

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R

Aim :- Reshaping data using pivot_longer()/pivot_wider() (R).

Input :



RStudio interface showing R code for data reshaping:

```
library(dplyr)
library(tidyr)

# 1. SETUP: Load billboard dataset
# Load the dataset from tidyr
df <- tidyr::billboard %>%
  mutate(SongID = row_number()) %>% # Unique ID (like your ProductID)
  select(SongID, artist, track, date.entered, starts_with("wk"))

print("--- 1. Original Wide Data ---")
print(head(df))
# Note: wk1, wk2, wk3 ... are currently in wide format.

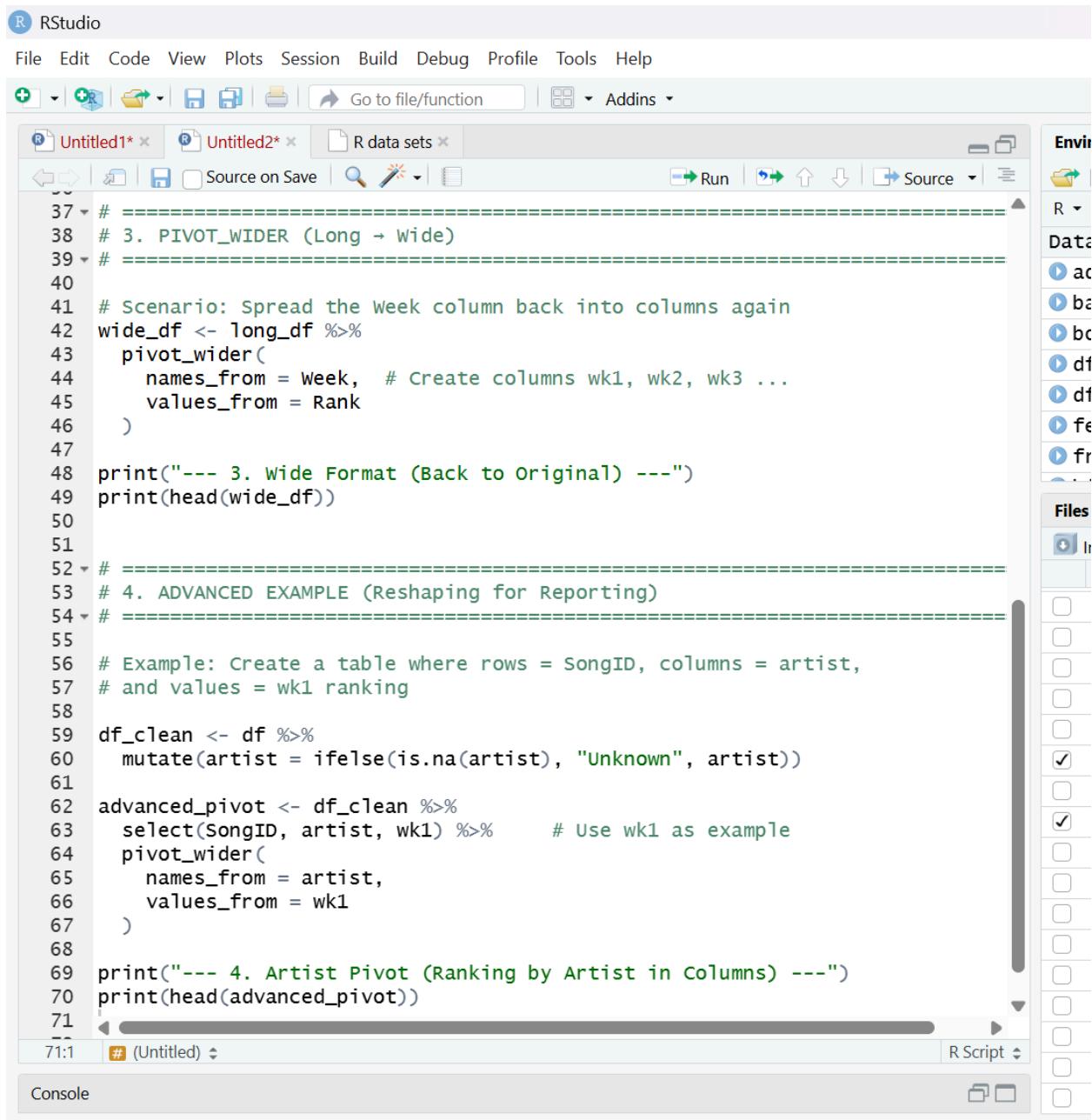
# 2. PIVOT_LONGER (wide + Long)
long_df <- df %>%
  pivot_longer(
    cols = starts_with("wk"),      # wk1, wk2, wk3, ...
    names_to = "Week",             # New column containing the old header names
    values_to = "Rank",            # New column containing the numbers
    values_drop_na = TRUE          # Remove NA week-rank combinations
  )

print("--- 2. Long Format (pivot_longer) ---")
print(head(long_df, 10))
# Notice: SongID appears many times (once per chart week)
```

