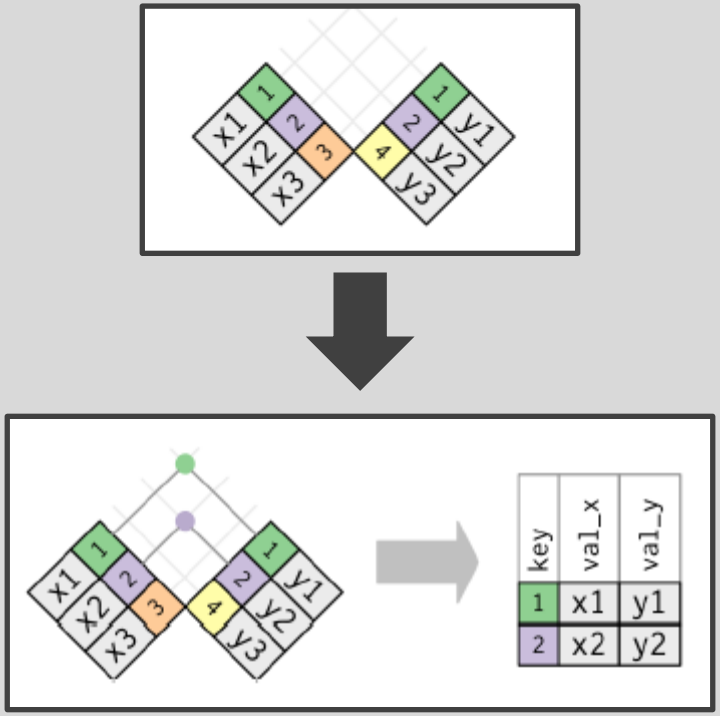
Joins II

Graphical Depiction

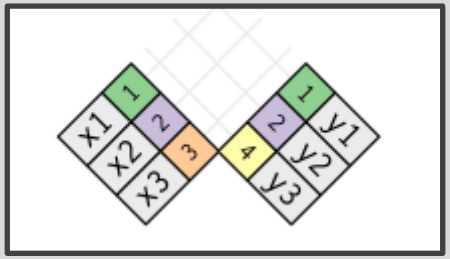
- Inner Join

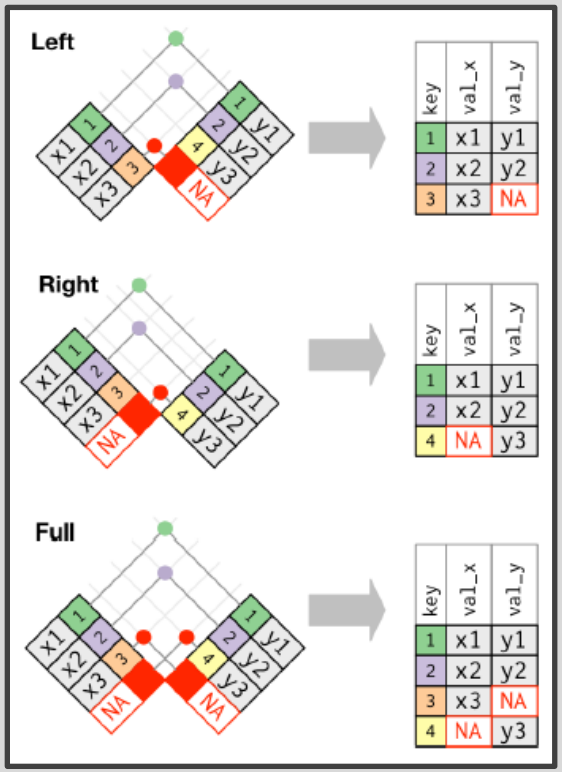


- Notes on Inner Join
  - Generally Inappropriate
  - Easy to Lose Observations

Graphical Depiction

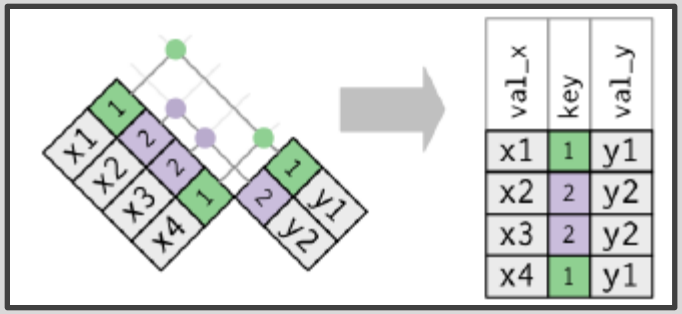
- Outer Join



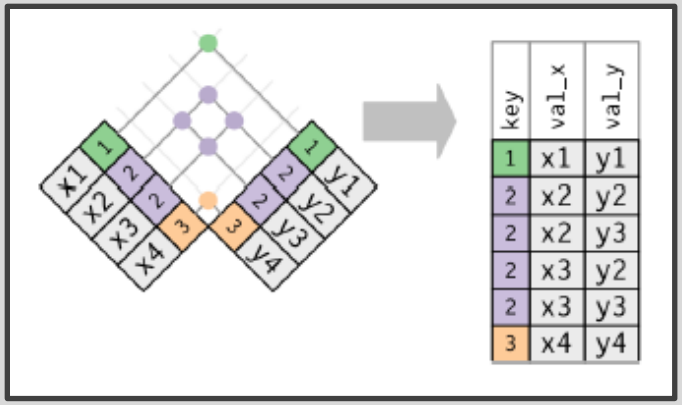


Graphical Depiction

- Duplicate Keys
  - Using Left Join
  - One Table Has Duplicates



  - Both Tables Have Duplicates



Joins Assignment

- Instructions
  - Download Analysis 2 File
  - Open Analysis 2 Rmd File
  - Knit to HTML
  - Read Introduction
- Three Part Assignment
  - Part 1: Scrape GDP Data
  - Part 2: Scrape Additional Data
  - Part 3: Merge Datasets
- All Data in HTML Tables

Part 1: GDP

- GDP Data by Country

#	Country	GDP (nominal, 2017)	GDP (abbrev.)	GDP growth	Population (2017)	GDP per capita	Share of World GDP
1	United States	$19,485,394,000,000	$19.485 trillion	2.27%	325,084,756	$59,939	24.08%
2	China	$12,237,700,479,375	$12.238 trillion	6.90%	1,421,021,791	$8,612	15.12%
3	Japan	$4,872,415,104,315	$4.872 trillion	1.71%	127,502,725	$38,214	6.02%
4	Germany	$3,693,204,332,230	$3.693 trillion	2.22%	82,658,409	$44,680	4.56%
