

ALY6000 Introduction to Analytics

Project Report 1

Assignment: Introduction to Problem-Solving with R

Submission Date: 31st January 2024, Wednesday

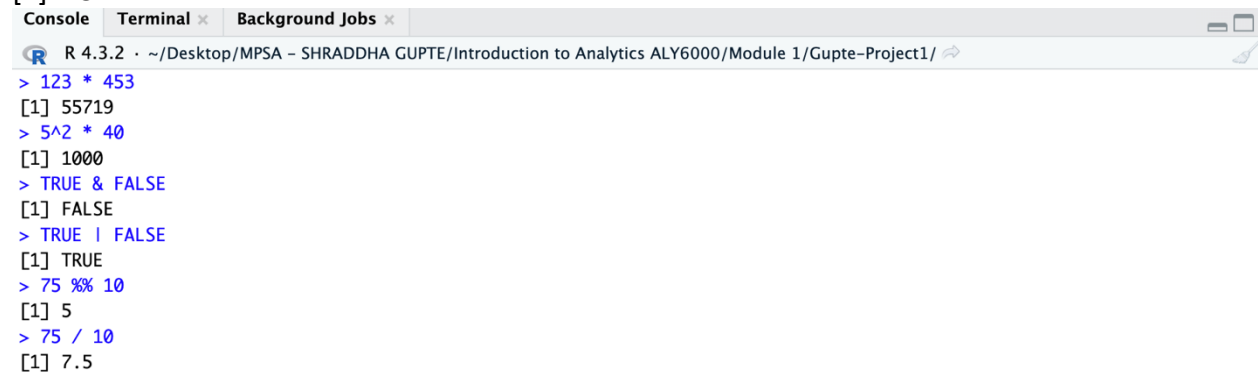
Submitted by

Shraddha P Gupte

Introduction to Problem-Solving with R

Question 1

```
> 123 * 453
[1] 55719
> 5^2 * 40
[1] 1000
> TRUE & FALSE
[1] FALSE
> TRUE | FALSE
[1] TRUE
> 75 %% 10
[1] 5
> 75 / 10
[1] 7.5
```

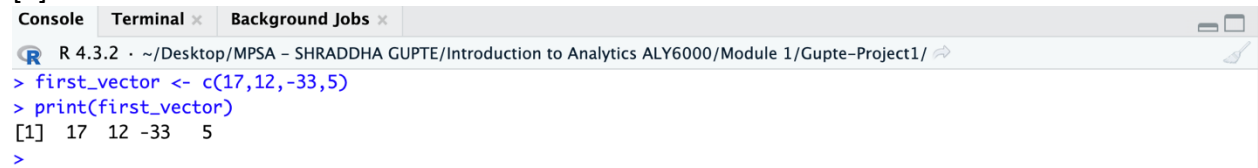


The screenshot shows an R console window with the following text:

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> 123 * 453
[1] 55719
> 5^2 * 40
[1] 1000
> TRUE & FALSE
[1] FALSE
> TRUE | FALSE
[1] TRUE
> 75 %% 10
[1] 5
> 75 / 10
[1] 7.5
```

Question 2

```
> first_vector <- c(17,12,-33,5)
> print(first_vector)
[1] 17 12 -33 5
```



The screenshot shows an R console window with the following text:

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> first_vector <- c(17,12,-33,5)
> print(first_vector)
[1] 17 12 -33 5
>
```

Question 3

```
> counting_by_fives <- c(5,10,15,20,25,30,35)
> print(counting_by_fives)
[1] 5 10 15 20 25 30 35
```

Introduction to Problem-Solving with R

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> counting_by_fives <- c(5,10,15,20,25,30,35)
> print(counting_by_fives)
[1] 5 10 15 20 25 30 35
> |
```

Question 4

```
> second_vector <- (20:1)
> print(second_vector)
[1] 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> second_vector <- (20:1)
> print(second_vector)
[1] 20 19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1
> |
```

Question 5

