

Exploratory Data Analysis for Learning Analytics

Newcastle University - CSC8631 Summative Assignment

December 2020

Summary

words

Enrolment analysis

A dataset of numerous files in a variety of formats, compressed into a single folder, was provided by the Newcastle programme team for analysis. The material related to a series of online cyber security training courses, entitled Cyber Security: Safety at Home, Online, in Life. This report summarises the data management and exploratory analysis undertaken using the enrolment files. Given that the training subject was online security I was interested to determine students' willingness to share information relating to gender, employment and education.

Seven enrolment files were provided for the courses that ran over a period of two years. The first student enrolled on 29 March 2016, the last student enrolled on 01 November 2018 indicating that the course series for a period of over two and a half years.

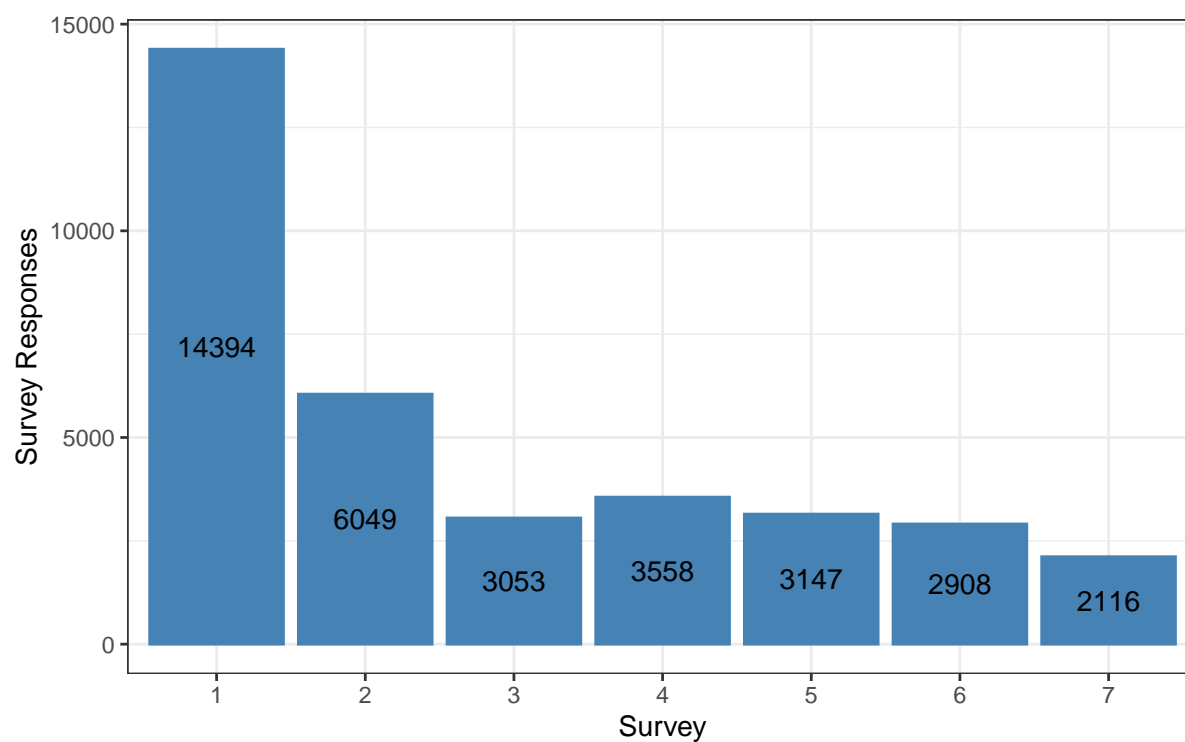
Enrolment - key statistics

In total, 35225 unique student enrolment records were assessed from seven survey files. The first survey file had the most enrolment records. The 'enrolled at' dates within the file indicate the first intake ran for the longest time.

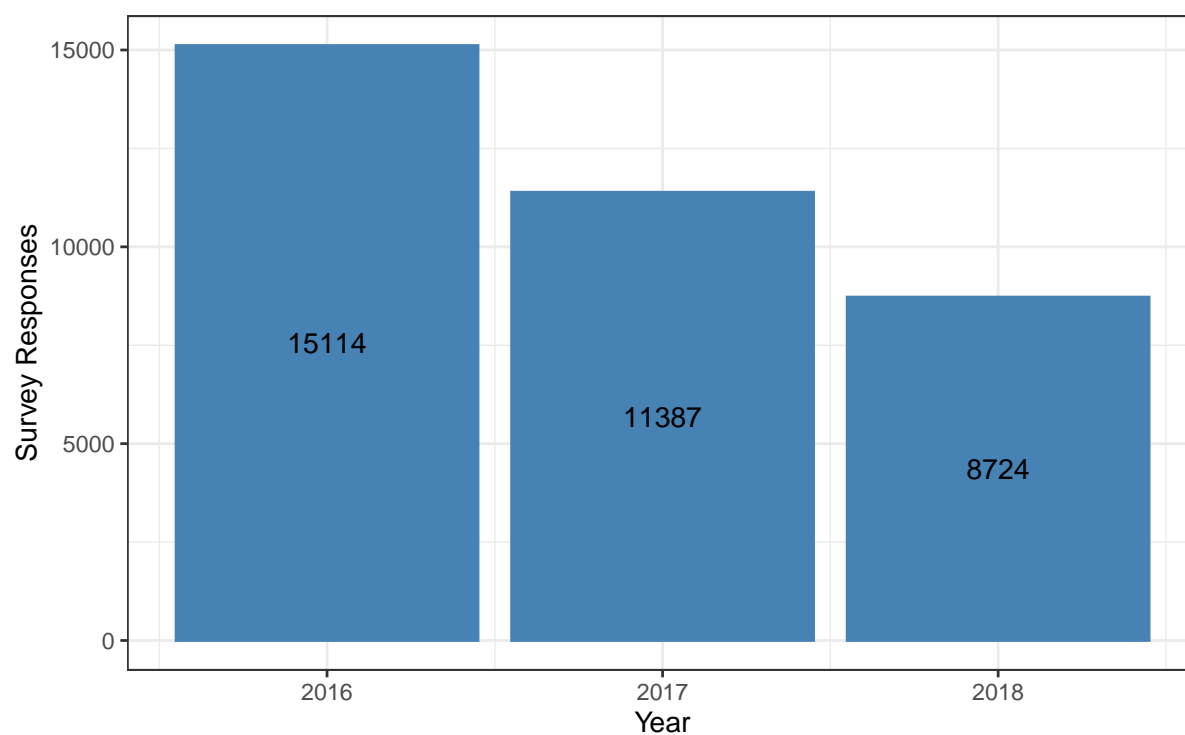
Survey No.	Date First Student Enrolled	Date Last Student Enrolled	Duration (days)	No. of Students
1	29 March 2016	07 September 2017	527 days	14394
2	05 December 2016	13 July 2017	220 days	6049
3	02 July 2017	26 February 2018	239 days	3053
4	27 July 2017	25 January 2018	182 days	3558
5	15 December 2017	09 September 2018	268 days	3147
6	08 April 2018	11 August 2018	125 days	2908
7	25 June 2018	01 November 2018	129 days	2116

