

BC COMS 1016: Intro to Comp Thinking & Data Science

Lecture 8 – Pivot/Joins & Randomness, Conditionals, Iterations

Announcements



- Lab 03 - Functions and Visualizations
 - Due Wed (11/11)
- HW03 - Functions, Histograms, and Groups
 - Due Thursday (11/12)
- Checkpoint/Project 1:
 - Paired assignment that covers the previous section of the course material
 - Released tonight (11/12) and due Wednesday (11/18)

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



Pivot Example

```
sky = Table.read_table('skyscrapers_v2.csv')
```



Pivot Example

```
sky = Table.read_table('skyscrapers_v2.csv')
```

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



Pivot Example

```
sky = Table.read_table('skyscrapers_v2.csv')
```

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.pivot('material', 'city')
```



Pivot description

`Tbl.pivot(col1, col2)`

1. **string**: name of column whose unique values will make up columns of pivot table
2. **string**: name of column whose unique values will make up rows of pivot table



Pivot description

`Tbl.pivot(col1, col2)`

1. **string**: name of column whose unique values will make up columns of pivot table
2. **string**: name of column whose unique values will make up rows of pivot table

`sky.pivot('material', 'city')`



Pivot description

sky.pivot('material', 'city')

	name	material	city
	Willis Tower	steel	Chicago
	Trump International Hotel & Tower	concrete	Chicago
	Aon Center	steel	Chicago
	John Hancock Center	steel	Chicago
	Bank of America Plaza	mixed/composite	Atlanta
	U.S. Bank Tower	steel	Los Angeles
	The Franklin - North Tower	mixed/composite	Chicago
	JPMorgan Chase Tower	mixed/composite	Houston
	Two Prudential Plaza	concrete	Chicago
	Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago			
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago			
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago			1
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago			1
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1		1
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1		2
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta			
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1		3
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta		1	
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1		3
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
JPMorgan Chase Tower	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta		1	
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1		3
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta		1	
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1	1	3
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Franklin - North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta		1	
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	1	1	3
Cincinnati			
Cleveland			
Columbus			

name	material	city
Willis Tower	steel	Chicago
Trump International Hotel & Tower	concrete	Chicago
Aon Center	steel	Chicago
John Hancock Center	steel	Chicago
Bank of America Plaza	mixed/composite	Atlanta
U.S. Bank Tower	steel	Los Angeles
The Financial North Tower	mixed/composite	Chicago
JPMorgan Chase Tower	mixed/composite	Houston
Two Prudential Plaza	concrete	Chicago
Wells Fargo Plaza	steel	Houston



Pivot description

sky.pivot('material', 'city')

city	concrete	mixed/composite	steel
Atlanta		1	
Austin			
Baltimore			
Boston			
Charlotte			
Chicago	2	1	3
Cincinnati			
Cleveland			
Columbus			

	name	material	city
	Willis Tower	steel	Chicago
	Trump International Hotel & Tower	concrete	Chicago
	Aon Center	steel	Chicago
	John Hancock Center	steel	Chicago
	Bank of America Plaza	mixed/composite	Atlanta
	U.S. Bank Tower	steel	Los Angeles
	The Franklin - North Tower	mixed/composite	Chicago
	JPMorgan Chase Tower	mixed/composite	Houston
	Two Prudential Plaza	concrete	Chicago
	Wells Fargo Plaza	steel	Houston

Pivot Tables



