**CX 4240 Introduction to Computing for Data Analysis**

Instructor: Guy Lebanon, School of Computational Science and Engineering

**Overview**

The course will introduce students to the computational techniques needed for data analysis. Topics include programming in C, R, and Python, accessing SQL databases, multidimensional arrays, basic numerical computing, and graphing data. The course will emphasize practical techniques and hands-on applications and case studies. It is primarily aimed at undergraduate students with some programming exposure.

**Prerequisites**

* CS 1371 – Computing for Engineers or a different programming course

**Grade Composition**

30% homework, 30% midterm exam, 40% final exam

**Topics**

1. **Introduction to C (3 week)**

Variables, pointers and arrays, control flow, functions, compiling and linking multiple files

1. **Introduction to Python (3 week)**

The Python shell, strong vs. weak typing, string processing, multidimensional arrays, numeric libraries, interfacing the World Wide Web

1. **Introduction to R (3 weeks)**

The R shell, differences between R, C and Python, lists and data-frames, R routines for data analysis, R packages, interfacing C code from R, vectorizing code, data manipulation including the reshape package and plyr’s split-apply-combine.

1. **SQL Databases (2 week)**

Introduction to databases, SQL syntax, accessing databases from R and Python

1. **Numerical Computing (1 week)**

Floating point and fixed point representations, underflow and overflow, stable vs. non-stable error propagation

1. **Graphing Data (3 weeks)**

Graphing data with R, strip plots, histograms, scatter plots, kernel smoothing, under-smoothing vs. over-smoothing, facets, contour plots, the quantile function, box plots and qq-plots. Case studies involving real world datasets.