

dplyr

1.0.0

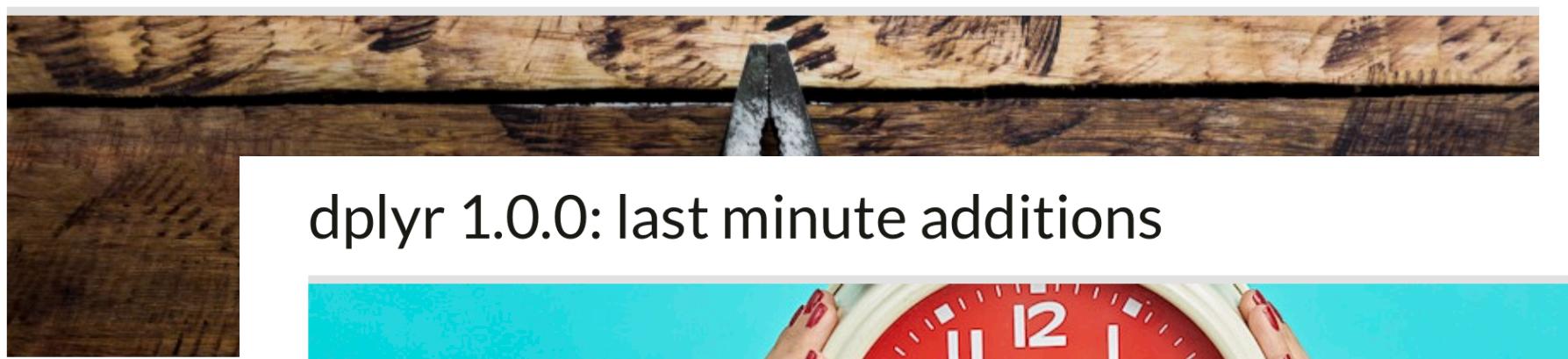
dplyr.tidyverse.org

enRum 2020

Romain

Romain Francois
@romain-francois

dplyr 1.0.0 available now!



2020/06/01

Hadley Wickham

I'm very excited to announce that dplyr 1.0.0 is now available on CRAN and GitHub. It's been a long time coming, and I'm grateful to all the contributors who have helped make it happen.

dplyr 1.0.0: last minute additions



dplyr 1.0.0 for package developers



2020/05/29

Hadley Wickham

dplyr 1.0.0 and vctrs



2020/04/29

Hadley Wickham

dplyr 1.0.0: working within rows



2020/04/27

Hadley Wickham

dplyr 1.0.0: working across columns



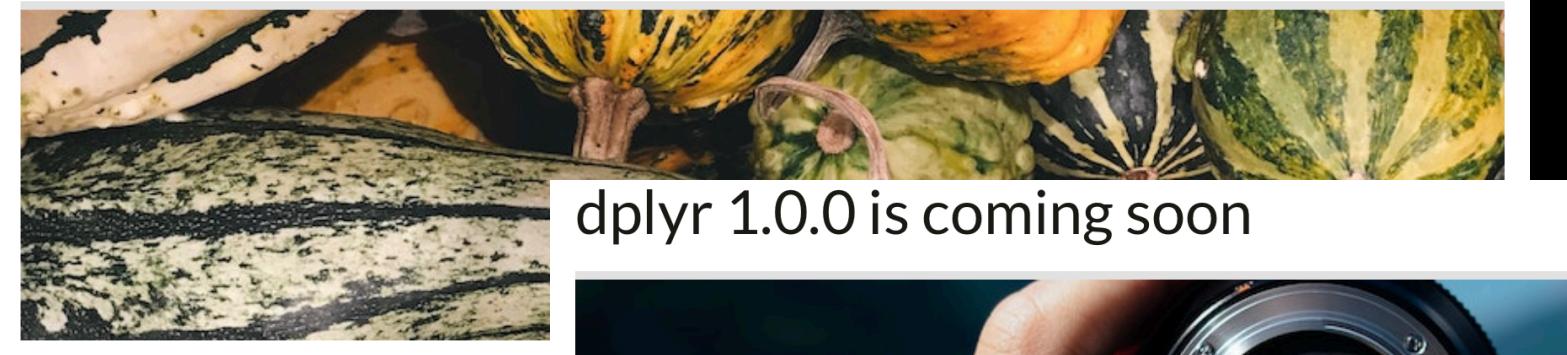
2020/04/10

Hadley Wickham

dplyr 1.0.0: select, rename, relocate



dplyr 1.0.0: new summarise() features



2020/03/27

Hadley Wickham

dplyr 1.0.0 is coming soon



2020/03/20

tidyverse.org/articles

summarised

CHINSTRAP!

