The R Project: A Quantitative Dossier of Impact

Verifiable evidence of how a statistical platform became the computational backbone of modern science

Balasubramanian Narasimhan

2026-03-04

Table of contents

1 Summary of evidence

The R project is an open-source language and environment for statistical computing that has grown, over three decades, from an academic experiment into infrastructure on which a substantial fraction of the world's quantitative research depends. The numbers below, each drawn from a verifiable primary source and dated at time of access, convey the scale.

Table 1: Key metrics at a glance (all retrieved 2026-03-03 unless noted). [a]CRAN. [b]CRAN Task Views. [c]Bioconductor 3.22 release. [d]Bioconductor download stats. [e]Michelaki et al. (2025). [f]Lai et al. (2023). [g]Crossref. [h]r-devel/r-svn.

Indicator	Value
CRAN packages[a]	23,304
CRAN Task Views[b]	49
Bioconductor packages (rel. 3.22)[c]	2,361
Bioconductor downloads, 2025[d]	98.8 M
Bioconductor distinct IPs, 2025[d]	1.88 M
Econometrics R packages[e]	207
R usage in biodiversity journals, 2022[f]	70.6%
Crossref citations of Ihaka & Gentleman (1996)[g]	8,813
R core SVN commits[h]	63,934

These figures reflect a single innovation—a free, extensible platform for statistical computation—whose reach now spans genomics, econometrics, clinical trials, ecology, finance, spatial science, psychometrics, sports analytics, and virtually every other discipline that relies on data.

2 The R project as a statistical innovation

R began in the early 1990s as a reimplementation of the S language (Ihaka & Gentleman, 1996), itself a pioneering environment for data analysis designed at Bell Laboratories (Chambers, 1998; Chambers & Hastie, 1992). What transformed R from an academic reimplementation into a global innovation was a combination of deliberate design decisions: the language was released under a free software license; it was designed to be extended by users through a package system; and it was governed by a small, technically expert core team that maintained rigorous standards for the base system while leaving the contributed ecosystem radically open.

The R Core Team and the R Foundation have sustained this governance model for over twenty-five years, publishing an official contributor list (R Foundation for Statistical Computing, 2026) and maintaining a disciplined release cadence of annual major releases with patch releases as needed (R Developer Page, 2026). Cross-sector institutional support has been organized through the R Consortium, a Linux Foundation project whose mission is to fund technical and social infrastructure for the R ecosystem (R Consortium, 2026). The public SVN mirror records 63,934 commits, a transparent trail of sustained engineering (r-devel organization, 2026).

3 CRAN: a curated ecosystem without parallel

The Comprehensive R Archive Network (CRAN) is the primary distribution channel for contributed R packages. As of March 2026, it hosts **23,304 packages** (Comprehensive R Archive Network, 2026a). What distinguishes CRAN from a mere repository is its quality-control regime: every submitted package must pass automated checks on multiple operating systems (Linux, macOS, Windows), and these check results are published continuously (Comprehensive R Archive Network, 2026a). This is not a rubber stamp; packages that fail checks are archived. The result is an ecosystem where a researcher can install a package and have reasonable confidence that it builds, loads, and passes its own tests on the major platforms.

3.1 Writing R Extensions: the on-ramp for package development

A critical but easily overlooked piece of infrastructure is the *Writing R Extensions* manual, maintained by R Core and shipped with every R release (R Core Team, 2026). This document specifies, in exacting detail, how to structure a package: the `DESCRIPTION` file, the namespace mechanism, documentation via Rd files, vignettes, test suites, compiled code interfaces, and the metadata required for CRAN submission. It is not a tutorial; it is a contract. Any statistician who develops a new method and wishes to distribute it through CRAN must conform to this specification, and the automated check system enforces compliance.

