

PRACTICAL NO 3

AIM:- Exploring data: View() or print() (R)

OUTPUT :-

1.The first few rows

The screenshot displays the RStudio interface with the following components:

- Script Editor:** Contains the following R code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read.csv("Breast_Cancer.csv")
7
8 # View the first few rows
9 head(my_data)
```
- Console:** Shows the output of the code execution:

```
R - R 4.5.2 - C:/Users/as993/DATA SET/
> # View the first few rows
> head(my_data)
  Age Race Marital.Status T.Stage N.Stage X6th.Stage differentiate Grade A.Stage
1  68 White      Married    T1      N1      IIA      Poorly differentiated  3 Regional
2  50 White      Married    T2      N2      IIIA Moderately differentiated  2 Regional
3  58 White    Divorced    T3      N3      IIIC Moderately differentiated  2 Regional
4  58 White      Married    T1      N1      IIA      Poorly differentiated  3 Regional
5  47 White      Married    T2      N1      IIB      Poorly differentiated  3 Regional
6  51 White    Single      T1      N1      IIA Moderately differentiated  2 Regional
Tumor.Size Estrogen.Status Progesterone.Status Regional.Node.Examined Regionol.Node.Positive
1         4      Positive          Positive                24                    1
2        35      Positive          Positive                14                    5
3        63      Positive          Positive                14                    7
4        18      Positive          Positive                 2                    1
5        41      Positive          Positive                 3                    1
6        20      Positive          Positive                18                    2
Survival.Months Status
1         60    Alive
2         62    Alive
3         75    Alive
4         84    Alive
5         50    Alive
6         89    Alive
```
- Environment:** Shows the loaded data frame: `Breast_Cancer` with 4024 observations and 16 variables. The variables listed are Age, Race, Marital.Status, T.Stage, N.Stage, X6th.Stage, differentiate, Grade, and A.Stage.
- Files:** Lists the files in the current directory, including `Breast_Cancer.csv`, `DATA.SET.Rproj`, `PRACTICAL NO 1.R`, and `SALARY.xlsx`.

2.The last few rows

The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read.csv("Breast_Cancer.csv")
7
8
9 # View the last few rows
10 tail(my_data)
11
```

The console on the bottom left shows the output of the `tail(my_data)` command, displaying the last 10 rows of the dataset. The Environment pane on the right shows the `Breast_Cancer` data frame with 4024 observations and 16 variables.

	Age	Race	Marital.Status	T.Stage	N.Stage	X6th.Stage	differentiate	Grade
4019	64	White	Married	T1	N1	IIA	Moderately differentiated	2
4020	62	Other	Married	T1	N1	IIA	Moderately differentiated	2
4021	56	White	Divorced	T2	N2	IIIA	Moderately differentiated	2
4022	68	White	Married	T2	N1	IIB	Moderately differentiated	2
4023	58	Black	Divorced	T2	N1	IIB	Moderately differentiated	2
4024	46	White	Married	T2	N1	IIB	Moderately differentiated	2

3.The dimensions (rows and columns)

The screenshot shows the RStudio interface. The script editor on the left contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read.csv("Breast_Cancer.csv")
7
8
9 #the dimensions (rows and columns)
10 dim(my_data)
11
12
```

The console on the bottom left shows the output of the `dim(my_data)` command, displaying the dimensions of the dataset: 4024 rows and 16 columns.

```
Dimensions (Rows, Columns): 4024 16
```

The Environment pane on the right shows the `Breast_Cancer` data frame with 4024 observations and 16 variables.

4.The structure (variable types and number of observations)

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read.csv("Breast_Cancer.csv")
7
8
9
10 # Get the structure (variable types and number of observations)
11 str(my_data)
12
```

The console output shows the structure of the data frame:

```
'data.frame': 4024 obs. of 16 variables:
 $ Age      : int  68 50 58 58 47 51 51 40 40 69 ...
 $ Race     : chr  "white" "white" "white" "white" ...
 $ Marital.Status : chr  "Married" "Married" "Divorced" "Married" ...
 $ T.Stage  : chr  "I1" "I2" "I3" "I1" ...
 $ N.Stage  : chr  "N1" "N2" "N3" "N1" ...
 $ X6th.Stage : chr  "IIA" "IIIA" "IIIC" "IIA" ...
 $ differentiate : chr  "Poorly differentiated" "Moderately differentiated" "Moderately differentiated" ...
 $ Grade    : chr  "3" "2" "3" ...
 $ A.Stage  : chr  "Regional" "Regional" "Regional" "Regional" ...
 $ Tumor.Size : int  4 35 63 18 41 20 8 30 103 32 ...
 $ Estrogen.Status : chr  "Positive" "Positive" "Positive" "Positive" ...
 $ Progesterone.Status : chr  "Positive" "Positive" "Positive" "Positive" ...
 $ Regional.Node.Examined : int  24 14 14 2 3 18 11 9 20 21 ...
 $ Regio1.Node.Positive : int  1 5 7 1 1 2 1 1 18 12 ...
 $ Survival.Months : int  60 62 75 84 50 89 54 14 70 92 ...
 $ Status    : chr  "Alive" "Alive" "Alive" "Alive" ...
```

The Environment pane shows the 'Breast_Cancer' data frame with 4024 observations and 16 variables.

