Which of the following are examples of predictive analysis? Select two answers. A: Dashboards. B: Pandemic trends prediction. C: Returning a summary of descriptive statistics for data. D: Understanding human languages.	Pandemic trends prediction. Understanding human languages.
Which of the following statements about numeric and integer values are true? Select three values. A: The converted value of a numeric to an integer is always equal to the original numeric value. B: The converted value of an integer to a numeric is always equal to the original integer value. C: A numeric value can be converted to an integer. D: An integer can be converted to a numeric value.	The converted value of an integer to a numeric is always equal to the original integer value. A numeric value can be converted to an integer. An integer can be converted to a numeric value.
What is the result of the R expression 4 + 3 * 25? A: 79 B: 103 C: 74 D: 175	79
Which R function saves a workspace to a .RData file? A: save.data() B: save.workspace() C: save.image() D: save.file()	save.image()
In RStudio, which of the following statements about writing code in the File Editor and the Console are true? Select two answers. A: Only files containing R code can be	

edited in the File Editor B: You write code in the Console when you want to try out R commands or to run a few lines of code. C: You write code, usually for multiple lines of code, in the File Editor and execute them in batch mode. D: Code in the File Editor executes immediately as you type it so you can see the results quickly.	You write code in the Console when you want to try out R commands or to run a few lines of code. You write code, usually for multiple lines of code, in the File Editor and execute them in batch mode.
Complete the statement: A Jupyter Notebook is made up of a series of _____ that you can use to write, run, and interact with your code. A: Workspaces B: Objects C: Cells D: Files	Cells
R can perform several forms of statistical computation. What is an example of hypothesis testing? A: Inferring an unknown mean value of a population from its samples. B: Compute and visualize a correlation matrix among four different variables to see if they are correlated. C: Obtaining a representative subset of data. D: Testing if the mean values of two groups are statistically different.	Testing if the mean values of two groups are statistically different.
Which of the following data type conversions may be not allowed in R? A: logical (like TRUE or FALSE) to numeric B: integer (like 1L or 2L) to numeric C: character (like `1`, `A`, or `test`) to numeric D: numeric (like 1 or 2) to integer	character (like `1`, `A`, or `test`) to numeric

What is the result of the R expression 100 * (5 - 3)? A: 200 B: 497 C: 500 D: 503	200
After you write code in an R script file or the R Console, what component of the R environment parses the code into objects in memory? A: R data files B: R variables, functions, and datasets C: R Workspace D: R Interpreter	R Interpreter
Which features of RStudio help facilitate code writing? Select two answers. A: Code auto completion B: File Explorer C: Syntax highlighting D: Workspace visualization	Code auto completion Syntax highlighting
True or False: Execution order does not matter when executing cells in a Jupyter notebook A: True B: False	False
What is the difference between the expression c(1, 2, 3, 4, 5) and the expression c(5:1)? A: The two expressions produce the same result. B: They both produce a vector with five numbers but the first is in ascending order and the second is in descending order. C: They both produce a factor with five numbers but the first is in ascending order and the second is in descending order.	They both produce a vector with five numbers but the first is in ascending order and the second is in descending order.

D: One produces a factor and the other produces a vector.	
Assume that the variable test_result contains the vector c(25, 35, 40, 50, 75). What is the result of the expression test_result[test_result < 50]? A: [1] 25 35 40 B: [1] TRUE TRUE TRUE FALSE FALSE C: [1] TRUE TRUE TRUE TRUE FALSE D: [1] 25 35 40 50	[1] 25 35 40
What is the main difference between a list and a vector? A: A list is a multi-dimensional array of values, while a vector is a single dimensional array of values. B: It is not possible to add or remove items from a list, but you can do this with a vector. C: A list can contain different types of data, while a vector may only contain one type of data. D: A list can contain nominal or ordinal values, while a vector cannot.	A list can contain different types of data, while a vector may only contain one type of data.
What are three types of data you can store in an array or matrix? Select three answers. A: Numeric valus B: Strings C: Vectors D: Integers	Numeric valus Strings Integers
In a data frame, each column is represented by a _____ of values of the same data type. A: Vector B: Variable C: List D: Matrix	Vector

What is a nominal factor? A: A factor with any type or number of elements. B: A factor with no implied order. C: A factor with ordering. D: A factor that contains numeric data.	A factor with no implied order.
Assume that the variable test_result contains the vector c(25, 35, 40, 50, 75).What is the result of the expression mean(test_result)? A: 45 B: 40 C: 50 D: 35	45
Assume you have variable called employee that contains the expression list(name = "Juan", age = 30). What is the correct command to change the contents of the age item to 35? A: employee["age"] <- 35 B: employee[age] = 35 C: employee[age] <- 35 D: employee["age"] == 35	employee["age"] <- 35
What is the main difference between a matrix and an array? A: A matrix must be two dimensional, but an array can be single, two dimensional, or more than two dimensional. B: A matrix can contain multiple types of data, but an array can only contain data of the same type. C: A matrix can contain vectors, but an array can only contain strings, characters, or integers. D: A matrix can be arranged by rows or columns, but an array is always arranged by columns.	A matrix must be two dimensional, but an array can be single, two dimensional, or more than two dimensional.

Assume that you have a data frame called employee that contains three variables: name, age, and title. If you want to return all the values in the title variable, what command should you use? A: employee$title B: employee.title C: employee[[3]] D: employee[title]	employee$title
Which looping statement should you use if you want to continue to perform an operation if the value of an expression is true? A: Do While loop B: For Each loop C: For loop D: While loop	While loop
What happens if a user-defined function is missing the return statement? A: The function returns an error. B: The function returns the last evaluated expression in the function. C: The function returns a NULL value. D: The function returns the first evaluated expression it encounters and then immediately exits the function.	The function returns the last evaluated expression in the function.
Assume you have a variable called word_string that contains the string "The quick red fox jumped over the lazy dog." Which function can you use to replace the word "dog" with the word "cat" in the word_string variable? A: strsplit() B: substr() C: nchar() D: chartr()	chartr()
Which of the following statements about regular expressions are true? Select two	

answers. A: Regular expressions are primarily used for data extraction. B: Regular expressions evaluate a condition to determine if it is true or false. C: Regular expressions perform mathematical operations on numeric values. D: Regular expressions are used to match patterns in strings and text.	Regular expressions are primarily used for data extraction. Regular expressions are used to match patterns in strings and text.
How do you convert a UNIX date/time format to an R Date object? A: Pass the UNIX date/time value to the as.POSIXct() function and then pass that value to the as.Date() function. B: Pass the UNIX date/time value to the is.Date() function. C: Pass the UNIX date/time value to the as.Date() function. D: Pass the UNIX date/time value to the as.POSIXct() function and then pass that value to the is.Date() function.	Pass the UNIX date/time value to the as.POSIXct() function and then pass that value to the as.Date() function.
What type of statement should you add to your code if you know that an error or warning might occur? A: A tryError statement. B: Incorrect distractor C: A tryCatch statement. D: A catchError statement.	A tryCatch statement.
What is the result of the conditional statement 25 > 15 | 99 >= 100? A: TRUE B: FALSE	TRUE
How do you define a global variable in a function? A: Use the -> assignment operator. B: Use the <- assignment operator. C: Use the <<- assignment operator. D: Use the == assignment operator.	Use the <<- assignment operator.

You can use the str_sub() function to form a substring by counting back from the last position. This function is part of which package? A: stringr B: tidyr C: purrr D: readr	stringr
Assume you have a data frame that contains a string variable called 'phone'. The phone numbers in this variable appear in (###) ###-#### or ###-###-#### format. Which feature of R can you use to isolate the area code (the three numbers between the parentheses or the first three numbers)? A: A regular expression. B: A mathematical operation. C: It is not possible to do this using R. D: A string operation.	A regular expression.
When you convert a date in string format to a Date object, what information do you need to pass to the as.Date() function? Select two answers. A: The string containing the date. B: The UNIX format of the string. C: The number of days since January 1, 1970. D: The date format of the string.	The string containing the date. The date format of the string.
What is the difference between an error and a warning in you R code? A: An error halts code execution, while a warning does not. B: A warning halts code execution, while an error does not. C: You can catch a warning, but you cannot catch an error.	An error halts code execution, while a warning does not.

