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
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 Coll 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

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

“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.

“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.

“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.

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

**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 x 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

- 36,757 workouts, 2,424 users
- April 17th, 2010 – November 21st, 2012
- 36.6 x 10^6 GPS observations
- 82% contained valid time stamps
- 65% of users tracked 5 or less workouts and 87% tracked 20 or less workouts
Data cyclicity - Helsinki

Yearly distribution

Monthly distribution

Daily distribution

Diurnal distribution
Sports Tracker vs. Traffic surveys

HELSINKI REGION

- Helsingin seutu
- Oulu
- Lahti
- Jyväskylä
- Koko maa

Weekdays

Frequency

0e-00  1e+05  2e+05  3e+05  4e+05

0  2  4  6  8  10  12  14  16  18  20  22  24

Frequency

0  5000  15000  25000  35000


Kantakaupungin ja Helsingin niemen rajan ylittäneet pyöräilijät arkivuorokautena

Kantakaupungin raja

Helsingin niemen raja

Date


Koeve: Helsingin kaupunkisuunniteluvuoteen liikennesuunnitelmannasto
Helsinki Bicycle Count 2013

VOLUME + HELMET USE (F/M)
Density of workouts and diversity of users

>2,5 km from central railway station

<2,5 km from central railway station
Method - ppDIV

- Simpsons diversity index $D$ and ppKDE:

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

$$ppDIV(s) = ppKDE(s) \times D(s)$$

- Example: $D = 0.24$

Example: 15 trajectories, 3 users
Density of workouts and diversity of users

ppDIV
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


e-mail - cecilia.bergman@nls.fi
twitter - @cykelpige

http://supra.fgi.fi

Supra