

# 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 belonged to Learning Analytics and related to a series of an online training course, entitled Cyber Security: Safety at Home, Online, in Life. This report summarises the findings and insights relating to the survey students completed upon enrolment. Given that the training subject was online security I was interested to determine students' attitudes before the course began, and their willingness to provide personal data.

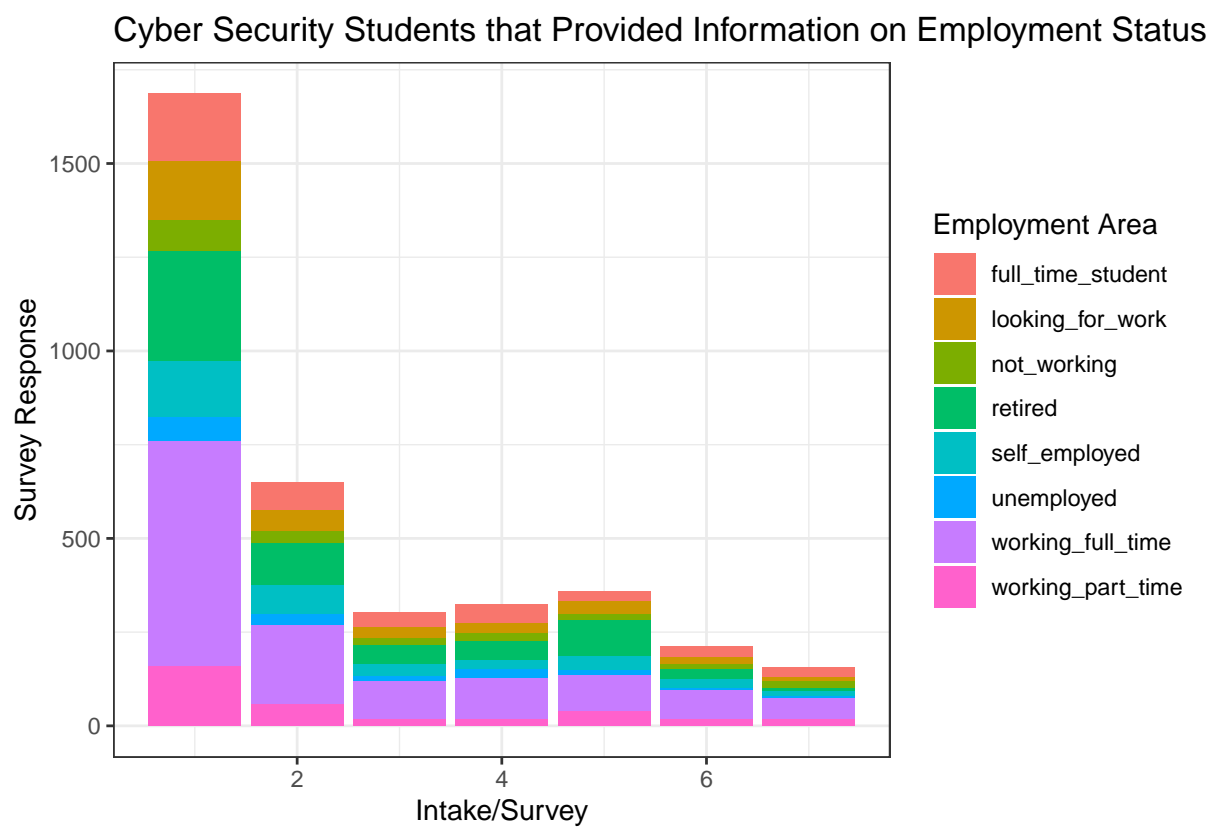
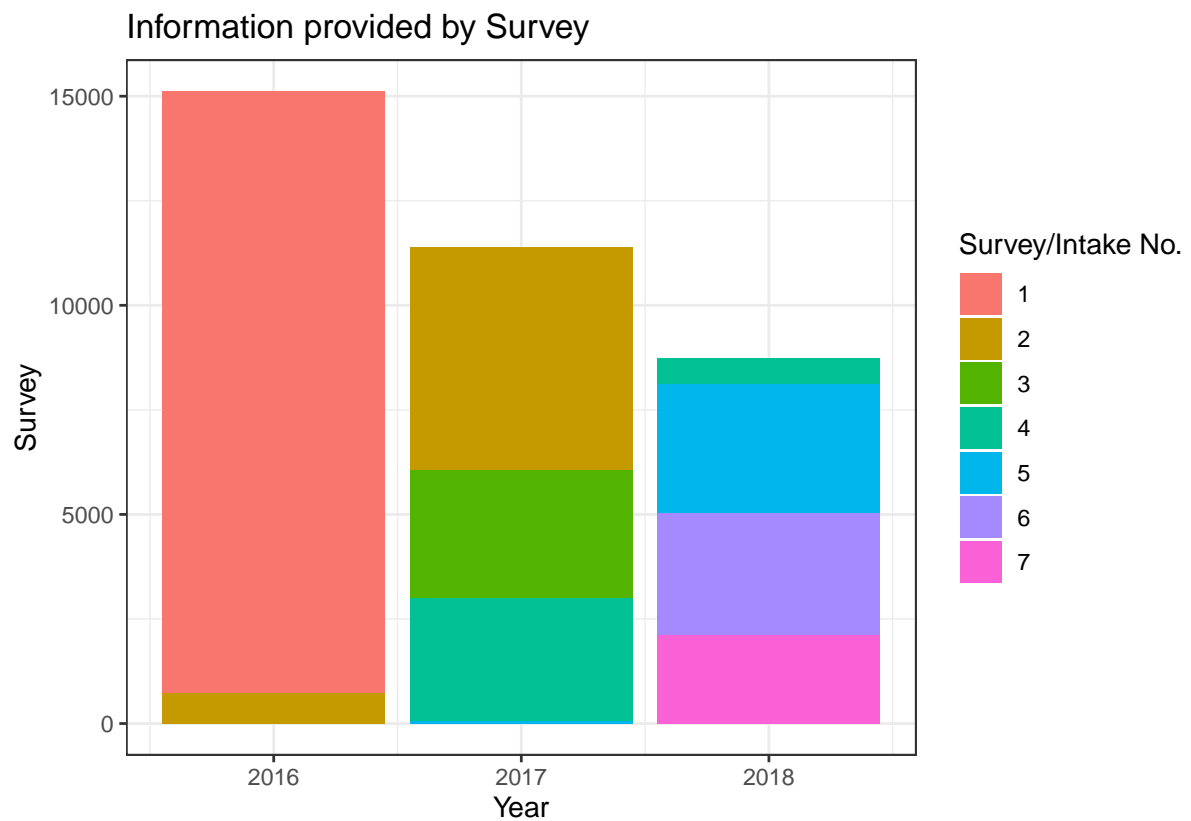
Seven intakes of the course ran over a period of two years, with the first student enrolled on 29 March 2016, the last student enrolled on 01 November 2018 meaning the seven course series ran over a period of over two and a half years.

