

# Section - A SUMMATIVE ASSESSMENT

Thrisha Rajkumar

2024-11-26

## SECTION A - SUMMATIVE ASSESSMENT

General instruction: In this part of your assessment, you will perform a data wrangling task using R programming. Note that clarity is highly important. Be careful to make sure you clearly explain each step in your answer. You should also include comments within your code when necessary. In addition, make the structure of your answer clear through the use of headings. You should also make sure your code is clean by making careful use of Tidyverse methods in R.

### Data Wrangling Task & Tidyverse Methods in R

(Q1). First download the files entailed “debt\_data.csv”, “country\_data.csv” and “indicator\_data.csv” which are available within the Assessment section within Blackboard. Download the csv files - “debt\_data.csv”, “country\_data.csv” and “indicator\_data.csv” and Loaded it to the folder - “Section - A Summative Assessment” The file “debt\_data.csv” contains debt data for different countries under different indicators, from 1960 to 2023. The indicators are represented by indicator codes (for example, NY.GNP.MKTP.CD). The file “indicator\_data.csv” contains a list of the indicator names as well as their associated indicator codes. The file “country\_data.csv” contains information about the country code, income levels, and regions for each country.

-> First, Loading the file “debt\_data.csv” into an R data frame called “debt\_df”, load the file “country\_data.csv” into an R data frame called “country\_df”, and load the file “indicator\_data.csv” into a data frame called “indicator\_df” using library readr.

*#First required package to be used to read the csv files and then load it into a R data frame.*

```
library(readr)
```

*#Loading the file "debt\_data.csv" into a R data frame called "debt\_df"*

```
debt_df <- read_csv("debt_data.csv")
```

```
## Rows: 13824 Columns: 63
```

```
## -- Column specification -----
```

```
## Delimiter: ","
```

```
## chr (2): Country.Code, Year
```

```
## dbl (61): NY.GNP.MKTP.CD, FI.RES.TOTL.MO, FI.RES.TOTL.DT.ZS, FI.RES.TOTL.CD,...
```

```
##
```

```
## i Use `spec()` to retrieve the full column specification for this data.
```

```
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
#Loading the file "country_data.csv" into an R data frame called "country_df"
country_df <- read_csv("country_data.csv", show_col_types = FALSE)

#Loading the file "indicator_data.csv" into a data frame called "indicator_data.csv"
indicator_df <- read_csv("indicator_data.csv", show_col_types = FALSE)

#printing or displaying the first few rows of the data frames - debt_df, country_df, and indicator_df
head(debt_df)
```

```
## # A tibble: 6 x 63
##   Country.Code Year      NY.GNP.MKTP.CD FI.RES.TOTL.MO FI.RES.TOTL.DT.ZS
##   <chr>         <chr>          <dbl>          <dbl>          <dbl>
## 1 ABW          year_1960          NA            NA            NA
## 2 ABW          year_1961          NA            NA            NA
## 3 ABW          year_1962          NA            NA            NA
## 4 ABW          year_1963          NA            NA            NA
## 5 ABW          year_1964          NA            NA            NA
## 6 ABW          year_1965          NA            NA            NA
## # i 58 more variables: FI.RES.TOTL.CD <dbl>, DT.TDS.MLAT.PG.ZS <dbl>,
## #   DT.TDS.MLAT.CD <dbl>, DT.TDS.DPPG.XP.ZS <dbl>, DT.TDS.DPPG.GN.ZS <dbl>,
## #   DT.TDS.DPPG.CD <dbl>, DT.TDS.DPPF.XP.ZS <dbl>, DT.TDS.DIMF.CD <dbl>,
## #   DT.TDS.DECT.GN.ZS <dbl>, DT.TDS.DECT.EX.ZS <dbl>, DT.TDS.DECT.CD <dbl>,
## #   DT.ODA.ODAT.PC.ZS <dbl>, DT.ODA.ODAT.GN.ZS <dbl>, DT.ODA.ODAT.CD <dbl>,
## #   DT.NFL.RDBN.CD <dbl>, DT.NFL.RDBC.CD <dbl>, DT.NFL.PRVT.CD <dbl>,
## #   DT.NFL.PROP.CD <dbl>, DT.NFL.PNGC.CD <dbl>, DT.NFL.PNGB.CD <dbl>, ...
```

```
head(country_df)
```

```
## # A tibble: 6 x 5
##   Country.Code Region      IncomeGroup SpecialNotes Country.Name
##   <chr>         <chr>          <chr>      <chr>      <chr>
## 1 ABW          Latin America & Caribbean High income <NA>      Aruba
## 2 AFG          South Asia      Low income "The report~ Afghanistan
## 3 AGO          Sub-Saharan Africa Lower middl~ "The World ~ Angola
## 4 ALB          Europe & Central Asia Upper middl~ <NA>      Albania
## 5 AND          Europe & Central Asia High income <NA>      Andorra
## 6 ARE          Middle East & North Africa High income <NA>      United Arab~
```

```
head(indicator_df)
```

```
## # A tibble: 6 x 2
##   INDICATOR_CODE INDICATOR_NAME
##   <chr>         <chr>
## 1 NY.GNP.MKTP.CD GNI (current US$)
## 2 FI.RES.TOTL.MO Total reserves in months of imports
## 3 FI.RES.TOTL.DT.ZS Total reserves (% of total external debt)
## 4 FI.RES.TOTL.CD Total reserves (includes gold, current US$)
## 5 DT.TDS.MLAT.PG.ZS Multilateral debt service (% of public and publicly guarant~
## 6 DT.TDS.MLAT.CD Multilateral debt service (TDS, current US$)
```

-> Second, use R to check the number of columns and the number of rows that the data frame “debt\_df” has. Display your results.

