Introduction to dplyr

Dahun Seo

2025-05-13

Contents

1 Setup

```
rm(list = ls())

suppressPackageStartupMessages({
    library(dplyr)
    library(tidyr) # for billboard dataset
    library(stringr) # for stringr::str_replace
})
```

```
utils::packageVersion("dplyr")
#> [1] '1.1.4'
```

2 Introduction

2.1 Why `dplyr`?

- `dplyr` is `R` package for data pre-processing.

- Very fast and intuitive usage using pipe operator

2.2 The pipe operator

All of the `dplyr` functions take a data frame (or tibble) as the first argument. Rather than forcing the user to either save intermediate objects or nest functions, dplyr provides the `%>%` operator from `magrittr`. `x %>% f(y)` turns into `f(x, y)` so the result from one step is then "piped" into the next step.

Warm up: dplyr::near

```
sqrt(3)^2 == 3
#> [1] FALSE
dplyr::near(sqrt(3)^2, 3)
#> [1] TRUE
```

Example data: `starwars`

```
dim(starwars) # starwars is in dplyr
#> [1] 87 14
starwars
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

3 Functions in `dplyr`

3.1 dplyr::glimpse

```
starwars %>% dplyr::glimpse()
#> Rows: 87
#> Columns: 14
#> $ name       <chr> "Luke Skywalker", "C-3PO", "R2-D2", "Darth Vader", "Leia Or~
#> $ height     <int> 172, 167, 96, 202, 150, 178, 165, 97, 183, 182, 188, 180, 2~
#> $ mass       <dbl> 77.0, 75.0, 32.0, 136.0, 49.0, 120.0, 75.0, 32.0, 84.0, 77.~
#> $ hair_color <chr> "blond", NA, NA, "none", "brown", "brown, grey", "brown", N~
#> $ skin_color <chr> "fair", "gold", "white, blue", "white", "light", "light", "~
#> $ eye_color  <chr> "blue", "yellow", "red", "yellow", "brown", "blue", "blue",~
#> $ birth_year <dbl> 19.0, 112.0, 33.0, 41.9, 19.0, 52.0, 47.0, NA, 24.0, 57.0, ~
#> $ sex        <chr> "male", "none", "none", "male", "female", "male", "female",~
#> $ gender     <chr> "masculine", "masculine", "masculine", "masculine", "femini~
#> $ homeworld  <chr> "Tatooine", "Tatooine", "Naboo", "Tatooine", "Alderaan", "T~
#> $ species    <chr> "Human", "Droid", "Droid", "Human", "Human", "Human", "Huma~
#> $ films      <list> <"A New Hope", "The Empire Strikes Back", "Return of the J~
#> $ vehicles   <list> <"Snowspeeder", "Imperial Speeder Bike">, <>, <>, <>, "Imp~
#> $ starships  <list> <"X-wing", "Imperial shuttle">, <>, <>, "TIE Advanced x1",~
```

Basic `R` code:

```
str(starwars)
```

3.2 dplyr::filter

Return rows with matching conditions

```
starwars %>%
    dplyr::filter(skin_color == "light", eye_color == "brown")
#> # A tibble: 7 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Leia Org~    150    49 brown      light      brown             19 fema~ femin~
#> 2 Biggs Da~    183    84 black      light      brown             24 male  mascu~
#> 3 Padmé Am~    185    45 brown      light      brown             46 fema~ femin~
#> 4 Cordé        157    NA brown      light      brown             NA <NA>  <NA>
#> # i 3 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::filter(skin_color == "light" & eye_color == "brown") # same
#> # A tibble: 7 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Leia Org~    150    49 brown      light      brown             19 fema~ femin~
#> 2 Biggs Da~    183    84 black      light      brown             24 male  mascu~
#> 3 Padmé Am~    185    45 brown      light      brown             46 fema~ femin~
#> 4 Cordé        157    NA brown      light      brown             NA <NA>  <NA>
#> # i 3 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

Basic R code:

```
starwars[starwars$skin_color == "light" & starwars$eye_color == "brown", ] # same
#> # A tibble: 7 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Leia Org~    150    49 brown      light      brown             19 fema~ femin~
#> 2 Biggs Da~    183    84 black      light      brown             24 male  mascu~
#> 3 Padmé Am~    185    45 brown      light      brown             46 fema~ femin~
#> 4 Cordé        157    NA brown      light      brown             NA <NA>  <NA>
#> # i 3 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
base::subset(starwars, skin_color == "light" & eye_color == "brown") # same
#> # A tibble: 7 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Leia Org~    150    49 brown      light      brown             19 fema~ femin~
#> 2 Biggs Da~    183    84 black      light      brown             24 male  mascu~
#> 3 Padmé Am~    185    45 brown      light      brown             46 fema~ femin~
#> 4 Cordé        157    NA brown      light      brown             NA <NA>  <NA>
#> # i 3 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
starwars[which(starwars$skin_color == "light" & starwars$eye_color == "brown"), , drop = FALSE] # sam
#> # A tibble: 7 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
```

```
#> 1 Leia Org~    150    49 brown      light      brown            19 fema~ femin~
#> 2 Biggs Da~    183    84 black      light      brown            24 male  mascu~
#> 3 Padmé Am~    185    45 brown      light      brown            46 fema~ femin~
#> 4 Cordé        157    NA brown      light      brown            NA <NA>  <NA>
#> # i 3 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

```
starwars %>%
    dplyr::filter(row_number() == 1)
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue              19 male  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::filter(row_number() == n())
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Captain ~     NA    NA none       none       unknown           NA fema~ femin~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::filter(between(row_number(), 2, 5))
#> # A tibble: 4 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 2 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 3 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> 4 Leia Org~    150    49 brown      light      brown           19   fema~ femin~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

