# BST430: Introduction to Statistical Computing
# Fall 2017

Room: SRB 1.402
Time: MW 11:00am – 12:30pm
Lead Instructor: Matthew N. McCall, PhD
Other Instructors: Christine A. Brower, MPH; Shirley Eberly, MS; Shan Gao, MS; Donald Harrington, MS
Phone: (585) 273-3177
Email:   matthew_mccall@urmc.rochester.edu
          Christine.Brower@chet.rochester.edu
          shirley_eberly@urmc.rochester.edu
          shan_gao@urmc.rochester.edu
          donald_harrington@urmc.rochester.edu
Office Hours: Mon 5-6pm
Course Website: learn.rochester.edu
Prerequisites: An advanced undergraduate course in Statistical Inference or permission from the instructor.

## Basic Information
*Classes will be held twice weekly in 90-minute sessions. Office hours will be held weekly at a time and location to be determined jointly by the students and instructors. Additional assistance can be arranged via email or phone.*

## Course Description
*The purpose of this course is to provide a strong foundation in the computational skills needed for graduate coursework and research in Statistics. The topics covered in this course include an introduction to statistical programming in R/RStudio and SAS, reading and parsing data from a variety of sources, handling of very large data, data visualization, reproducible research, version control, collaborative coding, R function and package development, uses of and approaches to simulating data, and command line tools. Topics in statistical data analysis and interpretation provide working examples.*

*The course will be taught via lectures and interactive sessions. The emphasis of the course will be on mastering the computational skills and techniques upon which subsequent coursework and research will build.*

## Course Aims and Objectives
*Students will gain an in-depth knowledge of modern statistical computing. Upon completion of this course students will be able to perform and clearly describe statistical analyses in R and SAS that include:*
*1. Loading data directly from the primary source.*
*2. Performing statistical analyses.*
*3. Creating clear visualizations that accurately convey the underlying data / results.*
*4. Generating a reproducible report that describes the analysis, results, and conclusions.*
*Additionally, students will be able to use version control and collaborative coding tools to track and share their work.*

**Course Policies and Expectations**

*Students will be responsible for completing weekly assignments and respectfully participating in classroom discussions. One lecture each week will typically be devoted to interactive problem solving. Attendance in the course is mandatory; students missing more than 2 lectures without pre-approval or documented illness will receive a reduced grade proportional to the number of absences.*

*Students are encouraged to work together on assignments; however, sharing of text (either code or interpretation) is not permitted. The text of all submitted assignments should be unique to each student. Assignments will not be accepted after the due date without prior approval from the instructor.*

**Materials and Access**

swirl: http://swirlstats.com
R Package Development: http://r-pkgs.had.co.nz
Writing R extensions: https://cran.r-project.org/doc/manuals/R-exts.html
BioC package guidelines: http://bioconductor.org/developers/package-guidelines
RStudio: http://rstudio.org
R Markdown: http://rmarkdown.rstudio.com
knitr: http://yihui.name/knitr
SAS Version 9 help: http://support.sas.com/documentation/index.html
SAS Graph: https://support.sas.com/documentation/cdl/en/graphref/65389/HTML/default/viewer.htm
SAS procedures:
http://support.sas.com/documentation/cdl/en/allprodsproc/63883/HTML/default/viewer.htm

Peng, Roger D. **Exploratory Data Analysis with R** https://leanpub.com/exdata
Peng, Roger D. **R Programming for Data Science** https://leanpub.com/rprogramming
Delwiche, Lora D., and Slaughter, Susan J. **The Little SAS Book: A Primer, Fourth or Fifth Edition.**

**Assignments and Grading Procedures**

*Attendance and participation: 10%*
*Weekly assignments: 90%*

**Academic Integrity**

Academic integrity is a core value of the University of Rochester. Students who violate the University of Rochester University Policy on Academic Honesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since academic dishonesty harms the individual, other students, and the integrity of the University, policies on academic dishonesty are strictly enforced. For further information on the University of Rochester Policy on Academic Honesty, please visit the following website:
*http://www.rochester.edu/college/honesty/docs/Academic_Honesty.pdf*

**Accommodations for Students with Disabilities**

Students needing academic adjustments or accommodations because of a documented disability must contact the Disability Resource Coordinator for the school in which they are enrolled:
http://www.rochester.edu/eoc/DisabilityCoordinators.html

**Course Schedule**

| Date | Topics | Instructor | Assignments |
|------|--------|------------|-------------|
| 8/30 | Overview of R and Reproducible Research Introduction to RStudio, knitr, and Rmarkdown | McCall | PS 1 Assigned |
| 9/4 | LABOR DAY – NO CLASS | | |
| 9/6 | Reading and formatting data in R | McCall | |
| 9/11 | Interactive problem solving | McCall | |
| 9/13 | Summary statistics, two group comparisons, stratification, apply functions | McCall | PS 1 Due PS 2 Assigned |
| 9/18 | Interactive problem solving | McCall | |
| 9/20 | Exploratory data analysis and visualization in R Comparison of base graphics and ggplot2 | McCall | PS 2 Due PS 3 Assigned |
| 9/25 | Interactive problem solving | McCall | |
| 9/27 | R functions and package development | McCall | PS 3 Due PS 4 Assigned |
| 10/2 | Interactive problem solving | McCall | |
| 10/4 | Comparison of base R and the tidyverse | McCall | PS 4 Due PS 5 Assigned |
| 10/9 | Interactive problem solving | McCall | |
| 10/11 | Version control and collaborative coding Introduction to git and github | McCall | PS 5 Due PS 6 Assigned |
| 10/16 | Interactive problem solving | McCall | |
| 10/18 | Distributions, simulations, and sampling | McCall | PS 6 Due PS 7 Assigned |
| 10/23 | Interactive problem solving | McCall | |
| 10/25 | Principles of a complete data analysis | McCall | PS 7 Due PS 8 Assigned |
| 10/30 | Interactive problem solving | McCall | |
| 11/1 | Coding style, table-driven analysis, and programming on the language | Harrington | PS 9 Assigned |
| 11/6 | Interactive problem solving | Harrington | |
| 11/8 | Introduction to SAS, variables, input statements, and informats | Brower | PS 9 Due PS 10 Assigned |
| 11/13 | Mixed input, datasets, dates and formats | Brower | |
| 11/15 | Sorting and combining datasets | Brower | PS 10 Due PS 11 Assigned |
| 11/20 | Frequencies and Means | Eberly | |
| 11/22 | NO CLASS | | |
| 11/27 | Working with SAS datasets | Gao | PS 11 Due PS 12 Assigned |
| 11/29 | Summary statistics and exploratory plots | Gao | |
| 12/4 | Data distributions, macros, and arrays | Gao | PS 12 Due PS 13 Assigned |
| 12/6 | Regression | Eberly | PS 13 Due |

| | | | PS 14 Assigned |
|-------|------------------------------|--------|----------------|
| 12/11 | Survival Analysis | Eberly | |
| 12/13 | Command line tools & BlueHive | McCall | PS 8 Due<br>PS 14 Due |