



SESSION 5: Data Management Using R

Assignment 1

Table of Contents

1. Introduction	2
2. Objective	2
3. Prerequisites	2
4. Associated Data Files	2
5. Problem Statement.....	2
7. Approximate Time to Complete Task	6

6.Expected Output
3

1. Introduction

This assignment will help you understand the concepts learnt in the session.

2. Objective

This assignment will test your skills on Data Management functions using R.

3. Prerequisites

Not applicable.

4. Associated Data Files

Not applicable.

5. Problem Statement

1. How many vowels are there in the names of USA States?
2. Visualize the vowels distribution.

1.

```
vector of vowels vowels = c("a", "e", "i", "o", "u") # vector for storing results num_vowels =  
vector(mode = "integer", length = 5)  
# calculate number of vowels in each name for (j in seq_along(vowels)) {
```

Data Analytics

```
num_aux = str_count(tolower(states), vowels[j])
```

```
num_vowels[j] = sum(num_aux) }
```

```
# add vowel name
```

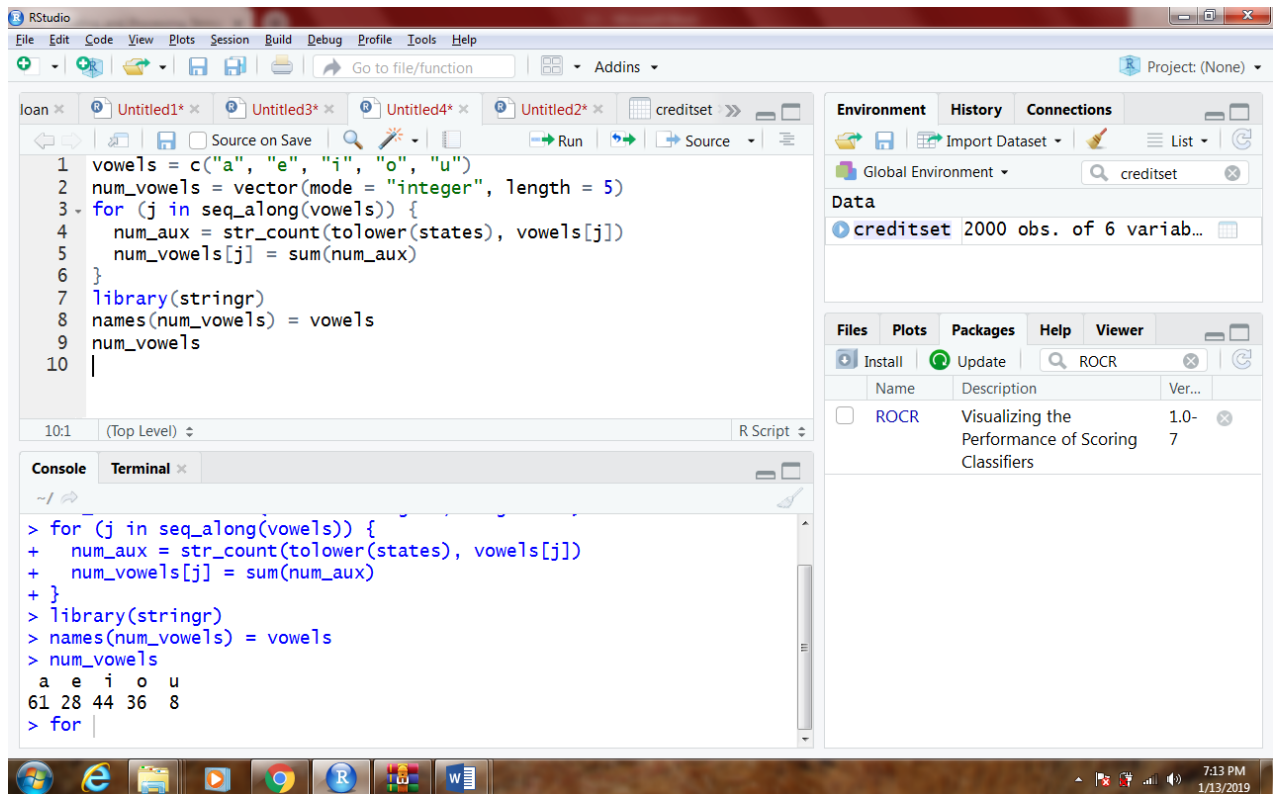
```
names(num_vowels) = vowels
```

```
# total number of vowels
```

```
num_vowels
```

```
## a e i o u
```

```
## 61 28 44 36 8
```



