



Week 1 Answers - R Programming

Question 1

R was developed by statisticians working at...

Answer

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?

Answer

The freedom to sell the software for any price.

Question 3

In R the following are all atomic data types EXCEPT

Answer

matrix

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?

Answer



numeric

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

Question 5

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

Answer

numeric

```
x <- c(4, 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 `cbind(x, y)`?

Answer

a 3 by 2 numeric matrix

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

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[[1]]` give me?

Answer

a numeric vector containing the element 2

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

class(x[[1]])
```

Question 9

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

Answer

a numeric vector with elements 3, 4, 5, 6.

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

class(x + y)
```

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?

Answer

```
x[x >= 11] <- 4

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



x

Question 11

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

Answer

Ozone, Solar.R, Wind, Temp, Month, Day

```
# install package if doesnt exist
install.packages("data.table")
library("data.table")

# Reading in data
quiz_data <- fread('hw1_data.csv')

# Column names of the dataset
names(quiz_data)
```

Question 12

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

Answer

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

```
# First two rows
quiz_data[c(1,2),]
```

OR

```
# First two rows
head(quiz_data,2)
```



Question 13

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

Answer

153

```
nrow(quiz_data)
```

Question 14

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

Answer

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

```
tail(quiz_data, 2)
```

Question 15

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

Answer

21

```
quiz_data[47, 'Ozone']
```

Question 16

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



Answer

37

```
# Going back to data.frame because dont it hasnt been taught yet in this
specialization
quiz_data = read.csv('hw1_data.csv')
sub = subset(quiz_data, is.na(Ozone))
nrow(sub)
```

```
# Can also remmove Missing Values using Something like This
quiz_data[complete.cases(quiz_data),]
```

Question 17

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

Answer

42.1

Explanation

The 'mean' function can be used to calculate the mean.

```
hw1 = read.csv('hw1_data.csv')
sub = subset(hw1, !is.na(Ozone), select = Ozone)
apply(sub, 2, mean)
```

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?

Answer

212.8



```
quiz_data = read.csv('hw1_data.csv')
sub = subset(quiz_data, Ozone > 31 & Temp > 90, select = Solar.R)
apply(sub, 2, mean)
```

Question 19

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

Answer

79.1

Explanation

```
quiz_data = read.csv('hw1_data.csv')
sub = subset(hw1, Month == 6, select = Temp)
apply(sub, 2, mean)
```

Question 20

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

Answer

115

Explanation

```
quiz_data = read.csv('hw1_data.csv')
sub = subset(quiz_data, Month == 5 & !is.na(Ozone), select = Ozone)
apply(sub, 2, max)
```

NOTE: If we wanted more columns to display for sub we can achieve this by the following:

```
sub = subset(quiz_data, Month == 5 & !is.na(Ozone), select = c(Ozone, Solar.R, Day))
```



Week 2 Answers - R Programming

Question 1

Suppose I define the following function in R

```
cube <- function(x, n) {  
  x^3  
}
```

What is the result of running `cube(3)` in R after defining this function?

Answer

The number 27 is returned

Question 2

The following code will produce a warning in R.

```
x <- 1:10  
if(x > 5) {  
  x <- 0  
}
```

Why?

Answer

'x' is a vector of length 10 and 'if' can only test a single logical statement.

Question 3

Consider the following function

```
f <- function(x) {  
  g <- function(y) {  
    y + z  
  }  
}
```



```
    }  
    z <- 4  
    x + g(x)  
}
```

If I then run in R

```
z <- 10  
f(3)
```

What value is returned?

Answer

10

Question 4

Consider the following expression:

```
x <- 5  
y <- if(x < 3) {  
    NA  
} else {  
    10  
}
```

What is the value of 'y' after evaluating this expression?

Answer

10

Question 5

Consider the following R function

```
h <- function(x, y = NULL, d = 3L) {  
    z <- cbind(x, d)  
    if(!is.null(y))  
        z <- z + y  
}
```



```
else
  z <- z + f
g <- x + y / z
if(d == 3L)
  return(g)
g <- g + 10
g
}
```

Which symbol in the above function is a free variable?

Answer

f

Question 6

What is an environment in R?

Answer

a collection of symbol/value pairs

Question 7

The R language uses what type of scoping rule for resolving free variables?

