# T2WML Table To Wikidata Mapping Language

Daniel Garijo, Pedro Szekely
Information Sciences Institute and
Department of Computer Science

@dgarijov, @szeke {dgarijo, pszekely}@isi.edu

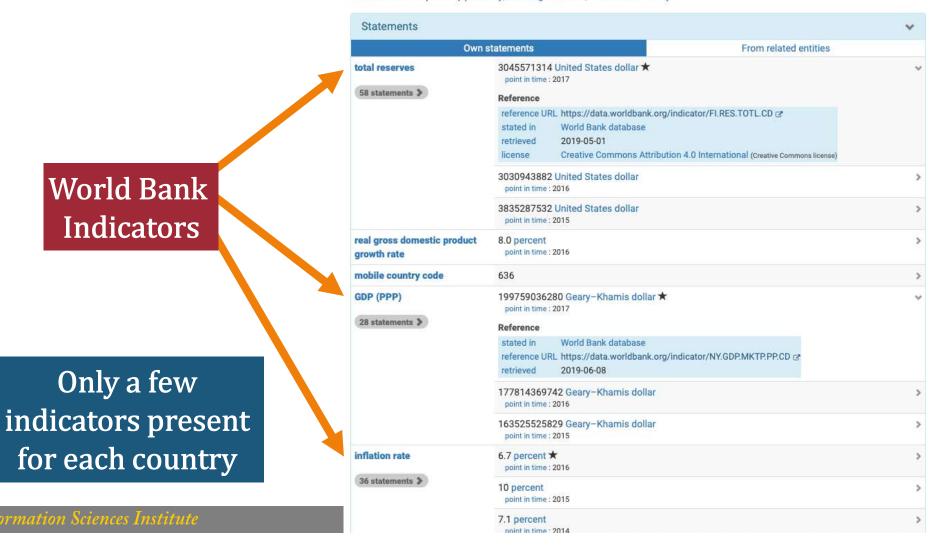
# **Motivating Example**

### Ethiopia (Q115)

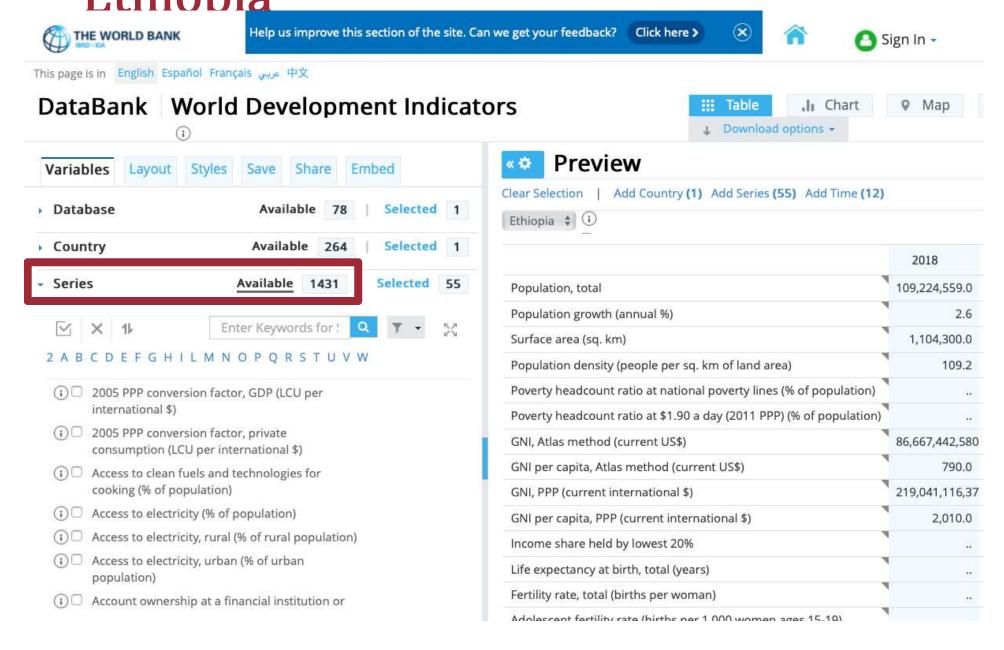
Federal Democratic Republic of Ethiopia | et | III | ETH