GENTOO!

ADÉLIE!



@allison_horst

@allison_horst

CULMEN: RIDGE ALONG THE TOP PART OF A BIRD'S BILL



@allison_horst

Kristen Gorman



Kristen Gorman



Allison Horst

```
library(palmerpenguins)
glimpse(penguins)
#> Rows: 344
#> Columns: 7
#> $ species <fct> Adelie, Adelie, Adelie, Adelie, A
#> $ island <fct> Torgersen, Torgersen, Torgersen,
#> $ bill_length_mm <dbl> 39.1, 39.5, 40.3, NA, 36.7, 39.3,
#> $ bill_depth_mm <dbl> 18.7, 17.4, 18.0, NA, 19.3, 20.6,
#> $ flipper_length_mm <int> 181, 186, 195, NA, 193, 190, 181,
#> $ body_mass_g <int> 3750, 3800, 3250, NA, 3450, 3650,
#> $ sex <fct> male, female, female, NA, female,
```

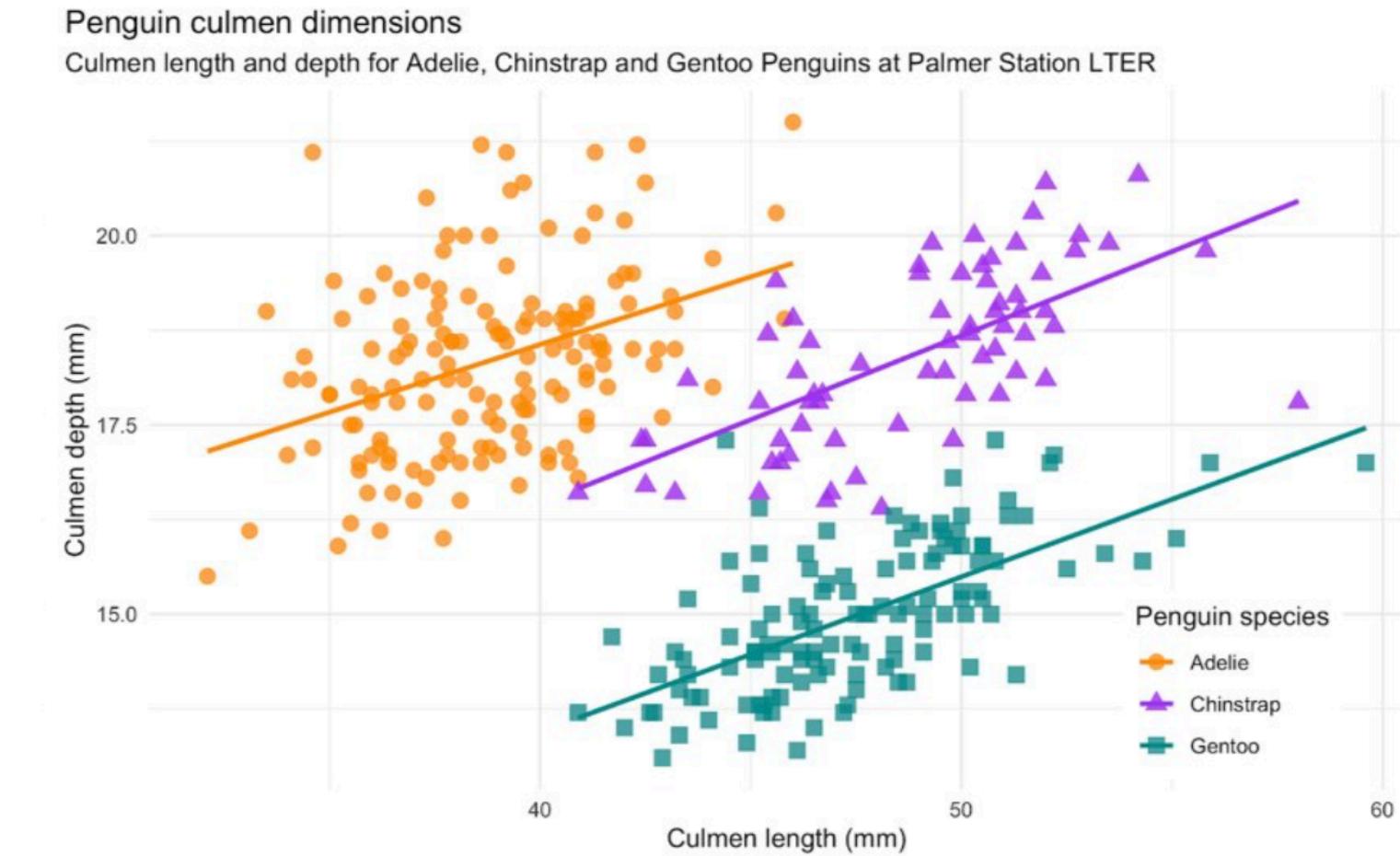
multiple rows

```
penguins %>%  
  group_by(species) %>%  
  summarise(  
    prob = c(.25, .75),  
    length = quantile(bill_length_mm, prob, na.rm = TRUE),  
    depth = quantile(bill_depth_mm, prob, na.rm = TRUE)  
)  
#> `summarise()` regrouping output by 'species' (override with '.groups' argument)  
#> # A tibble: 6 x 4  
#> # Groups:   species [3]  
#>   species     prob   length   depth  
#>   <fct>     <dbl>    <dbl>    <dbl>  
#> 1 Adelie     0.25    36.8    17.5  
#> 2 Adelie     0.75    40.8     19  
#> 3 Chinstrap  0.25    46.3    17.5  
#> 4 Chinstrap  0.75    51.1    19.4  
#> 5 Gentoo    0.25    45.3    14.2  
#> 6 Gentoo    0.75    49.6    15.7
```

