INVISIBLE CITIES: EXPLORING PSYCHOLOGICAL URBAN DATA

JOÃO PAULO BARROS COTTA PESCE

INVISIBLE CITIES: EXPLORING PSYCHOLOGICAL URBAN DATA

Dissertation presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Virgílio Almeida
Co-Advisor: Daniele Quercia

Belo Horizonte
February 2014

Pesce, João Paulo Barros Cotta

P473i Invisible Cities: Exploring Psychological Urban Data / João Paulo Barros Cotta Pesce. — Belo Horizonte, 2014

xxii, 47 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais

Orientador: Virgílio Almeida

1. Computação – Teses. 2. Mídias sociais. 3. Web science. 4. Análise espacial (Estatística). 5. Informática urbana. 6. Cidades do futuro. I. Orientador. II. Título.

CDU 519.6*06(043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Invisible cities: exploring psychological urban data

JOÃO PAULO BARROS COTTA PESCE

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. VIRGÍLIO AUGUSTO FERNANDES ALMEIDA - Orientador
Departamento de Ciência da Computação - UFMG

PROF. CLODOVEU AUGUSTO DAVIS JÚNIOR
Departamento de Ciência da Computação - UFMG

DR. DANIELE QUERCIA
YAHOO LABS

PROF. WAGNER MEIRA JÚNIOR
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 14 de novembro de 2014.

Aos meus pais, Lairson e Tina e aos meus irmãos Pedro e Lucas.

Acknowledgments

Primeiramente, gostaria de agradecer meu orientador, Professor Virgílio. Seu conhecimento e compreensão foram inestimáveis para o meu desenvolvimento acadêmico e pessoal. Tenho muito orgulho de ter sido seu aluno e agradeço por despertar meu amor pela pesquisa.

Ao meu co-orientador Daniele Quercia, pela oportunidade de trabalho em conjunto, intercambio cultural, inestimáveis contribuições e ensinamentos.

Ao amigo e pesquisador Diego Las Casas pela parceria. Sem você, nada disto existiria.

Aos amigos dos laboratórios CAMPS e LSL, por alegrar meus dias.

À minha família, pelo amor, incentivo e apoio incondicional.

"Truth is ever to be found in simplicity,
and not in the multiplicity and confusion of things."
(Isaac Newton)

Abstract

Planners and social psychologists have suggested that the recognizability of the urban environment is linked to people's socio-economic well-being. We build a web game that puts the recognizability of London's streets to the test. It follows as closely as possible one experiment done by Stanley Milgram in 1972. Each participant dedicates only few minutes to the task (as opposed to 90 minutes in Milgram's). We collect data from 2,255 participants (one order of magnitude a larger sample) and build a recognizability map of London based on their responses. We find that some boroughs have little cognitive representation; that recognizability of an area is explained partly by its exposure to Flickr and Foursquare users and mostly by its exposure to subway passengers; and that areas with low recognizability do not fare any worse on the *economic* indicators of income, education, and employment, but they do significantly suffer from *social* problems of housing deprivation, poor living conditions, and crime. These results could not have been produced without analyzing life off- and online: that is, without considering the interactions between urban places in the physical world and their virtual presence on platforms such as Flickr and Foursquare.

We then use the results of this experiment, along with other urban data, to tackle the problem of identifying interesting and memorable pictures in photo sharing sites. Past proposals for identifying such pictures have relied on either metadata (e.g., likes) or visual features. In practice, techniques based on those two inputs do not always work: metadata is sparse (only few pictures have considerable number of likes), and extracting visual features is computationally expensive. In mobile solutions, geo-referenced content becomes increasingly important. The premise behind this work is that pictures of a neighborhood is linked to the way the neighborhood is perceived by people: whether it is, for instance, distinctive and beautiful or not. Since 1970s, urban theories proposed by Lynch, Milgram and Peterson aimed at systematically capturing the way people perceive neighborhoods. Here we tested whether those theories could be put to use for automatically identifying appealing city pictures. We did so by gathering geo-referenced Flickr pictures in the city of London; selecting six urban qualities associated with those urban theories; computing proxies for those qualities from online social media data; and ranking Flickr pictures based on those proxies. We find that our proposal enjoys three main desirable properties: it is *effective*, *scalable*, and aware of *contextual* changes such as time of day and weather condition. All this suggests new promising research directions for multi-modal learning approaches that automatically identify appealing city pictures.

Palavras-chave: Social Media, Web Science, Spatial Analysis, Urban Informatics, Future Cities.

Resumo

Urbanistas e psicólogos sociais sugeriram que a facilidade de reconhecimento de um certo ambiente urbano está ligado ao bem estar social-econômico de seus habitantes. Nós criamos um jogo *online* que testa o reconhecimento das regiões de Londres. Ele segue, da forma mais próxima possível, um experimento conduzido por Stanley Milgram em 1972. Cada participante dedica apenas alguns minutos nessa tarefa (contrastando com os 90 minutos do experimento de Milgram). Nós coletamos resultados de 2.255 participantes (amostra com uma ordem de grandeza maior) e, baseando-se nas respostas, construímos o mapa de reconhecimento de Londres. Nós descobrimos que alguns bairros possuem baixa representação cognitiva; que o reconhecimento de uma área é explicada parcialmente pela sua exposição no Flickr e Foursquare e, mais fortemente, ao fluxo de pessoas no metrô; e que áreas com baixo reconhecimento não se mostram piores em indicadores *econômicos* como renda, educação e emprego, mas sofrem significativamente de problems *sociais* como privação habitacional, condições precárias de vida e crime. Esses resultados não seriam possíveis sem uma análise dos ambiente *online* e *offline*: isto é, sem considerar as interações de locais no mundo real e a sua presença virtual em plataformas como o Flickr e Foursquare.

Em seguida, nós usamos os resultados desse experimento, juntamente com outros dados urbanos, para contribuir para o problema de identificação de imagens interessantes e memoráveis em sites de compartilhamento de fotos. Até então, propostas para identificação dessas imagens basearam-se ou em metadados (*e.g.*, curtidas) ou características visuais. Na prática, técnicas baseadas nesses dois métodos não funcionam sempre: metadados são esparsos (apenas uma pequena porção de imagens tem um número considerável de curtidas), e a extração de características visuais é computacionalmente custoso. Em soluções móveis, conteúdos geo-referenciados têm se tornado cada vez mais importantes. A premissa por trás deste trabalho é que as fotos de uma região estão ligadas com o modo como essa região é percebida: por exemplo, se ela é vista como bonita e característica ou não. Desde os anos 70, teorias urbanas propostas por Lynch, Milgram e Peterson visam capturar sistematicamente o modo como as pessoas entendem suas cidades. Neste estudo, nós testamos se essas teorias poderiam ser usada para identificar automaticamente figuras atraentes em cidades. Para isso, coletamos fotos geo-referenciadas de Londres no Flickr; selecionamos seis qualidades urbanas associadas às teorias urbanas mencionadas anteriormente; computamos meios para prever essas qualidades a partir de dados de redes sociais; e, com base nisso, ordenamos as fotos do Flickr. Nós descobrimos que nossa proposta possui três propriedades desejadas nesse contexto: ela é *efetiva*, *escalável*, e sensível a mudanças de *contexto*, como hora do dia e condição climática. Esses resultados sugerem novas direções de pesquisa para abordagens de aprendizado multimodal que automaticamente identificam fotos urbanas relevantes em serviços de compartilhamento de imagens.

Palavras-chave: Mídias Sociais, Web Science, Análise Espacial, Informática Urbana, Cidades do Futuro.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

The share of the world's population living in cities has recently surpassed 50 percent. By 2025, we will see another 1.2 billion people living in cities. The world is in the midst of an immense population shift from rural areas to cities, not least because urbanization is powered by the potential for enormous economic benefits. Those benefits will be only realized, however, if we are able to manage the increased complexity that comes with larger cities. The 'smart city' agenda, in which this work falls into, is about the use of technological advances to collect and analyze urban data from different sources to create better cities.

In a general way, this research contributes to both the urban and computing research communities by combining early urban studies of how citizens perceive their cities with modern techniques of data science. In one side, the urban research community benefit from the advances in data collection and analysis and, on the other hand, the computer science community can explore new ways of applying psychological theories to otherwise typically computational problems.

This research is deeply rooted in early urban studies (Milgram et al. [1972]; Milgram and Jodelet [1976]; Lynch [1960]), but also taps into recent computing research, especially research on "games with a purpose", whereby one outsources certain activities (e.g., labeling images) to humans in an entertaining way (Von Ahn and Dabbish [2008]); research on large-scale urban dynamics (Crandall et al. [2009]; Cranshaw et al. [2012]; Noulas et al. [2012]); research on how location-based services affect people's behavior (Bentley et al. [2012]; Cramer et al. [2011]; Lindqvist et al. [2011]); and research on predicting popularity of pictures based on social features (van Zwol et al. [2010]).

1.1 Psychological Maps

The geographic map of a city consists of, among other entities, streets and buildings and reflects an *objective* representation of the city. By contrast, a psychological map (also known as mental map) is a *subjective* representation that each city dweller carries around in his/her head. Lynch [1960], in the seminal book *The Image of the City* states:

> "Most often our perception of the city is not sustained, but rather partial, fragmentary, mixed with other concerns. Nearly every sense is in operation, and the image is the composite of them all."

To understand the evolution of this perception, consider a tourist. Tourists who just arrived in a new city start with few reference points (*e.g.*, a hotel, the main street, a couple of tube stations) and then, as they expand the representation in their minds, they slowly begin to build a more thorough picture.

Over the years, cities have been built and maintained in a way that it is imaginable, *i.e.*, that mental maps of the city are clear and economical of mental effort. That is because studies have posited that good imaginability allows city dwellers to feel at home and increase their community well-being (Lynch [1960]). People generally feel at home in cities whose neighborhoods are recognizable. Comfort resulting from little effort, the argument goes, would impact individual and ultimately collective well-being.

The good news is that the concept of imaginability is *quantifiable*. Since Stanley Milgram's work in New York and Paris (Milgram et al. [1972]; Milgram and Jodelet [1976]), researchers have drawn recognizability maps by recruiting city dwellers, showing them scenes of their city, and testing whether they could recognize where those scenes were: depending on which places are correctly recognized, one could draw a collective psychological map of the city. The problem is that such an experiment takes time (in Milgram's, each participant spent 90 minutes for the recognition task), is costly (because of paid participants), and cannot be conducted at scale (so far the largest one had 200 participants). That is why the link between recognizability of a place and well-being of its residents has been hypothesized, qualitatively shown, but has never been quantitatively tested at scale.

On the first part of this work we test whether the recognizability of a place makes it a more desirable part of the city to live. We make the following contributions:

- We build a crowdsourcing web game that puts the recognizability of London's streets to the test (Chapter 3). It picks up random locations from Google Street

View and tests users to see if they can determine in which subway location (or borough or region) the scene is. In a period of five months, we have collected data from 2,255 users, have built a collective recognizability map of London based on their responses, and quantified the recognizability of different parts of the city.

- By analyzing the recognizability of London regions (Section 4.1), we find that the general conclusions drawn by Milgram for New York hold for London with impressive consistency, suggesting external validity of our results. Central London is the most recognizable region, while South London has little cognitive coverage. Londoners would answer "West London" when unsure, making the most incorrect guesses for that region - hence a West London response bias. We also find that the mental map of London changes depending on where respondents are from - London, UK, or rest of the world.

- We test to which extent an area's recognizability is explained by the area's exposure to people (Section 4.2). In particular, we study exposure to users of three social media services and to underground passengers. By collecting 1.2M Twitter messages, 224K Foursquare check-ins, 76.6M underground trips, and 1.3M Flickr pictures in London, we find that, the more a social media platform's content is geographically salient (*e.g.*, Flickr's), the better proxy it offers for recognizability.

- Upon census data showing the extent to which areas are socially deprived or not, we test whether recognizability of an area is negatively related with the area's socio-economic deprivation (Section 4.3). We find that recognizability is indeed low in areas that suffer from housing deprivation, poor living conditions, and crime.

