**Session 1: Introduction**

*Philosophy*

I have never taken a course or workshop in machine learning (ML). I've read papers, Stack Overflow threads, python and R documentations on ML implementations and followed other scientists' ML projects on open source Github repositories. I've started by taking code from these sources and modifying it to get it to do what I wanted. As I spent more and more time reading and learning how to implement ML tools to my research, I realized that there is a lot of bad code and bad ML practices out there. My goal became to avoid common mistakes, to make my code reproducible and to create reliable ML models. I also wanted to share what I have learned the hard way with other scientists in the microbiome field so that we could avoid the misconceptions that surround the use of ML in the microbiome field.

This is why I prepared a self-study guide for ML practicioners in the microbiome field. This tutorial will feature a series of lessons with video lectures, real-world case studies and hands-on practice exercises.

*Learning objectives*

By the end of these tutorials, we will know how to:

1. Define a ML problem
2. Decide which ML model to use
3. Construct our dataset
4. Transform our dataset
5. Train a ML model
6. Use the ML model to make predictions
7. Evaluate if our ML solved our problem

*What do you need to do these tutorials*

This tutorial does not presume or require prior knowledge in machine learning. However, to follow the lessons, we recommend learners to meet the following prerequisites.

- Be comfortable with intro-level algebra.
- Proficiency in programming basics with some experience coding in R.
- Downloaded in your computer:
    - R
    - RStudio
    - Text editor (e.g. Atom, Sublime, Notepad++)
    - `Machine_Learning` unzipped folder in your desktop.

*Setup a new R project named Machine_Learning*

1. If you look at Finder window or Window Explorer window, you should have `code`, `data` and `results` directories in `Machine_Learning`. In `data` you should also have `raw` and `process` directories.
2. Now, to make life easier, you should start with `RStudio`. Open RStudio and do "File->New Project->Existing Directory".
3. Once you're in the "Create Project" dialog click on the "Existing Directory" link.
4. Use the "Browse" button to find `Machine_Learning`.
5. My copy of `Machine_Learning` is on the desktop and it lists my "Project working directory" as `~/Desktop/Machine_Learning`. Click "Create Project".
6. In the lower right corner of the RStudio program window you will see that the "Files" tab is selected. In the panel it will have a file called `Machine_Learning.Rproj` and `code`, `data` and `results` directories.
7. Quit RStudio.
8. Use your finder to navigate to your `Machine_Learning` directory.
9. Double click on `Machine_Learning.Rproj`. This is probably the quickest way to have RStudio open up in your desired working directory.

*Working through tutorials*

As you go through the tutorials you should be saving your code as an R script. Save your R scripts in your `Machine_Learning/code` directory.

*Install packages*

We will use several R packages throughout the lessons. Let install the packages; `tidyverse`, `caret`, `pROC`, `LiblineaR`, `kernlab`, `rpart`, `randomForest`, `reshape2`, `ggplot2` and `cowplot`.