

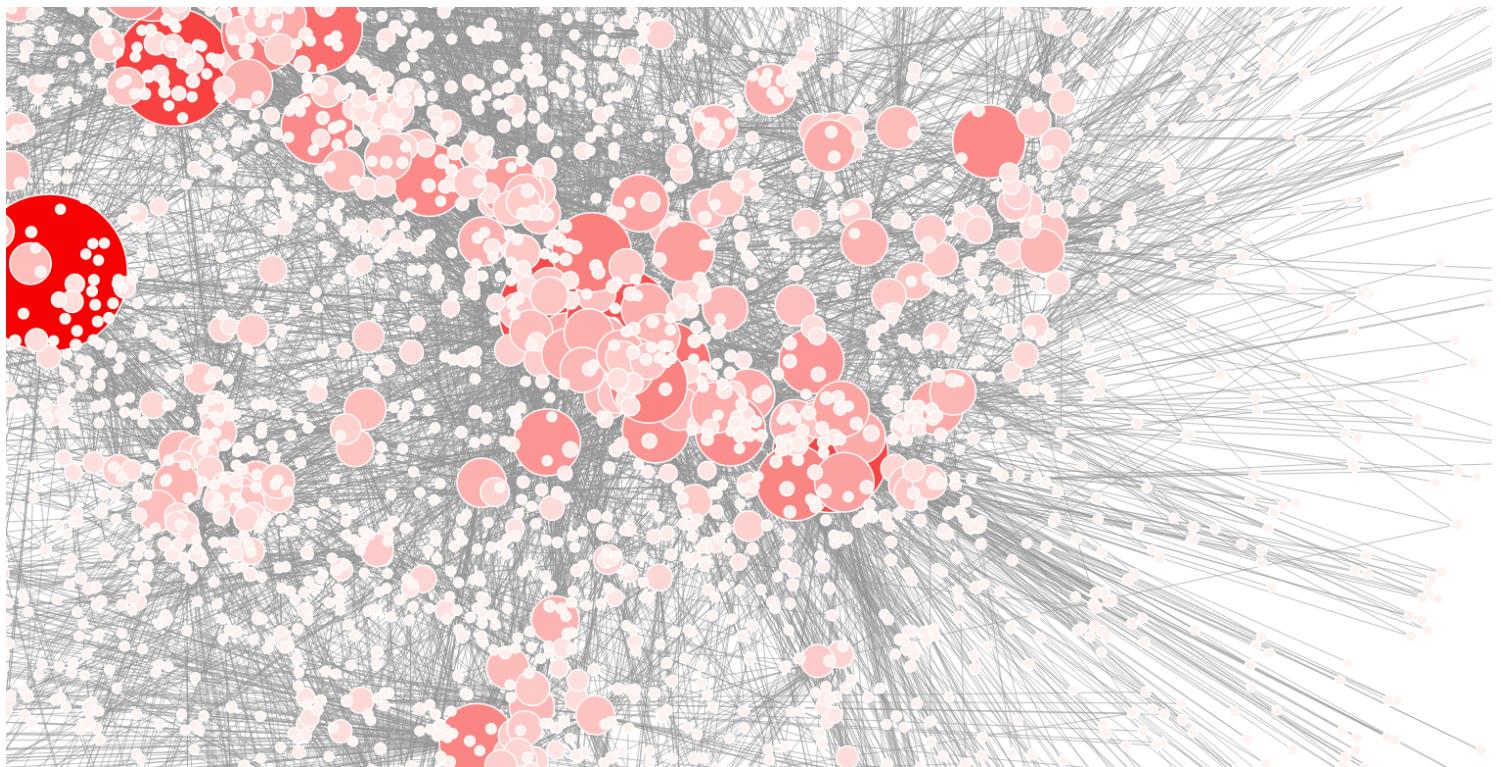
I ranked every Intro to Data Science course on the internet, based on thousands of data points



David Venturi

[Follow](#)

Jan 25, 2017 · 13 min read



Data visualization by Alanah Ryding

A year ago, I dropped out of one of the best computer science programs in Canada. I started creating my own [data science master's program](#) using online resources. I realized that I could learn everything I needed through edX, Coursera, and Udacity instead. And I could learn it faster, more efficiently, and for a fraction of the cost.

