# Can R Notebook help with reproducibility?

## Introduction

This text will focus on how R Notebook can help with reproducibility. The text is based on reproducibility as a standard for general research being published, as well as reproducibility in the field of economical studies. R Notebook will be discussed as a solution to meet the criteria of reproducibility.

## Literature review

**"Reproducible Research in Computational Science", by Peng 2011.**

Susann:

Roger D. Peng presents reproducibility as the bare minimum standard when evaluating all published scientific findings. The reasoning being the change in technology leading to large public data sets/databases used by multiple scientist to produce independent research. This change making it harder to investigate the credibility of different studies/papers.

According to Peng, the studies/papers needs to have data and code public to make it possible for it to be reproducible. An issue following this requirement is that the computer code often is not available due to software systems and packages being private.

As the main goal presented in the article is to develop a culture that require reproducibility for computational science to be publish, two solutions are mentioned:

1. Authors make their data and code available for everyone by using the free code repositories. Such as Github and SourceForge.
    a. This would make it possible to provide a minimum code as online material which would be informative, and could be used to reveal any potential problems.
    b. The next step would be to be able to provide datasets in durable non-proprietary format, but this would require additional costs.

2. The scientific community creates a single place for authors in all fields to make their studies reproducible.
    a. In order to accomplish this the community would need to create a common DataMed central, CodeMed central and PubMed central where publishers/authors data, metadata, and code can be stored with their studies. This would require additional costs and would need to be coordinated and supported by the government.

Kine:

According to the article Reproducible Research in Computational Science written by Roger D. Peng is that reproducibility research has an achievable minimum standard. This standard for reproducibility requires that the data and the computer code are published for others.

A goal for the reproducibility standard is to fill the hole between the replication of a study and where there is no replication in the scientific evidence-generating. Between the replication of study and no replication is it opportunities, where a study could be more reproducibility than the other. This depends on which data and code are made available.

The reproducibility standard is based on that every calculation experiment has a detailed log of every action that the computer does. A critic problem in reproducibility in many cases is that the computer code isn´t available longer.

In order for reproducibility to come more into the spotlight, there must be more contributions from several directions. Journals can have a meaning here. Roger D. Peng says "The journal Biostatistics, for which I am an associate editor, has implemented a policy for encouraging authors of accepted papers to make their work reproducible by others (11)".

The reproducibility of an analyze cannot guarantee the validity, quality and correctness when it is published to others.

Susann:

**"Do economics journal archives promote replicable research?", by McCullough.**

McCullough's article puts focus on reproducibility and replication in the field of economical studies. In this article the ability replicate a stydy is put as the gold standard to evaluate the reliability of a scientific claim, and should be aimed for in all archives. Replicability meaning being able to collect new data and get the same results/conclusions.

The article shines light on the fact that long-standing archives of economics do not accustom to reproducibility. And some of those who claim to be data-only archives fail in reality due to few authors contributing with the data. Meaning they are not promoting replicable research.

Additionally, the article shows that editors will defend published work even if it is not replicable with the argument of the papers being published before the requirements for data availability. This only showcases the importance stricter data+code requirements.

# Discussion

# Conclusion

# References