

Data Mining the Israeli Population Census

Yuval Adam

<http://bit.ly/28c3-dmic>

חיפוש

חיפוש לפי

ת.ז.

שם משפחה

שם פרטי

טלפון

עיר

רחוב

מספר

לידה

חפש

ת.ז.

שם משפחה ופרטי

עיר, רחוב, מספר

עיר, שם פרטי

עיר, תאריך לידה

טלפון

תאריך לידה

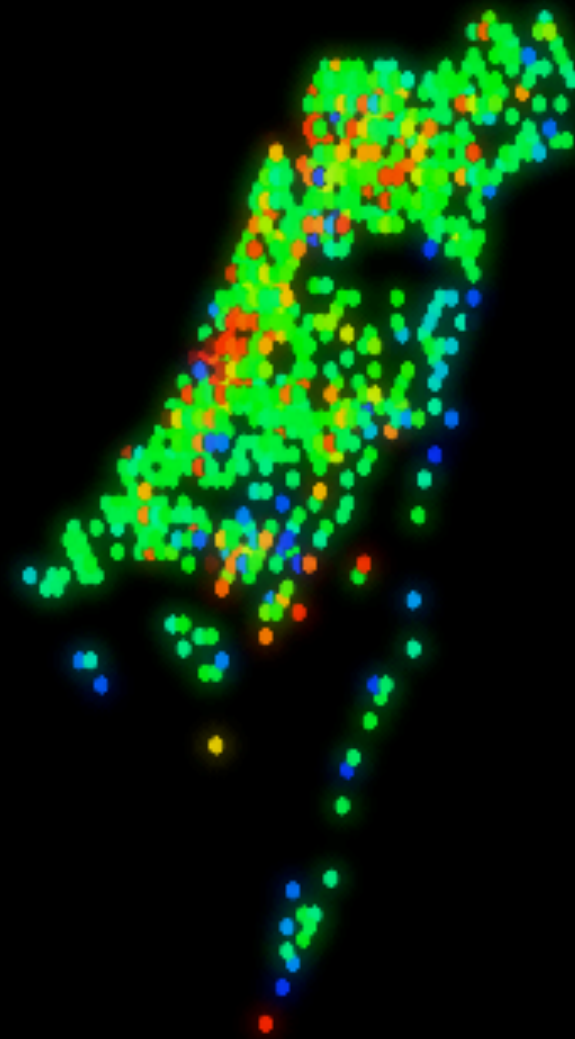
קד	טלפון	לידה	אמא	אבא	מס	מיקוד	רחוב	עיר	נוסף	פרטי	משפחה	ת.ז.
		260513		ישראל	0	10855	תמדיה	תמדיה		רובי	אנטי	0006
		380717		ספי	0	93105	עמק רפאים	ירושלים		אתנס	אנטי	0010
		190630		אנט	0	93105	עמק רפאים	ירושלים		גורג	אנטי	0010
		300728		אנט	0	93105	עמק רפאים	ירושלים		ניקול	אנטי	0010
09	53661	431221	אינ	מנח	48	43465	חנקין	רעננה		אורה	אנטי	0012
		380206		אבא	24	43221	רובצקי	רעננה		עזרא	אנטי	0024
		190000	רחל	שמון	7	53303	המבוא	גבעתיים		מיכא	אנטי	0029
08	74469	430000		אלי	40	79850	נווה מבטח	נווה מבטח	רוזל	רמון	אנטי	0032
		260301	שון	בן צ	126	49333	רוטשילד	פתח תקוה		ניצה	אנטי	0038
04	53315	410727	יפה	מנח	27	34604	אינשטין	חיפה		יהודי	אנטי	0047
08	74469	430608		עזר	40	79850	נווה מבטח	נווה מבטח		יחזק	אנטי	0051
		281004	שון	יוסן	7	53303	המבוא	גבעתיים		מלכה	אנטי	0051
03	30360	470000	מלנ	מנש	13	75431	החוצית	ראשון לציון	זילקו	דוד	אנטי	0051
		430830	אסו	משר	51	76286	גורדון	רחובות	נירנב	ניצה	אנטי	0052
03	53871	360318	רחל	יהונ	24	75318	ביל"ו	ראשון לציון		בן צי	אנטי	0052
		400209	חנה	בן צ	51	76286	גורדון	רחובות		נוח	אנטי	0052
03	47427	110000		יהונ	12	62644	קלונימוס	תל אביב יפו		יגאל	אנטי	0052
		360905	שרו	אלי	5	32583	מימון	חיפה		אהוב	אנטי	0057

