

# Cognitive Dimensions of Between-Table Context Support in Data Wrangling Applications

Steve Kasic

Dec. 10, 2019

# Introduction

- Cognitive Dimensions
- Between-Table Context Support
- Data Wrangling

# What is data wrangling?



```
wrangler_crime.txt - Notepad
File Edit Format View Help
Reported crime in Alabama,
,
2004,4029.3
2005,3900
2006,3937
2007,3974.9
2008,4081.9
,
Reported crime in Alaska,
,
2004,3370.9
2005,3615
2006,3582
2007,3373.9
2008,2928.3
,
Reported crime in Arizona,
,
2004,5073.3
2005,4827
2006,4741.6
2007,4502.6
2008,4087.3
,
Reported crime in Arkansas,
,
2004,4033.1
2005,4068
2006,4021.6
2007,3945.5
2008,3843.7
,
Reported crime in California,
,
```

```
year,state,reports
2004,Alabama,4029.3
2005,Alabama,3900
2006,Alabama,3937
2007,Alabama,3974.9
2008,Alabama,4081.9
2004,Alaska,3370.9
2005,Alaska,3615
2006,Alaska,3582
2007,Alaska,3373.9
2008,Alaska,2928.3
2004,Arizona,5073.3
2005,Arizona,4827
2006,Arizona,4741.6
2007,Arizona,4502.6
2008,Arizona,4087.3
2004,Arkansas,4033.1
2005,Arkansas,4068
2006,Arkansas,4021.6
2007,Arkansas,3945.5
2008,Arkansas,3843.7
2004,California,3423.9
2005,California,3321
2006,California,3175.2
2007,California,3032.6
2008,California,2940.3
2004,Colorado,3918.5
2005,Colorado,4041
2006,Colorado,3441.8
2007,Colorado,2991.3
2008,Colorado,2856.7
2004,Connecticut,2684.9
2005,Connecticut,2579
2006,Connecticut,2575
```

# Wrangling applications



Google Sheets



Microsoft Excel



Tableau Prep



Workbench



OpenRefine



Trifacta Wrangler



Google Cloud Dataprep

# Wrangling applications



Google Sheets



Microsoft Excel



Tableau Prep



Workbench



**OpenRefine**



Trifacta Wrangler



**Google Cloud Dataprep**

**OpenRefine** all longterm managed care [Permalink](#) Open... Export Help

Facet / Filter Undo / Redo 25 / 25 **3769 rows** Extensions: Wikidata

Show as: rows records Show: 5 10 25 50 rows « first < previous 1 - 50 next > last »

**Using facets and filters**

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? [Watch these screencasts](#)

	All	plan name	plan type	report date	county	enrollment
1.	★	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01T00:00:00Z	NEW YORK	2184
2.	★	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01T00:00:00Z	WESTCHESTER	170
3.	★	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	ALBANY	1
4.	★	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	SCHENECTADY	108
5.	★	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	SCHOHARIE	1
6.	★	INDEPENDENT LIVING FOR SENIORS	PACE	2009-01-01T00:00:00Z	MONROE	246
7.	★	PACE CNY	PACE	2009-01-01T00:00:00Z	CHAUTAUQUA	1
8.	★	PACE CNY	PACE	2009-01-01T00:00:00Z	ONONDAGA	317
9.	★	AMERIGROUP	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	431
10.	★	CCM SELECT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	1040
11.	★	CCM SELECT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	WESTCHESTER	17
12.	★	ELANT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	ORANGE	104
13.	★	ELANT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	ROCKLAND	34
14.	★	GUILDNET	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NASSAU	519
15.	★	GUILDNET	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	5810
16.	★	GUILDNET	PARTIAL CAPITATION	2009-01-01T00:00:00Z	SUFFOLK	217
17.	★	HHH CHOICES	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	706
18.	★	HOMEFIRST	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	2976
19.	★	INDEPENDENT CARE SYSTEMS	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK	1331

