

# Exploring learning city inclusion with diverse education data

Transcript from [webinar video recording](#)

1

00:00:00,880 --> 00:00:04,880

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

2

00:00:04,880 --> 00:00:11,200

with diverse education data with Professor Catherine Lido, ably assisted by Rachel Cassar.

3

00:00:12,640 --> 00:00:16,880

I would like to remind participants that this session is being recorded and will be uploaded on

4

00:00:16,880 --> 00:00:22,960

our UBDC website as an accessible format for those who could not attend the session, or if you want to

5

00:00:22,960 --> 00:00:28,640

have another look at it. Further details will be provided on the UBDC website, but you may wish to

6

00:00:29,440 --> 00:00:33,680

adjust your name as it appears on the screen. Cameras will be turned off and microphones muted

7

00:00:33,680 --> 00:00:38,960

to aid privacy and also for bandwidth reasons. Feel free to introduce yourself in the chat as Catherine's

8

00:00:38,960 --> 00:00:44,320

been saying but please do not include any personal information such as phone numbers or emails.

9

00:00:44,320 --> 00:00:49,920

But please do say hi and check out our website for further webinars and other resources

10

00:00:49,920 --> 00:00:54,880

including how to access data and other training and events delivered by UBDC.

11

00:00:55,600 --> 00:01:00,480

Please use the Q&A facility to ask questions. These will be collated, and responses will be

12

00:01:00,480 --> 00:01:08,240

provided in the Q&A session. Session structure - Catherine will provide a presentation, Rachel will

13

00:01:08,240 --> 00:01:15,200

do a demonstration of the data and the statistical analysis and then we will hold our Q&A session.

14

00:01:15,200 --> 00:01:19,120

So, I'll hand over to Catherine and she'll say a little bit about the

15

00:01:19,840 --> 00:01:26,080

presentation to come. So, thank you Catherine. [Catherine Lido] Thank you so much

16

00:01:26,080 --> 00:01:32,640

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

17

00:01:39,120 --> 00:01:46,080

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

18

00:01:46,640 --> 00:01:55,680

I am, as Muir said, Professor Catherine Lido and I'm ably assisted by Rachel Cassar who is a research

19

00:01:55,680 --> 00:02:03,280

assistant for our VisNET project. I'll just launch right in, but any questions along the way you

20

00:02:03,280 --> 00:02:09,920

just go ahead and pop them in the Q&A box. So, we're both psychologists and you might be thinking

21

00:02:09,920 --> 00:02:16,000

what can psychologists tell us about education and learning inclusion? Aren't psychologists people who

22

00:02:16,000 --> 00:02:23,280

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

23

00:02:23,280 --> 00:02:29,040

psychologists and psychology as a discipline is interested in people and how they think, act and

24

00:02:29,040 --> 00:02:36,480

react. We come from a social psychology background, which is not just how people act on their own but

25

00:02:36,480 --> 00:02:44,000

how we act in groups. And learning, adult learning, is a really powerful way to look at humans and

26

00:02:44,000 --> 00:02:50,160

humans in their context. So, this research is about life-wide learning inclusion and how we can learn

27

00:02:50,160 --> 00:02:57,280

more using some of the skills that psychologists have in terms of quantitative data, but also skills

28

00:02:57,280 --> 00:03:04,000

that come from urban planning and other disciplines to try and maybe get more

29

00:03:04,000 --> 00:03:11,840

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

30

00:03:11,840 --> 00:03:16,720

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

31

00:03:16,720 --> 00:03:22,560

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

32

00:03:22,560 --> 00:03:28,560

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

33

00:03:28,560 --> 00:03:34,880

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

34

00:03:34,880 --> 00:03:42,400

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

35

00:03:42,400 --> 00:03:48,400

but it might be quite realistic data - to try and blur the boundaries between qualitative and

36

00:03:48,400 --> 00:03:56,080

quantitative and instead come up with stories that can really help us improve the lives of people

37

00:03:56,080 --> 00:04:02,000

around us. And so, part of that is my work here with Urban Big Data Centre. So, if you want to read more

38

00:04:02,000 --> 00:04:10,240

about this, I was in The Psychologist magazine on the cover talking about big data in the big city.

39

00:04:10,240 --> 00:04:17,040

But all of this is couched, this whole webinar series is offered in the wider scope of Urban Big Data

40

00:04:17,040 --> 00:04:23,360

Centre as a research centre, which is basically a free resource for you all to use to try and

41

00:04:23,360 --> 00:04:30,400

get people to use more big and novel data and research methods to try and tackle some of these

42

00:04:30,400 --> 00:04:37,120

big social and environmental crises we're facing today. So, thank you so much for joining us. Just by

43

00:04:37,120 --> 00:04:42,400

being here you're really supporting our work and helping us get the word out and the message out

44

00:04:43,440 --> 00:04:47,280

that we want people to access our data. And if you're particularly

45

00:04:47,280 --> 00:04:53,360

interested in educational disadvantage please get in touch with one of our team. You met

46

00:04:54,000 --> 00:04:59,360

Muir Houston there but we've also got Professor Mike Osborne, Professor Keith Kintrea,

47

00:04:59,360 --> 00:05:04,640

Dr Phil Mason and a couple of wonderful PhDs - Brittney Nathaniel and Barry Black.

48

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

49

00:05:11,280 --> 00:05:18,960

is better for our health, for our occupational outcomes, for engaging citizens and for actually

50

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

51

00:05:23,840 --> 00:05:31,200

areas such as place-based inequalities in schools' attainment, further education and

52

00:05:31,200 --> 00:05:37,280

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

53

00:05:37,280 --> 00:05:42,400

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

54

00:05:42,400 --> 00:05:50,000

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

55

00:05:51,840 --> 00:05:58,000

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

56

00:05:58,000 --> 00:06:03,200

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

57

00:06:03,200 --> 00:06:11,680

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

58

00:06:12,240 --> 00:06:18,880

an interesting alternative to the smart future cities idea, which says yes we need smart cities,



59

00:06:18,880 --> 00:06:24,880

yes we need future cities but we have an aging population, more and more people living in cities.