מספר רשומות שנקראו 401

2001

Civil Registry

- Official Israeli government database
- Personal details of every Israeli citizen, alive or deceased
- Started shortly after the declaration of independence in 1948
- Last leaked 2006



Data Schema

<u>ID</u>	<u>Name</u>	<u>Date of Birth</u>	<u>Sex</u>	<u>Status</u>	<u>Address</u>	<u>Phone #</u>
1234	Yuval Adam	14/3/1980	M	Single	1 St. Tel Aviv	3-5551234

<u>ID</u>	<u>Maiden Name</u>	<u>Father ID</u>	<u>Mother ID</u>	<u>Country of Birth</u>	<u>Spouse ID</u>
1234	-	12	34	Israel	-

count (*) = ~9,200,000

חיפוש

עיר

רחוב

מספר

ת.ז.

שם משפחה

שם פרטי

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

לידה

חפש

סלפון

```
with open('id_numbers.csv') as f:
    for line in f:
        id, f_id, m_id = line.strip().split(',')

        block = id[:4]
        d[block] += 1

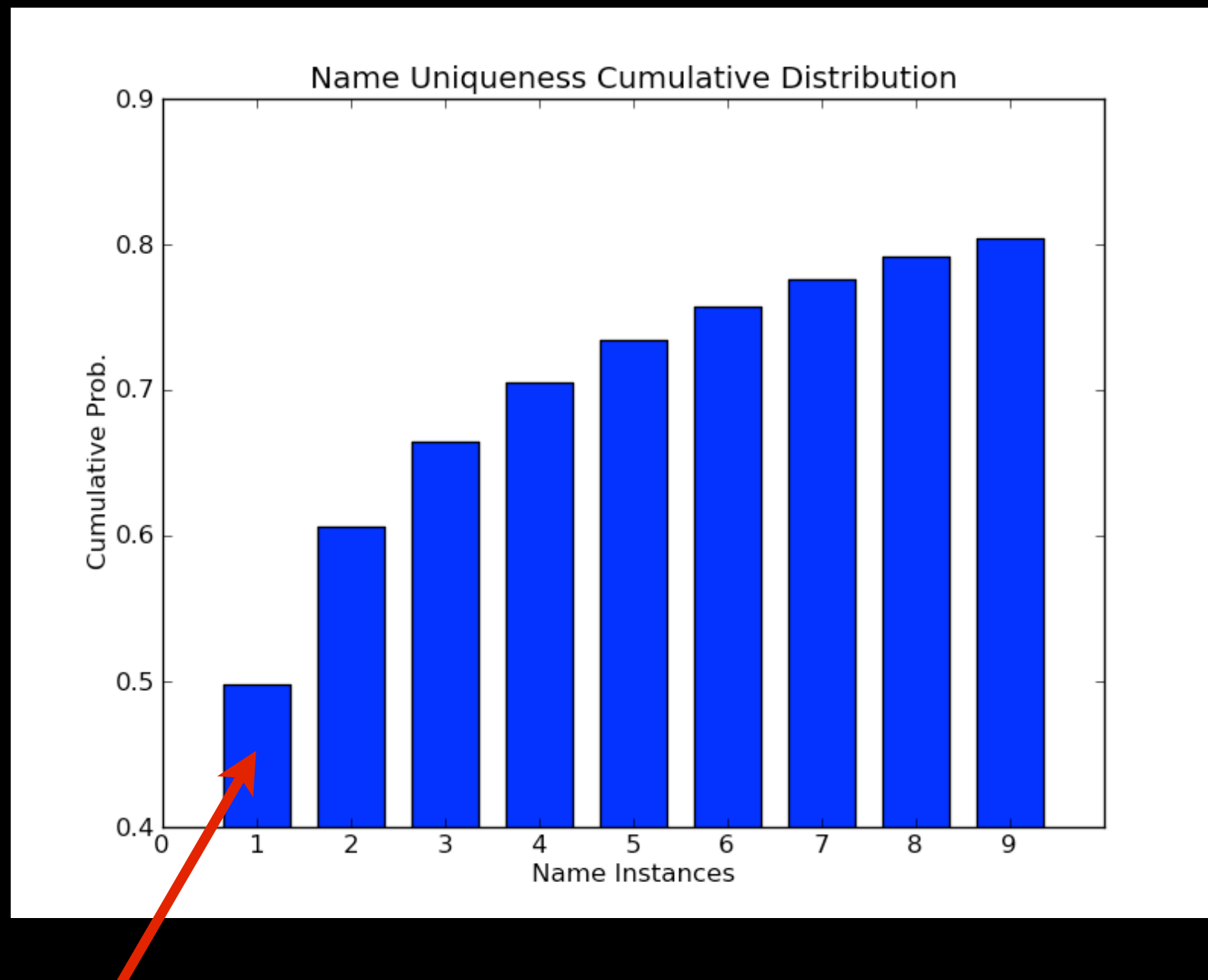
blocks = sorted(d.items(), key=lambda i: i[1])

for block in blocks:
    print block
```

