R FOR LUNCH

Reproducible workflows

John Little
Center for Data & Visualization Sciences
Duke University Libraries

2024-02-19

TODAY'S TOPICS

- RStudio (Projects & Setting/preferences)
- Generate reports from code (Quarto Scientific Publishing system)
- Publishing reports

You should have the latest version of Quarto!!

REPRODUCIBILITY

What is a reproducible workflow and why does it matter?

STREAMING PREVIOUS WORKSHOPS

R for Lunch: a lunchtime learning series

- IDE and Import data (RStudio IDE, Import data, Code notebook)
- Wrangle data {dplyr}
- Visualizing with {ggplot2}
- Mapping and Spatial Analysis

See Also: Online Resources: (Rfun | CDVS resources)

HOUSEKEEPING

- Drew / Lauren / breakout rooms
- CDVS
  - Themes
    - Data Management (Plans, Reproducibility, Repositories)
    - Data Science
    - Data Visualization
    - GIS and Spatial Analysis
    - Data Sources

HOUSEKEEPING CONTINUED

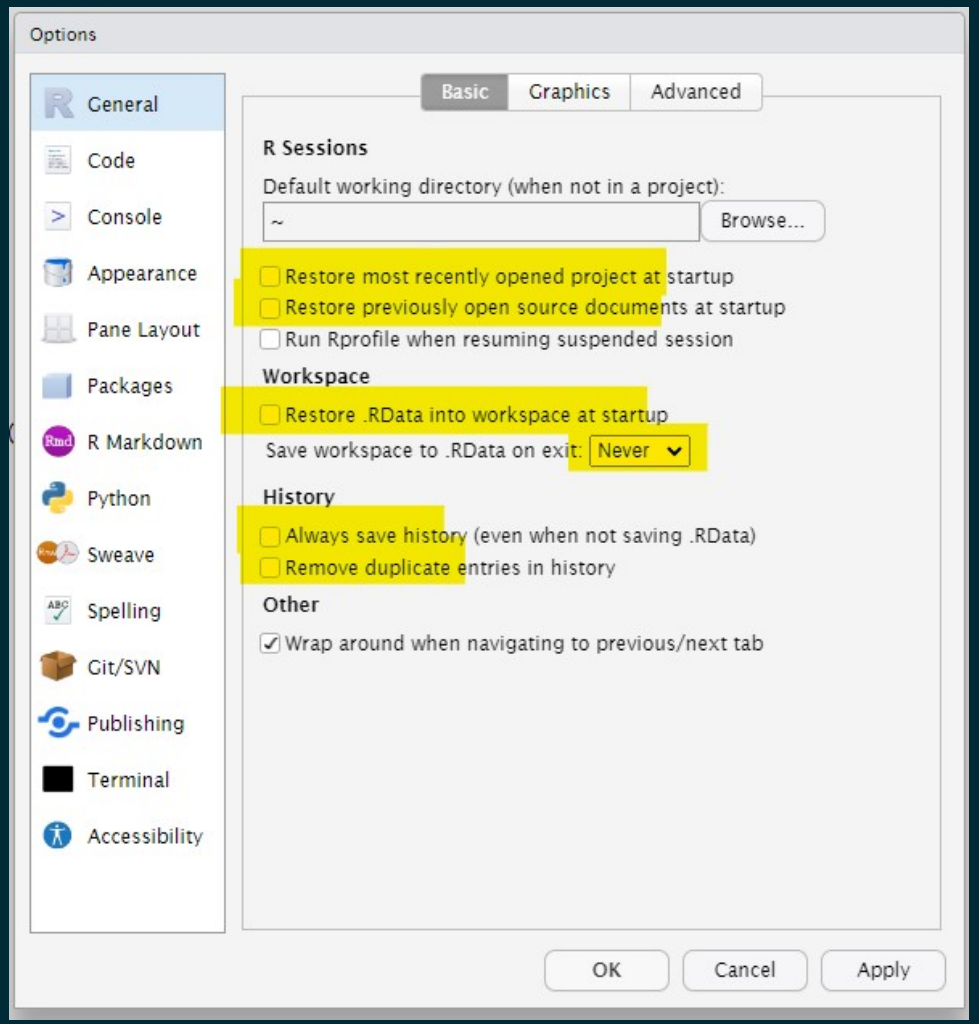
- Website - https://library.duke.edu/data
- Workshops
  - https://library.duke.edu/data/workshops
- Consulting in the Lab
  - askData@duke.edu
  - my schedule: https://is.gd/littleconsult

