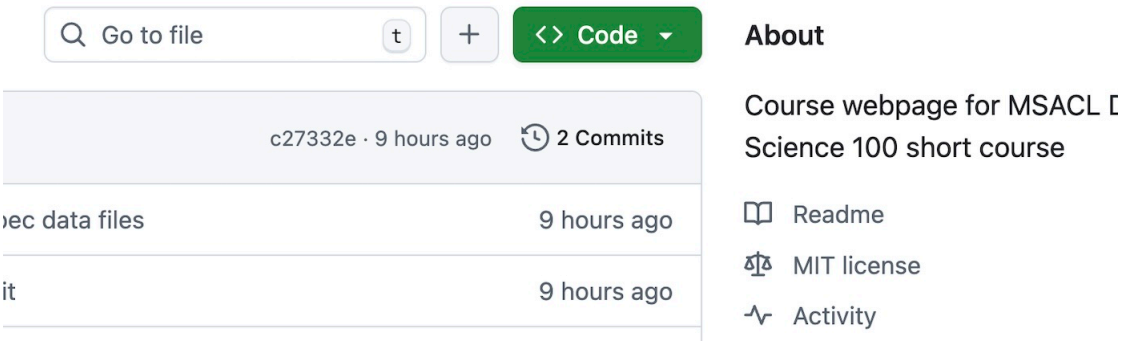
Lesson 1 Exercises

Exercises Setup

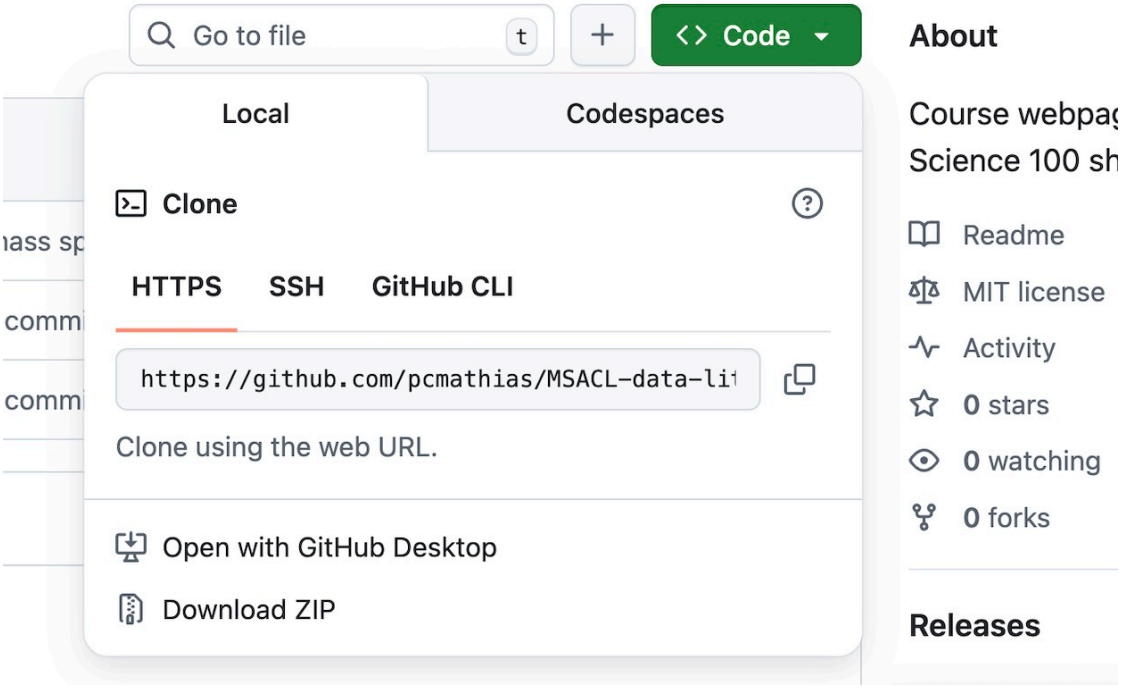
The materials for this course are hosted on a public Github website: https://github.com/pcmathias/MSACL-data-literacy. Github is a website that hosts repositories (repos), which are collections of files that are often organized into a project. You can think of each Github repository as a folder with files and subfolders. Github is a popular way to share code and projects with a broader community, who can contribute back to the repository. It also utilizes the git version control system, which allows users to track specific changes throughout the life cycle of a project.

To make accessing the course materials as easy as possible, we recommend downloading the repository for this course to your computer by following these steps.

