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

Practical No: 7

Aim: Selecting and dropping variables using select() in R. import dataset.

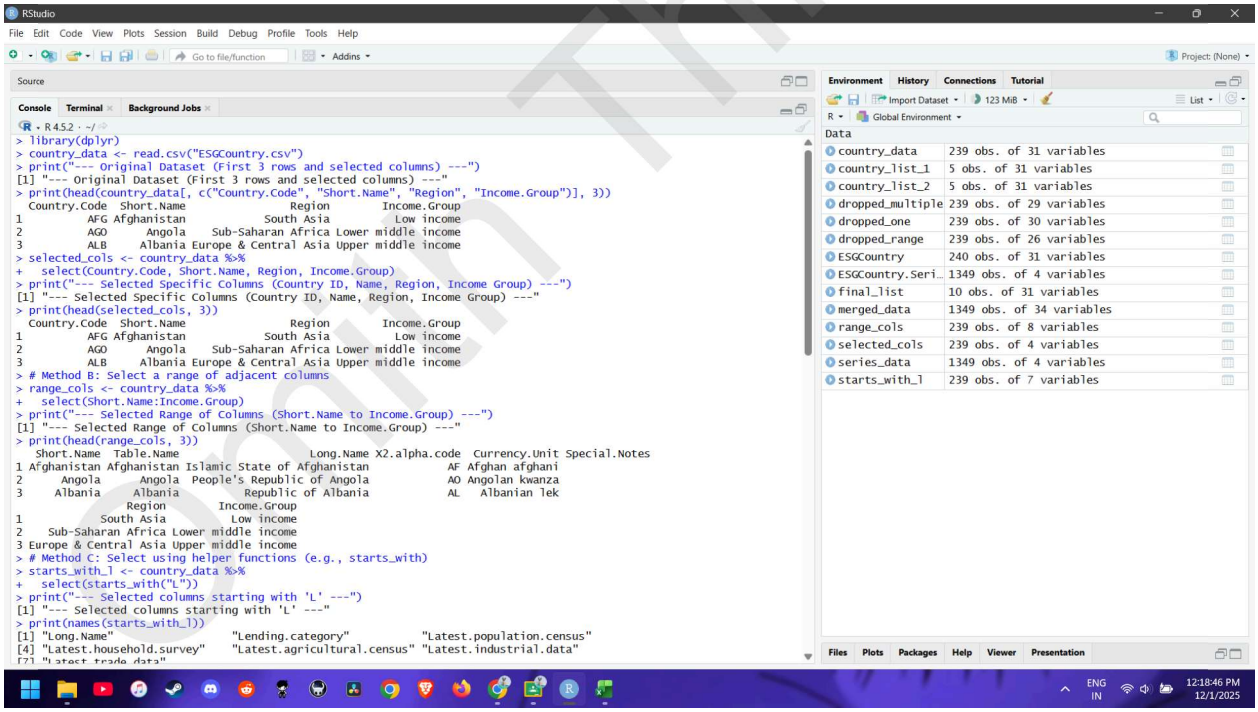
Code:

```
# 7. Selecting and dropping variables using select() in R.

library(dplyr)

# 1. IMPORT DATASET

country_data <- read.csv("ESGCountry.csv")

print("--- Original Dataset (First 3 rows and selected columns) ---")
print(head(country_data[, c("Country.Code", "Short.Name", "Region", "Income.Group")], 3))

# 2. SELECTING VARIABLES

selected_cols <- country_data %>%
  select(Country.Code, Short.Name, Region, Income.Group)

print("--- Selected Specific Columns (Country ID, Name, Region, Income Group) ---")
print(head(selected_cols, 3))

# Method B: Select a range of adjacent columns
range_cols <- country_data %>%
  select(Short.Name:Income.Group)

print("--- Selected Range of Columns (Short.Name to Income.Group) ---")
print(head(range_cols, 3))

# Method C: Select using helper functions (e.g., starts_with)
starts_with_l <- country_data %>%
  select(starts_with("L"))

print("--- Selected columns starting with 'L' ---")
print(names(starts_with_l))
print(head(starts_with_l, 3))

# 3. DROPPING VARIABLES (Removing Columns)

# FIX APPLIED: The blank column header is the 31st column, so it is named X31.
dropped_one <- country_data %>%
  select(-31)
```

