Q-STEP ACCESSING DATA 'HOW TO' GUIDES:

USING UK CENSUS DATA FOR SOCIAL SCIENCE RESEARCH

Creator: Patrick English

The UK Census provides a fantastic data resource for social scientists using a 'sample' of the entire UK population to produce an enormous amount of information on social-demographic, cultural, religious, and economic life (among many other topics) in Great Britain and Northern Ireland.

This guide will walk users through the UK Census, what it is used for, and how data can be accessed and analysed for social science research. It is split into the following sections:

1. **What is the UK Census?**
2. **How can the UK Census be useful for social science researchers?**
3. **Where is the best place to access UK Census data?**
4. **Accessing and Downloading UK Census Data using NOMIS**

Users familiar with the UK Census may want to move straight to section three.

This guide will also run users through examples of where UK Census data is used in social science research, including in high-impact areas such as election analysis and prediction and government statistics and reporting.

Q-Step
A step-change in quantitative social science skills
Funded by the Nuffield Foundation, ESRC and HEFCE
Nuffield Foundation
Sheffield Methods Institute.
The University Of Sheffield.

1. What is the UK Census?

The UK Census is a population survey carried out by the government of the United Kingdom every 10 years since the early 1800s. It is sent to every household in the country and its completion is mandatory by law. One person from each household is 'nominated' to fill out information on behalf of everyone living there.

Population censuses are very common across the world and have been for many years. Though only regular since the 19th century, the very first British census was actually carried out by King William I (William the Conqueror) in 1086. In order to raise money for an army to defend his lands from a potential Danish invasion, William wanted a full and thorough assessment of the population, demographics, and wealth of the country to inform a new taxation policy.

William's Royal Officers returned from all corners of the country with their notes and findings. Though sporadic, inconsistent (some officers measured wealth of places in 'goats', others in land), and often abbreviated beyond any recognition, the records were collated and written into the now famous Doomsday book – the very first census database!

(Note: You can read more about the Doomsday book and its fascinating history at http://www.nationalarchives.gov.uk/domesday/discover-domesday/)

The modern UK Census is delivered by post, rather than by royal men on horseback, by the Office for National Statistics. It asks all kinds of questions from basic information such as the age and sex of every member of the household, to detailed questions about the nominated respondents' socio- economic, employment, and health circumstances.

The British government has digitised Census records going back to 1981, with information aggregated to levels as small as postcodes to as large as countries. This gives us an easily accessible yet incredibly in-depth record of demographic, socio-economic, and cultural change in the United Kingdom over the past forty or so years.

2. How can the UK Census be useful for social science researchers?

UK Census data is used for many different purposes and specialist analysis and is often the backbone of some very high impact and important research. Its main advantages are its enormous size, scope, and detail, which allow researchers to conduct investigations and projects on UK Census data which would not be possible with other data sources.

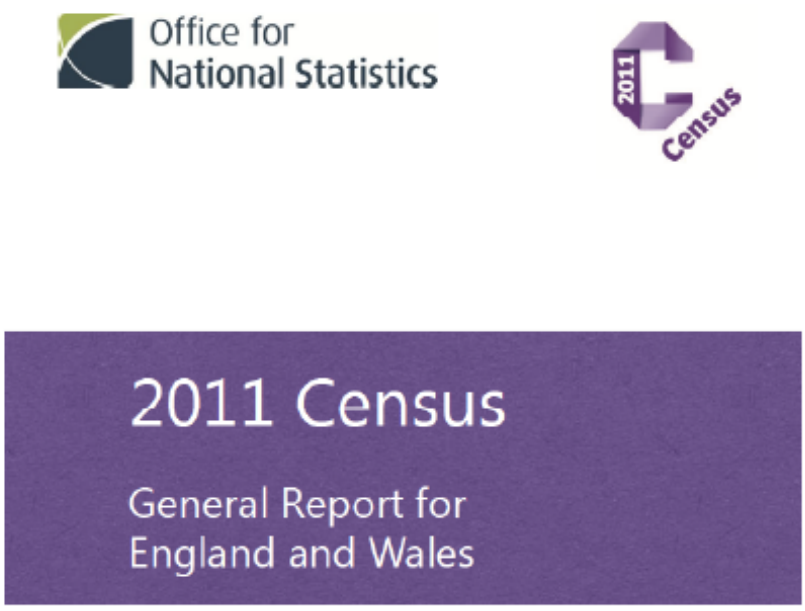
Its comprehensive account of the country in terms of population, demographics, culture, health, religiosity, and other factors also provides interest and accessibility for an enormous range of potential users. In short – there is undoubtedly something in UK Census data for everyone!

Below are some examples of UK Census data in action across the research world.

i. Government reports and population analysis

The most obvious use of UK Census data is for government reports and third- party analysis of population and demographic change across the country.

Examples include the 2011 Census General Report, which was presented to parliament in 2015 to help inform the country's politicians about the society they serve. The report (and its predecessors) and its findings were discussed online, in newspapers, on television, and on the radio all around the country.



Reports such as this cover how the UK population has grown and changed over the 10 years since the last UK Census was conducted. This includes things like ethnicity, religiosity, age structure, social class and industries of employment, among many others. This can be done at many levels of aggregation – from country-wide statistics right down to postcode areas.

ii.BBC elections analysis

For each major election held in Great Britain, a team of psephologists are hired by the BBC to work on the analysis (and sometimes predictions via an Exit Poll) of the results as they come in over the night and into the morning. The team is led by Professor Sir John Curtice, pictured below.