country in East Africa

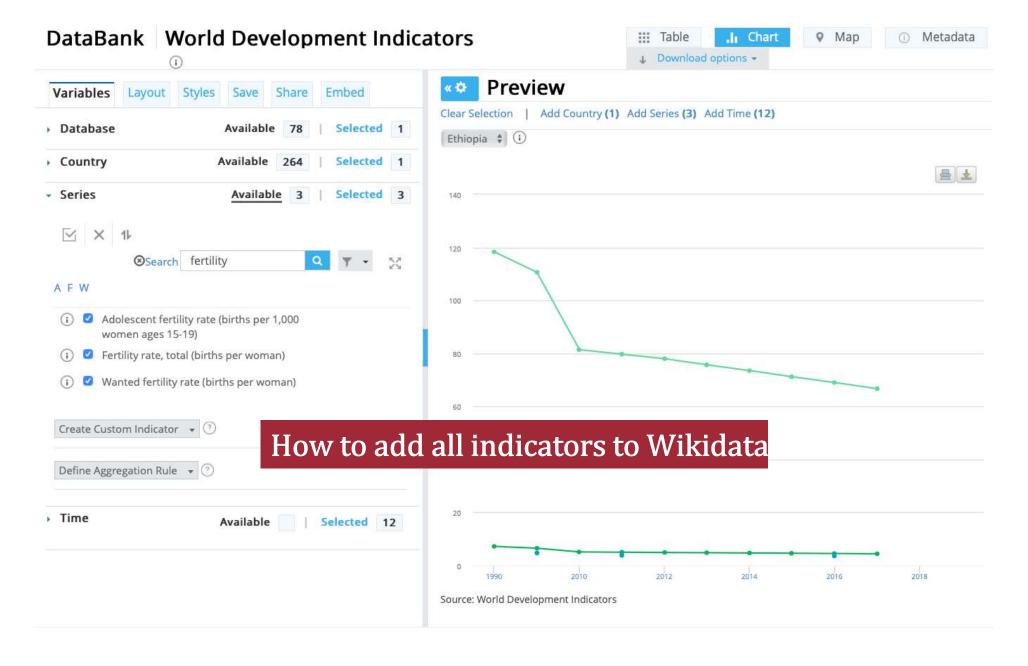
instance of: Ethiopia is a(n) country, sovereign state 1, landlocked country



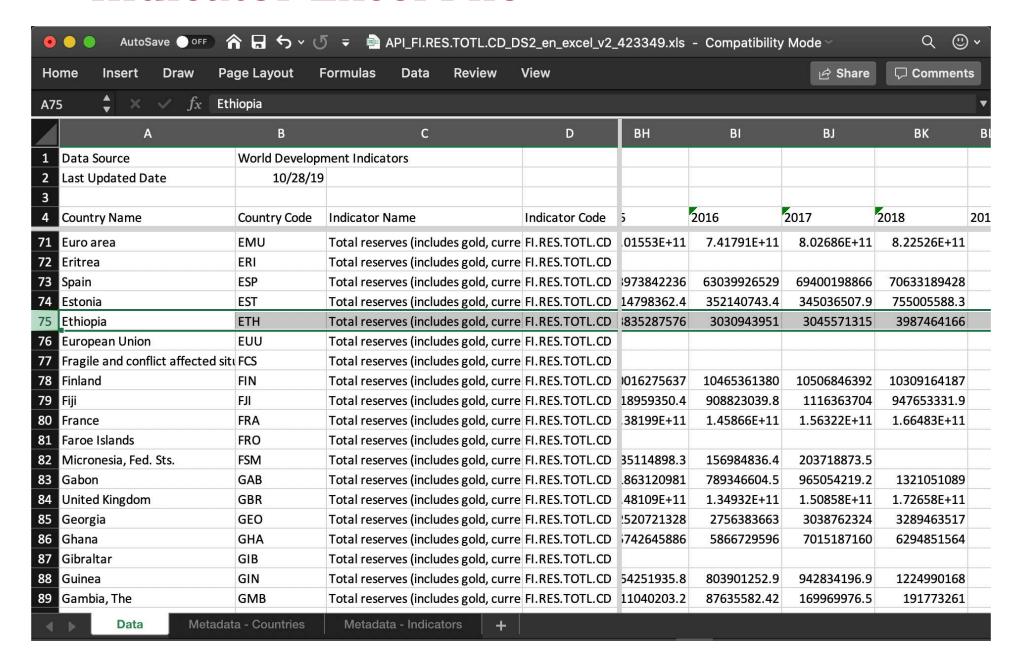
# World Bank: 1,431 Indicators For Ethionia



