Measuring Conditional Anonymity — A Global Study

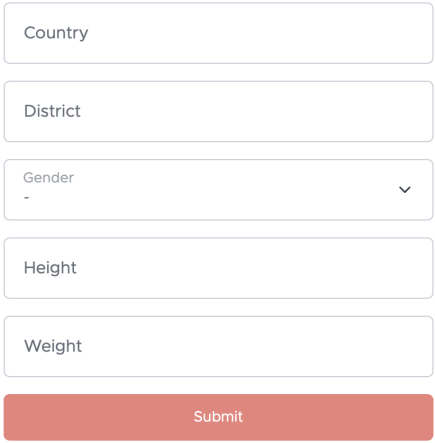
Pascal Berrang, **Paul Gerhart**, and Dominique Schröder

UNIVERSITY OF BIRMINGHAM

FAU

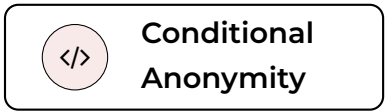
TU WIEN

Prologue: Modern Health App



- Your data is safe
- We do not store personalized data
- EU GDPR
- California Consumer Privacy Act
- Secure by Design
- HTTPS protocol encrypted with TLS (Transport Layer Security)

Our Contribution



Starting Point: Matching Attacks

Anonymized dataset containing **confidenital information**

Zip	Age	Sex	Confidential
15XX	70-75	F	…
12XX	25-30	M	…
95XX	65-70	F	…
11XX	15-20	M	…
12XX	45-50	F	…
⋮	⋮	⋮	⋮

Unanonimized dataset containing **no confidential information**

Identity	Zip	Age	Sex
Alice	1161	19	F
Bob	1234	27	M
Charly	4854	45	F
Dave	1277	28	M
Eve	9584	68	F
⋮	⋮	⋮	⋮

Adversarial goal: match the databases

Open Question

How can we **access the unanonymized dataset**?

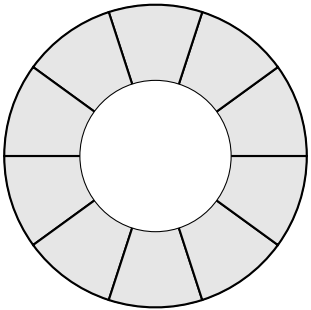
Idea: Conditional Anonymity

- We gather publicly available statistical data.
- Using **population statistics**, we estimate the anonymity set size $\psi(\vec{a})$.
- We refine the set size by each **auxiliary information** we have.
- We define the conditional anonymity set for attributes $\vec{a}$ and $\vec{b}$ via

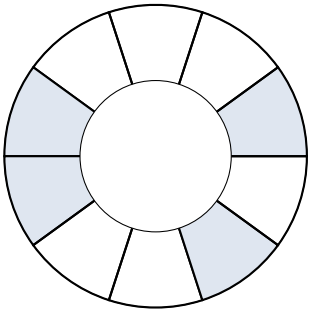
$$\mathsf{A}_{\mathcal{P}}(\vec{a} \mid \vec{b}) = \psi(\vec{a}) \cdot \Pr\left[\vec{b} \mid \vec{a}\right].$$

Conditional Anonymity Sets

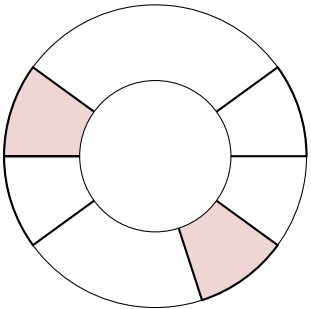
$$\mathsf{A}_{\mathcal{P}}(\vec{a} \mid \vec{b} \mid \vec{c}) = \psi(\vec{a}) \cdot \Pr\left[\vec{b} \mid \vec{a}\right] \cdot \Pr\left[\vec{c} \mid \vec{a} \wedge \vec{b}\right]$$



$\psi(\vec{a}) = 10$



$\Pr\left[\vec{b} \mid \vec{a}\right] = 0.4$



$\Pr\left[\vec{c} \mid \vec{a} \wedge \vec{b}\right] = 0.5$

$$\mathsf{A}_{\mathcal{P}}(\vec{a} \mid \vec{b} \mid \vec{c}) = 10 \cdot 0.4 \cdot 0.5 = 2$$

