Exploring learning city inclusion with diverse education data

Transcript from webinar video recording

[Muir Houston] Hi everyone, welcome to this session on exploring learning city inclusion.

I would like to remind participants that this session is being recorded and will be uploaded on our UBDC website as an accessible format for those who could not attend the session, or if you want to have another look at it. Further details will be provided on the UBDC website, but you may wish to
adjust your name as it appears on the screen. Cameras will be turned off and microphones muted to aid privacy and also for bandwidth reasons. Feel free to introduce yourself in the chat as Catherine's been saying but please do not include any personal information such as phone numbers or emails.

But please do say hi and check out our website for further webinars and other resources including how to access data and other training and events delivered by UBDC.

Please use the Q&A facility to ask questions. These will be collated, and responses will be provided in the Q&A session. Session structure - Catherine will provide a presentation, Rachel will do a demonstration of the data and the statistical analysis and then we will hold our Q&A session.
So, I'll hand over to Catherine and she'll say a little bit about the presentation to come. So, thank you Catherine. [Catherine Lido] Thank you so much.

Muir for that lovely introduction. Let me see if I can just get my screen share going.

So, welcome everybody from all over the world. Super nice to see you all and have you here.

I am, as Muir said, Professor Catherine Lido and I'm ably assisted by Rachel Cassar who is a research assistant for our VisNET project. I'll just launch right in, but any questions along the way you just go ahead and pop them in the Q&A box. So, we're both psychologists and you might be thinking
what can psychologists tell us about education and learning inclusion? Aren’t psychologists people who

make you sit on a couch and talk about your mother? [laughs] Maybe some psychologists do, but we’re research

psychologists and psychology as a discipline is interested in people and how they think, act and react. We come from a social psychology background, which is not just how people act on their own but how we act in groups. And learning, adult learning, is a really powerful way to look at humans and humans in their context. So, this research is about life-wide learning inclusion and how we can learn more using some of the skills that psychologists have in terms of quantitative data, but also skills
that come from urban planning and other disciplines to try and maybe get more

holistic pictures. You can go ahead and join me on LinkedIn. I'm on Facebook, I'm on Twitter. I might

whiz through these slides, but they'll be made available to you afterwards, so you don't need

to read through them as I go. I just wanted to give you a little bit of the context of my background.

I work from particular frameworks like social identity theory and Bourdieusian notions of social

and cultural capital. But I've come to this newer concept of symphonic social sciences,

which is about using novel data - I hate to say big data because it's not always large in number,

but it might be quite realistic data - to try and blur the boundaries between qualitative and
quantitative and instead come up with stories that can really help us improve the lives of people around us. And so, part of that is my work here with Urban Big Data Centre. So, if you want to read more about this, I was in The Psychologist magazine on the cover talking about big data in the big city.

But all of this is couched, this whole webinar series is offered in the wider scope of Urban Big Data Centre as a research centre, which is basically a free resource for you all to use to try and get people to use more big and novel data and research methods to try and tackle some of these big social and environmental crises we're facing today. So, thank you so much for joining us. Just by
being here you’re really supporting our work and helping us get the word out and the message out

00:04:43,440 --> 00:04:47,280
that we want people to access our data. And if you’re particularly

00:04:47,280 --> 00:04:53,360
interested in educational disadvantage please get in touch with one of our team. You met

00:04:54,000 --> 00:04:59,360
Muir Houston there but we’ve also got Professor Mike Osborne, Professor Keith Kintrea,

00:04:59,360 --> 00:05:04,640
Dr Phil Mason and a couple of wonderful PhDs - Brittney Nathaniel and Barry Black.

00:05:04,640 --> 00:05:11,280
Who are all doing amazing work in the area of how keeping people actively engaged in learning

00:05:11,280 --> 00:05:18,960
is better for our health, for our occupational outcomes, for engaging citizens and for actually

00:05:18,960 --> 00:05:23,840
having happy healthier citizens who live longer. So, if you’re interested in any of these research
areas such as place-based inequalities in schools’ attainment, further education and

whether it's fulfilling its purpose, levels of inclusion in our higher education universities

and more widely in inclusive learning cities, these are all different strands that we have

researchers working on to kind of look at this pipeline through the learning experience.