I'm almost finished now. I've taken many data science-related courses and audited portions of many more. I know the options out there, and what skills are needed for learners preparing for a data analyst or data scientist role. A few months ago, I started creating a review-driven

guide that recommends the best courses for each subject within data science.

For the first guide in the series, I recommended a few coding classes for the beginner data scientist. Then it was statistics and probability classes.

Now onto introductions to data science.

(Don't worry if you're unsure of what an intro to data science course entails. I'll explain shortly.)

For this guide, I spent 10+ hours trying to identify every online intro to data science course offered as of January 2017, extracting key bits of information from their syllabi and reviews, and compiling their ratings. For this task, I turned to none other than the open source Class Central community and its database of thousands of course ratings and reviews.

The screenshot shows the homepage of Class Central. At the top, there is a large search bar with the placeholder "Find the best online courses from top universities". Below the search bar are two buttons: "MEET YOUR NEXT COURSE" and "BROWSE SUBJECTS". A note below the buttons says "no sign up required". Underneath the search bar, there is a section titled "SPOTLIGHT" featuring five course cards:

- The Nature of Code**: This class focuses on the programming strategies and techniques behind computer simulations of natural systems.
- Modern & Contemporary American Poetry ("ModPo")**: Get a fast-paced introduction to modern and contemporary U.S. poetry, with an emphasis on experimental verse.
- Learning How to Learn: Powerful mental tools to help you master tough subjects**: Learn powerful mental tools that will help you learn more effectively in tough-to-master subjects.
- Medical Neuroscience**: This course will provide you with the foundational knowledge you will need in basic neuroscience and clinical neuroanatomy.
- He Named Me Malala: Girls in STEM**: This course brings Girls in STEM home.

Class Central's homepage.

Since 2011, [Class Central](#) founder [Dhawal Shah](#) has kept a closer eye on online courses than arguably anyone else in the world. Dhawal personally helped me assemble this list of resources.

How we picked courses to consider

Each course must fit three criteria:

1. **It must teach the data science process.** More on that soon.
2. **It must be on-demand or offered every few months.**
3. **It must be an interactive online course, so no books or read-only tutorials.** Though these are viable ways to learn, this guide focuses on courses.

We believe we covered every notable course that fits the above criteria. Since there are seemingly hundreds of courses on [Udemy](#), we chose to

consider the most-reviewed and highest-rated ones only. There's always a chance that we missed something, though. So please let us know in the comments section if we left a good course out.

How we evaluated courses

We compiled average rating and number of reviews from Class Central and other review sites to calculate a weighted average rating for each course. We read text reviews and used this feedback to supplement the numerical ratings.

We made subjective syllabus judgment calls based on two factors:

1. **Coverage of the data science process.** Does the course brush over or skip certain subjects? Does it cover certain subjects in too much detail? See the next section for what this process entails.
2. **Usage of common data science tools.** Is the course taught using popular programming languages like Python and/or R? These aren't necessary, but helpful in most cases so slight preference is given to these courses.

