# Supercharge Your Data Science Career!

88 Free, Hand-Picked Resources Every Data Scientist Should Have

EliteDataScience

# INTRODUCTION

Welcome to the Elite Data Science community!

We want to say a hearty **thank you from the bottom of our hearts** for joining.

You see, one of the challenges we've personally experienced as data scientists is knowing what to learn next to advance our careers...

First, there are so many topics that we need to master. Many companies simply see us as "data magicians." *Here's a bunch a data, now can you make us a lot more money?*

Second, because this is a relatively new field, so many hidden gem resources are scattered across the web, waiting to be found.

Therefore, the onus is on *us* as data professionals to keep our skills sharp and continue to round out our skill-set, and we hope to help you on that quest.

In the next few pages, we're going to share a hand-picked list of some of the best *free online resources* we've enjoyed and benefited from ourselves. A couple things...

**The list is not complete in breadth.** Data science is a rich and diverse field that's growing rapidly, and we're learning alongside everyone else. If you've come across a great resource that's not on here, we'd love for you to share it with us!

**This list is not complete in depth.** Some topics on here have entire sub-industries devoted to them. Let this be a curated sampling across a variety of topics and a quickstart reference guide for you as you "choose your own adventure." We invite you to visit **elitedatascience.com** for regularly updates.

Without further ado... onward we go!

# TABLE OF CONTENTS

# CHAPTER 1

# STATISTICS & PROBABILITY

### Statistics and Probability (Khan Academy)

Practical introduction to statistics and probability from Khan Academy. Recommended for getting up to speed quickly.

### Harvard Stats 110: Probability (Video Series)

Rigorous treatment of probability theory from Harvard. Recommended for building deeper mastery.

### Think Stats: Probability and Statistics for Programmers (PDF)

Excellent resource for those with programming backgrounds. Quote: “The thesis of this book is that if you know how to program, you can use that skill to help you understand probability and statistics.”

### Crash Course on Basic Statistics (PDF)

Short PDF that covers a whirlwind review of key topics. We like this review sheet because it has simple intuitive explanations for each concept.

### Stanford CS229: Probability Review (PDF)

Short PDF that covers a whirlwind review of key topics needed for machine learning. Assumes knowledge of linear algebra and calculus.

### Introduction to Probability (PDF)

*Reference text*. Textbook that covers every major topic in probability and statistics.

### Introduction to Probability and Statistics using R (PDF)

*Reference text*. Probability textbook with applications in R.

# CHAPTER 2

# PROGRAMMING

**Python**

Learn Python the Hard Way (Online Book)
Designed for beginners who want a complete course in programming with Python.

Introduction to Python for Data Science (Online Course)
Recommended for those with programming experience who only need a crash course on the basic Python tools needed for data science.

LearnPython.org (Interactive Tutorial)
Short, interactive tutorial for those who just need a quick way to pick up Python syntax.

How to Think Like a Computer Scientist (Interactive Tutorial), (PDF Version)
Interactive "Computer Science 101" course taught in Python that really focuses on the art of problem solving. Wonderful gem.

PythonChallenge.com (Online Puzzle)
Fun online puzzle with 33 levels that you can solve with Python programming.

**R**

Swirl (Interactive R Package)
Very cool R package that you can install and learn the language directly from inside RStudio (the most common interface used to run R).

R for Data Science (Online Book)
Recommended for beginners who want a complete course in data science with R.

Introduction to Data Science with R (Video Series)
For those who learn better by watching someone else walk through the steps.

# CHAPTER 3

# LINEAR ALGEBRA

### Linear Algebra Review for Machine Learning (Video Series)

These are the optional linear algebra review videos for Andrew Ng's machine learning course (more on this later). The entire 6-part series can be watched in under 1 hour. Recommended if you've taken linear algebra before and just need a quick review.

### Linear Algebra (MIT OpenCourseWare)

Rigorous linear algebra class from MIT. Recommended for those who intend to apply for R&D-heavy data scientist positions.

### Linear Algebra (Khan Academy)

Practical linear algebra lessons from Khan Academy. Recommended for those who intend to apply for application-heavy data scientist positions (because it's quicker to complete).

### Matrix Algebra Review (PDF)

Review of matrix algebra. Great to use as a reference.

# CHAPTER 4

# MULTIVARIABLE CALCULUS

### Multivariable Calculus Review (Video)

This is quick review of multivariable calculus in the format of solving practice problems. Recommended if you've taken multivariable calculus before and just need a quick review.

### Multivariable Calculus (MIT OpenCourseWare)

Rigorous multivariable calculus class from MIT. Recommended for those who intend to apply for R&D-heavy data scientist positions.