Which brings us to the learning cities agenda. I realise from the attendees that some people

who are here are probably unfamiliar with the learning cities agenda and are maybe just more

interested in data and research methods. But if you're not familiar with it, it's really quite

an interesting alternative to the smart future cities idea, which says yes we need smart cities,
yes we need future cities but we have an aging population, more and more people living in cities.

We really need to meet these challenges to social inclusion, technology, knowledge economy, diversity, and sustainability. And in order to do that we probably need to keep people engaged and upskilled and working longer towards transforming learning cities. So, it's really an idea of engaging all citizens across all ages. And we do a lot of that work with the PASCAL International Observatory in our School of Education, of which I'm the Deputy Director for the European Hub. So, if you're interested in learning more about PASCAL
do see the Glasgow website. I know we have some PASCAL members here, welcome welcome.

And I know Mike Osborne would be very happy to hear from you.

So, the learning cities framework comes from UNESCO and this is from the Beijing publication in 2013.

It's available online if you want to learn more. But you can see how it sort of has what's fundamental to be in place before the learning city vision is enacted and that includes resources, governance participation and political will. And then the columns are really some of the things that we can measure in terms of do we have inclusive participation from early years to adult learning and so on. What about mobilising our online learning technologies?
What about having a vibrant culture of learning through our lives and through our cities? And so this document actually came up with 42 indicators against which you could benchmark a city's sort of success as a learning city. And this is just an example, so you might be able to measure empowering individuals and promoting social cohesion say through civic participation like volunteering and voting and so on. You might be able to assess inclusive learning just by looking at diversity and demographics pre-nursery, pre-kindergarten, provision and so on, funded adult learning. And you might be able to examine something like literacies and not just reading, we'll talk about other types of literacies today.
So, I'll just quickly operationalise a couple of key terms before we launch into the present project.

You've probably heard of lifelong learning. If you haven't it's unhappy tagline is "cradle to grave", which I feel is incredibly depressing. So, I tend to think more about lifewide learning, which is this little box here, that formal learning leading to a qualification is such a small part of the learning that we do every day. I mean, you're all here engaging with us because you want to learn, and you might have different motivations for being here but at the heart of it is you've all stepped out of the formal learning environment to perhaps, I would call this
maybe a non-formal learning environment. It is structured. I apologise that I’m speaking

at you. It’s not the ideal conditions for knowledge exchange. But you’re moving into

the less formal realms of knowledge exchange and that is where it gets harder to measure.

So today I'll be talking to you about how maybe we can measure learning cities and these

less tangible lifewide literacies like health literacy, financial literacy, eco literacy and

digital literacy. So, my final definition comes from UNESCO in that literacy is not seen

by this framework as just reading. It's the ability to identify, understand, interpret,

create, communicate and compute using printed and written materials associated with varying contexts.
Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential and to participate fully in their community and wider society. You can see that there's something sort of empowering or enabling about these less formal types of literacies and learning. So that is the entire background and framework that brought us to the Integrated Multimedia City Data project, which I'm talking about today. The iMCD, or i mc d as I like to call it, is the first data product that was created at Urban Big Data Centre as a free open data resource. Anybody can apply to use it, and anybody can apply for any one of the strands.
The first strand is a 1,500-household survey. The second strand is that we followed up about 600
or so participants to wear GPS trackers around the city for one week so we know what they're
doing, where they're going. I'll talk to you about the ethics of all of this maybe just very briefly
but that's probably another talk. Lifelogging cameras they wore around their neck, that was about
300 or so participants and it took pictures every few seconds as they went about their
daily life for 48 hours. And lastly, we had one year's worth of social media capture.
Largely Twitter data but also traffic camera feeds, ScotRail feeds and so on.
So, I'll just hone into the survey first, but just to say that we did treat the ethics of this
very carefully. Not all of that data is equally open. Some of the data is obviously much more sensitive and so much more limited in terms of access. But the Understanding Glasgow survey is very open. We have an open version that you can apply to use and all you really need to do is go on to the website, fill in like a very brief form and you can have this data to look at yourself.

So today we're going to introduce you to the survey data. Rachel will take you through a demo and the hope is that you might access it yourself to practice the types of skills that we're talking about or to use it in your classes as a data example or to use it for your own personal research because it's really an under tapped resource. There's still plenty in here.
that’s never been looked at. So, what did we do for the survey? Well, we reviewed national and
EU surveys as well as these UNESCO indicators to try to measure attitudes (what do you feel about
something), behaviours (what do you do about something), and literacies (what do you know about
something) in the domains of education, sustainability, transport, cultural and civic engagement, and
ICT/technology.

