

[Outline](#) > 1. The Big Picture > Lecture: Data Science and Analysis > Video

Video

[Bookmark this page](#)

Introduction to Data Science and Data Analysis

[Start of transcript. Skip to the end.](#)

We live in an amazing age.
There's more collective data available to us now than there has ever been. Since the dawn of mankind, incidentally most of that data consist of cat videos and Netflix streams. But irrespective, since the dawn of mankind it's a great time to be studying the field of data

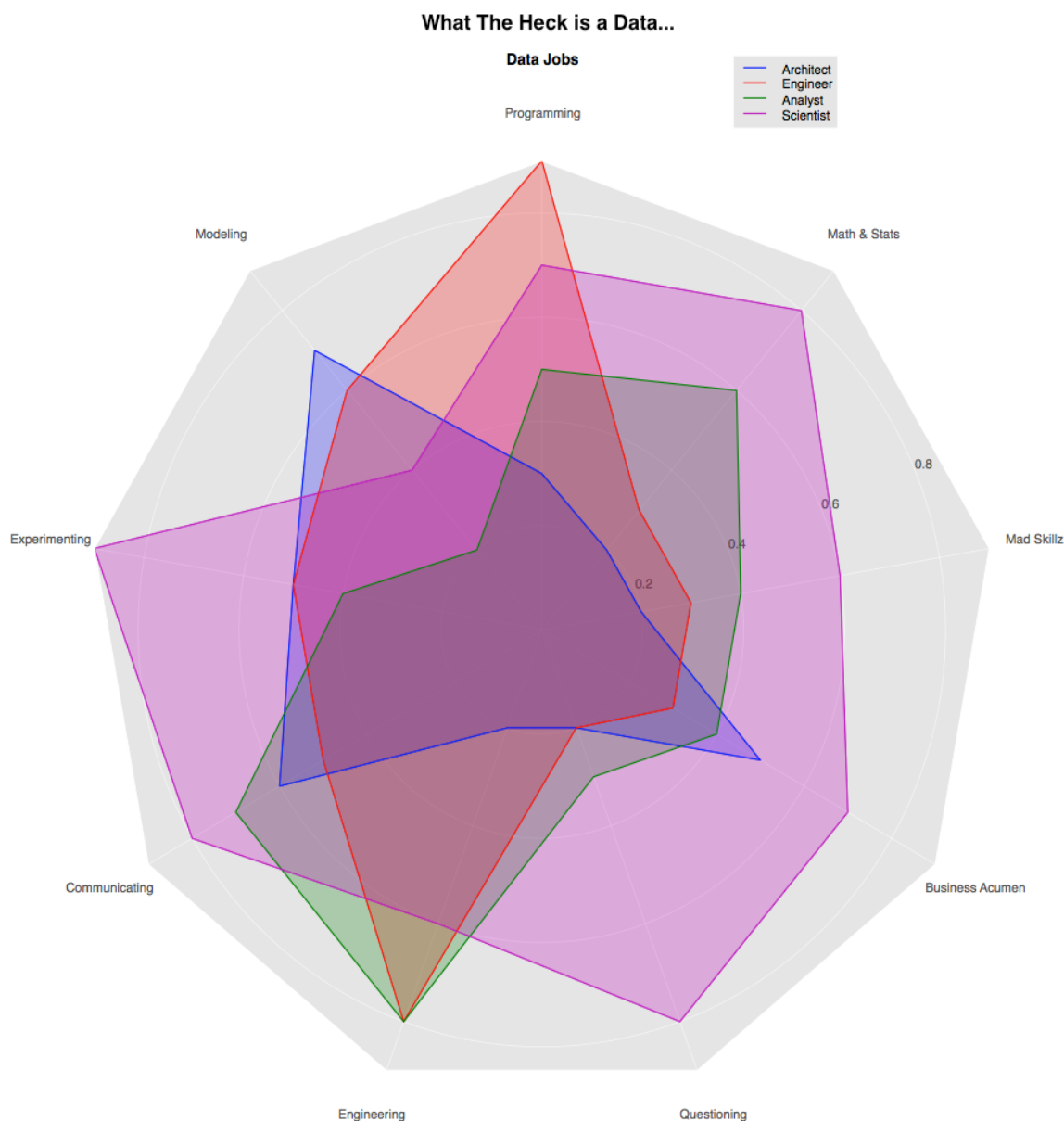
Video

[Download video file](#)

Transcripts

[Download SubRip \(.srt\) file](#)[Download Text \(.txt\) file](#)

There are so many definitions for *data scientists* out there that it's honestly a bit tiring keeping up. With the added case of slightly conflicting explanations, who's to say who has the right interpretation? When in doubt, combine them all and examine the data:



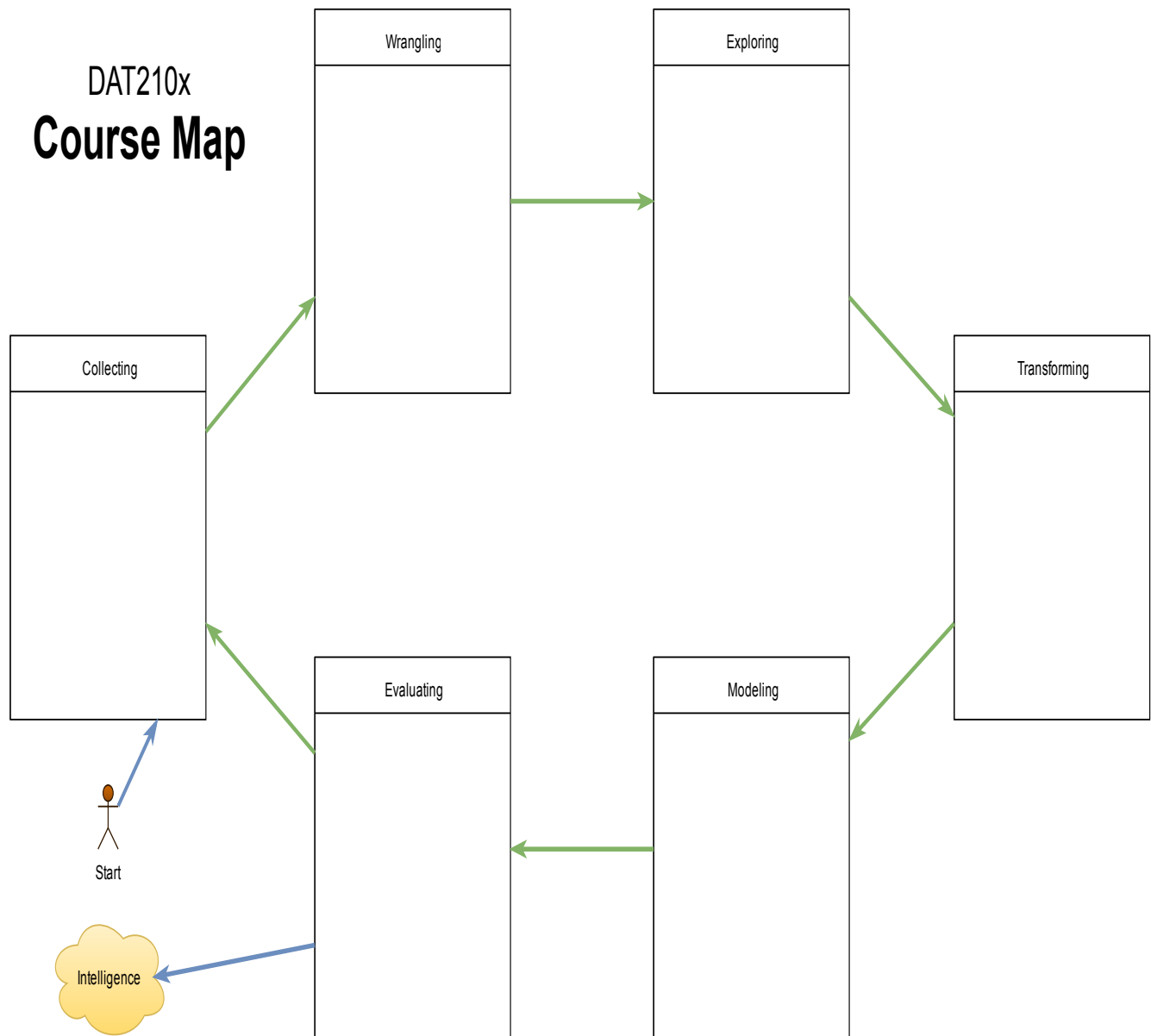
We live in an age where more information exist than ever before, and it's also easier and cheaper to collect it, store it, and process mass quantities of it. Due to this, data related jobs are popping up by the minute. In this course, you're going to touch on many of the facets of the job of data analysts, architects, engineers, and scientists... with a focus on analysis. At least from the chart above, it seems that analysts are mini-scientist in the making.

Data analysts carry out their job of engaging data using a well refined, multi-phased process. The goal of their job is to interpret raw data by converting it into useful, and actionable intelligence. What does this process entail exactly?

- **Collecting Data** from various sources
- **Wrangling Data** to make it more reliable
- **Exploring Data** using statistics and visualizations
- **Transforming Data** to prepare it for modeling

- **Modeling Data** using the right machine learning algorithms
- **Evaluating** the results of the data models

Our goal is to have you become comfortable with these steps. To make the learning process more interesting, let's gamify things a bit. Here is a course-map of the data analysis process, and your goal is to get to the golden **intelligence** spot by the end of the course:



A PDF of this map has been included in the course's GitHub repository.

As you progress through the course, be sure to annotate the boxes in the map with everything you are learning. Once you've completed the course, this map might serve as a source of inspiration for how far you've come. It can also act as a template for when you wish to apply data analysis again in the future.

For now, let's discuss the first step on the chart: **collecting data**. As you do that, you should have a question in mind to drive your data collection process. That's a topic in and of itself, so we'll discuss it further in the next lecture.

To be effective at analysis, you are going to need data and a lot of it. Simply collect as much data as you possibly can for whatever it is you are trying to study. Fine-tuning can and will come later. If you fail now at collecting at least the minimal amount of data necessary to run through the analysis process, you'll find yourself repeating data collection and analysis work instead of arriving at actionable intelligence.

Data might be collected from a variety of sources in the physical world, such as thermostats, smart cars, satellite transmissions, cameras, logs, and the Internet. As a data scientist, you'll usually be given data by your clients and supervisors. But if you work on your own passion projects, it'll be your own responsibility to amass data for analysis.

As you run data through the process outlined above, whether the data was spoon-fed to you or collected by you yourself, keep in mind that you might discover you actually need more of it before you can get to intelligence.

© All Rights Reserved



© 2012-2017 edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc. | 粤ICP备17044299号-2

POWERED BY
OPENedX®

