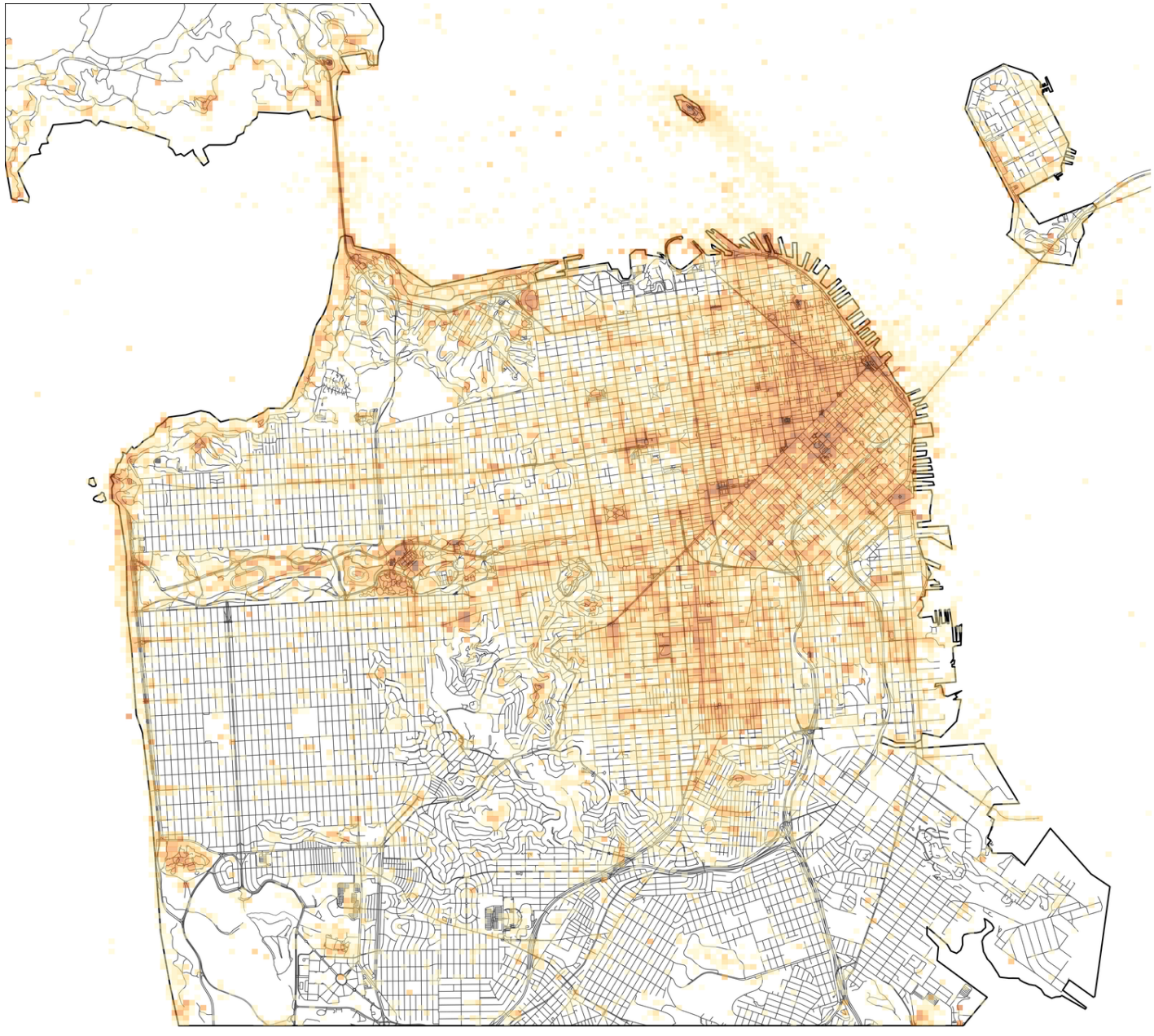
Aalto University
School of Science
and Technology

Social Map Annotation

Géraud Le Falher

November 26, 2013



Photos distribution

Pre processing

- Avoid spurious photos by removing tags of the same user set almost in the same place at almost the same time.
- Suppress 25% of tags but does not significantly affect the tags distribution.

Some statistics

- We want to find tag that are discriminative descriptor of location.
- Divide the city in a $g \times g$ grid and count tags in each cell.
- Entropy of tag t: $H_t = -\sum_{c=1}^{g^2} t(c) \log t(c)$

Lowest	Highest
californiaacademyofsciences	instagramapp
conservatoryofflowers	square
sfmoma	squareformat
deyoung	iphoneography
deyoungmuseum	unitedstates
ucsfschoolofdentistry	iphone
attpark	foundinsf
cityhall	california
japantown	sanfrancisco

More statistics

Kullback Leibler divergence of tag t with respect to the background distribution b: $D(t||b) = -\sum_{c=1}^{g^2} t(c) \log \frac{t(c)}{b(c)}$

Highest	Lowest
zoo	san
treasureisland	ca
conservatoryofflowers	instagramapp
oceanbeach	squareformat
ucsfschoolofdentistry	square
ucsf	iphoneography
japantown	unitedstates
sfmoma	usa
castro	sf
deyoung	california
attpark	sanfrancisco

