Data Science Project Report – Neha Jain

# Glossary

## SQA

SQA is accreditation and qualification awarding body in Scotland. More details can be found on their website <https://www.sqa.org.uk/sqa/5659.8271.html>

## SQCF

The SCQF is a way of comparing Scottish qualifications to other qualification frameworks. It does this by giving each qualification a **level** and several **credit points**.

* The level of a qualification shows how difficult the learning is.
* The credit points show how much learning is involved in achieving that qualification.

It covers achievements from school, college, university, and many work-based qualifications.

More details can be found on <https://scqf.org.uk/>

## STEM

STEM stands for science, technology, engineering, and mathematics. The related subjects used in analysis are:

* Mathematics
* Physics
* Business Management
* Computing Science
* Engineering Science
* Graphic Communication
* Mathematics of Mechanics
* Statistics

## GCSE

Source - <https://en.wikipedia.org/wiki/General_Certificate_of_Secondary_Education>

The General Certificate of Secondary Education (GCSE) is an academic qualification in a particular subject, taken in England, Wales and Northern Ireland.

All GCSE subjects are listed here: <https://www.bbc.co.uk/bitesize/levels/z98jmp3>

## Applied GCSE

You can also study for an applied GCSE that concentrates on work-related aspects of subjects. Applied GCSEs are offered in:

* applied art and design
* applied business
* applied ICT
* applied science
* engineering
* health and social care
* leisure and tourism
* manufacturing.

One applied GCSE is equivalent to two conventional GCSEs, depending on options chosen.

