Bayesian Insights into Aerial Bombing Strategies: An Ordered Logistic Regression Analysis of WWII Target Prioritization Against Germany*

Yunzhao Li

April 18, 2024

During World War II, the strategic decisions about where and how the Allies bombed German industries played a crucial role in the Allied campaign. This study employs Bayesian ordered logistic regression to methodically analyze how different factors—such as the industry type, the launching country, the bomb load, and the number of aircraft—shaped these decisions. Focusing on the preference for targeting crucial oil refineries, the analysis sheds light on the calculated approach to disrupt Germany's war capabilities effectively. The insights from this research not only clarify the strategic rationale behind the Allies' bombing priorities but also enhance our understanding of their impact on the course of the war.

1 Introduction

World War II reshaped global power dynamics, with the strategic aerial bombing campaigns playing a crucial role in determining the outcome of the conflict. The Allied forces' focused bombardment of German industries was pivotal, aiming to dismantle the economic backbone essential for Germany's war efforts. While considerable research has been conducted on the impact of these bombings, less attention has been given to the systematic selection process behind targeting specific industrial assets. This paper addresses this understudied aspect by exploring the intricate decision-making process that guided the Allied forces in targeting German industries.

The crux of this research lies in a detailed examination of a sampled subset of over 5,000 missions from comprehensive wartime records, representing a strategic selection designed to

*Code and data are available at: https://github.com/yunzhaol/aerial_bomb_priority.git.

analyze the impact of various factors on target prioritization. These factors include the type of industry targeted, the leading country of the mission, the volume of bombs dropped, and the scale of aircraft involvement. Prior literature has often provided broad overviews without delving into the interplay of these specific variables. By applying a Bayesian ordered logistic regression model, this study meticulously quantifies the influence of each factor, offering a clearer picture of the strategic considerations that shaped the Allied bombing strategy.

Our findings indicate a deliberate emphasis on disabling critical oil refineries and other vital industrial infrastructures, which were deemed essential for sustaining the German war machine. The analysis underscores that missions with larger bomb payloads and more aircraft were preferentially deployed against these high-value targets. This targeted approach was strategic, aiming not just at destruction but at crippling Germany's ability to sustain its military operations effectively.

The significance of these insights extends beyond the historical narrative of World War II, offering lessons on the allocation of military resources and strategic target selection that are relevant to modern military strategies. Furthermore, this paper enriches the ongoing academic debate about the strategic efficacy and ethical considerations of aerial bombing in warfare.

The remainder of this paper is structured as follows. Section 2….

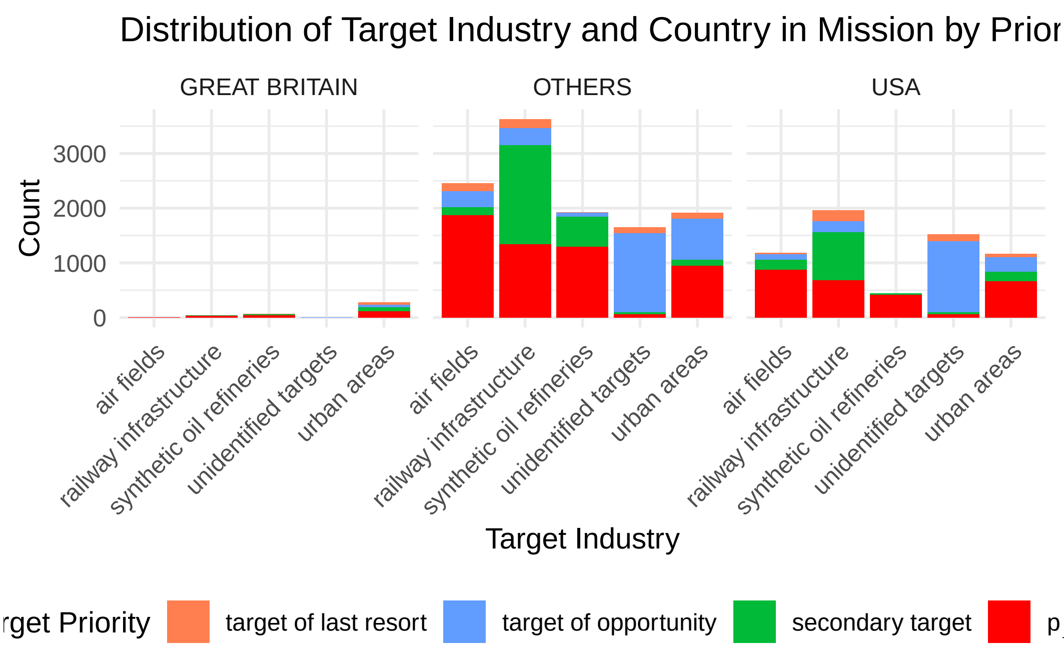
(R Core Team 2023), (Apache Arrow 2021), (Xie 2021), (Zhu 2021), (Arel-Bundock 2021), (Team 2021), (Wickham et al. 2021), (Wickham 2021), (Gelman, Gabry, et al. 2021), (Robinson, Hayes, et al. 2021), (Auguie 2021), (Firke 2021), (Wickham, François, et al. 2021), (Alexander 2023) R Core Team (2023).

