*Welcome to Data Scientist Tool Box*

Welcome to Week 1 of the Data Scientist's Toolbox! This course is an introduction to the tools and ideas that you will see throughout the rest of the Data Science Specialization.

We believe that the key word in Data Science is "science". Our course track is focused on providing you with three things: (1) an introduction to the key ideas behind working with data in a scientific way that will produce new and reproducible insight, (2) an introduction to the tools that will allow you to execute on a data analytic strategy, from raw data in a database to a completed report with interactive graphics, and (3) on giving you plenty of hands on practice so you can learn the techniques for yourself.

This course focuses primarily on getting you set up with the appropriate tools and accounts you will need for the rest of the specialization and on giving you a solid grounding in the key conceptual ideas. If you feel like the material is basic, that is ok, you will see much more in depth treatment of each topic in the subsequent courses in the track.

We are excited about the opportunity to attempt to scale Data Science education. We intend for the courses to be self contained, fast paced, and interactive.

One important note is that as part of this class you will be required to set up a Github account. Github is a tool for collaborative code sharing and editing. During this course and other courses in the track you will be submitting links to files you publicly place in your Github account as part of your Course Projects. If you are concerned about preserving your anonymity you should set up an anonymous Github account and be careful not to include any information you do not want made available to peer evaluators.

Please see the course syllabus for information about the quizzes, the Course Project, and grading. Don't forget to say hi on the forums. The community developed around these courses is one of the best places to learn and the best things about taking a MOOC!

*Syllabus*

**Course Title:** Data Scientist's Toolbox

**Course Instructor:**Jeff Leek

Course Description:

In this course you will get an introduction to the main tools and ideas in the data scientists toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, Github, R, and Rstudio.

This course focuses primarily on getting you set up with the appropriate tools and accounts you will need for the rest of the specialization and on giving you a solid grounding in the key conceptual ideas. If you feel like the material is basic, that is ok, you will see much more in depth treatment of each topic in the subsequent courses in the track.

Course Content:

* Track motivation
* Getting help
* Introduction to basic tools
* R
* Rstudio
* Git
* Github
* Types of data questions
* Steps in a data analysis
* Putting the science in data science

Weekly quizzes

* There are three weekly quizzes.
* You must earn a grade of at least 80% to pass a quiz
* You may attempt each quiz up to 3 times in 8 hours.
* The score from your most successful attempt will count toward your final grade.

*Welcome to the Data Scientist’s Toolbox*

The Course Project

In the Course Project, you will demonstrate that you've set up all of the necessary accounts for the tools we'll be using.

You are required to evaluate and grade at least four of your classmates' projects. For this course, the project can be evaluated with a series of yes/no answers to determine whether people completed the required installations.

Part of the course project includes submitting a screenshot to demonstrate you have installed the relevant software. Be sure not to take a screenshot with other applications open that may reveal personal information or anything else you don't want others to see.

Grading policy

You must score at least 80% on all assignments (Quizzes & Project) to pass the course.

Your final grade will be calculated as follows:

* Quiz 1 = 20%
* Quiz 2 = 20%
* Quiz 3 = 20%
* Course project = 40%

Differences of opinion

Keep in mind that currently data analysis is as much art as it is science - so we may have a difference of opinion - and that is ok! Please refrain from angry, sarcastic, or abusive comments on the message boards. Our goal is to create a supportive community that helps the learning of all students, from the most advanced to those who are just seeing this material for the first time.

Plagiarism

Johns Hopkins University defines plagiarism as "...taking for one's own use the words, ideas, concepts or data of another without proper attribution. Plagiarism includes both direct use or paraphrasing of the words, thoughts, or concepts of another without proper attribution." We take plagiarism very seriously, as does Johns Hopkins University.

We recognize that many students may not have a clear understanding of what plagiarism is or why it is wrong. Please see the JHU referencing guide for more information on plagiarism.

It is critically important that you give people/sources credit when you use their words or ideas. If you do not give proper credit -- particularly when quoting directly from a source -- you violate the trust of your fellow students.

