






# Teaching data science authentically

using open source educational resources

Tiffany A. Timbers

2021-08-03

# Why create and use open educational resources in general?

- cost savings for learners 
- quick iteration and extension on materials for instructors  + 
- and raising the quality of the resources by facilitating collaboration 
- your teaching materials (and pedagogies) become more discoverable 

# Why create and use open educational resources in data science?

*All the reasons I just stated, and...*

*... it mirrors the practices, tools and workflows used when practising data science.*

**Practice what you preach!**

## Example of OER's we have built using Data Science Open Source tools

1. *Data Science: A First Introduction* - an open textbook aimed at undergraduates students taking their first course in data science
2. Syllabi, lecture notes, labs and lecture videos from courses for a professional Master's in Data Science program
3. *Key Capabilities in Data Science online courses* - Interactive online learning modules aimed at mid-career learners

# Data Science: A First Introduction

Aimed at first year undergraduates, from any discipline.



Tiffany Timbers



Trevor Campbell



Melissa Lee

Data Science: A First Introduction

1 R, Jupyter, and the tidyverse

1.1 Chapter learning objectives

1.2 Jupyter notebooks

1.3 Loading a spreadsheet-like data...

1.4 Assigning value to a data frame

1.5 Creating subsets of data frames ...

1.6 Exploring data with visualizations

2 Reading in data locally and from the ...

2.1 Overview

2.2 Chapter learning objectives

2.3 Absolute and relative file paths

2.4 Reading tabular data from a plai...

2.5 Reading data from an Microsoft ...

2.6 Reading data from a database

2.7 Writing data from R to a .csv file

2.8 Scraping data off the web using R

2.9 Additional resources

3 Cleaning and wrangling data

3.1 Overview

3.2 Chapter learning objectives

3.3 Vectors and Data frames

3.4 Tidy Data

3.5 Combining functions using the p...

3.6 Iterating over data with group\_...

3.7 Using summary statistics

Data Science: A First Introduction

Tiffany-Anne Timbers

Trevor Campbell

Melissa Lee

2021-01-12

Chapter 1 R, Jupyter, and the tidyverse

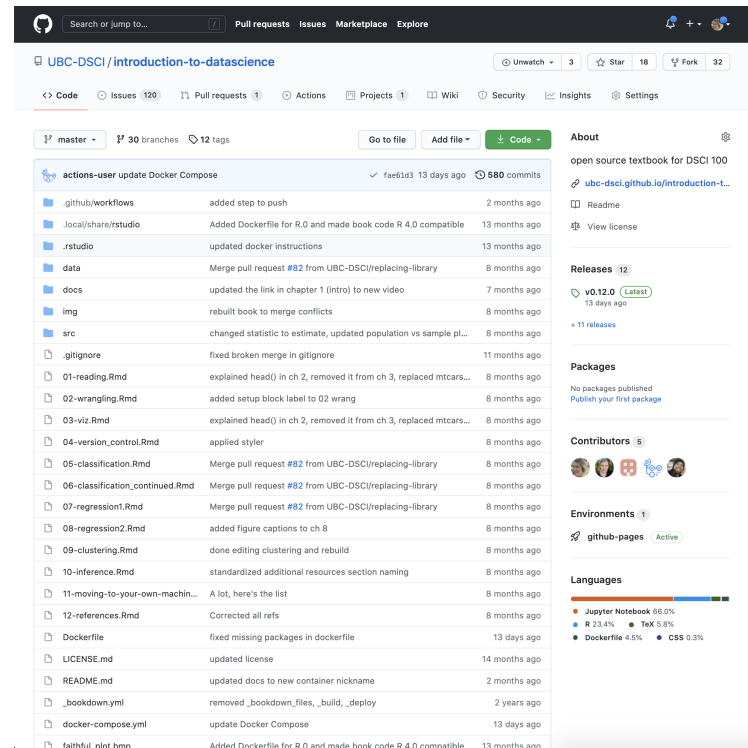
This is an open source textbook aimed at introducing undergraduate students to data science. It was originally written for the University of British Columbia's [DSCI 100 - Introduction to Data Science](#) course. In this book, we define data science as the study and development of reproducible, auditable processes to obtain value (i.e., insight) from data.