The 'enrolled at' dates within the files indicate that the courses overlapped in duration and varied in length, with the volumes of students applying for each course gradually reducing. Whilst enrolments within file four started after the third intake the last enrolment in the fourth file was prior to the last enrolment in file three.

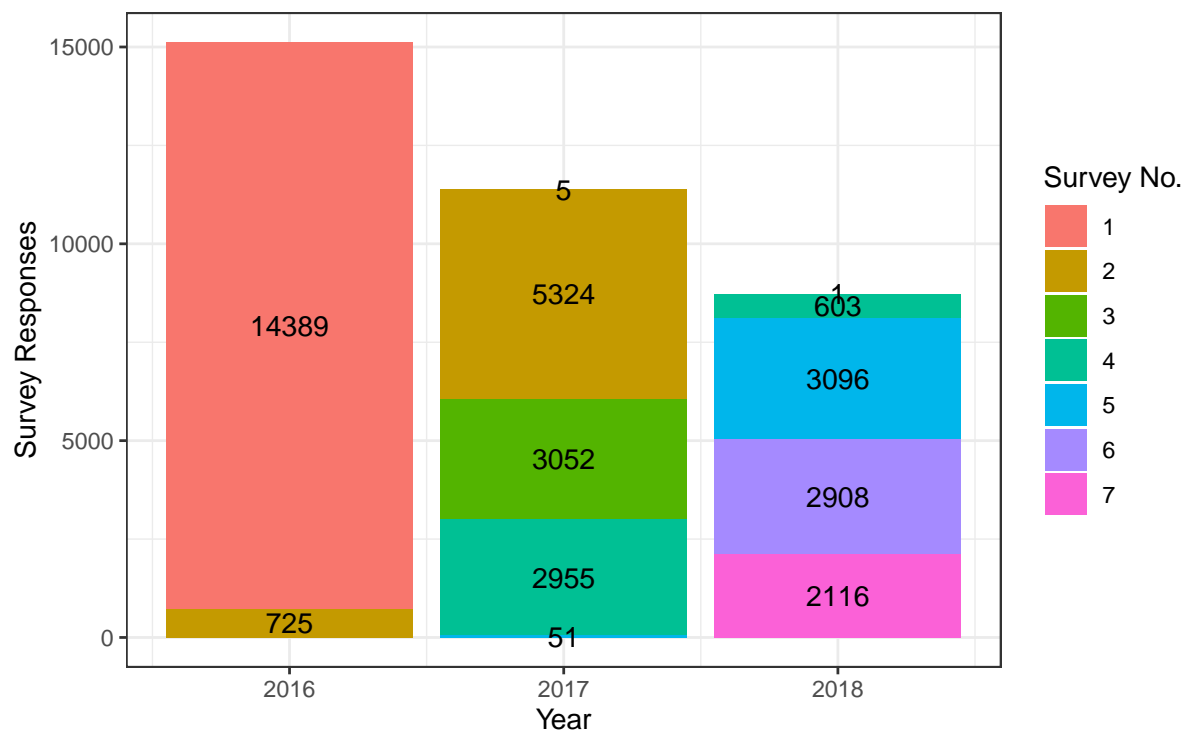
Enrolment surveys by course



Enrolment surveys by year



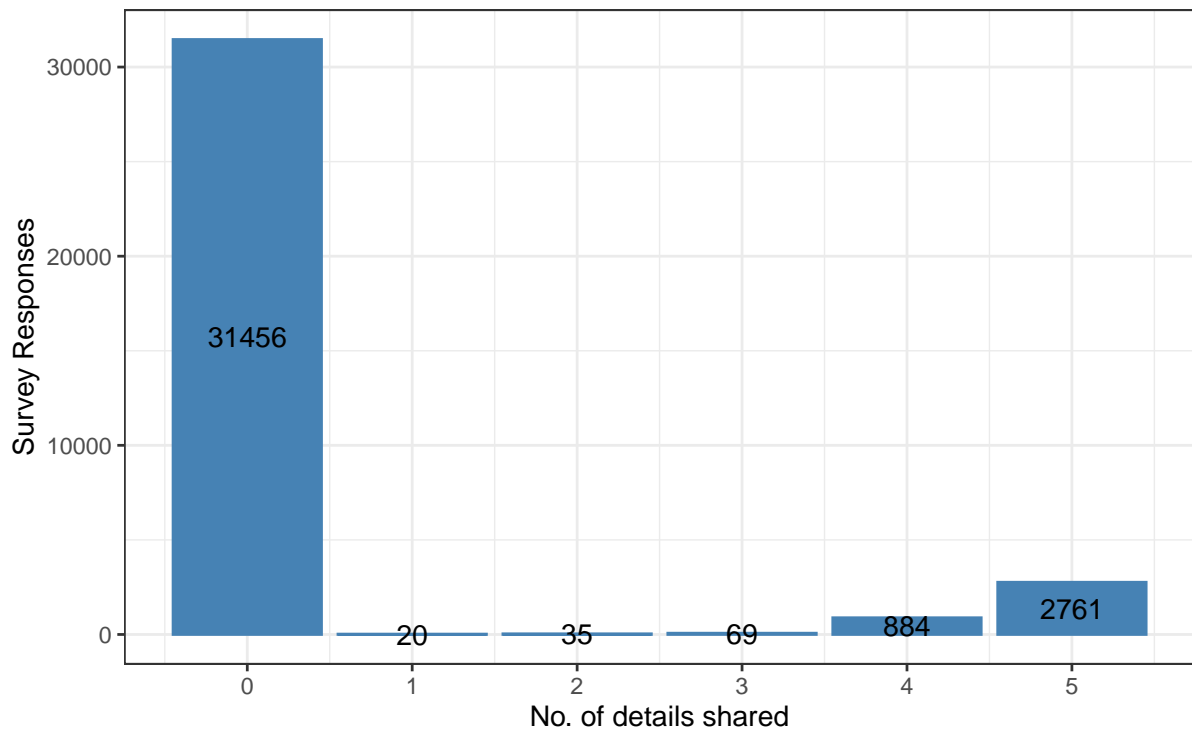
Survey responses by course and year



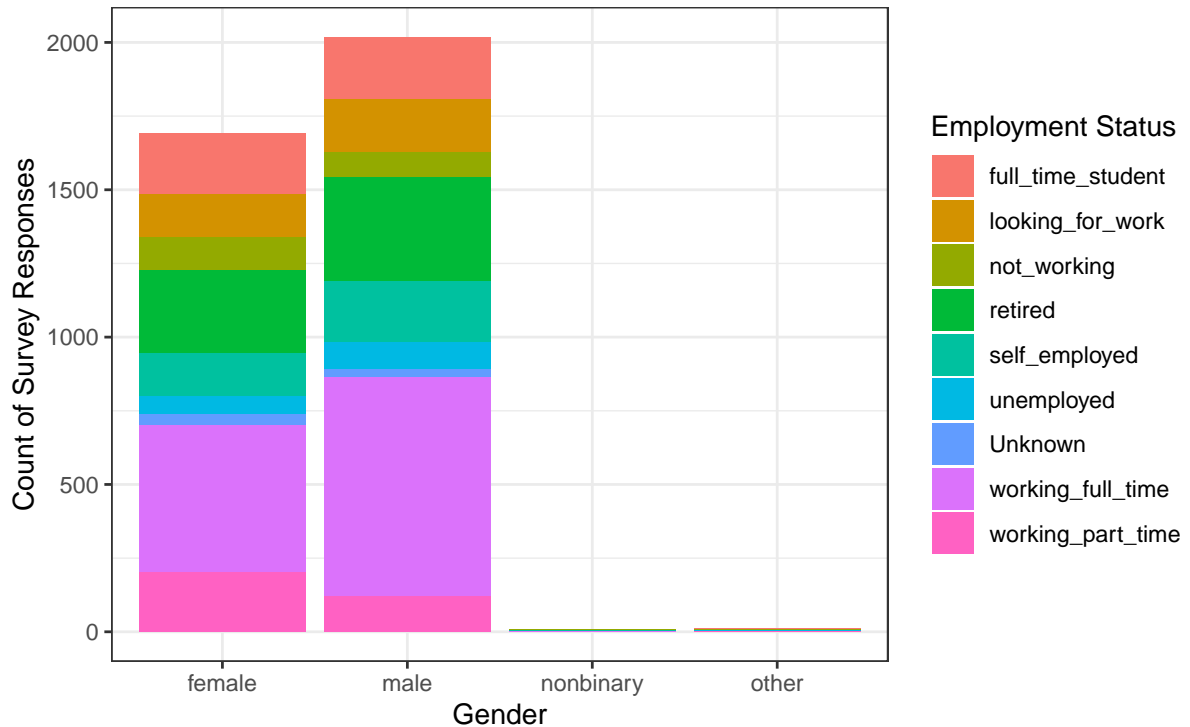