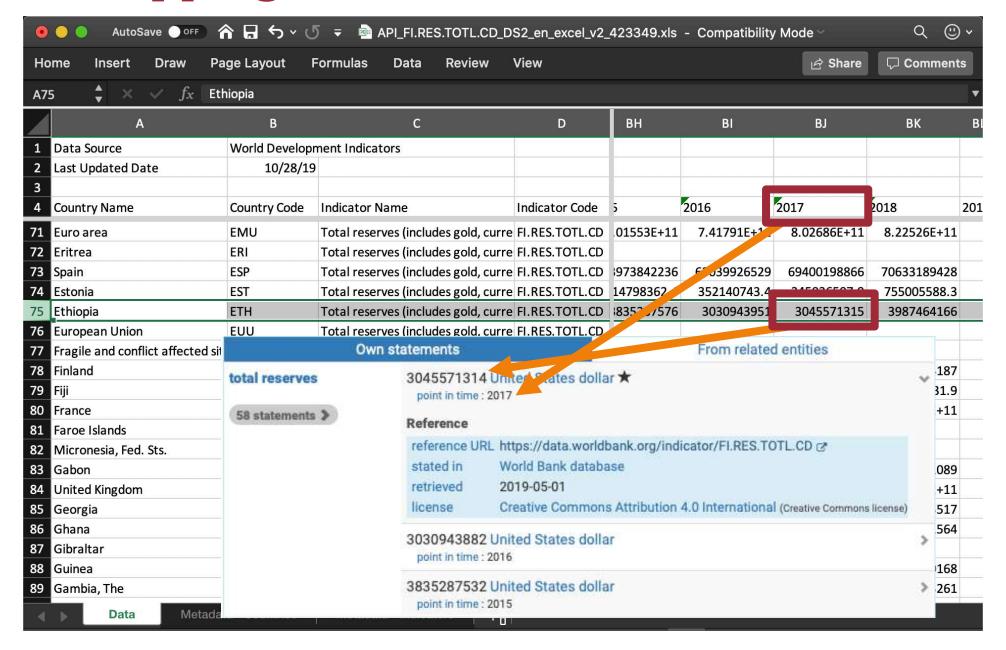
# **Indicator Examples**



# **Indicator Excel File**



# Mapping Data To Wikidata Statements



# Challenge: Different Shapes & Layouts

Country	Source		2000	2001	2002
Burundi	SDG	Females	1	2	2
Burundi	SDG	Males	4	5	6
Comoros	GHD Estimate	Females	2	1	3
Comoros	GHD Estimate	Males	4	4	8
Djibouti	GHD Estimate	Females	2	1	3
Djibouti	GHD Estimate	Males	1	1	

Country	Source	Population	Homicides					
2000								
Burundi	SDG	Females	1					
	SDG	Males	3					
Comoros	GHD Estimate	Females	2					
	GHD Estimate	Males	4					
Djibouti	GHD Estimate	Females	2					
	GHD Estimate	Males	1					
	200	)1						
Burundi	SDG	Females	2					
	SDG	Males	3					
Comoros	GHD Estimate	Females	1					
	GHD Estimate	Males	-					
Djibouti	GHD Estimate	Females	1					
	CUD Fatimata	N 4 - 1	1					

<b>UNODC Name</b>	Source		2000	2001	2002
Burundi	SDG	Females	1	2	2
Burundi	SDG	Males	4	5	6
Burundi	SDG	Total	5	7	8
Comoros	GHD Estimate	Females	2	1	n/a
Comoros	GHD Estimate	Males	4	n/a	8
Comoros	GHD Estimate	Total	6	1	8
Djibouti	GHD Estimate	Females	2	1	3
Djibouti	GHD Estimate	Males	1	1	n/a
Djibouti	GHD Estimate	Total	3	2	3