1.2 Ranking of Pictures in Photo-Sharing Websites

On the second part of the work, we use the data collected, along with other information, to contribute to the problem of identifying interesting and memorable geo-referenced pictures in photo sharing sites.

Currently, to determine which pictures are interesting and memorable, researchers have heavily explored solutions based on either metadata (*e.g.*, likes, comments) or visual features, or the combination of both. The main idea is that interesting pictures are the ones that have received a considerable number of likes or that contain the visual cues people often perceive to be interesting. Unfortunately, as we shall see in Section 5.1, metadata happens to be sparse (only few pictures have considerable number of likes), and visual extraction is computationally expensive and needs to be augmented with additional classes of features to guarantee good levels of accuracy.

To complement those solutions, we set out to consider a key element that has been hitherto left out: the idea of neighborhood. Pictures taken in a neighborhood reflect the neighborhood itself and people's idea of it. Since urban sociology has already dealt with those psychological aspects, we use prominent urban theories that aimed at explaining, for example, why a neighborhood is recognizable and distinctive (Lynch [1960]; Milgram et al. [1972]), and why it is perceived to be beautiful, quiet, and happy (Peterson [1967]). In so doing, we make the following main contributions:

- We gather geo-referenced Flickr pictures and contextual variables (*e.g.*, weather conditions) in the city of London (Section 5.2).
- We identify six main qualities that describe the way a city is psychologically perceived (Section 5.3) and quantify those qualities using proxies computed from Flickr and Foursquare data (Section 5.4).
- We rank Flickr pictures based on those proxies and find that such a ranking enjoys three main desirable properties (Section 5.5). First, it is *effective*, in that, the ranked results are similar yet complementary to the results produced by existing metadata-based solutions. Second, it is computationally inexpensive and, as such, *scalable*: our proxies are defined at the level of city rather than of picture and can be computed offline (no need for real-time updates). Third, it is aware of *contextual* factors (Section 5.6): different values of the same proxy can be computed as a function of, for example, the time of day or weather condition.

As we shall conclude in Section 4.4, these results suggest that, to offer a better mobile experience, future multi-modal learning research should further explore the idea of combining traditional features with domain-specific ones.

Chapter 2

Background

2.1 Hand-drawn maps of the city

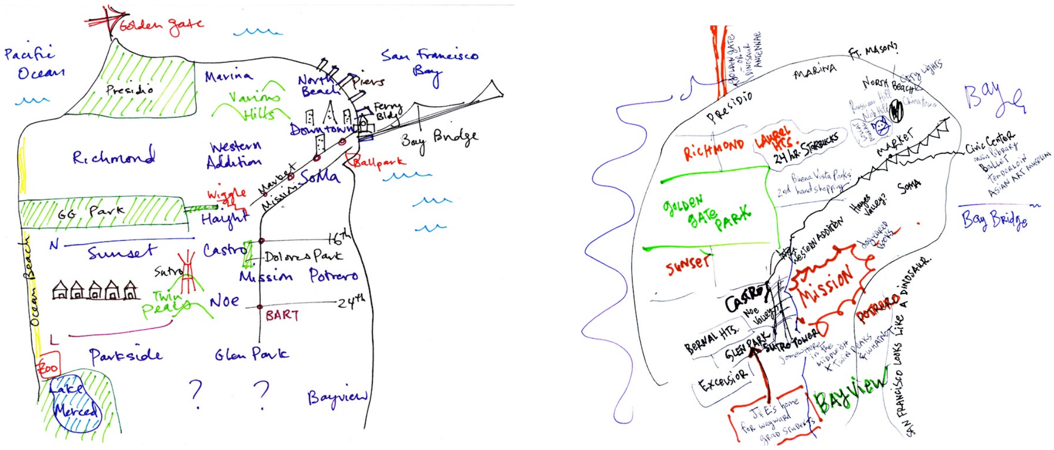


Figure 2.1: Hand-drawn maps of San Francisco from Annechino and Cheng [2011] experiment

In "The Image of the City", Lynch [1960] created a psychological map of Boston by interviewing Bostonians. Based on hand-drawn maps of what participants' "versions of Boston" looked like, he found that few central areas were know to almost all Bostonians, while vast parts of the city were unknown to its dwellers. More than ten years later, Stanley Milgram repeated the same experiment in a variety of other cities (e.g., Paris, New York). Milgram was an American social psychologist who conducted various studies, including a controversial study on obedience to authority and the original small world (six degree of separation) experiment (Milgram [1977]). Milgram was interested in understanding mental models of the city, and he turned to Paris to study them: his participants drew maps of what "their versions of Paris" looked like, and these maps were combined to identify the intelligible and recognizable parts of the city. Since then, researchers have collected people's opinions about neighborhoods in the form of hand-drawn maps in different cities, including (more recently) San Francisco (Annechino and Cheng [2011]) and Chicago (Bentley et al. [2012]).

2.2 Identification of city scenes

Figure 2.2: Scenes from New York from Milgram et al. [1972] experiment

The problem with the hand-drawn map experiment is that it takes time and it is not clear how to aggregate the variety of unique configurations of answers that are bound to appear. One way of fixing that problem is to place a number of constraints on the participants when externalizing their maps. In this vein, before his experience with Paris, Milgram constrained the experiment to the point of reducing it to a simple question: "If an individual is placed at random at a point in the city, how likely is he to know where he is?" (Milgram et al. [1972]). The idea is that one can measure the relative "imaginability" of cities by finding the proportion of residents who recognize sampled geographic points. That simply translates into showing participants scenes of their city and testing whether they can recognize where the scenes are. Milgram did setup and successfully run such an experiment in lecture theaters. Each participant spent roughly 90 minutes on the task, and he collected responses from as many as 200 participants for New York. Hitherto the experimental setup in which maps are drawn has been widely replicated (Annechino and Cheng [2011]; Bentley et al. [2012]), while that in which scenes are recognized has received far less attention. Next, we try to re-create the latter experimental setup at scale by building an online crowdsourcing platform in which each participant plays a one-minute game, a game with a purpose.

Chapter 3

Urbanopticon London

We have created an online game that asks users to identify Google Street View (Panorama) scenes [1] (an example is shown in Figure 3.1) of London. The project aims to learn how its players mentally map different locations around the city, ultimately creating a London-wide map of recognizability.

Figure 3.1: Example of Google Street View Scene

3.1 Mechanics

For each round, the game shows the player a randomly-selected scene in London and ask him/her to guess the nearest subway station, or generally what section of the city (borough/region) (s)he is seeing. Answers shouldn't be too hard, and that is why we choose the finest-grained answer to be subway stations as those are the most widely-used point of references among Londoners and visitors alike. To avoid sparsity problems

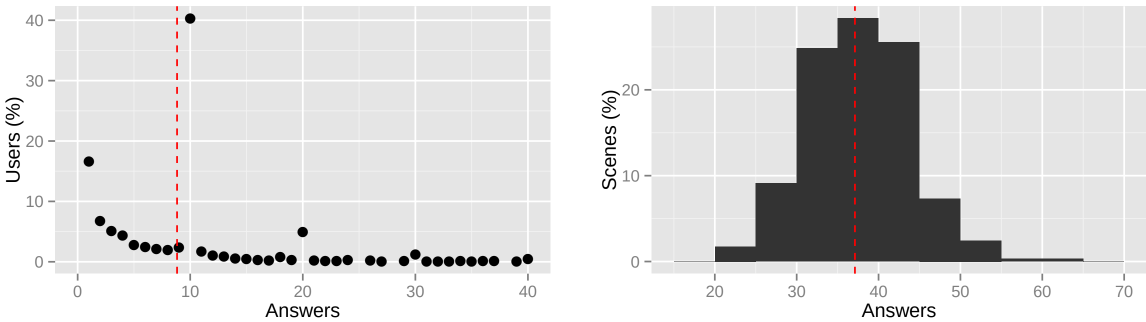
[1] http://bit.ly/SebbCj

(too few answers per picture), a random scene is selected within a 300-meter radius from a tube station but not next to it (to avoid easing recognizability). The idea is that, by collecting a large number of responses across a large number of participants, we can determine which areas are recognizable. By testing which places are remarkable and unmistakable and which places represent faceless sprawl, we are able to draw the recognizability map of London.

3.2 Gamification

Having the game at hand, now the question is whether individuals are likely to adopt and play with it. To increase such a likelihood, we borrow engagement strategies from what is known as "gamification". This relies on identifying the techniques that make video games enjoyable and applying them to other kinds of "more serious" activities, from training soldiers to tackling difficult scientific problems Werbach [2012]. The strategies we implemented in our game include:

Giving Points. Von Ahn and Dabbish [2008] states: "One of the most direct methods for motivating players in games is to grant points for each instance of successful output produced. Using points increases motivation by providing a clear connection among effort in the game, performance, and outcomes". When playing the game, each player receives a score that increases with the number of correct guesses of where a given picture was taken. To enhance the gaming experience, we reward not only strictly correct guesses (which are the only ones considered for experimental sake) but also "geographically close" ones by awarding points based on the Euclidian distance d between a user's guess and the correct answer. The idea is that guesses within a radius of 300 meters still amount to reasonable scores, while those outside it are severely and increasingly penalized depending on how far they are from the correct answer. To reduce the number of random guesses, we allow for an "I Don't Know" answer, which still rewards players with 15 points. After being presented with 10 pictures, the player has completed one round and (s)he can share the resulting score on Facebook or Twitter with only one click. The score is supposed to facilitate the player's assessment of his/her performance against previous game rounds or against other players (Von Ahn and Dabbish [2008]). From the distribution of number answers for each player (Figure 3.2a), we find multiples of 10 to be outliers, suggesting that players do tend to complete at least one round. After the first round, each player is also shown a small questionnaire (e.g., age, gender, location) (s)he is asked to complete. Participants engaging in multiple rounds are identified through browser cookies, which uniquely identify users [2].



(a) Number of Answers per User (b) Number of Answers per Scene

Figure 3.2: Number of Answers for Each User/Scene. *(a)* Each game round consists of 10 pictures - that is why outliers are multiples of 10. The analysis considers the 40% of users who have completed one round. *(b)* The number of answers each scene has received is normally distributed thanks to randomization.

Social Media Integration. Players can post their scores on the two social media platforms of Facebook and Twitter after each round with a default message of the form "How well do you know London? My score . . . ". The goal of such a message is to make Facebook and Twitter users aware of the game.

Randomness. "Games with a purpose should incorporate randomness. For example, inputs for a particular game session are typically selected at random from the set of all possible inputs. Because inputs are randomly selected, their difficulty varies, thus keeping the game interesting and engaging for expert and novice players alike." (Von Ahn and Dabbish [2008]). In our game, pictures are chosen randomly with the hope of creating a sense of freshness and increasing replay value. In addition, randomizing the selection of picture is a good idea for experimental sake. Randomization reduces spatial biases and leads to reliable results, producing a distribution of answers for each picture that is not skewed. In our analysis, such a distribution will turn out to be distributed around a mean of 37.11 (Figure 3.2b), and no scene has less than 20 answers.

Overall, by providing a clear sense of progression and goals that are challenging enough to maintain interest but not so hard as to put players off, we hope to capture that sense of engagement typical of gamification platforms.

[2]Unless two users use the same computer and the same username on it. This situation should represent a minority of cases though.

3.3 From beta to final

We build a working prototype featuring those desirable engagement properties and ascertain the extent to which it works in a controlled beta test involving more than 45 urban planners, architects, and computer scientists. We receive four main feedbacks:

Ease. The game is found to be difficult and, as such, frustrating to play. That is because random pictures from every (remote) part of London are shown. One player said: “I’ve been living in London for the past 35 years and I felt like a tourist. There were so many places I had no clue where they were. It is frustrating to get a score of 200 [out of 1000]! ”. To fix this problem, we manually add pictures of easily recognizable places (e.g., spots that are touristic or close to subway stations) and show them together with the randomly selected places from time to time. For the purpose of study, these “fake” pictures are ignored - they are just meant to improve the gaming experience and retention rate.