The RStudio interface includes the Environment pane showing various datasets like advanced_pivot, bats_information, boston, etc., and the Packages pane listing packages such as stringr, sys, systemfonts, textshaping, tidyverse, timechange, tinytex, tzdb, utf8, uuid, vctrs, viridisLite, vroom, and withr.

Name - Mithil Kadam
Roll No - S083

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R

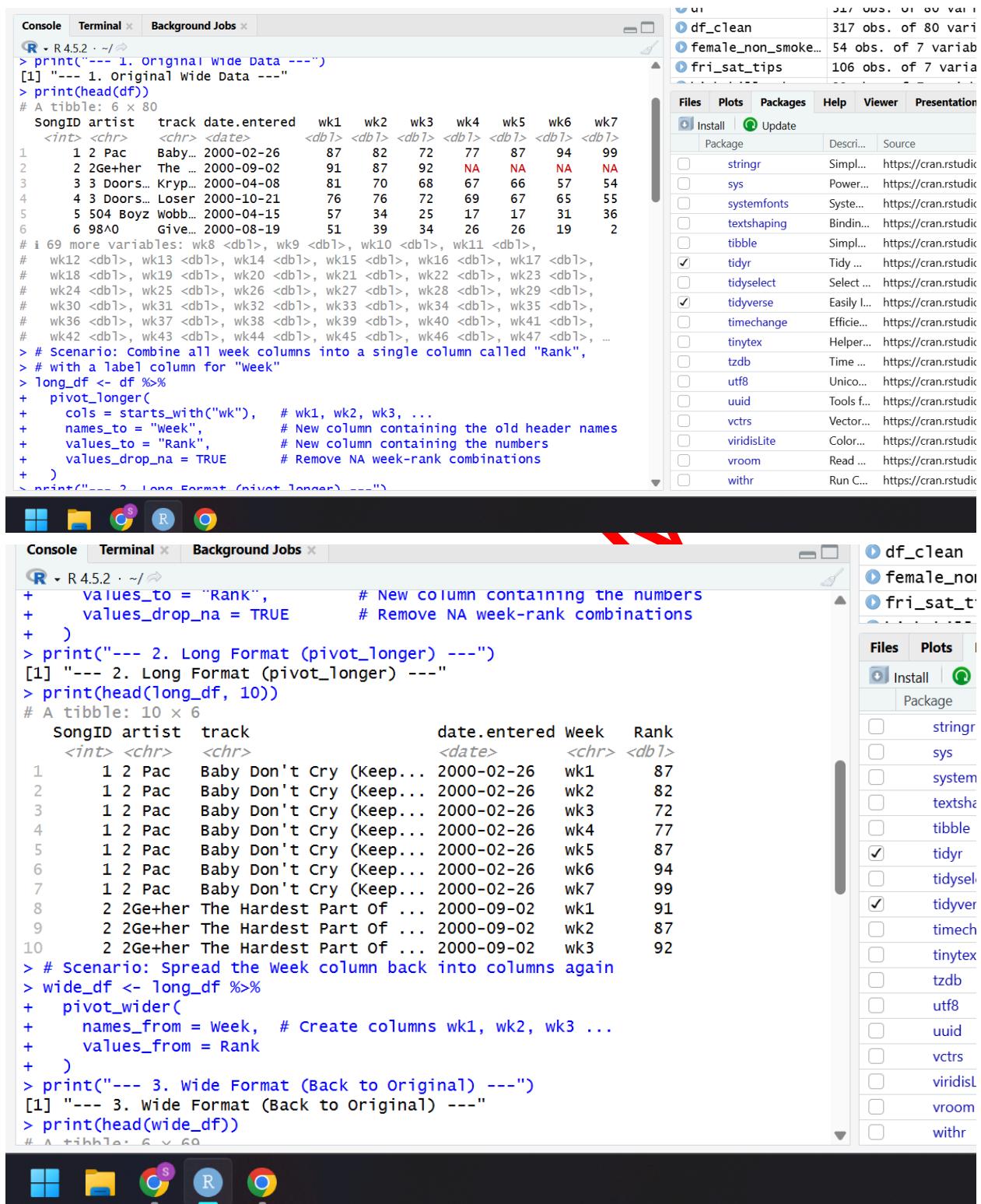


The screenshot shows the RStudio interface with the following details:

- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- Toolbar:** Includes icons for New, Open, Save, Run, Source, and Addins.
- Environment Tab:** Shows a list of objects: Envir, R, Data, ac, ba, bc, d1, d1, fe, fr,
- Code Editor:** Displays R code for data manipulation, specifically pivoting data frames. The code includes sections for "3. PIVOT_WIDER" and "4. ADVANCED EXAMPLE".
- Console Tab:** Labeled "Console" at the bottom.
- Output Tab:** Labeled "Output:" at the bottom, with a large red "M" drawn over it.

Name - Mithil Kadam
Roll No - S083

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R



The image shows two RStudio sessions running side-by-side. Both sessions are connected to R version 4.5.2.

Session 1 (Top):

```

> print("--- 1. Original Wide Data ---")
[1] "--- 1. Original Wide Data ---"
> print(head(df))
# A tibble: 6 × 80
SongID artist track date.entered wk1 wk2 wk3 wk4 wk5 wk6 wk7
<int> <chr> <chr> <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 2 Pac Baby... 2000-02-26 87 82 72 77 87 94 99
2 2 2Ge+her The ... 2000-09-02 91 87 92 NA NA NA NA
3 3 3 Doors... Kryp... 2000-04-08 81 70 68 67 66 57 54
4 4 3 Doors... Loser 2000-10-21 76 76 72 69 67 65 55
5 5 504 Boyz Wobb... 2000-04-15 57 34 25 17 17 31 36
6 6 98% Give... 2000-08-19 51 39 34 26 26 19 2
# i 69 more variables: wk8 <dbl>, wk9 <dbl>, wk10 <dbl>, wk11 <dbl>,
# wk12 <dbl>, wk13 <dbl>, wk14 <dbl>, wk15 <dbl>, wk16 <dbl>, wk17 <dbl>,
# wk18 <dbl>, wk19 <dbl>, wk20 <dbl>, wk21 <dbl>, wk22 <dbl>, wk23 <dbl>,
# wk24 <dbl>, wk25 <dbl>, wk26 <dbl>, wk27 <dbl>, wk28 <dbl>, wk29 <dbl>,
# wk30 <dbl>, wk31 <dbl>, wk32 <dbl>, wk33 <dbl>, wk34 <dbl>, wk35 <dbl>,
# wk36 <dbl>, wk37 <dbl>, wk38 <dbl>, wk39 <dbl>, wk40 <dbl>, wk41 <dbl>,
# wk42 <dbl>, wk43 <dbl>, wk44 <dbl>, wk45 <dbl>, wk46 <dbl>, wk47 <dbl>, ...
> # Scenario: Combine all week columns into a single column called "Rank",
> # with a label column for "week"
> long_df <- df %>%
+   pivot_longer(
+     cols = starts_with("wk"),      # wk1, wk2, wk3, ...
+     names_to = "Week",            # New column containing the old header names
+     values_to = "Rank",          # New column containing the numbers
+     values_drop_na = TRUE        # Remove NA week-rank combinations
+   )
> print("--- 2. Long Format (pivot_longer) ---")

