

# Exploring and analysing open transport data (with a focus on SCOOT data and bikeshare data)

Transcript from webinar video recording

1

00:00:01,643 --> 00:00:03,353

This is about, as I've said,

2

00:00:03,453 --> 00:00:08,862

this is about exploring and analysing the open transport data sets

3

00:00:09,700 --> 00:00:14,119

that are easily available to us these days.

4

00:00:14,338 --> 00:00:18,469

And, of course, it's not going to be exhaustive.

5

00:00:20,678 --> 00:00:23,079

What I'm trying to do is give some examples

6

00:00:23,179 --> 00:00:26,120

so that we know where to look

7

00:00:26,220 --> 00:00:30,919

when we would like to find out more about our environment

8

00:00:31,019 --> 00:00:34,499

and the transport issues around us.

9

00:00:34,689 --> 00:00:35,950

So, that's the goal.

10

00:00:36,050 --> 00:00:38,587

It's not like I can give an exhaustive list of

11

00:00:38,687 --> 00:00:40,716

all these different data sets.

12

00:00:40,816 --> 00:00:43,495

That's not the case. But I try to introduce

13

00:00:43,595 --> 00:00:48,124

some in a way such that it becomes more accessible.

14

00:00:49,133 --> 00:00:51,273

So, okay, let's get started.

15

00:00:51,373 --> 00:00:55,173

And then, the overview of the session.

16

00:00:55,273 --> 00:00:58,303

I plan to not use up the whole two hours.

17

00:00:58,403 --> 00:01:02,833

I find that to be a bit too much for both the audience and myself, of course.

18

00:01:03,300 --> 00:01:05,778

So, the first part,

19

00:01:05,878 --> 00:01:10,608

I will try to keep to time and finish before 10:45.

20

00:01:10,708 --> 00:01:15,407

And we will focus on this presentation

21

00:01:15,755 --> 00:01:20,734

where I'll try to introduce some of the sources of open data

22

00:01:21,183 --> 00:01:26,133

and then, how do we obtain them

23

00:01:26,233 --> 00:01:29,193

and what do they look like in general.

24

00:01:29,353 --> 00:01:33,361

And then, very briefly, I will go through a few applications that

25

00:01:33,461 --> 00:01:36,378

I came across in these days.

26

00:01:36,587 --> 00:01:39,846

You know, we all live in a pretty strange time these days

27

00:01:39,946 --> 00:01:41,965

with COVID and everything.

28

00:01:44,739 --> 00:01:48,769

For these applications, I'll focus on the later two,

29

00:01:48,869 --> 00:01:51,034

which are about COVID and mobility

30

00:01:51,134 --> 00:01:54,163

and then this is actually demonstrated,

31

00:01:55,609 --> 00:02:00,649

the use of this data set was demonstrated by a blog post

32

00:02:00,858 --> 00:02:04,258

on the website of the Urban Big Data Centre

33

00:02:04,358 --> 00:02:06,239

written by our colleague.

34

00:02:07,100 --> 00:02:09,720

And then, the other one would be

35

00:02:09,820 --> 00:02:16,939

a paper that is, I think, quite close to publication

36

00:02:17,039 --> 00:02:21,848

and it will be about the shared bikes in Glasgow.

37

00:02:22,016 --> 00:02:24,785

And then, the trip data that we use

38

00:02:24,885 --> 00:02:29,532

and how from there we turned it into a paper.

39

00:02:29,774 --> 00:02:31,713

So, that's what we are going to do.

40

00:02:33,858 --> 00:02:37,507

Then, after a short break, we will go into the practical session.

41

00:02:37,748 --> 00:02:42,431

And then, if you have asked to do, you are welcome to follow along.

42

00:02:42,531 --> 00:02:44,334

If you don't, it's still fine.

43

00:02:45,048 --> 00:02:48,715

I will share the script and the presentation afterwards.

44

00:02:49,495 --> 00:02:51,601

So, okay, let's get started.

45

00:02:51,926 --> 00:02:55,463

A little bit about me, not too much.

46

00:02:55,563 --> 00:02:57,512

I'm currently a research associate

47

00:02:57,612 --> 00:03:01,338

in transport analytics at the Urban Big Data Centre of

48

00:03:01,500 --> 00:03:03,046

University of Glasgow.

49

00:03:03,395 --> 00:03:07,253

So, during my PhD and after my PhD,

50

00:03:07,481 --> 00:03:11,796

what I've been interested in is mainly transport taxes such as

51

00:03:11,896 --> 00:03:18,024

road pricing, and then optimal public transport supply,

52

00:03:18,124 --> 00:03:22,821

and also accessibility, which I'm currently working on.

53

00:03:23,360 --> 00:03:27,956

And a few publications that I had

54

00:03:28,056 --> 00:03:33,544

with my colleagues are also about these issues, mainly.

55

00:03:34,093 --> 00:03:36,410

The first one is about rural areas.

56

00:03:39,584 --> 00:03:44,205

Our goal was to evaluate whether it is a good idea

57

00:03:44,305 --> 00:03:48,133

to have these kinds of very expensive rail systems

58

00:03:48,233 --> 00:03:52,320

in rural areas with very low-density populations.

59

00:03:52,420 --> 00:03:57,076

And then, sometimes replacing it with buses is actually a better idea.

60



00:03:57,283 --> 00:04:01,490

And then, the case study was from Sweden

61

00:04:01,590 --> 00:04:06,457

because one of our co-authors is from Sweden

62

00:04:06,557 --> 00:04:08,543

so we get access to the data.

63

00:04:09,109 --> 00:04:11,336

And then, the other one, which is

64

00:04:12,274 --> 00:04:15,679

also using Swedish data,

65

00:04:17,534 --> 00:04:22,260

we evaluated whether buses

66

00:04:23,955 --> 00:04:30,087

and cyclists, who are hindering who.

67

00:04:30,235 --> 00:04:32,862

It's very difficult to say because

68

00:04:33,097 --> 00:04:35,004

if you imagine that,

69

00:04:36,021 --> 00:04:39,548

I think some drivers are pretty annoyed by cyclists.

70

00:04:40,513 --> 00:04:42,470

But then, in the case of buses

71

00:04:42,570 --> 00:04:47,457

in a city area, like the city centre of Stockholm,

72

00:04:47,642 --> 00:04:51,488

with a bus line, they have to go to the bus stops

73

00:04:51,588 --> 00:04:54,675

and then stop at the bus stops.

74

00:04:54,775 --> 00:04:57,241

And what happened is that

75

00:04:58,981 --> 00:05:01,577

the cycle lane suddenly disappeared,

76

00:05:01,677 --> 00:05:05,894

so the cyclists would have to move around, if you see what I mean.

77

00:05:05,994 --> 00:05:08,520

They would have to switch to another lane.

78

00:05:12,853 --> 00:05:18,443

Then they would have to interfere with the other traffic in the outer lanes.

79

00:05:18,755 --> 00:05:19,877

So, that's the idea.

80

00:05:19,977 --> 00:05:22,954

And then we tried to evaluate this interaction.

81

00:05:23,460 --> 00:05:26,297

And then, in this paper, what we do is

82

00:05:26,397 --> 00:05:30,204

we usually use a more aggregate model.

83

00:05:30,400 --> 00:05:34,276

So, the key as to whether these models

84

00:05:34,376 --> 00:05:38,322

work well or not would be

85

00:05:38,422 --> 00:05:42,274

to evaluate the interactions between different modes.

86

00:05:42,374 --> 00:05:43,780

So, that's usually the key.

87

00:05:43,880 --> 00:05:45,186

How would one mode...

88

00:05:45,286 --> 00:05:48,743

Like, okay, what if there is road pricing?

89

00:05:49,150 --> 00:05:54,308

How will the ridership of public transport react?

90

00:05:54,400 --> 00:05:56,167

So, that's the key idea.

91

00:05:56,267 --> 00:05:58,504

The interaction between modes.

92

00:05:59,890 --> 00:06:02,407

Why do I talk so much about this?

93

00:06:02,507 --> 00:06:04,834

Because it is actually related to

94

00:06:04,934 --> 00:06:07,440

the paper that I'm going to talk about a bit later,

95

00:06:07,540 --> 00:06:10,383

about the interaction between

96

00:06:11,118 --> 00:06:15,089

the subway system and the shared bike system in Glasgow.

97

00:06:15,669 --> 00:06:18,705

So, okay, a bit about open data.

98

00:06:22,130 --> 00:06:26,418

This is just a very brief introduction to it.

99

00:06:27,678 --> 00:06:29,873

In the title of this session,

100

00:06:29,973 --> 00:06:33,249

I actually wrote open data, I think.

101

00:06:33,338 --> 00:06:35,954

But then, you need to be really careful

102

00:06:36,054 --> 00:06:40,065

that not all of these are Open Government License

103

00:06:40,165 --> 00:06:42,250

where you can pretty much,

104

00:06:44,140 --> 00:06:46,794

roughly speaking, you can reuse them,

105

00:06:46,894 --> 00:06:50,240

you can basically redistribute them most of the time.

106

00:06:50,755 --> 00:06:55,060

But then, you have to be really careful with the licensing terms

107

00:06:55,160 --> 00:06:57,655

when it comes to some of the data

108

00:06:57,755 --> 00:07:02,221

that does not necessarily have an open licence.

109

00:07:02,514 --> 00:07:06,279

So, yeah, you can see that many of them would be

110

00:07:06,455 --> 00:07:09,264

Open Government Licence later on.

111

00:07:11,483 --> 00:07:17,285

So, I think that many of these data sources

112

00:07:17,385 --> 00:07:20,651

that we are going to talk about,

113

00:07:21,976 --> 00:07:27,494

it's usually quite straightforward in that we know what they are.

114

00:07:27,800 --> 00:07:29,819

These tools, some are very straightforward,

115

00:07:29,919 --> 00:07:31,828

like Department for Transport,

116

00:07:31,928 --> 00:07:35,827

Transport for London, they have a very, very good page

117

00:07:35,927 --> 00:07:40,095

with many different data sets.

118

00:07:40,594 --> 00:07:43,932

And then, for some of the city council portals,

119

00:07:45,360 --> 00:07:47,939

some of them I'm also going to

120

00:07:48,268 --> 00:07:51,513

talk more about in the second half of the session,

121

00:07:54,572 --> 00:07:57,123

sometimes you might need to register

122



00:07:57,223 --> 00:08:01,101

to get a key or a login name to use them.

123

00:08:01,201 --> 00:08:07,915

And it's not exactly that easy because you might not

124

00:08:08,015 --> 00:08:12,283

get it in time when you try to use the data.

125

00:08:12,383 --> 00:08:15,991

And sometimes the site might not be working very well.

126

00:08:16,198 --> 00:08:19,428

So, let's see how we get on with it today.

127

00:08:19,528 --> 00:08:22,776

I believe that we included a link

128

00:08:22,991 --> 00:08:28,003

for one of the APIs for the Glasgow City Council data.

129

00:08:28,103 --> 00:08:33,041

But then I found out that link might not work.

130

00:08:33,141 --> 00:08:34,889

So, let's see.

131

00:08:35,557 --> 00:08:39,479

Another very important source, of course, is

132

00:08:39,579 --> 00:08:44,668

the Urban Big Data Centre collection of data that

133

00:08:44,768 --> 00:08:48,046

if you take a look,

134

00:08:48,521 --> 00:08:49,807

very briefly.

135

00:08:49,907 --> 00:08:51,885

And you can see that

136

00:08:57,040 --> 00:09:00,741

we acquire a pretty good collection of data,

137

00:09:00,841 --> 00:09:03,630

such as this CCTV data

138

00:09:03,730 --> 00:09:06,217

where they process the image and the video

139

00:09:06,317 --> 00:09:11,407

such that you could, if you specified the location,

140

00:09:11,507 --> 00:09:15,124

you could get a count of the pedestrians or cyclists

141

00:09:15,224 --> 00:09:16,982

or vehicles and so on.

142

00:09:17,640 --> 00:09:21,266

And some of the very new stuff, other than that,

143

00:09:22,117 --> 00:09:24,916

related to transport is this Huq data set,

144

00:09:25,232 --> 00:09:27,693

which is mobile phone data

145

00:09:27,920 --> 00:09:30,779

and very comprehensive and pretty detailed

146

00:09:30,879 --> 00:09:35,118

whenever the users use a certain app.

147

00:09:35,373 --> 00:09:38,124

Not a certain app. It's a collection of apps

148

00:09:38,224 --> 00:09:42,532

that the company works with

149

00:09:42,632 --> 00:09:46,209

so that we could have the location of

150

00:09:46,309 --> 00:09:48,767

the users.

151

00:09:50,265 --> 00:09:52,444

So, this could be pretty useful

152

00:09:52,762 --> 00:09:55,701

in terms of mobility research.

153

00:09:56,270 --> 00:09:58,608

And you could go through this

154

00:09:58,815 --> 00:10:01,783

when you have time and applied.

155

00:10:03,029 --> 00:10:05,250

Many of them are not open licence

156

00:10:05,350 --> 00:10:07,678

but then, of course, you can apply to use them

157

00:10:07,778 --> 00:10:09,856

from the Urban Big Data Centre

158

00:10:09,956 --> 00:10:12,140

and then be careful with the licencing part

159

00:10:12,240 --> 00:10:14,158

and see what you can do with them.

160

00:10:14,688 --> 00:10:18,206

And also, one thing that I found is

161

00:10:18,306 --> 00:10:19,335

GitHub.

162

00:10:21,904 --> 00:10:23,571

At first, I didn't realise.

163

00:10:23,671 --> 00:10:25,268

I know that GitHub exists

164

00:10:25,368 --> 00:10:27,107

but then I found out that now,

165

00:10:28,071 --> 00:10:30,930

sometimes when you browse through GitHub,

166

00:10:31,030 --> 00:10:34,077

you find different packages and different scripts.

167

00:10:34,177 --> 00:10:39,494

And that's led to some surprising discoveries

168

00:10:39,594 --> 00:10:41,143

with data sets.

169

00:10:41,299 --> 00:10:42,490

That's my opinion.

170

00:10:42,590 --> 00:10:44,529

So, you might look through that later.

171

00:10:44,763 --> 00:10:47,289

And for the map data,

172

00:10:48,047 --> 00:10:50,796

Google Maps, of course, they have a lot of

173

00:10:53,075 --> 00:10:56,534

different data that you could use,

174

00:10:56,634 --> 00:10:58,414

like distance matrices,

175

00:10:58,514 --> 00:11:00,113

you could specify the route

176

00:11:00,213 --> 00:11:02,321

and you can calculate the travel time

177

00:11:02,421 --> 00:11:05,259

and report it in pretty much real-time.

