



Tidy data reshaping & summaries

demonstration

Athanasia Monika Mowinckel

Sept. 15th 2020

LCBC
LIFESPAN CHANGES
in brain and cognition

pivots

```
penguins %>%
  pivot_longer(contains("_"))
```

```
## # A tibble: 1,376 x 6
##   species island    sex    year name      value
##   <fct>    <fct>    <fct> <int> <chr>    <dbl>
## 1 Adelie   Torgersen male    2007 bill_length_mm    39.1
## 2 Adelie   Torgersen male    2007 bill_depth_mm    18.7
## 3 Adelie   Torgersen male    2007 flipper_length_mm  181
## 4 Adelie   Torgersen male    2007 body_mass_g      3750
## 5 Adelie   Torgersen female  2007 bill_length_mm    39.5
## 6 Adelie   Torgersen female  2007 bill_depth_mm    17.4
## 7 Adelie   Torgersen female  2007 flipper_length_mm  186
## 8 Adelie   Torgersen female  2007 body_mass_g      3800
## 9 Adelie   Torgersen female  2007 bill_length_mm    40.3
## 10 Adelie  Torgersen female  2007 bill_depth_mm     18
## # ... with 1,366 more rows
```

```
penguins %>%
  pivot_longer(contains("_"), # select the columns
               names_to = c("body_part", "measure", "unit"), # break them into these columns
               names_sep = "_") # break the column names on this character
```

```
## # A tibble: 1,376 x 8
##   species island    sex    year body_part measure unit   value
##   <fct>    <fct>    <fct> <int> <chr>      <chr>  <chr>  <dbl>
## 1 Adelie  Torgersen male    2007 bill      length mm     39.1
## 2 Adelie  Torgersen male    2007 bill      depth mm     18.7
## 3 Adelie  Torgersen male    2007 flipper length mm    181
## 4 Adelie  Torgersen male    2007 body      mass  g    3750
## 5 Adelie  Torgersen female  2007 bill      length mm     39.5
## 6 Adelie  Torgersen female  2007 bill      depth mm     17.4
## 7 Adelie  Torgersen female  2007 flipper length mm    186
## 8 Adelie  Torgersen female  2007 body      mass  g    3800
## 9 Adelie  Torgersen female  2007 bill      length mm     40.3
## 10 Adelie Torgersen female  2007 bill      depth mm      18
## # ... with 1,366 more rows
```

```
penguins %>%
  pivot_longer(contains("_"),
               names_to = c("body_part",
                           "measure",
                           "unit"),
               names_sep = "_") %>%

  ggplot(aes(x = value, fill = species)) +
  geom_density() +
  facet_wrap(~ body_part, scales = "free") +
  scale_fill_viridis_d(alpha = .5) +
  theme(legend.position = "bottom")
```

```

penguins_long <- penguins %>%
  pivot_longer(contains("_"),
               names_to = c("body_part",
                           "measure",
                           "unit"),
               names_sep = "_")

penguins_long %>%
  pivot_wider(names_from = c("body_part", "measure", "unit"), # pivot these columns
              values_from = "value", # take the values from here
              names_sep = "_") # separate names_from with this character

```

```

## # A tibble: 35 x 8
##   species island sex    year bill_length_mm bill_depth_mm flipper_length_...
##   <fct>    <fct> <fct> <int> <list>          <list>          <list>
## 1 Adelie  Torge... male   2007 <dbl [7]>        <dbl [7]>        <dbl [7]>
## 2 Adelie  Torge... fema... 2007 <dbl [8]>        <dbl [8]>        <dbl [8]>
## 3 Adelie  Torge... <NA>    2007 <dbl [5]>        <dbl [5]>        <dbl [5]>
## 4 Adelie  Biscoe fema... 2007 <dbl [5]>        <dbl [5]>        <dbl [5]>
## 5 Adelie  Biscoe male   2007 <dbl [5]>        <dbl [5]>        <dbl [5]>
## 6 Adelie  Dream  fema... 2007 <dbl [9]>        <dbl [9]>        <dbl [9]>
## 7 Adelie  Dream  male   2007 <dbl [10]>       <dbl [10]>       <dbl [10]>
## 8 Adelie  Dream  <NA>    2007 <dbl [1]>        <dbl [1]>        <dbl [1]>
## 9 Adelie  Biscoe fema... 2008 <dbl [9]>        <dbl [9]>        <dbl [9]>
## 10 Adelie Biscoe male   2008 <dbl [9]>        <dbl [9]>        <dbl [9]>
## # ... with 25 more rows, and 1 more variable: body_mass_g <list>

