

ISE CERTIFICATION COURSE DETAILS

NAME:	ANUSHA	USN:	4AL16IS006
SEMESTER:	8	MENTOR:	Mr. SUDHARSHAN K
COURSE NAME:	Best datascience and data analytics course	DATE:	13/06/2020

SCREENSHOT:

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

```

13. Subset Command
81 which(letters=="s")
82
83 #using data frame WOMEN
84 women
85 which(women$height==66)
86 which(women$height>=70)
87
88 #which for subsetting vector and using conditions
89 #Example1
90 j = c(2,3,4,5,6,7,8,9,23,24,22,1,10)
91 j>6
92 which(j>6)
93 j[which(j > 6)]
94
95 LETTERS
96 which(LETTERS == "K") #11th alphabet
97
98 #Example2
99 marks = c(10, 30, 40, 60)

```

The console shows the output of the code:

```

9 66 139
10 67 142
11 68 146
12 69 150
13 70 154
14 71 159
15 72 164
> which(women$height>=70)
[1] 13 14 15
> #which for subsetting vector and using conditions
#Example1
> j[j>6]
[1] FALSE FALSE FALSE FALSE FALSE TRUE TRUE TRUE TRUE TRUE TRUE FALSE TRUE

```

The Environment pane on the right shows the objects in the global environment:

- a: int [1:3, 1:3] 1 2 3
- df: 3 obs. of 3 variables
- x: num [1:6, 1:6] -0.21
- x.df: 6 obs. of 6 variables
- y: num [1:13] 2 3 4 5 6
- y: num [1:6] 1 1 2 3 4 5

BRIEF REPORT: (POINT-WISE)

- 1) Subsetting data in R – subsetting is used to select or exclude variable and observation from our data. It makes our data more precise and as per need for analysis.
- 2) Subset commands - It returns the vectors, matrix or dataframes which meet a particular condition.
- 3) Aggregate functions is used to compare a statistic summary of subsetting dataframes.
- 4) Data manipulation in R using dplyr packages – dplyr package is grammar of data manipulation is also known as data wrangling and data cleaning.
- 5) Summarizing the data graphically in R- Scatter box, Box plot, Histogram , bar graph, line graph , pie chart.
- 6) Simple linear regression in R used to find relationship between dependent and independent variable.
- 7) Cluster analysis in R – It is a unsupervised machine learning technique of data segmentation that partitions the data into several groups .
- 8) Kmeans model – hierarchical clustering.