D: You can catch an error, but you cannot catch a warning.	
Which package do you need to install before reading an Excel file in R? A: readExcel B: readxlsx C: readxl D: No package is needed. This functionality is built into R.	readxl
What is the purpose of the readLines() function? A: Reads each text line (with a line break) in a text file and returns a data frame. B: Reads each text line (with a line break) in a text file and returns a character variable. C: Reads each text line (ending with a line break) in a text file and returns a character vector. D: Reads each text line (ending with a line break) in a text file and returns a character factor.	Reads each text line (ending with a line break) in a text file and returns a character vector.
Which data structure is more suitable to be exported into a text file rather than a .csv or Excel file? A: String or characters B: Matrices C: Data frames D: Datasets	String or characters
Complete the sentence: _____ is a communication protocol for fetching web resources for clients from servers on the Internet. A: HTTP B: REST C: TCP D: DNS	HTTP

Complete the sentence: In an HTML page, if the <head> and <body> nodes have the same parent, <html>, they are said to be _____ to each other. A: Sibling nodes B: Nested nodes C: Root nodes D: Child nodes	Sibling nodes
Assume you have read a .csv file into a data frame variable called employee. It has 20 rows of data and three variables: name, age, and title. What is the correct statement to use to return the fifth row of data in the name and title columns? A: employee[c("name", "title"), 5] B: employee[5, c("name", "title")] C: employee[5, 2:3] D: employee[2:3, 1:5]	employee[5, c("name", "title")]
How do you return the number of characters in each paragraph of a text file that has been read into a character vector? A: Use the file.size() function. B: Use the length() function. C: Use the nchar() function. D: Use the scan() function.	Use the nchar() function.
Which package do you need to install before writing to an Excel file in R? A: No package is needed. This functionality is built into R. B: writexl C: xlsx D: writexlsx	xlsx
You want to get a resource by its URL using an HTTP request and assign the HTTP response containing status code, headers, response body to a response variable. Which function should you use?	

A: response <-PUT("https://www.mysite.com") B: response <-GET("https://www.mysite.com") C: response <-POST("https://www.mysite.com") D: response <-HEAD("https://www.mysite.com")	response <-GET("https://www.mysite.com")
After reading an HTML page from a URL, what must you do to get the <body> node from the root <html> node? A: Use the html_text() function to return the <html> node. B: Use the html_text() function to return the <body> node of the HTML. C: Use the html_node() function to return the <body> as a child node of <html> node. D: Use the html_node() function to return the <html> node.	Use the html_node() function to return the <body> as a child node of <html> node.
Which of the following is a typical way that developers use the R language? A: Systems programming B: Web page interactivity C: Video game development D: Predictive analysis	Predictive analysis
In R, what is the result of the function as.numeric(TRUE)? A: 4 B: 1 C: FALSE D: NA	1
One movie is 150 minutes long, and another is 90 minutes long. Using R, which of the following commands would correctly calculate the difference in length, in seconds, between the two films? A: 150 - (90 * 60)	(150 - 90) * 60

B: (150 - 90 *60) C: (150 - 90) * 60 D: 150 - 90 * 60	
Which command in R would return the following numeric vector? 5 4 3 2 1 A: c(5-\>1) B: c(1,2,3,4,5) C: c(5:1) D: c(1:5)	c(5:1)
In R, assume you have a vector named "age," and each element in the vector is the age of one person in a group. Which command must you use to reorder the ages from youngest to oldest? A: rank(age) B: order(age) C: call(age) D: sort(age)	sort(age)
Assume the array books_array contains 6 elements. The array has three rows and two columns and appears as follows: [,1] [,2] [1,] "It" "Dr. Sleep" [2,] "Misery" "Carrie" [3,] "The Shining" "The Mist" If you input the books_array[,1] com-mand, what will be the output? A: "It" "Misery" "The Shining" B: "The Shining" C: "It" "Dr.Sleep" D: "It"	"It" "Misery" "The Shining"
In R, which command should you use to insert a new row into a data frame? A: rbind B: tail C: integrate D: head	rbind

Assume that the function add is defined as follows: add <- function(x,y) { (x + y) return (x - y) temp <<- (x * y) return (x / y) } What will be the output if you issue the command add(10,5)? A: 15 B: 2 C: 5 D: 50	5
What is the first step you must take before you can read an Excel spreadsheet in R? A: Run the c function. B: Install the readxl library. C: Run the read_excel function. D: Convert the spreadsheet's dataset.	Install the readxl library.
After installing and calling the httr library in R, which command can you use to request information about https://www.google.com? A: PUT("https://www.google.com/") B: BROWSE("https://www.google.com/") C: GET("https://www.google.com/") D: PATCH("https://www.google.com/")	GET("https://www.google.com/")
Which of the following statements are correct about databases? A: There are different types of databases - Relational, Hierarchical, No SQL, etc. B: A database is a repository of data C: All of the above D: A database can be populated with data and be queried	All of the above
True or False: A SELECT statement is used to retrieve data from a table. A: True B: False	True

You are working on a Film database, with a FilmLocations table. You want to retrieve a list of films that were released in 2019.You run the following query but find that all the films in the FilmLocations table are listed. SELECT Title,ReleaseYear, Locations FROMFilmLocations; What is missing? A: Nothing, the query is correct. B: A WHERE clause to limit the results to films released in 2019. C: A LIMIT clause to limit the results to films released in 2019. D: A DINSTINCT clause to specify a distinct year.	A WHERE clause to limit the results to films released in 2019.
Which of the following statements would you use to add a new instructor to the Instructor table. A: SELECT Instructor(ins_id, lastname, firstname, city, country) FROM VALUES(4, 'Doe', 'John', 'Sydney', 'AU'); B: INSERT INTO Instructor(ins_id, lastname, firstname, city, country) VALUES(4, 'Doe', 'John', 'Sydney', 'AU'); C: ADD INTO Instructor(ins_id, lastname, firstname, city, country) VALUES(4, 'Doe', 'John', 'Sydney', 'AU'); D: UPDATE Instructor(ins_id, lastname, firstname, city, country) WITH VALUES(4, 'Doe', 'John', 'Sydney', 'AU');	INSERT INTO Instructor(ins_id, lastname, firstname, city, country) VALUES(4, 'Doe', 'John', 'Sydney', 'AU');
What is the function of a WHERE clause in an UPDATE statement? A: A WHERE clause enables you to specify a new table to receive the updates.	

B: A WHERE clause is never used with an UPDATE statement. C: A WHERE clause enables you to specify which rows will be updated. D: A WHERE clause enables you to list the column and data to be updated.	A WHERE clause enables you to specify which rows will be updated.
True or False: The SELECT statement is called a query, and the output we get from executing the query is called a result set. A: True B: False	True
True or False: The INSERT statement can be used to insert multiple rows in a single statement. A: True B: False	True
Assume there exists an INSTRUCTOR table with several columns including FIRSTNAME, LASTNAME, etc. Which of the following is the most likely result set for the following query: SELECT DISTINCT(FIRSTNAME) FROM INSTRUCTOR A: LEON LEON PAUL PAUL B: LEON PAUL JOE C: LEON PAUL LEON JOE D: LEON KATSNELSON PAUL ZIKOPOLOUS JOE SANTARCANGELO	LEON PAUL JOE

