

PRACTICAL NO:5

AIM: Sorting data using arrange() in R.

RStudio Environment

```

> # Fixed & robust dplyr script for your uploaded csv
> library(dplyr)
> # file path (your uploaded csv)
> file_path <- "/mnt/data/StudentsPerformance - StudentsPerformance.csv"
> # Load data
> students <- read_csv("My web Sites/StudentsPerformance - StudentsPerformance.csv")
Rows: 1000 Columns: 8
-- Column specification ---

Delimited by ","
chr (5): gender, race/ethnicity, parental level of education, lunch, test pr...
dbl (3): math score, reading score, writing score

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.
> # Helper: normalized column names (lowercase, remove non-alnum)
> norm <- function(x) tolower(gsub("[^a-z0-9]", "", x))
> cols <- colnames(students)
> norm_cols <- setNames(norm(cols), cols) # named vector: original -> normalized
> # Helper to find best match from candidate names
> find_col <- function(candidates) {
+   cand_norm <- sapply(candidates, function(x) tolower(gsub("[^a-z0-9]", "", x)))
+   for (cn in cand_norm) {
+     match_idx <- which(norm_cols == cn)
+     if (length(match_idx) > 0) return(names(norm_cols)[match_idx[1]])
+   }
+   return(NA_character_)
+ }
> # Candidate lists for columns (add more variants if you like)
> math_candidates <- c("math.score", "math_score", "math score", "math", "maths", "maths.score")
> reading_candidates <- c("reading.score", "reading_score", "reading score", "reading", "readingscore")
> writing_candidates <- c("writing.score", "writing_score", "writing score", "writing", "writingscore")
> gender_candidates <- c("gender", "sex")

```

RStudio Files

- high_math_sorted 50 obs. of 8 variables
- high_price_subset 8 obs. of 13 variables
- Housing 545 obs. of 13 variables
- Housing 545 obs. of 13 variables
- insurance 1338 obs. of 7 variables
- large_highprice_s... 4 obs. of 13 variables
- mainroad_basement 168 obs. of 13 variables
- many_bathrooms 1 obs. of 13 variables

RStudio Status Bar

31°C Sunny 13:11 24-11-2025

RStudio Environment

```

> # Example 1: Sort by math score ascending
> students_sorted_math <- students %>
+   arrange(.data[[math.col]])
> cat("First 5 rows sorted by", math.col, "(ascending):\n")
First 5 rows sorted by math score (ascending):
> print(head(students_sorted_math, 5))
# A tibble: 5 x 8
  gender `race/ethnicity` parental level of education-1 lunch test preparation course-2
  <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 female group C some high school free~ none
2 female group B high school free~ none
3 female group B some high school free~ none
4 female group B some college stan~ none
5 female group C some college free~ none
# i abbreviated names: 1: 'parental level of education',
# i 2: 'test preparation course'
# i 3: more variables: `math score` <dbl>, `reading score` <dbl>,

```

RStudio Files

- high_math_sorted 50 obs. of 8 variables
- high_price_subset 8 obs. of 13 variables
- Housing 545 obs. of 13 variables
- Housing 545 obs. of 13 variables
- insurance 1338 obs. of 7 variables
- large_highprice_s... 4 obs. of 13 variables
- mainroad_basement 168 obs. of 13 variables
- many_bathrooms 1 obs. of 13 variables

RStudio Status Bar

31°C Sunny 13:12 24-11-2025

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

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
> # Example 2: Sort by reading score descending
> students_sorted_reading_desc <- students %>
+   arrange(desc(.data[[reading_col]]))
> cat("nFirst 5 rows sorted by", reading_col, "(descending):\n")
First 5 rows sorted by reading score (descending):
> print(head(students_sorted_reading_desc, 5))
# A tibble: 5 x 8
  gender `race/ethnicity` parental level of educa-1 lunch test preparation cou-2
  <chr>  <chr>          <chr>    <chr> <chr> <chr> <chr> <chr>
1 female group D master's degree stan~ none
2 female group E bachelor's degree stan~ completed
3 male group E associate's degree free~ completed
4 female group C bachelor's degree stan~ completed
5 female group E some high school stan~ completed
# i abbreviated names: 1: 'parental level of education',
# i 2: 'test preparation course'
# i 3 more variables: 'math score' <dbl>, 'reading score' <dbl>,
# i 'writing score' <dbl>
> # Example 3: Sort by gender (if found) then by writing score descending
> if (!is.na(gender_col)) {
+   students_sorted_gender_writing <- students %>
+     arrange(.data[[gender_col]], desc(.data[[writing_col]]))
+   cat("nFirst 10 rows sorted by", gender_col, "and then", writing_col, "(desc):\n")
+   print(head(students_sorted_gender_writing, 10))
+ } else {
+   cat("\ngender column not detected; skipping gender+writing sort example.\n")
+ }
```

First 10 rows sorted by gender and then writing score (desc):
A tibble: 10 x 8
 gender `race/ethnicity` parental level of educa-1 lunch test preparation cou-2
 <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
1 female group D master's degree stan~ none
2 female group E bachelor's degree stan~ completed

Environment History Connections Tutorial

Import Dataset 257 MB ✓

Global Environment

- high_math_sorted 50 obs. of 8 variables
- high_price_subset 8 obs. of 13 variables
- Housing 545 obs. of 13 variables
- Housing 545 obs. of 13 variables
- insurance 1338 obs. of 7 variables
- large_highprice_s... 4 obs. of 13 variables
- mainroad_basement 168 obs. of 13 variables
- many_bathrooms 1 obs. of 13 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home My Web Sites

Name Size Modified

- Housing.csv 29.3 KB Nov 18, 2025, 11:51 AM
- insurance.csv 54.3 KB Nov 24, 2025, 11:39 AM
- StudentsPerformance - StudentsPerformance.csv 55.7 KB Nov 18, 2025, 2:37 PM
- WebSite1

31°C Sunny 13:12 24-11-2025

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
> # Example 4: Filter (high math > 90) then arrange by writing score ascending
> # (adjust threshold 90 if your scoring range differs)
> high_math_sorted <- students %>
+   filter(.data[[math_col]] > 90) %>%
+   arrange(.data[[writing_col]])
> cat("nTop 5 high math scorers (> 90 ) with lowest writing scores:\n")
Top 5 high math scorers (> 90 ) with lowest writing scores:
> print(head(high_math_sorted %>% select(all_of(c(gender_col, math_col, writing_col)))), 5))
# A tibble: 5 x 3
  gender `math score` `writing score`
  <chr>      <dbl>        <dbl>
1 male         94          71
2 male         91          76
3 male         94          78
4 male         92          78
5 male         91          79
```

Environment History Connections Tutorial

Import Dataset 257 MB ✓

Global Environment

- high_math_sorted 50 obs. of 8 variables
- high_price_subset 8 obs. of 13 variables
- Housing 545 obs. of 13 variables
- Housing 545 obs. of 13 variables
- insurance 1338 obs. of 7 variables
- large_highprice_s... 4 obs. of 13 variables
- mainroad_basement 168 obs. of 13 variables
- many_bathrooms 1 obs. of 13 variables

Files Plots Packages Help Viewer Presentation

New Folder New File Delete Rename More

Home My Web Sites

Name Size Modified

- Housing.csv 29.3 KB Nov 18, 2025, 11:51 AM
- insurance.csv 54.3 KB Nov 24, 2025, 11:39 AM
- StudentsPerformance - StudentsPerformance.csv 55.7 KB Nov 18, 2025, 2:37 PM
- WebSite1

31°C Sunny 13:12 24-11-2025