



# Managing Your Data Pipeline

FROM WRANGLING TO VISUALIZING YOUR DATA IN R

Liz Beard - 01/29/2021

# WHY DO WE NEED TO THINK ABOUT DATA MANAGEMENT?

## MACRO:

Good data management practices enhance scientific rigor by simplifying reproducibility efforts, promoting collaboration, etc.

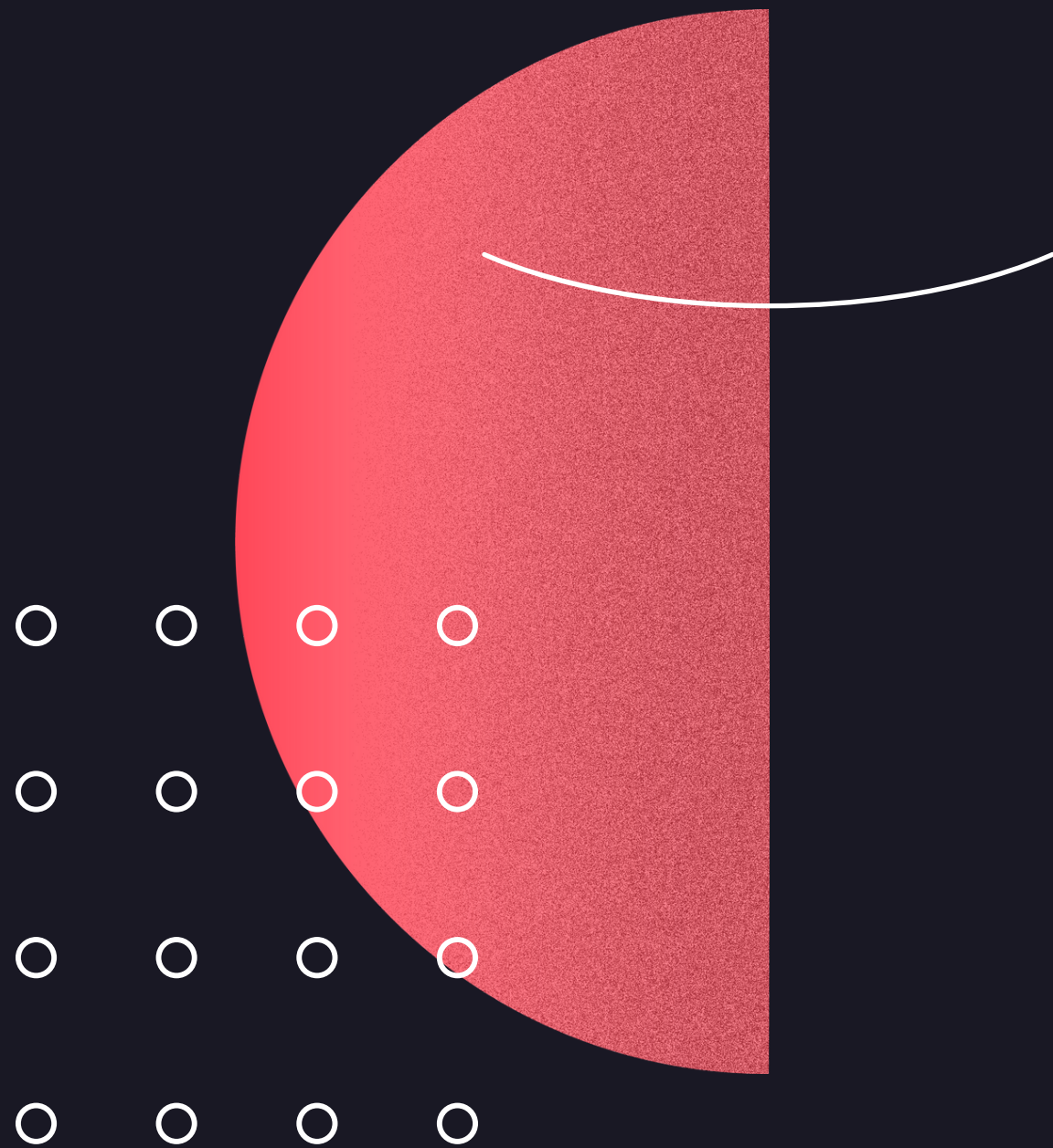
## MESO:

Larger-scale projects are often passed down from student to student. Pls work across projects with lots of different students.

## MICRO:

Your #1 collaborator is you from the past. Be nice to your favorite collaborator!

