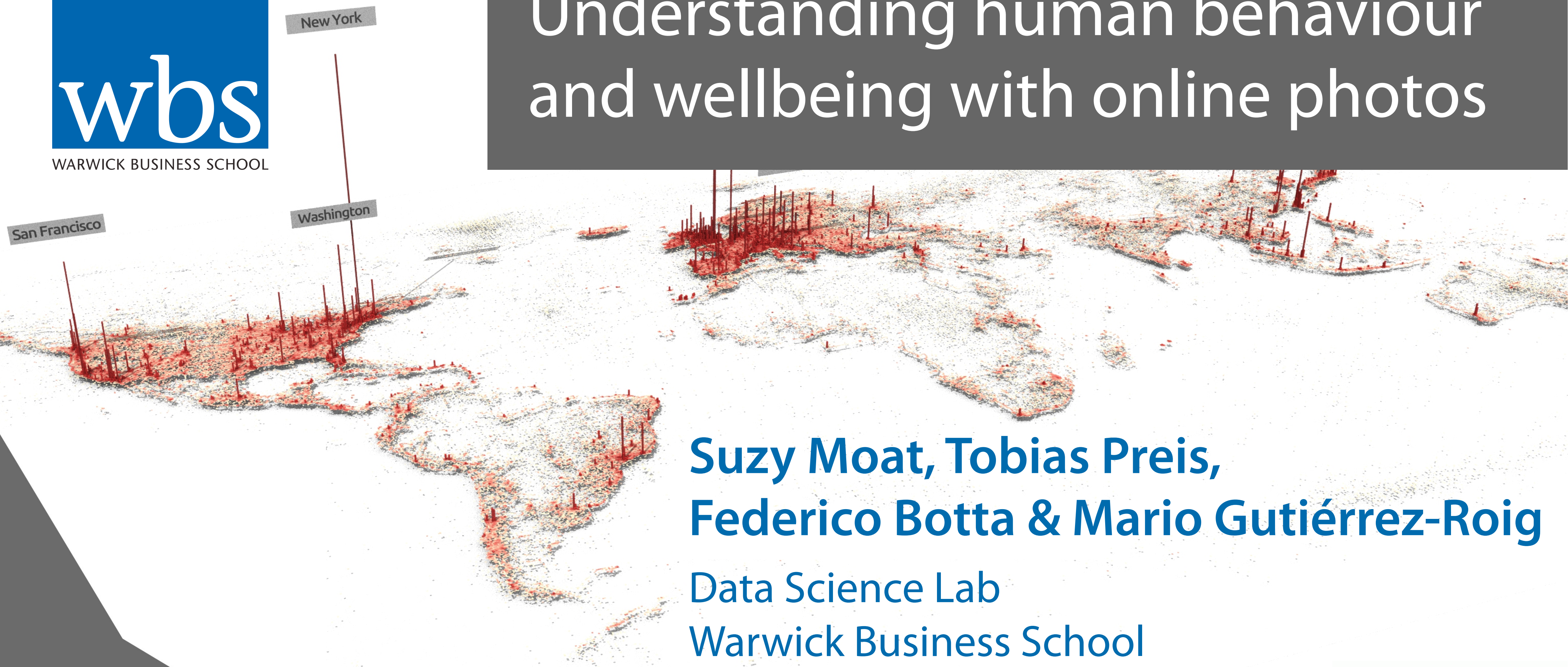
Understanding human behaviour and wellbeing with online photos



Suzy Moat, Tobias Preis, Federico Botta & Mario Gutiérrez-Roig

Data Science Lab
Warwick Business School

@thedatascilab

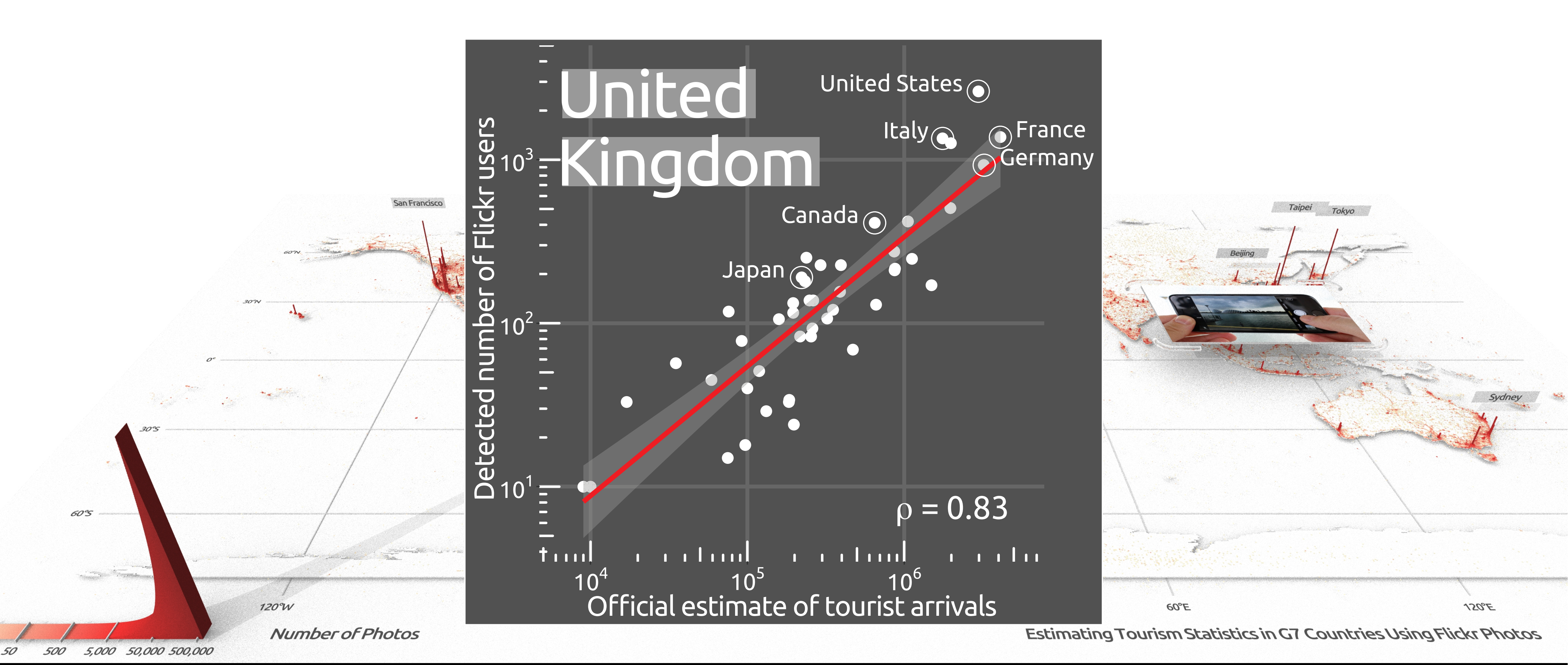
WARWICK BUSINESS SCHOOL

The Alan Turing Institute

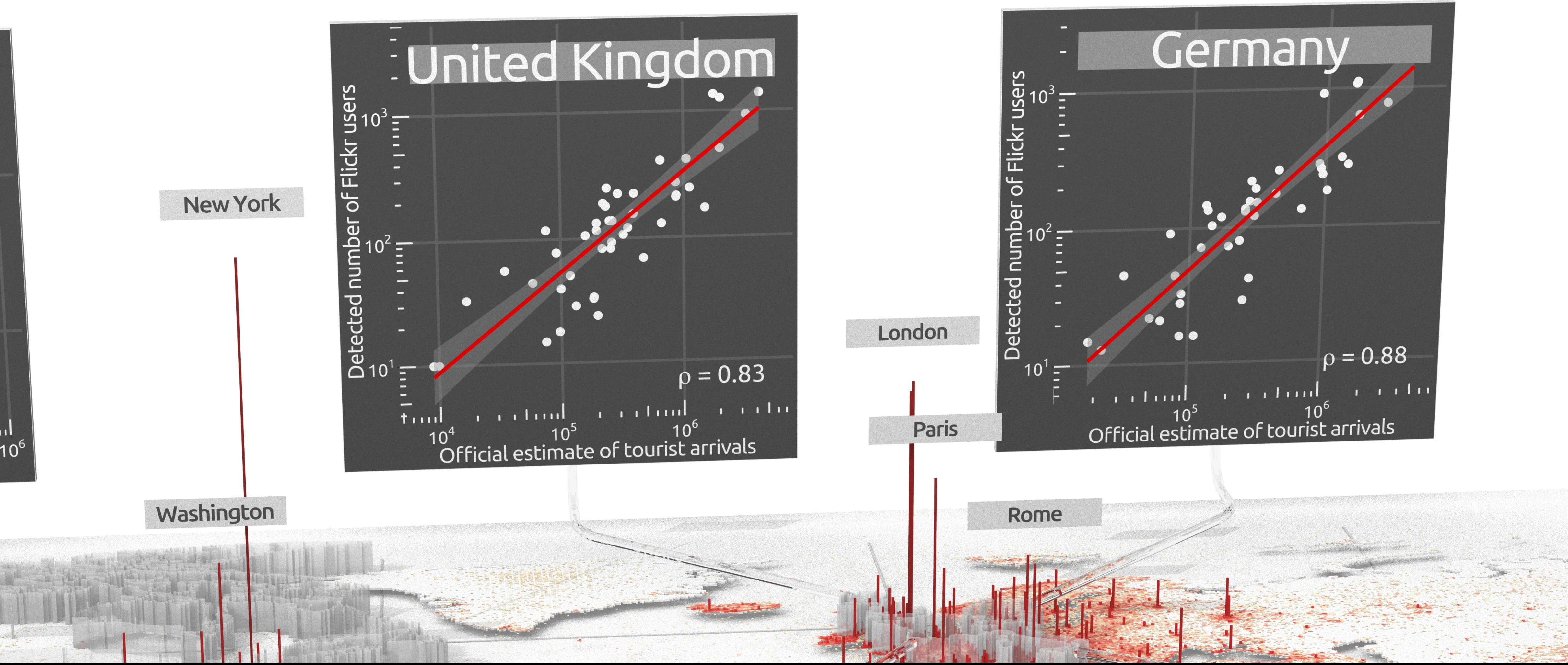
THE UNIVERSITY OF WARWICK


Beijing
Tokyo
Taipei
HDR
PANO
SQUARE
PHOTO
VIDEO
SLO-MO

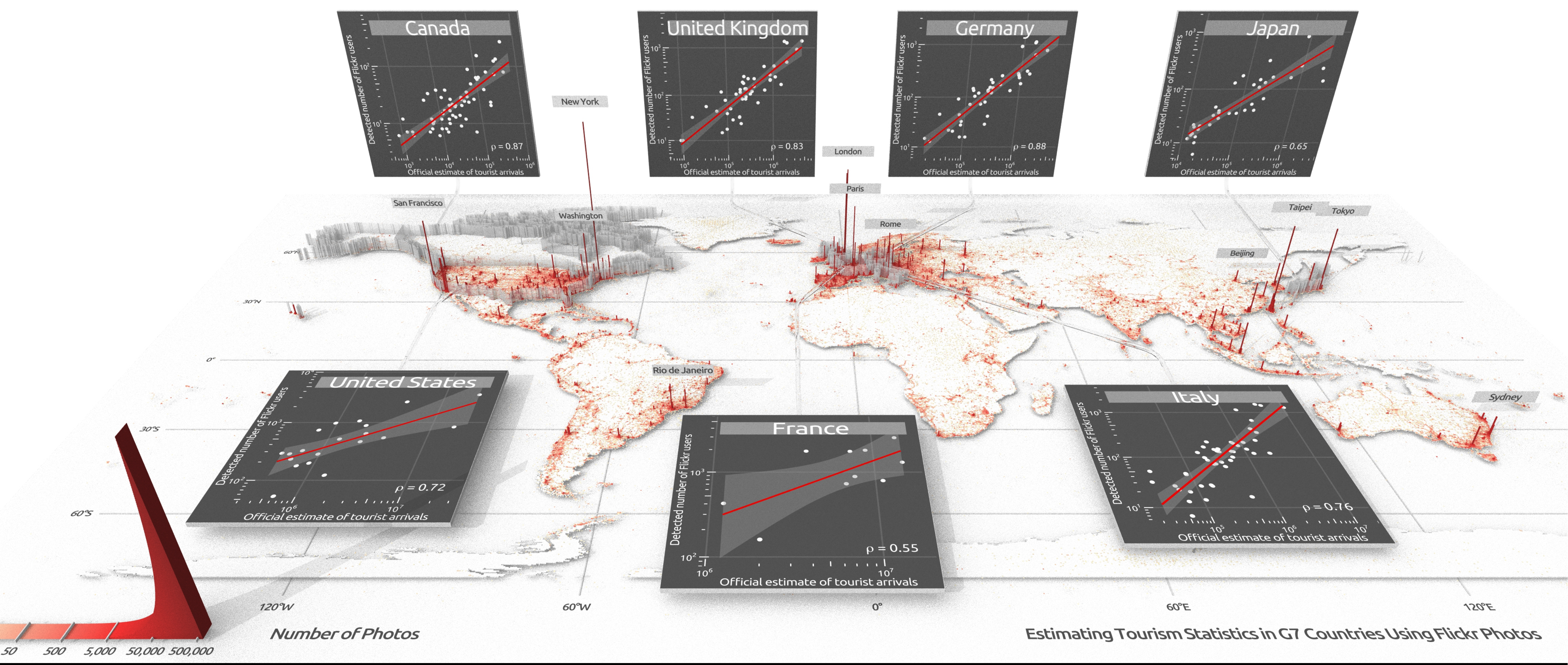

Beijing
Tokyo
Taipei
SQUARE
PHOTO
VIDEO
SLO-MO
HDR



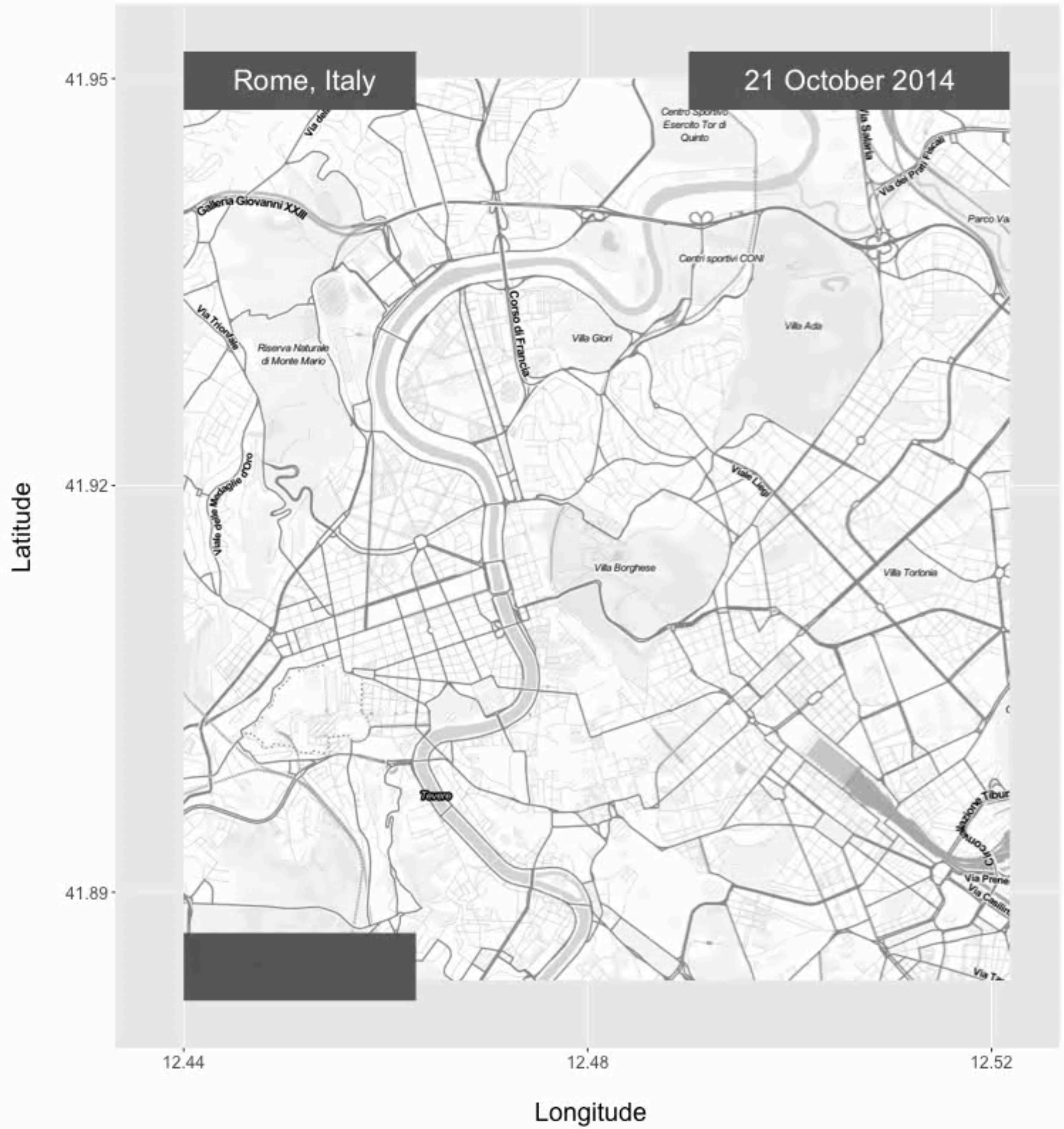
Barchiesi, Moat, Alis, Bishop & Preis (2015)

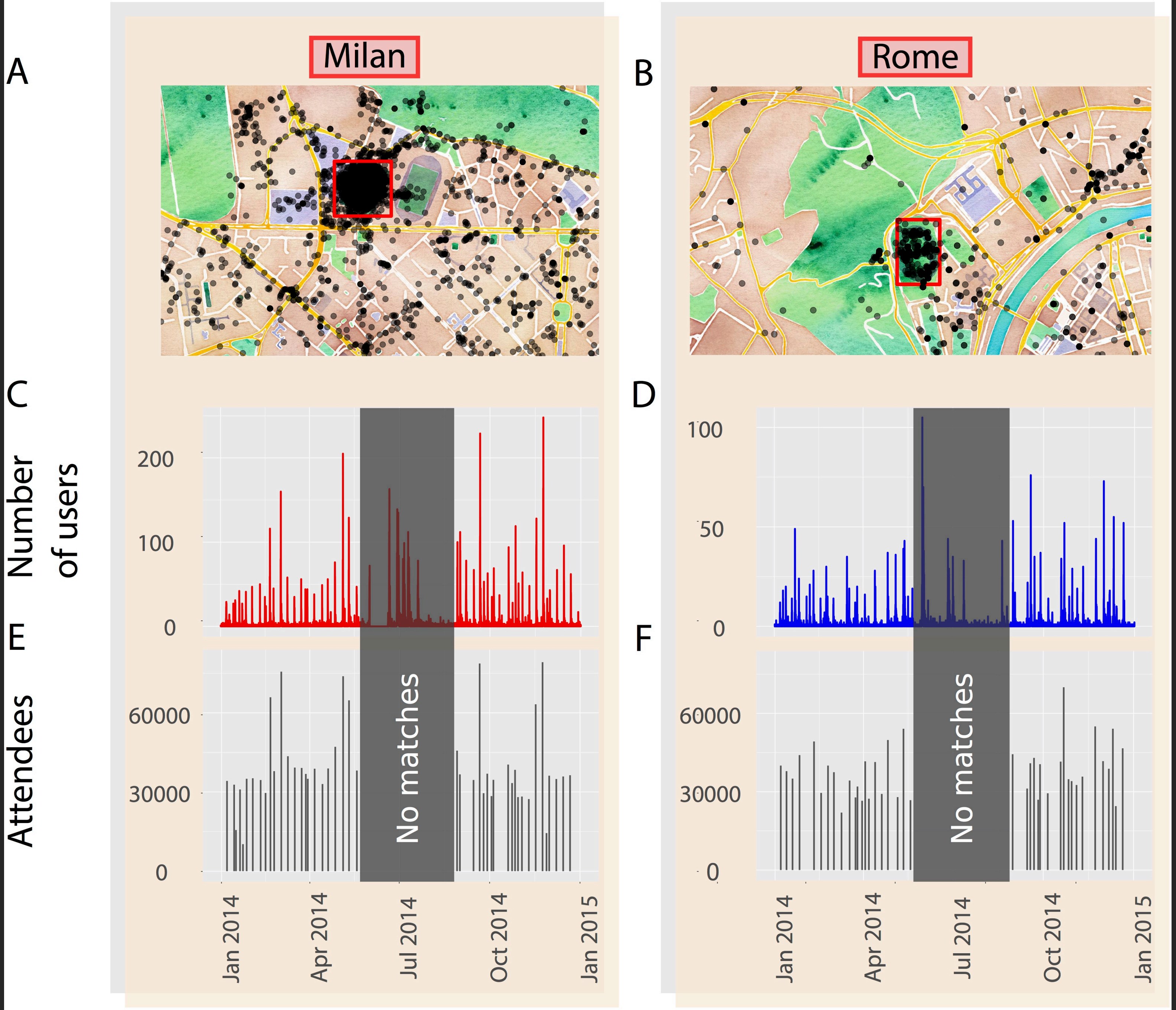


United Kingdom
Detected number of Flickr users
Official estimate of tourist arrivals
ρ = 0.83
Germany
Detected number of Flickr users
Official estimate of tourist arrivals
ρ = 0.88
New York
Washington
London
Paris
Rome



