

List of packages used in R

- **dplyr** for data transformation¹
- **impute** for imputing missing values, using k-nearest neighbor averaging²
- **candisc** for the canonical correlation analysis³
- **car** for calculating variance inflation factor⁴
- **caret** for implementing k-fold cross-validations and training/evaluating logistic model⁵
- **foreach** for parallel execution of jackknife iterations⁶
- **doParallel** for registering a parallel backend for **foreach** and **caret**⁷
- **reshape2** for converting wide dataframes into long dataframes to ease plotting⁸
- **cluster** for identifying agglomerative coefficients for different combinations of linkage and distance⁹
- **NbClust** for determining optimal number of clusters¹⁰
- **fpc** for assessing cluster stability¹¹
- **MASS** to generate multivariate normal distribution for the significance testing of cluster assignments¹²
- **MKinfer** for performing permuted t-tests¹³
- **pROC** for calculating ROC-AUC for the logistic model¹⁴
- **coin** to perform permuted test of the logistic model¹⁵
- **ggplot2** for general visualisation¹⁶
- **extrafont** for editing fonts in visualisations¹⁷
- **ggdendro** for converting hierarchical clustering object to a format compatible with **ggplot2**¹⁸
- Other packages integrated in the R environment, including **The R Base Package**, **parallel** and **stats**¹⁹

References

1. Wickham H, François R, Henry L, Müller K, Vaughan D. dplyr: A grammar of data manipulation. R package version 1.1.4 ed2023.
2. Hastie T, Tibshirani R, Narasimhan B, Chu G. impute: Imputation for microarray data. R package version 1.78.0 ed2024.
3. Friendly M, Fox J. candisc: Visualizing generalized canonical discriminant and canonical correlation analysis. R package version 0.9.0 ed2024.
4. Fox J. An R companion to applied regression. 2nd ed. Thousand Oaks, Calif: Sage, 2011.
5. Kuhn M. Building predictive models in R using the caret package. J Stat Softw 2008;28(5):1-26. DOI: 10.18637/jss.v028.i05.
6. Microsoft Corporation, Weston S. foreach: Provides foreach looping construct. R package version 1.5.2 ed2022.
7. Microsoft Corporation, Weston S. doParallel: Foreach parallel adaptor for the ‘parallel’ package. R package version 1.0.17 ed2022.
8. Wickham H. Reshaping data with the reshape package. J Stat Softw 2007;21(12):1-20. DOI: 10.18637/jss.v021.i12.
9. Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K. cluster: Cluster analysis basics. R package version 2.1.6.2023.
10. Charrad M, Ghazzali N, Boiteau V, Niknafs A. Nbclust: An R package for determining the relevant number of clusters in a data set. J Stat Softw 2014;61(6):1-36. DOI: 10.18637/jss.v061.i06.
11. Hennig C. fpc: Flexible procedures for clustering. R package version 2.2-13 ed2024.
12. Venables WN. Modern applied statistics with S. 4th ed. New York: Springer, 2002.
13. Kohl M. MKinfer: Inferential statistics. R package version 1.2 ed2024.
14. Robin X, Turck N, Hainard A, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12(1):77. DOI: 10.1186/1471-2105-12-77.
15. Hothorn T, Hornik K, van de Wiel MA, Zeileis A. A Lego system for conditional inference. Am Stat 2006;60(3):257-263. DOI: 10.1198/000313006X118430.
16. Wickham H. ggplot2: Elegant graphics for data analysis. 2nd edition. ed. Cham: Springer International Publishing; 2016.
17. Chang W. extrafont: Tools for using fonts. R package version 0.19 ed2023.
18. de Vries A, Ripley BD. ggdendro: Create dendrogram and tree diagrams using ‘ggplot2’. R package version 0.2.0 ed2024.
19. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024.