Local area profiles using charity data

There is often a wealth of information available that can be used to profile a local area. Data from the Census, for example, can be used to build up a picture of the demographics of an area - who lives there, what jobs do they do, what does health and wellbeing look like. One part of the local picture that is often harder to measure is what the local voluntary sector looks like - how many charities and other non-profit organisations work locally? What do they do? How many people are involved, whether as staff, volunteers, trustees or users of the services?

Local voluntary sector infrastructure bodies, like Councils for Voluntary Service, will often spot these data gaps and try to fill them. In some areas local commissioners will ask for this kind of research as part of funding agreements with local infrastructure. Often this will come in the form of a survey of their members, which will collect some basic data about the organisations (how big are they, where do they get their income from) alongside more opinion-based data about their work (is demand or need for their services increasing, what are the barriers they face).

As a data scientist I've spent over fifteen years working with administrative data from the Charity Commission and other sources, and I believe that these data sources can fill some of these data gaps without needing to survey organisations. These data sources can work in tandem with surveys to fill in all the data gaps, and can enhance the results of local surveys by providing context to their findings.

This guide aims to take you through the process of using administrative data about charities to analyse and understand the voluntary sector in a local area, and use that data to produce aggregate information. The guide is based on analysing data for one local authority area, though it could be adapted to look at other types of areas.

Using charity regulator data

The first thing to consider when using data from Charity regulators is the area you want to cover. For charities in the UK there are three main regulators to consider:

- Charity Commission for England and Wales (CCEW)
- Scottish Charity Regulator (OSCR)
- Charity Commission for Northern Ireland (CCNI)

Note that some charities may be registered with more than one of these regulators, depending on where they operate. Each regulator produces a different set of data, so what is available for analysis depends on where you are looking at. It's also important to remember that each regulator also has slightly different criteria for what types and sizes of organisations need to be registered as a charity.

This guide assumes that you are analysing data within England and Wales, so are mostly using CCEW data. The principles for the analysis will be the same in Scotland and Northern Ireland, but the exact data available will vary - generally there is more data available from CCEW than the other two regulators, so some of the detailed analysis that is possible in England and Wales can't be done elsewhere.

The guide also assumes you're using Microsoft Excel to carry out the data analysis. The CCEW data download operates at the limit of what is possible with Excel, and in some cases might go over the limits (for example some of the tables in the Charity Commission download exceed the 1,000,000 row limit in the latest versions of Excel - although we can use Excel Power Query to get around some of these limits). So for some use cases other software (for example holding the data in a separate database) might be preferable.

Basic - Using the Charity Commission search

The simplest way to use Charity Commission data is through the Commission's advanced search interface. This helps get around some of the issues caused by the size of the dataset by making sure you only download the data you need. However, there are a number of limitations to the search tool that it is important to be aware of.

The main way to search geographically through the register search is using the "Charities working locally in England and Wales" filters in the Advanced Search page. These allow you to select charities based on their area of operation, as they describe it to the charity commission. However, there are limitations to this field which means the search may not include all the organisations you expect it to. Charities select these areas manually when they do their annual return, but they are only allowed to select up to 10 areas they work in. If they work in more than 10 then they can choose the "Throughout England and Wales" or similar options.

There is also an interface through which you can download all the charities in a particular parliamentary constituency. This search is based on the registered postcode of each organisation.

Once you've created your search criteria you can then download a list of charities as an excel spreadsheet.

The main limitation of this method is that the download contains a limited number of fields. This means that lots of the fields needed for analysing different aspects of the charities are not available in the download. You will be able to produce a basic outline of the number of charities in an area, perhaps looking at different sizes available, but more detailed analysis, particularly on financial data, is not possible with this data.

Advanced - Using the Charity Commission data download

The Charity Commission data can be downloaded from the register website, at the following link: https://register-of-charities.charitycommission.gov.uk/register/full-register-download

There are 13 separate files available in either text or JSON format. Not all the files are relevant for the analysis described in this guide. Generally the text version is a better choice for data analysis in Excel.

The text files themselves are wrapped in ZIP files to make them smaller. Your operating system should be able to extract the text files from inside the ZIP, or you can use a program like 7zip for this.

