**Ch.7**

Interviewing data

When talking to journalists about using data, many trainers draw the comparison with interviewing a spreadsheet as you would interview a person. It’s an analogy which holds up well.

Firstly, data is just a source of information in the same way that a press release is a source, and so is an interview. What’s more, an interview can be probing, or go into depth. An interview may be with a whistleblower who wants to tell you something but doesn’t know how, or you may be trying to get answers out of someone who doesn’t want to tell you what they know. Interviewing a dataset is very similar. Some data is so well organised that finding out what you need to know is simple, some is messy and disorganised, and it takes a lot of extra work to dig the information out.

Most importantly, the comparison with interviewing a person should serve as a reminder that we need to decide how much to trust them, we will almost always need to talk to other sources.

And, just as we do with interviews with people, we can go in unprepared, or well-prepared. We can interview someone just because they are there – perhaps a famous person who is offered to us by their PR team, or we can track someone down and ask them searching questions we have spent a long time preparing.

That said, when we ask questions of a dataset we will usually be doing the following:

* sorting
* filtering
* summarising

The key at this trick is to translate the questions we want to ask into the operations we need to perform to ask them. So:

* sorting – “what’s the biggest, the smallest, the oldest, newest, most expensive…..?”
* filtering – “I only want to talk about what happened in 2020”, or “I am only interested in events in Newcastle, between July and September last year” or “I need to know about purchases worth more than £50k, made by your company before 2015”
* summarising – “what was the total value of purchases worth more than £50k, made by your company before 2015”, “how much did the Ministry spend on the top 5 items in total in 2020?”

In spreadsheets and analysis programs sorting and filtering are often called just that. Summarising is most easily and usually done by making a pivot table – a way of grouping together all records in your chosen category, adding up the numbers in them, and displaying them in a compact table.

Remember – the most important thing in all “computer aided reporting” is getting the computer to do as much of the work for you as you can – freeing your schedule and your brain to work on the story itself.

So – if you’re doing something boring and repetitive – you’re probably doing it wrong!

That is – for many processes there’s a built-in tool or operator which will do it for you; if you don’t know it yet, then use google search or youtube to find a more efficient way of doing what you need done.

If you move on to script-based programs like Python or R you will start to learn what you yourself can automate to get the job done.

PIVOT TABLES­

What is a pivot table?

It’s a feature embedded in most spreadsheet programs – notably MS Excel and Google sheets – which enables you to choose which variables (columns) to group together in order to summarise their contents. The name “pivot” refers to the notion of swiveling a table around your chosen point.

A pivot table takes the original data, but doesn’t edit it. Instead it makes it possible for you to reshape it to meet your needs without having to do any copying or pasting of the data. Think of it as a kind of lens – you can focus on details, but you don’t alter the data themselves.

Beware – when you first meet pivot tables and start to appreciate their power you may be tempted to pivot just about every datasheet you come across. But it’s a bit like that saying “to someone with a hammer, everything looks like a nail”. You can’t, for example, pivot a pivot table – that is, you can’t summarise something which is already a summary. (Well, you can – but you can’t learn anything meaningful by doing this!)

How can you tell that something has already been summarised/pivoted? It’s important to be able to see the difference between raw data and summary data. For example, if you can only see entries which are totals of something else then the chances are that you are looking at someone else’s pivot table. Raw data will have the original entries, and, importantly, could usefully be pivoted.

For example – which of these tables show raw data, and which are the summaries/pivots?

Graphical user interface, application

Description automatically generated with medium confidence

Table

Description automatically generated

A picture containing text, receipt

Description automatically generated

Table

Description automatically generated

That’s right – the last two tables show raw data. Even without the names, the 3rd table still contains only raw details – whereas the first two contain summaries. So trying to pivot the first two tables is not going to get you very far – somebody else has already done that step.

Take another example - you open a table that you have downloaded and you see something like this –

Graphical user interface, table

Description automatically generated

Ask yourself – even if I could pivot it, what would the result look like? Put another way – what questions could I ask it? That’s right – someone else, in this case the Office of National Statistics, has already summarised the original data for you. Just about the only thing your pivot of this table could tell you would be the grand totals for each row and column – which you can work out from the table itself.

