

REPRODUCIBLE RESEARCH IN R

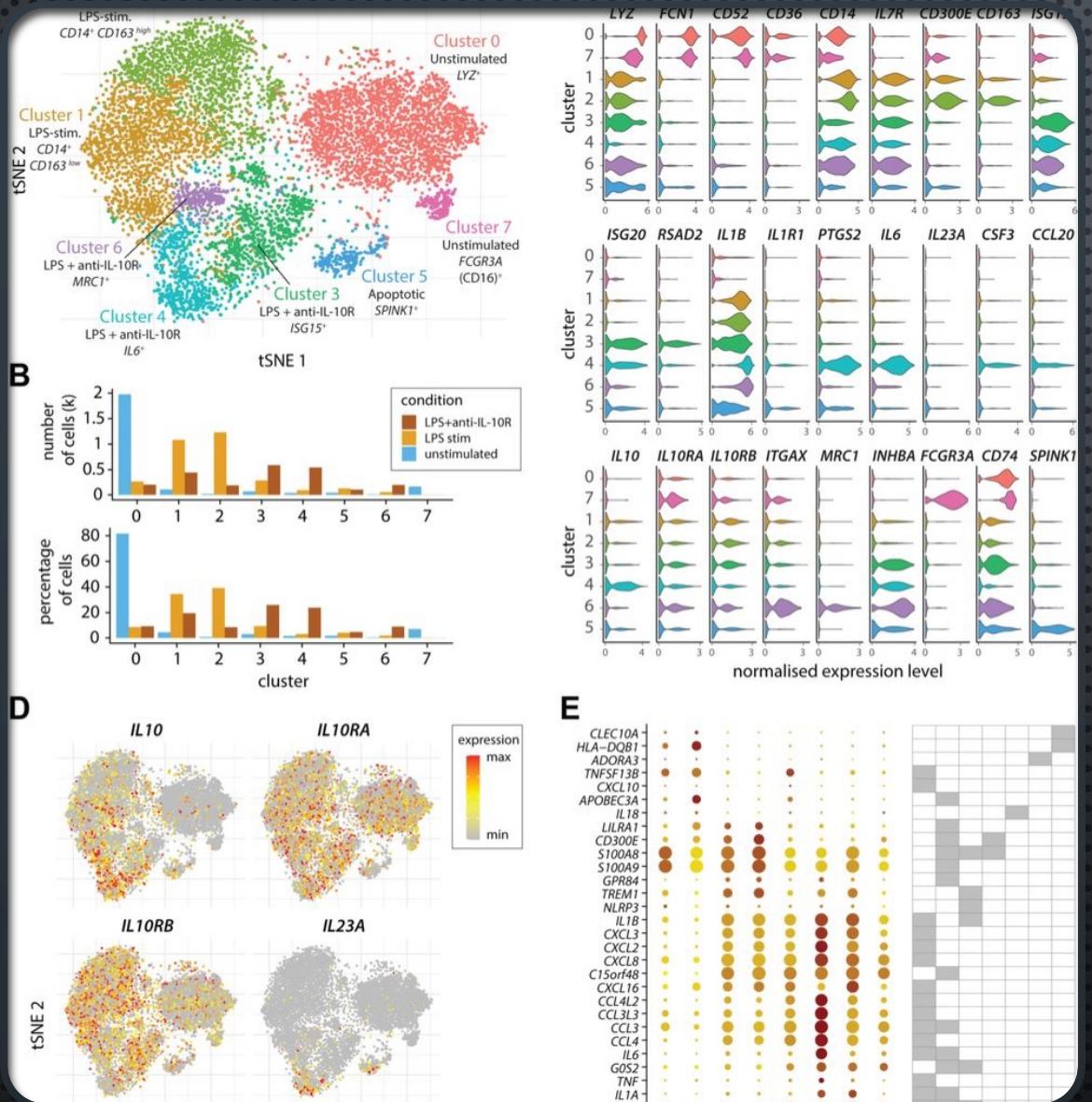
SOUMYA BANERJEE

RATIONALE

- YOUR DATA, YOUR MODEL DECISIONS, PARAMETERS AND YOUR DATA FILTERING DECISIONS WILL KEEP ON CHANGING. HOW DO YOU KNOW 6 MONTHS LATER WHAT HAS CHANGED? DOCUMENT YOUR CODE AND YOUR OUTPUT AND YOUR DESIGN DECISIONS ALL IN ONE PLACE.
- REPRODUCIBLE PIPELINE
- KNOW EXACTLY WHAT CHANGED AND WHEN
- KNOW HOW TO RERUN THE ANALYSIS AND GET THE (SAME) RESULTS
- THIS IS LIKE YOUR RESEARCH NOTEBOOK

