Worksheet

Rohan Alexander

2025-09-08

1. Inspired by those presented in the lecture, write down your own statistical workflow.

2. For each aspect of your workflow, rate how confident you are about your ability.

3. Exchange with another member of the class and make a suggestion for how their workflow might be improved.
4. Review the plan - https://rohanalexander.com/courses/statistical_data_science.html - for the class and note any concerns or suggestions.

5. Decide on a partner, and pick a week that you would like to own. Make a GitHub account if needed, then branch and make a PR to the class repo - https://github.com/RohanAlexander/STA4101 - adding a folder for that week (nicely formatted e.g. week_10), with a README file that contains your names.
6.

- Download and install uv: https://docs.astral.sh/uv/getting-started/installation/.
- Download and install VS Code.
- Download and install Quarto.
- Use uv to install Python: https://docs.astral.sh/uv/guides/install-python/.
- Use uv to run a script: https://docs.astral.sh/uv/guides/scripts/.

These are all from Timbers et al. (2025)

1. Which of these is untrue about the Git and GitHub version control software?

    a. Allows you to view and/or retrieve older snapshots of the files and directories in a project.
    b. Automatically snapshots your work every 2 minutes.
    c. Provides transparency on who made what changes to files and directories in a document.
    d. Can act as a way to back-up your work.

2. GitHub is the software you use locally on your computer (i.e., your laptop) to commit changes to the version control history. True or False?

    a. True
    b. False
    c. Neither true or false.

3. You changed two files (notes.txt and eda.ipynb) but you only want to commit changes to one of them (eda.ipynb) to the version control history. Which Git command allows you to specify this?

    a. Add
    b. Commit
    c. Push
    d. Push

4. At a minimum, how often should you push your work to GitHub?

    a. Every 5 min.
    b. Every 30 min.
    c. At the end of every work session.
    d. Once a week.

5. You try to push your most recent commit from your locale version control repository to your remote repository on GitHub and it fails because Git says the remote contains work that you do not have locally. What do should you do next?

    a. Commit the changes you made recently in your working directory.
    b. Force push your changes.
    c. Pull the changes from the remote repository that you do not have locally.

6. You pull changes that exist in your remote version control repository on GitHub that you do not have in your local version control repository, and you get the message

    Auto-merging in CONFLICT (content): Merge conflict in Automatic merge failed
    fix conflicts and then commit the result.

What do you need to do?

a. Push the changes from the local repository that you do not have remotely.
b. Force pull the changes.
c. Manually open the file with the conflict and edit it to have the desired version of the changes, as well as remove the special Git syntax used to identify the merge conflict.

References

Timbers, Tiffany A., Joel Ostblom, Florencia D'Andrea, Rodolfo Lourenzutti, and Daniel Chen. 2025. *Reproducible and Trustworthy Workflows for Data Science*.