

Data Analysis with R

A TO Z ACADEMY
info@atozacademy.nl

January 21, 2018

Course Overview

Data Analysis with R is a 15 hour intensive program wherein the participants will “learn by doing“. In this regard, two case studies will be shared, which will be prepared from

- <https://www.kaggle.com/datasets>

The idea is that participants apply the techniques that they learn during the first six sessions, on the cases and prepare a presentation of their findings for the last two sessions. Also the participants will be expected to submit their R code before the last two sessions. Each case will require the participants to spend approximately 20-25 hours.

Course Plan

1. **Introduction to R (3 hours)** In the introduction sessions we will cover the following

- Setup R and Introduction to RStudio
- Reading files in R. Introduction to data frames and basic operation on Data frames

2. **Data Processing in R (1.5 hours)**

In these sessions we will cover different control structures in R. The participants will learn to write their first “R function” and ways to clean, aggregate and transform data

3. **Exploratory Data Analysis and Data Visualization in R (3 hours)**

In these sessions the participants will learn how to compute summary statistics of data in R, visualize datasets and learn to understand and explain “the story of the dataset”

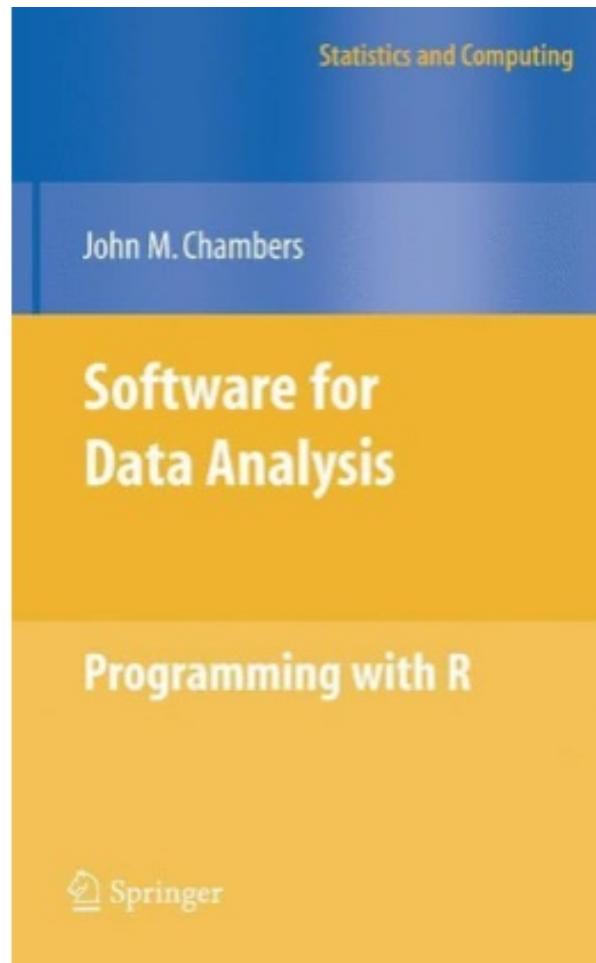
4. **Statistical Modelling in R (1.5 hours)**

In these sessions, the participants will learn how to build a linear regression model on a multivariate dataset and interpret its results.

5. **Case Analysis (6 hours)**

In these sessions, the participants will present, analyse and brainstorm on the application of different methods that they have been introduced to in the previous sessions, namely data processing, data exploration, data visualization and predictive modelling to two real world datasets.

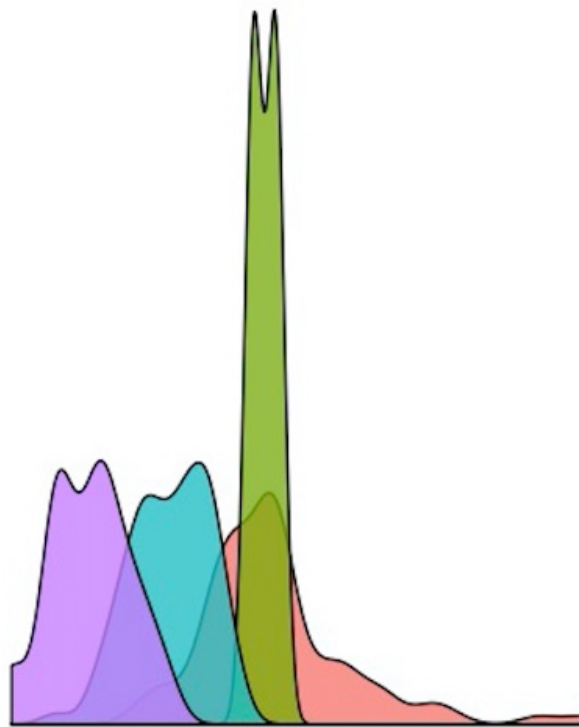
The Book



Session	Chapters in the Book
1	1, 2, 6.1, 6.2
2	3, 6.3, 6.4, 6.5, 6.6, 6.7
3	7
4	6.8, 6.9, 6.10

Session 1 - Introduction to R

The Mission

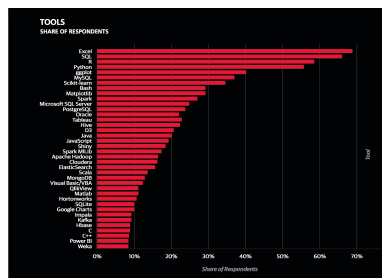


“Data Exploration”

Data Exploration entails the following activities:

- Formulate meaningful questions
- Organize data in a way to answer the questions

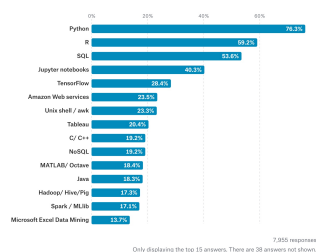
Why R?



What tools are used at work?

Python was the most commonly used data analysis tool across employed data scientists overall, but more [Statisticians](#) are still loyal to R.

[Company Size](#) [Industry](#) [Job Title](#)



Only displaying the top 15 answers. There are 38 answers not shown.

A Brief History of R

Concepts for Programming with R

Functional Programming

Classes and Methods

Data Frames