What does the following SQL statement do? UPDATE INSTRUCTOR SET LAST-NAME = 'Brewster' WHERE LASTNAME = 'Smith' A: Changes all rows for the instructor with a last name of Smith to have a last name of Brewster. B: Change the row for the instructor with a last name of Brewster to have a last name of Smith. C: Change all rows in the table to have a last name of Smith. D: Change all rows in the table to have a last name of Brewster.	Changes all rows for the instructor with a last name of Smith to have a last name of Brewster.
Which of the following SQL statements will delete the authors with IDs of A10 and A11? A: DELETE FROM AUTHOR WHERE AUTHOR_ID IS ('A10', 'A11') B: DELETE FROM AUTHOR WHERE AUTHOR_ID IN ('A10', 'A11') C: DELETE ('A10', 'A11') FROM AU-THOR D: D: DELETE AUTHOR_ID IS ('A10', 'A11') FROM AUTHOR	DELETE FROM AUTHOR WHERE AU-THOR_ID IN ('A10', 'A11')
What is the function of a primary key? A: The primary key is used to identify any rows in the table that contain NULL values. B: The primary key is used to grant ac-cess to a table. C: The primary key enables you to add data to columns. D: The primary key uniquely identifies each row in a table.	The primary key uniquely identifies each row in a table.
True or False: Data Manipulation Lan-guage statements like INSERT, SE-	

LECT, UPDATE, and DELETE are used to read and modify data. A: True B: False	True
Data Definition Language (or DDL) statements are used to define, change, or delete database objects such as tables. Which of the following statements are all DDL statements? A: SELECT, INSERT, UPDATE B: CREATE, ALTER, DROP C: INSERT and UPDATE D: SELECT and DELETE	CREATE, ALTER, DROP
Which of the following queries will change the data type of an existing column (phone) to the varchar data type? A: ALTER TABLE author ALTER COLUMN phone SET DATA TYPE VARCHAR(20); B: ALTER TABLE author ALTER COLUMN phone DATA TYPE = VARCHAR(20); C: ALTER COLUMN phone SET DATA TYPE VARCHAR(20); D: ALTER TABLE author ALTER COLUMN phone SET TYPE VARCHAR(20);	ALTER TABLE author ALTER COLUMN phone SET DATA TYPE VARCHAR(20);
The five basic SQL commands are: A: CREATE, INSERT, RETRIEVE, MODIFY, DELETE B: None of the above C: SELECT, COPY, PASTE, INSERT, ALTER D: CREATE, SELECT, INSERT, UPDATE, DELETE	CREATE, SELECT, INSERT, UPDATE, DELETE
The primary key of a relational table uniquely identifies each _____ in a table. A: row	row

B: column C: relation D: attribute	
Which of the following statements about a database is/are correct? A: A database is a logically coherent collection of data with some inherent meaning B: Data can only be added and queried from a database, but not modified. C: Only SQL can be used to query data in a database. D: All of the above	A database is a logically coherent collection of data with some inherent meaning
Attributes of an entity become _____ in a table. A: rows B: columns C: constraints D: keys	columns
What are the basic categories of the SQL language based on functionality? A: Data Definition Language B: Data Manipulation Language C: Both of the above D: None of the above	Both of the above
The CREATE TABLE statement is a.... A: DML statement B: DDL statement C: Both of the above	DDL statement
You want to retrieve a list of employees in alphabetical order of Lastname from the Employees table. Which SQL statement should you use? A: SELECT * FROM Employees GROUP BY Lastname; B: SELECT * FROM Employees ORDER BY Lastname DESC;	SELECT * FROM Employees ORDER BY Lastname;

C: SELECT * FROM Employees ORDER BY Lastname; D: SELECT * FROM Employees SORT BY Lastname;	
Which keyword is used to set a condition for a GROUP BY clause? A: HAVING B: SELECT C: ORDER BY D: WHERE	HAVING
You want to retrieve a list of authors from Australia, Canada, and India from the table Authors. Which SQL statement is correct? A: SELECT * FROM Author WHERE Country LIST ('CA', 'IN'); B: SELECT * FROM Author IF Country ('Australia', 'Canada', 'India'); C: SELECT * FROM Author WHERE Country BETWEEN('Australia', 'Canada', 'India'); D: SELECT * FROM Author WHERE Country IN ('Australia', 'Canada', 'India');	SELECT * FROM Author WHERE Country IN ('Australia', 'Canada', 'India');
You want to retrieve a list of books priced above $10 and below $25 from the table Book. What are the two ways you can specify the range? A: SELECT Title, Price FROM Book WHERE Price BETWEEN 10 and 25; B: SELECT Title, Price FROM Book WHERE Price 10 to 25; C: SELECT Title, Price FROM Book WHERE Price IN (10, 25); D: SELECT Title, Price FROM Book WHERE Price >= 10 and Price <= 25;	SELECT Title, Price FROM Book WHERE Price BETWEEN 10 and 25; SELECT Title, Price FROM Book WHERE Price >= 10 and Price <= 25;
You want to retrieve Salary information for an employee called Ed from the Employee table. You write the following	

statement: SELECT Firstname, Lastname, Salary FROM Employees You see all the employees listed, and it's hard to find Ed's information. Which clause should you add to reduce the number of rows returned? A: ORDER BY Firstname; B: GROUP BY Firstname = 'Ed'; C: WHERE Firstname = 'Ed'; D: WHERE Employees = 'Ed';	WHERE Firstname = 'Ed';
You want to select author's last name from a table, but you only remember the author's last name starts with the letter B, which string pattern can you use? A: SELECT lastname from author where lastname like 'B#' B: SELECT lastname from author where lastname like 'B%' C: SELECT lastname from author where lastname like 'B$' D: None of the above	SELECT lastname from author where lastname like 'B%'
In a SELECT statement, which SQL clause controls how the result set is displayed? A: ORDER BY clause B: ORDER IN clause C: ORDER WITH clause	ORDER BY clause
Which of the following can be used in a SELECT statement to restrict a result set? A: HAVING B: WHERE C: DISTINCT D: All of the above	All of the above
When querying a table called Author that contains a list of authors and their country of residence, which of the following	

queries will return the number of authors from each country? A: SELECT Country, count(Country) FROM Author GROUP BY Country B: SELECT Distinct(Country) FROM Author C: SELECT Country, count(Country) FROM Author D: SELECT Country, distinct(Country) FROM Author GROUP BY Country	SELECT Country, count(Country) FROM Author GROUP BY Country
You want to retrieve a list of books that have between 450 and 600 pages. Which clause would you add to the following SQL statement: SELECT Title, Pages FROM Book _____ A: IF Pages >= 450 and Pages <= 600 B: WHERE Pages = 450 C: WHERE Pages 450 - 600 D: WHERE Pages >= 450 and pages <= 600	WHERE Pages >= 450 and pages <= 600
Which of the following statements about built-in database functions is correct? A: Built-in database functions may increase network bandwidth consumed. B: Built-in database functions must be called from a programming language like Python. C: Built-in database functions may increase processing time. D: Built-in database functions reduce the amount of data that is retrieved.	Built-in database functions reduce the amount of data that is retrieved.
Which of the following SQL queries would return the day of the week each dog was rescued? A: SELECT DAYOFWEEK(RescueDate) From PetRescue WHERE Animal =	

'Dog'; B: SELECT DAYOFWEEK(RescueDate) From PetRescue; C: SELECT RescueDate From PetRescue WHERE Animal = 'Dog'; D: SELECT DAY(RescueDate) From PetRescue WHERE Animal = 'Dog';	SELECT DAYOFWEEK(RescueDate) From PetRescue WHERE Animal = 'Dog';
Which of the following queries will return the employees who earn less than the average salary? A: SELECT * FROM Employees WHERE Salary < (SELECT AVG(Salary)) B: SELECT AVG(Salary) FROM Employees WHERE Salary < AVG(Salary) C: SELECT * FROM Employees WHERE Salary < AVG(Salary) D: SELECT * FROM Employees WHERE Salary < (SELECT AVG(Salary) FROM Employees);	SELECT * FROM Employees WHERE Salary < (SELECT AVG(Salary) FROM Employees);
What are the three ways to work with multiple tables in the same query? A: Built-in functions, implicit joins, JOIN operators B: Sub-queries, Implicit joins, normalization. C: Sub-queries, APPEND, JOIN operators D: Sub-queries, Implicit joins, JOIN operators	Sub-queries, Implicit joins, JOIN operators
Which of the following will retrieve the LOWEST value of SALARY in a table called EMPLOYEES? A: SELECT MAX(SALARY) FROM EMPLOYEES B: SELECT LOWEST(SALARY) FROM EMPLOYER C: SELECT SALARY FROM EMPLOYEES WHERE MINIMUM(SALARY) =	SELECT MIN(SALARY) FROM EMPLOYEES

