

Course 2 Weekly Quiz 1

Zhenzhuo A. C.

25.03.22

Question 3: In R the following are all *atomic data* types EXCEPT:

- **Answer:** Array, list, data frame, matrix, table
- **PS:** use class() function can help you identify the data type

Question 4: If I execute the expression `x <- 4` in R, what is the class of the object ``x``?

```
x <- 4
class(x)
```

```
## [1] "numeric"
```

- **Answer:** numeric

Question 5: What is the class of the object defined by the expression `x <- c(4, "a", TRUE)`?

```
x <- c(4, "a", TRUE)
class(x)
```

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

- **Answer:** character

Question 6: If I have two vectors `x <- c(1,3, 5)` and `y <- c(3, 2, 10)`, what is produced by the expression `rbind(x, y)`?

```
x <- c(1,3, 5)
y <- c(3, 2, 10)
rbind(x,y)
```

```
##      [,1] [,2] [,3]
## x      1   3   5
## y      3   2  10
```

- **Answer:** This is a 2*3 matrix.

Additionally, with `cbind(x,y)`

```
x <- c(1,3, 5)
y <- c(3, 2, 10)
cbind(x,y)
```

```
##      x  y
## [1,] 1  3
## [2,] 3  2
## [3,] 5 10
```

- **Answer:** We can get a 3*2 matrix.

Question 7: A key property of vectors in R is that

- **Answer:** Elements of a vector all must be of the same class

Question 8: Suppose I have a list defined as `x <- list(2, "a", "b", TRUE)`. What does `x[[2]]` give me? Select all that apply.

```
x <- list(2, "a", "b", TRUE)
x[[2]]
```

```
## [1] "a"
```

- **Answer:** A character vector containing the letter "a".

Question 9: Suppose I have a vector `x <- 1:4` and a vector `y <- 2`. What is produced by the expression `x + y`?

```
x <- 1:4
y <- 2
x+y
```

```
## [1] 3 4 5 6
```

- **Answer:** 1:4 means 1 or 2 or 3 or 4. The answer is a numeric vector with elements 3, 4, 5, 6.A numeric vector with elements 3, 4, 5, 6.

Question 10: Suppose I have a vector `x <- c(17, 14, 4, 5, 13, 12, 10)` and I want to set all elements of this vector that are greater than 10 to be equal to 4. What R code achieves this? Select all that apply.

```
x <- c(17, 14, 4, 5, 13, 12, 10)
x[x > 10] <- 4
```

- **Answer:** assign for under the requirement given in []

Question 12: Extract the first 2 rows of the data frame and print them to the console. What does the output look like?

```
dataset1 <- read.csv("/Users/zhenzhuo.chen/Desktop/hw1_data.csv")
dataset1[1:2,]
```

| | Ozone <int> | Solar.R <int> | Wind <dbl> | Temp <int> | Month <int> | Day <int> |
|---|----------------|------------------|---------------|---------------|----------------|--------------|
| 1 | 41 | 190 | 7.4 | 67 | 5 | 1 |
| 2 | 36 | 118 | 8.0 | 72 | 5 | 2 |

2 rows

- **PS:** See this example for extracting specific columns from a dataset. Do not forget the comma.

Question 13:How many observations (i.e. rows) are in this data frame?

```
nrow(dataset1)
```

```
## [1] 153
```

- **Answer:** 153
- **PS:** Use fuction `nrow()` with `dataset_name` to count the number of rows in a dataset.

Question 14: Extract the last 2 rows of the data frame and print them to the console. What does the output look like?

```
dataset1[152:153,]
```

| | Ozone <int> | Solar.R <int> | Wind <dbl> | Temp <int> | Month <int> | Day <int> |
|-----|----------------|------------------|---------------|---------------|----------------|--------------|
| 152 | 18 | 131 | 8.0 | 76 | 9 | 29 |
| 153 | 20 | 223 | 11.5 | 68 | 9 | 30 |

2 rows

Question 15: What is the value of Ozone in the 47th row?

```
dataset1$Ozone[47]
```

```
## [1] 21
```

- **Answer:** 21
- **PS:** `Dataset_name$column[number_of_the_row]`. Using `$` to select a specific column.

Question 16: How many missing values are in the Ozone column of this data frame?

```
sum(is.na(dataset1$Ozone))
```

```
## [1] 37
```

- **Answer:** 37
- **PS:** Use `sum(is.na(dataset_nname))` to count missing values.

Question 17: What is the mean of the Ozone column in this dataset? Exclude missing values (coded as NA) from this calculation.

```
mean(dataset1$Ozone, na.rm=TRUE)
```

```
## [1] 42.12931
```

- **Answer:** 42.12931
- **PS:** `na.rm()` is the function to remove missing values NA in the calculation.

Question 18 : Extract the subset of rows of the data frame where Ozone values are above 31 and Temp values are above 90. What is the mean of Solar.R in this subset?

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##      filter, lag
```

```
## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```
subset_dataset1<-dataset1 %>%
  filter(dataset1$Ozone > 31, Temp>90)
mean(subset_dataset1$Solar.R)
```

```
## [1] 212.8
```

- **Answer:** 212.8
- **PS:** I created a new data frame with `filter()` function from the `dplyr` package.

Question 19: What is the mean of “Temp” when “Month” is equal to 6?

```
mean(dataset1[which(dataset1$Month == 6),]$Temp)
```

```
## [1] 79.1
```

- **Answer:** 79.1
- **PS:** `which()` is a very powerful function that prevents extra steps of creating a new data frame.

Question 20: What was the maximum ozone value in the month of May (i.e. Month is equal to 5)?

```
max(dataset1[which(dataset1$Month==5),]$Ozone, na.rm = TRUE)
```

```
## [1] 115
```

- **Answer:** 115