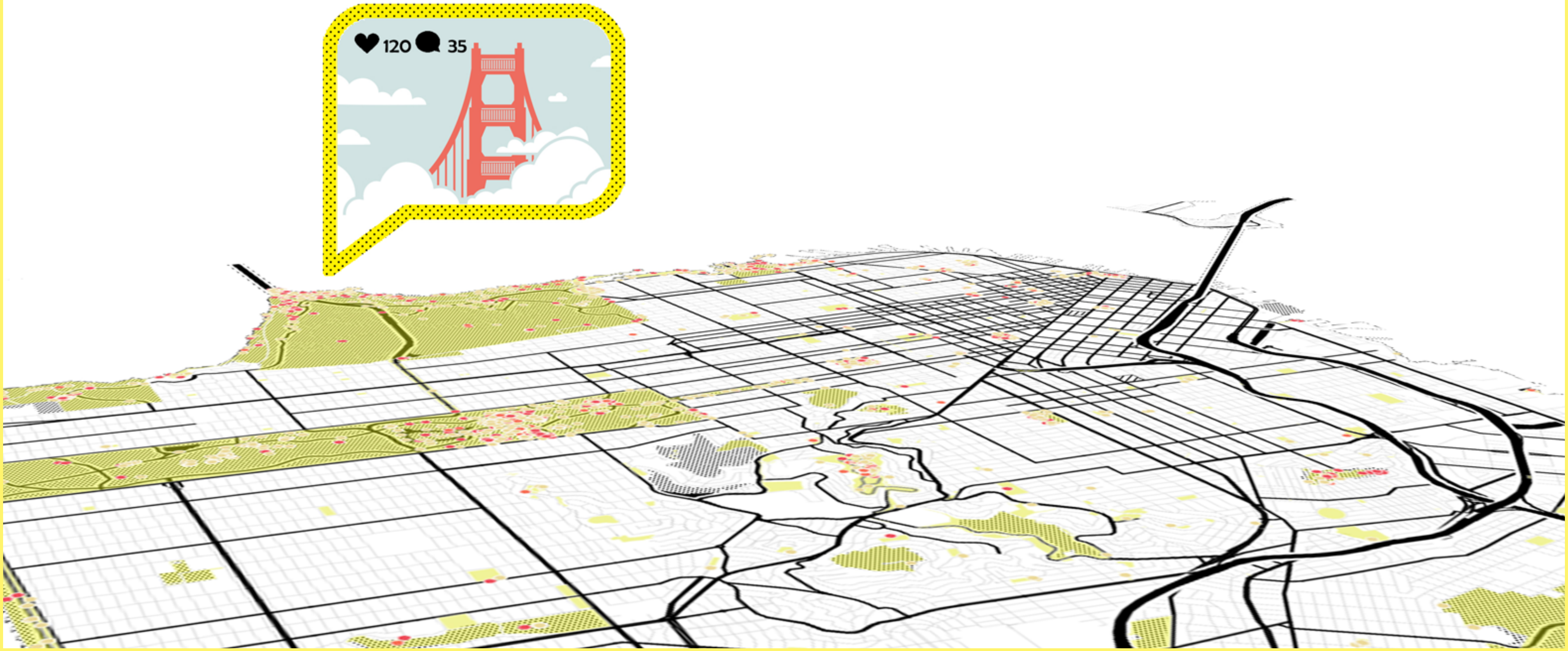
CROWDSOURCING

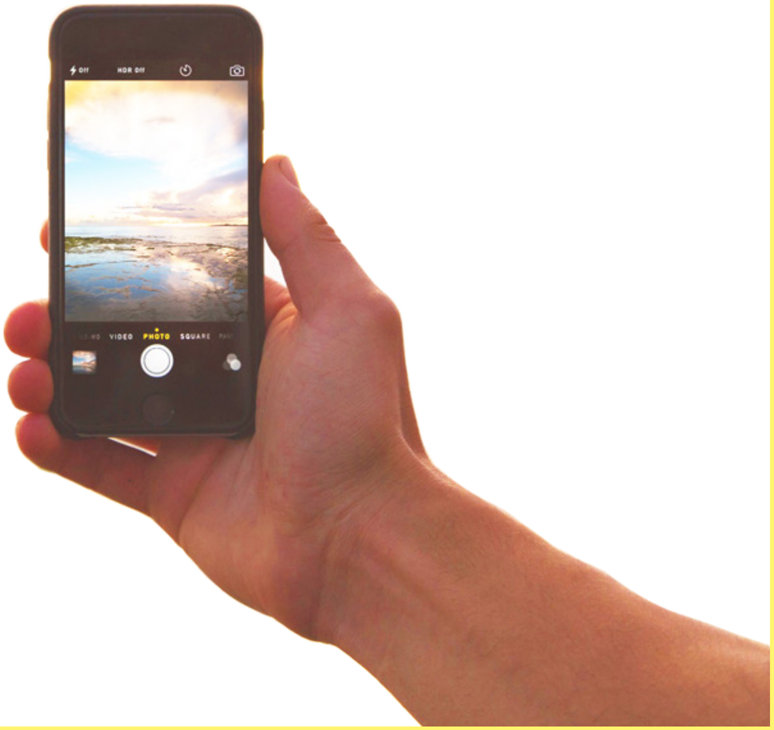
Using Geotagged Photos from Flickr and Instagram to Study Urban Green Space in San Francisco



CONTENTS

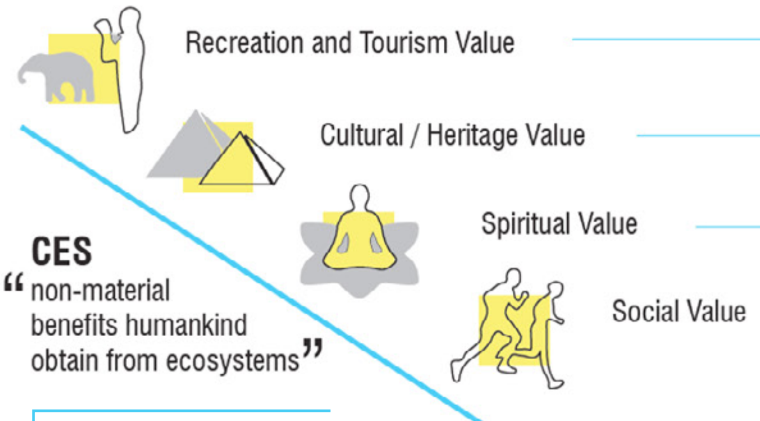
ABBREVIATION

UGS - Urban Green Space
CES - Cultural Ecosystem Services
API - Application Programming Interface
SNS - Social Network Services
HCA - Hierarchical Clustering Algorithm



01 INTRODUCTION

Urban green spaces (UGS) provide a diversity of benefits for human beings, among which, cultural ecosystem service (CES) is an important category. However, measurement and mapping of city-wide CES provided by UGS is underdeveloped though it is important for both urban planners and landscape designers.



An example of urban green space
Conservatory of Flowers in Golden Gate Park (source: Google Places)

UNDERDEVELOPED MAPPING

- Lack of city-wide spatially referenced data
- Quality indicators for assessing CES are unclear

MEASUREMENT AND MAPPING CITY-WIDE CES

- **Important for urban planners** Promote plannings for sustainable development, especially land-usemanagement, tourism planning, and conservation planning.
- **Important for landscape designers** Understand people’s interests in nature, people’s behavior and preferences in urban green space.