The Coursera Honor code includes an explicit statement about plagiarism:

I will register for only one account. My answers to homework, quizzes and exams will be my own work (except for assignments that explicitly permit collaboration). I will not make solutions to homework, quizzes or exams available to anyone else. This includes both solutions written by me, as well as any official solutions provided by the course staff. I will not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.

Reporting plagiarism on course projects

One of the criteria in the project rubric focuses on plagiarism. Keep in mind that some components of the projects will be very similar across terms and so answers that appear similar may be honest coincidences. However, we would appreciate if you do a basic check for obvious plagiarism and report it during your peer assessment phase.

It is currently very difficult to prove or disprove a charge of plagiarism in the MOOC peer assessment setting. We are not in a position to evaluate whether or not a submission actually constitutes plagiarism, and we will not be able to entertain appeals or to alter any grades that have been assigned through the peer evaluation system.

But if you take the time to report suspected plagiarism, this will help us to understand the extent of the problem and work with Coursera to address critical issues with the current system.

*Specialization Motivation*

[Welcome to the Johns Hopkins Data Science Track.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [I'm incredibly excited to tell you a little bit about the track](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [and about where you're going to be going over the next nine months.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [My name is Jeff Leek, and I'm a professor](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [in the Johns Hopkins Bloomberg School of Public Health.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [I thought I'd lead off this introductory video with a](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [quote by one of my favorite US Presidents, Teddy Roosevelt.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [He said it's not the critic who counts.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It's not the person who points out how the person](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [who's actually doing things is doing them wrong or messing up.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It's the person who's actually trying to get things](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [done, even when there are obstacles in the way.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And a lot of data science right now is being able to push through a](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [lot of the difficulties that you have when](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [you're dealing with either large or messy data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It includes collecting the data clean them up and then](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [building new announced techniques that exploring new information about that data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so, all of those steps are a little bit complicated and sometimes](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [it opens you to criticism when you're](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [trying to do something new and interesting.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so I wanted to lead with a quote that said it's important to strive](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [the valiantly do these sorts of things,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [even if you're going to take some criticism.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So the key challenge in data science is actually](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [really nicely summed up in this quote by Dan Myer.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [He says, ask yourselves, what problem you, have you ever solved, ever, that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [was worth solving, where you knew all of the given information in advance?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Where you didn't have a surplus of information and have to filter some of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [it out, or you didn't have insufficient information and have to go find some?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[1:24](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[And so, I think that this is a kind of a critical](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [quote because, in data science, this is usually what is going on.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [You're either in a situation where you really don't have](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [enough data to answer the question that you're interested in, and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [you have to go out and try to search for it,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [find it on the web, or find it in other places.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Or you're in a situation where you are overwhelmed with a surplus of data and you](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [have to filter out all of the irrelevant](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [information to try to narrow in on your question.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And you'll notice that I said question in both of those cases.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And I think this goes to the heart of our philosophy about data science.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [We're interested in answering questions with data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [We think the question should come first and then the data should follow after.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And that actually makes it more challenging,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [because sometimes, you can answer a question with](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [some data but you might not be able to answer your question with some data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So this track is about refocusing on answering the question](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [that you're interested in solving with the data that you have.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So I thought I'd tell you a little bit about the instructors](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [that you'll be hearing from throughout the course of this, course track.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So we are all faculty in the Johns Hopkins](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Blumberg School of Public Health in the Biostatistics Department.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And you could say that we all](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [do data intensive statistics in biology and medicine.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[2:33](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[Brian Caffo works on the statistics of brain, analyzing brain imaging data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And I work on the statistics of analyzing genomics data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And Roger Peng works on the statistics of analyzing fine particulate matter.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[2:47](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[All of us work on problems where the data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [aren't always clean and nice and easy to handle.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [All of us work on problems where the questions that we want](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [to answer are complicated and you have to break them down into parts.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And all of us, sort of, work on questions where we're very passionate about](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [trying to get the right answer so that we can help people in human health.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But the techniques that you're going to be](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [learning about are not exclusive to biology and medicine.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [That's just one area where there's been a](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [recent upsurge in the amount of data that's available.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[3:19](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So why data science?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Why should you take this program?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [This is a cover of The Economist now.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It's a little bit old I guess ancient history from a couple of years ago.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But it talks about the data deluge and it's really true.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Over the last several years data has become much, much cheaper to collect.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It's much easier to store.