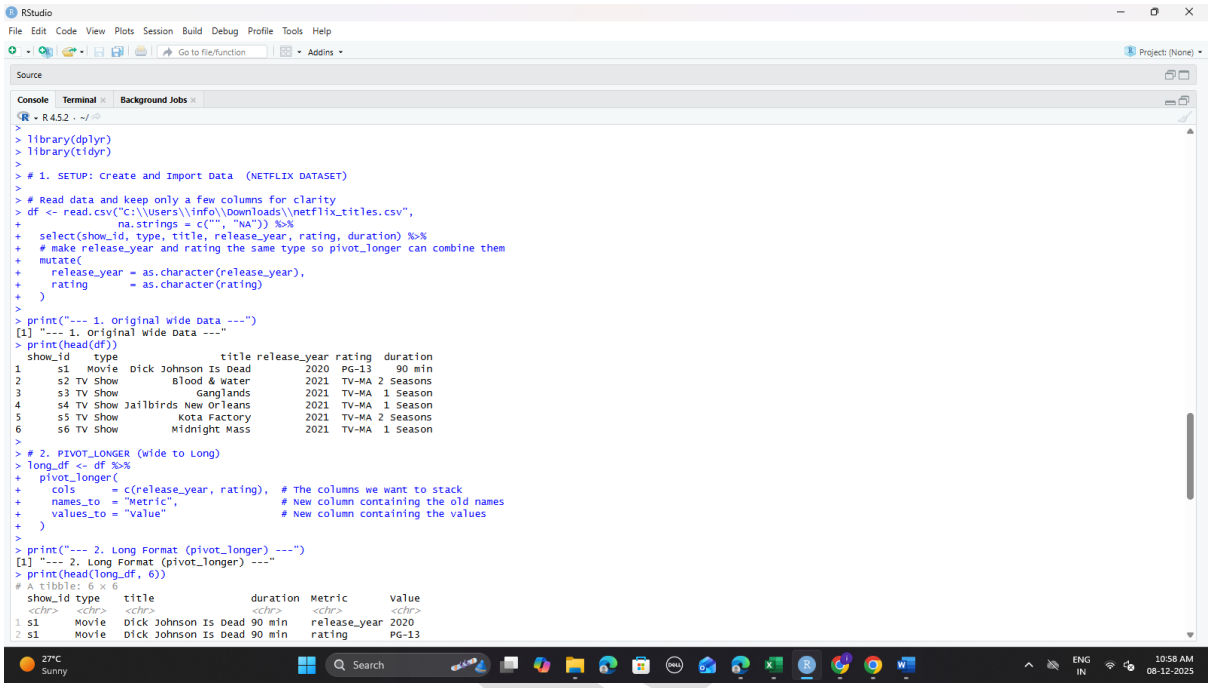
SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R

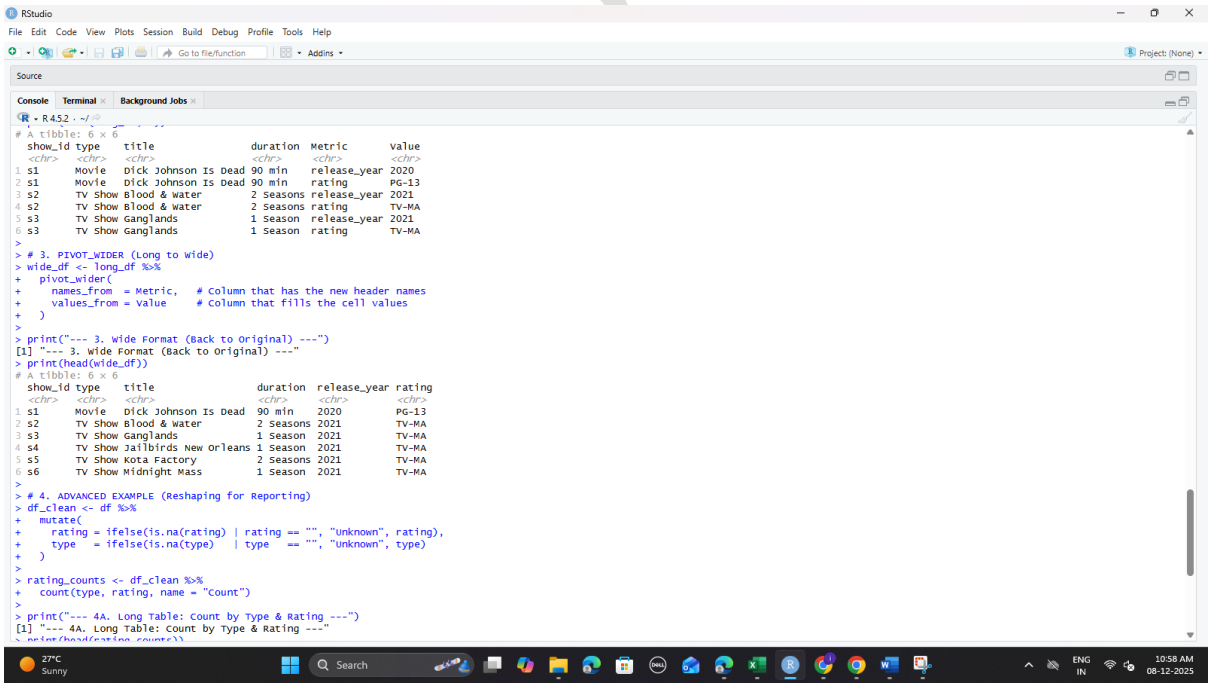
PRACTICAL – 11

AIM:- Reshaping data using pivot_longer() and pivot_wider() (R).

