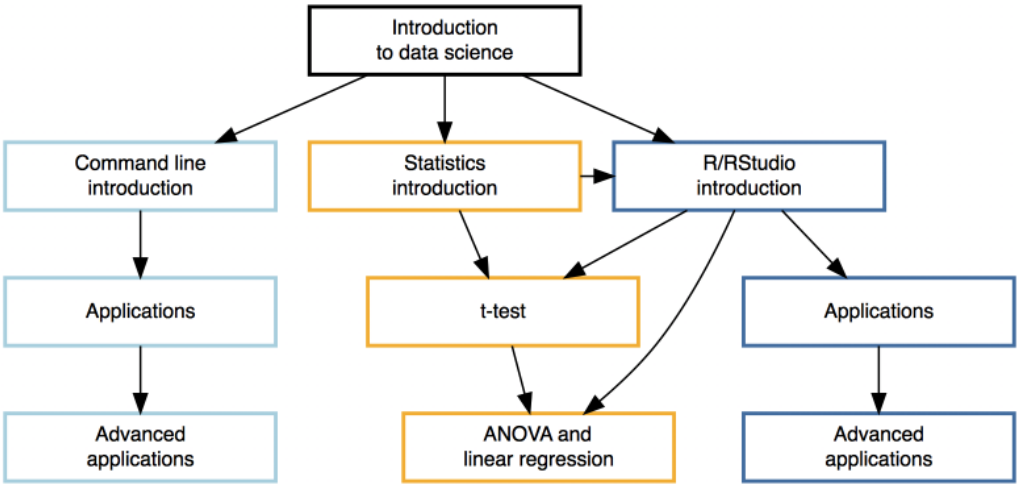
Data science curriculum across M&I

Kim Dill-McFarland, Steven Hallam

version September 22, 2018

Module outline



Introduction to data science

- Align student exceptions in data science and provide motivation by highlighting applications in future courses and careers (MICB 301, 322)

Command line

- **Introduction**: Define and use common Unix functions
- **Applications**: Simple functions like BLAST (MICB 301), remotely accessing a server (MICB 405, 425)
- **Advanced applications**: More complete and/or collaborative functions like microbiome sequence analysis (MICB 405, 425), version control in Git/GitHub (MICB 425)

R/RStudio

- **Introduction**: Download, install, and navigate the program
- **Applications**: Basic plotting and/or statistics of tabular data (MICB 301, 322, 323)
- **Advanced applications**: Data manipulation, visualization, and statistical tests using a variety of data types (MICB 405, 425)

Statistics

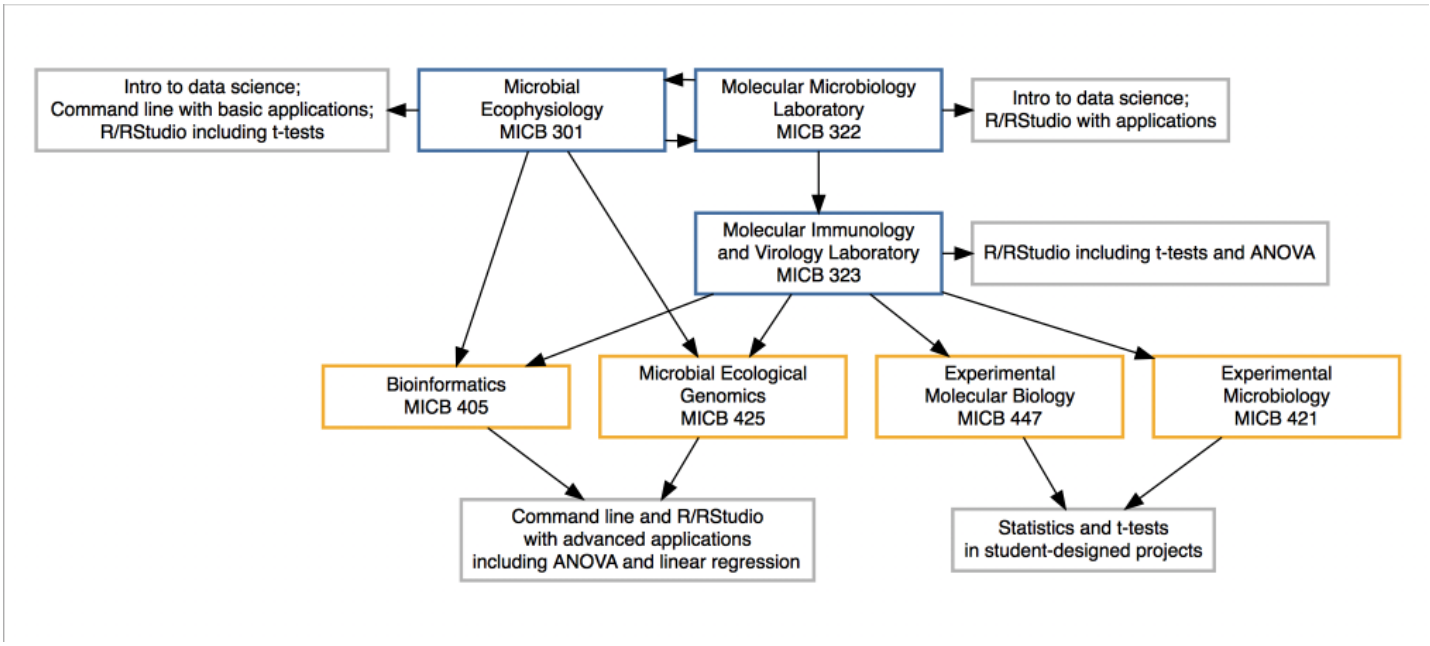
- **Introduction**: Define foundational terms (sample, population, statistic, parameter, p-value, etc)
- **t-test**: Define hypotheses, assumptions, uses, and limitations (MICB 301, 323, 405, 421, 425, 447)
- **ANOVA and linear regression**: Define hypotheses, assumptions, uses, and limitations (MICB 405, 425)
- **Implementation in R/RStudio**: Statistical tests in R (MICB 301, 323, 405, 425)

Custom

- Interested in incorporating data science into your course? Contact EDUCE (info.educe@ubc.ca)!

All module and workshop content is hosted on the EDUCE GitHub (https://github.com/EDUCE-UBC)

Course outline



MICB 301 (5 hrs)

- Introduction to data science in microbiology
- Command line introduction with application to BLAST of a large microbiome data set
- R/RStudio introduction with application to microbiome data including data import and simple visualizations
- Statistics introduction including t-tests and implementation in R/RStudio

MICB 322 & 323 (2.5 hrs)

- Review MICB 301 content (homework)
- Introduction to data science in course-specific areas
- R/RStudio application to student-generated sequencing data from transposon mutagenesis (322) or ELISA (323) including data import, intermediate visualizations, and statistics

MICB 421 & 447 (3 hrs)

- Introduction to data science focused on experimental design
- Statistics introduction including t-tests and limitations of student designed projects

MICB 405 (full integration)

- Command line introduction with application to accessing a remote server
- Advanced application of command line tools for microbiome sequence data (genomic, metagenomic)
- R/RStudio advanced applications in microbiome sequence data manipulation and visualization
- Statistics applications to microbiome data up to linear regression

MICB 425 (full integration)

- Command line introduction with application to accessing a remote server (Amazon Web Services) and version control (Git)
- Advanced application of command line tools for microbiome sequence data (metagenomic, metatranscriptomic)
- R/RStudio advanced applications in microbiome sequence data manipulation and visualization
- Statistics applications to microbiome data up to linear regression