Country	Source	Males			Females		
		2000	2001	2002	2000	2001	2002
Burundi	SDG	4	5	6	1	2	2
Comoros	GHD Estimate	4	-	8	2	1	
Djibouti	GHD Estimate	1	1		2	1	3

Same Data

# Wikidata Statements

Country	Source		2000	2001	2002		UNODC Name	Source		2000	2001	2002
Burundi	SDG	Females	1	2	2		Burundi	SDG	Females	1	2	2
Burundi	SDG	Burun	di (G	967	)				⁄lales	4	5	6
Comoros	<b>GHD Estimate</b>	Murder ar	nd non	neglig	gent m	an	slaughter (P1	00024): 1	otal	5	7	8
Comoros	<b>GHD Estimate</b>	- point		,	,		2000		emales	2	1	n/a
Djibouti	<b>GHD Estimate</b>	- appli				01)			⁄lales	4	n/a	8
Djibouti	GHD Estimate	- state	d in (F	248):			Institute fo		otal	6	1	8
							Metrics an	d Evaluation	emales	2	1	3
Country	y Source	_	11 /0	007	\				⁄lales	1	1	n/a
eeuntry	Source	Burun	dı (C	1967	)				otal	3	2	3

Sustainable Development

Country	Source	Burundi (Q967)	
		Murder and nonnegligent mans	aughter (P100024): 4
Burundi	SDG	- point in time (P585):	2000
	SDG	- applies to people (P6001):	
Comoros	GHD Estima		Institute for Health
	<b>GHD</b> Estima		Metrics and Evaluation
Djibouti	GHD Estima		
	<b>GHD Estima</b>	Comoros(Q970)	

les		Females					
001	2002	2000	2001	2002			
5	6	1	2	2			
	8	2	1				
1		2	1	3			

	GUD ESTILIA	001110102	(4370)		
		Murder and no	onnegligent	mans	aughter (P100024): 2
Burundi	SDG	- point in ti	me (P585):		2000
	SDG	<ul> <li>applies to</li> </ul>	people (P	6001):	female
Comoros	GHD Estima	- stated in	(P248):		Sustainable Developm
	<b>GHD Estima</b>	ce  iviales			Goals
Djibouti	GHD Estima	te Females	1	Q	
9	CUD Estima	to Malaa	1		

# Table To Wikidata Mapping Language



### T2WML GUI v1.4 (beta)



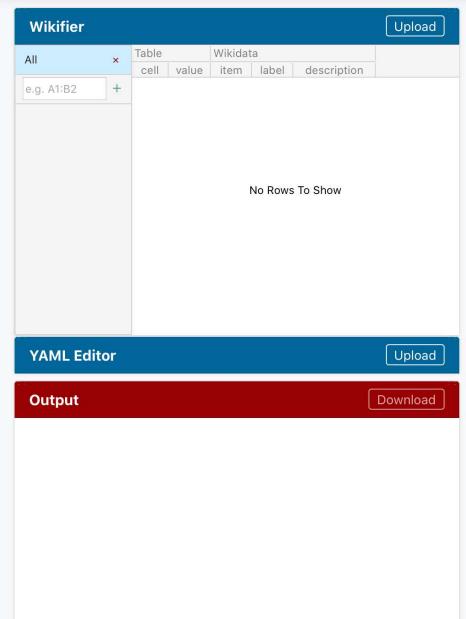


Step 1

Load the data

?

table-1a table-1b table-2a table-2b table-3a



### T2WML GUI v1.4 (beta)

6

7

8

9

10

11

12

13

14

15

Comoros





### Males GHD Esti... GHD Esti... 2001 Step 2

Males

Females

SDG

SDG

SDG

SDG

SDG

Link cell values to Wikidata

Burundi Q967 Comoros Q970 Djibouti Q977 SDG Q7649586 **GHD** Estimate Q6039400 **Females** Q6581072 Males Q6581097

