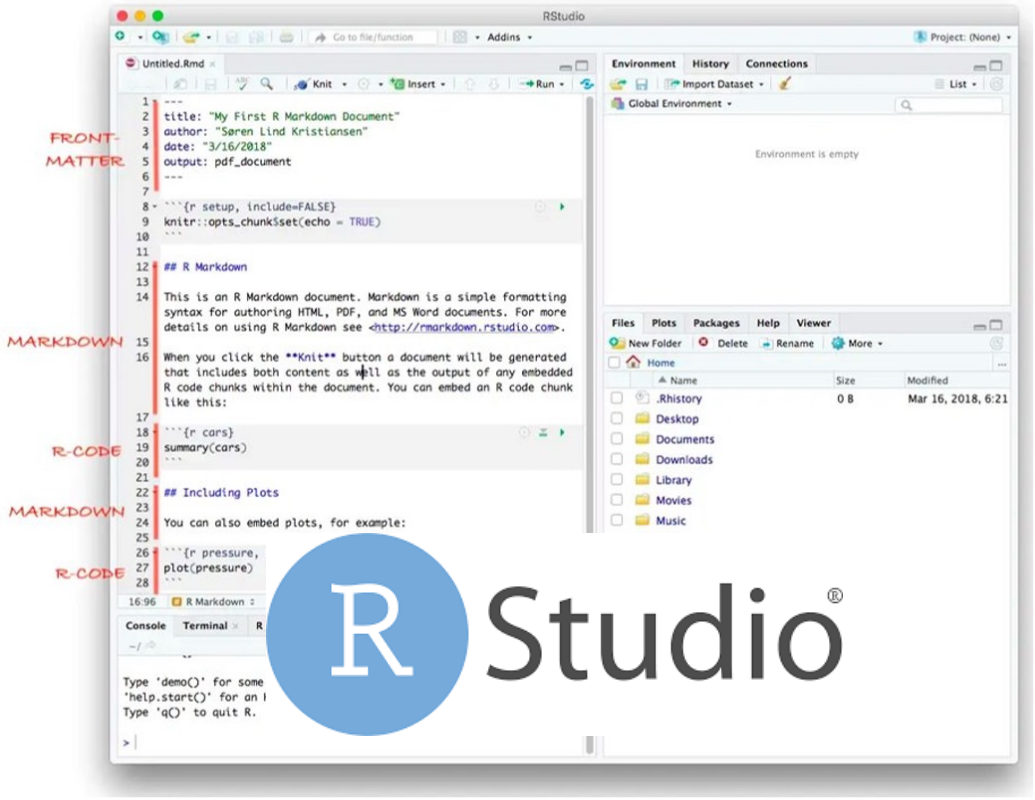
SIH Masterclass Series

Getting started with Git in RStudio: a hands-on guide

27th April 2023

Dr. Nandan Deshpande

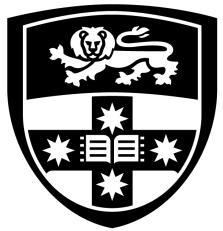


THE UNIVERSITY OF SYDNEY

Sydney Informatics Hub

Sydney Informatics Hub

“SIH is a Core Research Facility of the University of Sydney enabling excellence in computational and data-driven research through advanced digital infrastructure, expert data consultancy and analytics training”



THE UNIVERSITY OF SYDNEY

Sydney Informatics Hub

https://sydney.edu.au/sydney-informatics-hub

Sydney Informatics Hub

Empowering researchers with modern data & computational methods

Research Computing

- High performance computing
- Bioinformatics & genomics
- Modelling & visualisation
- Computing training

Statistical Consulting

- 1-on-1 consultancy for HDR-level and above.
- Experiment and survey design.
- Statistical model development and testing
- Statistics training

Data Science & SWE

Consulting and project collaboration providing analysis & software development for data-driven research.

