

# PAIKKATIEDOT JA BIG DATA

MASSADATOILLA  
NÄKÖKULMIA  
IHMISTEN LIIKUMISEEN



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI

TUULI TOIVONEN /  
DIGITAL GEOGRAPHY LAB

# PAIKKATIEDOT JA BIG DATA

MASSADATOILLA  
NÄKÖKULMIA  
IHMISTEN LIIKUMISEEN



HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI



**Tuuli Toivonen**

@TuuliToivonen

Geographer/a.prof of geoinformatics  
@Digigeolab @HelsinkiUni. Novel  
(big/open) data sources + spatial analyses  
for urban, land use and conservation  
planning.

📍 Helsinki, Suomi

🔗 [tuhat.halvi.helsinki.fi/portal/fi/pers...](https://tuhat.halvi.helsinki.fi/portal/fi/pers...)

# YHÄ DYNAMISEMPI YHTEISKUNTA



# YHÄ DYNAAMISEMPI YHTEISKUNTA

Kaupunki

## Apulaispormestari Nasima Razmyar haluaa Helsinkiin ”yöpormestarin” – ”Täällä ajatellaan edelleen, että yöllä ihmiset nukkuvat”

Maailman suurissa kaupungeissa on herätty siihen, minkälainen kulttuurinen ja taloudellinen arvo yöllä kaupunki voi olla, jos siitä oltaisiin kiinnostuneita. Nyt yöpormestari löytyy jo New Yorkista ja Lontoostakin.



Helsingin kulttuurin ja vapaa-ajan apulaispormestari Nasima Razmyar haluaa, että Helsinki kiinnittäisi katseensa siihen, mitä yöllä on Helsingille annettavana. (KUVA: NIKLAS MELTIO)

Lari Malmberg HS

Julkaistu: 28.10.2017 2:00 , Päivitetty: 28.10.2017 10:04



### Luetuimmat

JUURI NYT PÄIVÄ VI

1. Ennennäkemätön Kimi Räikkönen myös kansikuvapoika ja maailmanlaajuisen lehden uuden numeron pääaihe
2. Räikkösen kunnioitusta huokuva palkintojenjaossa kerää ylistystä: todellista tyylitajua”
3. Nico Rosberg nosti itsensä suom. F1-voittajien listan ykköseksi – far tily vastaanotto
4. Säässä täyskäännös tällä viikolla: pyryttää jopa 15 senttiä Etelä-Suom. myöten – Meteorologi suosittelee vaihtamaan talvirenkaat
5. Kuinka nopeasti nuuskan käyttö r hammaslääkärillä? Voiko nuuska l reian poskeen? Asiantuntijat käy 15 netistä tuttua väitettä Tilaa
6. **Nyt.fi:** Kommentti: Putouksen sketsihahmokilpailun piti olla kail muuta kuin se nyt on – Yksi hahm tiivistää sen, mikä vanhassa suunnitelmassa meni mönkään

# LIKKUMISEN LISÄÄNTYMINEN



# KESTÄVYYDEN HAASTEET



# IHMISTEN TOIMINNAN JA DYNAMIIKAN YMMÄRTÄMINEN YHÄ TÄRKEÄMPÄÄ



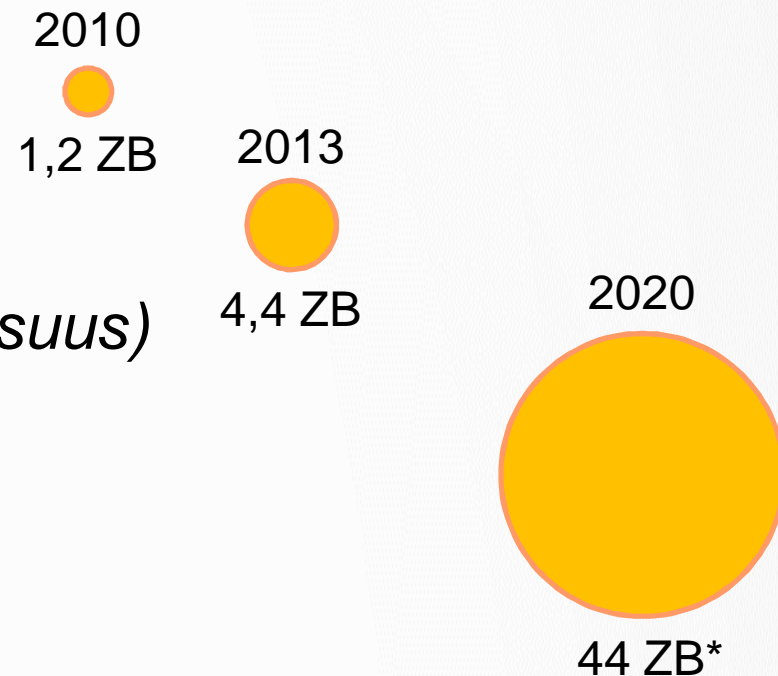


# BIG DATA – MASSADATAT?

## Perusominaisuudet:

- **Volume** (valtava määrä)
- **Variety** (vaihtoehtoja paljon)
- **Velocity** (tosiaikaisuus)
- **Veracity** (vinoumat)
- **Variability** (epäjohdonmukaisuus)

**Datamäärä  
maailmassa kasvaa 40%  
vuodessa**





### 40 ZETTABYTES

[ 43 TRILLION GIGABYTES ]  
of data will be created by 2020, an increase of 300 times from 2005



6 BILLION PEOPLE have cell phones



WORLD POPULATION: 7 BILLION

## Volume SCALE OF DATA

It's estimated that **2.5 QUINTILLION BYTES** [ 2.3 TRILLION GIGABYTES ] of data are created each day



Most companies in the U.S. have at least **100 TERABYTES** [ 100,000 GIGABYTES ] of data stored



# The FOUR V's of Big Data

From traffic patterns and music downloads to web history and medical records, data is recorded, stored, and analyzed to enable the technology and services that the world relies on every day. But what exactly is big data, and how can these massive amounts of data be used?

As a leader in the sector, IBM data scientists break big data into four dimensions: **Volume, Velocity, Variety and Veracity**

Depending on the industry and organization, big data encompasses information from multiple internal and external sources such as transactions, social media, enterprise content, sensors and mobile devices. Companies can leverage data to adapt their products and services to better meet customer needs, optimize operations and infrastructure, and find new sources of revenue.

By 2015 **4.4 MILLION IT JOBS** will be created globally to support big data, with 1.9 million in the United States



As of 2011, the global size of data in healthcare was estimated to be

**150 EXABYTES** [ 161 BILLION GIGABYTES ]



**30 BILLION PIECES OF CONTENT** are shared on Facebook every month



## Variety DIFFERENT FORMS OF DATA

By 2014, it's anticipated there will be **420 MILLION WEARABLE, WIRELESS HEALTH MONITORS**

**4 BILLION+ HOURS OF VIDEO** are watched on YouTube each month



**400 MILLION TWEETS** are sent per day by about 200 million monthly active users



The New York Stock Exchange captures

**1 TB OF TRADE INFORMATION** during each trading session



Modern cars have close to **100 SENSORS** that monitor items such as fuel level and tire pressure



## Velocity ANALYSIS OF STREAMING DATA

By 2016, it is projected there will be

**18.9 BILLION NETWORK CONNECTIONS**

— almost 2.5 connections per person on earth



**1 IN 3 BUSINESS LEADERS** don't trust the information they use to make decisions



Poor data quality costs the US economy around

**\$3.1 TRILLION A YEAR**



**27% OF RESPONDENTS**

in one survey were unsure of how much of their data was inaccurate

## Veracity UNCERTAINTY OF DATA

Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTec, QAS





# MASSADATAN LÄHTEET

## Suunnitelmallisesti kerätyt

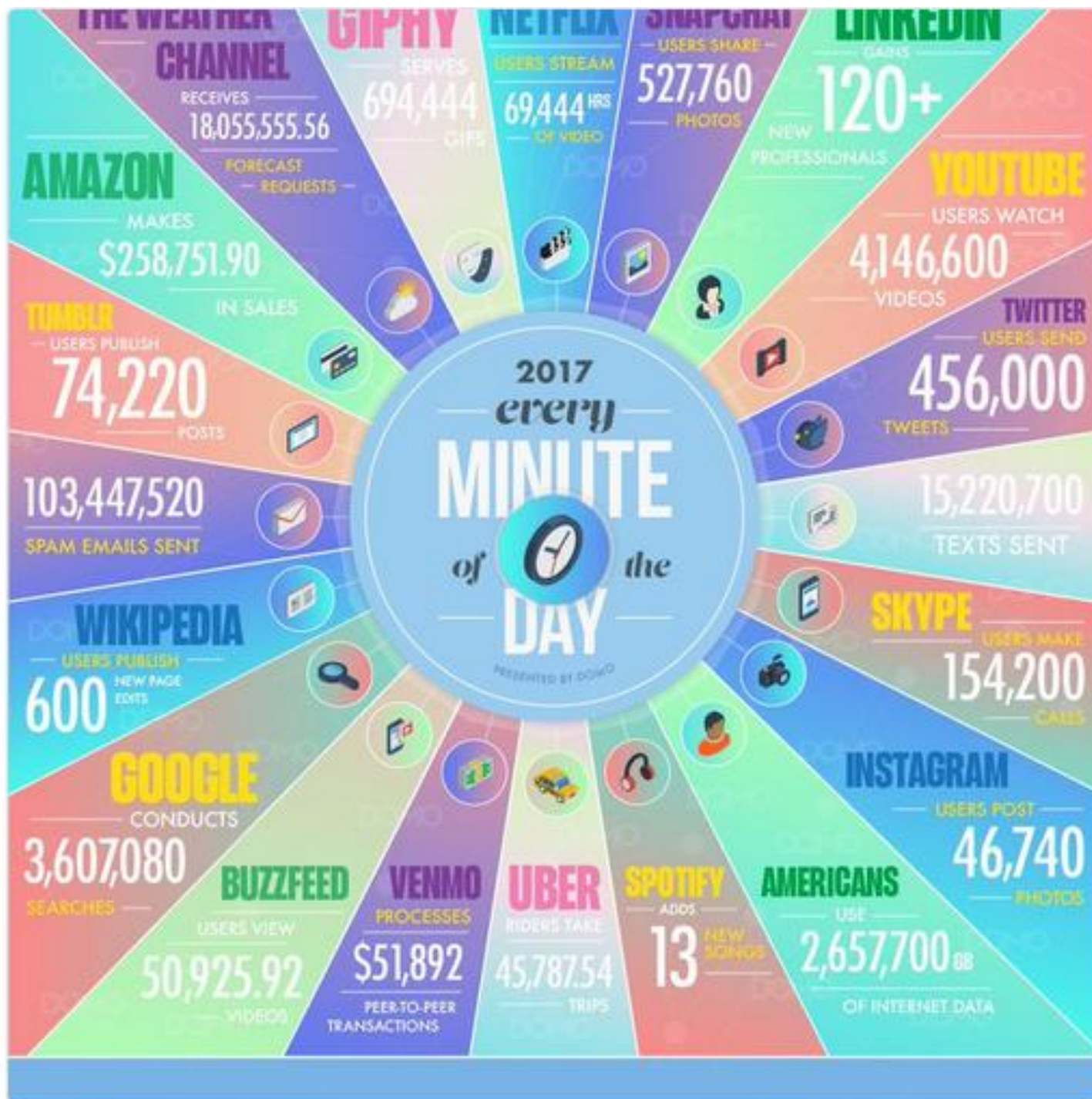
- Kaukokartoitus, laserkeilaus
- Valvontakamerat (CCTV; ANPR)
- Kävijälaskurit
- Tunnistautuminen (rajatarkastukset, rekisteröitymiset)

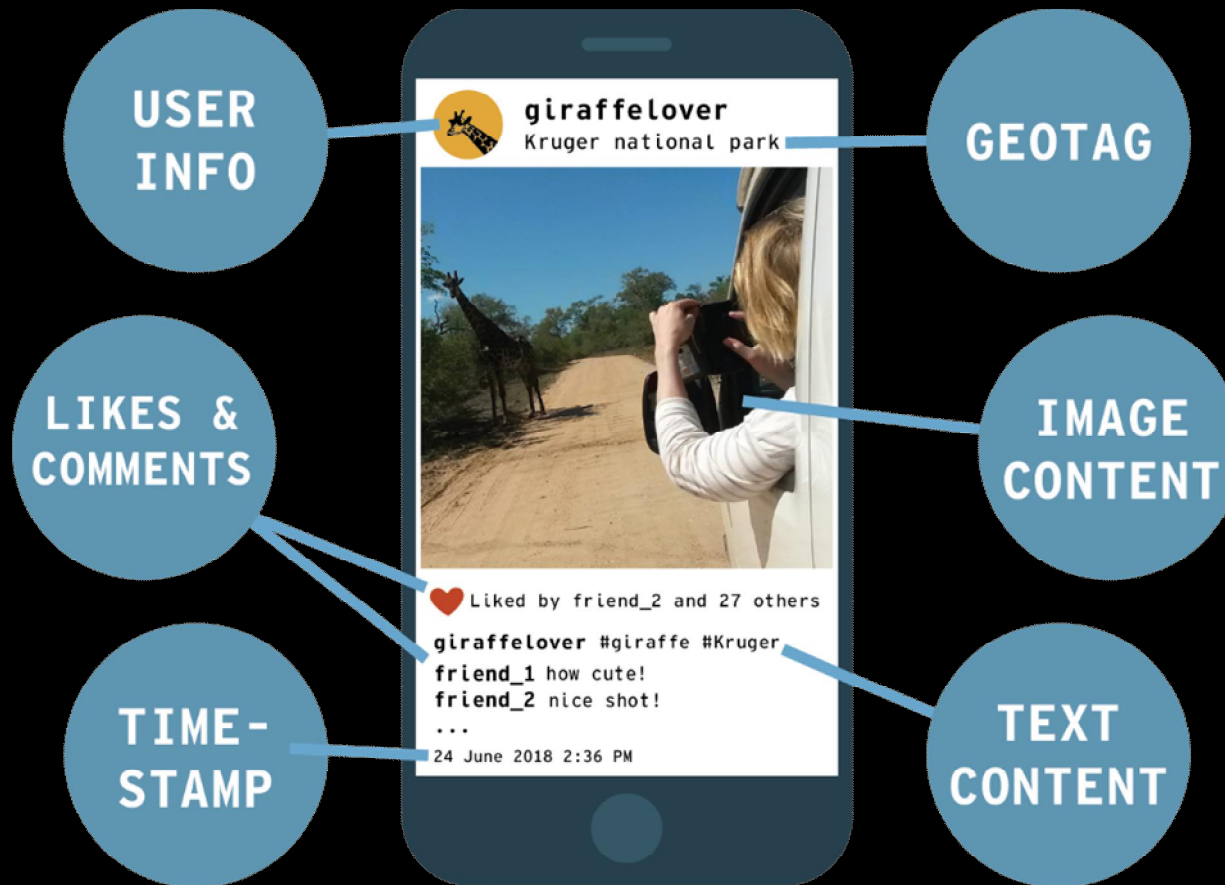
## Sivutuotteena syntyvät

- Matkapuhelinten käyttö
- Asiakaskorttien käyttö
- Verkon käyttö (hakukoneet, klikkaukset)
- Mobiiliappien käyttö

