**Communicating results**

* **In short**
  + - People are busy
    - Results of data analyses are sometimes presented in oral form, but often the 1st cut is presented via email
    - Useful to break down the results into chunks
* **Hierarchy of Information: Research Paper**
  + - Title/author list
    - Abstract
    - Body/results
    - Supplementary materials/the gory details
    - Code/data/really gory details
* **Hierarchy of Information: Email Presentation**
  + - Subject line/sender info
      * Minimum: include one
      * Summarize findings in one sentence
    - E-mail body
      * Brief description of problem, recall what was proposed and executed, summarize findings/results; 1-2 paragraphs
      * Suggest some options for actions if needed
      * If questions need to be asked, make them Yes/No
    - Attachments
      * R Markdown file
      * Knitr report
      * Stay concise
    - Links to supplementary materials
      * Code/software/data
      * GitHub repository/Project website

**RPubs**

* Rpubs.com

**Reproducible Research Checklist**

* **DO**
  + - Start with good science
      * Coherent, focused question simplifies many problems
      * Working with good collaborators reinforces good practices
      * Something that’s interesting to you 🡺 better
* **DON’T**
  + - Do things by hand
    - Editing spreadsheets of data to “clean it up”
      * Removing outliers
      * QA/QC
      * Validating
    - Editing tables or figures – rounding, formatting
    - Downloading data from a website (clicking links in a web browser)
    - Moving data around your computer, splitting/reformatting data files
  + Things done by hand need to be precisely documented (harder than it seems)
    - Point and click – do NOT
* **DO**
  + - Teach a Computer
      * If something needs to be done, try to teach your computer to do it
      * In order to give your computer instructions, you need to “write” down exactly what you mean to do and how it should be done 🡺 reproducibility guaranteed
    - Use Version Control
      * Slow things down
      * Add changes in small chunks
      * Track/tag snapshots
      * GitHub/BitBucket/SourceForge
    - Keep track of your software environment, version numbers

sessionInfo()

* **DON’T**
  + - Save any output
      * Avoid saving data analysis output (tables, figures, summaries, processed data, etc.)
      * Save the data + code that generated the output, rather than the output itself
* **DO**
  + - Set Your Seed