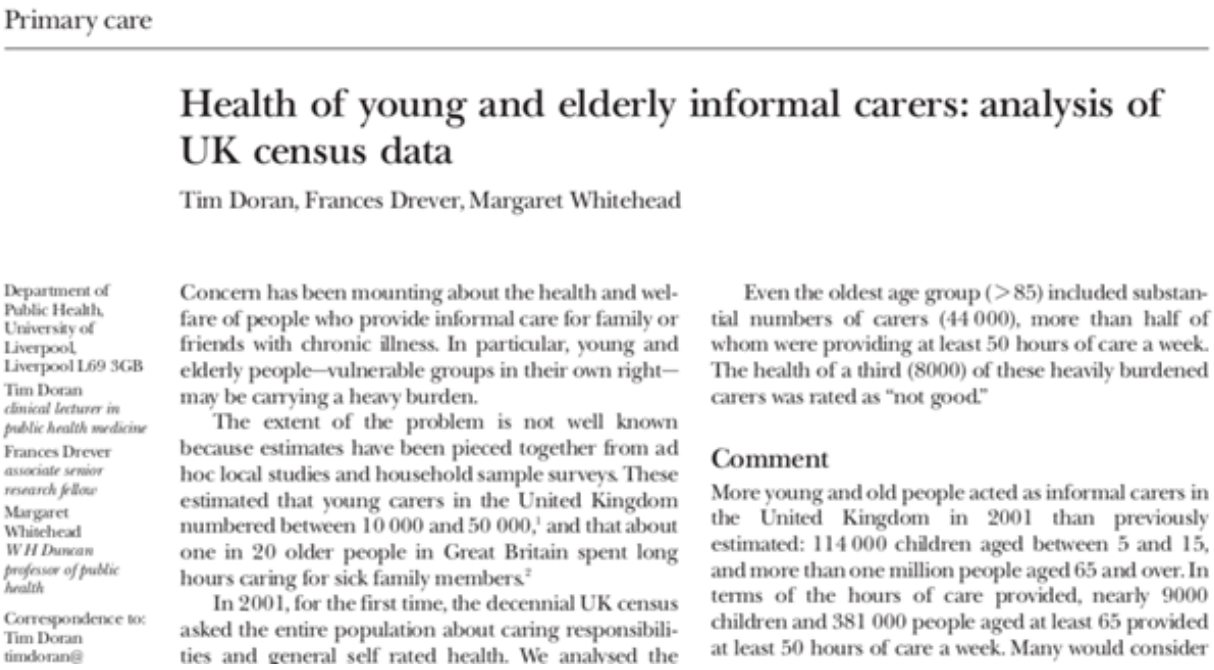
[Professor Sir John Curtice analysing the 2015 British General Election. Picture courtesyof the BBC]

The BBC elections team make extensive use of UK Census data to predict the outcome and narrate the results of each election. They do so by investigating the relationship between results in each area and their socio-demographic profiles according to the UK Census. This allows them to demonstrate how, for example, Labour did much better in areas with many students and young people in the 2017 General Election. Or for another example, how the Conservative advanced strongly in areas with more pensioners in the 2015 General Election.

iii. Health and inequality research

Census information is also used extensively in public health research, as it provides an unparalleled size of data on the lives, habits, subjective health views, and socio-economic conditions of people all around the country.

Examples include the report by Doran, Drever and Whitehead (2003) in the British Medical Journal on the health of informal carers.



Primary care

Health of young and elderly informal carers: analysis of UK census data

Tim Doran, Frances Drever, Margaret Whitehead

Department of Public Health, University of Liverpool, Liverpool L69 3GB
Tim Doran *clinical lecturer in public health medicine*
Frances Drever *associate senior research fellow*
Margaret Whitehead *WH Duncan professor of public health*

Correspondence to: Tim Doran timdoran@

Concern has been mounting about the health and welfare of people who provide informal care for family or friends with chronic illness. In particular, young and elderly people—vulnerable groups in their own right—may be carrying a heavy burden.

The extent of the problem is not well known because estimates have been pieced together from ad hoc local studies and household sample surveys. These estimated that young carers in the United Kingdom numbered between 10 000 and 50 000,[1] and that about one in 20 older people in Great Britain spent long hours caring for sick family members.[2]

In 2001, for the first time, the decennial UK census asked the entire population about caring responsibilities and general self rated health. We analysed the

Even the oldest age group (>85) included substantial numbers of carers (44 000), more than half of whom were providing at least 50 hours of care a week. The health of a third (8000) of these heavily burdened carers was rated as "not good."

Comment

More young and old people acted as informal carers in the United Kingdom in 2001 than previously estimated: 114 000 children aged between 5 and 15, and more than one million people aged 65 and over. In terms of the hours of care provided, nearly 9000 children and 381 000 people aged at least 65 provided at least 50 hours of care a week. Many would consider

Their report demonstrated how of the nearly 6 million people providing informal care for another person (usually in their household), only 56% described themselves as in 'good health' versus 70% of people who were not providing any informal care.

iv. Methodological publications

UK Census data is used in a range of academic work – including the above on pubic health – but is also very useful for methodological studies. This is because the sample it provides is so large and comprehensive, making it possible to run experiments and tests on methodological approaches and assumptions which it often not possible (or reliable) to do on survey or polling data (for example).

The example below is from Tranmer and Steel (2001) in the Environment and Planning journal.

DOI:10.1068/a3317

Ignoring a level in a multilevel model: evidence from UK census data†

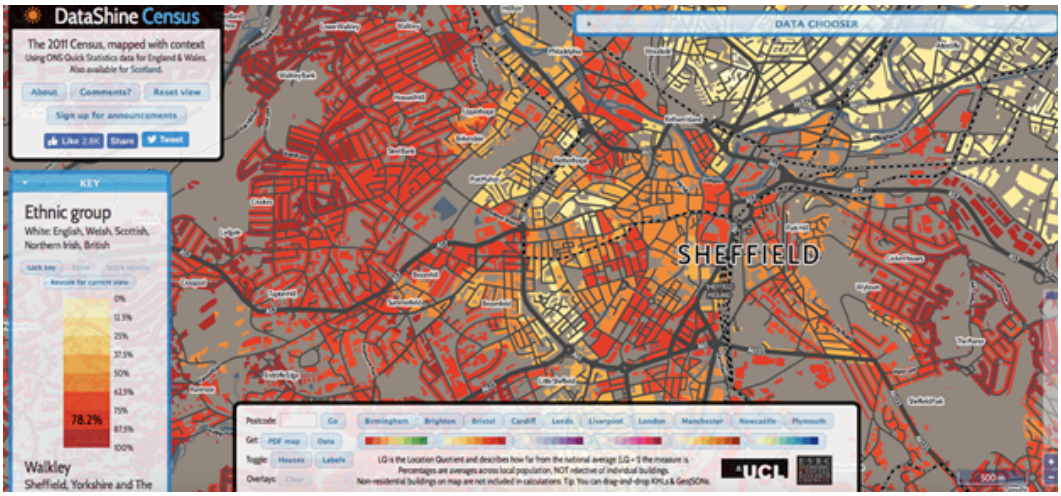
Mark Tranmer
Centre for Census and Survey Research, University of Manchester, Manchester M13 9PL, England; e-mail: mark.tranmer@man.ac.uk
David G Steel
School of Mathematics and Applied Statistics, University of Wollongong, Wollongong, NSW 2522, Australia; e-mail: David_Steel@uow.edu.au
Received 14 February 2000; in revised form 5 February 2001

Abstract. Because of the inherent multilevel nature of census data, it is often appropriate to use multilevel models to investigate relationships between census variables. For a local population, the data available from the census allow a three-level nested model to be assumed, with an individual level (level 1), an enumeration district (ED) level (level 2), and a ward level (level 3). The consequences of ignoring one of the three levels in this model are assessed here theoretically. Empirical results, based on 1991 UK Census data, are also provided, comparing the variance components estimated from the three-level model with analyses based on models where the ED or ward level are ignored. The results show how the variation that occurs at the level not included in the models is redistributed to the other levels that the models do include.

