

MRTT XIV. köz- és vándorgyűlése, Nagyvárad
2016. szeptember 15-16.



Városi differenciák vizsgálata big data környezetben

Jakobi Ákos

**Eötvös Loránd Tudományegyetem
Regionális Tudományi Tanszék**

Privacy, Big Data, and the Public Good

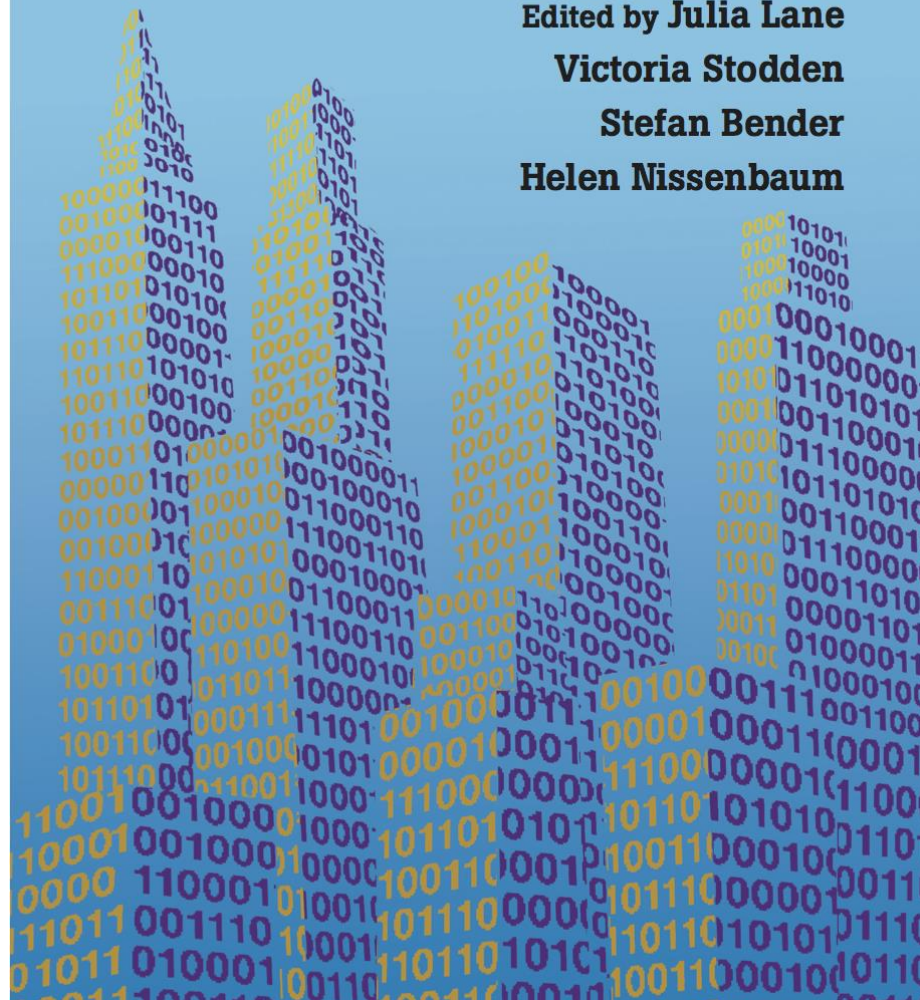
Frameworks for Engagement

Edited by Julia Lane

Victoria Stodden

Stefan Bender

Helen Nissenbaum



Mi az a big data?

- A „big data” (szabad fordításban „óriási adathalmaz”) arra a hatalmas adatmennyiségre utal, amely információs világunkban gyors tempóban és folyamatosan keletkezik
- Feldolgozása a hagyományos kapacitásokkal és módszerekkel már-már megoldhatatlan kihívást jelent
- Nem csak mennyiségi értelemben vett adatrobbanás, de minőségi is
- A nagy adatbázisok jellemzői:
 - Nagy adatmennyiség. Maga a méret a probléma lényege (Loukides 2010).
 - Változatosság, az adatfajták és források sokfélesége
 - Sebesség: az adatok keletkezésének gyorsaságára, a hasznosításhoz rendelkezésre álló időre vonatkozóan

Mi az a big data?

- Eleve digitálisan keletkezik, passzív módon termelődik a mindennapi digitális interakciók melléktermékeként
- A sokáig csak virtuális melléktermékként számon tartott napi információhalom épp akkor válik értékké, amikor a különböző adatokat sikerül összekötni, köztük összefüggéseket, felismerhető mintázatokat találni, s mindebből értékelhető következtetéseket levonni.
- Eltolódás az eseti megfigyelésektől a térben és időben folytonos megfigyelésig
- A társadalomtudományok számára valóságos aranybánya (vásárlási, munkabajárási, közlekedési és egyéb szokásaink adatai)

Big data források

- állami vagy egyéb intézményi források (pl.: elektronikus egészségügyi nyilvántartások, kórházi látogatások, biztosítási nyilvántartások, iskolai adatok, banki adatok, stb.)
- kereskedelmi vagy tranzakciós eredetűek (pl.: bankkártya tranzakciók, online tranzakciók, mobiltelefonos fizetések, stb.)
- fizikai szenzoros eredetűek is (pl. műholdképek, forgalomfigyelők, időjárás figyelők adatai, beléptető kapuk, stb.)
- nyomkövető eszközökből származók (pl. útvonal/követési adatok mobiltelefonoktól, GPS, stb.)
- online tartalmakból kiolvasott információk (pl. honlap látogatottság, online keresések, közösségi média adatok, önkéntes tartalmak, stb.)
- ...

Big data források

Az ENSZ Európai Gazdasági Bizottságának (2014) statisztikai munkacsoportja szerint

- Közösségi adatok (humán-eredetű információk), avagy „people to people” típusú adatok.
- Ezek az adatok csak lazán strukturáltak és gyakran irányítási kontrol nélküliek.
 - 1100. Közösségi hálók: Facebook, Twitter, Tumblr stb.
 - 1200. Blogok, hozzászólások
 - 1300. Személyes dokumentumok
 - 1400. Képek: Instagram, Flickr, Picasa stb.
 - 1500. Videók: Youtube stb.
 - 1600. Internetes keresések
 - 1700. Mobil adattartalom: szöveges üzenetek
 - 1800. Felhasználó által generált térképek
 - 1900. E-Mail

Big data források

Az ENSZ Európai Gazdasági Bizottságának (2014) statisztikai munkacsoportja szerint

- Hagyományos üzleti rendszerek adatai (folyamat által közvetített adatok), avagy „people to machine” típusú adatok
- Az adatokat határozottabb strukturáltság jellemzi, kapcsolati táblákkal, metaadatokkal
 - 21. Közhivatalok által szolgáltatott adatok
 - 2110. Orvosi/egészségügyi nyilvántartások
 - 22. Kereskedelem által létrehozott adatok
 - 2210. Kereskedelmi tranzakciók
 - 2220. Bank-/készletnyilvántartás
 - 2230. E-kereskedelem
 - 2240. Bankkártya/hitelkártya

Big data források

Az ENSZ Európai Gazdasági Bizottságának (2014) statisztikai munkacsoportja szerint

- Automatikus rendszerek adatai (gépek által közvetített adatok), avagy „machine to machine” típusú adatok
- Alapvetően a fizikai világ megfigyeléséből származnak és legtöbb esetben jól strukturáltak, ám méretük és keletkezési sebességük a tradicionális megközelítéseken messze túlmutat.

Big data források

Az ENSZ Európai Gazdasági Bizottságának (2014) statisztikai munkacsoportja szerint

31. Szenzor adatok

311. Rögzített szenzorok

3111. Otthon-automatizálás

3112. Időjárási, szennyezési szenzorok

3113. Közlekezési szenzorok, webkamerák

3114. Tudományos célú szenzorok

3115. Biztonsági videók, képek

312. Mobil szenzorok (nyomkövetés)

3121. Személyes (mobil helymeghatározás)

3122. Közúti (autók, teherszállítás)

3123. Vasúti (vonatok)

3124. Légi (repülőgépek)

3125. Vízi (hajók)

313. Műholdas adatok

3131. Topográfiai

3132. Hőmérsékleti

3133. Megfigyelési

3134. Meteorológiai

3135. Egyéb

32. Számítógépes rendszerekből származó adatok

3210. Naplók (log-ok)

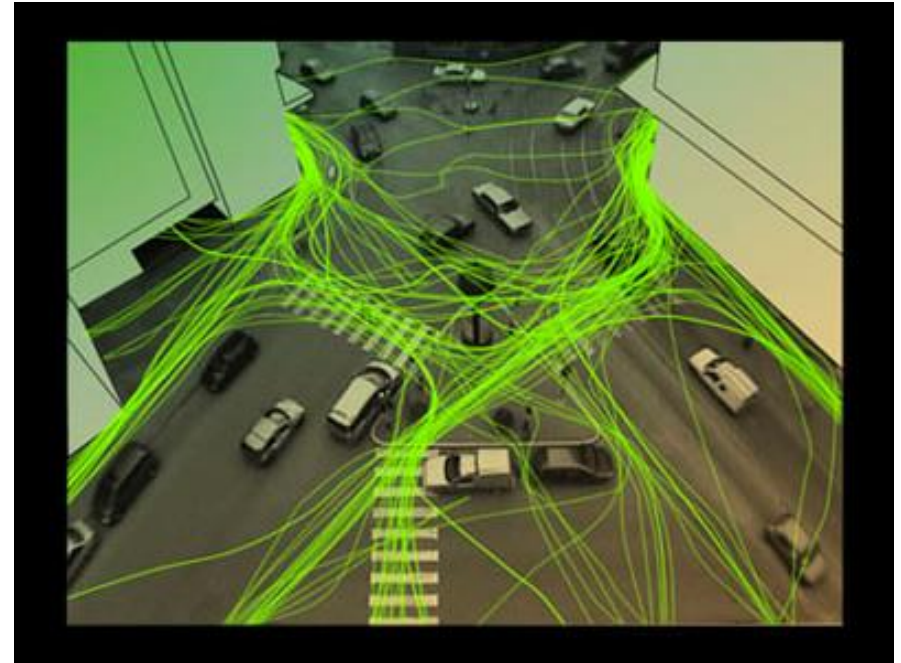
3220. Webes naplók (web log-ok)