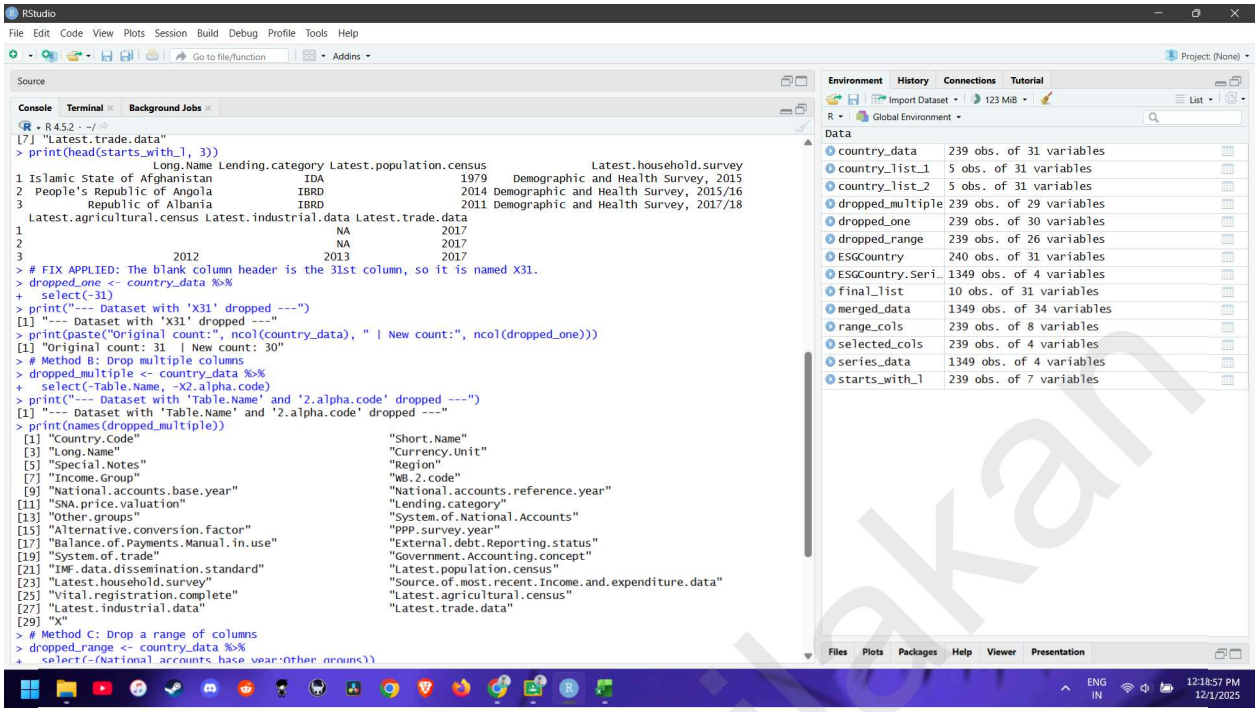
Omith Thilakan
S097

```
print("--- Dataset with 'X31' dropped ---")

print(paste("Original count:", ncol(country_data), " | New count:", ncol(dropped_one)))

# Method B: Drop multiple columns
dropped_multiple <- country_data %>%
  select(-Table.Name, -X2.alpha.code)

print("--- Dataset with 'Table.Name' and '2.alpha.code' dropped ---")
print(names(dropped_multiple))

# Method C: Drop a range of columns
dropped_range <- country_data %>%
  select(-(National.accounts.base.year:Other.groups))

print("--- Dataset with range 'National.accounts.base.year' to 'Other.groups' dropped ---")
print(names(dropped_range))
```

Output:

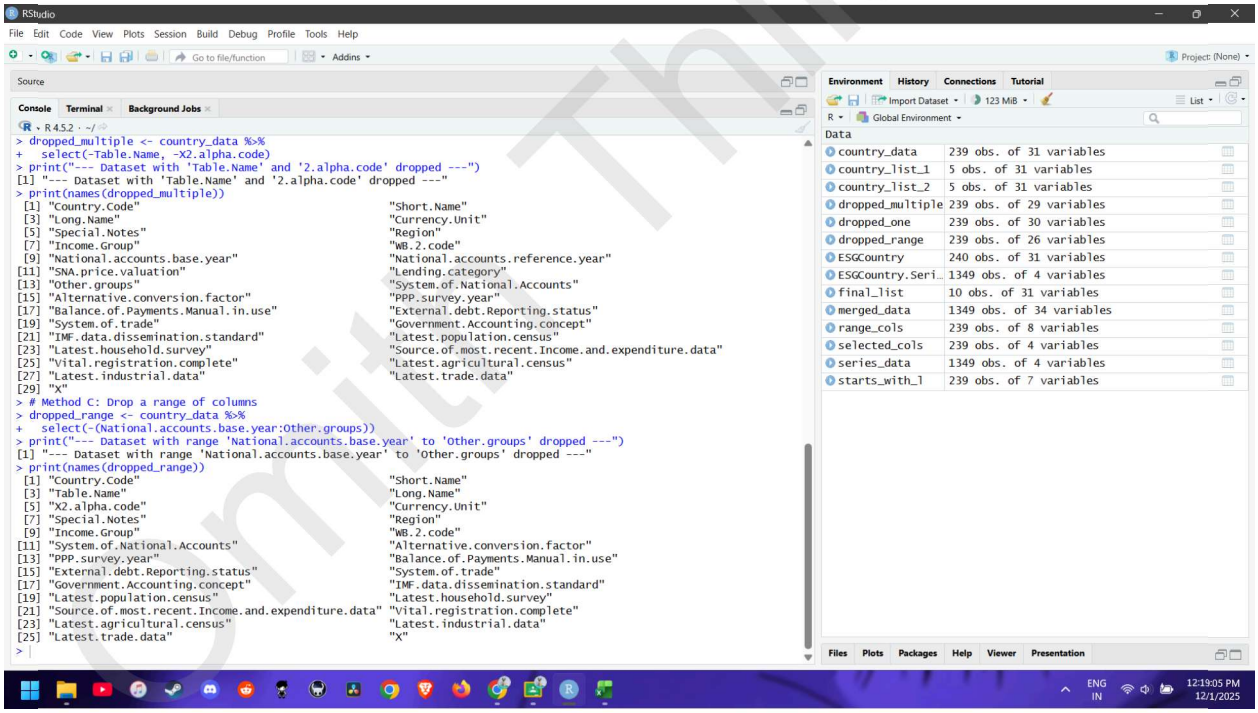


Omith Thilakan
S097



```
[7] "Latest.trade.data"
> print(head(starts_with_l, 3))
                  Long.Name Lending.category Latest.population.census                Latest.household.survey
1 Islamic State of Afghanistan              IDA                    1979    Demographic and Health Survey, 2015
2  People's Republic of Angola            IBRD                    2014 Demographic and Health Survey, 2015/16
3         Republic of Albania            IBRD                    2011 Demographic and Health Survey, 2017/18
  Latest.agricultural.census Latest.industrial.data Latest.trade.data
1                                                NA              2017
2                                                NA              2017
3                       2012                   2013              2017
> # FIX APPLIED: The blank column header is the 31st column, so it is named X31.
> dropped_one <- country_data %>%
+   select(-31)
> print("--- Dataset with 'X31' dropped ---")
[1] "--- Dataset with 'X31' dropped ---"
> print(paste("Original count:", ncol(country_data), " | New count:", ncol(dropped_one)))
[1] "Original count: 31  | New count: 30"
> # Method B: Drop multiple columns
> dropped_multiple <- country_data %>%
+   select(-Table.Name, -X2.alpha.code)
> print("--- Dataset with 'Table.Name' and '2.alpha.code' dropped ---")
[1] "--- Dataset with 'Table.Name' and '2.alpha.code' dropped ---"
> print(names(dropped_multiple))
 [1] "Country.Code"                                "Short.Name"
 [3] "Long.Name"                                   "Currency.Unit"
 [5] "Special.Notes"                               "Region"
 [7] "Income.Group"                                "WB.2.code"
 [9] "National.accounts.base.year"                 "National.accounts.reference.year"
[11] "SNA.price.valuation"                         "Lending.category"
[13] "Other.groups"                                "System.of.National.Accounts"
[15] "Alternative.conversion.factor"               "PPP.survey.year"
[17] "Balance.of.Payments.Manual.in.use"           "External.debt.Reporting.status"
[19] "System.of.trade"                             "Government.Accounting.concept"
[21] "IMF.data.dissemination.standard"             "Latest.population.census"
[23] "Latest.household.survey"                     "Source.of.most.recent.Income.and.expenditure.data"
[25] "Vital.registration.complete"                 "Latest.agricultural.census"
[27] "Latest.industrial.data"                      "Latest.trade.data"
[29] "X"
> # Method C: Drop a range of columns
> dropped_range <- country_data %>%
+   select(-(National.accounts.base.year:Other.groups))
```

