

# Explaining the “Why” of Urban Perception at Scale: A Context-Sensitive Generative AI Framework

Tim Guangyu Wu, ChengHe Guan

NYU Wagner | NYU Shanghai

Oct 2025

# Opportunity to Predict Urban Perception at Scale

Deep Learning the City: Quantifying Urban Perception at a Global Scale — Dubey et al. 2016

Which place looks **more beautiful** ?

For this question: 174,046 clicks collected      Goal: 500,000 clicks

SEE REAL-TIME RANKINGS

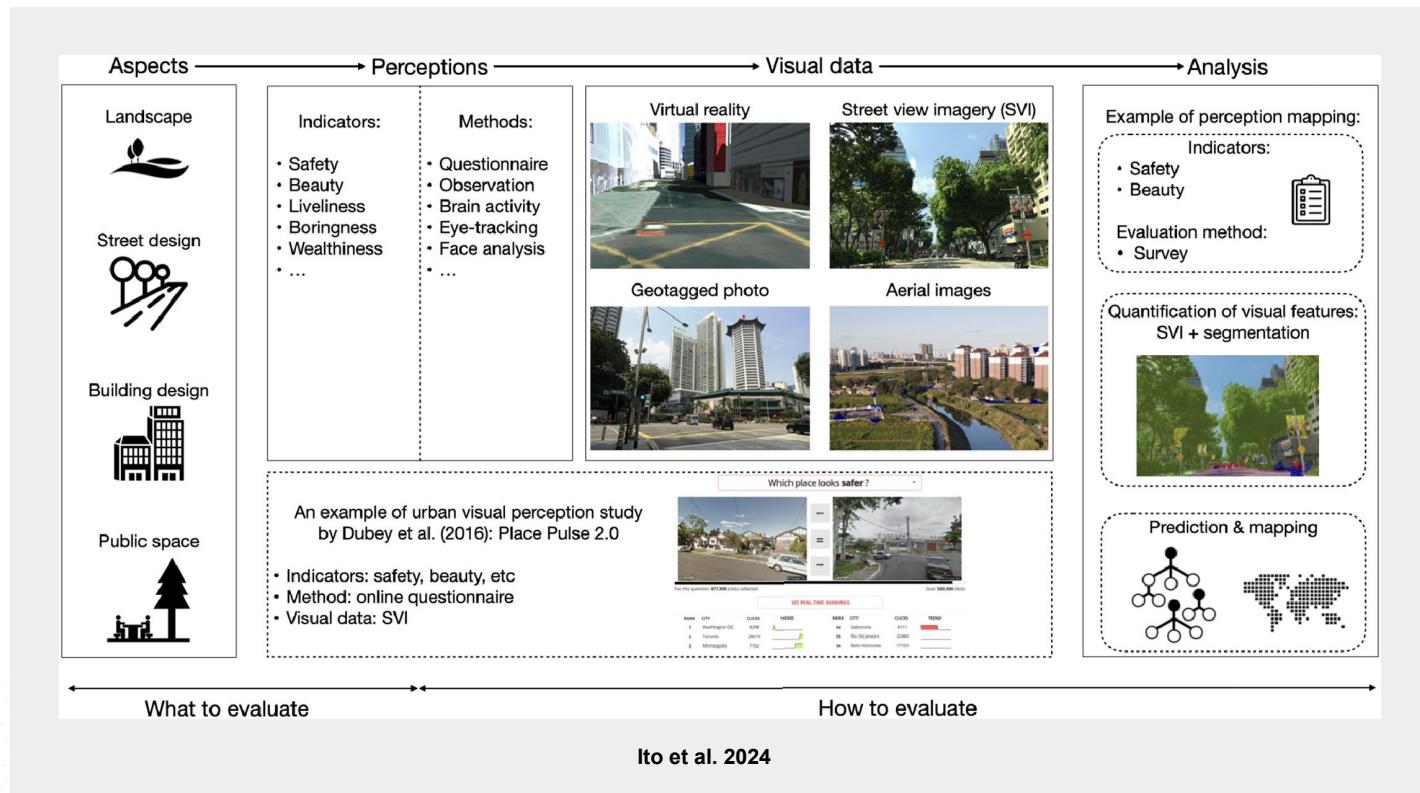
RANK	CITY	CLICKS	TREND	RANK	CITY	CLICKS	TREND
1	Atlanta	12638		54	Sao Paulo	9453	
2	Washington DC	3045		55	Belo Horizonte	6107	
3	Singapore	8439		56	Rio De Janeiro	11475	

(a) Statistics on Images		(b) Statistics on Pairwise Comparisons (PC)		
Continent	#Cities	#Images	Question	#PC
Asia	7	11,342	Safe	370,134
Africa	3	5,069	Lively	268,494
Australia	2	6,082	Beautiful	166,823
Europe	22	38,636	Wealthy	137,688
North America	15	33,691	Depressing	114,755
South America	7	16,168	Boring	111,184
Total	56	110,988	Total	1,169,078
				16.73

Dubey et al. 2016

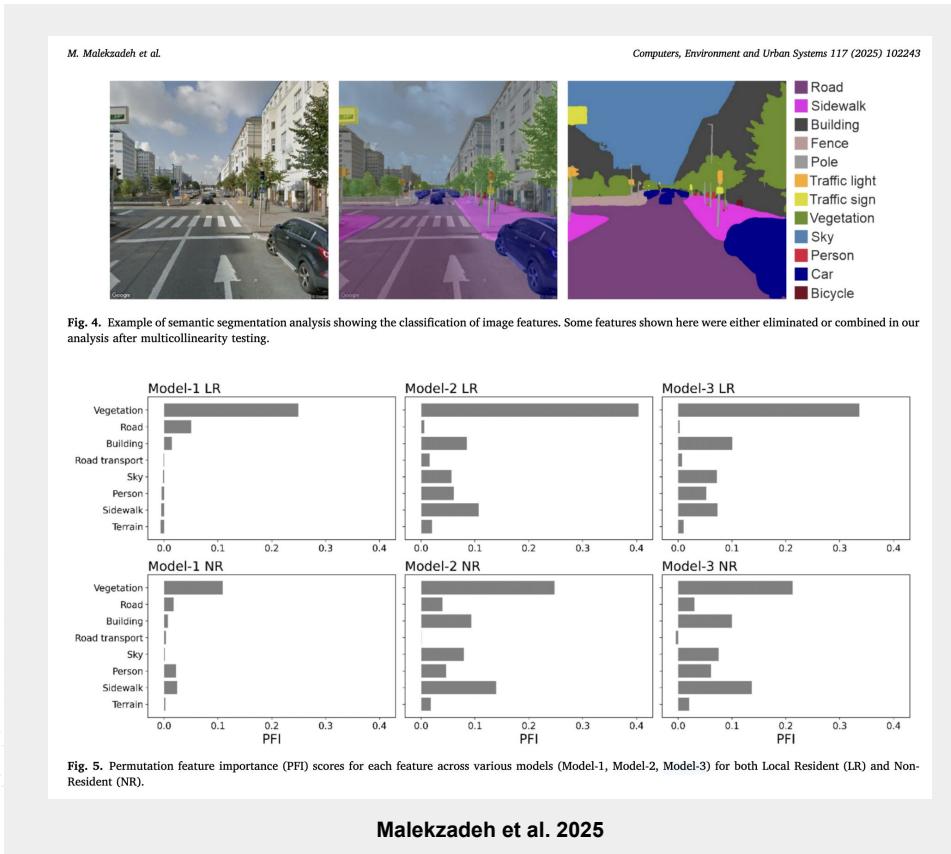
## Mainstreaming of Perception Research and Need for Explainability

