

Data analytics for personalized genomics and precision medicine

Data & Python

Lecturer: Yu LI (李煜) from CSE

Liyu95.com, liyu@cse.cuhk.edu.hk Friday, 9 September 2022

Expected outcomes:

- Different data types
- Introduction to Python programming

Pre-course survey results:

- Question 1: If we do two scribing, which one would be graded for the 10%? the first one, the second one or the best one?
 - Answer: The best one would be graded.
- Question 2: Will tutorial be recorded as well? How can we access video recording?
 - Answer: Tutorial will be recorded and TA will teach how to access the video.
- Comments from students:
 - Recommend focusing more on data rather than on biology
 - The course is with detailed and excellent information
- Recommendations from the lecturer:
 - Do the Scribing until 12 September

Data types

- Sequential data
 - Dynamic programming
- Data matrix
 - consists of collection of records and fixed set of attributes
 - genome assembly and mapping
- Spatial data
 - geographic locations and spatial information involved
- Temporal data
 - data involving time
- Graph or networks
 - objects and connections
- Text
 - short or long sentences or documents
- Multi-modality data
 - eg. video/ electronic health records/ spatial transcriptomics
- Unknown data type

- eg. diet and exercise

Python

- Computer programming
 - “ is the process of designing and building an executable computer program to accomplish a specific computing result or to perform a specific task.”---
Wikipedia
- Python
 - “is an interpreted high-level general-purpose programming language.”---
Wikipedia
- Library
 - Numpy/ Scipy/ Pandas
- Codes

```
• import numpy # import the numpy library
•
• a = [1,2,3,4,4,5,5,5,6,7,8,9] # create an array called "a"
• a_mean = numpy.mean(a) #4.916666666666667 <- calculate the mean
• a_std = numpy.std(a) #2.253084305765962 <- calculate the
• a_med = numpy.median(a) #5.0 <- calculate the median
• a_max = numpy.max(a) #9 <- calculate the maximum number among the array
• print(a) #[1,2,3,4,5,5,5,6,7,8,9] <- print out the array
•
• print("The a array is ", a) #The a array is [1,2,3,4,5,5,5,6,7,8,9]
• print("Its mean is ", a_mean) #Its mean is 4.916666666666667
• print("Its mean is ", numpy.mean(a)) #Its mean is 4.916666666666667
```