

KDnuggets

[Subscribe to KDnuggets News](#)



- [Blog/News](#)
- [Opinions](#)
- [Tutorials](#)
- [Top stories](#)
- [Companies](#)
- [Courses](#)
- [Datasets](#)
- [Education](#)
- [Events \(online\)](#)
- [Jobs](#)
- [Software](#)
- [Webinars](#)



[JMP Discovery Summit Online: Exploring Data, Inspiring Innovation. Oct 12-16. Register Now](#)

Topics: [Coronavirus](#) | [AI](#) | [Data Science](#) | [Deep Learning](#) | [Machine Learning](#) | [Python](#) | [R](#) | [Statistics](#)

[KDnuggets Home](#) » [News](#) » [2020](#) » [Oct](#) » [Opinions](#) » Data Science Minimum: 10 Essential Skills You Need to Know to Start Doing Data Science
(20:n38)

Data Science Minimum: 10 Essential Skills You Need to Know to Start Doing Data Science

[<= Previous post](#)

[Next post =>](#)

Like 63

Share 63

Tweet

Share

Share

329

Tags: [Algorithms](#), [Communication](#), [Data Preprocessing](#), [Data Science](#), [Data Science Skills](#), [Data Visualization](#), [Ethics](#), [Mathematics](#), [Python](#), [R](#)

Data science is ever-evolving, so mastering its foundational technical and soft skills will help you be successful in a career as a Data Scientist, as well as pursue advance concepts, such as deep learning and artificial intelligence.



Open for Innovation

Successful Data Science Teams with KNIME

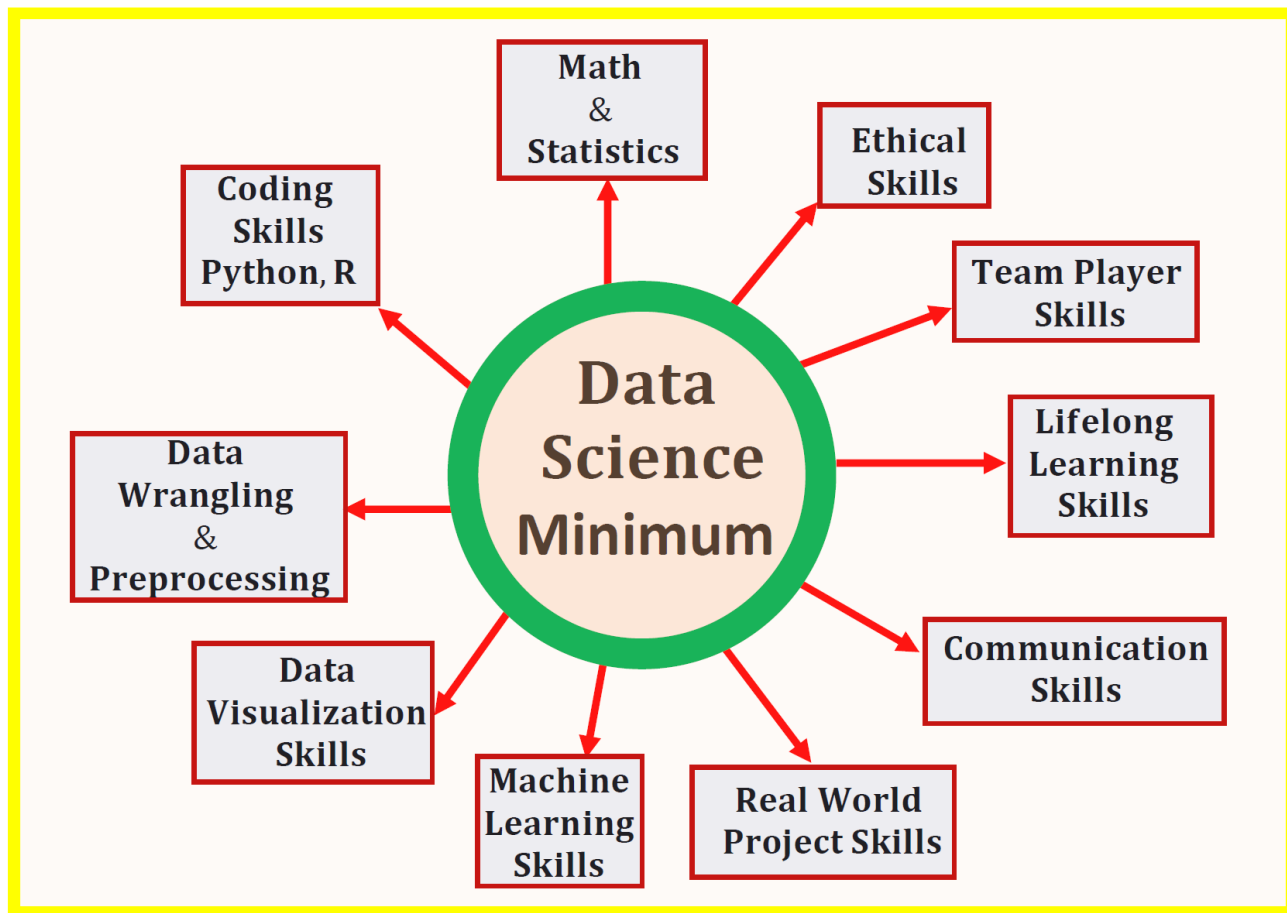
Live webinar on October 20

[Register now](#)

