Data Science

Educators' Guide to the National Progression Award

CONTENTS

ABOUT THIS GUIDE

This **Educators Guide** has been produced with an accompanying **Learners Guide**. Both guides have been created to support the core units of the National Progression Award (NPA) in Data Science at SCQF Levels 4, 5 and 6.

The **Educators Guide** covers information for teachers and lecturers when they are first considering selecting the NPA in Data Science, advice for planning for delivering the course, and information they may find useful when delivering the two core units of the NPA.

The **Learners Guide** is a summary document covering the core concepts that learners will need to know in order to learn about Data Science and to undertake the assessments. It can be used by educators to introduce each topic, or as a summary or revision document prior to assessments.

These two guides have also been produced alongside an online guide developed by the University of Edinburgh that details all the concepts in more depth. The University are also working with schools and colleges to develop and trial teaching materials. You can find more information about this in the 'Support and Resources' section of this Educators Guide.

There are many exciting and engaging contexts for learning about data science, and many different tools that can be used to gain the practical skills involved in the course. The Learners Guide will not include practical tasks for particular tools. Instead, this guide will give advice on the best tools that can be used in teaching and learning, based on the experience of educators, the level and interests of learners, and any technological constraints in your school or college.

About the authors

Kate Farrell is an experienced Computing Science teacher. She works for the Data Education in Schools project at the University of Edinburgh's Moray House School of Education and Sport. She was the Lead Developer for the NPA in Data Science at SCQF Levels 4, 5 and 6, and wrote the SOLAR unit assessments for the core units.

Dr Jo Watts is an experienced Data Scientist and is founder of Effini, a data science company. She wrote the NPA Data Science core units, the Data Science project units, and the SOLAR unit assessments for the core and project units. She was lead developer for the PDA in Data Science at SCQF Levels 7, 8 and 9.

Support and resources

These guides have been written with the University of Edinburgh's Data Education in Schools team. The Data Education in Schools project aims to work with schools and colleges that are delivering this course. The project is developing and adapting resources and are keen to support centres to work together in partnership. To date, they have worked with every school delivering this qualification, providing professional learning, facilitating sharing of resources, and working together to review materials and share the development workload.

Teaching materials will feature local data sets and case studies from industry. The Data Education in Schools team are actively working with a range of industry sectors, academics and researchers across the University, and with schools and colleges to trial resources.

Visit **www.dataschools.education** for more information about support materials.
Visit **dataed.in/NPADS** for more information about the qualification.

NPA DATA SCIENCE

National Progression Awards (NPAs) are available in a variety of subject areas and are aimed at assessing a defined set of skills and knowledge in specialist vocational areas.

Learners demonstrate their knowledge and understanding by internal assessments throughout the course. Practical skills are assessed by performing a practical assessment rather than by using a written final exam.

NPAs are available at SCQF Levels 2-6 and can be delivered in partnership between schools, colleges, and employers. Although schools are increasingly offering and delivering NPAs, they are mainly used by colleges for short-study programmes, such as return-to-work courses.

This NPA in Data Science is available at three levels, **SCQF levels 4, 5 and 6**. Learners can progress through the levels of the National Progression Award. If they wish to continue learning about Data Science, they can undertake the **Professional Development Award (PDA)** in Data Science at **SCQF Levels 7, 8 and 9** at a Further Education College or apply for one of the many degree programmes available in Universities across Scotland.

> **Group award codes**
>
> **NPA Level 4:** GP8N 44
> **NPA Level 5:** GP8P 45
> **NPA Level 6:** GP8R 46

STRUCTURE AND CONTENT

The NPA in Data Science consists of two core units at every level. There is a core unit in Data Citizenship and a core unit in Data Science.

The Level 4 course only consists of the two core units. This has been designed to allow learners the time to secure their learning at this level, particularly for learners who have 'dropped down' from Level 5.

Levels 5 and 6 consist of three units: the two core units and one optional unit.

There are a selection of units to choose from at both Level 5 and 6. This has been designed to fit flexibly into a range of educational contexts and allow educators the ability to select an optional unit that matches their expertise and the interests of their learners.

Core Units

Data Citizenship

"The purpose of this unit is to provide an overview on the place of data in society, how data can be used and misused, and the steps we can take to understand and use data responsibly. This unit will help learners become responsible, data literate citizens who participate in the decisions that affect people and society.

"Learners will gain a range of practical skills and acquire relevant underpinning knowledge. They will learn how to interpret meaning from visualisations, such as graphs and charts, and to create visualisations from data. They will learn about how data can be used in society for positive and negative effects. They will also learn about data security and their rights and responsibilities as data subjects and data owners.

"On completion of this unit, learners will have gained confidence in their use of data, and be aware of their rights and responsibilities as data citizens."

Unit Specification

The **Data Citizenship** core unit involves understanding how data is used. Learners learn data literacy and basic statistics. They will learn how to interpret data in different formats to find out interesting things from the data, to investigate why unusual results or trends, and think about the impact or behaviour change resulting from the analysis. Learners will investigate how data can have both a positive and negative effect on society, such as when biased data is used in decision making or when data is misrepresented to influence people.

Data Science

"The purpose of this unit is to introduce learners to data science in today's world. The unit focuses on the tools and techniques involved in data science, the main methods of data analysis, and provides an opportunity for learners to apply this knowledge in a practical context.

"The unit covers a variety of topics relating to data science including: the reasons for the emergence of data science as a distinct discipline, the uses and misuses of data and data science, the data science life cycle and common methods of data analysis. Learners will also gain practical skills in using software to identify patterns and trends in data.

At the completion of this unit, learners will appreciate the basic principles of data science and be able to apply this knowledge to solve routine problems using data analysis software."

Unit Specification

The **Data Science** core unit involves learners gathering data from different sources then analysing it by exploring, modelling and validating the data. Learners then visualise the results and present on their findings, reporting on what they have found and how it can make a difference to themselves or others.

OPTIONAL UNITS
Data Security

> "The purpose of this unit is to introduce concepts around personal and corporate data security, including aspects of legal and ethical obligations. Learners will discuss examples of real-life data security breaches, and examine the reputational and financial damage caused by poor data security practice. A specific aim of this unit is to place data security within the context of the real world. This includes the legal and ethical considerations, and the practical methods to protect personal and corporate data."
>
> *Unit Specification*

The Data Security unit provides a relatively gentle introduction into the field of cyber security.

This unit is one of the core units in the NPA in Cybersecurity at Levels 4, 5 and 6. This means if learners already have the Cybersecurity NPA then they do not need to do an optional unit in the Data Science NPA. Schools and colleges can offer the two qualifications together and learners will only need to achieve five units rather than six.

If a school or college is experienced in delivering the Data Security unit, then this would be a natural choice of optional unit, at least for the first year or so until educators are familiar with the other units and are interested in trying other options.

The Data Security unit is also available at Level 4, so if you have a multi-level class with some Level 4 learners, they could pick up this unit as an additional unit (although they do not need to achieve this to gain the NPA in Data Science).

Computer Programming

> "The purpose of this unit is to provide programming skills and knowledge of the principles of computer programming... Learners will gain a range of practical skills and acquire relevant underpinning knowledge. They will learn how to write code in a contemporary high-level language and appreciate programming concepts and techniques, and develop their computational thinking skills. On completion of this unit, learners will know how to write programs to solve real-world problems."
>
> *Unit Specification*

The Computer Programming unit covers writing algorithms to solve problems, explaining programming concepts and writing computing programs. The programming language used is not specified, so if learners are going to use Python in the Data Science unit (particularly at Level 6) then covering this unit first to introduce learners to the Python language would seem a sensible approach.

This would be a good optional unit choice if the course is being delivered by a Computing Science teacher or lecturer with experience of teaching programming.

The Computer Programming unit is also available at Level 4, so if you have a multi-level class with some Level 4 learners, they could pick up this unit as an additional unit (although they do not need to achieve this to gain the NPA in Data Science).

Data Science Project

"The purpose of this unit is to allow learners to complete an end-to-end data science project within a team, using pre-existing knowledge and skills.

This capstone project will introduce the types of routine problems that can be solved using data science techniques, reinforce the learner's knowledge of data science and provide a meaningful opportunity to apply data science tools and techniques. It will also provide an opportunity for learners to demonstrate the 'soft skills' that are important in a real-world workplace, such as team work, collaboration, good time management, problem-solving and effective communication skills.

The activities include identifying the problem in a project brief, analysing the collected data and communicating the results. The focus of the unit is on clarity and communication throughout each of the steps rather than the level of analysis delivered."

Unit Specification

The Data Science Project unit at Level 5 is a group project, with learners working together to source and analyse data and then communicate their findings and recommendations. At Level 6 this is an individual project. Learners will be investigating a problem that they have chosen themselves.

This unit has a lot of potential to be a hugely engaging and fun experience for learners, particularly if they are prompted to investigate projects that can make a real difference in their schools and communities. For example, learners could use IoT sensors to investigate environmental conditions in their school building or in their neighbourhood, perhaps combined with available climate data. Learners having a sense of ownership over the project and self-efficacy will be beneficial for the success of the projects.

Although it is anticipated that this unit will be a popular choice with schools and colleges, it might not be the best choice for an inexperienced teacher or lecturer the first time this course is run in a centre. The second time around, it would be a fun option for an educator who has gained confidence with the core units, particularly the Data Science unit.

Statistics

"The general aim of this unit is to develop knowledge, skills and understanding in statistical methods and techniques that can be applied to a variety of real-life contexts which may be new to the learner. This includes skills in interpreting and analysing graphs and statistical diagrams, applying skills to the normal distribution and determining the equation of linear regression and using it for prediction. Learners who complete this unit will be able to use statistical skills in real-life contexts and produce a statistical analysis on given data sets."

Unit Specification

This unit has been popular in Maths departments. If a school or college is experienced in delivering this Statistics unit, then this would be a natural choice of optional unit, at least for the first year or so until educators are familiar with the other units and are interested in trying other options. This would be a good optional unit choice if the course is being delivered by a Mathematics teacher with experience of teaching statistics.

Gaining this single unit qualifies learners for a Statistics Award, in addition to the NPA in Data Science. There are support materials available in the form of 'Unit Support Notes' and Understanding Standards notes. These are available at: **dataed.in/SQAstats**.

It should be noted that even though this is a single unit group award, the SQA does not automatically accredit candidates for the group award if they have gained the unit. Learners should be entered for both the unit and the group award.

Data Science Statistics

"The purpose of the **Level 5** unit is to introduce learners to the fundamental statistical concepts required in the field of data science... The unit introduces basic statistical methods that are fundamental to data science, and applies that knowledge using simple data analysis tools. Although the focus is statistics as it relates to data science, general statistical techniques are introduced when these underpin more specialist knowledge.

The purpose of the **Level 6** unit is to develop learners' knowledge of statistics as they relate to data science... The Level 6 unit explains statistical concepts and theorems that are important in data science including hypothesis testing and Bayes' Theorem. It prepares learners for carrying out a statistical study and then shows learners how to carry out the study using contemporary data analysis tools."

Unit Specification

This would be a suitable optional unit choice if the course is being delivered by a Mathematics teacher or lecturer with experience of teaching statistics in the field of data science. It would require more experience than the Statistics unit and covers more in-depth content. It may possibly require more preparation and development time than some of the other optional unit choices.

It would be more suitable for a single-level class, rather than a bi-level cohort due to the unit content being sufficiently different at Level 5 and 6. The Level 6 unit builds upon the content of the Level 5 unit.

Machine Learning

"The purpose of the **Level 5** unit is to provide a straightforward introduction to machine learning and its applications. It covers the a wide range of knowledge and skills including: the broad purpose of machine learning and its applications in business, health and science; the machine learning workflow; supervised and unsupervised learning; the role of algorithms; training and test datasets; fitting a classifier model and interpreting the results; under-fitting and over-fitting; and the ethical implications of machine learning.

The purpose of the **Level 6** unit is to provide a grounding in some of the computational approaches that pertain to machine learning, along with an appreciation of methods to prepare and select data to facilitate model development and use. It will develop skills in fitting and evaluating a predictive model, and introduce strategies to measure and improve model performance. The unit covers the following knowledge and skill: data scaling and normalisation; feature engineering; model validation; linear regression algorithms; gradient descent; logistic regression for binary classification; interpretation of algorithm outputs; measurement and improvement of model performance."

Unit Specification

This would be a suitable optional unit choice if the course is being delivered by an experienced Computing Science educator with experience of machine learning techniques. It would probably require more preparation and development time than some of the other optional unit choices.

It would be more suitable for a single-level class, rather than a bi-level cohort due to the unit content being sufficiently different at Level 5 and 6. The Level 6 unit builds upon the content of the Level 5 unit.

DELIVERY MODELS

The NPA in Data Science has been designed to fit in flexibly to a wide variety of delivery models within schools and colleges to address the realities of staff time, expertise and enthusiasms as well as learner demand and local industry requirements.