Néhány példa ízelítőül

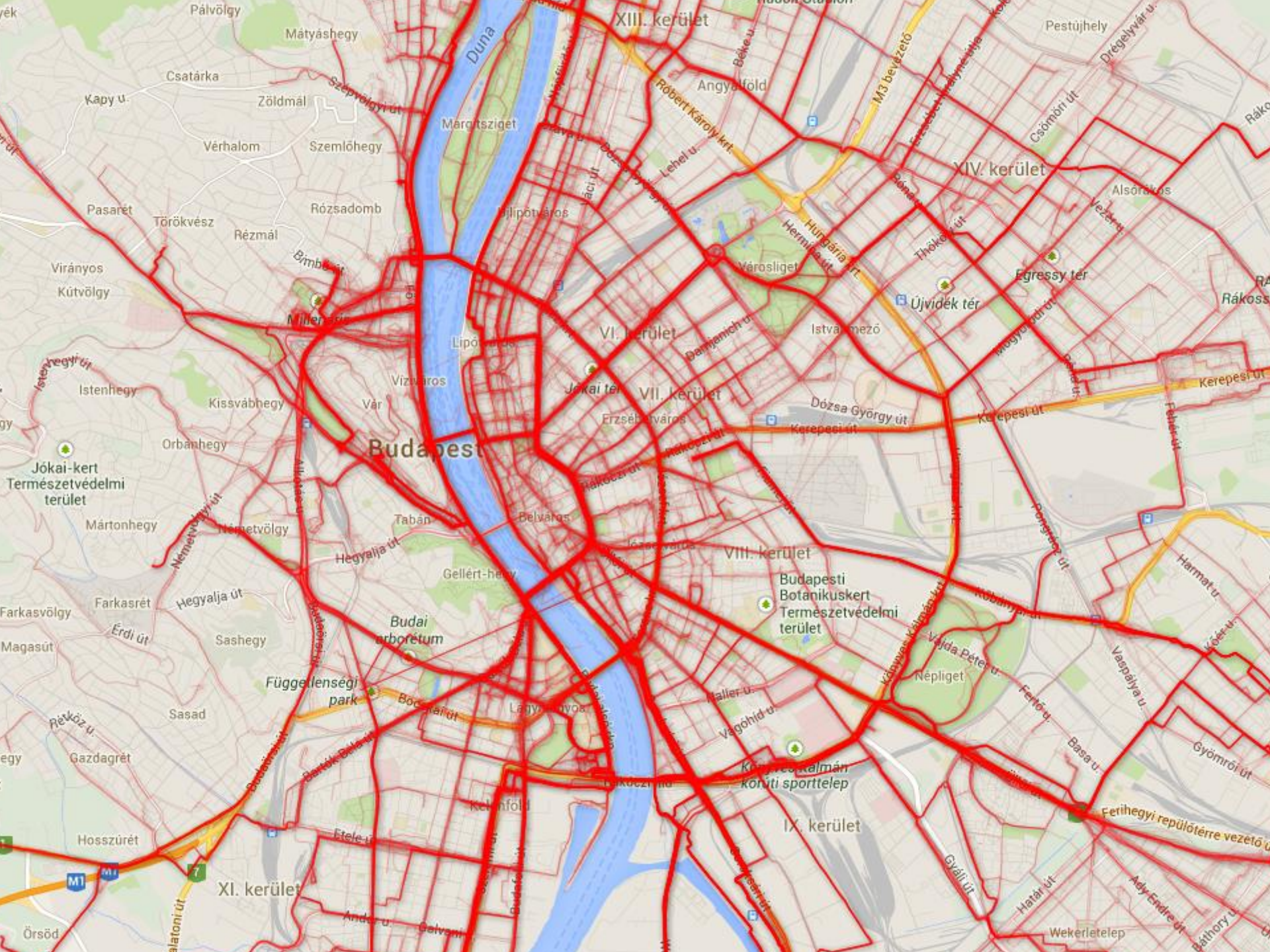
- Az áramszolgáltató tudja, hogy a fogyasztó reggel mikor kel fel és este mikor fekszik le.
- Digitális TV szolgáltató: tudjuk mit nézel
- GPS, adatrögzítők az autóban: tudjuk merre jártál
- Fizetőkapuk: tudjuk mikor voltál ott
- Weboldal statisztikák: preferenciák modellezése

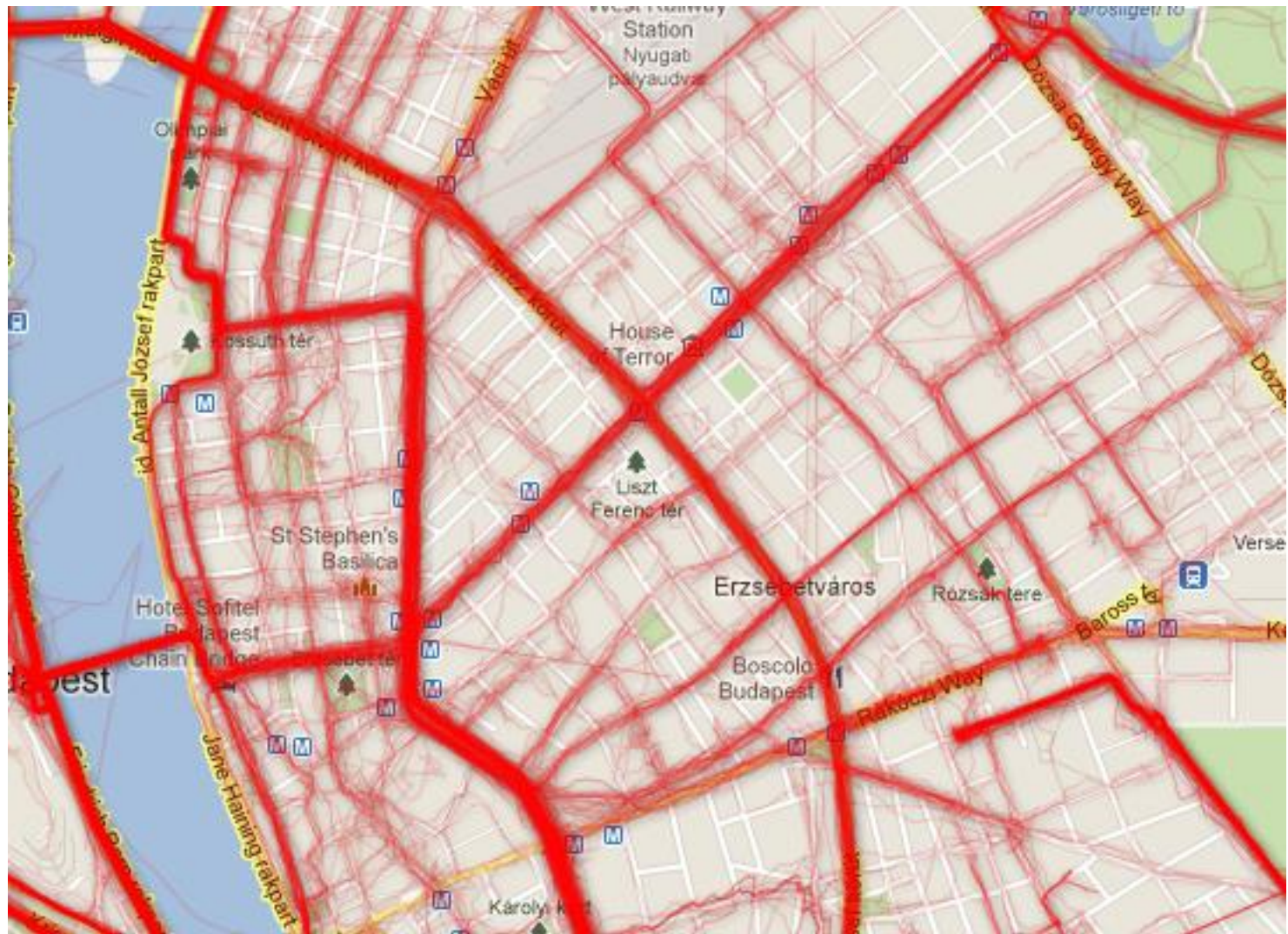
A területi (városi) kutatások lehetőségei

- Úton-útfélen hagyott térbeli tartalommal is rendelkező direkt vagy indirekt digitális nyomok
- Direkt digitális nyomok:
 - okostelefonok helyalkalmazásai
 - egyes honlapok forrása kiegészítve célzott geotag-ekkel
 - ...
- Indirekt digitális nyomok
 - nem szándékoltan, de mégis nagy számban keletkező spontán téradat



