DTU Biosustain hands-on data analysis training course 2018

28-29 November 2018, 9-15 both days

Instructors:

Jabus Tyerman, Principal Data Scientist, https://delvbio.com, jabus@delvbio.com,

Steve van Dien, CTO, Persephone Biosciences, San Diego, CA

Course Objective: To provide an introduction and overview of the software package R, and an introduction to statistical data analysis with a focus on working with quantitative biological data (small-scale screening, fermentation data).

Event details: https://www.eventbrite.com/e/dtu-biosustain-hands-on-data-analysis-training-course-tickets-52049217611

"Data are not taken for museum purposes; they are taken as a basis for doing something. If nothing is to be done with the data, then there is no use in collecting any. The ultimate purpose of taking data is to provide a basis for action or a recommendation for action." -W. Edwards Deming

Preparation (before the course)

Download and install R version 3.5.1 from https://www.r-project.org/ or https://mran.microsoft.com/open

Download and install RStudio Desktop (free version > 1.1.463) from https://www.rstudio.com/

Make sure you have permissions set on your computer to install packages; this may involve running R as an administrator (Windows).

Install (and explore!) required R packages: tidyverse, car

It would be useful to have familiarity and access to a text editor for writing and editing R scripts, and data files (e.g., on Windows: Notepad or vi).

It would be useful to have familiarity and access to MS Excel, or other spreadsheet software that can save or export data in a .csv format file format.

If you have time, visit this github repository and download the files (R scripts, data, and handout).

https://github.com/jabus/stat

Data & R-scripts

The data and R scripts for this course can be found on the web here: https://github.com/jabus/stat

A brief tour of R (30 min)

A whirlwind tour of Base R, RStudio, Jupyter notebooks

See Table 1 for a list of common R commands.

R script(s): https://github.com/jabus/stat/blob/master/0_my_first_script_in.R

Small-scale screening data

The data and statistical methods will illustrate how to support the decision about what strains to promote within and across cycles of the Design-Build-Test-Learn process (Figure 1). These methods can be generalized to other research questions (and decisions).

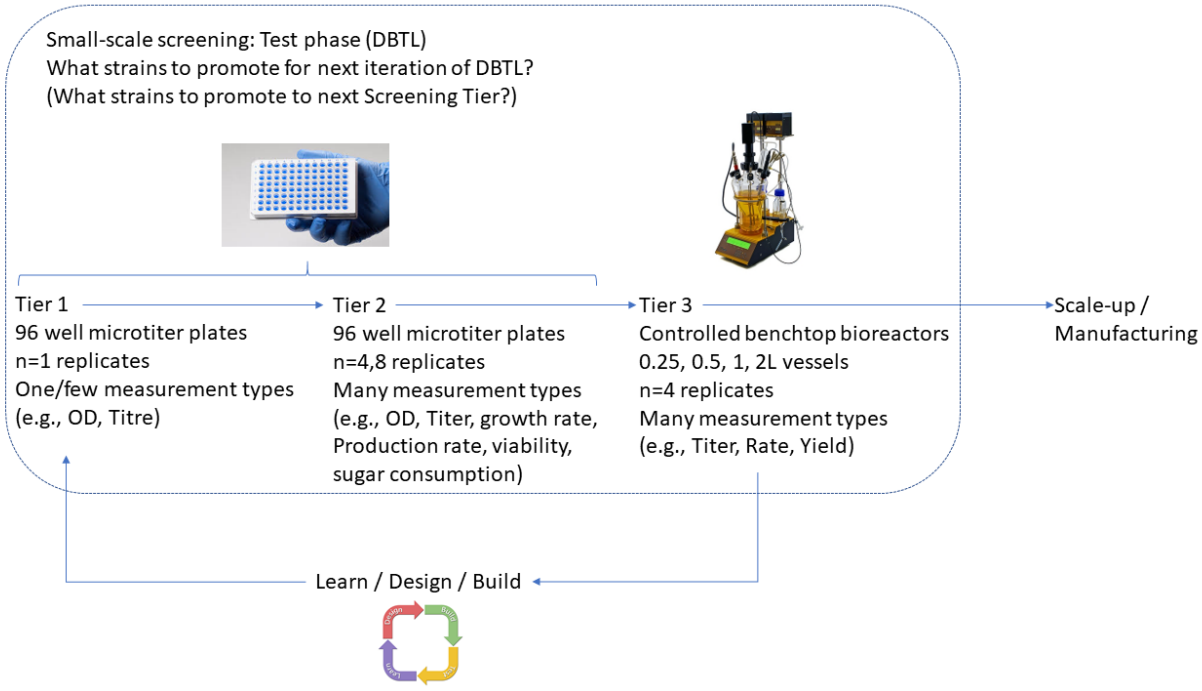

Figure 1. Small-scale screening pipeline in industrial biotech.

Table 1. A brief list of commands (expressions and functions) to try in the R console. Just enough to get you into trouble…

Command	Description
q()	Quit & close R software
quit()	Same
x <- 1	Assign (<-) the value 1 to the object with label x
x -> y	Assign the value in object labeled x to the object labeled y
x	Examine (print) the object x to console
x+y	Do the calculation and print to console
z <- x + y	Assing the value of x + y to z
x:y	expand series x to y
ls()	List objects in memory
install.packages("tidyverse")	Install the R package "tidyverse"
library(tidyverse)	Load the package tidyverse into R environment
↑	[up arrow] cycle through command history (down arrow cycles forward through command history)
?plot	help on <plot>
str(x)	Structure of object x
df <- as.data.frame(x)	Coerce object x to a dataframe object
is.data.frame(df)	Boolean test if df is a dataframe
df[i,j]	df element by row i, col j
df$X1[i]	df element by column heading X1, row i
write.table(…)	Write dataframe to file
read.csv(…)	Read in .csv data file
plot(…)	Plot object

Loading data in R (15-30 min), and Descriptive statistics, Univariate statistics (15-30 min)

Data preparation & working with data R script:
https://github.com/jabus/stat/blob/master/1_load_and_clean_data_in.R

Regression & Machine learning (45-60 min)

Data, R files: https://raw.githubusercontent.com/jabus/stat/master/Tier3_yields.csv

https://raw.githubusercontent.com/jabus/stat/master/3_Regression_models_in_R.Rmd

Graphics with ggpot2 (30-45 min)

Data, R files: https://github.com/jabus/stat/blob/master/5_graphics.R

https://raw.githubusercontent.com/jabus/stat/master/joined_data.csv