

SHETH LUJ AND SIR MV COLLEGE

Subject: Data Analysis with SAS / SPSS / R

Practical No: 3

Aim: Exploring data: View() or print() (R).

1) head:

```
install.packages(c("readr", "psych"))
library(readr)
library(psych)
Caffeine_mg <- read.csv("Daily.Coffee.Intake.vs.Sleep.Duration.csv")
head(Daily.Coffee.Intake.vs.Sleep.Duration)
```

ID	Age	Gender	Country	Coffee_Intake	Caffeine_mg	Sleep_Hours	Sleep_Quality	BMI	Heart_Rate
1	40	Male	Germany	3.5	328.1	7.5	Good	24.9	78
2	33	Male	Germany	1.0	94.1	6.2	Good	20.0	67
3	42	Male	Brazil	5.3	503.7	5.9	Fair	22.7	59
4	53	Male	Germany	2.6	249.2	7.3	Good	24.7	71
5	32	Female	Spain	3.1	298.0	5.3	Fair	24.1	76
6	32	Male	Mexico	3.4	326.4	6.4	Good	27.0	82

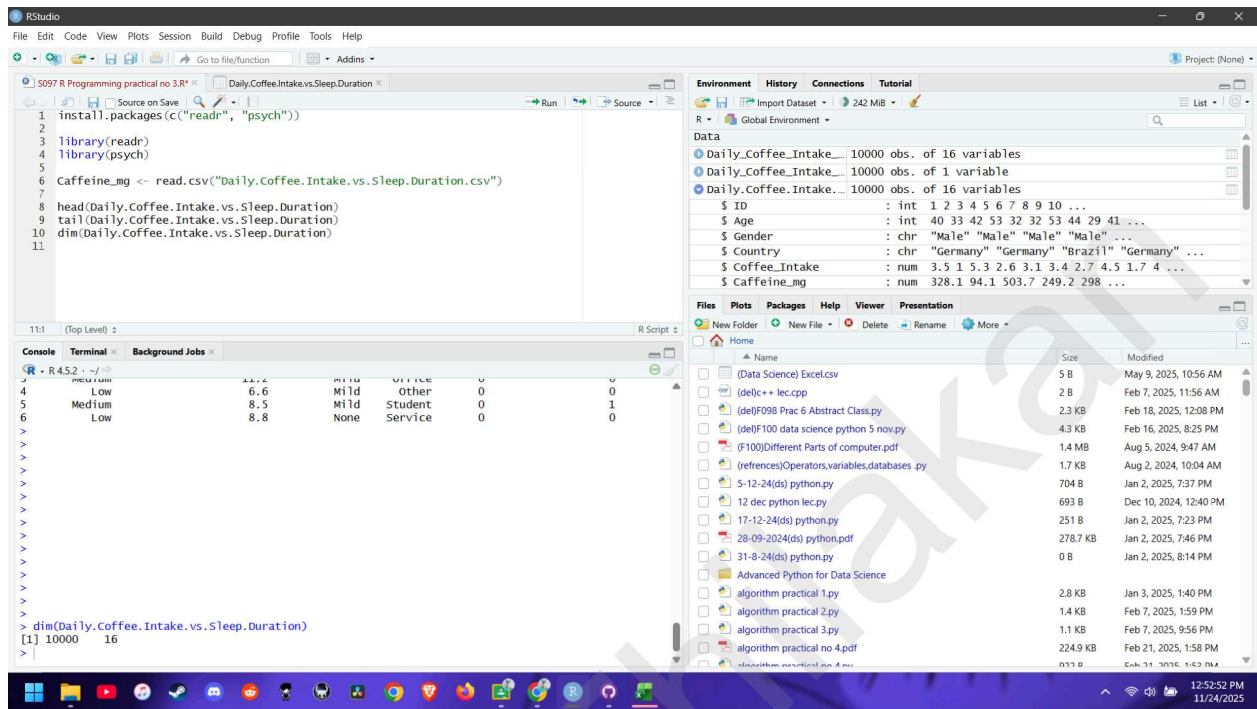
2) tail:

```
tail(Daily.Coffee.Intake.vs.Sleep.Duration)
```