MANAGED LONG TERM CARE FLOW > **Managed Long Term Care - 3** Full Data Run Job

48 Categories 2 Categories Jan 2008 - Apr 2012 45 Categories 0 - 24.21k

RBC	plan name	RBC	plan type	report date	RBC	county	#	enrollment
-	COMPREHENSIVE CARE MGT	PACE		JANUARY 2009		NEW YORK		2184
-	COMPREHENSIVE CARE MGT	PACE		JANUARY 2009		WESTCHESTER		170
-	EDDY SENIOR CARE	PACE		JANUARY 2009		ALBANY		1
-	EDDY SENIOR CARE	PACE		JANUARY 2009		SCHENECTADY		108
-	EDDY SENIOR CARE	PACE		JANUARY 2009		SCHOHARIE		1
-	INDEPENDENT LIVING FOR SENIORS	PACE		JANUARY 2009		MONROE		246
-	PACE CNY	PACE		JANUARY 2009		CHAUTAUQUA		1
-	PACE CNY	PACE		JANUARY 2009		ONONDAGA		317
-	AMERIGROUP	PARTIAL CAPITATION		JANUARY 2009		NEW YORK		431
-	CCM SELECT	PARTIAL CAPITATION		JANUARY 2009		NEW YORK		1040
-	CCM SELECT	PARTIAL CAPITATION		JANUARY 2009		WESTCHESTER		17
-	ELANT	PARTIAL CAPITATION		JANUARY 2009		ORANGE		104
-	ELANT	PARTIAL CAPITATION		JANUARY 2009		ROCKLAND		34
-	GUILDNET	PARTIAL CAPITATION		JANUARY 2009		NASSAU		519

**Details**

**RBC county**

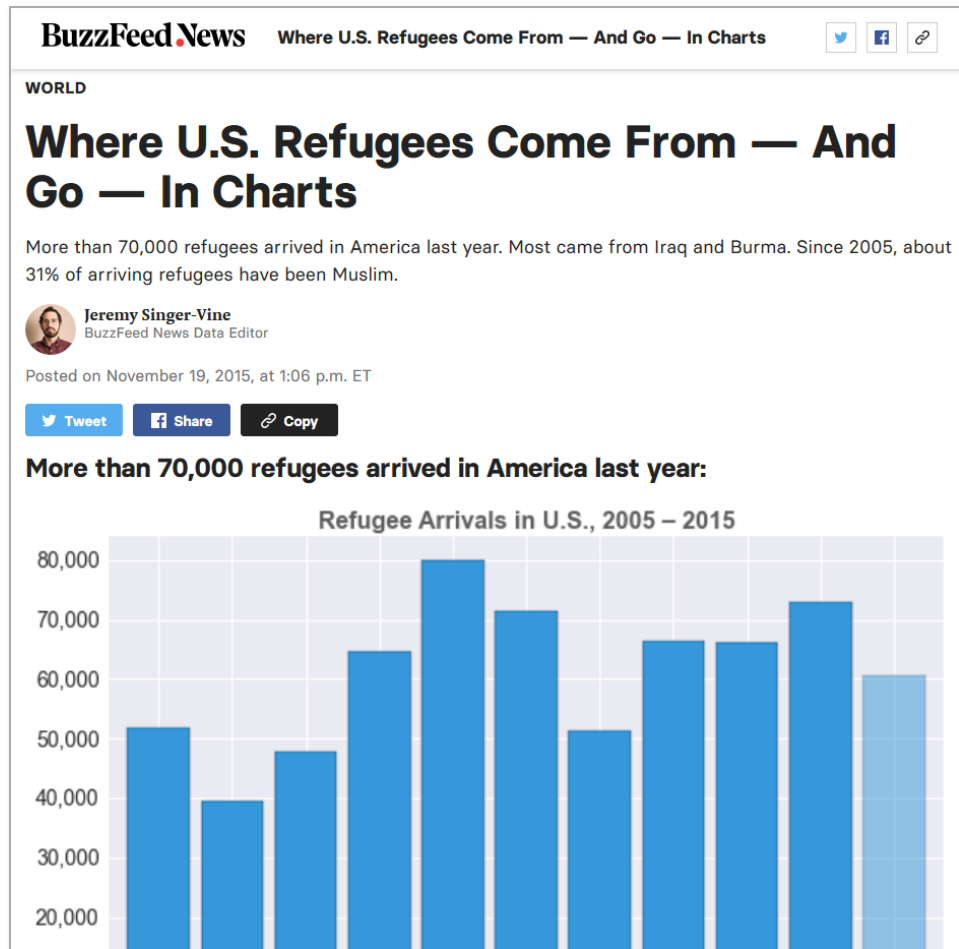
**Unique Values**

Search...

