dplyr tutorial

Grace Skinner

19/May/2022

Kaggle - Dive into dplyr website

Intro

- For cleaning and exploring data
- Can feed manipulated data into ggplot2 package for visualization
- Use the Palmer Penguins dataset, which I have downloaded

Set up our environment

Either load whole tidyverse package or the packages we're interested in

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.8
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
```

```
## Warning: package 'tibble' was built under R version 4.0.5
```

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

```
## Warning: package 'readr' was built under R version 4.0.5
```

```
## Warning: package 'purrr' was built under R version 4.0.5
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
## Warning: package 'stringr' was built under R version 4.0.5
```

```
## Warning: package 'forcats' was built under R version 4.0.5
```

```
## -- Conflicts --------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Conflicts are due to multiple packages using the same name for their functions. E.g. the filter() function from the dplyr package masks the filter() function from the stats package. To use the filter() function in the stats package, we will have to use stats::filter() syntax.

Import our data

```
penguins <- read_csv("penguins_size.csv")
```

```
## Rows: 344 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): species, island, sex
## dbl (4): culmen_length_mm, culmen_depth_mm, flipper_length_mm, body_mass_g
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# The read_csv is a function in readr (package in tidyverse)
# Don't worry about the text appearing being in red
# Could equally just use the read.csv function
penguins
```

```
## # A tibble: 344 x 7
##    species island  culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>   <chr>              <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie  Torger~             39.1            18.7              181        3750
##  2 Adelie  Torger~             39.5            17.4              186        3800
##  3 Adelie  Torger~             40.3            18                195        3250
##  4 Adelie  Torger~             NA              NA                 NA          NA
##  5 Adelie  Torger~             36.7            19.3              193        3450
##  6 Adelie  Torger~             39.3            20.6              190        3650
##  7 Adelie  Torger~             38.9            17.8              181        3625
##  8 Adelie  Torger~             39.2            19.6              195        4675
##  9 Adelie  Torger~             34.1            18.1              193        3475
## 10 Adelie  Torger~             42              20.2              190        4250
## # ... with 334 more rows, and 1 more variable: sex <chr>
```

Check out our data

```
# glimpse()
# Includes dataframe structure, variables and their data type, and a look at the first few rows of each
glimpse(penguins)
```

```
## Rows: 344
## Columns: 7
## $ species           <chr> "Adelie", "Adelie", "Adelie", "Adelie", "Adelie", "A~
## $ island            <chr> "Torgersen", "Torgersen", "Torgersen", "Torgersen", ~
## $ culmen_length_mm  <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3, 38.9, 39.2, 34.1, ~
## $ culmen_depth_mm   <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6, 17.8, 19.6, 18.1, ~
## $ flipper_length_mm <dbl> 181, 186, 195, NA, 193, 190, 181, 195, 193, 190, 186~
## $ body_mass_g       <dbl> 3750, 3800, 3250, NA, 3450, 3650, 3625, 4675, 3475, ~
## $ sex               <chr> "MALE", "FEMALE", "FEMALE", NA, "FEMALE", "MALE", "F~
```

```
# head()
# Shows first few rows of dataframe
# Good for seeing data is formatted consistently/correctly
head(penguins)
```

```
## # A tibble: 6 x 7
##   species island   culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>   <chr>               <dbl>           <dbl>            <dbl>       <dbl>
## 1 Adelie  Torgers~             39.1            18.7              181        3750
## 2 Adelie  Torgers~             39.5            17.4              186        3800
## 3 Adelie  Torgers~             40.3            18                195        3250
## 4 Adelie  Torgers~             NA              NA                 NA          NA
## 5 Adelie  Torgers~             36.7            19.3              193        3450
## 6 Adelie  Torgers~             39.3            20.6              190        3650
## # ... with 1 more variable: sex <chr>
```

```
# summary()
# summary statistics for the dataset
# Can get idea of spread, or how much missing data there is
summary(penguins)
```

```
##    species             island          culmen_length_mm culmen_depth_mm
##  Length:344         Length:344         Min.   :32.10    Min.   :13.10
##  Class :character   Class :character   1st Qu.:39.23    1st Qu.:15.60
##  Mode  :character   Mode  :character   Median :44.45    Median :17.30
##                                        Mean   :43.92    Mean   :17.15
##                                        3rd Qu.:48.50    3rd Qu.:18.70
##                                        Max.   :59.60    Max.   :21.50
##                                        NA's   :2        NA's   :2
##  flipper_length_mm  body_mass_g       sex
##  Min.   :172.0     Min.   :2700   Length:344
##  1st Qu.:190.0     1st Qu.:3550   Class :character
##  Median :197.0     Median :4050   Mode  :character
##  Mean   :200.9     Mean   :4202
##  3rd Qu.:213.0     3rd Qu.:4750
##  Max.   :231.0     Max.   :6300
##  NA's   :2         NA's   :2
```

```
# names()
# to get names of all variables in dataset
names(penguins)
```

```
## [1] "species"           "island"            "culmen_length_mm"
## [4] "culmen_depth_mm"   "flipper_length_mm" "body_mass_g"
## [7] "sex"
```