table-1a

table-1b

table-2a

table-2b

table-3a

2

2

3

4

6

2

2001

2002

2000

2001

2002

2001

2002

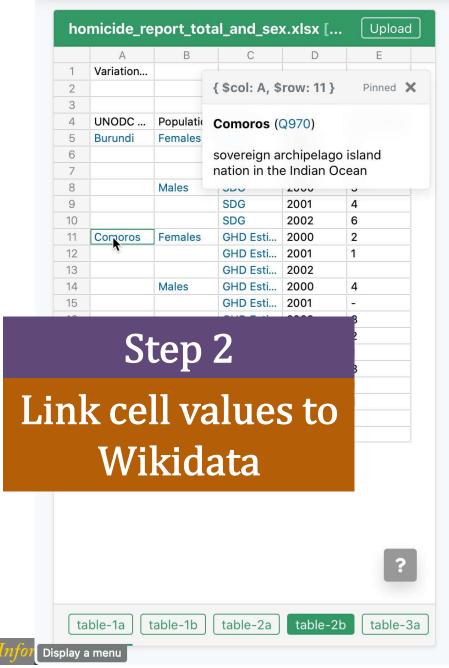
2000

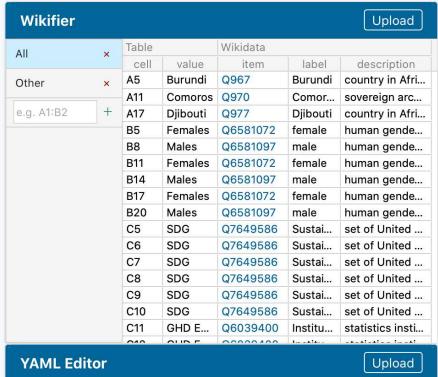
GHD Esti... 2000

GHD Esti...

GHD Esti...



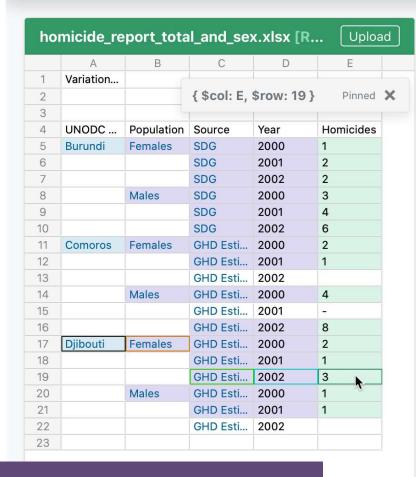




**Output** Download

erbi







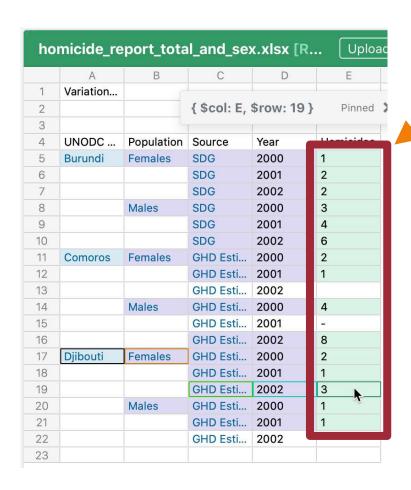
# Step 3 Map table to Wikidata

?