178

00:11:05,639 --> 00:11:08,408

But then, you will have to be very careful

179

00:11:08,508 --> 00:11:10,986

because you have to enter your credit card

180

00:11:11,524 --> 00:11:14,593

in order to use the API

181

00:11:14,693 --> 00:11:17,817

and you would not want to request too much

182

00:11:17,917 --> 00:11:21,135

and end up having to pay a whole lot for that.

183

00:11:22,025 --> 00:11:24,163

And also, OpenStreetMap.

184



00:11:24,863 --> 00:11:27,648

That's also a very good source

185

00:11:27,748 --> 00:11:32,176

for things like points of interest and so on.

186

00:11:32,573 --> 00:11:34,977

And if you were from academia

187

00:11:35,077 --> 00:11:38,295

or you were a student, if I remember correctly,

188

00:11:38,395 --> 00:11:41,372

you can also register to use Digimap.

189

00:11:42,583 --> 00:11:44,161

So, these kinds of data sources,

190

00:11:44,261 --> 00:11:47,438

they could be real-time or historical.

191

00:11:47,552 --> 00:11:50,100

Historical is simpler most of the time

192

00:11:50,200 --> 00:11:51,837

if you know it's historical.

193

00:11:51,937 --> 00:11:56,002

Of course, they would just dump the data on the site

194

00:11:56,102 --> 00:11:58,769

and then you can just click download and you get it.

195

00:11:59,346 --> 00:12:01,609

It's usually more straightforward.

196

00:12:01,709 --> 00:12:03,157

But then, for real-time data,

197

00:12:03,257 --> 00:12:05,584

usually it's more complicated because

198

00:12:05,793 --> 00:12:08,658

it gets updated a lot.

199

00:12:09,248 --> 00:12:11,846

Like, every minute or every five minutes.

200

00:12:12,146 --> 00:12:14,143

And if you want to get it

201

00:12:14,243 --> 00:12:15,750

or research your purpose,

202

00:12:16,807 --> 00:12:19,047

then you will have to make use of

203

00:12:20,305 --> 00:12:22,383

some simple tools in order to

204

00:12:22,581 --> 00:12:24,830

just get this data over time.

205

00:12:25,300 --> 00:12:27,168

And how do we get it?

206

00:12:27,268 --> 00:12:30,165

Of course, yeah, when I was saying about historical data,

207

00:12:30,265 --> 00:12:31,904

most of the time you can download them.

208

00:12:32,004 --> 00:12:34,011

And then, for the real-time ones,

209

00:12:34,111 --> 00:12:35,637

most of the time you will need

210

00:12:35,737 --> 00:12:37,984

to make use of API

211

00:12:38,084 --> 00:12:40,322

or, if you are familiar with it,

212

00:12:40,422 --> 00:12:42,404

you can do some web scraping.

213

00:12:42,554 --> 00:12:44,683

So, in the second half of the session,

214

00:12:44,783 --> 00:12:46,859

we will try to use R Studio

215

00:12:47,486 --> 00:12:52,340

to get access to some of this real-time data,

216

00:12:52,440 --> 00:12:54,487

like every five minutes,

217

00:12:54,587 --> 00:12:58,213

what is the count of traffic in certain locations.

218

00:13:00,662 --> 00:13:02,041

Okay. So, yeah.

219

00:13:02,141 --> 00:13:05,809

Cars. We briefly talked about this.

220

00:13:05,909 --> 00:13:07,525

The traffic count.

221

00:13:07,697 --> 00:13:08,763

We have it.

222

00:13:09,900 --> 00:13:12,357

The reasons that I put Brussels here is

223

00:13:12,457 --> 00:13:16,754

because I planned to use

224

00:13:16,854 --> 00:13:21,576

the Glasgow traffic counts

225

00:13:21,676 --> 00:13:24,623

and the data sets as a demonstration,

226

00:13:24,723 --> 00:13:26,317

but then I found out that

227

00:13:27,618 --> 00:13:30,889

you may not be able to get an API key in time.

228

00:13:30,977 --> 00:13:34,265

So, I think that, to avoid that issue,

229

00:13:34,801 --> 00:13:36,245

to avoid registration,

230

00:13:36,441 --> 00:13:39,114

we could try the Brussels one later

231

00:13:39,295 --> 00:13:41,142

because it doesn't require a key.

232

00:13:41,242 --> 00:13:43,118

You can just request it

233

00:13:44,086 --> 00:13:46,691

using two lines in R

234

00:13:47,353 --> 00:13:49,110

and then you get the count.

235

00:13:50,206 --> 00:13:52,064

Okay. And also,

236

00:13:52,430 --> 00:13:54,257

this is also real-time, obviously.

237

00:13:54,357 --> 00:13:56,423

You want to know about parking availability

238

00:13:56,523 --> 00:13:58,430

and you want to know it,

239

00:13:58,906 --> 00:14:02,717

to be updated in real-time.

240

00:14:03,697 --> 00:14:07,575

Sometimes, traffic events might also be interesting

241

00:14:07,773 --> 00:14:13,356

and you might want to know what happened in a certain part of

242

00:14:13,456 --> 00:14:16,284

the motorway, what happened with that part.

243

00:14:16,853 --> 00:14:17,967

If there were accidents

244

00:14:18,067 --> 00:14:20,568

or some sort of closure

245

00:14:21,067 --> 00:14:22,725

that we might want to know about.

246



00:14:23,065 --> 00:14:25,082

And then, the next thing would be

247

00:14:25,182 --> 00:14:27,020

some Variable Message System.

248

00:14:29,300 --> 00:14:32,106

That's just the message boards you can see

249

00:14:32,206 --> 00:14:37,144

when you are on the motorway.

250

00:14:39,609 --> 00:14:44,540

When we try to look at them in the second half of the session,

251

00:14:44,640 --> 00:14:46,928

then we could also show some of

252

00:14:47,028 --> 00:14:49,716

these messages, what they say.

253

00:14:49,976 --> 00:14:52,443

Like, over time, during COVID time,

254

00:14:52,543 --> 00:14:55,705

they ask you to stay safe and stay home and so on.

255

00:14:56,044 --> 00:14:57,682

And then, one thing that I was...

256

00:14:58,682 --> 00:15:02,610

There's actually a picture of the Variable Message sign here.

257

00:15:02,710 --> 00:15:05,047

You can see you can specify the location

258

00:15:05,665 --> 00:15:10,414

and then you get a message about the roadworks at a certain point.

259

00:15:11,023 --> 00:15:16,121

And another thing that I have been quite interested in are the potholes.

260

00:15:16,221 --> 00:15:18,129

And yeah, I find that quite interesting.

261

00:15:18,229 --> 00:15:23,860

FixMyStreet is actually not only about potholes.

262

00:15:24,550 --> 00:15:28,258

But then, I actually laughed a bit when I saw this one.

263

00:15:28,358 --> 00:15:30,907

You can see that people are reporting potholes

264

00:15:31,007 --> 00:15:32,395

because they are so pissed off.

265

00:15:32,495 --> 00:15:36,164

And they report these kinds of dangerous potholes.

266

00:15:36,264 --> 00:15:37,920

And they take pictures of them.

267

00:15:38,020 --> 00:15:40,239

And then, if you read it carefully,

268

00:15:40,339 --> 00:15:41,517

if you try to read this,

269

00:15:41,617 --> 00:15:43,675

you can even see that update.

270

00:15:44,259 --> 00:15:47,547

I thought that it would be from the city council or something, but no.

271

00:15:47,647 --> 00:15:49,780

It's actually from another user,

272

00:15:49,880 --> 00:15:53,308

saying that there's not a chance they'll repair it or deal with it.

273

00:15:53,616 --> 00:15:55,695

So, it has been quite interesting.

274

00:15:55,795 --> 00:15:58,642

So, if you were looking at a road network,

275

00:15:58,742 --> 00:16:02,806

I suppose this information about

276

00:16:02,906 --> 00:16:05,373

the quality of the road would also be

277

00:16:05,473 --> 00:16:08,280

quite useful.

278

00:16:09,514 --> 00:16:15,851

So, of course, accident reports are also

279

00:16:15,951 --> 00:16:17,530

part of the picture.

280

00:16:17,958 --> 00:16:19,405

And, yeah.

281

00:16:19,505 --> 00:16:21,432

Then we move on to another mode

282

00:16:21,532 --> 00:16:23,929

with bikes and pedestrians.

283

00:16:24,422 --> 00:16:28,429

Then we have this, again, from UBDC.

284

00:16:28,529 --> 00:16:32,967

I think we saw that when I was showing the website of UBDC.

285

00:16:33,296 --> 00:16:39,023

Where you have this API that you can request

286

00:16:40,248 --> 00:16:44,909

the count of pedestrians and other vehicles or cyclists

287

00:16:45,018 --> 00:16:47,034

over a certain period of time.

288

00:16:49,352 --> 00:16:53,691

You could also go over with cycling infrastructure.

289

00:16:53,987 --> 00:16:56,333

These are links. But I don't think I will

290

00:16:56,433 --> 00:16:59,505

show every one of them, but I will share it later.

291

00:16:59,975 --> 00:17:04,989

And then, one focus that we will be

292

00:17:06,124 --> 00:17:09,351

talking about in the second half of this session

293

00:17:09,451 --> 00:17:13,139

and also, it would be related to one of my papers is

294

00:17:14,442 --> 00:17:16,663

the use of shared bike data.

295

00:17:17,231 --> 00:17:19,250

And there are usually,

296

00:17:19,350 --> 00:17:22,367

if you go on a site of a shared bike company

297

00:17:22,467 --> 00:17:23,695

about their data sets,

298

00:17:23,795 --> 00:17:25,652

there would be two different links.

299

00:17:25,752 --> 00:17:27,471

Usually, they have this data dump

300

00:17:27,571 --> 00:17:30,759

where they have a lot of files

301

00:17:30,859 --> 00:17:33,907

about the details of every trip.

302

00:17:34,007 --> 00:17:37,276

When a bike is unlocked and when it is returned.

303

00:17:37,595 --> 00:17:38,723

And so on.

304

00:17:40,500 --> 00:17:43,879

I find this to be the most useful one

305

00:17:43,979 --> 00:17:47,008

and then we make use of this

306

00:17:47,108 --> 00:17:48,726

in one of our papers.

307

00:17:49,093 --> 00:17:52,113

And the API for shared bike is

308



00:17:52,382 --> 00:17:54,067

actually quite easy to use.

309

00:17:54,167 --> 00:17:59,094

And if you are interested in the availability of these kinds of bike,

310

00:18:00,713 --> 00:18:04,143

or how these bikes are arranged

311

00:18:04,243 --> 00:18:07,862

or moved by the company,

312

00:18:08,874 --> 00:18:12,134

or how the demand is different,

313

00:18:12,234 --> 00:18:14,790

then this is also quite useful.

314

00:18:14,957 --> 00:18:18,006

But then, from this API, as you can see later,

315

00:18:18,106 --> 00:18:22,624

when we actually tried it, we could see the location,

316

00:18:23,341 --> 00:18:25,459

the co-ordinates of the bike station,

317

00:18:25,559 --> 00:18:27,288

and also how many bikes are there

318

00:18:27,388 --> 00:18:29,596

or how many spaces are there,

319

00:18:31,255 --> 00:18:33,683

instead of other trip information

320

00:18:34,044 --> 00:18:37,003

from the file that you can download as a data dump.

321

00:18:39,442 --> 00:18:42,631

I find that it's also an interesting point

322

00:18:42,731 --> 00:18:45,629

if you are interested in bike sharing.

323

00:18:46,896 --> 00:18:51,674

If we look at this from a different point of view.

324

00:18:51,774 --> 00:18:56,043

Like, okay, this is in Edinburgh.

325

00:18:57,820 --> 00:19:01,040

You can see that you can download the trip data.

326

00:19:01,482 --> 00:19:06,386

And this is what I was referring to as a data dump.

327

00:19:07,700 --> 00:19:10,039

You can see you have all this information.

328

00:19:10,139 --> 00:19:12,168

Like stop and start station,

329

00:19:12,575 --> 00:19:13,853

description,

330

00:19:14,053 --> 00:19:17,760

co-ordinates, and the time, and so on.

331

00:19:18,190 --> 00:19:21,710

So, you have very detailed information for each trip.

332

00:19:22,141 --> 00:19:25,259

So, if you go on this data,

333

00:19:25,359 --> 00:19:27,597

that's what you get, and then you also get

334

00:19:27,697 --> 00:19:31,721

real-time data about the station's availability

335

00:19:31,821 --> 00:19:33,430

that I was referring to.

336

00:19:34,036 --> 00:19:39,522

But then, if you click on some of the US bike share data,

337

00:19:39,622 --> 00:19:41,942

I think it's also quite interesting because

338

00:19:43,204 --> 00:19:44,852

it looks very similar.

339

00:19:44,952 --> 00:19:46,997

But then, what strikes me is that

340

00:19:47,080 --> 00:19:51,846

they usually provide more information.

341

00:19:51,946 --> 00:19:54,235

I guess this is partly due to

342

00:19:55,233 --> 00:19:56,412

GDPR.

343

00:19:56,512 --> 00:20:00,580

That's my guess. I haven't really looked into it.

344

00:20:00,700 --> 00:20:04,069

But then, for the ones outside Europe,

345

00:20:04,169 --> 00:20:08,148

you usually get something more detailed.

346

00:20:08,248 --> 00:20:11,186

Such as the user type, even the birth year

347

00:20:11,286 --> 00:20:13,255

and the gender of the user.

348

00:20:14,743 --> 00:20:18,428

So, it depends on what you need in your research.

349

00:20:20,088 --> 00:20:23,055

That's a point that is worth noting.

350

00:20:24,533 --> 00:20:26,603

So, public transport.

351

00:20:27,044 --> 00:20:30,271

It's actually, for me, a bit painful

352

00:20:30,371 --> 00:20:34,309

because it's not very easy to obtain data.

353

00:20:34,409 --> 00:20:39,425

It's mostly not shared amongst the public.

354

00:20:42,101 --> 00:20:46,689

And most of the time, you will need to obtain them directly from the company.

355

00:20:46,789 --> 00:20:49,278

And then, as you can see,

356

00:20:49,378 --> 00:20:56,938

if they are not in a working relationship with

357

00:20:57,038 --> 00:20:58,606

you as a researcher

358

00:20:58,706 --> 00:21:02,824

or if you are not a consultant working for them,

359

00:21:02,924 --> 00:21:06,259

then it's not very easy to get sensitive information.