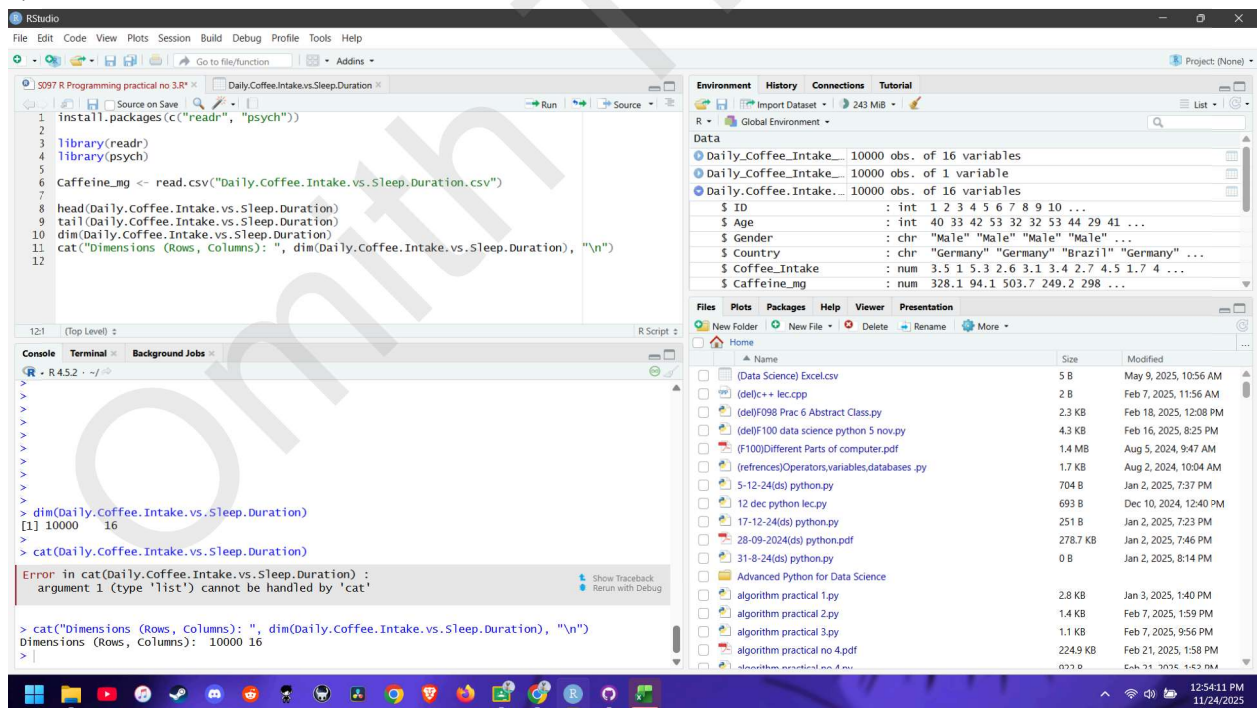
Stress_Level	Physical_Activity_Hours	Health_Issues	Occupation	Smoking	Alcohol_Consumption
1	Low	14.5	None	Other	0
2	Low	11.0	None	Service	0
3	Medium	11.2	Mild	Office	0
4	Low	6.6	Mild	Other	0
5	Medium	8.5	Mild	Student	1
6	Low	8.8	None	Service	0

SHETH LUJ AND SIR MV COLLEGE
Subject: Data Analysis with SAS / SPSS /R

3) dim:



4) cat:



SHETH LUJ AND SIR MV COLLEGE
Subject: Data Analysis with SAS / SPSS / R

5) str:

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

```
1 install.packages(c("readr", "psych"))
2
3 library(readr)
4 library(psych)
5
6 Caffeine_mg <- read.csv("Daily.Coffee.Intake.vs.Sleep.Duration.csv")
7
8 head(Daily.Coffee.Intake.vs.Sleep.Duration)
9 tail(Daily.Coffee.Intake.vs.Sleep.Duration)
10 dim(Daily.Coffee.Intake.vs.Sleep.Duration)
11 cat("Dimensions (Rows, Columns): ", dim(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
12 str(Daily.Coffee.Intake.vs.Sleep.Duration)
```

The console output shows the dimensions and structure of the data frame:

```
> cat("Dimensions (Rows, Columns): ", dim(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
Dimensions (Rows, Columns): 10000 16
> str(Daily.Coffee.Intake.vs.Sleep.Duration)
'data.frame': 10000 obs. of 16 variables:
 $ ID      : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Age     : int  40 33 42 53 32 32 53 44 29 41 ...
 $ Gender  : chr  "Male" "Male" "Male" "Male" ...
 $ Country : chr  "Germany" "Germany" "Brazil" "Germany" ...
 $ Coffee_Intake : num  3.5 1.5 3.2 6.3 1.3 4.2 7.4 5.1 7.4 ...
 $ Caffeine_mg : num  328.1 94.1 503.7 249.2 298 ...
 $ Sleep_Hours : num  7.5 6.2 5.9 7.3 5.3 6.4 7.8 5.5 7.1 6.4 ...
 $ Sleep_Quality : chr  "Good" "Good" "Fair" "Good" ...
 $ BMI     : num  24.9 20 22.7 24.7 24.1 27 24.3 15.8 21.7 30.4 ...
 $ Heart_Rate : int  78 67 59 71 76 82 58 62 60 69 ...
 $ Stress_Level : chr  "Low" "Low" "Medium" "Low" ...
 $ Physical_Activity_Hours : num  14.5 11 11.2 6.6 8.5 8.8 1.0 7.2 2.2 11.9 ...
 $ Health_Issues : chr  "None" "None" "Mild" "Mild" ...
 $ Occupation  : chr  "Other" "Service" "Office" "Other" ...
 $ Smoking     : int  0 0 0 0 0 1 1 1 0 ...
 $ Alcohol_Consumption : int  0 0 0 1 0 0 1 0 ...
```