60

00:06:25,760 --> 00:06:30,080

We really need to meet these challenges to social inclusion, technology,

61

00:06:31,360 --> 00:06:36,480

knowledge economy, diversity, and sustainability. And in order to do that we probably need to keep

62

00:06:37,040 --> 00:06:47,840

people engaged and upskilled and working longer towards transforming learning cities. So, it's

63

00:06:47,840 --> 00:06:54,160

really an idea of engaging all citizens across all ages. And we do a lot of that work with the

64

00:06:54,160 --> 00:07:00,560

PASCAL International Observatory in our School of Education, of which I'm the Deputy Director

65

00:07:01,360 --> 00:07:04,960

for the European Hub. So, if you're interested in learning more about PASCAL

66

00:07:05,840 --> 00:07:10,240

do see the Glasgow website. I know we have some PASCAL members here, welcome welcome.

67

00:07:10,240 --> 00:07:13,600

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

68

00:07:14,240 --> 00:07:21,840

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

69

00:07:21,840 --> 00:07:29,040

It's available online if you want to learn more. But you can see how it sort of has what's

70

00:07:29,040 --> 00:07:36,080

fundamental to be in place before the learning city vision is enacted and that includes resources,

71

00:07:36,080 --> 00:07:42,320

governance participation and political will. And then the columns are really

72

00:07:42,880 --> 00:07:50,160

some of the things that we can measure in terms of do we have inclusive participation from

73

00:07:50,160 --> 00:07:57,680

early years to adult learning and so on. What about mobilising our online learning technologies?

74

00:07:57,680 --> 00:08:03,760

What about having a vibrant culture of learning through our lives and through our cities? And so

75

00:08:04,800 --> 00:08:10,880

this document actually came up with 42 indicators against which you could benchmark a city's

76

00:08:10,880 --> 00:08:17,040

sort of success as a learning city. And this is just an example, so you might be able to

77

00:08:17,040 --> 00:08:23,040

measure empowering individuals and promoting social cohesion say through civic participation

78

00:08:23,040 --> 00:08:29,040

like volunteering and voting and so on. You might be able to assess inclusive learning

79

00:08:29,040 --> 00:08:36,240

just by looking at diversity and demographics pre-nursery, pre-kindergarten,

80

00:08:36,240 --> 00:08:42,320

provision and so on, funded adult learning. And you might be able to examine something

81

00:08:42,320 --> 00:08:47,360

like literacies and not just reading, we'll talk about other types of literacies today.

82

00:08:49,840 --> 00:08:57,040

So, I'll just quickly operationalise a couple of key terms before we launch into the present project.

83

00:08:57,040 --> 00:09:03,040

You've probably heard of lifelong learning. If you haven't it's unhappy tagline is "cradle to

84

00:09:03,040 --> 00:09:09,280

grave", which I feel is incredibly depressing. So, I tend to think more about lifewide learning, which

85

00:09:09,280 --> 00:09:14,960

is this little box here, that formal learning leading to a qualification is such a small

86

00:09:15,600 --> 00:09:20,400

part of the learning that we do every day. I mean, you're all here engaging with us because

87

00:09:20,400 --> 00:09:26,000

you want to learn, and you might have different motivations for being here but at the heart of it

88

00:09:26,000 --> 00:09:32,000

is you've all stepped out of the formal learning environment to perhaps, I would call this

89

00:09:32,000 --> 00:09:36,640

maybe a non-formal learning environment. It is structured. I apologise that I'm speaking

90

00:09:36,640 --> 00:09:42,480

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

91

00:09:42,480 --> 00:09:49,520

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

92

00:09:49,520 --> 00:09:55,200

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

93

00:09:56,320 --> 00:10:03,520

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

94

00:10:03,520 --> 00:10:10,960

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

95

00:10:11,920 --> 00:10:18,240

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

96

00:10:18,240 --> 00:10:24,800

create, communicate and compute using printed and written materials associated with varying contexts.

97

00:10:24,800 --> 00:10:29,600

Literacy involves a continuum of learning in enabling individuals to achieve their goals,

98

00:10:29,600 --> 00:10:35,040

to develop their knowledge and potential and to participate fully in their community and wider

99

00:10:35,040 --> 00:10:43,360

society. You can see that there's something sort of empowering or enabling about these less formal

100

00:10:43,360 --> 00:10:51,760

types of literacies and learning. So that is the entire background and framework that brought us

101

00:10:51,760 --> 00:11:00,400

to the Integrated Multimedia City Data project, which I'm talking about today. The iMCD, or i mc d  
as

102

00:11:00,400 --> 00:11:08,080

I like to call it, is the first data product that was created at Urban Big Data Centre as a free

103

00:11:08,640 --> 00:11:15,840

open data resource. Anybody can apply to use it, and anybody can apply for any one of the strands.

104

00:11:15,840 --> 00:11:23,120

The first strand is a 1,500-household survey. The second strand is that we followed up about 600

105

00:11:23,120 --> 00:11:30,800

or so participants to wear GPS trackers around the city for one week so we know what they're

106

00:11:30,800 --> 00:11:35,840

doing, where they're going. I'll talk to you about the ethics of all of this maybe just very briefly

107

00:11:35,840 --> 00:11:41,840

but that's probably another talk. Lifelogging cameras they wore around their neck, that was about

108

00:11:42,400 --> 00:11:47,600

300 or so participants and it took pictures every few seconds as they went about their

109

00:11:48,400 --> 00:11:53,840

daily life for 48 hours. And lastly, we had one year's worth of social media capture.

110

00:11:54,480 --> 00:12:00,560

Largely Twitter data but also traffic camera feeds, ScotRail feeds and so on.

111

00:12:02,480 --> 00:12:07,840

So, I'll just hone into the survey first, but just to say that we did treat the ethics of this

112

00:12:07,840 --> 00:12:15,760

very carefully. Not all of that data is equally open. Some of the data is obviously much more

113

00:12:15,760 --> 00:12:22,160

sensitive and so much more limited in terms of access. But the Understanding Glasgow survey is

114

00:12:22,160 --> 00:12:28,800

very open. We have an open version that you can apply to use and all you really need to do is go

115

00:12:28,800 --> 00:12:35,280

on to the website, fill in like a very brief form and you can have this data to look at yourself.

