LEVERAGING LINKED DATA: SEMANTIC WEB TECHNOLOGIES APPLIED TO ACS DATA

Jonathan Ortiz

May 12, 2017

OUR PURPOSE

We're building the most **meaningful**, **collaborative**, and **abundant** data resource in the world by dismantling the barriers between data and people.

Our platform helps people who work with data solve problems faster by creating new ways to discover, prep, and collaborate.

data.world

There are a **HUGE** NUMBER of OPEN DATA SETS

data.world

TOO MUCH OF DATA'S GROWTH IS HAPPENING IN SILOS

DATA

data.world

FINDING *it*
UNDERSTANDING *it*
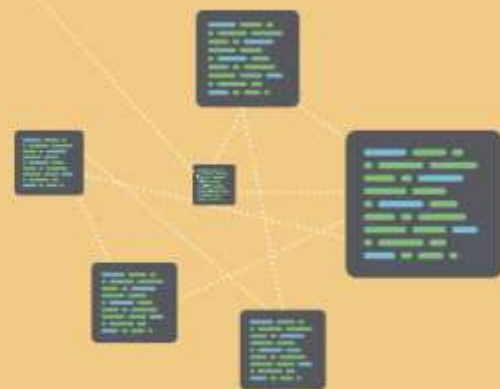*and* **USING** *it can be a challenge*

data.world

So much HUMAN EFFORT is wasted on the WORKING & REWORKING of the SAME DATA

data.world

What is **LINKED DATA**?

It's applying the SAME architecture as the **WWW**
of linked documents to... DATA

data.world

First, break DATA into ATOMIC FACTS

(SUBJECT, PREDICATE, OBJECT)

data.world

(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)

(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)

(Ankara, "is a", City)

(SUBJECT, PREDICATE, OBJECT)

(Turkey, "is a", Country)

(Ankara, "is a", City)

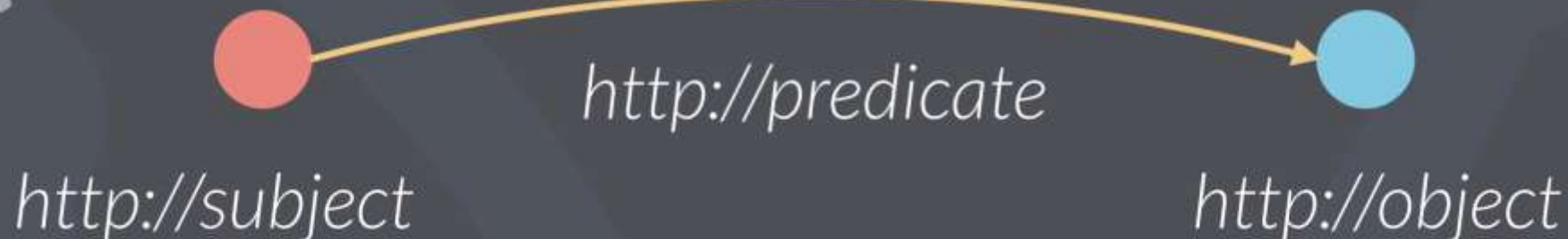(Ankara, "is the capital of", Turkey)

THE TRIPLE

City

Country

"is a"

"is a"

"is the capital of"

Ankara

Turkey

Refer to ENTITIES *and* RELATIONSHIPS *via* **URIs** *so their* **MEANINGS** *can be discussed*

THE CODE

```
<http://dbpedia.org/resource/Turkey>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://dbpedia.org/ontology/Country> .

<http://dbpedia.org/resource/Ankara>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://dbpedia.org/ontology/City> .

<http://dbpedia.org/resource/Ankara>
  <http://dbpedia.org/ontology/capital>
  <http://dbpedia.org/resource/Turkey> .
```