First loading the necessary package - Tidyverse to find the number of rows and columns in the data using ncol() and nrow().

```
# Loading the tidyverse required package  
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v purrr      1.0.2  
## v forcats    1.0.0      v stringr    1.5.1  
## v ggplot2    3.5.1      v tibble     3.2.1  
## v lubridate  1.9.3      v tidyr      1.3.1  
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
# To find the Number of Rows each data frame has we can use "nrow()"  
# To find the Number of Columns of each data frame we can use "ncol()"
```

```
#Data Frame - debt_df  
#Number of columns in debt_df  
ncol(debt_df)
```

```
## [1] 63
```

```
#Number of rows in debt_df  
nrow(debt_df)
```

```
## [1] 13824
```

```
#Data Frame - country_df  
#Number of columns in country_df  
ncol(country_df)
```

```
## [1] 5
```

```
#Number of rows in country_df  
nrow(country_df)
```

```
## [1] 216
```

```
#Data Frame - country_df  
#Number of columns in country_df  
ncol(indicator_df)
```

```
## [1] 2
```

```
#Number of rows in country_df
nrow(indicator_df)
```

```
## [1] 61
```

(Q2). Update “debt\_df” by reordering its rows such that the values of the indicator “DT.NFL.BLAT.CD” is in descending order. Display a subset of the updated “debt\_df” consisting of the first 4 rows and the columns “Country.Code”, “Year”, “NY.GNP.MKTP.CD”, and “DT.NFL.BLAT.CD”.

```
#Firstly, we need to update "debt_df" by reordering its rows with the values of indicators in descending order.
# Before proceeding to that step let us print all the column headers in the data
```

```
# Checking ofr the column names of the data frame
colnames(debt_df)
```

```
## [1] "Country.Code"      "Year"              "NY.GNP.MKTP.CD"
## [4] "FI.RES.TOTL.MO"    "FI.RES.TOTL.DT.ZS" "FI.RES.TOTL.CD"
## [7] "DT.TDS.MLAT.PG.ZS" "DT.TDS.MLAT.CD"    "DT.TDS.DPPG.XP.ZS"
## [10] "DT.TDS.DPPG.GN.ZS" "DT.TDS.DPPG.CD"    "DT.TDS.DPPF.XP.ZS"
## [13] "DT.TDS.DIMF.CD"    "DT.TDS.DECT.GN.ZS" "DT.TDS.DECT.EX.ZS"
## [16] "DT.TDS.DECT.CD"    "DT.ODA.ODAT.PC.ZS" "DT.ODA.ODAT.GN.ZS"
## [19] "DT.ODA.ODAT.CD"    "DT.NFL.RDBN.CD"    "DT.NFL.RDBC.CD"
## [22] "DT.NFL.PRVT.CD"    "DT.NFL.PROP.CD"    "DT.NFL.PNGC.CD"
## [25] "DT.NFL.PNGB.CD"    "DT.NFL.PCBO.CD"    "DT.NFL.PCBK.CD"
## [28] "DT.NFL.PBND.CD"    "DT.NFL.OFFT.CD"    "DT.NFL.NIFC.CD"
## [31] "DT.NFL.MOTH.CD"    "DT.NFL.MLAT.CD"    "DT.NFL.MIDA.CD"
## [34] "DT.NFL.MIBR.CD"    "DT.NFL.IMFN.CD"    "DT.NFL.IMFC.CD"
## [37] "DT.NFL.DPNG.CD"    "DT.NFL.BOND.CD"    "DT.NFL.BLAT.CD"
## [40] "DT.DOD.PVLX.GN.ZS" "DT.DOD.PVLX.EX.ZS" "DT.DOD.PVLX.CD"
## [43] "DT.DOD.MWBG.CD"    "DT.DOD.MIDA.CD"    "DT.DOD.MIBR.CD"
## [46] "DT.DOD.DSTC.ZS"    "DT.DOD.DSTC.XP.ZS" "DT.DOD.DSTC.IR.ZS"
## [49] "DT.DOD.DSTC.CD"    "DT.DOD.DPPG.CD"    "DT.DOD.DPNG.CD"
## [52] "DT.DOD.DLXF.CD"    "DT.DOD.DIMF.CD"    "DT.DOD.DECT.GN.ZS"
## [55] "DT.DOD.DECT.CD"    "BX.TRF.PWKR.CD.DT" "BX.PEF.TOTL.CD.WD"
## [58] "BX.KLT.DINV.CD.WD" "BX.GSR.TOTL.CD"    "BX.GRT.TECH.CD.WD"
## [61] "BX.GRT.EXTA.CD.WD" "BN.CAB.XOKA.CD"    "BM.GSR.TOTL.CD"
```

```
#Arranging the Column name "DT.NFL.BLAT.CD" in descnding order using arrange(desc())
```

```
# debt_df %>% arrange(desc(DT.NFL.BLAT.CD))
```

```
debt_df <- debt_df %>% arrange(desc(DT.NFL.BLAT.CD))
```

```
#Subset of the updated "debt_df" consisting of the first 4 rows and the columns "Country.Code", "Year",
```

```
subset_debt_df <- debt_df %>%
  select(Country.Code, Year, NY.GNP.MKTP.CD, DT.NFL.BLAT.CD) %>%
  head(4)
```

```
#Displaying subset_debt_df
subset_debt_df
```

```
## # A tibble: 4 x 4
##   Country.Code Year      NY.GNP.MKTP.CD DT.NFL.BLAT.CD
##   <chr>         <chr>          <dbl>          <dbl>
## 1 MEX          year_1995      366827000000    9398190731
## 2 EGY          year_2013      281028000000    7233642176
## 3 BRA          year_2017      202494000000    6506490468
## 4 PAK          year_2018      350691000000    6201281870
```

```
#Displaying debt_df
debt_df
```

```
## # A tibble: 13,824 x 63
##   Country.Code Year      NY.GNP.MKTP.CD FI.RES.TOTL.MO FI.RES.TOTL.DT.ZS
##   <chr>         <chr>          <dbl>          <dbl>          <dbl>
## 1 MEX          year_1995      366827000000      2.83          10.2
## 2 EGY          year_2013      281028000000      2.73          35.5
## 3 BRA          year_2017      202494000000     14.9          68.9
## 4 PAK          year_2018      350691000000      1.91          11.8
## 5 EGY          year_2016      327970000000      3.89          34.2
## 6 CHN          year_1995      722711000000      6.33          67.9
## 7 BGD          year_2022      478361000000      4.18          34.8
## 8 THA          year_1997      145706000000      4.05          24.5
## 9 EGY          year_2015      323666000000      2.60          31.8
## 10 BGD         year_2021      438182000000      6.29          50.5
## # i 13,814 more rows
## # i 58 more variables: FI.RES.TOTL.CD <dbl>, DT.TDS.MLAT.PG.ZS <dbl>,
## #   DT.TDS.MLAT.CD <dbl>, DT.TDS.DPPG.XP.ZS <dbl>, DT.TDS.DPPG.GN.ZS <dbl>,
## #   DT.TDS.DPPG.CD <dbl>, DT.TDS.DPPF.XP.ZS <dbl>, DT.TDS.DIMF.CD <dbl>,
## #   DT.TDS.DECT.GN.ZS <dbl>, DT.TDS.DECT.EX.ZS <dbl>, DT.TDS.DECT.CD <dbl>,
## #   DT.ODA.ODAT.PC.ZS <dbl>, DT.ODA.ODAT.GN.ZS <dbl>, DT.ODA.ODAT.CD <dbl>,
## #   DT.NFL.RDBN.CD <dbl>, DT.NFL.RDBC.CD <dbl>, DT.NFL.PRVT.CD <dbl>, ...
```

```
#debt_df_subset <- debt_df %>% arrange(desc(DT.NFL.BLAT.CD), na.rm = TRUE) %>%
#   select(Country.Code, Year, NY.GNP.MKTP.CD, DT.NFL.BLAT.CD) %>%
#   head(4)

#print(debt_df_subset)
```