Preparing the data

For the rest of this guide, I'm assuming that you've downloaded all the tables from the Charity Commission website, and extracted them all to a single folder called "ccew". You might have to move them from the default extracted location in order to do this. This will give you the following files:

- ccew/publicextract.charity.txt
- ccew/publicextract.charity_annual_return_history.txt
- ccew/publicextract.charity_annual_return_parta.txt
- ccew/publicextract.charity_annual_return_partb.txt
- ccew/publicextract.charity_area_of_operation.txt
- ccew/publicextract.charity_classification.txt
- ccew/publicextract.charity_event_history.txt
- ccew/publicextract.charity_governing_document.txt
- ccew/publicextract.charity_other_names.txt
- ccew/publicextract.charity_other_regulators.txt
- ccew/publicextract.charity_policy.txt
- ccew/publicextract.charity_published_report.txt
- ccew/publicextract.charity_trustee.txt

The most important table is the "charity" table, this contains the main record for each charity, along with some key details about them. An important part of understanding this table (which also applies to some of the other tables in the list) is to know the difference between the "registered_charity_number", "linked_charity_number" and "organisation_number" fields.

- **registered_charity_number**: This field is the registered charity number that is most commonly used by charities to display their charitable status publicly, and should be found on their website and in other places. This is generally the best way to distinguish between different charities.
- **linked_charity_number**: This number represents different subsidiary charities underneath the main registered charities. These subsidiaries are unlikely to represent

actual subsidiary activities, but instead represent administrative relics kept when, for example, a charity has absorbed another in a merger. The main charity is indicated with a “0” while the others have a number greater than 0. For almost all purposes only charities with a linked_charity_number of 0 should be used, and it is best practice to discard the rest.

- **organisation_number**: This is an internal Charity Commission identifier that allows for linking of charities between different tables. It is unique for every charity, even across the linked_charity_number. It’s important not to confuse this with the registered_charity_number as they often look similar.

Remember that you’ll want to only use charities where the linked_charity_number field is zero - the best course of action is to remove any charities with a linked_charity_number that isn’t zero from the data.