## Education in Scotland

Source: <https://en.wikipedia.org/wiki/Education_in_Scotland>

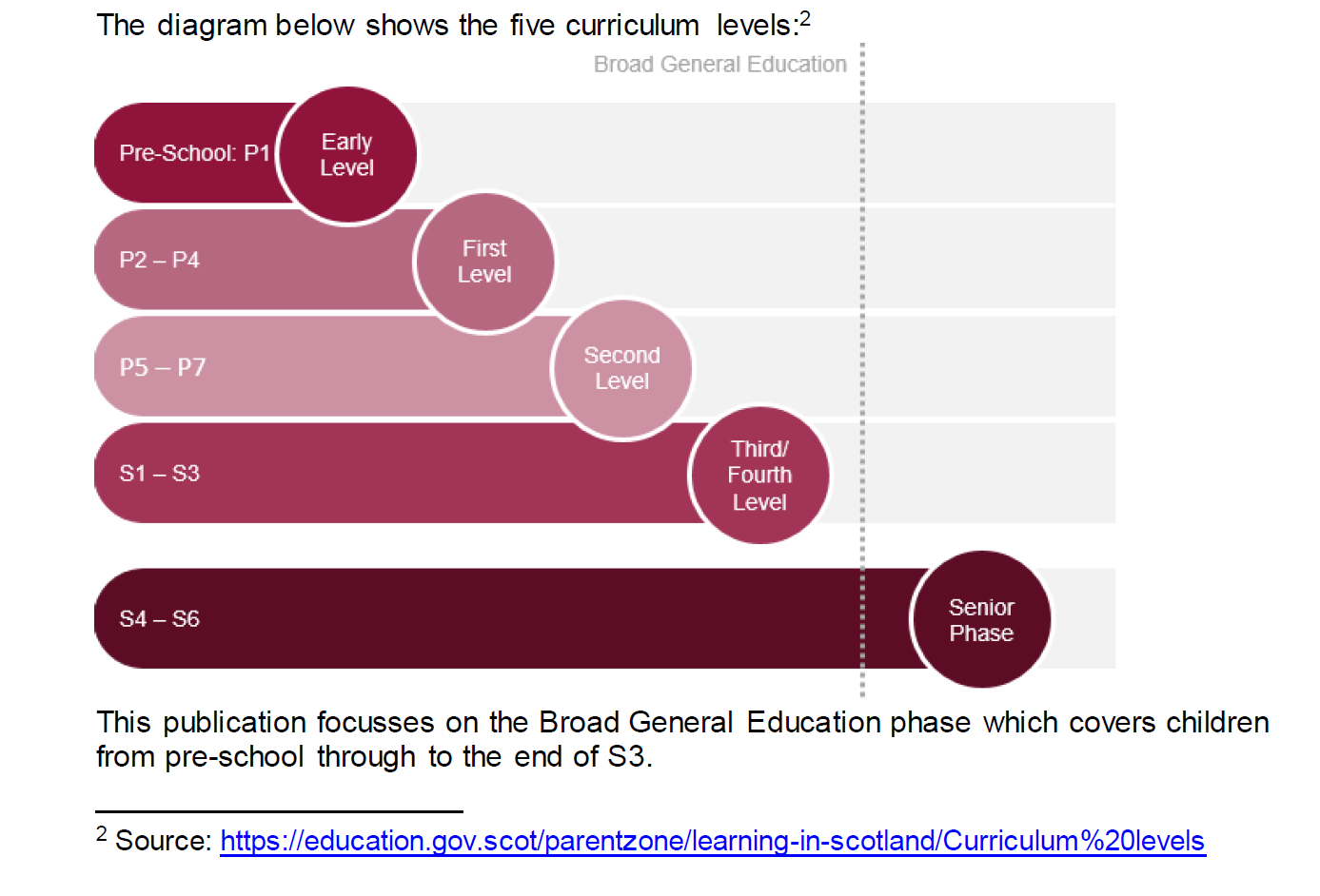


# Computing in schools

The original aim was to analyse how computing curriculum is impacting gender imbalance in schools. Publications from UK government and other international surveys seem to suggest that Scotland curriculum is revised periodically and is in-line with current industry requirements. More details can be found under PISA section in references.

The focus then shifted to analyse the intake of STEM subjects in school and ‘Computing Science’ in particular.

## Curriculum



SQA is responsible for developing the curriculum for various levels of its qualifications.

Computing science is not a standalone subject in Scotland, but forms an element of Technologies, one of eight key areas that make up the country’s Curriculum for Excellence. Its three core goals are to ensure learners can:

* Understand the world by means of computational thinking
* Comprehend and analyse computing technology
* Design, build and test computing solutions.

Other Technologies subjects, which are mandatory for students between the ages of three and 15, comprise:

* Digital literacy
* Technological developments in society
* Craft, design engineering and graphics
* Food and textiles

The last four areas are meant to provide suitable contexts for pupils to develop their technological skills and knowledge. In other words, the aim of this approach is to provide them both with specific training in computing science skills, such as coding, and experience of how to use them on a cross-curriculum basis.

### National 5

<https://www.sqa.org.uk/files_ccc/ComputingScienceCourseSpecN5.pdf>

### Higher

<https://www.sqa.org.uk/files_ccc/HigherCourseSpecComputingScience.pdf>

### Advanced Higher

<https://www.sqa.org.uk/files_ccc/AHCourseSpecComputingScience.pdf>

### Junior

Junior groups have a curriculum too. Some of the details can be found on this link [https://education.gov.scot/Documents/Technologies-es-os.pdf](https://protect-us.mimecast.com/s/oLe7CJ61oAupRDn6iGlESr?domain=education.gov.scot). This might have more of an impact as they won't choose it as a National if they don't enjoy it when starting high school.

### STEM

<https://education.gov.scot/parentzone/learning-in-scotland/curriculum-areas/stem>

### School specific

Some schools have made their curriculum available online e.g., Royal High School in Edinburgh <https://royalhigh.wordpress.com/curriculum/subjects/computing/>

### England Curriculum

Scotland curriculum is different from England, Wales, and Northern Ireland.

<https://www.informaticsforall.org/wp-content/uploads/2019/07/Informatics-Education-in-England-Presentation.pdf>

### GCSE

#### Computer Science

<https://www.bbc.co.uk/bitesize/subjects/z34k7ty>

#### Design and Technology

<https://www.bbc.co.uk/bitesize/subjects/zvg4d2p>

#### Digital Technology

<https://www.bbc.co.uk/bitesize/subjects/z9qy6yc>

#### ICT

<https://www.bbc.co.uk/bitesize/subjects/zqmtsbk>

## Trend in schools

This section focuses on analysing the intake of computing related subjects in primary and secondary schools in Scotland.

