Reproduction of Verbeke und Simon (2023)

Listening to accents: Comprehensibility, accentedness and intelligibility of native and non-native English speech

Rose Hörsting (rhoerst1), Zihang Su (zsu1), Ali Yıldız (ayildiz8)

2025-06-04

Table of contents

1 About the study

The aim of the study by Verbeke and Simon (2023) is to examine the multidialectal listening skills of proficient English learners, i.e. how well English learners (EFL) understand different accents of English. They establish three concepts: **comprehensibility** refers to the self-reported ease of understanding a speaker, **intelligibility** describes the actual understanding

measured by the performance in a transcription task, and **accentedness** reports how strong the listener perceived the accent to be.

Verbeke and Simon (2023: 5) apply the model of 'World Englishes' which was constructed by Kachru (1985). This model sorts the English-speaking and English-using world into three circles: The **Inner Circle** contains native English-speaking countries, the **Outer Circle** includes countries where English is an official language and that have a history of English occupation (largely colonisation, e.g. India, Nigeria, Kenya). Lastly, in countries of the **Expanding Circle**, English is not the first, dominant, or official language, but instead as an international language.

1.1 Research questions

These are the research questions defined by Verbeke and Simon (2023: 5-6):

- RQ1: How comprehensible are speakers with different native and non-native accents of English to EFL learners in higher education and to what extent do the perceived strength of the speaker's accent and listeners' familiarity with the speaker's accent impact on their comprehensibility ratings?
- RQ2: How intelligible are speakers with native and non-native accents of English to EFL learners enrolled at an institute of higher education in Flanders?
- RQ3: To what extent are EFL learners' comprehensibility ratings of speakers with native and non-native accents of English related to their intelligibility scores for these speakers?

1.2 Hypotheses

These are the hypotheses defined by Verbeke and Simon (2023: 5-6):

- H1: listeners' judgements of comprehensibility vary as function of which circle speakers represent, i.e. speakers with Inner circle accents are more comprehensible than speakers with non-Inner Circle accents
- H2: speakers with Expanding Circle accents are slightly easier to understand than speakers with Outer Circle accents, hypothesised to be rated more accented because further removed from expectation
- H3: a positive correlation between comprehensibility rating and intelligibility score of speakers, especially when comprehensibility rating is high

2 Method

2.1 Data provenance

Data used in this study were collected by Verbeke and Simon (2023) based on an experimental survey (consisting of a questionnaire followed by two listening tasks) developed in LimeSurvey (Version 2.73.1). All questions and instructions were provided in English. Participants completed the online experimental survey in a single session in a quiet room of their choice, which lasted on average 30 minutes (SD = 12 minutes). In the first part of the survey, listeners completed the questionnaire about their demographic and linguistic background. In the second part, listeners were asked to rate the speech of the eight speakers on comprehensibility and accentedness using nine-point scales. Clear and concise definitions of both terms were provided to the participants. Comprehensibility was defined as how easy or how difficult it is to understand what a speaker is saying (1 = easy to understand; 9 = hard to understand). Accentedness was defined as the degree to which a speaker's speech sounds different from a listener's expectation of English (1 = no accent; 9 = strong accent). The selected speech samples were randomized across all participants and were only played once. After completion of the comprehensibility and accentedness judgement task, listeners could take a self-paced break. In the final task, listeners were presented with 40 short sentences which they needed to transcribe orthographically. They were encouraged to transcribe as many words as possible. Participants heard each sentence only once and stimulus presentation was pseudo-randomized, in that no two sentences of the same speaker could immediately follow each other. After 20 sentences, participants could take a short break. Details regarding the composition of participants, questionnaire and stimuli are given as the following:

2.1.0.1 Listener:

- 33 (27 Female, 6 Male) Belgian Dutch students with self-reported normal or corrected-to-normal hearing (M = 18.9 years, SD = 1.3, range = 17–22 years)
- 6 listeners self-rated their listening skills in English as intermediate (B1)
- 16 listeners self-rated their listening skills in English as upper-intermediate (B2)
- 11 listeners self-rated their listening skills in English as advanced (C1-C2)

2.1.0.2 Questionnaire:

A questionnaire was administered to gather information about the demographic and linguistic background of each listener, and about the difficulties that these EFL learners experience when listening in English. Listeners were also asked to indicate how often they were exposed to the eight accents used in the present study on a 5-point scale (1 = never, 5 = very often).