5	India	$2,650,725,335,364	$2.651 trillion	6.68%	1,338,676,785	$1,980	3.28%
6	United Kingdom	$2,637,866,340,434	$2.638 trillion	1.79%	66,727,461	$39,532	3.26%

- Webscrape Data Using rvest
- Clean Data Using
  - str_replace()
  - str_replace_all()
  - Special Character: “$”

Part 2: More Data

- Education Data

Rank	Country	Education Index	Expected years of schooling	Mean years of schooling	HDI rank	Continent[6]
1	Australia	0.939	20.4	13.2	2	Oceania
2	Denmark	0.923	19.2	12.7	5	Europe
3	New Zealand	0.917	19.2	12.5	13	Oceania
4	Norway	0.916	17.7	12.7	1	Europe
5	Germany	0.914	17.1	13.2	4	Europe
6	Ireland	0.910	18.6	12.3	8	Europe
7	Iceland	0.906	19.0	12.2	9	Europe

- Coronavirus Data

COUNTRY	CONFIRMED	DEATHS	CASE-FATALITY	DEATHS/100K POP.
San Marino	723	42	5.8%	124.32
Peru	768,895	31,369	4.1%	98.06
Belgium	105,226	9,955	9.5%	87.16
Andorra	1,681	53	3.2%	68.83
Bolivia	131,453	7,693	5.9%	67.76
Spain	682,267	30,904	4.5%	66.14
Brazil	4,591,364	138,105	3.0%	65.93
Chile	448,523	12,321	2.7%	65.79

Part 3: Merges

- Two Different Merges
  - Education Into GDP
  - GDP Into COVID
- Use Appropriate Join -> “Into”
- Answer All Follow-Up Questions
- Fix Specific Instances

```
> ifelse(VAR==“OLD”, “NEW”, VAR)
```

Closing

Disperse and Make Reasonable Decisions