Consultation

Grant application support

Data Science, ML, modelling & analytics

Research compute platforms

Hacky Hour

Training

https://sydney.edu.au/sydney-informatics-hub

Learning Objectives

The primary objective: Use RStudio environment to backup/version control R Markdown-projects using GitHub

1. Introduce R/RStudio/R Markdown

2. Why backup and do version control?

3. GitHub for backup/ version control

4. Set up a GitHub repository, make local changes and back up using RStudio.

Programming in R

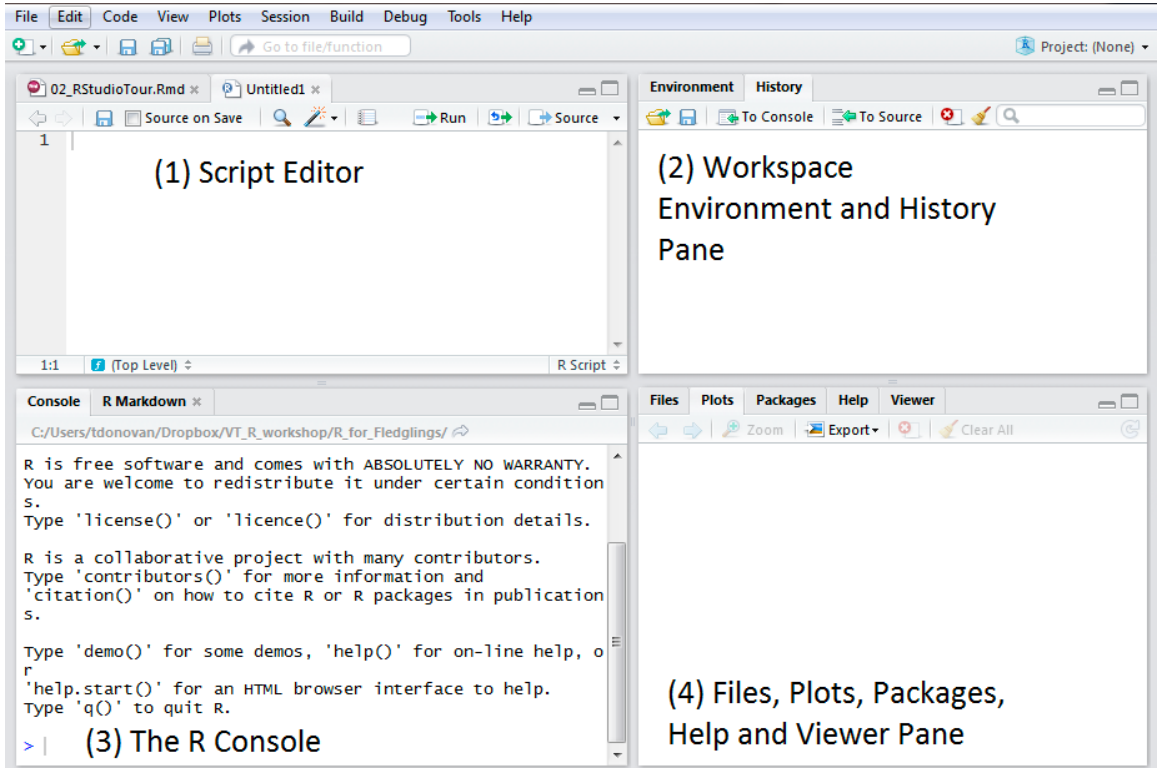
R is a programming language and software environment which can be used for a wide range of tasks, including:

- Data exploration and manipulation
- Statistical modeling and analysis
- Data visualization
- Machine learning
- Scientific computing and simulations
- Reproducible research

RStudio

- RStudio is an Integrated Development Environment (IDE) for the R programming language.

- It provides a user-friendly interface for writing and executing R code.