Understanding urban perception with visual data: A systematic review — Ito et al. 2024



# Semantic Segmentation Approach Limited by Broad Preset Categories

Urban attractiveness according to ChatGPT: Contrasting AI and human insights — Malekzadeh et al. 2025



# Large Language Model Approach Biased by Prompt/Protocol Design

Criteria used in each ChatGPT query prompt with their main relevant references.		
Prompt	Criteria	References
Model-1 (Prompt 1)	Overall attractiveness	
Model-2 (Prompt 2)	<b>Enduring Physical Features</b> - Sidewalk Features for Pedestrian Activity - Street Design for Traffic and Activity - Tree Canopy and Greenery - Physical Indicators of Human Activity - Permanent Lighting	<a href="#">(Clifton et al., 2007; Ewing &amp; Handy, 2009; Pilkora et al., 2002; Wimbardana et al., 2018)</a>
Model-3 (Prompt 3)	<b>Enduring Physical Features</b> - Sidewalk Features for Pedestrian Activity - Street Design for Traffic and Activity - Tree Canopy and Greenery - Physical Indicators of Human Activity - Permanent Lighting <b>Urban Design Qualities</b> - Imageability - Legibility - Enclosure - Human Scale - Transparency - Linkage - Complexity - Coherence <b>Subjective Reactions</b>	<a href="#">(Clifton et al., 2007; Ewing &amp; Handy, 2009; Pilkora et al., 2002; Wimbardana et al., 2018)</a> <a href="#">(Alexander, 1977; Arnold, 1980; Craig et al., 2002; Evans et al., 1982; Ewing &amp; Handy, 2009; Gehl, 1987; Jacobs, 1993; Lynch, 1960; Moudon &amp; Lee, 2003; Rapoport, 2013)</a> <a href="#">(Birenboim, 2018; Ewing &amp; Handy, 2009; Zhang, Zheng, &amp; Wang, 2022)</a>

Malekzadeh et al. 2025 - Comput. Environ. Urban Syst.

- 1) Mystery: Is this an interesting scene to explore further?
- 2) Complexity: Does the scene contain diverse elements and features?
- 3) Coherence: Do the visual elements of the scene fit together well?
- 4) Legibility: Is it easy to find your way around the environment in the scene?
- 5) Preference: Overall, do you like this scene?

Tung et al. 2025 - Landscape and Urban Planning

perception. For example, when asking about overall walkability, we wrote the following prompt: "How do you rate the Walkability of this environment for both photos from 1–10? Based on the photo you rated higher, why do you think it is more Walkable? (Please explain your opinion in at least 20 words). Overall Walkability: This measures the ease and appeal of walking around the area shown in the image.

Wedyan et al. 2025 - Plos One

You are an urban environment expert. Here is the definition of safe and dangerous for city scenes: {C}. Now help me to compare the two input images and tell me which one is safer. Give me a choice from A: First Image or B: Second Image. C: Unable to compare. You also need to briefly explain your choice.

Zhang et al. 2025 - Cities

# Research Questions

- 1. How do fine-grained perceptual differences within broad visual categories shape human perception of urban environments?**
  
- 2. What novel perceptual features will emerge when explanations are produced without constraints from predefined evaluation protocols?**
  
- 3. How does the perceptual influence of these features vary across cities, revealing global heterogeneity in urban perception?**

# Methodology and Workflow

## Image Perception Data Preparation

- MIT Place Pulse 2.0 dataset
- Removing duplicate images (based on location and visual similarity)
- Applying minimum vote threshold for reliability

## Stage 1: Explanation Generation

- Prompting the model with image and target perception dimension
- Describing perceived experience of the scene
- Generating free-form explanatory reasons
- Generating an overall perception score

## Stage 2: Perceptual Unit Synthesis

- Extracting main features from generated reasons
- Parsing features into base elements and descriptive qualifiers
- Developing a taxonomy of perceptual units
- Filtering units based on minimum occurrences

## Quantitative Estimation

- Linking perceptual units to perception scores
- Estimating each unit's contribution to perception scores using value backpropagation
- Re-estimating unit contributions using city-specific subsets

## Validation and Sensitivity Check

- Manual review of sampled explanations for plausibility
- Validating feature presence using open-vocabulary segmentation
- Conducting sensitivity analyses on threshold choices

**110,988 images** from 56 cities across 28 countries

**3.47 million reasons** and 426 thousand experiences

**3066 perceptual units** with  $\geq 100$  occurrences

**Contribution estimations** and global heterogeneity

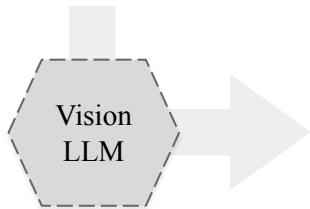
**Accuracy metrics** for scene feature presence

# Stage 1

Output of Stage 1



Perceptual Dimension: ***Beautiful***



## Experience Description

- ❖ Standing on this charming street, one is immediately enveloped by a sense of quaint urban beauty. The narrow, gently curving cobblestone road invites a leisurely stroll ...

Input to Stage 2

## Reasons that enhance ***Beauty***

- ❖ The mature trees lining the street provide natural beauty and shade, enhancing the visual appeal and comfort of the space.
- ❖ The canal with bicycles parked alongside adds a picturesque and quintessentially local charm, contributing to the scene's uniqueness.
- ❖ The historic brick buildings with their detailed facades and large windows offer architectural interest and a sense of place.
- ❖ The cobblestone street surface adds texture and a traditional aesthetic that complements the overall historic ambiance.
- ❖ The presence of people walking and cycling adds life and a sense of community.

## Reasons that detract from ***Beauty***

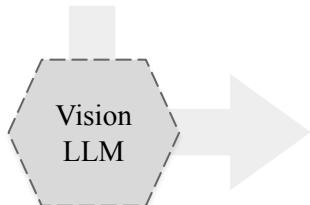
- ❖ The street is somewhat narrow and enclosed, which might limit expansive views and a sense of openness that some might associate with beauty.
- ❖ The presence of parked bicycles and some street clutter, while adding character, might be seen as visual distractions by some viewers.

Predicted ***Beauty Score*** 80

# Stage 1



Perceptual Dimension: *Lively*



## Output of Stage 1

### Experience Description

- ❖ Standing on this bustling city street, you are immediately immersed in a vibrant urban atmosphere. The wide sidewalk is crowded with people, many gathered around ...

### Input to Stage 2

#### Reasons that enhance *Liveliness*

- ❖ The dense crowd of pedestrians gathered around storefronts and street vendors indicates high human activity and social interaction.
- ❖ Visible commercial establishments with open displays and signage suggest active business and consumer engagement.
- ❖ The compact arrangement of buildings and narrow sidewalks create a sense of enclosure that intensifies the feeling of busyness.
- ❖ Bright daylight encourages outdoor presence, contributing to the street's vibrancy.