116

00:12:35,280 --> 00:12:41,040

So today we're going to introduce you to the survey data. Rachel will take you through a demo

117

00:12:41,040 --> 00:12:46,880

and the hope is that you might access it yourself to practice the types of skills that we're talking

118

00:12:46,880 --> 00:12:56,000

about or to use it in your classes as a data example or to use it for your own

119

00:12:56,000 --> 00:13:01,520

personal research because it's really an under tapped resource. There's still plenty in here



120

00:13:01,520 --> 00:13:06,960

that's never been looked at. So, what did we do for the survey? Well, we reviewed national and

121

00:13:06,960 --> 00:13:14,240

EU surveys as well as these UNESCO indicators to try to measure attitudes (what do you feel about

122

00:13:14,240 --> 00:13:20,160

something), behaviours (what do you do about something), and literacies (what do you know about something) in

123

00:13:20,160 --> 00:13:26,640

the domains of education, sustainability, transport, cultural and civic engagement, and ICT/technology.

124

00:13:27,600 --> 00:13:33,440

This was delivered door-to-door via a company called Ipsos MORI. And so, here's some of the

125

00:13:33,440 --> 00:13:41,120

little statistics that we have here. 2,095 adults took part. Age range 16 to 102, average age near

126

00:13:41,120 --> 00:13:46,480

50. So it's quite good for looking at adult learning and older adult learning because we

127

00:13:46,480 --> 00:13:53,440

have a range up to 102. And it's roughly, although it's the Greater Glasgow area - that's

128

00:13:54,080 --> 00:13:59,040

eight local authorities - it's broadly representative of Scotland in terms

129

00:13:59,040 --> 00:14:03,440

of all of the demographics we collected. So, it's representative of census data.

130

00:14:04,560 --> 00:14:10,240

The education questions, Rachel will talk us through some examples using them,

131

00:14:10,960 --> 00:14:17,520

but we tried to capture formal learning leading to qualifications, any formal learning engaged in in

132

00:14:17,520 --> 00:14:23,920

the last 12 months, any informal learning engaged in in the last 12 months that's structured but

133

00:14:23,920 --> 00:14:29,760

not leading to a formal qualification and then all the other types of learning that goes on. Sorry I

134

00:14:29,760 --> 00:14:36,000

swapped those, it's not my fault it's the way they are. Non-formal is structure but not leading to

135

00:14:36,640 --> 00:14:42,720

a qualification and informal is all other self-learning. I've got this lovely little pyramid

136

00:14:42,720 --> 00:14:48,880

here if you get confused about informal and formal learning. And also, family learning,

137

00:14:48,880 --> 00:14:53,840

which is helping anybody in your family to learn in the last 12 months and whether they were older

138

00:14:53,840 --> 00:15:00,480

or younger than you - a variable which nobody has analysed yet. So I'll talk you through other things

139

00:15:00,480 --> 00:15:07,120

but we have if they were engaged in learning, how? - online or face-to-face, how many hours, motivation

140

00:15:07,120 --> 00:15:12,720

for study. And this is just a list of some of the national measures we reviewed before we took

141

00:15:12,720 --> 00:15:20,960

Potter. Ok, so from this, even though we collected the data interestingly with no research questions

142

00:15:20,960 --> 00:15:27,360

and no aims in mind and that was really really tricky - I can talk you through that in another talk.

143

00:15:27,360 --> 00:15:33,280

I think the first pilot of the survey took six hours. I didn't do it all at once but we piloted

144

00:15:33,280 --> 00:15:38,000

bits of it and we added it all up and it was six hours' worth of questions that all the different

145

00:15:38,000 --> 00:15:43,920

researchers wanted represented in this survey. We got it down to about, I think it was, 45 minutes.

146

00:15:45,360 --> 00:15:51,280

And so, from that we then engaged in our own research and this is a publication we did about

147

00:15:51,280 --> 00:15:58,320

older learner engagement in the modern city. And it looked particularly at older adults, which was

148

00:15:58,320 --> 00:16:04,800

actually 60 plus not 65 plus. And you can read about in our paper that you can see that

149

00:16:04,800 --> 00:16:12,240

as you aged learning engagement drops basically. So

150

00:16:12,800 --> 00:16:17,600

our older adults were engaged in less forms of all types of learning, which is not surprising

151

00:16:17,600 --> 00:16:24,320

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

152

00:16:24,320 --> 00:16:29,280

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

153

00:16:29,280 --> 00:16:36,160

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

154

00:16:36,160 --> 00:16:43,280

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

155

00:16:43,280 --> 00:16:50,400

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

156

00:16:50,400 --> 00:16:54,320

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

157

00:16:54,320 --> 00:17:00,960

in the little green here that's the learning engaged older adults 60 plus compared to

158

00:17:00,960 --> 00:17:05,120

the orange. So those are those two graphs, and you can read about that in the paper.

159

00:17:06,400 --> 00:17:12,960

And then from this we did a second paper which was targeting a little bit more into lifewide

160

00:17:12,960 --> 00:17:20,320

literacies. So, here's just some examples that financial literacy had three questions taken from

161

00:17:20,320 --> 00:17:26,720

Lusardi and Mitchell. They have a standardised short form financial literacy. Suppose you had 100 pounds

162

00:17:26,720 --> 00:17:32,160

in a savings account and the interest rate was two percent per year, after five years how much

163

00:17:32,160 --> 00:17:37,120

do you think you would have? And it was quite a simple answer because it was, I think, "more than

164

00:17:37,120 --> 00:17:44,960

102", "less than 102" or "the same" and the answer was obviously "more than". So, there was three

165

00:17:44,960 --> 00:17:50,080

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.

173

00:18:36,000 --> 00:18:40,480

How often, if at all, do you need to have someone help you understand instruction pamphlets or

174

00:18:40,480 --> 00:18:44,880

other written material from your doctor or pharmacy? And how confident are you filling out

175

00:18:44,880 --> 00:18:50,720

health related forms? There is a longer health literacy task which is, see the woman up there

176

00:18:50,720 --> 00:18:58,240

trying to interpret the bottle? There was a task which was about calculating the maximum

177

00:18:58,800 --> 00:19:02,240

number of pills you could take in 12 hours but we did not use

178

00:19:02,880 --> 00:19:10,400

that item, but it's a good health item. And from this we ran some regressions. Well first

179

00:19:10,400 --> 00:19:16,080

we ran some correlations to look at what was going on with the lifewide literacies. They all

180

00:19:16,080 --> 00:19:22,880

correlated negatively with deprivation. So greater deprivation, lower levels of lifewide literacies.