## Enrolment - key statistics

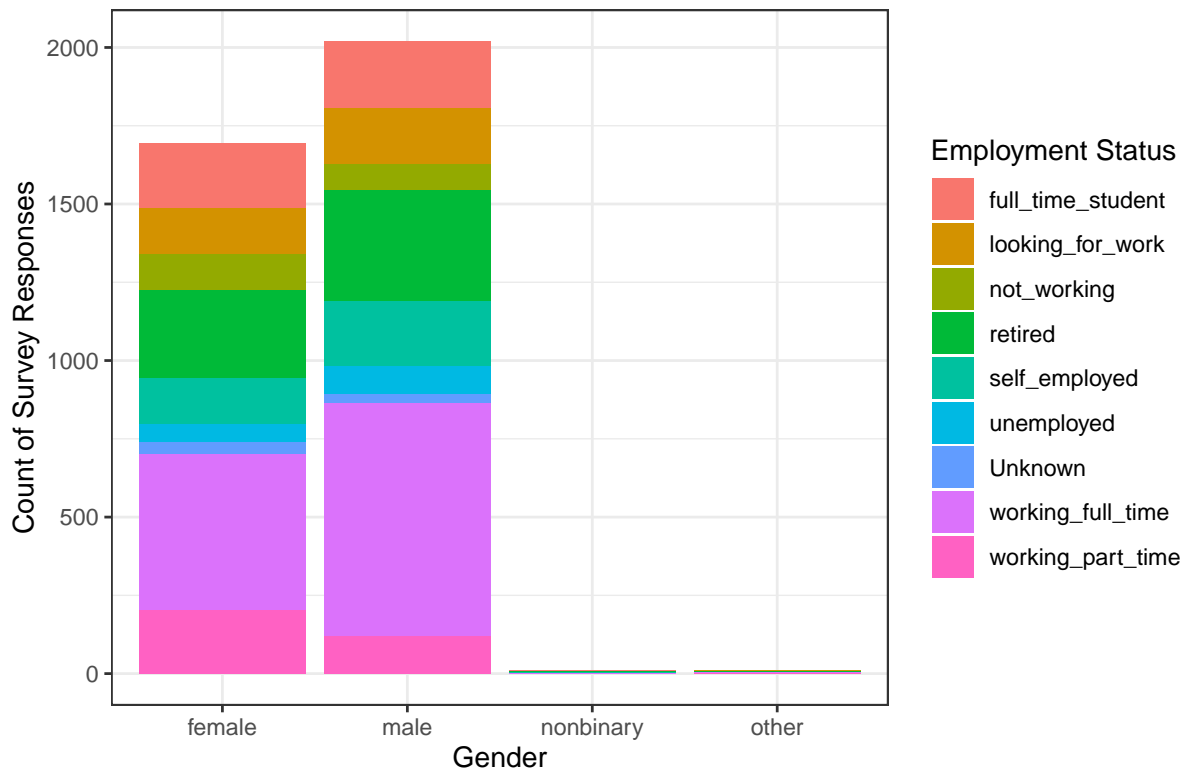
The first course intake had the most student enrolment survey records but was also the longest in duration. In total 35,225 unique student enrolment records were assessed across the seven intakes. The courses overlapped in duration and varied in length, with the volumes of students applying for each course gradually reducing.

Survey	First Student Enrolled	Last Student Enrolled	Duration (days)	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

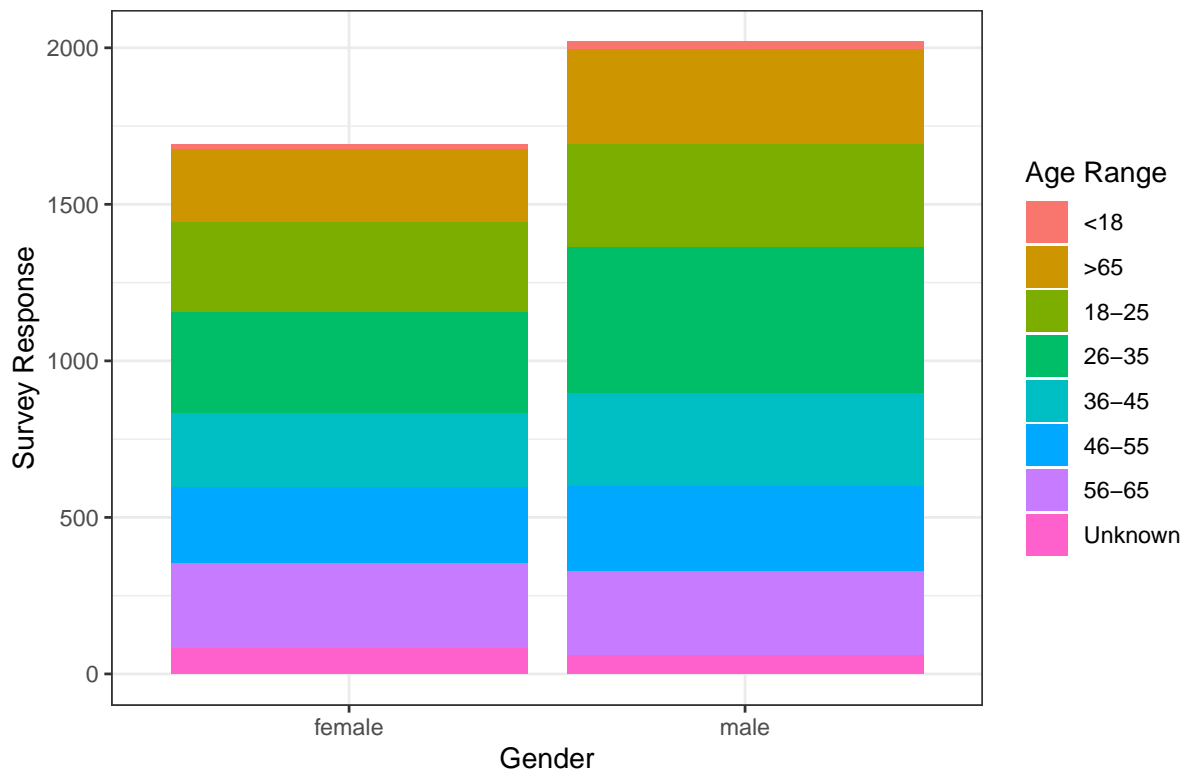
Intake volumes for the Cyber Security course by year and survey grouping are provided below:



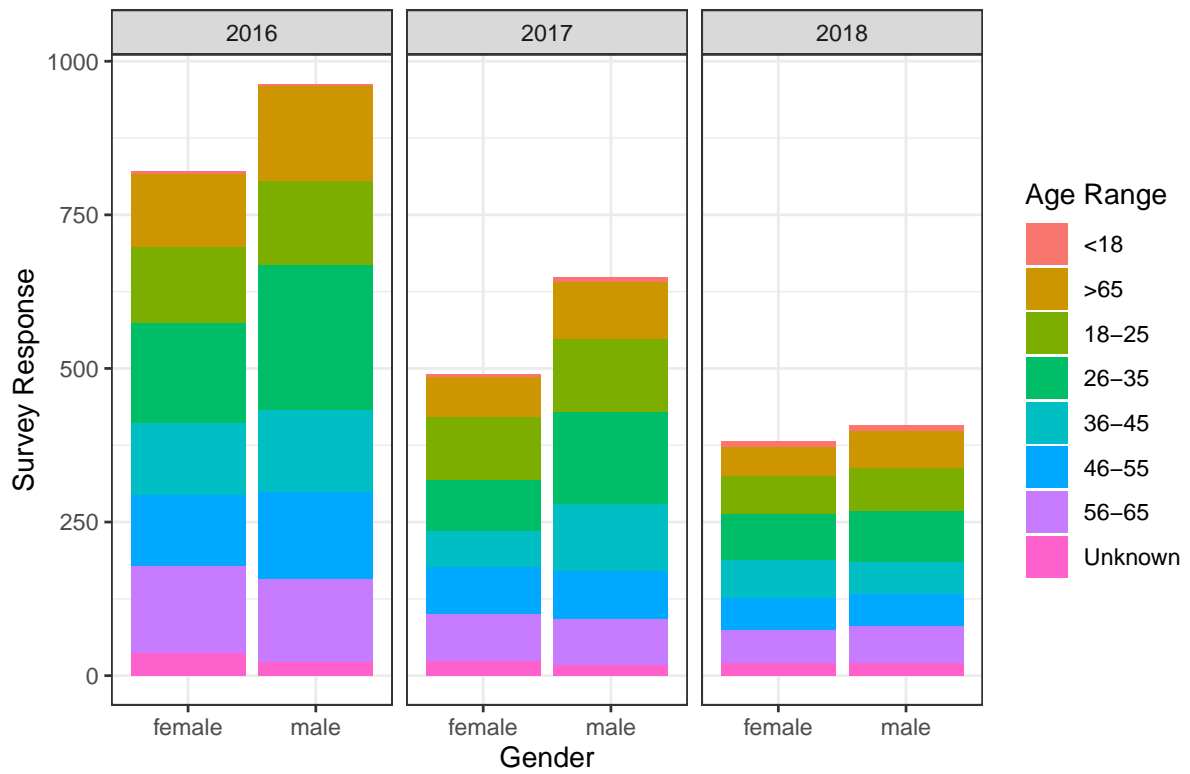
Cyber Security Students that Provided Information on Gender