2 Data

The data for this study was meticulously curated from declassified military records, comprising numerous raids, their targets, and associated strategic variables.

Table 1: Aerial Bombing missions against Germany on the main target industries in WWII

tgt_priority	tgt_industry	country_mission	bomb_tons	aircraft_attack
secondary target	railway infrastructure	usa	42	17
secondary target	railway infrastructure	usa	20	8
target of last resort	unidentified targets	usa	3	1
target of last resort	unidentified targets	usa	3	1
target of last resort	unidentified targets	usa	3	1
secondary target	railway infrastructure	usa	5	3



{#fig-industry and country by priority}

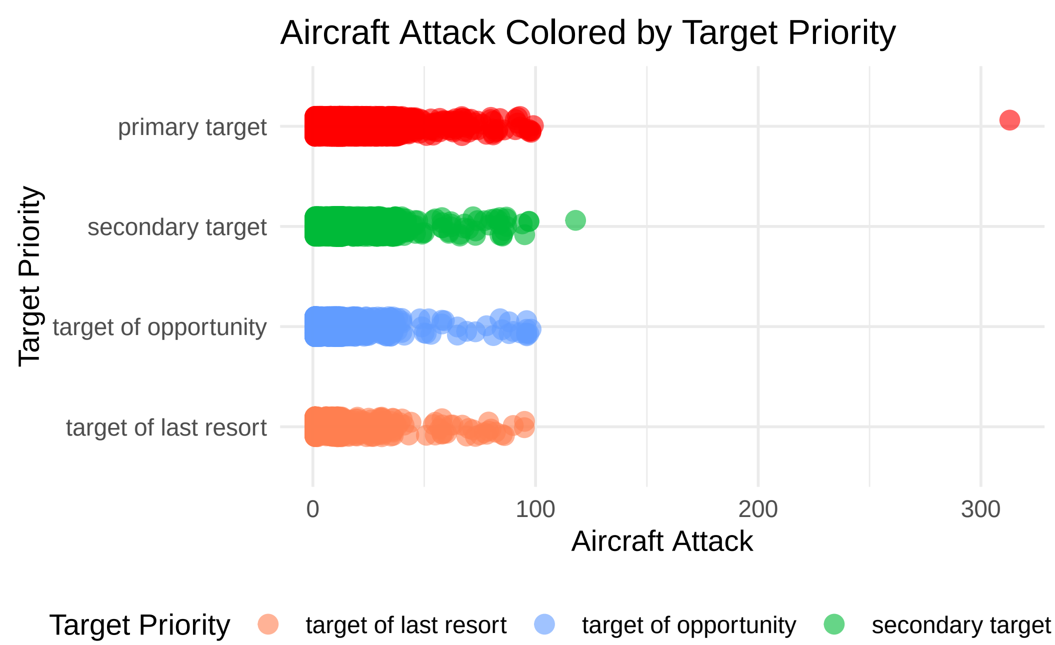


Figure 1: Distribution of target industry and country in mission by Priority

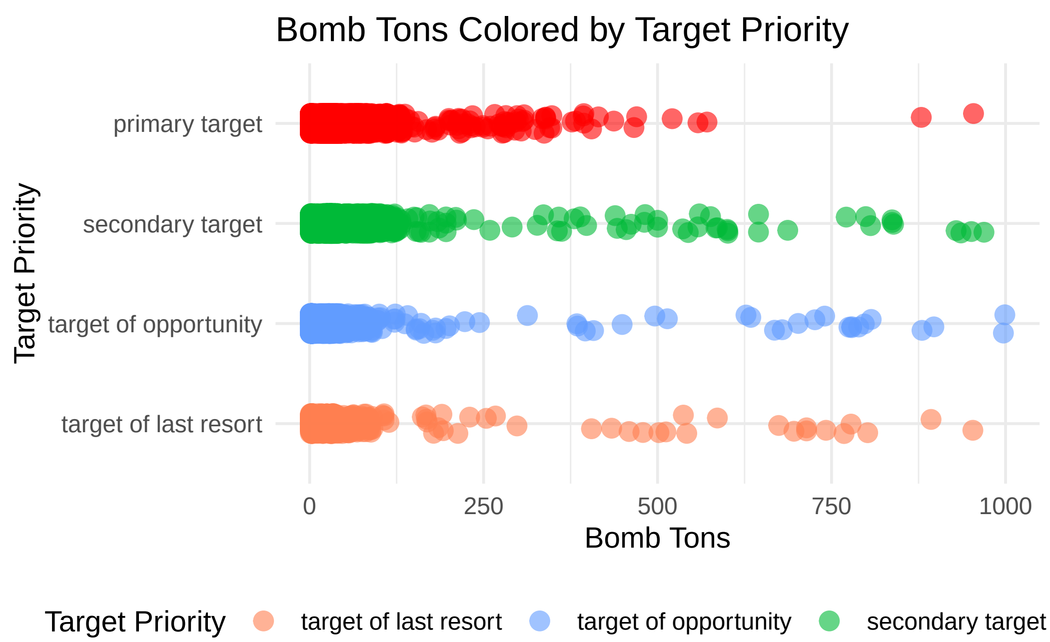


Figure 2: Distribution of target industry and country in mission by Priority

2.1 Variable

Our analysis focuses on the following variables: target priority, target industry, country of the mission's origin, total tons of bombs dropped, and the number of aircraft partaking in the mission.

2.2 Justification

The selected variables are integral for understanding the strategic matrix that guided target selection, enabling a thorough investigation into the strategic landscape of WWII aerial campaigns.

3 Model

The goal of our modelling strategy is twofold. F This study employs a Bayesian ordered logistic regression model to analyze the priorities assigned to different target industries during aerial bombing missions in World War II.

The model integrates several predictors including the type of target industry (`tgt_industry`), the country executing the mission (`country_mission`), the tonnage of bombs dropped (`bomb_tons`), and the number of aircraft involved in the attack (`aircraft_attack`).

The model uses a logistic cumulative link function, draws on a posterior sample size of 4000, and is based on a total of 5000 observations, utilizing the rstanarm package to accommodate the ordinal nature of target priority levels.

Our model statistically infers the relationship between target priority and various strategic factors, providing a probabilistic assessment of their impacts.

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.

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

We run the model in R (R Core Team 2023) using the `rstanarm` package of (**rstanarm?**). 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…