**Városi
térpályák**

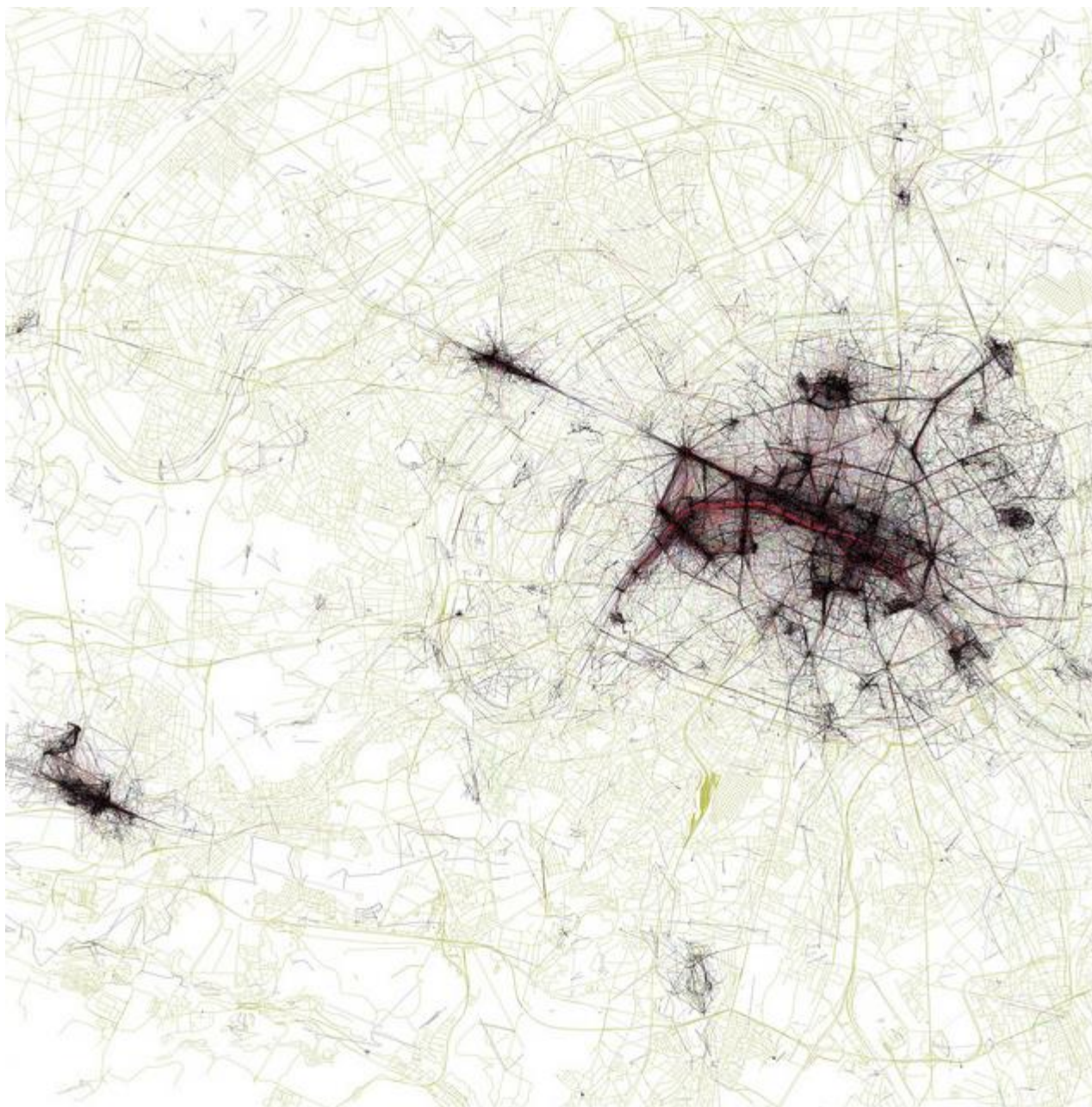




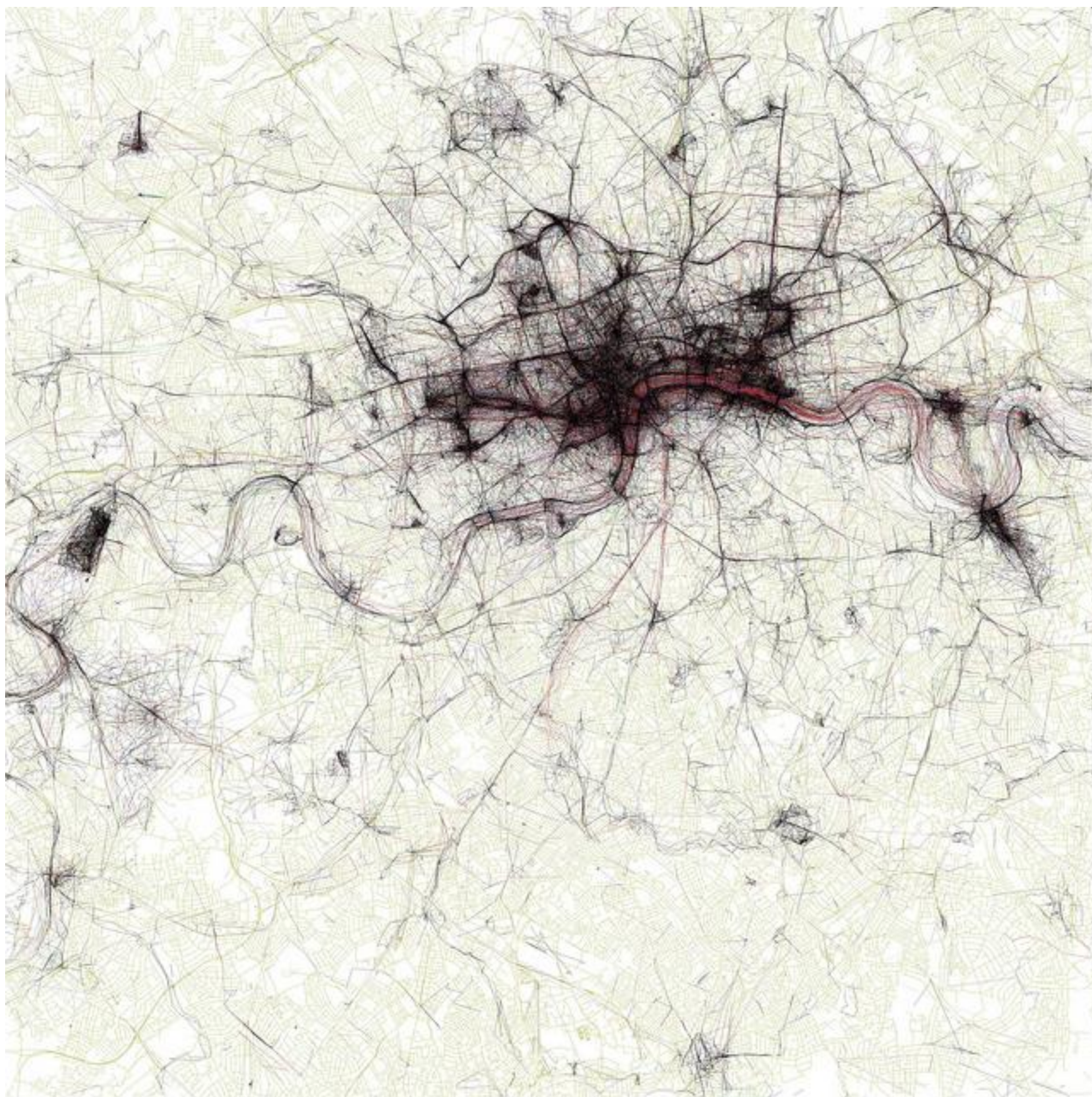
Csoportos térpályák „big data” adatok alapján



Fotókat készítő turisták
mozgási útvonalainak területi mintázata
Manhattan déli részén
(Crandall et. al., 2009)