2011

A word cloud of names in various colors and sizes, representing a multicultural community. The names are arranged in a dense, overlapping pattern. The largest names are Abraham, Moses, Jacob, Joseph, Muhammad, and David. Other prominent names include Miriam, Rachel, Michael, Shlomo, Rivkah, Isaac, Hannah, Sarah, and Esther. The colors range from light blue to dark blue, with some names in white or light grey.

Name Uniqueness



~50% chance a given (name, surname) is unique

Uniqueness

Lookup Fields

Uniqueness

name, surname, city

86.9%

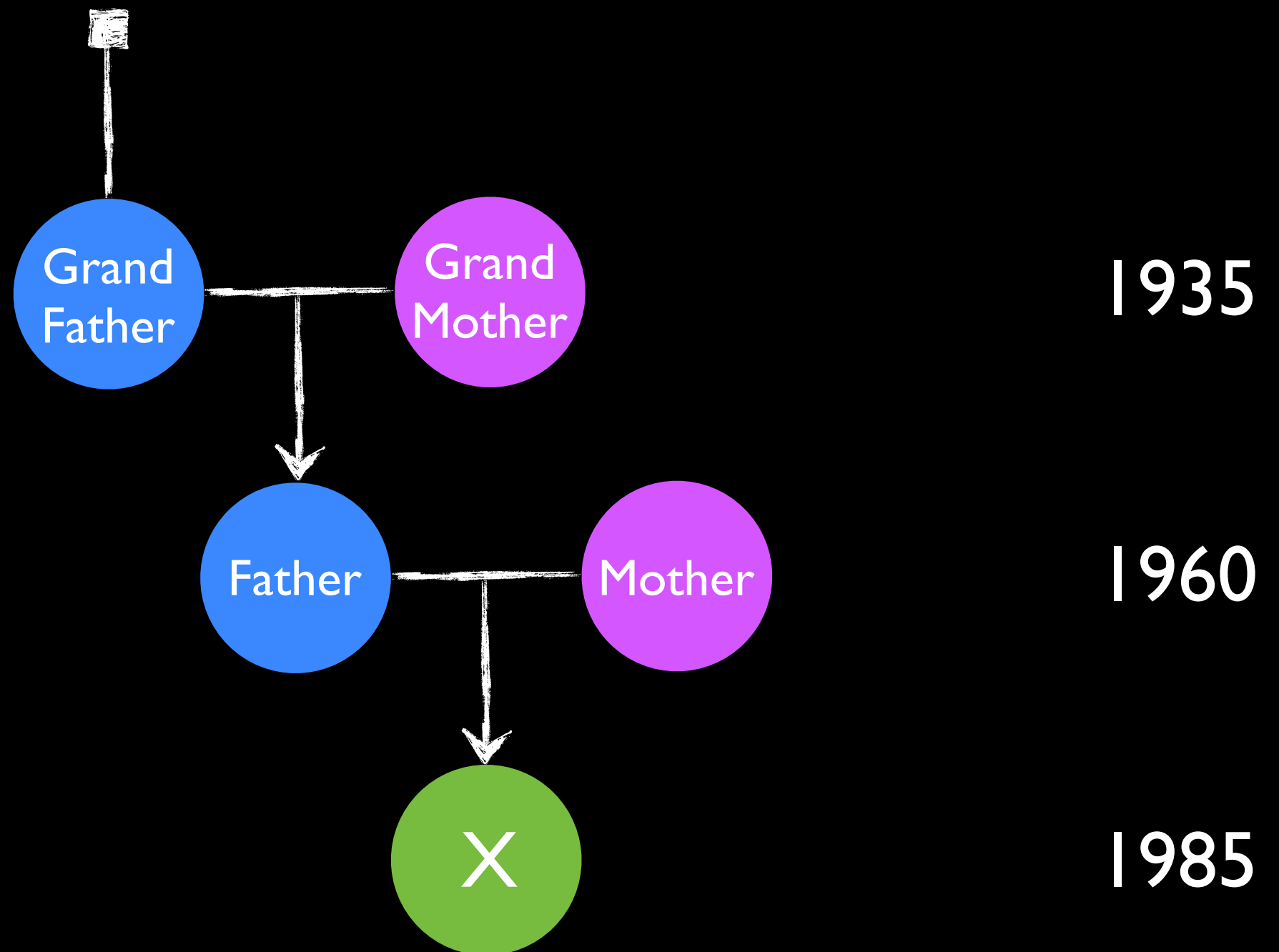
name, surname, date of birth

99.4%

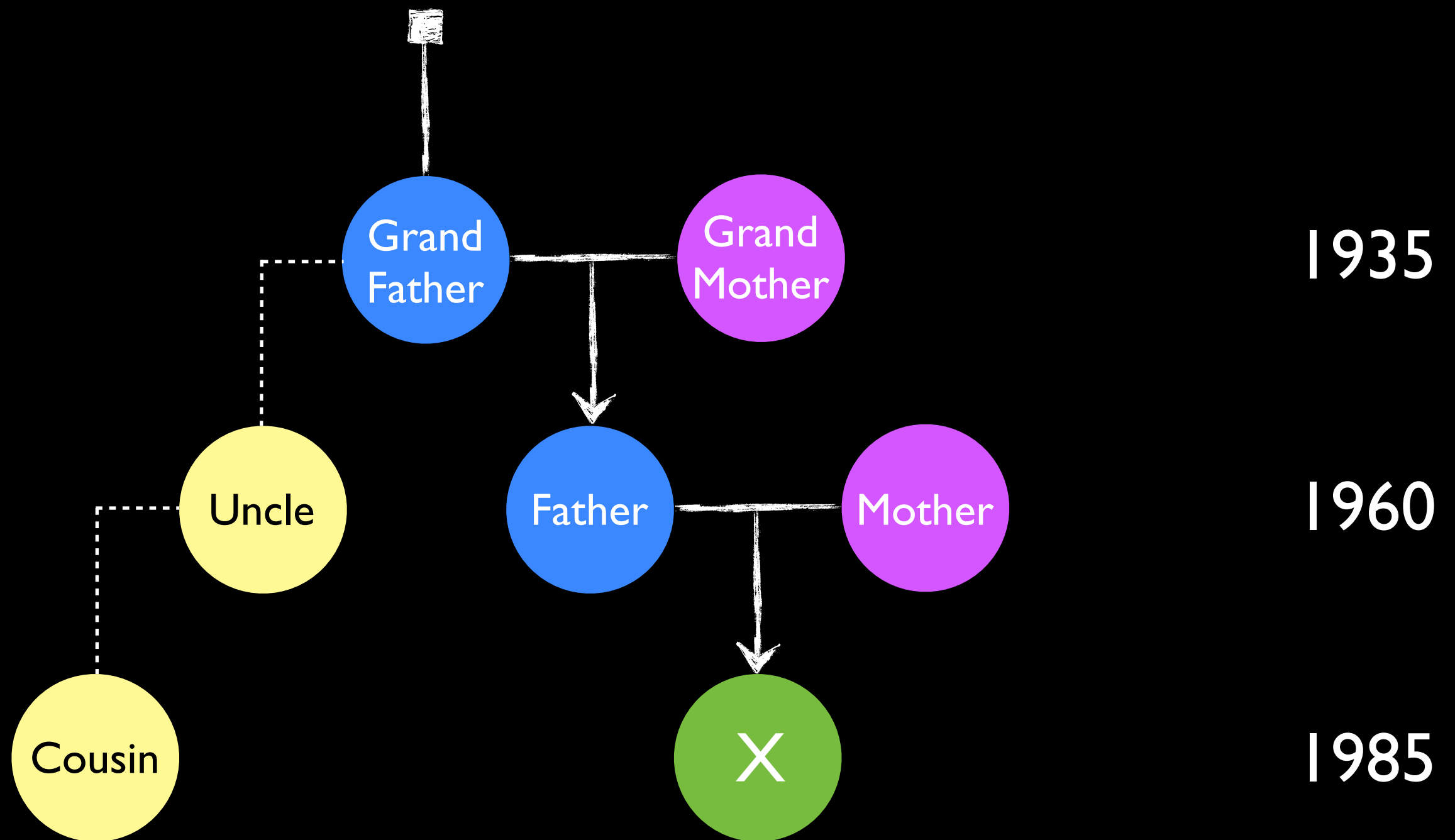
Family Tree

<u>ID</u> Primary Key	<u>Father ID</u> Foreign Key	<u>Mother ID</u> Foreign Key	<u>Spouse ID</u> Foreign Key
1234	12	34	—
12	1	2	34
34	3	4	12

Genogram



Genogram



Graph Properties

- ~9M nodes
- ~420K connected components