**Making a pivot table**

A quick reminder of how to make a pivot table. Slightly annoyingly, Excel and Googlesheets do the same thing in slightly different ways, so here are the two different methods –

**Excel**

From your main data worksheet, click on Insert, and then choose Pivot Table – the icon on the far left near the top of the screen. (NB – don’t be tempted by “Pivot Chart” in the middle of the same menu: it will take you on a rather confusing route to the same place!)

You should see a little dialogue box as below – make sure the cells detailed in “Table/Range” are the whole table, not a part of it. You cannot edit this box by hand: if the entry is incorrect, you have to cancel and return to your datasheet, and start again. Make sure you only highlight one cell in the data you intend to pivot.

Table

Description automatically generated

When you click OK you should see a new worksheet which looks like this –

Graphical user interface, application, table, Excel

Description automatically generated

The key to remaining sane, and knowing at all times what variables you have put where is the PIvot Table Fields selector:

Graphical user interface, application

Description automatically generated

To summarise all payments to political parties (“entities”) we drag the “Regulated Entity Name” entry in the PivotTable Fields chooser to Rows. And “Value” into the Values box:

Graphical user interface, application

Description automatically generated

Then to sort the new Sum of Value column into descending order, right click on any one of the totals in that column to bring up a little menu containing the option to “sort highest to lowest”.

Remember – summary totals like this are clickable – double clicking on a total in a pivot table will ask the program to make a new sheet made up of all the numbers which make up that total – along with the other information associated with those payments.

To go back over the basics of pivot tables with a demo, and an opportunity to watch short video demonstrations of the various techniques, click [here](https://www.dropbox.com/s/z3rb1elh0hio310/donations2019_2020_Exercise.xlsx?dl=0)

**To make a Pivot Table in Googlesheets**

From your main data worksheet, click on Data, and then choose Pivot Table – about 2/3 of the way down the menu which appears.

Graphical user interface, application, table, Excel

Description automatically generated

You should see a little dialogue box as below – make sure the cells detailed in “Table/Range” are the whole table, not a part of it. You cannot edit this box by hand: if the entry is incorrect, you have to cancel and return to your datasheet, and start again. Make sure you only highlight one cell in the data you intend to pivot.

Graphical user interface, application

Description automatically generated

When you click OK you should see a new worksheet which looks like this –

Graphical user interface, application

Description automatically generated

The key to remaining sane, and knowing at all times what variables you have put where is the PIvot Table Editor

Graphical user interface, application

Description automatically generated

To summarise all payments to political parties (“entities”) click the Add button next to Rows in the Pivot Table Editor (right hand side of screen) and select “Regulated Entity Name”. Next to the Values box, we click on the Add button and select “Value”.

Graphical user interface, application, table

Description automatically generated

You should see a screen like this. To sort the results in descending order we need to go back to the Pivot Table Editor and change “Order”, in the box called RegulatedEntityName to Descending.

Graphical user interface, application

Description automatically generated

“Sort by” needs to be changed to “SUM of Value” :

Graphical user interface, application

Description automatically generated

Remember – summary totals like this are clickable – double clicking on a total in a pivot table will ask the program to make a new sheet made up of all the numbers which make up that total – along with the other information associated with those payments.

**Further exercises**

To go back over the basics of pivot tables with a demo, and an opportunity to watch short video demonstrations of the various techniques, click [here](https://www.dropbox.com/s/z3rb1elh0hio310/donations2019_2020_Exercise.xlsx?dl=0)

For practice, do try to think of questions for your data – either the example in the exercise above, or your own. Don’t just wade in and start clicking – make notes of questions as you would in preparing an interview with a person. If you don’t, it’s just too easy to get lost down rabbit holes and dead ends.

If you need more help on Pivot Tables don’t forget that youtube is full of tutorials. The quality is mixed, of course, but you will almost always get visual help on the problem you need to solve.