

Working with tidy data in R: dplyr

Fundamental actions on data tables:

- select rows — `filter()`
- select 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:

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

filter (): pick rows

[illegible]

filter(): pick rows

[illegible]

`filter()`: pick rows



select (): pick columns

[illegible]

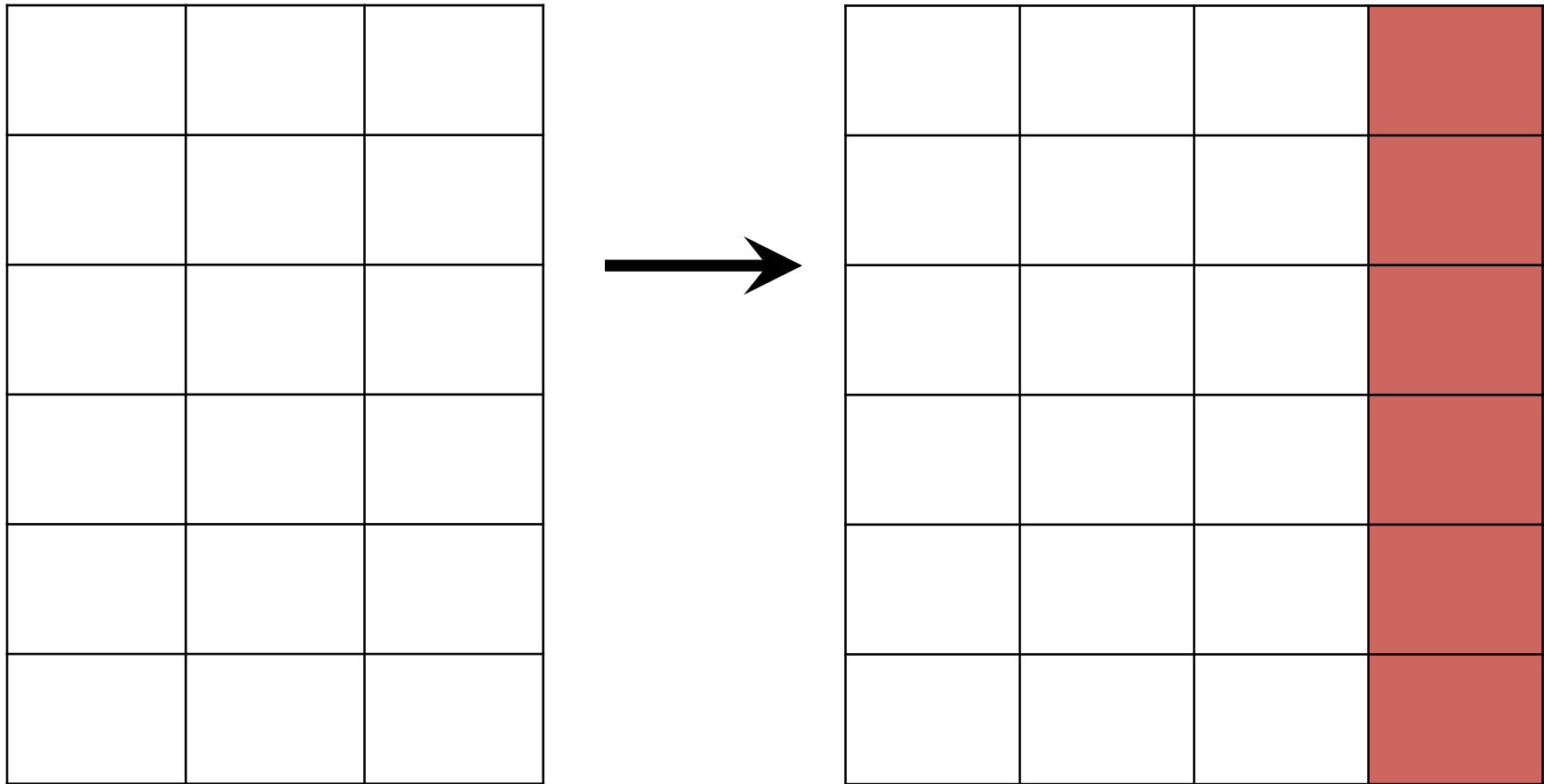
`select()`: pick columns



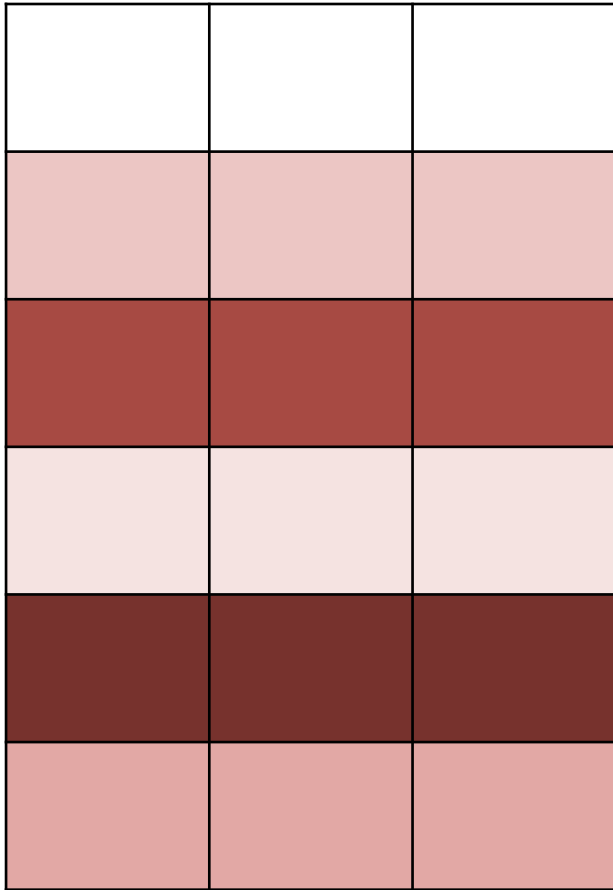
mutate(): make new columns

[illegible]

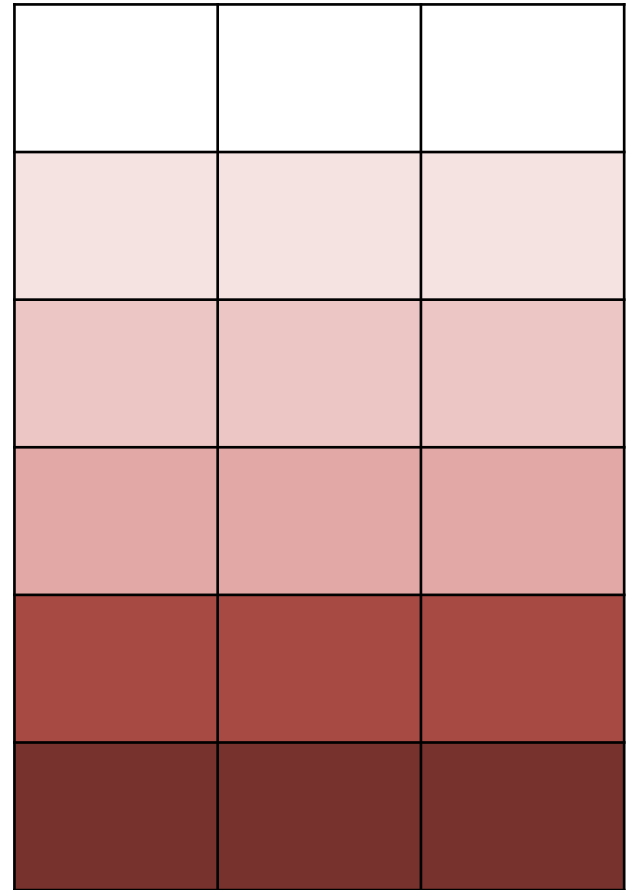
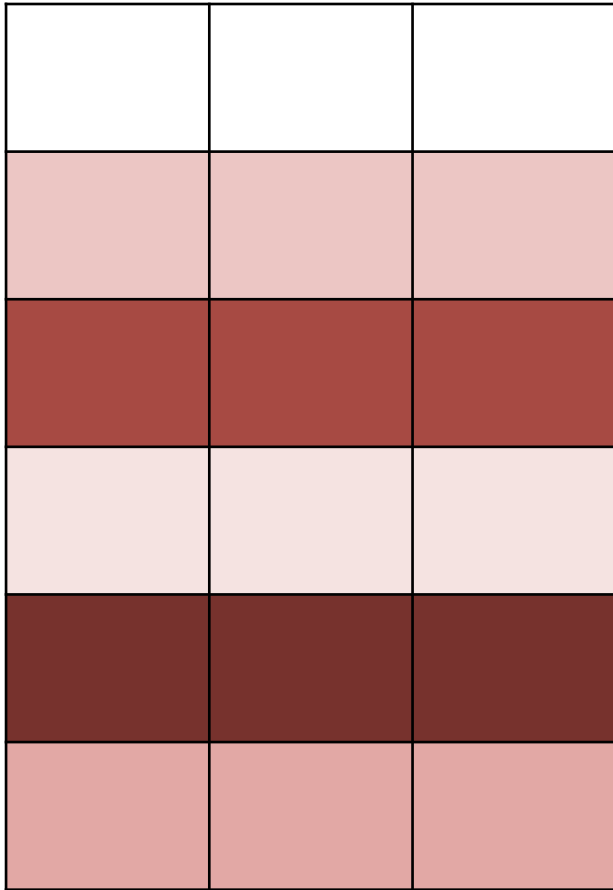
mutate(): make new columns



