

A Guide to the LearnPCA Package

David T. Harvey*

Bryan A. Hanson†

2024-04-25

This vignette is based upon `LearnPCA` version 0.3.4.

`LearnPCA` provides the following vignettes:

- Start Here
- A Conceptual Introduction to PCA
- Step By Step PCA
- Understanding Scores & Loadings
- Visualizing PCA in 3D
- The Math Behind PCA
- PCA Functions
- Notes
- To access the vignettes with R, simply type `browseVignettes("LearnPCA")` to get a clickable list in a browser window.

Vignettes are available in both pdf (on CRAN) and html formats (at Github).

Audience

Simply put, the audience for this package is people who want to learn about and more deeply understand Principal Components Analysis (PCA). Let's acknowledge that we are not talking to mathematicians and statisticians. We are talking to people in most any other discipline that want to improve their understanding of PCA. PCA is used in many fields, so we have included examples from several different fields. As PCA is the foundation of a number of related methods, people wanting to learn those variants will need a solid understanding of PCA before continuing.

Why a Package?

In our careers as researchers, we have needed to use PCA, and hence, to understand PCA. In our careers as professors of chemistry, we have taught PCA to students. We can confirm that trying to teach PCA, and eventually getting better at it, is one of the best ways to learn PCA.

Even so, we do not pretend to be experts in PCA, though we are not novices either. We have learned PCA like many of you, as a journey of gradually increasing understanding. We feel like that gives us an advantage in explaining PCA to new learners, as we have been learners in the recent past. In that sense PCA is like a lot of complex topics, it just cannot be fully understood in a single sitting. So plan to read, practice, reflect and repeat.

So why put this material in package? In the course of our own learning and teaching, we have tried to identify the best ways of teaching and illustrating this material. `LearnPCA` is our attempt to gather those materials in an integrated, user-friendly way. And we get a lot of questions from students. We are glad to answer those questions, but we hope that this package will be useful to a wider range of PCA learners.

*Professor of Chemistry & Biochemistry, DePauw University, Greencastle IN USA., harvey@depauw.edu

†Professor Emeritus of Chemistry & Biochemistry, DePauw University, Greencastle IN USA., hanson@depauw.edu

About the Authors

Bryan Hanson is a freelance R consultant and software developer. In a previous life he taught organic chemistry and biochemistry at DePauw University for 32 years. He blames co-author David Harvey for dragging him down the R rabbit hole. You can learn all about him at his web site.

David Harvey is a faculty member in the Department of Chemistry & Biochemistry at DePauw University where he has worked since 1986. He claims no knowledge of rabbit holes and maintains his innocence in luring others into R. You can learn more about his interests at his web site.

Acknowledgements

We have learned what we know from many articles, a few books, many blog posts and questions on StackExchange. Where we rely heavily on particular sources, we will cite them in the appropriate place. We will also point out additional useful resources for further study.

Works Consulted

There are a great many tutorials, case studies and questions about PCA available for study. We have looked at many of them! In the course of preparing this package, we have found the following resources, given in no particular order, especially useful.

Many of these are questions asked at Cross Validated, which form a tremendous corpus for learning. When studying these, be sure to look at the original question, the answers, and the comments. Some answers are more helpful than others, and the dialog can be enlightening as well.

Finally, as you study, remember that a lot of the confusion comes from jargon that differs from discipline-to-discipline. For example, what mathematicians call an “orthonormal basis” is an “orthogonal axis” to statisticians and a “frame of reference” to physicists (with right angles assumed, because, well, of course). Finding these translations really helps to advance one’s understanding.

- Making sense of principal component analysis, eigenvectors & eigenvalues
- What exactly is called principal component in PCA?
- An intuitive guide to PCA: Ideas behind Principal Component Analysis Note: behind paywall. Good discussion of frame of reference concepts.
- Principal component analysis (PCA) in R: which function to use?
- PCA and proportion of variance explained
- Loadings vs eigenvectors in PCA: when to use one or another?
- Simple SVD algorithms: Naive ways to calculate SVD Note: behind paywall.
- Relationship between SVD and PCA. How to use SVD to perform PCA?
- How to reverse PCA and reconstruct original variables from several principal components?
- *Chemometrics: A Practical Guide* by Kenneth R. Beebe, Randy J. Pell and Mary Beth Seasholtz. Wiley, New York, 1998. ISBN 0-471-12451-6
- “A Tutorial on Principal Components Analysis” Jonathan Shlens. 7 April 2014. access.
- Dimensionality Reduction For Dummies — Part 1: Intuition Note: behind paywall.
- You Don’t Know SVD: Truly Understanding SVD — The Intuitive Core Idea Note: behind paywall.
- The Math of Principal Components Analysis
- Principal components analysis Note: behind paywall. Be sure to get the corrected version. There’s also R scripts available.