3.3 dplyr::arrange

```
starwars %>%
    dplyr::arrange(height, mass) # first is priority
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yoda          66    17 white      green      brown            896 male  mascu~
#> 2 Ratts Ty~     79    15 none       grey, blue unknown           NA male  mascu~
#> 3 Wicket S~     88    20 brown      brown      brown             8 male  mascu~
#> 4 Dud Bolt      94    45 none       blue, grey yellow           NA male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::arrange(mass, height) # different
#> # A tibble: 87 x 14
```

```
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Ratts Ty~     79    15 none       grey, blue unknown           NA male  mascu~
#> 2 Yoda          66    17 white      green      brown            896 male  mascu~
#> 3 Wicket S~     88    20 brown      brown      brown              8 male  mascu~
#> 4 R2-D2         96    32 <NA>       white, bl~ red               33 none  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

Basic R code:

```
starwars[order(starwars$mass, starwars$height), ] # same
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Ratts Ty~     79    15 none       grey, blue unknown           NA male  mascu~
#> 2 Yoda          66    17 white      green      brown            896 male  mascu~
#> 3 Wicket S~     88    20 brown      brown      brown              8 male  mascu~
#> 4 R2-D2         96    32 <NA>       white, bl~ red               33 none  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

desc(): descending order

```
starwars %>%
    dplyr::arrange(desc(height)) # desc(): descending order
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yarael P~    264    NA none       white      yellow            NA male  mascu~
#> 2 Tarfful      234   136 brown      brown      blue              NA male  mascu~
#> 3 Lama Su      229    88 none       grey       black             NA male  mascu~
#> 4 Chewbacca    228   112 brown      unknown    blue             200 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

Basic R code:

```
starwars[order(starwars$height, decreasing = TRUE), ] # same
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yarael P~    264    NA none       white      yellow            NA male  mascu~
#> 2 Tarfful      234   136 brown      brown      blue              NA male  mascu~
#> 3 Lama Su      229    88 none       grey       black             NA male  mascu~
#> 4 Chewbacca    228   112 brown      unknown    blue             200 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
starwars[order(-starwars$height), ] # same
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yarael P~    264    NA none       white      yellow            NA male  mascu~
#> 2 Tarfful      234   136 brown      brown      blue              NA male  mascu~
```

```
#> 3 Lama Su      229    88 none       grey       black            NA male  mascu~
#> 4 Chewbacca    228   112 brown      unknown    blue            200 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

3.4 dplyr::slice

```
starwars %>%
    dplyr::slice(5:10)
#> # A tibble: 6 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Leia Org~    150    49 brown      light      brown             19 fema~ femin~
#> 2 Owen Lars    178   120 brown, gr~ light      blue              52 male  mascu~
#> 3 Beru Whi~    165    75 brown      light      blue              47 fema~ femin~
#> 4 R5-D4         97    32 <NA>       white, red red               NA none  mascu~
#> # i 2 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice_head(n = 3)
#> # A tibble: 3 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue              19 male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow           112 none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red               33 none  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice_head() # default is n = 1
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue              19 male  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice(-(2:n())) # same
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue              19 male  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice_tail()
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
```

```
#> 1 Captain ~     NA    NA none       none       unknown           NA fema~ femin~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice(n()) # same
#> # A tibble: 1 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Captain ~     NA    NA none       none       unknown           NA fema~ femin~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice_sample(n = 5)
#> # A tibble: 5 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Bib Fort~    180    NA none       pale       pink              NA male  mascu~
#> 2 Saesee T~    188    NA none       pale       orange            NA male  mascu~
#> 3 Finis Va~    170    NA blond      fair       blue              91 male  mascu~
#> 4 Roos Tar~    224    82 none       grey       orange            NA male  mascu~
#> # i 1 more row
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::slice_sample(prop = 0.1)
#> # A tibble: 8 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Adi Gall~    184    50 none       dark       blue            NA   fema~ femin~
#> 2 Obi-Wan ~    182    77 auburn, w~ fair       blue-gray       57   male  mascu~
#> 3 Anakin S~    188    84 blond      fair       blue            41.9 male  mascu~
#> 4 Jango Fe~    183    79 black      tan        brown           66   male  mascu~
#> # i 4 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::filter(!is.na(height)) %>%
    dplyr::slice_max(height, n = 3)
#> # A tibble: 3 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yarael P~    264    NA none       white      yellow            NA male  mascu~
#> 2 Tarfful      234   136 brown      brown      blue              NA male  mascu~
#> 3 Lama Su      229    88 none       grey       black             NA male  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars %>%
    dplyr::filter(!is.na(height)) %>%
    dplyr::slice_min(height, n = 3)
#> # A tibble: 3 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
```

```
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Yoda          66    17 white      green      brown            896 male  mascu~
#> 2 Ratts Ty~     79    15 none       grey, blue unknown           NA male  mascu~
#> 3 Wicket S~     88    20 brown      brown      brown              8 male  mascu~
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

3.5 dplyr::select

```
starwars %>% dplyr::select(name, height)
#> # A tibble: 87 x 2
#>   name           height
#>   <chr>           <int>
#> 1 Luke Skywalker    172
#> 2 C-3PO             167
#> 3 R2-D2              96
#> 4 Darth Vader       202
#> # i 83 more rows
```

Basic `R` code:

```
starwars[c("name", "height")] # same
#> # A tibble: 87 x 2
#>   name           height
#>   <chr>           <int>
#> 1 Luke Skywalker    172
#> 2 C-3PO             167
#> 3 R2-D2              96
#> 4 Darth Vader       202
#> # i 83 more rows
base::subset(starwars, select = c(name, height)) # same
#> # A tibble: 87 x 2
#>   name           height
#>   <chr>           <int>
#> 1 Luke Skywalker    172
#> 2 C-3PO             167
#> 3 R2-D2              96
#> 4 Darth Vader       202
#> # i 83 more rows
```

Variation:

```
# Select all columns except those from hair_color to eye_color
starwars %>%
    dplyr::select(!(hair_color:eye_color))
#> # A tibble: 87 x 11
#>   name     height  mass birth_year sex   gender homeworld species films vehicles
#>   <chr>     <int> <dbl>      <dbl> <chr> <chr>  <chr>     <chr>   <lis> <list>
#> 1 Luke Sk~    172    77       19   male  mascu~ Tatooine  Human   <chr> <chr>
#> 2 C-3PO       167    75      112   none  mascu~ Tatooine  Droid   <chr> <chr>
#> 3 R2-D2        96    32       33   none  mascu~ Naboo     Droid   <chr> <chr>
#> 4 Darth V~    202   136       41.9 male  mascu~ Tatooine  Human   <chr> <chr>
#> # i 83 more rows
#> # i 1 more variable: starships <list>

# Select all columns ending with color
starwars %>%
```

```
    dplyr::select(ends_with("color"))
#> # A tibble: 87 x 3
#>   hair_color skin_color  eye_color
#>   <chr>      <chr>       <chr>
#> 1 blond      fair        blue
#> 2 <NA>       gold        yellow
#> 3 <NA>       white, blue red
#> 4 none       white       yellow
#> # i 83 more rows

# Select all columns starting with h
starwars %>%
    dplyr::select(starts_with("h"))
#> # A tibble: 87 x 3
#>   height hair_color homeworld
#>    <int> <chr>      <chr>
#> 1    172 blond      Tatooine
#> 2    167 <NA>       Tatooine
#> 3     96 <NA>       Naboo
#> 4    202 none       Tatooine
#> # i 83 more rows

starwars[stringr::str_subset(names(starwars), "^h")] # same
#> # A tibble: 87 x 3
#>   height hair_color homeworld
#>    <int> <chr>      <chr>
#> 1    172 blond      Tatooine
#> 2    167 <NA>       Tatooine
#> 3     96 <NA>       Naboo
#> 4    202 none       Tatooine
#> # i 83 more rows

# Both
starwars %>%
    dplyr::select(starts_with("h") | ends_with("color"))
#> # A tibble: 87 x 5
#>   height hair_color homeworld skin_color  eye_color
#>    <int> <chr>      <chr>     <chr>       <chr>
#> 1    172 blond      Tatooine  fair        blue
#> 2    167 <NA>       Tatooine  gold        yellow
#> 3     96 <NA>       Naboo     white, blue red
#> 4    202 none       Tatooine  white       yellow
#> # i 83 more rows

starwars %>%
    dplyr::select(!starts_with("h") & ends_with("color"))
#> # A tibble: 87 x 2
#>   skin_color  eye_color
#>   <chr>       <chr>
#> 1 fair        blue
#> 2 gold        yellow
#> 3 white, blue red
#> 4 white       yellow
#> # i 83 more rows

# Select all columns containing _
```

```
starwars %>%
    dplyr::select(contains("_"))
#> # A tibble: 87 x 4
#>   hair_color skin_color  eye_color birth_year
#>   <chr>      <chr>       <chr>          <dbl>
#> 1 blond      fair        blue            19
#> 2 <NA>       gold        yellow         112
#> 3 <NA>       white, blue red             33
#> 4 none       white       yellow          41.9
#> # i 83 more rows

# Select all columns using regular expression
starwars %>%
    dplyr::select(matches("^(hair)"))
#> # A tibble: 87 x 1
#>   hair_color
#>   <chr>
#> 1 blond
#> 2 <NA>
#> 3 <NA>
#> 4 none
#> # i 83 more rows

starwars %>%
    dplyr::select(starts_with("hair")) # same
#> # A tibble: 87 x 1
#>   hair_color
#>   <chr>
#> 1 blond
#> 2 <NA>
#> 3 <NA>
#> 4 none
#> # i 83 more rows

# Select with num_range()
billboard %>% select(starts_with("wk"))
#> # A tibble: 317 x 76
#>     wk1   wk2   wk3   wk4   wk5   wk6   wk7   wk8   wk9  wk10  wk11  wk12  wk13
#>   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
#> 1    87    82    72    77    87    94    99    NA    NA    NA    NA    NA    NA
#> 2    91    87    92    NA    NA    NA    NA    NA    NA    NA    NA    NA    NA
#> 3    81    70    68    67    66    57    54    53    51    51    51    51    47
#> 4    76    76    72    69    67    65    55    59    62    61    61    59    61
#> # i 313 more rows
#> # i 63 more variables: wk14 <dbl>, wk15 <dbl>, wk16 <dbl>, wk17 <dbl>,
#> #   wk18 <dbl>, wk19 <dbl>, wk20 <dbl>, wk21 <dbl>, wk22 <dbl>, wk23 <dbl>,
#> #   wk24 <dbl>, wk25 <dbl>, wk26 <dbl>, wk27 <dbl>, wk28 <dbl>, wk29 <dbl>,
#> #   wk30 <dbl>, wk31 <dbl>, wk32 <dbl>, wk33 <dbl>, wk34 <dbl>, wk35 <dbl>,
#> #   wk36 <dbl>, wk37 <dbl>, wk38 <dbl>, wk39 <dbl>, wk40 <dbl>, wk41 <dbl>,
#> #   wk42 <dbl>, wk43 <dbl>, wk44 <dbl>, wk45 <dbl>, wk46 <dbl>, wk47 <dbl>, ...
billboard %>% select(num_range("wk", seq(1, 5, by = 2)))
#> # A tibble: 317 x 3
#>     wk1   wk3   wk5
#>   <dbl> <dbl> <dbl>
#> 1    87    72    87
#> 2    91    92    NA
```