360

00:21:06,359 --> 00:21:08,745

But then, one thing that is open, at least, is

361

00:21:08,845 --> 00:21:12,169

the timetable or even, most of the time,

362

00:21:14,996 --> 00:21:21,123

they will share the real-time location of their vehicles.

363

00:21:21,223 --> 00:21:23,001

So, that's also possible.

364

00:21:24,284 --> 00:21:27,291

And for the prices of their fares,

365

00:21:27,589 --> 00:21:30,257

it's a bit easier because

366

00:21:30,357 --> 00:21:34,595

you can basically scrape them directly from the web.

367

00:21:34,976 --> 00:21:38,914

Or in the case of rail here,

368

00:21:39,014 --> 00:21:40,202

Rail Delivery Group,

369

00:21:40,302 --> 00:21:44,790

you can also register and get this information from them.

370



00:21:45,289 --> 00:21:49,729

So, I think this is the most constrained area

371

00:21:49,829 --> 00:21:53,463

in terms of getting data that is available to the public.

372

00:21:53,993 --> 00:21:56,241

We don't know that much about them.

373

00:21:56,926 --> 00:22:00,676

Especially in this country, with the market structure of

374

00:22:01,244 --> 00:22:02,652

the operators.

375

00:22:02,752 --> 00:22:07,641

That's pretty unfortunate because it's a big missing piece of the puzzle,

376

00:22:07,791 --> 00:22:08,860

I would say.

377

00:22:09,499 --> 00:22:12,298

Someone said freight data is even more constrained.

378

00:22:12,604 --> 00:22:15,923

Yeah. Yeah, that's even more unfortunate.

379

00:22:19,101 --> 00:22:21,130

Okay. So, I'm just thinking that,

380

00:22:21,230 --> 00:22:24,139

okay, with Glasgow, with these kinds of data,

381

00:22:24,778 --> 00:22:26,575

if we put them all together,

382

00:22:26,675 --> 00:22:27,863

we have different modes.

383

00:22:27,963 --> 00:22:30,991

And, of course, it's not like a complete picture.

384

00:22:31,091 --> 00:22:37,056

But we have a much better picture if we gather them together

385

00:22:37,156 --> 00:22:39,795

and try to build

386

00:22:40,899 --> 00:22:43,789

models, including different modes.

387

00:22:44,487 --> 00:22:46,683

And for the mobile phone data,

388

00:22:47,552 --> 00:22:49,467

again, I highlighted it because

389

00:22:49,567 --> 00:22:53,385

UBDC, they are having this mobile phone data set.

390

00:22:54,256 --> 00:22:56,498

So, yeah, perhaps if you are interested,

391

00:22:56,598 --> 00:22:59,096

just check it out and apply to use some.

392

00:23:02,510 --> 00:23:05,850

I will move on to some of the applications.

393

00:23:06,237 --> 00:23:08,295

One thing that I've found about

394

00:23:09,043 --> 00:23:15,702

this type of data available to the public is

395

00:23:15,802 --> 00:23:21,110

this I found during times of lockdown.

396

00:23:21,498 --> 00:23:25,195

I read an article, I think it was in the Financial Times,

397

00:23:26,353 --> 00:23:28,132

or The Economist.

398

00:23:28,628 --> 00:23:31,249

Both of them did something pretty similar.

399

00:23:31,349 --> 00:23:34,006

They tried to compare the hotspots

400

00:23:34,106 --> 00:23:37,054

during lockdown and so on.

401

00:23:37,489 --> 00:23:40,448

It was very early in the lockdown,

402

00:23:40,548 --> 00:23:43,846

like, last year, in March or so.

403

00:23:44,992 --> 00:23:47,550

So, there is actually this...

404

00:23:47,948 --> 00:23:50,954

When you search for a place on Google,

405

00:23:51,621 --> 00:23:54,168

for example, this is Glasgow Botanic Gardens,

406

00:23:54,268 --> 00:23:57,219

which is like five minutes from where I am now,

407

00:23:57,461 --> 00:24:01,470

what you will get is you get some popular times

408

00:24:01,570 --> 00:24:09,052

and you see what times at this place are

409

00:24:09,152 --> 00:24:11,062

usually the busier times.

410

00:24:11,870 --> 00:24:17,293

And then, what I did is that I also found something pretty interesting,

411

00:24:17,393 --> 00:24:20,705

which is a GitHub repository,

412

00:24:21,072 --> 00:24:25,502

which I didn't write,

413

00:24:25,673 --> 00:24:31,213

that enabled us to download this data

414

00:24:31,313 --> 00:24:34,685

in real-time so we get an idea about

415

00:24:35,017 --> 00:24:38,600

how busy a certain location is.

416

00:24:38,700 --> 00:24:41,440

As long as you can get the location ID,

417

00:24:41,596 --> 00:24:43,046

the details are on the GitHub.

418

00:24:43,146 --> 00:24:44,827

It's the link here.

419

00:24:45,500 --> 00:24:46,690

I'll just press it.

420

00:24:47,450 --> 00:24:49,091

It's something that looks like this.

421

00:24:51,351 --> 00:24:53,792

So, as long as you can get...

422

00:24:54,680 --> 00:24:57,620

Again, you need this API key

423

00:24:58,420 --> 00:24:59,629

to set it up.

424

00:24:59,970 --> 00:25:01,879

But then, what happens is that,

425

00:25:01,979 --> 00:25:04,500

if you get the ID of the park,

426

00:25:05,662 --> 00:25:11,354

then you could do something that I was trying.

427

00:25:11,943 --> 00:25:14,724

I got a list of the IDs of some parks

428

00:25:14,824 --> 00:25:20,334

and I tried to compare them to see whether they became more popular

429

00:25:20,434 --> 00:25:22,400

during lockdown or not.

430

00:25:22,901 --> 00:25:25,692

And I find that this is actually quite interesting

431

00:25:26,083 --> 00:25:28,444

because you can see that, on the left,

432



00:25:28,600 --> 00:25:30,270

there is this one with...

433

00:25:30,340 --> 00:25:32,551

The blue bars actually represent,

434

00:25:34,101 --> 00:25:36,320

during normal times,

435

00:25:37,352 --> 00:25:39,772

how popular the park is.

436

00:25:39,872 --> 00:25:44,243

And then, the two lines represent the lockdown.

437

00:25:44,503 --> 00:25:49,562

So, you see that some parks are actually less popular,

438

00:25:49,662 --> 00:25:51,892

some of them are more popular

439

00:25:52,438 --> 00:25:53,746

during the lockdown.

440

00:25:53,846 --> 00:25:56,001

And then, when I look at...

441

00:25:56,101 --> 00:25:58,120

I don't know Glasgow that well.

442

00:25:58,220 --> 00:26:00,048

Actually, I wasn't here for that long.

443

00:26:00,148 --> 00:26:02,146

But then, the interesting thing is that...

444

00:26:14,822 --> 00:26:17,391

Where was my file?

445

00:26:18,400 --> 00:26:19,538

But then... Never mind.

446

00:26:20,230 --> 00:26:21,849

I was just talking about that.

447

00:26:22,896 --> 00:26:25,529

I could do that without showing that, actually.

448

00:26:26,111 --> 00:26:29,038

So, the parks, about the parks,

449

00:26:29,406 --> 00:26:31,265

which have this pattern

450

00:26:31,365 --> 00:26:34,063

where they become less popular during the lockdown.

451

00:26:34,163 --> 00:26:37,202

They are actually the city centre's smaller parks.

452

00:26:37,629 --> 00:26:40,716

One of the ones that is becoming more popular is

453

00:26:40,816 --> 00:26:46,145

actually in the peripheral area of the city.

454

00:26:46,455 --> 00:26:48,123

So, I found that to be quite interesting.

455

00:26:48,223 --> 00:26:51,931

Although, it seems to be a pretty trivial observation

456

00:26:52,109 --> 00:26:55,037

but then, yeah, you can see all sorts of things when you try to

457

00:26:55,675 --> 00:26:57,933

carry out this type of analysis.

458

00:26:58,033 --> 00:26:59,341

So, that's the first one.

459

00:27:00,267 --> 00:27:05,446

And then, this one we will also get into more detail with a bit later,

460

00:27:05,546 --> 00:27:09,215

although not exactly using the API of

461

00:27:09,315 --> 00:27:12,222

the SCOOT data in Glasgow,

462

00:27:13,341 --> 00:27:15,370

in the second half of the session

463

00:27:15,470 --> 00:27:16,838

where I try to...

464

00:27:17,047 --> 00:27:19,374

If you read the map carefully,

465

00:27:19,573 --> 00:27:23,273

you can see the green icons.

466

00:27:24,872 --> 00:27:25,950

This one, it says,

467

00:27:26,050 --> 00:27:28,578

"Check your mirrors when moving between lanes".

468

00:27:28,762 --> 00:27:31,421

So, this is actually a map

469

00:27:32,468 --> 00:27:34,107

showing all the...

470

00:27:34,786 --> 00:27:37,156

If you read it on the legend.

471

00:27:37,566 --> 00:27:41,525

This one shows the Variable Message System.

472

00:27:41,655 --> 00:27:42,835

The messages.

473

00:27:43,063 --> 00:27:45,572

Here, you also have, "Look once, look twice",

474

00:27:45,819 --> 00:27:47,618

"Think bike", and so on.

475

00:27:47,940 --> 00:27:49,758

And then, in the middle,

476

00:27:50,128 --> 00:27:54,916

the orange ones are actually about parking availability.

477

00:27:55,164 --> 00:28:00,671

And then, the blue ones are the count locations of

478

00:28:02,289 --> 00:28:03,346

traffic.

479

00:28:03,941 --> 00:28:05,229

So, these types of things,

480

00:28:05,776 --> 00:28:08,774

you could see that if we get hold of

481

00:28:08,874 --> 00:28:09,913

the real-time data,

482

00:28:10,013 --> 00:28:13,781

we can easily visualise them and download this data

483

00:28:13,881 --> 00:28:15,659

for research purposes.

484

00:28:16,869 --> 00:28:20,078

And the third one is actually about

485

00:28:20,973 --> 00:28:24,971

the paper I'm going to discuss

486

00:28:25,071 --> 00:28:27,978

and I'm trying to introduce it very briefly

487

00:28:28,605 --> 00:28:33,733

so that we could go through some of the steps later

488

00:28:33,833 --> 00:28:36,276

in the second part of the session.

489

00:28:37,317 --> 00:28:40,265

So, this paper is about examining

490

00:28:40,634 --> 00:28:43,851

the effects of a temporary subway closure

491

00:28:44,506 --> 00:28:47,428

on cycling in Glasgow.

492

00:28:48,548 --> 00:28:53,046

So, we used this bike sharing data from

493

00:28:53,231 --> 00:28:56,545

Nextbike which operates in Glasgow,

494



00:28:57,515 --> 00:29:03,342

and then what we want to evaluate is,

495

00:29:04,388 --> 00:29:09,857

there was a temporary closure of

496

00:29:10,594 --> 00:29:13,842

over one month in 2016

497

00:29:14,961 --> 00:29:16,779

for the whole subway system.

498

00:29:16,953 --> 00:29:19,191

So, that's quite amazing to me

499

00:29:19,291 --> 00:29:22,128

because I've rarely heard of anything like that.

500

00:29:22,845 --> 00:29:24,362

You close a whole system

501

00:29:26,603 --> 00:29:28,703

for a whole month

502

00:29:29,476 --> 00:29:33,314

for renovation and then upgrading and so on.

503

00:29:33,799 --> 00:29:37,067

And then, since the system is quite small

504

00:29:37,167 --> 00:29:40,525

and then the distances between the stations,

505

00:29:40,625 --> 00:29:45,023

the subway stations, are actually pretty short.

506

00:29:45,778 --> 00:29:48,326

So, I was thinking that, maybe,

507

00:29:49,454 --> 00:29:53,442

people will replace these subway trips

508

00:29:53,799 --> 00:29:55,916

with cycling trips.

509

00:29:56,134 --> 00:29:57,702

So, that's the idea.

510

00:29:57,802 --> 00:30:00,949

And then, that's why we evaluated

511

00:30:01,049 --> 00:30:02,846

using the bike sharing data.

512

00:30:04,677 --> 00:30:07,564

So, again, this is the same objective

513

00:30:07,664 --> 00:30:11,170

as I was saying about some of my previous work.

514

00:30:11,447 --> 00:30:14,173

To study how different modes interact

515

00:30:14,603 --> 00:30:19,511

when there is this disruption of the public transport mode,

516

00:30:19,869 --> 00:30:21,326

subway in this case.

517

00:30:21,426 --> 00:30:25,664

How will people react in terms of their cycling behaviour?

518

00:30:25,963 --> 00:30:27,202

That's the idea.

519

00:30:28,573 --> 00:30:31,040

So, very briefly then,

520

00:30:31,547 --> 00:30:34,524

we narrowed down the study period to here

521

00:30:34,624 --> 00:30:37,450

because you have this red segment

522

00:30:37,550 --> 00:30:42,737

which is the actual suspension from July to some time in August.

523

00:30:43,703 --> 00:30:50,044

And then, these phases are just the introduction of the bike stations,

524

00:30:50,831 --> 00:30:52,323

the number of bike stations

525

00:30:52,423 --> 00:30:57,280

that Nextbike are having

526

00:30:57,608 --> 00:30:58,719

during that time.

527

00:31:00,616 --> 00:31:02,754

I excluded some of it because

528

00:31:03,194 --> 00:31:06,563

the growing number of bike stations might have

529

00:31:07,032 --> 00:31:09,002

affected the results.

530

00:31:09,600 --> 00:31:11,650

So, that's the idea.

531

00:31:11,916 --> 00:31:15,094

We study a part within the month.

532

00:31:15,200 --> 00:31:17,009

And then, as I've said,

533

00:31:17,109 --> 00:31:19,818

the system is actually pretty small.

534

00:31:19,918 --> 00:31:22,026

So, this is just to show that.

535

00:31:23,415 --> 00:31:26,193

These stations are just...

536

00:31:26,931 --> 00:31:29,410

If you're not familiar with the system,

537

00:31:29,510 --> 00:31:32,959

I found it very amusing when I first came to Glasgow

538

00:31:33,059 --> 00:31:34,866

because it's just a circle

539

00:31:35,344 --> 00:31:39,421

and you go clockwise or you go anticlockwise.

540

00:31:40,374 --> 00:31:42,465

That's it.

You can never get lost.

541

00:31:42,592 --> 00:31:44,159

And it's very, very short.

542

00:31:44,490 --> 00:31:46,886

So, that is why I think that it is possible that

543

00:31:47,972 --> 00:31:51,278

it can be substituted with bike rides.

544

00:31:52,608 --> 00:31:54,566

So, what is this about?