Párizs

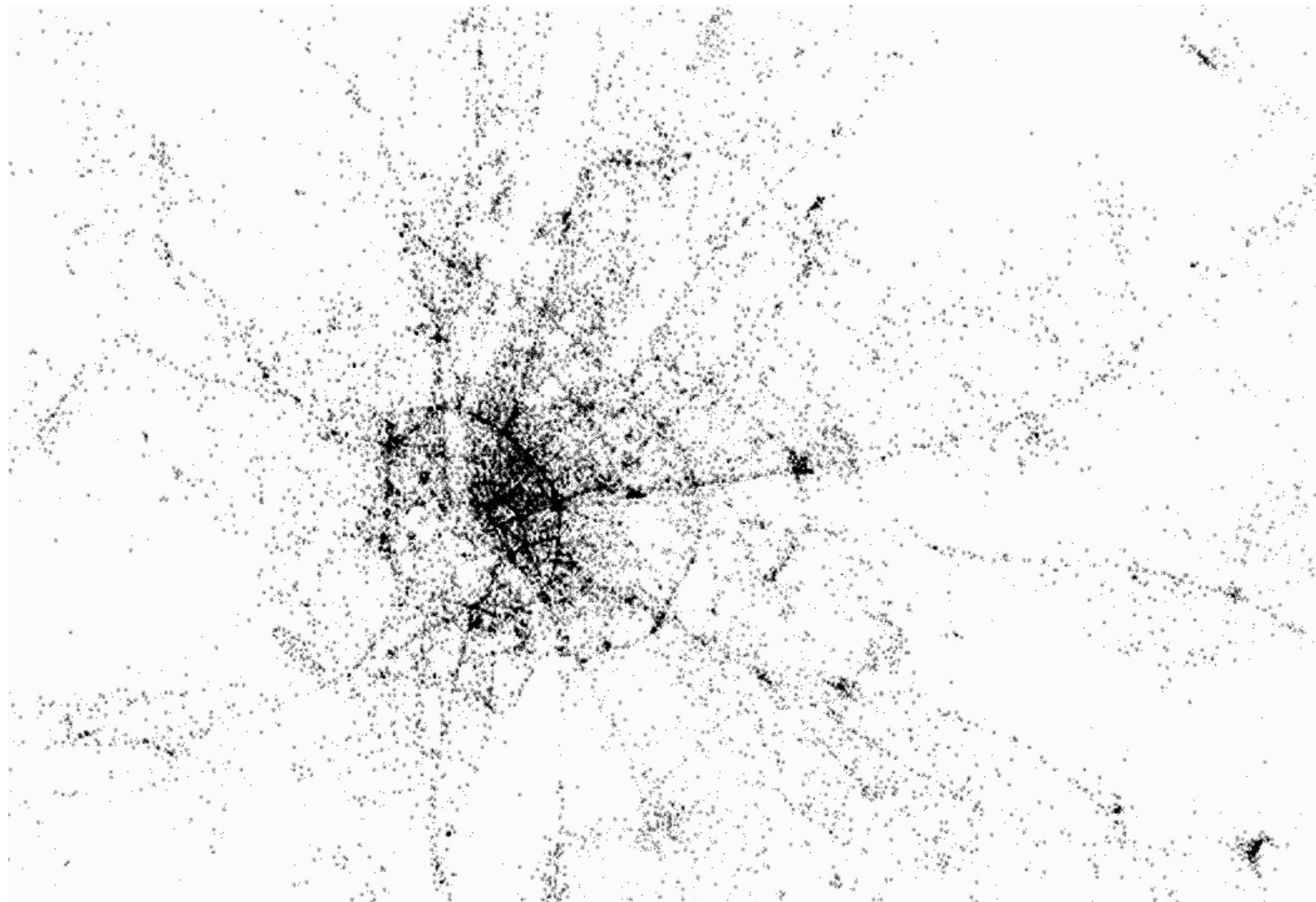


London



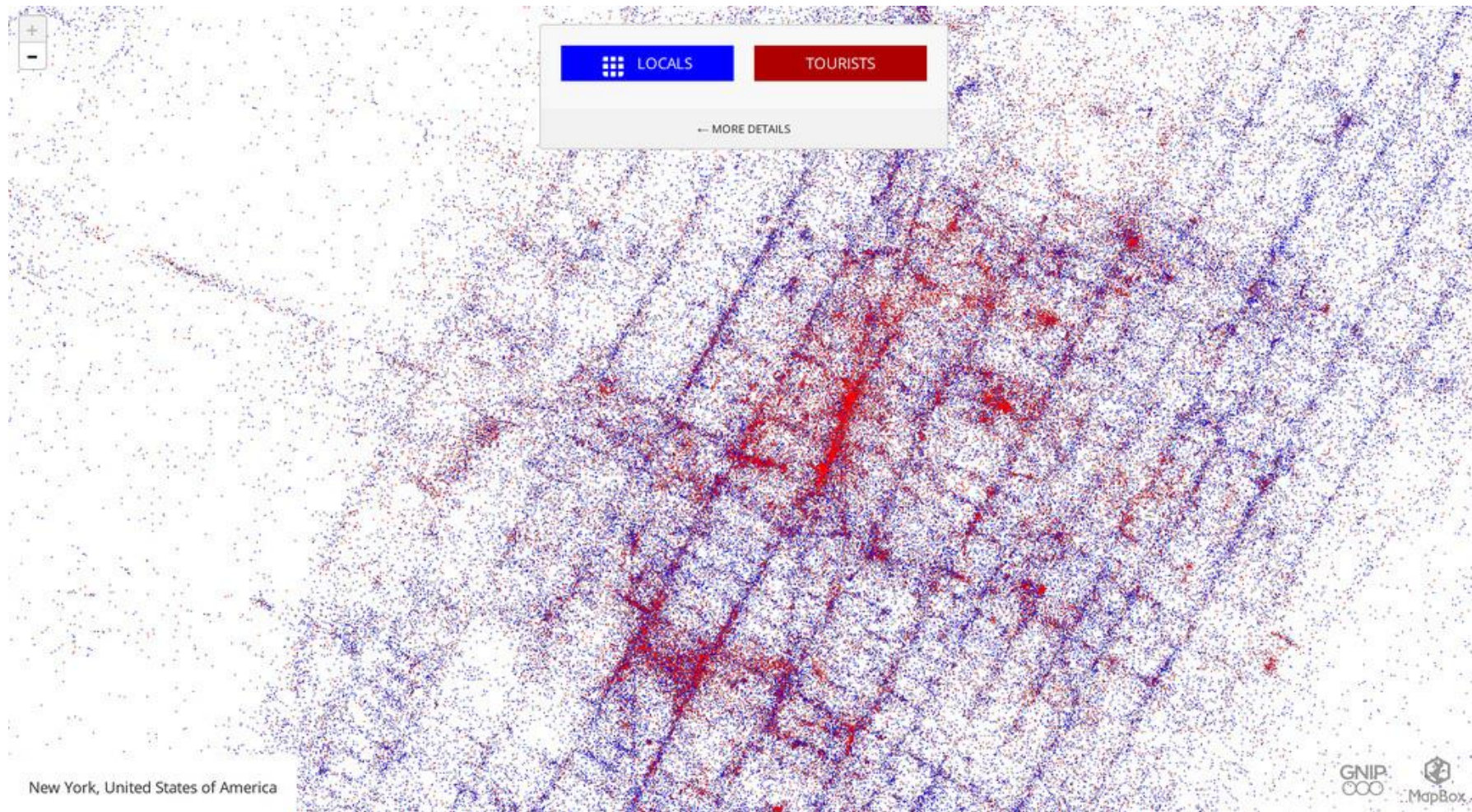
Budapest

Geotagelt Twitter bejegyzések sűrűsége Budapesten

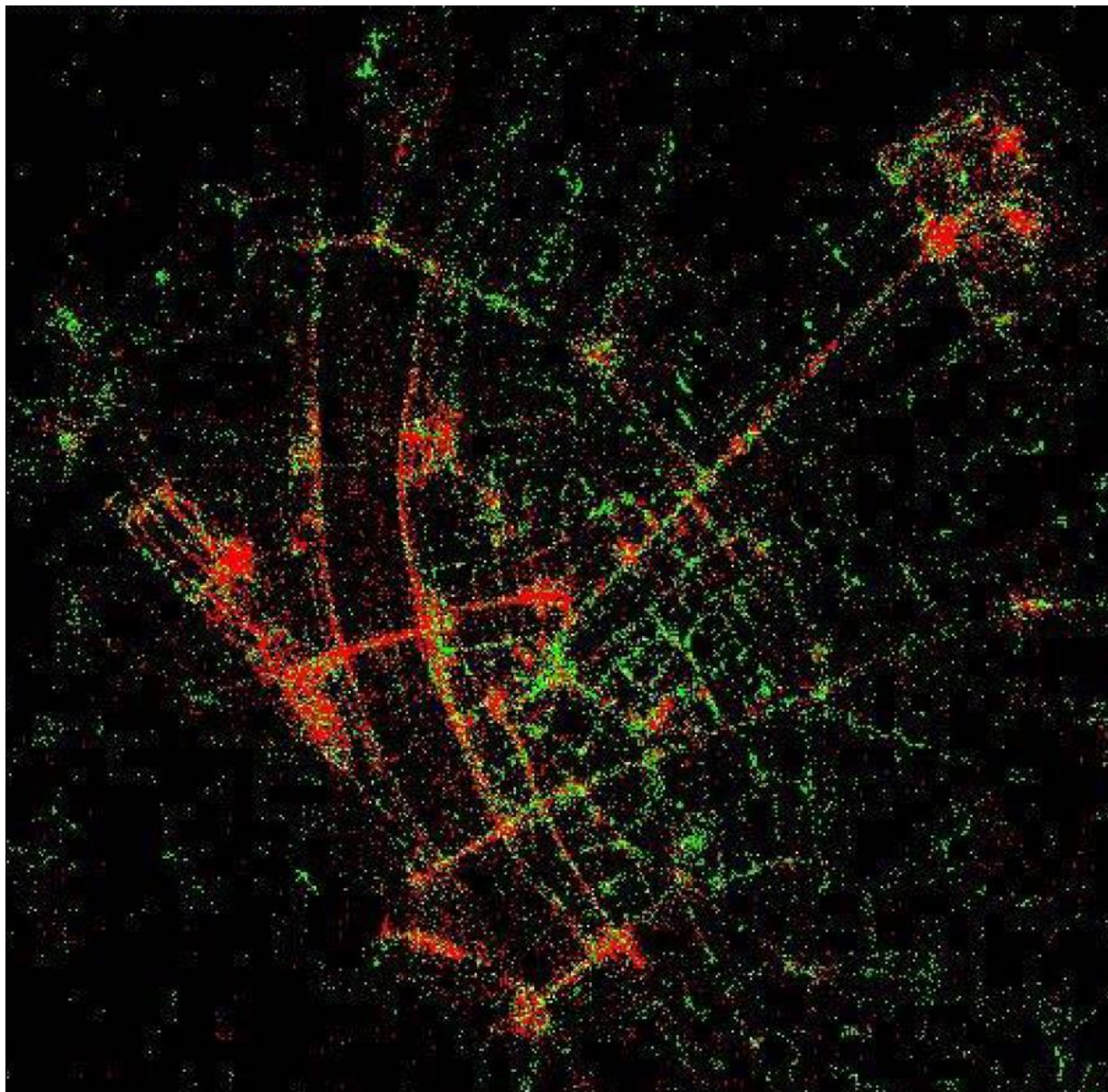


Fischer (2013)

Helyi és nem helyi lakosok (turisták) Flickr fotó-bejegyzéseinek sűrűsége Manhattan-ben