### Data Sets

The data sets used are from SQA <https://www.sqa.org.uk/sqa/57518.html>. Following national qualifications are used for data analysis:

* National 5 (SQCF Level 5)
* Higher (SQCF Level 6)
* Advanced Higher (SQCF Level 7)

#### Years chosen

The statistics used are from year 2015 till 2019. There were no exams conducted in 2020.

#### Information in data sets

The data sets published by SQA has following information:

* Trend in entries in all subjects across years
* Trend in grades A-C attainment rates in subjects
* Entries and attainment of grades A-D
* Entries and attainment by sex
* Entries and grades per learner

#### Subject/s used in data analysis

The main subject used for data visualisation is ‘Computing’/’Computing Science’. Some other subjects are also included as they also aim to instil following skills in students:

* Analytical and critical thinking
* Statistics (without advanced calculus or physics)
* Problem solving
* Creativity
* Collaboration
* Data-driven decision making
* Intellectual curiosity

Subjects:

* Mathematics
* Physics
* Business Management
* Computing Science
* Engineering Science
* Graphic Communication
* Mathematics of Mechanics
* Statistics

### Data Bias

1. The motivation behind the analysis is to look at female participation in computer science. And the belief is males outweigh females. So, cherry picking data to prove that belief can contribute to data bias.
2. No data analysis is done on population. It is possible that number of girls born are less than boys. Hence, there is a smaller number of girls in schools.
3. Data set does not provide any information on kids who do home schooling.
4. The trend seen in females in computer science when combined with male does not change. There is downward trend seen in both male and female in computer related subjects. So, don’t lose sight of bigger picture.
5. There is no comparison done with other countries. The data set is limited to Scotland.

### Data Ethics

1. The data sets used from SQA website are not confidential and are freely available for anyone to use.
2. The code to generate results/graphs is available to all team members for review.
3. The tools used in this project are provided by course providers.

### Data Analysis

Python is used analyse the data to produce the graphs below.

#### Libraries

The libraries used in data cleaning are mentioned below.

1. Pandas
2. Numpy
3. Matplotlib
4. Seaborn

#### Reading data

The data sets used are in excel format and Pandas read\_excel function is used to import in dataframe for further analysis. Only the sheets with meaningful data for this project are used.

#### Data cleaning and filtering

Python lists are used to filter the data. Few examples are listed below.

1. Filtering subjects of interest (e.g., STEM)
2. Filtering male and female students
3. Filtering grades
4. Cleaning data across years to match subject names

#### Data plotting

Matplotlib library provides plot function which is extensively used to generate plots used in this project.

#### Source code

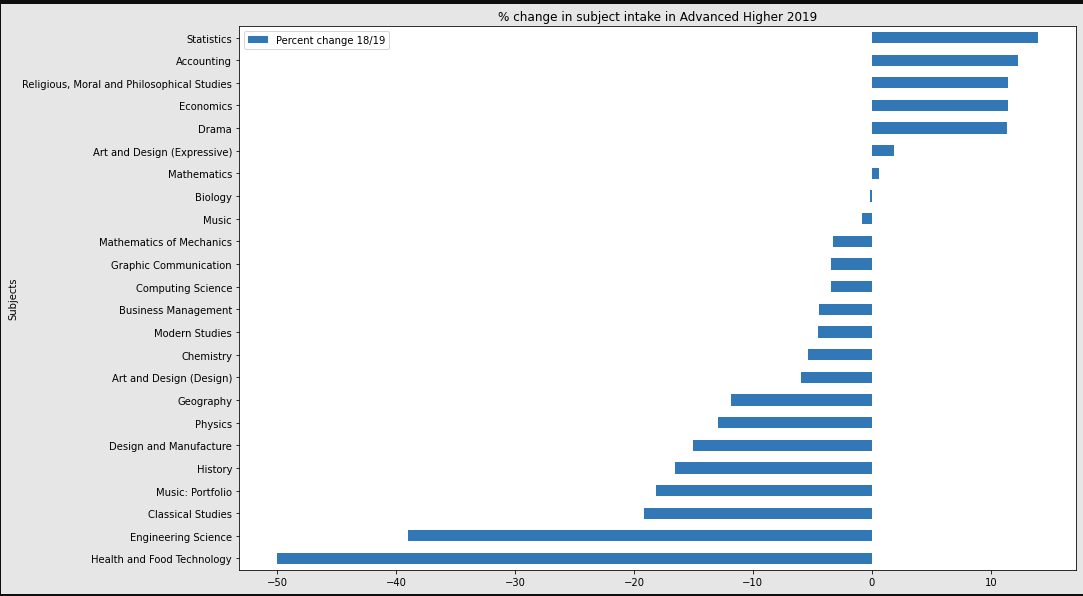
The source code used for data analysis can be found on GitHub <link to Git>

