## **Data Science: Getting Started**

Data Science is a multidisciplinary field covering at the very minimum - statistics, programming, machine learning [Drew Conway's venn diagram](http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram) or [Cheat Sheet of a Modern Data Scientist](http://www.marketingdistillery.com/2014/08/30/data-science-skill-set-explained/).=

### **Begin**

* [Data Science Pipeline](http://machinelearningmastery.com/wp-content/uploads/2014/05/Overview-of-the-Applied-Machine-Learning-Process.png) - Detailed overview of data pipeline from [MachineLearningMastery.com](http://machinelearningmastery.com/)
* [Data Science In colab notebooks](https://sanjeevsahu.xyz/Data-science-tutorials-68d35131cfa54e94b85029b5eccf1602) - A series of colab notebooks and Jupyter notebooks to help you with DS pipeline.
* [Intro to ipython](https://github.com/ipython/ipython/wiki/A-gallery-of-interesting-IPython-Notebooks/_edit#entire-books-or-other-large-collections-of-notebooks-on-a-topic) - A curation of Ipython Notebooks great for introductory level to python, programming, comp sci, data science and other topics.
* [How do I Become a Data Scientist?](http://www.quora.com/How-do-I-become-a-data-scientist) - Some more great starting points from William Chen.

### **Data Science Courses:**

* [Coursera](https://www.coursera.org/specialization/jhudatascience/1) - Data Science Specialization at Coursera - many other courses available as well.
* [Udacity](https://www.udacity.com/courses#!/data-science) - Online MOOCs that are the Data Science related courses. by I
* [Data Science Bootcamps](http://yet-another-data-blog.blogspot.com/2014/04/data-science-bootcamp-landscape-full.html) - A collection of all bootcamps currently on the market as of April 5, 2014 by Ikechukwu Okonkwo.
* [Coursera Machine Learning Course](https://www.coursera.org/course/ml) - Andrew Ng's pinnacle Machine Learning course.
* [Edx](https://www.edx.org/course/mitx/mitx-6-00-2x-introduction-computational-2836#.VEANx9TF-tw) - EDX courses related to data science.
* [Mlcourse.ai](http://mlcourse.ai) - Free courses on complete DS pipeline.
* [Complete lecture notes of the Stanford/Coursera Machine Learning class by Andrew Ng](http://www.holehouse.org/mlclass/)
* [200 universities just launched 560 free online courses. Here’s the full list.](https://medium.freecodecamp.org/200-universities-just-launched-560-free-online-courses-heres-the-full-list-d9dd13600b04)
* [Artificial Intelligence | MIT OpenCourseWare](https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/index.htm)
* [Dashboard | MIT Professional Education Digital Programs](https://mitprofessionalx.mit.edu/dashboard)
* [Data Science A-Z™: Real-Life Data Science Exercises Included | Udemy](https://www.udemy.com/datascience/)
* [Data Science Essentials | edX](https://www.edx.org/course/data-science-essentials-microsoft-dat203-1x-2?source=aw&awc=6798_1489913955_d9818a031ea60b9e133f81baa8e0fcbb&utm_source=aw&utm_medium=affiliate_partner&utm_content=text-link&utm_term=315645_LearnDataSci)

## **Data Pipeline & Tools**

* [Python @ Codecademy](http://www.codecademy.com/en/tracks/python) - If you have never used Python, right this way..
* [The Python Wiki](https://wiki.python.org/moin/FrontPage) - Good resource with lots of info about Python.
* [Python for Data Science Tutorial - Kaggle](https://www.kaggle.com/wiki/GettingStartedWithPythonForDataScience) - Stepping into Data Science with Kaggle and installing some libraries.
* [Introduction to Data Processing with Python](http://opentechschool.github.io/python-data-intro/) - Just as the name says - some introductory level information and exercises.
* [Git tutorial](https://try.github.io/levels/1/challenges/1) - Git for Version Control. Simple tutorial for Git from Github.
* [Git Tips](http://www.alexkras.com/19-git-tips-for-everyday-use/) - 19 git tips for everyday use.
* [Anyone Can Code](http://dhruvbird.com/61.html) - Languages, tutorials, cheat sheets, algorithms and data structures
* [Data Cleaning](https://towardsdatascience.com/your-ultimate-data-manipulation-cleaning-cheat-sheet-731f3b14a0be)

### **Data Structures & CS Topics**

* [Algorithms & Data Structures](http://www.bogotobogo.com/Algorithms/algorithms.php) - Binary trees, hash tables, linked lists, big(O) notation and more.
* [Algorithm & Data Structures](http://interactivepython.org/courselib/static/pythonds/index.html) - Well organized detailed and digestible site full of content covering data structures, algorithms, recursion and assignments!
* [Big O Notation](http://interactivepython.org/courselib/static/pythonds/AlgorithmAnalysis/BigONotation.html) - Great details and visual of big-O notation.
* [Visualizations of Data Structures](http://www.cs.usfca.edu/~galles/visualization/Algorithms.html) - Collection of different algorithms (graph problems) and data structures (queues, heaps, hashes) that walks through the visualization to get a better intuitive understanding.
* [Data Structures CheatSheet & Big Oh Notation](http://bigocheatsheet.com/)
* [Data Structures CheatSheet -smaller more readable](https://www.clear.rice.edu/comp160/data_cheat.html)
* [Coursera: Stanford Algorithms Design & Analysis](https://class.coursera.org/algo-006) - Course on algorithm design & analysis

**Statistics**

Some primers on understanding statistics and other resources to get a deeper understanding.