multiple columns

```
penguins %>%
  group_by(species) %>%
  summarise(
    broom::tidy(lm(bill_depth_mm ~ bill_length_mm))
  )
```

```
#> `summarise()` regrouping output by 'species' (override with `groups` argument)
#> # A tibble: 6 x 6
#> # Groups:   species [3]
#>   species   term       estimate std.error statistic p.value
#>   <fct>     <chr>      <dbl>     <dbl>      <dbl>    <dbl>
#> 1 Adelie (Intercept)  11.4      1.34      8.52  1.61e-14
#> 2 Adelie bill_length_mm 0.179     0.0344     5.19  6.67e- 7
#> 3 Chinstrap (Intercept)  7.57     1.55      4.88  6.99e- 6
#> 4 Chinstrap bill_length_mm 0.222     0.0317     7.01  1.53e- 9
#> 5 Gentoo   (Intercept)  5.25      1.05      4.98  2.15e- 6
#> 6 Gentoo   bill_length_mm 0.205     0.0222     9.24  1.02e-15
```



across(<selection>, <action>)

```
penguins %>%  
  group_by(species) %>%  
  summarise(  
    across(starts_with("bill"), min, na.rm = TRUE)  
)  
#> `summarise()` ungrouping output (override with `groups` argument)  
#> # A tibble: 3 x 3  
#>   species    bill_length_mm    bill_depth_mm  
#>   <fct>          <dbl>            <dbl>  
#> 1 Adelie         32.1             15.5  
#> 2 Chinstrap      40.9             16.4  
#> 3 Gentoo         40.9             13.1
```

across(<selection>, <actions>)

```
penguins %>%  
  group_by(species) %>%  
  summarise(  
    across(starts_with("bill"),  
           list(min = min, max = max),  
           na.rm = TRUE  
    )  
  )  
#> `summarise()` ungrouping output (override with ` `.groups` argument)  
#> # A tibble: 3 x 5  
#>   species    bill_length_mm_min bill_length_mm_max bill_depth_mm_min bill_depth_mm_max  
#>   <fct>          <dbl>            <dbl>            <dbl>            <dbl>  
#> 1 Adelie        32.1             46               15.5            21.5  
#> 2 Chinstrap     40.9             58               16.4            20.8  
#> 3 Gentoo        40.9             59.6              13.1            17.3
```

```
summarise(.groups = )
```

