ITA0435 - Statistics with R Program

Lab Experiments Output

Name: Srinidhi A
Reg.no: 192324277

1. Create numeric, character, and logical vectors and display type and content.

```
> source("~/Documents/R program Lab/exp1.R")
[1] 1 2 3 4 5 6
[1] "Class: numeric"
[1] "Type: double"
[1] "Apple"  "Banana" "Orange"
[1] "Class: character"
[1] "Type: character"
[1]  TRUE FALSE  TRUE FALSE
[1] "Class: logical"
[1] "Type: logical"
>
```

2. Create labeled matrices (5×4, 3×3, 2×2) filled by row/column.

```
> a=matrix(1:20,nrow=5,ncol=4)
> a
     [,1] [,2] [,3] [,4]
[1,]    1    6   11   16
[2,]    2    7   12   17
[3,]    3    8   13   18
[4,]    4    9   14   19
[5,]    5   10   15   20
> b=matrix(1:9,nrow=3,ncol=3)
> b
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
[3,]    3    6    9
> c=matrix(1:4,nrow=2,ncol=2)
> c
     [,1] [,2]
[1,]    1    3
[2,]    2    4
>
```

3. Write an R program to create and display a 3D array with specified rows, columns, and tables.

```
, , 1

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

, , 2

     [,1] [,2] [,3]
[1,]    7    9   11
[2,]    8   10   12

, , 3

     [,1] [,2] [,3]
[1,]   13   15   17
[2,]   14   16   18

, , 4

     [,1] [,2] [,3]
[1,]   19   21   23
[2,]   20   22   24
```

4. Create arrays from vectors with dimension names, print specific elements.

```
> source("~/Documents/R program Lab/exp4.R")
, , T1

   C1 C2
R1  1  4
R2  2  5
R3  3  6

, , T2

   C1 C2
R1  7 10
R2  8 11
R3  9 12

Element [R2,C1,T1]: 2
Element [R3,C2,T2]: 12
```

5. Create and manipulate factor variables (e.g., women's dataset heights, random LETTERS sample).

```
> source("~/Documents/R program Lab/exp5.R")
[1] Short  Medium Tall   Medium Tall   Short
Levels: Medium Short Tall
[1] C A B D C B B B
Levels: A B C D
>
```

6. Create an R list containing vectors, matrices, and functions; display contents.

```
> source("~/Documents/R program Lab/exp6.R")
$Vector
[1] 1 2 3 4

$Matrix
     [,1] [,2] [,3]
[1,]    1    3    5
[2,]    2    4    6

$Function
function(x) x^2
```

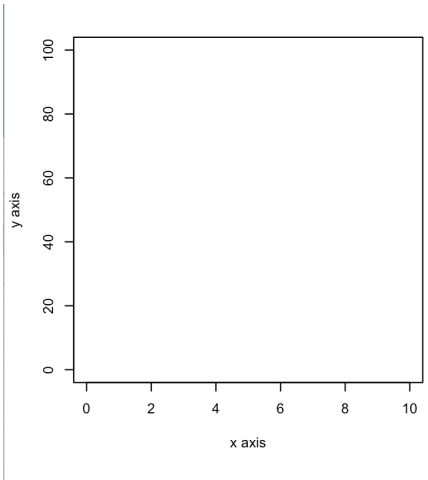
7. Write R programs for basic tasks: Factors of a number, generate a vector of 10 random integers between -50 and 50, print numbers 1–100 with FizzBuzz logic.

```
> source("~/Documents/R program Lab/exp7.R")
[1] "Factors of the number:"
[1]  1  2  3  4  6 12
[1] "10 random integers:"
 [1] -20  28   0 -37  16  -9  -1  -8  50 -37
[1] 1
[1] 2
[1] "Fizz"
[1] 4
[1] "Buzz"
[1] "Fizz"
[1] 7
[1] 8
[1] "Fizz"
[1] "Buzz"
[1] 11
[1] "Fizz"
[1] 13
[1] 14
```

8. Generate random numbers from a normal distribution; count occurrences.

```
> source("~/Documents/R program Lab/exp8.R")
  Var1 Freq
1   -2    4
2   -1   21
3    0   45
4    1   22
5    2    8
>
```

