

R Within-group Ascending and Descending Sort, Selection, and Differencing

Fan Wang

2022-07-15

Contents

1	Group, Sort and Slice	1
1.1	Sort in Ascending and Descending Orders	1
1.2	Get Highest Values from Groups	1
1.3	Differences in Within-group Sorted Value	2

1 Group, Sort and Slice

Go to the [RMD](#), [R](#), [PDF](#), or [HTML](#) version of this file. Go back to [fan's REconTools](#) research support package, [R4Econ](#) examples page, [PkgTestR](#) packaging guide, or [Stat4Econ](#) course page.

1.1 Sort in Ascending and Descending Orders

We sort the `mtcars` dataset, sorting in *ascending* order by `cyl`, and in *descending* order by `mpg`. Using `arrange`, `desc(dis)` means sorting the `disp` variable in descending order. In the table shown below, `cyl` is increasing, and `disp` id decreasing within each `cyl` group.

```
kable(mtcars %>%
  arrange(cyl, desc(dis)) %>%
  # Select and filter to reduce display clutter
  select(cyl, disp, mpg)) %>%
kable_styling_fc()
```

1.2 Get Highest Values from Groups

There is a dataframe with a grouping variable with N unique values, for example N classes. Find the top three highest scoring students from each class. In the example below, group by `cyl` and get the cars with the highest and second highest `mpg` cars in each `cyl` group.

```
# use mtcars: slice_head gets the lowest sorted value
df_groupby_top_mpg <- mtcars %>%
  rownames_to_column(var = "car") %>%
  arrange(cyl, desc(mpg)) %>%
  group_by(cyl) %>%
  slice_head(n=3) %>%
  select(car, cyl, mpg, disp, hp)

# display
kable(df_groupby_top_mpg) %>% kable_styling_fc()
```

	cyl	disp	mpg
Merc 240D	4	146.7	24.4
Merc 230	4	140.8	22.8
Volvo 142E	4	121.0	21.4
Porsche 914-2	4	120.3	26.0
Toyota Corona	4	120.1	21.5
Datsun 710	4	108.0	22.8
Lotus Europa	4	95.1	30.4
Fiat X1-9	4	79.0	27.3
Fiat 128	4	78.7	32.4
Honda Civic	4	75.7	30.4
Toyota Corolla	4	71.1	33.9
Hornet 4 Drive	6	258.0	21.4
Valiant	6	225.0	18.1
Merc 280	6	167.6	19.2
Merc 280C	6	167.6	17.8
Mazda RX4	6	160.0	21.0
Mazda RX4 Wag	6	160.0	21.0
Ferrari Dino	6	145.0	19.7
Cadillac Fleetwood	8	472.0	10.4
Lincoln Continental	8	460.0	10.4
Chrysler Imperial	8	440.0	14.7
Pontiac Firebird	8	400.0	19.2
Hornet Sportabout	8	360.0	18.7
Duster 360	8	360.0	14.3
Ford Pantera L	8	351.0	15.8
Camaro Z28	8	350.0	13.3
Dodge Challenger	8	318.0	15.5
AMC Javelin	8	304.0	15.2
Maserati Bora	8	301.0	15.0
Merc 450SE	8	275.8	16.4
Merc 450SL	8	275.8	17.3
Merc 450SLC	8	275.8	15.2

car	cyl	mpg	disp	hp
Toyota Corolla	4	33.9	71.1	65
Fiat 128	4	32.4	78.7	66
Honda Civic	4	30.4	75.7	52
Hornet 4 Drive	6	21.4	258.0	110
Mazda RX4	6	21.0	160.0	110
Mazda RX4 Wag	6	21.0	160.0	110
Pontiac Firebird	8	19.2	400.0	175
Hornet Sportabout	8	18.7	360.0	175
Merc 450SL	8	17.3	275.8	180

1.3 Differences in Within-group Sorted Value

We first take the largest N values in M groups, then we difference between the ranked top values in each group.

We have N classes, and M students in each class. We first select the 3 students with the highest scores from each class, then we take the difference between 1st and 2nd, and the difference between the 2nd and the 3rd students.

Note that when are using descending sort, so *lead* means the next value in descending sequencing, and *lag* means the last value which was higher in descending order.

```
# We use what we just created in the last block.
df_groupby_top_mpg_diff <- df_groupby_top_mpg %>%
  group_by(cyl) %>%
  mutate(mpg_diff_higher_minus_lower = mpg - lead(mpg)) %>%
  mutate(mpg_diff_lower_minus_higher = mpg - lag(mpg))

# display
kable(df_groupby_top_mpg_diff) %>% kable_styling_fc()
```

car	cyl	mpg	disp	hp	mpg_diff_higher_minus_lower	mpg_diff_lower_minus_higher
Toyota Corolla	4	33.9	71.1	65	1.5	NA
Fiat 128	4	32.4	78.7	66	2.0	-1.5
Honda Civic	4	30.4	75.7	52	NA	-2.0
Hornet 4 Drive	6	21.4	258.0	110	0.4	NA
Mazda RX4	6	21.0	160.0	110	0.0	-0.4
Mazda RX4 Wag	6	21.0	160.0	110	NA	0.0
Pontiac Firebird	8	19.2	400.0	175	0.5	NA
Hornet Sportabout	8	18.7	360.0	175	1.4	-0.5
Merc 450SL	8	17.3	275.8	180	NA	-1.4