[Vipin Tyagi](https://www.quora.com/profile/Vipin-Tyagi-9), Did several Data Science Projects in infrastructure arena

[Answered Nov 15, 2015](https://www.quora.com/Are-there-any-good-YouTube-channels-that-an-amateur-data-scientist-should-follow/answer/Vipin-Tyagi-9)

As Data science is a very comprehensive topic, we will divide into various sections as YouTube videos can be separately listed:

**1. Machine learning**  
**2.  Statistics**  
**3. Python**  
**4. R**  
**5. Databases and SQL**  
**6. Spark, Hadoop and distributed computing**

Let us discuss one by one:  
  
**Machine Learning:**  
These resources are suitable for beginners as well as experienced data scientists.

* [Introduction to Recommender Systems](https://www.youtube.com/watch?v=bLhq63ygoU8) is a 4-hour lecture of the 2014 Machine Learning Summer School at CMU. You can find other interesting machine learning lectures from the same summer school and other programs in [Alex Smola’s YouTube channel](https://www.youtube.com/user/smolix/playlists).
* [This PyData talk](https://www.youtube.com/watch?v=W9_SNGymRwo) is a good introduction to deep neural networks in general and to convolution networks in particular.
* “[Machine Learning Gremlins](https://www.youtube.com/watch?v=tleeC-KlsKA)” is a good presentation on common machine learning mistakes by Ben Hamner (Kaggle).
* Because we mayn't always need precise answers, this [introduction to stream mining](https://www.youtube.com/watch?v=u08tNk2jgPY) by Mikio Braun can be very useful to you.
* In order to widen your horizon about artificial intelligence, [these lectures](https://www.youtube.com/playlist?list=PLnvKubj2-I2LhIibS8TOGC42xsD3-liux) from the AI course taught at MIT by Patrick Winston.
* The lectures of the course “CS273a: Introduction to Machine Learning” by Prof. Alex Ihler (UCI) are [available on Youtube](https://www.youtube.com/playlist?list=PLaXDtXvwY-oDvedS3f4HW0b4KxqpJ_imw).
* If you want to know more about the state of the art of neural networks and deep learning, you can find tutorials and workshops of the NIPS 2014 conference in [this YouTube channel](https://www.youtube.com/user/NeuralInformationPro/feed). You can also find this [summary of the conference](http://blogs.technet.com/b/machinelearning/archive/2014/12/16/machine-learning-trends-from-nips-2014.aspx) by John Platt (Microsoft Research).

Introductory Data science courses on YouTube can be accessed:  
  
**Intro to Data Science** [UW / Coursera](http://bit.ly/uwintrodatascience)

* *Topics:* Python NLP on Twitter API, Distributed Computing Paradigm, MapReduce/Hadoop & Pig Script, SQL/NoSQL, Relational Algebra, Experiment design, Statistics, Graphs, Amazon EC2, Visualization.

**Data Science** / Harvard [Video Archive](http://bit.ly/harvarddatasciencevideos) & [Course](http://bit.ly/harvarddatasciencecourse)

* *Topics:* Data wrangling, data management, exploratory data analysis to generate hypotheses and intuition, prediction based on statistical methods such as regression and classification, communication of results through visualization, stories, and summaries.

**Statistics:**

* [Top 10 big ideas covered in the Probability course at Harvard](https://www.quora.com/What-are-the-top-10-big-ideas-in-Statistics-110-Introduction-to-Probability-at-Harvard) by Joe Blitzstein. You can also [watch on Youtube](https://www.youtube.com/playlist?list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo) the lectures of this course.
* Brian Caffo is well known lecturer of the Data Science specialization on Coursera and his YouTube channel is full of [resources to learn statistics](https://www.youtube.com/user/bcaffo/playlists).