* [Statistics Without the Agonizing Pain](https://www.youtube.com/watch?v=5Dnw46eC-0o) - John Rauser's really great video on statistics - funny and engaging with a good message.
* [Probability Programming and Bayesian Methods for Hackers](http://nbviewer.ipython.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Prologue/Prologue.ipynb) - full book all online through ipython notebooks.
* [Probabilistic Programming and Bayesian Methods for Hackers](https://github.com/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers) - Github Repo for the book above.
* [Statistics Cheat Sheet in Ipython Notebook](http://nbviewer.ipython.org/url/trust.sce.ntu.edu.sg/~gguo1/blogs/Statistics/Statistics.ipynb)
* [The only probability Cheatsheet you'll ever need](https://bayesrule.files.wordpress.com/2014/07/probability_cheatsheet_140718.pdf) - Self explanatory - (thanks William Chen @ <http://datastories.quora.com/>) for pointing me this great cheat sheet out - wish I had that back at college.
* [Khan Academy: Statistics](https://www.khanacademy.org/#statistics) - Tons of videos to help learn statistics concepts.
* [Statistical Distributions in iPython Notebook](http://nbviewer.ipython.org/urls/gist.github.com/mattions/6113437/raw/c5468ea930d6960225d83e112d7f3d00d9c13398/Exploring+different+distribution.ipynb) - Discrete, Bernoulli, Poisson, Binomial, Alpha, Beta etc. The descriptions are mathematical - will find another resource to explain.

Data Acquisition Libraries that are very helpful for abstracting away some of the complications of scraping or working with HTTP.

* [BeautifulSoup](http://www.crummy.com/software/BeautifulSoup/) - A python library to make web-scraping HTML easier.
  + [Beautiful Soup Cheat Sheet](http://youkilljohnny.blogspot.com/2014/03/beautifulsoup-cheat-sheet-parse-html-by.html)
* [Requests](http://docs.python-requests.org/en/latest/) - HTTP for Humans - python library that makes working with http and api's more effortless

**Databases/Frameworks**

A collection of databases & frameworks that are helpful for data management and are the industry standard.

* [SQL](http://www.postgresql.org/) - SQL Database - I linked to Postgres since that is the version I use.
* [Psycopg](http://initd.org/psycopg/) - Python <> Postgres. Able to adapt PostgreSQL for the python environment.
  + [SQL Cheet Sheet](http://www.sql-tutorial.net/sql-cheat-sheet.pdf)
  + [SQLZoo](http://sqlzoo.net/wiki/Main_Page) - Develop your skills
  + [SQLSchool](http://sqlschool.modeanalytics.com/) - Develop your skills [MongoDB](http://www.mongodb.org/) - NoSQL database
* [Mode.com](https://mode.com/sql-tutorial/) - Best tutorial on data analysis. It also has its own software that can be used for free.

**Machine Learning**

There is a lot of information available online about the theory, mathematical intuition, tuning for this discipline. Here are some tools that are currently available.

* [A visual introduction to Machine Learning](http://www.r2d3.us/visual-intro-to-machine-learning-part-1/) - Awesome d3 visualization to help understand machine learning.
* [SciKit-Learn](http://scikit-learn.org/stable/) - Simple and efficient machine learning tools for data mining and data analysis
* [NLTK](http://www.nltk.org/) - Natural Language Toolkit to work with human languages data.
* [Tour of Machine Learning Algorithms](http://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/) - Blog post about some of the high level ML methods
* [VIDEO - How to get started w/mL](https://www.youtube.com/watch?v=uBorfxosVYs) - Melanie Warrick @ PyCon 2014.
* [Some ML methods classified](http://nyghtowlblog.files.wordpress.com/2014/04/ml_algorithms.png?w=535&h=311) - Classification for some sample ML algorithms by Melanie Warrick.
* [SciKit-image](http://scikit-image.org/) - Algorithms for image processing.
* [Machine Learning CheatSheet](https://github.com/soulmachine/machine-learning-cheat-sheet) - I would actually say this is more than just a cheat sheet given that there are > 100 pages of notes.
* [Awesome Machine Learning](https://github.com/josephmisiti/awesome-machine-learning) - List of machine learning libraries in all languages and also Kaggle competition source code by Joseph Misiti.

**Machine Learning Theory**

* [MathematicalMonk ML videos](https://www.youtube.com/playlist?list=PLD0F06AA0D2E8FFBA) - Amazingly concise and digestible videos detailing how different machine learning algorithms function (e.g. logistical, sums, knn, Bayes, etc.)
* [Logistic Regression Explained](http://www.appstate.edu/~whiteheadjc/service/logit/intro.htm#hypothesis) - Detailed explanation of how logistic regression works.
* [Video explaining how Random Forests Algorithm works](https://www.youtube.com/watch?v=o7iDkcpOr_g) - Random Forests Algorithm explained.
* [Random Forest Explained](http://citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics/) - Write up about Random Forest in layman's terms.
* [Machine Learning 101](http://www.erogol.com/large-set-machine-learning-resources-beginners-mavens/) - Large set of ML resources for beginners.

