

Wait! There is more ... What about Privacy ?

Due to privacy concerns, **you are not allowed to see all the identifying information during the user study**

We have attempted to convey sufficient information for the record linkage task

The upcoming pages will give you an understanding of how the data is disclosed on a need to know basis

Checkmarks: Identical values



Identical values are represented by checkmarks

1	1234567891	∞	BRIAN	TIPTON	∞	9/9/1960	W
	1234561291	25	BRIANNA	TIPTON	∞	9/9/1960	W



1	*****@*	∞	*****	✓	∞	✓	✓
	*****&*	25	*****&	✓	∞	✓	✓

Stars (*) are used for Similar Values

When two items are similar, the characters that are the **same are represented by stars (*)**. Only the characters that are different are revealed.

1	1234567891	∞	BRIAN	TIPTON	∞	9/9/1960	W
	1234561291	25	BRIANNA	TIPTON	∞	9/9/1960	W



1	*****@**	∞	*****	✓	∞	✓	✓
	*****&*	25	*****&	✓	∞	✓	✓

@ and & for different characters

When two items have one of insert/replace/transpose operations, the differentiating characters are represented by @ and &s

1	1234567891	∞	BRIAN	TIPTON	∞	9/9/1960	W
	1234561291	25	BRIANNA	TIPTON	∞	9/9/1960	W



1	*****@@**	∞	*****	✓	∞	✓	✓
	*****&&**	25	*****&&	✓	∞	✓	✓

@@@ and &&& when a value is missing


When one of the values in a pair is missing, the other one is represented by either @@@ or &&& depending on the row in which the value is present


7	0000018335	∞	PATSY	CALLAHAN	...	11/13/1948	B
	?	∞	PATSY	CALLAHAN	...	?	B




7	aaaaaaaaaa	∞	✓	✓	...	aa/aa/aaaa	✓
	?	∞	✓	✓	...	?	✓

13	6556368585 DIFF	①	WILL DIFF	GREENE	∞	07/03/1950	B DIFF
	1092091430	①	DAVE	GREENE	∞	07/03/1950	W

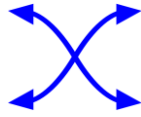






13	aaaaaaaaaa DIFF	①	aaaaa	✓	∞	✓	@ DIFF
	bbbbbbbbbb	①	bbbb	✓	∞	✓	&

Swaps



When there are swaps, the swapped values are represented by &&& and @@@@

5	0000006947	①	BRYANT		MADELINE	①	02/05/1962	W
	0000006947	25	MADELINE		BRYANT	∞	05/02/1962	W



5	✓	①	@@@@@		@@@@@@@	①	@@/@@/****	✓
	✓	25	@@@@@@@		@@@@@	∞	@@/@@/****	✓

Understand the mapping?

Take a moment to look at the next slide carefully and understand how the data is getting mapped.

Pair	ID	FFreq	First name	Last name	LFreq	DoB (M/D/Y)	Race
1	8000002767	①	JUDE	WILLIAM	①	09/09/1906	W
	8000003567	①	JUDE	WILLIAM JR	①	09/09/1960	DIFF B
2	0000006947	①	BRYANT	MADELINE	①	05/02/1962	W
	0000006947	25	MADELINE	BRYANT	∞	05/02/1962	W
3	9000018540	...	SALLY	BYRD	∞	07/04/1960	W
	6000008928	∞	JOHN	BYRD	∞	04/07/1960	?

Pair	ID	FFreq	First name	Last name	LFreq	DoB (M/D/Y)	Race
1	*****@*	①	✓	*****	①	**/**/**@	@
	*****&*	①	✓	*****&	①	**/**/**&	DIFF &
2	✓	①	aaaaaa	aaaaaaaa	①	✓	✓
	✓	25	aaaaaaaa	aaaaaa	∞	✓	✓
3	aaaaaaaaaaaa	...	aaaaa	✓	∞	@@/@@/****	@
	aaaaaaaaaaaa	∞	aaaa	✓	∞	&&/@@/****	?



Try doing a few practice problems again as you would see them in the real study