Data Cleaning With R: Takeaways

by Dataquest Labs, Inc. - All rights reserved © 2021

Syntax

MANIPULATING COLUMNS USING THE DPLYR PACKAGE:

- Convert a single column to numeric:

```
data_frame <- data_frame %>%
mutate(`col name` = as.numeric(`col name`))
```

- Converting multiple columns to numeric with a condition:

```
data_frame <- data_frame %>%
mutate(across(ends_with("this character"), as.numeric))
```

- Converting consecutive columns to numeric with column names:

```
data_frame <- data_frame %>%
mutate(across(`col name 1`: `col name 5`, as.numeric))
```

- Converting consecutive columns to numeric with column indexes:

```
data_frame <- data_frame %>%
mutate(across(1:5, as.numeric)) #`beginning index`: `ending index`
```

- Filtering a DataFrame:

```
data_frame <- data_frame %>%
filter(`col name` > condition)
```

- Grouping a DataFrame:

```
data_frame <- data_frame %>%
group_by(`col name 1`, `col name 2`)
```

- Summing up columns:

```
data_frame <- data_frame %>%
mutate(`col name` = `col name 1` + `col name 2`)
```

- Selecting variables from a DataFrame:

```
data_frame <- data_frame %>%
select(`col name 1`, `col name 2`, `col name 3`)
```

- Selecting variables from a condition:

```
data_frame <- data_frame %>%
select(`first col`, starts_with("col"))
```

- Removing a column from a DataFrame:

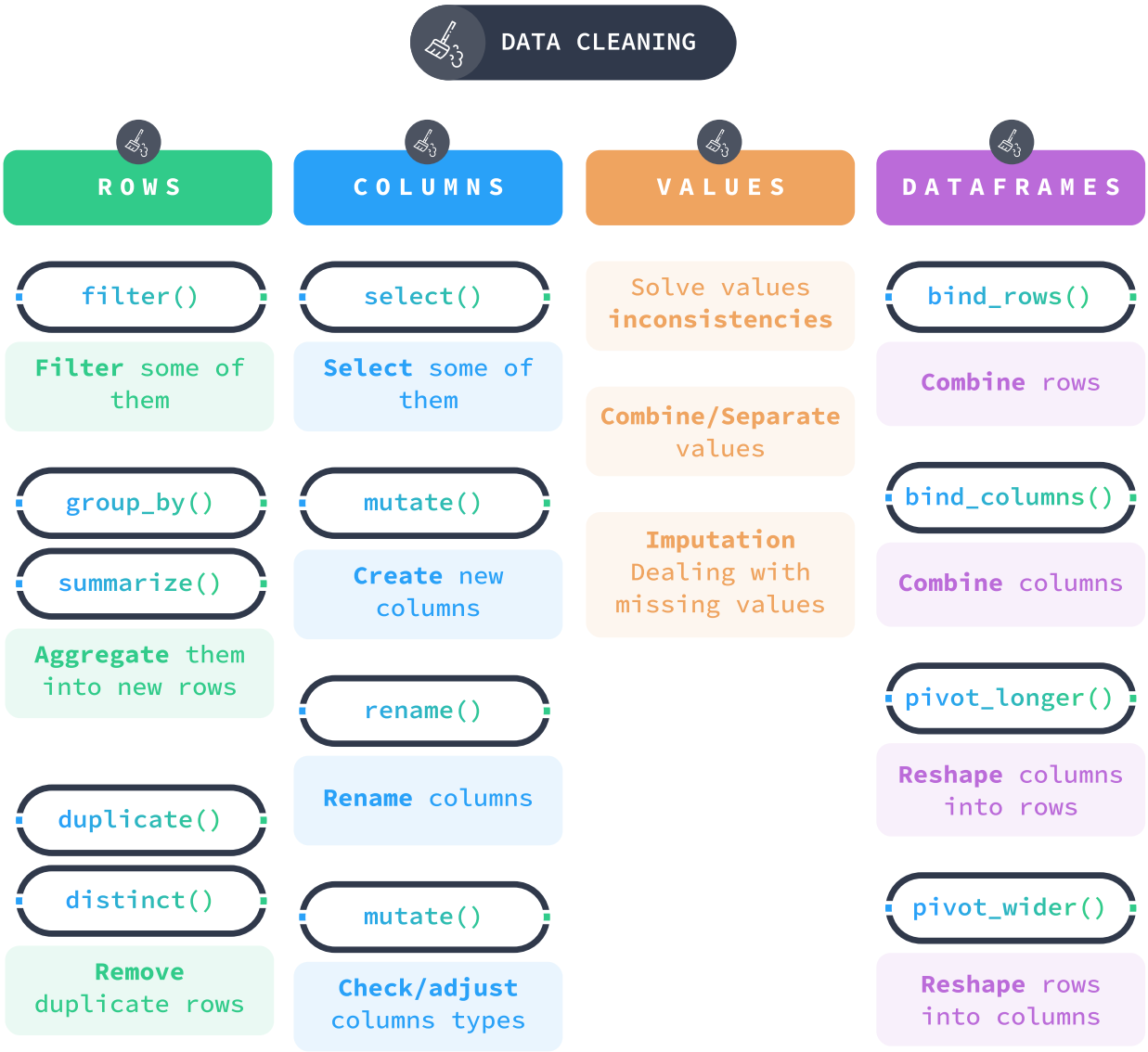
```
graduation <- graduation %>%
select(-the_name_of_column_to_remove) #note the presence of the symbol -
```

- Renaming a column in a DataFrame:

```
data_frame %>%
rename(new_column_name = old_column_name)
```

Concepts

- Much of the data we will encounter in the real world requires data cleaning. Data cleaning includes the following processes:
    - Removing data we don't need for analysis
    - Removing duplicate data
    - Dealing with missing data and outliers
    - Creating new variables where necessary
    - Combining separate datasets
- Metadata refers to any available descriptions of the datasets.
- Tick marks (``) are necessary when referring to variable names that contain spaces.



Resources

- Preparing data analysis
- Duplicated function
- Six steps to data cleaning

Takeaways by Dataquest Labs, Inc. - All rights reserved © 2021