Data Science Certificate Sample

Michael Espero

Email:
michael@perceptioninterval.com

Cell:
+1 909 267 8989

TABLE OF CONTENTS

ABOUT DATACAMP **3**

CERTIFICATIONS **4**

DATA SCIENTIST 4
COURSE CERTIFICATES 7-79

COURSE SYLLABI **80-148**

DATA SCIENTIST TRACK DETAILS **139-140**

PREVIOUS CURRICULUM 139
PROOF OF COMPLETION OF UPDATED CURRICULUM 140

DataCamp | What would you like to learn today? | Learn | Pricing | For Business | My Dashboard

From data to insights

DataCamp helps companies and individual learners answer their most challenging questions by making better use of data. Our learners build and maintain data fluency on the world's most advanced online learning platform for data science and analytics.

4,220,000	1,200+	844	236
Aspiring data scientists →	Companies upskilling their teams →	Courses, projects and practice exercises →	Instructors →

Modern learning experience

DataCamp makes it easy to build data science and analytics skills at your own pace with interactive courses, expert instruction from the world's top data scientists, and real-time feedback.

Continuous learning

DataCamp is designed for every stage of your career, and available on demand whenever you need to build or deepen your data skills. We are constantly expanding our curriculum so that you can keep up with the latest in data science and analytics.

Data science and analytics at scale

DataCamp enables managers to close the talent gap and embed data fluency across an entire organization, regardless of size or structure. We have more than 4.2 million learners around the world at leading companies.

Spreading data fluency

DataCamp is building the most comprehensive online learning resource for data science and analytics. We define success as helping individuals and companies to become more data fluent. Read more here.

Our investors

SPECTRUM EQUITY

ACCOMPLICE

arthur ventures

Our offices

NEW YORK, USA
Empire State Building
350 5th Ave, Floor 77
NY 10118

LEUVEN, BE
Martelarenlaan 38
3010 Kessel-Lo

LONDON, UK
WeWork, The Bower
207 Old Street
London EC1V 9NR

Contact Us

DataCamp

DataCamp offers interactive R, Python, Sheets, SQL and shell courses. All on topics in data science, statistics and machine learning. Learn from a team of expert teachers in the comfort of your browser with video lessons and fun coding challenges and projects.

ABOUT THE COMPANY

© 2019 DataCamp Inc.

#5,703

CERTIFICATE NUMBER

STATEMENT OF ACCOMPLISHMENT

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Scientist with R Track

DataCamp

#4,526

CERTIFICATE NUMBER

STATEMENT OF ACCOMPLISHMENT

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Analyst with R Track

DataCamp

#4,799

CERTIFICATE NUMBER

STATEMENT OF ACCOMPLISHMENT

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

R Programmer Track

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7839249

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Advanced Deep Learning with Keras in Python

COMPLETED ON

Dec 14, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6647511

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Bayesian Modeling with RJAGS

COMPLETED ON

Sep 02, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6920701

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Bayesian Regression Modeling with rstanarm

COMPLETED ON

Sep 05, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6250328

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Building Dashboards with shinydashboard

COMPLETED ON

Aug 19, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#8816333

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

ChIP-seq Workflows in R

COMPLETED ON

Mar 08, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3179641

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Cleaning Data in R

COMPLETED ON

Jun 06, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6158948

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Cluster Analysis in R

COMPLETED ON

May 31, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7937059

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Convolutional Neural Networks for Image Processing

COMPLETED ON

Dec 19, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3189566

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Correlation and Regression

COMPLETED ON

Jun 13, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3241680

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Manipulation in R with dplyr

COMPLETED ON

Jun 13, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6149135

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Privacy and Anonymization in R

COMPLETED ON

Jun 08, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3719290

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Visualization in R with lattice

COMPLETED ON

Aug 18, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3189650

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Visualization in R

COMPLETED ON

Jun 16, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3189624

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Visualization with ggplot2 (Part 1)

COMPLETED ON

Jun 07, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3268110

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Visualization with ggplot2 (Part 2)

COMPLETED ON

Jun 20, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3296418

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Data Visualization with ggplot2 (Part 3)

COMPLETED ON

Jul 06, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6450937

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Deep Learning in Python

COMPLETED ON

Jul 29, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6371610

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Democratizing Data Science Within Your Company

COMPLETED ON

Jun 26, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#8947736

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Differential Expression Analysis in R with limma

COMPLETED ON

Apr 05, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6617644

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Experimental Design in R

COMPLETED ON

Jul 29, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3305724

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Exploratory Data Analysis in R: Case Study

COMPLETED ON

Jun 21, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3285891

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Exploratory Data Analysis

COMPLETED ON

Jun 20, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3683169

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Forecasting Using R

COMPLETED ON

Aug 15, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3266756

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Foundations of Inference

COMPLETED ON

Jun 17, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3601044

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Foundations of Probability in R

COMPLETED ON

Aug 04, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6568834

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Fundamentals of Bayesian Data Analysis in R

COMPLETED ON

Jul 30, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6995687

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Generalized Linear Models in R

COMPLETED ON

Sep 19, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#5931221

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Hierarchical and Mixed Effects Models

COMPLETED ON

May 28, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7964379

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Hyperparameter Tuning in R

COMPLETED ON

Dec 31, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3214766

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Importing & Cleaning Data in R: Case Studies

COMPLETED ON

Jun 12, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3144240

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Importing Data in R (Part 1)

COMPLETED ON

Jun 01, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3179423

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Importing Data in R (Part 2)

COMPLETED ON

Jun 09, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6680953

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intermediate Python for Data Science

COMPLETED ON

Aug 09, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3133499

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intermediate R: Practice

COMPLETED ON

May 31, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3104202

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intermediate R

COMPLETED ON

May 30, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3646176

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intro to Statistics with R: Moderation and Mediation

COMPLETED ON

Aug 10, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3771266

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intro to Statistics with R: Multiple Regression

COMPLETED ON

Nov 27, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3243341

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Data

COMPLETED ON

Jun 14, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3535629

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Machine Learning

COMPLETED ON

Aug 09, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3525042

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Python

COMPLETED ON

Aug 03, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#1331175

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to R

COMPLETED ON

May 26, 2017

STATEMENT OF ACCOMPLISHMENT

#3349117

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Intro to SQL for Data Science

COMPLETED ON

May 29, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#4564909

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to the Tidyverse

COMPLETED ON

Dec 18, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3534993

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Time Series Analysis

COMPLETED ON

Jul 31, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3251818

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Joining Data in R with dplyr

COMPLETED ON

Jun 15, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6304460

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Joining Data in SQL

COMPLETED ON

Jun 22, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7348477

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Linear Algebra for Data Science in R

COMPLETED ON

Feb 12, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7097608

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Machine Learning in the Tidyverse

COMPLETED ON

Sep 21, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3349088

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Machine Learning Toolbox

COMPLETED ON

Jun 30, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#8866673

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Mixture Models in R

COMPLETED ON

Mar 16, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#4644540

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Multiple and Logistic Regression

COMPLETED ON

Dec 22, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3334714

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Object-Oriented Programming in R: S3 and R6

COMPLETED ON

Jun 26, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6352728

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Python for R Users

COMPLETED ON

Mar 30, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3296447

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Reporting with R Markdown

COMPLETED ON

Jun 22, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#8648022

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

RNA-Seq Differential Expression Analysis

COMPLETED ON

Mar 01, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6675078

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Single-Cell RNA-Seq Workflows in R

COMPLETED ON

Mar 16, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3397890

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Spatial Statistics in R

COMPLETED ON

Jul 06, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3439462

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Statistical Modeling in R (Part 1)

COMPLETED ON

Jul 19, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3510738

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Statistical Modeling in R (Part 2)

COMPLETED ON

Jul 21, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3316763

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

String Manipulation in R with stringr

COMPLETED ON

Jun 26, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6251044

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Structural Equation Modeling with lavaan in R

COMPLETED ON

Jun 10, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#5930862

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Supervised Learning in R: Case Studies

COMPLETED ON

Jun 12, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#4564794

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Supervised Learning in R: Classification

COMPLETED ON

Jun 03, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3638007

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Supervised Learning in R: Regression

COMPLETED ON

Aug 14, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7257091

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Support Vector Machines in R

COMPLETED ON

Oct 06, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#9240689

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Survival Analysis in R

COMPLETED ON

Apr 14, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3414903

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Text Mining: Bag of Words

COMPLETED ON

Jul 09, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3378138

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Unsupervised Learning in R

COMPLETED ON

Jun 30, 2017

DataCamp

STATEMENT OF ACCOMPLISHMENT

#7146184

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Visualization Best Practices in R

COMPLETED ON

Oct 07, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#6177866

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Working with Dates and Times in R

COMPLETED ON

Jun 15, 2018

DataCamp

STATEMENT OF ACCOMPLISHMENT

#10042694

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Relational Databases in SQL

COMPLETED ON

Jul 11, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#10100720

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Introduction to Shell for Data Science

COMPLETED ON

Jul 12, 2019

DataCamp

STATEMENT OF ACCOMPLISHMENT

#3133538

HAS BEEN AWARDED TO

Michael Espero

FOR SUCCESSFULLY COMPLETING

Writing Functions in R

COMPLETED ON

Jun 05, 2017

DataCamp

INTERACTIVE COURSE

Democratizing Data Science Within Your Company

Replay Course

0 hours | 1 Videos | 3 Exercises | 31 Participants | 200 XP

Course Description

In this webinar, DataCamp's Chief Data Scientist David Robinson will show you several examples of how learning R or Python can make your staff more effective, increase their satisfaction, and free up your data team to focus on executing more challenging and innovative projects. He'll share what he's learned from following this philosophy within DataCamp, where everyone at the company is encouraged to take several DataCamp courses and contribute to making us a data-driven company. You'll then try DataCamp for yourself by completing the first 2 exercises of David's course Introduction to the Tidyverse.

1 Data Science Isn't Just for Data Scientists

100%

The Webinar video followed by a quick product demo!

- Democratizing Data Science Within Your Company — ✓ 50 xp
- Thanks for watching! Now, here's a quick product demo — ✓ 100 xp
- Understanding a data frame — ✓ 50 xp

HIDE CHAPTER DETAILS ^

✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

David Robinson
Chief Data Scientist, DataCamp

Dave works on the data science behind DataCamp's product and curriculum development. He has worked as a data scientist at Stack Overflow and received his PhD in Quantitative and Computational Biology from Princeton University. His interests include statistics, data analysis, education, and programming in R. Follow him at @drob on Twitter or on his blog, Variance Explained.

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Bayesian Regression Modeling with rstanarm

Replay Course

4.4 (186 Ratings)

4 hours | 15 Videos | 45 Exercises | 1,553 Participants | 3,400 XP

Course Description

Bayesian estimation offers a flexible alternative to modeling techniques where the inferences depend on p-values. In this course, you'll learn how to estimate linear regression models using Bayesian methods and the rstanarm package. You'll be introduced to prior distributions, posterior predictive model checking, and model comparisons within the Bayesian framework. You'll also learn how to use your estimated model to make predictions for new data.

STATEMENT OF ACCOMPLISHMENT

View

1 Introduction to Bayesian Linear Models FREE 100%

A review of frequentist regression using lm(), an introduction to Bayesian regression using stan_glm(), and a comparison of the respective outputs.

- Non-Bayesian Linear Regression ✓ 50 xp
- Exploring the data ✓ 100 xp
- Fitting a frequentist linear regression ✓ 100 xp
- Bayesian Linear Regression ✓ 50 xp
- Fitting a Bayesian linear regression ✓ 100 xp
- Convergence criteria ✓ 50 xp
- Assessing model convergence ✓ 50 xp
- Comparing frequentist and Bayesian methods ✓ 50 xp
- Difference between frequentists and Bayesians ✓ 50 xp
- Creating credible intervals ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Modifying a Bayesian Model 100%

Learn how to modify your Bayesian model including changing the number and length of chains, changing prior distributions, and adding predictors.

- What's in a Bayesian Model? ✓ 50 xp
- Altering chains ✓ 100 xp
- Do I have enough iterations? ✓ 50 xp
- Prior Distributions ✓ 50 xp
- Determine Prior Distributions ✓ 100 xp
- Calculate Adjusted Scales ✓ 100 xp
- Unadjusted Priors ✓ 100 xp
- User Specified Priors ✓ 50 xp
- Changing Priors ✓ 100 xp
- Specifying informative priors ✓ 100 xp
- Consequences of informative priors ✓ 50 xp
- Tuning Model for Stability ✓ 50 xp
- Altering the Estimation ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Assessing Model Fit 100%

In this chapter, we'll learn how to determine if our estimated model fits our data and how to compare competing models.

- Using the R Squared Statistic ✓ 50 xp
- Calculating Frequentist R-squared ✓ 100 xp
- R-squared for a Bayesian Model ✓ 100 xp
- Posterior Predictive Model Checks ✓ 50 xp
- Predicted score distributions ✓ 100 xp
- Distributions for a single observation ✓ 50 xp
- Model Fit With Posterior Predictive Model Checks ✓ 50 xp
- R-squared Posterior ✓ 100 xp
- Posterior Predictive Testing ✓ 100 xp
- Bayesian Model Comparisons ✓ 50 xp
- Calculating the LOO estimate ✓ 100 xp
- Comparing models ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Presenting and Using a Bayesian Regression 100%

In this chapter, we'll learn how to use the estimated model to create visualizations of your model and make predictions for new data.

- Visualizing a Bayesian Model ✓ 50 xp
- Plotting a Bayesian model ✓ 100 xp
- Plotting Model Uncertainty ✓ 100 xp
- Making Predictions ✓ 50 xp
- Popularity for Observed Songs ✓ 100 xp
- Popularity for New Songs ✓ 70 xp
- Visualizing Predictions ✓ 50 xp
- Format prediction posteriors ✓ 100 xp
- Visualize New Predictions ✓ 100 xp
- Conclusion ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

Jake Thompson
Psychometrician, ATLAS, University of Kansas

Jake is a Psychometrician at the Center for Accessible Teaching, Learning, and Assessment Systems (ATLAS) and received his PhD in Educational Psychology and Research. His interests are include educational assessment, diagnostic classification modeling, and Bayesian inference. Follow him at @wjakethompson on Twitter or on his blog.

See More

COLLABORATOR(S)

Chester Ismay
David Campos
Shon Inouye

PREREQUISITES

- Data Visualization with ggplot2 (Part 1)
- Multiple and Logistic Regression
- Bayesian Modeling with RJAGS

DATASETS

- Spotify dataset

© 2019 DataCamp Inc.

FREE PROJECT

Phyllotaxis: Draw Flowers Using Mathematics

Use R to make art and create imaginary flowers inspired by nature.

Replay Project | Download Notebook

9 tasks | 11,830 participants | 1,500 XP

Project Description

R is a tool for doing serious statistics and data analysis. But not everything in life can be serious, life is also beautiful, and R can make beautiful things too. R can make art.

The arrangement of leaves on a plant stem is ruled by spirals. This fact is called phyllotaxis and it is a nice example of how mathematics can describe patterns in nature. In this project, we will *invent* flowers using this fact.

This R project assumes you have familiarity with the `ggplot2` package. If you don't know `ggplot2` we recommend you take either of the courses Introduction to the Tidyverse or Data Visualization with ggplot2 (Part 1). If you want to see more examples of how you can use R to make art, you should check out the Fronkonstin blog created by Antonio Sánchez Chinchón.

Project Tasks

1. Patterns in nature
2. Warming up: drawing points on a circle
3. Make it harmonious with the Golden Angle
4. Remove everything unnecessary
5. A bit of makeup: size, color and transparency
6. Play with aesthetics: the dandelion
7. Put all it together: the sunflower
8. What if you modify the angle?
9. All together now: imaginary flowers

Antonio Sánchez Chinchón
Data Scientist at Telefónica

Antonio Sánchez Chinchón is mathematician and works as data scientist at Telefónica, where he tries hard to extract value from data every day. He is the creator of Fronkonstin, an unclassifiable blog of mathematical experiments, data science, data art and R programming. He plays the banjo in a rock band. You can follow him on Twitter @aschinchon.

See More

TECHNOLOGY

R

TOPICS

DATA VISUALIZATION

CASE STUDIES

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Building Dashboards with shinydashboard

Replay Course

4.4 (507 Ratings)

4 hours | 13 Videos | 45 Exercises | 10,290 Participants | 3,750 XP

Course Description

Once you've started learning tools for building interactive web applications with shiny, this course will translate this knowledge into building dashboards. Dashboards, a common data science deliverable, are pages that collate information, often tracking metrics from a live-updating data source. You'll gain more expertise using shiny while learning to build and design these dynamic dashboards. In the process, you'll pick up tips to optimize performance as well as best practices to create a visually appealing product.

1 Building Static Dashboards FREE — 100%

In this chapter you will learn the basic structure of a Shiny Dashboard and how to fill it with static content.

Dashboard structure overview	50 xp
Create empty Header, Sidebar, and Body	100 xp
Create an empty Shiny Dashboard	100 xp
Dashboard Header overview	50 xp
Create message menus	100 xp
Create notification menus	100 xp
Create task menus	100 xp
Dashboard Sidebar and Body overview	50 xp
Create Sidebar tabs	100 xp
Create Body tabs	100 xp
Create tab boxes	100 xp

HIDE CHAPTER DETAILS — Completed

2 Building Dynamic Dashboards — 100%

In this chapter you will learn how to add dynamic content to your Shiny Dashboard.

Reactive expression refresher	50 xp
Review selectInput and sliderInput	100 xp
Reactive expression practice	70 xp
Server-side dynamic how-to	50 xp
Read in real-time data	100 xp
View real-time data	100 xp
Optimizing performance	50 xp
How many times will this code run? (1)	50 xp
How many times will this code run? (2)	50 xp
Optimize this	100 xp
UI dynamic how-to	50 xp
Create reactive menu items	0 xp
Create reactive boxes	0 xp

HIDE CHAPTER DETAILS — Completed

3 Customizing Style — 100%

In this chapter you will focus on customizing the style of your Shiny Dashboard.

Bootstrap explanation	50 xp
Create body with row-based layout	100 xp
Create body with column-based layout	100 xp
Create body with mixed layout	100 xp
Customizing the appearance	50 xp
Change the appearance of the dashboard	100 xp
Customize the body with CSS	100 xp
Icons, statuses, colors	50 xp
Incorporate icons	100 xp
Add some life to your layouts	100 xp

HIDE CHAPTER DETAILS — Completed

4 Case Study — 100%

In this chapter you will participate in a case study, practicing the skills you have acquired in the previous chapters.

Introduction to the NASA fireball data set	50 xp
Examine the variables in the data set	100 xp
Create a value box for the maximum velocity	100 xp
Create a value box for the maximum impact	100 xp
Create a value box for the maximum energy	100 xp
Dynamic refresher	50 xp
Make the value boxes dynamic	100 xp
Allow the user to input an alert threshold	70 xp
Create a dynamic plot of the location of fireballs	100 xp
Update the look of your new application	100 xp
Thank you!	50 xp

HIDE CHAPTER DETAILS — Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Shiny Fundamentals with R

Lucy D'Agostino McGowan
Postdoctoral fellow, Johns Hopkins Department of Biostatistics

Lucy D'Agostino McGowan is a postdoctoral fellow in the Biostatistics Department at Johns Hopkins. She received her PhD in biostatistics from Vanderbilt University. Her research focuses primarily on causal inference techniques as well as data science pedagogy. She is an enthusiastic leader in the R-Ladies movement and active in the R community. Lucy has been using Shiny Dashboards to communicate about data, results, and advanced statistical models for many years, and is excited to pass on what she has learned!

See More

COLLABORATOR(S)

Chester Ismay

Nick Solomon

PREREQUISITES

Building Web Applications in R with Shiny

DATASETS

NASA fireball dataset

Starwars dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Advanced Deep Learning with Keras in Python

Replay Course

4.3 (1,040 Ratings)

4 hours | 13 Videos | 46 Exercises | 5,755 Participants | 3,950 XP

Course Description

This course shows you how to solve a variety of problems using the versatile Keras functional API. You will start with simple, multi-layer dense networks (also known as multi-layer perceptrons), and continue on to more complicated architectures. The course will cover how to build models with multiple inputs and a single output, as well as how to share weights between layers in a model. We will also cover advanced topics such as category embeddings and multiple-output networks. If you've ever wanted to train a network that does both classification and regression, then this course is for you!

1 The Keras Functional API FREE 100%

In this chapter, you'll become familiar with the basics of the Keras functional API. You'll build a simple functional network using functional building blocks, fit it to data, and make predictions.

Keras input and dense layers	50 xp
Input layers	100 xp
Dense layers	100 xp
Output layers	100 xp
Build and compile a model	50 xp
Build a model	100 xp
Compile a model	100 xp
Visualize a model	100 xp
Fit and evaluate a model	50 xp
Fit the model to the tournament basketball data	100 xp
Evaluate the model on a test set	100 xp

HIDE CHAPTER DETAILS ^ Completed

2 Two Input Networks Using Categorical Embeddings, Shared Layers, and Merge Layers 100%

In this chapter, you will build two-input networks that use categorical embeddings to represent high-cardinality data, shared layers to specify re-usable building blocks, and merge layers to join multiple inputs to a single output. By the end of this chapter, you will have the foundational building blocks for designing neural networks with complex data flows.

Category embeddings	50 xp
Define team lookup	100 xp
Define team model	100 xp
Shared layers	50 xp
Defining two inputs	100 xp
Lookup both inputs in the same model	100 xp
Merge layers	50 xp
Output layer using shared layer	100 xp
Model using two inputs and one output	100 xp
Predict from your model	50 xp
Fit the model to the regular season training data	100 xp
Evaluate the model on the tournament test data	100 xp

HIDE CHAPTER DETAILS ^ Completed

3 Multiple Inputs: 3 Inputs (and Beyond!) 100%

In this chapter, you will extend your 2-input model to 3 inputs, and learn how to use Keras' summary and plot functions to understand the parameters and topology of your neural networks. By the end of the chapter, you will understand how to extend a 2-input model to 3 inputs and beyond.

Three-input models	50 xp
Make an input layer for home vs. away	100 xp
Make a model and compile it	100 xp
Fit the model and evaluate	100 xp
Summarizing and plotting models	50 xp
Model summaries	100 xp
Plotting models	100 xp
Stacking models	50 xp
Add the model predictions to the tournament data	100 xp
Create an input layer with multiple columns	100 xp
Fit the model	100 xp
Evaluate the model	100 xp

HIDE CHAPTER DETAILS ^ Completed

4 Multiple Outputs 100%

In this chapter, you will build neural networks with multiple outputs, which can be used to solve regression problems with multiple targets. You will also build a model that solves a regression problem and a classification problem simultaneously.

Two-output models	50 xp
Simple two-output model	100 xp
Fit a model with two outputs	100 xp
Inspect the model (I)	100 xp
Evaluate the model	100 xp
Single model for classification and regression	50 xp
Classification and regression in one model	100 xp
Compile and fit the model	100 xp
Inspect the model (II)	100 xp
Evaluate on new data with two metrics	100 xp
Wrap-up	50 xp

HIDE CHAPTER DETAILS ^ Completed

STATEMENT OF ACCOMPLISHMENT

View in

Zachary Deane-Mayer

Automation First Data Scientist at DataRobot

Zach is a Data Scientist at DataRobot and co-author of the caret R package. He's fascinated by predicting the future and spends his free time competing in predictive modeling competitions. He's currently one of top 500 data scientists on Kaggle and took 9th place in the Heritage Health Prize as part of the Analytics Inside team.

See More

COLLABORATOR(S)

Sumedh Panchadhar

PREREQUISITES

Deep Learning in Python

DATASETS

Basketball data

Basketball models

© 2019 DataCamp Inc.

INTERACTIVE COURSE

String Manipulation in R with stringr

Replay Course

4.3 (908 Ratings)

4 hours | 17 Videos | 60 Exercises | 14,029 Participants | 5,850 XP

Course Description

Character strings can turn up in all stages of a data science project. You might have to clean messy string input before analysis, extract data that is embedded in text or automatically turn numeric results into a sentence to include in a report. Perhaps the strings themselves are the data of interest, and you need to detect and match patterns within them. This course will help you master these tasks by teaching you how to pull strings apart, put them back together and use stringr to detect, extract, match and split strings using regular expressions, a powerful way to express patterns.

1 String basics FREE 100%

You'll start with some basics: how to enter strings in R, how to control how numbers are transformed to strings, and finally how to combine strings together to produce output that combines text and nicely formatted numbers.

Welcome!	50 xp
Quotes	70 xp
What you see isn't always what you have	100 xp
Escape sequences	100 xp
Turning numbers into strings	50 xp
Using format() with numbers	100 xp
Controlling other aspects of the string	100 xp
formatC()	100 xp
Putting strings together	50 xp
Annotation of numbers	0 xp
A very simple table	100 xp
Let's order pizza!	70 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Introduction to stringr 100%

Time to meet stringr! You'll start by learning about some stringr functions that are very similar to some base R functions, then how to detect specific patterns in strings, how to split strings apart and how to find and replace parts of strings.

Introducing stringr	50 xp
Putting strings together with stringr	100 xp
String length	100 xp
Extracting substrings	100 xp
Hunting for matches	50 xp
Detecting matches	100 xp
Subsetting strings based on match	100 xp
Counting matches	100 xp
Splitting strings	50 xp
Parsing strings into variables	0 xp
Some simple text statistics	100 xp
Replacing matches in strings	50 xp
Replacing to tidy strings	0 xp
Review	100 xp
Final challenges	0 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Pattern matching with regular expressions 100%

In this chapter you'll learn about regular expressions, a language for describing patterns in strings. By combining regular expressions with the stringr functions you'll greatly increase your power to manipulate strings.

Regular expressions	50 xp
Matching the start or end of the string	0 xp
Matching any character	0 xp
Combining with stringr functions	0 xp
More regular expressions	50 xp
Alternation	0 xp
Character classes	0 xp
Repetition	0 xp
Shortcuts	50 xp
Hunting for phone numbers	0 xp
Extracting age and gender from accident narratives	0 xp
Parsing age and gender into pieces	0 xp

HIDE CHAPTER DETAILS ✓ Completed

4 More advanced matching and manipulation 100%

Now for two advanced ways to use regular expressions along with stringr: selecting parts of a match (a.k.a capturing) and referring back to parts of a match (a.k.a back-referencing). You'll also learn to deal with and strings or patterns that contain Unicode characters (e.g. é).

Capturing	50 xp
Capturing parts of a pattern	0 xp
Pulling out parts of a phone number	70 xp
Extracting age and gender again	0 xp
Backreferences	50 xp
Using backreferences in patterns	0 xp
Replacing with regular expressions	100 xp
Replacing with backreferences	100 xp
Unicode and pattern matching	50 xp
Matching a specific code point or code groups	100 xp
Matching a single grapheme	0 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Case studies 100%

Practice your string manipulation skills on a couple of case studies. You'll also learn a few new skills, reading strings into R and handling problems of case (e.g. A versus a).

A case study, reading a play	50 xp
Getting the play into R	0 xp
Identifying the lines, take 1	0 xp
Identifying the lines, take 2	0 xp
A case study on case	50 xp
Changing case to ease matching	0 xp
Ignoring case when matching	0 xp
Fixing case problems	100 xp
Wrapping up	50 xp
An interview with Hadley Wickham	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

R Programmer

Text Mining with R

Charlotte Wickham
Assistant Professor at Oregon State University

Charlotte is an Assistant Professor in the Department of Statistics at Oregon State University and an avid R programmer with a passion for teaching. Her interests lie in spatiotemporal data, statistical graphics and computing, and environmental statistics.

See More

COLLABORATOR(S)

Tom Jeon

Richie Cotton

PREREQUISITES

Introduction to R

Intermediate R

DATASETS

DNA sequences from the genome of Yersinia pestis

Narratives

Adverbs

Importance of being earnest

Cat-related accidents

© 2019 DataCamp Inc.

INTERACTIVE COURSE

ChIP-seq Workflows in R

Replay Course

4.2 (56 Ratings)

4 hours | 13 Videos | 46 Exercises | 864 Participants | 3,650 XP

Course Description

ChIP-seq analysis is an important branch of bioinformatics. It provides a window into the machinery that makes the cells in our bodies tick. Whether it is a brain cell helping you to read this web page or an immune cell patrolling your body for microorganisms that would make you sick, they all carry the same genome. What differentiates them are the genes that are active at any given time. Which genes these are is determined by a complex system of proteins that can activate and deactivate genes. When this regulatory machinery gets out of control, it can lead to cancer and other debilitating diseases. ChIP-seq analysis allows us to understand the function of regulatory proteins, how they can contribute to disease and can provide insights into how we may be able to intervene to prevent cells from spinning out of control. In this course, you will explore a real dataset while learning how to process and analyze ChIP-seq data in R.

STATEMENT OF ACCOMPLISHMENT

View

1 Introduction to ChIP-seq FREE

100%

Introduction to ChIP-seq experiments. Why are they interesting? What sort of phenomena can be studied with ChIP-seq and what can we learn from these experiments.

What is ChIP-seq?	50 xp
ChIP-seq recap	50 xp
Sequencing data	100 xp
Peak calls	100 xp
ChIP-seq Workflow	50 xp
Heat map	100 xp
Genes that make a difference	100 xp
ChIP-seq results summary	50 xp
Understanding ChIP-seq data	50 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Back to Basics - Preparing ChIP-seq data

100%

Now the ChIP-seq analysis begins in earnest. This chapter introduces Bioconductor tools to import and clean the data.

Importing data	50 xp
Reading BAM files	100 xp
Reading BED files	100 xp
Taking a closer look at peaks	50 xp
Plotting a region in detail	100 xp
Adding Annotations	50 xp
Cleaning ChIP-seq data	50 xp
Removing blacklisted regions	100 xp
Filtering reads	100 xp
Compare filtered data to raw reads	50 xp
Assessing enrichment	50 xp
Computing coverage	100 xp
Peaks vs background	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Comparing ChIP-seq samples

100%

This chapter introduces techniques to identify and visualise differences between ChIP-seq samples.

Introduction to differential binding	50 xp
Do these samples look the same to you?	50 xp
Clustering samples	100 xp
Visualising differences in protein binding	100 xp
Testing for differential binding	50 xp
Loading Read Counts	50 xp
Setting-up the model	100 xp
Fitting the model	100 xp
Revisiting PCA and Heat map	100 xp
A closer look at differential binding	50 xp
MA plot	100 xp
Volcano plot	100 xp
Summarising differences in protein binding	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 From Peaks to Genes to Function

100%

Being able to identify differential binding between groups of samples is great, but what does it mean? This chapter discusses strategies to interpret differential binding results to go from peak calls to biologically meaningful insights.

Interpreting ChIP-seq peaks	50 xp
Consolidating peaks	100 xp
Using Annotations	100 xp
Annotating peaks	100 xp
Which peaks are different?	100 xp
Interpreting Gene lists	50 xp
Associating peaks with genes	100 xp
Finding common themes	100 xp
Understanding the impact on pathways	0 xp
A closer look at pathways	100 xp
Advanced ChIP-seq analyses	50 xp

HIDE CHAPTER DETAILS ✓ Completed

Peter Humburg
Statistician

Peter Humburg has extensive experience in the analysis of genomic data as a Bioinformatician. His doctoral work focused on the development of statistical methods for the analysis of ChIP-seq data. Peter now works as a Statistician at Macquarie University in Sydney, where he advises researchers on diverse aspects of data analysis and teaches R and Python as a Carpentries Instructor.

See More

COLLABORATOR(S)

Richie Cotton
David Campos
Shon Inouye

PREREQUISITES

Intermediate R
Introduction to Bioconductor

DATASETS

Androgen Receptor ChIP-seq Peaks dataset
Chromosome 20 dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Cleaning Data in R

Practice Now | Replay Course

4.6 (5,490 Ratings)

4 hours | 15 Videos | 58 Exercises | 90,553 Participants | 4,700 XP

Course Description

It's commonly said that data scientists spend 80% of their time cleaning and manipulating data and only 20% of their time actually analyzing it. For this reason, it is critical to become familiar with the data cleaning process and all of the tools available to you along the way. This course provides a very basic introduction to cleaning data in R using the tidyr, dplyr, and stringr packages. After taking the course you'll be able to go from raw data to awesome insights as quickly and painlessly as possible!

1 Introduction and exploring raw data FREE

100%

This chapter will give you an overview of the process of data cleaning with R, then walk you through the basics of exploring raw data.

Introduction to Cleaning Data in R	50 xp
The data cleaning process	50 xp
Here's what messy data look like	100 xp
Here's what clean data look like	100 xp
Exploring raw data	50 xp
Getting a feel for your data	100 xp
Viewing the structure of your data	100 xp
Exploring raw data (part 2)	50 xp
Looking at your data	100 xp
Visualizing your data	100 xp

HIDE CHAPTER DETAILS | ✓ Completed

2 Tidying data

100%

This chapter will give you an overview of the principles of tidy data, how to identify messy data, and what to do about it.

Introduction to tidy data	50 xp
Principles of tidy data	50 xp
Common symptoms of messy data	50 xp
Introduction to tidyr	50 xp
What kind of messy are the BMI data?	50 xp
Gathering columns into key-value pairs	100 xp
Spreading key-value pairs into columns	100 xp
Introduction to tidyr (part 2)	50 xp
Functions in tidyr	50 xp
Separating columns	100 xp
Uniting columns	100 xp
Column headers are values, not variable names	70 xp
Variables are stored in both rows and columns	100 xp
Multiple values are stored in one column	100 xp

HIDE CHAPTER DETAILS | ✓ Completed

3 Preparing data for analysis

100%

This chapter will teach you how to prepare your data for analysis. We will look at type conversion, string manipulation, missing and special values, and outliers and obvious errors.