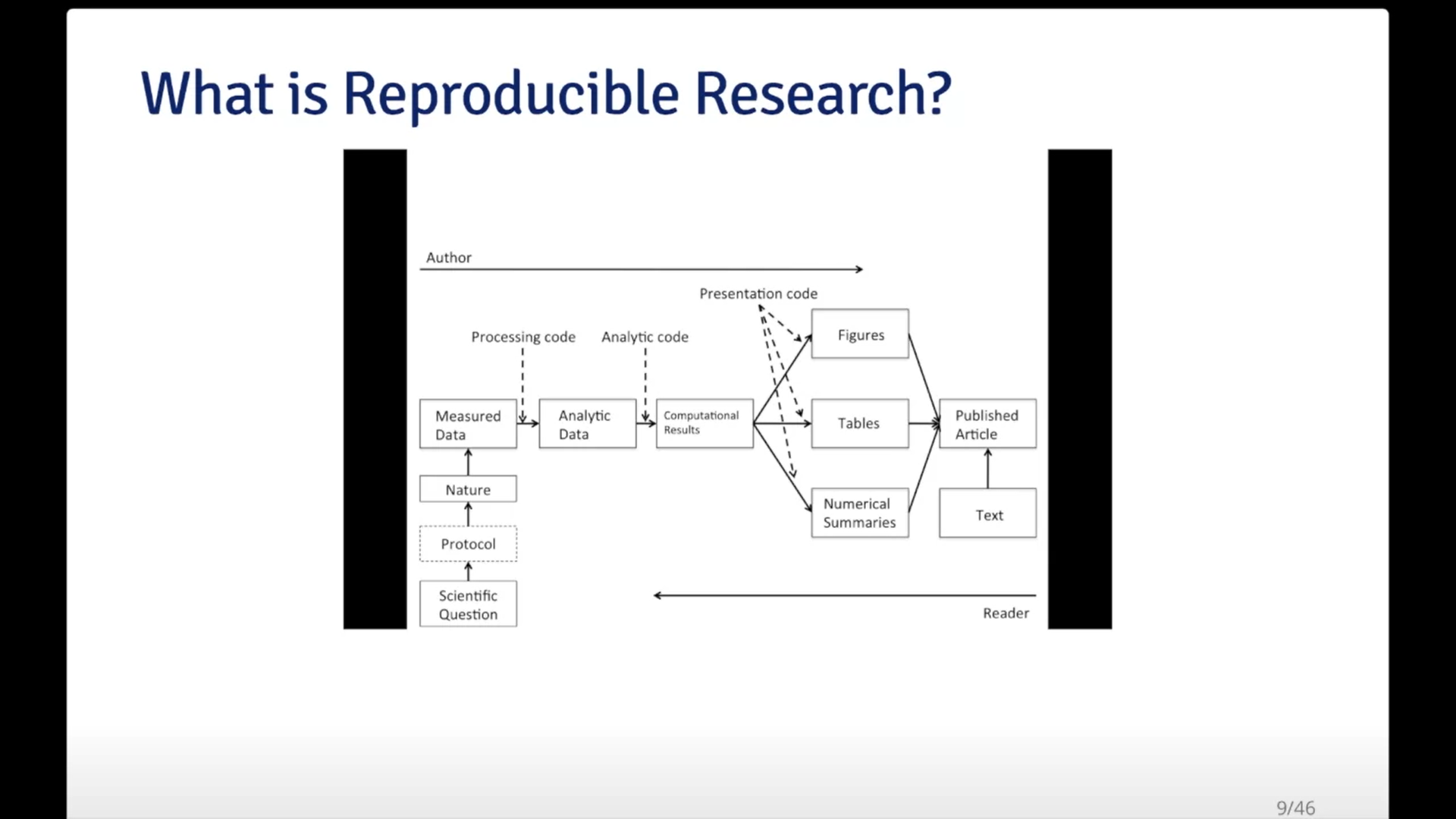
set.seed()