Feedbacks. The beta version does not show any feedback about which are the correct answers. A large number of testers feel that the game could be an opportunity to learn more about London. That is why, for each incorrect guess, the final version of the game also shows the right answer.

Sense of purpose. The site does not contain any explanation about the research aims behind the game. Yet, our testers feel that providing a sense of purpose to players was essential. The final version contains a short explanation of how the game is designed for purposes beyond pure entertainment, and how it might be used to promote urban interventions where needed.

Beyond one type of answer. The game asks players to guess the correct subway station. Many testers feel the need for coarser-grained answers. “I know this is Westminster [a borough in London], but I have no idea of the exact tube station!”, says one player. The final version thus allows for multiple types of answers: not only subway stations but also boroughs (50 points) or regions such as Central London and South London (25 points).

To sum up, the final version of the game works by giving a player ten (random plus morale boosting) images in Greater London (Figure 3.3). The player can either guess the tube station, borough, region, or click “Don’t know” to move ahead. At the end of the round, the player is given a total score based on the fraction of correct answers. The score can be automatically shared on Facebook or Twitter, and the player is presented with a survey that asks for personal details like birth location, place of employment, and familiarity with the city itself.

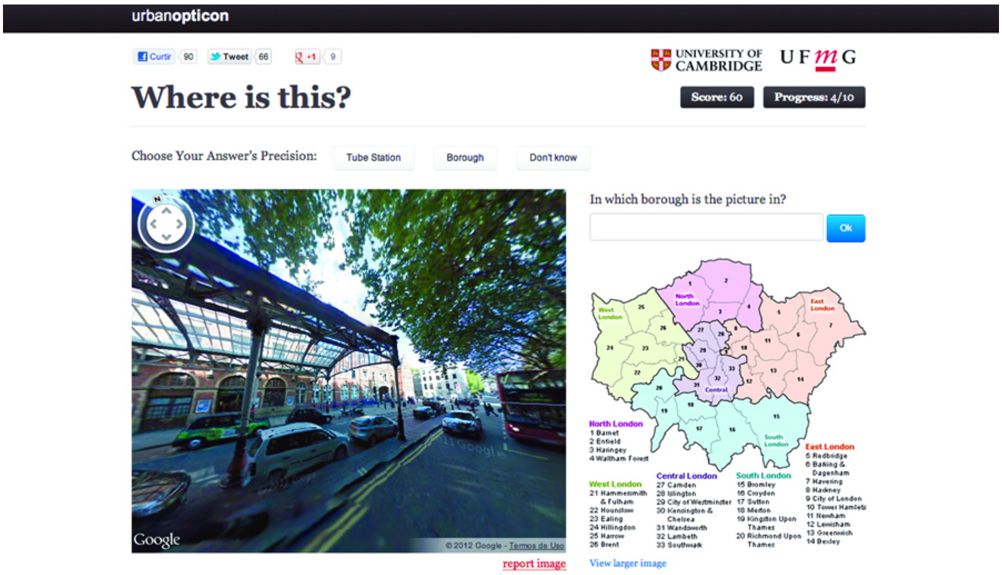


Figure 3.3: Screenshot of the Crowdsourcing Game. The city scene is on the left, and the answer box on the right.

3.4 Launch

We make the final version of the game publicly available and issue a press release in April 2012 (Figure 3.4). Shortly after that, the game is featured in major newspapers, including The Independent (UK national newspaper) and New Scientist. After 5 months, we collect data from as many as 2,255 participants: 739 connecting from London (IP addresses), 973 from the rest of UK, and 543 outside UK. A fraction of those participants (287) specified their personal details. The percentage of male-female participants overall is 60%-40% and slightly changes depending on one's location: it stays 60%-40% in London but changes to 65%-35% in UK and 45%-55% outside it. Also, across locations, average age does not differ from London's, which is 36.4 years old. As for geographic distribution of respondents, we find a strong correlation between London population and number of respondents across regions ($r = 0.82$). Having this data at hand, we are now ready to analyze it.

	London	UK	World	Total
Answers	7,238	8,705	3,972	19,915
Users	739	973	543	2,255
Gender (%)				
Male	59.13	64.34	46.51	59.58
Female	40.87	35.66	53.49	40.42
Age (%)				
<18	0.87	0.78	0.00	0.70
18-24	16.52	24.81	9.30	19.16
25-34	41.74	38.76	51.16	41.81
35-44	16.52	13.95	20.93	16.03
45-54	13.91	13.95	6.98	12.89
55-64	5.22	6.20	9.30	6.27
65+	5.22	1.55	2.33	3.14
Mean (years)	36.39	33.88	34.52	34.98

Table 3.1: Statistics of Participants. Gender and age are available for those 287 participants (13%) who have been willing to provide personal details.

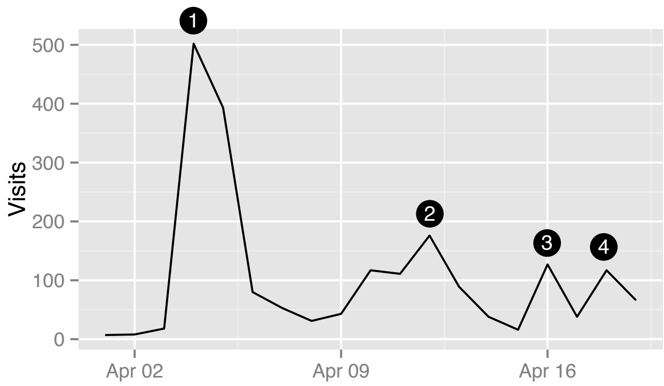


Figure 3.4: Initial Visitors on the Site. Peaks are registered for: (1) Cambridge press release; (2) The Independent article; (3) New Scientist article; and (4) residual sharing activity on Facebook and Twitter.

Chapter 4

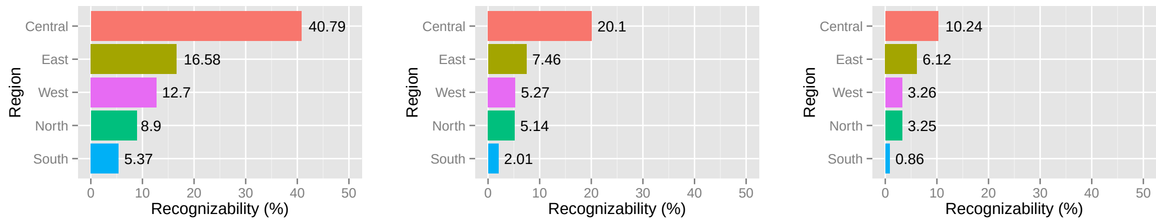
Recognizability Results

4.1 Relative recognizability

The goal of this project is to *quantify* the relative recognizability of different parts of London. Since familiarity with different parts of the city might depend on place of residence, we filter away participants outside London and consider Londoners first. According to the Greater London Authority's division, London is divided into five different city (sub) regions [1]. Thus, our first research question is to determine which proportion of the Google Street View scenes from each region were correctly attributed to the region. Since users were asked to name either the borough of each scene or the subway station closest to it, we consider an answer to be correct, if the region of the scene is the same as the region of the answered borough/subway station. For each of the five regions, we compute the region's percentage *recognizability* by summing the number of correct answers and then dividing by the total number of answers. Figure 4.1 reports the results. Clearly, Central London emerges as the most recognizable of the five regions, with about two and a half as many correct placements as the others. Conventional wisdom holds that Central London is better known than other parts of the city, as it hosts the main squares, major railway and subway stations, and most popular touristic attractions and night-life "hotspots". Interestingly, the East Region is twice as recognizable than the North Region. It is difficult to draw conclusions on why this is. However, the three most likely explanations are:

Sample Bias. It might depend on the distribution of the home and work addresses of our participants. However, that is unlikely, as the correlation between London population across regions and number of participants who answered the survey is as

[1] http://bit.ly/UewEtu



(a) *Region*-level recognizability (b) *Borough*-level recognizability (c) *Subway station*-level

Figure 4.1: Recognizability Across London Regions. This is computed based on whether scenes are recognized at: *(a) region* level; *(b) borough* level; or *(c) subway station* level.

high as $r = 0.82$.

Large volume of visitors. High recognizability for the East part of the city can be explained by an experiential effect. Large numbers of people are expected to visit that part of the city: workers at Canary Wharf, visitors to Olympic Park, Excel, City airport, and O2 arena. A recent study of Londoners' whereabouts on Foursquare conducted by Bawa-Cavia [2011] found them to be skewed towards mostly Central London and partly East London - especially the central east part. The north parts are unlikely to have been visited by similar volumes of people. In Section 4.2, we will see that there is a significant correlation between recognizability of an area and the area's exposure to specific subgroups of individuals. For example, we will see that the more passengers use an area's subway station, the more recognizable the area ($r = 0.45$).

Distinctiveness of the built environment. The East region includes most of the City and Canary Wharf (financial area with skyscrapers), as well as the O2 arena and Docklands region, all more visually recognizable areas than comparable parts of the North region. Also, East London has been affected by large homogeneous post-war housing projects that make the area quite distinctive (Glendinning and Muthesius [1994]).

Next, we adopt a more stringent criterion of recognition, that is, we determine what proportion of the scenes in each of the five regions were placed in the correct *borough*. By analyzing the answers at borough-level, we find substantial differences (Figure 4.1b). A scene placed in Central London is almost three times more likely to be placed in the correct borough than a scene in East London, and are four times more likely than a scene in West or North London.

When we then apply even a more stringent criterion of recognition (subway sta-

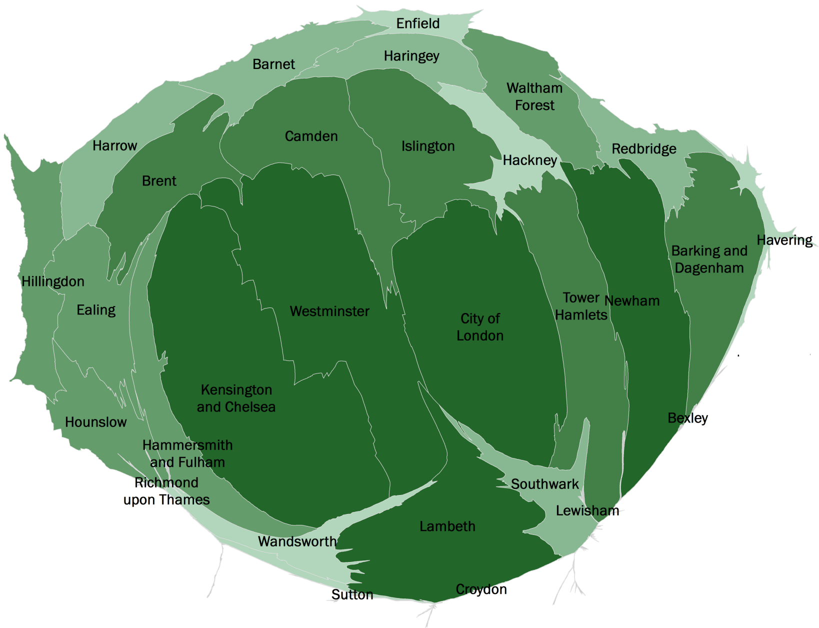


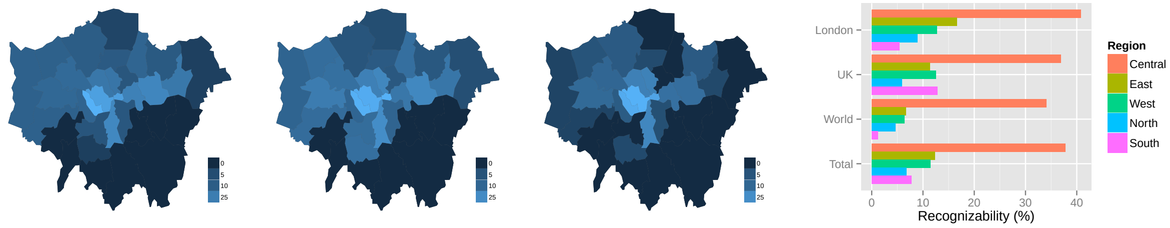
Figure 4.2: Cartogram of London Boroughs. The geographic area is distorted based on borough's recognizability.

tion), the correct guesses are drastically reduced (Figure 4.1c), as one expects. Interestingly, the information value of Central London is less pronounced. Central London scenes are only one and a half time more likely to be associated with the correct subway station than a scene in East London. Guessing the correct subway stations is hard, the more so in the central part of the city where stations are close to each other. During post-game interviews, one participants noted: "Perhaps people know where places are, but have difficulty identifying which of the [subway stations] it is actually close to." Despite these differences, the *relative* recognizability (ranked recognizability of the five regions) does not change. Figure 4.2 shows the cartogram of London boroughs. The geometry of the map is distorted based on recognizability scores. Central London dominates, while South London is relegated at the bottom.