arrange () : change column order



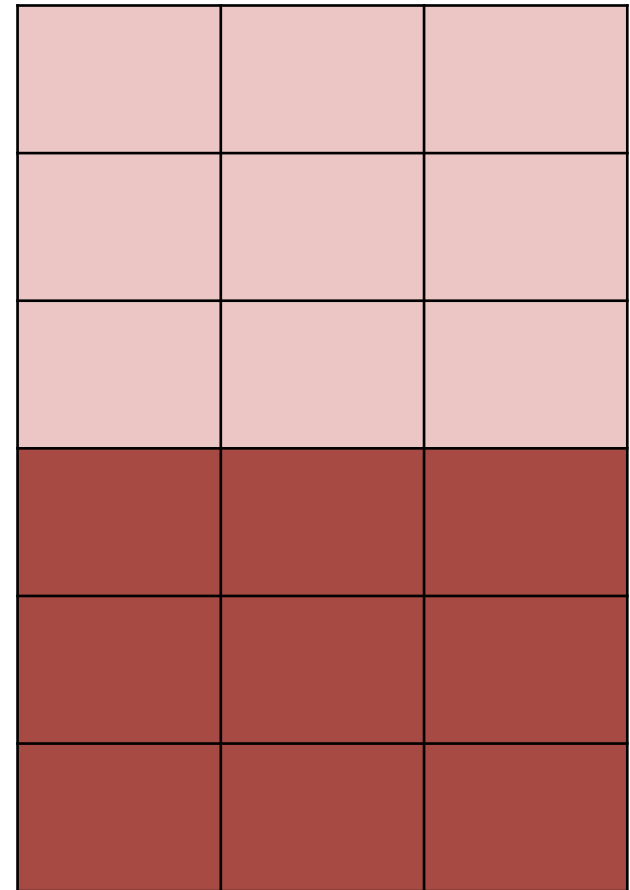
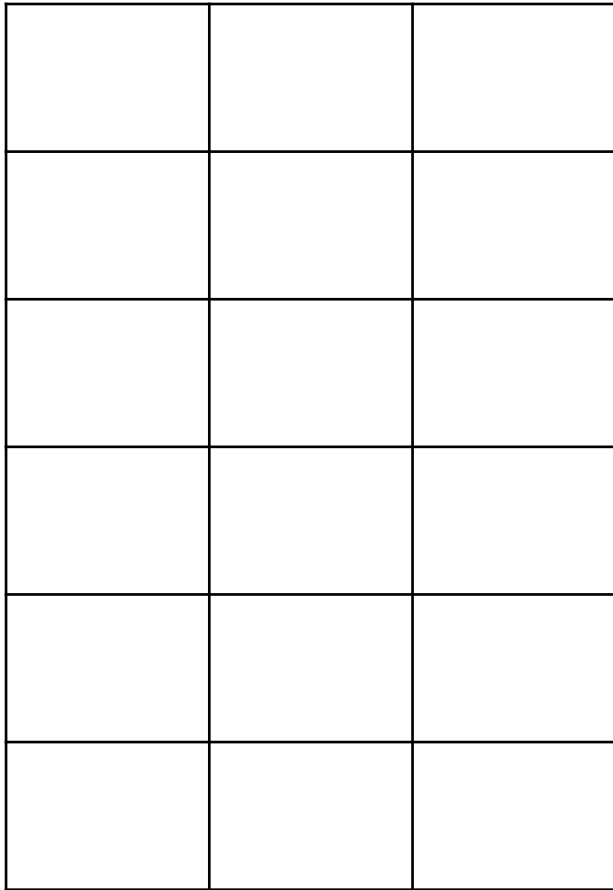
`arrange()`: change column order



