



R in 3 Months: Week 7 **(Tidy Data)**



Agenda

1. Housekeeping
2. Group work
3. More about Tidy Data
4. Next Week



Housekeeping

- Charlie's office hours on Friday instead of Wednesday



Questions on Tidy Data

- Blayne's question on ungrouping



Group Work



Group Work

Copy the code from

<https://gist.github.com/dgkeyes/1d5ceb8b607266f1d197a218dd5c6882> into a new R script file. Then tidy the `gapminder_wide` data.



Tidy Data



Tidy Data is more than just the `pivot_*()` functions

If you're interested in the philosophy behind tidy data then [read Hadley Wickham's 2014 paper](#) that introduced the concept.

The most **concise** description of tidy data is also provided by Hadley Wickham:

"Tidy datasets are all alike, but every messy dataset is messy in its own way"

Hadley Wickham



Other functions for tidying data



separate()

The `separate()` function is useful when a column contains **multiple variables**.

```
location_data <- tribble(  
  ~id, ~address,  
  1, "Las Vegas, USA",  
  2, "Bristol, UK",  
  3, "Kassala, Sudan"  
)  
location_data
```

id	address
<dbl>	<chr>
1	Las Vegas, USA
2	Bristol, UK
3	Kassala, Sudan

3 rows



separate()

```
location_data %>%  
  separate(col = address,  
           into = c("city",  
                    "country"),  
           sep = ", ")
```

id city		country
<dbl> <chr>		<chr>
1	Las Vegas	USA
2	Bristol	UK
3	Kassala	Sudan
3 rows		



separate_rows()

The `separate_rows()` function is useful when a column contains **multiple observations of the same variable**.



Which of these devices do you own? Please select all that apply.

```
device_ownership <- tribble(
  ~name, ~devices_owned,
  "Charlie", "Smart TV, Cell phone",
  "Mohammad", "Cell phone",
  "Christina", "Smart TV, Games Console, Cell phone"
)
device_ownership
```

name	devices_owned
<chr>	<chr>
Charlie	Smart TV, Cell phone
Mohammad	Cell phone
Christina	Smart TV, Games Console, Cell phone
3 rows	



```
device_ownership %>%  
  separate_rows(devices_owned,  
                sep = ", ")
```

name	devices_owned
<chr>	<chr>
Charlie	Smart TV
Charlie	Cell phone
Mohammad	Cell phone
Christina	Smart TV
Christina	Games Console

1-5 of 6 rows

Previous **1** [2](#) [Next](#)



Data Tidying Example

If we have time (otherwise, David will record a video)



Next Week

There are two big topics in the next week:

1. Reducing repetitive code with user-defined functions
2. Joining together datasets by shared columns (i.e. data merging)

These skills will equip you to deal with more complex and larger scale data wrangling

Project assignment: Create a function that can help you do your work more efficiently