Is there an association between art and changes in the economic conditions of urban neighbourhoods? While the popular media and policymakers commonly believe this to be the case, quantitative evidence remains lacking. Here, we use metadata of geotagged photographs uploaded to the popular image-sharing platform Flickr to quantify the presence of art in London neighbourhoods. We estimate the presence of art in neighbourhoods by determining the proportion of Flickr photographs which have the word ‘art’ attached. We compare this with the relative gain in residential property prices for each Inner London neighbourhood. We find that neighbourhoods which have a higher proportion of ‘art’ photographs also have greater relative gains in property prices. Our findings demonstrate how online data can be used to quantify aspects of the visual environment at scale and reveal new connections between the visual environment and crucial socio-economic measurements.

1. Background

The story of art playing a central role in the transformation of deprived urban neighbourhoods is a compelling one that has greatly engaged popular media and policymakers alike. Researchers of urban policy maintain that the creative industries are key to building a successful city economy [1,2]. Governments around the world make significant investments in the arts in order to incentivize gentrification and regeneration in specific neighbourhoods [3–5].

However, quantitative evidence linking the presence of art with changing economic conditions of urban neighbourhoods is lacking. Several studies focus on specific neighbourhoods that fit this narrative, rather than carrying out broader investigations.
to determine whether this pattern holds in general [6–10]. Other studies use broad definitions to examine
the role that artists or creative people play in the economic improvement of urban neighbourhoods,
thereby including groups of people with differing agendas [11]. For example, studies often include
professions such as engineering, mathematics and computer science [2,12], or categories such as
‘Architectural Services’ and ‘Dance Companies’ [13]. Other research attempting to use art galleries and
arts organizations to quantify arts-led redevelopment in urban neighbourhoods suggests that once art
institutions have moved into a neighbourhood, redevelopment is already taking place [14,15] or a high
level of income is already present [16].

A key challenge for studies to date has been the identification of a reliable metric that quantifies
the presence of art in an area. Traditional studies have relied upon categories of creative occupations
[2,12,13] or the presence of art institutions [14–16], which are potentially constraining measures that may
not accurately capture the level of art present in a particular area.

Online activity could help us trace the presence of art in neighbourhoods in a less restrictive fashion.
Today, we have access to a new source of information on human behaviour: patterns of activity on
platforms such as Google, Flickr, Twitter and Wikipedia, which have already led to several insights into real-
world human behaviour [17–32]. Specifically, such studies demonstrate that we can use these new data
sources to quantify aspects of human behaviour that have previously been too costly, time-consuming or
awkward to measure. For example, Seresinhe, Preis and Moat found a link between scenic environments
and people’s wellbeing [30], by analysing novel crowdsourced data garnered through an online game
called Scenic-or-Not, alongside data from the Census for England and Wales. Using this online game,
over 1.5 million ratings of images covering 95% of Great Britain at a granularity of 1 km² have already
been collected.

Here, we propose to create an indicator of the presence of art by exploiting photographs uploaded to
the popular image sharing website Flickr. We focus on London as our area of study, as artists have long
thought to be changing the economic landscape of London—from James Whistler and Oscar Wilde in
Chelsea, to the 1990s ‘Britart’ movement in Hoxton [33].

2. Material and methods

We draw on images uploaded to Flickr between 2004, the year Flickr was released to the public, and 2013.
As we are primarily interested in observing the presence of art in urban areas, we only include images
geotagged as being located in Inner London.

In order to identify and separate areas within Inner London for analysis, we use the system of
postcodes maintained by the Royal Mail. Postcodes are alphanumeric references comprising an ‘outward
code’ of two to four characters and an ‘inward code’ of three characters (e.g. CV4 7AL). Postcodes
beginning with NW, N, W, WC, E, EC, SW and SE are deemed as being located in Inner London. As
the inward code is specific to a single section of a street, we only use the outward code of the postcode
(e.g. ‘EC1’)—the postcode district—to locate neighbourhoods in Inner London. To ensure consistency
between the size of postcode areas, in the EC and WC areas, we aggregate the postcode districts by
ignoring the further subdivision created by the additional letter in the outward code. For example, the
postcode ‘EC1A’ and ‘EC1M’ are considered part of the area ‘EC1’, and data for these postcodes are
therefore, analysed together.

When photographers upload images to Flickr, they can include metadata such as a title, description
and tags (e.g. ‘art’, ‘sky’, ‘modern’). We deem a photograph to be an ‘art’-related image if there is a
mention of ‘art’ in this textual metadata associated with Flickr photographs for each Inner London
postcode. In this analysis, we restrict ourselves to classifying images using textual data only. Future
analysis could consider image recognition techniques to classify images as ‘art’-related [34], or could
work with citizen scientists to determine whether images contain art, following the example of such
initiatives as Galaxy Zoo where volunteers classify images of galaxies [35].

