Under the Radar: Simplifying the Representation of Latent Class Characteristics

jimi adams¹ and Adam M. Lippert¹

Abstract

In this visualization, we demonstrate how to use radar plots to represent the class-specific posterior response probabilities from latent class analysis results. These plots allow for a simple representation of the class differences in the distributions across the modeled indicators. We demonstrate the utility of this approach with results from a published model of women’s employment and family life circumstances. In doing so, we demonstrate how to avoid some of the pitfalls common to radar plots and provide example code to allow other researchers to readily adapt this approach to present their own results.

Keywords

latent class analysis, data visualization, radar plots, work/family

Latent Class Analysis

Sociologists have increasingly utilized a variety of cluster-based analytic strategies, especially those that incorporate the estimation of relationships between latent traits and observed behaviors. One particularly common form of this type of model is latent class analysis, or LCA (McCutcheon 1987).

LCA uses observable indicators to yield unobserved, or latent, probabilities of endorsing $y_1 \ldots y_k$ response patterns accounting for the covariance among observed indicators. Its basic form is defined by

$$P(Y = y) = \sum_{t} P(X = t) P(Y = y | X = t),$$

where $P(X = t)$ denotes the probability of belonging to class $t$ and $P(Y = y | X = t)$ is the probability of having response pattern $y$ conditional on membership in class $t$.

In reporting LCA results, it is common to disclose both the probability of class membership (gammas) and the distribution of class-specific item-response probabilities (rhos). For the latter, it is difficult to concisely and clearly represent the full pattern of response probabilities across classes, any time there are more than a few variables with a limited number of response items, combined with few classes.

Alternatives for presenting LCA results have typically included (a) tables with each class-item-response category probabilities reported, sometimes highlighting patterns across the observed variable distributions and item probabilities; (b) line/bar charts that group classes together; and occasionally (c) approaches that summarize some of that information, for example, with ternary plots separating out the differences in response patterns across classes (Bakk and Roux 2017). While rhos inherently include three dimensions of interest—(1) response-item distributions, (2) class memberships, and (3) the differences in (1) by (2), none of the existing approaches for presenting these results optimize across all three of these dimensions simultaneously, instead prioritizing some over the others. For example, tables can provide detailed item-response distributions but make seeing the class differences across those difficult. Simple line or bar charts can make the distributions visible but can only cluster classes together (making comparisons across them difficult) or variables together (making class compositions difficult to see). Alternatively, while ternary plots make the differences between classes easily perceptible, the underlying item-response distributions are obfuscated.

¹University of Colorado Denver, Denver, CO, USA

Corresponding Author:

jimi adams, Department of Health & Behavioral Sciences, University of Colorado Denver, Campus Box 188, PO Box 173364, Denver, CO, USA. Email: jimi.adams@ucdenver.edu
Visualizing LCA Results with Radar Plots

Here, we suggest radar plots (Nakazawa 2018) accomplish all three of these aims simultaneously, demonstrate their use as applied to a previously published result, and provide replicable code (in ways that incorporate these optimizations) for others to adapt for presenting their own results.

The radar plots presented in Figure 1 use the LCA solution from Lippert and Damaske (2019) on young adult women’s work and family formation trajectories. For presenting results in the radar plot format, we choose to employ two options. First, we normalize each variable’s value into a single index ranging from 0 to 1. Normalization requires multiplying the item-specific response probabilities by a scaling factor determined by the order and number of categories within each item. For illustration purposes, the variables in the presented model include dichotomous, trichotomous, and other ordered variables. Second, we reverse code indicators as necessary to ensure all normalized values are oriented in the same direction.

For the seven classes identified, four are characterized by full-time employment (left panel) and three by lower engagement with paid labor (right panel). Two classes—professional workers with and without children—were similar with respect to their employment status, propensity for well-paid high-skilled work, and job decision latitude but differed by their relationship status and whether they had children in their care (see blue vs. orange polygons in the left panel). Further distinctions between the identified classes are visible by comparing across the distribution of item posterior response probabilities between the classes.

Our approach provides a means for visually representing LCA results that allows readers to easily compare class-level differences in the pattern across distributions of the

---

1For the complete table of the posterior probabilities included in this solution, see the .csv file in the Appendices.

2These options can be turned on/off in the provided code.

3In the case presented here, this required subtracting the result of the normalization step from 1 for the variable indicating number of children in one’s care.

4This approach is likely also readily adaptable to the presentation of results from other clustering-based approaches.
item-specific response categories. We hope that the companion code will allow future researchers to more easily present LCA results in their own research.

**ORCID iD**

jimi adams [2] https://orcid.org/0000-0001-5043-1149

**Supplemental Material**

Data to replicate Figure 1, and code for adapting the approach to your own results are available at: https://github.com/jimiadams/LCA-Viz.

**References**

Bakk, Zsuzsa, and Niel J. le Roux. 2017. “Visualizing Latent Class Models with Analysis-of-Distance Biplots.” *Sociological Methodology* 47(1):345–78.

Lippert, Adam M., and Sarah Damaske. 2019. “Finding Jobs, Forming Families, and Stressing out? Work, Family, and Stress among Young Adult Women in the United States.” *Social Forces*. https://doi.org/10.1093/sf/soy117.

McCutcheon, Allen L. 1987. *Latent Class Analysis*. Thousand Oaks, CA: SAGE.

Nakazawa, Minato. 2018. “fmsb: Functions for Medical Statistics Book with Some Demographic Data.” https://CRAN.R-project.org/package=fmsb.

**Author Biographies**

jimi adams is an associate professor of health and behavioral sciences at the University of Colorado Denver. His work focuses on how network structure promotes or constrains diffusion of diseases and ideas through a population. He is the author of *Gathering Social Network Data*.

Adam M. Lippert is a sociologist and demographer whose work focuses on health and the intersections of work and family, influences of school and neighborhood contexts on youth health behaviors, and how families manage austerity to protect health. Findings from his research can be found in *Social Forces, Journal of Health & Social Behavior*, and *American Journal of Epidemiology*. 