- 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 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`

Group vs Pivot



Pivot

- One combo of grouping variables **per entry**
- **Two** grouping variables: columns and rows
- Aggregate values of **values column**
- Missing combos = **0 (or empty string)**

Group

- One combo of grouping variables **per row**
- **Any number** of grouping variables
- Aggregate values of **all other columns** in table
- Missing combos **absent**



Joining Two Tables

tblA.join(colA, tblB, colB)

tblA.join(colA, tblB)

Table Review



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

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

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

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

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

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

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

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

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

<https://coms1016.barnard.edu/python-reference.html>



Comparisons



Comparison Operators

Operator	Table predicate
<code>==</code>	<code>are.equal_to</code>
<code>!=</code>	<code>are.not_equal_to</code>
<code>></code>	<code>are.above</code>
<code>>=</code>	<code>are.above_or_equal_to</code>
<code><</code>	<code>are.below</code>
<code><=</code>	<code>are.below_or_equal_to</code>

The result of a comparison expression is a **bool** value:
True, False



Comparison Operators

The result of a comparison expression is a **bool** value

`x = 2`

`y = 3`

Comparison Operators



The result of a comparison expression is a **bool** value

x = 2

y = 3

Assignment
Statements

Comparison Operators



The result of a comparison expression is a **bool** value

`x = 2`

`y = 3`

Assignment
Statements

`x > 1`

`x > y`

`y >= 3`

`x == y`

`x != 2`

`2 < x < 5`

Comparison Operators



The result of a comparison expression is a **bool** value

`x = 2`

`y = 3`

Assignment
Statements

`x > 1`

`x > y`

`y >= 3`

`x == y`

`x != 2`

`2 < x < 5`

Comparison
Expressions

Combining Comparisons



The result of a comparison expression is a **bool** value

a = True

b = False

not b

a or b

a and not b

a and b

not (a or b)

b and b

Combining Comparisons



The result of a comparison expression is a **bool** value

$a = \text{True}$

$b = \text{False}$

Evaluate to True

$\text{not } b$

$a \text{ or } b$

$a \text{ and not } b$

$a \text{ and } b$

$\text{not } (a \text{ or } b)$

$b \text{ and } b$

Evaluate to False

Aggregating Comparisons



Summing an array or list of **bool** values count the number of **True** values

$$1 + 0 + 1 == 2$$

$$\text{True} + \text{False} + \text{True} == 2$$

$$\text{sum}([1, 0, 1]) == 2$$

$$\text{sum}([\text{True}, \text{False}, \text{True}]) == 2$$



Control Statements

Control Statements



These statements *control* the sequence of computations that are performed

- The keywords **if** and **for** begin control statements
- The purpose of **if** is to define functions that choose different behavior based on their arguments



Random Selection

Random Selection



`np.random.choice`

- Selects at random
- With replacement
- From an array
- A specific number of times

`np.random.choice(some_array, sample_size)`



Appendix

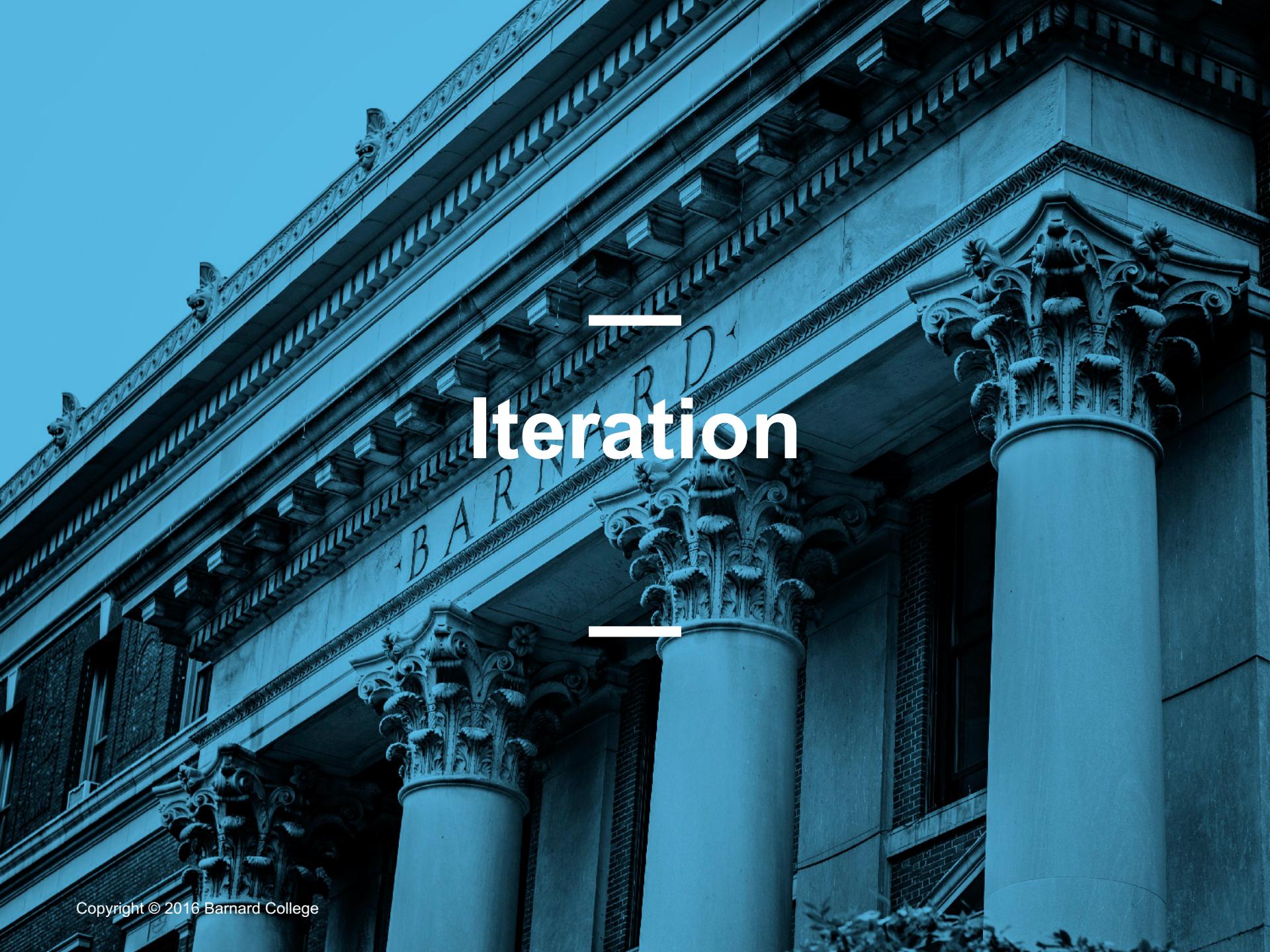
Appending Arrays



A Longer Array

- `np.append(array_1, value):`
 - new array with value appended to array_1
 - value has to be of the same type as elements of array_1

- `np.append(array_1, array_2):`
 - new array with array_2 appended to array_1
 - Elements of array_2 have to be of the same type as elements of array_1



Iteration



for statements

- **for** is a keyword that begins a control statement
- The purpose of **for** is to perform a computation for every element in a list or array