- Families of 20 people (average)
- Using other meta-data and heuristics, graph connection can be improved

Data Versions

1998

001011
011101
110011
010011
110101

2001

001111
010101
111011
010011
101101

2002

010110
010101
100010
011011
101111

2004

110110
010101
000010
010101
101001

2006

010100
010101
101010
011101
101000

Data Versions

1998	2001	2002	2004	2006
001011	001111	010110	110110	010100
011101	010101	010101	010101	010101
110011	111011	100010	000010	101010
010011	010011	011011	010101	011101
110101	101101	101111	101001	101000

```
diff 2001.csv 2006.csv
```


New Records

2001

<u>ID</u>	<u>Name</u>	<u>DoB</u>



2006

<u>ID</u>	<u>Name</u>	<u>DoB</u>
9876	Jane Doe	5/5/2005



<u>ID</u>	<u>Name</u>	<u>DoB</u>

<u>ID</u>	<u>Name</u>	<u>DoB</u>
1234	John Doe	6/6/1966

Updates

2001

<u>ID</u>	<u>Name</u>	<u>DoB</u>
1234	Yuval Adam	14/3/1980



2006

<u>ID</u>	<u>Name</u>	<u>DoB</u>
1234	John Doe	14/3/1980



<u>ID</u>	<u>Name</u>	<u>Status</u>
5678	John Doe	Single

<u>ID</u>	<u>Name</u>	<u>Status</u>
5678	John Doe	Deceased

Redactions

2001

<u>ID</u>	<u>Name</u>	<u>DoB</u>
5678	John Doe	14/3/1980



2006

<u>ID</u>	<u>Name</u>	<u>DoB</u>

Redactions

2001

<u>ID</u>	<u>Name</u>	<u>DoB</u>
5678	John Doe	14/3/1980



2006

<u>ID</u>	<u>Name</u>	<u>DoB</u>

???

Redactor's Dilemma^[1]

Google Maps



Walla Maps



[1] <http://www.juliansanchez.com/2009/12/08/the-redactors-dilemma/>

So what's the problem?

- Sensitive private data has leaked
- Social engineering is easier
- But this data *has* been out for 10 years
- What's done is done...

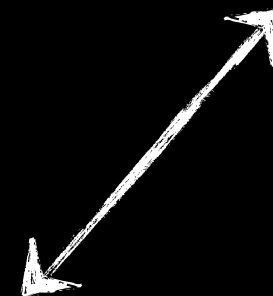
Biometric Data Law

- Passed in early 2011
- Regulate creation of 'Smart ID cards'
- Enable biometric data collection
- Stronger authentication and mitigation of fake 'double IDs'