Another aspect to consider is that one is likely to recognize areas closer to where one lives or works. Based on our survey respondents, we find that there is no relationship between recognizability of a scene and a respondent's self-reported home location. On the contrary, participants are more likely to recognize scenes in Central London rather than scenes in their own boroughs.

The recognizability of each region does not change depending on which parts of the city Londoners live, but does change depending on whether participants are in UK or not. Based on our participants' IP addresses [2], we infer the cities where they

[2] There might be cases of misclassification of cities and of people who use VPNs. However, at the



(a) Londoners (b) UK (c) Outiside UK (d) All

Figure 4.3: Recognizability Across London Regions by Respondent Location. On the maps of London, lighter colors correspond to more recognizable boroughs.

Region actually is	**But identified as** C	E	W	N	S	Combined Errors	Don't Know
Central	**40.79**	4.52	4.33	1.03	2.13	12.02	47.19
East	6.97	**16.58**	6.80	6.30	7.46	27.53	55.89
West	10.10	6.42	**12.70**	5.77	5.92	28.21	59.09
North	6.85	4.79	12.67	**8.90**	7.53	31.85	59.25
South	6.04	5.37	11.41	3.36	**5.37**	26.17	68.46
Response Bias*	29.96	21.1	**35.21**	16.46	23.04		

* popular among wrong guesses

Table 4.1: Matrix of Correct Classifications and Misclassifications.

are connecting from, and compute aggregate correct guesses by respondent location - that is, by whether participants connect from London, from the rest of UK, or outside UK (Figure 4.3). As expected, the number of correct guesses drastically decreases for participants outside London - but with two exceptions. First, scenes of South London are more recognizable for participants in the rest of UK than for Londoners themselves. That is because Londoners tend to know Southfields (known as "The Grid", which a series of parallel roads that consist almost entirely of Edwardian terrace houses), while people in the rest of UK recognize scenes not only in Southfields, but also in Clapham South, Balham, South Wimbledon, and Tooting Broadway in that order. Second, recognizability of Central London remains the same across participants from all over: participants outside UK are as good as those inside it at recognizing scenes in Central London. Hosting the most popular tourist attractions in the world, Central London is vividly present in the world's collective psychological map.

three coarse-grained levels of London *vs.* rest of UK *vs.* rest of the world, misclassification should have a negligible effect.

So far we have focused on correct guesses. Now we turn to errors that respondent often make, looking for widely-shared sources of confusion. We wish to know in which regions (e.g., North, South) a scene from, say, East London is often misplaced. To this end, Table 4.1 shows a matrix reporting both the percentage of correct guesses and that of wrong ones for each region. Central London is pre-eminent in Londoners' shared psychological maps as it is hardly confused with any other region. At times, instead, South and North London are thought to be West. It seems that, if respondents do not know where to place a scene, they would preferentially opt for West London. Indeed, the West part of the city is the most popular answer for those who end up guessing wrongly (last row in Table 4.1). We found a *West London response bias*, as Milgram et al. [1972] would put it [3].

Summary. Taken together, the results suggest two generalizable principles on why people recognize an area. They do so because they are exposed to it (Central London attracts dwellers from all over the city), and because the area offers a distinctive architecture (e.g., stadium, tower building) or cultural life (as the central part of East London notoriously does). Milgram found the very same two principles to hold for New York as well in 1972. So much so that Milgram hypothesized that the extent to which a scene will be recognized can be described by $R = f(C \cdot D)$, where R is recognition (our recognizability), C centrality of population flow (in the next section, we will see how to compute flow of subway passengers), and D is the social or architectural distinctiveness. It follows that, with simplifying assumptions (e.g., f is a linear relationship), one could derive an area's social or architectural distinctiveness by simply dividing recognizability by subway passenger flow. Since we are interested in the *relative* recognizability and flow, we take the rank values for these two quantities, compute their ratio, and report the results in Table 4.2. The most distinctive area is Blackfriars. It should be no coincidence that its older parts happen to "have regularly been used as a filming location in film and television, particularly for modern films and serials set in Victorian times, notably Sherlock Holmes and David Copperfield" [4]. In line with Milgram et al. [1972]'s experiment with New Yorkers , we find that the acquisition of a mental map is not necessarily a direct process but can also be indirect through, for example, movies. The following quote from one of our participants is telling: "I've done the quiz 3 or 4 times in the last couple of days, and am surprised how well I am doing - not just because I live in New Zealand, many thousands of kilometres from London (although I did live there for 10 years), but mainly because I am getting good scores on parts of London I

[3] Milgram found that New Yorkers would opt for answering "Queens" when unsure - hence he referred to a "Queens response bias"

[4] `http://en.wikipedia.org/wiki/Blackfriars,_London`

name	R	C	r_R	r_C	D
Blackfriars	9.09	4583	30	2	15.00
Park Royal	20.00	13119	61	5	12.20
Pinner	10.00	13823	37	6	6.17
Royal Oak	10.00	16681	37	8	4.63
Westbourne Park	16.66	24593	54	13	4.15
Hornchurch	7.14	11988	16	4	4.00
Essex Road	5.55	2027	4	1	4.00
Oakwood	11.11	22321	41	11	3.73
Hillingdon	6.67	9482	11	3	3.67
Acton Town	40.00	33022	73	22	3.32

Table 4.2: Subway Stations of Socially/Architecturally Distinctive Areas. For each area, R is the recognizability, C is the flow centrality (number of unique subway passengers), r_R and r_C are the corresponding ranked values, and D is the normalized distinctiveness.

have never been to. North and Eastern boroughs like Brent and Haringey seem to be recognisable, even though I have never knowingly gone there. Possibly some recognition from TV programs, or just - could it be - that there is something intrinsically North London about certain types of houses? ”

4.2 Recognizability and Exposure

4.2.1 Digital Data for Exposure

The goal of the game is to quantify the recognizability of the different parts of the city. It has been shown by Milgram et al. [1972] that New Yorkers are able to recognize an area partly because they were exposed to it. Thus, to quantify the extent to which it is so in London, we measure the exposure that an area receives by computing the number of overall unique individuals who happen to be in the area. These individuals are of four subgroups: those who post Twitter messages while in the area, those who visit locations (e.g., restaurants, bars) and say so on Foursquare, those who take pictures of the area and post them on Flickr, and those who catch a train in the closest subway station. We are thus able to associate the recognizability of an area with the area's exposure both in the physical and virtual worlds.

Twitter geo-enabled users. Our goal is to retrieve as large and unbiased a sample of geo-referenced tweets as possible. To do this, we use the public streamer API, which connects to a continuous feed of a random sample of all ever shared tweets, and crawl geo-referenced tweets within the bounding box of Greater London. During the period that goes from December 25th 2011 to January 12th 2012, we retrieve 1,238,339 geo-referenced tweets posted by 57,615 different users.

Foursquare users. *Gowalla, Facebook Places*, and *Foursquare* are popular mobile social-networking applications with which users share their whereabouts with friends. In this work, we consider the most used social-networking site in London - Foursquare (Bawa-Cavia [2011]). Users can check-in to locations (e.g., restaurants) and share their whereabouts. We consider the geo-referenced tweets collected by Cheng et al. [2011a]. They collected Twitter updates (single tweets) that report Foursquare check-ins all over the world. We take the 224,533 check-ins that fall into Greater London. Those check-ins are posted by 8,735 users.

Flickr users. We collect photo metadata from `Flickr.com` using the site's public search API. To collect all publicly available geo-referenced pictures in the Greater London area, we divide the area into 30K cells, search for photos in each of them, and retrieve metadata (e.g., tags, number of comments, and annotations). The final dataset contains metadata for 1,319,545 London pictures geo-tagged by 37,928 users. This reflects a complete snapshot of all pictures in the city as of December 21st 2011.

Subway passengers. In 2003, the public transportation authority in London introduced an RFID-based technology, known as Oyster card, which replaced traditional paper-based magnetic stripe tickets. We obtain an anonymized dataset containing a record of every journey taken on the London rail network (including the London Underground) using an Oyster card in the whole month of March 2010. A record registers that a traveler did a trip from station a at time t_a, to station b at time t_b. In total, the dataset contains 76.6 million journeys made by 5.2 million users, and is available upon request from the public transportation authority.

Demographics of the individuals under study. Activity analyzed in this work clearly relates only to certain social groups, and the exclusionary aspect of certain segments of the population should be acknowledged. It would be thus interesting to compare the demographics of the different types of individuals we are studying here. From a recent Ignite report on social media (Ignite [2012]), global demographics of Foursquare and Twitter show a pronounced skew towards university educated 25-34 year old women (66% women for Foursquare and 61% for Twitter), while those of

Flickr show a pronounced skew towards university educated 35-44 year old women (54% women). The demographics of subway passengers is by far the most representative but is also slightly skewed towards male with above-average income in the two age groups of 25-44 and 45-59 (TfL [2011]). Instead, demographics of our London gamers show a skew towards 25-34 year old men (60% men). Thus, compared to social media users, our gamers reflect similar age groups but are more likely to be men. This demographic comparison should inform the interpretation of our results.

4.2.2 Recognizability and Exposure

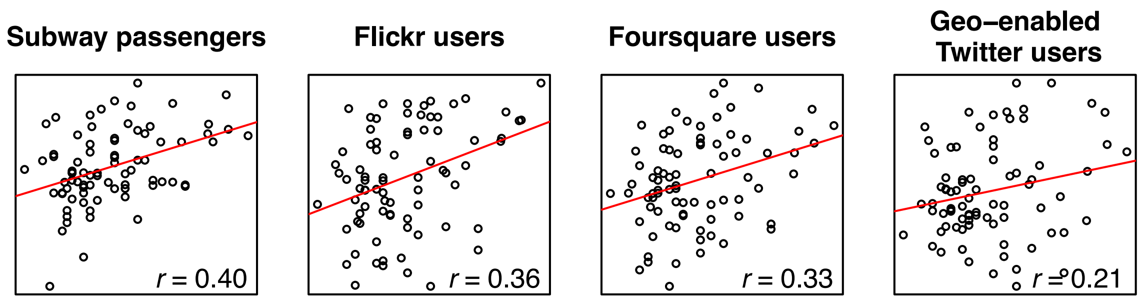


Figure 4.4: Area recognizability vs. Exposure to Four Classes of Individuals. Correlations are computed at *borough* level.

After computing each area's exposure to people of four subgroups[5] (i.e., to users of the three main social media sites and to underground passengers), we are now ready to relate the area's exposure to its recognizability. We compute the Pearson's product-moment correlation between recognizability and exposure, for all four classes of individuals. Pearson's correlation $r \in [-1, 1]$ is a measure of the linear relationship between two variables. We expect that the more a given class of individuals is representative of the general population, the higher its correlation with recognizability. When computing the correlation, if necessary (e.g., because of skewness), variables undergo a logarithmic transformation. Figure 4.4 shows the relationship between recognizability and exposure to the four classes of individuals, with corresponding correlation coefficients (which are all significant at level $p < 0.001$). To put results into context, we should say that the exposure measures derived from the three social media sites all show very similar pair-wise correlations with exposure to subway passengers ($r \approx 0.60$), yet their correlations with recognizability show telling differences. Given that subway passengers are slightly more representative of the general population than social media users (TfL

[5]By area, we mean UK census area also known as Lower Super Output Area, which we will introduce in Section 4.3.