```
#> 3    81    68    66
#> 4    76    72    67
#> # i 313 more rows
```

```
# Select columns with variable (any_of is useful)
target <- c("height", "mass", "kkk")

# starwars %>%
#     dplyr::select(target) # warning

starwars %>%
    dplyr::select(any_of(target)) # no warning
#> # A tibble: 87 x 2
#>   height  mass
#>    <int> <dbl>
#> 1    172    77
#> 2    167    75
#> 3     96    32
#> 4    202   136
#> # i 83 more rows
```

```
# Select with where
starwars %>%
    dplyr::select(where(is.numeric))
#> # A tibble: 87 x 3
#>   height  mass birth_year
#>    <int> <dbl>      <dbl>
#> 1    172    77       19
#> 2    167    75      112
#> 3     96    32       33
#> 4    202   136       41.9
#> # i 83 more rows

base::Filter(is.numeric, starwars) # same
#> # A tibble: 87 x 3
#>   height  mass birth_year
#>    <int> <dbl>      <dbl>
#> 1    172    77       19
#> 2    167    75      112
#> 3     96    32       33
#> 4    202   136       41.9
#> # i 83 more rows

starwars %>%
    dplyr::select(where(~ is.numeric(.x))) # same
#> # A tibble: 87 x 3
#>   height  mass birth_year
#>    <int> <dbl>      <dbl>
#> 1    172    77       19
#> 2    167    75      112
#> 3     96    32       33
#> 4    202   136       41.9
#> # i 83 more rows

starwars %>%
    dplyr::select(where(function(x) is.numeric(x))) # same
#> # A tibble: 87 x 3
```

```
#>   height  mass birth_year
#>    <int> <dbl>      <dbl>
#> 1    172    77       19
#> 2    167    75      112
#> 3     96    32       33
#> 4    202   136       41.9
#> # i 83 more rows

starwars %>%
    dplyr::select(where(~ is.numeric(.x) && mean(.x, na.rm = TRUE) > 100))
#> # A tibble: 87 x 1
#>   height
#>    <int>
#> 1    172
#> 2    167
#> 3     96
#> 4    202
#> # i 83 more rows

# Select columns with changed name
starwars %>%
    dplyr::select(homeworld)
#> # A tibble: 87 x 1
#>   homeworld
#>   <chr>
#> 1 Tatooine
#> 2 Tatooine
#> 3 Naboo
#> 4 Tatooine
#> # i 83 more rows

starwars %>%
    dplyr::select(home_world = homeworld)
#> # A tibble: 87 x 1
#>   home_world
#>   <chr>
#> 1 Tatooine
#> 2 Tatooine
#> 3 Naboo
#> 4 Tatooine
#> # i 83 more rows

# Select all columns
starwars %>%
    dplyr::select(everything())
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
```

```r
# Select last columns
starwars %>%
    dplyr::select(last_col())
#> # A tibble: 87 x 1
#>   starships
#>   <list>
#> 1 <chr [2]>
#> 2 <chr [0]>
#> 3 <chr [0]>
#> 4 <chr [1]>
#> # i 83 more rows

starwars %>%
    dplyr::select(last_col(1))
#> # A tibble: 87 x 1
#>   vehicles
#>   <list>
#> 1 <chr [2]>
#> 2 <chr [0]>
#> 3 <chr [0]>
#> 4 <chr [0]>
#> # i 83 more rows

height <- 5
starwars %>%
    dplyr::select(height)
#> # A tibble: 87 x 1
#>   height
#>    <int>
#> 1    172
#> 2    167
#> 3     96
#> 4    202
#> # i 83 more rows

starwars %>%
    dplyr::select(identity(height))
#> # A tibble: 87 x 1
#>   skin_color
#>   <chr>
#> 1 fair
#> 2 gold
#> 3 white, blue
#> 4 white
#> # i 83 more rows
```

3.6 dplyr::distinct

```r
starwars %>% dplyr::distinct(eye_color)
#> # A tibble: 15 x 1
#>   eye_color
#>   <chr>
#> 1 blue
#> 2 yellow
#> 3 red
#> 4 brown
```

```
#> # i 11 more rows
```

Basic R code:

```
base::unique(starwars["eye_color"]) # same
#> # A tibble: 15 x 1
#>   eye_color
#>   <chr>
#> 1 blue
#> 2 yellow
#> 3 red
#> 4 brown
#> # i 11 more rows
```

3.7 dplyr::pull

