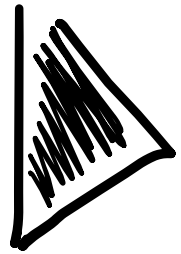
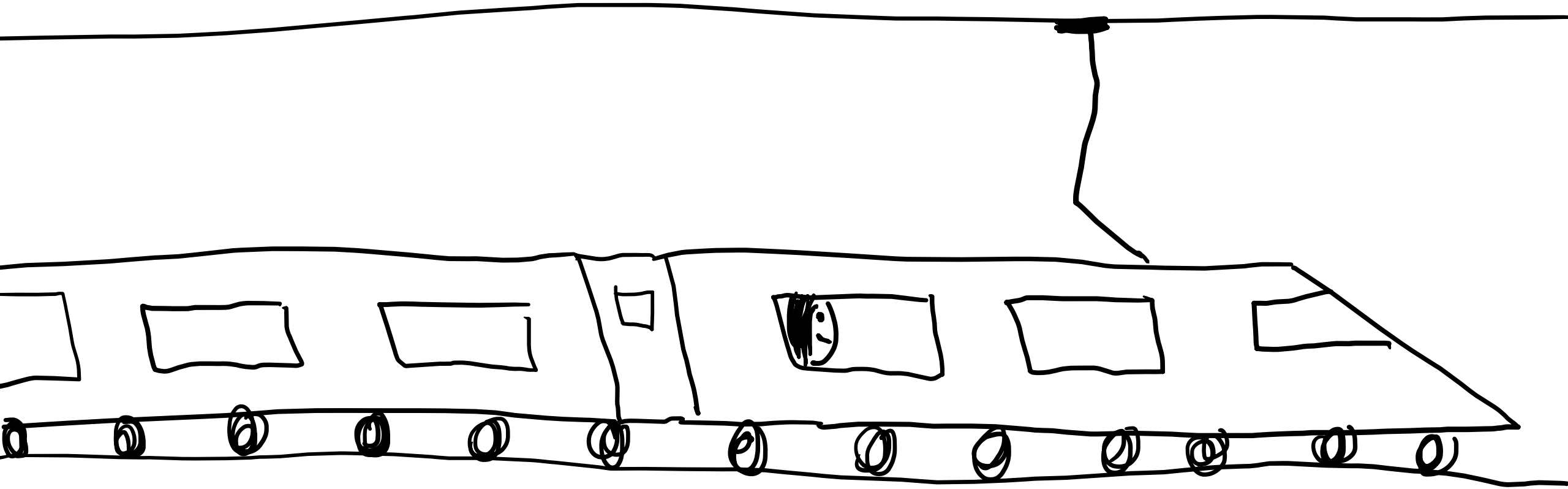
Differential privacy
— when, why and how?

CryptoParty
26/11 2021

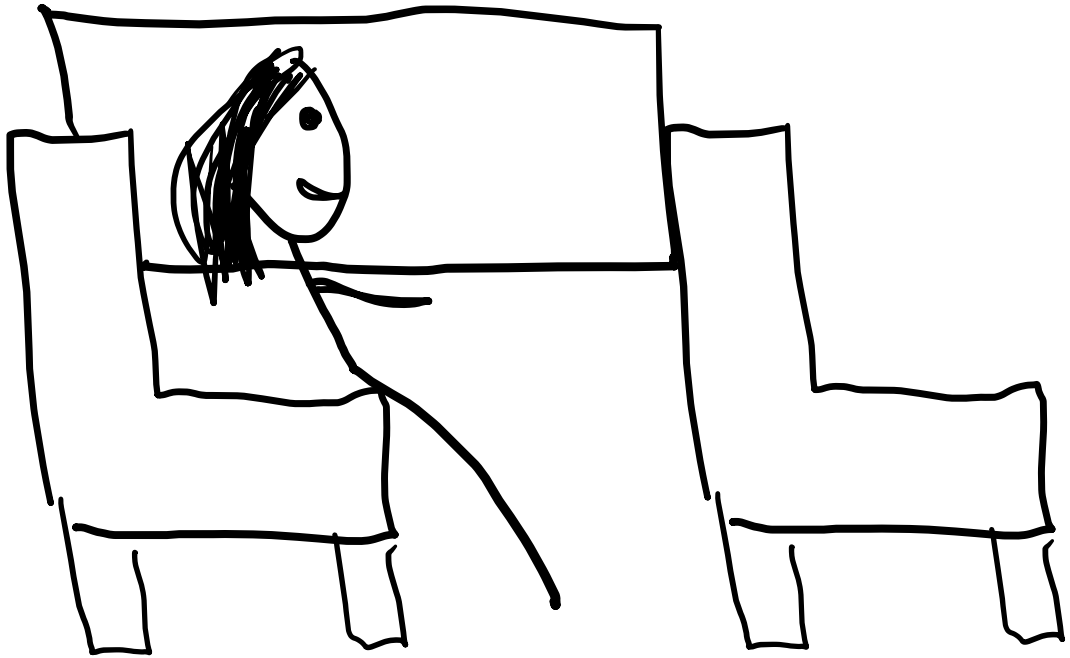
Boel Nelson



Once upon a train ride...

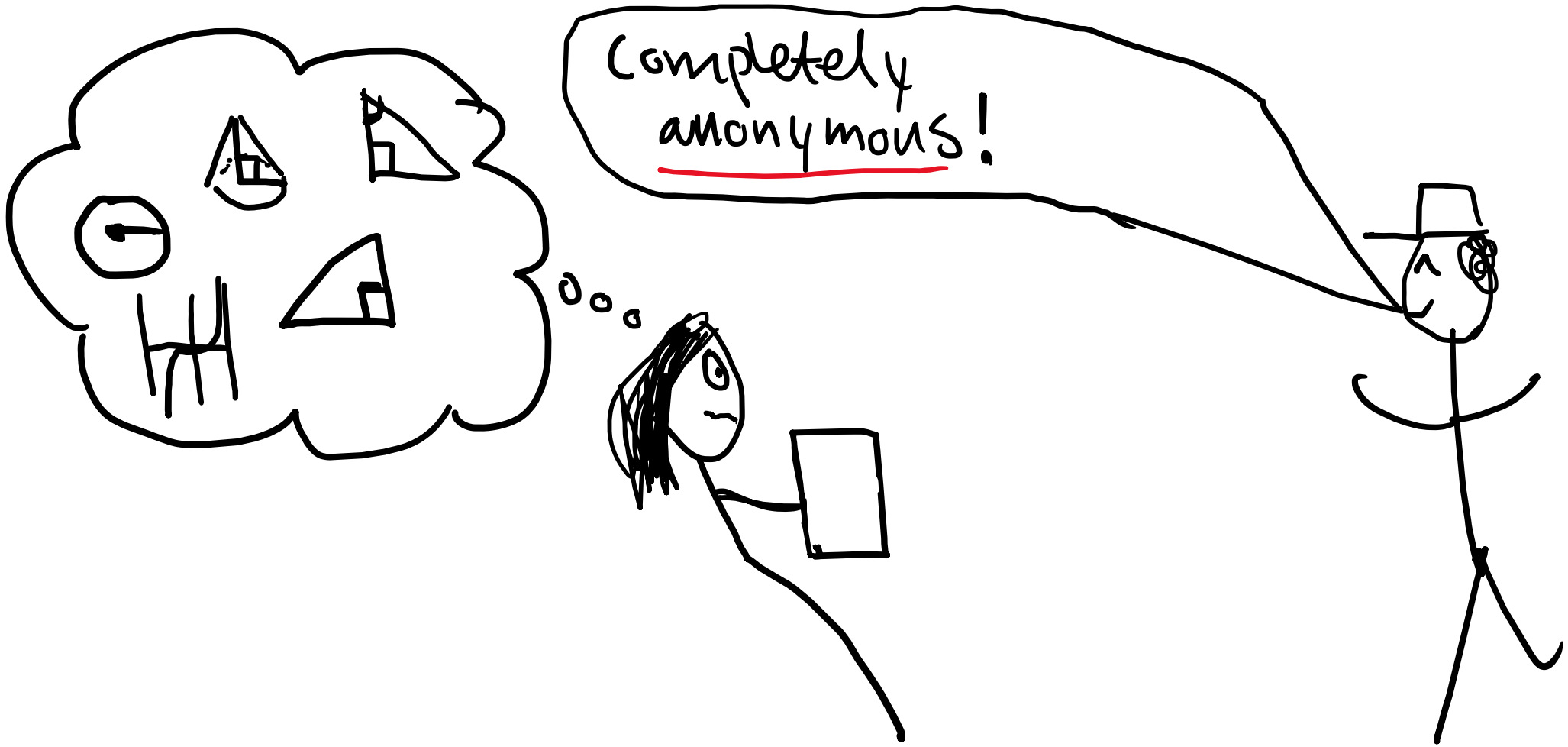


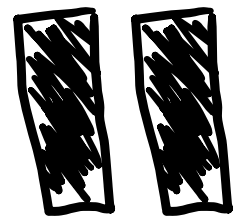
When suddenly...