Exposure	Central	East	West	North	London
Subway	0.87	0.65	0.63	0.95	0.73
Flickr	0.62	0.50	0.27	0.89	0.63
Foursquare	0.72	0.36	0.22	0.97	0.58
Twitter	0.56	0.28	0.11	0.97	0.52

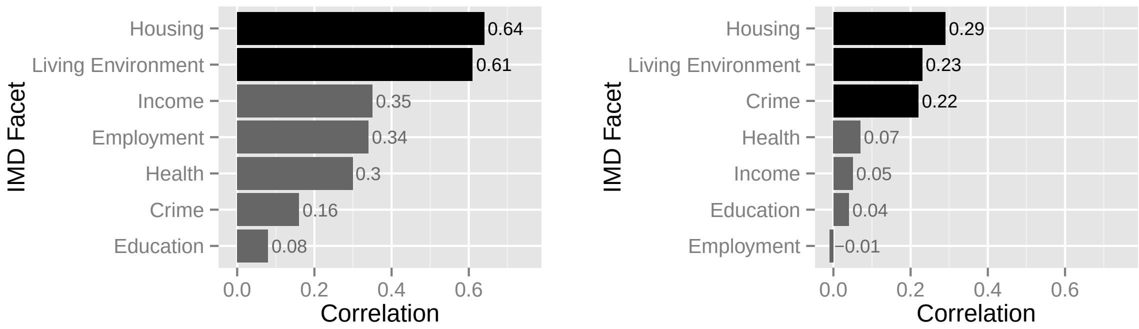
Table 4.3: Correlations between recognizability and Exposure by Region. Correlations are computed at *region* level. South London does not have enough subway stations to attain statistically significant correlations.

[2011]), it comes as no surprise that they show the highest correlation ($r = 0.40$). Both Flickr and Foursquare users are also associated with robust correlations ($r = 0.36$ and $r = 0.33$). By contrast, having the least geographically salient content, Twitter shows a moderate correlation ($r = 0.21$). If we break the results down to regions (Table 4.3) and show which regions' recognizability is easy to predict from exposure and which not, we see that exposure to *any* social media subgroup of individuals would predict the recognizability of North London ($r = 0.95$). By contrast, the subgroup whose exposure correlates with recognizability the most in Central London is Foursquare ($r = 0.72$), and in East London is Flickr' ($r = 0.50$). That is largely because Foursquare activity is skewed towards Central London.

4.3 Recognizability and Well-being

As already mentioned in the introduction, Lynch [1960] outlined a theory connecting urban recognizability to a person's well-being. To test this theory, we now gather census data on an area's socio-economic well-being and relate it to the area's recognizability.

Facets of Socio-economic Well-being. Since 2000, the UK Office for National Statistics has published, every three or four years, the Indices of Multiple Deprivation (IMD), a set of indicators which measure deprivation of small census areas in England known as Lower-layer Super Output Areas (Mclennan et al. [2011]). These census areas were designed to have a roughly uniform population distribution so that a fine-grained *relative* comparison of different parts of England is possible. As per formulation of IMD, deprivation is defined in such a way that it captures the effects of several different factors. More specifically, IMD consists of seven components: 1. *Income* deprivation (e.g., number of people claiming income support, child tax credits or asylum); 2. *Employment* deprivation (e.g., number of claimants of jobseeker's allowance or incapacity benefit); 3. *Health* deprivation (e.g., including a standard measure of premature death, rate of



(a) Borough Level (b) Area Level

Figure 4.5: Correlation between recognizability and Deprivation at the Level of *(a)* Borough or *(b)* Area. For both, the composite index IMD is not shown as it does not correlate. Also, correlations significant at level $p < 0.001$ are shown with black (as opposed to grey) bars.

adults suffering mood and anxiety disorders); 4. *Education* deprivation (e.g., education level attainment, proportion of working adults with no qualifications); 5. barriers to *Housing* and services (e.g., homelessness, overcrowding, distance to essential services); 6. *Crime* (e.g., rates of different kinds of criminal act); 7. *Living* Environment Deprivation (e.g., housing condition, air quality, rate of road traffic accidents); and finally a composite measure known as IMD which is the weighted mean of the seven domains.

Recognizability and Well-being. We start at borough level, correlate each facet of deprivation with recognizability, and obtain the results shown in Figure 4.5a [6]. We find that the composite score IMD does not correlate with recognizability at all. Neither does income, education, or (un)employment. What correlates are aspects less related to economic well-being and more related to social well-being: boroughs with low recognizability tend to suffer from housing deprivation ($r = 0.64$) and poor living environment ($r = 0.62$). Given the strong correlations [7], one could easily predict which boroughs suffer from housing deprivation and poor living conditions based on relative recognizability scores.

One might now wonder whether that would also be possible from social media data. We correlate each of the deprivation facets with exposure to the four subgroups

[6] To ease the interpretation of the correlation coefficients, we transformed (inverted) the deprivation scores, in that, the higher they are (e.g., high crime index), the better it is (e.g., low-crime area). We would thus expect the correlations between transformed deprivation scores and recognizability to be generally positive.

[7] Unless otherwise noted all correlations are significant at level $p < 0.001$.

(subway passengers plus users of three social media). For housing, we see that data on recognizability is hardly replaceable by social media data. Boroughs not suffering from housing deprivation (as per log-transformed score) are more recognizable ($r = 0.64$), and yet do not seem to be more exposed to our subgroups - all correlations between housing deprivation and exposure are not statistically significant. Instead, for living environment, we see that data on recognizability can be replaced by social media data. Boroughs with good living conditions are more recognizable ($r = 0.61$), and do tend to be more exposed to subway passengers ($r = 0.56$), Flickr users ($r = 0.57$), Foursquare users ($r = 0.52$), and Twitter users ($r = 0.46,\ p < 0.01$).

Here we are not claiming that each census area in a borough is the same. If we were to say that, we would commit an ecological fallacy. For indicators that show high variability within a borough, however, there is a danger of committing such a fallacy. We therefore investigate correlations at the lower geographic level of census area. We correlate each facet of deprivation with recognizability and obtain the results shown in Figure 4.5b. Again, the composite score IMD does not correlate with recognizability, while housing, living environment, and crime all do: areas with low recognizability tend to suffer from housing deprivation ($r = 0.29$), poor living environment ($r = 0.23$), and crime ($r = 0.22$). Crime has been added to the list of indicators associated with recognizability, and that is because crime is one of the deprivation facets that varies the most within a borough among the seven.

To sum up, from the previous results, we might say that, based on recognizability scores of *areas*, we could predict *whether* an area suffers from crime or not. Instead, based on recognizability scores of *boroughs*, one could predict not only *whether* but also *to which extent* a borough suffers from poor living conditions and housing deprivation. By contrast, social media data could only be used to identify boroughs with poor living conditions.

4.4 Discussion

This work is deeply rooted in early urban studies but also taps into recent computing research, especially research on "games with a purpose", whereby one outsources certain activities (e.g., labeling images) to humans in an entertaining way (Von Ahn and Dabbish [2008]); research on large-scale urban dynamics (Crandall et al. [2009]; Cranshaw et al. [2012]; Noulas et al. [2012]); and research on how location-based services affect people's behavior (Bentley et al. [2012]; Cramer et al. [2011]; Lindqvist et al. [2011]). Initially, with this study, we were aiming at informing social media research in the urban context by establishing which social media data could be used as proxy for recognizability and exposure (key aspects in studies of urban dynamics). It turns out that the answer is complex, suggesting a word of caution on researchers not to take social media data at face value. However, there is one generalizable finding: the more the content is geographically salient (e.g., Foursquare's whereabouts *vs.* Twitter messages), the more it is fit for purpose.

Chapter 5

Ranking City Pictures Using Urban Data

In this chapter, we use the data collected from the previous experiment, along with other urban data, to contribute to the problem identifying interesting geo-referenced pictures in photo sharing sites.

5.1 Related Work

To identify the pictures users tend to like, researchers have often used metadata. This is generally of two types. The first is *textual* metadata and is the most widely used: it consists of comments and tags users have annotated a picture with (van Zwol et al. [2010]). The second type of metadata consists of *social* features and has received less attention. van Zwol et al. [2010], for example, used the communication and social network of Flickr users for predicting the number of likes (favorites) a picture has received. They found that social features alone yielded a good baseline performance, but the addition of textual features resulted in greatly improved precision and recall.

Despite showing good accuracies, approaches that rely on metadata suffer from coverage. That is because the frequency distributions of tags, comments, or any other social feature are power law: few pictures are heavily annotated, while many have little (if any) annotation (Sigurbjörnsson and van Zwol [2008]). As such, approaches solely relying on metadata do not work for most of the pictures.

In those situations, researchers have explored the use of visual categorization. The most effective method is called bag-of-word model (Datta et al. [2006]). This computes descriptors at specific points in an image. It has been shown that, given an image's descriptors, machine learning algorithms are able to predict whether people tend to find the image interesting and appealing (Redi and Merialdo [2012]). The problem with visual categorization is that it is computationally expensive: it might take weeks to process 380 hours of video frames (van de Sande et al. [2011]). To fix that, research effort has gone into designing faster methods and building new parallel computing architectures.

Within the multimedia research community, a considerable number of research papers have been proposing the *combined* use of metadata and visual features. These works employ multi-modal machine learning approaches that model topical, visual, and social signals together. Their goal has mainly been to predict which pictures users find appealing and aesthetically pleasing (van Zwol et al. [2010]).

Those previous solutions have been designed to fit the general-purpose scenario of web ranking. However, when considering how pictures will be consumed on mobile phones, one might find that location becomes key: ranking pictures in location-based services might consider whether the neighborhoods in which the pictures were taken have high recognizability, are highly visited, beautiful, or quiet. We set out to do just that by identifying desirable urban qualities from seminal work done in the 1970s.

5.2 Datasets

Within the bounding box of the city of London, we crawled 1.2M geo-referenced pictures using the Flickr public API. We also crawled their metadata, which includes: latitude and longitude points, number of comments, tags, upload date, taken date, number of favorites (those are Flickr's equivalent of likes), and number of views. The last two values have been used by past research as a signal of user preference for pictures (Yildirim and Süsstrunk [2013]): the higher a picture's ratio of number of favorites to number of views, the more the picture's views have been converted into user likes. Figure 5.1 shows the density of photos in our dataset across London.

In addition to geo-referenced pictures, we collect data about two contextual factors. The first is 'time of the day' and is computed based on the time each picture was taken: if it was taken between 6am and 10pm, we consider it to be taken during the 'day' (similar to Martínez and Santamaría [2012]); otherwise, we consider it to be taken at 'night'. This results in 79.5% of the pictures being taken at 'day' and 20.5% at 'night' (Figure 5.2). Alternative temporal segmentations could have been chosen. We explored a variety of them and they all resulted in comparable percentages for day *vs.* night. The imbalance for number of pictures between day *vs.* night is natural as people tend to take more pictures during the day. However, this imbalance does not

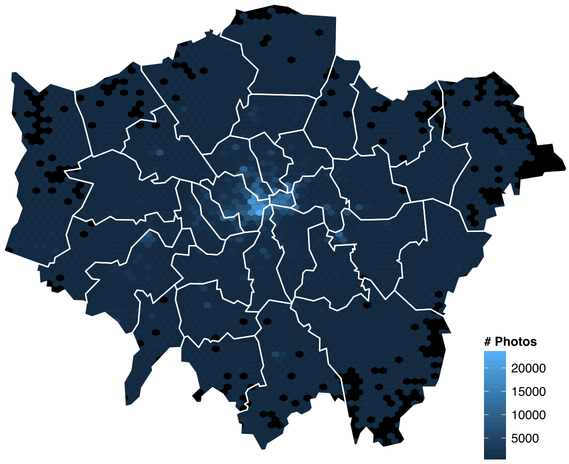


Figure 5.1: London Density Map of Photos in our Dataset.

compromise any of our results as there are enough pictures at night to ensure statistically significance. The temporal span of the pictures in our dataset goes from 2002 to 2013.