Exploring our data

%>% is the pipe operator, which allows us to push our data through sequential functions.

filter()

```
# Take the penguins dataset and then filter for all penguins that live on the Torgersen island
penguins %>%
  filter(island == "Torgersen")
```

```
## # A tibble: 52 x 7
##    species island  culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>   <chr>              <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie  Torger~             39.1            18.7              181        3750
##  2 Adelie  Torger~             39.5            17.4              186        3800
##  3 Adelie  Torger~             40.3            18                195        3250
##  4 Adelie  Torger~             NA              NA                 NA          NA
##  5 Adelie  Torger~             36.7            19.3              193        3450
##  6 Adelie  Torger~             39.3            20.6              190        3650
##  7 Adelie  Torger~             38.9            17.8              181        3625
##  8 Adelie  Torger~             39.2            19.6              195        4675
##  9 Adelie  Torger~             34.1            18.1              193        3475
## 10 Adelie  Torger~             42              20.2              190        4250
## # ... with 42 more rows, and 1 more variable: sex <chr>
```

```
penguins %>%
  filter(island == "Torgersen") %>%
  head()
```

```
## # A tibble: 6 x 7
##   species island   culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>   <chr>               <dbl>           <dbl>            <dbl>       <dbl>
## 1 Adelie  Torgers~             39.1            18.7              181        3750
## 2 Adelie  Torgers~             39.5            17.4              186        3800
## 3 Adelie  Torgers~             40.3            18                195        3250
## 4 Adelie  Torgers~             NA              NA                 NA          NA
## 5 Adelie  Torgers~             36.7            19.3              193        3450
## 6 Adelie  Torgers~             39.3            20.6              190        3650
## # ... with 1 more variable: sex <chr>
```

arrange()

- Arranges/organsises data in ascending order
- Provide a single argument to the arrange() function

```r
# numeric data
penguins %>%
  arrange(culmen_length_mm) %>%
  head()
```

```
## # A tibble: 6 x 7
##   species island   culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>   <chr>               <dbl>           <dbl>            <dbl>       <dbl>
## 1 Adelie  Dream                32.1            15.5              188        3050
## 2 Adelie  Dream                33.1            16.1              178        2900
## 3 Adelie  Torgers~             33.5            19                190        3600
## 4 Adelie  Dream                34              17.1              185        3400
## 5 Adelie  Torgers~             34.1            18.1              193        3475
## 6 Adelie  Torgers~             34.4            18.4              184        3325
## # ... with 1 more variable: sex <chr>
```

```r
# character data
penguins %>%
  arrange(species)
```

```
## # A tibble: 344 x 7
##    species island  culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>   <chr>              <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie  Torger~             39.1            18.7              181        3750
##  2 Adelie  Torger~             39.5            17.4              186        3800
##  3 Adelie  Torger~             40.3            18                195        3250
##  4 Adelie  Torger~             NA              NA                 NA          NA
##  5 Adelie  Torger~             36.7            19.3              193        3450
##  6 Adelie  Torger~             39.3            20.6              190        3650
##  7 Adelie  Torger~             38.9            17.8              181        3625
##  8 Adelie  Torger~             39.2            19.6              195        4675
##  9 Adelie  Torger~             34.1            18.1              193        3475
## 10 Adelie  Torger~             42              20.2              190        4250
## # ... with 334 more rows, and 1 more variable: sex <chr>
```