181

00:19:22,880 --> 00:19:28,560

And also, with proxy measures of precarity, such as a greater number of persons over fewer bedrooms,

182

00:19:28,560 --> 00:19:32,960

lower income over greater benefits and having internet access.

183

00:19:34,160 --> 00:19:44,960

So, there's definitely something going on in terms of wider issues of capital and inclusion.

184

00:19:45,760 --> 00:19:52,000

So, we started to do some regressions in order to explore the predictive relationship of how these

185

00:19:52,000 --> 00:19:58,560

lifewide literacies might matter for positive life outcomes like general health. And I promise

186

00:19:59,200 --> 00:20:03,200

I'll make the regression bit very brief so that I can hand over to

187

00:20:03,760 --> 00:20:07,600

Rachel to actually do some. I think that's probably why you're here.

188

00:20:07,600 --> 00:20:13,520

But if you don't have a background in regression modeling I'll keep it super super simple. From

189

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

190

00:20:20,080 --> 00:20:26,480

job done? Linear regression is based on a linear model. It comes from the Classical or Frequentist

191

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,

192

00:20:32,240 --> 00:20:38,240

like Bayesian statistics is really very popular. But basically, in this model you're looking to

193

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

194

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

195

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

196

00:20:59,200 --> 00:21:07,760

trying to predict, is comprised of a constant - how high up the line starts on your graph - plus

197

00:21:07,760 --> 00:21:14,560

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

198

00:21:14,560 --> 00:21:20,880

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

199

00:21:20,880 --> 00:21:26,160

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

200

00:21:26,720 --> 00:21:31,840

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

201

00:21:31,840 --> 00:21:36,960

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

202

00:21:36,960 --> 00:21:46,320

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

203

00:21:47,440 --> 00:21:53,680

this one has less sleep, greater grumpiness, yes, in the predicted direction, lovely.

204

00:21:53,680 --> 00:22:01,840

And in addition to this you should also have overall an indicator of whether that predictor

205

00:22:01,840 --> 00:22:08,240

is significant and how much variance in grumpiness is explained by sleep. And if

206

00:22:08,240 --> 00:22:13,680

you have a multiple regression you'll have how much variance and grumpiness is explained by

207

00:22:13,680 --> 00:22:18,640

a variety of factors, so not just my sleep but whether I'm having a good hair day or not,

208

00:22:18,640 --> 00:22:24,160

how my children slept and so on. But there are some caveats. We need to look out for issues like

209

00:22:24,720 --> 00:22:32,320

outliers, multi-collinearity and strange patterns in our residual errors. So that's the background

210

00:22:32,320 --> 00:22:37,280

to what we did in the paper and here's just the regressions that we present in that paper.

211

00:22:39,360 --> 00:22:45,280

Literacies do in fact matter and so, what I've presented here is eco literacy,

212

00:22:45,280 --> 00:22:50,720

financial literacy and health literacy are all significant individual predictors and I've got

213

00:22:50,720 --> 00:22:57,040

the standardised slope there and I've got the stars that indicate the significance of the t-test.

214

00:22:57,040 --> 00:23:02,960

Is this a good predictor or not? Very, very broadly speaking. And together they explain

215

00:23:02,960 --> 00:23:08,960

15.1 percent of the variance in general health. And so, we did two other regressions there that

216

00:23:08,960 --> 00:23:15,040

say also area relationship matters in predicting general health and so does social engagement. And

217

00:23:15,040 --> 00:23:21,840

all of these models in statistics are more fully explained in the notes of the slide.

218

00:23:22,400 --> 00:23:28,080

But I don't need to go too much into that because you can read that paper if you're

219

00:23:28,080 --> 00:23:35,360

interested in it. The last thing I want to say is how relationships can get a little bit complicated

220

00:23:35,360 --> 00:23:42,480

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

221

00:23:42,480 --> 00:23:49,760

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

222

00:23:49,760 --> 00:23:57,280

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

223

00:23:57,280 --> 00:24:03,120

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

224

00:24:03,120 --> 00:24:10,160

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

225

00:24:10,160 --> 00:24:16,320

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

226

00:24:16,320 --> 00:24:21,040

just a direct relationship but you've got this indirect effect of grades boost your

227

00:24:21,040 --> 00:24:27,440

self-esteem and the self-esteem contributes to greater motivation. Or it could be a moderator.

228

00:24:27,440 --> 00:24:33,680

My past educational qualifications or experience are going to act as an accelerator to have either

229

00:24:33,680 --> 00:24:39,840

greater or reduced present grades, presumably, it could work that way. And then that

230

00:24:40,400 --> 00:24:47,120

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

231

00:24:47,120 --> 00:24:53,520

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

232

00:24:55,120 --> 00:25:00,320

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

233

00:25:00,320 --> 00:25:05,840

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

234

00:25:05,840 --> 00:25:13,360

general health. But what happens is that health literacy is also a third mediating variable.

235

00:25:13,360 --> 00:25:20,400

So greater social support, greater health empowerment, better self-reported empowerment

236

00:25:20,400 --> 00:25:28,000

and action over your health behaviours, which means that that third mediator wipes out or makes not

237

00:25:28,000 --> 00:25:36,480

significant the initial relationship. And so that is how we have used, that's the whole story from

238

00:25:36,480 --> 00:25:42,960

the frameworks to the data that we collected to how we chose to use the data. But there's so

239

00:25:42,960 --> 00:25:51,680

so much more that could be done with this data. And so, I will present over to Rachel to go ahead

240

00:25:51,680 --> 00:25:58,960

and show you. But just to say that we are going to walk you through some logistic regression. So

241

00:25:58,960 --> 00:26:06,400

maybe Rachel can just say a little bit more as she runs it. But there are many alternatives to linear

242



00:26:06,400 --> 00:26:13,120

regression and logistic regression is just one and furthermore binary logistic regression is just

243

00:26:13,120 --> 00:26:18,960

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

244

00:26:18,960 --> 00:26:26,400

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

245

00:26:26,400 --> 00:26:31,360

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

246

00:26:32,000 --> 00:26:38,720

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

247

00:26:38,720 --> 00:26:45,760

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

248

00:26:45,760 --> 00:26:50,880

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

249

00:26:50,880 --> 00:26:57,120

because obviously there's no linear relationship. So, it looks at log transformations and odd ratios

250

00:26:57,120 --> 00:27:02,160

and it's still frequentest classical statistics because you're looking at the

251

00:27:02,160 --> 00:27:08,400

probability of this outcome event happening at different levels of the IV. So, it uses something

252

00:27:08,400 --> 00:27:14,560

called maximum likelihood estimation. But just to say we're aware that there's other types of

253

00:27:14,560 --> 00:27:20,000

approaches like Probit regression and Bayesian regression approaches. But this is just the one

254

00:27:20,000 --> 00:27:27,680

that we're going to demo for you today. And with that I will stop my screen share because it's demo time.