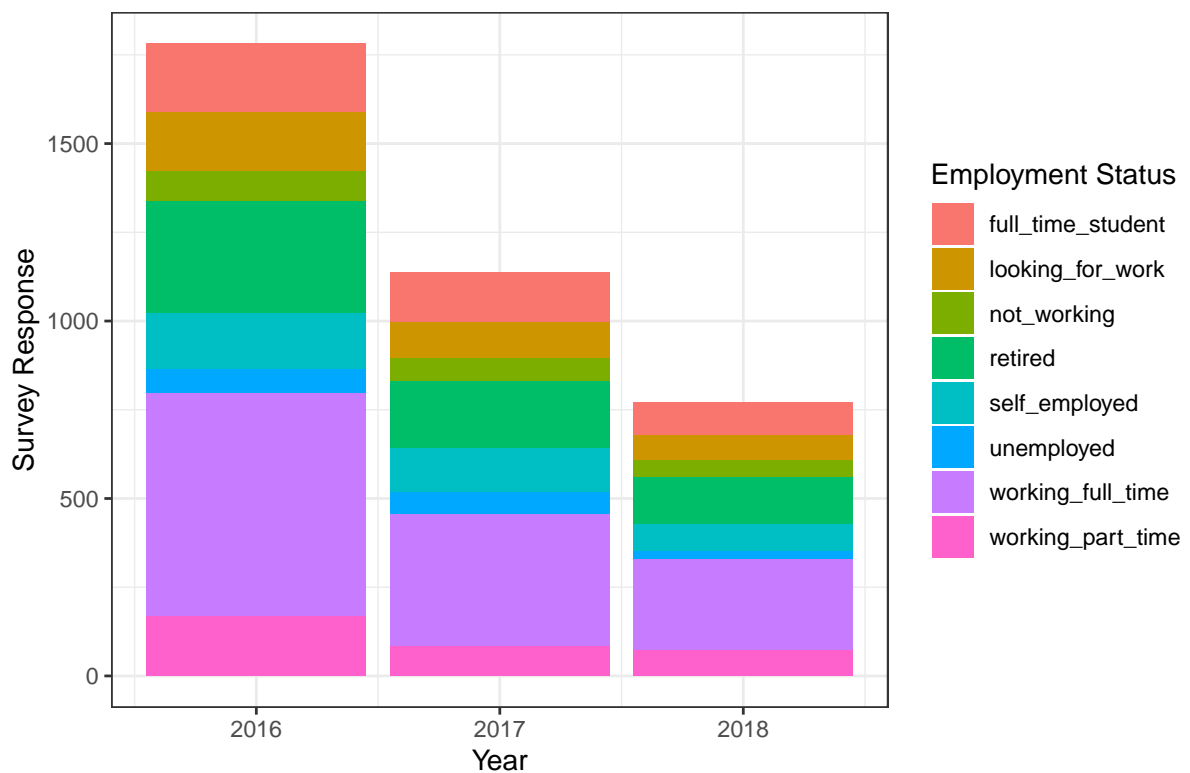
Male or Female Cyber Security Students that Provided Age Range



Male or Female Cyber Security Students that Provided Age Range



Cyber Security Students that Provided Information on Employment Status



## Data and assumptions

Working 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.

```
##
##      2  3  4  5  6  7
## 2016 99  0  0  0  0  0
## 2017 340 308 385 58  0  1
## 2018  0  0  49 339 267 225
```

- 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.

## Assumptions

Assumption	Value
Discount rate	#r paste(100*disc, "% p.a.", sep="")
...	...
...	...

## Data summary

I look forward to discussing this report with you.

**Issy Middleton - C1000051**

**Newcastle University - CSC8631 Summative Assignment**

**December 2020**

## Notes

Eum et tamquam patrioque, has an meliore imperdiet, audiam docendi delicata his ad. Ad solet perfecto vituperata his, ignota appetere mea ad. Id mea mazim consequat, in mei dignissim cotidieque, sed eloquentiam definitiones no. Persius democritum ne mea, duo an mutat blandit deterruisset. Assum deterruisset an vix.

Ius omnium argumentum ea, has an nibh eius, ei mel dicta choro singulis. Mel no iriure sententiae, dolorem menandri patrioque in eum, saepe singulis atomorum vim id. Vim te epicurei reformidans, eum at posidonium complectitur. Lucilius legendos sit ei. Duo ut denique epicurei vituperatoribus.