Subset

```r
# creating a random subset of the penguins dataset
set.seed(406)

penguins_subset <- penguins %>%
  sample_n(12)  # another dplyr function!

penguins_subset
```

```
## # A tibble: 12 x 7
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>              <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie   Torge~             41.4            18.5              202        3875
##  2 Chinstr~ Dream              50.1            17.9              190        3400
```

```
##  3 Gentoo   Biscoe             50.5           15.9              222        5550
##  4 Chinstr~ Dream              49             19.6              212        4300
##  5 Chinstr~ Dream              43.5           18.1              202        3400
##  6 Gentoo   Biscoe             51.5           16.3              230        5500
##  7 Adelie   Biscoe             40.5           17.9              187        3200
##  8 Gentoo   Biscoe             43.5           15.2              213        4650
##  9 Adelie   Dream              36.3           19.5              190        3800
## 10 Adelie   Torge~             39             17.1              191        3050
## 11 Adelie   Biscoe             41.6           18                192        3950
## 12 Gentoo   Biscoe             47.6           14.5              215        5400
## # ... with 1 more variable: sex <chr>
```

```
penguins_subset %>%
  arrange(species)
```

```
## # A tibble: 12 x 7
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>             <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie   Torge~             41.4            18.5              202        3875
##  2 Adelie   Biscoe             40.5            17.9              187        3200
##  3 Adelie   Dream              36.3            19.5              190        3800
##  4 Adelie   Torge~             39              17.1              191        3050
##  5 Adelie   Biscoe             41.6            18                192        3950
##  6 Chinstr~ Dream              50.1            17.9              190        3400
##  7 Chinstr~ Dream              49              19.6              212        4300
##  8 Chinstr~ Dream              43.5            18.1              202        3400
##  9 Gentoo   Biscoe             50.5            15.9              222        5550
## 10 Gentoo   Biscoe             51.5            16.3              230        5500
## 11 Gentoo   Biscoe             43.5            15.2              213        4650
## 12 Gentoo   Biscoe             47.6            14.5              215        5400
## # ... with 1 more variable: sex <chr>
```

```
# Can also nest desc() inside arrange() to get data in descending order
# numeric data arranged in descending order
penguins_subset %>%
  arrange(desc(culmen_length_mm))
```

```
## # A tibble: 12 x 7
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>             <dbl>           <dbl>            <dbl>       <dbl>
##  1 Gentoo   Biscoe             51.5            16.3              230        5500
##  2 Gentoo   Biscoe             50.5            15.9              222        5550
##  3 Chinstr~ Dream              50.1            17.9              190        3400
##  4 Chinstr~ Dream              49              19.6              212        4300
##  5 Gentoo   Biscoe             47.6            14.5              215        5400
##  6 Chinstr~ Dream              43.5            18.1              202        3400
##  7 Gentoo   Biscoe             43.5            15.2              213        4650
##  8 Adelie   Biscoe             41.6            18                192        3950
##  9 Adelie   Torge~             41.4            18.5              202        3875
## 10 Adelie   Biscoe             40.5            17.9              187        3200
## 11 Adelie   Torge~             39              17.1              191        3050
## 12 Adelie   Dream              36.3            19.5              190        3800
## # ... with 1 more variable: sex <chr>
```

```r
# character data arranged in descending - reverse alphabetical - order
penguins_subset %>%
  arrange(desc(species))
```

```
## # A tibble: 12 x 7
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>             <dbl>           <dbl>             <dbl>       <dbl>
##  1 Gentoo   Biscoe             50.5            15.9               222        5550
##  2 Gentoo   Biscoe             51.5            16.3               230        5500
##  3 Gentoo   Biscoe             43.5            15.2               213        4650
##  4 Gentoo   Biscoe             47.6            14.5               215        5400
##  5 Chinstr~ Dream              50.1            17.9               190        3400
##  6 Chinstr~ Dream              49              19.6               212        4300
##  7 Chinstr~ Dream              43.5            18.1               202        3400
##  8 Adelie   Torge~             41.4            18.5               202        3875
##  9 Adelie   Biscoe             40.5            17.9               187        3200
## 10 Adelie   Dream              36.3            19.5               190        3800
## 11 Adelie   Torge~             39              17.1               191        3050
## 12 Adelie   Biscoe             41.6            18                 192        3950
## # ... with 1 more variable: sex <chr>
```