Type conversions	50 xp
Types of variables in R	0 xp
Common type conversions	100 xp
Working with dates	100 xp
String manipulation	50 xp
Trimming and padding strings	100 xp
Upper and lower case	100 xp
Finding and replacing strings	100 xp
Missing and special values	50 xp
Types of missing and special values in R	50 xp
Finding missing values	100 xp
Dealing with missing values	100 xp
Outliers and obvious errors	50 xp
Identifying outliers and obvious errors	50 xp
Dealing with outliers and obvious errors	100 xp
Another look at strange values	100 xp

HIDE CHAPTER DETAILS | ✓ Completed

4 Putting it all together

100%

In this chapter, you will practice everything you've learned from the first three chapters in order to clean a messy dataset using R.

Time to put it all together!	50 xp
Get a feel for the data	100 xp
Summarize the data	100 xp
Take a closer look	100 xp
Let's tidy the data	50 xp
Column names are values	70 xp
Values are variable names	100 xp
Prepare the data for analysis	50 xp
Clean up dates	100 xp
A closer look at column types	100 xp
Column type conversions	100 xp
Missing, extreme, and unexpected values	50 xp
Find missing values	100 xp
An obvious error	70 xp
Another obvious error	0 xp
Check other extreme values	100 xp
Finishing touches	100 xp
Your data are clean!	50 xp

HIDE CHAPTER DETAILS | ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View | in

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Importing & Cleaning Data with R

Nick Carchedi

Product Manager at DataCamp

Nick is a Product Manager at DataCamp, responsible for the learning and teaching experience. Prior to joining DataCamp, he earned his master's degree at Johns Hopkins Biostatistics and worked as a data scientist for McKinsey. Nick's passion for teaching data science began in graduate school, where he was heavily involved in tutoring fellow students, developing the Johns Hopkins Data Science Specialization, and building the swirl R package.

See More

COLLABORATOR(S)

Jeff Paadre

PREREQUISITES

Introduction to R

DATASETS

- Messy weather data
- BMI data
- Census data
- Student data (with dates)

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Manipulation in R with dplyr

Practice Now | Replay Course

This course has been archived. If you are taking this course you will be able to finish it. If you are working on a track containing this course it will remain available until you complete the track.

4 hours | 11 Videos | 42 Exercises | 104,336 Participants | 3,400 XP

Course Description

In this interactive tutorial, you will learn how to perform sophisticated dplyr techniques to carry out your data manipulation with R. First you will master the five verbs of R data manipulation with dplyr: select, mutate, filter, arrange and summarise. Next, you will learn how you can chain your dplyr operations using the pipe operator of the magrittr package. In the final section, the focus is on practicing how to subset your data using the group_by function, and how you can access data stored outside of R in a database. All said and done, you will be familiar with data manipulation tools and techniques that will allow you to efficiently manipulate data.

STATEMENT OF ACCOMPLISHMENT

View

Team RStudio

Instructor

COLLABORATOR(S)

Filip Schouwenaars

PREREQUISITES

Introduction to R

Intermediate R

1 Introduction to dplyr and tbls FREE — 100%

Introduction to the dplyr package and the tbl class. Learn the philosophy that guides dplyr, discover some useful applications of the dplyr package, and meet the data structures that dplyr uses behind the scenes.

Section 1 - Introduction to dplyr	50 xp
Load the dplyr and hflights package	100 xp
Explore the data set	50 xp
Section 2 - tbl, a special type of data.frame	50 xp
Convert data.frame to tibble	100 xp
Changing labels of hflights, part 1 of 2	100 xp
Changing labels of hflights, part 2 of 2	100 xp

HIDE CHAPTER DETAILS — Completed

2 Select and mutate — 100%

Get familiar with dplyr's manipulation verbs. Meet the five verbs and then practice using the mutate and select verbs.

Section 3 - The five verbs and select in more detail	50 xp
The five verbs and their meaning	50 xp
Choosing is not losing! The select verb	100 xp
Helper functions for variable selection	100 xp
Comparison to base R	100 xp
Section 4 - The second of five verbs: mutate	50 xp
Mutating is creating	100 xp
Add multiple variables using mutate	100 xp
Recap on mutate and select	50 xp

HIDE CHAPTER DETAILS — Completed

3 Filter and arrange — 100%

Learn how to search through the observations in your data set (and extract useful observations) with the filter function. Rearrange the observations in your data set with the arrange verb.

Section 5 - The third of five verbs: filter	50 xp
Logical operators	100 xp
Combining tests using boolean operators	100 xp
Blend together what you've learned!	100 xp
Recap on select, mutate and filter	35 xp
Section 6 - Almost there: the arrange verb	50 xp
Arranging your data	100 xp
Reverse the order of arranging	100 xp
Recap on select, mutate, filter and arrange	50 xp

HIDE CHAPTER DETAILS — Completed

4 Summarize and the pipe operator — 100%

Master the data manipulation verb summarize, and practice combining the five verbs to solve advanced data manipulation tasks. Learn to chain the operators together with the piping operator.

Section 7 - Last but not least: summarize	50 xp
The syntax of summarize	70 xp
Aggregate functions	100 xp
dplyr aggregate functions	100 xp
Section 8 - Chaining your functions: the pipe operator	50 xp
Overview of syntax	100 xp
Drive or fly? Part 1 of 2	100 xp
Drive or fly? Part 2 of 2	100 xp
Advanced piping exercise	100 xp

HIDE CHAPTER DETAILS — Completed

5 Group_by and working with databases — 100%

Complete your mastery of data manipulation with group-wise operations and databases. Learn to use group_by to group your data into subsets of observations, and use dplyr to access data stored outside of R in a database.

Section 9 - Get group-wise insights: group_by	50 xp
Unite and conquer using group_by	100 xp
Combine group_by with mutate	100 xp
Advanced group_by exercises	70 xp
Section 10 - dplyr and databases	50 xp
dplyr deals with different types	100 xp
dplyr and mySQL databases	100 xp
Talk with Hadley Wickham	50 xp

HIDE CHAPTER DETAILS — Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Importing & Cleaning Data in R: Case Studies

Replay Course

4.5 (2,136 Ratings)

4 hours | 0 Videos | 35 Exercises | 30,465 Participants | 3,500 XP

Course Description

Running exciting analyses on interesting datasets is the dream of every data scientist. But first, some importing and cleaning must be done. In this series of four case studies, you'll revisit key concepts from our courses on importing and cleaning data in R.

1 Ticket Sales Data FREE

100%

Hone your skills by importing and cleaning some wonderfully messy online ticket sales data.

Importing the data	100 xp
Examining the data	100 xp
Summarizing the data	100 xp
Removing redundant info	100 xp
Information not worth keeping	70 xp
Separating columns	70 xp
Dealing with warnings	100 xp
Identifying dates	0 xp
More warnings!	100 xp
Combining columns	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 MBTA Ridership Data

100%

Boston's public transit system needs your help! The T wants to do some data analysis and you've been asked to clean their ridership data.

Using readxl	100 xp
Examining the data	100 xp
Removing unnecessary rows and columns	100 xp
Observations are stored in columns	100 xp
Type conversions	100 xp
Variables are stored in both rows and columns	100 xp
Separating columns	100 xp
Do your values seem reasonable?	100 xp
Dealing with entry error	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 World Food Facts

100%

We all know that you are what you eat, so what exactly are you? In this chapter, you'll import and clean some data about food products from around the world.

Importing the data	70 xp
Examining the data	100 xp
Inspecting variables	100 xp
Removing duplicate info	100 xp
Removing useless info	100 xp
Finding columns	0 xp
Replacing missing values	70 xp
Dealing with messy data	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 School Attendance Data

100%

Use all of the tools you've learned to import and clean a gnarly dataset containing information on average school attendance in the US.

Importing the data	100 xp
Examining the data	100 xp
Removing unnecessary rows	70 xp
Removing useless columns	100 xp
Splitting the data	100 xp
Replacing the names	100 xp
Cleaning up extra characters	100 xp
Some final type conversions	100 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

This course is part of these tracks:

Importing & Cleaning Data with R

Nick Carchedi

Product Manager at DataCamp

Nick is a Product Manager at DataCamp, responsible for the learning and teaching experience. Prior to joining DataCamp, he earned his master's degree at Johns Hopkins Biostatistics and worked as a data scientist for McKinsey. Nick's passion for teaching data science began in graduate school, where he was heavily involved in tutoring fellow students, developing the Johns Hopkins Data Science Specialization, and building the swirl R package.

See More

COLLABORATOR(S)

Jim Looney

Tom Jeon

PREREQUISITES

- Importing Data in R (Part 1)
- Importing Data in R (Part 2)
- Cleaning Data in R

DATASETS

- Sales
- MBTA
- Food
- Attendance

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Convolutional Neural Networks for Image Processing

Replay Course

4.5 (897 Ratings)

4 hours | 13 Videos | 45 Exercises | 7,713 Participants | 3,650 XP

Course Description

Deep learning methods use data to train neural network algorithms to do a variety of machine learning tasks, such as classification of different classes of objects. Convolutional neural networks are deep learning algorithms that are particularly powerful for analysis of images. This course will teach you how to construct, train and evaluate convolutional neural networks. You will also learn how to improve their ability to learn from data, and how to interpret the results of the training.

STATEMENT OF ACCOMPLISHMENT

View | in

1 Image Processing With Neural Networks FREE 100%

Convolutional neural networks use the data that is represented in images to learn. In this chapter, we will probe data in images, and we will learn how to use Keras to train a neural network to classify objects that appear in images.

Introducing convolutional neural networks	✓ 50 xp
Images as data: visualizations	✓ 100 xp
Images as data: changing images	✓ 70 xp
Classifying images	✓ 50 xp
Using one-hot encoding to represent images	✓ 100 xp
Evaluating a classifier	✓ 100 xp
Classification with Keras	✓ 50 xp
Build a neural network	✓ 100 xp
Compile a neural network	✓ 100 xp
Fitting a neural network model to clothing data	✓ 100 xp
Cross-validation for neural network evaluation	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Using Convolutions 100%

Convolutions are the fundamental building blocks of convolutional neural networks. In this chapter, you will be introducted to convolutions and learn how they operate on image data. You will also see how you incorporate convolutions into Keras neural networks.

Convolutions	✓ 50 xp
One dimensional convolutions	✓ 100 xp
Image convolutions	✓ 0 xp
Defining image convolution kernels	✓ 91 xp
Implementing image convolutions in Keras	✓ 50 xp
Convolutional network for image classification	✓ 100 xp
Training a CNN to classify clothing types	✓ 100 xp
Evaluating a CNN with test data	✓ 100 xp
Tweaking your convolutions	✓ 50 xp
Add padding to a CNN	✓ 100 xp
Add strides to a convolutional network	✓ 100 xp
Calculate the size of convolutional layer output	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Going Deeper 100%

Convolutional neural networks gain a lot of power when they are constructed with multiple layers (deep networks). In this chapter, you will learn how to stack multiple convolutional layers into a deep network. You will also learn how to keep track of the number of parameters, as the network grows, and how to control this number.

Going deeper	✓ 50 xp
Creating a deep learning network	✓ 100 xp
Train a deep CNN to classify clothing images	✓ 100 xp
What is special about a deep network?	✓ 50 xp
How many parameters?	✓ 50 xp
How many parameters in a CNN?	✓ 50 xp
How many parameters in a deep CNN?	✓ 100 xp
Pooling operations	✓ 50 xp
Write your own pooling operation	✓ 100 xp
Keras pooling layers	✓ 100 xp
Train a deep CNN with pooling to classify images	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Understanding and Improving Deep Convolutional Networks 100%

There are many ways to improve training by neural networks. In this chapter, we will focus on our ability to track how well a network is doing, and explore approaches towards improving convolutional neural networks.

Tracking learning	✓ 50 xp
Plot the learning curves	✓ 100 xp
Using stored weights to predict in a test set	✓ 100 xp
Regularization	✓ 50 xp
Adding dropout to your network	✓ 100 xp
Add batch normalization to your network	✓ 100 xp
Interpreting the model	✓ 50 xp
Extracting a kernel from a trained network	✓ 100 xp
Shape of the weights	✓ 50 xp
Visualizing kernel responses	✓ 100 xp
Next steps	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

Ariel Rokem
Senior Data Scientist, University of Washington

Ariel Rokem is a Data Scientist at the University of Washington eScience Institute. He received a PhD in neuroscience from UC Berkeley, and postdoctoral training in computational neuroimaging at Stanford. In his work, he develops data science algorithms and tools, and applies them to analysis of neural data. He is also a contributor to multiple open-source software projects in the scientific Python ecosystem.

See More

COLLABORATOR(S)

Lore Dirick
Eunkyung Park
Sumedh Panchadhar

PREREQUISITES

Deep Learning in Python

DATASETS

Shutterstock straight

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Privacy and Anonymization in R

Replay Course

4.2 (305 Ratings)

4 hours | 13 Videos | 45 Exercises | 1,774 Participants | 3,650 XP

Course Description

With social media and big data everywhere, data privacy has been a growing, public concern. Recognizing this issue, entities such as Google, Apple, and the US Census Bureau are promoting better privacy techniques; specifically differential privacy, a mathematical condition that quantifies privacy risk. In this course, you will learn to code basic data privacy methods and a differentially private algorithm based on various differentially private properties. With these tools in hand, you will learn how to generate a basic synthetic (fake) data set with the differential privacy guarantee for public data release.

1 Introduction to Data Privacy FREE

100%

This chapter covers some basic data privacy techniques that statisticians use to anonymize data. You'll first learn how to remove identifiers and then generate synthetic data from probability distributions.

Intro to Anonymization (I)	✓ 50 xp
Removing Names	✓ 100 xp
Rounding Salaries	✓ 100 xp
Intro to Anonymization (II)	✓ 50 xp
Generalization	✓ 100 xp
Bottom Coding	✓ 0 xp
summarize_at()	✓ 100 xp
count()	✓ 100 xp
Data Synthesis	✓ 50 xp
Binomial Distribution	✓ 100 xp
Normal Distribution	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Introduction to Differential Privacy

100%

After covering the basic data privacy techniques, you'll learn conceptually about differential privacy as well as how to implement the most popular and common differentially private algorithm called the Laplace mechanism.

Differential Privacy	✓ 50 xp
Zero Privacy Budget	✓ 50 xp
Changing the Privacy Budget	✓ 50 xp
Less Privacy Protection	✓ 50 xp
Less Information Leakage	✓ 50 xp
Global Sensitivity	✓ 50 xp
Sensitivity of Counting and Proportion Queries	✓ 100 xp
Sensitivity of Mean and Variance Queries	✓ 100 xp
Laplace mechanism	✓ 50 xp
Counting Query	✓ 100 xp
Proportion Query	✓ 100 xp
Mean and Variance Queries	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Differentially Private Properties

100%

In this chapter, you will learn the various properties of differential privacy, such as the combination rules and post-processing, to properly implement the Laplace mechanism for various kinds data questions.

Sequential Composition	✓ 50 xp
Value of Epsilon per the Sequential Composition	✓ 100 xp
Sequential Composition and Laplace mechanism	✓ 100 xp
Parallel Composition	✓ 50 xp
Value of Epsilon per the Parallel Composition	✓ 100 xp
Parallel Composition and Laplace mechanism	✓ 100 xp
Post-processing	✓ 50 xp
Setting Epsilon and Global Sensitivity	✓ 70 xp
Post-processing and Laplace mechanism	✓ 100 xp
Impossible and Inconsistent Answers	✓ 50 xp
Inconsistent answers	✓ 100 xp
Normalizing noise	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Differentially Private Data Synthesis

100%

In this chapter, you will learn how to release simple data sets publicly using differentially private data synthesis techniques.

Laplace Sanitizer	✓ 50 xp
Prepping for the Laplace sanitizer	✓ 100 xp
Applying the Laplace sanitizer	✓ 100 xp
Normalizing Noise and Generating Synthetic Data	✓ 50 xp
Parametric Approaches	✓ 50 xp
Prepping for the DP Binomial Distribution	✓ 100 xp
Generating Binomial Synthetic Data	✓ 100 xp
Prepping for the DP Normal Distribution	✓ 100 xp
Generating Normal Synthetic Data	✓ 100 xp
Wrap-up	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

This course is part of these tracks:

R Programmer

Claire Bowen

Postdoctoral Researcher at the Los Alamos National Laboratory

Claire McKay Bowen is a Postdoctoral Researcher in the Statistical Science Group at the Los Alamos National Laboratory. She conducts research in uncertainty quantification with physics-informed Bayesian Model updating and data privacy via differentially private data synthesis methods. Her other interests include statistical computing, scientific communication, and STEM outreach.

See More

COLLABORATOR(S)

Chester Ismay

Sumedh Panchadhar

PREREQUISITES

Intermediate R

Foundations of Probability in R

Introduction to the Tidyverse

DATASETS

Data sets

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Exploratory Data Analysis

Practice Now | Replay Course

4.5 (3,147 Ratings)

4 hours | 15 Videos | 54 Exercises | 33,947 Participants | 3,950 XP

Course Description

When your dataset is represented as a table or a database, it's difficult to observe much about it beyond its size and the types of variables it contains. In this course, you'll learn how to use graphical and numerical techniques to begin uncovering the structure of your data. Which variables suggest interesting relationships? Which observations are unusual? By the end of the course, you'll be able to answer these questions and more, while generating graphics that are both insightful and beautiful.

1 Exploring Categorical Data FREE — 100%

In this chapter, you will learn how to create graphical and numerical summaries of two categorical variables.

Exploring categorical data	50 xp
Bar chart expectations	50 xp
Contingency table review	100 xp
Dropping levels	100 xp
Side-by-side barcharts	100 xp
Bar chart interpretation	50 xp
Counts vs. proportions	50 xp
Conditional proportions	50 xp
Counts vs. proportions (2)	100 xp
Distribution of one variable	50 xp
Marginal barchart	100 xp
Conditional barchart	100 xp
Improve piechart	100 xp

HIDE CHAPTER DETAILS ^ Completed

2 Exploring Numerical Data — 100%

In this chapter, you will learn how to graphically summarize numerical data.

Exploring numerical data	50 xp
Faceted histogram	100 xp
Boxplots and density plots	100 xp
Compare distribution via plots	50 xp
Distribution of one variable	50 xp
Marginal and conditional histograms	100 xp
Marginal and conditional histograms interpretation	50 xp
Three binwidths	100 xp
Three binwidths interpretation	50 xp
Box plots	50 xp
Box plots for outliers	100 xp
Plot selection	100 xp
Visualization in higher dimensions	50 xp
3 variable plot	100 xp
Interpret 3 var plot	50 xp

HIDE CHAPTER DETAILS ^ Completed

3 Numerical Summaries — 100%

Now that we've looked at exploring categorical and numerical data, you'll learn some useful statistics for describing distributions of data.

Measures of center	50 xp
Choice of center measure	50 xp
Calculate center measures	100 xp
Measures of variability	50 xp
Choice of spread measure	50 xp
Calculate spread measures	100 xp
Choose measures for center and spread	100 xp
Shape and transformations	50 xp
Describe the shape	50 xp
Transformations	100 xp
Outliers	50 xp
Identify outliers	100 xp

HIDE CHAPTER DETAILS ^ Completed

4 Case Study — 100%

Apply what you've learned to explore and summarize a real-world dataset in this case study of email spam.

Introducing the data	50 xp
Spam and num_char	100 xp
Spam and num_char interpretation	50 xp
Spam and !!!	0 xp
Spam and !!! interpretation	50 xp
Check-in 1	50 xp
Collapsing levels	100 xp
Image and spam interpretation	50 xp
Data Integrity	100 xp
Answering questions with chains	100 xp
Check-in 2	50 xp
What's in a number?	100 xp
What's in a number interpretation	50 xp
Conclusion	50 xp

HIDE CHAPTER DETAILS ^ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Statistics Fundamentals with R

Andrew Bray

Assistant Professor of Statistics at Reed College

Andrew Bray is an Assistant Professor of Statistics at Reed College and lover of all things statistics and R.

See More

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

- Introduction to R
- Introduction to Data

DATASETS

- Cars data
- Comics data
- Immigration data
- Raw life expectancy data
- Names data
- Raw U.S. income data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Correlation and Regression

Practice Now | Replay Course

4.5 (3,800 Ratings)

4 hours | 18 Videos | 58 Exercises | 40,635 Participants | 4,200 XP

Course Description

Ultimately, data analysis is about understanding relationships among variables. Exploring data with multiple variables requires new, more complex tools, but enables a richer set of comparisons. In this course, you will learn how to describe relationships between two numerical quantities. You will characterize these relationships graphically, in the form of summary statistics, and through simple linear regression models.

1 Visualizing two variables FREE

100%

In this chapter, you will learn techniques for exploring bivariate relationships.

Visualizing bivariate relationships	50 xp
Scatterplots	100 xp
Boxplots as discretized/conditioned scatterplots	100 xp
Characterizing bivariate relationships	50 xp
Creating scatterplots	100 xp
Characterizing scatterplots	50 xp
Transformations	100 xp
Outliers	50 xp
Identifying outliers	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Correlation

100%

This chapter introduces correlation as a means of quantifying bivariate relationships.

Quantifying the strength of bivariate relationships	50 xp
Understanding correlation scale	50 xp
Understanding correlation sign	50 xp
Computing correlation	100 xp
The Anscombe dataset	50 xp
Exploring Anscombe	100 xp
Perception of correlation	50 xp
Perception of correlation (2)	100 xp
Interpretation of correlation	50 xp
Interpreting correlation in context	50 xp
Correlation and causation	50 xp
Spurious correlation	50 xp
Spurious correlation in random data	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Simple linear regression

100%

With the notion of correlation under your belt, we'll now turn our attention to simple linear models in this chapter.

Visualization of linear models	50 xp
The "best fit" line	100 xp
Uniqueness of least squares regression line	100 xp
Understanding the linear model	50 xp
Regression model terminology	50 xp
Regression model output terminology	50 xp
Fitting a linear model "by hand"	100 xp
Regression vs. regression to the mean	50 xp
Regression to the mean	100 xp
"Regression" in the parlance of our time	50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Interpreting regression models

100%

This chapter looks at how to interpret the coefficients in a regression model.

Interpretation of regression coefficients	50 xp
Interpretation of coefficients	50 xp
Interpretation in context	50 xp
Fitting simple linear models	100 xp
Units and scale	50 xp
Your linear model object	50 xp
The lm summary output	100 xp
Fitted values and residuals	100 xp
Tidying your linear model	100 xp
Using your linear model	50 xp
Making predictions	100 xp
Adding a regression line to a plot manually	100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Model Fit

100%

In this final chapter, you'll learn how to assess the "fit" of a simple linear regression model.

Assessing model fit	50 xp
RMSE	50 xp
Standard error of residuals	100 xp
Comparing model fits	50 xp
Assessing simple linear model fit	100 xp
Interpretation of R^2	50 xp
Linear vs. average	100 xp
Unusual points	50 xp
Leverage	70 xp
Influence	100 xp
Dealing with unusual points	50 xp
Removing outliers	100 xp
High leverage points	100 xp
Conclusion	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Statistics Fundamentals with R

Ben Baumer

Assistant Professor at Smith College

Ben is an Assistant Professor in the Statistical & Data Sciences Program at Smith College. He completed his Ph.D. in Mathematics at the Graduate Center of the City University of New York in 2012. He is an Accredited Professional Statistician™ by the American Statistical Association and was previously the Statistical Analyst for the Baseball Operations department of the New York Mets.

See More

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

- Introduction to R
- Introduction to Data
- Exploratory Data Analysis

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Importing Data in R (Part 1)

Practice Now | Replay Course

4.5 (6,493 Ratings)

3 hours | 11 Videos | 42 Exercises | 87,185 Participants | 3,550 XP

Course Description

Importing data into R to start your analyses—it should be the easiest step. Unfortunately, this is almost never the case. Data come in all sorts of formats, ranging from CSV and text files and statistical software files to databases and HTML data. Knowing which approach to use is key to getting started with the actual analysis.

In this course, you will get started with learning how to read CSV and text files in R. You will then cover the readr and data.table packages to easily and efficiently import flat file data. After that you will learn how to read XLS files in R using readxl and gdata.

1 Importing data from flat files with utils FREE — 100%

Lots of data comes in the form of flat files: simple tabular text files. Learn how to import all common formats of flat file data with base R functions.

Introduction & read.csv	50 xp
read.csv	100 xp
stringsAsFactors	100 xp
Any changes?	50 xp
read.delim & read.table	50 xp
read.delim	100 xp
read.table	100 xp
Arguments	100 xp
Column classes	100 xp
Final Thoughts	50 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 readr & data.table — 100%

Next to base R, there are also dedicated packages to easily and efficiently import flat file data. We'll talk about two such packages: readr and data.table.

readr: read_csv & read_tsv	50 xp
read_csv	100 xp
read_tsv	100 xp
readr: read_delim	50 xp
read_delim	100 xp
skip and n_max	100 xp
col_types	100 xp
col_types with collectors	100 xp
data.table: fread	50 xp
fread	100 xp
fread: more advanced use	100 xp
Dedicated classes	50 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Importing Excel data — 100%

Excel is a very widely used data analysis tool. If you prefer to do your analyses in R, though, you'll need an understanding of importing CSV data into R. This chapter will show you how to use readxl and gdata to do so.

readxl (1)	50 xp
List the sheets of an Excel file	100 xp
Import an Excel sheet	100 xp
Reading a workbook	100 xp
readxl (2)	50 xp
The col_names argument	100 xp
The skip argument	100 xp
gdata	50 xp
Import a local file	100 xp
read.xls() wraps around read.table()	100 xp
Work that Excel data!	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 Reproducible Excel work with XLConnect — 100%

Next to importing data from Excel, you can take things one step further with XLConnect. Learn all about it and bridge the gap between R and Excel!

Reading sheets	50 xp
Connect to a workbook	100 xp
List and read Excel sheets	100 xp
Customize readWorksheet	100 xp
Adapting sheets	50 xp
Add worksheet	100 xp
Populate worksheet	100 xp
Renaming sheets	100 xp
Removing sheets	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View | in

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Importing & Cleaning Data with R

Filip Schouwenaars
Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

PREREQUISITES

- Introduction to R

DATASETS

- Hotdogs
- Potatoes (CSV)
- Potatoes (TSV)
- Swimming pools
- Urban population (XLS)
- Urban population (XLSX)

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Differential Expression Analysis in R with limma

Replay Course

4.3 (141 Ratings)

4 hours | 15 Videos | 47 Exercises | 1,337 Participants | 3,900 XP

Course Description

Functional genomic technologies like microarrays, sequencing, and mass spectrometry enable scientists to gather unbiased measurements of gene expression levels on a genome-wide scale. Whether you are generating your own data or want to explore the large number of publicly available data sets, you will first need to learn how to analyze these types of experiments. In this course, you will be taught how to use the versatile R/Bioconductor package limma to perform a differential expression analysis on the most common experimental designs. Furthermore, you will learn how to pre-process the data, identify and correct for batch effects, visually assess the results, and perform enrichment testing. After completing this course, you will have general analysis strategies for gaining insight from any functional genomics study.

1 Differential Expression Analysis FREE 100%

To begin, you'll review the goals of differential expression analysis, manage gene expression data using R and Bioconductor, and run your first differential expression analysis with limma.

Differential expression analysis	50 xp
Applications of differential expression analysis	50 xp
Differential expression data	50 xp
Create a boxplot	100 xp
The ExpressionSet class	50 xp
Create an ExpressionSet object	100 xp
Create a boxplot with an ExpressionSet object	100 xp
The limma package	50 xp
Specify a linear model to compare 2 groups	100 xp
Test for differential expression between 2 groups	100 xp

HIDE CHAPTER DETAILS ^ Completed

2 Flexible Models for Common Study Designs 100%

In this chapter, you'll learn how to construct linear models to test for differential expression for common experimental designs.

Flexible linear models	50 xp
Design matrix for group-means model	100 xp
Contrasts matrix for group-means	100 xp
Test for differential expression for group-means	100 xp
Studies with more than two groups	50 xp
Design matrix for 3 groups	100 xp
Contrasts matrix for 3 groups	100 xp
Test for differential expression for 3 groups	100 xp
Factorial experimental design	50 xp
Design matrix for 2x2 factorial	100 xp
Contrasts matrix for 2x2 factorial	100 xp
Test for differential expression for 2x2 factorial	100 xp

HIDE CHAPTER DETAILS ^ Completed

3 Pre- and post-processing 100%

Now that you've learned how to perform differential expression tests, next you'll learn how to normalize and filter the feature data, check for technical batch effects, and assess the results.

Normalizing and filtering	50 xp
Normalize	100 xp
Filter genes	100 xp
Accounting for technical batch effects	50 xp
Visualize batch effects	100 xp
Remove batch effects	0 xp
Visualizing the results	50 xp
Histogram of p-values	100 xp
Volcano plot	100 xp
Enrichment testing	50 xp
KEGG pathways	100 xp
Gene ontology categories	100 xp

HIDE CHAPTER DETAILS ^ Completed

4 Case Study: Effect of Doxorubicin Treatment 100%

In this final chapter, you'll use your new skills to perform an end-to-end differential expression analysis of a study that uses a factorial design to assess the impact of the cancer drug doxorubicin on the hearts of mice with different genetic backgrounds.

Pre-process the data	50 xp
Pre-process features	100 xp
Boxplot of Top2b	100 xp
Check sources of variation	100 xp
Model the data	50 xp
Design matrix	100 xp
Contrasts matrix	100 xp
Test for differential expression	100 xp
Inspect the results	50 xp
Histogram of p-values	100 xp
Volcano plot	100 xp
Pathway enrichment	100 xp
Conclusion	50 xp

HIDE CHAPTER DETAILS ^ Completed

STATEMENT OF ACCOMPLISHMENT

View in

John Blischak
Postdoctoral Scholar at University of Chicago

John Blischak is a researcher in the Department of Human Genetics at the University of Chicago. He has years of experience using linear models to generate insight from functional genomics experiments, and he is excited to teach you how you can do the same.

See More

COLLABORATOR(S)

Richie Cotton
David Campos
Shon Inouye

PREREQUISITES

Introduction to R
Introduction to Data

DATASETS

Doxorubicin dataset
Leukemia dataset
Hypoxia dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Intermediate R: Practice

Replay Course

4.3 (1,580 Ratings)

4 hours | 0 Videos | 52 Exercises | 54,749 Participants | 4,800 XP

Course Description

This follow-up course on Intermediate R does not cover new programming concepts. Instead, you will strengthen your knowledge of the topics in Intermediate R with a bunch of new and fun exercises.

1 Conditionals and Control Flow FREE — 100%

If conditionals are your thing, these exercises will be a walk in the park. Else, let the feedback guide you and add these vital elements of R to your toolkit!

Grades analysis in R	100 xp
Explore your data	100 xp
Understand your data	50 xp
Basic queries	100 xp
Build aggregates	100 xp
Logical operator	100 xp
Build aggregates (2)	100 xp
if, else	50 xp
if, else: DIY	100 xp
else if	100 xp
Embed if-else clauses	100 xp
Operations and controls expertise	70 xp

HIDE CHAPTER DETAILS — Completed

2 Loops — 100%

Looping through data structures is something you'll often do. While and for loops help you do this. Get more practice on them by analyzing the log data from a chemical plant.

Scanning Logs in R	70 xp
Understand the logs	50 xp
While: start easy	100 xp
Adapt the while loop	100 xp
While: different approach	70 xp
The for loop	100 xp
Going through the list	0 xp
Adapt the logs list	100 xp
Collect all failures	100 xp

HIDE CHAPTER DETAILS — Completed

3 Functions — 100%

Functions make R powerful: you can isolate chunks of code, wrap them in a function and use them whenever you want. In this set of exercises, you'll practice more on using functions and writing your own functions.

Using functions	100 xp
Optional Arguments	100 xp
Which call is valid?	50 xp
Extract log information (1)	70 xp
Extract log information (2)	100 xp
Extract log information (3)	100 xp
Extract log information (4)	100 xp
Extract log information (5)	100 xp
Over to you	0 xp

HIDE CHAPTER DETAILS — Completed

4 The apply family — 100%

lapply, sapply and vapply are all members of R's apply family: they provide a fast and intuitive alternative to the while and for loops you've learned about before. Become an apply pro with some more practice!

lapply refresher	100 xp
lapply on logs (1)	100 xp
lapply on logs (2)	100 xp
lapply on logs (3)	100 xp
sapply refresher	100 xp
sapply on logs (1)	100 xp
sapply on logs (2)	100 xp
lapply or sapply?	50 xp
vapply refresher	70 xp
vapply on logs (1)	100 xp
Loop it the way you want it	70 xp

HIDE CHAPTER DETAILS — Completed

5 Utilities — 100%

To finish off these supplementary exercises, you can exercise some more with often-used functions in R, regular expressions and manipulating dates and times.

Titanic	100 xp
Observations and variables	50 xp
Exploratory queries	100 xp
Infer gender from name (1)	100 xp
Infer gender from name (2)	100 xp
Infer gender from name (3)	100 xp
Reformat passenger names	100 xp
Choose the pattern	50 xp
Add birth dates	100 xp
Average age	70 xp
Age in perspective	50 xp

HIDE CHAPTER DETAILS — Completed

STATEMENT OF ACCOMPLISHMENT

View

Filip Schouwenaars
Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

PREREQUISITES

- Introduction to R
- Intermediate R

DATASETS

- The 1912 Titanic ship disaster
- Chemical company log files

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Intro to Statistics with R: Multiple Regression

Replay Course

This course has been archived. If you are taking this course you will be able to finish it. If you are working on a track containing this course it will remain available until you complete the track.

