The Journal of Open Source Software

tidyhte: Tidy Estimation of Heterogeneous Treatment Effects in R

Drew Dimmery [1]

1 Hertie School Data Science Lab

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor: Open Journals

Reviewers:

- @openjournals

Submitted: 01 January 1970
Published: unpublished

Summary

Heterogeneous treatment effects (HTE) describe how the impact of an intervention varies across individuals or subgroups. Understanding this variation is crucial for targeting interventions, optimizing resource allocation, and gaining insights into mechanisms of action. `tidyhte` is an R package that provides a principled, user-friendly framework for estimating heterogeneous treatment effects using modern machine learning methods. The package implements the doubly-robust learner of Kennedy (2023), which combines outcome modeling and propensity score estimation with cross-validation to produce valid statistical inference on treatment effect heterogeneity.

The package uses a "recipe" design that makes it straightforward to scale analyses from single outcomes and moderators to many outcomes and moderators. Users specify machine learning algorithms for nuisance function estimation (treatment propensity and outcome models), define moderators of interest, and `tidyhte` handles cross-validation, model selection, diagnostics, and construction of relevant quantities of interest. The tidy data principles make it easy to chain operations and integrate with the broader R data science ecosystem.

Statement of Need

While heterogeneous treatment effect estimation has become increasingly important in applied research, existing tools often require substantial statistical expertise to implement correctly. Researchers must navigate complex decisions about cross-validation strategies, model selection, doubly-robust estimation, and valid inference. Many implementations are scattered across different packages with inconsistent interfaces, making it difficult to apply these methods reliably and at scale.

`tidyhte` addresses these challenges by providing a unified interface for HTE estimation that:

1. Implements state-of-the-art doubly-robust methods with automatic cross-validation
2. Uses intuitive "recipe" semantics familiar to R users from packages like `recipes` and `parsnip`
3. Handles both experimental and observational data
4. Scales easily from single to multiple outcomes and moderators
5. Provides built-in diagnostics for model quality and effect heterogeneity
6. Returns results in tidy data formats for easy downstream analysis and visualization

The package is designed for applied researchers in social sciences, policy evaluation, medicine, and technology companies who need to understand treatment effect variation but may not be experts in the underlying statistical methods. By automating many technical details while maintaining statistical rigor, `tidyhte` makes modern HTE methods accessible to a broader audience.

Key Features

- **Doubly-robust estimation**: Combines outcome and propensity score models using Kennedy (2023) methodology
- **Flexible ML integration**: Works with any SuperLearner-compatible algorithm
- **Multiple moderator types**: Supports stratified, kernel smooth, and custom moderator specifications
- **Built-in diagnostics**: Includes model quality metrics and tests for effect heterogeneity
- **Variable importance**: Optional variable importance measures for identifying key moderators
- **Tidy semantics**: Integrates with dplyr and tidyverse workflows
- **Cross-validation**: Automatic sample splitting with user control over fold numbers

Research Applications

`tidyhte` has been designed to support research in fields where understanding treatment effect heterogeneity is important:

- **Clinical trials**: Identifying patient subgroups that benefit most from treatments
- **Policy evaluation**: Understanding which populations are most affected by policy interventions
- **Technology**: Optimizing product interventions for different user segments
- **Economics**: Studying heterogeneous effects of economic policies across demographics
- **Education**: Evaluating differential impacts of educational interventions

It was created specifically as the HTE estimation software for use in the high-profile series of publications on the effects of Facebook and Instagram on the US 2020 Presidential election and was stress tested for this application which involved estimation of HTEs on around ten moderators across dozens of outcomes (Guess et al., 2023a, 2023b; Nyhan et al., 2023).

Related Work

The package builds on the theoretical framework of Kennedy (2023) for doubly-robust HTE estimation. It leverages the SuperLearner ensemble learning framework (Van der Laan et al., 2007) for flexible machine learning. The tidy design philosophy follows Wickham (2014) and integrates with the broader tidyverse ecosystem (Wickham et al., 2019).

Other R packages for HTE estimation include `grf` (Athey et al., 2019) for generalized random forests and `bcf` (Hahn et al., 2020) for Bayesian additive regression trees. `tidyhte` differentiates itself through its doubly-robust methodology, recipe interface, and focus on scaling to multiple outcomes and moderators.

Acknowledgements

We acknowledge contributions from early users who provided feedback on the package design and functionality.

References

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., & others. (2023a). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, *381*(6656), 398–404.

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., & others. (2023b). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, *381*(6656), 404–408.

Hahn, P. R., Murray, J. S., & Carvalho, C. M. (2020). Bayesian regression tree models for causal inference: Regularization, confounding, and heterogeneous effects. *Bayesian Analysis*, *15*(3), 965–1056. https://doi.org/10.1214/19-BA1195

Kennedy, E. H. (2023). Towards optimal doubly robust estimation of heterogeneous causal effects. *Electronic Journal of Statistics*, *17*(2), 3008–3049. https://doi.org/10.1214/23-EJS2157

Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., & others. (2023). Like-minded sources on facebook are prevalent but not polarizing. *Nature*, *620*(7972), 137–144.

Van der Laan, M. J., Polley, E. C., & Hubbard, A. E. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1). https://doi.org/10.2202/1544-6115.1309

Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, *59*(10), 1–23. https://doi.org/10.18637/jss.v059.i10

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686