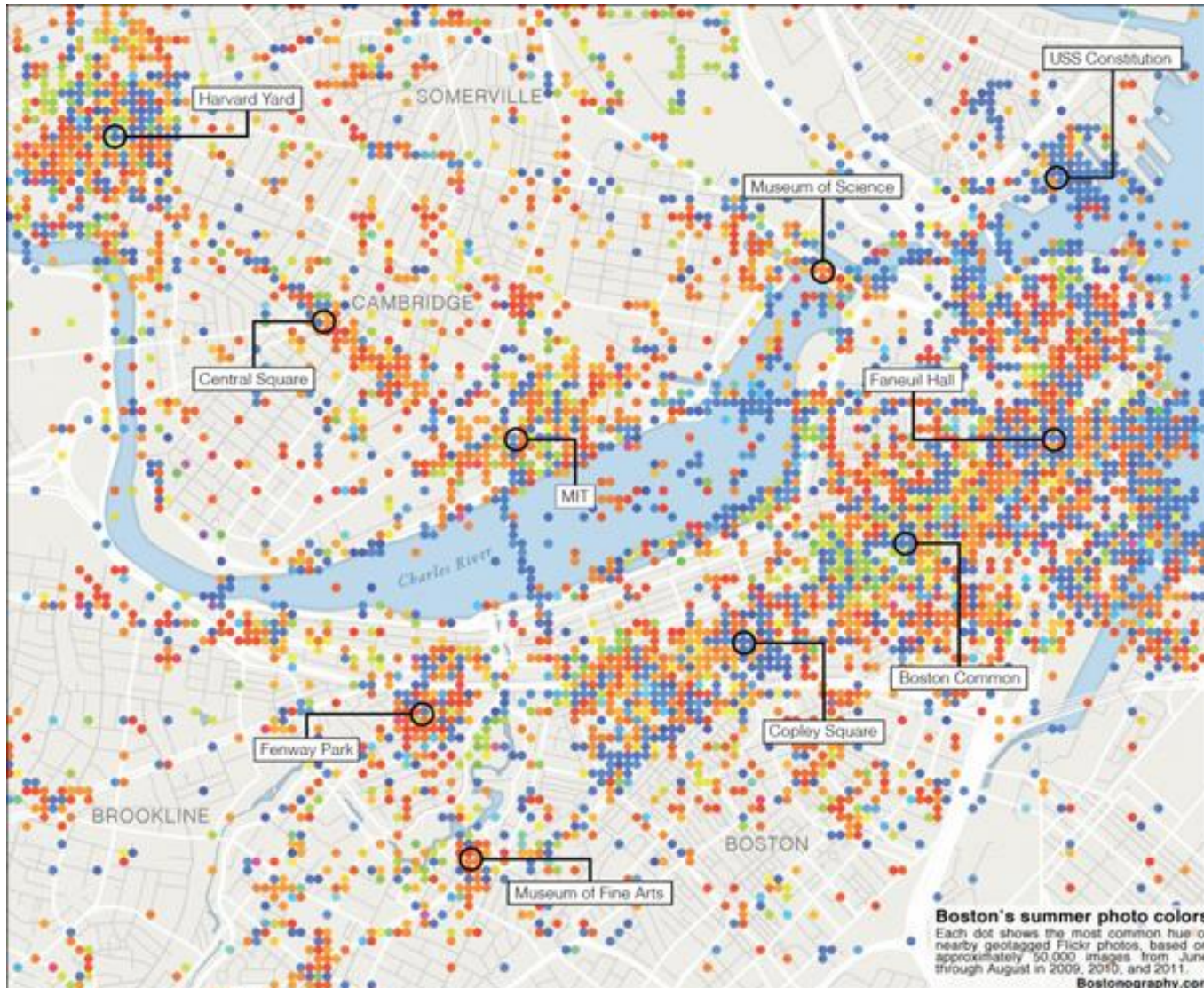


Helyi és nem helyi lakosok (turisták) Flickr fotó-bejegyzéseinek sűrűsége Budapesten



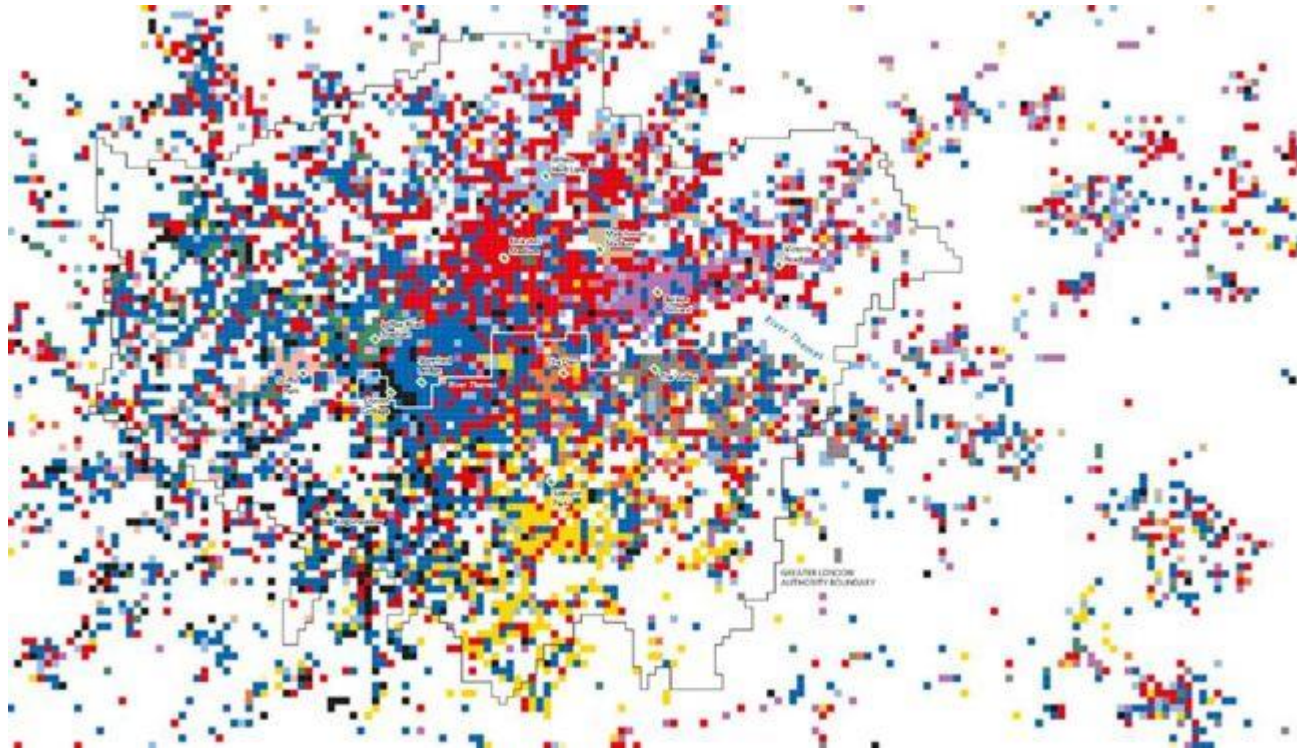
Forrás:
Kádár B. – Gede M. 2012

Geolokalizált Flickr fényképek jellemző színárnyalata Boston városának egyes körzeteiben (június-augusztusban)



Woodruff (2011)

A legnépszerűbb futball klubok a Twitter bejegyzések alapján Londonban



PREMIER LEAGUE

CHELSEA

Stamford Bridge

2013-14 record: 25-7-6 (3rd place)

WINS
LOSSES
DRAWS

ARSENAL

Emirates Stadium

2013-14 record: 24-7-7 (4th)

TOTTENHAM

White Hart Lane

2013-14 record: 21-6-11 (6th)

CRYSTAL PALACE

Selhurst Park

2013-14 record: 13-6-19 (11th)

WEST HAM

Boleyn Ground

2013-14 record: 11-7-20 (13th)

FULHAM

Craven Cottage

2013-14 record: 9-5-24 (19th)

FOOTBALL LEAGUE CHAMPIONSHIP

Queens Park Rangers

Charlton Athletic

Millwall

LEAGUE ONE

Brentford

Leyton Orient

LEAGUE TWO

Dagenham & Redbridge

AFC Wimbledon

Egy big spatial data empíria: webről geokódolt tartalmak térbeli eloszlásvizsgálata

- Weboldalak és azok tartalmainak földrajzi azonosítása
- Az internet strukturálatlan tartalmainak térbeli értékelése
- A világhálón számos olyan tartalom lelhető fel, amely földrajzilag azonosítható ugyan, de geotag nem kapcsolódik hozzá
- Adatbányászati technikákkal földrajzi helyzetet jelölő adatok keresése, így a hozzá kapcsolódó tartalom földrajzi azonosítása

Kulcsszó-keresés és címazonosítás

halászat

1011 BUDAPEST, ANGELO ROTTA RAKPART, MAGYARORSZÁG



Haltermelők Országos Szöv...

Address: 1126 budapest vöröskő u. 4/b, 1222 budapest nagyotétnyi út 48.
Coordinates: 47.495752, 19.009661
Website: <http://mezogazdasag.iclközepmagyarorszag.hu>
Category: It
Quality: undefined

CONTENT

FEEDBACK

CLOSE

Köcsöztes Mezőgazdaság Növénytermesztés Oktatás, szakkepzés, tanfolyam Raktárház, rendezvényszervezés Rendezvények, programok, szórakozás Rovarirtás, takarítás Sport, sporteszközök Szállítmányozás, fuvarozás, utaztatás Szolgáltatás Szórakozás, társkeresés Utazás, üdülés Webáruház, webshop, kereskedelem Linkajánló uszodatechnika penztargepek szerveze szállás Veszprém megye háztartási gépek, beépített sütő Sportegyesületek. sportszövetségek - Szombathelyi Haladás Vasutas Football Kft. Klíma, fűtés, hűtés, árnyékolás - Bakus Hűtő-Klíma Kereskedelem fordítás angolra rágcsálóirtás használt konténer Állateledel, táplálék kiegészítők - Bell Állateledel Bolt budapest nightclub Rlp professzionális iroda takarítás Lomtalanítás - TECHNOÉPSZER Szolgáltató és Kereskedelmi Kft. Haltermelők Országos Szövetsége - Halászat

Cím: 1126 Budapest Vöröskő u. 4/b Telefon: (1)355-7019 Fax:(1)214-2643 Web: <http://www.haltermosz.hu> Megye Bács-Kiskun megye Baranya megye Békés megye Borsod-Abaúj-Zemplén megye Budapest Csongrád megye Fejér megye Győr-Moson-Sopron megye Hajdú-Bihar megye Heves megye Jász-Nagykun-Szolnok megye Komárom-Esztergom megye Nógrád megye Pest megye Somogy megye Szabolcs-Szatmár-Bereg megye Tolna megye Vas megye Veszprém megye Zala megye Magyarország halászat céglista | Észak-Magyarország halászat céglista | Dunántúl halászat céglista | Alföld halászat céglista | © 2008. ICLKözépMagyarország.hu

Keywords
mezőgazdaság, halászat, haltermelők országos szövetsége, internet céglista, céglista, cégladatbázis