RESEARCH QUESTIONS

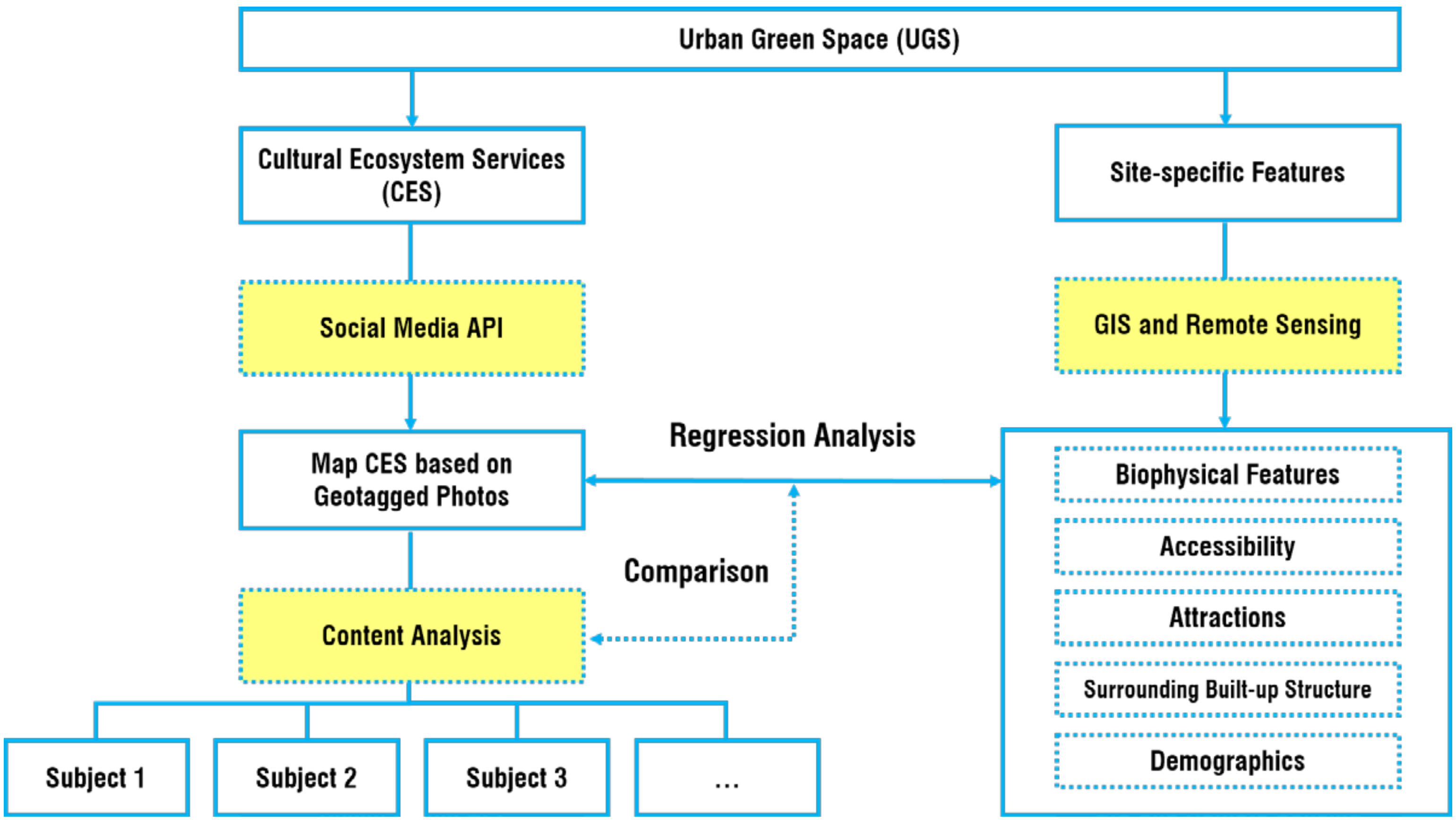
1 ***How to map and assess city-wide CES provided by UGS?***

2 ***What’s the relationship between CES provision and site-specific features of UGS?***

02 DATA & METHODS

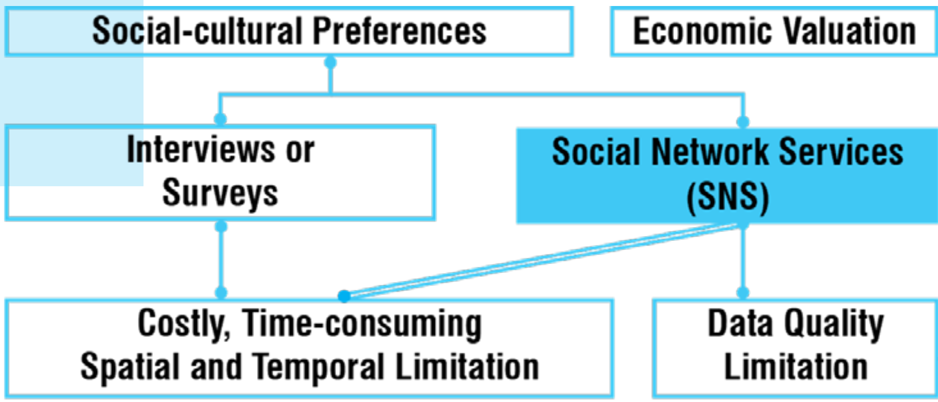
METHODOLOGY

Geotagged photos were collected using social media Application Programming Interface (API). To answer the first research question, static and dynamic mapping methods were applied and the results of different approaches and photo groups (Instagram and Flickr) were compared to get conclusions. Site-specific features of UGS were described by different categories of objective indicators to answer the second question. Hypotheses about the relationship between each indicator and CES provision were made based on observations, literature review, as well as the results of CES mapping and photo content analysis. Then I applied regression analysis to figure out the driven factors of the spatial pattern of CES, and compared the results with hypotheses to get conclusions.

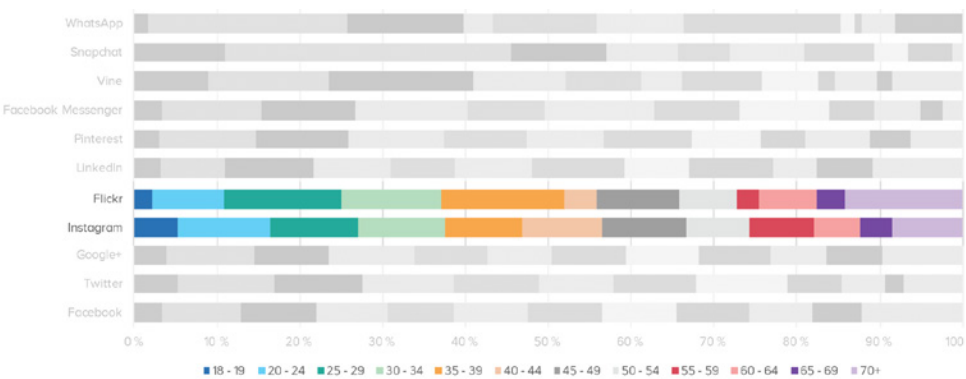


02 DATA & METHODS

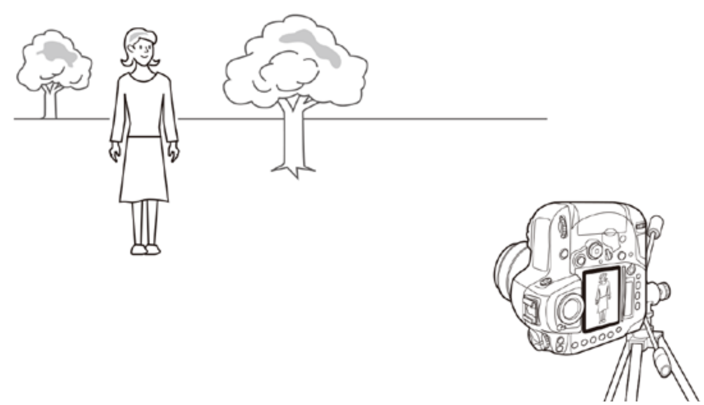
ADVANTAGES & LIMITATIONS



- The density of geotagged photos is not a perfect indicator of CES, mainly reflects recreation and aesthetic value
- Bias of user groups is a big concern
- Biased behavior on social media → not geo-tagged, or inaccurate location
- Social media privacy policy → few demographic data about users
- Photographic orientation is rarely recorded: geo-tagged locations ≠ target sites



"Bias of user group is a big concern" (different age groups represented by different social media, here taking Flickr and Instagram as an example)