SALARY D: SELECT MIN(SALARY) FROM EMPLOYEES	
Which of the following queries will return the first name of the employee who earns the highest salary? A: SELECT FIRST_NAME FROM EMPLOYEES WHERE SALARY IS HIGHEST B: SELECT FIRST_NAME, MAX(SALARY) FROM EMPLOYEES GROUP BY F_NAME C: SELECT MAX(SALARY) FROM EMPLOYEES D: SELECT FIRST_NAME FROM EMPLOYEES WHERE SALARY = (SELECT MAX(SALARY) FROM EMPLOYEES)	SELECT FIRST_NAME FROM EMPLOYEES WHERE SALARY = (SELECT MAX(SALARY) FROM EMPLOYEES)
Which of the following queries will return the data for employees who belong to the department with the highest value of department ID. A: SELECT * FROM EMPLOYEES WHERE DEP_ID = (SELECT DEPT_ID_DEP FROM DEPARTMENTS WHERE DEPT_ID_DEP IS MAX) B: SELECT * FROM EMPLOYEES WHERE DEP_ID = MAX(DEP_ID) C: SELECT * FROM EMPLOYEES WHERE DEPT_ID_DEP = MAX (SELECT DEPT_ID_DEP FROM DEPARTMENTS) D: SELECT * FROM EMPLOYEES WHERE DEP_ID = (SELECT MAX(DEPT_ID_DEP) FROM DEPARTMENTS)	SELECT * FROM EMPLOYEES WHERE DEP_ID = (SELECT MAX(DEPT_ID_DEP) FROM DEPARTMENTS)

A DEPARTMENTS table contains DEP_NAME, and DEPT_ID_DEP columns and an EMPLOYEES table contains columns called F_NAME and DEP_ID. We want to retrieve the Department Name for each Employee. Which of the following queries will correctly accomplish this? A: SELECT E.F_NAME, D.DEP_NAME FROM EMPLOYEES, DEPARTMENTS B: SELECT F_NAME, DEP_NAME FROM EMPLOYEES E, DEPARTMENTS D WHERE E.DEPT_ID_DEP = D.DEP_ID C: SELECT D.F_NAME, E.DEP_NAME FROM EMPLOYEES E, DEPARTMENTS D WHERE DEPT_ID_DEP = DEP_ID D: SELECT F_NAME, DEP_NAME FROM EMPLOYEES, DEPARTMENTS WHERE DEPT_ID_DEP = DEP_ID	SELECT D.F_NAME, E.DEP_NAME FROM EMPLOYEES E, DEPARTMENTS D WHERE DEPT_ID_DEP = DEP_ID
You are writing a query that will give you the total cost to the Pet Rescue organization of rescuing animals. The cost of each rescue is stored in the Cost column. You want the result column to be called "Total_Cost". Which of the following SQL queries is correct? A: SELECT SUM(Cost) FROM PetRescue B: SELECT SUM(Cost) AS Total_Cost FROM PetRescue C: SELECT SUM(Total_Cost) From PetRescue D: SELECT Total_Cost FROM PetRescue	SELECT SUM(Cost) AS Total_Cost FROM PetRescue
How can a relational database help R handle memory issues?	

A: Using a relational database increases memory use in R. B: Using a relational database with R has no impact on memory. C: Running SQL queries in a relational database reduces memory demands in the R client. D: Loading all the data from the relational database into a dataframe reduces memory demands.	Running SQL queries in a relational database reduces memory demands in the R client.
Which R function saves a single data structure to a .Rda file? A: save.image() B: save() C: saveRDS() D: readRDS()	saveRDS()
A dataframe in R is like which of the following relational database concept? A: Schema B: Table C: Database D: Column	Table
When mapping data types between R and a database, you should consider converting which of the following to strings? A: Logical B: Date C: Double D: Factor	Date
When designing a database and making decisions, which potential issue is not helped by normalization? A: Performance issues B: Lack of security C: Data redundancy D: Transaction processing problems	Lack of security

Which amongst the following is the simplest way to update individual observations in a dataframe? A: Convert the dataframe into a text file and make the changes there. B: There is no way to update individual observations directly in a dataframe. C: Store the data in a relational database instead and make the updates there. D: Make the updates in the dataframe and then store the results in a binary format.	Store the data in a relational database instead and make the updates there.
Which R function loads multiple R data structures from a .Rda file? A: save.image() B: save() C: load() D: readRDS()	load()
A variable in R is like which of the following relational database concept? A: Schema B: Column (or attribute) C: Row (or tuple) D: Table	Column (or attribute)
Which R variable holds the platform numeric limits for your R environment? A: .Computer B: .Precision C: .Numeric D: .Machine	.Machine
What is declarative referential integrity? A: Validates data normalization. B: Manages transactions by adhering to the ACID properties. C: Protects databases from the corruption, destruction, or removal of data. D: Manages dependency relationships between two tables.	Manages dependency relationships between two tables.

Which of the following statements about database packages for R are true? Select two answers. A: A DBI-based package can be used with many different database servers. B: The RODBC package can be used with many different database servers. C: The RODBC package can only be used with Microsoft database server products. D: DBI-based packages are database server specific.	The RODBC package can be used with many different database servers. DBI-based packages are database server specific.
Which RJDBC function retrieves the content of the result set in the form of a dataframe? A: retrieve B: print C: get_data D: fetch	fetch
The ODBC Driver Manager is: A: Platform specific B: Vendor specific C: Package specific D: Database server specific	Platform specific
What are the two steps you must complete outside of R before installing the RODBC package? Select two answers. A: Install the ODBC driver for your database server. B: List the names of the registered DSNs on your database server. C: Configure the ODBC DSN. D: Install the RODBC package from the CRAN website.	Install the ODBC driver for your database server. Configure the ODBC DSN.
Which RODBC function returns metadata that can help you preserve data integrity when working with database data	

in R? A: sqlColumns() B: sqlTables() C: sqlER D: sqlTypeInfo()	sqlTypeInfo()
What are the two categories of relational database access packages in R? A: JDBC B: ISO SQL/CLI C: ODBC D: Database Interface-Based (DBI)	ODBC Database Interface-Based (DBI)
What is the first step you must take before using the RJDBC package for R? A: Query the database using a SELECT statement. B: Load the DB2 JDBC type 4 driver and create a driver object. C: Load the RJDBC library. D: Create a connection object for a database on a remote server.	Load the RJDBC library.
Which of the following is one of the two components of ODBC? A: ODBC Runtime Environment B: ODBC Driver Manager C: Registered DNS D: Data sources	ODBC Driver Manager
Which SQL command returns the list of registered DSNs? A: print() B: registered() C: odbcDataSources() D: names()	names()
Which RODBC function does NOT help you learn about the schema of your database? A: sqlTypeInfo() B: odbcGetInfo()	odbcGetInfo()