Students were asked for information on gender, age range, employment status, employment area and highest education level. Analysis was conducted on these five details to determine willingness to share.

No. of students that shared data: 3769 For students who shared data, they shared an average of 0.5007239 items.
 No. of students sharing gender: 3733 No. of students sharing age range: 3620 No. of students sharing employment status: 3689 No. of students sharing employment area: 2881 No. of students sharing highest education level: 3715

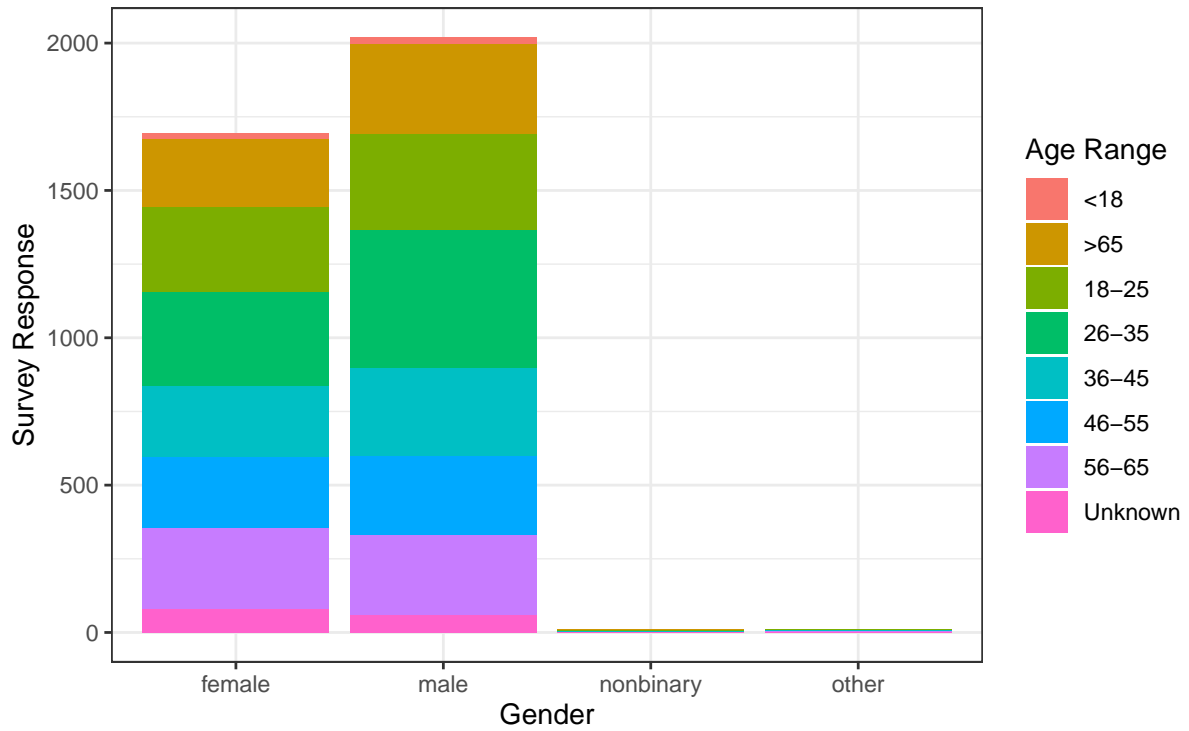
Amount of information shared by students