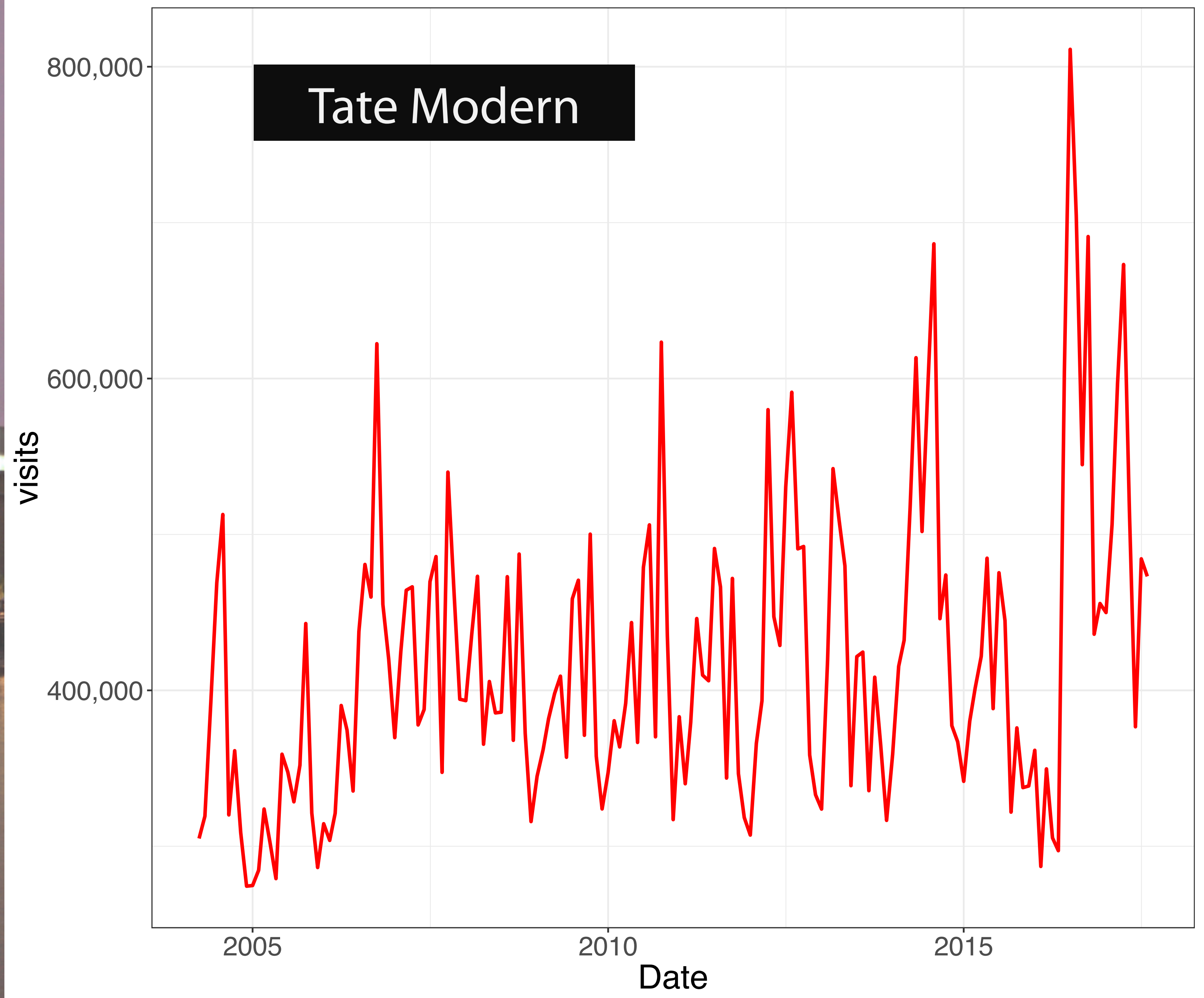
Canada
ρ = 0.87
United Kingdom
ρ = 0.83
Germany
ρ = 0.88
Japan
ρ = 0.65
United States
ρ = 0.72
France
ρ = 0.55
Italy
ρ = 0.76
Detected number of Flickr users
Official estimate of tourist arrivals
New York
London
Paris
Rome
San Francisco
Washington
Taipei
Tokyo
Beijing
Rio de Janeiro
Sydney
60°N
30°N
0°
30°S
60°S
120°W
60°W
0°
60°E
120°E
Number of Photos
50
500
5,000
50,000
500,000
Estimating Tourism Statistics in G7 Countries Using Flickr Photos

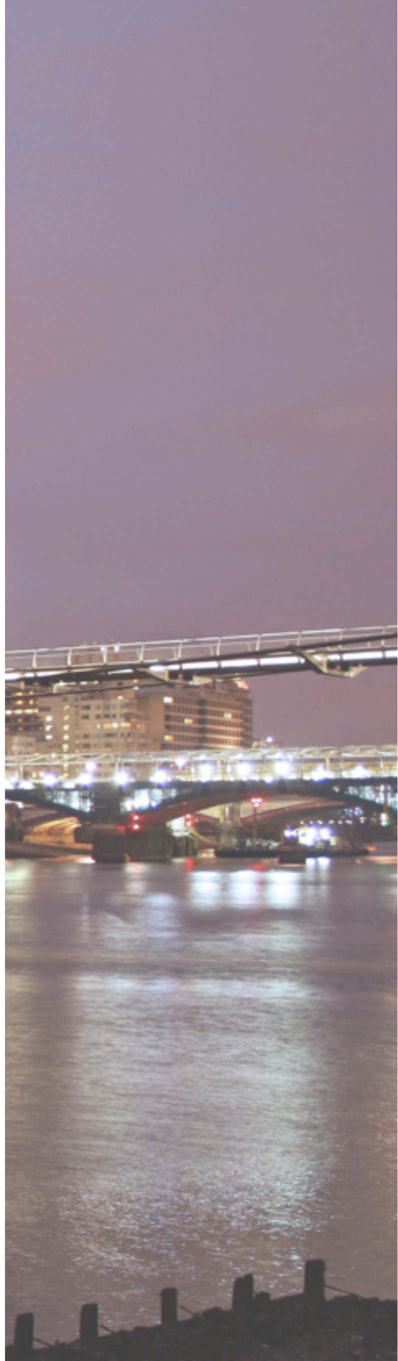

Rome, Italy
21 October 2014
41.95
41.92
41.89
Latitude
12.44
12.48
12.52
Longitude
Galleria Giovanni XXIII
Riserva Naturale di Monte Mario
Corso di Francia
Villa Glori
Villa Ada
Centri sportivi CONI
Centro Sportivo Esercito Tor di Quinto
Viale Liegi
Villa Borghese
Villa Torlonia
Tevere

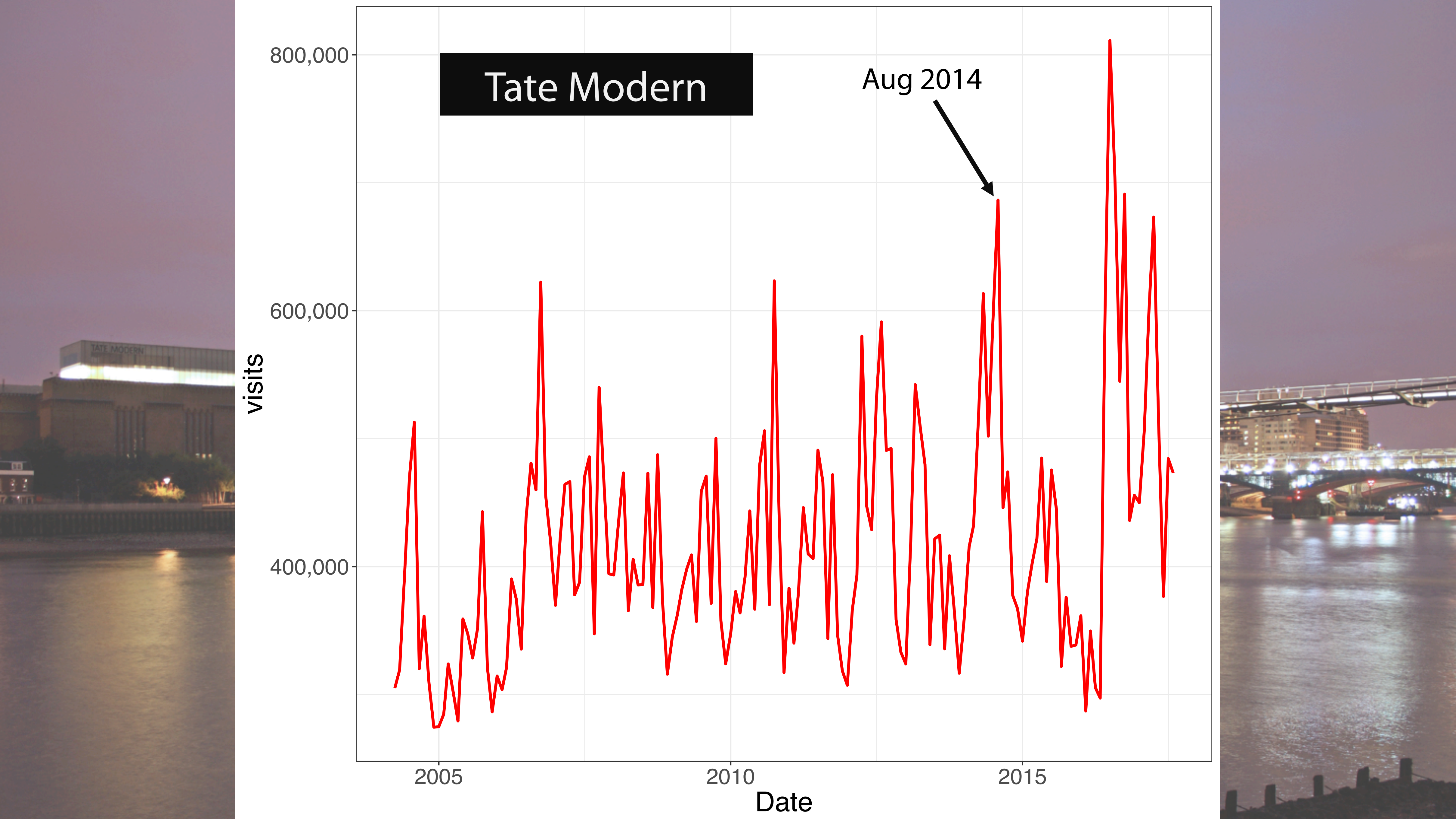


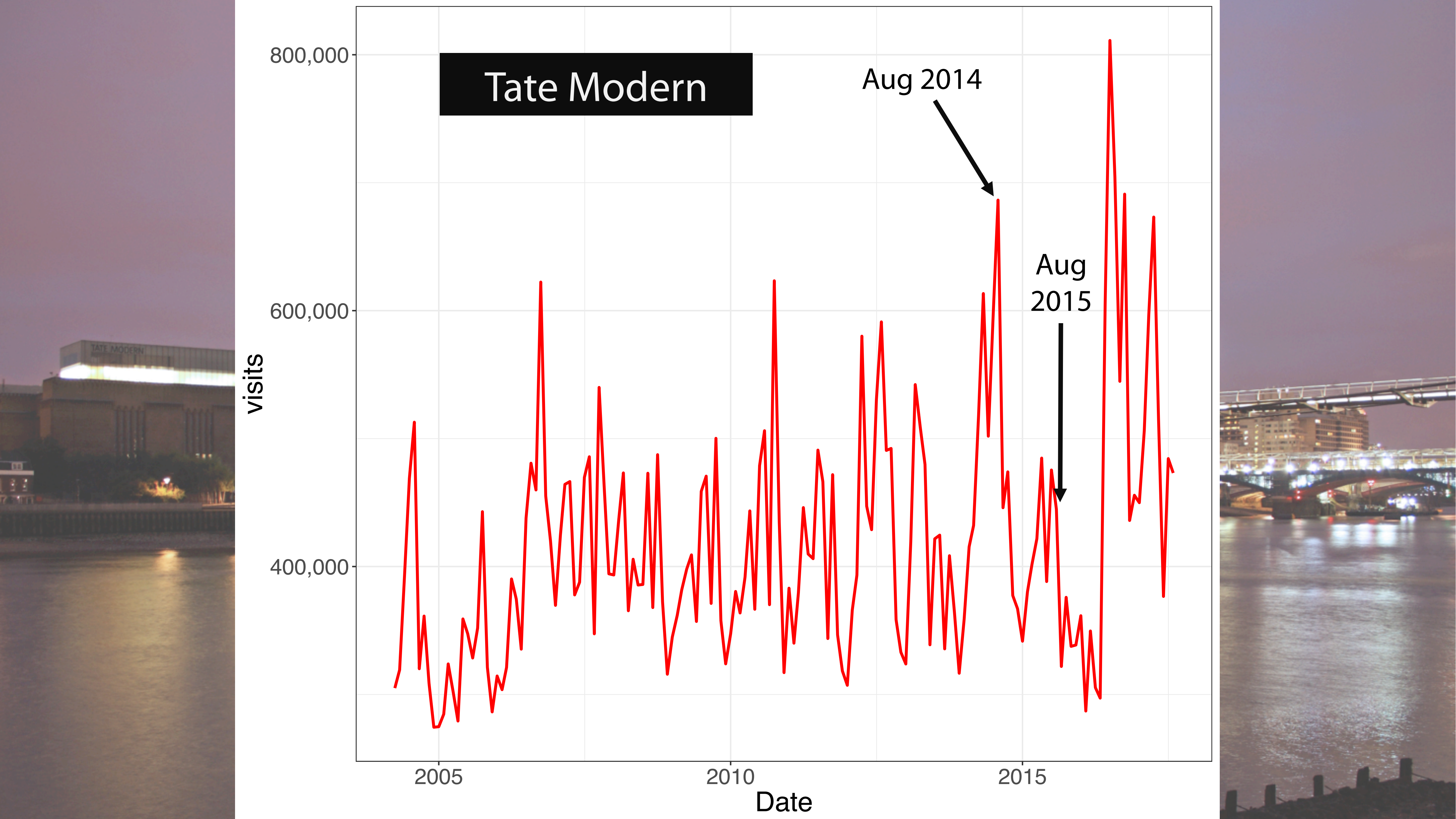
A
Milan
B
Rome
C
D
Number of users
200
100
0
100
50
0
E
F
Attendees
60000
30000
0
No matches
No matches
Jan 2014
Apr 2014
Jul 2014
Oct 2014
Jan 2015

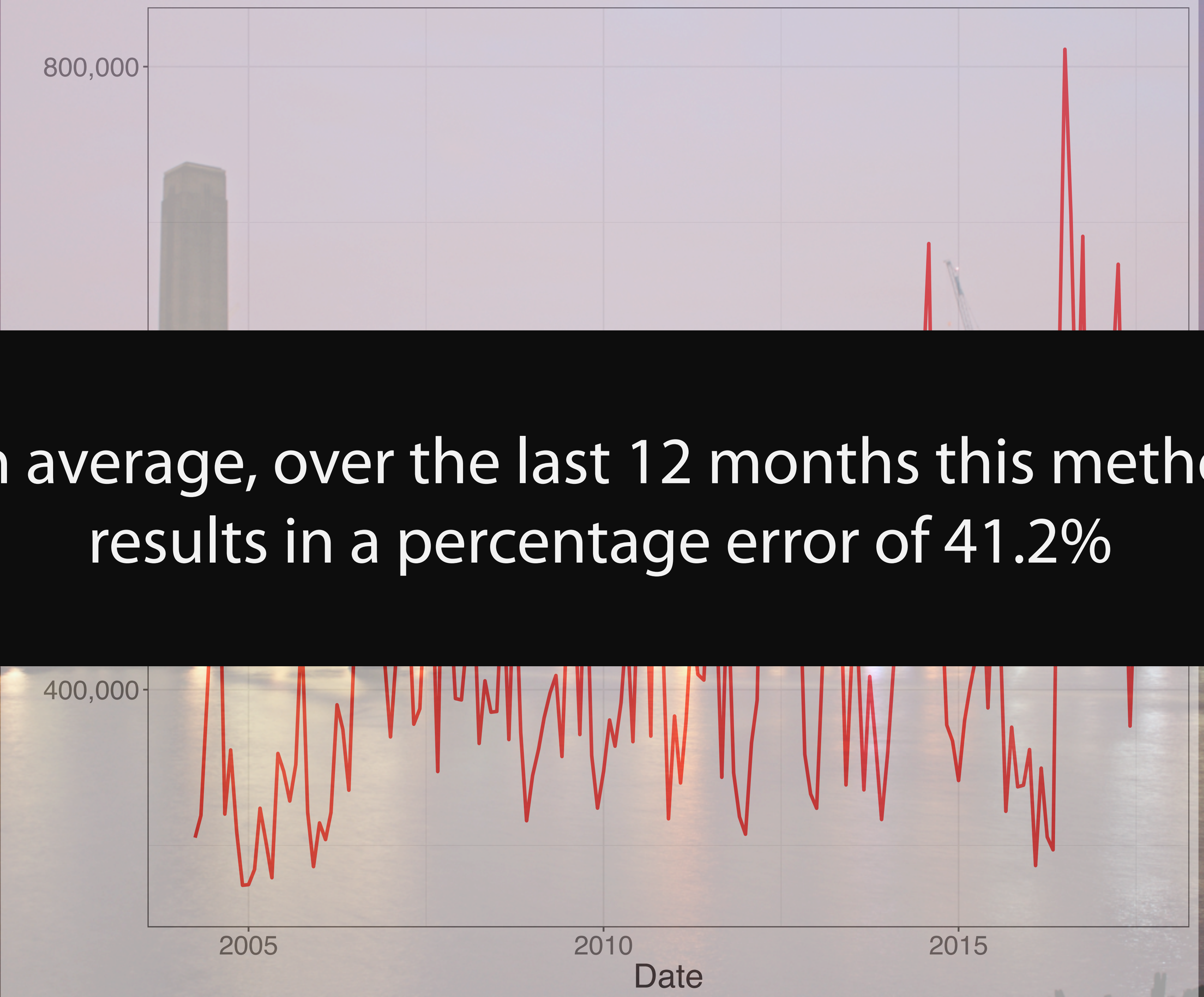
Department for
Digital, Culture
Media & Sport


Tate Modern
800,000
600,000
400,000
visits
2005
2010
2015
Date

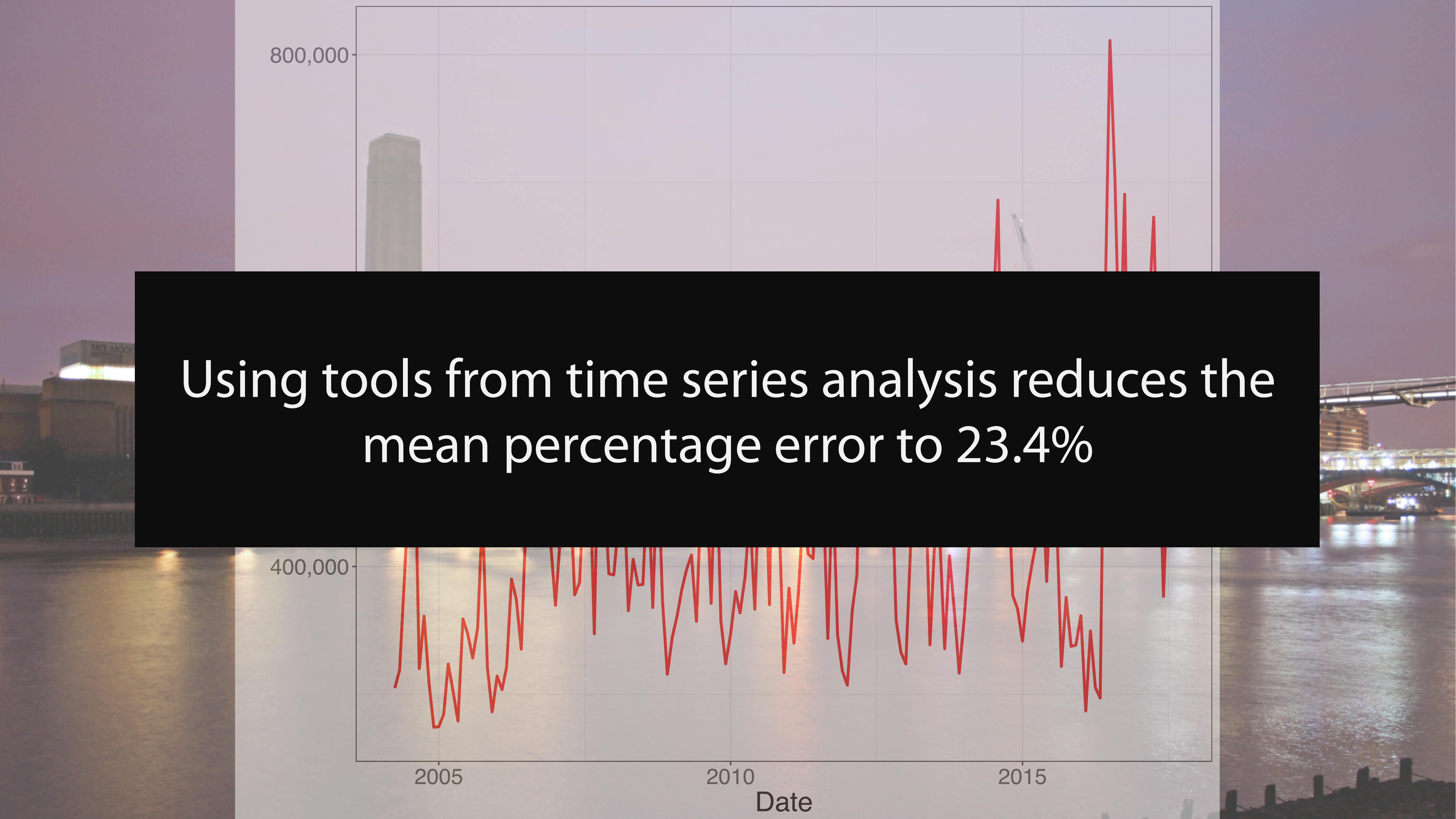



Tate Modern
Aug 2014
visits
800,000
600,000
400,000
2005
2010
2015
Date

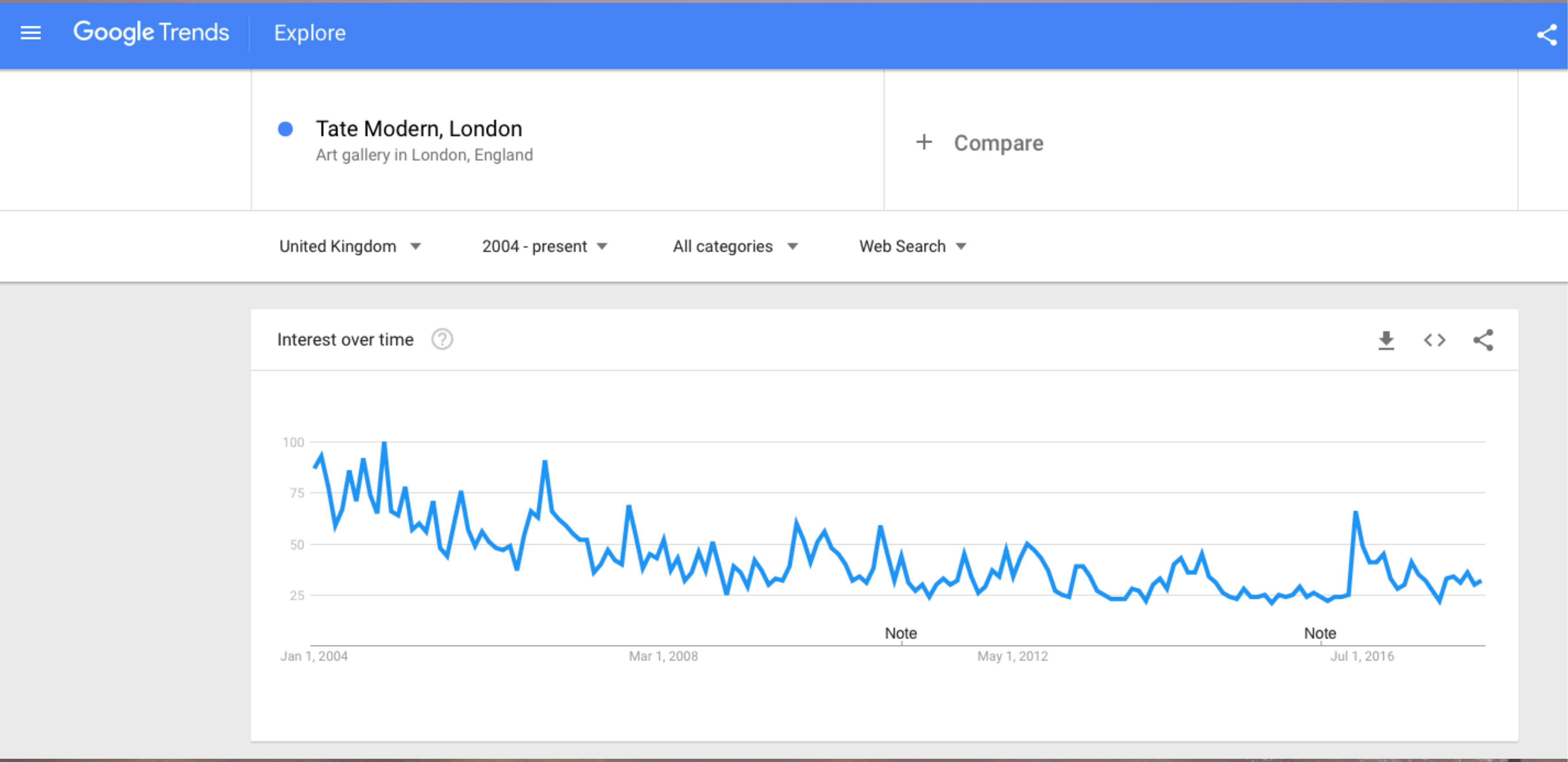

Tate Modern
Aug 2014
Aug
2015
800,000
600,000
400,000
visits
2005
2010
2015
Date



On average, over the last 12 months this method results in a percentage error of 41.2%



Using tools from time series analysis reduces the mean percentage error to 23.4%


Google Trends
Explore
Tate Modern, London
Art gallery in London, England
+ Compare
United Kingdom
2004 - present
All categories
Web Search
Interest over time
100
75
50
25
Note
Note
Jan 1, 2004
Mar 1, 2008
May 1, 2012
Jul 1, 2016


Google Trends
Explore
Tate Modern, London
Art gallery in London, England
+ Compare
Including data from Google Trends further reduces the mean percentage error to 14.5%
50
25
Note
Note
Jan 1, 2004
Mar 1, 2008
May 1, 2012
Jul 1, 2016

Data Science Lab

datasciencelab.co.uk

Estimates of museum and gallery visits using online data

November 2017

Federico Botta, Suzy Moat and Tobias Preis

Estimates for November 2017

Estimates are generated using a combination of figures from past visits to museums and galleries, together with data on how many people have been searching for those museums and galleries on the popular search engine *Google*. Confidence intervals represent the interval within which the estimates fall with 95% probability.

