Introduction to Excel Handout

www.weahsn.net/academy

Video available at https://vimeo.com/556277841/33ca32ac14

Contents

0:00 Introduction	2
2:52 Orientation	5
Download the dataset	5
Ribbon	5
Workbook structure	5
Moving around	6
Setting up views	6
Sorting and filtering	7
Duplicates	8
Review	8
Orientation to your data	8
16:37 Processing data	9
Download the dataset	9
Formulas: MAX, MIN, AVERAGE, COUNTIF	10
Calculating percentages	11
Absolute and relative references	12
=COUNTIF	12
Review	12
Pivot tables	13
Number format	14
Refreshing data	14
Slicers	14
Show report filter pages	15
Grouping data in a pivot table	15
40:11 Cleaning data	15
Download and orientating to the data	16
Structuring data for a pivot table	16
Paste special	16
Number formats	17
Text manipulation	17
Review	18
Working with dates	18
Text to columns	20
Review	21
1:09 Presenting data	21
Choose the right chart	22
Reduce data clutter	23
1:15 Takeaways	23
Further information	24
Thank you!	24

0:00 Introduction

Hello, my name is Nathalie Delaney. I'm the Patient Safety Programme Manager at the West of England Academic Health Science Network.

This video is for attendees of the Making Data Count course to learn a bit more about Excel.



This video is in four parts we're going to cover orientation to Excel and your data, how to process your data, cleaning up your data to get the best from it, and some ways of presenting your data.

The time stamps for each of these sections will be below this video.



Let's get you set up to get the most from this video. I really recommend that you have a copy of Excel open and follow along on the screen.

Download the datasets that we will be using so that you can work with real data.

You may want to have a pen, paper and something to drink to hand.

And please do take your time; feel free to pause, rewind or take a break at any point.



There's a few things that we won't be covering in this video.

I'm not going to teach you about quality improvement or how to choose the right measures but I've recommended two excellent videos on these topics.

- Introduction to Quality Improvement <u>https://youtu.be/Rk5wHFmAthA</u>
- How to choose the right measures https://youtu.be/Za1o77jAnbw

Don't worry there's going to be no statistical science or complex maths; everything in this is GCSE at most.

This is not going to be a comprehensive guide to everything Excel can do, but the key things and some quick tips that you may need to know when working with data for improvement.

I'm using Excel 2010 for windows. I know Office 365 has some shiny new features and I'm looking forward to getting stuck into them but since most people are still using 2010 I'm going to be using that.

I'm not going to be using any macros or visual basic or coding.

Let's get started.

We're going to be using three dataset and a set of tools from the Making Data Count FutureNHS site. You may find it helpful to open the links now so, you have them ready.

- Orientation EPCN data <u>https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data</u>
- Processing your data CQC care home directory <u>https://www.cqc.org.uk/about-us/transparency/using-cqc-data</u>
- Cleaning your data Covid Oximeter Weekly Report
 <u>https://digital.nhs.uk/data-and-information/publications/statistical/mi-covid-oximeter-weekly-report</u>
- Presenting your data <u>https://future.nhs.uk/MDC/grouphome</u>



I'd like you to reflect for a moment on where you're starting from.

- Are you comfortable navigating around Excel and orientating yourself to data? If not, start at very beginning with orientation.
- Do you know how to use formulas and pivot tables to ask and answer questions? If you don't processing your data is the right place to start.
- Once you know that, are you able to tidy up your data so, you can work with it more easily? We're going to be covering that in the third section.
- And finally, in presenting your data we're going to be covering the technical aspects including checking for errors and using the right file format.

2:52 Orientation



Let's get started with orientation.

Download the dataset

We're going to be working with data from the NHS digital organisation data service.

https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data

Scroll down to download data and then go to GP and GP practice related data.

The file you want is all the way at the bottom it's called **EPCN** and click on that and download it to your desktop.

Open the Excel file for the data we're going to be working with in this first section.

Ribbon

At the top of the screen, you can see the **Ribbon**.

It has a Home tab, Insert. Page layout. Formulas. Data. Review and View.

You can show and hide the ribbon by clicking on this **arrow**. That can be really helpful when you want more space on the screen.

And a lot of the useful things that we're going to be using or in this ribbon

Workbook structure

Next is workbook structure. A workbook is what we call a file in Excel. This workbook has two worksheets or tabs: PCN details and PCN Core Partner Details.

Moving around.

There are different ways of moving around. You can use your mouse. here I'm using my mouse clicking on different cells. Clicking on a column. And scrolling up and down.

As well as your mouse, you can use the keyboard. I've clicked into the first cell here: A1. Cells are what we call the boxes. I'm moving left and right using the arrow keys and up and down using the page up and page down buttons.

When you're working with a large table, if you hold down **CTRL (control)** and use an arrow key to go left or right, or up and down, you can see that I'm moving to the edge of my data, so, I've moved all the way to the right and now I'm going to move down. And then I'm going to move across back to the left.

Now did you see instead of going all the way to the first column, I actually skipped over one. That's because there's a blank cell there. You may sometimes need to go one or two times with the arrow key.

Try it for yourself, moving around your table. If you hold down the **SHIFT** key and the **CTRL** key, you can see that it selects all the data in that place. that's a quick way to select the whole table by going across, holding down, **CTRL SHIFT** and using one arrow key and then going down to the bottom of your data.

