Lecture 10, September 16

Classifying Categorical Data

Slides created by Ani Adhikari and John DeNero

DATA 8
Fall 2016

Announcements

- **Project 1** will be released next week. We're working to take care of those annoying server problems.

- **John DeNero** will give the lecture on Monday 9/19. Show up!

- 469 students filled out the survey. **Thank you!**

- According to the survey, today is the birthday of two students. Happy birthday!

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
    - **np.mean** — mean of all grouped values

```
table_name.group('column_label', function)
```

(Demo)

Groups

The **groups** method aggregates all rows that share the combination of values in multiple columns

- First argument: Which columns to group by
- Second argument: (Optional) How to combine values

```
table_name.groups(column_labels_array, function)
```

(Demo)

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 (both must be specified, or neither)
  - **values**=’column_label_to_aggregate’
  - **collect**=function_with_which_to_aggregate