

Get a data science job or your money back [Click Here](#)

OCT 26, 2016

109 Commonly Asked Data Science Interview Questions

👤 Michael Rundell 💬 0 📁 DATA SCIENCE
([HTTPS://WWW.SPRINGBOARD.COM/BLOG/CATEGORY/DATA-SCIENCE/](https://www.springboard.com/blog/category/data-science/))



A Curated List of Data Science Interview Questions

Preparing for an interview is not easy – naturally there is a large amount of uncertainty regarding the data science interview questions you will be asked. No matter how much work experience or technical skill you have, an interviewer can throw you off with a set of questions that you didn't expect. For a data science interview, an interviewer will ask questions spanning a wide range of topics, requiring strong technical knowledge and communication skills from the part of the interviewee. Your statistics, programming, and data modeling skills will be put to the test through a variety of questions and question styles – intentionally designed to keep you on

Get a data science job or your money back [Click Here](#)

(<https://www.springboard.com/workshops/data-science-career-track/>).

This guide contains all of the data science interview questions an interviewee should expect when interviewing for a position as a data scientist. At [Springboard](http://springboard.com) (<http://springboard.com>), we teach data science through our [mentored data science workshops](https://www.springboard.com/workshops/data-science-intensive) (<https://www.springboard.com/workshops/data-science-intensive>). They're a great way to learn data science and get expert guidance on [how to get a data science job](https://www.springboard.com/resources/guide-to-data-science-jobs) (<https://www.springboard.com/resources/guide-to-data-science-jobs>). We did our due diligence to comb through the internet to find real questions asked to data science interview candidates. We had built a [data science interview guide](https://www.springboard.com/resources/data-scientist-interview-guide) (<https://www.springboard.com/resources/data-scientist-interview-guide>), yet we still felt we had more to explore.

We set off to curate, create and edit different data science interview questions and provided answers for some. From this list of **data science interview questions**, an interviewee should be able to prepare for the tough questions, learn what answers will positively resonate with an employer, and develop the confidence to ace the interview. We've broken the data science interview questions into six different categories: statistics, programming, modeling, behavior, culture, and problem-solving.

Table of Contents

1. **Statistics**
2. **Programming**
 1. *General*
 2. *Big Data*
 3. *Python*
 4. *R*
 5. *SQL*
3. **Modeling**
4. **Behavioral**
5. **Culture Fit**

Get a data science job or your money back [Click Here](#)

1. Statistics

Statistical computing is the process through which data scientists take raw data and create predictions and models backed by the data. Without an advanced knowledge of statistics it is difficult to succeed as a data scientist – accordingly it is likely a good interviewer will try to probe your understanding of the subject matter with statistics-oriented data science interview questions. Be prepared to answer some fundamental statistics questions as part of your data science interview.

Here are examples of rudimentary statistics questions we've found:

1. What is the Central Limit Theorem and why is it important?

Answer (<https://spin.atomicobject.com/2015/02/12/central-limit-theorem-intro/>)

2. What is sampling? How many sampling methods do you know?
3. What is the difference between Type I vs Type II error?
4. What is linear regression? What do the terms P-value, coefficient, R-Squared value mean?
What is the significance of each of these components?

Answer (<http://www.statisticssolutions.com/assumptions-of-linear-regression/>), Answer (<http://blog.minitab.com/blog/adventures-in-statistics/how-to-interpret-regression-analysis-results-p-values-and-coefficients>)

5. What are the assumptions required for linear regression?

There are four major assumptions: 1. There is a linear relationship between the dependent variables and the regressors, meaning the model you are creating actually fits the data, 2. The errors or residuals of the data are normally distributed and independent from each other, 3. There is minimal multicollinearity between explanatory variables, and 4. Homoscedasticity. This means the variance around the regression line is the same for all values of the predictor variable.

6. What is a statistical interaction?

Answer (<http://icbseverywhere.com/blog/mini-lessons-tutorials-and-support-pages/statistical-interactions/>)

7. What is selection bias?

Get a data science job or your money back [Click Here](#)

Examples of similar data science interview questions found from Glassdoor:



Data Scientist at State Farm was asked...

Apr 19, 2016

What is a p-value? Would your interpretation of p-value change, if you had a different (much bigger, 3 mil records for ex.) data set?

[1 Answer](#) ▾



Data Scientist at Facebook was asked...

