

REPRODUCIBLE RESEARCH IN R

SOUMYA BANERJEE

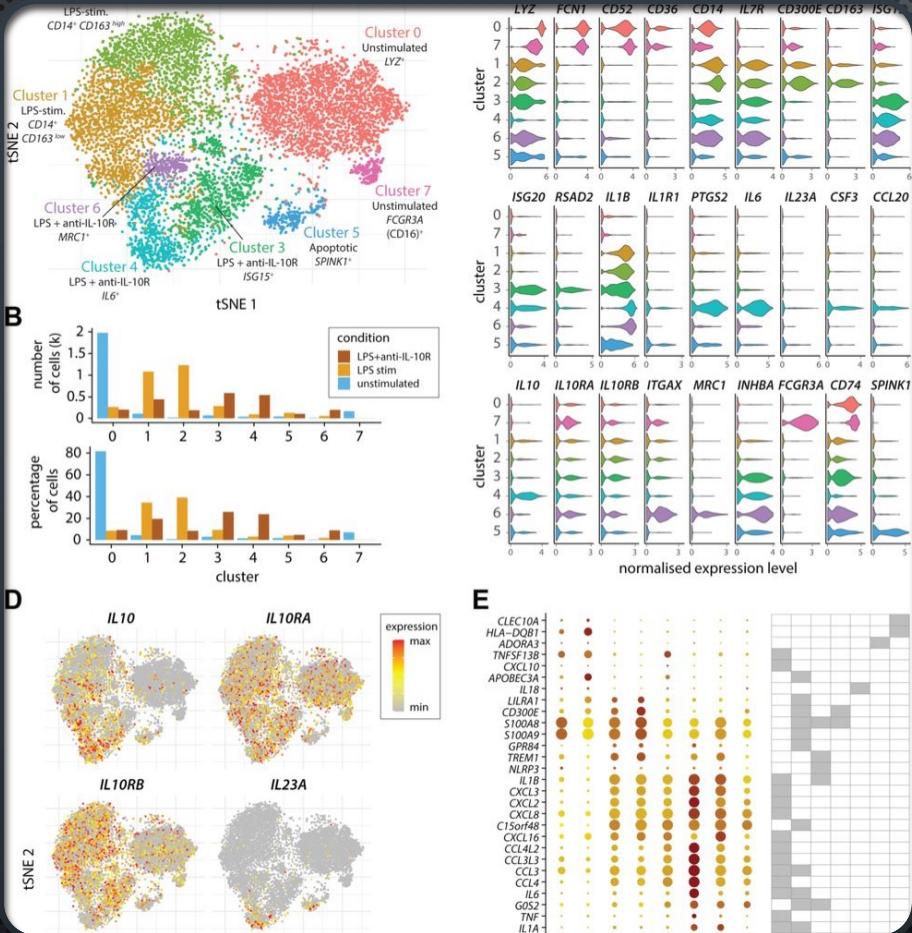
RATIONALE

- 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

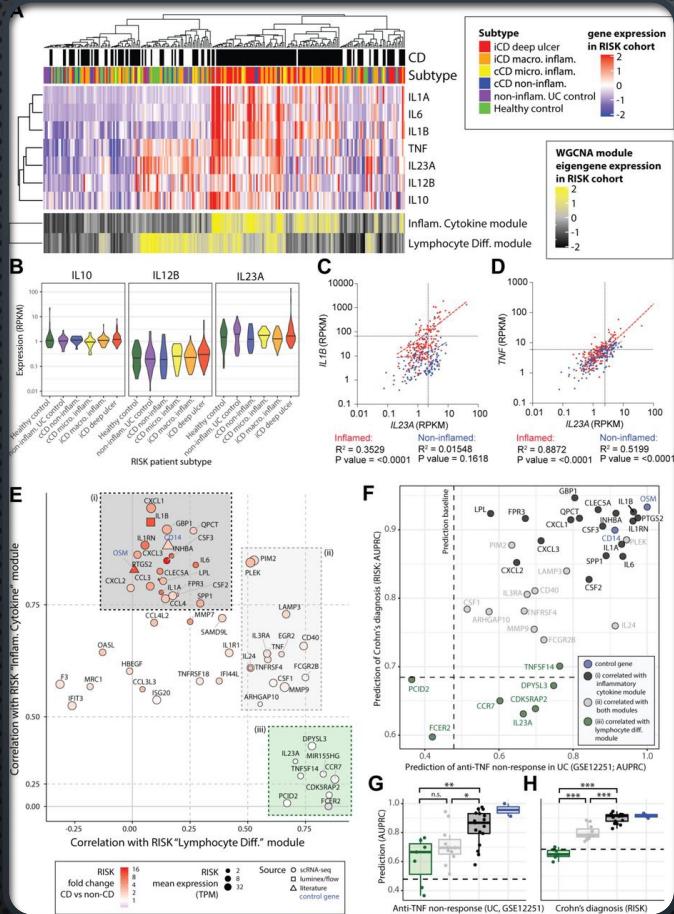
RATIONALE

- 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 COLLABORATORS/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).