2.1.0.3 Speaker

- 8 male speakers with 8 distinct accents of English (M = 48.4 years, SD = 18.6 years, range = 34-88 years)
- 1 speaker of General British English (Inner Circle, non-regional)
- 1 speaker of General American English (Inner Circle, non-regional)
- 1 speaker of Newcastle English (Inner Circle, regional)
- 1 speaker of Texan English (Inner Circle, regional)
- 1 speaker of Indian (Outer Circle)
- 1 speaker of Nigerian (Outer Circle)
- 1 speaker of Chinese (Expanding Circle)
- 1 speaker of Spanish (Expanding Circle)

2.1.0.4 Stimuli for the comprehensibility and accentedness judgement task:

- 8 passages (1 for each of the selected accents of English) of spoken language drawn from interviews and talk shows
- instances of backchanneling, false starts, throat clearings and salient background noises removed from the audio using Praat software
- root-mean-square amplitude scaled to 70 dB to minimize the potential effect of intensity differences on listeners' judgements
- total duration (M = 17.7 seconds, SD = 2.1 seconds), number of syllables (M = 80, SD = 13) and speaking rate (M = 4.5; SD = 0.7, range = 3.8–5.4 syllables per second) controlled

2.1.0.5 Stimuli for the orthographic transcription task:

- 40 short sentences (5 for each speaker of English) selected from the same talk shows or interviews but different from the excerpts selected for the comprehensibility and accentedness judgement task
- each selected sentence contained only moderate to high frequency content words
- all sentences scaled to an average intensity of 70 dB to increase inter-stimulus similarity
- total duration (M = 2.45 seconds, SD = 0.51 seconds), number of syllables (M = 11.8, SD = 1.95) and speaking rate (M = 4.90, SD = 0.63) controlled

Verbeke and Simon (2023) issued a data availability statement along with the publication of their paper. The data and their code is available at the Tromsø Repository of Language and Linguistics.

2.2 Reading and Pre-processing the Data

In providing their data and code, as well as detailed supplementary material, Verbeke and Simon (2023) hugely facilitated reproduction of their study. However, since they fit several complex models, the code takes a long time to fully run (more than one hour) and requires a lot of computational power. In this paper, we therefore do not aim to exactly reproduce the graphs and results of Verbeke and Simon (2023), but instead use the data to create our own analysis on a smaller scale. We decided to use the program `Quarto` instead of `R Markdown` as it promises to be the "next generation version of R Markdown from Posit" (https://quarto.org). Posit is the development company of RStudio and like `R Markdown`, `Quarto` renders documents via knitr. `Quarto` is free-of-cost and can be downloaded here.

First, we read in the data provided by the authors of the study.

In the `ComprAcc.data`, comprehensibility and accentedness are rated from the EFL learners on a scale from 1 to 10 (for comprehensibility: 1 = easy to understand; 9 = hard to understand; for accentedness: 1 = no accent; 9 = strong accent) (Verbeke and Simon 2023: 8)). The intelligibility data (`Intell.data`) represents the transcription performance of the EFL learners when listening to each accent (accuracy, so whether the word was correctly or falsely transcribed). The dataframe `ComprIntell.data` includes comprehensibility ratings and intelligibility scaled on the range of 0 to 1 (seemingly means: how much percent did the participant transcribe correctly in that accent).