Sep 12, 2013

You're about to get on a plane to Seattle. You want to know if you should bring an umbrella. You call 3 random friends of yours who live there and ask each independently if it's raining. Each of your friends has a 2/3 chance of telling you the truth and a 1/3 chance of messing with you by lying. All 3 friends tell you that "Yes" it is raining. What is the probability that it's actually raining in Seattle?

[27 Answers](#) ^



Data Scientist at Microsoft was asked...

Mar 25, 2015

How to compute an inverse matrix faster by playing around with some computational tricks?

[Be the first to answer this question](#)



Data Scientist at Spotify was asked...

Apr 1, 2015

given n samples from a uniform distribution $[0, d]$, how to estimate d ?

[4 Answers](#) ▾

2. Programming

To test your programming skills, employers will ask two things during their data science interview questions: they'll ask how you would solve programming problems in theory without writing out the code, and then they will also offer whiteboarding exercises for you to code on the spot. For the latter types of questions we will cover a few examples below, but if you're looking for in-depth practice solving coding challenges, visit Interview Cake (<https://www.interviewcake.com/all-questions/python>). They have an in-browser module for typing code, and they can walk you through tricky problems – all absolutely free.

2.1 General

Get a data science job or your money back [Click Here](#)

3. Tell me about an original algorithm you've created.
4. Describe a data science project in which you worked with a substantial programming component. What did you learn from that experience?
5. Do you contribute to any open source projects?
6. How would you clean a dataset in (insert language here)?
7. Tell me about the coding you did during your last project?

2.2 Big Data

1. What are the two main components of the Hadoop Framework?
2. Explain how MapReduce works as simply as possible.
Answer (<http://stackoverflow.com/questions/28982/simple-explanation-of-mapreduce>)
3. How would you sort a large list of numbers?
4. Here is a big dataset. What is your plan for dealing with outliers? How about missing values? How about transformations?

2.3 Python

1. What modules/libraries are you most familiar with? What do you like or dislike about them?
2. What are the supported data types in Python?
3. What is the difference between a tuple and a list in Python?

For additional Python questions that focus on looking at specific snippets of code, check out this useful resource created by Toptal (<https://www.toptal.com/python/interview-questions>).

2.4 R

1. What are the different types of sorting algorithms available in R language?
There are insertion, bubble, and selection sorting algorithms.
2. What are the different data objects in R?
3. What packages are you most familiar with? What do you like or dislike about them?
4. How do you access the element in the 2nd column and 4th row of a matrix named M?
5. What is the command used to store R objects in a file?

Get a data science job or your money back [Click Here](#)

8. Write a function in R language to replace the missing value in a vector with the mean of that vector.

2.5 SQL

Often, SQL questions are case-based, meaning that an employer will task you with solving an SQL problem in order to test your skills from a practical standpoint. For example, you could be given a table and be asked to extract relevant data, filter and order the data as you see fit, and report your findings. If you do not feel ready to do this in an interview setting, Mode Analytics (<https://community.modeanalytics.com/sql/tutorial/introduction-to-sql/>) has a delightful introduction to using SQL that will teach you these commands through an interactive SQL environment.

1. What is the purpose of the group functions in SQL? Give some examples of group functions.
2. Group functions are necessary to get summary statistics of a dataset. COUNT, MAX, MIN, AVG, SUM, and DISTINCT are all group functions
3. Tell me the difference between an inner join, left join/right join, and union.

Answer (<https://www.quora.com/SQL-What-is-the-difference-between-inner-join-left-join-right-join-and-full-join>)

4. What does UNION do? What is the difference between UNION and UNION ALL?
5. What is the difference between SQL and MySQL or SQL Server?
6. If a table contains duplicate rows, does a query result display the duplicate values by default? How can you eliminate duplicate rows from a query result?

For additional SQL questions that focus on looking at specific snippets of code, check out this useful resource created by Toptal (<https://www.toptal.com/sql/interview-questions>).

Examples of similar data science interview questions found from Glassdoor:



Data Scientist Intern at LinkedIn was asked...

Feb 25, 2012

Find the second largest element in a Binary Search Tree

13 Answers ▾

Get a data science job or your money back [Click Here](#)

problem could be solved by analytical queries

[Be the first to answer this question](#)

3. Modeling

Data modeling is where a data scientist provides value for a company. Turning data into predictive and actionable information is difficult, talking about it to a potential employer even more so. Practice describing your past experiences building models – what were the techniques used, challenges overcome, and successes achieved in the process? The group of questions below are designed to uncover that information, as well as your formal education of different modeling techniques. If you can't describe the theory and assumptions associated with a model you've used, it won't leave a good impression.