THE CODE

```
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
@prefix dbo: <http://dbpedia.org/ontology/> .

dbpedia:Turkey
   rdf:type dbo:Country .
dbpedia:Ankara
   rdf:type dbo:City ;
   dbo:capital dbpedia:Turkey .
```

PUTTING IT TOGETHER

Country

City

"is a"

"is a"

(dbpedia:Turkey rdf:type dbo:Country)

(dbpedia:Ankara rdf:type dbo:City)

"is the capital of"

Ankara

Turkey

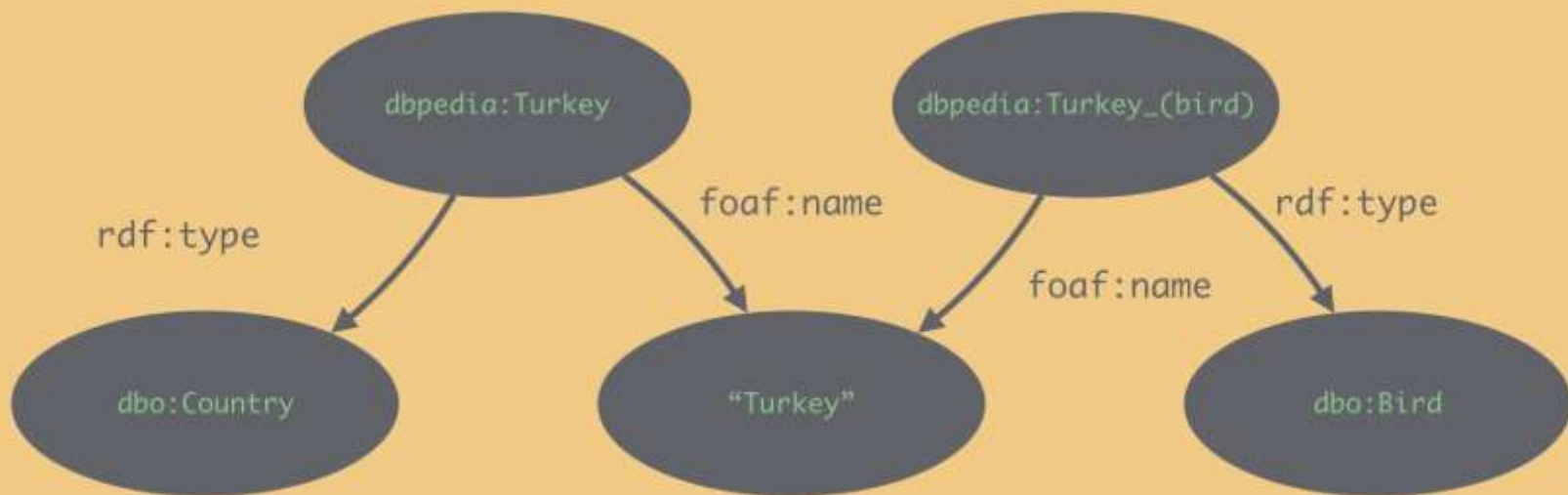(dbpedia:Ankara dbo:capital dbpedia:Turkey)

data.world

Turkey

Turkey

TURKEY *vs* TURKEY

```
(dbpedia:Turkey, rdf:type, dbo:Country)
(dbpedia:Turkey_(bird), rdf:type, dbo:Bird)
(dbpedia:Turkey, foaf:name, "Turkey")
(dbpedia:Turkey_(bird), foaf:name, "Turkey")
```

FIVE-⭐ OPEN DATA

★ OL

★★ OL RE

★★★ OL RE OF

★★★★ OL RE OF URI

★★★★★ OL RE OF URI LD

PDF

XLS

CSV

RDF

LOD

WHY IT MATTERS

- Usability *and* availability

- Sharing knowledge *with* data

- Better starting point

data.world

VIRTUAL STRATEGY MAGAZINE
virtual-strategy.com

Home | Features | Articles | Why Semantics & Data Linking is Vital to Artificial Intelligence

