Variation of [v] in Cook Islands Māori

Quartz Colvin^Rutgers University^

1 Rutgers University

Variation of [v] in Cook Islands Māori

# Introduction

In this paper, we will do a statistical analysis of [v] across a sample of islands in Cook Islands Māori. It’s known that in many dialects and other varieties of Māori, this phoneme can be realized as [w] or [v]. This paper aims to take a statistical approach to this generalization.

This paper has three main questions to investigate. First, are [w]-realizing dialects associated with more voicing than [v]-realizing dialects? Second, how does intensity of [v/w] vary by island? Lastly, how does duration of the target phoneme vary by island?

# Methods

Participants/material/procedure **???**

The data set this study uses was acquired by **???**

# Data

This data set contains a lot of information, so we will need to tidy it and only keep the relevant data for the analysis. The following code chunk shows how the data was tidied and what was kept from the untidy data set. I don’t show a preview of the untidy data set here because it had 17 columns and most of the columns will be taken out. In addition, it includes some metadata that I did not ask permission to share.

In the new data set, we need to fix the names that we’ve converted to values under ‘percent’. This is shown in the following code chunk.

Now that we have tidied this data set, here is a preview showing what it looks like.

| island | word | percent | hz | duration | intensity |
| --- | --- | --- | --- | --- | --- |
| Atiu | ava | 20 | 128.5916 | 0.08 | 54.08202 |
| Atiu | ava | 50 | 127.0908 | 0.08 | 54.08202 |
| Atiu | ava | 80 | 127.4540 | 0.08 | 54.08202 |
| Atiu | ava | 20 | 160.6464 | 0.07 | 58.06061 |
| Atiu | ava | 50 | 157.4397 | 0.07 | 58.06061 |
| Atiu | ava | 80 | 158.2820 | 0.07 | 58.06061 |

There are many more rows than this, but if all rows were included, it would be too long of a table to show here.

# Analysis

This section readdresses the three questions. First, are [w]-realizing dialects associated with more voicing than [v]-realizing dialects? Second, how does intensity of [v/w] vary by island? Lastly, how does duration of the target phoneme vary by island?