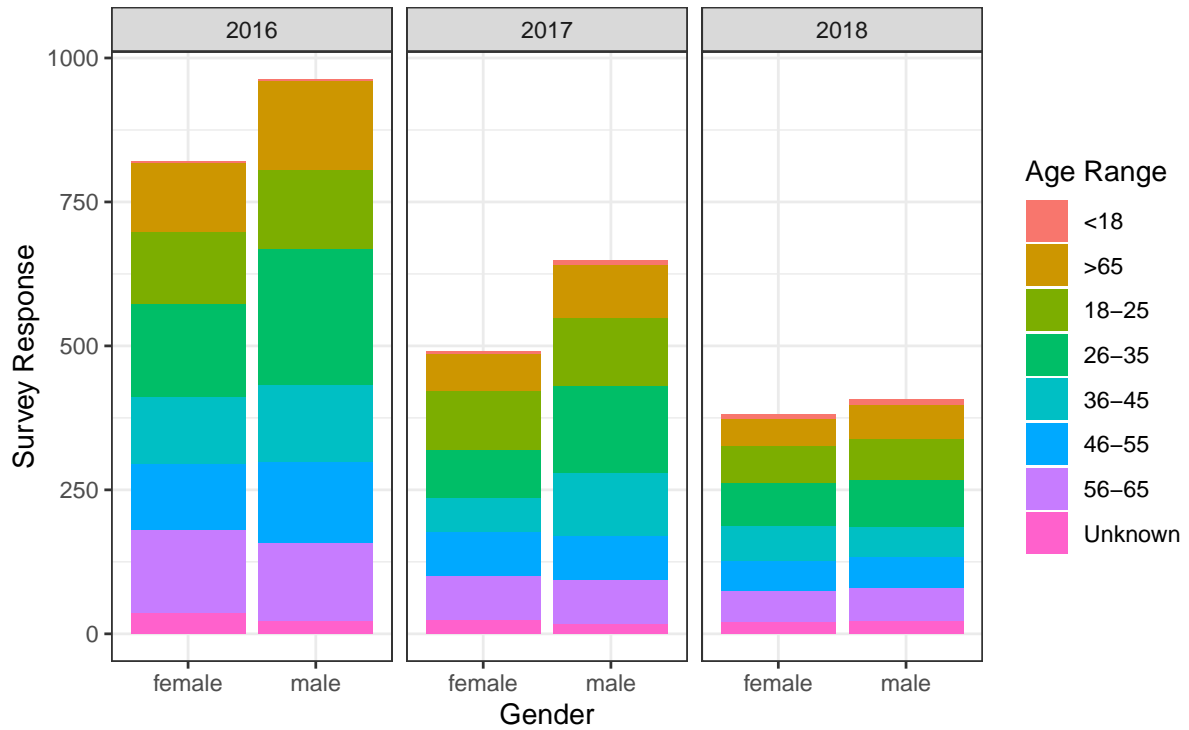
Students that provided gender



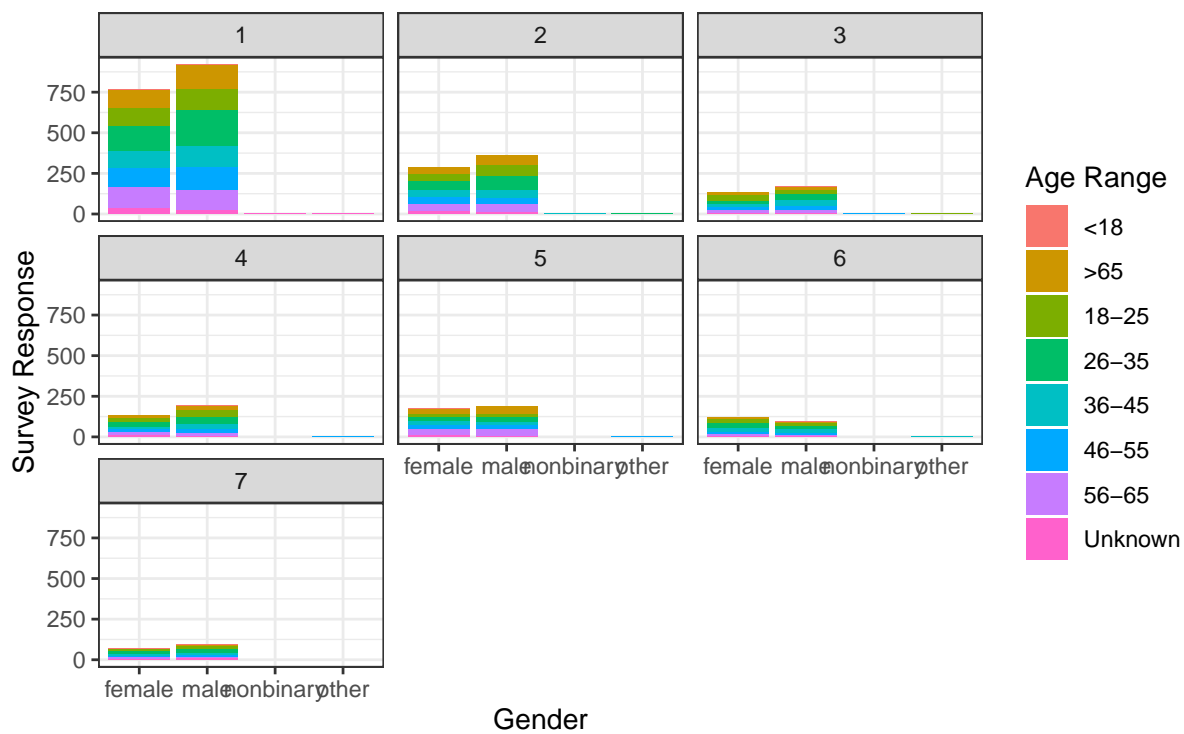
Students that provided gender and age range



