# Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

# Chapter 1: Preparing Data for Analysis and Visualization in R: The R-Team and the Pot Policy Problem

## Test Bank

### Multiple Choice

1. What are two benefits of using R?

A. It's free and open source.

B. It's inexpensive and open source.

C. It's free and includes responsive customer service.

D. It's open source and a single platform.

Ans: A

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Choosing and Learning R

Difficulty Level: Easy

2. What does "open source" mean?

A. The source code is locked, and you cannot change anything.

B. "Open" is the function for reading in source files.

C. the command to open R on your computer

D. Anyone with an idea of how to contribute open source material can do so.

Ans: D

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Choosing and Learning R

Difficulty Level: Medium

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

3. How many R packages are available on the CRAN (The Comprehensive R Archive Network)?

A. 150

B. 1,000

C. more than 10,000

D. millions

Ans: C

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Choosing and Learning R

Difficulty Level: Easy

4. Which of the following best describes a package in R?

A. a list of R tutorials for beginners

B. a matrix of complicated data

C. a script of code to produce publishable graphs

D. a collection of R functions, data, and code to solve specific questions

Ans: D

Achievement: 1-1: Observations and variables.

Cognitive Domain: Comprehension

Answer Location: Choosing and Learning R

Difficulty Level: Medium

5. For a survey that asks a group of people the amount of their income and whether they purchase organic food, each person is a(n) _____ .

A. variable

B. observation

C. function

D. object

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Ans: B

Achievement: 1-1: Observations and variables.

Cognitive Domain: Application

Answer Location: Defining Observations and Variables

Difficulty Level: Easy

6. For a survey that asks a group of people the amount of their income and whether they purchase organic food, income, and their purchasing behavior are two _____ .

A. variables

B. observations

C. functions

D. objects

Ans: A

Achievement: 1-1: Observations and variables.

Cognitive Domain: Application

Answer Location: Defining Observations and Variables

Difficulty Level: Easy

7. What are the key concepts in data science?

A. observations and variables

B. graphs and functions

C. R and the packages in R

D. computational methods

Ans: A

Achievement: 1-1: Observations and variables.

Cognitive Domain: Comprehension

Answer Location: Defining Observations and Variables

Difficulty Level: Easy

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

8. R stores information as _____.

A. values

B. matrices

C. objects

D. packages

Ans: C

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Entering and Storing Variables in R

Difficulty Level: Easy

9. In R, assigning an object to a name is done by using _____.

A. name <- object

B. name -> object

C. name :> object

D. name <= object

Ans: A

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Entering and Storing Variables in R

Difficulty Level: Easy

10. Which of the following tabs in R holds all of the code one has run?

A. Environment

B. History

C. Console

D. Jobs

Ans: B

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Entering and Storing Variables in R

Difficulty Level: Easy

11. In R, you can comment by using the sign _____.

A. *

B. #

C. //

D. %

Ans: B

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Knowledge

Answer Location: Using Comments to Organize and Explain Code

Difficulty Level: Easy

12. Reproducibility of a study means _____.

A. the data are the same

B. the statistical methods are valid

C. you produce the same conclusion as other studies

D. you get the same results from reanalyzing the data

Ans: D

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Using Comments to Organize and Explain Code

Difficulty Level: Medium

13. In R, the text following the # sign on the same line means the text _____.

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

A. will not be computed

B. is important

C. is shared across script files

D. is a function

Ans: A

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Using Comments to Organize and Explain Code

Difficulty Level: Easy

14. Which of the following describes a prolog?

A. It is a list of functions used in the script file.

B. It is generally at the end of the code file to document the results.

C. It is a set of comments at the top of a code file that provide information about what is in the code file.

D. It is the license file to run the code file.

Ans: C

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Including a Prolog to Introduce a Script File

Difficulty Level: Easy

15. Which of the following is a good option to name a code file?

A. project_author.R

B. author_result.R

C. date_result.R

D. date_project_author.R

Ans: D

Achievement: 1-2: Using reproducible research practices.

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Cognitive Domain: Comprehension

Answer Location: Including a Prolog to Introduce a Script File

Difficulty Level: Easy

16. What are class() and rm() examples of?

A. objects

B. arguments

C. packages

D. functions

Ans: D

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Naming Functions

Difficulty Level: Medium

17. What does a function act on?

A. library

B. argument

C. package

D. help

Ans: B

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Knowledge

Answer Location: Naming Functions

Difficulty Level: Easy

18. Which of the following data types is similar to numeric but only contains whole numbers?

A. character

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

B. integer

C. logical

D. factor

Ans: B

Achievement: 1-3: Understanding and changing data types.