3 hours | 7 Videos | 31 Exercises | 27,334 Participants | 2,400 XP

Course Description

Multiple regression is a powerful statistical technique, and here you will discover why and how to use it. Part of the course will focus on matrix algebra since it is essential if you want to start estimating regression coefficients in the regression equation. The final chapter will introduce dummy coding as a technique to handle categorical variables.

1 A gentle introduction to the principles of multiple regression

FREE 100%

The first chapter of the module will start with introducing the multiple regression equation, and the multiple correlation coefficient. You will visualize relationships between variables, and learn how to interpret the outcomes of the model.

A gentle introduction to the principles of multiple regression	✓ 50 xp
Multiple regression: starting off	✓ 50 xp
Multiple regression: visualization of the relationships	✓ 0 xp
Multiple regression: model selection	✓ 100 xp
Multiple regression: beware of redundancy	✓ 0 xp
Multiple regression: interpretation	✓ 50 xp
Multiple regression: interpretation regression constants	✓ 50 xp
Multiple regression: interpretation regression coefficients	✓ 50 xp
Multiple regression: strongest predictor variable	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Intuition behind estimation of multiple regression coefficients 100%

This chapter is especially for those that haven't done matrix algebra before, or for those that need to do a quick refresh on it. If you want to have a basic understanding on how the regression coefficients are estimated all at once in a multiple regression, you need matrix algebra. Step-by-step this chapter will show you how you go in R from a raw matrix data frame to the correlation matrix and the corresponding regression coefficients.

Definition of matrices	✓ 100 xp
Addition, subtraction and multiplication of matrices	✓ 100 xp
Which matrices might be multiplied?	✓ 50 xp
Correlation matrix of a raw dataframe	✓ 50 xp
Row vector of sums	✓ 100 xp
Row vector of means and matrix of means	✓ 0 xp
Matrix of deviation scores	✓ 100 xp
Sum of squares and sum of cross products matrix	✓ 100 xp
Calculating the correlation matrix	✓ 0 xp
Estimation of regression coefficients	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Dummy coding 100%

Dummy coding is used to code categorical variables in a regression analysis. Furthermore, dummy coding will also play an important role once you start doing more complex multiple regression analysis like in moderation (module 7). Conceptually, this chapter is not that hard, but dummy coding can become tedious and you have to be careful not to get tricked when doing your analysis. This chapter will show you how to avoid the most common traps.

An introduction to dummy coding	✓ 50 xp
Starting off	✓ 100 xp
Creating dummy variables (1)	✓ 100 xp
Dummy variables in multiple regressions	✓ 50 xp
Creating dummy variables (2)	✓ 100 xp
Model selection: ANOVA	✓ 100 xp
Model selection: interpretation	✓ 50 xp
Discrepancy between actual and predicted means	✓ 100 xp
Discrepancy between actual and predicted means: get intuitive!	✓ 50 xp
An introduction to effects coding	✓ 50 xp
Unweighted effects coding	✓ 100 xp
Weighted effects coding	✓ 0 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

Andrew Conway
Instructor

Andrew Conway is a Psychology Professor in the Division of Behavioral and Organizational Sciences at Claremont Graduate University in Claremont, California. He has been teaching introduction to statistics for undergraduate students and advanced statistics for graduate students for 20 years, at a variety of institutions, including the University of South Carolina, the University of Illinois in Chicago, and Princeton University.

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Bayesian Modeling with RJAGS

Replay Course

4.5 (147 Ratings)

4 hours | 15 Videos | 58 Exercises | 1,759 Participants | 4,650 XP

Course Description

The Bayesian approach to statistics and machine learning is logical, flexible, and intuitive. In this course, you will engineer and analyze a family of foundational, generalizable Bayesian models. These range in scope from fundamental one-parameter models to intermediate multivariate & generalized linear regression models. The popularity of such Bayesian models has grown along with the availability of computing resources required for their implementation. You will utilize one of these resources - the rjags package in R. Combining the power of R with the JAGS (Just Another Gibbs Sampler) engine, rjags provides a framework for Bayesian modeling, inference, and prediction.

1 Introduction to Bayesian Modeling FREE 100%

Bayesian models combine prior insights with insights from observed data to form updated, posterior insights about a parameter. In this chapter, you will review these Bayesian concepts in the context of the foundational Beta-Binomial model for a proportion parameter. You will also learn how to use the rjags package to define, compile, and simulate this model in R.

The prior model	✓ 50 xp
Simulating a Beta prior	✓ 100 xp
Comparing & contrasting Beta priors	✓ 100 xp
Which prior?	✓ 50 xp
Data & the likelihood	✓ 50 xp
Simulating the dependence of X on p	✓ 70 xp
Approximating the likelihood function	✓ 70 xp
Interpreting the likelihood function	✓ 50 xp
The posterior model	✓ 50 xp
Define, compile, and simulate	✓ 89 xp
Updating the posterior	✓ 100 xp
Influence of the prior & data on the posterior	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Bayesian Models & Markov Chains 100%

The two-parameter Normal-Normal Bayesian model provides a simple foundation for Normal regression models. In this chapter, you will engineer the Normal-Normal and define, compile, and simulate this model using rjags. You will also explore the magic of the Markov chain mechanics behind rjags simulation.

The Normal-Normal Model	✓ 50 xp
Normal-Normal priors	✓ 100 xp
Sleep study data	✓ 100 xp
Insights from the prior and data	✓ 50 xp
Simulating the Normal-Normal in RJAGS	✓ 50 xp
Define, compile, & simulate the Normal-Normal	✓ 0 xp
Posterior insights on sleep deprivation	✓ 50 xp
Markov chains	✓ 50 xp
Storing Markov chains	✓ 100 xp
Markov chain trace plots	✓ 100 xp
Markov chain density plots	✓ 100 xp
Markov chain diagnostics & reproducibility	✓ 50 xp
Multiple chains	✓ 100 xp
Naive standard errors	✓ 100 xp
Reproducibility	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Bayesian Inference & Prediction 100%

In this chapter, you will extend the Normal-Normal model to a simple Bayesian regression model. Within this context, you will explore how to use rjags simulation output to conduct posterior inference. Specifically, you will construct posterior estimates of regression parameters using posterior means & credible intervals, you will test hypotheses using posterior probabilities, and you will construct posterior predictive distributions for new observations.

A simple Bayesian regression model	✓ 50 xp
Regression priors	✓ 100 xp
Visualizing the regression priors	✓ 0 xp
Weight & height data	✓ 100 xp
Bayesian regression in RJAGS	✓ 50 xp
Define, compile, & simulate the regression model	✓ 66 xp
Regression Markov chains	✓ 50 xp
Posterior estimation & inference	✓ 50 xp
Posterior point estimates	✓ 100 xp
Posterior credible intervals	✓ 70 xp
Posterior probabilities	✓ 100 xp
Posterior prediction	✓ 50 xp
Inference for the posterior trend	✓ 100 xp
Calculating posterior predictions	✓ 100 xp
Posterior predictive distribution	✓ 0 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Multivariate & Generalized Linear Models 100%

In this final chapter, you will generalize the simple Normal regression model for application in broader contexts. You will incorporate categorical predictors, engineer a multivariate regression model with two predictors, and finally extend this methodology to Poisson multivariate regression models for count variables.

Bayesian regression with a categorical predictor	✓ 50 xp
RailTrail sample data	✓ 70 xp
RJAGS simulation with categorical variables	✓ 0 xp
Interpreting categorical coefficients	✓ 50 xp
Inference for volume by weekday	✓ 70 xp
Multivariate Bayesian regression	✓ 50 xp
Re-examining the RailTrail data	✓ 100 xp
RJAGS simulation for multivariate regression	✓ 60 xp
Interpreting multivariate regression parameters	✓ 50 xp
Posterior inference for multivariate regression	✓ 0 xp
Bayesian Poisson regression	✓ 50 xp
RJAGS simulation for Poisson regression	✓ 90 xp
Plotting the Poisson regression model	✓ 0 xp
Inference for the Poisson rate parameter	✓ 100 xp
Poisson posterior prediction	✓ 100 xp
Conclusion	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

Alicia Johnson
Associate Professor, Macalester College

Professor Johnson's primary research interest is in Markov chain Monte Carlo (MCMC) methods. The focus of her research has been on the convergence rates of chains corresponding to MCMC algorithms on general state spaces. In addition, she enjoys the unlimited applications of statistics. As a consultant for a division of the Centers for Disease Control and Prevention and the University of Minnesota, she has collaborated on projects in entomology, forestry, primatology, public health, and others. In other words, her work in statistics allows her to keep learning a little about a lot!

See More

COLLABORATOR(S)

Chester Ismay
Nick Solomon
Eunkyung Park

PREREQUISITES

Fundamentals of Bayesian Data Analysis in R
Introduction to the Tidyverse

DATASETS

Sleep study data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Reporting with R Markdown

Replay Course

This course has been archived. If you are taking this course you will be able to finish it. If you are working on a track containing this course it will remain available until you complete the track.

3 hours | 8 Videos | 31 Exercises | 64,506 Participants | 2,550 XP

Course Description

Learn how to write a data report quickly and effectively with the R Markdown package, and share your results with your friends, colleagues or the rest of the world. Learn how you can author your own R Markdown reports, and how to automate the reporting process so that you have your own reproducible reports. By the end of the interactive data analysis reporting tutorial, you will be able to generate reports straight from your R code, documenting your work — and its results — as an HTML, pdf, slideshow or Microsoft Word document.

STATEMENT OF ACCOMPLISHMENT

View in

Team RStudio
Instructor

COLLABORATOR(S)

Filip Schouwenaars

PREREQUISITES

Introduction to R

1 Authoring R Markdown Reports FREE

100%

Begin with a real life case study written in R code and then learn to narrate the code, adding interpretations, explanations, and descriptions with Markdown, the text syntax at the heart of R Markdown.

Section 1 - Introduction	✓ 50 xp
The R Markdown Exercise interface	✓ 100 xp
Explore R Markdown	✓ 100 xp
Section 2 - R code for your report	✓ 50 xp
Prepare the workspace for preliminary analysis	✓ 100 xp
Prepare your data	✓ 100 xp
Experiment with plot generation	✓ 100 xp
Prepare a model component	✓ 100 xp
Section 3 - Markdown	✓ 50 xp
Styling narrative sections	✓ 100 xp
Lists in R Markdown	✓ 100 xp
LaTeX equations	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Embedding Code

100%

Weave your code and narration into a single document and then render the document to create a finished report that includes both the code and its output. Understand how to customize this process, and open the door for automated, targeted reporting.

Section 4 - Knitr	✓ 50 xp
R code chunks	✓ 100 xp
Customize R code chunks	✓ 100 xp
Popular chunk options	✓ 100 xp
Inline R code	✓ 100 xp
Know your options	✓ 50 xp
Labeling and reusing code chunks	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Compiling Reports

100%

This chapter will show you how to generate an HTML, pdf, or Microsoft Word version of your report, as well as a pdf or HTML slideshow. Discover how to customize details of the output, and how to combine R Markdown with Shiny and ggvis.

Section 5 - Pandoc	✓ 50 xp
Alternative output formats	✓ 100 xp
Create slideshows	✓ 100 xp
Specify knitr and pandoc options	✓ 100 xp
Brand your reports with style sheets	✓ 100 xp
Section 6 - Shiny	✓ 50 xp
Shiny to make your reports interactive	✓ 100 xp
Interactive ggvis graphics	✓ 100 xp
Interview with Yihui Xie	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Configuring R Markdown (optional)

100%

The optional video tutorial and exercises in this chapter will help you configure your personal computer to use the R Markdown package.

Section 7 - Set up	✓ 50 xp
Software for R Markdown	✓ 50 xp
Prepare your system to use R Markdown	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Visualization in R with lattice

Replay Course

4.3 (230 Ratings)

4 hours | 17 Videos | 60 Exercises | 3,135 Participants | 4,950 XP

Course Description

Visualization is an essential component of interactive data analysis in R. Traditional (base) graphics is powerful, but limited in its ability to deal with multivariate data. Trellis graphics is the natural successor to traditional graphics, extending its simple philosophy to gracefully handle common multivariable data visualization tasks. This course introduces the lattice package, which implements Trellis graphics for R, and illustrates its basic use.

1 Basic plotting with lattice FREE — 100%

Introduction to some basic plotting functions in lattice. Draw histograms, scatter plots, density plots, and box and whisker plots.

Introduction	✓ 50 xp
Formula interface in histogram() and xyplot()	✓ 50 xp
Create a histogram	✓ 100 xp
Create a scatterplot	✓ 100 xp
Differences between base R graphics and lattice histogram	✓ 50 xp
Optional arguments	✓ 50 xp
Y-axis value in histogram()	✓ 100 xp
Labels	✓ 100 xp
Density plots	✓ 100 xp
Box and whisker plots and reordering levels	✓ 50 xp
Box and whisker plot	✓ 100 xp
Whisker extents and outliers	✓ 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 Conditioning and the formula interface — 100%

These exercises will teach you to create "conditioned" plots consisting of multiple panels using the formula interface.

Conditioning	✓ 50 xp
A conditioned plot of the airquality dataset	✓ 100 xp
Conditioning in density plots	✓ 100 xp
Data summary and transformation, grouping	✓ 50 xp
Superposition in density plots	✓ 100 xp
Enhancing the legend	✓ 100 xp
Incorporating external data sources	✓ 50 xp
Conditioned scatter plot	✓ 100 xp
The as.table argument	✓ 100 xp
Custom strips labels for box and whisker plots	✓ 0 xp
The trellis object	✓ 50 xp
Modifying strip labels	✓ 100 xp
Matrix-like subsetting	✓ 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Controlling scales and graphical parameters — 100%

Learn how to control and customize axis limits and visual appearance.

Combining scales	✓ 50 xp
Explicit limits	✓ 100 xp
Per-panel explicit limits	✓ 100 xp
Customizing axis annotation	✓ 100 xp
Logarithmic scales	✓ 50 xp
Annotating logarithmic scales	✓ 100 xp
Finding patterns by sex	✓ 50 xp
Log scales in economic data	✓ 100 xp
Graphical parameters	✓ 50 xp
ggplot2 styling with par.settings	✓ 100 xp
Changing graphical parameters in histogram	✓ 70 xp
Changing plotting characters	✓ 100 xp
Using simpleTheme()	✓ 50 xp
Changing plotting characters revisited	✓ 100 xp
Changing parameters in grouped density plot	✓ 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 Customizing plots using panel functions — 100%

Learn to use panel and prepanel functions to enhance existing displays or create new ones.

Panel Functions	✓ 50 xp
Scatter plot with rugs	✓ 100 xp
Alternative panel function - panel.violin()	✓ 100 xp
Alternative panel functions - panel.smoothScatter()	✓ 100 xp
Prepanel Functions to control limits	✓ 50 xp
Prepanel functions with scales	✓ 100 xp
Optional arguments of default panel functions	✓ 50 xp
Interaction plot of residuals	✓ 50 xp
Optional arguments of panel.bwplot()	✓ 100 xp
Using emojis as plotting symbols	✓ 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

5 Extensions and the lattice ecosystem — 100%

The lattice package is not just meant to be used as a standalone collection of plotting functions. Rather, it is a framework that is used as a base by many other packages. Some of these are very specialized and beyond the scope of this course. Here we give a brief survey of extensions that are generally useful to enhance displays or create new ones.

New methods	✓ 50 xp
The xyplot() method for time series objects	✓ 100 xp
The xyplot() method with panel.horizonplot()	✓ 100 xp
New high-level functions	✓ 50 xp
Map projections	✓ 100 xp
Confidence bands using the segplot() function	✓ 100 xp
Hexagonal binning of bivariate data	✓ 100 xp
Manipulation of trellis objects	✓ 50 xp
The directlabels package	✓ 100 xp
Adding ggplot2-style layers	✓ 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Visualization with R

Deepayan Sarkar

Member of R-Core & the creator of lattice

Deepayan is an Associate Professor at the Indian Statistical Institute, and a member of R-Core. He is the creator of the lattice package, and the author of the corresponding book Lattice - Multivariate Data Visualization with R.

See More

COLLABORATOR(S)

Tom Jeon

Richie Cotton

Sumedh Panchadhar

PREREQUISITES

Introduction to R

Data Visualization in R

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Deep Learning in Python

Practice Now | Replay Course

4.5 (4,434 Ratings)

4 hours | 17 Videos | 50 Exercises | 123,637 Participants | 3,500 XP

Course Description

Deep learning is the machine learning technique behind the most exciting capabilities in diverse areas like robotics, natural language processing, image recognition and artificial intelligence (including the famous AlphaGo). In this course, you'll gain hands-on, practical knowledge of how to use deep learning with Keras 2.0, the latest version of a cutting edge library for deep learning in Python.

1 Basics of deep learning and neural networks FREE 100%

In this chapter, you'll become familiar with the fundamental concepts and terminology used in deep learning, and understand why deep learning techniques are so powerful today. You'll build simple neural networks yourself and generate predictions with them.

- Introduction to deep learning — 50 xp
- Comparing neural network models to classical regression models — 50 xp
- Forward propagation — 50 xp
- Coding the forward propagation algorithm — 100 xp
- Activation functions — 50 xp
- The Rectified Linear Activation Function — 70 xp
- Applying the network to many observations/rows of data — 0 xp
- Deeper networks — 50 xp
- Forward propagation in a deeper network — 50 xp
- Multi-layer neural networks — 0 xp
- Representations are learned — 50 xp
- Levels of representation — 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Optimizing a neural network with backward propagation 100%

Here, you'll learn how to optimize the predictions generated by your neural networks. You'll do this using a method called backward propagation, which is one of the most important techniques in deep learning. Understanding how it works will give you a strong foundation to build from in the second half of the course.

- The need for optimization — 50 xp
- Calculating model errors — 50 xp
- Understanding how weights change model accuracy — 50 xp
- Coding how weight changes affect accuracy — 0 xp
- Scaling up to multiple data points — 70 xp
- Gradient descent — 50 xp
- Calculating slopes — 100 xp
- Improving model weights — 70 xp
- Making multiple updates to weights — 100 xp
- Backpropagation — 50 xp
- The relationship between forward and backward propagation — 50 xp
- Thinking about backward propagation — 50 xp
- Backpropagation in practice — 50 xp
- A round of backpropagation — 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Building deep learning models with keras 100%

In this chapter, you'll use the keras library to build deep learning models for both regression as well as classification! You'll learn about the Specify-Compile-Fit workflow that you can use to make predictions and by the end of this chapter, you'll have all the tools necessary to build deep neural networks!

- Creating a keras model — 50 xp
- Understanding your data — 50 xp
- Specifying a model — 100 xp
- Compiling and fitting a model — 50 xp
- Compiling the model — 100 xp
- Fitting the model — 100 xp
- Classification models — 50 xp
- Understanding your classification data — 50 xp
- Last steps in classification models — 100 xp
- Using models — 50 xp
- Making predictions — 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Fine-tuning keras models 100%

Here, you'll learn how to optimize your deep learning models in keras. You'll learn how to validate your models, understand the concept of model capacity, and experiment with wider and deeper networks. Enjoy!

- Understanding model optimization — 50 xp
- Diagnosing optimization problems — 50 xp
- Changing optimization parameters — 100 xp
- Model validation — 50 xp
- Evaluating model accuracy on validation dataset — 100 xp
- Early stopping: Optimizing the optimization — 100 xp
- Experimenting with wider networks — 100 xp
- Adding layers to a network — 100 xp
- Thinking about model capacity — 50 xp
- Experimenting with model structures — 50 xp
- Stepping up to images — 50 xp
- Building your own digit recognition model — 100 xp
- Final thoughts — 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

This course is part of these tracks:

Data Scientist with Python

Machine Learning with Python

Dan Becker

Data Scientist and contributor to Keras and TensorFlow libraries

Dan Becker is a Data Scientist at Kaggle with expertise in deep learning. He has contributed to the Keras and Tensorflow libraries for deep learning, finished 2nd (out of 1353 teams) in $3million Heritage Health Prize data mining competition, supervised data science consulting projects for 6 companies in the Fortune 100 and taught deep learning workshops at events and conferences such as ODSC.

See More

COLLABORATOR(S)

Hugo Bowne-Anderson

Yashas Roy

PREREQUISITES

- Introduction to Python
- Intermediate Python for Data Science
- Supervised Learning with scikit-learn

DATASETS

- Hourly wages
- MNIST
- Titanic

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Cluster Analysis in R

Practice Now | Replay Course

4.6 (1,933 Ratings)

4 hours | 16 Videos | 52 Exercises | 14,249 Participants | 3,800 XP

Course Description

Cluster analysis is a powerful toolkit in the data science workbench. It is used to find groups of observations (clusters) that share similar characteristics. These similarities can inform all kinds of business decisions; for example, in marketing, it is used to identify distinct groups of customers for which advertisements can be tailored. In this course, you will learn about two commonly used clustering methods - hierarchical clustering and k-means clustering. You won't just learn how to use these methods, you'll build a **strong intuition** for how they work and how to interpret their results. You'll develop this intuition by exploring three different datasets: soccer player positions, wholesale customer spending data, and longitudinal occupational wage data.

1 Calculating distance between observations FREE 100%

Cluster analysis seeks to find groups of observations that are similar to one another, but the identified groups are different from each other. This similarity/difference is captured by the metric called distance. In this chapter, you will learn how to calculate the distance between observations for both continuous and categorical features. You will also develop an intuition for how the scales of your features can affect distance.

What is cluster analysis?	✓ 50 xp
When to cluster?	✓ 50 xp
Distance between two observations	✓ 50 xp
Calculate & plot the distance between two players	✓ 100 xp
Using the dist() function	✓ 100 xp
Who are the closest players?	✓ 50 xp
The importance of scale	✓ 50 xp
Effects of scale	✓ 100 xp
When to scale data?	✓ 50 xp
Measuring distance for categorical data	✓ 50 xp
Calculating distance between categorical variables	✓ 100 xp
The closest observation to a pair	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Hierarchical clustering 100%

This chapter will help you answer the last question from chapter 1 - how do you find groups of similar observations (clusters) in your data using the distances that you have calculated? You will learn about the fundamental principles of hierarchical clustering - the linkage criteria and the dendrogram plot - and how both are used to build clusters. You will also explore data from a wholesale distributor in order to perform market segmentation of clients using their spending habits.

Comparing more than two observations	✓ 50 xp
Calculating linkage	✓ 100 xp
Revisited: The closest observation to a pair	✓ 50 xp
Capturing K clusters	✓ 50 xp
Assign cluster membership	✓ 100 xp
Exploring the clusters	✓ 100 xp
Validating the clusters	✓ 50 xp
Visualizing the Dendrogram	✓ 50 xp
Comparing average, single & complete linkage	✓ 100 xp
Height of the tree	✓ 50 xp
Cutting the tree	✓ 50 xp
Clusters based on height	✓ 100 xp
Exploring the branches cut from the tree	✓ 100 xp
What do we know about our clusters?	✓ 50 xp
Making sense of the clusters	✓ 50 xp
Segment wholesale customers	✓ 100 xp
Explore wholesale customer clusters	✓ 100 xp
Interpreting the wholesale customer clusters	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 K-means clustering 100%

In this chapter, you will build an understanding of the principles behind the k-means algorithm, learn how to select the right k when it isn't previously known, and revisit the wholesale data from a different perspective.

Introduction to K-means	✓ 50 xp
K-means on a soccer field	✓ 100 xp
K-means on a soccer field (part 2)	✓ 100 xp
Evaluating different values of K by eye	✓ 50 xp
Many K's many models	✓ 100 xp
Elbow (Scree) plot	✓ 100 xp
Interpreting the elbow plot	✓ 50 xp
Silhouette analysis: Observation level performance	✓ 50 xp
Silhouette analysis	✓ 100 xp
Making sense of the K-means clusters	✓ 50 xp
Revisiting wholesale data: "Best" k	✓ 100 xp
Revisiting wholesale data: Exploration	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Case Study: National Occupational mean wage 100%

In this chapter, you will apply the skills you have learned to explore how the average salary amongst professions have changed over time.

Occupational wage data	✓ 50 xp
Initial exploration of the data	✓ 50 xp
Hierarchical clustering: Occupation trees	✓ 100 xp
Hierarchical clustering: Preparing for exploration	✓ 100 xp
Hierarchical clustering: Plotting occupational clusters	✓ 100 xp
Reviewing the HC Results	✓ 50 xp
K-means: Elbow analysis	✓ 100 xp
K-means: Average Silhouette Widths	✓ 100 xp
The "best" number of clusters	✓ 50 xp
Review K-means Results	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Scientist with R

Unsupervised Machine Learning with R

Dmitriy Gorenshteyn

Lead Data Scientist at Memorial Sloan Kettering Cancer Center

Dmitriy is a Lead Data Scientist in the Strategy & Innovation department at Memorial Sloan Kettering Cancer Center. At MSK he develops predictive models for programs aimed at improving patient care. Prior to this role, Dmitriy completed his Doctorate in Quantitative & Computational Biology at Princeton University. With a passion for teaching and for R, he regularly holds cross-departmental R training sessions within MSK. His core teaching philosophy is centered on building intuition and understanding for the methods and tools available.

See More

COLLABORATOR(S)

Yashas Roy

Richie Cotton

PREREQUISITES

Intermediate R

DATASETS

Soccer player positions

Occupational Employment Statistics (OES)

Wholesale customer spending

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Experimental Design in R

Replay Course

4.3 (627 Ratings)

4 hours | 12 Videos | 52 Exercises | 3,423 Participants | 4,400 XP

Course Description

Experimental design is a crucial part of data analysis in any field, whether you work in business, health or tech. If you want to use data to answer a question, you need to design an experiment! In this course you will learn about basic experimental design, including block and factorial designs, and commonly used statistical tests, such as the t-tests and ANOVAs. You will use built-in R data and real world datasets including the CDC NHANES survey, SAT Scores from NY Public Schools, and Lending Club Loan Data. Following the course, you will be able to design and analyze your own experiments!

1 Introduction to Experimental Design FREE 100%

An introduction to key parts of experimental design plus some power and sample size calculations.

Intro to Experimental Design	50 xp
A basic experiment	100 xp
Randomization	100 xp
Replication	100 xp
Blocking	100 xp
Hypothesis Testing	50 xp
One sided vs. Two-sided tests	100 xp
pwr package Help Docs exploration	50 xp
Power & Sample Size Calculations	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Basic Experiments 100%

Explore the Lending Club dataset plus build and validate basic experiments, including an A/B test.

Single & Multiple Factor Experiments	50 xp
Exploratory Data Analysis (EDA) Lending Club	100 xp
How does loan purpose affect amount funded?	100 xp
Which loan purpose mean is different?	100 xp
Multiple Factor Experiments	100 xp
Model Validation	50 xp
Pre-modeling EDA	100 xp
Post-modeling validation plots + variance	100 xp
Kruskal-Wallis rank sum test	100 xp
A/B Testing	50 xp
Which post-A/B test test?	50 xp
Sample size for A/B test	100 xp
Basic A/B test	100 xp
A/B tests vs. multivariable experiments	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Randomized Complete (& Balanced Incomplete) Block Designs 100%

Use the NHANES data to build a RCBD and BIBD experiment, including model validation and design tips to make sure the BIBD is valid.

Intro to NHANES dataset & Sampling	50 xp
NHANES dataset construction	100 xp
NHANES EDA	100 xp
NHANES Data Cleaning	0 xp
Resampling NHANES data	100 xp
Randomized Complete Block Designs (RCBD)	50 xp
Which is NOT a good blocking factor?	50 xp
Drawing RCBDs with Agricolae	100 xp
NHANES RCBD	100 xp
RCBD Model Validation	100 xp
Balanced Incomplete Block Designs (BIBD)	50 xp
Is a BIBD even possible?	50 xp
Drawing BIBDs with agricolae	100 xp
BIBD - cat's kidney function	100 xp
NHANES BIBD	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Latin Squares, Graeco-Latin Squares, & Factorial experiments 100%

Evaluate the NYC SAT scores data and deal with its missing values, then evaluate Latin Square, Graeco-Latin Square, and Factorial experiments.

Latin Squares	50 xp
NYC SAT Scores EDA	100 xp
Dealing with Missing Test Scores	100 xp
Drawing Latin Squares with agricolae	100 xp
Latin Square with NYC SAT Scores	100 xp
Graeco-Latin Squares	50 xp
NYC SAT Scores Data Viz	100 xp
Drawing Graeco-Latin Squares with agricolae	100 xp
Graeco-Latin Square with NYC SAT Scores	100 xp
Factorial Experiments	50 xp
NYC SAT Scores Factorial EDA	100 xp
Factorial Experiment with NYC SAT Scores	100 xp
Evaluating the NYC SAT Scores Factorial Model	100 xp
What's next in Experimental Design	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Statistics Fundamentals with R

kaelen medeiros
Data Scientist, Product at DataCamp

Kaelen is the Data Scientist for the Product team at DataCamp. They are an admin for the R-Ladies Global community. Kaelen received a MS in Biostatistics from Louisiana State University Health Sciences Center, where they worked at the Louisiana Tumor Registry. Before DataCamp, they designed experiments (and more!) for the American College of Surgeons, HERE Technologies, and HealthLabs. If you meet them, you will undoubtedly hear about their cat, Scully, within the first 3 minutes. Other favorite topics include aliens, popcorn, podcasts, and nail polish.

See More

COLLABORATOR(S)

Richie Cotton

Becca Robins

PREREQUISITES

Introduction to Data

Introduction to the Tidyverse

DATASETS

sample of Lending Club data

NHANES Body Measures

NHANES Demographics

NHANES final combined dataset

NHANES Medical Conditions

NYC SAT Scores

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to Python

Practice Now | Replay Course

4.6 (34,138 Ratings)

4 hours | 11 Videos | 57 Exercises | 1,534,694 Participants | 4,700 XP | PROGRESS ON MOBILE: 24%

Course Description

Python is a general-purpose programming language that is becoming more and more popular for doing data science. Companies worldwide are using Python to harvest insights from their data and get a competitive edge. Unlike any other Python tutorial, this course focuses on Python specifically for data science. In our Intro to Python class, you will learn about powerful ways to store and manipulate data as well as cool data science tools to start your own analyses. Enter DataCamp's online Python curriculum.

1 Python Basics — 100%

An introduction to the basic concepts of Python. Learn how to use Python both interactively and through a script. Create your first variables and acquaint yourself with Python's basic data types.

Hello Python!	✓ 50 xp
The Python Interface	✓ 100 xp
When to use Python?	✓ 50 xp
Any comments?	✓ 100 xp
Python as a calculator	✓ 100 xp
Variables & Types	✓ 50 xp
Variable Assignment	✓ 100 xp
Calculations with variables	✓ 100 xp
Other variable types	✓ 100 xp
Guess the type	✓ 50 xp
Operations with other types	✓ 100 xp
Type conversion	✓ 100 xp
Can Python handle everything?	✓ 50 xp

HIDE CHAPTER DETAILS ^ | ✓ Completed

2 Python Lists — 100%

Learn to store, access and manipulate data in lists: the first step towards efficiently working with huge amounts of data.

Lists, what are they?	✓ 50 xp
Create a list	✓ 100 xp
Create list with different types	✓ 100 xp
Select the valid list	✓ 50 xp
List of lists	✓ 100 xp
Subsetting lists	✓ 50 xp
Subset and conquer	✓ 100 xp
Subset and calculate	✓ 100 xp
Slicing and dicing	✓ 100 xp
Slicing and dicing (2)	✓ 100 xp
Subsetting lists of lists	✓ 50 xp
List Manipulation	✓ 50 xp
Replace list elements	✓ 100 xp
Extend a list	✓ 100 xp
Delete list elements	✓ 50 xp
Inner workings of lists	✓ 100 xp

HIDE CHAPTER DETAILS ^ | ✓ Completed

3 Functions and Packages — 100%

To leverage the code that brilliant Python developers have written, you'll learn about using functions, methods and packages. This will help you to reduce the amount of code you need to solve challenging problems!

Functions	✓ 50 xp
Familiar functions	✓ 100 xp
Help!	✓ 50 xp
Multiple arguments	✓ 100 xp
Methods	✓ 50 xp
String Methods	✓ 100 xp
List Methods	✓ 100 xp
List Methods (2)	✓ 100 xp
Packages	✓ 50 xp
Import package	✓ 100 xp
Selective import	✓ 100 xp
Different ways of importing	✓ 50 xp

HIDE CHAPTER DETAILS ^ | ✓ Completed

4 NumPy — 100%

NumPy is a Python package to efficiently do data science. Learn to work with the NumPy array, a faster and more powerful alternative to the list, and take your first steps in data exploration.