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And there's so many free computing tools out](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [there right now, that you can actually do](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [something with this entire data deluge that's sort](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [of assaulting all different areas of science and business.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[3:48](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So the other thing is that you've probably heard the term big data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so we'll hear a little bit more about what we think about](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [big data throughout the course of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [this particular course, the Data Scientist's Toolbox.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But big data is, sort of a new frontier in the sense that,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [we have data in areas that we didn't used to have that data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [We didn't have access to information about GPS](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [coordinates from cars from everybody in the entire world.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It wasn't possible to sequence everybody's genome.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And now that's all possible.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So we have access to this data and it](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [allows us to answer questions we never could before.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, it's an incredibly exciting time, and you're somebody who can](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [get in there and use that data to answer those questions.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[4:25](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So why statistical data science?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [You'll notice that we're, all of your](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [instructors are biostatistics professors and so this](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [will, this data science track will obviously](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [have a little bit of a statistical bend.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [I think that that's appropriate given that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [statistics is the science of learning from data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, data is very, very, it's very rare that you'll get a data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [set where all of the answers are really clear, and there's no uncertainty.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [In any case where there is uncertainty,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [that's where statistics comes and plays a role.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, this is a again, a little bit older New](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [York Times article now, but it talks about how the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [key word for a lot of graduates to open the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [door for a lot of jobs, is to learn about statistics.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[5:05](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So why are you lucky?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [You're lucky because this moment, right now, in time is](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [sort of like the moment that Jeff Bezos discovered the internet.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [He got into building a internet company at the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [time when there was this explosive growth in internet](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [usage and it just opened the door for the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [opportunity to build something amazing and huge and wonderful.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And sort of, that's the right, that's what the time is right now for data.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It sort of there's an explosive growth of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [data in every possible area you can imagine.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so it's the opportunity right now to sort of jump on a rocket and, and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [find out something interesting, and, and sort of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [carry it off into a, a really major endeavor.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[5:44](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[You're also lucky because tools and competitions and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [websites have all been developed around the idea of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [helping to learn about data, but also getting](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [involved in projects that have super high profile results.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, one example is the Heritage Health Prize,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [which I'm showing you a picture of here.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [The Heritage Health Prize was a $3 million](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [contest for people who could analyze data and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [come up with a better predictor of who](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [would be admitted to a hospital in another year.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So you can see that's a huge amount of money that's being](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [invested in these ideas of algorithm](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [development and data science of prediction.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So it gives you an exciting opportunity to get involved in](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [projects that, sort of, weren't happening five or ten years ago.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[6:26](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[This course track will focus almost exclusively](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [on the use of the R programming language.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so I thought it was appropriate to talk](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [a little bit about why we like R so much.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So we like R obviously, because we all use it.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But it's also sort of increasingly the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [most commonly used language for data science.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [There are other languages that are also very useful.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And we won't be talking about them a lot in this](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [course but they're obviously good complements to the R programming language.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Like, Python, in this class we'll be focusing on R because](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [it has a broad range of packages that allow you to go](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [from the rawest of raw files, all the way to interactive reports and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [documents and web apps that you can share with your collaborators.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[7:12](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So, some more reasons why we might use r is because it's free, it has a](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [comprehensive set of packages, like I mentioned, for](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [all the processes that are involved in data science.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It has one of the best development](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [environments of any programming language, in our studio.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [It also has an amazing ecosystem of developers.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And what I mean by that is there are](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [a lot of people that are developing our packages.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And they're also available to get in touch with](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [on mailing lists or by email or on stack overflow.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so it's really possible to learn about](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [the cutting edge of packages that are being developed.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [There also very easy to install and play nicely together, which is a, a feature](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [that doesn't always happen in a lot of the languages that are used for data science.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So the next thing I thought I would mention is, who is a data scientist?](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, we're going to be talking about data science a lot.