




RgentAI: An AI-Powered Agent System for Guided Statistical Analysis in R

Nathan Bresette ¹ and Dane Winterboer ²

¹ Independent Researcher ² Independent Researcher

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

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

Editor: 

Submitted: 30 September 2025

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

Summary

Statistical analysis in R requires extensive domain knowledge, creating barriers for researchers. RgentAI provides an AI assistant with four specialized agents (data cleaning, transformation, statistical analysis, and machine learning) plus three core features: conversational chat, plot analysis, and error debugging.

The system integrates with RStudio through an R package, offering real-time assistance. Features include: (1) conversational AI chat, (2) automatic plot analysis, (3) intelligent error debugging, (4) guided workflow automation, and (5) statistical guidance. RgentAI accelerates RStudio workflows through intelligent automation.

The software targets domain scientists, students, and researchers seeking to accelerate analytical workflows. Early evaluation demonstrates improvements in analysis efficiency and coding speed. The system operates through a managed cloud service ensuring sustainability while maintaining academic accessibility.

Statement of need

Modern research relies heavily on statistical analysis, yet choosing appropriate methods creates barriers for domain experts. Researchers must navigate data preprocessing, statistical tests, effect size calculations, and modeling approaches—requiring specialized knowledge outside their expertise (Wickham, 2014; Wilson et al., 2014).

Existing solutions are either too general (base R functions) or too specialized (domain-specific packages). General tools require extensive statistical knowledge, while specialized packages lack complete workflow integration. This forces researchers to piece together solutions from multiple sources.

RgentAI fills this gap by providing intelligent AI assistance throughout the analytical process. Unlike existing tools, it combines workflow automation with conversational AI support to accelerate development while ensuring methodological appropriateness. The target audience includes domain scientists, students, and researchers seeking to maximize analytical productivity.

Software description

RgentAI implements an AI assistance system for R programming with general AI support features and specialized analytical agents:

Core AI Features

Conversational AI Chat provides assistance for R programming questions, code explanation, debugging help, and methodological guidance. Users ask natural language questions about R

syntax, statistical concepts, or data analysis approaches, receiving contextual responses with executable code examples.

Analyze Last Plot identifies and analyzes the most recent plot command from R history, providing intelligent interpretation of visualizations including patterns, outliers, distribution characteristics, and relationships. The system suggests visualization improvements and identifies potential data quality issues.

Debug Last Error monitors R console errors and provides intelligent diagnosis and solutions. When R code fails, the system analyzes the error message, examines code context, and suggests specific fixes with explanations.

Specialized Analytical Agents

Data Cleaning Agent automates detection and handling of missing values, outliers, duplicates, and data type inconsistencies. The agent provides intelligent recommendations based on data characteristics, such as median imputation for skewed numeric variables or mode imputation for categorical variables (Wickham, 2014).

Transformation Agent provides automated data transformations, including mathematical transformations based on skewness analysis (automatic log transformation for highly skewed data), new variable creation through feature engineering, and data reshaping using tidyverse principles (Wickham et al., 2019).

Statistical Analysis Agent offers intelligent method selection for hypothesis testing, effect size calculation, and exploratory analysis. The system automatically selects appropriate tests based on data types and sample characteristics (e.g., Shapiro-Wilk for normality testing with small samples, Anderson-Darling for larger samples).

Modeling Agent provides smart algorithm recommendations based on target variable characteristics, automatically suggesting logistic regression for binary classification, multinomial regression for multi-class problems, and linear regression for continuous outcomes. The agent handles feature engineering, model interpretation, and performance evaluation.

The system's intelligence lies in context-aware decision making. For example, when analyzing categorical variables, the system examines unique categories and automatically suggests binary classification methods for two-category variables and multi-class methods for three or more categories, preventing methodological errors while educating users.

Architecture: The software consists of an open-source R package providing the agent framework and user interface, integrated with a cloud-based AI service for natural language processing and code generation. The R package functions independently for basic workflows, with AI enhancement available through managed access codes.

Installation and Usage: RgentAI is installed from GitHub using `devtools::install_github("NathanBresette/rgentai")`, `force = TRUE`, `upgrade = "never"`). Users launch the assistant with `run_rgent()` and enter their access code to activate AI features. The system integrates seamlessly with RStudio (RStudio Team, 2024), providing real-time assistance through the familiar development environment.

Comparison to existing tools: Unlike general statistical packages (R base (R Core Team, 2024), SPSS) requiring extensive expertise, or specialized tools (mlr3, caret) focusing on specific domains, RgentAI provides comprehensive AI assistance across all aspects of R programming and analysis. The system offers conversational support, visual analysis, and error debugging—capabilities not found in traditional statistical software.

Integration: The software integrates with RStudio through a native add-in, providing comprehensive AI assistance. All AI interactions occur within the RStudio interface, accelerating coding workflows. Generated code can be inserted directly into R scripts with a single click, eliminating copy-paste overhead. The system automatically monitors R console errors and

85 analyzes recent plot commands, providing instant contextual assistance without interrupting
86 development flow.

87 Impact and ongoing research

88 RgentAI enhances productivity and accelerates development workflows in R programming and
89 statistical analysis. Early adoption demonstrates improvements in coding speed, analytical
90 efficiency, and development cycle times. The system's productivity enhancements are valuable
91 for time-sensitive research projects and intensive data analysis workflows.

92 The AI-powered architecture provides a foundation for extending intelligent assistance to
93 other aspects of scientific computing. Current development focuses on expanding productivity
94 features, reducing time-to-insight, and improving integration with the R ecosystem. This
95 aligns with reproducible research practices (Peng, 2011) by providing tools that enhance both
96 productivity and methodological rigor.

97 The open-source R package encourages community contributions and customization, while
98 the managed cloud service ensures sustainable operation and continuous improvement of AI
99 capabilities.

100 Acknowledgements

101 We acknowledge the R community for creating the foundation that makes this work possible,
102 and the developers of the statistical packages that inform our methodological recommendations.

103 References

- 104 Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060),
105 1226–1227. <https://doi.org/10.1126/science.1213847>
- 106 R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation
107 for Statistical Computing. <https://www.R-project.org/>
- 108 RStudio Team. (2024). *RStudio: Integrated development environment for r*. RStudio, PBC.
109 <http://www.rstudio.com/>
- 110 Wickham, H. (2014). Tidy data. *Journal of Statistical Software*, 59(10), 1–23. <https://doi.org/10.18637/jss.v059.i10>
- 112 Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Golemund,
113 G., Hayes, A., Henry, L., Hester, J., & others. (2019). Welcome to the tidyverse. *Journal*
114 *of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- 115 Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H.,
116 Huff, K. D., Mitchell, I. M., Plumbley, M. D., & others. (2014). Best practices for scientific
117 computing. *PLoS Biology*, 12(1), e1001745. <https://doi.org/10.1371/journal.pbio.1001745>