```

summaries

```
penguins %>%  
  summarise(mean(bill_length_mm, na.rm = TRUE))
```

```
## # A tibble: 1 x 1  
##   `mean(bill_length_mm, na.rm = TRUE)`  
##                                     <dbl>  
## 1                                     43.9
```



```
penguins %>%  
  group_by(species) %>%  
  summarise(m_bill_length = mean(bill_length_mm, na.rm = TRUE))
```

```
## # A tibble: 3 x 2  
##   species    m_bill_length  
##   <fct>         <dbl>  
## 1 Adelie         38.8  
## 2 Chinstrap      48.8  
## 3 Gentoo         47.5
```

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(m_bill_length = mean(bill_length_mm, na.rm = TRUE))
```

```
## # A tibble: 5 x 3  
## # Groups:   species [3]  
##   species    island    m_bill_length  
##   <fct>      <fct>      <dbl>  
## 1 Adelie    Biscoe         39.0  
## 2 Adelie    Dream         38.5  
## 3 Adelie    Torgersen     39.0  
## 4 Chinstrap Dream         48.8  
## 5 Gentoo    Biscoe         47.5
```

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(across(bill_length_mm,  
                    mean  
  ))
```

```
## # A tibble: 5 x 3  
## # Groups:   species [3]  
##   species    island bill_length_mm  
##   <fct>      <fct>          <dbl>  
## 1 Adelie     Biscoe             39.0  
## 2 Adelie     Dream              38.5  
## 3 Adelie     Torgersen          NA  
## 4 Chinstrap Dream             48.8  
## 5 Gentoo     Biscoe             NA
```

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(across(bill_length_mm,  
                    list(mean, sd)  
  ))
```

```
## # A tibble: 5 x 4  
## # Groups:   species [3]  
##   species    island bill_length_mm_1 bill_length_mm_2  
##   <fct>      <fct>          <dbl>          <dbl>  
## 1 Adelie    Biscoe             39.0             2.48  
## 2 Adelie    Dream              38.5             2.47  
## 3 Adelie    Torgersen          NA              NA  
## 4 Chinstrap Dream             48.8             3.34  
## 5 Gentoo    Biscoe             NA              NA
```

```
penguins %>%
  group_by(species, island) %>%
  summarise(across(bill_length_mm,
                    list(mean = mean, sd = sd)
  ))
```

```
## # A tibble: 5 x 4
## # Groups:   species [3]
##   species    island bill_length_mm_mean bill_length_mm_sd
##   <fct>      <fct>          <dbl>          <dbl>
## 1 Adelie     Biscoe             39.0             2.48
## 2 Adelie     Dream              38.5             2.47
## 3 Adelie     Torgersen          NA              NA
## 4 Chinstrap Dream             48.8             3.34
## 5 Gentoo     Biscoe             NA              NA
```

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(across(bill_length_mm,  
                    list(mean = mean, sd = sd),  
                    .names = "{.fn}")  
  ))
```

```
## # A tibble: 5 x 4  
## # Groups:   species [3]  
##   species    island      mean    sd  
##   <fct>     <fct>    <dbl> <dbl>  
## 1 Adelie   Biscoe     39.0  2.48  
## 2 Adelie   Dream      38.5  2.47  
## 3 Adelie   Torgersen  NA    NA  
## 4 Chinstrap Dream      48.8  3.34  
## 5 Gentoo   Biscoe     NA    NA
```

```

penguins %>%
  group_by(species, island) %>%
  summarise(across(contains("_"),
                    list(Mean = mean,
                         SD = sd,
                         Min = min,
                         Max = max),
                    na.rm = TRUE))

```

```

## # A tibble: 5 x 18
## # Groups:   species [3]
##   species island bill_length_mm_... bill_length_mm_... bill_length_mm_...
##   <fct>    <fct>          <dbl>          <dbl>          <dbl>
## 1 Adelie  Biscoe           39.0           2.48           34.5
## 2 Adelie  Dream            38.5           2.47           32.1
## 3 Adelie  Torge...         39.0           3.03           33.5
## 4 Chin... Dream            48.8           3.34           40.9
## 5 Gentoo  Biscoe           47.5           3.08           40.9
## # ... with 13 more variables: bill_length_mm_Max <dbl>, bill_depth_mm_Mean <dbl>,
## #   bill_depth_mm_SD <dbl>, bill_depth_mm_Min <dbl>, bill_depth_mm_Max <dbl>,
## #   flipper_length_mm_Mean <dbl>, flipper_length_mm_SD <dbl>,
## #   flipper_length_mm_Min <int>, flipper_length_mm_Max <int>,
## #   body_mass_g_Mean <dbl>, body_mass_g_SD <dbl>, body_mass_g_Min <int>,
## #   body_mass_g_Max <int>

