Why Git, GitHub and R Are Good Friends

Roopam Das | 401060784 | das.roopam@stud.hs-fresenius.de

Table of contents

Introduction

This presentation explains why Git, GitHub and R work effectively when combined in software development and analytical workflows. These tools integrate into a structured, collaborative, and reproducible system. Git is a version manager, GitHub supports collaboration and accessibility, and R is a powerful statistical and data-analysis environment. Together, they support documentation, clarity in decision-making, and efficient project flow.


The Power of Integrated Data Science Workflow
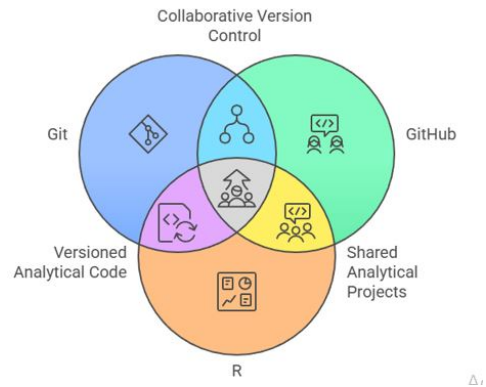
Understanding Git

Git is a distributed version control system that tracks changes, protects analytical work, and allows experimentation safely.

Key Functions

- Maintains historical snapshots
- Supports branching for experimentation
- Prevents conflicts in collaborative work
- Enables restoration of previous versions
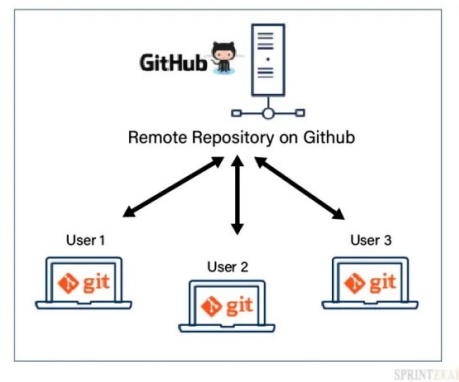- Provides a transparent audit trail
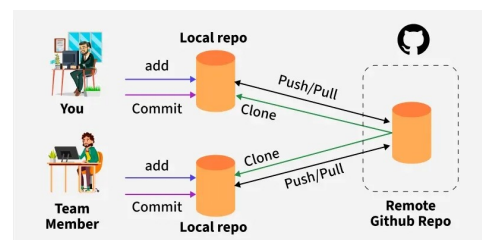
Understanding GitHub

GitHub is a cloud-based platform built on Git, providing collaboration, automation, and version-controlled project management.

Collaboration Features

- Pull requests
- Issue tracking
- Project boards
- Automated testing
- Cloud repository hosting

Understanding R

R is an open-source analytic environment for statistical modelling, visualization, and reproducible research.

Key Features

- Thousands of packages for modelling and automation
- High-quality visualizations
- Tidyverse, RMarkdown, and Quarto support workflow organization Atkins et al. (2025)
- Strong emphasis on transparency and reproducibility
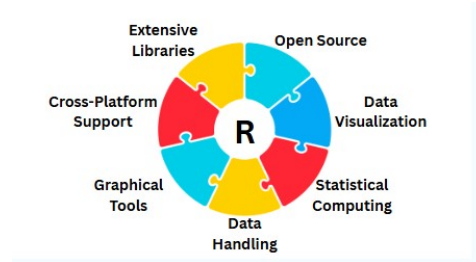
Why These Tools Complement Each Other

Git, GitHub, and R each serve unique purposes but integrate seamlessly for effective workflows.