We converted the character variable "Accent" in all three data frames to factor to get its levels.

```
[1] "ChinEng" "GAE"     "GBE"     "IndEng"  "NBE"     "NigEng"  "SAE"
[8] "SpanEng"
```

Next, we renamed the levels of the factor for more transparent labeling.

```
[1] "Chinese English"          "General American English"
[3] "General British English"  "Indian English"
[5] "Newcastle English"        "Nigerian English"
[7] "Texan English"            "Spanish English"
```

We then reordered the levels according to Kachru (1985) ' model of 'World Englishes' (Inner Circle: Standard British, Standard American, Dialect British, Dialect American; Outer circle: Indian English, Nigeran English; Expanding Circle: Chinese English, Spanish English).

```
[1] "General British English"  "General American English"
[3] "Newcastle English"        "Texan English"
[5] "Indian English"           "Nigerian English"
[7] "Chinese English"          "Spanish English"
```

3 Results and Discussion

3.1 Accentedness

Figure 1 visualises the accentedness ratings for each accent. It illustrates that General American English was rated the least accented; General British English was rated second least accented but more than General American. Participants assessed Newcastle English (British dialect) as slightly more accented than General British English while Texan, Indian, Nigerian, Chinese and Spanish English were all rated similarly as the most accented.

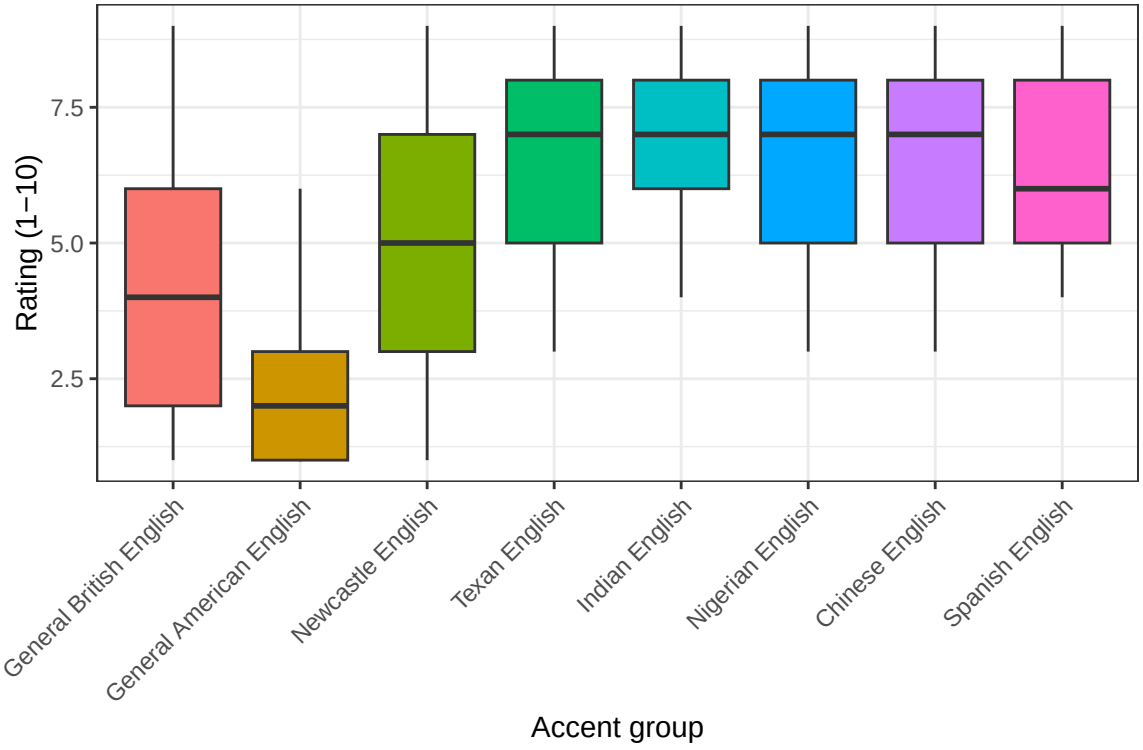

Figure 1: Rated Accentedness Grouped by Accent

3.2 Intelligibility

The second research question of the study examined how intelligible speakers with native and non-native English accents are to EFL learners enrolled at a higher education institution in Flanders.