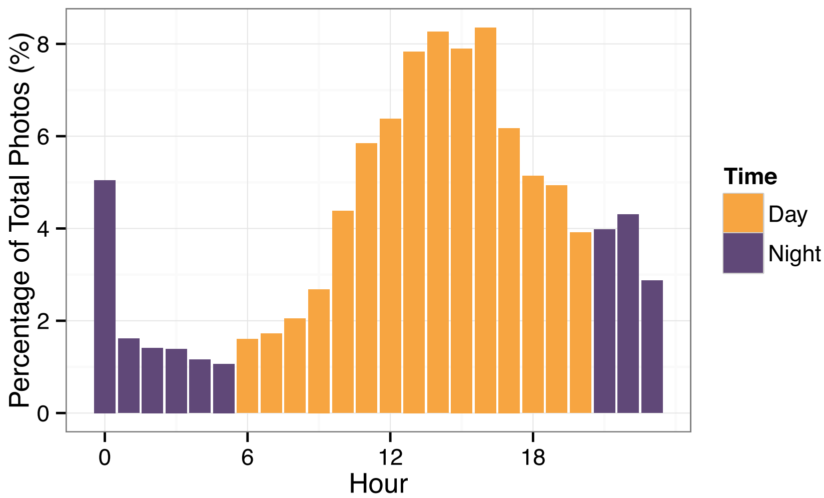


Figure 5.2: Fraction of Photos in each Hour of the Day ('day' is $[6am - 10pm]$). We have 79.5% of the pictures being taken during the 'day' and 20.5% during 'night'.

The second contextual feature for which we collect data concerns weather conditions. We collect weather data from the British Atmospheric Data Centre for 11 years (2002-2013)[1]. This consists of hourly observation and amounts to roughly 10GB of data. We classify weather conditions as follows: cloudy *vs.* not-cloudy; hot *vs.* cold; humid *vs.* dry; high visibility *vs.* low visibility; windy *vs.* not-windy.

[1] http://badc.nerc.ac.uk/data/ukmo-midas/WH_Table.html

5.3 Urban Qualities

Before mining those datasets, we need to identify the urban qualities that reflect people's psychological perceptions of the city. In addition to **Recognizability** and **Distinctiveness** (Section 4.1), we also make use of two other qualities:

Eventfulness Yildirim and Süsstrunk [2013] have partly shown that routinely visited places (*e.g.*, the daily street from home to the train station) are expected to be associated with geo-referenced content that is less interesting than that associated with places that are visited in exceptional circumstances (*e.g.* places visited only on weekends or holidays).

To capture that intuition, we compute a measure that we call 'routine score'. We do so on a Foursquare dataset released by Cheng et al. [2011b]: 22,387,930 Foursquare check-ins collected from September 2010 to January 2011. From these check-ins, we extracted those that happen to be in London: 230,785 check-ins in 8,197 places from 8,895 distinct users. To avoid computing anomalous scores, we filter out users with less than 10 check-ins and places which were visited by less than 10 distinct users. Then, for each user, we compute the fraction of times (s)he visits each location. By aggregating those user scores at each location (we used a geometric average as scores are skewed), we are able to compute a location's routine score in the range $[0, 1]$: the higher it is, the more routine visits the location enjoys. To ease illustration, from the routine score, we derive its complementary measure, which we call 'eventfulness score' and is just 1 minus the routine score.

Beauty, Quiet, and Happiness Not only mental maps but also aesthetically pleasing environments are associated with community well-being. Researchers in environmental aesthetics have widely studied the relationship between well-being and the ways urban dwellers perceive their surroundings (Nasar [1994]; Taylor [2009]; Weber et al. [2008]). In 1967, Peterson [1967] proposed a methodology for quantifying people's perceptions of a neighborhood's visual appearance: he selected ten dimensions that reflected visual appearance (e.g., preference for the scene, greenery, open space, safety, beauty) and had 140 participants rate 23 pictures of urban scenes taken in Chicago along those dimensions. Based on his analysis, he concluded that preferences for urban scenes are best captured by asking questions concerning the beauty and safety of those scenes: beauty is synonymous with visual pleasure and appearance. To capture visual pleasure, the concept of *beauty* is thus key, and that is why it is our first perception quality. Beauty is indeed one of the three dimensions that recent work concerned with urban aesthetics has tried to quantify (Quercia et al. [2013]). In this work, researchers collected votes on the extent to which hundreds of London urban scenes were perceived to be beautiful, quiet, and happy by more than 3.3K crowdsourcing participants. We get hold of the scores for beauty, quiet, and happiness at both subway and borough levels.

The researchers chose *quiet* because of popular discussions on 'city life'. Sound artist Jason Sweeney proposed a platform where people crowdsource and geo-locate quiet spaces, share them with their social networks, and take audio and visual snapshots. It is called Stereopublic[2] and is "an attempt to both promote 'sonic health' in our cities and offer a public guide for those who crave a retreat from crowds" - both for those in need of quietness and for people with disabilities, like autism and schizophrenia.

The remaining quality is that of *happiness*. This quality reflects the ultimate goal behind the 1970s research we have referred to: Milgram, Lynch and colleagues were after understanding which urban elements help to create intelligible spaces and would ultimately make residents happy.

Overall, we consider the three qualities of **beauty**, **quiet**, and **happiness** plus **recognizability**, **distinctiveness**, and **eventfulness**. Each of those qualities is defined at the two geographic levels of study: subway and borough levels.

5.4 Modeling Urban Qualities

To see how our urban qualities change depending on contextual factors, we need to build predictive models for each of them. To see why, consider our urban quality of beauty as an example. Its values could be represented on a heat map of London: darker squares (larger values) contain crowdsourced pictures considered to be beautiful, while lighter squares (smaller values) contain pictures considered to be less beautiful. One could then build a predictive model for beauty that estimates the extent to which those squares are dark (or light) on input of, say, Flickr or Foursquare metadata (e.g., likes on pictures, check-ins in Foursquare venues). By having this model at hand and stratifying the input metadata according to, say, time of day (e.g., number of favorites for photos taken at night), one could test which squares the model predicts to be beautiful at night, assuming that its predictions do not dramatically change with the contextual factors. We will test the validity of this assumption in Section 5.6.

The input features are derived from Flickr and Foursquare. These features include number of views, number of favorites, number of comments, number of tags, number of

[2] `http://www.stereopublic.net/`

photos, number of unique Flickr users, number of unique Foursquare users, and number of check-ins. Since the urban qualities are defined at the levels of subway station and borough, we aggregate those features at the two levels. Then, if skewed, the features are log-transformed and, as such, their averages are not arithmetic but geometric.

On input of those features, we put the following models to test: linear model (least squares), decision tree regressor, support vector regression, ADA boost regressor, gradient boosting regressor, extra trees regressor and random forest regressor. For all the models, we have tried different parameter values and found that the default ones specified in the scikit-learn library[3] produced reasonable results. For brevity, we report only those results.

The predictive accuracies of the models are expressed with two measures: i) Mean Squared Error (MSE), which reflects the differences between the values predicted by the model under test and the actual values; and ii) Spearman's rank correlation ρ between two ordered lists of areas: in one list, areas are ranked by the model's predicted values; in the other list, areas are ranked by the actual values; ρ ranges from -1 to 1: it is 0 if the two lists are dissimilar, +1 if the two lists are exactly the same (best match), and -1 if the two lists are exactly reversed.

Figure 5.3 shows the models' error values (left panel) and accuracy values (right panel) for "in sample" predictions[4]. The large pink area reflects the statistical significance of the baseline being extremely low. The more sophisticated models (e.g., ADA boost, Gradient Boost) perform exceptionally well, yet simpler models (e.g., linear model, decision tree) show competitive performance: for all qualities other than quiet, the squared errors are below 0.03. The same goes for Spearman's ρ, which is always above 0.50 for all models. If we reduce the number of input features from 12 to 6, those results do not significantly change, suggesting that overfitting has little to do with such good prediction accuracies. To further reinforce this last point, we will now see to which extent such predicted values are associated with actual appealing content.

5.5 Rankings by Urban Qualities

We have just established how accurately off-the-shelf models can predict the urban qualities from Flickr and Foursquare metadata. However, we have not yet ascertained whether the predictions of those models would ultimately result into the selection of appealing geo-referenced pictures. To ascertain that, we need to determine which pictures are to be considered appealing. We do so by resorting to the widely-used

[3] `http://scikit-learn.org/stable/`

[4] We could not use cross-validation given the limited number of subway stations or boroughs.

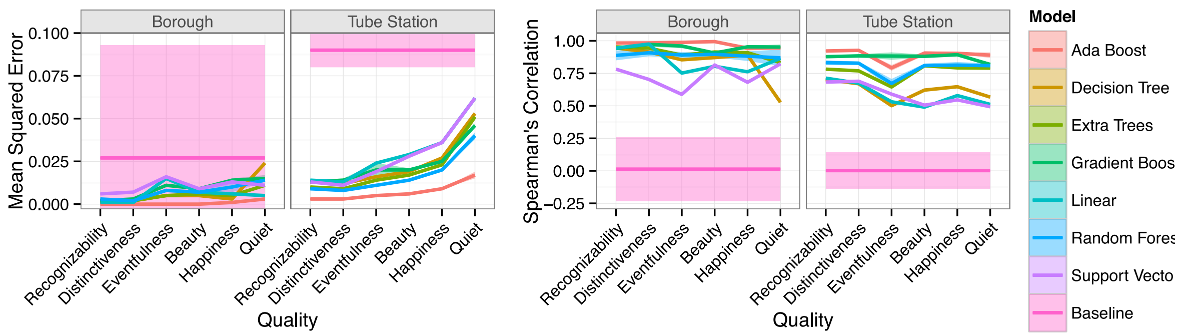


Figure 5.3: Mean Squared Error (left panel) and Spearman's Correlation ρ (right panel) for Area Rankings Produced by Seven Models plus Baseline. Each panel shows the results at both borough and subway station levels.

normalized measure of community (user) appeal of picture i used in Yildirim and Süsstrunk [2013]:

$$\text{appeal}_i = \frac{\text{number of favorites}_i}{\text{number of views}_i}$$

The higher a picture's ratio of number of favorites to number of views, the more the picture's views have led to user likes. Pictures with few views do not need to be filtered away as their presence does not affect the overall ranking: pictures with many favorites and views will still be highly ranked.

We use the appeal measure to produce lists of geo-referenced pictures. Each list orders areas in a different way (we will see how) and, for each area, top k pictures ordered by appeal are, in turn, shown. Given that pictures are always ordered by appeal, the desirability of such a list depends on the ordering of areas. We produce two lists with two distinct orderings. In the first, areas are ordered at random (*baseline list*). In the second list, areas are ordered by a predicted urban quality (e.g., *beauty list*) [5]. As a result, both lists contain pictures that Flickr users have liked, but the order of areas in one list differs from that in other list. As such, by comparing the two lists, one can establish whether the urban qualities are useful for ranking city pictures or not. If there is no difference between the ways the two lists fare, then either the urban quality of, say, beauty does not happen to promote appealing geo-reference photos or its predicted values do not accurately reflect beautiful areas.

To quantitatively ascertain whether each of those two lists return appealing content, we build a third one, which we call *ideal list*: in it, pictures are ordered by appeal without any consideration for the areas in which they were taken. The more similar the *beauty list* to the *ideal list*, the more the urban quality of beauty is able to promote

[5] We use an urban quality's predicted values and not the actual values to test to which extent our predictions are reasonable and whether they could be used in realistic scenarios.

pictures that users have liked on Flickr. To measure the similarity of the two lists, we, again, use Spearman's rank correlation ρ.

Figure 5.4 shows the results, which suggest two noteworthy considerations. The first is that the *baseline list* greatly differs from the *ideal list* (as the red line shows) and differs from the remaining lists related to our urban qualities (suggesting that the ordering of *areas* matters). The second consideration is that the working hypothesis behind our work holds true: ordering areas by a given urban quality tends to preferentially promote city pictures that are indeed appealing. The quality that most successfully promotes appealing content is that of beauty ($\rho = 0.69$), followed by recognizability ($\rho = 0.58$), eventfulness ($\rho = 0.53$) and distinctiveness ($\rho = 0.47$). These results are further confirmed by visually inspecting the set of pictures ranked by each urban quality (Appendix A).