9. Create empty plots with specified axis limits.



10. Create and explore a data frame exam_data with name, score, attempts, and qualify fields. Perform extract, add row/column, sort, save to file.

```
> source("~/Documents/R program Lab/exp10.R")
    name score attempts qualify grade
3 Charan    90        2     yes    A+
5   Esha    88        1     yes     A
1    Anu    85        1     yes     A
4  Divya    75        2      no     B
6 Farhan    70        3      no     B
2   Bala    62        3      no     C
>
```

11. Write an R program to read a .csv file and display contents.

```
> source("~/Documents/R program Lab/exp11.R")
[1] "Contents of the CSV file:"
     name score attempts qualifying grade
1  Vikram    95        1       TRUE     A
2  Raghul    92        2       TRUE     A
3    Arun    88        3       TRUE     B
4   Gokul    85        1       TRUE     B
5 Dhanush    78        1      FALSE     C
>
```

12. Perform data reshaping on airquality dataset: melt, cast, compute monthly averages for Ozone, Solar.R, Wind, and Temperature.

```
> source("~/Documents/R program Lab/exp12.R")
[1] "Monthly average values:"
  Month    Ozone  Solar.R      Wind     Temp
1     5 24.12500 182.0417 11.504167 66.45833
2     6 29.44444 184.2222 12.177778 78.22222
3     7 59.11538 216.4231  8.523077 83.88462
4     8 60.00000 173.0870  8.860870 83.69565
5     9 31.44828 168.2069 10.075862 76.89655
>
```

13. Combine multiple arrays row-wise.

```
> source("~/Documents/R program Lab/exp13.R")
[1] "Combined arrays (row-wise):"
  [,1] [,2] [,3]
a    1    2    3
b    4    5    6
c    7    8    9
>
```

14. Explore and manipulate ChickWeight dataset (sorting, melting, casting by Diet).

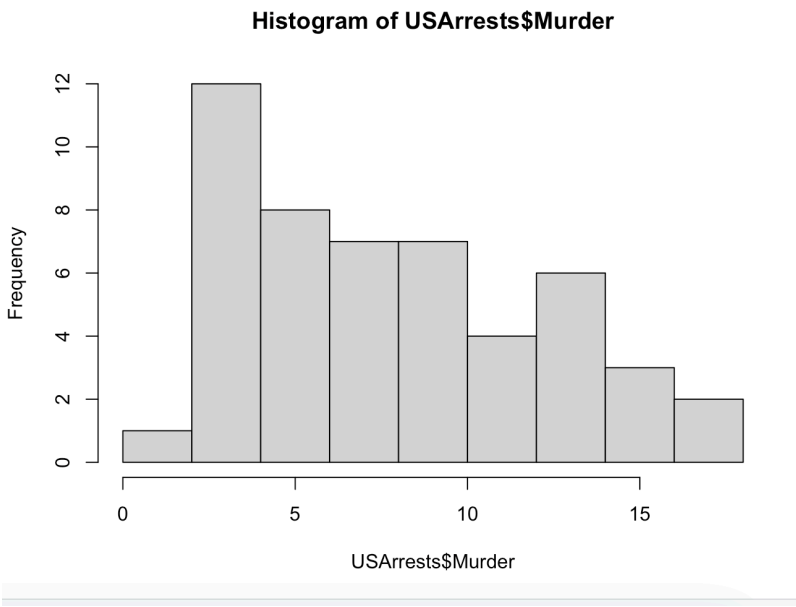
```
> source("~/Documents/R program Lab/exp14.R")
    weight Time Chick Diet
196     35    2    18    1
26      39    2     3    1
195     39    0    18    1
293     39    0    27    2
305     39    0    28    2
317     39    0    29    2
365     39    0    33    3
401     39    0    36    3
543     39    0    48    4
13      40    0     2    1
221     40    0    21    2
269     40    0    25    2
519     40    0    46    4
555     40    0    49    4
49      41    0     5    1
61      41    0     6    1
73      41    0     7    1
108     41    0    10    1
132     41    0    12    1
144     41    0    13    1
156     41    0    14    1
168     41    0    15    1
176     41    0    16    1
209     41    0    20    1
233     41    0    22    2
353     41    0    32    3
377     41    0    34    3
389     41    0    35    3
413     41    0    37    3
425     41    0    38    3
449     41    0    40    3
507     41    0    45    4
531     41    0    47    4
```

15. Perform EDA on iris dataset: dimensions, summary, standard deviation, quantiles, grouping by Species, pivot table, categorical grouping with Sepal.Length categories.