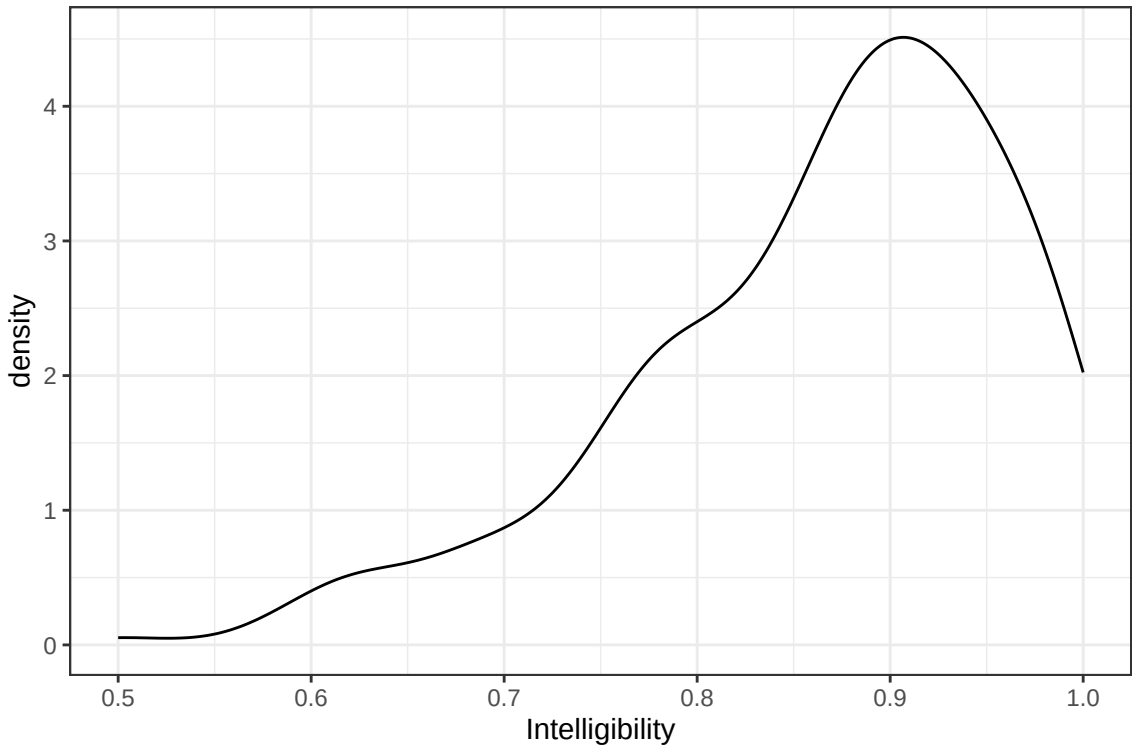

Figure 2: Density of Participants' Intelligibility

We attempted to reproduce the analysis on intelligibility data, and visualize our output as seen on Figure 2. The figure shows a bell-shaped curve but is left-skewed: overall, participants performed above chance in the transcription task.

As Figure 3 demonstrates, participants performed better than chance for all accents, although there was some variation between languages that match the Circle model. Most errors were made for Nigerian English and Indian English, which belong to the Outer Circle according to Kachru (1985), i.e. participants had most trouble understanding these accents. Participants had least trouble transcribing General American English & Texan English, i.e. participants understood these accents best.

Notably, there were more errors in understanding Standard British English than American English (British English similar to Expanding Circle Accents). For American English, tran-

sciption performance for dialect speech (Texan English) was slightly worse than for the Standard variant (though Accuracy was still very high) while for British English, the Newcastle dialect improved intelligibility slightly compared to General British English.