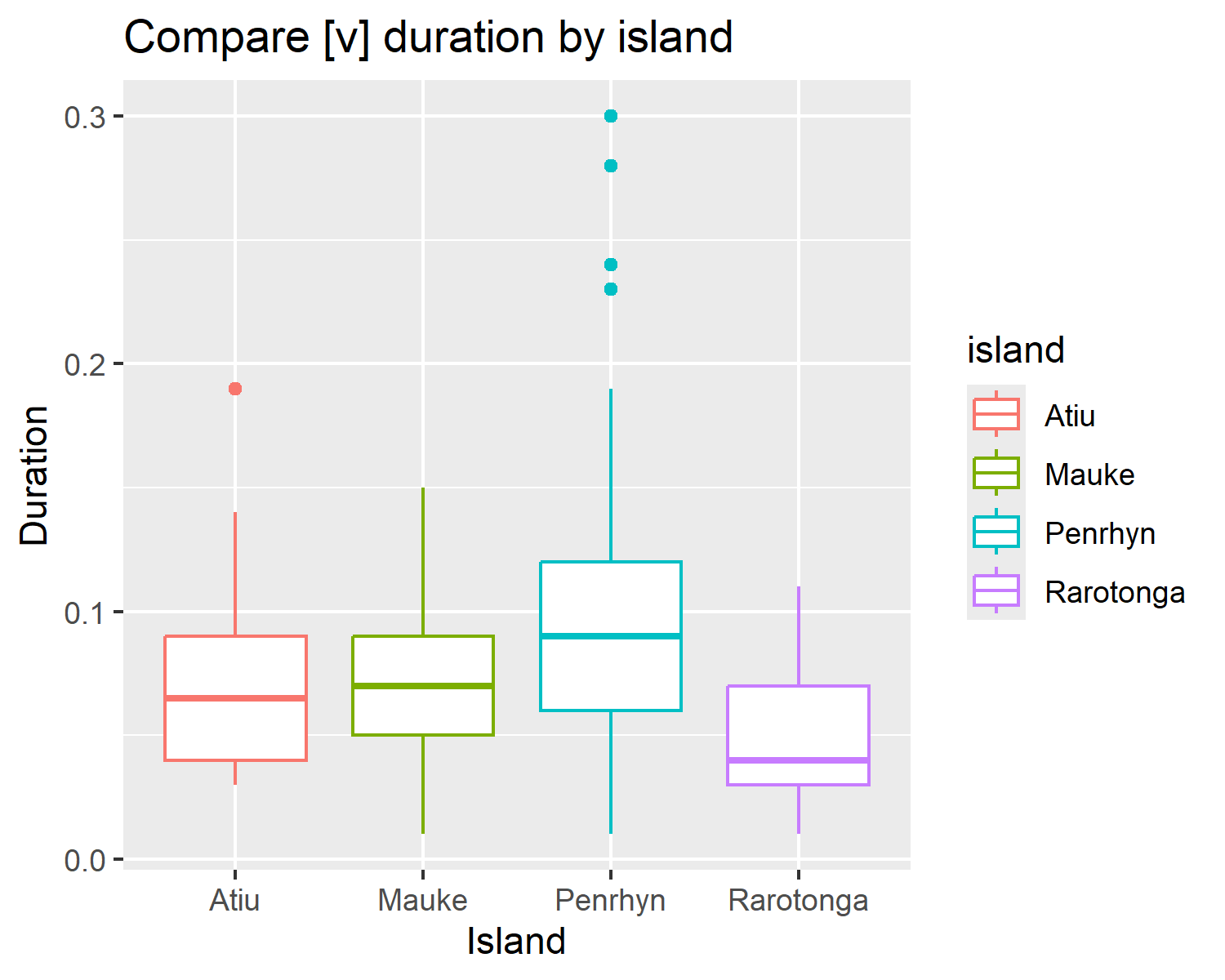
## Duration by island

Next, we can compare the duration of [v] across the four islands. The following table shows a boxplot of the durations by island.

**model-assumptions-and-diagnostics???**

**how-you-assessed-main-effects-and-interactions???**

**decision-rules-aka-alpha???**



A few things are visually clear from this comparative boxplot. Most obviously, Penrhyn had the largest range of [v] durations and Rarotonga had the smallest range. Rarotonga also had the lowest average (=mean) duration for [v] and Penrhyn had the highest average duration. Interestingly, Mauke and Rarotonga didn’t seem to have outliers in [v] durations, but Atiu and Penrhyn had a few outlier data points.