Cognitive Domain: Knowledge

Answer Location: Integer Data Type

Difficulty Level: Easy

19. Which of the following data types only includes the values of TRUE and FALSE?

A. character

B. integer

C. logical

D. factor

Ans: C

Achievement: 1-3: Understanding and changing data types.

Cognitive Domain: Knowledge

Answer Location: Logical Data Type

Difficulty Level: Easy

20. Which of the following data types includes the data that cannot be logically included in calculations (such as zip code and letters)?

A. character

B. integer

C. numeric

D. factor

Ans: A

Achievement: 1-3: Understanding and changing data types.

Cognitive Domain: Knowledge

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Answer Location: Character Data Type

Difficulty Level: Easy

21. Variables measured in categories are _____ data type

A. character

B. integer

C. numeric

D. factor

Ans: D

Achievement: 1-3: Understanding and changing data types.

Cognitive Domain: Comprehension

Answer Location: Factor Data Type

Difficulty Level: Medium

22. A vector in R is _____.

A. a list of only numerical data

B. a group of functions

C. a package with summary of results

D. a set of data elements that are saved together as the same type

Ans: D

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Achievement 4: Entering or Loading Data into R

Difficulty Level: Easy

23. The c() function stands for _____.

A. concatenate

B. conclude

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

C. complete

D. contrast

Ans: A

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Creating Vectors for Different Data Types

Difficulty Level: Medium

24. Which of the following codes combines numbers 1, 10, and 100 into a single vector?

A. t(1, 10, 100)

B. c(1, 10, 100)

C. rm(1, 10, 100)

D. vec(1, 10, 100)

Ans: B

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Application

Answer Location: Creating Vectors for Different Data Types

Difficulty Level: Medium

25. For the data frame format, _____.

A. it can be only imported from Excel spreadsheet

B. it contains only numerical data type

C. the rows are observations and the columns are variables

D. no more than one data frame can be saved in the environment at the same time

Ans: C

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Comprehension

Answer Location: Creating a Data Frame

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Difficulty Level: Medium

26. Which of the following codes specifies the variable named “income” within the data frame named “survey”?

A. income%survey

B. survey!income

C. income#survey

D. survey$income

Ans: D

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Creating a Data Frame

Difficulty Level: Medium

27. Once a package is installed, you can open and then use the package with _____.

A. open()

B. library()

C. c()

D. package()

Ans: B

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Importing a Comma Separated Values (CSV) File.

Difficulty Level: Easy

28. In R, missing values are recorded as _____.

A. NO

B. NA

C. MV

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

D. MIS

Ans: B

Achievement: 1-5: Identifying and treating missing values.

Cognitive Domain: Knowledge

Answer Location: Achievement 5: Identifying and Treating Missing Values

Difficulty Level: Easy

29. Which of the following codes refers to the pipe feature available in packages such as tidyverse?

A. %>%

B. #>#

C. $>$

D. @>@

Ans: A

Achievement: 1-5: Identifying and treating missing values.

Cognitive Domain: Comprehension

Answer Location: Recoding Missing Values to NA.

Difficulty Level: Medium

30. From the code below, _____ is the data frame, and _____ is the variable.

“

```
gss.2016.cleaned <- gss.2016 %>%
  mutate(grass = as.factor(grass)) %>%
  mutate(grass = na_if(x = grass, y = "DK"))
```

“

A. gss.2016.clean; DK

B. grass; gss.2016

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

C. gss.2016; DK

D. gss.2016; grass

Ans: D

Achievement: 1-5: Identifying and treating missing values.

Cognitive Domain: Application

Answer Location: Recoding Missing Values to NA.

Difficulty Level: Hard

31. To produce custom graphs in R, you can use the _____ package.

A. plot_verse

B. ggplot2

C. tidyplot

D. plot

Ans: B

Achievement: 1-6: Building a basic bar chart.

Cognitive Domain: Knowledge

Answer Location: Achievement 6: Building a Basic Bar Chart

Difficulty Level: Easy

**True/False**

1. Since R is open source, everyone can write packages in the R language.

Ans: T

Achievement: 1-1: Observations and variables.

Cognitive Domain: Knowledge

Answer Location: Choosing and Learning R

Difficulty Level: Easy

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

2. You have to install the packages in R every time before using it.

Ans: F

Achievement: 1-1: Observations and variables.

