Title:

Reproducibility Under Scrutiny: A Critical Assessment of RStudio and Jupyter Notebooks in Dutch Supply Chain Sustainability Innovation Labs

Background & Rationale

The need for reproducible research is urgent in both science and business, as irreproducible findings undermine trust, waste resources, and can lead to poor decision-making (Baker, 2016; Goodman et al., 2016). In the context of supply chain sustainability, reproducibility is essential for validating claims about environmental and social impacts, ensuring that innovations are genuinely effective and not the result of flawed or opaque analysis (Montabon et al., 2016). The reproducibility crisis has shown that even computational research is often not independently verifiable, threatening the credibility of sustainability initiatives (Stodden et al., 2018; Collberg & Proebsting, 2016). Without reproducibility, businesses risk implementing unsound practices, regulators lack reliable evidence for policy, and academic progress is stalled by untrustworthy results (Ioannidis, 2005; Sandve et al., 2013).

Open-source tools like RStudio and Jupyter Notebooks offer a pathway to address these challenges by enabling transparent, shareable, and repeatable workflows (Rule et al., 2019). However, their real-world impact depends on overcoming technical, organizational, and contextual barriers, especially in business environments where data confidentiality and complexity are significant (Freire et al., 2012).

Research Objectives

- Empirically test the reproducibility of a selected supply chain sustainability intervention using RStudio and Jupyter Notebooks.
- Identify and document technical, organizational, and contextual barriers to reproducibility in this setting.
- Develop practical recommendations for researchers and practitioners on overcoming these barriers or setting realistic expectations for reproducibility in business innovation research.

Research Questions

1. To what extent can RStudio and Jupyter Notebooks enable reproducible research in real-world supply chain sustainability projects?
2. What are the main technical, organizational, and contextual obstacles to achieving true reproducibility in this domain?

3. How do issues such as data confidentiality, tool limitations, and context specificity affect the transferability and credibility of reproducible research findings?

Methodology

- Case Study Selection: Choose a Dutch supply chain sustainability innovation lab or project with partial data/process access.
- Workflow Implementation: Attempt to document and analyze the intervention using RStudio and Jupyter Notebooks. Rigorously record all steps, including environment setup, data cleaning, analysis, and reporting. Use version control (e.g., Git) and containerization (e.g., Docker) to address dependency and environment issues.
- Reproducibility Audit: Invite an independent researcher to reproduce the analysis using the provided materials. Document where and why reproducibility fails or succeeds.
- Stakeholder Interviews: Conduct interviews with project participants, data stewards, and IT staff to understand practical barriers (e.g., data sharing policies, technical skills, resource constraints).
- Critical Analysis: Compare findings with recent literature on reproducibility in computational science and business research. Reflect on broader implications for supply chain sustainability research and innovation.

Expected Outcomes

- A transparent account of what works and what doesn’t when applying reproducible research tools in a real business innovation context.
- A set of practical guidelines and realistic expectations for academics and practitioners aiming to implement reproducible research in supply chains.
- Recommendations for future tool development, institutional policies, and research practices.
- Evidence that reproducible research can improve the credibility, impact, and scalability of sustainability innovations, supporting better business decisions and policy-making.

Academic and Practical Relevance

- For Academics: Addresses the reproducibility crisis and provides evidence-based recommendations for improving research credibility and impact in business innovation.
- For Practitioners: Demonstrates how reproducible research can validate sustainability claims, reduce risk, and support compliance with regulatory and market demands, leading to more sustainable and trustworthy supply chains.

Deliverables

- Thesis report detailing methodology, findings, obstacles, and recommendations.

- Publicly available (as far as confidentiality allows) RStudio and Jupyter Notebooks, with clear documentation of any reproducibility gaps.
- Presentation summarizing lessons learned for both academic and industry audiences.

Sources

- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454. doi:10.1038/533452a
- Goodman, S. N., Fanelli, D., & Ioannidis, J. P. (2016). What does research reproducibility mean? *Science Translational Medicine*, 8(341), 341ps12. doi:10.1126/scitranslmed.aaf5027
- Stodden, V., Seiler, J., & Ma, Z. (2018). An empirical analysis of journal policy effectiveness for computational reproducibility. *PNAS*, 115(11), 2584–2589. doi:10.1073/pnas.1708290115
- Collberg, C., & Proebsting, T. A. (2016). Repeatability in computer systems research. *Communications of the ACM*, 59(3), 62–69. doi:10.1145/2812803
- Rule, A., Birmingham, A., Zuniga, C., et al. (2019). Ten simple rules for writing and sharing computational analyses in Jupyter Notebooks. *PLoS Computational Biology*, 15(7), e1007007. doi:10.1371/journal.pcbi.1007007
- Montabon, F., Pagell, M., & Wu, Z. (2016). Making sustainability sustainable. *Journal of Supply Chain Management*, 52(2), 11-27. doi:10.1111/jscm.12103
- Freire, J., Bonnet, P., & Shasha, D. (2012). Computational reproducibility: State-of-the-art, challenges, and database research opportunities. *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, 593–596. doi:10.1145/2213836.2213916
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9(10), e1003285. doi:10.1371/journal.pcbi.1003285
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. doi:10.1371/journal.pmed.0020124