NumPy	✓ 50 xp
Your First NumPy Array	✓ 100 xp
Baseball players' height	✓ 100 xp
Baseball player's BMI	✓ 100 xp
Lightweight baseball players	✓ 100 xp
NumPy Side Effects	✓ 50 xp
Subsetting NumPy Arrays	✓ 100 xp
2D NumPy Arrays	✓ 50 xp
Your First 2D NumPy Array	✓ 100 xp
Baseball data in 2D form	✓ 100 xp
Subsetting 2D NumPy Arrays	✓ 100 xp
2D Arithmetic	✓ 100 xp
NumPy: Basic Statistics	✓ 50 xp
Average versus median	✓ 100 xp
Explore the baseball data	✓ 100 xp
Blend it all together	✓ 100 xp

HIDE CHAPTER DETAILS ^ | ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Scientist with Python
- Python Programmer
- Python Programming

Filip Schouwenaars
Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

Hugo Bowne-Anderson
Data Scientist at DataCamp

Hugo is a data scientist, educator, writer and podcaster and DataCamp. His main interests are promoting data & AI literacy, helping to spread data skills through organizations and society and doing amateur stand up comedy in NYC. If you want to know what he likes to talk about, definitely check out DataFramed, the DataCamp podcast, which he hosts and produces: https://www.datacamp.com/community/podcast

See More

COLLABORATOR(S)

Vincent Vankrunkelsven

Patrick Varilly

DATASETS

- MLB (baseball)
- FIFA (soccer)

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Generalized Linear Models in R

Replay Course

4.2 (475 Ratings)

4 hours | 14 Videos | 56 Exercises | 3,379 Participants | 4,300 XP

Course Description

Linear regression serves as a workhorse of statistics, but cannot handle some types of complex data. A generalized linear model (GLM) expands upon linear regression to include non-normal distributions including binomial and count data. Throughout this course, you will expand your data science toolkit to include GLMs in R. As part of learning about GLMs, you will learn how to fit model binomial data with logistic regression and count data with Poisson regression. You will also learn how to understand these results and plot them with ggplot2.

STATEMENT OF ACCOMPLISHMENT

View

1 GLMs, an extension of your regression toolbox FREE 100%

This chapter teaches you how generalized linear models are an extension of other models in your data science toolbox. The chapter also uses Poisson regression to introduce generalize linear models.

- Limitations of linear models — 50 xp
- Assumptions of linear models — 50 xp
- Refresher on fitting linear models — 100 xp
- Poisson regression — 50 xp
- Fitting a Poisson regression in R — 100 xp
- Comparing linear and Poisson regression — 100 xp
- Intercepts-Comparisons versus means — 100 xp
- Basic lm() functions with glm() — 50 xp
- Applying summary, print, and tidy to glm — 100 xp
- Extracting coefficients from glm() — 100 xp
- Predicting with glm() — 100 xp

HIDE CHAPTER DETAILS — Completed

2 Logistic Regression 100%

This chapter covers running a logistic regression and examining the model outputs.

- Overview of logistic regression — 50 xp
- Fitting a logistic regression — 100 xp
- Examining logistic regression outputs — 100 xp
- Interpreting logistic regression outputs — 50 xp
- Bernoulli versus binomial distribution — 50 xp
- Bernoulli versus binomial — 50 xp
- Simulating binary data — 100 xp
- Long-form logistic regression input — 100 xp
- Wide-form input logistic regression — 100 xp
- Comparing logistic regression outputs — 50 xp
- Link functions-Probit compared to logit — 50 xp
- Probit versus to logit — 50 xp
- Fitting probits and logits — 100 xp
- Simulating a logit — 100 xp
- Simulating a probit — 100 xp

HIDE CHAPTER DETAILS — Completed

3 Interpreting and visualizing GLMs 100%

This chapter teaches you about interpreting GLM coefficients and plotting GLMs using ggplot2.

- Poisson regression coefficients — 50 xp
- Poisson link — 50 xp
- lm vs. Poisson coeffients — 100 xp
- Comparing lm and Poisson coeffcients — 50 xp
- Plotting Poisson regression — 50 xp
- Poisson regression plotting — 100 xp
- Understanding output from logistic regression — 50 xp
- Understanding odds ratios — 50 xp
- Extracting odds-ratios — 100 xp
- Interpreting odds-ratios — 50 xp
- Odds-ratios in the Tidyverse — 100 xp
- Interpreting odds-ratio confidence intervals — 50 xp
- ggplot2 and binomial regression — 50 xp
- Default trend lines — 100 xp
- Methods for trend lines — 100 xp
- Comparing probits and logits — 100 xp

HIDE CHAPTER DETAILS — Completed

4 Multiple regression with GLMs 100%

In this chapter, you will learn how to do multiple regression with GLMs in R.

- Multiple logistic regression — 50 xp
- Fitting a multiple logistic regression — 100 xp
- Building two models — 100 xp
- Comparing regression outputs — 50 xp
- Comparing variable order — 100 xp
- Formulas in R — 50 xp
- Multiple slopes — 100 xp
- Intercepts — 100 xp
- Multiple intercepts — 100 xp
- Assumptions of multiple logistic regression — 50 xp
- Simpson's paradox — 100 xp
- Results from Berkeley data — 50 xp
- Non-linear logistic regression — 100 xp
- Conclusion — 50 xp

HIDE CHAPTER DETAILS — Completed

Richard Erickson
Quantitative Ecologist

Richard helps people to experience and understand their increasingly numerical world. For his day job he develops new quantitative methods for monitoring and controlling invasive species as well as helping other scientists analyze and understand their data. He has worked on diverse datasets ranging from continent wide species distributions to pesticides in playa wetlands. After hours, he teaches SCUBA Diving as a NAUI Instructor. He has been a "UserR" since 2007.

See More

COLLABORATOR(S)

- Chester Ismay
- David Campos
- Shon Inouye

PREREQUISITES

- Multiple and Logistic Regression

DATASETS

- Bus Commuter dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Intro to SQL for Data Science

Practice Now | Replay Course

4.7 (21,695 Ratings)

4 hours | 1 Videos | 41 Exercises | 316,556 Participants | 3,450 XP | PROGRESS ON MOBILE: 80%

Course Description

The role of a data scientist is to turn raw data into actionable insights. Much of the world's raw data--from electronic medical records to customer transaction histories--lives in organized collections of tables called *relational databases*. Therefore, to be an effective data scientist, you must know how to wrangle and extract data from these databases using a language called SQL (pronounced ess-cue-ell, or sequel). This course teaches syntax in SQL shared by many types of databases, such as PostgreSQL, MySQL, SQL Server, and Oracle.This course teaches you everything you need to know to begin working with databases today!

1 Selecting columns — 100%

This chapter provides a brief introduction to working with relational databases. You'll learn about their structure, how to talk about them using database lingo, and how to begin an analysis by using simple SQL commands to select and summarize columns from database tables.

Welcome to the course!	50 xp
Onboarding | Tables	50 xp
Onboarding | Query Result	50 xp
Onboarding | Errors	100 xp
Onboarding | Bullet Exercises	90 xp
Beginning your SQL journey	50 xp
SELECTing single columns	90 xp
SELECTing multiple columns	120 xp
SELECT DISTINCT	90 xp
Learning to COUNT	50 xp
Practice with COUNT	150 xp

HIDE CHAPTER DETAILS | Completed

2 Filtering rows — 100%

This chapter builds on the first by teaching you how to filter tables for rows satisfying some criteria of interest. You'll learn how to use basic comparison operators, combine multiple criteria, match patterns in text, and much more.

Filtering results	50 xp
Simple filtering of numeric values	90 xp
Simple filtering of text	120 xp
WHERE AND	90 xp
WHERE AND OR	50 xp
WHERE AND OR (2)	90 xp
BETWEEN	50 xp
BETWEEN (2)	100 xp
WHERE IN	100 xp
Introduction to NULL and IS NULL	50 xp
NULL and IS NULL	100 xp
LIKE and NOT LIKE	100 xp

HIDE CHAPTER DETAILS | Completed

3 Aggregate Functions — 100%

This chapter builds on the first two by teaching you how to use aggregate functions to summarize your data and gain useful insights. Additionally, you'll learn about arithmetic in SQL, and how to use aliases to make your results more readable!

Aggregate functions	120 xp
Aggregate functions practice	120 xp
Combining aggregate functions with WHERE	120 xp
A note on arithmetic	50 xp
It's AS simple AS aliasing	90 xp
Even more aliasing	90 xp

HIDE CHAPTER DETAILS | Completed

4 Sorting, grouping and joins — 100%

This chapter provides a brief introduction to sorting and grouping your results, and briefly touches on the concept of joins.

ORDER BY	50 xp
Sorting single columns	90 xp
Sorting single columns (2)	90 xp
Sorting single columns (DESC)	90 xp
Sorting multiple columns	120 xp
GROUP BY	50 xp
GROUP BY practice	120 xp
GROUP BY practice (2)	150 xp
HAVING a great time	50 xp
All together now	150 xp
All together now (2)	100 xp
A taste of things to come	60 xp

HIDE CHAPTER DETAILS | Completed

STATEMENT OF ACCOMPLISHMENT

View | in

This course is part of these tracks:

- Data Analyst with Python
- Data Analyst with R
- Data Scientist with R
- Data Scientist with Python
- SQL Fundamentals

Nick Carchedi
Product Manager at DataCamp

Nick is a Product Manager at DataCamp, responsible for the learning and teaching experience. Prior to joining DataCamp, he earned his master's degree at Johns Hopkins Biostatistics and worked as a data scientist for McKinsey. Nick's passion for teaching data science began in graduate school, where he was heavily involved in tutoring fellow students, developing the Johns Hopkins Data Science Specialization, and building the swirl R package.

See More

COLLABORATOR(S)

Colin Ricardo
Filip Schouwenaars

DATASETS

IMDb Film data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Foundations of Probability in R

Replay Course

4.6 (838 Ratings)

4 hours | 13 Videos | 54 Exercises | 10,876 Participants | 4,350 XP

Course Description

Probability is the study of making predictions about random phenomena. In this course, you'll learn about the concepts of random variables, distributions, and conditioning, using the example of coin flips. You'll also gain intuition for how to solve probability problems through random simulation. These principles will help you understand statistical inference and can be applied to draw conclusions from data.

1 The binomial distribution FREE

100%

One of the simplest and most common examples of a random phenomenon is a coin flip: an event that is either "yes" or "no" with some probability. Here you'll learn about the binomial distribution, which describes the behavior of a combination of yes/no trials and how to predict and simulate its behavior.

Flipping coins in R	50 xp
Simulating coin flips	100 xp
Simulating draws from a binomial	100 xp
Density and cumulative density	50 xp
Calculating density of a binomial	100 xp
Calculating cumulative density of a binomial	70 xp
Varying the number of trials	100 xp
Expected value and variance	50 xp
Calculating the expected value	100 xp
Calculating the variance	100 xp

HIDE CHAPTER DETAILS ^ Completed

2 Laws of probability

100%

In this chapter you'll learn to combine multiple probabilities, such as the probability two events both happen or that at least one happens, and confirm each with random simulations. You'll also learn some of the properties of adding and multiplying random variables.

Probability of event A and event B	50 xp
Solving for probability of A and B	50 xp
Simulating the probability of A and B	100 xp
Simulating the probability of A, B, and C	100 xp
Probability of A or B	50 xp
Solving for probability of A or B	50 xp
Simulating probability of A or B	100 xp
Probability either variable is less than or equal to 4	70 xp
Multiplying random variables	50 xp
Expected value of multiplying a random variable	50 xp
Simulating multiplying a random variable	100 xp
Variance of a multiplied random variable	100 xp
Adding two random variables	50 xp
Solving for the sum of two binomial variables	50 xp
Simulating adding two binomial variables	100 xp
Simulating variance of sum of two binomial variables	100 xp

HIDE CHAPTER DETAILS ^ Completed

3 Bayesian statistics

100%

Bayesian statistics is a mathematically rigorous method for updating your beliefs based on evidence. In this chapter, you'll learn to apply Bayes' theorem to draw conclusions about whether a coin is fair or biased, and back it up with simulations.

Updating with evidence	50 xp
Updating	50 xp
Updating with simulation	100 xp
Updating after 16 heads	50 xp
Updating with simulation after 16 heads	100 xp
Prior probability	50 xp
Updating with priors	100 xp
Updating with three coins	100 xp
Bayes' theorem	50 xp
Updating with Bayes theorem	100 xp
Updating for other outcomes	0 xp
More updating with priors	100 xp

HIDE CHAPTER DETAILS ^ Completed

4 Related distributions

100%

So far we've been talking about the binomial distribution, but this is one of many probability distributions a random variable can take. In this chapter we'll introduce three more that are related to the binomial: the normal, the Poisson, and the geometric.

The normal distribution	50 xp
Approximating a binomial to the normal	50 xp
Simulating from the binomial and the normal	100 xp
Comparing the cumulative density of the binomial	100 xp
Comparing the distributions of the normal and binomial for low n	100 xp
The Poisson distribution	50 xp
Approximating a binomial with a Poisson	50 xp
Simulating from a Poisson and a binomial	100 xp
Density of the Poisson distribution	70 xp
Sum of two Poisson variables	100 xp
The geometric distribution	50 xp
Waiting for first coin flip	100 xp
Using replicate() for simulation	100 xp
Simulating from the geometric distribution	100 xp
Probability of a machine lasting X days	70 xp
Graphing the probability that a machine still works	0 xp

HIDE CHAPTER DETAILS ^ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Probability and Distributions with R

David Robinson
Chief Data Scientist, DataCamp

Dave works on the data science behind DataCamp's product and curriculum development. He has worked as a data scientist at Stack Overflow and received his PhD in Quantitative and Computational Biology from Princeton University. His interests include statistics, data analysis, education, and programming in R. Follow him at @drob on Twitter or on his blog, Variance Explained.

See More

COLLABORATOR(S)

Nick Carchedi
Tom Jeon
Nick Solomon

PREREQUISITES

Introduction to R

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Visualization with ggplot2 (Part 2)

Replay Course

4.2 (2,771 Ratings)

5 hours | 11 Videos | 55 Exercises | 43,776 Participants | 4,750 XP

Course Description

This ggplot2 tutorial builds on your knowledge from the first course to produce meaningful explanatory plots. We'll explore the last four optional layers. Statistics will be calculated on the fly and we'll see how Coordinates and Facets aid in communication. Publication quality plots will be produced directly in R using the Themes layer. We'll also discuss details on data visualization best practices with ggplot2 to help make sure you have a sound understanding of what works and why. By the end of the course, you'll have all the tools needed to make a custom plotting function to explore a large data set, combining statistics and excellent visuals.

STATEMENT OF ACCOMPLISHMENT

View in

This course is part of these tracks:

Data Visualization with R

Rick Scavetta

Rick Scavetta is a co-founder of Scavetta Academy.

Rick Scavetta is a biologist, workshop trainer, freelance data scientist and cofounder of Scavetta Academy, a company dedicated to helping scientists better understand and visualize their data. Rick's practical, hands-on exposure to a wide variety of datasets has informed him of the many problems scientists face when trying to visualize their data.

See More

COLLABORATOR(S)

Vincent Vankrunkelsven

Filip Schouwenaars

PREREQUISITES

Introduction to R

Intermediate R

Data Visualization with ggplot2 (Part 1)

DATASETS

CHIS adult response dataset, 2009

1 Statistics FREE 100%

In this chapter, we'll delve into how to use R ggplot2 as a tool for graphical data analysis, progressing from just plotting data to applying a variety of statistical methods. This includes a variety of linear models, descriptive and inferential statistics (mean, standard deviation and confidence intervals) and custom functions.

Stats and Geoms	50 xp
Smoothing	100 xp
Grouping variables	100 xp
Modifying stat_smooth	100 xp
Modifying stat_smooth (2)	100 xp
Quantiles	100 xp
Sum	100 xp
Stats outside Geoms	50 xp
Preparations	0 xp
Plotting variations	100 xp
Custom Functions	100 xp
Custom Functions (2)	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Coordinates and Facets 100%

The Coordinates and Facets layers offer specific and very useful tools for efficiently and accurately communicating data. In this chapter we'll look at the various ways of effectively using these two layers.

Coordinates Layer	50 xp
Zooming In	100 xp
Aspect Ratio	100 xp
Pie Charts	100 xp
Facets Layer	50 xp
Facets: the basics	100 xp
Many variables	100 xp
Dropping levels	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Themes 100%

Now that you've built high-quality plots, it's time to make them pretty. This is the last step in the data viz process. The Themes layer will enable you to make publication quality plots directly in R.

Themes from Scratch	50 xp
Rectangles	100 xp
Lines	100 xp
Text	100 xp
Legends	100 xp
Positions	100 xp
Recycling Themes	50 xp
Updating Themes	100 xp
Exploring ggthemes	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Best Practices 100%

Once you have the technical skill to make great visualizations, it's important that you make them as meaningful as possible. In this chapter we'll go over three plot types that are mostly discouraged in the data viz community - heat maps, pie charts and dynamite plots. We'll understand what the problems are with these plots and what the alternatives are.

Best Practices: Bar Plots	50 xp
Bar Plots (1)	100 xp
Bar Plots (2)	100 xp
Bar Plots (3)	100 xp
Best Practices: Pie Charts	50 xp
Pie Charts (1)	100 xp
Pie Charts (2)	100 xp
Plot Matrix (1)	50 xp
Plot Matrix (2)	50 xp
Best Practices: Heat Maps	50 xp
Heat Maps	100 xp
Heat Maps Alternatives (1)	100 xp
Heat Maps Alternatives (2)	100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Case Study 100%

In this case study, we'll explore the large, publicly available California Health Interview Survey dataset from 2009. We'll go step-by-step through the development of a new plotting method - a mosaic plot - combining statistics and flexible visuals. At the end, we'll generalize our new plotting method to use on a variety of datasets we've seen throughout the first two courses.

CHIS - Descriptive Statistics	50 xp
Exploring Data	100 xp
Unusual Values	50 xp
Default Binwidths	50 xp
Data Cleaning	70 xp
Multiple Histograms	100 xp
Alternatives	70 xp
Do Things Manually	100 xp
CHIS - Mosaic Plots	50 xp
Marimekko/Mosaic Plot	100 xp
Adding statistics	100 xp
Adding text	100 xp
Generalizations	100 xp

HIDE CHAPTER DETAILS ✓ Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to Data

Practice Now | Replay Course

4.4 (4,794 Ratings)

4 hours | 15 Videos | 46 Exercises | 56,015 Participants | 3,200 XP

Course Description

Scientists seek to answer questions using rigorous methods and careful observations. These observations—collected from the likes of field notes, surveys, and experiments—form the backbone of a statistical investigation and are called data. Statistics is the study of how best to collect, analyze, and draw conclusions from data. It is helpful to put statistics in the context of a general process of investigation: 1) identify a question or problem; 2) collect relevant data on the topic; 3) analyze the data; and 4) form a conclusion. In this course, you'll focus on the first two steps of the process.

1 Language of data FREE

100%

This chapter introduces terminology of datasets and data frames in R.

- Welcome to the course! ✓ 50 xp
- Loading data into R ✓ 100 xp
- Types of variables ✓ 50 xp
- Identify variable types ✓ 100 xp
- Categorical data in R: factors ✓ 50 xp
- Filtering based on a factor ✓ 100 xp
- Complete filtering based on a factor ✓ 100 xp
- Discretize a variable ✓ 50 xp
- Discretize a different variable ✓ 100 xp
- Combining levels of a different factor ✓ 100 xp
- Visualizing numerical data ✓ 50 xp
- Visualizing numerical and categorical data ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Study types and cautionary tales

92%

In this chapter, you will learn about observational studies and experiments, scope of inference, and Simpson's paradox.

- Observational studies and experiments ✓ 50 xp
- Identify type of study: Reading speed and font ✓ 50 xp
- Identify type of study: Countries ✓ 100 xp
- Random sampling and random assignment ✓ 50 xp
- Random sampling or random assignment? ✓ 50 xp
- Identify the scope of inference of study ✓ 50 xp
- Simpson's paradox ✓ 50 xp
- Number of males and females admitted ✓ 100 xp
- Proportion of males admitted overall ✓ 100 xp
- Proportion of males admitted for each department ✓ 0 xp
- Admission rates for males across departments ✓ 50 xp
- Recap: Simpson's paradox ✓ 50 xp
- Identify type of study: Countries [new] 50 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Sampling strategies and experimental design

100%

This chapter defines various sampling strategies and their benefits/drawbacks as well as principles of experimental design.

- Sampling strategies ✓ 50 xp
- Sampling strategies, determine which ✓ 50 xp
- Sampling strategies, choose worst ✓ 50 xp
- Sampling in R ✓ 50 xp
- Simple random sample in R ✓ 100 xp
- Stratified sample in R ✓ 100 xp
- Compare SRS vs. stratified sample ✓ 50 xp
- Principles of experimental design ✓ 50 xp
- Identifying components of a study ✓ 50 xp
- Experimental design terminology ✓ 50 xp
- Connect blocking and stratifying ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Case study

100%

Apply terminology, principles, and R code learned in the first three chapters of this course to a case study looking at how the physical appearance of instructors impacts their students' course evaluations.

- Beauty in the classroom ✓ 50 xp
- Inspect the data ✓ 100 xp
- Identify type of study ✓ 50 xp
- Sampling / experimental attributes ✓ 50 xp
- Variables in the data ✓ 50 xp
- Identify variable types ✓ 100 xp
- Recode a variable ✓ 0 xp
- Create a scatterplot ✓ 100 xp
- Create a scatterplot, with an added layer ✓ 100 xp
- Congratulations! ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Statistics Fundamentals with R

Mine Cetinkaya-Rundel

Associate Professor at Duke University & Data Scientist and Professional Educator at RStudio

Mine is the Director of Undergraduate Studies and an Associate Professor of the Practice in the Department of Statistical Science at Duke University as well as a Professional Educator at RStudio. Her work focuses on innovation in statistics pedagogy, with an emphasis on computation, reproducible research, open-source education, and student-centered learning. She is the author of three open-source introductory statistics textbooks as part of the OpenIntro project and teaches the popular Statistics with R MOOC on Coursera.

See More

COLLABORATOR(S)

- Nick Carchedi
- Tom Jeon

PREREQUISITES

- Introduction to R

DATASETS

- Course evaluation
- UC Berkeley admissions
- US state regions

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Visualization with ggplot2 (Part 1)

Replay Course

4.5 (6,640 Ratings)

5 hours | 14 Videos | 62 Exercises | 117,595 Participants | 5,250 XP

Course Description

The ability to produce meaningful and beautiful data visualizations is an essential part of your skill set as a data scientist. This course, the first R data visualization tutorial in the series, introduces you to the principles of good visualizations and the grammar of graphics plotting concepts implemented in the ggplot2 package. ggplot2 has become the go-to tool for flexible and professional plots in R. Here, we'll examine the first three essential layers for making a plot - Data, Aesthetics and Geometries. By the end of the course you will be able to make complex exploratory plots.

1 Introduction FREE

100%

In this chapter we'll get you into the right frame of mind for developing meaningful visualizations with R. You'll understand that as a communications tool, visualizations require you to think about your audience first. You'll also be introduced to the basics of ggplot2 - the 7 different grammatical elements (layers) and aesthetic mappings.

Introduction	50 xp
Explore and Explain	50 xp
Exploring ggplot2, part 1	100 xp
Exploring ggplot2, part 2	100 xp
Grammar of Graphics	50 xp
Exploring ggplot2, part 3	100 xp
Understanding Variables	50 xp
ggplot2	50 xp
Exploring ggplot2, part 4	100 xp
Exploring ggplot2, part 5	100 xp
Understanding the grammar, part 1	100 xp
Understanding the grammar, part 2	100 xp

HIDE CHAPTER DETAILS

Completed

2 Data

100%

The structure of your data will dictate how you construct plots in ggplot2. In this chapter, you'll explore the iris dataset from several different perspectives to showcase this concept. You'll see that making your data conform to a structure that matches the plot in mind will make the task of visualization much easier through several R data visualization examples.

Objects and Layers	50 xp
base package and ggplot2, part 1 - plot	100 xp
base package and ggplot2, part 2 - lm	100 xp
base package and ggplot2, part 3	100 xp
ggplot2 compared to base package	50 xp
Proper Data Format	50 xp
Plotting the ggplot2 way	50 xp
Tidy Data	50 xp
Variables to visuals, part 1	100 xp
Variables to visuals, part 1b	100 xp
Variables to visuals, part 2	100 xp
Variables to visuals, part 2b	100 xp

HIDE CHAPTER DETAILS

Completed

3 Aesthetics

100%

Aesthetic mappings are the cornerstone of the grammar of graphics plotting concept. This is where the magic happens - converting continuous and categorical data into visual scales that provide access to a large amount of information in a very short time. In this chapter you'll understand how to choose the best aesthetic mappings for your data.

Visible Aesthetics	50 xp
All about aesthetics, part 1	100 xp
All about aesthetics, part 2	100 xp
All about aesthetics, part 3	100 xp
All about attributes, part 1	100 xp
All about attributes, part 2	100 xp
Going all out	100 xp
Aesthetics for categorical and continuous variables	50 xp
Modifying Aesthetics	50 xp
Position	100 xp
Setting a dummy aesthetic	100 xp
Aesthetics Best Practices	50 xp
Overplotting 1 - Point shape and transparency	100 xp
Overplotting 2 - alpha with large datasets	100 xp

HIDE CHAPTER DETAILS

Completed

4 Geometries

100%

A plot's geometry dictates what visual elements will be used. In this chapter, we'll familiarize you with the geometries used in the three most common plot types you'll encounter - scatter plots, bar charts and line plots. We'll look at a variety of different ways to construct these plots.

Scatter Plots	50 xp
Scatter plots and jittering (1)	100 xp
Scatter plots and jittering (2)	100 xp
Bar Plots	50 xp
Histograms	100 xp
Position	100 xp
Overlapping bar plots	100 xp
Overlapping histograms	100 xp
Bar plots with color ramp, part 1	100 xp
Bar plots with color ramp, part 2	100 xp
Overlapping histograms (2)	100 xp
Line Plots - Time Series	50 xp
Line plots	100 xp
Periods of recession	100 xp
Multiple time series, part 1	100 xp
Multiple time series, part 2	100 xp

HIDE CHAPTER DETAILS

Completed

5 qplot and wrap-up

100%

In this chapter you'll learn about qplot; it is a quick and dirty form of ggplot2. It's not as intuitive as the full-fledged ggplot() function but may be useful in specific instances. This chapter also features a wrap-up video and corresponding data visualization exercises.

qplot	50 xp
Using qplot	100 xp
Using aesthetics	100 xp
Choosing geoms, part 1	100 xp
Choosing geoms, part 2 - dotplot	0 xp
Wrap-up	50 xp
Chicken weight	100 xp
Titanic	100 xp

HIDE CHAPTER DETAILS

Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Data Visualization with R

Rick Scavetta

Rick Scavetta is a co-founder of Scavetta Academy

Rick Scavetta is a biologist, workshop trainer, freelance data scientist and cofounder of Scavetta Academy, a company dedicated to helping scientists better understand and visualize their data. Rick's practical, hands-on exposure to a wide variety of datasets has informed him of the many problems scientists face when trying to visualize their data.

See More

COLLABORATOR(S)

Vincent Vankrunkelsven

Filip Schouwenaars

PREREQUISITES

- Introduction to R
- Intermediate R

DATASETS

- Subset of 1,000 diamonds
- Fish datasets
- Iris datasets
- Recession

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Exploratory Data Analysis in R: Case Study

Replay Course

4.6 (2,055 Ratings)

4 hours | 15 Videos | 58 Exercises | 22,926 Participants | 4,800 XP

Course Description

Once you've started learning tools for data manipulation and visualization like dplyr and ggplot2, this course gives you a chance to use them in action on a real dataset. You'll explore the historical voting of the United Nations General Assembly, including analyzing differences in voting between countries, across time, and among international issues. In the process you'll gain more practice with the dplyr and ggplot2 packages, learn about the broom package for tidying model output, and experience the kind of start-to-finish exploratory analysis common in data science.

1 Data cleaning and summarizing with dplyr FREE 100%

The best way to learn data wrangling skills is to apply them to a specific case study. Here you'll learn how to clean and filter the United Nations voting dataset using the dplyr package, and how to summarize it into smaller, interpretable units.

The United Nations voting dataset	50 xp
Filtering rows	100 xp
Adding a year column	100 xp
Adding a country column	100 xp
Grouping and summarizing	50 xp
Summarizing the full dataset	100 xp
Summarizing by year	100 xp
Summarizing by country	100 xp
Sorting and filtering summarized data	50 xp
Sorting by percentage of "yes" votes	100 xp
Filtering summarized output	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Data visualization with ggplot2 100%

Once you've cleaned and summarized data, you'll want to visualize them to understand trends and extract insights. Here you'll use the ggplot2 package to explore trends in United Nations voting within each country over time.

Visualization with ggplot2	50 xp
Choosing an aesthetic	50 xp
Plotting a line over time	100 xp
Other ggplot2 layers	100 xp
Visualizing by country	50 xp
Summarizing by year and country	100 xp
Plotting just the UK over time	100 xp
Plotting multiple countries	100 xp
Faceting	50 xp
Faceting by country	100 xp
Faceting with free y-axis	100 xp
Choose your own countries	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Tidy modeling with broom 100%

While visualization helps you understand one country at a time, statistical modeling lets you quantify trends across many countries and interpret them together. Here you'll learn to use the tidyr, purrr, and broom packages to fit linear models to each country, and understand and compare their outputs.

Linear regression	50 xp
Linear regression on the United States	100 xp
Finding the slope of a linear regression	50 xp
Finding the p-value of a linear regression	50 xp
Tidying models with broom	50 xp
Tidying a linear regression model	100 xp
Combining models for multiple countries	100 xp
Nesting for multiple models	50 xp
Nesting a data frame	100 xp
List columns	100 xp
Unnesting	100 xp
Fitting multiple models	50 xp
Performing linear regression on each nested dataset	100 xp
Tidy each linear regression model	100 xp
Unnesting a data frame	100 xp
Working with many tidy models	50 xp
Filtering model terms	100 xp
Filtering for significant countries	100 xp
Sorting by slope	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Joining and tidying 100%

In this chapter, you'll learn to combine multiple related datasets, such as incorporating information about each resolution's topic into your vote analysis. You'll also learn how to turn untidy data into tidy data, and see how tidy data can guide your exploration of topics and countries over time.

Joining datasets	50 xp
Joining datasets with inner_join	100 xp
Filtering the joined dataset	100 xp
Visualizing colonialism votes	100 xp
Tidy data	50 xp
Tidy data observations	50 xp
Using gather to tidy a dataset	100 xp
Recoding the topics	100 xp
Summarize by country, year, and topic	100 xp
Visualizing trends in topics for one country	100 xp
Tidy modeling by topic and country	50 xp
Nesting by topic and country	100 xp
Interpreting tidy models	100 xp
Steepest trends by topic	50 xp
Checking models visually	100 xp
Conclusion	50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Manipulation with R
- Data Scientist with R

David Robinson
Chief Data Scientist, DataCamp

Dave works on the data science behind DataCamp's product and curriculum development. He has worked as a data scientist at Stack Overflow and received his PhD in Quantitative and Computational Biology from Princeton University. His interests include statistics, data analysis, education, and programming in R. Follow him at @drob on Twitter or on his blog, Variance Explained.

See More

COLLABORATOR(S)

- Nick Carchedi
- Tom Jeon

PREREQUISITES

- Introduction to R
- Data Visualization with ggplot2 (Part 1)

DATASETS

- United Nations voting dataset
- Topic information for each country (Descriptions)

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Fundamentals of Bayesian Data Analysis in R

Replay Course

4.6 (759 Ratings)

4 hours | 23 Videos | 58 Exercises | 7,675 Participants | 4,450 XP

Course Description

Bayesian data analysis is an approach to statistical modeling and machine learning that is becoming more and more popular. It provides a uniform framework to build problem specific models that can be used for both statistical inference and for prediction. This course will introduce you to Bayesian data analysis. What it is, how it works, and why it is a useful tool to have in your data science toolbox.

STATEMENT OF ACCOMPLISHMENT

View

Rasmus Bååth

Senior Data Scientist at King (Activision Blizzard)

I'm a Data Scientist at King (Activision/Blizzard) and previously I've been an instructor at DataCamp. I once did a PhD in Cognitive Science at Lund University. I'm passionate about Bayesian statistics, good graphs and free coffee. Follow me @rabaath on Twitter or check out my blog, Publishable Stuff.

See More

COLLABORATOR(S)

Chester Ismay

Nick Solomon

PREREQUISITES

Introduction to R

1 What is Bayesian Data Analysis? FREE — 100%

This chapter will introduce you to Bayesian data analysis and give you a feel for how it works.

1	A first taste of Bayes	50 xp
2	Unknowns and ice creams	50 xp
3	Let's try some Bayesian data analysis	50 xp
4	Coin flips with prop_model	100 xp
5	Zombie drugs with prop_model	100 xp
6	Samples and posterior summaries	50 xp
7	Looking at samples from prop_model	100 xp
8	Summarizing the zombie drug experiment	100 xp
9	You've done some Bayesian data analysis!	50 xp

HIDE CHAPTER DETAILS — Completed

2 How does Bayesian inference work? — 100%

In this chapter we will take a detailed look at the foundations of Bayesian inference.

1	The parts needed for Bayesian inference	50 xp
2	Take a generative model for a spin	100 xp
3	Take the binomial distribution for a spin	100 xp
4	Using a generative model	50 xp
5	How many visitors could your site get (1)?	100 xp
6	Representing uncertainty with priors	50 xp
7	Adding a prior to the model	100 xp
8	Bayesian models and conditioning	50 xp
9	Update a Bayesian model with data	100 xp
10	How many visitors could your site get (3)?	100 xp
11	What have we done?	50 xp

HIDE CHAPTER DETAILS — Completed

3 Why use Bayesian Data Analysis? — 100%

This chapter will show you four reasons why Bayesian data analysis is a useful tool to have in your data science tool belt.

1	Four good things with Bayes	50 xp
2	Explore using the Beta distribution as a prior	100 xp
3	Pick the prior that best captures the information	50 xp
4	Change the model to use an informative prior	100 xp
5	Contrasts and comparisons	50 xp
6	Fit the model using another dataset	100 xp
7	Calculating the posterior difference	100 xp
8	Decision analysis	50 xp
9	A small decision analysis 1	100 xp
10	A small decision analysis 2	100 xp
11	Change anything and everything	50 xp
12	The Poisson distribution	100 xp
13	Clicks per day instead of clicks per ad	76 xp
14	Bayes is optimal, kind of...	50 xp

HIDE CHAPTER DETAILS — Completed

4 Bayesian inference with Bayes' theorem — 100%

Learn what Bayes theorem is all about and how to use it for statistical inference.