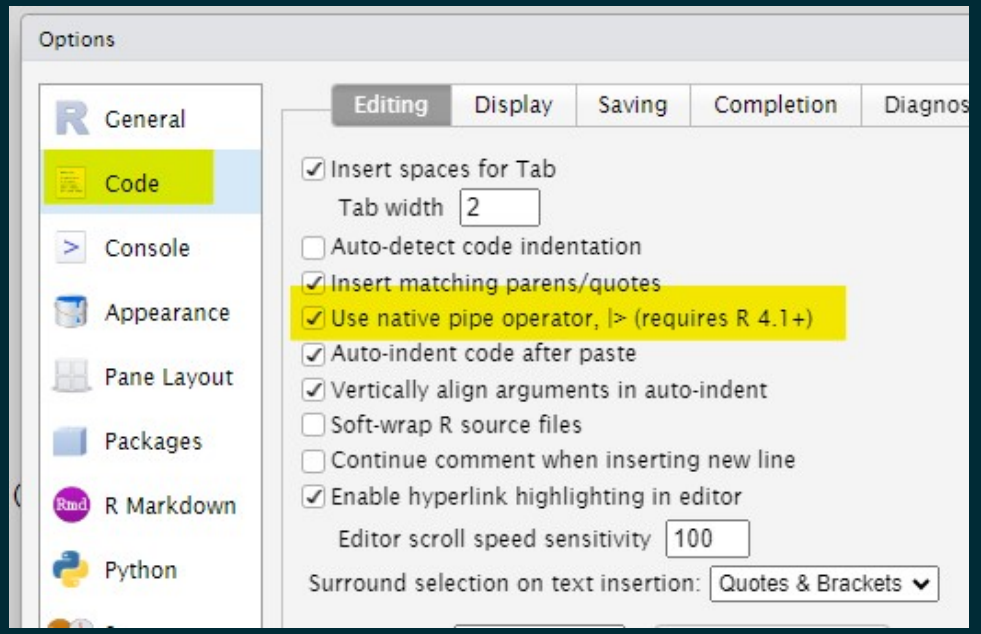
IDEAL: R AS A PRACTICAL REPRODUCIBLE WORKFLOW

- Code in RStudio
- Report types: A select list of Quarto output formats
  - these slides (Live | Code and PDF on GitHub)
  - the *Introduction to R/Tidyverse/Quarto* text.
  - Manuscript: a framework for writing and publishing scholarly articles (live example)
- Use Git and GitHub for version control, code sharing, and collaboration

REPRODUCIBILITY SETTINGS FOR YOUR IDE

- RStudio (Projects & Setting/preferences)





Global options > Code ^^

<- Global options > General

PUBLISH

- Full Instructions
- Steps (simplified)
  1. In **CLI** (terminal, powershell, bash, etc.): `quarto publish` (For select Quarto project types)
  2. Choose a destination (e.g. quarto.pub or netlify.com)

Or, drop and drag to Netlify (A very simple alternative Netlify targeting approach)

Or, GitHub Pages YMMV. Different options. Most advanced leverages GitHub Actions

Or, other targets (see Instructions link, above)

MANUSCRIPT DEMO

- Live Demonstrations
  - my take | documented at GitHub
  - Posit documentation | documented

INSTRUCTIONS SIMPLIFIED

1. In **CLI** (e.g. RStudio Terminal, PowerShell, Bash, etc.):
   `quarto create` > `project` > `manuscript`
   - Identify LFS path / location
2. Go to that location; double-click <new_proj_name>.Rproj file (if using R)
3. `useths::use_git()`
4. in `_quarto.yml`: change `jats: default` to `pdf: default`

{RENV}

Create reproducible environments for your R projects

- Get Started
  - `renv::init()`
  - `renv::snapshot()`
  - and `renv::restore()`

BINDER (CONTAINERS)

The Binder project makes it easy to share compute environments. This is documented very well at Quarto.

In essence:

- CLI terminal: `quarto use binder`

Recommended: While sharing your compute environments consider sharing your GitHub Releases while simultaneously minting DOIs. This makes your content citable and places your milestones into the Zenodo archive for posterity.

POLLS

WE ARE HERE TO HELP

- askData@duke.edu
- https://library.duke.edu/data
- https://is.gd/littleconsult

LET'S DO IT

TWO THINGS FOR TODAY

- five essential {dplyr} data wrangling verbs
- data pipes inside code-chunks
- https://intro2r.library.duke.edu/wrangle.html

EXERCISES

1. https://intro2r.library.duke.edu/ > Exercises > Link out > Green **Code** button > Download ZIP
2. Then, Unzip (i.e. Expand) the folder (on your local file system)
3. Then, double click the `rforlunch_exercises.Rproj` file
4. From **RStudio** the Files tab, open the `01_dplyr.qmd`
    - The answer file is in the **RStudio** `rforlunch_exercises` project > `Files Tab > Answers` folder

CLOSING

CITATION MANAGEMENT

RStudio > Quarto Notebook > Insert > Citation

Example DOI: 10.18637/jss.v059.i10

WHERE TO FIND

- These slides
  - A PDF copy
- Code for above

BYE FOR NOW

- askData@duke.edu
- https://is.gd/littleconsult
- https://library.duke.edu/data