

Course 1 Section 2.14 - Tidy data

Jiaying Wu

14/09/2020

```
library(tidyverse)
```

```
kb <- read_csv("https://raw.githubusercontent.com/datascienceprogram/ids_course_data/master/koala_bilby")
kb
```

```
## # A tibble: 5 x 5
##   ID      koala_NSW koala_VIC bilby_NSW bilby_VIC
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 grey         23         43         11         8
## 2 cream        56         89         22        17
## 3 white        35         72         13         6
## 4 black        28         44         19        16
## 5 taupe        25         37         21        12
```

Exercise 1

Once you've read the data into your R session, gather the data into long form, naming the two new variables, label and count.

```
kb_long <- gather(kb, key = "label", value = "count", -ID)
kb_long
```

```
## # A tibble: 20 x 3
##   ID      label      count
##   <chr> <chr>      <dbl>
## 1 grey  koala_NSW    23
## 2 cream koala_NSW    56
## 3 white koala_NSW    35
## 4 black koala_NSW    28
## 5 taupe koala_NSW    25
## 6 grey  koala_VIC    43
## 7 cream koala_VIC    89
## 8 white koala_VIC    72
## 9 black koala_VIC    44
## 10 taupe koala_VIC    37
## 11 grey  bilby_NSW    11
## 12 cream bilby_NSW    22
## 13 white bilby_NSW    13
## 14 black bilby_NSW    19
## 15 taupe bilby_NSW    21
```

```
## 16 grey   bilby_VIC      8
## 17 cream bilby_VIC     17
## 18 white  bilby_VIC      6
## 19 black  bilby_VIC     16
## 20 taupe  bilby_VIC     12
```

Exercise 2

Separate the labels into two new variables: animal and state.

```
kb_separate <- separate(kb_long,
                        col = label,
                        into = c("animal", "state"),
                        sep = "_")
kb_separate
```

```
## # A tibble: 20 x 4
##   ID    animal state count
##   <chr> <chr>   <chr> <dbl>
## 1 grey   koala   NSW     23
## 2 cream koala   NSW     56
## 3 white  koala   NSW     35
## 4 black  koala   NSW     28
## 5 taupe  koala   NSW     25
## 6 grey   koala   VIC     43
## 7 cream koala   VIC     89
## 8 white  koala   VIC     72
## 9 black  koala   VIC     44
## 10 taupe koala   VIC     37
## 11 grey   bilby   NSW     11
## 12 cream bilby   NSW     22
## 13 white  bilby   NSW     13
## 14 black  bilby   NSW     19
## 15 taupe  bilby   NSW     21
## 16 grey   bilby   VIC      8
## 17 cream bilby   VIC     17
## 18 white  bilby   VIC      6
## 19 black  bilby   VIC     16
## 20 taupe  bilby   VIC     12
```

Exercise 3

Spread the long form data into wide form, where the columns are the states.

```
kb_wide <- spread(kb_separate, key = "state", value = "count")
kb_wide
```

```
## # A tibble: 10 x 4
##   ID    animal  NSW  VIC
##   <chr> <chr>   <dbl> <dbl>
## 1 black bilby    19    16
## 2 black koala    28    44
```

```
## 3 cream bilby      22    17
## 4 cream koala     56    89
## 5 grey  bilby      11     8
## 6 grey  koala      23    43
## 7 taupe bilby      21    12
## 8 taupe koala      25    37
## 9 white bilby      13     6
## 10 white koala     35    72
```

Exercise 4

Spread the long form data into wide form, where the columns are the animals.

```
kb_animal <- spread(kb_separate, key = "animal", value = "count")
kb_animal
```

```
## # A tibble: 10 x 4
##   ID    state bilby koala
##   <chr> <chr> <dbl> <dbl>
## 1 black NSW      19     28
## 2 black VIC      16     44
## 3 cream NSW      22     56
## 4 cream VIC      17     89
## 5 grey  NSW      11     23
## 6 grey  VIC       8     43
## 7 taupe NSW      21     25
## 8 taupe VIC      12     37
## 9 white NSW      13     35
## 10 white VIC       6     72
```