The effect has been profound. By codifying package structure in a single, authoritative reference, *Writing R Extensions* lowered the barrier for domain scientists—who are statisticians or biologists or economists, not software engineers—to produce distributable, documented, testable software. The 23,304 packages on CRAN are, in a real sense, the downstream output of this manual. No other statistical computing environment has produced a comparable on-ramp from "I wrote a useful function" to "my method is a citable, installable, cross-platform package."

3.2 Packages as citable scholarly objects: the CRAN DOI system

CRAN assigns a persistent Digital Object Identifier (DOI) to each package—for example, `https://doi.org/10.32614/CRAN.package.lme4` for `lme4` (Comprehensive R Archive Network, 2026b). This DOI identifies the *package itself as a scholarly artifact*, distinct from the DOI of any journal article that describes the package's methodology. The distinction matters: it means that a researcher who uses `lme4` can cite the software (via the CRAN DOI) and the method (via the JSS paper by Bates et al. (2015)) separately, giving appropriate credit to both the implementation and the theory.

This two-DOI structure—one for the method paper, one for the software—embodies a principle that the R ecosystem pioneered and that the broader scientific community is only now catching up to: **statistical software is a first-class research contribution**, not merely an appendage to a journal article. CRAN's DOI policy makes this principle machine-readable and citation-trackable.

3.3 Task Views: disciplinary maps across modern science

CRAN's 49 Task Views are curated, versioned topic pages, each maintained by a domain expert, that organize the package landscape by application area (CRAN, 2026a; Zeileis, 2025). Their breadth is itself evidence of R's penetration across disciplines:

- **Medicine and biostatistics:** ClinicalTrials, Epidemiology, MedicalImaging, Pharmacokinetics, Survival, MetaAnalysis, Omics
- **Environment and earth sciences:** Environmetrics, Hydrology, Paleontology, Spatial, SpatioTemporal
- **Social sciences and psychology:** OfficialStatistics, Psychometrics, SportsAnalytics
- **Economics and finance:** Econometrics, Finance
- **Biology:** Phylogenetics, Agriculture
- **Core methodology:** Bayesian, CausalInference, Cluster, Distributions, ExperimentalDesign, ExtremeValue, FunctionalData, GraphicalModels, MissingData, MixedModels, Robust, CompositionalData
- **Machine learning:** MachineLearning, AnomalyDetection, NaturalLanguageProcessing
- **Computation and infrastructure:** DifferentialEquations, HighPerformanceComputing, NumericalMathematics, Optimization, Databases, WebTechnologies
- **Visualization and reproducibility:** DynamicVisualizations, ReproducibleResearch, TeachingStatistics, ModelDeployment
- **Other domains:** ChemPhys, ActuarialScience, TimeSeries, Tracking, NetworkAnalysis

It is difficult to name a quantitative discipline that does *not* have a Task View. This is not happenstance; it reflects three decades of methodologists packaging their innovations as freely available R software, creating a feedback loop between statistical research and statistical practice.

4 Scholarly infrastructure: The R Journal and the Journal of Statistical Software

4.1 The R Journal

The R Journal (ISSN 2073-4859) is the peer-reviewed, open-access journal of the R project, published by the R Foundation (The R Foundation for Statistical Computing, 2026). It succeeded *R News* (2001–2008) and publishes articles describing new R packages, statistical methodology implemented in R, and practical computing techniques. The journal serves a distinctive function in the scientific ecosystem: it is the venue where a statistical innovation, packaged as software, receives formal peer review not just of the method but of the implementation. This tight coupling between method and code is central to the reproducibility culture that R has fostered.

4.2 Journal of Statistical Software

The Journal of Statistical Software (JSS; ISSN 1548-7660), founded in 1996 by Jan de Leeuw, is the premier open-access venue for papers describing statistical software (Foundation for Open Access Statistics, 2026). With an impact factor of 8.11 (2024) and an h-index of 187, JSS is among the highest-impact journals in statistics—a remarkable fact for a journal whose articles are about *software*. The vast majority of JSS publications describe R packages: `lme4` (Bates et al., 2015), `ggplot2` (Wickham, 2016), `glmnet` (Friedman et al., 2010), `sandwich` (Zeileis, 2004, 2006), `plm` (Croissant & Millo, 2008), `coin` (Hothorn et al., 2008), and `highfrequency` (Boudt et al., 2022) are representative examples. The existence of JSS has created a norm—now widely imitated—in which releasing a statistical method without a peer-reviewed software paper feels incomplete.