(Q3). In the data frame “debt\_df”, the indicators are represented by their associated indicator codes rather than by their names. The data frame “indicator\_df” contains a list of indicator names and their corresponding indicator codes. Create a new data frame called “debt\_df2” by combining the data from the two data frames “debt\_df” and “indicator\_df”. The new data frame “debt\_df2” should be equivalent to “debt\_df” except that “debt\_df2” now contains indicator names codes. The indicator names in “debt\_df2” should match the indicator codes in “debt\_df” according to their correspondence described in “indicator\_df”. Display a subset of “debt\_df2” consisting of the first 5 rows and the three columns “Country.Code”, “Year”, and “Net financial flows, others (NFL, current US\$)”.

```
library(dplyr)
library(tidyr)
```

```

# In this question we need to rename the indicator code by their corresponding indicator names from "in
#Let us print the column headers of inddicator_df and the head to understand the dataframe
# creating a vector called "rename_map" for renaming the columns in indicator_df data frame
colnames(indicator_df)

## [1] "INDICATOR_CODE" "INDICATOR_NAME"

head(indicator_df)

```

```

## # A tibble: 6 x 2
##   INDICATOR_CODE    INDICATOR_NAME
##   <chr>           <chr>
## 1 NY.GNP.MKTP.CD    GNI (current US$)
## 2 FI.RES.TOTL.MO    Total reserves in months of imports
## 3 FI.RES.TOTL.DT.ZS Total reserves (% of total external debt)
## 4 FI.RES.TOTL.CD    Total reserves (includes gold, current US$)
## 5 DT.TDS.MLAT.PG.ZS Multilateral debt service (% of public and publicly guarant~
## 6 DT.TDS.MLAT.CD    Multilateral debt service (TDS, current US$)

```

```

# renaming_vector is a vector which is mapping the indicstor code to its corresponding indicator name h
renaming_vector <- setNames(indicator_df$INDICATOR_NAME, indicator_df$INDICATOR_CODE)

# Using the vector "renaming_vector" now we can rename the columns in debt_df
#.cols is the columns we need to rename in debt_df
debt_df2 <- debt_df %>%
  rename_with(~ renaming_vector[.x], .cols = names(renaming_vector))

#Diaplaying the colnames to check if the Indicator codes are renamed by the indicator names.
colnames(debt_df2)

```

```

## [1] "Country.Code"
## [2] "Year"
## [3] "GNI (current US$)"
## [4] "Total reserves in months of imports"
## [5] "Total reserves (% of total external debt)"
## [6] "Total reserves (includes gold, current US$)"
## [7] "Multilateral debt service (% of public and publicly guaranteed debt service)"
## [8] "Multilateral debt service (TDS, current US$)"
## [9] "Public and publicly guaranteed debt service (% of exports of goods, services and primary income)"
## [10] "Public and publicly guaranteed debt service (% of GNI)"
## [11] "Debt service on external debt, public and publicly guaranteed (PPG) (TDS, current US$)"
## [12] "Debt service (PPG and IMF only, % of exports of goods, services and primary income)"
## [13] "IMF repurchases and charges (TDS, current US$)"
## [14] "Total debt service (% of GNI)"
## [15] "Total debt service (% of exports of goods, services and primary income)"
## [16] "Debt service on external debt, total (TDS, current US$)"
## [17] "Net ODA received per capita (current US$)"
## [18] "Net ODA received (% of GNI)"
## [19] "Net official development assistance received (current US$)"