They experimented with ignoring (missing out) levels in multilevel models. Their findings suggested that, in this particular example at least, variation from omitted levels is redistributed to other levels. This implies, they argued, that effects estimated at a particular level may include contributions from other, non- estimated levels, which ought to be routinely accounted for when interpreting analysis of effect at each given level.

v. Visualisation – DataShine

The DataShine mapping platform was developed as part of the ESRC-funded project "Big Open Data: Mining and Synthesis (BODMAS)". The aim was to promote the use of big data among social scientists by providing a platform where large and open datasets could be visualised and investigated in incredible detail, focusing on developing and hosting interactive maps. The first phase of the project used UK Census data from 2011. An example below shows the DataShine map covering ethnicity in Sheffield according to the 2011 UK Census data.



DataShine allows researchers to load a range of different variables onto a fully interactive, post-code level map of UK Census Data.

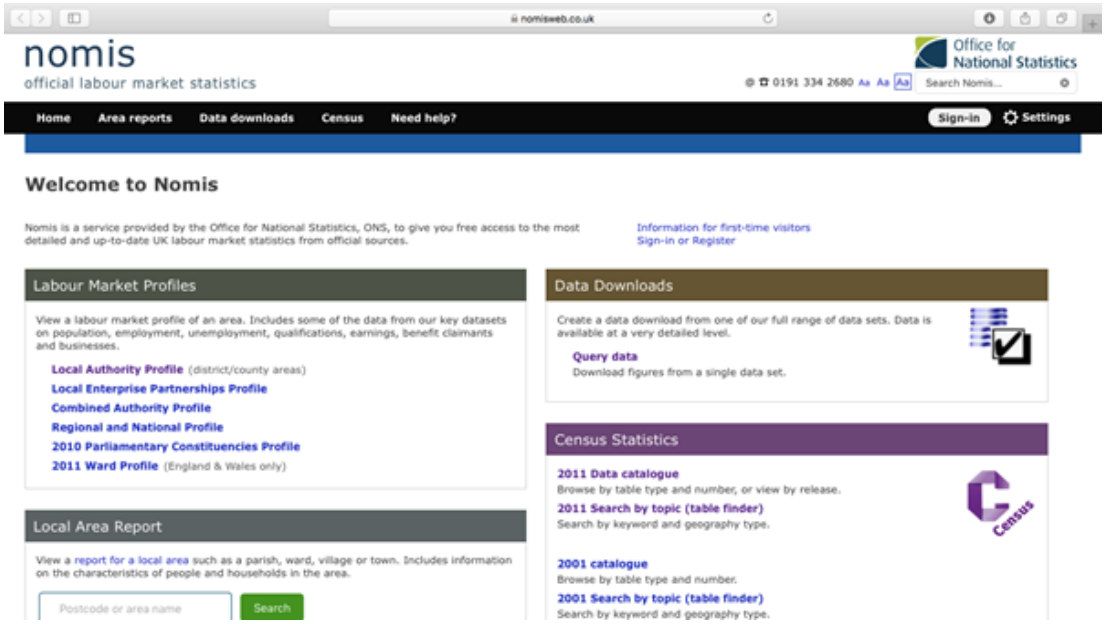
Users can move around the country easily, and helpful graphics explain and articulate the scales and underlying data used to generate the maps. Other data can also be loaded onto the maps including commuter data, bike-route data, election results, and consumer data.

3. Where is the best place to access UK Census data?

Now that we know what the UK Census is and how useful it can be to social science researchers, we need to be able access and download the data for our own purposes. The Office for National Statistics provides easy, open, and free access to Census data. The quickest and easiest way to access UK Census data is through the NOMIS portal: https://www.nomisweb.co.uk

NOMIS is run by the Office of National Statistics themselves, and they use it to publish a wide variety of official statistics – not just UK Census data. NOMIS provides easily selectable and downloadable formats for official data with very little restrictions in terms of what you can do with it (just be sure to reference the service and the databases).

Upon visiting the NOMIS website you will be greeted with a home page looking like this:

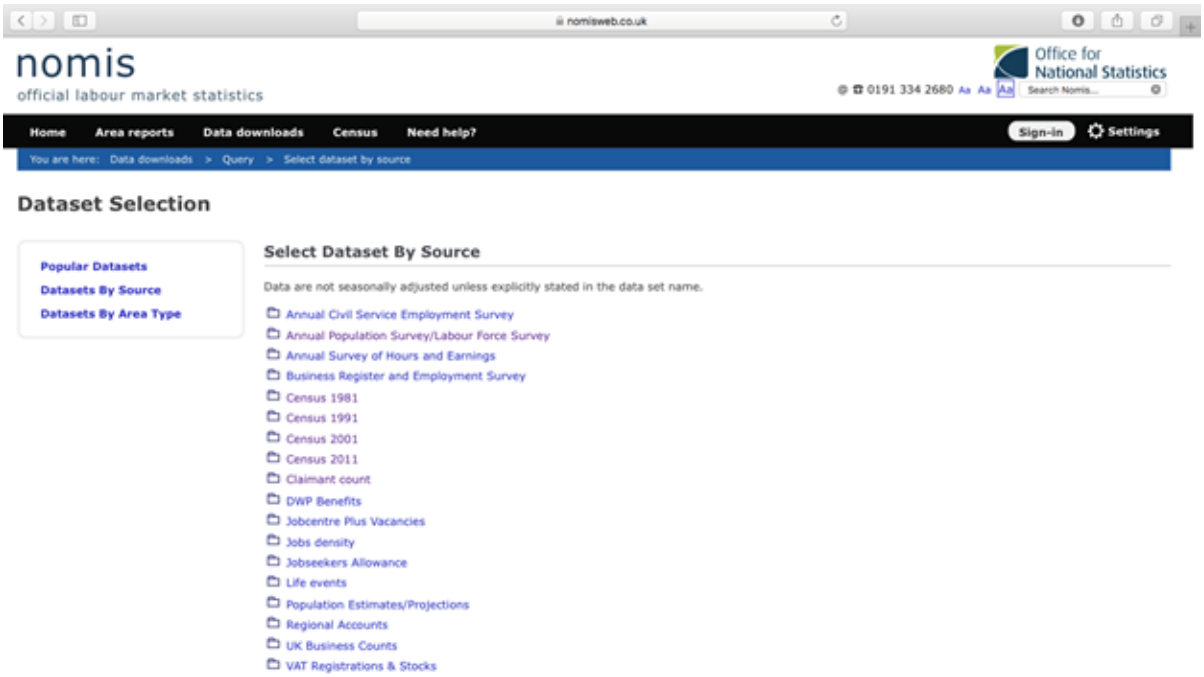


There are a range of guides and further information on the service itself available on the 'Information for first-time visitors' page. Do explore everything when you get the chance – there is so much more very useful and detailed information hosted on this site than just Census data.