255

00:27:32,480 --> 00:27:42,080

[Rachel Cassar] Hi everyone. So, I have prepared a rather simple script using the

256

00:27:42,080 --> 00:27:48,640

iMCD data. So, I'm going to go over that with you just now. For those of you who are not

257

00:27:48,640 --> 00:27:54,960

familiar with R, there's the script over here and over here we have the global environment,

258

00:27:54,960 --> 00:28:02,560

which is where our outputs are going to be. So, our data sets, every object that we're going

259

00:28:02,560 --> 00:28:09,520

to create is going to appear over here in our global environment. So, the data can be accessed

260

00:28:09,520 --> 00:28:17,520

from this link. It should be in the chat now as well. There's an application procedure that you

261

00:28:17,520 --> 00:28:22,720

need to follow in order to obtain the data. It's fairly simple, you just fill in a couple of forms

262

00:28:23,440 --> 00:28:31,120

and the data will be in your inbox to use. So, for this webinar we're going to be using the iMCD

263

00:28:31,120 --> 00:28:38,240

household survey data that Catherine spoke about earlier. And for general housekeeping it's best

264

00:28:38,240 --> 00:28:43,840

to keep all your data and the markdown file that you're going to use - so this is the markdown file -

265

00:28:44,720 --> 00:28:48,640

all together in the working directory. So, I'll just show you,

266

00:28:49,920 --> 00:28:57,040

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

267

00:28:57,040 --> 00:29:04,080

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

268

00:29:04,080 --> 00:29:12,640

libraries. These are the packages that I'm going to be using Tidyverse, readxl and lmerTest.

269

00:29:13,360 --> 00:29:21,680

These packages basically allow you to use different functions throughout your

270

00:29:21,680 --> 00:29:28,480

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

271

00:29:28,480 --> 00:29:34,800

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

272

00:29:34,800 --> 00:29:38,960

becomes code and you can run that. But I'm going to leave the hashtag on because I already have them

273

00:29:38,960 --> 00:29:45,840

installed. So that's the libraries installed and next we're going to load the raw data.

274

00:29:47,440 --> 00:29:52,960

And this is the data, the iMCD data, so we can have a look at these. They've appeared over here on the

275

00:29:52,960 --> 00:30:00,560

global environment. Let's have a look through the code book and over here you can see these are the

276

00:30:00,560 --> 00:30:07,520

codes of the variables which you will also find in the raw data. So, these are your variables up here

277

00:30:07,520 --> 00:30:14,480

but in the data file you can see this is the actual data collected. The code book is just explaining

278

00:30:14,480 --> 00:30:25,920

to you what each variable consists of, any notes and how it is coded. You can see, for example,

279

00:30:28,000 --> 00:30:38,080

let's pick one over here, so this is household relationship - how person one is related to the HIH.

280

00:30:38,720 --> 00:30:45,520

And you can see value one would be husband or wife, partner, civil partner. Value two is

281

00:30:45,520 --> 00:30:51,120

son or daughter. So, you get an idea of the coding system that was used. So, it's handy to keep it

282

00:30:51,120 --> 00:30:58,640

over here because you can refer to it. Ok, so go back to the code, we've loaded the data and

283

00:30:58,640 --> 00:31:07,760

now we're going to select and tidy the data. So, for this demo we are interested in how

284

00:31:08,640 --> 00:31:14,800

different variables predict an engagement in learning, so our dependent variables

285

00:31:14,800 --> 00:31:20,000

are formal learning, informal learning, and self-learning. We have the codes over here,

286

00:31:20,000 --> 00:31:27,200

which you can find in the code book. So, I'll look up just to give you an idea. Formal one...

287

00:31:29,760 --> 00:31:38,640

So, this is our variable for formal education, so that's the ID over there, and we have

288

00:31:38,640 --> 00:31:52,080

it in the data as well. Let's look this up. Let's see, oh right of course that's not in this line.

289

00:31:53,280 --> 00:31:59,600

Ok, so these are our dependent variables and our independent variables that we're going to be

290

00:31:59,600 --> 00:32:06,720

looking at. So, our predictors are age, area safe - so do the participants feel safe walking at night -

291

00:32:07,360 --> 00:32:09,760

belonging to area and general health.

292

00:32:10,560 --> 00:32:18,240

So, we're going to create a new data object over here, so it will always keep the raw data separate.

293

00:32:18,880 --> 00:32:25,600

And that's just handy to have in case you make a mistake with the code. You can always refer back to

294

00:32:25,600 --> 00:32:34,320

your raw data. So, we have our new data folder here, data frame here, and this includes the

295

00:32:34,320 --> 00:32:41,840

unique ID for each individual, formal education, informal education, self-learn, age, areasafe. So, all our

296

00:32:41,840 --> 00:32:48,000

variables. So, it's a trimmed down version, this is what's relevant for our analysis.

297

00:32:49,680 --> 00:32:57,040

So, if we have a look, we can see that, so we look at formal education, so the minimum is

298

00:32:57,040 --> 00:33:05,840

one and the maximum is two. So, we know that it's coded for ones and twos. We've got 23 NAs.

299

00:33:07,440 --> 00:33:13,920

In the informal variable we can see the minimum is one and the maximum is three, so we've got one,

300

00:33:13,920 --> 00:33:20,320

two and three. So, this just gives you an idea of how the variables are coded because we

301

00:33:20,320 --> 00:33:27,520

need to understand the data in order to analyse that of course. So, this is a breakdown of what

302

00:33:27,520 --> 00:33:37,840

we gather from this. Now if we have a look through the code book, let's pick formal learning again.



