r4ds Ex 5.4.1

MW

2019/05/22

5.4.1

1

> Brainstorm as many ways as possible to select `dep_time`, `dep_delay`, `arr_time`, and `arr_delay`from flights.

The simplest way is as follows;

```
flights %>% select(dep_time, dep_delay, arr_time, arr_delay)
```

```
## # A tibble: 336,776 x 4
##    dep_time dep_delay arr_time arr_delay
##       <int>     <dbl>    <int>     <dbl>
##  1      517         2      830        11
##  2      533         4      850        20
##  3      542         2      923        33
##  4      544        -1     1004       -18
##  5      554        -6      812       -25
##  6      554        -4      740        12
##  7      555        -5      913        19
##  8      557        -3      709       -14
##  9      557        -3      838        -8
## 10      558        -2      753         8
## # ... with 336,766 more rows
```

Others are as follows;

```
tmp <- c("dep_time", "dep_delay", "arr_time", "arr_delay")
flights %>% select(tmp)
```

```
## # A tibble: 336,776 x 4
##    dep_time dep_delay arr_time arr_delay
##       <int>     <dbl>    <int>     <dbl>
##  1      517         2      830        11
##  2      533         4      850        20
##  3      542         2      923        33
##  4      544        -1     1004       -18
##  5      554        -6      812       -25
##  6      554        -4      740        12
##  7      555        -5      913        19
##  8      557        -3      709       -14
##  9      557        -3      838        -8
## 10      558        -2      753         8
## # ... with 336,766 more rows
```

```
flights %>% select(contains("_time"), contains("_delay"))
```

```
## # A tibble: 336,776 x 7
##    dep_time sched_dep_time arr_time sched_arr_time air_time dep_delay
```

```
##       <int>          <int>    <int>          <int>    <dbl>     <dbl>
##  1      517            515      830            819      227         2
##  2      533            529      850            830      227         4
##  3      542            540      923            850      160         2
##  4      544            545     1004           1022      183        -1
##  5      554            600      812            837      116        -6
##  6      554            558      740            728      150        -4
##  7      555            600      913            854      158        -5
##  8      557            600      709            723       53        -3
##  9      557            600      838            846      140        -3
## 10      558            600      753            745      138        -2
## # ... with 336,766 more rows, and 1 more variable: arr_delay <dbl>
```

2

What happens if you include the name of a variable multiple times in a 'select() call?

It is selected only once no matter how many times you call.

3

What does the `one_of()` function do? Why might it be helpful in conjunction with this vector?

This is useful when you have the column names you want to select as a string vector.

4

Does the result of running the following code surprise you? How do the select helpers deal with case by default? How can you change that default?

```
select(flights, contains("TIME"))
```

```
## # A tibble: 336,776 x 6
##    dep_time sched_dep_time arr_time sched_arr_time air_time
##       <int>          <int>    <int>          <int>    <dbl>
##  1      517            515      830            819      227
##  2      533            529      850            830      227
##  3      542            540      923            850      160
##  4      544            545     1004           1022      183
##  5      554            600      812            837      116
##  6      554            558      740            728      150
##  7      555            600      913            854      158
##  8      557            600      709            723       53
##  9      557            600      838            846      140
## 10      558            600      753            745      138
## # ... with 336,766 more rows, and 1 more variable: time_hour <dttm>
```

By default, `contains()` ignores case. If you don't want to, you should use `ignore.case=FALSE`.