Comparison To Prior Work

Matching Attacks

Identity	Zip	Age	Sex
Bob	1234	27	M
Dave	1277	28	M
⋮	⋮	⋮	⋮

Use an **exact database**

Counting all matching entries

Conditional Anonymity

Attribute	Value	Perc.	Set
Zip	11**	100 %	**2000**
Age	25 - 30	15%	**300**
Sex	M	50%	**150**

Use of **statistical data**

Compute an abstract group size

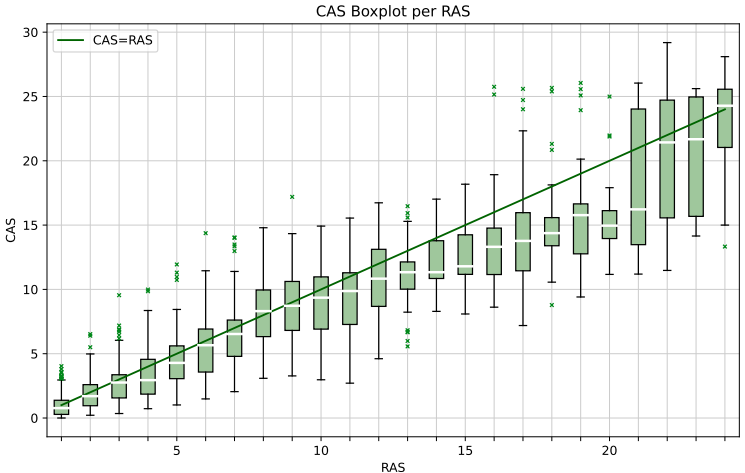
Accuracy of CAS: AnonLand

Accuracy of CAS: AnonLand

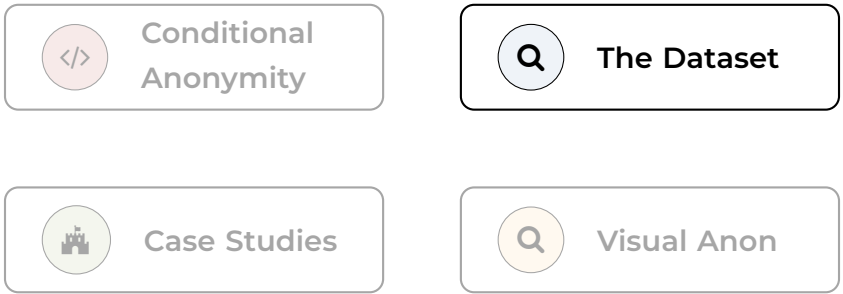
AnonLand has 102.5 million citizens

- The age distribution is modeled using a trapezoidal shape
- Gender is distributed uniformly random
- Height and weight follow a normal distribution
  - 5 metropolises of five million inhabitants each
  - 25 cities of one million inhabitants each
  - 250 counties of 100,000 inhabitants each
  - 2,500 areas of 10,000 inhabitants each
  - 2,500 villages of one thousand inhabitants each

Accuracy of CAS



Our Contribution



Data Request

Paul Gerhart 22. November 2021 at 11:56

Consensus-data request for researching purposes

Hide

Bcc: census.customerservices@ons.gov.uk, Eurostat Helpdesk_EN, User Information Services Stats SA, leosanni@nigerianstat.gov.ng, STATCAN.infostats-infostats.STATCAN@canada.ca, Atencion a Usuarios, ibge@ibge.gov.br, info, Stat, info@stats.gov.cn, ddu.rgi@nic.in, pbs@pbs.gov.pk, client.services@abs.gov.au, info@stats.govt.nz, nstac-info@nstac.go.jp, statistics@un.org, statistics@afdb.org, nfo@tuik.gov.tr, Dominique Schröder, Pascal Berrang

To whom it may concern,
My name is Paul Gerhart, and I am part of a privacy research team at the chair of applied cryptography of the Friedrich-Alexander-University Erlangen-Nürnberg in cooperation with the University of Birmingham.

My team and I are working on a web app to inform people about the anonymity set they are currently living in. That is the number of people who fit in the same data bucket created by several data points one may provide voluntarily without worries. With our app, we want to create awareness of how sensitive personal data is to help people protect their privacy.

Our work is based on the paper Pandemic Privacy by Berrang and Schröder, but we want to stress the insights to a worldwide dataset.

Therefore, we are interested in census data that gives insights into the population count by postcode separated by age groups and sex. Moreover, we are interested in the distribution of height and weight by age group and sex.