- CHAUTAUQUA 51
- DUTCHESS 50
- ULSTER 42
- RENSSELAER 41
- ALLEGANY 40
- NIAGARA 30
- COLUMBIA 27
- SULLIVAN 23
- MONTGOMERY 21
- PUTNAM 21
- SCHOHARIE 20



# Journalists publish data and analysis code



BuzzFeedNews / 2015-11-refugees-in-the-united-states

Code Issues Pull requests Actions Projects Wiki Security Insights

Branch: master 2015-11-refugees-in-the-united-states / notebooks / us-refugee-analysis.ipynb

jsvine Fix Wyoming omission f1f5ae4 on Nov 19, 2015

1 contributor

966 lines (966 sloc) 377 KB

### Analysis of U.S. Refugee Data

The refugee data in this notebook comes from the [Department of State's Refugee Processing Center's data portal](#). It was retrieved on November 18, 2015.

#### Load the data

```
In [1]: import pandas as pd
import matplotlib as mpl
import seaborn as sb
import itertools
%matplotlib inline

In [2]: by_destination = pd.read_csv("../data/WRAPS-arrivals-by-destination-2005-2015-clean.csv")
by_religion = pd.read_csv("../data/WRAPS-arrivals-by-religion-2005-2015-clean.csv")

In [3]: # The totals for the two data-reports are off by just four arrivals.
print(by_destination["arrivals"].sum())
print(by_religion["arrivals"].sum())

672518
672522

In [4]: state_populations = pd.read_csv("../data/census-state-populations.csv").set_index("state")

In [5]: colors = itertools.cycle([ "#3498db", "#2ecc71", "#9b59b6" ])
```

#### Set up a few data-processing and chart-making functions

```
In [6]: def arrivals_by_year(origin=None, dest_state=None, dest_city=None):
    return by_destination[
        (by_destination["origin"] == (origin if origin else by_destination["origin"])) &
        (by_destination["dest_state"] == (dest_state if dest_state else by_destination["dest_stat
e"])) &
```

# Raw data

	A	B	C
1	<b>MANAGED LONG TERM CARE ENROLLMENT</b>		
2	<b>BY PLAN, COUNTY, AND PROGRAM</b>		
3			
4	NYS JANUARY, 2009		
5			<b>TOTAL</b>
6	<b>MLTC PACE PLAN NAME</b>	<b>COUNTY</b>	<b>ENROLLMENT</b>
7	COMPREHENSIVE CARE MGMT CORP	NEW YORK	2,184
8		WESTCHESTER	170
9		Total	2,354
10	EDDY SENIOR CARE	ALBANY	1
11		SCHENECTADY	108
12		SCHOHARIE	1
13		Total	110
14	INDEPENDENT LIVING FOR SENIORS	MONROE	246
15		Total	246
16	PACE CNY	CHAUTAUQUA	1
17		ONONDAGA	317
18		Total	318
19	Total MLTC PACE Enrollment		3,028
20			
21	<b>MLTC PARTIAL CAPITATION PLAN NAME</b>		
22	AMERIGROUP	NEW YORK	431
23		Total	431
24	CCM SELECT	NEW YORK	1,040
25		WESTCHESTER	17
26		Total	1,057
27	ELANT	ORANGE	104
28		ROCKLAND	34
29		Total	138
30	GUILDNET	NASSAU	519
31		NEW YORK	5,810