Cognitive Domain: Comprehension

Answer Location: Importing a Comma Separated Values (CSV) File. | Nancy's fancy code: working with R packages

Difficulty Level: Medium

3. You can save the R code written in the Console from the History pane.

Ans: F

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Using Comments to Organize and Explain Code

Difficulty Level: Medium

4. When naming objects in R, it is better to use words and abbreviations that are common mathematical terms so that others understand the meanings of the object.

Ans: F

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Naming objects

Difficulty Level: Medium

5. You can create new objects and print them in console by adding parentheses around the code that creates the object.

Ans: T

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Creating Vectors for Different Data Types

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Difficulty Level: Easy

6. You can use dots (.) as part of the variable names in R.

Ans: T

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Knowledge

Answer Location: Creating Vectors for Different Data Types

Difficulty Level: Easy

7. NA, Na, and na are all reserved terms in R to indicate missing values.

Ans: F

Achievement: 1-5: Identifying and treating missing values.

Cognitive Domain: Knowledge

Answer Location: Achievement 5: Identifying and Treating Missing Values

Difficulty Level: medium

8. drop_na() is one of the commands to remove the missing values from the graph.

Ans: T

Achievement: 1-6: Building a basic bar chart.

Cognitive Domain: Knowledge

Answer Location: Omitting NA From a Graph

Difficulty Level: Easy

9. The graphs produced with ggplot2 have a very limited set of color choices.

Ans: F

Achievement: 1-6: Building a basic bar chart.

Cognitive Domain: Knowledge

Answer Location: Working With Color in a Bar Chart

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

Difficulty Level: Easy

10. The “gg” in ggplot2 stands for the “grammar of graphics.”

Ans: T

Achievement: 1-6: Building a basic bar chart.

Cognitive Domain: Knowledge

Answer Location: Achievement 6: Building a Basic Bar Chart

Difficulty Level: Medium

**Essay**

1. What are the main benefits of using R compared to other traditional statistics software options?

Ans: First, R is free while most other software options can be costly. Second, R is open source which results in an extensive number of contributed packages. Third, the R community is active and inclusive.

Achievement: 1-1: Observations and variables.

Cognitive Domain: Comprehension

Answer Location: Choosing and Learning R

Difficulty Level: Medium

2. What are the benefits of writing R code in the Script file?

Ans: You can edit and save code by writing in the Script file. You can also write short explanation in the script while coding in the script file. You can share the script file with others easily.

Achievement: 1-2: Using reproducible research practices.

Cognitive Domain: Comprehension

Answer Location: Using Comments to Organize and Explain Code

Difficulty Level: Medium

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

3. When does a namespace conflict occur and how do you solve it?

Ans: The namespace conflict occurs when there are function names that are the same in two different packages and both packages are opened at the same time in the R file. Thus, R cannot decide which function to use. You can solve the namespace conflict by specifying one of the packages using the code "::".

Achievement: 1-4: Entering or loading data into R.

Cognitive Domain: Comprehension

Answer Location: Importing a Comma Separated Values (CSV) File.

Difficulty Level: Medium

4. Please explain the logic behind the code below. (Hint: What is the main output of the code? Which package does the code use? Why are there more than one line of code?)

"

```
ggplot(aes(x = grass)) +
  geom_bar()
```

"

Ans: The code is used to build graph with ggplot2 package. The graph is built with layers with ggplot() function. The first line of code is the first layer of the graph, which contains the basic information and which variable to include (variable "grass" in this case). The second line of the code gives the geometry of this graph, which geom_bar() means a bar graph. The second line of the code acts as the second layer of the plot and is added with a plus sign.

Achievement: 1-6: Building a basic bar chart.

Cognitive Domain: Analysis

Answer Location: Achievement 6: Building a Basic Bar Chart

Difficulty Level: Hard

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

# Chapter 1

## Solutions to end-of-chapter multiple choice questions

*__Note__: answers are marked with a yellow asterisk*

Q1: Which R data type is most appropriate for a categorical variable?

a. Numeric

*b. Factor

c. Integer

d. Character

Q2: Which of the following opens the **`ggplot2`** library?

a. `install.packages("ggplot2")`

*b. `library(package = "ggplot2")`

c. `summary(object = ggplot2)`

d. `open(x = ggplot2)`

Q3: The block of text at the top of a code file that introduces the project is called

a. library.

b. summary.

*c. prolog.

d. pane.

Q4: In a data frame containing information on the age and height of 100 people, the people are the _____ and age and height are the _____.

*a. observations, variables

b. variables, observations

c. data, factors

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris

d. factors, data

Q5: The results of running R code show in which pane?

a. Source

b. Environment

c. History

*d. Console

Access Complete Test Bank for Statistics With R Solving Problems Using Real-World Data 1st Harris