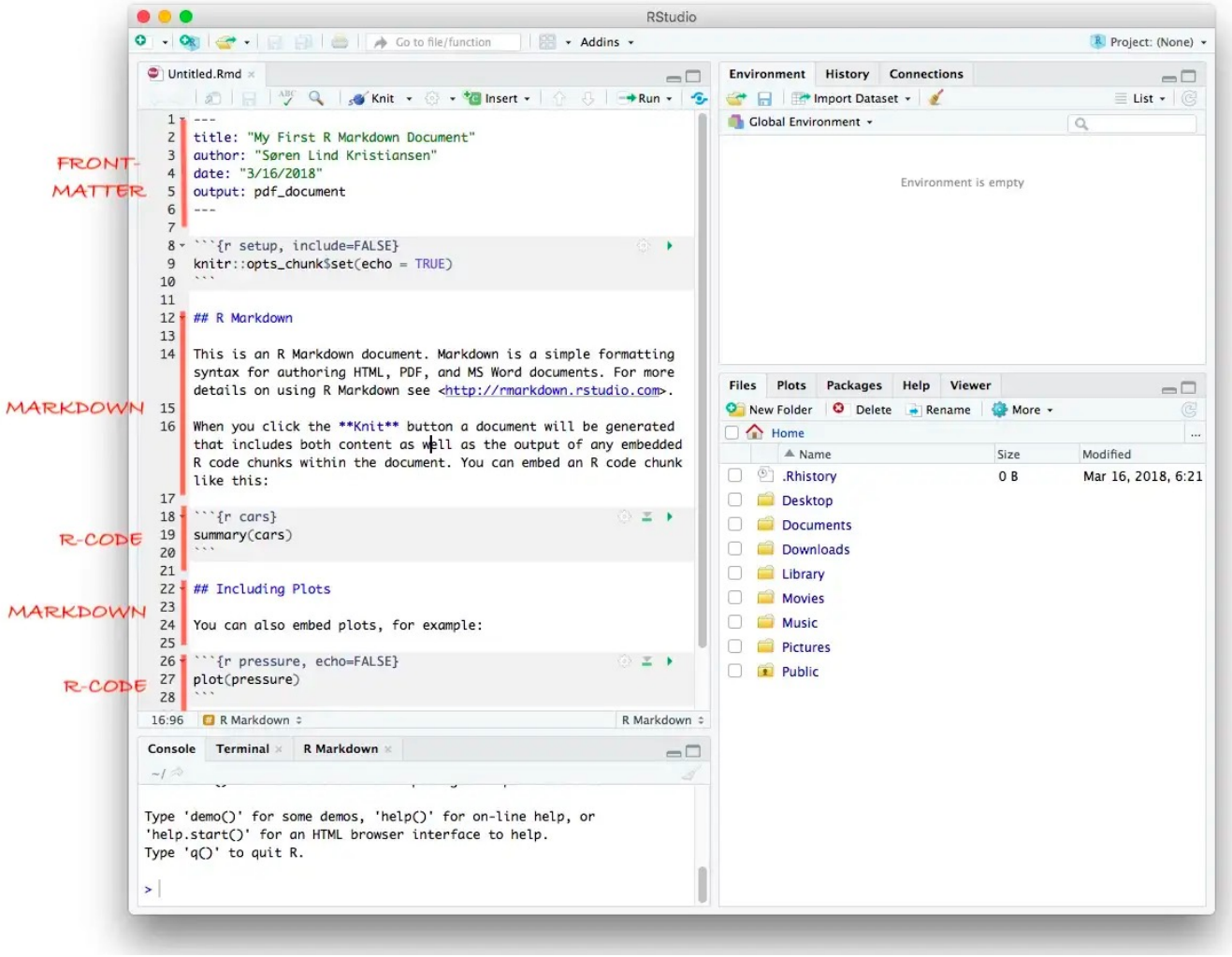


- It also provides several tools to make data analysis and visualization easier.

- It makes working with R much more efficient and enjoyable.