	A	B	C	D	E
1	<b>plan name</b>	<b>plan type</b>	<b>report date</b>	<b>county</b>	<b>enrollment</b>
2	COMPREHENSIVE CARE MGMT	PACE	2009-01-01T00:00:00	NEW YORK	2184
3	COMPREHENSIVE CARE MGMT	PACE	2009-01-01T00:00:00	WESTCHESTE	170
4	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00	ALBANY	1
5	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00	SCHENECTAC	108
6	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00	SCHOHARIE	1
7	INDEPENDENT LIVING FOR SENI	PACE	2009-01-01T00:00:00	MONROE	246
8	PACE CNY	PACE	2009-01-01T00:00:00	CHAUTAUQUA	1
9	PACE CNY	PACE	2009-01-01T00:00:00	ONONDAGA	317
10	AMERIGROUP	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	431
11	CCM SELECT	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	1040
12	CCM SELECT	PARTIAL CAPITAT	2009-01-01T00:00:00	WESTCHESTE	17
13	ELANT	PARTIAL CAPITAT	2009-01-01T00:00:00	ORANGE	104
14	ELANT	PARTIAL CAPITAT	2009-01-01T00:00:00	ROCKLAND	34
15	GUILDNET	PARTIAL CAPITAT	2009-01-01T00:00:00	NASSAU	519
16	GUILDNET	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	5810
17	GUILDNET	PARTIAL CAPITAT	2009-01-01T00:00:00	SUFFOLK	217
18	HHH CHOICES	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	706
19	HOMEFIRST	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	2976
20	INDEPENDENCE CARE SYSTEMS	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	1331
21	PARTNERS IN COMMUNITY CAR	PARTIAL CAPITAT	2009-01-01T00:00:00	ORANGE	157
22	PARTNERS IN COMMUNITY CAR	PARTIAL CAPITAT	2009-01-01T00:00:00	ROCKLAND	91
23	SENIOR HEALTH PARTNERS INC	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	1375
24	SENIOR NETWORK HEALTH	PARTIAL CAPITAT	2009-01-01T00:00:00	HERKIMER	36
25	SENIOR NETWORK HEALTH	PARTIAL CAPITAT	2009-01-01T00:00:00	ONEIDA	319
26	TOTAL AGING IN PLACE PROGF	PARTIAL CAPITAT	2009-01-01T00:00:00	ERIE	142
27	VNS CHOICE	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	6783
28	WELLCARE	PARTIAL CAPITAT	2009-01-01T00:00:00	NEW YORK	158
29	ARCHCARE SENIOR LIFE	PACE	2010-01-01T00:00:00	NEW YORK	10
30	CHS BUFFALO LIFE	PACE	2010-01-01T00:00:00	ERIE	9
31	COMPREHENSIVE CARE MGMT	PACE	2010-01-01T00:00:00	NASSAU	8
32	COMPREHENSIVE CARE MGMT	PACE	2010-01-01T00:00:00	NEW YORK	2265
33	COMPREHENSIVE CARE MGMT	PACE	2010-01-01T00:00:00	SUFFOLK	61
34	COMPREHENSIVE CARE MGMT	PACE	2010-01-01T00:00:00	WESTCHESTE	175
35	EDDY SENIOR CARE	PACE	2010-01-01T00:00:00	SCHENECTAC	101
36	INDEPENDENT LIVING FOR SENI	PACE	2010-01-01T00:00:00	MONROE	264
37	PACE CNY	PACE	2010-01-01T00:00:00	CHAUTAUQUA	1
38	PACE CNY	PACE	2010-01-01T00:00:00	ONONDAGA	341

# #2: Wrangling contexts



Within-table















Between-table

# #3: Cognitive Dimensions (CDs)

- Viscosity
- Visibility
- Premature commitment
- Hidden dependencies
- Role-expressiveness
- Abstraction
- Secondary notation
- Closeness of mapping
- Consistency
- Diffuseness/Terseness
- Provisionality
- Progressive evaluation
- Error-proneness
- Hard mental operations
- Juxtaposability/side-by-side ability

# #3: Cognitive Dimensions (CDs)

- Viscosity
- **Visibility**
- Premature commitment
- Hidden dependencies
- Role-expressiveness
- Abstraction
- **Secondary notation**
- Closeness of mapping
- Consistency
- Diffuseness/Terseness
- Provisionality
- Progressive evaluation
- **Error-proneness**
- Hard mental operations
- Juxtaposability/side-by-side ability

