"WHY WON'T THIS RUN AGAIN?!"

Andrew Heiss, PhD

Brigham Young University

May 14, 2018

@andrewheiss

andhs.co/rr

PLAN FOR TONIGHT

Why focus on reproducibility?

Reproducibility vs. replicability

The continuum of reproducibility

Demonstrations

WHY FOCUS ON REPRODUCIBILITY?

Andrew,

The models you ran look great but my RA just found a newer version of the data. Can you make all the figures again? Sorry!

Judith

The results in Table 1 aren't the same as what's in Figure 2

Where did we get this data file?

Why did I filter out those rows?

How did I make that figure?

In what order do I run these scripts?

```
Warning message:
`data_frame()` is deprecated, use `tibble()`.
```

REPRODUCIBILITY VS. REPLICABILITY

SCIENCE

Psychology's Replication Crisis Is Running Out of Excuses

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

ED YONG   NOV 19, 2018

Reproducibility	Replicability
Someone with the same code and data can recreate your findings	Someone who conducts another study with the same methods will find similar results

		Data	
		Same	Different
Code	Same	Reproducible	Replicable
	Different	Robust	Generalisable

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

CONSENSUS STUDY REPORT

Reproducibility and Replicability in Science

May 7, 2019

New Report Examines Reproducibility and Replicability in Science, Recommends Ways to Improve Transparency and Rigor in Research

WASHINGTON – While computational reproducibility in scientific research is generally expected when the original data and code are available, lack of ability to replicate a previous study -- or obtain consistent results looking at the same scientific question but with different data -- is more nuanced and occasionally can aid in the process of scientific discovery, says a new congressionally mandated report from the National Academies of Sciences, Engineering, and Medicine. *Reproducibility and Replicability in Science* recommends ways that researchers,

THE CONTINUUM OF REPRODUCIBILITY

Gold standard of reproducibility

Anyone* can recreate your results at any point in the future on any computing platform

* yourself included

This is hard!
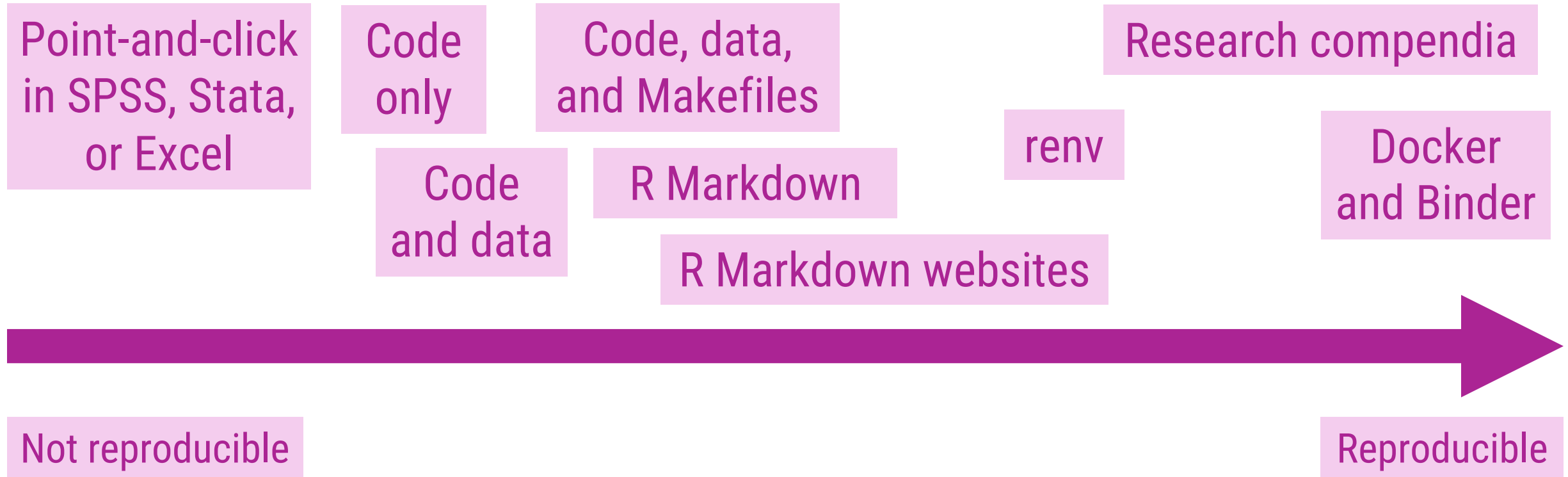
A little bit reproducible is better than not reproducible

Try to make each project a bit better organized and reproducible than the last

Again, you shouldn't try to do these things all at once; start with one, or part of one. Then in your next project, do that plus another thing.

- Organize your data and code
- Everything with a script
- Automate the process
- Turn scripts into reproducible reports
- Turn repeated code into functions
- Package functions for reuse
- Use version control
- License your software

- Other resources

http://kbroman.org/steps2rr/

The Reproducibility Continuum

Point-and-click in SPSS, Stata, or Excel

Code only

Code, data, and Makefiles

Research compendia

renv

Code and data

R Markdown

Docker and Binder

R Markdown websites

Not reproducible

Reproducible

There's no one right way to do this

DEMONSTRATIONS

https://andhs.co/rr