### Data Visualisation

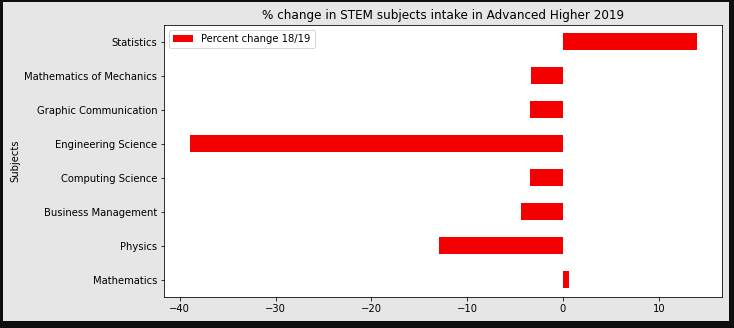
The graphs below focus mainly on datasets for Advanced Higher.

#### Trend in subject intake in Advanced Higher

The graph below shows the trend in year 2018 and 2019. Bars on right hand side show positive trend and left-hand side shows negative trend.

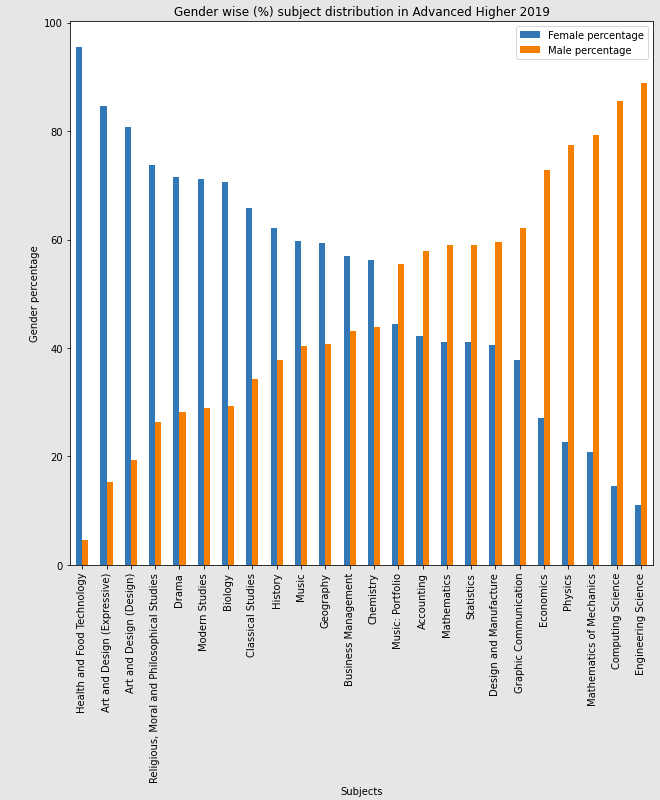


The trend in STEM subjects is highlighted in the graph below. Except statistics all other subjects seem to lose popularity with engineering science being least popular.