1	Probability rules	50 xp
2	Cards and the sum rule	100 xp
3	Cards and the product rule	100 xp
4	Calculating likelihoods	50 xp
5	From rbinom to dbinom	100 xp
6	Calculating probabilities with dbinom	100 xp
7	Bayesian calculation	50 xp
8	Calculating a joint distribution	100 xp
9	Conditioning on the data (again)	100 xp
10	A conditional shortcut	100 xp
11	Bayes' theorem	50 xp
12	A Poisson model description	50 xp

HIDE CHAPTER DETAILS — Completed

5 More parameters, more data, and more Bayes — 100%

Learn about using the Normal distribution to analyze continuous data and try out a tool for practical Bayesian analysis in R.

1	The temperature in a Normal lake	50 xp
2	rnorm, dnorm, and the weight of newborns	76 xp
3	A Bayesian model of water temperature	50 xp
4	A Bayesian model of Zombie IQ	70 xp
5	Eyeballing the mean IQ of zombies?	50 xp
6	Answering the question: Should I have a beach party?	50 xp
7	Sampling from the zombie posterior	93 xp
8	But how smart will the next zombie be?	100 xp
9	A practical tool: BEST	50 xp
10	The BEST models and zombies on a diet	100 xp
11	BEST is robust	100 xp
12	What have you learned? What did we miss?	50 xp

HIDE CHAPTER DETAILS — Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Foundations of Inference

Replay Course

4.3 (791 Ratings)

4 hours | 17 Videos | 58 Exercises | 15,087 Participants | 4,350 XP

Course Description

One of the foundational aspects of statistical analysis is inference, or the process of drawing conclusions about a larger population from a sample of data. Although counter intuitive, the standard practice is to attempt to disprove a research claim that is not of interest. For example, to show that one medical treatment is better than another, we can assume that the two treatments lead to equal survival rates only to then be disproved by the data. Additionally, we introduce the idea of a p-value, or the degree of disagreement between the data and the hypothesis. We also dive into confidence intervals, which measure the magnitude of the effect of interest (e.g. how much better one treatment is than another).

1 Introduction to ideas of inference FREE 100%

In this chapter, you will investigate how repeated samples taken from a population can vary. It is the variability in samples that allow you to make claims about the population of interest. It is important to remember that the research claims of interest focus on the population while the information available comes only from the sample data.

Welcome to the course!	50 xp
Hypotheses (1)	50 xp
Hypotheses (2)	50 xp
Randomized distributions	50 xp
Working with the NHANES data	0 xp
Calculating statistic of interest	0 xp
Randomized data under null model of independence	0 xp
Randomized statistics and dotplot	0 xp
Randomization density	70 xp
Using the randomization distribution	50 xp
Do the data come from the population?	100 xp
What can you conclude?	50 xp
Study conclusions	50 xp

HIDE CHAPTER DETAILS ^ Completed

2 Completing a randomization test: gender discrimination 100%

In this chapter, you will gain the tools and knowledge to complete a full hypothesis test. That is, given a dataset, you will know whether or not is appropriate to reject the null hypothesis in favor of the research claim of interest.

Example: gender discrimination	50 xp
Gender discrimination hypotheses	50 xp
Summarizing gender discrimination	100 xp
Step-by-step through the permutation	100 xp
Randomizing gender discrimination	0 xp
Distribution of statistics	50 xp
Reflecting on analysis	50 xp
Critical region	100 xp
Two-sided critical region	100 xp
Why 0.05?	50 xp
How does sample size affect results?	50 xp
Sample size in randomization distribution	0 xp
Sample size for critical region	100 xp
What is a p-value?	50 xp
Calculating the p-values	100 xp
Practice calculating p-values	0 xp
Calculating two-sided p-values	100 xp
Summary of gender discrimination	50 xp

HIDE CHAPTER DETAILS ^ Completed

3 Hypothesis testing errors: opportunity cost 100%

You will continue learning about hypothesis testing with a new example and the same structure of randomization tests. In this chapter, however, the focus will be on different errors (type I and type II), how they are made, when one is worse than another, and how things like sample size and effect size impact the error rates.

Example: opportunity cost	50 xp
Summarizing opportunity cost (1)	100 xp
Plotting opportunity cost	100 xp
Randomizing opportunity cost	0 xp
Summarizing opportunity cost (2)	100 xp
Opportunity cost conclusion	50 xp
Errors and their consequences	50 xp
Different choice of error rate	50 xp
Errors for two-sided hypotheses	50 xp
p-value for two-sided hypotheses: opportunity costs	100 xp
Summary of opportunity costs	50 xp

HIDE CHAPTER DETAILS ^ Completed

4 Confidence intervals 94%

As a complement to hypothesis testing, confidence intervals allow you to estimate a population parameter. Recall that your interest is always in some characteristic of the population, but you only have incomplete information to estimate the parameter using sample data. Here, the parameter is the true proportion of successes in a population. Bootstrapping is used to estimate the variability needed to form the confidence interval.

Parameters and confidence intervals	50 xp
What is the parameter?	50 xp
Hypothesis test or confidence interval?	50 xp
Bootstrapping	50 xp
Resampling from a sample	0 xp
Visualizing the variability of p-hat	100 xp
Always resample the original number of observations	50 xp
Variability in p-hat	50 xp
Empirical Rule	100 xp
Bootstrap t-confidence interval	100 xp
Bootstrap percentile interval	0 xp
Interpreting CIs and technical conditions	50 xp
Sample size effects on bootstrap CIs	100 xp
Sample proportion value effects on bootstrap CIs	100 xp
Percentile effects on bootstrap CIs	100 xp
Summary of statistical inference	50 xp

HIDE CHAPTER DETAILS ^ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Statistical Inference with R

Jo Hardin

Professor at Pomona College

Jo Hardin is a professor of mathematics and statistics at Pomona College. Her statistical research focuses on developing new robust methods for high throughput data. Recently, she has also worked closely with the statistics education community on ways to integrate data science early into a statistics curriculum. When not working with students or on her research, she loves to put on a pair of running shoes and hit the road.

See More

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

- Introduction to R
- Introduction to Data
- Exploratory Data Analysis
- Correlation and Regression

DATASETS

- All polls
- Polling data
- Big discrimination dataset
- New discrimination dataset
- Small discrimination dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Importing Data in R (Part 2)

Practice Now | Replay Course

4.5 (5,640 Ratings)

3 hours | 10 Videos | 48 Exercises | 41,157 Participants | 3,950 XP

Course Description

In this second part to Importing Data in R, you will take a deeper dive into the wide range of data formats out there. More specifically, you'll learn how to import data from relational databases and how to import and work with data coming from the web. Finally, you'll get hands-on experience with importing data from statistical software packages such SAS, STATA and SPSS.

1 Importing data from databases (Part 1) FREE 100%

Many companies store their information in relational databases. The R community has also developed R packages to get data from these architectures. You'll learn how to connect to a database and how to retrieve data from it.

- Connect to a database ✓ 50 xp
- Establish a connection ✓ 100 xp
- Inspect the connection ✓ 50 xp
- Import table data ✓ 50 xp
- List the database tables ✓ 100 xp
- Import users ✓ 100 xp
- Import all tables ✓ 100 xp
- How do the tables relate? ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Importing data from databases (Part 2) 100%

Importing an entire table from a database while you might only need a tiny bit of information seems like a lot of unnecessary work. In this chapter, you'll learn about SQL queries, which will help you make things more efficient by performing some computations on the database side.

- SQL Queries from inside R ✓ 50 xp
- Query tweater (1) ✓ 100 xp
- Query tweater (2) ✓ 100 xp
- Query tweater (3) ✓ 100 xp
- Query tweater (4) ✓ 100 xp
- Join the query madness! ✓ 50 xp
- DBI internals ✓ 50 xp
- Send - Fetch - Clear ✓ 100 xp
- Be polite and ... ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Importing data from the web (Part 1) 100%

More and more of the information that data scientists are using resides on the web. Importing this data into R requires an understanding of the protocols used on the web. In this chapter, you'll get a crash course in HTTP and learn to perform your own HTTP requests from inside R.

- HTTP ✓ 50 xp
- Import flat files from the web ✓ 100 xp
- Secure importing ✓ 100 xp
- Downloading files ✓ 50 xp
- Import Excel files from the web ✓ 100 xp
- Downloading any file, secure or not ✓ 100 xp
- Reading a text file from the web ✓ 50 xp
- HTTP? httr! (1) ✓ 100 xp
- HTTP? httr! (2) ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Importing data from the web (Part 2) 100%

Importing data from the web is one thing; actually being able to extract useful information is another. Learn more about the JSON format to get one step closer to web domination.

- APIs & JSON ✓ 50 xp
- From JSON to R ✓ 100 xp
- Quandl API ✓ 100 xp
- OMDb API ✓ 70 xp
- JSON & jsonlite ✓ 50 xp
- JSON practice (1) ✓ 100 xp
- JSON practice (2) ✓ 100 xp
- toJSON() ✓ 100 xp
- Minify and prettify ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Importing data from statistical software packages 100%

Next to R, there are also other commonly used statistical software packages: SAS, STATA and SPSS. Each of them has their own file format. Learn how to use the haven and foreign packages to get them into R with remarkable ease!

- haven ✓ 50 xp
- Import SAS data with haven ✓ 100 xp
- Import STATA data with haven ✓ 100 xp
- What does the graphic tell? ✓ 50 xp
- Import SPSS data with haven ✓ 100 xp
- Factorize, round two ✓ 100 xp
- foreign ✓ 50 xp
- Import STATA data with foreign (1) ✓ 100 xp
- Import STATA data with foreign (2) ✓ 100 xp
- Do you know your data? ✓ 50 xp
- Import SPSS data with foreign (1) ✓ 100 xp
- Excursion: Correlation ✓ 50 xp
- Import SPSS data with foreign (2) ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- Importing & Cleaning Data with R

Filip Schouwenaars

Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

PREREQUISITES

- Importing Data in R (Part 1)

DATASETS

- Education equality data
- Employee data
- Florida election data
- International socio-economic data
- Latitude (XLS)
- Latitude (XLSX)
- Big Five data
- Potatoes
- Sales data
- Swimming pools
- Sugar import data
- Water data
- Wine data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to the Tidyverse

Practice Now | Replay Course

4.7 (10,487 Ratings)

4 hours | 16 Videos | 50 Exercises | 69,853 Participants | 4,150 XP | ALSO AVAILABLE ON MOBILE

Course Description

This is an introduction to the programming language R, focused on a powerful set of tools known as the "tidyverse". In the course you'll learn the intertwined processes of data manipulation and visualization through the tools dplyr and ggplot2. You'll learn to manipulate data by filtering, sorting and summarizing a real dataset of historical country data in order to answer exploratory questions. You'll then learn to turn this processed data into informative line plots, bar plots, histograms, and more with the ggplot2 package. This gives a taste both of the value of exploratory data analysis and the power of tidyverse tools. This is a suitable introduction for people who have no previous experience in R and are interested in learning to perform data analysis.

1 Data wrangling FREE — 100%

In this chapter, you'll learn to do three things with a table: filter for particular observations, arrange the observations in a desired order, and mutate to add or change a column. You'll see how each of these steps lets you answer questions about your data.

The gapminder dataset	50 xp
Loading the gapminder and dplyr packages	100 xp
Understanding a data frame	50 xp
The filter verb	50 xp
Filtering for one year	100 xp
Filtering for one country and one year	100 xp
The arrange verb	50 xp
Arranging observations by life expectancy	100 xp
Filtering and arranging	100 xp
The mutate verb	50 xp
Using mutate to change or create a column	100 xp
Combining filter, mutate, and arrange	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 Data visualization — 100%

You've already been able to answer some questions about the data through dplyr, but you've engaged with them just as a table (such as one showing the life expectancy in the US each year). Often a better way to understand and present such data is as a graph. Here you'll learn the essential skill of data visualization, using the ggplot2 package. Visualization and manipulation are often intertwined, so you'll see how the dplyr and ggplot2 packages work closely together to create informative graphs.

Visualizing with ggplot2	50 xp
Variable assignment	100 xp
Comparing population and GDP per capita	100 xp
Comparing population and life expectancy	100 xp
Log scales	50 xp
Putting the x-axis on a log scale	100 xp
Putting the x- and y- axes on a log scale	100 xp
Additional aesthetics	50 xp
Adding color to a scatter plot	100 xp
Adding size and color to a plot	100 xp
Faceting	50 xp
Creating a subgraph for each continent	100 xp
Faceting by year	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Grouping and summarizing — 100%

So far you've been answering questions about individual country-year pairs, but we may be interested in aggregations of the data, such as the average life expectancy of all countries within each year. Here you'll learn to use the group by and summarize verbs, which collapse large datasets into manageable summaries.

The summarize verb	50 xp
Summarizing the median life expectancy	100 xp
Summarizing the median life expectancy in 1957	100 xp
Summarizing multiple variables in 1957	100 xp
The group_by verb	50 xp
Summarizing by year	100 xp
Summarizing by continent	100 xp
Summarizing by continent and year	100 xp
Visualizing summarized data	50 xp
Visualizing median life expectancy over time	100 xp
Visualizing median GDP per capita per continent over time	100 xp
Comparing median life expectancy and median GDP per continent in 2007	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 Types of visualizations — 100%

You've learned to create scatter plots with ggplot2. In this chapter you'll learn to create line plots, bar plots, histograms, and boxplots. You'll see how each plot needs different kinds of data manipulation to prepare for it, and understand the different roles of each of these plot types in data analysis.

Line plots	50 xp
Visualizing median GDP per capita over time	100 xp
Visualizing median GDP per capita by continent over time	100 xp
Bar plots	50 xp
Visualizing median GDP per capita by continent	100 xp
Visualizing GDP per capita by country in Oceania	0 xp
Histograms	50 xp
Visualizing population	100 xp
Visualizing population with x-axis on a log scale	100 xp
Boxplots	50 xp
Comparing GDP per capita across continents	100 xp
Adding a title to your graph	100 xp
Conclusion	50 xp

HIDE CHAPTER DETAILS — ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View | in

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- R Programmer
- Tidyverse Fundamentals with R

David Robinson
Chief Data Scientist, DataCamp

Dave works on the data science behind DataCamp's product and curriculum development. He has worked as a data scientist at Stack Overflow and received his PhD in Quantitative and Computational Biology from Princeton University. His interests include statistics, data analysis, education, and programming in R. Follow him at @drob on Twitter or on his blog, Variance Explained.

See More

COLLABORATOR(S)

Yashas Roy

 Chester Ismay

DATASETS

Gapminder

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Hyperparameter Tuning in R

Replay Course

4.4 (135 Ratings)

4 hours | 14 Videos | 47 Exercises | 1,835 Participants | 3,500 XP

Course Description

For many machine learning problems, simply running a model out-of-the-box and getting a prediction is not enough; you want the best model with the most accurate prediction. One way to perfect your model is with hyperparameter tuning, which means optimizing the settings for that specific model. In this course, you will work with the caret, mlr and h2o packages to find the optimal combination of hyperparameters in an efficient manner using grid search, random search, adaptive resampling and automatic machine learning (AutoML). Furthermore, you will work with different datasets and tune different supervised learning models, such as random forests, gradient boosting machines, support vector machines, and even neural nets. Get ready to tune!

STATEMENT OF ACCOMPLISHMENT

View

1 Introduction to hyperparameters FREE

100%

Why do we use the strange word "hyperparameter"? What makes it hyper? Here, you will understand what model parameters are, and why they are different from hyperparameters in machine learning. You will then see why we would want to tune them and how the default setting of caret automatically includes hyperparameter tuning.

- Parameters vs hyperparameters — 50 xp
- Model parameters vs. hyperparameters — 100 xp
- Hyperparameters in linear models — 50 xp
- What are the coefficients? — 100 xp
- Recap of machine learning basics — 50 xp
- Machine learning with caret — 100 xp
- Resampling schemes — 50 xp
- Hyperparameter tuning in caret — 50 xp
- Hyperparameters in Stochastic Gradient Boosting — 50 xp
- Changing the number of hyperparameters to tune — 100 xp
- Tune hyperparameters manually — 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Hyperparameter tuning with caret

100%

In this chapter, you will learn how to tune hyperparameters with a Cartesian grid. Then, you will implement faster and more efficient approaches. You will use Random Search and adaptive resampling to tune the parameter grid, in a way that concentrates on values in the neighborhood of the optimal settings.

- Hyperparameter tuning in caret — 50 xp
- Finding hyperparameters — 50 xp
- Cartesian grid search in caret — 90 xp
- Plot hyperparameter model output — 100 xp
- Grid vs. Random Search — 50 xp
- Grid search with range of hyperparameters — 100 xp
- Find train() option for random search — 50 xp
- Random search with caret — 100 xp
- Adaptive Resampling — 50 xp
- Advantages of Adaptive Resampling — 50 xp
- Adaptive Resampling with caret — 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Hyperparameter tuning with mlr

100%

Here, you will use another package for machine learning that has very convenient hyperparameter tuning functions. You will define a Cartesian grid or perform Random Search, as well as advanced techniques. You will also learn different ways to plot and evaluate models with different hyperparameters.

- Machine learning with `mlr` — 50 xp
- Machine Learning with mlr — 50 xp
- Modeling with mlr — 100 xp
- Grid and Random Search with `mlr` — 50 xp
- Random search with mlr — 100 xp
- Perform hyperparameter tuning with mlr — 76 xp
- Evaluating hyperparameters with `mlr` — 50 xp
- Why to evaluate tuning? — 50 xp
- Evaluating hyperparamter tuning results — 100 xp
- Advanced tuning with `mlr` — 50 xp
- Define advanced tuning controls — 50 xp
- Define aggregated measures — 100 xp
- Setting hyperparameters — 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Hyperparameter tuning with h2o

100%

In this final chapter, you will use h2o, another package for machine learning with very convenient hyperparameter tuning functions. You will use it to train different models and define a Cartesian grid. Then, You will implement a Random Search use stopping criteria. Finally, you will learn AutoML, an h2o interface which allows for very fast and convenient model and hyperparameter tuning with just one function.

- Machine learning with h2o — 50 xp
- Prepare data for modelling with h2o — 100 xp
- Modeling with h2o — 100 xp
- Grid and random search with h2o — 50 xp
- Grid search with h2o — 100 xp
- Random search with h2o — 100 xp
- Stopping criteria — 100 xp
- Automatic machine learning with h2o — 50 xp
- AutoML in h2o — 100 xp
- Scoring the leaderboard — 50 xp
- Extract h2o models and evaluate performance — 76 xp
- Wrap-up — 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

Shirin Elsinghorst (formerly Glander)

Data Scientist @ codecentric

I'm Shirin, a biologist turned bioinformatician turned data scientist. During my PhD and Postdoc I worked with Next Generation Sequencing data to analyze diseases like arthritis. However, I then chose to become a data scientist for a German IT company, called codecentric. In this capacity, I have been working on many different projects, e.g. building fraud detection models, creating a chatbot, implementing predictive maintenance, and more. My tool of choice for data analysis so far has been R but I also work with Python. I am also very passionate about teaching and sharing knowledge, so I give workshops, talk at conferences or meetups, write blog posts and organize the MünsteR R-users group.

See More

COLLABORATOR(S)

Chester Ismay

Hadrien Lacroix

PREREQUISITES

- Introduction to the Tidyverse
- Supervised Learning in R: Classification
- Machine Learning Toolbox

DATASETS

- Bc train data
- Breast cancer data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Linear Algebra for Data Science in R

Replay Course

4.3 (275 Ratings)

4 hours | 15 Videos | 56 Exercises | 3,387 Participants | 4,000 XP

Course Description

Linear algebra is one of the most important set of tools in applied mathematics and data science. In this course, you'll learn how to work with vectors and matrices, solve matrix-vector equations, perform eigenvalue/eigenvector analyses and use principal component analysis to do dimension reduction on real-world datasets. All analyses will be performed in R, one of the world's most-popular programming languages.

1 Introduction to Linear Algebra FREE 100%

In this chapter, you will learn about the key objects in linear algebra, such as vectors and matrices. You will understand why they are important and how they interact with each other.

Motivations	50 xp
Creating Vectors in R	100 xp
The Algebra of Vectors	100 xp
Creating Matrices in R	100 xp
Matrix-Vector Operations	50 xp
Matrix-Vector Compatibility	50 xp
Matrix Multiplication as a Transformation	100 xp
Reflections	100 xp
Matrix-Matrix Calculations	50 xp
Matrix Multiplication Compatibility	50 xp
Matrix Multiplication - Order Matters	100 xp
Intro to The Matrix Inverse	70 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Matrix-Vector Equations 100%

Many machine learning algorithms boil down to solving a matrix-vector equation. In this chapter, you learn what matrix-vector equations are trying to accomplish and how to solve them in R.

Motivation for Solving Matrix-Vector Equations	50 xp
The Meaning of Ax = b	50 xp
Exploring WNBA Data	100 xp
Matrix-Vector Equations - Some Theory	50 xp
Why is a Matrix Not Invertible?	50 xp
Understanding a Linear System's Three Outcomes	50 xp
Understanding the Massey Matrix	50 xp
Adjusting the Massey Matrix	100 xp
Inverting the Massey Matrix	100 xp
Solving Matrix-Vector Equations	50 xp
An Analogy with Regular Algebra	50 xp
2017 WNBA Ratings!	100 xp
Who Was the Champion?	50 xp
Other Considerations for Matrix-Vector Equations	50 xp
Other Methods for Matrix-Vector Equations	50 xp
Alternatives to the Regular Matrix Inverse	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Eigenvalues and Eigenvectors 100%

Matrix operations are complex. Eigenvalue/eigenvector analyses allow you to decompose these operations into simpler ones for the sake of image recognition, genomic analysis, and more!

Intro to Eigenvalues and Eigenvectors	50 xp
Interpreting Scalar Multiplication	50 xp
Scaling Different Axes	100 xp
Definition of Eigenvalues and Eigenvectors	50 xp
Why "Eigen"?	50 xp
Finding Eigenvalues in R	0 xp
Scalar Multiplies of Eigenvectors are Eigenvectors	100 xp
Computing Eigenvalues and Eigenvectors in R	50 xp
How Many Eigenvalues?	50 xp
Verifying the Math on Eigenvalues	100 xp
Computing Eigenvectors in R	100 xp
Some More on Eigenvalues and Eigenvectors	50 xp
Eigenvalue Ordering	50 xp
Markov Models for Allele Frequencies	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Principal Component Analysis 100%

"Big Data" is ubiquitous in data science and its applications. However, redundancy in these datasets can be problematic. In this chapter, we learn about principal component analysis and how it can be used in dimension reduction.

Intro to the Idea of PCA	50 xp
What Does "Big Data" Mean?	50 xp
Finding Redundancies	0 xp
The Linear Algebra Behind PCA	50 xp
Covariance Explored	50 xp
Standardizing Your Data	100 xp
Variance/Covariance Calculations	100 xp
Eigenanalyses of Combine Data	100 xp
Where's the Variance?	50 xp
Performing PCA in R	50 xp
Scaling Data Before PCA	100 xp
Summarizing PCA in R	100 xp
Does Subsetting Change Things?	50 xp
Wrap-Up	50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

Eric Eager
Data Scientist at Pro Football Focus

Eric Eager is a data scientist for Pro Football Focus, where he analyzes data for all 32 National Football League teams and over 40 college football teams. Before joining PFF in 2018, he was a professor in the Department of Mathematics and Statistics at the University of Wisconsin – La Crosse, where he published over 20 papers in mathematical biology and the scholarship of teaching and learning while securing more than $300,000 in National Science Foundation funding for undergraduate mentorship.

See More

COLLABORATOR(S)

Chester Ismay

David Campos

Shon Inouye

PREREQUISITES

- Introduction to R
- Introduction to the Tidyverse

DATASETS

- NFL Player dataset
- WNBA Massey Matrix dataset
- WNBA Point Differentials dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Multiple and Logistic Regression

Replay Course

4.4 (2,033 Ratings)

4 hours | 19 Videos | 59 Exercises | 17,840 Participants | 4,250 XP

Course Description

Career

It take your skills with simple linear regression to the next level. By learning multiple and logistic ques you will gain the skills to model and predict both numeric and categorical outcomes using multiple input variables. You'll also learn how to fit, visualize, and interpret these models. Then you'll apply your skills to learn about Italian restaurants in New York City!

1 Parallel Slopes FREE 100%

In this chapter you'll learn about the class of linear models called "parallel slopes models." These include one numeric and one categorical explanatory variable.

What if you have two groups?	50 xp
Fitting a parallel slopes model	100 xp
Reasoning about two intercepts	50 xp
Visualizing parallel slopes models	50 xp
Using geom_line() and augment()	100 xp
Interpreting parallel slopes coefficients	50 xp
Intercept interpretation	50 xp
Common slope interpretation	50 xp
Three ways to describe a model	50 xp
Syntax from math	100 xp
Syntax from plot	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Evaluating and extending parallel slopes model 100%

This chapter covers model evaluation. By looking at different properties of the model, including the adjusted R-squared, you'll learn to compare models so that you can select the best one. You'll also learn about interaction terms in linear models.

Model fit, residuals, and prediction	50 xp
R-squared vs. adjusted R-squared	100 xp
Prediction	100 xp
Understanding interaction	50 xp
Thought experiments	50 xp
Fitting a model with interaction	100 xp
Visualizing interaction models	70 xp
Simpson's Paradox	50 xp
Consequences of Simpson's paradox	50 xp
Simpson's paradox in action	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Multiple Regression 100%

This chapter will show you how to add two, three, and even more numeric explanatory variables to a linear model.

Adding a numerical explanatory variable	50 xp
Fitting a MLR model	100 xp
Tiling the plane	100 xp
Models in 3D	100 xp
Conditional interpretation of coefficients	50 xp
Coefficient magnitude	50 xp
Practicing interpretation	50 xp
Adding a third (categorical) variable	50 xp
Visualizing parallel planes	100 xp
Parallel plane interpretation	50 xp
Higher dimensions	50 xp
Interpretation of coefficient in a big model	50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Logistic Regression 100%

In this chapter you'll learn about using logistic regression, a generalized linear model (GLM), to predict a binary outcome and classify observations.

What is logistic regression?	50 xp
Fitting a line to a binary response	100 xp
Fitting a line to a binary response (2)	100 xp
Fitting a model	100 xp
Visualizing logistic regression	50 xp
Using geom_smooth()	100 xp
Using bins	70 xp
Three scales approach to interpretation	50 xp
Odds scale	0 xp
Log-odds scale	70 xp
Interpretation of logistic regression	50 xp
Using a logistic model	50 xp
Making probabilistic predictions	100 xp
Making binary predictions	100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Case Study: Italian restaurants in NYC 100%

Explore the relationship between price and the quality of food, service, and decor for Italian restaurants in NYC.

Italian restaurants in NYC	50 xp
Exploratory data analysis	50 xp
SLR models	100 xp
Incorporating another variable	50 xp
Parallel lines with location	50 xp
A plane in 3D	100 xp
Higher dimensions	50 xp
Parallel planes with location	100 xp
Interpretation of location coefficient	50 xp
Impact of location	100 xp
Full model	50 xp
Wrap-up	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Scientist with R

Statistics Fundamentals with R

Ben Baumer

Assistant Professor at Smith College

Ben is an Assistant Professor in the Statistical & Data Sciences Program at Smith College. He completed his Ph.D. in Mathematics at the Graduate Center of the City University of New York in 2012. He is an Accredited Professional Statistician™ by the American Statistical Association and was previously the Statistical Analyst for the Baseball Operations department of the New York Mets.

See More

COLLABORATOR(S)

Nick Solomon

PREREQUISITES

Correlation and Regression

DATASETS

Average SAT scores by state

New York City Zagat restaurant reviews

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Joining Data in SQL

Replay Course

4.5 (12,814 Ratings)

5 hours | 13 Videos | 53 Exercises | 76,311 Participants | 4,400 XP

Course Description

Now that you've learned the basics of SQL in our Intro to SQL for Data Science course, it's time to supercharge your queries using joins and relational set theory! In this course you'll learn all about the power of joining tables while exploring interesting features of countries and their cities throughout the world. You will master inner and outer joins, as well as self-joins, semi-joins, anti-joins and cross joins - fundamental tools in any PostgreSQL wizard's toolbox. You'll fear set theory no more, after learning all about unions, intersections, and except clauses through easy-to-understand diagrams and examples. Lastly, you'll be introduced to the challenging topic of subqueries. You will see a visual perspective to grasp the ideas throughout the course using the mediums of Venn diagrams and other linking illustrations.

1 Introduction to joins FREE 100%

In this chapter, you'll be introduced to the concept of joining tables, and explore the different ways you can enrich your queries using inner joins and self-joins. You'll also see how to use the case statement to split up a field into different categories.

- Introduction to INNER JOIN — 50 xp
- Inner join — 91 xp
- Inner join (2) — 100 xp
- Inner join (3) — 100 xp
- INNER JOIN via USING — 50 xp
- Review inner join using on — 50 xp
- Inner join with using — 100 xp
- Self-ish joins, just in CASE — 50 xp
- Self-join — 100 xp
- Case when and then — 100 xp
- Inner challenge — 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Outer joins and cross joins 100%

In this chapter, you'll come to grips with different kinds of outer joins. You'll learn how to gain further insights into your data through left joins, right joins, and full joins. In addition to outer joins, you'll also work with cross joins.

- LEFT and RIGHT JOINs — 50 xp
- Left Join — 100 xp
- Left join (2) — 100 xp
- Left join (3) — 91 xp
- Right join — 100 xp
- FULL JOINs — 50 xp
- Full join — 89 xp
- Full join (2) — 100 xp
- Full join (3) — 100 xp
- Review outer joins — 50 xp
- CROSSing the rubicon — 50 xp
- A table of two cities — 70 xp
- Outer challenge — 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Set theory clauses 100%

In this chapter, you'll learn more about set theory using Venn diagrams and you will be introduced to union, union all, intersect, and except clauses. You'll finish by investigating semi-joins and anti-joins, which provide a nice introduction to subqueries.

- State of the UNION — 50 xp
- Union — 100 xp
- Union (2) — 100 xp
- Union all — 100 xp
- INTERSECTional data science — 50 xp
- Intersect — 100 xp
- Intersect (2) — 100 xp
- Review union and intersect — 50 xp
- EXCEPTional — 50 xp
- Except — 100 xp
- Except (2) — 100 xp
- Semi-joins and Anti-joins — 50 xp
- Semi-join — 70 xp
- Relating semi-join to a tweaked inner join — 50 xp
- Diagnosing problems using anti-join — 35 xp
- Set theory challenge — 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Subqueries 100%

In this closing chapter, you'll learn how to use nested queries to add some finesse to your data insights. You'll also wrap all of the content covered throughout this course into solving three challenge problems.

- Subqueries inside WHERE and SELECT clauses — 50 xp
- Subquery inside where — 50 xp
- Subquery inside where (2) — 100 xp
- Subquery inside select — 50 xp
- Subquery inside FROM clause — 50 xp
- Subquery inside from — 85 xp
- Advanced subquery — 59 xp
- Subquery challenge — 70 xp
- Subquery review — 50 xp
- Course review — 50 xp
- Final challenge — 0 xp
- Final challenge (2) — 70 xp
- Final challenge (3) — 0 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

This course is part of these tracks:

- Data Analyst with Python
- Data Analyst with R
- Data Scientist with R
- Data Scientist with Python
- SQL Fundamentals
- SQL Server Fundamentals

Chester Ismay

Head of Content Development (Belgium) at DataCamp

Chester builds (and helps instructors build) R, Python, SQL, and Spreadsheets courses for DataCamp. He obtained a PhD in Statistics from Arizona State University and has taught courses and led workshops in mathematics, computer science, statistics, data science, and sociology. He is co-author of the fivethirtyeight R package and author of the thesistown R package. He is also a co-author of ModernDive, an open-source textbook for introductory statistics and data science students using R.

See More

COLLABORATOR(S)

- Colin Ricardo
- Filip Schouwenaars

PREREQUISITES

- Intro to SQL for Data Science

DATASETS

- Countries
- Leaders
- Diagrams

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Forecasting Using R

Replay Course

4.6 (1,056 Ratings)

5 hours | 18 Videos | 55 Exercises | 22,940 Participants | 4,450 XP

Course Description

Forecasting involves making predictions about the future. It is required in many situations: deciding whether to build another power generation plant in the next ten years requires forecasts of future demand; scheduling staff in a call centre next week requires forecasts of call volumes; stocking an inventory requires forecasts of stock requirements. Forecasts can be required several years in advance (for the case of capital investments), or only a few minutes beforehand (for telecommunication routing). Whatever the circumstances or time horizons involved, forecasting is an important aid to effective and efficient planning. This course provides an introduction to time series forecasting using R.

1 Exploring and visualizing time series in R FREE 100%

The first thing to do in any data analysis task is to plot the data. Graphs enable many features of the data to be visualized, including patterns, unusual observations, and changes over time. The features that are seen in plots of the data must then be incorporated, as far as possible, into the forecasting methods to be used.