```

Session 2 (Bottom):

```

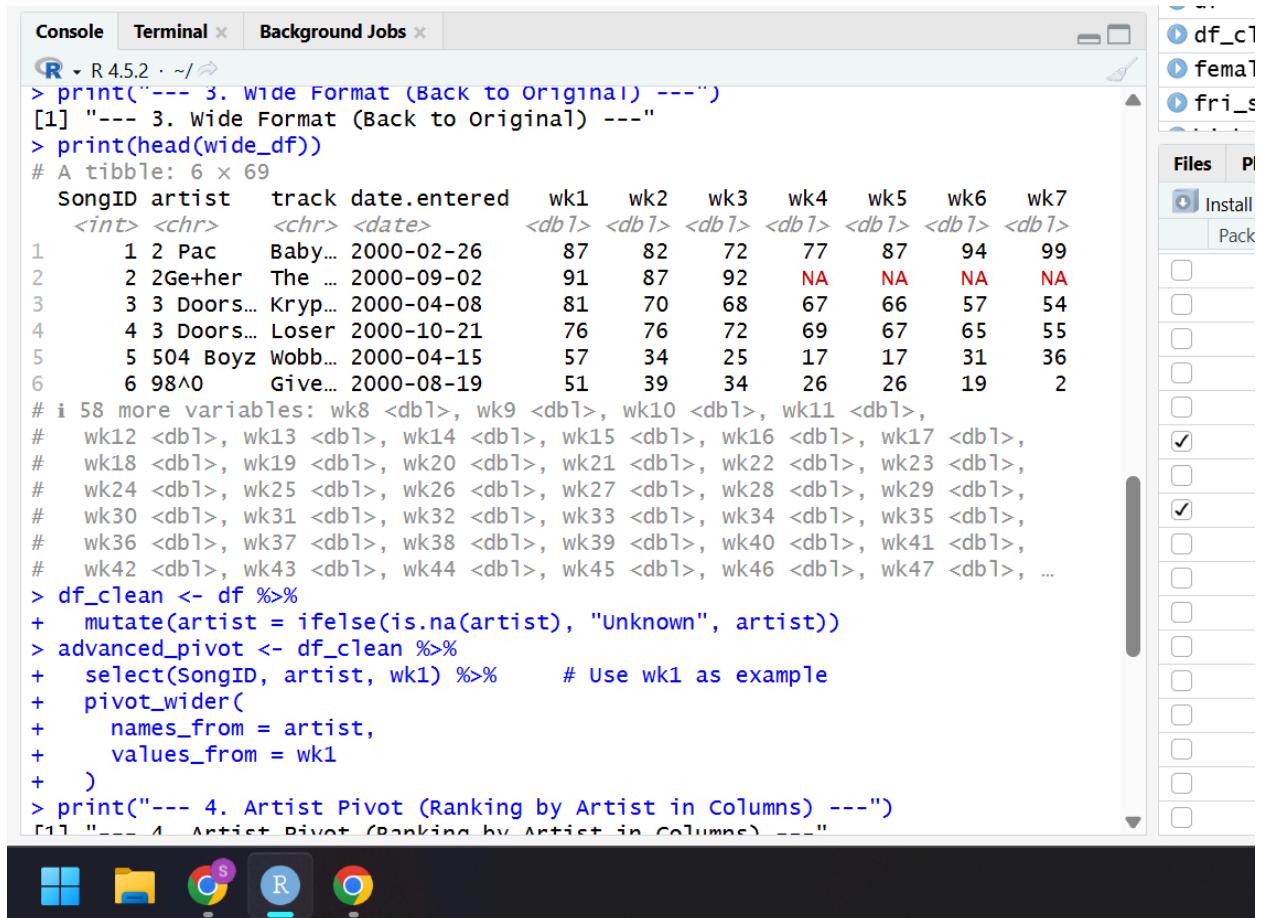
+   values_to = "Rank",           # New column containing the numbers
+   values_drop_na = TRUE         # Remove NA week-rank combinations
+ )
> print("--- 2. Long Format (pivot_longer) ---")
[1] "--- 2. Long Format (pivot_longer) ---"
> print(head(long_df, 10))
# A tibble: 10 × 6
SongID artist track                  date.entered week   Rank
<int> <chr> <chr> <date> <chr> <dbl>
1 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk1    87
2 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk2    82
3 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk3    72
4 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk4    77
5 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk5    87
6 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk6    94
7 1 2 Pac Baby Don't Cry (Keep... 2000-02-26 wk7    99
8 2 2Ge+her The Hardest Part Of ... 2000-09-02 wk1    91
9 2 2Ge+her The Hardest Part Of ... 2000-09-02 wk2    87
10 2 2Ge+her The Hardest Part Of ... 2000-09-02 wk3   92
> # Scenario: Spread the Week column back into columns again
> wide_df <- long_df %>%
+   pivot_wider(
+     names_from = week,       # Create columns wk1, wk2, wk3 ...
+     values_from = Rank
+   )
> print("--- 3. Wide Format (Back to Original) ---")
[1] "--- 3. Wide Format (Back to Original) ---"
> print(head(wide_df))
# A tibble: 6 × 60
SongID artist track date.entered wk1 wk2 wk3 wk4 wk5 wk6 wk7
<int> <chr> <chr> <date> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1 2 Pac Baby... 2000-02-26 87 82 72 77 87 94 99
2 2 2Ge+her The ... 2000-09-02 91 87 92 NA NA NA NA
3 3 3 Doors... Kryp... 2000-04-08 81 70 68 67 66 57 54
4 4 3 Doors... Loser 2000-10-21 76 76 72 69 67 65 55
5 5 504 Boyz Wobb... 2000-04-15 57 34 25 17 17 31 36
6 6 98% Give... 2000-08-19 51 39 34 26 26 19 2
# i 69 more variables: wk8 <dbl>, wk9 <dbl>, wk10 <dbl>, wk11 <dbl>,
# wk12 <dbl>, wk13 <dbl>, wk14 <dbl>, wk15 <dbl>, wk16 <dbl>, wk17 <dbl>,
# wk18 <dbl>, wk19 <dbl>, wk20 <dbl>, wk21 <dbl>, wk22 <dbl>, wk23 <dbl>,
# wk24 <dbl>, wk25 <dbl>, wk26 <dbl>, wk27 <dbl>, wk28 <dbl>, wk29 <dbl>,
# wk30 <dbl>, wk31 <dbl>, wk32 <dbl>, wk33 <dbl>, wk34 <dbl>, wk35 <dbl>,
# wk36 <dbl>, wk37 <dbl>, wk38 <dbl>, wk39 <dbl>, wk40 <dbl>, wk41 <dbl>,
# wk42 <dbl>, wk43 <dbl>, wk44 <dbl>, wk45 <dbl>, wk46 <dbl>, wk47 <dbl>, ...