This was delivered door-to-door via a company called Ipsos MORI. And so, here’s some of the
little statistics that we have here. 2,095 adults took part. Age range 16 to 102, average age near
50. So it’s quite good for looking at adult learning and older adult learning because we
have a range up to 102. And it's roughly, although it's the Greater Glasgow area - that's

eight local authorities - it's broadly representative of Scotland in terms of all of the demographics we collected. So, it's representative of census data.

The education questions, Rachel will talk us through some examples using them, but we tried to capture formal learning leading to qualifications, any formal learning engaged in in the last 12 months, any informal learning engaged in in the last 12 months that's structured but not leading to a formal qualification and then all the other types of learning that goes on. Sorry I swapped those, it's not my fault it's the way they are. Non-formal is structure but not leading to
a qualification and informal is all other self-learning. I've got this lovely little pyramid here if you get confused about informal and formal learning. And also, family learning, which is helping anybody in your family to learn in the last 12 months and whether they were older or younger than you - a variable which nobody has analysed yet. So I'll talk you through other things but we have if they were engaged in learning, how? - online or face-to-face, how many hours, motivation for study. And this is just a list of some of the national measures we reviewed before we took Potter. Ok, so from this, even though we collected the data interestingly with no research questions and no aims in mind and that was really really tricky - I can talk you through that in another talk.
I think the first pilot of the survey took six hours. I didn't do it all at once but we piloted bits of it and we added it all up and it was six hours’ worth of questions that all the different researchers wanted represented in this survey. We got it down to about, I think it was, 45 minutes. And so, from that we then engaged in our own research and this is a publication we did about older learner engagement in the modern city. And it looked particularly at older adults, which was actually 60 plus not 65 plus. And you can read about in our paper that you can see that as you aged learning engagement drops basically. So
our older adults were engaged in less forms of all types of learning, which is not surprising

except for a little blip with middle age adults and family learning. But we found other predictors

that mattered in terms of overall learning engagement such as feeling safe, belonging

to your area and even the local authority that you lived in. More importantly we found that

those older adults who were engaged though tended to be more highly engaged in everything

including

activism, boycotting things online, more reported better overall general health

and they were actually more physically mobile in and around their cities. You can see them

in the little green here that’s the learning engaged older adults 60 plus compared to
the orange. So those are those two graphs, and you can read about that in the paper.

And then from this we did a second paper which was targeting a little bit more into lifewide literacies. So, here's just some examples that financial literacy had three questions taken from Lusardi and Mitchell. They have a standardised short form financial literacy. Suppose you had 100 pounds in a savings account and the interest rate was two percent per year, after five years how much do you think you would have? And it was quite a simple answer because it was, I think, "more than 102", "less than 102" or "the same" and the answer was obviously "more than". So, there was three
questions about financial literacy, one was about compound interest, one was about risk and so on.

166
00:17:50,080 --> 00:17:56,880
We also had eco literacy, which was like do you know what the ozone is there to protect us from?

167
00:17:56,880 --> 00:18:04,800
What's the greatest risk to animal habitats? And health literacy I'll talk you through now.

168
00:18:06,640 --> 00:18:10,480
Keep in mind that if you get in contact, we can actually give you the survey.

169
00:18:11,520 --> 00:18:18,080
And you can see also in the survey where our standardised measures came from. So

170
00:18:18,080 --> 00:18:23,040
for health literacy it was again more than just health knowledge or health behaviour.

171
00:18:23,040 --> 00:18:29,520
It's sort of more about empowering people to access, understand, evaluate and communicate

172
00:18:29,520 --> 00:18:34,160
in order to make decisions about their own health. So, it's a critical empowerment strategy.
How often, if at all, do you need to have someone help you understand instruction pamphlets or other written material from your doctor or pharmacy? And how confident are you filling out health related forms? There is a longer health literacy task which is, see the woman up there trying to interpret the bottle? There was a task which was about calculating the maximum number of pills you could take in 12 hours but we did not use that item, but it's a good health item. And from this we ran some regressions. Well first we ran some correlations to look at what was going on with the lifewide literacies. They all correlated negatively with deprivation. So greater deprivation, lower levels of lifewide literacies.
And also, with proxy measures of precarity, such as a greater number of persons over fewer bedrooms, lower income over greater benefits and having internet access.

