Multiple Images of the City

Unveiling Group-Specific Urban Perceptions through a Crowdsourcing Game

David Candeia
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brazil
david.maia@ifpb.edu.br

Flavio Figueiredo
Universidade Federal de Minas Gerais
Belo Horizonte, Minas Gerais, Brazil
flaviovdf@dcc.ufmg.br

Nazareno Andrade
Universidade Federal de Campina Grande
Campina Grande, Paraíba, Brazil
nazareno@computacao.ufcg.edu.br

Daniele Quercia
Bell Labs
Cambridge, UK
daniele.quercia@gmail.com

ABSTRACT

Our perceptions of public spaces are central for our experience in the city. Understanding which factors shape this perception informs both urban planners, that aim at improving city life, as well as computational models that help us navigate in urban spaces. To understand cities at scale, crowdsourcing games have been employed successfully to evaluate citizens' opinions about cities and urban scenes. By analyzing human perceptions from residents of a mid-sized Brazilian city, this work brings three novel contributions. First, we consider theories from urban design to explore through crowdsourcing which high and low level features in an urban space are linked to perceptions of safety and pleasantness. Secondly, this paper leverages theory from urban sociology and anthropology to show how the sociodemographic profile of the citizens significantly mediate their perception of safeness and pleasantness of places. Finally, we show that features of the urban form proposed by urbanists can be combined with sociodemographics to improve the accuracy of machine learning models that predict which scene a person will find more safe or pleasant. This last result paves the road for more personalized recommendations in cold-start scenarios.

KEYWORDS

Urban Informatics; Urban Perception; Crowdsourcing

1 INTRODUCTION

In order to better understand the relationship between humans and urban environments, researchers have investigated urban perception through the years [3, 21, 24, 34], pointing that such perceptions influence our behavior, decisions and day to day lives [11, 16, 21] and how these perceptions vary across individuals [5, 8, 12, 17, 19, 36]. For instance, Lynch [21] demonstrates that different people depend on different urban elements (paths, edges, etc.) to guide themselves in cities. These past efforts, however, are usually unable to gather perceptions and perform studies at city scale, missing to reach large and diverse sociodemographic groups and understand the differences between these groups.

To offer broader insights at a large scale, crowdsourcing has been more recently employed as a tool to enlarge the participant pool [10, 18, 29, 31, 35]. One of the alternatives being used by crowdsourcing studies to gather perceptions is by designing web games that present urban scenes to the crowd [10, 29, 31, 35]. One of such studies [31] explores a method that translates preferences stated through pairwise comparison of urban scenes into an overall ranking, and other [29] investigates the influence of visual cues (e.g., image colors and texture) on the perception of urban scenes.

Nevertheless, two important factors were started to be investigated by [10, 35] and guide this paper. First, can we link the differences in the perception of how pleasant and how safe a urban scene is to high-level elements that are actionable by urban designers, such as the presence of trees, street width, or building height? Second, do people of different sociodemographic backgrounds (e.g., males versus females and/or people from different economic classes) differ in the ways they perceive the city [32]?

These two questions are considered in turn, with data collected through two crowdsourcing applications. The first is a game that captures perceptions about 108 urban scenes and sociodemographics of game players, based on previous work [10, 29, 31]. The second is a crowdsourcing task that makes use of human judgments to extract from urban scenes a set of high-level features that are linked with urban design (in this paper we will refer to them as urban design elements). Finally, our study is also conducted using scenes from a Brazilian mid-sized city, the city of Campina Grande, a city in a context which has received much less attention in the literature than economically more developed cities in the Northern hemisphere and, so, with markedly different characteristics.

In this context, our main results are:

- The analysis of perceptions collected in Campina Grande highlights similarities with previous work, such as green places being perceived as more pleasant;
- Our analysis points that urban design elements can help explain pleasantness and safety scores. Understanding such elements is interesting since they are measurable components that can be used to improve city life as more trees and fewer cars;

HT'17, July 4-7, 2017, Prague, Czech Republic.

DOI: http://dx.doi.org/10.1145/3078714.3078728

- The study on how urban perceptions differ across different sociodemographic groups points that while much scenes raise similar perceptions between participants, others raise perception differences. We've built a method that gathers collected perception, urban elements and sociodemographics to evaluate such differences. Gender, age and income were highlighted as important sociodemographic variables that mediate perception differences;
- We evaluate the novelty of urban elements combined with sociodemographic variables to predict scenes preferences in cold-start recommender settings. Classifiers achieved about 0.62 precision, around a 9.6% mean increase when compared to models that only explore urban elements.

Before continuing, in line with [10, 35] we point out that our work supports the possibility of evaluating divergences of different sociodemographic groups in terms of crowdsourced perception of city images. Also, our work provides evidences that these perception divergences are related to high-level urban elements. By uncovering these differences, we highlight the importance of residents background when results from crowdsourced games are used to motivate the decisions of urban planners.

2 RELATED WORK

The study of environment perception has been conducted over the years [3, 5, 10, 21, 31, 34, 35] in order to better understand the relation between humans and their environments. The psychological definition of perception [30] highlights that perception is built based on the individual, with his/her experiences, needs and emotions, on the thing being perceived and on the context. Over the years, evidences have pointed that the way we perceive our environments is influenced by our past experiences and way of life [3, 24] and by the aesthetics of different places [39]. Also, the way we perceive our environments influence our daily decisions [16, 21].

Regarding the effect of the individual, a large set of studies have investigated similarities [2, 10, 31, 35, 41] and differences [2, 5, 8, 10, 12, 17, 19, 32, 35, 36] in urban environment perceptions mediated by different sociodemographic aspects. Age [5, 19, 32, 35, 41], income [17, 19], gender [8, 12, 19, 32, 36], education levels [2, 5], social roles [41], hometown [35], nationality and culture [2, 12, 32, 36] are aspects that may influence urban perception.

Past studies were conducted using *small samples of* city images/photos [33, 34], personal interviews [32], as well as pen-and-paper questionnaires [19, 26]. Such methods used by previous researchers made it possible to understand perceptions of city environments and raised our knowledge on urban perceptions. However, such methods are mostly non-scalable when we consider medium to larger cities. In this direction, the recent development of computational technologies and crowdsourcing, combined with quantitative analysis, have created the opportunity to repeat and continue perception studies in a scalable way, at lower cost, higher speed and with larger pools of participants [10, 18, 20, 29, 31, 35]. With such tools there is the opportunity to capture perceptions using web games [10, 29, 31, 35] or even analyzing data from social media [18, 20]. This development has also created the opportunity to train machine learning models on perception data [9, 25] to learn urban preferences and then predict preferences for places not yet studied, producing perception data even faster.

