Syllabus

Course Title: Spatial Bioinformatics

Lead Instructors: Mary Blair (AMNH) and Rob Harbert (AMNH)

Other Course Instructor(s): [Peter Galante]

Course Description:

Spatial data and models are ubiquitous in modern comparative biology and ecology due to the vast amount of available data and ever developing modeling methods. This course will focus lectures on a series of "best-practices" in handling and modeling spatial biological data including data-mobilization, bias detection and reduction, geographic projection management, and comparative modeling frameworks. Labs will concentrate on demonstration of best-practices on a range of datasets including student's personal data. The course will culminate in student's working in "hackathon" style working groups to develop a spatial data analysis pipeline to address a question of mutual interest to be posted on appropriate code-sharing repositories.

Duration in weeks: 1.5

Number of lectures/seminars: 5. **Duration:** 2 hr.

Number of labs: 5, **Duration:** 3 hr.

Number of Workshops: 2**, Duration:** 3hr.

Hackathon: 1, **Duration:** 8 hr.

Location: RGGS Classroom

Prerequisites: General biology, basic knowledge of evolutionary biology, ecology, and of statistics. Interest in learning basic programming and computational methods. It is not necessary that students have prior experience coding in R, but an R primer will be offered prior to the start of this course for those who need an introduction or review of R basics.

Learning Objectives:
- To introduce students to concepts and best practices related to spatial representation and analysis of biological data.
- To present the different methods and data analyses associated with spatial representation and analyses of biological data.
- To present examples from comparative biology studies and of questions that can be answered by applying these methods.
- To provide students with tools to apply the methods for their own studies.

Bibliography:

Guisan, A., Thuiller, W. and Zimmermann, N.E., 2017. Habitat Suitability and Distribution Models: with Applications in R. Cambridge University Press.

Wegmann, M., Leutner, B. and Dech, S. eds., 2016. *Remote sensing and GIS for ecologists: using open source software*. Pelagic Publishing Ltd.

Grading: RGGS Comparative Biology students will be graded pass/fall; students enrolled in our partner programs will be assigned either a letter grad or pass/fall as required by their registrar.

Evaluations Basis: Students will be evaluated on attendance (20%), participation in class and on class exercises (40%) and on performance on course projects/writing (40%).

Credits: 1

Dates: May 2018 **
Daily schedule: Lecture/Workshop: 10:00-12:00; Break: 12:00-1:30; Working Lab: 1:30-4:30.

Content: (Days 1-9)

Day 1 - Lecture Topic: Introduction to programming. Lab: Unix command line, basic data handling in R.

Day 2 - Lecture Topic: Geography, coordinate systems, projections, georeferencing, GIS data types (Vector, Raster), sampling bias, spatial autocorrelation, and remote sensing; Lab: What's in a GPS coordinate? Comparative analysis of GPS vs. cell-phone locality information.

Day 3 - Lecture Topic: Occurrence data and distribution modeling; Lab: Comparative niche modeling. Covering - background/absence sampling, environmental layers, model evaluation and parameterization, niche differentiation and overlap.

Day 4 - Lecture Topic: Survey of advanced modeling topics (see lab topics); Lab: Advanced modeling exercises. (possibly including: past/future range estimation, biogeographic reconstruction, landscape genetics, landscape analysis, modeling biotic interactions, mechanistic models – may be tailored to class interests).

Day 5 - Lecture Topic: Literature Discussion of spatial data in Comparative Biology - Survey of 2-4 current papers. Lab: Advanced topics in R programming and an introduction to Git for code sharing and management. Will cover scripting, functions, and documentation. A primer in preparation for workshops and hackathon days.

Day 7-8 - BYOD (Bring Your Own Data) Workshops. (1-4PM). Students will bring or find data relevant to their own research (from i.e., Landsat, GBIF, iNaturalist) and will work through practical issues with their data analysis plans. These sessions will also serve as a brainstorming and guidance period for Hackathon project development.

Day 9 - Project/Hackathon day (9AM - 5PM): Students work in small groups a project that will develop a data analysis pipeline to address a question of interest to them. The goal of a hackathon is to create usable software with a specific focus by collaboration between interested parties. In this course the students will develop an R script, an example dataset, and a demonstration. Students will supplement this usable code with a write-up of the supporting literature and target audience for the code. Final code will be shared on the course GitHub to facilitate reuse and further development.

Eligibility: Must be enrolled in a doctoral degree program at RGGS, in the partner programs at Columbia University or CUNY, or in a NYCEP consortium university (with special permission). Graduate or post-graduate students from other U.S. institutions can be authorized to attend to this course by special arrangement and subject to available places, but will not receive graduate credit from RGGS. Students from the RGGS and its partner universities will have priority for registration.

Class size is limited to 10 students.

Students are required to bring their own laptops and will be asked to install a current version of RStudio and QGIS. Instructors will circulate details for obtaining and installing this software after course registration is complete including details on installation specifics for mac versus windows laptops. It is recommended but not required that laptops with a Unix (Mac/Linux) operating system are used for this course.

Statement on Academic Integrity: Each graduate student bears the responsibility to observe traditional canons of scholarly discourse, scientific research, and academic honesty. Plagiarism, cheating, and fraud in research will not be tolerated. Accordingly it is expected that students work individually unless specifically instructed to work in groups. The full Academic Integrity policy is in the student handbook.

Course Evaluations: Each student is required to complete an anonymous course evaluation at the end of the term. The course evaluation is a tool for faculty and administrators to improve the student learning experience.

Contact: mblair1@amnh.org**;** rharbert@amnh.org