```

```

penguins %>%
  group_by(species, island) %>%
  summarise(across(contains("_"),
                    list(Mean = mean,
                         SD = sd,
                         Min = min,
                         Max = max),
                    na.rm = TRUE,
                    .names = "{.fn}_{.col}"))

```

```

## # A tibble: 5 x 18
## # Groups:   species [3]
##   species island Mean_bill_length... SD_bill_length... Min_bill_length...
##   <fct>    <fct>          <dbl>          <dbl>          <dbl>
## 1 Adelie  Biscoe             39.0             2.48             34.5
## 2 Adelie  Dream              38.5             2.47             32.1
## 3 Adelie  Torge...           39.0             3.03             33.5
## 4 Chinst... Dream              48.8             3.34             40.9
## 5 Gentoo  Biscoe             47.5             3.08             40.9
## # ... with 13 more variables: Max_bill_length_mm <dbl>, Mean_bill_depth_mm <dbl>,
## #   SD_bill_depth_mm <dbl>, Min_bill_depth_mm <dbl>, Max_bill_depth_mm <dbl>,
## #   Mean_flipper_length_mm <dbl>, SD_flipper_length_mm <dbl>,
## #   Min_flipper_length_mm <int>, Max_flipper_length_mm <int>,
## #   Mean_body_mass_g <dbl>, SD_body_mass_g <dbl>, Min_body_mass_g <int>,
## #   Max_body_mass_g <int>

```



```

penguins %>%
  group_by(species, island) %>%
  summarise(across(contains("_"),
                    list(Mean = mean,
                         SD = sd,
                         Min = min,
                         Max = max),
                    na.rm = TRUE)) %>%
  pivot_longer(contains("_"),
               names_to = c("body_part", "measure", "unit", "stat"),
               names_sep = "_") %>%
  pivot_wider(names_from = stat, values_from = value)

```

```

## # A tibble: 20 x 9
## # Groups:   species [3]
##   species island body_part measure unit Mean SD Min Max
##   <fct> <fct> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>
## 1 Adelie Biscoe bill length mm 39.0 2.48 34.5 45.6
## 2 Adelie Biscoe bill depth mm 18.4 1.19 16 21.1
## 3 Adelie Biscoe flipper length mm 189. 6.73 172 203
## 4 Adelie Biscoe body mass g 3710. 488. 2850 4775
## 5 Adelie Dream bill length mm 38.5 2.47 32.1 44.1
## 6 Adelie Dream bill depth mm 18.3 1.13 15.5 21.2
## 7 Adelie Dream flipper length mm 190. 6.59 178 208
## 8 Adelie Dream body mass g 3688. 455. 2900 4650
## 9 Adelie Torgersen bill length mm 39.0 3.03 33.5 46
## 10 Adelie Torgersen bill depth mm 18.4 1.34 15.9 21.5
## 11 Adelie Torgersen flipper length mm 191. 6.23 176 210

```

```

penguins %>%
  # pivot all the columns we want to summarise
  pivot_longer(contains("_"),
    names_to = c("body_part", "measure", "unit", "stat"),
    names_sep = "_",
    values_drop_na = TRUE) %>%
  # Group by wanted grouping variables, including names of columns we made above
  group_by(species, island, body_part, measure, unit) %>%
  # Summarise and give columns just function name
  summarise(across(value,
    list(Mean = mean,
         SD = sd,
         Min = min,
         Max = max),
    .names = "{.fn}"))

```

```

## # A tibble: 20 x 9
## # Groups:   species, island, body_part, measure [20]
##   species  island  body_part measure unit   Mean    SD   Min   Max
##   <fct>    <fct>    <chr>    <chr>  <chr> <dbl> <dbl> <dbl> <dbl>
## 1 Adelie  Biscoe    bill     depth  mm     18.4  1.19   16    21.1
## 2 Adelie  Biscoe    bill     length mm     39.0  2.48  34.5   45.6
## 3 Adelie  Biscoe    body     mass   g    3710. 488.   2850   4775
## 4 Adelie  Biscoe    flipper  length mm     189.   6.73  172    203
## 5 Adelie  Dream     bill     depth  mm     18.3  1.13  15.5   21.2
## 6 Adelie  Dream     bill     length mm     38.5  2.47  32.1   44.1
## 7 Adelie  Dream     body     mass   g    3688. 455.   2900   4650
## 8 Adelie  Dream     flipper  length mm     190.   6.59  178    208