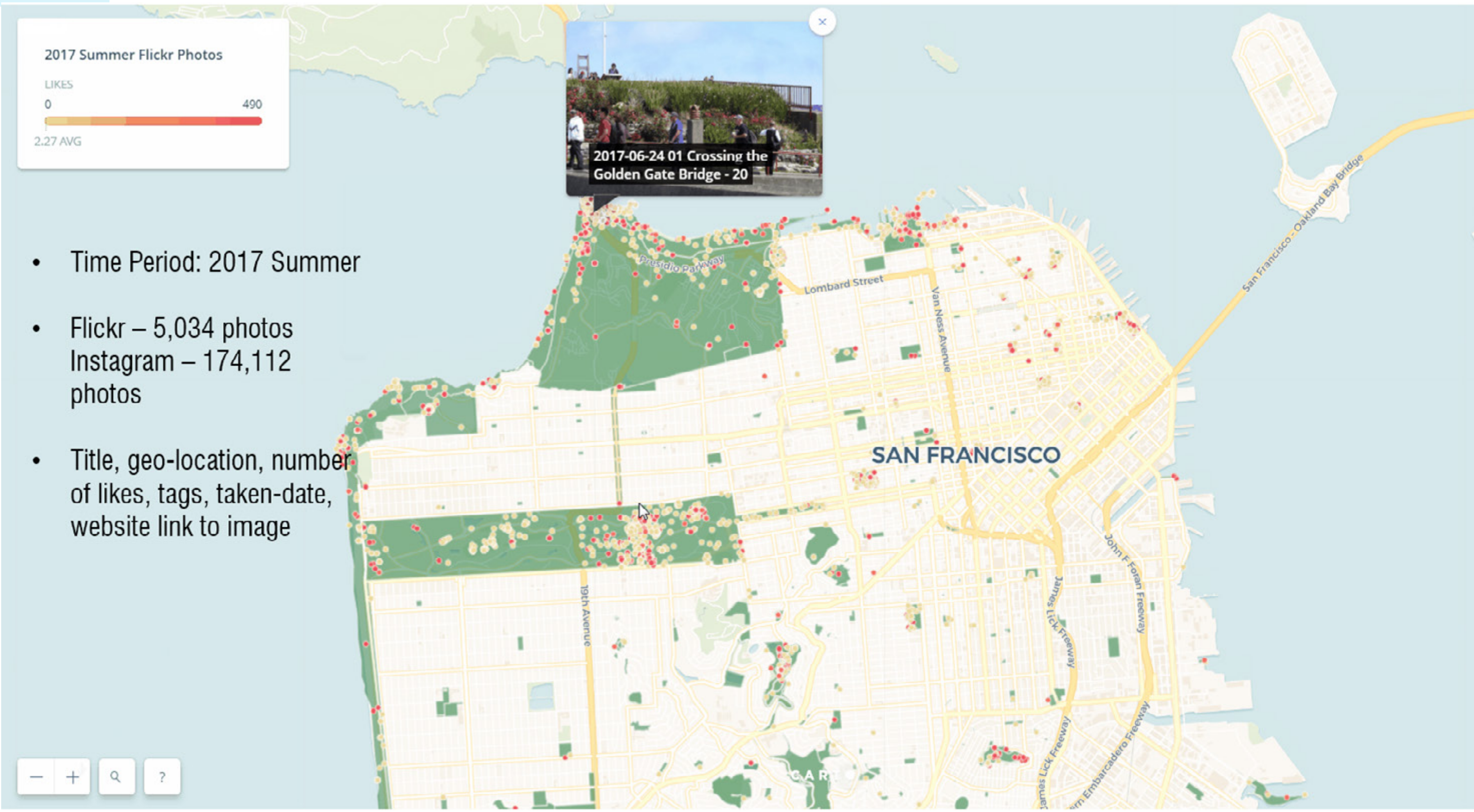


"Geotagged locations ≠ target sites"

02 DATA & METHODS

SITE SELECTION & PHOTO DATABASE

San Francisco is suitable for testing this methodology for three reasons. Firstly, there is a great amount of urban green spaces in San Francisco. Secondly, San Francisco urban green space includes a variety of types of spaces, providing various kinds of cultural ecosystem services. Thirdly, there is a great deal of spatial referenced data available because San Francisco is both a highly urbanized dense city and a world-famous tourist destination.



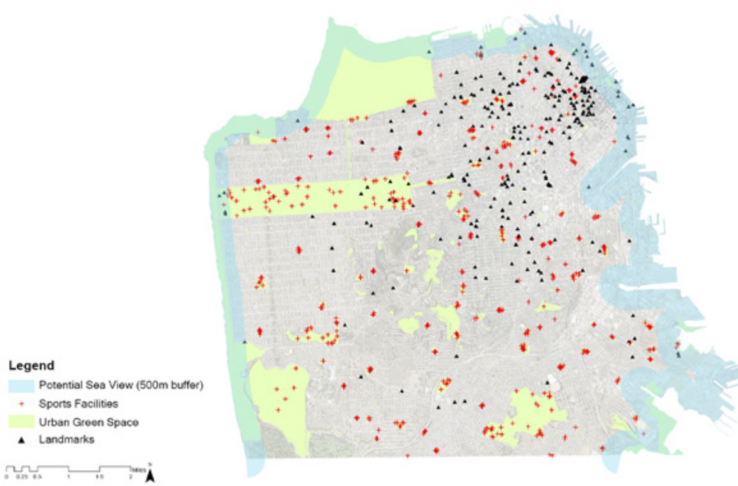
■ **Figure. The interactive map of Flickr geotagged photos taken in San Francisco urban green space during the summer of 2017.**

02 DATA & METHODS

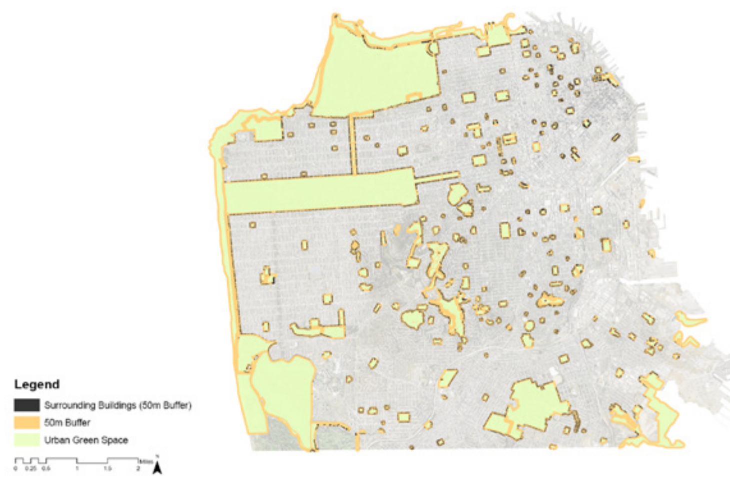
SITE-SPECIFIC FEATURES OF UGS

■ **Figure 2.1. San Francisco urban green space land cover map.**

■ **Figure 2.2. Network analysis to find the nearest distance from urban green spaces to BART stations.**



■ **Figure 2.3. The map of partial attractions in San Francisco urban green space.**



■ **Figure 2.4. The map of buildings in 50m buffer of urban green space in San Francisco.**

Site-specific features such as biophysical features, accessibility, and recreational facilities of UGS are highly related to the capacity of UGS to deliver cultural ecosystem services. According to surveys and interviews about park visitation in terms of CES, sports facilities, accessibility, and aesthetics are of relative importance though the results varied by age. Also, there is a link between spatially explicit indicators such as surrounding building height and density, and visitor's perceptions of UGS. Therefore, I used multiple categories of site-specific features including biophysical features, accessibility, attractions, demographics, and surrounding built-up structure to describe UGS.

02 DATA & METHODS

SITE-SPECIFIC FEATURES OF UGS

■ Table 2.1. Site-specific feature of urban green space in San Francisco.