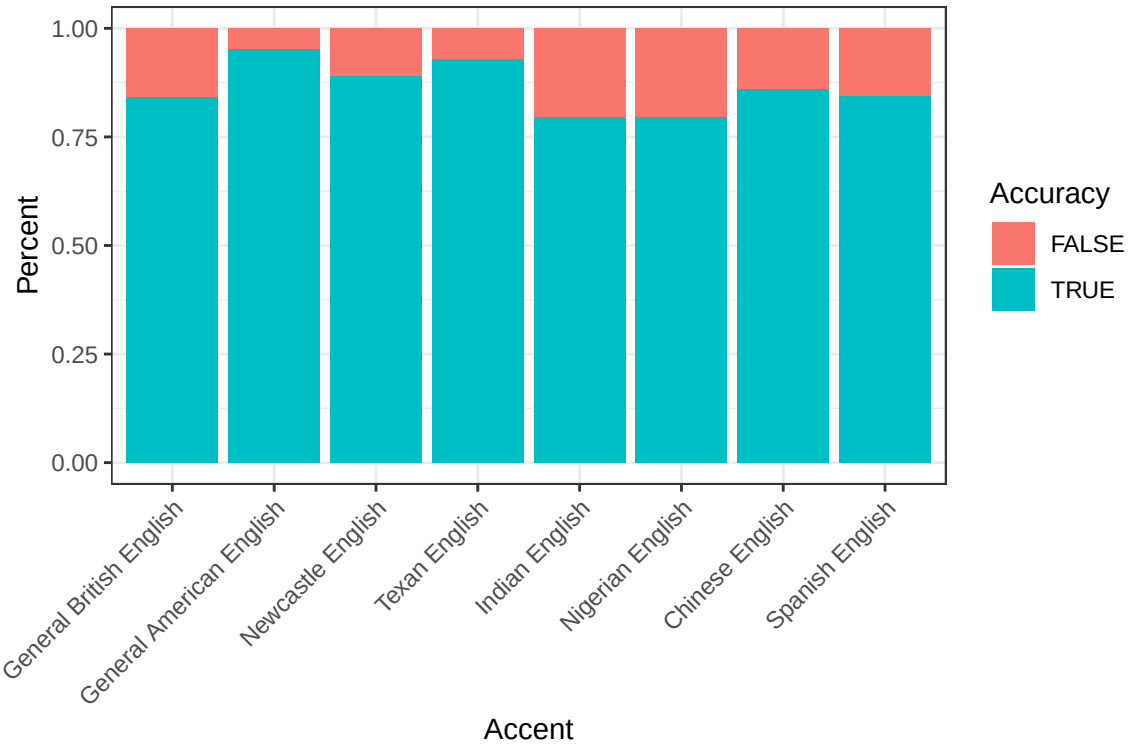

Figure 3

3.3 Intelligibility and Comprehensibility

Finally, we explored the correlation of intelligibility and comprehensibility to approach the question whether self-reported comprehensibility aligned with objective, measured intelligibility. (RQ3: To what extent are EFL learners' comprehensibility ratings of speakers with native and non-native accents of English related to their intelligibility scores for these speakers?)

In line with Verbeke and Simon (2023), we created a scatterplot to check the correlation of comprehensibility and intelligibility for each accent.

