The Journal of Open Source Software

Data Science for JavaScript with tangent/ds

Essi Parent [1,2]¶

1 Département des sols et du génie agroalimentaire, Université Laval, Québec, Canada **2** Ministère de l'Environnement, de la Lutte contre les Changements Climatiques, de la Faune et des Parcs, Québec, Canada ¶ Corresponding author

DOI: 10.xxxxxx/draft

Software

- Review
- Repository
- Archive

Editor:

Submitted: 13 February 2026
Published: unpublished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).

Summary

tangent/ds is a JavaScript library that provides a unified data science toolkit for browsers and server-side JavaScript environments. It implements multivariate analysis methods, data clustering, statistical modeling, machine learning, and in pure ESM JavaScript, filling a gap in the JavaScript ecosystem where data wrangling and visualization tools exist but statistical and analytical capabilities remain fragmented or absent. The library offers five integrated modules: `core` (linear algebra, optimization, formula parsing), `stats` (distributions, generalized linear models, hypothesis tests), `ml` (clustering, classification, regression, neural networks, cross-validation), `mva` (principal component analysis, linear discriminant analysis, redundancy analysis, canonical correspondence analysis, hierarchical clustering), and `plot` (Observable Plot configuration generators for diagnostic and ordination plots).

Statement of need

The JavaScript data science ecosystem already has mature libraries for data wrangling and visualization. ml-matrix (ml.js contributors, 2024b) provides fast matrix operations, simple statistics (MacWright, 2012) a collection of statistical testing, Arquero (Heer, 2021) and Danfojs (Bamigbade & Adeyemi, 2020) dplyr-inspired tabular transformations, as well as Observable Plot (Bostock & Observable, Inc., 2021) a concise, high-level visualization. However, a critical gap remains: none of these tools provide a comprehensive, one-stop API for statistical modeling, hypothesis testing, or multivariate analysis. Typically, a data scientist working in an Observable notebook can import, clean, reshape, and plot data entirely in JavaScript, leave the browser environment to an R or Python environment fit a generalized linear model, run a PCA, or perform an ANOVA, and either postprocess data away from JavaScript, either export everything to get back to Observable for another round of visualization. tangent/ds eliminates this context-switching by providing, directly in JavaScript, R-level statistical (R Core Team, 2025) and multivariate data analyses (Oksanen et al., 2024), as well as Scikit-Learn-level (Pedregosa et al., 2011) machine learning.

The library targets four audiences: (1) students, self-learners and teachers who need a reliable computing environment to learn and teach data science without complicated computer setups, (2) researchers using Observable notebooks in need of analytical methods beyond data wrangling, (3) developers building browser-based data applications that require embedded data computations without server round-trips, and (4) data scientists who prefer or require JavaScript environments (Deno, Node.js) for their analysis pipelines.

DRAFT

The Journal of Open Source Software

State of the field

Several JavaScript libraries address individual areas of the data science workflow. TensorFlow.js (Google Brain Team, 2018) brings deep learning to the browser with GPU-accelerated tensor operations, but targets neural network inference and training rather than classical statistics or exploratory data analysis. ml.js (ml.js contributors, 2024a) provides a collection of machine learning algorithms (k-nearest neighbors, PCA, naive Bayes) but does not offer formula-based modeling, generalized linear models, or statistical testing. Neither library provides the integrated analytical workflow, from hypothesis testing through model diagnostics, that R and Python users expect.

No existing JavaScript library offers the combination of GLM/GLMM fitting with formula syntax, statistical distributions with PDF/CDF/quantile functions, model comparison via AIC/BIC and likelihood ratio tests, ordination methods (PCA, LDA, RDA, CCA), and integrated diagnostic plotting, which together constitute the standard analytical workflow in R or Python. tangent/ds was built to fill this gap.

Software design

tangent/ds is structured as five modules that mirror the conceptual organization of a data analysis workflow, from data manipulation (`core`) through statistical modeling (`stats`), machine learning (`ml`), multivariate analysis (`mva`), and visualization (`plot`). API design decisions reflect the library's intended use in browser-based and Observable environments.