Based on this data alone, we cannot deanonymize people, but we can show how anonymity decreases by the publication of personal data that might seem irrelevant.

Hence we were hoping you could provide the desired data for us.

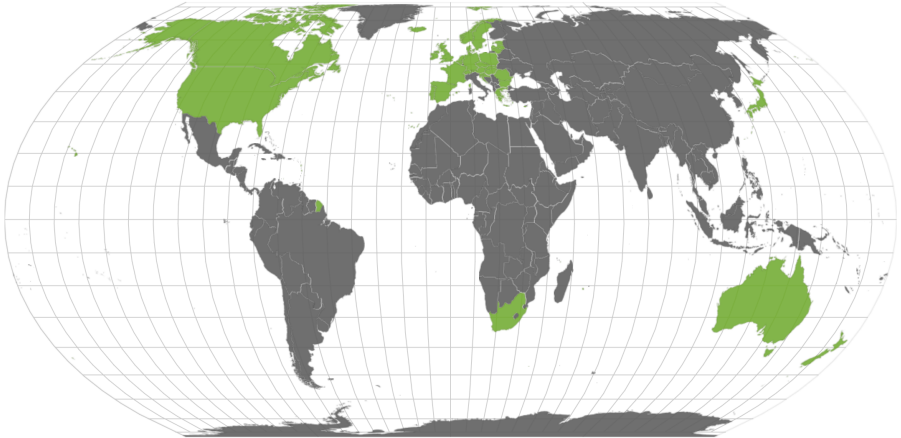
Best regards

Paul Gerhart

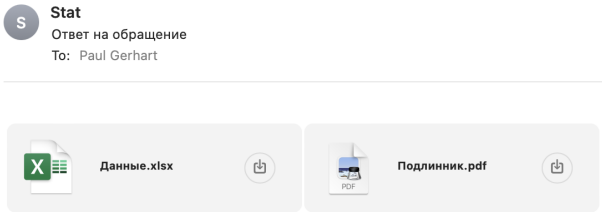
--
Paul Gerhart
paul.gerhart@fau.de
Lehrstuhl für Angewandte Kryptographie
Friedrich-Alexander-Universität Erlangen-Nürnberg

Our Dataset

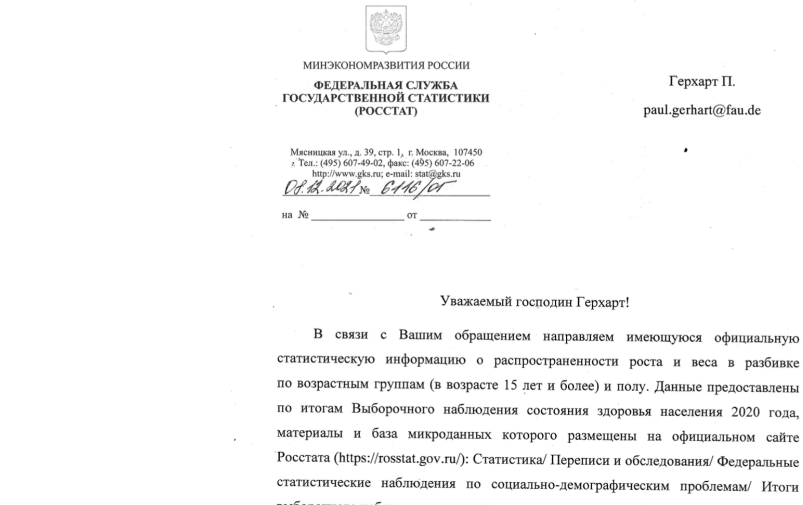
Currently, we can calculate anonymity sets for 1 084 230 346 people.



Data Response I



Data Response II



МИНЭКОНОМРАЗВИТИЯ РОССИИ

ФЕДЕРАЛЬНАЯ СЛУЖБА ГОСУДАРСТВЕННОЙ СТАТИСТИКИ (РОССТАТ)

Мясницкая ул., д. 39, стр. 1, г. Москва, 107450
г. Тел.: (495) 607-49-02, факс: (495) 607-22-06
http://www.gks.ru; e-mail: stat@gks.ru

06.12.2021 № СА1С/01

на № _____ от _____

Герхарт П.

paul.gerhart@fau.de

Уважаемый господин Герхарт!

