

Representativeness of crowdsourced cycle volumes:

Exploring the challenges and possibilities of
mobile sports tracking application data

Cecilia Bergman and Juha Oksanen

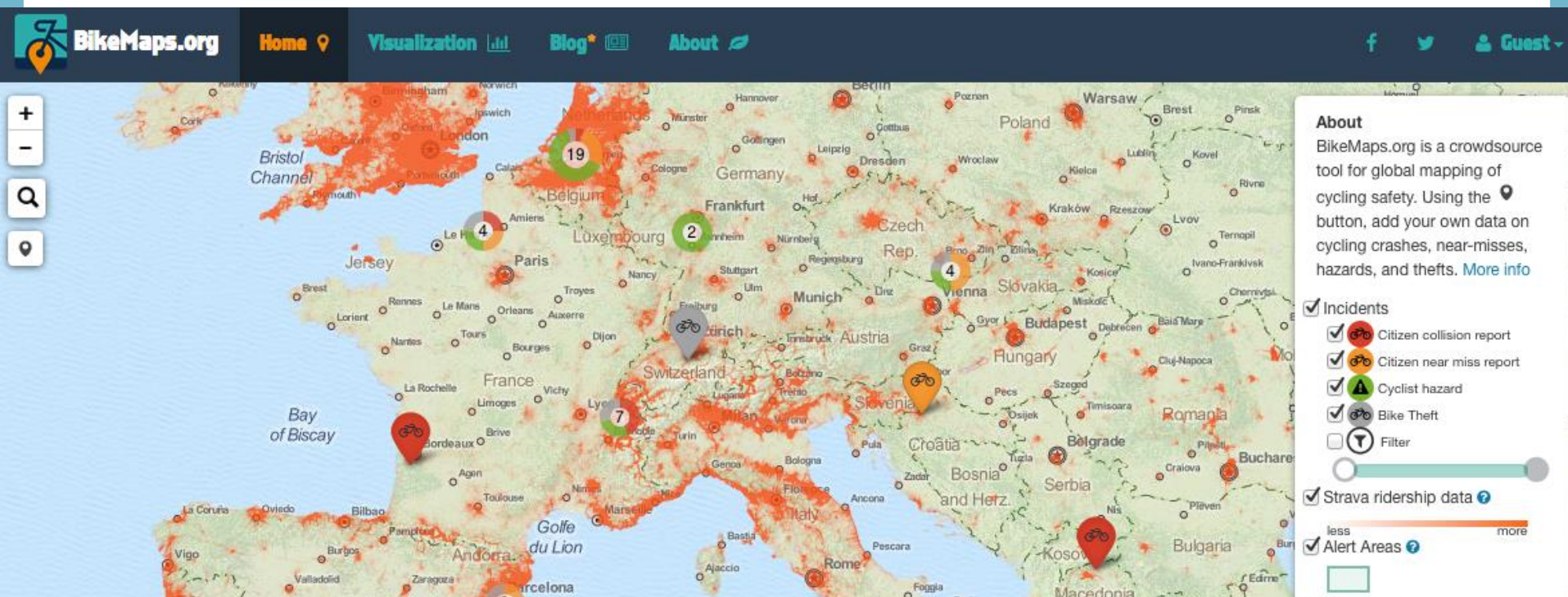
June 9th 2015, RICH-VGI workshop @ the 18th AGILE Conference,
Lisbon



Supra

Crowdsourcing cycling data

- "...data related to active transportation is often limited"
 - both spatially and temporally
- GPS traces (in situ data): two categories



1. Dedicated smartphone applications for the purpose of collecting data for research and planning to e.g. enhance cycling infrastructure with cost-effective interventions

Cities collecting data with CycleTracks

- Austin, TX (University Transportation Center for Mobility report: "[Using Smartphones to Collect Data in Texas](#)" and [CycleTracks Austin](#))
- Monterey, CA
- Raleigh, NC
- Fort Collins, CO
- Minneapolis/St. Paul, MN
- Seattle, WA
- Salt Lake City, Utah
- Los Angeles, CA
- Toronto, Ontario
- Lexington, KY

Interested in using CycleTracks in your city? Send an email to [info@cycletracks.org](#)

CycleTracks rebranded and improved by other cities

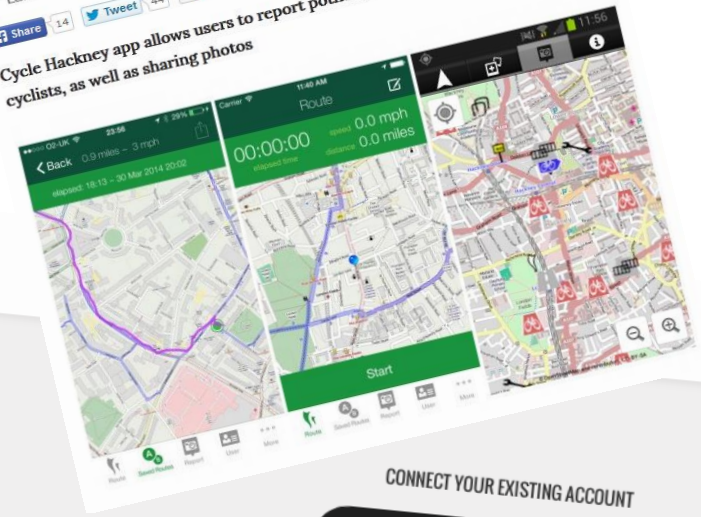
- Lane County, OR (LaneTracks)
- College Station, TX (AggieTracks)
- Charlottesville, VA ([C-Vill Bike mAPP](#))
- Hampton Roads, VA
- Atlanta, GA ([Cycle Atlanta](#))
- Montreal, Quebec ([My ResoVelo](#))
- Reno, NV ([RenoTracks](#))
- Philadelphia, PA ([CyclePhilly](#))

New app crowd sources information for cyclists

Laura Laker June 10, 2014 0 Comments

Share 14 Tweet 44 8+1 2 Pin it 0 Share 62

Cycle Hackney app allows users to report potholes and report problem areas for cyclists, as well as sharing photos



BIKE DATA PROJECT

DONATE YOUR
BIKE DATA

There are several ways you can contribute data to the project. If you already use one of the apps listed below you can easily connect them to The Bike Data Project. Once your app is connected you can just carry on as usual and your rides will automatically be uploaded to our service.

You can also download the "Bikes vs Cars" app for iPhone or Android. Bikes vs Cars started as a documentary film by the well renowned director Fredrik Gertten and turned into a global movement.

CONNECT YOUR EXISTING ACCOUNT



MOVES



RUNKEEPER



STRAVA

DOWNLOAD BIKES VS CARS



IPHONE



ANDROID

CONTACT

2. Mobile fitness applications for tracking own workouts with possibility to share them with friends or everyone and to keep an online workout diary.

ODOT embarks on “big data” project with purchase of Strava dataset

Posted by Jonathan Maus (Publisher/Editor) on May 1st, 2014 at 11:32 am

The Oregon Department of Transportation agency (ODOT) is the first state transportation agency in the United States to ink a deal with **Strava**, a leading website and smartphone app used by people to track their bike rides via GPS.

