LABORATÓRIO DE ESTATÍSTICA APLICADA

How to simplify your statistical reports using R Markdown

Marcus Nunes

April 17th, 2023

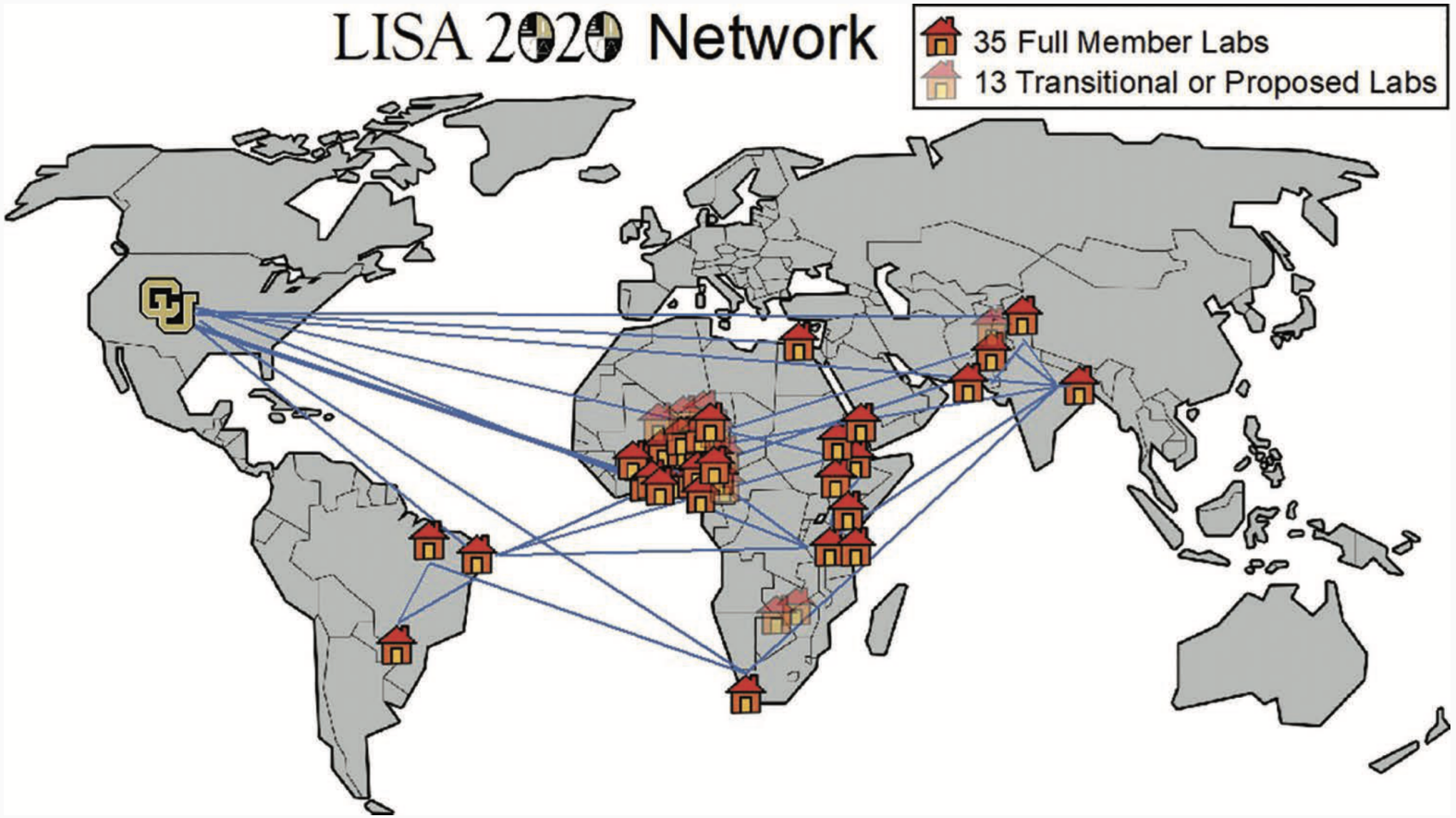
UFRN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE

Who am I?

Who am I?

- Marcus Nunes, Assistant Professor at Federal University of Rio Grande do Norte
- PhD in Statistics from Penn State
- Data science, machine learning, R programming, statistics education
- Director of Laboratório de Estatística Aplicada: lea.estatistica.ccet.ufrn.br
- Personal website: marcusnunes.me (in Portuguese, sorry)
- Talk material: github.com/mnunes/umat-lisa

Who am I


LISA 2020 Network
35 Full Member Labs
13 Transitional or Proposed Labs

What is R Markdown?

What is R Markdown?

- R Markdown is a file format for making dynamic reports
- You put together markdown (a lightweight markup language) and R (a programming language) to build reports
- Advanced users can add LaTeX in the mix and make their texts more sophisticated

What is R Markdown?

- **Word:** What You See Is What You Get (WYSIWYG)
- **R Markdown**: markup language, like LaTeX and HTML
- Focus on the content, not the design

Why R Markdown?

Why R Markdown?

- **Reproducibility:** using the same data and code to get the same conclusions
- **Replicability:** doing the experiment twice to get the same result
- We use it to write our own reports at LEA

Why R Markdown?

