Question – 1:

```r
# Load the diabetes dataset (assuming it is available in the current working directory)

diabetes <- read.csv("diabetes.csv")

# (a) Number of rows and columns

num_rows <- nrow(diabetes)

num_cols <- ncol(diabetes)

cat("Number of rows:", num_rows, "\n")

cat("Number of columns:", num_cols, "\n")


# (b) Column names

col_names <- names(diabetes)

cat("Column names:", col_names, "\n")


# (c) Average glyhb

average_glyhb <- mean(diabetes$glyhb)

cat("Average glyhb:", average_glyhb, "\n")


# (d) Column types

column_types <- sapply(diabetes, class)

cat("Column types:", column_types, "\n")


# (e) Summary of the dataframe

summary(diabetes)


# (f) Access the age column

age_column <- diabetes$age


# (g) Number of people whose age is greater than 40

num_people_age_gt_40 <- sum(diabetes$age > 40)

cat("Number of people with age > 40:", num_people_age_gt_40, "\n")
```

 # (h) Female diabetic patients of age >= 30

female_diabetic_age_30 <- subset(diabetes, diabetes$gender == "Female" & diabetes$age >= 30)

cat("Female diabetic patients with age >= 30:\n")

print(female_diabetic_age_30)

 # (i) Patients not from Louisa

patients_not_from_louisa <- subset(diabetes, diabetes$location != "Louisa")

cat("Patients not from Louisa:\n")

print(patients_not_from_louisa)

```
> # Load the diabetes dataset (assuming it is available in the current working directory)
> diabetes <- read.csv("diabetes.csv")
> # (a) Number of rows and columns
> num_rows <- nrow(diabetes)
> num_cols <- ncol(diabetes)
> cat("Number of rows:", num_rows, "\n")
Number of rows: 403
> cat("Number of columns:", num_cols, "\n")
Number of columns: 24
> # (b) Column names
> col_names <- names(diabetes)
> cat("Column names:", col_names, "\n")
Column names: X id chol stab.glu hdl ratio glyhb location age gender height weight frame bp.1s bp.1d bp.2s bp.2d waist
hip time.ppn insurance fh smoking dm
> # (c) Average glyhb
> average_glyhb <- mean(diabetes$glyhb)
> cat("Average glyhb:", average_glyhb, "\n")
Average glyhb: NA
> # (d) Column types
> column_types <- sapply(diabetes, class)
> cat("Column types:", column_types, "\n")
Column types: integer integer integer integer integer integer numeric numeric character integer character integer integer chara
cter integer integer integer integer integer integer integer integer integer character
> # (e) Summary of the dataframe
> summary(diabetes)
       X                id              chol          stab.glu          hdl             ratio           glyhb
 Min.   :  1.0   Min.   : 1000   Min.   : 78.0   Min.   : 48.0   Min.   : 12.00   Min.   : 1.500   Min.   : 2.68
 1st Qu.:101.5   1st Qu.: 4792   1st Qu.:179.0   1st Qu.: 81.0   1st Qu.: 38.00   1st Qu.: 3.200   1st Qu.: 4.38
 Median :202.0   Median :15766   Median :204.0   Median : 89.0   Median : 46.00   Median : 4.200   Median : 4.84
 Mean   :202.0   Mean   :15978   Mean   :207.8   Mean   :106.7   Mean   : 50.45   Mean   : 4.522   Mean   : 5.59
 3rd Qu.:302.5   3rd Qu.:20336   3rd Qu.:230.0   3rd Qu.:106.0   3rd Qu.: 59.00   3rd Qu.: 5.400   3rd Qu.: 5.60
 Max.   :403.0   Max.   :41756   Max.   :443.0   Max.   :385.0   Max.   :120.00   Max.   :19.300   Max.   :16.11
                                 NA's   :1                       NA's   :1        NA's   :1        NA's   :13
   location             age           gender             height          weight           frame
 Length:403        Min.   :19.00   Length:403        Min.   :52.00   Min.   : 99.0   Length:403
 Class :character   1st Qu.:34.00   Class :character   1st Qu.:63.00   1st Qu.:151.0   Class :character
 Mode  :character   Median :45.00   Mode  :character   Median :66.00   Median :172.5   Mode  :character
                    Mean   :46.85                      Mean   :66.02   Mean   :177.6
                    3rd Qu.:60.00                      3rd Qu.:69.00   3rd Qu.:200.0
                    Max.   :92.00                      Max.   :76.00   Max.   :325.0
                                                       NA's   :5       NA's   :1
```

```
      bp.1s           bp.1d           bp.2s           bp.2d            waist            hip            time.ppn
 Min.   : 90.0   Min.   : 48.00   Min.   :110.0   Min.   : 60.00   Min.   :26.0   Min.   :30.00   Min.   :   5.0
 1st Qu.:121.2   1st Qu.: 75.00   1st Qu.:138.0   1st Qu.: 84.00   1st Qu.:33.0   1st Qu.:39.00   1st Qu.:  90.0
 Median :136.0   Median : 82.00   Median :149.0   Median : 92.00   Median :37.0   Median :42.00   Median : 240.0
 Mean   :136.9   Mean   : 83.32   Mean   :152.4   Mean   : 92.52   Mean   :37.9   Mean   :43.04   Mean   : 341.2
 3rd Qu.:146.8   3rd Qu.: 90.00   3rd Qu.:161.0   3rd Qu.:100.00   3rd Qu.:41.0   3rd Qu.:46.00   3rd Qu.: 517.5
 Max.   :250.0   Max.   :124.00   Max.   :238.0   Max.   :124.00   Max.   :56.0   Max.   :64.00   Max.   :1560.0
 NA's   :5       NA's   :5        NA's   :262     NA's   :262      NA's   :2      NA's   :2       NA's   :3
    insurance            fh            smoking            dm
 Min.   :0.000   Min.   :0.0000   Min.   :1.000   Length:403
 1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:1.000   Class :character
 Median :1.000   Median :0.0000   Median :2.000   Mode  :character
 Mean   :1.055   Mean   :0.1886   Mean   :1.846
 3rd Qu.:2.000   3rd Qu.:0.0000   3rd Qu.:2.000
 Max.   :2.000   Max.   :1.0000   Max.   :3.000

> # (f) Access the age column
> age_column <- diabetes$age
> # (g) Number of people whose age is greater than 40
> num_people_age_gt_40 <- sum(diabetes$age > 40)
> cat("Number of people with age > 40:", num_people_age_gt_40, "\n")
Number of people with age > 40: 243
> # (h) Female diabetic patients of age >= 30
> female_diabetic_age_30 <- subset(diabetes, diabetes$gender == "Female" & diabetes$age >= 30)
> cat("Female diabetic patients with age >= 30:\n")
Female diabetic patients with age >= 30:
> print(female_diabetic_age_30)
 [1] X          id         chol       stab.glu   hdl        ratio      glyhb      location   age        gender     height
[12] weight     frame      bp.1s      bp.1d      bp.2s      bp.2d      waist      hip        time.ppn   insurance  fh
[23] smoking    dm
<0 rows> (or 0-length row.names)
> # (i) Patients not from Louisa
> patients_not_from_louisa <- subset(diabetes, diabetes$location != "Louisa")
> cat("Patients not from Louisa:\n")
Patients not from Louisa:
> print(patients_not_from_louisa)
     X   id chol stab.glu hdl ratio glyhb   location age gender height weight  frame bp.1s bp.1d bp.2s bp.2d waist
1    1 1000  203       82  56   3.6  4.31 Buckingham  46 female     62    121 medium   118    59    NA    NA    29
2    2 1001  165       97  24   6.9  4.44 Buckingham  29 female     64    218  large   112    68    NA    NA    46
3    3 1002  228       92  37   6.2  4.64 Buckingham  58 female     61    256  large   190    92   185    92    49
4    4 1003   78       93  12   6.5  4.63 Buckingham  67   male     67    119  large   110    50    NA    NA    33

   hip time.ppn insurance fh smoking  dm
1   38      720         1  0       3  no
2   48      360         0  0       2  no
3   57      180         2  0       2  no
4   38      480         1  0       3  no
5   41      300         0  0       3 yes
6   42      195         1  0       1  no
7   49      720         2  1       2  no
8   39     1020         0  0       2  no
9   40      300         2  1       1  no
10  50      240         2  0       2  no
20  51      540         0  0       2 yes
21  41     1020         2  0       2  no
22  46      480         0  0       2  no
23  55      240         1  1       2 yes
24  44      285         1  0       1  no
25  38      210         2  0       1  no
26  36      780         1  0       2  no
27  47      420         1  1       2  no
28  33      510         0  1       1  no
29  38      720         2  0       3  no
31  40      540         1  0       2  no
32  35      780         1  0       1  no
33  41      225         0  0       2 yes
34  43      330         0  0       3  no
35  34      720         1  0       2  no
36  44      780         2  1       2  no
37  35      720         1  0       1  no
38  49      240         2  0       2  no
39  36       30         0  0       3  no
40  48      840         2  0       3 yes
41  40      720         1  0       2  no
```