You can also go to the top of your dataset. Do you see this little box here? And if you click on that, it selects the whole data sheet.

Setting up views

Now let's talk about setting up views. You can see when I go down, I can no longer see the headers at the top of my columns. it's hard to know what each of these mean. The way that you can resolve that is using something called **freeze panes**, which you find in the **View** tab on the ribbon. You can freeze the top row by clicking on freeze panes and then choosing **freeze top row**. This keeps those headers the same. See now as we scroll down, that header row is staying in place. You can also **freeze first column**. This means that as we scrolled across that first column is always staying in place.

But what if you want to freeze the header row and the first column? To do that, you click in the cell that my mouse is in here. it's the cell underneath the top row. And it's the cell to the right of the first column. Let's highlight that cell so, you can see which one I'm talking about. Now when you go up to the view panel and freeze the panes it will freeze both the top row and the right column. Let's try that. Now go into view. **Freeze panes** and you can see as I scroll down, keeping the one at the top, and as I go to the right that's also keeping that first column. that's freeze panes, if you ever want a quick reference, it does show you in these icons here.

Sometimes you want to see all of your data on the screen at once. There are different ways to do it. You can use the **zoom** down in the bottom right hand corner in order to zoom in and zoom out. You can see there's a lot of blank space here, so, we can change the column width. Do you see how when I move my mouse here it goes from a white cross to a black double arrowhead and that if you hold down the mouse button and click that reduces the width of that column. We'll do it to the next one here, and one more. Now I can move the columns in, and I can now see much more of my data on screen. One of the other things you can use. If I reduce this one in size, you can see that it's cut off the edge of the partner name. If you double click, it will automatically make that column the width that it needs to be for all your data to be in there. Or if you prefer, select the column and then click **wrap text**. And what it will do is it will make that go over more than one line, and that's another way that you can have your data on the screen.

Sorting and filtering

The dataset has all the PCNs in England in it. But we are only interested in the ones from the South West region. To find those, we're going to need sorting and filtering.

Do you see these little grey boxes at the top of each of the headers? Those are the **filter** and you can use them to select different parts of your data. You can see that this one has an arrow pointing up against it and this one doesn't have an arrow that shows that the data is being sorted by one of these columns. The arrow shows that it's sorting A to Z: starting with the smallest and going to the largest.

If we sort it Z to A by clicking on the box and sorting Z to A you'll see it's now in reverse alphabetical order, starting with NHS Wirral CCG.

Let's go back to the original sort. By selecting your whole table and going to the sort function on the ribbon you can choose **custom sort**. You can sort by more than one thing at once. we could start by sorting with PCN parent CCG name and then adding in PCN name and then finally adding in the Partner name. We can see that all are being sorted A to Z and you have the options to choose different ones however you want to sort it.

Now we have NHS Barnsley first, then Barnsley PCN, and then inside that is alphabetical order for each of the practices there.

You can **clear a filter** if you want to remove everything or reapply if you've changed a few bits of your data and you want the filter to reapply again.

Now we're going to filter for the ones in our region. we click on CCG name. You might want to make this box for little bit bigger to help working on it. Unselect things and then go through looking for the ones from our region.

And there we go. These are the practices associated with PCNs in the NHS South West region.

I'm going to show you a quick tip now. Take all the ones that you have in your data that you filled out and highlight them in yellow. When you clear the filter will see that these ones still stay as yellow.

Now if we'd done some more manipulation of the data but wanted to go back to our filter, we'd have to go through that thing again by clicking all those different CCGs and making sure we had the right ones. And there's a risk that you'll miss one or get it wrong, but actually, because we put them all in one colour, you can go up to **filter by colour**. Choose yellow, which is the colour we picked and that will take you straight to all the ones in our region.

I find using filtering by colour is a really helpful way to make sure you're doing your data quickly, and that you're getting it right and the same every time.

Duplicates

Let's check we've selected the right data. Copy that column into a new tab. With your data still selected, go up to the top of your ribbon to the data tab. And from there select **remove duplicates** and click OK.

This takes all the selected data and removes any duplicates so now you have a list of what is the unique information in that column and you can see that we've got all the CCGs that are in the NHS South West region here.

The other thing you can do is select a column, and then if you go to the Home tab and go across to **conditional formatting, highlight cell rules** and **duplicate values**. You can highlight all cells that are duplicates in a column. you going to choose format cells are duplicate. OK. All cells that have duplicate values are now shaded in red.

But can you see although that Three Checkers medical practise is in this data and the filtered data. We also have the Wellspring surgery, which is in red, but there's no other Wellspring Surgery in the South West and that's because this is looking for duplicates in your whole dataset, not in the information is filtered.

If you **remove the filter** from your data you will be able to see that. we remove the filter. And then sort those A-Z. And you can see that Abbey Medical Centre, Abbey Medical Practice, Abbey Road Surgery, are popular names in different parts of England, and that's why you have the duplicate showing up.

Review

Let's review what we've covered so far.