So, there's definitely something going on in terms of wider issues of capital and inclusion. So, we started to do some regressions in order to explore the predictive relationship of how these lifewide literacies might matter for positive life outcomes like general health. And I promise I'll make the regression bit very brief so that I can hand over to Rachel to actually do some. I think that's probably why you're here.
But if you don't have a background in regression modeling I'll keep it super super simple. From

00:20:14,800 --> 00:20:20,080
a jobbing statistician type of view, which is basically what do I need to know to get the

00:20:20,080 --> 00:20:26,480
job done? Linear regression is based on a linear model. It comes from the Classical or Frequentest

00:20:26,480 --> 00:20:32,240
sort of view of statistics, which is how I was trained. I realise that there are now alternatives,

00:20:32,240 --> 00:20:38,240
like Bayesian statistics is really very popular. But basically, in this model you're looking to

00:20:38,240 --> 00:20:47,360
predict one outcome from one or more predictor variables. So, in this scenario it's assumed that

00:20:47,360 --> 00:20:53,840
the relationship between the variables that you're looking for is linear or linearish. And so that way

00:20:53,840 --> 00:20:59,200
regression is based on the formula for a line. So simply put y, which is the outcome you're
trying to predict, is comprised of a constant - how high up the line starts on your graph - plus

the slope - how steep that relationship is, either a positive correlation or a negative -

times the actual value for your predictor. In a simple linear regression, you've got one

predictor and one outcome and it might look like this. So, this is taken from an online

stats guide for R, which I thought was really quite helpful. If you want to look it up I've put

the web link in the notes. So, we'll circulate these slides to you. Navarro 2011, and this is

how grumpy I am via how much sleep I've had in my last few evenings. So, you can see that

this one has less sleep, greater grumpiness, yes, in the predicted direction, lovely.
And in addition to this you should also have overall an indicator of whether that predictor is significant and how much variance in grumpiness is explained by sleep. And if you have a multiple regression you'll have how much variance and grumpiness is explained by a variety of factors, so not just my sleep but whether I'm having a good hair day or not, how my children slept and so on. But there are some caveats. We need to look out for issues like outliers, multi-collinearity and strange patterns in our residual errors. So that's the background to what we did in the paper and here's just the regressions that we present in that paper. Literacies do in fact matter and so, what I've presented here is eco literacy,
financial literacy and health literacy are all significant individual predictors and I've got
the standardised slope there and I've got the stars that indicate the significance of the t-test.

Is this a good predictor or not? Very, very broadly speaking. And together they explain
15.1 percent of the variance in general health. And so, we did two other regressions there that
say also area relationship matters in predicting general health and so does social engagement. And
all of these models in statistics are more fully explained in the notes of the slide.

But I don't need to go too much into that because you can read that paper if you're
interested in it. The last thing I want to say is how relationships can get a little bit complicated.

by moderations and mediations. So basically, I was talking to you about just straightforward

variables. This predictor affects this outcome or this IV affects this DV. What can happen is a

third variable, such as a confound or a covariate, can actually affect that initial relationship. So

either it's a mediator and it comes in the middle, or it's a moderator which comes at the start but

acts like a foot on an accelerator. So, broadly speaking, your better grades might lead to

greater increased study motivation but we might find that a mediator is self-esteem, so it's not

just a direct relationship but you've got this indirect effect of grades boost your
self-esteem and the self-esteem contributes to greater motivation. Or it could be a moderator.

My past educational qualifications or experience are going to act as an accelerator to have either
greater or reduced present grades, presumably, it could work that way. And then that

makes the grades affect how the grades directly impact on study motivation. So, I think I've said

enough about that, but just to show you how we use that in our paper. We found that social support -

feeling like you had neighbours to turn to in a crisis, having talked to neighbours over

the past week, having diverse friendship groups - was a direct indicator of greater self-reported

general health. But what happens is that health literacy is also a third mediating variable.
So greater social support, greater health empowerment, better self-reported empowerment and action over your health behaviours, which means that that third mediator wipes out or makes not significant the initial relationship. And so that is how we have used, that's the whole story from the frameworks to the data that we collected to how we chose to use the data. But there's so much more that could be done with this data. And so, I will present over to Rachel to go ahead and show you. But just to say that we are going to walk you through some logistic regression. So maybe Rachel can just say a little bit more as she runs it. But there are many alternatives to linear
regression and logistic regression is just one and furthermore binary logistic regression is just

one. So, in this instance your outcome, what you're trying to predict, is no longer continuous or

normally distributed. It's usually like yes, no, pass, fail, although it could be ordinal or multinomial.