```
> counting_vector <- (5:15)
> print(counting_vector)
[1] 5 6 7 8 9 10 11 12 13 14 15
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> counting_vector <- (5:15)
> print(counting_vector)
[1] 5 6 7 8 9 10 11 12 13 14 15
> |
```

Question 6

```
> grades <- c(96,100,85,92,81,72)
> print(grades)
[1] 96 100 85 92 81 72
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> grades <- c(96,100,85,92,81,72)
> print(grades)
[1] 96 100 85 92 81 72
> |
```

Question 7

```
> bonus_points_added <- (grades + 3)
> print(bonus_points_added)
[1] 99 103 88 95 84 75
```

Introduction to Problem-Solving with R

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> bonus_points_added <- (grades + 3)
> print(bonus_points_added)
[1] 99 103 88 95 84 75
```

Question 8

```
> one_to_one_hundred <- c(1:100)
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> one_to_one_hundred <- c(1:100)
> print(one_to_one_hundred)
[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
[27] 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
[53] 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78
[79] 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100
```

Question 9

```
# + means "add"
```

```
second_vector + 20
```

```
# * means "multiply"
```

```
second_vector * 20
```

```
# >= means "less than equal to"
```

```
second_vector >= 20
```

```
# != means "not equal"
```

```
second_vector != 20
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> # + means "add"
> second_vector + 20
[1] 40 39 38 37 36 35 34 33 32 31 30 29 28 27 26 25 24 23 22 21
> # * means "multiply"
> second_vector * 20
[1] 400 380 360 340 320 300 280 260 240 220 200 180 160 140 120 100 80 60 40 20
> # >= means "less than equal to"
> second_vector >= 20
[1] TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[19] FALSE FALSE
> # != means "not equal"
> second_vector != 20
[1] FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
[19] TRUE TRUE
> |
```

Comments –

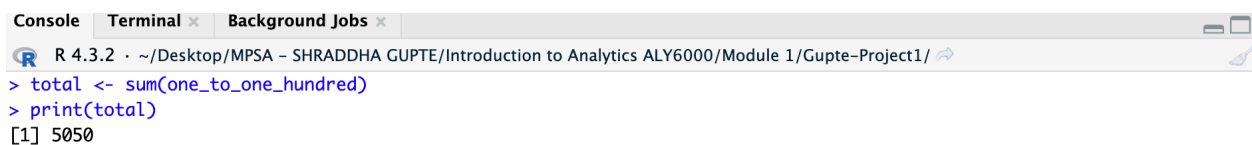
- `second_vector + 20` – This operation will add 20 to each element of the second vector. The + sign denotes addition to the values of second vector. For eg – The first element in second vector is 20. Hence the operation will add $20+20 = 40$
- `second_vector * 20` - This operation will multiply each element of the second vector by 20. The * sign denotes multiplication to the values of second vector. For eg – The first element in second vector is 20. Hence the operation will multiply $20*20 = 400$

Introduction to Problem-Solving with R

- `second_vector >= 20` - This operation will show the result for each element of the second vector that is more than or equal to 20. For eg – The first element in second vector is 20. Hence the operation is showing the return as 'TRUE'. But the following elements are not equal to 20. Hence the return is showing as 'FALSE'.
- `second_vector != 20` - This operation will show the result for each element of the second vector that is not equal to 20. For eg – The first element in second vector is 20. Hence the operation is showing the return as 'FALSE' as it is equal to 20. But the following elements are not equal to 20. Hence the return is showing as 'TRUE'.

Question 10

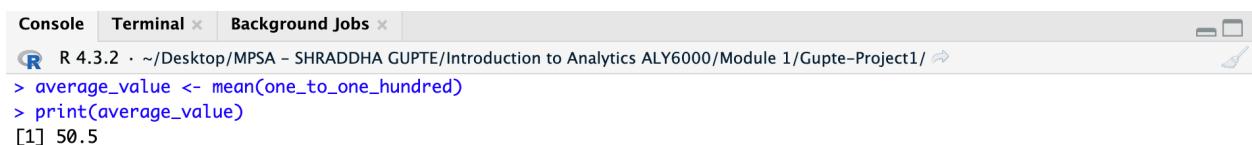
```
> total <- sum(one_to_one_hundred)
> print(total)
[1] 5050
```



The screenshot shows an R console window with the following text: R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/. The commands entered are `> total <- sum(one_to_one_hundred)` and `> print(total)`, resulting in the output `[1] 5050`.