Articles   Features

Why Semantics & Data Linking is Vital to Artificial Intelligence

🕐 December 6, 2016   👤 Craig Norvell   💬 0 Comments   👁 2973 Views

Artificial intelligence and cognitive computing seem to be in the news almost everyday – everything

STRATEGIC PARTNERS

Nutanix

NUTANIX

Call Us: 855-NUTANIX

Website: Visit Us
Directory: Click Here

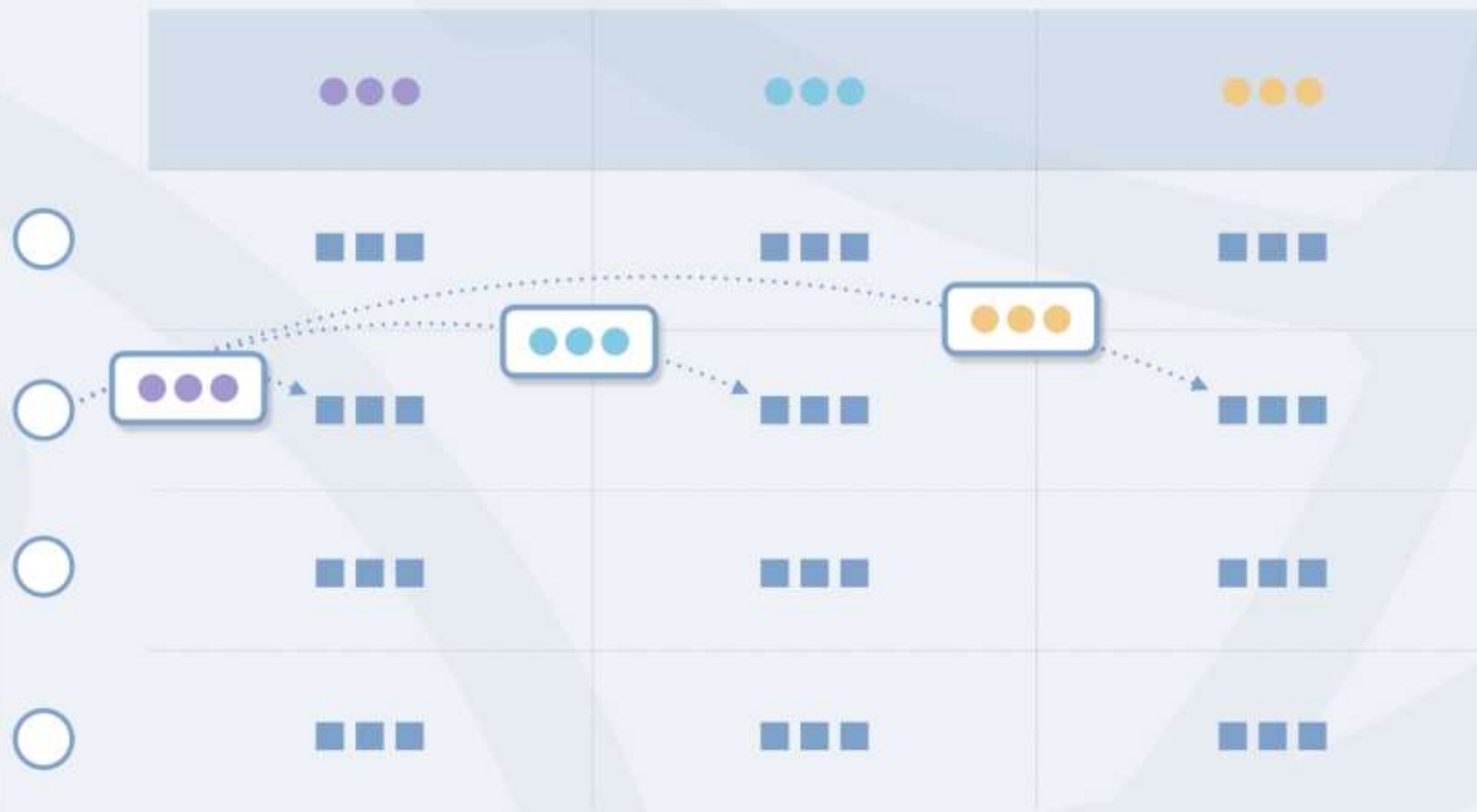Categories: Strategic Partners,Cloud,Hyper-Convergence,Management,Server,Storage