4. Accessing and Downloading UK Census Data using NOMIS

NOMIS offers hotlines straight to the 2011 and 2001 UK Census data catalogues on the home page see above. However, making use of the data query function allows for a much cleaner and quicker access (and will get you used to using the client if and when you want to access other kinds of data).

Select 'Query data' from the page as shown above. This will bring you to a page listing the datasets available on NOMIS by source. As you can see, there are quite some options available from the portal, which makes it a fantastic resource!

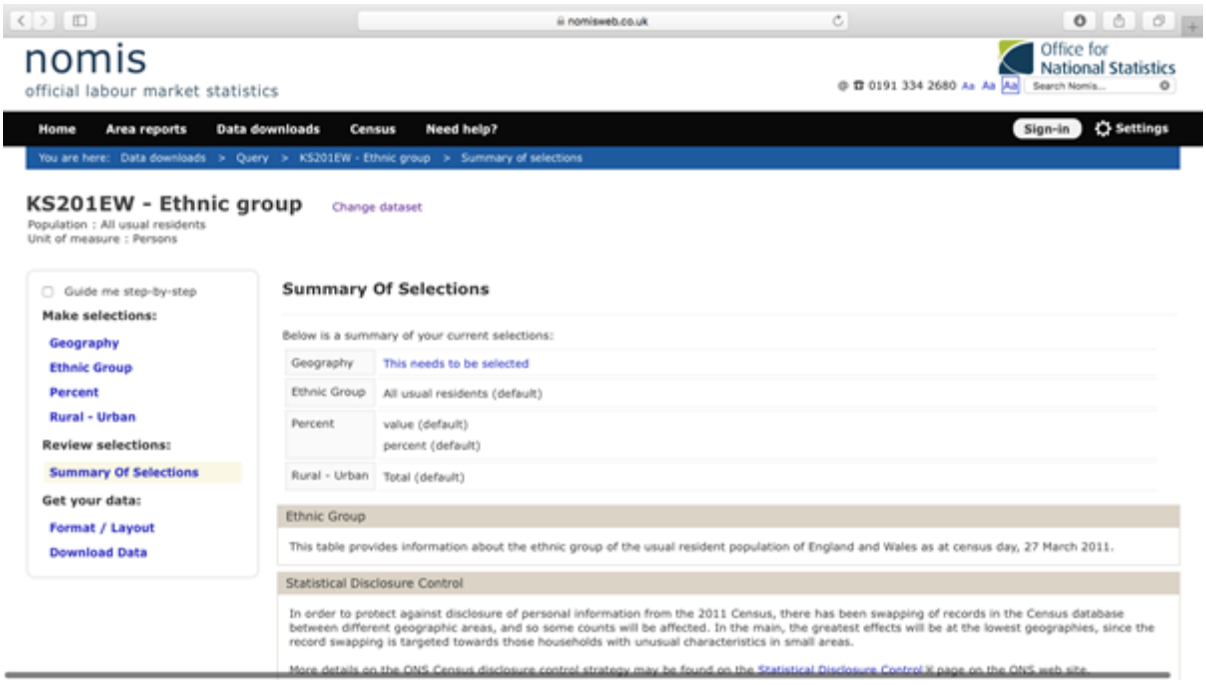


Select 'Census 2011' to see a drop-down menu of different types of statistics available to you from NOMIS from the 2011 UK Census. Most often, we will be using one of either 'Key Statistics' or 'Quick Statistics'. These are the most flexible databases which are also available at pretty much every aggregated level that a researcher could want. For now, select 'Key Statistics'.

We will use the ethnicity data for this example guide, so go ahead and select 'KS201EW – Ethnic Group' from the listing. The code is the unique identifier for the database (worth noting down for any correspondence or referencing you might need to do). The 'UK Key Statistics' contains the same information as the EW files (England and Wales), but it is also harmonised with Scottish Census data.

Responsibility for the Scottish and Northern Irish Censuses have been devolved to the respective regional parliaments, and there are some issues of comparability between the English and Welsh and Scottish data from 2001 and 2011. This is due to changes in questions, definitions and response categories made by the devolved administrations. This can make comparing between Census data from different UK regions quite tricky, but between using the NOMIS comparable datasets and the Census Comparability Report most issues can be ironed out.

Returning to our example, selecting the ethnicity database will bring you to the following page:

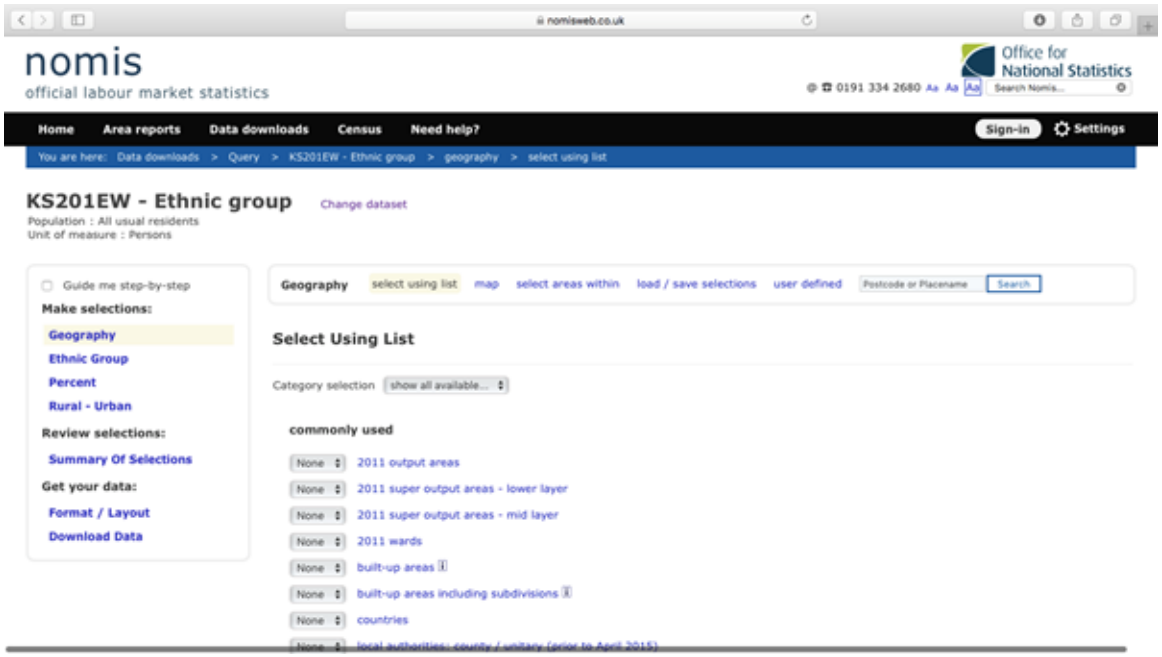


This page gives us a brief overview of the database selected – in our case, "information about the ethnic group of the usual resident population of England and Wales at census day, 27th March 2011". So, from this database we are able to look at the exact ethnic structure of the British population (every usual resident) as it was in spring 2011.