## Käyttäjän aktiivisesti tuottamat

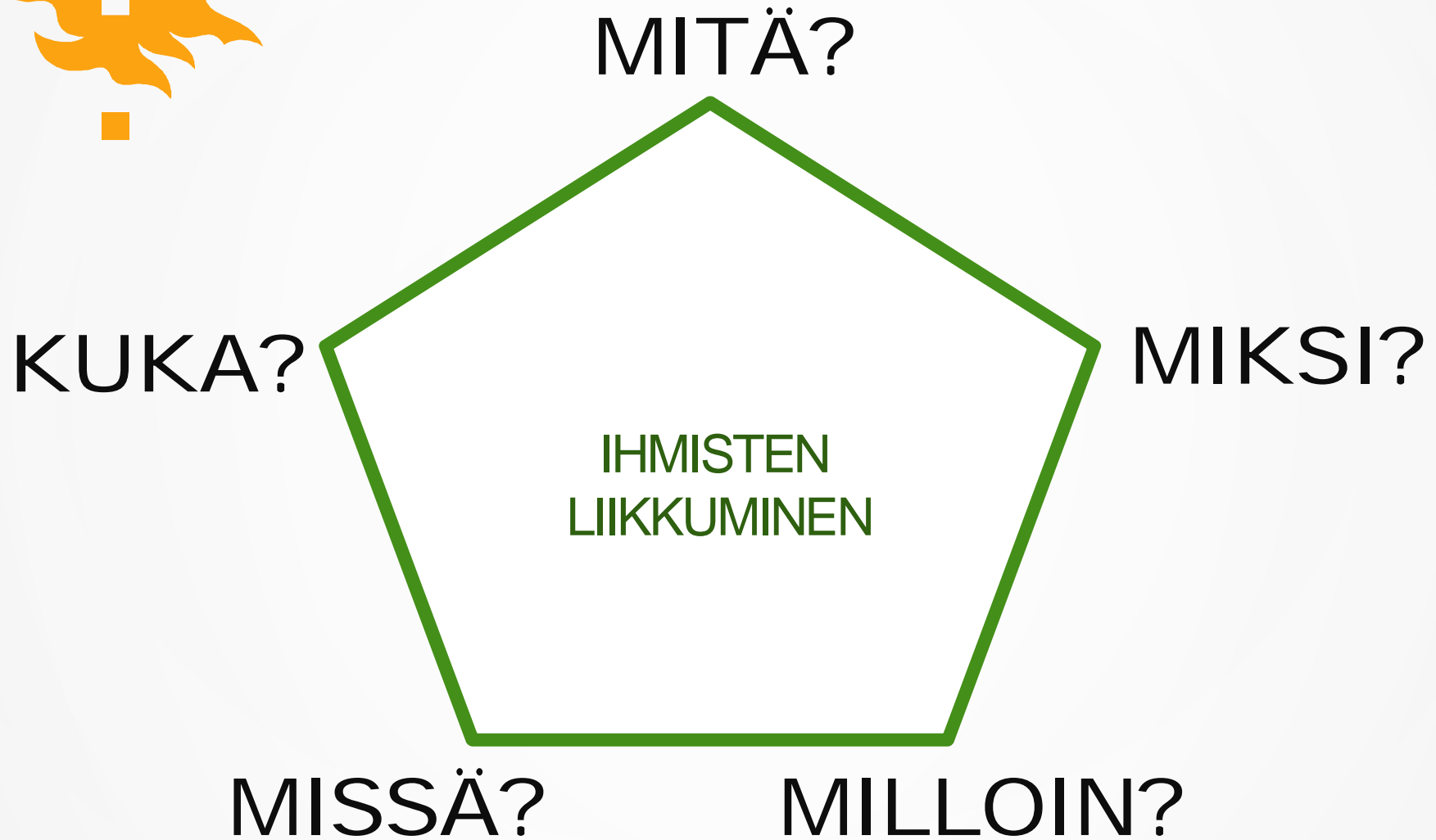
- Sosiaalinen media
- Urheilusovellukset

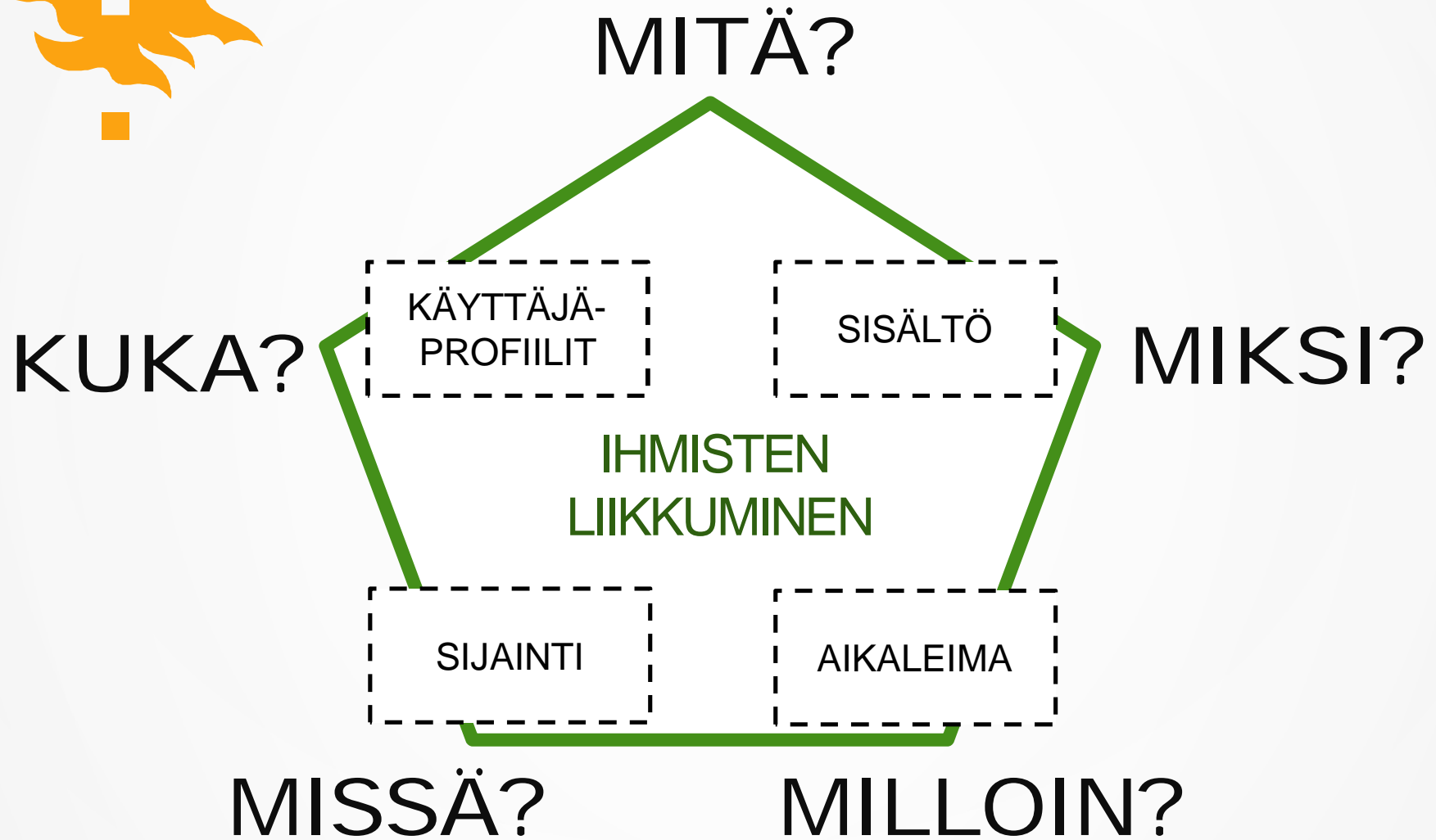


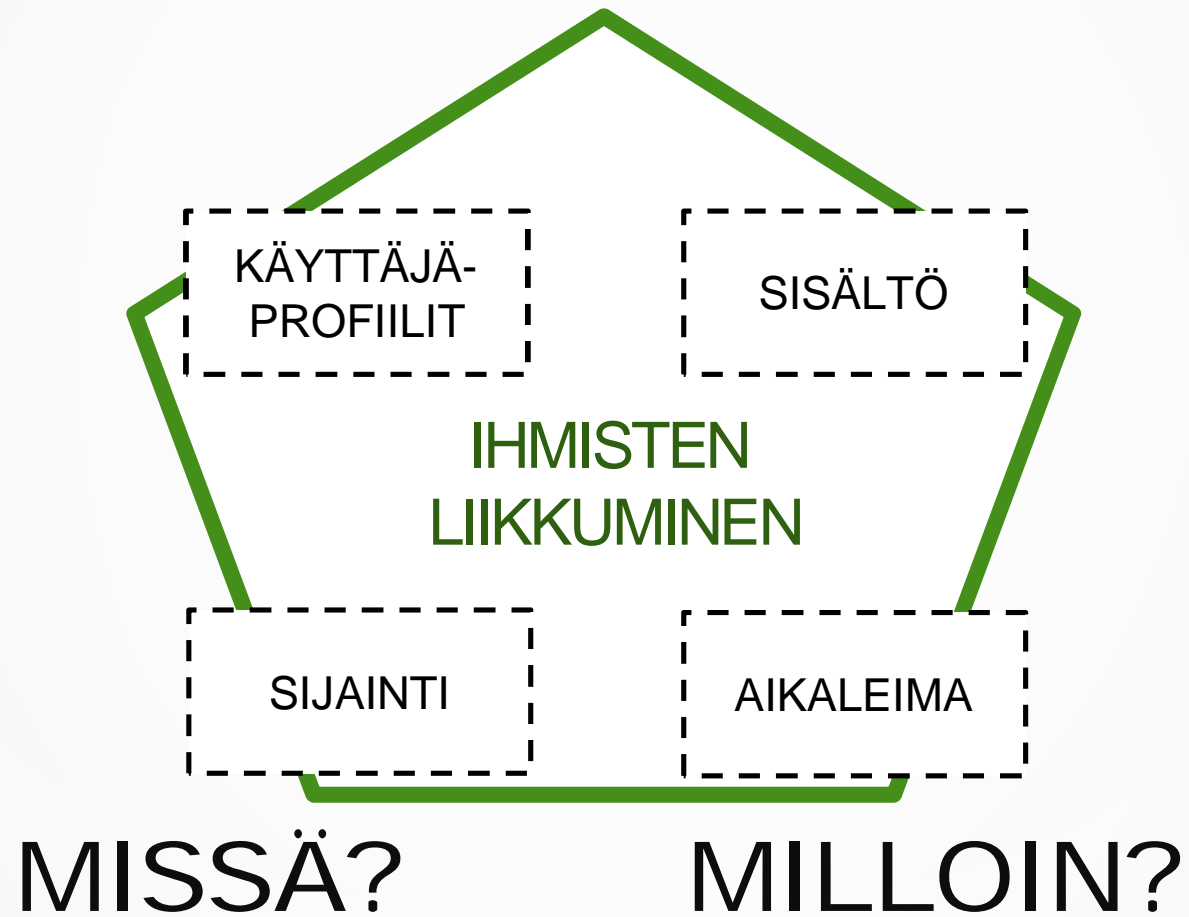


Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järvi, O., Tenkanen, H., Di Minin, E.  
*Social media data for conservation science: a methodological overview.* Biological Conservation (in press).







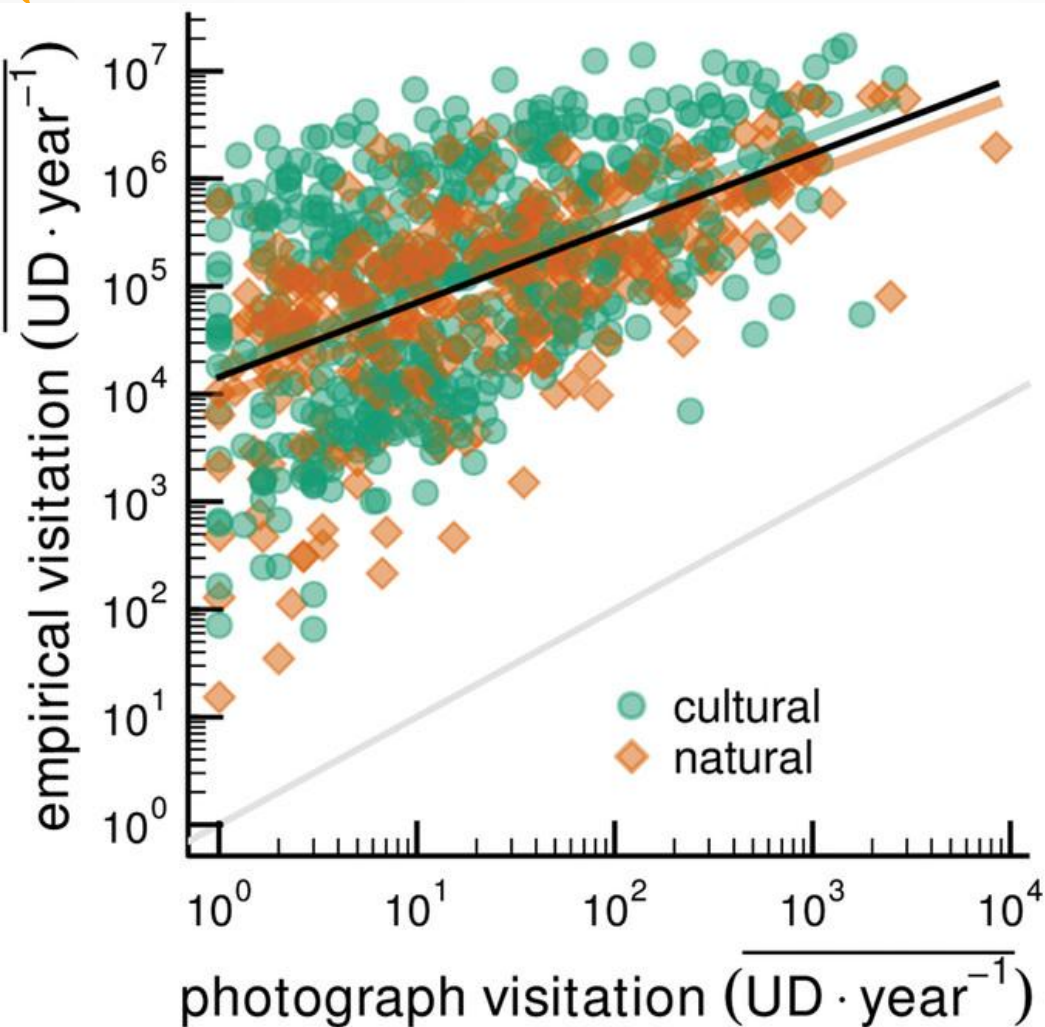






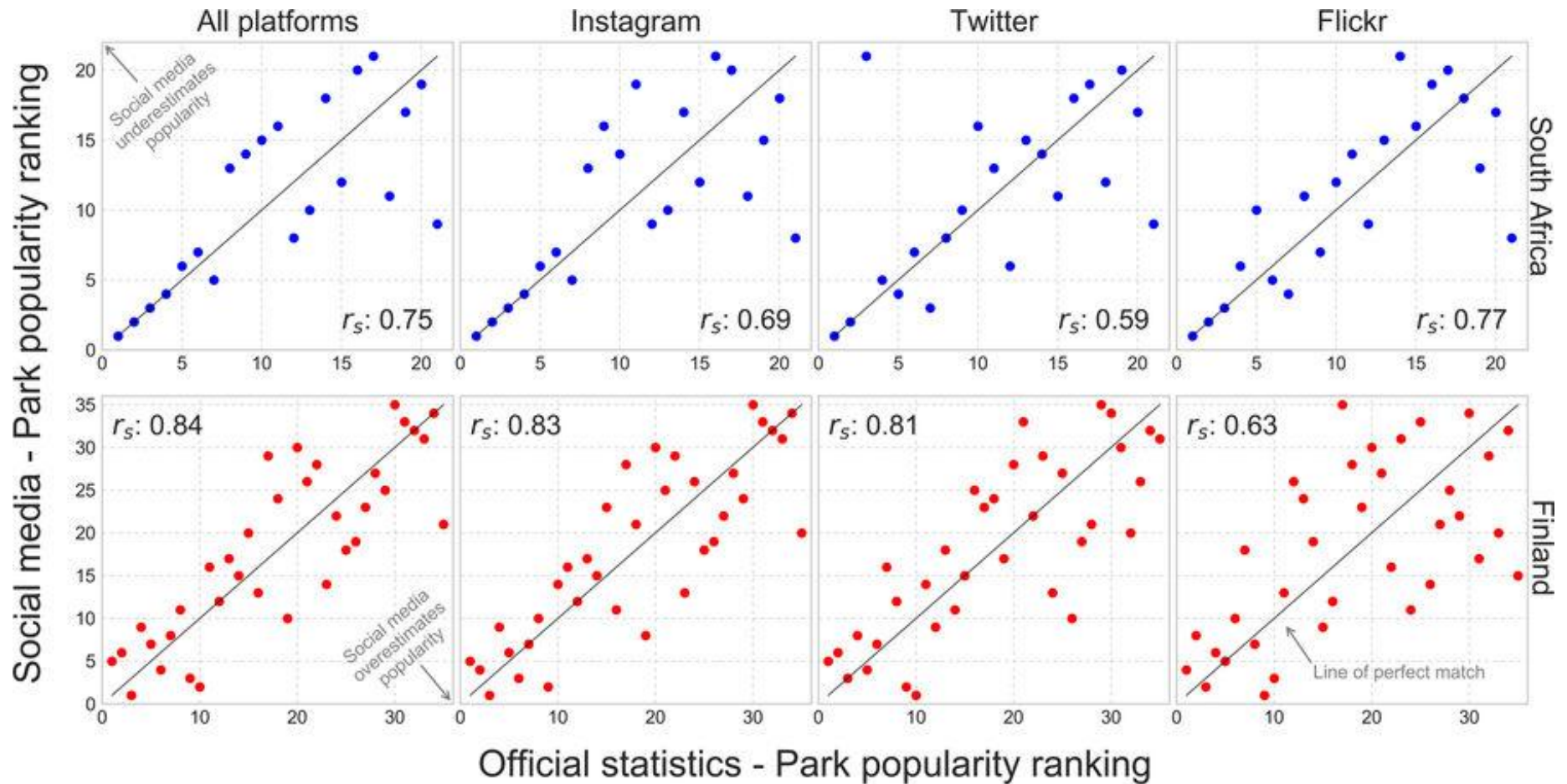


# FLICKR KUVAT TURISKOIHOIEN SUOSITUMMUUDEN MITTARINA



Wood et al. (2013)  
**Using social media  
to quantify  
nature-based  
tourism and  
recreation**

