

**Question 1**

R was developed by statisticians working at

- (i) Insightful
- (ii) Bell Labs
- (iii) The University of New South Wales
- (iv) The University of Auckland

## Question 2

The definition of free software consists of four freedoms (freedoms 0 through 3).

Which of the following is NOT one of the freedoms that are part of the definition?

- (i) The freedom to study how the program works, and adapt it to your needs.
- (ii) The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.
- (iii) The freedom to run the program, for any purpose.
- (iv) The freedom to sell the software for any price.

### Question 3

In **R** the following are all **atomic data types** EXCEPT

- (i) `logical`
- (ii) `complex`
- (iii) `list`
- (iv) `integer`

**Remark:** Atomic would mean *non divisible*, having no subsets.

#### Question 4

If I execute the expression `x <- 4` in `R` , what is the class of the object '`x`' as determined by the '`class()`' function?

- (i) `vector`
- (ii) `numeric`
- (iii) `complex`
- (iv) `real`

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

## Other Useful Commands

```
mode(x)
str(x)
dim(x)
length(x)
```

Try out these commands for some other objects, including inbuilt data sets **iris** and **Nile**.

```
iris
Nile
Y <- "R"
Z <- c(TRUE,FALSE,TRUE)
```

### Question 5

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

- (i) mixed
- (i) logical
- (i) character
- (i) integer

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

### 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)`?

- (i) a vector of length 2
- (ii) a 2 by 3 matrix
- (iii) a vector of length 3
- (iv) a 2 by 2 matrix

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

- Use the help file to find out what the commands `rbind()`, `cbind()` and `t()` do.
- The convention is to specify the number of rows, then number of columns.
- The tranpose operator “`t()`” is actually really useful for creating reports from data contained in data frames.

**Question 7**

A key property of vectors in **R** is that

- (i) the length of a vector must be less than 32,768
- (ii) elements of a vector all must be of the same class
- (iii) elements of a vector can only be character or numeric
- (iv) elements of a vector can be of different classes



### Question 8

Suppose I have a list defined as `x <- list(2, "a", "b", TRUE)`.

What does `x[[2]]` give me?

- (i) a list containing the number 2 and the letter "a".
- (ii) a character vector with the elements "a" and "b".
- (iii) a character vector containing the letter "a".
- (iv) a list containing a character vector with the elements "a" and "b".

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

- Try out `dim()` and `class()` on x also.

### Question 9

Suppose I have a vector  $\mathbf{x} \leftarrow 1:4$  and a vector  $\mathbf{y} \leftarrow 2$ .  
What is produced by the expression  $\mathbf{x} + \mathbf{y}$ ?

- (i) a numeric vector with elements 3, 4, 5, 6.
- (ii) an integer vector with elements 3, 2, 3, 4.
- (iii) a numeric vector with elements 3, 2, 3, 4.
- (iv) an integer vector with elements 3, 2, 3, 6.

### Question 10

Suppose I have a vector

```
x <- c(3, 5, 1, 10, 12, 6)
```

and I want to set all elements of this vector that are less than 6 to be equal to zero. What **R** code achieves this?

- (i) `x[x > 0] <- 6`
- (ii) `x[x < 6] <- 0`
- (iii) `x[x > 6] <- 0`
- (iv) `x[x == 6] <- 0`

This is called **Logical Indexing**. To get an idea of logical indexing, try out the following code snippets.

```
x <- c(3, 5, 1, 10, 12, 6)
x>0
x<6
X==6
```

## The airquality data set

**Some Useful Commands** As well as some of the commands we have seen earlier, it is worth getting to know the following commands also.

1 `head()` and `tail()`

2 `names()`, `rownames()` and `colnames()`

3 `summary()`

4 `complete.cases()`

5 `dim()`, `nrow()` and `ncol()`

Use the `help` command to find out what each of these commands do.

```
help(complete.cases)
```

```
> head(airquality)
Ozone Solar.R   Wind Temp Month Day
1      41    190   7.4   67    5   1
2      36    118   8.0   72    5   2
3      12    149  12.6   74    5   3
....
```

**Question 11**

In the dataset provided for this Quiz, what are the column names of the dataset?

- (i) Ozone, Solar.R, Wind, Temp, Month, Day
- (ii) Ozone, Solar.R, Wind
- (iii) Month, Day, Temp, Wind
- (iv) 1, 2, 3, 4, 5, 6

## Sequences and Numerical Indexing

A sequence of integers can be created using the colon symbol. The sequence may either be *count-up* or *count-down*.

Importantly, the sequence will return all integers between the upper and lower bound, and including both the upper and lower bound (*Other languages can be different in this respect*)

```
0:5  
1:6  
-4:5  
10:1
```

A contiguous group of rows from a data frame may be extracted using the appropriate sequence of values as indices. Likewise for a contiguous group of columns

```
iris[1:6,]      # First Six Rows  
iris[,3:4]      # Third and Fourth Columns  
iris[1:40,2:3]  # 40 Rows, Columns 2 and 3
```

### Question 12

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

```
Ozone Solar.R Wind Temp Month Day
1     41     190  7.4   67     5   1
2     36     118  8.0   72     5   2
```

```
Ozone Solar.R Wind Temp Month Day
1      7      NA  6.9   74     5  11
2     35     274 10.3   82     7  17
```

```
Ozone Solar.R Wind Temp Month Day
1     18     224 13.8   67     9  17
2     NA     258  9.7   81     7  22
```

```
Ozone Solar.R Wind Temp Month Day
1      9      24 10.9   71     9  14
2     18     131  8.0   76     9  29
```

### Question 13

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

- 129
- 153
- 160
- 45



### Question 14

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

```
Ozone Solar.R Wind Temp Month Day
152    11     44  9.7   62     5  20
153   108    223  8.0   85     7  25
```

```
Ozone Solar.R Wind Temp Month Day
152    34    307 12.0   66     5  17
153    13     27 10.3   76     9  18
```

```
Ozone Solar.R Wind Temp Month Day
152    18    131  8.0   76     9  29
153    20    223 11.5   68     9  30
```

```
Ozone Solar.R Wind Temp Month Day
152    31    244 10.9   78     8  19
153    29    127  9.7   82     6   7
```

**Question 15**

What is the value of Ozone in the 47th row?

- (i) 63
- (ii) 34
- (iii) 18
- (iv) 21

### Question 16

How many missing values (**NAs**) are in the *Ozone* column of this data frame?

- (i) 37
- (ii) 78
- (iii) 43
- (iv) 9

```
names(airquality)
attach(airquality)
is.na(Ozone)

!is.na(Ozone)

which(is.na(Ozone))

length(which(is.na(Ozone)))
detach(airquality)
```

**Question 17**

What is the mean of the Ozone column in this dataset? Exclude missing values (coded as NA) from this calculation. 42.1 18.0 53.2 31.5

### 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? 185.9  
212.8 334.0 205.0

**Question 19**

What is the mean of "Temp" when "Month" is equal to 6? 90.2 85.6 79.1 75.3

**Question 20**

What was the maximum ozone value in the month of May (i.e. Month = 5)? 100 115 18 97