Description

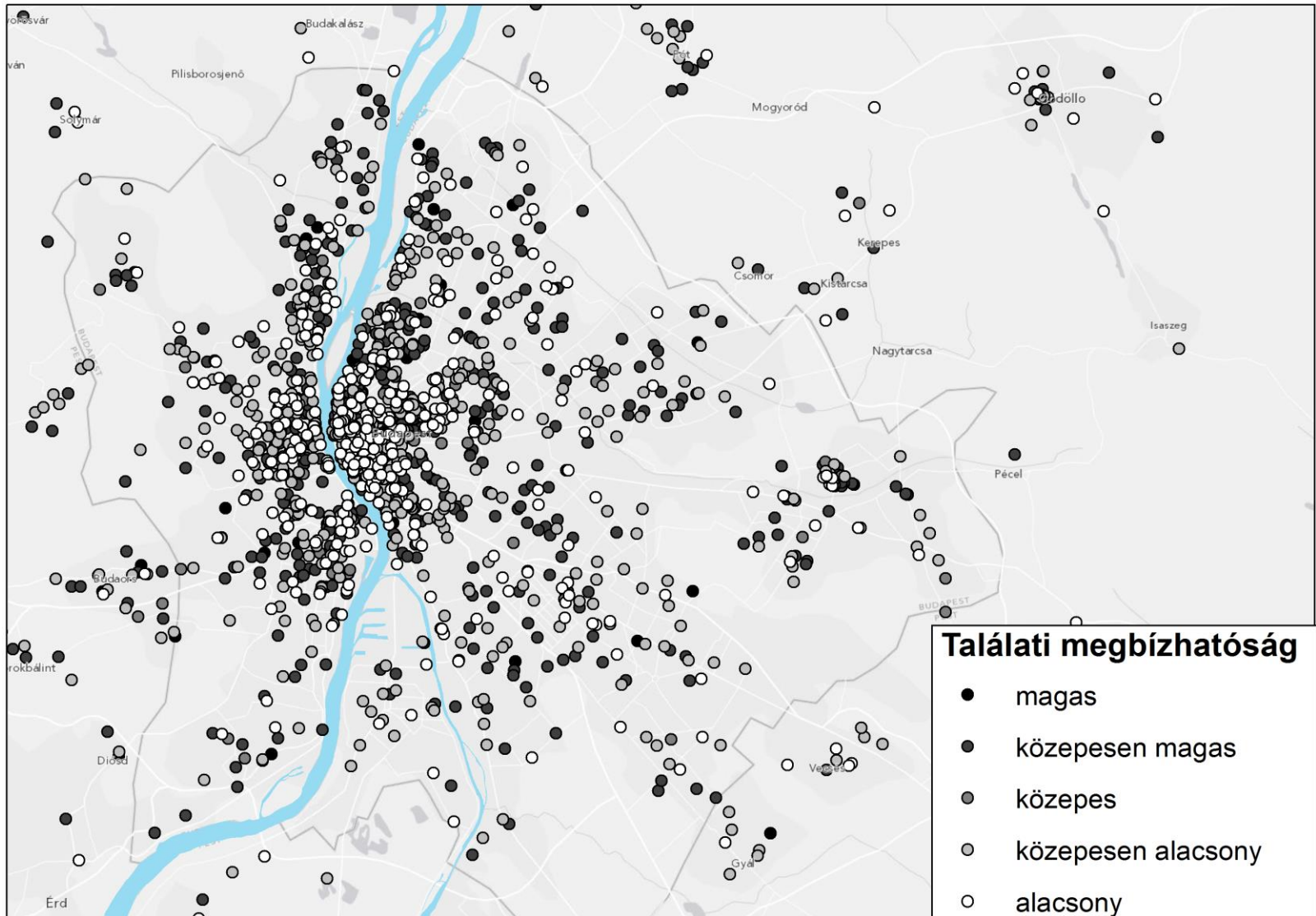


Google

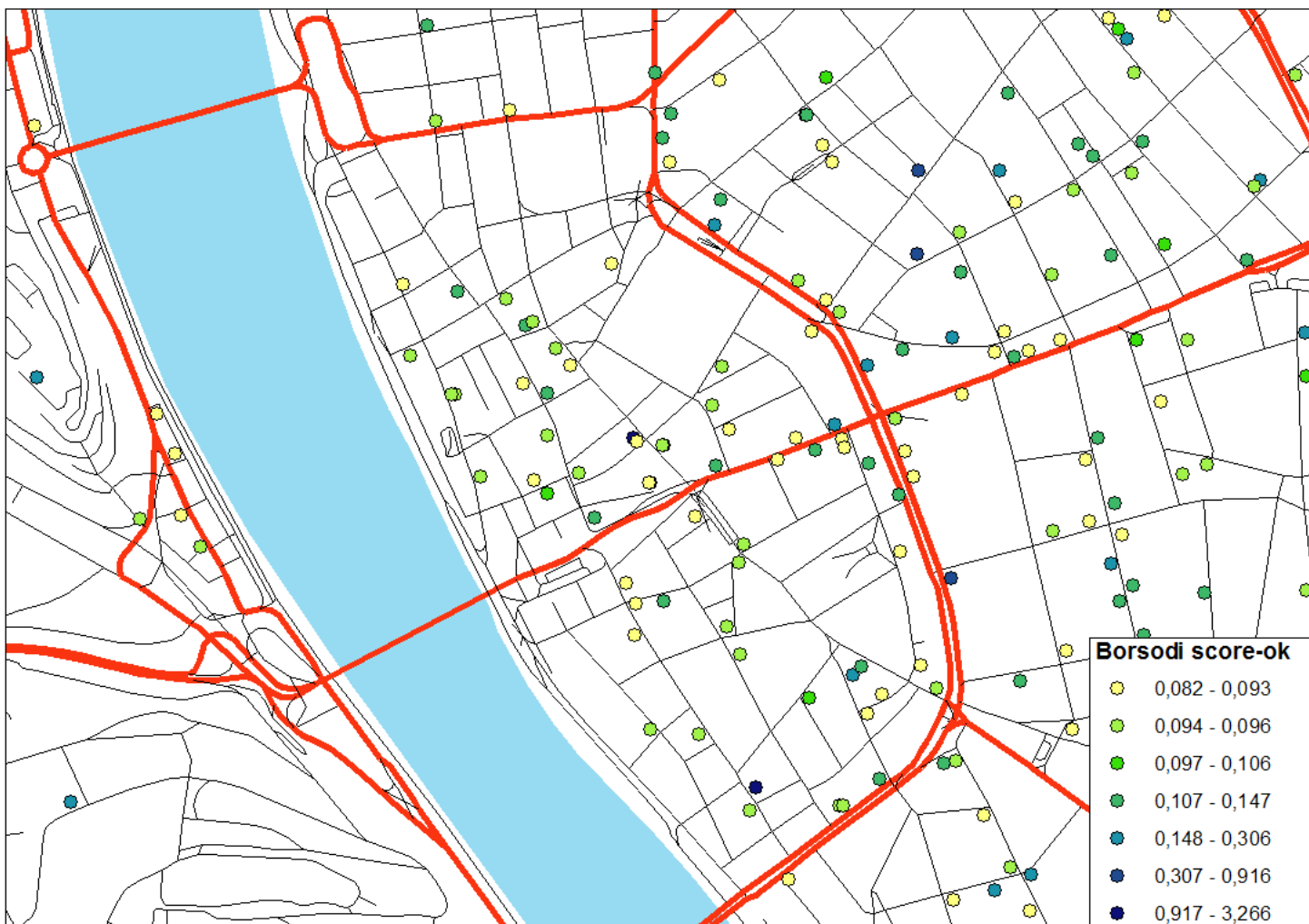


thumbnail coming soon

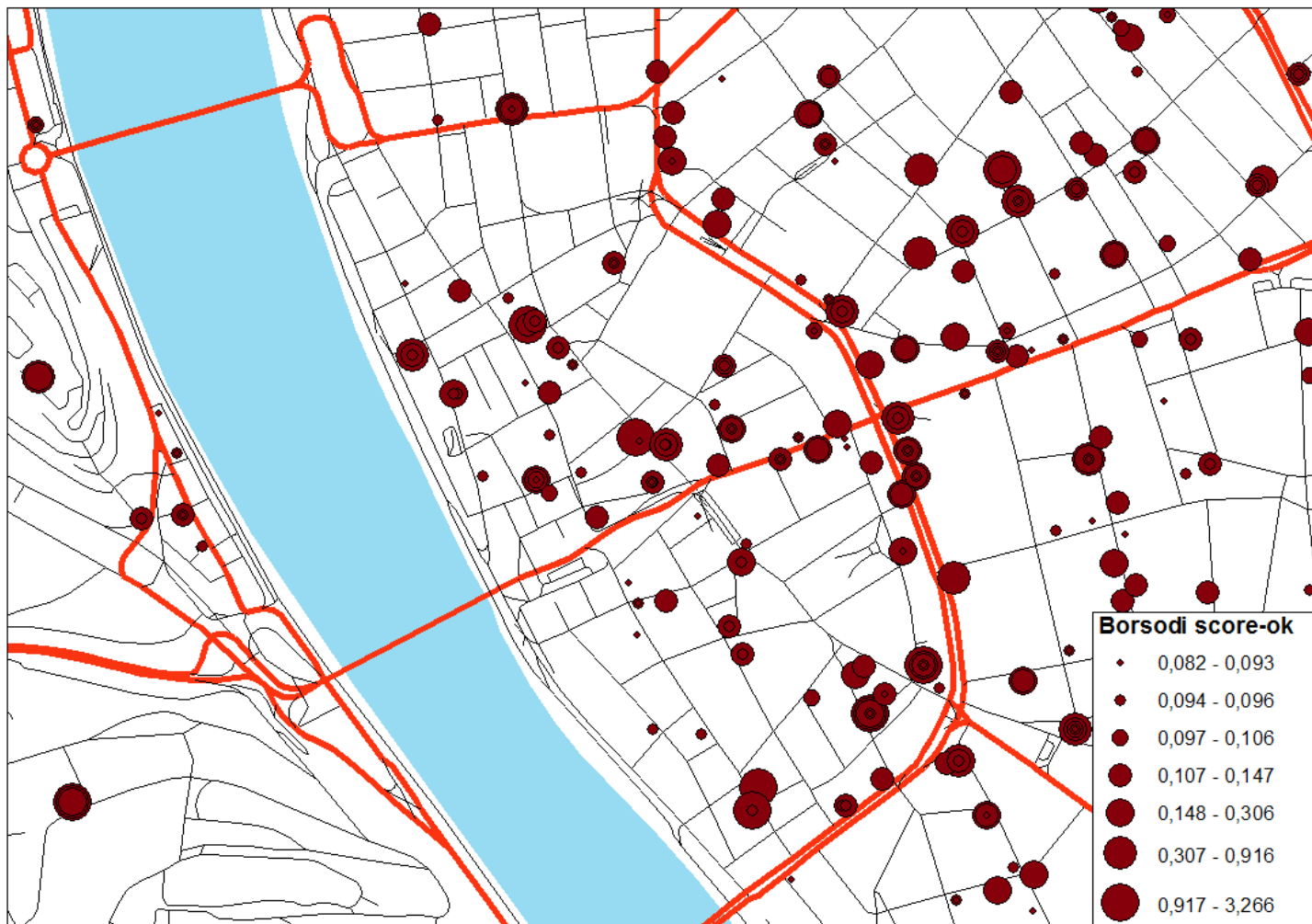
A „borsodi” kulcsszó előfordulásai a találatok megbízhatósága szerint Budapesten