- We talked about the ribbon, workbook structure and tabs.
- We've tried moving around using both the mouse, keyboard and keyboard shortcuts.
- We've talked about setting up your view using zoom, freezing panes, changing the column size, and using wrap text so, that text goes over more than one line in a cell.
- We've used sort and filter to slice and dice our data in different ways and we've looked at two different ways of working with duplicates: removing duplicates and using conditional formatting to see where the duplicates are in your data.

Orientation to your data

The other thing to think about is orientation to your data. Ask yourself the questions:

- Can I trust this data source?
- What am I really looking at with this data?

When you downloaded this dataset, there were two files. If you look at the second file, it gives you some useful information about how the data is collected and also what fields are recorded in the data.



Now we're going to try processing data using pivot tables and formulas in order to get some information out of the data.

Download the dataset

This time we're using a dataset from the CQC. Go to their website, click on **download our data** and then scroll down. Looking for the data that is **care directory with filter**. Click on it to download and open the file.

https://www.cqc.org.uk/about-us/transparency/using-cqc-data

You'll see that this dataset has useful information on the first tab marked as readme and the data itself is on the second tab: HSCA active locations. This is quite helpful if we need to come back and refer to it.

This is a large dataset. It goes across to column DO and if we scroll down to the bottom of our dataset, there are 50,100 rows. This is a lot of data and it can be hard to deal with a file that's so large particularly when you're manipulating it. let's start by reducing our data size to the data we want to work on.

Before you do anything, I recommend you start by saving. We go up to **save as** a new file. You can see that the file format of this is an open document spreadsheet. This is a format which can be opened by lots of different types of programmes, but because we're going to be using the Excel functions, we want to change it to an Excel workbook so that we can use all the functionality of Excel. We rename the file here: CQC data and save then we have our new file set up. Then if we need to go back to our original data, it's there for us.

Let's start by filtering out all the rows that are not a care home. Remember on readme it says to obtain care home locations filter for Y. What we going to do is delete the cells we're not interested

in. instead of selecting "Y" we're going to select, "N". This means these are all the rows that we're not interested in, and we can select all those rows and **delete** them.

That's reduced our data down to the care homes, we can check that there's nothing left in that column we saw that it only gave us the option "Y" for that.

We've now got about 15,000 rows which is much more manageable, but next we want to make sure we're looking at homes from the NHS South West region. Let's have a look. There's a location region: ah – location NHS region that looks like the helpful one. Let's see what it has... It does have a South West one. Click on that and then we go across. We can see it's got all the CCGs that come under this one.

Let's check that we've got all the right ones in there. Yes, that looks like this is the right field to filter on to get the information we need, sometimes with an unfamiliar dataset, you may need to have a look to find the right one.

We'll do the same thing, so, we're going to select everything that is not South West. And then we're going to delete these rows to give us a smaller dataset. Delete these rows, and now our dataset is much smaller, and save that before we go any further.

I'm interested in whether home is a nursing home or a residential home. That doesn't exist in the data, so, I'm going to need to create a new column in order to help me with that. I'm going to **insert column** and call that one "Nursing or residential?". And then I'm going to use the filter to help me find the data for this field.

First I'm going to select all the care homes that have nursing using "Y" put "nursing" into that column. And then **paste** that down. I'm going to clear the filter and select everything that's blank: everything that still doesn't have information in.

I'm going to do a quick check that's picked up all the other rows so in here, yes, everything in there has a "Y" in. That's all my data and I'm going to put residential in. When you **copy** that and you see how the cell has the little ants marching around it, that means that that cell has been copied to the clipboard.

I'm going to use my shortcuts to go down to the bottom and then paste them in and I'll save it there.

Formulas: MAX, MIN, AVERAGE, COUNTIF

I've made some pivot tables from the data for us to work with. Later on, we're going to be creating these pivot tables for ourselves, but for now, follow along on the screen.

First, we want to know what's the home which has the maximum number of beds. There are two ways of doing this. The easiest way is to take the column number of beds. Go up to the top where we had our sort function and sort largest to smallest.

But what if you don't want to change the order that your table is in? Well, we can use formulae to find out what the maximum, the minimum, and the average number of beds are in each of our areas. And we can also use a formula to find out what's the average number of beds per home in each area to see if it's different.

To make a formula, you click into the cell that you're going to put the formula into. And type in = to start your formula. The maximum you then type in **MAX**. Open brackets and select the cells you are going to be using, in this case C6: C12. Do you see there's now a blue box around these cells?

To finish your formula, you close the bracket and hit return. This tells us that the maximum value in that range is 12,630, which is NHS Devon.

To do a minimum, we did the same, so, type an = to start the formula type **MIN** for minimum. Open brackets, the same data source C6:C12, close the brackets and hit return.

The formula works in the same way for an average, so, we will go = AVERAGE (C6:C12)

To get the average number of beds per home, we need to do a little bit more data manipulation. if you go up to cell D6 type = we're going to want to divide the number of beds by the number of homes = C6 divided by that's the slash sign B6. **=C6/B6** and then I'm going to fill that formula down by copying and pasting it down. Let's tidy up that **number format** by choosing the number. I'm going to talk about number formats a little bit more in Part 3.

You can see I've got the average for each of the areas. It ranges somewhere between 23 and 30 beds per home. And we can do the same with the total at the bottom using the same formula. copy it down and that gives us the answer the average number of beds per home is 27.30.

