

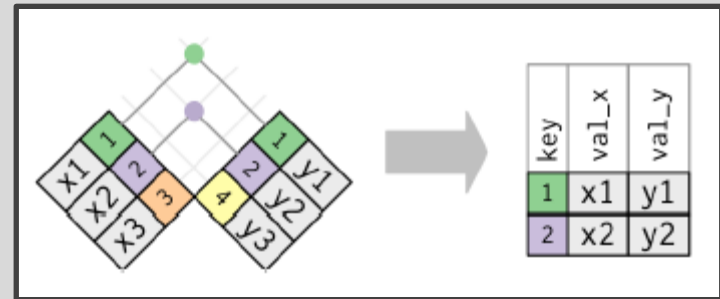
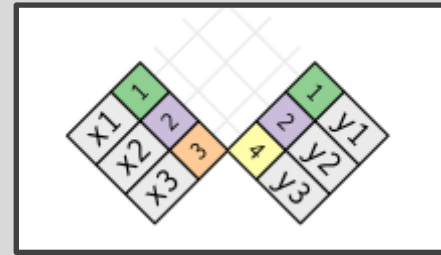


Joins II

Graphical Depiction



- Inner Join

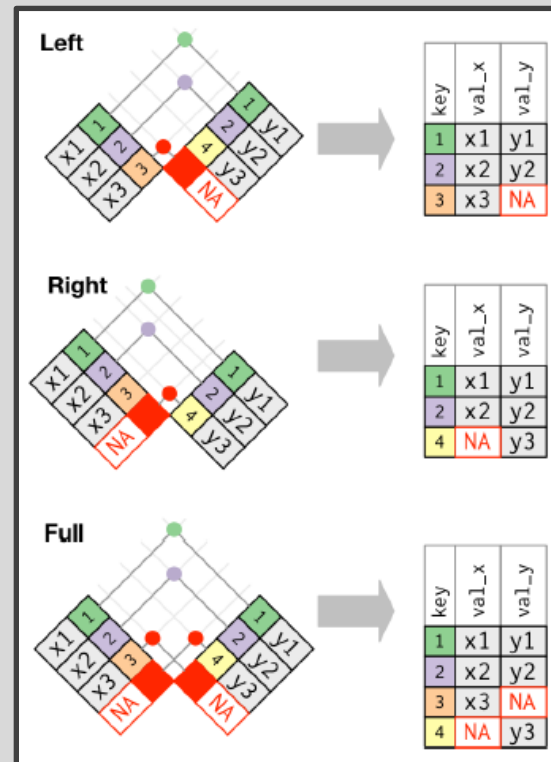
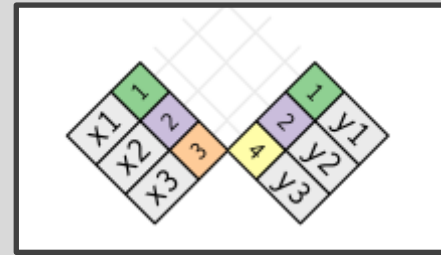


- Notes on Inner Join
 - Generally Inappropriate
 - Easy to Lose Observations

Graphical Depiction



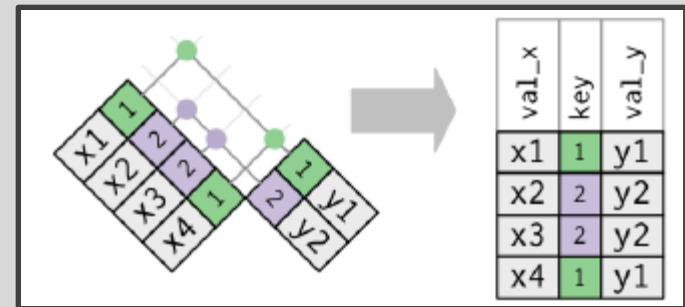
- Outer Join



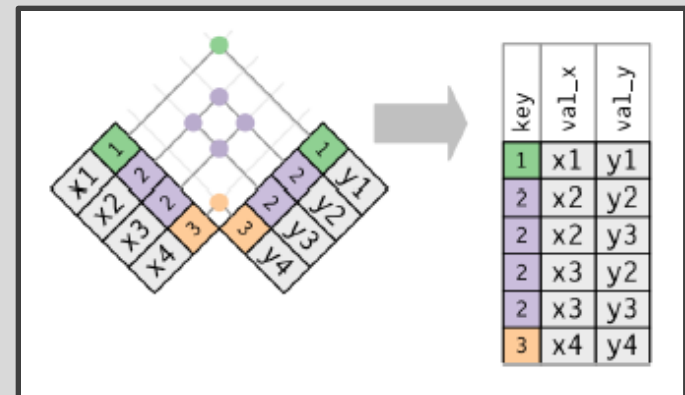
Graphical Depiction



- 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

Data [\[edit \]](#)

Worldwide education indexes are provided by the UNDP's Human Development Report derived from the UNESCO Institute for Statistics.

Country	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Afghanistan	0.122	0.133	0.145	0.156	0.168	0.179	0.19	0.202	0.213	0.224	0.235
Albania	0.583	0.588	0.557	0.542	0.528	0.55	0.557	0.569	0.579	0.584	0.586
Algeria	0.385	0.395	0.405	0.414	0.424	0.431	0.443	0.458	0.473	0.487	0.5
Andorra											0.63
Angola										0.276	0.288
Antigua and Barbuda											
Argentina	0.628	0.632	0.634	0.636	0.639	0.648	0.659	0.67	0.681	0.715	0.736
Armenia	0.637	0.64	0.64	0.647	0.64	0.631	0.633	0.646	0.666	0.674	0.671
Australia	0.873	0.874	0.874	0.875	0.875	0.894	0.895	0.895	0.895	0.895	0.895
Austria	0.675	0.682	0.692	0.699	0.704	0.709	0.711	0.715	0.72	0.727	0.731
Azerbaijan						0.618	0.614	0.616	0.625	0.633	0.64
Bahamas											0.7
Bahrain	0.574	0.586	0.586	0.599	0.609	0.619	0.624	0.628	0.632	0.635	0.643

- 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