```

The right sidebar of both sessions shows a package browser with the following packages listed:

- df_clean
- female_non_smoke...
- fri_sat_tips
- stringr
- sys
- systemfonts
- textshaping
- tibble
- tidy
- tidyselect
- tidyverse
- timechange
- tinytex
- tzdb
- utf8
- uuid
- vctrs
- viridisLite
- vroom
- withr

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R



The screenshot shows an RStudio interface with the following details:

- Console Tab:** Active tab.
- Terminal Tab:** Shows the R version: R 4.5.2 · ~/.
- Background Jobs Tab:** Shows a single job named "R 4.5.2 · ~/".
- Code Area:** Displays R code for data manipulation and visualization. It includes:
 - Printing a wide format dataset back to its original form.
 - Displaying the head of the dataset.
 - Describing the dataset as a tibble with 6 rows and 69 columns.
 - Handling missing values (NA) in the artist column.
 - Creating a clean dataset (df_clean) by mutating the artist column to "Unknown" if it's NA.
 - Performing an advanced pivot operation on df_clean.
 - Selecting specific columns (SongID, artist, wk1).
 - Pivoting the data to wider format, specifying names_from = artist and values_from = wk1.
 - Printing a message for step 4: "Artist Pivot (Ranking by Artist in columns)".
- File Explorer:** Shows files like df_clean.R, female.R, fri_s.R, and others.
- Toolbar:** Includes icons for Windows, File, Copy, Paste, and R.

MITHIL KADAM

SHETH L.U.J. & SIR M.V. COLLEGE OF SCIENCE
SUBJECT - Data Analysis with SAS / SPSS / R

```
+ pivot_wider<
+   names_from = artist,
+   values_from = wk1
+ )
> print(" --- 4. Artist Pivot (Ranking by Artist in Columns) ---")
[1] " --- 4. Artist Pivot (Ranking by Artist in Columns) ---"
> print(head(advanced_pivot))
# A tibble: 6 × 229
SongID `2 Pac` `2Ge+her` `3 Doors Down` `504 Boyz` `98^0` `A*Teens` Aaliyah
<int> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1     1     87     NA      NA     NA     NA     NA     NA
2     2     NA     91      NA     NA     NA     NA     NA
3     3     NA     NA      81     NA     NA     NA     NA
4     4     NA     NA      76     NA     NA     NA     NA
5     5     NA     NA      NA     57     NA     NA     NA
6     6     NA     NA      NA     NA     51     NA     NA
# i 221 more variables: `Adams, Yolanda` <dbl>, `Adkins, Trace` <dbl>,
# `Aguilera, Christina` <dbl>, `Alice Deejay` <dbl>, `Allan, Gary` <dbl>,
# `Amber` <dbl>, `Anastacia` <dbl>, `Anthony, Marc` <dbl>, `Avant` <dbl>,
# `BBMak` <dbl>, `Backstreet Boys, The` <dbl>, `Badu, Erkyah` <dbl>,
# `Baha Men` <dbl>, `Barenaked Ladies` <dbl>, `Beenie Man` <dbl>,
# `Before Dark` <dbl>, `Bega, Lou` <dbl>, `Big Punisher` <dbl>,
# `Black Rob` <dbl>, `Black, Clint` <dbl>, `Blaque` <dbl>, ...
# i Use `colnames()` to see all variable names
>
>
>
>
```



MITHIL KADAM