So, your predictor, what's going in the IV, can still be any level. It can be,

it can be grouping, it could be a continuous, it could be likert scale. But what's coming

out is now usually dichotomous or somehow categorical. And so, it works in a different way

to regression models. It doesn't use Ordinary Least Squares to model that line, that regression line,

because obviously there's no linear relationship. So, it looks at log transformations and odd ratios
and it's still frequentist classical statistics because you're looking at the probability of this outcome event happening at different levels of the IV. So, it uses something called maximum likelihood estimation. But just to say we're aware that there's other types of approaches like Probit regression and Bayesian regression approaches. But this is just the one that we're going to demo for you today. And with that I will stop my screen share because it's demo time.

Hi everyone. So, I have prepared a rather simple script using the iMCD data. So, I'm going to go over that with you just now. For those of you who are not
familiar with R, there's the script over here and over here we have the global environment,

which is where our outputs are going to be. So, our data sets, every object that we're going
to create is going to appear over here in our global environment. So, the data can be accessed
from this link. It should be in the chat now as well. There's an application procedure that you
need to follow in order to obtain the data. It's fairly simple, you just fill in a couple of forms
and the data will be in your inbox to use. So, for this webinar we're going to be using the iMCD
household survey data that Catherine spoke about earlier. And for general housekeeping it's best
to keep all your data and the markdown file that you're going to use - so this is the markdown file -
all together in the working directory. So, I'll just show you,

see I have over here the data and this is the script, the R markdown file. They're all saved

together in a folder so that will work nicely. Ok so we're going to start by loading the

These packages basically allow you to use different functions throughout your

script. And these were the ones that I tend to use, so it's a matter of preference really. If you don't

have them installed all you need to do is remove the hashtag over there and the line suddenly

becomes code and you can run that. But I'm going to leave the hashtag on because I already have them
installed. So that's the libraries installed and next we're going to load the raw data.

And this is the data, the iMCD data, so we can have a look at these. They've appeared over here on the global environment. Let's have a look through the code book and over here you can see these are the codes of the variables which you will also find in the raw data. So, these are your variables up here but in the data file you can see this is the actual data collected. The code book is just explaining to you what each variable consists of, any notes and how it is coded. You can see, for example, let's pick one over here, so this is household relationship - how person one is related to the HIH.
And you can see value one would be husband or wife, partner, civil partner. Value two is son or daughter. So, you get an idea of the coding system that was used. So, it's handy to keep it over here because you can refer to it. Ok, so go back to the code, we've loaded the data and now we're going to select and tidy the data. So, for this demo we are interested in how different variables predict an engagement in learning, so our dependent variables are formal learning, informal learning, and self-learning. We have the codes over here, which you can find in the code book. So, I'll look up just to give you an idea. Formal one...

So, this is our variable for formal education, so that's the ID over there, and we have
it in the data as well. Let's look this up. Let's see, oh right of course that's not in this line.

Ok, so these are our dependent variables and our independent variables that we're going to be looking at. So, our predictors are age, area safe - so do the participants feel safe walking at night - belonging to area and general health.

So, we're going to create a new data object over here, so it will always keep the raw data separate. And that's just handy to have in case you make a mistake with the code. You can always refer back to your raw data. So, we have our new data folder here, data frame here, and this includes the
unique ID for each individual, formal education, informal education, self-learn, age, areasafe. So, all our
variables. So, it's a trimmed down version, this is what's relevant for our analysis.

So, if we have a look, we can see that, so we look at formal education, so the minimum is
one and the maximum is two. So, we know that it's coded for ones and twos. We've got 23 NAs.

In the informal variable we can see the minimum is one and the maximum is three, so we've got one,
two and three. So, this just gives you an idea of how the variables are coded because we
need to understand the data in order to analyse that of course. So, this is a breakdown of what
we gather from this. Now if we have a look through the code book, let's pick formal learning again.
We have coded, 'yes' is coded as one, 'no' is coded as two and 'don't know' is coded as eight. So, we are only interested in the 'yes' and 'no' s and to go forward we'd like to recode that so that yeses are ones nos are zeros and we will recode the don't knows into zeros as well so that they won't be included as engaging in learning. So, we're going to go ahead and this bit of code is going to do exactly that. I'll run each line separately.