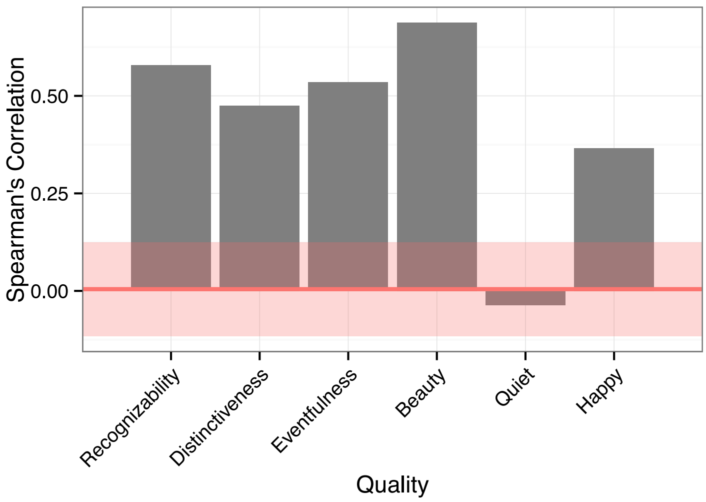


Figure 5.4: Similarity (Spearman's ρ) between the *Ideal List* and a List generated by one of our Urban Qualities. The similarity between baseline and the ideal list is shown in red with corresponding standard errors. For this barplot, the number of pictures per area is set to $k = 3$.

Figure 5.5 further shows that the Spearman correlation remains high as the user list of suggested pictures grows: suggesting five or even ten pictures in each area does not degrade the results at all. We also find that beautiful areas tend to be associated with appealing content, while quiet areas are not (the rank by quiet is comparable to the baseline). This might be because quiet areas either are not associated with appealing content or are difficult to predict out of the metadata we have used here. Perhaps, further investigation should go into enlarging the pool of metadata to include textual descriptors or even city-wide sound recordings [6].

[6] http://cs.everyaware.eu/event/widenoise

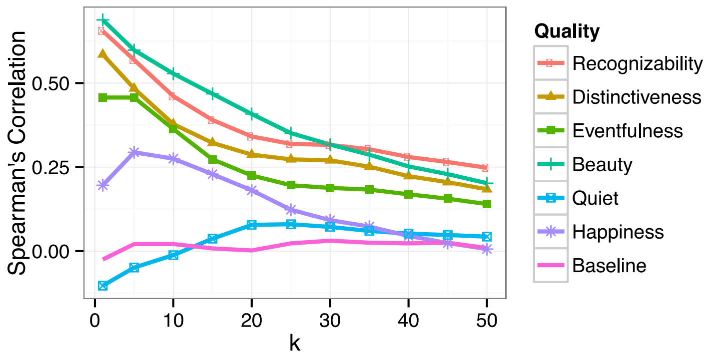


Figure 5.5: Similarity (Spearman's ρ) between the *Ideal List* and a List Generated by one of our Urban Qualities. The similarity varies with the number k of pictures per area (i.e., as the recommended list gets longer).

5.6 Contextual Factors

We now study how the predicted values of our urban qualities change depending on two contextual variables: time of day, and weather conditions.

To do so, in input of each of the models in the previous section, we give different features whose values change with the contextual variables. As we have mentioned in Section 5.4, this methodology is valid only if a model does not dramatically change with context. To test this assumption, we study whether the predictive accuracies of our models do not significantly change with time of day or weather, and we find this to be the case (Figure 5.6).

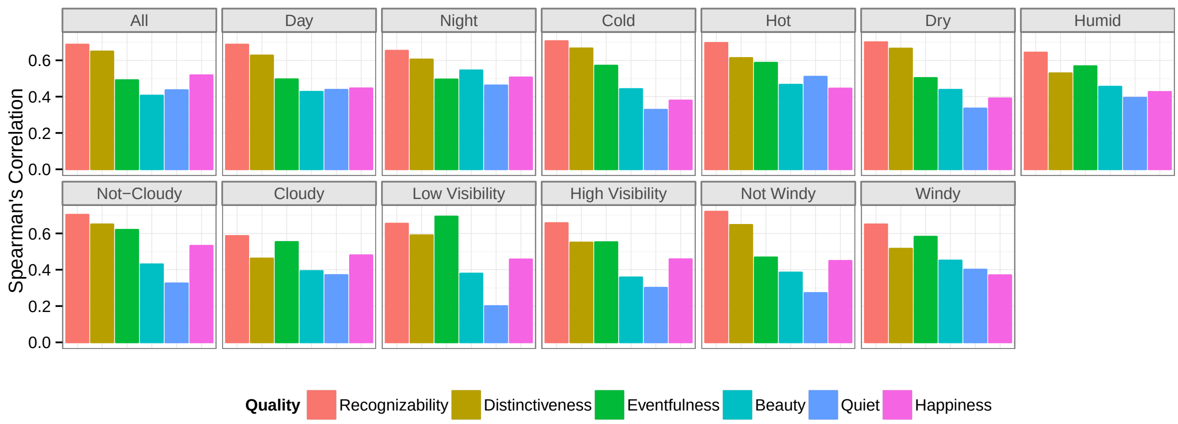


Figure 5.6: Accuracy of the Predicted Urban Qualities by Contextual Factors. Similarity (Spearman's ρ) between predicted and actual values for different contexts. The correlations do not significantly change for: day *vs.* night; cold *vs.* hot; dry *vs.* humid; not-cloudy *vs.* cloudy; low *vs.* high visibility; not-windy *vs.* windy.

5.6.1 Time of day

Using the definition of day *vs.* night in Section 5.2, Figure 5.7 shows the similarity (Spearman ρ) between the ideal list and a list generated by a given urban quality during different times of the day. The higher the similarity, the more the generated list contains appealing content. We find that beautiful areas tend to be associated with appealing content more during the day than during the night (the cerulean bar decreases from day to night). In a similar way, eventful areas are associated with appealing content during the day, which might reasonably suggest that people do not tend explore new parts of the city at night. Also, by visually inspecting the pictures in the top 5 most recognizable areas at day *vs.* those in the top 5 at night (Figure 5.8), one observes two distinctive sets of results, which speaks to the external validity of our approach.

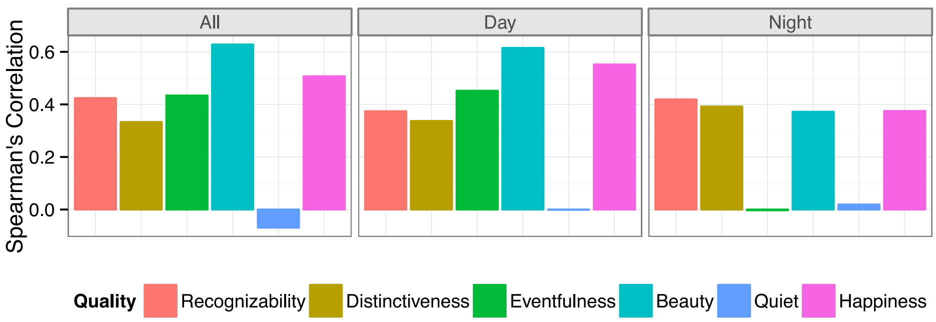


Figure 5.7: Rank Correlation (Spearman's ρ) Between the Ideal List and a List Generated by one of our Urban Qualities for Day *vs.* Night.

Figure 5.8: Pictures in Top 5 most Recognizable Areas at Day *vs.* Night.

5.6.2 Weather

For every day present in our weather dataset between 2002 and 2013, we discretize each of the five weather variables listed in Table 5.1 into lower class and upper class depending on whether their values are in the bottom or upper quartiles (Table 5.1 (b) shows the resulting thresholds). Depending on the weather condition of the day a picture was taken, we associate the five discretized values with the picture. For example, for a photo taken at 2007-06-09 17:05, its associated weather variables are: wind speed is $2knots$, air temperature is $24.7°C$, wet bulb temperature is $18.0°C$, cloud level is $6oktas$, and visibility is $12km$. That translates into associating the following discretized values with the picture: hot, humid, not-windy, low-visibility, and not-cloudy. Table 5.1 (c) shows the fraction of photos taken under different weather conditions: as one expects, photos are taken in non-cold and non-dry days; also, people tend to avoid cloudy days while preferring high visibility days.

	Air Temperature	Wet bulb temp	Wind speed	Cloud level	Visibility
(a) **Binary Discretization**					
Lower Condition	cold	dry	not-windy	not-cloudy	low-visibility
Upper Condition	hot	humid	windy	cloudy	high-visibility
(b) **Threshold Values**					
t_{lower}	7.2°C	5.8°C	$5.0knots$	$2.0oktas$	$12.0km$
t_{upper}	15.9°C	13.2°C	$11knots$	$8.0oktas$	$29.0km$
(c) **Distribution of Photos**					
$< t_{lower}$	16.4%	17.5%	24.2%	34.4%	18.9%
$> t_{upper}$	40.5%	36.0%	31.8%	26.8%	27.3%
In-between	43.1%	46.5%	44.0%	27.3%	53.8%

Table 5.1: Summary of the Binary Discretization of Five Weather Variables.

.

Figure 5.9 shows the similarity (Spearman ρ) between the ideal list and a list generated by a given urban quality under different weather conditions. We find that, with hot weather (which, in London, means a temperature above 16 degrees Celsius), any type of area (whether it is recognizable, distinctive, eventful, beautiful, or happy) is associated with appealing content. Dry and cold turn out to be the weather conditions that most negatively affect the production of appealing content. Again, ranking pictures during hot *vs.* cold days results in meaningful and inexpensive segmentations (Figure 5.10).

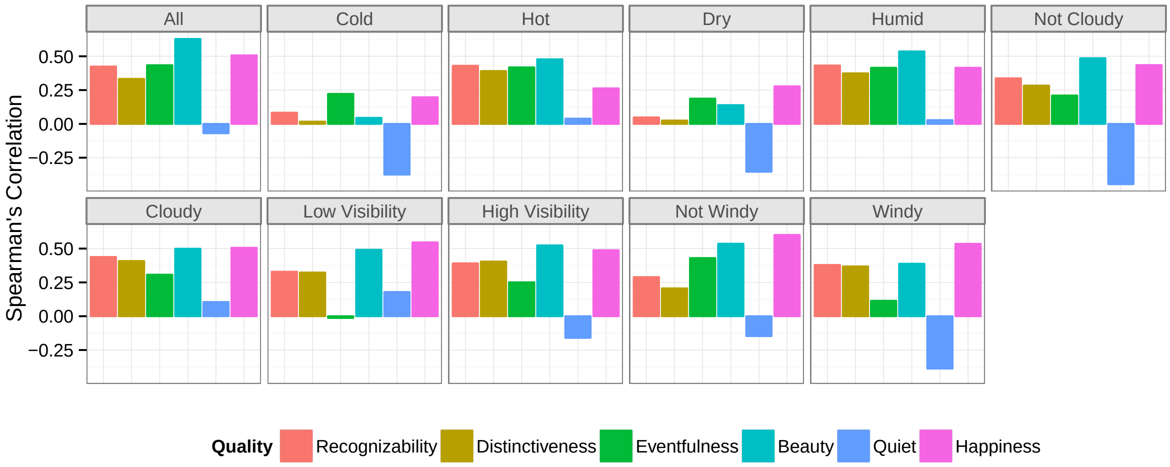


Figure 5.9: Rank Correlation (Spearman's ρ) Between the Ideal List and the Lists Generated by one of our Urban Qualities across Different Weather Conditions.

Figure 5.10: Pictures in the Top 5 Most Recognizable Areas During Hot *vs.* Cold Days.