A CS course for a range of settings

Although this course is in the SQA's 'Computing and IT' course catalogue, the course can be delivered in a range of different settings. In schools, it is already being delivered by CS, Maths and Geography teachers. In FE colleges, the NPA has so far been embedded into courses for Health and Social Care students and Car Mechanics students.

Data Science tools and techniques are being used in some many different fields of industry. The NPA is flexible enough that it can be taught using a wide range of examples, case studies and data sets, or it can be used to focus on a narrower field of application such as health and sports science, or geography and agriculture.

The optional unit can be selected to fit in best with the educator's experience and interests as well as the needs of the learners. For example, a Computing Science educator might choose to deliver the Computing Programming unit first to introduce the learners to the Python language that they can then use later in the Data Science unit. A Maths lecturer might choose to follow up the core units with the Statistics unit to develop learners' knowledge and skills beyond the basic statistics concepts covered in the core units.
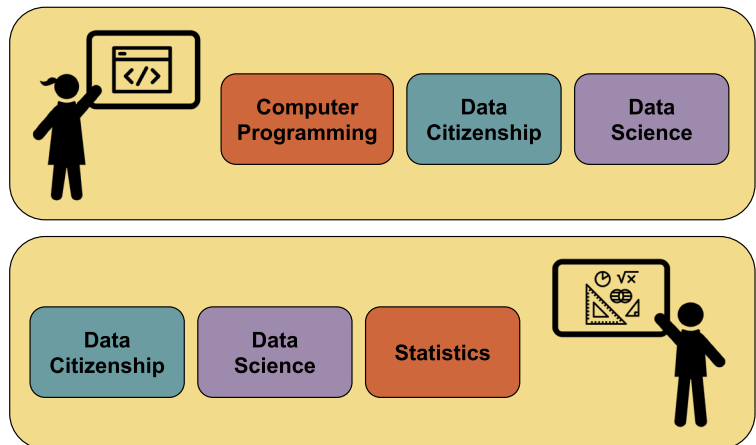
Interdisciplinary team delivery

Many units have been designed so they do not require to be delivered by educators with Computing Science expertise. This means the NPA can be taught across different departments, particularly given the flexibility of the optional units. For example, Data Citizenship could be delivered by Geography or Modern Studies teachers, Data Science by a Science or Computing Science teacher, and Statistics delivered by the Maths faculty. This allows for more flexible timetabling based on staff availability, expertise and interests.
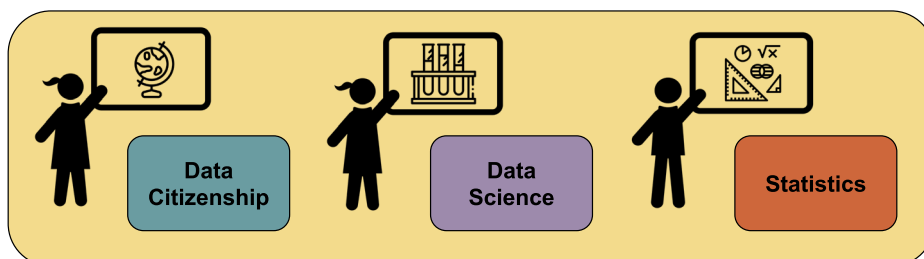
Engaging stand-alone units

Each of the units was written so that it could be delivered stand-alone. For example, the Data Citizenship unit could be delivered as part of a digital literacy course or an additional unit in a social studies course. The Data Science course could be matched to a core Numeracy unit to enhance a real-world maths course.
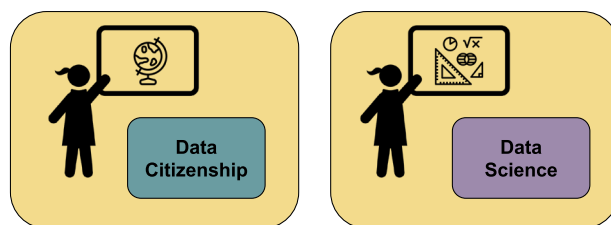
Designed to cope with multi-level classes

Multi-level classes are not ideal but they are a reality in many settings. The core units in the NPA have been written to aid educators who are teaching multiple levels of learners in one class. Many of the optional units in levels 5 and 6 have also been written to be hierarchical as well. This means that topics can be delivered once to all learners, with some learners given some additional information or responding to assessments in more depth.

Flexibility for Level 4 learners

Level 4 only consists of two core units. This potentially gives Level 4 learners the time and space to reinforce learning by going over familiar concepts in other contexts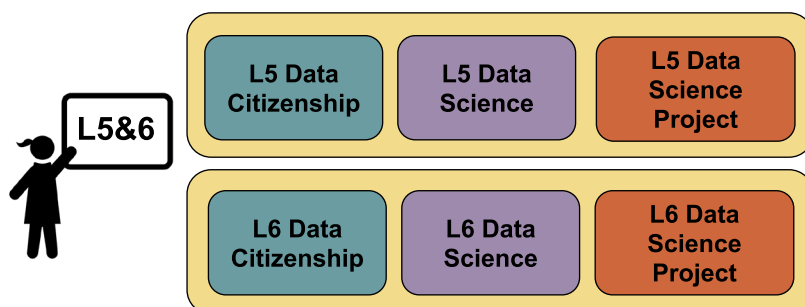. This is particularly beneficial for learners who have maybe started the course working at Level 5 but have moved down to Level 4 instead.
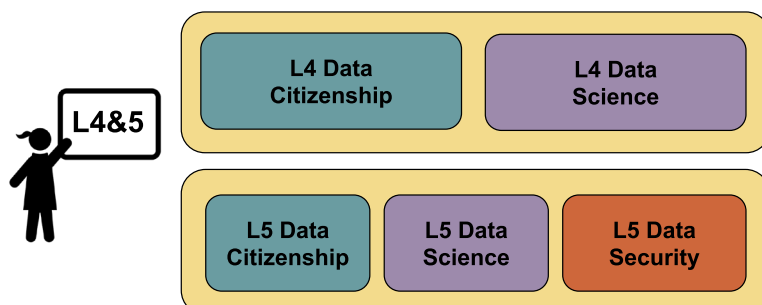
Alternatively in a multi-level class, Level 4 learners could pick up an additional unit if they have extra time, such as the Level 4 units in Computing Programming or Data Security, while the Level 5 learners undertake the Level 5 equivalent. This additional unit would not be required to complete the NPA, but it gives the learner an additional unit and additional credit points.
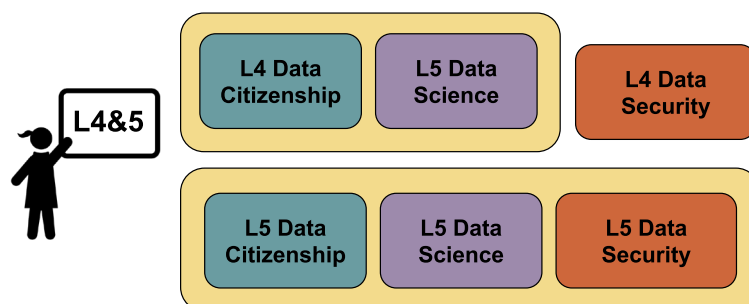
Some centres have considered using the Level 4 NPA as a short intensive course offering for learners not sitting exams during exam leave time. It may also fit in well with another unit or set of units on a specific topic, where an introduction to data science would supplement the subject knowledge and skills in the other unit(s).

Numeracy and Real-world Maths

The NPA at Level 4, with only the two core units, could slot into a timetable with a numeracy unit, such as the Numeracy Level 5 unit. This could be an attractive option for Maths departments looking for real-world maths courses that would engage learners.
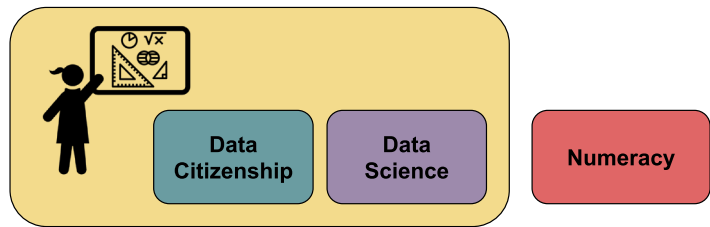
The Level 6 NPA may be an attractive alternative to Higher Maths for learners planning on going to university to study fields such as Geography or Sports Science.

Complementing other course offerings

As previously mentioned, the Data Security unit at Levels 5 and 6 is also part of the NPA in Cybersecurity, so learners could gain two qualifications by completing just five units.

Gaining the Statistics unit at Level 6 will also qualify learners for an Award in Statistics.
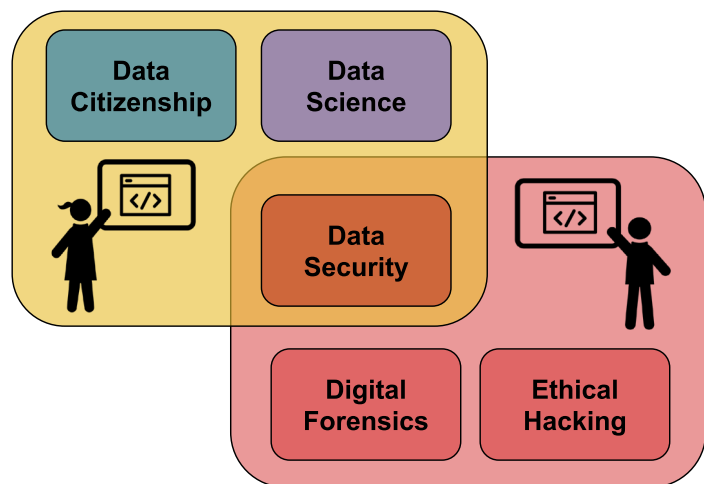
Planning for Progression

The National Progression Award in Data Science is available at SCQF Levels 4, 5 and 6. Learners successfully gaining the Award at Levels 4 or 5 would be able to move on to a higher level of the NPA.

There are a range of Data Science qualifications that learners can progress onto from the NPA. There is a Professional Development Award in Data Science available at SCQF Levels 7, 8 and 9. This PDA has been designed to provide a smooth progression from the NPA. However, as Level 7 is an introduction to the subject, learners who have achieved the NPA at Level 6 might find progressing on to the PDA at Level 8 would be a better transition for them.

There are HNC and HND qualifications in Data Science and Data Analytics, as well as many courses in Computing which involve a significant component of Data Science. Many Scottish Further Education Colleges offer smooth progression from NPA to HND Data Science, and then direct entry onto third year of Data Science BSc programs at Scottish universities.

CORE SKILLS

The NPA in Data Science has been designed to support the development of a range of core skills in learners.

ICT

Throughout the course, in all of the core and optional units, learners will develop their core skills in ICT. Learners will search for information online in the form of datasets, tables, graphs and visualisations. They will use software such as a spreadsheet package, a web tool such as CODAP or a programming language such as Python to interpret, manipulate, and analyse the information. They will then create visualisations and present the information to others.

> **Accessing information**
> **Providing and creating information**

Numeracy

The two core units involve a good grounding in a range of numeracy concepts. Learners will be looking at tabular data, analysing and creating graphical data, and producing summary statistics. Many of the optional units will also support the development of numeracy skills, particularly the Statistics, Data Science Statistics and Data Science Project units.

> **Using number**
> **Using graphical information**

Communication

Communication of findings is a crucial part of the data science lifecycle in the NPA. Learners are expected not only to gain insights from data, but to then present their findings to others. In the Visualisation and Storytelling topic, communicating information to an audience is a key learning point. Learners are expected to create visualisations with that audience in mind, in order to develop a persuasive argument.

> **Written (reading)**
> **Written (writing)**
> **Oral**

Problem Solving

Problem solving is another core component of the Data Science lifecycle, with learners encouraged to first think about the problem that needs to be solved before collecting and analysing data. The course has been developed with the aim that learners will be solving authentic real-world problems that impact their daily lives and their communities. As learners go through the **PPDAC cycle**, they will think about the **Problem** that needs solved, **Plan** their approach, gather the **Data** that will be required to resolve the problem, **Analyse** the data, and then **Communicate** their findings.

> **Critical thinking**
> **Planning and organising**
> **Reviewing and evaluating**

Working with Others

Learners will be encouraged to work co-operatively to solve problems with data. The Data Science Project unit at Level 5 is a group project and will involve learners working in a small group to identify a problem that interests them and then plan and carry out their analysis.

> **Working co-operatively with others**
> **Reviewing co-operative contribution**

SEQUENCING OF TOPICS

The core units have been written so that they can be delivered either independently or as part of the wider NPA qualification. As a result, there is some overlap in content and learning concepts that come up in both the Data Citizenship and Data Science units. For example, both the Data Citizenship and Data Science units involve learners understanding what data is and how it is represented and used, interpreting visualisations and carrying out simple summary statistics on data.

A progression pathway has been developed for teaching the two core units together in order to reduce repetition in content and to introduce learners to important concepts in a logical order. 13 main topics have been pulled together based on the content and concepts covered in the outcomes and performance criteria, with an additional topic that covers tools and languages.

