### **Inner Join**

# Method 3: inner\_join()

Joining columns based on an *id* matching from two different dataframes. The resulting dataframe only has the rows in which there is an *id* in both dataframes, if not the rows are dropped. The id variable must be a character.

id		id			id		
:4.422	+	id_123		=	id 122		
id_123		id_563			id_123		
id_789		id_956			id_789		
:4 425		id_235					
id_435		id_789					

# inner\_join()

There is a dataframe named **basic\_info\_teen\_fl\_df** for teenagers that have the cd and downloaded versions of Game 1.

```
basic_info_teen_f1_df
## # A tibble: 3 x 4
##
       id
            group_age
                               gender
                        age
## <chr>
              <chr>
                        <dbl>
                                <chr>
## 1 id 333
                         17
                                Female
             teenage
## 2 id_122
             teenage
                         13
                                Female
## 3 id_530
                         14
                                Male
             teenage
```

There is another dataframe named **basic\_info\_teen\_f2\_df** for teenages that have of Game 2.

```
basic_info_teen_f2_df
## # A tibble: 8 x 3
##
       id
           number_of_tournaments
                                        level
## <chr>
                     <dbl>
                                        <chr>
                       23
## 1 id_721
                                       competent
## 2 id_815
                        9
                                       expert
                       2
## 3 id_122
                                       expert
## 4 id_530
                       12
                                       novice
## 5 id_782
                       10
                                       competent
## 6 id_295
                       24
                                       novice
## 7 id_386
                       1
                                       novice
## 8 id_308
                       17
                                       novice
```

We can inner join **basic\_info\_teen\_fl\_df** and **basic\_info\_teen\_f2\_df** by using the following function and specifying a 'by' variable, id, where order just places one dataframe in front of another:

```
inner_join(basic_info_teen_f1_df,basic_info_teen_f2_df, by = 'id')
## # A tibble: 2 x 6
##
       id
                                           number_of_tournaments
                                                                      level
              group_age
                                  gender
                          age
## <chr>
                          <dbl>
                                  <chr>
                                                   <dbl>
                <chr>
                                                                     <chr>
                                                    2
## 1 id_122
               teenage
                           13
                                  Female
                                                                     expert
                           14
## 2 id_530
               teenage
                                  Male
                                                    12
                                                                     novice
```

This also could be done using the pipe function:

```
basic info teen f1 df %>%
inner_join(basic_info_teen_f2_df, by = 'id')
## # A tibble: 2 x 6
##
                                           number_of_tournaments
       id
              group_age
                          age
                                  gender
                                                                      level
## <chr>
                <chr>
                          <dbl>
                                  <chr>>
                                                   <dbl>
                                                                     <chr>
## 1 id_122
                           13
                                  Female
                                                    2
               teenage
                                                                     expert
## 2 id_530
                                                   12
              teenage
                           14
                                  Male
                                                                     novice
```

### Question about Inner Join

There are two dataframes named, **basic\_info\_adult\_f1\_df** and **basic\_info\_adult\_f2\_df**. Combine the rows of these dataframes based on the common ids, provide an appropriate name for the dataframe, and specify the variable types in the combined dataframe.

### **Left Join**

## Method 4: left\_join()

Joining columns based on an *id* matching from two different dataframes, however left\_join keeps all rows from the left dataframe. The resulting dataframe only has the rows from the left dataframe and information from the right dataframe where the id matches. When the *id* from both dataframes does not match, NA is given. The id variable must be a character.

id		id			id		
11.400		id_123		=	:4 122		
id_123	+	id_563			id_123		
id_789		id_956			id_789		
:4 425		id_235			id_435	NA	NA
id_435		id_789					

# left\_join()

There is a dataframe named **basic\_info\_teen\_fl\_df** for teenagers that have the cd and downloaded versions of Game 1.

```
basic_info_teen_f1_df
## # A tibble: 3 x 4
##
                                gender
       id
            group_age
                         age
                                <chr>
## <chr>
                        <dbl>
              <chr>
                                Female
## 1 id_333
              teenage
                          17
## 2 id 122
              teenage
                          13
                                Female
## 3 id_530
                          14
                                Male
              teenage
```

There is another dataframe named **basic\_info\_teen\_f2\_df** for teenages that have of Game 2.

```
basic_info_teen_f2_df
## # A tibble: 8 x 3
##
       id
           number_of_tournaments
                                         level
## <chr>
                      <dbl>
                                        <chr>
## 1 id_721
                       23
                                       competent
## 2 id_815
                        9
                                       expert
## 3 id_122
                        2
                                       expert
## 4 id_530
                       12
                                       novice
## 5 id_782
                                       competent
                       10
## 6 id_295
                       24
                                       novice
## 7 id_386
                        1
                                       novice
                       17
## 8 id_308
                                       novice
```

We can left join basic\_info\_teen\_fl\_df and basic\_info\_teen\_f2\_df by using the following function and specifying a 'by' variable,id, where order determines which is the left dataframe (BE CAREFUL):

```
left_join(basic_info_teen_f1_df,basic_info_teen_f2_df, by = 'id')
## # A tibble: 3 x 6
##
       id
                                 gender
              group_age
                          age
                                           number_of_tournaments
                                                                     level
## <chr>
                                                   <dbl>
                                                                    <chr>
               <chr>>
                                 <chr>
                         <dbl>
## 1 id_333
              teenage
                          17
                                 Female
                                                   NA
                                                                    <NA>
## 2 id 122
                                                   2
              teenage
                          13
                                 Female
                                                                    expert
## 3 id_530
                                                   12
                          14
                                 Male
              teenage
                                                                    novice
```