```

```

## [20] "Net financial flows, RDB nonconcessional (NFL, current US$)"
## [21] "Net financial flows, RDB concessional (NFL, current US$)"
## [22] "PPG, private creditors (NFL, US$)"
## [23] "PPG, other private creditors (NFL, current US$)"
## [24] "PNG, commercial banks and other creditors (NFL, current US$)"
## [25] "PNG, bonds (NFL, current US$)"
## [26] "Commercial banks and other lending (PPG + PNG) (NFL, current US$)"
## [27] "PPG, commercial banks (NFL, current US$)"
## [28] "PPG, bonds (NFL, current US$)"
## [29] "PPG, official creditors (NFL, US$)"
## [30] "IFC, private nonguaranteed (NFL, US$)"
## [31] "Net financial flows, others (NFL, current US$)"
## [32] "Net financial flows, multilateral (NFL, current US$)"
## [33] "Net financial flows, IDA (NFL, current US$)"
## [34] "Net financial flows, IBRD (NFL, current US$)"
## [35] "Net financial flows, IMF nonconcessional (NFL, current US$)"
## [36] "Net financial flows, IMF concessional (NFL, current US$)"
## [37] "Net flows on external debt, private nonguaranteed (PNG) (NFL, current US$)"
## [38] "Portfolio investment, bonds (PPG + PNG) (NFL, current US$)"
## [39] "Net financial flows, bilateral (NFL, current US$)"
## [40] "Present value of external debt (% of GNI)"
## [41] "Present value of external debt (% of exports of goods, services and primary income)"
## [42] "Present value of external debt (current US$)"
## [43] "IBRD loans and IDA credits (DOD, current US$)"
## [44] "PPG, IDA (DOD, current US$)"
## [45] "PPG, IBRD (DOD, current US$)"
## [46] "Short-term debt (% of total external debt)"
## [47] "Short-term debt (% of exports of goods, services and primary income)"
## [48] "Short-term debt (% of total reserves)"
## [49] "External debt stocks, short-term (DOD, current US$)"
## [50] "External debt stocks, public and publicly guaranteed (PPG) (DOD, current US$)"
## [51] "External debt stocks, private nonguaranteed (PNG) (DOD, current US$)"
## [52] "External debt stocks, long-term (DOD, current US$)"
## [53] "Use of IMF credit (DOD, current US$)"
## [54] "External debt stocks (% of GNI)"
## [55] "External debt stocks, total (DOD, current US$)"
## [56] "Personal remittances, received (current US$)"
## [57] "Portfolio equity, net inflows (BoP, current US$)"
## [58] "Foreign direct investment, net inflows (BoP, current US$)"
## [59] "Exports of goods, services and primary income (BoP, current US$)"
## [60] "Technical cooperation grants (BoP, current US$)"
## [61] "Grants, excluding technical cooperation (BoP, current US$)"
## [62] "Current account balance (BoP, current US$)"
## [63] "Imports of goods, services and primary income (BoP, current US$)"

```

*# According to the question Creating a subset with the first 5 rows and the three columns "Country.Code*

```

subset_debt_df2 <- debt_df2 %>%
  select(Country.Code, Year, `Net financial flows, others (NFL, current US$)` ) %>%
  head(5)

subset_debt_df2

```

```
## # A tibble: 5 x 3
```

```
## Country.Code Year `Net financial flows, others (NFL, current US$)`
## <chr> <chr> <dbl>
## 1 MEX year_1995 NA
## 2 EGY year_2013 -14314777.
## 3 BRA year_2017 -195705180.
## 4 PAK year_2018 321846510.
## 5 EGY year_2016 2141976215
```

(Q4). The data frame “country\_df” contains information about Region, Income groups, and country name for each country. Create a new data frame called “debt\_df3” by combining data from the two data frames “debt\_df2” and “country\_df”. The new data frame “debt\_df3” should contain a) all columns from “debt\_df2” and b) 3 columns from “country\_df” called “Region”, “IncomeGroup”, and “Country.Name”. Make sure that in each row of “debt\_df3”, the “Region”, “IncomeGroup”, and “Country.Name” match “Country.Code” according to their correspondence described in “country\_df”. Your data frames “debt\_df3” and “debt\_df2” should have the same numbers of rows, but “debt\_df3” has three more columns. Display a subset of “debt\_df3” consisting of the first three rows and 4 columns called “Country.Name”, “IncomeGroup”, “Year”, and “Total reserves in months of imports”.

```
# Checking of the column names of the data frame - "country_df"
colnames(country_df)
```

```
## [1] "Country.Code" "Region" "IncomeGroup" "SpecialNotes" "Country.Name"
```

```
# Create a new data frame called "debt_df3" by combining data from the two data frames "debt_df2" and "
# The data frame - "debt_df3" should contain
# a) all columns from "debt_df2"
# b) 3 columns from "country_df" called "Region", "IncomeGroup", and "Country.Name"
```

```
# Therefore we can duplicate the dataframe debt_df3 from debt_df2 and then left join to "country_df" to
# After the left join, we are selecting "Region", "IncomeGroup", and "Country.Name" for print or display
# First duplicating the data frame debt_df2 into debt_df3
```

```
debt_df3 <- debt_df2
```

```
debt_df3 <- debt_df3 %>%
  left_join(country_df %>% select(Country.Code, Region, IncomeGroup, Country.Name),
    by = "Country.Code")
```

```
#Printing the no of columns in debt_df2 and debt_df3 to check if 3 columns are added and no. of rows remain same
ncol(debt_df2)
```

```
## [1] 63
```

```
ncol(debt_df3)
```

```
## [1] 66
```



```
nrow(debt_df2)
```

```
## [1] 13824
```

```
nrow(debt_df3)
```

```
## [1] 13824
```