RStudio and R Markdown

- R Markdown combines text, code and results
- RStudio provides the flexibility to display them across multiple panels.
- R Markdown output can be easily rendered into a variety of formats (HTML, PDF, and Word documents)



Applications of R markdown

- Data Analysis and Reporting
- Scientific Research
- Business Intelligence
- Data Journalism
- Education
- Technical Documentation

Backup and version control

Backup and version control are important aspects of document management:

Backup: A process of making copies of data so that it can be restored in the event of a data loss.

Version control:

- A system that records changes to a file or set of files over time so that you can recall specific versions later.
- Allows multiple users to collaborate on a project.

Different software options for version control

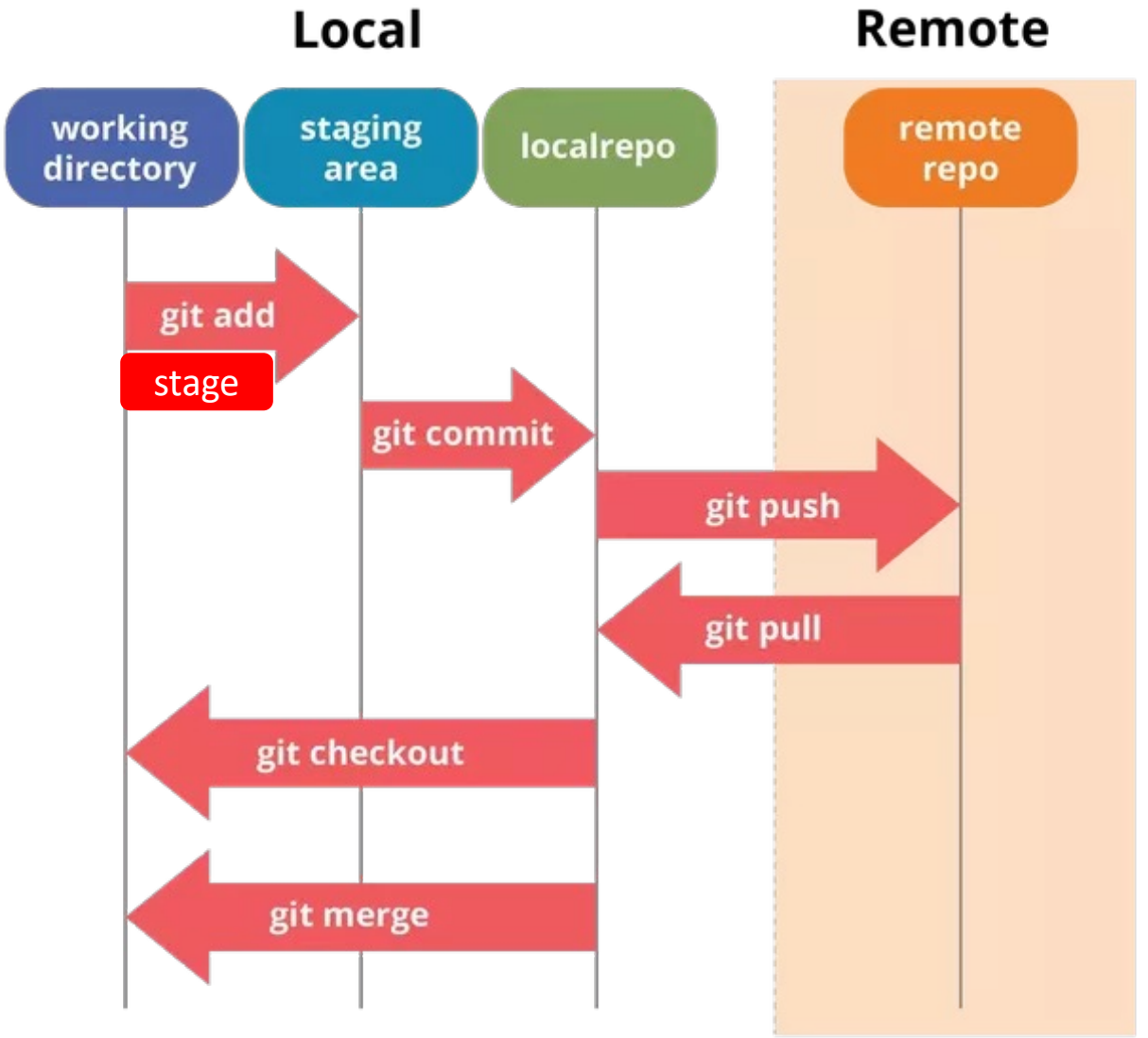
Git is a popular open-source distributed version control system.