Progression pathway

1. **What is Data?**
2. **Interpreting Data**
3. **What is Data Science?**
4. **Working with Data**
5. **Security**
6. **Privacy**
7. **Capturing Data**

8. **Data Manipulation**
9. **Statistics**
10. **Analysis**
11. **Visualisation and Storytelling**
12. **Quality and Management**
13. **Ethics and Bias**
14. **Tools and Language**

Although it is a valid option to teach all of the Data Citizenship unit before moving on to the Data Science unit, it does mean that potentially learners are a few months in to the NPA in Data Science before learning anything about data science. The aim of the progression pathway is to allow learners to learn about what data is and how to utilise it within the context of data science, exploring the various themes and topics throughout the course in a more coherent manner.

This is not a fixed pathway. Educators should feel free to introduce topics and concepts in a different order as appropriate to your learning context and situation. Some topics, such as tools and languages, security, privacy, and ethics and bias could be delivered at various points throughout the course. Some topics could be revisited in increasing complexity throughout the course, such as the interpreting data topic, where learners could be shown a variety of visualisations regularly as the course progresses (such as a 'graph of the week')

The 'Delivery' section of this Educators' Guide and the majority of the Learners' Guide is set out following this progression pathway. The topics are labelled to assist educators who are delivering one of the units separately.

Educators who are working as a team to deliver the units of the NPA may wish to divide up the topics between them as an alternative to having a split by core unit. This is particularly the case if there is an uneven division of time between educators, so for example one teacher with 1 period a week with the class and another with 3 periods. In this case, it might make the most sense for the first teacher to take topics like Security, Privacy, Ethics and Bias and cover a section of the course rather than trying to squeeze a whole unit into limited time.

Progression of knowledge and skills between levels

Educators should be aware that there is overlap of content across the levels of the core units. For example, learners at all levels will need to interpret data, but the depth and complexity of the concepts and the assessments will increase for Levels 5 and 6.

This is to aid the delivery of multi-level classes, and to allow learners to be placed at the most appropriate level without having to have previously gained the lower level of qualification. Learners can go straight into Level 5 without gaining Level 4, or into Level 6 without gaining Level 5 previously.

This means that although Level 5 and 6 learners who have not studied data science previously will need to be familiar with content at a lower level. For example, Level 6 learners need to be able to interpret box plots, density charts and sankey diagrams but they should also know how to interpret bar charts and line graphs as well.

Outcome and performance criteria mapping

The topics in the progression pathway have all been mapped to the Outcomes and Performance Criteria in the core units at all levels. The 'Deliver' section of the Educators' Guide has the relevant levels and core units indicated for each topic. For example, most content will be marked as applicable for all Level 4,5,6 learners, some may be marked for Level 5 and 6 learners, and a few areas will be marked for Level 6 only.

ASSESSMENT

Centres can set their own assessments for their learners, if they choose. This should be done by following the evidence requirements information in the Unit Specification documents. This might be more relevant to Colleges where the NPA is part of a wider course and educators want to theme the assessment to the course. Centres should get any assessments reviewed by the SQA prior to use.

Most centres will prefer to use the assessment materials that have been created for the core units at all three levels. These are available through the SQA's SOLAR system. There are also assessments available for Data Science Project unit, and assessment support packs available on SQA Secure for the Computer Programming units and the Data Security units.

Although it is acceptable to have one joint practical assessment that covers both the Data Citizenship and Data Science units, learners will gain more experience carrying out two separate assessments and will be able to secure their learning in different contexts. If a centre wants to run a single assessment that covers the two core units, it is recommended that they get these prior verified before using them.

The Statistics unit at Level 6 does not have assessments in SOLAR, but there are support materials available in the form of 'Unit Support Notes' and Understanding Standards notes. These are available at **dataed.in/SQAstats**.

SOLAR Assessments

SOLAR is SQA's secure quality assured e-assessment system that can be delivered across various devices. It is an online assessment tool that provides both summative and formative assessments. These e-assessments cover a wide range of subject areas from SCQF Levels 2-9.

> SOLAR offers two different types of e-assessments:
>
> **Computer Based Tests:** These are based on banks of quality assured question items that can either be set as dynamically generated or fixed version assessments. These can be fully automatically marked, online human marked or a combination of both.
>
> **Computer Based Projects:** Open ended assessments that support digitally submitted evidence that are tutor-marked online within SOLAR.

SQA approved centres can obtain access to the subjects they are delivering. If you are not a current SOLAR centre or if you want to check if you have already have existing SOLAR users within your centre (who can create your account and request new subject access) you can access this at **dataed.in/SOLAR**.

When accessing SOLAR using computers, then Adobe Flash Player is required on the machine (at least version 12 as a minimum). To use SOLAR using tablets, then delivery is available through the SurPass app (available from the Apple and Google App stores).

SOLAR is not linked to the SQA registration and certification systems, so candidates must be uploaded to the SOLAR system as well as submitted to SQA for registration and certification purposes.

Data Citizenship Unit, Level 4, 5, 6

Outcomes 1 and 2: At Levels 4 and 5, the assessment is a computer-based test, dynamically generated from a bank of questions that are all automatically marked. This is a closed book assessment consisting of 20 questions with a 60% pass mark.

At Level 6, this is a fixed assessment comprising of 11 extended response questions which are tutor-marked online through Solar using associated marking schemes. There is a 60% pass mark applied to this assessment.

Outcome 3: This computer-based project assessment supports digitally submitted evidence that is tutor-marked online through Solar using associated marking schemes. The assessment consists of 5 tasks at Level 4, 6 tasks at Level 5 or 7 tasks at Level 6. These are to be completed with a 60% pass mark across the assessment. This is an e-portfolio approach where candidates can access their project multiple times over a set period to upload their evidence.

Data Science Unit, Levels 4, 5, 6

Outcomes 1 and 2: At Levels 4 and 5, the assessment is a computer-based test, dynamically generated from a bank of questions that are all automatically marked. This is a closed book assessment consisting of 20 questions with a 60% pass mark.

At Level 6 for Outcomes 1 and 2, the assessment is dynamically generated from a bank of extended response questions. The assessment comprising of 12 extended response questions which are tutor-marked online through Solar using associated marking schemes. There is a 60% pass mark applied to this assessment.

Outcome 3: This computer-based project assessment supports digitally submitted evidence that is tutor-marked online through Solar using associated marking schemes. There are 6 tasks to be completed with a 60% pass mark across the assessment. This is an e-portfolio approach where candidates can access their project multiple times over a set period to upload their evidence.

Data Science: Project Unit, Level 5 and 6

This computer-based project assessment supports digitally submitted evidence that is tutor-marked online through Solar using associated marking schemes. There is 1 task per Outcome (4 in total) to be completed with a 60% pass mark across the assessment. This is an e-portfolio approach where candidates can access their project multiple times over a set period to upload their evidence.

Assessment checklists

There are assessment checklists for recording the progress of learners at the end of this Educators' Guide.

TOOLS AND SOFTWARE OPTIONS

The NPA has been designed to provide a great deal of flexibility in the software tools that can be used by learners for data gathering, analysis and visualisation. It was very important that every outcome could be achieved using a selection of different tools that are free (or free for educational use) or open source. It was also important that there were a range of tools that could be installed as well as a selection of online tools available.

In addition, the NPA can be achieved using only a basic spreadsheet tool, such as Microsoft Excel or Google Sheets. It is important that learners gain a secure grounding in their knowledge of data science and statistical concepts. Gaining experience of particular tools or programming language that might be favoured by industry is much less important to learners at this stage of their career.

Although it is not a requirement, Level 6 learners would benefit from carrying out data analysis and visualisation tasks using a programming language. This would give them an understanding of data science techniques using code and will be a better level of complexity. Level 6 can be taught using any of the tools though, and it is understood that using a programming language would be particularly challenging when teaching a multi-level class.
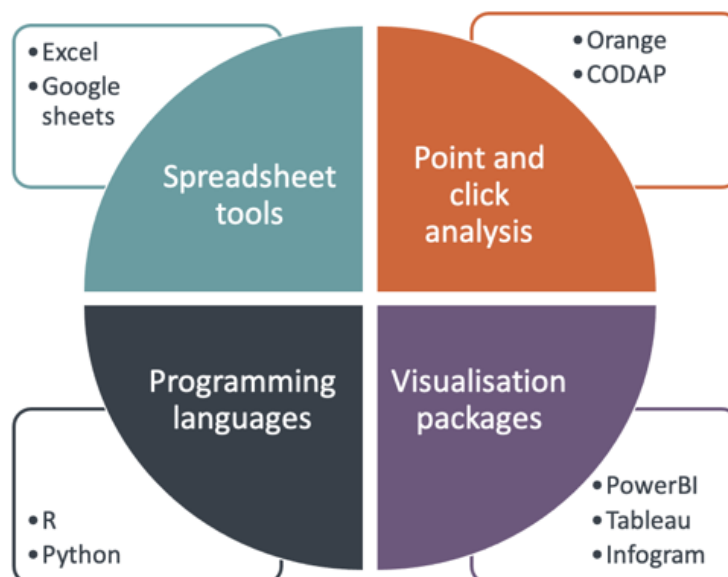
SPREADSHEET TOOLS

The two main spreadsheet options are **Microsoft Excel** (part of the Office 365 package at **Office.com**) and **Google Sheets** (part of the G Suite package at **dataed.in/sheets**)

The choice of which spreadsheet package to use will be largely guided by the local situation in the school or college. Schools in local authorities that mainly use Microsoft will likely opt for Excel, whereas schools in local authorities that use Google for Education and Google Classroom will prefer to use Sheets.

Google Sheets has a smaller size limit for datasets than Excel (5 million cells compared to 17 billion) but it is unlikely that learners will be using datasets this large. Both packages allow scripting, if required 9VBA in Excel, Apps Script in Sheets). Both packages allow multiple users to edit the same file, although Sheets handles this far better than Excel, which might be an important consideration for the Level 5 Data Science Project unit with its team project work. It is also clearer to see which user has made particular changes with the version control in Sheets than in Excel.

COMMERCIAL VISUALISATION PACKAGES

Some of the major visualisation tools used in industry offer educational versions or licences that cover school, college and university students and educators. The packages that might be of interest for the NPA are Microsoft's **Power BI**, **Tableau** and **Infogram**.

Power BI (**dataed.in/pbi**) is well supported with good training material, it works well online, and has an intuitive interface. Unfortunately, it doesn't come with data preparation tools, so learners will have to clean and manage the data in another tool such as Excel, before creating the visualisations in Power BI. Power BI should be available as part of Office 365 Educational licences.

Tableau (**dataed.in/tab**) Desktop and Tableau Prep are both available free to academic institutions. Educators need to apply for licences. There are support materials and webinars available on using Tableau with a variety of datasets (including movies and Eurovision). Tableau integrates well with databases and is fairly easy to use.

Infogram (**infogram.com**) is an online tool that has a limited free service available for educational use. Infogram makes it extremely easy to create dashboard visualisations. The basic package has limited functionality but should be sufficient for NPA learners.

POINT AND CLICK ANALYSIS AND VISUALISATION

There are two 'point and click' tools that would be suitable for NPA learners, **CODAP** and **Orange**.

CODAP (**dataed.in/codap**) is a free open-source online tool that has been specifically designed for education and for young people. It would be particularly suitable for Level 4 and 5 learners. The website contains an interesting variety of examples that can be used in class. The tool makes it very intuitive to explore a dataset, to the extent that you can click on a data point and the tool will highlight the row in the dataset for that value.

Orange (**dataed.in/orange**) is a slightly more advanced point and click tool for data analysis and visualisation. It is also open-source, like CODAP. It would require a slightly bigger learning curve than CODAP, however it has far greater functionality, including predictive modelling. Orange is not web-based and would need to be downloaded and installed.

Programming languages

For Level 6 learners, it is recommended (although not required) that they gain an experience of using a programming language for data analysis and visualisation. The two main languages are **R** and **Python**. Many centres' Computing Science departments are already using Python with learners, so this would be a reasonable choice for the NPA as well. Some centres already teach the Statistics unit in Maths using R, so for them R might be a better choice. Both languages are suitable for teaching the NPA at Level 6, so it would depend on local circumstances and preferences.

R (**dataed.in/R**) is a language designed for statistical analysis and visualisation. Add on tools, particularly **Rshiny**, can be used for more attractive visualisations, dashboards and web applications. Packages like **ggplot** can be used for enhanced data visualisation.

Python (**python.org**) is an object-oriented programming language that has easy to learn syntax which emphasises readability. It is a popular language in education and industry, and can be used for general purpose programming, games development, web development and data science. This makes is rather more challenging to find support materials for learning Python for Data Science specifically.

