R Notebook

Code

Exercise 11.1

Hide

```
library(tidyverse)
```

```
── Attaching core tidyverse packages ──────────────────────
✓ dplyr     1.1.3     ✓ readr     2.1.4
✓ forcats   1.0.0     ✓ stringr   1.5.0
✓ ggplot2   3.4.4     ✓ tibble    3.2.1
✓ lubridate 1.9.3     ✓ tidyr     1.3.0
✓ purrr     1.0.2
── Conflicts ────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to b
ecome errors
```

Hide

```
install.packages("dplyr")
```

```
Error in install.packages : Updating loaded packages
```

Hide

```
library(dplyr)
```

Exercise 11.2

Hide

```
scores <-
  tibble(
    name = c("mike", "carol", "greg", "marcia", "peter", "jan", "bobby", "cindy", "alice"),
    school = c("south", "south", "south", "south", "north", "north", "north", "south", "south"),
    teacher = c("johnson", "johnson", "johnson", "johnson",  "smith", "smith", "smith", "perry",
"perry"),
    sex = c("male", "female", "male", "female", "male", "female", "male", "female", "female"),
    math_score = c(4, 3, 2, 4, 3, 4, 5, 4, 5),
    reading_score = c(1, 5, 2, 4, 5, 4, 1, 5, 4)
)
```

Exercise 11.3

Hide

```
print(scores)
```

Exercise 11.4

Hide

```
scores_small <- scores |>
  slice(1:3)
```

Hide

```
scores_small
```

Exercise 11.5

Hide

```
scores |>
  arrange(desc(math_score))
```

Exercise 11.6

Hide

```
scores |>
  arrange(name)
```

Exercise 11.7

Hide

```
scores |>
  arrange(sex)
```

Exercise 11.8

Hide

```
scores |>
  arrange(school, teacher, sex, math_score, reading_score)
```

Exercise 11.9

Hide

```
scores |>
  select(name, math_score, reading_score)
```

Exercise 11.10

Hide

```
scores |>
  select(-sex)
```

Exercise 11.11

Hide

```
scores |>
  select(-sex, -reading_score)
```

Exercise 11.12

Hide

```
scores |>
  select(sex, everything()) ## everything just helps <I forget what it does>
```

Exercise 11.13

Hide

```
scores |>
  filter(sex == "male", school == "south")
```

Exercise 11.14

Hide

```
scores |>
  filter(math_score > mean(math_score))
```

Exercise 11.15

Hide

```
scores |>
  filter(math_score >= 4 & reading_score >= 3)
```

Exercise 11.16

Hide

```
scores |>
  filter(math_score <= 3 | reading_score <= 3)
```

Exercise 11.17

Hide

```
scores |>
  filter(reading_score > 1 & reading_score < 5)

## or
scores |>
  filter(reading_score %in% 2:4)
```

Exercise 11.18

Hide

```
scores |>
  filter(substr(name, "m"))
```

```
Error in `filter()`:
ℹ In argument: `substr(name, "m")`.
Caused by error in `substr()`:
! argument "stop" is missing, with no default
Backtrace:
 1. dplyr::filter(scores, substr(name, "m"))
 8. base::substr(name, "m")
```

Exercise 11.19

Hide

```
scores |>
  group_by(teacher) |>
  filter(max(math_score) == 5)
```

name	school	teacher	sex	math_score	reading_score
<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
peter	north	smith	male	3	5
jan	north	smith	female	4	4
bobby	north	smith	male	5	1
cindy	south	perry	female	4	5
alice	south	perry	female	5	4

5 rows

Exercise 11.20

Hide

```
scores |>
  group_by(sex) |>
  filter(mean(math_score) == 4)
```

name	school	teacher	sex	math_score	reading_score
<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
carol	south	johnson	female	3	5
marcia	south	johnson	female	4	4
jan	north	smith	female	4	4
cindy	south	perry	female	4	5
alice	south	perry	female	5	4

5 rows

Exercise 11.21

Hide

```
scores |>
  mutate(
    math_score = math_score * 10,
    reading_score = reading_score * 10
  )
```

name	school	teacher	sex	math_score	reading_score
<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>
mike	south	johnson	male	40	10
carol	south	johnson	female	30	50

name <chr>	school <chr>	teacher <chr>	sex <chr>	math_score <dbl>	reading_score <dbl>
greg	south	johnson	male	20	20
marcia	south	johnson	female	40	40
peter	north	smith	male	30	50
jan	north	smith	female	40	40
bobby	north	smith	male	50	10
cindy	south	perry	female	40	50
alice	south	perry	female	50	40

9 rows

Exercise 11.22

Hide

```
scores |>
  mutate(math_reading_avg = mean(reading_score + math_score))
```

name <chr>	school <chr>	teacher <chr>	sex <chr>	math_score <dbl>	reading_score <dbl>	math_reading_avg <dbl>
mike	south	johnson	male	4	1	7.222222
carol	south	johnson	female	3	5	7.222222
greg	south	johnson	male	2	2	7.222222
marcia	south	johnson	female	4	4	7.222222
peter	north	smith	male	3	5	7.222222
jan	north	smith	female	4	4	7.222222
bobby	north	smith	male	5	1	7.222222
cindy	south	perry	female	4	5	7.222222
alice	south	perry	female	5	4	7.222222

9 rows

Exercise 11.26

Hide

```
scores |>
  group_by(sex) |>
  mutate(math_score_centered_by_sex = math_score -  mean(math_score)) |>
  arrange(desc(math_score_centered_by_sex))
```

name	school	teacher	sex	math_score	reading_score	math_score_centered_by_sex
<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>
bobby	north	smith	male	5	1	1.5
alice	south	perry	female	5	4	1.0
mike	south	johnson	male	4	1	0.5
marcia	south	johnson	female	4	4	0.0
jan	north	smith	female	4	4	0.0
cindy	south	perry	female	4	5	0.0
peter	north	smith	male	3	5	-0.5
carol	south	johnson	female	3	5	-1.0
greg	south	johnson	male	2	2	-1.5

9 rows

Exercise 11.27

Hide

```
scores |>
  group_by(teacher) |>
  mutate(reading_score_centered_by_teacher = math_score - mean(math_score))
```

name	sch…	teacher	sex	math_score	reading_score	reading_score_centered_by_teac
<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<c
mike	south	johnson	male	4	1	0
carol	south	johnson	female	3	5	-0
greg	south	johnson	male	2	2	-1
marcia	south	johnson	female	4	4	0
peter	north	smith	male	3	5	-1
jan	north	smith	female	4	4	0
bobby	north	smith	male	5	1	1
cindy	south	perry	female	4	5	-0
alice	south	perry	female	5	4	0

Exercise 11.32

Hide

```
scores |>
  group_by(school) |>
  summarize(min_math_score = min(math_score))
```

school <chr>	min_math_score <dbl>
north	3
south	2

2 rows

Exercise 11.33

Hide

```
scores |>
  group_by(teacher) |>
  summarize(max_math_score = max(math_score))
```

teacher <chr>	max_math_score <dbl>
johnson	4
perry	5
smith	5

3 rows