FUNDAMENTALS OF DATA SCIENCE FOR SOCIAL SCIENTISTS

Full data science e-Learning solution for your institution

Upskilling a group of students, researchers, faculty or staff in data science can be daunting and time consuming. Finding quality courses with relevant, applicable content that suit groups with varying skill sets and schedules is no easy feat.

Fundamentals of Data Science for Social Scientists is a full suite of ten online courses designed to solve your institution’s pedagogical needs and get your team or group mobilized and practicing the latest computational methods. The learning path covers everything from the big picture of data science, getting started with programming skills in R and Python, through to more advanced ‘bytesize’ data management and analysis topics. There’s something to offer everyone.

SAGE Campus created the courses in partnership with the Social Science Data Lab (D-Lab) at the University of Berkeley, so you can get courses by leading expert instructors, with the trusted SAGE quality, and an unparalleled online learning experience.

COURSES ARE SPLIT INTO 3 SECTIONS

- **Introductory:** Introduction to Data Science for Social Scientists
  - Consists of 1 introductory course on the theory of data science and its applicability to social science and takes 5-9 hours to complete.

- **Beginner:** Programming in R and Programming in Python
  - Consists of 2 courses, where learners can choose to learn a programming language of their choice; Python (which takes 33 hours), or R (which takes 35 hours).

- **Intermediate:** Bytesize courses on core data management and analysis skills
  - Consists of 7 standalone “bytesize” courses so learners can choose topics most relevant to them. The courses take a total of 30 hours combined.
LIST OF THE COURSES

1. Introduction to Data Science
2. Introduction to Data Science with R
3. Introduction to Data Science with Python
4. Bytesize: Collecting Data from the Web
5. Bytesize: Cleaning Data and Preprocessing
6. Bytesize: Data Formats
7. Bytesize: Network Analysis
8. Bytesize: Data Visualization
10. Bytesize: Text Analysis

Pricing
Your institution can purchase a 6 subscription to Fundamentals of Data Science for Social Scientists. Pricing depends on number of learners and your institution's requirements, starting at a minimum of 10 learners for $5,000.

FREE DEMO ACCESS

Demo the course package here:
https://classroom.sagepub.com/course/view.php?id=101
Demo key: datasci-fun-DEMO

CONTACT US

RACHEL CROOKES
Head of SAGE Campus
Rachel.Crookes@sagepub.co.uk

AMY SPARROW
Marketing Manager
Amy.Sparrow@sagepub.co.uk

WEBSITE
https://campus.sagepub.com
THE SYLLABUS FOR EACH COURSE IN THE SUITE OF 10 IS OUTLINED ON THE FOLLOWING PAGES

1. Introduction to Data Science for Social Scientists (introductory)

IMODULE ONE: INTRODUCTION AND OVERVIEW TO DATA SCIENCE
This module will provide a social science perspective on data science, introducing you to the objectives of the course via a visual overview of each module. It will discuss how data science is changing social science and statistics, and will cover reliability, generalizability, and reproducibility.

MODULE TWO: ETHICS IN DATA SCIENCE
This module will teach you about the shortcomings and problems of data science in respect to the groups of people it affects, who it’s representing, and how to responsibly acknowledge these issues in research. Beyond problems in data collection and data sources are issues of privacy, sampling, population size, interpretation, and application. This module importantly emphasizes issues of deidentification and reidentification, and data security.

MODULE THREE: SURVEYS AND CROWDSOURCING DATA
This module will give an overview of how to construct a survey and crowdsource responses. Specifically, you will be introduced to Qualtrics and learn about other freely available tools for building surveys. We will discuss Amazon’s Mechanical Turk, which is becoming the new norm for online data collection in the social sciences.

MODULE FOUR: DATA SCIENCE TOOLS
This module will introduce you to the data science tools commonly used in social science research. We will discuss the value of open-source programming languages, specifically R and Python, for research of this nature and weigh the advantages and disadvantages of each.

This module introduces Jupyter notebooks, and concludes with a brief overview of Git and GitHub as they have become essential for collaborative research programming projects.
2. Introduction to Data Science with R (beginner)

INTRODUCTION & OVERVIEW OF R
An introduction to programming and the R language.

JUPYTER NOTEBOOKS
How do I use a Jupyter Notebook?

VARIABLES
How can I store data in programs?

DATA TYPES
What kinds of data do programs store and how can I convert one type to another?

DATA STRUCTURES
How do I organize several data types in an ordered manner and how can I modify this collection of data?

LOADING DATA
How do I open a file and read its contents into R, read a CSV file into a data frame, and how do I get back basic information about the shape of my data?

SUBSETTING
How do I only work with select columns and how can I reduce my data based on certain conditions?

MISSING DATA
What do I do if my data are incomplete?

MERGING AND RESHAPING
How do I combine data from two different sources and how do I change between 'long' and 'wide' formats?

DATA EXPLORATION
How do I get quick and informative descriptive statistics of my data?

PLOTTING
How do I visualize my data?