Notice that we're doing an average of the total we're not doing an average of those cells D6 to D12 because that would give us an incorrect average. Always make sure you are doing your average of your total values.

As a quick check you can also see some information at the bottom of your screen. You will see I selected some cells shown in blue and at the bottom of my screen there it says average 7869, count 7 and sum 55,084. That's another way to get some basic statistics if you're looking for that information quickly.

But, however, we wanted to know the average number bed per home, so, I'm going to put = and then D13, which is the cell that the answer is in. But that means that you're referring to that cell so that you don't have to copy and paste information across.

Calculating percentages

Now let's look at calculating some percentages. Let's say we want to know what percentage of homes in each area is split between nursing and residential.

I've got a pivot table here. Remember, we're going to try and build this ourselves later. I'm going to add in two headings, the first one is the percentage of homes that are nursing. And the next one is percentage of homes that are residential. And it's a similar formula I'm going to divide one number by another.

In this case I'm going to divide the number of nursing homes that's in cell B7. I'm going to divide it by the total for that CCG, so, that's D7 and I'm going to use the **number format** to make that a **percentage**. 29% of homes in Bath, North East Somerset, Swindon and Wiltshire are nursing homes.

We can copy that formula down in a number of ways. We can click and drag by using the black box on the corner of the cell. You'll see that it's copied the formula down and it's using the right one, so, it's gone from B7 to now referring to B11. So that was using the black square to drag it down. You can also copy and paste formula by selecting it all and then paste in the formula in or you can do **CTRL+ D** to fill a formula down.

Absolute and relative references

We've got the formula here. Now, let's copy it across into the cell to the right. Oh, dear those numbers don't look right, do they? What's going wrong?

If you double click on a cell, it will tell you which sells its pulling for its formula and we can see that while residential moved across grand total also moved across. That's not helpful. We want it to be using the same grand total for both calculations.

And the way you fix that is to use something called absolute and relative reference. We're using a **\$.** let's look at that now.

We go to that cell at the top and put a \$ in front of D. What it does is fixes that formula to that column. even when you copy it, it's still going to stay fixed to D. Let's try it. Let's copy that across and you can see that worked. If we click on one of those cells now you can see that it moved across the residential one C7, but it's fixed that grand total as D7 that it's looking for. Absolute reference uses a \$ to fix something a relative reference doesn't, and those move as you copy and paste them around.

But what if you want to look at a percentage of the whole grand total? Let's clear those cells. I'm going to be starting with cell B7 so = B7/D14 and this time we're going to put the \$ in front of both the D and the 14, because wherever this formula is copied, we want that cell to stay in the same. We're going to copy it down. Let's check. Yes, it's referring to D14, that's working.

Let's copy across to the right. Yes, that's working there.

=COUNTIF

Now this dataset has lots of columns where the data is structured that it's a "Y" if it applies to that row. This data can be quite tricky to work with as a pivot table and you don't really want to be filtering to find out how many in each column, but you can use a formula called **COUNTIF**.

Same way as before = for formula =COUNTIF (to choose the range, but this time we're going to add something else to the formula. "Y" and that "Y" is in quotation marks that shows that it's text that we're working with and then we close the brackets to finish the formula and then hit return, and that will count how many of the cells, how many of the rows in that column have a "Y".

Then you can copy that across and it will give you the same information for each one of those columns. And that's the quickest way to work it out.

Review

Let's do a quick recap. Remember, you can pause this video and rewind whilst you're working with the data if you need more time.

- We covered formulas calculating minimum and maximum.
- We calculated an average in two different ways
- We also looked at working with percentages

- When we were using those formulas, we used a \$ to separate out relative and absolute references.
- We also looked at a formula called COUNTIF which is a way of finding if something applies in a certain range.

Next, we're going to cover pivot tables.

Pivot tables

Start by selecting all your data that you want in your pivot table. Once you have all the data selected, go up to the top of the Ribbon under Styles and choose **Format as Table.**

We can pick any of the styles. This is my personal favourite. It will ask you to confirm where's the data for your table. This is what you selected and whether or not you have headers. Ours does, so let's click OK. Now your data is formatted as a table.

Go to the **Insert** tab on the ribbon. You can see we have two options to add a **pivot table** or a **pivot chart**. If we click add a pivot table it comes up with few options and asks us to confirm the table we're using and whether we want to add as a new or existing worksheet. Now we're ready to build our table.

The Pivot Table Builder has two sides.

- On the left hand side is a box where our pivot table will appear.
- On the right hand side we have a **field chooser** where we can choose fields to add the report. You'll see all the column headings have been brought in automatically.

There are four boxes to build. And then at the bottom, this box **defer layout update**. You would click this if you're working with a complex pivot table and you want to make sure that things don't change while you're working on building it.

For now, let's start by putting some fields in. We can start by putting CCG into the filter so that we can choose between the different CCGs. We also want to add in that field we put in our data earlier whether a home is a nursing or residential home.

Next we'll add in some row labels. **Care home name** would be a good one to start with, so let's add location names into the table.

It can be helpful to have one of the values set up as a field where there is some data in every row in your dataset so that you can count things correctly. Every home has a **location ID**, so this seems like a good field to use. Let's add it to the report. You can see we now have a count of 1 against each care home. This is going to be useful later.