545

00:31:54,666 --> 00:31:55,964

It's not super clear.

546

00:31:56,064 --> 00:31:59,763

But then, you can see the orange ones are

547

00:31:59,863 --> 00:32:01,007

the subway stations

548

00:32:01,107 --> 00:32:03,616

and then these little dots are

549

00:32:03,716 --> 00:32:05,354

the bike stations.

550

00:32:06,003 --> 00:32:08,491

So, the idea is that, of course,

551

00:32:08,591 --> 00:32:11,698

we try to compare the time

552

00:32:11,873 --> 00:32:15,374

during the suspension

553

00:32:15,474 --> 00:32:17,212

and after the suspension

554

00:32:18,240 --> 00:32:21,338

to before the suspension

555

00:32:21,438 --> 00:32:23,806

so you can see how behaviour changed.



556

00:32:24,100 --> 00:32:27,488

But then, that might be problematic

557

00:32:28,200 --> 00:32:31,518

because we do not make use of

558

00:32:34,057 --> 00:32:35,976

most of our information.

559

00:32:36,244 --> 00:32:38,363

In terms of bike rides,

560

00:32:38,611 --> 00:32:42,628

we know the starting stations, we know the ending stations,

561

00:32:42,728 --> 00:32:45,292

so we have more information than that.

562

00:32:45,392 --> 00:32:47,969

So, what we have done is that

563

00:32:48,069 --> 00:32:50,397

we draw this buffer area

564

00:32:50,725 --> 00:32:52,905

where it's like a catchment area,

565

00:32:53,283 --> 00:32:55,120

where we think that, okay,

566

00:32:55,220 --> 00:33:00,974

if these bike stations are inside this treatment area,

567

00:33:01,074 --> 00:33:02,442

this catchment area,

568

00:33:02,620 --> 00:33:07,258

then we will call these bike stations

569

00:33:07,358 --> 00:33:09,056

the treatment stations.

570

00:33:09,284 --> 00:33:11,792

The other ones are the control stations.

571

00:33:11,892 --> 00:33:15,736

Like, these ones that are very far away from the subway system,

572

00:33:15,836 --> 00:33:17,134

the subway stations.

573

00:33:17,522 --> 00:33:20,660

So, the idea is that

574

00:33:20,760 --> 00:33:25,046

these stations, the trips of them are affected by

575

00:33:25,146 --> 00:33:27,954

the subway suspension, whilst these are not.

576

00:33:28,107 --> 00:33:30,865

So, we get another comparison,

577

00:33:31,410 --> 00:33:33,630

instead of only over time.

578

00:33:33,730 --> 00:33:36,998

We compare the suspension and without the suspension.

579

00:33:37,098 --> 00:33:38,355

So, that's the idea.

580

00:33:39,887 --> 00:33:43,335

So, yeah, of course, it relates to

581

00:33:43,435 --> 00:33:46,032

some of the existing studies.

582

00:33:46,134 --> 00:33:49,581

But then, again, I will not go into detail about this.

583

00:33:49,909 --> 00:33:54,004

It's just that, if you are already familiar with the terms,

584

00:33:54,541 --> 00:33:57,040

then you will know this is only about

585

00:33:57,603 --> 00:34:00,337

the interactions between the different modes.

586

00:34:00,611 --> 00:34:03,637

So, to put it very simply,

587

00:34:03,737 --> 00:34:08,460

it's just that now we are suspending the service of the subway.

588

00:34:08,560 --> 00:34:11,973

That means, you can think about it

589

00:34:12,073 --> 00:34:14,436

as an increase in the full cost of

590

00:34:14,536 --> 00:34:15,969

taking a subway trip.

591

00:34:17,191 --> 00:34:19,966

And then, with this change,

592

00:34:20,635 --> 00:34:23,739

how would it affect the quantity of trips

593

00:34:23,923 --> 00:34:25,627

in bike sharing.

594

00:34:25,971 --> 00:34:26,975

So, that's the idea.

595

00:34:27,075 --> 00:34:30,839

And then, if we put together these types of measures,

596

00:34:30,939 --> 00:34:35,251

then we can compare these with different studies

597

00:34:35,351 --> 00:34:37,055

that are doing similar things.

598

00:34:37,254 --> 00:34:38,285

That's the idea.

599

00:34:39,400 --> 00:34:42,223

And then, sometimes, if you go into the literature,

600

00:34:42,628 --> 00:34:46,302

it's not a new thing that people keep wanting to

601

00:34:48,177 --> 00:34:50,640

find out about the substitutability of, or

602

00:34:50,740 --> 00:34:53,595

complementarity of trips.

603

00:34:53,854 --> 00:34:57,757

Like, they study Uber

604

00:34:57,951 --> 00:35:01,347

as a complement or as a substitute of

605

00:35:01,900 --> 00:35:03,012

other trips.

606

00:35:03,212 --> 00:35:07,266

And also, it is possible when you look at it

607

00:35:07,529 --> 00:35:10,092

about rail or bus and so on.

608

00:35:11,000 --> 00:35:13,913

Because if you see bike sharing

609

00:35:14,013 --> 00:35:17,265

as a substitute, that means it replaces the trip.

610

00:35:17,588 --> 00:35:19,531

But then, if it is as a complement,

611

00:35:19,631 --> 00:35:23,284

then it is like the second case here

612

00:35:23,566 --> 00:35:27,459

where it is a first-mile last-mile facilitator,

613

00:35:27,650 --> 00:35:32,687

which means you ride the bike to, let's say, the rail station

614

00:35:32,787 --> 00:35:34,760

and then you take a rail trip.

615

00:35:35,560 --> 00:35:37,183

So, that's the idea.

616

00:35:37,274 --> 00:35:38,907

And then, of course,

617

00:35:39,410 --> 00:35:42,253

this may look a bit intimidating



618

00:35:43,560 --> 00:35:45,494

if you are not used to this,

619

00:35:45,656 --> 00:35:47,779

but then, the idea is, as I've said,

620

00:35:47,879 --> 00:35:50,193

that you're making two comparisons.

621

00:35:50,293 --> 00:35:52,507

One is between the time periods

622

00:35:52,607 --> 00:35:56,240

when the suspension is there,

623

00:35:56,882 --> 00:36:01,488

and since we would want to see whether people are coming back to,

624

00:36:01,697 --> 00:36:04,029

again, the...

625

00:36:05,300 --> 00:36:09,793

Let's say, with the suspension,

they want to ride the bikes

626

00:36:10,555 --> 00:36:14,260

and afterwards, do they still want to keep on riding bikes?

627

00:36:14,360 --> 00:36:19,352

Or when the subway is back, they will switch right back to

628

00:36:19,452 --> 00:36:21,095

riding on the subway,

629

00:36:21,195 --> 00:36:23,321

So, that's the idea. So, that's why there is

630

00:36:23,484 --> 00:36:26,523

the post-suspension period also.

631

00:36:27,173 --> 00:36:29,755

And then, another comparison would be

632

00:36:29,855 --> 00:36:32,638

the treatment station and the control station,

633

00:36:33,010 --> 00:36:34,993

the bike stations that I was talking about.

634

00:36:35,194 --> 00:36:36,706

So, when you do the analysis,

635

00:36:36,806 --> 00:36:39,796

you create different variables and so on.

636

00:36:40,585 --> 00:36:41,887

I won't go into details.

637

00:36:43,969 --> 00:36:47,242

If you followed the study, then you would know that

638

00:36:47,792 --> 00:36:50,302

after some basic cleaning of the data,

639

00:36:50,402 --> 00:36:53,255

you would want to cut some of the trips that are

640

00:36:53,605 --> 00:36:56,607

perhaps way too long to be covered

641

00:36:57,007 --> 00:37:00,229

because we are talking about replacing subway trips

642

00:37:00,329 --> 00:37:04,068

so it won't make sense if they are riding bikes for five hours.

643

00:37:04,168 --> 00:37:07,881

Then I can't really claim that it is a commuting trip.

644

00:37:07,981 --> 00:37:10,193

So, that's why some of the trips are removed

645

00:37:10,293 --> 00:37:13,585

and you can see that the percentage is not that much

646

00:37:13,685 --> 00:37:15,527

if you draw the cut-off

647

00:37:16,127 --> 00:37:20,670

around 150 minutes or something like that.

648

00:37:20,982 --> 00:37:25,324

That's pretty useful in looking at

649

00:37:25,424 --> 00:37:26,916

the whole picture

650

00:37:27,016 --> 00:37:30,877

and deciding where you want to make the cut-off.

651

00:37:31,818 --> 00:37:33,510

And then, also you have the...

652

00:37:34,300 --> 00:37:37,888

This is a diagram for the justification of

653

00:37:37,988 --> 00:37:41,360

using this type of model where I have the treatment

654

00:37:41,460 --> 00:37:43,052

and the control and so on.

655

00:37:43,351 --> 00:37:44,592

You can see that,

656

00:37:45,233 --> 00:37:50,505

visually, the treatment group and the control group do behave

657

00:37:50,605 --> 00:37:53,863

quite differently during the suspension period.

658

00:37:53,974 --> 00:37:59,699

So, I think that's quite strong visual evidence of that.

659

00:38:00,908 --> 00:38:04,482

So, the hypothesis, to say it simply is that

660

00:38:05,543 --> 00:38:10,790

since Glasgow subway station is in a dense city area, we would expect that

661

00:38:10,990 --> 00:38:14,736

these two modes, the bike sharing mode and the subway,

662

00:38:14,836 --> 00:38:16,654

would be substitutes.

663

00:38:16,802 --> 00:38:21,630

So, when you evaluate it and you estimate the model,

664

00:38:21,829 --> 00:38:24,485

the coefficient would be positive,

665

00:38:24,585 --> 00:38:27,333

the representing coefficient of interest.

666

00:38:27,813 --> 00:38:30,332

So, yeah, I wouldn't go back to the notation

667

00:38:30,432 --> 00:38:33,661

but then the findings show us that, actually,

668

00:38:33,839 --> 00:38:37,019

despite the system being pretty small,

669

00:38:37,119 --> 00:38:42,277

we really get more incoming and more outgoing bike trips

670

00:38:43,316 --> 00:38:45,175

when compared to the control group

671

00:38:45,275 --> 00:38:49,543

and also when compared to the other times that

672

00:38:50,299 --> 00:38:52,598

were without a suspension.

673

00:38:53,089 --> 00:38:55,989

So, that kind of

674

00:38:56,467 --> 00:38:59,435

strengthens our argument for

675

00:38:59,612 --> 00:39:04,163

our initial anticipation for the results.

676

00:39:05,079 --> 00:39:08,274

And then, although the number doesn't look too large,

677

00:39:08,374 --> 00:39:11,282

it looks like, okay, it's not a large number of trips,

678

00:39:11,382 --> 00:39:15,146

but then the system, both the subway system

679

00:39:15,240 --> 00:39:18,558

and the bike sharing system are not that big



680

00:39:19,045 --> 00:39:22,860

so it already represents around a 20% increase

681

00:39:22,960 --> 00:39:24,269

when you compare it.

682

00:39:24,478 --> 00:39:26,532

So, it's actually quite big.

683

00:39:27,471 --> 00:39:29,268

But then, of course, some of them,

684

00:39:29,497 --> 00:39:32,546

after the resumption of the subway service,

685

00:39:32,646 --> 00:39:36,192

they disappear again and they go back to

686

00:39:37,861 --> 00:39:38,969

riding the subway.

687

00:39:39,515 --> 00:39:42,216

So, after this,

688

00:39:42,316 --> 00:39:43,864

then we will see that,

689

00:39:44,213 --> 00:39:46,170

with these findings,

690

00:39:46,987 --> 00:39:48,154

if we converted...

691

00:39:48,328 --> 00:39:49,805

If you remember that earlier

692

00:39:49,905 --> 00:39:52,669

I brought up this term, diversion factor,

693

00:39:53,308 --> 00:39:58,846

it's just that when we have one fewer subway trip,

694

00:39:58,946 --> 00:40:01,724

like, people are shifting away from the subway,

695

00:40:01,824 --> 00:40:06,222

how many of them ended up riding a shared bike?

696

00:40:06,405 --> 00:40:08,633

It would be very, very small.

697

00:40:08,733 --> 00:40:13,890

Less than 0.05, which is the suggested value

698

00:40:13,990 --> 00:40:15,707

in some of the literature.

699

00:40:15,874 --> 00:40:19,095

But then, of course, this could be due to many reasons,

700

00:40:19,195 --> 00:40:23,401

that we found a much smaller value than this.

701

00:40:24,012 --> 00:40:28,903

So, I think this will end the first half

702

00:40:29,003 --> 00:40:32,491

and then the second half, we will get more hands on with

703

00:40:33,800 --> 00:40:38,258

trying to obtain some of the data

704

00:40:38,358 --> 00:40:40,455

and do something with it.

705

00:40:40,802 --> 00:40:42,336

In the second half.

706

00:40:42,436 --> 00:40:45,282

And then, I will also try to show

707

00:40:46,022 --> 00:40:48,088

how we start from there

708

00:40:48,188 --> 00:40:51,625

and then what could be done potentially

709

00:40:51,725 --> 00:40:54,551

with the data set to create something that is

710

00:40:54,730 --> 00:40:58,188

pretty similar to what I have just shown you.

711

00:40:58,989 --> 00:41:04,376

Although, I couldn't really use the Nextbike trip data because

712

00:41:04,476 --> 00:41:06,672

it's obtained from the company

713

00:41:06,772 --> 00:41:10,175

and it's not on their website right now.

714

00:41:11,212 --> 00:41:13,800

But then, we would be using

715

00:41:13,900 --> 00:41:16,635

something that is freely available

716

00:41:18,220 --> 00:41:20,596

in the session.

717

00:41:21,016 --> 00:41:22,683

So, this is it.

718

00:41:23,192 --> 00:41:26,857

And yeah, we will come back to the second half.

719

00:41:26,957 --> 00:41:29,975

But then, do I have to...

720

00:41:30,831 --> 00:41:32,590

Shall I answer something?

721

00:41:32,960 --> 00:41:34,736

Some questions.

722

00:41:35,826 --> 00:41:38,603

So, thank you, thank you for your talk.

723

00:41:38,975 --> 00:41:41,344

I think the very first question,

724

00:41:41,444 --> 00:41:44,786

which is not a question, rather a comment, was

725

00:41:44,886 --> 00:41:50,484

about Norway's system for public transport

726

00:41:50,584 --> 00:41:52,172

public information portal.

727

00:41:52,518 --> 00:41:55,375

- It's compulsory, it's cool.

- Which I think is very...

728

00:41:56,696 --> 00:41:57,773

Yeah, it's...

