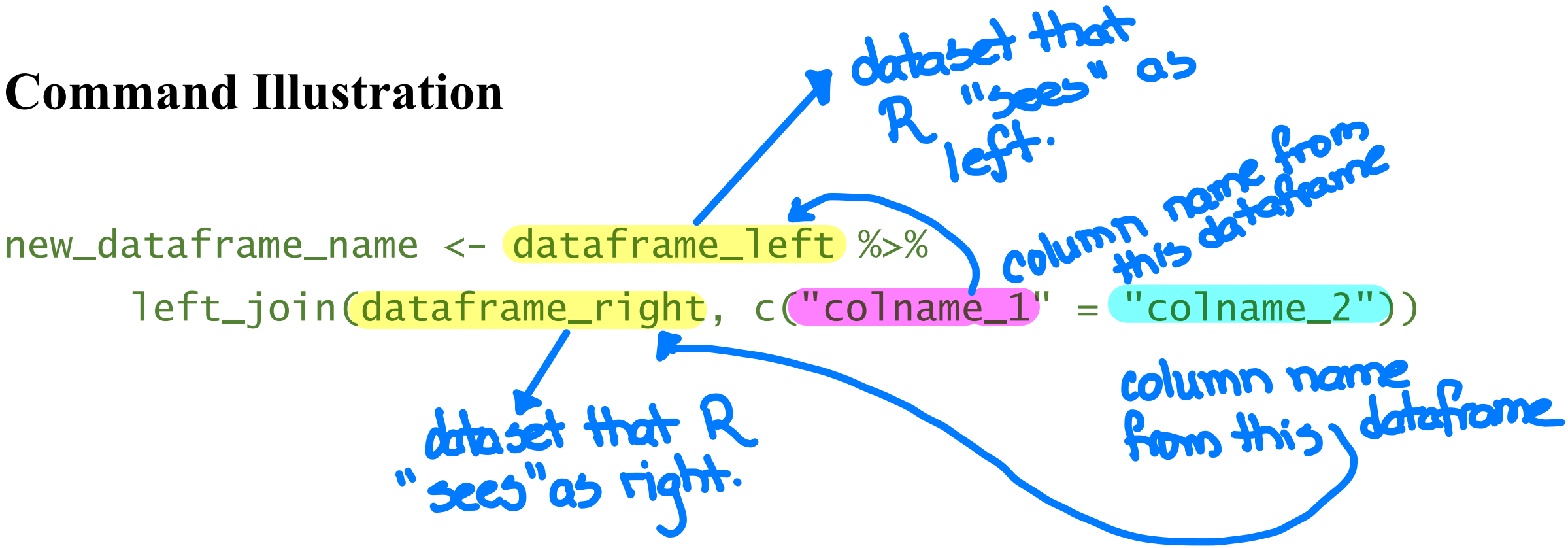
Joining Datasets
Part 2

`left_join()` Function

Command Illustration



Example 1

Left

Illustration_Data_1

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

Right

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"))
```

This is the "dominant" dataframe

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~~

Output will be a dataframe that looks like:

Name	Age	num_kids	Last_Name	Gender
Val	18	1	Chmerkovskiy	Male
Derek	25	0	Hough	Male
Whitney	30	2	Carson	Female
Daniela	45	1	Karagach	Female

Example 2

Left

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

Right

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"))
```

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

Output will be a dataframe that looks like:

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 3

Left

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

Right

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"))
```

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~~

Output will be a dataframe that looks like:

Name	Last_Name	Car	Age	num_kids
Val	Chmerkovskiy	Mercedes	18	1
Val	Chmerkovskiy	Tesla	18	1
Val	Chmerkovskiy	Audi	18	1
Derek	Hough	Ferrari	25	0
Lindsay	Arnold	Tesla	NA	NA
Mark	Ballas	BMW	NA	NA

Example 4

Left

Illustration_Data_1

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

Right

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"))
```