Another thing you might want to count is number of **care home beds**. Let's drag them into values.

Excel has been clever enough to realise that we want to add up all the care home beds. But it doesn't always do this, so let's see how we fix that. If you click on one of the values, say count of location ID. It brings up a menu and what we want to choose on here is **value field settings**. This will bring up a menu which has a number of useful features. The first one is you can choose how you want to summarise your values. Do you want to add them up? Do you want to count them to see

how many there are? You can also do other basic statistics, so averaging maximum minimum and others.

You can also do a calculation if you want to show something as a percentage of your row total column total. There's lots of options there. You can also rename the field. Count of location ID isn't terribly helpful, so let's change that to "number of homes".

Then we'll do the same for care home beds. Value field setting. "Number of beds"

Our pivot table is now the same as the one we used in the earlier section about manipulating data using formulas.

Have a go yourself building a pivot table on some data you're familiar with, have a play.

By moving two fields around we get to the table that we use when we were calculating percentages.

One thing to know about pivot tables if that if you double click on any of the **numbers** in the pivot table, you can go to the filtered list of that data. Let's try it by clicking on NHS Gloucestershire, CCG which is **219**. And it sets up a new tab which has all the data on there. Then we'll see if we go to the bottom there are 220 rows, so that's the 219 plus the header row.

Number format

Let's make these large numbers in number of beds look a bit friendlier. Go down to number of beds, value fields setting and click on **number format**. From here you can choose a different format to make it easier to see. I've chosen to add 1000 separator and have no decimal places. Click OK, OK. And now the data looks like this.

Refreshing data

A pivot table is linked to the data behind it, so we can make any change in that data. For example, I'm going to delete some rows at random here from our original dataset. delete a few words. Go back to your pivot table. And click on **options refresh**. You can see that the number of homes has reduced to reflect the change in that data.

Slicers

You can add interaction to your pivot table with a slicer. These make it much easier and faster to manipulate pivot tables. Go up to **options**, **insert Slicer**. You can choose any of the fields that are in your table. Let's add the ones we had before "nursing or residential home". move it to the side. Gives you two buttons nursing residential and you can resize the slicer to make it easier to use.

Add a second slicer for our CCGs. As before we've moved and resized the box to make it easier to read.

We can click on any of these blue buttons to filter the data by that. Let's try with nursing. This is now nursing homes in the data, and the summary there. Remove the filter by clicking the little funnel with the red cross. Choose "residential" and then clear again.

By holding down the **CTRL** key when you click the choices, you can choose more than one option. So now we're going to choose Bath and North East Somerset, Bristol and Gloucestershire: the areas in the West of England region. This gives this summary with these three.

Show report filter pages

With CCG already in your filter, go up to **options** in Pivot table tools. And then **options**, **show report filter pages**. And then click OK. This sets up a new tab for each one of the filters, in this case CCG, so that you have separate reports. This is a really powerful but slightly hidden function.

Grouping data in a pivot table

If we move some of the fields around so nursing up to filter, number of beds across to rows. We can now see how we group data in a pivot table. Down the row labels we've got the number of beds in the home and across here we have the number of homes so there are 14 homes with one bed, 33 homes with two, etc.

Under options, **pivot chart**, you can choose a chart to display this data, so I've removed total and total removing this to show the charts available here. We now have Number of beds on the bottom and number of homes along the side.

Hover over the rows and right click to get a **group** function. It goes from 1 to 169 and let's group them by groups of 25. put 25 there, go OK. And that has updated both our pivot table and our chart to make that data a bit clearer and easier to see.

40:11 Cleaning data



So far our datasets have been well structured and it's been easy to sort, filter and construct pivot tables from them. But what if you've extracted data from a system and it isn't in a format that's easy to use? That's what we're going to cover in this section.

I'm going to show you some different ways of cleaning up your data and re-structuring it to make it easier to work with. Let's get started.

Download and orientating to the data

In this section we going start by working with some of the COVID oximetry weekly reports from NHS Digital. If you go to the link <u>https://digital.nhs.uk/data-and-information/publications/statistical/mi-covid-oximeter-weekly-report</u> and then scroll down, you will be able to see some information about the data and then you can access the most recent report.

I'm going to be using the most recent data from the 20th of May. If you're watching this at a later date, you can find all the historical data underneath, so you may want to download the 20th of May data, or you may want to work with the most recent data set. It's up to you, but for now click and download.

As we learned in part one, let's start by orientating ourselves to the data. We've got lots of tables here and some contact details if we do need to contact anyone about the data. And helpfully it tells us when we can and can't reuse the data.

There's also a notes and definitions tab. This gives us guidance on anything that may or may not be in the data. And also, any particular symbols used. In particular, we can see that counts are rounded to the nearest 5. And any counts between one and seven inclusive are suppressed and shown by a star in the tables. This is information we may need to use later in cleaning up our data.

We're going to be working with table 1A. This is a count of patients onboarded to COVID oximetry at home programme and its weekly data and there's one row for each CCG.

Structuring data for a pivot table

When you structure data for a pivot table you want to have it so that when you add more data to your table you're adding it down so you're adding it as rows underneath previous ones.

But this dataset when it add more information, it's added in columns we can see in the weeks. If you think about next week's report, we'll have another column added between C and D to include the data.

