COMP20008 Workshop 1

COMP20008

Elements of Data Processing
Zijie Xu

Agenda

- Self-introduction
- First subject in Data Science
- Jupyter Notebook
- Pandas
  - Series
  - DataFrame
  - Group by

Introduction

- My name: Zijie Xu (also call me Jerry)
- Just completed Master of Computer Science
  - Research topic: Explainable ML for enzyme function prediction
- My (current) favourite emoji is： 🤪
- A fact about myself…

- Your turn!

Contact

Lectures

- 2 x 1-hr lectures on Friday
- Lecturers: James Bailey and Eduard Hovy

Workshops

- 1.5hr workshop
- Attendance strongly recommended (Bonus mark up for grabs!)

Assessment

Assignment 1

- Individual
- Release in about week 3?

Assignment 2

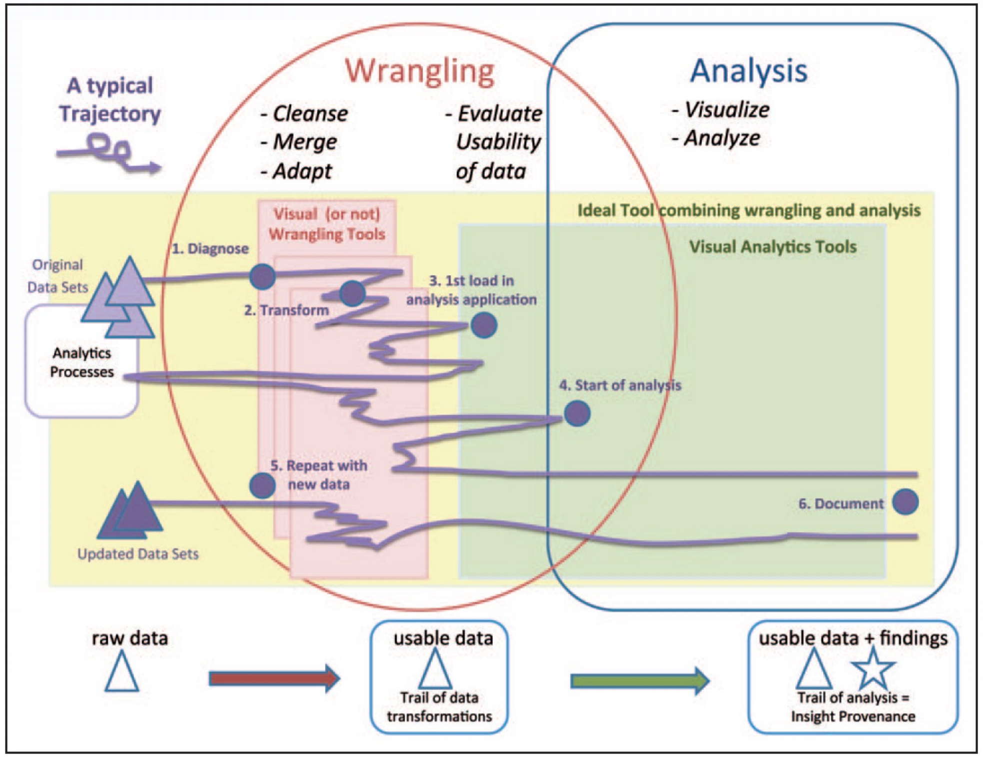
- Group project
- **<u>Team forming session today!!</u>**
- All group members must be in same workshop

Final exam

- 2hr in-person exam

First subject in Data Science

- “Foundations of Data Science”
- Prerequisite to COMP30027 Machine Learning
- Learn about the pipeline of data science



Research directions in Data Wrangling:
visualisations and transformations for crediible data. S. Kandel et al, Information Visualisation 10(4), 2011.

