

Workflow Languages

<https://dbsloan.github.io/TS2022/>

[illegible]

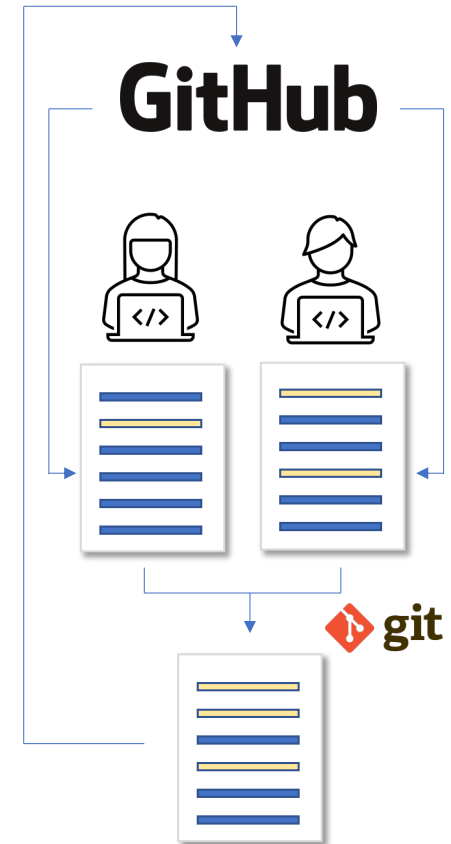
Figuring out how to do a particular analysis can be time consuming and it is common to have to repeat the analysis on the same dataset or on additional datasets.

Best practices in genomics data analysis include documentation, version control, reproducibility, rigor, and automation

How to Keep Track of Your Workflow

It is important to keep track of your computational experiments just as you do your wet lab experiments. This can be accomplished by using GitHub to document your code, changes to your code, issues with your code, notes on your code. An electronic lab notebook is also useful.

Best practices often include GitHub and lab notebooks



Workflows and Automation

Bash, R, Python, and Jupyter Notebooks can all be used to document data analysis steps and to chain together multiple steps in a workflow. They can also be used to automate workflows.

Workflow languages are becoming popular because they provide a very structured and reproducible approach to creating and running workflows.



nextflow



See this article on workflow languages: <https://www.nature.com/articles/d41586-019-02619-z>

Learning a workflow language will have tremendous payoff in terms of reproducibility, automation, and time savings.