What is the result grouped by?

```
penguins %>%
  group_by(species, island) %>%
  summarise(
    prob = c(.25, .75),
    length = quantile(bill_length_mm, prob, na.rm = TRUE),
    depth = quantile(bill_depth_mm, prob, na.rm = TRUE)
  )
```

> 1 rows

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(  
    prob = c(.25, .75),  
    length = quantile(bill_length_mm, prob, na.rm = TRUE),  
    depth = quantile(bill_depth_mm, prob, na.rm = TRUE)  
)
```

```
#> `summarise()` regrouping output by 'species', 'island' (override with `groups` argument)  
#> # A tibble: 10 x 5  
#> # Groups:   species, island [5]  
#>   species   island     prob  length  depth  
#>   <fct>     <fct>     <dbl>   <dbl>   <dbl>  
#> 1 Adelie     Biscoe     0.25    37.7    17.6  
#> 2 Adelie     Biscoe     0.75    40.7    19.0  
#> 3 Adelie     Dream      0.25    36.8    17.5  
#> 4 Adelie     Dream      0.75    40.4    18.8  
#> 5 Adelie     Torgersen  0.25    36.7    17.4  
#> 6 Adelie     Torgersen  0.75    41.1    19.2  
#> 7 Chinstrap  Dream      0.25    46.3    17.5  
#> 8 Chinstrap  Dream      0.75    51.1    19.4  
#> 9 Gentoo     Biscoe     0.25    45.3    14.2  
#> 10 Gentoo    Biscoe     0.75    49.6    15.7
```

What is the result grouped by?

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

~~====~~ 1 rows

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

```
#> `summarise()` regrouping output by 'species' (override with `groups` argument)
#> # A tibble: 5 x 4
#> # Groups:   species [3]
#>   species   island   length   depth
#>   <fct>     <fct>     <dbl>    <dbl>
#> 1 Adelie     Biscoe     39.0    18.4
#> 2 Adelie     Dream      38.5    18.3
#> 3 Adelie     Torgersen  39.0    18.4
#> 4 Chinstrap  Dream      48.8    18.4
#> 5 Gentoo    Biscoe     47.5    15.0
```

.groups = drop_last/drop/keep/rowwise