Question 11

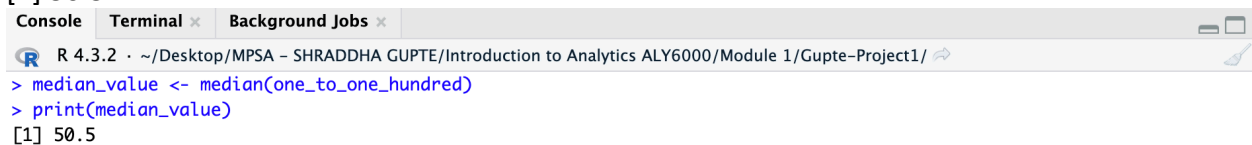
```
> average_value <- mean(one_to_one_hundred)
> print(average_value)
[1] 50.5
```



The screenshot shows an R console window with the following text: R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/. The commands entered are `> average_value <- mean(one_to_one_hundred)` and `> print(average_value)`, resulting in the output `[1] 50.5`.

Question 12

```
> median_value <- median(one_to_one_hundred)
> print(median_value)
[1] 50.5
```



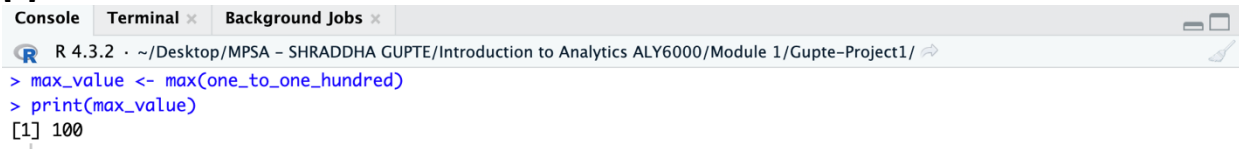
The screenshot shows an R console window with the following text: R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/. The commands entered are `> median_value <- median(one_to_one_hundred)` and `> print(median_value)`, resulting in the output `[1] 50.5`.

Question 13

```
> max_value <- max(one_to_one_hundred)
```

Introduction to Problem-Solving with R

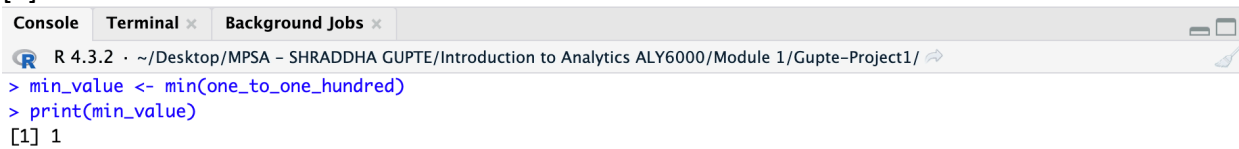
```
> print(max_value)
[1] 100
```

A screenshot of an R console window. The title bar shows 'R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/'. The console output shows the command 'max_value <- max(one_to_one_hundred)' followed by 'print(max_value)', resulting in the output '[1] 100'.