Welcome to the course!	50 xp
Creating time series objects in R	100 xp
Time series plots	100 xp
Seasonal plots	100 xp
Trends, seasonality and cyclicity	50 xp
Autocorrelation of non-seasonal time series	100 xp
Autocorrelation of seasonal and cyclic time series	100 xp
Match the ACF to the time series	50 xp
White noise	50 xp
Stock prices and white noise	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Benchmark methods and forecast accuracy 100%

In this chapter, you will learn general tools that are useful for many different forecasting situations. It will describe some methods for benchmark forecasting, methods for checking whether a forecasting method has adequately utilized the available information, and methods for measuring forecast accuracy. Each of the tools discussed in this chapter will be used repeatedly in subsequent chapters as you develop and explore a range of forecasting methods.

Forecasts and potential futures	50 xp
Naive forecasting methods	100 xp
Fitted values and residuals	50 xp
Checking time series residuals	100 xp
Training and test sets	50 xp
Evaluating forecast accuracy of non-seasonal methods	70 xp
Evaluating forecast accuracy of seasonal methods	100 xp
Do I have a good forecasting model?	50 xp
Time series cross-validation	50 xp
Using tsCV() for time series cross-validation	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Exponential smoothing 100%

Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation, the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in business.

Exponentially weighted forecasts	50 xp
Simple exponential smoothing	100 xp
SES vs naive	100 xp
Exponential smoothing methods with trend	50 xp
Holt's trend methods	100 xp
Exponential smoothing methods with trend and seasonality	50 xp
Holt-Winters with monthly data	100 xp
Holt-Winters method with daily data	100 xp
State space models for exponential smoothing	50 xp
Automatic forecasting with exponential smoothing	100 xp
ETS vs seasonal naive	100 xp
Match the models to the time series	50 xp
When does ETS fail?	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Forecasting with ARIMA models 100%

ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely-used approaches to time series forecasting, and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

Transformations for variance stabilization	50 xp
Box-Cox transformations for time series	100 xp
Non-seasonal differencing for stationarity	100 xp
Seasonal differencing for stationarity	70 xp
ARIMA models	50 xp
Automatic ARIMA models for non-seasonal time series	100 xp
Forecasting with ARIMA models	100 xp
Comparing auto.arima() and ets() on non-seasonal data	100 xp
Seasonal ARIMA models	50 xp
Automatic ARIMA models for seasonal time series	100 xp
Exploring auto.arima() options	100 xp
Comparing auto.arima() and ets() on seasonal data	100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Advanced methods 100%

The time series models in the previous chapters work well for many time series, but they are often not good for weekly or hourly data, and they do not allow for the inclusion of other information such as the effects of holidays, competitor activity, changes in the law, etc. In this chapter, you will look at some methods that handle more complicated seasonality, and you consider how to extend ARIMA models in order to allow other information to be included in the them.

Dynamic regression	50 xp
Forecasting sales allowing for advertising expenditure	100 xp
Forecasting electricity demand	70 xp
Dynamic harmonic regression	50 xp
Forecasting weekly data	100 xp
Harmonic regression for multiple seasonality	100 xp
Forecasting call bookings	70 xp
TBATS models	50 xp
TBATS models for electricity demand	100 xp
Your future in forecasting!	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Quantitative Analyst with R

Time Series with R

Rob J. Hyndman

Professor of Statistics at Monash University

Rob J. Hyndman is Professor of Statistics at Monash University, Australia, and Editor-in-Chief of the International Journal of Forecasting. Rob is the author of over 150 research papers and books in statistical science. In 2007, he received the Moran medal from the Australian Academy of Science for his contributions to statistical research, especially in the area of statistical forecasting. He is the author of about 20 R packages, including the popular forecast package.

See More

COLLABORATOR(S)

Lore Dirick

Davis Vaughan

PREREQUISITES

Introduction to R

Intermediate R

DATASETS

ExcelFile in the first exercise

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to R

Practice Now | Replay Course

4.6 (29,860 Ratings)

4 hours | 0 Videos | 62 Exercises | 1,256,139 Participants | 6,200 XP | ALSO AVAILABLE ON MOBILE

Course Description

In this introduction to R, you will master the basics of this beautiful open source language, including factors, lists and data frames. With the knowledge gained in this course, you will be ready to undertake your first very own data analysis. With over 2 million users worldwide R is rapidly becoming the leading programming language in statistics and data science. Every year, the number of R users grows by 40% and an increasing number of organizations are using it in their day-to-day activities. Leverage the power of R by completing this R online course today!

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- R Programmer
- R Programming

Jonathan Cornelissen

Co-founder of DataCamp

Jonathan Cornelissen is one of the co-founders of DataCamp, and is interested in everything related to data science, R, education and entrepreneurship. He holds a PhD in financial econometrics, and is the author of an R package for quantitative finance. DataCamp is the second education start-up he founded, and the first one that went international. In his spare time, he loves to do some programming, especially in Ruby on Rails.

See More

1 Intro to basics — 100%

In this chapter, you will take your first steps with R. You will learn how to use the console as a calculator and how to assign variables. You will also get to know the basic data types in R. Let's get started!

How it works	100 xp
Arithmetic with R	100 xp
Variable assignment	100 xp
Variable assignment (2)	100 xp
Variable assignment (3)	100 xp
Apples and oranges	100 xp
Basic data types in R	100 xp
What's that data type?	100 xp

HIDE CHAPTER DETAILS — Completed

2 Vectors — 100%

In this R course, we'll take you on a trip to Vegas, where you will learn how to analyze your gambling results using vectors in R! After completing this chapter, you will be able to create vectors in R, name them, select elements from them and compare different vectors.

Create a vector	100 xp
Create a vector (2)	100 xp
Create a vector (3)	100 xp
Naming a vector	100 xp
Naming a vector (2)	100 xp
Calculating total winnings	100 xp
Calculating total winnings (2)	100 xp
Calculating total winnings (3)	100 xp
Comparing total winnings	100 xp
Vector selection: the good times	100 xp
Vector selection: the good times (2)	100 xp
Vector selection: the good times (3)	100 xp
Vector selection: the good times (4)	100 xp
Selection by comparison - Step 1	100 xp
Selection by comparison - Step 2	100 xp
Advanced selection	100 xp

HIDE CHAPTER DETAILS — Completed

3 Matrices — 100%

In this chapter you will learn how to work with matrices in R. By the end of the chapter, you will be able to create matrices and to understand how you can do basic computations with them. You will analyze the box office numbers of Star Wars to illustrate the use of matrices in R. May the force be with you!

What's a matrix?	100 xp
Analyze matrices, you shall	70 xp
Naming a matrix	100 xp
Calculating the worldwide box office	100 xp
Adding a column for the Worldwide box office	100 xp
Adding a row	100 xp
The total box office revenue for the entire saga	100 xp
Selection of matrix elements	100 xp
A little arithmetic with matrices	100 xp
A little arithmetic with matrices (2)	100 xp

HIDE CHAPTER DETAILS — Completed

4 Factors — 100%

Very often, data falls into a limited number of categories. For example, human hair color can be categorized as black/brown/blonde/red/gray/white (and perhaps a few more options for people who dye it). In R, categorical data is stored in factors. Given the importance of these factors in data analysis, you should start learning how to create, subset and compare them now!

What's a factor and why would you use it?	100 xp
What's a factor and why would you use it? (2)	100 xp
What's a factor and why would you use it? (3)	100 xp
Factor levels	100 xp
Summarizing a factor	100 xp
Battle of the sexes	100 xp
Ordered factors	100 xp
Ordered factors (2)	100 xp
Comparing ordered factors	100 xp

HIDE CHAPTER DETAILS — Completed

5 Data frames — 100%

Most data sets you will be working with will be stored as data frames. By the end of this chapter focused on R basics, you will be able to create a data frame, select interesting parts of a data frame and order a data frame according to certain variables.

What's a data frame?	100 xp
Quick, have a look at your data set	100 xp
Have a look at the structure	100 xp
Creating a data frame	100 xp
Creating a data frame (2)	100 xp
Selection of data frame elements	100 xp
Selection of data frame elements (2)	100 xp
Only planets with rings	100 xp
Only planets with rings (2)	70 xp
Only planets with rings but shorter	100 xp
Sorting	100 xp
Sorting your data frame	100 xp

HIDE CHAPTER DETAILS — Completed

6 Lists — 100%

Lists, as opposed to vectors, can hold components of different types, just like your to-do list at home or at work. This intro to R chapter will teach you how to create, name and subset these lists.

Lists, why would you need them?	100 xp
Lists, why would you need them? (2)	100 xp
Creating a list	100 xp
Creating a named list	100 xp
Creating a named list (2)	70 xp
Selecting elements from a list	100 xp
Adding more movie information to the list	100 xp

HIDE CHAPTER DETAILS — Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Spatial Statistics in R

Replay Course

4.5 (153 Ratings)

4 hours | 16 Videos | 60 Exercises | 4,965 Participants | 4,950 XP

Course Description

Everything happens somewhere, and increasingly the place where all these things happen is being recorded in a database. There is some truth behind the oft-repeated statement that 80% of data have a spatial component. So what can we do with this spatial data? Spatial statistics, of course! Location is an important explanatory variable in so many things - be it a disease outbreak, an animal's choice of habitat, a traffic collision, or a vein of gold in the mountains - that we would be wise to include it whenever possible. This course will start you on your journey of spatial data analysis. You'll learn what classes of statistical problems present themselves with spatial data, and the basic techniques of how to deal with them. You'll see how to look at a mess of dots on a map and bring out meaningful insights.

1 Introduction FREE

100%

After a quick review of spatial statistics as a whole, you'll go through some point-pattern analysis. You'll learn how to recognize and test different types of spatial patterns.

Problems in spatial statistics	50 xp
Simple spatial principles	100 xp
Plotting areas	0 xp
Uniform in a circle	70 xp
Simulation and testing with spatstat	50 xp
Quadrat count test for uniformity	100 xp
Creating a uniform point pattern with spatstat	100 xp
Simulating clustered and inhibitory patterns	100 xp
Point pattern testing	50 xp
Further testing	50 xp
Nearest-neighbor distributions	100 xp
Other point pattern distribution functions	0 xp
Tree location pattern	50 xp

HIDE CHAPTER DETAILS

Completed

2 Point Pattern Analysis

100%

Point Pattern Analysis answers questions about why things appear where they do. The things could be trees, disease cases, crimes, lightning strikes - anything with a point location.

Bivariate point patterns	50 xp
Crime in Preston	0 xp
Violent crime proportion estimation	70 xp
Spatial segregation	50 xp
Bandwidth selection	100 xp
Segregation probabilities	100 xp
Mapping segregation	0 xp
Space-time data	50 xp
Sasquatch data	100 xp
Spatial pattern of bigfoot sightings	50 xp
Temporal pattern of bigfoot sightings	100 xp
Temporal pattern analysis	50 xp
Space-time clustering	50 xp
Preparing data for space-time clustering	100 xp
Monte-carlo test of space-time clustering	100 xp
Space-time clustering explanation	50 xp

HIDE CHAPTER DETAILS

Completed

3 Areal Statistics

100%

So much data is collected in administrative divisions that there are specialized techniques for analyzing them. This chapter presents several methods for exploring data in areas.

Areal statistics	50 xp
London EU referendum data	100 xp
Cartogram	100 xp
Spatial autocorrelation test	100 xp
Spatial health data	50 xp
London health data	100 xp
Binomial confidence intervals	100 xp
Exceedence probabilities	100 xp
Generalized linear models in space	50 xp
A Poisson GLM	100 xp
Residuals	100 xp
Correlation in spatial GLMs	50 xp
Fit a Bayesian GLM	100 xp
Adding a spatially autocorrelated effect	100 xp
Mapping the spatial effects	70 xp

HIDE CHAPTER DETAILS

Completed

4 Geostatistics

100%

Originally developed for the mining industry, geostatistics covers the analysis of location-based measurement data. It enables model-based interpolation of measurements with uncertainty estimation.

Geostatistical data	50 xp
Canadian geochemical survey data	100 xp
Fitting a trend surface	100 xp
Predicting from a trend surface	100 xp
The variogram	50 xp
Variogram estimation	100 xp
Variogram with spatial trend	100 xp
Variogram model fitting	100 xp
Kriging predictions	50 xp
Filling in the gaps	100 xp
Making a prediction grid	70 xp
Gridded predictions	100 xp
Automatic kriging	50 xp
Auto-kriging at point locations	100 xp
Auto-kriging over a grid	0 xp
Last Words	50 xp

HIDE CHAPTER DETAILS

Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Spatial Data with R

Barry Rowlingson

Research Fellow at Lancaster University

Barry Rowlingson is a Research Fellow in the Lancaster Medical School, part of the Faculty of Health and Medicine at Lancaster University. His primary field of research is statistical applications in epidemiology and other health-related areas. He plays assorted instruments, takes pictures in exotic places, and drives an old Land Rover.

See More

COLLABORATOR(S)

Tom Jeon

Richie Cotton

PREREQUISITES

Introduction to R

Intermediate R

Working with Geospatial Data in R

DATASETS

Canadian geological survey soil acidity

Bounding region for ca_geo.rds

Flu incidents by London borough, 2017

EU referendum voting by London borough, 2016

An OpenStreetMap map of Preston, UK

Crime in Preston, UK

Ph.grid

Sasquatch sightings, 1990 to 2016

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to Time Series Analysis

Replay Course

4.5 (1,323 Ratings)

4 hours | 16 Videos | 58 Exercises | 27,903 Participants | 4,600 XP

Course Description

Many phenomena in our day-to-day lives, such as the movement of stock prices, are measured in intervals over a period of time. Time series analysis methods are extremely useful for analyzing these special data types. In this course, you will be introduced to some core time series analysis concepts and techniques.

1 Exploratory time series data analysis FREE 100%

This chapter will give you insights on how to organize and visualize time series data in R. You will learn several simplifying assumptions that are widely used in time series analysis, and common characteristics of financial time series.

- Welcome to the Course! ✓ 50 xp
- Exploring raw time series ✓ 100 xp
- Basic time series plots ✓ 100 xp
- What does the time index tell us? ✓ 100 xp
- Sampling Frequency ✓ 50 xp
- Identifying the sampling frequency ✓ 100 xp
- When is the sampling frequency exact? ✓ 50 xp
- Missing values ✓ 100 xp
- Basic Time Series Objects ✓ 50 xp
- Creating a time series object with ts() ✓ 100 xp
- Testing whether an object is a time series ✓ 100 xp
- Plotting a time series object ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Predicting the future 100%

In this chapter, you will conduct some trend spotting, and learn the white noise (WN) model, the random walk (RW) model, and the definition of stationary processes.

- Trend Spotting! ✓ 50 xp
- Random or not random? ✓ 50 xp
- Name that trend ✓ 50 xp
- Removing trends in variability via the logarithmic transformation ✓ 100 xp
- Removing trends in level by differencing ✓ 100 xp
- Removing seasonal trends with seasonal differencing ✓ 100 xp
- The White Noise (WN) Model ✓ 50 xp
- Simulate the white noise model ✓ 100 xp
- Estimate the white noise model ✓ 100 xp
- The Random Walk (RW) Model ✓ 50 xp
- Simulate the random walk model ✓ 100 xp
- Simulate the random walk model with a drift ✓ 100 xp
- Estimate the random walk model ✓ 100 xp
- Stationary Processes ✓ 50 xp
- Stationary or not? ✓ 50 xp
- Are the white noise model or the random walk model stationary? ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Correlation analysis and the autocorrelation function 100%

In this chapter, you will review the correlation coefficient, use it to compare two time series, and also apply it to compare a time series with its past, as an autocorrelation. You will discover the autocorrelation function (ACF) and practice estimating and visualizing autocorrelations for time series data.

- Scatterplots ✓ 50 xp
- Asset prices vs. asset returns ✓ 100 xp
- Characteristics of financial time series ✓ 100 xp
- Plotting pairs of data ✓ 100 xp
- Covariance and Correlation ✓ 50 xp
- Calculating sample covariances and correlations ✓ 100 xp
- Guess the correlation coefficient ✓ 50 xp
- Autocorrelation ✓ 50 xp
- Calculating autocorrelations ✓ 70 xp
- The autocorrelation function ✓ 70 xp
- Visualizing the autocorrelation function ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Autoregression 100%

In this chapter, you will learn the autoregressive (AR) model and several of its basic properties. You will also practice simulating and estimating the AR model in R, and compare the AR model with the random walk (RW) model.

- The Autoregressive Model ✓ 50 xp
- Simulate the autoregressive model ✓ 100 xp
- Estimate the autocorrelation function (ACF) for an autoregression ✓ 100 xp
- Persistence and anti-persistence ✓ 50 xp
- Compare the random walk (RW) and autoregressive (AR) models ✓ 70 xp
- AR Model Estimation and Forecasting ✓ 50 xp
- Estimate the autoregressive (AR) model ✓ 100 xp
- Simple forecasts from an estimated AR model ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 A simple moving average 100%

In this chapter, you will learn the simple moving average (MA) model and several of its basic properties. You will also practice simulating and estimating the MA model in R, and compare the MA model with the autoregressive (AR) model.

- The Simple Moving Average Model ✓ 50 xp
- Simulate the simple moving average model ✓ 100 xp
- Estimate the autocorrelation function (ACF) for a moving average ✓ 100 xp
- MA Model Estimation and Forecasting ✓ 50 xp
- Estimate the simple moving average model ✓ 100 xp
- Simple forecasts from an estimated MA model ✓ 100 xp
- Compare the AR and MA Models ✓ 50 xp
- AR vs MA models ✓ 100 xp
- Name that model by time series plot ✓ 35 xp
- Name that model by ACF plot ✓ 50 xp
- Congratulations! ✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT View

This course is part of these tracks:

Quantitative Analyst with R

Time Series with R

David S. Matteson

Associate Professor at Cornell University

David S. Matteson is Professor of Statistical Science at Cornell University and co-author of Statistics and Data Analysis for Financial Engineering with R examples.

See More

COLLABORATOR(S)

Lore Dirick

Matt Isaacs

PREREQUISITES

Introduction to R

Intermediate R

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Intermediate R

Practice Now | Replay Course

★★★★★ 4.5 (13,919 Ratings)

4 hours | 14 Videos | 81 Exercises | 313,391 Participants | 6,950 XP

Course Description

The intermediate R course is the logical next stop on your journey in the R programming language. In this R training you will learn about conditional statements, loops and functions to power your own R scripts. Next, you can make your R code more efficient and readable using the apply functions. Finally, the utilities chapter gets you up to speed with regular expressions in the R programming language, data structure manipulations and times and dates. This R tutorial will allow you to learn R and take the next step in advancing your overall knowledge and capabilities while programming in R.

1 Conditionals and Control Flow FREE

100%

To be TRUE or not be TRUE, that's the question. In this chapter you'll learn about relational operators to see how R objects compare and logical operators to combine logicals. Next, you'll use this knowledge to build conditional statements.

Relational Operators	✓ 50 xp
Equality	✓ 100 xp
Greater and less than	✓ 100 xp
Compare vectors	✓ 100 xp
Compare matrices	✓ 100 xp
Logical Operators	✓ 50 xp
& and |	✓ 100 xp
& and | (2)	✓ 100 xp
Reverse the result: !	✓ 50 xp
Blend it all together	✓ 200 xp
Conditional Statements	✓ 50 xp
The if statement	✓ 100 xp
Add an else	✓ 100 xp
Customize further: else if	✓ 100 xp
Else if 2.0	✓ 50 xp
Take control!	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Loops

100%

Loops can come in handy on numerous occasions. While loops are like repeated if statements; the for loop is designed to iterate over all elements in a sequence. Learn all about them in this chapter.

While loop	✓ 50 xp
Write a while loop	✓ 100 xp
Throw in more conditionals	✓ 100 xp
Stop the while loop: break	✓ 100 xp
Build a while loop from scratch	✓ 100 xp
For loop	✓ 50 xp
Loop over a vector	✓ 100 xp
Loop over a list	✓ 100 xp
Loop over a matrix	✓ 100 xp
Mix it up with control flow	✓ 70 xp
Next, you break it	✓ 100 xp
Build a for loop from scratch	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Functions

100%

Functions are an extremely important concept in almost every programming language; R is not different. After learning what a function is and how you can use one, you'll take full control by writing your own functions.

Introduction to Functions	✓ 50 xp
Function documentation	✓ 100 xp
Use a function	✓ 100 xp
Use a function (2)	✓ 0 xp
Use a function (3)	✓ 100 xp
Functions inside functions	✓ 100 xp
Required, or optional?	✓ 50 xp
Writing Functions	✓ 50 xp
Write your own function	✓ 100 xp
Write your own function (2)	✓ 100 xp
Write your own function (3)	✓ 70 xp
Function scoping	✓ 50 xp
R passes arguments by value	✓ 50 xp
R you functional?	✓ 100 xp
R you functional? (2)	✓ 70 xp
R Packages	✓ 50 xp
Load an R Package	✓ 100 xp
Different ways to load a package	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

4 The apply family

100%

Whenever you're using a for loop, you might want to revise your code and see whether you can use the lapply function instead. Learn all about this intuitive way of applying a function over a list or a vector, and its variants sapply and vapply.

lapply	✓ 50 xp
Use lapply with a built-in R function	✓ 70 xp
Use lapply with your own function	✓ 100 xp
lapply and anonymous functions	✓ 100 xp
Use lapply with additional arguments	✓ 70 xp
Apply functions that return NULL	✓ 50 xp
sapply	✓ 50 xp
How to use sapply	✓ 100 xp
sapply with your own function	✓ 70 xp
sapply with function returning vector	✓ 100 xp
sapply can't simplify, now what?	✓ 100 xp
sapply with functions that return NULL	✓ 100 xp
Reverse engineering sapply	✓ 50 xp
vapply	✓ 50 xp
Use vapply	✓ 100 xp
Use vapply (2)	✓ 100 xp
From sapply to vapply	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Utilities

100%

Mastering R programming is not only about understanding its programming concepts. Also a solid knowledge of a wide range of R functions is useful. This chapter introduces you to a bunch of useful functions for data structure manipulation, regular expressions and working with times and dates.

Useful Functions	✓ 50 xp
Mathematical utilities	✓ 100 xp
Find the error	✓ 70 xp
Data Utilities	✓ 70 xp
Find the error (2)	✓ 100 xp
Beat Gauss using R	✓ 100 xp
Regular Expressions	✓ 50 xp
grepl & grep	✓ 100 xp
grepl & grep (2)	✓ 100 xp
sub & gsub	✓ 100 xp
sub & gsub (2)	✓ 50 xp
Times and Dates	✓ 50 xp
Right here, right now	✓ 100 xp
Create and format dates	✓ 100 xp
Create and format times	✓ 100 xp
Calculations with Dates	✓ 100 xp
Calculations with Times	✓ 100 xp
Time is of the essence	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with R
- Data Scientist with R
- R Programmer
- R Programming

Filip Schouwenaars
Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

PREREQUISITES

Introduction to R

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Python for R Users

Replay Course

4.5 (407 Ratings)

5 hours | 15 Videos | 57 Exercises | 5,926 Participants | 4,950 XP

Course Description

Python and R have seen immense growth in popularity in the "Machine Learning Age". They both are high-level languages that are easy to learn and write. The language you use will depend on your background and field of study and work. R is a language made by and for statisticians, whereas Python is a more general purpose programming language. Regardless of the background, there will be times when a particular algorithm is implemented in one language and not the other, a feature is better documented, or simply, the tutorial you found online uses Python instead of R. In either case, this would require the R user to work in Python to get his/her work done, or try to understand how something is implemented in Python for it to be translated into R. This course helps you cross the R-Python language barrier.

1 The Basics FREE

100%

Learn about some of the most important data types (integers, floats, strings, and booleans) and data structures (lists, dictionaries, numpy arrays, and pandas DataFrames) in Python and how they compare to the ones in R.

Introduction	✓ 50 xp
Assignment and data types	✓ 100 xp
Arithmetic with strings	✓ 100 xp
Containers	✓ 50 xp
Lists	✓ 100 xp
Dictionaries	✓ 100 xp
Functions, Methods, and Libraries	✓ 50 xp
Methods	✓ 100 xp
NumPy arrays	✓ 100 xp
Pandas DataFrames	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Control flow, Loops, and Functions

100%

This chapter covers control flow statements (if-else if-else), for loops and shows you how to write your own functions in Python!

Control flow and loops	✓ 50 xp
Control flow	✓ 100 xp
Loops	✓ 100 xp
Functions	✓ 50 xp
Individual binge drinking function	✓ 100 xp
General binge drinking function	✓ 100 xp
Lambda functions	✓ 70 xp
Comprehensions	✓ 50 xp
Mapping functions	✓ 100 xp
List comprehension	✓ 70 xp
Dictionary comprehension	✓ 0 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Pandas

100%

In this chapter you will learn more about one of the most important Python libraries, Pandas. In addition to DataFrames, pandas provides several data manipulation functions and methods.

Selecting data in pandas	✓ 50 xp
Selecting columns	✓ 100 xp
Selecting rows	✓ 100 xp
Selecting rows and columns	✓ 100 xp
Data types	✓ 50 xp
Integers and floats	✓ 100 xp
Strings	✓ 100 xp
Category	✓ 100 xp
Dates (I)	✓ 100 xp
Dates (II)	✓ 100 xp
Advanced Pandas	✓ 50 xp
Missing values	✓ 68 xp
Groupby	✓ 100 xp
Tidy data	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Plotting

100%

You will learn about the rich ecosystem of visualization libraries in Python. This chapter covers matplotlib, the core visualization library in Python along with the pandas and seaborn libraries.

Plotting directly using pandas	✓ 50 xp
Univariate plots in pandas	✓ 100 xp
Bivariate plots in pandas	✓ 100 xp
Seaborn	✓ 50 xp
Univariate plots in seaborn	✓ 100 xp
Bivariate plots in seaborn	✓ 100 xp
Facet plots in seaborn	✓ 100 xp
Matplotlib	✓ 50 xp
Univariate and bivariate plots in matplotlib	✓ 100 xp
Subfigures in matplotlib	✓ 100 xp
Working with axes	✓ 100 xp
Polishing up a figure	✓ 100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

5 Capstone

100%

As a final capstone, you will apply your Python skills on the NYC Flights 2013 dataset.

NYC flights data	✓ 50 xp
Load multiple data files	✓ 0 xp
Explore	✓ 100 xp
Visualize	✓ 100 xp
Manipulating data	✓ 50 xp
Recode dates	✓ 85 xp
Groupby aggregates	✓ 100 xp
Plots	✓ 100 xp
Dummy variables	✓ 100 xp
Wrap-up	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

Daniel Chen

Data Science Consultant at Lander Analytics

Daniel is a Software Carpentry instructor and a doctoral student in Genetics, Bioinformatics, and Computational Biology at Virginia Tech, where he works in the Social and Decision Analytics Laboratory under the Biocomplexity Institute. He received his MPH at the Mailman School of Public Health in Epidemiology and is interested in integrating hospital data in order to perform predictive health analytics and build clinical support tools for clinicians. An advocate of open science, he aspires to bridge data science with epidemiology and health care.

See More

COLLABORATOR(S)

Hugo Bowne-Anderson

Eunkyung Park

Sumedh Panchadhar

PREREQUISITES

Writing Functions in R

DATASETS

Air quality

Country timeseries

Flights (sample)

Inflammation

NYC Flights 2013

Tips

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Hierarchical and Mixed Effects Models

Replay Course

4.2 (310 Ratings)

4 hours | 14 Videos | 54 Exercises | 4,960 Participants | 4,600 XP

Course Description

This course begins by reviewing slopes and intercepts in linear regressions before moving on to random-effects. You'll learn what a random effect is and how to use one to model your data. Next, the course covers linear mixed-effect regressions. These powerful models will allow you to explore data with a more complicated structure than a standard linear regression. The course then teaches generalized linear mixed-effect regressions. Generalized linear mixed-effects models allow you to model more kinds of data, including binary responses and count data. Lastly, the course goes over repeated-measures analysis as a special case of mixed-effect modeling. This kind of data appears when subjects are followed over time and measurements are collected at intervals. Throughout the course you'll work with real data to answer interesting questions using mixed-effects models.

STATEMENT OF ACCOMPLISHMENT

View

1 Overview and introduction to hierarchical and mixed models

FREE 58%

The first chapter provides an example of when to use a mixed-effect and also describes the parts of a regression. The chapter also examines a a student test-score dataset with a nested structure to demonstrate mixed-effects.

What is a hierarchical model?	✓ 50 xp
Explore multi-level data: students	✓ 100 xp
Exploring multiple-levels: Classrooms and schools	100 xp
Parts of a regression	✓ 50 xp
Intercepts	✓ 100 xp
Slopes	✓ 0 xp
Random-effects in regressions	✓ 50 xp
Random-effect intercepts	100 xp
Random-effect slopes	100 xp
School data	✓ 50 xp
Analyzing and interpreting the school data	100 xp
Plotting and interpreting results	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Linear mixed-effect models

69%

This chapter providers an introduction to linear mixed-effects models. It covers different types of random-effects, describes how to understand the results for linear mixed-effects models, and goes over different methods for statistical inference with mixed-effects models using crime data from Maryland.

Linear mixed effect model- Birth rates data	✓ 50 xp
Building a lmer model with random effects	✓ 0 xp
Including a fixed effect	✓ 0 xp
Random-effect slopes	100 xp
Uncorrelated random-effect slope	100 xp
Fixed- and random-effect predictor	100 xp
Understanding and reporting the outputs of a lmer	✓ 50 xp
Comparing print and summary output	✓ 50 xp
Extracting coefficients	✓ 100 xp
Displaying the results from a lmer model	✓ 100 xp
Statistical inference with Maryland crime data	✓ 50 xp
Visualizing Maryland crime data	✓ 0 xp
Rescaling slopes	✓ 70 xp
Null hypothesis testing	100 xp
Controversies around P-values	✓ 50 xp
Model comparison with ANOVA	100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Generalized linear mixed-effect models

100%

This chapter extends linear mixed-effects models to include non-normal error terms using generalized linear mixed-effects models. By altering the model to include a non-normal error term, you are able to model more kinds of data with non-linear responses. After reviewing generalized linear models, the chapter examines binomial data and count data in the context of mixed-effects models.

Crash course on GLMs	✓ 50 xp
Logistic regression	✓ 0 xp
Poisson Regression	✓ 100 xp
Plotting GLMs	✓ 0 xp
Binomial data	✓ 50 xp
Toxicology data	✓ 0 xp
Marketing example	✓ 0 xp
Calculating odds-ratios	✓ 100 xp
Count data	✓ 50 xp
Internet click-throughs	✓ 0 xp
Chlamydia by age-group and county	✓ 0 xp
Displaying chlamydia results	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Repeated Measures

50%

This chapter shows how repeated-measures analysis is a special case of mixed-effect modeling. The chapter begins by reviewing paired t-tests and repeated measures ANOVA. Next, the chapter uses a linear mixed-effect model to examine sleep study data. Lastly, the chapter uses a generalized linear mixed-effect model to examine hate crime data from New York state through time.

An introduction to repeated measures	✓ 50 xp
Paired t-test	✓ 0 xp
Repeated measures ANOVA	✓ 0 xp
Sleep study	✓ 50 xp
Exploring the data	100 xp
Building models	100 xp
Comparing regressions and ANOVAs	100 xp
Plotting results	100 xp
Hate in NY state?	✓ 50 xp
Exploring NY hate data	100 xp
Building the model	100 xp
Interpreting model results	100 xp
Displaying the results	✓ 0 xp
Conclusion	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

Richard Erickson
Quantitative Ecologist

Richard helps people to experience and understand their increasingly numerical world. For his day job he develops new quantitative methods for monitoring and controlling invasive species as well as helping other scientists analyze and understand their data. He has worked on diverse datasets ranging from continent wide species distributions to pesticides in playa wetlands. After hours, he teaches SCUBA Diving as a NAUI Instructor. He has been a "UserR" since 2007.

See More

COLLABORATOR(S)

Chester Ismay

Nick Solomon

PREREQUISITES

Generalized Linear Models in R

DATASETS

Illinois chlamydia data
Maryland crime data
Classroom data
Birth rate data
New York hate crime data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Single-Cell RNA-Seq Workflows in R

Replay Course

4 hours | 12 Videos | 50 Exercises | 1,666 Participants | 4,100 XP

Course Description

Novel single-cell transcriptome sequencing assays allow researchers to measure gene expression levels at the resolution of single cells and offer the unprecedented opportunity to investigate fundamental biological questions at the cellular level, such as stem cell differentiation or the discovery and characterization of rare cell types. The majority of the computational methods to analyze single-cell RNA-Seq data are implemented in R making it a natural tool to start working with single-cell transcriptomic data. Using real single-cell datasets, this course provides a step-by-step tutorial to the methodology and associated R packages for the following four main tasks: (1) normalization, (2) dimensionality reduction, (3) clustering, (4) differential expression analysis.

STATEMENT OF ACCOMPLISHMENT

View in

Fanny Perraudeau

Senior Data Scientist, Whole Biome

Fanny Perraudeau is a Senior Data Scientist at Whole Biome where she manages, designs, and implements novel genomics algorithms and bioinformatics pipelines to further improve the analysis of of Whole Biome microbiome data. In addition, she runs statistical analyses to aid the company's therapeutic discovery efforts. She has a master from Ecole Polytechnique, France and a PhD in Biostatistics from University of California, Berkeley with a Designated Emphasis in Computational and Genomic Biology. Much of her work is motivated by the development and application of statistical methods and software for the analysis of biomedical and genomic data, especially metagenomics and single-cell RNASeq.

See More

COLLABORATOR(S)

Richie Cotton

David Campos

Shon Inouye

PREREQUISITES

Intermediate R

Introduction to Bioconductor

DATASETS

Mouse Epithelium dataset

Toy dataset

Tung dataset

1 What is Single-Cell RNA-Seq? FREE — 100%

In Chapter 1, you will learn what single-cell RNA-Seq is and why it is a such a powerful technique. By the end of this chapter, you'll also know how to load, create, and access single-cell datasets in R.