Ok and we're also going to remove the NAs.

So now we'll have a look at the data again and we can see that formal1 now has a minimum of zero and a maximum of one. Same with informal and self-learning. So now we've got
our variables cleaned as we wanted, and we'll do the same with the independent variables.

If we have a look at, for example, areasafe we'll go back to the code book.

See how that's coded areasafe. So, we've got a one for 'very safe', two for 'fairly safe', three 'a bit unsafe', 'very unsafe' and then we've got five which is 'never go out after dark'.

And so, we would like to recode that so that it is increasing with safety. This is just a preference for this analysis. We're going to do the same for belonging to area and general health.

Belonging to area has number five coded as 'don't know' and general health has number six coded as
'don't know' and number seven coded as 'refused'. And those are not relevant for this analysis so we're going to remove those as well. Here we go removing them. And now we can have a look at the data again.

And everything is exactly how we want it, so we can go ahead and change this to long format. So what do we mean by that? If we look at the data file, we've got separate columns for the different types of learning. Now we want all engagement and learning to be in a long format, so we want them all into one column and we'll have one column that describes which type of learning and another column which has the corresponding value.

This means that every participant will now have three rows, one for each type
of learning. So, we'll go ahead and do that and as you can see it's now in long format.

Ok, so now we can run a few descriptives based on our data. So now we've got 2,069 observations. This went down from 2,095. That's because we removed the NAs and missing data, anything that we didn't need.

average age. And if we have a look here, we've got a breakdown by the different types of learning. So just over 10 percent were engaging in formal learning.

Just over seven percent informal learning and 10.7 percent were engaging in self-learning.

We can plot the distributions of the predictors. So, if we have a look at age as expected, area safety
so now this is going from very unsafe to very safe. Likewise belonging to area,

no sense of belonging to a strong sense of belonging, and general health.

And these are some plots here. I've plotted some and I've plotted the effects of the predictors on learning. And just so you know these are, I treated the predictors as continuous variables just because it was simpler and it's easier to digest the information that the plot is trying to give up. So, age is continuous, so that's fine, but you'd probably want to plot them differently. That's up to you. So, we can have a look and we see that as feeling safe walking at night increases, you're more likely to engage in learning and
this is the same across all three types of learning. We can have a look at belonging to area.

See it moves differently across the different types of learning.

And general health.

So now we're going to go to the analysis. And for the analysis the categorical variables -

so that's general health, belonging to area and area is safe - these are

on a scale from one to five, one to four and one to five. So, we're going to convert these

into factors so that R knows to treat them as factors. We'll have a look
here. So, we can run this code to see what the variables are currently being treated as. So here we see that areasafe is treated as numeric, belonging to area is the same and general health is the same. So, we want to change that so that we can notify R that the levels actually mean something else. So, they're not just numbers. So, we're going to do that. Convert them into factors and then we can see the contrast coding for each variable. So, this is going to compare and so we see that group number one is the reference group for each of the categorical independent variables. So, if I'm looking at general health, for example, it's level number one that's going to be our reference group which means that
when we run the regression, the intercept is going to be representing this particular group.

Right so let's go ahead and run the regression. So

first, we're going to look at area is safe. I've broken these down separately for each

variable just for ease of showing you what you can do with the data. So, we're

going to look at area is safe, and this is based on total engagement. So formal, informal, and self.

We have a look at the results and we can see that area safety appears to be significant at least areas group three, four and five. So, on the scale here three, four and five. Those appear to be significantly different from the
reference group, which is the currently the intercept. Likewise, with the interaction between age and area safety, age is a very strong predictor as expected and so I decided to include that in there so that we can run the interaction as well. So, this gives us an idea that area safety may be a significant predictor. We'll break this down by the different types of learning as well. So here we are filtering the data to include just the formal learning, education formal learning engagement. And if we have a look now, area safety is no longer significant.