```
starwars %>% dplyr::pull(2)
#>  [1] 172 167  96 202 150 178 165  97 183 182 188 180 228 180 173 175 170 180  66
#> [20] 170 183 200 190 177 175 180 150  NA  88 160 193 191 170 185 196 224 206 183
#> [39] 137 112 183 163 175 180 178  79  94 122 163 188 198 196 171 184 188 264 188
#> [58] 196 185 157 183 183 170 166 165 193 191 183 168 198 229 213 167  96 193 191
#> [77] 178 216 234 188 178 206  NA  NA  NA  NA  NA
starwars %>% dplyr::pull(height)
#>  [1] 172 167  96 202 150 178 165  97 183 182 188 180 228 180 173 175 170 180  66
#> [20] 170 183 200 190 177 175 180 150  NA  88 160 193 191 170 185 196 224 206 183
#> [39] 137 112 183 163 175 180 178  79  94 122 163 188 198 196 171 184 188 264 188
#> [58] 196 185 157 183 183 170 166 165 193 191 183 168 198 229 213 167  96 193 191
#> [77] 178 216 234 188 178 206  NA  NA  NA  NA  NA
starwars %>% dplyr::pull("height")
#>  [1] 172 167  96 202 150 178 165  97 183 182 188 180 228 180 173 175 170 180  66
#> [20] 170 183 200 190 177 175 180 150  NA  88 160 193 191 170 185 196 224 206 183
#> [39] 137 112 183 163 175 180 178  79  94 122 163 188 198 196 171 184 188 264 188
#> [58] 196 185 157 183 183 170 166 165 193 191 183 168 198 229 213 167  96 193 191
#> [77] 178 216 234 188 178 206  NA  NA  NA  NA  NA

starwars[[1]] # numeric
#>  [1] "Luke Skywalker"        "C-3PO"                 "R2-D2"
#>  [4] "Darth Vader"           "Leia Organa"           "Owen Lars"
#>  [7] "Beru Whitesun Lars"    "R5-D4"                 "Biggs Darklighter"
#> [10] "Obi-Wan Kenobi"        "Anakin Skywalker"      "Wilhuff Tarkin"
#> [13] "Chewbacca"             "Han Solo"              "Greedo"
#> [16] "Jabba Desilijic Tiure" "Wedge Antilles"        "Jek Tono Porkins"
#> [19] "Yoda"                  "Palpatine"             "Boba Fett"
#> [22] "IG-88"                 "Bossk"                 "Lando Calrissian"
#> [25] "Lobot"                 "Ackbar"                "Mon Mothma"
#> [28] "Arvel Crynyd"          "Wicket Systri Warrick" "Nien Nunb"
#> [31] "Qui-Gon Jinn"          "Nute Gunray"           "Finis Valorum"
#> [34] "Padmé Amidala"         "Jar Jar Binks"         "Roos Tarpals"
#> [37] "Rugor Nass"            "Ric Olié"              "Watto"
#> [40] "Sebulba"               "Quarsh Panaka"         "Shmi Skywalker"
#> [43] "Darth Maul"            "Bib Fortuna"           "Ayla Secura"
#> [46] "Ratts Tyerel"          "Dud Bolt"              "Gasgano"
#> [49] "Ben Quadinaros"        "Mace Windu"            "Ki-Adi-Mundi"
#> [52] "Kit Fisto"             "Eeth Koth"             "Adi Gallia"
#> [55] "Saesee Tiin"           "Yarael Poof"           "Plo Koon"
#> [58] "Mas Amedda"            "Gregar Typho"          "Cordé"
```

```
#> [61] "Cliegg Lars"           "Poggle the Lesser"     "Luminara Unduli"
#> [64] "Barriss Offee"         "Dormé"                 "Dooku"
#> [67] "Bail Prestor Organa"   "Jango Fett"            "Zam Wesell"
#> [70] "Dexter Jettster"       "Lama Su"               "Taun We"
#> [73] "Jocasta Nu"            "R4-P17"                "Wat Tambor"
#> [76] "San Hill"              "Shaak Ti"              "Grievous"
#> [79] "Tarfful"               "Raymus Antilles"       "Sly Moore"
#> [82] "Tion Medon"            "Finn"                  "Rey"
#> [85] "Poe Dameron"           "BB8"                   "Captain Phasma"
starwars$height # numeric
#>  [1] 172 167  96 202 150 178 165  97 183 182 188 180 228 180 173 175 170 180  66
#> [20] 170 183 200 190 177 175 180 150  NA  88 160 193 191 170 185 196 224 206 183
#> [39] 137 112 183 163 175 180 178  79  94 122 163 188 198 196 171 184 188 264 188
#> [58] 196 185 157 183 183 170 166 165 193 191 183 168 198 229 213 167  96 193 191
#> [77] 178 216 234 188 178 206  NA  NA  NA  NA  NA
starwars["height"] # tibble
#> # A tibble: 87 x 1
#>   height
#>    <int>
#> 1    172
#> 2    167
#> 3     96
#> 4    202
#> # i 83 more rows
```

3.8 dplyr::relocate

```
starwars %>%
    dplyr::relocate(sex:homeworld, .before = height)
#> # A tibble: 87 x 14
#>   name       sex   gender homeworld height  mass hair_color skin_color eye_color
#>   <chr>      <chr> <chr>  <chr>      <int> <dbl> <chr>      <chr>      <chr>
#> 1 Luke Skyw~ male  mascu~ Tatooine     172    77 blond      fair       blue
#> 2 C-3PO      none  mascu~ Tatooine     167    75 <NA>       gold       yellow
#> 3 R2-D2      none  mascu~ Naboo         96    32 <NA>       white, bl~ red
#> 4 Darth Vad~ male  mascu~ Tatooine     202   136 none       white      yellow
#> # i 83 more rows
#> # i 5 more variables: birth_year <dbl>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

# to front
starwars %>% dplyr::relocate(sex, gender)
#> # A tibble: 87 x 14
#>   sex   gender    name   height  mass hair_color skin_color eye_color birth_year
#>   <chr> <chr>     <chr>   <int> <dbl> <chr>      <chr>      <chr>          <dbl>
#> 1 male  masculine Luke ~    172    77 blond      fair       blue            19
#> 2 none  masculine C-3PO     167    75 <NA>       gold       yellow         112
#> 3 none  masculine R2-D2      96    32 <NA>       white, bl~ red             33
#> 4 male  masculine Darth~    202   136 none       white      yellow          41.9
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
starwars[base::union(c("sex", "gender"), names(starwars))] # same
#> # A tibble: 87 x 14
#>   sex   gender    name   height  mass hair_color skin_color eye_color birth_year
```