# Task Example: Remove rows

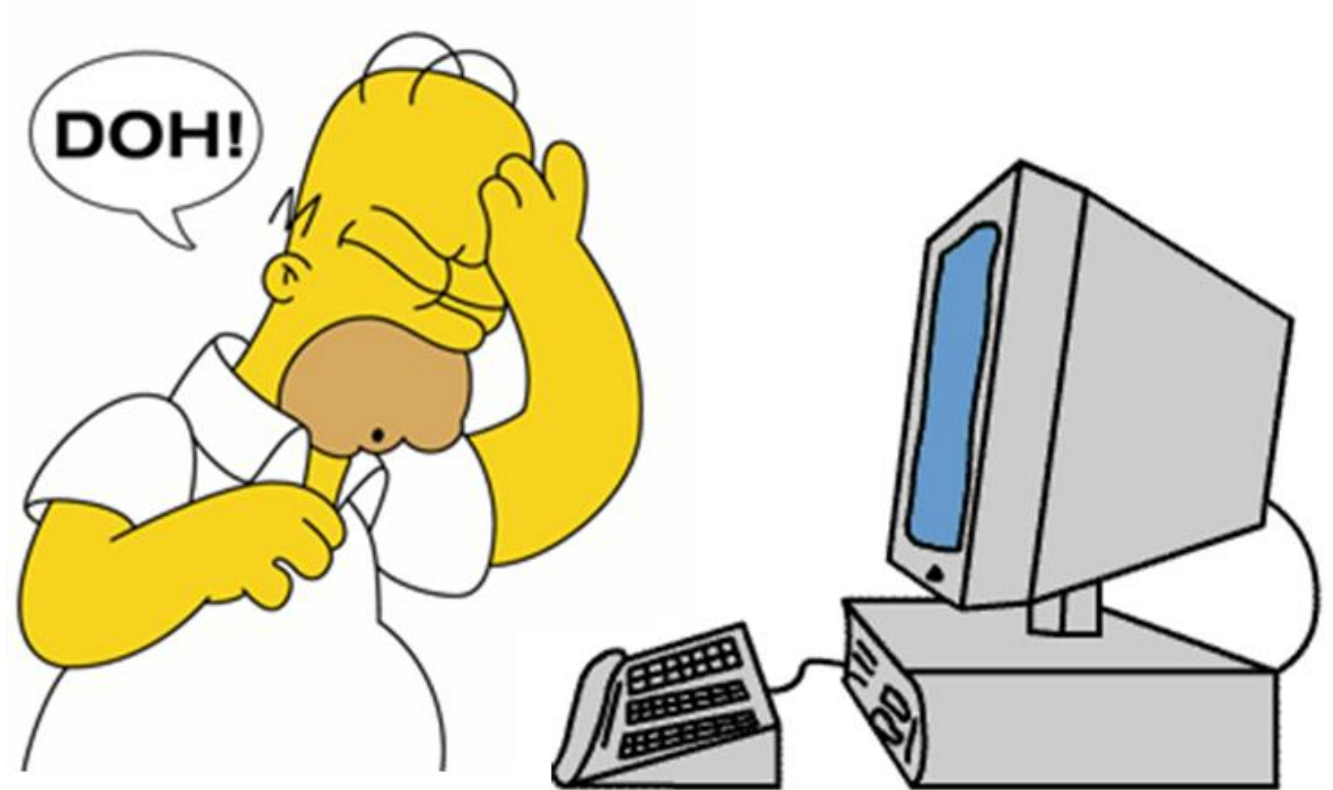
NYS · JANUARY, · · 2009	NYS · JANUARY, · · 2009	<i>null</i>	2009-01-01
	<i>null</i>	<i>null</i>	2009-01-01
MLTC · PACE · PLAN · NAME	MLTC · PACE · PLAN · NAME	PACE	2009-01-01
COMPREHENSIVE · CARE · MGMT · CORP	COMPREHENSIVE · CARE · MGMT · CORP	PACE	2009-01-01
	<i>null</i>	PACE	2009-01-01
	<i>null</i>	PACE	2009-01-01
EDDY · SENIOR · CARE	EDDY · SENIOR · CARE	PACE	2009-01-01
	<i>null</i>	PACE	2009-01-01
	<i>null</i>	PACE	2009-01-01

- Want to extract remove rows that start with “NYS”
  - `/^\s*NYS/` in Regular Expressions
- How can we reduce the *error-proneness* of this task?

# Error-proneness

“The notation invites mistakes and the system gives little protection.”

-- Blackwell et al., 2003



# Errors in row filtering

		<b>Expectation</b>	
		Deleted	Retained
<b>Reality</b>	Deleted	Success!!!	False positive
	Retained	False negative	Success!!!