Take a look at the questions below to practice. Not all of the questions will be relevant to your interview – you're not expected to be a master of all techniques. The best use of these questions is to re-familiarize yourself with the modeling techniques you've learned in the past.

1. Tell me about how you designed the model you created for a past employer or client.
2. What are your favorite data visualization techniques?
3. How would you effectively represent data with 5 dimensions?
4. How is kNN different from k-means clustering?

kNN, or k-nearest neighbors is a classification algorithm, where the k is an integer describing the the number of neighboring data points that influence the classification of a given observation. K-means is a clustering algorithm, where the k is an integer describing the number of clusters to be created from the given data. Both accomplish different tasks.

5. How would you create a logistic regression model?
6. Have you used a time series model? Do you understand cross-correlations with time lags?
7. Explain the 80/20 rule, and tell me about its importance in model validation.
8. Explain what precision and recall are. How do they relate to the ROC curve?

Answer (<http://www.kdnuggets.com/faq/precision-recall.html>). Recall describes what percentage of true positives are described as positive by the model. Precision describes what percent of positive predictions were correct. The ROC curve shows the relationship

Get a data science job or your money back [Click Here](#)

measures used to identify how useful a given classification model is.

9. Explain the difference between L1 and L2 regularization methods.
10. What is root cause analysis?
11. What are hash table collisions?
12. What is an exact test?
13. In your opinion, which is more important when designing a machine learning model: Model performance? Or model accuracy?

One approach to this question (https://en.wikipedia.org/wiki/Accuracy_paradox)

14. What is one way that you would handle an imbalanced dataset that's being used for prediction? (i.e. vastly more negative classes than positive classes.)
15. How would you validate a model you created to generate a predictive model of a quantitative outcome variable using multiple regression?
16. I have two models of comparable accuracy and computational performance. Which one should I choose for production and why?
17. How do you deal with sparsity?
18. Is it better to spend 5 days developing a 90% accurate solution, or 10 days for 100% accuracy?
19. What are some situations where a general linear model fails?
20. Do you think 50 small decision trees are better than a large one? Why?
21. When modifying an algorithm, how do you know that your changes are an improvement over not doing anything?
22. Is it better to have too many false positives, or too many false negatives?

Examples of similar data science interview questions found from Glassdoor:



Data Scientist at Apple was asked...

Nov 6, 2013

How do you take millions of users with 100's of transactions each, amongst 10k's of products and group the users together in a meaningful segments?

[2 Answers](#)

Get a data science job or your money back [Click Here](#)

what machine learning algorithm did I use, and why?

[Be the first to answer this question](#)



Data Scientist at Instacart was asked...

Jun 21, 2016

How would you tune a random forest?

[1 Answer](#) ✓



Data Scientist at Netflix was asked...

Jul 19, 2012

How do you know if one algorithm is better than other?

[1 Answer](#) ✓

4. Past Behavior

Employers love behavioral questions. They reveal information about the work experience of the interviewee as well as information about the demeanor of any potential team member. From these questions, an interviewer wants to see how a candidate has reacted to situations in the past, how well they can articulate what their role was, and what they learned from their experience.

There are several categories of behavioral questions you'll be asked:

1. Teamwork
2. Leadership
3. Conflict Management
4. Problem-solving
5. Failure

Before the interview, write down examples of work experience related to these topics to refresh your memory – you will need to recall specific examples to answer the questions. When asked about a prior experience, make sure you tell a story as well. Being able to concisely and logically craft a story to detail your experiences is important. For example – "I was asked X, I did A, B, and C, and decided that the answer is Y."

LEARN DATA SCIENCE WITH SPRINGBOARD

Get a data science job or your money back

Click Here

833
Shares

245

6

2



Break into Data Science

Get our free, comprehensive 70-page guide to getting your first data science job.

GET IT NOW

and the hiring company come from. Take a look at these examples and think about what your best answer would be, but keep in mind – it's important to be honest with these questions.

There's no reason to not be yourself. There are no right answers to these questions – but the best answers are communicated with confidence and a smile.

Michael Rundell (<https://www.springboard.com/blog/author/mrundell>)



Get a data science job or your money back [Click Here](#)



(<https://www.springboard.com/blog/author/mrundell/>)

www.springboard.com/blog/top-data-scientists-on-twitter/) on top data science

ers for some more insight on some of the top data scientists in the world.

