

Lab 3: Wrangling sales data

MinJae Jo

2025-09-18

Exercise 1

i.

`select(coffeeshop, !cogs) #!cogs` → Select all columns except for the cogs column

ii.

`select(coffeeshop, starts_with("prod")) #Starts_with("prod")` → product_line, select a column that begins with “prod”, such as product

iii.

`select(coffeeshop, contains("pe")) #contains("pe")` → Select total_expenses column with “pe” in column name

iv.

`select(coffeeshop, caffeineation = type) #Recall and rename the correction = type` → type column

Exercise 2

```
## # A tibble: 1,844 x 1
##   caffeineation
##   <chr>
## 1 Decaf
## 2 Decaf
## 3 Decaf
## 4 Decaf
## 5 Decaf
## 6 Decaf
```

```
## 7 Decaf
## 8 Decaf
## 9 Decaf
## 10 Decaf
## # i 1,834 more rows

## # A tibble: 181 x 2
##   product_line product
##   <chr>          <chr>
## 1 Leaves        Darjeeling
## 2 Leaves        Darjeeling
## 3 Leaves        Darjeeling
## 4 Leaves        Darjeeling
## 5 Leaves        Darjeeling
## 6 Leaves        Darjeeling
## 7 Leaves        Darjeeling
## 8 Leaves        Darjeeling
## 9 Leaves        Darjeeling
## 10 Leaves       Darjeeling
## # i 171 more rows
```

Exercise 3

```
## # A tibble: 1,844 x 10
##   cogs date      market product_line product sales state total_expenses type
##   <dbl> <date>    <chr> <chr>          <chr> <dbl> <chr>          <dbl> <chr>
## 1    95 2012-01-01 Centr~ Beans        Decaf ~    234 Colo~          38 Decaf
## 2    95 2012-01-01 Centr~ Beans        Decaf ~    234 Illi~          52 Decaf
## 3    72 2012-01-01 Centr~ Beans        Decaf ~    180 Colo~          55 Decaf
## 4   228 2012-01-01 Centr~ Beans        Decaf ~    456 Illi~          88 Decaf
## 5    58 2012-01-01 Centr~ Beans        Decaf ~    130 Ohio           56 Decaf
## 6    84 2012-01-01 East   Beans        Decaf ~    200 Flor~          49 Decaf
## 7    77 2012-01-01 East   Beans        Decaf ~    180 Flor~          53 Decaf
## 8    83 2012-01-01 South  Beans        Decaf ~    190 Texas          39 Decaf
## 9    54 2012-01-01 South  Beans        Decaf ~    134 Texas          26 Decaf
## 10   234 2012-01-01 West   Beans        Decaf ~    546 Cali~         109 Decaf
## # i 1,834 more rows
## # i 1 more variable: product_profit <dbl>
```

```
## # A tibble: 1,844 x 10
##   cogs date      market product_line product sales state total_expenses type
##   <dbl> <date>    <chr> <chr>          <chr> <dbl> <chr>          <dbl> <chr>
## 1    95 2012-01-01 Centr~ Beans        Decaf ~    234 Colo~          38 Decaf
## 2    95 2012-01-01 Centr~ Beans        Decaf ~    234 Illi~          52 Decaf
## 3    72 2012-01-01 Centr~ Beans        Decaf ~    180 Colo~          55 Decaf
## 4   228 2012-01-01 Centr~ Beans        Decaf ~    456 Illi~          88 Decaf
## 5    58 2012-01-01 Centr~ Beans        Decaf ~    130 Ohio           56 Decaf
```

```
## 6      84 2012-01-01 East    Beans      Decaf ~    200 Flor~      49 Decaf
## 7      77 2012-01-01 East    Beans      Decaf ~    180 Flor~      53 Decaf
## 8      83 2012-01-01 South   Beans      Decaf ~    190 Texas      39 Decaf
## 9      54 2012-01-01 South   Beans      Decaf ~    134 Texas      26 Decaf
## 10    234 2012-01-01 West    Beans      Decaf ~    546 Cali~     109 Decaf
## # i 1,834 more rows
## # i 1 more variable: 'sales - (cogs + total_expenses)' <dbl>
```

Exercise 4

```
## # A tibble: 1,844 x 11
##   cogs date      market product_line product sales state total_expenses type
##   <dbl> <date>      <chr> <chr>      <chr> <dbl> <chr>      <dbl> <chr>
## 1    95 2012-01-01 Centr~ Beans      Decaf ~    234 Colo~      38 Decaf
## 2    95 2012-01-01 Centr~ Beans      Decaf ~    234 Illi~      52 Decaf
## 3    72 2012-01-01 Centr~ Beans      Decaf ~    180 Colo~      55 Decaf
## 4   228 2012-01-01 Centr~ Beans      Decaf ~    456 Illi~      88 Decaf
## 5    58 2012-01-01 Centr~ Beans      Decaf ~    130 Ohio      56 Decaf
## 6    84 2012-01-01 East    Beans      Decaf ~    200 Flor~      49 Decaf
## 7    77 2012-01-01 East    Beans      Decaf ~    180 Flor~      53 Decaf
## 8    83 2012-01-01 South   Beans      Decaf ~    190 Texas      39 Decaf
## 9    54 2012-01-01 South   Beans      Decaf ~    134 Texas      26 Decaf
## 10   234 2012-01-01 West    Beans      Decaf ~    546 Cali~     109 Decaf
## # i 1,834 more rows
## # i 2 more variables: product_profit <dbl>, expiration_date <date>
```

Exercise 5

```
## # A tibble: 10 x 10
##   cogs date      market product_line product sales state total_expenses type
##   <dbl> <date>      <chr> <chr>      <chr> <dbl> <chr>      <dbl> <chr>
## 1    95 2012-01-01 Centr~ Beans      Decaf ~    234 Colo~      38 Decaf
## 2    95 2012-01-01 Centr~ Beans      Decaf ~    234 Illi~      52 Decaf
## 3    72 2012-01-01 Centr~ Beans      Decaf ~    180 Colo~      55 Decaf
## 4   228 2012-01-01 Centr~ Beans      Decaf ~    456 Illi~      88 Decaf
## 5    58 2012-01-01 Centr~ Beans      Decaf ~    130 Ohio      56 Decaf
## 6    84 2012-01-01 East    Beans      Decaf ~    200 Flor~      49 Decaf
## 7    77 2012-01-01 East    Beans      Decaf ~    180 Flor~      53 Decaf
## 8    83 2012-01-01 South   Beans      Decaf ~    190 Texas      39 Decaf
## 9    54 2012-01-01 South   Beans      Decaf ~    134 Texas      26 Decaf
## 10   234 2012-01-01 West    Beans      Decaf ~    546 Cali~     109 Decaf
## # i 1 more variable: product_profit <dbl>
```

Exercise 6

Exercise 7