WOMEN IN TECHNOLOGY –––––––––––––––––––––––ReGIONAL DIFFERENCES1

Origin of data, work methodology and visualizations

CS and Women in Scotland

Ester Giménez, 1 August 2021

Table of Contents

1. [**Introduction ..……………………………………………………………………………..… 2**](#_Toc65783358)
2. **Research of the websites……………………………………...………………………….. 3**
3. **Review the data sources the organization uses to gather the data …………..….. 3**
4. **Business intelligence and data driven decision making ..…………………..….….. 4**
5. **Domain knowledge and the business context …………...……………………….….. 4**
6. **Business processes involved and data flows ……………...………………….….….. 5**
7. **Review of data quality and data bias used in data decision making process ….. 6**
8. **Review of data tools ………………………………………………………………..……… 8**
9. **Dataset discusision - Cleaning…………………………………………………………… 9**
10. **Descriptive, diagnostic, predictive and prescriptive analysis…………………….. 11**
11. **Data management including security…………………………………..……………… 11**
12. **Ethical implications of business requirements………………………………………. 11**
13. **Document the results of the analysis………………………………………………….. 12**
14. **Document business intelligence gained………………………………………………. 14**
15. **Document additional research……….………………………………………………….. 15**
16. **References…………………………………………………………………………………... 16**

1. Introduction

Code Division, a charity dedicated to provide high quality computer science courses to people with no previous access to this subject, has requested some information on schools that offer CS in Scotland. The students of Code Division have the opportunity to learn computing, programming and a set of skills that allow them to begin a career in technology. The aim of the charity is to promote CS across the population of Scotland and to increase the number of underrepresented parts of the population in the CS working pool, like for example, women.

This necessity has been more noticeable during the pandemic provoked by the COVID-19 virus. Health boards needed people to collect data of the development of the infection, in order to make models and temporal projections for future care. The reduced number of professionals dedicated to data analysis was made apparent. At the same time, a high number of companies realized the necessity of collecting data for their own use and began looking for data analysts as well.

This report is part of the project “Women in Tech”, final project of the PDA in Data Analysis, Level 8 (first cohort, July 2021). This PDA, provided by Code Division, will help mitigate the urgent need for Data Analysts in Scotland.

The project has been divided in several parts, which will be taken by our team members. I will be the acting project manager/coordinator in this case. The divisions are:

* **University:** subjects teach at universities in Scotland in CS, comparisons of men/women that attend those courses, entry requirements, etc.
* **SQA board:** differences in qualifications between boys/girls across all levels of primary, secondary, vocational studies, etc.
* **Regional differences:** differences between schools across Scotland and which schools have CS as a subject. If possible, to compare boys/girls on those schools that take CS across the country.
* **Pipeline to tech:** comparison in the curriculums of CS between Scotland vs England and number of girls taking CS due to changes in curriculum. If possible, to compare them with other countries.
* **Role of charities / out of the pipeline organisations:** what is the impact of charities in CS in Scotland and specially in girls.
* **Role of colleges:** to add college numbers of CS subjects and girls studying CS across the country.
* **Further research ideas:** there is a lot to develop and study in this project. Our ideas will be added at the end of the final report.

The aim of this project is not to advise on procedures or to apply changes, but only to highlight points of interest that can be useful for future decision making. We leave Code Division the use of our conclusions at their lease.

My study will be a pre-study in “Regional Differences”, awaiting the reception of data regarding pupils that study CS in schools. This information was requested to the Scottish Government at the end of June 2021.

2. Research of the websites

The datasets that I will use were downloaded from the website of the Scottish Government, hence my review will be of this website alone.

My report also mentions SQA website and others. Those will be reviewed by other team members, because they will use them more than me.

**Values**

The website of the Education directorate, <https://www.gov.scot/about/how-government-is-run/directorates/learning/> lists what type of information is available (there is a list of policies in the right-hand side of the webpage). It does not need to state their “values” as such. Those are usually specified in each document and report.

**Purpose**

The purpose as per their webpage states that they “ensure success for Scotland's learners through an effective school system and wider learning environment”. This gives an idea of what the directorate is striving for.

**Their business**

The same webpage lists the items that the directorate is responsible for. This also included the issue of statistics related to education.

