Recap
Tidying Up GitHub
OpenRefine assignment
Feedback

Elizabeth Stregger

DATA 3101

Sep. 22, 2022

Summary of what we've covered

Tools for working with data:

- Installed Rstudio and Git
- Connected to MTA DATA 3101 Organization
- Installed OpenRefine

Data knowledge:

- Scholarship lifecycle
- Examples of data problems (Retraction watch)
- Finding data for reuse
- Examining data for inconsistencies and correcting them

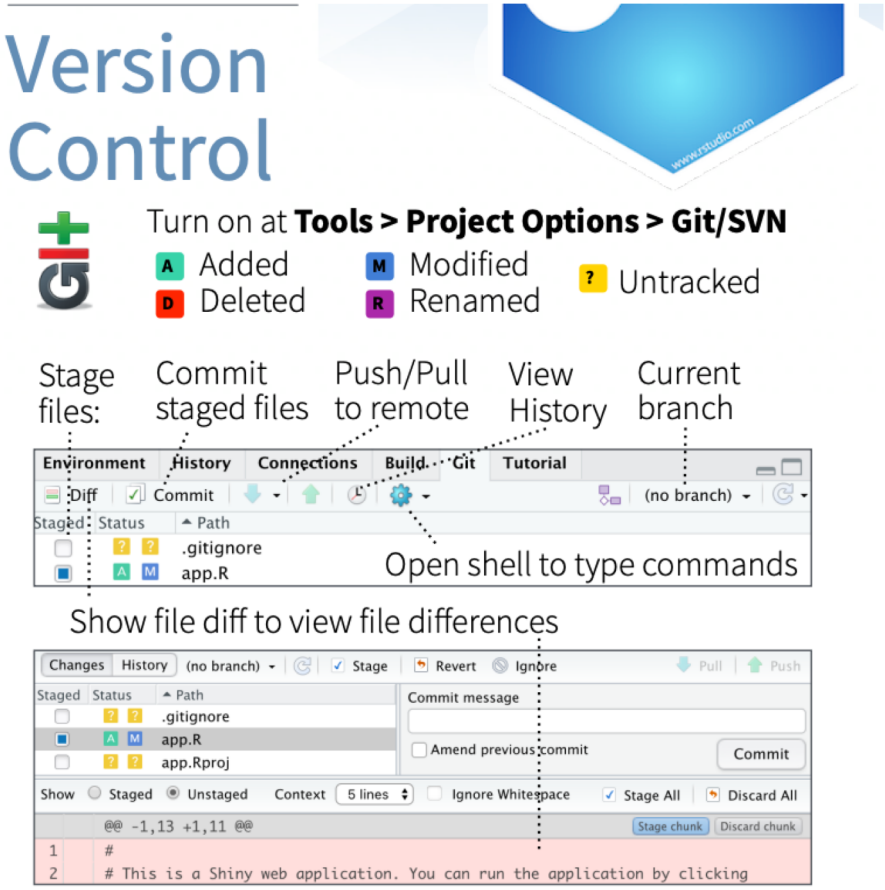
Assignments / in-class activities

- Data management reflection
- Retraction Watch example
- Finding a dataset
- OpenRefine (today)
- Feedback (today)

Question: What do the letters and icons in the Git pane in RStudio mean?

- See RStudio IDE Cheatsheet for letters and colours
- Column on right is changes on your computer
- Column on left is staged for commit

RStudio cheatsheets: https://www.rstudio.com/resources/cheatsheets/



Question: How do I add a file not created in RStudio to GitHub? (example: PNG on research)

- Creating the PNG:
  - Draw the data management flow diagram (on paper is fine!)
  - If on paper, take a photo
  - In your file viewer, export to PNG and give the file a useful name
- Adding to GitHub
  - Move the PNG into your repository folder on your computer
  - The new file should appear in your Git pane in Rstudio
  - Stage, commit, and push file

Working with GitHub

- Real life version of your project – messy, files on your computer
- Official version of your project – clean, committed history
- Can choose to “ignore” files you create but do not intend to push to GitHub

We made a bit of a mess when we tested that Rstudio could push to GitHub. Let’s clean it up today.

Tidying up GitHub

- Delete extra files in GitHub – with caution!!
  - Then in RStudio, pull changes
- Edit assignment files with RMarkdown or the Visual Editor and push changes
- Info about RMarkdown: https://rmarkdown.rstudio.com/lesson-1.html
- Create issues when you're ready for feedback

OpenRefine in-class assignment

- Download the data files you chose in the Finding Data class
- What format are the files? Can you open them in OpenRefine? OpenRefine can open tab separated (tsv), comma separated (csv), Excel (xls , xlsx), JSON, XML, RDF as XML, and Google Spreadsheets.
- If not, find a different dataset for this assignment
- Import a data file from the dataset you chose into OpenRefine. Facet, filter, and cluster some of the columns. Are there any data problems to be corrected? Do you have questions about the dataset? Write a summary and add it to your repository.

Feedback on course so far

- Create a new file in your repository to give us feedback on the course
- One up / one down
  - What is one thing you like about the course so far?
  - What is one thing that can be improved?

Next: Data in R

- Doug will be lead instructor for this section (2 - 3 classes)
- Classes will meet on Teams when Doug is teaching