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And I thought I'd mention that some people that I think are data scientists, that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [might not, either label themselves that way](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [or have other people label them that way.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So the first is Daryl Morey, who'd the general](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [manager of the Houston Rockets basketball team in the US.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So he uses data to analyze](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [basketball players and transactions and making trades.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so I would consider him to be a data scientist](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Because he's a person who uses data to answer questions about basketball.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[8:23](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[Another data scientist that you may, or may not, have heard of, is Hilary Mason.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, she used to be the Chief, Data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Scientist at Bentley, and now she's at Accel Partners.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so, she uses data to answer all sorts of questions about mining the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [web, and understanding that way that humans](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [interact with each other through social media.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, again she might not label herself a data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [scientist, but I think the way that she uses data,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [is a evocative of the sort of ideas, that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [we would like to convey in this data sciences track.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [If you're taking this course, you probably know who Daphne Koller is.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [She's the CEO of Coursera.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But she's also another person who's using all the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [data they're collecting through Coursera to better, to improve](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [the way that we do educational delivery and educational](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [assessment at this huge scale that Coursera is providing.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[9:08](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[And finally, Nate Silver is one of the most](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [famous data scientists, or statisticians in the world today.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So he used a large amount of totally free public data to make](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [predictions about who would win elections in](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [the United States, and was remarkably accurate.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [I'm going to finish with him, as the da-, as the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [last data scientist I'm going to illustrate because it's so amazing that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [he could use public free data and create such an amazing](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [product that so many people read about, and are excited about.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[9:36](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So our goal is to teach you about a bunch different](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [skills that will be useful for you as a data scientist.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, this is a Venn Diagram and some statisticians and data scientists](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [don't like Venn Diagrams but I'm going to get, show you one anyway.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so, this Venn Diagram has Data Science at the, sort](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [of, the center of this Venn diagram that intersects several different skills.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So, if you look right here there's](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [data science and it involves three different components.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [There's hacking skills, there's math and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [statistics knowledge, and there's substantive expertise.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[10:13](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[And so our data science track will focus a little bit on each of](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [these, but it will primarily focus on](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [math and statistics knowledge and hacking skills.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[10:21](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[And so math and statistics knowledge sort of speaks for itself.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [We're going to teach you a little bit](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [about math and a little bit about statistics.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Hacking skills involves two different components.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [One thing is we're going to teach you a little](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [bit about computer programming or at least computer programming](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [with R, which will allow you to access data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [and play around with it and analyze it, plot it.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [But hacking skills also has another component to it which](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [is the ability to go out and answer questions for yourself.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [One key component of a data scientist job right now is](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [that most of the answers aren't already outlined in the textbook.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [This is all new stuff that's happening.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [So what are the major skills of being a](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [data scientist is being to go to Google, and](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [go to Stack Overflow, and go to one of the other sites and look up what you need](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [to learn and figure out what answers you know](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [and what answers you don't know, and then figuring](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [out how you can use the information you have](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [to answer the question that you'd like to answer.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[11:13](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So another reason obviously is jobs.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [That might be the reason you're taking this course track.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And so you can see this is a plot of listings](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [are for data science jobs over time and of course it's exploding.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And we'll talk a little bit about why you shouldn't extrapolate,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [necessarily, from your data forever, but it does suggest that data](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [science is a hot area that's growing and I think obviously](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [we're very excited about it and hope you're excited about it too.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[11:37](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)

[So this course, Data Scientist's Toolbox, will](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [continue with lectures on the following three things.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [First, we're going to introduce you to the course track.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Then we're going to tell you a little bit about getting the tools that](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [you need to get set up and get installed, hopefully get you over that hump.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [And then we're going to give you the](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [basic background on data science sort of writ large,](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [so that you'll be ready to jump into](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [any of the individual classes and really take off.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation) [Looking forward to seeing you in the rest of the class.](https://www.coursera.org/learn/data-scientists-tools/lecture/VwuDp/specialization-motivation)