Museum	Oct 2017	Nov 2016	**Estimate Nov 2017**	**95% Confidence Interval**	Year-on-year change (%)	
Science Museum	276,299	213,130	204,634	*(177,000 - 232,268)*	↘	-4
British Museum	479,492	477,088	477,470	*(328,439 - 626,501)*	↗	0
Natural History	427,034	342,784	330,226	*(272,142 - 388,310)*	↘	-4
Tate Modern	512,262	436,004	421,087	*(287,684 - 554,490)*	↘	-3
Geffrye Museum	9,385	8,202	8,889	*(6,432 - 11,346)*	↗	8
Horniman Museum	85,513	45,291	52,566	*(38,410 - 66,722)*	↗	16
National Gallery	416,086	527,795	460,322	*(347,619 - 573,025)*	↘	-13
Victoria and Albert	350,904	228,418	273,130	*(216,200 - 330,061)*	↗	20
National Portrait Gallery	80,172	154,664	101,605	*(67,290 - 135,920)*	↘	-34

Comparison of estimates to true values for October 2017

Here, we report a comparison between our estimates for October 2017 and the actual figures recorded by the museums and galleries.

Museum	Sep 2017	Oct 2016	**Estimate Oct 2017**	**95% Confidence Interval**	Actual visits Oct 2017	Year-on-year change	
						Predicted	Actual
Science Museum	182,938	287,881	316,357	*(288,482 - 344,232)*	276,299	↗	↘
British Museum	422,497	541,954	473,923	*(324,652 - 623,194)*	479,492	↘	↘
Natural History	335,004	425,923	491,560	*(434,401 - 548,720)*	427,034	↗	↗
Tate Modern	406,788	681,188	520,418	*(386,115 - 654,720)*	512,262	↘	↘
Geffrye Museum	9,807	10,348	10,753	*(8,340 - 13,166)*	9,385	↗	↘
Horniman Museum	63,586	84,487	83,649	*(68,161 - 99,137)*	85,513	↘	↗
National Gallery	319,761	585,215	438,806	*(330,460 - 547,152)*	416,086	↘	↘
Victoria and Albert	495,227	252,484	354,660	*(293,452 - 415,868)*	350,904	↗	↗
National Portrait Gallery	83,437	184,001	125,777	*(92,893 - 158,661)*	80,172	↘	↘

Photo: Tom Richardson

SCENIC OR NOT

Rate Great Britain's pretty places

- Home
- FAQ
- Leaderboard

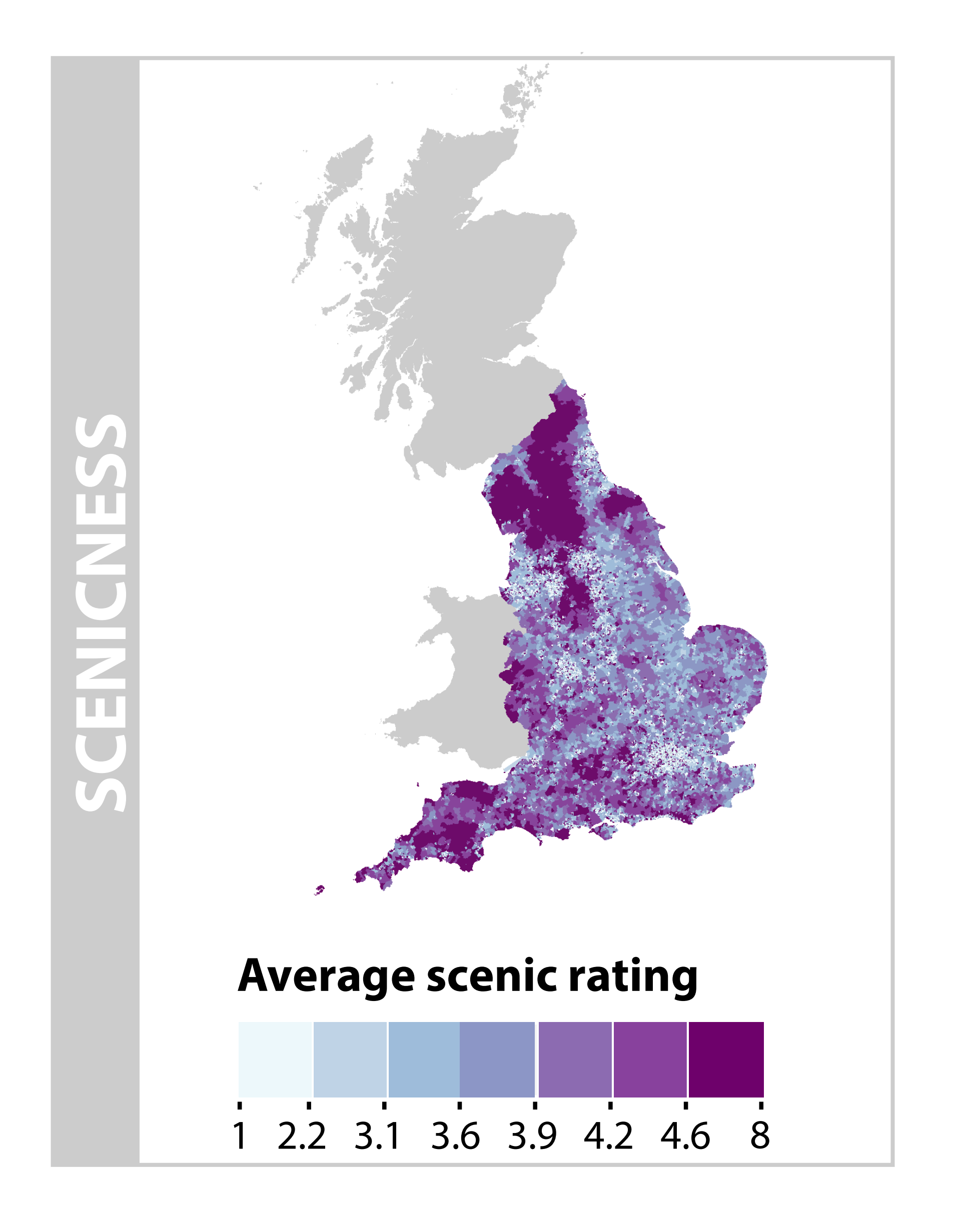
ScenicOrNot helps you to explore every corner of England, Scotland and Wales, all the while comparing your aesthetic judgements with fellow players.

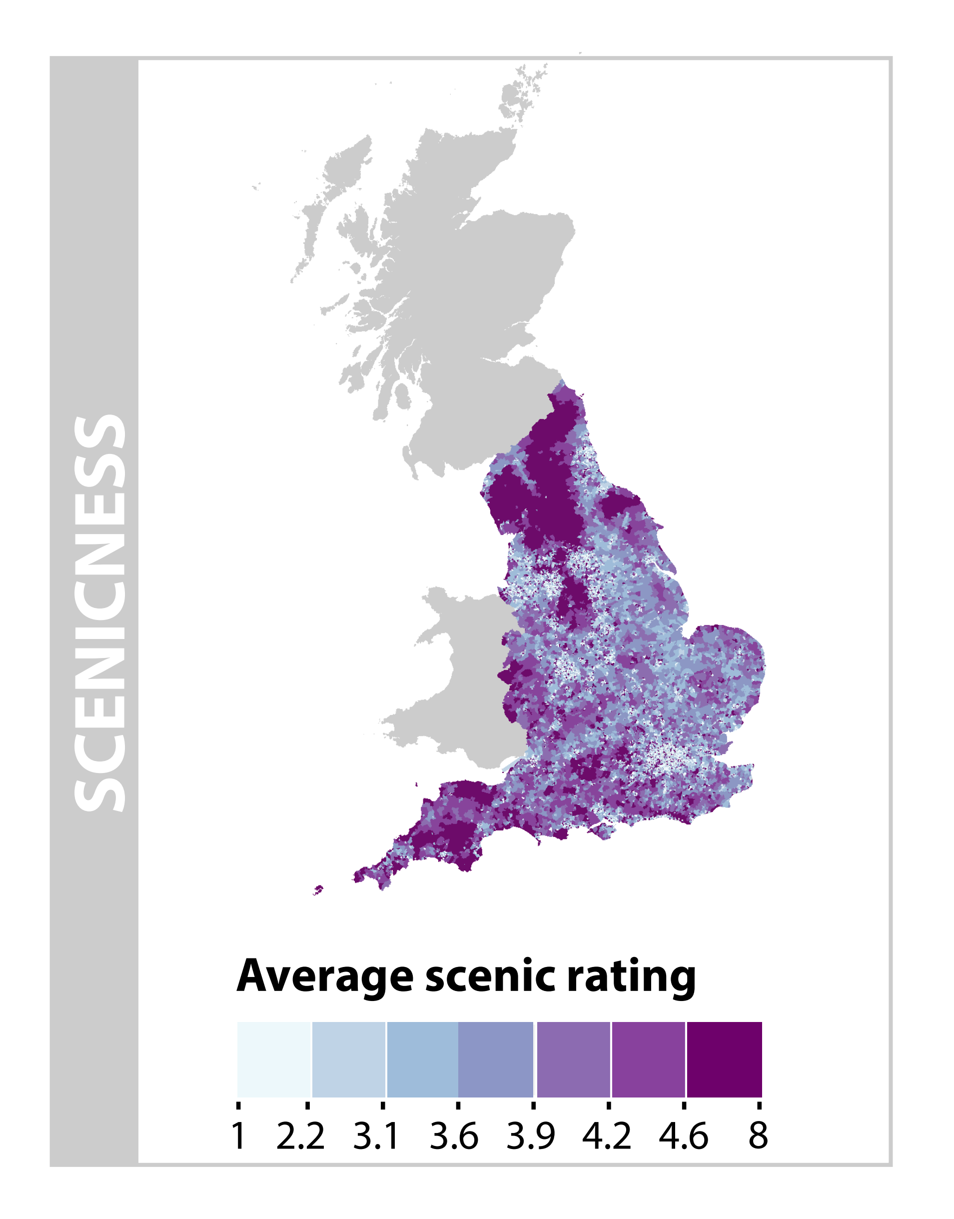
VOTING BOARD

- click numbers below to vote -

Not Scenic 1 2 3 4 5 6 7 8 9 10 Very Scenic

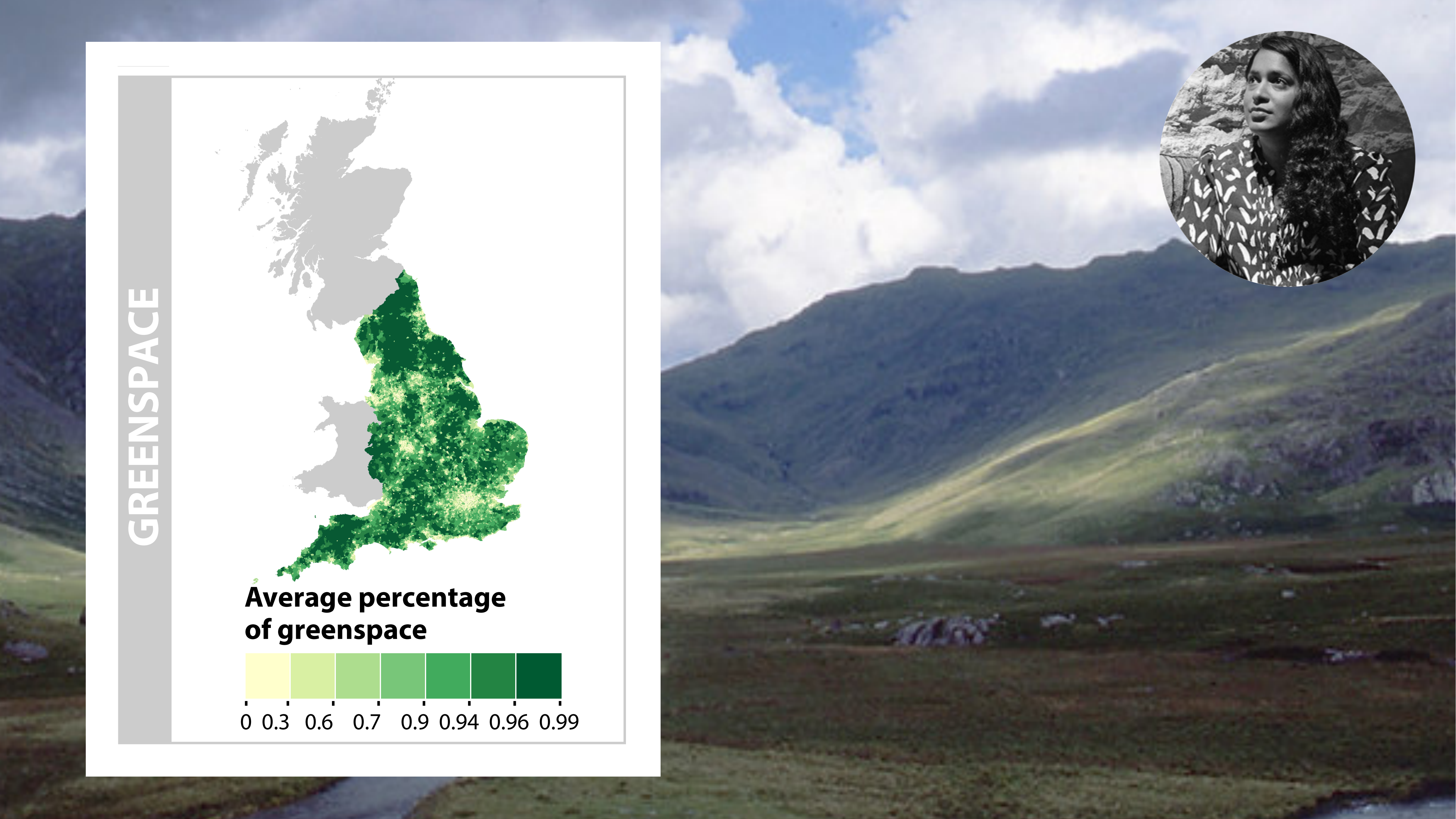
Photo by David Wild (Licence)