Last fall, the agency paid \$20,000 for one-year license of a dataset that includes the activities of about 17,700 riders and 400,000 individual bicycle trips totaling 5 million BMT (bicycle miles traveled) logged on Strava in 2013. The Strava bike “traces” are mapped to **OpenStreetMap**.

If all goes according to plan, the data could revolutionize how ODOT makes decisions about their policies, plans, and projects. At the very least, forging boldly into the realm of “big data” and pushing the boundaries of bicycle planning marks an important step for an agency that’s **facing a very different future** and actively **looking to shed** its old-school, highway-first reputation.



Dedicated apps vs. Fitness apps

- **Similarities:** information about travel mode, speed, distance, time, exact route...
- **Differences:**
 - **Dedicated apps:** potentially more demographic data about the cyclist and trip purpose, some control over data collection - people are contributing to a common/collaborative project
 - **VGI** (volunteered crowdsourced data)
 - **Fitness apps:** amount of data/larger population, no control over data collection – brand/target audience (utilitarian, recreational, sport cyclists), no common project, private/public division
 - **CGI** (contributed crowdsourced data)
- *VGI / CGI - Important for assessing data's fitness for use, identifying biases or inaccuracies. - F. Harvey (2015)*

Whom does fitness app data represent?



” Primarily the **generalizability** of MapMyFitness data must be thoughtfully considered before use in research.

” In-situ data may be [...] **biased toward specific types of users.**

” This application captures a quite **small portion of the population** as Strava users.

” Planners need to understand the **representation** of each dataset used, and **its limits to a particular application.**

” [...] where bicyclists ride, primarily for **fitness purposes.**

” Though the marketing focus of the Strava app is oriented towards fitness, **it is likely that its users log trips for other purposes, as well.**

Data	Demography/ generalizability	Other notes	Reference
Strava Travis County, Texas	Bias towards young and middle-aged males (< 25% female)	22% commuting rides	Griffin et al. 2015
Sports Tracker (global sample; cities with enough data)	Comparison to MapMyRide data: Pearson correlation 0.55 (std 0.37)	Number of tracks per city (> 200) is very small	Ferrari & Mamei 2013
Endomondo Global sample	-	Absolute numbers are not interesting without ref. data	Costés et al. 2014
MapMyFitness Winston-Salem, NC	Majority of users female (57.1%) and between ages of 18 and 44.	All workout types	Hirsch et al. 2014

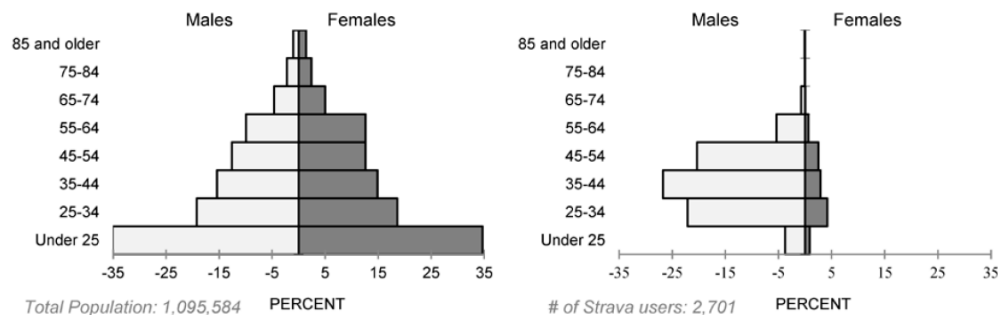
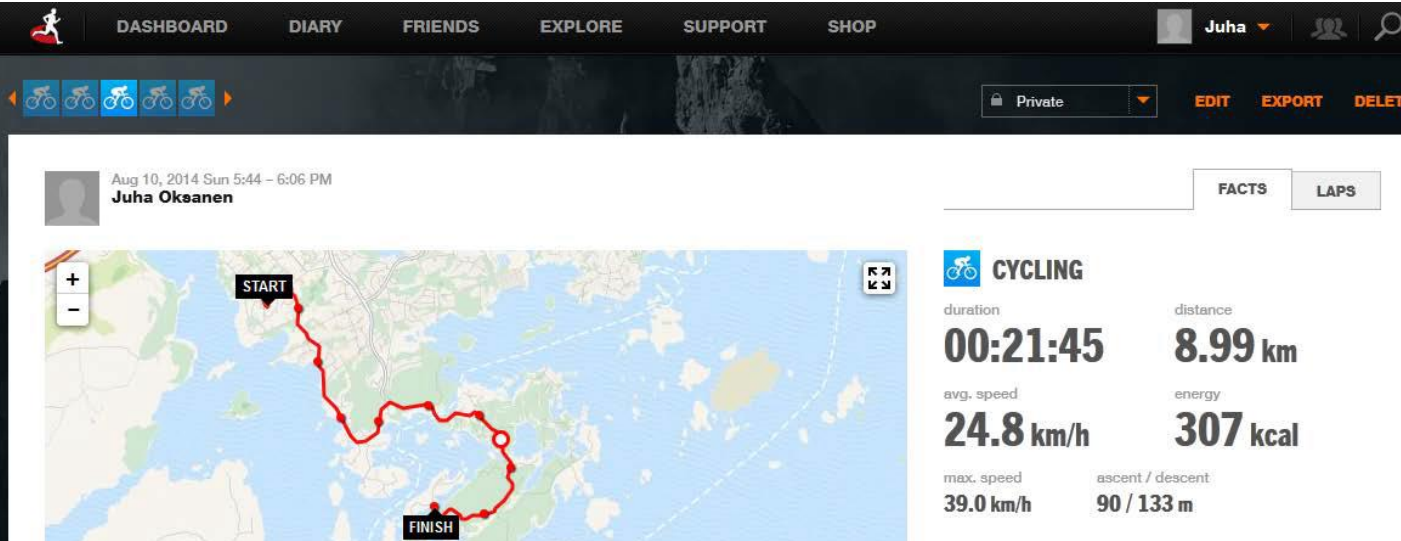
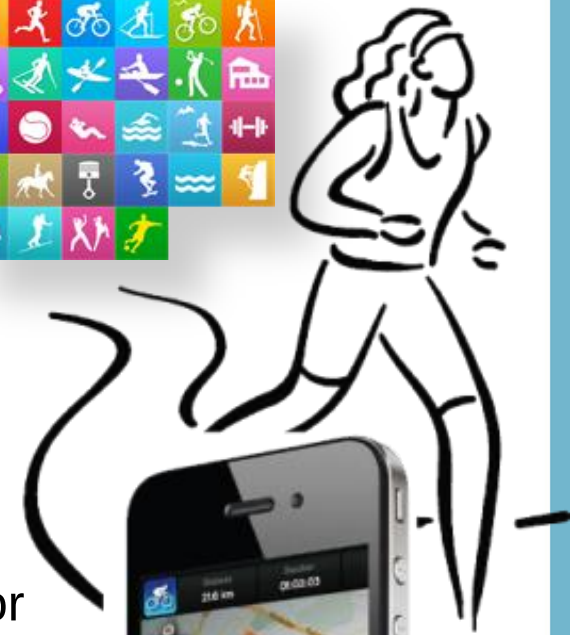