What we're going to need to do with this data is turn it around by 90 degrees and we may also need to do some other manipulation because of the dashes and stars that are in it.

The first task will be to filter it to the CCGs that we are interested in for our region. **Y58** is our region code, and once it's filtered, select all the data and then paste it into a new workbook.

CTRL+N is the shortcut for a new workbook

Paste special

Let's turn the data around. Select all the data in the table. Click into one of the cells underneath to paste into and click on the arrow underneath the **paste** icon on the ribbon. This gives you more options. The one we're going to use is **transpose**, which shows as this little icon here, T.

This has taken our data and transposed it, keeping all the formatting and formulas the same. I'm going to make it a bit easier for us to see what we're doing: Deleting the rows up the top. Selecting everything. Wrapping the text. Making everything at the top of the cell. putting everything into the same font size of 10. Dates are orientated up, so we can change it so that **format cell alignment**. They're now going to be rotated the right way round. And then last thing to make the columns all the same size and reduce the row size.

Now we have a table that we can see what's happening here and get ready to turn it into something that we can work with.

We're going to build our pivot table data underneath. I'm going to start by copying the dates and select everything from A7 down. Copying that and pasting it below. Let's add a header "Week commencing".

Our next column will be CCG and we need to take each of these columns and have them underneath each other.

Using **CTRL+ X** to **cut** rather than copy, let's move our data across. So **cut** the data and put it in the next cell across cell C42 and paste to paste it in. You can see that's now all gone from that column we going do the same.

I've sped up video, but I'm cutting and pasting all the information in all, I'll remove the extra rows at the top. And then fill down the CCG names and dates.

And there's our data ready to pivot. As before we'll format it as a table, up the top in the styles, there we go and then we'll insert on the pivot table.

On table tools. Design tab on the Ribbon is a place where you can rename your table. When you're working with lots of tables and pivot tables this can make it easier to refer to in formulas. Let's rename our table now: "SWonboard"

Number formats

Number formats up here on the Home tab in the ribbon include **general**, **number**, **currency** and **dates**. Short date looks like this 10/05/2021. You could also have a longer date. 10th May 2021.

Now we'll put our data into a pivot table, same as before and build it using the fields. There are only three fields in this pivot table at the moment. You can see this has count of onboarded. We want to change this so that it's a sum: **value field settings**, **sum** as above.

Remembering that our data is rounded to five, and also that data under 7 has been removed, so we may want to clean that up later. But now we've got our data in a reasonable pivotable format.

We can group our column labels by months and years. This smooths out if there are any differences in weekly submissions. We know some areas are only submitting data every other week.

You can also choose to add a date filter, for example the last three months.

Text manipulation

Let's say we want to remove any NHS from our CCG names. Add a new column and we're going to use a formula called MID which will return some text from the middle of a text string.

We select our CCG column, and then it's a start number so "NHS " you want the 5th letter K. And then we've said we want 40 characters after that. That's not long enough for some of our cells. It's truncated South Glos, so we want to change that to be a little bit longer. Let's make it up to 80.

If we increase the size. We can see that's in there now.

I've added a new column "onboard cleaned" because now we're going to clean the onboard data. I'm going to show you why. Let's move onboarded into the row label. We want to group as we did before with care home beds. Go to group. And we get an error message. We can't group that selection. That's because we've got a mixture and we've got these non-number cells in our data, so we're going to need to remove them.

Copy across the data into column D. Then use the filter. We'll start with the dashes. That's where there's no data, so we can delete them. We're not going to put zero in, we're going to leave it blank.

Next the *. These are between 1 and 7. So let's replace it with 5 as all other data is rounded to the nearest 5.

Of course, I'm using this dataset for demonstration purposes and this isn't something you would have to do if you using this particular data for analysis. But what I've done is show you how to clean up your data so that when we go to the pivot table and refresh it our two new columns have been added to the Pivot Table Field list. This is because we set it up as a table.

Let's add onboard cleaned into the rows and if you right click you'll see that we can now group this data.

Review

Let's review what we're covered so far in this section.

- We've used paste special transpose to restructure our data so we can make a pivot table from it.
- We saw the difference between cutting and pasting compared to copy.
- We did some basic text manipulation using a formula called MID.
- We looked at data formats for dates and numbers.
- We cleaned our data where there was a mixture of numbers and non-numbers, so the data was all one type.

We're going to continue to look at different ways of manipulating numbers and text now, working with our date field.

Working with dates

Now I'm going to show you a few formulas for working with dates and other text manipulation.

We start by copying our dates in this column, let's call that date. And make our columns little bit smaller so that we can work with things on the screen. Here we go.

You might want to separate your date into the day, the month, the year, and we might even want to know what's the weekday of your date. You can do all this using a date formula.

We go =DAY(date) and it returns the day, so 10th May 2021. Same for =MONTH, =YEAR and for weekday =WEEKDAY. This is less helpful for this dataset because everything is reported weekly. But if you're working with data that might have a weekly pattern, we might want to separate out whether something is on a particular day.