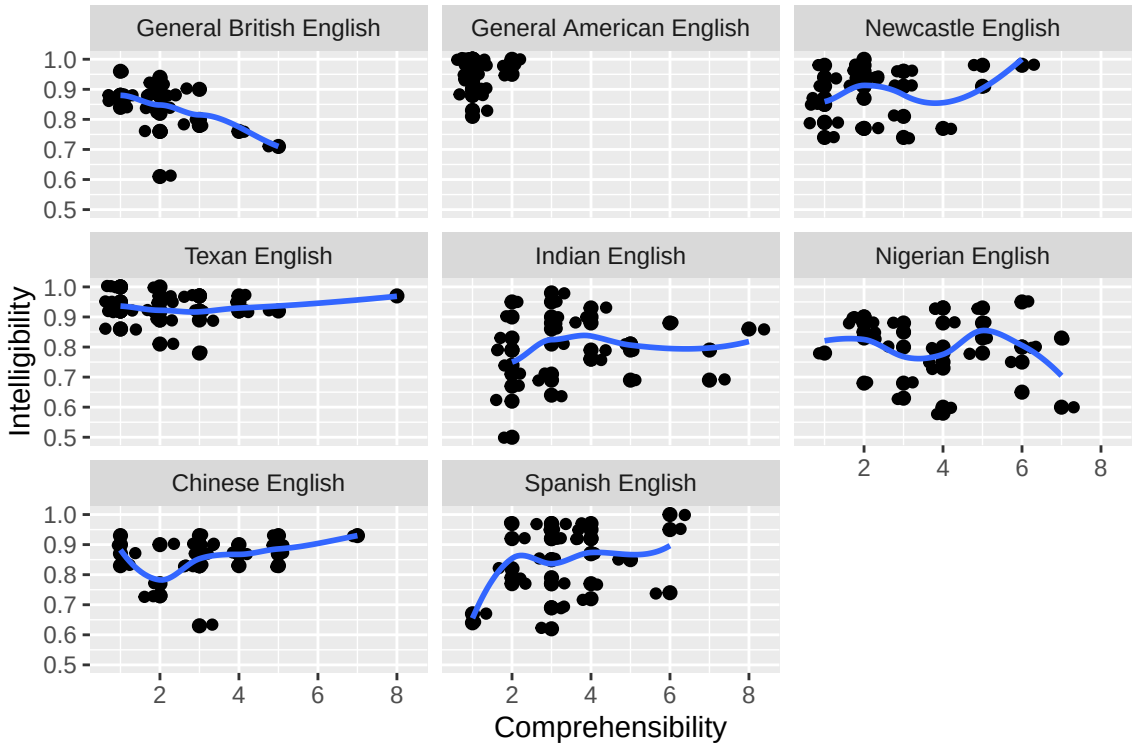

Figure 4: Scatterplot illustrating the correlation of Comprehensibility and Intelligibility

Figure 4 looks almost identical to the one from the paper. As dispersion of data points differs across accents, and no real pattern can be detected, there seems to no clear correlation of comprehensibility and intelligibility. Definitely not linear across accents. Therefore, self-reported ease of understanding (comprehensibility) and actual understanding (intelligibility) did not linearly and clearly relate to each other.

4 Conclusion

Our reproduction, just like the original study, has shown that for the most part, proficient EFL learners were able to understand accented English speech very well, but there were differences: General American English, which was rated the least accented and the most comprehensible was the easiest accent to understand. Comprehensibility and accentedness are positively correlated; comprehensibility ratings did statistically not align with actual transcription performance.

… (Please add!)

5 References

Kachru, Braj B. 1985. "Standards, Codification and Sociolinguistic Realism: The English Language in the Outer Circle." *English in the World: Teaching and Learning the Language and Literatures/Cambridge UP*.

Le Foll, Elen. 2025. "Data Analysis for the Language Sciences: A Very Gentle Introduction to Statistics and Data Visualisation in r." Open Educational Resource. https://elenlefoll.github.io/RstatsTextbook/.

Verbeke, Gil, and Ellen Simon. 2023. "Listening to Accents: Comprehensibility, Accentedness and Intelligibility of Native and Non-Native English Speech." *Lingua* 292: 103572. https://doi.org/10.1016/j.lingua.2023.103572.