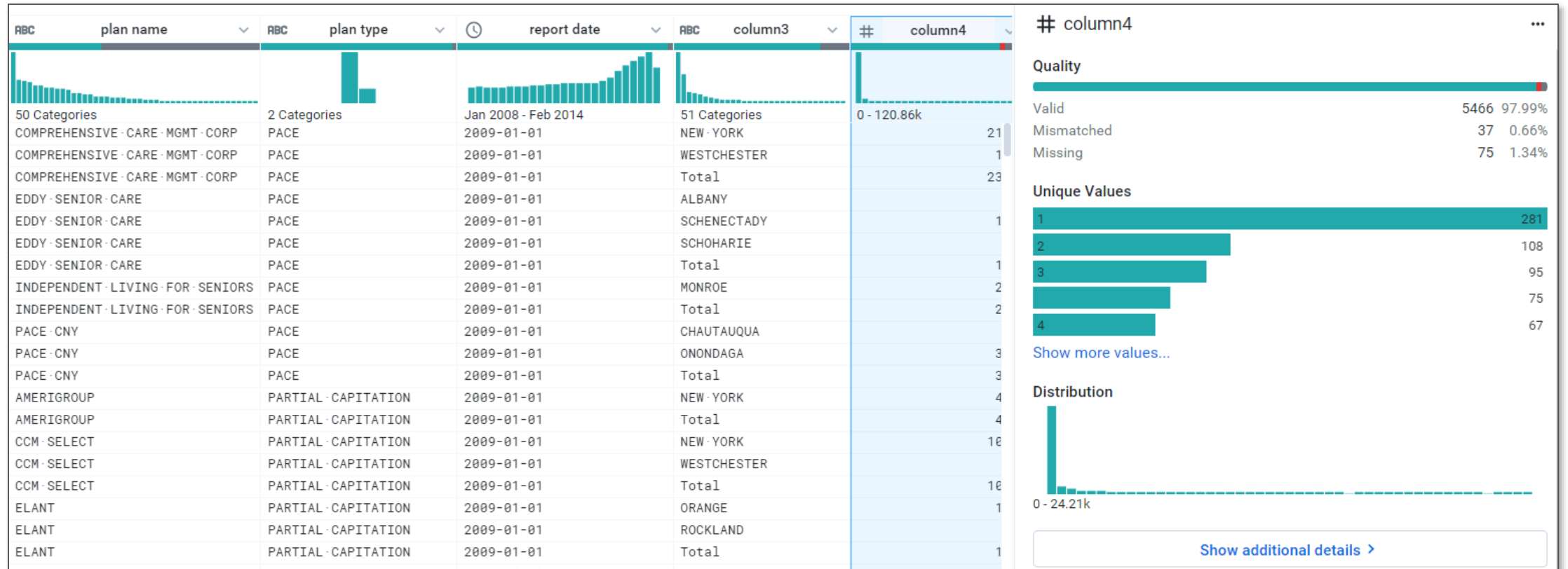


# Secondary Notation

“Extra information in means other than formal syntax”

-- Blackwell et al., 2003

# Dataprep: Visualization



# Dataprep: color stripes

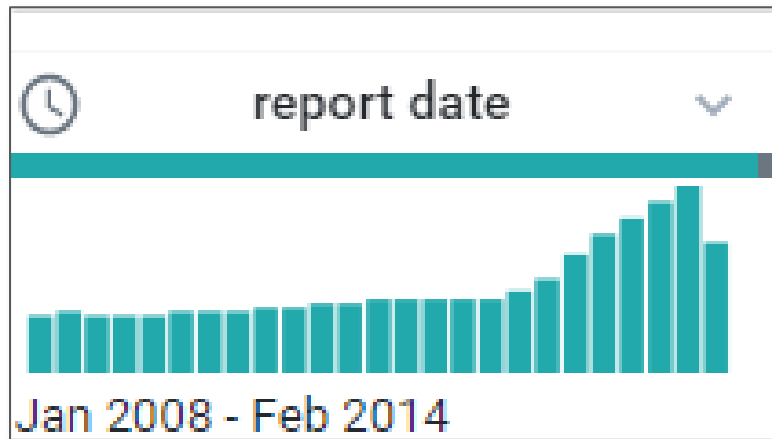
**What:** Proportion of data quality types in the column