More filtering

- Can use filter on single or multiple conditions

```r
# filter with a single numeric condition
penguins_subset %>%
  filter(culmen_depth_mm > 16.2)
```

```
## # A tibble: 9 x 7
##   species   island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>     <chr>             <dbl>           <dbl>             <dbl>       <dbl>
## 1 Adelie    Torge~             41.4            18.5               202        3875
## 2 Chinstrap Dream              50.1            17.9               190        3400
## 3 Chinstrap Dream              49              19.6               212        4300
## 4 Chinstrap Dream              43.5            18.1               202        3400
## 5 Gentoo    Biscoe             51.5            16.3               230        5500
## 6 Adelie    Biscoe             40.5            17.9               187        3200
## 7 Adelie    Dream              36.3            19.5               190        3800
## 8 Adelie    Torge~             39              17.1               191        3050
## 9 Adelie    Biscoe             41.6            18                 192        3950
## # ... with 1 more variable: sex <chr>
```

```r
# filter with a single character condition
penguins_subset %>%
  filter(island == "Dream")
```

```
## # A tibble: 4 x 7
##   species   island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>     <chr>             <dbl>           <dbl>             <dbl>       <dbl>
## 1 Chinstrap Dream              50.1            17.9               190        3400
```

```
## 2 Chinstrap Dream              49              19.6              212        4300
## 3 Chinstrap Dream              43.5            18.1              202        3400
## 4 Adelie    Dream              36.3            19.5              190        3800
## # ... with 1 more variable: sex <chr>
```

```
# filter with a single numeric condition between two values
penguins_subset %>%
  filter(between(culmen_depth_mm, 16.2, 18.1))
```

```
## # A tibble: 6 x 7
##   species   island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##   <chr>     <chr>             <dbl>           <dbl>             <dbl>       <dbl>
## 1 Chinstrap Dream              50.1            17.9               190        3400
## 2 Chinstrap Dream              43.5            18.1               202        3400
## 3 Gentoo    Biscoe             51.5            16.3               230        5500
## 4 Adelie    Biscoe             40.5            17.9               187        3200
## 5 Adelie    Torge~             39              17.1               191        3050
## 6 Adelie    Biscoe             41.6            18                 192        3950
## # ... with 1 more variable: sex <chr>
```

select()

- Pick which columns/variables we want to look at
- Use it to pull a subset of variables, or rearrange order of variables

```
# selecting species, flipper_length_mm, and sex columns
penguins_subset %>%
  select(species, flipper_length_mm, sex)
```

```
## # A tibble: 12 x 3
##    species   flipper_length_mm sex
##    <chr>                 <dbl> <chr>
##  1 Adelie                  202 MALE
##  2 Chinstrap               190 FEMALE
##  3 Gentoo                  222 MALE
##  4 Chinstrap               212 MALE
##  5 Chinstrap               202 FEMALE
##  6 Gentoo                  230 MALE
##  7 Adelie                  187 FEMALE
##  8 Gentoo                  213 FEMALE
##  9 Adelie                  190 MALE
## 10 Adelie                  191 FEMALE
## 11 Adelie                  192 MALE
## 12 Gentoo                  215 MALE
```

```
# selecting all character data
penguins_subset %>%
  select(where(is.character))
```

```
## # A tibble: 12 x 3
##    species   island    sex
```

```
##    <chr>     <chr>     <chr>
##  1 Adelie    Torgersen MALE
##  2 Chinstrap Dream     FEMALE
##  3 Gentoo    Biscoe    MALE
##  4 Chinstrap Dream     MALE
##  5 Chinstrap Dream     FEMALE
##  6 Gentoo    Biscoe    MALE
##  7 Adelie    Biscoe    FEMALE
##  8 Gentoo    Biscoe    FEMALE
##  9 Adelie    Dream     MALE
## 10 Adelie    Torgersen FEMALE
## 11 Adelie    Biscoe    MALE
## 12 Gentoo    Biscoe    MALE
```

```
# selecting all numeric data
penguins_subset %>%
  select(where(is.numeric))
```

```
## # A tibble: 12 x 4
##    culmen_length_mm culmen_depth_mm flipper_length_mm body_mass_g
##               <dbl>           <dbl>             <dbl>       <dbl>
##  1             41.4            18.5               202        3875
##  2             50.1            17.9               190        3400
##  3             50.5            15.9               222        5550
##  4             49              19.6               212        4300
##  5             43.5            18.1               202        3400
##  6             51.5            16.3               230        5500
##  7             40.5            17.9               187        3200
##  8             43.5            15.2               213        4650
##  9             36.3            19.5               190        3800
## 10             39              17.1               191        3050
## 11             41.6            18                 192        3950
## 12             47.6            14.5               215        5400
```

```
# selecting all character data by using "where not numeric" data
penguins_subset %>%
  select(!where(is.numeric))
```

```
## # A tibble: 12 x 3
##    species   island    sex
##    <chr>     <chr>     <chr>
##  1 Adelie    Torgersen MALE
##  2 Chinstrap Dream     FEMALE
##  3 Gentoo    Biscoe    MALE
##  4 Chinstrap Dream     MALE
##  5 Chinstrap Dream     FEMALE
##  6 Gentoo    Biscoe    MALE
##  7 Adelie    Biscoe    FEMALE
##  8 Gentoo    Biscoe    FEMALE
##  9 Adelie    Dream     MALE
## 10 Adelie    Torgersen FEMALE
## 11 Adelie    Biscoe    MALE
## 12 Gentoo    Biscoe    MALE
```

mutate()

