Lecture 09

Groups
Announcements
Example: Prediction
Sir Francis Galton

- 1822 - 1911 (knighted in 1909)
- A pioneer in making predictions
- Particular (and troublesome) interest in heredity
- Charles Darwin's half-cousin (Demo)
Apply with Multiple Columns
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)

```python
table_name.apply(one_arg_function, 'column_label')
```

```python
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)
Grouping by One Attribute
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

(Demo)
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
- Second argument: (Optional) How to combine values

(Demo)
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)
Challenge Question

Which NBA teams spent the most on their “starters” in 2017-2018?

Assume the “starter” for a team & position is the player with the highest salary on that team in that position.

<table>
<thead>
<tr>
<th>Player</th>
<th>Position</th>
<th>Team</th>
<th>Salary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stephen Curry</td>
<td>PG</td>
<td>Golden State Warriors</td>
<td>34.3826</td>
</tr>
<tr>
<td>LeBron James</td>
<td>SF</td>
<td>Cleveland Cavaliers</td>
<td>33.2857</td>
</tr>
<tr>
<td>Paul Millsap</td>
<td>PF</td>
<td>Denver Nuggets</td>
<td>31.2692</td>
</tr>
</tbody>
</table>
### Take-Home Question

Generate a table of the names of the starters for each team

<table>
<thead>
<tr>
<th>Team</th>
<th>C</th>
<th>PF</th>
<th>PG</th>
<th>SF</th>
<th>SG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta Hawks</td>
<td>Miles Plumlee</td>
<td>Mike Muscala</td>
<td>Dennis Schroder</td>
<td>Taurean Prince</td>
<td>Kent Bazemore</td>
</tr>
<tr>
<td>Boston Celtics</td>
<td>Aron Baynes</td>
<td>Al Horford</td>
<td>Kyrie Irving</td>
<td>Gordon Hayward</td>
<td>Jaylen Brown</td>
</tr>
<tr>
<td>Brooklyn Nets</td>
<td>Timofey Mozgov</td>
<td>Luis Scola</td>
<td>Jeremy Lin</td>
<td>DeMarre Carroll</td>
<td>Allen Crabbe</td>
</tr>
<tr>
<td>Charlotte Hornets</td>
<td>Dwight Howard</td>
<td>Marvin Williams</td>
<td>Kemba Walker</td>
<td>Michael Kidd-Gilchrist</td>
<td>Nicolas Batum</td>
</tr>
<tr>
<td>Chicago Bulls</td>
<td>Robin Lopez</td>
<td>Nikola Mirotic</td>
<td>Kris Dunn</td>
<td>Denzel Valentine</td>
<td>Justin Holiday</td>
</tr>
<tr>
<td>Cleveland Cavaliers</td>
<td>Kevin Love</td>
<td>Ante Zizic</td>
<td>George Hill</td>
<td>LeBron James</td>
<td>JR Smith</td>
</tr>
<tr>
<td>Dallas Mavericks</td>
<td>Dwight Powell</td>
<td>Josh McRoberts</td>
<td>J.J. Barea</td>
<td>Harrison Barnes</td>
<td>Wesley Matthews</td>
</tr>
<tr>
<td>Denver Nuggets</td>
<td>Mason Plumlee</td>
<td>Paul Millsap</td>
<td>Devin Harris</td>
<td>Wilson Chandler</td>
<td>Mike Miller</td>
</tr>
<tr>
<td>Detroit Pistons</td>
<td>Andre Drummond</td>
<td>Blake Griffin</td>
<td>Reggie Jackson</td>
<td>Stanley Johnson</td>
<td>Langston Galloway</td>
</tr>
<tr>
<td>Golden State Warriors</td>
<td>Zaza Pachulia</td>
<td>Draymond Green</td>
<td>Stephen Curry</td>
<td>Kevin Durant</td>
<td>Klay Thompson</td>
</tr>
</tbody>
</table>