Jupyter Notebook

- Great tool for displaying code projects along with text and visualisations (hence “notebook”)
  - De facto industry standard
  - A server-client app with a backend computing kernel + frontend website for editing

- A notebook consists of
  - Markdown cells: text, figures, HTML etc.
  - Code cells: runnable python code blocks

Pandas

- A software package for data analysis and manipulation in python
  - Open-sourced since 2009
  - Also de facto industry standard (for now)
  - Key benefits: fast, flexible & memory-efficient

```
import pandas as pd
```

Series and DataFrame

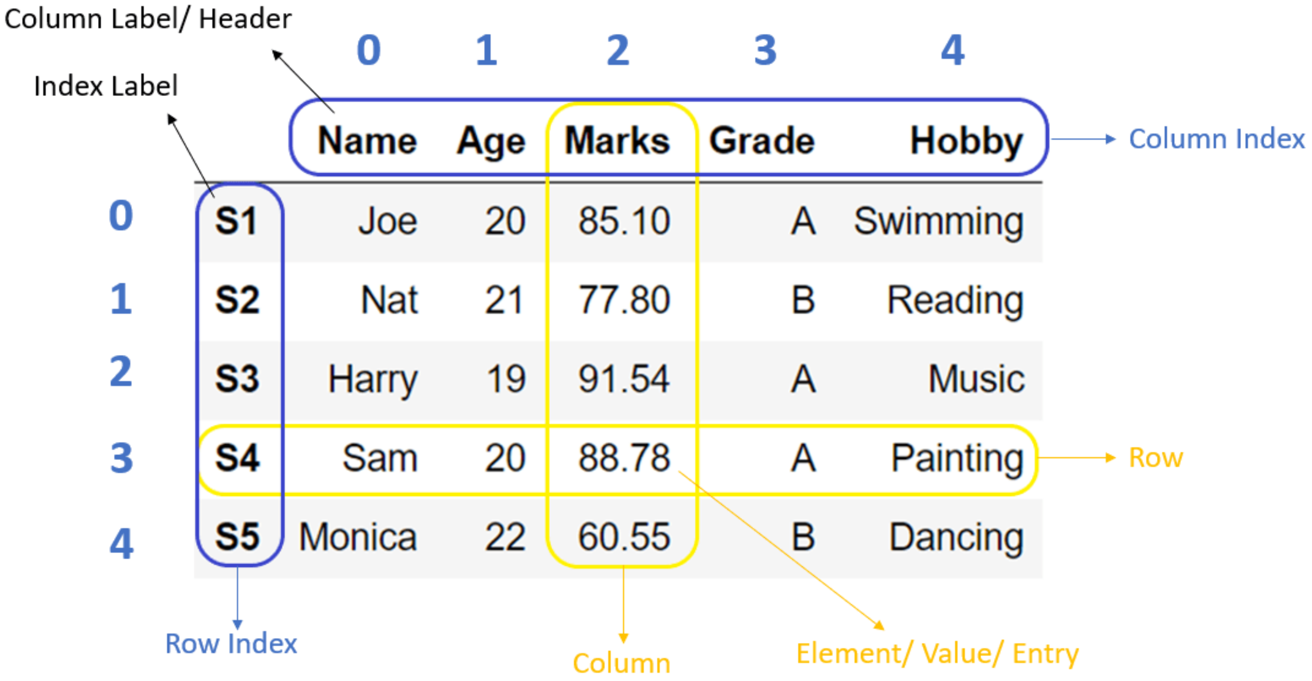
`pd.Series`

- A column of data

`pd.DataFrame`

- A spreadsheet
- Multiple columns of data

Check API documentation



	Name	Age	Marks	Grade	Hobby
S1	Joe	20	85.10	A	Swimming
S2	Nat	21	77.80	B	Reading
S3	Harry	19	91.54	A	Music
S4	Sam	20	88.78	A	Painting
S5	Monica	22	60.55	B	Dancing

THE UNIVERSITY OF MELBOURNE

Thank you

More Resources: Canvas