#### Reasons that detracts from *Liveliness*

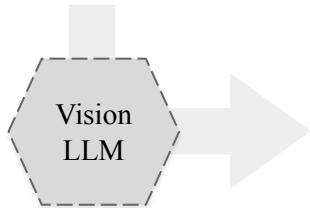
- ❖ The buildings appear somewhat aged and worn, which might detract from a sense of modern vibrancy or renewal.
- ❖ The street, while busy, does not show signs of festive or celebratory activity that might elevate liveliness to a higher level.
- ❖ The traffic is orderly and not chaotic, which might reduce the perception of energetic hustle.

Predicted *Liveliness Score* 75

# Stage 1



Perceptual Dimension: *Safe*



## Output of Stage 1

### Experience Description

- ❖ Standing on this street, the environment feels somewhat mixed in terms of safety. The road is moderately busy with a few vehicles, including a motorcycle and ...

### Input to Stage 2

#### Reasons that enhance Safety

- ❖ Presence of vehicles and a motorcyclist indicates some level of activity and human presence, which can deter crime and increase safety.
- ❖ Daylight conditions provide good visibility, reducing the risk of hidden dangers.
- ❖ Trees and greenery contribute to a calmer environment, which can psychologically enhance feelings of safety.

#### Reasons that detracts from Safety

- ❖ The buildings and storefronts look worn and poorly maintained, which can signal neglect and potentially higher risk areas.
- ❖ Sidewalks are uneven and partially obstructed, posing physical hazards to pedestrians.
- ❖ Lack of visible security measures such as cameras, police, or clear signage reduces reassurance.
- ❖ Sparse pedestrian presence might make the area feel isolated and less safe, especially for vulnerable individuals.
- ❖ The overhead power lines and general disorderly appearance contribute to a sense of neglect and potential risk.

Predicted Safety Score 45

## Stage 2

### Reasons that enhance *Beauty*

- ❖ The **mature trees** lining the street provide natural beauty and shade, enhancing the visual appeal and comfort of the space. → mature tree
- ❖ The **canal** with **bicycles parked** alongside adds a picturesque and quintessentially local charm, contributing to the scene's uniqueness. → canal → parked bicycle
- ❖ The **historic brick buildings** with their **detailed facades** and **large windows** offer architectural interest and a sense of place. → historical brick building → detailed facade → large window
- ❖ The **cobblestone street surface** adds texture and a traditional aesthetic that complements the overall historic ambiance. → cobblestone street surface
- ❖ The **presence of people** walking and cycling adds life and a sense of community. → people presence

### Perceptual unit

### Reasons that detracts from *Beauty*

- ❖ 'The **street** is somewhat **narrow and enclosed**, which might limit expansive views and a sense of openness that some might associate with beauty. → narrow street → enclosed street
- ❖ The presence of **parked bicycles** and some **street clutter**, while adding character, might be seen as visual distractions by some viewers. → parked bicycle → visual clutter

## Stage 2

### Reasons that enhance *Liveliness*

- ❖ The dense crowd of pedestrians gathered around storefronts and street vendors indicates high human activity and social interaction.
- ❖ Visible commercial establishments with open displays and signage suggest active business and consumer engagement.
- ❖ The compact arrangement of buildings and narrow sidewalks create a sense of enclosure that intensifies the feeling of busyness.
- ❖ Bright daylight encourages outdoor presence, contributing to the street's vibrancy.

### Perceptual unit

- dense crowd
- storefront
- street vendor
- commercial establishment
- open display
- signage
- compact arrange of buildings
- narrow sidewalk
- bright daylight

### Reasons that detracts from *Liveliness*

- ❖ The buildings appear somewhat aged and worn, which might detract from a sense of modern vibrancy or renewal.
- ❖ The street, while busy, does not show signs of festive or celebratory activity that might elevate liveliness to a higher level.
- ❖ The traffic is orderly and not chaotic, which might reduce the perception of energetic hustle.

- aged building
- worn building
- orderly traffic

## Stage 2

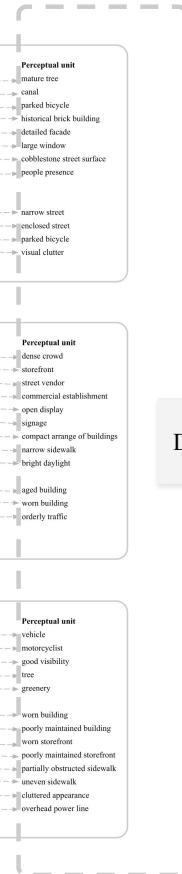
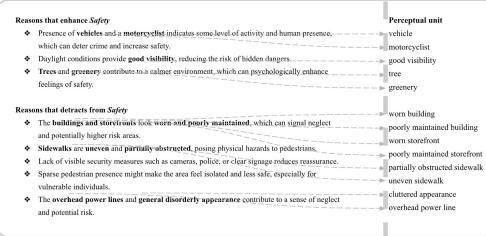
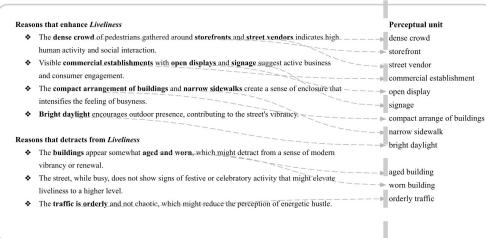
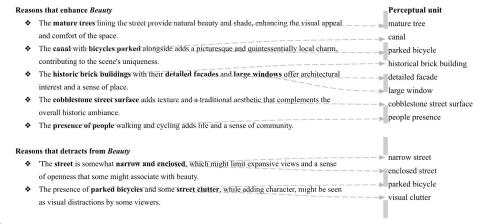
### Reasons that enhance Safety

- ❖ Presence of **vehicles** and a **motorcyclist** indicates some level of activity and human presence, → vehicle
- which can deter crime and increase safety. → motorcyclist
- ❖ Daylight conditions provide **good visibility**, reducing the risk of hidden dangers → good visibility
- ❖ **Trees and greenery** contribute to a calmer environment, which can psychologically enhance → tree
- feelings of safety. → greenery

