**NAME: SANDEEP PATIL**

**Email: saashasandeep@gmail.com**

**DATA ANALYTICS WITH R, EXCEL and TABLEAU**

**Session 5 – Assignment – 5.1**

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

**Answer:** View(USArrests)

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

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

US\_States

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

library(stringr)

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

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

177 Vowels

WE CAN ALTERNATIVELY MERGE THE 50 NAMES IN ONE STRING AND AVOID "SUM"

c<-paste(b,collapse = "")

str\_count(c,"V") # gives 177 straight away

str\_count(b,"C")

sum(str\_count(b,"C"))

245 consonants

ONE MORE WAY OF SOLVING THE SAME PROBLEM

USArrests # data set

names(USArrests) # to know what is the column name for State names

USA\_States <- rownames(USArrests) # names of states

USA\_States <- paste(USA\_States, collapse = "") # converting the names to a string

USA\_States

USA\_States <- tolower(strsplit(USA\_States, "")[[1]]) # converting to lower case and spliting each letter

USA\_States

USA\_States <- USA\_States[USA\_States %in% letters]

table(USA\_States) # Frequency of each letter

distribution <- as.data.frame(table(USA\_States)) # converted to data frame

names(distribution)

colnames(distribution) <- c("letters", "Freq") # changed column names

names(distribution)

library(dplyr)

filter(distribution, letters %in% c("a","e","i","o","u")) # finding number of vowels in the names of USA States

thus there are 61 a, 28 e , 44 i , 36 o and 8 u in the names of USA States

1. Visualize the vowels distribution.

**Answer:** type 1

vowel\_dist <- filter(distribution, letters %in% c("a","e","i","o","u"))

vowel\_dist

barplot(vowel\_dist$Freq, axes = TRUE, axisnames = TRUE, xlab = "Vowels", ylab = "frequency")

type 2

since we want vowel distribution, assiging colour to position of vowel

area.color <- c("withcolour",NA,NA,NA,"withcolour",NA,NA,NA,"withcolour",

NA,NA,NA,NA,NA,"withcolour",NA,NA,NA,NA,"withcolour",

NA,NA,NA,NA,NA)

area.color

library(ggplot2)

plot.vowel <- ggplot(data = distribution, aes(x=letters, y=Freq, fill=area.color))+

geom\_bar(stat = "identity") +

xlab(colnames(distribution)[1]) +

ylab(colnames(distribution)[2])

plot.vowel # vowel distribution is visualized with different colour