Recent studies aiming at larger scale data collection developed the Place Pulse [31], UrbanGems [29] and Street Seen [10] applications. Such applications pick geotagged images from different cities (e.g., New York, London), present pairs of images and ask perception questions such as "Which place looks safer?". Place Pulse [31] demonstrates the possibility of creating a quantitative measure, named Q-Score, that translates user votes on preferred urban scenes to a score of urban perception. UrbanGems [29] investigates the influence of visual cues (i.e., image colors, texture and visual words) on perception. Evans and Akar [10] relates cyclability perceptions with features of the urban form. Also, authors in [35] evaluate safety perception variations according to characteristics of people present in images (e.g., gender, ethnicity and facing). All such studies demonstrate the potential of crowdsourcing applications to capture urban perceptions, but only Evans and Akar [10] and Traunmueller, Marshall and Capra [35] focused on investigating, respectively, cyclability and safety perceptions through different groups of people. This is the focus of this study.

As a further step, the PlacePulse team has investigated the use of PlacePulse data [9, 25] to train machine learning models and predict urban preferences, with promising results. Such studies [9, 25] have investigated the use of low-level features of images, such as color histograms, to predict scenes preferences. We investigate if using high-level features, urban design elements that are closer to urbanists, combined with participants profile can achieve good prediction results. Also, considering our focus on a smaller city of Brazil, a less studied and underdeveloped country, we also deal with a different culture and habits when compared to such studies.

3 METHODS

To investigate how Campina Grande is perceived, we collected data in three steps: (1) Crowdsourcing the perceptions of urban scenes; (2) Crowdsourcing the extraction of urban design elements from the scenes; and (3) Extracting color patterns from the same scenes.

3.1 Crowdsourcing Urban Perceptions

We collected perceptions of participants who live in the city of Campina Grande, Brazil. The collection was enabled by the creation of a crowdsourcing application called "*Como é Campina?*" (How is Campina like?, in Portuguese).

Building a crowdsourcing platform. Our platform is based on other platforms such as UrbanGems [29], Place Pulse [31] and StreetSeen [10]. Each participant is asked to compare urban scenes randomly selected from a pool of scenes extracted from Google Street View. Each time, the user is asked to select two scenes (Figure 1): the most and least pleasant (or safe) scenes. The participant can also choose a "Can't tell" option. Each 4-image comparison provides information about six pairwise comparisons: the most pleasant (safest) scene outperforms the other three, which, in turn, outperform the least pleasant (safe) scene. After ten 4-image comparisons, the user is asked to answer sociodemographic questions about their age, gender, income, education level and marital status.

We evaluate scenes based on safety and pleasantness. This is aligned with previous studies that explored urban security [25, 26,

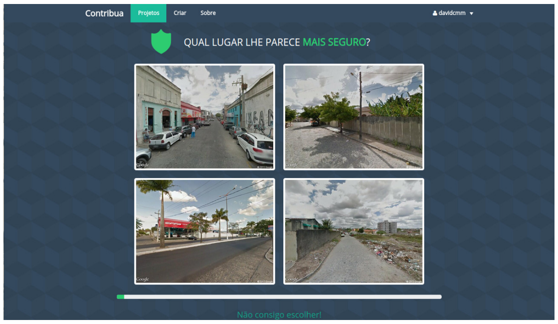


Figure 1: *Como é Campina?* crowdsourcing application.

33] and beauty [29, 31, 33]. The main difference in the design of our application compared to previous work [10, 29, 31] is the use of this 4-image comparison instead of pairwise comparisons. The use of four images provides considerable more data per task, and is inspired by MaxDiff [23] design, common in marketing studies, and known to have a comfortable cognitive load for participants[1].

Collecting urban scenes. Campina Grande is a city of 593.026 km^2 in the northeast of Brazil, with 49 neighborhoods and 385, 213 inhabitants [15]. To allow for the study of diverse socioeconomic conditions and land uses, we focus on three diverse neighborhoods: *Catolé* is residential and high-income; *Liberdade* is residential, has mostly low-/medium-income inhabitants, and a diverse land use, including commerce and residences; and *downtown* is mostly commercial and highly iconic. For each neighborhood, we selected two census areas (defined by the Brazilian Institute of Spatial Geography) such that one area would include main streets and commercial places, and the other area would include a residential portion of the neighborhood. In each area, we then select 10 random geographic points (latitude and longitude) that are at least 50 meters apart from each other. For each latitude and longitude point, we collect Google Street views in four headings (i.e., 0°, 90°, 180° and 270°), so as to create a comprehensive view of each place. After filtering the scenes that do not allow for a view of the public space (e.g., obstructed by walls and buildings), we were left with 108 urban scenes. It is important to note that using Google Street views helps making meaningful scenes comparison, as its methodology for data collection controls for weather and time of the day in order to publish images of a city under similar conditions and quality.

Recruiting participants. After publicizing our crowdsourcing platform, 304 participants answered a total of more than 5, 400 4-image comparisons, resulting in votes for a total of 32, 723 pairwise comparisons of scenes. The platform was advertised using Facebook campaigns focused on residents of Campina Grande. At times, the campaign was targeted to ensure the participation of less represented groups such as older women. The encouragement for participants to contribute aimed at tapping into intrinsic motivations, highlighting, for example, the enjoyment of a task and improvement of available information about the city.

[1]We have validated the 4-way comparison by comparing the Kendall correlation of the Q-Scores produced with this strategy with Q-Scores obtained from another version of our application that uses pairwise comparisons only. The correlation of two waves of participants using only pairwise comparisons and between participants using pairwise and 4-way comparison was very similar, and always greater than 0.78

Comparing urban scenes. Using scenes comparisons data, we analyze preferences for scenes in two ways. First, we follow the same steps as Salesses, Schechtner and Hidalgo [31] to derive Q-Scores for the scenes, which allow us to identify those evaluated as most and least pleasant or safe. Next, we consider the comparisons where participants preferred one scene over others to analyze what urban elements are associated with preferred scenes and whether sociodemographic variables moderate such preferences.