# Nobody's Perfect



- *Something* is better than nothing
- You can start incorporating data management practices at any point in your research cycle
- Do what works for you with the time/effort trade-off in mind
- Look into existing standards for data management
- Consistency is key

# DATA MANAGEMENT COMPONENTS

<https://guides.temple.edu/datamgmt/managingdata>

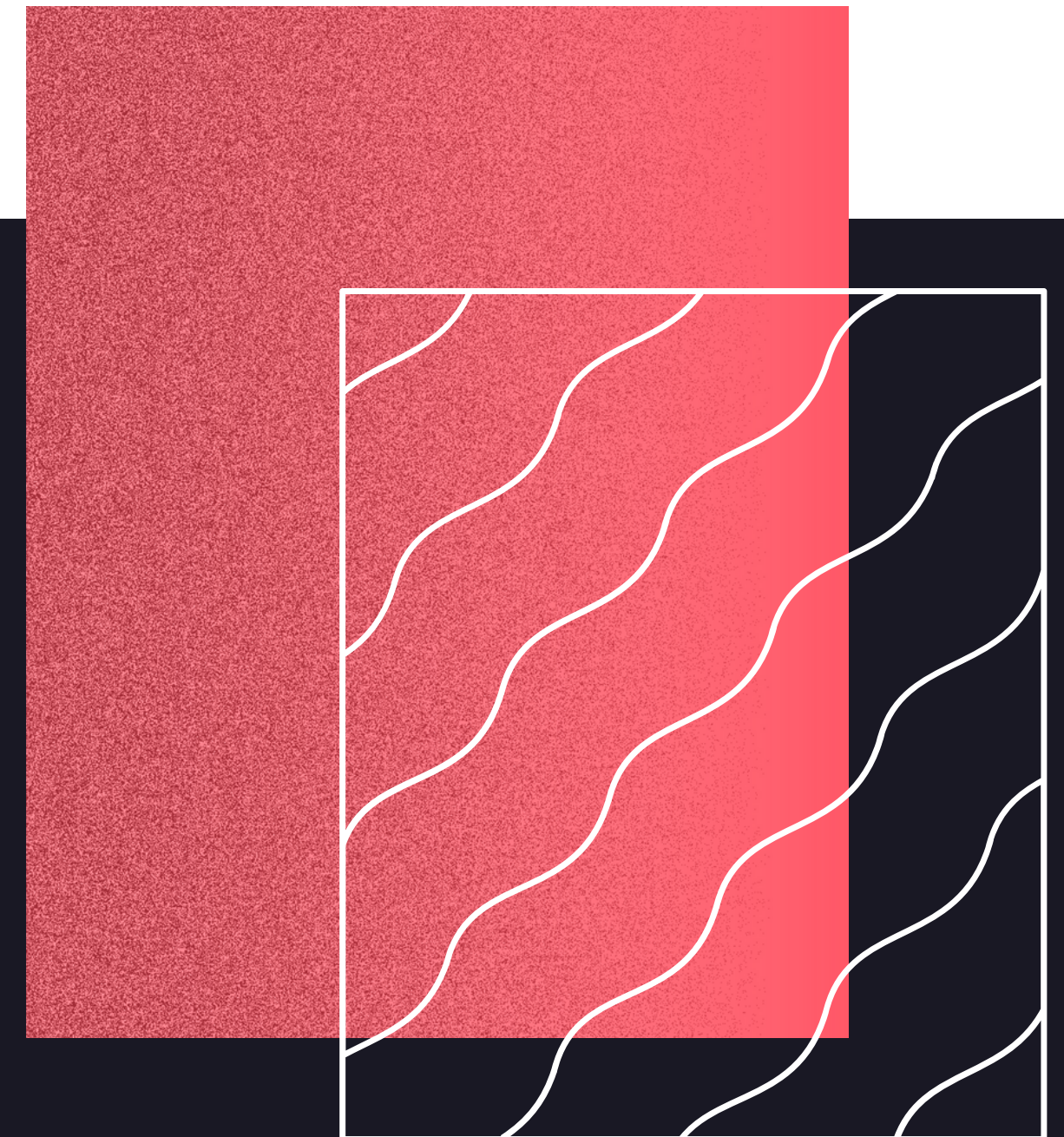
- Data dictionary / codebook
- File structures and naming conventions
- Metadata
- Storage and backups





# TODAY'S DEMO

We've collected all our subject level data, and want to run some preliminary models and visualizations.



# TOOLS USED IN TODAY'S DEMO



# OTHER USEFUL TOOLS





Liz Beard  
liz.beard@temple.edu  
[github.com/elizabethbeard](https://github.com/elizabethbeard)  
[@thetokenhipster](https://twitter.com/thetokenhipster)