The book is structured so that learners spend the first four chapters learning how to use the R programming language and Jupyter notebooks to load, wrangle/clean, and visualize data, while answering descriptive and exploratory data analysis questions. The remaining chapters illustrate how to solve four common problems in data science, which are useful for answering predictive and inferential data analysis questions:

1. Predicting a class/category for a new observation/measurement (e.g., cancerous or benign tumour)
2. Predicting a value for a new observation/measurement (e.g., 10 km race time for 20 year old females with a BMI of 25).
3. Finding previously unknown/unlabelled subgroups in your data (e.g., products commonly bought together on Amazon)
4. Estimating an average or a proportion from a representative sample (group of people or units) and using that estimate to generalize to the broader population (e.g., the proportion of undergraduate students that own an iphone)

# Data Science: A First Introduction

- Created via R + [bookdown](#) R package.
- Source code is available on GitHub: [github.com/UBC-DSCI/introduction-to-datascience](https://github.com/UBC-DSCI/introduction-to-datascience)



# Data Science: A First Introduction

- Book source and rendered HTML version openly licensed under [Creative Commons BY-NC-SA 4.0](#) and will remain that way indefinitely.
- Partnering with CRC Press to publish and sell a print version.



### Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

This is a human-readable summary of (and not a substitute for) the [license](#), [Disclaimer](#).

**You are free to:**

**Share** — copy and redistribute the material in any medium or format

**Adapt** — remix, transform, and build upon the material

The licensor cannot revoke these freedoms as long as you follow the license terms.

---

**Under the following terms:**

 **Attribution** — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

 **NonCommercial** — You may not use the material for [commercial purposes](#).

 **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the [same license](#) as the original.

**No additional restrictions** — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.



# UBC MDS syllabi, lecture notes, labs & lecture videos

Course resources from the UBC Professional Master of Data Science program.

## **Past and present core MDS teaching team members**

- Tomas Beuzen
- Vincenzo Coia
- Giulio Valentino Dalla Riva
- Mike Gelbart
- Varada Kolhatkar
- Rodolfo Lourenzutti
- Firas Moosvi
- Joel Ostblom
- Alexi Rodríguez-Arelis
- Tiffany Timbers



# UBC MDS syllabi, lecture notes, labs & lecture videos

- Course resources from a Master of Data Science program:  
[github.com/UBC-MDS/public](https://github.com/UBC-MDS/public)
- Program uses OER's from others, as well as creates its own
- Syllabi and course notes created via Jupyter, R Markdown & hosted on GitHub
- Videos of some courses hosted on YouTube

## UBC MDS public teaching materials

This repository contains publicly available versions of some of the teaching materials from the [UBC Master of Data Science program](#). See below for the currently available material. The links in the left column lead to course-specific repositories.

Course Number	Course Title	syllabus	lecture notes	labs	lecture videos
DSCI 511	Programming for Data Science	✓	✓		✓
DSCI 521	Computing Platforms for Data Science	✓	✓	✓	
DSCI 551	Descriptive Statistics and Probability for Data Science	✓	✓		
DSCI 542	Communication and Argumentation	✓			
DSCI 523	Programming for Data Manipulation	✓			
DSCI 512	Algorithms and Data Structures	✓	✓	✓	✓
DSCI 531	Data Visualization I	✓	✓	✓	
DSCI 552	Statistical Inference and Computation I	✓			
DSCI 522	Data Science Workflows	✓	✓	✓	
DSCI 513	Databases and Data Retrieval	✓			
DSCI 561	Regression I	✓			
DSCI 571	Supervised Learning I	✓			
DSCI 572	Supervised Learning II	✓	✓	✓	
DSCI 563	Unsupervised Learning	✓			
DSCI 573	Feature and Model Selection	✓			
DSCI 532	Data Visualization II	✓			
DSCI 562	Regression II	✓	✓	✓	
DSCI 524	Collaborative Software Development	✓	✓	✓	
DSCI 553	Statistical Inference and Computation II	✓			
DSCI 574	Spatial and Temporal Models	✓			
DSCI 541	Privacy, Ethics, and Security	✓			
DSCI 525	Web and Cloud Computing	✓			
DSCI 554	Experimentation and Causal Inference	✓			
DSCI 575	Advanced Machine Learning	✓			
DSCI 591	Capstone project	✓	n/a	n/a	

*Course instructors choose to share and distribute their resources using one of the available Creative Commons licenses.*

# Key Capabilities in Data Science online courses

Suite of online courses, and a certificate program, aimed at mid-career learners in British Columbia, Canada, but open to anyone.

UBC certificate program: <https://extendedlearning.ubc.ca/programs/key-capabilities-data-science>

- Course notes, videos and knowledge checks are openly available.
- Paid version of the courses includes a course facilitator, office hours, graded assignments and quizzes.



## Data Visualization

Welcome to Data Visualization! This course is part of the [Key Capabilities for Data Science program](#) and covers topics related to data visualization.