DEVELOP AN **RDF SCHEMA** FOR THE **ACS PUMS**

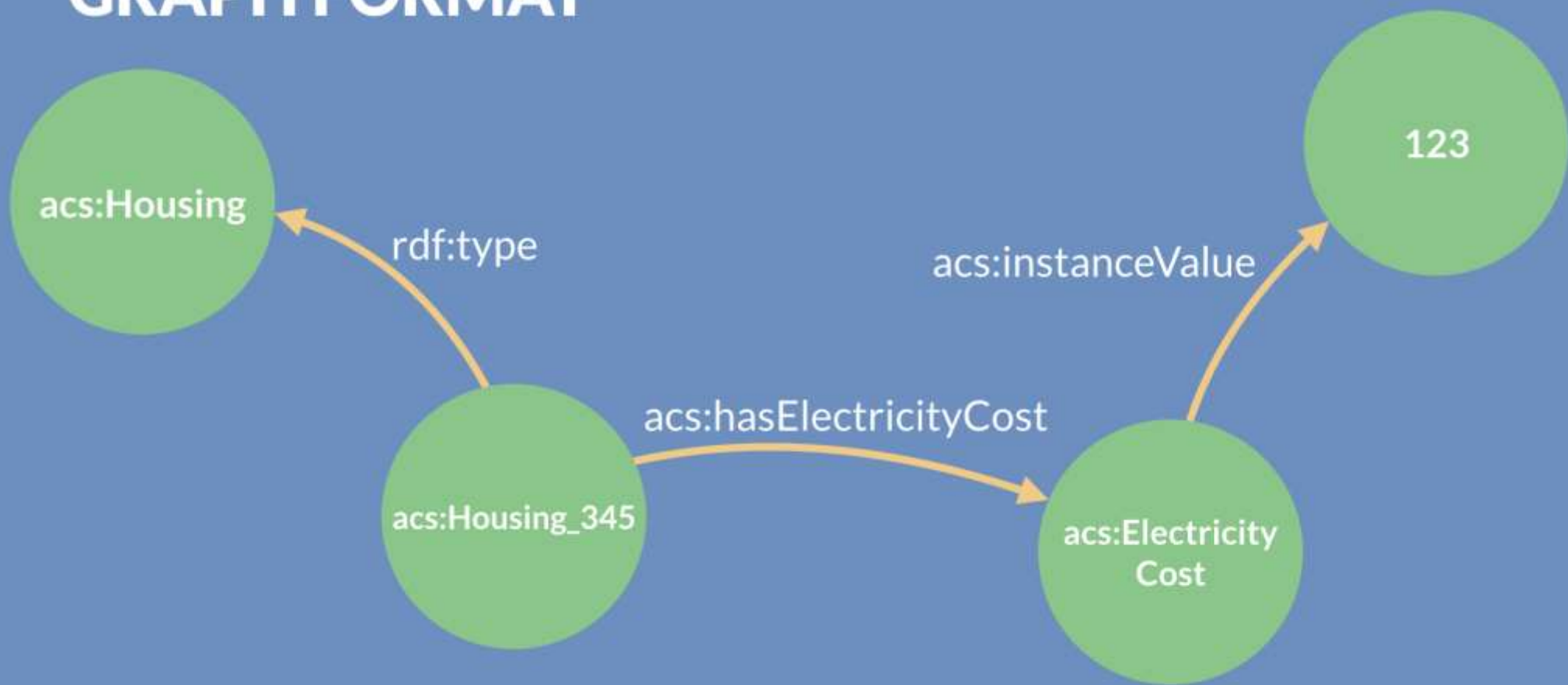TYPICAL PUMS DATA RELEASES ARE TABULAR

CSV

SAS

data.world

TABULAR DATA AS A GRAPH

GRAPH FORMAT

acs:Housing

rdf:type

acs:Housing_345

acs:hasElectricityCost

acs:Electricity Cost

acs:instanceValue

123

data.world

SEMANTIC MODELING PROVIDES BENEFITS

ORIGINAL	ENHANCED
Flat, tabular data	URIs denote linkable concepts
Complex data dictionary	Metadata built into the data itself
Unwieldy raw files	Secure, queryable database

To make **DISCOVERY** of the data easier

To make the data **INTEROPERABLE**

To help the machines learn **FASTER**

The full **LINKED ACS** is available as an **AWS PUBLIC DATASET** thanks to our partners

amazon
web services™

data.world

QUESTIONS

CONTACT INFO

data.world: @databeats
Twitter: @tweetsbyortiz
LinkedIn: linkedin.com/in/jonathanortiz
email: jonathan.ortiz@data.world

THANK YOU

data.world

APPENDIX

CODED PROPERTIES

```
2014 ACS PUMS DATA DICTIONARY
October 27, 2015

HOUSING RECORD

RT 1
Record Type
H .Housing Record or Group Quarters Unit

SERIALNO 7
Housing unit/GQ person serial number
0000001..9999999 .Unique identifier

DIVISION 1
Division code
0 .Puerto Rico
1 .New England (Northeast region)
2 .Middle Atlantic (Northeast region)
3 .East North Central (Midwest region)
4 .West North Central (Midwest region)
5 .South Atlantic (South region)
6 .East South Central (South region)
7 .West South Central (South Region)
8 .Mountain (West region)
9 .Pacific (West region)
```

NESTED PROPERTIES

```
RAC1P 1
Recoded detailed race code
1 .White alone
2 .Black or African American alone
3 .American Indian alone
4 .Alaska Native alone