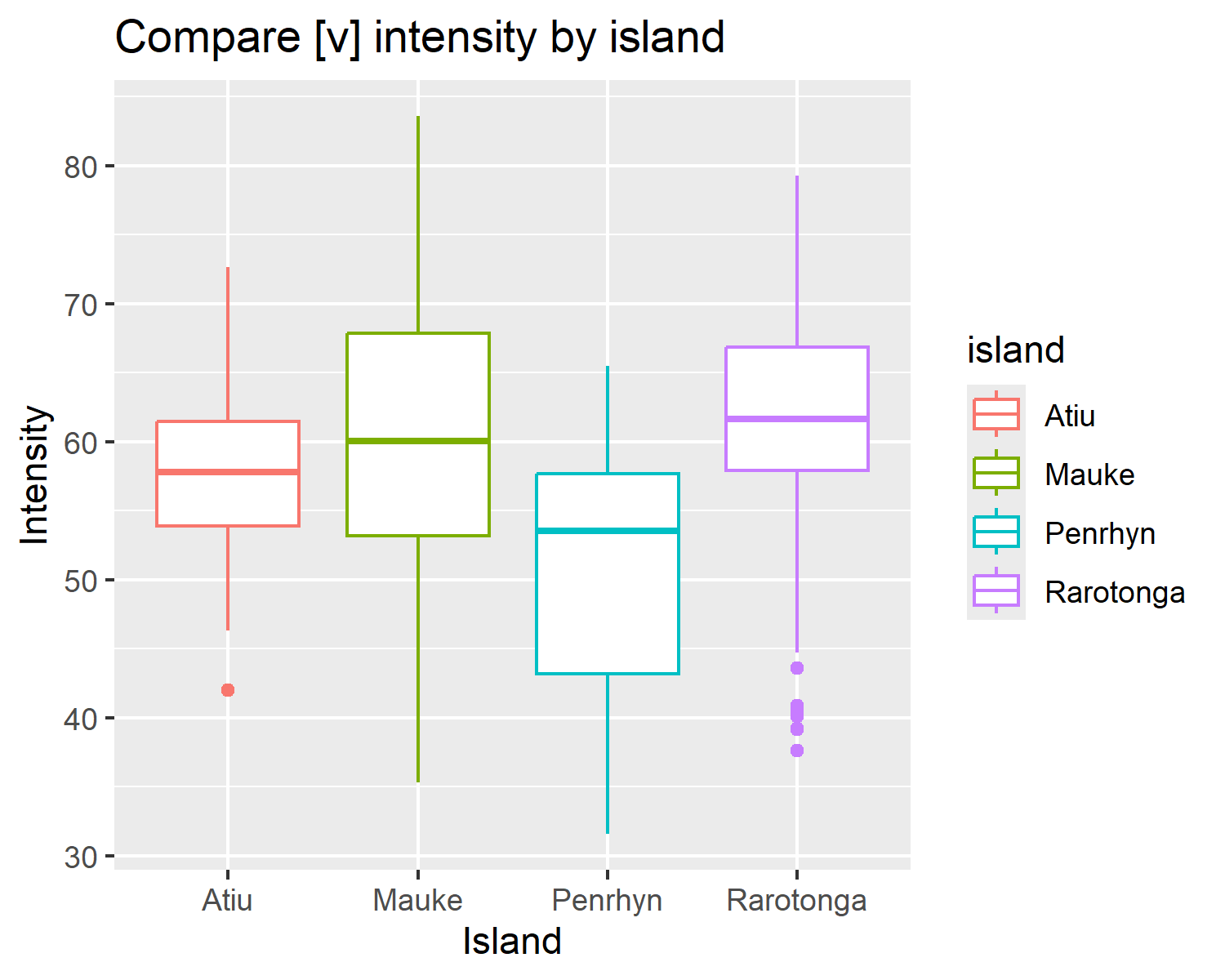
## Intensity by island

Next, in the same fashion we did to compare duration of [v] tokens, we can compare the intensity midpoints for the tokens across the islands.

**model-assumptions-and-diagnostics???**

**how-you-assessed-main-effects-and-interactions???**

**decision-rules-aka-alpha???**



Looking at the comparative boxplot, we can see how intensity midpoints vary across these four islands. The average intensity values were similar for all four, but Penrhyn had the lowest mean. These averages varied by less than 10, so they are quite similar. Mauke had the highest range for intensity midpoints, while Atiu had the smallest. Rarotonga had the most outliers and the other three islands had basically no outliers.

## Voicing by island

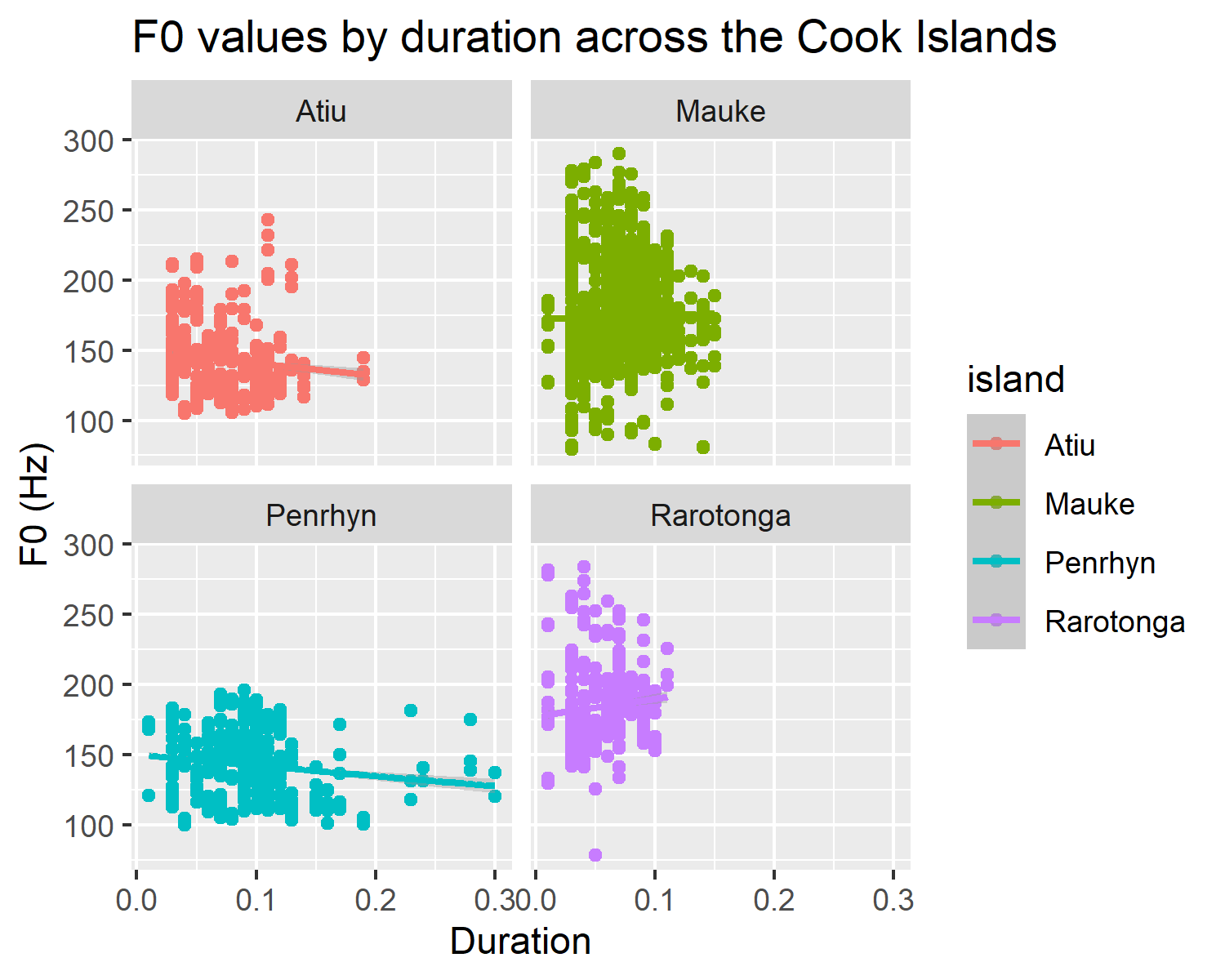
First we use f0 values for the target phoneme to see how voicing differs across the four islands in this data set. F0 values were extracted at the 20%, 50% and 80% points of the duration of the target phoneme.