In this course we will learn how to (and how not to) visualize data. We will learn about visualization grammar and how it's implemented in the "Altair" package to create our figures. In addition to common statistical visualizations, we will learn how to tell stories with data, create geographical visualizations, and bring out plots to life by adding interactive elements. Together, these skills will allow us to create effective data visualizations that strengthen our own exploratory data analysis skills and our ability to communicate insights to others.

Course prerequisites: [Programming in Python for Data Science](#)

### Module 0: Exploratory Data Visualization

Course introduction, summary of course learning outcomes and prerequisite validation.

### Module 1: Why Visualize Data?

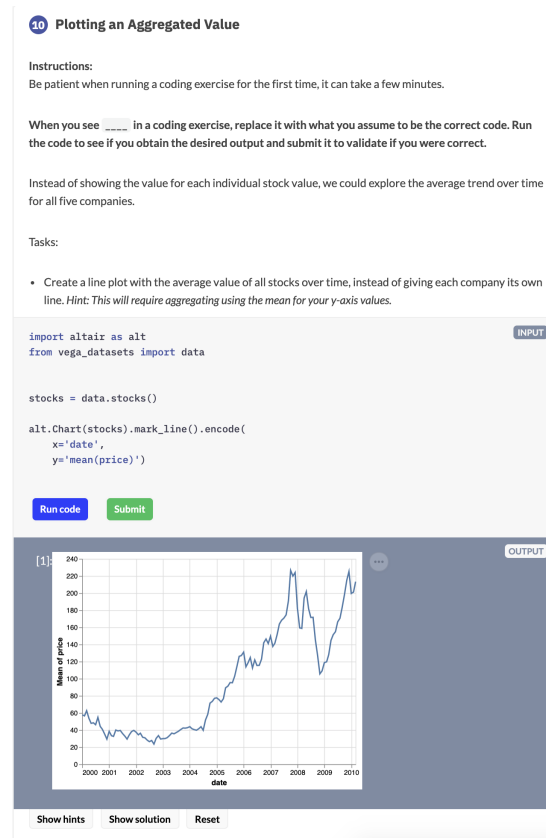
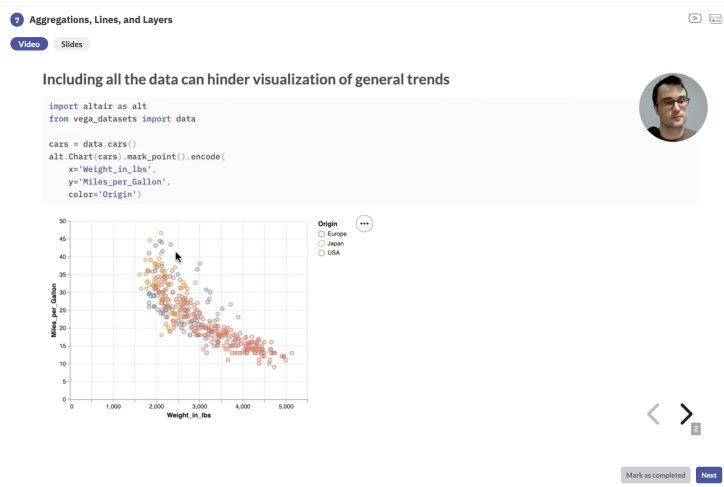
In this module we will be learning about the importance of data visualization and how a grammar of graphics can help us effectively visualize data.

### Module 2: Visual Encoding Channels

In this module we will be learning about the most and least effective ways to encode and present data graphically.

# Key Capabilities in Data Science online courses

Open content created via Ines Montani's open course framework, whose front-end is powered by Gatsby and Reveal.js and back-end code execution is powered by BinderHub.



# Key Capabilities in Data Science online courses

Course open content	Course source
Programming in Python for Data Science	<a href="https://github.com/UBC-MDS/programming-in-python-for-data-science">github.com/UBC-MDS/programming-in-python-for-data-science</a>
Data Visualization	<a href="https://github.com/UBC-MDS/exploratory-data-viz">github.com/UBC-MDS/exploratory-data-viz</a>
Introduction to Machine Learning	<a href="https://github.com/UBC-MDS/introduction-machine-learning">github.com/UBC-MDS/introduction-machine-learning</a>

*All content and source is openly licensed under Creative Commons Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).*

# Lessons learned from building, sharing and maintaining data science OER's

- don't do it alone
- don't be afraid to share
- use a license
- maintenance is needed to keep material current
- OER's which need interactive computation require money or infrastructure to share

# Impact of OER's

# Attribution



# Resources

---

Thanks