Add-on tools are required for Python to be used more easily for data analysis and visualisation. Popular options are:

- **Pandas** is a tool for data analysis and manipulation.
- **Numpy** is the Python library for mathematics, useful for performing operations on data.
- **Seaborn** is a library specialised in statistical data visualisation.
- **Matplotlib** is another library specialised in visualisation for Python.
- **Plotly** can be used for interactive plots

Installing R and Python

Python and R are free and open source. There are a number of different methods for installing the languages, so hopefully centres can find an installation option that suits local circumstances.   One of the easiest ways to install Python or R is to use Anaconda (**anaconda.com**). The Open source Individual Edition of Anaconda makes it easier to install both languages plus their associated data science packages and add-ons.

Using R and Python online

Both languages can be used online using interactive notebooks such as Jupyter using a free cloud-based service such as Binder (**mybinder.org**), Google Colab (**dataed.in/colab**) or Kaggle Kernels (**kaggle.com**). This means that software does not need to be installed on a local machine, allowing learners to be more flexible in where they work.

PEDAGOGY

Educators of the NPA may come from a variety of backgrounds.   Therefore, it might be helpful to look at pedagogy strategies from different subject areas to support the teaching of Data Science.

Pedagogy in Computing and Data Science

There are three phases to learning a new skill in Computing that apply equally well to Data Science. Firstly, learners will need to understand the concept or the theory behind the skill. Secondly, they will need to learn the 'language' of the tool that will be used to apply the skill. This might include seeing other people's programming code or seeing a demonstration of how to apply a skill. Lastly, learners can practice applying the skill themselves in different settings and circumstances.

For example, in order to teach how to load and view a dataset into Python (the practical skill), first we need to ensure learners know about the concepts of data and how it is stored in a structured format. Learners will then need to be shown the commands in Python and try loading up an example dataset, view the column headings and first few rows of the data. Only after they can do this in a few times in a limited and supported way (such as following a set of instructions) can learners go on to do this more confidently on their own and in different situations (such as finding their own dataset and loading it).

This three-phase approach enables teachers to identify and correct learner misconceptions early on - something which is notoriously difficult to do when Computing and Data Science education is centred on creation and coding.

This approach isn't just useful for programming languages but also for other skills. Learning the 'language' might be understanding the terminology and technology or understanding how to carry out a task using a piece of software. It is not just learning a set of steps, but understanding what those steps do, why those steps are in that order, why the skill is useful, what the limits are, and how it all works. This leads to far deeper learning with more secure knowledge that can be built upon with more advanced concepts and skills.

It is a spiral approach, where the learners will revisit concepts at increasing depth as they work through the course. This approach does not mean that learners must gain an understanding of all the concepts first and then all about the languages and tools before going on to develop and build and apply their new skills.

The Scottish 'Broad General Education' curriculum for Computing Science is based on this three-phase spiral. You can read more about the theory behind this curriculum in the Teach Computing Science guide for Secondary (**dataed.in/teachcs**)

The Centre for Computing Science Education at the University of Glasgow have produced an online course on 'Getting Started with Teaching Data Science' (**dataed.in/teachds**) that covers pedagogical issues. The College Development Network have a course for lecturers on data science (**dataed.in/CDN**). Although neither course covers the NPA, both courses would be useful and suitable to NPA educators, whichever setting they teach in.

Barriers to learning

There is a growing body of research into potential barriers to learning. Two issues that are most relevant to data science are cognitive load theory and misconceptions (also known as alternative conceptions).

Cognitive load theory

Cognitive load theory is based on the idea that stress on a learner's working memory reduces their ability to acquire new learning. If learners get bombarded with too many things to think about at once, then they can't process those concepts and ideas into long term memory.

The Pedagogy Quick Read for Cognitive Load (**ncce.io/qr01**) suggest some strategies for addressing cognitive load including using **worked examples** to provide scaffolding for novices, collaborative techniques such as **pair programming**, or annotating programs using comments identifying common sections or patterns, known as **subgoal labelling**.
(Additional information at **dataed.in/cog**)

Misconceptions

When learners have misconceptions or alternative conceptions, this can prevent them completing tasks, lead to frustration and can stop them learning about related concepts. The Teach Computing site has a great guide to avoiding and learning from misconceptions. It covers how misconceptions arise, different strategies for avoiding them, and links to collections of misconception in Computing. (Additional information at **dataed.in/miscon** and **dataed.in/altc**). The Science Teacher website talks about misconceptions within Science education **dataed.in/scimisc**

Strategies From Computing Science

The Teach Computing website explores a range of approaches that are supported by research. A few examples are listed below.

Peer instruction is an active learning technique that uses multiple choice questions are posed to pairs with the aim of forming a consensus and enabling educators to highlight and correct alternative conceptions. **ncce.io/qr04** and **dataed.in/qrpi** and **dataed.in/pi**

Worked examples are sample solutions are shared with learners and annotated with subgoal labels, providing a model for similar incomplete problems. **ncce.io/qr02** and **dataed.in/qrwe** and **dataed.in/we**

PRIMM is a framework that encourages students to begin by reading code and then to **P**redict, **R**un, **I**nvestigate, **M**odify, and **M**ake. **dataed.in/primm**
This deepens learners' understanding of new programming concepts and helps to addressing misconceptions.

Pair programming is when two learners work together on the same task, taking it in turns to 'drive' or 'navigate'. The driver controls the computing device, and the navigator provides support and direction. **ncce.io/qr03** and **dataed.in/qrpp**

Live coding is when a teacher develops the solution to a problem in front of the class for learners to follow. **ncce.io/qr05** and **dataed.in/qrlc**

Some other instructional techniques for teaching Computing are available at: **dataed.in/cstech**. You might also find it useful to look at the Teach Computing's 'pedagogy quick reads' and blog posts: **dataed.in/csped**.

There are also quick tips guides and short videos on a range of topics about teaching computing at **csteachingtips.org**, including pair programming, assessment, inclusivity and reducing bias.

Strategies from Science

The enquiry-based learning approach (**dataed.in/es4all**) works well for data education. There are different types of science enquiry activities:

- When learners **identify, classify & group** they make sense of the world. They can organise items into groups, observe similarities, and find additional matching items.
- By carrying out **comparative testing**, learners explore the relationship between variables. Using a tool such as CODAP will allow learners to carry out this comparative testing easily.
- Learners can collect and interpret data to **look for patterns** in the data.
- Through **observing over time** learners can identify and measure events and changes.
- Carrying out **research using secondary sources** learners will develop strategies to evaluate sources, recognise conflicting evidence and bias.

Strategies from Maths and Statistics

Cambridge Maths have brief guides to research and different pedagogies for teaching Information handling, statistics and other concepts. They have guides on several topics, such as:
- **exploratory data analysis**, an informal and exploratory approach to statistics, drawing different representations, searching for patterns and considering "what is going on here?" **dataed.in/cmeda**
- **teaching the concept of the mean**, describing the research important to consider when students are developing their concept of the mean **dataed.in/cmmean**
- effective ways to learn from **comparing data sets dataed.in/cmdata**

The American Statistical Association has an excellent report on 'Guidelines for Assessment and Instruction in Statistics Education' (GAISE) which provides a framework for how learners can develop their understanding of different concepts in statistics. **dataed.in/GAISE**

Dataspire have a set of professional learning mini-lectures for educators starting to teach statistics and data analysis (**dataed.in/minidata**) on topics like types of data, organising data, teaching graphs, and making effective visualisations.

TEACHING DATA SCIENCE

WHAT IS DATA?

Level 4

DC4.1a State the reasons for the growth of data
DC4.1d State common sources of public and private data

Level 5

DC5.1a Describe the reasons for the growth of data
DC5.1d Describe sources of public and private data and the concept of open data
DS5.1e Identify sources of public and private datasets

Level 6

DC6.1a Explain the technological, economic and societal reasons for the growth of data
DC6.1d Explain types and sources of large datasets and the philosophy of open data
DC6.2a Explain the concepts of data volume, variety, velocity, veracity and value
DS6.1f Explain the principle of open data and sources of open data

What learners should be able to do

L4-6: Find datasets and data visualisations from different public and private sources.
L6: Find datasets and data visualisations from different open sources.
L6: Assess data using the 'V's of data.

What learning could look like

Learners will be able to gain an understanding of sources of datasets and be able to identify situations where data is used. They will be able to explain where data about themselves might be stored and used.   They could think about their own data footprint (**dataed.in/datafoot**) such as how Google knows where they have been (**dataed.in/gmaps**) or their fitness levels (**dataed.in/gfit**).

In understanding the growth of data, learners could look at the history of data visualisations such as Florence Nightingale's rose visualisation, John Snow's cholera map, or the visualisation of Napoleon's march on Russia, as shown in the videos from 'Unlocking the World of Data' (**dataed.in/uwd1**)

In understanding the contemporary growth of data, learners could do the iDEA Awards' Big Data badge activity (**dataed.in/idea1**) and look at the 'A minute on the internet' graphs and the changes in recent years (**dataed.in/1min**). Learners could look at use of data in different sectors such as weather (**dataed.in/MET**) or local walking tours (**dataed.in/beer**).

As well as looking at personal data, learners can look at data about their community (**dataed.in/shine**), like looking at changing temperatures in their area (**showyourstripes.info**), life expectancy (**dataed.in/life**) or in surveys and the census (**dataed.in/census**).

Educators wanting to explore public, private and open data in a fun and engaging way could use the ODI's Datopolis board game (**dataed.in/game**).

Timings

The core concepts in this topic can be covered in two or three main lessons. One lesson could cover what data is, how it's being used and where it comes from, including sources of data that learners will have personally generated or encountered. One lesson for L6 learners will cover the importance and value of data, including the 7 'V's. Another lesson will cover sources of public and private data (and open data for L5/6).

However, the concepts in this topic will be reinforced throughout the course. Learners will continue to get data from a variety of sources and assess the value of each dataset they encounter, as well as the other 'V's. They will hopefully also continue to realise how data can be used to make an impact to their lives and to people in their community.

INTERPRETING DATA

Level 4

DC4.2c Describe types of data visualisations
DS4.2e Describe simple data visualisations

Level 5

DC5.2c Explain types of data visualisations and the best use of each type
DC5.3a Extract information from data visualisations and dashboards.

Level 6

DC6.3a Extract information from data visualisations and dashboards.

What learners should be able to do

L4-6: Interpret a range of data visualisations correctly
L4-6: Interpret bar charts, lollipop charts, histograms, line graphs, pie charts, donut charts
L5/6: Interpret overlaid histograms, time series graphs, slope graphs, scatterplots, bubble plots, heatmaps, treemaps, waffle charts, stacked bar chart, violin plots
L6: Interpret box plots, stacked area charts, density charts, sankey diagrams, waterfall charts, maps
L5/6: Be able to spot errors with 'bad graphs' and understand why they are misleading

What learning could look like

It is anticipated that after a few initial lessons, interpreting data could be taught throughout the course as a series of class opener activities, such regularly displaying a data visualisation as learners are setting into class. The Seeing Data project has videos and activities for teaching this at **seeingdata.org**. Learners could also explore **gapminder.org** and watch the videos there.

The New York Times have a range of activities (**dataed.in/nyt**) for schools, contemporary graphs with discussion prompts. They use a Notice and Wonder approach which lets learners begin analysing graphs simply using active questioning and curiosity. Learners can build confidence and acquire new conceptual understanding. Over time their critical thinking skills develop and their vocabulary grows.

These are the questions that you could ask learners about a graph:

What do you notice?

Every learner can notice something in a graph, whether it's a data point, a trend or pattern. As they hear each other's observations, they dig deeper. If learners make a claim, ask them to say what they noticed that supports their claim.

What do you wonder?

Learners can discuss what they are curious about that comes from what they have noticed earlier. By hearing other people's ideas, learners form more and deeper stories from the graph. For time series data, you could ask learners to predict what will continue to happen to the data in the future.

What impact does this topic have on you and your community?

A good visualisation should prompt action such as a behaviour change in the audience.

What's going on in this graph?

Just like photographs, graphs tell stories. Ask learners what's happening, what story can the graph tell? Ask learners to write a catchy headline that captures the main idea as a way of summarising their understanding.

The Turner's Graph of the Week (**dataed.in/gow**) also has a set of common questions to ask when interpreting graphs:

- Is there an upward or downward trend?
- Are there any sudden spikes in the graph?
- What is being compared in the graph?
- What prediction can I make for the future?
- What inferences can I make about the graph?

These approaches can be used with other sources of graphs. There are many great visualisation collections online: **informationisbeautiful.net**, **statista.com**, **chartr.co**, **ourworldindata.org**, **pudding.cool**, **public.tableau.com/en-gb/gallery** and **dataed.in/gap**.

When looking at bad and misleading graphs with L6 learners, the Computerphile series have a useful video explaining this (**dataed.in/comp**). There are many great examples at **viz.wtf** and in Reddit **dataed.in/bad**. The Spurious Correlations site is useful too **dataed.in/spur**.

Timings

There might be two to three lessons on interpreting data. A lesson on interpreting graphic information, refreshing chart types that learners will have encountered in Broad General Education. A second lesson could introduce some of the different chart types that learners will encounter throughout the course, discussing when these are best used. L5/6 could also have a lesson on bad graphs, looking at inappropriate or misleading visualisations.