**Time-Series**

* [ANN & Computational Intelligence Forecasting Competition](http://www.neural-forecasting-competition.com/index.htm)
* [Neural Networks for Time Series Slidedeck](http://www.cs.cmu.edu/afs/cs/academic/class/15782-f06/slides/timeseries.pdf)

**Model Selection Resources about how to decide on your model.**

* [SciKit Learn Flow Chart for Model Selection](http://scikit-learn.org/stable/tutorial/machine_learning_map/index.html) - A helpful for a starting point selecting SKlearn algorithms.

**Model Evaluation Resources to help with understanding model evaluation.**

* [Evaluating ML Algorithms](http://machinelearningmastery.com/how-to-evaluate-machine-learning-algorithms/) - Blog Post from MachineLearningMastery about how to evaluate your performance.
* [Cross-Validation](http://robjhyndman.com/hyndsight/crossvalidation/) - Critical concept to evaluate the performance of your models.
  + [K-fold & Grid Search in Scikitlearn](http://randomforests.wordpress.com/2014/02/02/basics-of-k-fold-cross-validation-and-gridsearchcv-in-scikit-learn/) - Demo on how to implement kfold cross validation and grid-search using scikit-learn.
  + [Scikit-learn Cross Validation doc](http://scikit-learn.org/stable/modules/cross_validation.html) - Self explanatory title.
  + [Cross Validation - how to select your final Kaggle Model](http://www.chioka.in/how-to-select-your-final-models-in-a-kaggle-competitio/) - Importance of cross-validation described specifically in how it effects Kaggle competition scores.

## **Statistics**

* [Common statistical tests are linear models (or: how to teach stats)](https://lindeloev.github.io/tests-as-linear/)
* [Introductory statistics - OpenText Library](https://saylordotorg.github.io/text_introductory-statistics/index.html)
* [Common statistical tests are linear models (or: how to teach stats)](https://lindeloev.github.io/tests-as-linear/)
* [Background: Markov chains](https://d18ky98rnyall9.cloudfront.net/_adadc80290e52a99b282ca9d7c1a41ee_background_MarkovChains.html)
* [OpenIntro Stats](https://www.openintro.org/index.php)
* [Regression Analysis Tutorial and Examples | Minitab](http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-tutorial-and-examples)
* [The 10 Statistical Techniques Data Scientists Need to Master](https://towardsdatascience.com/the-10-statistical-techniques-data-scientists-need-to-master-1ef6dbd531f7)
* [The Ultimate Guide to 12 Dimensionality Reduction Techniques (with Python codes)](https://medium.com/analytics-vidhya/the-ultimate-guide-to-12-dimensionality-reduction-techniques-with-python-codes-2c2afdbc09e3)
* [Thomas Bayes and the crisis in science – TheTLS](https://www.the-tls.co.uk/articles/public/thomas-bayes-science-crisis/)
* [Welcome to STAT 505! | STAT 505](https://onlinecourses.science.psu.edu/stat505/node/1)
* [Introduction to Bayesian Linear Regression – Towards Data Science](https://towardsdatascience.com/introduction-to-bayesian-linear-regression-e66e60791ea7)
* [Regression Analysis Tutorial and Examples | Minitab](http://blog.minitab.com/blog/adventures-in-statistics-2/regression-analysis-tutorial-and-examples)
* [The 10 Statistical Techniques Data Scientists Need to Master](https://towardsdatascience.com/the-10-statistical-techniques-data-scientists-need-to-master-1ef6dbd531f7)
* [Welcome to STAT 505! | STAT 505](https://onlinecourses.science.psu.edu/stat505/node/1)
* [Probability and Statistics Visually](https://seeing-theory.brown.edu/)

### **Data Visualization**

Collection of the best libraries that I know for easy and powerful data visualizations.

* [ggplot](http://ggplot.yhathq.com/) - ggplot for python ported by the team at yhat.
* [matplotlib](http://matplotlib.org/) - Awesome plotting library for python.
* [d3](http://d3js.org/) - Mike Bostock's viz library - the de facto gold standard for polished visualization - in js, steep learning curve but beautiful outcomes.
* [bokeh](http://bokeh.pydata.org/) - Interactive visualization library.
* [d3py](https://github.com/mikedewar/d3py) - Another library for data viz.
* [vincent](http://vincent.readthedocs.org/en/latest/) - Help with python for d3.
* [seaborn](http://web.stanford.edu/~mwaskom/software/seaborn/) - Clean statistical data visualization library.

**Design Theory**

The importance of design theory in data visualization, storytelling and presentations could not be understated. It can take great content and make it confusing or virtually unusable, or it can make content sing and connect with the audience. Through better understanding of design theory, UI principles, a data scientist (or anyone) can convey more understandable information to the intended audience and give a strong story to their content.