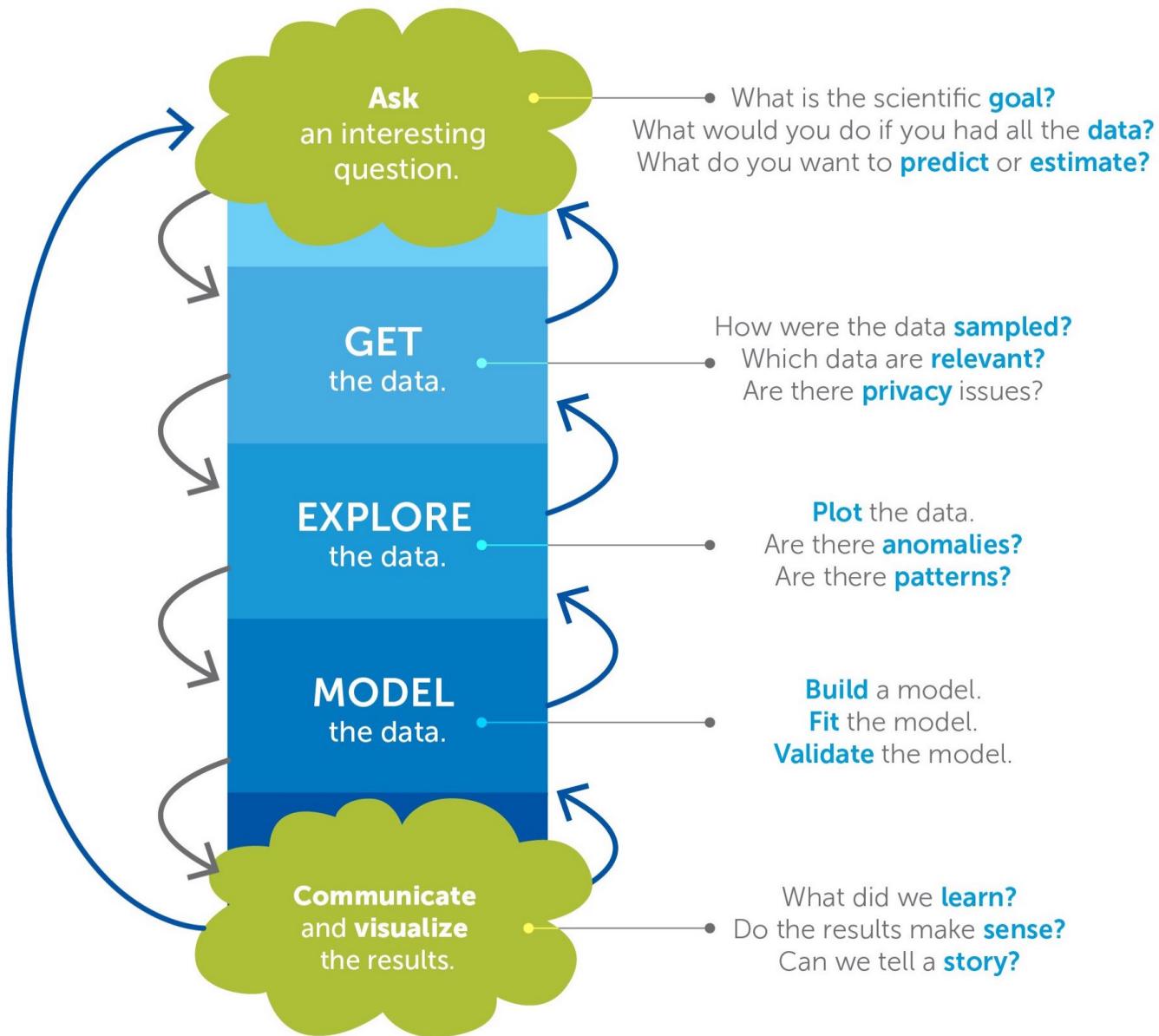


Python and R are the two most popular programming languages used in data science.

What is the data science process?

What is data science? What does a data scientist do? These are the types of fundamental questions that an intro to data science course should answer. The following infographic from Harvard professors Joe Blitzstein and Hanspeter Pfister outlines a typical **data science process**, which will help us answer these questions.

The Data Science Process



Derived from the work of Joe Blitzstein and Hanspeter Pfister,
originally created for the Harvard data science course <http://cs109.org/>.

Visualization from Opera Solutions.

Our goal with this introduction to data science course is to become familiar with the data science process. We don't want too in-depth

coverage of specific aspects of the process, hence the “intro to” portion of the title.

For each aspect, the ideal course explains key concepts within the framework of the process, introduces common tools, and provides a few examples (preferably hands-on).

We’re only looking for an introduction. This guide therefore won’t include full specializations or programs like Johns Hopkins University’s [Data Science Specialization](#) on Coursera or Udacity’s [Data Analyst Nanodegree](#). These compilations of courses elude the purpose of this series: to find the best **individual** courses for each subject to comprise a data science education. The final three guides in this series of articles will cover each aspect of the data science process in detail.

Basic coding, stats, and probability experience required

Several courses listed below require basic programming, statistics, and probability experience. This requirement is understandable given that the new content is reasonably advanced, and that these subjects often have several courses dedicated to them.

This experience can be acquired through our recommendations in the first two articles ([programming](#), [statistics](#)) in this Data Science Career Guide.

