

68 R Language interview questions (and answers) to hire top data analysts

Questions

1. Can you explain the difference between a data frame and a matrix in R?
2. How would you handle missing values in a dataset using R?
3. What is the purpose of the 'apply' family of functions in R?
4. How do you install and load a package in R?
5. Can you describe what a factor is and how it is used in R?
6. What are some common data visualization libraries available in R?
7. How do you perform a linear regression in R?
8. Can you explain how to merge two data frames in R?
9. What is the difference between 'lapply' and 'sapply' functions?
10. How do you create a custom function in R?
11. Can you explain what the 'tidyverse' is in R?
12. How would you describe the purpose of the 'ggplot2' package?
13. What do you understand by 'data wrangling' in R?
14. Can you explain the concept of 'vectorization' in R and its benefits?
15. How do you handle categorical variables in R?
16. What is the 'pipe' operator in R, and why is it useful?
17. Can you describe what 'data imputation' is and why it might be necessary?
18. How do you ensure reproducibility in your R projects?
19. How would you use the 'dplyr' package to perform group-wise operations on a dataset?
20. Can you explain the concept of 'lazy evaluation' in R and how it's beneficial?
21. What are the key differences between 'for' loops and 'while' loops in R?
22. How would you create and manipulate a list of lists in R?
23. Can you describe the purpose and usage of the 'aggregate' function in R?
24. What is the difference between 'rbind' and 'cbind' functions, and when would you use each?
25. How do you handle and analyze time-series data in R?
26. Can you explain the concept of 'recursion' and provide an example in R?
27. What are S3 and S4 object systems in R, and how do they differ?
28. How would you perform k-means clustering in R?
29. Can you explain what 'regular expressions' are and how to use them in R?
30. What is the purpose of the 'reshape2' package in R?
31. How would you create a custom plot using base R graphics?
32. Can you explain the concept of 'scope' in R and how it affects variable accessibility?
33. How do you optimize R code for better performance?
34. How would you explain the concept of hypothesis testing in R?
35. Can you describe how you would perform a correlation analysis in R?
36. How would you explain the concept of p-value in the context of statistical testing?
37. What is the role of AIC and BIC in model selection in R?
38. How do you interpret the results of an ANOVA test in R?
39. Can you describe how you would check for multicollinearity in a dataset using R?
40. What are some common ways to handle outliers in a dataset using R?
41. How would you explain the concept of a confidence interval to a non-technical person?
42. What is the purpose of conducting a chi-squared test in R, and how do you interpret its results?
43. Can you demonstrate how to filter rows in a data frame based on a condition using the 'dplyr' package?
44. How would you use the 'tidyr' package to reshape a data frame from wide to long format?
45. What is the use of the 'mutate' function in the 'dplyr' package, and can you provide an example?
46. How do you remove duplicate rows from a data frame in R?
47. Can you explain how to use the 'stringr' package for text data cleaning in R?
48. What techniques do you use in R to identify and handle outliers in a dataset?
49. How would you merge multiple data frames in R with different schemas?
50. Can you describe how to perform data normalization and scaling in R?
51. How do you handle inconsistent data formats in a column using R?
52. What is the purpose of the 'lubridate' package, and how do you use it to handle date-time data?
53. Can you demonstrate how to use the 'forcats' package to manipulate factor levels in R?
54. How would you use 'regex' in R for advanced data cleaning tasks?
55. Describe a time when you had to clean a highly messy dataset. What steps did you take using R?
56. How would you approach optimizing a slow-running R script that processes a large dataset?
57. Can you give an example of how you used R for a time-sensitive project? What were the challenges and how did you overcome them?
58. Imagine you are given a dataset with mixed data formats in a single column. How would you handle this using R?
59. Explain a situation where you had to choose between different R packages to solve a problem. What was your decision-making process?
60. Describe a project where you used R to integrate data from multiple sources. What were the key challenges and how did you address them?
61. How would you handle a situation where your R code produces unexpected results? What steps would you take to debug it?
62. Can you discuss a time when you had to use R to conduct an A/B test? What were the steps and how did you ensure statistical validity?
63. Have you ever had to work with unstructured data in R? How did you approach it and what tools did you use?
64. What strategies do you use to ensure that your R code is maintainable and understandable by other team members?
65. Describe a scenario where you had to use advanced R features to solve a complex data analysis task. What features did you use and why?
66. How have you automated repetitive data processing tasks in R? Can you provide a specific example?