* [Slidedeck on Data Storytelling & Visualization](https://www.notion.so/62ac7e8cc7ec4e98b58bdd540bf9e0d9) - Overview of different story structures and how to tell a story with data.
* [Accelerating Understanding Through Data Visualization](http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Accellerating-Understanding-Through-Data-Visualization.pdf) - Accenture White paper on Data Visualization

### Ipython Notebook Tutorials

Collection of ipython notebooks that are helpful as examples to either using tools or to explain certain topics.

* [Pandas Tutorial](http://nbviewer.ipython.org/github/twiecki/financial-analysis-python-tutorial/blob/master/1.%20Pandas%20Basics.ipynb) - Basic intro to Pandas in notebook form.
* [Pandas / Stats Tutorial](https://github.com/fonnesbeck/pytenn2014_tutorial) - Intermediate tutorial by Christopher Fonnesbeck Feb 2014.
* [Scipy Tutorial](http://nbviewer.ipython.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-3-Scipy.ipynb) - Basic Scipy Tutorial.
* [Numpy Tutorial](http://nbviewer.ipython.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-2-Numpy.ipynb) - Basic Numpy Tutorial.
* [Multiple Regressions using Statsmodels](http://nbviewer.ipython.org/urls/s3.amazonaws.com/datarobotblog/notebooks/multiple_regression_in_python.ipynb) - Using statsmodels for regression.
* [Intro to PyMC](http://nbviewer.ipython.org/github/fonnesbeck/Bios366/blob/master/notebooks/Section4_3-Introduction-to-PyMC.ipynb) - Intro to PyMC.
* [More on PyMC](http://nbviewer.ipython.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter2_MorePyMC/MorePyMC.ipynb) - More PyMC.
* [Kaggle Titanic Comp Tutorial](http://nbviewer.ipython.org/github/agconti/kaggle-titanic/blob/master/Titanic.ipynb) - Kaggle Titanic Tutorial using RandomForests.
* [Psycopg2 tutorial in Python](https://wiki.postgresql.org/wiki/Psycopg2_Tutorial) - How to use Psycopg2.
* [SQL in iPython](http://nbviewer.ipython.org/gist/catherinedevlin/6588378) - SQL in Python.
* [Beautiful Soup Tutorial](http://nbviewer.ipython.org/github/kcranston/2013-08-ku/blob/master/beautifulsoup/notebooks/00-BeautifulSoup.ipynb) - Beautiful Soup!
* [Sci-Kit Learn Basics](http://nbviewer.ipython.org/urls/raw2.github.com/yhat/DataGotham2013/master/notebooks/4%20-%20scikit-learn%20basics.ipynb?create=1) - Machine Learning Basics with scikit-learn.
* [MatPlotLib](http://nbviewer.ipython.org/github/jrjohansson/scientific-python-lectures/blob/master/Lecture-4-Matplotlib.ipynb) - Some of the possibilities of data-viz with MatPlotLib.
* [Choosing the right priors - Bayesian](http://nbviewer.ipython.org/github/CamDavidsonPilon/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/blob/master/Chapter6_Priorities/Priors.ipynb) - Bayesian statistics and prior selection.
* [Some Basic Data Analysis in Python](http://nbviewer.ipython.org/github/jvns/talks/blob/master/pyconca2013/pistes-cyclables.ipynb) - Basic data analysis with python.
* [Crash Course in Python for Scientists](http://nbviewer.ipython.org/gist/rpmuller/5920182) - Ipython Notebook for Scientists!
* [Regular Expressions](http://nbviewer.ipython.org/gist/rjweiss/7577022) - Regex to match patterns in strings - very powerful.

### **Data Sources**

Collection of sites to access data if you want to build out a project or just use some of the tools for EDA.

* [Data.Gov](https://www.data.gov/) - The US government portal to open data.
* [California Water Resources](http://www.water.ca.gov/data_home.cfm) - California's water resource data.
* [Data for Cool DS projects](http://101.datascience.community/2014/10/17/data-sources-for-cool-data-science-projects-part-1-guest-post/)
* [Academic Torrents](http://academictorrents.com/) - Sharing Data is hard, torrents make it easier for academics.
* [Data Basin](http://databasin.org/) - Science based mapping and analytics platform.
* [Open Energy Data Initiative](http://en.openei.org/wiki/Main_Page) - Over 800 data sets covering energy issues.
* [UCI Machine Learning Datasets](https://archive.ics.uci.edu/ml/datasets.html) - Data for machine learning - lots of labeled data and description of the problem types.
* [London Data Store](http://data.london.gov.uk/) - Lots of datasets on London, UK

### **New Data Tools**

Aim to keep track of developing trends and new tech that is helpful for the practicing Data Scientist. New might be a misnomer.