### Multivariable Calculus (Khan Academy)

Practical multivariable calculus lessons from Khan Academy. Recommended for those who intend to apply for application-heavy data scientist positions (because it's quicker to complete).

# CHAPTER 5

# DATA COLLECTION

**API Tutorials**

Python: requests Quickstart Guide (Tutorial)

How to use the **requests** library to request data from API's.

R: httr Quickstart Guide (Tutorial)

How to use the **httr** library to request data from API's.

**Web Scraping Tutorials**

Python: Basic HTML Scraping (Tutorial)

Basic web scraping with the **lxml** and **requests** libraries.

Python: Selenium (Tutorial)

Useful for scraping websites that have Javascript (**selenium** replaces requests).

Python: BeautifulSoup (Tutorial)

Popular library for parsing web pages (**beautifulsoup** replaces lxml).

R: rvest (Tutorial)

Basic web scraping with the **rvest** library.

# CHAPTER 6

# SQL

Intro to SQL by Khan Academy (Course)
Comprehensive video series that covers every important SQL topic.

sqlcourse.com (Interactive Tutorial)
Great to use review or a quick crash course.

SQL Fundamentals (Course)
Course that covers the basics of SQL. Includes quizzes along the way to test your understanding.

CHAPTER 7

# DATA VISUALIZATION

## Data Visualization in Python (Video Series)

Tutorial on using the **matplotlib** library in Python for data visualization.

## Data Visualization in R (Video Series)

Tutorial on using the **ggplot** library in R for data visualization.

# CHAPTER 8

# GENERAL MACHINE LEARNING

### Machine Learning by Andrew Ng (Video Series)

This is the gold standard when it comes to Machine Learning courses. You'll walk away with a firm understanding of the theoretical underpinnings as well as recommendations on when to use different algorithms in practice.

### Harvard's CS 109: Data Science (Course)

Fantastic end-to-end general-purpose data science course that covers several machine learning models (in slightly less depth) than Andrew Ng's course. The course is taught in Python.

### R: caret package webinar (Video)

Introduction to the caret package in R, which is how algorithms are often implemented in practice.

### Python: scikit-learn quickstart (Tutorial)

Introduction to the sklearn package in Python, which is how algorithms are often implemented in practice.

### Elements of Statistical Learning (PDF)

*Reference text*. This is one of the classic textbooks of the industry. It assumes you have a fairly high level of math background.

### An Introduction to Statistical Learning in R (PDF)

*Reference text*. Another classic textbook that's a gentler introduction than Elements of Statistical Learning.

# CHAPTER 9

# COMMUNICATION

### The best stats you've ever seen (TED Talk)

This is an iconic TED talk and a fun display of storytelling with data.

### Think Fast, Talk Smart (Video)

This is a workshop at the Stanford Graduate School of Business on how to overcome anxiety and speak spontaneously. Not only will this help you for the rest of your career, but it will also allow you to stand out during your interview.

### 7 Tips for Improving Communication (Video)

Simple, practical tips on how to communicate effectively on a daily basis.

### How to Win Friends and Influence People (PDF), (Free Audiobook Version)

This is a book we'd recommend for anyone, data scientist or not. While some of the verbiage is a bit dated, the teachings about interpersonal relationships are timeless.

# CHAPTER 10

# BUSINESS INTUITION

### Data Driven Decisions (Video)

How to take business objectives, extract testable hypotheses from them, and then design experiments to evaluate.

### How to be data driven and build great products by DJ Patil (Video)

Lecture by DJ Patil before he become Chief Data Scientist of the USA.

### Big Data: New Tricks for Econometrics by Hal Varian (PDF)

Hal Varian, Chief Economist at Google, gives an excellent overview of the technology and methodology landscape for data analysis.

### How data will transform business (TED Talk)

Thought-provoking discussion of the relationship between business strategy and technology. Explains why the two long-standing theories of business strategy have become invalidated by the rise of big data.

### Victor Cheng's Case Interview Workshop (Video Series)

Some employers like to ask consulting-style "case" questions during the interview. This is more common for Data Scientists in business operations, strategy, or analytics roles. This is an excellent crash course on tackling case interviews.

# CHAPTER 11

# CREATIVITY & INNOVATION

### Machine Intelligence and Data Products (Video)
Future-looking discussion of data products and data science.

### Machine Intelligence Landscape (Chart)
Venture capitalist's perspective on the landscape of machine intelligence applications.

### The art of innovation (TED Talk)
Masterclass on innovation by Guy Kawasaki.

### 7 steps of creative thinking (TED Talk)
Creative thinking tips from the perspective of a serial artist and entrepreneur.

### Working backwards to solve a problem (TED Talk)
Chess grand-master Maurice Ashley on how to see the endgame and work backwards.