OUTPUT:-



```
> library(dplyr)
> library(tidyr)
>
> # 1. SETUP: Create and Import Data  (NETFLIX DATASET)
>
> # Read data and keep only a few columns for clarity
> df <- read.csv("C:\\Users\\info\\Downloads\\netflix_titles.csv",
+                na.strings = c("", "NA")) %>%
+   select(show_id, type, title, release_year, rating, duration) %>%
+   # make release_year and rating the same type so pivot_longer can combine them
+   mutate(
+     release_year = as.character(release_year),
+     rating       = as.character(rating)
+   )
>
> print("--- 1. Original Wide Data ---")
[1] "--- 1. Original Wide Data ---"
> print(head(df))
  show_id    type                title release_year rating  duration
1      s1   Movie  Dick Johnson Is Dead         2020  PG-13    90 min
2      s2 TV Show          Blood & Water         2021  TV-MA 2 Seasons
3      s3 TV Show             Ganglands         2021  TV-MA  1 Season
4      s4 TV Show Jailbirds New Orleans         2021  TV-MA  1 Season
5      s5 TV Show           Kota Factory         2021  TV-MA 2 Seasons
6      s6 TV Show          Midnight Mass         2021  TV-MA  1 Season
>
> # 2. PIVOT_LONGER (Wide to Long)
> long_df <- df %>%
+   pivot_longer(
+     cols      = c(release_year, rating),  # The columns we want to stack
+     names_to  = "Metric",                 # New column containing the old names
+     values_to = "Value"                   # New column containing the values
+   )
>
> print("--- 2. Long Format (pivot_longer) ---")
[1] "--- 2. Long Format (pivot_longer) ---"
> print(head(long_df, 6))
# A tibble: 6 × 6
  show_id type    title               duration  Metric       Value
  <chr>   <chr>   <chr>               <chr>     <chr>        <chr>
1 s1      Movie   Dick Johnson Is Dead 90 min   release_year 2020
2 s1      Movie   Dick Johnson Is Dead 90 min   rating       PG-13
```