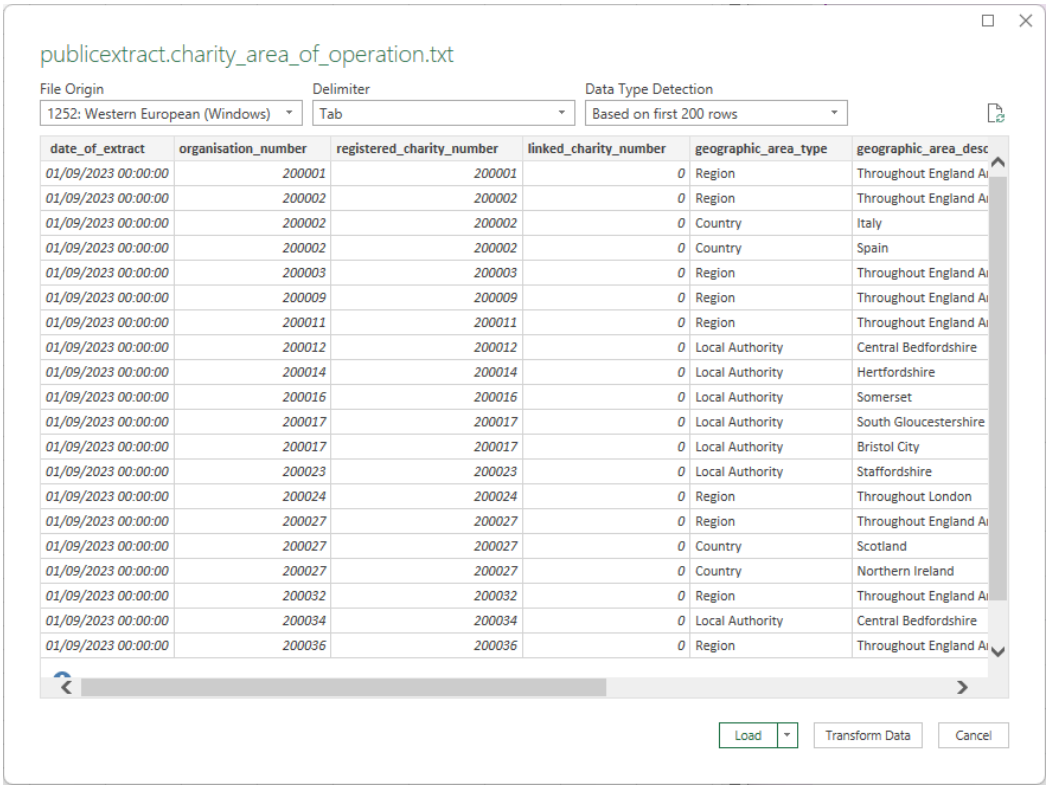
Using the data in Excel

The size of the data means that Excel is not the ideal solution for handling this data. If you are comfortable using other tools, you may find it easier to use a database or a scripted data analysis package (such as python or R) to do the extraction and analysis of the data.

If you are using Excel, then one way to increase the maximum number of rows you can fit in a sheet is through Excel Power Query. The workflow for this would look something like:

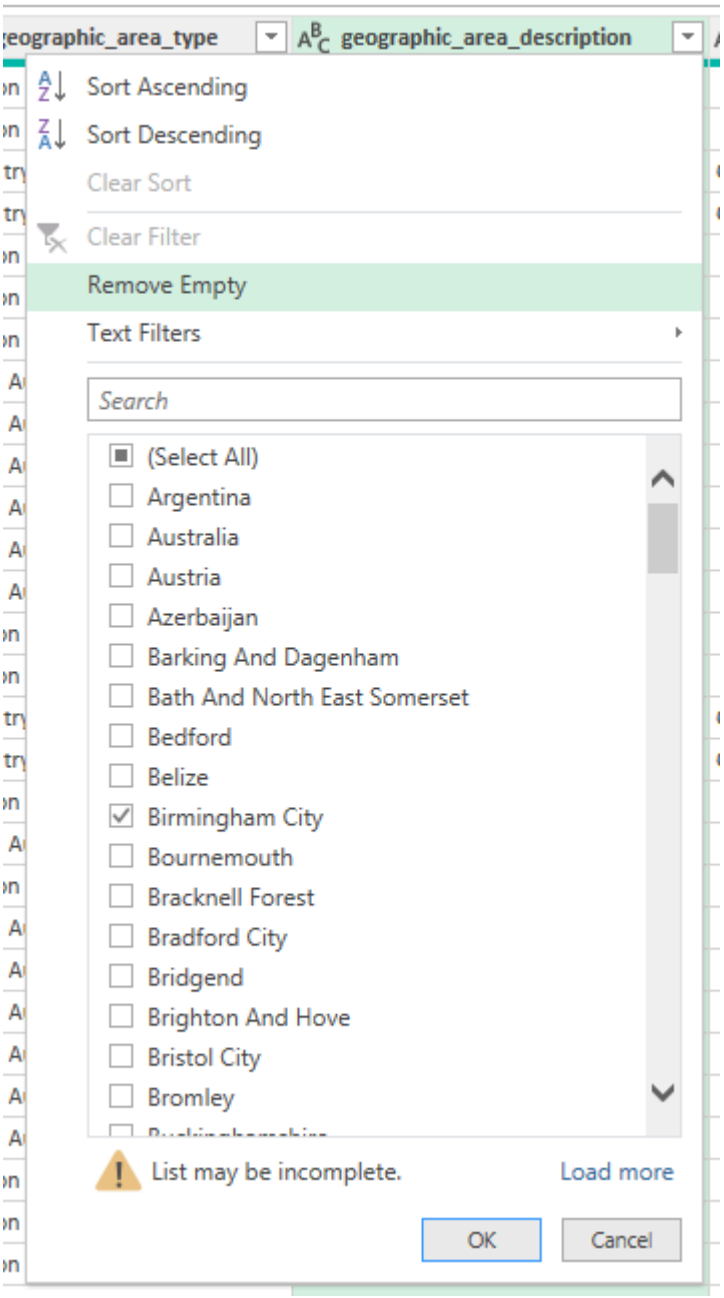
1. Extract the data downloaded out of the ZIP file and into a separate folder
2. In Excel, use “Data > Get & Transform data > From Text/CSV” to open the get data process
3. Select the extracted text file to import