WHAT IS DATA SCIENCE?

Level 4

DS4.1a State the reasons for the development of data science
DS4.1b Describe contemporary applications of data science
DS4.1c Describe the steps in solving a problem using data science

Level 5

DS5.1a Describe the reasons for the development and growth of data science
DS5.1b Describe contemporary application of data science
DS5.1c Describe the data science life cycle including the potential for bias at each stage
DS5.1f Describe the role of domain knowledge and subject matter experts in data science

Level 6

DC6.2e Explain the role of domain knowledge within data science
DS6.1a Explain the relationship between artificial intelligence, machine learning, big data and data science
DS6.1b Explain the technological, economic and societal reasons for the development and growth of data science
DS6.1c Describe contemporary applications of data science and the types of problem that data science can address
DS6.1d Explain the data science life cycle and the significance of domain expertise
DS6.1e Explain descriptive analytics and predictive analytics

What learners should be able to do

L4-6: Describe how data science has grown
L4-6: Describe problems that have been solved using data science
L4-6: Understand the data science lifecycle can be used to solve problems with data
L5/6: Explain the role of people in data science projects like data scientists and domain experts
L6: Explain the terms AI, machine learning, big data and data science

What learning could look like

Learners will gain an understanding of the types of problem that data science has successfully resolved. Learners could be introduced to the application of data science by finding out about how it is used in many different fields. The 'Unlocking the World of Data' video on ice cream sales is a gentle introduction (**dataed.in/uwd2**) as is this video of data professionals talking about their work (**dataed.in/work**).

The **databasic.io** tools are a great way to explore how powerful data analysis can be in a very accessible way with fun topics. They have lesson plans for guiding learners through exploring a CSV dataset.

The Digital World site has a guide to careers in Data Science in Scotland and the sort of tasks people in those roles would carry out. (**dataed.in/careers**). The MEI has also produced videos of data scientists in different fields talking about their work (**dataed.in/MEI**)

For L6 learners the Computerphile video (**dataed.in/comp1**) explains the difference between ML, AI, big data and data analysis, and the Friendly Guide to ML goes into more detail (**dataed.in/ML**).

Timings

It is anticipated there will be two lessons looking at applications of data science across different aspects of personal, business and government and looking at the data science lifecycle. In addition, there would be a lesson for L6 learners looking at data science in business and across AI and machine learning.

WORKING WITH DATA

Level 4

DS4.2a Describe common data types and data formats
DS4.2b Describe structured and unstructured data

Level 5

DS5.2a Describe common data types and data formats
DS5.2b Describe the composition of a structured dataset

Level 6

DS6.2a Describe common data types and data formats including structured and unstructured data
DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3a Define the required analyses and data models.
DS6.3b Create a relational data model from external sources of data.

What learners should be able to do

L4-6:	Describe the difference between qualitative and quantitative data
L5/6:	Give examples of qualitative data: nominal and ordinal
L5/6:	Give examples of quantitative data: continuous and discrete, interval and ratio
L4-6:	Give examples of structured and unstructured data
L5/6:	Use a tidy data structure for data
L4-6:	Understand the difference between data types, display formats and file formats

What learning could look like

It is anticipated that learners will be introduced to increasingly more complex data categories, types, structures and file formats throughout the course. Ideally learners should be introduced to each concept first before then learning how to use and manipulate that data type or structure in an application or programming language.

The 'Unlocking the World of Data' video on 'What is Data?' is a gentle introduction to different types of data with Edinburgh examples (**dataed.in/uwd3**). The Computerphile series has a clearly explained introduction video (**dataed.in/comp2**).

Some of the concepts can be introduced simply using sticky notes. Ask learners to write down answers to a few questions such as name, date of birth, shoe size, favourite colour, place of birth. Then ask them to sort them on a long wall (or online in a tool like **Padlet.com** or **jamboard.google.com**) by various categories, like ascending by height. Ask them to sort by favourite colour should lead to interesting discussions on how colour can be categorised and sorted (alphabetically if we think of it as text, numerically if we assign a RGB colour code).

Timings

Throughout the course learners will be introduced to increasingly more complex data categories, types, structures and file formats. When first being introduced to data types and formats, there could be a lesson on basic data categories and structures. When learning how to practically apply and use these data types in a software package or programming language, learners could then be introduced to different display formats using that tool.

SECURITY

Level 4

DC4.2d Describe simple methods of managing and securing data

Level 5

DC5.2d Explain methods of managing and securing data
DS5.2d Describe methods of securing and managing data

Level 6

DC6.2d Explain methods of data management and data security
DS6.2c Explain data management and data security techniques

What learners should be able to do

L4-6: Know how to keep devices and personal data safe with anti-virus software, firewalls, secure passwords and use of password managers
L5/6: Know how to use multi-factor authentication
L5/6: Know that biometrics, encryption and backups can be used to protect data

What learning could look like

There are an increasing number of external resources that are suitable for teaching data security to learners at this level. The Young Scot's 'DigiAye' website has some great resources and activities (**dataed.in/aye**). The SQA have guides and worksheets for the Data Security unit in the NPA Cybersecurity course (**secure.sqa.org.uk** login required) and there is online materials for this unit (**dataed.in/jhigh**). The Cyberskills live lessons site has a great activity called One Million Passwords (**cyberskillslesson.com**) based on the LinkedIn data breach.

There are many activities online for testing the strength of password, such as **howsecureismypassword.net**. You could also play password bingo with your learners (**dataed.in/bingo**). Actua Codemakers have a lesson on passwords (**dataed.in/actua**) that links to a good video on how companies encrypt your passwords (**dataed.in/passvid**). Learners could read about password breaches (**dataed.in/773m**) and check if their email addresses have been compromised (**haveibeenpwned.com**) and if their passwords have been leaked before (**dataed.in/pwnedpass**) L5/6 learners can research previous breaches (**dataed.in/pwned**) and find out how to crack passwords and about encryption from Computerphile (**dataed.in/comp3**)

Once learners have learners about password managers (**dataed.in/effpass**) they should be encouraged to set up a password manager app or account (such as **RememBear.com** or **1password.com**) and use a super-secure password for the account (**dataed.in/effdice**). L4/5 learners can find out about malware, hacking and pwning (**dataed.in/effprotect**) and encryption (**dataed.in/encrypt**). All learners could read the Electronic Frontier Foundation's guide on privacy for students (**dataed.in/effpriv**) and discuss other reasons to protect your data.

It is hoped that all learners would gain practical experience in setting up anti-virus and firewall software on a personal device, if they have one. They should also be encouraged to set up a VPN (such as **Tunnelbear.com**) and password manager app if they have a mobile device. Learners could read up on suitable free options and compare them. They should think about themselves as the product when using free services, could find out the security companies use to protect their users' data. If learners have accounts with companies using multi-factor authentication (such as Google, Epic, Amazon, Facebook) then they should be encouraged to set this up on their accounts, if possible.

Timings

The timing of lessons depends on how much practical experience it is possible to give learners. This will depend on availability of mobile devices or personal computers and access to web services that use multi-factor authentication (as local authorities or Colleges may have blocked these in web filtering). Some of this could be homework tasks for learners. It is anticipated there may be two or three lessons on this topic. One lesson could cover protecting accounts with passwords, password managers and multi-factor authentication. One lesson could cover protecting personal devices with anti-virus software, firewalls and VPNs. A third lesson for L5/6 learners could cover biometrics, encryption and backups.

PRIVACY

Level 4

DC4.1e State the rights and responsibilities of data subjects and data owners

Level 5

DC5.1e Describe the rights and responsibilities of data subjects and data owners

Level 6

DC6.1e Explain the rights and responsibilities of data subjects and data owners

What learners should be able to do

L4-6: Know about your data rights under GDPR
L6: Know about lawful processing of data under GDPR
L4-6: Know ways to view and manage your information online
L4-6: Know ways to maintain privacy online

What learning could look like

Learners could get an introduction to GDPR through the iDEA Awards badge (**dataed.in/idea2**). Learners could use their data rights to request their data from an online service or social media and inspect the data that company stores about them. They could also watch the Data Dollar Store video (**dataed.in/dollar**) about how we value our data. Discuss with learners about which companies and organisations they trust with their data. Data can be shared for good purposes too, such as for medical research (**registerforshare.org**)

Learners could investigate photo metadata through the Cyberskills live lesson on 'Every Picture Tells a Story' (**dataed.in/pic**) to track cyber criminals by learning basic Python commands to analyse geolocation data. They could see how geotagged data can accidentally reveal private information in **iknowwhereyourcatlives.com** or see data from their own images with **metapicz.com**. Learners could explore the issue of apps selling on location data about users (**dataed.in/location**)

Learners could think about the terms and conditions for apps and services by watching people's responses to reading them (**dataed.in/tc**), through a quiz (**dataed.in/bbctc**) or by reading simplified version of them (**dataed.in/simpletc**). There's also a great graphic novel on this topic (**dataed.in/gfxtc**)

Learners could find out about cookies and how their data is being shared with other companies (**dataed.in/cookies**) and could read more about it (**dataed.in/ads**). Learners could use the Blacklight tool (**dataed.in/blacklight**) to see what user-tracking exists on their favourite websites, or they could use the Track This service (**trackthis.link**) to confuse trackers.

There are more activities on the 'My Data and Privacy Online' toolkit (**dataed.in/mydata**), which includes instructions on how to change your privacy settings on social media (**dataed.in/privacy**). Educators could also play the Privacy Chicken game with learners (**dataed.in/chicken**).

Timings

Like the Data Security topic, the timing of lessons depends on how much practical experience it is possible to give learners. This will depend on access to web services and social media to export their data and secure privacy settings (as local authorities or Colleges may have blocked these in web filtering). Some of this could be homework tasks for learners.

It is anticipated there may be two or three lessons on this topic. One lesson could cover GDPR and their data rights under the law. A second lesson could cover the information they share online and how to protect their privacy.

MANUAL DATA CAPTURE

Level 5

DS5.3b Capture data from an external source.

Level 6

DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling

What learners should be able to do

L5/6: Manually gather data using a survey
L5/6: Get data from an external source and assess the quality

What learning could look like

Although this topic mainly covers only two performance criteria, a lot of background knowledge and advice is required to do data capture well while avoiding bias.

There are good resources for teaching surveying from the New Zealand Census resources (**dataed.in/nzc**). There is also a good video from the US Census Bureau on why we gather data (**dataed.in/usc**).

The Teach Computing website has a unit of work on data science that features data gathering in lesson 4 that would be suitable for L4/5 learners. It is based around the PPDAC data science lifecycle (**dataed.in/nccedata**).

There are also resources on sampling, which comes up later in the Statistics topic but is useful to think about when gathering data. The 'Unlocking the World of Data' series have videos on sampling (**dataed.in/uwd4**) and the Royal Statistical Society have activities on random sampling (**dataed.in/rssrandom**) and bias (**dataed.in/rssbias**). Dataspire have a video explaining sampling and bias (**dataed.in/spire**)

Timings

There could be one introductory lesson on good and bad survey design, best practice and avoiding bias. A second lesson could involve the learners designing and creating their own survey. Learners could then for homework ask other people to complete the survey (or send out an electronic survey link and wait for responses.) A third lesson could involve looking at the survey responses, entering them from paper if gathered by hand, and tidying the data.

DATA MANIPULATION

Level 4

DS4.2c Describe simple methods of cleaning and transforming data
DS4.3a Perform simple data cleaning and structuring.
DS4.3b Perform basic analyses including sort, filter, group and summarise.

Level 5

DS5.2c Describe methods of cleaning and transforming data
DS5.3c Perform routine data cleaning and structuring.
DS5.3d Perform analyses including query, sort, filter, consolidate, group and summarise.

Level 6

DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3c Perform data transformation to complete, correct and structure data.

What learners should be able to do

L4-6: Manipulate a single table to select or reorder columns, create new variables, reformat or extract columns, filter, sort and deduplicate rows
L4-6: Summarise data to get counts, totals, averages and min/max values
L5/6: Merge two datasets (left, right, inner or full)
L5/6: Reshape a dataset (wide or long)

What learning could look like

Dataspire have a set of resources on organising data and learning how to use Google Sheets (**dataed.in/spireorg**). Dataspire also have a mini-lecture video and slides on organising and reshaping data that is aimed at professional learning for teachers but these slides would be suitable to use with learners (**dataed.in/reshape**).

Learners could find out more about cleaning a chocolate dataset from Computerphile (**dataed.in/comp4**). The series also has useful videos on data transformation (**dataed.in/comp5**) and data reduction (**dataed.in/comp6**) which looks at reducing rows and columns in a music dataset.

Educators might be interested in Cambridge Maths' guide on teaching learners on how to explore and compare datasets (**dataed.in/compare**)

Timings

The timings of these lessons will depend on the prior learning of the learners. Some learners may have used the tools (such as Excel, Sheets or Python) before or have a familiarity through studying Computing Science, Business or Administration, but other learners may need a couple of lessons to introduce them to using the tool, entering data, opening existing datasets, exploring the data.

