

Can R Notebooks help with reproducibility?

Introduction

In this paper we will discuss the differences between replicability and reproducibility, the importance of reproducibility in science and how R Notebook may be a potential solution to this problem.

Reproducibility

Scientific research and methodologies have undergone tremendous technological evolution in the latest decades. However, the manner in which these scientific papers and studies are being published has remained the same. To an extent, one might claim that publication methods are fifty years behind (Gentleman and Lang 2007). For instance, one finds that the conclusion from many published scientific papers are not uphold when thoroughly replicated. Giles and Tedd's book in McCullough, McGeary, and Harrison (2008) illustrates an example of this matter around the Canadian underground economy. Their findings showed that the illegal economy grew tremendously in just under two decades from 3.6% to 15,6 % (McCullough, McGeary, and Harrison 2008). However, the model they used, MIMIC (multiple indicators, multiple causes), was later deemed useless by Breusch for this experiment (McCullough, McGeary, and Harrison 2008) .

Challenges

Science advances on findings of trustworthy research and results (en referanse hadde gjort seg her). A crucial factor in knowing whether these findings are reliable is whether it is possible to reproduce the results. The ability to reproduce research and get the same results is essential when making a trusted conclusion. Unfortunately, reproducibility is a big challenge in science. Many articles that are produced in peer-reviewed journals have problems with reproducibility (McNutt 2014). These challenges entail weak ability of investigators to build on prior work. The issues with reproducibility arise due to insufficient data and documentation of used procedures, software, and packages. Reasons why this deficiency occurs is old, lost or unpublished data. Data, software, and packages are kept private in many cases, making it impossible to duplicate the original research.

One might argue that even though scientific development has seen tremendous growth, especially in the 20th century, much of produced information is misleading. As a consequence, new methods should be implemented to set a basic standard for scientific claims. Not only to validate these claims but also to reduce the uncertainty and dangers of other scientists to become involved in a study, with money and resources, that prove false. Peng (2011) suggests that reproducibility in fact, should become a minimum requirement for scientific papers to validate their claims.

Literature review

Replication of a study is described by Jasny et al. (2011) as the “gold standard” in scientific research (McCullough, McGeary, and Harrison 2008). We can define replication as *“the ability of a researcher to duplicate the result of a prior study if the same procedures are followed, but new data are collected”* (Bollen

et al. 2015). One reason why it's considered the absolute gold standard for researcher is that they can conduct the experiment or study in a new way, with new data-sets, but still get the same results. Reproducibility on the other hand can be seen as *"the ability of a researcher to duplicate the results of a prior study using the same materials and procedures as were used by the original investigator"* (Bollen et al. 2015). Here the same procedures and data will be used to attempt to attain the same result as the original investigation. From the definitions we can see that replication is much wider than reproducibility, giving room for more flexibility in the research.

In economics, there have been different leading professional journals published by scientists. In 1933, Frisch started *Econometrica* that has been one of the leading journals in econometrics for many years (Frisch 1933). During this period, scientists would not publish original data in statistical and numerous work unless their volume were excessive or comprehensive (Her forstår jeg ikke hva dere mener).

In Frisch's journal, Ezekiel wrote an empirical work with included dataset. In this era, regression analysis was pervasive to implement even though the datasets were available (Ezekiel 1933). (Et viktig poeng er at datasettene var veldig små, ofte bare en liten tabelli artikkelen.) In 1960, scientists advanced to big economic models. However, notwithstanding bigger models, it was still unmanageable to move data from one mainframe to another. The predicaments arose because scientists produced the datasets in software and programs related to specific computers. The research results were published exclusively, without exposure of codes, datasets, or research procedures. This way of constructing research made reproduction nearly impossible.

The leading professional journal, "The journal of money, credit and banking project," founded by Dewald in 1982, attempted to replicate published articles (Dewald, Thursby, and Anderson 2020). The investigators conducted a study on the reproducibility of published articles by using datasets submitted with various published articles. This attempt concluded that only 2 of the 70 examined articles were possible to reproduce (Dewald, Thursby, and Anderson 2020). The results were generally disappointing due to missing data, software errors, no documentation, lost and missing investigators and data. In addition, many authors just ignored requests for data and code.

Discussion

While investigating the importance of reproducibility we found a few points of significance to the matter that we will discuss further:

1. Documentation of used procedures

- Data, codes, packages and software
- Quality, correctness and validity

2. Individual preferences

- Demand for data and code archives
- Policies, penalties

3. Culture within science community

R Markdown is a file format that allows production of dynamic documents, while R notebook is a document type that implements a notebook interface (Riederer, n.d.). When performing research in R notebooks all codes used in the study is saved and documented. Accordingly, R notebooks helps to reduce one of the barriers to reproducibility by making codes available. Published and available codes give other researchers and scientist the ability to follow the same procedures when duplicating research. For instance the function

“*SessionInfo*” can be helpful and informative when duplicating procedures as it documents information about which packages and R versions was used by the original investigator.

