Working with tidy data in R: dplyr

Fundamental actions on data tables:

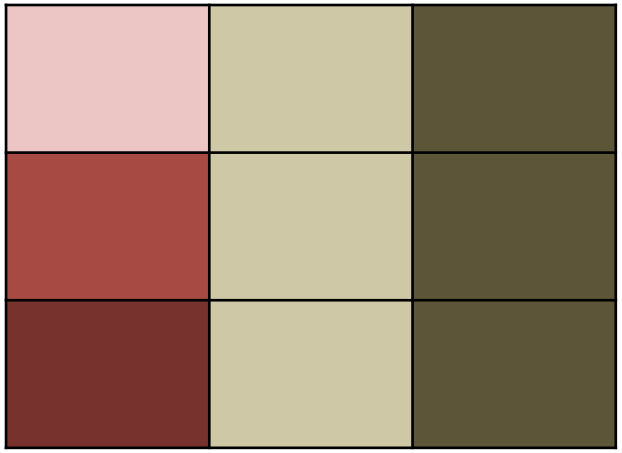
- choose rows — `filter()`
- choose columns — `select()`
- make new columns — `mutate()`
- arrange rows — `arrange()`
- calculate summary statistics — `summarize()`
- work on groups of data — `group_by()`

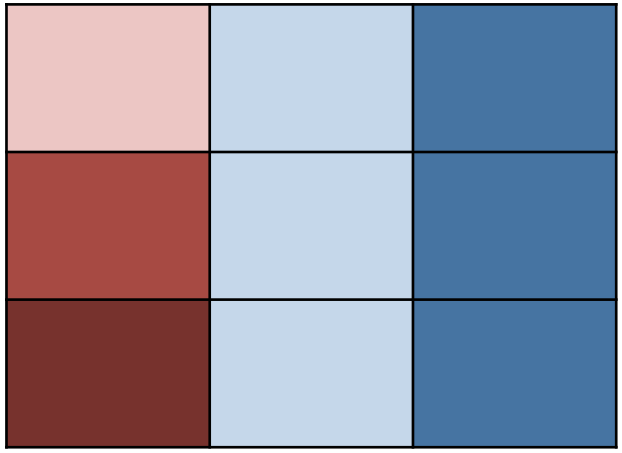
Working with tidy data in R: dplyr

Fundamental actions on data tables:

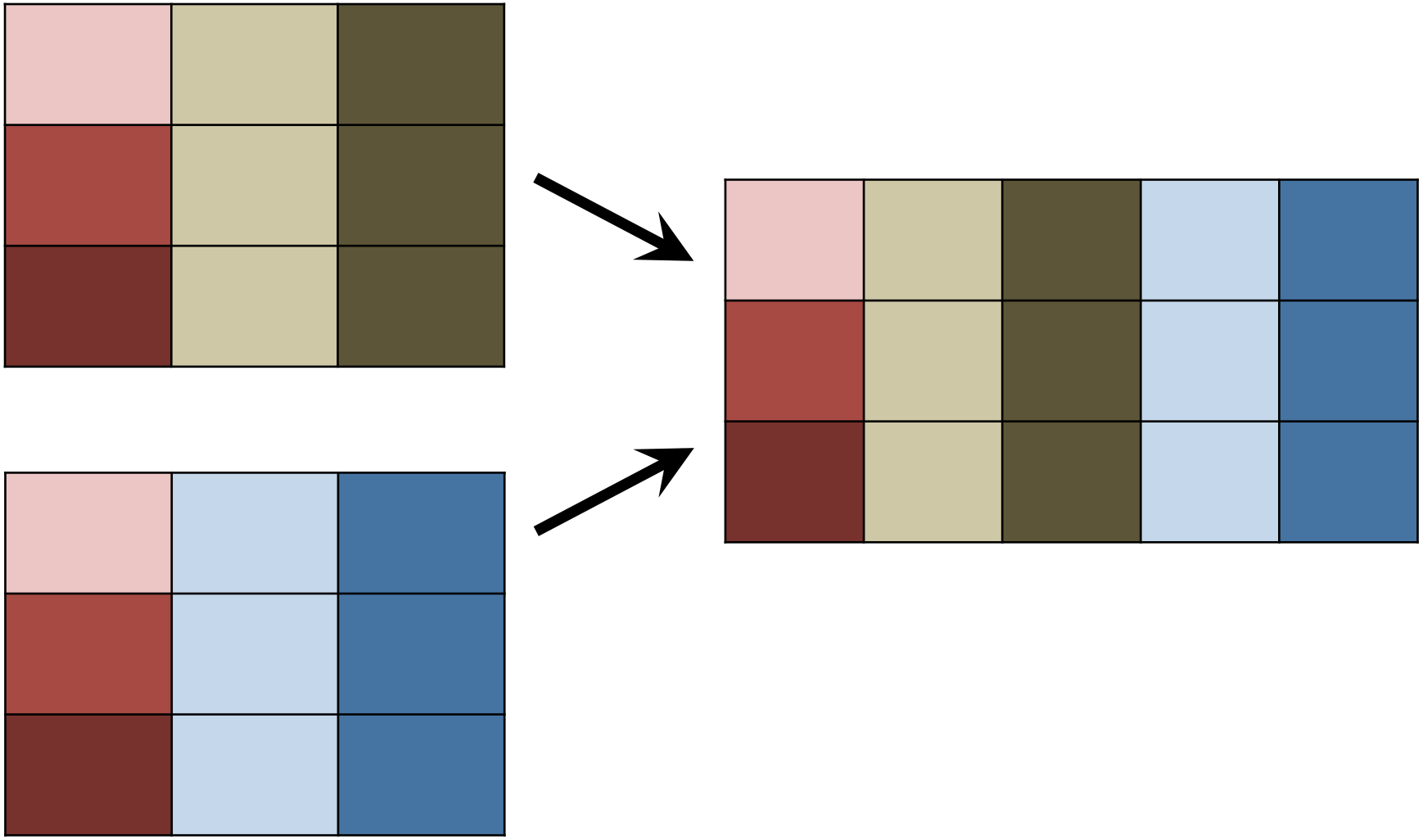
- choose rows — `filter()`
- choose columns — `select()`
- make new columns — `mutate()`
- arrange rows — `arrange()`
- calculate summary statistics — `summarize()`
- work on groups of data — `group_by()`
- combine tables — `left_join(), ...`

`left_join()`: combine two tables





`left_join()`: combine two tables



Example: Joining tables

Let’s extract two tables from msleep:

Example: Joining tables

Let's extract two tables from msleep:

```
> order_table <- select(msleep, name, order)
> order_table
                                    name           order
1                                Cheetah       Carnivora
2                             Owl monkey         Primates
3                         Mountain beaver         Rodentia
4      Greater short-tailed shrew    Soricomorpha
5                                    Cow    Artiodactyla
6                     Three-toed sloth          Pilosa
7                    Northern fur seal       Carnivora
8                         Vesper mouse         Rodentia
9                                    Dog       Carnivora
10                            Roe deer    Artiodactyla
```

Example: Joining tables

Let's extract two tables from msleep:

```
> awake_table <- select(msleep, name, awake)
> awake_table
                                    name awake
1                                Cheetah 11.90
2                             Owl monkey  7.00
3                         Mountain beaver  9.60
4          Greater short-tailed shrew  9.10
5                                    Cow 20.00
6                       Three-toed sloth  9.60
7                     Northern fur seal 15.30
8                           Vesper mouse 17.00
9                                    Dog 13.90
10                             Roe deer 21.00
```

Example: Joining tables

And put them back together:

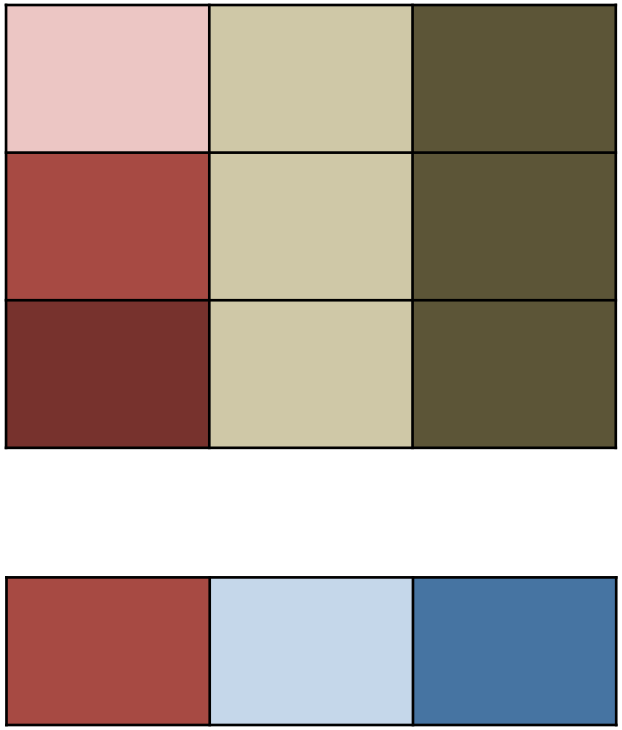
```
> left_join(order_table, awake_table)
```

Example: Joining tables

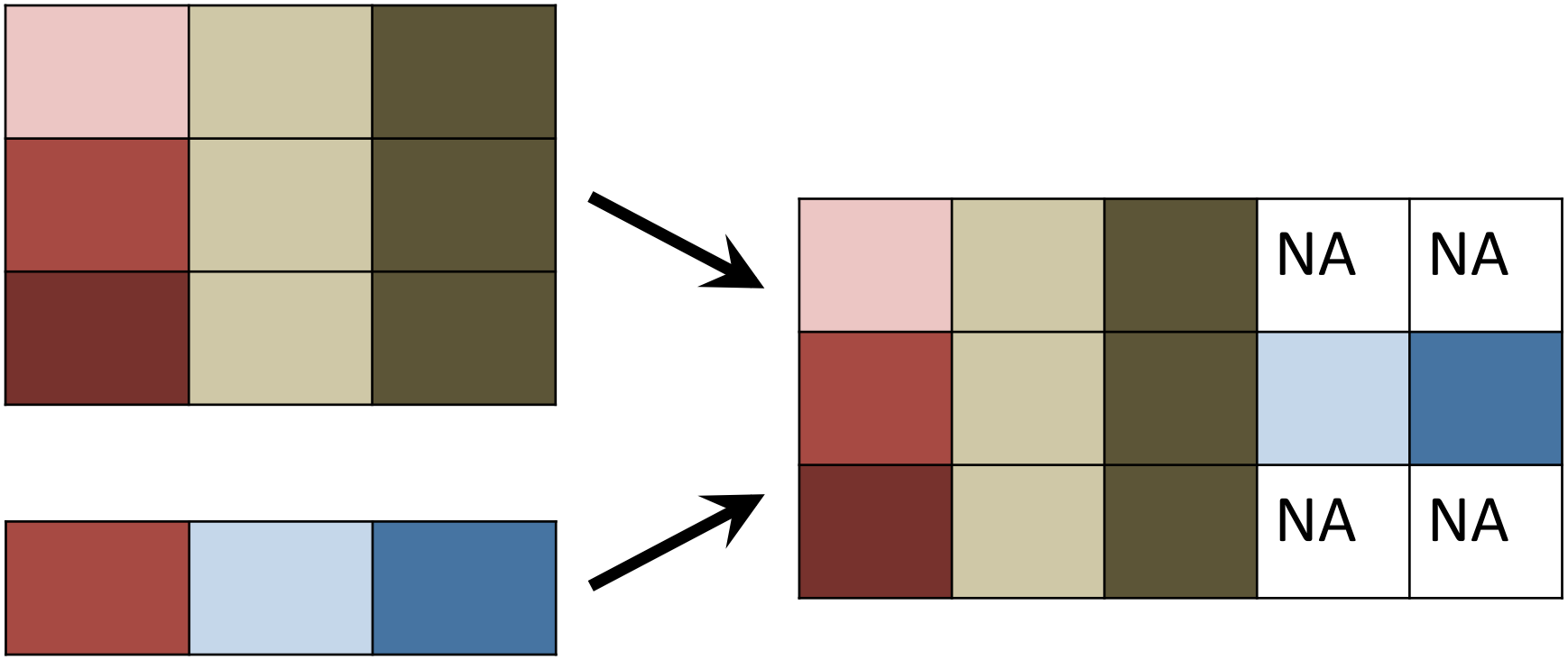
And put them back together:

```
> left_join(order_table, awake_table)
Joining by: "name"
                                  name          order awake
1                              Cheetah      Carnivora 11.90
2                           Owl monkey       Primates  7.00
3                       Mountain beaver       Rodentia  9.60
4      Greater short-tailed shrew    Soricomorpha  9.10
5                                  Cow    Artiodactyla 20.00
6                    Three-toed sloth         Pilosa  9.60
7                  Northern fur seal      Carnivora 15.30
8                        Vesper mouse       Rodentia 17.00
9                                  Dog      Carnivora 13.90
10                            Roe deer    Artiodactyla 21.00
```