Why They Work Together

- Git organizes analytical code
- GitHub manages collaboration
- R performs modelling and visualization Vélez et al. (2023)
- Integration reduces errors
- Ensures reproducible workflows
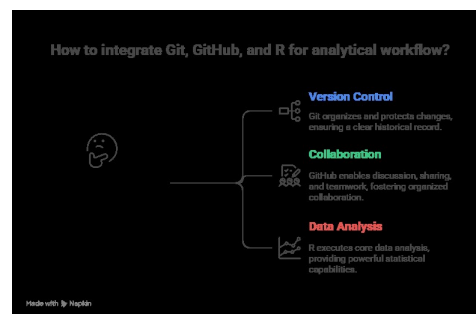
Enhancing Reproducible Research

Version control workflows strengthen transparency and long-term collaboration

Reproducibility is essential in modern analytics

- Git records every analytical decision
- GitHub shares decisions with teams
- R executes computation transparently
- R Markdown and Quarto ensure synchronized documentation
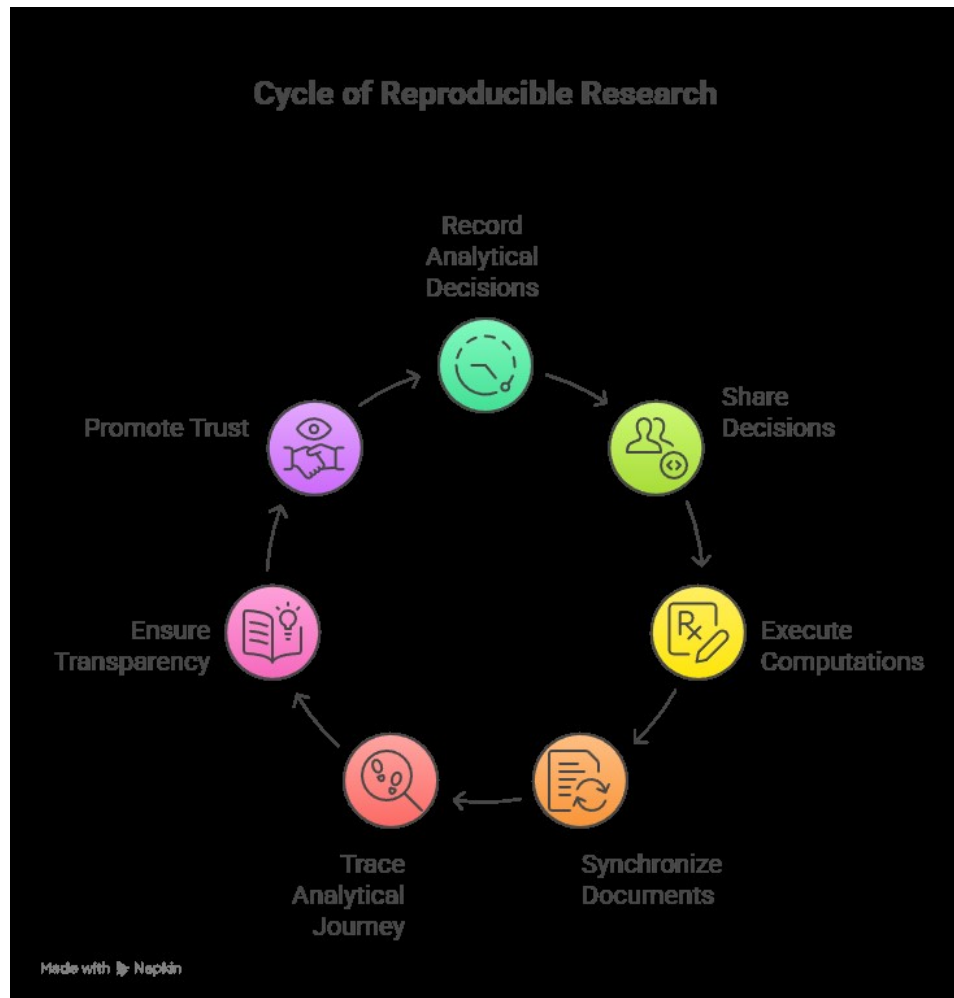
Supporting Collaboration and Teamwork

Git and GitHub provide structure and accountability.

Team Advantages

- Parallel development via branching
- Peer-review via pull requests
- Issue tracking improves communication
- Encourages coordinated development Braga et al. (2023)
- Maintains project consistency
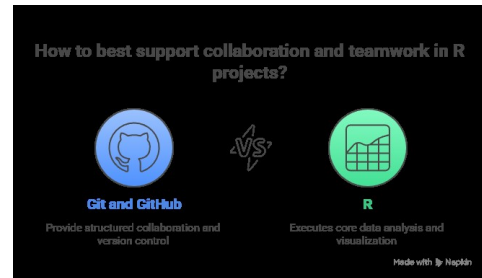
Improving Project Organization

Git and GitHub ensuring clear version tracking, seamless collaboration, and a well-organized analytical environment.