table-3a



# T2WML: Two Ideas



1) define data region

### 2) define mapping for one cel

### Djibouti (Q977)

P100024

3

- point in time: 2002-01-

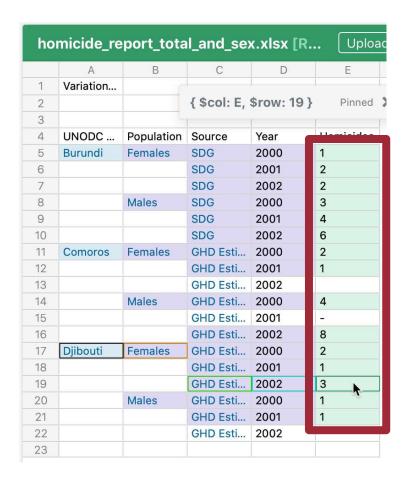
01T00:00:00

- applies to people: female

- publisher: Institute for Health

Metrics and Evaluation

# 1) define data region



### region defined using constants

### region:

- left: D

right: F

top: 4

bottom: \$bottom

# 1) define data region



### region defined using constants

```
region:
- left: D
```

right: F top: 4

bottom: \$bottom

### region robust to variable # of header row

```
region:
```

- left: D right: F

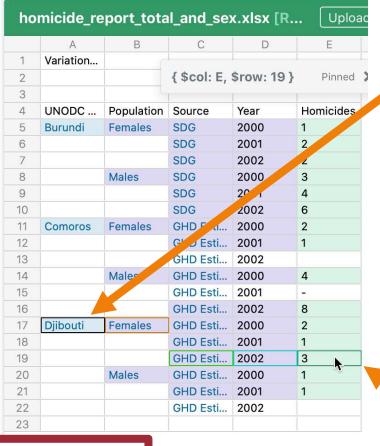
top: value(B/1+n) =

"Population" -> 1+n

bottom: \$bottom

- in column B, start at row 1
- search down for a cell containing "Population"
- return the row where it was found

# 2a) define mapping for one cell



for cell in \$col/\$row

item: item(A/\$row-n)

in column A, go up until you find a cell mapped to a Wikidata item

property: P100024 # murder

the property is P100024 for all cells

value: value(\$col/\$row)

the value is the contents of the current cell

Djibouti (Q977)

P100024

3

- point in time: 2002-01-

01T00:00:00

- applies to people: female

- publisher: Institute for Health

# 2a) define mapping for one cell

	А	В	C	D	Е
1	Variation				
2			{ \$col: E,	\$row: 19 }	Pinned
3		1			
4	UNODC	Population	Source	Year	Homicides
5	Burundi	Females	SDG	2000	1
6			SDG	2001	2
7			SDG	2002	2
8		Males	SDG	2000	3
9			SDG	2001	4
10			SDG	2002	6
11	Comoros	Females	GHD Esti	2000	2
12			GHD Esti	2001	1
13			GHD Esti	2002	
14		Males	GHD Esti	2000	4
15			GHD Esti	2001	-
16			GHD Esti	2002	8
17	Djibouti	Females	GHD Esti	2000	2
18			GHD Esti	2001	1
19			GHD Esti	2002	3
20		Males	GHD Esti	2000	1
21			GHD Esti	2001	1
22			GHD Esti	2002	
23					

### for cell in \$col/\$row

item: item(A/\$row-n)

property: P100024 # murder

value: value(\$col/\$row)

Djibouti (Q977)

P100024

3

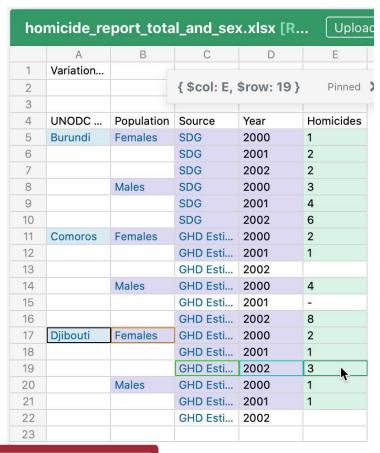
- point in time: 2002-01-

01T00:00:00

- applies to people: female

- publisher: Institute for Health

# 2b) define "point in time" qualifier



for cell in \$col/\$row

property: P585 # point in time

value: value(\$col-1/\$row)

the year is in the column to the lef

calendar: Q1985727

precision: year

time\_zone: 0

format: "%Y"

Djibouti (Q977)

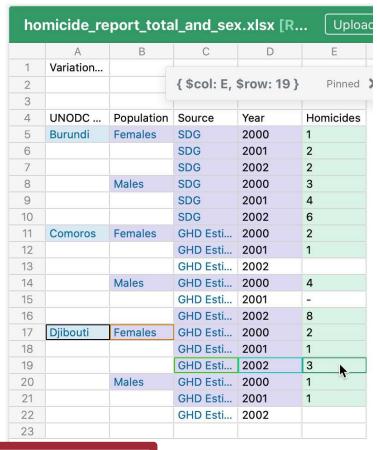
P100024



- applies to people: female

- publisher: Institute for Health

# 2c) define "applies to people" qualifier



for cell in \$col/\$row

property: P6001 # applies to # people

value: item(B/\$row-n)

- the value is in column B
- need to go up from the current row until finding an item mapped to Wikidata

