

The impact of physical characteristics on Penguin flight*

Hoang Viet Nguyen

September 24, 2024

This study investigates the relationship between width, length, weight and flying time in penguins. Using a dataset that includes variables such as flying time, width, and length. We found that increased weight is associated lower flying time.

1 Introduction

In this study, we investigate the relationship between various physical characteristics of penguins, particularly focusing on weight and flying time as the main variables. Understanding how these variables interact is crucial in the fields of biology and ecology, as they provide insights into the adaptive traits of penguins and their capabilities in flight.

Weight is a critical factor that influences the physical performance of penguins. Heavier penguins may face limitations in their flying capabilities due to increased energy requirements for flight. Consequently, flying time can vary significantly based on weight, as heavier individuals may struggle to remain airborne for extended periods.

In addition to weight, we explore the effects of wing dimensions, including wing width and length, on flying time. By analyzing a dataset that incorporates these dimensions along with weight, we aim to explain the dynamics of flight in penguins and contribute to the understanding of how physical characteristics affect their ability to fly.

The remainder of this paper is structured as follows. In Section 2, we describe the data used in our analysis, followed by the modeling approach in Section 3. Section 4 presents the results, and Section 5 discusses the implications of our findings. # Data {#sec-data}

Some of our data is of penguins (Figure 1), from Horst, Hill, and Gorman (2020) which provides insights into the characteristics of different penguins species. The dataset includes

*Code and data are available at: [LINK](#).

measurements such as length and width which serve as useful indicators of various biological attributes.

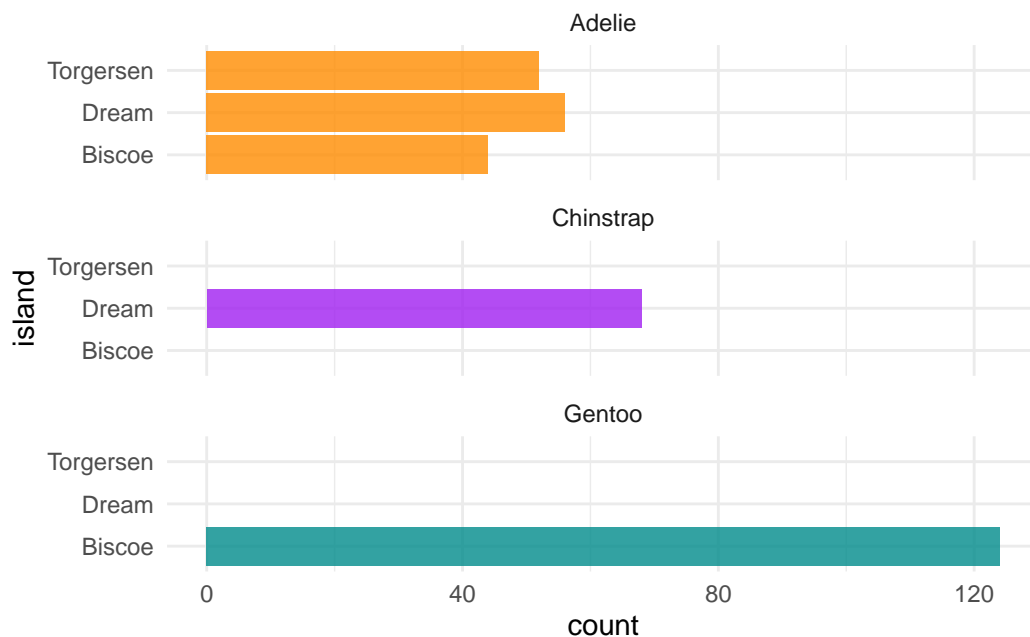


Figure 1: Bills of penguins

In addition to penguin data, a scatter plot below shows the relationship between width and length of the penguins which allows us to analyze furthermore.