 # output isn't shown completely here… execute it!

Question – 2:

 # Create a 2x3 matrix M with randomly selected values between 10 and 30

M <- matrix(sample(10:30, 6, replace = TRUE), nrow = 2, byrow = TRUE)

 # Create a vector V with three values

V <- c(20, 25, 30)

 # Add V as the third row of M

M <- rbind(M, V)

 # Print the matrix M

print(M)

```
> source("~/.active-rstudio-document")
  [,1] [,2] [,3]
    13   12   26
    27   29   18
V   20   25   30
```

Question – 3:

 # Create a numeric vector

x <- c(1, 2, 3, 4, 5)

 # Add zeros at starting and ending of x

x <- c(0, 0, x, 0, 0)

 # Print the vector x

print(x)

```
> source("~/.active-rstudio-document")
[1] 0 0 1 2 3 4 5 0 0
```

Question – 4:

```
# Load the iris dataset

library(datasets)

data(iris)

# Set the number of clusters

k <- 3

# Initialize the cluster centroids

centroids <- kmeans(iris[, 1:4], centers = k, nstart = 10)$centers

# Assign each data point to a cluster

cluster_assignments <- kmeans(iris[, 1:4], centers = centroids)$cluster

#print(cluster_assignments)

# Plot the data points with their cluster assignments

plot(iris[, 1:4], col = c("red", "green", "blue")[cluster_assignments], pch = 19)

points(centroids, pch = 1, cex = 1.5, col = "black")
```