# SOSIAALINEN MEDIA VS. KÄVIJÄTILASTOT



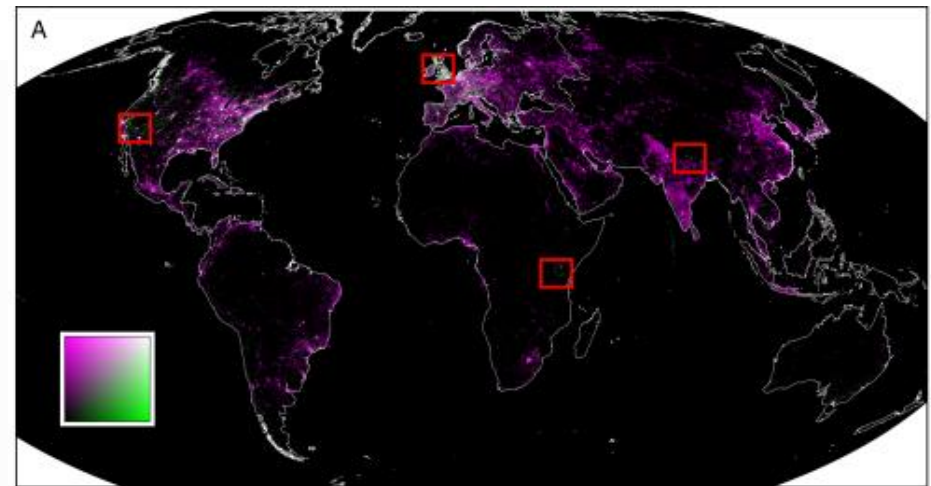
Tenkanen et al. (2017) Scientific Reports 7, Article number: 17615



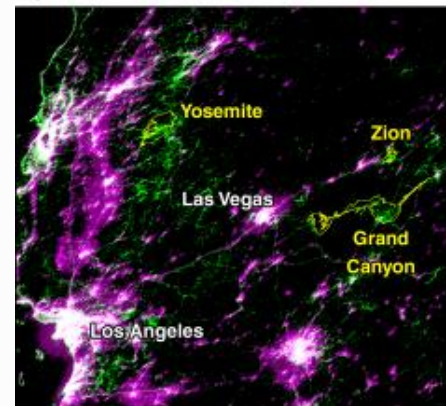
Levin et al. (2015) Where have all the people gone? Enhancing global conservation using night lights and social media.

Ecological Applications  
Volume 25, Issue 8, pages 2153-2167, 1 DEC  
2015 DOI: 10.1890/15-0113.1  
<http://onlinelibrary.wiley.com/doi/10.1890/15-0113.1/full#i1051-0761-25-8-2153-f04>

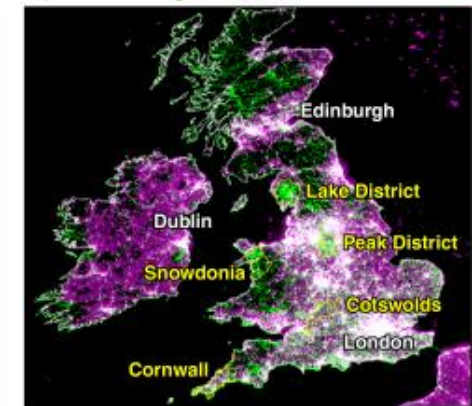
**HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI**



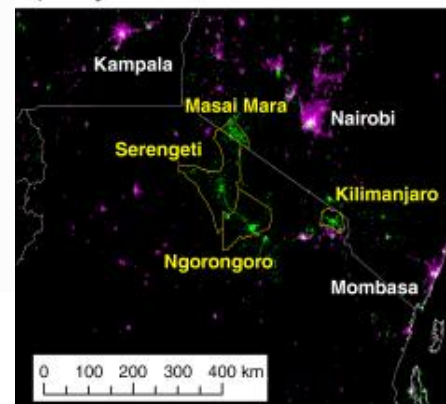
B) Southwestern United States



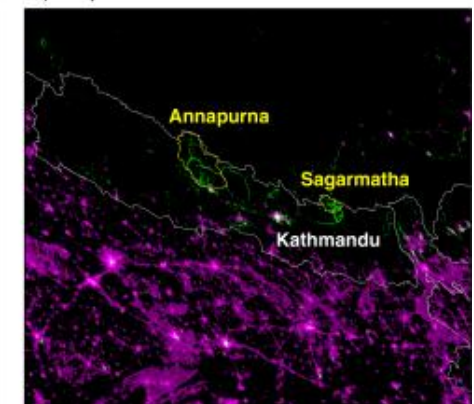
C) United Kingdom



D) Kenya and Tanzania

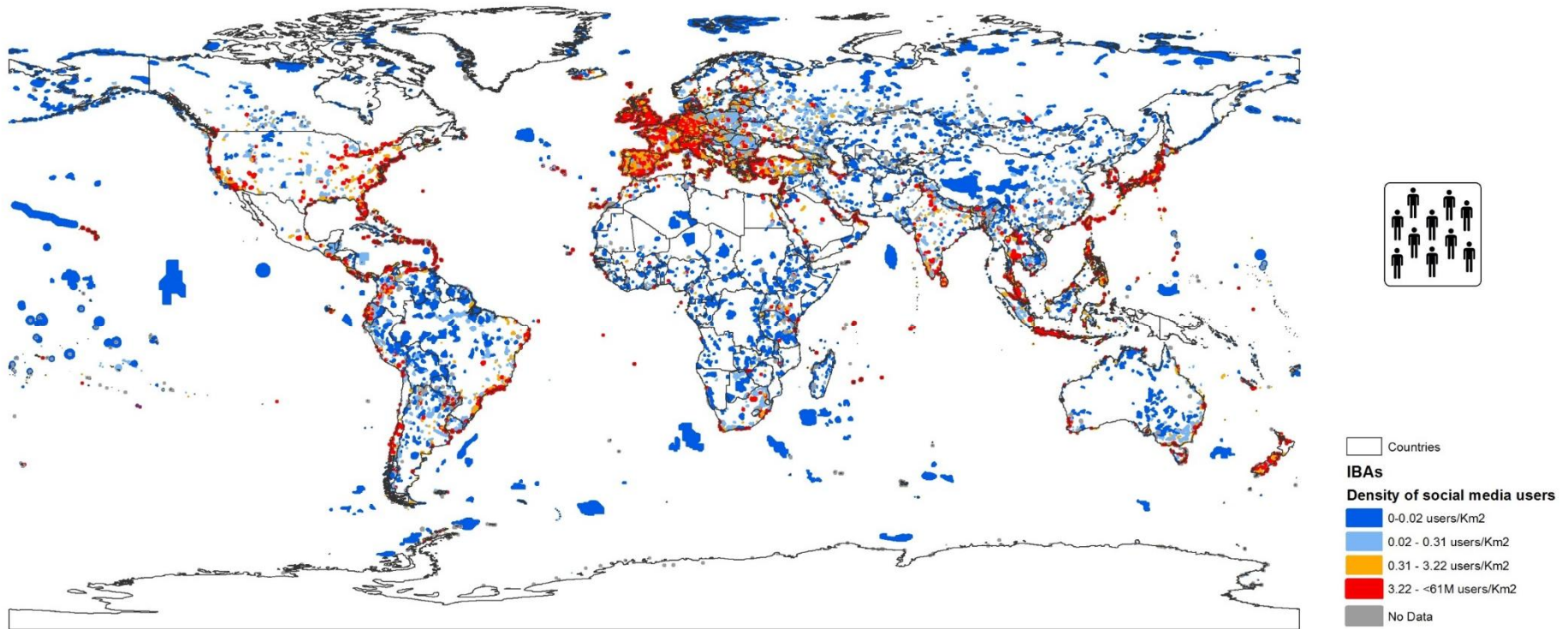


E) Nepal

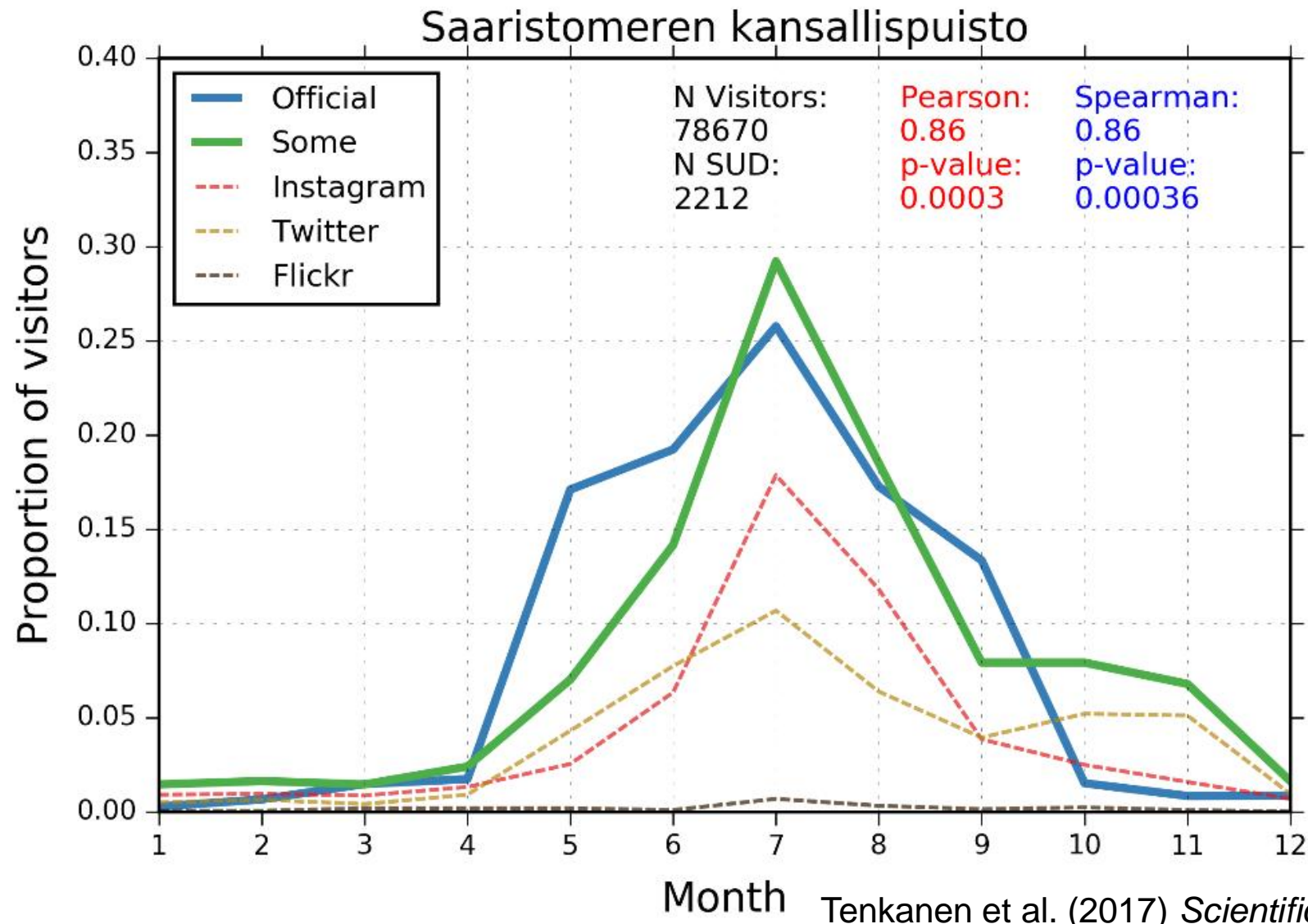


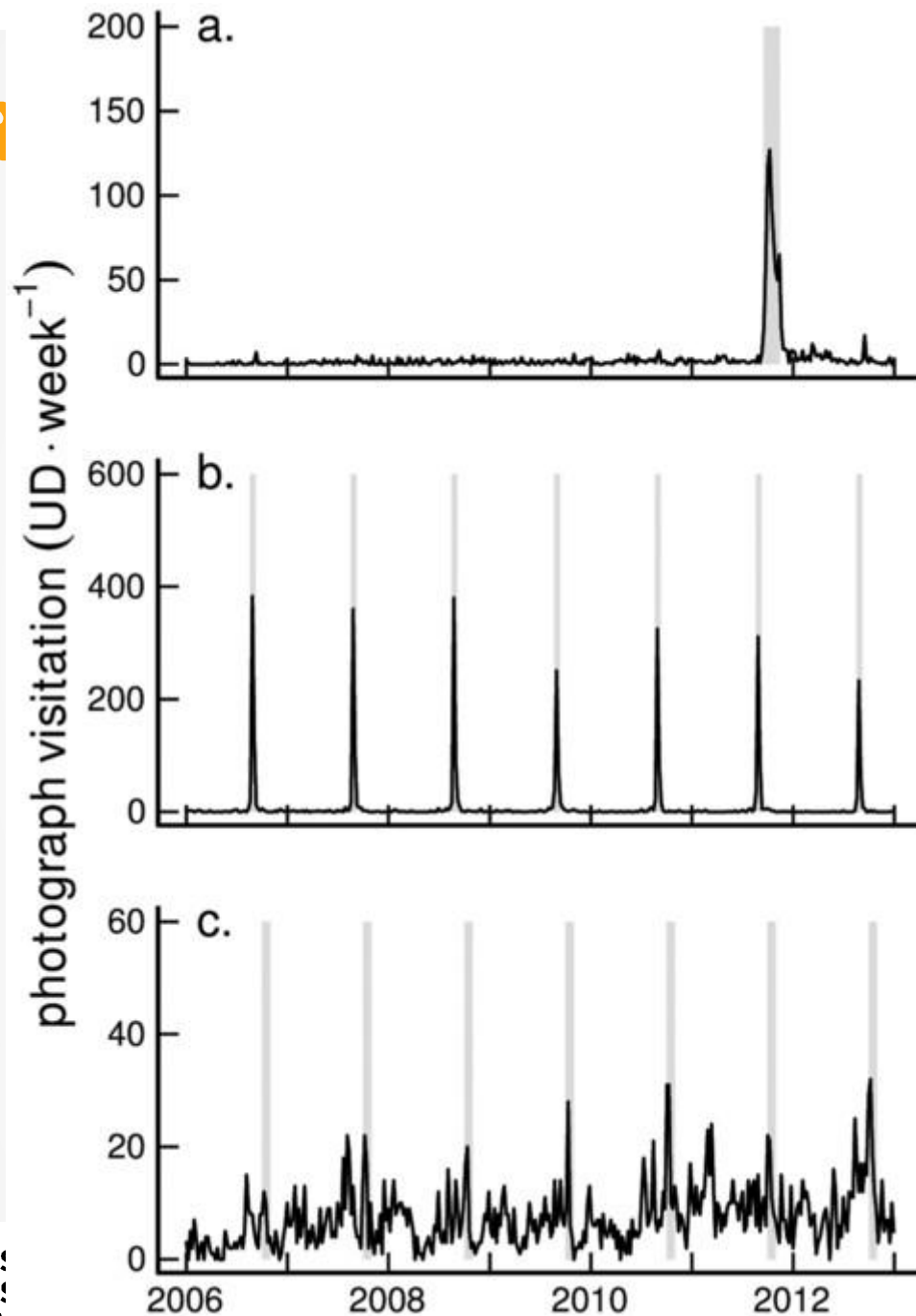


# DENSITY OF SOCIAL MEDIA USERS IN IBA'S



# PUISTOJEN KÄYTÖN TEMPORAALISUUS - VERTAILU





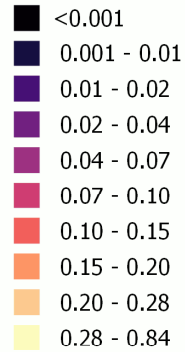
**TOTAL PHOTO-USER-DAYS EACH WEEK FROM 2006–2012 IN (A) ZUCCOTTI PARK, THE SITE OF THE OCCUPY PROTEST, (B) BLACK ROCK DESERT, THE SITE OF THE BURNING MAN FESTIVAL, AND (C) SOUTHERN VERMONT, AN AREA POPULAR FOR VIEWING AUTUMN FOLIAGE. GREY SHADING INDICATES TIME PERIODS FROM THE START OF THE PROTEST UNTIL PEOPLE WERE BARRED FROM THE PARK (A), THREE WEEKS SPANNING THE ANNUAL WEEK-LONG FESTIVAL (B), OR THE MONTH OF OCTOBER EACH YEAR (C).**