*#Creating the subset\_debt\_df3 so that the original debt\_df3 can be used for the further questions below*

*# Creating a subset\_debt\_df3 to display 3 rows of only the headers - Country.Name, IncomeGroup, Year,*

```
subset_debt_df3 <- debt_df3 %>%
```

```
  select(Country.Name, IncomeGroup, Year, `Total reserves in months of imports`) %>%
```

```
  head(3)
```

```
print(subset_debt_df3)
```

```
## # A tibble: 3 x 4
```

```
##   Country.Name      IncomeGroup      Year      Total reserves in months of i-1
```

```
##   <chr>            <chr>          <chr>          <dbl>
```

```
## 1 Mexico          Upper middle income year_1995          2.83
```

```
## 2 Egypt, Arab Rep. Lower middle income year_2013          2.73
```

```
## 3 Brazil          Upper middle income year_2017          14.9
```

```
## # i abbreviated name: 1: `Total reserves in months of imports`
```

**(Q5). Rename the following 5 columns from their original names to the new names specified below**

Original column names	New column names
Total reserves in months of imports	Total_reserves
External debt stocks, total (DOD, current US\$)	External_debt
Net financial flows, bilateral (NFL, current US\$)	Financial_flow
Imports of goods, services and primary income (BoP, current US\$)	Imports
IFC, private nonguaranteed (NFL, US\$)	IFC

*# Using rename() function to rename the column headers from their original names to new names*

*# Rename columns in debt\_df3 for clarity*

*# Checking of the column names of the data frame- "debt\_df3"*

```
colnames(debt_df3)
```

```
## [1] "Country.Code"
```

```
## [2] "Year"
```

```
## [3] "GNI (current US$)"
```

```
## [4] "Total reserves in months of imports"
```

```
## [5] "Total reserves (% of total external debt)"
```

```
## [6] "Total reserves (includes gold, current US$)"
```

```
## [7] "Multilateral debt service (% of public and publicly guaranteed debt service)"
```

```
## [8] "Multilateral debt service (TDS, current US$)"
```

## [9] "Public and publicly guaranteed debt service (% of exports of goods, services and primary income)"  
 ## [10] "Public and publicly guaranteed debt service (% of GNI)"  
 ## [11] "Debt service on external debt, public and publicly guaranteed (PPG) (TDS, current US\$)"  
 ## [12] "Debt service (PPG and IMF only, % of exports of goods, services and primary income)"  
 ## [13] "IMF repurchases and charges (TDS, current US\$)"  
 ## [14] "Total debt service (% of GNI)"  
 ## [15] "Total debt service (% of exports of goods, services and primary income)"  
 ## [16] "Debt service on external debt, total (TDS, current US\$)"  
 ## [17] "Net ODA received per capita (current US\$)"  
 ## [18] "Net ODA received (% of GNI)"  
 ## [19] "Net official development assistance received (current US\$)"  
 ## [20] "Net financial flows, RDB nonconcessional (NFL, current US\$)"  
 ## [21] "Net financial flows, RDB concessional (NFL, current US\$)"  
 ## [22] "PPG, private creditors (NFL, US\$)"  
 ## [23] "PPG, other private creditors (NFL, current US\$)"  
 ## [24] "PNG, commercial banks and other creditors (NFL, current US\$)"  
 ## [25] "PNG, bonds (NFL, current US\$)"  
 ## [26] "Commercial banks and other lending (PPG + PNG) (NFL, current US\$)"  
 ## [27] "PPG, commercial banks (NFL, current US\$)"  
 ## [28] "PPG, bonds (NFL, current US\$)"  
 ## [29] "PPG, official creditors (NFL, US\$)"  
 ## [30] "IFC, private nonguaranteed (NFL, US\$)"  
 ## [31] "Net financial flows, others (NFL, current US\$)"  
 ## [32] "Net financial flows, multilateral (NFL, current US\$)"  
 ## [33] "Net financial flows, IDA (NFL, current US\$)"  
 ## [34] "Net financial flows, IBRD (NFL, current US\$)"  
 ## [35] "Net financial flows, IMF nonconcessional (NFL, current US\$)"  
 ## [36] "Net financial flows, IMF concessional (NFL, current US\$)"  
 ## [37] "Net flows on external debt, private nonguaranteed (PNG) (NFL, current US\$)"  
 ## [38] "Portfolio investment, bonds (PPG + PNG) (NFL, current US\$)"  
 ## [39] "Net financial flows, bilateral (NFL, current US\$)"  
 ## [40] "Present value of external debt (% of GNI)"  
 ## [41] "Present value of external debt (% of exports of goods, services and primary income)"  
 ## [42] "Present value of external debt (current US\$)"  
 ## [43] "IBRD loans and IDA credits (DOD, current US\$)"  
 ## [44] "PPG, IDA (DOD, current US\$)"  
 ## [45] "PPG, IBRD (DOD, current US\$)"  
 ## [46] "Short-term debt (% of total external debt)"  
 ## [47] "Short-term debt (% of exports of goods, services and primary income)"  
 ## [48] "Short-term debt (% of total reserves)"  
 ## [49] "External debt stocks, short-term (DOD, current US\$)"  
 ## [50] "External debt stocks, public and publicly guaranteed (PPG) (DOD, current US\$)"  
 ## [51] "External debt stocks, private nonguaranteed (PNG) (DOD, current US\$)"  
 ## [52] "External debt stocks, long-term (DOD, current US\$)"  
 ## [53] "Use of IMF credit (DOD, current US\$)"  
 ## [54] "External debt stocks (% of GNI)"  
 ## [55] "External debt stocks, total (DOD, current US\$)"  
 ## [56] "Personal remittances, received (current US\$)"  
 ## [57] "Portfolio equity, net inflows (BoP, current US\$)"  
 ## [58] "Foreign direct investment, net inflows (BoP, current US\$)"  
 ## [59] "Exports of goods, services and primary income (BoP, current US\$)"  
 ## [60] "Technical cooperation grants (BoP, current US\$)"  
 ## [61] "Grants, excluding technical cooperation (BoP, current US\$)"  
 ## [62] "Current account balance (BoP, current US\$)"

```
## [63] "Imports of goods, services and primary income (BoP, current US$)"
## [64] "Region"
## [65] "IncomeGroup"
## [66] "Country.Name"
```