The Q-Score metric works as follows [31]. Each urban scene goes through a series of pairwise disputes extracted from a 4-way comparison. Each dispute has three possible outcomes: the scene is selected as the best and wins the dispute; it is selected as the worst and loses the dispute; or there is a draw. For each scene i, its win (W) and loss (L) ratios are computed, and then i's Q-Score is function of all the other scenes' (against which i has been compared) win and loss ratios. More formally, we computed i's win (W) and loss (L) ratios for each of our two questions q (one about safety and the other about pleasantness):

$$W_{i,q} = \frac{w_{i,q}}{w_{i,q} + l_{i,q} + t_{i,q}}; L_{i,q} = \frac{l_{i,q}}{w_{i,q} + l_{i,q} + t_{i,q}}$$

where $w_{i,q}$ is the number of times scene i won a dispute, l is the number of times the scene lost a dispute, and t is the number of times that the scene ended in a draw. Then, we compute i's Q-Score:

$$Q_{i,q} = \frac{10}{3}(W_{i,q} + \frac{1}{n_i^w}\sum_{j1=1}^{n_i^w} W_{j1,q} - \frac{1}{n_i^l}\sum_{j2=1}^{n_i^l} L_{j2,q} + 1)$$

where n_i^w is the total number of scenes i was preferred over, n_i^l is the total number of scenes i was not preferred over. As a result, a value in the range $[0, 10]$ is obtained, where 10 represents the best evaluation of a scene (i.e., the scene was preferred over scenes that were also preferred in their disputes) and 0 represents the worst evaluation of a scene (i.e., the scene was not preferred over scenes that were also not preferred in their disputes). This way of scoring urban scenes circumvents the need to perform all possible pairwise comparisons of scenes. Experimentally, it has been shown that each scene needs to go through $22 - 32$ disputes to obtain a stable Q-Score [31]. Each of our scenes went through 25 disputes, and each dispute had answers from at least 3 application users of diverse sociodemographic conditions.

3.2 Crowdsourcing Urban Elements

Our second data source aims at informing us about high-level urban design elements in the scenes, such as street width or number of trees in the scene. Such elements were extracted from the *Como é Campina?* scenes using the CrowdFlower[2] crowdsourcing platform. To identify an expressive and comprehensive list of visible urban elements, we resorted to the most authoritative study in the field. In their "Measuring Urban Design" book, Ewing and Clemente [11] summarized decades of research on the identification of desirable urban design qualities, which include imageability, complexity, and human scale, and their relation to walkability and urban elements. As walkable places may relate to pleasant and safe places, we devised a series of human computation tasks to extract

[2]http://www.crowdflower.com/

31 relevant urban elements from our scenes, based on [11]: **mean building height**; **street width**; **number of trees**; **number of moving cars; number of parked cars; number of moving cyclists; number of buildings with identifiers; number of different buildings**; **number of people on streets**; **presence of graffiti**; **maintenance**, debris and pavement conditions, which are grades from 1 to 5; sidewalk width; number of major landscape features; number of non-rectangular buildings; number of basic building colors; number of accent building colors; number of long sight lines; presence of outside dining; number of public lights; number of street furniture; number of courtyards/plazas/parks; number of pieces of public art; number of small planters; presence of buildings with different ages; proportion of sky; proportion of street wall; proportion of active use buildings; proportion of street-level facade that is covered by windows; proportion of historic buildings. To control for accuracy and reliability of our estimates of such elements, two approaches were used:

i) throughout the execution of tasks, a worker is presented to test questions already answered by authors, and worker's answers accuracy is calculated from these test questions. A reliable worker has to achieve at least a 70% accuracy and only answers of reliable workers are considered;

ii) Krippendorff's alpha coefficients were calculated to filter urban elements whose estimations presented agreement rates of at least 0.6. Other four elements (a grade, street width, presence of graffiti and number of buildings), were also considered in spite of lower agreement rates, following a manual inspection of estimates that revealed reasonable values.

After controlling for accuracy and reliability we were able to extract estimates for the 11 elements highlighted in bold above.

3.3 Extracting Image Colors

In addition to extracting visual elements related to urban planning, we are also interested in investigating the relation of colors with pleasantness and safety perceptions. To do so, we extracted color information from our scenes in a way similar to Quercia, O'Hare and Cramer [29]. Each scene was associated with its average RGB triplet (r,g,b) and 64-bin color histograms. The OpenCV[3] library was used to calculate all color features.

4 OVERALL PERCEPTIONS

Our first analysis uses the same method as previous work [10, 29, 31] to evaluate the overall perception of our participants. We then link this perception with high-level urban design elements.

As one expects, perceptions of safety and pleasantness are related (Figure 2), but do not match perfectly. The corresponding linear regression ($pleasantness = 0.81 + 0.83 \cdot safety + error$) results in an R^2 of 0.55. Our values are somewhat larger than some values found in previous studies. Previous work found $R^2 = 0.35$ between perceptions of safety and uniqueness [31], $R^2 = 0.37$ between uniqueness and economic-class [31], and $r = 0.64$ between beauty and happiness [29]. Although this small increase may simply be a consequence of the different questions asked, we also conjecture that it may stem from the reality of some Brazilian cities. Segregated areas tend to be both those with less urban development [4] as well as more criminality [28, 38]. Crowdsourced urban perceptions can thus be used by urban planners to tackle and understand segregation issues.

[3] http://opencv.org/

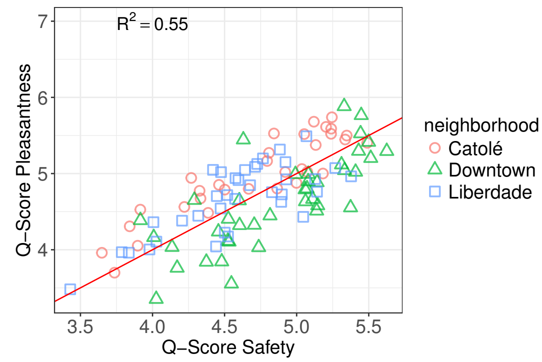
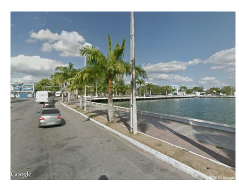


Figure 2: Q-Scores: Pleasantness versus Safety.

4.1 Best and Worst Scenes Evaluations

To make evaluations concrete, we inspect in Figure 3 the 3 most and least pleasant urban scenes. The most pleasant places are well-maintained and have vegetation, while the least pleasant present physical disorder (e.g., dirty, wastelands, lack of maintenance). Figure 4 shows the 3 safest and least safe urban scenes. The safest places are well-maintained and have people, while the least safe show signs of physical disorder. Those findings are in line with previous work in which pleasantness (or beauty) has been related to greenery [2, 29, 33] and well-maintained places [2, 33], and in which safety has been related to physical order [26, 32, 33, 40, 41].



(a) Q-Score = 5.88

(b) Q-Score = 5.76

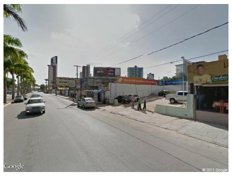
(c) Q-Score = 5.73

(d) Q-Score = 3.55

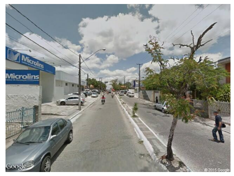
(e) Q-Score = 3.48

(f) Q-Score = 3.35

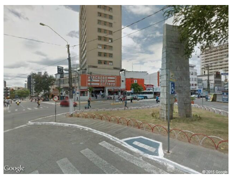
Figure 3: The three most and least pleasant urban scenes.



