# Course outline

Biochemistry 9545 Macromolecular Informatics. The course will typically meet twice per week Mondays and Wednesdays at 9:30-11:30. There will be interruptions in the course, and additional sessions will be scheduled to make this time up.

# Instructor:

Greg Gloor, [ggloor@uwo.ca](mailto:ggloor@uwo.ca), 519-661-3526 (office)

#### NOTE: This is intended to be focused largely on theoretical aspects of analysis, and practical information on how to manage and keep track of your work

# Requirements

Students will be expected to bring a functional laptop to class with the R programming environment already loaded and functional. Students should be familiar with the file system of their computer and be able to locate files associated with R on their own. Students should have a working plain text editor installed. For Macs this could include Textwrangler, for PCs this could include Notepad++. Atom is also a good choice. Students should have a copy of *"Analyzing compositional data with R"* (<https://link.springer.com/book/10.1007%2F978-3-642-36809-7>) loaded on their computer for reference. We will use many of the principles outlined in that book. In addition, I encourage students to get a copy of *"R in Action"* --- you will find this invaluable. There is a companion web site <http://www.statmethods.net>.

I expect that students will learn outside of class, by doing the readings and doing at least some of their own troubleshooting. We will not be installing much, but what we do install will help you to keep track and present your work in the future.

## Sessions

1. An introduction to high throughput sequencing and data types generated in high throughput sequencing. An introduction to R, markdown and knitr (functional note taking and reproducible data analysis)
2. A further introduction to R, data types, installing packages, simple plots, exploring multivariate data
3. An introduction to Bayesian thinking. Monte Hall problem, Regression to the mean, Gambler's fallacy
4. An introduction to the structure of multivariate datasets. An introduction to compositional data.
5. An introduction to the error structure in high throughput sequencing. PCA and compositional biplots.
6. Correlations in compositional data

## Assessments

1. The student will be required to submit a proper R program. Marks will be assessed based on ability to properly comment, ability to make a functional script, conciseness and proper form. This must be submitted in the form of a .Rmd document without external dependencies. There will be an oral component to the grading for this assignment
2. The student will be required to make and properly interpret a compositional biplot from a dataset provided. The student will be required to solve and explain some simple problems based on probabilistic thinking. There may be an oral component to the grading for this assignment
3. The student will be required to examine and interpret the correlation and pairwise abundance of parts of a high-throughput sequencing dataset

The course is largely theoretical. Assessments are to be worked on and handed in individually, and are equally weighted. Students will be expected to conduct independent learning to practice the concepts and tools. Here is where you will find the R in Action book or the companion website invaluable.

# Sources

This should be all you need, aside from some R packages

* R installation: <http://cran.utstat.utoronto.ca>
* Markdown: <http://daringfireball.net/projects/markdown/>
* R studio and markdown: <http://rmarkdown.rstudio.com>

# Readings:

* how does multiple testing correction work? Nat. Biotech. 2009 27:1135
* how to make more published research true. PLoS Medicine. 2014 11:e1001747
* it's the effect size stupid. <http://www.leeds.ac.uk/educol/documents/00002182.htm>
* the fickle P value generates irreproducible results. Nat. Meth. 2015. 12:179
* It's all relative: analyzing microbiome data as compositions, 2016, 26:322
* Proportionality: a valid alternative to correlation for relative data, PLoS Comp. Bio. (2015)11:e1004075

# Due dates

All assignments are due one week after they are given. All three assignments must be handed in and passed to obtain course credit.

# Statement on Academic Offences

The statement: Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following Web site: <http://www.uwo.ca/univsec/handbook/appeals/scholastic_discipline_grad.pdf> Academic Handbook, Exam, Course Outlines Page 4 Issued: 2011 02

Additionally, "All required papers may be subject to submission for textual similarity review to the commercial plagiarism-detection software under license to the University for the detection of plagiarism. All papers submitted for such checking will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com (<http://www.turnitin.com>)."