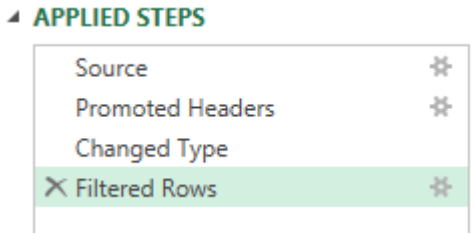
4. Accept the default options on the import screen and click “Transform” to open the Power Query editor.



date_of_extract	organisation_number	registered_charity_number	linked_charity_number	geographic_area_type	geographic_area_desc
01/09/2023 00:00:00	200001	200001	0	Region	Throughout England Au
01/09/2023 00:00:00	200002	200002	0	Region	Throughout England Au
01/09/2023 00:00:00	200002	200002	0	Country	Italy
01/09/2023 00:00:00	200002	200002	0	Country	Spain
01/09/2023 00:00:00	200003	200003	0	Region	Throughout England Au
01/09/2023 00:00:00	200009	200009	0	Region	Throughout England Au
01/09/2023 00:00:00	200011	200011	0	Region	Throughout England Au
01/09/2023 00:00:00	200012	200012	0	Local Authority	Central Bedfordshire
01/09/2023 00:00:00	200014	200014	0	Local Authority	Hertfordshire
01/09/2023 00:00:00	200016	200016	0	Local Authority	Somerset
01/09/2023 00:00:00	200017	200017	0	Local Authority	South Gloucestershire
01/09/2023 00:00:00	200017	200017	0	Local Authority	Bristol City
01/09/2023 00:00:00	200023	200023	0	Local Authority	Staffordshire
01/09/2023 00:00:00	200024	200024	0	Region	Throughout London
01/09/2023 00:00:00	200027	200027	0	Region	Throughout England Au
01/09/2023 00:00:00	200027	200027	0	Country	Scotland
01/09/2023 00:00:00	200027	200027	0	Country	Northern Ireland
01/09/2023 00:00:00	200032	200032	0	Region	Throughout England Au
01/09/2023 00:00:00	200034	200034	0	Local Authority	Central Bedfordshire
01/09/2023 00:00:00	200036	200036	0	Region	Throughout England Au

5. Filter the data using the dropdowns next to the column headings. For example, if you’re using the “charity_area_of_operation” file, select the areas you are interested in from the “geographic_area_description” field. Once you’ve made your selection, click OK. This will add the filter to the “Applied steps” list on the right hand side.





6. If the file has a “linked_organisation_number” field then make sure you add a filter at this point to only include those charities with “0” in this field.
7. Click “Close and Load” in the power query menu to return to your spreadsheet.

Once you’re comfortable with Power Query you can develop a workflow to put just the data you need into a single Excel workbook. The workflow may look like:

1. Create a list of charity numbers you want to include. You could do this through the area_of_operation file, or the main charity table or through using the Charity Commission basic search list.
2. Copy and paste this list of charity numbers into a separate file.
3. Begin the import of a CCEW file using the process described above. When you reach step 6, instead of filtering the data, click “New Source” to add the list of charity numbers you have saved.
4. In Power Query Editor, with the CCEW file selected, click “Merge Queries” to open the Merge process window.
5. Select your charity list (I’ve called it Sheet1) from the dropdown, and select both the registered_charity_number columns. Under “Join Kind”, select “Inner (only matching rows)”. This will ensure that the resulting data is only included if it’s in our charity number list. Then click Ok.
6. If the file has a “linked_charity_number” field then make sure only those with 0 are included.
7. Click “Close and Load” to add the data to the spreadsheet.
8. Once the “publicextract charity” (or other sheet name) has loaded, select the table and click “Query > Delete”. This will remove the query and leave the data as it is.
9. Repeat this process for each of the tables.

This should leave you with a copy of the CCEW database containing only the data relating to a particular area, which should be a more manageable size for Excel.

Choosing your population

The most crucial part of the analysis process is defining the population of organisations that you will be examining. There are a number of parts to this definition:

- Which charities operate/are registered in the area?
- Which types of charities are in scope?
- How do we account for charities that also operate outside the area?

