

# Psyc 7765-001

## Fundamentals of Statistical Computing for Behavioral and Social Scientists

### Course Syllabus

**Meeting Time:** Fridays, 2:00-3:15 PM

**Room:** Gilmer 166

**Instructor:** Jeffrey R. Spies

**Office:** Millmont 210

**Office Hours:** TBA or by appointment

**Email:** [jspies@virginia.edu](mailto:jspies@virginia.edu)

**Course Web Page:** <http://people.virginia.edu/~js6ew/#7765>

## 1 Summary

This course will provide students with a basic understanding of statistical computing and programming using the R language. Students will learn methods of integrating the computational skills they acquire into a workflow making the process from analysis to publication more efficient with examples emphasizing areas where R has specific advantages over packages like Excel, SPSS, and SAS. These topics will include data management and cleaning, task automation, and visualizations.

## 2 Background

R is a free, community-developed, open-source programming language and statistical environment. Because of its unparalleled flexibility, it is the software used to demonstrate analyses in every statistics course currently taught in the Department of Psychology at the University of Virginia, including the yearlong introductory statistics sequence required for all graduate students. UVA is one of many academic institutions following a trend in the behavioral and social sciences to adopt R as the primary software for data analyses taught to incoming students as it is widely used both across academia and industry.

It is often the case that the questions researchers seek to ask are limited by the capabilities of the statistical software they use. However, because R is also a programming language, users can both perform common analyses or develop more advanced methods that suit their purposes using the same platform. This is a task becoming increasingly common as psychology progresses as a science and researchers ask more complicated, computationally intensive questions of their data.

In order to offer the flexibility required to not inhibit research, R tends to come with a learning curve for those without programming experience. When a person learns R, even at the most basic level, they are essentially learning how to program because that is the method the user interacts with the system rather than point-and-click menus, like Microsoft Excel. To take full advantage of the flexibility that comes with a programming language, a user must learn not only the software tool itself, but the concepts that accompany programming in general. For many psychologists, this is a daunting task as most undergraduate psychology programs do not include programming or computational courses.

## 3 Focus

The goal of this course is to provide students with a more basic understanding of the computational properties of R as a programming language. While R has many intrinsic benefits that point to a long

future (e.g. free, community-developed, open-source, easily extendable), computational techniques will be taught in such a way that generalize to other programming languages. The emphasis will not be on the syntax used to accomplish particular analyses or the statistics involved in those analyses, but instead on the concepts and functionality that the language allows one to specify. Examples making use of these computational techniques will be relevant to the behavioral and social scientist and provide them with a base of knowledge that could be applied to their own research. These examples will include data cleaning and management, automating analyses, visualization generation, and integrating R into a workflow making the research process more efficient.

## 4 Prerequisites

It is recommended that students have taken or are currently enrolled in an introductory statistics course. No programming experience is required.

Students should bring laptops to each class in order to follow along with the in-class presentation and take part in exercises. R should be installed on their system. Information for downloading and installing R can be found at: <http://people.virginia.edu/~js6ew/#7765>.

## 5 Student Evaluation

Evaluation and course grades will be based on attendance and completion of weekly assignments (3-4 hours per week outside of class); both are required. Regarding attendance, only University-accepted excuses for being absent will be allowed, and the instructor should be notified before being absent or, in the event of an emergency, as soon as possible. When allowed, students can work with the instructor to make up missed classes.

## 6 Textbooks and Handouts

TBA

## 7 Timeline

	Date	Description	Assignment
	08/23	SEMESTER BEGINS	
1	08/26	Intro to programming: Data types (numeric, string, logical), primitive expressions, functions, documentation	TBA
2	09/02	More on functions and documentation Intro to collection data types (vectors, lists) Basic importing of data into R R workflow style and tips	TBA
3	09/09	Data relevant data types (1): data frame, factors, representing missingness Example: linear regression	TBA
4	09/16	Data relevant data types (2): tables	TBA
5	09/23	Logical expressions and selection vectors(1)	TBA
6	09/30	Selection vectors (2)Example: dataset	TBA
7	10/07	Writing functions - advanced	TBA
8	10/14	Apply functions and intro to Plyr	TBA
9	10/21	Apply functions Example: data aggregation	TBA
0	10/28	Loops and their relation to apply functions	TBA
1	11/04	Nested loops	TBA
2	11/11	Advanced data management in R	TBA
3	11/18	Programming visualizations using the Lattice library Example: xyplot	TBA
4	11/25	Advanced workflow	TBA
4	12/02	TBA	TBA
	12/06	Courses end	