Learners will need a lesson or two to cover filtering and sorting rows and columns, selecting and reordering and reformatting columns. They will need a couple of lessons on summarising data and creating new variables.

L5/6 learners will need a couple of lessons on merging datasets, including time to practice different types of join. They will also need a couple of lessons on reshaping data and tidy data, including practical exercises in changing datasets from long to wide (and from wide to long).

STATISTICS

Level 4

DS4.2d Describe basic descriptive statistics used to summarise a dataset

Level 5

DS5.2e Describe descriptive statistics used to summarise a dataset including measures of central tendency, dispersion and correlation

Level 6

DS6.2d Explain statistical techniques involved in data science.

What learners should be able to do

L4-6: Describe population and samples
L4-6: Summarise data with basic descriptive statistics, including mean, mode and median
L5/6: Describe the skewness of data and the effect of outliers on the dispersion
L5/6: Describe distributions of data
L6: Describe dispersion, variance and standard deviation

What learning could look like

For a comprehensive overview of why statistics are useful to us, learners could watch the 'Joy of Stats' documentary (**dataed.in/joy**) presented by Hans Rosling from Gapminder. The 'Unlocking the World of Data' series have well-presented videos explaining averages and measuring spread (**dataed.in/uwd5**). The Data to Insight course videos explain a range of concepts clearly and would be good for flipped or blended learning (**dataed.in/nzstat**). Khan Academy also have a range of videos explaining different topics in statistics (**dataed.in/khan**). The 'Calling Bull' course has a video on spotting when mean and median are used badly that would be suitable for L6 learners (**dataed.in/calling**)

The MEI have some ideas on practical and fun activities for teaching statistics (**dataed.in/MEIstats**). There is also a great activity for making 'human bar charts' with learners to teach normal distribution (**dataed.in/rixmas2**) on the Royal Institution Christmas lectures.

There are lots of web-based tools for demonstrating and exploring different statistics concepts. The Art of Statistics site has web apps for playing with categorical data, quantitative data, time series, mean vs median, correlations, and distributions (**dataed.in/art**). The Book of Apps for Statistics Teaching (or BOAST) site has apps on bias, correlation, descriptive statistics, outliers, variance, time series, sampling, reshaping and visualisation (**dataed.in/boast**). The 'Statlets' site (**dataed.in/statlets**) is more basic but covers averages, correlation, sampling, distributions, variance and Simpson's paradox.

Timings

During this topic, learners will be learning and securing their understanding of different statistical concepts. It is important that they have a chance to explore these concepts before moving on to applying their knowledge with real datasets and tools. Web apps are particularly valuable for exploration, and hopefully highlighting any learner misconceptions.

All learners will need two or three lessons covering population and samples and summarising data. Level 4/5 learners may need longer to cover mean, mode and median, particularly if they have not covered this in Level 4 Information Handling in Broad General Education (MTH 4-20b). Population and sampling should have been covered (in MTH 3-20b) but a refresher may be required.

L5/6 learners will need a couple of lessons to cover skewness, outliers and distributions and L6 learners will need a lesson or two on dispersion, variance and standard deviation. Although dispersion is included in L5, this would not be in any great depth or detail.

DATA ANALYSIS

Level 4

DC4.2b Describe how data can be analysed
DS4.2c Describe simple methods of cleaning and transforming data
DS4.3a Perform simple data cleaning and structuring.
DS4.3b Perform basic analyses including sort, filter, group and summarise.

Level 5

DC5.2b Explain how data can be analysed
DS5.2c Describe methods of cleaning and transforming data
DS5.3a Define the required analyses.
DS5.3c Perform routine data cleaning and structuring.
DS5.3d Perform analyses including query, sort, filter, consolidate, group and summarise.

Level 6

DC6.2b Explain how data can be analysed and the tools that can be used to perform analysis
DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3a Define the required analyses and data models.
DS6.3d Perform descriptive and predictive analyses on the data.

What learners should be able to do

L4-6: Analyse a dataset to get size, shape and summary statistics
L4-6: Clean and tidy a dataset
L5/6: Deal with missing values, outliers and duplicates
L4-6: Manipulate and transform the data
L5/6: Identify any patterns and correlations in the data
L6: Analyse a dataset to find the standard deviation, interquartile range, number of missing and distinct values

What learning could look like

Much of the learning in this topic will centre around the tool that has been chosen by the educator and learners. It is expected that for each of the activities listed above, learners will first learn about the concept (if not already covered previously in the course). They will then learn how to apply that concept simply in the tool of choice (such as learning how to tidy a simple dataset), before then applying that skill more independently in a more complex dataset.

For examples of simple data analysis activities, L4/5 learners could analyse how much KPop bands earn on YouTube (**dataed.in/BTS**) or analyse data on streaming media services (**dataed.in/stream**)

There are more complex environmental and science data analysis activities and lessons from Australia (**dataed.in/aus**). Closer to home, the Institute for Research in Schools (IRIS) have a regular series of projects for learners to work with authentic research (**dataed.in/iris**) which could involve analysing science data. The College Development Network in Scotland have work-through data analysis examples and videos using Excel (**dataed.in/cdnex**)

The Girls in Data project have run a series of engaging data analysis and visualisation challenges on the topics of film and TV (**dataed.in/girlstv**), music streaming (**dataed.in/girlsmusic**) and social media influencers (**dataed.in/girlsmedia**). The Stat Wars also have a film and TV challenge as well as a climate change challenge (**statwarscompetition.com**)

Timings

As stated above, the learning in this topic will depend on the tool selected. The amount of time required will depend on the prior learning and experience of the learners as well.

VISUALISATION AND STORYTELLING

Level 4

DC4.3a Create visualisation to identify patterns and trends in the data.
DC4.3b Draw conclusions from data.
DC4.3c Make recommendations based on conclusions and communicate findings.
DS4.3c Visualise the data to provide basic insights.
DS4.3d Create a simple report to communicate insights.

Level 5

DC5.3a Extract information from data visualisations and dashboards.
DC5.3b Interpret data to identify patterns and trends and draw conclusions.
DC5.3c Create appropriate visualisations from data.
DC5.3d Communicate findings and make recommendations based on conclusions.
DS5.2f Describe the selection of data visualisations to illustrate different types of data
DS5.3e Visualise the data to provide insights.
DS5.3f Create an interactive data dashboard to identify patterns and trends.

Level 6

DC6.2c Explain data visualisations and data storytelling
DC6.3a Extract information from data visualisations and dashboards.
DC6.3c Interpret data to identify patterns and trends and draw conclusions.
DC6.3d Create appropriate visualisations from data.
DC6.3e Communicate findings and make recommendations based on conclusions.
DS6.2e Explain techniques for data visualisation, data dashboards and data storytelling
DS6.3e Create data visualisations and data dashboards to provide insights.

What learners should be able to do

L4-6: Choose an appropriate visualisation type to communicate findings
L4-6: Create visualisations that clearly communicate meaning and insights
L4-6: Be able to create visualisations such as bar charts, lollipop charts, histograms, line graphs, pie charts, donut charts
L5/6: Be able to create visualisations such as overlaid histograms, time series graphs, slope graphs, scatterplots, bubble plots, heatmaps, treemaps, waffle charts, stacked bar chart, violin plots
L6: Be able to create visualisations such as box plots, stacked area charts, density charts, sankey diagrams, waterfall charts, maps
L5/6: Create a dashboard visualisation

What learning could look like

Learners could explore visualisations through drawing visualisations in the style of the Dear Data book (**dataed.in/deardata**) or by creating them using objects (**dataed.in/objects**) by using crafting skills (**dataed.in/craftvis**), or even food (**dataed.in/foodvis**). This allows learners to focus on the concepts they are learning rather than the tool being used.

The Information is Beautiful site (**informationisbeautiful.net**) have well designed visualisation examples, and their founder David McCandless has a TED talk video and lesson which would be a good introduction for learners (**dataed.in/iibted**). The iDEA Awards badge on Data Visualisation would also be great to introduce the topic to learners (**dataed.in/idea3**).

The MET Office have an activity on visualising data for different audiences (**dataed.in/meto**). The Data 101 Toolkit has a presentation and activities for introducing visualisation as a topic (**dataed.in/101vis**). They also have a workshop on storytelling with data (**dataed.in/101story**).

There are resources online for helping learners chose the best visualisation. The FT have a poster on github of different types of visualisations and when best to use each type (**dataed.in/ft**) as do the

DataViz Project (**datavizproject.com**). The Narrative patterns (or 'napa') cards support data-driven storytelling (**napa-cards.net**). Dataspire have a matching activity for learners (**dataed.in/vismatch**) and other resources (**dataed.in/spirevis**). Computerphile have a video that would be suitable for L5/6 learners (**dataed.in/comp7**)

When looking at storytelling, The Pudding visual essays are excellent examples (**pudding.cool**). The Storytelling with Data have regular challenges that might be of interest to learners (**dataed.in/swd**). Data comics are also good examples of using visualisations in an accessible way (**datacomics.net**)

In this topic learners will be creating visualisations that they have previously encountered in the 'interpreting data' topic. Some of these visualisations are at a suitable level for learners to have learned how to read, but may be too complex to create for learners of that level. For example, L6 learners can interpret sankey diagrams, but will likely find it too difficult to create these themselves. Learners should not be asked to create visualisations in this topic that were beyond their level to interpret earlier on in the course.

L5/6 learners are expected to create an interactive dashboard in the performance criteria. To clarify this, in the context and levels of the NPA, dashboard means two or more related visualisations, and interactive means being able to click on points on those visualisations for more information. A tool such as CODAP (**dataed.in/codap**) or Infogram (**infogram.com**) would be excellent for creating a dashboard of a suitable complexity level for NPA learners.

Timings

Learners will need to be shown how to create different visualisation types in the tool that has been selected for the class. The amount of time required will depend on the tool choice as well as the prior learning and experience of the learners.

Learners will need to be taught about producing visualisations in the tool that has been chosen for the class. L5/6 learners will also need to learn about dashboards and metrics. Learners will also need to be taught about the communication of insights and presenting visual information.

DATA QUALITY AND MANAGEMENT

Level 4

DC4.2d Describe simple methods of managing and securing data

Level 5

DC5.1e Describe the rights and responsibilities of data subjects and data owners
DC5.2a Explain the characteristics of high quality data
DC5.2d Explain methods of managing and securing data
DS5.2d Describe methods of securing and managing data

Level 6

DC6.1e Explain the rights and responsibilities of data subjects and data owners
DC6.2d Explain methods of data management and data security
DC6.3b Evaluate a dataset in terms of its quality including potential bias.
DS6.2c Explain data management and data security techniques

What learners should be able to do

L4-6: Assess the quality of a dataset using the data quality dimensions
L5/6: Look at ways that organisations care for their data
L5/6: Give examples of metadata
L6: Give examples of reference and master data
L6: Understand how businesses manage and secure data

What learning could look like

There are very few suitable external resources currently for teaching this area of the course.

Once learners have covered the core concepts in data quality they could select and assess a dataset against the data quality dimensions and present their assessment to their class or group.

After learning about caring for data, L5/6 learners could discuss what is meant by "data as an asset". Think about how companies might value data on their balance sheet? Perhaps discuss the value of tangible vs intangible assets to different companies such Uber and AirBnB.

L5/6 learners could find examples of metadata. They could create a data dictionary for a dataset of their choice.

L6 learners could discuss why these data management areas are important. Why couldn't you do without these? What could happen if an organisation fails to manage their data in each of the areas? Can learners identify cases where companies have failed to manage their data?

Timings

It is anticipated that one lesson would cover the core concepts on data quality. L5/6 learners would have a lesson on caring for data as well as metadata. L6 learners would need a third lesson that covers data management.

ETHICS AND BIAS

Level 4

DC4.1b State how data is used and misused by individuals, organisations and society
DC4.1c State the types of data bias and its impact on society
DS4.1d Identify sources of bias in data science including historical bias

Level 5

DC5.1b Describe how data is used and misused by individuals, organisations and society
DC5.1c Describe types of data bias and its impact on individuals and society
DS5.1c Describe the data science life cycle including the potential for bias at each stage

Level 6

DC6.1b Explain how data is used and misused by individuals, organisations and society
DC6.1c Explain types of bias and its impact on individuals and society
DC6.2f Explain the concept of data ethics
DC6.3b Evaluate a dataset in terms of its quality including potential bias
DS6.1g Explain data ethics, including data bias, with reference to national and international standards and frameworks
DS6.3f Identify potential sources of bias in the analysis

What learners should be able to do

L6: Data ethics, ethical frameworks and ethical challenges
L4-6: Sample and exclusion bias
L5/6: Measurement, confirmation, stereotype and survivorship bias
L6: Simpson's Paradox and correlation bias
L4-6: Impacts of bias in data science and ways to mitigate bias

What learning could look like

Learners could try to make an AI fairer than a judge in a courtroom algorithm game (**dataed.in/crime**) that has an article and explorable explanation on racism in predictive algorithms and the impacts of bias. The Royal Institution Christmas lectures have a section on bias in legal decisions and image recognition, including tricking algorithms with makeup (**dataed.in/RIxmas**)

'AI, Ain't I A Woman?' is a poem and video by Joy Buolamwini (**dataed.in/aigender**) that demonstrates the problem of facial recognition algorithms that have been trained using images of mainly white skin tones. Learn more about how she's fighting bias in algorithms in her TED talk (**dataed.in/aited**). Caroline Criado Perez has written about biased data or the lack of data about gender in her book 'Invisible Woman'. Learners could watch her summarise her main points (**dataed.in/women**). Google have an explorables explanation on hidden bias in data using the example of student grades (**dataed.in/exambias**)

The UnBias project has a useful set of awareness cards and activities that can used to discuss causes of bias and reflect on the impacts (**dataed.in/unbias**). Learners could then research online to find examples of bias, such as facial recognition in shops to catch shoplifters (**dataed.in/aishop**), mass surveillance in China using a phone app (**dataed.in/China**) or data from a charity app being used by the government to deport people sleeping rough on the streets (**dataed.in/homeless**)

Themes of privacy and surveillance are looked at in the short fictional film 'Frames' (**dataed.in/Frames**). The ethics of facial recognition and tracking could be discussed by L6 learners after watching the film.