The Environment pane on the right shows the loaded data frame: `Daily.Coffee.Intake...` with 10000 observations and 16 variables.

6) summary:

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

```
13 dim(Daily.Coffee.Intake.vs.Sleep.Duration)
14
15 cat("Dimensions (Rows, Columns): ", dim(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
16
17 str(Daily.Coffee.Intake.vs.Sleep.Duration)
18
19 summary(Daily.Coffee.Intake.vs.Sleep.Duration)
20
```

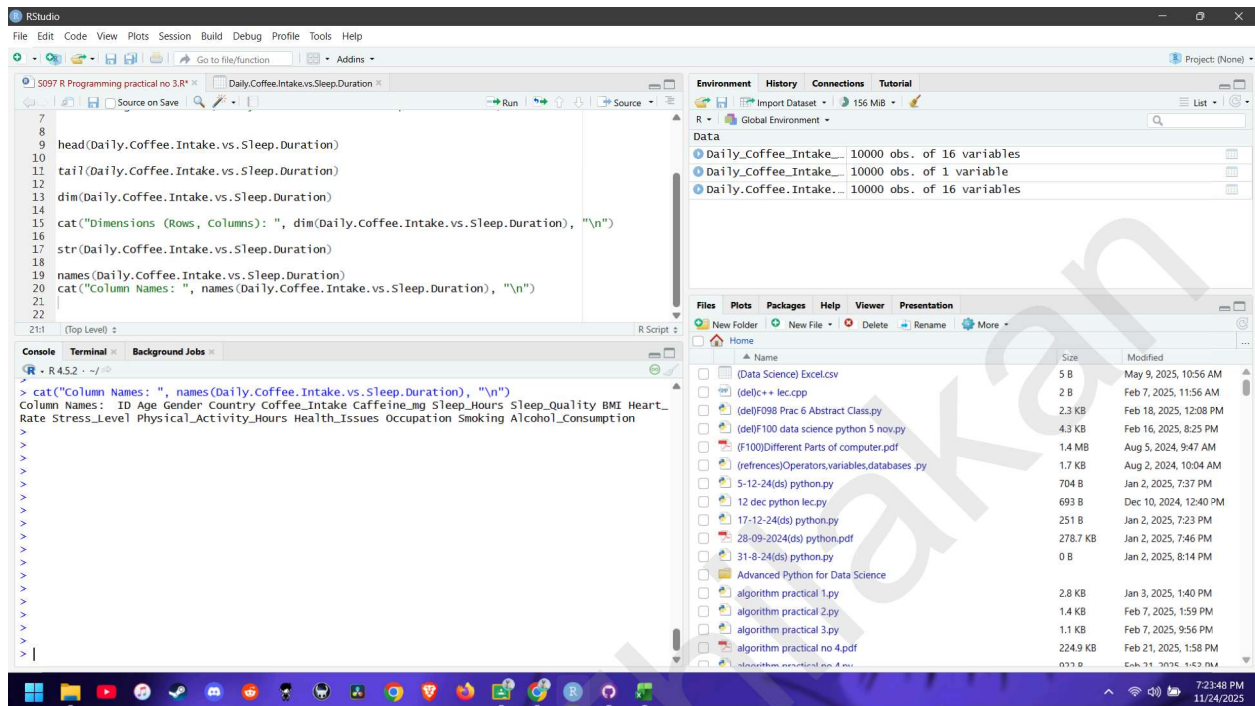
The console output shows the summary statistics for the data frame:

```
> summary(Daily.Coffee.Intake.vs.Sleep.Duration)
      ID      Age      Gender      Country      Coffee_Intake
Min.   : 1      Min.   :18.00      Length:10000      Length:10000      Min.   :0.000
1st Qu.:2501    1st Qu.:26.00      Class :character      Class :character      1st Qu.:1.500
Median :5000    Median :34.00      Mode  :character      Mode  :character      Median :2.500
Mean   :5000    Mean   :34.95
3rd Qu.:7500    3rd Qu.:43.00
Max.   :10000   Max.   :80.00
Caffeine_mg Sleep_Hours Sleep_Quality BMI      Heart_Rate
Min.   : 0.0      Min.   : 3.000      Length:10000      Min.   :15.00      Min.   : 50.00
1st Qu.:138.8    1st Qu.: 5.800      Class :character      1st Qu.:21.30      1st Qu.: 64.00
Median :235.4    Median : 6.600      Mode  :character      Median :24.00      Median : 71.00
Mean   :238.4    Mean   : 6.636      Mean   :23.99      Mean   : 70.62
3rd Qu.:332.0    3rd Qu.: 7.500      3rd Qu.:26.60      3rd Qu.: 77.00
Max.   :780.3    Max.   :10.000      Max.   :38.20      Max.   :109.00
Stress_Level Physical_Activity_Hours Health_Issues Occupation
Length:10000      Min.   : 0.000      Length:10000      Length:10000
Class :character      1st Qu.: 3.700      Class :character      Class :character
Mode  :character      Median : 7.500      Mode  :character      Mode  :character
Mean   : 7.487
3rd Qu.:11.200
Max.   :15.000
Smoking      Alcohol_Consumption
Min.   :0.0000      Min.   :0.0000
1st Qu.:0.0000      1st Qu.:0.0000
Median :0.0000      Median :0.0000
Mean   :0.2004      Mean   :0.3007
3rd Qu.:0.0000      3rd Qu.:1.0000
Max.   :1.0000      Max.   :1.0000
```

