

My title*

My subtitle if needed

First author

Another author

October 29, 2024

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

1 Introduction

Overview paragraph

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

R

2.2 Raw data

raw	data	52 variables	15891	variables	url,	varibales	variables	description
		appendix						

*Code and data are available at: https://github.com/RohanAlexander/starter_folder.

Table 1: Main column descriptions of the raw data

Variable	Description
poll_id	Unique identifier for each poll conducted.
numeric_grade	A numeric rating given to the pollster to indicate their quality or reliability .
pollscore	A numeric value representing the score or reliability of the pollster in question .
methodology	The method used to conduct the poll .
transparency_score	A score reflecting the pollster’s transparency about their methodology .
start_date	The date the poll began .
end_date	The date the poll ended .
sample_size	The total number of respondents participating in the poll .
population	The abbreviated description of the respondent group, typically indicating their voting status .
hypothetical	Indicates whether the poll is about a hypothetical match-up.
pct	The percentage of the vote or support that the candidate received in the poll .

variable na

Table 2: Number of Missing Values and Percentages for Variables

Variable	Missing_Values	Percentage_Missing
poll_id	0	0.00
pollster_id	0	0.00
sponsor_ids	8236	51.83
pollster_rating_id	0	0.00
numeric_grade	1895	11.92
pollscore	1881	11.84
methodology	993	6.25
transparency_score	3209	20.19
state	7552	47.52
start_date	0	0.00
end_date	0	0.00
sponsor_candidate_id	15562	97.93
sponsor_candidate_party	15562	97.93
question_id	0	0.00
sample_size	139	0.87
population	0	0.00
tracking	14363	90.38

Table 2: Number of Missing Values and Percentages for Variables

Variable	Missing_Values	Percentage_Missing
created_at	0	0.00
notes	15621	98.30
source	15690	98.74
internal	13427	84.49
partisan	14513	91.33
race_id	0	0.00
cycle	0	0.00
office_type	0	0.00
election_date	0	0.00
stage	0	0.00
nationwide_batch	0	0.00
ranked_choice_reallocated	0	0.00
ranked_choice_round	15872	99.88
hypothetical	0	0.00
party	0	0.00
answer	0	0.00
pct	0	0.00