This also could be done using the pipe function:

```
basic_info_teen_f1_df %>%
left_join(basic_info_teen_f2_df, by = 'id')
## # A tibble: 3 x 6
##
                                 gender
       id
                                          number_of_tournaments
              group_age
                          age
                                                                    level
## <chr>
               <chr>>
                         <dbl>
                                 <chr>
                                                  <dbl>
                                                                   <chr>
## 1 id_333
                          17
                                 Female
                                                  NA
                                                                   <NA>
              teenage
## 2 id_122
                          13
                                 Female
                                                   2
              teenage
                                                                   expert
## 3 id_530
                                 Male
                                                  12
              teenage
                          14
                                                                   novice
```

### Question about Left Join

There are two dataframes named, **basic\_info\_adult\_fl\_df** and **basic\_info\_adult\_f2\_df**. Combine the rows of these dataframes based on the common ids where your resulting dataframe should have the rows from **basic\_info\_adult\_f2\_df**, provide an appropriate name for the dataframe, and specify the variable types in the combined dataframe.

### **Full Join**

### Method 5: full\_join()

Joining columns based on an *id* matching from two different dataframes, however full\_join keeps all rows from both dataframes. The resulting dataframe has the rows from the left and right dataframes however, creates 1 row for rows that match. When the *id* from both dataframes does not match, NA is given. The id variable must be a character.

					id			
id	+	id		=				
:4 422		id_123			id_123			
id_123		id_563			id_789			
id_789		id_956			id_435		NA	NA
	5	id_235						
id_435		id_789			id_563	NA		
					id_956	NA		
					id_235	NA		

# full\_join()

There is a dataframe named **basic\_info\_teen\_fl\_df** for teenagers that have the cd and downloaded versions of Game 1.

```
basic_info_teen_f1_df
## # A tibble: 3 x 4
##
       id
            group_age
                        age
                                gender
                        <dbl>
## <chr>
              <chr>
                                <chr>
## 1 id_333
             teenage
                          17
                                Female
## 2 id 122
                                Female
             teenage
                          13
## 3 id_530
                          14
                                Male
             teenage
```

There is another dataframe named **basic\_info\_teen\_f2\_df** for teenages that have of Game 2.

```
basic_info_teen_f2_df
basic_info_teen_f2_df
## # A tibble: 8 x 3
##
           number_of_tournaments
                                        level
       id
## <chr>
                     <dbl>
                                        <chr>
## 1 id_721
                       23
                                       competent
                        9
## 2 id 815
                                       expert
                        2
## 3 id_122
                                       expert
## 4 id_530
                       12
                                       novice
## 5 id_782
                       10
                                       competent
## 6 id_295
                       24
                                       novice
## 7 id_386
                       1
                                       novice
## 8 id_308
                       17
                                       novice
```

We can full join basic\_info\_teen\_fl\_df and basic\_info\_teen\_f2\_df by using the following function and specifying a 'by' variable,id, where order determines which is the left dataframe (BE CAREFUL):

full_join(basic_info_teen_f1_df,basic_info_teen_f2_df, by = 'id') ## # A tibble: 9 x 6								
## id g	group_age	age	gender	number_of_tournaments	level			
## <chr></chr>	<chr></chr>	<dbl></dbl>	<chr></chr>	<dbl></dbl>	<chr></chr>			
## 1 id_333	teenage	17	Female	NA	<na></na>			
## 2 id_122	teenage	13	Female	2	expert			
## 3 id_530	teenage	14	Male	12	novice			
## 4 id_721	<NA $>$	NA	<NA $>$	23	competent			
## 5 id_815	<NA $>$	NA	<NA $>$	9	expert			
## 6 id_782	<NA $>$	NA	<NA $>$	10	competent			
## 7 id_295	<NA $>$	NA	<NA $>$	24	novice			
## 8 id_386	<NA $>$	NA	<NA $>$	1	novice			
## 9 id 308	<NA $>$	NA	<NA $>$	17	novice			

This also could be done using the pipe function:

```
basic_info_teen_f1_df %>%
full_join(basic_info_teen_f2_df, by = 'id')
## # A tibble: 9 x 6
## id
                                        number_of_tournaments
           group_age
                               gender
                                                                       level
                        age
## <chr>
                                              <dbl>
            <chr>
                       <dbl>
                               <chr>
                                                                       <chr>
## 1 id_333
            teenage
                        17
                               Female
                                                NA
                                                                       <NA>
## 2 id_122
            teenage
                        13
                               Female
                                                2
                                                                       expert
## 3 id_530
                        14
                               Male
                                               12
            teenage
                                                                       novice
## 4 id_721
            <NA>
                       NA
                               <NA>
                                               23
                                                                       competent
## 5 id_815
                                                9
            <NA>
                               <NA>
                                                                       expert
                       NA
## 6 id_782
            <NA>
                               <NA>
                                               10
                       NA
                                                                       competent
## 7 id_295
            <NA>
                       NA
                               <NA>
                                               24
                                                                       novice
## 8 id_386
            <NA>
                               <NA>
                                                1
                       NA
                                                                       novice
## 9 id_308 <NA>
                       NA
                               <NA>
                                                17
                                                                       novice
```

### Question about Full Join

There are two dataframes named, **basic\_info\_adult\_f1\_df** and **basic\_info\_adult\_f2\_df**. Combine the rows of these dataframes based on the common ids where your resulting dataframe should have the rows from both original dataframes, provide an appropriate name for the dataframe, and specify the variable types in the combined dataframe.