



DATA 1202
Spring 2024

Lecture 10

Pivots and Joins

Announcements

- **Lab 6** on Thursday due 2/16 @ 5pm
 - **HW 6** due Wednesday 2/21 @ 11pm
 - **Project 1** due Friday 3/8
 - Checkpoint due Friday 2/23
 - [World population through data](#)
 - Covers topics in lecture through today
 - **Midterm** in-class, Wednesday 2/21
-

Lists

Lists are Generic Sequences

A list is a sequence of values (just like an array),
but the values can have different types

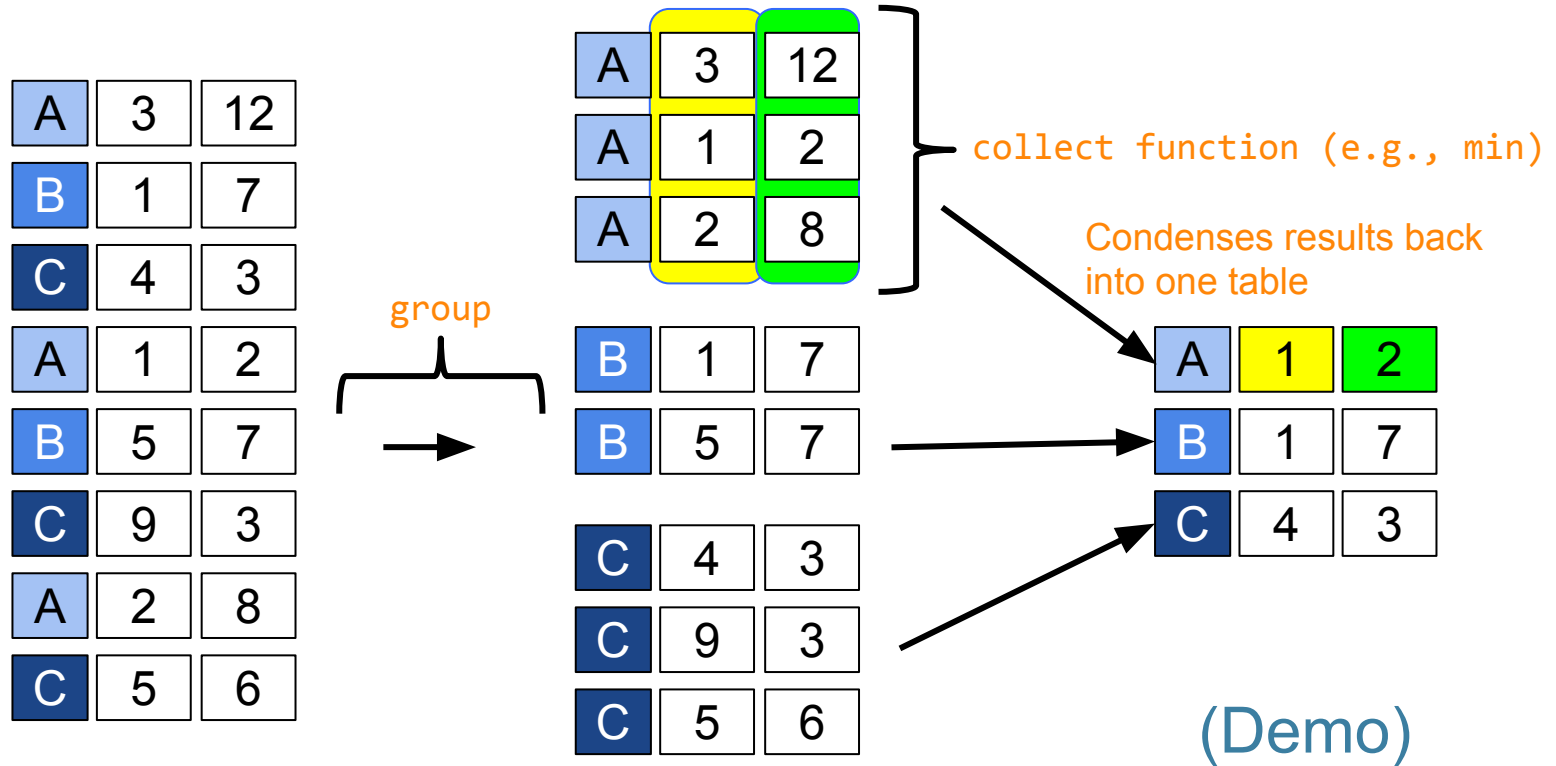
```
[2+3, 'four', Table().with_column('K', [3, 4])]
```