(a) Q-Score = 5.62



(b) Q-Score = 5.51



(c) Q-Score = 5.49

(d) Q-Score = 3.73

(e) Q-Score = 3.64

(f) Q-Score = 3.42

Figure 4: The three safest and least safe urban scenes.

4.2 Perceptions and Colors

To then determine which colors tend to be present in safe and pleasant urban scenes, we computed linear regressions between safety (pleasantness) Q-Scores and RGB values of colors. To compute a percent importance of each color [29], we divide each color's β coefficient by the sum of the absolute values of all the β coefficients. Very much in line with previous work [29], pleasantness was associated with less red (−40.4%) and more green (48.8%) in scenes. Safety was also associated with less red (−53.9%). Similar results are found when regressing safety (pleasantness) with the color histograms: pleasantness was associated with variations of green and red, and safety with variations of yellow.

The confirmation of previous findings on a Brazilian city serves as evidence that the relation of colors with pleasantness (and also safety to some extent) is universal. Though outside of the scope of our study, we notice that novel datasets [9] that cover perceptions of multiples cities can be used to further validate these findings.

4.3 Perceptions and Urban Design Elements

We now turn to investigate the relation between the crowdsourced perceptions of safety and pleasantness in each scene and urban elements extracted through CrowdFlower. A similar first effort was done by Evans and Akar [10] by relating urban elements and cyclability perception. Different from colors, urban elements capture simple and measurable entities (e.g., more trees or less cars) that can be explored to improve city life for urban dwellers. Our results here can thus be viewed from a practical perspective on how to improve urban settings. More importantly, our findings in this section will be complemented by a study on how sociodemographic variables moderate the perceptions of some urban elements in Section 5.

To perform our analysis, a new dependent variable called **rank evaluation** was created for this task by first applying a rank transformation to the scenes according to their Q-Scores for pleasantness and safety, and then inverting the sign of such ranks. Such transformations make linear models more readily applicable, and result in a configuration where the best evaluated scene has the highest value of the dependent variable.

Table 1 shows regression models relating **rank evaluation** for safety and pleasantness, urban elements and the neighborhood where each picture was taken. Positive coefficients indicate that a better image ranking is positively related to the predictor value.

As expected, there is a significant positive relation between pleasantness and trees (vegetation) and good maintenance condition in a scene. Also, there is a positive relation between perception of safety and the presence of more people (number of people, number of parked cars) and good maintenance conditions. The importance of people on streets and good maintenance for walkability was demonstrated by Ewing and Clemente [11], and walkable places may relate with more pleasant and safe places. As pointed before, the relation of greenery [2, 29, 33] and well-maintained places [2, 33] with beauty/pleasantness, and the link of visual signs of physical order [26, 32, 33, 40, 41] with perception of safety were demonstrated. Regarding safety, the positive relation with proximity to human elements in parks [33] was also highlighted and may indicate that presence of cars may be interpreted as presence of people. Such results and ours supports the intuition that places where there are eyes on the street are perceived as safer.

Also, a few unexpected results were found. First, such high values for R^2 are not expected, since perceptions deal with past experiences and personal background, and our models consider only objective features from images of places. Second, indicators of people presence are associated with safe scenes but not necessarily with pleasant scenes. Finally, the presence of graffiti is not strongly associated with pleasantness or safety [8, 33]. This may happen because graffiti is typically present in places with poor maintenance conditions in our sample, and so the effect of graffiti may be indistinguishable in this data.

The positive evaluation of cars in terms of safety may point a divergent perception of Brazilian culture compared to developed countries. Brazil has a markedly car-oriented culture, and the car is seen as an object associated with wealth, social status, greater sense of security and avoidance of poor public transport systems [7]. Also, there is a predominant culture in government city planning that typically focuses on enlarging streets for cars instead of investing in public transportation systems or pedestrian streets [37].

5 GROUP DIFFERENCES IN URBAN PERCEPTION

After evaluating the overall perception of the city, we now investigate perception differences. Initially we note that from the set of 304 participants who contributed in our experiment, 211 of them (69%) answered the sociodemographic questions. Considering the groups with largest numbers of participants, we divided them in terms of: gender, age, and income. The percentage of participants in each of these groups are: i) 66.89% men and 33.11% women; ii) 45.89% young (below 25 years), and 54.11% adult (25 – 62 years);

Table 1: Linear regression models of scene ranking for pleasantness and safety controlling for neighborhood. β and standard error values are shown.

	Pleas.	Saf.
(Intercept)	$-39.40^{\cdot}$	-56.48^{**}
	(19.99)	(20.09)
Street width (ft)	0.00	0.07
	(0.38)	(0.38)
Number of moving cars	−1.24	2.68
	(1.74)	(1.75)
Number of parked cars	1.01	1.98^{*}
	(0.76)	(0.77)
Number of moving cyclists	1.39	−0.90
	(4.06)	(4.08)
Maintenance condition	23.38^{***}	20.74^{***}
	(3.46)	(3.48)
Number of building with identifiers	$-1.84^{\cdot}$	−1.35
	(0.99)	(0.99)
Number of trees	2.66^{*}	1.53
	(1.17)	(1.18)
Log of mean building height (feet)	0.62	2.96
	(3.37)	(3.39)
Number of people	0.00	2.50^{*}
	(1.14)	(1.14)
Image is in downtown	0.72	8.26
	(7.57)	(7.61)
Image is in Catolé	$11.41^{\cdot}$	7.77
	(6.18)	(6.21)
Number of different buildings	1.00	2.01
	(1.51)	(1.52)
Presence of graffiti	−8.14	$-15.51^{\cdot}$
	(8.39)	(8.43)
R^2	0.51	0.51
Adj. R^2	0.45	0.44

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

iii) 45.05% low income (classes E and D according to the Brazilian census authority, IBGE), and 54.95% high income (classes A, B and C in the same classification).

First, we examine whether there are striking differences in rankings created from preferences of different groups. For that, we calculate Q-Scores for the scenes according only to answers from each sociodemographic group. Overall, there is a high correlation in the ranking of scenes as evaluated by the different groups. In general, we found Kendall correlation values ranging from 0.60 (in the case of Low vs High income for safety) to 0.72 (in the case of Men vs Women for pleasantness).

Given that the overall perception of the city is similar across groups, we now turn to pinpoint the situations where differences occur and to examine which urban elements are associated with such differences. All 11 urban elements extracted reliably from Crowdflower are considered in this analysis since they are related to walkability. Also, although only some of them were significantly related to pleasantness and safety overall ranking of scenes (Table 1), different elements may be related to specific groups differences.