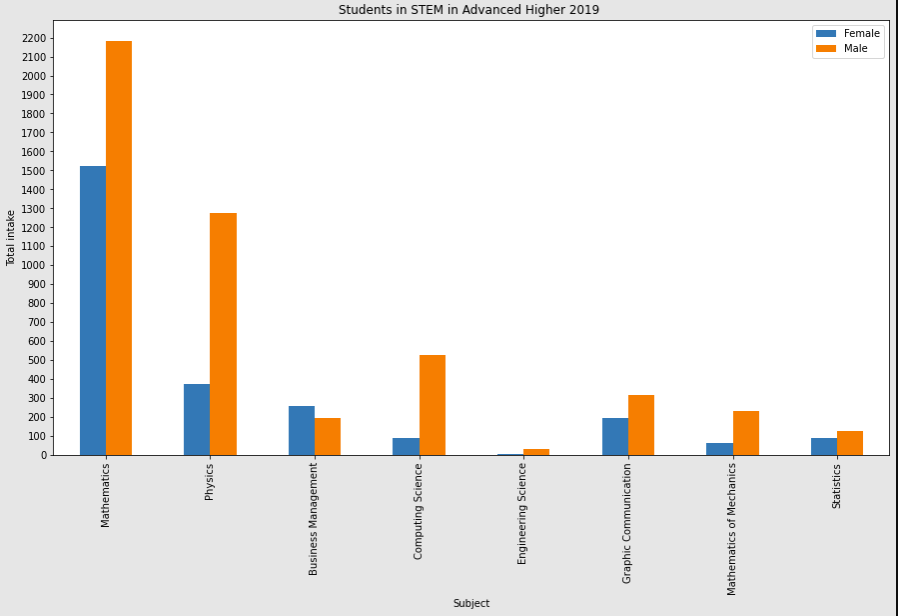
#### Gender distribution in subjects

The graph below shows the percentage of male and female taking various subjects in Advanced Higher for the year 2019. For sake of simplicity, language subjects are removed. From the graph it is quite evident that male tend to take more core subjects related to STEM as compared to female.



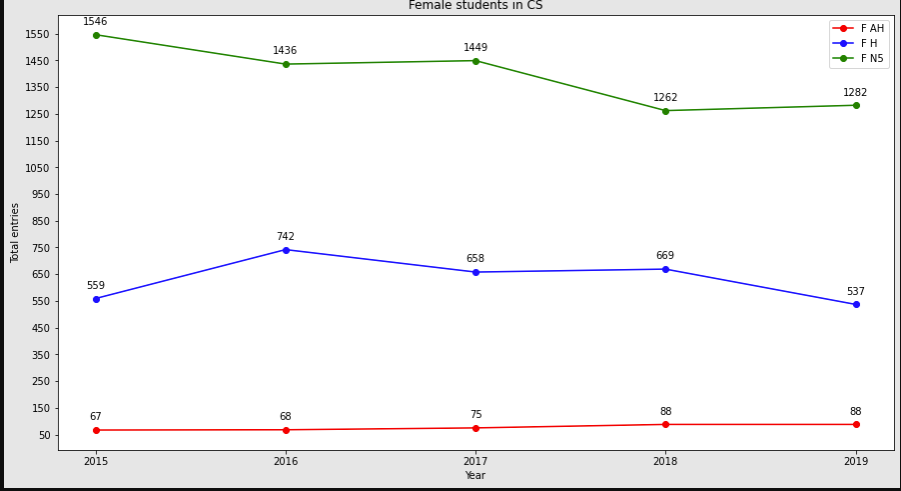
#### Total intake in STEM subjects

The graph below shows the number of students taking STEM subjects. Only 614 students enrolled in in Computing Science out of which only 14% were female.



#### Trend in Computer Science in females

The graph below shows the number of female students taking ‘Computing Science’ subject in N5, Higher and Advanced Higher. There is steep drop in female students taking computing science from N5 to Higher and Advanced Higher.