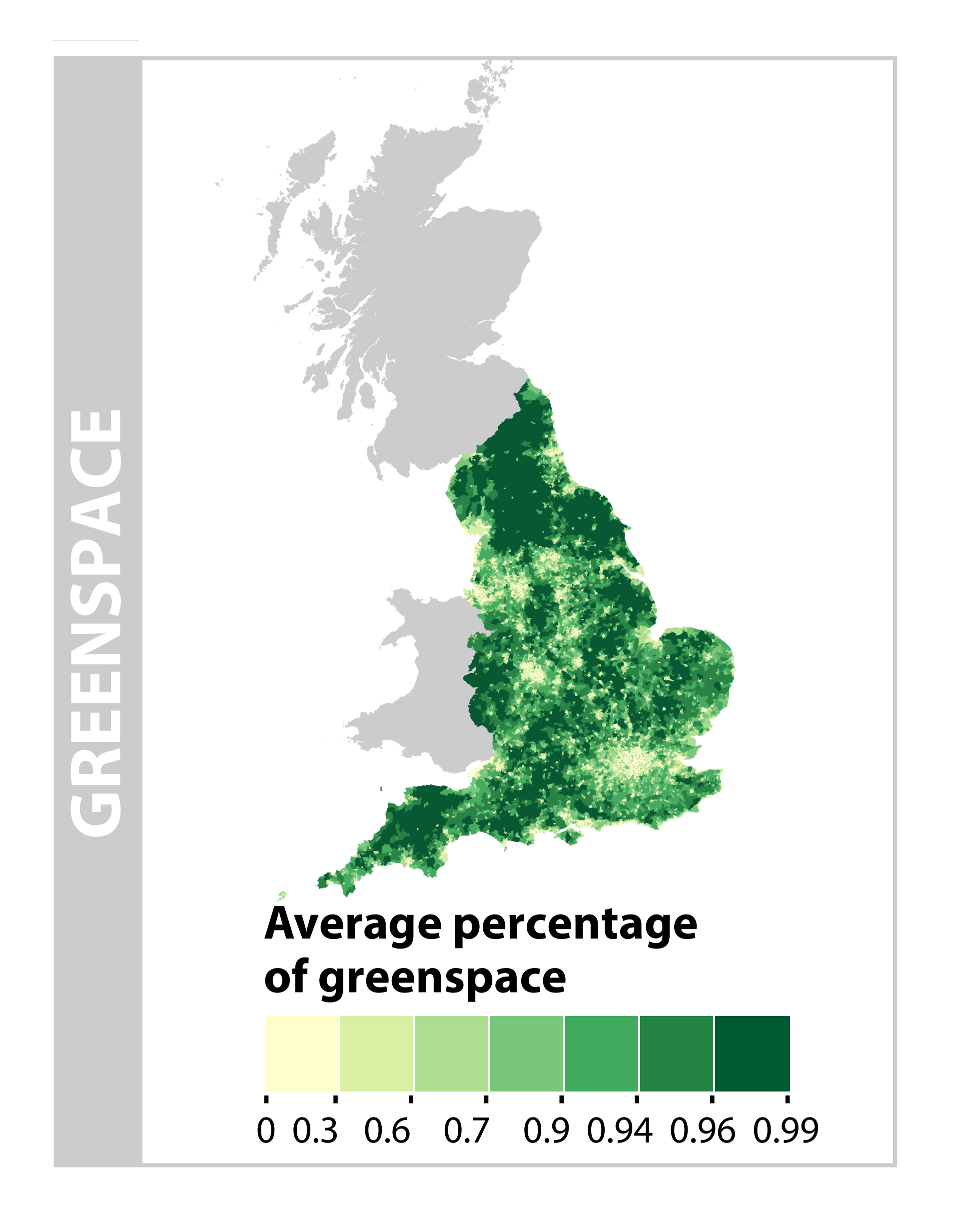

SCENICNESS
Average scenic rating
1
2.2
3.1
3.6
3.9
4.2
4.6
8

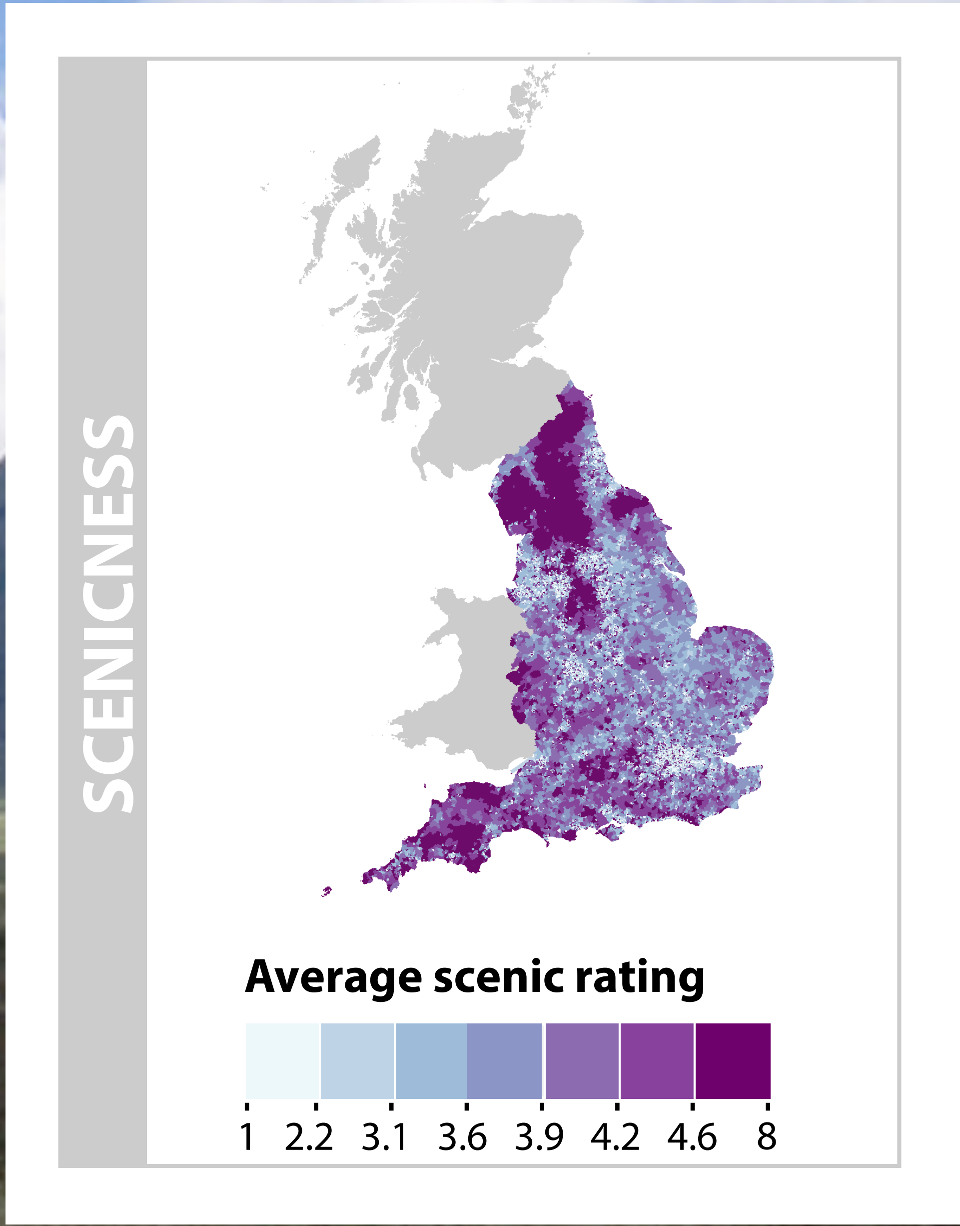


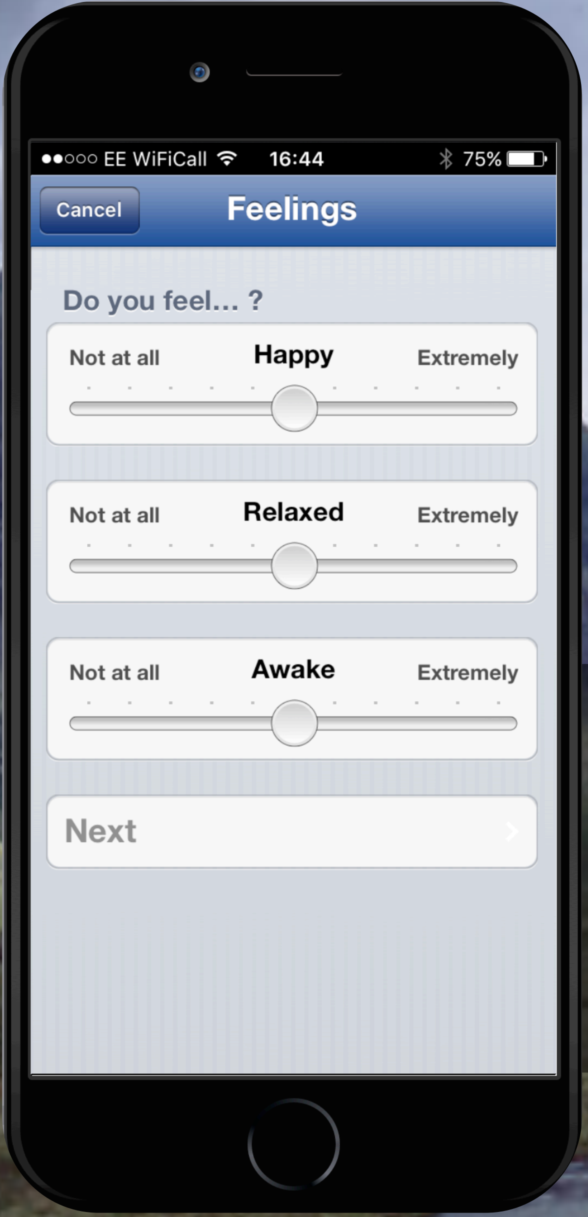
People who live in more scenic locations report better health

Seresinhe, Preis & Moat (2015)


GREENSPACE
Average percentage
of greenspace
0 0.3 0.6 0.7 0.9 0.94 0.96 0.99

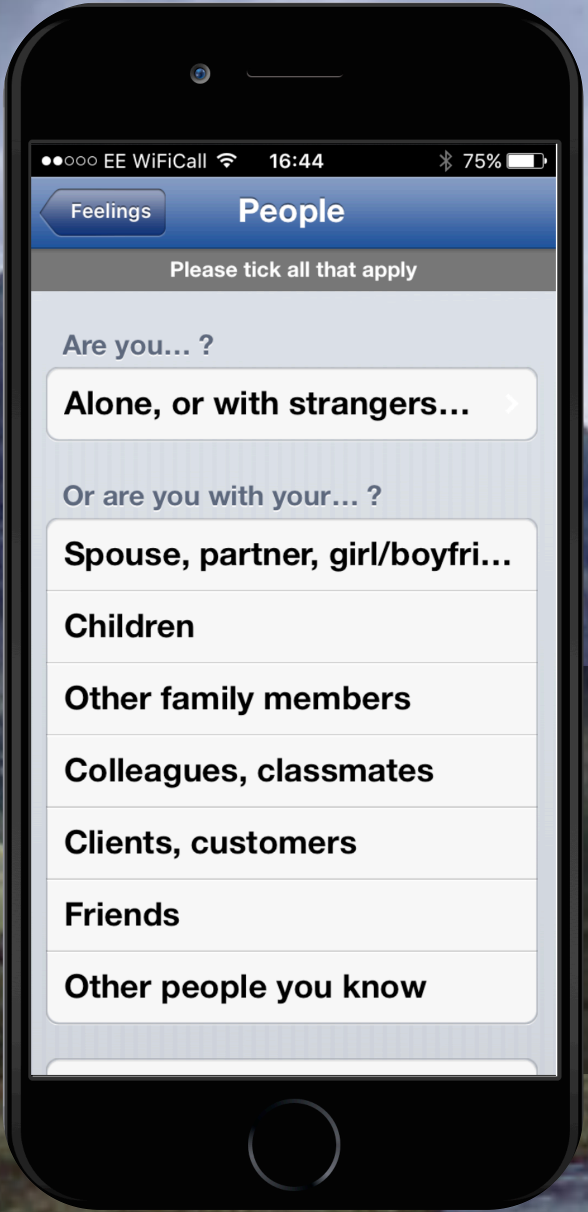

GREENSPACE
Average percentage of greenspace
0 0.3 0.6 0.7 0.9 0.94 0.96 0.99


SCENICNESS
Average scenic rating
1 2.2 3.1 3.6 3.9 4.2 4.6 8



George MacKerron's Mappiness

mappiness.org.uk

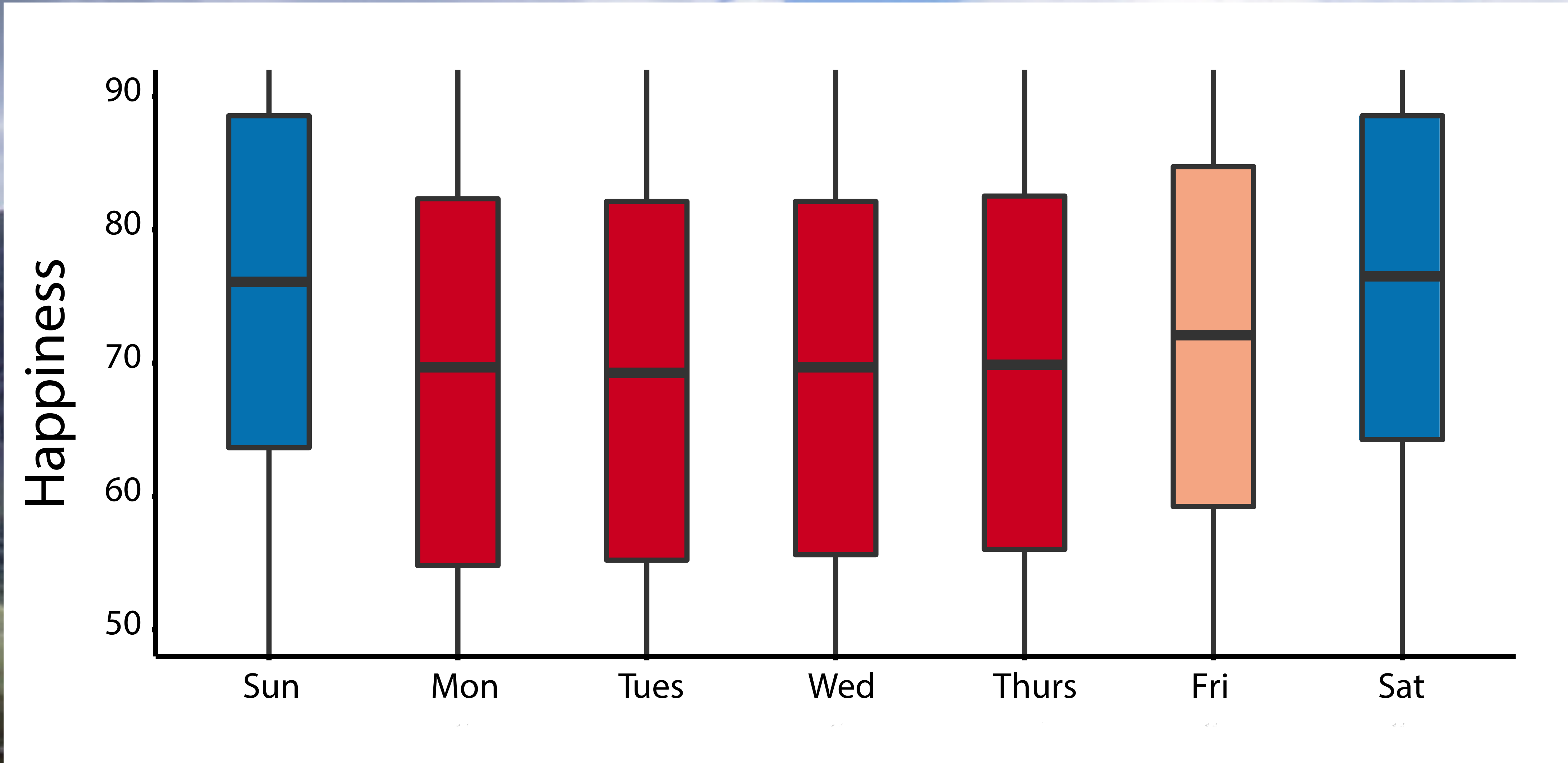


George MacKerron's Mappiness

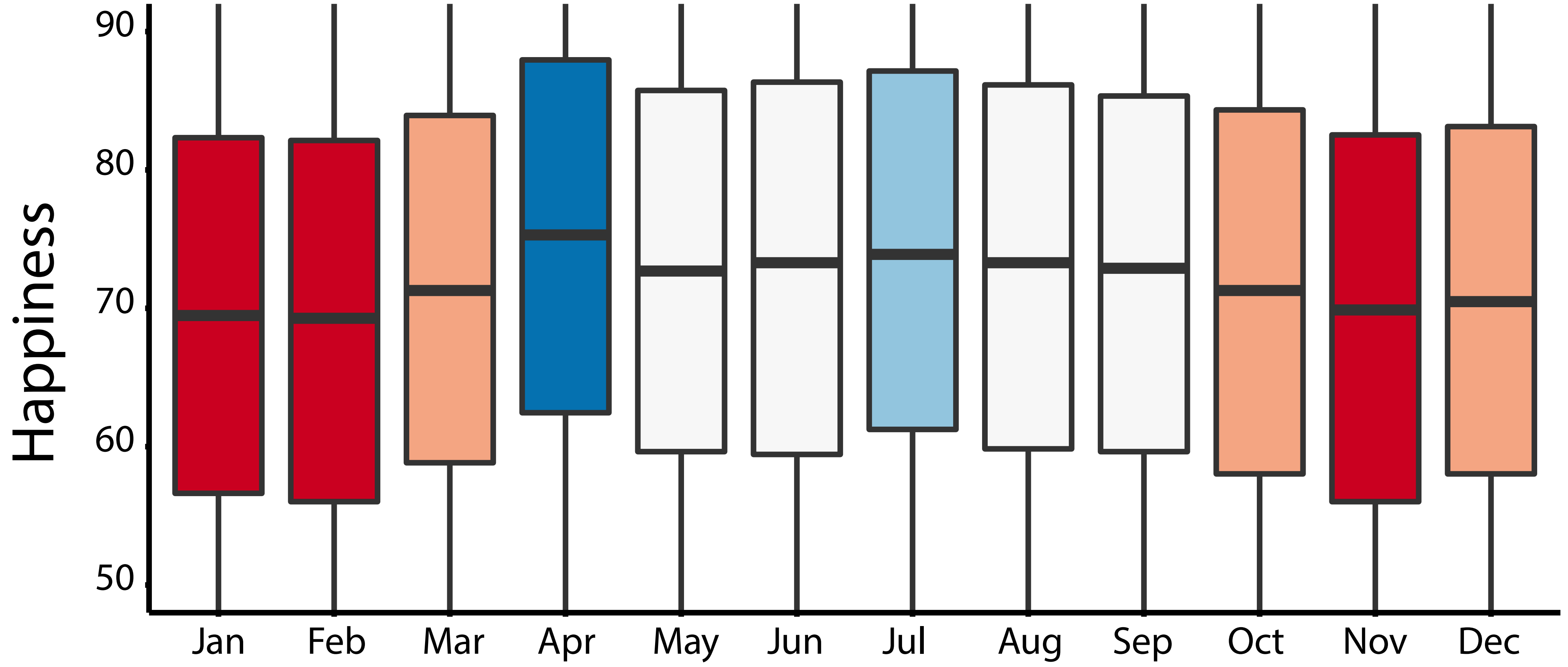
mappiness.org.uk

George MacKerron's Mappiness

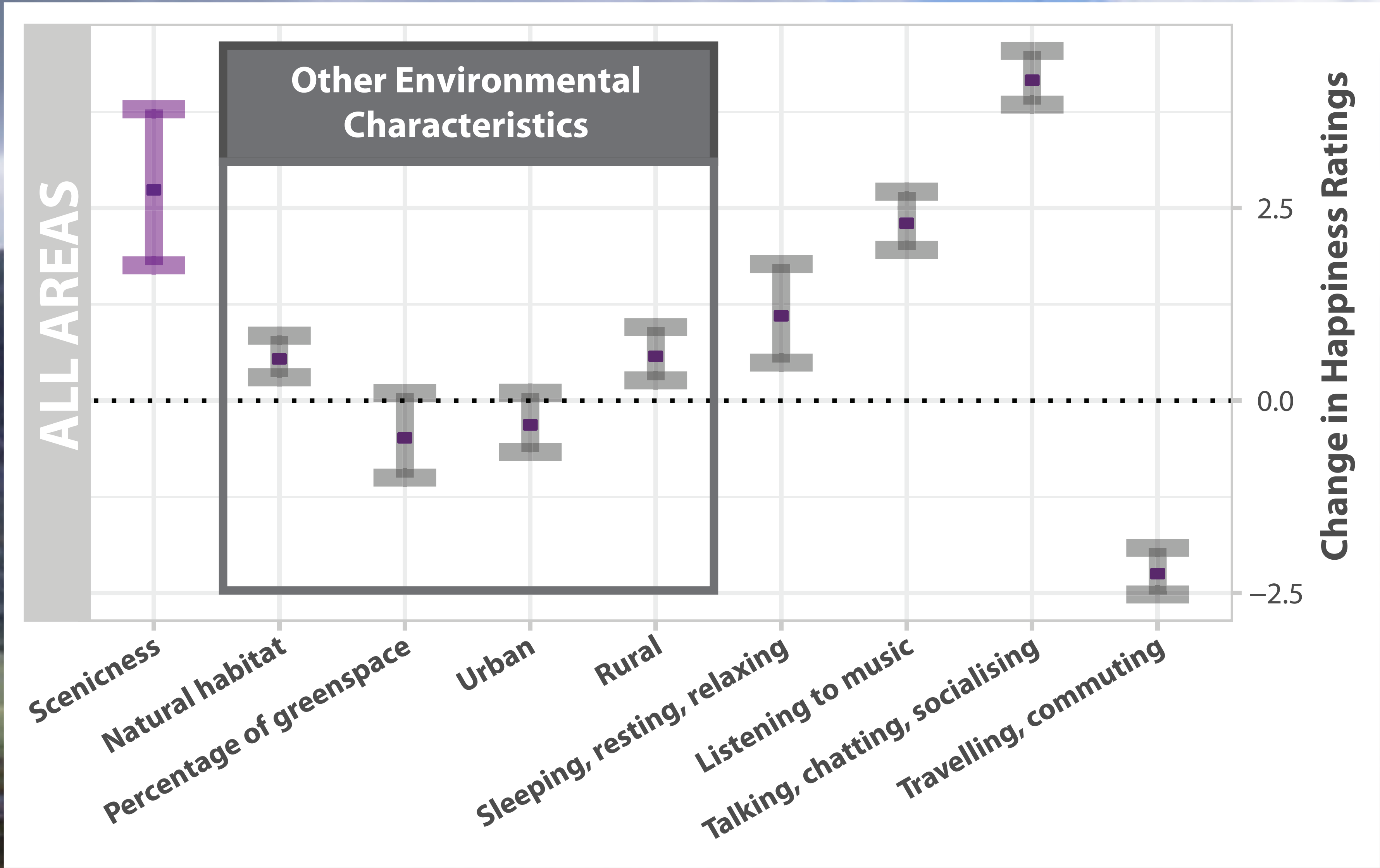
mappiness.org.uk



Seresinhe, Preis, MacKerron & Moat (under review)



Seresinhe, Preis, MacKerron & Moat (under review)



Seresinhe, Preis, MacKerron & Moat (under review)

Places365 Categories	*0.293*	Valley
	0.203	Lake Natural
	0.128	Mountain
SUN Scene Attributes	*0.856*	Natural Light
	0.081	Open Area
	0.058	Sailing / Boating

Seresinhe, Preis & Moat (2017)


Lake natural
+
Seresinhe, Preis
& Moat (2017)


Lake natural
+
Valley
+
Seresinhe, Preis
& Moat (2017)

Lake natural

+

Industrial area

-

Valley

+

Seresinhe, Preis & Moat (2017)

Lake natural

+

Industrial area

-

Valley

+

Hospital

-

Seresinhe, Preis & Moat (2017)


Cottage
+
Seresinhe, Preis
& Moat (2017)

Cottage

Viaduct

Seresinhe, Preis & Moat (2017)

Cottage

Trees

Viaduct

Seresinhe, Preis & Moat (2017)