This scatter plot below takes two variables which are weight and flying time of the penguins. By observing the distribution we can notice the relationship between weight and flying time. Therefore, observing how likely the weight would affect the flying time.

```
simulated_data <- read_csv(here::here("/Users/nguyenviet/Documents/STA304 - paper 1/data/raw/"))
```

```
Rows: 100 Columns: 2
```

```
-- Column specification -----
```

```
Delimiter: ","
```

```
dbl (2): flying_time, weight
```

```
i Use `spec()` to retrieve the full column specification for this data.
```

```
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

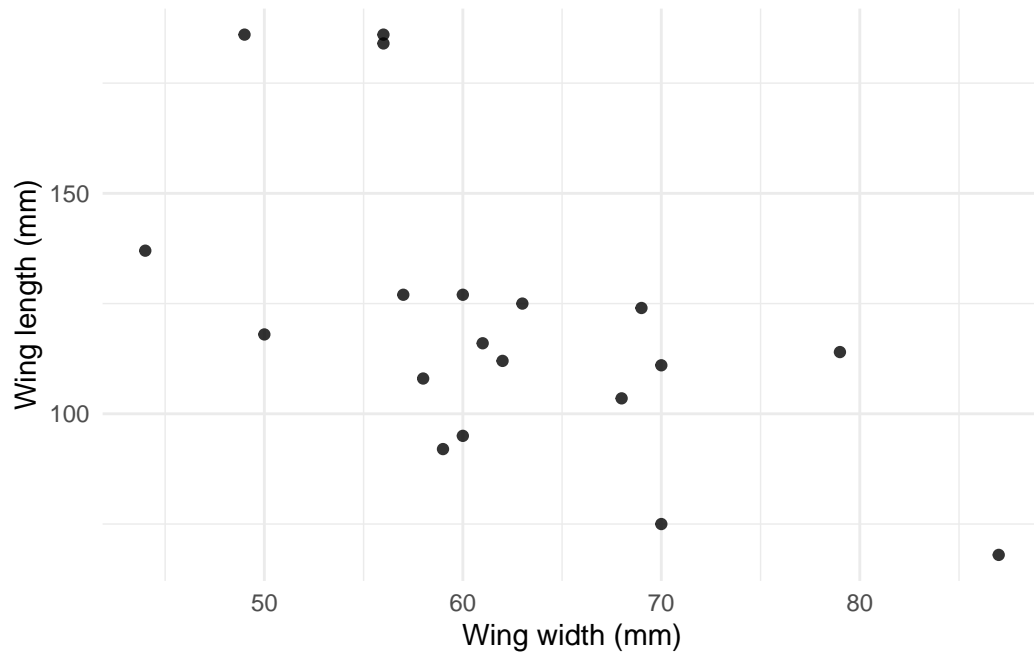
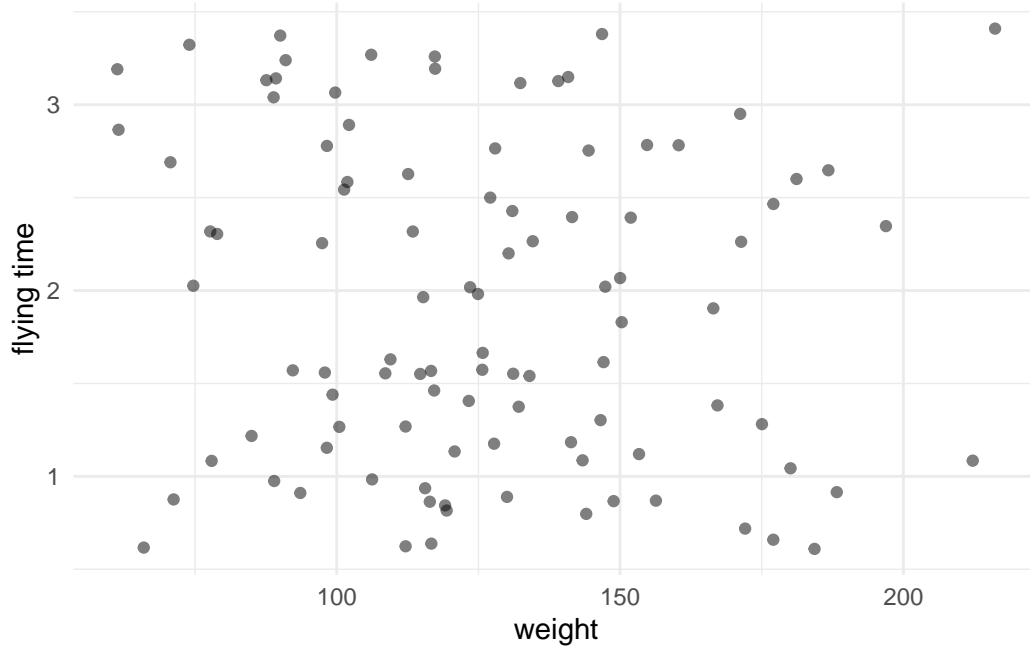


Figure 2: Relationship between wing length and width

