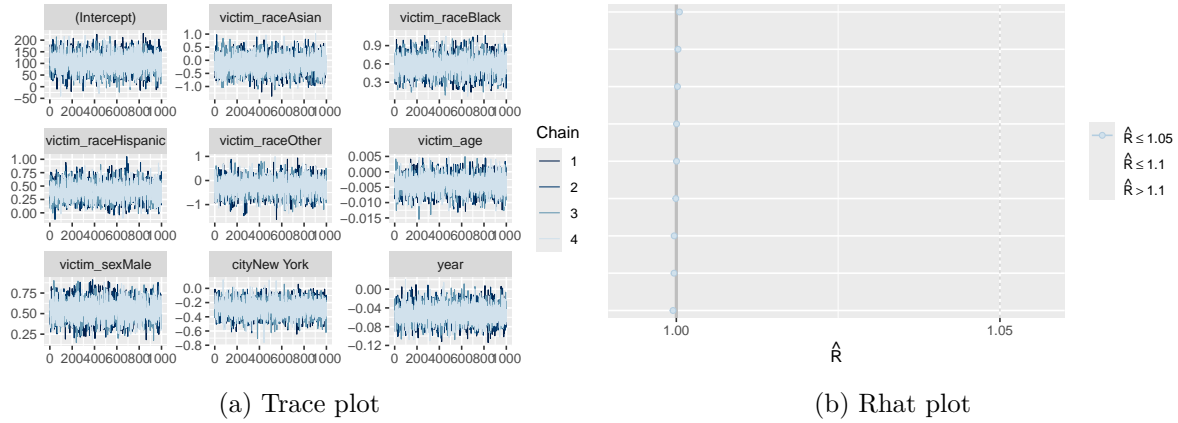Figure 10b is a Rhat plot. It shows... This suggests...

(a) Trace plot

(b) Rhat plot

Figure 10: Checking the convergence of the MCMC algorithm

References

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

Hester, Jim, Florent Angly, Russ Hyde, Michael Chirico, Kun Ren, Alexander Rosenstock, and Indrajeet Patil. 2024. *Lintr: A 'Linter' for r Code.* https://CRAN.R-project.org/packag e=lintr.

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code.* https://CRAN.R-project.org/package=styler.

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

The Washington Post. 2018. *Unsolved Homicide Database.* https://github.com/washingtonp ost/data-homicides.