- Ease of use
- Short learning curve when compared to other markup languages
- Makes documentation simpler
- Collaborative tool
- Long run time saver

Typical Problems without R Markdown

Typical Problems without R Markdown

- Data scattered in different folders
- Many versions of the same dataset
- Missing steps in the code
- Go to statistical software, run the analysis, export the figures, insert them in the paper, they do not look good, remove, go back to statistical software, change the figures, export them again, insert in the paper, they still do not look good, rinse and reapeat until exhaustion

Typical Solutions with R Markdown

- Code and data are in the same parent folder (usually)
- The dataset is processed everytime you run your code
- The code must be complete; otherwise, the report is not generated
- Run the analysis and report at the same time, so everything is always in sync

Syntax

Syntax

Every R Markdown document has three parts:

1. Header: document definitions
2. Chunks: snippets with R code
3. Text: analysis explanation

Syntax

- The best way to start working with R Markdown is by practicing
- Open the file `examples/report.Rmd`
- Let’s work on it together, figuring out what can be done in this language

Other Styles

Other Styles

- Like in LaTeX, it is very easy to change the look and feel of our documents
- You can check what styles are available in your computer in the menu `File > New File > R Markdown...`

Other Styles

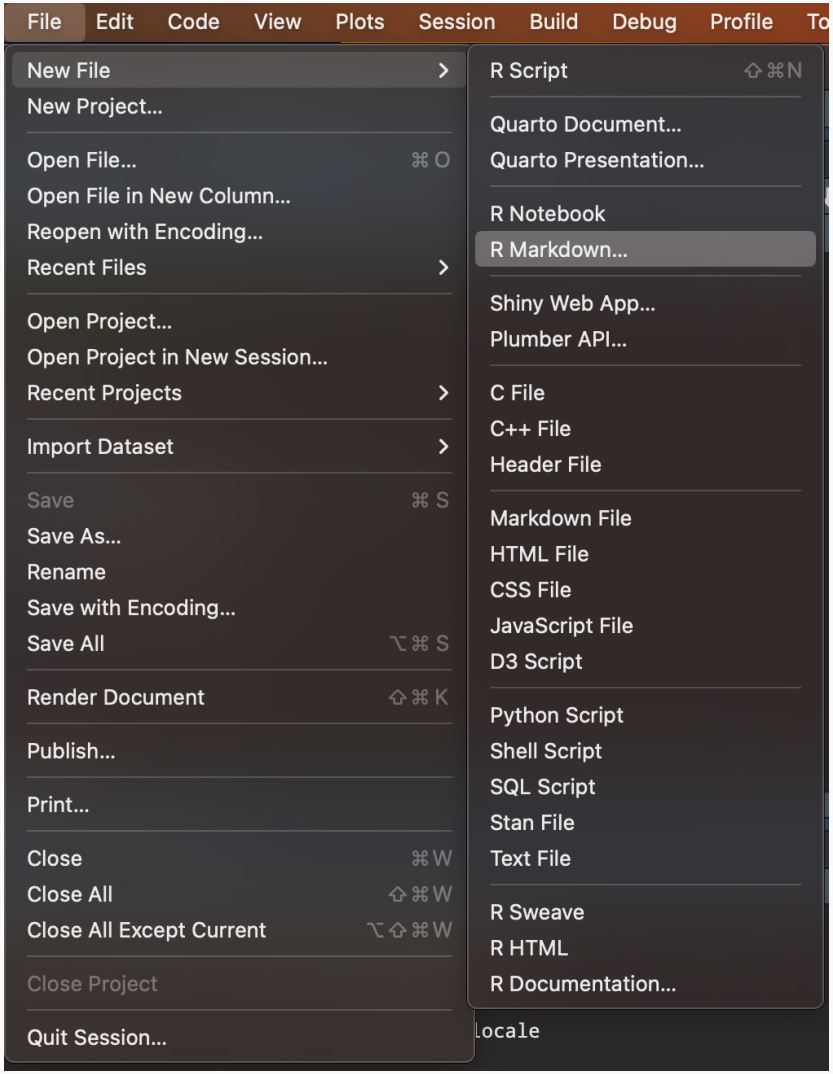


Figura 1: New file screenshot

Practice

Practice

- You can practice what you learned today
- Open the file `practice/practice.Rmd` and follow the instructions
- Notice the solutions are in `practice/solutions.Rmd`, but avoid to look at them and try to figure out the solutions by yourself
- We will have no time to check the results, but feel free to email your questions to `marcus.nunes@ufrn.br`

References i

Yihui Xie. knitr: A comprehensive tool for reproducible research in R. In Victoria Stodden, Friedrich Leisch, and Roger D. Peng, editors, *Implementing Reproducible Computational Research*. Chapman and Hall/CRC, 2014. ISBN 978-1466561595.

Yihui Xie. *Dynamic Documents with R and knitr*. Chapman and Hall/CRC, Boca Raton, Florida, 2nd edition, 2015. URL `https://yihui.org/knitr/`. ISBN 978-1498716963.

Yihui Xie. *knitr: A General-Purpose Package for Dynamic Report Generation in R*, 2023. URL `https://yihui.org/knitr/`. R package version 1.42.