How to interpret Big Data in cycling context?

Dublin // 28th of June 2019

Dipl.-Ing. Sven Lißner
Institute for Transportation Planning and Road Traffic - Technische Universität Dresden
What is Big Data in Cycling?

Data Dimensions

- Hours of counting
- People in a cycling study
- People answering a household survey
- Hours of counting (365d)
- Bike Trips (citywide)
- Bike Trips (nationwide)
- GPS - Points (citywide)
- GPS - Points (nationwide)
- Acceleration Points (nationwide)
- GPS - Points (Strava)

![Data Dimensions Chart](image-url)
Benefits of Big Data in Cycling

- spot
- spatial
- static
- variable
- counts
- routes
Limitations of Big Data

Crowdsourcing means self-selection!
Limitations of Big Data

...and not selection by chance!

Age distribution of Strava users in Dresden

<table>
<thead>
<tr>
<th>Age in Years</th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td>under 25</td>
<td>12%</td>
<td>88%</td>
</tr>
<tr>
<td>25-34</td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>45-54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75-84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85-94</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

absolute Number
Limitations of Big Data

Self Selection possibly leads to

- Skilled...
- Digital...
- Middle-aged...
- Male.. cyclists

Unlike our Cycling Population?

COMPARISON OF STRAVA AND HOUSEHOLD SURVEY

<table>
<thead>
<tr>
<th>AGE GROUPS</th>
<th>UNDER 25</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65-74</th>
<th>&gt;74</th>
</tr>
</thead>
<tbody>
<tr>
<td>PERC. SHARE</td>
<td>34%</td>
<td>19%</td>
<td>21%</td>
<td>14%</td>
<td>7%</td>
<td>3%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Gender Share Household Survey (Cyclists)

- Household Survey
- Strava Data

Limitations of Big Data

- 34%
- 19%
- 21%
- 14%
- 7%
- 5%
- 1%
- 0%
The big question

Does this bias lead to substantially different cycling behaviour?
How to Interpret Big Data in Cycling Context – Results of a representative Cycling Study
Technische Universität Dresden / Sven Lißner
Velo City 2019 „Cycling for the Ages“ // Dublin // 2019/06/28

Different Behaviour...

Speed distribution Elberadewg at Waldschlösschenbrücke

Proportion of Strava trip length

...but not really comparable
What did we do?

Step 1: Conduct a Survey!

- Twitter, Facebook, Mailing Lists, Radio, Newspapers, Universities....
- “hire your Grandparents”
- Conducted in december 2017
- 11.000 participants
What did we do?

Step 2: Typology of Cyclists

- ambitious
- passionate
- pragmatic
- functional

Types of cyclists:
- weather and comfort
- rule violation
- frequency of use
- distance
- affective motivation
- instrumental motivation
- symbolic motivation
- subjective safety
- identification

What did we do?

- passionate
- pragmatic
- functional
- ambitious
What did we do?

Step 3: Conduct a field experiment

The Setup:
- 200 People
- 100 female
- 50 of each type
- 5 age classes
- oldest 88y

Duration:
- 2 Waves
- 2 Weeks in Mai/June
Step 4: Getting the Data
Step 4: getting the Data
What did we do?

Step 5: Data pre-processing
Step 5: processed Data
How to Interpret Big Data in Cycling Context – Results of a representative Cycling Study
Technische Universität Dresden / Sven Lißner
Velo City 2019 „Cycling for the Ages“ // Dublin // 2019/06/28

Evaluation

- 187 Cyclists with usable Data

- Temperature warm, hardly rainy
Evaluation

AVERAGE SHARE OF DAILY BICYCLE USAGE IN STUDY PERIOD

<table>
<thead>
<tr>
<th>TYPE OF CYCLIST</th>
<th>ambitious</th>
<th>functional</th>
<th>passionate</th>
<th>pragmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>10%</td>
<td>13%</td>
<td>13%</td>
<td>17%</td>
<td>25%</td>
</tr>
<tr>
<td>max</td>
<td>87%</td>
<td>93%</td>
<td>100%</td>
<td>93%</td>
</tr>
<tr>
<td>min</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
<td>7%</td>
</tr>
<tr>
<td>90%</td>
<td>83%</td>
<td>80%</td>
<td>83%</td>
<td>87%</td>
</tr>
<tr>
<td>Average</td>
<td>52%</td>
<td>51%</td>
<td>58%</td>
<td>58%</td>
</tr>
</tbody>
</table>
Evaluation