numerical variable

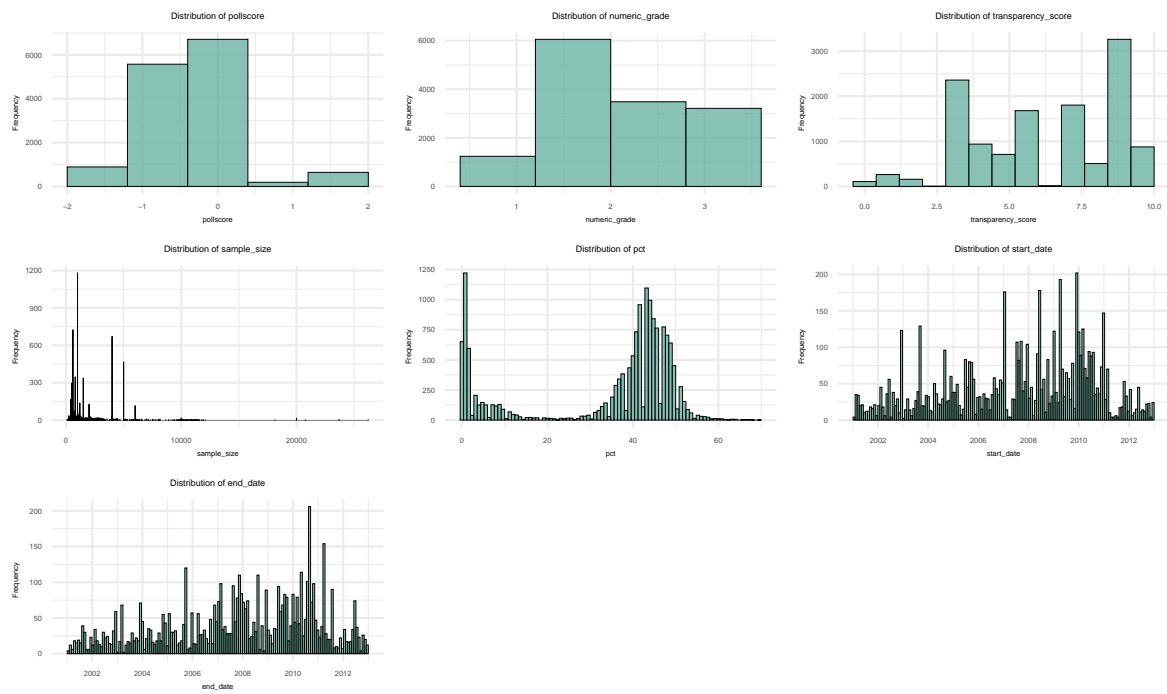


Figure 1: Distribution of numerical varibales

categorical

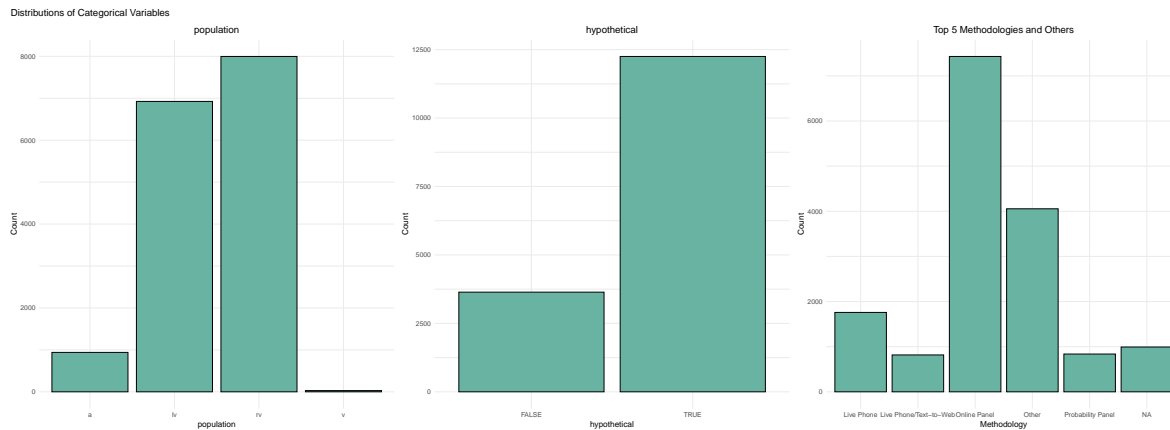


Figure 2: Distribution of categorical variables

appendix

2.3 Measurement

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

2.4 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 Horst, Hill, and Gorman (2020).

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.

2.5 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](#).

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.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\sigma \sim \text{Exponential}(1) \tag{6}$$

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`.

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 [?@tbl-modelresults](#).

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should 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 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

raw data variabile

Table 3: Remaining column descriptions of the raw data

Table 3: Descriptions of Remaining Variables

Variable	Description
pollster_id	Unique identifier for the polling organization conducting the poll.
sponsor_ids	Unique identifier(s) for the sponsor(s) of the poll, typically organizations that fund the poll.
pollster_rating_id	Unique identifier for the pollster’s rating within a rating system.
state	The U.S. state where the poll was conducted or focused, if applicable.
sponsor_candidate_id	Unique identifier for the candidate sponsored by the sponsoring organization (if applicable).
sponsor_candidate_party	The political party of the candidate sponsored by the sponsor (if applicable).
question_id	Unique identifier for the question asked in the poll.
tracking	Indicates whether the poll is part of a tracking series .
created_at	The timestamp when the poll data was created or entered into the system .
notes	Any additional notes or comments related to the poll.
source	The source from where the poll data was derived.
internal	Indicates whether the poll is conducted internally by a campaign or organization.
partisan	Indicates whether the poll has partisan sponsorship or is conducted by a partisan organization.
race_id	A unique identifier for the political race being polled .
cycle	The election cycle in which the poll is conducted .
office_type	The type of political office being polled .
election_date	The date of the election the poll is related to .
stage	The stage of the election being polled .
nationwide_batch	Indicates whether the poll is part of a nationwide batch.
ranked_choice_reallocations	Indicates if ranked-choice voting reallocations have been applied in the results.
ranked_choice_round	The round of ranked-choice voting, if applicable.
party	The political party of the candidate in the poll .
answer	The response or answer choice given in the poll .

raw data

```
#to be done
```

B Model details

B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

References

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