Timings

Learners will need one lesson looking at the different causes of bias and researching and discussing examples of these. They will need a second lesson to investigate the impacts of bias and how to mitigate these, ideally looking at examples. More time will be needed to debate and explore these areas in more detail or for learners to do their own research. L6 learners will need to spend at least a couple of lessons on different ethical risks, ethical frameworks and examples of ethical issues in data. More time will be needed if learners will be searching for recent examples themselves than if they are provided a set of examples.

TOOLS AND LANGUAGES

Level 5

DS5.1d Describe the tools that can be used at each stage in the life cycle

Level 6

DC6.2b Explain how data can be analysed and the tools that can be used to perform analysis

What learners should be able to do

L5/6: Learners should be able to make a reasoned choice about the software packages and tools they use for different stages of the data science life cycle.

What learning could look like

Learners need to understand there are different tools and different methods that can be used to achieve the same task. Some software is multi-purpose, allowing the user to use the same tool all the way through the data science life cycle. Other tools have a narrower focus but are designed to do that job well, be more advanced or be much easier to use.

Learners should be able to use one software package or one set of tools well for going through the data science life cycle. They should also be aware of alternatives tools and have tried some of them. Ideally learners should be able to make their own choice of tools to use, particularly for project work.

After learners have been introduced to a range of tools and packages, they could compare them and discuss their preferred tool choices for different tasks.

Resources and Links

More information about tools and software is listed separately in the 'Tools and Software Options' section of this guide.

Microsoft Excel (**office.com**)
Google Sheets (**dataed.in/sheets**)
Power BI (**dataed.in/pbi**)
Tableau (**dataed.in/tab**)
Infogram (**infogram.com**)
CODAP (**dataed.in/codap**)

Orange (**dataed.in/orange**)
R (**dataed.in/R**)
Python (**python.org**)
Binder (**mybinder.org**)
Google Colab (**dataed.in/colab**)
Kaggle Kernels (**kaggle.com**)

Timings

It is not anticipated that learners will spend time learning about specific packages, but rather be introduced to software and tools through the context of learning the other topics in the progression pathway.

OTHER RESOURCES

Further resources

The Data Education in Schools project at the University of Edinburgh have an online collection of resources for the NPA, sorted by course topic and including resources for professional learning and useful datasets (**dataed.in/NPA**). This will include all the resources in this guide and additional links that have been developed since publishing.

As the project develops more teaching resources tailored to the NPA, these will be available at **teachdata.science**.

Whole Course resources

There are many resources that are useful across several topics in the progression pathway. As mentioned previously, the Teach Computing course on data science with lesson plans, presentations, activities and video lessons for flipped or blended learning (**dataed.in/nccedata**). Runestone Academy have an online textbook on 'How to Think Like a Data Scientist' (**dataed.in/rune**). The European Journalism Centre have a clearly written 'Data Journalism Handbook' that has great case studies and examples of storytelling with data (**dataed.in/journalism**)

The Institute for Research in Schools (IRIS) have a regular series of projects for learners to work with authentic research (**dataed.in/iris**) which could involve working across different stages of data science, from gathering to analysing and visualising science data.

For professional learning, the MOOC online course on 'Getting started with teaching data science' (**dataed.in/teachds**) and the College Development Network course on data science (**dataed.in/CDN**) would both be useful and suitable to NPA educators.

Datasets

The online resources collection collated by the Data Education in Schools team featuring over a hundred different datasets or dataset collections on a variety of themes that will appeal and engage learners (**dataed.in/NPA**).

SUPPLEMENTARY MATERIALS

UNIT AND AWARD CODES

Group award codes

NPA Level 4: GP8N 44
NPA Level 5: GP8P 45
NPA Level 6: GP8R 46

Core Unit codes

Data Citizenship	Level 4: J2HN 44	Level 5: J2HN 45,	Level 6: J2HN 46
Data Science	Level 4: J2G2 44,	Level 5: J2G2 45,	Level 6: J2G2 46

Optional Unit codes

Computer Programming	Level 5: HY2C 45,	Level 6: HY2C 46
Data Science Project	Level 5: J2GT 45,	Level 6: J2GT 46
Data Science Statistics	Level 5: J2G8 45,	Level 6: J2G8 46
Data Security	Level 5: H9E2 45,	Level 6: H9E2 46
Machine Learning	Level 5: J2G6 45,	Level 6: J2G6 46
Statistics	Level 6: H95Y 46 (unit),	GK8Y 46 (Group Award)

NPA Data Science Level 4

- Data Citizenship Level 4
- Data Science Level 4

NPA Data Science Level 5

- Data Citizenship Level 5
- Data Science Level 5

and one of

- Data Security Level 5
- Computer Programming Level 5
- Machine Learning Level 5
- Data Science Project Level 5
- Data Science Statistics Level 5

NPA Data Science Level 6

- Data Citizenship Level 6
- Data Science Level 6

and one of

- Data Security Level 6
- Computer Programming Level 6
- Machine Learning Level 6
- Data Science Project Level 6
- Data Science Statistics Level 6
- Statistics Level 6

DATA CITIZENSHIP UNITS

OUTCOMES AND PERFORMANCE CRITERIA

Level 4

DC4.1 State the use of data in society

DC4.1a State the reasons for the growth of data.
DC4.1b State how data is used and misused by individuals, organisations and society.
DC4.1c Describe types of data bias and its impact on society.
DC4.1d State common sources of public and private data.
DC4.1e State the rights and responsibilities of data subjects and data owners.

DC4.2 Describe data literacy concepts

DC4.2a Describe the characteristics of high quality data.
DC4.2b Describe how data can be analysed.
DC4.2c Describe types of data visualisations.
DC4.2d Describe simple methods of managing and securing data.

DC4.3 Interpret simple data

DC4.3a Create visualisation to identify patterns and trends in the data.
DC4.3b Draw conclusions from data.
DC4.3c Make recommendations based on conclusions and communicate findings.

Level 5

DC5.1 Describe the use of data in society

DC5.1a Describe the reasons for the growth of data.
DC5.1b Describe how data is used and misused by individuals, organisations and society.
DC5.1c Describe types of bias and its impact on individuals and society.
DC5.1d Describe sources of public and private data and the concept of open data.
DC5.1e Describe the rights and responsibilities of data subjects and data owners.

DC5.2 Explain data literacy concepts

DC5.2a Explain the characteristics of high quality data.
DC5.2b Explain how data can be analysed.
DC5.2c Explain types of data visualisations and the best use of each type.
DC5.2d Explain methods of managing and securing data.

DC5.3 Interpret data

DC5.3a Extract information from data visualisations and dashboards.
DC5.3b Interpret data to identify patterns and trends and draw conclusions.
DC5.3c Create appropriate visualisations from data.
DC5.3d Communicate findings and make recommendations based on conclusions.

Level 6

DC6.1 Explain the use of data in society

DC6.1a Explain the technological, economic and societal reasons for the growth of data.
DC6.1b Explain how data is used and misused by individuals, organisations and society.
DC6.1c Explain types of bias and its impact on individuals and society.
DC6.1d Explain types and sources of large datasets and the philosophy of open data.
DC6.1e Explain the rights and responsibilities of data subjects and data owners.

DC6.2 Explain data literacy concepts

DC6.2a Explain the concepts of data volume, variety, velocity, veracity and value.
DC6.2b Explain how data can be analysed and the tools that can be used to perform analysis.
DC6.2c Explain data visualisations and data storytelling.
DC6.2d Explain methods of data management and data security.
DC6.2e Explain the role of domain knowledge within data science.
DC6.2f Explain the concept of data ethics.

DC6.3 Interpret complex data

DC6.3a Extract information from data visualisations and dashboards.
DC6.3b Evaluate a dataset in terms of its quality including potential bias.
DC6.3c Interpret data to identify patterns and trends and draw conclusions.
DC6.3d Create appropriate visualisations from data.
DC6.3e Communicate findings and make recommendations based on conclusions.

DATA SCIENCE UNITS

OUTCOMES AND PERFORMANCE CRITERIA

Level 4

DS4.1 Describe data science

DS4.1a State the reasons for the development of data science.
DS4.1b Describe contemporary applications of data science.
DS4.1c Describe the steps in solving a problem using data science.
DS4.1d Identify sources of bias in data science including historical bias.

DS4.2 Describe simple ways of analysing data

DS4.2a Describe common data types and data formats.
DS4.2b Describe structured and unstructured data.
DS4.2c Describe simple methods of cleaning and transforming data.
DS4.2d Describe basic descriptive statistics used to summarise a dataset.
DS4.2e Describe simple data visualisations.

3: Analyse a small dataset to identify patterns (DS4.3)

DS4.3a Perform simple data cleaning and structuring.
DS4.3b Perform basic analyses including sort, filter, group and summarise.
DS4.3c Visualise the data to provide basic insights.
DS4.3d Create a simple report to communicate insights.

Level 5

DS5.1 Describe the tools and techniques of data science

DS5.1a Describe the reasons for the development and growth of data science.
DS5.1b Describe contemporary applications of data science.
DS5.1c Describe the data science life cycle including the potential for bias at each stage.
DS5.1d Describe the tools that can be used at each stage in the life cycle.
DS5.1e Identify sources of public and private datasets.
DS5.1f Describe the role of domain knowledge and subject matter experts in data science.

DS5.2 Describe methods of routine data analysis

DS5.2a Describe common data types and data formats.
DS5.2b Describe the composition of a structured dataset.
DS5.2c Describe methods of cleaning and transforming data.
DS5.2d Describe methods of securing and managing data.
DS5.2e Describe descriptive statistics used to summarise a dataset including measures of central tendency and dispersion.
DS5.2f Describe the selection of data visualisations to illustrate different types of data.

DS5.3 Analyse a dataset to identify patterns and trends

DS5.3a Define the required analyses.
DS5.3b Capture data from an external source.
DS5.3c Perform routine data cleaning and structuring.
DS5.3d Perform analyses including query, sort, filter, consolidate, group and summarise.
DS5.3e Visualise the data to provide insights.
DS5.3f Create an interactive data dashboard to identify patterns and trends.

Level 6

DS6.1 Explain the principles of data science

DS6.1a Explain the relationship between artificial intelligence, machine learning, big data and data science.
DS6.1b Explain the technological, economic and societal reasons for the development and growth of data science.
DS6.1c Describe contemporary applications of data science and the types of problem that data science can address.
DS6.1d Explain the data science life cycle and the significance of domain expertise.
DS6.1e Explain descriptive analytics and predictive analytics.
DS6.1f Explain the principle of open data and sources of open data.
DS6.1g Explain data ethics, including data bias, with reference to national and international standards and frameworks.

DS6.2 Explain data science techniques

DS6.2a Describe common data types and data formats including structured and unstructured data.
DS6.2b Explain techniques for data capture, cleaning and transformation including data modelling.
DS6.2c Explain data management and data security techniques.
DS6.2d Explain statistical techniques involved in data science.
DS6.2e Explain techniques for data visualisation, data dashboards and data storytelling.

