
MATH1401

Fall 2021

Lecture 10

Groups

Class Checklist

- **HW 3 - Due Date** : Friday: 9/24 – 9 PM
 - Graded Questions: 1.1-1.7, 2.1-2.5, 2.7, 3.1, 3.3, 4.3-4.5
- **Lab 3 – Due Date** : Thursday 9/30 – 9 PM
- **Quiz 7** – Thursday: 9/23 – Covers Chapter 8

Review of Lecture 9

Histogram

- Distribution of numerical variable
 - Horizontal axis is numerical: bins can be unequal
 - **Area of bars** proportional to the percent of individuals
 - **Height of bars** measures density
-

Area and Height

$$\text{Area of bar} = \% \text{ in bin} = \frac{\text{Number of individuals in bin} \times 100}{\text{Total number of individuals}}$$

$$\text{Height} = \frac{\% \text{ in bin}}{\text{width of bin}} = \frac{\text{Area of bin}}{\text{width of bin}}$$

Defining Functions

Def Statements

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

(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')
```

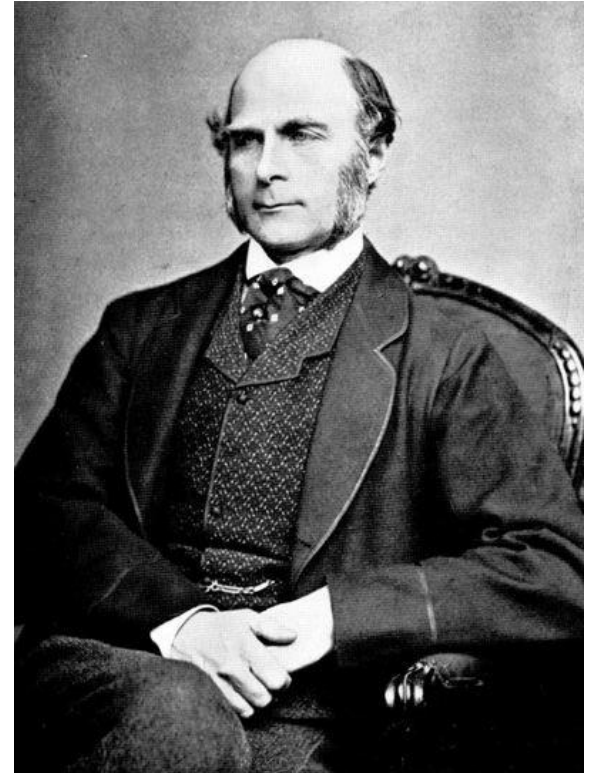
(Demo)

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)



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

(Demo)
