

Challenge-3

Guan Ziwen

2023-08-30

I. Questions

Question 1: Emoji Expressions Imagine you're analyzing social media posts for sentiment analysis. If you were to create a variable named "postSentiment" to store the sentiment of a post using emojis (for positive, for neutral, for negative), what data type would you assign to this variable? Why? (*narrative type question, no code required*)

Solution:

As the data (variable) collected do not contain numerical values and have a natural ordering, I will classify it as ordinal variable. I will assign 'character' data type to this variable.

Question 2: Hashtag Havoc In a study on trending hashtags, you want to store the list of hashtags associated with a post. What data type would you choose for the variable "postHashtags"? How might this data type help you analyze and categorize the hashtags later? (*narrative type question, no code required*)

Solution:

As the data (variable) collected do not contain numerical values and do not have a natural ordering, I will classify it as nominal variable. I will assign 'character' data type to this variable. Each hashtag will contain a specific meaning and can be categoric on its own. By assigning character data type to hashtags, I can easily retrieve information by looking at the variable values alone.

Question 3: Time Traveler's Log You're examining the timing of user interactions on a website. Would you use a numeric or non-numeric data type to represent the timestamp of each interaction? Explain your choice (*narrative type question, no code required*)

Solution:

As timestamp is discrete values that contain ':' character that is not purely numerical, I will assign 'character' data type to this variable.

Question 4: Event Elegance You're managing an event database that includes the date and time of each session. What data type(s) would you use to represent the session date and time? (*narrative type question, no code required*)

Solution:

As dates are discrete numerical values, I will assign 'integer' data type to represent the session date. As time of each session is discrete and may contain ':' character other than numbers, I will assign 'character' data type to represent the session time.

Question 5: Nominee Nominations You're analyzing nominations for an online award. Each participant can nominate multiple candidates. What data type would be suitable for storing the list of nominated candidates for each participant? (*narrative type question, no code required*)

Solution:

As the nominated names (variable) collected do not contain numerical values and do not have a natural ordering, I will classify it as nominal variable. I will assign ‘character’ data type to this variable.

Question 6: Communication Channels In a survey about preferred communication channels, respondents choose from options like “email,” “phone,” or “social media.” What data type would you assign to the variable “preferred Channel”? (*narrative type question, no code required*)

Solution:

As the options do not contain numerical values and do not have a natural ordering, data (variable) collected will be nominal and non-numerical. I will assign ‘character’ data type to the variables.

Question 7: Colorful Commentary In a design feedback survey, participants are asked to describe their feelings about a website using color names (e.g., “warm red,” “cool blue”). What data type would you choose for the variable “feedbackColor”? (*narrative type question, no code required*)

Solution:

As the feedback consist of purely descriptions and do not contain numerical values, data (variable) collected will be nominal and non-numerical. I will assign ‘character’ data type to the variables.

Question 8: Variable Exploration Imagine you’re conducting a study on social media usage. Identify three variables related to this study, and specify their data types in R. Classify each variable as either numeric or non-numeric.

Solution:

Variable 1: Daily duration of social media usage. Data type - ‘double’ Numeric

Variable 2: Social media platforms engaged. Data type - ‘character’ Non-numeric

Variable 3: Number of exposures to violent content per day. Data type - ‘integer’ Numeric

Question 9: Vector Variety Create a numeric vector named “ages” containing the ages of five people: 25, 30, 22, 28, and 33. Print the vector.

Solution:

```
# Enter code here
ages <- c(25L,30L,22L,28L,33L)
ages
```

```
## [1] 25 30 22 28 33
```

Question 10: List Logic Construct a list named “student_info” that contains the following elements:

- A character vector of student names: “Alice,” “Bob,” “Catherine”
- A numeric vector of their respective scores: 85, 92, 78
- A logical vector indicating if they passed the exam: TRUE, TRUE, FALSE

Print the list.

Solution:

```
# Enter code here
```

```
student_info <- list(names = c("Alice","Bob","Catherine"), scores = c(85,92,78), logical_info = c(TRUE,TRUE,FALSE))
student_info
```

```
## $names
## [1] "Alice"      "Bob"        "Catherine"
##
## $scores
## [1] 85 92 78
##
## $logical_info
## [1] TRUE TRUE FALSE
```

Question 11: Type Tracking You have a vector “data” containing the values 10, 15.5, “20”, and TRUE. Determine the data types of each element using the `typeof()` function.

Solution:

```
# Enter code here
```

```
typeof(10)
```

```
## [1] "double"
```

```
typeof(15.5)
```

```
## [1] "double"
```

```
typeof("20")
```

```
## [1] "character"
```

```
typeof(TRUE)
```

```
## [1] "logical"
```

Question 12: Coercion Chronicles You have a numeric vector “prices” with values 20.5, 15, and “25”. Use explicit coercion to convert the last element to a numeric data type. Print the updated vector.

Solution:

```
# Enter code here
```

```
prices <- c(20.5,15,"25")
prices <- as.numeric(prices)
prices
```

```
## [1] 20.5 15.0 25.0
```

Question 13: Implicit Intuition Combine the numeric vector `c(5, 10, 15)` with the character vector `c("apple", "banana", "cherry")`. What happens to the data types of the combined vector? Explain the concept of implicit coercion.

Solution:

```
# Enter code here
numeric_vector <- c(5,10,15)
character_vector <- c("apple", "banana", "cherry")
combined_vector <- c(numeric_vector, character_vector)
typeof(combined_vector)
```

```
## [1] "character"
```

The data type of the combined vector was converted from numeric and character to character only. As vector is a collection of elements of the same type, R will automatically convert the vector to one data type when the input is of mixed data type. When R converts the data type based on the vector's content automatically, implicit coercion occurs.

