

The stantargets R package: a workflow framework for efficient reproducible Stan-powered Bayesian data analysis pipelines

William Michael Landau¹

¹ Eli Lilly and Company

DOI: [10.21105/joss.03193](https://doi.org/10.21105/joss.03193)

Software

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

Editor: [Kyle Niemeyer](#) ↗

Reviewers:

- [@kyleniemeyer](#)

Submitted: 15 April 2021

Published: 21 April 2021

License

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

Statement of need

Real-world Bayesian data analysis is a thorough process of investigation and experimentation. Statisticians iteratively refine and compare multiple models to improve inference and understand model behavior ([Gelman et al., 2020](#)). Intermediate results inform downstream model-building decisions, and the final models do not always agree with the initial choices. A typical step of this empirical development process requires a computational method such as Markov chain Monte Carlo to approximate the posterior distribution of the model parameters given the data ([Gelman et al., 2014](#)). Even with fast and flexible probabilistic programming languages like Stan ([Stan Development Team, 2012](#)), this computation can take several minutes or hours to complete, and successive iterations become expensive enough to obstruct the research.

Summary

The [stantargets](#) R package ([Landau, 2021a](#)) reduces the practical burdens of developing and maintaining Bayesian data analysis workflows with Stan. It expresses the models, datasets, and inferential results as interdependent components of a formal pipeline, tracks these components for changes, and automatically reruns the affected steps in response to these changes, optionally with distributed computing on a cluster. If a step is already up to date with its upstream dependencies, [stantargets](#) automatically skips it, potentially saving hours of runtime. When the whole pipeline is up to date, the user has tangible evidence that the output matches the underlying code and data, which affirms reproducibility.

The [stantargets](#) package is an extension of [cmdstanr](#) ([Gabry & Češnovar, 2021](#)), a lightweight interface to Stan, and [targets](#) ([Landau, 2021b](#)), a general-purpose pipeline toolkit for reproducible research and high-performance computing. [stantargets](#) builds [targets](#)-powered pipelines specifically tailored to Bayesian statistics with [cmdstanr](#), from single-run workflows to large-scale simulation studies. Using domain knowledge, [stantargets](#) abstracts away burdensome low-level configuration details and streamlines pipeline construction, freeing Bayesian statisticians to focus less on software development and more on model development. [stantargets](#) is part of the [R Targetopia](#) ([Landau, 2020](#)), an emerging ecosystem of R packages to democratize reproducible analysis pipelines across multiple domains of Statistics and data science.

References

- Gabry, J., & Češnovar, R. (2021). *Cmdstanr: R interface to 'CmdStan'*. <https://mc-stan.org/cmdstanr>.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (F. Dominici, J. J. Faraway, M. Tanner, & J. Idek, Eds.; Third). CRC Press.
- Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., & Modrák, M. (2020). *Bayesian workflow*. <http://arxiv.org/abs/2011.01808>
- Landau, W. M. (2020). *The R Targetopia: An R package ecosystem for democratized reproducible pipelines at scale*. <https://wlandau.github.io/targetopia>.
- Landau, W. M. (2021a). *Stantargets: Targets for Stan workflows*. <https://docs.ropensci.org/stantargets/>.
- Landau, W. M. (2021b). The targets R package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. *Journal of Open Source Software*, 6(57), 2959. <https://doi.org/10.21105/joss.02959>
- Stan Development Team. (2012). *Stan: A C++ library for probability and sampling*. <https://mc-stan.org>.