2. # sort them in decreasing order

```
sort(num_vowels, decreasing = TRUE)
```

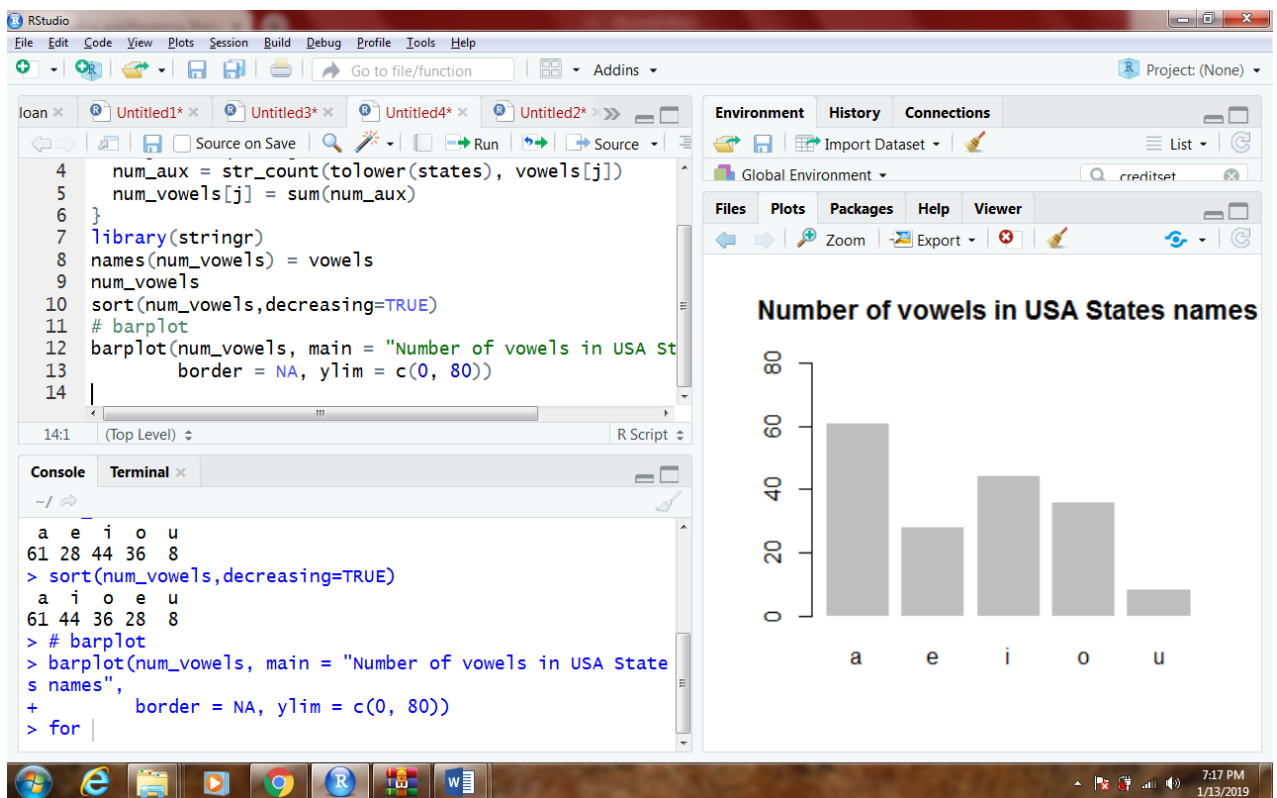
```
## a i o e u
```

```
## 61 44 36 28 8
```

And finally, we can visualize the distribution with a barplot:

```
# barplot
```

```
barplot(num_vowels, main = "Number of vowels in USA States names", border = NA, ylim = c(0, 80))
```



6. Expected Format

Data Analytics

1. R file should be submitted where applicable.
2. R file should be in PDF or in .r format
3. Proper screenshots of the outputs should be submitted as well
4. The r codes, if submitted in any other format, will be subjected to deduction in marks

Note: Your solution will not be entertained if it is any other format, e.g., .zip, .doc, .rtf etc.

7. Approximate Time to Complete Task

30 mins.