3. Review the data sources the organization uses to gather the data

The law requires all schools to be registered every year in a database. The schools have to inform if they are active or not, some descriptive statistics of their teachers and pupils, what subjects they teach, etc. The data goes through a validating software called ProcXed.net. This information is used to make reports and statistics and it is available for parents at the website <https://education.gov.scot/parentzone/find-a-school>.

The data is as good as the information given by the schools. Some schools do not have contact details, or do not provide the grade level of their pupils, just to mention a couple of examples. Some information can be collected manually, looking in internet or calling the schools. For the making of this report, schools will not be contacted.

4. Business intelligence and data-driven decision making

One of the objectives of Code Division, as stated in the Introduction, is to improve the reach of their courses. In the last years, there have been a number of articles and studies (please see the reference section), explaining the status of CS in schools and in the life of work. It has been highlighted, that women have a very reduced presence in CS in general. Code Division decided to help increase the number of women that go into CS, and in order to do this, they need to know where are the main gaps.

This report will help visualize differences in the number of girls vs number of boys in CS in public schools. The idea is to develop the following:

* + Distribution of pupils by postcode areas and by local authorities.
  + Distribution of gender of the pupils by postcode areas and local authorities.
  + How postcode areas relate to local authorities and vice versa.
  + Comparison by location of the school in Scotland (regions, cities, rural and urban).
  + Schools with/without computer science subjects (current and historical).
  + Historical development of CS in schools and regional differences between boys/girls.

The completion of the above points depends on the availability of data and time restrictions.

5. Domain knowledge and the business context

The datasets used in this report come from the Scottish Government website [https://www.gov.scot/publications/](https://www2.gov.scot/Topics/Statistics/Browse/School-Education/Datasets/contactdetails) .

The Scottish Government receives information directly from schools, colleges and universities in Scotland on a yearly basis: contact details of the premises, demographic data of teachers and pupils, information about subjects given, etc. These data help the Government organize the education needs across the country, in terms of budget, moving pupils across schools, changes in the curriculum, etc.

The Government also requests specific data collections in specific areas. This is why in the lasts years it was detected an increase in external workforce being hired by companies (workers that come from England or other countries, instead of Scottish citizens). Some surveys were able to highlight an important decrease in pupils taking CS in schools in Scotland, which lead to a lack of adult professionals and the companies not being able to cope (private and public).

Furthermore, it was assessed that very few girls take this career path in comparison with boys. The Government believes that having a more balanced genre numbers in CS will be an important advantage for Scotland. On the other hand, having more young people in CS will benefit the future of computer science, because the country will be less dependent on external resources.

6. Business processes involved and data flows

The Government publishes some datasets in their webpage of the Education Directorate. The data comes from the centres, who have to fill in annual register forms.

The number of public schools (primary, secondary and special) can vary from year to year from 2400 to 2600, depending on closings, openings or joining of centres. In 2020 there were 357 secondary schools.

Independent schools can be members of the SCIS. There are currently 71 schools in this association. Their website publishes the list of its members, <https://www.scis.org.uk/find-a-school/>). Another register of independent schools can be found at https://www.gov.scot/publications/independent-schools-in-scotland-register/. In total, there are 102 independent schools in Scotland.

* + I compiled manually the **independent schools** that are members of the SCIS. The website does not offer this type of list and clients need to browse for every single school, in order to obtain the data.
  + This list is not complete, because every school needs to be contacted to obtain information about their CS courses and how many girls attend them. This contact list can be useful for future studies. The compiled file is called “8—Sec-Ind-Schools.xlsx”.

For a **comprehensive list of all schools** in Scotland (primary, secondary, special) is better to download the dataset of the Government website, https://www.gov.scot/publications/pupil-census-supplementary-statistics/.

**Colleges:** there are 27, listed in the website [www.collegesscotland.ac.uk](http://www.collegesscotland.ac.uk). The information provided does not specify CS courses, or if they have teachers, or how many pupils or girls take CS. For this, every college needs to be consulted separately. Another member of our team is investigating this.

My main datasets would be downloaded from the **Education Directorate:**

* + School level summary statistics and summary statistics schools bulleting tables
  + School contacts list

Apart from this, I need the geographical data of the local authorities in Scotland (pub\_las.shp of the [*https://data.spatialhub.scot/dataset/local\_authority\_boundaries-is*](https://data.spatialhub.scot/dataset/local_authority_boundaries-is)).