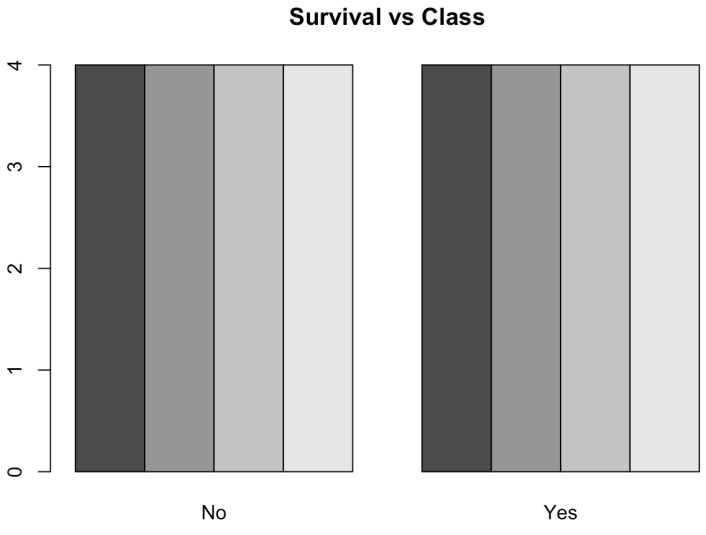
```
> data("iris")
> dim(iris)
[1] 150   5
> summary(iris)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
> sapply(iris[,1:4], sd)
Sepal.Length  Sepal.Width Petal.Length  Petal.Width
   0.8280661    0.4358663    1.7652982    0.7622377
> quantile(iris$Sepal.Length)
  0%  25%  50%  75% 100%
 4.3  5.1  5.8  6.4  7.9
> aggregate(. ~ Species, iris, mean)
     Species Sepal.Length Sepal.Width Petal.Length Petal.Width
1     setosa        5.006       3.428        1.462       0.246
2 versicolor        5.936       2.770        4.260       1.326
3  virginica        6.588       2.974        5.552       2.026
>
```

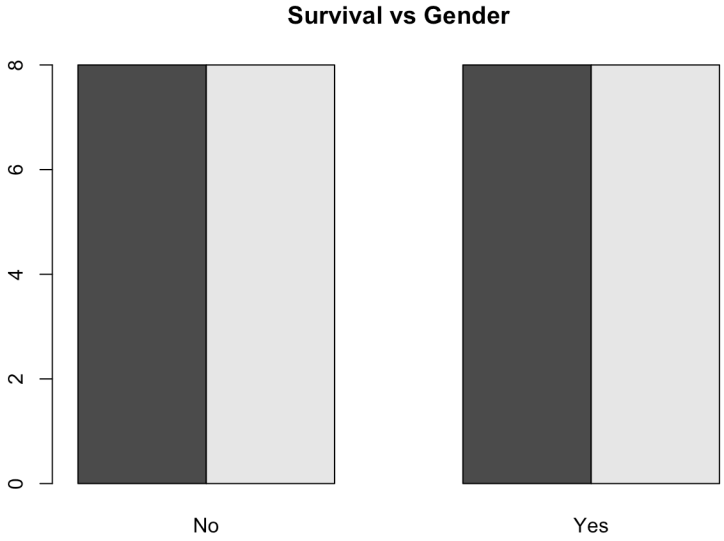
16. Explore USArrests dataset: summary statistics, state with largest rape arrests, max & min murder rates, correlation among features, states above median assault arrests and bottom 25% for murder, visualization with histogram, density, scatterplots, bar graphs.

```
> data("USArrests")
> summary(USArrests)
     Murder          Assault         UrbanPop          Rape
 Min.   : 0.800   Min.   : 45.0   Min.   :32.00   Min.   : 7.30
 1st Qu.: 4.075   1st Qu.:109.0   1st Qu.:54.50   1st Qu.:15.07
 Median : 7.250   Median :159.0   Median :66.00   Median :20.10
 Mean   : 7.788   Mean   :170.8   Mean   :65.54   Mean   :21.23
 3rd Qu.:11.250   3rd Qu.:249.0   3rd Qu.:77.75   3rd Qu.:26.18
 Max.   :17.400   Max.   :337.0   Max.   :91.00   Max.   :46.00
> which.max(USArrests$Rape)
[1] 28
> max(USArrests$Murder)
[1] 17.4
> min(USArrests$Murder)
[1] 0.8
> cor(USArrests)
             Murder   Assault   UrbanPop      Rape
Murder   1.00000000 0.8018733 0.06957262 0.5635788
Assault  0.80187331 1.0000000 0.25887170 0.6652412
UrbanPop 0.06957262 0.2588717 1.00000000 0.4113412
Rape     0.56357883 0.6652412 0.41134124 1.0000000
> hist(USArrests$Murder)
```