```
simulated_data |>
  ggplot(aes(x = weight, y = flying_time)) +
  geom_point(alpha = 0.5) +
  theme_minimal() +
  labs(x = "weight", y = "flying time")
```



2 Model

The goal of our modelling strategy is twofold. Firstly, we aim to investigate the relationship between weight and flying time in penguins, contributing to a deeper understanding of how physical characteristics impact their flight dynamics. Secondly, we have developed a simulation model to create synthetic data that allows us to explore these relationships under controlled conditions.

To simulate our data set, I generated 100 samples for two key variables: flying time and weight. The flying time was derived from a uniform distribution on the existing flying time data in our analysis data, ensuring that the simulated values lie within the range of observed values.

For the weight variable, I generated values from a normal distribution, utilizing the mean and standard deviation of the length variable in the analysis data. This decision was made based on the assumption that weight might have a distribution pattern that is reflective of the lengths of the penguins in our data set.

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

2.1 Model set-up

We aim to establish a statistical model that explains the variation in flying time based on the simulated weight of penguins. The relationship can be formally described as:

$$y_i = \beta_0 + \beta_1 \cdot \text{weight}_i + \epsilon_i \quad (1)$$

$$\beta_0 \text{ is the intercept,} \quad (2)$$

$$\beta_1 \text{ represents the effect of weight on flying time,} \quad (3)$$

$$\epsilon_i \text{ is the error term, assumed to be normally distributed.} \quad (4)$$

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.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (5)$$

$$\mu_i = \alpha + \beta_i + \gamma_i \quad (6)$$

$$\alpha \sim \text{Normal}(0, 2.5) \quad (7)$$

$$\beta \sim \text{Normal}(0, 2.5) \quad (8)$$

$$\gamma \sim \text{Normal}(0, 2.5) \quad (9)$$

$$\sigma \sim \text{Exponential}(1) \quad (10)$$

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`.

2.1.1 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 .

3 Results

Our results are summarized in Table 1.

```
lm(flying_time ~ weight, data=simulated_data)
```

Table 1: 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

