SESSION 5

AASIGNMENT 1

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

Solution:

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

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

ANSWER:

letters Freq

1 a 61

2 e 28

3 i 44

4 o 36

5 u 8

2. Visualize the vowels distribution.

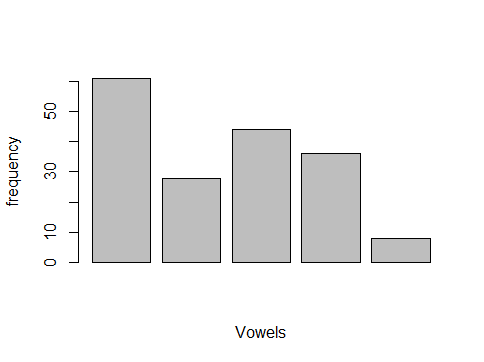
Solution:

# 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

# assiging colour to position of vowel

Install.pacakages(“ggplot2”)

Library(“ggplot2”)

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

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