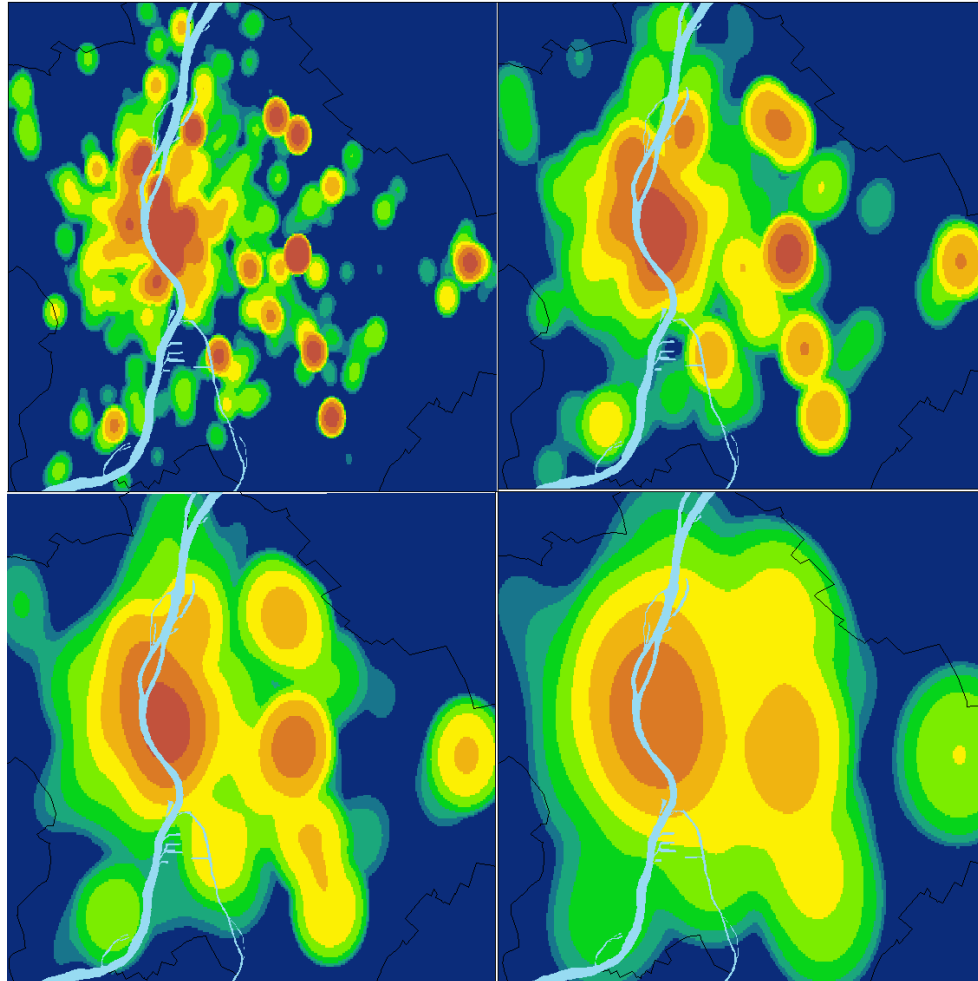
A „borsodi” kulcsszó előfordulásai a találatok megbízhatósága szerint Budapest belvárosában



***Egy címen több találatot is feljegyezhattünk:
A „borsodi” kulcsszó előfordulásai Budapest belvárosában***

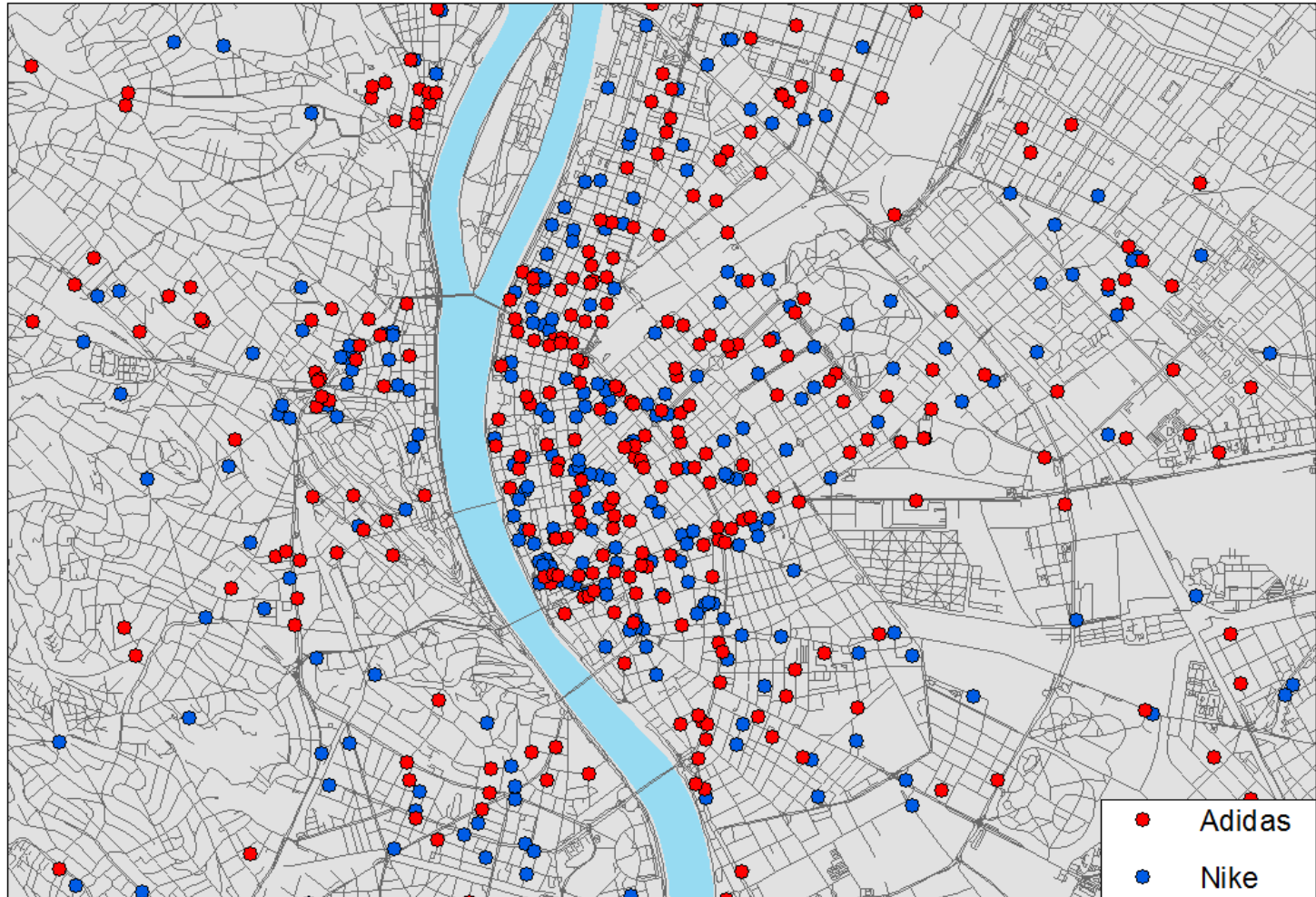


***Sűrűségterképek különböző kernel-beosztások mellett:
A „Dreher” kulcsszó-előfordulások különböző interpolált modelljei Budapesten***