```
Call:
lm(formula = flying_time ~ weight, data = simulated_data)
```

```
Coefficients:
(Intercept)      weight
  2.276291    -0.002813
```

4 Discussion

4.1 First discussion point

The analysis of the simulated penguin data, particularly focusing on the relationship between weight and flying time, has led to several key insights. The model suggests a significant association between penguin weight and their flying time, where increases in weight tend to reduce the time spent. This findings aligns with intuitive biological expectations, as heavier penguins likely face greater maintaining prolonged flight compared to their lighter counterparts.

The linear model. which modeled flying time as a function of weight, demonstrated that even small changes in weight can have a measurable impact on flight duration. These results provide a quantitative basis for understanding the constraints on fligth that weight imposes, potentially influencing the energy expenditure and behavior of penguins during flight. ##
Second discussion point

4.2 Third discussion point

4.3 Weaknesses and next steps

The data set used for this analysis is simulated, meaning the results might not reflect actual biological processes in penguins. Although the simulation aimed to mimic realistic relationships between the variables, empirical data would provide stronger evidence for the conclusions drawn. In addition, the variable weight is not an original measurement of the penguins, it was derived by the using the length variable from the analysis data. While weight is a critical factor for understanding the bio mechanics of flight in real-world scenarios, in this case, its simulation introduces a potential source of bias. The weight does not have direct empirical support from real measurements, and any patterns or correlations identified between weight and flying time may be artifacts of the way the data was generated. This could affect the generalizability of the results to real-world penguin flight dynamics.

Additionally, because weight and length are related in the simulated data, the true variability in weight that might exist in real penguin populations is not captured. This limitation needs to

be considered when interpreting the findings, as the strength of the relationship between weight and flying time might be overstated. Future work should focus on collecting data actual weight data to better understand its role in flying time.

Appendix

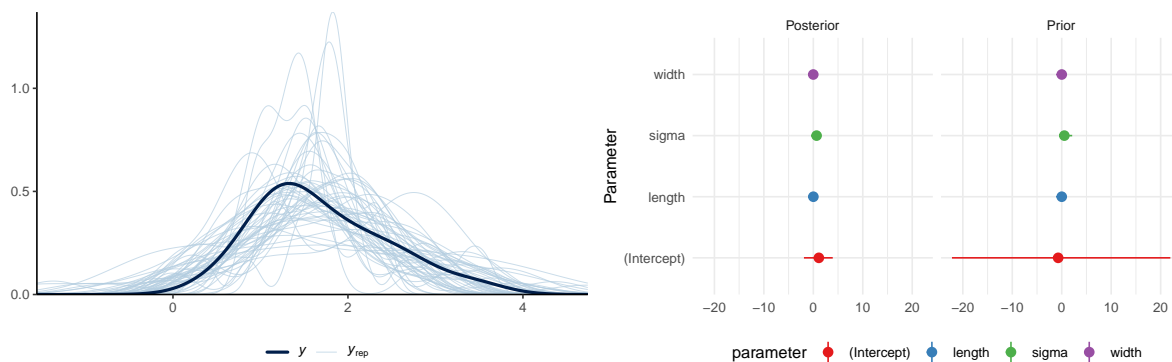
A Additional data details

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...

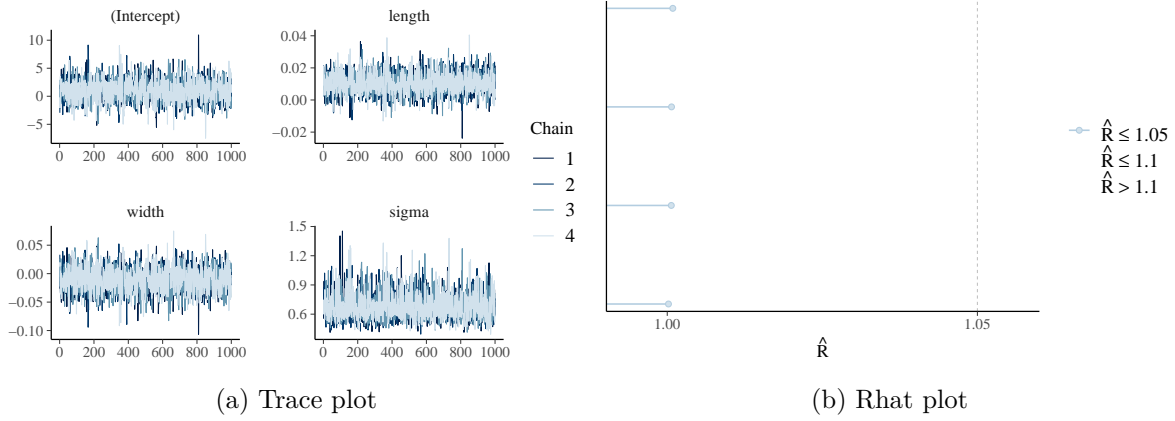


Figure 4: Checking the convergence of the MCMC algorithm

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. *Palmer penguins: 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/>.