If you want to turn this weekday into something that's more easily readable by human. Go down to all **number format**s. Choose **custom** and then type in ddd if you want to have a short word like Mon

Tue Wed and dddd if you want to have the full date. I'm going to use ddd now. I'm going to show you how that changes by changing this cell here. I'm going to make it yellow so you see which cell I'm changing so if I change this to 11/5/21 that weekday over here. It's changed to Tuesday. 12/5/21 21 changes to Wednesday and so on. Let's put that back to 10/5/21.

If you wanted take this information and turn it into a new field, we will go over here. And we're going to change the format of our date. I want to view my data with a date code "Mon 10-5-21". I'm going to show you a few things to get to that date code. Let's put the date code in here.

First we need to create a short year by using RIGHT this time, going to the year, 2. This is the short year. = (clicks on weekday cell) we're going to add in an & (day) & ... to add a dash we use quotation marks to show you're using text, "-" & (month) & "-" & and finally add in our (short year).

Now this doesn't look quite right, does it? Because we've got this at the front actually I'm going to tidy this up by adding space in here and there,

So you can see that this number here is a 2 to fix that we're going to make a weekday in text by using something called VLOOKUP.

Let's go here and build it: weekday number, weekday text. We start on a Sunday... I'm going to format it as a table, then go, and I'm going to call this "Day of the Week". This is a lookup table. This is column 1 and this is column 2. I'm going to ask you to remember this number here: column 2, because that's what we're going to be looking for.

It can be helpful with the look up to sort these smallest largest if they're not, but ours already is.

We go back to a data and start our formula =VLOOKUP(our weekday, ...Now we add in the name of our table. Remember we called it "Day of the Week" and because we've named our table is showing up here "Day of the Week". Double click. Do you remember I asked you to remember a column, that's column 2? That's where we're looking for returning and then finally, we choose true or false, whether it's an approximate match (true) or an exact match (false). Type in 0 and then that will make sure it's looking for an exact match. Hit return that gives you all the weekdays down here.

Do you notice something different between weekday here and weekday here? So over here everything is aligned to the right. That's because it's a number. We can see the same for ... this and this they're all numbers.

But here they're text we've used textual manipulation on this and on this so it's made them left align because they're text and you can see this also on "Kernow". As long as none of these options are chosen, text will always be to the left and numbers will always be to the right.

Anyway. Let's go back to our date code and go to change this weekday for our new column, weekday text and hit return. And we now have our date code matches above "Mon 10-5-21".

To show you how a VLOOKUP matches, let's change this date over here to the 12th. It changed Wednesday, that's changed to Wednesday and it changed it to Wednesday here. Move that back to the 10th.

Now let's do a little bit more work to show you some other things that may come up when cleaning your data. Let's say this date code was the date code that you had in the dataset, and we can't pivot or group or do anything fancy with this. So we going to clean our date.

Start by copying everything, pasting over here. What I'm going to do is paste it as values. So this means everything is now a value. It's no longer a formula, so here is a formula. Over here is the value. **Paste as value**.

Then I'm going to use MID to get the date string. Remember =MID(text, 4, ... let's have 15 characters just to make sure we get everything. Now the reason I made it 4 not 5 is because it's given us something which has a trailing space. Sometimes you need to **TRIM** your data to get rid of a training space. =TRIM (select your text string) and that just removes that trailing space.

Almost there with getting this to a date code. We have one last thing to do which is convert this **text to a number**. So this is a number and it's stored as text. One thing you can do just going to paste this in as values, you can see that there is an error which says "this is text date with two digital year" we can choose to convert it, that's pretty tedious clicking on each one. You can go into a cell and hit return. If you want to do this really quickly, hitting F2 will get you into the cell, return F2 return, F2 return. But we've got 227 rows to work with so that's going to take.

Here's another quick tip. Put the number 1 in any cell and copy it to your clipboard. Then select all the cells where you have a number stored as text. We go to the **paste special** menu that we've been working with; go to this one here, **paste special** and choose option **multiply**,

I'm going to recap that: we put 1 in a cell. We copied it to the clipboard. We selected all the text that was a number stored as text. We've gone up to **paste special**. We chose that extra optional bottom, and now we've got multiply. There we go OK.

What it's done is it's converted all those numbers into a date as a number and then our last thing to do. Go up here. Change the number format to **short date**. And we've got back to our short dates.

Obviously, this has been a little bit complicated this section because I've been showing you how to manipulate text in different ways.

We had our date here. And we have a date here. Now to confirm that these things are the same, I'm going to check dates. And I'm going to use an IF function for that, =IF ... I'm going to open this one. This first date equals. This second date, which has been through all our manipulation. If it is true, that's "YES" and if it's not true "NO".

Is that date the same? YES NO. True, hit return you can see that all our dates are the same they're all yes.

I'm going to change one of these. Because this field is no longer linked to this field, if I change one of these, so I'll change this one again. We're going be looking at this one here. If I change this to 12/5/21 we can see that changes to "NO". That is because these ones are no longer linked. I broke this around here when I pasted it in as a value.

Text to columns

The last thing to look at is text to column. So I'm going to copy this column across, go up to **data** and choose **text to columns** what I want to do is split this into first name and surname. This does it very quickly, we've got spaces between so I choose a space, next, text and then go finish. This splits at every space into a new column. That's nice and easy, we can just cut "Ross Geller" across because he had "Dr" in his name field so all the data is the same.