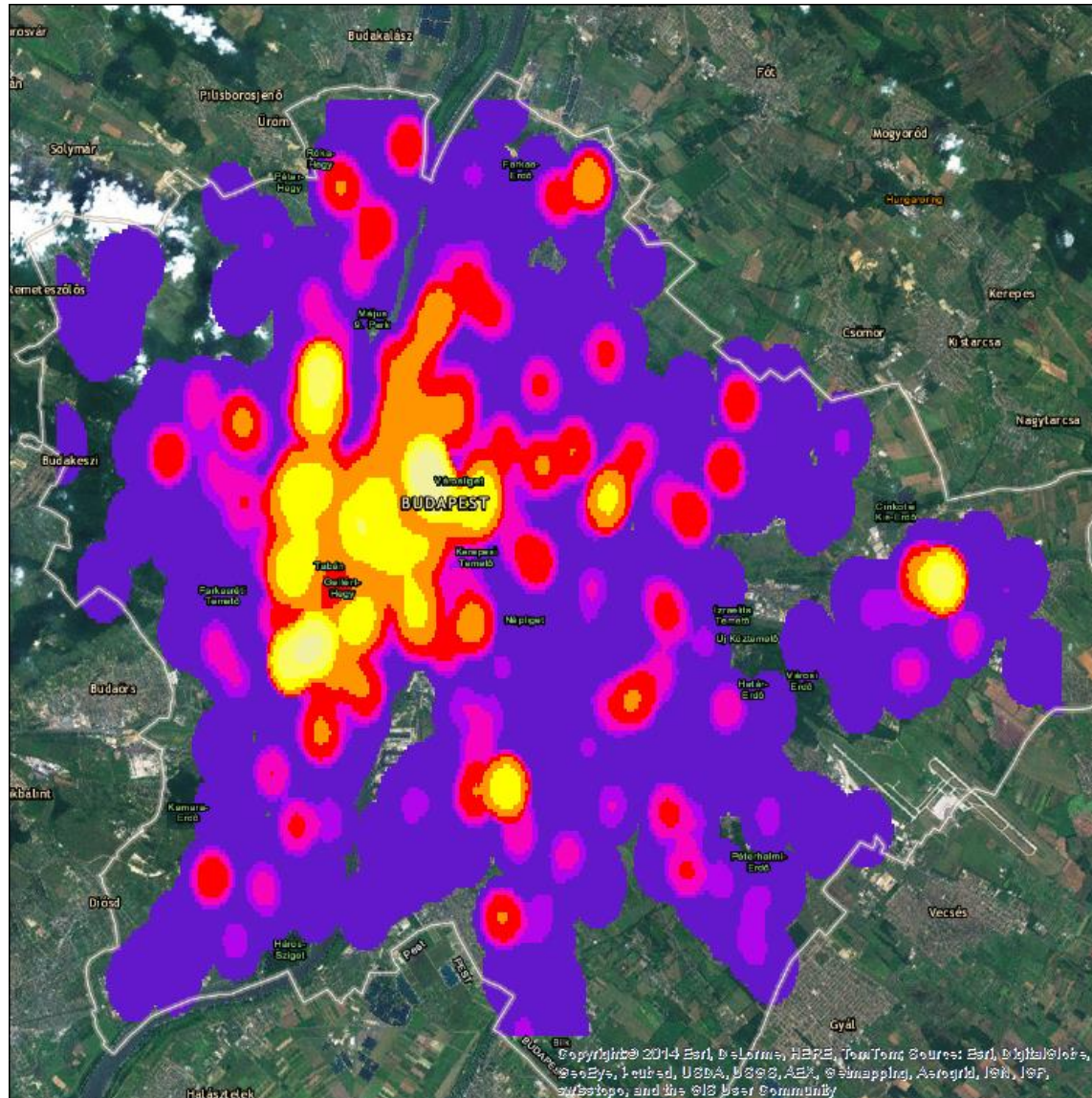


(kernel 0,01; 0,02; 0,03; 0,05 fok, weight population = találati score)

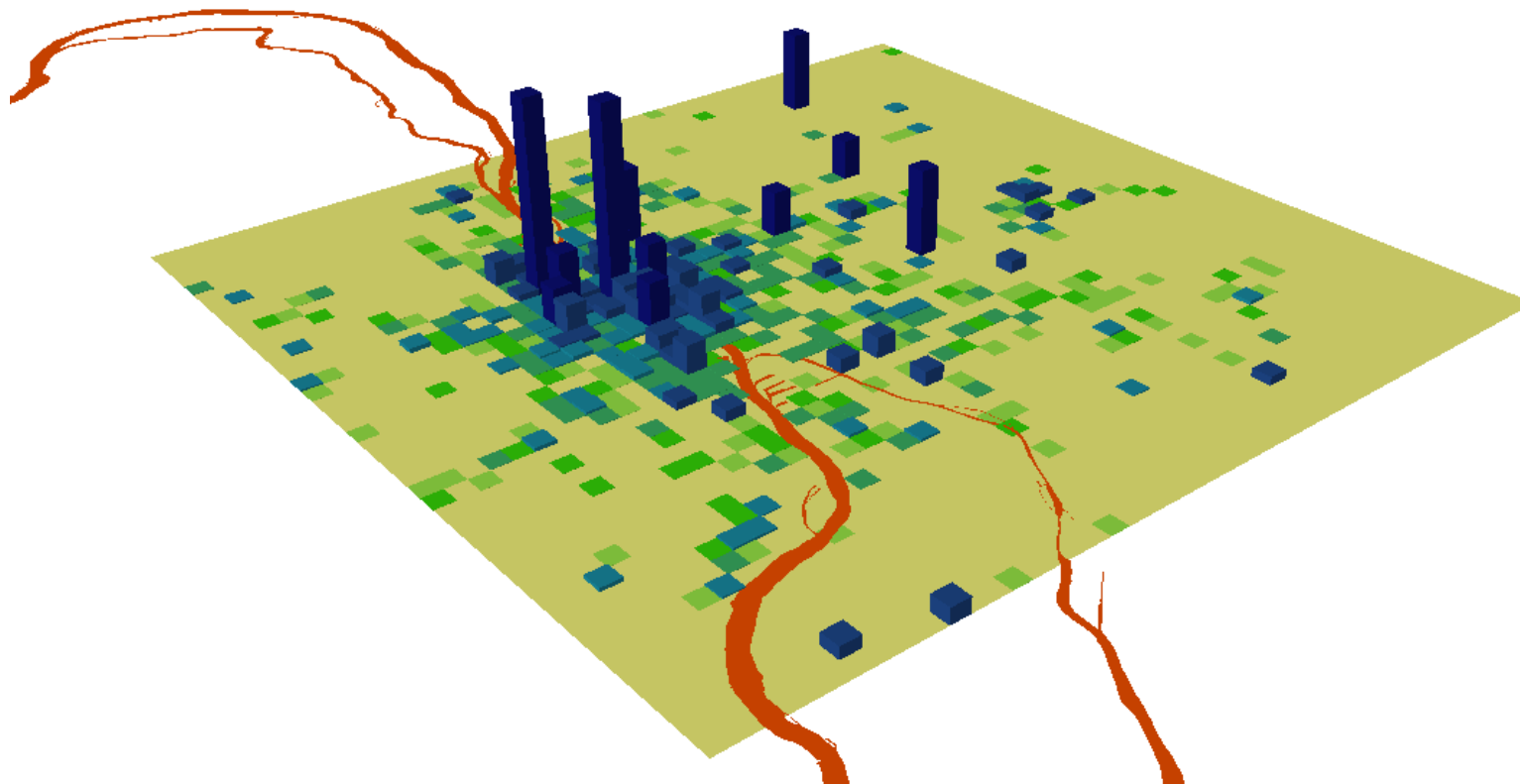
***Brand wars:
Adidas vs. Nike Budapest***



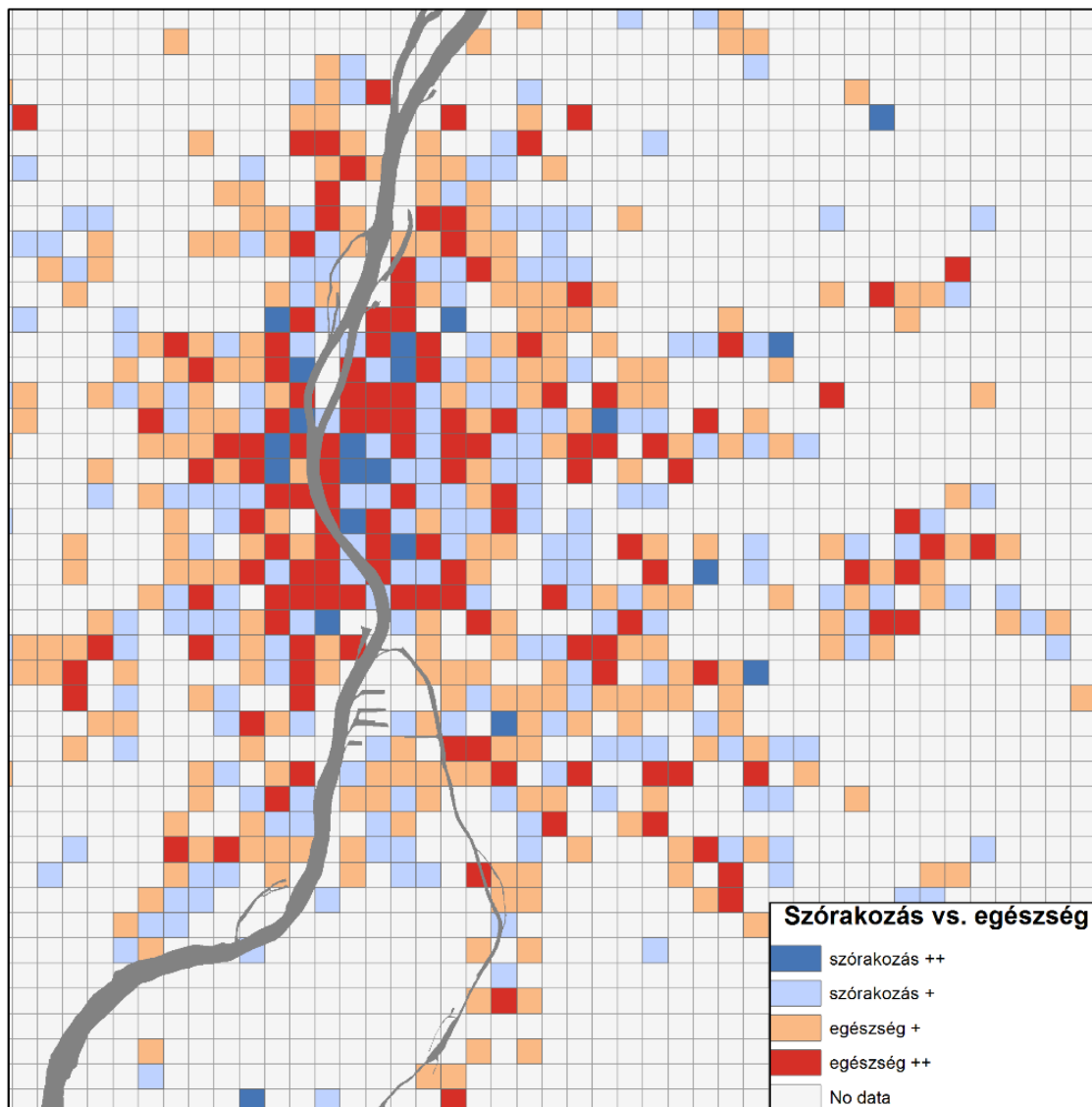
Általános keresőkifejezés: az „egészség”kulcsszó-előfordulások interpolált modellje Budapesten



***Az „egészség” kulcsszó-előfordulások száma Budapesten
(1x1 km-es térbeli rácsmodell)***



A „szórakozás” és az „egészség” kulcsszavak előfordulásainak dominancia-viszonyai a főváros környékén (1x1 km-es körzetekben)



Köszönöm a figyelmet!