Generally this process is done by starting with a large list of charities based on broad criteria, and then coming up with criteria to exclude particular charities.

Charities that operate or are registered in the area

The first step is to find charities that are connected with the local area. The CCEW data offers two main ways to do this: by finding those charities that say they operate in an area (known as the “area of operation”) and by looking at charities whose registered address is in an area.

The area of operation is stored in the “publicextract.charity_area_of_operation.txt” table. Charities can choose to represent the area they operate in in a number of ways:

- They can choose any number of overseas countries in which they operate. For this purpose Scotland and Northern Ireland are counted as overseas countries (these

rows have a “geographic_area_type” of “Country” in the Charity Commission area of operation file).
- They can choose up to ten “Local Authorities” that they operate in, within England and Wales. These are upper tier local authorities (i.e. counties and boroughs) so districts are not shown. The list is also not necessarily using the latest list of local authorities when there have been changes.
- There is also a selection of “regions” they can select. These are:
    - Throughout England and Wales
    - Throughout England
    - Throughout Wales
    - Throughout London

Filtering using this list will give you a list of charities that say they operate in your target Local Authority. However, there are a number of things to consider when using this data:

- In practice the data quality of this field can be poor, and not accurately map to where the charity works. Some charities may not know the name of the area they work in, or may tick a wider area than their actual area of operation.
- It contains no details about the scale of operation in each area. For example a charity might tick two adjoining local authorities, but actually it operates mostly in one of them with a small project in the other.
- It won’t capture charities that work across a wider area - for example national charities or those that operate at a regional level.

Charities whose registered address is in the area

The second method for selecting charities is to use the registered address of the charity. This is found in the main “charity” table, in fields starting with “charity_contact”. The easiest way to do this is to add local authority information to the “charity_contact_postcode” field and then filter based on the local authorities. You can also directly filter for particular postcodes if you wanted to look at, for example, everything with a “KT” postcode - although it’s important to remember that postcode areas don’t correspond to local authority boundaries.

There are a number of things to bear in mind when using this method:

- This introduces a headquarters effect, where we select charities based on their headquarters rather than where they actually work. This might mean that you include large national or international charities that happened to be registered to an address located in a local authority, when actually their work takes place elsewhere. It might also mean that you miss out on charities that work locally, but are registered to an address somewhere else. Bear in mind that registered contact addresses can include solicitors, accountants and other “virtual” addresses that may not be the charity’s main operating location.
- Postcodes aren’t available for removed charities, so your list won’t include charities who previously operated in the area but have since closed.

Removing National, International and Regional charities

Once you have a population of charities, you may want to remove national, international and/or regional charities. This is particularly pertinent when you have used the registered address to choose charities, as this takes no account of where the charity actually works. Many national or international charities have their headquarters in large cities around the UK - this is a particular problem in central London but it does apply across the UK.

There is no one definitive record of the scale a charity operates at (and the definition itself is fuzzy), but there are a number of places you can look to help decide:

- As described above, the “publicextract.charity_area_of_operation.txt” table allows a charity to select overseas countries or “Throughout England and Wales” as their area of operation. Applying some rules to this table can help work out the scale at which charities work.
- The “publicextract.charity_governing_document.txt” includes a “area_of_benefit” field. This is slightly different to the area of operation - it describes the area in which a charity can legally operate within its charitable objects. In some cases this provides useful information - it may give the name of a local parish or city - but it can also give a generic description.

Removing charities that aren’t relevant

As you look at the list of charities you’ve created, you will probably start thinking about which of them are actually relevant to your analysis. I would recommend sorting the list so the largest charities are first (there’s a field called latest_expenditure in the “publicextract.charity.txt” table that works well for this) - as larger charities are more likely to be outside of your scope, and also have a bigger impact on the final figures.

It’s important to know that the definition of what can register as a charity can often differ from the intuitive understanding of what a “charity” is. NCVO’s “general charities” definition provides a good guide to some of the decisions you need to think about when deciding what to include or exclude.