Alongside the left-hand panel of this page is a menu of selections. This is the main user interface through which we will select and eventually download our data for analysis. The first thing which needs determining is the geography.

i. Geography

Select 'Geography' from the menu, and the following page appears:



This is the same Geography page which will be presented to users for all UK Census data queries. It is asking us to select the aggregated level to which we want the data to be reported. There is a very diverse range of possibilities, which is what makes the NOMIS portal so useful. If we were interested in the ethnic breakdown of the country by local council, we would select one of the 'local authorities' options (depending on whether we were interested in rural – county/unitary – or urban – district – councils). Also available is aggregation to post codes, local government wards, government regions, and many more – including specialised Census Output Areas (note: These are useful for more specialist analyses – in particular those analyses requiring a different level of aggregation to what is currently already available on NOMIS. A number of resources exist which allow you to map Census output areas to other types of boundaries. You can read more on the output areas and how to use them here and also here).

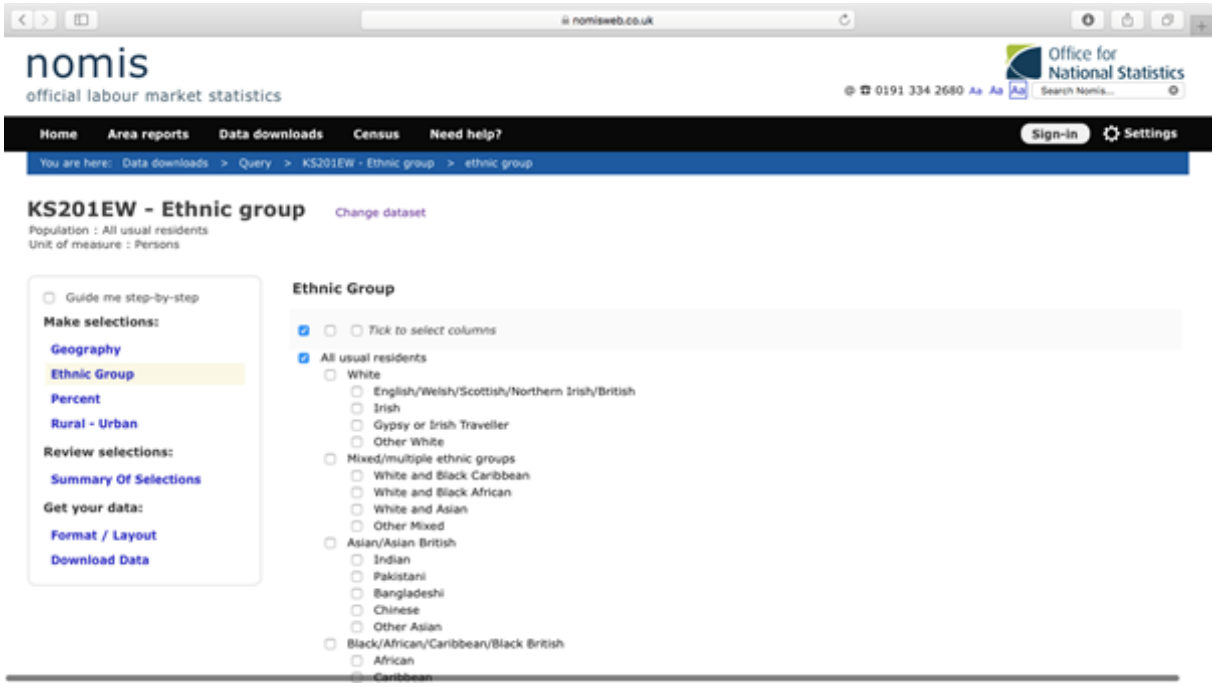
Let's assume that we are interested in the ethnic make-up of parliamentary constituencies in England and Wales. We might be interested in examining the relationship between party vote share and ethnicity, for instance. Or maybe we want to construct a density plot showing how high-ethnic minority constituencies are clustered around the country.

To pick a type of area, we select either 'some' (if we are interested in getting this information for only a few areas) or 'all' from the drop-down menu beside each respective locality. 'Parliamentary constituencies 2010' (this simply means that constituencies are reported as they were defined by the 2010 boundary review and used in each subsequent General Election) is second from bottom on the 'commonly used' list.

Generally speaking, it is advisable to download information for 'all' locations within your preferred type of area; it is always possible to pull out' the ones you are interested in from the larger file, and you will almost certainly want the same information for other areas at a later date. With 'all' downloaded, you have each possible location already to hand (and so you do not have to go through the selection and download process all over again).

ii. Ethnic Group (variable selection)

With the geography selected (parliamentary constituencies), the next stage is to select the variable(s) of interest. Selecting 'Ethnic Group' brings up the following screen:

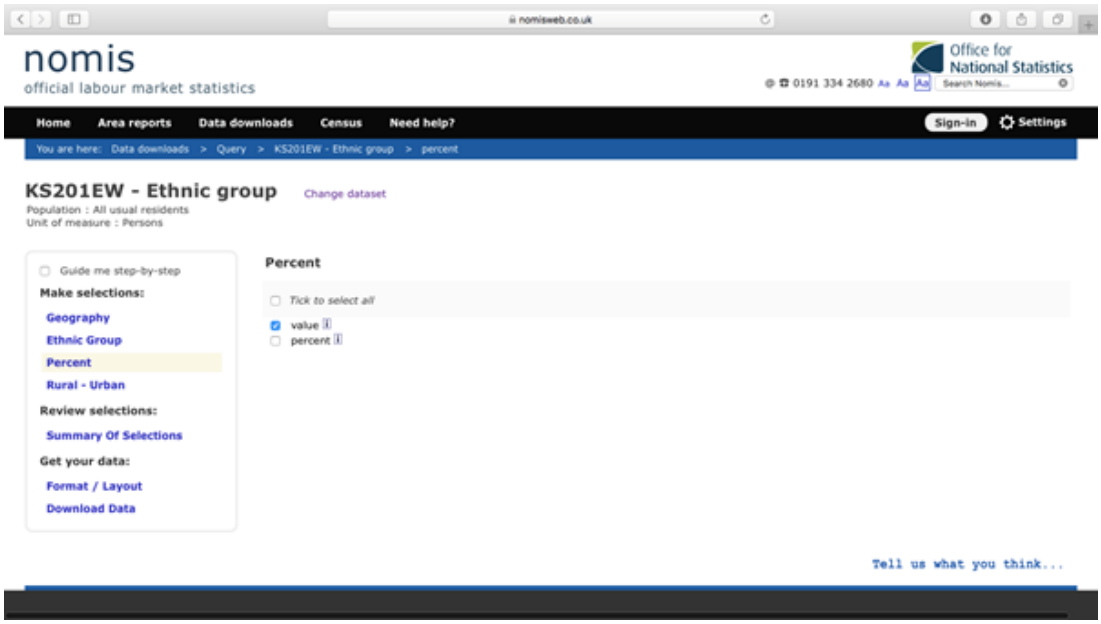


In the UK Census, respondents are asked to classify themselves according to one of nearly 20 ethnic categories. Of these, white can be aggregate to 'White', four to 'Mixed', five to 'Asian/Asian British', and there are three categories of Black ethnic identification and two further categories of 'Other'.

Let's say we are interested in the presence of ethnic Asian/Asian British people in English and Welsh constituencies – regardless of their specific ethnic identification according to the five sub-groups of Asian/Asian British identity. Select 'Asian' to isolate this particular set of data. Be sure to keep the 'All usual residents' box checked as a baseline – this is a good habit to get into, especially if you want to use count data.

iii. Percent

The NOMIS interface then asks us whether we are interested in count data or proportional data (or indeed both). Selecting 'Percent' brings up the following screen:



While count data can of course be useful – for example if we wanted to add up and generate proportions to aggregations at different levels than currently available on NOMIS – it is highly advisable to use proportional data when comparing and contrasting between units (in our case, 2010 parliamentary constituencies) and over time. This is because proportional data tells us something about the *relative size* of a characteristic, group, or any other type of cluster within a *broader population*, whereas count data only provides us with the raw, *absolute* figures.

The danger with using raw, absolute figures for comparison is that it can cause disingenuous and misleading results and interpretations. For example, 10,000 Asian/Asian British people in Constituency A may by a bigger population than the 7,500 in Constituency B, but if Constituency A contains 120,000 people while

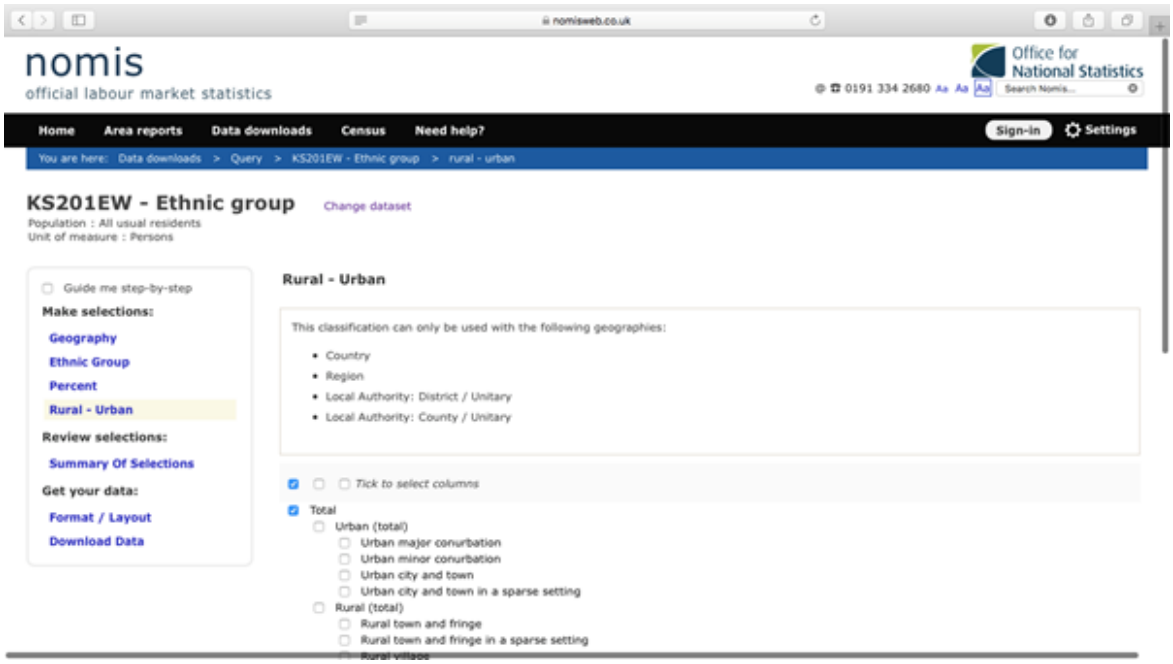
Constituency B contains only 70,000, then the *relative size* of the Asian/Asian British population is much bigger in Constituency B and so they will have a much greater *relative impact* in terms of voting and *relative visibility* on life in Constituency B than in Constituency A.

Equally, without a sense of proportion then changes over time can also be misinterpreted. For instance, let's imagine that from the 2001 Census to the 2011 Census there is a change of minus 500 Asian/Asian British people in Constituency A. This does constitute an *absolute decline* in the Asian/Asian British population, but not necessarily a *relative decline*. Let's also imagine that the overall population Constituency A declines by 30,000 people from 2001 to 2011. If there are 500 less Asian/Asian British people in Constituency A, but there are 30,000 less people overall, then the *relative size* of the Asian/Asian British population in Constituency A goes from 7% in 2001 to 8.3% in 2011 – a *relative increase* of almost 1.5%. So, while the absolute numbers are smaller, given the overall population change we can expect that the smaller Asian/British Asian population to actually have relatively *more impact* and *increased visibility* in 2011 than 2001.

