

Introduction to Data Analysis

with R and Python

**KAP Competencies**

Skills

Knowledge

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Develop and  Grow Professional Capabilities | Leverage Abilities in the Non-Profit Network | Embody the Guiding Principles to Foster a Pro­ductive Culture | Defend a Free Society as Essen­tial to Well-Being  ­ | Use MBM & Five Dimension Think­ing as a Frame­work to Make Better Decisions |
| ­√ | √ | √ | √ | √ |

Course Description

|  |
| --- |
| Course Information  February 12 – March 5, 2015  3:30 – 4:45pm  Instructor  David Ryan Data Scientist  Elder Research, Inc.  Phone  540-521-5909  Email  dtr5a@virginia.edu |

This main goal of this course is to provide instruction that will strengthen participant knowledge of and competency in data analysis. Course sessions will focus on learning how to harness Python and R, two open-source statistical programming languages and software platforms, for analysis in projects of interest. Class time will be divided between lectures and hands-on programming practice. Participants will have the opportunity to use the skills acquired through the course to practice exploring and analyzing relevant data sets.

Objectives

* Deepen understanding of data analysis principles
* Learn R programming basics and their application in data analysis
* Learn Python programming basics and their application in data analysis
* Gain experience using R and Python programming to explore relevant data

Course Materials

* Presentations
* Example data
* Example R scripts
* *TBD*

Week-By-Week

February 12

This session will begin with a discussion about the principles of data analytics. R will be introduced and demonstrated. Basic data exploration features will be highlighted.

**Deliverables:** 1) Understanding of the principles of data analytics and why they are important

2) Introduction to R programming language and software platform

3) Practice using basic data exploration features in R

February 19

Basic R functionality will be reviewed and new features will be introduced for data analysis. Class participants will have the opportunity to work on in-class practice problems with relevant data. An overview of R’s modeling packages and functionality will be presented.

**Deliverables:** 1) Deeper understanding of how R can be used for data exploration and modeling

2) Practice using R to analyze relevant data

February 26

This session will wrap up R instruction with a discussion of R’s reporting capabilities. Python programming languages and software tools will be introduced and demonstrated. Class participants will be able to use Python to explore relevant datasets.

**Deliverables:** 1) Understanding of R’s reporting capabilities

2) Introduction to Python programming language and software platform

3) Practice using basic Python functionality

March 5

The final session will explore Python in greater depth, focusing on its use in data exploration. Advanced features, such as modeling tools, will also be introduced, and students will be able to practice Python with in-class problems. A concluding discussion will briefly compare R and Python, as demonstrated through the course, and allow time for final questions.

**Deliverables:** 1) Deeper understanding of how Python can be used for data exploration and modeling

2) Practice using Python to analyze relevant data

3) Understanding of the basic differences between R and Python for data analysis