```
#>   <chr> <chr>     <chr>   <int> <dbl> <chr>      <chr>      <chr>          <dbl>
#> 1 male  masculine Luke ~    172    77 blond      fair       blue            19
#> 2 none  masculine C-3PO     167    75 <NA>       gold       yellow         112
#> 3 none  masculine R2-D2      96    32 <NA>       white, bl~ red             33
#> 4 male  masculine Darth~    202   136 none       white      yellow          41.9
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

# to back
starwars %>% dplyr::relocate(sex, gender, .after = last_col())
#> # A tibble: 87 x 14
#>   name         height  mass hair_color skin_color eye_color birth_year homeworld
#>   <chr>         <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr>
#> 1 Luke Skywal~    172    77 blond      fair       blue            19   Tatooine
#> 2 C-3PO           167    75 <NA>       gold       yellow         112   Tatooine
#> 3 R2-D2            96    32 <NA>       white, bl~ red             33   Naboo
#> 4 Darth Vader     202   136 none       white      yellow          41.9 Tatooine
#> # i 83 more rows
#> # i 6 more variables: species <chr>, films <list>, vehicles <list>,
#> #   starships <list>, sex <chr>, gender <chr>
to_back <- c("sex", "gender")
starwars[c(base::setdiff(names(starwars), to_back), to_back)] # same
#> # A tibble: 87 x 14
#>   name         height  mass hair_color skin_color eye_color birth_year homeworld
#>   <chr>         <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr>
#> 1 Luke Skywal~    172    77 blond      fair       blue            19   Tatooine
#> 2 C-3PO           167    75 <NA>       gold       yellow         112   Tatooine
#> 3 R2-D2            96    32 <NA>       white, bl~ red             33   Naboo
#> 4 Darth Vader     202   136 none       white      yellow          41.9 Tatooine
#> # i 83 more rows
#> # i 6 more variables: species <chr>, films <list>, vehicles <list>,
#> #   starships <list>, sex <chr>, gender <chr>
```

3.9 dplyr::rename, rename_with

```
starwars %>%
    dplyr::rename(home_world = homeworld)
#> # A tibble: 87 x 14
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: home_world <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>

starwars2 <- starwars
colnames(starwars2)
#>  [1] "name"       "height"     "mass"       "hair_color" "skin_color"
#>  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
#> [11] "species"    "films"      "vehicles"   "starships"
names(starwars2)[1] <- "aaaaaaaa"
```

```r
colnames(starwars2)
#>  [1] "aaaaaaaa"   "height"     "mass"       "hair_color" "skin_color"
#>  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
#> [11] "species"    "films"      "vehicles"   "starships"

names(starwars2)[names(starwars2) == "height"] <- "bbbbbbbb"
colnames(starwars2)
#>  [1] "aaaaaaaa"   "bbbbbbbb"   "mass"       "hair_color" "skin_color"
#>  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
#> [11] "species"    "films"      "vehicles"   "starships"

starwars2 %>% dplyr::rename_with(toupper) %>% colnames()
#>  [1] "AAAAAAAA"   "BBBBBBBB"   "MASS"       "HAIR_COLOR" "SKIN_COLOR"
#>  [6] "EYE_COLOR"  "BIRTH_YEAR" "SEX"        "GENDER"     "HOMEWORLD"
#> [11] "SPECIES"    "FILMS"      "VEHICLES"   "STARSHIPS"
stats::setNames(starwars2, toupper(names(starwars2))) %>% colnames()
#>  [1] "AAAAAAAA"   "BBBBBBBB"   "MASS"       "HAIR_COLOR" "SKIN_COLOR"
#>  [6] "EYE_COLOR"  "BIRTH_YEAR" "SEX"        "GENDER"     "HOMEWORLD"
#> [11] "SPECIES"    "FILMS"      "VEHICLES"   "STARSHIPS"

colnames(starwars2)
#>  [1] "aaaaaaaa"   "bbbbbbbb"   "mass"       "hair_color" "skin_color"
#>  [6] "eye_color"  "birth_year" "sex"        "gender"     "homeworld"
#> [11] "species"    "films"      "vehicles"   "starships"
starwars2 %>%
    dplyr::rename_with(~ stringr::str_replace(.x, "_", "____")) %>%
    colnames()
#>  [1] "aaaaaaaa"      "bbbbbbbb"      "mass"          "hair____color"
#>  [5] "skin____color" "eye____color"  "birth____year" "sex"
#>  [9] "gender"        "homeworld"     "species"       "films"
#> [13] "vehicles"      "starships"
```

3.10 dplyr::mutate, transmute

```r
starwars %>%
    dplyr::mutate(height_m = height / 100)
#> # A tibble: 87 x 15
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 6 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>, height_m <dbl>

starwars %>%
    dplyr::mutate(height_m = height / 100) %>%
    dplyr::select(height_m, height, everything())
#> # A tibble: 87 x 15
#>   height_m height name     mass hair_color skin_color eye_color birth_year sex
#>      <dbl>  <int> <chr>   <dbl> <chr>      <chr>      <chr>          <dbl> <chr>
#> 1     1.72    172 Luke S~    77 blond      fair       blue            19   male
#> 2     1.67    167 C-3PO      75 <NA>       gold       yellow         112   none
```

