

---

## Repo Purpose

Welcome to your community-sourced data science repo! The overarching goal here is to provide anyone interested in learning data science with a wealth of open source, industry-best learning materials and learning tracks.

This repo is a work in progress. Please check back for updates. @momiji15, @Annu-07, and I are collaborating on the structure for this repo. If you would like to be involved in that process, please file an issue in this repo and we will add you to our Slack channel.

This repo is motivated by recent incidents. The data science community deserves better, and this repo is an attempt to provide a platform for the excellent learning resources available.

## Guided Data Science Resources

- Dataquest
- Software Carpentry Lessons
- Data Carpentry Lessons
- Chromebook Data Science
- Business Science University

## Direct Course Replacements

Many instructors have admirably advocated against taking their own DataCamp courses. Often, these instructors have suggested other ways in which learners can access the same material. The suggested replacements for their courses are listed below:

### R Courses

Introduction to R

- Also see here

RYouWithMe

Ready for R

Intermediate R

Working with the RStudio IDE (Part 1)

---

Working with the RStudio IDE (Part 2)

Cleaning Data in R

- Also see here

Importing & Cleaning Data in R: Case Studies

- See Guided Project: NYC Schools Perceptions

Working with Dates and Times in R

Categorical Data in the Tidyverse

- Also see here

Writing Functions in R

- Next, learn about iteration

Data Manipulation in R with dplyr

Data Analysis in R, the data.table Way

- See here for updating
- See here for indexing

Building Processing Pipelines in `data.table`

Developing R Packages

- Also see the `usethis` documentation

Foundations of Probability in R

- See weeks 3 and 4
- Also see here

Dealing With Missing Data in R

Dimensionality Reduction in R

Advanced Dimensionality Reduction in R

Foundations of Inference

Correlation and Regression

Fundamentals of Bayesian Data Analysis in R

---

Structural Equation Modeling with lavaan in R

Introduction to Machine Learning

Supervised Machine Learning: Case Studies in R

Unsupervised Learning in R

Machine Learning Toolbox

- Also see a book on the [caret](#) package

Differential Expression Analysis in R with limma

Bayesian Regression Modeling with [rstanarm](#)

- Also see a walkthrough article and a practical example

Forecasting Using R

Introduction to Time Series Analysis

- Also see here

ARIMA Modeling with R

Forecasting Product Demand in R

- Also see here

Nonlinear Modeling in R with GAMs

Marketing Analytics in R: Choice Modeling

- Please see Chapter 13

Hyperparameter Tuning in R \* Also see the [mlr](#) package docs and the [h2o](#) package docs

Exploratory Data Analysis

Exploratory Data Analysis in R: Case Study

- Please see week 4

Visualization Best Practices in R

Data Visualization with ggplot2 (Part 1)

Data Visualization with ggplot2 (Part 2)

Building Dashboards with [shinydashboard](#)

---

Building Dashboards with [flexdashboard](#)

Interactive Data Visualization with [rbokeh](#)

Interactive Maps with [leaflet](#) in R

Working with Geospatial Data in R

Building Web Applications in R with Shiny

Building Web Applications in R with Shiny: Case Studies

Introduction to Text Analysis in R

Text Mining with R

- Also see the associated blogs and tutorials

Sentiment Analysis in R

Sentiment Analysis in R: The Tidy Way

Analyzing Election and Polling Data in R

Analyzing US Census Data in R

Single-Cell RNA-Seq Workflows in R

Data science for the medical and biomedical sciences (ds4biomed)

## **Python Courses**

Introduction to Python

- Also see here

Python for R Users

Python for MATLAB Users

Introduction to Data Science in Python

Elements of Data Science

Intermediate Python for Data Science

Object-Oriented Programming in Python

Writing Efficient Python Code

Analyzing Police Activity with pandas

---

Interactive Data Visualization with Bokeh

Advanced NLP with spaCy

Intro to Python for Finance

## **SQL Courses**

Intro to SQL for Data Science

- Also see here (**requires subscription**)

Intermediate SQL

- Please see the Intermediate SQL section

Intermediate SQL Server

Joining Data in SQL

- Also see here (**requires subscription**)

## **Git Courses**

Introduction to Git for Data Science

- Also see here
- Also see git branching

## **Shell Courses**

Introduction to Shell for Data Science

- Also see here
- Also see here (**requires subscription**)

## **Spreadsheet Courses**

Spreadsheet Basics

---

## **Contributing**

Please feel free to submit a pull request. The full list of DC courses can be found [here](#)