* [BigML](https://bigml.com/) - machine learning for the everyday user, also useful for EDA.
* [GraphLab](http://graphlab.com/) - graph-based, high performance, distributed computation framework. They just implemented deep learning onto their platform.
* [ModeAnalytics](https://modeanalytics.com/) - platform to share analysis/data science.
* [Apache Mahout](https://mahout.apache.org/) - Scalable machine learning library. Not in python.
* [Apache Hadoop](http://hadoop.apache.org/) - Open-source software for reliable, scalable, distributed computing. Not really new (10 years old at this point)
* [KNIME Analytics Platform | KNIME](https://www.knime.com/knime-analytics-platform) - No code way to do Data Science (makes easy to do DS)

**Product Metrics**

Understanding product, user behavior, and product metrics is helpful for data scientists in industry. Being able to help your product manager and team execute on strategies by understanding the problem, metrics and what they understand facilitates a more fruitful relationsh

* [Analytics for Product Managers](http://www.mindtheproduct.com/2013/02/everything-a-product-manager-needs-to-know-about-analytics/) - Everything a PM needs to know about analytics - or the minimum amount your PM should know about analytics as a Data Scientist.
* [Web analytics 101-](https://holistic-currant-5f0.notion.site/Web-Analytics-Quick-Guide-by-Sanjeev-Sahu-1538406f9c3d409f92bc0e6fcd3d1884)
* [Startups, you are doing data science wrong!](https://gigaom.com/2013/09/28/notice-to-startups-you-are-doing-data-science-wrong/) - High level explanation about how to use data science in a start-up company.
* [Product Psychology](http://www.productpsychology.com/category/user-behavior/) - Understanding user behavior.
* [Understanding Cohort Analysis](http://amarnath14.tumblr.com/post/69790103060/understanding-cohort-analysis) - Blog about cohort analysis, conversions, customer lifetime value, etc. Great starting point understanding product metrics.
* [Tech Product Management](http://techproductmanagement.com/) - More product focused than Data Science but can provide a good sense to view product management.
* [Mind The Product](http://www.mindtheproduct.com/) - Another solid PM blog.
* [Product market Fit](https://growthuniversity.io/measuring-product-market-fit/) - On PMF
* [E-commerce Analytics](https://holistic-currant-5f0.notion.site/Ecommerce-analytics-101-How-to-drive-more-online-sales-with-data-73a626b6ceff423dbce3a328987052a0)

### **Types of Data Scientists**

Not all Data Scientists are the same and it's critical for organizations to understand what it is they need, and how best to fill those roles and/or complement the skills of their team. Finding the organizational structure that enables the data scientists/data engineers within the organization and generates better results is also crucial. It should be given thorough consideration.

* [Kind's of Data Scientist](http://radar.oreilly.com/2013/06/theres-more-than-one-kind-of-data-scientist.html) - O'Reilly's classification of 4 different data scientists.
* [Data Science For Startups](http://tomtunguz.com/data-science-types/) - Which of the Five Types of DS does your startup need? Different classification from O'Reilly.
* [Building Data Science Teams](http://radar.oreilly.com/2011/09/building-data-science-teams.html) - posted from 2011 about how to build data science teams.
* [Data Science Team Building - The Power of Collaborative Analytics](http://www.experfy.com/blog/data-science-team-building-power-collaborative-analytics/) - Post post about different team org structures, difference between DS & BI.

### **Data Science Applications/Use Cases**

Data Science has so many different applications and use cases within industry - many are continuously discovered. These resources provide some potential ideas.

* [Kaggle Data Science Use Cases](https://www.kaggle.com/wiki/DataScienceUseCases) - Helpful to generate ideas for new uses in different industries
* [Data Science for each Industry](http://www.mastersindatascience.org/industry/) - Description of uses for different industries.
* [Big Data Analytics News - use Cases](http://bigdataanalyticsnews.com/big-data-use-cases/) - For Big Data but that's almost synonymous with Data Science.

### **Data Science Websites/Books**

More resources for community based information or hard copy books.

* [Data Science Handbook](https://medium.com/@pericarus/introducing-the-data-science-handbook-b2bfa216bf4b) - Not yet released but should be interesting providing stories from academia and industry about data science - go read the post for a better description!
* [CrossValidated](http://stats.stackexchange.com/) - A question and answer site for people interested in statistics, machine learning, data analysis, data mining, and data visualization.
* [StackOverflow](https://github.com/jonathan-bower/DataScienceResources/blob/master/stackoverflow.com) - Language-independent collaboratively edited question and answer site for programmers.
* [Kaggle](http://www.kaggle.com/) - Model building competition and great resources for training and data.
* [O'Reilly Media](http://shop.oreilly.com/category/get/data-science-kit.do) - A lot of content rich books available and tutorials on using the tools.
* [Quora](http://www.quora.com/) - Question and answer site - lots of data science content and career content.
* [Data Science @ StackExchange](http://datascience.stackexchange.com/) - Still in beta.