### Perceptual unit

### Reasons that detracts from Safety

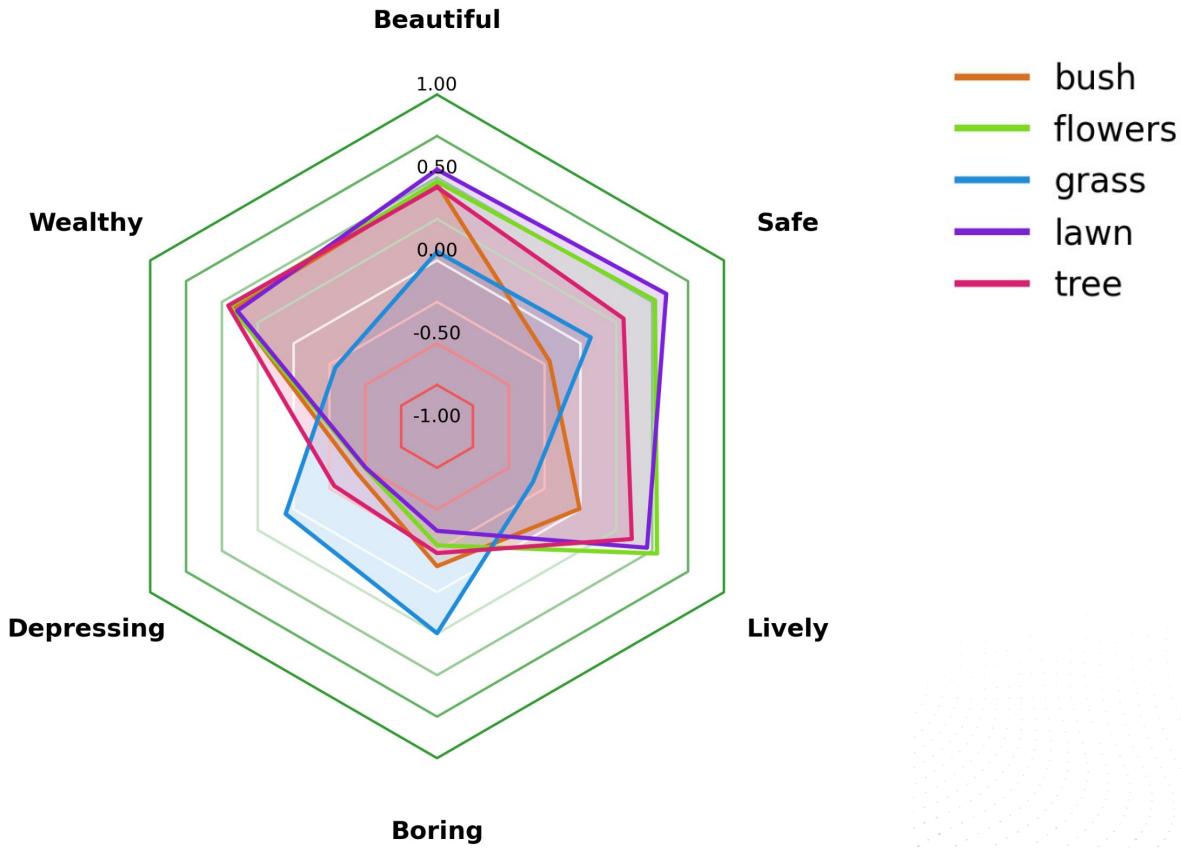
- ❖ The **buildings and storefronts** look **worn and poorly maintained**, which can signal neglect → worn building  
and potentially higher risk areas. → poorly maintained building
- ❖ **Sidewalks** are **uneven** and **partially obstructed**, posing physical hazards to pedestrians. → worn storefront
- ❖ Lack of visible security measures such as cameras, police, or clear signage reduces reassurance. → poorly maintained storefront
- ❖ Sparse pedestrian presence might make the area feel isolated and less safe, especially for → partially obstructed sidewalk
- vulnerable individuals. → uneven sidewalk
- ❖ The **overhead power lines** and **general disorderly appearance** contribute to a sense of neglect → cluttered appearance
- and potential risk. → overhead power line



