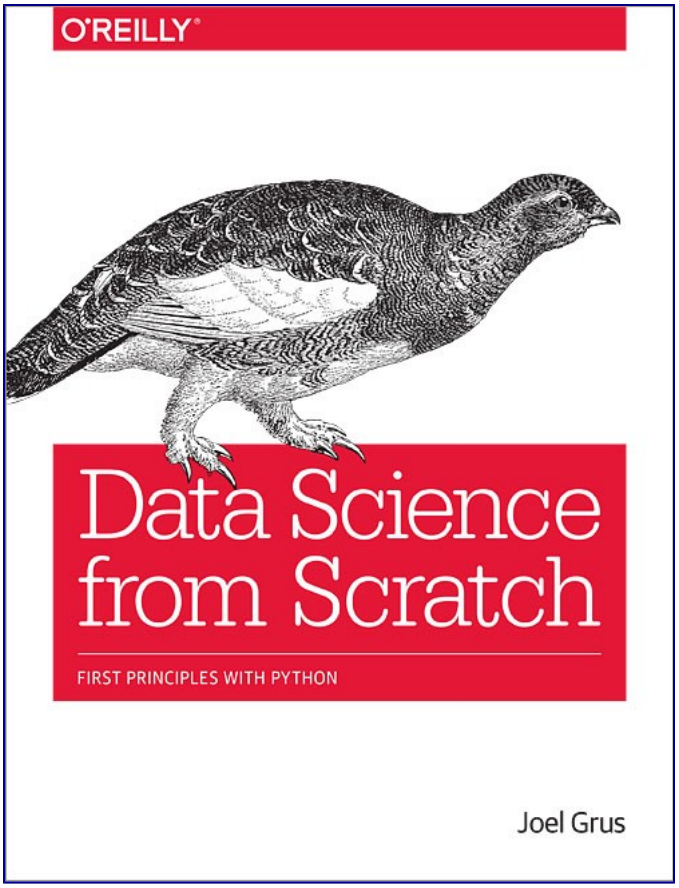
joelgrus.com/2015/04/26/data-science-from-scratch-first-principles-with-python/

Data Science From Scratch: First Principles with Python

I am super-excited to announce that the book I've been working on for more than the last year, *Data Science from Scratch: First Principles with Python* is finally available! (buy from O'Reilly, use discount code AUTHD to save some money) (buy from Amazon).

My experience learning and teaching data science was that there were two primary paths:

1. The Math Path: "So you want to be a data scientist? Sure, the first thing you need to know is *matrix decompositions*. How well do you remember your measure theory?"
2. The Tools Path: "So you want to be a data scientist? Great, here's the most important libraries to know. How well do you know R?"



Although I am myself a "math person", the first approach never resonated with me. The fun of data science for me has always been *working with data*. At the same time, I've never been thrilled with the second approach -- it's a good way to start doing data science without ever really understanding what you're doing.

My ideal would be a "third way" between these approaches:

1. understanding the behavior of the most common tools by working through a solid-but-less-than-textbook-rigorous understanding of the math behind them, and
2. implementing simplified versions of them *from scratch* to understand exactly what it is they're doing.

Hence *Data Science from Scratch*. It's got math, but only as much as is totally necessary. It's got scraping and cleaning and munging. It's got machine learning. It's got databases and MapReduce. Necessarily it doesn't go *deep* into any of these, but I like to think it establishes a broad, solid foundation for someone who knows some math and some programming but is not (necessarily) an expert at either.

Many technical books (I won't name names) explain things in their text and then dump pages of hard-to-follow code at you that you are expected to puzzle through. I spent a lot of time trying to write clean code that illuminated the concepts *on its own* and that *reinforced* the ideas from the text. As is the current fashion these days, all of the code and data is on GitHub, if you'd like to get a sense of what the book is about.

If you are interested in the topic, I encourage you to check it out, write a review, and let me know what you think! (You can see the full table of contents on the O'Reilly page.)

Posted on: 2015-04-26

Category: Life, Data Science, Writing

Data Science from Scratch
First Principles with Python
By Joel Grus
Publisher: O'Reilly Media
Final Release Date: April 2015
Pages: 330

Data science libraries, frameworks, modules, and toolkits are great for doing data science, but they’re also a good way to dive into the discipline without actually understanding data science. In this book, you’ll learn how many of the most fundamental data science tools and algorithms work by implementing them *from scratch*.

If you have an aptitude for mathematics and some programming skills, author Joel Grus will help you get comfortable with the math and statistics at the core of data science, and with hacking skills you need to get started as a data scientist. Today’s messy glut of data holds answers to questions no one’s even thought to ask. This book provides you with the know-how to dig those answers out.

- Get a crash course in Python
- Learn the basics of linear algebra, statistics, and probability—and understand how and when they're used in data science
- Collect, explore, clean, munge, and manipulate data
- Dive into the fundamentals of machine learning
- Implement models such as k-nearest Neighbors, Naive Bayes, linear and logistic regression, decision trees, neural networks, and clustering
- Explore recommender systems, natural language processing, network analysis, MapReduce, and databases

joelgrus.com/2015/04/26/data-science-from-scratch-first-principles-with-python/

Chapter 7 Hypothesis and Inference

Chapter 8 Gradient Descent

Chapter 9 Getting Data

Chapter 10 Working with Data

joelgrus.com/2015/04/26/data-science-from-scratch-first-principles-with-python/