[Wood et al. \(2013\)](#)

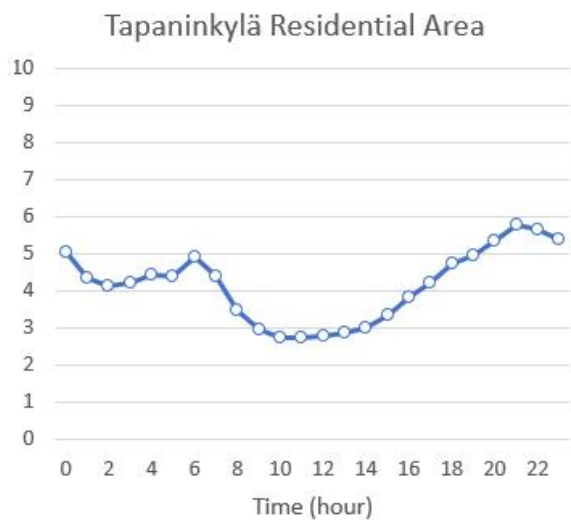
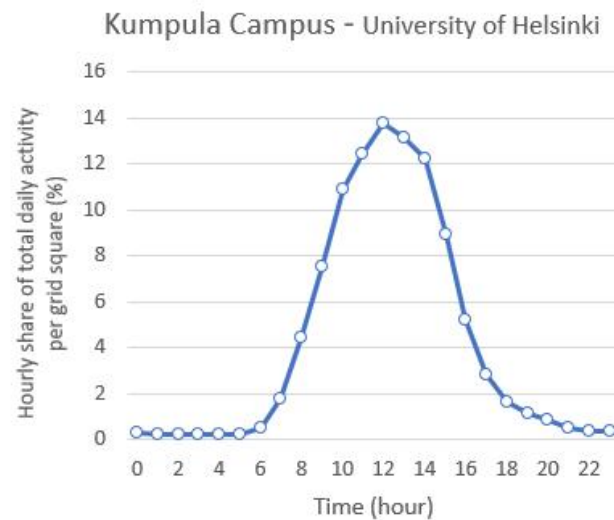
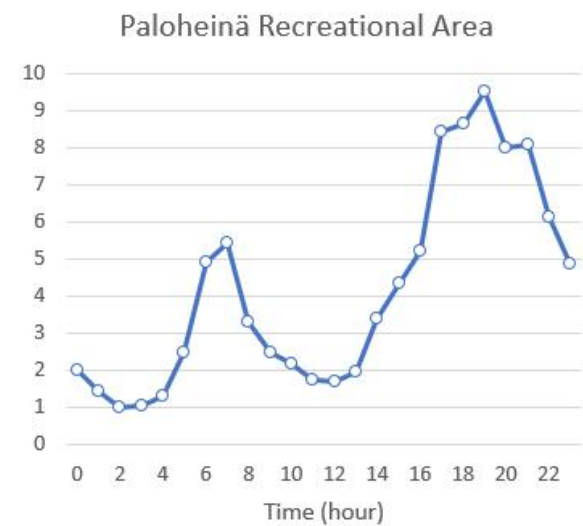
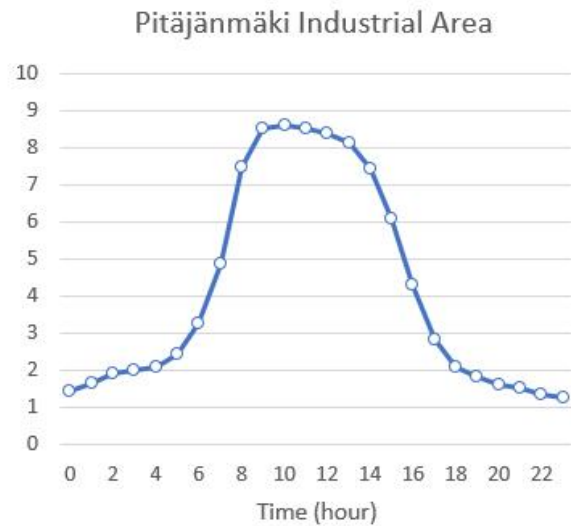
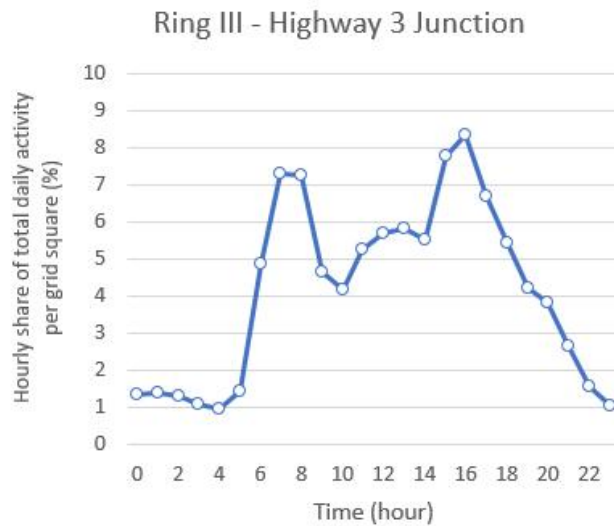
<http://www.nature.com/articles/srep02976>

*Ongoing work by Bergroth et al. 2018*

Relative share of population  
in a 250 m grid square [%]



**Pääkaupunkiseudun väestöpulssi –  
matkapuhelinaineiston perusteella**



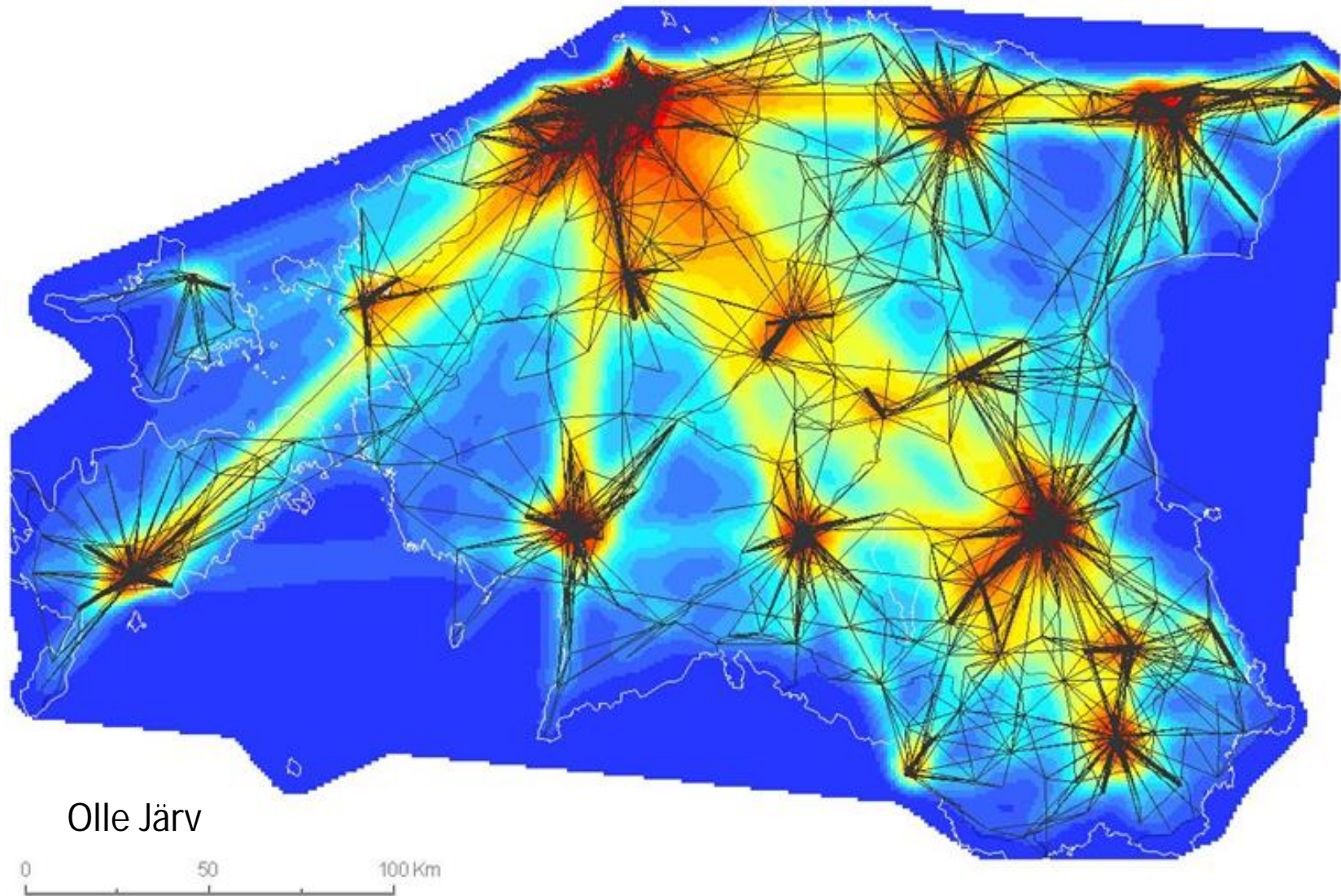
©Claudia Bergroth, Olle Järv, Henrikki Tenkanen, Tuuli Toivonen - Digital Geography Lab, University of Helsinki, 2018. CC-BY-4.0.  
[blogs.helsinki.fi/accessibility/2018/10/09/the-24-h-population-dynamics-of-the-finnish-capital-region-uncovered](https://blogs.helsinki.fi/accessibility/2018/10/09/the-24-h-population-dynamics-of-the-finnish-capital-region-uncovered)

*Ongoing work by Bergroth et al. 2018*

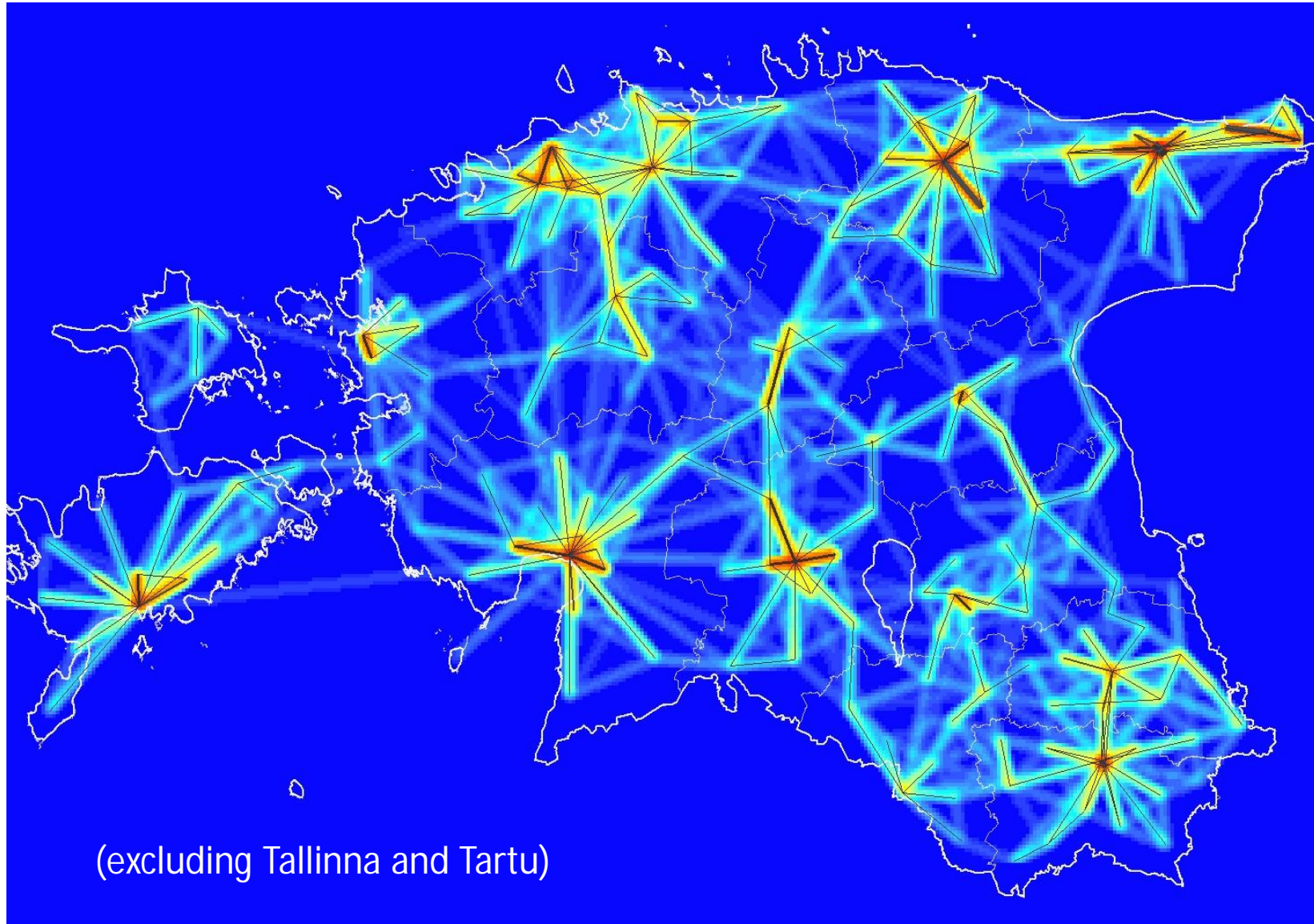
# Väestön vuorokauden aikainen dynamiikka eri paikoissa



