**PRACTICAL 4**

**Dataset:** Basketball data from March Madness

**Download Link:**<http://becomingvisual.com/rfundamentals/march_madness.csv>

**Data Dictionary:**

| **Variable** | **Description** |
| --- | --- |
| Rank | Team Ranking |
| Previous | Previous Team Ranking |
| School | Name of the College or University |
| Conference | NCAA Conference (30 +) |
| Record | Overall Record |
| Neutral | Record with games in a neutral location |
| Home | Record with games at home |
| Non Div I | Record with non-divison 1 games |

**Write a R script to do the following:**

1. Set working directory – Hint: setwd()

| Solution:  setwd("/home/itmbu/Desktop/NEERAJ/RPD/datasheets")  > getwd()  [1] "/home/itmbu/Desktop/NEERAJ/RPD/datasheets" |
| --- |

1. Import the csv file

| Solution:  > march\_madness. <- read.csv("~/Desktop/NEERAJ/RPD/datasheets/march\_madness .csv") |
| --- |

1. View the file

| Solution:  > View(march\_madness\_)  . |
| --- |

1. Print number of rows and columns – Hint: dim()

| Solution:  > dim(march\_madness\_)  [1] 349 9  > nrow(march\_madness.)  [1] 349  > ncol(march\_madness.) |
| --- |

1. Print columns names

| Solution:  > names(march\_madness.)  [1] "RANK" "PREVIOUS" "SCHOOL" "CONFERENCE"  [5] "RECORD" "ROAD" "NEUTRAL" "HOME"  [9] "NON.DI" |
| --- |

1. Change column names to lower case so it is easier to use Hint: names(df\_name) <- tolower(names(df\_name))

| Solution:  > names(march\_madness.)=tolower(names(march\_madness.))  > names(march\_madness.)  [1] "rank" "previous" "school" "conference" "record" "road" "neutral"  [8] "home" "non.di" |
| --- |

1. Explore the variable types. – Hint: str()

| Solution:  > str(march\_madness.)  'data.frame': 349 obs. of 9 variables:  $ rank : int 1 2 3 4 5 6 7 8 9 10 ...  $ previous : int 2 1 3 4 5 6 8 11 7 10 ...  $ school : Factor w/ 349 levels "A&M-Corpus Christi",..: 87 11 131 340 326 343 124 327 72 61 ...  $ conference: Factor w/ 33 levels "AAC","America East",..: 26 24 6 20 7 10 6 4 4 7 ...  $ record : Factor w/ 123 levels "13-16","13-17",..: 80 78 60 81 73 67 67 74 66 67 ...  $ road : Factor w/ 102 levels "0-10","0-11",..: 75 63 37 84 75 63 28 49 29 49 ...  $ neutral : Factor w/ 34 levels "0- 0","0- 1",..: 32 20 19 32 17 20 34 33 24 18 ...  $ home : Factor w/ 111 levels "0-12","01-Nov",..: 107 110 101 103 101 97 101 100 107 103 ...  $ non.di : Factor w/ 10 levels "0-0","0-1","01-Jan",..: 1 1 1 7 1 1 1 1 1 1 ... |
| --- |

1. How many different conferences are there?

| Solution:  > march\_madness.$conference<-as.factor(march\_madness.$conference)  > conference<-as.factor(march\_madness.)  > conference<-as.factor(march\_madness.$conference)  > conference |
| --- |

1. Let’s look at the difference in values of first two columns:
2. Compute a new vector called “diff” and calculate the difference in rank and previous
3. Print count and list of schools that changed 3 or more places Hint: create subset that satisfies criteria

| Solution: |
| --- |

1. Import the GDP dataset and compute the difference in GDP between 2007 and 2017 for each country.

| Solution: |
| --- |

\*\*\*\*\*\*\*\*