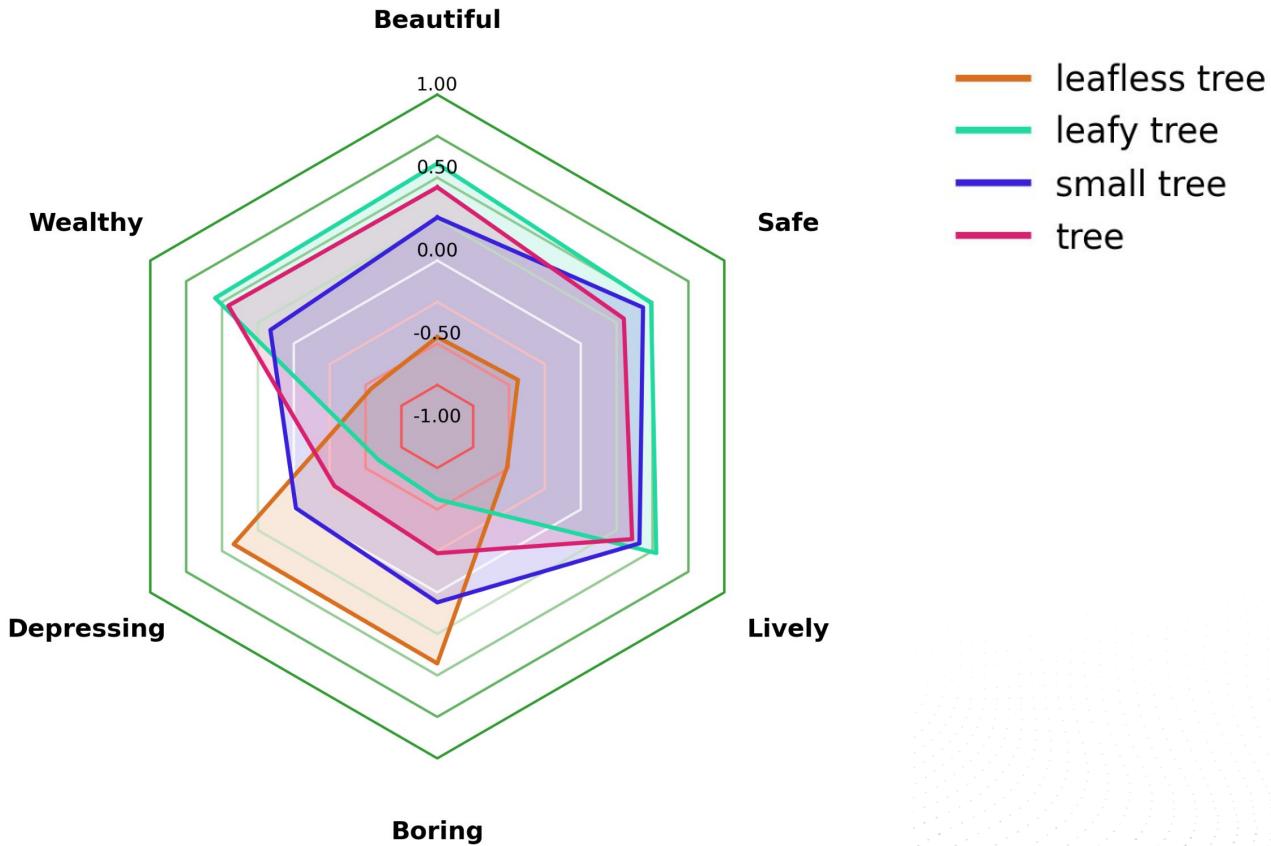
Influence  
Estimated

## RQ1: Quantifying the Fine-grained Perceptual Differences Within Broad Categories

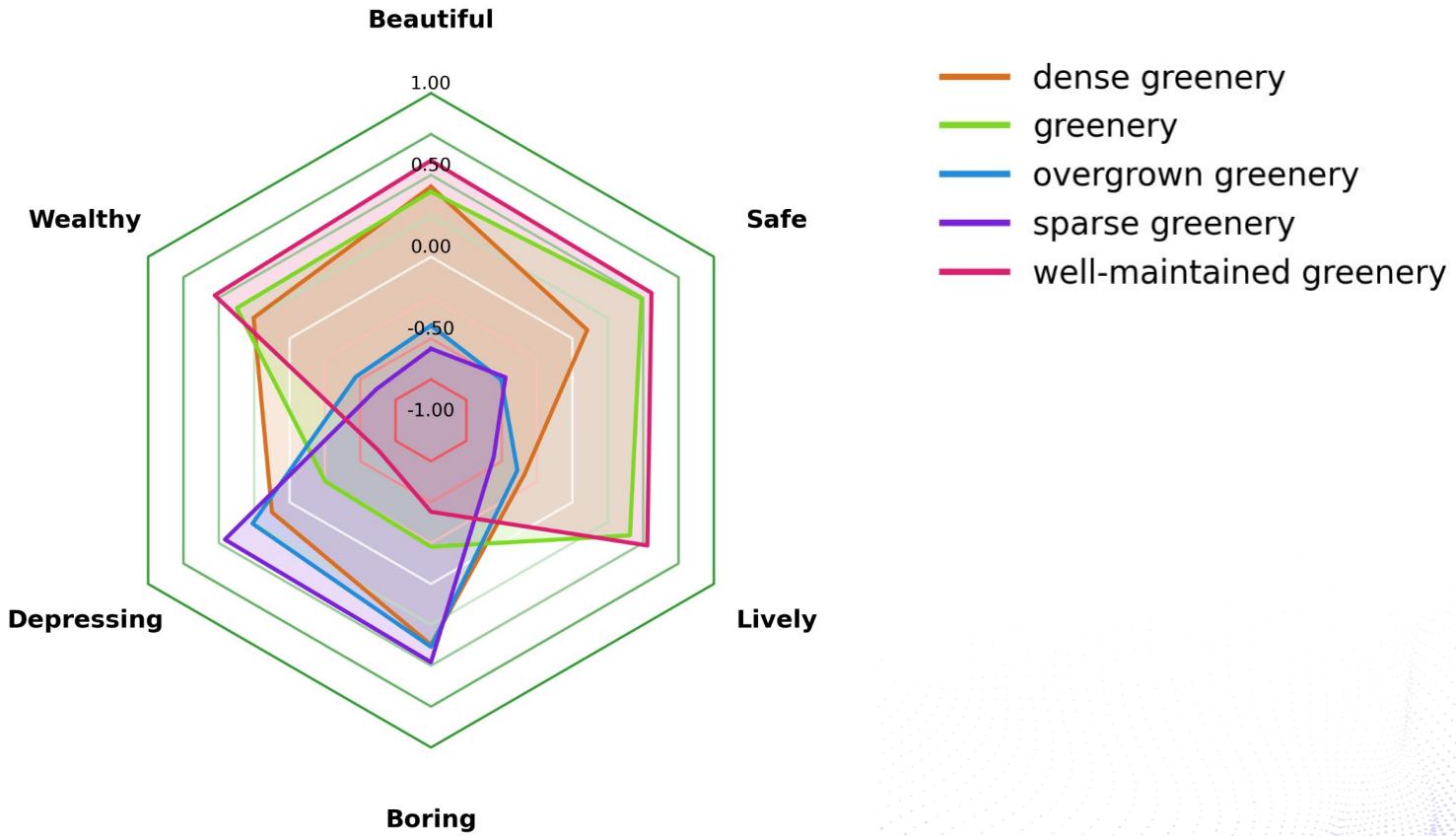
# Variations in Base Elements



# Variations in Qualifiers

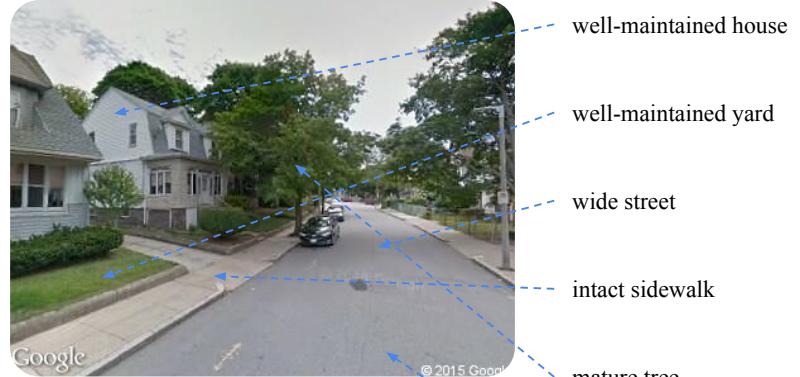
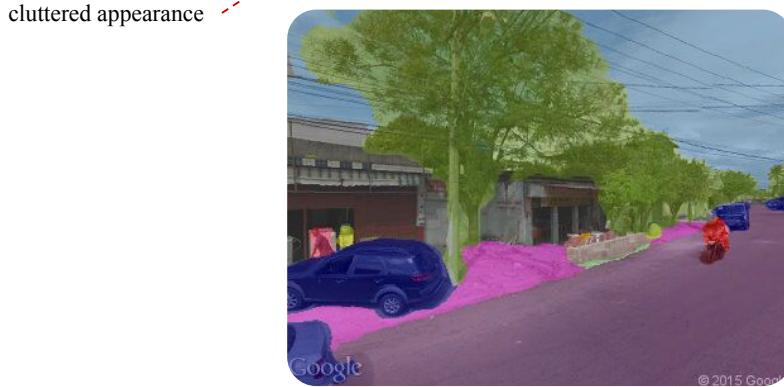