5 The textbook–package paradigm

A phenomenon essentially unique to R is the tight coupling between canonical statistics textbooks and companion CRAN packages. This model, in which a textbook ships with a package containing datasets, functions, and reproducible examples, has transformed how statistics is taught and practiced worldwide. Notable examples include:

Table 2: Canonical statistics textbooks with companion R packages.

Textbook	Authors	Package	Note
Modern Applied Statistics with S	Venables & Ripley	`MASS`	Shipped with base R
Mixed-Effects Models in S and S-PLUS	Pinheiro & Bates	`nlme`	Shipped with base R
An Introduction to Statistical Learning (2nd ed.)	James, Witten, Hastie & Tibshirani	`ISLR2`	Used in hundreds of university courses worldwide
Applied Econometrics with R	Kleiber & Zeileis	`AER`	100+ datasets
An R Companion to Applied Regression (3rd ed.)	Fox & Weisberg	`car`	
Generalized Additive Models (2nd ed.)	Wood	`mgcv`	Shipped with base R
Modeling Survival Data	Therneau & Grambsch	`survival`	Shipped with base R
Bootstrap Methods and Their Application	Davison & Hinkley	`boot`	Shipped with base R
Lattice: Multivariate Data Visualization with R	Sarkar	`lattice`	Shipped with base R
Introductory Statistics with R (2nd ed.)	Dalgaard		
The Elements of Statistical Learning (2nd ed.)	Hastie, Tibshirani & Friedman	`ElemStatLearn`	Freely available online
Pattern Recognition and Neural Networks	Ripley	`nnet`, `MASS`	`nnet` shipped with base R

Several of these packages—`MASS`, `nlme`, `survival`, `boot`, `mgcv`, `lattice`, `nnet`—are so fundamental that they ship with every R installation. This means that a student who installs R immediately has access to the reference implementations underlying some of the most widely used textbooks in statistics. The textbook–package model has no parallel in any other computing ecosystem: it ensures that the canonical teaching of statistical methods is inseparable from executable, testable code.

The *Introduction to Statistical Learning* (ISLR) deserves special mention. Freely available online and adopted in hundreds of university courses, it has become the entry point for a generation of data scientists. Its companion package `ISLR2` (James et al., 2021) embeds every dataset in the book, and the book's code examples are in R. This single textbook has done more to establish R as the language of statistical education than perhaps any other publication since Ihaka & Gentleman (1996).

6 Bioconductor: R as the platform for computational biology

Bioconductor is the largest domain-specific extension of R, providing a coordinated ecosystem for computational biology and bioinformatics built around interoperable data structures, standardized annotation resources, and rigorous package review (Gentleman et al., 2004; Huber et al., 2015).

Bioconductor 3.22 (October 2025) ships 2,361 software packages, 435 experiment-data packages, 926 annotation packages, 29 workflows, and 6 books (Bioconductor, 2025a). Growth has been sustained since the project's founding in 2002 (Bioconductor, 2025b). In 2025 alone, the Bioconductor repository recorded **98.8 million downloads** from **1.88 million distinct IP addresses**, with public mirror traffic excluded (Bioconductor, 2026).

The scholarly impact of Bioconductor's platform papers is exceptional. The founding paper (Gentleman et al., 2004) and the 2015 *Nature Methods* paper (Huber et al., 2015) are among the most cited articles in computational biology (National Library of Medicine, 2004, 2015). Their citation trajectories track the adoption of high-throughput genomics itself: as sequencing became ubiquitous, R—via Bioconductor—became the default analytical environment.