2. What do you think makes a good data scientist?
3. How did you become interested in data science?
4. Give a few examples of "best practices" in data science.
5. What/when is the latest data science book / article you read? What/when is the latest data mining conference / webinar / class / workshop / training you attended?
If you haven't read a good data science book recently, Springboard compiled a list of the best data science books to read (<https://www.springboard.com/blog/eleven-best-data-science-books/>)!
6. What's a project you would want to work on at our company?
7. What unique skills do you think you'd bring to the team?
8. What data would you love to acquire if there were no limitations?
9. Have you ever thought about creating a startup? Around which idea / concept?
10. What can your hobbies tell me that your resume can't?
11. What are your top 5 predictions for the next 20 years?
12. What did you do today? Or what did you do this week / last week?
13. If you won a million dollars in the lottery, what would you do with the money?
14. What is one thing you believe that most people do not?
15. What personality traits do you butt heads with?
16. What are you passionate about?

Examples of similar data science interview questions found from Glassdoor:



Data Scientist at Capital One was asked...

Jul 1, 2015

What is your career plan?

[Be the first to answer this question](#)



Senior Data Scientist at Apple was asked...

Aug 25, 2015

If you could have one superpower, what would it be?

[Be the first to answer this question](#)

Get a data science job or your money back [Click Here](#)

Interviewers will, at some point during the interview process, want to test your problem-solving ability through data science interview questions. Often these tests will be presented as an open-ended question – How would you do “X”? In general, that ‘X’ will be a task or problem specific to the company you are applying with. For example, an interviewer at Yelp may ask a candidate how they would create [a system to detect fake Yelp reviews](#)

(<https://www.springboard.com/blog/eat-rate-love-an-exploration-of-r-yelp-and-the-search-for-good-indian-food/>). Some quick tips: Don’t be afraid to ask questions. Employers want to test your critical thinking skills – and asking questions that clarify points of uncertainty are a great way to show that you know how to ask the right questions (a trait that any data scientist should have). Also, if the problem offers an opportunity to show off your white-board coding skills or to create schematic diagrams – use that to your advantage. It shows technical skill, and helps to communicate your thought process through a different medium of communication. Always communicate your thought process — process is often more important than the results themselves for the interviewer.

1. How would you come up with a solution to identify plagiarism?
2. How many “useful” votes will a Yelp review receive?
3. How do you detect individual paid accounts shared by multiple users?
4. You are about to send one million emails. How do you optimize delivery? How do you optimize response?
5. You have a dataset containing 100K rows and 100 columns, with one of those columns being our dependent variable for a problem we’d like to solve. How can we quickly identify which columns will be helpful in predicting the dependent variable. Identify two techniques and explain them to me as though I were 5 years old.
6. How would you detect bogus reviews, or bogus Facebook accounts used for bad purposes?

This is an opportunity to showcase your knowledge of machine learning algorithms; specifically, sentiment analysis and text analysis algorithms. Showcase your knowledge of fraudulent behavior – what are the abnormal behaviors (<https://www.cs.uic.edu/~liub/FBS/fake-reviews.html>) that can typically be seen from fraudulent accounts?

Get a data science job or your money back [Click Here](#)

similar these two keywords are? How would you create this 10 million data points table in the first place?

8. How would you optimize a web crawler to run much faster, extract better information, and better summarize data to produce cleaner databases?

Examples of similar data science interview questions found from Glassdoor:



Data Scientist at Symantec was asked...

Sep 2, 2015

Suppose you have a coffee store, how do you do to increase the number of customers?

[1 Answer](#) ✓



Data Scientist at Massachusetts General Hospital was asked...

May 19, 2015

Given an existing set of purchases, how do you predict the next item to purchase of a new basket?

[1 Answer](#) ✓



Senior Data Scientist at Natera was asked...

May 18, 2015

Imagine you have N pieces of rope in a bucket. You reach in and grab one end-piece, then reach in and grab another end-piece, and tie those two together. What is the expected value of the number of loops in the bucket?

[2 Answers](#) ✓

Conclusion

There is no exact formula for preparing for data science interview questions, but hopefully by reviewing these common interview questions you will be able to walk into your interviews well-practiced and confident. If you have any suggestions for questions, feel free to comment below! Good luck.

Related reading:

Our guide to [data science interviews](https://www.springboard.com/resources/data-scientist-interview-guide/) (<https://www.springboard.com/resources/data-scientist-interview-guide/>).