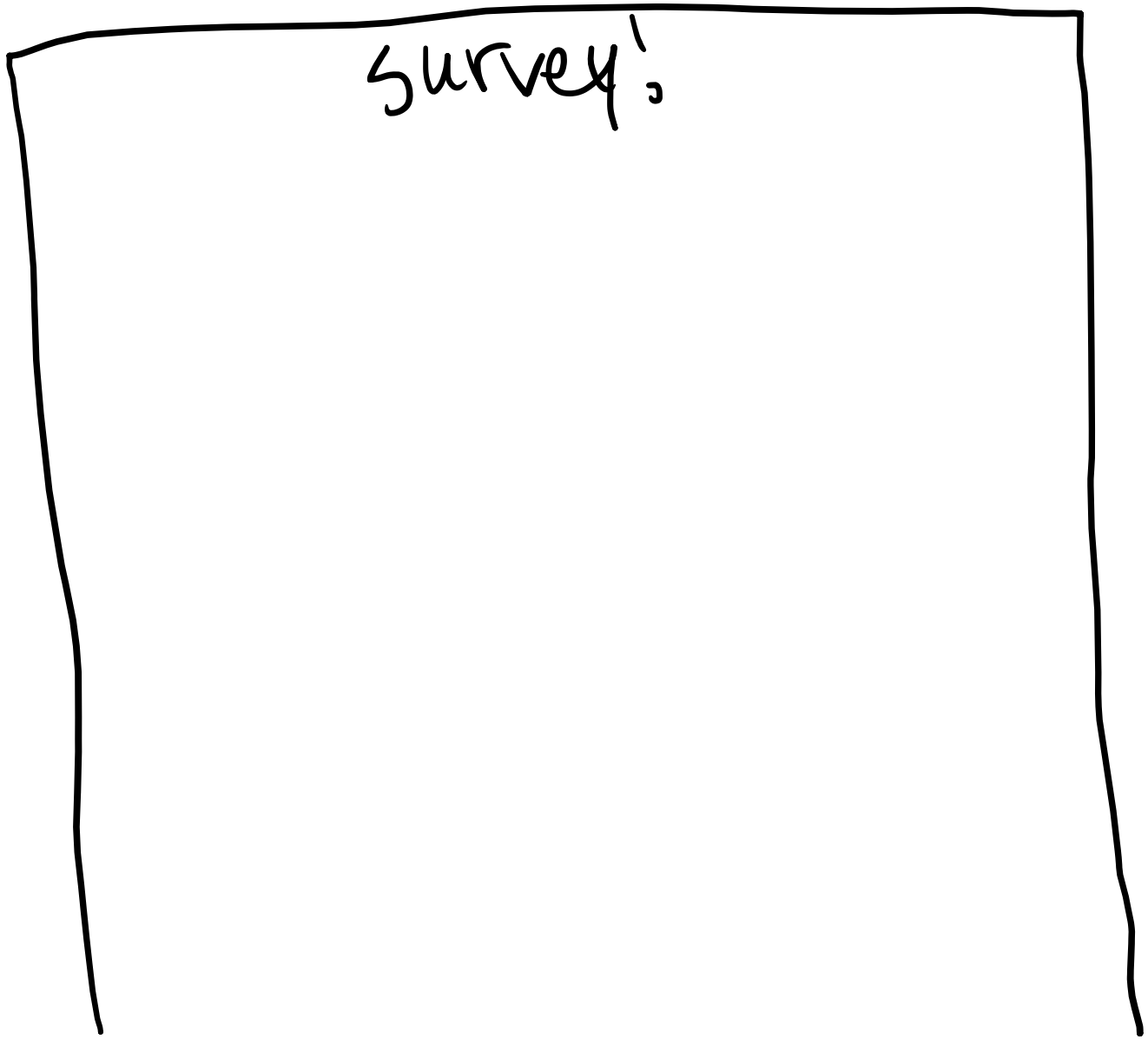



Wild
survey
appears


Completely anonymous!


Completely anonymous!




survey!

Survey:

- Source
- Destination
- Purpose of trip
- Travel frequency
- Sex
- Age
- ZIP code

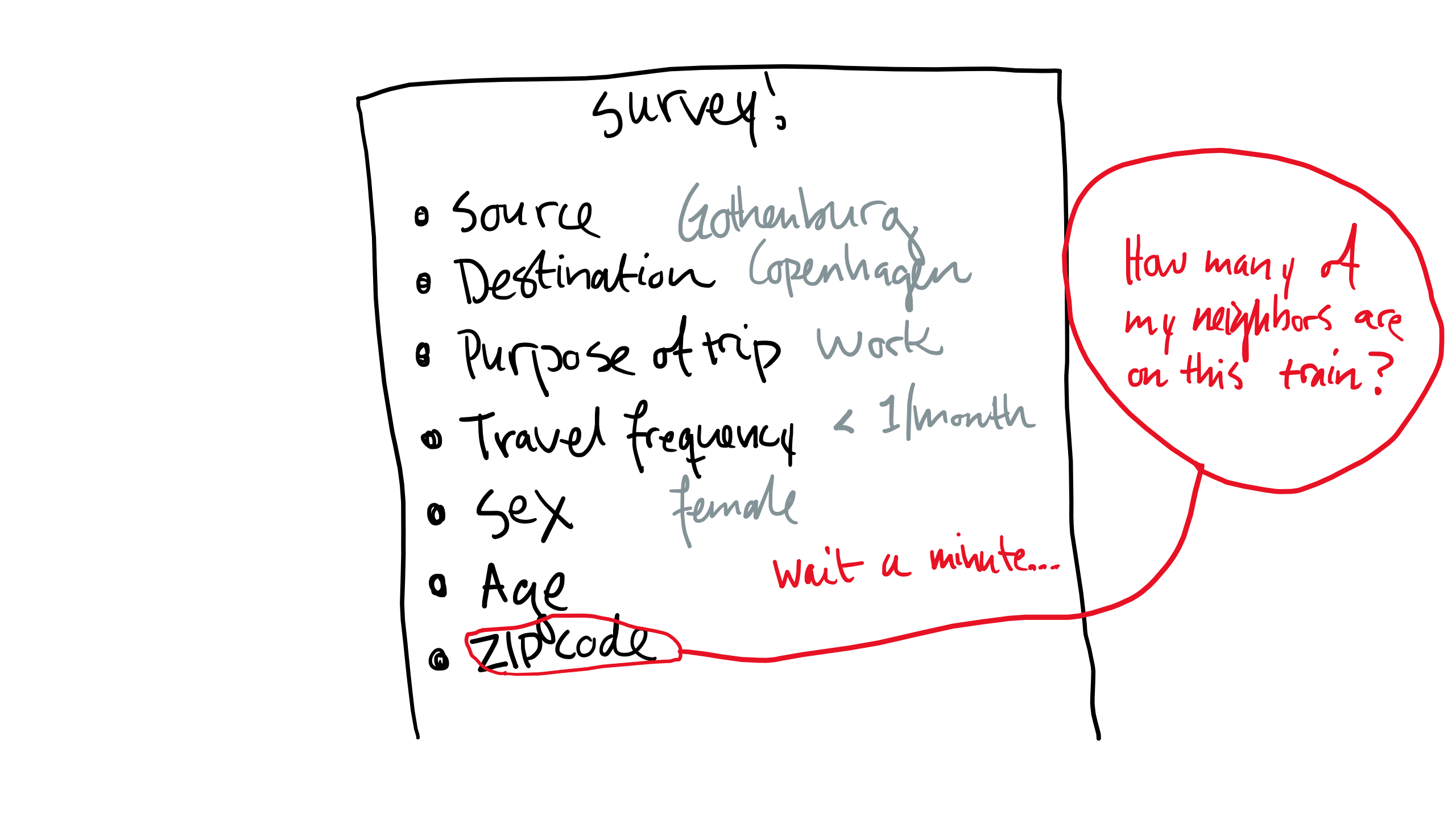
Survey:

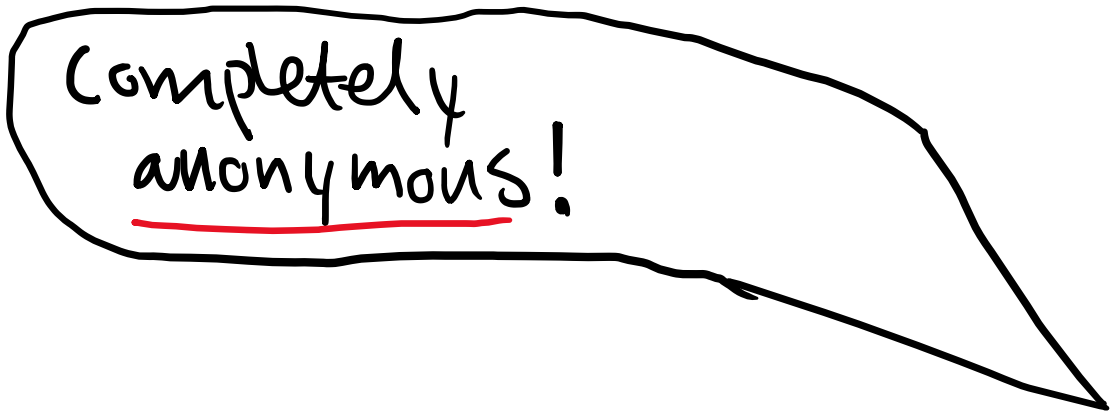
- Source — Gothenburg
- Destination — Copenhagen
- Purpose of trip — work
- Travel frequency — < 1/month
- Sex — female
- Age
- ZIP code

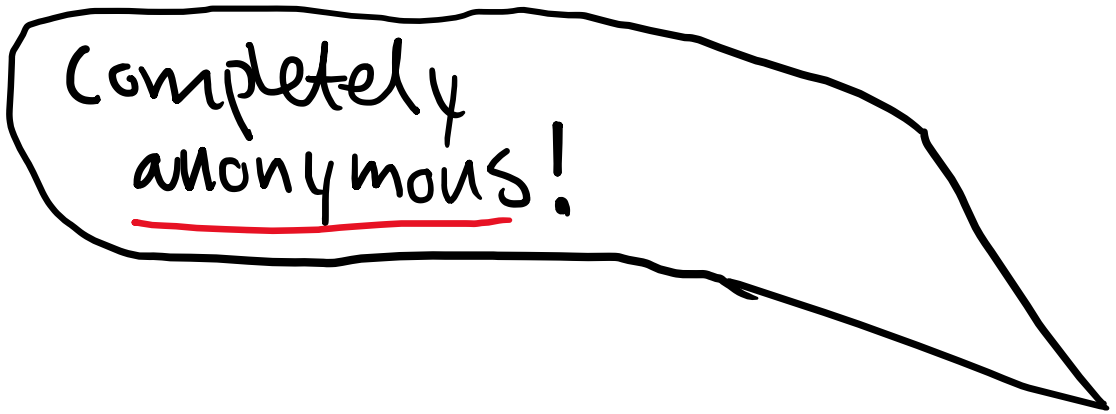
Survey:

- Source — Gothenburg
- Destination — Copenhagen
- Purpose of trip — work
- Travel frequency — < 1/month
- Sex — female
- Age
- ZIP code

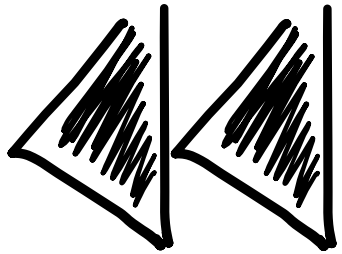
Wait a minute...


survey:
Source
Gothenburg
Destination
Copenhagen
Purpose of trip
work
Travel frequency
< 1/month
Sex
female
Age
ZIP code
wait a minute...
How many of my neighbors are on this train?


Completely
anonymous!



but how?



Anonymization

Gov. of Massachusetts

Gov. of Massachusetts

Medical data

Gov. of Massachusetts

Medical data

SSN
(personnr)

Gov. of Massachusetts

Medical data

SSN Name

(personnr)

Gov. of Massachusetts

Medical data

SSN Name Ethnicity

(personnr)

Gov. of Massachusetts

Medical data

SSN Name Ethnicity DoB

(personnr) (dateofbirth)

Gov. of Massachusetts

Medical data

SSN (personnr) Name Ethnicity DoB (dateofbirth) Sex

Gov. of Massachusetts

Medical data

SSN (personnr) | Name | Ethnicity | DoB (dateofbirth) | Sex | ZIPcode (postnr)

Gov. of Massachusetts

Medical data

SSN (personnr) | Name | Ethnicity | DoB (dateofbirth) | Sex | ZIPcode (postnr) | Marital status

Gov. of Massachusetts

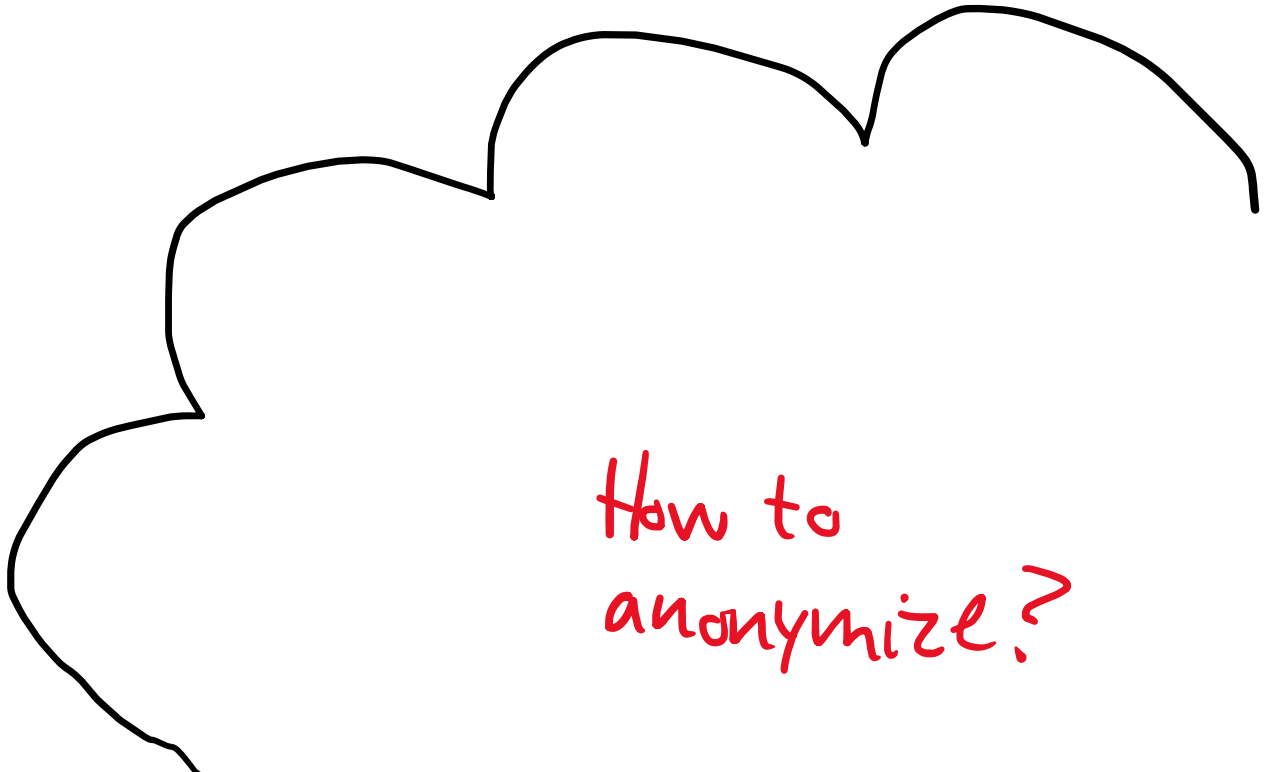
Medical data →

SSN (personnr)	Name	Ethnicity	DoB (dateofbirth)	Sex	ZIPcode (postnr)	Marital status	Problem

Gov. of Massachusetts

Medical data

SSN (personnr) | Name | Ethnicity | DoB (dateofbirth) | Sex | ZIPcode (postnr) | Marital status | Problem



Gov. of Massachusetts

Medical data →

SSN (personnr)	Name	Ethnicity	DoB (dateofbirth)	Sex	ZIPcode (postnr)	Marital status	Problem

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Gov. of Massachusetts

Medical data →

[redacted] (personnr) | [redacted] | Ethnicity | DoB (date of birth) | Sex | ZIPcode (postnr) | Marital status | Problem

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Gov. of Massachusetts

Medical data →

[redacted] (personnr) | [redacted] | Ethnicity | DoB (date of birth) | Sex | ZIPcode (postnr) | Marital status | Problem

Enter: background data

Voter registration list

Gov. of Massachusetts

Medical data →

[redacted] (personnr) | [redacted] | Ethnicity | DoB (dateofbirth) | Sex | ZIPcode (postnr) | Marital status | Problem

