**Introduction to Data Analysis in R**

DATIC 2018: Workshop Curriculum

*Instructor:* *Randi L. Garcia*

**Session 1: *June 7* *-* *June 8,* *2018***

* prior to Multilevel Modeling with R Workshop, held 6/11-6/15

**Session 2: *June 21* *-* *June 22,* *2018***

* prior to Dyadic Data Analysis with R workshop, held 6/25-6/29

**Should I Attend?**

Are you curious about using R for data analysis? Have you been thinking about making the switch to R, but don’t know where to start? This two-day workshop is the perfect quick start guide to analyzing your data with R. We will cover the fundamentals of data analysis in R with a special focus on translating your existing knowledge and skills from other software (e.g., SPSS) into R. The first day will introduce participants to the RStudio software and cover data management, descriptive statistics, and data exploration including graphical displays. Day 2 will cover familiar statistical analysis procedures including, but not limited to reliability tests, t-tests, chi-square tests, ANOVA, and linear regression. You will have time to work with your own data with the instructor present and optional homework will be provided.

**Overview**

This two-day workshop on Introduction to Data Analysis in R will be held at the University of Connecticut from Thursday, June 7, through Friday, June 8, 2018, then a second session will be held from Thursday, June 21, through Friday, June 22, 2018. The workshop focuses on converting your knowledge of data analysis in another software program—for example, SPSS—to R. All analyses, including data cleaning and visualization, will be done in R via the RStudio graphical interface. RStudio is rapidly becoming popular in many fields, mainly because it is free and therefore widely accessible. This feature makes it particularly helpful for teaching statistics because students will have access to the software at all times, and thus, will not need to go to a computing lab to complete assignments. No prior familiarity with R is required, we will be starting from the very beginning.

The goal of this workshop is to develop proficiency in R for data preparation and preliminary data analysis. We will build confidence in importing data from different sources into RStudio and getting that data ready for any advanced technique you might then employ. Among the topics to be covered are intro to the RStudio environment, packages, and RMarkdown, data manipulation, data visualization, correlations, reliability tests, basic inference tests, ANOVA, linear regression, Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and more. Instruction on the specific statistics and statistical models will be minimal to zero. It is assumed that you already know how to do these analyses, but you want to see how to do them in R. If time permits, more instruction of the less well-known topics (e.g., EFA and CFA) will be given. See the schedule below for a complete list of topics.

Routines for completing specific data tasks, for example, filtering your data, live inside of R packages. In this workshop, you will be introduced to a set of packages which together are referred to as the **tidyverse**. They were written by the Chief Data Scientist for RStudio, Hadley Wickham, and there are even built in cheatsheets for these packages in RStudio. The tidyverse is rapidly gaining popularity due its ease of use, streamlined syntax, and powerful applications. Further, support is easily found for these packages in online community forums such as Stackoverflow. The open source computing nature of R offers both opportunities and challenges. There are currently over 10,000 packages freely available for you to use in R—many of which were written to do the exact same tasks! This overabundance of options for analyzing your data can feel like a bottomless pit, but learning the **tidyverse** offers a bounded experience for working with data in R.

Another important feature of this workshop is that we will use using RMarkdown for all analyses. RMarkdown is a file type, similar to an SPSS syntax file or R script, but it integrates code with written text. In this workshop you will be given all analysis code as RMarkdown files, complete with detailed text explanations, to which you can add your own notes for later. The main benefit of RMarkdown is that, ideally, writing about your analyses can now be seamlessly integrated with the code for your analyses. Further, entire APA style research reports can be written in RMarkdown where the numerical information is automatically generated within the text. Thus, RMarkdown allows for a fully reproducibility research report. On Day 2 of the workshop, some instruction will be given on how to write research reports using RMarkdown.

Participants are encouraged, but not required, to bring their own data so that they can apply these new R skills to their own data. There will be time at the end of each day to work with your own data while the instructor is present for questions/individual meetings. If you are attending a week-long DATIC workshop, it might also be good idea to bring two datasets, 1) a dataset you know very well—to convince yourself that R is doing what you expect, and 2) a dataset that you are preparing for one of the other DATIC workshops: Multilevel Modeling with R or Dyadic Data Analysis with R.

**Benefits:**

The following materials and events are included in the cost of the workshop:

* Booklet with workshop outline, computer setups, and outputs for physical notetaking.
* R code written in RMarkdown, for easy notetaking within RStudio.
* Meals: Continental breakfast each morning

**Recommended Follow-up Reading:**

R for Data Science by Garrett Grolemund and Hadley Wickham

**Schedule:**

* Thursday and Friday:
* 8:30am – 9:00am: Continental Breakfast
* 9:00am – 3:30pm: Workshop, including computer workshops; Lunch on your own
* 3:30pm – 5:00pm: Work on one’s own data; do an optional lab assignment; individual meeting time with workshop instructor available

**Day 1: Working with Data in R**

* Intro to the RStudio Environment, packages, and RMarkdown
* Intro to data manipulation with the **dplyr** package
* Descriptive statistics
* Data transformations with **tidyr**
* Data visualization with **ggplot2**
* Bivariate correlations and reliability tests with **psych**

**Day 2: Statistical Modeling in R**

* Basic inference tests: T-tests and chi-square test
* Analysis of Variance (ANOVA)
* Linear regression
* Logistic regression
* Advanced RMarkdown: Preparing APA style research reports in R with **papaja**
* Exploratory Factor Analysis (EFA)
* Confirmatory Factor Analysis (CFA) with **lavaan**