5.1 Group-Specific Perceptions and Urban Design Elements

We now focus on determining the effect of sociodemographic variables and urban elements on safety and pleasantness perceptions. To achieve our goal, we focus on situations where there are a marked preference for one scene over another. That is, we resort to pairwise disputes instead of evaluating rankings. This approach allows us to more accurately model and isolate multiple moderation effects, as building a ranking implies in aggregating scenes preferences from people with diverse backgrounds without controlling for the inherent variation that can be attributed to the different individuals. Thus, we can analyze better the relation of the visual content of scenes and participants backgrounds using the pairwise comparisons.

More specifically, we first filter the pairwise scenes disputes (scene A x scene B) to consider only disputes in which our participants indicated a preference (they picked either A or B as the most pleasant/safe scene). In other words, we removed draws from our disputes. For each selected dispute we compute the differences in each urban element between both scenes. For example, we calculate the number of trees for scene A minus the quantity of trees for scene B. As a result, there is for each dispute a preferred scene A or B, the differences in urban elements between A and B, and the sociodemographic profile of the participant. It is important to note that, while the decision of which picture is A or B is arbitrary (we choose so randomly and our findings do not change with multiple executions of our experiments), our evaluations will be interpreted in relation to the difference between urban elements from A to B, being A the reference scene, as we now describe.

Our models consider the selected scene as our dependent variable (i.e., a binary variable indicating that selecting A over B is the positive class, the opposite is the negative class), urban elements differences (i.e., elements in A minus those in B) as independent variables, and age, gender and income as moderators. We model the sociodemographic group as moderators in order to capture how a certain group, say men, potentially moderates (or perceives) an element such as trees. This moderation is used to compare the perception of trees between men and women. Finally, we also consider the neighborhoods of each image as explanatory variables. Here, we encode the neighborhood of both scenes in the comparison as a single indicator variable that captures the different pairs of *distinct neighborhoods*. With this variable we can investigate if there are inherent preferences of one neighborhood over another.

We resort to logistic regression models [14] to investigate the impact of sociodemographic variables on choosing A over B. We shall validate these results with more advanced machine learning models later on. With this consideration, two logistic regression models were built, one for pleasantness and another for safety.

Positive coefficients (β) in the logistic regression models will indicate that higher number of urban elements (e.g., trees) in scene A (our reference scene) will bias participants to preferring A over B. That is, if A has more trees, for instance, and β is positive, the model indicates that an increase in trees will lead to a higher chance of preferring A. In contrast, negative coefficients indicate that the higher number of urban elements in scene A will bias to choosing

Table 2: Logistic regression of the preference towards scene A for pleasantness. Relevant predictors are sorted according to their coefficients, which are each in the scale of the independent variable. Neighborhood comparison coefficients and intercepts are shown in the bottom rows of the table.

Term	β	std. error
1. Maintenance diff.	0.51^{***}	0.05
2. Cyclists diff. x adult	-0.11^{*}	0.05
3. Maintenance diff. x men	0.10^{*}	0.04
4. Maintenance diff. x high inc.	0.08^{*}	0.04
5. Moving cars diff. x high inc.	-0.08^{***}	0.01
6. People diff. x adult	0.05^{***}	0.01
7. Trees diff. x adult	0.04^{**}	0.01
8. Trees diff.	0.04^{**}	0.01
9. Buildings identifiers diff. x adult	-0.02^{*}	0.01
10. Parked cars diff.	0.02^{*}	0.01
11. Street width diff.	0.01^{*}	0.00
12. Street width diff x adult	-0.01^{*}	0.00
(Intercept)	-0.18^{***}	0.03
Catolé x Downtown	0.71^{***}	0.06
Catolé x Liberdade	0.52^{***}	0.06
Liberdade x Downtown	0.17^{**}	0.06
Num. groups: participant 282		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 3: Logistic regression of the preference towards scene A for safety. Relevant predictors are sorted according to their coefficients, which are each in the scale of the independent variable. Neighborhood comparison coefficients and intercepts are shown in the bottom rows of the table.

Term	β	std. error
1. Maintenance condition	0.39^{***}	0.04
2. Moving cars diff.	0.20^{***}	0.02
3. Maintenance diff. x men	0.13^{**}	0.04
4. Maintenance diff. x high inc.	0.11^{**}	0.04
5. Moving cars diff. x men	-0.08^{***}	0.02
6. Buildings diff. x adult	0.08^{***}	0.01
7. Moving cars diff. x adult	-0.06^{**}	0.02
8. Buildings diff. x high inc.	-0.05^{***}	0.01
9. People diff. x high inc.	0.05^{***}	0.01
10. People diff.	0.05^{***}	0.01
11. Trees diff.	0.04^{**}	0.01
12. Parked cars diff.	0.04^{***}	0.01
13. Trees diff. x high inc.	0.03^{*}	0.01
14. Trees diff. x adult	-0.03^{*}	0.01
15. Street width diff. x adult	0.01^{***}	0.00
(Intercept)	-0.15^{***}	0.03
Catolé x Downtown	0.36^{***}	0.06
Downtown x Catolé	0.34^{***}	0.07
Liberdade x Catolé	-0.16^{**}	0.06
Downtown x Liberdade	0.16^{*}	0.06
Num. groups: participants 279		

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

image B as their preferred image. That is, a positive coefficient indicates a preference towards an urban element, whereas a negative one represent a rejection.

It is also important to point out that we present our results considering a random effects approach related to each participant. We account for the fact that participants provided answers to multiple comparisons by adding the participant as a random effect in the models. In our setting, random effects will relate to the expected variability across participants. This random effect contributed to significantly varying the intercept for pleasantness ($p < .001$) and was marginally significant for safety ($p < .1$). Finally, results without random effects lead to similar qualitative findings but less accurate models (greater AIC and BIC scores [14])[4]. The logistic regression models considering random effects are shown in Tables 2 and 3, which we now discuss. For clarity, our tables only present coefficients deemed statistically significant ($p < .05$ at least) and moderations are represented with an x symbol indicating the preferences of a certain group (i.e., men) for a certain urban element in comparison to the other corresponding group (i.e., women).

Some of our previous results are confirmed by analyzing non-moderated predictors in Tables 2 and 3. Better maintenance condition and amount of trees lead to scenes being preferred for pleasantness. Moreover, better maintenance, greater numbers of parked cars and people bias the safety perception (also positively). In addition to such predictors, some neighborhoods comparison, in contrast to places in the same neighborhood, were also highlighted as important predictors. Wider streets also impact pleasantness and more moving cars and trees impact safety. These last new effects are expected according to the culture of cars previously discussed, as well as the idea that wider streets may be related with more comfortable streets and that a positive relation between vegetation and safety was already highlighted [22].