Our pick for the best intro to data science course is...

- [Data Science A-Z™: Real-Life Data Science Exercises Included](#)
(Kirill Eremenko/Udemy)

Kirill Eremenko’s [Data Science A-Z™](#) on Udemy is the clear winner in terms of breadth and depth of coverage of the data science process of the 20+ courses that qualified. It has a 4.5-star weighted average rating over 3,071 reviews, which places it among the highest rated and most reviewed courses of the ones considered.

It outlines the full process and provides real-life examples. At 21 hours of content, it is a good length. Reviewers love the instructor’s delivery

and the organization of the content. The price varies depending on Udemy discounts, which are frequent, so you may be able to purchase access for as little as \$10.

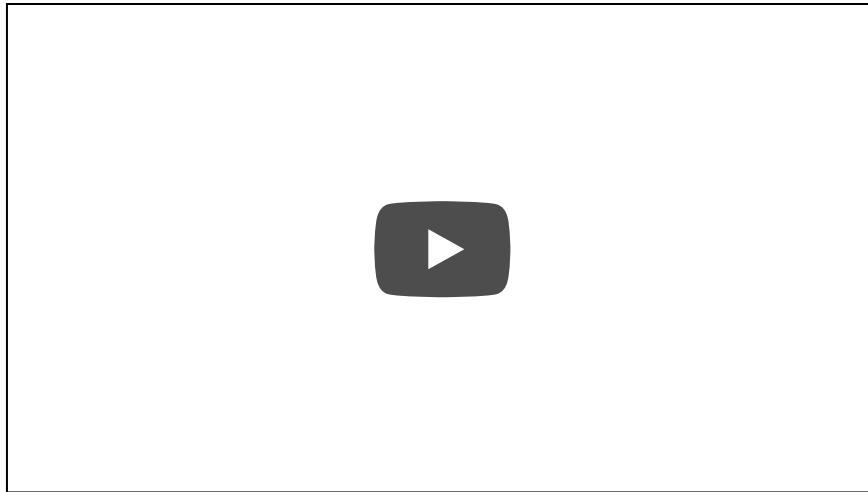
Though it doesn't check our "usage of common data science tools" box, the non-Python/R tool choices (gretl, Tableau, Excel) are used effectively in context. Eremenko mentions the following when explaining the gretl choice (gretl is a statistical software package), though it applies to all of the tools he uses (emphasis mine):

In gretl, we will be able to do the same modeling just like in R and Python but we won't have to code. That's the big deal here. Some of you may already know R very well, but some may not know it at all. My goal is to show you how to build a robust model and give you a framework that you can apply in any tool you choose. gretl will help us avoid getting bogged down in our coding.

One prominent reviewer noted the following:

Kirill is the best teacher I've found online. He uses real life examples and explains common problems so that you get a deeper understanding of the coursework. He also provides a lot of insight as to what it means to be a data scientist from working with insufficient data all the way to presenting your work to C-class management. I highly recommend this course for beginner students to intermediate data analysts!





The preview video for Data Science A-Z™.

A great Python-focused introduction

- [Intro to Data Analysis](#) (Udacity)

Udacity's [Intro to Data Analysis](#) is a relatively new offering that is part of Udacity's popular [Data Analyst Nanodegree](#). It covers the data science process clearly and cohesively using Python, though it lacks a bit in the modeling aspect. The estimated timeline is 36 hours (six hours per week over six weeks), though it is shorter in my experience. It has a 5-star weighted average rating over two reviews. It is free.

The videos are well-produced and the instructor (Caroline Buckey) is clear and personable. Lots of programming quizzes enforce the concepts learned in the videos. Students will leave the course confident in their new and/or improved NumPy and Pandas skills (these are popular Python libraries). The final project—which is graded and reviewed in the Nanodegree but not in the free individual course—can be a nice add to a portfolio.





Udacity instructor Caroline Buckey outlining the data analysis process (also known as the data science process).

An impressive offering with no review data

- [Data Science Fundamentals](#) (Big Data University)

Data Science Fundamentals is a four-course series provided by IBM's Big Data University. It includes courses titled [Data Science 101](#), [Data Science Methodology](#), [Data Science Hands-on with Open Source Tools](#), and [R 101](#).