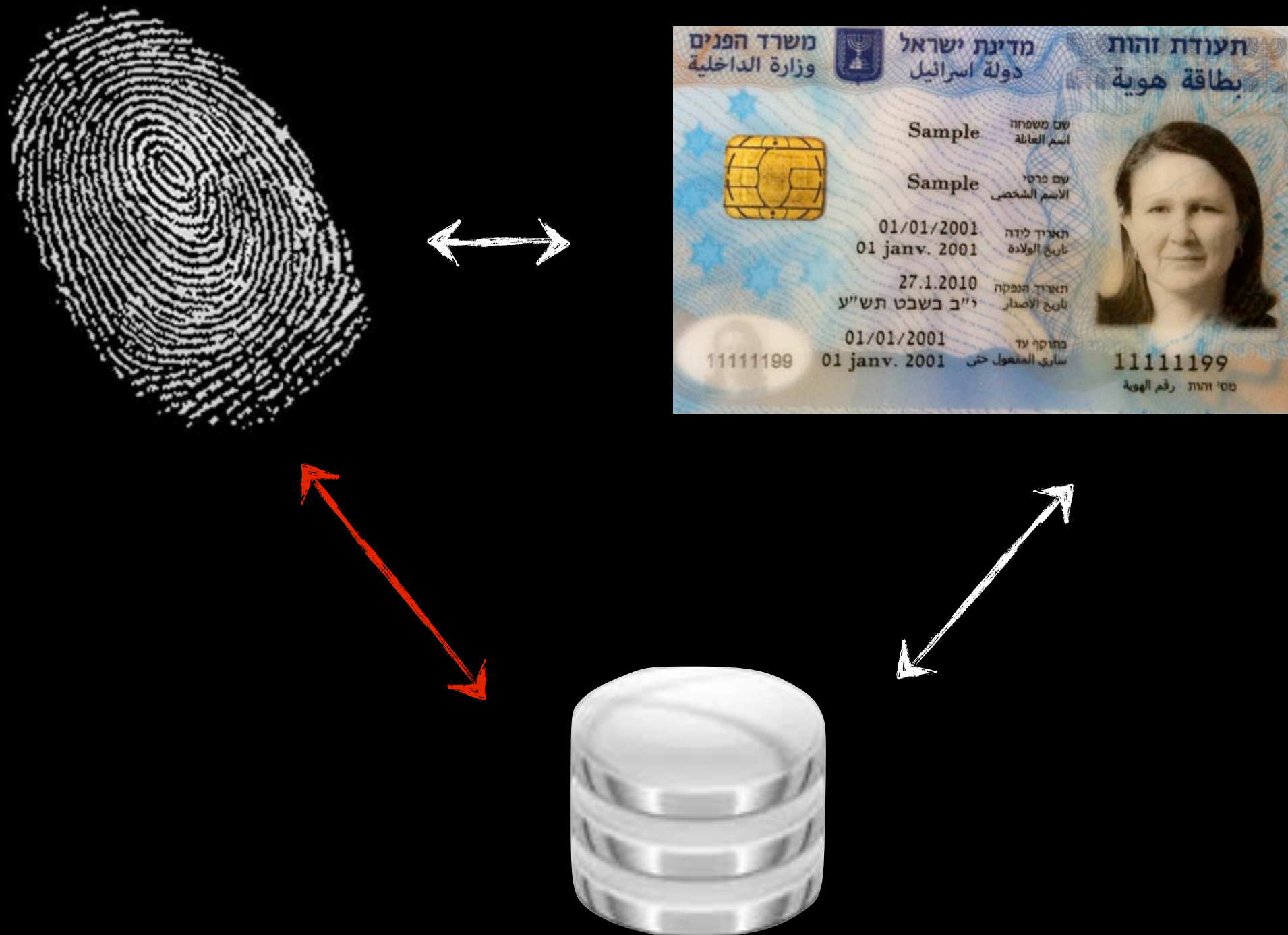
Biometric Data



Biometric Data



Biometric Data



Bleak Future

- Many offices suddenly *need* access
- Israeli public has expressed concern over biometric data collection
- Parliament has not addressed privacy concerns, no due process
- Rejected feedback from Prof. Adi Shamir (RSA, 2002 ACM Turing Award)

Review

- Every Israeli citizens data is exposed
- Not enough has been done to prevent recurring leaks
- Society should closely monitor government data collection policies

Questions?

<http://y3xz.com>