Even more statistics

Distribution of the distance from tags to their center of gravity $G_t = (\bar{x}_i, \bar{y}_i)$: $\mathcal{D}(t) = \{||p_i - G_t||^2\}_{i=1,...,T}$

std($\mathcal{D}(t)$) as a function of mean($\mathcal{D}(t)$)

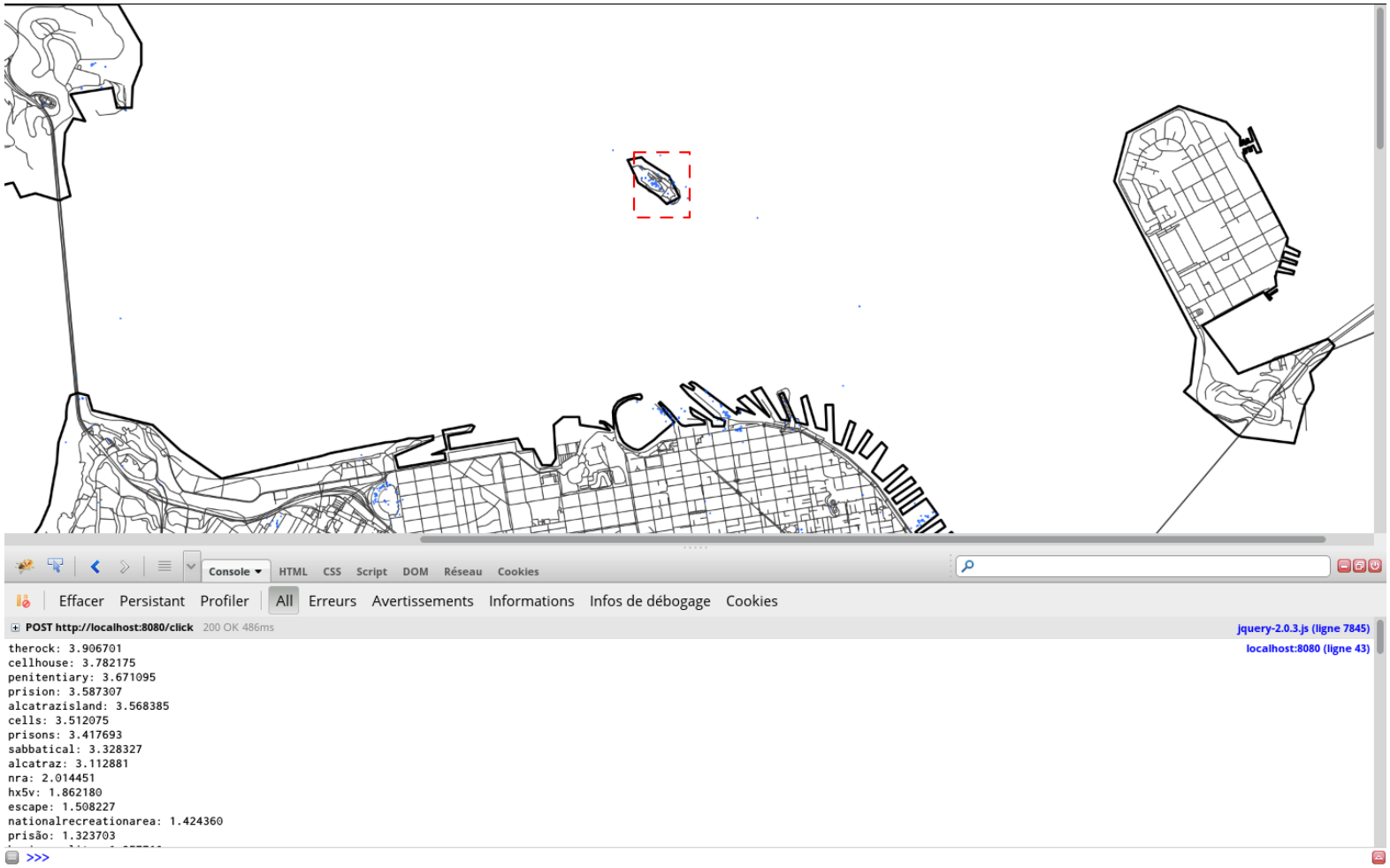
One last statistics

Kulldorff Spatial Scan Statistic:
$d(t(R), b(R)) = t(R) \log \frac{t(R)}{b(R)} + (1 - t(R)) \log \frac{1-t(R)}{1-t(R)}$ where R is one of the g^4 possible rectangular region.

Highest	Lowest
sutrobaths	sanfrancisco
legionofhonor	amaro
missiondolores	rise
byobw	california
palaceoffinearts	sierra
twinpeaks	xproii
dolorespark	sf
conservatoryofflowers	iphoneography
flickrhq	square
surf	squareformat
japaneseteagarden	instagramapp
lombardstreet	normal
exploratorium	usa

How to use this

Compute statistics for all tags with enough support and retrieve the most significant ones when a user query a region.



Next

- Improve the demonstration, maybe by weighting different statistics.
- Cover the map with tags.
- Try different parameter for computing the statistics.

Conservatory of flowers

Thank you