C: sqlColumns() D: sqlTables()	
You are preparing to analyze some sales data and have a large amount of information in an Excel spreadsheet. You have decided to convert the Excel spreadsheet to a relational database. What is your first step? A: Create a logical and physical database design. B: Create the physical database objects. C: Clean and split the data into load files. D: Get the data into the database.	Create a logical and physical database design.
What is a referential constraint? A: This occurs when one table can't find a matching entry in another table. B: This occurs when there is a many-to-many relationship between tables. C: This occurs when there is a one-to-many relationship between tables. D: This occurs when there is a primary key/foreign key relationship between two tables.	This occurs when there is a primary key/foreign key relationship between two tables.
Which RODBC function can you use to create a new table in a database from R? A: CREATE TABLE B: You cannot do this from R. You need to use a database management tool to do this. C: sqlQuery() D: sqlAddTable()	sqlQuery()
Why is the SQL LOAD command recommended over the IMPORT command for large amounts of data? A: The LOAD command bypasses the database transactional logging mecha-	

nism, making it fast and efficient. B: The LOAD command performs SQL INSERT statements for a group of rows and commits them periodically, saving transaction process time. C: The LOAD command uses transactional logging, making it fast and efficient. D: The LOAD command is not recommended for large amounts of data. IMPORT is a much better choice.	The LOAD command bypasses the database transactional logging mechanism, making it fast and efficient.
After you query a database, how do you load query results into a dataframe so you can perform data analysis? Select two answers. A: Use the sqlCommit() function. B: Use the sqlQuery() function. C: Use the sqlFetch() function. D: Use the sqlLoad() function.	Use the sqlQuery() function. Use the sqlFetch() function.
What are two reasons to map an existing data source, like pre-existing database tables, database dump files, or raw data, to a relational database design? Select two answers. A: Limit the number of concurrent users to just you so it's secure. B: There's no reason to do this. It just adds a lot of extra work with little benefit. C: Eliminate redundancy in the data. D: Address issues with data normalization.	Eliminate redundancy in the data. Address issues with data normalization.
You have two tables in your database design: Customers, which lists all your customers, and Orders, which lists all the sales transactions that your customers have made over the years. The two tables each have a field called Customer_ID. Which of the following correct-	

ly describe the relationship between the two tables? A: Both the Customer_ID field in the Customers table and the Customer_ID field in the Orders table are Primary Keys. B: The Customer_ID field in the Customers table is a Foreign Key and the Customer_ID field in the Orders table is a Primary Key. C: Both the Customer_ID field in the Customers table and the Customer_ID field in the Orders table are Foreign Keys. D: The Customer_ID field in the Customers table is a Primary Key and the Customer_ID field in the Orders table is a Foreign Key.	The Customer_ID field in the Customers table is a Primary Key and the Customer_ID field in the Orders table is a Foreign Key.
What is the SQL DDL command can be used for adding primary keys to an existing table in a database? A: CREATE TABLE B: SQL QUERY C: ALTER TABLE D: UPDATE TABLE	ALTER TABLE
What is the recommend SQL command for loading small to medium amounts of data into a database? A: LOAD command B: LOAD or IMPORT (there is no difference) C: IMPORT command	IMPORT command
What are two ways to limit database movement and increase performance when querying a database? A: Load all the data directly into a dataframe to reduce the number of times you must revisit the database.	

B: Use the sqlQuery() or sqlFetch() commands. C: Use SQL functions provided by the database vendor whenever possible. D: Use stored procedures when possible.	Use SQL functions provided by the database vendor whenever possible. Use stored procedures when possible.
What is wrong with the following INSERT statement? INSERT INTO Movie (MOVIE_ID, MOVIE TITLE, RELEASE_YEAR) VALUES (52, 'Princess Bride', 1987), ('When Harry Met Sally', 1989); A: The column names are in the incorrect order. B: It is missing a value. C: The word "INTO" should be removed. D: It is missing the keyword "COLUMNS."	It is missing a value.
Which of the following displays the correct general syntax for the SELECT statement that also includes a predicate? A: WHERE <logical statement> SELECT * FROM <tablename>; B: IF <logical statement> THEN SELECT * FROM <tablename>; C: SELECT * FROM <tablename> WHERE <logical statement>; D: SELECT * FROM <tablename> IF <logical statement>;	SELECT * FROM <tablename> WHERE <logical statement>;
What is a primary key? A: It creates a link between two or more tables. B: It uniquely defines the columns in a table. C: It is a unique value that defines a table. D: It is a unique value for each row in a table.	It is a unique value for each row in a table.

Which two statements are true regarding data manipulation language (DML) operations? A: They are often referred to as create, read, update, and delete (CRUD) operations. B: They are used to read and modify data in tables. C: They are used to define, change, or drop database objects such as tables. D: They are used to define relationships among tables.	They are often referred to as create, read, update, and delete (CRUD) operations. They are used to read and modify data in tables.
Which of the following SQL statements is equivalent to "SELECT title, pages FROM Book WHERE pages >= 290 AND pages <= 300;"? A: SELECT title, pages FROM Book WHERE pages BETWEEN 290 AND 300;. B: SELECT pages BETWEEN 290 AND 300 FROM Book; C: SELECT title, pages FROM Book WITH pages BETWEEN 290 AND 300,•. D: None of the above. There is not another equivalent form of that SQL statement.	SELECT title, pages FROM Book WHERE pages BETWEEN 290 AND 300;.
Which statement is the correct way to retrieve the average number of points scored by each basketball team? A: SELECT TeamName, AVG(Points) FROM Scores GROUP BY TeamName; B: SELECT Scores FROM TeamName, AVG(Points) GROUP BY TeamName; C: SELECT TeamName, AVG(Points) FROM Scores GROUP BY Scores; D: SELECT Scores FROM TeamName, AVG(Points) GROUP BY Scores;	SELECT TeamName, AVG(Points) FROM Scores GROUP BY TeamName;

What feature in a relational database management system is similar to a "namespace" in R? A: Tables B: Primary keys C: Schemas D: Attributes	Schemas
How should dates and times be mapped from R to a relational database management system (RDBMS)? A: As strings B: Using the toDate() method C: As objects D: As integers	As strings
Which RODBC method is most closely associated with the SQL TRUNCATE statement? A: sqlClear() B: sqlDrop() C: sqlTrim() D: sqlRemove()	sqlClear()
Data that needs transformed into a factor is generally transferred from an R object database management system (RODBMS) into R as what data type? A: An object B: A float C: A Boolean D: A character	A character
Data analysis plays an important role in which of the following scenarios? Select 3 answers. A: Answering questions. B: Finding data. C: Discovering useful information. D: Predicting the future.	Answering questions. Discovering useful information. Predicting the future.

Complete the sentence: In a dataset, a _____ is also referred to as a variable, feature, or attribute. A: Row B: Column C: Data frame D: Observation	Column
Which tidyverse package is used for data import and management? A: ggplot2 B: tidyr C: readr D: dplyr	readr
What is the next step must you perform after you download a dataset file from a URL? A: Use the write_csv() function to read the dataset into a data frame B: Download the file using the download.file() function. C: Unzip the file using the untar() function. D: Use the read_csv() function to read the dataset into a data frame.	Unzip the file using the untar() function.
You are checking your data using the glimpse() function before beginning your analysis and determine that the data type of a variable called TimeStamp is in a character format. What should you do next? A: Drop that column since it is too tricky to handle. B: Immediately change the character type to a date type. C: Evaluate how you plan to use this variable in your data analysis. D: Nothing. R functions can handle dates	Evaluate how you plan to use this variable in your data analysis.

and times in character format without modification.	
What is the purpose of the Data Asset eXchange? A: Provides data that you can explore to conduct data analysis. B: Helps you exchange data with others. C: Provides data that you can use for a small fee. D: Provides data that is only useful for learning purposes.	Provides data that you can explore to conduct data analysis.
In the Airline Performance dataset from the Asset Data eXchange, which of the following variables is a target for predicting on-time arrivals? A: Distance B: ArrDelay C: SecurityDelay D: CarrierDelay	ArrDelay
What is the purpose of the pipe (%>%) operator? A: Assigns a value to a global variable. B: Combines two functions into a single operation. C: Assigns a value to a variable. D: Combines multiple functions into a single operation.	Combines multiple functions into a single operation.
Which function can you use to read a text file that uses the "%" character as a delimiter? A: read_delim() B: read_csv() C: read_any() D: read_tsv()	read_delim()
What is the main similarity between the summarize() and group_by() functions? A: Both return a statistical summary of	