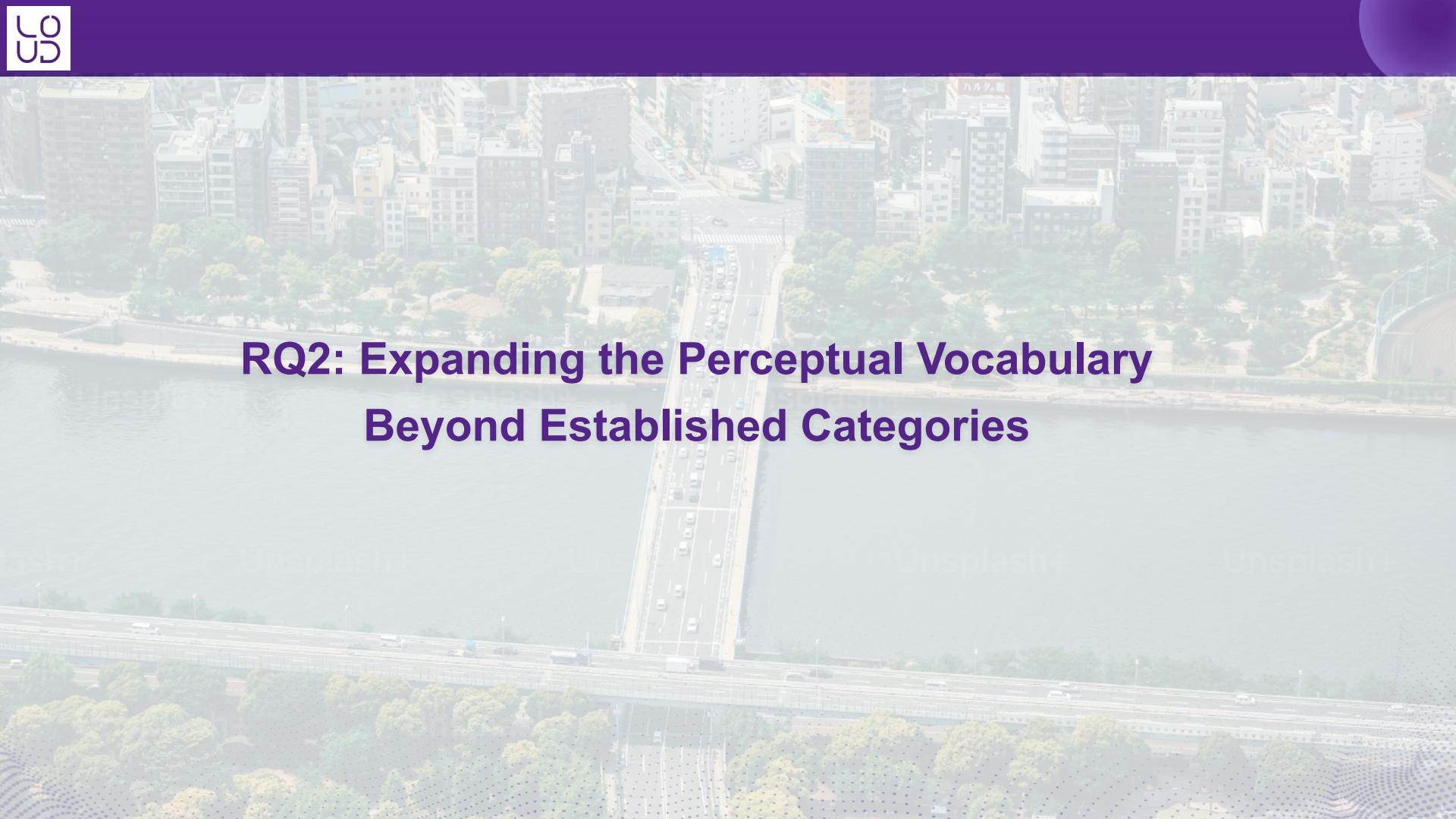
# Variations in Qualifiers



# Variations in Other Segmentation Categories

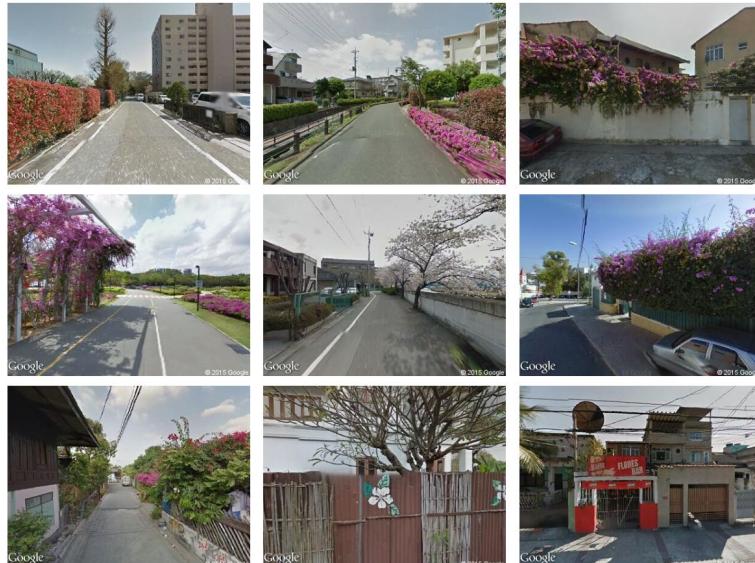




The background image shows a dense urban landscape from an aerial perspective. In the foreground, there's a river with a large, multi-lane highway bridge spanning it. The city is filled with numerous buildings of varying heights, mostly modern apartment complexes and office buildings. Some green spaces and trees are visible between the buildings. The overall scene is a typical representation of a major city's infrastructure and urban sprawl.

## RQ2: Expanding the Perceptual Vocabulary Beyond Established Categories

## Detect Novel Visual Features Relevant to Urban Perception



Flower / Blossom



Overhead Power Lines

## Describe Fine-grained Elements within Segmented Areas



Roller Shutter Door

European Style Window

## Characterize Quality and Aesthetic Nuances of Specific Instances



Manicured Lawn

Overgrown Lawn

## Identify Social Behaviors and Demographic Presence



Children Playing



Crowd Shopping

## Capture the Impact of Transient and Contextual Factors



Gloomy Weather

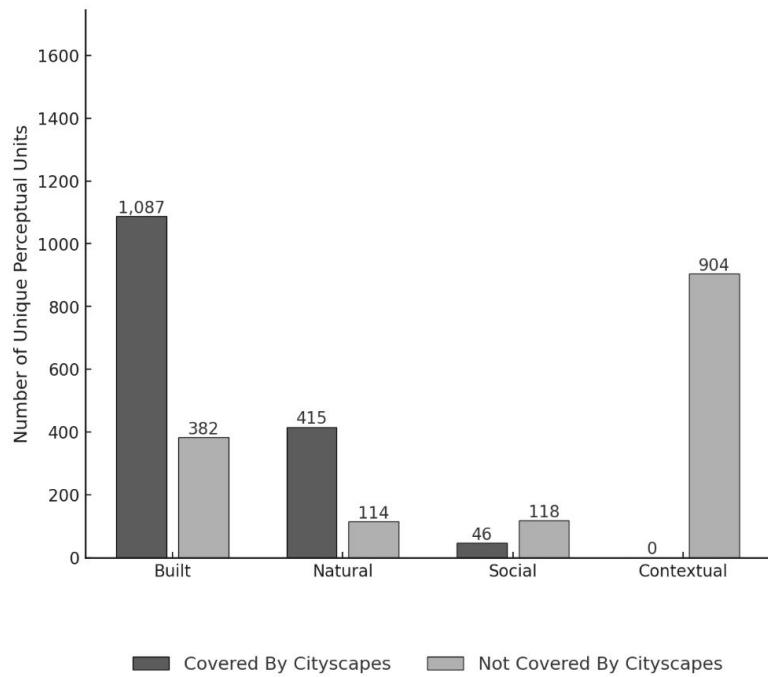
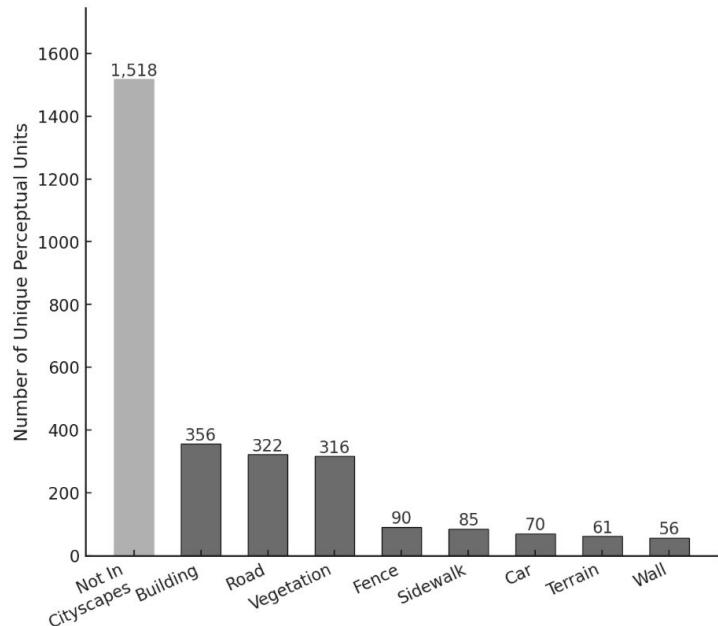
Sunny Weather