It covers the full data science process and introduces Python, R, and several other open-source tools. The courses have tremendous production value. 13–18 hours of effort is estimated, depending on if you take the “R 101” course at the end, which isn’t necessary for the purpose of this guide. Unfortunately, it has no review data on the major review sites that we used for this analysis, so we can’t recommend it over the above two options yet. It is free.



What is Data Science? [Data Science 101]

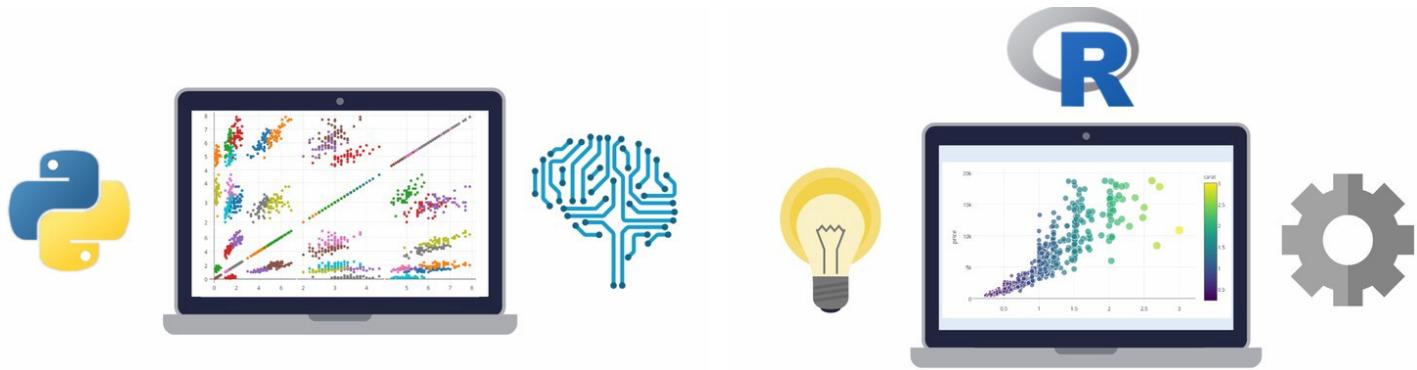


A video from the first module of the Big Data University's Data Science 101 (which is the first course in the Data Science Fundamentals series).

The competition

Our #1 pick had a weighted average rating of 4.5 out of 5 stars over 3,068 reviews. Let's look at the other alternatives, sorted by descending rating. Below you'll find several R-focused courses, if you are set on an introduction in that language.

- [Python for Data Science and Machine Learning Bootcamp](#) (Jose Portilla/Udemy): Full process coverage with a tool-heavy focus (Python). Less process-driven and more of a very detailed intro to Python. Amazing course, though not ideal for the scope of this guide. It, like Jose's R course below, can double as both intros to Python/R and intros to data science. 21.5 hours of content. It has a **4.7-star** weighted average rating over 1,644 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Data Science and Machine Learning Bootcamp with R](#) (Jose Portilla/Udemy): Full process coverage with a tool-heavy focus (R). Less process-driven and more of a very detailed intro to R. Amazing course, though not ideal for the scope of this guide. It, like Jose's Python course above, can double as both intros to Python/R and intros to data science. 18 hours of content. It has a **4.6-star** weighted average rating over 847 reviews. Cost varies depending on Udemy discounts, which are frequent.



Jose Portilla has two Data Science and Machine Learning Bootcamps on Udemy: one for Python and one for R.