Git and GitHub strengthen project structure for R workflows.

- Versioning preserves scripts, figures, and data
- GitHub offers planning, reviews, and discussions
- Encourages frequent, meaningful commits
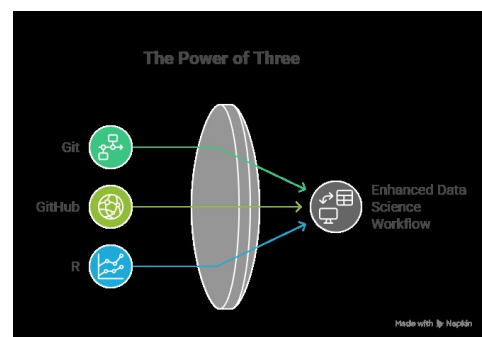- Supports scalable and complex projects
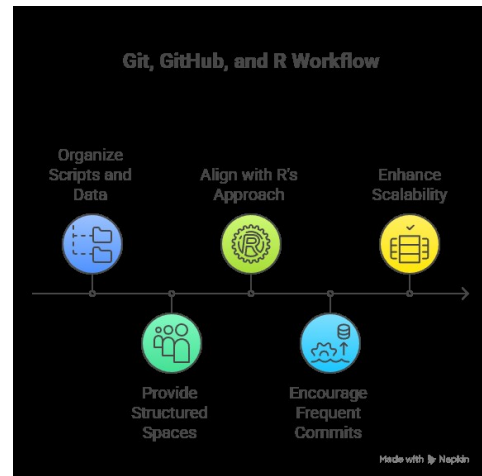
Real-World Applications

Git, GitHub, and R support data-driven processes in many sectors.

Common Use Cases

- Reproducible research
- Forecasting and dashboards
- Government data transparency
- Open-source R package development Weberpals & Wang (2024)
- Education in analytics

Conclusion

Git, GitHub, and R form a complementary ecosystem for analytical work. Together, they create a reproducible, transparent, and collaborative environment that strengthens data science practice across industries.

Real-World Applications of Git, GitHub, and R
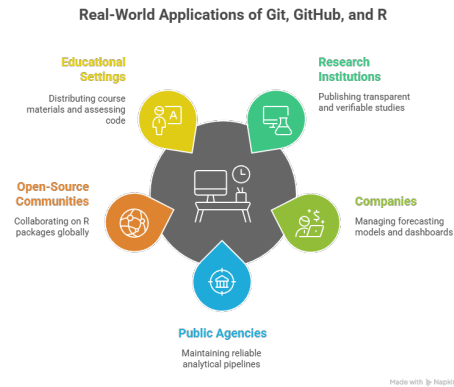
References

Atkins, J. W., Aho, K. S., Chen, X., Elmore, A. J., Fiorella, R., Luo, W., & Richardson, A. D. (2025). Recommendations for developing, documenting, and distributing data products derived from NEON data. *Ecosphere*, *16*(1), e70159.

Braga, P. H. P., Hébert, K., Hudgins, E. J., Scott, E. R., Edwards, B. P., Sanchez Reyes, L. L., & Crystal-Ornelas, R. (2023). Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. *Methods in Ecology and Evolution*, *14*(6), 1364–1380.

Vélez, J., McShea, W., Shamon, H., Castiblanco-Camacho, P. J., Tabak, M. A., Chalmers, C., & Fieberg, J. (2023). An evaluation of platforms for processing camera-trap data using artificial intelligence. *Methods in Ecology and Evolution*, *14*(2), 459–477.

Weberpals, J., & Wang, S. V. (2024). The FAIRification of research in real-world evidence: A practical introduction to reproducible analytic workflows using git and r. *Pharmacoepidemiology and Drug Safety*, *33*(1), e5740.