Indicator		Type	Data Source	Method
Biophysical Features	Size (hectare)	Continuous	CPAD (California Protected Area Database)	Calculating geometry in ArcGIS
	Landscape Shape index (LSI)	Continuous		Calculating $LSI = \frac{P}{2 \cdot \sqrt{A \cdot \pi}}$ in Jupyter Notebook
	Percentage of woody vegetation	Continuous	NAIP Image, CPAD, and the layer of building footprints derived from LiDar data (city GIS data portal)	OBIA Classification in ArcGIS and Calculation in Jupyter Notebook
	Percentage of vegetation (woody vegetation and grass/meadow)	Continuous		
	Percentage of impervious surface (with and without building footprints)	Continuous		
	Waterbody	Binomial		
	Simpson's Diversity Index (SIDI)	Continuous		Calculating $SIDI = 1 - \sum_{i=1}^{m} P_i^2$ in Jupyter Notebook
Accessibility	The number of bus stops	Continuous	SFMTA routes and stops data (city GIS data portal)	Creating a 300m buffer for each UGS to see how many bus stops are in the buffer in ArcGIS
	Nearest distance to BART Stations	Continuous	SF Bay Area BART Stations (county GIS data portal)	Creating a 15m buffer for each UGS to figure out all street junctions, and then finding nearest BART station using network analysis in ArcGIS
	Bike Parking Facilities	Binomial	Bike Parking database (city GIS data portal)	Creating a 100m buffer for each UGS to see whether there is bike parking facilities in ArcGIS
Attractions	The number of landmarks	Continuous	Landmarks dataset (city GIS data portal)	Creating a 50m buffer for each UGS to see the number of landmark in ArcGIS
	Sports facilities	Binomial	Cultural and recreation facilities owned by the city (city GIS data portal) and Google Map	Filtering facilities within the group including tennis courts, basketball field, recreation centers/pools, stadiums, children's play area, do play area, picnic area, and so on. Manually identifying whether there are sports facilities for other UGS not owned by the city
	Cultural & entertainment spots	Binomial	Google Map	Manually checking whether there are museums, art galleries, public libraries, shopping center/mall, or popular restaurants/bars/bee gardens in 50m buffer of UGS
	Shore View	Binomial	San Francisco boundary (city GIS data portal)	Creating a 500m buffer inside the coast to see which UGS are overlapped in ArcGIS
Surrounding Built-up Structure	Average height of surrounding buildings (meter)	Continuous	Building footprints derived from LiDar data (city GIS data portal)	Creating a 50m buffer for each UGS in ArcGIS and calculating all the variables in Jupyter Notebook
	Density of surrounding buildings (100 square meters of building area per hectare)	Continuous		
Demographics	Population Density (people per hectare)	Continuous	2016 Planning Database and San Francisco boundary (city GIS data portal)	Calculating the average population density in 300m buffer of UGS based on the estimation of population at census group level

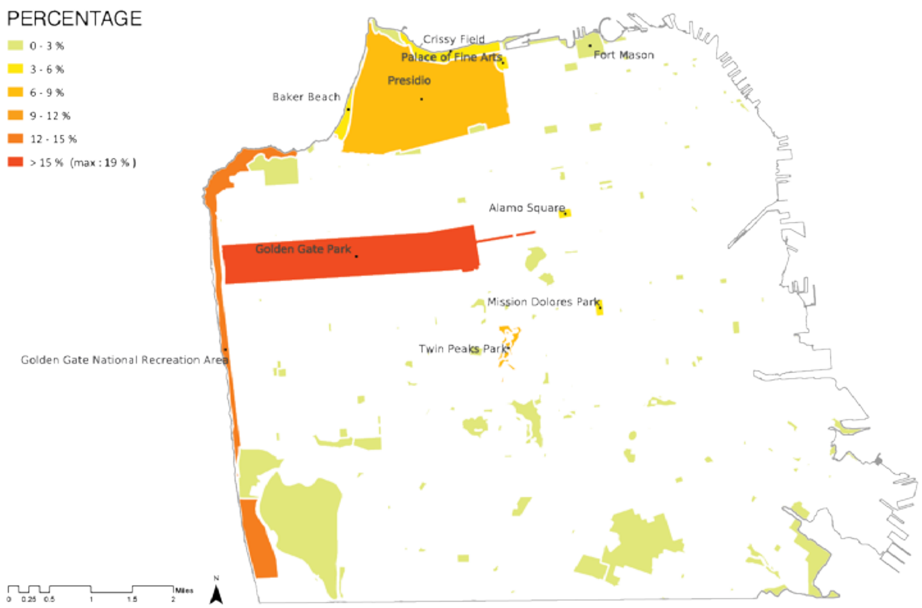
03 ANALYSIS & RESULTS

SPATIAL VARIATION OF CES INDICATED BY GEOTAGGED PHOTOS

Choropleth maps were used to visualize the geotagged photos counts, the geotagged photos counts normalized by the size of UGS, and the number of likes gained by geotagged photos in San Francisco UGS. All the three ways assign each UGS average values, which can be the indicators of its CES at such a big scale. The main results show as following:

- *Mapping CES in different ways is necessary and meaningful. The maps of photo counts and the maps of number of likes on photos are relatively consistent, while the maps of photo counts per hectare are quite different.*
- *UGS size and closeness to shoreline appear to be important factors.*
- *Comparing results of Flickr and Instagram groups, the spatial patterns of CES mapped in the three ways look similar, with little difference reflecting the different preferences of two represented groups.*

Instagram



Flickr

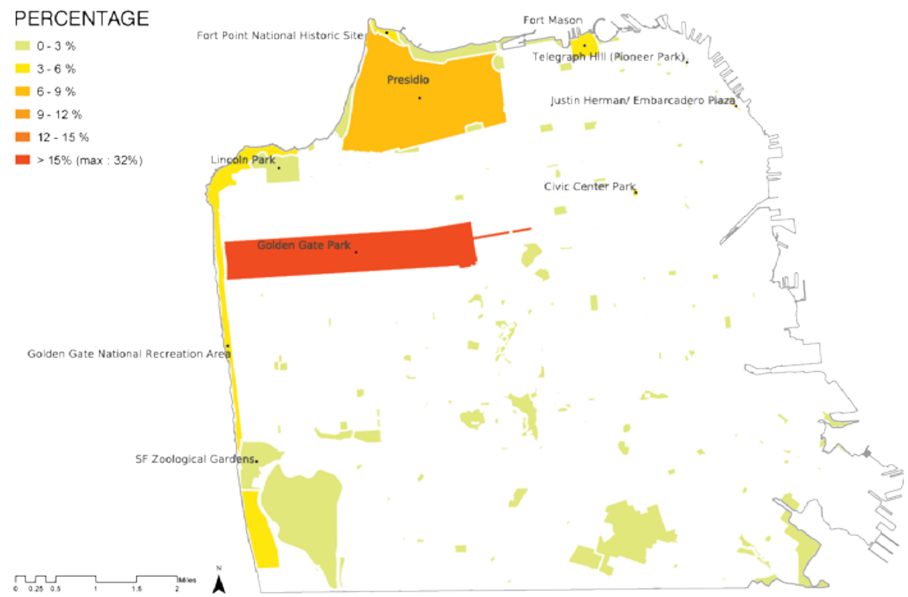
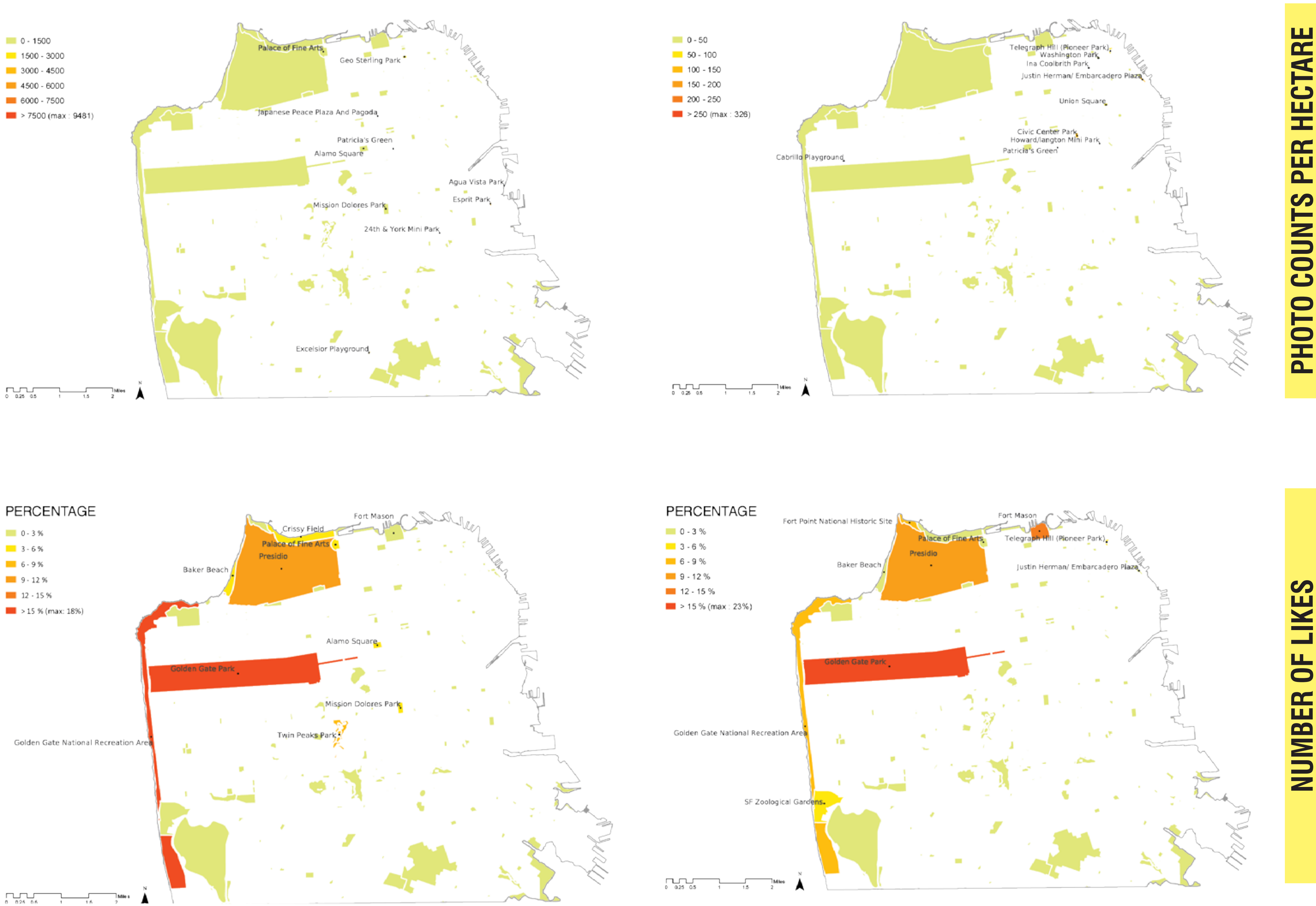