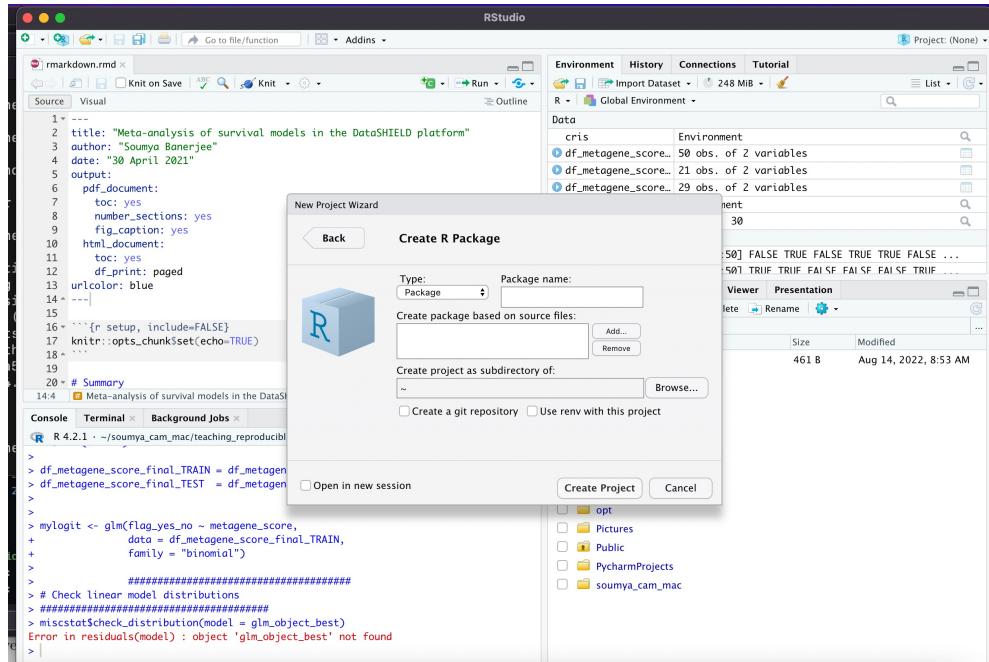
CASE STUDY



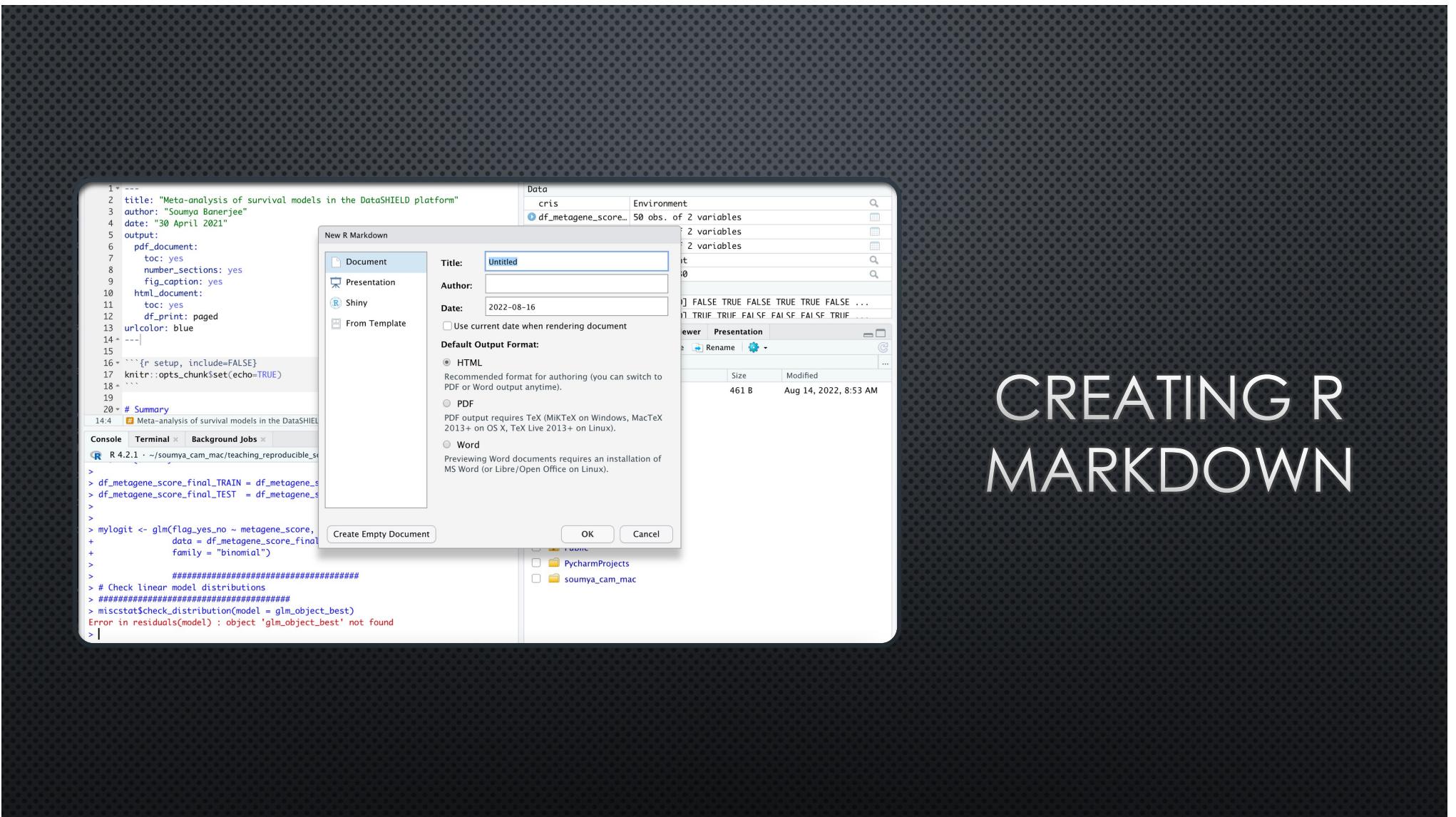
CASE STUDY



CREATING PACKAGES



CREATING R MARKDOWN



DEMO

The screenshot shows the RStudio interface with the following components:

- Left Panel (Code Editor):** Displays the content of an R Markdown file named "rmarkdown.rmd". The code includes R code chunks and explanatory text. For example, it shows how to include plots in a document.
- Middle Panel (Environment):** Shows the R environment. It displays the message "Environment is empty".
- Bottom Panel (File Explorer):** Shows the local file system structure under "Home". The visible files and folders include .History, cam_project, Desktop, Documents, Downloads, Library, Movies, Music, opt, Pictures, Public, PycharmProjects, and soumya_cam_mac.

R MARKDOWN BASICS

- ````R { } ```
- INCLUDE = FALSE PREVENTS CODE AND RESULTS FROM APPEARING IN THE FINISHED FILE. R MARKDOWN STILL RUNS THE CODE IN THE CHUNK, AND THE RESULTS CAN BE USED BY OTHER CHUNKS.
- ECHO = FALSE PREVENTS CODE, BUT NOT THE RESULTS FROM APPEARING IN THE FINISHED FILE. THIS IS A USEFUL WAY TO EMBED FIGURES.