303

00:33:39,840 --> 00:33:47,760

We have coded, 'yes' is coded as one, 'no' is coded as two and 'don't know' is coded as

304

00:33:47,760 --> 00:33:55,440

eight. So, we are only interested in the 'yes' and 'no's and to go forward we'd like to

305

00:33:56,000 --> 00:34:03,280

recode that so that yeses are ones nos are zeros and we will recode the don't knows

306

00:34:03,280 --> 00:34:10,160

into zeros as well so that they won't be included as engaging in learning. So, we're

307

00:34:10,160 --> 00:34:16,400

going to go ahead and this bit of code is going to do exactly that. I'll run each line separately.

308

00:34:19,360 --> 00:34:22,240

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

309

00:34:23,360 --> 00:34:30,080

So now we'll have a look at the data again and we can see that formal1 now has a

310

00:34:30,080 --> 00:34:36,160

minimum of zero and a maximum of one. Same with informal and self-learning. So now we've got

311

00:34:36,160 --> 00:34:44,960

our variables cleaned as we wanted, and we'll do the same with the independent variables.

312

00:34:46,000 --> 00:34:52,320

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

313

00:34:54,320 --> 00:35:04,560

See how that's coded areasafe. So, we've got a one for 'very safe', two for 'fairly safe', three 'a bit

314

00:35:04,560 --> 00:35:08,800

unsafe', 'very unsafe' and then we've got five which is 'never go out after dark'.

315

00:35:09,920 --> 00:35:17,360

And so, we would like to recode that so that it is increasing with safety. This is just

316

00:35:17,360 --> 00:35:22,880

a preference for this analysis. We're going to do the same for belonging to area and general health.

317

00:35:22,880 --> 00:35:30,480

Belonging to area has number five coded as 'don't know' and general health has number six coded as

318

00:35:30,480 --> 00:35:35,840

'don't know' and number seven coded as 'refused'. And those are not relevant for this analysis so we're

319

00:35:35,840 --> 00:35:43,840

going to remove those as well. Here we go removing them. And now we can have a look at the data again.

320

00:35:44,720 --> 00:35:52,400

And everything is exactly how we want it, so we can go ahead and change this to long format. So

321

00:35:52,400 --> 00:35:57,760

what do we mean by that? If we look at the data file, we've got separate columns

322

00:35:57,760 --> 00:36:03,040

for the different types of learning. Now we want all engagement and learning to be

323

00:36:04,000 --> 00:36:09,600

in a long format, so we want them all into one column and we'll have one column that

324

00:36:09,600 --> 00:36:15,920

describes which type of learning and another column which has the corresponding value.

325

00:36:15,920 --> 00:36:20,480

This means that every participant will now have three rows, one for each type

326

00:36:20,480 --> 00:36:28,080

of learning. So, we'll go ahead and do that and as you can see it's now in long format.

327

00:36:30,160 --> 00:36:42,080

Ok, so now we can run a few descriptives based on our data. So now we've got 2,069 observations.  
This

328

00:36:42,080 --> 00:36:51,040

went down from 2,095. That's because we removed the NAs and missing data, anything that we didn't need

329

00:36:52,080 --> 00:37:00,240

average age. And if we have a look here, we've got a breakdown by the different types of learning.  
So

330

00:37:01,040 --> 00:37:04,320

just over 10 percent were engaging in formal learning.

331

00:37:05,040 --> 00:37:12,160

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

332

00:37:14,880 --> 00:37:25,120

We can plot the distributions of the predictors. So, if we have a look at age as expected, area safety

333

00:37:25,680 --> 00:37:34,080

so now this is going from very unsafe to very safe. Likewise belonging to area,

334

00:37:34,080 --> 00:37:39,200

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

335

00:37:42,000 --> 00:37:48,720

And these are some plots here. I've plotted some and I've plotted the effects of the predictors

336

00:37:48,720 --> 00:37:56,800

on learning. And just so you know these are, I treated the predictors as continuous variables

337

00:37:56,800 --> 00:38:05,040

just because it was simpler and it's easier to digest the information that

338

00:38:05,040 --> 00:38:10,480

the plot is trying to give up. So, age is continuous, so that's fine, but you'd

339

00:38:10,480 --> 00:38:16,080

probably want to plot them differently. That's up to you. So, we can have a look and we see that as

340

00:38:16,080 --> 00:38:24,640

feeling safe walking at night increases, you're more likely to engage in learning and

341

00:38:24,640 --> 00:38:31,840

this is the same across all three types of learning. We can have a look at belonging to area.

342

00:38:33,200 --> 00:38:38,480

See it moves differently across the different types of learning.

343

00:38:40,880 --> 00:38:42,080

And general health.

344

00:38:44,240 --> 00:38:51,520

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

345

00:38:51,520 --> 00:38:57,440

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

346

00:38:58,160 --> 00:39:06,080

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

347

00:39:06,080 --> 00:39:14,240

into factors so that R knows to treat them as factors. We'll have a look

348

00:39:14,240 --> 00:39:20,080

here. So, we can run this code to see what the variables are currently being

349

00:39:20,880 --> 00:39:28,720

treated as. So here we see that areasafe is treated as numeric, belonging to area is the same

350

00:39:28,720 --> 00:39:33,920

and general health is the same. So, we want to change that so that we can notify R that

351

00:39:33,920 --> 00:39:39,200

the levels actually mean something else. So, they're not just numbers. So, we're going

352

00:39:39,200 --> 00:39:45,840

to do that. Convert them into factors and then we can see the contrast coding for each variable.

353

00:39:47,360 --> 00:39:54,480

So, this is going to compare and so we see that group number one

354

00:39:54,480 --> 00:40:00,560

is the reference group for each of the categorical independent variables. So, if I'm looking at general

355

00:40:00,560 --> 00:40:05,600

health, for example, it's level number one that's going to be our reference group which means that

356

00:40:05,600 --> 00:40:13,360

when we run the regression, the intercept is going to be representing this particular group.