What is Single-Cell RNA-Seq, and why is it useful?	50 xp
Bulk versus Single-cell RNA-Seq	50 xp
Toy scRNA-Seq dataset	100 xp
Explore dataset	100 xp
Typical workflow	50 xp
GC content	100 xp
Library size	50 xp
Nesting between batches and biology	100 xp
Load, create, and access single-cell datasets in R	50 xp
SCE object from counts matrix	100 xp
SCE object from SummarizedExperiment	100 xp
Load a single-cell dataset in R	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 Quality Control and Normalization — 100%

In Chapter 2, we go over the first steps of the workflow to analyze single-cell RNA-seq data, which include quality control and normalization. These two steps should get all the technical issues and biases out of the way so that in the next chapters we can focus on the biological signal of interest.

Quality Control	50 xp
Explore Tung dataset	100 xp
Calculate QC metrics	100 xp
Filter cells with small library size	100 xp
Quality Control (continued)	50 xp
Filter cells by number of expressed genes	100 xp
Use of positive controls	100 xp
Filter genes mainly not expressed	100 xp
Normalization	50 xp
Batch effect	100 xp
Correlation between PC1 and library size	100 xp
Compute size factors	100 xp
Normalize SCE object	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Visualization and Dimensionality Reduction — 100%

When studying single-cell data at the cellular level, the number of dimensions is the number of genes. The goal of dimensionality reduction is to reduce the number of dimensions to a smaller number either to visualize the data in 2 dimensions or to prepare the dataset for subsequent steps like clustering.

Mouse Epithelium Dataset	50 xp
Explore dataset	100 xp
Nested experiment design	100 xp
Cell differentiation	50 xp
Visualization	50 xp
Plot PCA of counts	100 xp
Plot PCA of log counts	100 xp
Plot t-SNE of log counts	100 xp
Dimensionality Reduction	50 xp
Subset sce	100 xp
Perform PCA on log counts	100 xp
Plot PCA using ggplot	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 Cell Clustering and Differential expression analysis — 100%

In Chapter 4, we cluster cells with similar gene expression profiles and then perform differential expression (DE) analysis to find genes differentially expressed between known groups of cells. We then visualize DE genes with volcano plots and heatmaps.

Clustering methods for scRNA-Seq	50 xp
Create Seurat object	100 xp
Perform PCA on Seurat object	100 xp
Perform clustering	50 xp
Refine clustering	50 xp
Differential expression analysis	50 xp
Fit zero-inflated regression using MAST	100 xp
Create result table	100 xp
Compute adjusted p-values	100 xp
Visualization of DE genes	50 xp
Plot volcano plot	100 xp
Understand volcano plot	50 xp
Plot heatmap	100 xp

HIDE CHAPTER DETAILS — ✓ Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Object-Oriented Programming in R: S3 and R6

Replay Course

4.4 (386 Ratings)

4 hours | 18 Videos | 56 Exercises | 9,786 Participants | 4,250 XP

Course Description

Object-oriented programming (OOP) lets you specify relationships between functions and the objects that they can act on, helping you manage complexity in your code. This is an intermediate-level course, providing an introduction to OOP, using the S3 and R6 systems. S3 is a great day-to-day R programming tool that simplifies some of the functions that you write. R6 is especially useful for industry-specific analyses, working with web APIs, and building GUIs. The course concludes with an interview with Winston Chang, creator of the R6 package.

1 Introduction to Object-Oriented Programming FREE 100%

Learn what object-oriented programming (OOP) consists of, when to use it, and what OOP systems are available in R. You'll also learn how R identifies different types of variable, using classes, types, and modes.

The Objectives of Object-Oriented Programming	✓ 50 xp
Should I OOP?	✓ 50 xp
You've Already Been Working With Objects	✓ 100 xp
The Nine Systems	✓ 50 xp
Which Systems Should I Use?	✓ 50 xp
How Does R Distinguish Types of Variable?	✓ 50 xp
What's my type?	✓ 100 xp
Assigning Classes and Implicit Classes	✓ 50 xp
Make it Classy (1)	✓ 100 xp
Make it Classy (2)	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Using S3 100%

S3 is a very simple object-oriented system that lets you define different behavior for functions, depending upon their input argument. This chapter explains how to use S3, and how generics and methods work.

Function Overload	✓ 50 xp
What's in a Name?	✓ 50 xp
Creating a Generic Function	✓ 100 xp
Creating an S3 Method (1)	✓ 70 xp
Creating an S3 method (2)	✓ 100 xp
Methodical Thinking	✓ 50 xp
Finding Available Methods (1)	✓ 100 xp
Finding Available Methods (2)	✓ 50 xp
S3 and Primitive Functions	✓ 50 xp
Method Lookup for Primitive Generics	✓ 100 xp
Who is Calling?	✓ 50 xp
Too Much Class	✓ 50 xp
Very Classy	✓ 100 xp
Writing the Next Method	✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Using R6 100%

Learn how to define R6 classes, and to create R6 objects. You'll also learn about the structure of R6 classes, and how to separate the user interface from the implementation details.

The Object Factory	✓ 50 xp
Specifying the Microwave Oven Class	✓ 100 xp
Making Microwave Ovens	✓ 100 xp
Hiding Complexity with Encapsulation	✓ 50 xp
Learning to Cook	✓ 100 xp
Close the Door	✓ 70 xp
First Thing's First	✓ 100 xp
Getting and Setting with Active Bindings	✓ 50 xp
Read the Rating	✓ 100 xp
Control the Power	✓ 70 xp

HIDE CHAPTER DETAILS ✓ Completed

4 R6 Inheritance 100%

Learn how to inherit from an R6 class, and how the relationship between parent and child classes works.

Propagating Functionality with Inheritance	✓ 50 xp
Specifying a Fancy Microwave Oven	✓ 100 xp
Who's the Daddy?	✓ 50 xp
Making a Fancy Microwave Oven	✓ 100 xp
Embrace, Extend, Override	✓ 50 xp
Extending the Cooking Capabilities	✓ 100 xp
Overriding the Cooking Capabilities	✓ 100 xp
Multiple Levels of Inheritance	✓ 50 xp
Exposing your Parent	✓ 100 xp
Over-Overriding the Cooking Capabilities	✓ 70 xp
Remembering Your Origins	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

5 Advanced R6 Usage 100%

Complete your mastery of R6 by learning about advanced topics such as copying by reference, shared fields, cloning objects, and finalizing objects. The chapter concludes with an interview with Winston Chang, creator of the R6 package.

Environments, Reference Behaviour, and Static Fields	✓ 50 xp
Working with Environments (1)	✓ 0 xp
Working with Environments (2)	✓ 0 xp
Static Electricity	✓ 70 xp
Cloning Objects	✓ 50 xp
Attack of the Clones (1)	✓ 100 xp
Attack of the Clones (2)	✓ 100 xp
Shut it Down	✓ 50 xp
The Final Countdown	✓ 50 xp
This is the End (of an R6 Object)	✓ 0 xp
Interview with Winston Chang	✓ 50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

R Programmer

Richie Cotton
Head of Content Quality at DataCamp

Richie runs the Content Quality team at DataCamp. He has been using R since 2004, in the fields of proteomics, debt collection, and chemical health and safety. He has released almost 30 R packages on CRAN and Bioconductor – most famously the assertive suite of packages – as well as creating and contributing to many others. He also has written two books on R programming, Learning R and Testing R Code.

See More

PREREQUISITES

Introduction to R
Intermediate R

DATASETS

Cooking times (SQLite file)

© 2019 DataCamp Inc.

INTERACTIVE COURSE

RNA-Seq Differential Expression Analysis

Replay Course

4.4 (269 Ratings)

4 hours | 16 Videos | 44 Exercises | 2,349 Participants | 3,150 XP

Course Description

RNA-Seq is an exciting next-generation sequencing method used for identifying genes and pathways underlying particular diseases or conditions. As high-throughput sequencing becomes more affordable and accessible to a wider community of researchers, the knowledge to analyze this data is becoming an increasingly valuable skill. Join us in learning about the RNA-Seq workflow and discovering how to identify which genes and biological processes may be important for your condition of interest! We will start the course with a brief overview of the RNA-Seq workflow with an emphasis on differential expression (DE) analysis. Starting with the counts for each gene, the course will cover how to prepare data for DE analysis, assess the quality of the count data, and identify outliers and detect major sources of variation in the data. The DESeq2 R package will be used to model the count data using a negative binomial model and test for differentially expressed genes. Visualization of the results with heatmaps and volcano plots will be performed and the significant differentially expressed genes will be identified and saved.

STATEMENT OF ACCOMPLISHMENT

View

1 Introduction to RNA-Seq theory and workflow FREE

100%

In this chapter we explore what we can do with RNA-Seq data and why it is exciting. We learn about the different steps and considerations involved in an RNA-Seq workflow.

Introduction to RNA-Seq	50 xp
Core Concepts	50 xp
RNA-Seq Packages	100 xp
RNA-Seq Workflow	50 xp
Read Alignments	50 xp
Exploring the raw count matrix	100 xp
Differential Gene Expression Theory	50 xp
DGE Theory	50 xp
DGE Theory: Metadata	100 xp

HIDE CHAPTER DETAILS

Completed

2 Exploratory data analysis

100%

In this chapter, we perform quality control on the RNA-Seq count data using heatmaps and principal component analysis. We explore the similarity of the samples to each other and determine whether there are any sample outliers.

Introduction to differential expression analysis	50 xp
Practice with the DESeq2 vignette	50 xp
Organizing the data for DESeq2	50 xp
Matching metadata and counts data	70 xp
Count normalization	50 xp
Normalizing counts with DESeq2	100 xp
Hierarchical heatmap	50 xp
Hierarchical heatmap by condition	100 xp
Hierarchical heatmap analysis	50 xp
Principal component analysis	50 xp
PCA analysis	100 xp
PCA practice: exploring variations	50 xp
PCA practice: exploring additional variations	50 xp

HIDE CHAPTER DETAILS

Completed

3 Differential expression analysis with DESeq2

100%

In this chapter, we execute the differential expression analysis, generate results and identify the differentially expressed genes.

DE analysis	50 xp
Creating the DE object	100 xp
DESeq2 model - dispersion	50 xp
DESeq2 model - exploring dispersions	100 xp
Interpreting the dispersion plot	50 xp
DESeq2 model - contrasts	50 xp
DESeq2 model - extracting results	100 xp
DESeq2 results - LFC shrinkage	100 xp
DESeq2 results	50 xp
DESeq2 results exploration	100 xp
Summarizing DESeq2 results	100 xp
DESeq2 significant results	100 xp

HIDE CHAPTER DETAILS

Completed

4 Exploration of differential expression results

100%

In this final chapter we explore the differential expression results using visualizations, such as heatmaps and volcano plots. We also review the steps in the analysis and summarize the differential expression workflow with DESeq2.

Visualization of results	50 xp
DESeq2 visualizations - MA and volcano plots	100 xp
DESeq2 visualizations - heatmap	70 xp
RNA-Seq DE analysis summary - setup	50 xp
RNA-Seq DE analysis - experimental planning	50 xp
RNA-Seq DE workflow summary	100 xp
RNA-Seq DE analysis summary	50 xp
DE analysis	100 xp
DE analysis results	100 xp
RNA-Seq next steps	50 xp

HIDE CHAPTER DETAILS

Completed

Mary Piper

Bioinformatics Consultant and Trainer

Mary Piper serves dual roles as research analyst and bioinformatics trainer in the Department of Biostatistics at the Harvard T.H. Chan School of Public Health. However, her primary role is the development and instruction of bioinformatics workshops focused on the analysis of next-generation sequencing data. She has a PhD in cellular and molecular biology from the University of Michigan and a background in science education. Her passion for bioinformatics research and teaching led to her desire to pursue bioinformatics as a career and to share that knowledge with the community.

See More

COLLABORATOR(S)

Richie Cotton

David Campos

Shon Inouye

PREREQUISITES

Introduction to Bioconductor

Data Visualization with ggplot2 (Part 1)

DATASETS

Fibrosis raw counts dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Science for Managers

Replay Course

4 hours | 14 Videos | 51 Exercises | 519 Participants | 3,350 XP

Course Description

What is data science and how can you use it to strengthen your organization? This course will teach you about the skills you need on your data team, and how you can structure that team to meet your organization's needs. Data is everywhere! This course will provide you with an understanding of data sources your company can use and how to store that data. You'll also discover ways to analyze and visualize your data through dashboards and A/B tests. To wrap up the course, we'll discuss exciting topics in machine learning, including clustering, time series prediction, natural language processing (NLP), deep learning, and explainable AI! Along the way, you'll learn about a variety of real-world applications of data science and gain a better understanding of these concepts through practical exercises.

1 Introduction to Data Science FREE — 100%

We'll start the course by defining what data science is. We'll cover the data science workflow, and how data science is applied to real-world business problems. We'll finish the chapter by learning about ways to structure your data team to meet your organization's needs.

What is Data Science?		50 xp
Customer Segmentation Workflow		100 xp
Building a customer service chatbot		100 xp
Improving OKRs		50 xp
Applications of Data Science		50 xp
Assigning data science projects		100 xp
Investment research		50 xp
Building a data science team		50 xp
Interpreting a team sprint		50 xp
Editing a job posting		50 xp
Matching skills to jobs		100 xp
Classifying data tasks		100 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 Data Collection and Storage — 100%

Now that we understand the data science workflow, we'll dive deeper into the first step: data collection. We'll learn about the different data sources your company can draw from, and how to store that data once it's collected.

Data sources and risks		50 xp
Classifying data for security		100 xp
Creating web data events		50 xp
Protecting PII		50 xp
Solicited data		50 xp
Identifying question purpose		100 xp
Validating focus group feedback		50 xp
Net Promoter Score		50 xp
Collecting additional data		50 xp
Sorting data sources		100 xp
Asthma frequency		50 xp
Data storage and retrieval		50 xp
Cloud platforms		50 xp
Querying a database		50 xp
Which type of database?		100 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Analysis and Visualization — 100%

In this chapter, we'll discuss ways to explore and visualize data through dashboards. We'll discuss the elements of a dashboard and how to make a directed request for a dashboard. This chapter will also cover making ad hoc data requests and A/B tests, which are a powerful analytics tool that de-risk decision-making.

Dashboards		50 xp
Classifying dashboard elements		50 xp
Improving a dashboard		50 xp
Choosing the right dashboard		100 xp
Ad hoc analysis		50 xp
Filling out an ad hoc request		50 xp
Classifying requests		100 xp
A/B Testing		50 xp
Creating an A/B testing workflow		100 xp
Sample size		50 xp
Intermediate results		50 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 Prediction — 100%

In this final chapter, we'll discuss the buzziest topic in data science: machine learning! We'll cover supervised and unsupervised machine learning, and clustering. Then, we'll move on to special topics in machine learning, including time series prediction, natural language processing, deep learning, and explainable AI!

Supervised machine learning		50 xp
When to use Supervised Learning		100 xp
Features and labels		50 xp
Model evaluation		50 xp
Clustering		50 xp
Supervised vs. unsupervised		100 xp
Cluster size selection		50 xp
Special topics in Machine Learning		50 xp
Classifying machine learning tasks		100 xp
Sentiment Analysis		50 xp
Deep Learning and Explainable AI		50 xp
Finding the correct solution		100 xp
Should I use Deep Learning?		50 xp

HIDE CHAPTER DETAILS — ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View in

Mari Nazary
VP of Content at DataCamp

Mari Nazary is a global EdTech executive who partners with educators and subject matter experts to build and scale effective, outcomes-focused, SaaS learning solutions. After spending over a decade working in EdTech for multimillion dollar brands and startups, Mari knows what truly closes the skills gap across the world—and it's not mastering the marketing flavor of the week. It's how well you understand your learners' needs in order to help them measure and achieve real-world success. Mari has designed digital learning solutions for worldwide audiences including Rosetta Stone, EF Education First, and DataCamp. In addition to her instructional design and curriculum development expertise, Mari is a certified Agile scrum master, Python programmer, and data analyst. Mari holds an MA in Linguistics from Middlebury College and a BA from Barnard College in Classics.

See More

Michael Chow
Assessment Research Lead at DataCamp

Michael is a data scientist at DataCamp, where he develops models for adaptive assessment. He has programmed in python and R for a little over a decade, and received a PhD in cognitive psychology from Princeton University. His research interests include statistical methods, skill acquisition, and human memory. You can follow him on twitter @chowthedog.

See More

kaelen medeiros
Data Scientist, Product at DataCamp

Kaelen is the Data Scientist for the Product team at DataCamp. They are an admin for the R-Ladies Global community. Kaelen received a MS in Biostatistics from Louisiana State University Health Sciences Center, where they worked at the Louisiana Tumor Registry. Before DataCamp, they designed experiments (and more!) for the American College of Surgeons, HERE Technologies, and HealthLabs. If you meet them, you will undoubtedly hear about their cat, Scully, within the first 3 minutes. Other favorite topics include aliens, popcorn, podcasts, and nail polish.

See More

Ramnath Vaidyanathan
VP of Product Research at DataCamp

Ramnath Vaidyanathan is the VP of Product Research at DataCamp, where he drives product innovation and data-driven development. He has 10+ years experience doing statistical modeling, machine learning, optimization, retail analytics, and interactive visualizations. He brings a unique perspective to product development, having worked in diverse industries like management consulting, academia, and enterprise software. Prior to joining DataCamp, he worked as a data scientist at Alteryx, leading the roadmap for interactive visualizations and dashboards for predictive analytics. Prior to Alteryx, he was an Assistant Professor of Operations Management in the Desautels Faculty of Management at McGill University. His research primarily focused on the application of predictive analytics and optimization methodologies to improve operational decisions in retailing. He got his Ph.D. in Operations Management from the Wharton School.

See More

COLLABORATOR(S)

Amy Peterson

Hillary Green-Lerman

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Visualization with ggplot2 (Part 3)

Replay Course

4.5 (500 Ratings)

6 hours | 19 Videos | 86 Exercises | 12,081 Participants | 7,550 XP

Course Description

In this third ggplot2 course, we'll dive into some advanced topics including grams commonly used in maths and sciences, strategies for handling large data sets, a variety of specialty plots, and some useful features of ggplot2 internals.

1 Statistical plots FREE — 100%

Actually, all the plots you've explored in the first two ggplot2 courses can be considered 'statistical plots'. Here, however, you'll consider those that are intended for a specialist audience that is familiar with the data: box plots and density plots.

- Introduction — 50 xp
- Aesthetics review — 50 xp
- Refresher (1) — 100 xp
- Refresher (2) — 100 xp
- Refresher (3) — 100 xp
- Box Plots — 50 xp
- Transformations — 100 xp
- Cut it up! — 100 xp
- Understanding quartiles — 50 xp
- Density Plots — 50 xp
- geom_density() — 100 xp
- Combine density plots and histogram — 100 xp
- Adjusting density plots — 100 xp
- Multiple Groups/Variables — 50 xp
- Box plots with varying width — 100 xp
- Multiple density plots — 100 xp
- Multiple density plots (2) — 70 xp
- Weighted density plots — 100 xp
- 2D density plots (1) — 100 xp
- 2D density plots (2) — 100 xp

HIDE CHAPTER DETAILS — Completed

2 Plots for specific data types (Part 1) — 100%

In this chapter, you'll explore useful specialty plots for specific data types such as ternary plots, networks and maps. You'll also look at how to use ggplot2 to convert typical base package plots that are used to evaluate the results of statistical methods. Finally, you'll take a look at a couple ways in which you can make and appropriately use animations.

- Graphics of Large Data — 50 xp
- Pair plots and correlation matrices — 100 xp
- Create a correlation matrix in ggplot2 — 100 xp
- Ternary Plots — 50 xp
- Proportional/stacked bar plots — 100 xp
- Producing ternary plots — 100 xp
- Adjusting ternary plots — 100 xp
- Network Plots — 50 xp
- Build the network (1) — 100 xp
- Build the network (2) — 100 xp
- Adjusting the network — 100 xp
- Diagnostic Plots — 50 xp
- Autoplot on linear models — 100 xp
- ggfortify - time series — 100 xp
- Distance matrices and Multi-Dimensional Scaling (MDS) — 100 xp
- Plotting K-means clustering — 100 xp

HIDE CHAPTER DETAILS — Completed

3 Plots for specific data types (Part 2) — 100%

In this chapter, we'll continue our discussion of plots for specific data types by diving into the world of maps. You'll also have a look at animations to make your data come to life!

- Choropleths — 50 xp
- Working with maps from the maps package: USA — 100 xp
- Working with maps from the maps package: adding points — 100 xp
- State choropleth — 100 xp
- Map from shapefiles — 100 xp
- Choropleth from shapefiles — 100 xp
- Cartographic Maps — 50 xp
- Different templates — 100 xp
- Mapping points onto a cartographic map — 100 xp
- Combine cartographic and choropleth maps — 100 xp
- Animations — 50 xp
- The population pyramid — 100 xp
- Animations with gganimate — 100 xp

HIDE CHAPTER DETAILS — Completed

4 ggplot2 Internals — 100%

In this chapter, you'll delve into ggplot2 internals, exploring the grid package and ggproto. You'll learn how to use these tools to create unique plots.

- Grid Graphics — 50 xp
- Viewport basics (1) — 100 xp
- Viewport basics (2) — 100 xp
- Build a plot from scratch (1) — 100 xp
- Build a plot from scratch (2) — 100 xp
- Modifying a plot with grid.edit — 100 xp
- Grid Graphics in ggplot2 — 50 xp
- Exploring the gTable — 100 xp
- Modifying the gTable — 100 xp
- ggplot Objects — 50 xp
- Exploring ggplot objects — 100 xp
- ggplot_build and ggplot_gtable — 100 xp
- Extracting details — 100 xp
- gridExtra — 50 xp
- Arranging plots (1) — 100 xp
- Arranging plots (2) — 100 xp

HIDE CHAPTER DETAILS — Completed

5 Data Munging and Visualization Case Study — 100%

In this chapter, you'll draw on some of the many tools for effective data visualization that we've covered over the three ggplot2 courses and combine them with some data munging techniques.

- Case Study I - Bag Plot — 50 xp
- Base package bag plot — 100 xp
- Multilayer ggplot2 bag plot — 100 xp
- Creating ggproto functions — 100 xp
- Creating stat_bag() — 100 xp
- Use stat_bag() — 100 xp
- Case Study II - Weather (Part 1) — 50 xp
- Step 1: Read in data and examine — 100 xp
- Step 2: Summarize history — 100 xp
- Step 3: Plot history — 100 xp
- Step 4: Plot present — 100 xp
- Step 5: Find new record highs — 100 xp
- Step 6: Efficiently calculate record highs and lows — 70 xp
- Step 7: Custom legend — 100 xp
- Case Study II - Weather (Part 2) — 50 xp
- Step 1: clean_weather() — 100 xp
- Step 2: Historical data — 100 xp
- Step 3: Present data — 70 xp
- Step 4: Extremes — 100 xp
- Step 5: Re-use plotting style — 70 xp
- Wrap-up — 50 xp

HIDE CHAPTER DETAILS — Completed

STATEMENT OF ACCOMPLISHMENT

View

Rick Scavetta

Rick Scavetta is a co-founder of Scavetta Academy.

Rick Scavetta is a biologist, workshop trainer, freelance data scientist and cofounder of Scavetta Academy, a company dedicated to helping scientists better understand and visualize their data. Rick's practical, hands-on exposure to a wide variety of datasets has informed him of the many problems scientists face when trying to visualize their data.

See More

COLLABORATOR(S)

Filip Schouwenaars

PREREQUISITES

- Introduction to R
- Intermediate R
- Data Visualization with ggplot2 (Part 1)
- Data Visualization with ggplot2 (Part 2)

DATASETS

- Movies (subset of 10000 observations)
- Test datasets
- Mammals
- Africa
- US Cities
- US States
- Germany unemployment data
- Population of Japan
- Shape files
- Paris weather data
- Reykavik weather data
- New York weather data
- London weather data

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Intermediate Python for Data Science

Practice Now | Replay Course

4.6 (32,417 Ratings)

4 hours | 18 Videos | 87 Exercises | 344,527 Participants | 7,400 XP | ALSO AVAILABLE ON MOBILE

Course Description

The intermediate python course is crucial to your data science curriculum. Learn to visualize real data with matplotlib's functions and get to know new data structures such as the dictionary and the Pandas DataFrame. After covering key concepts such as boolean logic, control flow and loops in Python, you're ready to blend together everything you've learned to solve a case study using hacker statistics.

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

- Data Analyst with Python
- Data Scientist with Python
- Python Programmer
- Python Programming

Filip Schouwenaars

Data Science Instructor at DataCamp

Filip is the passionate developer behind several of DataCamp's interactive courses, covering both R and Python. Under the motto 'Eat your own dog food', he has used the techniques DataCamp teaches its students to perform data analysis for DataCamp. Filip holds degrees in Electrical Engineering and Artificial Intelligence.

See More

COLLABORATOR(S)

- Vincent Vankrunkelsven
- Patrick Varilly
- Florian Goossens

PREREQUISITES

- Introduction to Python

DATASETS

- Gapminder
- Cars
- BRICS

1 Matplotlib FREE — 100%

Data Visualization is a key skill for aspiring data scientists. Matplotlib makes it easy to create meaningful and insightful plots. In this chapter, you will learn to build various types of plots and to customize them to make them more visually appealing and interpretable.

Basic plots with matplotlib	✓ 50 xp
Line plot (1)	✓ 100 xp
Line Plot (2): Interpretation	✓ 50 xp
Line plot (3)	✓ 100 xp
Scatter Plot (1)	✓ 100 xp
Scatter plot (2)	✓ 100 xp
Histograms	✓ 50 xp
Build a histogram (1)	✓ 100 xp
Build a Histogram (2): bins	✓ 100 xp
Build a histogram (3): compare	✓ 100 xp
Choose the right plot (1)	✓ 50 xp
Choose the right plot (2)	✓ 50 xp
Customization	✓ 50 xp
Labels	✓ 100 xp
Ticks	✓ 100 xp
Sizes	✓ 100 xp
Colors	✓ 100 xp
Additional Customizations	✓ 100 xp
Interpretation	✓ 50 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

2 Dictionaries & Pandas — 100%

Learn about the dictionary, an alternative to the Python list, and the Pandas DataFrame, the de facto standard to work with tabular data in Python. You will get hands-on practice with creating, manipulating and accessing the information you need from these data structures.

Dictionaries, Part 1	✓ 50 xp
Motivation for dictionaries	✓ 100 xp
Create dictionary	✓ 100 xp
Access dictionary	✓ 100 xp
Dictionaries, Part 2	✓ 50 xp
Dictionary Manipulation (1)	✓ 100 xp
Dictionary Manipulation (2)	✓ 100 xp
Dictionarception	✓ 70 xp
Pandas, Part 1	✓ 50 xp
Dictionary to DataFrame (1)	✓ 100 xp
Dictionary to DataFrame (2)	✓ 100 xp
CSV to DataFrame (1)	✓ 100 xp
CSV to DataFrame (2)	✓ 100 xp
Pandas, Part 2	✓ 50 xp
Square Brackets (1)	✓ 100 xp
Square Brackets (2)	✓ 100 xp
loc and iloc (1)	✓ 100 xp
loc and iloc (2)	✓ 100 xp
loc and iloc (3)	✓ 70 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

3 Logic, Control Flow and Filtering — 100%

Boolean logic is the foundation of decision-making in your Python programs. Learn about different comparison operators, how you can combine them with boolean operators and how to use the boolean outcomes in control structures. You'll also learn to filter data from Pandas DataFrames using logic.

Comparison Operators	✓ 50 xp
Equality	✓ 100 xp
Greater and less than	✓ 100 xp
Compare arrays	✓ 100 xp
Boolean Operators	✓ 50 xp
and, or, not (1)	✓ 100 xp
and, or, not (2)	✓ 50 xp
Boolean operators with Numpy	✓ 100 xp
if, elif, else	✓ 50 xp
Warmup	✓ 50 xp
if	✓ 100 xp
Add else	✓ 100 xp
Customize further: elif	✓ 100 xp
Filtering Pandas DataFrame	✓ 50 xp
Driving right (1)	✓ 100 xp
Driving right (2)	✓ 100 xp
Cars per capita (1)	✓ 70 xp
Cars per capita (2)	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

4 Loops — 100%

There are several techniques to repeatedly execute Python code. While loops are like repeated if statements; the for loop is there to iterate over all kinds of data structures. Learn all about them in this chapter.

while loop	✓ 50 xp
while: warming up	✓ 50 xp
Basic while loop	✓ 100 xp
Add conditionals	✓ 70 xp
for loop	✓ 50 xp
Loop over a list	✓ 100 xp
Indexes and values (1)	✓ 0 xp
Indexes and values (2)	✓ 70 xp
Loop over list of lists	✓ 70 xp
Looping Data Structures, Part 1	✓ 50 xp
Loop over dictionary	✓ 100 xp
Loop over Numpy array	✓ 70 xp
Looping Data Structures, Part 2	✓ 50 xp
Loop over DataFrame (1)	✓ 100 xp
Loop over DataFrame (2)	✓ 70 xp
Add column (1)	✓ 100 xp
Add column (2)	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

5 Case Study: Hacker Statistics — 100%

This chapter blends together everything you've learned up to now. You will use hacker statistics to calculate your chances of winning a bet. Use random number generators, loops and matplotlib to get the competitive edge!

Random Numbers	✓ 50 xp
Random float	✓ 100 xp
Roll the dice	✓ 100 xp
Determine your next move	✓ 100 xp
Random Walk	✓ 50 xp
The next step	✓ 70 xp
How low can you go?	✓ 100 xp
Visualize the walk	✓ 100 xp
Distribution	✓ 50 xp
Simulate multiple walks	✓ 100 xp
Visualize all walks	✓ 100 xp
Implement clumsiness	✓ 100 xp
Plot the distribution	✓ 100 xp
Calculate the odds	✓ 35 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Data Visualization in R

Replay Course

4.2 (609 Ratings)

4 hours | 15 Videos | 60 Exercises | 32,262 Participants | 5,250 XP

Course Description

This course provides a comprehensive introduction on how to plot data with R's default graphics system, base graphics.

After an introduction to base graphics, we look at a number of R plotting examples, from simple graphs such as scatterplots to plotting correlation matrices. The course finishes with exercises in plot customization. This includes using R plot colors effectively and creating and saving complex plots in R.

Base Graphics Background

R supports four different graphics systems: base graphics, grid graphics, lattice graphics, and ggplot2. Base graphics is the default graphics system in R, the easiest of the four systems to learn to use, and provides a wide variety of useful tools, especially for exploratory graphics where we wish to learn what is in an unfamiliar dataset.

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Visualization with R

1 A quick introduction to base R graphics FREE

100%

This chapter gives a brief overview of some of the things you can do with base graphics in R. This graphics system is one of four available in R and it forms the basis for this course because it is both the easiest to learn and extremely useful both in preparing exploratory data visualizations to help you see what's in a dataset and in preparing explanatory data visualizations to help others see what we have found.

- The world of data visualization — 50 xp
- Creating an exploratory plot array — 100 xp
- Creating an explanatory scatterplot — 100 xp
- The plot() function is generic — 100 xp
- A preview of some more and less useful techniques — 50 xp
- Adding details to a plot using point shapes, color, and reference lines — 100 xp
- Creating multiple plot arrays — 100 xp
- Avoid pie charts — 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

2 Different plot types

100%

This chapter introduces several Base R supported plot types that are particularly useful for visualizing important features in a dataset. We start with simple tools like histograms and density plots for characterizing one variable at a time, move on to scatter plots and other useful tools for showing how two variables relate, and finally introduce some tools for visualizing more complex relationships in our dataset.

- Characterizing a single variable — 50 xp
- The hist() and truehist() functions — 100 xp
- Density plots as smoothed histograms — 70 xp
- Using the qqPlot() function to see many details in data — 0 xp
- Visualizing relations between two variables — 50 xp
- The sunflowerplot() function for repeated numerical data — 100 xp
- Useful options for the boxplot() function — 100 xp
- Using the mosaicplot() function — 100 xp
- Showing more complex relations between variables — 69 xp
- Using the bagplot() function — 0 xp
- Plotting correlation matrices with the corrplot() function — 100 xp
- Building and plotting rpart() models — 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

3 Adding details to plots

100%

Most base R graphics functions support many optional arguments and parameters that allow us to customize our plots to get exactly what we want. In this chapter, we will learn how to modify point shapes and sizes, line types and widths, add points and lines to plots, add explanatory text and generate multiple plot arrays.

- The plot() function and its options — 50 xp
- Introduction to the par() function — 100 xp
- Exploring the type option — 100 xp
- The surprising utility of the type "n" option — 100 xp
- Adding lines and points to plots — 50 xp
- The lines() function and line types — 100 xp
- The points() function and point types — 100 xp
- Adding trend lines from linear regression models — 100 xp
- Adding text to plots — 50 xp
- Using the text() function to label plot features — 70 xp
- Adjusting text position, size, and font — 70 xp
- Rotating text with the srt argument — 100 xp
- Adding or modifying other plot details — 50 xp
- Using the legend() function — 100 xp
- Adding custom axes with the axis() function — 100 xp
- Using the supsmu() function to add smooth trend curves — 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

4 How much is too much?

100%

As we have seen, base R graphics provides tremendous flexibility in creating plots with multiple lines, points of different shapes and sizes, and added text, along with arrays of multiple plots. If we attempt to add too many details to a plot or too many plots to an array, however, the result can become too complicated to be useful. This chapter focuses on how to manage this visual complexity so the results remain useful to ourselves and to others.