FIGURE 2. Travis County population estimate by age and sex at left (U.S. Census Bureau 2012), Strava fitness app bicyclists at right (Kitchel & Riordan 2014)

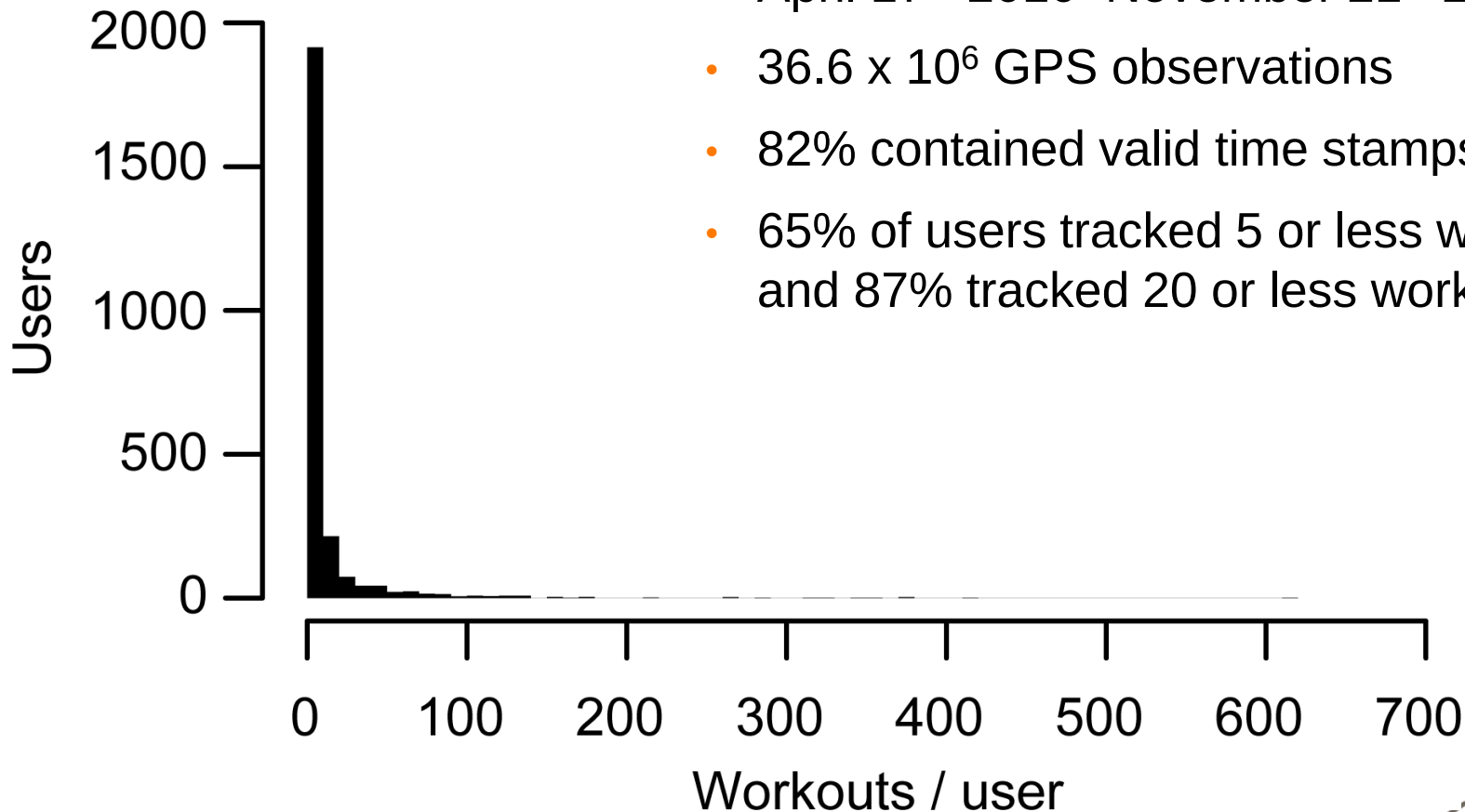
What is Sports Tracker?

- A free mobile GPS tracking app aimed at keeping diary about sports activities (=workouts)
- Online service sports-tracker.com
- Social network
- Started already in 2004 inside Nokia, now independent with millions of application downloads for iOS, Android, WP, Nokia N9 and Symbian
- Top 5 sports app in 25 countries



Characteristics of the data

- 36 757 workouts, 2424 users
- April 17th 2010–November 21st 2012
- 36.6×10^6 GPS observations
- 82% contained valid time stamps
- 65% of users tracked 5 or less workouts and 87% tracked 20 or less workouts



Characteristics of the data

The 90-9-1 Rule for Participation Inequality in Social Media and Online Communities

by JAKOB NIELSEN on October 9, 2006

Topics: Social UX

Summary: In most online communities, 90% of users are lurkers who never contribute, 9% of users contribute a little, and 1% of users account for almost all the action.

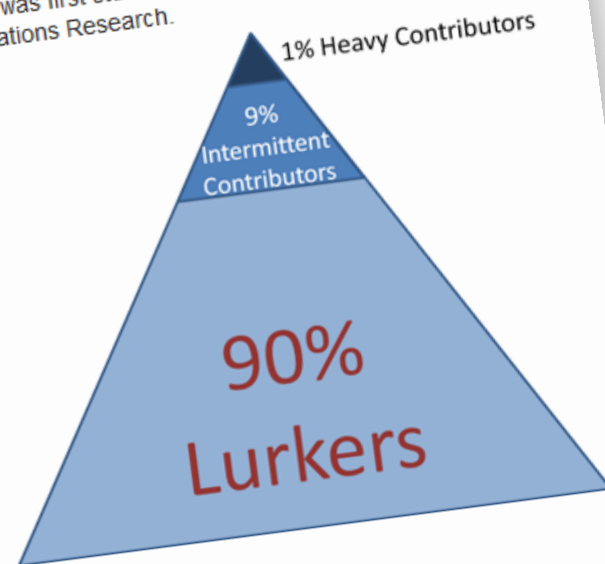
All large-scale, multi-user communities and online social networks that rely on users to contribute content or build services share one property: **most users don't participate** very much. Often, they simply **lurk** in the background.

In contrast, a tiny minority of users usually accounts for a disproportionately large amount of the content and other system activity. This phenomenon of **participation inequality** was first studied in depth by Will Hill in the early '90s, when he worked down the hall from me at Bell Communications Research.

When you plot the amount of activity for each user, the result is a [Zipf curve](#), which shows as a straight line in a [log-log diagram](#).

User participation often more or less follows a **90-9-1 rule**:

- **90% of users are lurkers** (i.e., read or observe, but don't contribute).
- **9% of users contribute from time to time**, but other priorities dominate their time.
- **1% of users participate a lot and account for most contributions**: it can seem as if they don't have lives because they often post just minutes after whatever event they're commenting on occurs.