Get a data science job or your money back [Click Here](#)

selection of data science interview questions.

Sources

User-submitted interview questions from data science interview questions across the United States

Glassdoor – Data Scientist Interview Questions (https://www.glassdoor.com/Interview/data-scientist-interview-questions-SRCH_KO0,14.htm)

Lists of general data science interview questions

KDnuggets (<http://www.kdnuggets.com/2016/02/21-data-science-interview-questions-answers.html>)

DeZyre (<https://www.dezyre.com/article/100-data-science-interview-questions-and-answers-general-for-2016/184>)

Udacity (<http://blog.udacity.com/2015/04/data-science-interview-questions.html>)

Data Science Central – 66 Interview Questions for Data Scientists
(<http://www.datasciencecentral.com/profiles/blogs/66-job-interview-questions-for-data-scientists>)

AnalyticsVidhya – 40 Interview Questions asked at Startups in Machine Learning/Data Science
(<https://www.analyticsvidhya.com/blog/2016/09/40-interview-questions-asked-at-startups-in-machine-learning-data-science/>)

Sources for all programming and coding related data science questions

Workable – Data Scientist Coding Interview Questions (<https://resources.workable.com/data-scientist-coding-interview-questions>)

Get a data science job or your money back [Click Here](#)

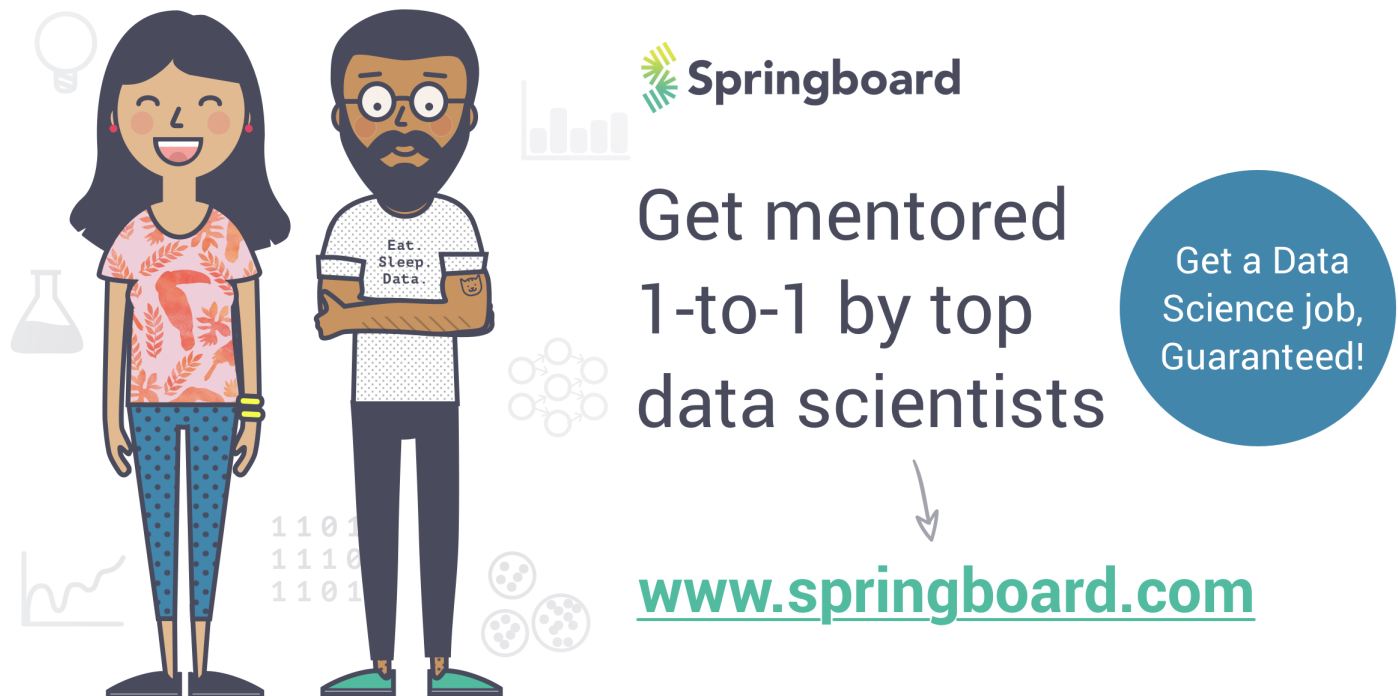
MaxNoy – Coding Interviews (<http://maxnoy.com/interviews.html>)