PHOTO COUNTS

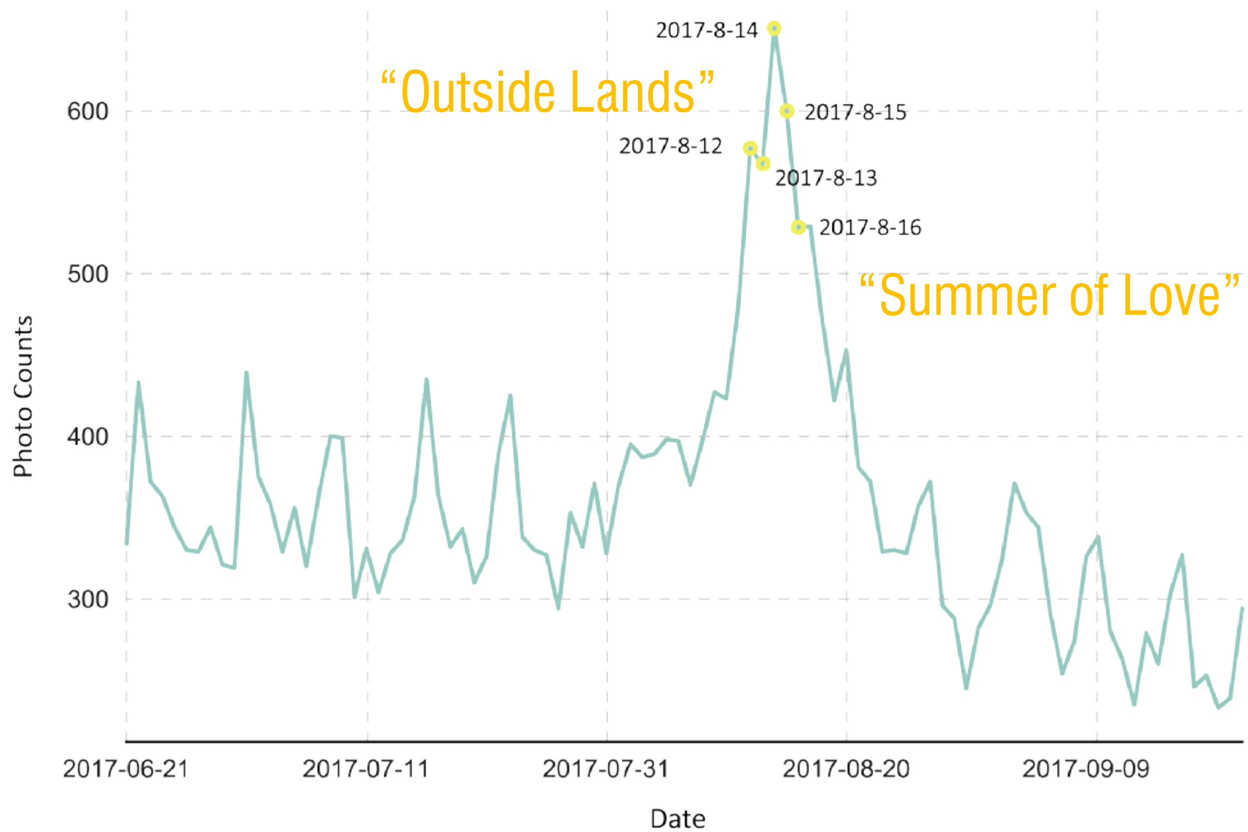


■ **Figure 3.1. Three ways to map cultural ecosystem services provided by urban green space using geotagged photos in San Francisco during the summer of 2017. The top 10 UGS are annotated in each map. Maps on the left are based on Instagram geotagged photos, and the right part are based on Flickr geotagged photos.**

03 ANALYSIS & RESULTS

TEMPORAL VARIATION IN GEOTAGGED PHOTO COUNTS

Social event, such as the musical festival and the bike race, is an indispensable part of recreational value in CES. Timestamps and tags of photos make it possible to perform temporal analysis to detect social events. In this study, I used Instagram photo datasets to do a simple test. After inspecting the time series and box plots of daily photo counts in a few parks, I extracted suspect outliers which had photo counts more than the upper inner fence, 1.5 times the Interquartile Range (IQR) above the third quartile. For each outlier date, I sorted unique tags for all the photos to find the top 5 tags. In this way, I mapped the number, scale, and types of possible big social events in all UGS during the summer of 2017.