The Environment pane on the right shows the loaded data frame: `Daily.Coffee.Intake...` with 10000 observations and 16 variables.

SHETH LUJ AND SIR MV COLLEGE
Subject: Data Analysis with SAS / SPSS / R

7) column names:



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

```
7 head(Daily.Coffee.Intake.vs.Sleep.Duration)
8
9
10 tail(Daily.Coffee.Intake.vs.Sleep.Duration)
11
12
13 dim(Daily.Coffee.Intake.vs.Sleep.Duration)
14
15 cat("Dimensions (Rows, Columns): ", dim(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
16
17 str(Daily.Coffee.Intake.vs.Sleep.Duration)
18
19 names(Daily.Coffee.Intake.vs.Sleep.Duration)
20 cat("Column Names: ", names(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
21
22
```

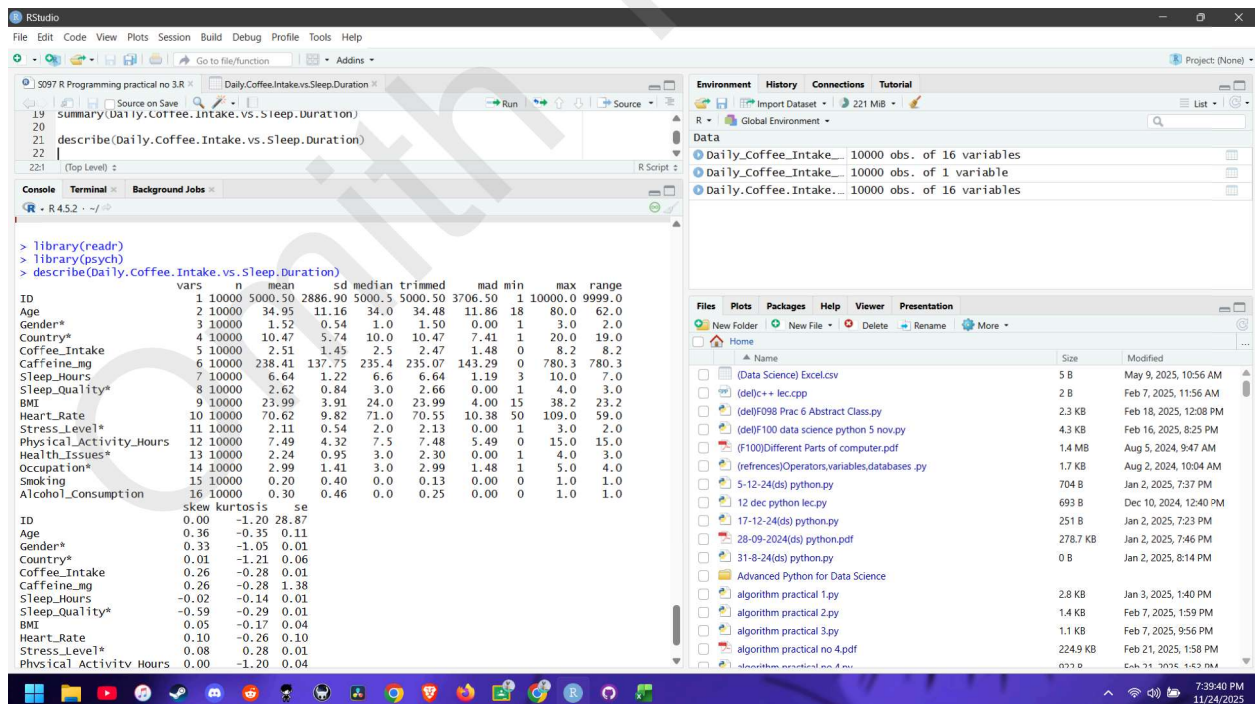
The console output shows the column names:

```
> cat("Column Names: ", names(Daily.Coffee.Intake.vs.Sleep.Duration), "\n")
Column Names: ID Age Gender Country Coffee_Intake Caffeine_mg Sleep_Hours Sleep_Quality BMI Heart_Rate
Stress_Level Physical_Activity_Hours Health_Issues Occupation Smoking Alcohol_Consumption
>
```