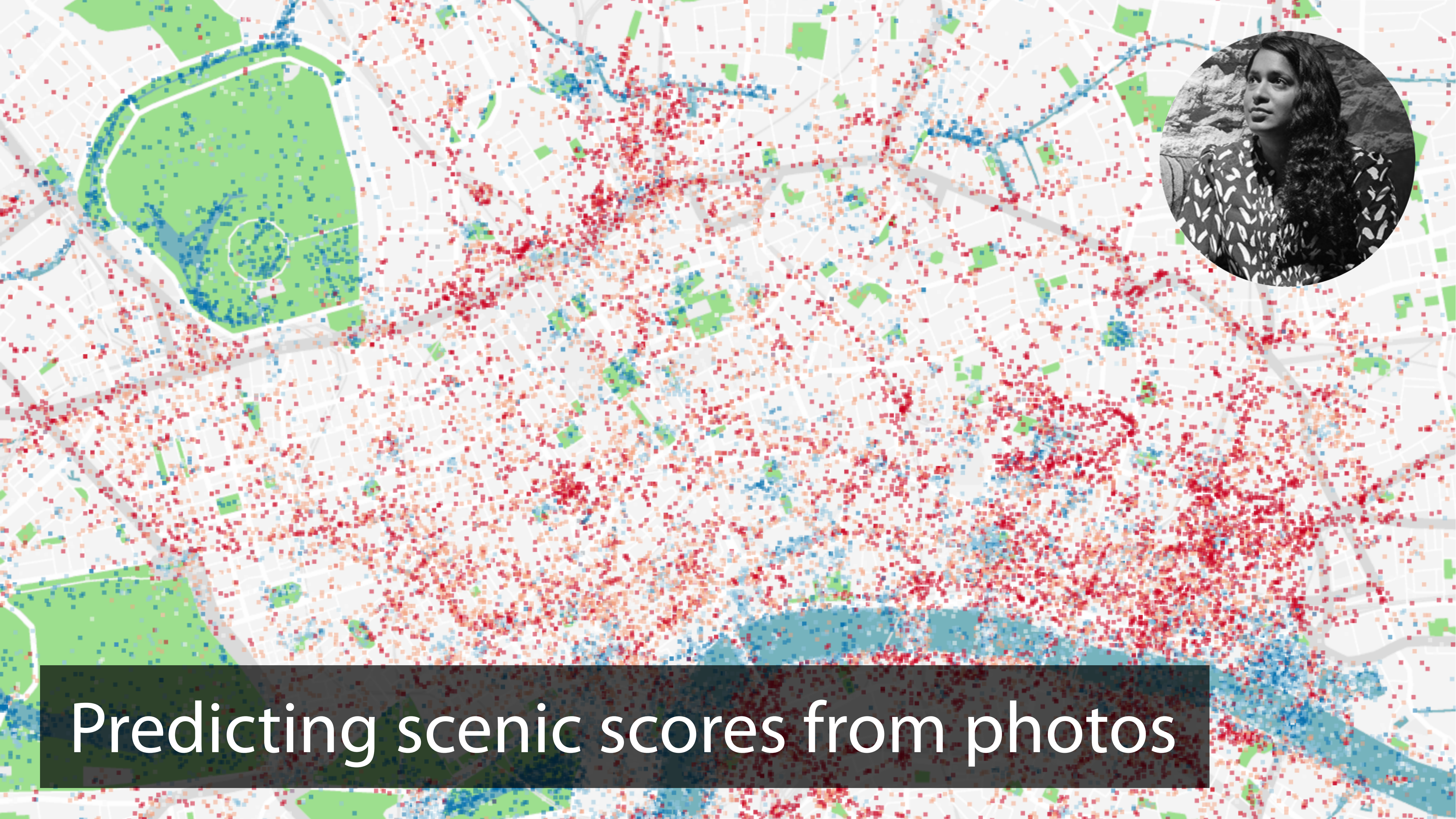
Cottage

Trees

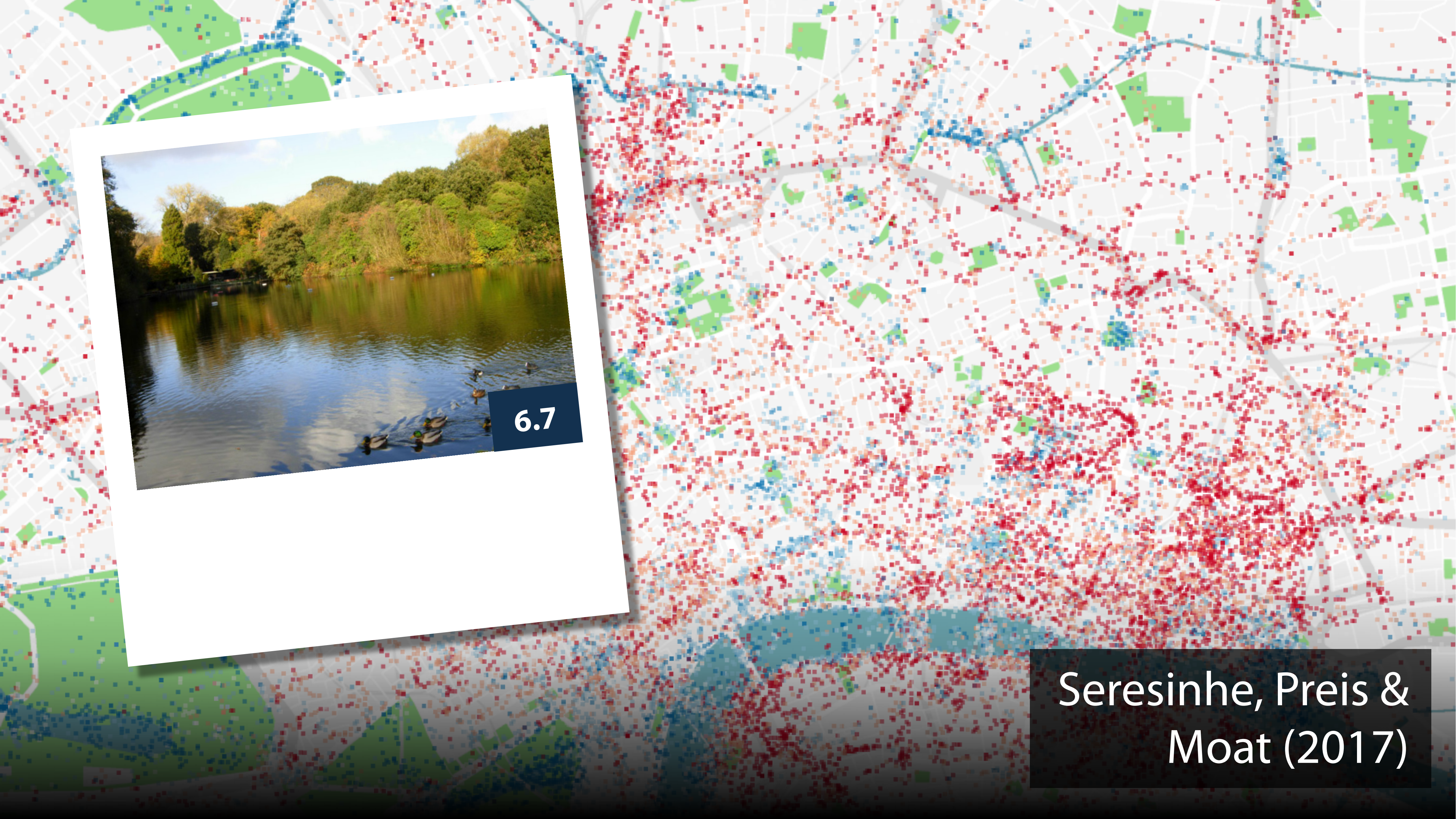
Viaduct

Grass

Seresinhe, Preis & Moat (2017)



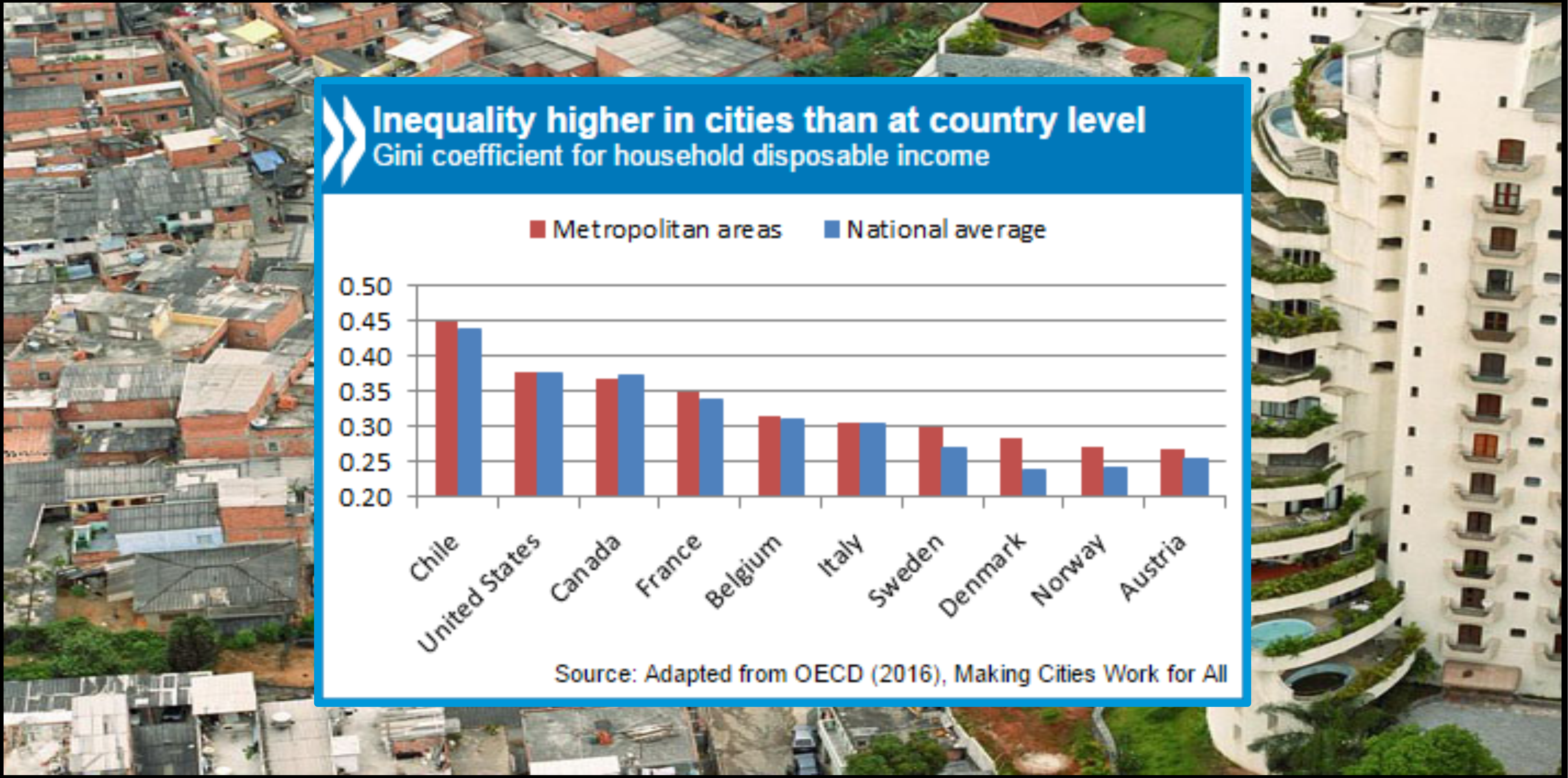
Predicting scenic scores from photos



Seresinhe, Preis & Moat (2017)


4.5
Seresinhe, Preis & Moat (2017)


4.7
Seresinhe, Preis & Moat (2017)



Inequality higher in cities than at country level
Gini coefficient for household disposable income
Metropolitan areas
National average
0.50
0.45
0.40
0.35
0.30
0.25
0.20
Chile
United States
Canada
France
Belgium
Italy
Sweden
Denmark
Norway
Austria
Source: Adapted from OECD (2016), Making Cities Work for All

Making Cities Work for All

DATA AND ACTIONS FOR INCLUSIVE GROWTH

OECD

How do we monitor inclusive growth in cities?

Policies for more inclusive growth in cities need to be supported by a solid evidence base. Measuring inclusive growth in cities, however, is no easy task, mainly for two reasons. First, inclusiveness filters through many dimensions beyond income and any measurement of it needs to include a wide array of variables, such as jobs, education, health or environment. However, such data are very scarce at the city level, even in advanced economies. Recent OECD work has mapped well-being outcomes according to 11 dimensions (material conditions and quality of life), both in countries and within countries in the 395 subnational OECD regions (OECD, 2011; 2014b). Building on the

How's Life in Your Region framework, Chapter 2 of this report sets about filling the evidence gap by providing data on selected well-being outcomes at the metropolitan spatial scale. These data allow a comparison of how OECD cities fare on income, jobs, education, environment and income inequality. But much still needs to be done to improve the availability of statistics at the city level. Alternative sources of data – such as administrative records, open government data and big data – will help overcome the current sampling limitations of national household surveys by increasing the amount of data on households and individuals at smaller geographical scales, which can then be aggregated up to the geography of interest.

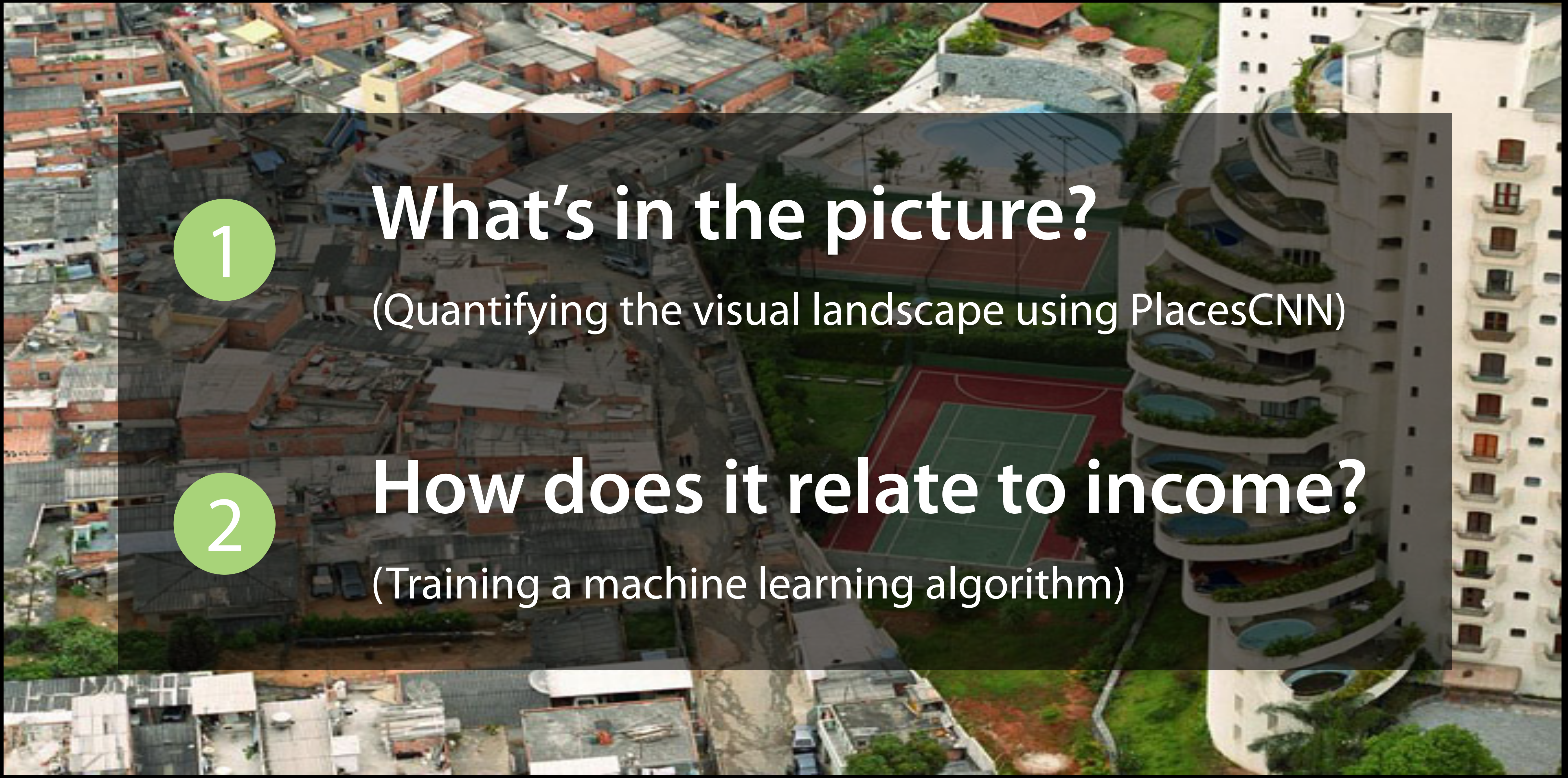
Second, measuring inclusive growth in cities requires taking into account several

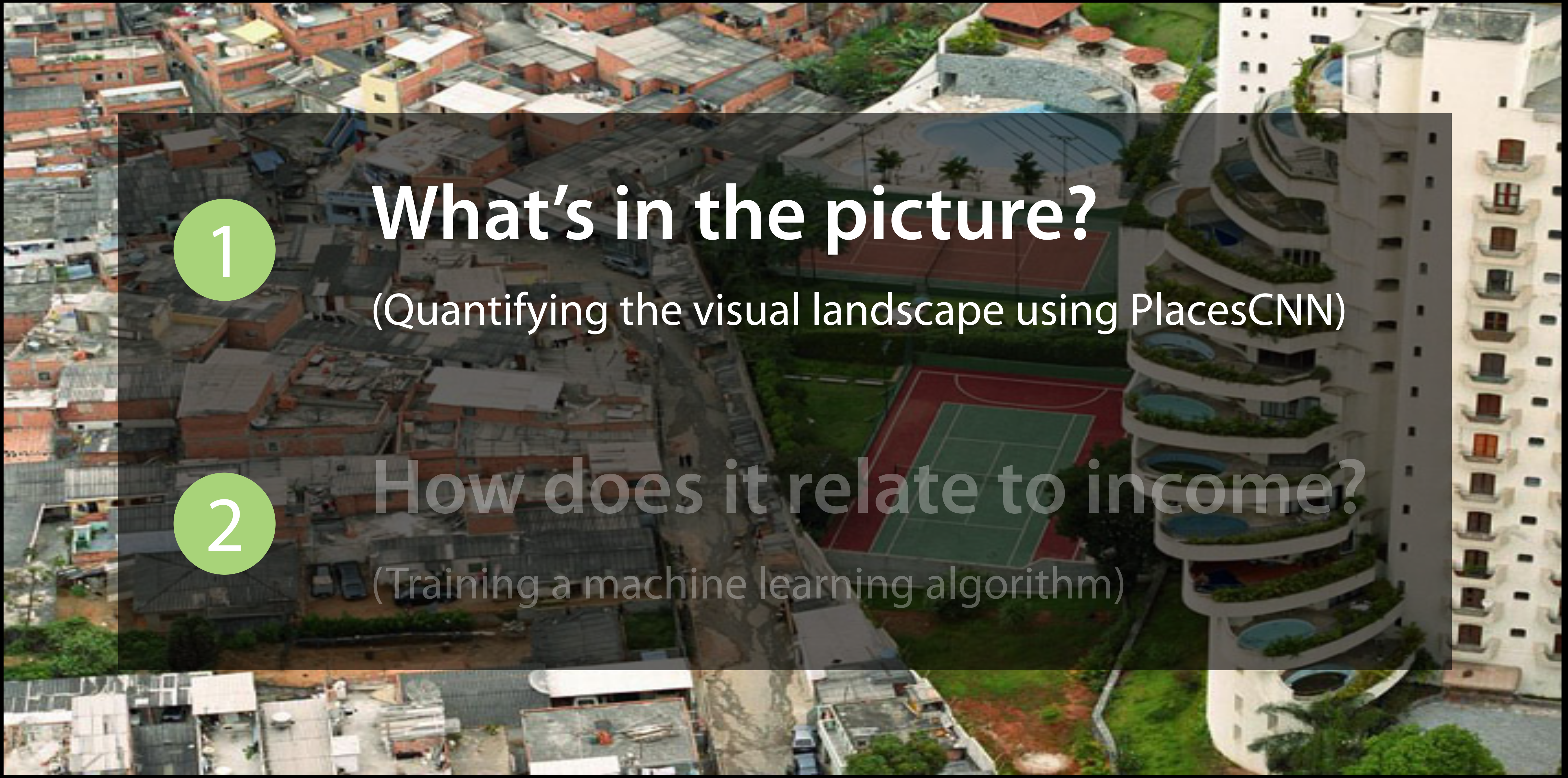
Using Google Street View pictures

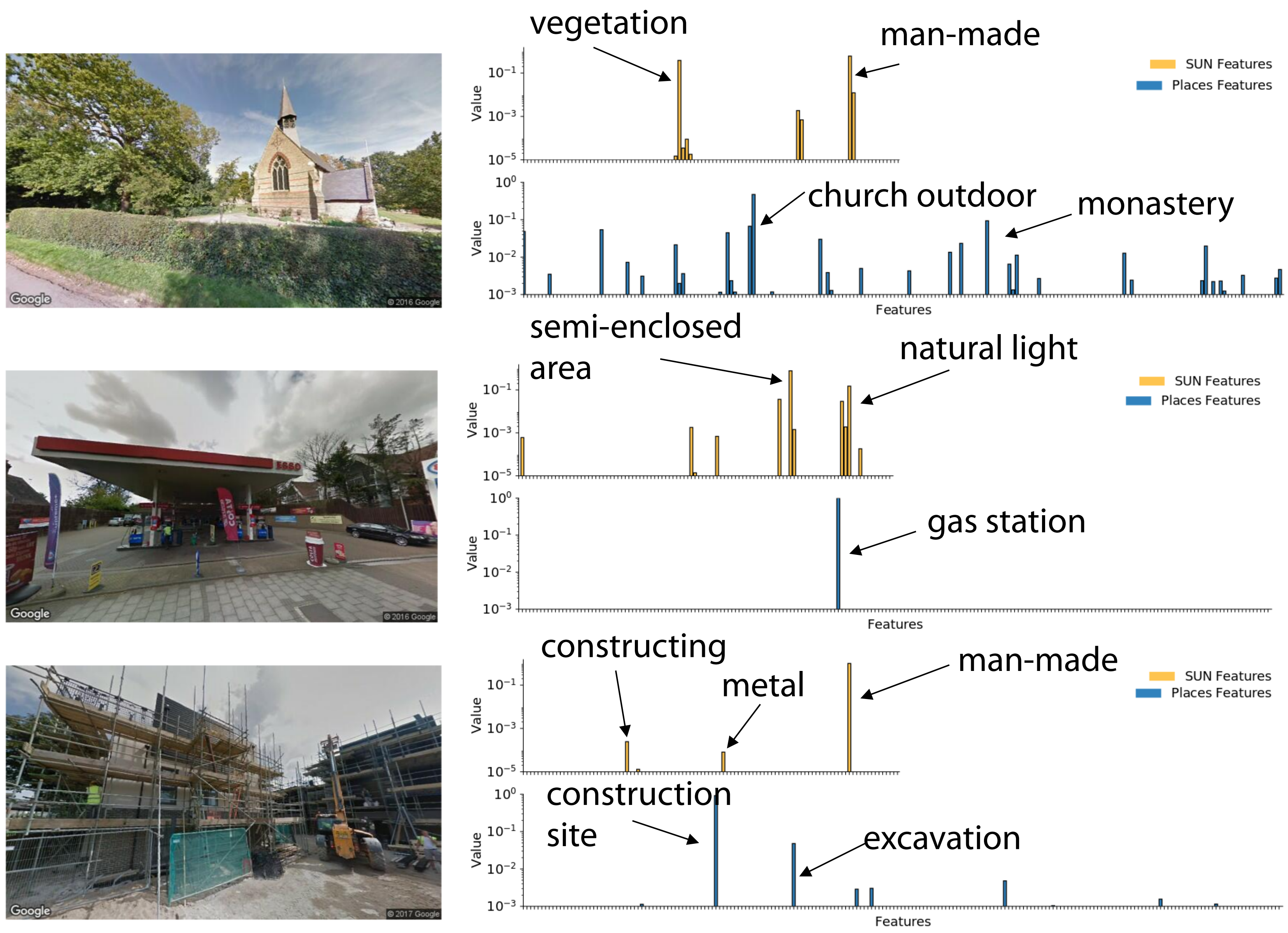
Which ones correspond to highest / lowest income areas?

Using Google Street View pictures

Which ones correspond to highest / lowest income areas?