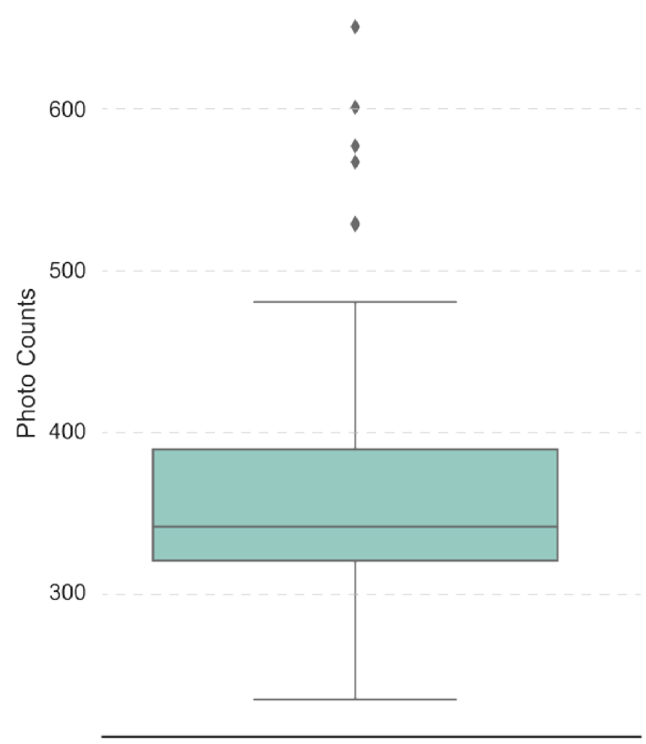
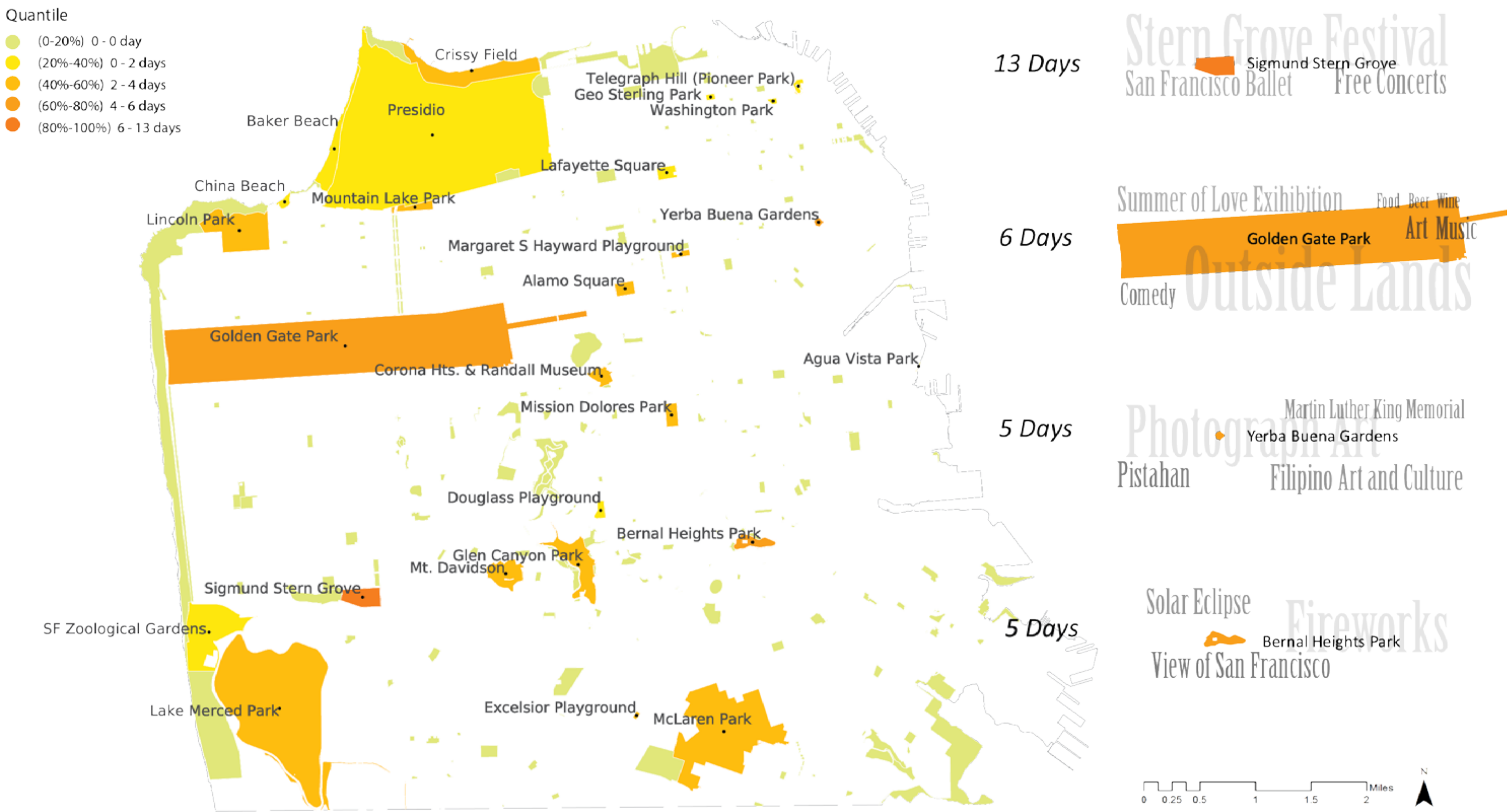


Figure 3.2. The example for Golden Gate Park big event detection in 2017 Summer using Instagram geotagged photos. (left: daily geotagged photo counts time series in Golden Gate Park; right: the box-plot for geotagged photo counts in Golden Gate Park; highly repeated tags were annotated on the left figure).

The main results show as following:

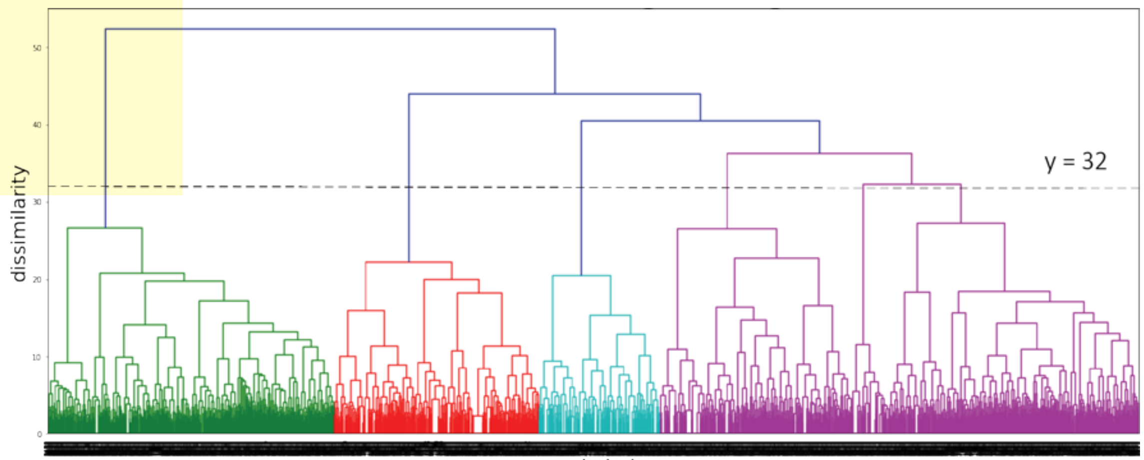
- ***26 out of 251 UGS (10.3%) had big events in the summer of 2017. The capability of holding big events is not limited by size of UGS, but the scale of events is affected by the size.***
- ***Events are diverse but mostly about art, music and culture.***



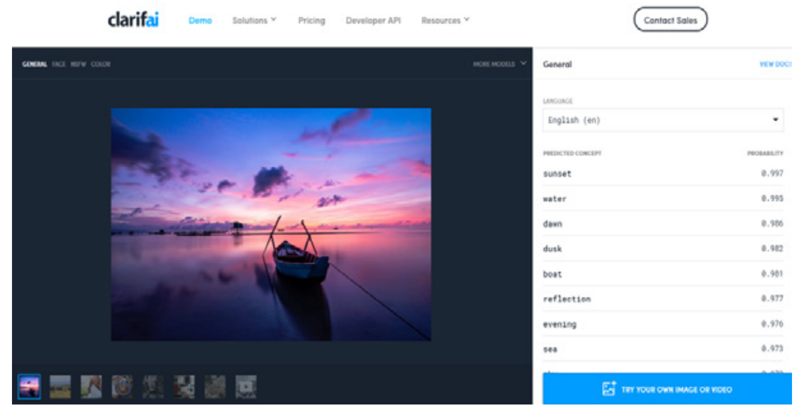
■ **Figure 3.3. The map of the numbers of big event days in San Francisco urban green space during the summer of 2017 based on Instagram geotagged photos. On the right are top 4 urban green space annotated with some highly repeated tags the users posted.**

03 ANALYSIS & RESULTS

PHOTO CONTENT AND CLASSIFICATION



■ Figure 3.5. The dendrogram of hierarchical clustering of 5034 Flickr photos into 6 clusters.

