

INTRODUCTION TO VISUAL MARKUP

Common Issues with Data about People

Watch out for common issues

Data are expressed differently

- Nick Names (Elizabeth & Beth)

Data change over time

- Women get married and change their last name

Data are not unique attributes

- John Smith (there are different people that have the same name)
- Twins & Family members have similar identifying information such as DOB & last name
- Same names in Families with different suffix (Jr and Sr)

Data are sometimes missing

- SSN are often missing

Data have errors

- Inserting/deleting extra characters
- Typing in the wrong character
- Transposing two characters
- First name and last name are mixed up
- Day and month is mixed up

Missing Values



Data are sometimes missing.

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
7	0000018335	PATSY	CALLAHAN	11/13/1948	F	B
	?	PATSY	CALLAHAN	?	F	B




Added or Deletions Characters

Insertion (or deletion) of characters are common typing errors

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
1	8000001276	JAYDEN	TIPTON	09/09/1960	M	W
	8000002768	JADEN	TIPTON	09/09/1960	M	W

Different Characters

Mistyping can lead to certain characters replacing others

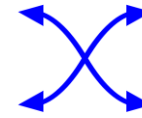
Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
3	9000018540 	SAL	BYRD 	04/07/1960 	F	W
	9000018870	SAL	BIRD	04/09/1960	F	W

Switched Characters

Two characters can be interchanged by mistake

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
11	1719582520	ROGRES	HYLEMON	07/15/1924	M	W
	1719852520	ROGERS	HYLEMON	07/15/1942	M	W

Column Swaps



Sometimes whole values are swapped as well:

Date Swaps

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
9	0000020502	SAMANTHA	MORGAN	02/11/1958	F	W
	0000020502	SAMANTHA	MORGAN	11/02/1958	F	W




Name Swaps

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
5	0000006947	BRYANT	MADELINE	09/22/1926	F	W
	0000006947	MADELINE	BRYANT	09/22/1926	F	W

Different



This icon is shown if the values in a column are very different.

Pair	ID	First name	Last name	DoB (M/D/Y)	Sex	Race
13	6556368585 	WILL 	GREENE	07/03/1950	M	B 
	1092091430	DAVE	GREENE	07/03/1950	M	W

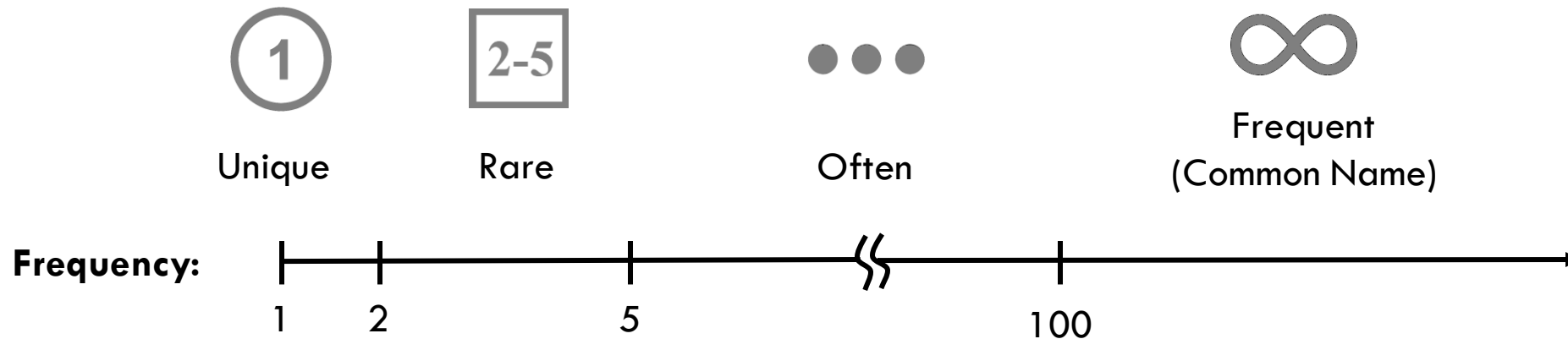
Michelle Williams ?

Name Frequency

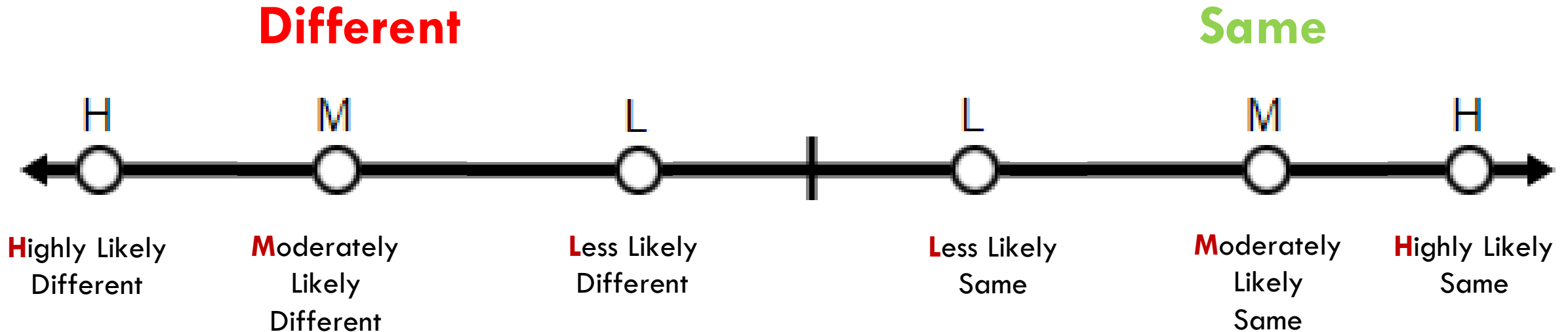


It would not be surprising for two people to have the same **common name**, but it might be unlikely for two people to have the same **rare names**.

Frequency icons indicate how many times a given name occurred in the data source



The Answer Buttons



If you think the rows are the **same person**, click one of the choices on the **right side**. Pick one of L, M, H depending on your confidence level.

If you think the rows are for **different people**, click one of the choices on the **left side**. Pick one of L, M, H depending on your confidence level.