

PRACTICAL NO 5

AIM : Sorting data using arrange() in R.

```

R - R4.1.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Console Terminal Background Jobs
> # Step 0: Load the dataset
> housing <- read_csv("American_Housing_Data_20231209.csv")
rows: 39981 columns: 14
-- Column specification -----
Delimiter: ","
chr (4): Address, City, State, County
dbl (10): Zip Code, Price, Beds, Baths, Living Space, Zip Code Population, Zip C...

# Use 'spec()' to retrieve the full column specification for this data.
# Specify the column types or set 'show_col_types = FALSE' to quiet this message.
> # Inspect the first few rows
> head(housing)
# A tibble: 6 x 14
  Zip Code Price Beds Baths Living Space Address City State
  <dbl> <dbl> <dbl> <dbl> <dbl> <chr> <chr> <chr>
1 10013 3990000 2 3 1967 74 GRAND ST APT 3 New York New ~
2 10013 3990000 2 3 1967 74 GRAND ST APT 3 New York New ~
3 10014 1650000 1 1 718 140 CHARLES ST APT 4D New York New ~
4 10014 760000 3 2 1538 38 JONES ST New York New ~
5 10014 1100000 1 1 600 81 BEDFORD ST APT 3F New York New ~
6 10017 764900 1 1 643 145 E 48TH ST APT 11E New York New ~

# 6 more variables: 'Zip Code Population' <dbl>, 'Zip Code density' <dbl>,
# 'County' <chr>, 'Median household income' <dbl>, 'Latitude' <dbl>, 'Longitude' <dbl>
> # Example 1: Sort by Price (ascending)
> #-----
> housing_sorted_price <- housing[order(housing$Price), ]
> cat("First 5 cheapest houses:\n")
First 5 cheapest houses:
> head(housing_sorted_price[, c("Zip Code", "Price", "Beds", "Baths", "Living Space")],
5)
# A tibble: 5 x 5
  Zip Code Price Beds Baths Living Space
  <dbl> <dbl> <dbl> <dbl> <dbl>
1 10308 1800 1 1 700
2 33150 1950 2 1 1365
3 93650 1995 3 2 1332
4 70123 2000 3 2 1467
5 10309 2200 2 1 1100

```

```

R - R4.1.2 - ~/
File Edit Code View Plots Session Build Debug Profile Tools Help
Source Console Terminal Background Jobs
> head(housing_sorted_price[, c("Zip Code", "Price", "Beds", "Baths", "Living Space")],
5)
# A tibble: 5 x 5
  Zip Code Price Beds Baths Living Space
  <dbl> <dbl> <dbl> <dbl> <dbl>
1 10308 1800 1 1 700
2 33150 1950 2 1 1365
3 93650 1995 3 2 1332
4 70123 2000 3 2 1467
5 10309 2200 2 1 1100
> # Example 2: Sort by Price (descending)
> #-----
> housing_sorted_price_desc <- housing[order(-housing$Price), ]
> cat("\nTop 5 most expensive houses:\n")
Top 5 most expensive houses:
> head(housing_sorted_price_desc[, c("Zip Code", "Price", "Beds", "Baths", "Living Space")],
5)
# A tibble: 5 x 5
  Zip Code Price Beds Baths Living Space
  <dbl> <dbl> <dbl> <dbl> <dbl>
1 90027 38000000 9 11 20000
2 94115 32000000 10 7 11155
3 20007 31900000 5 9 16250
4 33133 31500000 7 9 14720
5 90069 28500000 4 6 9052
> # Example 3: Multi-column sort
> # Sort first by Beds (descending), then Price (descending)
> #-----
> housing_multi_sort <- housing[order(-housing$Beds, -housing$Price), ]
> cat("\nTop 10 houses by Beds and Price:\n")

```

The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains R code for data manipulation. The code includes:


```

      Top 10 houses by Beds and Price:
      > head(housing_multi_sort[, c("Zip Code", "Beds", "Price", "Living Space")], 10)
      # A tibble: 10 x 4
      Zip code Beds Price Living Space
      <dbl> <dbl> <dbl> <dbl>
    1 93728 54 1899998 33308
    2 93726 46 3500000 19768
    3 93726 44 5875000 28728
    4 90802 37 11495000 35429
    5 90046 36 7300000 22251
    6 90003 36 4995000 17947
    7 75214 34 3277025 9536
    8 38104 32 4950000 3400
    9 90006 30 5690000 40000
    10 90806 30 4750000 14814

      > # Example 4: Filter and sort
      > # Filter for houses with more than 3 beds, then sort by Living Space (ascending)
      > #-----
      > large_houses <- subset(housing, Beds > 3)
      > large_houses_sorted <- large_houses[order(large_houses$Living.Space), ]

      Error in order(large_houses$Living.Space) : argument 1 is not a vector

      In addition: Warning message:
      Unknown or uninitialised column: 'Living.Space'.

      > cat("\nTop 5 large houses (Beds > 3) by smallest Living Space:\n")

      Top 5 large houses (Beds > 3) by smallest Living Space:
      > head(large_houses_sorted[, c("Zip Code", "Beds", "Living Space", "Price")], 5)

      Error in head(large_houses_sorted[, c("Zip Code", "Beds", "Living Space", :
      
```
- Environment:** Lists loaded objects:
 - Global Environment: 1000 obs. of 10 variables
 - high_income_subset: 281 obs. of 35 variables
 - housing: 39981 obs. of 14 variables
 - housing_multi_sort: 39981 obs. of 14 variables
 - housing_sorted_pr...: 39981 obs. of 14 variables
 - housing_sorted_pr...: 39981 obs. of 14 variables
 - hr: 1470 obs. of 35 variables
 - large_houses: 13190 obs. of 14 variables
- Files:** Shows a file explorer view of the Home directory with files like .Rhistory, American_Housing_Data_20231209.csv, and Bills.pdf.
- Terminal:** Displays the output of the R commands, including the error and warning messages.