DS6.3 Analyse a dataset to make predictions

DS6.3a Define the required analyses and data models.
DS6.3b Create a relational data model from external sources of data.
DS6.3c Perform data transformation to complete, correct and structure data.
DS6.3d Perform descriptive and predictive analyses on the data.
DS6.3e Create data visualisations and data dashboards to provide insights.
DS6.3f Identify potential sources of bias in the analysis

DATA CITIZENSHIP UNITS

Performance Criteria to Progression Pathway

	1 What is Data?	2 Interpreting Data	3 What is Data Science?	4 Working with Data	5 Security	6 Privacy	7 Capturing Data	8 Data Manipulation	9 Statistics	10 Analysis	11 Visualisation and Storytelling	12 Quality and Management	13 Ethics and Bias	14 Tools and Languages
DC4.1a	█													
DC4.1b													█	
DC4.1c													█	
DC4.1d	█													
DC4.1e						█						█		
DC4.2a												█		
DC4.2b										█				
DC4.2c		█												
DC4.2d					█							█		
DC4.3a											█			
DC4.3b											█			
DC4.3c											█			
DC5.1a	█													
DC5.1b													█	
DC5.1c													█	
DC5.1d	█													
DC5.1e						█						█		
DC5.2a														
DC5.2b										█				
DC5.2c		█												
DC5.2d					█							█		
DC5.3a		█									█			
DC5.3b											█			
DC5.3c											█			
DC5.3d											█			
DC6.1a	█													
DC6.1b													█	
DC6.1c													█	
DC6.1d	█													
DC6.1e						█						█		
DC6.2a	█													
DC6.2b										█				█
DC6.2c												█		
DC6.2d					█							█		
DC6.2e			█											
DC6.2f													█	
DC6.3a		█									█			
DC6.3b											█	█		
DC6.3c											█			
DC6.3d											█			
DC6.3e											█			

DATA SCIENCE UNITS

Performance Criteria to Progression Pathway

	1 What is Data?	2 Interpreting Data	3 What is Data Science?	4 Working with Data	5 Security	6 Privacy	7 Capturing Data	8 Data Manipulation	9 Statistics	10 Analysis	11 Visualisation and Storytelling	12 Quality and Management	13 Ethics and Bias	14 Tools and Languages
DS4.1a			X											
DS4.1b			X											
DS4.1c			X											
DS4.1d													X	
DS4.2a				X										
DS4.2b				X										
DS4.2c								X		X				
DS4.2d									X					
DS4.2e		X												
DS4.3a								X		X				
DS4.3b								X		X				
DS4.3c											X			
DS4.3d											X			
DS5.1a			X											
DS5.1b			X											
DS5.1c			X										X	
DS5.1d														X
DS5.1e	X													
DS5.1f			X											
DS5.2a				X										
DS5.2b				X										
DS5.2c								X		X				
DS5.2d					X							X		
DS5.2e									X					
DS5.2f											X			
DS5.3a										X				
DS5.3b							X							
DS5.3c								X		X				
DS5.3d								X		X				
DS5.3e											X			
DS5.3f											X			
DS6.1a			X											
DS6.1b			X											
DS6.1c			X											
DS6.1d			X											
DS6.1e			X											
DS6.1f	X													
DS6.1g													X	
DS6.2a				X										
DS6.2b				X			X	X		X				
DS6.2c					X							X		
DS6.2d									X					
DS6.2e											X			
DS6.3a				X						X				
DS6.3b				X										
DS6.3c								X						
DS6.3d										X				
DS6.3e											X			
DS6.3f													X	

DATA CITIZENSHIP AND DATA SCIENCE UNITS

Progression Pathway to Performance Criteria

1.   What Is Data?

DC4.1a	State the reasons for the growth of data
DC4.1d	State common sources of public and private data
DC5.1a	Describe the reasons for the growth of data
DC5.1d	Describe sources of public and private data and the concept of open data
DC6.1a	Explain the technological, economic and societal reasons for the growth of data
DC6.1d	Explain types and sources of large datasets and the philosophy of open data
DC6.2a	Explain the concepts of data volume, variety, velocity, veracity and value
DS5.1e	Identify sources of public and private datasets
DS6.1f	Explain the principle of open data and sources of open data

2.   Interpreting Data

DC4.2c	Describe types of data visualisations
DC5.2c	Explain types of data visualisations and the best use of each type
DC5.3a	Extract information from data visualisations and dashboards.
DC6.3a	Extract information from data visualisations and dashboards.
DS4.2e	Describe simple data visualisations

3.   What Is Data Science?

DC6.2e	Explain the role of domain knowledge within data science
DS4.1a	State the reasons for the development of data science
DS4.1b	Describe contemporary applications of data science
DS4.1c	Describe the steps in solving a problem using data science
DS5.1a	Describe the reasons for the development and growth of data science
DS5.1b	Describe contemporary application of data science
DS5.1c	Describe the data science life cycle including the potential for bias at each stage
DS5.1f	Describe the role of domain knowledge and subject matter experts in data science
DS6.1a	Explain the relationship between artificial intelligence, machine learning, big data and data science
DS6.1b	Explain the technological, economic and societal reasons for the development and growth of data science
DS6.1c	Describe contemporary applications of data science and the types of problem that data science can address
DS6.1d	Explain the data science life cycle and the significance of domain expertise
DS6.1e	Explain descriptive analytics and predictive analytics

4.   Working with Data

DS4.2a	Describe common data types and data formats
DS4.2b	Describe structured and unstructured data
DS5.2a	Describe common data types and data formats
DS5.2b	Describe the composition of a structured dataset
DS6.2a	Describe common data types and data formats including structured and unstructured data
DS6.2b	Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3a	Define the required analyses and data models.
DS6.3b	Create a relational data model from external sources of data.

5.   Security

DC4.2d	Describe simple methods of managing and securing data
DC5.2d	Explain methods of managing and securing data
DC6.2d	Explain methods of data management and data security
DS5.2d	Describe methods of securing and managing data
DS6.2c	Explain data management and data security techniques

6.   Privacy

DC4.1e	State the rights and responsibilities of data subjects and data owners
DC5.1e	Describe the rights and responsibilities of data subjects and data owners
DC6.1e	Explain the rights and responsibilities of data subjects and data owners

7.   Capturing Data

DS5.3b	Capture data from an external source.
DS6.2b	Explain techniques for data capture, cleaning and transformation including data modelling

8.   Data Manipulation

DS4.2c	Describe simple methods of cleaning and transforming data
DS4.3a	Perform simple data cleaning and structuring.
DS4.3b	Perform basic analyses including sort, filter, group and summarise.
DS5.2c	Describe methods of cleaning and transforming data
DS5.3c	Perform routine data cleaning and structuring.
DS5.3d	Perform analyses including query, sort, filter, consolidate, group and summarise.
DS6.2b	Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3c	Perform data transformation to complete, correct and structure data.

9. Statistics

DS4.2d	Describe basic descriptive statistics used to summarise a dataset
DS5.2e	Describe descriptive statistics used to summarise a dataset including measures of central tendency, dispersion and correlation
DS6.2d	Explain statistical techniques involved in data science.

10. Analysis

DC4.2b	Describe how data can be analysed
DC5.2b	Explain how data can be analysed
DC6.2b	Explain how data can be analysed and the tools that can be used to perform analysis
DS4.2c	Describe simple methods of cleaning and transforming data
DS4.3a	Perform simple data cleaning and structuring.
DS4.3b	Perform basic analyses including sort, filter, group and summarise.
DS5.2c	Describe methods of cleaning and transforming data
DS5.3a	Define the required analyses.
DS5.3c	Perform routine data cleaning and structuring.
DS5.3d	Perform analyses including query, sort, filter, consolidate, group and summarise.
DS6.2b	Explain techniques for data capture, cleaning and transformation including data modelling
DS6.3a	Define the required analyses and data models.
DS6.3d	Perform descriptive and predictive analyses on the data.

11. Visualisation and Storytelling

DC4.3a	Create visualisation to identify patterns and trends in the data.
DC4.3b	Draw conclusions from data.
DC4.3c	Make recommendations based on conclusions and communicate findings.
DC5.3a	Extract information from data visualisations and dashboards.
DC5.3b	Interpret data to identify patterns and trends and draw conclusions.
DC5.3c	Create appropriate visualisations from data.
DC5.3d	Communicate findings and make recommendations based on conclusions.
DC6.2c	Explain data visualisations and data storytelling
DC6.3a	Extract information from data visualisations and dashboards.
DC6.3c	Interpret data to identify patterns and trends and draw conclusions.
DC6.3d	Create appropriate visualisations from data.
DC6.3e	Communicate findings and make recommendations based on conclusions.
DS4.3c	Visualise the data to provide basic insights.
DS4.3d	Create a simple report to communicate insights.
DS5.2f	Describe the selection of data visualisations to illustate different types of data
DS5.3e	Visualise the data to provide insights.
DS5.3f	Create an interactive data dashboard to identify patterns and trends.
DS6.2e	Explain techniques for data visualisation, data dashboards and data storytelling
DS6.3e	Create data visualisations and data dashboards to provide insights.

12.   Quality and Management

DC4.2d	Describe simple methods of managing and securing data
DC5.1e	Describe the rights and responsibilities of data subjects and data owners
DC5.2a	Explain the characteristics of high quality data
DC5.2d	Explain methods of managing and securing data
DC6.1e	Explain the rights and responsibilities of data subjects and data owners
DC6.2d	Explain methods of data management and data security
DC6.3b	Evaluate a dataset in terms of its quality including potential bias.
DS5.2d	Describe methods of securing and managing data
DS6.2c	Explain data management and data security techniques

13.   Ethics and Bias

DC4.1b	State how data is used and misused by individuals, organisations and society
DC4.1c	State the types of data bias and its impact on society
DC5.1b	Describe how data is used and misused by individuals, organisations and society
DC5.1c	Describe types of data bias and its impact on individuals and society
DC6.1b	Explain how data is used and misused by individuals, organisations and society
DC6.1c	Explain types of bias and its impact on individuals and society
DC6.2f	Explain the concept of data ethics
DC6.3b	Evaluate a dataset in terms of its quality including potential bias.
DS4.1d	Identify sources of bias in data science including historical bias
DS5.1c	Describe the data science life cycle including the potential for bias at each stage
DS6.1g	Explain data ethics, including data bias, with reference to national and international standards and frameworks
DS6.3f	Identify potential sources of bias in the analysis

14.   Tools and Languages

DC6.2b	Explain how data can be analysed and the tools that can be used to perform analysis
DS5.1d	Describe the tools that can be used at each stage in the life cycle

NPA DATA SCIENCE LEVEL 4 ASSESSMENT RECORD

Candidate Name: ....

Candidate Number: .... **Class:** ....

DATA CITIZENSHIP UNIT
(Unit code: GP8N 44)

Completion date: __ / __ / _____

SOLAR computer-based test (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Date of second sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: __ %   Pass mark 60%

DATA SCIENCE UNIT
(Unit code: J2G2 44)

Completion date: __ / __ / _____

SOLAR computer-based test (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Date of second sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: __ %   Pass mark 60%

NPA DATA SCIENCE, LEVEL 4
(Group Award code: GP8N 44)

Completion date: __ / __ / _____

NPA DATA SCIENCE, LEVEL 5 ASSESSMENT RECORD

Candidate Name: ....

Candidate Number: .... **Class:** ....

DATA CITIZENSHIP UNIT
(Unit code: J2HN 45)

Completion date: __ / __ / _____

SOLAR computer-based test (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Date of second sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: ___ %   Pass mark 60%

DATA SCIENCE UNIT
(Unit code: J2G2 45)

Completion date: __ / __ / _____

SOLAR computer-based test (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Date of second sitting: __ / __ / _____   Result: __ / 20   Pass mark 12/20

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: ___ %   Pass mark 60%

OPTIONAL UNIT

Completion date: __ / __ / _____

- [ ] **Computer Programming** (HY2C 45)
- [ ] **Data Science Project** (J2GT 45)
- [ ] **Data Science Statistics** (J2G8 45)
- [ ] **Data Security** (H9E2 45)
- [ ] **Machine Learning** (J2G6 45)

Result: **Pass / Fail**

NPA DATA SCIENCE, LEVEL 5
(Group Award code: GP8P 45)

Completion date: __ / __ / _____

NPA DATA SCIENCE, LEVEL 6
ASSESSMENT RECORD

Candidate Name: ....

Candidate Number: .... **Class:** ....

DATA CITIZENSHIP UNIT
(Unit code: J2HN 46)

Completion date: __ / __ / _____

SOLAR extended response questions (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: ___ %   Pass mark 60%

Date of second sitting: __ / __ / _____   Result: ___ %   Pass mark 60%

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: ___ %   Pass mark 60%

DATA SCIENCE UNIT
(Unit code: J2G2 46)

Completion date: __ / __ / _____

SOLAR extended response questions (Outcomes 1 and 2)

Date of first sitting: __ / __ / _____   Result: ___ %   Pass mark 60%

Date of second sitting: __ / __ / _____   Result: ___ %   Pass mark 60%

Practical assessment (Outcome 3)

Completion date: __ / __ / _____   Result: ___ %   Pass mark 60%

OPTIONAL UNIT

Completion date: __ / __ / _____

☐ **Computer Programming** (HY2C 46)
☐ **Data Science Project** (J2GT 46)
☐ **Data Science Statistics** (J2G8 46)
☐ **Data Security** (H9E2 46)
☐ **Machine Learning** (J2G6 46)
☐ **Statistics** (H95Y 46)

Result: **Pass / Fail**

NPA DATA SCIENCE, LEVEL 6
(Group Award code: GP8R 46)

Completion date: __ / __ / _____

Data Education
in Schools

SQA