- [Data Science and Machine Learning with Python—Hands On!](#)
 (Frank Kane/Udemy): Partial process coverage. Focuses on statistics and machine learning. Decent length (nine hours of content). Uses Python. It has a **4.5**-star weighted average rating over 3,104 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Introduction to Data Science](#) (Data Hawk Tech/Udemy): Full process coverage, though limited depth of coverage. Quite short (three hours of content). Briefly covers both R and Python. It has a **4.4**-star weighted average rating over 62 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Applied Data Science: An Introduction](#) (Syracuse University/Open Education by Blackboard): Full process coverage, though not evenly spread. Heavily focuses on basic statistics and R. Too applied and not enough process focus for the purpose of this guide. Online course experience feels disjointed. It has a **4.33**-star weighted average rating over 6 reviews. Free.
- [Introduction To Data Science](#) (Nina Zumel & John Mount/Udemy): Partial process coverage only, though good depth in the data preparation and modeling aspects. Okay length (six hours of content). Uses R. It has a **4.3**-star weighted average rating over 101 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Applied Data Science with Python](#) (V2 Maestros/Udemy): Full process coverage with good depth of coverage for each aspect of

the process. Decent length (8.5 hours of content). Uses Python. It has a **4.3**-star weighted average rating over 92 reviews. Cost varies depending on Udemy discounts, which are frequent.

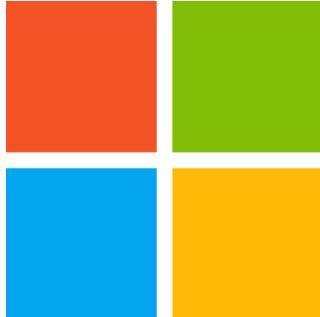


The Big Data Science Experts

V2 Maestros has two versions of their "Applied Data Science" course: one for Python and one for R.

- [Want to be a Data Scientist?](#) (V2 Maestros/Udemy): Full process coverage, though limited depth of coverage. Quite short (3 hours of content). Limited tool coverage. It has a **4.3**-star weighted average rating over 790 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Data to Insight: an Introduction to Data Analysis](#) (University of Auckland/FutureLearn): Breadth of coverage unclear. Claims to focus on data exploration, discovery, and visualization. Not offered on demand. 24 hours of content (three hours per week over eight weeks). It has a **4**-star weighted average rating over 2 reviews. Free with paid certificate available.
- [Data Science Orientation](#) (Microsoft/edX): Partial process coverage (lacks modeling aspect). Uses Excel, which makes sense given it is a Microsoft-branded course. 12–24 hours of content (two-four hours per week over six weeks). It has a **3.95**-star weighted average rating over 40 reviews. Free with Verified Certificate available for \$25.
- [Data Science Essentials](#) (Microsoft/edX): Full process coverage with good depth of coverage for each aspect. Covers R, Python, and Azure ML (a Microsoft machine learning platform). Several 1-star reviews citing tool choice (Azure ML) and the instructor's

poor delivery. 18–24 hours of content (three-four hours per week over six weeks). It has a **3.81**-star weighted average rating over 67 reviews. Free with Verified Certificate available for \$49.



Microsoft

The above two courses are from Microsoft's Professional Program Certificate in Data Science on edX.

- [Applied Data Science with R](#) (V2 Maestros/Udemy): The R companion to V2 Maestros' Python course above. Full process coverage with good depth of coverage for each aspect of the process. Decent length (11 hours of content). Uses R. It has a **3.8**-star weighted average rating over 212 reviews. Cost varies depending on Udemy discounts, which are frequent.
- [Intro to Data Science](#) (Udacity): Partial process coverage, though good depth for the topics covered. Lacks the exploration aspect, though Udacity has a great, [full course](#) on exploratory data analysis (EDA). Claims to be 48 hours in length (six hours per week over eight weeks), but is shorter in my experience. Some reviews think the set-up to the advanced content is lacking. Feels disorganized. Uses Python. It has a **3.61**-star weighted average rating over 18 reviews. Free.
- [Introduction to Data Science in Python](#) (University of Michigan/Coursera): Partial process coverage. No modeling and visualization, though courses #2 and #3 in the [Applied Data Science with Python Specialization](#) cover these aspects. Taking all three courses would be too in depth for the purpose of this guides. Uses Python. Four weeks in length. It has a **3.6**-star weighted average rating over 15 reviews. Free and paid options available.