729

00:41:58,281 --> 00:42:01,768

So, in Norway,

there's compulsory sharing of

730

00:42:01,868 --> 00:42:06,156

the information of every stop and timetable for public transport.

731

00:42:06,256 --> 00:42:07,505

That's pretty cool.

732

00:42:07,605 --> 00:42:11,234

I think that, if I understand correctly,

733

00:42:13,309 --> 00:42:16,607

we are also going towards that direction.

734

00:42:16,707 --> 00:42:18,194

But we are not there yet.

735

00:42:18,585 --> 00:42:20,232

From what I understand.

736

00:42:21,444 --> 00:42:25,123

Yeah.

I would encourage everybody to look

737

00:42:25,223 --> 00:42:29,607

in the chat room, there is a useful link

738

00:42:29,707 --> 00:42:32,485

pointing to English development pages,

739

00:42:32,954 --> 00:42:34,108

if anybody is interested.

740

00:42:34,208 --> 00:42:37,003

So, public information portal for Norway,



741

00:42:37,770 --> 00:42:39,507

which is based on open data.

742

00:42:40,510 --> 00:42:42,631

Right. Let's move to the other questions.

743

00:42:43,916 --> 00:42:45,176

A question from Alan.

744

00:42:45,276 --> 00:42:47,923

How are you accounting for the other counterfactuals

745

00:42:48,023 --> 00:42:49,719

which have affected the treatment group?

746

00:42:53,128 --> 00:42:55,834

I think that we... Yeah, there are,

747

00:42:55,934 --> 00:42:58,033

of course, there are a lot of things

748

00:42:58,133 --> 00:43:00,852

that we couldn't really take into account.

749

00:43:00,952 --> 00:43:02,970

Then, the model we used.

750

00:43:03,886 --> 00:43:05,414

First thing, it was,

751

00:43:05,983 --> 00:43:10,072

we tried to take in the differences of stations

752

00:43:10,172 --> 00:43:14,051

in terms of using the fixed effects,

753

00:43:14,151 --> 00:43:15,310

that's the first thing,

754

00:43:15,988 --> 00:43:17,246

for different stations.

755

00:43:17,563 --> 00:43:19,835

And also,

756

00:43:19,935 --> 00:43:22,507

we included the weather data

757

00:43:22,607 --> 00:43:24,975

and then some other variables.

758

00:43:28,071 --> 00:43:31,500

And at the same time, of course it won't be perfect

759

00:43:31,600 --> 00:43:33,649

but we tried to make sure that,

760

00:43:33,749 --> 00:43:36,638

at the same time, there were no particular things that were

761

00:43:36,795 --> 00:43:40,584

particularly affecting the treatment group only.

762

00:43:42,553 --> 00:43:46,902

So, I think that's probably a simplification

763

00:43:47,002 --> 00:43:50,148

but that was the best that we could do

764

00:43:51,036 --> 00:43:55,675

from that simple model and the information that we had.

765

00:43:56,185 --> 00:43:58,303

So, it was a panel fixed effect

766

00:43:58,850 --> 00:44:01,078

and also, the other co-variants

767

00:44:01,178 --> 00:44:04,986

and trying to take out the period,

768

00:44:05,434 --> 00:44:09,293

as I have shown,

769

00:44:09,393 --> 00:44:11,661

the diagram with different phases.

770

00:44:11,759 --> 00:44:13,877

So, we tried to identify,

771

00:44:14,176 --> 00:44:16,243

we tried to cut out the time period

772

00:44:16,343 --> 00:44:19,260

where there were a lot of things happening with stations

773

00:44:19,839 --> 00:44:23,177

because there had been an increase in the number of stations.

774

00:44:23,367 --> 00:44:26,560

So, that was what we were trying to do.

775

00:44:27,299 --> 00:44:31,528

I hope that answers the question, partly at least.

776

00:44:32,459 --> 00:44:36,066

Okay.

Sharing the presentation as a PDF?

777

00:44:37,163 --> 00:44:38,332

Yes, I will.

778

00:44:38,670 --> 00:44:42,808

Actually, the scripts that we are going to use

779

00:44:42,908 --> 00:44:45,446

and the presentation will be shared.

780

00:44:46,031 --> 00:44:48,118

But I'm not sure if it will be on my personal GitHub

781

00:44:48,218 --> 00:44:50,964

or the Urban Big Data Centre's GitHub.

782

00:44:51,064 --> 00:44:53,102

I will try to find it out.

783

00:44:53,241 --> 00:44:56,189

And then, when we share this recording,

784

00:44:56,289 --> 00:44:58,266

it will also be available.

785

00:44:59,654 --> 00:45:01,643

Can I just quickly jump in with a comment?

786

00:45:01,990 --> 00:45:03,680

In a previous session, I think,

787

00:45:03,957 --> 00:45:09,358

every registered participant received a link to a recording

788

00:45:09,458 --> 00:45:11,132

and supporting materials at the very end.

789

00:45:11,232 --> 00:45:12,649

- So, I think it's...

- Oh, I see.

790

00:45:13,224 --> 00:45:14,773

- That is cool.

- This time as well.

791

00:45:15,161 --> 00:45:17,011

That is cool. I haven't uploaded it yet

792

00:45:17,111 --> 00:45:20,699

but I will tidy it up after today.

793

00:45:23,734 --> 00:45:25,086

So, just to answer the question.

794

00:45:25,296 --> 00:45:27,534

It will be shared in some form.

795

00:45:27,720 --> 00:45:30,737

- Yeah.

- To registered participants.

796

00:45:30,987 --> 00:45:32,016

Thank you.

797

00:45:35,120 --> 00:45:39,378

If you want to follow along for some of these apps,

798

00:45:39,745 --> 00:45:40,842

I'm trying to...

799

00:45:41,081 --> 00:45:44,467

Do you see my screen with RStudio?

800

00:45:44,729 --> 00:45:45,744

Yes.

801

00:45:46,023 --> 00:45:47,182

Brilliant.



802

00:45:47,322 --> 00:45:53,701

So, I think that the best way is that...

803

00:45:55,480 --> 00:45:57,475

Some of these scripts, of course,

804

00:45:58,593 --> 00:46:01,000

because you don't have the script with you,

805

00:46:01,100 --> 00:46:02,769

I haven't shared it yet.

806

00:46:02,957 --> 00:46:06,770

So, it would be unreasonable to ask you to

807

00:46:06,870 --> 00:46:11,760

follow it, unless you type in a super speedy way.

808

00:46:11,958 --> 00:46:16,586

But then, I think if you want to follow along,

809

00:46:16,686 --> 00:46:20,285

for some of the steps, I will share

810

00:46:20,385 --> 00:46:24,403

the particular command in the chat box

811

00:46:24,941 --> 00:46:29,979

so that we could try running it for some parts.

812

00:46:30,147 --> 00:46:34,872

But then, this first part, I would not expect that

813

00:46:34,972 --> 00:46:38,240

because it's a bit long and you can try it later.

814

00:46:38,628 --> 00:46:41,436

I guess. Because this part is

815

00:46:41,536 --> 00:46:46,335

actually using the API for Glasgow City Council

816

00:46:49,313 --> 00:46:52,982

on the information that they share,

817

00:46:53,082 --> 00:46:57,009

as you can see here very briefly.

818

00:46:57,397 --> 00:46:59,514

It would be about

819

00:47:03,637 --> 00:47:06,944

the movement, the traffic count,

820

00:47:07,478 --> 00:47:08,726

the carparks,

821

00:47:08,826 --> 00:47:11,814

the events that might be happening on the highways,

822

00:47:11,914 --> 00:47:13,271

the locations,

823

00:47:13,371 --> 00:47:18,628

and the Variable Message System that I was talking about.

824

00:47:18,914 --> 00:47:22,312

So, you will need an API key

825

00:47:22,412 --> 00:47:29,180

and then, we provided a link to access that,

826

00:47:29,392 --> 00:47:30,819

to sign up for that

827

00:47:32,156 --> 00:47:33,413

for this session.

828

00:47:33,513 --> 00:47:38,558

But then, I believe the website was dead for quite some time.

829

00:47:39,753 --> 00:47:42,131

I don't know if it was my problem or not.

830

00:47:42,231 --> 00:47:46,810

If any of you could actually get a key or could actually sign up.

831

00:47:47,047 --> 00:47:49,435

So, it doesn't really matter. I will still share the script

832

00:47:49,535 --> 00:47:53,830

and you could try it later, after you are able to get an account.

833

00:47:53,930 --> 00:47:56,918

But we could see what it looks like.

834

00:47:57,966 --> 00:48:00,974

So, the very first thing is

835

00:48:01,074 --> 00:48:05,511

that there would be some libraries that we might need

836

00:48:05,611 --> 00:48:06,859

most of the time.

837

00:48:07,549 --> 00:48:12,466

Some libraries like jsonlite and then something for tidying the tables.

838

00:48:12,566 --> 00:48:16,633

And then, one particular library I would be using would be

839

00:48:17,051 --> 00:48:22,556

Leaflet, just to show some locations information

840

00:48:22,753 --> 00:48:24,531

and then it looks pretty nice.

841

00:48:24,631 --> 00:48:28,719

The HTML tools are also related to that.

842

00:48:29,679 --> 00:48:32,116

And then, yeah, you could sign up later.

843

00:48:35,963 --> 00:48:40,323

You would be able to, if you got access to the site,

844

00:48:40,423 --> 00:48:45,161

you would be able to sign up for a standard account,

845

00:48:45,261 --> 00:48:47,408

or something like, start an account.

846

00:48:48,055 --> 00:48:53,915

And then, if you would like a lot, of course,

847

00:48:54,015 --> 00:48:55,112

over time,

848

00:48:55,212 --> 00:49:00,631

like, making a request every minute or whatever

849

00:49:00,814 --> 00:49:02,395

that you would like,

850

00:49:02,612 --> 00:49:04,510

then you would need a business account

851

00:49:04,610 --> 00:49:07,397

but then, yeah, you could do that when you sign up.

852

00:49:08,167 --> 00:49:12,235

And then, there are a few URLs.

853

00:49:12,335 --> 00:49:17,251

And the key here is actually just entering the key.

854

00:49:18,066 --> 00:49:20,843

I didn't show my key, I have already entered it.

855

00:49:22,129 --> 00:49:27,026

So, this URL is provided from their website because

856

00:49:27,284 --> 00:49:29,381

it's, from what I know,

857

00:49:31,029 --> 00:49:33,015

it might not be accessible right now.

858

00:49:33,115 --> 00:49:35,122

So, I would just not show it.

859

00:49:35,263 --> 00:49:37,670

But then, the flow is a bit like this.

860

00:49:38,039 --> 00:49:40,399

You get the key, you get the URL,

861

00:49:40,618 --> 00:49:45,215

and then you request that using this command

862

00:49:45,315 --> 00:49:48,306

by placing the key if it is necessary.

863

00:49:49,362 --> 00:49:51,521

And then, the next part would be



864

00:49:51,621 --> 00:49:55,265

more problematic because you will want to see,

865

00:49:55,597 --> 00:49:57,821

after what you have requested,

866

00:49:58,948 --> 00:50:02,647

how it looks and whether you can just use it directly

867

00:50:02,833 --> 00:50:07,098

for your, like, will it look like a very nice data set

868

00:50:07,198 --> 00:50:08,906

that you can use right away.

869

00:50:09,178 --> 00:50:13,207

So, yeah, it works, which is a good sign.

870

00:50:14,497 --> 00:50:15,824

So, from this request,

871

00:50:15,924 --> 00:50:19,074

we can see what we actually have.

872

00:50:19,663 --> 00:50:22,782

And then, if you look at it, then you can see...

873

00:50:22,882 --> 00:50:25,992

Okay, this doesn't look like what I was expecting.

874

00:50:26,243 --> 00:50:27,882

I was looking for traffic counts.

875

00:50:27,982 --> 00:50:30,861

And then, if you open it,

876

00:50:31,318 --> 00:50:35,738

there is some information about the whole schema.

877

00:50:36,174 --> 00:50:40,674

And then, you continue with it.

878

00:50:42,563 --> 00:50:45,435

Then you will start seeing the information

879

00:50:45,535 --> 00:50:46,713

that you would want.

880

00:50:46,893 --> 00:50:51,073

Because here, let's say this one is number one,

881

00:50:51,173 --> 00:50:56,842

and then you will start seeing the level of the traffic.

882

00:51:00,162 --> 00:51:03,001

Yeah, you see the traffic flow of

883

00:51:03,101 --> 00:51:05,500

that particular location.

884

00:51:05,770 --> 00:51:09,739

So, you can see that the site information is here.

885

00:51:09,839 --> 00:51:12,397

Each of them is a different site.

886

00:51:13,106 --> 00:51:15,511

So, this is not tidy at all

887

00:51:15,611 --> 00:51:19,529

and most of the time, when it comes to counts and locations,

888

00:51:19,966 --> 00:51:22,155

the counter information, the data,

889

00:51:22,469 --> 00:51:24,019

that you try to grab,

890

00:51:24,220 --> 00:51:27,339

there is no simple,

891

00:51:27,467 --> 00:51:29,307

as far as I know, most of the time,

892

00:51:29,407 --> 00:51:33,336

there are no simple commands that you can just use

893

00:51:33,845 --> 00:51:37,494

to tidy up this type of data,

894

00:51:37,594 --> 00:51:40,873

data frame that you get right away.

895

00:51:41,073 --> 00:51:45,462

So, of course, I would think that

896

00:51:46,891 --> 00:51:50,121

it would take too much time for us to do this

897

00:51:50,623 --> 00:51:52,412

like this here.

898

00:51:52,612 --> 00:51:55,170

But then, I guess the point would be

899

00:51:55,270 --> 00:51:56,838

that if you click it

900

00:51:56,938 --> 00:52:01,097

and you can see that there are different levels and layers.

901

00:52:04,116 --> 00:52:07,075

And then... Yeah, I will have to

902

00:52:07,275 --> 00:52:09,763

move this a little bit.

903

00:52:30,043 --> 00:52:32,071

So, that was the request.

904

00:52:32,171 --> 00:52:34,789

And then, we tidy it up.

905

00:52:34,974 --> 00:52:37,854

I did some of the following steps to tidy it up.

906

00:52:37,963 --> 00:52:42,032

And then, it looks a bit weird

907

00:52:42,132 --> 00:52:43,959

but then, the point is that

908

00:52:44,059 --> 00:52:47,088

I tried to get the movement information

909

00:52:47,188 --> 00:52:48,818

from those different layers

910

00:52:49,139 --> 00:52:55,757

and put it back into a data frame how I would like it to look.