```
penguins %>%  
  group_by(species, island) %>%  
  summarise(  
    length = mean(bill_length_mm, na.rm = TRUE),  
    depth = mean(bill_depth_mm, na.rm = TRUE)  
)  
#> `summarise()` regrouping output by 'species' (override with `.`groups`  
argument)  
#> # A tibble: 5 x 4  
#> # Groups:   species [3]  
#>   species     island   length  depth  
#>   <fct>       <fct>     <dbl>   <dbl>  
#> 1 Adelie      Biscoe     39.0    18.4  
#> 2 Adelie      Dream      38.5    18.3  
#> 3 Adelie      Torgersen  39.0    18.4  
#> 4 Chinstrap   Dream      48.8    18.4  
#> 5 Gentoo     Biscoe     47.5    15.0
```



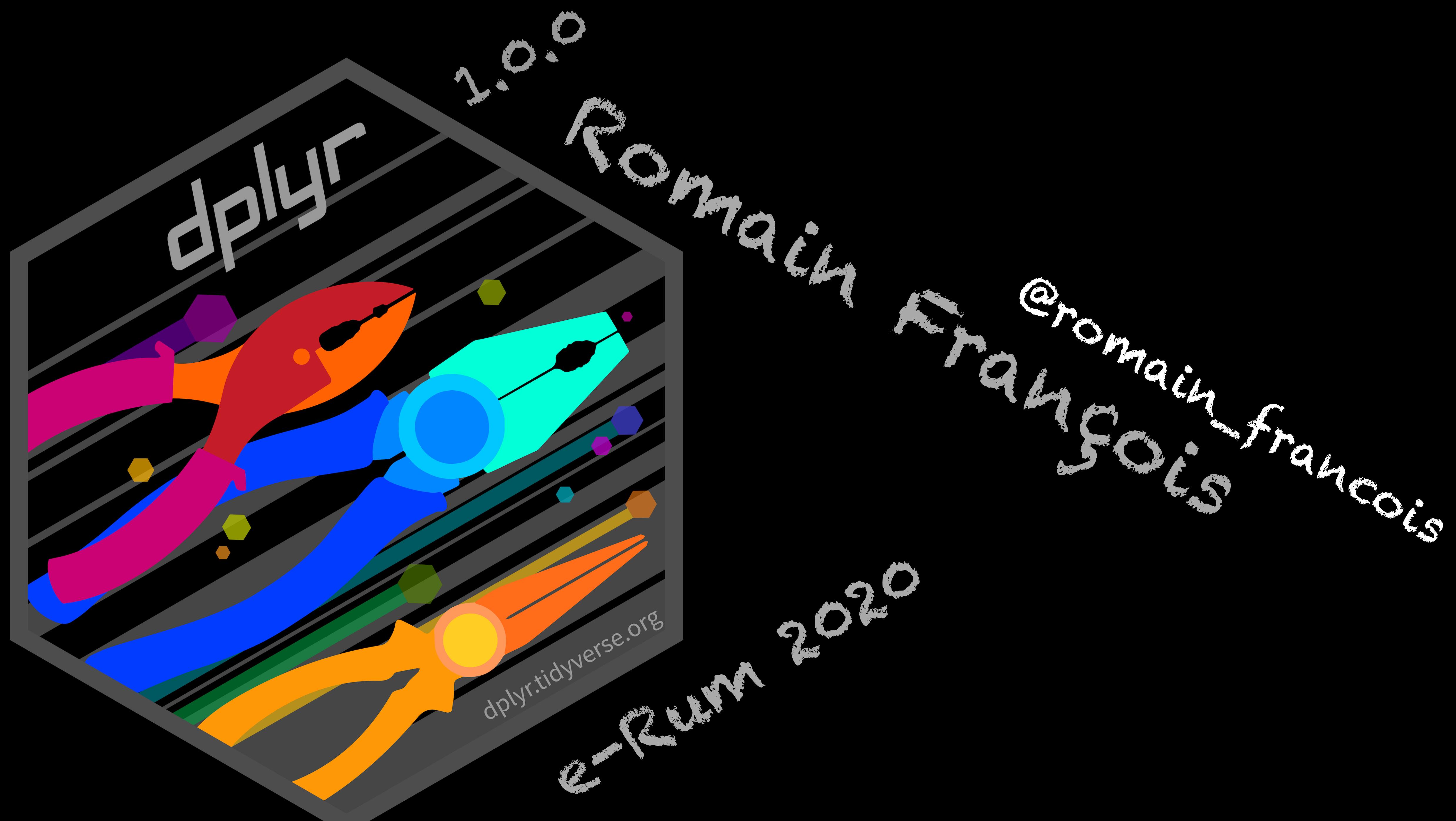
```
summarise(.groups = "drop")
```

```
penguins %>%
  group_by(species, island) %>%
  summarise(.groups = "drop",
            length = mean(bill_length_mm, na.rm = TRUE),
            depth = mean(bill_depth_mm, na.rm = TRUE))
#> # A tibble: 5 x 4
#>   species    island    length  depth
#>   <fct>      <fct>     <dbl>   <dbl>
#> 1 Adelie     Biscoe     39.0    18.4
#> 2 Adelie     Dream      38.5    18.3
#> 3 Adelie     Torgersen  39.0    18.4
#> 4 Chinstrap  Dream      48.8    18.4
#> 5 Gentoo    Biscoe     47.5    15.0
```

```
summarise(.groups = "keep")
```

```
penguins %>%
  group_by(species, island) %>%
  summarise(.groups = "keep",
            length = mean(bill_length_mm, na.rm = TRUE),
            depth = mean(bill_depth_mm, na.rm = TRUE))
)
```

```
#> # A tibble: 5 x 4
#> # Groups:   species, island [5]
#>   species     island    length  depth
#>   <fct>       <fct>     <dbl>   <dbl>
#> 1 Adelie      Biscoe     39.0    18.4
#> 2 Adelie      Dream      38.5    18.3
#> 3 Adelie      Torgersen  39.0    18.4
#> 4 Chinstrap   Dream      48.8    18.4
#> 5 Gentoo     Biscoe     47.5    15.0
```



dplyr

1.0.0

Romain

dplyr.tidyverse.org

enRum 2020

Romain Francois
@romain-francois