[Successful Data Science Teams with KNIME Live Webinar on Oct 20 Register now](#)

By [Benjamin Obi Tayo, Ph.D.](#), DataScienceHub.

[comments](#)



Data Science is such a broad field that includes several subdivisions like data preparation and exploration, data representation and transformation, data visualization and presentation, predictive analytics, and machine learning, etc. For beginners, it's only natural to raise the following question: **What skills do I need to become a data scientist?**

This article will discuss 10 essential skills that are necessary for practicing data scientists. These skills could be grouped into 2 categories, namely, **technological skills** (Math & Statistics, Coding Skills, Data Wrangling & Preprocessing Skills, Data Visualization Skills, Machine Learning Skills, and Real World Project Skills) and **soft skills** (Communication Skills, Lifelong Learning Skills, Team Player Skills, and Ethical Skills).

Data science is a field that is ever-evolving, however mastering the foundations of data science will provide you with the necessary background that you need to pursue advanced concepts such as deep learning, artificial intelligence, etc. This article will discuss 10 essential skills for practicing data scientists.

1. Mathematics and Statistics Skills

(i) Statistics and Probability

Statistics and Probability is used for visualization of features, data preprocessing, feature transformation, data imputation, dimensionality reduction, feature engineering, model evaluation, etc. Here are the topics you need to be familiar with:

- a) Mean
- b) Median
- c) Mode
- d) Standard deviation/variance
- e) Correlation coefficient and the covariance matrix
- f) Probability distributions (Binomial, Poisson, Normal)
- g) p-value
- h) MSE (mean square error)
- i) R² Score
- j) Baye's Theorem (Precision, Recall, Positive Predictive Value, Negative Predictive Value, Confusion Matrix, ROC Curve)
- k) A/B Testing
- l) Monte Carlo Simulation

(ii) Multivariable Calculus

Most machine learning models are built with a data set having several features or predictors. Hence, familiarity with multivariable calculus is extremely important for building a machine learning model. Here are the topics you need to be familiar with:

- a) Functions of several variables
- b) Derivatives and gradients
- c) Step function, Sigmoid function, Logit function, ReLU (Rectified Linear Unit) function
- d) Cost function
- e) Plotting of functions
- f) Minimum and Maximum values of a function

(iii) Linear Algebra

Linear algebra is the most important math skill in machine learning. A data set is represented as a matrix. Linear algebra is used in data preprocessing, data transformation, and model evaluation. Here are the topics you need to be familiar with:

- a) Vectors
- b) Matrices
- c) Transpose of a matrix
- d) The inverse of a matrix
- e) The determinant of a matrix
- f) Dot product
- g) Eigenvalues
- h) Eigenvectors

(iv) Optimization Methods

Most machine learning algorithms perform predictive modeling by minimizing an objective function, thereby learning the weights that must be applied to the testing data in order to obtain the predicted labels. Here are the topics you need to be familiar with:

- a) Cost function/Objective function
- b) Likelihood function
- c) Error function
- d) Gradient Descent Algorithm and its variants (e.g., Stochastic Gradient Descent Algorithm)

Find out more about the gradient descent algorithm here: [Machine Learning: How the Gradient Descent Algorithm Works](#).

2. Essential Programming Skills

Programming skills are essential in data science. Since Python and R are considered the two most popular programming languages in data science, essential knowledge in both languages are crucial. Some organizations may only require skills in either R or Python, not both.

(i) Skills in Python

Be familiar with basic programming skills in python. Here are the most important packages that you should master how to use:

- a) Numpy
- b) Pandas
- c) Matplotlib
- d) Seaborn
- e) Scikit-learn
- f) PyTorch

(ii) Skills in R

- a) Tidyverse
- b) Dplyr
- c) Ggplot2
- d) Caret
- e) Stringr

(iii) Skills in Other Programming Languages

Skills in the following programming languages may be required by some organizations or industries:

- a) Excel
- b) Tableau
- c) Hadoop
- d) SQL
- e) Spark

3. Data Wrangling and Preprocessing Skills

Data is key for any analysis in data science, be it inferential analysis, predictive analysis, or prescriptive analysis. The predictive power of a model depends on the quality of the data that was used in building the model. Data comes in different forms, such as text, table, image, voice, or video. Most often, data that is used for analysis has to be mined, processed, and transformed to render it to a form suitable for further analysis.