Enter: background data

Voter registration list

Name

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Voter registration list

Name Address

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Voter registration list

Name Address City

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

([redacted]: personnr) (DoB: date of birth) (ZIPcode: postnr)

Enter: background data

Voter registration list

Name Address City Party

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Voter registration list

Name Address City Party DoB

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

([redacted]: personnr) (DoB: dateofbirth) (ZIPcode: postnr)

Enter: background data

Voter registration list

Name Address City Party DoB Sex

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Voter registration list

Name Address City Party DoB Sex ZIPcode

Gov. of Massachusetts

Medical data →

[redacted] [redacted] Ethnicity DoB Sex ZIPcode Marital status Problem

(personnr) (dateofbirth) (postnr)

Enter: background data

Voter registration list

Name Address City Party DoB Sex ZIPcode

Gov. of Massachusetts

Medical data →

DoB (date of birth) Sex ZIPcode (post nr)

Problem

87% of US population uniquely identified by {DoB, sex, ZIP}

[Sweeney]

Name Address City Party DoB Sex ZIPcode

Gov. of Massachusetts

Medical data →

DoB (date of birth) Sex ZIPcode (post nr)

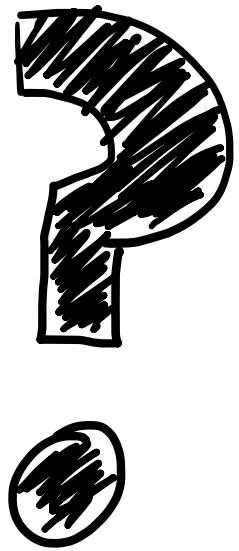
Problem

87% of US population uniquely identified by {DoB, sex, ZIP}

[Sweeney]

Name Address City Party DoB Sex ZIPcode

William Weld



Unfortunately, it keeps happening...

Unfortunately, it keeps happening...

2008 IEEE Symposium on Security and Privacy

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Netflix + IMDb

Unfortunately, it keeps happening...

2008 IEEE Symposium on Security and Privacy

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Netflix + IMDb

DE GRUYTER OPEN Proceedings on Privacy Enhancing Technologies ; 2016 (1):34–51

Miro Enev*, Alex Takakuwa, Karl Koscher, and Tadayoshi Kohno

Automobile Driver Fingerprinting

In-car sensors

Unfortunately, it keeps happening...

2008 IEEE Symposium on Security and Privacy

Robust De-anonymization of Large Sparse Datasets

Arvind Narayanan and Vitaly Shmatikov
The University of Texas at Austin

Netflix + IMDb

DE GRUYTER OPEN Proceedings on Privacy Enhancing Technologies ; 2016 (1):34–51

Miro Enev*, Alex Takakuwa, Karl Koscher, and Tadayoshi Kohno

Automobile Driver Fingerprinting

In-car sensors

De-anonymizing Web Browsing Data with Social Networks

Jessica Su
Stanford University
jtysu@stanford.edu

Ansh Shukla
Stanford University
anshukla@stanford.edu

Sharad Goel
Stanford University
scgoel@stanford.edu

Arvind Narayanan
Princeton University
arvindn@cs.princeton.edu

Twitter links

Proposed defenses

Proposed defenses

- Inference control [1983]

Proposed defenses

- Inference control [1983]
- ⋮
- Differential privacy [2006]

Inference control

Inference control

Example query

Inference control

Example query

```
SELECT * FROM employees
```

Inference control

Example query

```
SELECT * FROM employees
WHERE sex == female
```

Inference control

Example query

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsen
```

Inference control

Example query

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

Inference control

Example query

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

How many matches?

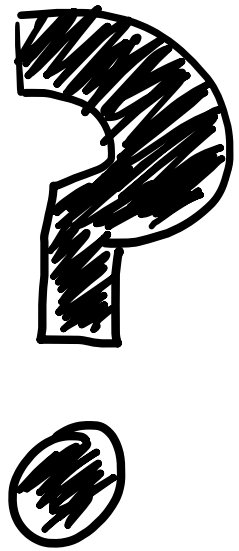
Inference control

Example query

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

How many matches?

Rule:
don't release
if smaller than
k



Intersection attacks

Intersection attacks

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

= 1

Intersection attacks

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

$= n - 1$

Intersection attacks

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

$= n - 1$

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

$= n$

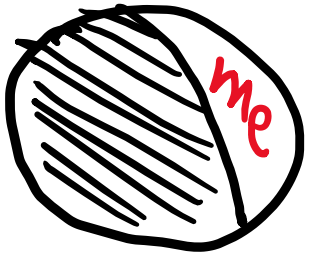
Intersection attacks

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

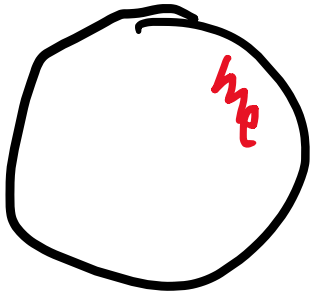
= n - 1

```
SELECT * FROM employees
WHERE sex == female AND
      group == logsem AND
      role == postdoc
```

= n



∩

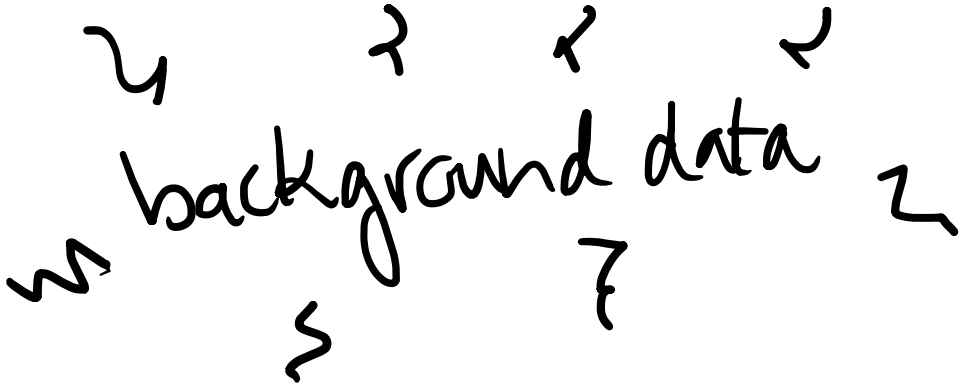


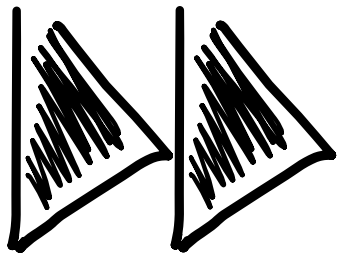
= me

Proposed defenses

- ~~Inference control [1963]~~
- ⋮
- Differential privacy [2006]

Everyone's historical weakness:





2006

Differential privacy

Differential privacy

What?

Differential privacy

What?

- Property of algorithm

Differential privacy

What?

- Property of algorithm
- Quantifies "privacy loss", ϵ

Differential privacy

What?

- Property of algorithm
- Quantifies "privacy loss", ε

Ex

An algorithm is differentially private

Differential privacy

What?

- Property of algorithm
- Quantifies "privacy loss", ϵ

Ex

An algorithm is differentially private

The result/output of a query is not differentially private

Differential privacy

Differential privacy

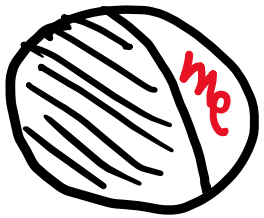
Differential privacy

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsen AND
      role == postdoc
```

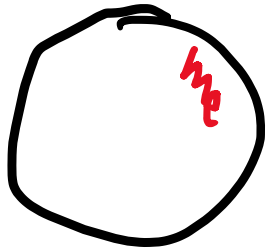
= n - 1

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsen AND
      role == postdoc
```

= n



∩



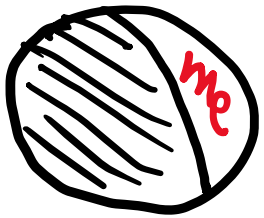
= me

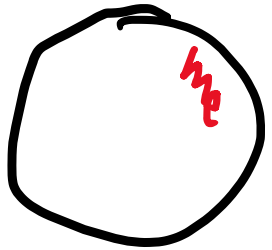
Differential privacy

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

≈

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```



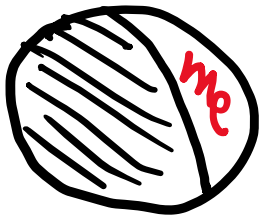


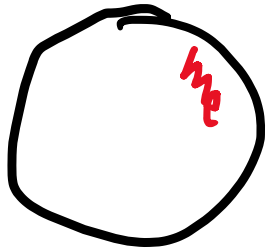
Differential privacy

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

≈

