

FIT1043 Introduction to Data Science

Week 3: Python Aggregation

Dr. Sicily Ting

School of Information Technology
Monash University Malaysia

With materials from Wray Buntine, Mahsa Salehi

Learning Outcomes

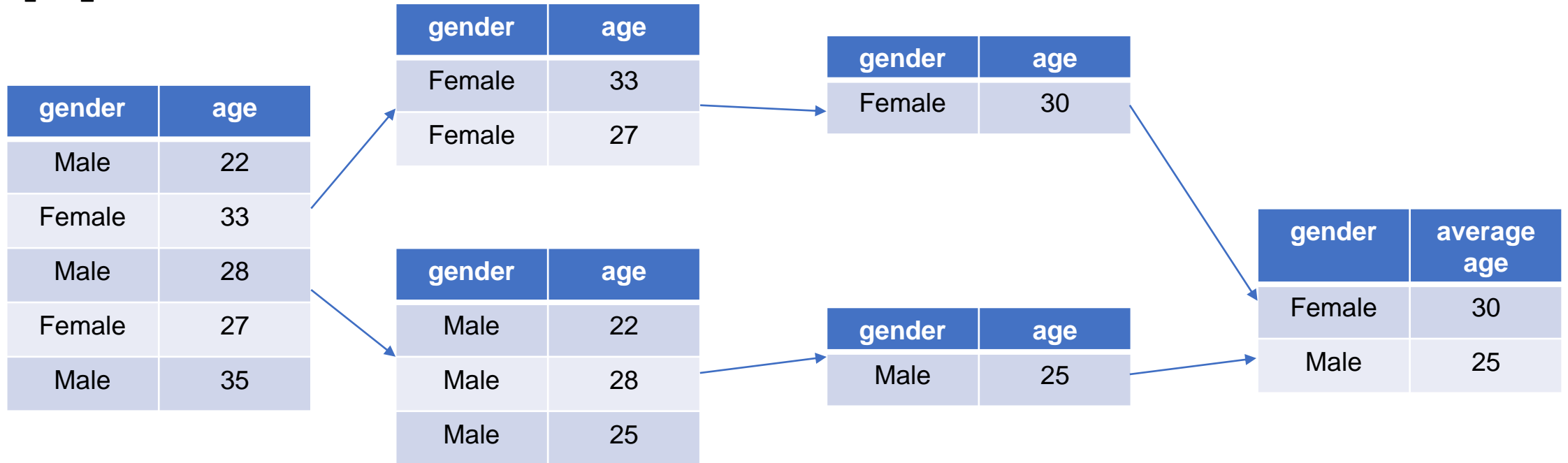
Week 3

By the end of this week you should be able to:

- **Comprehend more sophisticated group-by operations and graphing in Python**
- Comprehend the power/importance of data visualisation
- Differentiate between **approaches for data visualisation**, and explain where each approach is appropriate to be used
- Explain/differentiate different concepts in **descriptive statistics**

Python Aggregation

`groupby()`



`groupby('gender')['age']`

`groupby('gender')`

`groupby('gender')['age'].mean()`

Input

Split

Apply (mean)

Combine

Python Aggregation

A more complicated aggregation example

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck	embark_town	alive	alone
0	0	3	male	22.0	1	0	7.2500	S	Third	man	True	NaN	Southampton	no	False
1	1	1	female	38.0	1	0	71.2833	C	First	woman	False	C	Cherbourg	yes	False
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False	NaN	Southampton	yes	True
3	1	1	female	35.0	1	0	53.1000	S	First	woman	False	C	Southampton	yes	False
4	0	3	male	35.0	0	0	8.0500	S	Third	man	True	NaN	Southampton	no	True



```
titanic.groupby('class')['who'].count()
```

age	
class	
First	186
Second	173
Third	355

Python Aggregation

A more complicated aggregation example

We can use the `agg()` function

<https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.agg.html>

```
titanic.groupby('class').agg()
```

What can you put in the `agg()` function? Let's relook at the previous, where we have the `count()` of the number of entries for the `age` column. It can be written as

```
titanic.groupby('class')[who].agg('count')
```