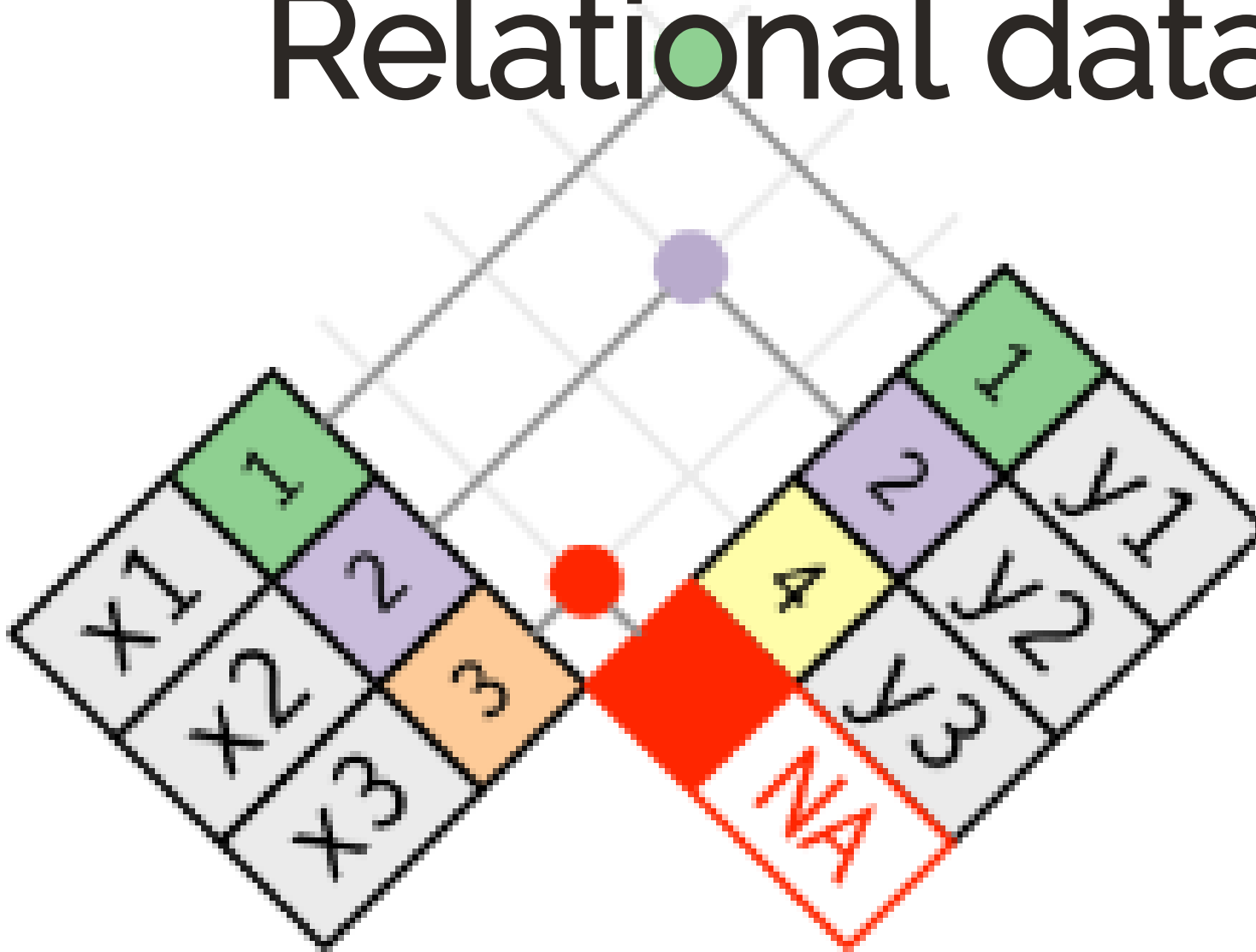# Relational data

Andrew Jones | Strategy Unit

Relational data

Rare to find all necessary data in a single table.

A common "key" variable is found across two (or more) tables, allowing you to link tables by observation.

Use joins in SQL (or VLOOKUP in Excel)

Relational data

Here, we'll focus on left (outer) joins. The syntax is similar for other types of join.

left_join

Keep structure of table_1

`table_1 %>%`

`left_join(table_2, by = "x")`

...and match to rows in table_2

Variable common to both tables

Relational Data

We're going to join two tables – one with cases of tuberculosis by country, one with population by country. From this new table we can derive a rate.

cases   *pop*

W.H.O. data

Keep the original structure of the cases dataframe
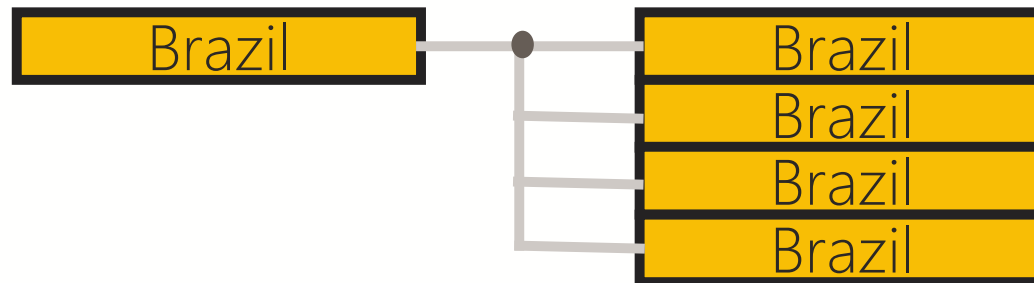
```
cases %>%

    left_join(pop, by = "country")
```

*...then match
to rows in pop*

*based on
"country"
value*

Duplicates!

```
cases %>%
  left_join(pop, by = "country")
```


For every value of Brazil in tb_cases, there are 4 in pop...

Joining on multiple rows

match on two variables

```
cases %>%

    left_join(pop, by = c("country" , "year"))
```

c stands for 'combine'

Joining with different names

Two tables have different name for same variable:

```
tb_cases %>%

    left_join(bad_names,

    by = c("country" = "Place" , "year" = "Yr"))
```

name in cases

name in bad_names

Some other dplyr joins

a			b	
x1	**x2**		**x1**	**x3**
A	1		A	T
B	2	**+**	B	F
C	3		D	T

=

x1	**x3**	**x2**
A	T	1
B	F	2
D	T	NA

dplyr::**right_join(a, b, by = "x1")**
Join matching rows from a to b.

x1	**x2**	**x3**
A	1	T
B	2	F

dplyr::**inner_join(a, b, by = "x1")**
Join data. Retain only rows in both sets.

x1	**x2**	**x3**
A	1	T
B	2	F
C	3	NA
D	NA	T

dplyr::**full_join(a, b, by = "x1")**
Join data. Retain all values, all rows.

Image taken from: https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

This work is licensed as

Creative Commons

Attribution-ShareAlike 4.0

International

To view a copy of this license, visit

https://creativecommons.org/licenses/by-sa/4.0/

End