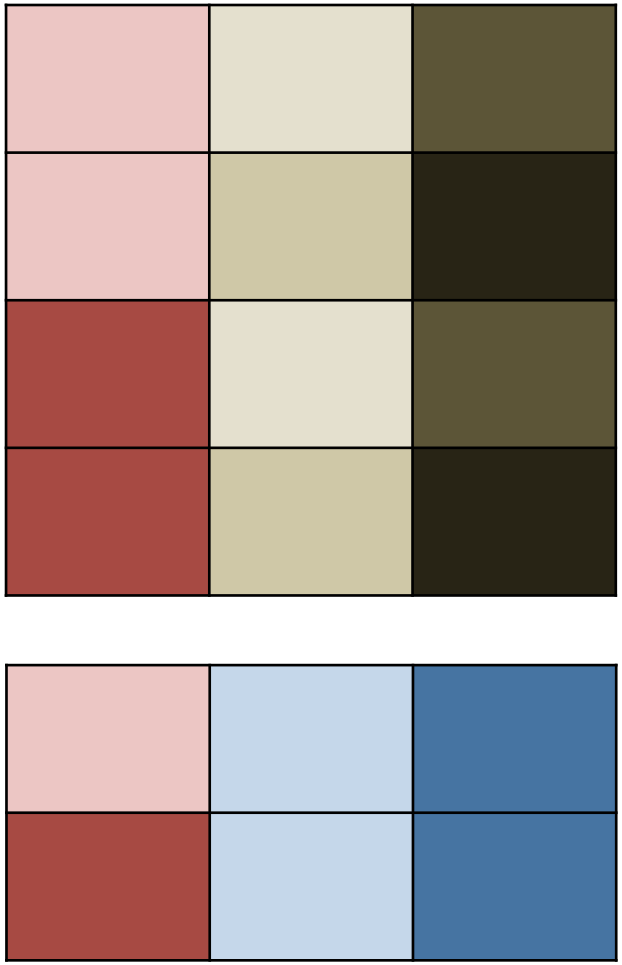
`left_join()`: missing values in 2nd table are set to NA