```
#> 3     0.96     96 R2-D2      32 <NA>       white, bl~ red             33   none
#> 4     2.02    202 Darth ~   136 none       white      yellow          41.9 male
#> # i 83 more rows
#> # i 6 more variables: gender <chr>, homeworld <chr>, species <chr>,
#> #   films <list>, vehicles <list>, starships <list>

starwars %>%
    dplyr::mutate(
        height_m = height / 100,
        BMI = mass / (height_m^2)
    ) %>%
    dplyr::select(BMI, everything())
#> # A tibble: 87 x 16
#>     BMI name       height  mass hair_color skin_color eye_color birth_year sex
#>   <dbl> <chr>       <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr>
#> 1  26.0 Luke Skyw~    172    77 blond      fair       blue            19   male
#> 2  26.9 C-3PO         167    75 <NA>       gold       yellow         112   none
#> 3  34.7 R2-D2          96    32 <NA>       white, bl~ red             33   none
#> 4  33.3 Darth Vad~    202   136 none       white      yellow          41.9 male
#> # i 83 more rows
#> # i 7 more variables: gender <chr>, homeworld <chr>, species <chr>,
#> #   films <list>, vehicles <list>, starships <list>, height_m <dbl>

starwars %>%
    dplyr::transmute(
        height_m = height / 100,
        BMI = mass / (height_m^2)
    )
#> # A tibble: 87 x 2
#>   height_m   BMI
#>      <dbl> <dbl>
#> 1     1.72  26.0
#> 2     1.67  26.9
#> 3     0.96  34.7
#> 4     2.02  33.3
#> # i 83 more rows

starwars2$BMI <- starwars2$mass / (starwars2$bbbbbbbb / 100)^2
starwars2["BMI"]
#> # A tibble: 87 x 1
#>     BMI
#>   <dbl>
#> 1  26.0
#> 2  26.9
#> 3  34.7
#> 4  33.3
#> # i 83 more rows

# base::transform(starwars2, ccccccc = bbbbbbbb  / 100)
```

3.11 dplyr::summarise

```
starwars %>%
    dplyr::summarise(
        height = mean(height, na.rm = TRUE),
```

```
        mass = mean(mass, na.rm = TRUE),
        number_of_rows = n()
    )
#> # A tibble: 1 x 3
#>   height  mass number_of_rows
#>    <dbl> <dbl>          <int>
#> 1   175.  97.3             87
```

3.12 dplyr::group_by

```
# grouping doesn't change how the data looks
dplyr::group_by(starwars, sex)
#> # A tibble: 87 x 14
#> # Groups:   sex [5]
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
dplyr::group_by(starwars, sex = as.factor(sex))
#> # A tibble: 87 x 14
#> # Groups:   sex [5]
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <fct> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 5 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>
dplyr::group_by(starwars, height_binned = cut(height, 3))
#> # A tibble: 87 x 15
#> # Groups:   height_binned [4]
#>   name      height  mass hair_color skin_color eye_color birth_year sex   gender
#>   <chr>      <int> <dbl> <chr>      <chr>      <chr>          <dbl> <chr> <chr>
#> 1 Luke Sky~    172    77 blond      fair       blue            19   male  mascu~
#> 2 C-3PO        167    75 <NA>       gold       yellow         112   none  mascu~
#> 3 R2-D2         96    32 <NA>       white, bl~ red             33   none  mascu~
#> 4 Darth Va~    202   136 none       white      yellow          41.9 male  mascu~
#> # i 83 more rows
#> # i 6 more variables: homeworld <chr>, species <chr>, films <list>,
#> #   vehicles <list>, starships <list>, height_binned <fct>

starwars %>%
    dplyr::filter(!is.na(height)) %>%
    dplyr::group_by(height_binned = cut(height, 3)) %>%
    dplyr::summarise(mean_of_height = mean(height, na.rm = TRUE))
#> # A tibble: 3 x 2
#>   height_binned mean_of_height
#>   <fct>                  <dbl>
#> 1 (65.8,132]              94.4
```

```
#> 2 (132,198]             178.
#> 3 (198,264]             220.

starwars %>%
    dplyr::filter(!is.na(height)) %>%
    dplyr::group_by(sex, height_binned = cut(height, 3)) %>%
    dplyr::summarise(mean_of_height = mean(height, na.rm = TRUE))
#> `summarise()` has grouped output by 'sex'. You can override using the `.groups`
#> argument.
#> # A tibble: 10 x 3
#> # Groups:   sex [5]
#>   sex            height_binned mean_of_height
#>   <chr>          <fct>                  <dbl>
#> 1 female         (132,198]              168.
#> 2 female         (198,264]              213
#> 3 hermaphroditic (132,198]              175
#> 4 male           (65.8,132]              93.5
#> # i 6 more rows
```

3.13 dplyr::count

```
starwars %>% dplyr::count(sex)
#> # A tibble: 5 x 2
#>   sex                n
#>   <chr>          <int>
#> 1 female            16
#> 2 hermaphroditic     1
#> 3 male              60
#> 4 none               6
#> # i 1 more row

# be careful with factor type
starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
    dplyr::count(sex)
#> # A tibble: 5 x 2
#>   sex                n
#>   <fct>          <int>
#> 1 male              60
#> 2 none               6
#> 3 female            16
#> 4 hermaphroditic     1
#> # i 1 more row

starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
    dplyr::count(sex, .drop = FALSE)
#> # A tibble: 6 x 2
#>   sex                n
#>   <fct>          <int>
#> 1 male              60
#> 2 none               6
#> 3 female            16
#> 4 hermaphroditic     1
#> # i 2 more rows
```