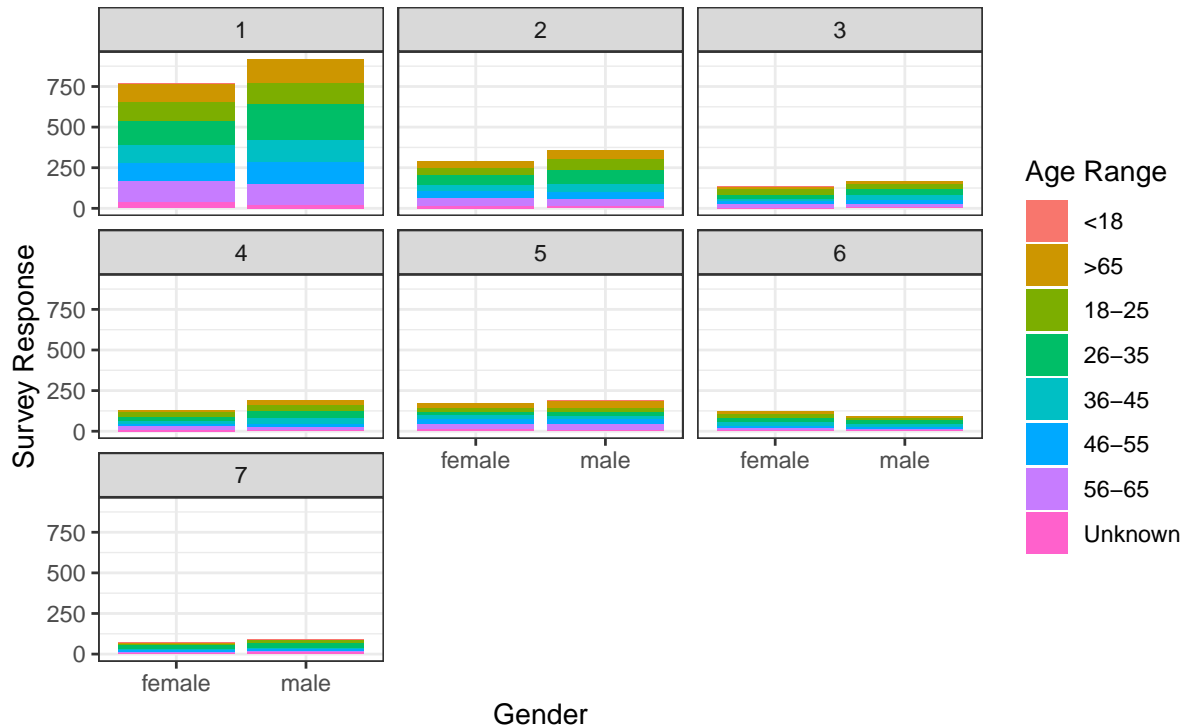
Male or female students that provided age range