7. Review of data quality and data bias used in data decision making process

**DATA QUALITY**

After giving a look at the datasets, I believe that, in general, all datasets are of high quality. The Government provides corrections of data when needed and gives explanations and rationale for the changes. This is very welcome and gives credibility to the data.

The datasets provided by the government contain a lot of discrepancies in the addresses and locations of the schools. Some of them lack the contact data or the address, or they have been assigned to the wrong geographical area. For example:

* The field “Address line3” is for the town/city. It contained lots of blank cells. I corrected as many as I was able to, manually on the dataset.
* The phone number / website address was unknown or TBC in some cases: just for completeness, I looked those schools in Google and completed the information manually on the dataset.
* Roll by SIMD (Scottish Index of Multiple Deprivation): there are 6 columns dedicated to this index, however only 14 schools are classified, being the rest of the fields, ceros. This important lack of data means that the columns cannot be used, unless the data is found and imported from a different source.
* Urban/rural classification by the pupil's home address:
* This could be important to add in a map of schools, as it would show distances that the pupils need to travel each day to attend classes. When travel was not possible, it would be highlighted. Unfortunately, there is not enough information in the database or in the website of the government to know where the pupil lives and which school attends.
* It has to be taken into account, that a pupil is not able to choose school (unless it has enough economical freedom to attend an independent school outside the residence postcode). By law, the pupil has to attend the school of the area of residence. This can be a source of bias, since independent schools are not included in the government datasets and maybe, those schools are providing more CS than public schools.
* Currently, there is no availability of boundary geographical data for urban/rural classifications in Scotland, that can be used with Geopandas. This means, that postcodes have to be divided in 2 maps, one for rural and another for urban, if the postcodes of the pupils were known.
* 9 rows show as unclassified: all belong to the Grant Aided authority.
  + Grant Aided is not a real Local Authority. Those are special schools that receive grants from the government and they are usually boarding type schools.
  + 7 rows are special schools located in the Central Belt and South Lanarkshire.
  + 2 rows belong to Jordanhill School in Glasgow, one row for the primary school and another for the secondary school. This is correct, because there is a column in the dataset for Primary / Secondary education, which only accepts yes/no and if the school offers both levels, there should be 2 rows. For this case, we could have a mistake in the Primary School pupil data, because the two rows show numbers and also different ones.

**DATA BIAS**

Regarding **data bias**, apart from the discussion on urban/rural classification above, it is worth discussing the availability of data for this project. Information can be obtained from the Education Directorate of the Scottish Government website and from the SQA website. The issue is that it is not possible to connect all the data. For example:

* + If there is information about school’ s location and pupils’ distribution by gender/age, the subjects taught are not included (no information about which schools offer CS or which subjects).
  + If there is information about subjects and pupils by gender/age, the schools are not included. The age can be obtained by the certification year they are in, using the SQA datasets for every SQCF qualification and every year.
  + School information compared with subject taught is lacking. Taking into account that some schools, mostly secondary schools and independent ones, teach the English system or even external systems not related to the Scottish SQCF, the information about which subjects they teach, especially CS, becomes more important.

I would not like to say that there is a bias here, but certainly there is some information that has not been compiled properly or that it is not showing in the Government database and/or the websites of the colleges, independent schools, etc. That would be, which schools offer CS (now and in the past) and how many boys/girls attended those courses and for which grades.

After requesting this information for the public schools to the Government, it has not been provided. They argue that the information is confidential and perhaps can be available by local authority instead of by school. That could be useful, however it would hinder the deployment of CS actions in specific schools. On the other hand, the reputation of the schools is guaranteed, because nobody knows how many pupils pass the grades in each school and thus, the centres cannot be qualified as good or bad. This again, means losing opportunities, as it is not known which schools would benefit from CS actions.

8. Review of Data Tools

In order to manipulate this dataset, there are several tools that I used:

* **Excel:**
* The dataset was composed of less than 2500 rows and around 64 columns, far less than the limits of Excel.
* I used it for completing the summary of schools with data from the schools list. Through vlookups and coping/pasting the results, I obtained the final dataset, called

0—School+level+summary+statistics+2020-compiled – values.xlsx.