### **Data Science Blogs**

* [Data Stories @ Quroa](http://datastories.quora.com/) - William Chen's (DS@Quora) blog about data science.
* [FastML](http://fastml.com/)
* [FiveThirtyEight Blog](http://fivethirtyeight.com/) - Nate Silver's blog.
* [Data Science Hanbook](http://www.datasciencehandbook.me/) - Data Science Handbook Project (not quite a blog but it fits here).
* [Simply Statistics Blog](http://simplystatistics.org/)
* [All The Things Tech](http://nyghtowl.io/)
* [Musings in Data Science](http://deblivingdata.net/)
* [Zipfian Data Science Blog](http://www.zipfianacademy.com/blog/) - Zipfian Academy DS Blog.
* [Machine Learning Mastery](http://machinelearningmastery.com/)
* [DataTau](http://datatau.com/) - Hackernews for Data Science.
* [HackerNews](https://news.ycombinator.com/)
* [Quora](http://quora.com/) - Q&A site with lots of information about Data Science.
* [ThreeStoryBlog](http://blog.threestory.com/) - Design blog
* [7-Step Guide to Become a Machine Learning Engineer in 2021](https://www.dezyre.com/article/7-step-guide-to-become-a-machine-learning-engineer-in-2021/409) - Works well
* [Reducing the Need for Labeled Data in Generative Adversarial Networks](https://ai.googleblog.com/2019/03/reducing-need-for-labeled-data-in.html)
* [Jason's Google ML 101 deck](https://docs.google.com/presentation/d/1kSuQyW5DTnkVaZEjGYCkfOxvzCqGEFzWBy4e9Uedd9k/edit)
* [10 Free Must-Read Books for Machine Learning and Data Science](https://www.kdnuggets.com/2017/04/10-free-must-read-books-machine-learning-data-science.html?utm_content=buffer5a67a&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer)
* [Advice to aspiring data scientists: start a blog – Variance Explained](http://varianceexplained.org/r/start-blog/)
* [Brandon Roher Blog](https://brohrer.github.io/blog.html)
* [Chris Albon - Data Science, Machine Learning, and Artificial Intelligence](https://chrisalbon.com/#Python)
* [Data Science Stack Exchange](http://datascience.stackexchange.com/)
* [Data Skeptic](https://dataskeptic.com/)
* [DataTau](http://www.datatau.com/)
* [explained.ai - Deep explanations of machine learning and related topics](https://explained.ai/)
* [FlowingData](http://flowingdata.com/)
* [Here Are (Approximately) 3000 Free Data Sources You Can Use Right Now](https://www.forbes.com/sites/metabrown/2017/06/30/here-are-approximately-3000-free-sources-for-data-you-can-use-right-now/amp/?utm_content=bufferef401&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer)
* [If you want to learn Data Science, take a few of these statistics classes](https://medium.freecodecamp.com/if-you-want-to-learn-data-science-take-a-few-of-these-statistics-classes-9bbabab098b9)
* [Learn Data Science - Infographic (article) - DataCamp](https://www.datacamp.com/community/tutorials/learn-data-science-infographic)
* [LIGO Gravity Wave GW150914\_tutorial](https://losc.ligo.org/s/events/GW150914/GW150914_tutorial.html)
* [O.R. & Analytics Success Stories - INFORMS](https://www.informs.org/Impact/O.R.-Analytics-Success-Stories)
* [OpenAI Blog](https://blog.openai.com/)
* [Paul Ford: What Is Code? | Bloomberg](https://www.bloomberg.com/graphics/2015-paul-ford-what-is-code/)
* [Science Isn’t Broken | FiveThirtyEight](https://fivethirtyeight.com/features/science-isnt-broken/#part1)
* [Scientifically Sound](https://scientificallysound.org/)
* [AIspace](http://aispace.org/)
* [Top 28 Cheat Sheets for Machine Learning, Data Science, Probability, SQL & Big Data](https://www.analyticsvidhya.com/blog/2017/02/top-28-cheat-sheets-for-machine-learning-data-science-probability-sql-big-data/?utm_content=buffer9e308&utm_medium=social&utm_source=linkedin.com&utm_campaign=buffer)
* [GitHub Python Data Science Spotlight: AutoML, NLP, Visualization, ML Workflows](https://www.kdnuggets.com/2018/08/github-python-data-science-spotlight.html)

### **Start-Up Resources**

* [How to Start a Start-up](http://startupclass.samaltman.com/) - Series of lectures from successful entrepreneurs (i.e. Y comb, SV angels, etc.) on how to start a start up.
* [Network effects bible](https://www.dropbox.com/s/fm9fdmy0l2nnm2o/The%20Network%20Effects%20Bible%20%28eBook%29.pdf?dl=0&__hstc=37863228.57719403a6edb95763b1c7882661ea52.1632655410747.1632655410747.1632655410747.1&__hssc=37863228.1.1632655410749&__hsfp=3777543435) - All about network effects
* [Startups Are The Ultimate “Mental Game” — Part I](https://www.nfx.com/post/mental-models-part-one/#:~:text=Startups%20Are%20The%20Ultimate%20%E2%80%9CMental%20Game%E2%80%9D%20%E2%80%94%20Part,of%20leverage%20in%20a%20company%E2%80%99s%20success%20or%20failure.)
* [Part II](https://www.nfx.com/post/mental-models-part-two/)

##Open Source Data Science Resources While the name might sound redundant this section represents other sites or repos that have aggregated information covering similar topics. Tons of great content on these sites - definitely go check them out.

### Other Open Source Data Science Content

There are some really great resources linked within this section covering all of Data Science, the entire data pipeline, machine-learning, statistics, python, etc. Go check them out.