Djibouti (Q977)

P100024

3

- point in time: 2002-01-

04700.00.00

- applies to peopl

female

publisher: Institut<mark>e for Healt</mark>

# **T2WML Specification**

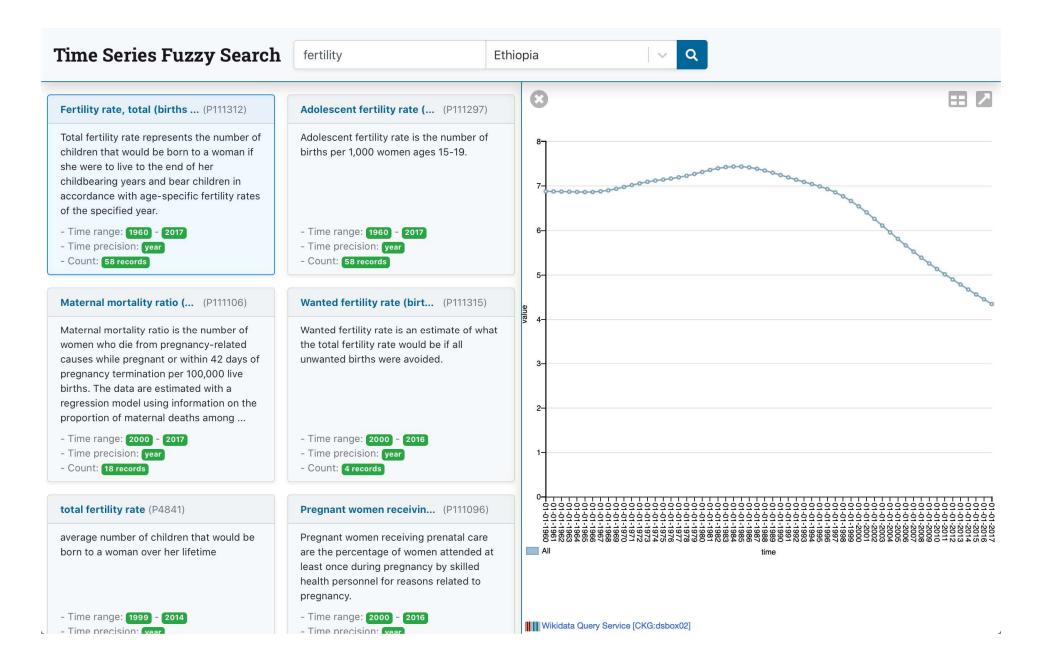
	А	В	С	D	Е
1	Variation				
2			{ \$col: E,	\$row: 19 }	Pinned
3					
4	UNODC	Population	Source	Year	Homicides
5	Burundi	Females	SDG	2000	1
6			SDG	2001	2
7			SDG	2002	2
8		Males	SDG	2000	3
9			SDG	2001	4
10			SDG	2002	6
11	Comoros	Females	GHD Esti	2000	2
12			GHD Esti	2001	1
13			GHD Esti	2002	
14		Males	GHD Esti	2000	4
15			GHD Esti	2001	-
16			GHD Esti	2002	8
17	Djibouti	Females	GHD Esti	2000	2
18			GHD Esti	2001	1
19			GHD Esti	2002	3
20		Males	GHD Esti	2000	1
21			GHD Esti	2001	1
22			GHD Esti	2002	
23					

### Djibouti (Q977)

# P100024 3 - point in time: 2002-0101T00:00:00 - applies to people: female - publisher: Institute for Health

```
statementMapping:
  region:
    - left: D
      right: F
      top: value(B/1+n) = "Population" -> 1+n
      bottom: $bottom
  template:
    item: item(A/$row-n)
    property: P100024 # murder (new property)
    value: value($col/$row)
    unit: D1002 (a new property)
    qualifier:
      - property: P585 # point in time
        value: value($col-1/$row)
        calendar: Q1985727
        precision: year
        time zone: 0
        format: "%Y"
      - property: P6001 # applies to people
        value: item(B/$row-n)
      - property: P123 # source
        value: item(C/$row)
```

# World Bank Indicators in Wikidata



# World Bank Indicators in Wikidata

