Version control Episode Notebook

Part of FAIR in (bio) practice, <https://carpentries-incubator.github.io/fair-bio-practice>

Type your name and institution:

-

-

-

**Exercise 1.** **Problems with a change**

Which of these issues can you relate to? (Type +1 next to each statement if you relate to it)

* I have fifteen versions of this file - which one do I use?
* I can’t remake this figure from last year
* I slightly modified my code in one place, everything stopped working
* I have several copies of the same directory because I'm worried about breaking something
* Somebody added erroneous records in a shared file with samples, cannot find who and why
* You remember seeing a data file but cannot find it anymore: is it deleted? Moved away?
* I tried multiple analysis and I don't remember which one I chose to generate my output data
* I have to merge changes to our manuscript from 10 different emails with collaborators
* I accidentally deleted a part of my work
* I came to an old project and forgot where I left/saved it
* I reorganized my data files, but broke the analysis pipeline and don't know how to fix it
* I have trouble to find the source of a mistake in an experiment
* My directory is polluted with a lot of unused/temporary/old folders because I'm afraid of losing something important
* I made a lot of changes to my paper but want to bring back one paragraph

DONE:

**Exercise 2.** **Manual Versioning**

Which of these issues does manual versioning help to tackle?

(Type +1 next to each statement if manual versioning helps to tackle)

* I have fifteen versions of this file and I don't know which one to use
* I can't remake this figure from last year
* I slightly modified my code in one place, everything stopped working
* I have several copies of the same directory because I'm worried about breaking something
* Somebody added erroneous records in a shared file with samples, I cannot find who and why
* I tried multiple analysis and I don't remember which one I chose to generate my output data
* I have to merge changes to our manuscript from 10 different emails from collaborators
* I made a lot of changes to my paper but want to bring back one paragraph

DONE:

-

**Exercise 3.** **Changelog in action**

Have a look at one of the example Github repositories and how they track changes\*:

* [data from E.R. Ballou et al. 2020](https://github.com/ewallace/pseudonuclease_evolution_2020/commits/master)  
  https://github.com/ewallace/pseudonuclease\_evolution\_2020/commits/master
* [data from I. Boehm et al. 2020](https://github.com/BioRDM/nmj-pig/commits/main)  
  <https://github.com/BioRDM/nmj-pig/commits/main>

Give examples of:

* what makes them a good changelog
* what could be improved

Think which are the most difficult features to replicate with manual version control?

-------------------------------------------------

Five reasons to use a version control system in research, give +1 to the two most important for you

* Tell the story: The history of your commit messages will describe your project progress.
* Travel back in time: a version control system makes it easy to compare different time points of your project smoothly. If you want to compare the stage of your project a year ago from now, it only takes one command-line of code.
* Experiment with changes: if you want to make changes in a script you can first make a “snapshot” of the project status before experimenting with changes.
* Backup your work: by being able to link your local repository (folder) to a distant online host (GitHub), a version control system backs up your precious work instantly
* Collaborate easily on projects: having a web-hosted synchronised version of your project will encourage collaboration with other researchers. Think about a colleague of yours being able to add a script to make a figure for your publication for instance. The repository will track their contribution and evidence their participation in the project

DONE

**Exercise 4.** **Semantic versioning quiz**

1. Which of the library version is the latest?

* a) 0.12.4
* b) 1.1.0
* c) 1.12.3-alpha
* d) 1.12.2

1. You use library fair.2.3.1 to automatically upload your data after analysis. You found out that there is available new version of the library fair.4.3.1. Which situation is the most probable?

* a) fair.4.3.1 uploads data twice as fast as fair.2.3.1
* b) you need to reserve time to test the new version and adapt your analysis pipeline to work with it
* c) both versions offer the same features and bug fixes as they have the same minor/path version 3.1

1. You developed code that helps to plot detrended and normalized data, the last release has version 1.2.1. You added Z-score to the available normalization methods and fixed the spelling errors in the user messages. Which version number should you give to the new release:

* a) 2.0.0
* b) 1.3.0
* c) 1.3.2
* d) 1.2.2
* e) 2.2.1

1. Your group releases data about studied cancer patients following semantic versioning. The last data sets are released as 1.11, you added data from a new patient,  
   what version should you use:

* a) 1.12
* b) 2.11
* c) 2.0

1. You were asked to better anonymize the above cancer data (v1.11) in future releases. Instead of data and time of patient visit you release only the year, the hospital names have been encoded and patient age has been obscured by random noise.  
   What version should you use:

* a) 1.12
* b) 2.11
* c) 2.0

DONE:

**Feedback**

**Feedback**

1.      How do you feel about the presented topics after this session (type

+1 next to the statement that best describes your feeling):

•       I am more confused:

•       I have a better understanding of them now:

•       My knowledge has not changed much:

2.      Thinking of your knowledge of the lesson topic and its presentation,

which one of the statements best characterize your experience (type +1

next to the statement)

•       I am a novice, and I found the course useful/informative:

•       I am a novice, but I think the course should be improved:

•       I have experience in the presented area, but I found the course

useful/informative:

•       I have experience in the presented area, and I think the course could

be improved:

3. How was the pace of the lesson:

•       Too fast:

•       About right:

•       Too slow:

4. If the lesson had to be 5 minutes shorter, what would you remove:

-

-

-

5. If the lesson could be 5 minutes longer, what would you add or spend

more time on:

-

-

-