Data	
country_data	239 obs. of 31 variables
country_list_1	5 obs. of 31 variables
country_list_2	5 obs. of 31 variables
dropped_multiple	239 obs. of 29 variables
dropped_one	239 obs. of 30 variables
dropped_range	239 obs. of 26 variables
ESGCountry	240 obs. of 31 variables
ESGCountry.Seri...	1349 obs. of 4 variables
final_list	10 obs. of 31 variables
merged_data	1349 obs. of 34 variables
range_cols	239 obs. of 8 variables
selected_cols	239 obs. of 4 variables
series_data	1349 obs. of 4 variables
starts_with_l	239 obs. of 7 variables



```
> dropped_multiple <- country_data %>%
+   select(-Table.Name, -X2.alpha.code)
> print("--- Dataset with 'Table.Name' and '2.alpha.code' dropped ---")
[1] "--- Dataset with 'Table.Name' and '2.alpha.code' dropped ---"
> print(names(dropped_multiple))
 [1] "Country.Code"                                "Short.Name"
 [3] "Long.Name"                                   "Currency.Unit"
 [5] "Special.Notes"                               "Region"
 [7] "Income.Group"                                "WB.2.code"
 [9] "National.accounts.base.year"                 "National.accounts.reference.year"
[11] "SNA.price.valuation"                         "Lending.category"
[13] "Other.groups"                                "System.of.National.Accounts"
[15] "Alternative.conversion.factor"               "PPP.survey.year"
[17] "Balance.of.Payments.Manual.in.use"           "External.debt.Reporting.status"
[19] "System.of.trade"                             "Government.Accounting.concept"
[21] "IMF.data.dissemination.standard"             "Latest.population.census"
[23] "Latest.household.survey"                     "Source.of.most.recent.Income.and.expenditure.data"
[25] "Vital.registration.complete"                 "Latest.agricultural.census"
[27] "Latest.industrial.data"                      "Latest.trade.data"
[29] "X"
> # Method C: Drop a range of columns
> dropped_range <- country_data %>%
+   select(-(National.accounts.base.year:Other.groups))
> print("--- Dataset with range 'National.accounts.base.year' to 'Other.groups' dropped ---")
[1] "--- Dataset with range 'National.accounts.base.year' to 'Other.groups' dropped ---"
> print(names(dropped_range))
 [1] "Country.Code"                                "Short.Name"
 [3] "Table.Name"                                  "Long.Name"
 [5] "X2.alpha.code"                               "Currency.Unit"
 [7] "Special.Notes"                               "Region"
 [9] "Income.Group"                                "WB.2.code"
[11] "System.of.National.Accounts"                 "Alternative.conversion.factor"
[13] "PPP.survey.year"                             "Balance.of.Payments.Manual.in.use"
[15] "External.debt.Reporting.status"              "System.of.trade"
[17] "Government.Accounting.concept"               "IMF.data.dissemination.standard"
[19] "Latest.population.census"                    "Latest.household.survey"
[21] "Source.of.most.recent.Income.and.expenditure.data" "Vital.registration.complete"
[23] "Latest.agricultural.census"                  "Latest.industrial.data"
[25] "Latest.trade.data"                           "X"
>
```

Data	
country_data	239 obs. of 31 variables
country_list_1	5 obs. of 31 variables
country_list_2	5 obs. of 31 variables
dropped_multiple	239 obs. of 29 variables
dropped_one	239 obs. of 30 variables
dropped_range	239 obs. of 26 variables
ESGCountry	240 obs. of 31 variables
ESGCountry.Seri...	1349 obs. of 4 variables
final_list	10 obs. of 31 variables
merged_data	1349 obs. of 34 variables
range_cols	239 obs. of 8 variables
selected_cols	239 obs. of 4 variables
series_data	1349 obs. of 4 variables
starts_with_l	239 obs. of 7 variables

Omith Thilakan
S097