- Managing visual complexity — 50 xp
- Too much is too much — 100 xp
- Deciding how many scatterplots is too many — 100 xp
- How many words is too many? — 100 xp
- Creating plot arrays with the mfrow parameter — 50 xp
- The Anscombe quartet — 100 xp
- The utility of common scaling and individual titles — 100 xp
- Using multiple plots to give multiple views of a dataset — 100 xp
- Creating plot arrays with the layout() function — 50 xp
- Constructing and displaying layout matrices — 100 xp
- Creating a triangular array of plots — 100 xp
- Creating arrays with different sized plots — 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

5 Advanced plot customization and beyond

100%

This final chapter introduces a number of important topics, including the use of numerical plot details returned invisibly by functions like barplot() to enhance our plots, and saving plots to external files so they don't vanish when we end our current R session. This chapter also offers some guidelines for using color effectively in data visualizations, and it concludes with a brief introduction to the other three graphics systems in R.

- Creating and saving more complex plots — 50 xp
- Some plot functions also return useful information — 100 xp
- Using the symbols() function to display relations between more than two variables — 100 xp
- Saving plot results as files — 100 xp
- Using color effectively — 50 xp
- Ilinsky and Steele's 12 recommended colors — 100 xp
- Using color to enhance a bubbleplot — 70 xp
- Using color to enhance stacked barplots — 100 xp
- Other graphics systems in R — 50 xp
- The tabplot package and grid graphics — 100 xp
- A lattice graphics example — 100 xp
- A ggplot2 graphics example — 100 xp

HIDE CHAPTER DETAILS — ✓ Completed

Ronald Pearson

PhD in Electrical Engineering and Computer Science from M.I.T.

Ron has been actively involved in data analysis and predictive modeling in a variety of technical positions, both academic and commercial, including the DuPont Company, the Swiss Federal Institute of Technology (ETH Zurich), the Tampere University of Technology in Tampere, Finland, the Travelers Companies and DataRobot. He holds a PhD in Electrical Engineering and Computer Science from M.I.T. and has written or co-written five books, including **Exploring Data in Engineering, the Sciences, and Medicine** (Oxford University Press, 2011) and **Nonlinear Digital Filtering with Python** (CRC Press, 2016, with Moncef Gabbouj). Ron is the author and maintainer of the GoodmanKruskal R package, and one of the authors of the datarobot R package.

See More

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

Introduction to R

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Mixture Models in R

Replay Course

4.5 (153 Ratings)

4 hours | 14 Videos | 47 Exercises | 1,621 Participants | 3,600 XP

Course Description

Mixture modeling is a way of representing populations when we are interested in their heterogeneity. Mixture models use familiar probability distributions (e.g. Gaussian, Poisson, Binomial) to provide a convenient yet formal statistical framework for clustering and classification. Unlike standard clustering approaches, we can estimate the probability of belonging to a cluster and make inference about the sub-populations. For example, in the context of marketing, you may want to cluster different customer groups and find their respective probabilities of purchasing specific products to better target them with custom promotions. When applying natural language processing to a large set of documents, you may want to cluster documents into different topics and understand how important each topic is across each document. In this course, you will learn what Mixture Models are, how they are estimated, and when it is appropriate to apply them!

1 Introduction to Mixture Models FREE 100%

In this chapter, you will be introduced to fundamental concepts in model-based clustering and how this approach differs from other clustering techniques. You will learn the generating process of Gaussian Mixture Models as well as how to visualize the clusters.

- Introduction to model-based clustering ✓ 50 xp
- Clustering approaches ✓ 50 xp
- Explore gender data ✓ 100 xp
- Gaussian distribution ✓ 50 xp
- Sampling a Gaussian distribution ✓ 100 xp
- (not so good) Estimations of the mean and sd ✓ 100 xp
- Gaussian mixture models (GMM) ✓ 50 xp
- Simulate a mixture of two Gaussian distributions ✓ 100 xp
- Plot histogram of Gaussian Mixture ✓ 100 xp
- Mixture of three Gaussian distributions ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Structure of Mixture Models and Parameters Estimation 100%

In this chapter, you will be introduced to the main structure of Mixture Models, how to address different data with this approach and how to estimate the parameters involved. To accomplish the estimation, you will learn an iterative method called Expectation-Maximization algorithm.

- Structure of mixture models ✓ 50 xp
- Which probability distribution? ✓ 50 xp
- Handwritten digits dataset ✓ 100 xp
- Parameters estimation ✓ 50 xp
- Estimation given the probabilities ✓ 100 xp
- Calculating the probabilities ✓ 100 xp
- EM algorithm ✓ 50 xp
- Expectation function ✓ 100 xp
- Maximization function ✓ 100 xp
- Apply the two steps ✓ 100 xp
- Plot the estimated clusters ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Mixture of Gaussians with `flexmix` 100%

This chapter shows how to fit Gaussian Mixture Models in 1 and 2 dimensions with `flexmix` package. The data used is formed by 10.000 observations of people with their weight, height, body mass index and informed gender.

- Univariate Gaussian Mixture Models ✓ 50 xp
- Number of clusters ✓ 50 xp
- Number of parameters ✓ 50 xp
- Univariate Gaussian Mixture Models with `flexmix` ✓ 50 xp
- Univariate case with flexmix ✓ 100 xp
- Extracting Parameters for Univariate Case ✓ 100 xp
- Visualizing Univariate Gaussian Mixture Model ✓ 100 xp
- Compare the results ✓ 100 xp
- Bivariate Gaussian Mixture Models ✓ 50 xp
- Cross-term from covariance matrix ✓ 50 xp
- Parameters in the bivariate case ✓ 50 xp
- Bivariate Gaussian Mixture Models with `flexmix` ✓ 50 xp
- Fit the model with cross-terms ✓ 100 xp
- Get the components ✓ 100 xp
- Create the ellipses ✓ 100 xp
- Visualize the clusters ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Mixture Models Beyond Gaussians 100%

In this module, you will learn how Mixture Models extends to consider probability distributions different from the Gaussian and how these models are fitted with `flexmix`. The datasets used are handwritten digits images and the number of crimes in Chicago city. For the first dataset you will find clusters that summarize the handwritten digits and for the second dataset, you will find clusters of communities where is more or less dangerous to live in.

- Bernoulli Mixture Models ✓ 50 xp
- Binary images ✓ 100 xp
- How many values? ✓ 50 xp
- Bernoulli Mixture Models with `flexmix` ✓ 50 xp
- Handwritten digits with `flexmix` ✓ 100 xp
- Poisson Mixture Models ✓ 50 xp
- Discover the lambda ✓ 50 xp
- Sample from Poisson distribution ✓ 100 xp
- Poisson Mixture Models with `flexmix` ✓ 50 xp
- Crimes data with `flexmix` ✓ 100 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Probability and Distributions with R

Victor Medina
Doctoral Researcher at The University of Edinburgh

Victor is a doctoral researcher at the University of Edinburgh. With a background in Applied Statistics, he has worked researching and developing predictive risk models and mathematical tools for the Financial System. With a passion for teaching, he was a professor of Econometrics at the Universidad de Chile and, currently, works with the School of Political Sciences at Universidad Central to create applications to extract and analyze data from social media. Victor holds an MSc in Statistics, BEng, BSc in Physics and R is his second language.

See More

COLLABORATOR(S)

- Chester Ismay
- David Campos
- Benjamin Feder
- Shon Inouye

PREREQUISITES

- Intermediate R
- Introduction to the Tidyverse
- Foundations of Probability in R

DATASETS

- Chicago Crimes dataset
- Digits dataset
- Gender dataset

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Machine Learning Toolbox

Replay Course

4.5 (1,809 Ratings)

4 Hours | 24 Videos | 88 Exercises | 31,166 Participants | 6,200 XP

Course Description

Machine learning is the study and application of algorithms that learn from and make predictions on data. From search results to self-driving cars, it has manifested itself in all areas of our lives and is one of the most exciting and fast growing fields of research in the world of data science. This course teaches the big ideas in machine learning: how to build and evaluate predictive models, how to tune them for optimal performance, how to preprocess data for better results, and much more. The popular `caret` R package, which provides a consistent interface to all of R's most powerful machine learning facilities, is used throughout the course.

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Scientist with R

Machine Learning Fundamentals in R

Zachary Deane-Mayer

Automation First Data Scientist at DataRobot

Zach is a Data Scientist at DataRobot and co-author of the caret R package. He's fascinated by predicting the future and spends his free time competing in predictive modeling competitions. He's currently one of top 500 data scientists on Kaggle and took 9th place in the Heritage Health Prize as part of the Analytics Inside team.

See More

Max Kuhn

Software Engineer at RStudio and creator of caret.

Dr. Max Kuhn is a Software Engineer at RStudio. He is the author or maintainer of several R packages for predictive modeling including caret, AppliedPredictiveModeling, Cubist, C50 and SparseLDA. He routinely teaches classes in predictive modeling at Predictive Analytics World and UseR! and his publications include work on neuroscience biomarkers, drug discovery, molecular diagnostics and response surface methodology.

See More

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

Introduction to R

Intermediate R

Correlation and Regression

DATASETS

Diamonds

Sonar

Wine

Overfit data

Breast Cancer

Blood-brain

Churn

1 Regression models: fitting them and evaluating their performance

FREE 100%

In the first chapter of this course, you'll fit regression models with `train()` and evaluate their out-of-sample performance using cross-validation and root-mean-square error (RMSE).

Welcome to the Toolbox	50 xp
In-sample RMSE for linear regression	✓ 50 xp
In-sample RMSE for linear regression on diamonds	100 xp
Out-of-sample error measures	50 xp
Out-of-sample RMSE for linear regression	✓ 50 xp
Randomly order the data frame	100 xp
Try an 80/20 split	100 xp
Predict on test set	100 xp
Calculate test set RMSE by hand	100 xp
Comparing out-of-sample RMSE to in-sample RMSE	✓ 50 xp
Cross-validation	✓ 50 xp
Advantage of cross-validation	✓ 50 xp
10-fold cross-validation	100 xp
5-fold cross-validation	100 xp
5 x 5-fold cross-validation	100 xp
Making predictions on new data	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

2 Classification models: fitting them and evaluating their performance

100%

In this chapter, you'll fit classification models with `train()` and evaluate their out-of-sample performance using cross-validation and area under the curve (AUC).

Logistic regression on sonar	✓ 50 xp
Why a train/test split?	✓ 50 xp
Try a 60/40 split	100 xp
Fit a logistic regression model	100 xp
Confusion matrix	✓ 50 xp
Confusion matrix takeaways	✓ 50 xp
Calculate a confusion matrix	100 xp
Calculating accuracy	✓ 50 xp
Calculating true positive rate	✓ 50 xp
Calculating true negative rate	✓ 50 xp
Class probabilities and predictions	50 xp
Probabilities and classes	✓ 50 xp
Try another threshold	100 xp
From probabilities to confusion matrix	100 xp
Introducing the ROC curve	50 xp
What's the value of a ROC curve?	✓ 50 xp
Plot an ROC curve	100 xp
Area under the curve (AUC)	50 xp
Model, ROC, and AUC	✓ 50 xp
Customizing trainControl	100 xp
Using custom trainControl	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

3 Tuning model parameters to improve performance

96%

In this chapter, you will use the `train()` function to tweak model parameters through cross-validation and grid search.

Random forests and wine	✓ 50 xp
Random forests vs. linear models	✓ 50 xp
Fit a random forest	100 xp
Explore a wider model space	50 xp
Advantage of a longer tune length	✓ 50 xp
Try a longer tune length	100 xp
Custom tuning grids	✓ 50 xp
Advantages of a custom tuning grid	✓ 50 xp
Fit a random forest with custom tuning	✓ 0 xp
Introducing glmnet	✓ 50 xp
Advantage of glmnet	✓ 50 xp
Make a custom trainControl	100 xp
Fit glmnet with custom trainControl	100 xp
glmnet with custom tuning grid	✓ 50 xp
Why a custom tuning grid?	✓ 50 xp
glmnet with custom trainControl and tuning	100 xp
Interpreting glmnet plots	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

4 Preprocessing your data

94%

In this chapter, you will practice using `train()` to preprocess data before fitting models, improving your ability to making accurate predictions.

Median imputation	✓ 50 xp
Median imputation vs. omitting rows	✓ 50 xp
Apply median imputation	100 xp
KNN imputation	✓ 50 xp
Comparing KNN imputation to median imputation	✓ 50 xp
Use KNN imputation	100 xp
Compare KNN and median imputation	✓ 50 xp
Multiple preprocessing methods	✓ 50 xp
Order of operations	✓ 50 xp
Combining preprocessing methods	✓ 0 xp
Handling low-information predictors	✓ 50 xp
Why remove near zero variance predictors?	✓ 50 xp
Remove near zero variance predictors	100 xp
preProcess() and nearZeroVar()	✓ 50 xp
Fit model on reduced blood-brain data	100 xp
Principle components analysis (PCA)	50 xp
Using PCA as an alternative to nearZeroVar()	100 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

5 Selecting models: a case study in churn prediction

88%

In the final chapter of this course, you'll learn how to use `resamples()` to compare multiple models and select (or ensemble) the best one(s).

Reusing a trainControl	✓ 50 xp
Why reuse a trainControl?	✓ 50 xp
Make custom train/test indices	100 xp
Reintroducing glmnet	50 xp
glmnet as a baseline model	✓ 50 xp
Fit the baseline model	100 xp
Reintroducing random forest	50 xp
Random forest drawback	✓ 50 xp
Random forest with custom trainControl	100 xp
Comparing models	✓ 50 xp
Matching train/test indices	✓ 50 xp
Create a resamples object	100 xp
More on resamples	✓ 50 xp
Create a box-and-whisker plot	100 xp
Create a scatterplot	100 xp
Ensembling models	100 xp
Summary	✓ 50 xp

HIDE CHAPTER DETAILS ^ ✓ Completed

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Machine Learning in the Tidyverse

Replay Course

4.6 (473 Ratings)

5 hours | 15 Videos | 52 Exercises | 2,791 Participants | 4,300 XP

Course Description

This course will teach you to leverage the tools in the "tidyverse" to generate, explore, and evaluate machine learning models. Using a combination of tidyr and purrr packages, you will build a foundation for how to work with complex model objects in a "tidy" way. You will also learn how to leverage the broom package to explore your resulting models. You will then be introduced to the tools in the test-train-validate workflow, which will empower you evaluate the performance of both classification and regression models as well as provide the necessary information to optimize model performance via hyperparameter tuning.

1 Foundations of "tidy" Machine learning FREE 100%

This chapter will introduce you to the backbone of machine learning in the tidyverse, the List Column Workflow (LCW). The LCW will empower you to work with many models in one dataframe.
This chapter will also introduce you to the fundamentals of the broom package for exploring your models.

Foundations of "tidy" machine learning	50 xp
Nesting your data	100 xp
Unnesting your data	100 xp
Explore a nested cell	100 xp
The map family of functions	50 xp
Mapping your data	100 xp
Expecting mapped output	100 xp
Mapping many models	70 xp
Tidy your models with broom	50 xp
The three ways to tidy your model	50 xp
Extracting model statistics tidily	100 xp
Augmenting your data	100 xp

HIDE CHAPTER DETAILS ✓ Completed

2 Multiple Models with broom 100%

This chapter leverages the List Column Workflow to build and explore the attributes of 77 models. You will use the tools from the broom package to gain a multidimensional understanding of all of these models.

Exploring coefficients across models	50 xp
Tidy up the coefficients of your models	100 xp
What can we learn about these 77 countries?	50 xp
Evaluating the fit of many models	50 xp
Glance at the fit of your models	100 xp
Best and worst fitting models	100 xp
Visually inspect the fit of many models	50 xp
Augment the fitted values of each model	100 xp
Explore your best and worst fitting models	50 xp
Improve the fit of your models	50 xp
Build better models	100 xp
Predicting the future	50 xp

HIDE CHAPTER DETAILS ✓ Completed

3 Build, Tune & Evaluate Regression Models 100%

In this chapter you will learn how to use the List Column Workflow to build, tune and evaluate regression models. You will have the chance to work with two types of models: linear models and random forest models.

Training, test and validation splits	50 xp
The test-train split	100 xp
Cross-validation dataframes	100 xp
Measuring cross-validation performance	50 xp
Build cross-validated models	0 xp
Preparing for evaluation	100 xp
Evaluate model performance	100 xp
Building and tuning a random forest model	50 xp
Build a random forest model	100 xp
Evaluate a random forest model	100 xp
Fine tune your model	100 xp
The best performing parameter	100 xp
Measuring the test performance	50 xp
Build & evaluate the best model	100 xp

HIDE CHAPTER DETAILS ✓ Completed

4 Build, Tune & Evaluate Classification Models 100%

In this chapter you will shift gears to build, tune and evaluate classification models.

Logistic regression models	50 xp
Prepare train-test-validate parts	100 xp
Build cross-validated models	100 xp
Evaluating classification models	50 xp
Predictions of a single model	100 xp
Performance of a single model	100 xp
Prepare for cross-validated performance	0 xp
Calculate cross-validated performance	100 xp
Random forest for classification	50 xp
Tune random forest models	100 xp
Random forest performance	100 xp
Build final classification model	0 xp
Measure final model performance	100 xp
Wrap-up	50 xp

HIDE CHAPTER DETAILS ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View

This course is part of these tracks:

Data Scientist with R

Intermediate Tidyverse Toolbox

Dmitriy Gorenshteyn

Lead Data Scientist at Memorial Sloan Kettering Cancer Center

Dmitriy is a Lead Data Scientist in the Strategy & Innovation department at Memorial Sloan Kettering Cancer Center. At MSK he develops predictive models for programs aimed at improving patient care. Prior to this role, Dmitriy completed his Doctorate in Quantitative & Computational Biology at Princeton University. With a passion for teaching and for R, he regularly holds cross-departmental R training sessions within MSK. His core teaching philosophy is centered on building intuition and understanding for the methods and tools available.

See More

COLLABORATOR(S)

Chester Ismay

Sumedh Panchadhar

Eunkyung Park

PREREQUISITES

Modeling with Data in the Tidyverse

DATASETS

Gapminder

Attrition

© 2019 DataCamp Inc.

INTERACTIVE COURSE

Introduction to Machine Learning

Replay Course

4.4 (1,365 Ratings)

6 hours | 15 Videos | 81 Exercises | 79,706 Participants | 6,500 XP

Course Description

This online machine learning course is perfect for those who have a solid basis in R and statistics, but are complete beginners with machine learning. After a broad overview of the discipline's most common techniques and applications, you'll gain more insight into the assessment and training of different machine learning models. The rest of the course is dedicated to a first reconnaissance with three of the most basic machine learning tasks: classification, regression and clustering.

STATEMENT OF ACCOMPLISHMENT

View

Vincent Vankrunkelsven
Data Science Instructor at DataCamp

Vincent has a Master's degree in Artificial Intelligence, and has more than 3 years of experience with machine learning problems of different kinds. He experienced first-hand the difficulties that come with building and assessing machine learning systems. This made him passionate about teaching people how to do machine learning the right way.

See More

Gilles Inghelbrecht
Doctoral Student at Vrije Universiteit Brussel

Even though Gilles has recently graduated with a degree in Fundamental Mathematics, he knows that there's more to be done than mathematics. With a solid knowledge in classical statistics, he now pursues a PhD in parallelizing regression modeling techniques.

See More

COLLABORATOR(S)

Filip Schouwenaars

Sebastian Perez Saaibi

PREREQUISITES

- Introduction to R
- Intermediate R
- Correlation and Regression

DATASETS

- Cars
- Emails
- Titanic
- Air
- Seeds
- Income
- Kangaroos
- World Bank data
- School results
- Olympic run records
- Crime data

1 What is Machine Learning FREE

100%

In this first chapter, you get your first intro to machine learning. After learning the true fundamentals of machine learning, you'll experiment with the techniques that are explained in more detail in future chapters.

Machine Learning: What's the challenge?	50 xp
Acquainting yourself with the data	100 xp
What is, what isn't?	50 xp
What is, what isn't? (2)	50 xp
Basic prediction model	100 xp
Classification, Regression, Clustering	50 xp
Classification, regression or clustering?	50 xp
Classification: Filtering spam	100 xp
Regression: LinkedIn views for the next 3 days	100 xp
Clustering: Separating the iris species	100 xp
Supervised vs. Unsupervised	50 xp
Getting practical with supervised learning	100 xp
How to do unsupervised learning (1)	100 xp
How to do unsupervised learning (2)	100 xp
Tell the difference	50 xp

HIDE CHAPTER DETAILS

Completed

2 Performance measures

100%

You'll learn how to assess the performance of both supervised and unsupervised learning algorithms. Next, you'll learn why and how you should split your data in a training set and a test set. Finally, the concepts of bias and variance are explained.

Measuring model performance or error	50 xp
The Confusion Matrix	100 xp
Deriving ratios from the Confusion Matrix	100 xp
The quality of a regression	100 xp
Adding complexity to increase quality	100 xp
Let's do some clustering!	100 xp
What to do with all these performance measures?	50 xp
Training set and test set	50 xp
Split the sets	100 xp
First you train, then you test	100 xp
Using Cross Validation	100 xp
How many folds?	50 xp
Bias and Variance	50 xp
Overfitting the spam!	100 xp
Increasing the bias	100 xp
Interpretability	50 xp

HIDE CHAPTER DETAILS

Completed

3 Classification

100%

You'll gradually take your first steps to correctly perform classification, one of the most important tasks in machine learning today. By the end of this chapter, you'll be able to learn and build a decision tree and to classify unseen observations with k-Nearest Neighbors.

Decision trees	50 xp
Learn a decision tree	100 xp
Understanding the tree plot	50 xp
Classify with the decision tree	100 xp
Pruning the tree	100 xp
Interpreting the tree	50 xp
Splitting criterion	100 xp
k-Nearest Neighbors	50 xp
Preprocess the data	100 xp
The knn() function	100 xp
K's choice	100 xp
Interpreting a Voronoi diagram	50 xp
The ROC curve	50 xp
Creating the ROC curve (1)	100 xp
Creating the ROC curve (2)	100 xp
The area under the curve	100 xp
Interpreting the curves	50 xp
Comparing the methods	100 xp

HIDE CHAPTER DETAILS

Completed

4 Regression

100%

Although a traditional subject in classical statistics, you can also consider regression from a machine learning point of view. You'll learn more about the predictive capabilities and performance of regression algorithms. At the end of this chapter you'll be acquainted with simple linear regression, multi-linear regression and k-Nearest Neighbors regression.

Regression: simple and linear!	50 xp
Simple linear regression: your first step!	100 xp
Performance measure: RMSE	100 xp
Performance measure: R-squared	100 xp
Another take at regression: be critical	100 xp
Non-linear, but still linear?	100 xp
Interpreting R-squared	50 xp
Multivariable Linear Regression	50 xp
Going all-in with predictors!	100 xp
Are all predictors relevant?	100 xp
Are all predictors relevant? Take 2!	100 xp
Interpreting the residuals and p-values	50 xp
k-Nearest Neighbors and Generalization	50 xp
Does your model generalize?	100 xp
Your own k-NN algorithm!	100 xp
Parametric vs non-parametric!	100 xp

HIDE CHAPTER DETAILS

Completed

5 Clustering

100%

As an unsupervised learning technique, clustering requires a different approach than the ones you have seen in the previous chapters. How can you cluster? When is a clustering any good? All these questions will be answered; you'll also learn about k-means clustering and hierarchical clustering along the way. At the end of this chapter and our machine learning video tutorials, you'll have a basic understanding of all the main principles.

Clustering with k-means	50 xp
k-means: how well did you do earlier?	100 xp
The influence of starting centroids	100 xp
Making a scree plot!	0 xp
What is your optimal k?	50 xp
Performance and scaling issues	50 xp
When to standardize your data?	50 xp
Standardized vs non-standardized clustering (1)	70 xp
Standardized vs non-standardized clustering (2)	100 xp
Hierarchical Clustering	50 xp
Did you get it all?	50 xp
Single Hierarchical Clustering	100 xp
Complete Hierarchical Clustering	100 xp
Hierarchical vs k-means	100 xp
Interpreting Dunn's Index	50 xp
Clustering US states based on criminal activity	100 xp

HIDE CHAPTER DETAILS

Completed

© 2015 DataCamp Inc.

INTERACTIVE COURSE

Joining Data in R with dplyr

Practice Now | Replay Course

This course has been archived. If you are taking this course you will be able to finish it. If you are working on a track containing this course it will remain available until you complete the track.

4 hours | 20 Videos | 84 Exercises | 31,572 Participants | 6,550 XP

Course Description

This course builds on what you learned in *Data Manipulation in R with dplyr* by showing you how to combine data sets with dplyr's two table verbs. In the real world, data comes split across many data sets, but dplyr's core functions are designed to work with single tables of data. In this course, you'll learn the best ways to combine data sets into single tables. You'll learn how to augment columns from one data set with columns from another with mutating joins, how to filter one data set against another with filtering joins, and how to sift through data sets with set operations. Along the way, you'll discover the best practices for building data sets and troubleshooting joins with dplyr. Afterwards, you'll be well on your way to data manipulation mastery!

1 Mutating joins FREE — 100%

Mutating joins add new variables to one dataset from another dataset, matching observations across rows in the process. This chapter will explain the various ways you can join datasets together and what happens when you do.

Welcome to the course!	✓ 50 xp
The advantages of dplyr	✓ 50 xp
Keys	✓ 50 xp
Primary keys	✓ 50 xp
Secondary keys	✓ 50 xp
Multi-variable keys	✓ 50 xp
Joins	✓ 50 xp
A basic join	✓ 100 xp
A second join	✓ 100 xp
A right join	✓ 100 xp
Variations on joins	✓ 50 xp
Inner joins and full joins	✓ 100 xp
Pipes	✓ 100 xp
Practice with pipes and joins	✓ 100 xp
Choose your joins	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

2 Filtering joins and set operations — 100%

Filtering joins and set operations combine information from datasets without adding new variables. Filtering joins filter the observations of one dataset based on whether or not they occur in a second dataset. Set operations use combinations of observations from both datasets to create a new dataset.

Semi-joins	✓ 50 xp
Apply a semi-join	✓ 100 xp
Exploring with semi-joins	✓ 100 xp
A more precise way to filter?	✓ 50 xp
Anti-joins	✓ 50 xp
Apply an anti-join	✓ 100 xp
Apply another anti-join	✓ 100 xp
Which filtering join?	✓ 100 xp
Set operations	✓ 50 xp
How many songs are there?	✓ 100 xp
Greatest hits	✓ 100 xp
Live! Bootleg songs	✓ 100 xp
Multiple operations	✓ 70 xp
Unique values	✓ 50 xp
Comparing datasets	✓ 50 xp
Apply setequal	✓ 100 xp
Apply setequal again	✓ 100 xp
Comparing albums	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

3 Assembling data — 100%

This chapter will show you how to build datasets from basic elements: vectors, lists, and individual datasets that do not require a join. dplyr contains a set of functions for assembling data that work more intuitively than base R's functions. The chapter will also look at when dplyr does and does not use data type coercion.

Binds	✓ 50 xp
Differences between dplyr and base R	✓ 50 xp
Which bind?	✓ 100 xp
Bind rows	✓ 70 xp
Bind columns	✓ 100 xp
Danger	✓ 50 xp
Build a better data frame	✓ 50 xp
data_frame	✓ 50 xp
Make a data frame	✓ 100 xp
Lists of columns	✓ 100 xp
Lists of rows (data frames)	✓ 70 xp
Working with data types	✓ 50 xp
Atomic data types	✓ 50 xp
dplyr's coercion rules	✓ 50 xp
dplyr and coercion	✓ 50 xp
Determining type	✓ 50 xp
Results	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

4 Advanced joining — 100%

Now that you have the basics, let's dive deep into the mechanics of joins. This chapter will show you how to spot common join problems, how to join based on multiple or mismatched keys, how to join multiple tables, and how to recreate dplyr's joins with SQL and base R.

What can go wrong?	✓ 50 xp
Spot the key	✓ 100 xp
Non-unique keys	✓ 50 xp
Two non-unique keys	✓ 50 xp
Missing keys	✓ 70 xp
Defining the keys	✓ 50 xp
Which keys?	✓ 50 xp
A subset of keys	✓ 100 xp
Mis-matched key names	✓ 70 xp
More mismatched names	✓ 70 xp
Joining multiple tables	✓ 50 xp
Multiple joins	✓ 50 xp
Join multiple tables	✓ 100 xp
Filter multiple tables	✓ 70 xp
Other implementations	✓ 50 xp
SQL	✓ 50 xp
Base R	✓ 100 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

5 Case study — 100%

You know the ins and outs of two-table verbs with dplyr, but your knowledge is untried! Let's cement what you've learned with a real world application.

Lahman's Baseball Database	✓ 50 xp
Universal keys?	✓ 70 xp
Common keys	✓ 100 xp
playerID	✓ 100 xp
Salaries	✓ 50 xp
Who are the players?	✓ 100 xp
Missing salaries	✓ 100 xp
Unpaid games?	✓ 100 xp
How many games?	✓ 100 xp
How many at-bats?	✓ 100 xp
Introducing the hall of fame	✓ 50 xp
Hall of fame nominations	✓ 100 xp
Hall of fame inductions	✓ 100 xp
Awards	✓ 100 xp
Salary	✓ 0 xp
Retirement	✓ 100 xp
Congratulations!	✓ 50 xp

HIDE CHAPTER DETAILS ^ — ✓ Completed

STATEMENT OF ACCOMPLISHMENT

View | in

Garrett Grolemund
Data Scientist at RStudio

Garrett is the author of Hands-On Programming with R and R for Data Science from O'Reilly Media. He is a Data Scientist at RStudio and holds a Ph.D. in Statistics, but specializes in teaching. He's taught people how to use R at over 50 government agencies, small businesses, and multi-billion dollar global companies and he's designed RStudio's training materials for R, Shiny, dplyr and more and is a frequent contributor to the RStudio blog. He wrote the popular lubridate package for R...

COLLABORATOR(S)

Nick Carchedi

Tom Jeon

PREREQUISITES

- Introduction to R
- Intermediate R
- Data Manipulation in R with dplyr

DATASETS

- Aerosmith
- The Eagles
- Elvis Presley
- Hank Williams
- Jimi Hendrix
- Julie Andrews
- Michael Jackson
- Frank Sinatra and Bing Crosby
- Musicals
- The Dark Side of the Moon (Pink Floyd)
- Top-selling albums in the US
- The Complete Studio Recordings
- The Song Remains the Same
- The Definitive Collection
- Lahman Names
- Live! Bootleg
- Supergroups

© 2019 DataCamp Inc.

CAREER TRACK

Data Scientist with R

TRACK UPDATE AVAILABLE

A Data Scientist combines statistical and machine learning techniques with R programming to analyze and interpret complex data.

✓ Completed

R Language | 94 hours | 23 Courses | Estimated Salary: $118,000/year

1. Introduction to R — Practice Now
2. Intermediate R — Practice Now
3. Introduction to the Tidyverse — Practice Now
4. Importing Data in R (Part 1) — Practice Now
5. Importing Data in R (Part 2) — Practice Now
6. Cleaning Data in R — Practice Now
7. Importing & Cleaning Data in R: Case Studies — Share | Replay Course
8. Writing Functions in R — Practice Now
9. Data Manipulation in R with dplyr — Practice Now
10. Joining Data in R with dplyr — Practice Now
11. Intro to SQL for Data Science — Practice Now
12. Data Visualization with ggplot2 (Part 1) — Share | Replay Course
13. Data Visualization with ggplot2 (Part 2) — Share | Replay Course
14. Working with Dates and Times in R — Share | Replay Course
15. Introduction to Data — Practice Now
16. Exploratory Data Analysis — Practice Now
17. Exploratory Data Analysis in R: Case Study — Share | Replay Course
18. Correlation and Regression — Practice Now
19. Cluster Analysis in R — Practice Now
20. Supervised Learning in R: Classification — Practice Now
21. Unsupervised Learning in R — Share | Replay Course
22. Joining Data in SQL — Share | Replay Course
23. Reporting with R Markdown — Share | Replay Course

STATEMENT OF ACCOMPLISHMENT

View

TRACK UPDATE AVAILABLE

This is an old version of the track. Click here to access the newest version.

STATEMENT OF ACCOMPLISHMENT

View

INSTRUCTORS

- Jonathan Cornelissen — Co-founder of DataCamp
- Filip Schouwenaars — Data Science Instructor at DataCamp
- David Robinson — Chief Data Scientist, DataCamp
- Nick Carchedi — Product Manager at DataCamp

See all instructors

© 2019 DataCamp Inc.

CAREER TRACK

Data Scientist with R

A Data Scientist combines statistical and machine learning techniques with R programming to analyze and interpret complex data.

✓ Completed

R Language | 94 hours | 22 Courses | Estimated Salary: $118,000/year

- ✓ Introduction to R — Practice Now
- ✓ Intermediate R — Practice Now
- ✓ Introduction to the Tidyverse — Practice Now
- ✓ Importing Data in R (Part 1) — Practice Now

See all 22 courses

STATEMENT OF ACCOMPLISHMENT

View

INSTRUCTORS

Jonathan Cornelissen
Co-founder of DataCamp

Filip Schouwenaars
Data Science Instructor at DataCamp

David Robinson
Chief Data Scientist, DataCamp

Nick Carchedi
Product Manager at DataCamp

See all instructors

© 2019 DataCamp Inc.