Histograms, Data Density, Functions

Prof. Abby Stylianou

Logistics

- Homework 2 due next THURSDAY
  - want to make sure we have another office hour before it's due

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')
```

Apply w/ Multiple Arguments

- The `apply` method creates an array by calling a function on every element in input column(s)

```
table_name.apply(no_arg_function)
table_name.apply(one_arg_function, 'column_label')
table_name.apply(two_arg_function,
                 'column_label_for_first_arg',
                 'column_label_for_second_arg')
```

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

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

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

Prediction: Sir Francis Galton

- 1822 – 1911 (knighted in 1909)
- A pioneer in making predictions
- Particular interest in heredity (father of eugenics)
- Charles Darwin's half-cousin