MATH1401

Fall 2021

Lecture 12

Joins

Class Checklist

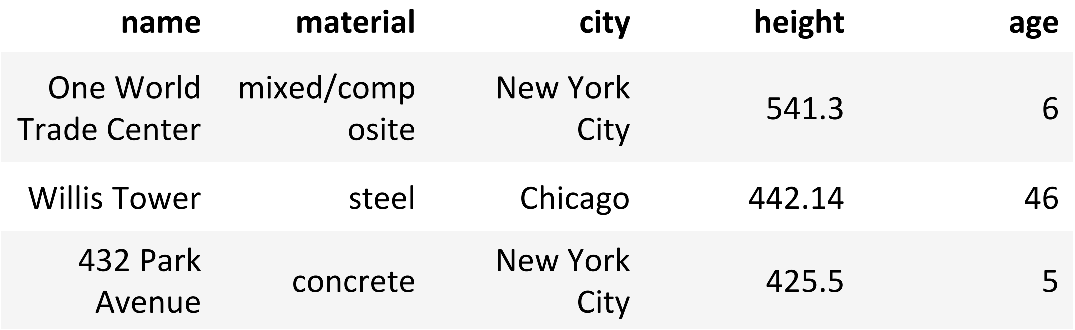
- **Lab 4 - Due Date** : Thursday: 9/30 – 9 PM
  - Graded Questions:1.1-1.5, 2.1-2.4, 3.1-3.6, 4.2

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.

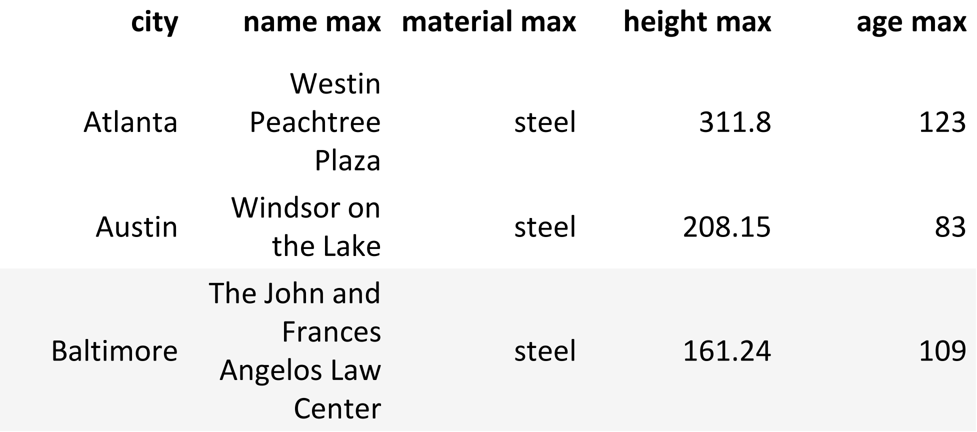
- First argument: Which column to group by
- Second argument: (Optional) How to combine values

Grouping by One Column



name	material	city	height	age
One World Trade Center	mixed/composite	New York City	541.3	6
Willis Tower	steel	Chicago	442.14	46
432 Park Avenue	concrete	New York City	425.5	5

```
sky.group('city', max)
```



city	name max	material max	height max	age max
Atlanta	Westin Peachtree Plaza	steel	311.8	123
Austin	Windsor on the Lake	steel	208.15	83
Baltimore	The John and Frances Angelos Law Center	steel	161.24	109

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
- Second argument: (Optional) How to combine values

Grouping by multiple Columns

name	material	city	height	age
One World Trade Center	mixed/composite	New York City	541.3	6
Willis Tower	steel	Chicago	442.14	46
432 Park Avenue	concrete	New York City	425.5	5

```
sky.group(['city', 'material'], max)
```

city	material	height max
Atlanta	concrete	264.25
Atlanta	mixed/composite	311.8
Atlanta	steel	169.47

Pivot

- Cross-classifies according to two categorical variables
- Two required arguments:
  - First: variable that forms column labels of grid
  - Second: variable that forms row labels of grid
- Two optional arguments (include **both** or **neither**)
  - `values`=’column_label_to_aggregate’
  - `collect`=function_to_aggregate_with

Pivot

- Cross-classifies according to two categorical variables

- **Table_name.pivot(‘label1’,’label2’)** – Creates pivot table and displays distribution

- **Table_name.pivot(‘label1’,’label2’,’numerical’,function)** – Applys function to numerical value for each group defined by label1 and label 2

Join

Joining Two Tables

```
drinks.join('Cafe', discounts, 'Location')
```

Match rows in this table ...

… using values in this column ...

… with rows in that table ...

… using values in that column.

Columns from both tables

drinks

Drink	Cafe	Price
Milk Tea	Asha	5.5
Espresso	Strada	1.75
Latte	Strada	3.25
Espresso	FSM	2

discounts

Coupon	Location
10%	Asha
25%	Strada
5%	Asha

Cafe	Drink	Price	Coupon
Asha	Milk Tea	5.5	10%
Asha	Milk Tea	5.5	5%
Strada	Espresso	1.75	25%
Strada	Latte	3.25	25%

The joined column is sorted automatically

Examples