* The final excel file of values was later manually filled in with data of some locations, postcodes, etc, just for completeness and to be used for the mapping of schools. This is easier done in the dataset in excel than in any other tool.
* **Python (Jupiter Notebooks and Pandas, Folium, Geopandas libraries):**
  + For mapping and graphing purposes, I chose Jupyter Notebooks, because there is a library called “Geopandas” that allows to color geographical areas in maps of Scotland. I believed that this way of working is very visual.
  + Jupiter has been very useful for showing some visual graphs of demographic comparisons of pupils versus postcode areas and its differences with local authorities.
  + There is another library, called “Folium” that can be used in the future to map schools together with demographic data or CS data. I believe that even if the use of Folium seems attractive, the demographic data has been sufficiently graphed with Geopandas.

**9. Dataset discussion - Cleaning**

**a) School Level Summary Statistics 2020.xlsx**

This file contains information about pupils attending each level in every school in Scotland.

The information is of demographic type and it does not have any insights into subjects taught.

Column D, School Type, has some discrepancies, as some schools offer primary and secondary tuition, however this is not showing in this dataset. This can be caused by the absence of the independent schools, which offer both, or simply a simplification in the amount of information offered. The number of schools listed is 2477.

**b) Schools contacts list.xlsx**

This file does not contain independent schools. The number of schools listed is 2467.

My idea was to join both files, however, due to the difference in the number of lines showing in both files, I decided to make a copy of the addresses into the file of the summary statistics using the Excel vlookup function.

Firstly, I moved the column of the Seed Code in the Summary Statistics.2020 file from column C to A (the same as the contact list file). I moved the comments written at the end of the table in the Summary Statistics file to the sheet of the Background notes.

Secondly, I inserted the necessary columns in the file of the Summary Statistics after the School name column.

Third, I used a vlookup function to look for the values of the same Seed Code. I tried first with the name of the school, but it does not work. Then, I copied the address in the file of Summary Statistics:

**=VLOOKUP(A3,'[2--School+Contacts+List+April+2021.xlsx]Open Schools'!$A$1:$S$2468,5,FALSE)**

Column 5 is the first address line in the file of contacts list. This formula can be used with the rest of the columns, modifying the number 5 for 6, 7, 8… until 15.

Checking upon the Sort function for this new created column with addresses, there are 3 cells with =N/A. Their seed codes are:

8244022

8240922

8108625

These 3 schools lacked contact details. In order to find them, I looked for the name of the school in internet. For the UPRN numbers, I went to <https://uprn.uk> and used the postcode. Barshare Primary School has 2 indistinguishable numbers, while the others were easily found. The email addresses were not found and in one case, not even the website. I left them highlighted in blue, just in case in the future I needed further working on them.

In total I obtained data on 2477 schools in Scotland, without independent schools and colleges.

**c) Denominations of the schools**

If this database is used in the future for further comparisons that are out of scope of this project, I verified the denominations of the schools used in the file of “Schools Contact List” and the “Summary Statistics 2020”. There are discrepancies, highlighted by a comparison made in column BL. The “false” rows are in pink. I believe that the best column to use for denominations should be the one in the contact list file, added in column R.

**d) Computer subjects in the datasets**

None of the datasets that I checked in the website of the Scottish Government contain data on the subjects taught in schools. There is a pool of information in teachers and pupils in other datasets of the Government, but the numbers are added for the whole country and not specified by school. On the other hand, the SQA website has data on the subjects taught, but no information about regions, areas or schools.

In order to find out how many pupils studied CS, I consulted the website of the SQA, <https://www.sqa.org.uk/sqa/91419.html>. The information for the academic year 2020/2021 was posted the first week of July 2021. All data are divided by the type of certification and/or level obtained as per the Curriculum for Excellence. Due to this, the data for only CS / Information Technology and similar subjects, needs to be manually extracted from each one of those files. There are 38 files for each year. The tables needed are Table 3 and Table 4 of each file. The numbers are in percentage, so they need to be transformed back into totals, in order to represent them (percentages can be the same for different initial amounts. For example, 5000 students one year can yield 50% women and 50% men, and the next year, 500 students can have the same percentage, however the initial data shows a big decrease from 5000 to 500). The totals for Scotland can be extracted from the Scottish Government website “Pupils in Scotland 2020” census. This information, however, does not tell us anything about regional differences and would not be used in this section of the project.