Illustration_Data_1

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

Illustration_Data_3

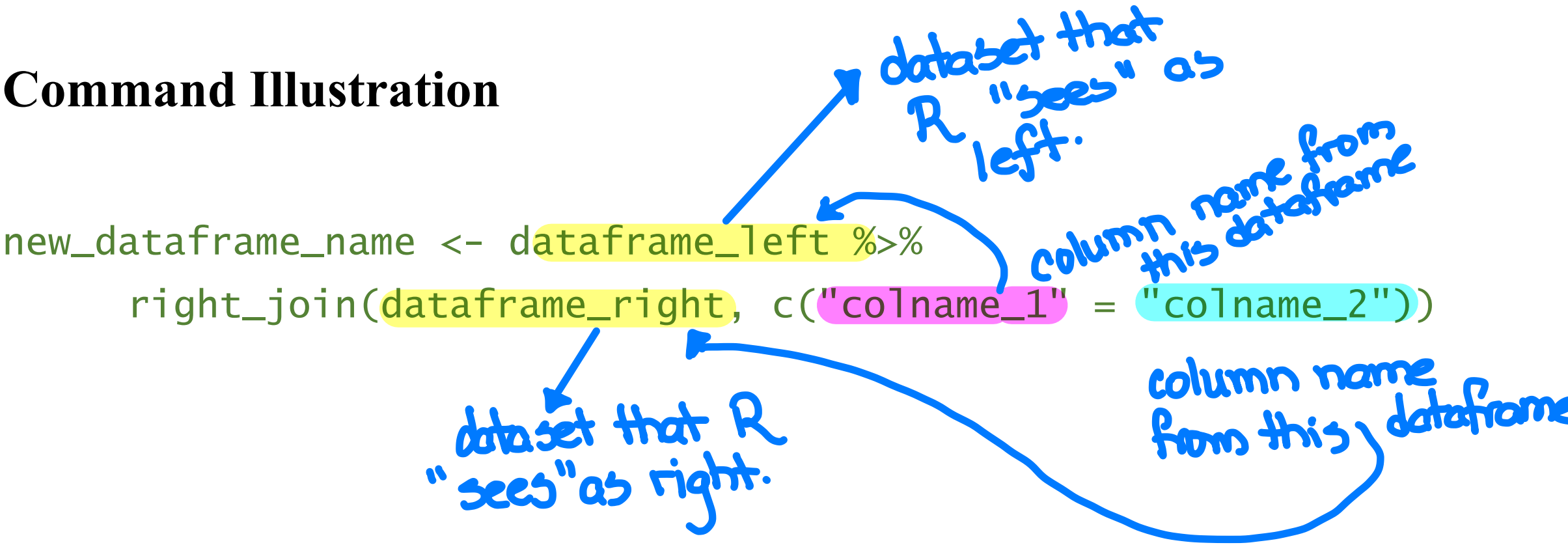
Name	Last_Name	Car
Val	Chmerkovskiy	Mercedes
Val	Chmerkovskiy	Tesla
Val	Chmerkovskiy	Audi
Derek	Hough	Ferrari
~~Lindsay~~	~~Arnold~~	~~Tesla~~
~~Mark~~	~~Ballas~~	~~BMW~~

Output will be a dataframe that looks like:

Name	Age	num_kids	Last_Name	Car
Val	18	1	Chmerkovskiy	Mercedes
Val	18	1	Chmerkovskiy	Tesla
Val	18	1	Chmerkovskiy	Audi
Derek	25	0	Hough	Ferrari
Whitney	30	2	NA	NA
Daniella	45	1	NA	NA

`right_join()` Function

Command Illustration



Example 5

Left

Illustration_Data_1

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

This is the dominant dataset with right join

Right

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"))
```

This is the dominant dataframe

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

Note: in R, when doing right joining, this is NOT the order that the columns appear.

Output will be a dataframe that looks like:

Name	Last_Name	Car	Age	num_kids
Val	Chmerkovskiy	Mercedes	18	1
Val	Chmerkovskiy	Tesla	18	1
Val	Chmerkovskiy	Audi	18	1
Derek	Hough	Ferrari	25	0
Lindsay	Arnold	Tesla	NA	NA
Mark	Ballas	BMW	NA	NA

Summarizing NA's

We summarize the # of NA's for the columns that were added to the left dataset.

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)))
```

counts # of rows

counts # of NA in column Age

counts # of not NA in column Age

num_people	num_na	num_not_na
7	3	4