20250513_01

May 13, 2025

[3]:
```
library(tidyverse)
```

```
  Attaching core tidyverse packages                tidyverse
2.0.0
  dplyr     1.1.4       readr     2.1.5
  forcats   1.0.0       stringr   1.5.1
  ggplot2   3.5.2       tibble    3.2.1
  lubridate 1.9.4       tidyr     1.3.1
  purrr     1.0.4
  Conflicts
tidyverse_conflicts()
  dplyr::filter() masks stats::filter()
  dplyr::lag()    masks stats::lag()
  Use the conflicted package
(<http://conflicted.r-lib.org/>) to force all conflicts to
become errors
```

[45]:
```
students = tibble(id = 1:5,
                  name = c("Ann", "Ben", "Cathy", "Dan", "Eva"),
                  math = c(78, 92, 85, 66, 88),
                  english = c(70, 85, 90, 75, 80),
                  passed = c(TRUE, TRUE, TRUE, FALSE, TRUE))

students
```

A tibble: 5 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
1	Ann	78	70	TRUE
2	Ben	92	85	TRUE
3	Cathy	85	90	TRUE
4	Dan	66	75	FALSE
5	Eva	88	80	TRUE

[53]:
```
students %>% select(name, math)
```

A tibble: 5 × 2

name	math
<chr>	<dbl>
Ann	78
Ben	92
Cathy	85
Dan	66
Eva	88

[55]:
```
students %>% select(-id)
```

A tibble: 5 × 4

name	math	english	passed
<chr>	<dbl>	<dbl>	<lgl>
Ann	78	70	TRUE
Ben	92	85	TRUE
Cathy	85	90	TRUE
Dan	66	75	FALSE
Eva	88	80	TRUE

[57]:
```
students %>% filter(math > 80)
```

A tibble: 3 × 5

id	name	math	english	passed
<int>	<chr>	<dbl>	<dbl>	<lgl>
2	Ben	92	85	TRUE
3	Cathy	85	90	TRUE
5	Eva	88	80	TRUE

[59]:
```
students %>% filter(math > 80 & english < 90)
```

A tibble: 2 × 5

id	name	math	english	passed
<int>	<chr>	<dbl>	<dbl>	<lgl>
2	Ben	92	85	TRUE
5	Eva	88	80	TRUE

[61]:
```
students %>% filter(math > 80 | passed == FALSE)
```

A tibble: 4 × 5

id	name	math	english	passed
<int>	<chr>	<dbl>	<dbl>	<lgl>
2	Ben	92	85	TRUE
3	Cathy	85	90	TRUE
4	Dan	66	75	FALSE
5	Eva	88	80	TRUE

[63]:
```
students %>% filter(math > 80) %>% select(name, math)
```

A tibble: 3 × 2

name	math
<chr>	<dbl>
Ben	92
Cathy	85
Eva	88

[65]:
```
# Exercise
# 1. Show students with english > 80
# 2. Show students with math + english > 170
# 3. Only show 'name' and 'passed'

students %>% filter(english > 80)

students %>% filter(english + math > 170)

students %>% select(name, passed)
```

A tibble: 2 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
2	Ben	92	85	TRUE
3	Cathy	85	90	TRUE

A tibble: 2 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
2	Ben	92	85	TRUE
3	Cathy	85	90	TRUE

A tibble: 5 × 2

name <chr>	passed <lgl>
Ann	TRUE
Ben	TRUE
Cathy	TRUE
Dan	FALSE
Eva	TRUE

[67]:
```
students %>% mutate(avg_score = (math + english)/2)
```

A tibble: 5 × 6

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	avg_score <dbl>
1	Ann	78	70	TRUE	74.0
2	Ben	92	85	TRUE	88.5
3	Cathy	85	90	TRUE	87.5
4	Dan	66	75	FALSE	70.5
5	Eva	88	80	TRUE	84.0

[69]:
```
students %>% mutate(status = ifelse(passed == TRUE, 'Passed', 'Failed'))
```

A tibble: 5 × 6

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	status <chr>
1	Ann	78	70	TRUE	Passed
2	Ben	92	85	TRUE	Passed
3	Cathy	85	90	TRUE	Passed
4	Dan	66	75	FALSE	Failed
5	Eva	88	80	TRUE	Passed

[71]:
```
students %>% mutate(avg_score = (math + english) / 2,
                    grade = ifelse(avg_score >= 90, 'A',
                                   ifelse(avg_score >= 80, 'B', 'C')))
```

A tibble: 5 × 7

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	avg_score <dbl>	grade <chr>
1	Ann	78	70	TRUE	74.0	C
2	Ben	92	85	TRUE	88.5	B
3	Cathy	85	90	TRUE	87.5	B
4	Dan	66	75	FALSE	70.5	C
5	Eva	88	80	TRUE	84.0	B

[77]:
```
students %>% arrange(math)

students %>% arrange(desc(english))

students %>% arrange(desc(passed), desc(english))
```

A tibble: 5 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
4	Dan	66	75	FALSE
1	Ann	78	70	TRUE
3	Cathy	85	90	TRUE
5	Eva	88	80	TRUE
2	Ben	92	85	TRUE

A tibble: 5 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
3	Cathy	85	90	TRUE
2	Ben	92	85	TRUE
5	Eva	88	80	TRUE
4	Dan	66	75	FALSE
1	Ann	78	70	TRUE

A tibble: 5 × 5

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>
3	Cathy	85	90	TRUE
2	Ben	92	85	TRUE
5	Eva	88	80	TRUE
1	Ann	78	70	TRUE
4	Dan	66	75	FALSE

[89]:
```
# Exercise
# 1. Arrnage based on avg_score from highest score to least
# 2. Arrange based passed then avg_score
# 3. Arrange based on grade.

students %>% mutate(avg_score = (math + english)/2) %>% arrange(desc(avg_score))
```

```
students %>% mutate(avg_score = (math + english)/2) %>% arrange(desc(passed),␣
↪desc(avg_score))

students %>% mutate(avg_score = (math + english)/2,
                    grade = ifelse(avg_score > 85, 'A',
                                   ifelse(avg_score > 80, 'B', 'C'))) %>%␣
↪arrange(grade)
```

A tibble: 5 × 6

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	avg_score <dbl>
2	Ben	92	85	TRUE	88.5
3	Cathy	85	90	TRUE	87.5
5	Eva	88	80	TRUE	84.0
1	Ann	78	70	TRUE	74.0
4	Dan	66	75	FALSE	70.5

A tibble: 5 × 6

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	avg_score <dbl>
2	Ben	92	85	TRUE	88.5
3	Cathy	85	90	TRUE	87.5
5	Eva	88	80	TRUE	84.0
1	Ann	78	70	TRUE	74.0
4	Dan	66	75	FALSE	70.5

A tibble: 5 × 7

id <int>	name <chr>	math <dbl>	english <dbl>	passed <lgl>	avg_score <dbl>	grade <chr>
2	Ben	92	85	TRUE	88.5	A
3	Cathy	85	90	TRUE	87.5	A
5	Eva	88	80	TRUE	84.0	B
1	Ann	78	70	TRUE	74.0	C
4	Dan	66	75	FALSE	70.5	C