5 .American Indian and Alaska Native tribes
specified; or American Indian or Alaska
Native, not specified and no other races
6 .Asian alone
7 .Native Hawaiian and Other Pacific
Islander alone
8 .Some Other Race alone
9 .Two or More Races
```

```
RAC2P 2
Recoded detailed race code
01 .White alone
02 .Black or African American alone
03 .Apache alone
04 .Blackfeet alone
05 .Cherokee alone
06 .Cheyenne alone
07 .Chickasaw alone
08 .Chippewa alone
09 .Choctaw alone
10 .Comanche alone
11 .Creek alone
12 .Crow alone
13 .Hopi alone
14 .Iroquois alone
15 .Lumbee alone
16 .Mexican American Indian alone
17 .Navajo alone
18 .Pima alone
...
...
68 .Two or More Races
```

COMBINED PROPERTIES

```
ESP 1
Employment status of parents
b .N/A (not own child of householder, and not
  .child in subfamily)
  .Living with two parents:
1 .Both parents in labor force
2 .Father only in labor force
3 .Mother only in labor force
4 .Neither parent in labor force
  .Living with one parent:
  .Living with father:
5 .Father in the labor force
6 .Father not in labor force
  .Living with mother:
7 .Mother in the labor force
8 .Mother not in labor force
```

COMBINED PROPERTIES

```
ESP 1
Employment status of parents
b .N/A (not own child of householder, and not child in subfamily)
    .Living with two parents:
        1 .Both parents in labor force
        2 .Father only in labor force
        3 .Mother only in labor force
        4 .Neither parent in labor force
    .Living with one parent:
        .Living with father:
            5 .Father in the labor force
            6 .Father not in labor force
        .Living with mother:
            7 .Mother in the labor force
            8 .Mother not in labor force
```

