

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

Practical No: 5

Aim: Sorting data using arrange() in R.

Code:

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Go to file/function Addins
Source
1 install.packages("dplyr")
2
3 library(dplyr)
4 library(readr)
5
6 house_data <- read_csv("House Price Prediction.csv")
7
8 head(house_data)
9
10 #Example 1: Sorting by a Single Variable
11
12 house_sorted_price <- house_data |>
13   arrange(price)
14
15 head(house_sorted_price, 5)
16
17 #Example 2: Sorting by a Single Variable
18
19 house_sorted_sqft_desc <- house_data |>
20   arrange(desc(sqft_living))
21
22 head(house_sorted_sqft_desc, 5)
23
24 #Example 3: Sorting by Two Variables
25
26 house_multi_sort <- house_data |>
27   arrange(view, desc(price))
28
29 head(house_multi_sort, 10)
30
31 #Example 4: Combined Filter and Sort
32
33 large_homes_by_year <- house_data |>
34   filter(bedrooms > 4) |>
35   arrange(yr_built)
36
37 cat("Top 5 large homes (5+ bedrooms) by oldest build year:\n")
38 print(large_homes_by_year |> select(bedrooms, yr_built, price) |> head(5))
39
231 | (Top Level) |
Console
Files Plots Packages Help Viewer Presentation
Environment History Connections Tutorial
R - Global Environment
Data
brazil_or_caffe_ 1853 obs. of 16 variables
coffee_data 10000 obs. of 16 variables
Daily_Coffee_In_ 10000 obs. of 16 variables
Daily_Coffee_In_ 10000 obs. of 1 variable
Daily.Coffee.In_ 10000 obs. of 16 variables
eu_subset_filter 983 obs. of 16 variables
female_low_caffe_ 859 obs. of 16 variables
high_sleep_subs_ 1268 obs. of 16 variables
house_data 4600 obs. of 13 variables
house_multi_sort 4600 obs. of 13 variables
house_sorted_pr_ 4600 obs. of 13 variables
house_sorted_sq_ 4600 obs. of 13 variables
House.Price.Pre_ 4600 obs. of 13 variables
large_homes_by_ 431 obs. of 13 variables
relaxed_subset 1268 obs. of 16 variables
young_people_fi_ 2037 obs. of 16 variables
```

Output:

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Go to file/function Addins
Source
Console Terminal Background Jobs
R - R 4.5.2 - /-
> house_data <- read_csv("House Price Prediction.csv")
Rows: 4600 Columns: 13
Column specification
Delimiter: ","
chr (2): street, city
dbl (11): price, bedrooms, sqft_living, sqft_lot, floors, view, condition, sqft_above, sqft_basement, yr_built, ...
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.
> head(house_data)
# A tibble: 6 x 13
  price bedrooms sqft_living sqft_lot floors view condition sqft_above sqft_basement yr_built yr_renovated street
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 313000 3 1340 2912 1.5 0 3 1340 0 1955 2005 18810.
2 2384000 5 3650 9050 2 4 5 3370 280 1921 0 709 W.
3 342000 3 1930 11947 1 0 4 1930 0 1966 0 26206.
4 420000 3 2000 8030 1 0 4 1000 1000 1963 0 857 I.
5 550000 4 1940 10900 1 0 4 1140 800 1976 1992 9105.
6 490000 2 880 6380 1 0 3 880 0 1938 1994 522 N.
# 11 more variable: city <chr>
> house_sorted_price <- house_data |>
+   arrange(price)
# A tibble: 5 x 13
  price bedrooms sqft_living sqft_lot floors view condition sqft_above sqft_basement yr_built yr_renovated street
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 0 3 1490 10125 1 0 4 1490 0 1962 0 3911 S.
2 0 4 2600 5390 1 0 4 1300 1300 1960 2001 2120 31.
3 0 6 3200 9200 1 2 4 1600 1953 1983 12271 W.
4 0 5 3480 36615 2 0 4 2490 990 1983 0 21809 S.
5 0 5 1500 7112 1 0 5 760 740 1920 0 14901-L.
# 11 more variable: city <chr>
> house_sorted_sqft_desc <- house_data |>
+   arrange(desc(sqft_living))
> head(house_sorted_sqft_desc, 5)
# A tibble: 5 x 13
  price bedrooms sqft_living sqft_lot floors view condition sqft_above sqft_basement yr_built yr_renovated street
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 2280000 7 13540 307752 3 4 3 9410 4130 1999 0 26408.
2 7062500 5 10040 32325 2 2 3 7680 2360 1940 2001 4442.
3 4668000 5 6640 13068 1 4 3 4870 4870 1083 7000 S.M.A.
Files Plots Packages Help Viewer Presentation
Environment History Connections Tutorial
R - Global Environment
Data
brazil_or_caffe_ 1853 obs. of 16 variables
coffee_data 10000 obs. of 16 variables
Daily_Coffee_In_ 10000 obs. of 16 variables
Daily_Coffee_In_ 10000 obs. of 1 variable
Daily.Coffee.In_ 10000 obs. of 16 variables
eu_subset_filter 983 obs. of 16 variables
female_low_caffe_ 859 obs. of 16 variables
high_sleep_subs_ 1268 obs. of 16 variables
house_data 4600 obs. of 13 variables
house_multi_sort 4600 obs. of 13 variables
house_sorted_pr_ 4600 obs. of 13 variables
house_sorted_sq_ 4600 obs. of 13 variables
House.Price.Pre_ 4600 obs. of 13 variables
large_homes_by_ 431 obs. of 13 variables
relaxed_subset 1268 obs. of 16 variables
young_people_fi_ 2037 obs. of 16 variables
```

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

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation, including variable selection, sorting, and filtering.
- Console:** Shows the output of the R commands, including data frames and tibbles.
- Environment:** Lists the objects in the global environment, such as `brazil_or_caffe`, `coffee_data`, `house_data`, and `house_multi_sort`.
- Files:** Shows the file explorer with the current project files.

```
R - R4.5.2 - ~/RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins

Source
Console Terminal Background Jobs

# 1 more variable: city <chr>
> house_multi_sort <- house_data |>
+ arrange(view, desc(price))
> head(house_multi_sort, 10)
# A tibble: 10 x 13
  price bedrooms sqft_living sqft_lot floors view condition sqft_above sqft_basement yr_built yr_renovated street
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <chr>
1 2280000 7 13540 302752 3 4 3 9410 4130 1999 0 26408...
2 7062500 5 10040 32325 2 2 3 2680 2360 1940 2001 4442...
3 4668000 5 9640 13068 1 4 3 4820 4820 1983 2009 5044...
4 2888000 5 8670 64033 2 4 3 6120 2550 1965 2003 1291...
5 0 5 8020 21738 2 0 3 8020 0 2001 0 2 cre...

# 1 more variable: city <chr>
> large_homes_by_year <- house_data |>
+ filter(bedrooms > 4) |>
+ arrange(yr_built)
> cat("Top 5 large homes (5+ bedrooms) by oldest build year:\n")
Top 5 large homes (5+ bedrooms) by oldest build year:
> print(large_homes_by_year |> select(bedrooms, yr_built, price) |> head(5))
# A tibble: 5 x 3
  bedrooms yr_built price
<dbl> <dbl> <dbl>
1 5 1900 550000
2 6 1900 660000
3 6 1900 1297000
4 7 1901 599000
5 5 1902 650000
```