



**ACADGILD**

# SESSION 5: Data Management Using R

---

Assignment 1

## PROBLEM STATEMENT

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

**The R-script for the given problem is as follows:**

# 1. How many vowels are there in the names of USA States?

#USA states vowels

USArrests      # data set

library(stringi)

library(stringr)

str\_count(states,"a")

str\_count(states,"e")

str\_count(states,"i")

str\_count(states,"o")

str\_count(states,"u")

vowela<-str\_count(states,"a")

vowele<-str\_count(states,"e")

voweli<-str\_count(states,"i")

vowelo<-str\_count(states,"o")

vowelu<-str\_count(states,"u")

sum(vowela,vowele,voweli,vowelo,vowelu)

sum(vowela)

sum(vowele)

sum(voweli)

```
sum(vowelo)
```

```
sum(vowelu)
```

**The output of the R-Script (from Console window) is given as follows**

```
> USArrests      # data set
```

	Murder	Assault	UrbanPop	Rape
Alabama	13.2	236	58	21.2
Alaska	10.0	263	48	44.5
Arizona	8.1	294	80	31.0
Arkansas	8.8	190	50	19.5
California	9.0	276	91	40.6
Colorado	7.9	204	78	38.7
Connecticut	3.3	110	77	11.1
Delaware	5.9	238	72	15.8
Florida	15.4	335	80	31.9
Georgia	17.4	211	60	25.8
Hawaii	5.3	46	83	20.2
Idaho	2.6	120	54	14.2
Illinois	10.4	249	83	24.0
Indiana	7.2	113	65	21.0
Iowa	2.2	56	57	11.3
Kansas	6.0	115	66	18.0
Kentucky	9.7	109	52	16.3
Louisiana	15.4	249	66	22.2
Maine	2.1	83	51	7.8
Maryland	11.3	300	67	27.8
Massachusetts	4.4	149	85	16.3
Michigan	12.1	255	74	35.1
Minnesota	2.7	72	66	14.9
Mississippi	16.1	259	44	17.1
Missouri	9.0	178	70	28.2
Montana	6.0	109	53	16.4
Nebraska	4.3	102	62	16.5
Nevada	12.2	252	81	46.0
New Hampshire	2.1	57	56	9.5
New Jersey	7.4	159	89	18.8
New Mexico	11.4	285	70	32.1
New York	11.1	254	86	26.1
North Carolina	13.0	337	45	16.1
North Dakota	0.8	45	44	7.3
Ohio	7.3	120	75	21.4
Oklahoma	6.6	151	68	20.0
Oregon	4.9	159	67	29.3
Pennsylvania	6.3	106	72	14.9

Rhode Island	3.4	174	87	8.3
South Carolina	14.4	279	48	22.5
South Dakota	3.8	86	45	12.8
Tennessee	13.2	188	59	26.9
Texas	12.7	201	80	25.5
Utah	3.2	120	80	22.9
Vermont	2.2	48	32	11.2
Virginia	8.5	156	63	20.7
Washington	4.0	145	73	26.2
West Virginia	5.7	81	39	9.3
Wisconsin	2.6	53	66	10.8
Wyoming	6.8	161	60	15.6

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

Source

Console Terminal x

~/

```

Utah          3.2    120    80 22.9
Vermont       2.2     48    32 11.2
Virginia      8.5    156    63 20.7
Washington    4.0    145    73 26.2
West Virginia 5.7     81    39  9.3
Wisconsin     2.6     53    66 10.8
Wyoming       6.8    161    60 15.6
> library(stringi)
> library(stringr)
> str_count(states,"a")
[1] 4 3 2 3 2 1 0 2 1 1 2 1 0 2 1 0 2 1 2 2 1 1 0 2 2 2 1 0 0 0 2 2 0 2 0 2 1 2 2 0 1 1 0 1 1 1 0 0
> str_count(states,"e")
[1] 0 0 0 0 0 0 0 1 2 0 1 0 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1 2 3 2 1 0 0 0 0 1 1 1 1 0 0 4 1 0 1 0 0 1 0 0
> str_count(states,"i")
[1] 0 0 1 0 2 0 1 0 1 1 2 1 3 2 1 0 2 1 0 0 2 1 4 0 0 0 1 0 1 0 1 0 1 0 0 1 1 1 0 0 0 0 0 3 1 3 2 1
> str_count(states,"o")
[1] 0 0 1 0 1 3 1 0 1 1 0 1 1 0 1 0 1 0 0 0 0 1 0 1 0 0 0 0 1 1 2 2 2 2 2 0 1 2 2 0 0 0 1 0 1 0 1 1
> str_count(states,"u")
[1] 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1 0 0 0 0 0 0
> vowel_a<-str_count(states,"a")
> vowel_e<-str_count(states,"e")
> vowel_i<-str_count(states,"i")
> vowel_o<-str_count(states,"o")
> vowel_u<-str_count(states,"u")
> sum(vowel_a,vowel_e,vowel_i,vowel_o,vowel_u)
[1] 171
> sum(vowel_a)
[1] 59
> sum(vowel_e)
[1] 28
> sum(vowel_i)
[1] 42
> sum(vowel_o)
[1] 35
> sum(vowel_u)
[1] 7
>

```

## 2. Visualize the vowels distribution.

**The R-script for the given problem is as follows:**

```
vowel_dist <- filter(distribution, USA_States %in% c("a","e","i","o","u"))
vowel_dist
barplot(vowel_dist$Freq, axes = TRUE, axisnames = TRUE, xlab = "Vowels", ylab
= "frequency")
```

**The output of the R-Script (from Console window) is given as follows**

```
> vowel_dist <-filter(distribution, USA_States %in% c("a","e","i","o","u"))
> vowel_dist
USA_States Freq
1      a      61
2      e      28
3      i      44
4      o      36
5      u       8
> barplot(vowel_dist$Freq, axes = TRUE, axisnames = TRUE, xlab = "Vowels",
ylab = "frequency" )
```