With this in mind, select 'percent' and untick 'value' from the options shown on the screen.

iv. Rural – Urban

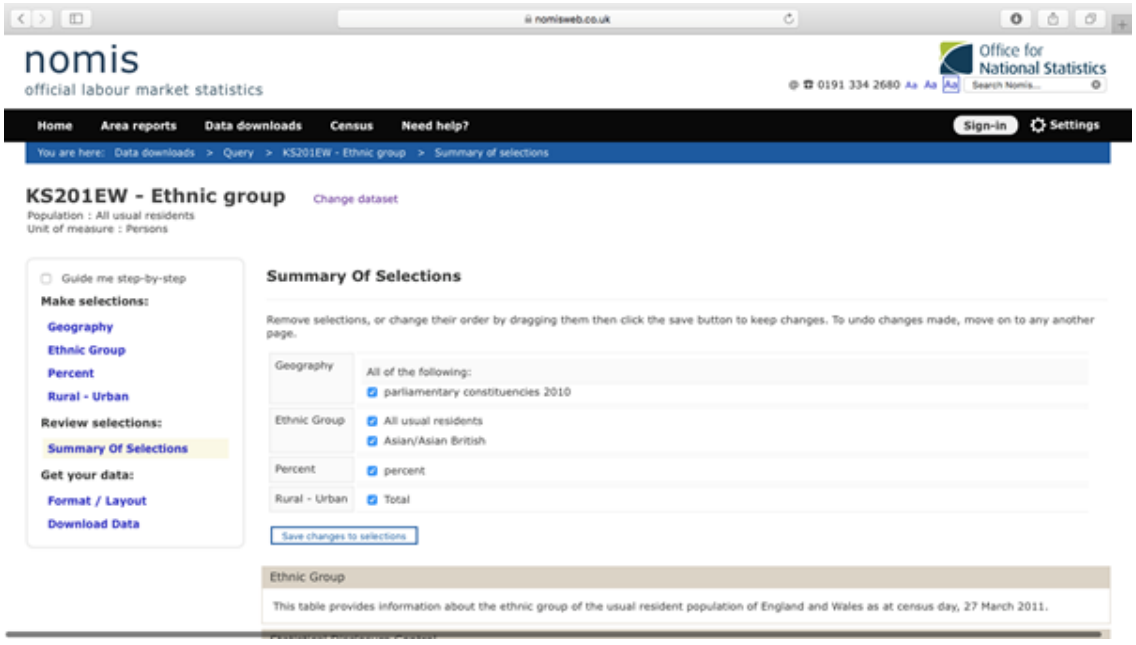
NOMIS also allows you in some circumstances to create separate figures for urban versus rural areas. When selecting 'Rural – Urban' from the panel on the left, the following screen appears:



As we can see, the rural/urban split is only available if we are looking at countries, regions, or local authorities. Given that we are researching ethnicity in parliamentary constituencies, we can bypass this option and proceed directly to the summary of selections.

v. Summary of Selections

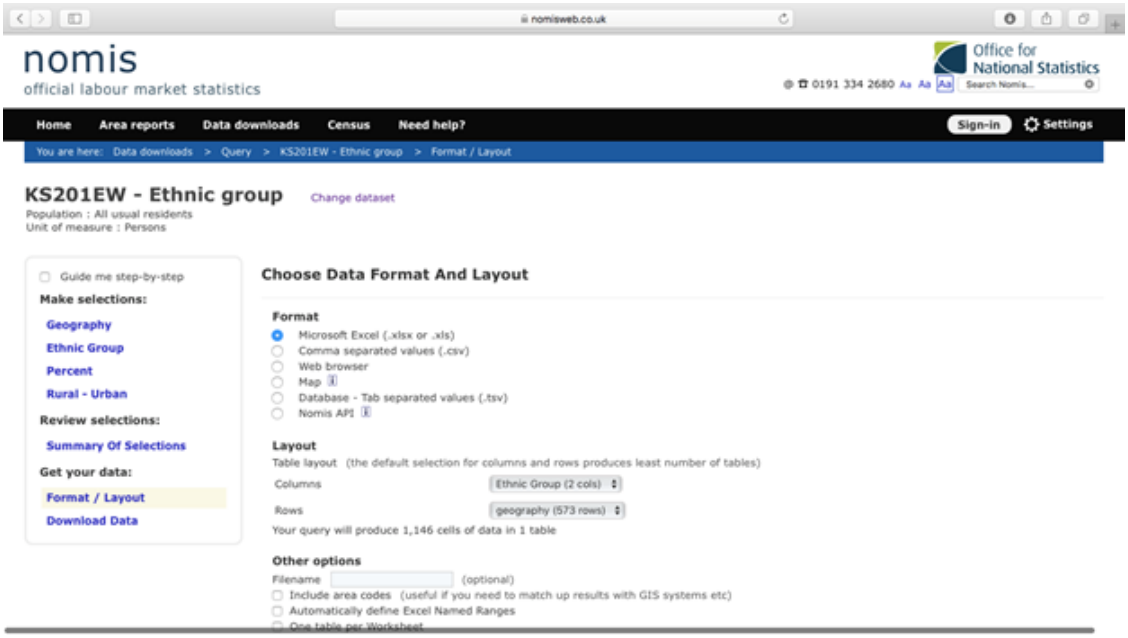
If the guide has been followed to the letter, then the 'Summary of Selections' screen should look like this:



This screen is useful for providing a quick check that we have made all the selections properly. Once we are happy, we can move to formatting the data and then downloading it.

i.Format / Layout

Selecting the 'Format / Layout' link will bring up the following screen(providing all the selections have been consistent with this guide):



NOMIS offers six different formats for data download, but the most often used will be the Excel worksheet (.xls), comma separated text (.csv) file, or Map (which uses the 'Leaflet' package in R to generate interactive browser-based maps).

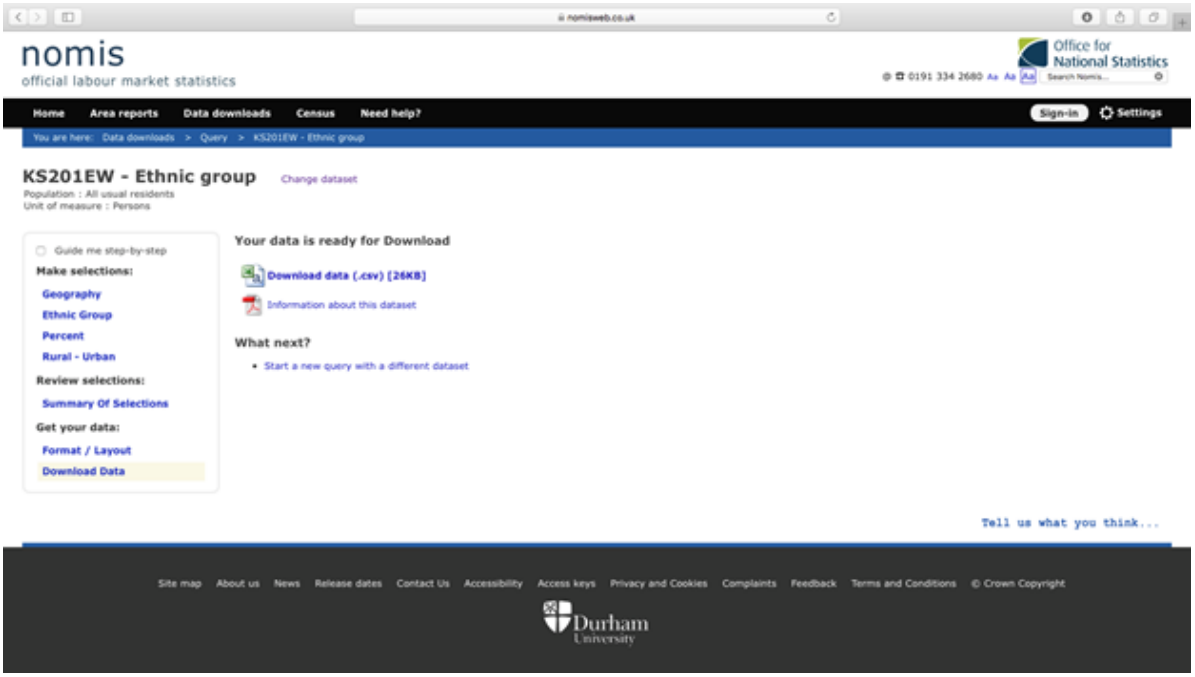
Comma separated files are the most flexible and easily-readable formats (accessible to Excel users and users of statistical programmes such as STATA, SPSS, and R). They also avoid compatibility issues sometimes raised with Excel files. Indeed, after the selected number of variables and locations generates over a certain number of cells, Excel downloads are not possible anyway. Select 'Comma separated values (.csv)' from the options presented.

