Week 03
Intro to Reproducible Research

API209: Summer Math Camp

Rony Rodrigo Maximiliano Rodriguez-Ramirez

rrodriguezramirez@g.harvard.edu

Harvard University

August 26, 2024

Why Reproducibility?

Are We in a Crisis?

- The **replication crisis** in social sciences has highlighted significant issues in the credibility of research findings.

- Many high-profile studies have failed to replicate, raising concerns about the **reliability** of published results.

- The crisis has prompted a call for greater **transparency** and **rigor** in research practices.

The Replication Crisis

What Went Wrong?

- **Selective Reporting**: Only significant findings get published, leading to publication bias.

- **P-Hacking**: Manipulating data and analyses until nonsignificant results become significant.

- **Lack of Transparency**: Opaque methodologies that others cannot replicate or verify.

The Importance of Reproducibility

Building Trust in Research

- Reproducibility ensures that research findings are not just a result of **chance** or **specific conditions**.

- It allows others to **verify results** and build upon them, fostering cumulative knowledge.

- **Transparent reporting** of data and methods strengthens the credibility and utility of research.

How Can We Improve Reproducibility?

Adopting Best Practices

- **Pre-registration**: Outlining the study design and analysis plan before data collection.

- **Open Data and Code**: Sharing data and analysis scripts for others to verify and use.

- **Reproducible Workflows**: Using tools like Quarto to create dynamic documents that combine analysis and narrative.

Quarto: A Tool for Reproducible Research

What is Quarto?

Quarto is an open-source scientific and technical publishing system that enables researchers to create dynamic documents, reports, presentations, and websites.

Why Quarto?

The Need for Reproducible Research

- **Reproducibility** is a cornerstone of scientific research.

- Quarto ensures that your analysis and outputs (tables, figures, etc.) can be reproduced by others, enhancing the credibility of your work.

- **Integrated with R, Python, Julia, and Observable**: Quarto supports multiple languages, making it versatile for various research needs.

Key Features of Quarto

1. **Dynamic Documents**: Create documents that are automatically updated with the latest data and analysis.

2. **Multiple Outputs**: Generate reports, presentations, blogs, and books from a single source.

3. **Version Control**: Integrates seamlessly with Git for version control, tracking changes, and collaboration.

4. **Cross-Platform**: Works with RStudio, VSCode, or directly from the command line.

Why Use Quarto for Your Problem Sets?

Consistency and Organization

- Quarto helps you **organize your code, analysis, and narrative** in a single document.

- It ensures that your problem sets are **well-documented** and **easily understandable**.

Hands-On Practice

What We'll Do This Week

- **Set up Quarto**: Install and configure Quarto on your systems.

- **Create Your First Quarto Document**: Practice by creating a simple document that integrates text, code, and visualizations.

- **Reproducible Problem Sets**: Learn how to structure your problem sets to be fully reproducible.

Getting Started with Quarto

Installing Quarto

1. **Download and Install**: Visit quarto.org and download the latest version for your operating system.

2. **Setting Up in RStudio**: Quarto integrates directly into RStudio. Once installed, you can start creating Quarto documents right from the IDE.

Creating Your First Quarto Document

A Quick Demo

- **Open RStudio** (or your preferred editor).

- **New Quarto Document**: Go to File > New File > Quarto Document.

- **Choose Format**: Decide between HTML, PDF, Word, or a presentation format like Reveal.js.

- **Add Content**: Start writing text, and insert code chunks to see how Quarto integrates code and narrative.

Reproducible Research with Quarto

Integrating Code and Narrative

- **Code Chunks**: Embed your analysis directly in the document.

- **Automatic Updates**: Quarto will automatically re-run code and update results when the data changes.

- **Easy Collaboration**: Share your Quarto files with others, and they can re-run the entire analysis with a single command.

Best Practices for Quarto Documents

Structuring Your Document

1. **Use Sections**: Organize your content into clear sections and subsections.

2. **Comment Your Code**: Make sure that each code chunk is well-commented for clarity.

3. **Version Control**: Regularly commit changes to your Quarto documents using Git.

Quarto in Action

Real-World Examples

- **Research Papers**: Automate the creation of research papers with Quarto, ensuring all analysis and plots are reproducible.

- **Presentations**: Create dynamic presentations like this one, where code can be executed live during the talk.

- **Blogs and Websites**: Share your research with the world by publishing Quarto documents as websites or blogs.

Summary and Next Steps

Why Quarto is Essential

- Ensures reproducibility, consistency, and clarity in your research.

- Simplifies the creation of dynamic documents and presentations.

- Helps you produce high-quality, professional outputs that can be shared and reproduced.

Let's Practice!

- **Lab Activity**: In our lab sessions this week, we'll apply what we've learned by creating and refining Quarto documents for your problem sets.

Questions?

Feel free to ask any questions as we dive into Quarto this week!