- **Minimal dependencies.** The library relies on well-established JavaScript packages (ml-matrix for linear algebra, simple-statistics for basic statistics and Arquero for data manipulation) and requires no WebAssembly compilation or bundler configuration.
- **Observable Plot integration.** The `plot` module generates configuration objects for Observable Plot rather than rendering graphics directly. This design composes naturally with Observable notebooks, where the user calls `Plot.plot(config)` to render, and avoids coupling the library to a specific rendering backend.
- **Tidy and formula syntax**. The API is inspired by Scikit-Learn's pipeline approach (Pedregosa et al., 2011), and by tidyverse (Wickham et al., 2019) chaining ethos. The GLM class accepts Wilkinson-Rogers formula notation (`y ~ x1 + x2`), lowering the barrier for researchers transitioning from R. Formulas are parsed at fit time and resolved against a data object, enabling a concise and familiar modeling interface.
- **Compositional data awareness.** Often overlooked but of important impact in data science, log-ratio transformations for compositional variables are implemented in the core module, reflecting the author's research in geosciences and ecology where compositional data analysis (Pawlowsky-Glahn et al., 2015) is routine but rarely supported in general-purpose libraries.
- **Numerical validation against R and Python.** The test suite includes cross-language validation tests (`tests_compare-to-R/`, `tests_compare-to-python/`) that verify numerical outputs against R's base packages (R Core Team, 2025), R's vegan package (Oksanen et al., 2024) and Scikit-Learn (Pedregosa et al., 2011), ensuring that results are reproducible across ecosystems.

For example, fitting a linear model uses a syntax familiar to R users.

```
import { GLM } from "@tangent.to/ds";
const simple_lm = new GLM({ family: "gaussian", link: "identity" })
  .fit("`Body Mass (g)` ~ `Beak Length (mm)`", penguinsData);
console.log(simple_lm.summary());
```

DRAFT

Research impact statement

tangent/ds is currently used as lecture notes in the Observable collection Data science with `tangent/ds`, and by the author in research in agroenvironmental science at Université Laval. The library enables fully browser-based analytical workflows in Observable notebooks for research applications including soil compaction modeling and compositional analysis of environmental data.

The library is published on npm as `@tangent.to/ds` and on JSR as `@tangent/ds`, with documentation hosted at tangent-to.github.io/ds/. The repository includes working examples, a complete API reference, and cross-language validation tests. While the library is at an early stage of community adoption, it addresses a documented gap in the JavaScript ecosystem (Gans et al., 2020) and is designed for immediate use in Observable notebooks, one of the most widely adopted platforms for browser-based data analysis.

AI usage disclosure

Claude Code was used during development for code generation assistance and debugging. All AI-generated code was reviewed, tested, and validated against R and Python reference implementations. This paper was reviewed by AI for language and references. All scientific content and claims were authored and verified by the author.

References

Bamigbade, R. O., & Adeyemi, S. O. (2020). *Danfo.js: JavaScript library for data analysis*. https://github.com/javascriptdata/danfojs

Bostock, M., & Observable, Inc. (2021). *Observable Plot: A high-level visualization library*. https://observablehq.com/plot/

Gans, M., Hodges, T., & Wilson, G. (2020). *JavaScript for data science*. https://third-bit.com/js4ds/

Google Brain Team. (2018). *TensorFlow.js: Machine learning for JavaScript*. https://www.tensorflow.org/js

Heer, J. (2021). Arquero: A JavaScript library for query processing of array-backed data tables. *Journal of Open Source Software*, *6*(65), 3501. https://doi.org/10.21105/joss.03501

MacWright, T. (2012). *Simple-statistics: Statistical methods in readable JavaScript*. https://github.com/simple-statistics/simple-statistics

ml.js contributors. (2024a). *Ml.js: Machine learning tools in JavaScript* (Version 6.11.0). https://github.com/mljs/ml

ml.js contributors. (2024b). *Ml-matrix: Matrix manipulation tools* (Version 6.11.0). https://github.com/mljs/matrix

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., … Weedon, J. (2024). *Vegan: Community ecology package*. https://CRAN.R-project.org/package=vegan

Pawlowsky-Glahn, V., Egozcue, J. J., & Tolosana-Delgado, R. (2015). *Modelling and analysis of compositional data* (1st ed.). Wiley. https://doi.org/10.1002/9781119003144

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python.

DRAFT

Journal of Machine Learning Research, *12*(85), 2825–2830. https://jmlr.org/papers/v12/pedregosa11a.html

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.2) [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

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

DRAFT