1
What's in the picture?
(Quantifying the visual landscape using PlacesCNN)
2
How does it relate to income?
(Training a machine learning algorithm)


1
What's in the picture?
(Quantifying the visual landscape using PlacesCNN)
2
How does it relate to income?
(Training a machine learning algorithm)



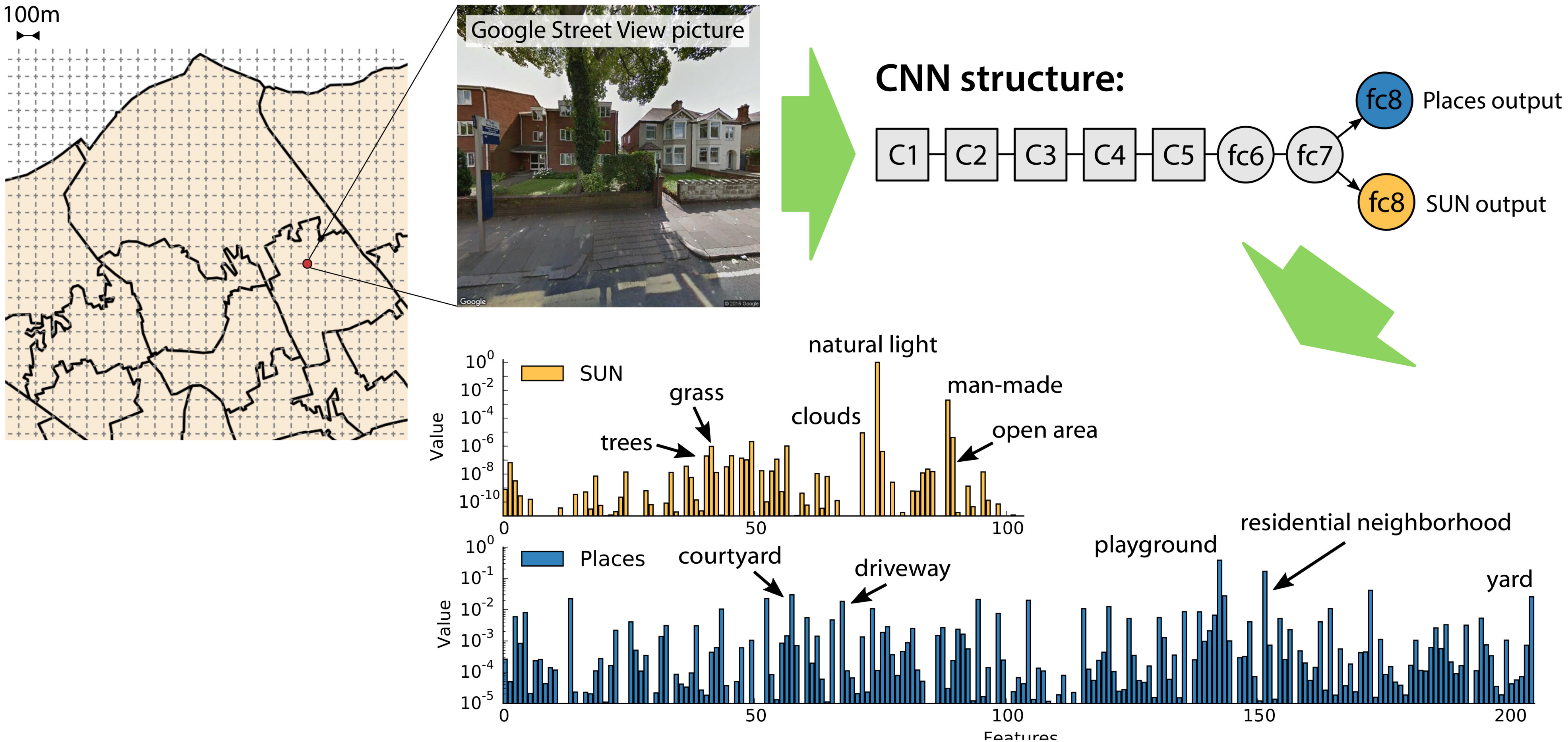
Places CNN

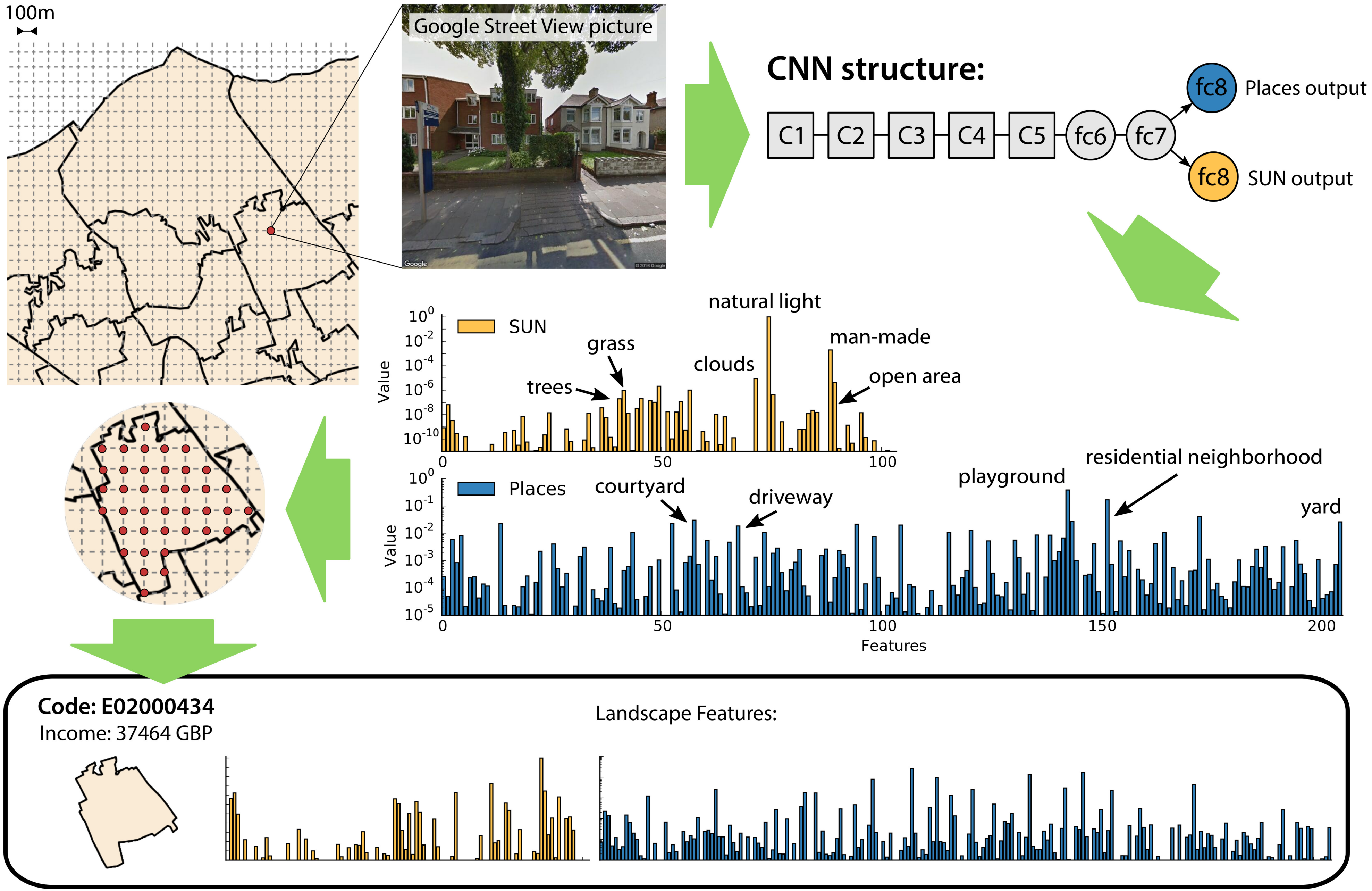
MIT

London

(MSOA)

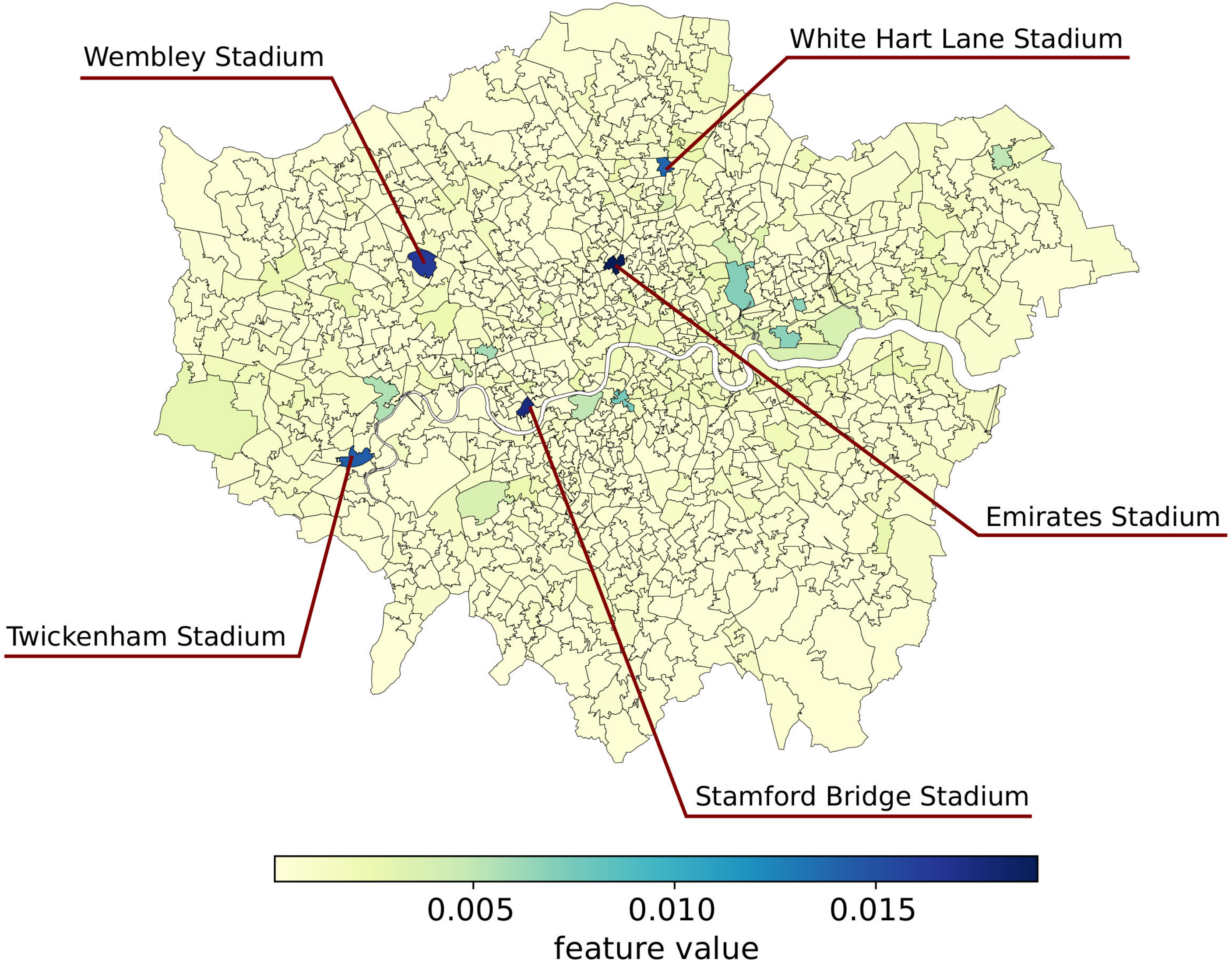

100m
Google Street View picture
Google
©2016 Google


100m
Google Street View picture
CNN structure:
C1
C2
C3
C4
C5
fc6
fc7
fc8
Places output
fc8
SUN output
SUN
natural light
grass
clouds
man-made
open area
trees
Value
Places
courtyard
driveway
playground
residential neighborhood
yard
Features

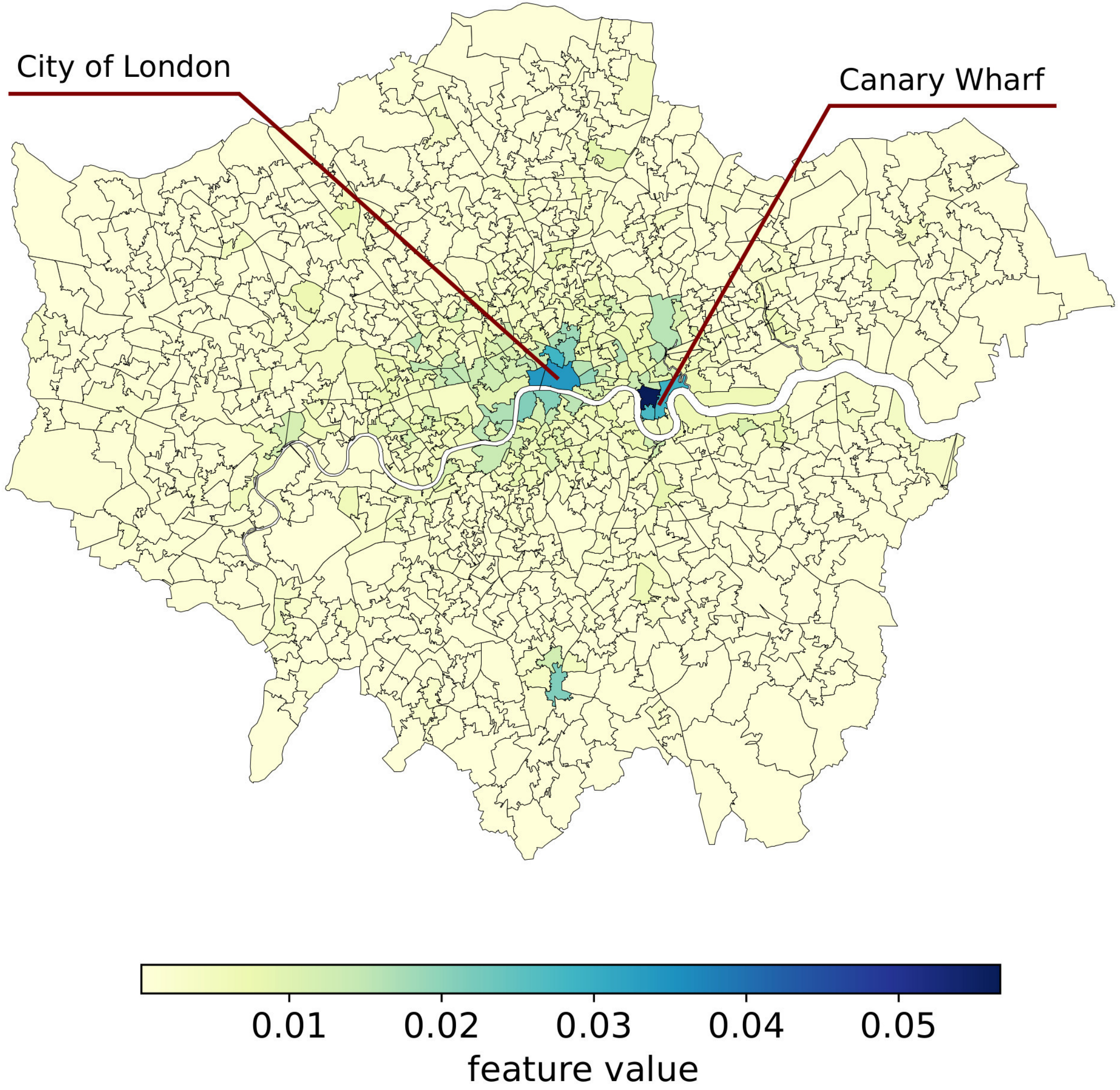


100m
Google Street View picture
Google
CNN structure:
C1
C2
C3
C4
C5
fc6
fc7
fc8
Places output
fc8
SUN output
SUN
natural light
man-made
grass
clouds
open area
trees
Value
Places
courtyard
driveway
playground
residential neighborhood
yard
Features
Code: E02000434
Income: 37464 GBP
Landscape Features:

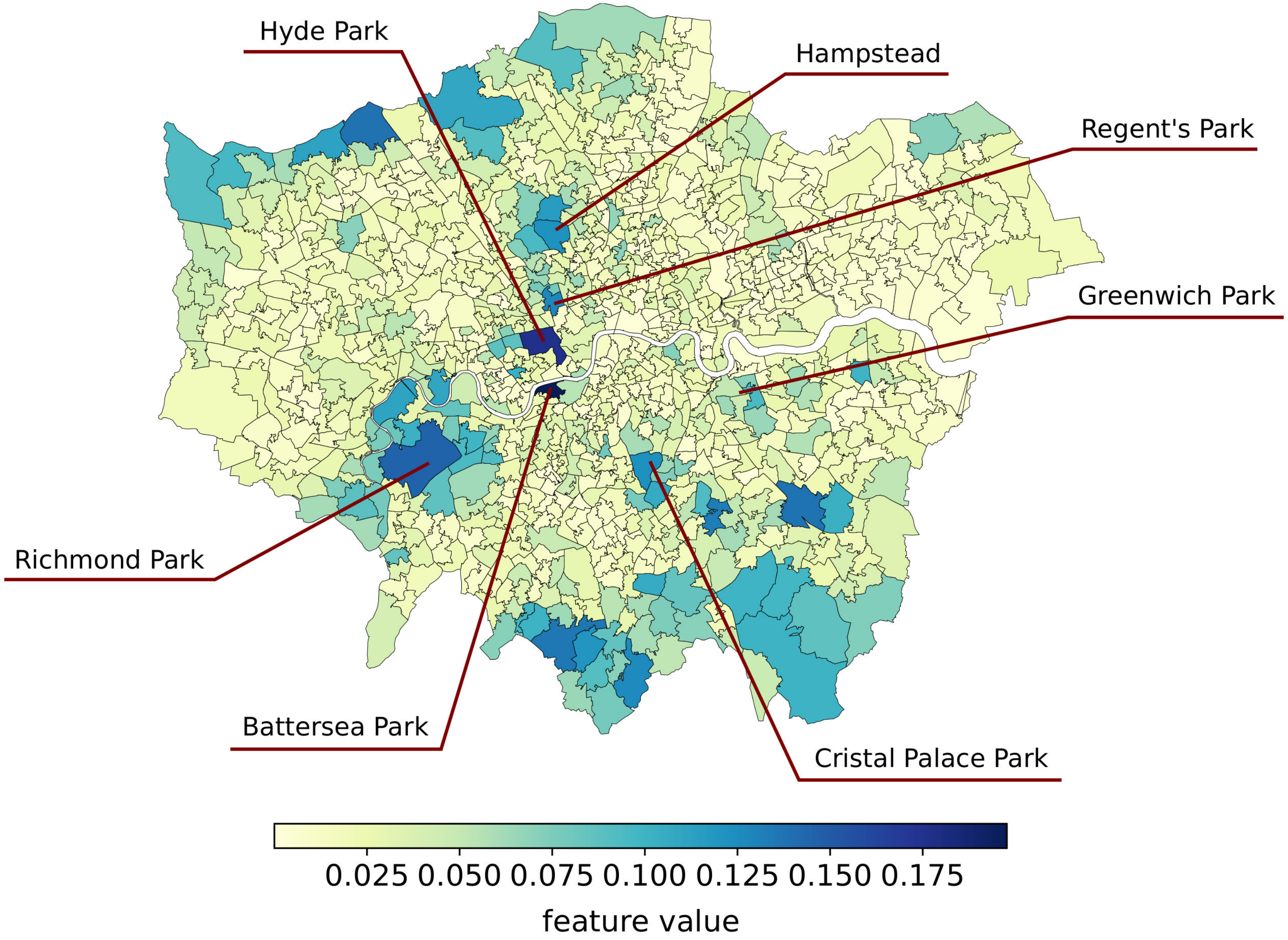
Feature: Stadium/football



Feature: Skyscraper



Feature: Trees



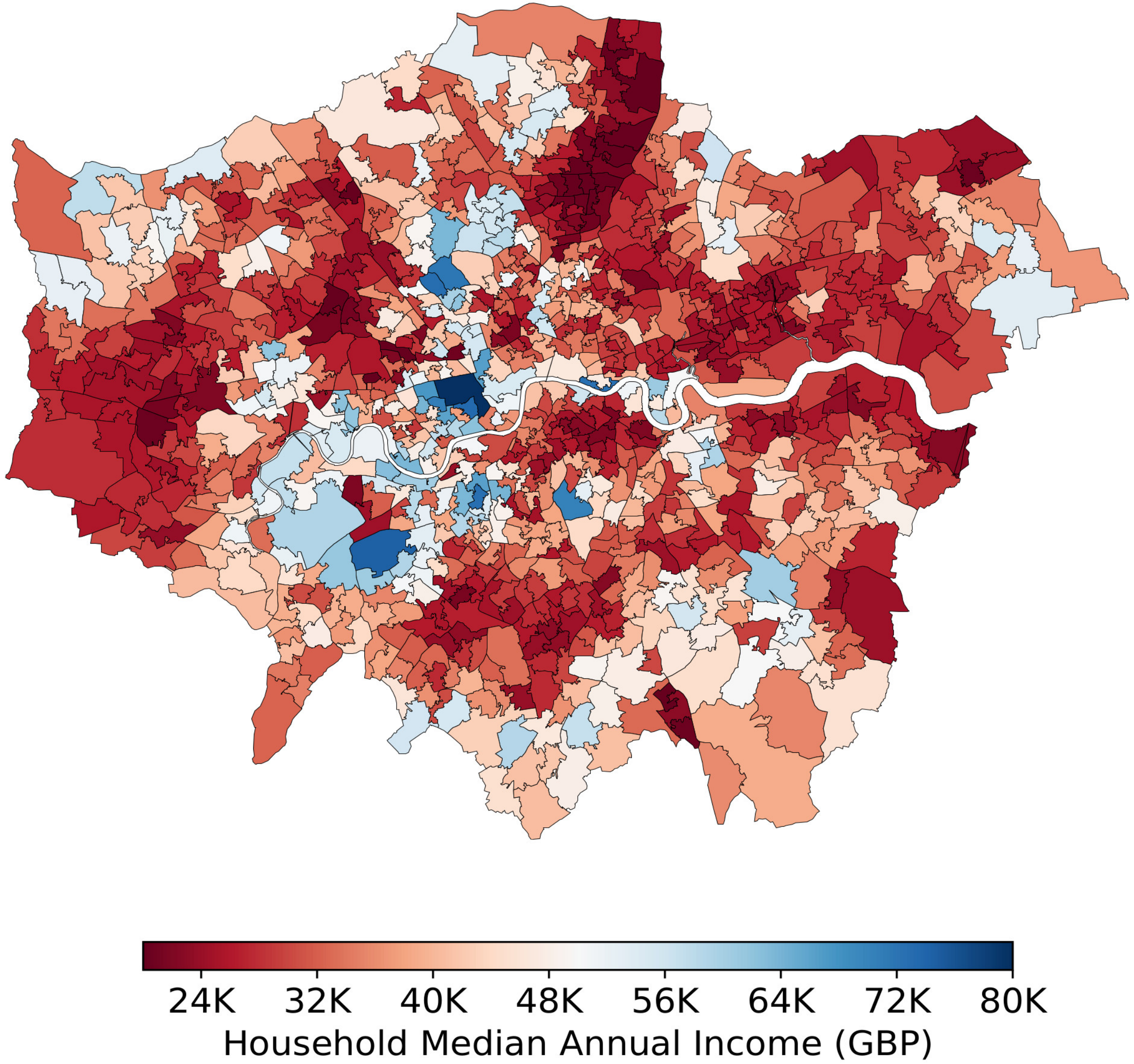
1 What's in the picture?