Answer

lexical scoping

Question 8

How are free variables in R functions resolved?

Answer



The values of free variables are searched for in the environment in which the function was defined

Question 9

What is one of the consequences of the scoping rules used in R?

Answer

All objects must be stored in memory Correct 1.00

Question 10

In R, what is the parent frame?

Answer

It is the environment in which a function was called

Week 3 Answers - R Programming

Question 1

Take a look at the 'iris' dataset that comes with R. The data can be loaded with the code:

```
library(datasets)
data(iris)
```

A description of the dataset can be found by running

```
?iris
```



There will be an object called 'iris' in your workspace. In this dataset, what is the mean of 'Sepal.Length' for the species virginica? (Please only enter the numeric result and nothing else.)

Answer

6.588

```
# if you don't have data.table installed
# install.packages("data.table")

library(data.table)
iris_dt <- as.data.table(iris)
iris_dt[Species == "virginica", round(mean(Sepal.Length)) ]
```

Question 2

Continuing with the 'iris' dataset from the previous Question, what R code returns a vector of the means of the variables 'Sepal.Length', 'Sepal.Width', 'Petal.Length', and 'Petal.Width'?

Answer

```
apply(iris[, 1:4], 2, mean)
```

Question 3

Load the 'mtcars' dataset in R with the following code

```
library(datasets)
data(mtcars)
```

There will be an object names 'mtcars' in your workspace. You can find some information about the dataset by running

```
?mtcars
```

How can one calculate the average miles per gallon (mpg) by number of cylinders in the car (cyl)?



Answer

```
with(mtcars, tapply(mpg, cyl, mean))
tapply(mtcars$mpg, mtcars$cyl, mean)
sapply(split(mtcars$mpg, mtcars$cyl) , mean)
```

Question 4

Continuing with the 'mtcars' dataset from the previous Question, what is the absolute difference between the average horsepower of 4-cylinder cars and the average horsepower of 8-cylinder cars?

Answer

126.5779

```
mtcars_dt <- as.data.table(mtcars)
mtcars_dt <- mtcars_dt[, .(mean_cols = mean(hp)), by = cyl]
round(abs(mtcars_dt[cyl == 4, mean_cols] - mtcars_dt[cyl == 8, mean_cols]))
```

Question 5

If you run

```
debug(ls)
```

what happens when you next call the 'ls' function?

Answer

Execution of 'ls' will suspend at the beginning of the function and you will be in the browser.

Week 4 Answers - R Programming



Question 1

What is produced at the end of this snippet of R code?

```
set.seed(1)
rpois(5, 2)
```

Answer

A vector with the numbers 1, 1, 2, 4, 1

Question 2

What R function can be used to generate standard Normal random variables?

Answer

rnorm

Question 3

When simulating data, why is using the `set.seed()` function important?

Answer

It ensures that the sequence of random numbers starts in a specific place and is therefore reproducible.

Question 4

Which function can be used to evaluate the inverse cumulative distribution function for the Poisson distribution?

Answer

qpois



Explanation

Probability distribution functions beginning with the q prefix are used to evaluate the quantile function.

Question 5

What does the following code do?

```
set.seed(10)
x <- rbinom(10, 10, 0.5)
e <- rnorm(10, 0, 20)
y <- 0.5 + 2 * x + e
```

Answer

Generate data from a Normal linear model

Question 6

What R function can be used to generate Binomial random variables?

Answer

rbinom

Question 7

What aspect of the R runtime does the profiler keep track of when an R expression is evaluated?

Answer

the function call stack

Question 8



Consider the following R code

```
library(datasets)
Rprof()
fit <- lm(y ~ x1 + x2)
Rprof(NULL)
```

(Assume that `y`, `x1`, and `x2` are present in the workspace.) Without running the code, what percentage of the run time is spent in the `lm` function, based on the `by.total` method of normalization shown in `summaryRprof()`?

Answer

100%

Explanation

When using `by.total` normalization, the top-level function (in this case, `lm()`) always takes 100% of the time.

Question 9

When using `system.time()`, what is the user time?

Answer

It is the time spent by the CPU evaluating an expression

Question 10

If a computer has more than one available processor and R is able to take advantage of that, then which of the following is true when using `system.time()`?

Answer

Elapsed time may be smaller than user time