**e) STEM website datasets**

Lastly, the STEM Strategy for Education and Training in Scotland reports (last one issued in March 2020), available from the [www.gov.scot](http://www.gov.scot) website, shows comparisons by gender and apprenticeships, festival attendance, science centres support, etc. KPI IIIa shows a gender comparison, however it has no relation with regions or subjects taught. KPI IIIb shows a gender and subject comparison, but no regions.

The “KPI Ic by LA” shows which regions in Scotland offer IT as IT hardware system support or software development within the STEM Foundation Apprenticeship framework. “KPI Ic” gives the total number of pupils that completed the apprenticeship in the last 3 cohorts by subject. Joining the data of both sheets would be very challenging and it would be better to ask the Government for the proper datasets if we are going to show comparisons on this specific apprenticeship.

I believe this is out of scope for this project, as it does not relate to schools, so this report will not be included in this study.

Other cleaning issues have been mentioned in the quality section.

**10. Descriptive, diagnostic, predictive and prescriptive analysis**

With the data available, it is possible to provide a descriptive analysis of the school’s dataset (median, mode, mean, standard deviation, etc) and gender distribution by location (local authority and postcode area). However, diversity is too big (there are schools with 2 pupils and others with more than 2000) and is not comparable with the geographical location of the schools.

A map can be drawn of the geographic distribution of schools in Scotland, however it is more useful to include a maps of pupil distribution by local authorities, as those are the ones that the pupils are divided into and also shows how the schools are assigned throughout the country.

The results and the maps can be used in the future to add data of CS per school.

**11. Data management including security**

All data will be treated as per the GDPR conditions in the respective websites.

The final report and the datasets used will be forwarded to Code Division for further consideration.

**12. Ethical implications of business requirements**

Pupil’s individual data is confidential and does not show in the datasets of the government. The official datasets include general demographic information.

The school information is public up to a certain degree, because some schools do not provide information on their pupils or teachers. This can be a drawback when making comparisons. On the other hand, we are more interested in knowing which schools offer CS, in which stages is it offered and how many girls attend, which is not provided. There is a possibility that if the information is made public, some schools will be qualified less than others. This can be a disadvantage, because charities would have to obtain this information directly from the schools, and there are more than 2000 in the country.

13. Document the results of the analysis

The tabs/graphs shown in the Jupiter notebook are the following:

There are 2476 rows in the dataset, which would imply as many schools, however there are only 2304 unique postcodes distributed in 980 locations (towns/ cities). In fact, there are 2304 schools in Scotland, with 2476 – 2304 = 172 schools with primary and secondary stages (that form duplicated rows in the database).

**13.1 STATISTICS**

* The minimum number of pupils/teachers is cero: probably, some schools do not want to disclose their data.
* The biggest academic centre (the maximum) has 2196 pupils, 1064 females 1132 males.
* The means of girls and boys is biased towards boys. Comparing this demographic data with Europe, the mean is usually directed towards girls (51% women, 49% men).
* There are 33 local authorities, including the exception of "Grant aided", that is not really a local authority. Grant Aided includes 7 schools, one with primary and secondary stages, located throughout the Central Belt of Scotland.

**13.2 RELATION BETWEEN LOCAL AUTHORITIES AND POSTCODE AREAS**

The geographical distribution of the postcode areas of the schools and the education local authorities do not match exactly. Some local authorities deal with several postcode areas. For example:

* **Argyll & Bute**: G - Glasgow, PA - Paisley
* **Highland**: IV - Inverness, PA - Paisley, KW - Kirkwall, PH - Perth
* **North Lanarkshire**: G - Glasgow, ML - Motherwell
* **Perth and Kinross**: DD - Dundee, PH - Perth, KY - Kirkaldy
* **Scottish Borders**: TD - Galashiels, EH - Edinburgh
* **South Lanarkshire**: G - Glasgow, ML - Motherwell
* **West Lothian**: ML - Motherwell, EH - Edinburgh
* **Stirling**: FK - Falkirk, G – Glasgow
* **North Ayrshire**: PA – Paisley, KA - Kilmarnock

On the other way round, the postcodes can be present at more than one local authority. For example, postcodes with GL - Glasgow appear in 9 local authorities, from Glasgow City to Lanarkshire, Dunbartonshire, Renfrewshire, Argyll, etc.

