



DSFA
Spring 2021

Lecture 10

Groups

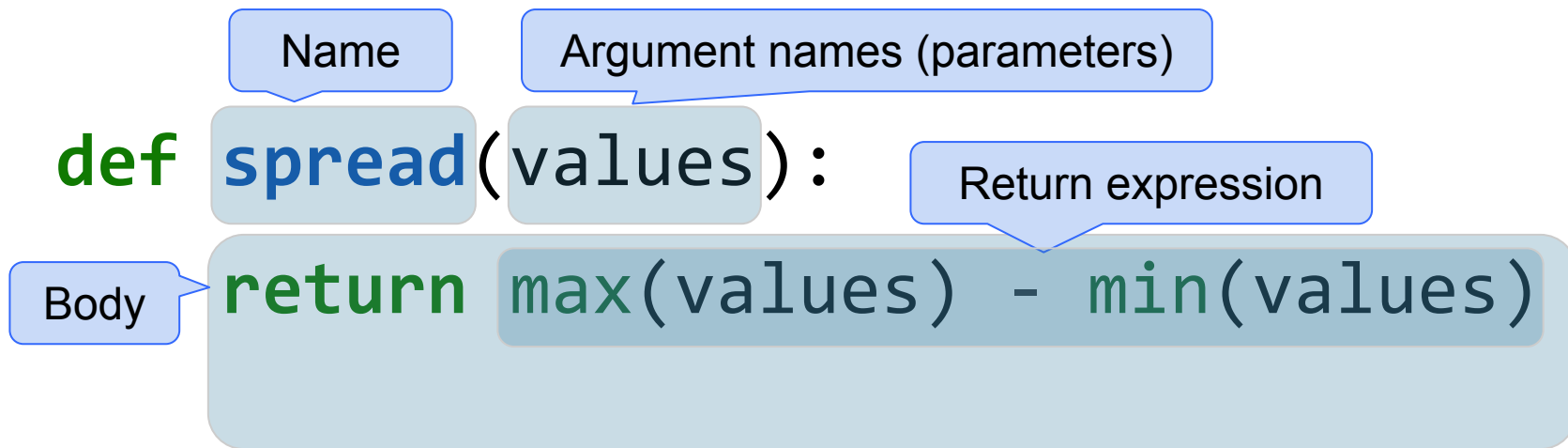
Announcements

- Project 1 out this Friday
 - Go to section this week to find a partner!
-

Defining Functions

Def Statements

User-defined functions give names to blocks of code



(Demo)

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

The `apply` method creates an array by calling a function on every element in one or more input columns

- 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)

Example: Prediction

Grouping Rows

Group

The **group** method aggregates all rows with the same value for a column into a single row in the result

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

(Demo)

Grouping By Two Columns

The `groups` 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

(Demo)

Challenge Question

Which NBA teams spent the most on their starters in 2016?

- Each team has one *starter* per position
- Assume the starter for a team & position is the player with the highest salary on that team in that position

team	player	position	salary
Atlanta Hawks	Brandon Goodwin	PG	0.522738
Atlanta Hawks	Bruno Fernando	C	1.4
Atlanta Hawks	Cam Reddish	SG	4.24572
Atlanta Hawks	Chandler Parsons	SF	25.1025
Atlanta Hawks	Damian Jones	C	2.30506
Atlanta Hawks	DeAndre' Bembry	SG	2.60398

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 grid
 - Second: variable that forms row labels of grid
- Two optional arguments (include both or neither)
 - `values='column_label_to_aggregate'`
 - `collect=function_with_which_to_aggregate`

(Demo)

When poll is active, respond at PolleEv.com/dsfa

Text **DSFA** to **22333** once to join

The topic giving me the most trouble...

Tables, table manipulation

Arrays

Histograms

Functions/Apply

Other

I'm good, thanks

