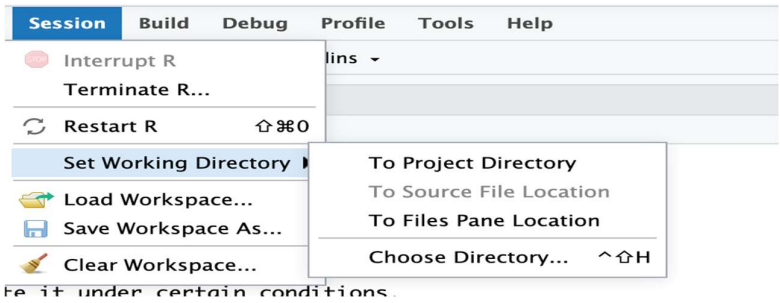
Download dataset for the examination and set the working directory.

1. Download 1st-Internal dataset from website (https://sitmbadept.github.io/BDTM/)
2. Create new folder with your name/USN on Desktop
3. Copy Paste Downloaded file in the folder created in Step#2
4. Unzip downloaded file(Right click > Extract All) or using Winzip software
5. Set Unzip folder as a project working directory in R Studio
   (Click on Session Menu → Set Working Directory → Choose Directory)



Q1. Write R Code for creating numeric vector and display all vector elements [5]

101, 102, 103, 104, 105, 106, 107, 108, 109, 110

Ans.

```
# Vector creation
x <- c(101, 102, 103, 104, 105, 106, 107, 108, 109, 110)

# Display vector element
print(x)
```

Q2. Write R Code for following statement. [5]

1. Read SPSS exam.sav file in R
2. Display the structure and dimension of data

Note: Download & load external packages to read SPSS file

```
install.packages("haven", dep=TRUE)
library(haven)
```

Ans.

```
install.packages("haven", dep=TRUE)
library(haven)

data = read_sav("exam.sav")

str(data)
dim(data)
```

Q3. Read Marketing Spend (marketing-spend.csv) data, Perform simple linear regression and display summary of regression model in R.

Dataset contains 12 months of Monthly Spends & Sales details, whereas Spends & Month are independent variables and Sales is dependent variable.

[10]

```
# Import data
data = read.csv("marketing-spend.csv")

# Prepare Formula & Regression
model_1 = lm(Sales~Spend, data)

# Regression Summary
summary(model_1)
```

Q4. Create SQL Statement for below tables [15]

emp_id	**emp_name**	**emp_city**	**Emp_education**
E1	John	USA	B.Tech
E2	Dom	UK	BSc
E3	Nick	USA	BCom
E4	Sager	CHINA	BCA
E5	Paul	UK	M.tech

a) Create above table in SQL using CREATE TABLE
b) Insert table records in SQL using INSERT INTO
c) Display all the records

Ans.

[a]

```
CREATE TABLE employee(
        emp_id text,
        emp_name text,
        emp_city text,
        Emp_education text
);
```

[b]

```
INSERT INTO employee VALUES ('E1', 'John', 'USA', 'B.Tech');
INSERT INTO employee VALUES ('E2', 'Dom', 'UK', 'BSc');
INSERT INTO employee VALUES ('E3', 'Nick', 'USA', 'BCom');
INSERT INTO employee VALUES ('E4', 'Sager', 'CHINA', 'BCA');
INSERT INTO employee VALUES ('E5', 'Paul', 'UK', 'M.tech');
```

[c]

```
SELECT * FROM employee;
```

Q5. Read Credit Card churn dataset and write R code for following statements. [15]

About Dataset:

- **CLIENTNUM**: Client number. Unique identifier for the customer holding the account
- **Attrition_Flag**: Internal event/Churn (customer activity) variable - if the account is closed then 1 else 0
- **Customer_Age**: Demographic variable - Customer's Age in Years
- **Gender**: Demographic variable - M=Male, F=Female
- **Dependent_count**: Demographic variable - Number of dependents
- **Education_Level**: Educational Qualification of the account holder (example: high school, college graduate, etc.)
- **Marital_Status**: Demographic variable - Married, Single, Divorced, Unknown
- **Income_Category**: Demographic variable - Annual Income Category of the account holder
- **Card_Category**: Product Variable - Type of Card (Blue, Silver, Gold, Platinum)
- **Months_on_book**: Period of relationship with bank
- **Total_Relationship_Count**: Total no. of products held by the customer
- **Months_Inactive_12_mon**: No. of months inactive in the last 12 months
- **Contacts_Count_12_mon** No. of Contacts in the last 12 months
- **Credit_Limit**: Credit Limit on the Credit Card
- **Total_Revolving_Bal**: Total Revolving Balance on the Credit Card
- **Avg_Open_To_Buy**: Open to Buy Credit Line (Average of last 12 months)
- **Total_Amt_Chng_Q4_Q1**: Change in Transaction Amount (Q4 over Q1)
- **Total_Trans_Amt**: Total Transaction Amount (Last 12 months)
- **Total_Trans_Ct**: Total Transaction Count (Last 12 months)
- **Total_Ct_Chng_Q4_Q1**: Change in Transaction Count (Q4 over Q1)
- **Avg_Utilization_Ratio** Average Card Utilization Ratio

Write R Code for following statements :

A. Read (BankChurners.csv) CSV file
B. Display the structure and dimension of data
C. Display top 6 records from the Dataset
D. Display bottom 6 records from the Dataset
E. Display quick summary of all the columns
F. Display frequency of Attrition_Flag using table function
G. Display frequency of Marital_Status & Attrition_Flag using table function
H. Display Unique values of Card_Category column in data.

```
#A.Read (BankChurners.csv) CSV file
data = read.csv("BankChurners.csv")

#B.Display the structure  and dimension of data
str(data)
dim(data)

#C.Display top 6 records from the Dataset
head(data,6)

#D.Display bottom 6 records from the Dataset
```

```
tail(data,6)

#E.Display quick summary of all the columns
summary(data)

#F.Display frequency of Attrition_Flag using table function
table(data$Attrition_Flag)

#G.Display frequency of Marital_Status & Attrition_Flag using table function
table(data$Marital_Status, data$Attrition_Flag)

#H.Display Unique values of Card_Category column in data.
unique(data$Card_Category)
```