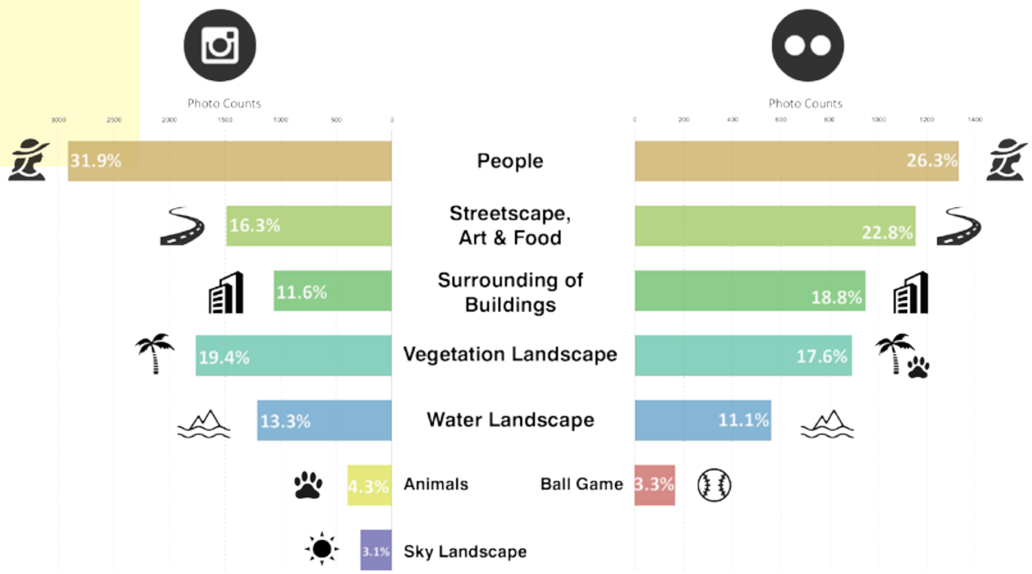


■ Figure 3.4. Clarifai pre-built “General” model demonstration.

■ Figure 3.6. Interpretation of 6 retrieved clusters of Flickr photos based on top 5 labels in each cluster.

To understand the relationship between CES provision and site-specific features, the indices such as photos counts and density used in mapping city-wide CES in Fig. 3.1 are not enough. Photo content analysis can help understand human interactions with different features in the site and the specific activities people engaged in. However, manually categorization of photos is arduous and can result in some errors when the image dataset is extremely large. Therefore, in this study, I used Clarifai (Fig. 3.4), an image recognition tool helping rapidly analyze the content of photos. Geotagged photos were assigned top 5 labels according to the result of content analysis and then clustered based on their similarity in labels using Hierarchical Clustering Algorithm (HCA) (Fig. 3.5). At last, I interpreted these clusters based on the barplots of top 5 labels (Fig. 3.6) and sample photo check.

03 ANALYSIS & RESULTS



■ **Figure 3.7. Bar charts of subjects of clustered Flickr and Instagram geo-tagged photos.**

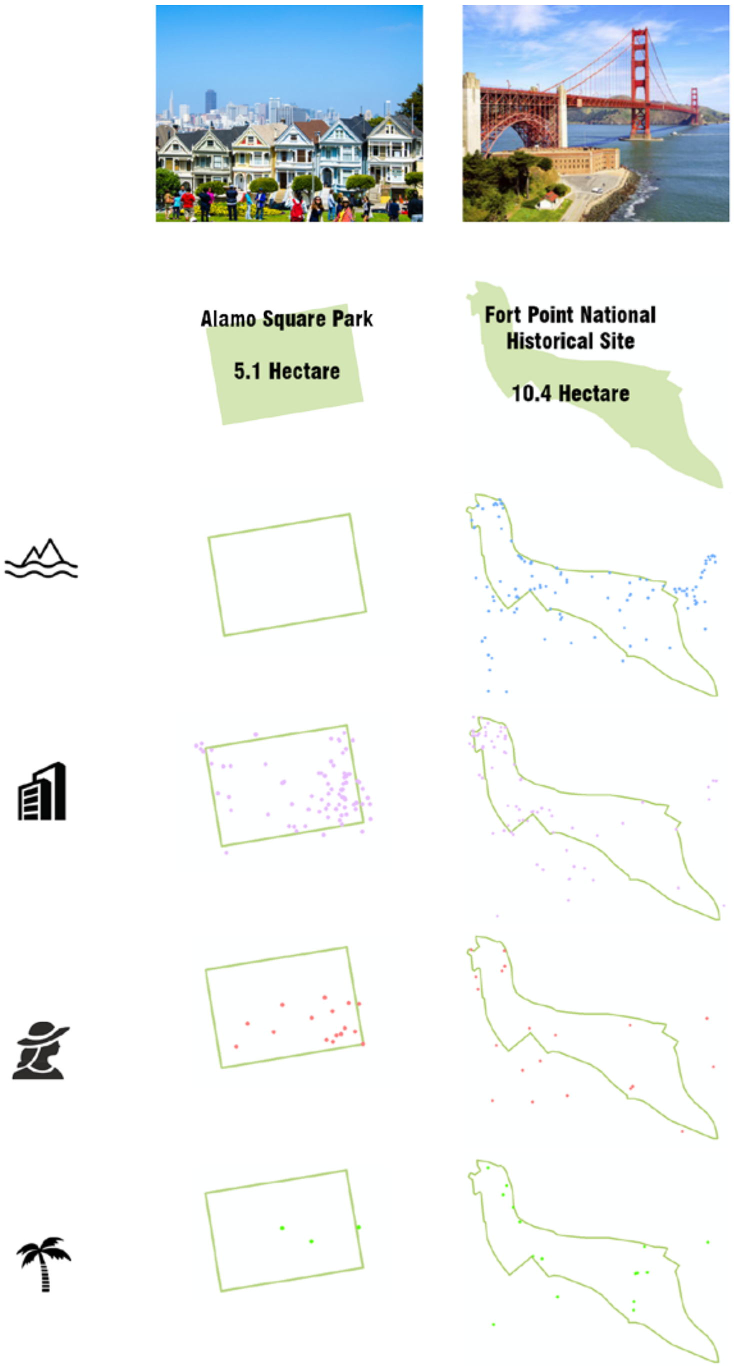
■ **Table 3.1. Interpretation of subjects of clustered Flickr and Instagram geotagged photos**

Subjects	Descriptions
People	Portraits and group photos related with some recreational activities
Surroundings of Buildings	Architecture is the principle part of the photo, with some surroundings including trees, flowers, the lawn, and so on.
Vegetation Landscape	Natural elements including trees, flowers, and the grass is the main focus in the photo.
Water Landscape	The sea and other open water body is the main part of the photo, with some surroundings such as seashore, sand beach, and bridge.
Animals	Wildlife and pets are the main focus in the photo.
Sky Landscape	Sky is the main topic of the photo, including sunset, dusk, dawn and nightscape.
Ball Game	Depicting people playing sports, especially the baseball game.
Streetscape, Art and Food	Some small other subjects such as food, art (posters, exhibitions, and etc.) and streetscape (cars, bicycles, roads, and etc.).

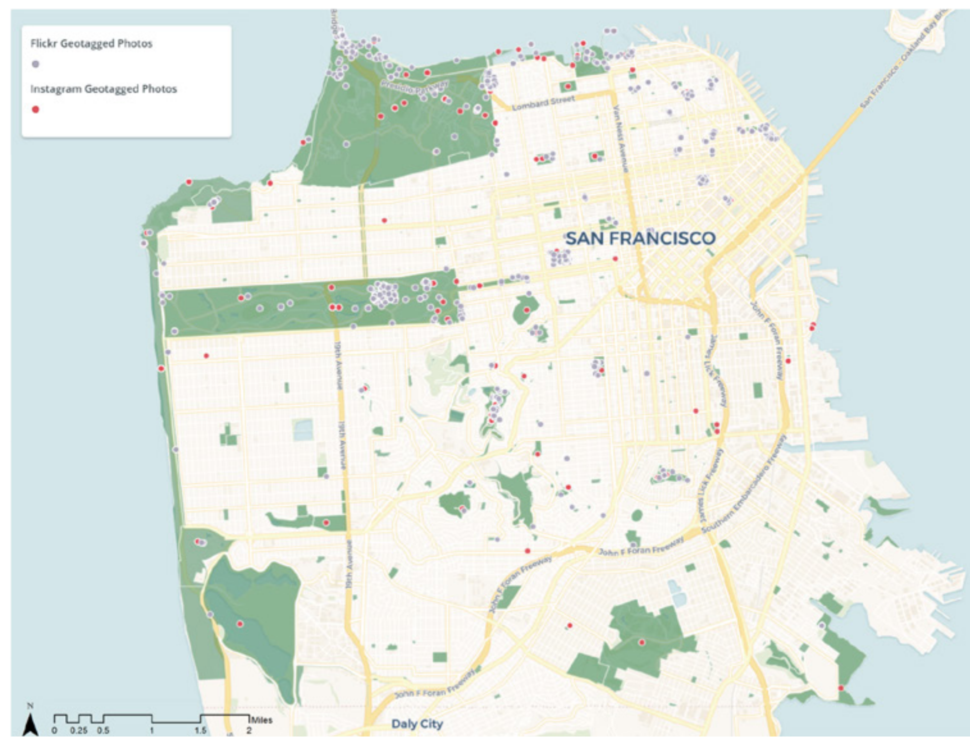
PHOTO CONTENT AND CLASSIFICATION

The results of photo content analysis show as following:

- ***The subjects of clustered Flickr and Instagram photos in the summer of 2017 were quite similar, though there was a little difference about small subjects.***
- ***The subjects of posted photos reflect what types of landscape may provide more aesthetic and recreation value.***
- ***The spatial pattern of different subject photos indicates cultural ecosystem services provided by each UGS are different due to different site-specific features.***



WATER LANDSCAPE SUBJECT



SURROUNDINGS OF BUILDINGS SUBJECT

■ Figure 3.8. The example of comparing CES contributed by different site-specific features at individual UGS scale. The compositions of geotagged photos of different subjects were different in Alamo Square Park and Fort Point National Historical Site.

Table 3.2. Hypotheses of the relationship between site-specific features and CES provision indicators.

	Independent Variables	Relationship with Dependent Variables (+/-)		References
		Photo counts	Photo counts per hectare	
Biophysical Features	Size (hectare)	+		Result 3.1 and Scopelliti et al., 2016; Bolund and Hunhammar, 1999; Kaczynski et al., 2008
	Landscape Shape index (LSI)	-		Frank et al., 2013
	Percentage of woody vegetation	+		Kaczynski et al., 2008; McGinlay et al., 2017; Breuste et al., 2013
	Percentage of vegetation (woody vegetation and grass/meadow)	+		
	Percentage of impervious surface (with building footprints)	-		
	Waterbody	+		
	Simpson's Diversity Index (SIDI)	+		de la Fuente de Val et al., 2006; Frank et al., 2013
Accessibility	The number of bus stops	+		Veitch et al., 2017; Timperio et al., 2008; Kaczynski et al., 2008; Xu et al., 2017; Tan and Samsudin, 2017
	Nearest distance to BART Stations	+		
	Bike Parking Facilities	+		
Attractions	The number of landmarks	+		Giedych and Maksymiuk, 2017
	Sports facilities	+		Veitch et al., 2017; Timperio et al., 2008; Kaczynski et al., 2008
	Cultural & entertainment spots	+		Giedych and Maksymiuk, 2017
	Shore View	+		Result 3.1
Surrounding Built-up Structure	Average height of surrounding buildings (meter)	+		Canter, 1983; Bonaiuto et al. , 2003
	Density of surrounding buildings (100 square meters of building area per hectare)	+		
Demographics	Population Density (people per hectare)	+		Cohen et al., 2010; Andersson et al., 2015

Based on previous analysis, observation and literature review, I proposed a series of hypotheses about the relationship between the site-specific features (Table 2.1) and CES provision in UGS quantified by photo counts and photo counts per hectare (Table 3.2). To find the driven factors of the spatial patterns of CES, I constructed multivariate ordinary least squares (OLS) linear regression models with different sets of indicators (Table 2.1) in Jupyter Notebook (Project Jupyter, 2014). Photo counts and photo counts per hectare as proxies of CES provided by UGS were two dependent variables in the models. Only Instagram geotagged photos were used in statistical analysis for fewer zero photo counts in UGS contributed by the larger number of photos.

■ **Table 3.3. Multivariate ordinary least squares (OLS) regression models for two metrics of cultural ecosystem service (CES) provided by urban green space (UGS) within 4 units of AICc from the lowest AICc OLS model.**

Number	Models	AICc	R^2	Ra^2	Δ AICc
Dependent variable:	***Photo Counts in each UGS***				
1	size + waterbody + the number of landmarks	4487.7	0.68	0.67	0.00
2	size + waterbody + the number of landmarks + LSI + sports facilities	4488.9	0.68	0.68	1.16
3	size + waterbody + the number of landmarks + Sqrt(the number of bus stops) + density of surrounding buildings	4489.7	0.68	0.68	1.97
4	size + waterbody + the number of landmarks + Sqrt(the number of bus stops) + Sqrt(population density)	4489.9	0.68	0.68	2.15
5	size + waterbody + the number of landmarks + Sqrt(the number of bus stops) + sports facilities	4490.0	0.68	0.68	2.30
Dependent variable:	***Photo Counts in each UGS (without "size" as an independent variable)***				
1	waterbody + the number of landmarks + LSI + cultural & entertainment spots - density of surrounding buildings	4564.6	0.57	0.57	0.00
2	waterbody + the number of landmarks + LSI + cultural & entertainment spots - Sqrt(population density)	4566.6	0.57	0.56	1.98
3	waterbody + the number of landmarks + LSI - density of surrounding buildings	4567.1	0.56	0.56	2.53
4	waterbody + the number of landmarks + LSI + cultural & entertainment spots - Sqrt (% impervious surface - buildings included)	4567.3	0.57	0.56	2.74
5	waterbody + the number of landmarks + shore view	4567.6	0.55	0.55	2.99
6	waterbody + the number of landmarks + LSI - Sqrt(population density)	4568.2	0.56	0.55	3.57
7	waterbody + the number of landmarks + LSI + cultural & entertainment spots - average height of surrounding buildings	4568.2	0.57	0.56	3.60
8	waterbody + the number of landmarks + cultural & entertainment spots	4568.3	0.55	0.55	3.67
9	waterbody + the number of landmarks	4568.5	0.55	0.54	3.95
Dependent variable:	***Photo Counts per Hectare in each UGS***				
1	waterbody + density of surrounding buildings	4072.0	0.09	0.09	0.00
2	waterbody + Sqrt(population density)	4075.0	0.08	0.07	2.98
3	density of surrounding buildings	4075.4	0.07	0.06	3.32

The main results of regression analysis show as following:

- ***Size of UGS is the dominant independent variable to interpret the spatial pattern of CES, which reconfirms the observation in the first part of exploration.***

- ***"Waterbody", "number of landmarks", "landscape shape index " (LSI), "cultural & entertainment spots", "shore view" are also positive and statistically significant for photo counts without "size" as an independent variable, while "density of surrounding buildings", "population density" and "average height of surrounding buildings" are negatively associated with photo counts.***

- ***"Waterbody", "density of surrounding buildings" and "population density" are positive and statistically significant photo counts per hectare though they are unable to interpret the changes of the dependent variable well ($R^2 \leq 0.09$).***

CONCLUSIONS & CONTRIBUTION

- *Geotagged photos from social media have increasingly been used for mapping and assessing landscape values and CES that in turn can support environmental planning and landscape design. The analysis of geotagged photos can also be used as a complementary technique of traditional revealed preferences approaches such as interviews and surveys to assess CES.*

- *Dynamic mapping is beneficial to detect CES changes and anomalies at different time scales. Detecting the big event in UGS, for example, helps planners know quantitatively about recreation and social value in the field of CES.*

- *The inconsistency between the spatial pattern of CES and extreme event days in UGS, as well as various compositions of different subjects of photos in UGS demonstrated the CES composition across all UGS in San Francisco varied a lot due to different site-specific features in UGS. UGS located in less dense area with a larger size, more disaggregated shape, more landmarks, waterbody, cultural and recreation spots, as well as a potential sea view were probably to get more photo counts, while UGS with waterbody, higher building and population density in surrounding neighborhoods could have more photo counts per hectare in San Francisco according to the results of regression analysis.*

- *The effects of different site-specific features on CES provision were measured quantitatively and could be compared with each other, which help landscape planners to know the trade-offs across different options of landscape changes.*