1. Navigate to the course website in your browser
2. Press the "Code" button near the top right of the page



3. Click "Download ZIP" to download a .zip file that contains the repository



4. Place the .zip file in a location you will remember on your computer and unzip it. The best practice is to create a "Projects" folder in which you create a new folder for each project.

Orientation to the Data Set

Throughout this short course we will refer to a mass spectrometry data set that includes three files. The batch data file includes data that capture information about six analytes that are measured on 20 batches per day over the course of a month. The sample data file includes information about each sample in each batch, which is composed of 52 samples. Each sample has 24 peaks with data on each peak captured in the peak data file.

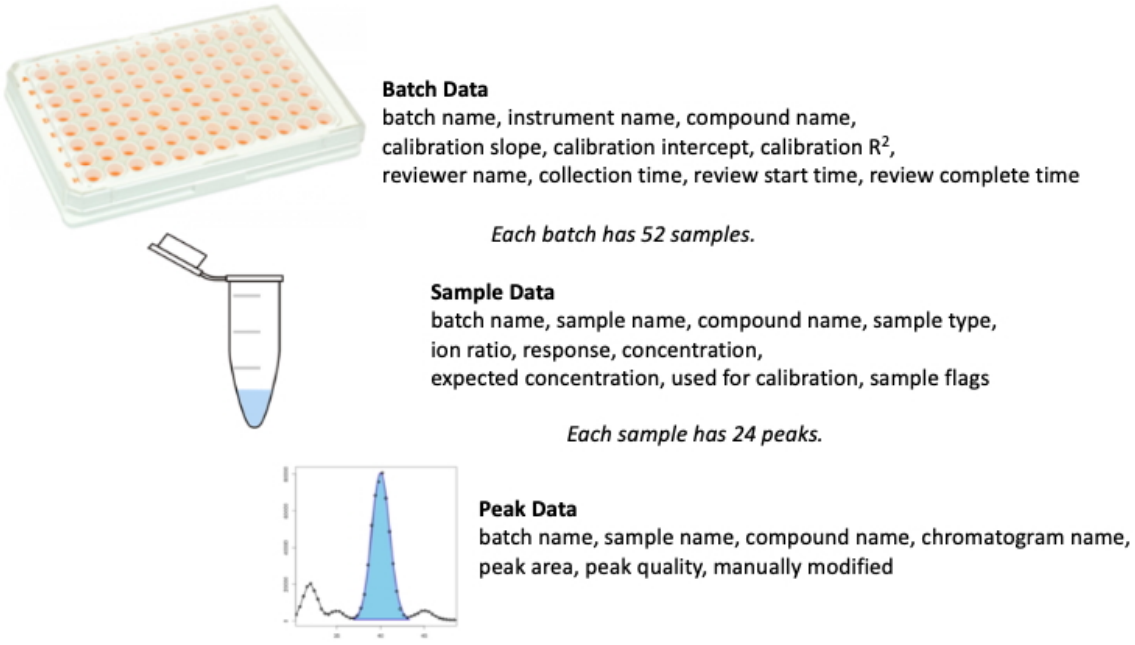


Figure 1: Summary of mass spec data set

Exercise 1: Data Types

We will explore data types in our mass spec data set and how different approaches to interacting with the data set may provide different views or perspectives of the data.

1. In the *exercises* folder within the main project folder, open the **Lesson 1 Exercise 1 Worksheet.xlsx** file.
2. Navigate to the *data* folder within the main project folder. Right click the **batch_data.csv** file and open the file with a text application such as Notepad for Windows or TextEdit for Macs.
3. Right click the **batch_data.csv** file and open with Excel. (Note that there is a similarly named **batch_data_ts.csv** file in the data folder as well. Do not open this yet.) If there is a pop-up message that prompts you to take action, read the message carefully and click "Convert."
4. For each of the column names (separated by commas in the text application), complete the **Lesson 1 Exercise 1 Worksheet.xlsx** based on the data types that you manually infer based on what we just learned about data types from the text file, and based on the format window in Excel.
5. Working the batch_data.csv file, click the File menu at the top left and click "Save As…" The file format to save should be defaulted to "Comma Separated Values (.csv)" - if

it is not, select that and update the file name to "batch_data_excel.csv." Open the **batch_data_excel.csv** file using your text application. How is it different from the original **batch_data.csv** file that you previously opened within the text application?

6. Open the **batch_data_ts.csv** file in Excel. What is different in this file compared to the previous file?