Regarding groups differences, it's important to highlight that the majority of significant predictors, and 7 of the top 10 predictors (i.e., greater β values) in our models were moderated by age, gender or income. Figure 5 summarizes the moderations presented on the tables. Age was the most influential sociodemographic variable perceived for pleasantness in our data, while age and income were equally important for safety and gender was the least important one for both pleasantness and safety. These moderations increase or decrease the impact of urban elements in scenes evaluation for some groups, and in some cases turn some elements as relevant. For example, the number of people in streets was not relevant for pleasantness in general, but for adults (line 6 of Table 2) this element was important to prefer a scene as more pleasant. The same can be observed for the number of moving cyclists (line 2 of Table 2). Maintenance condition, on the other hand, is a relevant element in both pleasantness and safety general evaluation, but it's effect is increased for men and high income groups.

We can also evaluate the coefficients in Tables 2 and 3 in light of the "divide by 4" rule [14]. This rule indicates that dividing the β by 4 we have an estimate of the maximum difference in probability corresponding to a unit difference in the predictor. For example, considering Table 2 we can point that more trees (line 8) are associated with a greater chance of a scene being preferred, and one more tree increases this probability by a maximum of 1%. In Table 3 we can point that more moving cars (line 2) are associated

[4]AIC and BIC capture the trade-off between accurate (log likelihood) and complex models. Lower values indicate a good accuracy with fewer complexity [14].

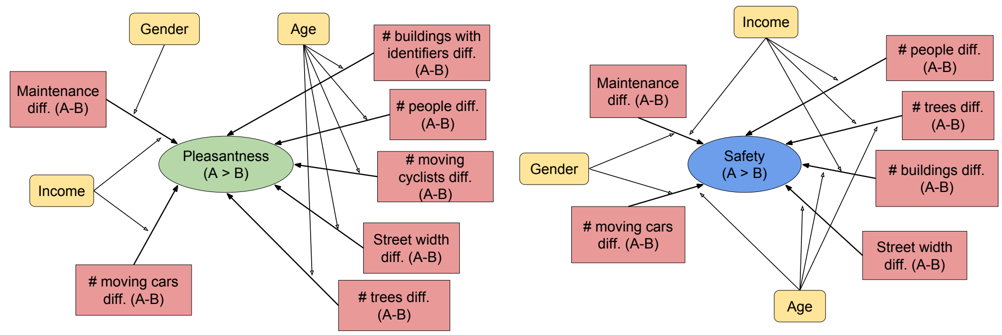


Figure 5: Significant moderations of predictors in logistic models. Moderators are shown in yellow.

with a greater chance of a scene being preferred, and one more car increases this probability by a maximum of 5%.

It is important to note that, given that moderator variables are binary (e.g., 1 will indicate men while 0 will indicate women), to interpret the effect of a sociodemographic group the β value of the urban element must be added to the value of the same element moderated by a group. In addition to general preference of more trees (including young people) for pleasantness (line 8 of Table 2), adults prefer trees even more (positive β in the moderation, line 7). So, one more tree increases the adult probability of preferring a scene in 1% more than young people by maximum. Another additive example is that, for safety (Table 3), better maintenance is preferred by the whole group of participants (line 1). In addition to this preference, men place a higher value to maintenance condition of places than women (line 3). A greater number of moving cars is preferred for safety by the whole group of participants (line 2), but men do not prefer more cars as women (line 5).

It's also important to notice that some moderations are present both for pleasantness and safety, this is expected since pleasantness and safety scores were related (Section 4). Maintenance condition is moderated by gender and income, while number of trees and street width are moderated by age for both perceptions. While gender and income influence both perceptions in the same way, age impacts both perceptions differently. Regarding gender and income, better maintained places are generally preferred for pleasantness and safety, but men and high income people prefer even more well maintained places. Regarding age, trees are generally preferred for pleasantness and safety, but adults evaluate places with more trees as even more pleasant than young people do, but not safe. Wider streets are generally preferred for pleasant places and its effect was not significant for safety perception, however, adults do not prefer wider streets as young people do for pleasantness and for safety the relation is exactly the opposite, with adults preferring wider streets more than young people.

Finally, analyzing significant coefficients, moderated or not, and their association with urban qualities [11] we can point Human Scale and Complexity as the most important urban qualities for both pleasantness and safety. This importance is expected since these qualities describe, respectively, proportions and visual richness of spaces, which can be related to pleasantness, and they are also related to intimidation and pedestrians stimulation, which can be related to safety. Although the qualities are the same, the most impacting elements associated with these qualities are number of trees for pleasantness and number of moving cars for safety.

5.2 Predicting Scene Preferences

Once relevant effects of user profiles are established through logistic models, our next step is an experiment to evaluate the predictive capability of using urban elements and participant profile information to predict scenes preferences.

Our experiment here is motivated by a recommendation scenario. For example, on a navigational system a recommender can predict pairwise preferences of urban places to create rankings of places for a user to visit or pass through. Motivated by this setting, we look into how sociodemographic variables can be explored in cold-start settings [1]. In a cold-start setting, previous evaluations of participants are not available. Once these evaluations are present, classic approaches as collaborative filtering [1] may be able to capture user preferences. When not present, the cold-start case, classifiers will usually rely on other user features as the sociodemographic groups.

To represent the cold-start, we remove each participant from the training dataset in a turn (and we also remove the scenes evaluated by him/her). This single participant's answers become our test set. We trained multiple classifiers with the votes of other participants and then tested the prediction on the removed participant. By removing the participant and every pair with a scene s/he evaluated, we filter out the possibility of the classifier learning users' latent preferences or factors related to individual scenes. So, the classifier will only explore sociodemographics and urban elements.

We evaluated four classifiers using the scikit-learn framework [27]: KNN, RBF SVM, Naive Bayes and Extra Trees. For each pair of scenes, classifiers were trained considering the urban elements (of each scene in the pair) as real numbers and sociodemographic information as indicator variables. In order to tune classifiers, a grid-search on hyperparameters, in a 3-fold cross validation, was performed by further breaking the training set into training and validation. Overall, Extra Trees led to the best results and we focus our analysis on this classifier in isolation (our goal is not to compare different classifiers).

We resort to accuracy, precision, recall and F1 scores to gain a better understanding of the classifier ability to perform correct predictions in both positive and negative classes. In Figure 6 we show such metrics for a classifier that consider sociodemographic variables (participant profile) and for other that explores only urban elements. By comparing both, we unveil the impact of sociodemographic variables in cold-start settings. The figure presents the average of each score (one per left-out user) and corresponding 95% confidence intervals, and, also, a random classifier for comparison.