Exactly what you want to include or exclude will depend on your own definition and the circumstances in the area you’re looking at. But some of the types of organisations you might want to consider include:

- **Universities and colleges**. While most universities are exempt from charity registration, those in Wales and Oxford & Cambridge colleges are included, as well as some others. These are often large employers and centres of volunteer activity, but they can skew the figures for a local area so may want to be removed. Similarly, a local university Student Union may be registered as a charity.
- **Independent schools**. There is some controversy & debate about the charitable status of independent schools, but most are registered as charities.
- **Former Local Authority services**. Sometimes a local authority will have hived off its care or cultural services into a separate registered charity. You may decide that these are not independent from the local authority and should therefore be excluded.

- **National/International charities**. If you're looking through the list of the largest charities, you may also spot charities that operate nationally or internationally.

There are also groups of organisations that could be included, but you would only get a partial picture using Charity Commission data, as not all of these are registered charities. This may mean you want to exclude all of them, rather than have only part of the story. These include:

- **Housing Associations**. Whilst some of these, or their subsidiaries, are registered as charities, most are instead Registered Societies, registered on the Financial Services Authority Mutuals Register.
- **Places of Worship**. A large number of these are on the charity register, but there are also a large number that are excepted from registration.

Analysis

Once you have your population of organisations, ideally in a single list, there are a number of pieces of analysis you can do on these organisations. It's best to start with simple analysis that shows the number and size of organisations (by income or expenditure) compared by some grouping variables.

Generally the analysis that is performed falls into the following areas:

- **Size and scope of charities**. This might involve grouping charities by their income band (the standard groups used can be found in the NCVO Almanac). This is a good place to show the largest charities in an area.
- **Geography**. If your area covers a number of smaller areas (for example you might look at wards within a local authority) you can group by those wards. Remember that the postcode a charity is registered at may not reflect where it actually operates.
- **Classification**. The Charity Commission data contains a number of classification details in the "publicextract.charity_classification.txt" table. Charities can choose more than one of these categories. You can also use data from charityclassification.org.uk which contains a number of automatically-generated classifications
- **Changes over time**. If you're confident that you have good coverage of removed organisations in your area then you can look at how the number of organisations or their total income has changed over time. You may want to take into account how inflation has affected these figures.
- **People**. While the number of charity employees is only available for charities with more than £500,000 income, there is data available on the number of volunteers, plus it is possible to work out the number of trustees that charities have.
- **Income from government**. The "publicextract.charity_annual_return_parta.txt" contains details of income from government grants and contracts. This data is very useful, but it is worth applying some caution as previous analysis has found that there are data quality issues with this field.

Which time period to choose?

Generally it's best to pick a single financial year to analyse data from. The "latest income" and "latest expenditure" fields in the main charity table cover different financial years, some of them quite old. So you may be best to pick a consistent set of data using the "publicextract.charity_annual_return_history.txt" table.

Charities have ten months to submit accounts, and most charities have a financial year end in March. This means that if you were looking at a May to April financial year, you should expect sufficient data to be available in the following January or February. This means that for accounts in the year ending March 2023 (the 2022-23 financial year), you would expect to have a full (or complete enough for analysis) dataset by January 2024.

I would recommend a financial year covering 1st May to 30th April - a number of charities have accounts ending in April, and it makes sense to include them in the previous financial year.

Detailed financial data

The data analysis described above covers the data available for all organisations. For larger organisations - those with income over £500,000 - there is more detailed financial information available in the "publicextract.charity_annual_return_partb.txt" table. This information covers:

- **Income types** - whether the income is donated (from public, trusts & foundations, government or companies), earned (from fundraising or providing services) or from investments.
- **Expenditure categories** - whether spent on charitable activities, grants to institutions, fundraising or governance
- **Balance sheet** - details of the assets and liabilities of charities
- **Number of employees**

Descriptions of the different fields can be found in this guide.

Comparison with national figures or other areas

To give context to your data it is useful to compare it with national data. The best source for this is NCVO's Civil Society Almanac, which is released yearly. It's possible to compare your own statistics with the national aggregates produced by NCVO, although it's worth bearing in mind that NCVO makes decisions around what the population is that might not match the decisions that you have made - so the data may not be exactly comparable.

For comparison with other areas, there are a number of ways to approach it. You could run a similar analysis for other areas - particularly if you prepare your analysis in a way that makes it easy to repeat it. The difficulty may be in the time & knowledge needed to apply the same population data, to ensure you have comparable data.