*#Renaming the column headers*

```
debt_df3 <- debt_df3 %>%
  rename(
    Total_reserves = "Total reserves in months of imports",
    External_debt = "External debt stocks, total (DOD, current US$)",
    Financial_flow = "Net financial flows, bilateral (NFL, current US$)",
    Imports = "Imports of goods, services and primary income (BoP, current US$)",
    IFC = "IFC, private nonguaranteed (NFL, US$)"
  )
```

*#Print the head of debt\_df3 to check.*

```
head(debt_df3)
```

```
## # A tibble: 6 x 66
##   Country.Code Year   `GNI (current US$)` Total_reserves Total reserves (% of~1
##   <chr>         <chr>         <dbl>         <dbl>         <dbl>
## 1 MEX          year_1~          366827000000          2.83          10.2
## 2 EGY          year_2~          281028000000          2.73          35.5
## 3 BRA          year_2~          202494000000          14.9          68.9
## 4 PAK          year_2~          350691000000          1.91          11.8
## 5 EGY          year_2~          327970000000          3.89          34.2
## 6 CHN          year_1~          722711000000          6.33          67.9
## # i abbreviated name: 1: `Total reserves (% of total external debt)`
## # i 61 more variables: `Total reserves (includes gold, current US$)` <dbl>,
## #   `Multilateral debt service (% of public and publicly guaranteed debt service)` <dbl>,
## #   `Multilateral debt service (TDS, current US$)` <dbl>,
## #   `Public and publicly guaranteed debt service (% of exports of goods, services and primary income)` <dbl>,
## #   `Public and publicly guaranteed debt service (% of GNI)` <dbl>,
## #   `Debt service on external debt, public and publicly guaranteed (PPG) (TDS, current US$)` <dbl>,
```

*#creating a subset\_debt\_renamed\_columns to print and check if the column headers are renamed by only s*

```
subset_debt_renamed_columns <- debt_df3 %>%
  select(Total_reserves, External_debt, Year, Financial_flow, Imports, IFC) %>%
  head(4)

print(subset_debt_renamed_columns)
```

```
## # A tibble: 4 x 6
##   Total_reserves External_debt Year   Financial_flow Imports IFC
##   <dbl>         <dbl> <chr>         <dbl>         <dbl> <dbl>
## 1      2.83    166734000000 year_1995    9398190731  72391910000      0
## 2      2.73     46534987115 year_2013    7233642176  72685700000 -42864095
## 3     14.9    543000000000 year_2017    6506490468  301961000000 397855350
## 4      1.91    100199000000 year_2018    6201281870  74555877000 11389136
```

(Q6). Next generate a summary data frame called “debt\_summary” from the data frame “debt\_df3” with the following properties: Your summary data frame “debt\_summary” should contain 7 rows corresponding to the 7 different Regions, and it should also have 5 columns:

“Region” - the names of the 7 different regions including “East Asia & Pacific”, “Europe & Central Asia” etc.  
 “TR\_mn” - the average of “Total\_reserves” in each region. “ED\_md” - the median of “External\_debt” in each region. “FF\_quantile” - the 0.2 quantile of “Financial\_flow” in each region. “IFC\_sd” - the standard deviation of “IFC” in each region. All missing values should be discarded when computing the summary data.

```
# Checking of the column names of the data frame- "debt_df3" to generate a summary
colnames(debt_df3)
```

```
## [1] "Country.Code"
## [2] "Year"
## [3] "GNI (current US$)"
## [4] "Total_reserves"
## [5] "Total reserves (% of total external debt)"
## [6] "Total reserves (includes gold, current US$)"
## [7] "Multilateral debt service (% of public and publicly guaranteed debt service)"
## [8] "Multilateral debt service (TDS, current US$)"
## [9] "Public and publicly guaranteed debt service (% of exports of goods, services and primary income)"
## [10] "Public and publicly guaranteed debt service (% of GNI)"
## [11] "Debt service on external debt, public and publicly guaranteed (PPG) (TDS, current US$)"
## [12] "Debt service (PPG and IMF only, % of exports of goods, services and primary income)"
## [13] "IMF repurchases and charges (TDS, current US$)"
## [14] "Total debt service (% of GNI)"
## [15] "Total debt service (% of exports of goods, services and primary income)"
## [16] "Debt service on external debt, total (TDS, current US$)"
## [17] "Net ODA received per capita (current US$)"
## [18] "Net ODA received (% of GNI)"
## [19] "Net official development assistance received (current US$)"
## [20] "Net financial flows, RDB nonconcessional (NFL, current US$)"
## [21] "Net financial flows, RDB concessional (NFL, current US$)"
## [22] "PPG, private creditors (NFL, US$)"
## [23] "PPG, other private creditors (NFL, current US$)"
## [24] "PNG, commercial banks and other creditors (NFL, current US$)"
## [25] "PNG, bonds (NFL, current US$)"
## [26] "Commercial banks and other lending (PPG + PNG) (NFL, current US$)"
## [27] "PPG, commercial banks (NFL, current US$)"
## [28] "PPG, bonds (NFL, current US$)"
## [29] "PPG, official creditors (NFL, US$)"
## [30] "IFC"
## [31] "Net financial flows, others (NFL, current US$)"
## [32] "Net financial flows, multilateral (NFL, current US$)"
## [33] "Net financial flows, IDA (NFL, current US$)"
## [34] "Net financial flows, IBRD (NFL, current US$)"
## [35] "Net financial flows, IMF nonconcessional (NFL, current US$)"
## [36] "Net financial flows, IMF concessional (NFL, current US$)"
## [37] "Net flows on external debt, private nonguaranteed (PNG) (NFL, current US$)"
## [38] "Portfolio investment, bonds (PPG + PNG) (NFL, current US$)"
## [39] "Financial_flow"
## [40] "Present value of external debt (% of GNI)"
## [41] "Present value of external debt (% of exports of goods, services and primary income)"
```

```
## [42] "Present value of external debt (current US$)"
## [43] "IBRD loans and IDA credits (DOD, current US$)"
## [44] "PPG, IDA (DOD, current US$)"
## [45] "PPG, IBRD (DOD, current US$)"
## [46] "Short-term debt (% of total external debt)"
## [47] "Short-term debt (% of exports of goods, services and primary income)"
## [48] "Short-term debt (% of total reserves)"
## [49] "External debt stocks, short-term (DOD, current US$)"
## [50] "External debt stocks, public and publicly guaranteed (PPG) (DOD, current US$)"
## [51] "External debt stocks, private nonguaranteed (PNG) (DOD, current US$)"
## [52] "External debt stocks, long-term (DOD, current US$)"
## [53] "Use of IMF credit (DOD, current US$)"
## [54] "External debt stocks (% of GNI)"
## [55] "External_debt"
## [56] "Personal remittances, received (current US$)"
## [57] "Portfolio equity, net inflows (BoP, current US$)"
## [58] "Foreign direct investment, net inflows (BoP, current US$)"
## [59] "Exports of goods, services and primary income (BoP, current US$)"
## [60] "Technical cooperation grants (BoP, current US$)"
## [61] "Grants, excluding technical cooperation (BoP, current US$)"
## [62] "Current account balance (BoP, current US$)"
## [63] "Imports"
## [64] "Region"
## [65] "IncomeGroup"
## [66] "Country.Name"
```

*#The column headers - "Region", "Total\_reserves", "External\_debt", "Financial\_flow", "IFC" are all present*

*#Now creating a summary according to the question*

```
debt_summary <- debt_df3 %>% #Creating a data frame debt_summary
  group_by(Region) %>% #grouping the data by Region
  summarise(
    TR_mn = mean(Total_reserves, na.rm = TRUE), #Average is the mean of the Total_reserves, Using the mean() function
    ED_md = median(External_debt, na.rm = TRUE), #Using the median() function in External_debt also specifies na.rm = TRUE
    FF_quantile = quantile(Financial_flow, probs = 0.2, na.rm = TRUE), #Using the quantile() function in Financial_flow also specifies na.rm = TRUE
    IFC_sd = sd(IFC, na.rm = TRUE) #Using the sd() function for calculating the standard deviation also specifies na.rm = TRUE
  )
```

*# Displaying the "debt\_summary" data frame*

```
print(debt_summary)
```

```
## # A tibble: 7 x 5
##   Region          TR_mn      ED_md FF_quantile    IFC_sd
##   <chr>          <dbl>      <dbl>      <dbl>      <dbl>
## 1 East Asia & Pacific  5.19 2248479410 -2357020. 52498519.
## 2 Europe & Central Asia  3.58 8237728122 -53631246. 50820255.
## 3 Latin America & Caribbean  3.84 4159662669 -25144268. 62054545.
## 4 Middle East & North Africa  7.72 7481954468 -92269932. 21414719.
## 5 North America      1.99      NA      NA      NA
## 6 South Asia         4.94 4940329805 -373253. 76630044.
## 7 Sub-Saharan Africa  3.32 1709094992 -1673594. 24748455.
```