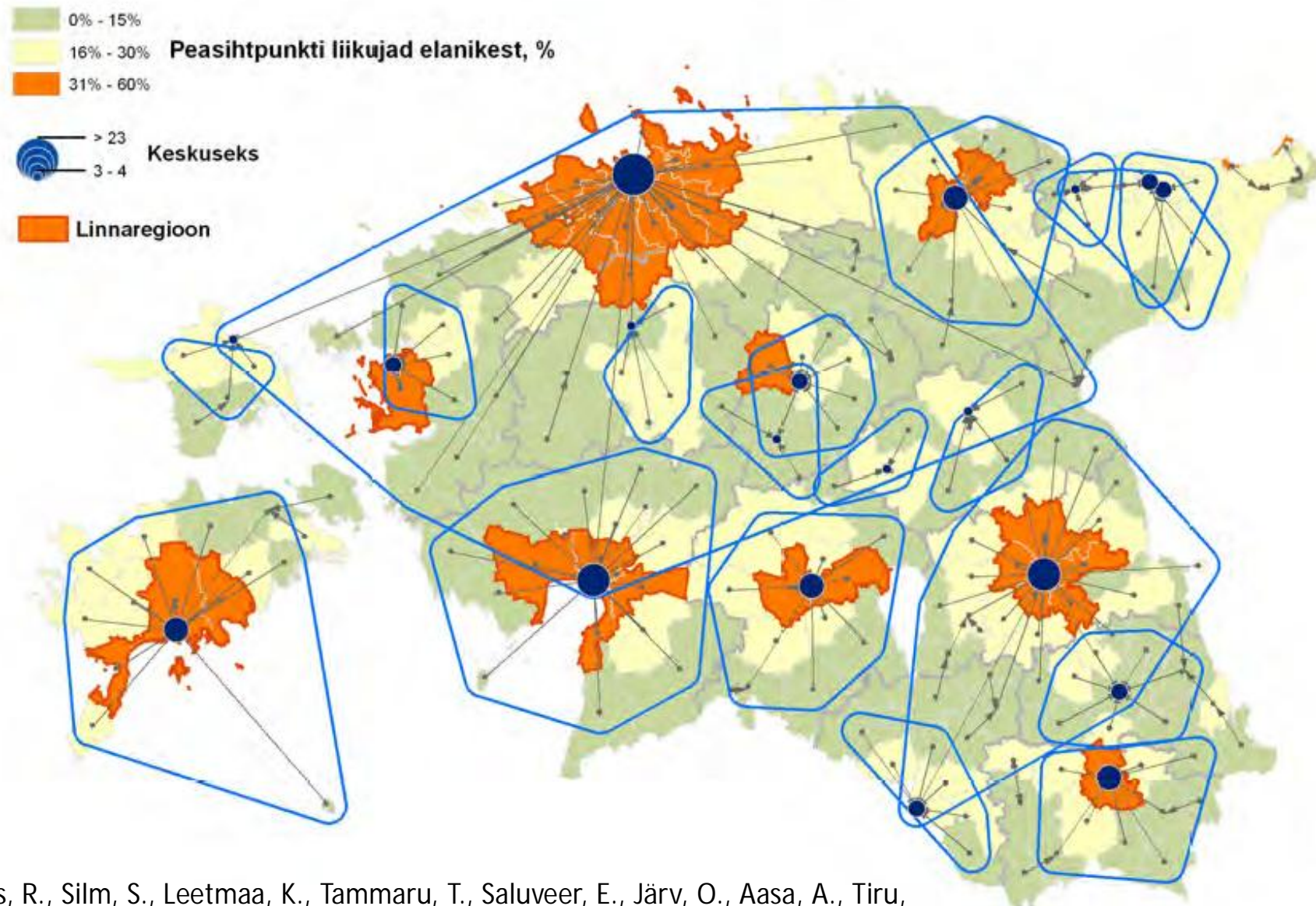
# Liikkumisen rakenteet Virossa



# Päivittäisen pendelöinnin rakenteet Virossa



# Toiminnalliset alueet liikkumisen pohjalta



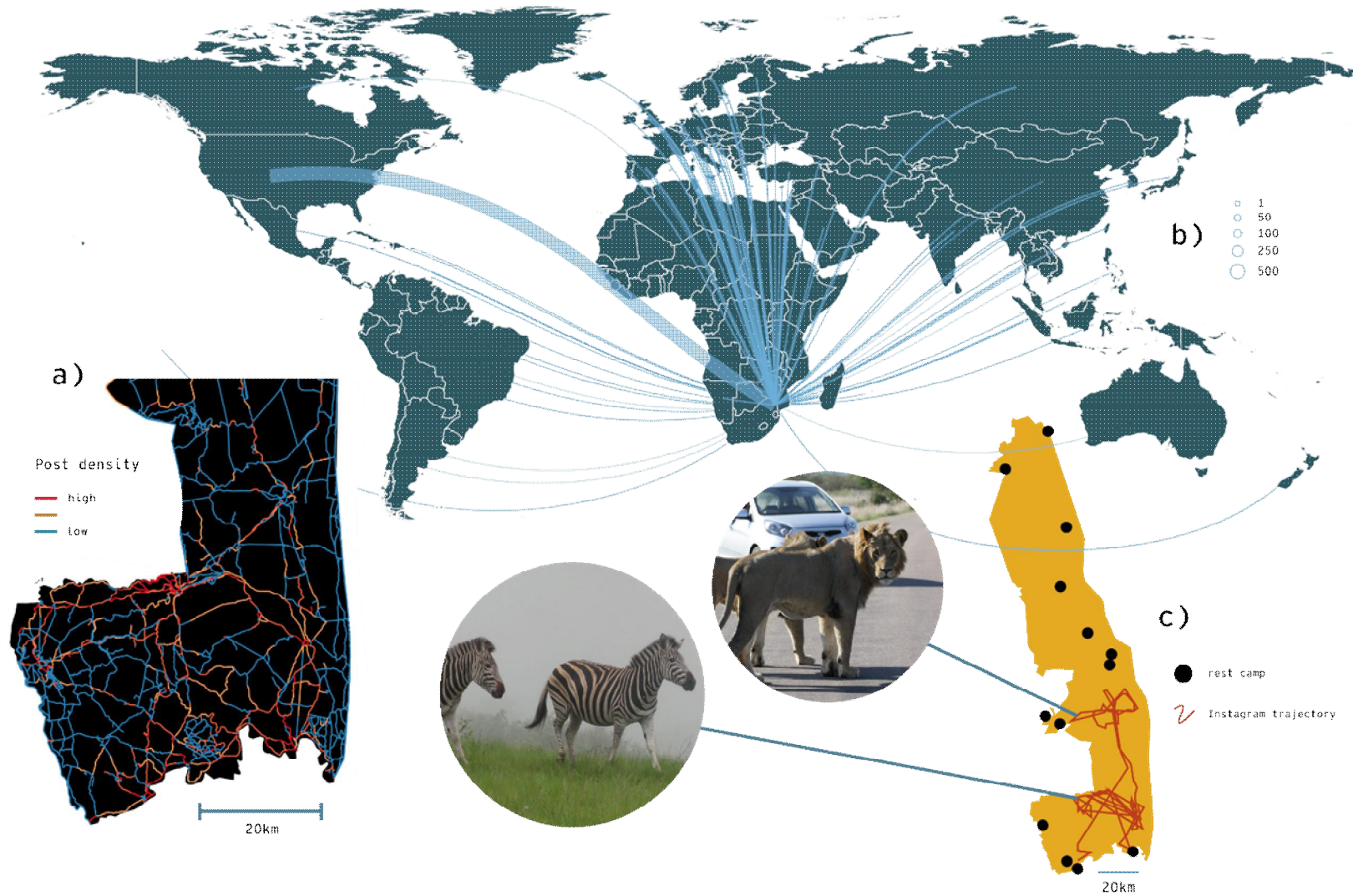
Ahas, R., Silm, S., Leetmaa, K., Tammaru, T., Saluveer, E., Järv, O., Aasa, A., Tiru, M. (2010). *The Regional Commuting Study*. Estonian Ministry of the Interior.

# KANSALLISESTA KANSAINVÄLISEEN



Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järvi, O., Tenkanen, H., Di Minin, E.  
*Social media data for conservation science: a methodological overview*. *Biological Conservation* (in press).

# PAIKALLISESTA KANSAINVÄLISEEN



Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järvi, O., Tenkanen, H., Di Minin, E.  
*Social media data for conservation science: a methodological overview*. *Biological Conservation* (in press).

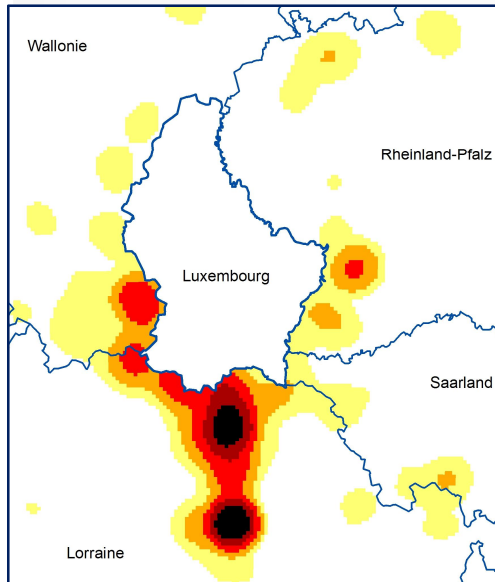


# YLIKANSALLINEN LIIKKUMINEN SOSIAALISEN MEDIAN POHJALTA

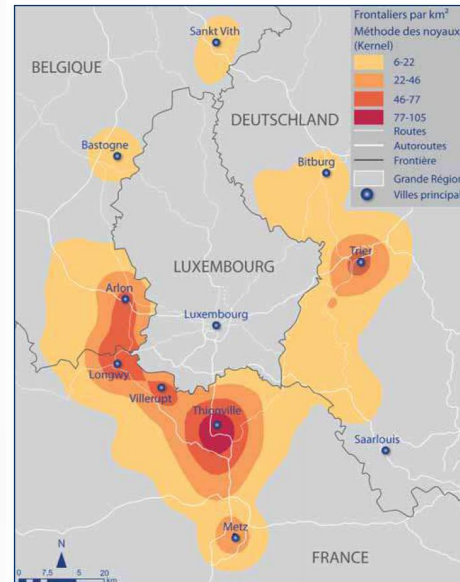
Homes locations of cross-border commuters of Luxembourg

Cross-border mobility  
Based on Twitter data

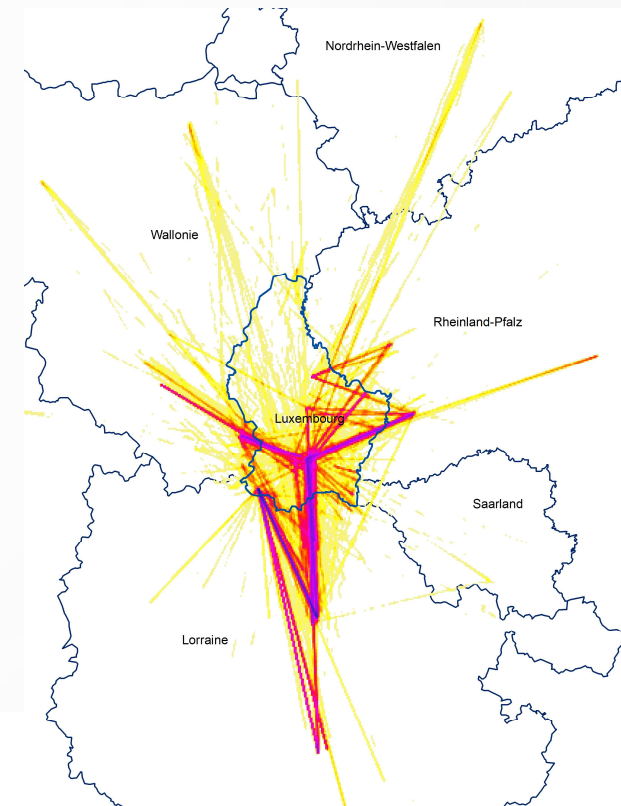
Twitter data 2013-2018 vs Register data 2010



Authors: O. Järvi, S. Massinen. DGL  
2019

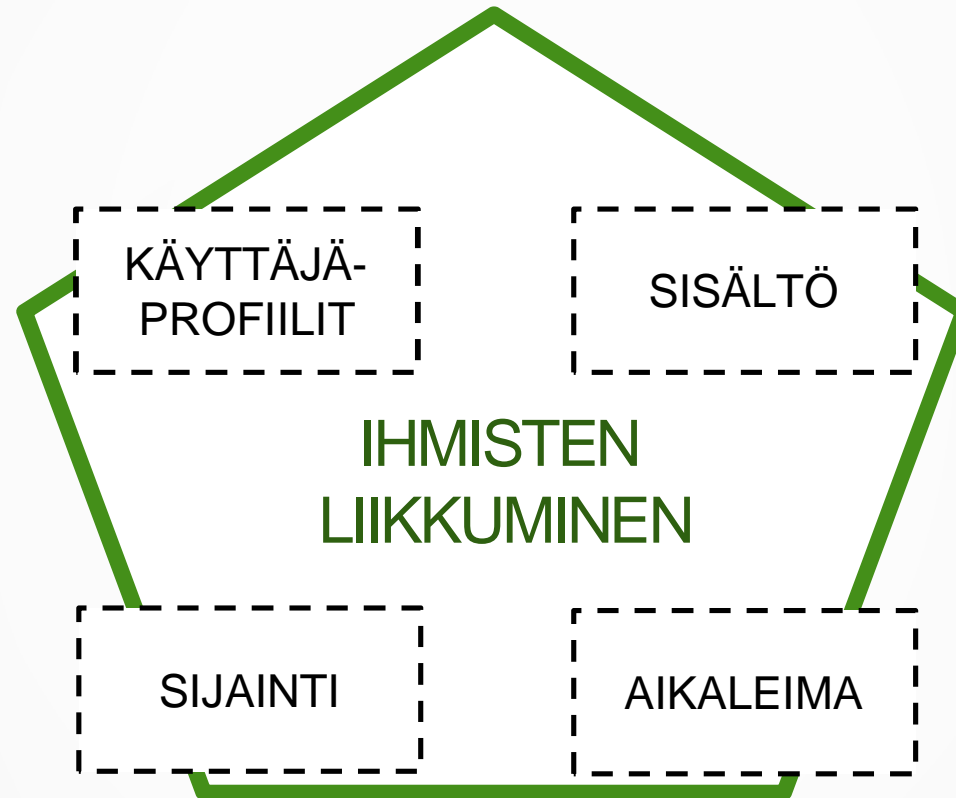


Cartographie : G. Drevon, P. Gerber, M. Schneider, CEPS/INSTEAD, 2012  
Données statistiques : IGS55, 1994, 1999, 2010  
Fond de carte : © EuroGeographics 2002 pour les limites administratives



