Data Processing

**Original files**

*AY1415 Evening Prelims by Instructor (xlsx)*

* Number of prelims for each distinct course
* Date of the prelims

(for determining possible bin input in the new model)

*qrySP15PrelimEnrollmentFile (xlsx)*

* Course index

Using Combination type, combination id and Course id to identify distinct courses and assigning indexes.

* Course size

Used for capacity constraint

* Individual enrollment *(Module.xlsx)*

Replacing the student id and course with indexes.

(to derive the course pairwise overlap matrix.)

**Results**

*Reference for courses*

* Course ID
* Course index
* Number of prelims for each course

*Overlap Matrix*

* Student id
* Course id

Indexing Course ID

Step 1: find distinct courses

* Course type Blank: use course id
* Course type Cross list (C): use course id
* Course type Co-meet (W): use combined section id
* Course type Both (B): use combined section id

Step 2: indexing courses

* Index type C and Blank first
* Index type W and B

Step 3: combine indexed courses and make the reference file

Indexing Student ID

Step 1: remove duplicates in AnonID column

Step 2: indexing as student ID and match with the course ID