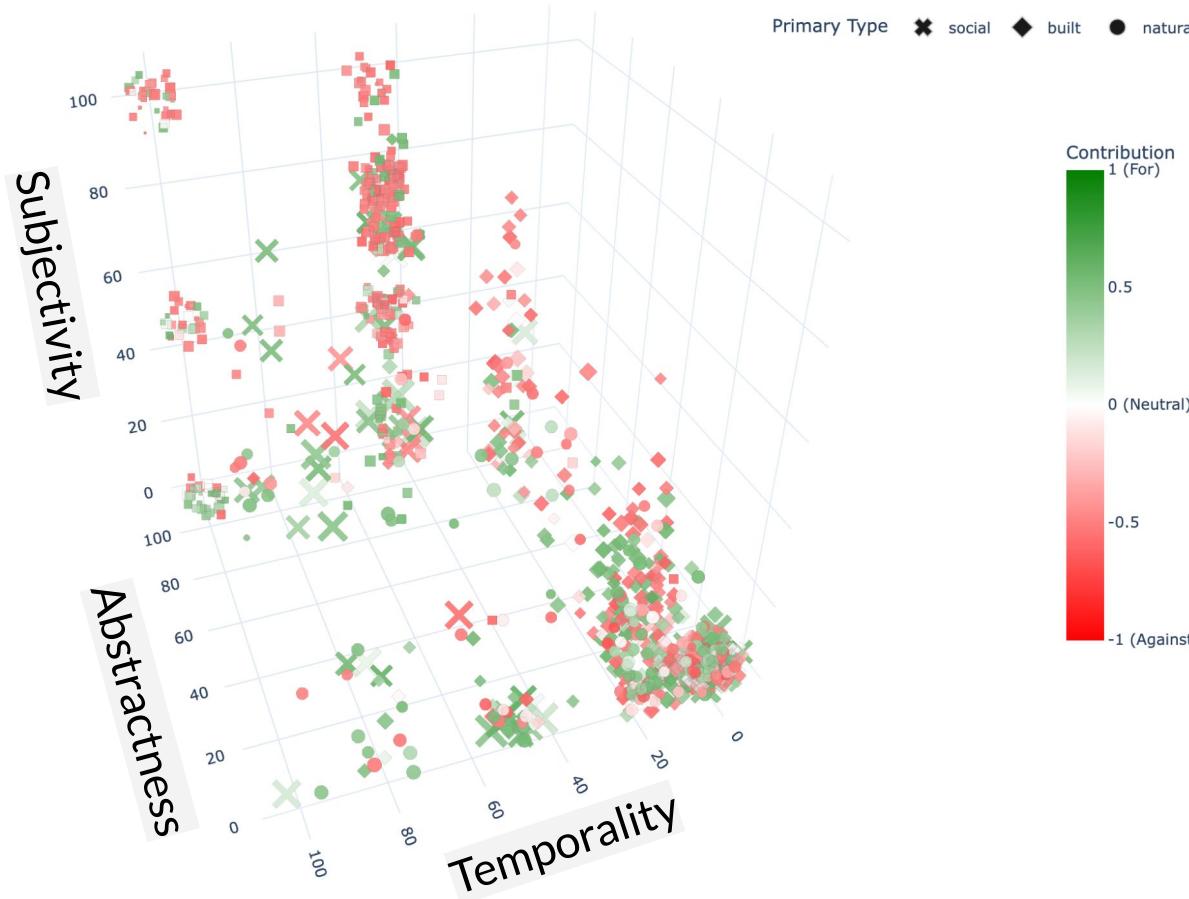
Maybe it's just bad weather ...

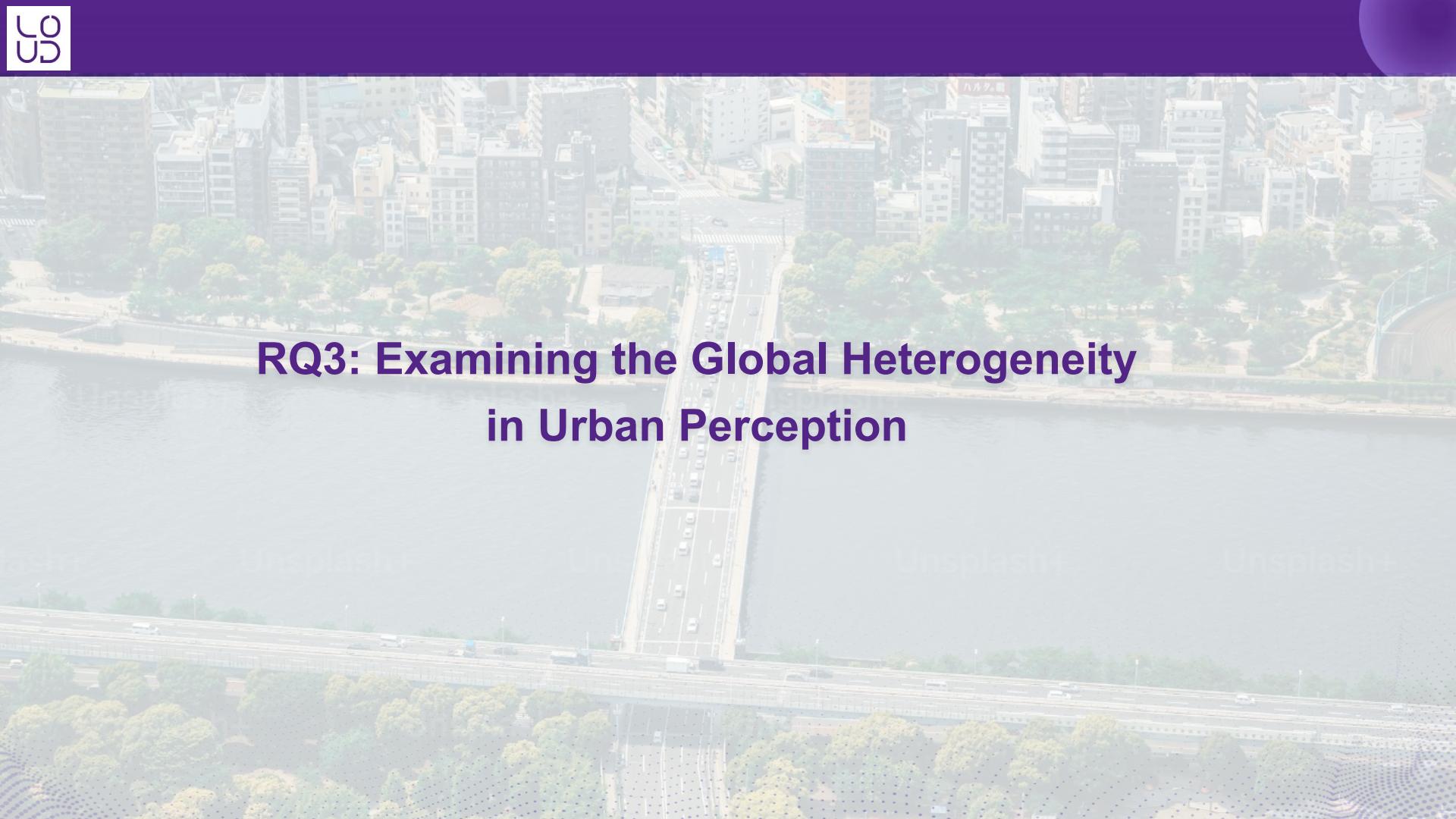


# Nearly Half of the New Features Fall Outside Conventional Taxonomies



# A Navigable Feature Space to Organize the Diverse Features



An aerial photograph of a cityscape, likely Tokyo, showing a dense cluster of buildings, a river, and a multi-lane highway bridge spanning across it. The image has a semi-transparent purple overlay.