With the number of postcodes by local authority, we have from maximum to minimum the first 4 positions:

* Glasgow City: almost 200 postcodes
* Aberdeenshire (AB)
* Fife (KY)
* Highland: it deals with a high number of postcodes because it covers more than 1/3 of the surface of the country, from Paisley to Kirkwall. Highland should be an exception, as it does not have that many pupils as other local authorities (approximately 30,000) spread through the country.

The postcode concentration by local authority is focused in the Central Belt of Scotland and Aberdeen.

**13.3 RELATION BETWEEN POSTCODE AREAS AND PUPIL ROLLS**

The majority of schools are concentrated in 6 main postcode areas. Ordered from big to small are:

* Glasgow - GL
* Edinburgh - EH
* Aberdeen - AB
* Motherwell - ML
* Kirkaldy - KY
* Kilmarnock - KA

**13.4 RELATION BETWEEN LOCAL AUTHORITIES AND PUPIL ROLLS**

The 6 main local authorities, ordered from big to small are:

* Glasgow
* City of Edinburgh
* Fife
* North Lanarkshire: it includes some PA and G postcodes, which increases the number of pupils.
* South Lanarkshire
* Aberdeenshire: its position is not that high, because schools have been divided between Aberdeen City local authority and Aberdeenshire local authority.

**13.5 RELATION BETWEEN LOCAL AUTHORITIES / POSTCODE AREAS AND PUPIL GENDER DISTRIBUTION**

In general, there are more boys than girls in all postcodes, except in the islands (Lewis, Orkney, Shetlands, etc), where the proportion seems to be closer to 50:50.

14. Document business intelligence gained

Given the results of the graphs, I believe that:

As per the education law, pupils are asked to go to schools in their residence areas. If the families have enough means, children are sent to private schools, that can be located outside the residence area.

* Local Authorities are the education councils. They have splitted the postcode areas of the country in a way that some massified areas become less pupils per class. For example:
  + Pupils from Glasgow go to schools of local authorities in: Glasgow, Paisley, North/South Lanarkshire and Kilmarnock.
  + Pupils from Edinburgh go to schools in: Midlothian, Falkirk and Kilkenny (Fife).
* This explains why there is a high population of pupils in certain postcode areas (cities like Glasgow, Edinburgh, Aberdeen) that matches with more than one local authority, as the pupils attend schools in councils around the cities. Also, the schools are easier to manage if there are not concentrated in a few local authorities but divided between several of them.
* In general, there are more boys than girls attending school. The difference is not big, however it provides more boys for CS. In this way, if CS subjects have been on the increase in the last years, it could be because there are more boys in the general population, and not that more girls are motivated to take CS subjects.
  + This difference seems not to exist in the islands (Lewis, Orkney, Shetlands, etc).

If we want other comparisons, we need to know which schools offer CS subjects. At this moment, demographic information is the only output available.

The maps and the instructions provided in this Jupyter notebook can be used when the data is received.

15. Document additional research

My initial idea was to plot a map with all the schools’ locations that taught CS subjects. This data has not arrived, but in the meantime, I have been investigating how to do this. The summary statistics of schools downloaded from the Scottish Government website contains the address of all public schools in Scotland. The independent school’s addresses can be obtained from their website, as mentioned earlier.

These addresses need to be transformed into geographical coordinates. Those have to be obtained for every school, maybe through a GPS provider (geolocator service). There is a library in Python that does the requests to obtain the geolocators, through Google, Bing or others. The service is paid if the manual searches go over 40,000 a month. There are other options, like pinpoint the schools in Google maps or Bing maps and take note of the coordinates. The time necessary to collate the information would be significative.

Due to time limitations, this is out of scope. In the future it could be useful to plot this type of map, as it would give an idea of where are the schools located and which areas of the country are more covered with CS than others.