7 Domain adoption: quantitative evidence

7.1 Biodiversity and ecology

A bibliometric study of eight leading biodiversity conservation journals found that 42.3% of articles published between 2008 and 2022 explicitly used R, with usage rising from 11.1% in 2008 to **70.6% in 2022** (Lai et al., 2023). The ten most-used packages—`vegan`, `lme4`, `MuMIn`, `nlme`, `mgcv`, `raster`, `MASS`, `ggplot2`, `car`, and `dismo`—reflect the full spectrum from mixed models and model selection to spatial analysis and visualization.

7.2 Econometrics

A 2025 systematic review identifies and analyzes **207 econometrics-related R packages** using CRAN and the Econometrics Task View, examining maintenance patterns, documentation quality, reverse dependencies, and publication outcomes (Michelaki et al., 2025). Canonical econometrics packages have peer-reviewed JSS papers: `plm` for panel data (Croissant & Millo, 2008), `sandwich` for robust covariance estimation (Zeileis, 2004, 2006), and `strucchange` for structural-change monitoring (Zeileis et al., 2005). The textbook *Applied Econometrics with R* and its companion package `AER` (CRAN, 2026b; Kleiber & Zeileis, 2008) anchor this ecosystem in pedagogy.

7.3 Finance

CRAN's Empirical Finance Task View, established around 2004, curates packages for portfolio analytics, risk modeling, volatility estimation, and high-frequency data (Eddelbuettel, 2009, 2026). The R/Finance conference series, documented in *The R Journal*, draws practitioners from portfolio management, market microstructure, and quantitative risk (Ulrich, 2013). Peer-reviewed package papers such as `highfrequency` (Boudt et al., 2022) illustrate the depth of R's finance tooling.

7.4 Spatial statistics

R has become the reference platform for spatial data analysis. Bivand (2022) provides a comparative case study demonstrating R's package ecosystem for areal data. The Spatial and SpatioTemporal Task Views organize hundreds of packages for geostatistics, point pattern analysis, remote sensing, and spatial econometrics.

7.5 Sports analytics

A systematic PRISMA-compliant review identifies 81 sport-related R packages on CRAN, covering basketball, soccer, physical activity, and more (Casals et al., 2023). The SportsAnalytics Task View further organizes this growing domain.

8 R's influence on the Python data-science ecosystem

R's impact extends beyond its own user base: it has shaped the design of Python's data-science stack in ways that the Python tool authors themselves have documented.

The DataFrame concept. Wes McKinney's `pandas` library—the foundation of tabular data analysis in Python—was designed with explicit reference to R's `data.frame`. In his founding paper at SciPy 2010, McKinney discusses pandas with "comparisons with the R language" and the keyword list includes "R" (McKinney, 2010). The `DataFrame` name itself is borrowed from R. The `groupby` pattern in pandas mirrors R's split-apply-combine paradigm.

The formula interface. The `patsy` library, which powers model specification in Python's `statsmodels`, describes itself as "closely inspired by and compatible with the formula mini-language used in R and S" (Smith, 2026). The `y ~ x1 + x2` notation that R inherited from S (Chambers & Hastie, 1992) has thus propagated into the Python statistics ecosystem.

The grammar of graphics. `plotnine`, the leading declarative visualization library in Python, describes itself as "an implementation of a grammar of graphics in Python based on ggplot2" (Kibirige, 2026). R's `ggplot2` (Wickham, 2016), itself an implementation of Wilkinson's grammar (Wilkinson, 2005), has become the template for how high-level visualization APIs are designed across languages.

Community peer review and software as scholarship. rOpenSci pioneered open peer review of scientific software packages, treating the package itself—not just a companion paper—as a reviewable scholarly contribution (Boettiger et al., 2015; Ram et al., 2019). Reviewers evaluate code quality, documentation, API design, and test coverage, producing feedback that improves the software and signals trustworthiness to users. pyOpenSci, founded in 2018, adopted this model wholesale, explicitly crediting rOpenSci's process as its template (pyOpenSci, 2026) and extending R's innovation in software peer review into the Python ecosystem.

