R for Beginners: an intensive introduction to exploratory data analysis

Kyle Christie and Marshall McMunn
kchristie@ucdavis.edu; msmcmunn@ucdavis.edu
Population Biology Graduate Group, UC Davis

The Problem
One of the biggest challenges for many graduate students in STEM fields is connecting with their data. Did I input my data accurately? How can I summarize my findings? Regardless of the degree program or scientific emphasis, exploring one’s data is the first step towards analyzing it, yet many graduate students lack the proficiency to efficiently manipulate, summarize, or visualize their data.

Workshop Description
The intensive week-long workshop was designed for students with no previous experience with R or other programming languages. Students gained basic fluency in R through exploration of four major topics:

1) data manipulation and quality control
2) data summary
3) graphing and visualization
4) basic automation of repetitive tasks

Evaluation and Assessment
A hands-on and active learning environment equipped students with the requisite R knowledge and confidence needed to explore their own data and tackle graduate-level statistics courses.

Before the workshop, students generally self-ranked their R proficiency as 1 out of 10, after the 4-day workshop, their R proficiency as 4 out of 10.

Participants were primarily graduate students and postdoctoral scholars; staff and faculty also participated. Participants following a short lecture before beginning an independent programming exercise.

Graduate Student Assistants
Fifteen (15) graduate students generously volunteered as assistant instructors throughout the course. Thank you!

Future Improvements
Feedback from the course was generally very positive; a few consistent suggestions included:

• Assess completion of exercises more effectively, to better pace the class.
• Recruit additional assistant instructors.
• Provide an additional “advanced R” course.