R FOR LUNCH

Reproducible workflows

John Little iD

Center for Data & Visualization Sciences

Duke University Libraries
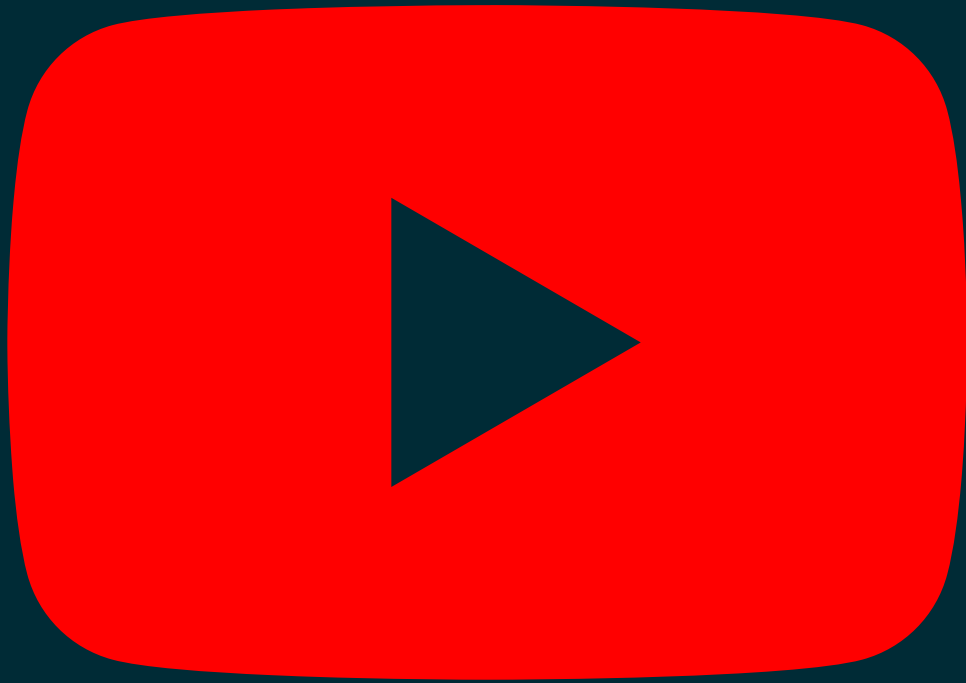
2024-02-22

R fun

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

YouTube playlist

Sophia Lafferty-Hess expertly explains the foundation of reproducibility and why it matters.

R fun

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)