In order to create an indicator of ‘art’-related images, we search all the textual elements of each image,
including the title, description and tags, using the regular expression ‘\<art\>’, which ensures that only
the whole word ‘art’ is found. We note that in regular expressions in R, the symbols ‘\<’ and ‘\>’ match
the empty string at the beginning and end of a word. While this process ensures we capture the word
‘art’ and no other words that contain ‘art’ (such as ‘apart’), this method does have a minor drawback in
that words such as ‘arty’ or art-related compound words that do not have spaces (such as ‘streetart’) are
not discovered. We count only a single occurrence of ‘art’ for each image, even if it has ‘art’ mentioned
several times in the metadata.
Figure 1. Proportion of ‘art’ photographs uploaded to Flickr from 2004 to 2013 in Inner London postcode areas. We deem a photograph as ‘art’ related if there is a mention of ‘art’ in the photograph’s textual metadata. As tourist areas in London will attract more images on Flickr than other areas, we divide the number of ‘art’ images by the total number of images for each postcode for the whole time period (2004–2013), in order to calculate the proportion of ‘art’ photographs per Inner London neighbourhood. We define neighbourhoods on the basis of postcodes, as explained in detail in the main text. With this measure, we find that the proportion of ‘art’ related photographs is highest in East and Southeast London, with some heightened activity in Southwest and North London. All photos from Flickr are licensed for reuse under the Creative Commons Attribution-Non-Commercial-ShareAlike 2.0 Generic Licence (https://creativecommons.org/licenses/by-nc-sa/2.0/). Postal Boundaries © GeoLytix copyright and database right 2012. Contains Ordnance Survey data © Crown copyright and database right 2012. Contains Royal Mail data © Royal Mail copyright and database right 2012. Contains National Statistics data © Crown copyright and database right 2012.

Certain areas in London will attract more images on Flickr than others (e.g. tourists generate many images of notable locations such as Big Ben and the London Eye). Therefore, in order to meaningfully compare one area with another, we divide the number of ‘art’ images by the total number of images for each postcode for the whole time period (2004–2013), in order to create a normalized indicator of ‘art’ per area.

We underline that with this approach, images that are in reality ‘art’ images but do not have the word ‘art’ in their textual metadata will not be identified. In addition, images which have the word ‘art’ in their metadata but do not in reality contain any art will count towards our indicator. Indeed, our indicator is susceptible to any outside influence which may cause photographers to upload a higher proportion of photographs with the word ‘art’ in the textual metadata—for example, the presence of more photographers who are interested in art, rather than an increased presence of art.

With these limitations in mind, our indicator allows us to estimate which neighbourhoods in London may have relatively more ‘art’, or photographers interested in art, over this time period. Visual inspection...
of the proportion of ‘art’ images in Inner London postcodes reveals that art is most prevalent in East and Southeast London, with some sparse heightened activity in Southwest, North and East Central (figure 1).

Next, as one indicator of the economic transformation of Inner London neighbourhoods, we use a measure of the relative change in mean residential property prices. We track residential property prices in Inner London using data from residential property sales registered at the Land Registry from 2004 to 2013 (https://www.gov.uk/government/collections/price-paid-data). To understand the relative increase in house price in comparison with other areas, we first calculate the mean residential property price per Inner London postcode. (Note that as residential property prices have only been recorded since 2013 for the newly created postcode ‘E20’, this area has not been included in the analysis.) We then rank the postcodes according to mean residential property price, where a rank of 1 indicates the highest mean residential property price. Finally, we calculate the relative changes in property price by determining the difference between the 2013 rank and the 2004 rank for each Inner London postcode. Thus, negative changes in rank signify an area becoming relatively more expensive, and positive changes in rank signify an area becoming relatively less expensive.

We note that other measures of change in house price are strongly correlated with the average house prices in an area. For example, areas with the highest mean house price are those with the highest change in mean house price. This is not true for change in rank mean house price. An alternative approach to resolving this issue would be to normalize changes in mean house price, by dividing the change by the original mean house price for an area.

3. Results

Visual inspection of the change in rank of mean residential property prices reveals that areas in central East and Southeast London have become relatively more expensive while areas in outer East London and North London have become relatively less expensive (figure 2a). A comparison of the relative change of mean residential property price and the proportion of ‘art’ images suggests that the higher the proportion of ‘art’ images, the greater the relative gain in house price, as measured by the change in rank of mean residential property prices (figure 2b). A Kendall rank correlation test provides further evidence of this relationship ($r = -0.23$, $p < 0.001$, $n = 119$, Kendall’s rank correlation). This is initial evidence in support of our hypothesis that the presence of art is associated with improvements in economic conditions of urban neighbourhoods.