357

00:40:16,240 --> 00:40:19,520

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

358

00:40:19,520 --> 00:40:26,720

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

359

00:40:26,720 --> 00:40:33,360

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

360

00:40:33,360 --> 00:40:40,320

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

361

00:40:43,200 --> 00:40:46,400

We have a look at the results and we can see that

362

00:40:47,360 --> 00:40:57,520

area safety appears to be significant at least areas group three, four and five. So, on the scale

363

00:40:58,320 --> 00:41:04,640

here three, four and five. Those appear to be significantly different from the



364

00:41:06,080 --> 00:41:12,400

reference group, which is the currently the intercept. Likewise, with the interaction

365

00:41:12,400 --> 00:41:21,120

between age and area safety, age is a very strong predictor as expected and so I decided to include

366

00:41:21,120 --> 00:41:28,960

that in there so that we can run the interaction as well. So, this gives us an idea

367

00:41:28,960 --> 00:41:36,400

that area safety may be a significant predictor. We'll break this down by the different types of

368

00:41:36,400 --> 00:41:45,040

learning as well. So here we are filtering the data to include just the formal learning, education

369

00:41:45,040 --> 00:41:52,960

formal learning engagement. And if we have a look now, area safety is no longer significant.

370

00:41:54,720 --> 00:42:02,480

If we look at informal learning, it's approaching significance when interacting with

371

00:42:02,480 --> 00:42:11,600

age. And if we look at the self-learning, we see that in area five and the interaction

372

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

373

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

374

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

375

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 age.

376

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

And we're going to compare the two and we see that area safety is in fact an important variable to

377

00:42:48,000 --> 00:42:54,160

look into further. We'll do the same with belonging to area. So first looking at total engagement.

378

00:42:54,800 --> 00:43:00,720

We have here total engagement predicted by age and belonging to area and the interaction.

379

00:43:02,320 --> 00:43:09,760

And we see that it is not a significant predictor, so we'll move on, there's no need to explore

380

00:43:09,760 --> 00:43:17,520

further. And likewise, with general health. So that just gives you an idea of what you can do

381

00:43:18,800 --> 00:43:26,000

with the data. But obviously you can play around and make your own more

382

00:43:26,000 --> 00:43:33,440

sophisticated regressions and models. I would strongly recommend looking through the

383

00:43:33,440 --> 00:43:40,080

code book because this is where you can see all the different types of variables that you can use.

384

00:43:40,080 --> 00:43:48,800

There's 537 odd variables on here at least. So, there's definitely so many different variables

385

00:43:48,800 --> 00:43:55,280

that you can play around with and the sample size is really quite large so it should really be

386

00:43:55,280 --> 00:44:02,320

explored further. I think that's it from me. We're going to make the script available

387

00:44:02,320 --> 00:44:09,040

so you'll be able to play around with it like this and hopefully go on further. [Catherine Lido] Thank you so much

388

00:44:09,040 --> 00:44:18,080

Rachel. So we've had some great questions that I've been trying to answer as we go. Waltraut did

389

00:44:18,080 --> 00:44:24,240

notice that there is actually a slight labelling issue with what

390

00:44:24,240 --> 00:44:30,960

the data set labels as formal, informal, and self-learner. Actually formal, non-formal

391

00:44:30,960 --> 00:44:38,080

and informal in terms of education speak. But you can absolutely apply to get the survey. I copied

392

00:44:38,080 --> 00:44:44,560

some of the actual wording of the questions over but if you use that UBDC link not only can you

393

00:44:44,560 --> 00:44:50,400

apply for the data but you can actually get the survey itself and the technical report that

394

00:44:50,400 --> 00:44:56,400

Ipsos MORI prepared for us. So, you can absolutely use it for educational purposes, which some of

395

00:44:56,400 --> 00:45:03,280

you expressed interest in doing. So, thanks so much Rachel. I'll just take back control and finish up

396

00:45:03,280 --> 00:45:09,920

by showing what you can do with some of the other data as well. You can actually apply to look

397

00:45:09,920 --> 00:45:18,240

at the GPS data. Again, I'm not quite sure how that works in terms of the data preparation

398

00:45:18,240 --> 00:45:24,480

for use by third parties. But here's a little visualisation that was done with the iMCD

399

00:45:24,480 --> 00:45:31,680

data to just look at the patterns by gender. I hope I can get it to play but it obviously might not

400

00:45:31,680 --> 00:45:35,600

when you want it to. But it just shows the different patterns of men

401

00:45:35,600 --> 00:45:40,880

and women around the city centre. And I apparently can't get it to play. Ok

402

00:45:41,840 --> 00:45:49,760

here we go, yes I can. So, this is the centre of Glasgow and we have a terribly gendered blue for

403

00:45:49,760 --> 00:45:55,680

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

404

00:45:58,880 --> 00:46:08,080

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

405

00:46:08,080 --> 00:46:14,880

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

406

00:46:14,880 --> 00:46:23,840

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?

407

00:46:26,400 --> 00:46:34,240

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

408

00:46:34,240 --> 00:46:41,280

and around the city centre in a 24-hour period versus women and so on. But you could use the

409

00:46:41,280 --> 00:46:47,680

GPS data to combine with the survey data and by then you could look at like demographic

410

00:46:47,680 --> 00:46:54,320

characteristics such as persons with disabilities or people who are using more active transportation

411

00:46:54,320 --> 00:46:59,520

modes, people in different housing conditions or even with different political affiliations.

