DATA 1202
Spring 2024

Lecture 9

Groups

Announcements
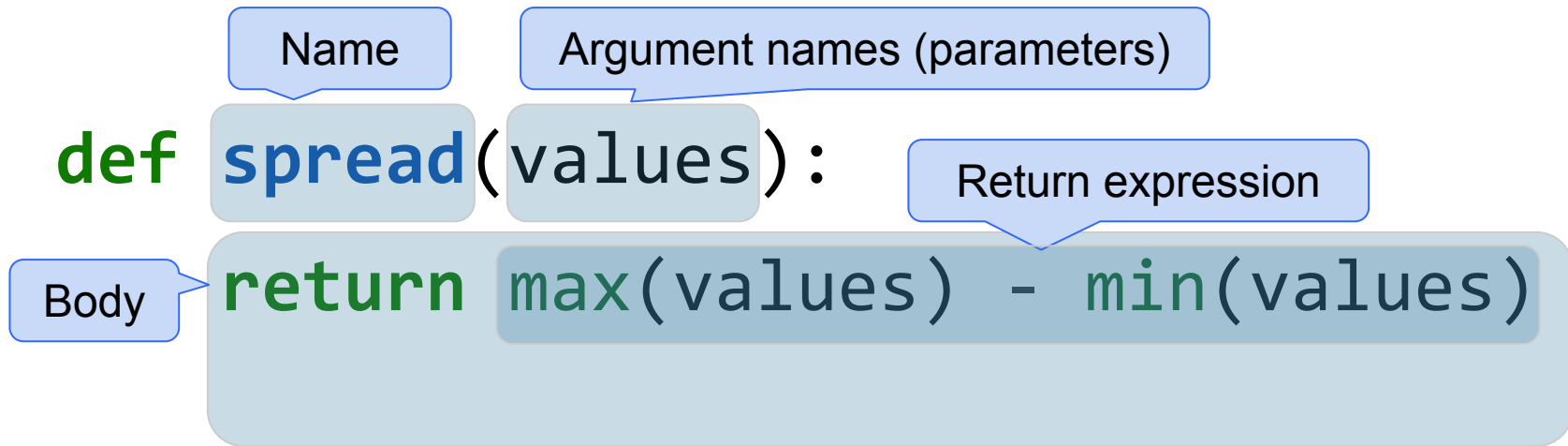
- **HW 4** due Wednesday 2/7 at **5pm**

- **NO LAB on Thursday**

Review:
Functions and Apply

Def Statements

User-defined functions give names to blocks of code

Name

Argument names (parameters)

```
def spread(values):
        return max(values) - min(values)
```

Return expression

Body

Apply

apply

1. Calls a function on every element in the input column(s)
2. Produces an array containing the output of the function on each input column element.

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

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

Prediction (using Apply)

(Demo)

Discussion Questions

1. How could we take sex into account when making predictions?
2. Do we make smaller errors on average when we do this?

group

Grouping

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

Grouping and Collection

collect function (e.g., min)

Condenses results back into one table

group

(Demo)
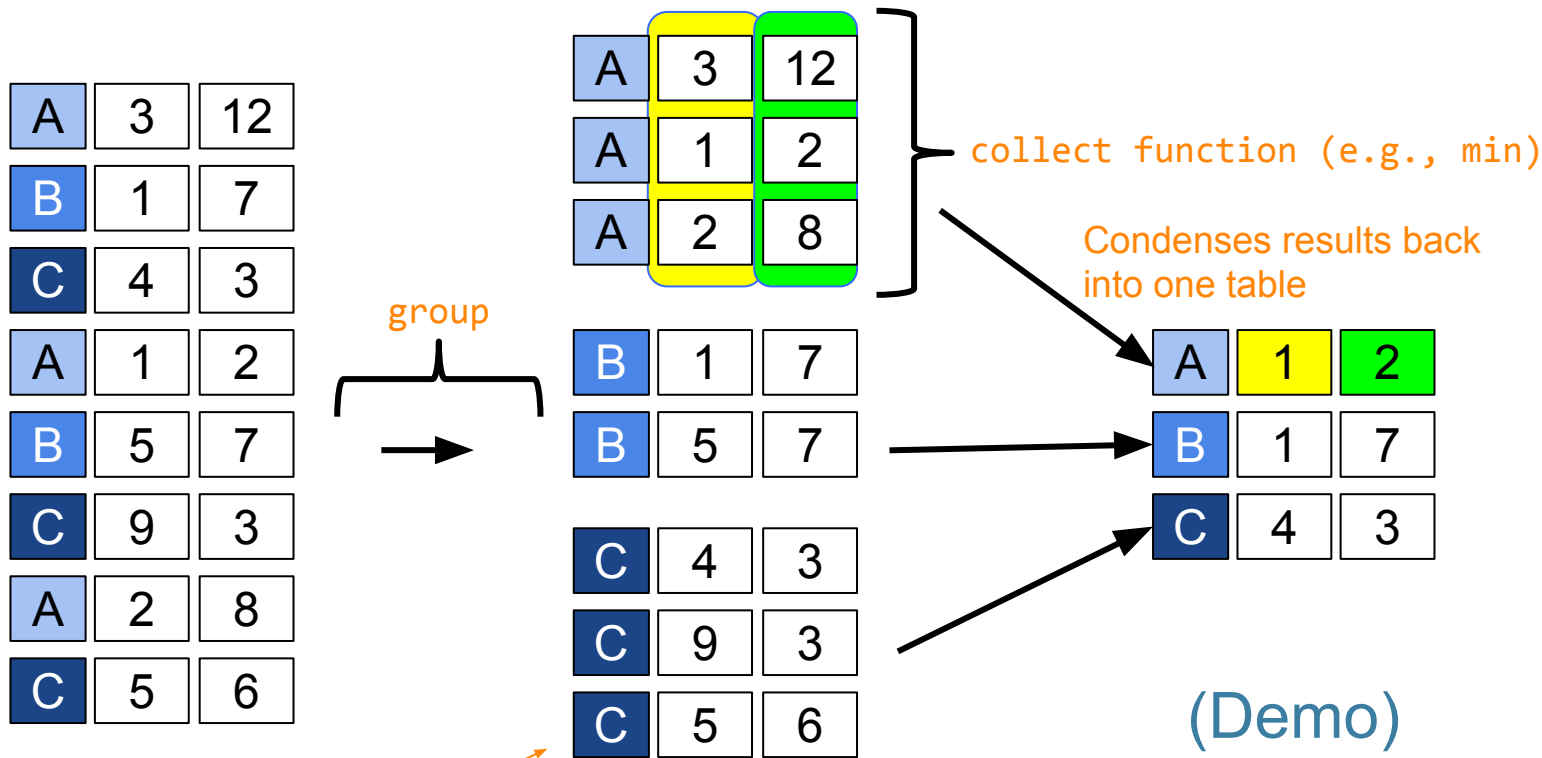
Can think of as temporary (A,B,C) sub-tables

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)