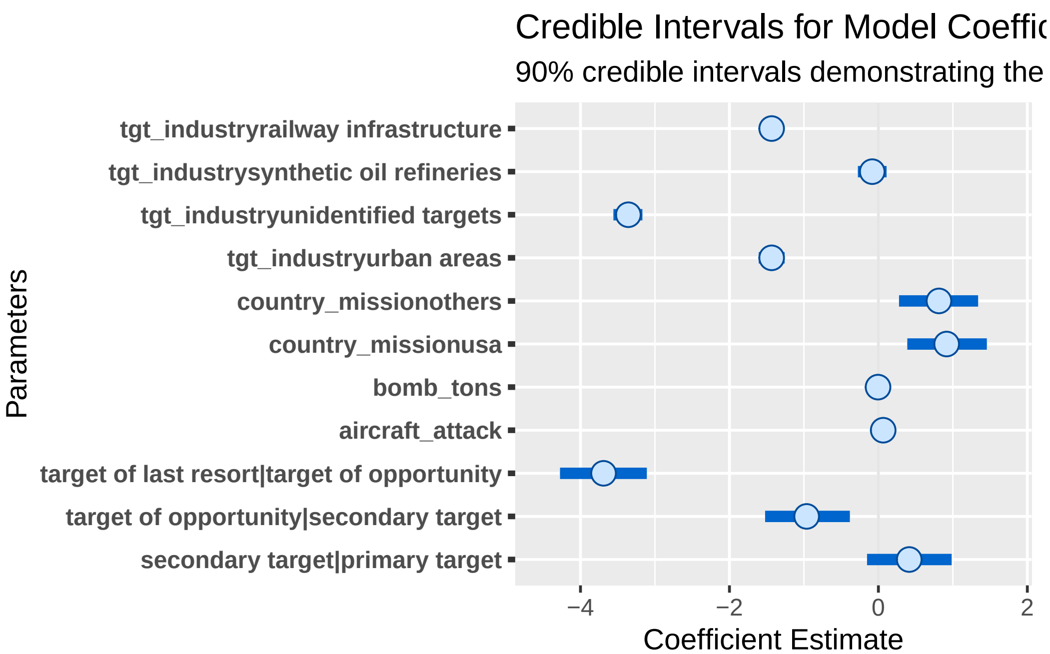
We can use maths by including latex between dollar signs, for instance θ.

Table 2: Explanatory models of flight time based on wing width and wing length

```
Model Info:
 function:     stan_polr
 family:       ordered [logistic]
 formula:      tgt_priority ~ tgt_industry + country_mission + bomb_tons + aircraft_attack
 algorithm:    sampling
 sample:       4000 (posterior sample size)
 priors:       see help('prior_summary')
 observations: 5000

Estimates:
                                            mean   sd   10%   50%   90%
tgt_industryrailway infrastructure         -1.4    0.1 -1.6  -1.4  -1.3
tgt_industrysynthetic oil refineries       -0.1    0.1 -0.2  -0.1   0.1
tgt_industryunidentified targets           -3.4    0.1 -3.5  -3.4  -3.2
tgt_industryurban areas                    -1.4    0.1 -1.6  -1.4  -1.3
country_missionothers                       0.8    0.3  0.4   0.8   1.2
country_missionusa                          0.9    0.3  0.5   0.9   1.3
bomb_tons                                   0.0    0.0  0.0   0.0   0.0
aircraft_attack                             0.1    0.0  0.1   0.1   0.1
target of last resort|target of opportunity -3.7    0.4 -4.1  -3.7  -3.2
target of opportunity|secondary target     -1.0    0.3 -1.4  -1.0  -0.5
secondary target|primary target             0.4    0.3  0.0   0.4   0.9

Fit Diagnostics:
                               mean   sd   10%   50%   90%
mean_PPD:target of last resort 0.0    0.0  0.0   0.0   0.1
mean_PPD:target of opportunity 0.2    0.0  0.2   0.2   0.3
mean_PPD:secondary target      0.2    0.0  0.2   0.2   0.2
mean_PPD:primary target        0.5    0.0  0.5   0.5   0.5

The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics
                                            mcse Rhat n_eff
tgt_industryrailway infrastructure          0.0  1.0  3583
tgt_industrysynthetic oil refineries        0.0  1.0  4392
tgt_industryunidentified targets            0.0  1.0  3608
tgt_industryurban areas                     0.0  1.0  3491
country_missionothers                       0.0  1.0  5105
country_missionusa                          0.0  1.0  4972
bomb_tons                                   0.0  1.0  5818
aircraft_attack                             0.0  1.0  3987
target of last resort|target of opportunity 0.0  1.0  5262
target of opportunity|secondary target      0.0  1.0  5068
secondary target|primary target             0.0  1.0  4950
mean_PPD:target of last resort              0.0  1.0  3779
mean_PPD:target of opportunity              0.0  1.0  3433
mean_PPD:secondary target                   0.0  1.0  3451
mean_PPD:primary target                     0.0  1.0  4147
log-posterior                               0.1  1.0   721
```