412

00:46:59,520 --> 00:47:08,560

Because not only is there the GPS data but there's a written travel diary that tracks the trips

413

00:47:08,560 --> 00:47:14,400

that people have made in the days preceding their survey participation. So, there is some

414

00:47:14,400 --> 00:47:20,240

work here to be done in terms of sedentary patterns say for certain demographic

415

00:47:20,800 --> 00:47:30,960

categories. And in addition to the transportation related data, like the GPS

416

00:47:30,960 --> 00:47:35,440

and the travel diary, there's also the Twitter dashboard which I've just stuck

417

00:47:35,440 --> 00:47:41,680

the link here. I don't know if Rachel wants to copy it over. But it's a very easy to use

418

00:47:41,680 --> 00:47:49,760

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

419

00:47:49,760 --> 00:47:55,440

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

420

00:47:56,080 --> 00:48:02,880

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

421

00:48:02,880 --> 00:48:08,640

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

422

00:48:08,640 --> 00:48:14,720

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

423

00:48:14,720 --> 00:48:20,720

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

424

00:48:20,720 --> 00:48:26,560

have been removed or have restricted access. But it can give you sample tweets, basic statistics,



425

00:48:26,560 --> 00:48:32,320

interactive maps, the density of tweeting in a specific region and you can do things like

426

00:48:32,320 --> 00:48:38,560

a sentiment analysis - you know, happy, sad, and so on. So that's something else you could have a look at

427

00:48:38,560 --> 00:48:46,000

for your own research or for students. We took some of this data and we tried to create a story

428

00:48:46,000 --> 00:48:52,800

about the meaning of lifewide literacies for better mental health, well-being

429

00:48:52,800 --> 00:48:59,440

and physical health and we took it out to places like IKEA, the Mugdock Country Park for MoSSFest,

430

00:48:59,440 --> 00:49:05,280

for all kinds of social science festivals. And we created these little lifewide literacies

431

00:49:05,280 --> 00:49:11,360

badges. So, you've got here this dinosaur, reading, theatre. So, you've got regular reading literacy,

432

00:49:11,360 --> 00:49:18,560

you've got the creative arts or cultural literacy and foreign language literacy.

433

00:49:18,560 --> 00:49:24,400

And this one is holding a green leaf, that's environmental literacy. And so, children and

434

00:49:24,400 --> 00:49:30,320

adults were able to customise their badges and create their own literacy person. But it was just

435

00:49:30,320 --> 00:49:36,640

a way of moving from the data to actually communicating our messages with people. So

436

00:49:37,200 --> 00:49:43,280

in a sense the whole story of our research and what we're doing here today is really about the

437

00:49:43,280 --> 00:49:50,800

data literacy. And I told you we've measured ICT, computer use and attitudes towards technology.

438

00:49:50,800 --> 00:49:58,240

But data literacy, again, is an empowerment strategy. We wrote a briefing paper for UNESCO for their

439

00:49:58,240 --> 00:50:03,680

last learning cities conference and that should be released as a publication shortly. And it's about

440

00:50:03,680 --> 00:50:10,560

inclusion in data literacy. As we've had to move increasingly online, we really need to look more

441

00:50:10,560 --> 00:50:19,520

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

442

00:50:19,520 --> 00:50:25,840

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

443

00:50:26,400 --> 00:50:31,760

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

444

00:50:31,760 --> 00:50:38,720

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

445

00:50:38,720 --> 00:50:45,200

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

446

00:50:45,200 --> 00:50:51,520

with the push for openness. So, we are mindful that this is possibly more ethically fraught than

447

00:50:51,520 --> 00:50:57,760

we're making it out to be today, but we just hope that we've shared with you some ways in which

448

00:50:57,760 --> 00:51:03,520

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

449

00:51:03,520 --> 00:51:09,680

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

450

00:51:09,680 --> 00:51:14,400

education data acquisition. We've got some higher education data including some UCAS

451

00:51:14,400 --> 00:51:23,600

lookup tables. But there's loads of training webinars and seminars going on at UBDC. So

452

00:51:23,600 --> 00:51:31,200

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

453

00:51:31,200 --> 00:51:38,400

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

454

00:51:38,400 --> 00:51:44,800

cycling and mobility data, we've got all types of things. So why not consider us to help support

455

00:51:44,800 --> 00:51:53,440

you in your work in future? So that is it from us as the team. I think I've really tried to tackle

456

00:51:53,440 --> 00:52:00,720

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

457

00:52:00,720 --> 00:52:08,400

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

458

00:52:08,400 --> 00:52:15,200

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

459

00:52:15,200 --> 00:52:21,520

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

460

00:52:21,520 --> 00:52:26,560

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

461

00:52:26,560 --> 00:52:32,720

and I should have referenced that as the work of Susan Halford, and I've put a link to their work

462

00:52:32,720 --> 00:52:41,440

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 resource.

464

00:52:49,120 --> 00:52:54,960

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

00:52:54,960 --> 00:52:59,840

the telephone and you do one of those complaint forms, it's just that quick and simple. What data

466

00:52:59,840 --> 00:53:04,400

acquisition are you interested in? if you're interested in iMCD what strand of it is it? -

467

00:53:04,400 --> 00:53:10,160

the survey, the travel diary, the survey and the travel diary, the GPS, the lifelogging images

468

00:53:10,160 --> 00:53:16,720

or some more in-depth twitter data that you can't get off the Twitter dashboard. So be sure, as Muir

469

00:53:16,720 --> 00:53:23,920

says, to look at the other two webinars that are happening. We've had a small attendance but

470

00:53:23,920 --> 00:53:30,640

really insightful questions. And so, when this video gets up, why not apply for the data

471

00:53:30,640 --> 00:53:36,640

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

472

00:53:36,640 --> 00:53:41,200

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

473

00:53:41,200 --> 00:53:45,680

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

474

00:53:45,680 --> 00:53:51,360

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

475

00:53:51,360 --> 00:53:57,040

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

476

00:53:57,040 --> 00:54:03,280

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

477

00:54:03,280 --> 00:54:10,240

and I think we've helped you make that start. So, we're bang on the one hour. I will just say

478

00:54:10,240 --> 00:54:14,880

thank you so much for attending. Muir or Rachel, do you want to say anything to conclude?

479

00:54:16,640 --> 00:54:22,640

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

480

00:54:22,640 --> 00:54:29,600

Strava cycling data in the Glasgow city area. And, as Catherine has said, we will be putting these

481

00:54:29,600 --> 00:54:35,920

videos online but for accessibility purposes we need to provide them with a transcript and various

482

00:54:35,920 --> 00:54:42,480

accessibility measures to ensure compliance with current regulation. So, it might take us a wee

483

00:54:42,480 --> 00:54:47,520

bit to do the technical background stuff for that sort of stuff. But, as I've said, there's two more

484

00:54:47,520 --> 00:54:54,960

coming up in the series. If you look on the UBDC site and again, as with this one, free registration.

485



00:54:54,960 --> 00:55:01,680

So, thanks for taking part. [Catherine Lido] Thank you everybody, have a great day, thanks so much.

486

00:55:01,680 --> 00:55:02,200

Bye.