MIXED PROPERTIES

```
ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)
```

MIXED PROPERTIES

```
ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)
```

```
SEMP 6
Self-employment income past 12 months (signed)
bbbbbb  .N/A (less than 15 years old)
000000  .None
-10000..-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001  .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)
```

MIXED PROPERTIES

```
ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)
```

```
SEMP 6
Self-employment income past 12 months (signed)
bbbbbb  .N/A (less than 15 years old)
000000  .None
-10000..-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001  .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)
```

```
CONP 4
Condo fee (monthly amount)
bbbb .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 - $9999 (Rounded and top-coded)
```

MIXED PROPERTIES

```
ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)
```

```
SEMP 6
Self-employment income past 12 months (signed)
bbbbbb  .N/A (less than 15 years old)
000000  .None
-10000..-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001  .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)
```

```
CONP 4
Condo fee (monthly amount)
bbbb .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 - $9999 (Rounded and top-coded)
```

```
POVPIP 3
Income-to-poverty ratio recode
bbb N/A
000..500 .Below 501 percent
501 .501 percent or more
```

MIXED PROPERTIES

```
ELEP 3
Electricity (monthly cost)
bbb .N/A (GQ/vacant)
001 .Included in rent or in condo fee
002 .No charge or electricity not used
003..999 .$3 to $999 (Rounded and top-coded)
```

```
SEMP 6
Self-employment income past 12 months (signed)
bbbbbb  .N/A (less than 15 years old)
000000  .None
-10000..-00001 .Loss of $1 to $10000 (Rounded and bottom-coded)
000001  .$1 or break even
000002..999999 .$2 to $999999 (Rounded and top-coded)
```

```
CONP 4
Condo fee (monthly amount)
bbbb .N/A (GQ/vacant/not owned or being bought)
0000 .Not condo
0001..9999 .$1 - $9999 (Rounded and top-coded)
```

```
POVPIP 3
Income-to-poverty ratio recode
bbb N/A
000..500 .Below 501 percent
501 .501 percent or more
```

```
FULP 4
Fuel cost(yearly cost for fuels other than gas and electricity)
bbbb .N/A (GQ/vacant)
0001 .Included in rent or in condo fee
0002 .No charge or these fuels not used
0003..9999 .$3 to $9999 (Rounded and top-coded)
```

DATA INACCURACIES

```
ST 2                          29 .Missouri/MO
State Code                    30 .Montana/MT
01 .Alabama/AL                31 .Nebraska/NE
02 .Alaska/AK                 32 .Nevada/NV
04 .Arizona/AR                33 .New Hampshire/NH
05 .Arkansas/AR               34 .New Jersey/NJ
06 .California/CA             35 .New Mexico/NM
08 .Colorado/CO               36 .New York/NY
09 .Connecticut/CT            37 .North Carolina/NC
10 .Delaware/DE               38 .North Dakota/ND
11 .District of Columbia/DC   39 .Ohio/OH
12 .Florida/FL                40 .Oklahoma/OK
13 .Georgia/GA                41 .Oregon/OR
15 .Hawaii/HI                 42 .Pennsylvania/PA
16 .Idaho/ID                  44 .Rhode Island/RI
17 .Illinois/IL              45 .South Carolina/SC
18 .Indiana/IN                46 .South Dakota/SD
19 .Iowa/IA                   47 .Tennessee/TN
20 .Kansas/KS                 48 .Texas/TX
21 .Kentucky/KY               49 .Utah/UT
22 .Louisiana/LA              50 .Vermont/VT
23 .Maine/ME                  51 .Virginia/VA
24 .Maryland/MD               53 .Washington/WA
25 .Massachusetts/MA          54 .West Virginia/WV
26 .Michigan/MI               55 .Wisconsin/WI
27 .Minnesota/MN              56 .Wyoming/WY
28 .Mississippi/MS            72 .Puerto Rico/PR
```

data.world