My title*

My subtitle if needed

First author Another author

October 30, 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

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.

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

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 (Figure 1), from Horst, Hill, and Gorman (2020).

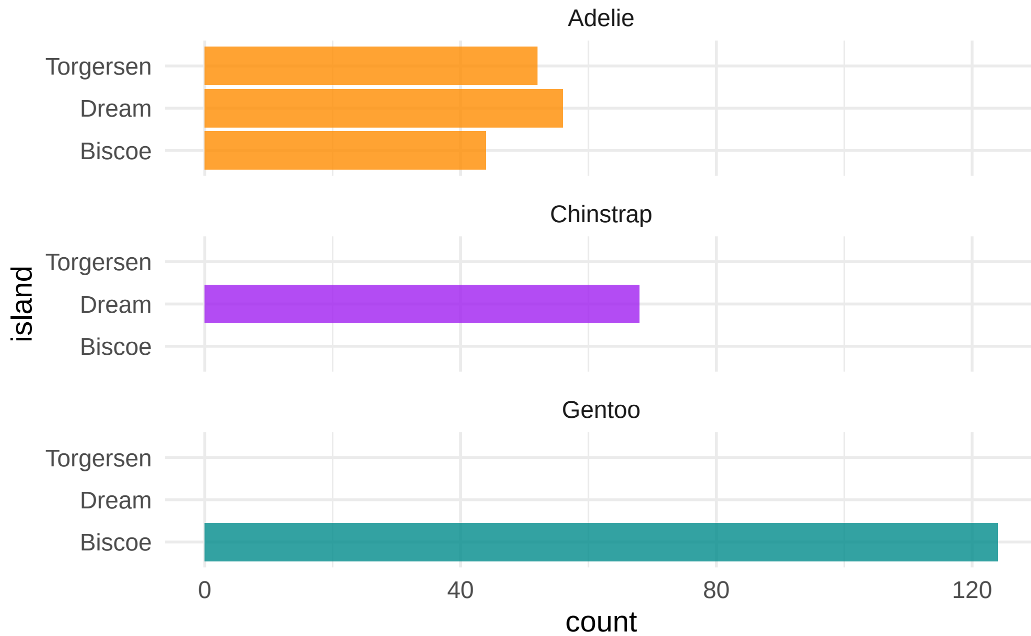

Figure 1: Bills of penguins

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.

```
library(rstanarm)
```

```
Loading required package: Rcpp

This is rstanarm version 2.32.1

- See https://mc-stan.org/rstanarm/articles/priors for changes to default priors!

- Default priors may change, so it's safest to specify priors, even if equivalent to the defa

- For execution on a local, multicore CPU with excess RAM we recommend calling

  options(mc.cores = parallel::detectCores())
```

```
model_date <-
  readRDS(file = here::here("models/model_date.rds"))
model_date_pollster <-
   readRDS(file = here::here("models/model_date_pollster.rds"))
```

3 Model

```
modelsummary(models = list("Model 1" = model_date, "Model 2" = model_date_pollster))
```

3.0.1 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

	Model 1	Model 2
(Intercept)	−1064.538	−1135.824
	(66.474)	(64.877)
end_date	0.057	0.061
	(0.003)	(0.003)
pollsterBeacon/Shaw		0.043
		(4.601)
pollsterCNN/SSRS		−0.518
		(4.635)
pollsterEchelon Insights		4.605
		(4.309)
pollsterEmerson		5.435
		(4.322)
pollsterIpsos		−5.180
		(4.555)
pollsterMarist		9.720
		(6.325)
pollsterMarquette Law School		−0.453
		(4.370)
pollsterMonmouth		3.046
		(4.488)
pollsterQuinnipiac		3.369
		(4.500)
pollsterSelzer		−14.457
		(5.867)
pollsterSiena		4.110
		(5.554)
pollsterSiena/NYT		2.495
		(5.199)
pollsterSuffolk		−7.449
		(4.599)
pollsterSurveyUSA		8.794
		(5.866)
pollsterU. North Florida		−9.385
		(5.854)
pollsterUniversity of Massachusetts Lowell/YouGov		−12.803
		(7.193)
pollsterYouGov		2.066
		(4.212)
pollsterYouGov Blue		−5.799
		(6.333)

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 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

Polling Methodology for CBS News/YouGov Survey (October 11-16, 2024)

1. Population, Frame, and Sample This CBS News/YouGov survey took place from October 11-16, 2024, with 1,439 registered voters in Arizona. The survey focused on registered voters in Arizona, and the sample was weighted to match key demographics like gender, age, race, and education. The weights were based on data from the U.S. Census American Community Survey, the U.S. Census Current Population Survey, and voter turnout data from the 2020 Presidential election.