Firstly, the use of urban elements and participant profile improves scores in comparison to a random classifier (our baseline model), with mean gains of 22.9% for pleasantness scores and 18.9% for safety scores. Then, we compare the use of both participant and urban elements information to the scenario in which only urban elements are used. In these scenarios, the mean accuracy gains were about of 7.2% and 8.2% for pleasantness and safety, respectively. For precision, the mean gains were of 10.2% and 8.9%, respectively.

The gains in precision and accuracy further validate our previous find that the sociodemographic background of participants impact their perceptions of urban scenes. Statistically speaking (i.e., considering the confidence intervals), it is also important to point that using only urban elements information was not sufficient to improve recall values. Nevertheless, there are small gains in average. This last result stems from the fact that many other social and cognitive factors, as well as other urban elements, impact participants perceptions. Because of this, our classifiers will not correctly predict every possible evaluation, we only explore a small number of complex human background features. Investigating other factors that impact perceptions is left as future work.

6 IMPLICATIONS

Our results lead to implications for researchers, practitioners, and for theory. First, and in line with [10], some of our results link citizens' pleasantness and safety perceptions to urban elements that can be used by urban planners to understand what interventions may be applied to areas perceived as less pleasant or safe. Moreover, using crowdsourcing, this evaluation can be done at scale.

A second important aspect is the comparison of urban elements that are most related to safety and pleasantness in Campina Grande and in previous work. On the one hand, this examination contributes to a body of results that can lead us to understand invariants affecting how people perceive safety and pleasantness in the city. On the other hand, the fact that moving and/or parked cars are positively associated with these perceptions in Campina Grande calls for further study of this phenomenon in Brazil.

Our main result with implications for both theory and practice is the significant moderation that age, income and – to a lesser extent – gender exert on how different groups perceive urban elements in terms of pleasantness and safety perceptions. This results calls for attention to planning the city for diversity. Moreover, unveiling which elements lead to discordance seems to us as a promising pointer for future work that leverages this discordance for productive debates about the city. It is also relevant that these sociodemographic characteristics can be used to improve the accuracy of predictive models that help one navigate in the city. Future work providing recommendation methods for more pleasant routes or routes perceived as safer may take these results into account.

Finally, this work provides to methodological contributions for researchers. First, the design of our crowdsourcing game shows that it is feasible to compare scenes four at a time, instead of the commonly used pairwise comparisons [10, 29, 31]. This design speeds up data production, and should be considered by researchers. Second, our method for modeling the effect of moderators on pairwise scene preferences may be relevant for further research, as it provides an intuitive and tractable framework for the analysis considering both the effects of urban elements in individual scenes, and the multilevel modeling of random effects in participant preferences.

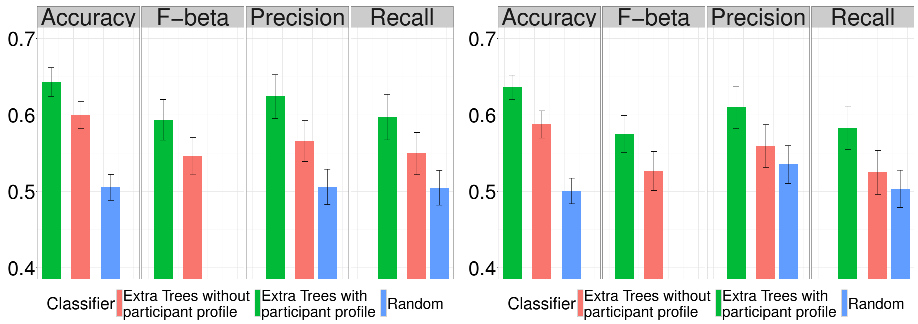
7 LIMITATIONS

The way we captured urban perceptions is partly biased for five main reasons. First, perceptions depend on our participants' past experiences, and we have no systematic way of capturing those experiences. However, the higher the number of participants, the more randomized such effects become. Second, we have elicited perceptions from images. But a place's image does not fully capture the place - for example, how it smells, sounds, and changes over time. Third, the perceptions of a place might drastically change over the course of a day. This study has investigated relations between urban design elements and scenes preferences captured from crowdsourcing solutions. For finer-grained analyses and the development of a robust recommendation system, future work should capture additional information to further stratify perceptions across, for example, times of the day. Also, as opposed to other images of urban scenes, Google Street views tend to control for factors such as time of the day, presence of people, and weather conditions. So, using images from other sources that help to capture such variations is necessary. Fourth, it is unclear whether our findings generalize to other cities. To ascertain that, again, additional data has to be collected. Nevertheless, the methodology presented in this paper can be readily applied to different cities for further comparisons if need be. Finally, although considering residents of Campina Grande may incur in biases of recognizing places of the city, residents know the particularities of their city, how it works, how safe and pleasant places are like. We try to minimize individual bias by capturing data from at least 3 participants for each 4-image comparison task.

8 CONCLUSION

Considering the growth of urban population, local authorities need to evaluate and implement solutions to manage the complexity, problems and expectations that come with larger cities in order to improve their citizens well-being. This work contributes towards this goal in the emerging theme of urban informatics [13], in particular associated with social computing and urban dynamics [6].

Our aim has been to test whether urban perceptions of safety and pleasantness change across different "classes" of people. In order to do so we developed a crowdsourcing web game based on [10, 29, 31] to gather urban scenes perception of residents of the city of Campina Grande, Brazil. We compared the overall perception of our participants with previous crowdsourcing [29, 31] and urbanism works in order to validate our findings. Then, we evaluated collected perceptions considering different sociodemographic groups (age, gender and income) and found that different groups perceived about 60% to 72% of scenes in similar ways, but other scenes raised perception differences. A few urban elements were related with perception differences being mediated by income, age and gender.



(a) Pleasantness **(b) Safety**

Figure 6: Confidence intervals for classifiers Accuracy, Precision, Recall and F-beta scores

The maintenance condition is an important one since it contributes to explain differences in pleasantness and safety perceptions.

ACKNOWLEDGMENTS

This research is supported by EU-BR BigSea project (MCTI/RNP 3rd Coordinated Call).