```
SELECT * FROM employees
WHERE [redacted] AND
      group == logsem AND
      role == postdoc
```





Originally

"ε-indistinguishability"

Differential privacy

Protection:

Differential privacy

Protection: essentially, OPT-IN and OPT-OUT are the same

under a small ε

Differential privacy

Differential privacy

How?

Differential privacy

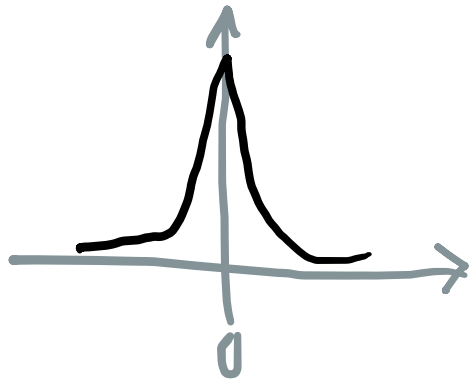
How?

- Laplace mechanism

Differential privacy

How?

- Laplace mechanism



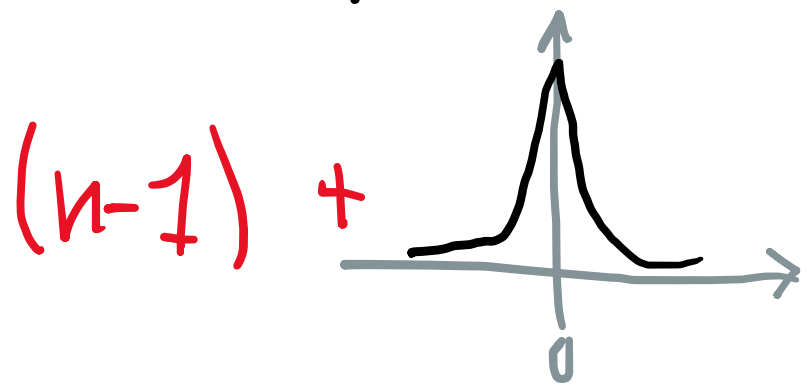
Differential privacy

How?

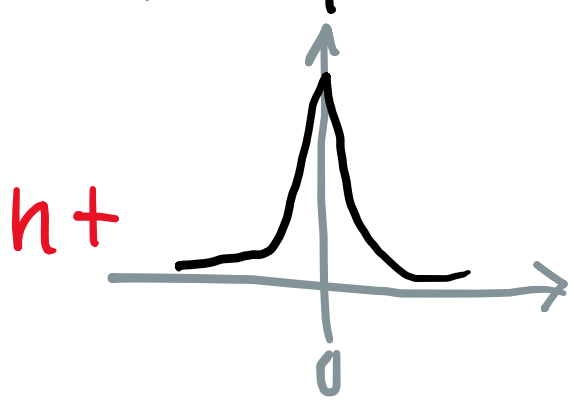
- Laplace mechanism

```
SELECT * FROM employees
WHERE sex != female AND
      group == logsem AND
      role == postdoc
```

= n - 1




```
SELECT * FROM employees
WHERE [sex != female AND] = n
      group == logsem AND
      role == postdoc
```



Differential privacy

How?

- Laplace mechanism

```
SELECT * FROM employees
WHERE sex != female AND
    group == logsen AND
    role == postdoc
```

= n - 1