DeZyre – 100 Hadoop Interview Questions and Answers (<https://www.dezyre.com/article/top-100-hadoop-interview-questions-and-answers-2016/159>)

Interactive tutorials for practicing Python and SQL for an interview

Tutorials Point – Python Interview Questions
(https://www.tutorialspoint.com/python/python_interview_questions.htm)

Tutorials Point – SQL Interview Questions
(https://www.tutorialspoint.com/sql/sql_interview_questions.htm)



The advertisement features a woman on the left and a man on the right. The woman has long dark hair and is wearing a pink floral shirt and blue polka-dot pants. The man has a beard and glasses, wearing a white t-shirt with the text 'Eat. Sleep. Data.' and dark pants. They are surrounded by various data science icons: a lightbulb, a flask, a bar chart, a line graph, binary code (1101), and molecular structures. The Springboard logo is in the top right. The main text reads 'Get mentored 1-to-1 by top data scientists'. A blue circle on the right contains the text 'Get a Data Science job, Guaranteed!'. An arrow points down to the URL www.springboard.com.

Springboard

Get mentored
1-to-1 by top
data scientists

Get a Data
Science job,
Guaranteed!

www.springboard.com

(https://www.springboard.com/workshops/data-science-career-track?utm_source=blog&utm_campaign=springboardbottombanner&utm_medium=blog)

Data Science Career Track

Get a data science job or your money back [Click Here](#)

Get a data science job or your money back [Click Here](#)

Five Awesome Data Science Projects
from Bootcamp Graduates
(<https://www.springboard.com/blog/data-science-projects-bootcamp-graduates/>)

50 of the Best Data Science Blogs
(<https://www.springboard.com/blog/data-science-blogs/>)

4 Comments

Springboard Blog

 Login ▾

 Recommend

 Share

Sort by Best ▾



Join the discussion...

LOG IN WITH

OR SIGN UP WITH DISQUS 

Name



Yongkai Wong • 7 months ago

A minor correction on the first assumption required for linear regression: it should be that there is a linear relationship between the dependent variable and the regressors, not the explanatory variables. For example, suppose we have one explanatory variable X . It is possible to run $Y = b_0 + b_1 X + b_2 X^2$ or $Y = b_0 + b_1 \log(X)$ and etc.

Get a data science job or your money back [Click Here](#)



Roger Huang • Yongkai Wong • 7 months ago

Thanks Yongkai--made an edit :)

^ | v • Reply • Share >



paulartois • 10 months ago

There are multiple problems with this article. Some of the proffered answers are incorrect. For instance, in the linear regression section:

- The data is normally distributed

It's not the **data** that is assumed to be normal, but the **errors** or residuals. It doesn't make sense for the data itself to be normally distributed.

- There is no auto-correlation

Autocorrelation is a concept strictly related to time-series data, where there is an ordered sequence. It does not apply to general linear regression. The correct notion is i.i.d., that is, the errors are independent and identically distributed.

^ | v • Reply • Share >



Roger Huang → paulartois • 10 months ago

Hi Paul,

Thanks for catching those. I've made some edits to correct those errors.

Cheers,

Roger

^ | v • Reply • Share >

ALSO ON SPRINGBOARD BLOG

Data Mining in Python: A Guide

2 comments • a year ago•

Charles Sutton — Dude. This rocks. Really helped me with some Python driven marketing intelligence. Nice job.

How Bayes Theorem, Probability, Logic and Data Interact

1 comment • a year ago•

micalael — What a simple and interesting way to explain Bayes Theorem. Thanks for writing it.

A comprehensive introduction to data wrangling

7 comments • 8 months ago•

Roger Huang — Haha, nice catch both of you. I've fixed it in the article as well :)

31 Free Data Visualization Tools

10 comments • a year ago•

Jared Bloom — Thanks for the up-to-date list. May I suggest another free chart editor for non-programmers – <http://charte.ca> ?

Get a data science job or your money back [Click Here](#)

Millions of students are improving their skills by using our courses

LEARN MORE

(/workshops/)

Stay in the know!

I Want To Learn...



Enter your email

GET UPDATES

Workshops



Resources



Company



© 2017 SlideRule Labs Inc. All rights reserved. [Terms \(/terms/\)](#) | [Privacy \(/privacy/\)](#)

[\(https://www.facebook.com/springboard/\)](https://www.facebook.com/springboard/)