911

00:52:55,964 --> 00:52:59,493

So, these are a bit annoying

912

00:52:59,593 --> 00:53:00,817

when you try to do it.

913

00:53:00,917 --> 00:53:04,455

But then, we will try another one later.

914

00:53:04,683 --> 00:53:08,144

But then, this is how it will look

915

00:53:08,244 --> 00:53:09,582

at the very beginning.

916

00:53:11,739 --> 00:53:15,359

So, we could see that this is just requested.

917

00:53:15,459 --> 00:53:18,237

So, if you click that open,

918

00:53:18,337 --> 00:53:23,116

you will see that the time is just now, not far from now.

919

00:53:23,705 --> 00:53:29,394

And then, if I run these,

920

00:53:34,302 --> 00:53:38,770

then I will be able to get a data frame called movement count,

921

00:53:38,870 --> 00:53:41,599

which is the camera

922

00:53:41,699 --> 00:53:44,237

and the number of the count,

923

00:53:45,137 --> 00:53:49,802

so that I could make use of it more easily.

924

00:53:50,012 --> 00:53:51,919

And then, let's say if...

925

00:53:52,452 --> 00:53:56,311

I will just try it with the parking data,



926

00:53:56,411 --> 00:53:58,166

doing almost the same thing.

927

00:53:58,266 --> 00:54:01,233

Requesting through the URL and then tidying it up.

928

00:54:03,355 --> 00:54:04,836

Okay. I run it.

929

00:54:06,065 --> 00:54:09,712

Then what I'm getting here is that...

930

00:54:10,919 --> 00:54:12,487

Okay. We'll try to open it

931

00:54:12,587 --> 00:54:15,865

and then we will see that there are

932

00:54:15,965 --> 00:54:17,973

the timestamps,

933

00:54:18,642 --> 00:54:24,070

and then you can see the parking locations,

934

00:54:24,777 --> 00:54:27,265

and also, whether there are

935

00:54:27,365 --> 00:54:31,405

enough spaces and how much of it is occupied.

936

00:54:32,169 --> 00:54:36,291

So, this sort of information is

937

00:54:36,859 --> 00:54:39,745

what we could get

938

00:54:41,022 --> 00:54:45,270

from this particular URL.

939

00:54:49,080 --> 00:54:51,157

It is a bit annoying because,

940

00:54:51,257 --> 00:54:54,251

I'm sorry, the bar keeps blocking me.

941

00:54:55,719 --> 00:54:57,348

We can move it.

942

00:54:58,819 --> 00:55:00,216

Yeah. Finally.

943

00:55:01,502 --> 00:55:04,381

So, the parking co-ordinates,

944

00:55:04,669 --> 00:55:05,816

they look like this.

945

00:55:05,916 --> 00:55:09,234

And then, there's the traffic events.

946

00:55:09,334 --> 00:55:12,192

And then, I could do similar things to it.

947

00:55:12,982 --> 00:55:16,134

And also, the Variable Message boards.

948

00:55:16,910 --> 00:55:19,627

And you can see that after you've requested it,

949

00:55:19,727 --> 00:55:24,256

it's also a bit of a mess if you try.

950

00:55:24,813 --> 00:55:26,331

This is the message board.

951

00:55:26,471 --> 00:55:31,381

And then, this is the thing that we have requested

952

00:55:31,481 --> 00:55:33,940

and we try to press into it.

953

00:55:34,340 --> 00:55:39,408

And then, we will see something, like,

954

00:55:43,008 --> 00:55:45,488

the information, the status,

955

00:55:46,566 --> 00:55:49,602

and then the message that is actually there.

956

00:55:50,942 --> 00:55:53,180

So, what we have to do is

957

00:55:53,280 --> 00:55:58,760

that we have to extract the useful information from this.

958

00:55:59,218 --> 00:56:01,896

And then, that's what we get.

959

00:56:02,096 --> 00:56:04,575

It would be after the steps.

960

00:56:05,947 --> 00:56:08,515

Because these are actually just columns.

961

00:56:08,728 --> 00:56:11,332

You go into one column and then you go into another column.

962

00:56:11,432 --> 00:56:14,059

So, I didn't go into great detail about this.

963

00:56:14,159 --> 00:56:16,087

And then, I renamed these columns

964

00:56:16,187 --> 00:56:18,413

so that it will look nicer.

965

00:56:18,513 --> 00:56:20,710

And then, the data frame, if we tidy it up,

966

00:56:20,810 --> 00:56:24,153

it will look like, okay, we have these five locations

967

00:56:24,253 --> 00:56:26,090

and then these are the times.

968

00:56:26,190 --> 00:56:30,622

And then, the message is just something like this.

969

00:56:30,722 --> 00:56:32,823

"Check your mirrors when moving".

970

00:56:33,455 --> 00:56:34,456

And so on.

971

00:56:34,556 --> 00:56:37,337

Sometimes there is more interesting information

972

00:56:37,437 --> 00:56:39,438

but not right now.

973

00:56:40,972 --> 00:56:43,199

And, of course, these message boards

974

00:56:43,299 --> 00:56:44,419

also have their location.

975

00:56:44,519 --> 00:56:46,919

And similarly, we could request this.

976

00:56:47,400 --> 00:56:51,430

So, what I was doing with this is...

977

00:56:52,602 --> 00:56:56,731

The next thing would be to combine these message boards

978

00:56:56,831 --> 00:56:59,618

with the location, because the message board data set

979

00:56:59,718 --> 00:57:01,538

and the location data set are

980

00:57:01,780 --> 00:57:05,141

in two different data frames right now.

981

00:57:05,321 --> 00:57:08,813

So, if we try to combine that,

982

00:57:09,385 --> 00:57:12,165

I have to rearrange a little bit

983

00:57:12,705 --> 00:57:17,479

because some of them, they have extra characters

984

00:57:17,579 --> 00:57:22,590

if you click into that particular name of the data frame.

985

00:57:22,860 --> 00:57:26,301

So, I remove some of them and I try to combine them.

986

00:57:26,441 --> 00:57:30,062

And then, in the end, what I aim to do is

987

00:57:30,162 --> 00:57:33,303

nothing too complicated.



988

00:57:33,403 --> 00:57:35,595

It's just that I was trying to

989

00:57:36,106 --> 00:57:38,937

put together these data sets

990

00:57:39,419 --> 00:57:43,420

and then place them into a nice map.

991

00:57:43,520 --> 00:57:49,313

So, I will make use of the Leaflet package now

992

00:57:49,874 --> 00:57:53,483

and then I add different markers,

993

00:57:54,415 --> 00:57:59,933

such as I supplied the location and what kind of label I am going to use.

994

00:58:01,555 --> 00:58:04,984

And then I save it as an HTML file.

995

00:58:05,084 --> 00:58:07,051

So, this is for the parking map

996

00:58:07,151 --> 00:58:08,612

and so on.

997

00:58:08,932 --> 00:58:12,322

And I try to choose different icons for that.

998

00:58:13,082 --> 00:58:18,374

So, I think the most interesting one that I'm going to show is

999

00:58:18,474 --> 00:58:22,302

the one where I included the different markers

1000

00:58:22,402 --> 00:58:23,560

on one map.

1001

00:58:23,878 --> 00:58:27,799

Like the one for the parking locations

1002

00:58:27,899 --> 00:58:30,974

and the one for the count of traffic

1003

00:58:31,233 --> 00:58:34,003

and also, the Variable Message Systems.

1004

00:58:34,842 --> 00:58:38,812

And it will look something like this.

1005

00:58:41,331 --> 00:58:42,969

So, you have something like,

1006

00:58:43,337 --> 00:58:45,766

okay, the messages are showing here.

1007

00:58:46,322 --> 00:58:50,961

And then, if you go around, if you move around.

1008

00:58:51,790 --> 00:58:53,388

This is way too slow.

1009

00:58:53,933 --> 00:58:58,172

You will see the count on the map

1010

00:58:59,590 --> 00:59:03,827

along the way. Or when you hover over the parking ones.

1011

00:59:04,563 --> 00:59:08,138

I've got a flag. I couldn't find a parking logo.

1012

00:59:08,378 --> 00:59:09,826

And then, you will see that

1013

00:59:10,447 --> 00:59:14,448

I am showing the message whether there would be enough space.

1014

00:59:14,792 --> 00:59:19,729

So, I think that could be something that would be

1015

00:59:19,829 --> 00:59:23,917

of interest if you want to combine them and show them all.

1016

00:59:24,124 --> 00:59:25,532

You find that too chaotic?

1017

00:59:25,760 --> 00:59:27,638

Then you could remove them.

1018

00:59:28,193 --> 00:59:31,500

You could overlay each one of them

1019

00:59:31,600 --> 00:59:33,939

or you could just remove these.

1020

00:59:34,239 --> 00:59:39,063

And this is done using the Leaflet package.

1021

00:59:42,974 --> 00:59:44,541

Then the next thing would be,

1022

00:59:44,641 --> 00:59:46,926

I think that we could try something

1023

00:59:47,375 --> 00:59:51,322

with this one on Brussels' traffic count.

1024

00:59:51,856 --> 00:59:55,123

So, let me go to that one.

1025

00:59:56,373 --> 01:00:01,640

This API, because it's open without a key.

1026

01:00:01,740 --> 01:00:05,538

So, I think if you have R with you,

1027

01:00:05,638 --> 01:00:07,536

you could just try it.

1028

01:00:08,616 --> 01:00:10,395

Let me share this.

1029

01:00:11,863 --> 01:00:13,731

Where's the chat box?

1030

01:00:27,304 --> 01:00:29,152

Okay. There you go.

1031

01:00:33,425 --> 01:00:35,545

So, if you go to this.

1032

01:00:43,981 --> 01:00:45,458

This is the site.

1033

01:00:45,658 --> 01:00:48,756

Then you should be able to see that

1034

01:00:48,935 --> 01:00:51,518

this is actually more complicated

1035

01:00:52,757 --> 01:00:56,853

than the one that we had for the Glasgow City Council

1036

01:00:57,685 --> 01:01:00,442

because of what they supplied.

1037

01:01:00,879 --> 01:01:04,019

We mainly have the vehicle count for that

1038

01:01:04,119 --> 01:01:07,901

and this one, we actually have way more information.

1039

01:01:08,900 --> 01:01:10,902

You can read it in detail later,

1040

01:01:11,002 --> 01:01:12,474

but then the point is that

1041

01:01:14,102 --> 01:01:17,353

you can have this kind of live data

1042

01:01:17,453 --> 01:01:22,205

about the count and speed and even the occupancy of the road,

1043

01:01:22,652 --> 01:01:28,350

and the timestamps and how they get hold of these numbers.

1044

01:01:28,450 --> 01:01:33,431

So, if you look here, you get a link

1045

01:01:33,531 --> 01:01:36,121

with the livestream counts.

1046

01:01:36,302 --> 01:01:39,682

And also, this one, if you click on this one,

1047

01:01:39,782 --> 01:01:45,314

you would be able to download a CSV with the locations of the devices.

1048

01:01:45,512 --> 01:01:49,224

So, that would be useful if you want to put them on a map

1049

01:01:49,324 --> 01:01:52,905

and if you want to know the exact location of these devices.



1050

01:01:53,286 --> 01:01:56,296

So, let's click on this.

1051

01:01:56,547 --> 01:01:57,937

And then you see,

1052

01:01:59,126 --> 01:02:03,486

when you place a live request, it will mostly look like this

1053

01:02:03,607 --> 01:02:07,718

for a lot of the cases you will look into.

1054

01:02:08,016 --> 01:02:10,138

First of all, you get the request date.

1055

01:02:10,238 --> 01:02:12,208

This is updated.

1056

01:02:13,548 --> 01:02:17,568

They are in Brussels, so they are one hour faster.

1057

01:02:17,878 --> 01:02:20,958

And, of course, it's close to this time.

1058

01:02:21,058 --> 01:02:24,899

And then, you get the device

1059

01:02:25,659 --> 01:02:29,327

and you get it for one minute.

1060

01:02:29,854 --> 01:02:32,002

And you get the count.

1061

01:02:32,267 --> 01:02:34,708

And then the speed and the occupancy

1062

01:02:34,997 --> 01:02:36,125

and all that.

1063

01:02:36,225 --> 01:02:38,113

It's not only one minute you get.

1064

01:02:38,213 --> 01:02:41,891

I think you get five minutes, fifteen minutes, and so on.

1065

01:02:42,089 --> 01:02:44,479

But then, the main problem would be,

1066

01:02:44,579 --> 01:02:47,947

of course, to get this and tidy this up

1067

01:02:49,394 --> 01:02:51,681

in your script.

1068

01:02:53,430 --> 01:02:56,780

First things first, we just try some of this

1069

01:02:56,880 --> 01:03:01,198

and with the URL that we just had.

1070

01:03:02,131 --> 01:03:03,680

Which is this one.

1071

01:03:04,871 --> 01:03:09,278

You can either save this page and then import it to R

1072

01:03:09,378 --> 01:03:12,234

or you can copy this one.

1073

01:03:12,334 --> 01:03:15,480

I think I just shared that too.

1074

01:03:15,908 --> 01:03:17,898

Yeah, but I can do that again.

1075

01:03:21,083 --> 01:03:24,375

I think we can put this up, sorry.

1076

01:03:26,715 --> 01:03:30,482

Put this in R and try to import that.

1077

01:03:32,827 --> 01:03:35,858

Yeah, like, okay, we start from here.

1078

01:03:35,958 --> 01:03:38,892

URL. So, we enter this URL

1079

01:03:39,399 --> 01:03:41,848

and then we do this request.

1080

01:03:47,121 --> 01:03:49,040

If you have RStudio with you,

1081

01:03:49,140 --> 01:03:50,989

you could also try.

1082

01:03:51,089 --> 01:03:53,347

And then, if you run it,

1083

01:03:53,447 --> 01:03:59,996

then you will see that we have the whole request for Brussels.

1084

01:04:01,185 --> 01:04:03,673

This is the data that we are looking at

1085

01:04:03,773 --> 01:04:08,701

and you can see that's the device

1086

01:04:08,801 --> 01:04:11,981

and then the results involved with this device

1087

01:04:12,081 --> 01:04:17,510

and then you have this different type of data

1088

01:04:17,610 --> 01:04:18,629

that you get.

1089

01:04:19,236 --> 01:04:21,007

You have the count and so on.

1090

01:04:21,107 --> 01:04:23,237

There's none for this one.

1091

01:04:23,636 --> 01:04:26,566

Yeah,

Perhaps, again, another one.

1092

01:04:35,466 --> 01:04:38,434

Then you see this kind of information.

1093

01:04:38,534 --> 01:04:40,941

That's the type of thing you can extract.