```
starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
    dplyr::count(sex, .drop = TRUE)
#> # A tibble: 5 x 2
#>   sex                n
#>   <fct>          <int>
#> 1 male              60
#> 2 none               6
#> 3 female            16
#> 4 hermaphroditic     1
#> # i 1 more row

starwars %>% dplyr::count(species, sort = TRUE)
#> # A tibble: 38 x 2
#>   species     n
#>   <chr>   <int>
#> 1 Human      35
#> 2 Droid       6
#> 3 <NA>        4
#> 4 Gungan      3
#> # i 34 more rows
starwars %>% dplyr::count(sex, gender, sort = TRUE)
#> # A tibble: 6 x 3
#>   sex    gender        n
#>   <chr>  <chr>     <int>
#> 1 male   masculine    60
#> 2 female feminine     16
#> 3 none   masculine     5
#> 4 <NA>   <NA>          4
#> # i 2 more rows
```

3.14 dplyr::tally

```
starwars %>% dplyr::tally()
#> # A tibble: 1 x 1
#>       n
#>   <int>
#> 1    87

# be careful with factor type
starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
    dplyr::group_by(sex) %>%
    dplyr::tally()
#> # A tibble: 5 x 2
#>   sex                n
#>   <fct>          <int>
#> 1 male              60
#> 2 none               6
#> 3 female            16
#> 4 hermaphroditic     1
#> # i 1 more row

starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
```

```
    dplyr::group_by(sex, .drop = FALSE) %>%
    dplyr::tally()
#> # A tibble: 6 x 2
#>   sex                n
#>   <fct>          <int>
#> 1 male              60
#> 2 none               6
#> 3 female            16
#> 4 hermaphroditic     1
#> # i 2 more rows

starwars %>%
    dplyr::mutate(sex = factor(sex, levels = c(unique(sex), "kkkk"))) %>%
    dplyr::group_by(sex, .drop = FALSE) %>%
    dplyr::tally(sort = TRUE)
#> # A tibble: 6 x 2
#>   sex        n
#>   <fct>  <int>
#> 1 male      60
#> 2 female    16
#> 3 none       6
#> 4 <NA>       4
#> # i 2 more rows
```

3.15 across

```
# same function on multiple columns
# or multiple functions on multiple columns
starwars %>%
    dplyr::select(height, mass) %>%
    dplyr::mutate(across(c(height, mass), function(x) x * 100))
#> # A tibble: 87 x 2
#>   height  mass
#>    <dbl> <dbl>
#> 1  17200  7700
#> 2  16700  7500
#> 3   9600  3200
#> 4  20200 13600
#> # i 83 more rows

starwars %>%
    dplyr::select(height, mass) %>%
    dplyr::mutate(across(c(height, mass), ~ .x * 100))
#> # A tibble: 87 x 2
#>   height  mass
#>    <dbl> <dbl>
#> 1  17200  7700
#> 2  16700  7500
#> 3   9600  3200
#> 4  20200 13600
#> # i 83 more rows

starwars %>%
    dplyr::summarise(across(where(is.numeric), \(x) mean(x, na.rm = TRUE)))
#> # A tibble: 1 x 3
#>   height  mass birth_year
```

```
#>    <dbl> <dbl>      <dbl>
#> 1   175.  97.3       87.6
```

3.16 quasiquotation syntax

```
old_name <- "mass"
new_name <- "weight"

starwars %>%
    dplyr::rename(!!new_name := !!sym(old_name)) %>%
    select(name, !!sym(new_name))
#> # A tibble: 87 x 2
#>   name           weight
#>   <chr>           <dbl>
#> 1 Luke Skywalker     77
#> 2 C-3PO              75
#> 3 R2-D2              32
#> 4 Darth Vader       136
#> # i 83 more rows

target <- c("height", "mass", "kkk")

starwars %>%
    dplyr::select(any_of(target)) # no warning
#> # A tibble: 87 x 2
#>   height  mass
#>    <int> <dbl>
#> 1    172    77
#> 2    167    75
#> 3     96    32
#> 4    202   136
#> # i 83 more rows

# starwars %>%
#     dplyr::select(!!!syms(target)) # error

target <- c("height", "mass")

starwars %>%
    dplyr::select(!!!syms(target))
#> # A tibble: 87 x 2
#>   height  mass
#>    <int> <dbl>
#> 1    172    77
#> 2    167    75
#> 3     96    32
#> 4    202   136
#> # i 83 more rows

new_var <- "height_cm"

starwars %>%
    dplyr::mutate(!!new_var := height * 1) %>%
    dplyr::select(name, height, !!sym(new_var))
#> # A tibble: 87 x 3
#>   name           height height_cm
```

```
#>   <chr>           <int>     <dbl>
#> 1 Luke Skywalker    172       172
#> 2 C-3PO             167       167
#> 3 R2-D2              96        96
#> 4 Darth Vader       202       202
#> # i 83 more rows

var <- "mass"

starwars %>%
    dplyr::filter(!!sym(var) > 100) %>%
    dplyr::select(name, !!sym(var))
#> # A tibble: 10 x 2
#>   name                  mass
#>   <chr>                <dbl>
#> 1 Darth Vader            136
#> 2 Owen Lars              120
#> 3 Chewbacca              112
#> 4 Jabba Desilijic Tiure  1358
#> # i 6 more rows

target <- "eye_color"

starwars %>%
    dplyr::distinct(!!sym(target))
#> # A tibble: 15 x 1
#>   eye_color
#>   <chr>
#> 1 blue
#> 2 yellow
#> 3 red
#> 4 brown
#> # i 11 more rows
```