If we look at informal learning, it's approaching significance when interacting with
age. And if we look at the self-learning, we see that in area five and the interaction

00:42:11,600 --> 00:42:18,400
and some are approaching significance as well. So, we'll run a likelihood ratio test that compares

00:42:19,040 --> 00:42:28,720
the model that includes the area safety variable with a model that doesn't include it. So

00:42:28,720 --> 00:42:33,840
we can test whether it's a relevant variable at all, whether what we're seeing is actually

00:42:33,840 --> 00:42:40,480
something. So here we're going to create model none. So, this doesn't have area safety, it just has

00:42:40,480 --> 00:42:48,000
age.

00:42:48,000 --> 00:42:54,160
And we're going to compare the two and we see that area safety is in fact an important variable to

00:42:54,800 --> 00:43:00,720
look into further. We'll do the same with belonging to area. So first looking at total engagement.
And we see that it is not a significant predictor, so we'll move on, there's no need to explore further. And likewise, with general health. So that just gives you an idea of what you can do with the data. But obviously you can play around and make your own more sophisticated regressions and models. I would strongly recommend looking through the code book because this is where you can see all the different types of variables that you can use.

There's 537 odd variables on here at least. So, there's definitely so many different variables that you can play around with and the sample size is really quite large so it should really be explored further. I think that's it from me. We're going to make the script available.
so you’ll be able to play around with it like this and hopefully go on further. [Catherine Lido] Thank you so much

Rachel. So we’ve had some great questions that I’ve been trying to answer as we go. Waltraut did notice that there is actually a slight labelling issue with what the data set labels as formal, informal, and self-learner. Actually formal, non-formal and informal in terms of education speak. But you can absolutely apply to get the survey. I copied some of the actual wording of the questions over but if you use that UBDC link not only can you apply for the data but you can actually get the survey itself and the technical report that
Ipsos MORI prepared for us. So, you can absolutely use it for educational purposes, which some of you expressed interest in doing. So, thanks so much Rachel. I'll just take back control and finish up by showing what you can do with some of the other data as well. You can actually apply to look at the GPS data. Again, I'm not quite sure how that works in terms of the data preparation for use by third parties. But here's a little visualisation that was done with the iMCD data to just look at the patterns by gender. I hope I can get it to play but it obviously might not when you want it to. But it just shows the different patterns of men and women around the city centre. And I apparently can't get it to play. Ok
here we go, yes I can. So, this is the centre of Glasgow and we have a terribly gendered blue for

for men and a purplish colour for women, at least it’s not pink! So, you can kind of see

Oh sorry, a lovely visualisation. Am I just on my own here? Oh, the webinar was going so well until

that moment. I’m watching a very nice visualisation here which none of you can see. Here we go, let’s

try this again. Can you see it now Muir? [Muir Houston] Yes! [Catherine Lido] Another minute to see if it will play. Can you see that now?

Is it moving my screen? [Muir Houston] Yes! [Catherine Lido] Ok, so now you can see the blue pattern is for men walking in

and around the city centre in a 24-hour period versus women and so on. But you could use the
GPS data to combine with the survey data and by then you could look at like demographic
characteristics such as persons with disabilities or people who are using more active transportation
modes, people in different housing conditions or even with different political affiliations.

Because not only is there the GPS data but there's a written travel diary that tracks the trips
that people have made in the days preceding their survey participation. So, there is some
work here to be done in terms of sedentary patterns say for certain demographic
categories. And in addition to the transportation related data, like the GPS
and the travel diary, there's also the Twitter dashboard which I've just stuck
the link here. I don’t know if Rachel wants to copy it over. But it’s a very easy to use

searchable bespoke tool to search the Twitter data that we captured. And we used

those domains that I presented at the start in terms of education, sustainability, transportation.

And one percent of it is geo-located within the city of Glasgow. So, they’re Glasgow related or

Glasgow geolocated over that year and what we can do is you can search through this corpus

of 2.9 terabytes of Twitter data by time, place, or keyword. And then it goes through and it uses

the ideas to hydrate them. Of course, that means it can’t access tweets that no longer exist or

have been removed or have restricted access. But it can give you sample tweets, basic statistics,
interactive maps, the density of tweeting in a specific region and you can do things like sentiment analysis - you know, happy, sad, and so on. So that's something else you could have a look at for your own research or for students. We took some of this data and we tried to create a story about the meaning of lifewide literacies for better mental health, well-being and physical health and we took it out to places like IKEA, the Mugdock Country Park for MoSSFest, for all kinds of social science festivals. And we created these little lifewide literacies badges. So, you've got here this dinosaur, reading, theatre. So, you've got regular reading literacy,
you've got the creative arts or cultural literacy and foreign language literacy.