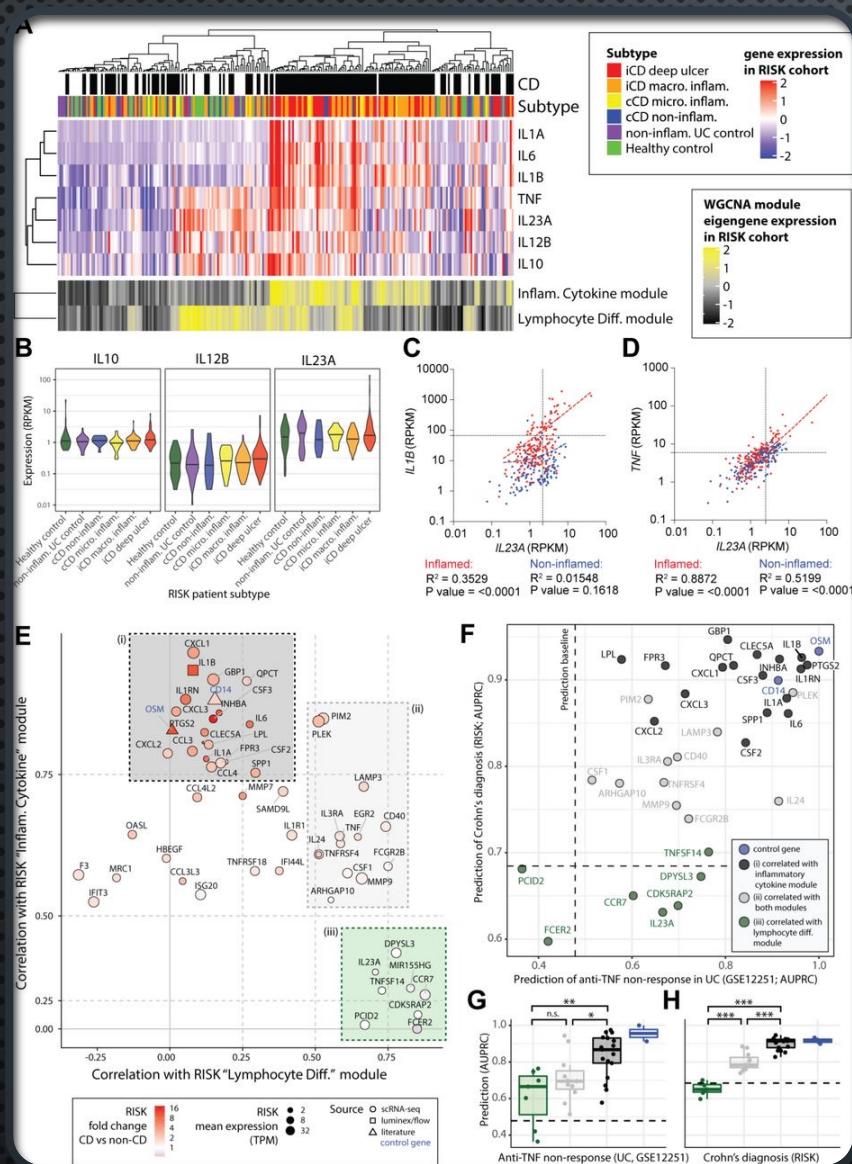
RATIONALE

- EXPERIENCES/CASE STUDIES OF USING RMARKDOWN NOTEBOOKS AND HELPING BIOLOGISTS USE THEM TO ANALYZE THEIR OWN DATA
- WHEN YOU ARE DEEP IN YOUR WORK, IT CAN BE DIFFICULT TO MAKE CODE PRETTY, COMMENT IT AND MAKE IT REPRODUCIBLE.
- BUT YOU WILL REGRET NOT DOING THIS WHEN YOU PARK THE WORK AND 6 MONTHS LATER YOUR COLLOABORATORS/REVIEWERS ASK FOR ADDITIONAL ANALYSIS OR CHANGING SOME ASSUMPTION, ETC.
- YOUR CODE SHOULD THEN BE READY (YOU SHOULD BE ABLE TO CLICK A BUTTON AND REPRODUCE THE FIGURES FOR YOUR PAPER).