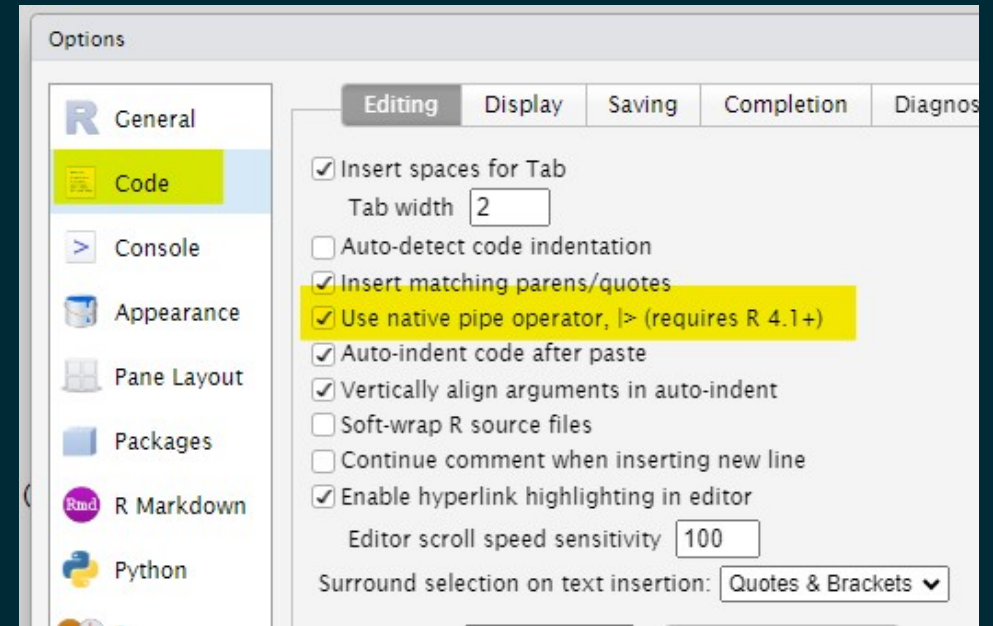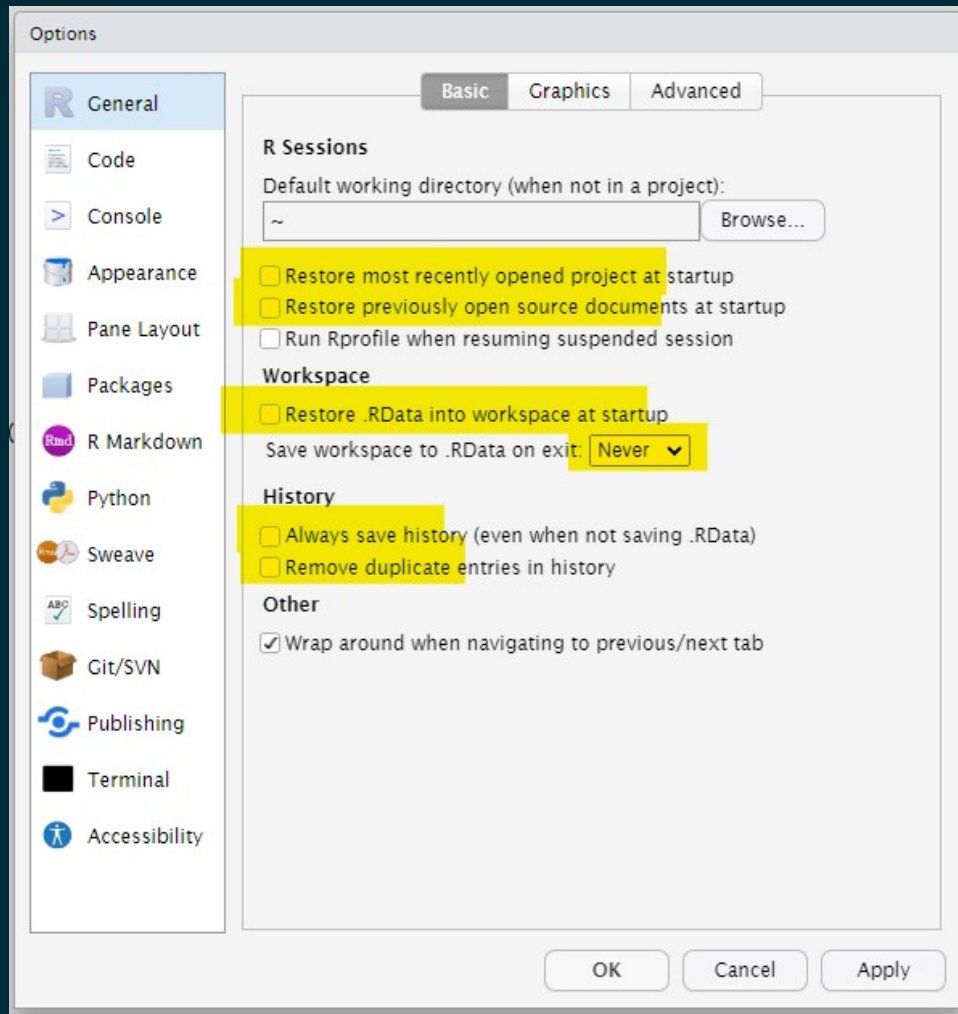- Use Git and GitHub for version control, code sharing, and collaboration

R fun

PROJECTS

- Use the projects feature of RStudio

`New Project > Quarto Project`

- **Do use code notebooks**

- **Do use** *relative file paths* and {`here`}

- **Do use** *Restart R and Run All Chunks*

- *Do not rewrite code* for different computers

- *Do not use* `setwd()`

- *Do not use* `rm(list=ls())`

REPRODUCIBILITY SETTINGS FOR YOUR IDE

Global options > Code ^^
<- Global options > General

CITATION MANAGEMENT

Use a bibliography manager such as Zotero

Then,

RStudio > Quarto Notebook > Insert > Citation

Example DOI: 10.18637/jss.v059.i10

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)

R fun

MANUSCRIPT DEMO

- Live Demonstrations

  - Live example for this workshop | code at GitHub

    - Based on Posit demo | documentation

INSTRUCTIONS SIMPLIFIED

1. In **Rstudio**: `File` > `New project` > `New Directory` > `Quarto manuscript`

2. ☑ Create a git repository (**optional**)

3. in `_quarto.yml`: change `jats: default` to `pdf: default` (**optional**)

JATS creates a MECA bundle, i.e. a standard manuscript exchange format.

{RENV}

Create reproducible environments for your R projects by storing the exact versions of the code interpreter and r packages within the version-control repo

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

BINDER (CONTAINERS)

The Binder project makes it easy to share live 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.

WE ARE HERE TO HELP

- askData@duke.edu

- https://library.duke.edu/data

- https://is.gd/littleconsult

CLOSING

WHERE TO FIND

- These slides
  - A PDF copy

- Code for above

R fun

BYE FOR NOW

- askData@duke.edu

- https://is.gd/littleconsult

- https://library.duke.edu/data