5.1 Session Info

```
R version 4.4.1 (2024-06-14)
Platform: aarch64-apple-darwin20
Running under: macOS 15.5

Matrix products: default
BLAS:   /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRblas.0.dylib
LAPACK: /Library/Frameworks/R.framework/Versions/4.4-arm64/Resources/lib/libRlapack.dylib;

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

time zone: Europe/Berlin
tzcode source: internal

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
 [1] kableExtra_1.4.0 here_1.0.1       lubridate_1.9.3  forcats_1.0.0
 [5] stringr_1.5.1    dplyr_1.1.4      purrr_1.0.2      readr_2.1.5
 [9] tidyr_1.3.1      tibble_3.2.1     ggplot2_3.5.1    tidyverse_2.0.0

loaded via a namespace (and not attached):
 [1] utf8_1.2.4        generics_0.1.3    xml2_1.3.6        lattice_0.22-6
 [5] stringi_1.8.4     hms_1.1.3         digest_0.6.37     magrittr_2.0.3
 [9] evaluate_1.0.0    grid_4.4.1        timechange_0.3.0  fastmap_1.2.0
[13] Matrix_1.7-0      rprojroot_2.0.4   jsonlite_1.8.9    mgcv_1.9-1
[17] fansi_1.0.6       viridisLite_0.4.2 scales_1.3.0      cli_3.6.3
```

```
[21] rlang_1.1.4       splines_4.4.1     munsell_0.5.1     withr_3.0.1
[25] yaml_2.3.10       tools_4.4.1       tzdb_0.4.0        colorspace_2.1-1
[29] vctrs_0.6.5       R6_2.5.1          lifecycle_1.0.4   pkgconfig_2.0.3
[33] pillar_1.9.0      gtable_0.3.5      glue_1.8.0        systemfonts_1.1.0
[37] xfun_0.48         tidyselect_1.2.1  rstudioapi_0.16.0 knitr_1.48
[41] farver_2.1.2      nlme_3.1-164      htmltools_0.5.8.1 rmarkdown_2.28
[45] svglite_2.1.3     labeling_0.4.3    compiler_4.4.1
```

5.2 Package References

The code to generate these package references was written by Elen Le Foll (Le Foll 2025).

[1] G. Grolemund and H. Wickham. "Dates and Times Made Easy with lubridate". In: *Journal of Statistical Software* 40.3 (2011), pp. 1-25. https://www.jstatsoft.org/v40/i03/.

[2] K. Müller. *here: A Simpler Way to Find Your Files*. R package version 1.0.1. 2020. https://here.r-lib.org/.

[3] K. Müller and H. Wickham. *tibble: Simple Data Frames*. R package version 3.2.1. 2023. https://tibble.tidyverse.org/.

[4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2024. https://www.R-project.org/.

[5] V. Spinu, G. Grolemund, and H. Wickham. *lubridate: Make Dealing with Dates a Little Easier*. R package version 1.9.3. 2023. https://lubridate.tidyverse.org.

[6] H. Wickham. *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0. 2023. https://forcats.tidyverse.org/.

[7] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. https://ggplot2.tidyverse.org.

[8] H. Wickham. *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1. 2023. https://stringr.tidyverse.org.

[9] H. Wickham. *tidyverse: Easily Install and Load the Tidyverse*. R package version 2.0.0. 2023. https://tidyverse.tidyverse.org.

[10] H. Wickham, M. Averick, J. Bryan, et al. "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43 (2019), p. 1686. DOI: 10.21105/joss.01686.

[11] H. Wickham, W. Chang, L. Henry, et al. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. R package version 3.5.1. 2024. https://ggplot2.tidyverse.org.

[12] H. Wickham, R. François, L. Henry, et al. *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. 2023. https://dplyr.tidyverse.org.

[13] H. Wickham and L. Henry. *purrr: Functional Programming Tools*. R package version 1.0.2. 2023. https://purrr.tidyverse.org/.

[14] H. Wickham, J. Hester, and J. Bryan. *readr: Read Rectangular Text Data*. R package version 2.1.5. 2024. https://readr.tidyverse.org.

[15] H. Wickham, D. Vaughan, and M. Girlich. *tidyr: Tidy Messy Data*. R package version 1.3.1. 2024. https://tidyr.tidyverse.org.

[16] Y. Xie. *Dynamic Documents with R and knitr*. 2nd. ISBN 978-1498716963. Boca Raton, Florida: Chapman and Hall/CRC, 2015. https://yihui.org/knitr/.

[17] Y. Xie. "knitr: A Comprehensive Tool for Reproducible Research in R". In: *Implementing Reproducible Computational Research*. Ed. by V. Stodden, F. Leisch and R. D. Peng. ISBN 978-1466561595. Chapman and Hall/CRC, 2014.

[18] Y. Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.48. 2024. https://yihui.org/knitr/.

[19] H. Zhu. *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.4.0. 2024. http://haozhu233.github.io/kableExtra/.