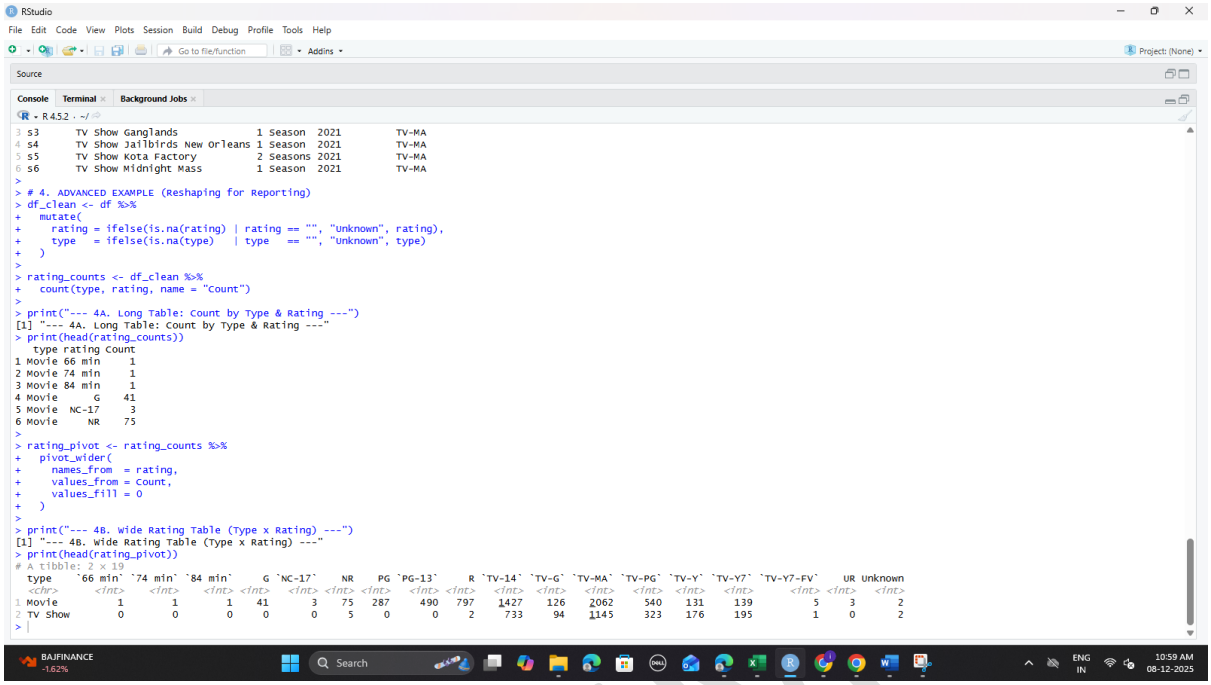
```
# A tibble: 6 × 6
  show_id type    title               duration  Metric       Value
  <chr>   <chr>   <chr>               <chr>     <chr>        <chr>
1 s1      Movie   Dick Johnson Is Dead 90 min   release_year 2020
2 s1      Movie   Dick Johnson Is Dead 90 min   rating       PG-13
3 s2      TV Show Blood & Water        2 Seasons release_year 2021
4 s2      TV Show Blood & Water        2 Seasons rating       TV-MA
5 s3      TV Show Ganglands            1 Season  release_year 2021
6 s3      TV Show Ganglands            1 Season  rating       TV-MA
>
> # 3. PIVOT_WIDER (Long to Wide)
> wide_df <- long_df %>%
+   pivot_wider(
+     names_from  = Metric,   # Column that has the new header names
+     values_from = Value     # Column that fills the cell values
+   )
>
> print("--- 3. Wide Format (Back to Original) ---")
[1] "--- 3. Wide Format (Back to Original) ---"
> print(head(wide_df))
# A tibble: 6 × 6
  show_id type    title               duration  release_year rating
  <chr>   <chr>   <chr>               <chr>     <chr>        <chr>
1 s1      Movie   Dick Johnson Is Dead  90 min    2020         PG-13
2 s2      TV Show Blood & Water         2 Seasons 2021         TV-MA
3 s3      TV Show Ganglands             1 Season  2021         TV-MA
4 s4      TV Show Jailbirds New Orleans 1 Season  2021         TV-MA
5 s5      TV Show Kota Factory          2 Seasons 2021         TV-MA
6 s6      TV Show Midnight Mass         1 Season  2021         TV-MA
>
> # 4. ADVANCED EXAMPLE (Reshaping for Reporting)
> df_clean <- df %>%
+   mutate(
+     rating = ifelse(is.na(rating) | rating == "", "Unknown", rating),
+     type   = ifelse(is.na(type)   | type   == "", "Unknown", type)
+   )
>
> rating_counts <- df_clean %>%
+   count(type, rating, name = "Count")
>
> print("--- 4A. Long Table: Count by Type & Rating ---")
[1] "--- 4A. Long Table: Count by Type & Rating ---"
```

SUMEET JITENDRA YADAV

S124

SHETH L.U.J AND SIR M.V. COLLEGE

SUBJECT :- DATA ANALYSIS WITH SAS/SPSS/R



```
RStudio
File  Edit  Code  View  Plots  Session  Build  Debug  Profile  Tools  Help
Source
Console  Terminal  Background Jobs
R 4.5.2 · ~/
3 s3      TV Show Ganglands             1 Season  2021         TV-MA
4 s4      TV Show Jailbirds New Orleans 1 Season  2021         TV-MA
5 s5      TV Show Kota Factory          2 Seasons 2021         TV-MA
6 s6      TV Show Midnight Mass         1 Season  2021         TV-MA
>
> # 4. ADVANCED EXAMPLE (Reshaping for Reporting)
> df_clean <- df %>%
+   mutate(
+     rating = ifelse(is.na(rating) | rating == "", "unknown", rating),
+     type   = ifelse(is.na(type)   | type   == "", "unknown", type)
+   )
>
> rating_counts <- df_clean %>%
+   count(type, rating, name = "Count")
>
> print("--- 4A. Long Table: Count by Type & Rating ---")
[1] "--- 4A. Long Table: Count by Type & Rating ---"
> print(head(rating_counts))
  type rating Count
1 Movie 66 min     1
2 Movie 74 min     1
3 Movie 84 min     1
4 Movie      G    41
5 Movie  NC-17     3
6 Movie     NR    75
>
> rating_pivot <- rating_counts %>%
+   pivot_wider(
+     names_from  = rating,
+     values_from = Count,
+     values_fill = 0
+   )
>
> print("--- 4B. Wide Rating Table (Type x Rating) ---")
[1] "--- 4B. Wide Rating Table (Type x Rating) ---"
> print(head(rating_pivot))
# A tibble: 2 x 19
  type    `66 min` `74 min` `84 min`     G `NC-17`    NR    PG `PG-13`     R `TV-14` `TV-G` `TV-MA` `TV-PG` `TV-Y` `TV-Y7` `TV-Y7-FV`    UR Unknown
  <chr>      <int>    <int>    <int> <int>   <int> <int> <int>   <int> <int>   <int>  <int>   <int>   <int>  <int>   <int>      <int> <int>   <int>
1 Movie          1        1        1    41       3    75   287     490   797    1427    126    2062     540    131     139          5     3       2
2 TV Show        0        0        0     0       0     5     0       0     2     733     94    1145     323    176     195          1     0       2
>
```


SUMEET JITENDRA YADAV

S124