i) **Data Wrangling:** The process of data wrangling is a critical step for any data scientist. Very rarely is data easily accessible in a data science project for analysis. It's more likely for the data to be in a file, a database, or extracted from documents such as web pages, tweets, or PDFs. Knowing how to

wrangle and clean data will enable you to derive critical insights from your data that would otherwise be hidden.

ii) **Data Preprocessing:** Knowledge about data preprocessing is very important and include topics such as:

- a) Dealing with missing data
- b) Data imputation
- c) Handling categorical data
- d) Encoding class labels for classification problems
- e) Techniques of feature transformation and dimensionality reduction, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

4. Data Visualization Skills

Understand the essential components of good data visualization.

- a) **Data Component:** An important first step in deciding how to visualize data is to know what type of data it is, e.g., categorical data, discrete data, continuous data, time series data, etc.
- b) **Geometric Component:** Here is where you decide what kind of visualization is suitable for your data, e.g., scatter plot, line graphs, bar plots, histograms, qqplots, smooth densities, boxplots, pair plots, heatmaps, etc.
- c) **Mapping Component:** Here you need to decide what variable to use as your x-variable and what to use as your y-variable. This is important, especially when your dataset is multi-dimensional with several features.
- d) **Scale Component:** Here you decide what kind of scales to use, e.g., linear scale, log scale, etc.
- e) **Labels Component:** This includes things like axes labels, titles, legends, font size to use, etc.
- f) **Ethical Component:** Here, you want to make sure your visualization tells the true story. You need to be aware of your actions when cleaning, summarizing, manipulating, and producing a data visualization and ensure you aren't using your visualization to mislead or manipulate your audience.

5. Basic Machine Learning Skills

Machine Learning is a very important branch of data science. It is important to understand the machine learning framework: Problem Framing, Data Analysis, Model Building, Testing & Evaluation, and Model Application. Find out more about the machine learning framework from here: [The Machine Learning Process](https://www.kdnuggets.com/2020/10/data-science-minimum-10-essential-skills.html).

The following are important machine learning algorithms to be familiar with.

i) Supervised Learning (Continuous Variable Prediction)

- a) Basic regression
- b) Multiregression analysis
- c) Regularized regression

ii) Supervised Learning (Discrete Variable Prediction)

- a) Logistic Regression Classifier
- b) Support Vector Machine Classifier
- c) K-nearest neighbor (KNN) Classifier
- d) Decision Tree Classifier
- e) Random Forest Classifier

iii) Unsupervised Learning

- a) KMeans clustering algorithm

[<= Previous post](#)[Next post =>](#)

Top Stories Past 30 Days

Most Popular

1. [If I had to start learning Data Science again, how would I do it?](#)
2. [Free From MIT: Intro to Computer Science and Programming in Python](#)
3. [Automating Every Aspect of Your Python Project](#)
4. [Top Online Masters in Analytics, Business Analytics, Data Science Updated](#)
5. [Introduction to Time Series Analysis in Python](#)
6. [Autograd: The Best Machine Learning Library You're Not Using?](#)
7. [Machine Learning from Scratch: Free Online Textbook](#)

Most Shared

1. [Modern Data Science Skills: 8 Categories, Core Skills, and Hot Skills](#)
2. [Top Online Masters in Analytics, Business Analytics, Data Science – Updated](#)
3. [Machine Learning from Scratch: Free Online Textbook](#)
4. [How to Evaluate the Performance of Your Machine Learning Model](#)
5. [Deep Learning's Most Important Ideas](#)
6. [Statistics with Julia: The Free eBook](#)
7. [Free From MIT: Intro to Computer Science and Programming in Python](#)

Latest News

- [Data Protection Techniques Needed to Guarantee Privacy](#)
- [5 Concepts Every Data Scientist Should Know](#)
- [Comparing the Top Business Intelligence Tools: Power BI...](#)
- [10 Days With "Deep Learning for Coders"](#)
- [Data Science Minimum: 10 Essential Skills You Need to K...](#)
- [Understanding Transformers, the Data Science Way](#)

Top Stories Last Week

Most Popular

1. [Introduction to Time Series Analysis in Python](#)



2. [Machine Learning from Scratch: Free Online Textbook](#)
3. [How I Consistently Improve My Machine Learning Models From 80% to Over 90% Accuracy](#)
4. [New Poll: What Python IDE / Editor you used the most in 2020?](#)
5. [Automating Every Aspect of Your Python Project](#)
6. [If I had to start learning Data Science again, how would I do it?](#)
7. [Making Python Programs Blazingly Fast](#)