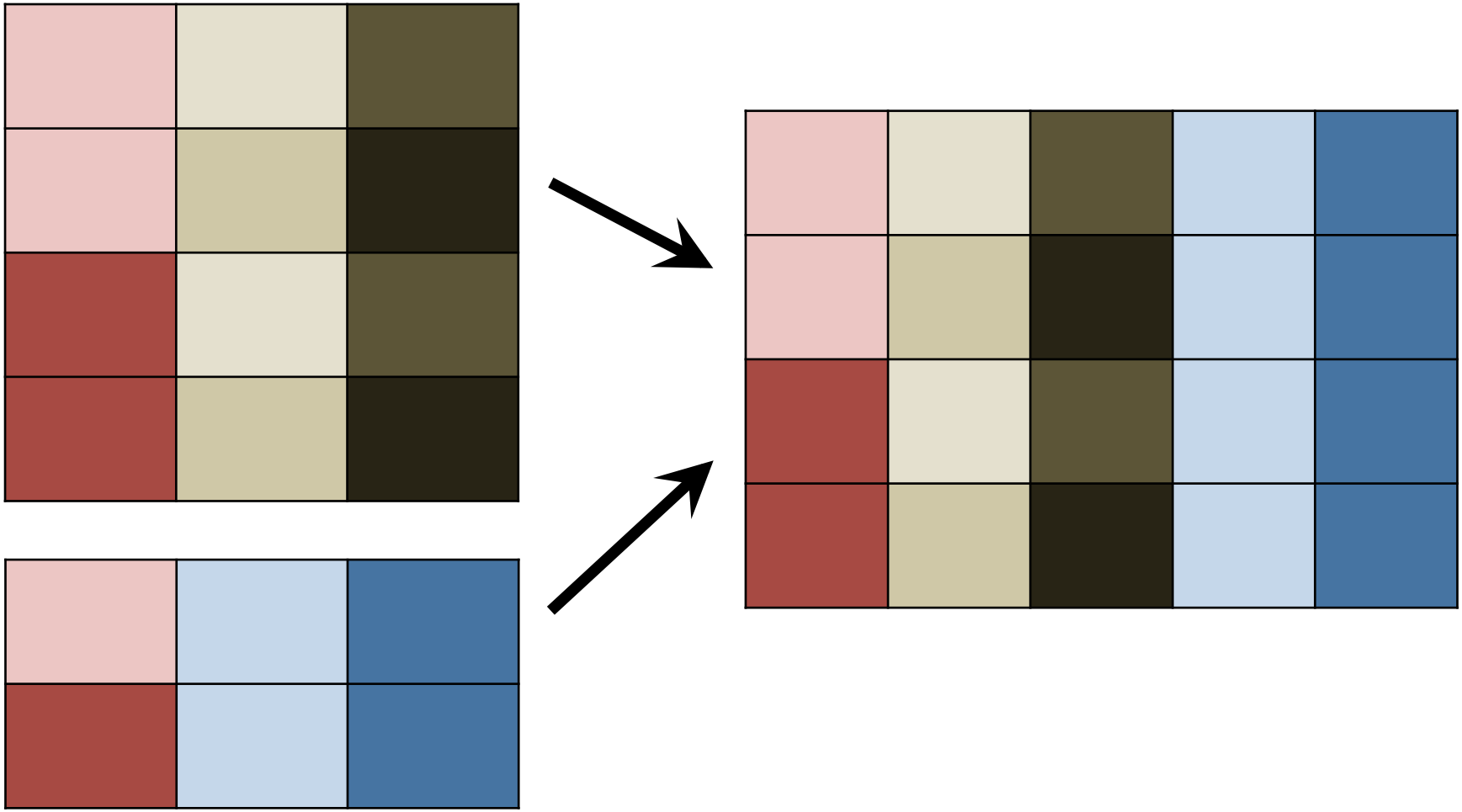
`left_join()`: missing values in 2nd table are set to NA



`left_join()`: values from 2nd table are duplicated where necessary



`left_join()`: values from 2nd table are duplicated where necessary



Several different join functions are available

- `left_join()`
- `right_join()`
- `inner_join()`
- `semi_join()`
- `full_join()`
- `anti_join()`