The following plots show how F0 (measured in Hz) changes over the duration of the [v] phoneme. Each facet shows this for each island so that the reader can compare the islands visually.

**model-assumptions-and-diagnostics???**

**how-you-assessed-main-effects-and-interactions???**

**decision-rules-aka-alpha???**



As noted earlier, duration clearly differs between these four islands. The slopes for Atiu and Penrhyn show that over the duration of the phoneme, f0 decreases. This slope isn’t steep, but it is present. Mauke and Rarotonga have upwards slopes, showing that over the duration of [v], the f0 tends to increase. Rarotongan Māori has the sharpest increase in Hz over the duration of the phoneme.

# Formal results

**report-descriptive-stats???**

**model fit, main effects (usually NMC), interactions (usually, NMC), interpretations (directionality, effect size, uncertainty)**

**beta, SE, CI, p-val**

### Data analysis.

We used R (Version 4.5.0; R Core Team, 2024) and the R-packages *dplyr* (Version 1.1.4; Wickham, François, Henry, Müller, & Vaughan, 2023), *forcats* (Version 1.0.0; Wickham, 2023a), *ggplot2* (Version 3.5.2; Wickham, 2016), *here* (Version 1.0.1; Müller, 2020), *lme4* (Version 1.1.37; Bates, Mächler, Bolker, & Walker, 2015), *lmerTest* (Version 3.1.3; Kuznetsova, Brockhoff, & Christensen, 2017), *lubridate* (Version 1.9.4; Grolemund & Wickham, 2011), *Matrix* (Version 1.7.3; Bates, Maechler, & Jagan, 2025), *papaja* (Version 0.1.3; Aust & Barth, 2024), *purrr* (Version 1.0.4; Wickham & Henry, 2025), *readr* (Version 2.1.5; Wickham, Hester, & Bryan, 2024), *stringr* (Version 1.5.1; Wickham, 2023b), *tibble* (Version 3.2.1; Müller & Wickham, 2023), *tidyr* (Version 1.3.1; Wickham, Vaughan, & Girlich, 2024), *tidyverse* (Version 2.0.0; Wickham et al., 2019) and *tinylabels* (Version 0.2.5; Barth, 2023) for all our analyses.

# Discussion

**fill-in???**

# References

Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>

Barth, M. (2023). *tinylabels: Lightweight variable labels*. Retrieved from <https://cran.r-project.org/package=tinylabels>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

Bates, D., Maechler, M., & Jagan, M. (2025). *Matrix: Sparse and dense matrix classes and methods*. <https://doi.org/10.32614/CRAN.package.Matrix>

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*(3), 1–25. Retrieved from <https://www.jstatsoft.org/v40/i03/>

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>

Müller, K. (2020). *Here: A simpler way to find your files*. <https://doi.org/10.32614/CRAN.package.here>

Müller, K., & Wickham, H. (2023). *Tibble: Simple data frames*. <https://doi.org/10.32614/CRAN.package.tibble>

R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wickham, H. (2023a). *Forcats: Tools for working with categorical variables (factors)*. <https://doi.org/10.32614/CRAN.package.forcats>

Wickham, H. (2023b). *Stringr: Simple, consistent wrappers for common string operations*. <https://doi.org/10.32614/CRAN.package.stringr>

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *Dplyr: A grammar of data manipulation*. <https://doi.org/10.32614/CRAN.package.dplyr>

Wickham, H., & Henry, L. (2025). *Purrr: Functional programming tools*. <https://doi.org/10.32614/CRAN.package.purrr>

Wickham, H., Hester, J., & Bryan, J. (2024). *Readr: Read rectangular text data*. <https://doi.org/10.32614/CRAN.package.readr>

Wickham, H., Vaughan, D., & Girlich, M. (2024). *Tidyr: Tidy messy data*. <https://doi.org/10.32614/CRAN.package.tidyr>