* [Open Data Science Masters](http://datasciencemasters.org/) - Clare Corthell's Open Source online blog/github with lots of resources available for data science.
* [A Practical Intro to Data Science](http://www.zipfianacademy.com/blog/post/46864003608/a-practical-intro-to-data-science) - Zipfian Academy's collection of excellent resources available.
* [LearnDataScience](https://github.com/nborwankar/LearnDataScience) - Nitin Borwankar's collection of IpythonNotebooks for Linear Regression, Logistic Regression, Random Forests, K-Means Clustering
* [FreeDataScienceBooks](https://github.com/chaconnewu/free-data-science-books/blob/master/free-data-science-books.md) - Yu Wu's free open sourced online data science books.
* [Gallery of Ipython Notebooks](https://github.com/ipython/ipython/wiki/A-gallery-of-interesting-IPython-Notebooks) - iPython's introduction to Python, Data Science, Economics, Comp Sci, Linguistics, and much more.
* [Data Science 45 Min Intros](https://github.com/DrSkippy/Data-Science-45min-Intros) - The team @ Gnip have a collection of repos to introduce data science topics in roughly 45 minutes per topic.
* [Awesome Data Science](https://github.com/okulbilisim/awesome-datascience) - Collection of bloggers, twitter accounts, facebook accounts, MOOC's, datasets, tools.
* [Awesome Big Data](https://github.com/onurakpolat/awesome-bigdata) - Onur Akpolat's curated list of awesome big data frameworks, resources and papers.
* [Mining the Social Web](https://github.com/ptwobrussell/Mining-the-Social-Web-2nd-Edition) - Matthew Russell's repo related to his book that focuses on working with the Twitter, Facebook, etc.
* [Harvard CS109 Github Repo](https://github.com/cs109/)
* [Pete Warden's Data Science Toolkit](https://github.com/petewarden/dstk) - Collection of open data sets and open-source tools for data science in ruby but has python.
* [Course Materials for Data Science Specialization](https://github.com/DataScienceSpecialization/courses) - Coursera course materials.
* [iPython Cookbook Materials](https://github.com/ipython-books/cookbook-code) - Excellent resources for high performance scientific computing and data science in python.

**No code tools**

* [Build an app from a Google Sheet in five minutes, for free](https://www.glideapps.com/)
* [The best way to build web apps without code | Bubble](http://bubble.io)