* + - Think About the Entire Pipeline

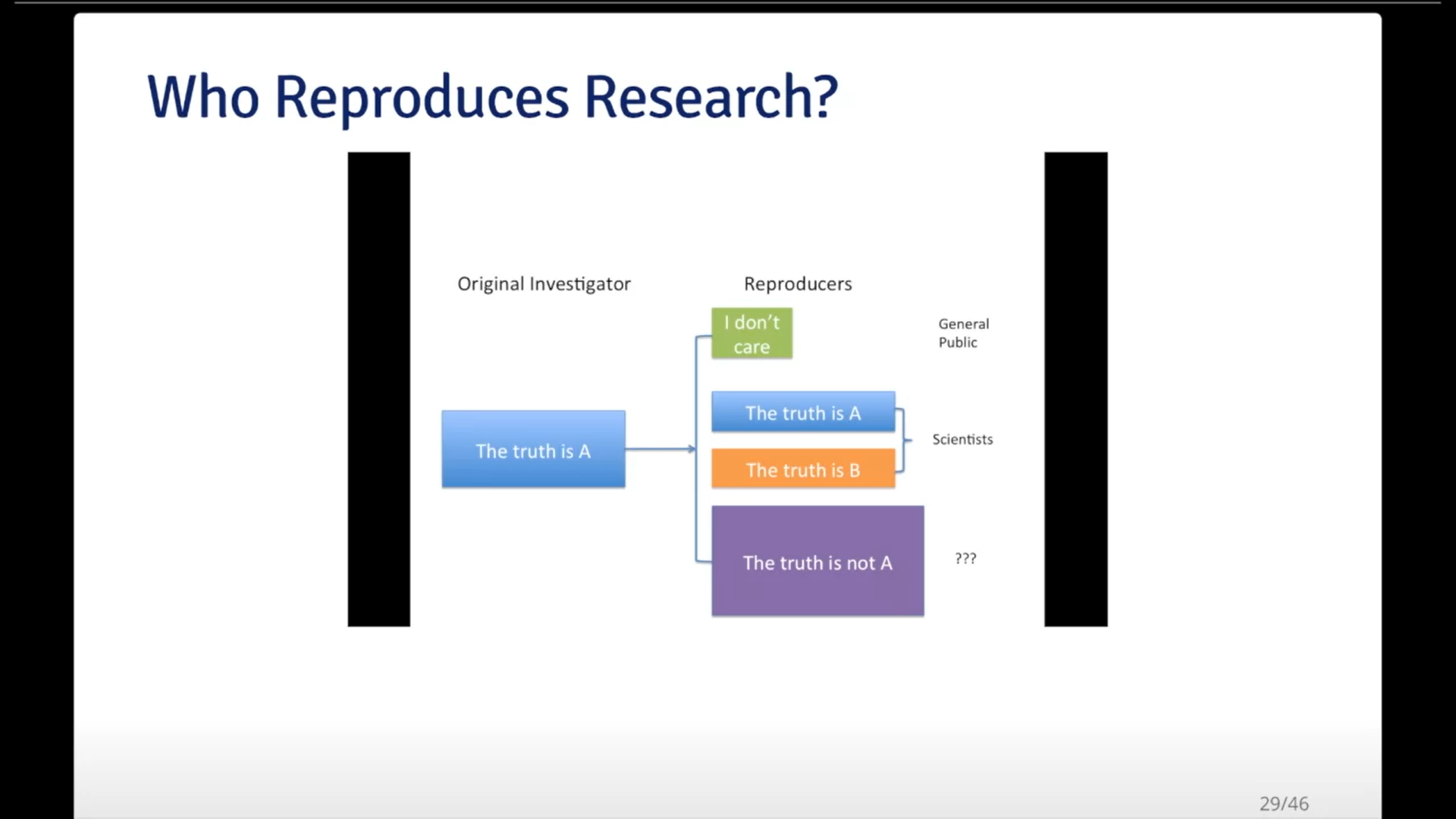
**Raw data 🡺 processed data 🡺 analysis 🡺 report**

**Evidence-based Data Analysis**

* **Replication and Reproducibility**
  + - Replication
      * Focuses on the validity of the scientific claim
      * “is this claim true?”
      * The ultimate standard for strengthening scientific evidence
    - Reproducibility
      * Focusses on the validity of the data analysis
      * “can we trust this analysis?”
      * A minimum standard for any study
      * New investigators, same data, same methods
      * Important when replication is impossible
* **Background and Underlying Trends**
  + - Some studies cannot be replicated
    - Data is more and more complex/multidimensional
* **The results?**
  + - Even basic analyses are difficult to describe
    - Complicated analyses cannot be trusted
* **What is reproducible research?**



* **What problem does Reproducibility solve?**
  + - Pros
      * Transparency
      * Data availability
      * Software/methods availability
      * Improved transfer of knowledge
    - Cons
      * Validity/correctness of the analysis not verified, an analysis can be reproducible and still wrong
* **Who reproduces research?**
  + - To check validity, need to
      * Rerun the analysis, check if results match
      * Check code for bugs/errors
      * Try alternate approaches, check sensitivity
    - Then who does it?



* **Evidence-based Data Analysis**
  + - Most data analyses involve stringing together many different tools and methods
    - Some methods may be standard for some fields, but others are often applied ad hoc
    - We should apply thoroughly studied, mutually agreed methods to analyze data whenever possible
    - There should be evidence to justify the application of a given method

**Summary**

* Reproducible research is important, but does not justify whether an analysis is trustworthy
* Evidence-based data analysis would provide standardized, best practices for given scientific areas and questions
* Gives reviewers an important tool without dramatically increase the burden on them
* More effort should be put into improving the quality of research, as REPRODUCIBILITY != RIGHT