These are not indirect or arguable influences. In each case, the Python tool's own documentation or founding publication names R as the direct inspiration.

9 Reproducible research: from Sweave to Quarto

R has been at the center of the reproducible-research movement since Friedrich Leisch introduced Sweave in 2002 (Leisch, 2002), building on Knuth's literate-programming concept (Knuth, 1984). `knitr` generalized the approach to multiple output formats (Xie, 2014), and R Markdown became the de facto standard for combining code, prose, and results in a single document. The latest evolution, Quarto—which this very document is written in—was designed by Posit (formerly RStudio) and supports R, Python, Julia, and Observable, but its intellectual lineage runs directly through R's reproducibility toolchain. The expectation that a statistical analysis should be a reproducible narrative document, not a set of disconnected scripts and screenshots, is an idea that R's community normalized.

10 Software supply-chain evidence

Empirical software-engineering research has begun to study R's package ecosystem as a scientific software supply chain. Malviya-Thakur et al. (2023) surveyed R developers to understand how they choose packages, finding that documentation quality, maintenance signals, and community reputation drive adoption—all features that CRAN's infrastructure (check pages, reverse-dependency counts, vignettes) makes visible. This study, published at ACM/IEEE ESEM, treats R's ecosystem as a case study in how scientific communities build and maintain software trust at scale.

11 Citation signals

Citation counts are imperfect, but their order of magnitude is informative. Crossref reports **8,813** "is-referenced-by" records for Ihaka & Gentleman (1996) as of March 2026 (Crossref, 2026). The Bioconductor platform papers (Gentleman et al., 2004; Huber et al., 2015) each carry thousands of PubMed citations (National Library of Medicine, 2004, 2015). Li et al. (2017) examine R citation practices in PLoS journals and highlight the challenge of consistently citing software at the right granularity (language vs. package vs. function), suggesting that formal citation counts understate actual usage.

12 Regulated industries: pharma and finance

The R Validation Hub, a cross-industry consortium, supports the use of R in GxP-regulated pharmaceutical environments (R Validation Hub, 2024). Major pharmaceutical companies (Roche, Novartis, Pfizer) have adopted R for clinical-trial analysis, and the FDA accepts R-based submissions. In finance, R is used in risk modeling and regulatory reporting across banks and central banks. These adoptions reflect the maturity and trustworthiness of R as infrastructure, not merely as a convenience.

13 Conclusion

The evidence assembled here—23,304 CRAN packages, 49 Task Views spanning virtually every quantitative discipline, nearly 99 million Bioconductor downloads in a single year, 70.6% adoption in biodiversity journals, 207 econometrics packages in a systematic review, 8,813 Crossref citations of the founding paper, and documented influence on the Python data-science stack—describes an innovation whose impact on statistical practice is both wide and deep. This is not a tool that serves a niche; it is infrastructure on which modern quantitative science runs. The small group of individuals who built and sustained this platform created something whose value, measured in any terms one chooses—scientific output, educational reach, industrial adoption, or influence on other technologies—is extraordinary.

14 Appendix: refreshable metrics (optional code)

The following chunks show how to refresh selected metrics from primary sources. They are optional and require internet access.

To enable them, render with `quarto render r_project_impact.qmd -P refresh_metrics:true` or set `refresh_metrics: true` under `params:` in the YAML.

14.1 CRAN package count (refresh)

```r
nrow(available.packages(repos="cloud.r-project.org"))
```

14.2 Bioconductor repository-wide downloads (refresh)

```r
# Bioconductor provides a tab file; we download and summarize the most recent year.
tab_url <- "https://bioconductor.org/packages/stats/bioc/bioc_2025_stats.tab"
bioc_2025 <- read_tsv(tab_url, show_col_types = FALSE)

bioc_2025 |>
  summarise(
    distinct_ips = sum(Nb_of_distinct_IPs, na.rm = TRUE),
    downloads = sum(Nb_of_downloads, na.rm = TRUE)
  )
```

14.3 Count packages listed in the Finance Task View (refresh)

```r
# install.packages(c("ctv", "utils"))
library(ctv)

# Number of packages is the number of rows in data frame
nrow(ctv("Finance")$packagelist)
```

References

Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314. https://doi.org/10.1080/10618600.1996.10474713

Chambers, J. M., & Hastie, T. J. (1992). *Statistical models in S*. Wadsworth & Brooks/Cole.

