## STA242 Introduction to Statistical Programming (Graduate)

The intent of the class is to ensure that graduate students have (i) a vocabulary for the basic computational tasks that they will encounter in their research and career as well as other courses, and (ii) knowledge and experience to approach future computational tasks intelligently.

Students will initially work on small programming exercises, for example, doing simulations or writing functions for generating random numbers (e.g. acceptance/ rejection sampling). We will proceed to deal with reading data into *R* from standard and various more complex formats (e.g. tables, CSV, text, ragged records) and then exploring it with graphical and numerical summaries and exploring potential models. We will cover how to report the results and also deal with elementary aspects of computationally intensive tasks. The students will learn how to use the computing environment for its existing facilities and also to create new functionality for non-standard tasks.

The computational topics in this course will be presented in the context of analyses of real scientific/social studies. We will use a variety of statistical topics as examples and exercises, including linear models and "modern" techniques such as machine learning methods (e.g. CART, SVMs) and Bayesian computations (e.g. MCMC).

We will focus on the *R* language and cover additional tools (e.g. SAS) in less detail.

Week 1.	Overview of Statistical Computing
	Survey of Programming Languages
	Choosing the Best Language for a Task
Week 2.	Introduction to R
Week 3.	Programming in <i>R</i>
	Formula Language/Modeling
Week 4.	Data Visualization & Interactive Graphics
Week 5.	Advanced Data Input & Output
	Working with Data from Databases
Week 6.	Writing <i>R</i> Functions
Week 7.	Efficiency and Debugging
Week 8.	R Packages
Week 9.	The UNIX Shell
Week 10.	Interfacing to Other Languages