## RQ3: Examining the Global Heterogeneity in Urban Perception

# Influences of the Same Feature Estimated with Different City-specific Subsets

Grass



*Depressing contribution: +0.521*

Gaborone

*Depressing contribution: -0.487*

Minneapolis

# Influences of the Same Feature Estimated with Different City-specific Subsets

**Brick**



*Wealthy contribution:-0.373*

**Belo Horizonte**



*Wealthy contribution: +0.571*

**Atlanta**

# Influences of the Same Feature Estimated with Different City-specific Subsets

## Graffiti



***Safe contribution: -0.583***

**Rio De Janeiro**



***Safe contribution: -0.257***

**Berlin**

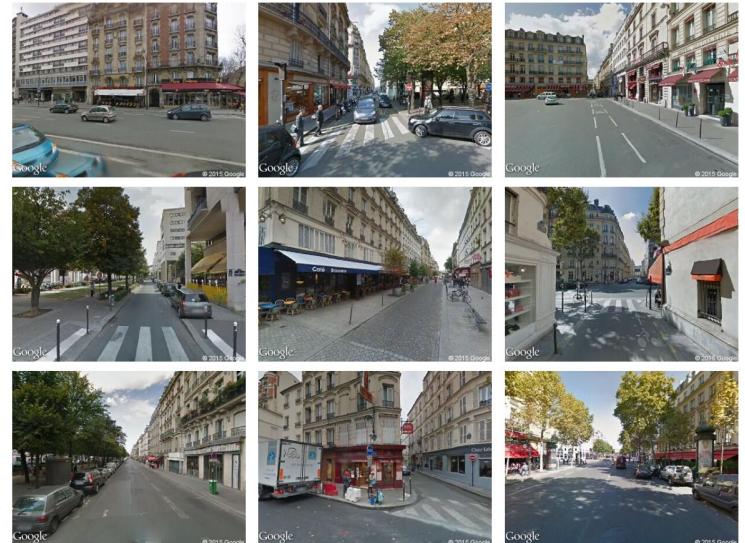
# Influences of the Same Feature Estimated with Different City-specific Subsets

## Awning



*Beautiful contribution: -0.076*

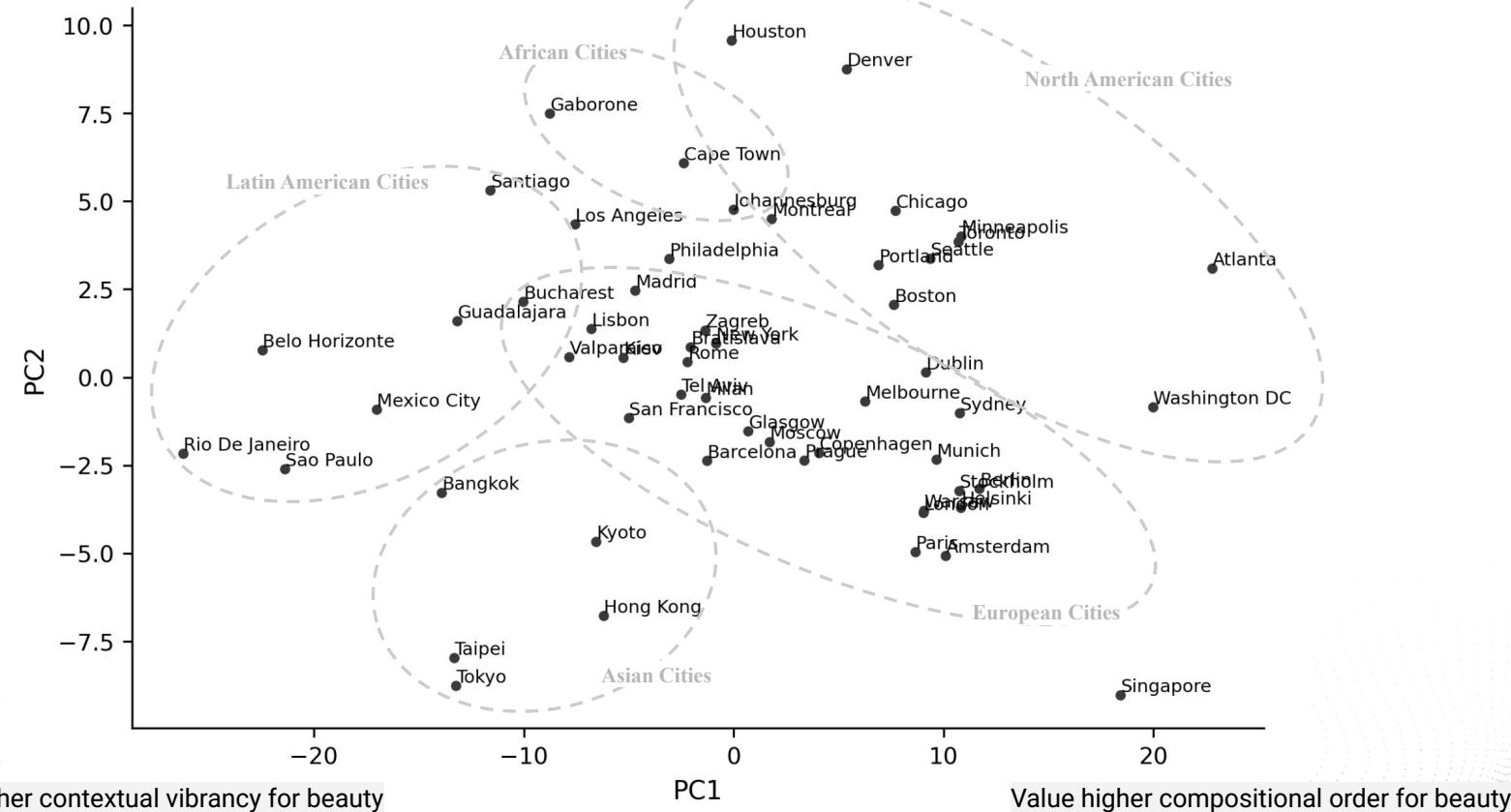
Taipei



*Beautiful contribution: +0.584*

Paris

# PCA Analysis of Perceptual Fingerprints of Cities Shows Emergent Patterns



# Summary

- Explainable pipeline: two-stage process, free-form reasons to perceptual units
- Fine-grained perception mapping: captures variations with base–qualifier structure
- Unconstrained discovery: social and contextual cues beyond traditional segmentation
- Global heterogeneity: city-specific weights, perceptual fingerprints showing clustering
- Practice-ready use: interpretable levers, locally tuned design, collaboration framework

# The Cityscapes Dataset for Semantic Urban Scene Understanding — Cordts et al. 2016



Group	Classes
flat	road · sidewalk · parking <sup>+</sup> · rail track <sup>+</sup>
human	person <sup>*</sup> · rider <sup>*</sup>
vehicle	car <sup>*</sup> · truck <sup>*</sup> · bus <sup>*</sup> · on rails <sup>*</sup> · motorcycle <sup>*</sup> · bicycle <sup>*</sup> · caravan <sup>*+</sup> · trailer <sup>*+</sup>
construction	building · wall · fence · guard rail <sup>+</sup> · bridge <sup>+</sup> · tunnel <sup>+</sup>
object	pole · pole group <sup>+</sup> · traffic sign · traffic light
nature	vegetation · terrain
sky	sky