

EdX and its Members use cookies and other tracking technologies for performance, analytics, and marketing purposes. By using this website, you accept this use. Learn more about these technologies in the [Privacy Policy](#).



PH125.1x: Data Science: R Basics - Course Syllabus

Course Instructor

Rafael Irizarry

Course Description

In this first course of nine in the [HarvardX Data Science Professional Certificate](#), we learn the basic building blocks of R.

The demand for skilled data science practitioners in industry, academia, and government is rapidly growing. The Harvard Data Science Series prepares you with the necessary knowledge base and skills to tackle real world data analysis challenges. We cover concepts such as probability, inference, regression and machine learning and develop skill sets such as R programming, data wrangling with **dplyr**, data visualization with **ggplot2**, file organization with unix, version control with GitHub, and reproducible document preparation with RStudio. Throughout the series, we use motivating case studies that ask specific questions and answer them through data analysis. Some of our assessments use code checking technology that will permit you to get hands-on practice during the courses.

Throughout the series, we will be using the R software environment for all our analysis. You will learn R, statistical concepts, and data analysis techniques simultaneously. In this course, we will introduce basic R syntax to get you going. However, rather than cover every R skill you need, we introduce just enough to get you going with the next courses in this series, which will provide more in depth coverage, building upon what you learn here. We believe that you can

better retain R knowledge when you learn it to solve a specific problem.

Using a motivating case study, we ask specific questions related to crime in the United States and provide a relevant dataset. You will learn some basic R skills to permit us to answer these questions.

What you'll learn:

- how to read, extract, and create datasets in R
- how to perform a variety of operations and analyses on datasets using R
- how to write your own functions/sub-routines in R

New to EdX?

Are you new to edX? Check out edX's [Demo Course!](#)

Need help? Visit edX Support via the Support tab or [visit the Help Center](#).

Course Structure

All material will be released when the course opens. You can progress through the material at your own pace.

Grading

There are 9 interactive programming exercises using the DataCamp platform, and 4 sets of exercises directly on the edX platform. These 13 sets of exercises together make up **100%** of your grade. 12 of the 13 sets of exercises are available to all learners; the 13th set of exercises is available to Verified learners only.

All other components of the course, such as the discussion boards, are not for credit.

Certification

In order to receive a Verified Certificate, you must sign up and pay for a Verified Certificate by the deadline on the course page and earn a passing grade of at least **70%**.

COURSE OUTLINE

Section 1: R Basics, Functions, and Data Types

You will get started with R and learn about R's functions and data types.

Section 2: Vectors and Sorting

You will learn to operate on vectors and advanced functions such as sorting.

Section 3: Indexing, Data Manipulation, and Plots

You will learn to wrangle, analyze and visualize data.

Section 4: Programming Basics

You will learn to use general programming features like 'if-else', and 'for loop' commands to write your own functions to perform various operations on datasets.

Installing R

One can download R freely from the [Comprehensive R Archive Network](#) (CRAN). It is relatively straightforward, but if you need further help you can check out [chapter 1 of the textbook](#).

FAQS

What is the deadline to sign up for a Verified Certificate?

The deadline is listed on the right side of the course landing page.

How do I earn a certificate?

To earn a certificate, you must sign up for a Verified Certificate by the deadline and earn a grade of at least 70%. When you achieve this score, a view your certificate button will appear on your dashboard. For more information, [click on this link](#).

How do I upgrade to a verified certificate?

Go to your edX Dashboard (by clicking the edX icon at the top left of this page). Under this course, click the "Challenge Yourself!" link.

How long does the course take?

That is up to you! It is 4 weeks of content. Just be aware of the course close date on the right side of the course landing page!

I am doing well on the assessments, but when I look under "Progress" I have a very low grade...why?

The grade is calculated based on all of the assessments you have completed **and** the assessments that you have not completed (edX says you have a "zero" on those assessments **until** you have attempted them). You will see your overall grade move up as you progress through the course.

How often will the courses be offered?

Courses in the program are offered frequently, with overlap - so if now isn't a good time for you to start one of the courses you need as a prerequisite or if you missed a deadline, there will be another offering of the course you need coming soon!

Please note that progress does not carry over from one offering to another.

Does the order of courses in the Professional Certificate Program matter?

Yes, order does matter, particularly for the first four courses in the sequence. For the later courses, depending on your previous experience, you may be able to swap the sequence of some of the courses. The courses are designed to be taken in the following order:

1. R Basics
2. Visualization
3. Probability
4. Inference and Modeling

5. Productivity Tools

6. Wrangling

7. Linear Regression

8. Machine Learning

9. Capstone

Do I need to register for all of the courses at once in order to be eligible for the Professional Certificate?

No! You can take courses individually - once you have obtained an ID Verified Certificate in each course, you will be eligible for the Professional Certificate. If you choose to pre-pay for the entire program, you receive a discount on the total registration cost.

Online R references

- R reference card (PDF) by Tom Short (more can be found under Short Documents and Reference Cards [here](#))
- Quick-R: quick online reference for data input, basic statistics and plots
- R programming class on Coursera, taught by Roger Peng, Jeff Leek and Brian Caffo

R books

- *Software for Data Analysis: Programming with R (Statistics and Computing)* by John M. Chambers (Springer)
- *S Programming (Statistics and Computing)* by Brian D. Ripley and William N. Venables (Springer)
- *Programming with Data: A Guide to the S Language* by John M. Chambers (Springer)

Research

HarvardX pursues the science of learning. When you participate in this course, you will also participate in research about learning. Read our [research](#)

statement to learn more.