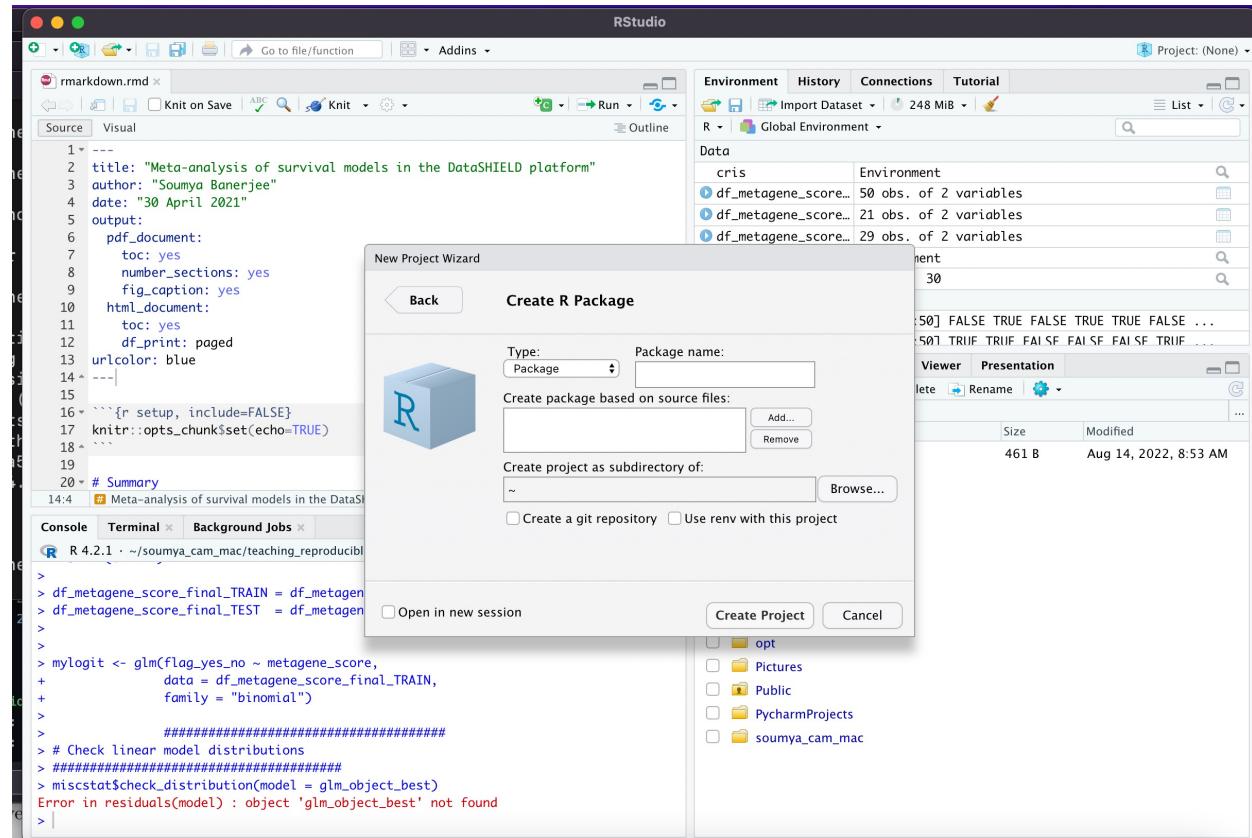


CASE STUDY

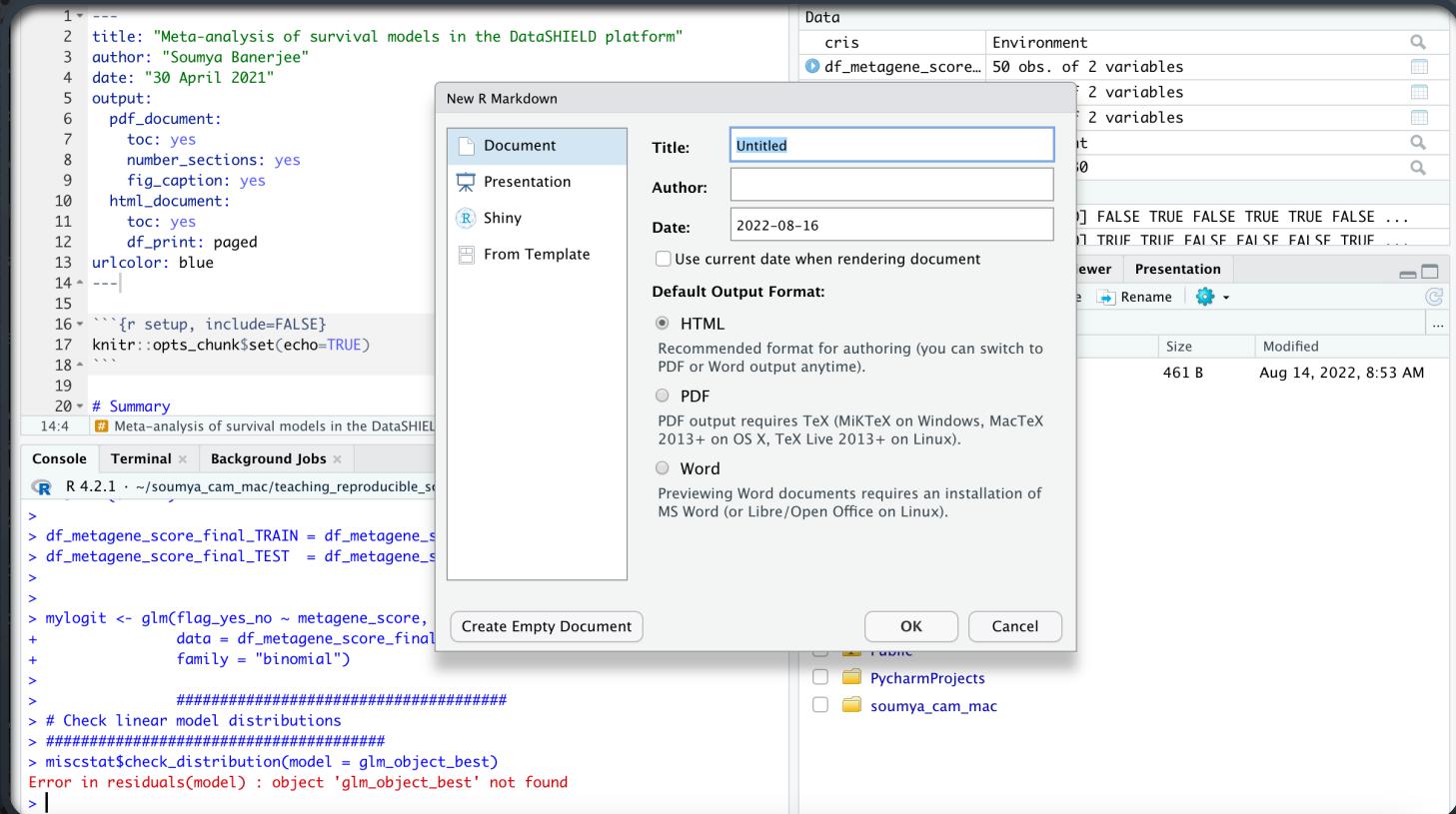
CASE STUDY



CREATING PACKAGES



CREATING R MARKDOWN



RATIONALE

- THE CONCEPTS ARE THE SAME IN ANY PROGRAMMING LANGUAGE (R/PYTHON)
- BOTTOMLINE: WE ARE ALL BUSY AND WE WOULD ALL RATHER PUBLISH PAPERS, BUT IN THE LONG TERM THESE BEST PRACTICES WILL MAKE US MORE PRODUCTIVE
- THIS IS LIKE PROTOCOLS (USED IN EXPERIMENTAL BIOLOGY) FOR COMPUTER SCIENTISTS. ALSO LIKE A LAB NOTEBOOK BUT FOR COMPUTATIONAL PEOPLE.