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> max_value <- max(one_to_one_hundred)
> print(max_value)
[1] 100
```

Question 14

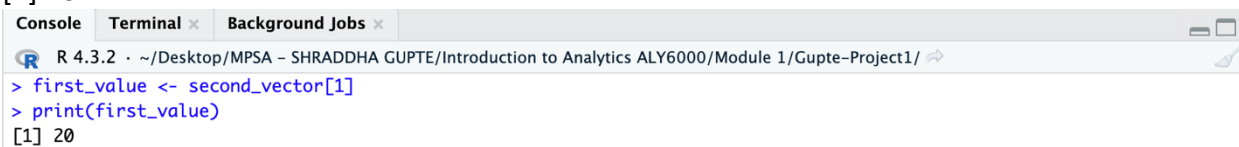
```
> min_value <- min(one_to_one_hundred)
> print(min_value)
[1] 1
```

A screenshot of an R console window. The title bar shows 'R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/'. The console output shows the command 'min_value <- min(one_to_one_hundred)' followed by 'print(min_value)', resulting in the output '[1] 1'.

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> min_value <- min(one_to_one_hundred)
> print(min_value)
[1] 1
```

Question 15

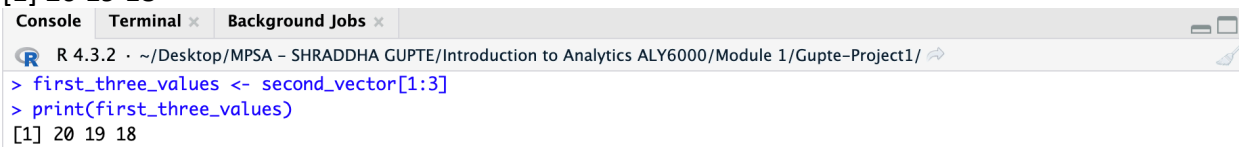
```
> first_value <- second_vector[1]
> print(first_value)
[1] 20
```

A screenshot of an R console window. The title bar shows 'R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/'. The console output shows the command 'first_value <- second_vector[1]' followed by 'print(first_value)', resulting in the output '[1] 20'.

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> first_value <- second_vector[1]
> print(first_value)
[1] 20
```

Question 16

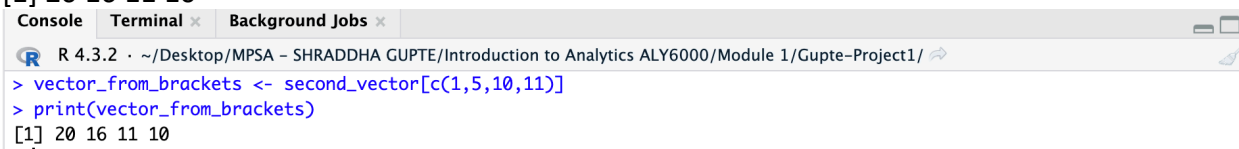
```
> first_three_values <- second_vector[1:3]
> print(first_three_values)
[1] 20 19 18
```

A screenshot of an R console window. The title bar shows 'R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/'. The console output shows the command 'first_three_values <- second_vector[1:3]' followed by 'print(first_three_values)', resulting in the output '[1] 20 19 18'.

```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> first_three_values <- second_vector[1:3]
> print(first_three_values)
[1] 20 19 18
```

Question 17

```
> vector_from_brackets <- second_vector[c(1,5,10,11)]
> print(vector_from_brackets)
[1] 20 16 11 10
```

A screenshot of an R console window. The title bar shows 'R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/'. The console output shows the command 'vector_from_brackets <- second_vector[c(1,5,10,11)]' followed by 'print(vector_from_brackets)', resulting in the output '[1] 20 16 11 10'.

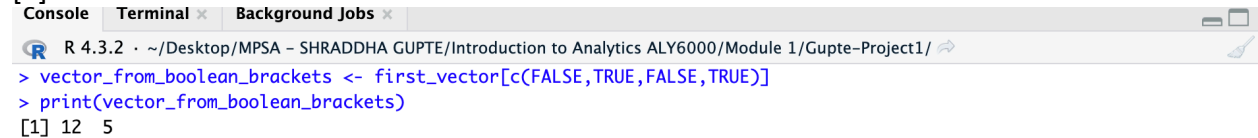
```
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> vector_from_brackets <- second_vector[c(1,5,10,11)]
> print(vector_from_brackets)
[1] 20 16 11 10
```

Introduction to Problem-Solving with R

Question 18

