Volpe Center R Training

Fall 2017

Introduction to R Training Course

This four-session course focuses on essential topics in the R coding language for statistical computing. Volpe Center staff will gain an understanding of how to use R to process data, conduct exploratory data analysis, and communicate results. This course assumes no prior statistical or coding knowledge. Each session will be 1.5 hours.

Instructors will give individual feedback on exercises to assist each participant, and have availability out of class to answer additional questions.

R is a free, powerful, and flexible statistical software environment. This workshop will introduce the essentials of the R language to Volpe Center staff. This workshop will focus on building skills with examples illustrating analysis of transportation-related data. The workshop will be conducted over an 8-week period, with one classroom session every two weeks. Classroom sessions will be 90 minutes to allow time for participants to actively work on exercises. Participants will be expected to complete exercises and write questions for discussion on a shared document on the workshop web page (https://github.com/danfbflynn/volpeR).

Goals

Through this workshop, participants should expect to be able to do the following:

- Learn essential concepts command-line programming
- Understand how R works as an object-oriented programming language
- Write R code to open, create, and organize data files
- Explore a data set of interest with graphical and tabular summaries
- Become familiar with basic statistical approaches in R

Expectations

Participants will be expected to come to each session prepared to actively participate in the hands-on learning. No coding experience is necessary; exposure to basic statistical approaches will be helpful, but not necessary.

Session	Description
1	**R Concepts** - Programming concepts - Workflow in R - Data input - R language: vectors and data frames - Supporting resources: Stackoverflow, R Users Group, QuickR
2	**R Essentials** - R language: factors and lists - Writing functions - Loops and vectorized functions - Logical evaluations
3	**Data exploration** - Reshaping data - Aggregating data - Matching - Basic data analysis - Saving work and sharing
4	**Communicating results** - Graphics in R - Histograms, scatterplots, bar plots, dot plots - ggplot2 and reports in RMarkdown - Tables

Sessions

1) R Concepts

This session will introduce the R language, and present the R approach to data analysis. The workflow will be compared with other software such as SPSS or Stata. Essential concepts in programming will be introduced, and good coding habits will be emphasized. Participants will be introduced to the basics of the R language, in particular the different types of R objects, including vectors and data frames.

Participants must at a minimum install R and RStudio on their laptops prior to the workshop. No coding knowledge will be assumed.

Data input from standard sources (Excel workbooks) will be demonstrated, and initial exploratory data analysis will be shown. Participants will learn how to summarize the distribution, extremes, and central tendencies of data of different types.

Supporting software tools will be introduced, including the use of the cross-platform interface to R, RStudio. Participants will be exposed to the range of resources available for self-education, including various R users' groups, Stackoverflow, and QuickR.

Exercise Participants will read in a data set and carry out initial steps of data input and exploratory data analysis.

Homework Participants will be directed to an R tutorial and be asked to complete supplemental reading on the core concepts of the R language.

2) R Essentials

Building off of the concepts presented in the first session, participants will be exposed to additional aspects of the R language, including factors and lists. The concept of writing a function to automate a task will be presented, and a simple function will be demonstrated.

Loops will be demonstrated, as another way to automate data management and analysis processes, and compared with vectorized processes.

The use of logical evaluations to control functions and loops, as well as to filter data, will be presented.

Exercise A simple function and simple loop will be demonstrated. Participants will create a loop to automate a task using an example data set.

Homework Participants will write an R script to define a function and perform that function in a looped fashion over an example data set.

3) Data exploration

R has numerous tools for processing data. The most common approaches to reshaping data using the base tools as well as new tools of `dplyr` and `tidyr` for data wrangling will be introduced, as well as basic tools for aggregating data in a way familiar to users of PivotTables in Excel.

Matching using regular expressions and other tools will be presented. R approaches to saving work and sharing in a cross-platform way (such as saving .csv files) will be demonstrated.

Exercise Participants will be given data from two sources and will need to import, perform basic data exploration, and save the results.

Homework Participants will write an R script to reshape, aggregate, and summarize an example data set.

4) Communicating results

This session will cover graphical tools and other approaches to presenting results. The base graphics tools are powerful, and there are additional tools which provide easy ways to have high-quality graphics, such as `ggplot2`.

Exercise Participants will prepare a short report using some of the tools introduced here to communicate the results of their example data analysis.

Homework Participants will write an R script to present the analysis and visuals (table, plots, and graphs) with an example data set using the reporting tool.