>

# Session 5 - Data Management using R - Assignment - 1

> #Problem Statement

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

> View(USArrests)

> names(USArrests) #to know the names of the columns

[1] "Murder" "Assault" "UrbanPop" "Rape"

> US\_States<-rownames(USArrests) # to get the row names, i.e the USA states

> US\_States

[1] "Alabama" "Alaska" "Arizona" "Arkansas"

[5] "California" "Colorado" "Connecticut" "Delaware"

[9] "Florida" "Georgia" "Hawaii" "Idaho"

[13] "Illinois" "Indiana" "Iowa" "Kansas"

[17] "Kentucky" "Louisiana" "Maine" "Maryland"

[21] "Massachusetts" "Michigan" "Minnesota" "Mississippi"

[25] "Missouri" "Montana" "Nebraska" "Nevada"

[29] "New Hampshire" "New Jersey" "New Mexico" "New York"

[33] "North Carolina" "North Dakota" "Ohio" "Oklahoma"

[37] "Oregon" "Pennsylvania" "Rhode Island" "South Carolina"

[41] "South Dakota" "Tennessee" "Texas" "Utah"

[45] "Vermont" "Virginia" "Washington" "West Virginia"

[49] "Wisconsin" "Wyoming"

> b<-gsub("[^aeiouAEIOU]","C",US\_States)#is basically "if not any of 'aeiouAEIOU' then 'C'

> #if aeiouAEIOU is NOT in [], its a complete pattern.

> # using ^ means, exclude what comes after this

> b<-gsub("[^C]","V",b)# is "if not 'C', then 'V

> b

[1] "VCVCVCV" "VCVCCV" "VCVCVCV" "VCCVCCVC"

[5] "CVCVCVCCVV" "CVCVCVCV" "CVCCVCCVCVC" "CVCVCVCV"

[9] "CCVCVCV" "CVVCCVV" "CVCVVV" "VCVCV"

[13] "VCCVCVVC" "VCCVVCV" "VVCV" "CVCCVC"

[17] "CVCCVCCC" "CVVVCVVCV" "CVVCV" "CVCCCVCC"

[21] "CVCCVCCVCVCCC" "CVCCVCVC" "CVCCVCVCV" "CVCCVCCVCCV"

[25] "CVCCVVCV" "CVCCVCV" "CVCCVCCV" "CVCVCV"

[29] "CVCCCVCCCCVCV" "CVCCCVCCVC" "CVCCCVCVCV" "CVCCCVCC"

[33] "CVCCCCCVCVCVCV" "CVCCCCCVCVCV" "VCVV" "VCCVCVCV"

[37] "VCVCVC" "CVCCCCCCVCVV" "CCVCVCVCCVCC" "CVVCCCCVCVCVCV"

[41] "CVVCCCCVCVCV" "CVCCVCCVV" "CVCVC" "VCVC"

[45] "CVCCVCC" "CVCCVCVV" "CVCCVCCCVC" "CVCCCCVCCVCVV"

[49] "CVCCVCCVC" "CCVCVCC"

> library(stringr)

> str\_count(b,"V") #counts charecter in a string

[1] 4 3 4 3 5 4 4 4 3 4 4 3 4 4 3 2 2 6 3 2 4 3 4 4 4 3 3 3 4 3 4 2 5 4 3 4 3 4 4 6

[41] 5 4 2 2 2 4 3 5 3 2

> sum(str\_count(b,"V")) # sums the count of "V" i.e vowels

[1] 177