Handout – Working with Geo-Data

NaWi-Workshop: Obtaining, linking and plotting geographic data

Markus Konrad

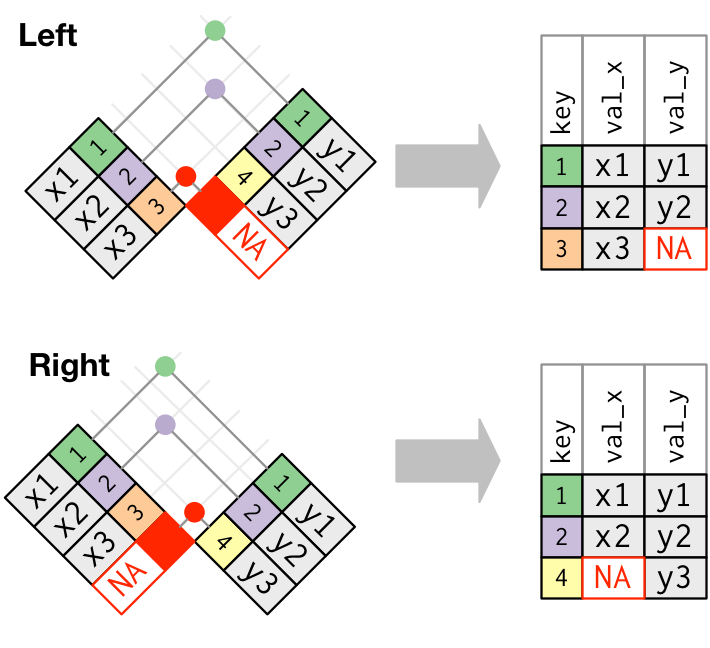
??? (TODO), 2019

# 1. Data linkage with *dplyr*

## Left and right (outer) joins

*Left and right outer joins* keep all observations on the left-hand or right-hand side data sets respectively. Unmatched rows are filled up with *NA*s:

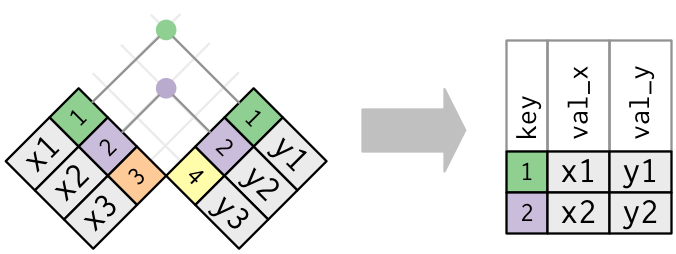
Syntax: inner\_join(a, b, by = <criterion>)



Left and right join. Source: Grolemund, Wickham 2017: R for Data Science

## Inner joins

An *inner join* matches keys that appear in both data sets and returns the combined observations:



Inner join. Source: Grolemund, Wickham 2017: R for Data Science

Syntax: inner\_join(a, b, by = <criterion>)

## Specifying matching criteria

Parameter by can be:

1. a character string specifying the key for both sides, e.g.: inner\_join(pm, city\_coords, by = 'city') will match city column in pm with city column in city\_coords;
2. a vector of character strings specifying several keys to match both sides, e.g.: inner\_join(pm, city\_coords, by = c('city', 'country') will match those rows, where city *and* country columns match;
3. a *named* character string vector like inner\_join(pm, city\_coords, by = c('cityname' = 'id'), which will match the column cityname in pm with the column id in city\_coords