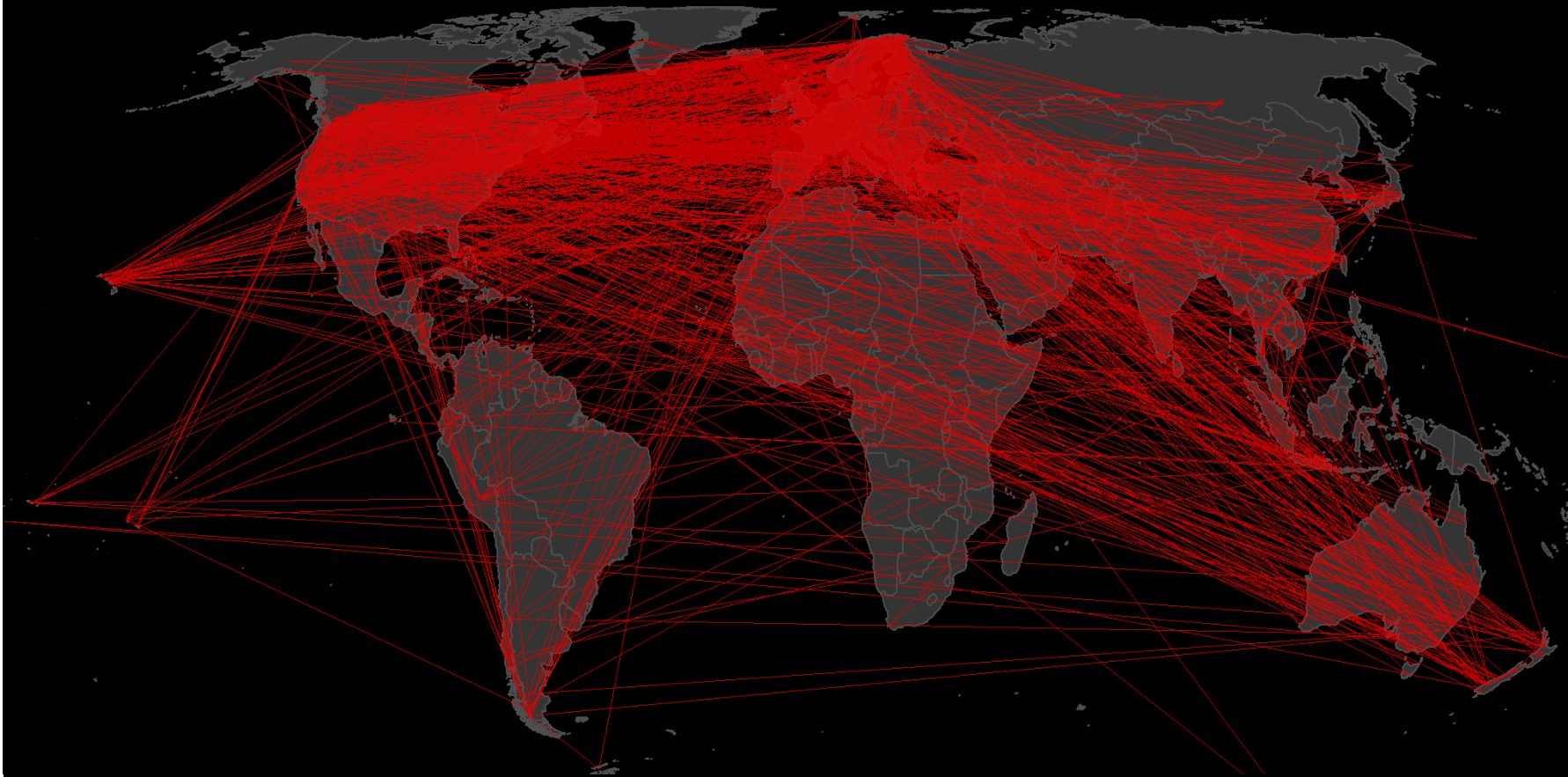


# MITÄ?



# MIKSI?

# PALLAS-YLLÄKSEN KANSALLISPUISTO



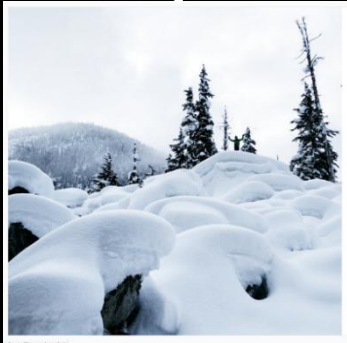
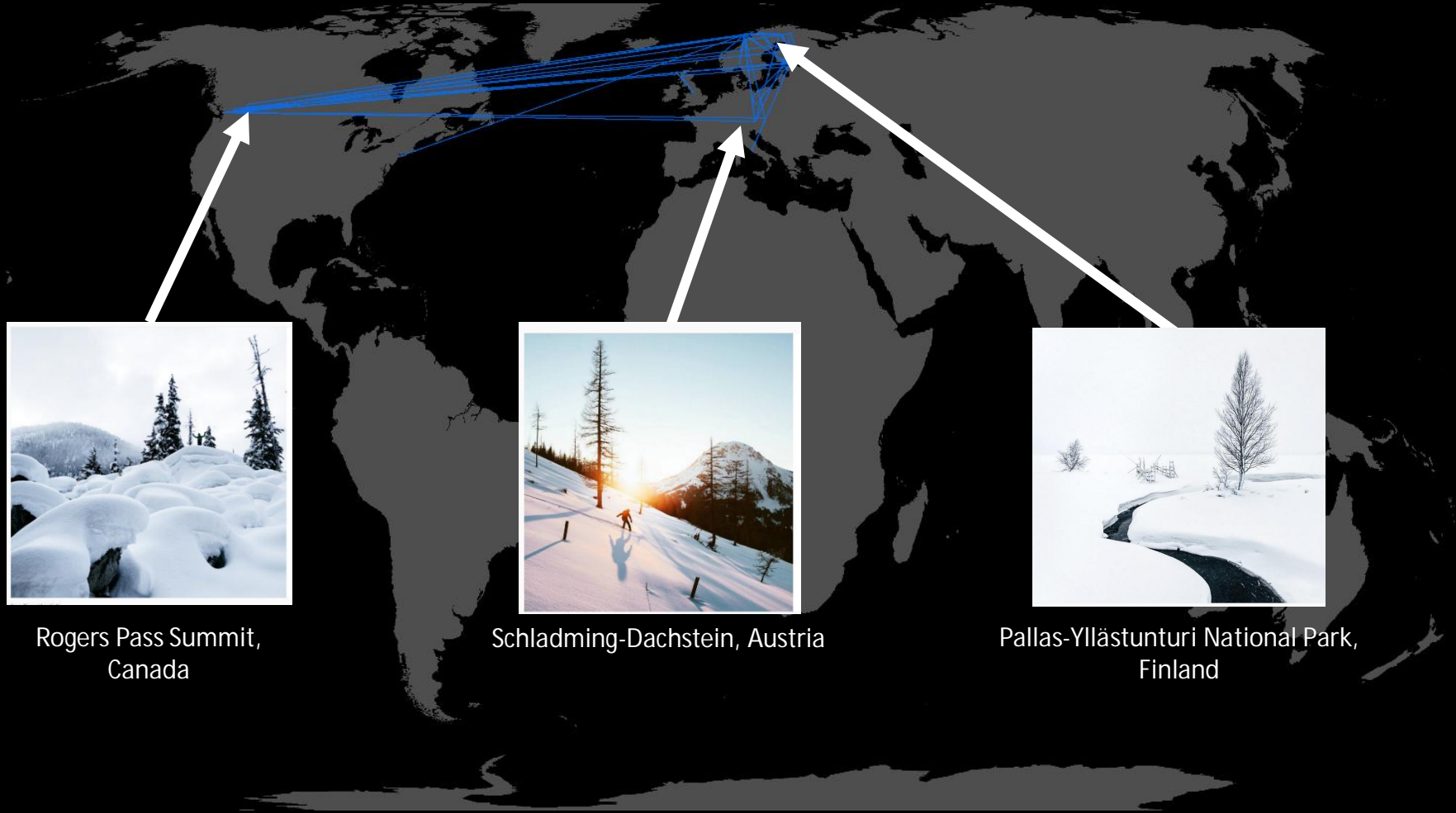


# PALLAS-YLLÄKSEN KANSALLISPUISTO

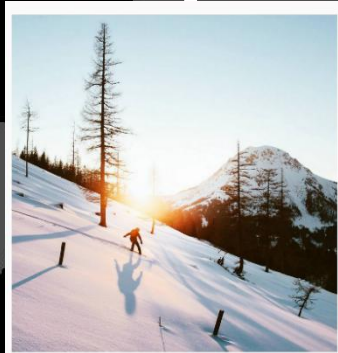


Posts from a Finnish Instagram-user. 107 posts in total.

# PALLAS-YLLÄKSEN KANSALLISPUISTO



Rogers Pass Summit,  
Canada



Schladming-Dachstein, Austria

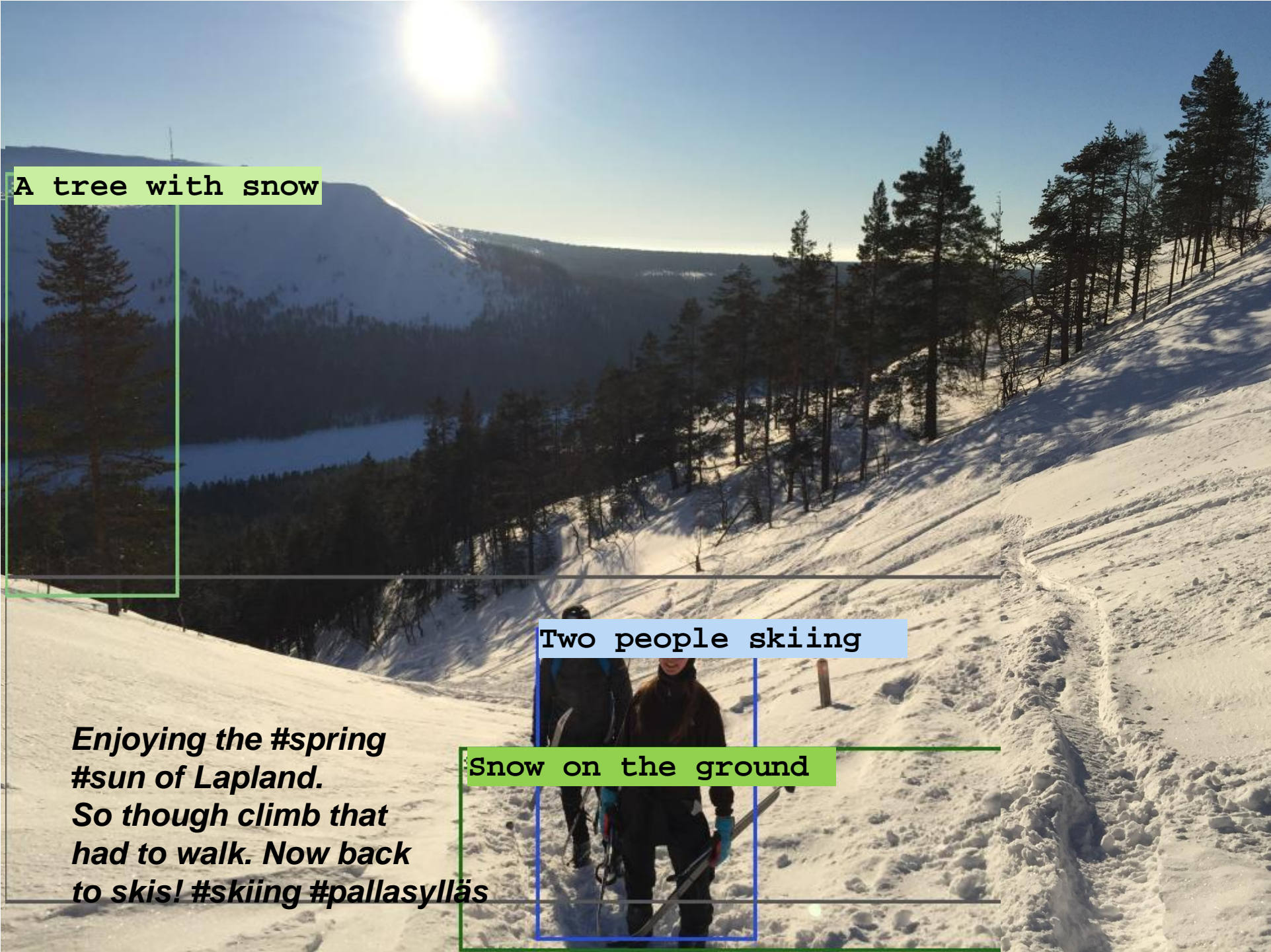


Pallas-Yllästunturi National Park,  
Finland

Posts from a Finnish Instagram-user. 107 posts in total.



***Enjoying the #spring  
#sun of Lapland.  
So though climb that  
had to walk. Now back  
to skis! #skiing #pallasylläs***



A tree with snow

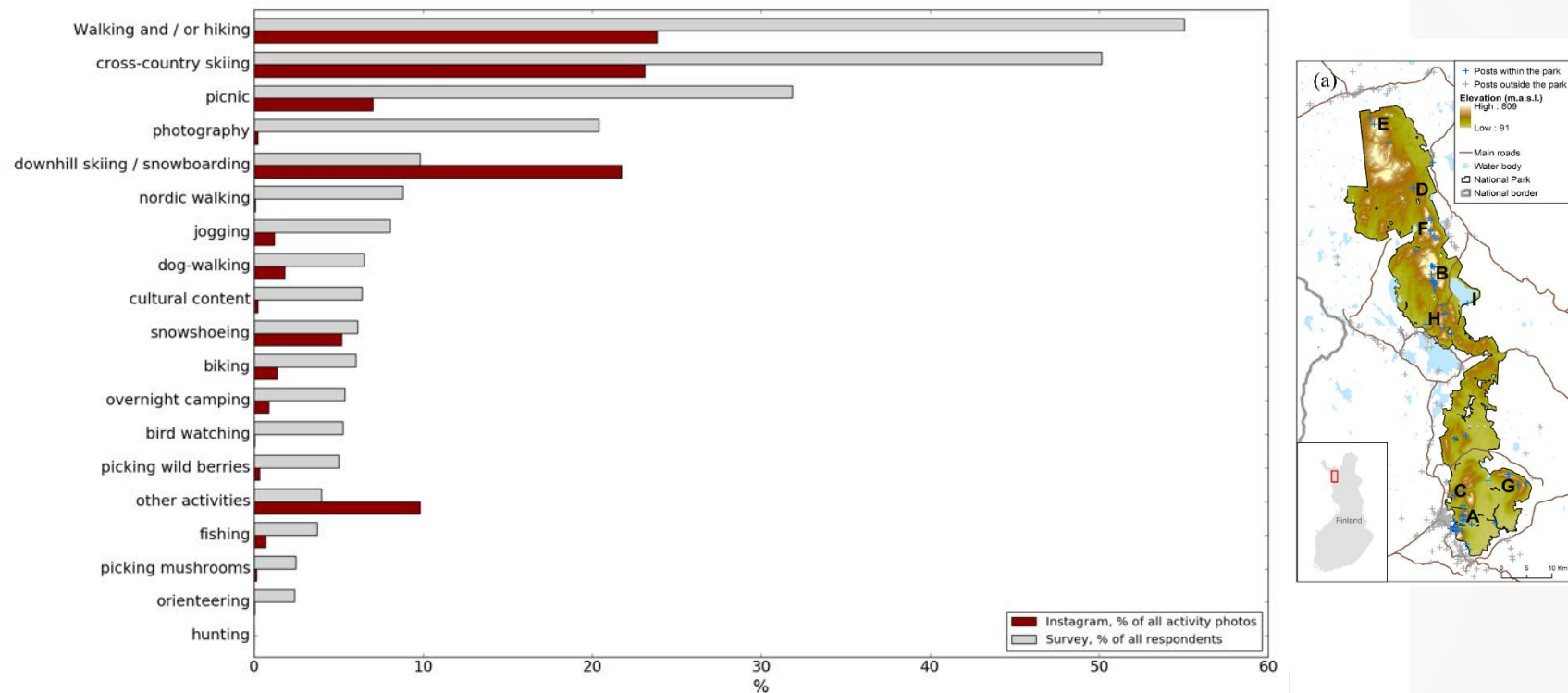
Two people skiing

*Enjoying the #spring  
#sun of Lapland.  
So though climb that  
had to walk. Now back  
to skis! #skiing #pallasylläs*

Snow on the ground

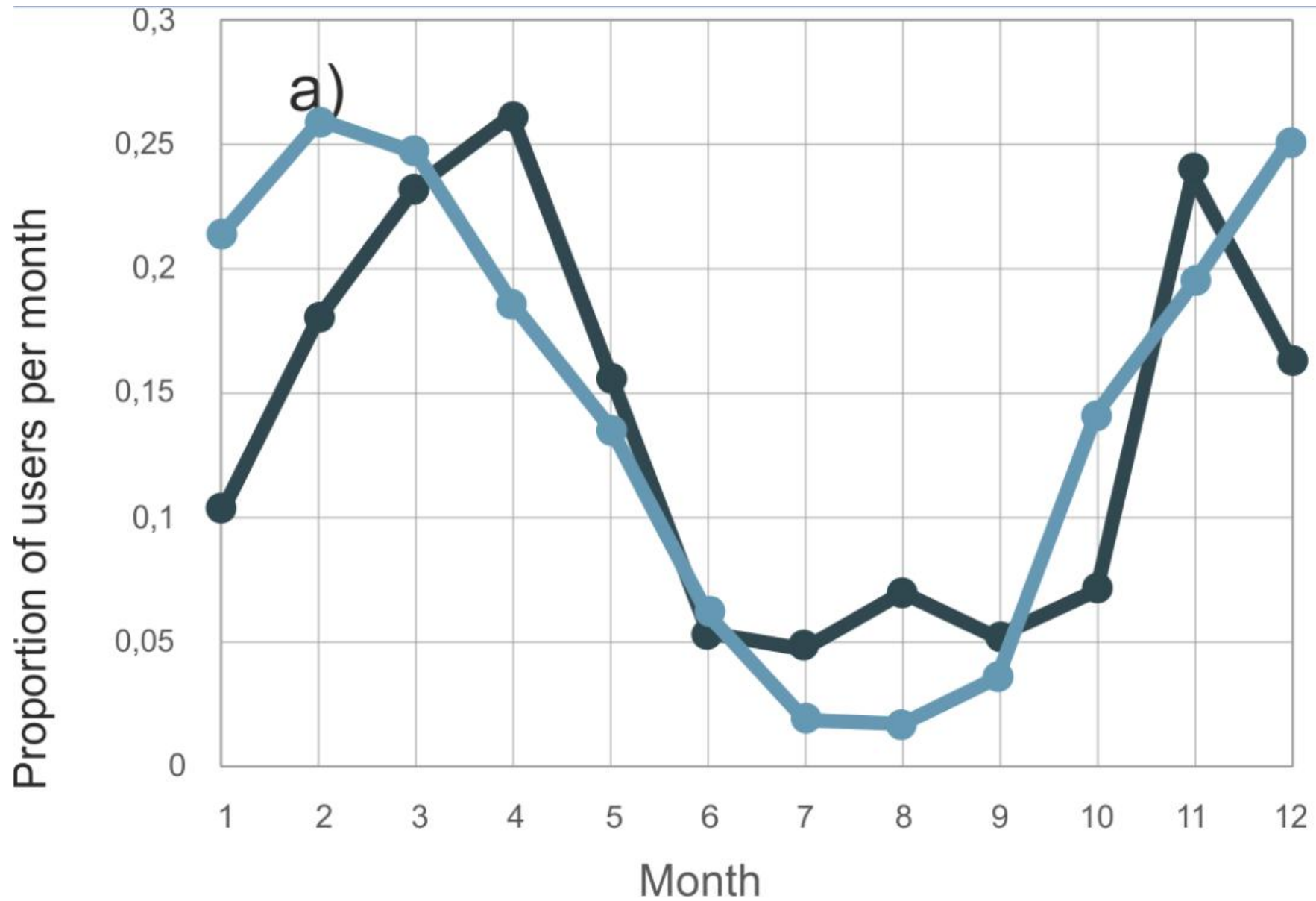


# ACTIVITIES: GROUND TRUTHING (PALLAS-YLLÄS)



**Figure 4.** Proportion of respondents / Instagram photos per activity in PY National Park. Activities are sorted by popularity in the survey.

(Heikinheimo et al. 2017, *ISPRS International Journal of Geo-Information*)



Heikinheimo et al. 2017,  
ISPRS International Journal of Geo-Information)

— snow  
— skiing



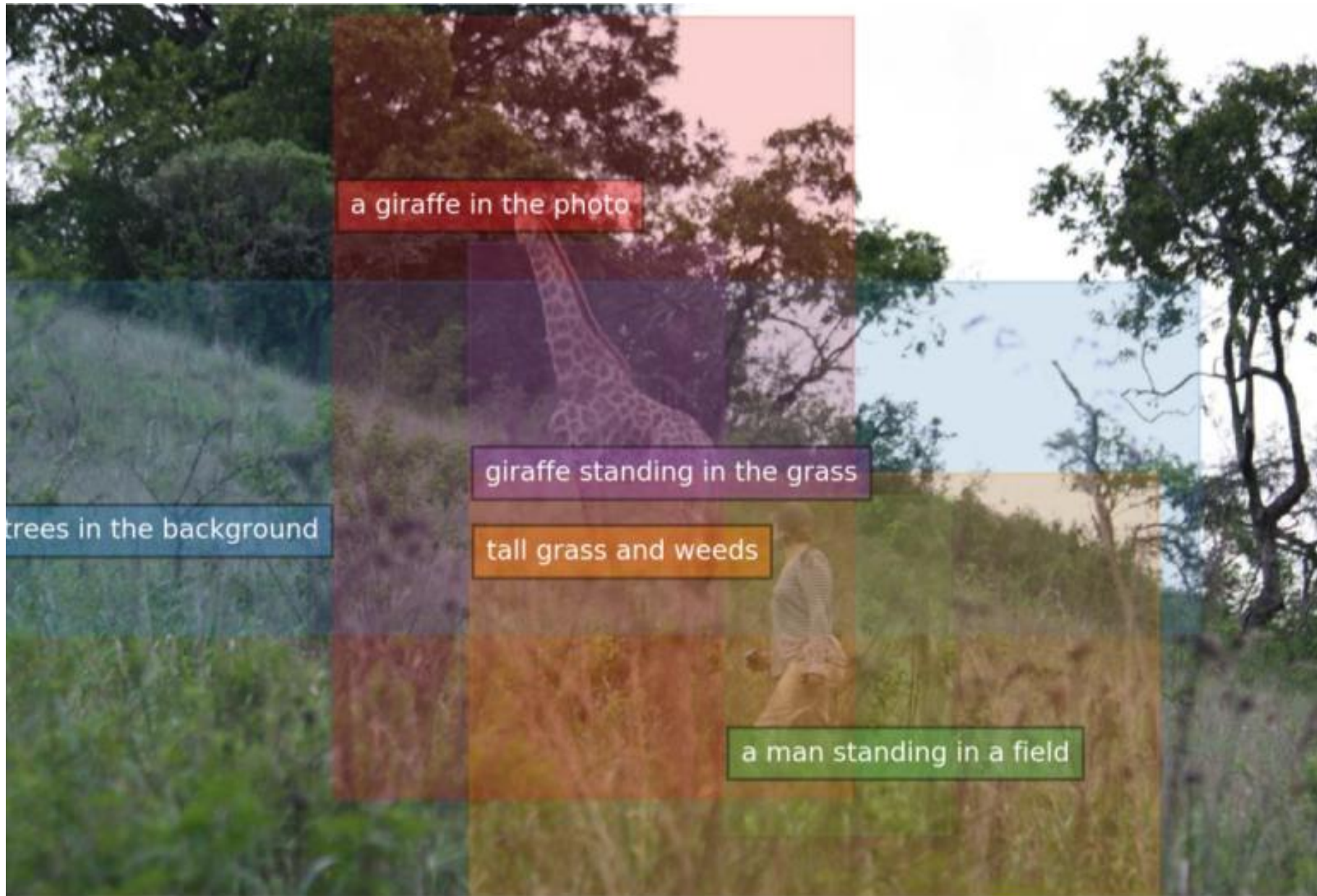
**HELSINGIN YLIOPISTO  
HELSINGFORS UNIVERSITET  
UNIVERSITY OF HELSINKI**



**HELSINGIN YLIOPISTO**  
**HELSINGFORS UNIVERSITET**  
**UNIVERSITY OF HELSINKI**

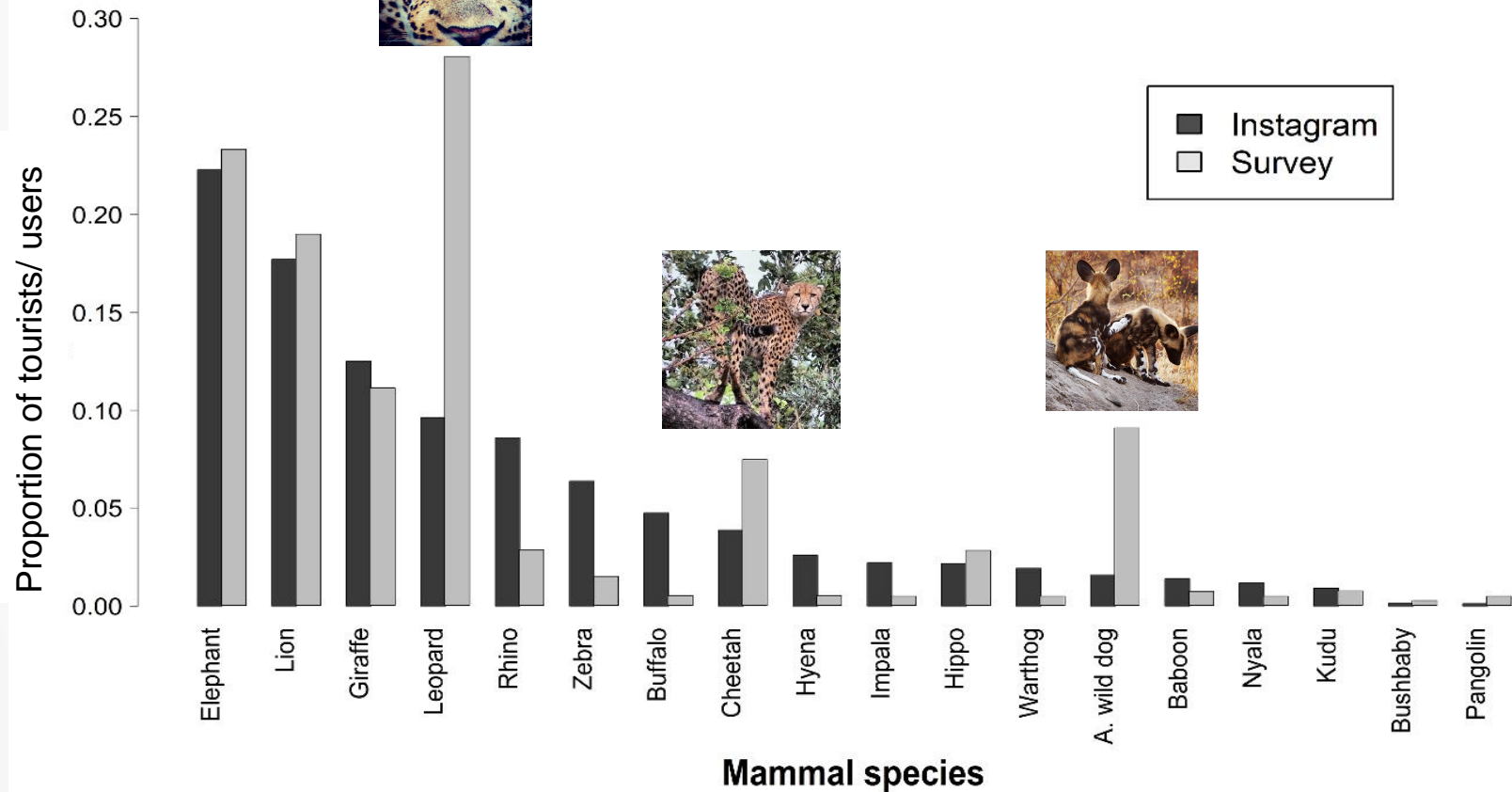
Instance segmentation using a pre-trained deep neural network, namely Mask R-CNN (He et al., 2017) with a ResNeXt X-101-32x8d-FPN backbone, which is available in the Detectron library (Girshick et al., 2018). Implemented by T. Hiippala.







# MATKAILIJOIDEN TOIVEET VS. SOSIAALINEN MEDIA (KRUGER)

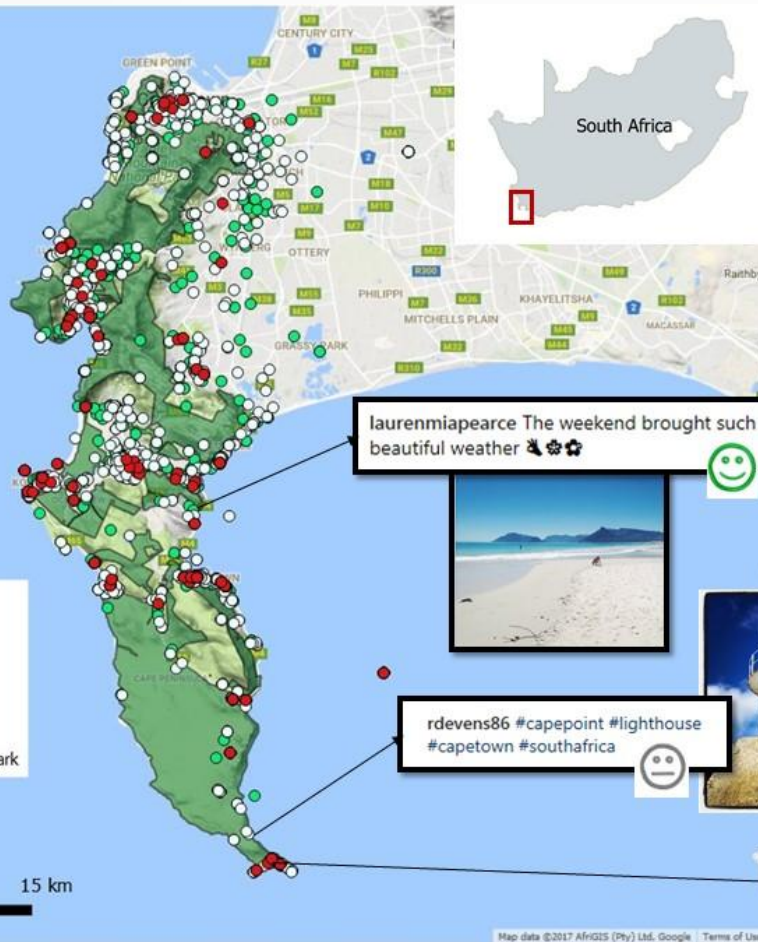




# SÄVYANALYYSSEILLÄ LISÄÄ YMMÄRRYSTÄ



Instagram posts  
N = 4013



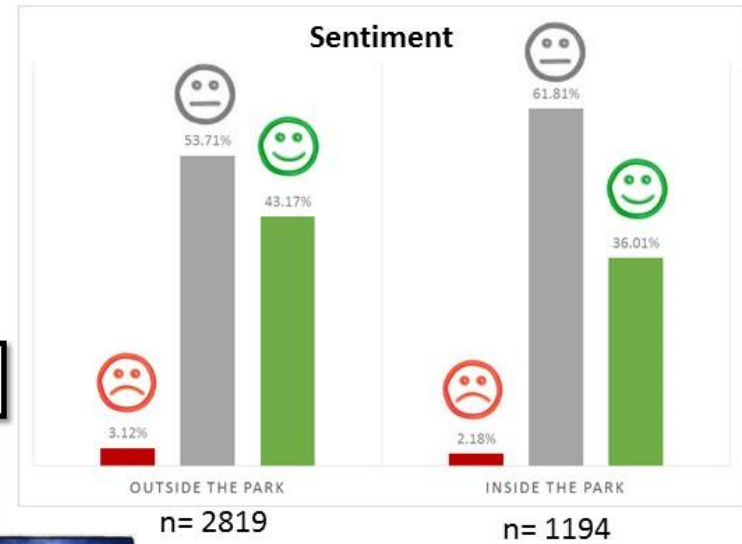
laurenmiapearce The weekend brought such beautiful weather 🌊🌴🌞



rdevens86 #capepoint #lighthouse #capetown #southafrica



deguinan Monkey Problems  
wellthatsfine Yo those things are terrifying





KUKA?

KÄYTTÄJÄ-  
PROFIILIT

SISÄLTÖ

IHMISTEN  
LIKKUMINEN

SIJAINTI

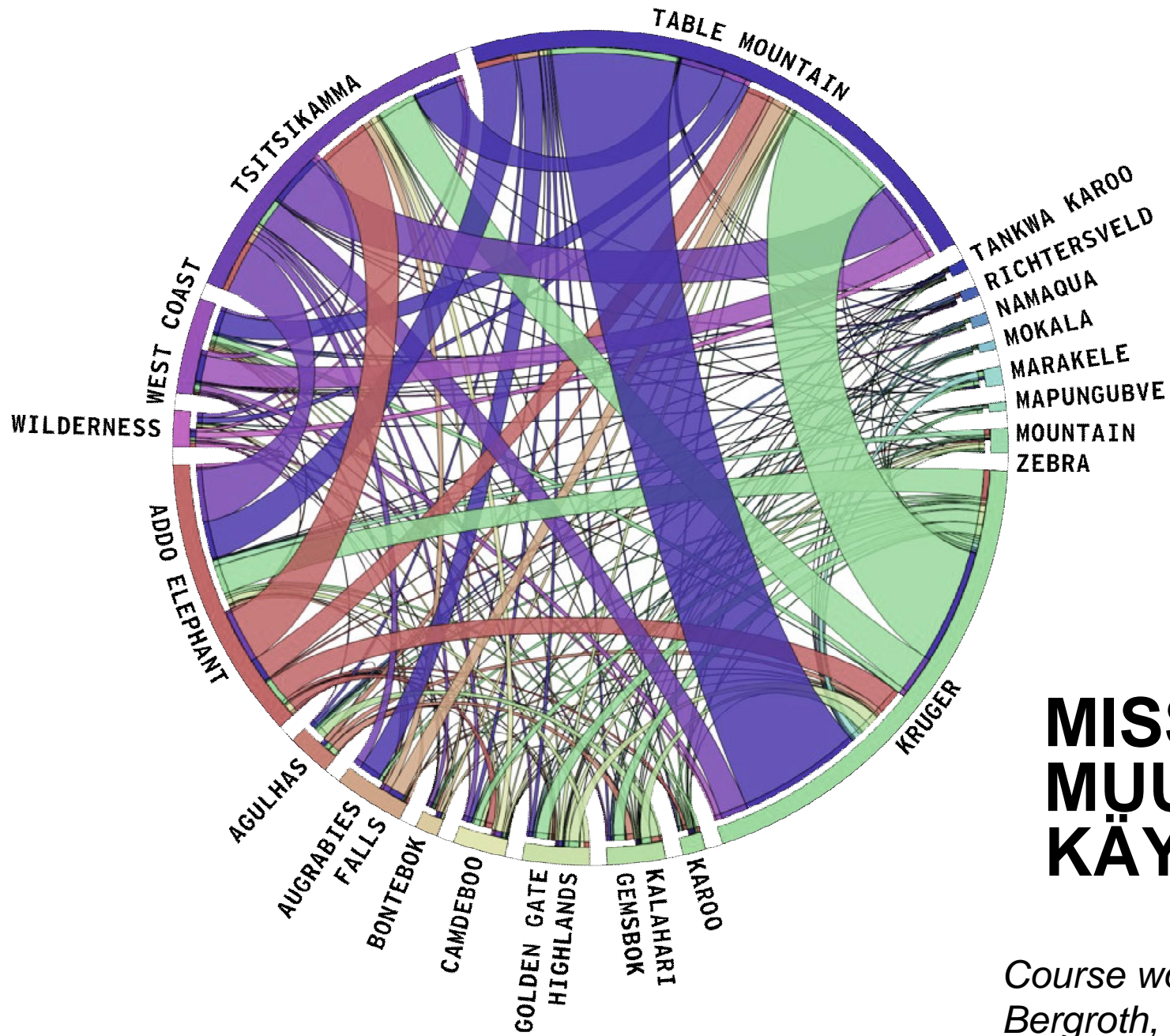
AIKALEIMA



# KANSALLISPUISTO: PAIKALLISIA VAI KANSAINVÄLISIÄ KÄVIJÖITÄ?

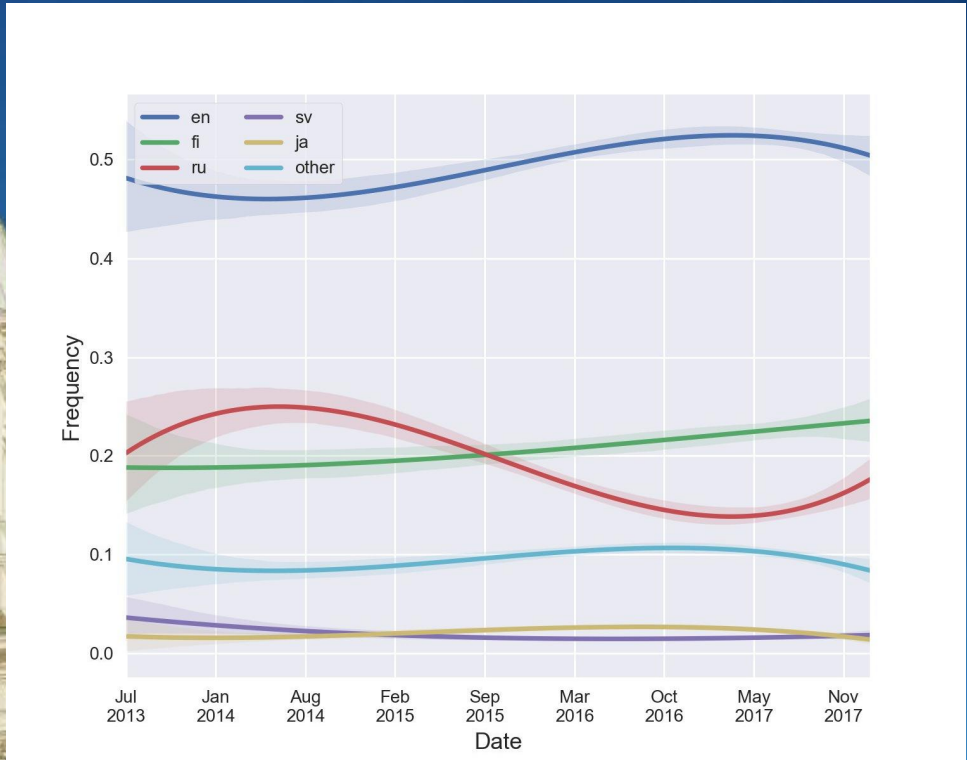


Detected home country of visitors to Kruger National Park, South Africa (Tenkanen, H.)