- Lists can be used to create table rows.
- If you create a table column from a list, it will be converted to an array automatically

(Demo)

Review: Grouping and Functions

Grouping and Collection



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
- `collect = ...` (Optional) How to combine values

Group by **one** column (and minimize each other column):

`t.group("label", collect=min)` or `t.group("label", min)`

Group by **two** columns (and minimize each other):

`t.group(["label 1", "label 2"], collect=min)` or
`t.group(["label 1", "label 2"], min)`

(Demo)



PIVOT

Pivot Tables

Pivot

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

Group or Pivot?

- When to **Group**:
 - aggregates of one categorical variable
 - aggregates of **many variables**
 - **Multiple** outputs (aggregate columns)
 - When to **Pivot**:
 - Aggregates of **exactly two variables**
 - **Few unique values** for column variable
 - Interested in **every combination** of values
-



JOIN

Joins

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

The joined column is
sorted automatically

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%

(Demo)

Table Review

Important Table Methods

`t.select(column, ...) or t.drop(column, ...)`

`t.take([row_num, ...]) or t.exclude([row_num, ...])`

`t.sort(column, descending=False)`

`t.where(column, are.condition(...))`

`t.apply(function_name, column, ...)`

`t.group(column) or t.group(column, function_name)`

`t.group([column, ...]) or t.group([column, ...], function_name)`

`t.pivot(cols, rows) or t.pivot(cols, rows, vals, function_name)`

`t.join(column, other_table, other_table_column)`

<https://pages.mtu.edu/~lebrown/data1202-s24/reference/index.html>

Let's Practice

How to solve a table manipulation problems:

1. **Understand the result** — what information is needed to fill in a particular value of the goal.
 2. **Describe the operations** — articulate (in English) the operations that will produce the result.
 3. **Write Python expressions** — express the operations using table methods, functions, & array arithmetic.
-

Discussion Question

1. For each city, what's the height of the tallest building for each material?
2. For each city, what's the height difference between the tallest steel building and the tallest concrete building?

sky

name	material	city	height	age
Metropolitan Tower	concrete	New York City	218.24	35
Paul Hastings Tower	steel	Los Angeles	213.06	49
Barclay Tower	concrete	New York City	205.06	13
Westin Peachtree Plaza	concrete	Atlanta	220.37	44
Wells Fargo Plaza	steel	Houston	302.37	37

(Demo)

Extra Example

Challenge Question

Generate a table of the names of the oldest buildings for each material for each city:

city	concrete	mixed/composite		steel
Atlanta	Westin Peachtree Plaza	One Atlantic Center		FlatironCity
Austin	One American Center		University of Texas Tower	
Baltimore	Charles Towers North Apartments			Emerson Tower
Boston	Harbor Towers I	Ellison Building	Marriott's Custom House	
Charlotte	Bank of America Corporate Center		Hearst Tower	Midtown Plaza
Chicago	The Powhatan	American Furniture Mart		The Rookery
Cincinnati	Kroger Building	Great American Tower at Queen City Square		PNC Tower
Cleveland	National City Center		55 Public Square	Huntington Bank Building
Columbus	Key Bank Building			Leveque Tower
Dallas	Reunion Tower	Bank of America Plaza		Three AT&T Plaza