1094

01:04:41,041 --> 01:04:43,339

Of course, then we will have to go through

1095

01:04:44,085 --> 01:04:49,596

the annoying steps of requesting them.

1096

01:04:49,995 --> 01:04:53,533

I didn't include them here because I will put them,

1097

01:04:55,013 --> 01:04:58,711

I would not want to go over them here

1098

01:04:58,811 --> 01:05:02,550

and I will include them later when I share the script.

1099

01:05:02,650 --> 01:05:04,398

But then, this is the idea.

1100

01:05:04,778 --> 01:05:06,626

And then, as I've said,

1101

01:05:06,726 --> 01:05:10,346

the location of the counter is actually a CSV file that

1102

01:05:10,446 --> 01:05:13,874

you could download from the page that I just shared.

1103

01:05:13,974 --> 01:05:16,524

Then, afterwards, I think the next step would be

1104

01:05:16,624 --> 01:05:18,876

to combine the counter location

1105

01:05:18,976 --> 01:05:20,299

and also the count.

1106

01:05:20,499 --> 01:05:23,180

And then you can do different things,

1107

01:05:23,658 --> 01:05:26,079

like putting it on a map or something like that.

1108

01:05:28,551 --> 01:05:31,103

Of course, that's the map.

1109

01:05:31,203 --> 01:05:35,109

But then, I didn't go into a lot of detail.

1110

01:05:44,059 --> 01:05:45,290

Where is it?

1111

01:05:49,245 --> 01:05:51,330

This is the Brussels map



1112

01:05:51,461 --> 01:05:55,302

where you can see these are the different counter devices

1113

01:05:55,402 --> 01:05:59,761

that I plotted from the previous CSV file.

1114

01:06:00,411 --> 01:06:03,271

So, you can show more information,

1115

01:06:03,371 --> 01:06:06,380

like the count, after you have combined the data sets.

1116

01:06:07,420 --> 01:06:11,057

Again, I will share this information,

1117

01:06:11,157 --> 01:06:13,206

this script, later.

1118

01:06:14,246 --> 01:06:16,855

But then, it would be interesting

1119

01:06:16,955 --> 01:06:22,253

if you really wanted to try to tackle the area of your interest

1120

01:06:22,353 --> 01:06:23,408

for traffic count.

1121

01:06:23,508 --> 01:06:26,687

You will have to try to do that on your own.

1122

01:06:26,906 --> 01:06:30,485

Because I found out that, sometimes, for different cities

1123

01:06:30,585 --> 01:06:31,905

or different countries

1124

01:06:32,137 --> 01:06:36,209

the way they share or structure the scripts would be

1125

01:06:36,309 --> 01:06:40,237

a bit different, and it's not entirely clear to me

1126

01:06:41,087 --> 01:06:44,136

if there's a more uniform way of doing it.

1127

01:06:44,365 --> 01:06:48,670

But from what I've found also on GitHub and from other researchers,

1128

01:06:48,770 --> 01:06:54,509

they do it in this kind of way, one by one.

1129

01:06:54,609 --> 01:06:57,228

So, it's a bit annoying

1130

01:06:57,328 --> 01:06:59,257

but then you get what you want.

1131

01:06:59,357 --> 01:07:03,086

After you set it up, you can use it,

1132

01:07:03,186 --> 01:07:05,154

probably, for a longer time

1133

01:07:06,083 --> 01:07:08,732

so the effort will be worth it.

1134

01:07:09,572 --> 01:07:12,721

Actually, there's another example for Hull.

1135

01:07:13,126 --> 01:07:18,556

And if you check, this one also doesn't...

1136

01:07:19,215 --> 01:07:21,084

It's super small, sorry.

1137

01:07:22,109 --> 01:07:25,700

This one also doesn't require an API key,

1138

01:07:28,082 --> 01:07:32,297

so you could also try to play with this one.

1139

01:07:42,077 --> 01:07:46,933

You can see that if you don't want to deal with

1140

01:07:47,033 --> 01:07:50,200

all the messy files that I just showed you,

1141

01:07:50,300 --> 01:07:52,268

you can also download the CSV.

1142

01:07:52,368 --> 01:07:57,937

But then, you can see that it's one location by location

1143

01:07:58,037 --> 01:08:01,385

so you have a ton of files if you want a lot of them.

1144

01:08:02,895 --> 01:08:04,621

There are quite a lot of them.

1145

01:08:07,377 --> 01:08:09,801

And then, if you scroll down,

1146

01:08:09,901 --> 01:08:14,061

then you can see that you go back to the original things.

1147

01:08:14,161 --> 01:08:18,329

scootdata.geojson file or .jsonfile.

1148

01:08:21,932 --> 01:08:24,971

Let's see if we can access this.

1149

01:08:28,551 --> 01:08:30,777

Then you get this URL.

1150

01:08:32,507 --> 01:08:37,115

And then you almost go back to something very similar that

1151

01:08:37,215 --> 01:08:39,534

you will have to read.

1152

01:08:46,543 --> 01:08:50,718

Okay. This is the one that I'm referring to.

1153

01:08:50,818 --> 01:08:52,194

It's a bit slow, sorry.

1154

01:08:55,656 --> 01:08:57,091

Why is it not moving?

1155

01:09:00,344 --> 01:09:02,584

Yeah, you go back to something like this.

1156

01:09:02,836 --> 01:09:05,095

So, again, I would say that

1157

01:09:05,305 --> 01:09:08,223

this is going to take some time.

1158

01:09:08,873 --> 01:09:12,293

And, yeah, I will share it later,

1159

01:09:12,393 --> 01:09:15,747

how to go fast through this because I don't think we have

1160

01:09:15,847 --> 01:09:17,815

enough time to do it right away.

1161

01:09:17,915 --> 01:09:20,203

But then, one thing that I would like to mention is that

1162

01:09:20,303 --> 01:09:22,932

these things are usually in real-time

1163

01:09:23,850 --> 01:09:29,289

and it's not much use if you just get the data at this point in time.

1164

01:09:29,389 --> 01:09:31,834

So, how do you want to use it?

1165

01:09:33,033 --> 01:09:37,049

I think that it would be quite interesting

1166

01:09:37,149 --> 01:09:38,957

if you want to try it.

1167

01:09:39,534 --> 01:09:44,800

You can connect through a PostgreSQL database

1168

01:09:44,900 --> 01:09:49,295

and then you save it every certain number of minutes

1169

01:09:49,395 --> 01:09:51,873

and then you build your database

1170

01:09:51,973 --> 01:09:53,962

will all this kind of information.

1171

01:09:54,062 --> 01:09:56,950

And then you can schedule your script

1172

01:09:57,050 --> 01:10:00,448

to be run in a certain time interval,

1173

01:10:00,734 --> 01:10:03,645

so that you can get this type of information.



1174

01:10:04,665 --> 01:10:07,503

So, yeah, this is what this part is doing.

1175

01:10:07,603 --> 01:10:12,894

And then, if we move on to some of the bike share data,

1176

01:10:12,994 --> 01:10:18,590

there are many lengths that some bike companies

1177

01:10:19,037 --> 01:10:21,917

that I have found over time.

1178

01:10:22,375 --> 01:10:25,952

And then, the thing about bike share is

1179

01:10:26,052 --> 01:10:31,039

that it's much nicer than what I've shown you about the traffic counters.

1180

01:10:31,300 --> 01:10:35,509

So, let's say for CitiBike.

1181

01:10:39,874 --> 01:10:43,572

Let's go to the site of CitiBike.

1182

01:10:49,542 --> 01:10:52,091

Why can't I click on the link?

1183

01:10:53,379 --> 01:10:54,748

It doesn't really matter.

1184

01:11:02,078 --> 01:11:04,316

CitiBike, here.

1185

01:11:17,267 --> 01:11:21,242

I think I have shown some of this in the previous presentation,

1186

01:11:21,342 --> 01:11:26,106

where you have all the trip data you can download as a data dump

1187

01:11:26,206 --> 01:11:31,533

and then you have this kind of real-time data in this form.

1188

01:11:32,275 --> 01:11:38,632

So, we try to obtain the real-time data here

1189

01:11:38,732 --> 01:11:40,489

on the script because...

1190

01:11:41,308 --> 01:11:44,046

About the trip data, I will briefly talk about it

1191

01:11:44,146 --> 01:11:47,874

a bit later using the London shared bike.

1192

01:11:48,700 --> 01:11:51,177

So, again, we do the same thing.

1193

01:11:51,277 --> 01:11:52,636

We obtain the link.

1194

01:11:52,736 --> 01:11:56,233

We don't need an API key so it's relatively simple.

1195

01:11:56,470 --> 01:11:57,596

We get this.

1196

01:11:58,316 --> 01:12:04,323

So, the response of the request would be called Citibike.

1197

01:12:04,680 --> 01:12:08,676

And let's see what Citibike looks like.

1198

01:12:10,448 --> 01:12:12,865

It's not too bad. You have the station IDs,

1199

01:12:12,965 --> 01:12:15,822

station names, and then the locations,

1200

01:12:15,922 --> 01:12:19,158

the co-ordinates, and the status and so on.

1201

01:12:19,238 --> 01:12:22,337

So, this is relatively simple.

1202

01:12:23,816 --> 01:12:28,188

And then, if you try to arrange it a little bit

1203

01:12:28,566 --> 01:12:30,793

from this Citibike data frame,

1204

01:12:30,893 --> 01:12:34,490

then I use it.

1205

01:12:35,847 --> 01:12:39,294

I called it the station list.

1206

01:12:41,283 --> 01:12:43,734

I arranged it a little bit.

1207

01:12:44,424 --> 01:12:46,118

Then you can see that

1208

01:12:47,215 --> 01:12:50,169

we could obtain this data frame

1209

01:12:50,269 --> 01:12:52,424

which is quite tidy,

1210

01:12:52,803 --> 01:12:54,547

if you try it like this.

1211

01:12:54,927 --> 01:13:00,230

I can also put this in the chat if you are trying.

1212

01:13:04,250 --> 01:13:07,514

You can just run this and then you can check the column links

1213

01:13:07,614 --> 01:13:10,018

and then you can see that this is quite neat.

1214

01:13:12,118 --> 01:13:13,753

Station, yeah.

1215

01:13:13,853 --> 01:13:14,967

Can check it.

1216

01:13:17,357 --> 01:13:21,628

You can see that you have the available docks

1217

01:13:21,728 --> 01:13:23,733

and also, latitude and longitude,

1218

01:13:25,562 --> 01:13:26,980

and the status.

1219

01:13:28,091 --> 01:13:31,215

And then also you can get the timestamps.

1220

01:13:32,201 --> 01:13:34,794

Oh my god, this is from 2016.

1221

01:13:34,894 --> 01:13:37,900

But it's supposed to be much more up-to-date.

1222

01:13:42,921 --> 01:13:45,166

Most of these different companies,

1223

01:13:45,266 --> 01:13:47,325

they work in very similar ways

1224

01:13:47,425 --> 01:13:48,833

and it's quite straightforward.

1225

01:13:48,933 --> 01:13:52,883

And then, because they have this combined way of

1226

01:13:52,983 --> 01:13:54,562

inputting bike data,

1227

01:13:54,771 --> 01:13:56,930

usually it's not as complicated,

1228

01:13:57,030 --> 01:13:59,648

and then you could try on your way.

1229

01:14:00,028 --> 01:14:03,878

And then, with Nextbike, the bike company that's

1230

01:14:03,978 --> 01:14:06,226

also operating shared biking in Glasgow,

1231

01:14:06,326 --> 01:14:10,034

we could also obtain that in a very similar way.

1232

01:14:11,011 --> 01:14:16,950

Like the station data and the status and the number of available bikes.

1233

01:14:17,670 --> 01:14:20,341

So, yeah, this could be done.

1234

01:14:23,218 --> 01:14:26,806

So, this brings us very briefly to

1235

01:14:26,906 --> 01:14:29,511

the last part.



1236

01:14:30,012 --> 01:14:35,579

Actually, this is a very simplified version of

1237

01:14:35,779 --> 01:14:37,827

what I... I intend to share

1238

01:14:37,927 --> 01:14:42,663

a very simplified version of the paper I was talking about earlier.

1239

01:14:44,404 --> 01:14:50,885

Which is something that I am trying to share.

1240

01:14:53,053 --> 01:14:55,318

Wait a second. It's a bit slow.

1241

01:15:05,500 --> 01:15:10,208

So, that was the analysis, the application that I was talking about

1242

01:15:10,308 --> 01:15:14,129

a bit earlier in the first half of the session.

1243

01:15:14,538 --> 01:15:16,795

So, the workflow is a bit like this.

1244

01:15:16,895 --> 01:15:21,344

Of course, I don't expect you to follow every step right now.

1245

01:15:21,444 --> 01:15:24,163

But then, of course, you import the libraries.

1246

01:15:24,887 --> 01:15:28,078

And then, because this is a data set

1247

01:15:28,178 --> 01:15:31,655

with the detailed trip data.

1248

01:15:32,680 --> 01:15:34,518

You can see it looks like this.

1249

01:15:35,057 --> 01:15:38,942

You have the start time, end time, duration,

1250

01:15:39,042 --> 01:15:41,070

and then also the station number.

1251

01:15:41,170 --> 01:15:42,949

And you can do all sorts of things,

1252

01:15:43,155 --> 01:15:47,853

including arranging these numbers

1253

01:15:47,953 --> 01:15:49,681

in terms of stations.

1254

01:15:51,933 --> 01:15:54,708

So, the idea is that

1255

01:15:55,019 --> 01:15:56,037

we could also...

1256

01:15:56,425 --> 01:15:58,855

Because when I was talking about a paper,

1257

01:15:58,955 --> 01:16:02,831

you remember that we had this suspension period and so on,

1258

01:16:03,151 --> 01:16:09,868

so that we could introduce some dummy variables

1259

01:16:13,766 --> 01:16:16,715

to mark the period during the suspension

1260

01:16:16,815 --> 01:16:20,051

and after the suspension, and so on.

1261

01:16:20,361 --> 01:16:23,018

And also introduce the treatment group

1262

01:16:23,118 --> 01:16:26,897

and the control group for the different stations.

1263

01:16:27,834 --> 01:16:31,846

So, the workflow is something like this.

1264

01:16:31,946 --> 01:16:34,700

Where you introduce the buffer

1265

01:16:34,800 --> 01:16:38,387

and then you introduce the dummy variables

1266

01:16:38,694 --> 01:16:41,343

for the treatment and control stations,

1267

01:16:41,443 --> 01:16:42,897

the suspension period,

1268

01:16:43,093 --> 01:16:47,791

and also clean up the data about longer trips

1269

01:16:48,048 --> 01:16:49,615

that you might want to draw up

1270

01:16:49,715 --> 01:16:53,652

or trips that you do not want to include in your analysis,

1271

01:16:53,752 --> 01:16:56,818

such as having the same origin and destination.