Chambers, J. M. (1998). *Programming with data: A guide to the S language*. Springer.

R Foundation for Statistical Computing. (2026). *R: contributors*. https://www.r-project.org/contributors.html

R Developer Page. (2026). *R developer page: Release plans*. https://developer.r-project.org/

R Consortium. (2026). *R consortium mission statement*. https://r-consortium.org/about/

r-devel organization. (2026). *R-devel/r-svn: Mirror of the r svn with CI for testing patches (GitHub repository)*. https://github.com/r-devel/r-svn

Comprehensive R Archive Network. (2026a). *CRAN: Contributed packages*. https://cran.r-project.org/web/packages/

R Core Team. (2026). *Writing R extensions*. https://cran.r-project.org/doc/manuals/R-exts.html

Comprehensive R Archive Network. (2026b). *CRAN DOI*. https://cran.r-project.org/web/packages/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

CRAN. (2026a). *CRAN task views*. https://cran.r-project.org/web/views/

Zeileis, A. (2025). CRAN task views. *The R Journal*. https://journal.r-project.org/articles/RJ-2025-011/

The R Foundation for Statistical Computing. (2026). *The R journal*. https://journal.r-project.org/

Foundation for Open Access Statistics. (2026). *Journal of statistical software*. https://www.jstatsoft.org/

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer. https://doi.org/10.1007/978-3-319-24277-4

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*(10), 1–17. https://doi.org/10.18637/jss.v011.i10

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, *16*(9), 1–16. https://doi.org/10.18637/jss.v016.i09

Croissant, Y., & Millo, G. (2008). Panel data econometrics in r: The plm package. *Journal of Statistical Software*, *27*(2), 1–43. https://doi.org/10.18637/jss.v027.i02

Hothorn, T., Hornik, K., van de Wiel, M. A., & Zeileis, A. (2008). Implementing a class of permutation tests: The coin package. *Journal of Statistical Software*, *28*(8), 1–23. https://doi.org/10.18637/jss.v028.i08

Boudt, K., Kleen, O., & Sjørup, E. (2022). Analyzing intraday financial data in r: The highfrequency package. *Journal of Statistical Software*, *104*(8). https://doi.org/10.18637/jss.v104.i08

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., … Zhang, J. (2004). Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biology*, *5*(10), R80. https://doi.org/10.1186/gb-2004-5-10-r80

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., … Morgan, M. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. *Nature Methods*, *12*(2), 115–121. https://doi.org/10.1038/nmeth.3252

Bioconductor. (2025a). *Bioconductor 3.22 released*. https://bioconductor.org/news/bioc_3_22_release/

Bioconductor. (2025b). *Bioconductor 2024 annual report*. https://www.bioconductor.org/about/annual-reports/AnnRep2024.pdf

Bioconductor. (2026). *Download stats for bioconductor software repository (all packages combined)*. https://bioconductor.org/packages/stats/bioc/

National Library of Medicine. (2004). *PubMed: Bioconductor: Open software development for computational biology and bioinformatics (citation record)*. https://pubmed.ncbi.nlm.nih.gov/15461798/

National Library of Medicine. (2015). *PubMed: Orchestrating high-throughput genomic analysis with bioconductor (citation record)*. https://pubmed.ncbi.nlm.nih.gov/25633503/

Lai, J., Cui, D., Zhu, W., & Mao, L. (2023). The use of R and R packages in biodiversity conservation research. *Diversity*, *15*(12), 1202. https://doi.org/10.3390/d15121202

