Files organisation Episode Notebook

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

Type your name and institution:

-

-

-

**Naming and sorting**

Have a look at the example files from a project, similar to the one from the previous metadata episode. All the files have been sorted by name and demonstrate consequences of different naming strategies.

For your information, to encode experimental details the following conventions were taken

* phyB/phyA are sample genotypes
* sXX is the sample number
* LD/SD are different light conditions (long or short day)
* on/off are different media (on sucrose, off sucrose)
* measurement date
* other details are timepoint and raw or normalized data
* 2020-07-14\_s12\_phyB\_on\_SD\_t04.raw.xlsx
* 2020-07-14\_s1\_phyA\_on\_LD\_t05.raw.xlsx
* 2020-07-14\_s2\_phyB\_on\_SD\_t11.raw.xlsx
* 2020-08-12\_s03\_phyA\_on\_LD\_t03.raw.xlsx
* 2020-08-12\_s12\_phyB\_on\_LD\_t01.raw.xlsx
* 2020-08-13\_s01\_phyB\_on\_SD\_t02.raw.xlsx
* 2020-7-12\_s2\_phyB\_on\_SD\_t01.raw.xlsx
* AUG-13\_phyB\_on\_LD\_s1\_t11.raw.xlsx
* JUL-31\_phyB\_on\_LD\_s1\_t03.raw.xlsx
* LD\_phyA\_off\_t04\_2020-08-12.norm.xlsx
* LD\_phyA\_on\_t04\_2020-07-14.norm.xlsx
* LD\_phyB\_off\_t04\_2020-08-12.norm.xlsx
* LD\_phyB\_on\_t04\_2020-07-14.norm.xlsx
* SD\_phyB\_off\_t04\_2020-08-13.norm.xlsx
* SD\_phyB\_on\_t04\_2020-07-12.norm.xlsx
* SD\_phya\_off\_t04\_2020-08-13.norm.xlsx
* SD\_phya\_ons\_t04\_2020-07-12.norm.xlsx
* ld\_phyA\_ons\_t04\_2020-08-12.norm.xlsx

Questions:

* What are the problems with having the date first?
* How do different date formats behave once sorted?
* Can you tell the importance of a leading 0 (zeros)?
* Is it equally easy to find all data from LD conditions as ON media?
* Can you spot the problem when using different cases (upper/lower)?
* Do you see benefits of keeping consistent lengths of the naming conventions?
* Do you see what happens when you mix conventions?

**Folders vs Files**

Have a look as these two different organization strategies:

(1) |-- Project

|-- |-- arab\_LD\_phyA\_off\_t04\_2020-08-12.metab.xlsx

(2) |-- Project

|-- |-- arabidopsis

|-- |-- |-- long\_day

|-- |-- |-- |-- phyA

|-- |-- |-- |-- |-- off\_sucrose\_2020-08-12

|-- |-- |-- |-- |-- |-- t04.metab.xlsx

Can you think of scenarios in which one is better suited than other? Hint: think of other files that could be present as well.

**Organization for computing**

Take a look at the folder structure recommended by the Good enough practices in scientific computing paper.

Why do you think this layout is recommended and suited for a computing project?

.

|-- CITATION

|-- README

|-- LICENSE

|-- requirements.txt

|

|-- data

| |-- birds\_count\_table.csv

|

|-- doc

| |-- notebook.md

| |-- manuscript.md

| |-- changelog.txt

|

|-- results

| |-- summarized\_results.csv

|

|-- src

| |-- sightings\_analysis.py

| |-- runall.py

|

**Version control Quiz**

**Still need Qs here**

**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:

-

-

-