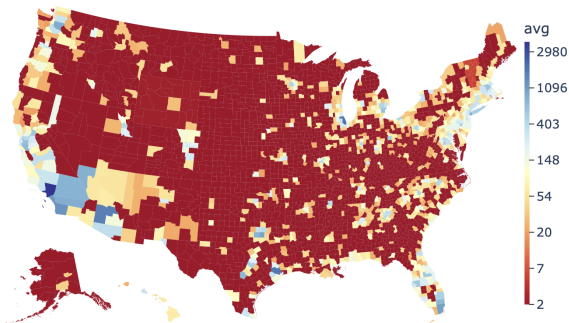
В связи с Вашим обращением направляем имеющуюся официальную статистическую информацию о распространенности роста и веса в разбивке по возрастным группам (в возрасте 15 лет и более) и полу. Данные предоставлены по итогам Выборочного наблюдения состояния здоровья населения 2020 года, материалы и база микроданных которого размещены на официальном сайте Росстата (https://rosstat.gov.ru/): Статистика/ Переписи и обследования/ Федеральные статистические наблюдения по социально-демографическим проблемам/ Итоги выборочного наблюдения состояния здоровья населения.

Our Contribution

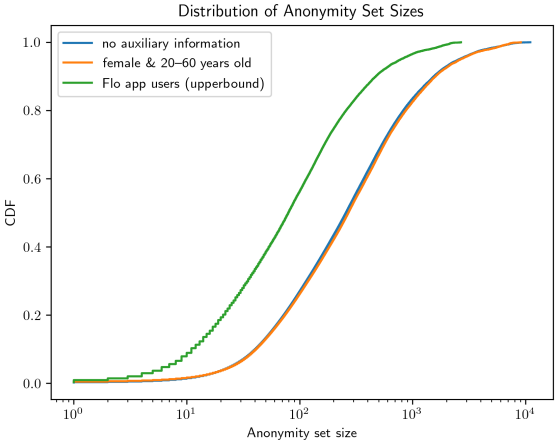
- Conditional Anonymity
- The Dataset
- **Case Studies**
- Visual Anon

Case Study: USA

- Avg. CAS: 77
- Avg. CAS in red area: 2
- Avg. CAS below 5 in 97% of the counties



Case Study: Flo app



- App users avg. CAS: 20
- Avg. CAS **below 5 in 97% of the US counties**

Our Contribution

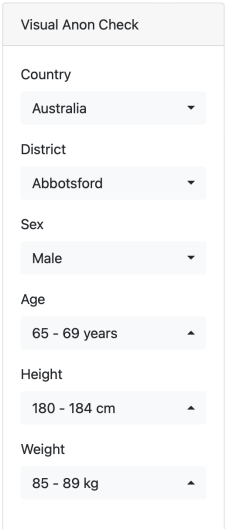
Conditional Anonymity

The Dataset

Case Studies

Visual Anon

Visual Anon



- 26 177 413 people live in Australia
- 7 860 of them in Abbotsford
- 3 981 are male

Visual Anon (Age)

Visual Anon Check

Country

Australia

District

Abbotsford

Sex

Male

Age

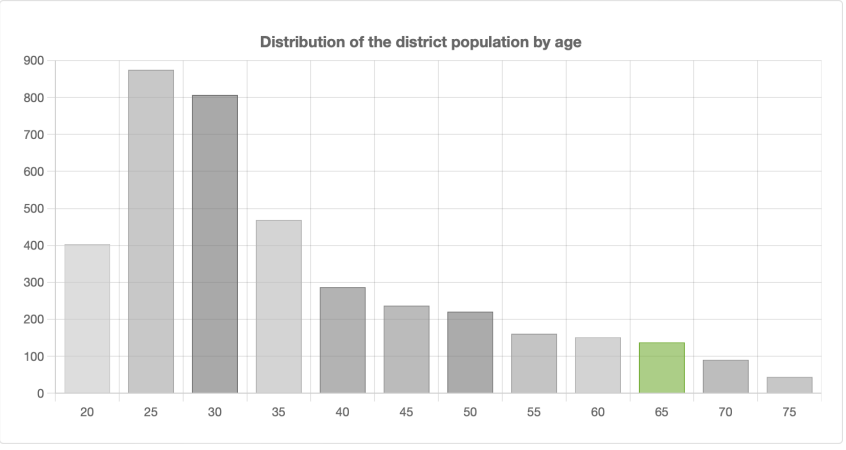
65 - 69 years

Height

180 - 184 cm

Weight

85 - 89 kg



Visual Anon (Height)