UNIVERSITY OF MICHIGAN

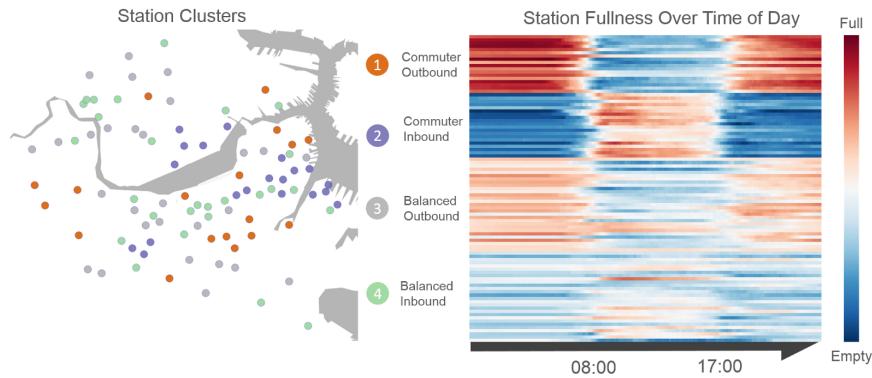
The University of Michigan teaches the Applied Data Science with Python Specialization on Coursera.

- Data-driven Decision Making (PwC/Coursera): Partial coverage (lacks modeling) with a business focus. Introduces many tools, including R, Python, Excel, SAS, and Tableau. Four weeks in length. It has a **3.5**-star weighted average rating over 2 reviews. Free and paid options available.
- A Crash Course in Data Science (Johns Hopkins University/Coursera): An extremely brief overview of the full process. Too brief for the purpose of this series. Two hours in length. It has a **3.4**-star weighted average rating over 19 reviews. Free and paid options available.
- The Data Scientist's Toolbox (Johns Hopkins University/Coursera): An extremely brief overview of the full process. More of a set-up course for Johns Hopkins University's Data Science Specialization. Claims to have 4–16 hours of content (one-four hours per week over four weeks), though one reviewer noted it could be completed in two hours. It has a **3.22**-star weighted average rating over 182 reviews. Free and paid options available.
- Data Management and Visualization (Wesleyan University/Coursera): Partial process coverage (lacks modeling). Four weeks in length. Good production value. Uses Python and SAS. It has a **2.67**-star weighted average rating over 6 reviews. Free and paid options available.

The following courses had no reviews as of January 2017.

- CS109 Data Science (Harvard University): Full process coverage in great depth (probably too in depth for the purpose of this series). A full 12-week undergraduate course. Course navigation is difficult since the course is not designed for online consumption.

Actual Harvard lectures are filmed. The above data science process infographic originates from this course. Uses Python. No review data. Free.



The featured viz on Harvard CS109's homepage.

- [Introduction to Data Analytics for Business](#) (University of Colorado Boulder/Coursera): Partial process coverage (lacks modeling and visualization aspects) with a focus on business. The data science process is disguised as the “Information-Action Value chain” in their lectures. Four weeks in length. Describes several tools, though only covers SQL in any depth. No review data. Free and paid options available.
- [Introduction to Data Science](#) (Lynda): Full process coverage, though limited depth of coverage. Quite short (three hours of content). Introduces both R and Python. No review data. Cost depends on Lynda subscription.

Wrapping it Up

This is the third of a six-piece series that covers the best online courses for launching yourself into the data science field. We covered programming in the [first article](#) and statistics and probability in the [second article](#). The remainder of the series will cover other data science core competencies: data visualization and machine learning.

If you want to learn Data Science, start with one of these programming classes

medium.freecodecamp.com





If you want to learn Data Science, take a few of these statistics classes

medium.freecodecamp.com



The final piece will be a summary of those articles, plus the best online courses for other key topics such as data wrangling, databases, and even software engineering.

If you're looking for a complete list of Data Science online courses, you can find them on Class Central's [Data Science and Big Data](#) subject page.

If you enjoyed reading this, check out some of [Class Central](#)'s other pieces:

Here are 250 Ivy League courses you can take online right now for free

250 MOOCs from Brown, Columbia, Cornell, Dartmouth, Harvard, Penn, Princeton, and Yale

medium.freecodecamp.com



The 50 best free online university courses according to data

When I launched Class Central back in November 2011, there were around 18 or so free online...

medium.freecodecamp.com



If you have suggestions for courses I missed, let me know in the responses!

If you found this helpful, click the  so more people will see it here on Medium.

This is a condensed version of my [original article published on Class Central](#), where I've included further course descriptions, syllabi, and multiple reviews.