In addition to Git, there are other version control systems such as:

- Subversion (SVN)
- Mercurial (Hg)
- CVS (Concurrent Versions System)

Using GitHub for R markdown

- Backup
- Version control
- Collaboration
- Reproducibility



Pre-workshop set-up

1. **Internet Connection**: You will need a stable internet connection.

2. **GitHub Account**: Sign up for a free account at https://github.com/.

3. **Git**: Git is a version control system that GitHub uses to manage files.

   Install Git using the link. https://git-scm.com/downloads.

4. **R and RStudio**: You must have R and RStudio installed on your local system to create and edit R Markdown and other files.

Details for all setup –related information has been provided to the attendees via the Menu option "Set-up" in the link: https://sydney-informatics-hub.github.io/RStudio_github_versioncontrol/

Which Github?

1.University of Sydney Enterprise GitHub :
https://github.sydney.edu.au/

2.GitHub.com:
https://github.com/

3.Knowledge base article explaining the difference between the university instance and the public GitHub:
https://sydneyuni.service-now.com/sm/?id=kb_article_view&sysparm_article=KB0010881

The link for masterclass: https://sydney-informatics-hub.github.io/RStudio_github_versioncontrol/

Collaborative changes, conflicts ...

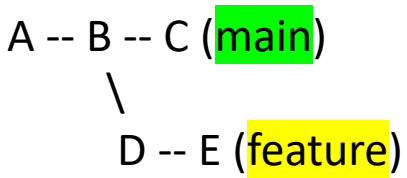
- Collaborate with other group members
- Use of Git branches
- Basic rules of thumb to avoid merge conflicts
  - Communicate often
  - Tell each other what you are working on
  - Pull before you change, commit or push
  - Commit often in small chunks.

A good workflow when working with git as a team:

Pull -> Edit -> Add -> Pull -> Commit -> Push

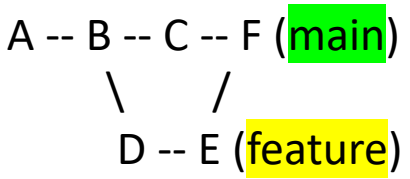
Collaborative changes and git branching

Two branches: main and feature.



- Main branch contains three commits (labelled A, B, and C).
- Feature branch contains two commits (labelled D and E).
  - To work on a new feature or fix a bug

The resulting branch history might look like this:



When the feature is complete, the developer can merge the changes from the feature branch back into the main branch

- The basic idea is that branches allow developers to work on changes to their project independently, and
- Merge those changes back into the main project when they are complete.

- Git branching can be much more complex.

Feedback

Thank you for attending the session! Please use the link or scan the QR code for registering your feedback.

https://tinyurl.com/ycymtppr

For more details on Git

- Online resource about using Git with RStudio https://happygitwithr.com/index.html
- Introduction to Git: https://sydney-informatics-hub.github.io/intro-git/
- Atlassian basic intro to version control and Git https://www.atlassian.com/git/tutorials/what-is-version-control

Thank you

- Giorgia Mori
- Meredith Tucker
- Michael Lynch
- Xinwei Luo
- Everyone at Sydney Informatics Hub ☺

Thanks for attending