**Other Miscellaneous Links**

* [OpenAI blog](https://blog.openai.com/)
* [AI thinks like a corporation—and that’s worrying - Open Voices](https://www.economist.com/open-future/2018/11/26/ai-thinks-like-a-corporation-and-thats-worrying)
* [AITopics](https://aitopics.org/search)
* [Does the Brain Store Information in Discrete or Analog Form?](https://medium.com/mit-technology-review/does-the-brain-store-information-in-discrete-or-analog-form-f0e169361c99)
* [Explainable Artificial Intelligence (Part 1) — The Importance of Human Interpretable Machine…](https://towardsdatascience.com/human-interpretable-machine-learning-part-1-the-need-and-importance-of-model-interpretation-2ed758f5f476)
* [Is The Singularity Coming? – Arc Digital](https://arcdigital.media/is-the-singularity-coming-ef8580d4ce97)
* [Michael I. Jordan NYSE Machine Learning Presentation](https://www.youtube.com/watch?time_continue=2&v=17cp8PLKvOc)
* [Some scientists fear superintelligent machines could pose a threat to humanity | The Washington Post](https://www.washingtonpost.com/sf/national/2015/12/27/aianxiety/?noredirect=on&utm_term=.c3ac6321c831)
* [The Four Waves of A.I. | LinkedIn](https://www.linkedin.com/pulse/four-waves-ai-kai-fu-lee/)[When algorithms go wrong we need power to fight back, say researchers - The Verge](https://www.theverge.com/2018/12/8/18131745/ai-now-algorithmic-accountability-2018-report-facebook-microsoft-google)
* [Solved end-to-end Data Science projects](https://www.dezyre.com/projects/data-science-projects)
* [Dive into Deep Learning (An interactive deep learning book with code, math, and discussions)](https://d2l.ai/index.html)
* [Towards an anti-fascist AI (from opendemocracy.net)](https://www.opendemocracy.net/en/digitaliberties/towards-anti-fascist-ai/)
* [Becoming a Level 3.0 Data Scientist](https://www.kdnuggets.com/2019/05/becoming-a-level-3-data-scientist.html)
* [The Third-wave of Data Scientist](https://towardsdatascience.com/the-third-wave-data-scientist-1421df7433c9)
* [46 Most Intellectually Stimulating Sites That Will Spark Your Inner Genius in 10 Minutes a Day](https://medium.com/swlh/in-less-than-10-minutes-a-day-these-46-intellectually-stimulating-sites-will-spark-your-inner-d96ee6fc8387)
* [Artificial Intelligence Learns to Learn Entirely on Its Own | Quanta Magazine](https://www.quantamagazine.org/artificial-intelligence-learns-to-learn-entirely-on-its-own-20171018/?utm_content=buffer578b7&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer)
* [Edward Witten Ponders the Nature of Reality | Quanta Magazine](https://www.quantamagazine.org/edward-witten-ponders-the-nature-of-reality-20171128/)
* [Engineers Shouldn’t Write ETL: A Guide to Building a High Functioning Data Science Department | Stitch Fix Technology – Multithreaded](https://multithreaded.stitchfix.com/blog/2016/03/16/engineers-shouldnt-write-etl/)
* [Foundations Built for a General Theory of Neural Networks - Quanta Magazine](https://www.quantamagazine.org/foundations-built-for-a-general-theory-of-neural-networks-20190131)
* [General Thinking Tools: 9 Mental Models to Solve Difficult Problems](https://www.fs.blog/general-thinking-tools/)
* [How Social Media Endangers Knowledge | WIRED](https://www.wired.com/story/wikipedias-fate-shows-how-the-web-endangers-knowledge/)
* [In These Small Cities, AI Advances Could Be Costly - MIT Technology Review](https://www.technologyreview.com/s/609076/in-these-small-cities-ai-advances-could-be-costly/?utm_campaign=Owned+Social&utm_source=Facebook&utm_medium=Owned+Social)
* [Machine Learning’s ‘Amazing’ Ability to Predict Chaos | Quanta Magazine](https://www.quantamagazine.org/machine-learnings-amazing-ability-to-predict-chaos-20180418/)
* [New Brain Maps With Unmatched Detail May Change Neuroscience | WIRED](https://www.wired.com/story/new-brain-maps-with-unmatched-detail-may-change-neuroscience/)
* [Pedro Domingos on the Arms Race in Artificial Intelligence - SPIEGEL ONLINE](http://www.spiegel.de/international/world/pedro-domingos-on-the-arms-race-in-artificial-intelligence-a-1203132.html)
* [Quantum Leaps in Quantum Computing? - Scientific American](https://www.scientificamerican.com/article/quantum-leaps-in-quantum-computing/?utm_source=facebook&utm_medium=social&utm_campaign=sa-editorial-social&utm_content&utm_term=physics_sa-magazine_text_free)
* [The Fragile State of the Midwest’s Public Universities - The Atlantic](https://www.theatlantic.com/business/archive/2017/10/midwestern-public-research-universities-funding/542889/?utm_source=vxfb)
* [The Future of Human Work Is Imagination, Creativity, and Strategy](https://hbr.org/2018/01/the-future-of-human-work-is-imagination-creativity-and-strategy?utm_campaign=hbr&utm_source=linkedin&utm_medium=social)
* [The Quantum Thermodynamics Revolution | Quanta Magazine](https://www.quantamagazine.org/the-quantum-thermodynamics-revolution-20170502?utm_content=buffere2607&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer)
* [What Is Code? | Paul Ford| Bloomberg](https://www.bloomberg.com/graphics/2015-paul-ford-what-is-code/)
* [The Economics Of Artificial Intelligence - How Cheaper Predictions Will Change The World](https://www.forbes.com/sites/bernardmarr/2018/07/10/the-economics-of-artificial-intelligence-how-cheaper-predictions-will-change-the-world/#5b3b146f5a0d)
* [OpenAI’s Dota 2 defeat is still a win for artificial intelligence - The Verge](https://www.theverge.com/2018/8/28/17787610/openai-dota-2-bots-ai-lost-international-reinforcement-learning)
* [Machine Learning Confronts the Elephant in the Room | Quanta Magazine](https://www.quantamagazine.org/machine-learning-confronts-the-elephant-in-the-room-20180920/)

**Extras**

* [ISLR class videos](<https://www.r-bloggers.com/in-depth-introduction-to-machine-learning-in-15-hours-of-expert-videos/>
* [Machine Learning Zero-to-Hero: Everything you need in order to compete on Kaggle for the first…](https://towardsdatascience.com/machine-learning-zero-to-hero-everything-you-need-in-order-to-compete-on-kaggle-for-the-first-time-18644e701cf1)
* [GOOGLE - Rules of Machine Learning: | Machine Learning Rules | Google Developers](https://developers.google.com/machine-learning/rules-of-ml/)
* [PySpark ML tutorial example](https://nbviewer.jupyter.org/github/anindya-saha/Data-Science-with-Spark/blob/master/predict-us-census-income-classification/predict-us-census-income.ipynb)
* [Python Generators Tutorial](https://www.dataquest.io/blog/python-generators-tutorial/)
* [R Markdown: The Definitive Guide](https://bookdown.org/yihui/rmarkdown/)
* [Understanding the GitHub Flow · GitHub Guides](https://guides.github.com/introduction/flow/)
* [How to Prepare for a Machine Learning Interview - Semantic Bits](https://semanti.ca/blog/?how-to-prepare-for-a-machine-learning-interview)
* [Cheat Sheets for AI, Neural Networks, Machine Learning, Deep Learning & Big Data](https://becominghuman.ai/cheat-sheets-for-ai-neural-networks-machine-learning-deep-learning-big-data-678c51b4b463)
* [AI Knowledge Map: How To Classify AI Technologies](https://www.forbes.com/sites/cognitiveworld/2018/08/22/ai-knowledge-map-how-to-classify-ai-technologies/#5878d667773f)