(Quantifying the visual landscape using PlacesCNN)

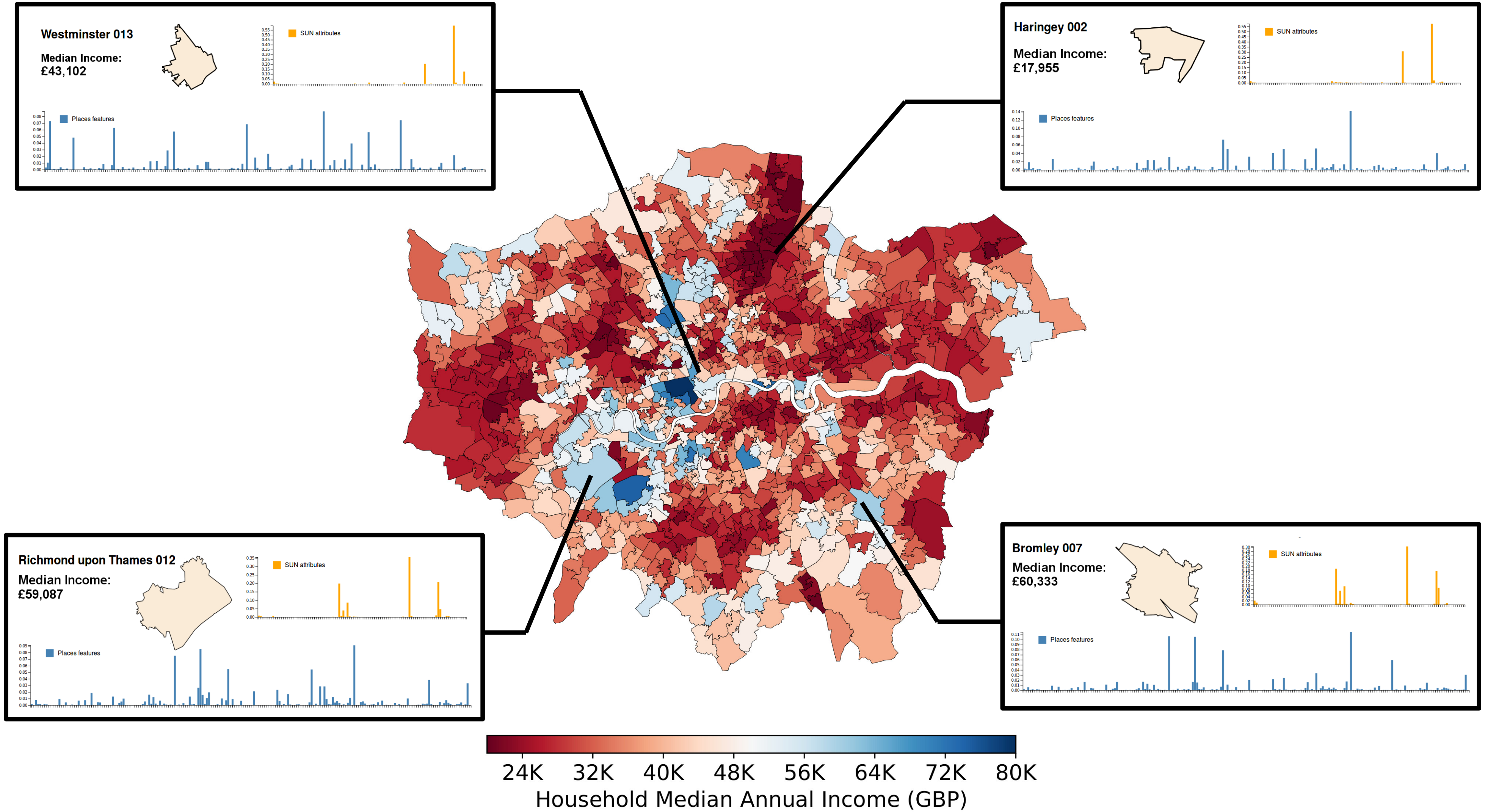
2 How does it relate to income?

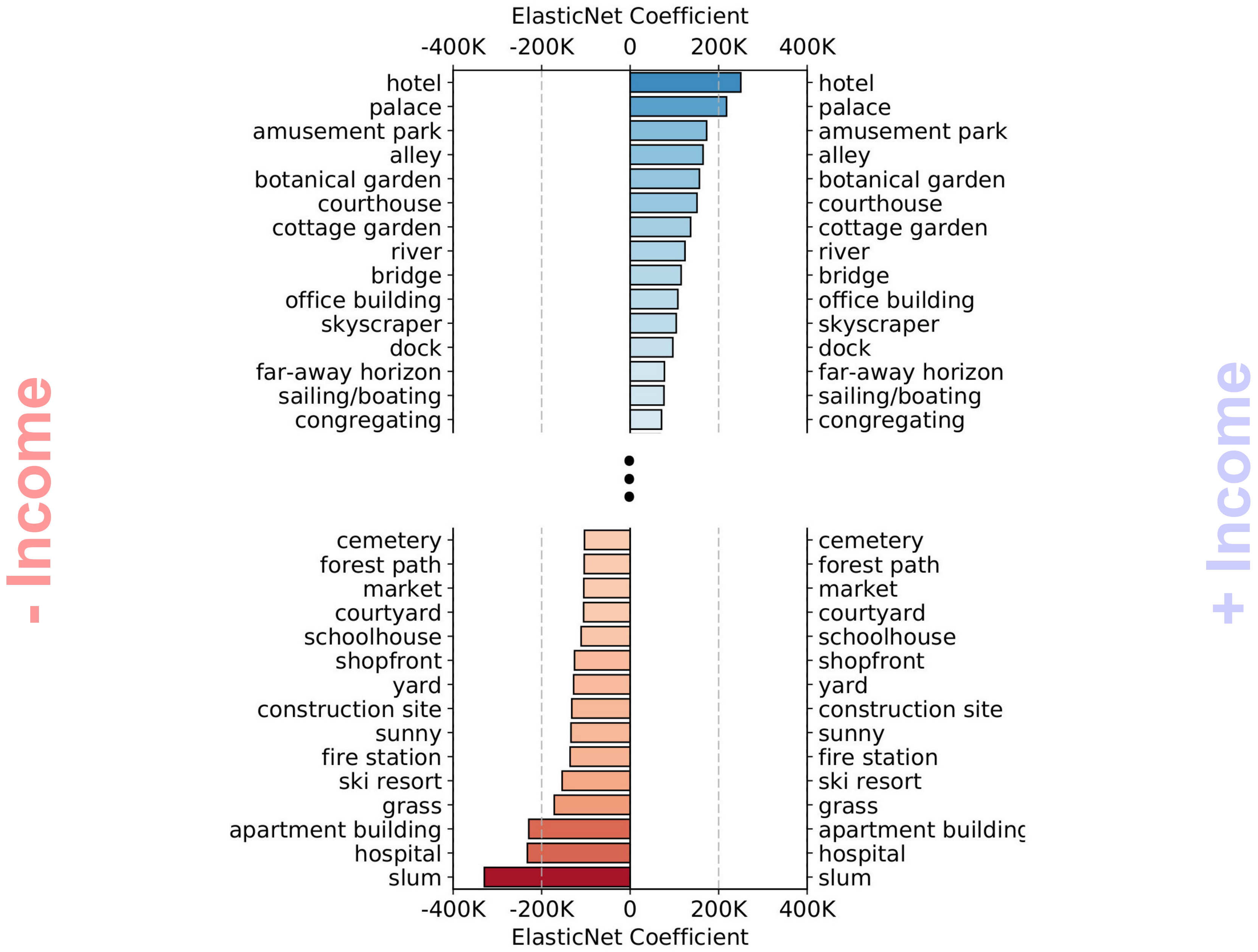
(Training a machine learning algorithm)

Income in London

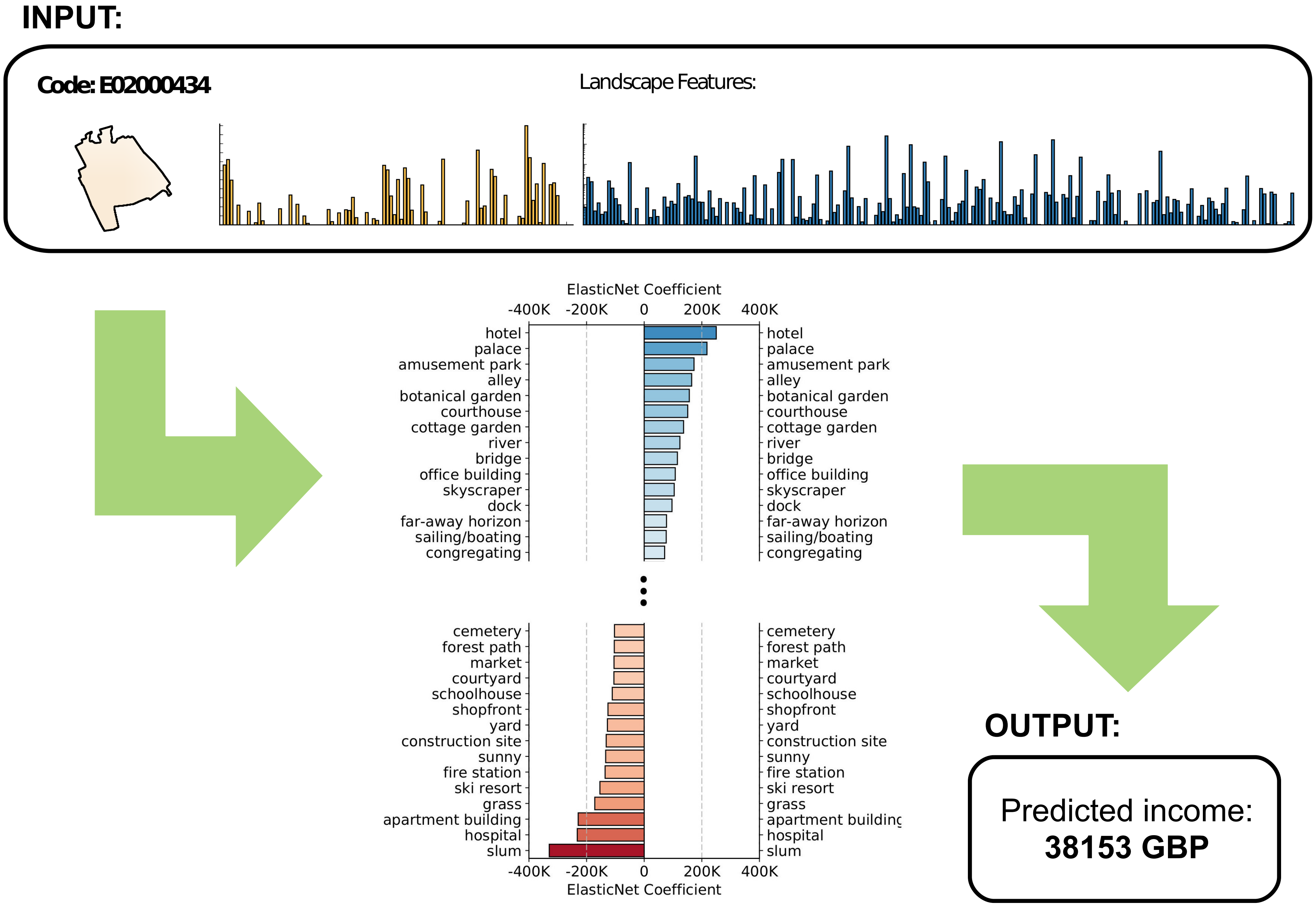


Training an elastic net



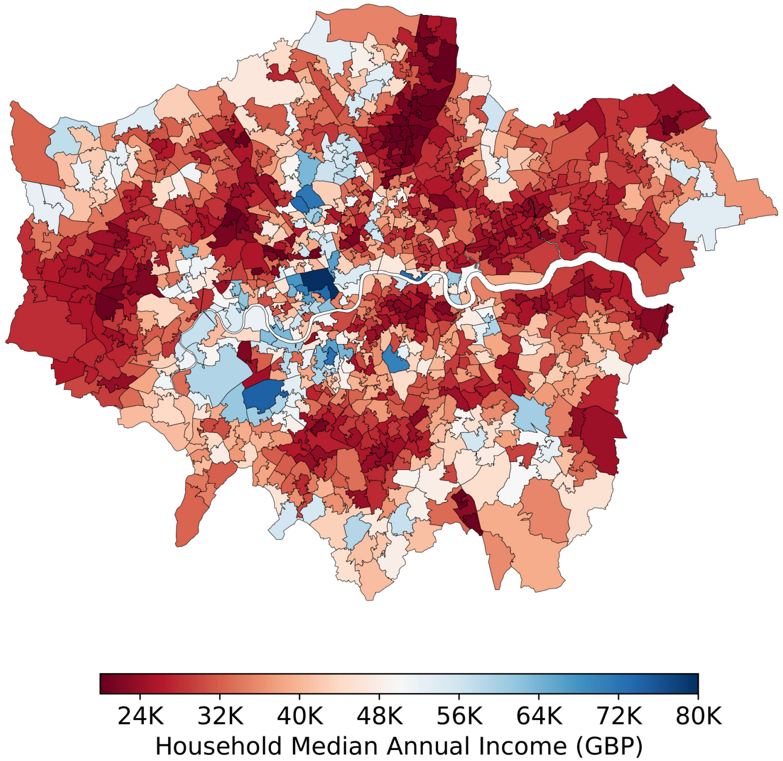


- Income
+ Income
ElasticNet Coefficient
-400K
-200K
0
200K
400K
hotel
palace
amusement park
alley
botanical garden
courthouse
cottage garden
river
bridge
office building
skyscraper
dock
far-away horizon
sailing/boating
congregating
cemetery
forest path
market
courtyard
schoolhouse
shopfront
yard
construction site
sunny
fire station
ski resort
grass
apartment building
hospital
slum

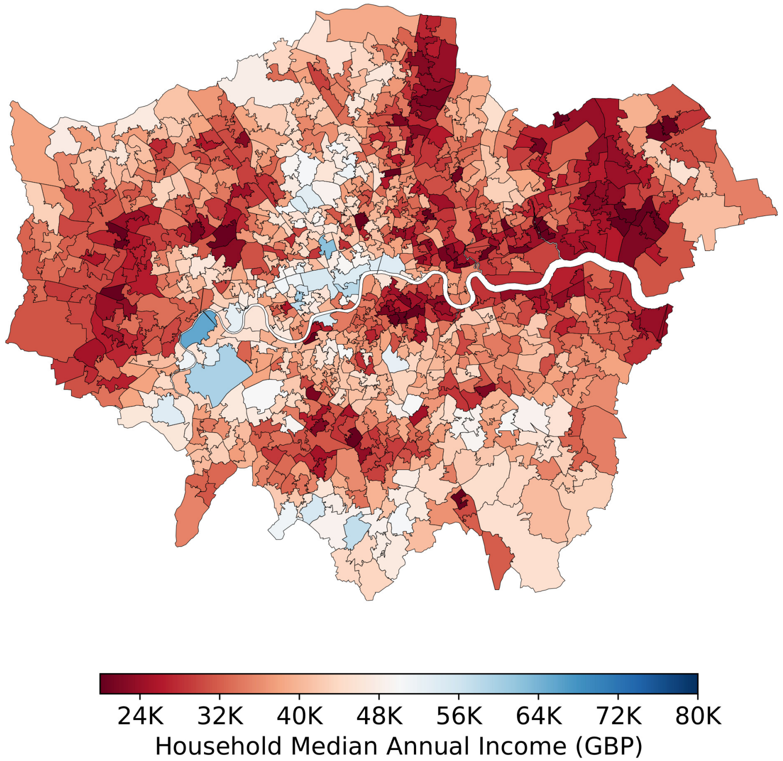

INPUT:
Code: E02000434
Landscape Features:
ElasticNet Coefficient
-400K
-200K
0
200K
400K
hotel
palace
amusement park
alley
botanical garden
courthouse
cottage garden
river
bridge
office building
skyscraper
dock
far-away horizon
sailing/boating
congregating
cemetery
forest path
market
courtyard
schoolhouse
shopfront
yard
construction site
sunny
fire station
ski resort
grass
apartment building
hospital
slum
OUTPUT:
Predicted income:
38153 GBP

London

Observed

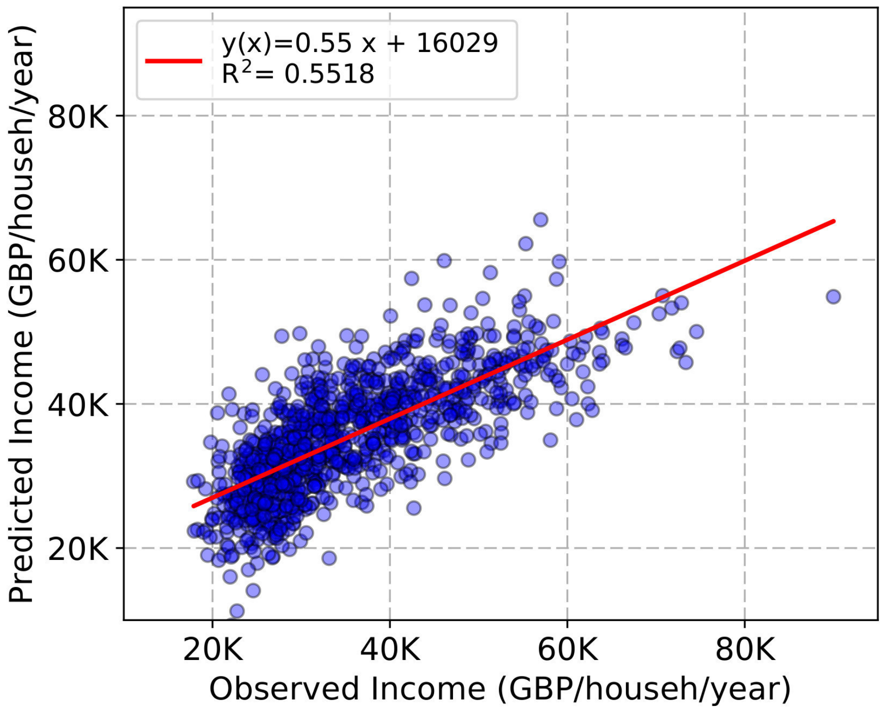


Predicted

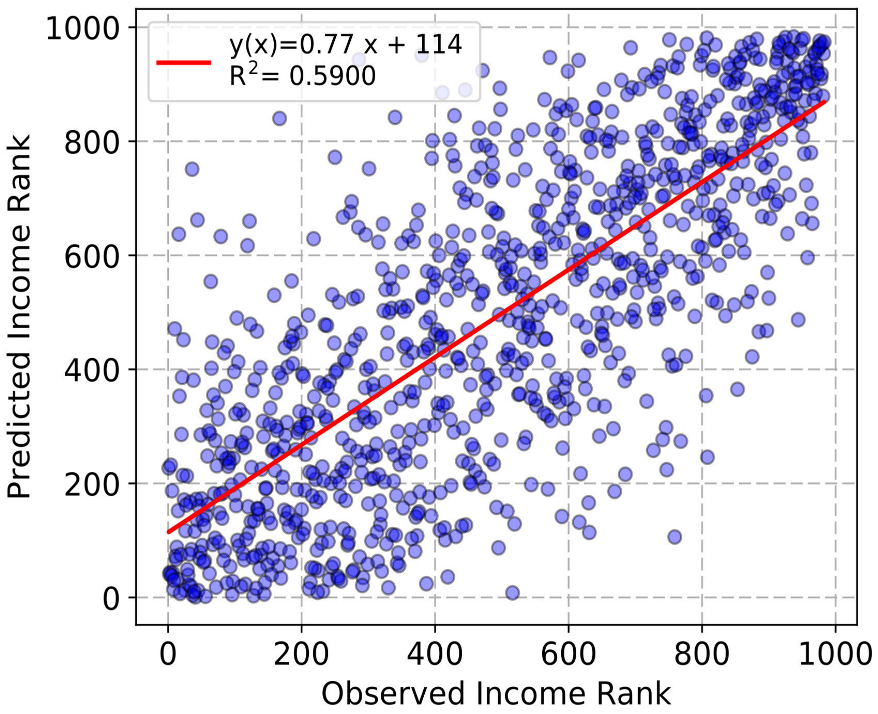


Prediction Accuracy

Income Value



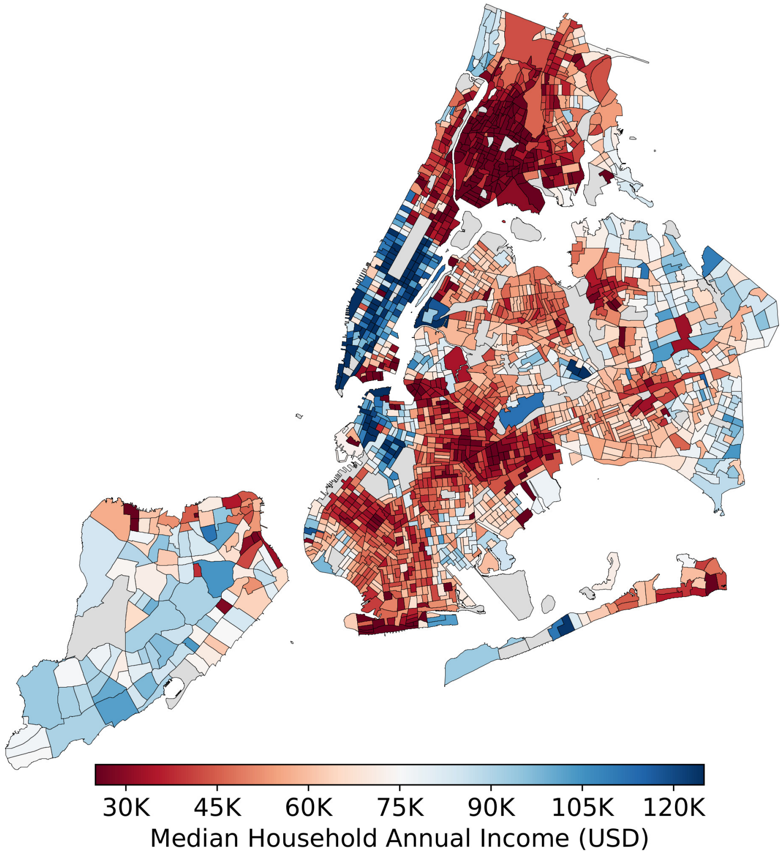
Income Rank



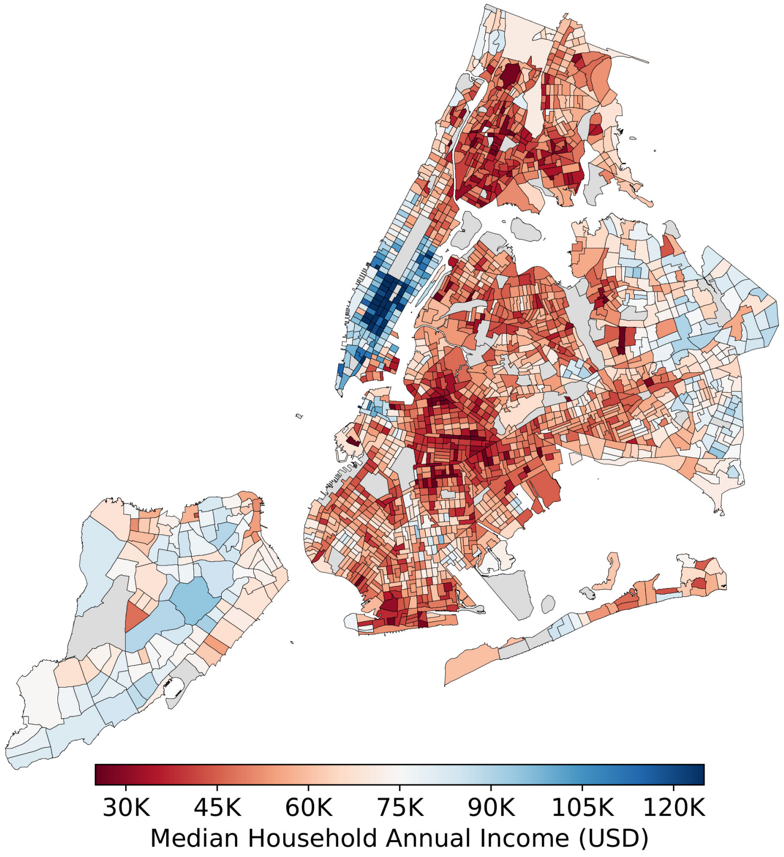

London
New York
Hong Kong
Rio de Janeiro
Johannesburg
Sydney
Greenland
Iceland
Finland
Sweden
Russia
Canada
Poland
Germany
Ukraine
France
Italy
Spain
Kazakhstan
Mongolia
North Pacific Ocean
United States
North Atlantic Ocean
Turkey
Japan
Iraq
Iran
Afghanistan
Pakistan
Algeria
Libya
Egypt
Saudi Arabia
Mexico
India
Thailand
Mali
Niger
Sudan
Chad
Nigeria
Ethiopia
Venezuela
Colombia
Kenya
DR Congo
Tanzania
Indonesia
Papua New Guinea
Peru
Bolivia
Namibia
Madagascar
Indian Ocean
Australia
South Pacific Ocean
Chile
South Atlantic Ocean
South Africa
Argentina
New Zealand
Southern Ocean
Southern

