BIG DATA IN UNDERSTANDING OUTDOOR RECREATION AND ACTIVE MOBILITY

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

BIG DATA IN UNDERSTANDING OUTDOOR RECREATION AND ACTIVE MOBILITY

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

DIGITAL GEOGRAPHY LAB

Tuuli Toivonen

Olle Järv

Petteri Muukkonen

Johanna Eklund

Tuomo Hiippala

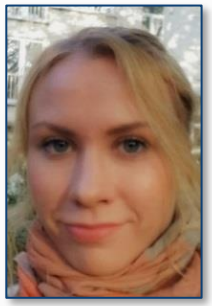
Age Poom

Janika Raun

Christoph Fink

Charlotte van der Lijn

Pengyuan Liu



Kerli Müürisepp

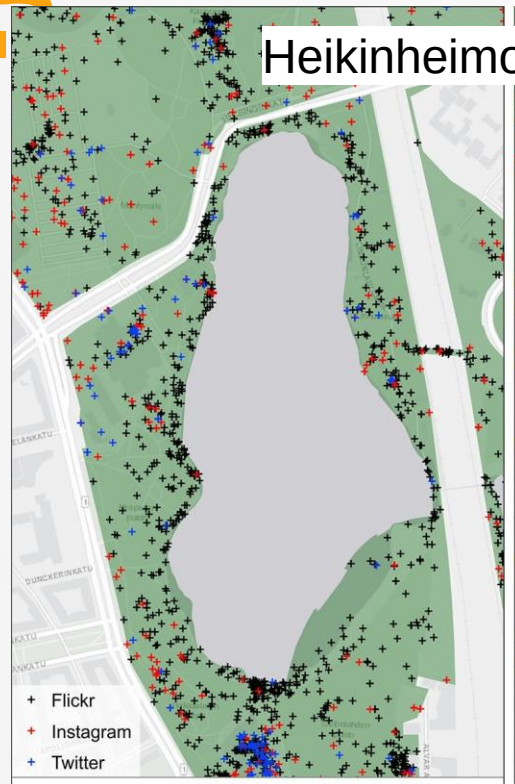
Elias Willberg

Tuomas Väisänen

Håvard Aagesen

USER-GENERATED MOBILE DATA

Heikinheimo et al. 2020 https://doi.org/10.1016/j.landurbplan.2020.103845



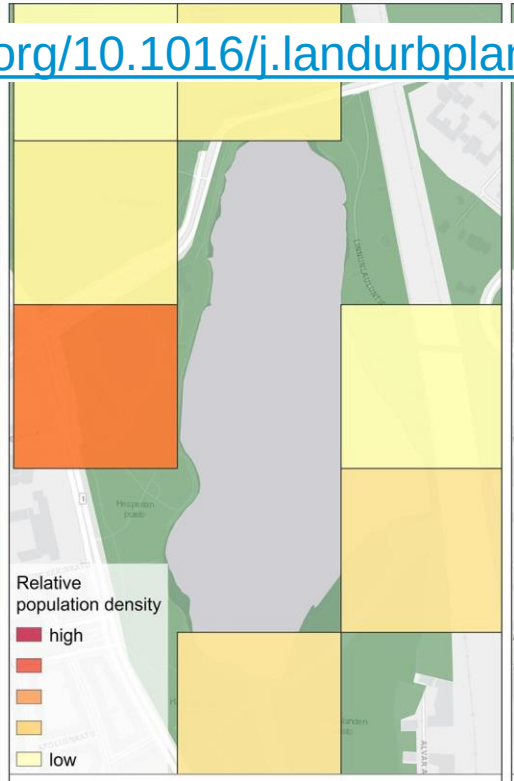
social media data
spatial unit: point

- user id
- post id
- timestamp
- link to media
- caption text
- comment text
- number of likes

sports tracking data
spatial unit: line

- way id
- timestamp
- number of people
- number of trips
- number of commute trips

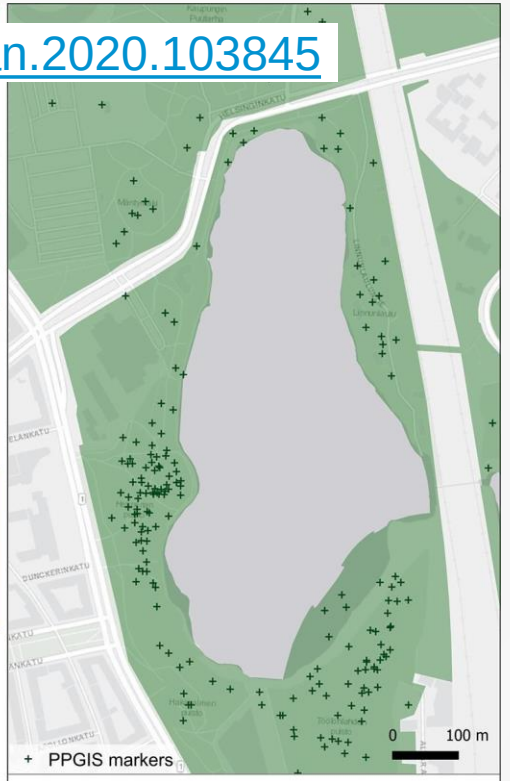
Example of sports tracking data attached to road segments.



mobile phone data
spatial unit: polygon

- grid id
- timestamp
- number of people

Example of mobile phone data interpolated to a grid.



PPGIS data
spatial unit: point

- user id
- marker id
- answers
- comments
- ...

Example of point markers from PPGIS.

TWO VIEWING ANGLES

1. Active urban mobility

2. Recreational use of green areas

TWO VIEWING ANGLES

1. Active urban mobility

Health

Sustainability

Equality

2. Recreational use of green areas

TWO VIEWING ANGLES

1. Active urban mobility

Health

Sustainability

Equality

2. Recreational use of green areas

MOBILE BIG DATA

BIG DATA AS AN INFORMATION SOURCE

- The emergence of big novel data sources (Kitchin, 2014) provide new opportunities – *digital footprints* – to understanding the realities of our environment and societies.

SOURCES OF SPATIAL BIG DATA ON PEOPLE

COLLECTION CAMPAIGNS

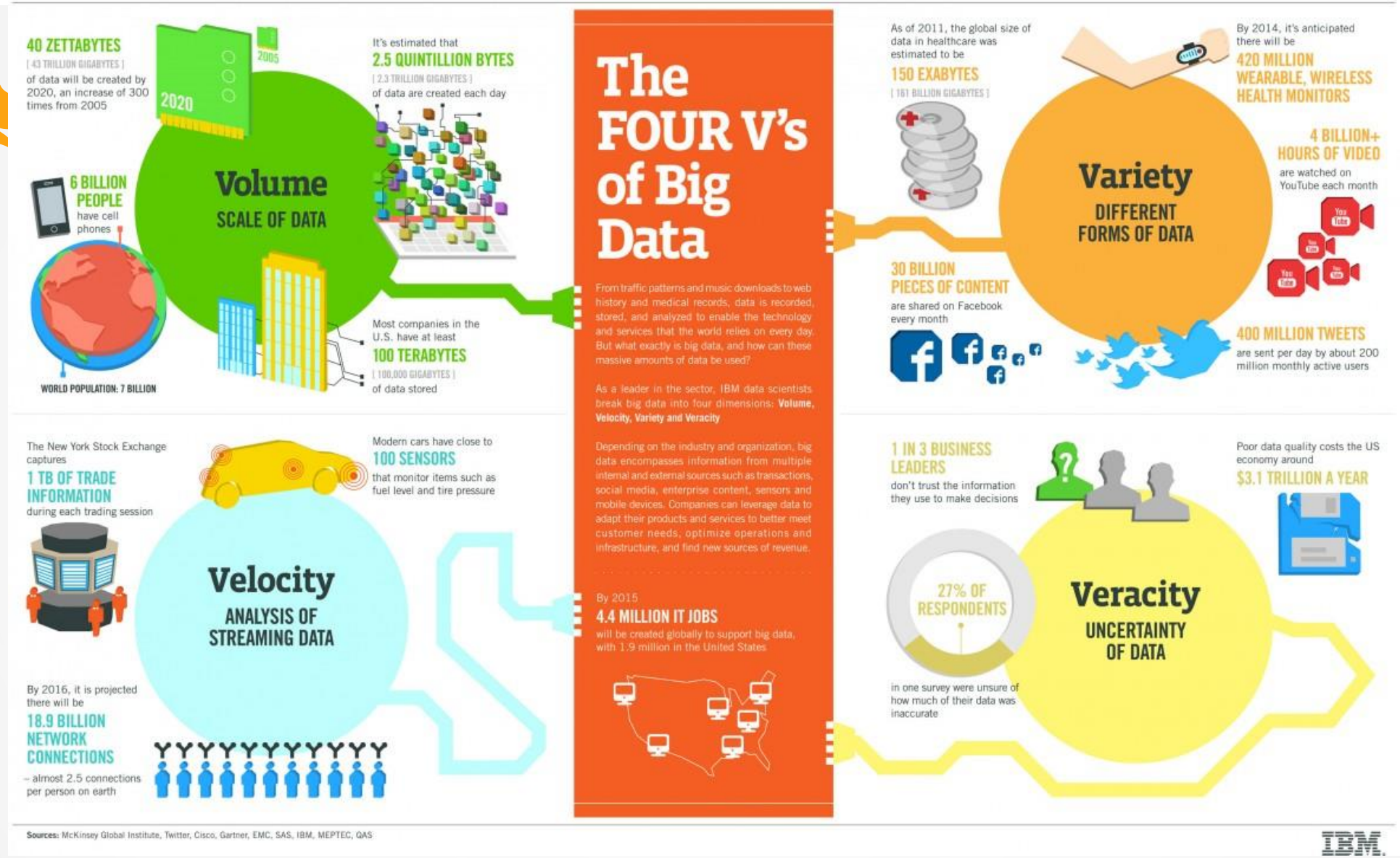
- Remote sensing, laser scanning
- Surveillance cameras (CCTV; ANPR)
- Visitor counters
- Registrations (border control)

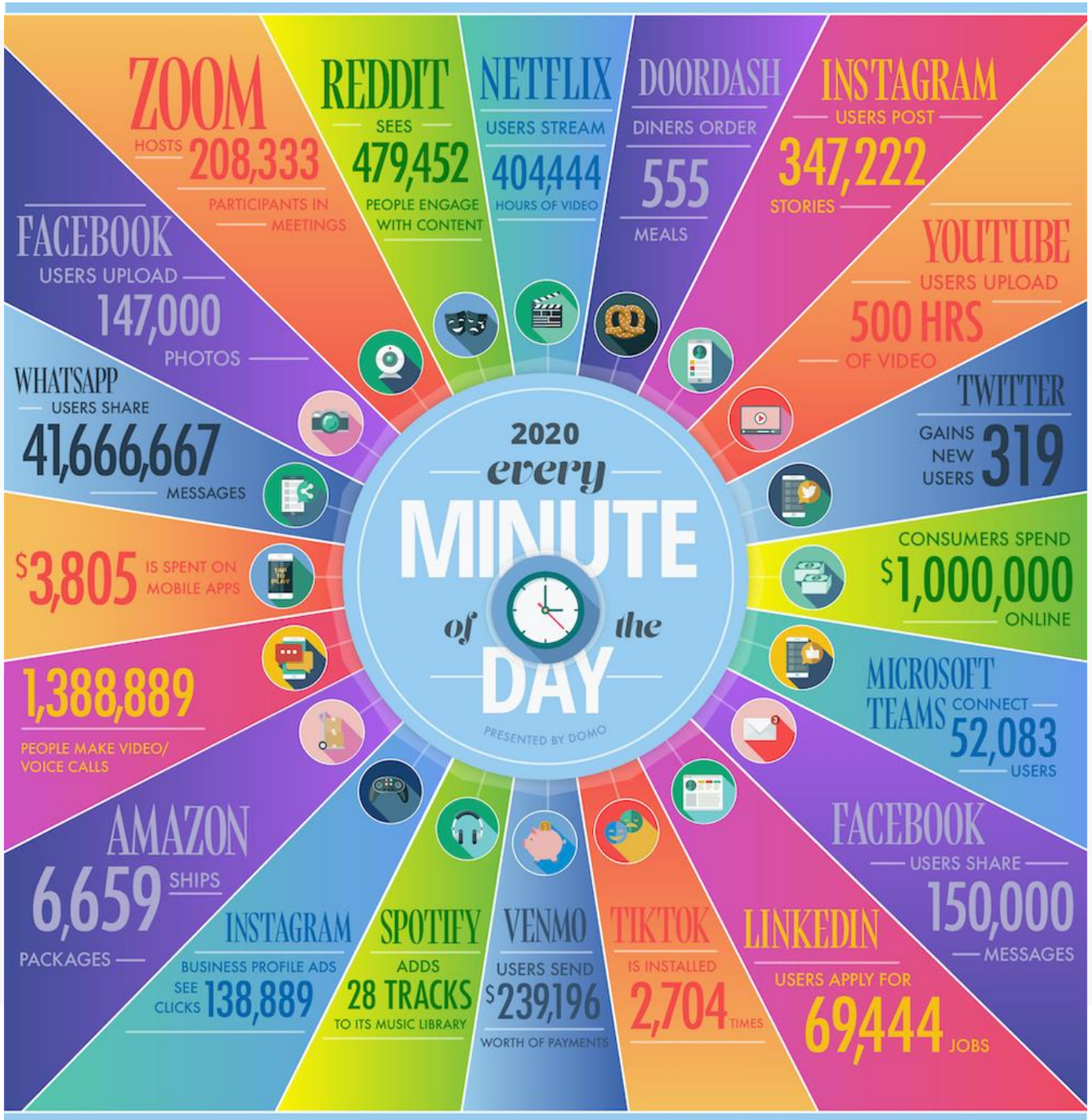
USER-GENERATED CONTENT (AS A SIDE PRODUCT)

- Use of smart phone
- Loyalty cards
- Internet use (WLAN, clicks)
- Use of mobile Apps

USER-GENERATED CONTENT (ACTIVE)

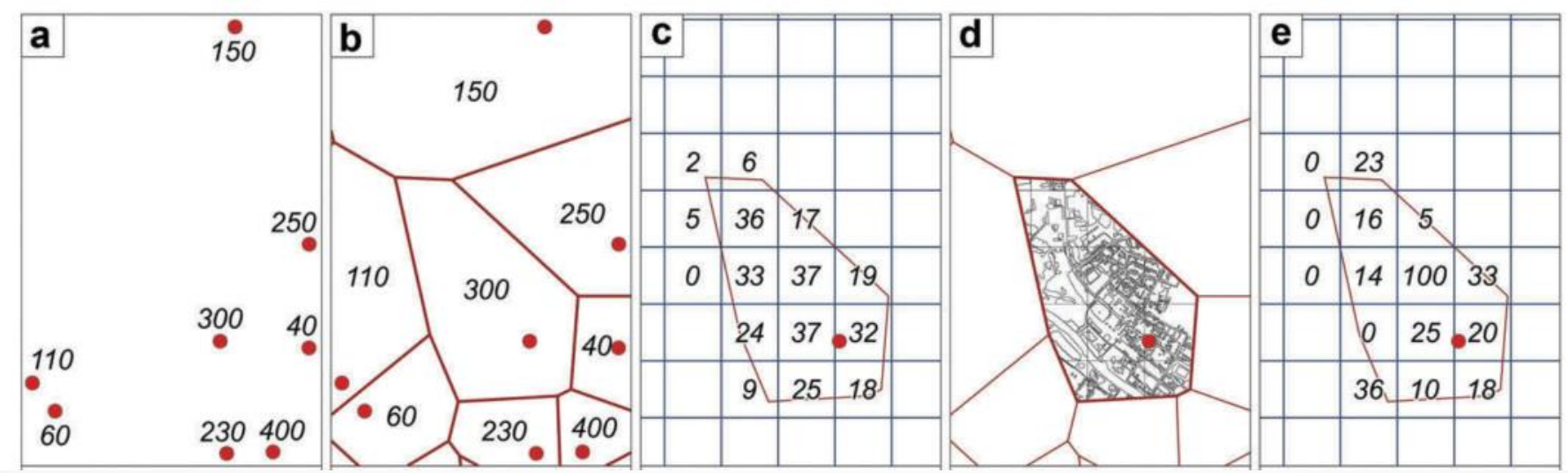
- Social Media
- Sports Applications


40 ZETTABYTES
[43 TRILLION GIGABYTES]
of data will be created by 2020, an increase of 300 times from 2005
2005
2020
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
27% OF RESPONDENTS
in one survey were unsure of how much of their data was inaccurate
Veracity
UNCERTAINTY OF DATA
Poor data quality costs the US economy around
$3.1 TRILLION A YEAR
Sources: McKinsey Global Institute, Twitter, Cisco, Gartner, EMC, SAS, IBM, MEPTEC, QAS
IBM.


2020
every
MINUTE
of the
DAY
PRESENTED BY DOMO
ZOOM
HOSTS 208,333
PARTICIPANTS IN MEETINGS
REDDIT
SEES 479,452
PEOPLE ENGAGE WITH CONTENT
NETFLIX
USERS STREAM 404444
HOURS OF VIDEO
DOORDASH
DINERS ORDER 555
MEALS
INSTAGRAM
USERS POST 347,222
STORIES
FACEBOOK
USERS UPLOAD 147,000
PHOTOS
YOUTUBE
USERS UPLOAD 500 HRS
OF VIDEO
WHATSAPP
USERS SHARE 41,666,667
MESSAGES
TWITTER
GAINS NEW USERS 319
$3,805 IS SPENT ON MOBILE APPS
CONSUMERS SPEND $1,000,000 ONLINE
1,388,889
PEOPLE MAKE VIDEO/ VOICE CALLS
MICROSOFT TEAMS CONNECT 52,083
USERS
AMAZON
6,659 SHIPS
PACKAGES
FACEBOOK
USERS SHARE 150,000
MESSAGES
INSTAGRAM
BUSINESS PROFILE ADS SEE CLICKS 138,889
SPOTIFY
ADDS 28 TRACKS
TO ITS MUSIC LIBRARY
VENMO
USERS SEND $239,196
WORTH OF PAYMENTS
TIKTOK
IS INSTALLED 2,704 TIMES
LINKEDIN
USERS APPLY FOR 69,444 JOBS

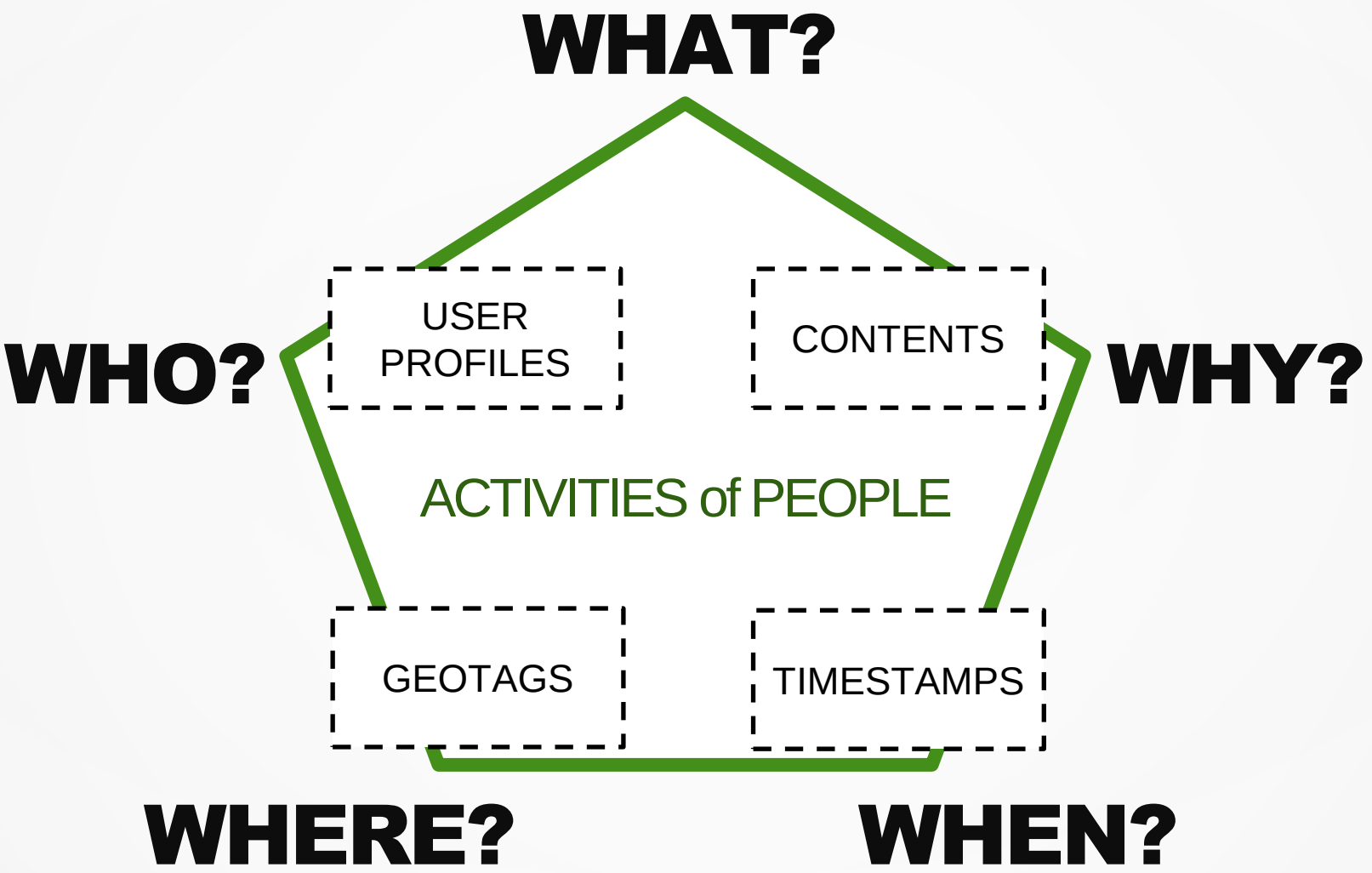
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E.
Social media data for conservation science: a methodological overview. Biological Conservation (2019).

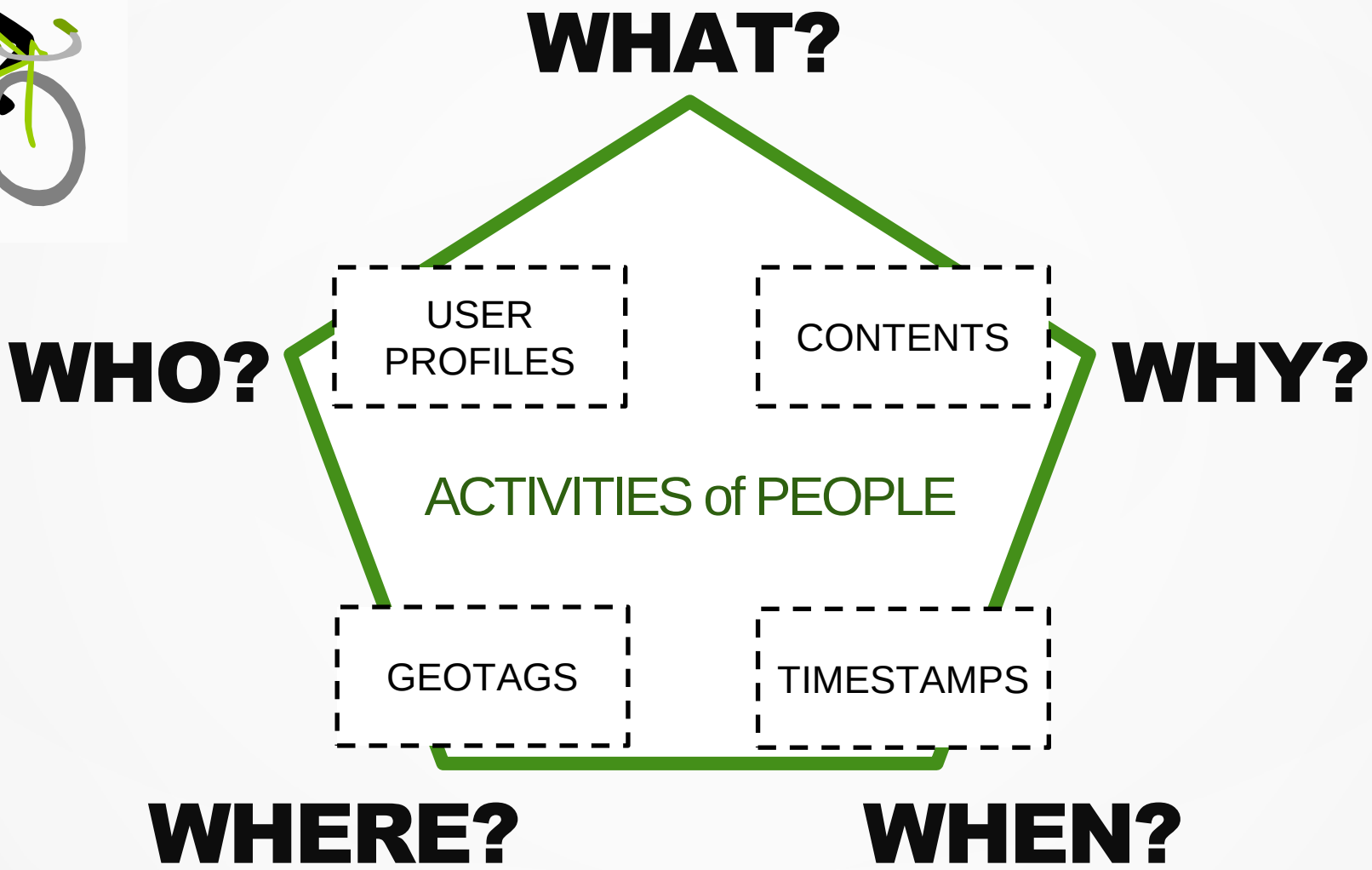


(Järv et al. 2017)

- ANTENNA LEVEL DATA
- CDR DATA
- EEA DATA

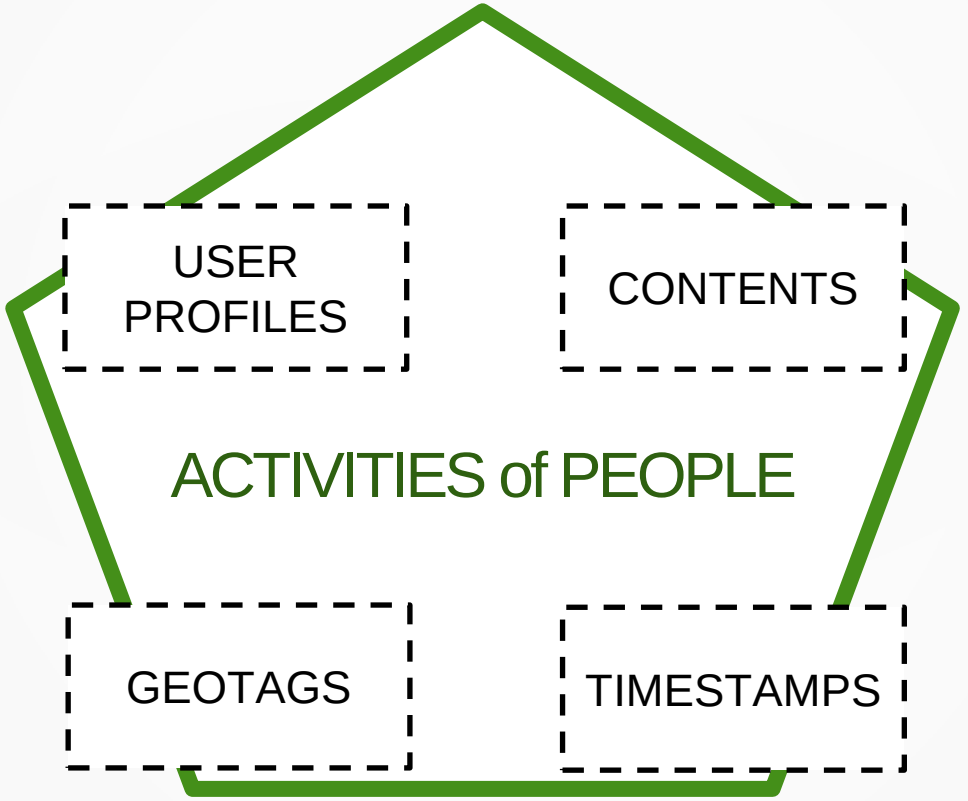


Di Minin, Tenkanen & Toivonen (2015)
Heikinheimo et al. 2020



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

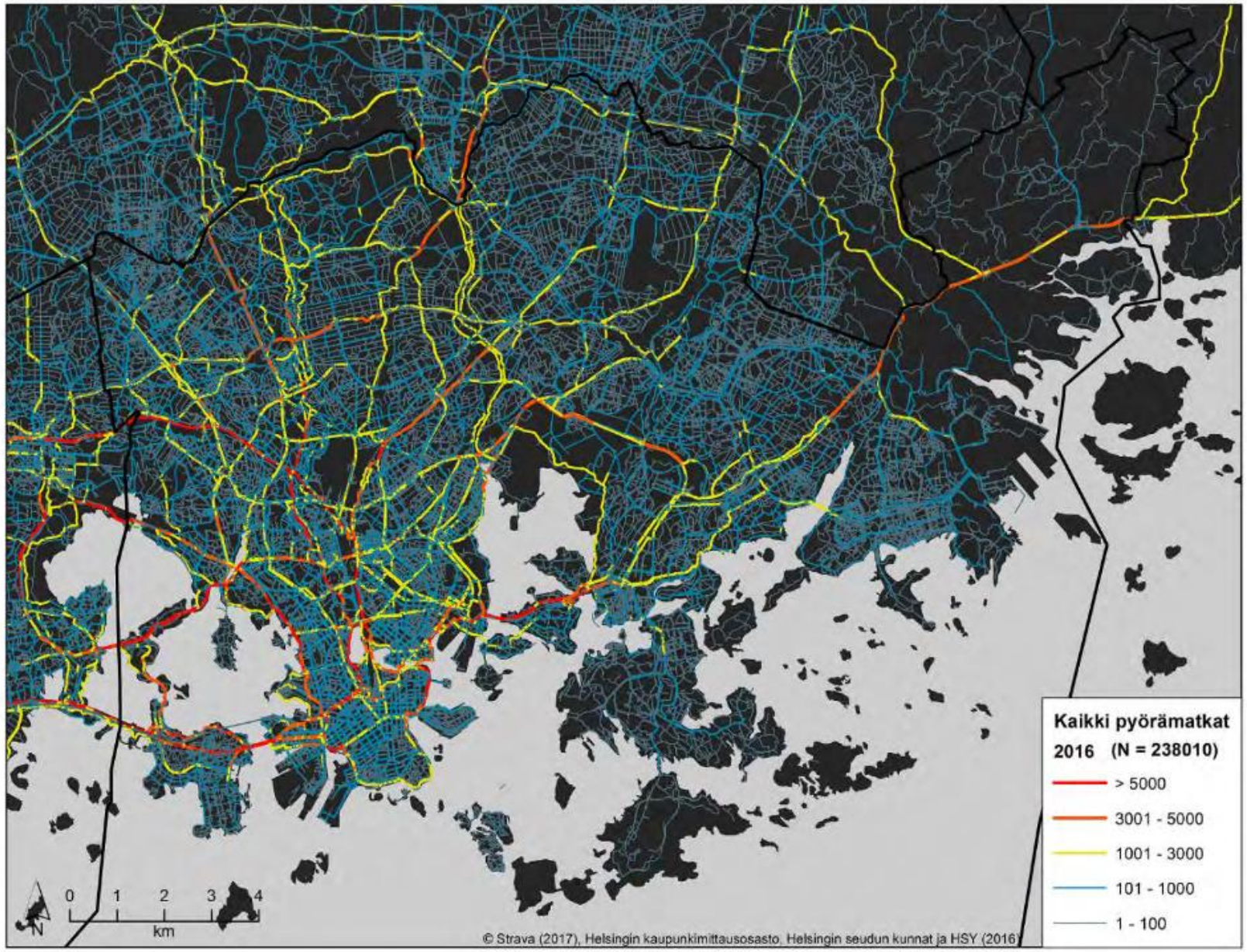
Di Minin, Tenkanen & Toivonen (2015)
Heikinheimo et al. 2020


USER PROFILES
CONTENTS
ACTIVITIES of PEOPLE
GEOTAGS
TIMESTAMPS

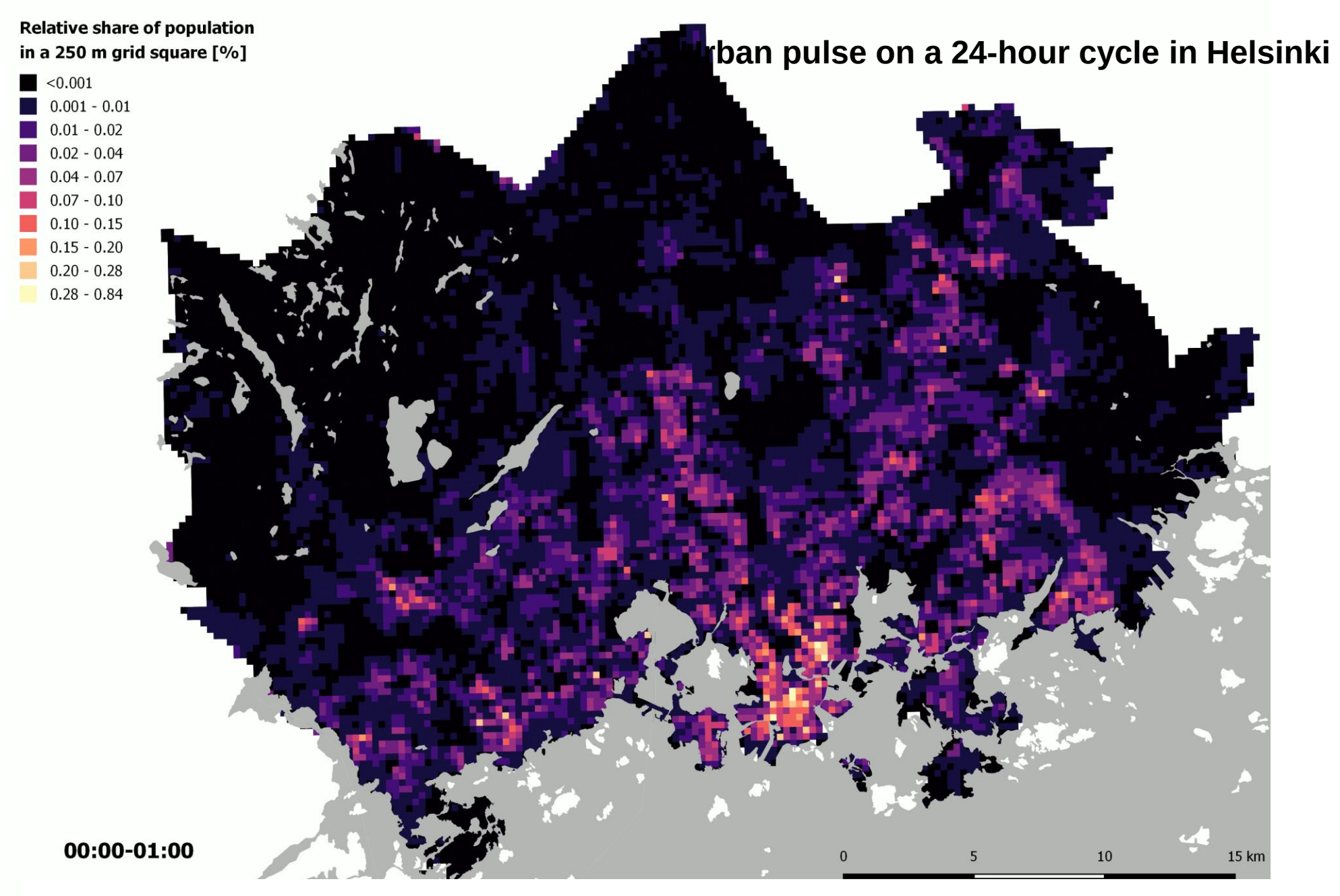
WHERE? **WHEN?**

BIKE TRIPS IN HELSINKI ACCORDING TO STRAVA

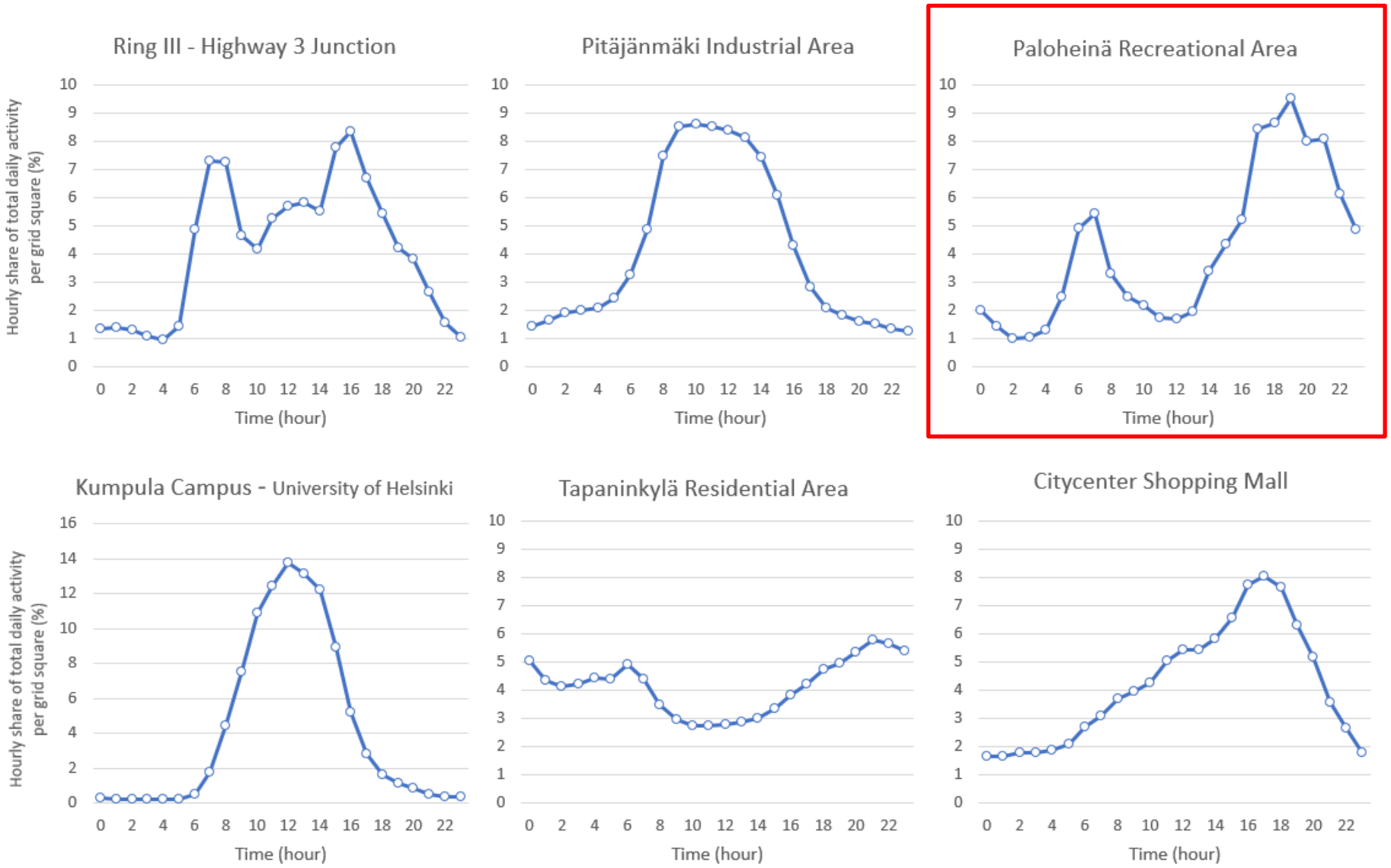
https://www.hel.fi/static/liitteet/kaupunkiymparisto/julkaisut/julkaisut/julkaisu-16-17.pdf



Kuva 26. Kaikki Strava-käyttäjien tallentamat pyörämatkat pääkaupunkiseudulla vuonna 2016.



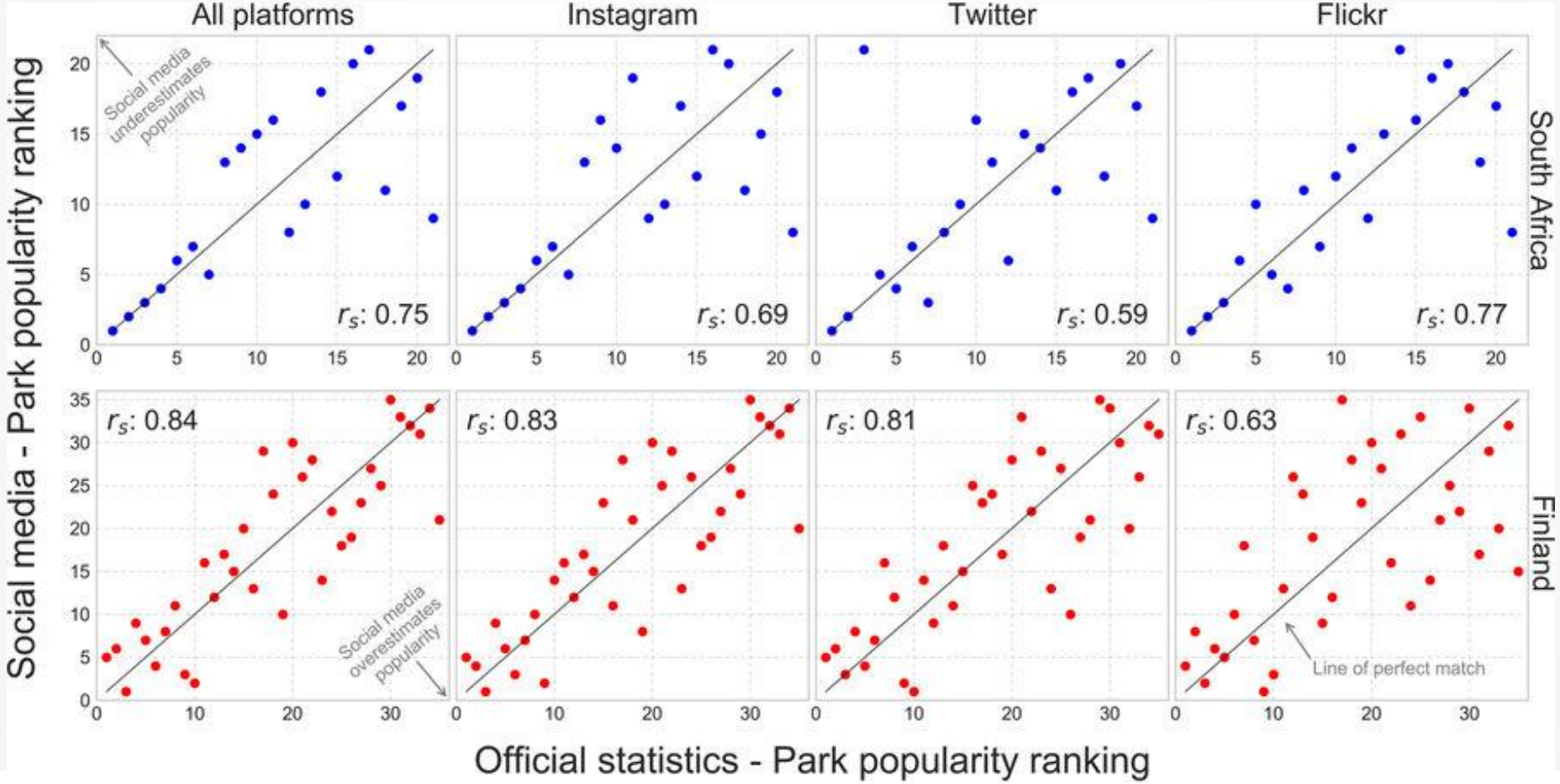
Bergroth 2019, https://helda.helsinki.fi/handle/10138/302228



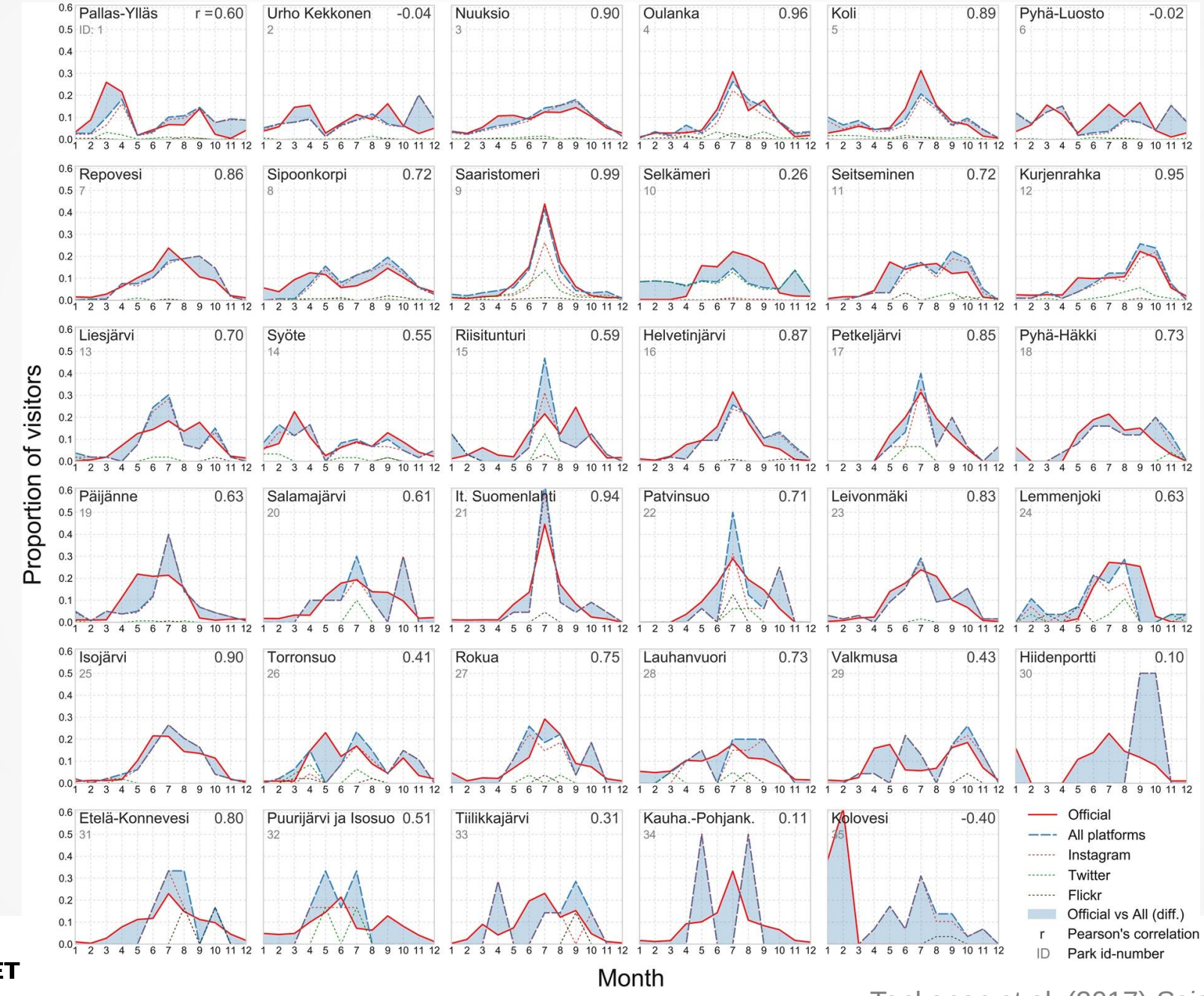
Urban pulse in different areas

Bergroth et al. 2021 (in press)

NATIONAL PARK VISITS VS. SOCIAL MEDIA USERS, FINLAND & SOUTH AFRICA



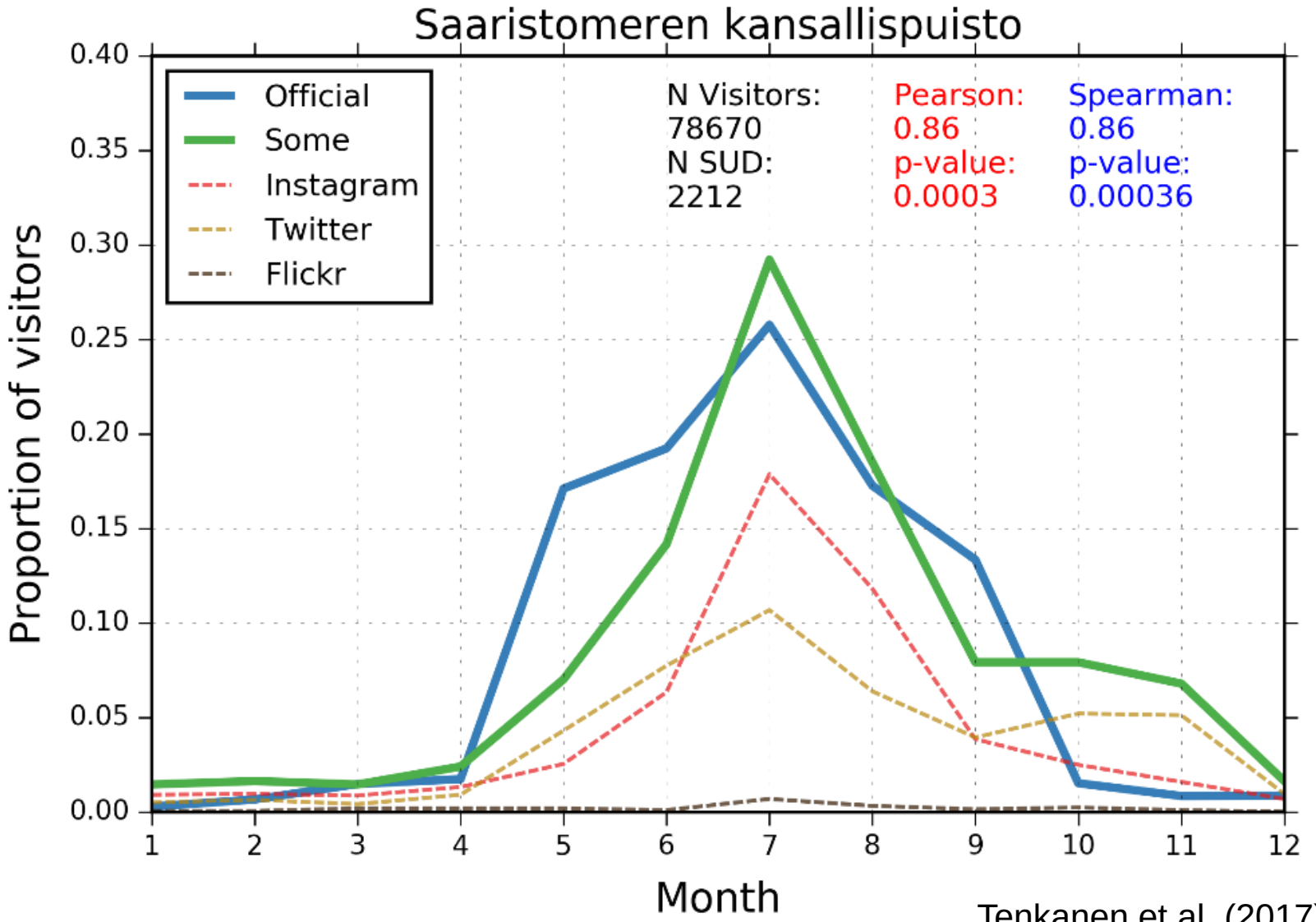
Tenkanen et al. (2017) Scientific Reports 7, Article number: 17615



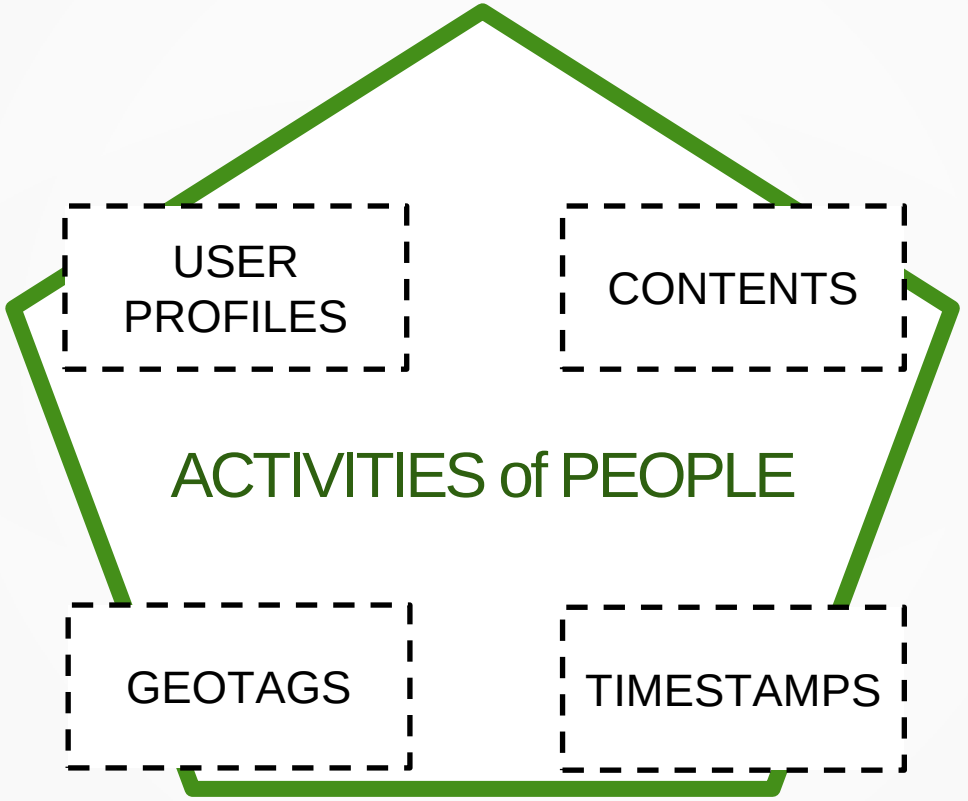
HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tenkanen et al. (2017) *Scientific reports*

FINNISH NATIONAL PARK VISITS VS. SOCIAL MEDIA USERS



Tenkanen et al. (2017) *Scientific reports*



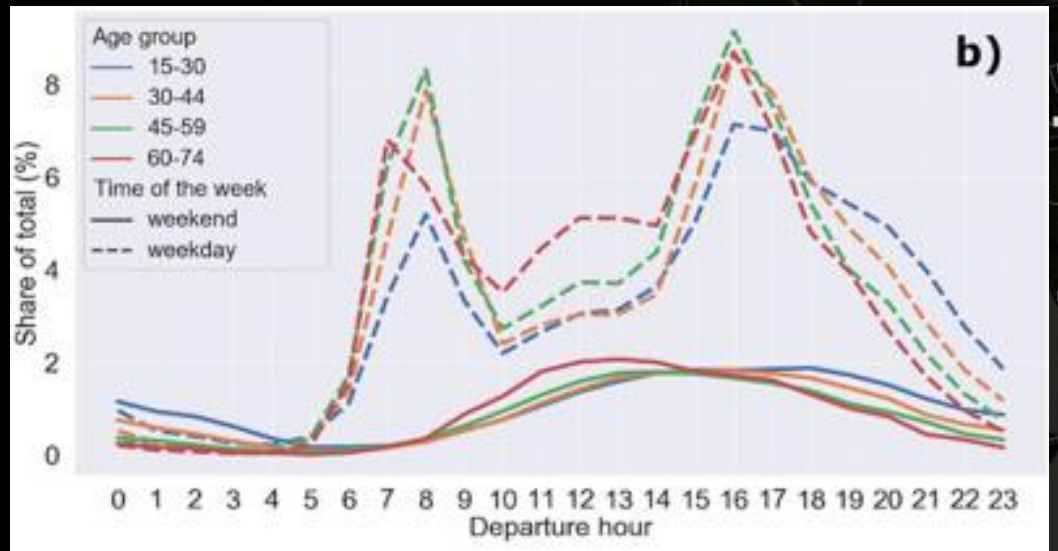
WHERE? **WHEN?**

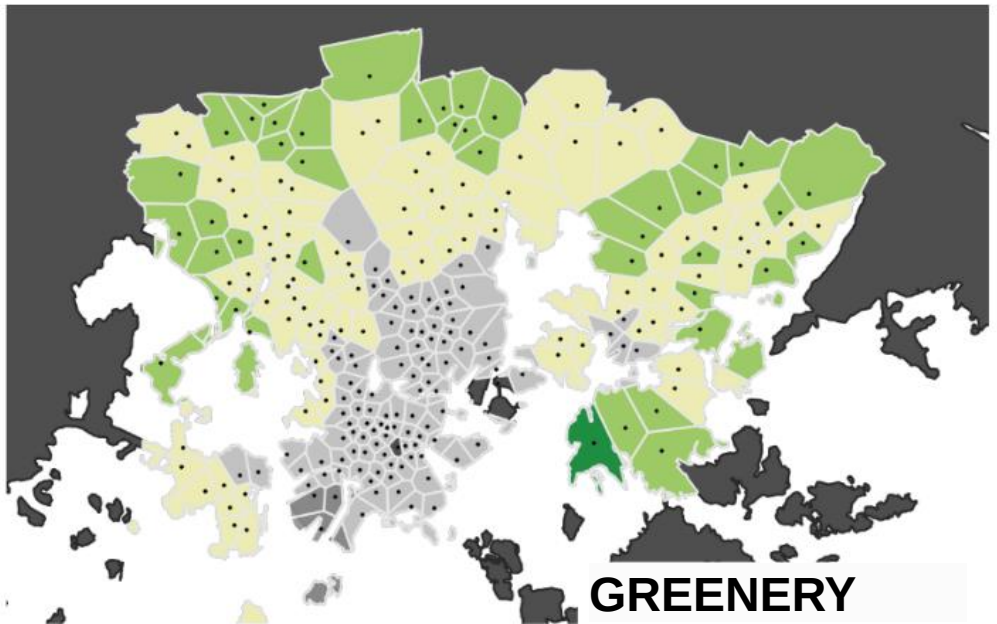
MOBILITY

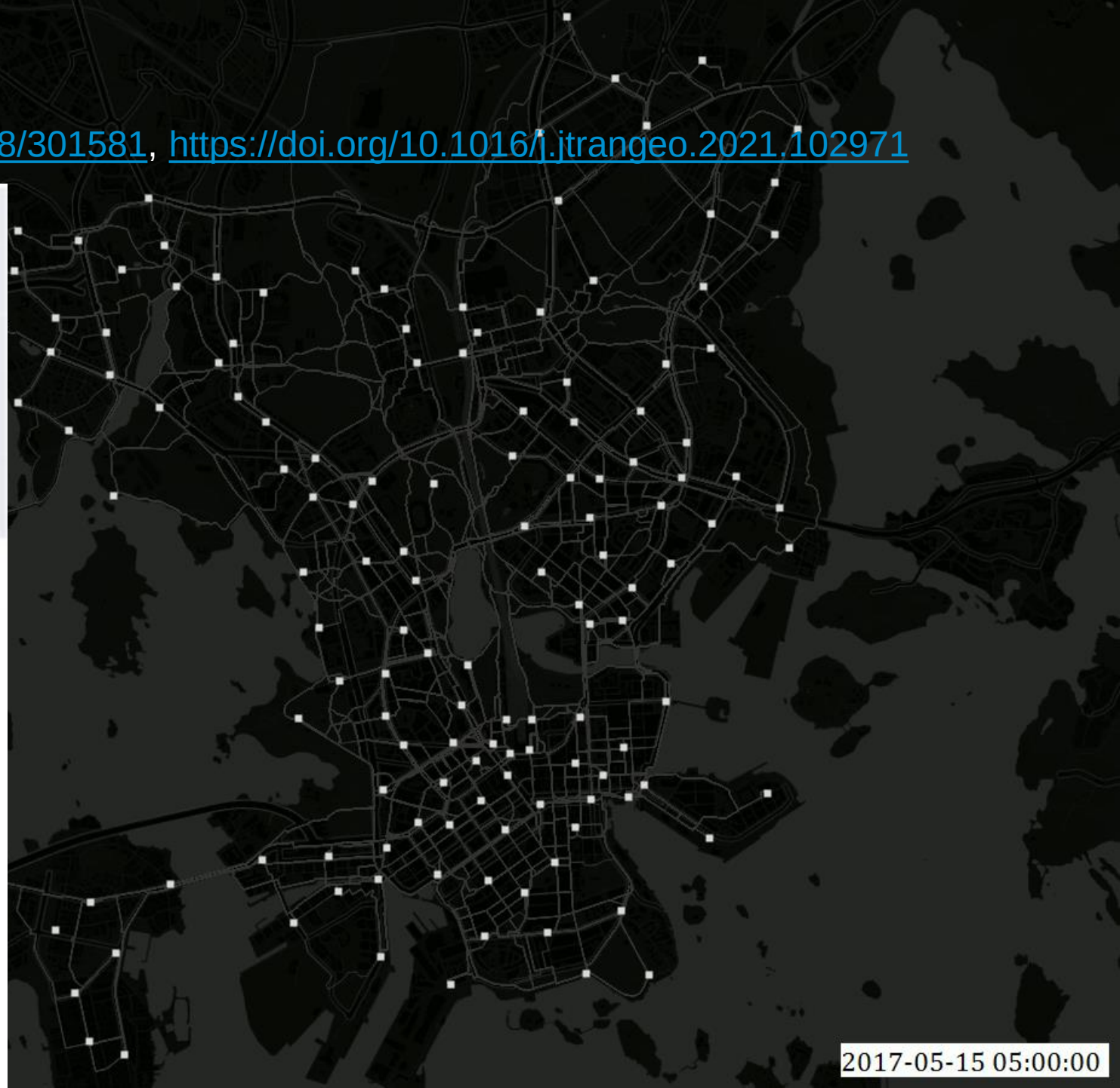
Bikesharing in Helsinki

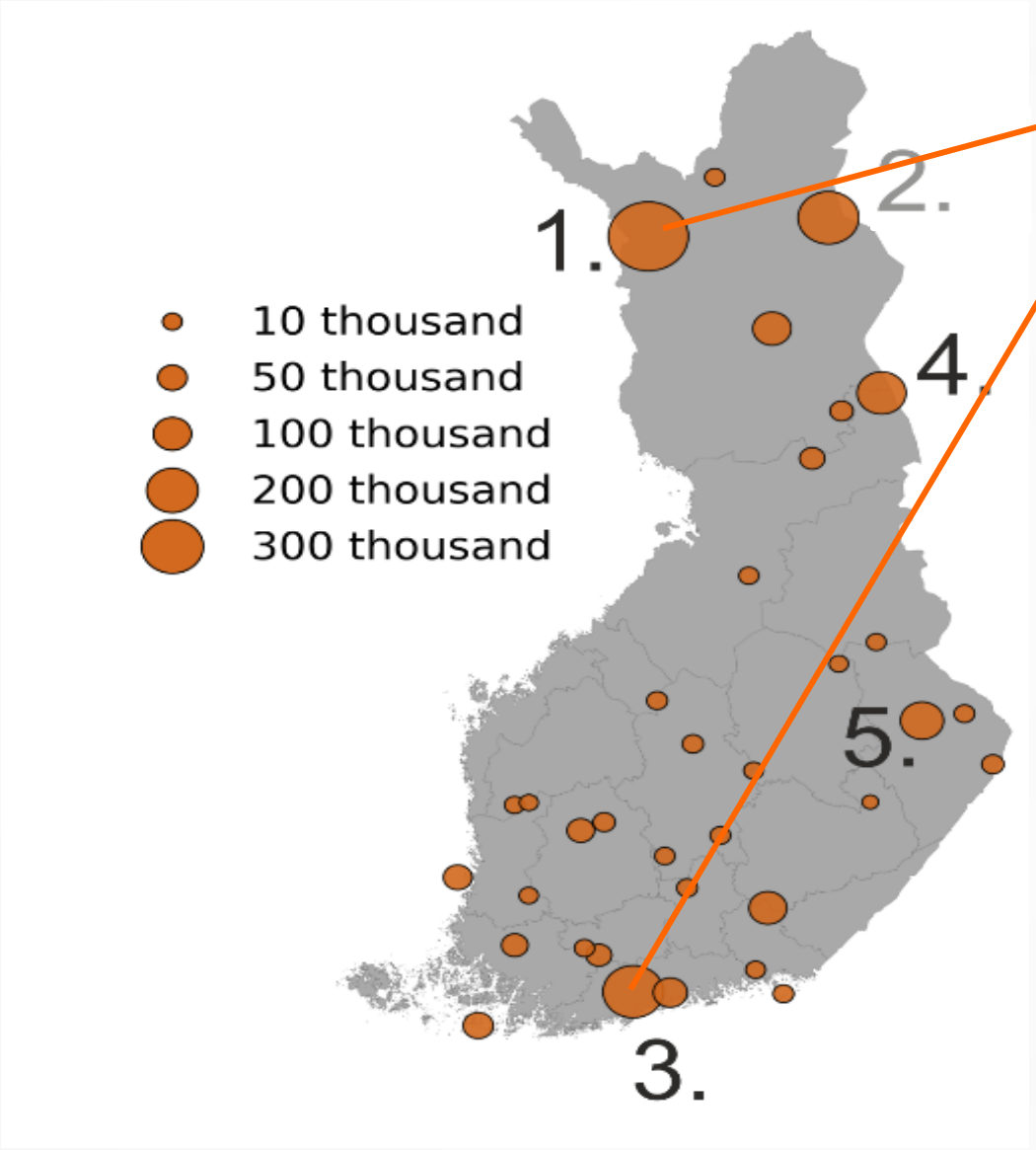
Elias Willberg 2019, Willberg et al. 2021

Read more: https://helda.helsinki.fi/handle/10138/301581, https://doi.org/10.1016/j.jtrangeo.2021.102971

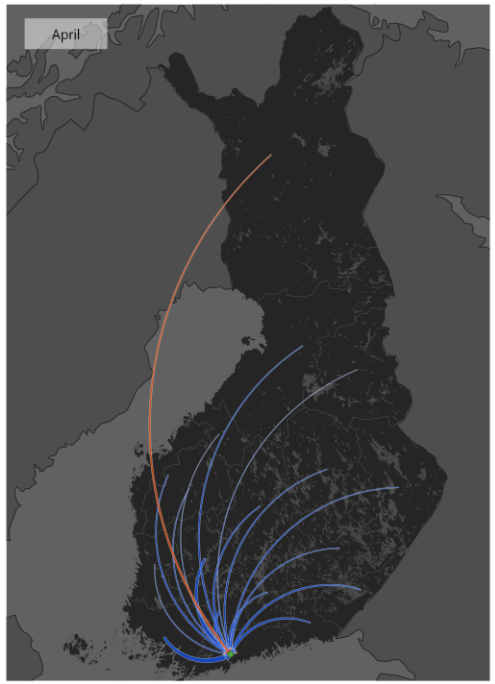


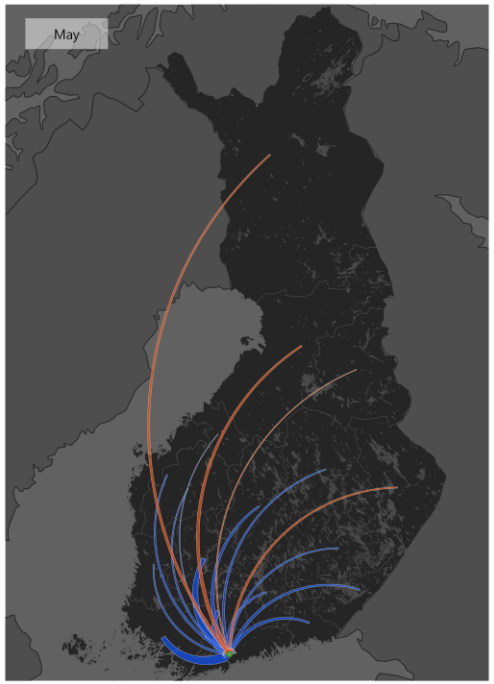


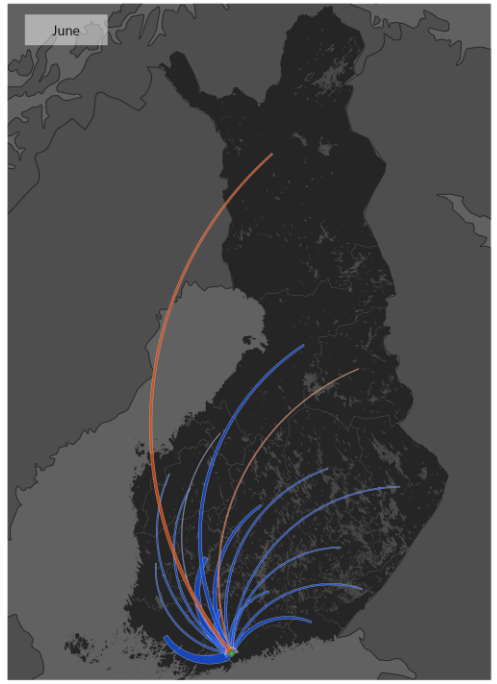




1. Pallas-Ylläs
3. Nuuksio

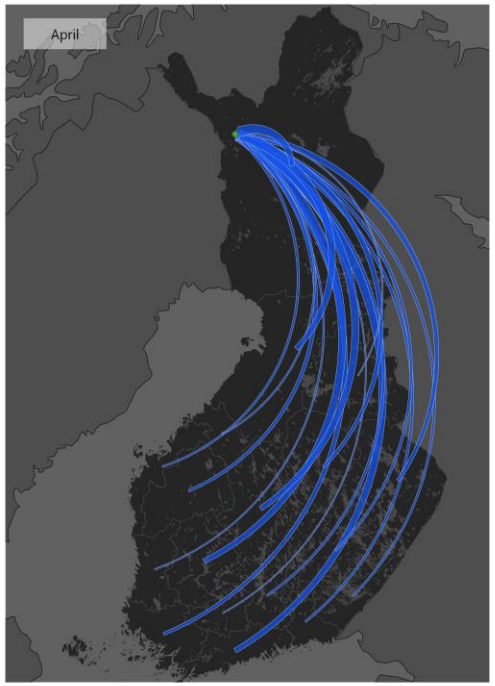


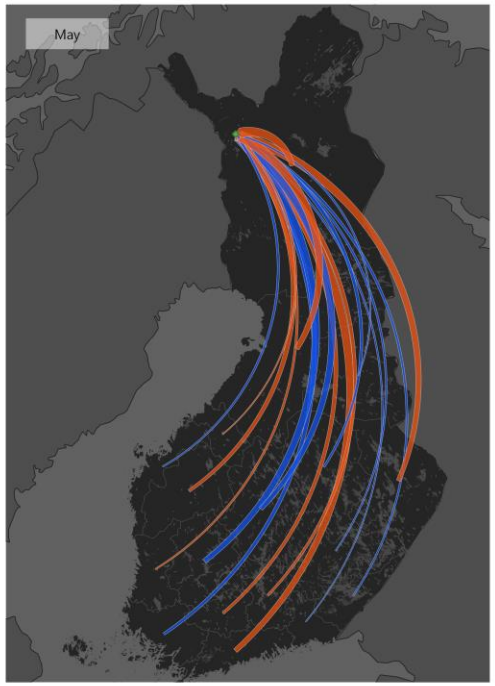


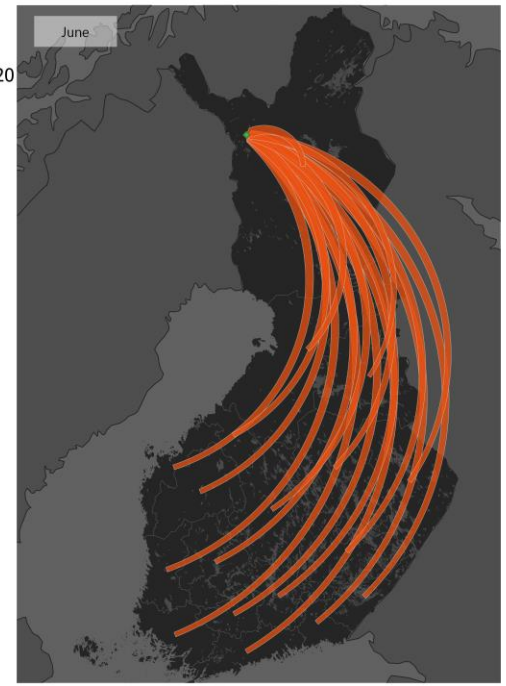


Visitation flows to Nuuksio by province in 4–6/2020

Visitation intensity (thickness)
- Increase from 2019
- Decrease from 2019
- Nuuksio







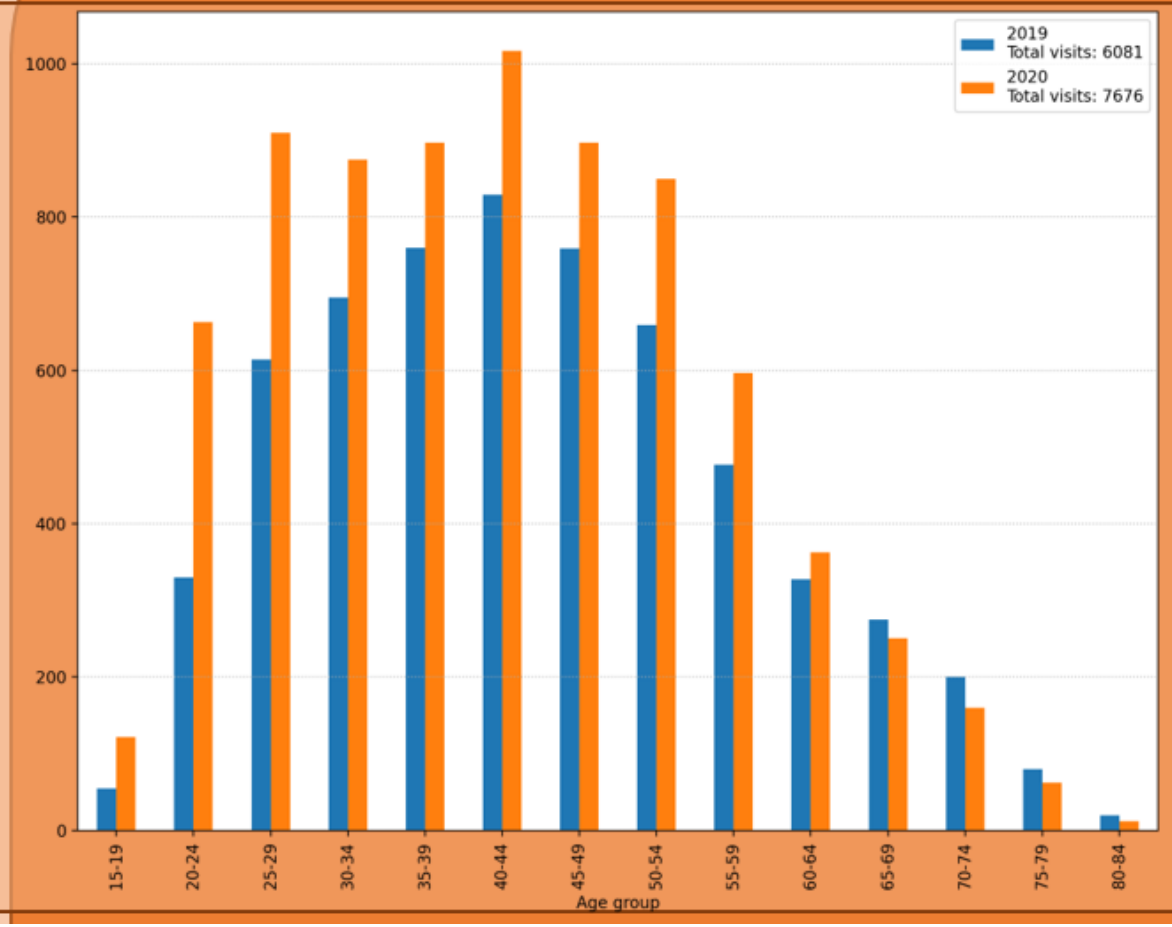
Visitation flows to Pallas-Ylläs by province in 4–6/2020

Visitation intensity (thickness)
- Increase from 2019
- Decrease from 2019
- Pallas-Ylläs

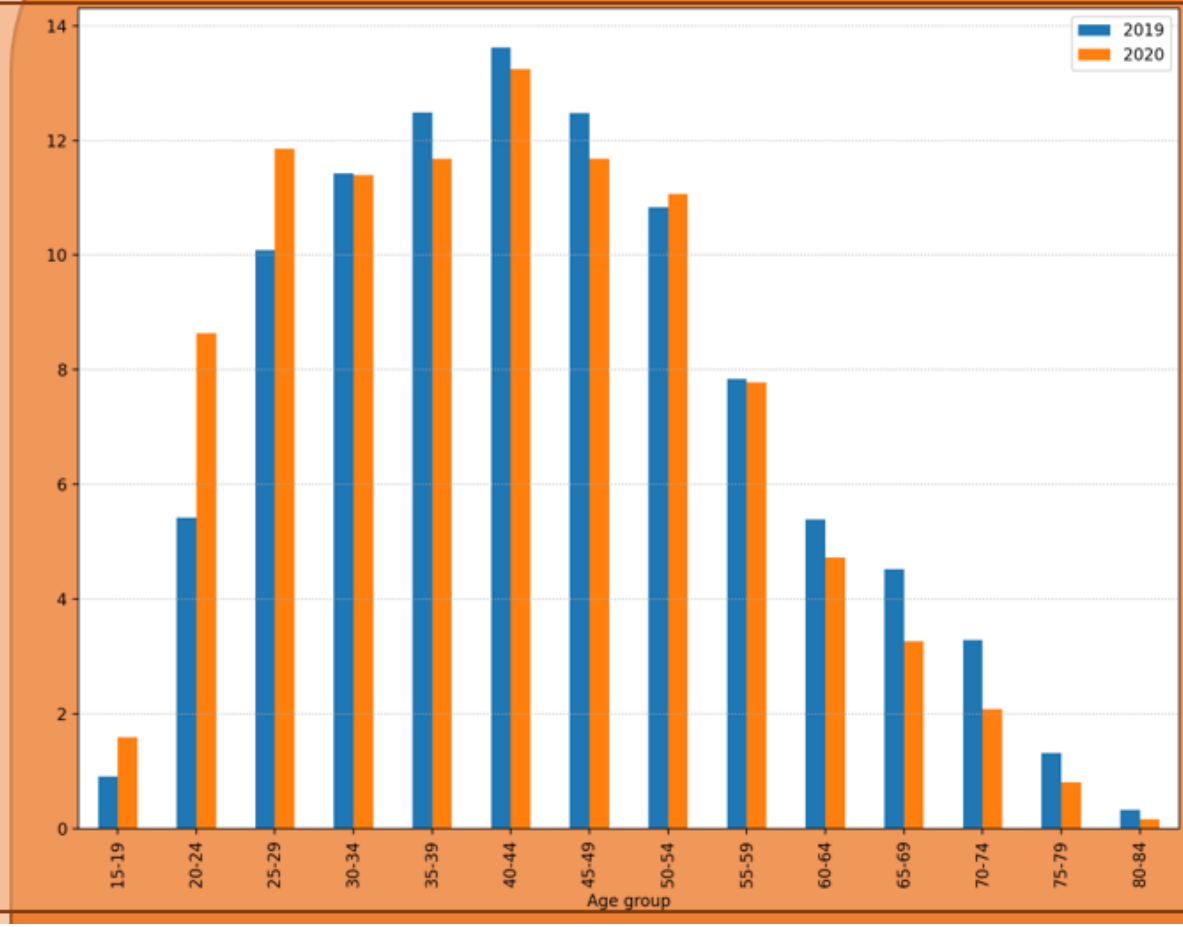
UNDERSTANDING WHO VISITS PARKS

April

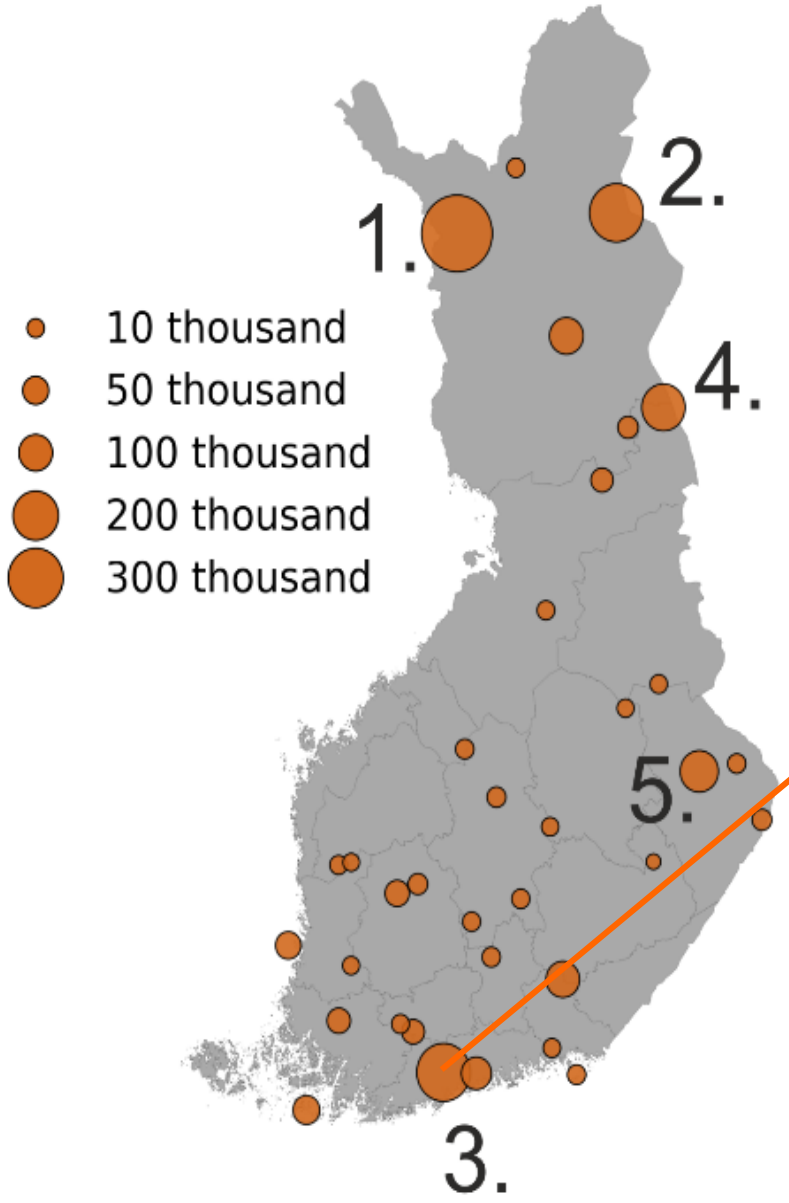
Count

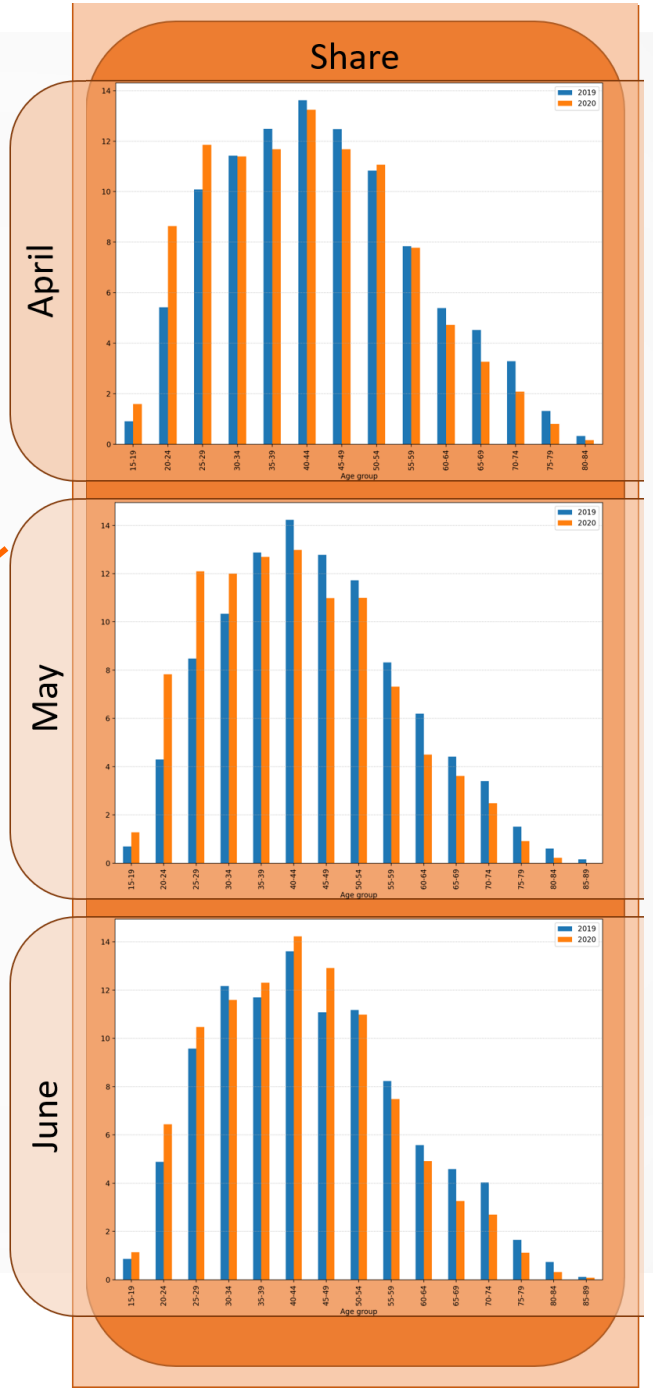


Share

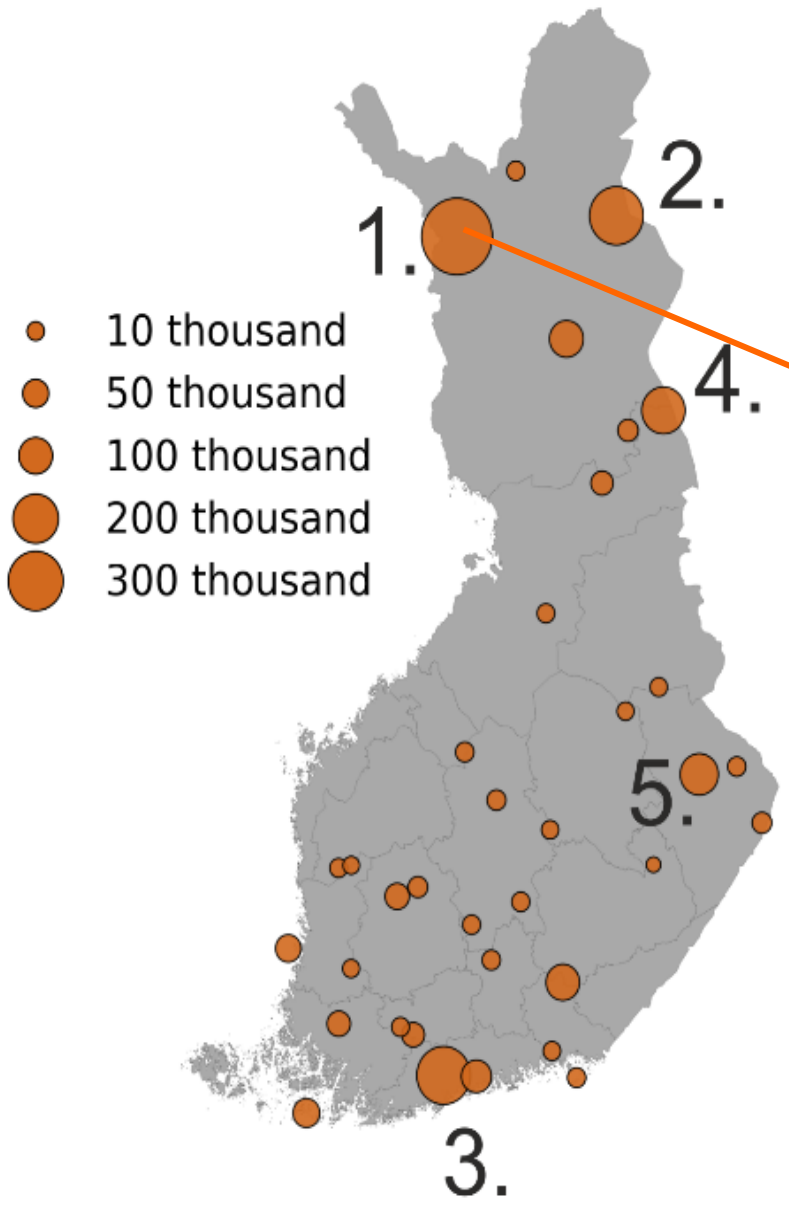


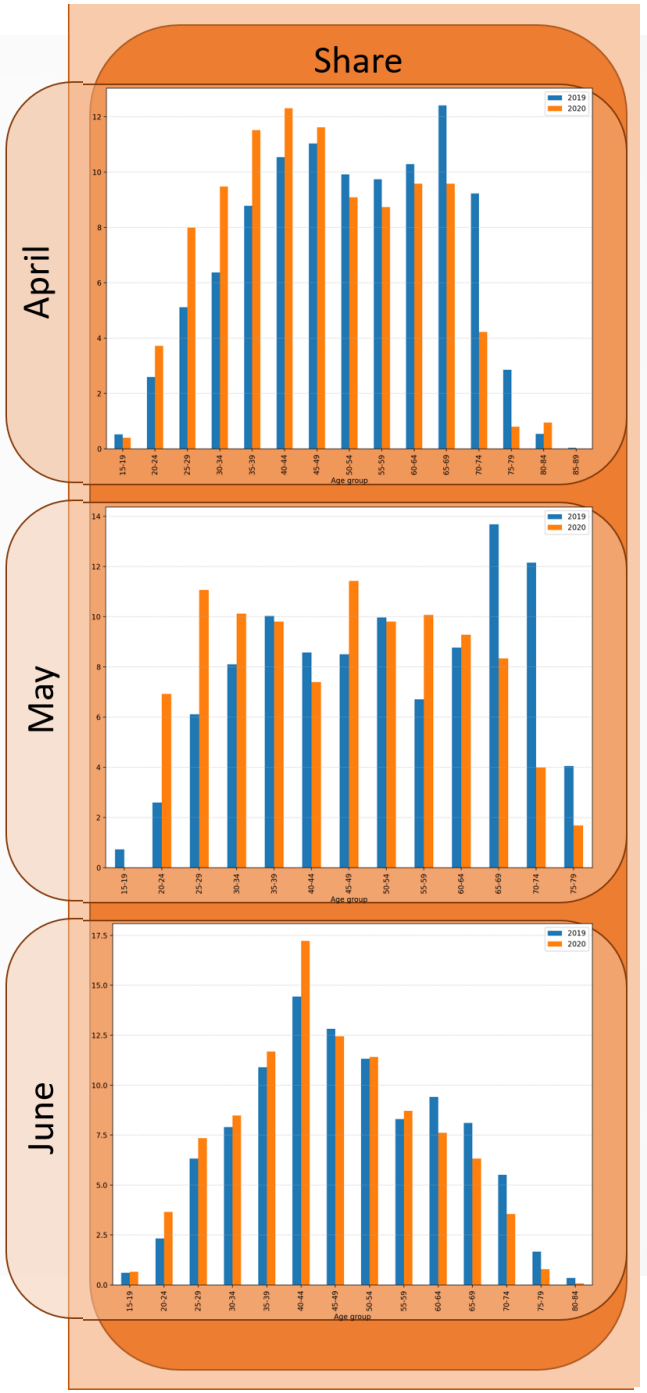
NUUKSIO


10 thousand
50 thousand
100 thousand
200 thousand
300 thousand
1.
2.
3.
4.
5.

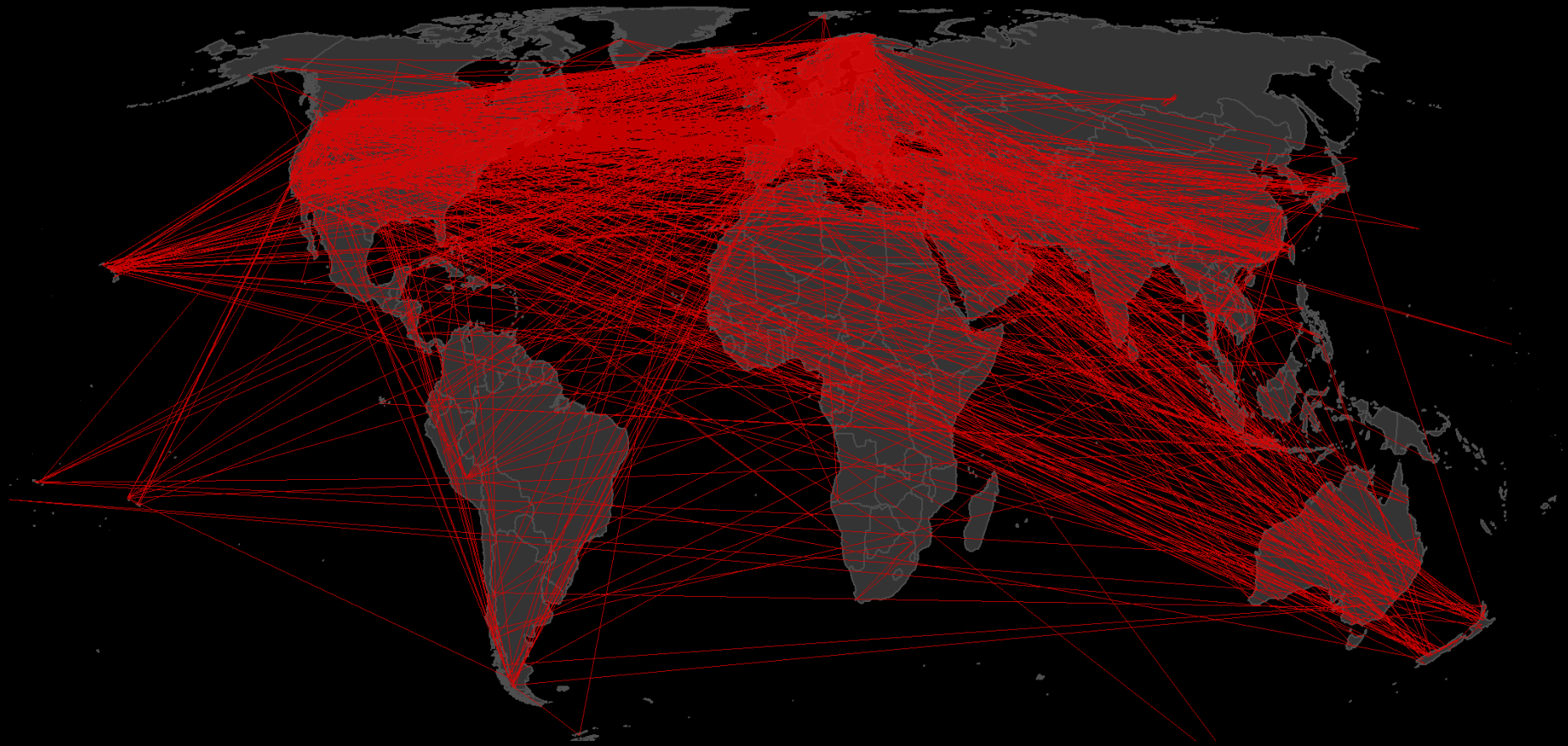

Share
April
May
June
2019
2020
Age group

PALLAS-YLLÄS

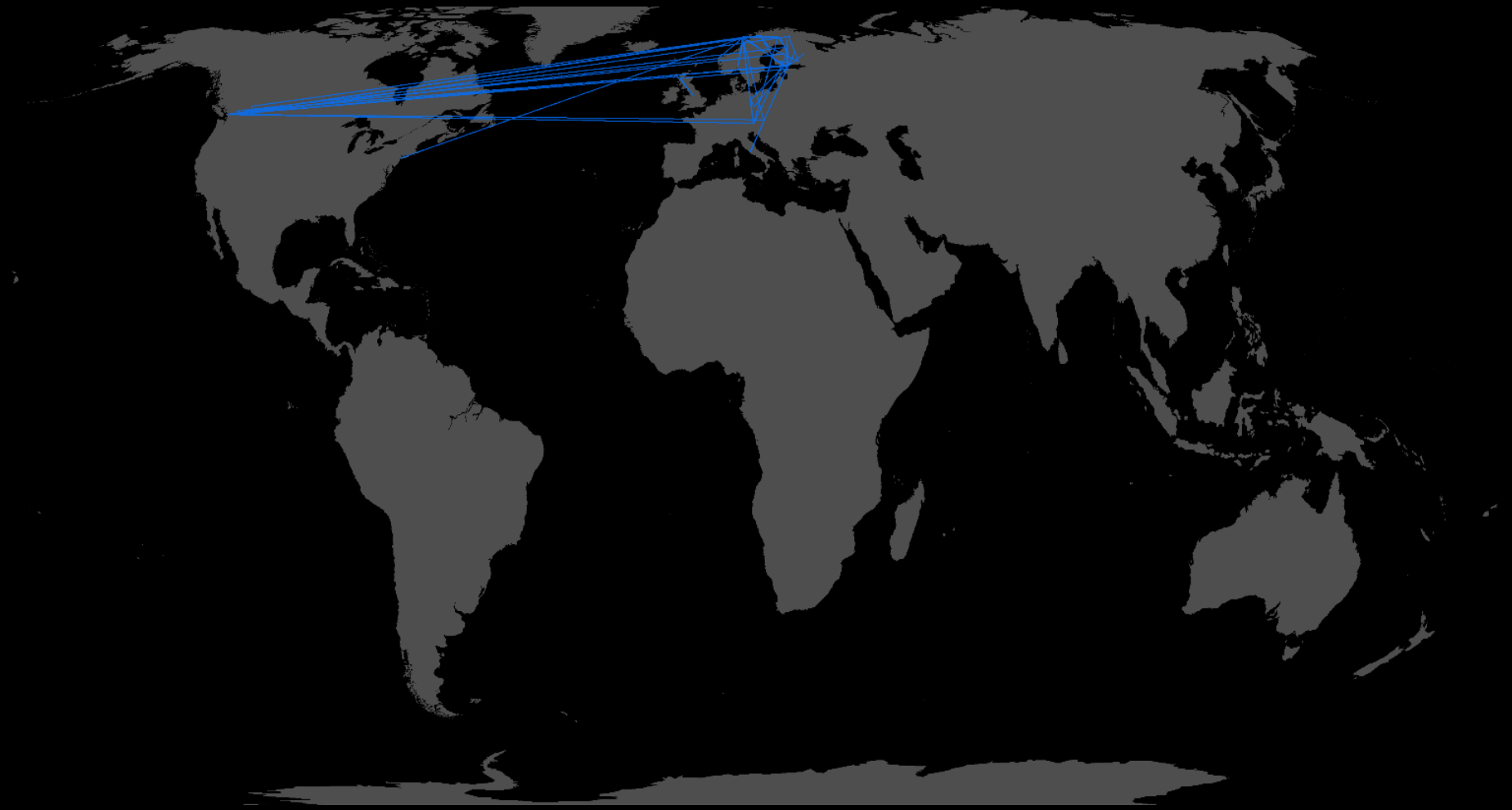




PALLAS-YLLÄS NATIONAL PARK



PALLAS-YLLÄS NATIONAL PARK



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

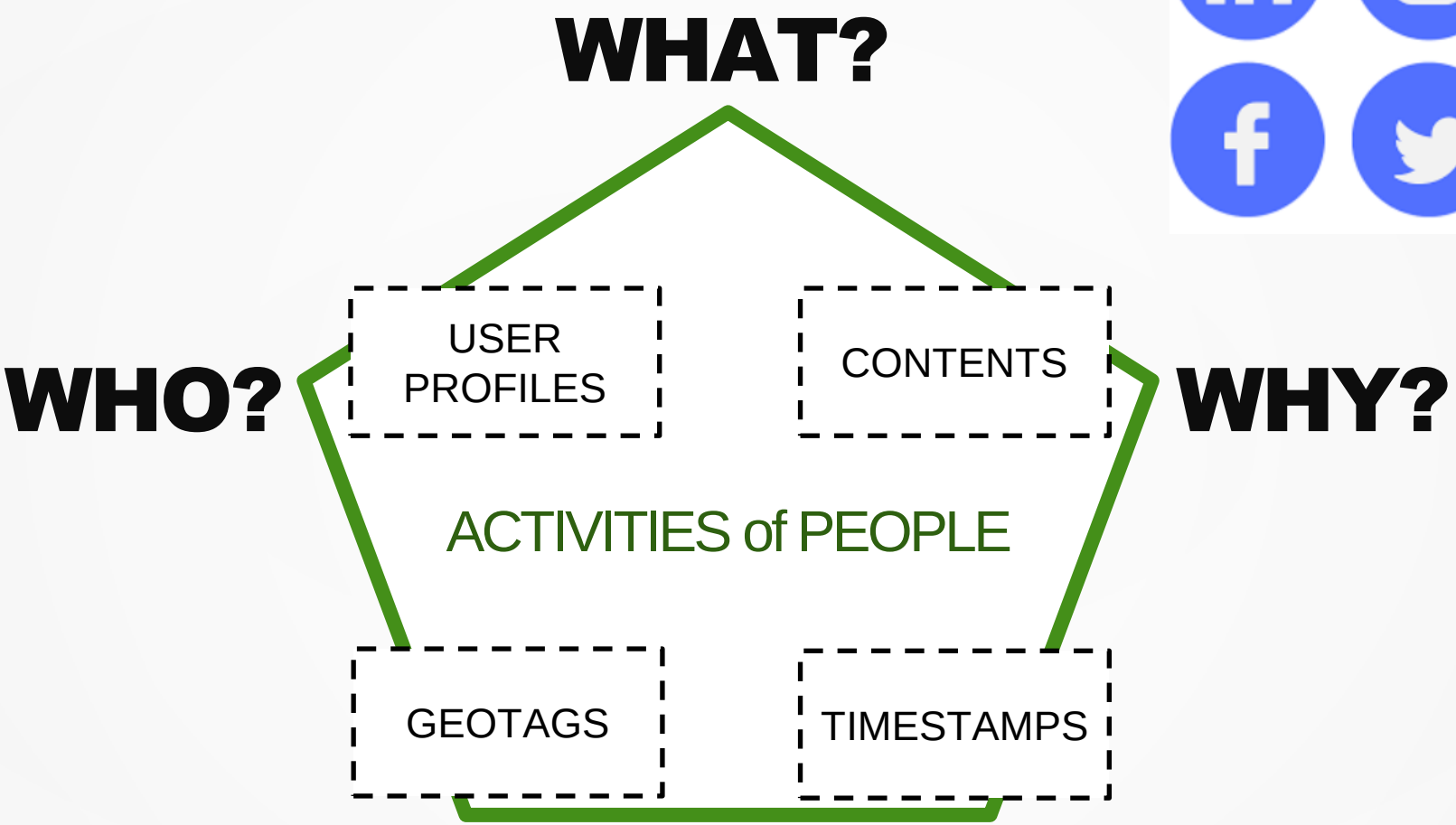
PALLAS-YLLÄS NATIONAL PARK

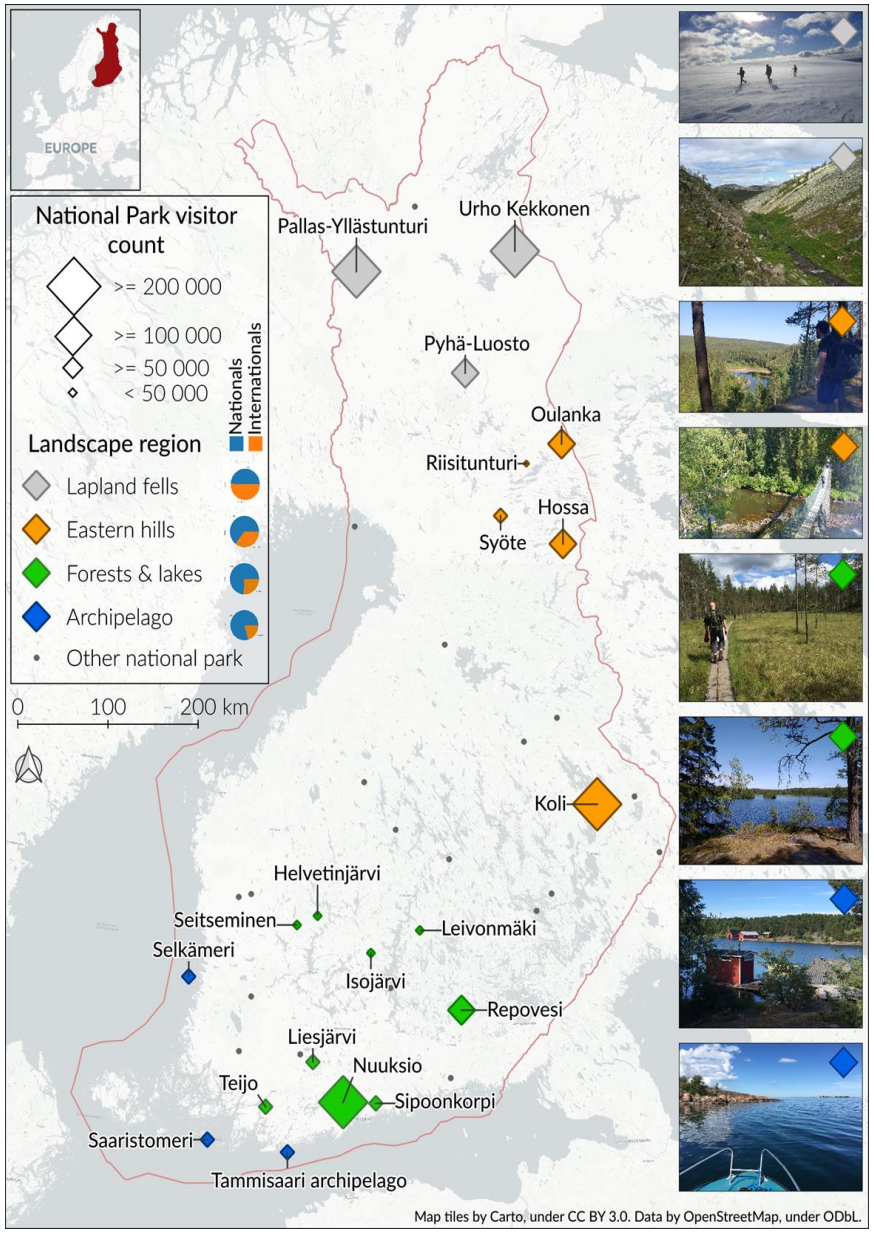
Rogers Pass Summit, Canada

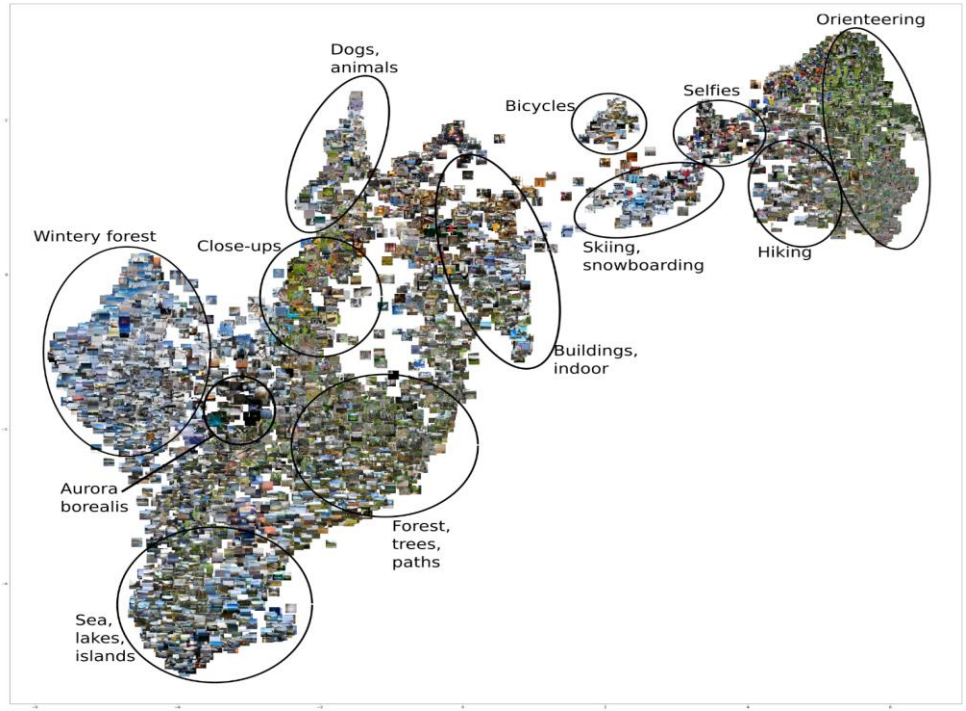
Schladming-Dachstein, Austria

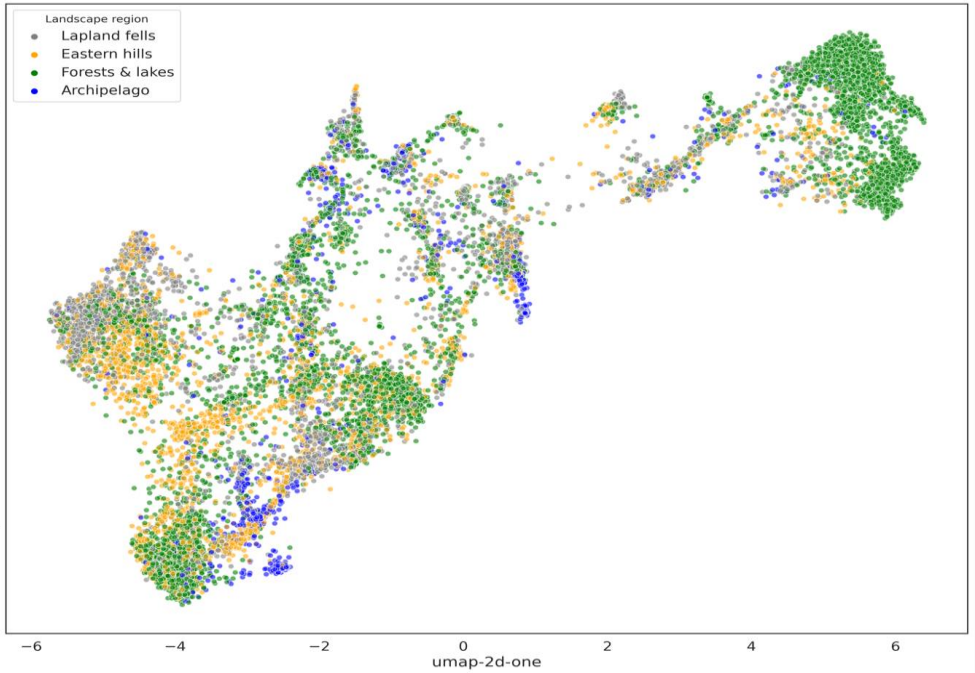
Pallas-Yllästunturi National Park, Finland

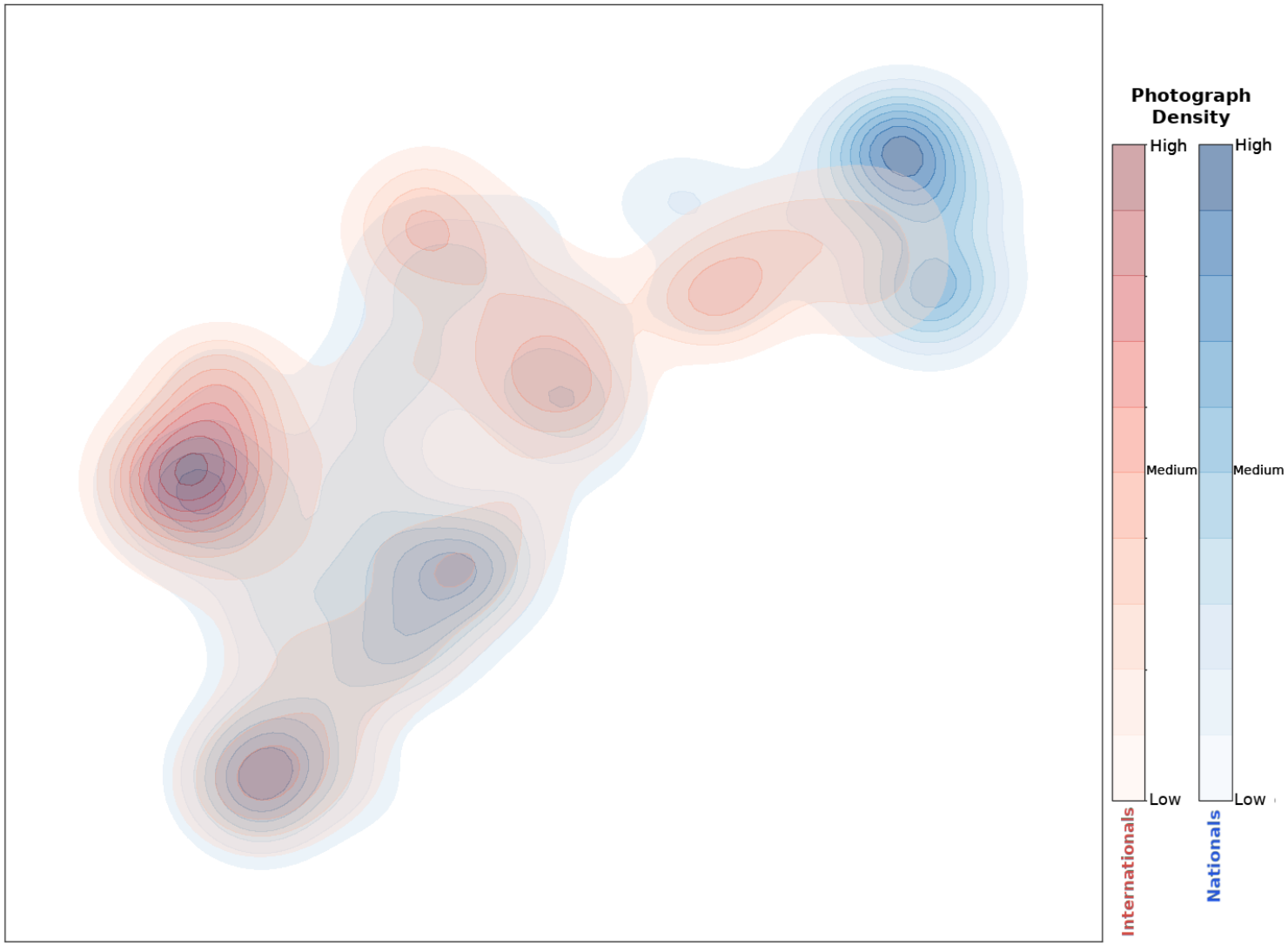
Posts from a Finnish Instagram-user. 107 posts in total.


WHAT?
WHO?
WHY?
USER PROFILES
CONTENTS
ACTIVITIES of PEOPLE
GEOTAGS
TIMESTAMPS

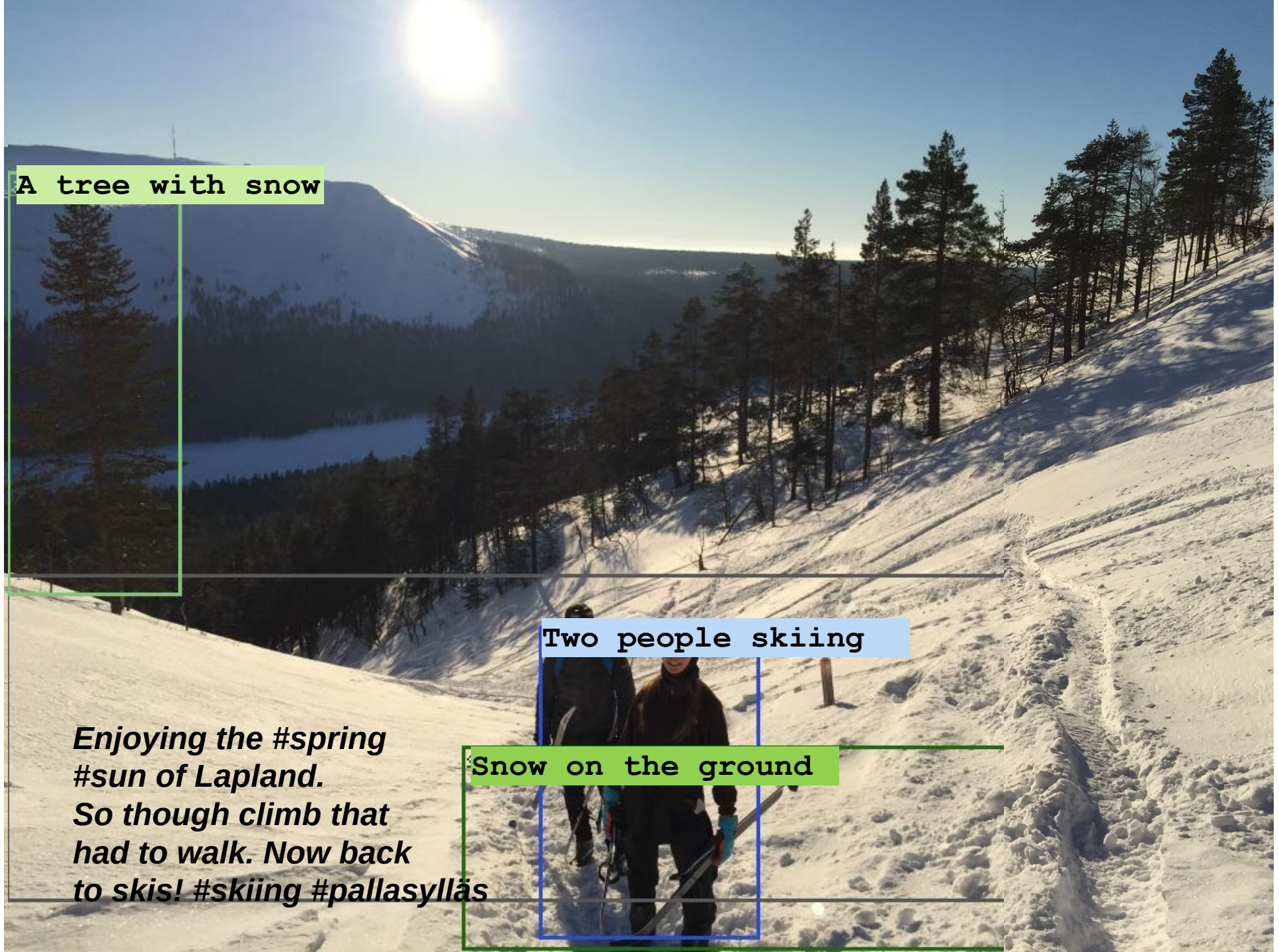

EUROPE
National Park visitor count
>= 200 000
>= 100 000
>= 50 000
< 50 000
Nationals
Internationals
Landscape region
Lapland fells
Eastern hills
Forests & lakes
Archipelago
Other national park
0
100
200 km
Pallas-Yllästunturi
Urho Kekkonen
Pyhä-Luosto
Oulanka
Riisitunturi
Hossa
Syöte
Koli
Helvetinjärvi
Seitseminen
Leivonmäki
Selkämeri
Isojärvi
Repovesi
Liesjärvi
Nuuksio
Teijo
Sipoonkorpi
Saaristomeri
Tammisaari archipelago
Map tiles by Carto, under CC BY 3.0. Data by OpenStreetMap, under ODbL.

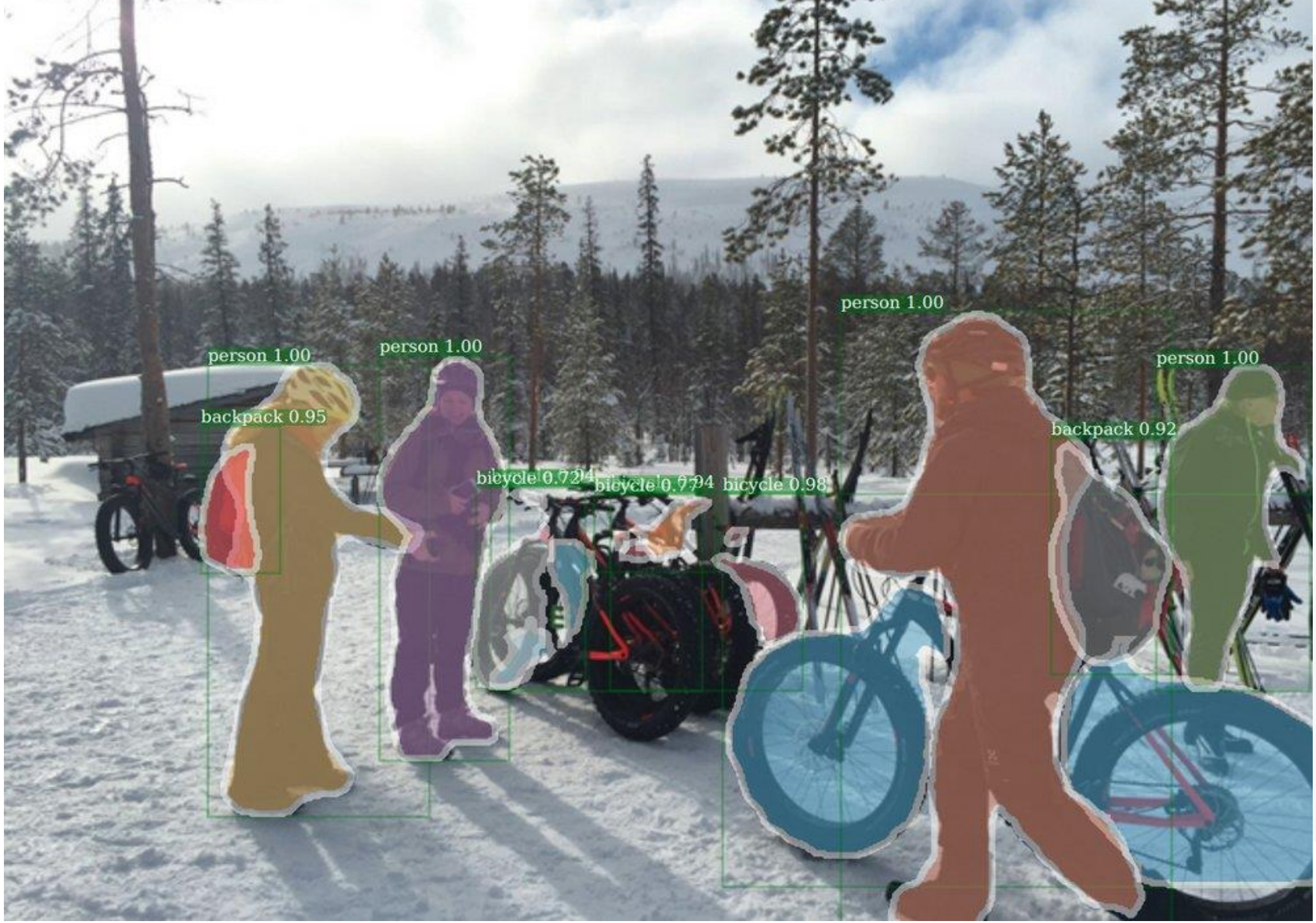

Orienteering
Dogs, animals
Bicycles
Selfies
Wintery forest
Close-ups
Skiing, snowboarding
Hiking
Buildings, indoor
Aurora borealis
Forest, trees, paths
Sea, lakes, islands

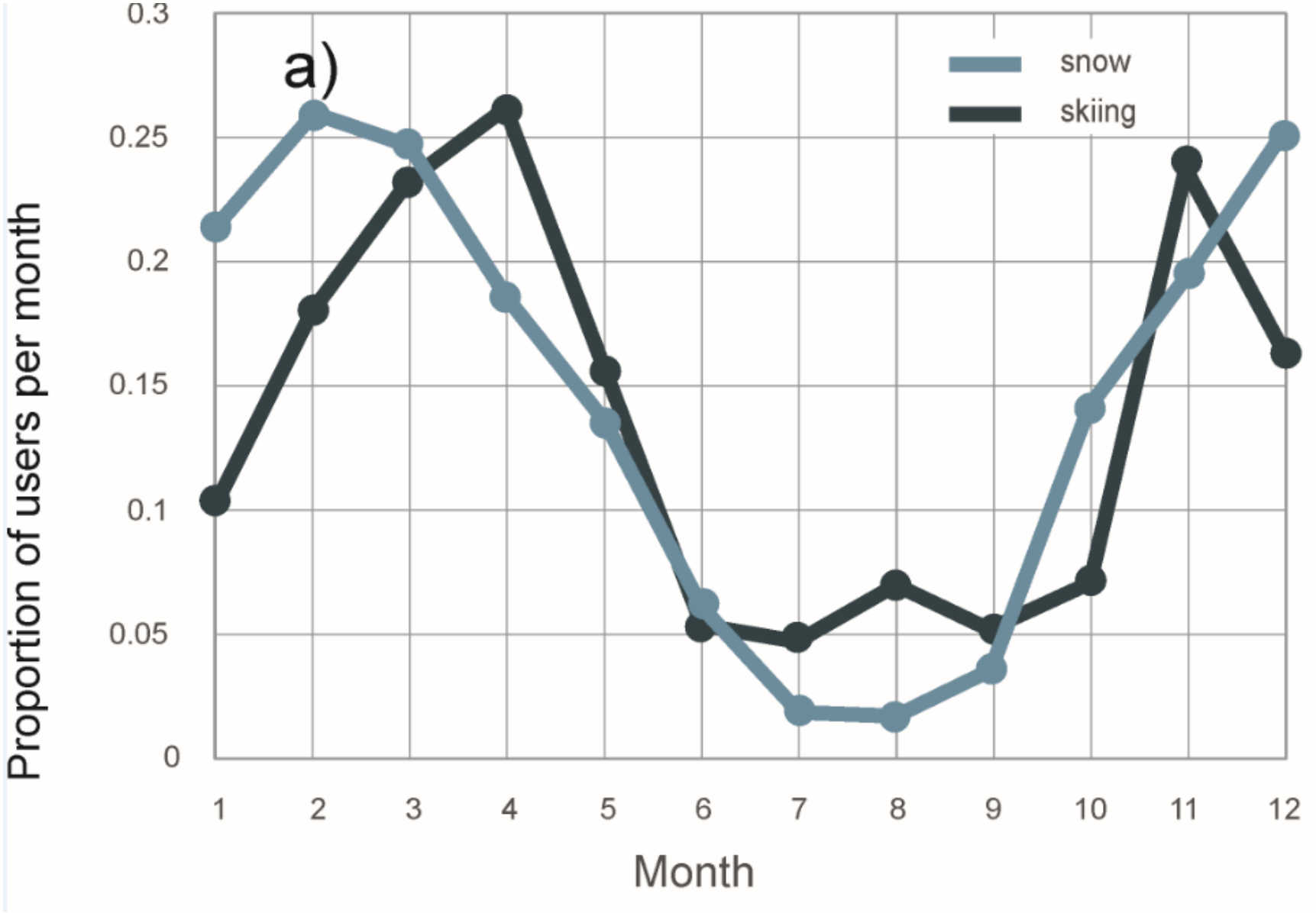

Landscape region
Lapland fells
Eastern hills
Forests & lakes
Archipelago
2
0
−2
−4
−6
−4
−2
0
2
4
6
umap-2d-one


Photograph Density
High
Medium
Low
High
Medium
Low
Internationals
Nationals


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

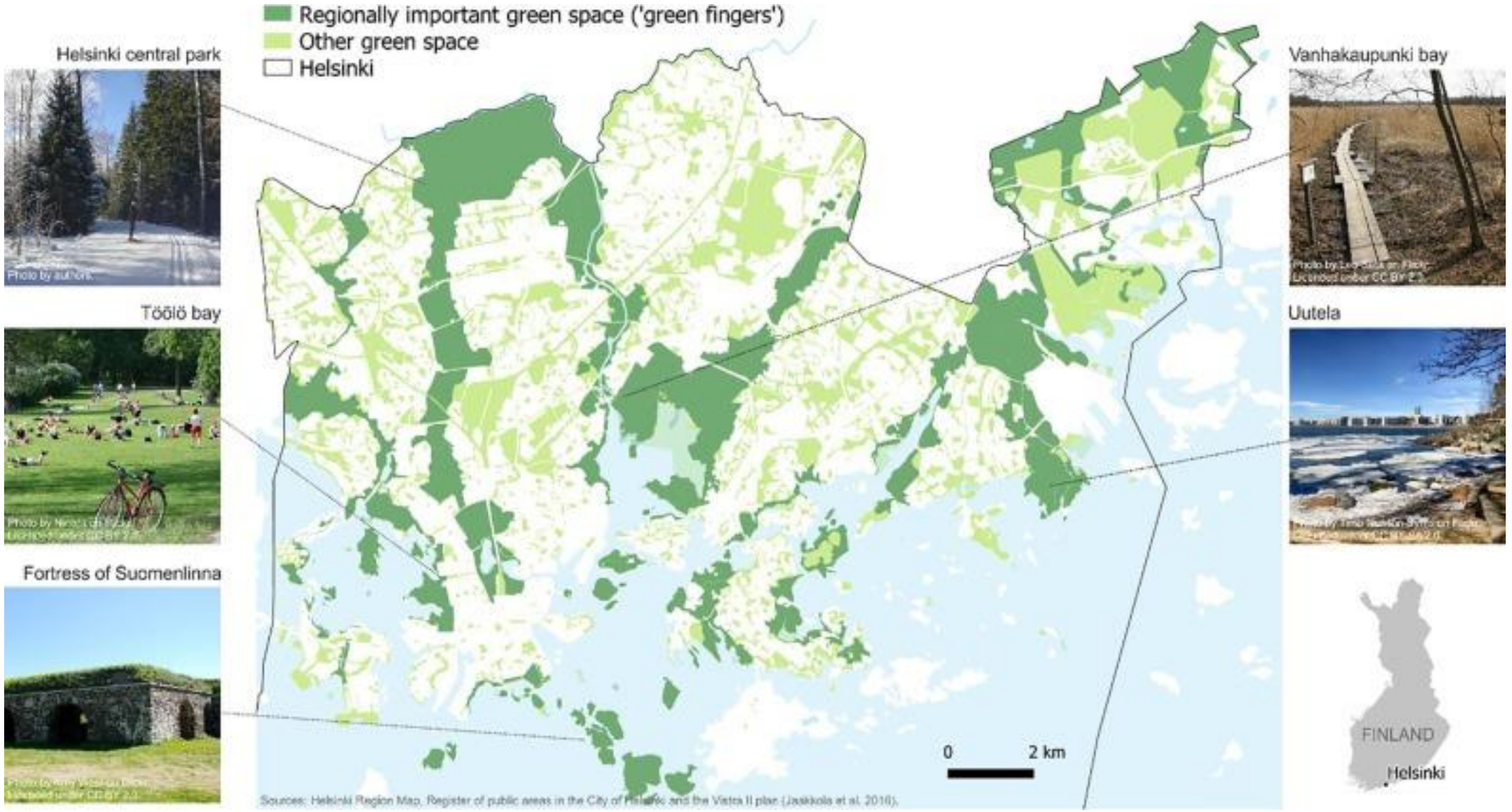

A tree with snow
Two people skiing
Snow on the ground
Enjoying the #spring #sun of Lapland. So though climb that had to walk. Now back to skis! #skiing #pallasylläs

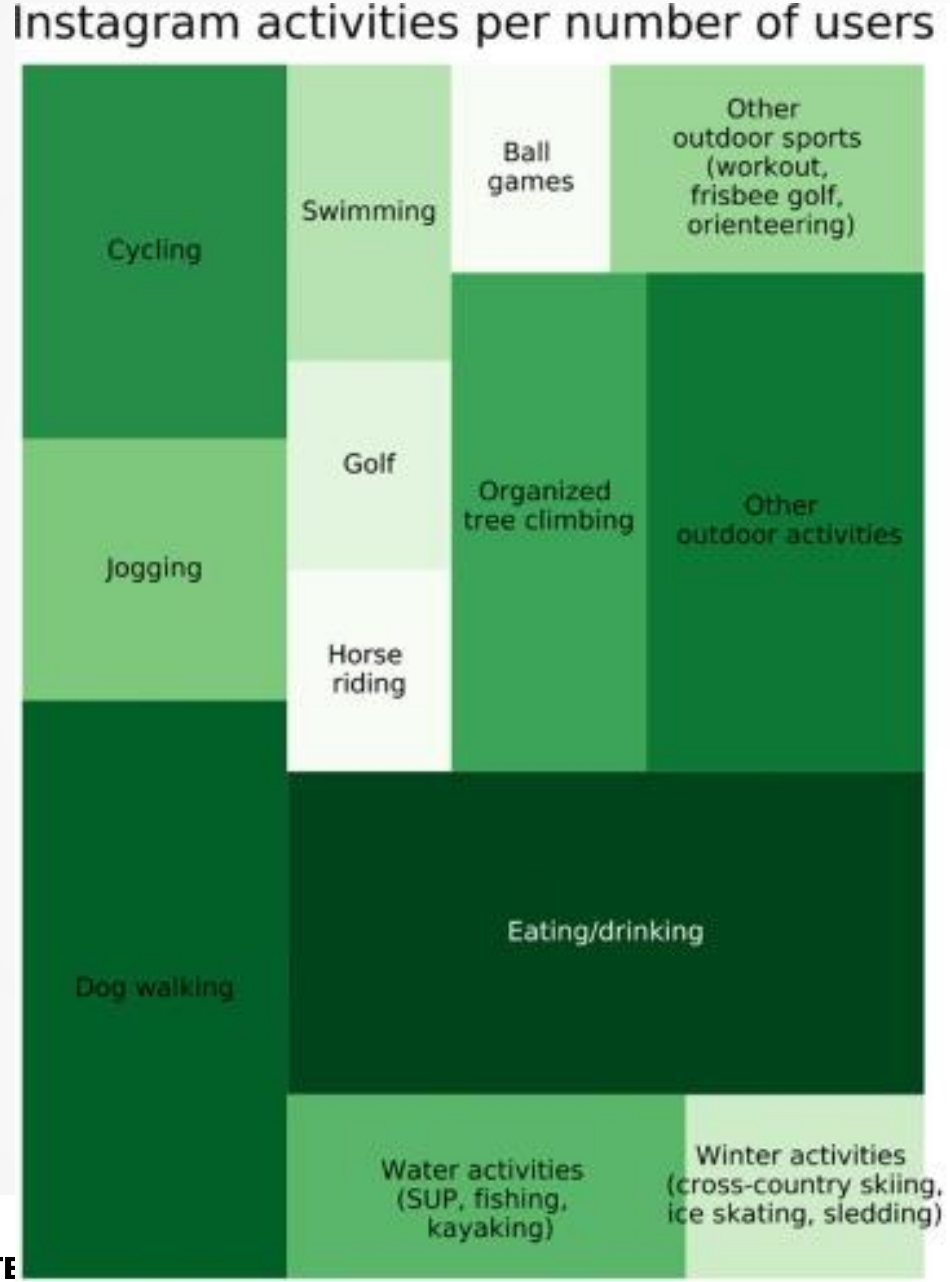

person 1.00
person 1.00
person 1.00
person 1.00
backpack 0.95
backpack 0.92
bicycle 0.72
bicycle 0.77
bicycle 0.98

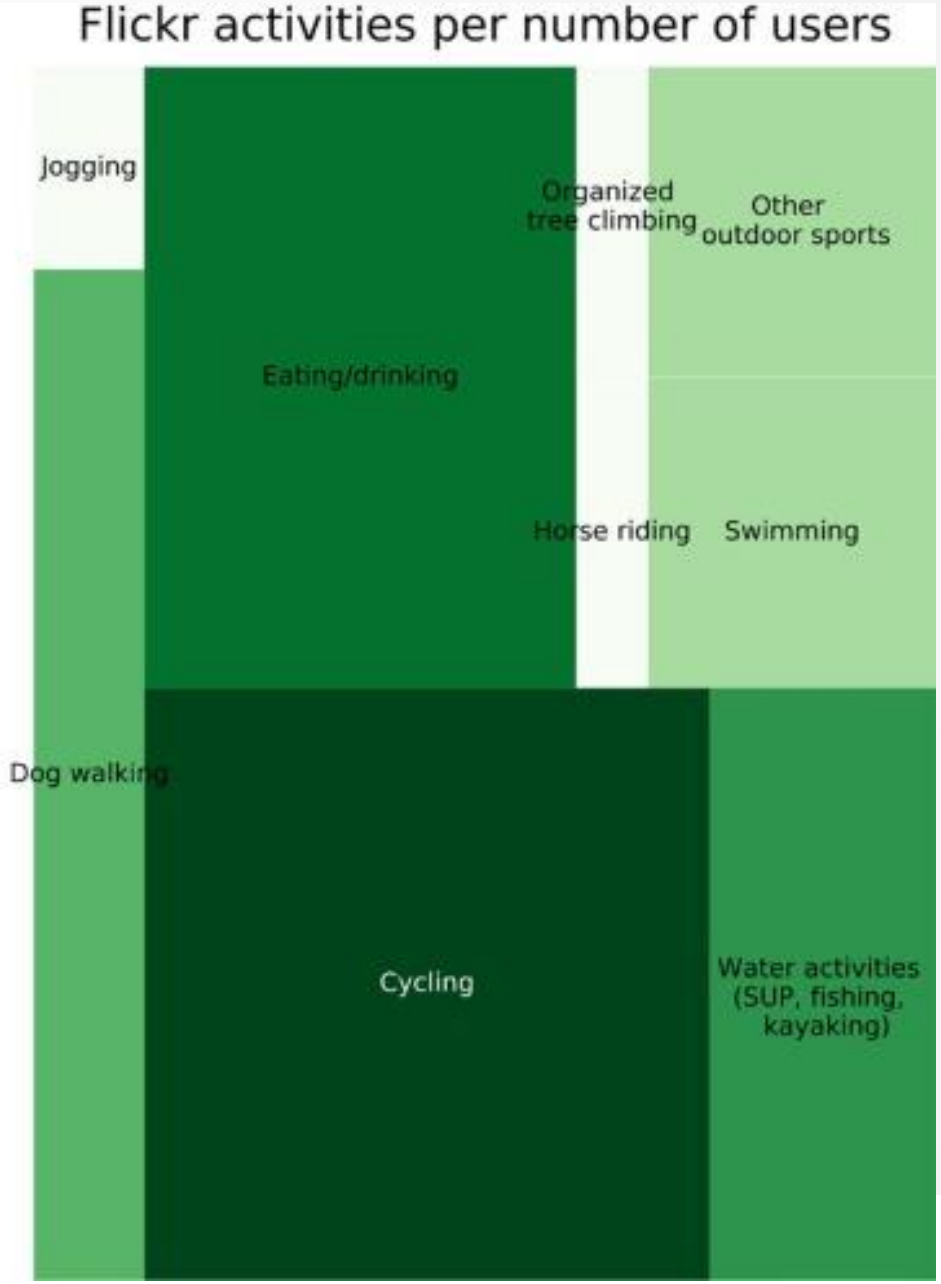


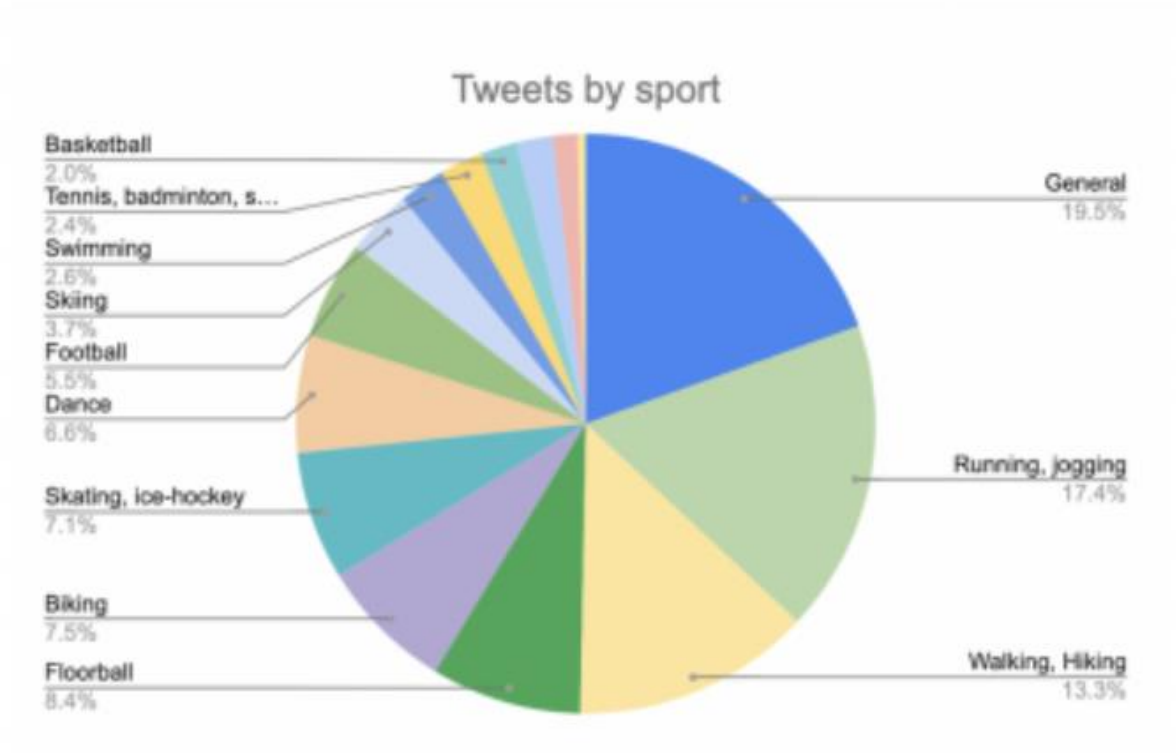
Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., Di Minin, E.
Social media data for conservation science: a methodological overview.
Biological Conservation (2019).

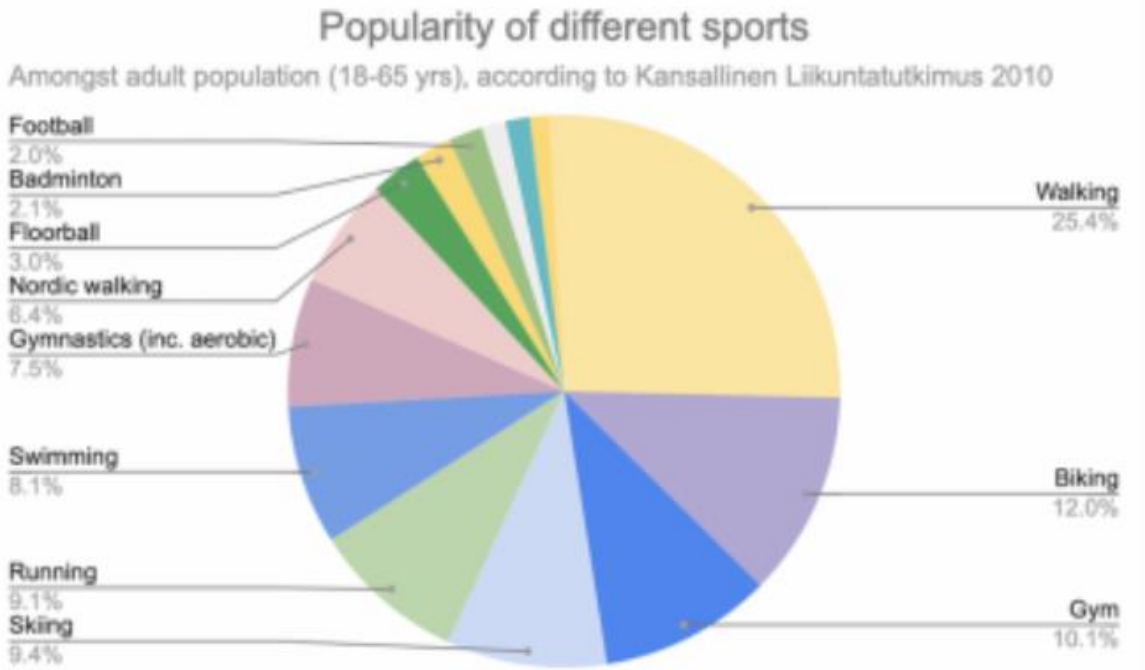
WHAT DO PEOPLE DO IN URBAN PARKS?




Instagram activities per number of users
Cycling
Swimming
Ball games
Other outdoor sports (workout, frisbee golf, orienteering)
Jogging
Golf
Organized tree climbing
Other outdoor activities
Horse riding
Dog walking
Eating/drinking
Water activities (SUP, fishing, kayaking)
Winter activities (cross-country skiing, ice skating, sledding)


Flickr activities per number of users
Jogging
Organized tree climbing
Other outdoor sports
Eating/drinking
Horse riding
Swimming
Dog walking
Cycling
Water activities (SUP, fishing, kayaking)





The different types of sports are distributed differently in the Metropolitan Area. Many sports are clustered to the center of Helsinki and to other residential and sports facilities hotspots.

Koivisto 2021
https://blogs.helsinki.fi/digital-geography/2021/08/25/creating-knowledge-about-exercising-in-the-helsinki-metropolitan-area-using-twitter-data/

Heikinheimo et al. in press.
Heikinheimo et al. (2017) ISPRS

Hausmann et al. (2020) People & Nature
Väisänen et al. (2021) Conservation Biology

WHAT?

WHO?

WHY?

USER PROFILES

CONTENTS

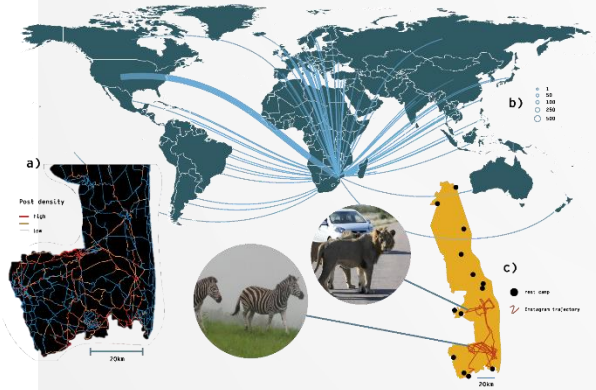
ACTIVITIES of PEOPLE

GEOTAGS

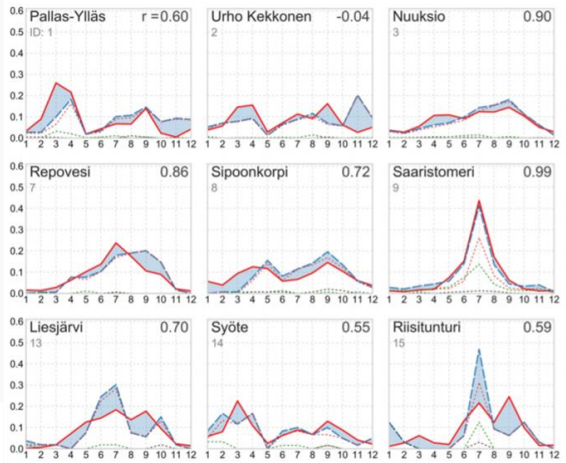
TIMESTAMPS

WHERE?

WHEN?

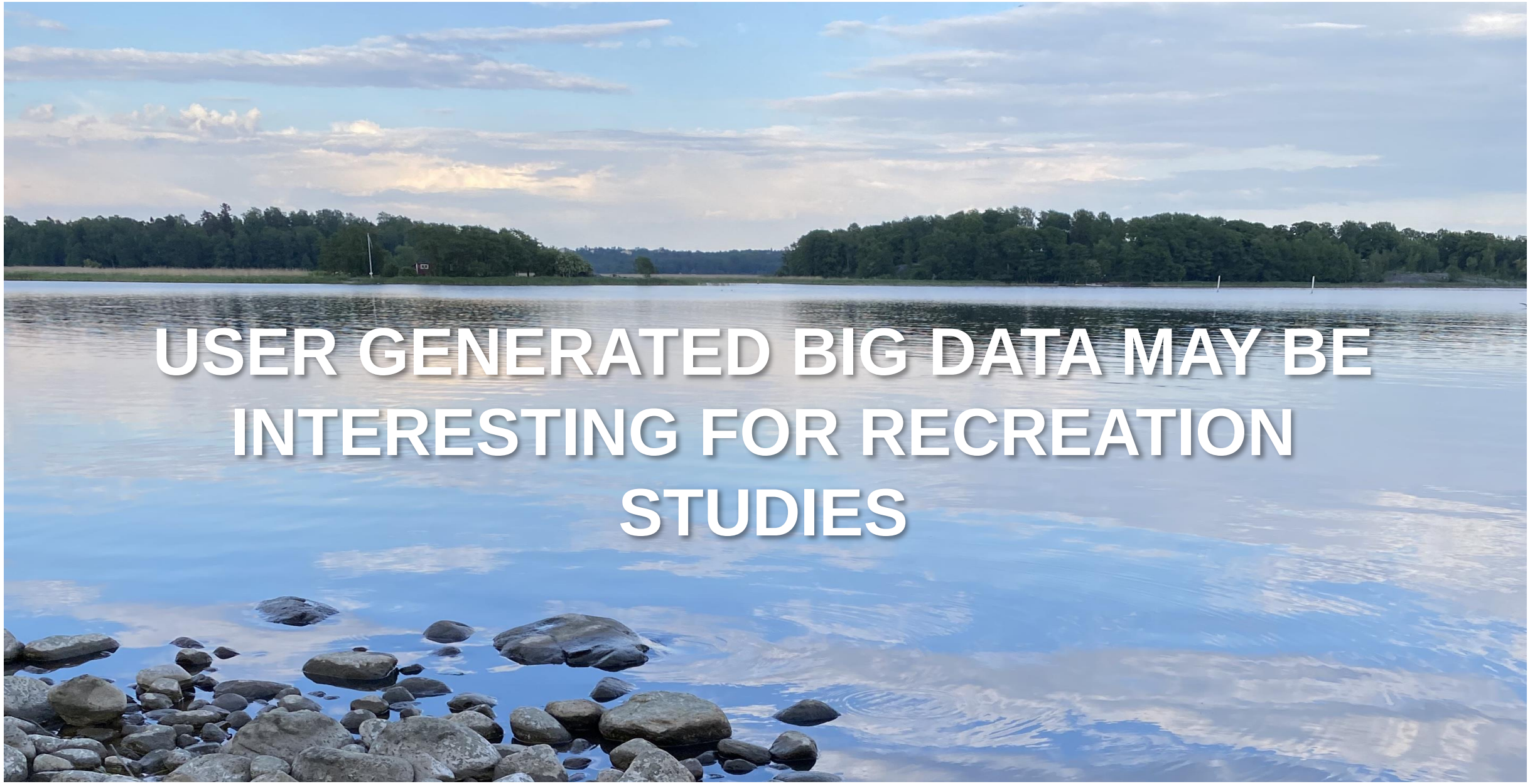


Toivonen et al. (2019) Biological Conservation
Heikinheimo et al. (2020) Landscape & Urban planning



Tenkanen et al. (2017) Scientific Reports

WHAT SHOULD WE LEARN FROM ALL THIS?

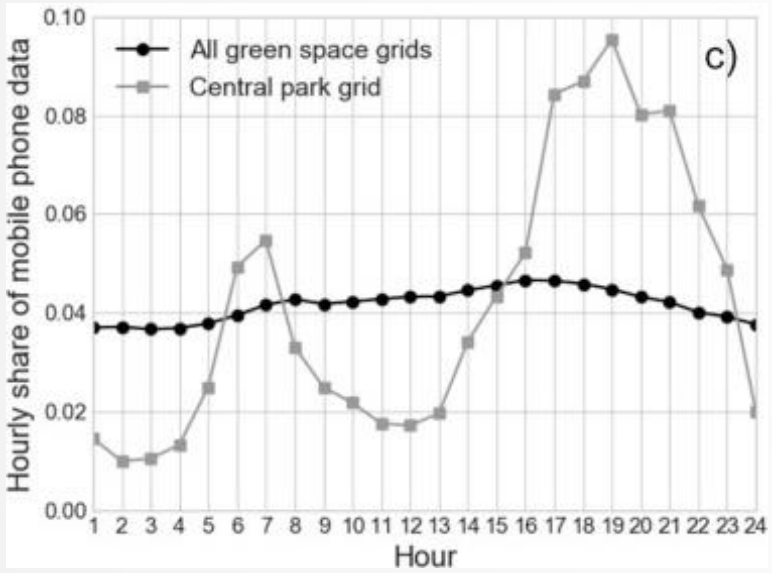


HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

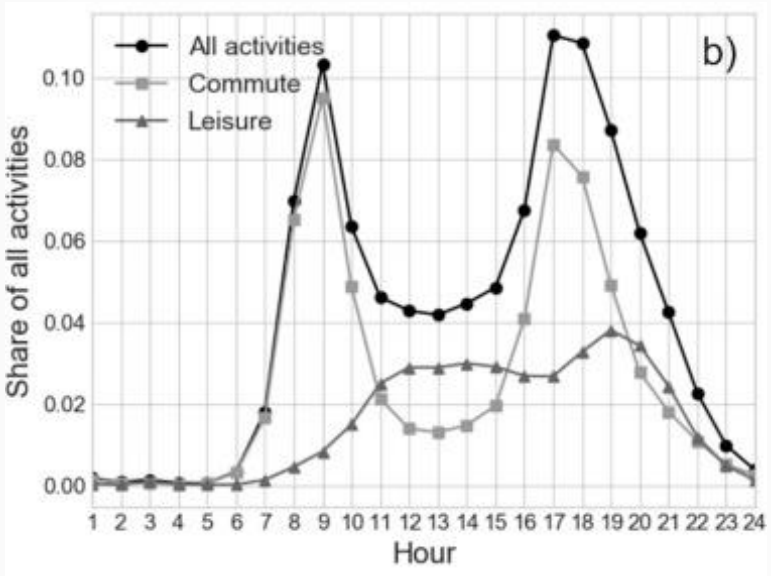

UNEXPECTED EVENTS LIKE THE PANDEMIC
HIGHLIGHT THE VALUE OF CONTINUOUSLY
COLLECTED DATA


DATA SOURCES ARE NOT THE SAME

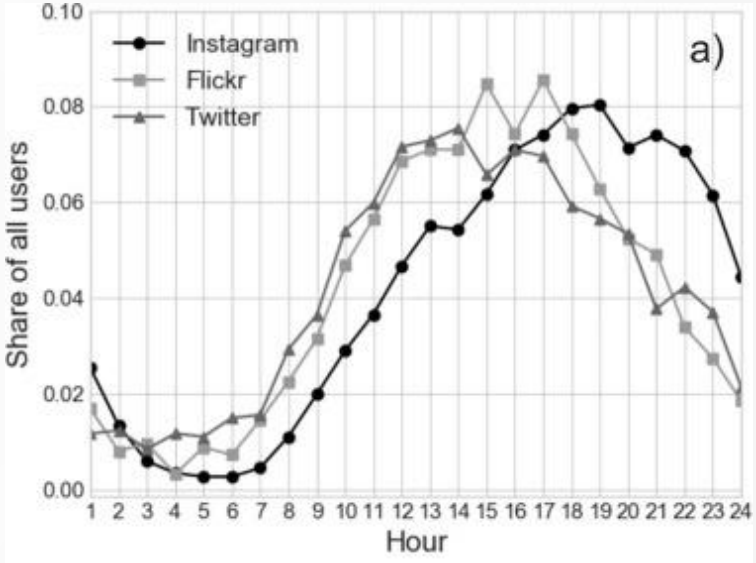
HOURLY USE OF GREEN AREAS OF HELSINKI ACCORDING TO VARIOUS BIG DATA



MOBILE PHONE

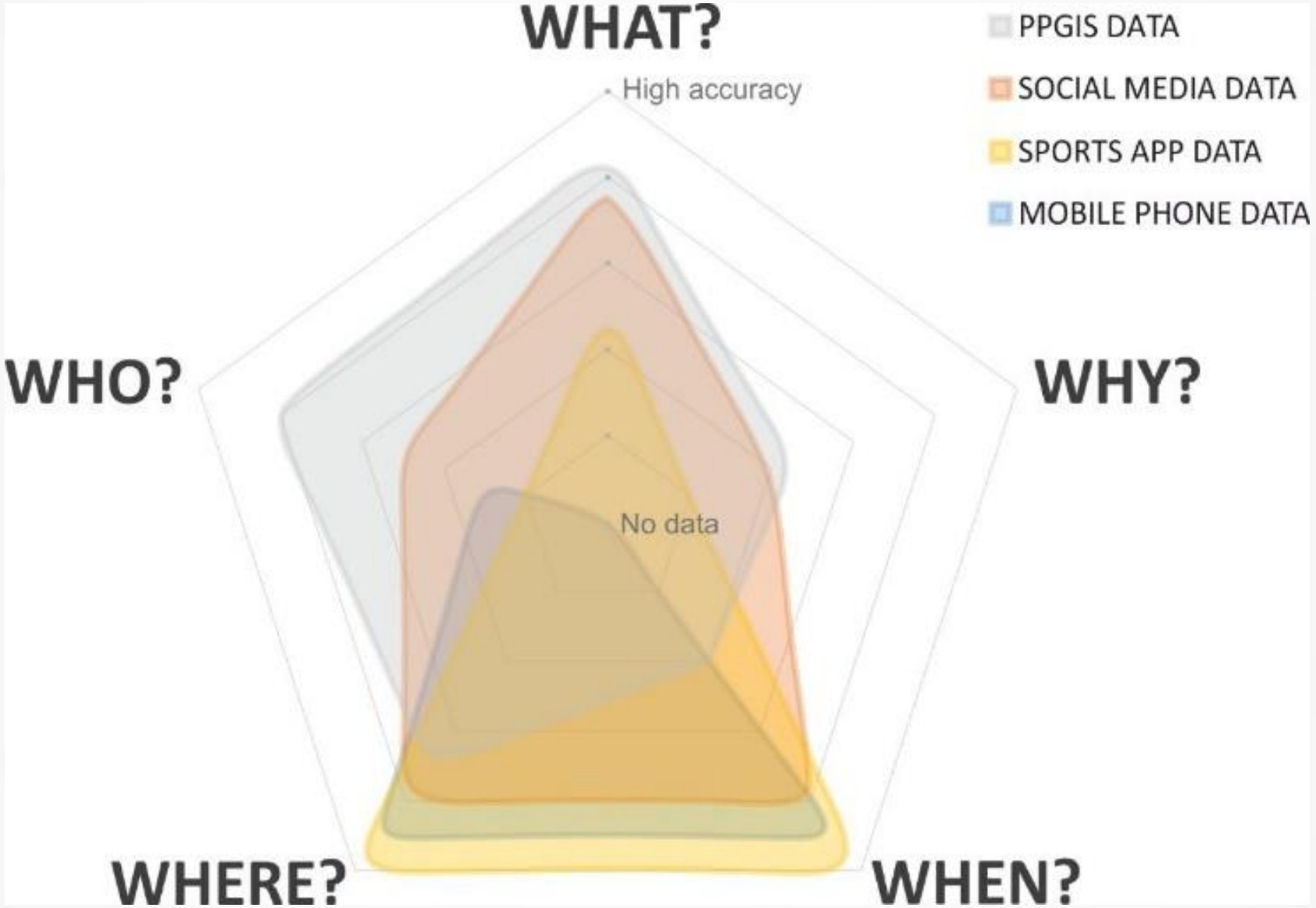


STRAVA



SOCIAL MEDIA

Heikinheimo et al. 2020 https://doi.org/10.1016/j.landurbplan.2020.103845



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Heikinheimo et al. 2020 https://doi.org/10.1016/j.landurbplan.2020.103845

WHAT DATA TO USE (ON CYCLING)?

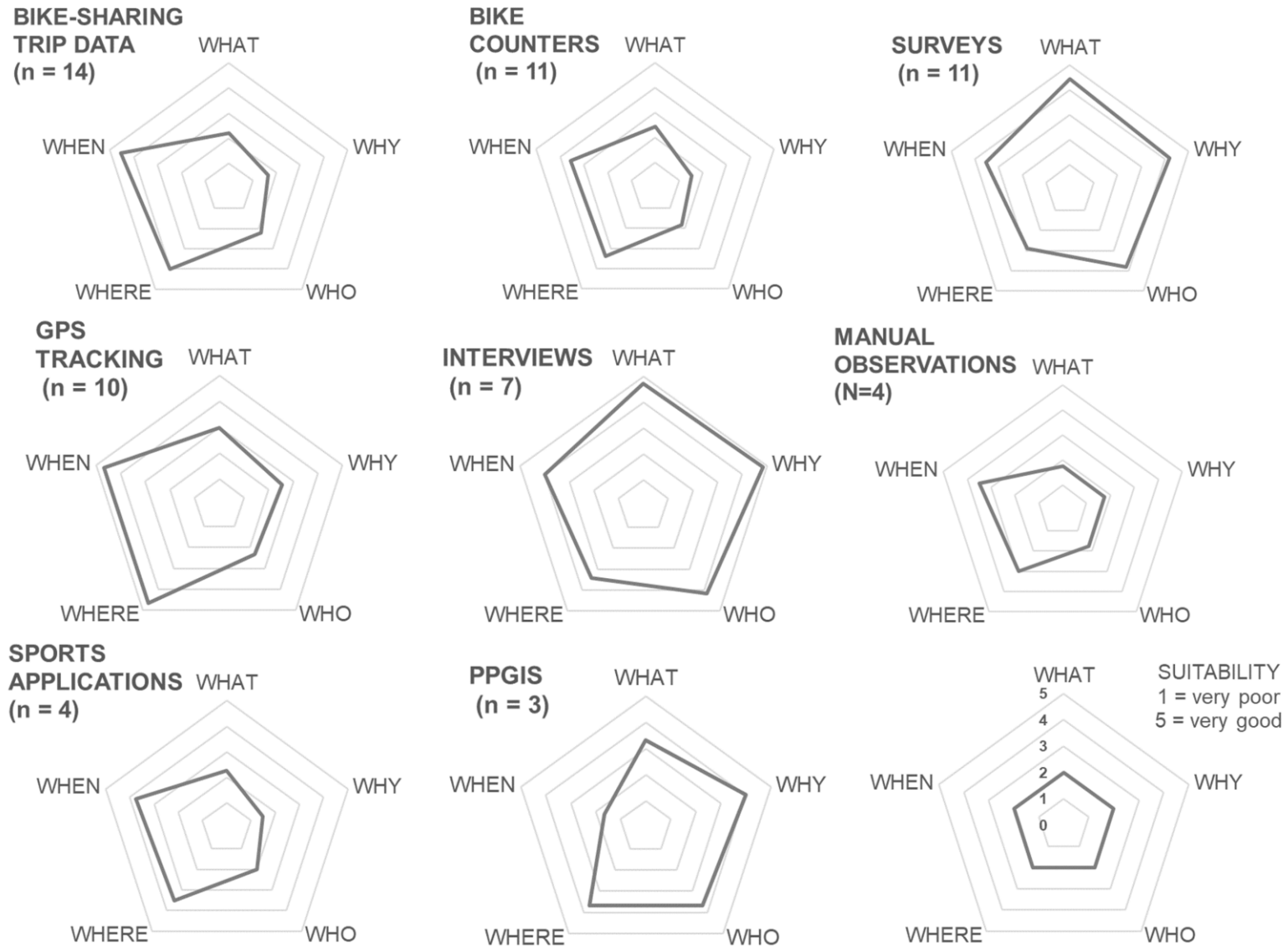
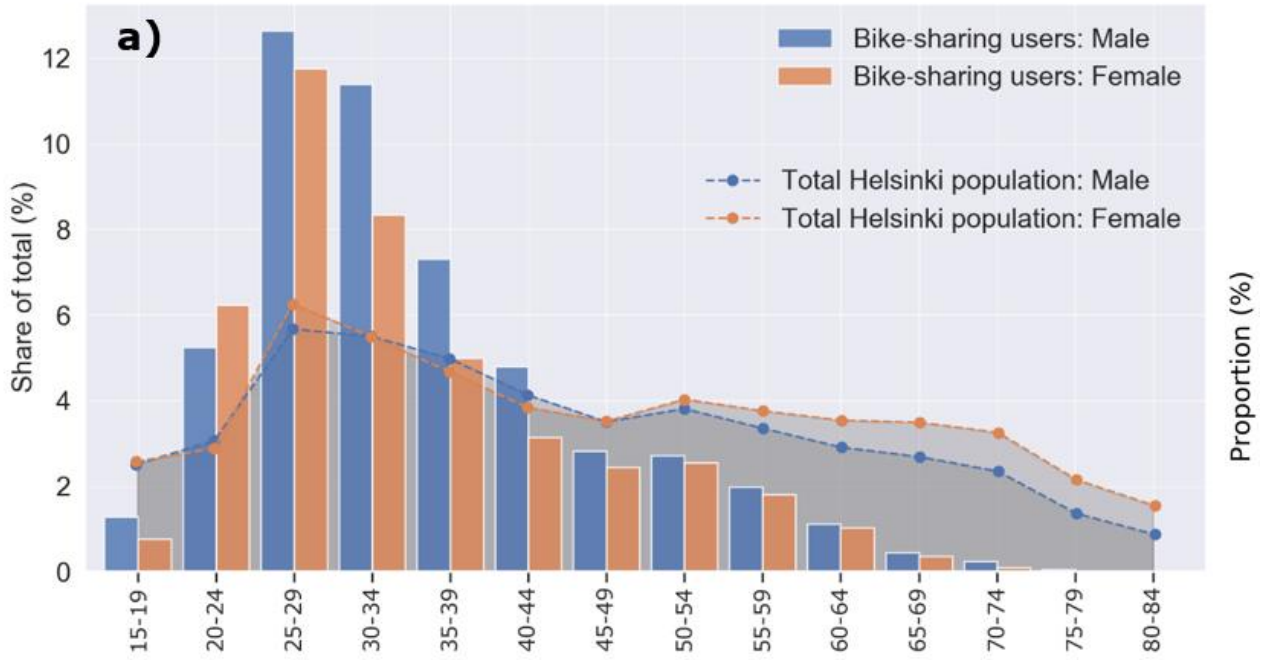


Figure from:
Willberg et al., 2021b)


NOT EVERYONE IS REPRESENTED



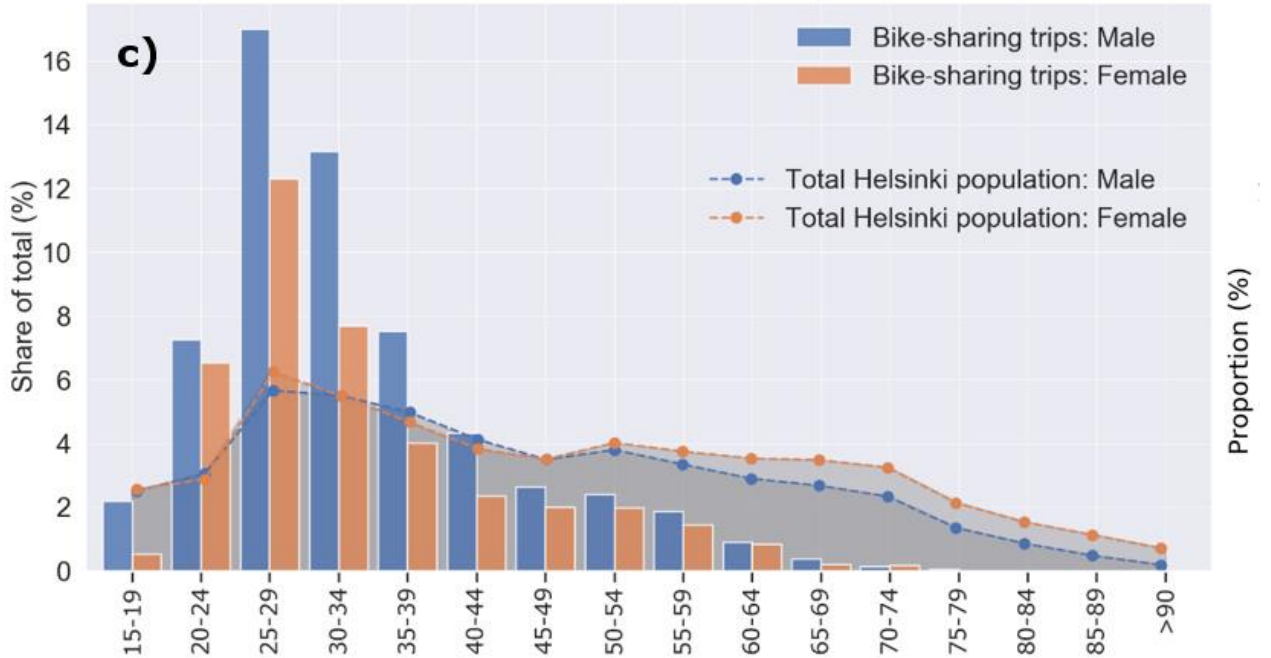


Figure from:
Willberg et al., 2021a


EXPECTATIONS ARE SOMETIMES
EXAGGERATED


BIG DATA MAY TURN OUT TO BE SMALL

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MANY THANKS

KONEEN SÄÄTIÖ

EMIL AALTOSEN SÄÄTIÖ

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

METSÄHALLITUS

@digigeolab

@TuuliToivonen

SOME REFERENCES

Toivonen, TK, Heikinheimo, VV, Fink, CA, Hausmann, A, Hiippala, T, Järv, O, Tenkanen, HTO & Di Minin, E 2019, 'Social media data for conservation science: a methodological overview', Biological Conservation, 233, 298-315. https://doi.org/10.1016/j.biocon.2019.01.023

Heikinheimo, V., Minin, E. Di, Tenkanen, H., Hausmann, A., Erkkonen, J., Toivonen, T., 2017. User-Generated Geographic Information for Visitor Monitoring in a National Park: A Comparison of Social Media Data and Visitor Survey. ISPRS Int. J. Geo-Information 6, 85. doi:10.3390/ijgi6030085

Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, Flickr, or Twitter: Assessing the usability of social media data for visitor monitoring in protected areas. Sci. Rep. 7, 17615. doi:10.1038/s41598-017-18007-4

Di Minin, E., Tenkanen, H., Toivonen, T., 2015. Prospects and challenges for social media data in conservation science. Front. Environ. Sci. 3. doi:10.3389/fenvs.2015.00063

Willberg, E., Salonen, M., & Toivonen, T. (2021a). What do trip data reveal about bike-sharing system users? Journal of Transport Geography, 91, 102971. https://doi.org/10.1016/j.jtrangeo.2021.102971

Willberg, E., Tenkanen, H., Poom, A., Salonen, M., & Toivonen, T. (2021b). Comparing spatial data sources for cycling studies – a review. In M. N. Mladenović, T. Toivonen, E. Willberg, & K. T. Geurs (Eds.), Transport in Human Scale Cities. Edwar Elgar Publishing Limited.

Willberg, E., Poom, A., Helle, J., Toivonen, T. (forthcoming. Exposure of urban cyclists during travel to air pollution, noise, and greenery: From spatio-temporal mapping to measuring the role of route-choice. Unpublished

For more, see https://www2.helsinki.fi/en/researchgroups/digital-geography-lab/publications