```

nested data

```
penguins %>%  
  nest_by(species, island)
```

```
## # A tibble: 5 x 3  
## # Rowwise: species, island  
##   species island data  
##   <fct>    <fct>    <list<tbl_df[,6]>>  
## 1 Adelie   Biscoe      [44 x 6]  
## 2 Adelie   Dream       [56 x 6]  
## 3 Adelie   Torgersen   [52 x 6]  
## 4 Chinstrap Dream      [68 x 6]  
## 5 Gentoo   Biscoe     [124 x 6]
```

```
penguins %>%
  nest_by(species, island) %>%
  mutate(lm_model = list(
    lm(bill_length_mm ~ bill_depth_mm, data = data)
  ))
```

```
## # A tibble: 5 x 4
## # Rowwise: species, island
##   species island data lm_model
##   <fct>   <fct>   <list<tbl_df[,6]>> <list>
## 1 Adelie  Biscoe      [44 x 6] <lm>
## 2 Adelie  Dream      [56 x 6] <lm>
## 3 Adelie  Torgersen [52 x 6] <lm>
## 4 Chinstrap Dream    [68 x 6] <lm>
## 5 Gentoo  Biscoe    [124 x 6] <lm>
```

```

model_penguins <- penguins %>%
  nest_by(species, island) %>%
  mutate(
    lm_model = list(
      lm(bill_length_mm ~ bill_depth_mm, data = data)
    ),
    table = list(broom::tidy(lm_model))
  )
model_penguins

```

```

## # A tibble: 5 x 5
## # Rowwise: species, island
##   species island data lm_model table
##   <fct>    <fct>    <list<tbl_df[,6]>> <list>    <list>
## 1 Adelie   Biscoe      [44 x 6] <lm>      <tibble [2 x 5]>
## 2 Adelie   Dream      [56 x 6] <lm>      <tibble [2 x 5]>
## 3 Adelie   Torgersen  [52 x 6] <lm>      <tibble [2 x 5]>
## 4 Chinstrap Dream      [68 x 6] <lm>      <tibble [2 x 5]>
## 5 Gentoo   Biscoe     [124 x 6] <lm>      <tibble [2 x 5]>

```

```
model_penguins %>%
  unnest(data)
```

```
## # A tibble: 344 x 10
## # Groups:   species, island [5]
##   species island bill_length_mm bill_depth_mm flipper_length... body_mass_g
##   <fct>    <fct>         <dbl>         <dbl>         <int>         <int>
## 1 Adelie  Biscoe           37.8           18.3           174           3400
## 2 Adelie  Biscoe           37.7           18.7           180           3600
## 3 Adelie  Biscoe           35.9           19.2           189           3800
## 4 Adelie  Biscoe           38.2           18.1           185           3950
## 5 Adelie  Biscoe           38.8           17.2           180           3800
## 6 Adelie  Biscoe           35.3           18.9           187           3800
## 7 Adelie  Biscoe           40.6           18.6           183           3550
## 8 Adelie  Biscoe           40.5           17.9           187           3200
## 9 Adelie  Biscoe           37.9           18.6           172           3150
## 10 Adelie Biscoe           40.5           18.9           180           3950
## # ... with 334 more rows, and 4 more variables: sex <fct>, year <int>,
## #   lm_model <list>, table <list>
```

```
model_penguins %>%
  unnest(table) %>%
  select(-lm_model, -data)
```

```
## # A tibble: 10 x 7
## # Groups:   species, island [5]
##   species  island term          estimate std.error statistic  p.value
##   <fct>    <fct>  <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 Adelie   Biscoe  (Intercept)    21.0      5.24      4.01 2.41e- 4
## 2 Adelie   Biscoe  bill_depth_mm  0.977     0.285     3.43 1.35e- 3
## 3 Adelie   Dream   (Intercept)    18.3      4.66      3.93 2.40e- 4
## 4 Adelie   Dream   bill_depth_mm  1.10      0.255     4.33 6.46e- 5
## 5 Adelie   Torgersen (Intercept)    28.6      5.77      4.95 9.15e- 6
## 6 Adelie   Torgersen bill_depth_mm  0.562     0.313     1.80 7.81e- 2
## 7 Chinstrap Dream   (Intercept)    13.4      5.06      2.66 9.92e- 3
## 8 Chinstrap Dream   bill_depth_mm  1.92      0.274     7.01 1.53e- 9
## 9 Gentoo   Biscoe  (Intercept)    17.2      3.28      5.25 6.60e- 7
## 10 Gentoo  Biscoe  bill_depth_mm  2.02      0.219     9.24 1.02e-15
```