Joining Datasets – Part 2

Left Join

Left Join: includes all rows from the left dataset and the matching rows from the right dataset. If there's no match, the columns from the right dataset will be filled with NA. Here the rows of the first table are always returned, regardless of whether there is a match in the second table.

Command Illustration

```
new_dataframe_name <- dataframe_left %>%
     left_join(dataframe_right, c("colname_1" = "colname_2"))
```

For Example 1, assume the two following dataframes:

Illustration_Data_1

Name	Age	num_kids
Val	18	1
Derek	25	0
Whitney	30	2
Daniella	45	1

Illustration_Data_2

First_Name	Last_Name	Gender
Val	Chmerkovskiy	Male
Derek	Hough	Male
Whitney	Carson	Female
Sasha	Farber	Male
Daniella	Karagach	Female
Lindsay	Arnold	Female
Mark	Ballas	Male

Example 1: Do a left join where Illustration_Data_1 is the left data and Illustration_Data_2 is the right data.

```
example_1 <- Illustration_Data_1 %>%
  left_join(Illustration_Data_2, by = c("Name" = "First_Name"))
```


For Example 2, assume the two following dataframes:

Illustration_Data_2

First_Name	Last_Name	Gender
Val	Chmerkovskiy	Male
Derek	Hough	Male
Whitney	Carson	Female
Sasha	Farber	Male
Daniella	Karagach	Female
Lindsay	Arnold	Female
Mark	Ballas	Male

Illustration_Data_1

Name	Age	num_kids
Val	18	1
Derek	25	0
Whitney	30	2
Daniella	45	1

Example 2: Do a left join where Illustration_Data_2 is the left data and Illustration_Data_1 is the right data.

```
example_2 <- Illustration_Data_2 %>%
  left_join(Illustration_Data_1, by = c("First_Name" = "Name"))
```


For Example 3, assume the two following dataframes:

Illustration_Data_3

Name	Last_Name	Car
Val	Chmerkovskiy	Mercedes
Val	Chmerkovskiy	Tesla
Val	Chmerkovskiy	Audi
Derek	Hough	Ferrari
Lindsay	Arnold	Tesla
Mark	Ballas	BMW

Illustration_Data_1

Name	Age	num_kids
Val	18	1
Derek	25	0
Whitney	30	2
Daniella	45	1

Example 3: Do a left join where Illustration_Data_3 is the left data and Illustration_Data_1 is the right data.

```
example_3 <- Illustration_Data_3 %>%
  left_join(Illustration_Data_1, by = c("Name" = "Name"))
```


For Example 4, assume the two following dataframes:

Illustration_Data_1

Name	Age	num_kids
Val	18	1
Derek	25	0
Whitney	30	2
Daniella	45	1

Illustration_Data_3

Name	Last_Name	Car
Val	Chmerkovskiy	Mercedes
Val	Chmerkovskiy	Tesla
Val	Chmerkovskiy	Audi
Derek	Hough	Ferrari
Lindsay	Arnold	Tesla
Mark	Ballas	BMW

Example 4: Do a left join where Illustration_Data_1 is the left and Illustration_Data_3 is the right.

```
example_4 <- Illustration_Data_1 %>%
  left_join(Illustration_Data_3, by = c("Name" = "Name"))
```


Right Join

Right Join: is the opposite of a left join. It includes all rows from the right dataset and the matching rows from the left dataset. A right join but this is much less common.

Command Illustration

```
new_dataframe_name <- dataframe_left %>%
     right_join(dataframe_right, c("colname_1" = "colname_2"))
```

For Example 5, assume the two following dataframes:

Illustration_Data_1

Name	Age	num_kids
Val	18	1
Derek	25	0
Whitney	30	2
Daniella	45	1

Illustration_Data_3

Name	Last_Name	Car
Val	Chmerkovskiy	Mercedes
Val	Chmerkovskiy	Tesla
Val	Chmerkovskiy	Audi
Derek	Hough	Ferrari
Lindsay	Arnold	Tesla
Mark	Ballas	BMW

Example 5: Do a right join where Illustration_Data_1 is the left and Illustration_Data_3 is the right.

```
example_5 <- Illustration_Data_1 %>%
  right_join(Illustration_Data_3, by = c("Name" = "Name"))
```


Summarizing NA's

Example 2 Output:

First_Name	Last_Name	Gender	Age	num_kids
Val	Chmerkovskiy	Male	18	1
Derek	Hough	Male	25	0
Whitney	Carson	Female	30	2
Sasha	Farber	Male	*NA*	*NA*
Daniella	Karagach	Female	45	1
Lindsay	Arnold	Female	*NA*	*NA*
Mark	Ballas	Male	*NA*	*NA*

Example 6: Summarize the NA's for the left joined data from Example 2

```
example_6 <- example_2 %>%
  summarize(num_people = n(),
            num_na = sum(is.na(Age)),
            num_not_na = sum(!is.na(Age)))
```
