

YData: An Introduction to Data Science

Lecture 10: Groups

Elena Khusainova & John Lafferty
Statistics & Data Science, Yale University
Spring 2021

Credit: data8.org



Announcements

- HW03 due Thursday by 11:59 PM
- The first project will be posted this Friday, with checkpoint the Friday after (3/5)

Apply

Apply

The `apply` method creates an array by calling a function on every element in input column(s)

- First argument: Function to apply
- Other arguments: The input column(s)

```
table_name.apply(function_name, 'column_label')
```

(DEMO)

Apply with Multiple Columns

Apply

The `apply` method creates an array by calling a function on every element in input column(s)

- First argument: Function to apply
- Other arguments: The input column(s)

```
table_name.apply(one_arg_function, 'column_label')  
table_name.apply(two_arg_function,  
                  'column_label_for_first_arg',  
                  'column_label_for_second_arg')
```

`apply` called with only a function applies it to each row

(DEMO)

Grouping by One Attribute

Grouping by One Column

The `group` method aggregates all rows with the same value for a column into a single row in the resulting table.

- First argument: Which column to group by
- Second argument: (Optional) How to combine values
 - `len` – number of grouped values (default)
 - `list` – list of all grouped values
 - `sum` – total of all grouped values

(DEMO)

Cross-Classification

Grouping By Multiple Columns

The `group` method can also aggregate all rows that share the combination of values in multiple columns

- First argument: A list of which columns to group by
- Other arguments: (Optional) How to combine values

(DEMO)