the data. B: Both group data by the specified variables. C: Both compute summary statistics. D: There is no similarity between the summarize() and group_by() functions.	Both return a statistical summary of the data.
The process of converting or mapping data from the initial raw form to another format to prepare it for further analysis goes by several names. What is this process commonly called? Select three answers. A: Data pre-processing B: Data cleaning C: Data wrangling D: Data formatting	Data pre-processing Data cleaning Data wrangling
What is the result of the following statement? sub_airline %>% map(~sum(is.na(.))) A: Counts the missing values in all columns in the dataset. B: Counts the missing values and returns the result only for columns in the dataset that have missing values. C: Counts all instances of zero in all columns in the dataset. D: Counts all instances of NA in all columns in the dataset.	Counts the missing values in all columns in the dataset.
Which functions do you use together to correct data types in all columns of your dataset? Select two answers. A: mutate() B: sapply() C: mutate_if() D: mutate_all()	mutate_if() mutate_all()
Which data normalization technique divides each value by the maximum value for that variable, resulting in new values	

that range between 0 and 1? A: Simple feature scaling B: Min-max C: Z-score	Simple feature scaling
With data binning, observations are often organized into defined intervals called quartiles. Which quartile is the median of the dataset? A: 4th quartile B: 3rd quartile C: 1st quartile D: 2nd quartile	2nd quartile
You want to access the "Date" column of a data frame called sales_data so you can perform an operation on it. What is the correct way to refer to this column? A: sales_data.Date B: sales_data#Date C: sales_data%Date D: sales_data$Date	sales_data$Date
Which function replaces missing values in a dataset? A: is.na() B: drop_na() C: drop_columns() D: replace_na()	replace_na()
You have a variable called "Status" that contains a status code in the format "error_type-severity_level", for example "10-07", and you want to reformat the column so that the "error_type" and "severity_level" are in different columns. What is the correct function to do this? A: dataframe %>% mutate_if(Status, sep = "-", into = c("error_type", "severity_level") B: dataframe %>% sapply(Status, sep = "-",	dataframe %>% separate(Status, sep = "-", into = c("error_type", "severity_level")

into = c("error_type", "severity_level") C: dataframe %>% mutate_all(Status, sep = "-", into = c("error_type", "severity_level") D: dataframe %>% separate(Status, sep = "-", into = c("error_type", "severity_level")	
What are two benefits of data normalization? A: Helps you better understand data distribution. B: Minimize the effects of outliers, which can influence the result more. C: Enables a fair comparison between the different features and making sure they have the same impact. D: Brings data into a common standard of expression that allows you to make meaningful comparisons.	Minimize the effects of outliers, which can influence the result more. Enables a fair comparison between the different features and making sure they have the same impact.
To visualize its distribution, binned data is often plotted in which of the following type of chart? A: Scatter plot B: Line chart C: Histogram D: Bar chart	Histogram
Which of the following can you accomplish using the spread() function? Select two answers. A: Convert categorical variables to dummy variables and assign the value of another variable to each category. B: Size down three variables into one. C: Convert categorical variables to dummy variables. D: Reformat the categorical variable that its contents are in two or more columns.	Convert categorical variables to dummy variables and assign the value of another variable to each category. Convert categorical variables to dummy variables.

When conducting exploratory data analysis, which visualizations are particularly useful for examining the distribution of numerical data and skewness through displaying the data quartiles (or percentiles) and averages? A: Heatmaps B: Scatter plots C: Boxplots D: Histograms	Boxplots
When grouping data and calculating the mean of each group as part of your exploratory data analysis, you typically use the group_by() function with which other function? A: arrange() B: summarize() C: desc() D: sort()	summarize()
Which of the following forms of exploratory data analysis is a statistical comparison of groups of data? A: Descriptive statistics B: Pearson correlation C: Correlation D: Analysis of variance (ANOVA)	Analysis of variance (ANOVA)
Which of the following statements describe a positive correlation between two variables? Select two answers. A: The correlation coefficient is greater than zero. B: The correlation coefficient is less than zero. C: Both variables move in opposite directions. D: Both variables move in the same direction.	The correlation coefficient is greater than zero. Both variables move in the same direction.

When using the Pearson method to evaluate the correlation between two variables, which set of numbers indicates a strong positive correlation? A: The correlation coefficient is -.85 and the P value is 1. B: The correlation coefficient is .85 and the P value is 0.00037. C: The correlation coefficient is .85 and the P value is 0.06. D: The correlation coefficient is -.85 and the P value is 0.00037.	The correlation coefficient is .85 and the P value is 0.00037.
Which of the following forms of exploratory data analysis generates short summaries about the sample and measures of the data? A: Pearson correlation B: Descriptive statistics C: Correlation D: Analysis of variance (ANOVA)	Descriptive statistics
When conducting exploratory data analysis, which visualizations are particularly useful for plotting the target variable over multiple variables to get visual clues of the relationship between these variables and the target. A: Histograms B: Boxplots C: Heatmaps D: Scatter plots	Heatmaps
Which of the following statements about the ANOVA F-test score are true? Select two answers. A: A large F-test score implies a poor correlation between variable categories and the target variable. B: A large F-test score implies a strong correlation between variable categories	A large F-test score implies a strong correlation between variable categories and the target variable. A small F-test score implies a poor cor-

and the target variable. C: A small F-test score implies a strong correlation between variable categories and the target variable. D: A small F-test score implies a poor correlation between variable categories and the target variable.	relation between variable categories and the target variable.
You can visualize the correlation between two variables by plotting them on a scatter plot and then doing which of the following? A: Add a correlation line. B: Add a regression line. C: You should not use a scatter plot for visualizing the correlation between two variables. D: Nothing. The scatter plot alone can show the correlation completely.	Add a regression line.
When using the Pearson method to evaluate the correlation between two variables, how do you know you can have strong certainty in the result? A: The P value is less than 0.05. B: The P value is less than 0.001. C: The P value is greater than 0.1. D: The P value is less than 0.1.	The P value is less than 0.001.
What are the key reasons to develop a model for your data analysis? Select three answers. A: Identify any special structures that may exist in the data. B: Determine the accuracy of your data. C: Understand how the data were generated. D: Determine the relationships between variables.	Identify any special structures that may exist in the data. Understand how the data were generated. Determine the relationships between variables.
There are four assumptions associated with a linear regression model. What	

is the definition of the assumption homoscedasticity? A: The variance of residual is the same for any value of X. B: The relationship between X and the mean of Y is linear. C: Observations are independent of each other. D: For any fixed value of X, Y is normally distributed.	The variance of residual is the same for any value of X.
What step must you take before you can obtain a prediction based on a fitted simple linear regression model? A: Use or create a data frame containing known predictor variables. B: Use or create a data frame containing never seen data. C: Use or create a data frame containing known target variables. D: Do nothing. Once you have a fitted simple linear regression model, you have all you need to make predictions.	Use or create a data frame containing never seen data.
Assume you have a dataset called "new_dataset", two predictor variables called X and Y, and a target variable called Z, and you want to fit a multiple linear regression model. Which command should you use? A: linear_model <- lm(Z ~ X + Y, data = new_dataset) B: linear_model <- lm(X + Y ~ Z, data = new_dataset) C: linear_model <- lm(Z ~ X ~ Y, data = new_dataset) D: linear_model <- lm(X + Y + Z, data = new_dataset)	linear_model <- lm(Z ~ X + Y, data = new_dataset)
Which plot types help you validate assumptions about linearity? Select two	

answers. A: Regression plot B: Q-Q plot C: Residual plot D: Scale-location plot	Regression plot Residual plot
True or False: When using the poly() function to fit a polynomial regression model, you must specify "raw = FALSE" so you can get the expected coefficients. A: True. B: False.	False.
Which performance metric for regression is the mean of the square of the residuals (error)? A: Root mean squared error (RMSE) B: R-squared (R2) C: Mean absolute error (MAE) D: Mean squared error (MSE)	Mean squared error (MSE)
When comparing the MSE of different models, do you want the highest or lowest value of MSE? A: Highest value of MSE B: Lowest value of MSE	Lowest value of MSE
In model development, you can develop more accurate models when you have which of the following? A: Larger quantities of data. B: Relevant data. C: More dependent variables. D: Fewer independent variables.	Relevant data.
Assume you have a dataset called "new_dataset", a predictor variable called X, and a target called Y, and you want to fit a simple linear regression model. Which command should you use? A: linear_model <- lm(Y ~ X, data =	

new_dataset) B: linear_model <- predict(X ~ Y, data = new_dataset) C: linear_model <- lm(X ~ Y, data = new_dataset) D: linear_model <- predict(Y ~ Z, data = new_dataset)	linear_model <- lm(Y ~ X, data = new_dataset)
When using the predict() function in R, what is the default confidence level? A: 90% B: 100% C: 95% D: 85%	95%
Which plot type helps you validate assumptions about normality? A: Scale-location plot B: Residual plot C: Q-Q plot D: Regression plots	Q-Q plot
A third order polynomial regression model is described as which of the following? A: Simple linear regression. B: Squared, meaning that the predictor variable in the model is squared. C: Quadratic, meaning that the predictor variable in the model is squared. D: Cubic, meaning that the predictor variable in the model is cubed.	Cubic, meaning that the predictor variable in the model is cubed.
How should you interpret an R-squared result of 0.89? A: There is a strong negative correlation between the variables. B: 89% of the response variable variation is explained by a linear model. C: The X variable causes the Y variable to positively change 89% of the time. D: 89% of the response variable variation is explained by a polynomial model.	89% of the response variable variation is explained by a linear model.