This is an example of a code-chunk which shows the packages used in the program.

```
sessionInfo()

## R version 4.1.1 (2021-08-10)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## loaded via a namespace (and not attached):
## [1] compiler_4.1.1    magrittr_2.0.1    tools_4.1.1      htmltools_0.5.1.1
## [5] yaml_2.2.1        stringi_1.7.3     rmarkdown_2.10   knitr_1.33
## [9] stringr_1.4.0     xfun_0.25         digest_0.6.27    rlang_0.4.11
## [13] evaluate_0.14
```

R Notebooks can be aligned with reproducibility as it enables documentation of used procedures. In this way potential mistakes in the procedure can be revealed. However, this is only the case if the scientist saves the versions of tools being used and results are interpreted in the R notebook. Peng (2011) emphasizes that even though a research is reproducible, quality, correctness and validity of the published results is not guaranteed.

One issue related to sharing code through R Notebook is that the original code may no longer be available. Loss of original codes will hinder the researcher’s ability to reproduce the content created earlier. Even though R markdown collaborates with the computational environment on most operating systems, the differences within OS software may challenge R markdown as a general solution. Using different versions of OS software may lead to different results when implementing the same procedure.

The results seems to be mixed when it comes to researcher incentives to engage in reproducible research. On the one hand, we can see a lack of penalties for not contributing to this matter. According to McCullough, McGeary, and Harrison (2008), auteurs who refused policies related to reproducibility and those who wouldn’t distribute their code and data to an archive did not get a penalty. A stricter penalty system, where authors would have to distribute their code and data, may promote the cause of reproducible research. On the other hand, we see multiple journals that have adopted a mandatory demand for data and code archives, such as “Econometrica” in 2005 (Frisch (1933) neppe den beste referansen her). Having these integrated adaptations in their journals can further incentivize reproducibility among researchers.

One of the most significant barriers to reproducible research seems to be the lack of an integrated culture within the science community. Despite many incremental steps towards this goal, it will most likely be a long process before we see this type of culture fully implemented. Peng (2011) describes this culture as a “culture of reproducibility,” and the effect this may have on the field of science can be immense. For example, we would see reproducibility as a bare minimum standard for all scientific papers, which would validate scientific claims and make the information spread out to the masses more reliable.

A potential solution

A potential solution to this issue is implementing a research compendium, which combines text, data, and additional software. Gentleman (2005) described a compendium as “a container for the different elements that make up the document and its computations, and as a means for distributing, managing and updating the collection.” The goal here is to achieve a greater way to construct a reproducible project with software such as R. A critical factor to the compendium is that it contains one or multiple dynamic documents, which can be continually updated and edited. In addition, dynamic documents can be converted into more traditional forms, such as a static document. Gentleman and Lang (2007) describes dynamic documents as an “ordered composition of code chunks and text chunks that describe and discuss a problem and its solution.” The code chunks in program languages, such as R, are sequences of commands. The ordering of these doesn’t have to be sequential but instead supports more complex structures in the document. Finally, scientists can utilize text chunks to describe the problem.

So how can the usage of a research compendium be a potential solution with reproducibility? A common idea around this is related to the factor of publication. Hence readers of the article or paper can now access the utilized code and data, which will result in easier management when reproducing results. In addition, using a compendium will make your work more efficient since it allows for more open communication between individual researchers and the ability to test and validate your claims. To summarize, employing a compendium will make the researchers content more available to others. It will be more accessible to validation, and it opens the possibility for others to continue the work on the project.

Conclusion

We believe that introducing the standard of making data and computer code available to others through R Notebook, will only partly solve the problem with reproducibility. From the research conducted in this assignment, we found that a few critical barriers prohibit reproducibility in the field of science. On the upside, technological innovations have made sharing detailed logs about all utilized codes much more manageable. They will most definitely be more precise than a journal using a natural language.

Reproducibility in the field of science is something that develops slowly but steadily towards the future. The increased awareness about this issue and the implementation of electronic journals, mandatory archives, and code-sharing platforms such as Github has made the possibility of engaging in reproducible research more achievable. With further implementation, the infrastructure can keep evolving. As stated earlier, the importance of a deeply integrated culture within science will be essential for development. If future papers are released, published, and distributed to others, it will reduce the spread of misinformation and make reproducibility achievable.