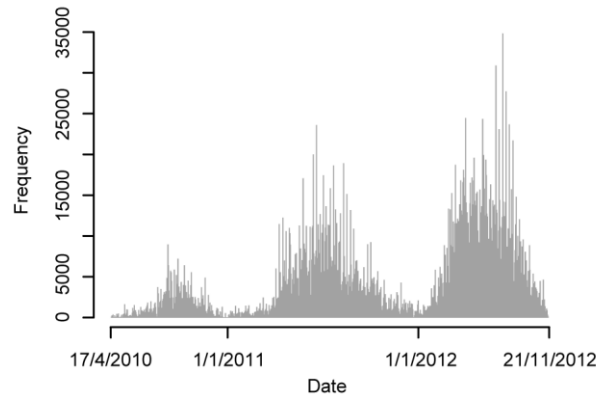
st 2012

ps

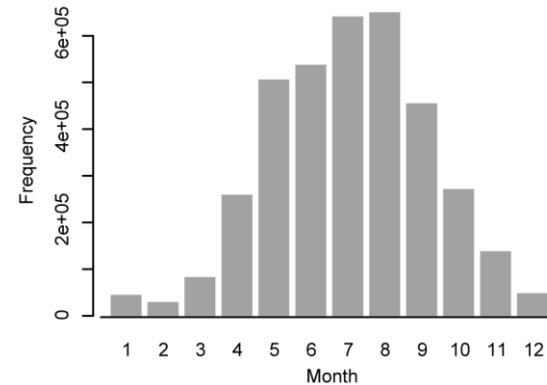
workouts
kouts

Data cyclicity - Helsinki

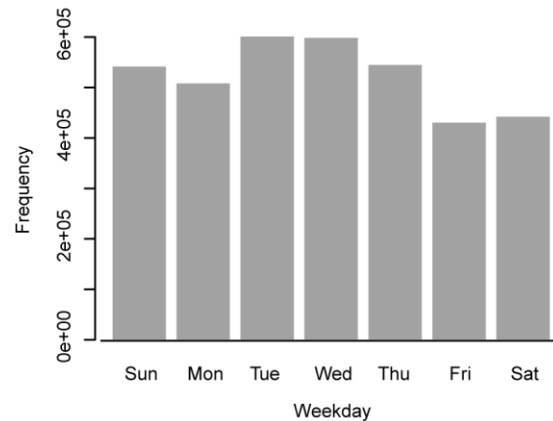
Yearly distribution



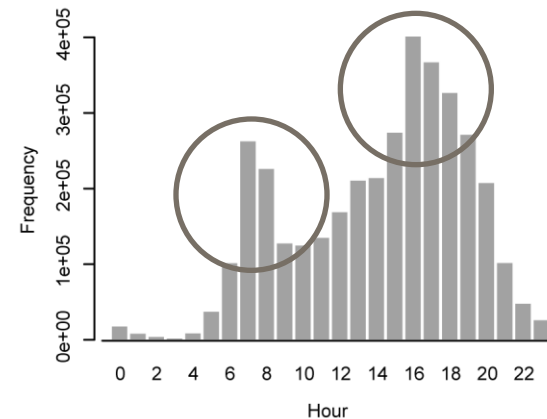
Monthly distribution



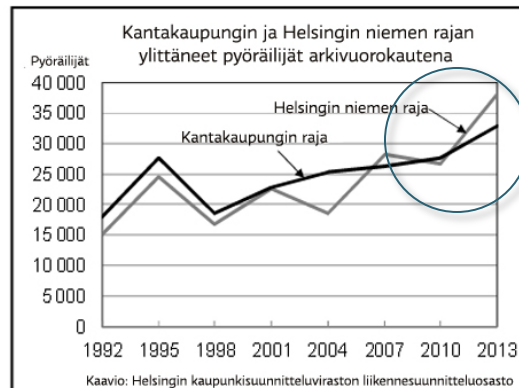
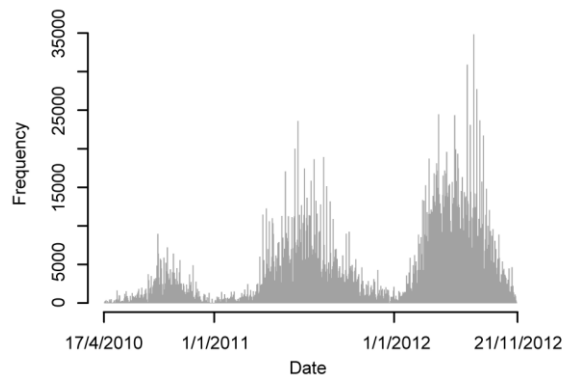
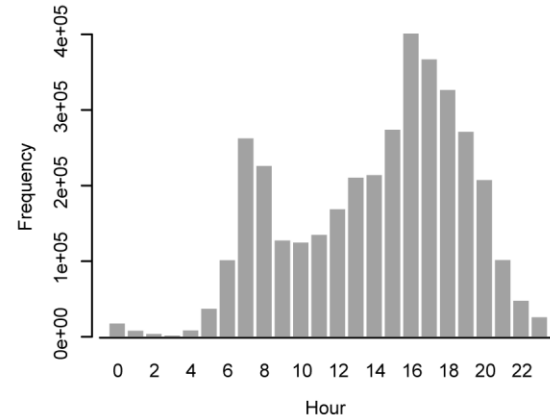
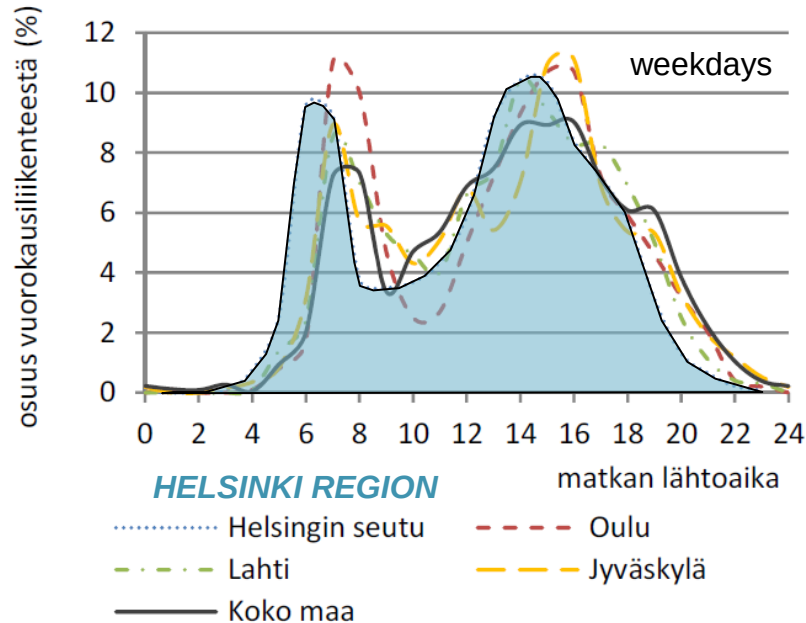
Daily distribution



Diurnal distribution

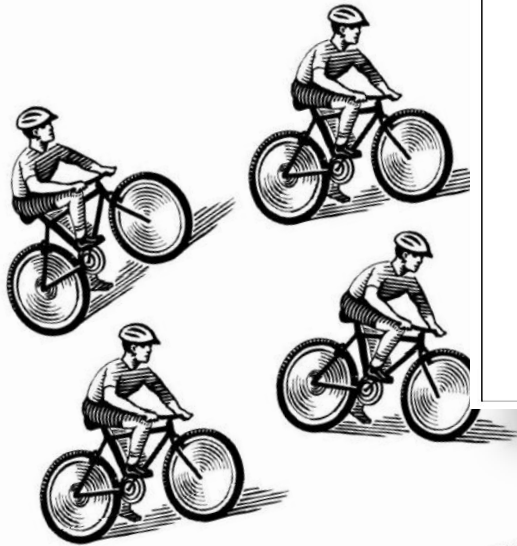


Sports Tracker vs. Traffic surveys

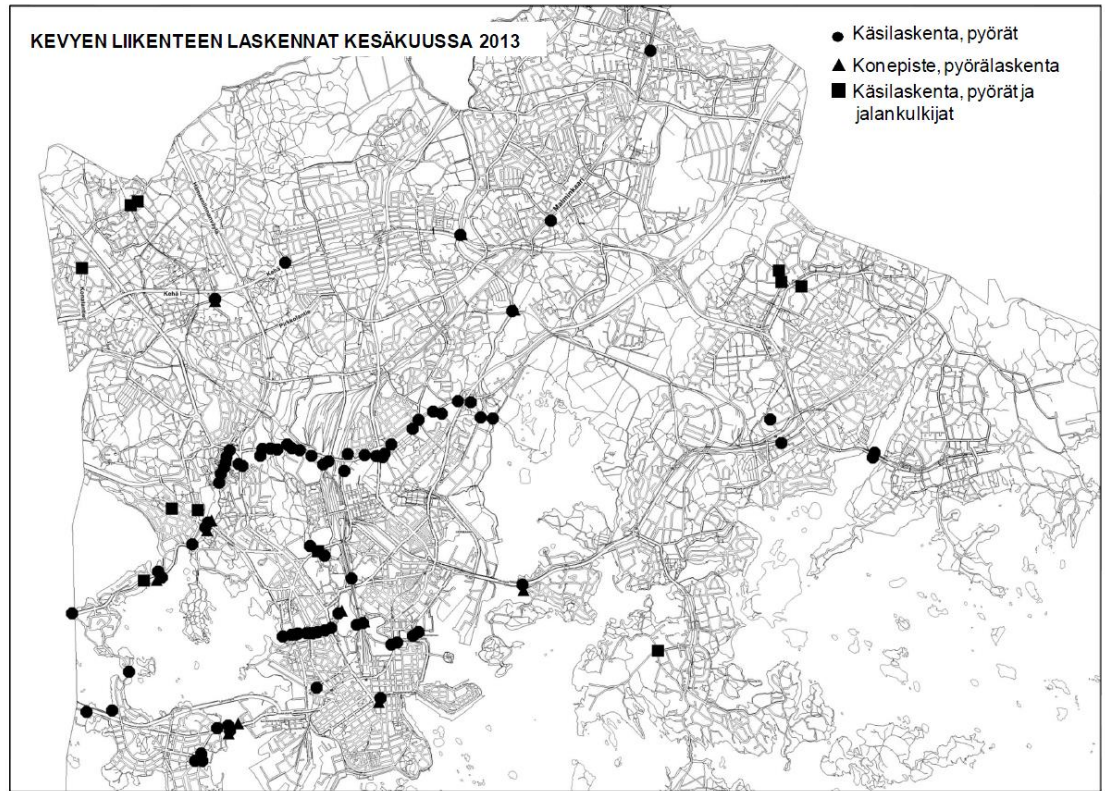


Helsinki Bicycle Count 2013

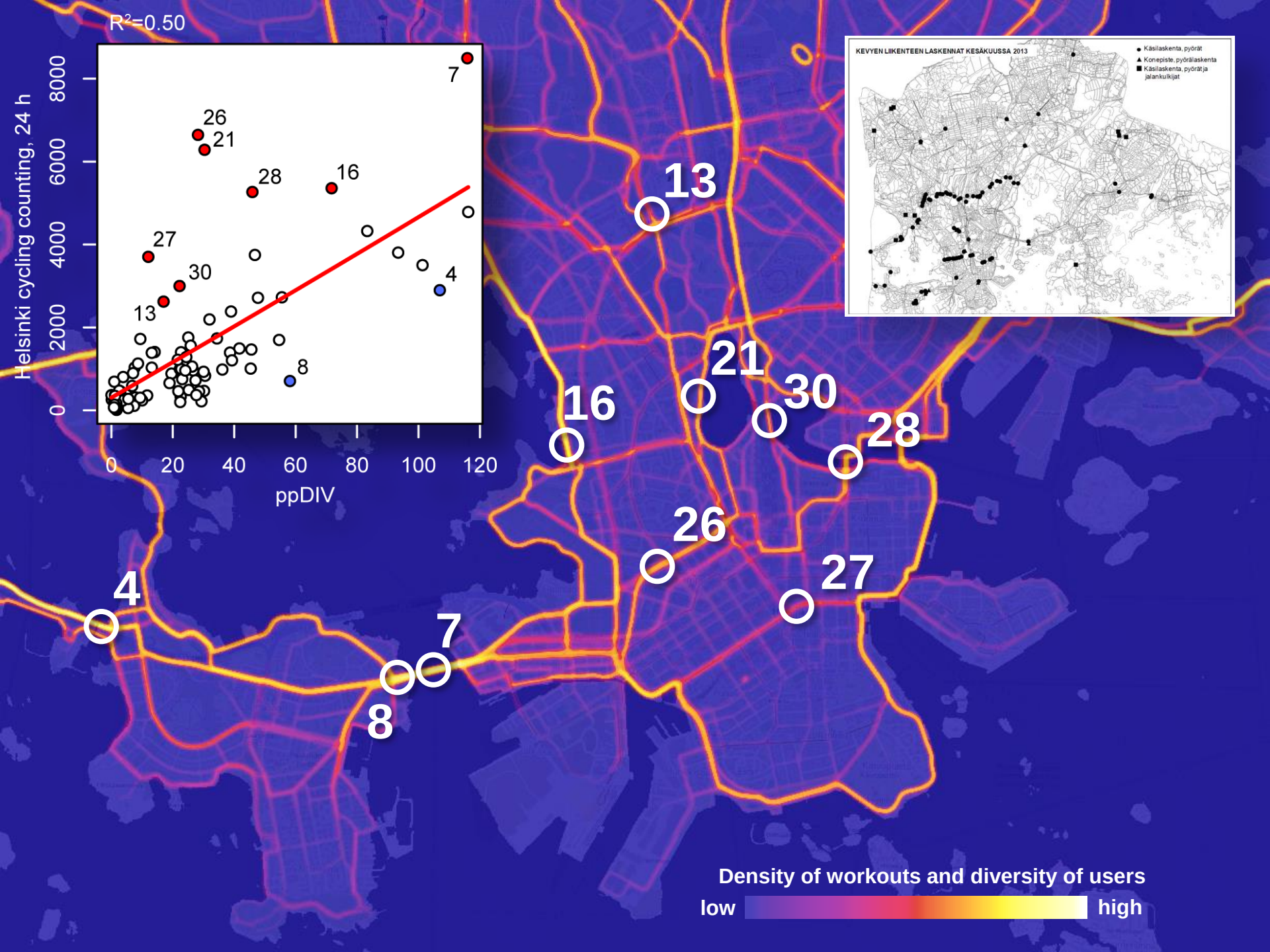
POLKUPYÖRÄLASKENNAT
HELSINGISSÄ
2013



Muistio 28.10.2013 / Tuija Hellman
Helsingin kaupunkisuunnitteluvirasto
Liikennesuunnitteluosasto

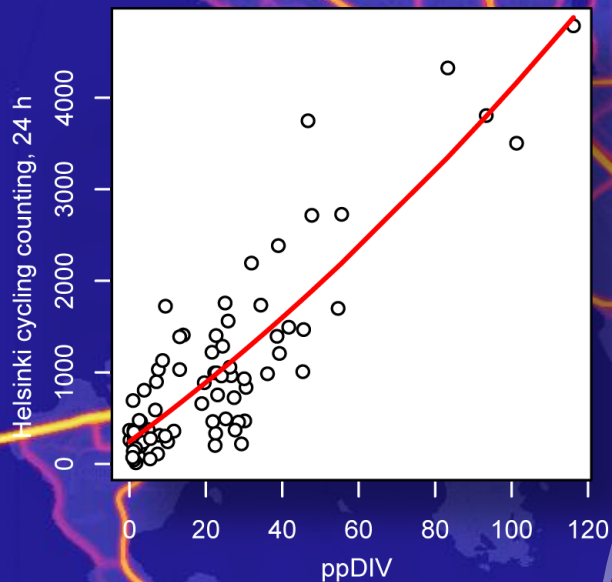


VOLUME + HELMET USE
(F/M)



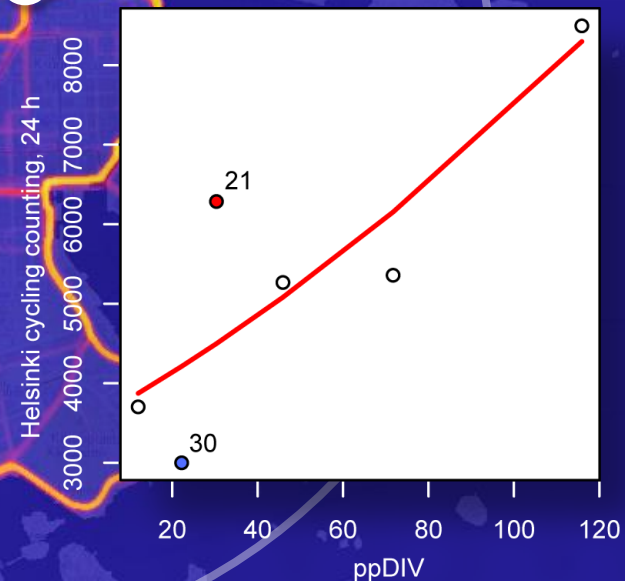
>2,5 km from central railway station

$R^2=0.76$



<2,5 km from central railway station

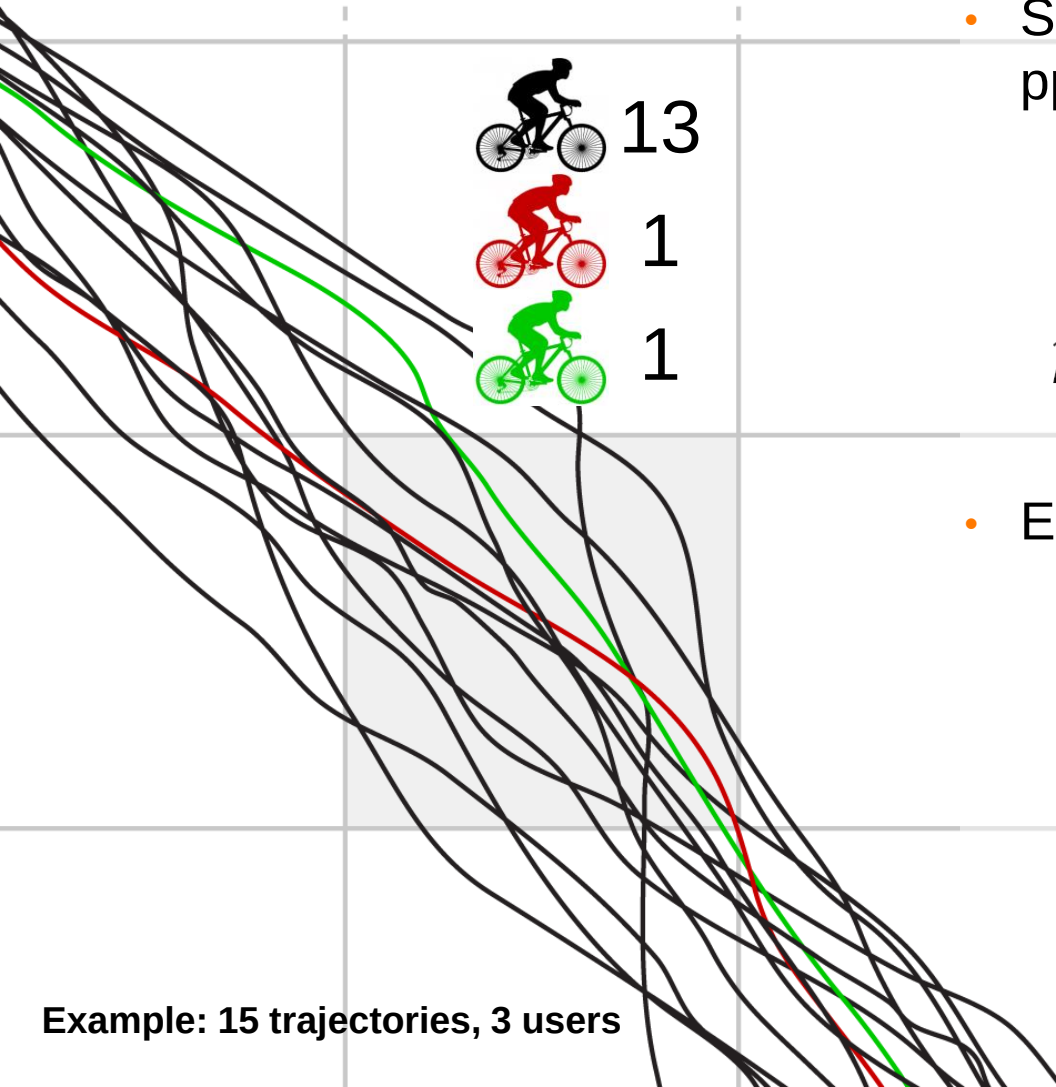
$R^2=0.72$



Density of workouts and diversity of users

low high

Method - ppDIV

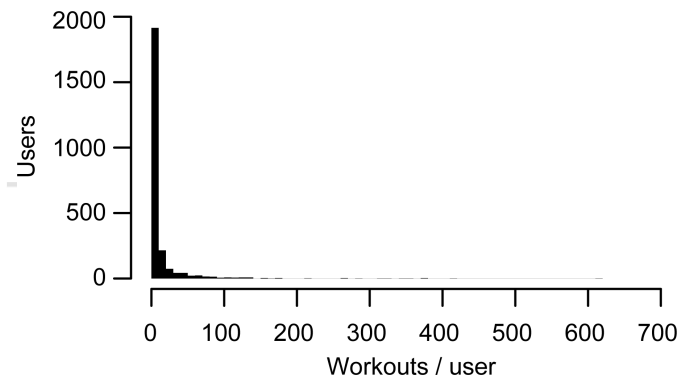


- Simpsons diversity index D and ppKDE:

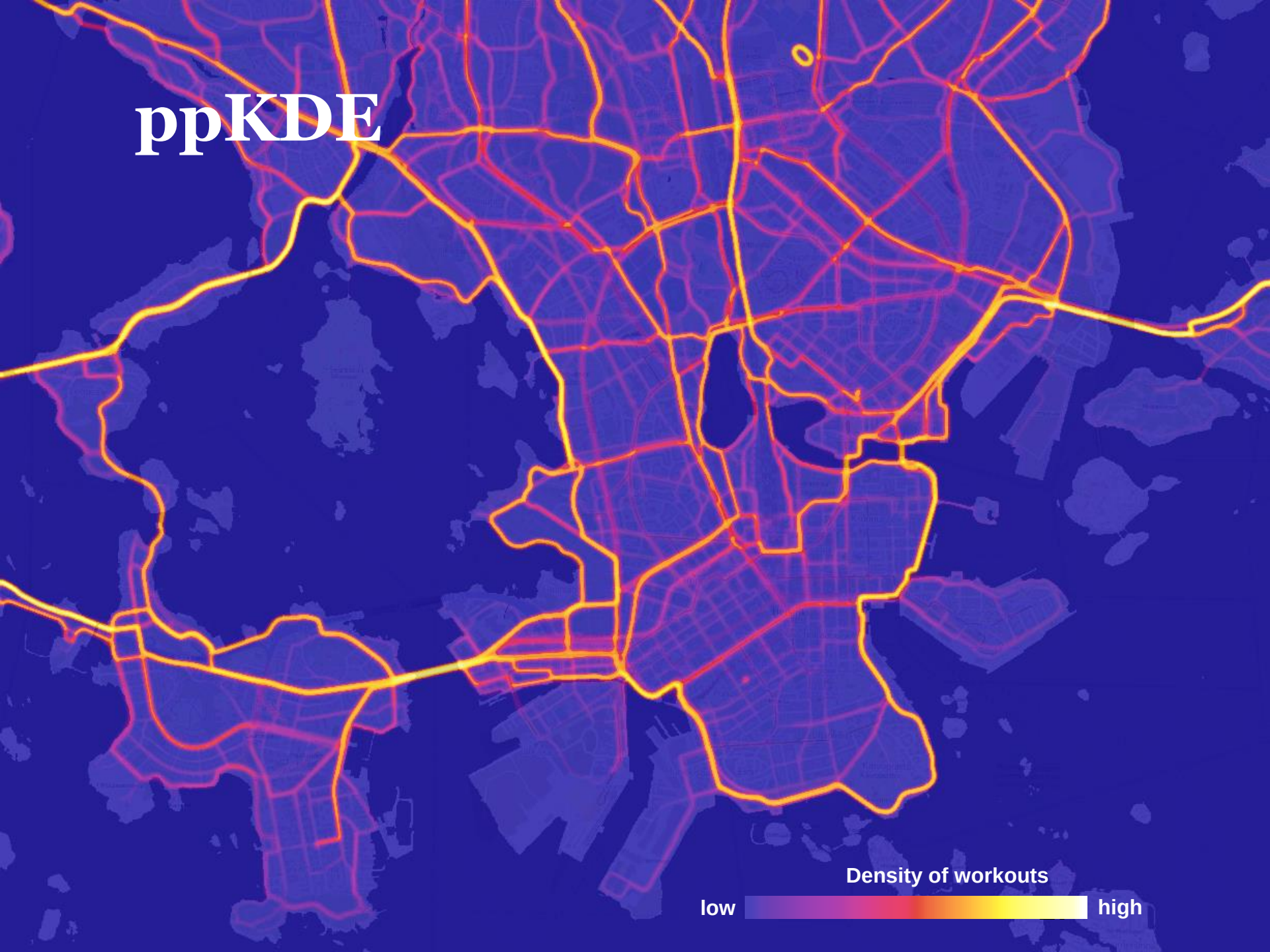
$$D(s) = 1 - \sum p(s)_i^2$$

$$ppDIV(s) = ppKDE(s) * D(s)$$

- Example: $D = 0.24$



ppKDE

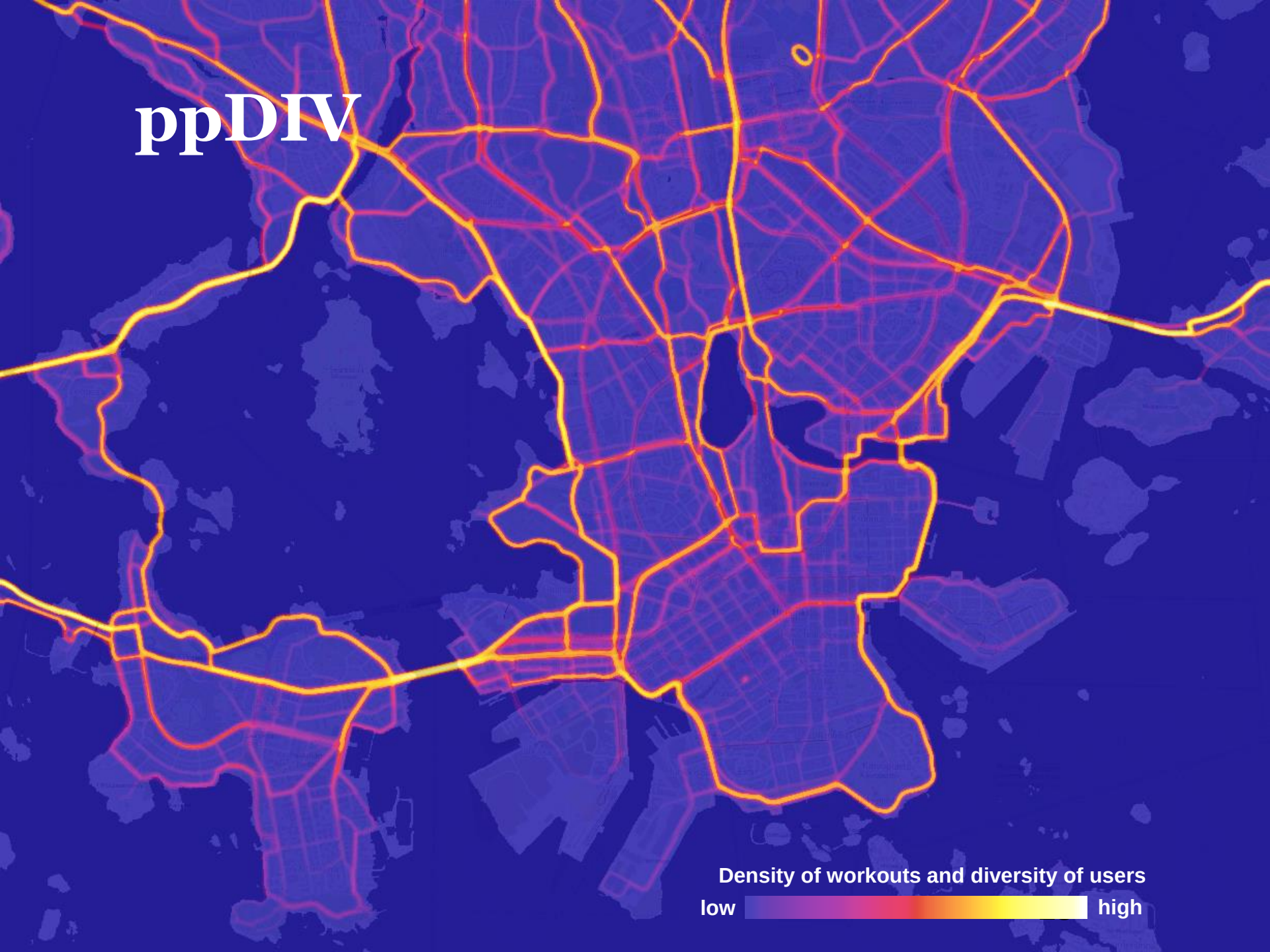


Density of workouts

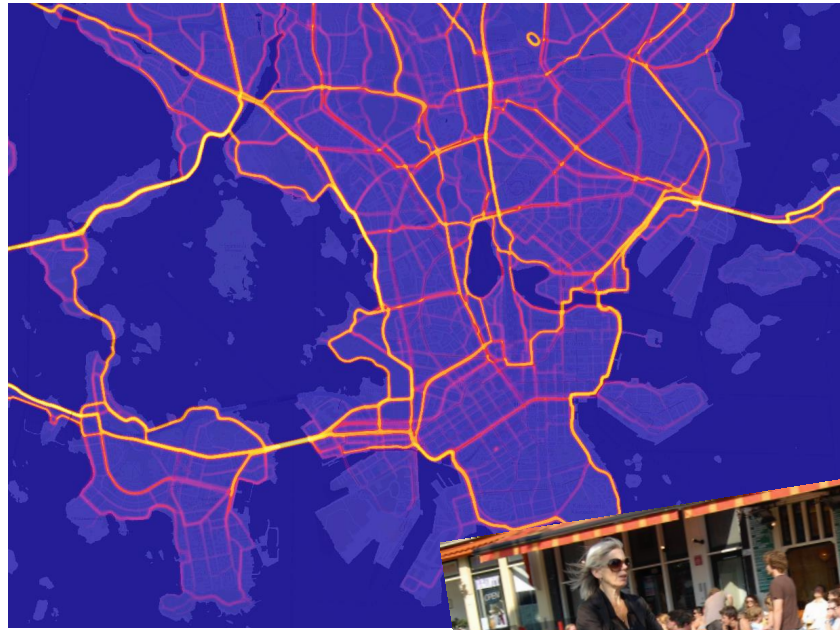
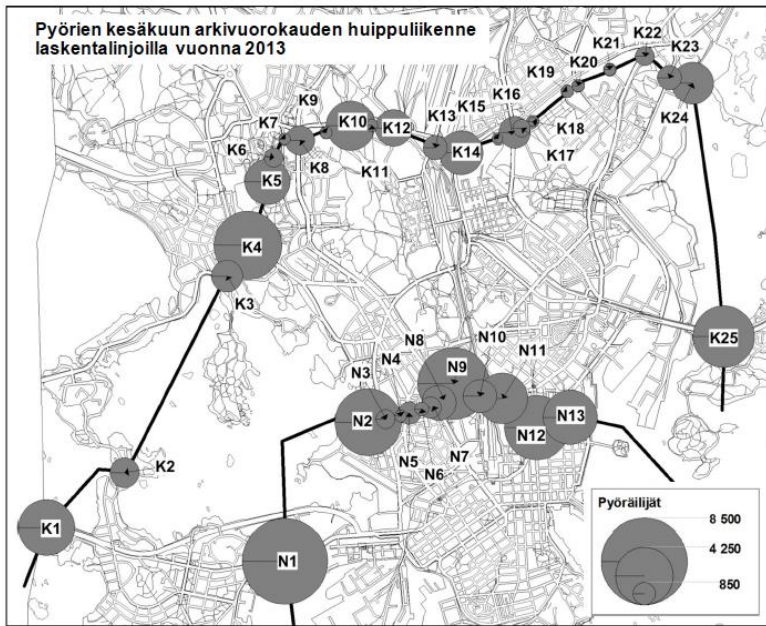
low

high

ppDIV



Density of workouts and diversity of users
low high



Does this tell something about..

- Cycling - physical activity?
- Sports Tracker application?
- Cycling in Helsinki?
- Once the dataset is large enough, the effect of biases gets eliminated ?



Conclusions

- Sports tracking apps provide a rich data source
- Users are not as limited to sport and recreational cyclists as we might expect (all cycling is physical activity)
- We need to understand what the data represents before drawing any conclusions based on it
- Diversity of applications
 - Dedicated apps vs. fitness (sports tracking) apps
 - Sports tracking applications have their differences
- 'Participation inequality' should not be neglected

Acknowledgements

SUPRA project, Tekes (40261/12)

Sports Tracking Technologies Ltd.

References

- Cortés R, Bonnaire X, Marin O, and Sens P. Sport Trackers and Big Data: Studying user traces to identify opportunities and challenges. *Research Report RR-8636, INRIA Paris*. 2014.
- Ferrari L and Mamei M. Identifying and understanding urban sport areas using Nokia Sports Tracker. *Pervasive and Mobile Computing* 9(5): 616-628, 2013.
- Griffin GB and Jiao J. Where does bicycling for health happen? Analysing volunteered geographic information through place and plexus. *Journal of Transport and Health*, 2015.
- Harvey F. To Volunteer or to Contribute Locational Information? Towards Truth in Labeling for Crowdsourced Geographic Information. In Sui D. et al. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. 2015
- Hirsch JA, James P, Robinson JRM, Eastman KM, Conley KD, Evenson KR, and Laden F. Using MapMyFitness to Place Physical Activity into Neighborhood Context. *Frontiers in Public Health* 2:19, 2014.

Oksanen J, Bergman C, Sainio J and Westerholm J. Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. Submitted.

Bergman C and Oksanen J. Conflation of OpenStreetMap and Mobile Sports Tracking Data for Automatic Bicycle Routing. Submitted.

email - cecilia.bergman@nls.fi
twitter - @cykelpige

<http://supra.fgi.fi>

Supra



NLS
FINNISH GEOSPATIAL
RESEARCH INSTITUTE
FGI