New York

Observed

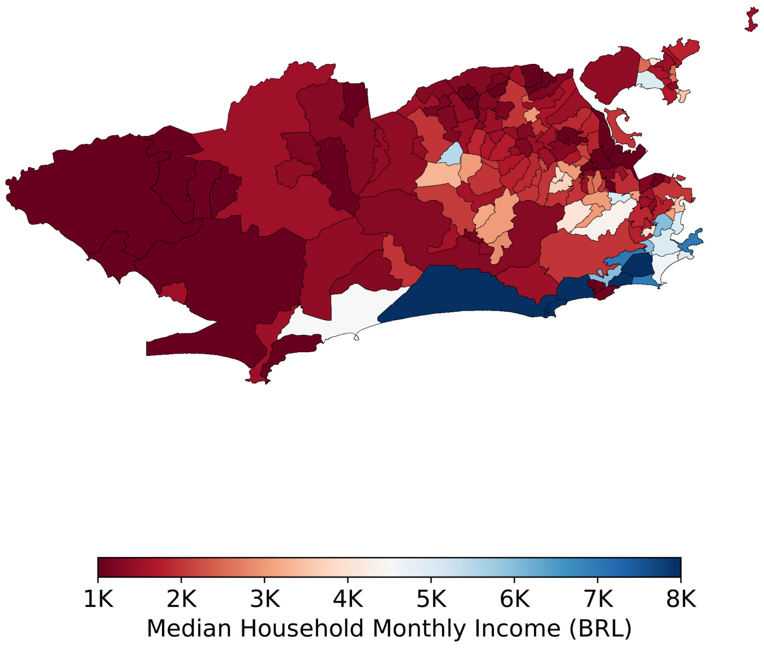


Predicted

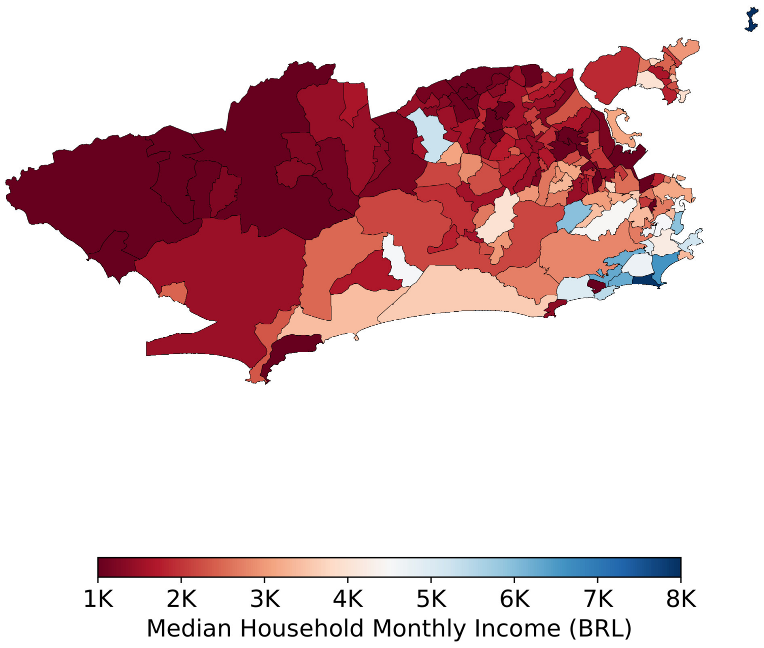


Rio de Janeiro

Observed

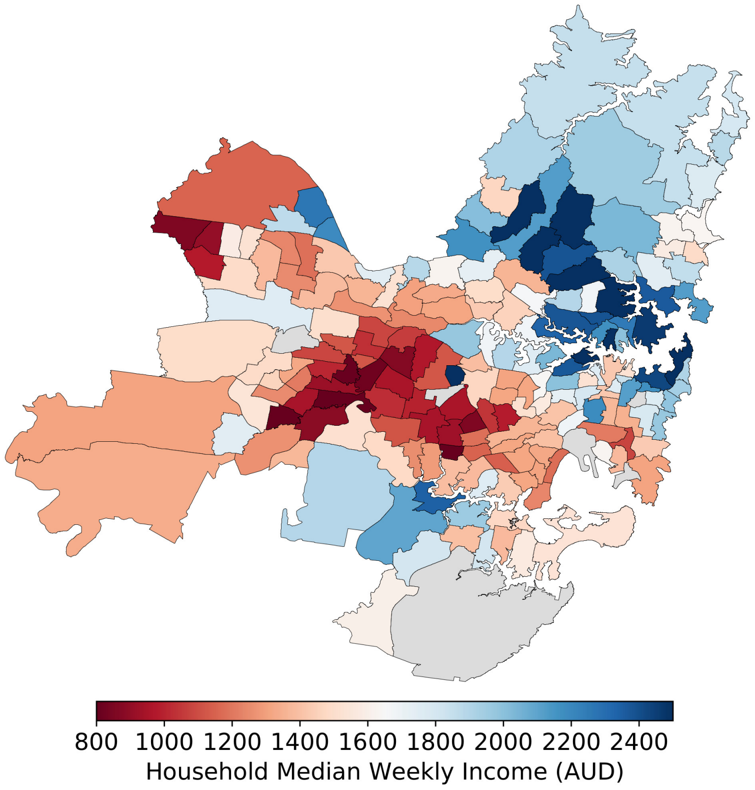


Predicted

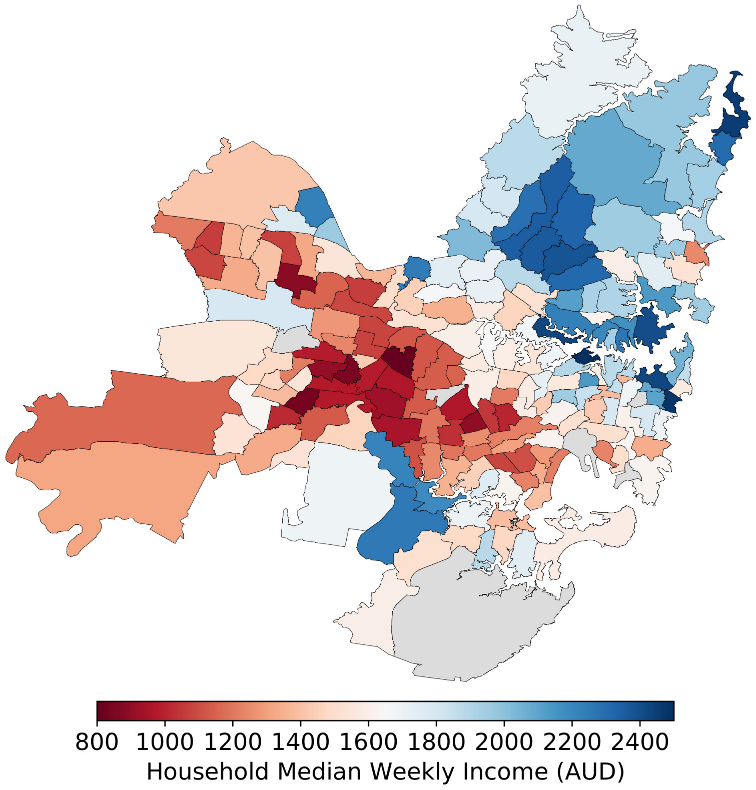


Sydney

Observed

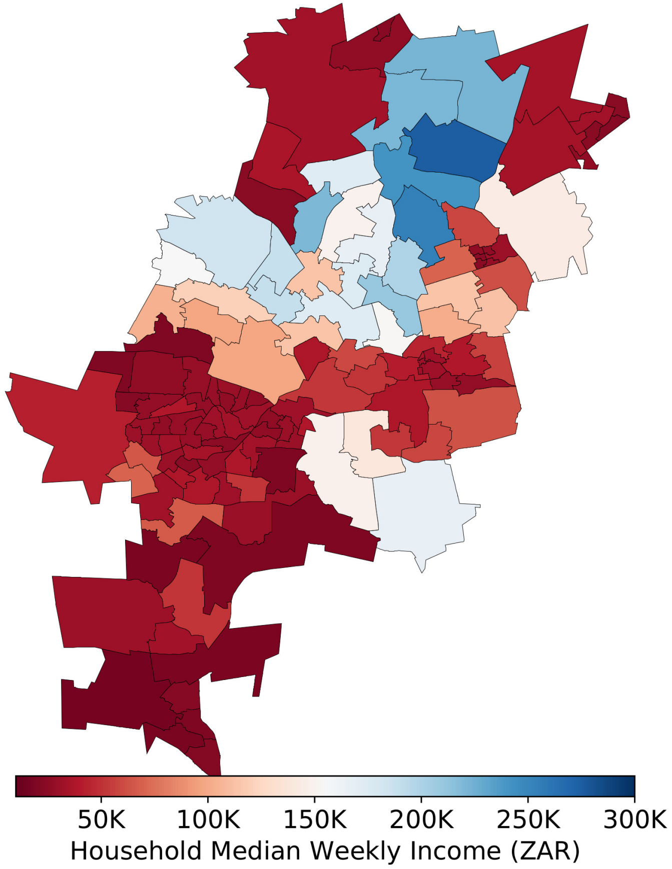


Predicted

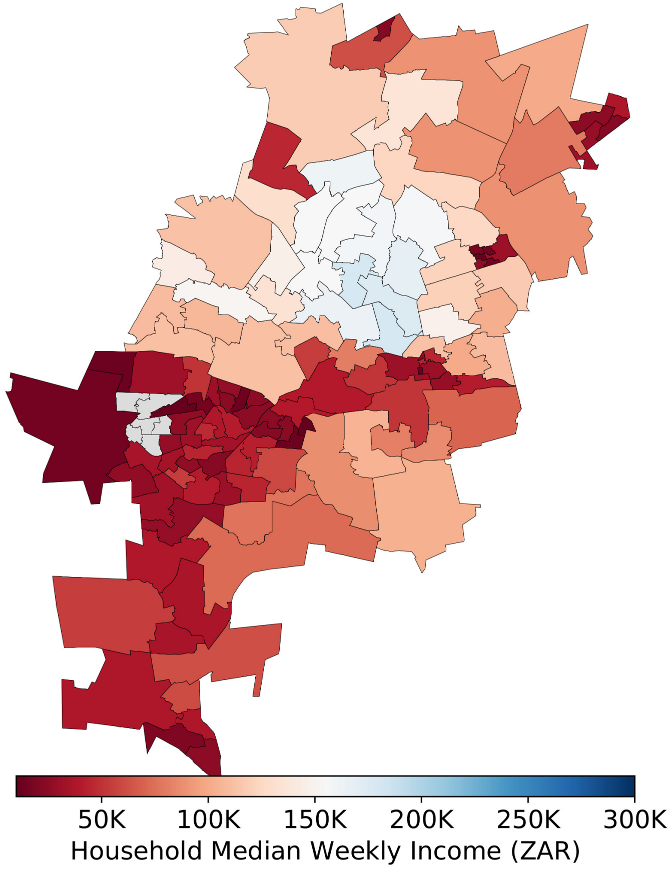


Johannesburg

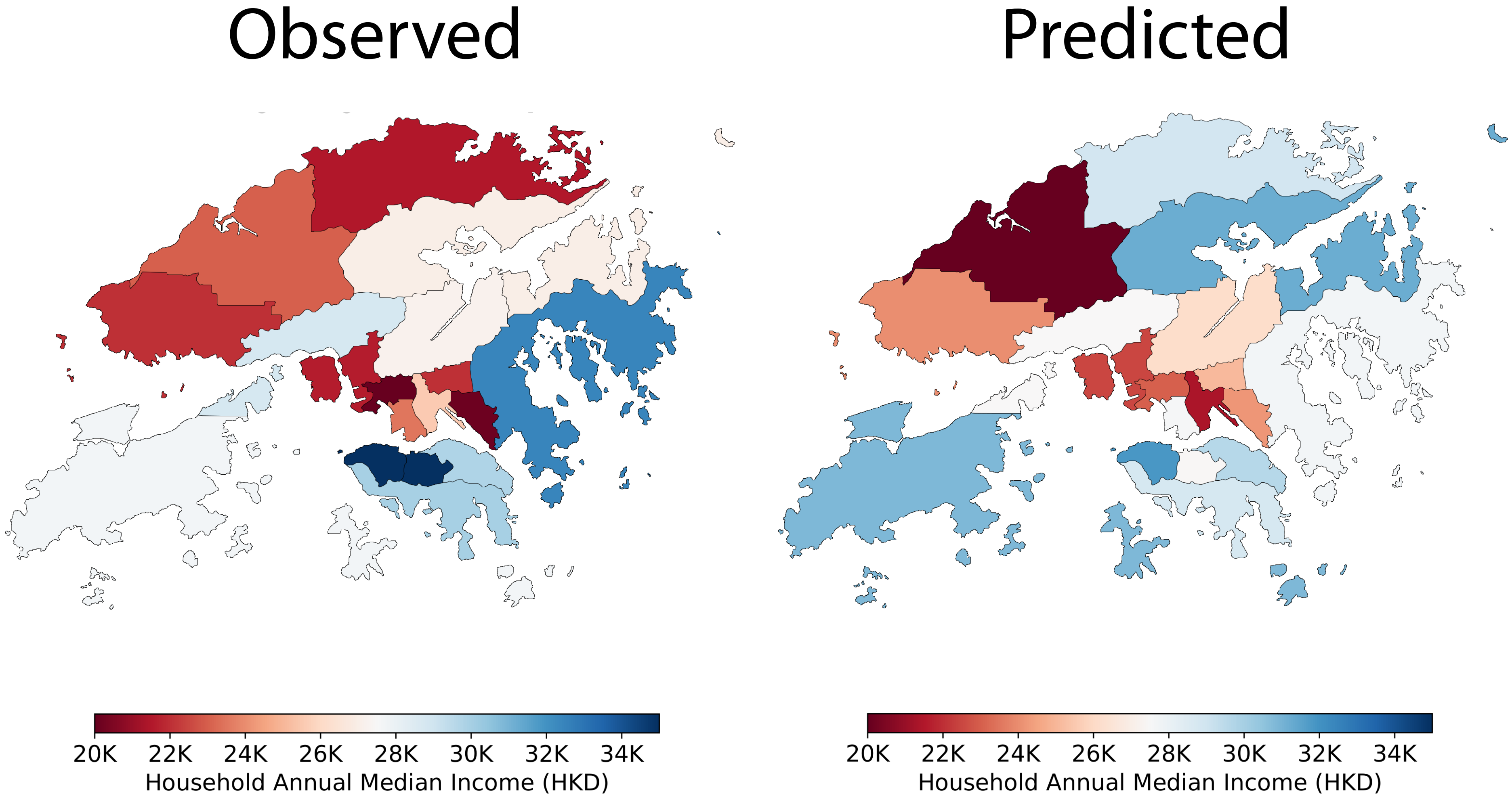
Observed

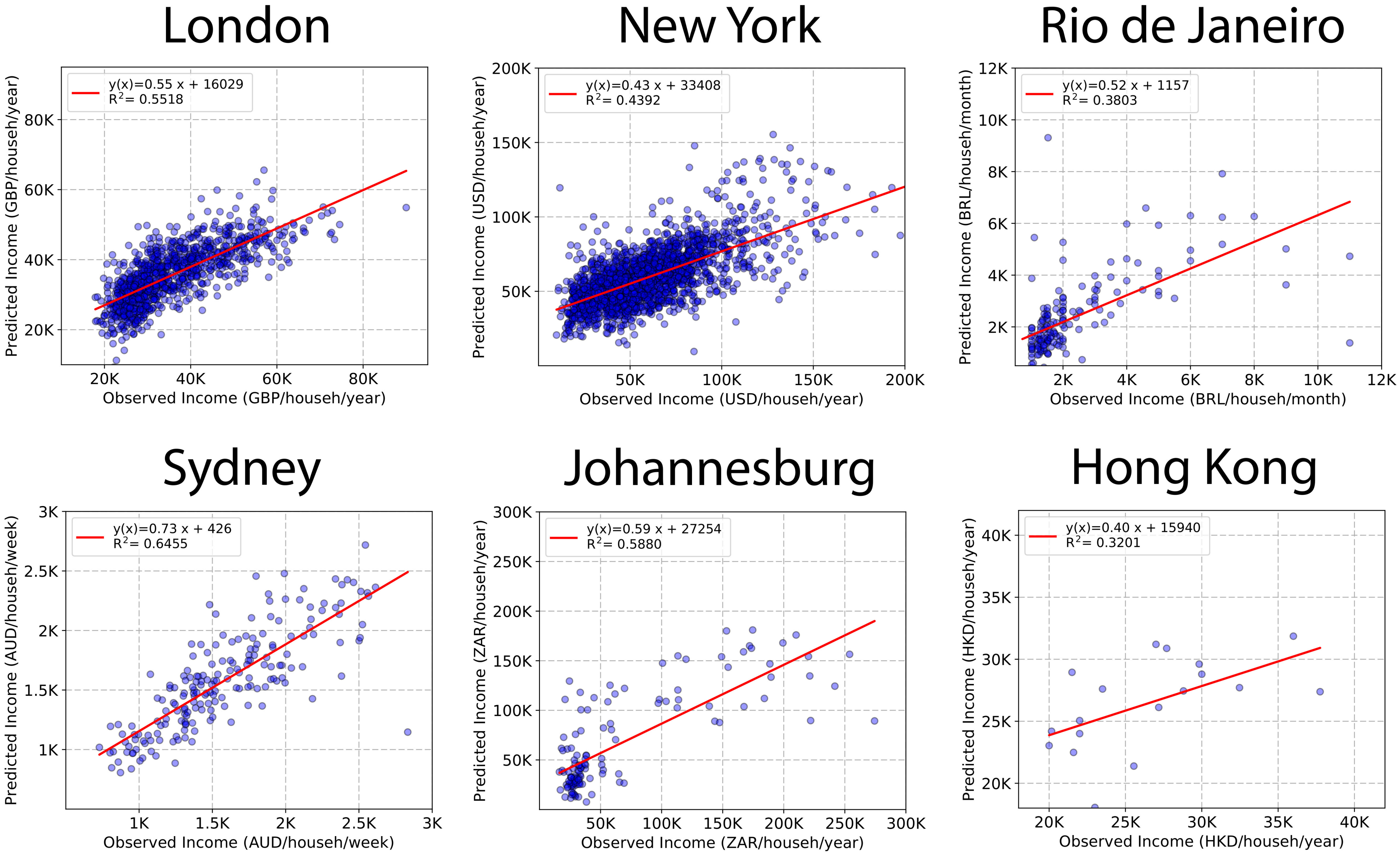


Predicted



Hong Kong


Observed
Predicted
20K 22K 24K 26K 28K 30K 32K 34K
Household Annual Median Income (HKD)
20K 22K 24K 26K 28K 30K 32K 34K
Household Annual Median Income (HKD)


London
y(x)=0.55 x + 16029
R^2= 0.5518
Predicted Income (GBP/househ/year)
Observed Income (GBP/househ/year)
New York
y(x)=0.43 x + 33408
R^2= 0.4392
Predicted Income (USD/househ/year)
Observed Income (USD/househ/year)
Rio de Janeiro
y(x)=0.52 x + 1157
R^2= 0.3803
Predicted Income (BRL/househ/month)
Observed Income (BRL/househ/month)
Sydney
y(x)=0.73 x + 426
R^2= 0.6455
Predicted Income (AUD/househ/week)
Observed Income (AUD/househ/week)
Johannesburg
y(x)=0.59 x + 27254
R^2= 0.5880
Predicted Income (ZAR/househ/year)
Observed Income (ZAR/househ/year)
Hong Kong
y(x)=0.40 x + 15940
R^2= 0.3201
Predicted Income (HKD/househ/year)
Observed Income (HKD/househ/year)


Urban visual landscape
can be used to
estimate income

Data from online photographs may help us better measure and understand human behaviour and wellbeing

wbs
WARWICK BUSINESS SCHOOL

Data Science Lab

The Alan Turing Institute

EPSRC
Engineering and Physical Sciences Research Council

Digital Economy
Transforming Business and Society

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL