

profileR: An R package for profile analysis

Christopher David Desjardins¹ and Okan Bulut²

1 St. Michael's College 2 University of Alberta

DOI: [10.21105/joss.01941](https://doi.org/10.21105/joss.01941)

Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

Editor: [Marcos Vital](#) ↗

Reviewers:

- [@wjakethompson](#)
- [@jrosen48](#)

Submitted: 21 November 2019

Published: 17 March 2020

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

Summary

A profile is a vector containing a person's or group's scores from an assessment. An assessment, defined here as a series of items used to measure one or more traits, is administered in a variety of contexts. For example, schools administer high-stakes standardized assessments for accountability purposes and to measure student aptitude and progress; a young adult exhibiting compensatory and binge eating behaviors may complete a brief assessment to ascertain the likelihood of an underlying eating disorder; and human resources may recommend whether to hire an applicant based on a personality assessment. A student's, a young adult's, and an applicant's scores on these assessments make up their profile.

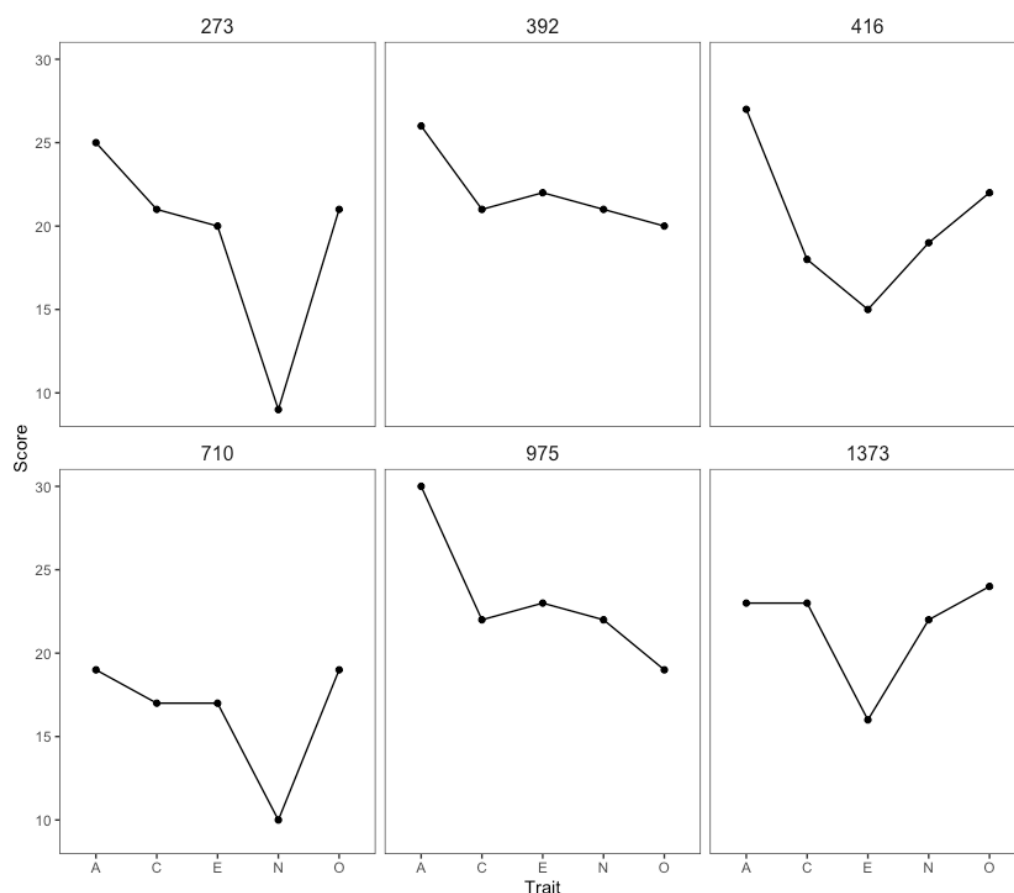


Figure 1: Six random profiles on a personality assessment.

Figure 1 shows the profiles of six random participants on an assessment of the five personality traits - agreeableness (A), conscientiousness (C), extraversion (E), neuroticism (N), and openness (O) (Revelle, 2018). Figure 1 shows that participant 975 had the highest average score across the five traits (23.6), while 710 had the lowest average score (16.4), that participants 273 and 710 and 392 and 975 share similar, potentially qualitatively important, patterns, and that participant 273's profile had the greatest variance (36.2), while participant 392's had the least (5.5). These features (i.e., level, pattern, and variability) in a profile often hold important clinical meaning and statistically assessing differences in these features is of importance to researchers, educators, and other practitioners.

A suite of techniques, collectively known as **profile analysis**, can be used to analyze and classify profiles. These techniques come from a variety of statistical frameworks including repeated-measures MANOVA, multidimensional scaling and factor analysis, cluster analysis, and machine learning. Applying these disparate methods in R can be challenging for novice users and practitioners with limited statistical training as these techniques are either unavailable, require interfacing with a myriad of packages, or require understanding the relationship between a technique and a more general statistical modeling framework. The R package `profileR` was designed to address this need and can be installed by typing the following in R:

```
install.packages("profileR")
```

or for the development version on GitHub

```
devtools::install_github("cddesja/profileR")
```

The R package `profileR`, currently in version 0.3-6 on the comprehensive R archive network (CRAN), implements profile reliability (Bulut, 2013; Bulut, Davison, & Rodriguez, 2017), criterion-related profile analysis (Davison & Davenport, 2002), profile analysis via multidimensional scaling (Davison, 1994), moderated profile analysis, profile analysis by group, and a within-person factor model to derive score profiles (Davison, Kim, & Close, 2009) as well as a variety of graphical methods to visualize profiles. Related profile analysis packages exist in R, such as `tidyLPA` (Rosenberg, Beymer, Anderson, Lissa, & Schmidt, 2018) and `mclust` (Scrucca, Fop, Murphy, & Raftery, 2016), as well as in `Mplus` (Muthén & Muthén, 2017), however, our approach does not rely on mixture modeling and aims to provide access to a wider range of profile analytic techniques. An extensive vignette documenting `profileR` (Bulut & Desjardins, 2020) is included with the package as well as vignette to ensure the functionality of the package locally.

The API for `profileR` was designed to provide a unified, consistent, and user-friendly R interface for these methods (see Figure 2). It uses the S3 class and many generics have been written to work with `profileR` objects. Any future statistical methods added to `profileR` will adhere to the existing API (such as the addition of Bayesian approaches and multilevel profile analysis).

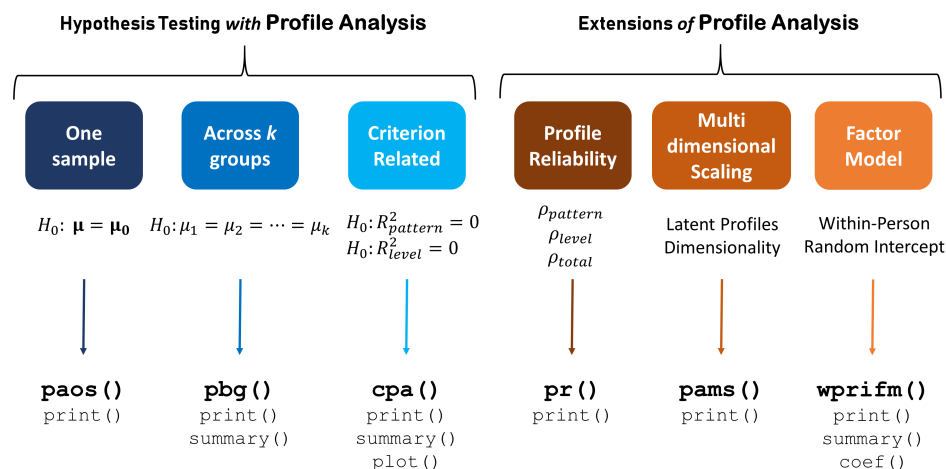


Figure 2: Primary functions and their generic functions available in profileR

profileR has been used in education, counseling, and to study learning behavior in mice. The package has been featured in a handbook on measurement and psychometrics in R (Desjardins & Bulut, 2018); used in workshops and graduate level courses; is a part of the psychometric CRAN task view; and is downloaded, on average, 752 per month from CRAN. Our focus on R novices and usability, should help to expand the reach of profile analysis into new scientific disciplines.

References

- Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional irt models*. (PhD thesis). University of Minnesota.
- Bulut, O., Davison, M. L., & Rodriguez, M. C. (2017). Estimating between-person and within-person subscore reliability with profile analysis. *Multivariate Behavioral Research*, 52(1), 86–104. doi:[10.1080/00273171.2016.1253452](https://doi.org/10.1080/00273171.2016.1253452)
- Bulut, O., & Desjardins, C. D. (2020). Profile analysis of multivariate data: A brief introduction to the profileR package. *PsyArXiv*. doi:[10.31234/osf.io/sgy8m](https://doi.org/10.31234/osf.io/sgy8m)
- Davison, M. L. (1994). Multidimensional scaling models of personality responding. In S. Strack & M. Lorr (Eds.), *Differentiating normal and abnormal personality*. Springer Publishing Co.
- Davison, M. L., & Davenport, E. C., Jr. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods*, 7(4), 468–484. doi:[10.1037/1082-989X.7.4.468](https://doi.org/10.1037/1082-989X.7.4.468)
- Davison, M. L., Kim, S. K., & Close, C. (2009). Factor analytic modeling of within person variation in score profiles. *Multivariate Behavioral Research*, 44, 668–687. doi:[10.1080/00273170903187665](https://doi.org/10.1080/00273170903187665)
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of educational measurement and psychometrics using R*. Chapman; Hall/CRC. doi:[10.1201/b20498](https://doi.org/10.1201/b20498)
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide*. Eighth edition. Muthén & Muthén.

- Revelle, W. (2018). *Psych: Procedures for psychological, psychometric, and personality research*. Evanston, Illinois: Northwestern University. Retrieved from <https://CRAN.R-project.org/package=psych>
- Rosenberg, J., Beymer, P., Anderson, D., Lissa, C. van, & Schmidt, J. (2018). TidyLPA: An r package to easily carry out latent profile analysis (lpa) using open-source or commercial software. *Journal of Open Source Software*, 3(30), 978. doi:[10.21105/joss.00978](https://doi.org/10.21105/joss.00978)
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 205–233. doi:[10.32614/rj-2016-021](https://doi.org/10.32614/rj-2016-021)