### Crunchbase (Database)
Database of the newest startups, searchable by keywords

# CHAPTER 12

# BUSINESS ANALYTICS

### Introduction to Business Analytics (Video)

Short and sweet intro to how businesses use analytics to make better decisions, including case studies.

### Marketing Metrics and Analytics (Video)

Introduction to common metrics and analytics methods using in marketing.

### Effective Cross-Selling using Market Basket Analysis (Tutorial)

How to do smarter cross-selling.

### An Intuitive Guide to A/B Testing (Video)

Overview of A/B testing and interpretation.

### 25 Examples of Business KPIs (Examples)

They say what gets measured gets managed. Here are 25 examples of business Key Performance Indicators (KPIs) that are commonly used to make better decisions.

### Analytics Academy by Google (Courses)

Practical courses on digital analytics, e-commerce analytics, and other topics.

# CHAPTER 13

# TEXT ANALYSIS (NLP)

Stanford NLP (Video Series)
Full course on “traditional” Natural Language Processing, including sentiment analysis, Naive Bayes models, n-grams, etc.

Deep Learning for Natural Language Processing (Course), (Course materials)
The current bleeding edge of Natural Language Processing. You should finish Andrew Ng's machine learning course first.

The Unreasonable Effectiveness of Recurrent Neural Networks (Tutorial)
Fantastic breakdown of recurrent neural networks, which are special applications of deep learning especially successful in natural language processing.

Recurrent NN in Keras (Tutorial)
Step-by-step tutorial of implementing a recurrent neural network in Python’s **keras** package.

# CHAPTER 14

# RECOMMENDATION SYSTEMS

### Recommendation engine tutorial (Video Series)

Introduction to collaborative filters using Python. Does a very nice job of explaining the intuition behind the algorithm.

### Recommender Systems (Video Series)

Discussion of the theory and math behind collaborative filters by Andrew Ng. More math-heavy, and it'll be easier to follow if you have some background with Linear Algebra.

### Collaborative Filtering with Python (Tutorial)

Reference tutorial that implements a music recommender system in Python.

### Collaborative Filtering with R (Tutorial)

The same tutorial as the previous one, except in R.

CHAPTER 15

# TIME SERIES ANALYSIS

### Time Series (Course Material)

Lecture slides, homework, and R Code for the Time Series course at Oregon State University.

### The Little Book of R for Time Series (Online Book)

Very practical step-by-step introduction to using R for time series analysis. Includes code and outputs for each step.

### Time Series Forecasting with Python (Tutorial)

Tutorial on performing time series visualization, analysis, and forecasting with Python.

### Seasonal ARIMA with Python (Tutorial)

Introduction to ARIMA models in Python. Includes all code.

### Statistical forecasting, Fuqua School of Business (Online Book)

Course notes from the statistical forecasting course taught at the Fuqua School of Business at Duke University.

# CHAPTER 16

# DEEP LEARNING

Neural Networks and Deep Learning (Online Book)
Relatively little-known hidden gem, but one of our favorite resources for learning about neural networks. Explanations are clear and intuitive.

Unsupervised Feature Learning and Deep Learning (Online Book)
Comprehensive online book that covers a wide range of topics in deep learning.

Tech Talks by Yann LeCun (Videos)
Tech talks by Yann LeCun, one of the "Godfathers" of modern deep learning.

Neural Networks for Machine Learning (Video Series)
Course taught by Geoff Hinton, one of the other "Godfathers" of modern deep learning.

Hacker's Guide to Neural Networks (Tutorial)
Neural networks and deep learning taught from the perspective of a computer scientist. Heavy on code and light on math.

Stanford 231n: Convolutional Neural Networks (Course Notes), (Lecture Videos)
Rigorous course on convolutional neural networks for computer vision.

Deep NN with Tensorflow (Tutorial)
Step-by-step tutorial for building deep neural networks with Google Tensorflow.

Build Your Own NN in R (Tutorial)
Building a neural network from scratch in R.

# CHAPTER 17

# ANOMALY DETECTION

### Anomaly Detection (Video Series)

Part of Andrew Ng's excellent machine learning course. We recommend starting here.

### Practical Machine Learning: A New Look at Anomaly Detection (PDF)

Short (66-page) textbook on anomaly detection. Excellent introduction with intuitive explanations.

### A Review of Machine Learning based Anomaly Detection Techniques (PDF)

Short academic overview of anomaly detection techniques. Useful to get a lay of the land.

### Novelty and Outlier Detection (Tutorial)

Tutorial using the **sklearn** library to perform anomaly detection in Python.

### Anomaly Detection in R (Tutorial)

Tutorial using the **AnomalyDetection** package in R.