Michelaki, D., Tsagris, M., & Adam, C. (2025). An overview of economics and econometrics related r packages. *Stats*, *8*(4), 85. https://doi.org/10.3390/stats8040085

Zeileis, A., Leisch, F., Hornik, K., & Kleiber, C. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, *20*(1), 99–121. https://doi.org/10.1002/jae.776

Kleiber, C., & Zeileis, A. (2008). *Applied econometrics with r*. Springer. https://doi.org/10.1007/978-0-387-77318-6

CRAN. (2026b). *AER: Applied econometrics with r (CRAN package)*. https://cran.r-project.org/package=AER

Eddelbuettel, D. (2026). *CRAN task view: Empirical finance*. https://cran.r-project.org/web/views/Finance.html

Eddelbuettel, D. (2009). *Using r in finance*. https://dirk.eddelbuettel.com/papers/mfa2009.pdf

Ulrich, J. (2013). *Conference report: R/finance 2013*. https://journal.r-project.org/news/RJ-2013-1-r-finance/

Bivand, R. (2022). R packages for analyzing spatial data: A comparative case study with areal data. *Geographical Analysis*, *54*(3), 488–518. https://doi.org/10.1111/gean.12319

Casals, M., Fernández, J., Martínez, V., Lopez, M., Langohr, K., & Cortés, J. (2023). A systematic review of sport-related packages within the R CRAN repository. *International Journal of Sports Science & Coaching*, *18*(2), 621–629. https://doi.org/10.1177/17479541221136238

McKinney, W. (2010). *Data structures for statistical computing in Python*. 56–61. https://doi.org/10.25080/Majora-92bf1922-00a

Smith, N. J. (2026). *Patsy: Describing statistical models in Python*. https://patsy.readthedocs.io/en/latest/overview.html

Kibirige, H. (2026). *Plotnine: A grammar of graphics for Python*. https://plotnine.org/

Wilkinson, L. (2005). *The grammar of graphics* (2nd ed.). Springer. https://doi.org/10.1007/0-387-28695-0

Boettiger, C., Chamberlain, S., Hart, E., & Ram, K. (2015). Building software, building community: Lessons from the rOpenSci project. *Journal of Open Research Software*, *3*(1), e8. https://doi.org/10.5334/jors.bu

Ram, K., Boettiger, C., Chamberlain, S., Ross, N., Salmon, M., & Butland, S. (2019). A community of practice around peer-review for long-term research software sustainability. *Computing in Science & Engineering*, *21*(2), 59–65. https://doi.org/10.1109/MCSE.2018.2882753

pyOpenSci. (2026). *pyOpenSci: Open source tools for open science*. https://www.pyopensci.org/

Leisch, F. (2002). *Sweave: Dynamic generation of statistical reports using literate data analysis*. 575–580. https://doi.org/10.1007/978-3-642-57489-4_89

Knuth, D. E. (1984). Literate programming. *The Computer Journal*, *27*(2), 97–111. https://doi.org/10.1093/comjnl/27.2.97

Xie, Y. (2014). knitr: A comprehensive tool for reproducible research in R. *Implementing Reproducible Research*, 3–32.

Malviya-Thakur, A., Mockus, A., Zaretzki, R., Bichescu, B., & Bradley, R. (2023). How R developers explain their package choice: A survey. *2023 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 1–12. https://doi.org/10.1109/ESEM56168.2023.10304869

Crossref. (2026). *Crossref REST API record for ihaka & gentleman (1996) with is-referenced-by-count*. https://api.crossref.org/v1/works/http%3A//dx.doi.org/10.1080/10618600.1996.10474713

Li, K., Yan, E., & Feng, Y. (2017). How is r cited in research outputs? Structure, impacts, and citation standard. *Journal of Informetrics*, *11*(4), 989–1002. https://doi.org/10.1016/j.joi.2017.08.003

R Validation Hub. (2024). *R validation hub: A cross-industry initiative*. https://www.pharmar.org/