Learning R - Session 1
Quiz 1

Viraj

June 14, 2016

Quiz Questions and Answers

Question 1

R was developed by statisticians working at

- The University of New South Wales
- Johns Hopkins University
- Insightful
- **The University of Auckland**

Notes: *R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand*

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?

- The freedom to redistribute copies so you can help your neighbor.
- The freedom to run the program, for any purpose.
- The freedom to improve the program, and release your improvements to the public, so that the whole community benefits.
- **The freedom to restrict access to the source code for the software.**

This is not part of the free software definition. Freedoms 1 and 3 require access to the source code.

Question 3

In R the following are all atomic data types EXCEPT

- complex
- numeric
- integer
- **matrix**

Atomic data are data elements that represent the lowest level of detail, which `matrix` is not.

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?

- integer
- **numeric**
- real
- matrix

```
x <- 4
class(x)

## [1] "numeric"
```

Question 5

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

- mixed
- numeric
- integer
- **character**

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

## [1] "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 `cbind(x, y)`?

- a vector of length 3
- a 2 by 3 matrix
- **a numeric matrix with 3 rows and 2 columns**
- a vector of length 2

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

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

class(cbind(x, y))

## [1] "matrix"
```

Question 7

A key property of vectors in R is that:

- a vector cannot have have attributes like dimensions
- the length of a vector must be less than 32,768
- **elements of a vector all must be of the same class**
- elements of a vector can only be character or numeric

Question 8

Suppose I have a list defined as `x <- list(2, "a", "b", TRUE)`. What does `x[[1]]` give me?

- a list containing the letter "a".
- **a numeric vector containing the element 2.**
- a list containing the number 2.
- a character vector containing the element "2".

```
x <- list(2,"a","b",TRUE)
x[[1]]

## [1] 2
```

Question 9

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

- a numeric vector with elements 3, 2, 3, 4.
- **a numeric vector with elements 3, 4, 5, 6.**
- a numeric vector with elements 3, 2, 3, 6.
- an integer vector with elements 3, 2, 3, 6.

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

## [1] 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?

- **x[x >= 11] <- 4**
- x[x > 10] == 4
- x[x == 4] > 10
- x[x == 10] <- 4

```r
x <- c(17, 14, 4, 5, 13, 12, 10)
x[x >= 11]  <-  4
x
```

```
## [1]  4  4  4  5  4  4 10
```

Question 11

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

- Month, Day, Temp, Wind
- Ozone, Solar.R, Wind
- **Ozone, Solar.R, Wind, Temp, Month, Day**
- 1, 2, 3, 4, 5, 6

```r
mydata <- read.csv("./datasets/NYair.csv", header = TRUE)
colnames(mydata)
```

```
## [1] "Ozone"   "Solar.R" "Wind"    "Temp"    "Month"   "Day"
```

Question 12

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

- Option 1

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

- Option 2

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

- Option 3

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

- **Option 4**

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

```r
head(mydata,2)
```

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

Question 13

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

- **153**
- 45
- 160
- 129

```
nrow(mydata)
```

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

Question 14

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

- Option 1

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

- Option 2

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

- **Option 3**

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

- Option 4

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

```
tail(mydata, 2)
```

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

Question 15

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

- **21**
- 63
- 18
- 34

```
mydata$Ozone[47]
```

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

```
mydata[47,]
```

```
##    Ozone Solar.R Wind Temp Month Day
## 47    21     191 14.9   77     6  16
```

Question 16

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

- 78
- 9
- **37**
- 43

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

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

```
sum(!complete.cases(mydata$Ozone))
```

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

Question 17

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

- 18.0
- **42.1**
- 53.2
- 31.5

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

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

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

```
q18Ans <- mydata[(mydata$Ozone > 31 & mydata$Temp > 90),]
q18Ans
```

```
##      Ozone Solar.R Wind Temp Month Day
## NA      NA      NA   NA   NA    NA  NA
## NA.1    NA      NA   NA   NA    NA  NA
## 69      97     267  6.3   92     7   8
## 70      97     272  5.7   92     7   9
## NA.2    NA      NA   NA   NA    NA  NA
## NA.3    NA      NA   NA   NA    NA  NA
## 120     76     203  9.7   97     8  28
## 121    118     225  2.3   94     8  29
## 122     84     237  6.3   96     8  30
## 123     85     188  6.3   94     8  31
## 124     96     167  6.9   91     9   1
## 125     78     197  5.1   92     9   2
## 126     73     183  2.8   93     9   3
## 127     91     189  4.6   93     9   4
```

*Using the [ ], we get records with NAs, to avoid this we can use `which()`

```
q18Ans <- mydata[which(mydata$Ozone > 31 & mydata$Temp > 90),]
q18Ans
```

```
##     Ozone Solar.R Wind Temp Month Day
## 69     97     267  6.3   92     7   8
## 70     97     272  5.7   92     7   9
## 120    76     203  9.7   97     8  28
## 121   118     225  2.3   94     8  29
## 122    84     237  6.3   96     8  30
## 123    85     188  6.3   94     8  31
## 124    96     167  6.9   91     9   1
## 125    78     197  5.1   92     9   2
## 126    73     183  2.8   93     9   3
## 127    91     189  4.6   93     9   4
```

Now we will pull out the mean

```
mean(q18Ans$Solar.R, na.rm = TRUE)
```

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

Alternatively, we do not have to create a subset, and can directly get the desired output as below:

```
mean(mydata$Solar.R[(mydata$Ozone > 31 &
                       mydata$Temp > 90)], na.rm = TRUE)

## [1] 212.8
```

Question 19

What is the mean of "Temp" when "Month" is equal to 6?

- **79.1**
- 75.3
- 85.6
- 90.2

```
mean(mydata$Temp[mydata$Month == 6], na.rm = TRUE)

## [1] 79.1
```

Question 20

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

- 97
- 100
- **115**
- 18

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

## [1] 115
```