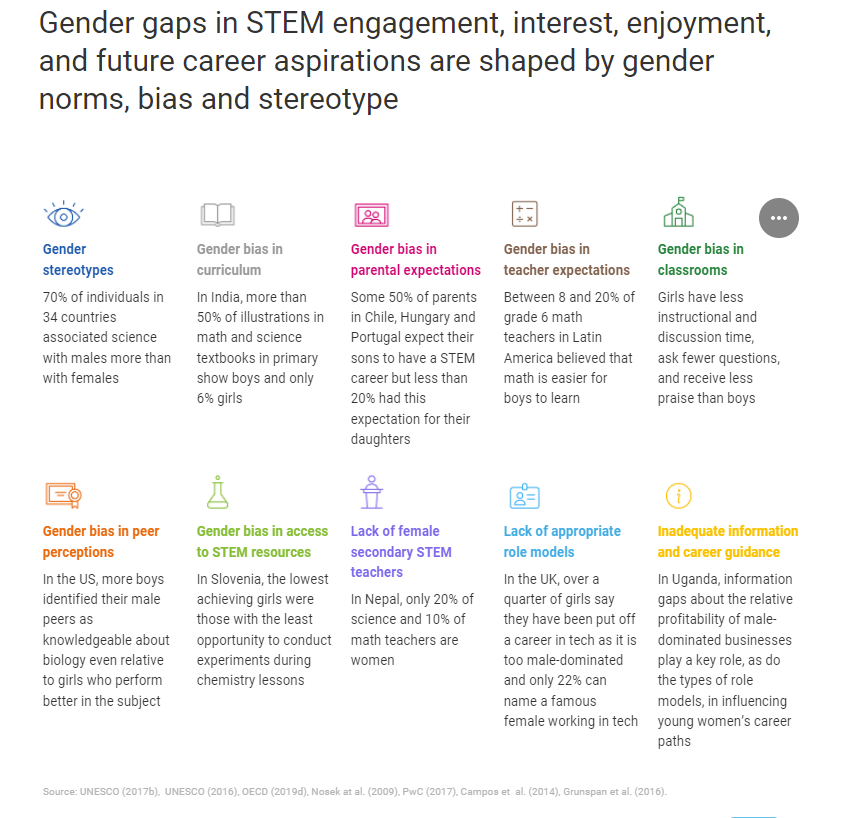
## Conclusion

There are not enough data literate people in jobs. There are not enough in school. The ones in school drop after N5. Why? Its not just CS but all STEM subjects. Its not just female but male too.

Gender imbalance in STEM at school level is very well covered by the link below. <https://www.unicef.org/globalinsight/stories/mapping-gender-equality-stem-school-work>

It shows that same trend is seen across developing and developed countries.

Some of the factors listed for gender bias at school level are pasted below.



### Students’ perspective

While working on the project, we got chance to talk to some of the girls in secondary school. All of them echoed similar challenges while studying computing science (CS) in school. Some of them are listed below.

1. Lack of infrastructure
   1. Not enough computers/laptops, internet bandwidth in schools
   2. Very little dedicated time to study CS
   3. Made to choose between core subjects and CS (overlapping timetable)
   4. Not enough dedicated teachers, some reported one CS teacher for 300 pupils
   5. Not all schools offer CS subjects
2. Lack of support
   1. Very limited to zero support in Advanced Higher
   2. Solely dependent on parents for guidance and support
   3. No visibility of career path
   4. In Scotland, students get one chance to score marks whereas in countries like USA there are regular assessments which gives chance to improve performance
3. Cultural
   1. Detrimental attitude of pupils towards strict teachers
   2. CS is still given step-motherly treatment and more focus is put on core subjects

## Future Work

1. The data used in analysis is limited to past 5 years. It can be expanded beyond those years to see better trend.
2. The trend in computer science intake in schools can be further divided across schools if such data is available. It can give an insight which schools are doing better and what others can learn from them.
3. The impact of teachers (availability/support etc.), family and school infrastructure can give more visibility into trends seen so far.
4. How is Scotland doing as compared to rest of the world?
5. Are students aware of various career paths from different subjects and what impact they can have on society?

## References

### PISA

<https://www.bbc.co.uk/news/uk-scotland-50642855>

<https://www.gov.scot/publications/programme-international-student-assessment-pisa-2018-highlights-scotlands-results/pages/9/>

### Gov Scot

<https://www.gov.scot/publications/stem-strategy-education-training-scotland-second-annual-report/pages/9/>

### Mark Logan report

[scottish-technology-ecosystem-review (7).pdf](https://codedivisiongroup.slack.com/files/U01PTT1F752/F023JH62CA2/scottish-technology-ecosystem-review__7_.pdf?origin_team=T01P6UTNUP9&origin_channel=C023QPYCW4F)

### Others

<https://services.google.com/fh/files/misc/unconscious-bias-in-the-classroom-report.pdf>

<https://www.skillsdevelopmentscotland.co.uk/media/42478/tackling-the-technology-gender-gap-together-2.pdf>

<https://royalsociety.org/topics-policy/projects/dynamics-of-data-science/>

<https://www.pure.ed.ac.uk/ws/portalfiles/portal/118843174/CS_Teacher_Shortage_Report_UoE_SDS_Nov_2019.pdf>

<https://www.bcg.com/en-gb/publications/2020/what-keeps-women-out-data-science>