- Can create new columns/variables
- Works well with group_by()

```
# converting grams to pounds
# notice how the order of our columns stays the same, and the new column, body_weight_pounds, gets plac
penguins_subset %>%
  mutate(body_weight_pounds = body_mass_g / 453.59237)
```

```
## # A tibble: 12 x 8
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>             <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie   Torge~             41.4            18.5              202        3875
##  2 Chinstr~ Dream              50.1            17.9              190        3400
##  3 Gentoo   Biscoe             50.5            15.9              222        5550
##  4 Chinstr~ Dream              49              19.6              212        4300
##  5 Chinstr~ Dream              43.5            18.1              202        3400
##  6 Gentoo   Biscoe             51.5            16.3              230        5500
##  7 Adelie   Biscoe             40.5            17.9              187        3200
##  8 Gentoo   Biscoe             43.5            15.2              213        4650
##  9 Adelie   Dream              36.3            19.5              190        3800
## 10 Adelie   Torge~             39              17.1              191        3050
## 11 Adelie   Biscoe             41.6            18                192        3950
## 12 Gentoo   Biscoe             47.6            14.5              215        5400
## # ... with 2 more variables: sex <chr>, body_weight_pounds <dbl>
```

```
# Combining select and mutate
penguins_subset %>%
  mutate(body_weight_pounds = body_mass_g / 453.59237) %>%
  select(everything()) # everything selects all variables
```

```
## # A tibble: 12 x 8
##    species  island culmen_length_mm culmen_depth_mm flipper_length_~ body_mass_g
##    <chr>    <chr>             <dbl>           <dbl>            <dbl>       <dbl>
##  1 Adelie   Torge~             41.4            18.5              202        3875
##  2 Chinstr~ Dream              50.1            17.9              190        3400
##  3 Gentoo   Biscoe             50.5            15.9              222        5550
##  4 Chinstr~ Dream              49              19.6              212        4300
##  5 Chinstr~ Dream              43.5            18.1              202        3400
##  6 Gentoo   Biscoe             51.5            16.3              230        5500
##  7 Adelie   Biscoe             40.5            17.9              187        3200
##  8 Gentoo   Biscoe             43.5            15.2              213        4650
##  9 Adelie   Dream              36.3            19.5              190        3800
## 10 Adelie   Torge~             39              17.1              191        3050
## 11 Adelie   Biscoe             41.6            18                192        3950
## 12 Gentoo   Biscoe             47.6            14.5              215        5400
## # ... with 2 more variables: sex <chr>, body_weight_pounds <dbl>
```

summarise()

- Can either use summarise() or summarize()

- Useful to use with group_by()

```
# summarising the average body mass of penguins, in grams
penguins_subset %>%
  summarise(avg_body_mass = mean(body_mass_g))
```

```
## # A tibble: 1 x 1
##   avg_body_mass
##           <dbl>
## 1         4173.
```

```
# since we're now summarising our data we can go ahead and use the full dataframe, since the printout w
penguins %>%
  summarise(avg_body_mass = mean(body_mass_g))
```

```
## # A tibble: 1 x 1
##   avg_body_mass
##           <dbl>
## 1            NA
```

```
# This doesn't work very well due to NAs in the data!!!
# For now we're going to use na.rm = TRUE, but you could use filter() from the dplyr package or drop_na

# summarising body mass on the entire penguins dataset while removing NA values from the calculation
penguins %>%
  summarise(avg_body_mass = mean(body_mass_g, na.rm = TRUE))
```

```
## # A tibble: 1 x 1
##   avg_body_mass
##           <dbl>
## 1         4202.
```

```
# now let's use the grouping function, group_by(), to look at the average body mass of penguins, in gra
penguins %>%
  group_by(species) %>%
  summarise(avg_species_body_mass = mean(body_mass_g, na.rm = TRUE))
```

```
## # A tibble: 3 x 2
##   species   avg_species_body_mass
##   <chr>                     <dbl>
## 1 Adelie                    3701.
## 2 Chinstrap                 3733.
## 3 Gentoo                    5076.
```

```
# now let's calculate the average body mass by species AND island
penguins %>%
  group_by(species, island) %>%
  summarise(avg_species_body_mass = mean(body_mass_g, na.rm = TRUE))
```

```
## `summarise()` has grouped output by 'species'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 5 x 3
## # Groups:   species [3]
##   species   island    avg_species_body_mass
##   <chr>     <chr>                     <dbl>
## 1 Adelie    Biscoe                    3710.
## 2 Adelie    Dream                     3688.
## 3 Adelie    Torgersen                 3706.
## 4 Chinstrap Dream                     3733.
## 5 Gentoo    Biscoe                    5076.
```