To summarize, we found several indications that computable documents such as R notebooks can increase the possibility of reproducibility. Computable documents are a possible way to solve issues connected to mandatory archives and code-sharing platforms. However, notwithstanding the possibilities R notebooks and computable documents entail, other predicaments need to be solved to optimize reproducibility. There will still be a need for development in other components of the domain. In conclusion, R notebook implies a step in the right direction but not a complete solution to this problem.

References

Bollen, Kenneth, John T. Cacioppo, Jon A. Krosnick, James L. Olds, and Robert M. Kaplan. 2015. “Social, Behavioral, and Economic Sciences Perspectives on Robust and Reliable Science.” Report of the Subcommittee on Replicability in Science Advisory Committee to the National Science Foundation Directorate for Social, Behavioral, and Economic Sciences. NSF.

- Dewald, William G, Jerry G Thursby, and Richard G Anderson. 2020. “Replication in Empirical Economics: The Journal of Money, Credit and Banking Project,” 18.
- Ezekiel, Mordecai. 1933. “Some Considerations on the Analysis of the Prices of Competing or Substitute Commodities.” *Econometrica* 1 (2): 172–80.
- Frisch, Ragnar. 1933. “Editor’s Note.” *Econometrica* 1 (1): 1–4.
- Gentleman, Robert. 2005. “Reproducible Research: A Bioinformatics Case Study.” *Statistical Applications in Genetics and Molecular Biology* 4 (1). <https://doi.org/10.2202/1544-6115.1034>.
- Gentleman, Robert, and Duncan Temple Lang. 2007. “Statistical Analyses and Reproducible Research.” *Journal of Computational and Graphical Statistics* 16 (1): 1–23. <https://doi.org/10.1198/106186007X178663>.
- Jasny, Barbara R., Gilbert Chin, Lisa Chong, and Sacha Vignieri. 2011. “Again, and Again, and Again.” *Science* 334 (6060): 1225–25. <https://doi.org/10.1126/science.334.6060.1225>.
- McCullough, B. D., Kerry Anne McGeary, and Teresa D. Harrison. 2008. “Do Economics Journal Archives Promote Replicable Research?” *Canadian Journal of Economics/Revue Canadienne d’économique* 41 (4): 1406–20. <https://doi.org/10.1111/j.1540-5982.2008.00509.x>.
- McNutt, Marcia. 2014. “Reproducibility.” *Science* 343 (6168): 229–29. <https://doi.org/10.1126/science.1250475>.
- Peng, Roger D. 2011. “Reproducible Research in Computational Science.” *Science* 334 (6060): 1226–27. <https://doi.org/10.1126/science.1213847>.
- Riederer, Emily, Christophe Dervieux. n.d. *R Markdown Cookbook*.

Appendix

We have included a screenshot of our git-history to show commits and branches used during writing this paper.

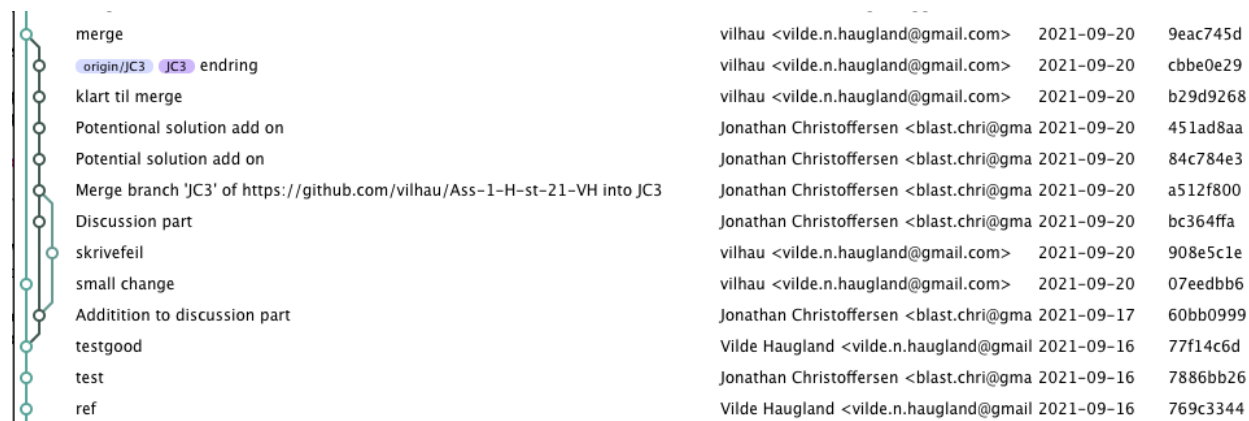


Figure 1: Git history.