The Environment pane on the right shows the loaded data:

Object	Class	Attributes
Daily_Coffee_Intake_...	data.frame	10000 obs. of 16 variables
Daily_Coffee_Intake_...	data.frame	10000 obs. of 1 variable
Daily_Coffee_Intake_...	data.frame	10000 obs. of 16 variables

8) describe:



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

```
19 summary(Daily.Coffee.Intake.vs.Sleep.Duration)
20
21 describe(Daily.Coffee.Intake.vs.Sleep.Duration)
22
```

The console output shows the summary and description of the data:

```
> library(readr)
> library(psych)
> describe(Daily.Coffee.Intake.vs.Sleep.Duration)

vars      n  mean  sd median trimmed  mad min  max range
ID      1 10000 5000.0 2886.90 5000.5 5000.50 3706.50 1 10000.0 9999.0
Age      2 10000  34.95  11.16  34.0  34.48  11.86  18  80.0  62.0
Gender*   3 10000  1.52   0.54   1.0   1.50   0.00   1   3.0   2.0
Country*  4 10000  10.47   5.74  10.0  10.47  7.41  1  20.0  19.0
Coffee_Intake 5 10000  2.51   1.45   2.5   2.47  1.48  0   8.2   8.2
Caffeine_mg 6 10000 238.41 137.75 235.4 235.07 143.29 0 780.3 780.3
Sleep_Hours 7 10000  6.64   1.22   6.6   6.64  1.19  3  10.0   7.0
Sleep_Quality* 8 10000  2.62   0.84   3.0   2.66  0.00  1   4.0   3.0
BMI      9 10000 23.99   3.91  24.0 23.99  4.00 15 38.2 23.2
Heart_Rate 10 10000 70.62   9.82  71.0 70.55 10.38 50 109.0 59.0
Stress_Level* 11 10000  2.11   0.54   2.0   2.13  0.00  1   3.0   2.0
Physical_Activity_Hours 12 10000  7.49   4.32   7.5   7.48  5.49  0  15.0  15.0
Health_Issues* 13 10000  2.24   0.95   3.0   2.30  0.00  1   5.0   3.0
Occupation* 14 10000  2.99   1.41   3.0   2.99  1.48  1   5.0   4.0
Smoking   15 10000  0.20   0.40   0.0   0.13  0.00  0   1.0   1.0
Alcohol_Consumption 16 10000  0.30   0.46   0.0   0.25  0.00  0   1.0   1.0

skew kurtosis se
ID      0.00  -1.20 28.87
Age      0.36  -0.35 0.11
Gender*  0.33  -1.05 0.01
Country* 0.01  -1.21 0.06
Coffee_Intake 0.26  -0.28 0.01
Caffeine_mg 0.26  -0.28 1.38
Sleep_Hours -0.02  -0.14 0.01
Sleep_Quality* -0.59  -0.29 0.01
BMI      0.05  -0.17 0.04
Heart_Rate 0.10  -0.26 0.10
Stress_Level* 0.08  0.28 0.01
Physical_Activity_Hours 0.00  -1.20 0.04
```

The Environment pane on the right shows the loaded data:

Object	Class	Attributes
Daily_Coffee_Intake_...	data.frame	10000 obs. of 16 variables
Daily_Coffee_Intake_...	data.frame	10000 obs. of 1 variable
Daily_Coffee_Intake_...	data.frame	10000 obs. of 16 variables