1272

01:16:58,612 --> 01:17:00,348

We also include weather data.

1273

01:17:00,545 --> 01:17:02,675

So, it also took some time

1274

01:17:02,775 --> 01:17:07,163

to just put in the weather data by day

1275

01:17:07,597 --> 01:17:12,114

and arranging them such that it could be combined with

1276

01:17:12,214 --> 01:17:16,483

the panel data set and also the public holidays and so on.

1277

01:17:17,621 --> 01:17:21,579

So, if we go back to the R Script.

1278

01:17:22,057 --> 01:17:29,452

Of course, because I couldn't share that particular data set.

1279

01:17:29,789 --> 01:17:33,647

So, that's why I think that we could do it like this.

1280

01:17:34,084 --> 01:17:36,843

You could try, if you want to try,

1281

01:17:36,843 --> 01:17:39,741

you can import a few libraries.

1282

01:17:39,865 --> 01:17:41,450

You might not need all of them.

1283

01:17:41,550 --> 01:17:44,499

This is just a start. I will share the whole thing later.

1284

01:17:45,584 --> 01:17:49,001

And because the London one,

1285

01:17:49,101 --> 01:17:54,227

if you try to get access to the London bike share data,

1286

01:17:56,015 --> 01:17:58,495

it's actually pretty comprehensive.

1287

01:17:59,463 --> 01:18:02,062

Let me try to get to it.

1288

01:18:02,781 --> 01:18:03,848

Yeah, here.

1289

01:18:26,709 --> 01:18:28,308

If you go there,

1290

01:18:31,867 --> 01:18:35,006

you should be able to see this bucket loading thing.

1291

01:18:36,000 --> 01:18:38,039

Because I was just creating the example,

1292

01:18:38,139 --> 01:18:42,869

I got a bit lazy and I saw this ZIP file

1293

01:18:43,438 --> 01:18:46,788

for 2015 so I just downloaded it

1294

01:18:47,366 --> 01:18:48,685

and then read it.

1295

01:18:49,773 --> 01:18:51,780

It's somewhere. But then you can see that

1296

01:18:52,178 --> 01:18:55,266

they have data for every single week.

1297

01:18:56,002 --> 01:18:57,616

And then, if you scroll down,



1298

01:18:57,716 --> 01:19:00,614

you can see the older data

1299

01:19:00,714 --> 01:19:02,940

in a ZIP file.

1300

01:19:03,101 --> 01:19:06,619

So, if you just want to try something to see if it works,

1301

01:19:06,719 --> 01:19:07,987

just download one of these.

1302

01:19:08,087 --> 01:19:10,135

And then, what I did is that

1303

01:19:10,235 --> 01:19:11,982

I set up a folder

1304

01:19:12,291 --> 01:19:14,439

to house this part.

1305

01:19:14,687 --> 01:19:16,444

And then, with this folder,

1306

01:19:16,713 --> 01:19:18,829

I created a file list

1307

01:19:18,888 --> 01:19:24,533

and then I read all these files.

1308

01:19:24,633 --> 01:19:28,347

Because if you actually look at the folder,

1309

01:19:28,615 --> 01:19:34,093

if for the whole 2015 you have one file

1310

01:19:34,193 --> 01:19:35,441

for each week,

1311

01:19:36,190 --> 01:19:39,407

you end up getting a pretty big data set.

1312

01:19:40,295 --> 01:19:42,843

So, yeah, let's do this.

1313

01:19:44,791 --> 01:19:46,389

It might take some time.

1314

01:19:53,123 --> 01:19:55,529

After you download it, you can try it.

1315

01:19:58,330 --> 01:20:02,928

I'll just put this here in case you want to follow right now.

1316

01:20:03,627 --> 01:20:06,519

If you have a fast enough connection.

1317

01:20:07,736 --> 01:20:09,885

Of course, this is just my folder path,

1318

01:20:09,985 --> 01:20:11,833

which you will have to change.

1319

01:20:13,854 --> 01:20:15,563

And then, if you follow this,

1320

01:20:15,663 --> 01:20:19,681

then it will be reading these files.

1321

01:20:20,110 --> 01:20:25,272

And it will be called London bike share.

1322

01:20:25,497 --> 01:20:26,848

If we click into it,

1323

01:20:27,185 --> 01:20:30,313

you can see that it's really pretty huge.

1324

01:20:31,518 --> 01:20:36,247

I think it's almost 10 million, right?

1325

01:20:36,347 --> 01:20:37,916

If I read correctly.

1326

01:20:39,253 --> 01:20:44,423

While the one for Nextbike is like 10% of this size, I think,

1327

01:20:44,523 --> 01:20:46,331

or even less.

1328

01:20:47,274 --> 01:20:51,365

So, it would be more effective

1329

01:20:51,465 --> 01:20:57,840

if you can read this using fread,

1330

01:20:58,074 --> 01:21:00,894

I suppose, if you're familiar with it.

1331

01:21:01,294 --> 01:21:03,862

And then, about the location of the station,

1332

01:21:03,962 --> 01:21:06,559

then it gets more interesting

1333

01:21:06,659 --> 01:21:10,685

because it actually took me some time to dig it out

1334

01:21:11,204 --> 01:21:16,271

and surprisingly, it's actually not from direct download.

1335

01:21:16,371 --> 01:21:21,029

It's more like from a freedom of information request,

1336

01:21:21,129 --> 01:21:25,437

that people were requesting the location of the stations.

1337

01:21:26,447 --> 01:21:29,355

Then I found the file of the locations of stations

1338

01:21:29,455 --> 01:21:30,542

around that time.

1339

01:21:30,642 --> 01:21:32,440

Then I read it.

1340

01:21:36,719 --> 01:21:37,875

Then I tried to,

1341

01:21:38,138 --> 01:21:42,178

because it's a bit messy.

1342

01:21:42,278 --> 01:21:43,426

If you can see.

1343

01:21:43,526 --> 01:21:45,782

I will just show it very briefly.

1344

01:21:45,882 --> 01:21:49,670

You can see that the station name, here you have the station name

1345

01:21:49,770 --> 01:21:54,058

and then you have some sort of local area

1346

01:21:54,717 --> 01:21:57,807

in your trip data set.

1347

01:21:58,045 --> 01:22:01,699

But then, if you look at the station data set,

1348

01:22:01,977 --> 01:22:03,066

I have it,

1349

01:22:04,586 --> 01:22:06,964

you can see that it's slightly different.

1350

01:22:07,172 --> 01:22:10,642

So, what I was doing here is

1351

01:22:10,742 --> 01:22:13,071

just renaming them

1352

01:22:13,171 --> 01:22:17,048

and trying to remove the part of

1353

01:22:17,148 --> 01:22:20,895

the station name so that they could match.

1354

01:22:20,995 --> 01:22:24,473

I tried to extract the part before the comma.

1355

01:22:26,763 --> 01:22:28,061

So, afterwards,

1356

01:22:28,852 --> 01:22:31,174

you get something that is quite neat

1357

01:22:31,274 --> 01:22:34,535

and, actually, you could try to arrange it,

1358

01:22:36,584 --> 01:22:39,487

you could try to combine them together.

1359

01:22:39,865 --> 01:22:43,869

And then, here, this is just something simple,



1360

01:22:43,969 --> 01:22:46,347

because if you look back to your station,

1361

01:22:46,792 --> 01:22:49,402

to your trip data set,

1362

01:22:49,502 --> 01:22:52,964

you can see these names are all separated by spaces

1363

01:22:53,064 --> 01:22:55,614

and sometimes spaces could cause problems.

1364

01:22:55,714 --> 01:22:57,662

So, I replaced them with dots.

1365

01:22:58,648 --> 01:23:04,950

I mean, it's usually up to what you want to do with them.

1366

01:23:05,050 --> 01:23:07,347

Or you can replace it with something else.

1367

01:23:08,307 --> 01:23:12,084

And then, the next one would be

1368

01:23:12,289 --> 01:23:14,532

just to merge them

1369

01:23:14,889 --> 01:23:17,936

so that we get a different data frame

1370

01:23:21,751 --> 01:23:23,152

that we could use.

1371

01:23:25,727 --> 01:23:29,905

Let's say, I will need the locations of these stations

1372

01:23:30,296 --> 01:23:33,467

to plot them on a map or something,

1373

01:23:34,017 --> 01:23:36,577

that would be pretty useful.

1374

01:23:36,677 --> 01:23:38,899

But the merging would take some time

1375

01:23:38,999 --> 01:23:40,780

so I would not run it.

1376

01:23:43,791 --> 01:23:46,631

And then, afterwards,

1377

01:23:47,242 --> 01:23:52,421

I would like to arrange the number of trips by day.

1378

01:23:52,997 --> 01:23:55,887

Actually, this is just to show that

1379

01:23:56,269 --> 01:23:59,057

sometimes we might want to do something like

1380

01:23:59,248 --> 01:24:01,517

sorting the number of trips by day

1381

01:24:01,617 --> 01:24:06,367

or we might want to plot the trip duration by day.

1382

01:24:06,597 --> 01:24:08,337

And then, if you think back,

1383

01:24:08,437 --> 01:24:12,357

this is actually, if you combine these data sets,

1384

01:24:12,457 --> 01:24:15,648

this is actually something that you find a lot

1385

01:24:16,217 --> 01:24:17,895

in places like Kaggle

1386

01:24:17,995 --> 01:24:21,283

where you can find the number of trips dataset

1387

01:24:21,383 --> 01:24:23,009

combined with the weather data set

1388

01:24:23,109 --> 01:24:26,408

where people are using these data sets to

1389

01:24:26,508 --> 01:24:28,165

run their algorithm from.

1390

01:24:28,636 --> 01:24:31,173

So, my point is that

1391

01:24:31,273 --> 01:24:35,052

if you get the more complicated data dump,

1392

01:24:35,271 --> 01:24:37,960

then you could easily reduce it back

1393

01:24:38,060 --> 01:24:40,700

to something that you could use very easily.

1394

01:24:41,160 --> 01:24:44,432

Because here, I didn't include everything.

1395

01:24:44,532 --> 01:24:49,181

But then, the next step that I would share would be

1396

01:24:49,281 --> 01:24:52,332

something like, after arranging them by day,

1397

01:24:52,432 --> 01:24:54,833

I will combine it.

1398

01:24:55,134 --> 01:24:57,500

Actually, I find this quite interesting because

1399

01:24:57,600 --> 01:25:02,312

I find daily restriction of

1400

01:25:02,523 --> 01:25:05,641

COVID lockdown restrictions in London.

1401

01:25:06,742 --> 01:25:08,252

So, potentially,

1402

01:25:08,352 --> 01:25:13,754

if I were using the 2020 and 2021 data sets,

1403

01:25:13,881 --> 01:25:17,121

then I could easily run some analysis

1404

01:25:17,632 --> 01:25:22,927

after I combine these for daily number of trips

1405

01:25:23,027 --> 01:25:25,648

and then I also get the station level data,

1406

01:25:25,748 --> 01:25:31,266

and also, every day, what the restrictions look like.

1407

01:25:31,507 --> 01:25:32,625

This is very interesting.

1408

01:25:32,725 --> 01:25:35,996

You can get it from the data for later.

1409

01:25:36,096 --> 01:25:41,726

But then, it's also on the Department for Transport site.

1410

01:25:42,865 --> 01:25:44,725

The reason that I highlight this is that

1411

01:25:44,825 --> 01:25:47,736

you can see you have the school closure dummy,

1412

01:25:47,926 --> 01:25:49,157

pub closure dummy,

1413

01:25:49,257 --> 01:25:51,928

shop closure dummy, and then by date.

1414

01:25:52,124 --> 01:25:55,049

And that's pretty useful information.

1415

01:25:55,149 --> 01:25:58,619

If you try to run it and relate it to

1416

01:25:58,719 --> 01:26:01,959

other data sets, all you need is the date to combine them.

1417

01:26:04,190 --> 01:26:07,322

I also mentioned that, in terms of the stations,

1418

01:26:07,422 --> 01:26:12,361

in terms of what you have between the treatment stations.

1419

01:26:12,822 --> 01:26:16,254

Let's say that you can identify that

1420

01:26:16,354 --> 01:26:18,444

some stations you have a hypothesis for

1421

01:26:18,725 --> 01:26:20,656

saying that some stations may be



1422

01:26:20,756 --> 01:26:23,585

more affected by these measures than the others.

1423

01:26:24,116 --> 01:26:27,466

Then, you can carry out the same type of analysis

1424

01:26:27,566 --> 01:26:32,126

that I was just talking about earlier on.

1425

01:26:32,616 --> 01:26:37,836

Similar to that research paper that I was talking about.

1426

01:26:38,731 --> 01:26:39,914

So, this is the idea.

1427

01:26:40,014 --> 01:26:42,100

Although, I didn't show every step

1428

01:26:45,260 --> 01:26:49,420

or the complete codes for now.

1429

01:26:49,721 --> 01:26:54,432

But I'm going to share it and I hope that you can see,

1430

01:26:55,214 --> 01:26:57,143

after you get some of this data,

1431

01:26:57,576 --> 01:27:02,336

how you can proceed to get more from them

1432

01:27:02,436 --> 01:27:05,425

by developing a data set

1433

01:27:05,525 --> 01:27:08,263

and a workflow that goes from there.

1434

01:27:08,451 --> 01:27:12,448

And there are numerous research possibilities for that

1435

01:27:12,548 --> 01:27:18,839

as a citizen or even perhaps as a student

1436

01:27:18,939 --> 01:27:22,829

to look through these kinds of things, or even as a researcher, you might

1437

01:27:23,470 --> 01:27:27,970

want to explore your research questions

1438

01:27:28,219 --> 01:27:30,480

with this type of data.

1439

01:27:31,170 --> 01:27:34,810

So, this is it for now.

1440

01:27:35,930 --> 01:27:41,710

I understand that I didn't really have much time to cover all the steps

1441

01:27:41,908 --> 01:27:45,640

so you might find it a bit puzzling

1442

01:27:45,740 --> 01:27:47,319

and a bit chaotic at times.

1443

01:27:47,419 --> 01:27:51,409

But I will try to share the materials

1444

01:27:51,509 --> 01:27:54,930

and the links here so you could try it.

1445

01:27:55,497 --> 01:27:58,657

And you could try it later.

1446

01:27:58,757 --> 01:28:01,226

So, this is it.

1447

01:28:01,977 --> 01:28:03,017

Thanks.