GGPLOT2
Is there an easier way to make plots and how can I better customize them?
2. Introduction to Data Science with R continued

**STATISTICAL TESTING**
How do I carry out hypothesis testing on my data and how can I model my data in a linear regression?

**FOR LOOPS**
How can I make a program do many things?

**CONDITIONALS**
How can programs do different things for different data?

**FUNCTIONS**
How can I avoid rewriting code that I will use again?

**MONTE CARLO**
How do I randomly sample from my data?

**THE BIRTHDAY PROBLEM**
How do I put this all together to answer a probability question?
3. Introduction to Data Science with Python (beginner)

INTRODUCTION & OVERVIEW OF PYTHON
An introduction to programming and the Python language

JUPYTER NOTEBOOKS
How do I use a Jupyter Notebook?

VARIABLE ASSIGNMENT
How can I store data in programs?

DATA TYPES & COERCION
What kinds of data do programs store and how can I convert one type to another?

STRINGS
How do I manipulate strings (text)?

BUILT-INS
How can I use built-in functions and find out what they do?

LISTS
How do I organize several data types in an ordered manner and how can I modify this collection of data?

LOOPS
How can I make a program do many things?

CONDITIONALS
How can programs do different things for different data?

FUNCTIONS
How can I avoid rewriting code that I will use again?

FUNCTIONS AND VARIABLE SCOPE
How do function calls actually work?

DICTIONARIES
How can I store more complicated data and retrieve it efficiently?

FILES
How do I open a file and read its contents and write a file with the variables I generated?
3. Introduction to Data Science with Python continued

LIBRARIES
How can I use software that other people have written and find out what it does?

ERRORS
How do I read, interpret and fix an error message, and what if I still can’t figure it out?

LIST COMPREHENSIONS
Is there a faster way to generate a new list by changing all the elements in an old one and how can I write cleaner, more compact code?

PROGRAMMING STYLE
How can I make my programs more readable and how do most programmers format their code?
4. **Bytesize: Collecting Data from the Web (intermediate)**

Teaches learners how to extract data from web resources appropriate to their research questions. Special attention will be given to how to obtain permission from hosts, and proper etiquette when using APIs and scraping. By the end of this bytesize course, learners will be able to:

- Explain in simple terms how the internet works.
- Define and use an API to collect data from the web.
- Explain the difference between using an API and web scraping.
- Recognize potential legal issues surrounding web scraping.
- Use a programming language to collect web data.

5. **Bytesize: Cleaning Data and Preprocessing (intermediate)**

Teaches learners how to prepare data so that it is in a format that can be recognized by the coding function in R or Python. By the end of this bytesize course, learners will be able to:

- Define what cleaning and preprocessing data are.
- Explain why these steps are necessary.
- Identify common cleaning tasks and possible solutions.
- Use regular expressions to standardize text.
- Join multiple data sources together.
- Incorporate best practices into your data science workflow.

6. **Bytesize: Data Formats (intermediate)**

Teaches learners what formats data comes in, and how they should structure their own data if they collect it themselves. By the end of this bytesize course, learners will be able to:

- Name the most popular data formats used today.
- Explain the difference between data formats.
- Summarize the reasons for storing data in these different formats.
- Decide which format a dataset should be kept in.
- Read and write these formats with Python or R.
7. Bytesize: Network Analysis (intermediate)
Teachers learners how to model explicit relationships, how to examine the statistical properties of relationships in co-mention networks, and how contextualize the statistical properties of a network. By the end of this bytesize course, learners will be able to:

- Understand what constitutes a social network.
- Identify and describe different levels of analysis.
- Use the open source software tool GEPHI to do your own network analyses.
- Correctly interpret network properties.

8. Bytesize: Data Visualization (intermediate)
Teachers learners effective presentation methods for various data types and variables, and how to create their own visualizations in Jupyter notebooks in Python or R. By the end of this bytesize course, learners will be able to:

- Explain two popular visualization theories: Marks and Visual Variables, and Gestalt.
- Enumerate a hierarchy of visual perception.
- Apply these theories to assess the effectiveness of visualizations.
- Describe how to convey an entire story using an infographic.
- Use visualizations for exploratory data analysis.
- Write code to produce geographic visualizations.

Teaches learners the basics of machine learning and core organizational concepts of classification and regression, data preprocessing and fitting a model to a training dataset. By the end of this bytesize course, learners will be able to:

- Identify machine learning applications.
- Explain the importance of preprocessing your data for machine learning.
- Explain the rationale for splitting data into training, cross-validation and test sets.
- Understand and apply basic ideas about algorithm construction and configuration settings.
- Situate ensemble methods in the broader machine learning environment.

10. Bytesize: Text Analysis (intermediate)
Teaches learners the building blocks that serve as the foundation for computational text analysis. By the end of this bytesize course, learners will be able to:

- List and justify or criticize common preprocessing steps.
- Explain the "bag of words" (BoW) model.
- Define TFIDF value.
- Define "n-gram" and explain how it improves our language model.
- Create features suitable for a classification model.
- Correctly interpret a topic model.