

# wireservice/lookup

Chris Groskopf • Quartz • @onyxfish

- **wireservice**: The new umbrella project for csvkit and other open source tools for journalists.
- **agate**: A python data analysis library designed for journalists.

We are all  
**wasting** so  
much time

Not on Twitter

Not reading the comments

Not on NICAR-L flame wars

But on...

Number of Customers by State by Sector 1000 2014								
	Year	State	Industry Sector		A	B	C	D
3	2014	US	Total Electric Industry	1	usps	state		
54	2014	CA	Total Electric Industry	2	AL	Alabama		
105	2014	TX	Total Electric Industry	3	AK	Alaska		
131	2014	FL	Total Electric Industry	4	AZ	Arizona		
178	2014	NY	Total Electric Industry		AR	Arkansas		
257	2014	PA	Total Electric Industry					

Mapping postal codes to their damned state names.

539	2014	VA	Total Electric Industry	11	FL	Florida		
617	2014	WA	Total Electric Industry	12	GA	Georgia		
631	2014	TN	Total Electric Industry	13	HI	Hawaii		
656	2014	IN	Total Electric Industry	14	ID	Idaho		
662	2014	MA	Total Electric Industry	15	IL	Illinois		
675	2014	MO	Total Electric Industry	16	IN	Indiana		
731	2014	WI	Total Electric Industry					
733	2014	AZ	Total Electric Industry					

$$\frac{\left( \sum_{n=1}^a 52w \cdot b \right)}{365 \cdot 24}$$

14.8  
years












The **madness** must stop.



# Hello, lookup

[github.com/wireservice/lookup](https://github.com/wireservice/lookup)

# A structured repository of lookup tables

 fips	Add population by year. Update metadata spec.
 iso2	Add iso2 and iso3 codes.
 iso3	Add iso2 and iso3 codes.
 naics	Add population by year. Update metadata spec.
 usps	Fix column specs.
 year	Add cpi and error log.
 .gitignore	Add naics 2007.
 ERRORS.md	Add cpi and error log.
 README.md	Updated README.md to add link to agate-lookup

# lookup/usps/state.csv

59 lines (58 sloc) | 780 Bytes

🔍 Search this file...

1	<b>usps</b>	<b>state</b>
2	AL	Alabama
3	AK	Alaska
4	AZ	Arizona
5	AR	Arkansas
6	CA	California
7	CO	Colorado
8	CT	Connecticut

# lookup/usps/state.csv.yml

14 lines (13 sloc) | 350 Bytes

```
1  data: United States Postal System state abbreviations and US state names
2  version: N/A
3  sources:
4    - United States Census Bureau, http://www.census.gov/geo/reference/ansi.html
5  contributors:
6    - Christopher Groskopf <chrisgroskopf@gmail.com>
7  columns:
8    usps:
9      name: USPS state abbreviation
10     type: Text
11   state:
12     name: State name
13     type: Text
```

# Hello, agate-lookup

[agate-lookup.rtfld.org](https://agate-lookup.rtfld.org)

-----+-----	
usps	starbucks
-----+-----	
WA	682
VA	347
NE	39
...	...
-----+-----	

```
table.lookup('usps', 'state')
```

-----+-----+-----		
usps	starbucks	state
-----+-----+-----		
WA	682	Washington
VA	347	Virginia
NE	39	Nebraska
...	...	...
-----+-----+-----		



-----+	
iso2	gini
-----+	
ZA	64.8
AM	30.5
BY	26.5
...	...
-----+	

```
table.lookup('iso2', 'country')
```

-----+-----+-----		
iso2	gini	country
-----+-----+-----		
ZA	64.8	South Africa
AM	30.5	Armenia
BY	26.5	Belarus
...	...	...
-----+-----+-----		

naics	employees
11	154,496
313320	7,294
522310	41,454
...	...

```
joined = table.lookup(  
    'naics',  
    'description',  
    version='2012'  
)
```

naics	employees	description
11	154,496	Agriculture, Fore...
313320	7,294	Fabric Coating Mills
522310	41,454	Mortgage and Nonm...
...	...	...

year	month	price
1985	3	4.45
2007	5	3.18
2011	6	5.21
...	...	...

```
cpi = agate.Table.from_lookup(['year', 'month'], 'cpi')

BASE_CPI = cpi.rows[('2000', '1')]['cpi']

table.compute([
    ('real_price', agate.Formula(agate.Number(), lambda r: r
['price'] * cpi.rows[(r['year'], r['month'])]['cpi'] /
BASE_CPI))
])
```



year	month	price	real_price
1985	3	4.45	2.807206142941523...
2007	5	3.18	3.883525694034258...
2011	6	5.21	6.919206202008269...
...	...	...	...

~~VLOOKUP~~

~~LEFT JOIN~~

~~Bespoke scripts~~  
**lookup**

Let's take our 14 years back.

# Join the wireservice

*Send your pull requests!*

[github.com/wireservice/lookup](https://github.com/wireservice/lookup)  
[agate-lookup.rtfld.org](https://agate-lookup.rtfld.org)