DEMO

The screenshot shows the RStudio interface with three panes:

- Left Pane (Source):** Displays the code of an R Markdown document named "rmarkdown.rmd". The code includes R Markdown syntax like `#` for sections, `##` for subsections, and code chunks with ````{r}` blocks. It also shows a summary of the `cars` dataset.
- Middle Pane (Environment):** Shows the R environment. The message "Environment is empty" is displayed. There are tabs for Environment, History, Connections, and Tutorial.
- Right Pane (Files):** A file browser showing the contents of the "Home" directory. The directory structure is as follows:

Name	Size	Modified
.Rhistory	461 B	Aug 14, 2022, 8:53 AM
cam_project		
Desktop		
Documents		
Downloads		
Library		
Movies		
Music		
opt		
Pictures		
Public		
PycharmProjects		
soumya_cam_mac		

Syntax	Becomes
Make a code chunk with three back ticks followed by an r in braces. End the chunk with three back ticks:	Make a code chunk with three back ticks followed by an r in braces. End the chunk with three back ticks:
<pre>```{r} paste("Hello", "World!") ```</pre>	<pre>paste("Hello", "World!")</pre>
Place code inline with a single back ticks. The first back tick must be followed by an R, like this `r paste("Hello", "World!")`.	Place code inline with a single back ticks. The first back tick must be followed by an R, like this Hello World!.
Add chunk options within braces. For example, `echo=FALSE` will prevent source code from being displayed:	Add chunk options within braces. For example, <code>echo=FALSE</code> will prevent source code from being displayed:
<pre>```{r eval=TRUE, echo=FALSE} paste("Hello", "World!") ```</pre>	<pre>## [1] "Hello World!"</pre>

DEMO

DEMO

- SEE THE LINK BELOW FOR MORE DETAILS
- [HTTPS://WWW.RSTUDIO.COM/WP-CONTENT/UPLOADS/2015/03/RMarkdown-Reference.pdf](https://www.rstudio.com/wp-content/uploads/2015/03/rmarkdown-reference.pdf)
- NOW HEAD OVER TO THE FILE NAMED RMarkdown.RMD
- [HTTPS://GITHUB.COM/NEELSOUMYA/TEACHING_REPRODUCIBLE_SCIENCE_R/BLOB/MAIN/RMARKDOWN.RMD](https://github.com/NeelSoumya/Teaching_Reproducible_Science_R/blob/main/rmarkdown.Rmd)
- RUNNING THIS WILL CREATE A REPORT LIKE THE FOLLOWING:
- [HTTPS://GITHUB.COM/NEELSOUMYA/TEACHING_REPRODUCIBLE_SCIENCE_R/BLOB/MAIN/RMARKDOWN.PDF](https://github.com/NeelSoumya/Teaching_Reproducible_Science_R/blob/main/rmarkdown.pdf)

Simple demo of rmarkdown

Soumya Banerjee

30 August 2022

Contents

1	Summary	1
2	Visualize data	1
3	Perform logistic regression	1

1 Summary

This is a document that outlines a demo of an R markdown.

2 Visualize data

todo histogram

3 Perform logistic regression

```
## [1] 0.3666667
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

DEMO

CONCEPTS ARE LANGUAGE AGNOSTIC

```
# This is a test  
  
Testing pandoc  
  
``` code  
import GPy

``` code  
print("Cat")  
```
```

# CONCEPTS ARE LANGUAGE AGNOSTIC

- [https://github.com/neelsoumya/teaching\\_reproducible\\_science\\_R/blob/main/tst.md](https://github.com/neelsoumya/teaching_reproducible_science_R/blob/main/tst.md)
- pandoc tst.md -o test.ipynb

# RESOURCES

- [HTTPS://GITHUB.COM/NEELSOUMYA/TEACHING\\_REPRODUCIBLE\\_SCIENCE\\_R/](https://github.com/neelsoumya/teaching_reproducible_science_R/)
- CODE, RESOURCES AND TEMPLATES
- EXERCISES WITH SAMPLE DATA
- INSTALLATION INSTRUCTIONS