

# BIOL-4606 / COMP-4550 / BIOL-7945 – Advanced Bioinformatics

## Course Outline - Winter 2015

**Instructor:** Dr. Lourdes Peña-Castillo

**Office:** ER-3014

**Office Hours:** Monday/Tuesday 1:00 – 2:00 pm

**Phone:** 864-6769

**Webpage:** [www.cs.mun.ca/~lourdes](http://www.cs.mun.ca/~lourdes)

**Email:** All emails to the instructor should be sent through the [Desire2Learn \(D2L\)](#) shell. D2L email will be replied within two working days.

Course-related questions may also be posted on the discussion forum in D2L.

**Lectures:** Slot 16 – T R 12:00 – 12:50 pm and F 1:00 – 1:50 pm (Location SN-2064)

**Labs:** Slot 61 - M 2:00 – 5:00 pm (Location SN-4117)

---

### Calendar Description:

This course provides students with the basis to analyze a variety of biological data within an integrated programming environment for data manipulation, calculation and graphical display. Students who succeed in this course should be able to extract meaningful information from data generated by high-throughput experimentation. The course will introduce one such integrated programming environment and will explore the computational and statistical foundations of the most commonly used biological data analysis procedures.

### Course Description:

The aim of the course is to introduce the students to techniques and methods for processing, visualizing, analyzing, learning classification models from, large-scale data sets within an integrated programming environment (R).

### Educational Goals:

1. Learn R programming to an intermediate level
2. Become familiar with Bioconductor
3. Be able to perform exploratory data analysis on large-scale data sets
4. Become familiar with analysis of real data from a variety of case studies

### Prerequisites:

No previous programming knowledge or experience in R is assumed.

PR: BIOL-3951 or COMP-3550, and STAT-2550 or STAT-2510 (or equivalent), or permission of the course instructor.

### Course Activities:

1. Nine labs & quizzes
2. Two midterm exams, and one final exam
3. In class participation
4. **BIOL7945 only:** a term project

## BIOL-4606 / COMP-4550 / BIOL-7945 – Advanced Bioinformatics Course Outline - Winter 2015

### Evaluation Scheme:

	BIOL-4606 / COMP-4550	BIOL-7945
Lab work and quizzes	30%	25%
Midterm exams (1 best out of 2)	30%	25%
Final exam	35%	25%
In-class participation	5%	5%
Term Project		20%

### Semester Plan:

Week	Monday	Tuesday	Wednesday	Thursday	Friday
<b>1</b>	<b>January 5</b> Lectures Begin No lab	6 <i>Introduction to R</i>		8 <i>Introduction to R</i>	9 <i>Vectors</i>
<b>2</b>	12 No lab	13 <i>Vectors</i>		15 <i>Calling functions</i>	16 <i>Matrices</i>
<b>3</b>	19 <i>Lab 1</i>	20 <i>Matrices</i>		22 <i>Lists</i>	23 <i>Lists</i>
<b>4</b>	26 <i>Lab 2</i>	27 <i>Data Frames</i>		29 <i>Factors &amp; Tables</i>	30 <i>Factors &amp; Tables</i>
<b>5</b>	<b>February 2</b> <i>Lab 3</i>	3 <b>1<sup>st</sup> Midterm</b>		5 <i>Review</i>	6 <i>Reading / Writing data</i>
<b>6</b>	9 <i>Lab 4</i>	10 <i>String manipulation</i>		12 <i>String manipulation</i>	13 <i>Conditionals / loops</i>
<b>7</b>	16 <b>Winter Break Begins</b>	17	18 <b>Winter Break Ends</b>	19 <i>Vectorization</i>	20 <i>Vectorization</i>
<b>8</b>	23 <b>Last day to Drop</b> <i>Lab 5</i>	24 <i>Writing functions</i>		26 <i>Writing functions</i>	27 <i>Plotting</i>
<b>9</b>	<b>March 2</b> <i>Lab 6</i>	3 <i>Plotting</i>		5 <i>Classification</i>	6 <i>Classification</i>
<b>10</b>	9 <i>Lab 7</i>	10 <i>Classification</i>		12 <i>Statistical testing</i>	13 <i>Statistical testing</i>
<b>11</b>	16 <i>Lab 8</i>	17 <b>2<sup>nd</sup> Midterm</b>		19 <i>Review</i>	20 <i>Intro. to Bioconductor</i>
<b>12</b>	23 <i>Lab 9</i>	24 <i>Expression data</i>		26 <i>Expression data</i>	27 <i>Expression data</i>
<b>13</b>	30 No lab	31 <i>Graphs</i>	<b>April 1</b>	2 <b>Lectures End</b> Q&A	3 <b>Good Friday. No classes</b>
<b>14</b>	6	7	8 <b>Final exams start</b>	9	10
<b>15</b>	13	14	15	16	17 <b>Final exams end</b>