5.Summary of the dataset

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read.csv("Breast_Cancer.csv")
7 # See a summary of the dataset
8 summary(my_data)
9
```

The console output shows the summary of the data frame:

```
Age      Race      Marital.Status  T.Stage  N.Stage
Min.   :30.00  Length:4024  Length:4024  Length:4024  Length:4024
1st Qu.:47.00  Class :character  Class :character  Class :character  Class :character
Median :54.00  Mode  :character  Mode  :character  Mode  :character  Mode  :character
Mean   :53.97
Max.   :69.00
X6th.Stage  differentiate  Grade  A.Stage  Tumor.Size
Length:4024  Length:4024  Length:4024  Length:4024  Min.   : 1.00
1st Qu.:1.00  1st Qu.:1.00  1st Qu.:1.00  1st Qu.:1.00  1st Qu.:16.00
Median :14.00  Median :14.00  Median :14.00  Median :14.00  Median :25.00
Mean   :14.36  Mean   :14.36  Mean   :14.36  Mean   :14.36  Mean   :30.47
3rd Qu.:19.00  3rd Qu.:19.00  3rd Qu.:19.00  3rd Qu.:19.00  3rd Qu.:38.00
Max.   :61.00  Max.   :61.00  Max.   :61.00  Max.   :61.00  Max.   :140.00

Estrogen.Status  Progesterone.Status  Regional.Node.Examined  Regio1.Node.Positive
Length:4024  Length:4024  Length:4024  Length:4024
Class :character  Class :character  Class :character  Class :character
Mode  :character  Mode  :character  Mode  :character  Mode  :character
Survival.Months  Status
Min.   : 1.0  Length:4024
1st Qu.:56.0  Class :character
Median :73.0  Mode  :character
Mean   :71.3
3rd Qu.:90.0
```

The Environment pane shows the 'my_data' data frame with 4024 observations and 16 variables.

6.The column names

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read_csv("Breast_Cancer.csv")
7
8 # Get the column names
9 names(my_data)
10 cat("Column Names: ", names(my_data), "\n")
```

The console output shows the column names: "Age Race Marital.Status T.Stage N.Stage X6th.Stage differentiate Grade A.Stage Tumor.Size Estrogen. Status Progesterone.Status Regional.Node.Examined Reginal.Node.Positive Survival.Months Status".

The Environment pane shows the data frame 'my_data' with 4024 observations and 16 variables. The variables are: Tumor.Size, Estrogen.Status, Progesterone.Status, Regional.Node.Examined, Reginal.Node.Positive, Survival.Months, Status, and 9 other variables.

7. 'psych' package for more detailed descriptive statistics

'describe()' provides: n, mean, sd, median, trimmed mean, mad, min, max, range, skew, kurtosis, and se.

The screenshot shows the RStudio interface. The script editor contains the following code:

```
1 install.packages(c("readr", "psych"))
2
3 library(readr) # For efficient data reading
4 library(psych) # For descriptive statistics
5
6 my_data <- read_csv("Breast_Cancer.csv")
7
8 # Use the 'psych' package for more detailed descriptive statistics
9 # 'describe()' provides: n, mean, sd, median, trimmed mean, mad, min, max, range, skew, kurtosis, and se.
10 describe(my_data)
```

The console output shows the following detailed descriptive statistics:

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
Age	1	4024	53.97	8.96	54	54.25	10.38	30	69	39	-0.22	-0.76	0.14
Race*	2	4024	2.78	0.56	3	2.94	0.00	1	3	2	-2.41	4.39	0.01
Marital.Status*	3	4024	2.37	1.06	2	2.27	0.00	1	5	4	1.15	0.36	0.02
T.Stage*	4	4024	1.78	0.77	2	1.70	1.48	1	4	3	0.72	0.06	0.01
N.Stage*	5	4024	1.44	0.69	1	1.30	0.00	1	3	2	1.28	0.21	0.01
X6th.Stage*	6	4024	2.32	1.27	2	2.15	1.48	1	5	4	0.81	-0.20	0.02
differentiate*	7	4024	1.69	1.02	1	1.49	0.00	1	4	3	1.42	0.74	0.02
Grade*	8	4024	3.13	0.64	3	3.17	0.00	1	4	3	-0.23	-0.21	0.01
A.Stage*	9	4024	1.98	0.15	2	2.00	0.00	1	2	1	-6.38	38.74	0.00
Tumor.Size	10	4024	30.47	21.12	25	27.11	14.83	1	140	139	1.74	3.62	0.33
Estrogen.Status*	11	4024	1.93	0.25	2	2.00	0.00	1	2	1	-3.47	10.02	0.00
Progesterone.Status*	12	4024	1.83	0.38	2	1.91	0.00	1	2	1	-1.72	0.97	0.01
Regional.Node.Examined	13	4024	14.36	8.10	14	13.85	7.41	1	61	60	0.83	1.64	0.13
Reginal.Node.Positive	14	4024	4.16	5.11	2	2.96	1.48	1	46	45	2.70	8.96	0.08
Survival.Months	15	4024	71.30	22.92	73	72.97	25.20	1	107	106	-0.59	0.02	0.36
Status*	16	4024	1.15	0.36	1	1.07	0.00	1	2	1	1.93	1.71	0.01