It is possible at this stage to alter the layout of the selections, moving the geography (parliamentary constituencies) from the top of the database (rows) to the side (columns), and the selected variable (percent Asian/British Asian ethnicity) vice-versa. Most often, this will not be desirable (particularly if we want to combine this data with other types and sources).

Further options allow us to 'Include area codes', 'Include row numbers', and to place the 'Row descriptions on the right'. The most useful of these options are the area codes, which allow users to merge data with other sources which also use (or can present data according to) the Office for National Statistics coding system. Most notably, this function allows you to combine UK Census data files together. For example, we could merge our ethnicity data with 2010 parliamentary constituency data on age, gender, religion, and anything else that might be of use. For now, tick only the 'Include area codes' checkbox.

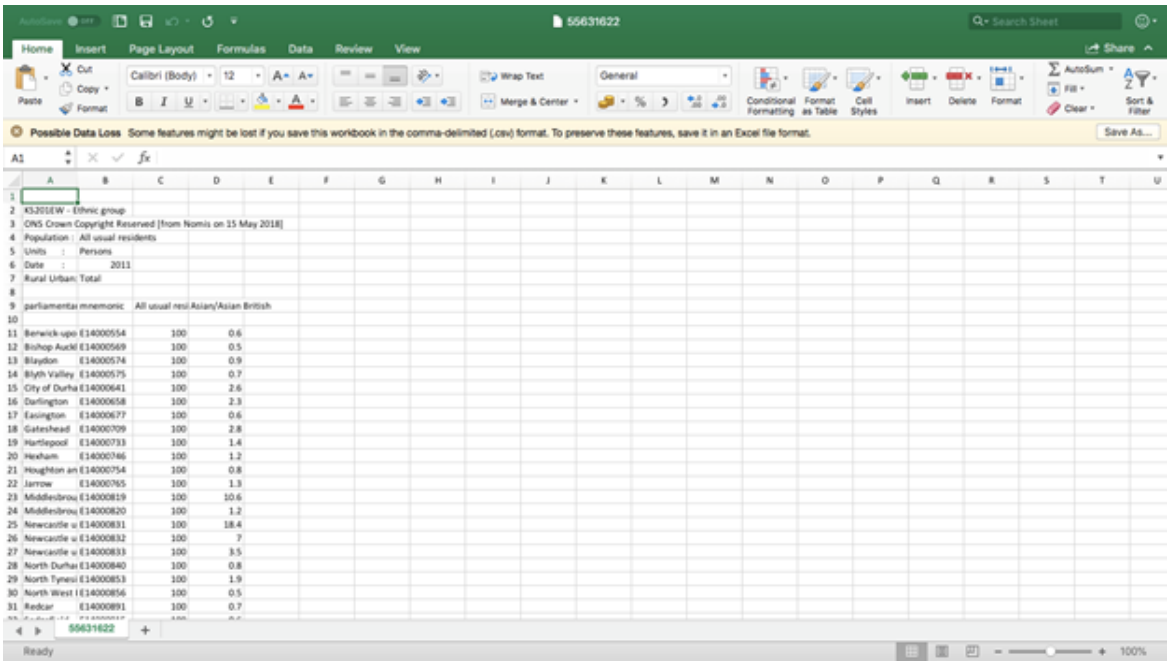
vii. Download Data

With the geography, ethnic group (variables), percent, rural – urban, and format / layout all completed, we are ready to download the data. Click 'download data' and you will, after a short wait while the request is processed, be presented with the following screen:



Simply left-click on the data to begin the download. There is also a PDF file which accompanies every user-generated data file. This contains the metadata associated with the database used to generate the data file. This is useful to store alongside the data for future reference.

If everything has worked as it should, the data file generated from the NOMISinterface should look something like this:



The first column contains the name of each 2010 parliamentary constituency in England and Wales, while the second column shows the area codes – unique to each constituency. The next column shows our 'total' (which is of course 100 when using proportional data), while the fourth column (column D) shows us the percentage of each constituency population who identified as Asian or British Asian in the 2011 UK Census.

We can sort this data (descending) to see where the largest Asian/British Asian populations were in 2011 (by constituency). We can see that in Leicester East and Ilford South, nearly 60% of the population identified as Asian/British Asian. Over half of all usual residents in the London constituencies of East Ham, Brent North, and Ealing Southall had a similar identity, while Bradford West is the only constituency in the North of England to surpass that 50% barrier.

Sorting in the opposite direction (ascending) shows us the 2010 parliamentary constituencies in England and Wales with the fewest Asian/British Asian self- identifiers. We can see that in Montgomeryshire, less than 0.5% of the population held this identity, making it the least Asian/British Asian constituency in England and Wales. In all, a full 94 constituencies – including nearly half of all the constituencies in Wales – had less than 1% of their population declaring such an identity in the 2011 UK Census.

And there we have it – how to find, access, and download UK Census data!

REFERENCES

Doran T, Drever F, Whitehead M. 2003. Health of young and elderly informal carers: analysis of UK census data. *BMJ,* 327:1388. doi:10.1136/bmj.327.7428.1388

Tranmer M, Steel DG. 2001. Ignoring a Level in a Multilevel Model: Evidence from UK Census Data. *Environment and Planning A: Economy and Space*, 33(5):941-948. doi:10.1068/a3317