Most Shared

1. [Machine Learning from Scratch: Free Online Textbook](#)
2. [Introduction to Time Series Analysis in Python](#)
3. [How I Consistently Improve My Machine Learning Models From 80% to Over 90% Accuracy](#)
4. [I'm a Data Scientist, Not Just The Tiny Hands that Crunch your Data](#)
5. [New Poll: What Python IDE / Editor you used the most in 2020?](#)
6. [The Most Complete Guide to PyTorch for Data Scientists](#)
7. [What an Argentine Writer and a Hungarian Mathematician Can Teach Us About Machine Learning Overfitting](#)

More Recent Stories

- [Understanding Transformers, the Data Science Way](#)
- [Top tweets, Sep 23-29: An Introduction to #AI – updat...](#)
- [AI in Healthcare: A review of innovative startups](#)
- [Machine Learning Model Deployment](#)
- [The Best Free Data Science eBooks: 2020 Update](#)
- [KDNuggets 20:n37, Sep 30: Introduction to Time Series Analy...](#)
- [Are Data Analytics and Data Science Two Separate Fields?](#)
- [Missing Value Imputation – A Review](#)
- [International alternatives to Kaggle for Data Science / Machin...](#)
- [How AI is Driving Innovation in Astronomy](#)
- [Top Stories, Sep 21-27: Introduction to Time Series Analysis i...](#)
- [Looking Inside The Blackbox: How To Trick A Neural Network](#)
- [Geographical Plots with Python](#)
- [The Online Courses You Must Take to be a Better Data Scientist](#)
- [Making Python Programs Blazingly Fast](#)
- [Create and Deploy your First Flask App using Python and Heroku](#)
- [Causal Inference: The Free eBook](#)
- [KDD 2020 Celebrates Recipients of the SIGKDD Best Paper Aw...](#)
- [Introduction to Time Series Analysis in Python](#)
- [The Most Complete Guide to PyTorch for Data Scientists](#)

[KDNuggets Home](#) » [News](#) » [2020](#) » [Oct](#) » [Opinions](#) » Data Science Minimum: 10 Essential Skills You Need to Know to Start Doing Data Science ([20:n38](#))



© 2020 KDNuggets. | [About KDNuggets](#) | [Contact](#) | [Privacy policy](#) | [Terms of Service](#)

6. Skills from Real World Capstone Data Science Projects

Skills acquired from course work alone will not make you a data scientist. A qualified data scientist must be able to demonstrate evidence of successful completion of a real-world data science project that includes every stage in data science and machine learning process such as problem framing, data acquisition and analysis, model building, model testing, model evaluation, and deploying models. Real-world data science projects could be found in the following:

- a) Kaggle Projects
- b) Internships
- c) From Interviews

7. Communication Skills

Data scientists need to be able to communicate their ideas with other members of the team or with business administrators in their organizations. Good communication skills would play a key role here to be able to convey and present very technical information to people with little or no understanding of technical concepts in data science. Good communication skills will help foster an atmosphere of unity and togetherness with other team members such as data analysts, data engineers, field engineers, etc.

8. Be a Lifelong Learner

Data science is a field that is ever-evolving, so be prepared to embrace and learn new technologies. One way to keep in touch with developments in the field is to network with other data scientists. Some platforms that promote networking are LinkedIn, GitHub, and Medium ([Towards Data Science](#) and [Towards AI](#) publications). The platforms are very useful for up-to-date information about recent developments in the field.

9. Team Player Skills

As a data scientist, you will be working in a team of data analysts, engineers, administrators, so you need good communication skills. You need to be a good listener, too, especially during early project development phases where you need to rely on engineers or other personnel to be able to design and frame a good data science project. Being a good team player will help you to thrive in a business environment and maintain good relationships with other members of your team as well as administrators or directors of your organization.

10. Ethical Skills in Data Science

Understand the implication of your project. Be truthful to yourself. Avoid manipulating data or using a method that will intentionally produce bias in results. Be ethical in all phases, from data collection and analysis to model building, analysis, testing, and application. Avoid fabricating results for the purpose of misleading or manipulating your audience. Be ethical in the way you interpret the findings from your data science project.

In summary, we've discussed 10 essential skills needed for practicing data scientists. Data science is a field that is ever-evolving, however mastering the foundations of data science will provide you with the necessary background that you need to pursue advance concepts such as deep learning, artificial intelligence, etc.

[Original](#). Reposted with permission.

Related:

- [Modern Data Science Skills: 8 Categories, Core Skills, and Hot Skills](#)
- [These Data Science Skills will be your Superpower](#)
- [Top 5 must-have Data Science skills for 2020](#)

What do you think?

Important Update

When you log in with Disqus, we process personal data to facilitate your authentication and posting of comments. We also store the comments you post and those comments are immediately viewable and searchable by anyone around the world.

Please access our Privacy Policy to learn what personal data Disqus collects and your choices about how it is used. All users of our service are also subject to our Terms of Service.

Proceed