REFERENCES

[1] Charu C Aggarwal. 2016. *Recommender Systems: The Textbook*. Springer.
[2] Tarik M Al-Soliman. 1990. The impact of the surrounding environment on people's perception of major urban environmental attributes. *Journal of King Saud University* (1990).
[3] Donald Appleyard. 1973. *Notes on urban perception and knowledge.*
[4] Florine Bos and Rivke Jaffe. 2015. A rejuvenated approach to urban development and inequality: Young people's perceptions and experiences in Rio de Janeiro. *Habitat International* 48 (2015), 106–112.
[5] Xinyu Jason Cao. 2015. How does neighborhood design affect life satisfaction? Evidence from Twin Cities. *Travel Behaviour and Society* (2015).
[6] Justin Cranshaw, Raz Schwartz, Jason I Hong, and Norman Sadeh. 2012. The livehoods project: Utilizing social media to understand the dynamics of a city. In *International AAAI Conference on Weblogs and Social Media*. 58.
[7] Roberto DaMatta. 2012. *Fé em Deus e pé na tábua: ou como e por que o trânsito enlouquece no Brasil.* Editora Rocco.
[8] Kristen Day, Cheryl Stump, and Daisy Carreon. 2003. Confrontation and loss of control: Masculinity and men's fear in public space. *Journal of Environmental Psychology* 23, 3 (2003), 311–322.
[9] Abhimanyu Dubey, Nikhil Naik, Devi Parikh, Ramesh Raskar, and César A Hidalgo. 2016. Deep learning the city: Quantifying urban perception at a global scale. In *European Conference on Computer Vision*. Springer, 196–212.
[10] Jennifer S Evans-Cowley and Gulsah Akar. 2014. Streetseen: factors influencing the desirability of a street for bicycling. In *Forthcoming in 93rd Annual Meeting of the Transportation Research Board.*
[11] Reid Ewing and Otto Clemente. 2013. *Measuring urban design: Metrics for livable places.* Island Press.
[12] S Farrell, J Bannister, J Ditton, and E Gilchrist. 2000. Social psychology and the fear of crime: Re-examining a speculative model. *British Journal of Criminology* 40 (2000), 399–413.
[13] Marcus Foth, Jaz Hee-jeong Choi, and Christine Satchell. 2011. Urban informatics. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 1–8.
[14] Andrew Gelman and Jennifer Hill. 2006. *Data analysis using regression and multilevel/hierarchical models.* Cambridge University Press.
[15] Instituto Brasileiro de Geografia e Estatística IBGE. 2011. Sinopse do censo demográfico 2010. (2011).
[16] William H Ittelson, Harold M Proshansky, Leanne G Rivlin, and Gary H Winkel. 1974. *An introduction to environmental psychology.* Holt, Rinehart & Winston.
[17] Haydea Izazola, Carolina Martínez, and Catherine Marquette. 1998. Environmental perceptions, social class and demographic change in Mexico City: a comparative approach. *Environment and Urbanization* 10, 1 (1998), 107–118.
[18] Andrew Jenkins, Arie Croitoru, Andrew T Crooks, and Anthony Stefanidis. 2016. Crowdsourcing a Collective Sense of Place. *PloS one* 11, 4 (2016), e0152932.
[19] CY Jim and Xizhang Shan. 2013. Socioeconomic effect on perception of urban green spaces in Guangzhou, China. *Cities* 31 (2013), 123–131.
[20] Matthew James Kelley. 2013. The emergent urban imaginaries of geosocial media. *GeoJournal* 78, 1 (2013), 181–203.
[21] Kevin Lynch. 1960. *The image of the city.* Vol. 11. MIT press.
[22] Jolanda Maas, Peter Spreeuwenberg, Marijke Van Winsum-Westra, Robert A Verheij, Sjerp de Vries, and Peter P Groenewegen. 2009. Is green space in the living environment associated with people's feelings of social safety? *Environment and planning. A* 41, 7 (2009), 1763.
[23] Anthony AJ Marley and Jordan J Louviere. 2005. Some probabilistic models of best, worst, and best–worst choices. *Journal of Mathematical Psychology* 49, 6 (2005), 464–480.
[24] Stanley Milgram. 1974. The experience of living in cities. *Crowding and behavior* 167 (1974), 41.
[25] Naren Naik, Jade Philipoom, Ramesh Raskar, and Cesar Hidalgo. 2014. Streetscore–Predicting the Perceived Safety of One Million Streetscapes. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on.* IEEE, 793–799.
[26] Segun Okunola and Dolapo Amole. 2012. Perception of safety, social participation and vulnerability in an urban neighbourhood, Lagos, Nigeria. *Procedia-Social and Behavioral Sciences* 35 (2012), 505–513.
[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
[28] Jennifer Peirce. 2008. Divided cities: crime and inequality in urban Brazil. *Paterson Review* 9 (2008), 85–98.
[29] Daniele Quercia, Neil Keith O'Hare, and Henriette Cramer. 2014. Aesthetic capital: what makes london look beautiful, quiet, and happy?. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 945–955.
[30] Alan Saks and Gary Johns. 2011. Perception, attribution, and judgment of others. *Organizational Behaviour: Understanding and Managing Life at Work* 7 (2011), 72–114.
[31] Philip Salesses, Katja Schechtner, and César A Hidalgo. 2013. The collaborative image of the city: mapping the inequality of urban perception. *PloS one* 8, 7 (2013), e68400.
[32] Robert J. Sampson and Stephen W. Raudenbush. 2004. Seeing Disorder: Neighborhood Stigma and the Social Construction of Broken Windows. *Social Psychology Quarterly* 67, 4 (2004).
[33] Herbert W Schroeder, LM Anderson, et al. 1984. Perception of personal safety in urban recreation sites. *Journal of Leisure Research* 16, 2 (1984), 178–194.
[34] WC Sullivan. 1994. Perceptions of the rural-urban fringe: citizen preferences for natural and developed settings. *Landscape and Urban Planning* 29, 2 (1994), 85–101.
[35] Martin Traunmueller, Paul Marshall, and Licia Capra. 2015. Crowdsourcing Safety Perceptions of People: Opportunities and Limitations. In *International Conference on Social Informatics*. Springer, 120–135.
[36] Adri Van der Wurff, Leendert Van Staalduinen, and Peter Stringer. 1989. Fear of crime in residential environments: Testing a social psychological model. *The Journal of social psychology* 129, 2 (1989), 141–160.
[37] Eduardo Alcântara Vasconcellos. 2014. *Urban Transport Environment and Equity: The case for developing countries.* Routledge.
[38] Julio Jacobo Waiselfiz. 2013. Mapa da Violência 2013: mortes matadas por armas de fogo. (2013).
[39] Ralf Weber, Jörg Schnier, and Thomas Jacobsen. 2008. Aesthetics of streetscapes: Influence of fundamental properties on aesthetic judgments of urban space 1, 2. *Perceptual and motor skills* 106, 1 (2008), 128–146.
[40] James Q Wilson and George L Kelling. 1982. Broken windows. *Atlantic monthly* 249, 3 (1982), 29–38.
[41] Sue-Ming Yang and Chih-Chao Pao. 2015. Do We "See" the Same Thing? An Experimental Look into the Black Box of Disorder Perception. *Journal of Research in Crime and Delinquency* 52, 4 (2015), 534–566.