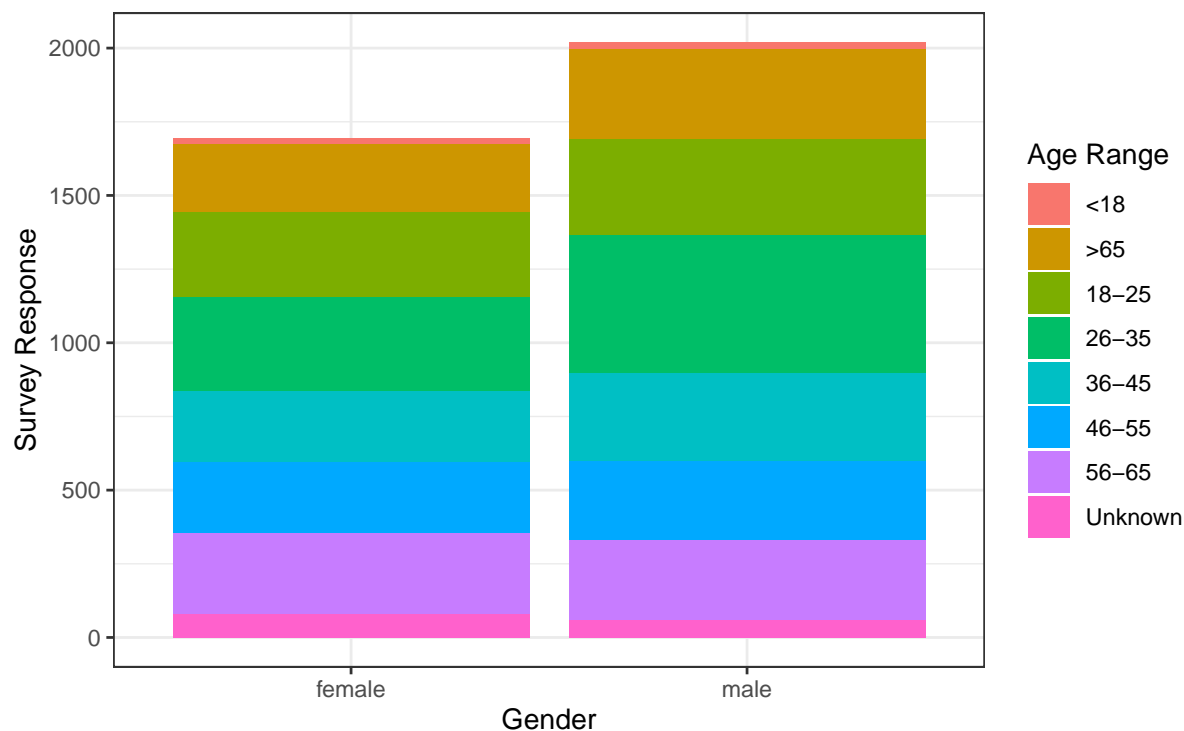
Students that provided gender and age



Male or female students that provided age range

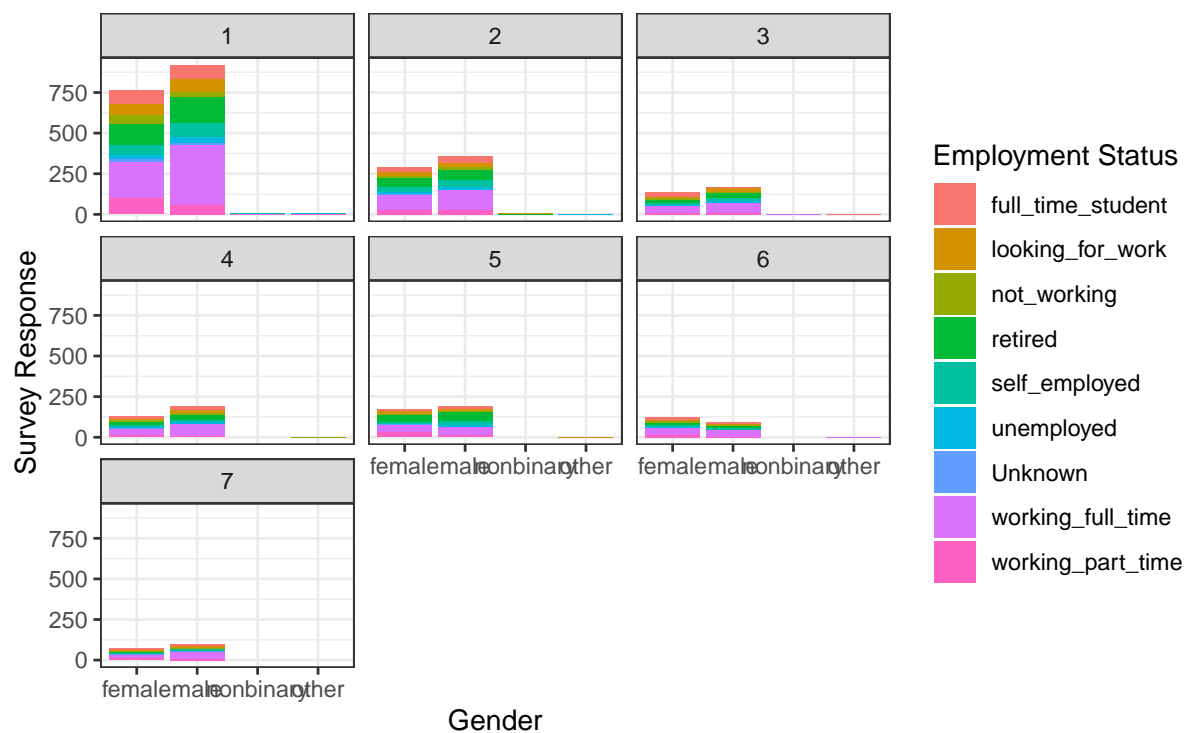


Male or female students that provided age range



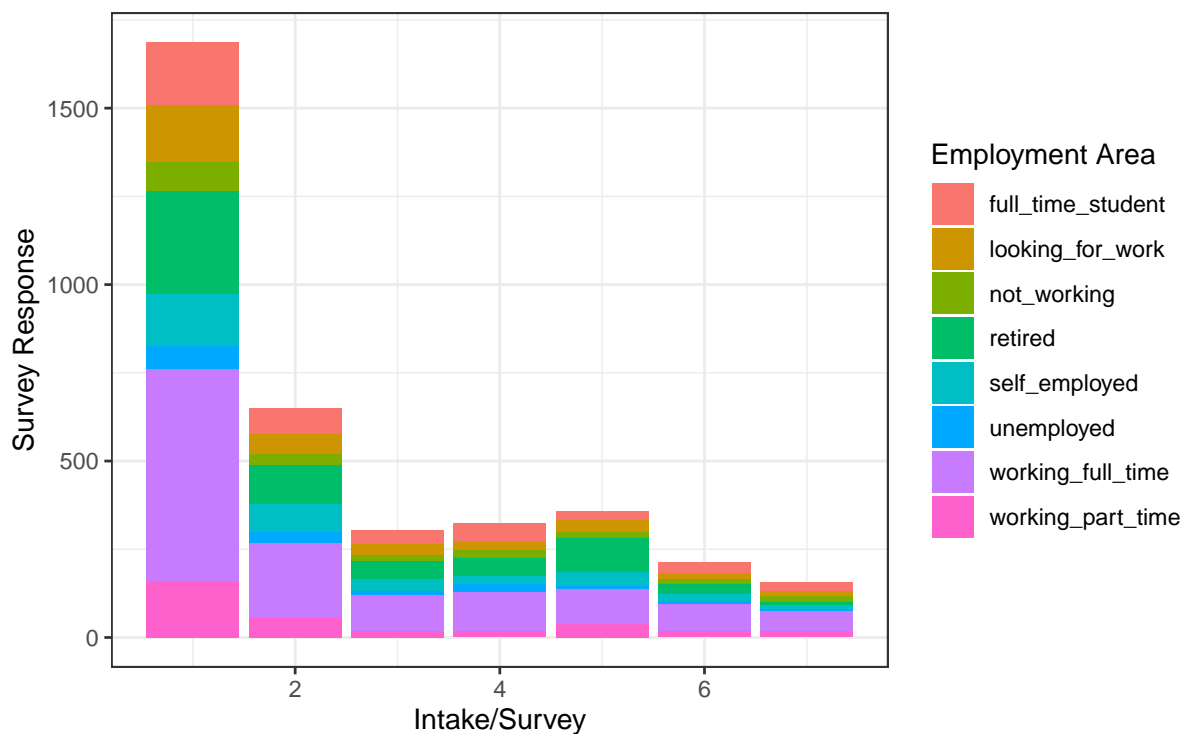