However, in general, measurements of social and physical characteristics can be clustered in geographical space, and thus spatially dependent, where nearby observations may have similar values simply due to their proximity [36,37]. We test whether such spatial clustering exists by first building a linear regression model to predict relative change in mean residential property price from the proportion of ‘art’ images. We determine which postcode areas are neighbours based on whether they share any boundary points, using boundary-line data for Great Britain from the Ordinance Survey (https://www.ordnancesurvey.co.uk/opendatadownload/products.html). Neighbours are defined in a binary fashion: either two areas are neighbours, or they are not. Using this neighbour structure, we confirm that our model contains measurements clustered in space by running a Moran’s I test on the residuals of the linear regression model (Moran’s $I = 0.281$, $p < 0.001$, $n = 119$).

The presence of clustering in the data means that we cannot guarantee that our observations are independent, a key requirement of linear regression. In order to address the problem of spatial autocorrelation in the residuals of our basic linear regression model, we build a conditional auto regressive (CAR) model. The CAR model is a linear regression model which additionally captures the spatial dependency between the error term for each area and the error terms for the neighbouring areas. Further details of this model can be found in the electronic supplementary material.

This CAR model is used to predict relative change in mean residential property price using the proportion of ‘art’ images. We find that areas with higher proportions of ‘art’ images also have greater relative gains in house price, as measured by the change in rank of mean residential property prices ($\beta = -90.92$, $p = 0.040$, $n = 119$).

We note that a small number of areas, such as Shoreditch (E2) and Dalston (E9) which are commonly thought to be associated with arts-led redevelopment, exhibit particularly high proportions of ‘art’-related photographs, as well as high relative gains in house price (figure 2b). We also find that the residuals from our model are not normally distributed and exhibit a slight skew. To ensure that the presence of this handful of cases and the distributional properties of our residuals are not driving our
results, we run a bootstrap analysis on our CAR model [38], in which we compare the original regression coefficient with a bootstrapped distribution of regression coefficients.

To carry out this bootstrap analysis, we resample the residuals of our CAR model 10,000 times. We use these bootstrapped residuals to generate bootstrapped changes in rank of mean residential property prices. Finally, we build CAR models to regress the bootstrapped changes in property price on the
proportions of ‘art’ photographs in each area, which remain the same as in the original CAR model, in order to create 10,000 bootstrapped regression coefficients. A comparison of the original regression coefficient with the bootstrapped distribution of regression coefficients provides further evidence that areas which have a higher proportion of art images have greater relative gains in house price, as measured by the change in rank of mean residential property prices (mean bootstrapped $\beta = -116.74$, 95% confidence interval $[-209.85, -32.19]$, $p = 0.044$, 10,000 bootstrap samples).

4. Conclusion

We consider whether geotagged photographs uploaded to Flickr may help us quantify the link between art and the changing economic conditions of urban neighbourhoods. We quantify art in a neighbourhood using the proportion of ‘art’ photographs uploaded to Flickr. We compare our art indicator with the relative change in mean residential property price for each Inner London neighbourhood. Once we take clustering of spatial measurements into account, relative increases in mean residential property prices are significantly associated with higher proportions of ‘art’ images per neighbourhood.

The nature, as well as the direction, of any causal link requires more investigation. Does the presence of art in an area remove a negative stigma attached to deprived neighbourhoods [9]? Do artists improve the aesthetics of deprived neighbourhoods through the creation of public artworks such as iconic sculptures and street murals [39–41]? Or, as people connected to the arts rely on socializing to advance their careers, are more ‘arty’ neighbourhoods attracting more cafes and restaurants, which in turn attract other groups of people to move into the neighbourhood [1,10]? We highlight once more that it is also possible that ‘self-selection’ may be influencing our measure of art in a neighbourhood. As property prices increase, such neighbourhoods might attract more affluent people with interests in art, who might then be more likely to search for art in their neighbourhood which they can photograph and display on Flickr.

We also underline that the study we present here considers only one aspect of urban economic change. Future studies may wish to extend this analysis to other related socio-economic variables. For example, as areas regenerate, artists themselves may be displaced [8,42,43]. Studies quantifying the presence of art over time alongside changing socio-economic factors could be vital in better understanding this process. Finally, the link between the presence of art and increasing property prices might be due to a combination of factors. It has previously been argued that in neighbourhoods where a restricted supply of housing is not pressuring property prices, artist-led economic development may not take place [11,44]. Caution should, therefore, be taken in singling out art alone as the driving force behind improving economic conditions of urban neighbourhoods.

Our results suggest that art is indeed associated with improving economic conditions of urban neighbourhoods. More generally, our analysis demonstrates that data on our online interactions can provide novel measurements of the environment in which we live. These measurements reveal that the visual environment may affect aspects of our life as crucial as economic development.