```
> vector_from_boolean_brackets <- first_vector[c(FALSE,TRUE,FALSE,TRUE)]  
> print(vector_from_boolean_brackets)
```

```
[1] 12 5
```



The screenshot shows the R console with the following text:
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> vector_from_boolean_brackets <- first_vector[c(FALSE,TRUE,FALSE,TRUE)]
> print(vector_from_boolean_brackets)
[1] 12 5

Comments - vector_from_boolean_brackets picks values from first_vector and assigns them in a logical R objects.

FALSE: denotes not to take the corresponding values from the first_vector. i.e 17 and -33

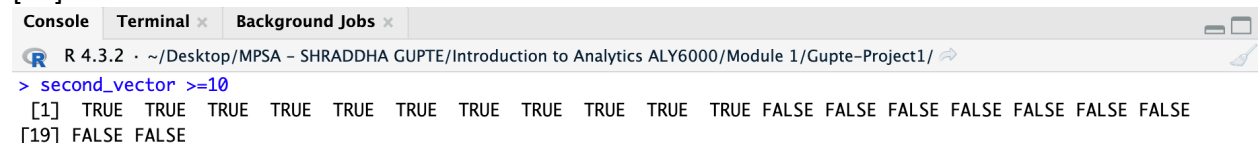
TRUE: denotes to take the corresponding values from the first_vector. i.e 12 and 5

Question 19

```
> second_vector >= 10
```

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

```
[19] FALSE FALSE
```

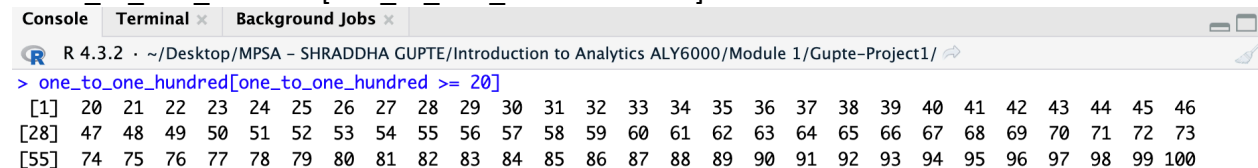


The screenshot shows the R console with the following text:
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> second_vector >= 10
[1] TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
[19] FALSE FALSE

Comments – This operation took all the elements in second_vector that are greater than or equal to 10 and assigned them for ‘TRUE’. The elements that were less or not equal to 10 are assigned to ‘FALSE’

Question 20

```
> one_to_one_hundred[one_to_one_hundred >= 20]
```



The screenshot shows the R console with the following text:
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> one_to_one_hundred[one_to_one_hundred >= 20]
[1] 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
[28] 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73
[55] 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

Comments – This operation showing all the elements that are greater than or equal to 20 from the vector named ‘one_to_one_hundred’.

Question 21

```
> lowest_grades_removed <- grades[grades > 85]
```

```
> print(lowest_grades_removed)
```

```
[1] 96 100 92
```

Introduction to Problem-Solving with R

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> lowest_grades_removed <- grades[grades >85]
> print(lowest_grades_removed)
[1] 96 100 92
```

Comments – This operation is showing elements that are more than 85 from the vector named as 'grades'.

Question 22

```
> middle_grades_removed <- grades[-c(3,4)]
> print(middle_grades_removed)
[1] 96 100 81 72
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> middle_grades_removed <- grades[-c(3,4)]
> print(middle_grades_removed)
[1] 96 100 81 72
```

Question 23

```
> fifth_vector <- second_vector[-c(5,10)]
> print(fifth_vector)
[1] 20 19 18 17 15 14 13 12 10 9 8 7 6 5 4 3 2 1
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> fifth_vector <- second_vector[-c(5,10)]
> print(fifth_vector)
[1] 20 19 18 17 15 14 13 12 10 9 8 7 6 5 4 3 2 1
```

Question 24

```
> set.seed(5)
> random_vector <- runif(n=10, min = 0, max = 1000)
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MP5A - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> set.seed(5)
> random_vector <- runif(n=10, min = 0, max = 1000)
> print(set.seed(5))
NULL
> print(random_vector)
[1] 200.2145 685.2186 916.8758 284.3995 104.6501 701.0575 527.9600 807.9352 956.5001 110.4530
```

Question 25

```
> sum_vector <- sum(random_vector)
> print(sum_vector)
[1] 5295.264
```


Introduction to Problem-Solving with R

```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MP5A – SHRADDHA GUPTÉ/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/ ↗
> sum_vector <- sum(random_vector)
> print(sum_vector)
[1] 5295.264
```

Question 26

```
> cumsum_vector <- cumsum(random_vector)
> print(cumsum_vector)
[1] 200.2145 885.4330 1802.3088 2086.7083 2191.3584 2892.4159 3420.3759 4228.3111
5184.8112 5295.2642
```

```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MP5A – SHRADDHA GUPTÉ/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/ ↗
> cumsum_vector <- cumsum(random_vector)
> print(cumsum_vector)
[1] 200.2145 885.4330 1802.3088 2086.7083 2191.3584 2892.4159 3420.3759 4228.3111 5184.8112 5295.2642
```

Question 27

```
> mean_vector <- mean(random_vector)
> print(mean_vector)
[1] 529.5264
```

```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MP5A – SHRADDHA GUPTÉ/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/ ↗
> mean_vector <- mean(random_vector)
> print(mean_vector)
[1] 529.5264
```

Question 28

```
> sd_vector <- sd(random_vector)
> print(sd_vector)
[1] 331.3606
```

```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MP5A – SHRADDHA GUPTÉ/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/ ↗
> sd_vector <- sd(random_vector)
> print(sd_vector)
[1] 331.3606
```

Question 29

```
> round_vector <- round(random_vector)
> print(round_vector)
[1] 200 685 917 284 105 701 528 808 957 110
```

Introduction to Problem-Solving with R

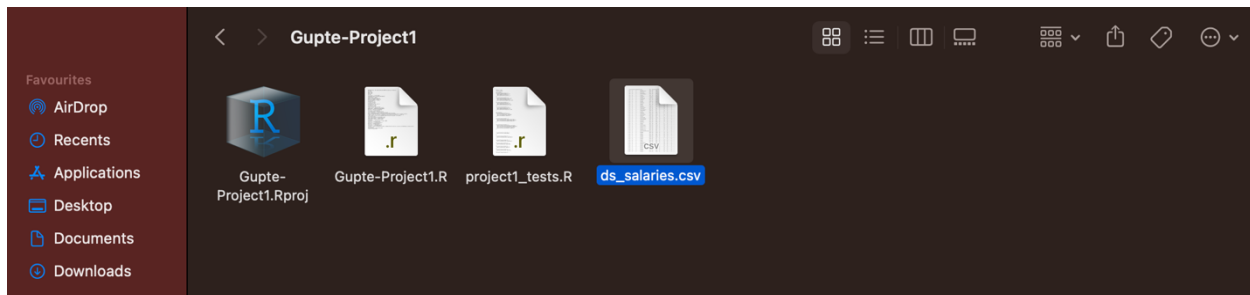
```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> round_vector <- round(random_vector)
> print(round_vector)
[1] 200 685 917 284 105 701 528 808 957 110
```

Question 30

```
> sort_vector <- sort(random_vector)
> print(sort_vector)
[1] 104.6501 110.4530 200.2145 284.3995 527.9600 685.2186 701.0575 807.9352 916.8758
956.5001
```

```
Console Terminal x Background Jobs x
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> sort_vector <- sort(random_vector)
> print(sort_vector)
[1] 104.6501 110.4530 200.2145 284.3995 527.9600 685.2186 701.0575 807.9352 916.8758 956.5001
```

Question 31



Question 32

```
> file_path <- "/Users/shraddha/Desktop/MPSA - SHRADDHA GUPTE/Introduction to Analytics
ALY6000/Module 1/Gupte-Project1/ds_salaries.csv"
> first_dataframe <- read.csv(file_path)
> print(first_dataframe)
```

Introduction to Problem-Solving with R

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> file_path <- "/Users/shraddha/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/ds_salaries.csv"
> first_dataframe <- read.csv(file_path)
> print(first_dataframe)
```

	X	work_year	experience_level	employment_type	job_title	salary
1	0	2020	MI	FT	Data Scientist	70000
2	1	2020	SE	FT	Machine Learning Scientist	260000
3	2	2020	SE	FT	Big Data Engineer	85000
4	3	2020	MI	FT	Product Data Analyst	20000
5	4	2020	SE	FT	Machine Learning Engineer	150000
6	5	2020	EN	FT	Data Analyst	72000
7	6	2020	SE	FT	Lead Data Scientist	190000
8	7	2020	MI	FT	Data Scientist	1100000
9	8	2020	MI	FT	Business Data Analyst	135000
10	9	2020	SE	FT	Lead Data Engineer	125000
11	10	2020	EN	FT	Data Scientist	45000
12	11	2020	MI	FT	Data Scientist	300000
13	12	2020	EN	FT	Data Scientist	35000
14	13	2020	MI	FT	Lead Data Analyst	87000
15	14	2020	MI	FT	Data Analyst	85000
16	15	2020	MI	FT	Data Analyst	8000
17	16	2020	EN	FT	Data Engineer	445000
18	17	2020	SE	FT	Big Data Engineer	100000
19	18	2020	EN	FT	Data Science Consultant	423000
20	19	2020	MI	FT	Lead Data Engineer	56000
21	20	2020	MI	FT	Machine Learning Engineer	299000
22	21	2020	MI	FT	Product Data Analyst	450000
23	22	2020	SE	FT	Data Engineer	42000
24	23	2020	MI	FT	BI Data Analyst	98000
25	24	2020	MI	FT	Lead Data Scientist	115000
26	25	2020	EX	FT	Director of Data Science	325000
27	26	2020	EN	FT	Research Scientist	42000
28	27	2020	SE	FT	Data Engineer	720000
29	28	2020	EN	CT	Business Data Analyst	100000

Question 33

```
> summary(first_dataframe)
```

```
Console Terminal Background Jobs
R 4.3.2 · ~/Desktop/MPSA - SHRADDHA GUPTA/Introduction to Analytics ALY6000/Module 1/Gupte-Project1/
> summary(first_dataframe)
```

	X	work_year	experience_level	employment_type	job_title	salary
Min.	: 0.0	Min. :2020	Length:607	Length:607	Length:607	Min. : 4000
1st Qu.:	:151.5	1st Qu.:2021	Class :character	Class :character	Class :character	1st Qu.: 70000
Median :	:303.0	Median :2022	Mode :character	Mode :character	Mode :character	Median : 115000
Mean :	:303.0	Mean :2021				Mean : 324000
3rd Qu.:	:454.5	3rd Qu.:2022				3rd Qu.: 165000
Max.	:606.0	Max. :2022				Max. :30400000
salary_currency		salary_in_usd	employee_residence	remote_ratio	company_location	company_size
Length:607		Min. : 2859	Length:607	Min. : 0.00	Length:607	Length:607
Class :character		1st Qu.: 62726	Class :character	1st Qu.: 50.00	Class :character	Class :character
Mode :character		Median :101570	Mode :character	Median :100.00	Mode :character	Mode :character
		Mean :112298		Mean : 70.92		
		3rd Qu.:150000		3rd Qu.:100.00		
		Max. :600000		Max. :100.00		

```
> |
```

Citation:

1. **Book**-Kabacoff, R.I. (2022). *R in action: Data analysis and graphics with R and tidyverse* (3rd edition). Manning Publications.

Introduction to Problem-Solving with R

2. **Website**-R Core Team.(2021). R: A Language and Environment for Statistical Computing <https://www.R-project.org/>