And this one is holding a green leaf, that's environmental literacy. And so, children and adults were able to customise their badges and create their own literacy person. But it was just a way of moving from the data to actually communicating our messages with people. So in a sense the whole story of our research and what we're doing here today is really about the data literacy. And I told you we've measured ICT, computer use and attitudes towards technology. But data literacy, again, is an empowerment strategy. We wrote a briefing paper for UNESCO for their last learning cities conference and that should be released as a publication shortly. And it's about
inclusion in data literacy. As we've had to move increasingly online, we really need to look more

holistically at people's experiences with data, with technology and also with who's being left out

of these data conversations. Who's being left out, who's being researched on and not researched with.

So, this is a really important strand of the work we're doing at Urban Big Data Centre. Breaking down

the quantitative qualitative divide. Working across disciplinary silos to create really impactful

uses of data and we realise and we acknowledge there's real tensions with GDPR privacy concerns,

with the push for openness. So, we are mindful that this is possibly more ethically fraught than
we're making it out to be today, but we just hope that we've shared with you some ways in which

Urban Big Data Centre could help support your research or your educational needs in terms of

providing access to data like iMCD. Providing we've got other data as well, we're working on a further

education data acquisition. We've got some higher education data including some UCAS lookup tables. But there's loads of training webinars and seminars going on at UBDC. So

there's more of these data dives, but there's also funded PhD studentships and opportunities

to work with us either using Glasgow data as a comparator, but we have national data, we've got

cycling and mobility data, we've got all types of things. So why not consider us to help support
you in your work in future? So that is it from us as the team. I think I've really tried to tackle

the questions and as we've gone through. So, I don't believe there's any unanswered questions.

But just to review the questions, and the answer is that, yes, you can apply to get the survey data

and the survey itself. And within that it will give you a greater idea of how these different types

of learning were operationalised and measured and how that links to the learning cities agenda.

Somebody else, I think it was David, asked a great question about symphonic social science

and I should have referenced that as the work of Susan Halford, and I've put a link to their work
in the chat box. But I think overall that's most of the questions covered. So, feel free to keep

463
00:52:41,440 --> 00:52:49,120
in touch with us. Yes, you’re right Jack, the travel diary is a very interesting and underused

464
00:52:49,120 --> 00:52:54,960
resource.

00:52:54,960 --> 00:52:59,840
So just use that UBDC form. It's as simple as if you got cut off, you know, when you get cut off on

465
00:52:59,840 --> 00:53:04,400
the telephone and you do one of those complaint forms, it's just that quick and simple. What data

466
00:53:04,400 --> 00:53:10,160
acquisition are you interested in? if you're interested in iMCD what strand of it is it? -

467
00:53:10,160 --> 00:53:16,720
the survey, the travel diary, the survey and the travel diary, the GPS, the lifelogging images

468
00:53:16,720 --> 00:53:23,920
or some more in-depth twitter data that you can't get off the Twitter dashboard. So be sure, as

469
00:53:23,920 --> 00:53:28,940
Muir

00:53:28,940 --> 00:53:31,600
says, to look at the other two webinars that are happening. We've had a small attendance but
really insightful questions. And so, when this video gets up, why not apply for the data

and go through what Rachel has done and try and replicate it. Because the biggest obstacle, I think,

for working with secondary open data sources is that you don't know what you're

looking at. You don't know how to clean it. You don't know how to approach it. You don't

know how it's operationalised. And I think what Rachel and I hopefully did here for you today

is to peek under the hood of some of our publications but also

to really practically think if I got this data set of 500 and something variables where would I start?

and I think we've helped you make that start. So, we're bang on the one hour. I will just say
thank you so much for attending. Muir or Rachel, do you want to say anything to conclude?

[Muir Houston] No just to reiterate that this is the second in a series of four webinars. The first one was on

Strava cycling data in the Glasgow city area. And, as Catherine has said, we will be putting these videos online but for accessibility purposes we need to provide them with a transcript and various accessibility measures to ensure compliance with current regulation. So, it might take us a wee bit to do the technical background stuff for that sort of stuff. But, as I've said, there's two more coming up in the series. If you look on the UBDC site and again, as with this one, free registration.
So, thanks for taking part. [Catherine Lido] Thank you everybody, have a great day, thanks so much.

Bye.