We can do something similar when the text field has a comma in it. Again we go **text to column** this time it's a comma, next and finish. In this case it's left a trailing space in so we can use the formula we had before =TRIM just to trim that space out of there. I'm going to copy that and paste it as values back where it says first name.

We can use the & symbol from earlier to add them back together = First name & " " & Second name. Fill that down and we have converted Surname, First Name back into First Name Surname.

Review

We covered a lot in that last section.

- We used date formulas like DAY, MONTH, YEAR and WEEKDAY to extract parts of a date.
- We used TRIM, RIGHT and MID to separate out parts of a text string.
- We used & to glue cells back together, remembering that any text we added was in "quotation marks".
- We used paste special, values to take the output of a formula and fix it into a value.
- We used a LOOKUP VLOOKUP to convert one thing into another.
- We saw three different ways to convert numbers stored as text.
- We used an IF function to check whether or not two numbers were the same.
- We saw how to convert text into columns.

Our final section will look at some ways to present our data and share it with others.

1:09 Presenting data



When presenting data there are a few questions you might want to ask yourself to make sure your data is accurate.

- Think about your data sources, are there any assumptions, missing data or caveats you might want to include in supporting information, like you've seen in the supporting data, notes and definitions included in the datasets we've worked with.
- Check your data for errors. We've looked at ways of checking as you go along, and it is worth a final check before sharing that everything looks right and makes sense.
- Make sure you are using a compatible file format that the person you are sharing with can use. This is an example of a safety issue that was caused by not checking for compatible file formats.
- Lastly always have in mind why are you collecting the data and why are you looking at it? Measurement for improvement is all about using the data to help us make changes for the better for our services and patients.



Some of these questions can help you think about your data as you get ready to share it with others:

- What do you see and notice?
- What questions does the data raise for you?
- How do you explain what you see?
- What actions should we take?

Get together your team to look at the data with you and work through these questions together.

Choose the right chart

Choose the right chart to present your data. This website has the most common chart types and what they are best for showing.

https://www.storytellingwithdata.com/chart-guide

If you've been on the Making Data Count course you can also access all the tools we've used on the FutureNHS platform under tools, the ones we've mostly used in our sessions have been the SPC singe chart and SPC multi-chart.

https://future.nhs.uk/MDC/grouphome

Reduce data clutter

Make your charts easy to read by maximising the data to ink ratio. Everything on your chart should have a purpose to make it easier for the reader to understand.

- Add a **title** to your chart explaining what it is showing. Add the data source and if you have any inclusions or exclusions so people know what is included in the dataset.
- If you are using a **legend** show it at the bottom to give the chart area more space.
- Add a **label** to your **axis**. You can either have this as a horizontal title showing information, or a vertical one showing it down. We've added one to the side of our chart and also added one to the horizontal axis. rotating that so we maximise the space for our data.
- You can change the font size to make it easier to read, if you reduce the size enough it will show all dates in your dataset.
- I like to make the line of my data stand out a bit more by taking the **gridlines** and bringing the colour down to a lighter shade of grey. This creates more contrast on screen.
- We have a line chart here, you can add in the dots for the data points by changing the chart type; **format selection** and add in a marker. If you want to replicate what you see on a plot the dots templates you can change a colour of a marker manually, click once to select the whole group then again to select that particular marker, and then change the shape fill to get an orange dot to show an outlier. Of course, the plot the dots tools do this automatically.
- You can **annotate** your chart using a text box.

Checklist for effective data visualisation: <u>https://stephanieevergreen.com/wp-content/uploads/2020/12/EvergreenDataVizChecklist.pdf</u>

1:15 Takeaways



Before you go take a few minutes to think about what you are taking away from this session.

- What tools can you take away from today? Which of things that we've looked at will be most useful to you working with your data?
- What barriers are there to you taking this forward? If you do get stuck check out the Excel help with has information on all the formulas and functions, and if you are really stuck have a search online there's a wealth of information out there and lots of people willing to share their skills and knowledge through blogs and short videos for free.
- Finally build on what's working well for you. What areas are you confident in and what areas could you improve and be even better if. Start small, try learning a new keyboard shortcut today so you don't need to use a mouse, or try a pivot table if you haven't used one before. get stuck in and give it a go!

Further information

Here are some resources you might find useful if you're interested in developing your skills further:

- Al Chen https://www.skillshare.com/user/alchen
- Storytelling with data Cole Nussbaum Knaflic
- #BetterPoster https://www.youtube.com/watch?v=SYk29tnxASs

Thank you!

Thank you for watching. I hope this was helpful and you learned something new. If you were new to Excel I really hope this has given you the confidence to try playing with Excel, trying out some of the things in this video, and if you were more familiar with Excel I hope you've picked up at least one tip or timesaver that will help you in the future.

Feel free to get in touch and let me know what you found useful about this video and anything that could be better: nathalie.delaney@nhs.net

I wish you all the best for your future work using Excel. Goodbye!

Formulas used in this video	Keyboard shortcuts
 =MAX =MIN =AVERAGE =COUNTIF =MID =LEFT =RIGHT =TRIM =DAY =MONTH =YEAR =WEEKDAY = & =IF =VLOOKUP 	 CTRL+N = New workbook CTRL+C = Copy CTRL+V = Paste CTRL+X = Cut CTRL+D = Fill down