# MISSÄ MUUALLA KÄYNEET?

*Course work by Claudia Bergroth, Elias Willberg and Juha Lähteenmäki*



**1**  
Gorgeous light and awesome views at the Helsinki Cathedral!

Helsinki Cathedral

**2**  
Letzter Tag in Helsinki. Leider sind wir schon auf dem Weg zum Flughafen ...

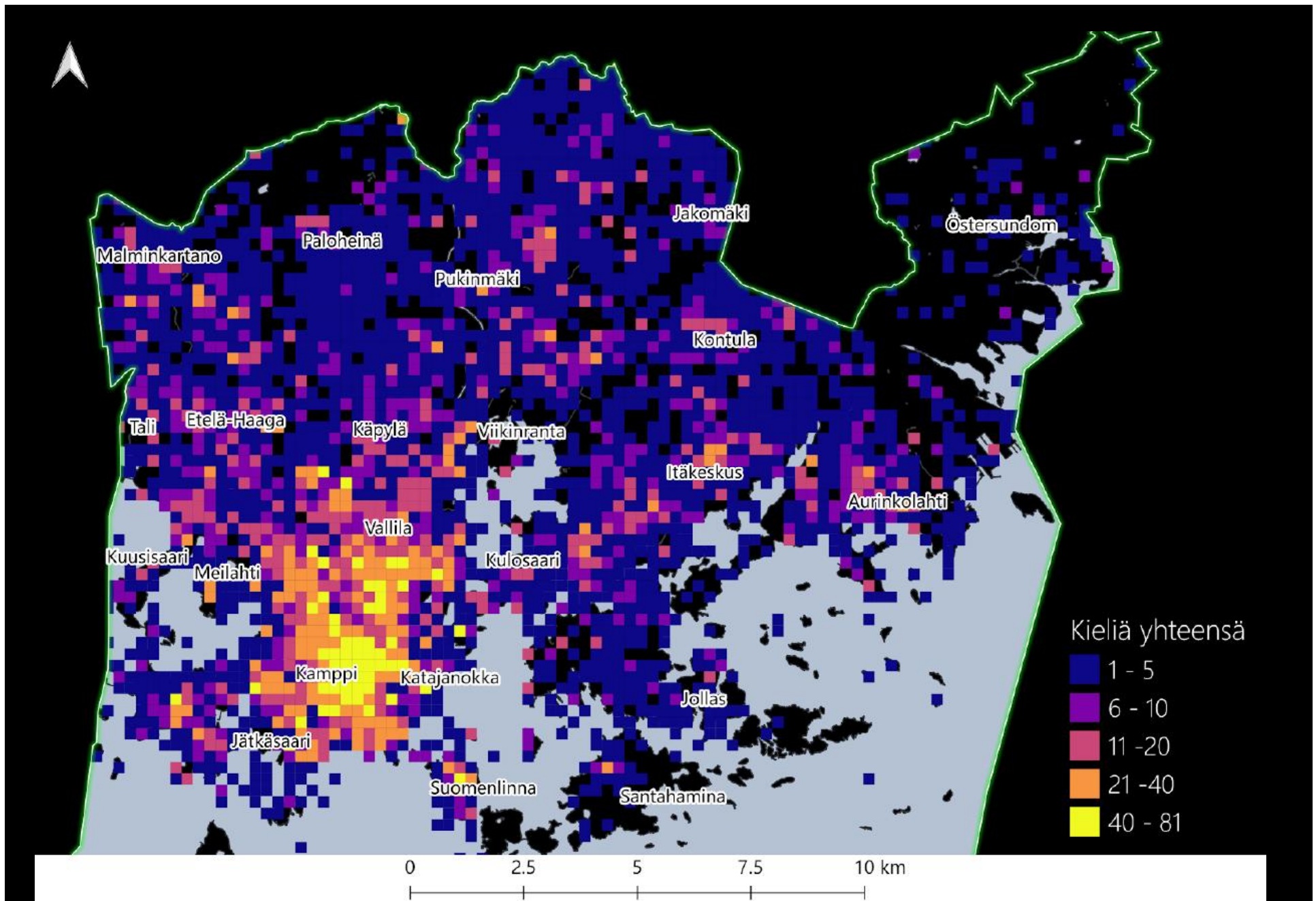
Helsinki, Finland (60.17...,24.95...)

**3**  
Сенатская площадь. Солнце светит!

(60.170117, 24.951721)

Hiippala et al. 2018

*Digital Scholarship in the Humanities*, fqy049, <https://doi.org/10.1093/llc/fqy049>



Tuomas Väisänen, 2018, MSc thesis



# KEITÄ IHMISET OVAT?



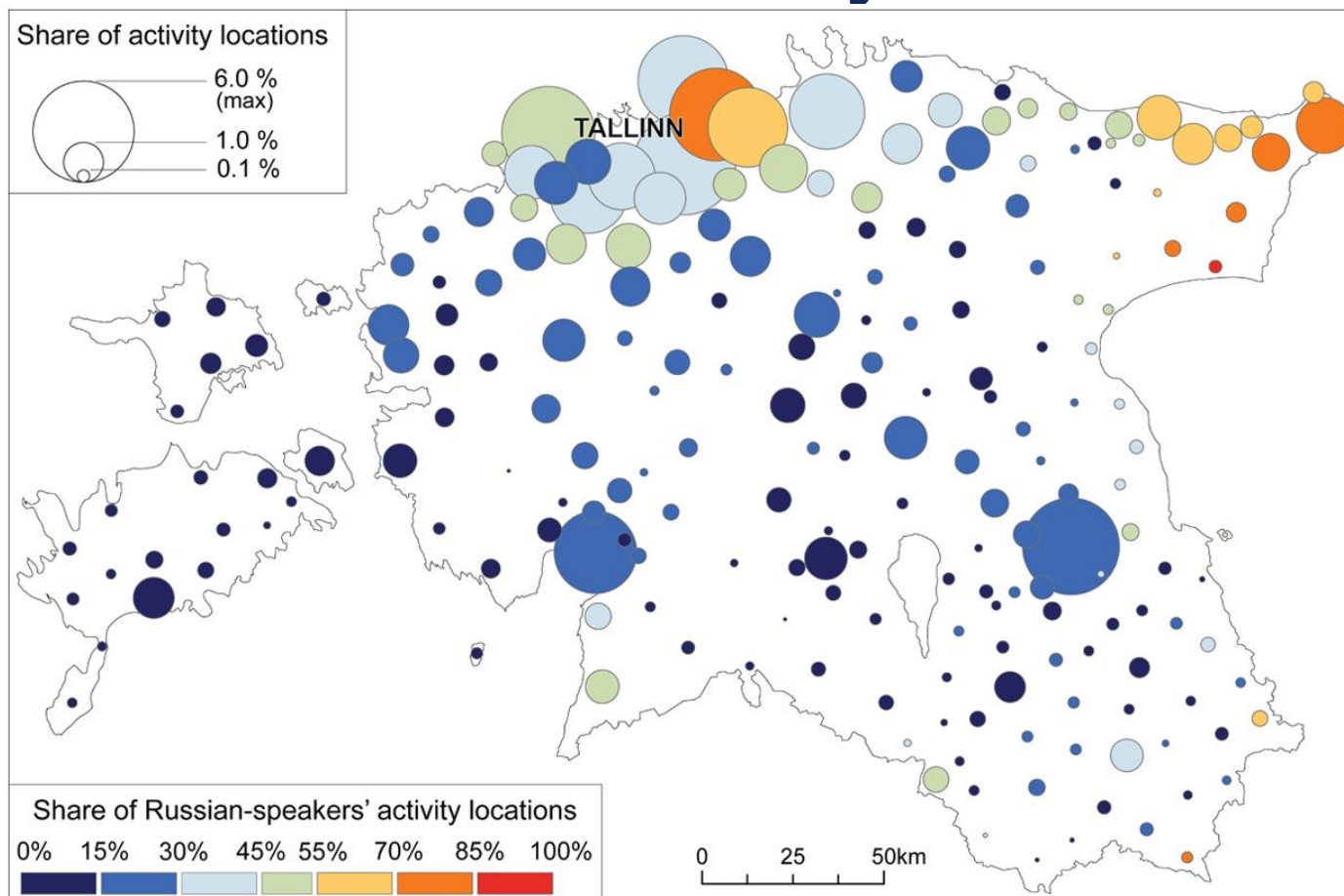
# KEITÄ IHMISET OVAT?



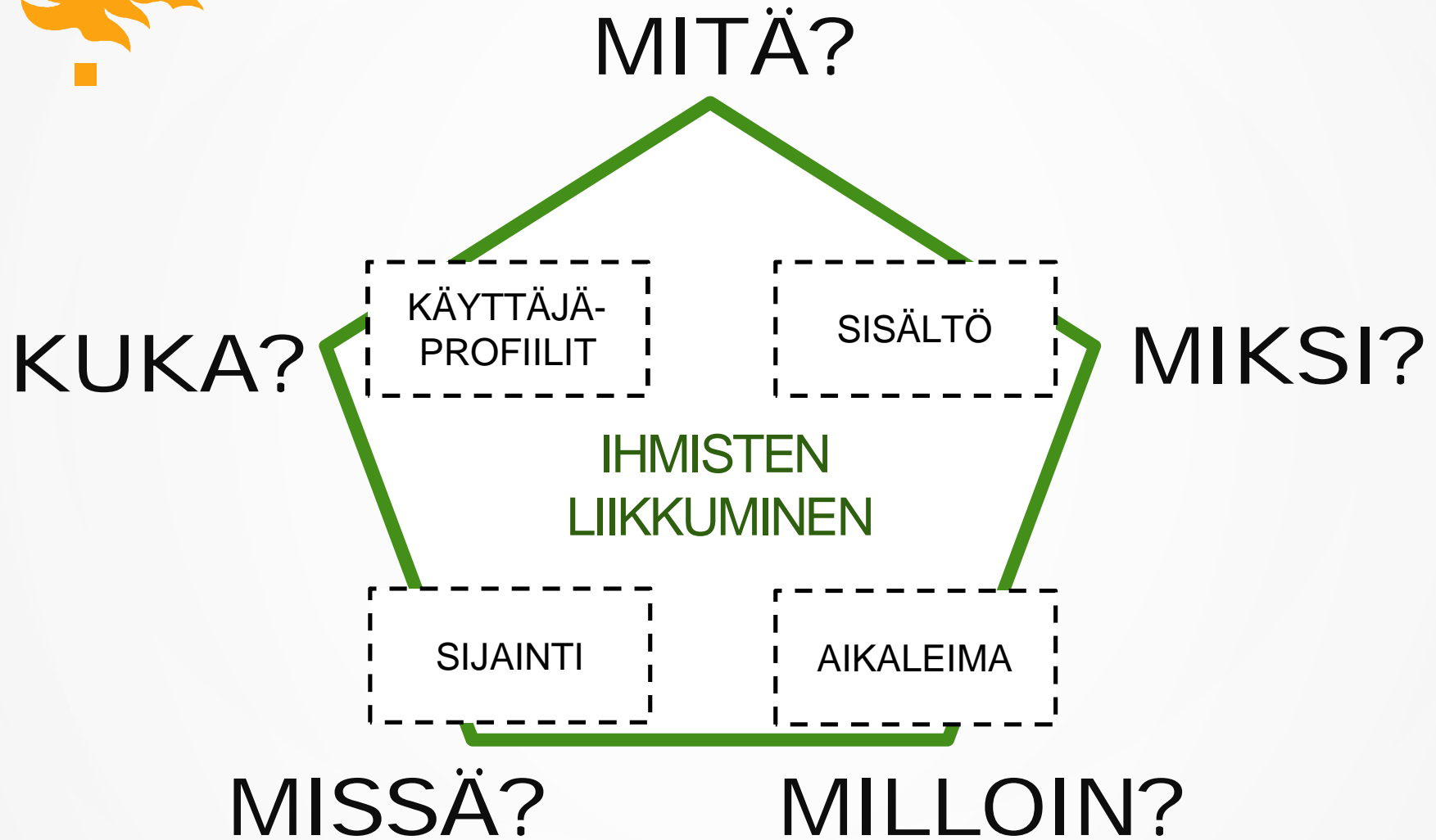


# **KUKA EI OLE MUKANA? MIKSI?**

# Taustan vaikutus tilankäyttöön virossa



Tallinnan ulkopuoliset vierailukohteet kuntatasolla:  
Symbolin koko kuvaa määrää ja väri etnistä taustaa.





**MASSADATAT  
TARJOAVAT AIVAN UUSIA  
NÄKÖKULMIA IHMISTEN  
TOIMINTAAN JA YHTEISKUNNAN  
DYNAMIIKKAAN**

**VOLUME**

**VELOCITY**

**VARIETY**



# ISO DATA VOI YLLÄTTÄÄ PIENUUDELLAAN

**VOLUME?**



# **AINEISTOJEN ANALYSOINTI ON VAATIVAA – YHTEISTYÖ JA JATKUVA KEHITYS ON TARPEEN**

**VARIETY**

**VARIABILITY**

**VERACITY**





# **VISUALISOINTI, ANALYSOINTI, VERIFIONTI**



# VISUALISOINTI, ANALYSOINTI, VERIFIONTI

**VOLUME**

**VERACITY**

**VELOCITY**



# VINOUMAT OVAT TODELLISIA, NIITÄ ON POHDITTAVA

VERACITY



**KENELLÄ ON OIKEUS DATAAN?**

**LÄPINÄKYVYYDEN IHANTEESTA  
SULJETTUIHIN ANALYYSSEIHIN?**



# MASSADATOISSAKIN SIJAINNITieto YHDISTÄÄ

**VOLUME**

**VELOCITY**

**VARIETY**



**VALTAVASTI MAHDOLLISUUKSIA!**

**HYVÄÄN?**

**VOLUME**

**VELOCITY**

**VARIETY**

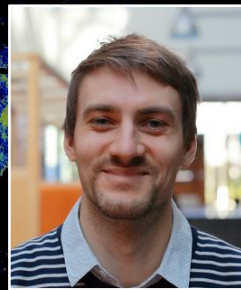
# DIGITAL GEOGRAPHY LAB+



Tuuli  
Toivonen



Enrico  
Di Minin



Olle  
Järvi



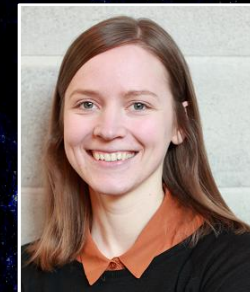
Henrikki  
Tenkanen



Anna  
Hausmann



Tuomo  
Hiippala



Vuokko  
Heikinheimo



Christoph  
Fink



Johanna  
Eklund



Gonzalo  
Cortés Capano



Joel  
Jalkanen



Kerli  
Mürisepp



Elias  
Willberg



Claudia  
Bergroth

# KIITOS!

@DigiGeoLab

@Tuuli Toivonen



**HELSUS**  
HELSINKI INSTITUTE OF SUSTAINABILITY SCIENCE



**KONE FOUNDATION**



**European Research Council**  
Established by the European Commission

**KIITOS!**