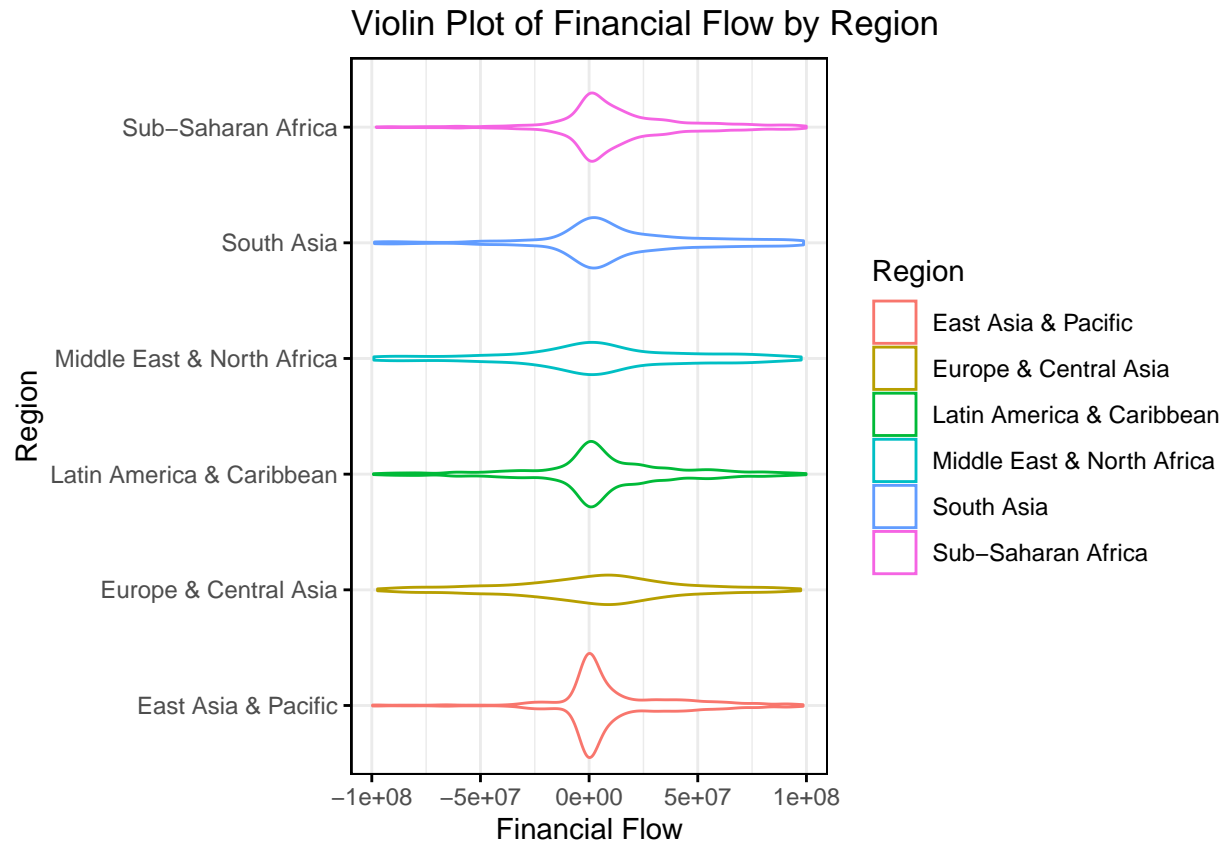
(Q7). Based on your data frame “debt\_df3”, create a violin plot of “Financial\_flow” for each of the regions. The violin plots should be displayed in the same figure and with different colors representing different regions. Ignore all missing values and all values that are smaller than -108 or bigger than 108.

```
# Using the package ggplot2 for visualization and plotting a violin plot of "Financial_flow" for each o

library(ggplot2)

# Ignoring and Filtering out all missing values and all values that are smaller than -108 or bigger than
debt_df3 %>%
  filter(Financial_flow > -10^8 & Financial_flow < 10^8, !is.na(Financial_flow)) %>% #this line is remo
#using ggplot to use violin plotting and specifying x axis as Financial_Flow, and y axis as Region an
  ggplot(aes(x = Financial_flow, y = Region, color = Region)) +
  geom_violin() + #Creating the violin plot
  labs(title = "Violin Plot of Financial Flow by Region", #Naming the title of the violin plot
        x = "Financial Flow",
        y = "Region") +
  theme_minimal() +
  theme( #Here we are adding the ticks mark to each value in the x axis and y axis
        panel.border = element_rect(color = "black", fill = NA, size = 0.7),
        axis.ticks = element_line(color = "black")
  )
```

```
## Warning: The `size` argument of `element_rect()` is deprecated as of ggplot2 3.4.0.
## i Please use the `linewidth` argument instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



(Q8). Based on the data frame “debt\_df3”, create a plot which displays the “Total\_reserves” as a function of the years (from 1960 to 2023), for each of the following countries: Italy, France, United Kingdom, Sudan, Afghanistan, and Brazil. Additionally, the values of “Total\_reserves” should be displayed in different panels according to the income groups of the countries. Use different colors to represent different countries.

```
library(tidyr)
library(ggplot2)
```

```
# printing head to understand the Year column in debt_df3
head(debt_df3$Year,5)
```

```
## [1] "year_1995" "year_2013" "year_2017" "year_2018" "year_2016"
```

```
# modifying the year column by using sub() to replace the "year_" by empty string"" and then converting
debt_df3$Year <- as.integer(sub("year_", "", debt_df3$Year))
```

```
# we can also remove by library(stringr) as.integer(str_remove(debt_df3$Year, "year_"))
```

```
# filtering and cleaning the data for the plotting
```

```
subset_debt_df3 <- debt_df3 %>%
```

```
  #filtering only countries: Italy, France, United Kingdom, Sudan, Afghanistan, and Brazil by using tid
```

```

filter(Country.Name %in% c("Italy", "France", "United Kingdom", "Sudan", "Afghanistan", "Brazil") & Y
# omitting the null or na values in Year and total_reserves
drop_na(Year, Total_reserves)

# Using a line and point graph and x-axis as Year and y axis as Total_reserves and color by Country.Name
ggplot(subset_debt_df3, aes(x= Year, y= Total_reserves, color = Country.Name)) + # Giving the plot an a
  labs( title = "Total Reserves Over Time (1960-2023) by Country and Income Group",
        x = "Year",
        y = "Total Reserves", color = "Country Name") +
  geom_line(size = 0.5) + # using geom_line size =0.6
  geom_point(size = 1.5) + # using geom_line size = 1.7
  facet_wrap(~ IncomeGroup, scale = "fixed") +
  theme_minimal() +
  theme(
    panel.background = element_rect(fill = "White"), #plot area as white
    #border around the plot as black ith the size=0.7
    panel.border = element_rect(color = "black", fill = NA, size = 0.7),
    #To replicate the deaired plot from question using strip_background and strip.text
    strip.background = element_rect(color = "black", fill = "grey90"),
    strip.text = element_text(size = 8),
    #Here we are adding the ticks mark to each value in the x axis and y axis
    axis.ticks = element_line(color = "black")
  )

```

```

## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

```



Total Reserves Over Time (1960–2023) by Country and Income Group