When comparing linear regression models, when will the mean squared error (MSE) be smaller? A: When using a simple linear regression (SLR) model. B: When using a multiple linear regression (MLR) model. C: This depends on your data. The model that fits the data better has the smaller MSE. D: When using a polynomial regression model.	This depends on your data. The model that fits the data better has the smaller MSE.
When evaluating models, what is the term used to describe a situation where a model fits the training data very well but performs poorly when predicting new data? A: Overfit B: Cross validation C: Underfit D: Small dataset	Overfit
An underfit model is said to have which of the following? A: High generalizability B: High bias C: High variance D: High complexity	High bias
What does regularization introduce into a model that results in a drop in variance? A: Noise B: Feature/variable C: Lambda D: Complexity	Noise
Complete the sentence: When tuning a model, a grid search attempts to find the value of a parameter that has the small-	

est _____. A: Variance B: Error C: Bias D: Lambda	Error
Which situations are helped by using the cross-validation method to train your model? Select two answers. A: Determining if a model can be generalized for a broader group. B: Working with models with small amounts of data. C: Working with models with large amounts of data. D: Working with models that are underfit.	Determining if a model can be generalized for a broader group. Working with models with small amounts of data.
What is a strategy you can employ to address an underfit model? A: Reduce the number of features in the training data. B: Use regularization. C: Increase model complexity. D: Reduce model complexity.	Increase model complexity.
What is the difference between Ridge and Lasso regression? A: Ridge regression penalizes the sum of the absolute values of the coefficients while Lasso regression penalizes the sum of squared coefficients. B: There is no major difference between Ridge and Lasso regression. C: Lasso regression penalizes the sum of the absolute values of the coefficients while Ridge regression penalizes the sum of squared coefficients. D: Lasso regression increases or decreases the value of Lambda to penalize complex models more or less.	Lasso regression penalizes the sum of the absolute values of the coefficients while Ridge regression penalizes the sum of squared coefficients.

Which tidymodels function do you use to create the grid for a grid search? A: tune() B: add_model() C: grid_regular() D: tune_grid()	grid_regular()
Which of the following is NOT a task facilitated by R? A: Data cleaning B: Model development C: Model evaluation D: Data generation	Data generation
Functions contained in packages such as dplyr are used to: A: Identify users of the data set B: Select a data set to use C: Prevent unwanted operations D: Perform common operations	Perform common operations
If you don't ensure that data is stored in the correct format (such as numeric or character), what can happen? A: It will not significantly affect your models. B: You will still be able to make meaningful comparisons. C: Valid data can be treated as missing data. D: Missing data may be imported.	Valid data can be treated as missing data.
Which of the following situations does NOT call for data normalization? A: When the scale of a feature causes it to have a disproportional impact on results B: When there are outliers that might skew the results C: When you want to compare numerical and character values	When you want to compare numerical and character values

D: When different data set features are in very different ranges	
Which of the following is NOT true of a scatter plot? A: It cannot suggest a linear relationship between two variables. B: Each observation is represented as a point. C: The predictor/independent variables are on the x-axis. D: It shows the relationship between two variables.	It cannot suggest a linear relationship between two variables.
What is the purpose of an ANOVA test? A: It helps find correlations between different groups of a categorical variable. B: It is not a useful test except in certain specific cases. C: It helps compare correlating categories in different data sets. D: It determines which variable is most statistically significant.	It helps find correlations between different groups of a categorical variable.
A positive correlation is one in which _____. A: both variables move in opposite directions B: a causative relationship is shown C: both variables move in the same direction D: only one variable moves	both variables move in the same direction
Which of the following is NOT true about a model? A: The more data you have, the more accurate your model will be. B: A model helps predict a value given one or more other values. C: Different types of models may be more accurate in different situations.	Models work by relating independent values.

D: Models work by relating independent values.	
Which is NOT true for comparing multiple linear regression (MLR) and simple linear regression (SLR)? A: The MSE for an MLR model will be smaller than the MSE for an SLR model. B: R2 will have a smaller MSE. C: A lower mean squared error (MSE) always implies a better fit. D: Polynomial regression will have a smaller MSE than regular regression.	A lower mean squared error (MSE) always implies a better fit.
A training set is _____. A: multiple data sets that have been run on the model B: a selected portion of the data set that is known to function well within the model C: a small portion of the data used to evaluate the performance of a model D: a large portion of a data set that is used to build a sound model	a large portion of a data set that is used to build a sound model
Which chart is a type of correlation chart? A: Histogram B: Scatterplot C: Pie chart D: Bar chart	Scatterplot
Which R statement creates a chart object based on the data frame "salesdata", but allows you to vary the aesthetics from one layer to another? A: qplot() B: ggplot() C: ggplot(salesdata, aes(x = feature1, y = feature2)) D: ggplot(salesdata)	ggplot(salesdata)

True or False: The qplot() function has no defaults so you have more control over the output. A: True B: False	False
Which R packages will this course use to create data visualizations? Select two answers. A: qplot B: None, you will use base R C: Leaflet D: ggplot2	Leaflet ggplot2
Which chart is a type of part to the whole chart? A: Horizontal bar chart B: Grouped bar chart C: Bar chart D: Stacked bar chart	Stacked bar chart
Which ggplot2 function can create a complete plot given the data, mappings, and geom as parameters? A: ggplot() B: ggplot2() C: qplot() D: geom()	qplot()
True or False: Numeric data can be qualitative or quantitative. A: True B: False	True
Which of the following statements about histograms is true? Select two answers. A: A histogram displays quantitative data, while a bar chart displays qualitative data. B: A histogram divides data into bins and then counts the number of times a data	A histogram displays quantitative data, while a bar chart displays qualitative data. A histogram divides data into bins and

point falls into each bin. C: A histogram displays qualitative data, while a bar chart displays quantitative data. D: A histogram counts the frequency of each individual number in the data set.	then counts the number of times a data point falls into each bin.
Complete the sentence: A pie chart is the same as a _____ in polar coordinates. A: Stacked bar chart B: Bar chart C: Grouped bar chart D: Horizontal bar chart	Stacked bar chart
Which parameter of the qplot() function changes the border color of the bars in a bar chart to blue? A: fill = I("blue") B: colour = I("blue") C: border = I("blue") D: outline = I("blue")	colour = I("blue")
How can you improve the smoothness of a histogram? A: Reduce the number of bins to increase the bin width. B: Changing the number of bins has no impact of the smoothness of the histogram. C: Always go with the default number of bins. D: Increase the number of bins to reduce the bin width.	Reduce the number of bins to increase the bin width.
What step must you take before you can add the coord_polar() function to ggplot() to create a pie chart? A: Add the geom_bar(position = "stack") command to the ggplot()	