Some other ideas for further research:

* **Schools that offer CS subjects in Scotland.** Examples of ideas that can be researched:
  + To map some comparisons of the number of boys and girls that were studying CS at public schools in 2020. Also, to add some graphs of pupil gender distribution in CS by local authority.
  + To identify local authorities with less CS, that would need more actions to increase the number of girls going to CS.
  + Comparisons of gender distribution in the historic data, since 2010 (if available) with schools that offer CS and divided by local authority.
  + If the historic data becomes available, it would be possible to make future projections on number of girls that will study CS subjects.
* **SIMD**:
* It is out of scope of this project, if we do not obtain information for all schools. This could be to manually relate the postcodes of all schools to a SIMD area.
* SIMD areas change every year.
* There is a Geopandas map of Scotland with SIMD areas in the Government website. For some unknown reason I am unable to download it.
* There is also a file called "SIMD+2020v2+-+datazone+lookup" that can be downloaded. This file is used to look for the SIMD area after manually entering a postcode. It can be used for the 2300 schools in Scotland.

* **Urban/rural distribution:**
* It will be out of scope for the moment. The dataset of the Government includes a column with a yes/no answer as to what the pupils attending a given school are rural or urban, however it is incomplete.
* A map of this distribution would be very welcomed to locate schools in rural areas, that would have more difficulties accessing CS.

**16.** **References**

Below there is a list of links, with information and datasets that were used or consulted in the elaboration of this report.

Scottish Index of Multiple Deprivation 2020

<https://www.gov.scot/collections/scottish-index-of-multiple-deprivation-2020/?utm_source=redirect&utm_medium=shorturl&utm_campaign=SIMD>

Consulted on 26/06/2021.

[Statistics 2019 - SQA](https://www.sqa.org.uk/sqa/91419.html)

<https://www.sqa.org.uk/sqa/91419.html>

Scottish Government publications and data consultation:

<https://education.gov.scot/parentzone/find-a-school>

<https://www.gov.scot/publications/foi-17-01802/>

<https://www.gov.uk/government/statistics/provisional-entries-for-gcse-as-and-a-level-summer-2021-exam-series/provisional-entries-for-gcse-as-and-a-level-summer-2021-exam-series>

SIMD webpages:

[www.gov.scot/SIMD](http://www.gov.scot/SIMD)

Statistics Scotland:

<http://statistics.gov.scot/>

STEM strategy for education and training: second annual report

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

All spatial data websites have been consulted in July 2021.

SpatialData

<https://www.spatialdata.gov.scot/geonetwork/srv/eng/catalog.search#/metadata/1cd57ea6-8d6e-412b-a9dd-d1c89a80ad62>

Updated April 2021

Free OS OpenData Map Downloads | Free Vector & Raster Map Data | OS Data Hub

<https://osdatahub.os.uk/downloads/open>

Community Council Boundaries

<https://data.spatialhub.scot/dataset/community_council_boundaries-is>

Newspaper articles

<https://www.thescottishsun.co.uk/news/4020266/scotland-secondary-schools-results-chart-worst-best-glasgow-edinburgh/>

[Computing at School](https://www.computingatschool.org.uk/about)

<https://www.computingatschool.org.uk/custom_pages/32-documents>

<https://learningandwork.org.uk/resources/research-and-reports/disconnected-exploring-the-digital-skills-gap/>

<https://blog.teachcomputing.org/digital-literacy-within-the-computing-curriculum/>

<https://www.gov.uk/government/publications/remote-education-research/remote-education-research>

<https://www.gov.scot/publications/scottish-technology-ecosystem-review/>

<https://royalsociety.org/-/media/policy/Publications/2019/21-08-19-policy-briefing-on-teachers-of-computing.pdf>

<https://www.bcs.org/media/2520/tracer-2018.pdf>

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

[https://community.computingatschool.org.uk/resources/155/single#](https://community.computingatschool.org.uk/resources/155/single)

[https://community.computingatschool.org.uk/resources/446/single#](https://community.computingatschool.org.uk/resources/446/single)

<https://www.informaticsforall.org/>

<https://info.hackerrank.com/rs/487-WAY-049/images/HackerRank_2019-2018_Developer-Skills-Report.pdf>

<https://info.hackerrank.com/rs/487-WAY-049/images/HackerRank_2019-2018_Women-In-Tech-Report.pdf>

<https://research.hackerrank.com/student-developer/2018>

<https://www.childrenscommissioner.gov.uk/wp-content/uploads/2018/01/Childrens-Commissioner-for-England-Life-in-Likes-3.pdf>