4 Results

The analysis revealed distinct patterns in target prioritization, highlighting the role of industry type and the mission's originating country in determining target selection, while the impact of tonnage and aircraft numbers was relatively subdued. Table 2.

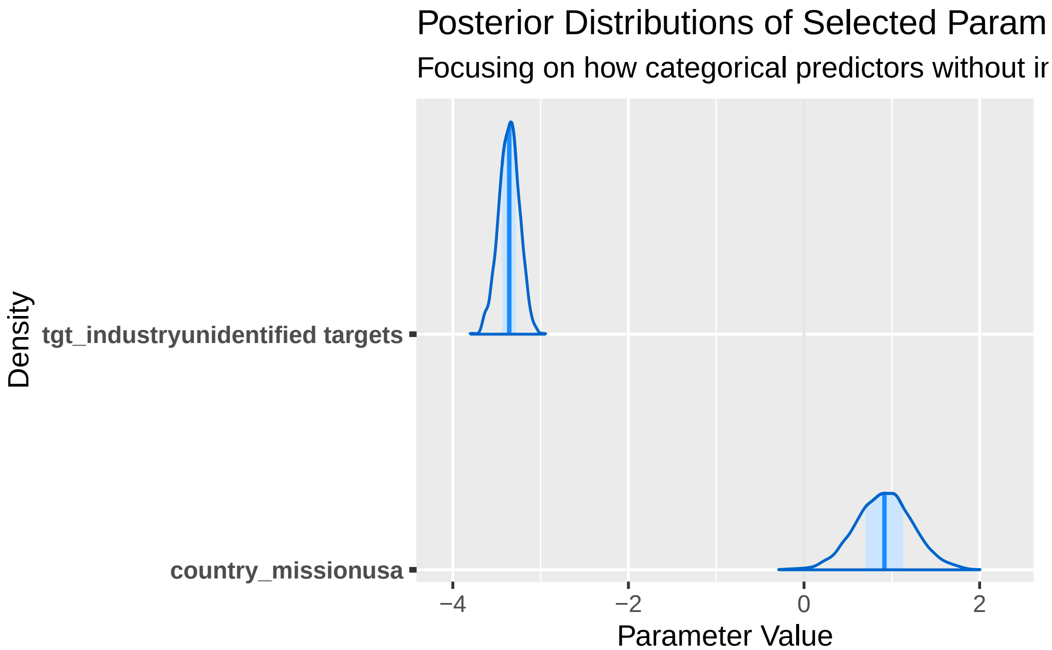


The Bayesian ordered logistic regression model's estimates are visualized in Figure X. Each point in the plot represents the posterior mean effect size of the predictor variables on the target priority ranking, while the lines indicate the 95% credible intervals. The estimates reveal several key insights into the Allied bombing strategy:

- The variable `tgt_industrySynthetic Oil Refineries` exhibits the largest positive effect, suggesting that synthetic oil refineries were assigned the highest priority for bombing missions, in line with the strategic objective to disrupt the German war effort's fuel supply.
- In contrast, the effects associated with `tgt_industryRailway Infrastructure` and `tgt_industryUrban Areas` are closer to zero, indicating a lower priority relative to other types of targets. This aligns with a strategy that placed less emphasis on disrupting transportation and civilian structures.
- Regarding the country of mission execution, `country_missionusa` shows a significant positive effect, reflecting the United States' prominent role in the strategic bombing campaign against Germany.
- The effects of `bomb_tons` and `aircraft_attack` are positive, suggesting that missions with greater bomb tonnage and more aircraft tend to have a higher target priority,

possibly reflecting the allocation of resources to strategically important missions.