Question 14: Coercion Challenges You have a vector “numbers” with values 7, 12.5, and “15.7”. Calculate the sum of these numbers. Will R automatically handle the data type conversion? If not, how would you handle it?

Solution:

```
# Enter code here
numbers <- c(7,12.5,"15.7")
sum(numbers)
```

Error in `sum()`: ! invalid ‘type’ (character) of argument

R will not automatically handle the data type conversion. I will then convert the character data type to numeric data type to facilitate calculation.

```
numbers <- as.numeric(c(7,12.5,"15.7"))
sum(numbers)
```

```
## [1] 35.2
```

Question 15: Coercion Consequences Suppose you want to calculate the average of a vector “grades” with values 85, 90.5, and “75.2”. If you directly calculate the mean using the `mean()` function, what result do you expect? How might you ensure accurate calculation?

Solution:

I will expect an error in `mean()` due to invalid ‘type’(character) of argument. To ensure accurate calculation, I will convert the character data type to numeric data type first.

```
# Enter code here
grades <- as.numeric(c(85,90.5,"75.2"))
mean(grades)
```

```
## [1] 83.56667
```

Question 16: Data Diversity in Lists Create a list named “mixed_data” with the following components:

- A numeric vector: 10, 20, 30
- A character vector: “red”, “green”, “blue”
- A logical vector: TRUE, FALSE, TRUE

Calculate the mean of the numeric vector within the list.

Solution:

```
# Enter code here
numeric_vector <- c(10,20,30)
character_vector <- c("red","green","blue")
logical_vector <- c(TRUE,FALSE,TRUE)
mixed_data <- list(numeric_vector,character_vector,logical_vector)
mixed_data
```

```
## [[1]]
## [1] 10 20 30
##
## [[2]]
## [1] "red" "green" "blue"
##
## [[3]]
## [1] TRUE FALSE TRUE
```

```
mean(mixed_data[[1]])
```

```
## [1] 20
```

Question 17: List Logic Follow-up Using the “student_info” list from Question 10, extract and print the score of the student named “Bob.”

Solution:

```
# Enter code here
student_info$scores[student_info$names=="Bob"]
```

```
## [1] 92
```

Question 18: Dynamic Access Create a numeric vector values with random values. Write R code to dynamically access and print the last element of the vector, regardless of its length.

Solution:

```
# Enter code here
x <- c(2,7,9,13,6)
tail(x,n=1)
```

```
## [1] 6
```

Question 19: Multiple Matches You have a character vector `words <- c("apple", "banana", "cherry", "apple")`. Write R code to find and print the indices of all occurrences of the word "apple."

Solution:

```
# Enter code here
words = c("apple", "banana", "cherry", "apple")
which(words == "apple")
```

```
## [1] 1 4
```

Question 20: Conditional Capture Assume you have a vector `ages` containing the ages of individuals. Write R code to extract and print the ages of individuals who are older than 30.

Solution:

```
# Enter code here
ages <- c(13,79,56,37,17)
ages[ages>30]
```

```
## [1] 79 56 37
```

Question 21: Extract Every Nth Given a numeric vector `sequence <- 1:20`, write R code to extract and print every third element of the vector.

Solution:

```
# Enter code here
#seq()
sequence <- 1:20
extracted_sequence <- sequence[seq(0,length(sequence),3)]
extracted_sequence
```

```
## [1] 3 6 9 12 15 18
```

Question 22: Range Retrieval Create a numeric vector `numbers` with values from 1 to 10. Write R code to extract and print the values between the fourth and eighth elements.

Solution:

```
# Enter code here
x <- 1:10
x[5:7]
```

```
## [1] 5 6 7
```

Question 23: Missing Matters Suppose you have a numeric vector `data <- c(10, NA, 15, 20)`. Write R code to check if the second element of the vector is missing (NA).

Solution:

```
# Enter code here
data <- c(10, NA, 15, 20)
is.na(data[2])
```

```
## [1] TRUE
```

Question 24: Temperature Extremes Assume you have a numeric vector `temperatures` with daily temperatures. Create a logical vector `hot_days` that flags days with temperatures above 90 degrees Fahrenheit. Print the total number of hot days.

Solution:

```
# Enter code here
temperatures <- c(35,70,95,15,93,99,23)
hot_days <- temperatures > 90
length(hot_days[hot_days == TRUE])
```

```
## [1] 3
```

Question 25: String Selection Given a character vector `fruits` containing fruit names, create a logical vector `long_names` that identifies fruits with names longer than 6 characters. Print the long fruit names.

Solution:

```
# Enter code here
fruits <- c("apple","grape","watermelon","honeydew")
name_length <- nchar(fruits)
long_names <- name_length > 6
fruits[long_names==TRUE]
```

```
## [1] "watermelon" "honeydew"
```

Question 26: Data Divisibility Given a numeric vector `numbers`, create a logical vector `divisible_by_5` to indicate numbers that are divisible by 5. Print the numbers that satisfy this condition.

Solution:

```
# Enter code here
numbers <- c(26,30,70,35,66,44)
divisible_by_5 <- as.logical(numbers%%5==0)
numbers[divisible_by_5]
```

```
## [1] 30 70 35
```

Question 27: Bigger or Smaller? You have two numeric vectors `vector1` and `vector2`. Create a logical vector `comparison` to indicate whether each element in `vector1` is greater than the corresponding element in `vector2`. Print the comparison results.

Solution:

```
# Enter code here
vector1 <- c(13,52,24,23,7)
vector2 <- c(2,67,12,33,4)
comparison <- (vector1 > vector2)
comparison

## [1]  TRUE FALSE  TRUE FALSE  TRUE
```