In conjunction with survey data

One advantage that using the administrative charity data gives is that you can use it to add context to results you get from a survey of organisations. Probably the most useful part of this is to understand which kinds of organisations were more or less likely to respond to the survey. For example, you could compare survey respondents by income band to the population by income band. You may find that larger organisations are more likely to respond.

Using this information you can add explanations and caveats to your survey findings. You can even “weight” the survey results to take into account different levels of response by size.

This process is easier if it is simple to work out which of your population has or hasn’t responded to the survey. One way to do this is to ask survey respondents to record their charity number - you can then use this to match the data. It’s important to be aware that this then makes survey responses identifiable to a single organisation, so you will need the appropriate reassurance and processes in place to ensure that respondents are comfortable sharing that data and are answering the other questions honestly.

Another use for the administrative data is as a population frame for the survey. You may have to supplement it with data on other organisations who are not registered charities and with contact details. There are some contact details within the Charity Commission data, but they are generally very low quality and users have reported low response rates when using them for survey work.

Useful Excel techniques for data analysis

The analysis being suggested above is mostly about doing some relatively simple counts, sums and lookups to produce aggregate statistics about the charities in an area. To do these, there are some techniques that are useful:

Using Tables

Tables, which are added by default if you use the Power Query method described above, are a good way of organising your data. They can help ensure that any formulas you use are copied correctly to every row in the table.

XLOOKUP (or the older equivalent VLOOKUP)

XLOOKUP is a function that is useful for combining data into one place. For example, if you wanted to add the charitable objects from the governing documents file to the main charity file, you could use XLOOKUP to add a record for each one.

COUNTIFS/SUMIFS

Like XLOOKUP, the COUNTIFS and SUMIFS functions are useful for combining data from two sheets. But these functions can be used to count and sum fields rather than just bringing their values across. Could be useful if you wanted to, for example, count how many charities were in a particular Charity Commission classification category.

Pivot Tables

Pivot tables make it easy to summarise the raw form of data. For example, you could look at how many organisations fall into each of the Charity Commission classification categories in the charity_classification file.

Other data sources

As well as data from the Charity Commission, there are a number of other data sources which can provide additional data for analysis of the local voluntary sector.

360Giving

360Giving is a charity that provides a data standard for publishing data on grant funding, and helps grant makers to publish data about grants they have made. 360Giving provides GrantNav, which brings this data together in one place, and allows searching and filtering of the data. Current publishers of 360Giving data include national lottery distributors, central government, community foundations and other grantmaking foundations.

Within GrantNav you can filter or search by local authority, based on information that grantmakers have provided with their data. It's important to bear in mind that not all funders publish location data, so you may need to do a number of searches to get the data you're looking for.

From GrantNav you can download a CSV export of search results, and analyse that file. In many cases funders will provide a charity number or other organisation identifier which can be used to connect the grants data with Charity Commission data.

Companies House

It is possible to use Companies House data to understand other forms of non-profit organisations that operate in your area, for example Community Interest Companies (CICs), Registered Societies and Companies Limited by Guarantee.

The Companies House data downloads contain over 6 million companies, split across 6 large files. This means it is much more difficult to deal with, and you may need specialist data analysis tools to handle it.

Gender Pay Gap data

Companies with over 250 employees are required to report data on their gender pay gap. This data is published in one place, which includes data downloads. If your population has a large number of larger organisations then this data may be useful to analyse.

The data does not contain charity numbers, but it does include Company Numbers which could be matched to the Company Number field in the Charity Commission data (in the main charity table).

Care Quality Commission

Data from the Care Quality Commission contains charity numbers, and so can be matched to Charity Commission data. This data enables two particular uses:

1. The data can be used to identify charities that are registered elsewhere but that operate in your local area. The list of individual locations regulated by the CQC is used for this purpose.
2. You can look at the relative performance of care providers in your area, based on the CQC ratings applied.

Additional resources

There are a number of additional resources which point to other data sources and guides which may be helpful. This include:

- **VSSN list of data resources**: https://www.vssn.org.uk/resources-for-research/quantitative-data-and-resources-for-researchers/ - VSSN is an association for voluntary sector researchers
- **Charity Sector data repository**: https://dataaboutthesector.softr.app/data