words

Students that provided gender and employment status

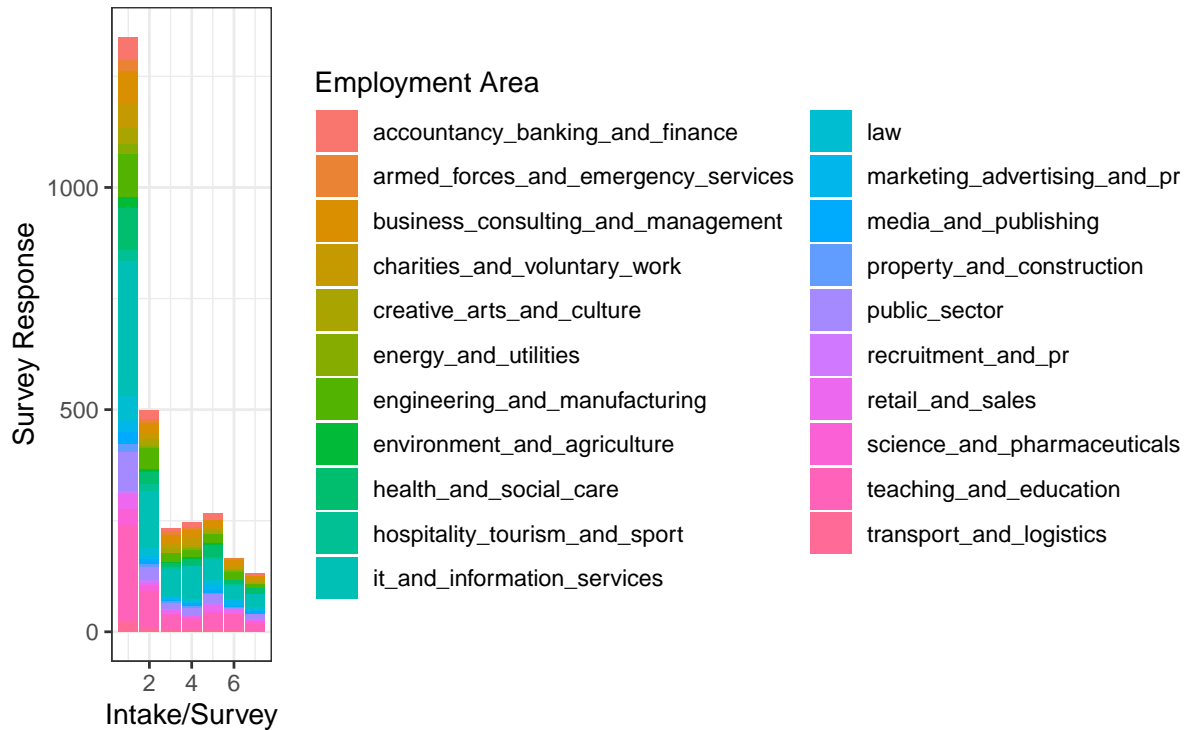


Words about employment graphs to follow

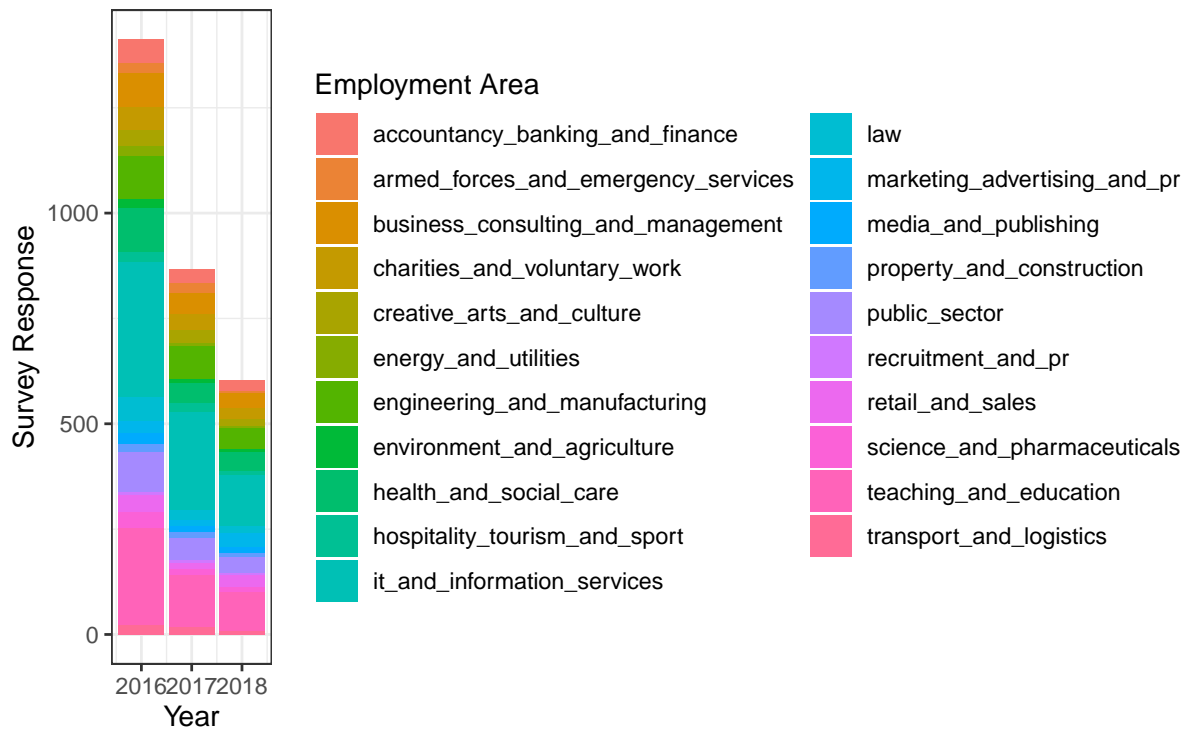
Students that Provided Information on employment status