```
SELECT * FROM employees
WHERE sex = female AND
    group == logsen AND
    role == postdoc
```

= n

Differential privacy

How?

- Laplace mechanism
- Randomized response

Differential privacy

How?

- Laplace mechanism
- Randomized response

Was the train crew accommodating?

Yes / No

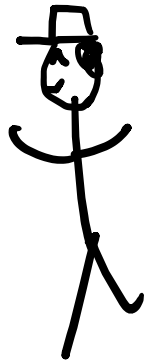
Differential privacy

How?

- Laplace mechanism
- Randomized response

Was the train crew accommodating?

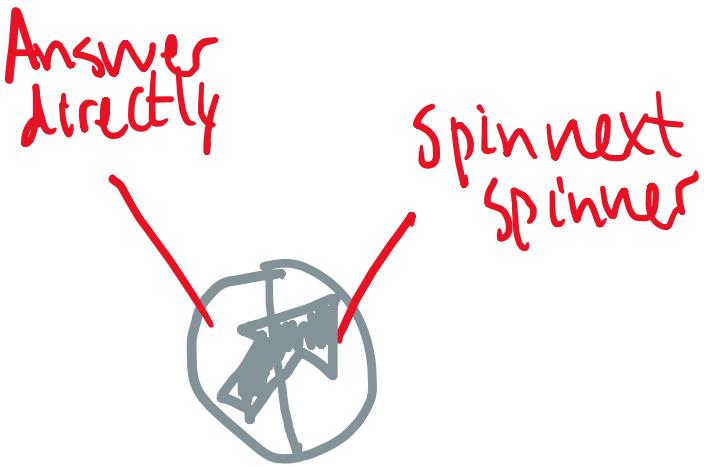
Yes / No

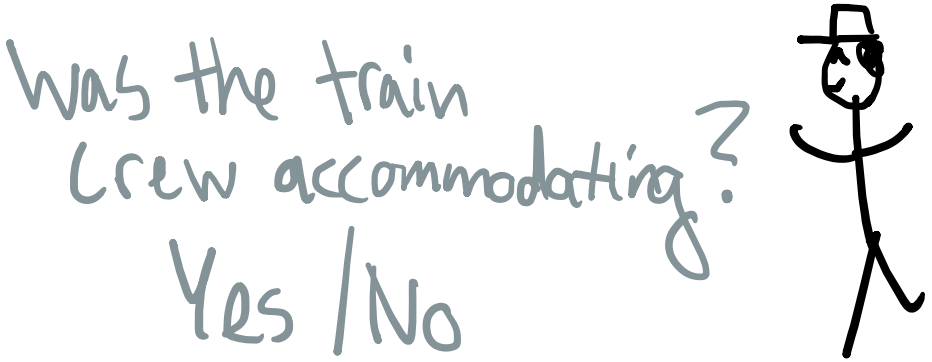


Differential privacy

How?

- Laplace mechanism
- Randomized response

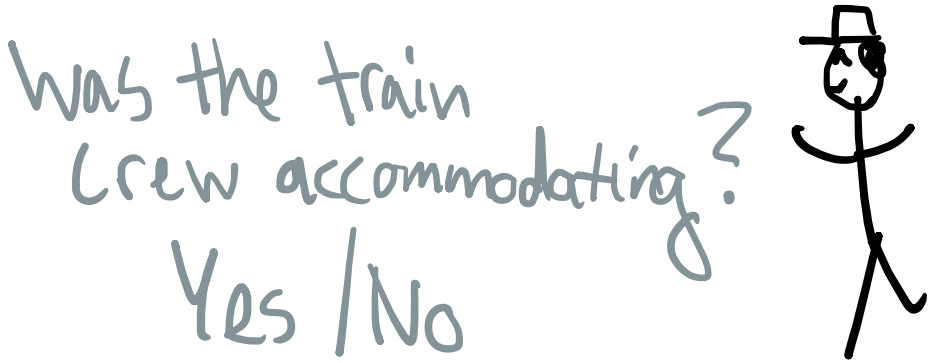


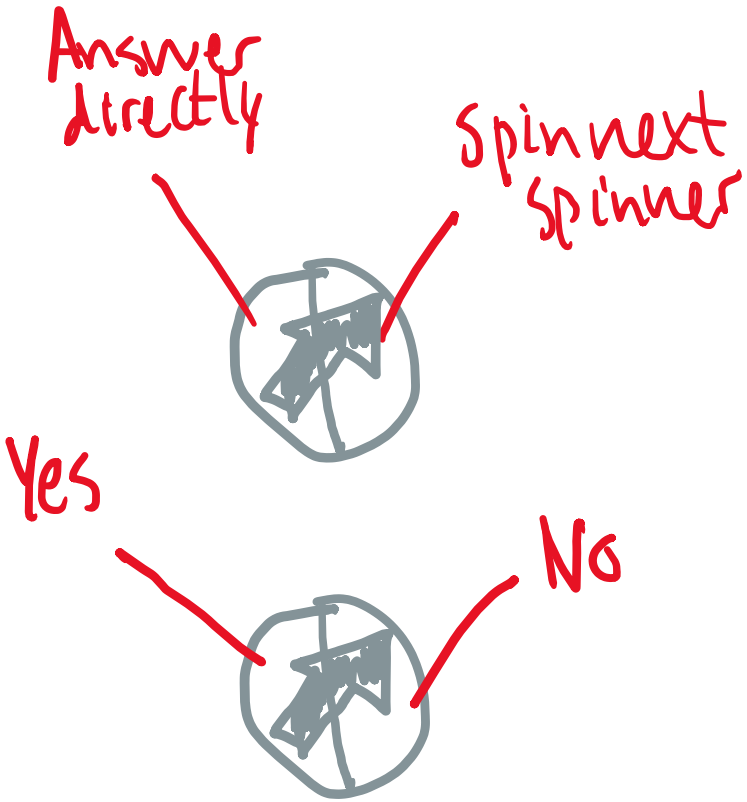


Differential privacy

How?

- Laplace mechanism
- Randomized response





Differential privacy

How?

- Laplace mechanism
- Randomized response

Differential privacy

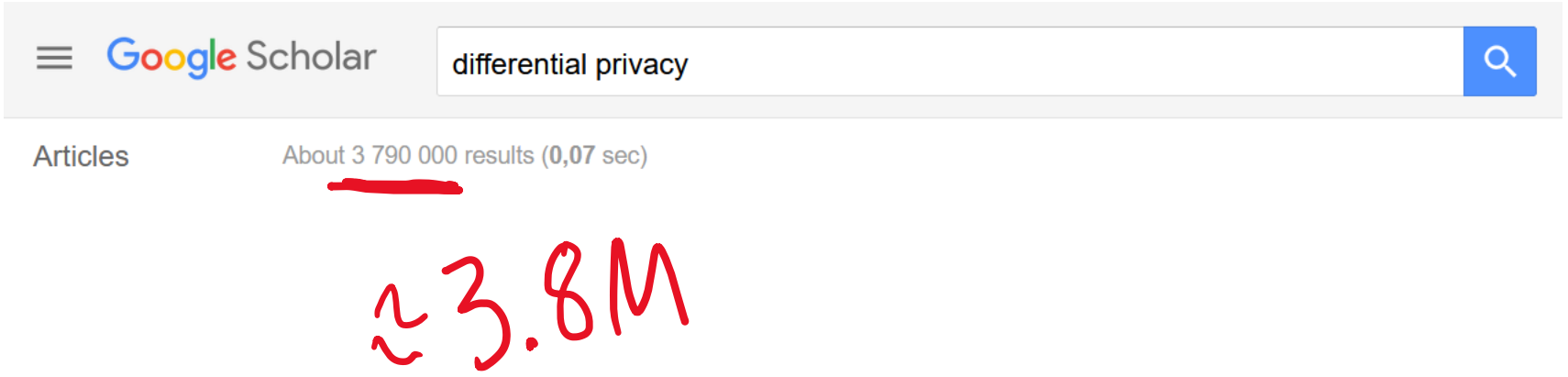
How?

- Laplace mechanism
- Randomized response
- Sooo many more!

Differential privacy

How?

- Laplace mechanism
- Randomized response
- Sooo many more!



≈3.8M

Bua...

But...

does anyone use it?

Real world use cases

Real world use cases

- Google Chrome [RAPPOR, deprecated]

Real world use cases

- Google Chrome [RAPPOR, deprecated]
- Apple iOS

Real world use cases

- Google Chrome [RAPPOR, deprecated]
- Apple iOS
- Uber [Chorus]

Real world use cases

- Google Chrome [RAPPOR, deprecated]
- Apple iOS
- Uber [Chorus]
- Microsoft since Win10

Real world use cases

- Google Chrome [RAPPOR, deprecated]
- Apple iOS
- Uber [Chorus]
- Microsoft since Win10
- US 2020 Census ← ≈"Statistiska centralbyrån" survey data!

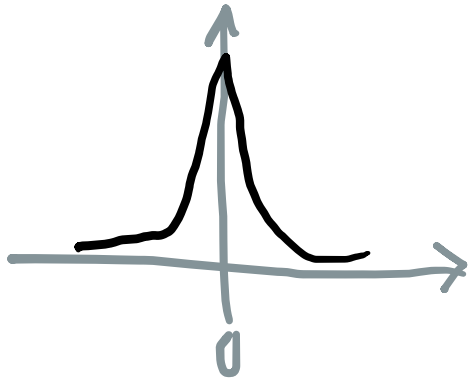
Why don't everyone use differential privacy?

Why don't everyone use differential privacy?

- Privacy "costs"

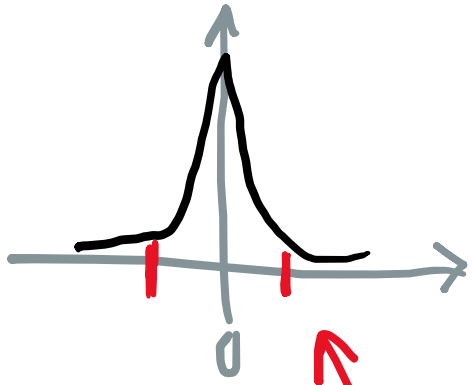
Why don't everyone use differential privacy?

- Privacy "costs"



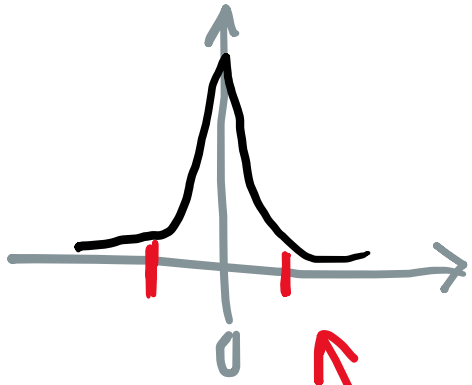
Why don't everyone use differential privacy?

- Privacy "costs"



Why don't everyone use differential privacy?