group_by () : set up groupings

[illegible]

group_by(): set up groupings



summarize(): collapse multiple rows

[illegible]

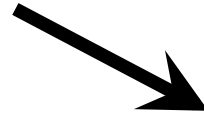
`summarize()`: collapse multiple rows



`left_join()`: combine two tables

`left_join()`: combine two tables

Light Pink	Light Tan	Dark Brown
Red	Light Tan	Dark Brown
Dark Red	Light Tan	Dark Brown



Light Pink	Light Blue	Dark Blue
Red	Light Blue	Dark Blue
Dark Red	Light Blue	Dark Blue



Light Pink	Light Tan	Dark Brown	Light Blue	Dark Blue
Red	Light Tan	Dark Brown	Light Blue	Dark Blue
Dark Red	Light Tan	Dark Brown	Light Blue	Dark Blue

dplyr example 1: What is total day time for each animal in `msleep`?

```
> msleep
```

dplyr example 1: What is total day time for each animal in msleep?

```
> msleep %>%  
  mutate(total_day_time = awake + sleep_total)
```

dplyr example 1: What is total day time for each animal in msleep?

```
> msleep %>%  
  mutate(total_day_time = awake + sleep_total) %>%  
  select(name, total_day_time)
```

dplyr example 1: What is total day time for each animal in msleep?

```
> msleep %>%  
  mutate(total_day_time = awake + sleep_total) %>%  
  select(name, total_day_time)
```

	name	total_day_time
1	Cheetah	24.00
2	Owl monkey	24.00
3	Mountain beaver	24.00
4	Greater short-tailed shrew	24.00
5	Cow	24.00
6	Three-toed sloth	24.00
7	Northern fur seal	24.00
8	Vesper mouse	24.00
9	Dog	24.00
10	Roe deer	24.00

dplyr example 2: What is the median awake time of different orders in `msleep`?

```
> msleep
```

dplyr example 2: What is the median awake time of different orders in msleep?

```
> msleep %>% group_by(order)
```


dplyr example 2: What is the median awake time of different orders in msleep?

```
> msleep %>% group_by(order) %>%  
  summarize(med_awake=median(awake))
```

dplyr example 2: What is the median awake time of different orders in msleep?

```
> msleep %>% group_by(order) %>%  
  summarize(med_awake=median(awake)) %>%  
  arrange(med_awake)
```

dplyr example 2: What is the median awake time of different orders in msleep?

```
> msleep %>% group_by(order) %>%  
  summarize(med_awake=median(awake)) %>%  
  arrange(med_awake)
```

Source: local data frame [19 x 2]

	order	med_awake
1	Chiroptera	4.20
2	Didelphimorphia	5.30
3	Cingulata	6.25
4	Afrosoricida	8.40
5	Pilosa	9.60
6	Rodentia	11.10
7	Diprotodontia	11.60
8	Soricomorpha	13.70
9	Carnivora	13.75
10	Erinaceomorpha	13.80

dplyr example 3: Joining tables

Let's extract two tables from msleep:

dplyr example 3: 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

dplyr example 3: 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

dplyr example 3: Joining tables

And put them back together:

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

dplyr example 3: 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