Chapter 6

Discussion

6.1 Psychological Maps

Limitations. To increase response rate, we kept the survey as short as possible. It asks a minimum number of questions from which controlled variables are derived. However, this choice has drawbacks. For example, the survey asks for home location but does not ask for any other information about one's urban recognizability reach (the parts of the city one better knows visually). The problem is that one would know better (apart from the area one lives) also areas near work and on the way back home. We acknowledge this limitation but also stress that these differences are likely to cancel themselves out in a big sample like ours because of randomization. Also, some pictures might be more revealing than others. The game has two kinds of pictures. The fake pictures (excluded from the analysis) are meant to increase retention rate and, as such, are easy to recognize - they depict touristic locations or well-known stations. The real pictures (included in the analysis) are instead less informative as they have been vetted by us. However, they might still contain clues that make them recognizable.

6.1.1 Smart Cities Meet Web Science

The share of the world's population living in cities has recently surpassed 50 percent. By 2025, we will see another 1.2 billion people living in cities. The world is in the midst of an immense population shift from rural areas to cities, not least because urbanization is powered by the potential for enormous economic benefits. Those benefits will be only realized, however, if we are able to manage the increased complexity that comes with larger cities. The 'smart city' agenda is about the use of technological advances in physical and computing infrastructure to manage that complexity and create better cities. We will now discuss the ways in which this work suggests that the future of web scientists is charged with great potentials.

Planning urban interventions. We have shown that the relationship between recognizability and specific aspects of socio-economic deprivation is strong enough to identify boroughs suffering from high housing deprivation and poor living conditions, and also areas affected by crime. There is strong demand for making cities smarter, and the ability to identify areas in need could provide real-time information to, for example, local authorities. They could receive early warnings and identify areas of high deprivation quickly and at little cost, which is beneficial for cash-strapped city councils when planning renewal initiatives. However, before making any policy recommendation, recognizability data (based on a convenience sample) needs to be supplemented by other types of data - for example, by underground data (Lathia et al. [2012]; Smith et al. [2013]).

Making experiments on the web. By turning the execution of the experiment into a game, we have applied principles from games to a serious task and and have been consequently able to harness thousands of human brains. This might be fascinating to social science researchers, who must usually pay people to participate in their experiments. The game we have presented inverts that rule: players will happily fork out time for the privilege of being allowed to test their knowledge of London. Indeed, participants were rewarded with being able to test how well they knew London. One participant added: "Yesterday we had few friends over for dinner. I started to play the game on my laptop, and that escalated into a ridiculous competition among all of us that left my husband - the only Londoner in the room - quite injured, so to speak".

Rewarding schemes. We should design and test alternative engagement strategies. For now, we have focused on intrinsic (as opposed to extrinsic) rewards (Werbach [2012]). That is because recent psychological experiments (summarized in Werbach [2012]) have suggested that "intrinsic rewards (the enjoyment of a task for its own sake) are the best motivators, whereas extrinsic rewards, such as badges, levels, points or even in some circumstances money, can be counter-productive" (TheEconomist [2012]). In this vein, it might be beneficial to build a similar game on crowdsourcing platforms where participants are paid (e.g., on Mechanical Turk) and test how different reward schemes affect the externalization of the mental map. Finally, more research has to go into determining which incentives make engagement sustainable.

6.2 Ranking of Pictures in Photo-Sharing Websites

Limitations. This work is the first step towards using urban features to identify appealing geo-referenced content. In the future, research should go into combining all classes of features together. One simple way of doing so is to order each area's pictures depending on how appealing they are (appeal can be derived from visual features). The second limitation is that new ways of presenting pictures other than segmenting them by city neighborhoods (which are politically-defined and might be arbitrary at times) are in order: one could, for example, show pictures by areas that emerge from location-based data. Cranshaw *et al.* Cranshaw et al. [2012] used Foursquare data to draw dynamic boundaries in the city: what they called 'livehoods'. However, any work that uses location-based data (including ours) should account for the limitation of the data itself: the geographic distribution of Foursquare check-ins is biased Rost et al. [2013] (e.g., a user is likely to check-in more at restaurants than at home), and that can greatly affect the computation of our routine scores. Finally, given our promising results, it might be beneficial to further explore the use of urban features in cold-start situations, which are increasingly common.

Complementary to existing approaches. This work has to be considered complementary to existing approaches. By no means, it is meant to replace ranking solutions based on metadata or on visual features. Instead, all these solutions can be used together considering that they work under different conditions: whenever pictures come with rich metadata, then that metadata could be used to rank them; by contrast, in cold-start situations, our lightweight ranking combined with visual features might well be the only option at hand. We have shown that this option is viable as it offers good baseline performance. More generally, our results speak to the importance of incorporating cross-disciplinary findings. This work heavily borrows from 1970s urban studies and is best placed within an emerging area of Computer Science research, which is often called 'urban informatics'. Researchers in this area have been studying large-scale urban dynamics Crandall et al. [2009]; Cranshaw et al. [2012]; Noulas et al. [2012], and people's behavior when using location-based services such as Foursquare Bentley et al. [2012]; Cramer et al. [2011]; Lindqvist et al. [2011].

Chapter 7

Conclusion

In the sixties, scholars started to design experiments that captured the psychological representations that dwellers had of their cities. In mid-2012, we have translated their experimental setup into a 1-minute web game with a purpose, and have began with a deployment in London. We have gained insights into the differing perceptions of London that are held by not only Londoners but also people in UK and the rest of the world. The pre-eminence of Central London in the world's collective psychological map speaks to the popularity of its landmarks and touristic locations. The acquisition of a mental map is a slow process that does not necessarily come from direct experience but might be indirectly learned from, for example, atlases or movies. It comes as no surprise that Blackfriars, having being often used as a filming location, turned out to be the most socially/architecturally distinctive area - that is, an area whose recognizability is explained less by exposure to people and more by its distinctiveness. We have been able to quantitatively show the extent to which Londoners' collective psychological map tallies with the socio-economic indicators of housing deprivation, living environment conditions, and crime. By then comparing different social media platforms, we have suggested that a platform's demographics and geographic saliency determine whether its content is fit for urban studies similar to ours or not. This is a preliminary yet useful guideline for the web community who has recently turned to the study of large-scale urban dynamics derived from social media data.

We have used these results, along with other urban data, to assist the problem of automatic identification of appealing pictures, which has been often casted as a ranking problem. By contrast, we posited that, in a geo-enabled environment, the research roadmap should differ and revolve around the concept of neighborhood. Before this work, we did not know whether and, if so, how some of the 1970s theories in urban sociology could be practically used to identify appealing city pictures. We have shown that, upon theories proposed by Lynch, Milgram and Peterson, one is indeed able to do so. We hope that these results will encourage further work on multi-modal machine learning approaches that combine traditional (e.g., visual, textual, and social) features with domain-specific urban features.

Bibliography

Annechino, R. and Cheng, Y.-S. (2011). Visualizing mental maps of san francisco. *School of Information, UC Berkeley.*

Bawa-Cavia, A. (2011). Sensing the urban: Using location-based social network data in urban analysis. In *Pervasive Urban Applications (PURBA).*

Bentley, F., Cramer, H., Hamilton, W., and Basapur, S. (2012). Drawing the city: differing perceptions of the urban environment. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI).*

Cheng, Z., Caverlee, J., and Lee, K. (2011a). Exploring millions of footprints in location sharing services. In *Proceedings of the AAAI International Conference on Webblogs and Social Media (ICWSM).*

Cheng, Z., Caverlee, J., Lee, K., and Sui, D. Z. (2011b). Exploring millions of footprints in location sharing services. In *ICWSM.*

Cramer, H., Rost, M., and Holmquist, L. E. (2011). Performing a check-in: emerging practices, norms and 'conflicts' in location-sharing using foursquare. In *Proceedings of ACM International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI).*

Crandall, D. J., Backstrom, L., Huttenlocher, D., and Kleinberg, J. (2009). Mapping the world's photos. In *Proceedings of ACM International Conference on World Wide Web (WWW).*

Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In *International AAAI Conference on Weblogs and Social Media (ICWSM).*

Datta, R., Joshi, D., Li, J., and Wang, J. Z. (2006). Studying aesthetics in photographic images using a computational approach. In *Proceedings of the 9th European Conference on Computer Vision (ECCV).*

Glendinning, P. and Muthesius, S. (1994). *Tower Block: Modern Public Housing in England, Scotland, Wales, and Northern Ireland*. Yale University Press.

Ignite (2012). 2012 social network analysis report. In *Social Media*.

Lathia, N., Quercia, D., and Crowcroft, J. (2012). The hidden image of the city : Sensing community well-being from urban mobility. In *Proceedings of the International Conference on Pervasive Computing*.

Lindqvist, J., Cranshaw, J., Wiese, J., Hong, J., and Zimmerman, J. (2011). I'm the mayor of my house: examining why people use foursquare - a social-driven location sharing application. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI)*.

Lynch, K. (1960). *The Image of the City*. Urban Studies. MIT Press.

Martínez, P. and Santamaría, M. (2012). Atnight: Visions through data. *Mass Context*, pages 172–183.

Mclennan, D., Barnes, H., Noble, M., Davies, J., and Garratt, E. (2011). *The English Indices of Deprivation 2010*. UK Office for National Statistics.

Milgram, S. (1977). *The individual in a social world: essays and experiments*. Addison-Wesley.

Milgram, S. and Jodelet, D. (1976). Psychological maps of paris. *Environmental Psychology*.

Milgram, S., Kessler, S., and McKenna, W. (1972). A psychological map of new york city. *American Scientist*.

Nasar, J. L. (1994). Urban design aesthetics: The evaluative qualities of building exteriors. *Environment and Behavior*, 26(3):377.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., and Mascolo, C. (2012). A tale of many cities: Universal patterns in human urban mobility. *PLoS ONE*.

Peterson, G. L. (1967). A model of preference: Quantitative analysis of the perception of the visual appearance of residential neighborhoods. *Journal of Regional Science*, 7(1):19--31.

Quercia, D., Ohare, N., and Cramer, H. (2013). Aesthetic capital: What makes london look beautiful, quiet, and happy?

Redi, M. and Merialdo, B. (2012). Where is the beauty?: Retrieving appealing video scenes by learning flickr-based graded judgments. In *Proceedings of the 20th ACM Conference on Multimedia (MM).*

Rost, M., Barkhuus, L., Cramer, H., and Brown, B. (2013). Representation and communication: Challenges in interpreting large social media datasets. In *16th ACM Conference on Computer Supported Cooperative Work and Social Computing.*

Sigurbjörnsson, B. and van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th ACM Conference on World Wide Web (WWW).*

Smith, C., Quercia, D., and Capra, L. (2013). Finger on the pulse: Identifying deprivation using transit flow analysis. In *Proceedings of ACM International Conference on Computer-Supported Cooperative Work (CSCW).*

Taylor, N. (2009). Legibility and aesthetics in urban design. *Journal of Urban Design*, 14(2):189--202.

TfL (2011). London travel demand survey (ltds). *Transport for London.*

TheEconomist (2012). More than just a game: Video games are behind the latest fad in management).

van de Sande, K. E., Gevers, T., and Snoek, C. G. (2011). Empowering Visual Categorization With the GPU. *IEEE Transactions on Multimedia*, 13(1). ISSN 1520-9210.

van Zwol, R., Rae, A., and Garcia Pueyo, L. (2010). Prediction of favourite photos using social, visual, and textual signals. In *Proceedings of ACM Conference on Multimedia (MM).*

Von Ahn, L. and Dabbish, L. (2008). Designing games with a purpose. *Communications of ACM.*

Weber, R., Schnier, J., and Jacobsen, T. (2008). Aesthetics of streetscapes: Influence of fundamental properties on aesthetic judgments of urban space 1, 2. *Perceptual and motor skills*, 106(1):128--146.

Werbach, K. (2012). *For the Win: How Game Thinking Can Revolutionize Your Business.* Wharton Press.

Yildirim, G. and Süsstrunk, S. (2013). Rare is interesting: Connecting spatio-temporal behavior patterns with subjective image appeal. In *Proceedings of the 2nd ACM International Workshop on Geotagging and Its Applications in Multimedia (GeoMM)*.

Appendix A

Pictures Ranking

For a visual, more subjective analysis we show here the Flickr pictures Ranked by Urban Qualities plus Baseline (last row). As an example, in the first row, the most appealing pictures in the top (bottom) five most recognizable areas are shown.