- Privacy "costs"



what if it's still big?

Ex. 10 + 853

Why don't everyone use differential privacy?

- Privacy "costs"

Why don't everyone use differential privacy?

- Privacy "costs"
- Community still struggle to interpret ε

Why don't everyone use differential privacy?

- Privacy "costs"
- Community still struggle to interpret ε

There's no silver bullet...

Take away

Take away

- Differential privacy is strong:

Take away

- Differential privacy is strong:
  - provable privacy

Take away

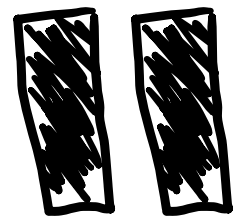
- Differential privacy is strong:
  - provable privacy
  - resistant to background data

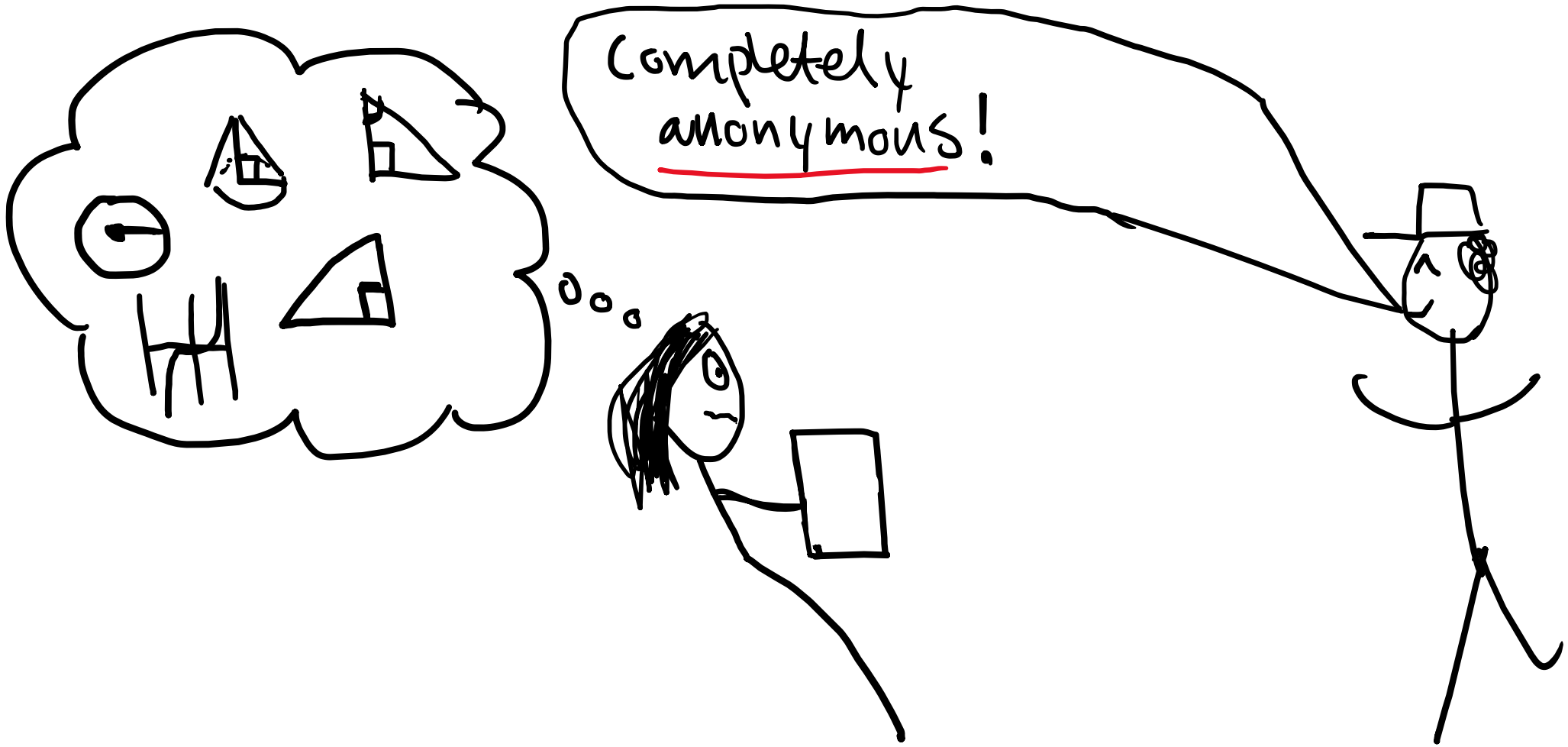
Take away

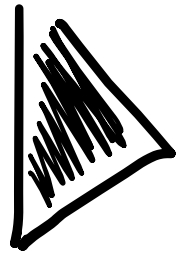
- Differential privacy is strong:
  - provable privacy
  - resistant to background data

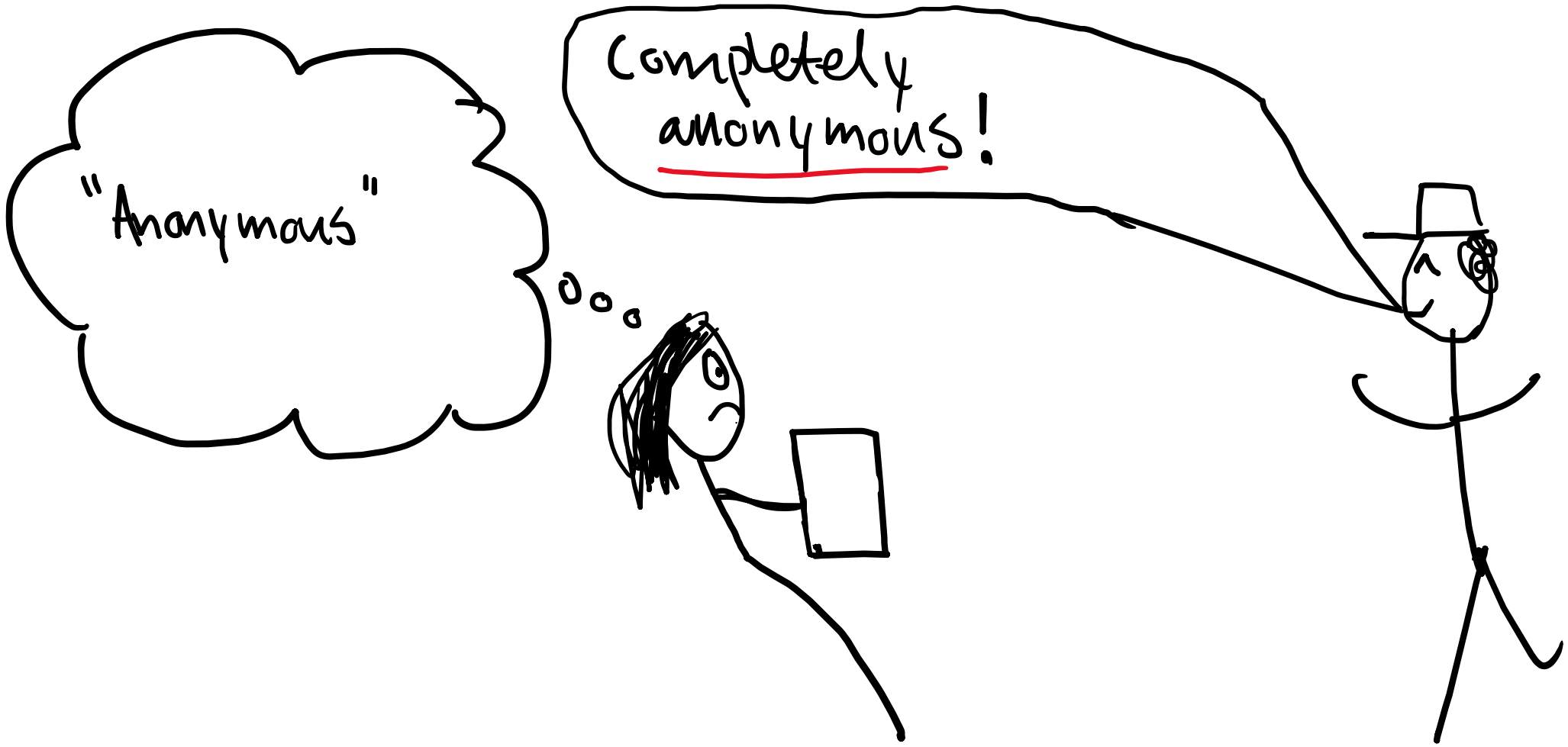
But: it's difficult...

- each use case needs tuning

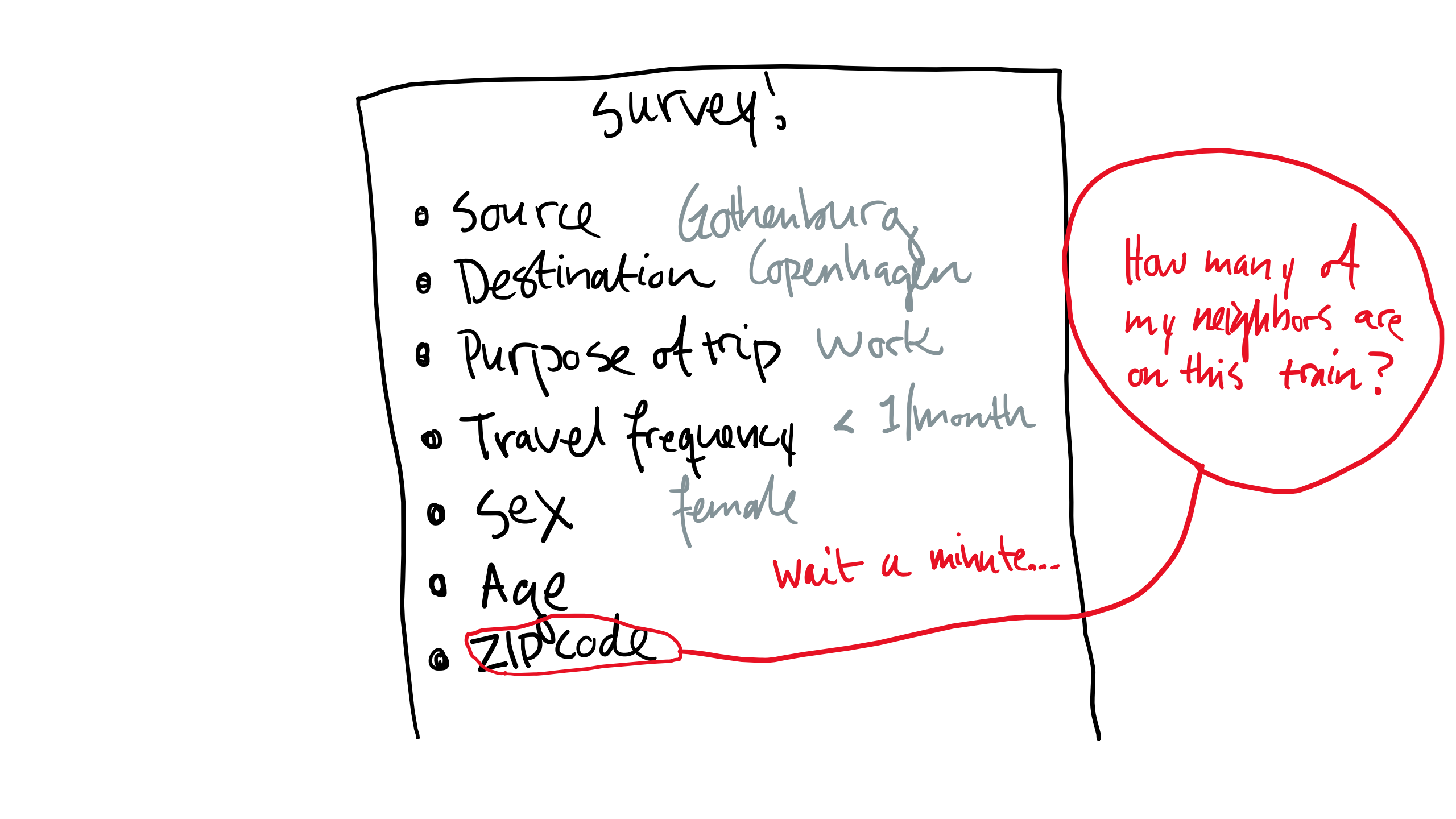



Completely anonymous!




"Anonymous"
Completely anonymous!

—"Am I unique?"


survey:
Source
Gothenburg
Destination
Copenhagen
Purpose of trip
work
Travel frequency
< 1/month
Sex
female
Age
ZIP code
wait a minute...
How many of my neighbors are on this train?

Background data

SJ has my ticket information ...

Survey:

- Source Gothenburg
- Destination Copenhagen
- Purpose of trip Work
- Travel frequency < 1/month
- Sex female
- Age
- ZIP code

How many of my neighbors are on this train?

Wait a minute...

Did I answer the survey?

Did I answer the survey?

= would you? 

What can we do?

What can we do?

- Be curious :)

What can we do?

- Be curious :)
- Ask!

What can we do?

- Be curious :)
- Ask!
  - How do you anonymize the data?
  - What privacy guarantees do I get?

What can we do?

- Be curious :)
- Ask!
  - How do you anonymize the data?
  - What privacy guarantees do I get?

If they can't or won't answer... Beware!

What can we do?

- Be curious :)
- Ask!
  - How do you anonymize the data?
  - What privacy guarantees do I get?