Visual Anon Check

Country

Australia

District

Abbotsford

Sex

Male

Age

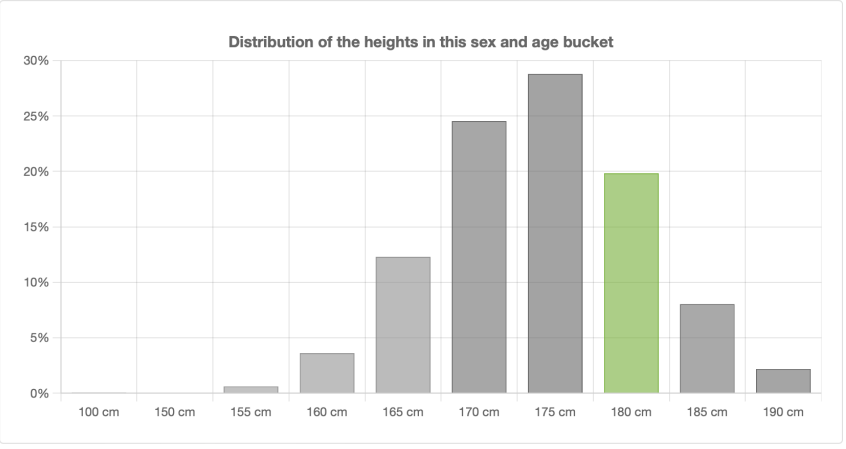
65 - 69 years

Height

180 - 184 cm

Weight

85 - 89 kg



Visual Anon (Weight)

Visual Anon Check

Country

Australia

District

Abbotsford

Sex

Male

Age

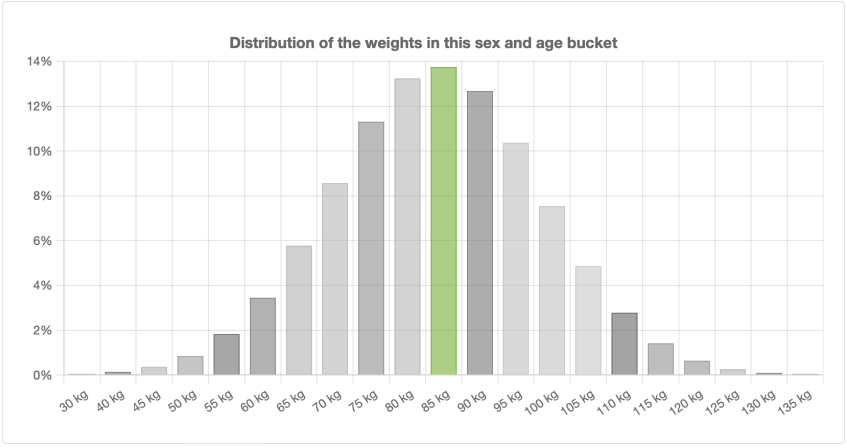
65 - 69 years

Height

180 - 184 cm

Weight

85 - 89 kg

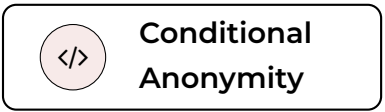


Visual Anon: Result

Providing this data you can be deanonymized up to **3** Persons.

26 177 413 People live in Australia	138 are in the ['65'] years bucket
7 860 of them in Abbotsford	27 of them are ['180'] cm high
3 981 are male	3 of them weight ['85'] kg

Conclusion


Conditional
Anonymity


The Dataset


Case Studies


Visual Anon

Measuring Conditional Anonymity — A Global Study

Pascal Berrang, **Paul Gerhart**, and Dominique Schröder

https://eprint.iacr.org/2024/998.pdf

OurPaper

VisualAnon

The Smallest CAS Size

Country	District	Minimal Set
Australia	Acton Alps - East	1 male 1 female
NZ	Chatham Islands Territory Chatham Islands Territory	3 males 3 females
Japan	Nakagusuku-son Nakagusuku-son	19 males 34 females
US	Bastrop County, Texas Riley County, Kansas	82 males 165 females

Gender-Specific CAS Differences

Country	CAS **(Male)**	CAS **(Female)**	**Male/Female**
Sweden	13,020.80	12,983.48	1.00
Iceland	4,381.61	4,357.64	1.01
Norway	7,465.53	7,350.08	1.02
Hong Kong	10,551.02	2,672.81	**3.95**