- MESSAGE = FALSE PREVENTS MESSAGES THAT ARE GENERATED BY CODE FROM APPEARING IN THE FINISHED FILE.
- WARNING = FALSE PREVENTS WARNINGS THAT ARE GENERATED BY CODE FROM APPEARING IN THE FINISHED.
- FIG.CAP = "..." ADDS A CAPTION TO GRAPHICAL RESULTS.

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

DEMO

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.

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*

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)

DEMO WITH RMARKDOWN FROM REAL-WORLD
PROJECT

EXERCISES WITH SYNTHETIC DATA

GRAPHICAL USER INTERFACES

- YOU CAN ALSO EASILY CREATE GRAPHICAL USER INTERFACES HERE IS A DEMO:
[HTTPS://SB2333MEDSCHL.SHINYAPPS.IO/SHINYAPP](https://sb2333medschl.shinyapps.io/shinyapp)
- CODE APP.R:
[HTTPS://GITHUB.COM/NEELSOUMYA/TEACHING_REPRODUCIBLE_SCIENCE_R/BLOB/MAIN/SHINYAPP/APP.R](https://github.com/neelsoumya/teaching_reproducible_science_R/blob/main/shinyapp/app.R)
- DEMO

RESOURCES

- [HTTPS://GITHUB.COM/NEELSOUMYA/TEACHING_REPRODUCIBLE_SCIENCE_R/](https://github.com/neelsoumya/TEACHING_REPRODUCIBLE_SCIENCE_R/)
- CODE, RESOURCES AND TEMPLATES
- EXERCISES WITH SYNTHETIC DATA
- INSTALLATION INSTRUCTIONS