2. Sample Recruitment The recruitment process focused on including respondents representative of Arizona's registered voter population by adjusting for demographic factors such as age, race, gender, and education. The sample was recruited primarily from various online panels, which included a mixture of respondents across demographic lines, to ensure a representative sample: 1,152 respondents were selected from YouGov's online panel. 212 respondents from Pure Spectrum's panel. 49 respondents from Dynata. 17 respondents from Cint's panel. 9 respondents from ROI Rocket's panel.

Surveys were conducted in both English and Spanish to account for language preferences among respondents. The weights applied to the data ranged from 0.1 to 5.0, with a mean of 1 and a standard deviation of 0.8, ensuring that the sample was representative of Arizona's voting population.

3. Sampling Approach and Trade-offs Sampling Approach: This survey employed stratified random sampling and applied post-survey weighting to ensure accurate representation of key demographic groups. Stratified random sampling is a technique that divides the overall population into several subgroups (or strata) based on specific attributes such as age, gender, race, and education level. Random samples are then drawn from each subgroup. The advantage of this method is that it ensures adequate representation across each stratum, preventing certain groups from being underrepresented or overlooked in the sample. After the survey was completed, weighting was applied to further adjust the sample to match the demographic distribution of registered voters in Arizona. This process involved adjusting the weights of individuals in the sample to better reflect the true composition of the overall voter population. This allows for a more accurate capture of how different demographic factors (such as gender, age, race, and education) influence voting behavior, thereby improving the external validity of the results. By using stratified random sampling and weighting, the researchers were able to minimize sampling bias and increase the accuracy of predicting voter tendencies. Trade-offs: One of the main limitations of this sampling method is its reliance on online panels, which may exclude individuals without internet access, thus introducing selection bias. While online surveys are cost-effective and convenient for collecting large samples, this reliance may result in certain groups (such as older individuals, low-income households, or voters living in remote areas) being left out due to lack of internet access. This means that some groups

might be underrepresented in the sample, which can affect the overall representativeness of the results. Although the weighting process can help adjust the sample to better reflect demographic differences, some errors are still difficult to completely eliminate. For example, when filling out surveys, respondents might overstate or understate their voting intentions due to social pressure or personal emotions—this is known as self-reporting bias. Even after weighting adjustments, such biases may persist and impact the accuracy of the final survey results. Therefore, while weighting can improve the representativeness of the results to a certain extent, systematic biases like non-response bias or selection bias may still leave traces in the final outcomes. Researchers need to interpret these potential errors with caution.

4. Regression Model A regression model was used to estimate each respondent's likelihood of voting. This model combined self-reported voting intentions with demographic and historical voting data, such as: Age, gender, race, and education. Voting history from past elections. The regression model allowed the survey to distinguish "likely voters" from the broader pool of registered voters, improving the accuracy of the predictions. By analyzing both individual and aggregate data, the model offered a more reliable estimate of actual voter turnout, thus increasing the precision of the survey's results.

5. Handling Non-response In surveys, non-response can cause bias because these people might have different voting behaviors or opinions. To fix this potential bias, researchers use weighting to adjust the sample data. Specifically, they assign different weights to respondents based on key demographic factors like gender, age, race, and education. The main goal of this weighting process is to make sure that even if some voters didn't respond, the final sample still accurately represents the overall population of registered voters in Arizona. This adjustment helps make the sample more representative and reduces the bias caused by non-response, improving the reliability and accuracy of the survey results. Besides that, weighting helps balance the proportion of different groups in the sample, ensuring that certain groups (like those with less access to the internet) are properly reflected in the results. In the end, this process helps make sure the survey results are more valid and can be applied more effectively to predict real voter behavior.

6. Strengths and Weaknesses of the Questionnaire Strengths: This survey effectively covered the key issues that Arizona voters care about the most, such as the economy, immigration, abortion, and the state of democracy. By combining demographic factors and historical voting data, the reliability of the results was improved, making it more accurate in reflecting the voting preferences of different groups. Weaknesses: Since the survey relies on self-reported voting intentions, this might introduce some bias, as respondents could overestimate or underestimate their likelihood of voting. Also, because it depends on online panels, voters without internet access may have been excluded, which could affect the external validity of the results. The margin of error is ± 3.3 points, showing that there's still some uncertainty in the findings.

7. Margin of Error The margin of error for this survey is ± 3.3 points, within a 95% confidence interval. The formula to calculate the margin of error is: $\hat{p} \pm 100 \times \sqrt{((1+CV^2)/n)}$

Where CV is the coefficient of variation of the sample weights, and (n) is the sample size. This formula calculates the sampling error, meaning that 95% of the sample results should fall within this range. It's important to note that this margin doesn't account for non-sampling errors, such as biases from panel selection or respondent behavior.

A Additional data details

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

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

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/.

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