AVERAGE DAILY TRIP LENGTH FOR DIFFERENT TYPES OF CYCLISTS

<table>
<thead>
<tr>
<th>TYPE OF CYCLIST</th>
<th>10%</th>
<th>max</th>
<th>min</th>
<th>90%</th>
<th>Mittelwert</th>
</tr>
</thead>
<tbody>
<tr>
<td>ambitious</td>
<td>7,343</td>
<td>51,194</td>
<td>5,417</td>
<td>35,288</td>
<td>20,449</td>
</tr>
<tr>
<td>functional</td>
<td>5,005</td>
<td>34,347</td>
<td>556</td>
<td>21,579</td>
<td>13,138</td>
</tr>
<tr>
<td>passionate</td>
<td>4,776</td>
<td>50,202</td>
<td>1,617</td>
<td>22,048</td>
<td>13,055</td>
</tr>
<tr>
<td>pragmatic</td>
<td>5,403</td>
<td>27,594</td>
<td>2,221</td>
<td>24,946</td>
<td>13,300</td>
</tr>
</tbody>
</table>

- 10%: Lower quartile
- max: Maximum
- min: Minimum
- 90%: Upper quartile
- Mittelwert: Mean
## Evaluation

### facility type: average share of trip length per type of cyclist

<table>
<thead>
<tr>
<th>Facility Type</th>
<th>Ambitious</th>
<th>Functional</th>
<th>Passionate</th>
<th>Pragmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sidewalks/others</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Shared Cycling/Walking</td>
<td>13%</td>
<td>13%</td>
<td>11%</td>
<td>11%</td>
</tr>
<tr>
<td>Sidewalk, cycling allowed</td>
<td>4%</td>
<td>4%</td>
<td>4%</td>
<td>5%</td>
</tr>
<tr>
<td>Cycleways</td>
<td>10%</td>
<td>10%</td>
<td>9%</td>
<td>10%</td>
</tr>
<tr>
<td>Side Roads on street</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
<td>1%</td>
</tr>
<tr>
<td>Main Street on street</td>
<td>11%</td>
<td>8%</td>
<td>12%</td>
<td>11%</td>
</tr>
<tr>
<td>Side road</td>
<td>42%</td>
<td>45%</td>
<td>42%</td>
<td>43%</td>
</tr>
<tr>
<td>Main street</td>
<td>12%</td>
<td>11%</td>
<td>13%</td>
<td>12%</td>
</tr>
</tbody>
</table>
How to Interpret Big Data in Cycling Context – Results of a representative Cycling Study
Technische Universität Dresden / Sven Lißner
Velo City 2019 „Cycling for the Ages“ // Dublin // 2019/06/28

Evaluation

DISTRIBUTION OF SPEED
AFTER TYPE AND AGE GROUP

V IN KM/H

<table>
<thead>
<tr>
<th>TYPE OF CYCLIST</th>
<th>DISTRIBUTION AFTER TYPE AND AGE GROUP</th>
</tr>
</thead>
<tbody>
<tr>
<td>ambitious</td>
<td>16-29: 27.6</td>
</tr>
<tr>
<td></td>
<td>10%: 38.3</td>
</tr>
<tr>
<td></td>
<td>10%: 22.1</td>
</tr>
<tr>
<td></td>
<td>10%: 24.3</td>
</tr>
<tr>
<td></td>
<td>10%: 24.3</td>
</tr>
</tbody>
</table>

10%  20%  30%  40%  50%  60%  70%  80%  90%  100%
Why does this matter?

- Differences between different types of Cyclists are not as big as we thought

- Distribution of trip length looks very similar to random sample from household survey

- This kind of distribution seems representative and is to be achieved in biased samples

![DISTRIBUTION OF TRIP LENGTH](chart-image)
THANK YOU!

Sven Lißner
Chair of Transportation Ecology
Sven.lissner@tu-dresden.de

Angela Francke
Chair of Transportation Psychology
Angela.Francke@tu-dresden.de

Juliane Anke
Chair of Transportation Psychology
Juliane.Anke@tu-dresden.de

Stefan Huber
Chair of Transportation Ecology
Stefan.Huber1@tu-dresden.de