These findings underscore the strategic considerations that guided the Allies' targeting decisions during World War II, highlighting the emphasis on industrial targets critical to the German war machine.



5 Discussion

5.1 First discussion point

The implications of our study stretch beyond historical curiosity, offering a lens through which modern strategic decisions can be evaluated, and the methodology refined.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

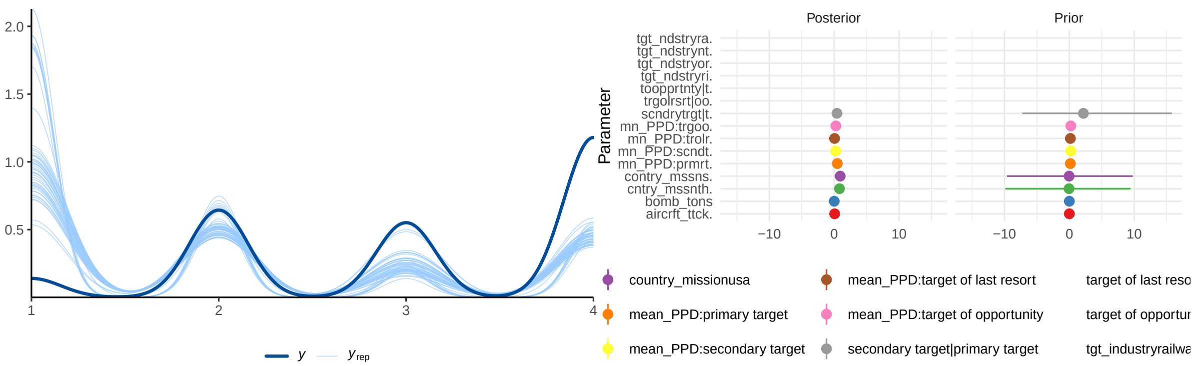
A Additional data details

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

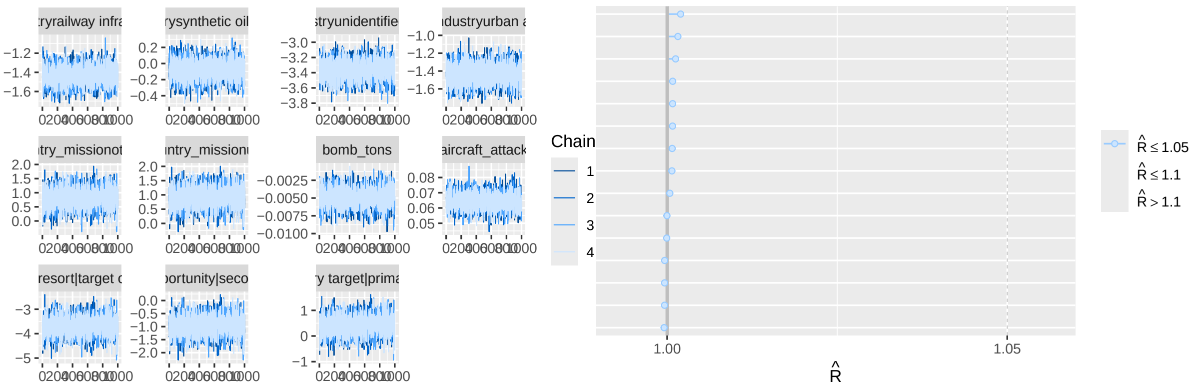
(b) Comparing the posterior with the prior

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.

Apache Arrow. 2021. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Arel-Bundock, Vincent. 2021. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://vincentarelbundock.github.io/modelsummary/.

Auguie, Baptiste. 2021. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. https://CRAN.R-project.org/package=gridExtra.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelman, Gabriel, Jonah Gabry, et al. 2021. *Bayesplot: Plotting for Bayesian Models*. https://mc-stan.org/bayesplot.

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

Robinson, David, Alex Hayes, et al. 2021. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://broom.tidymodels.org.

Team, Stan Development. 2021. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm.

Wickham, Hadley. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://ggplot2.tidyverse.org.

Wickham, Hadley et al. 2021. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://tidyverse.tidyverse.org.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.