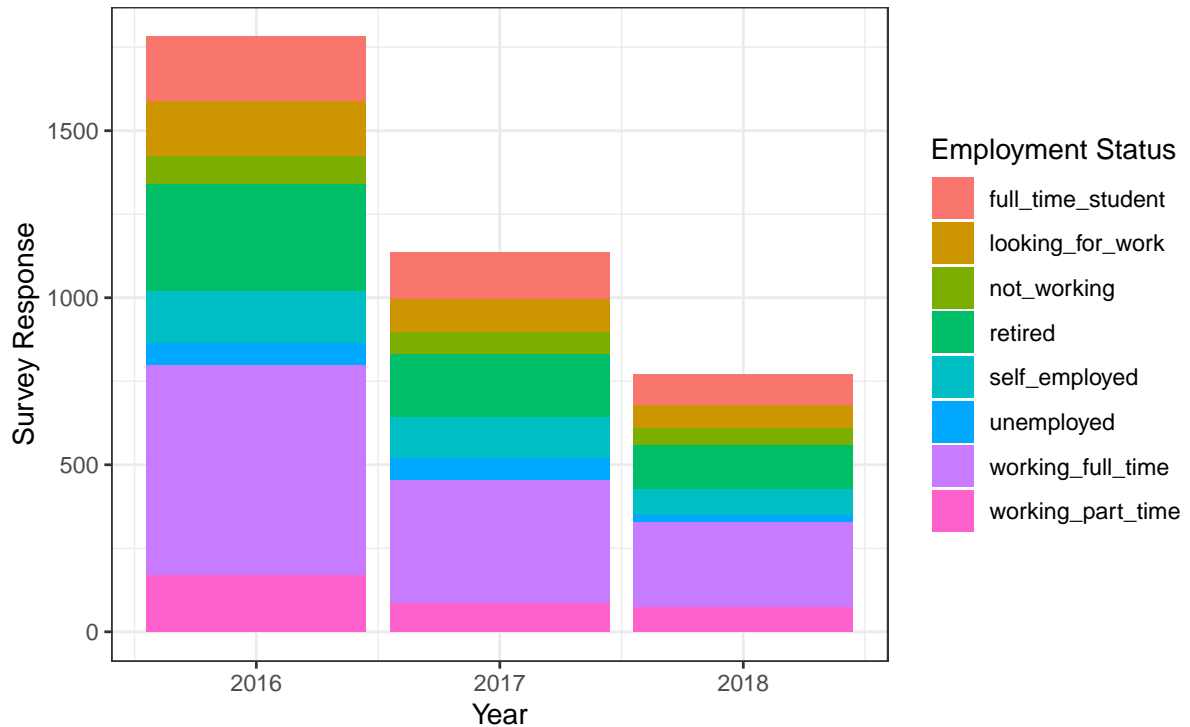
Students that provided employment area



Students that provided information on employment area



Students that provided information on employment status

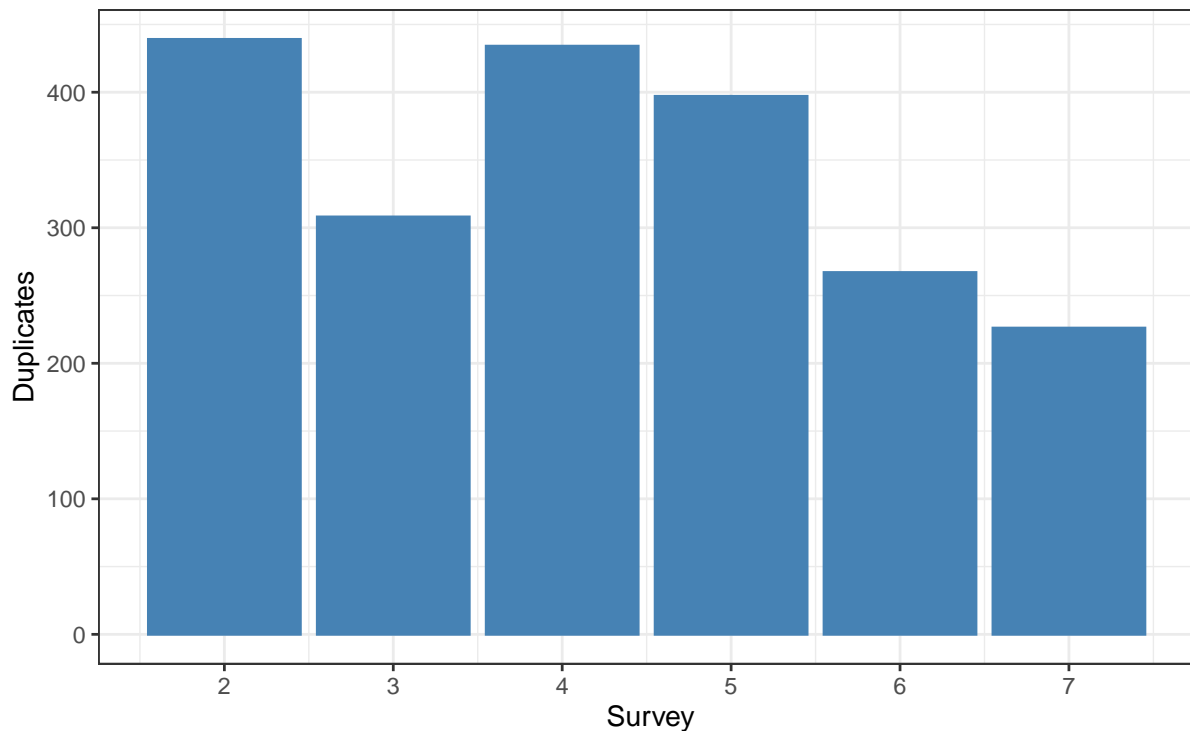


Data and assumptions

Assumptions

- It is assumed that duplicated learner_id records are exact copies of the same survey response. In total 2071 non-unique learner_id records were removed from the dataset prior to analysis. It was expected that the learner_id being 36 characters in length was expected to be unique and represents a single student. Example, 233b9253-e8e6-4734-8a7f-cb2c7845c85a.

Duplicate enrolment records by survey



- It is assumed that each record within the cyber security enrolments files corresponds to a student who provided details that truthfully reflected their gender, employment status and area, age range and highest educational level.
- It is assumed that the data contained within the cyber security enrolments files was provided before the training had commenced.
- It is assumed that each enrolment file relates to a course intake.

Data summary

Issy Middleton - C1000051
Newcastle University - CSC8631 Summative Assignment
December 2020

Notes

Words