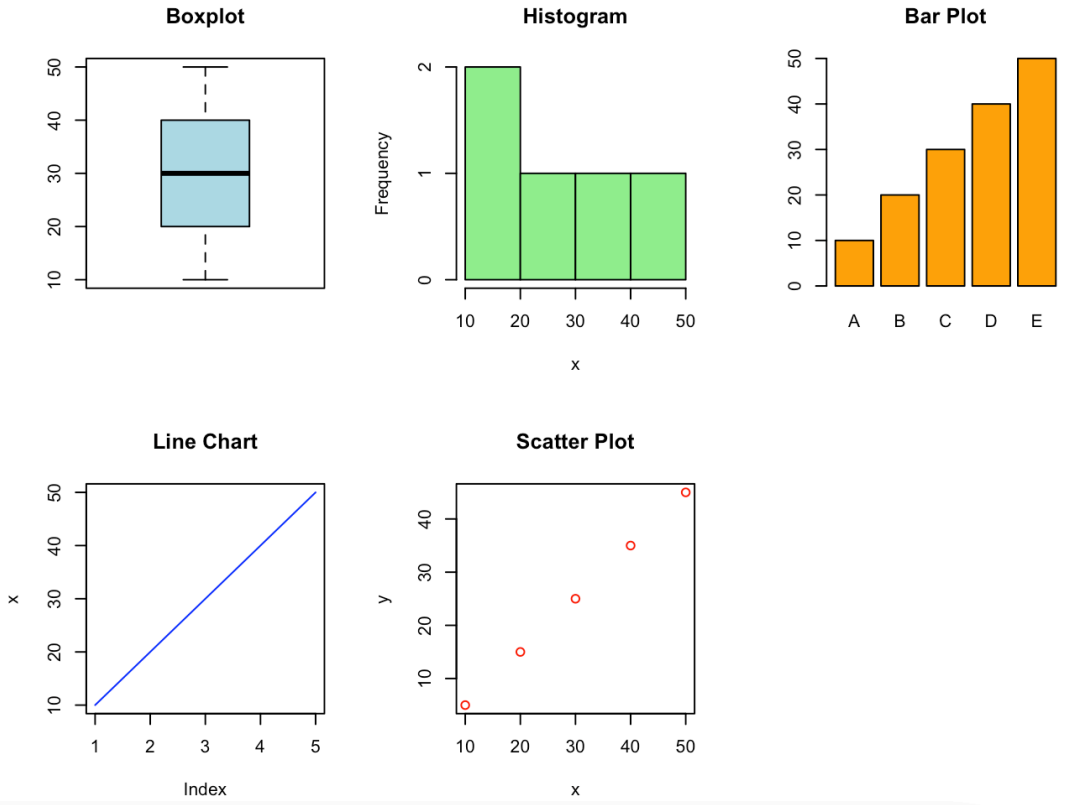


17. Explore Titanic dataset: bar chart of survival vs class, modify plot by gender, histogram of Age.





18. Create graphs in R: boxplot, histogram, bar plot, line chart, scatter plot.



19. Build a regression model on advertising dataset (Sales ~ Spend) and predict Sales.

```
> predict(model, data.frame(Spend=60))
 1
65
>
```

20. Create multiple regression model using ChickWeight dataset with "Time" and "Diet" as predictors; predict weight and compute model error.

```
> source("~/Documents/R program Lab/exp20.R")
> data("ChickWeight")
>
> m <- lm(weight ~ Time + Diet, data = ChickWeight)
> p <- predict(m)
> mean((ChickWeight$weight - p)^2)
[1] 1284.319
```

21. Randomly split iris dataset into train/test (80/20), build logistic regression (Species ~ Petal.Length + Petal.Width), predict, and evaluate with confusion matrix.

```
> p <- ifelse(predict(m, type="response") > 0.5, "versicolor", "setosa")
> table(p, iris$Species)

p            setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         50         0
>
```