**Python:**  
Python plays role of connecting various data science softwares and is very versatile and east to learn.

* Video tutorials to learn [how to use Python’s scikit-learn library to perform machine learning](http://ow.ly/PoAnU) by Kevin Markham is an excellent resource.
* [3h+ in-depth introduction to machine learning with scikit-learn](https://www.youtube.com/watch?v=80fZrVMurPM) by Kyle Kastner (Université de Montréal) and Andreas Mueller (NYU Center for Data Science).
* Make your first machine learning predictions using Python with [this Kaggle tutorial](https://www.kaggle.com/c/titanic/details/getting-started-with-python).
* NLTK is the most popular library for natural language processing in Python. [This presentation](http://www.slideshare.net/japerk/nltk-in-20-minutes) can give you a good overview of what you can do with it and this [1 hour tutorial](https://www.youtube.com/playlist?list=PLjzUxUKyAHG0UeXr09D7fP1srkx32MZQK) will show you what you can do with it.
* [PyDataTV](https://www.youtube.com/user/PyDataTV/playlists) is the YouTube channel of the PyData conferences. You can find keynotes, talks and workshops on how to use the PyData stack.

**R:**  
R appears to be very simple but has a harder learning curve afterwards. It is simple yet powerful for handling big data and great visualization.

* [Intro to R](https://www.youtube.com/playlist?list=PLOU2XLYxmsIK9qQfztXeybpHvru-TrqAP) is a playlist by Google Developers that explains all the basics of the language
* [An Introduction to Statistical Learning with Applications in R](http://www-bcf.usc.edu/%7Egareth/ISL/) is a wonderful free book full of examples.
* qdap is not only one of the best packages for natural language processing in R, but also one of the best documented one. Use [the vignette](https://trinker.github.io/qdap/vignettes/qdap_vignette.html) to get started with it and later on [the manual](http://chrome-extension:/oemmndcbldboiebfnladdacbdfmadadm/https:/dl.dropboxusercontent.com/u/61803503/qdap.pdf).
* It has been the experience of data scientists that R’s memory limitations can give you a headache. [These tricks](http://stackoverflow.com/questions/1358003/tricks-to-manage-the-available-memory-in-an-r-session) are sometimes an effective painkiller. Also the slides “[Taking R to the Limit: Large Datasets](http://ow.ly/xCZvs)” might help.
* [statsTeachR](http://ow.ly/wXwue) is a repository of lessons for teaching statistics using R.
* Try to make your first machine learning predictions using R with any of [these four tutorials](https://www.kaggle.com/c/titanic/details/new-getting-started-with-r).

**Databases and SQL**  
“If you’re doing data science/analysis, learn SQL. People hate on it, but it’s important. World of tech is built on it.” [Greg Reda](https://twitter.com/gjreda/status/627205410987053056)

* If you want to know how a DBMS works, check the [videos of the Standford Introduction to Databases course](https://www.youtube.com/playlist?list=PL6hGtHedy2Z4EkgY76QOcueU8lAC4o6c3) by Jennifer Widom.
* [Apache Calcite](https://www.youtube.com/watch?v=5_MyORYjq3w) is a query planner and optimizer for any kind of data sources. If you want to get familiar with how a 21st century query planner and optimizer looks like, this is a good starting point.
* If you are a beginner to SQL and MySQL is your choice, try this tutorial that claims to cover [95-98% of everything you’ll ever need to know in MySQL](https://www.youtube.com/watch?v=yPu6qV5byu4) or [this playlist](https://www.youtube.com/watch?index=33&list=PL32BC9C878BA72085&v=DOLL20iUUXg).

**Spark, Hadoop and Distributed Computing:**  
If you want to deal with big data then you can use these resources:

* [Data Science training with Spark](https://www.youtube.com/watch?v=oTOgaMZkBKQ), a 5+ hour video from the Spark Summit.
* [Introduction to Spark SQL](https://www.youtube.com/watch?v=GQSNJAzxOr8) and its rule-based optimizer by Michael Armbrust
* [Presentation on Parquet](https://www.youtube.com/watch?v=pFS-FScophU), the easiest way to columnar storage in Hadoop, at Hadoop Summit
* Apache Flink is more suitable than Spark for iterative and also streaming processes. The best resources to learn about it is the [official YouTube channel](https://www.youtube.com/channel/UCY8_lgiZLZErZPF47a2hXMA/videos) and [DataArtisans YouTube channel](https://www.youtube.com/channel/UCxt7B_fmLplq0OMRVxLiE8w/videos).

List is not exhaustive. I will keep on adding up as they come up.

7.3k Views · [View Upvoters](https://www.quora.com/api/mobile_expanded_voter_list?key=jkO1QmANysQ&type=answer)