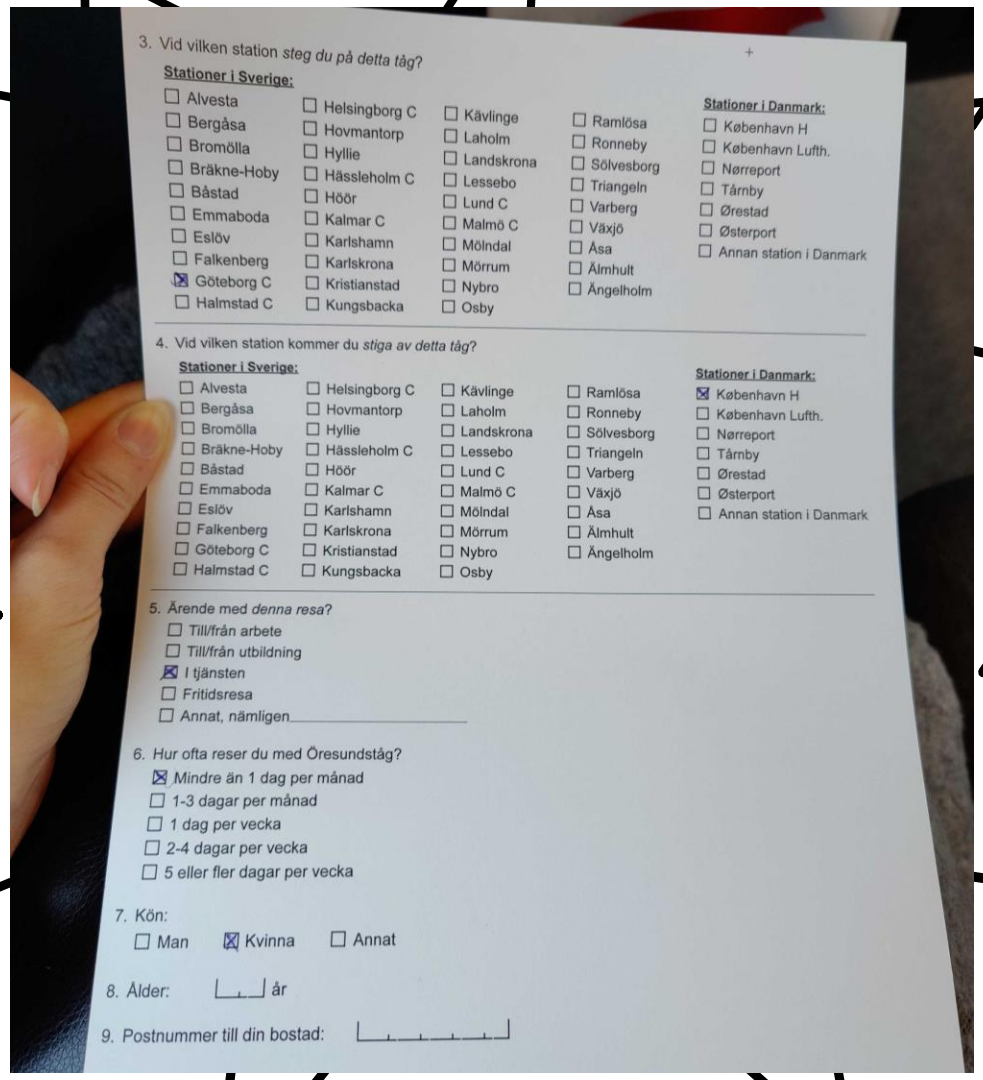
(they probably won't reply 'differential privacy')

If they can't or won't answer... Beware!

Thanks for listening! 

boel@cs.au.dk

starring: The survey



3. Vid vilken station *steg du på detta tåg?*

Stationer i Sverige:

- ☐ Alvesta
- ☐ Bergåsa
- ☐ Bromölla
- ☐ Bräkne-Hoby
- ☐ Båstad
- ☐ Emmaboda
- ☐ Eslöv
- ☐ Falkenberg
- ☒ Göteborg C
- ☐ Halmstad C
- ☐ Helsingborg C
- ☐ Hovmantorp
- ☐ Hyllie
- ☐ Hässleholm C
- ☐ Höör
- ☐ Kalmar C
- ☐ Karlshamn
- ☐ Karlskrona
- ☐ Kristianstad
- ☐ Kungsbacka
- ☐ Kävlinge
- ☐ Laholm
- ☐ Landskrona
- ☐ Lessebo
- ☐ Lund C
- ☐ Malmö C
- ☐ Mölndal
- ☐ Mörrum
- ☐ Nybro
- ☐ Osby
- ☐ Ramlösa
- ☐ Ronneby
- ☐ Sölvesborg
- ☐ Triangeln
- ☐ Varberg
- ☐ Växjö
- ☐ Åsa
- ☐ Älmhult
- ☐ Ängelholm

Stationer i Danmark:

- ☐ København H
- ☐ København Lufth.
- ☐ Nørreport
- ☐ Tårnby
- ☐ Ørestad
- ☐ Østerport
- ☐ Annan station i Danmark

4. Vid vilken station kommer du *stiga av detta tåg?*

Stationer i Sverige:

- ☐ Alvesta
- ☐ Bergåsa
- ☐ Bromölla
- ☐ Bräkne-Hoby
- ☐ Båstad
- ☐ Emmaboda
- ☐ Eslöv
- ☐ Falkenberg
- ☐ Göteborg C
- ☐ Halmstad C
- ☐ Helsingborg C
- ☐ Hovmantorp
- ☐ Hyllie
- ☐ Hässleholm C
- ☐ Höör
- ☐ Kalmar C
- ☐ Karlshamn
- ☐ Karlskrona
- ☐ Kristianstad
- ☐ Kungsbacka
- ☐ Kävlinge
- ☐ Laholm
- ☐ Landskrona
- ☐ Lessebo
- ☐ Lund C
- ☐ Malmö C
- ☐ Mölndal
- ☐ Mörrum
- ☐ Nybro
- ☐ Osby
- ☐ Ramlösa
- ☐ Ronneby
- ☐ Sölvesborg
- ☐ Triangeln
- ☐ Varberg
- ☐ Växjö
- ☐ Åsa
- ☐ Älmhult
- ☐ Ängelholm

Stationer i Danmark:

- ☒ København H
- ☐ København Lufth.
- ☐ Nørreport
- ☐ Tårnby
- ☐ Ørestad
- ☐ Østerport
- ☐ Annan station i Danmark

5. *Ärende med denna resa?*

- ☐ *Till/från arbete*
- ☐ *Till/från utbildning*
- ☒ *I tjänsten*
- ☐ *Fritidsresa*
- ☐ *Annat, nämligen*_____

6. *Hur ofta reser du med Öresundståg?*

- ☒ *Mindre än 1 dag per månad*
- ☐ *1-3 dagar per månad*
- ☐ *1 dag per vecka*
- ☐ *2-4 dagar per vecka*
- ☐ *5 eller fler dagar per vecka*

7. *Kön:*

☐ Man ☒ Kvinna ☐ Annat

8. Ålder: __ år

9. Postnummer till din bostad: _____