function. B: Add the geom_bar(position = "dodge") command to the ggplot() function. C: Set the x argument of the aes() function used in the ggplot() function to the factor. D: Add the geom_circle() command to the ggplot() function.	Add the geom_bar(position = "stack") command to the ggplot() function.
True or False: A scatter plot can only show how two variables relate to each other across the points of the dataset. A: True B: False	False
When you create a line plot using multiple geom_line() statements, the y axis label reflects the y variable for which geom_line() statement? A: The last geom_line() statement. B: The first geom_line() statement. C: None of the geom_line() statements are used for the y-axis label. D: You specify this using an argument of the geom_line() function.	The first geom_line() statement.
Which plot type summarizes the distribution of sorted numerical data? A: Line plots B: Scatter plots C: Bar charts D: Box plots	Box plots
In a scatter plot, what is the best way to change the color of the points based on a categorical variable? A: Assign the variable to the "color" argument of the aes() function within the ggplot() function. B: Assign the variable to the "color" argu-	

ment of the geom_point() function. C: Convert the categorical variable to a factor and then assign it to the "color" argument of the aes() function within the ggplot() function. D: Convert the categorical variable to a factor and then assign it to the "color" argument of the geom_point() function.	Convert the categorical variable to a factor and then assign it to the "color" argument of the aes() function within the ggplot() function.
Which plot type helps you visualize time series data? A: Box plots B: Line plots C: Histograms D: Scatter plots	Line plots
In a box plot, in which quartile does 50% of the sorted data fall below? A: First quartile B: Second quartile C: Third quartile D: Fourth quartile	Second quartile
Which functions can you use to change the title of a plot? Select two answers. A: ylab() B: ggtitle() C: xlab() D: labs()	ggtitle() labs()
What can you add to a plot if you want to emphasize important elements, such as outliers or spikes in your data? A: Axis label B: Annotation C: Theme D: Plot title	Annotation
You want to divide a plot into subplots based on a categorical variable called "quarters". Which function should	

you add to ggplot() to do this? A: facet(quarters) B: facet(~quarters) C: facet_wrap(quarters) D: facet_wrap(~quarters)	facet_wrap(~quarters)
What information do you need to provide to create a visualization using the Leaflet library? A: The date associated with each data point. B: The percentage that each data point represents. C: Two items to compare to each other. D: The latitude and longitude of each data point.	The latitude and longitude of each data point.
You added text labels to the data points on your plot, but now the plot looks messy because there are so many of them. What should you do? A: Set the check_overlap parameter of geom_text() to FALSE. B: Set the check_overlap parameter of geom_text() to TRUE. C: Set the overlap parameter of geom_text() to FALSE. D: Set the overlap parameter of geom_text() to TRUE.	Set the check_overlap parameter of geom_text() to TRUE.
If you do not specify a theme when creating a plot with ggplot2, which theme does it use by default? A: theme_light() B: theme_gray() C: theme_classic() D: theme_minimal()	theme_gray()
Using themes, you can change the colors and styles of the borders, backgrounds, lines, and text on a plot. What should you do if you want to	

completely remove one of these elements from the theme? A: Assign the element.delete() function to the element. B: Assign the element.empty() function to the element. C: Assign the element.blank() function to the element. D: Assign the element.remove() function to the element.	Assign the element.blank() function to the element.
In a Leaflet map, which two statements describe the difference between the addCircles() and addCircleMarkers() functions? A: Markers created with addCircles() can be rescaled. B: Markers created with addCircleMarkers() remain a constant size. C: Markers created with addCircleMarkers() can be rescaled. D: Markers created with addCircles() remain a constant size.	Markers created with addCircles() can be rescaled. Markers created with addCircleMarkers() remain a constant size.
Which two components of a dashboard happen on the back end? A: Serve B: Visualize C: Analyze D: Interact	Serve Analyze
Which is the preferred method for creating a Shiny app? A: One file for all server and UI code. B: Both methods work the same so one is not preferred over the other. C: Separate code into server.R and ui.R files.	Separate code into server.R and ui.R files.
True or False: You must include the shinyApp() function in the code for all Shiny apps.	True

A: True B: False	
Which function creates an empty layout? A: sidebarLayout() B: fluidPage() C: shinyUI() D: mainPanel()	fluidPage()
True or False: A Shiny app consists of two parts, the server that the user interacts with and the UI that powers the app. A: True B: False	False
Which two components of a dashboard happen on the front end? A: Serve B: Interact C: Analyze D: Visualize	Interact Visualize
Complete the sentence: You use the Layout functions to organize _____ containing user interface elements in the application. A: Layouts B: Inputs C: Panels D: Outputs	Panels
When defining the server logic for a Shiny app, you define a function that includes which of the following parameters? A: input, response B: input, output C: input, renderPlot D: input, plotOutput	input, output
If you have the command plotOutput("plot_histogram") in the UI-side	

code in your Shiny application, what is the name of the variable that you assign the plot to in the server-side code? A: output(plot_histogram) B: output$plot_histogram C: input$plot_histogram D: plot_histogram	output$plot_histogram
Can you publish a Shiny app to your shinyapps.io account from RStudio? A: No, you cannot publish a Shiny app to shinyapps.io from RStudio. B: Yes, but only if you install the rscon-nect package first. C: Yes, you can do this using the built-in capabilities of RStudio. D: Yes, but only if you install the shiny package first.	Yes, but only if you install the rsconnect package first.
What is the process to convert an R Markdown file to an HTML, PDF, or Microsoft Word document? A: Join B: Weave C: Knit D: Append	Knit
In a Shiny application, where do you add input widgets? A: A panel. B: A tabset panel. C: A layout. D: A title panel.	A panel.
Which deployment method should you select for your Shiny app if you do not want to run your own server? A: Shiny Server B: shinyapps.io C: RStudio Connect D: None of these options	shinyapps.io

What are the two main differences between an R Markdown document and a Shiny dashboard? A: A dashboard can be interactive, while an R Markdown document is static. B: A dashboard can contain text, images, plots, and other information, while an R Markdown document contains only easy-to-write plain text. C: A dashboard always reflects current data, while an R Markdown document produces a snapshot of the data at the time the report is generated. D: A dashboard is reusable, while an R Markdown file can only be generated once.	A dashboard can be interactive, while an R Markdown document is static. A dashboard always reflects current data, while an R Markdown document produces a snapshot of the data at the time the report is generated.
Which chart type shows the inner subdivision of a value among different categories or groups? A: Correlation charts B: Part to whole charts C: Relationship charts D: Comparison charts	Part to whole charts
Which chart is a type of comparison chart? A: Histogram B: Scatterplot C: Stacked bar chart D: Bar chart	Bar chart
Which data visualization displays information about the distribution of a population? A: Pie Chart B: Histogram C: Scatter Plot D: Line Plot	Histogram

Which package should be used to create a scatter plot? A: ggplot B: geom_plot C: ggplot2 D: qplot	ggplot2
Which statement is true regarding box plots? A: It displays categorical data. B: It divides the data set into quartiles. C: It obscures the outliers in the data set. D: It displays the mean of the data set.	It divides the data set into quartiles.
Which statement best describes facets? A: Facets are a type of trend chart. B: Facets subdivide an ordinal data set into panels based on a categorical or discrete variable. C: Facets can only be used with histograms. D: Facets are used to display ordinal data.	Facets subdivide an ordinal data set into panels based on a categorical or discrete variable.
Which function will return an object that represents a world map? A: leaflet.map() B: leaflet() C: addTiles() D: addMap()	leaflet()
Which of the following is a true statement with regards to functions found in the Shiny library? A: splitLayout() function lays out elements vertically, dividing the available vertical space into equal parts. B: flowLayout() arranges elements in a layout with a side bar and main area. C: fluidRow() creates a fluid page layout, which consists of rows that in turn include columns.	fluidRow() creates a fluid page layout, which consists of rows that in turn include columns.

D: The sidebarLayout() function scales components in real time to fill the entire browser width.	
Which function is used to construct an initial empty UI when creating a Shiny app? A: fluidPage() B: titlePanel() C: sidebarLayout() D: mainPanel()	fluidPage()
Which of the following is NOT a parameter to the varSelectInput() function? A: The data frame from which the column names will be retrieved B: The input ID that is used to access the value C: The data set variable name D: The display label for the control	The data set variable name