## **BIOL-4606 / COMP-4550 / BIOL-7945 – Advanced Bioinformatics Course Outline - Winter 2015**

### **Reference Texts:**

- The Art of R Programming. Norman Matloff. No Starch Press. 2011.
- R for everyone. Jared P. Lander. Addison Wesley. 2014.
- Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Robert Gentleman, Vincent J. Carey, Wolfgang Huber, Rafael A. Irizarry, and Sandrine Dudoit, editors. Springer. 2005.
- Bioconductor Case Studies. Florian Hahne, Wolfgang Huber, Robert Gentleman, Seth Falcon. Springer. 2008.
- R programming for Bioinformatics. Robert Gentleman. CRC Press. 2008.
- Statistical Bioinformatics with R. Sunil K. Mathur. Academic Press. 2010.
- Introductory Statistics with R. P Dalgaard. Springer, 2nd edition. 2008.
- Using R for Introductory Statistics. John Verzani. 2002.
- The Stat 390 R Primer. Christopher G. Green. 2004.
- Statistics and Data Analysis for Microarrays Using R and Bioconductor. Sorin Drăghici. CRC Press, 2<sup>nd</sup> edition. 2012.
- A Beginner's Guide to R. Alain F. Zuur, Elena N. Ieno, Erik H.W.G. Meesters. Springer. 2009.

### **Notes:**

- In the event of university closure on the day of a test, the test will be given in the next class meeting.
- All grades will be assigned according to the University Calendar ([Section 6.8.2 under University Regulations](#)).
- All assignments and lab submissions are to be submitted electronically through the D2L environment (no hard copy submission is required).
- All assignments and written reports are due prior to the beginning of the class on the specified dates. Late submissions will be penalized. 20% will be deducted to submissions received within 24 hours of the deadline, and an additional 10% will be deducted each extra 24 hours after the first 24 hours past the deadline.
- All written materials delivered must comply with the expectations set out in the University Calendar ([Section 6.8.3 under University Regulations](#))
- If, for special circumstances (such as medical or bereavement), you miss a deadline for a grade item, notify your professor as soon as possible, providing any necessary related documentation (if documentation is required). Failure to do this might result in a mark of 0% for that grade item. For more information regarding medical notes and the information to be in them, please see [Section 6.14.6 under University Regulations](#) in the University Calendar or consult the Registrar's Office. A bereavement note should include the officiating parties name, the date of the service and the relationship to the student.
- Course materials, news and communications will be available from the Desire2Learn system.
- Although changes to the course outline are not intended at this time, any part of this document can be subject to change, particularly during the first two weeks of classes. In this case, students will be notified in class.
- Each lab period will start with a quiz on material covered in the previous lab. Lab quiz has to be taken in the lab room. The quiz is the only means of obtaining credit for work done during the lab period.
- In-class participation will be credited by using TopHat.
- From section 6.11 of the University Calendar: “A student is expected to adhere to those principles which constitute proper academic conduct. Academic misconduct cannot be condoned or even appear to be condoned. A student has the responsibility to know which actions, as described under [Academic Offences](#), could be construed as dishonest or improper.”. Note that when two or more students submit identical or nearly identical work claiming it is their own, it is a clear sign of improper academic conduct.