**How:** Slacked line marks encode proportion. Color encodes category.

**Why:** To signal data quality issues



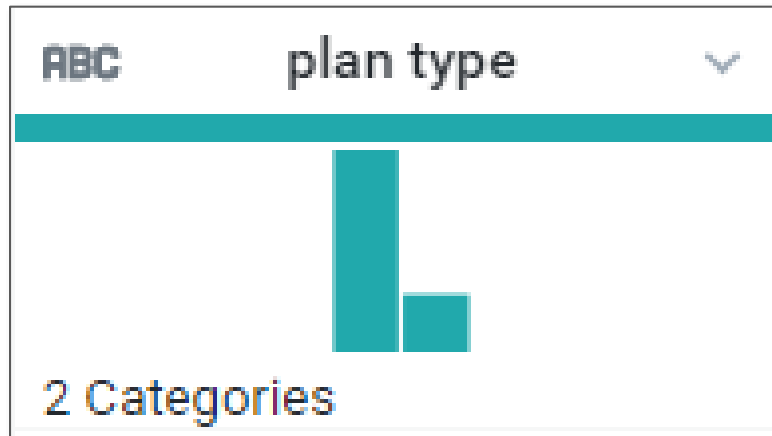
# Dataprep: bar charts and histograms



**What:** An ordered key attribute (bin), quantitative attribute (count)

**How:** Line marks where length denotes frequency and position denote bin

**Why:** Deviations from “normal”



**What:** frequency table within column

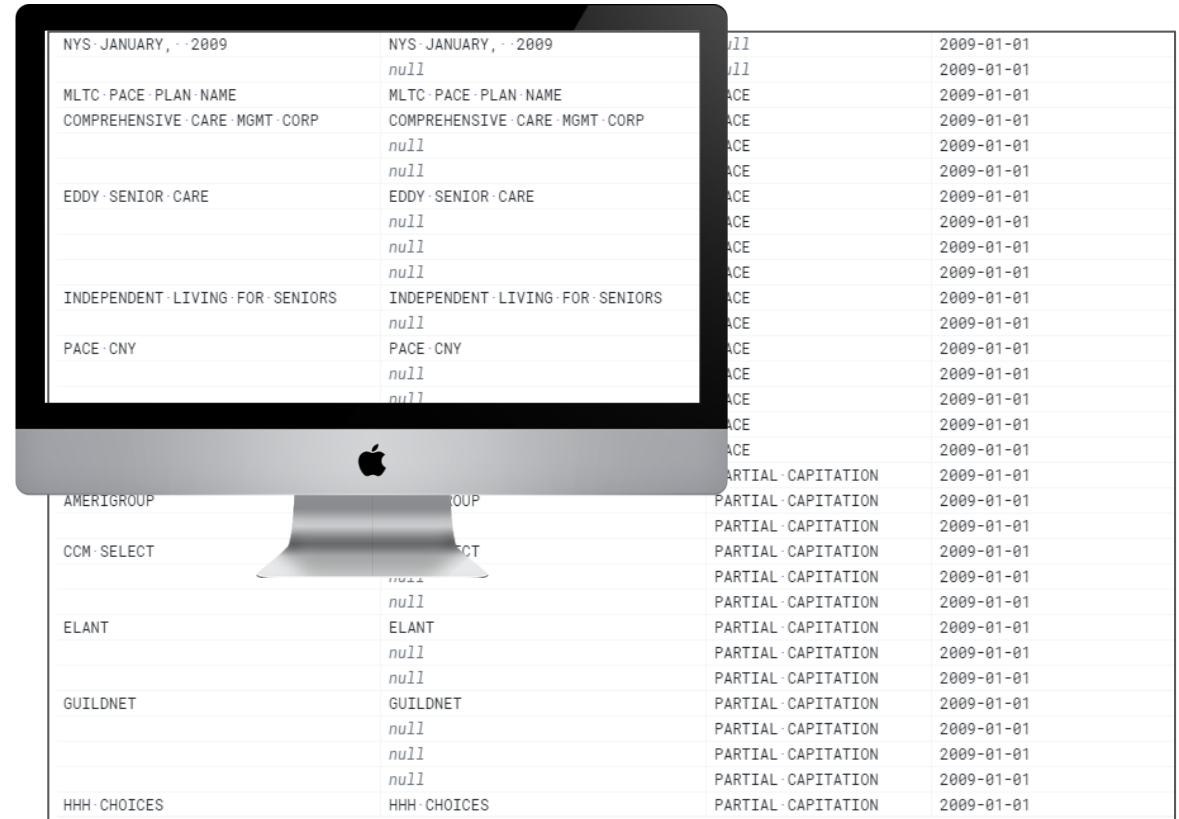
**How:** Line marks where length denotes frequency and position denote value

**Why:** Deviations from “normal”

# Visibility

“Ability to view components easily”

-- Blackwell et al., 2003



The image shows a computer monitor displaying a data table. The table is divided into two main sections. The top section is highlighted with a black border and contains the following data:

NYS - JANUARY, - 2009	NYS - JANUARY, - 2009	null	2009-01-01
MLTC - PACE - PLAN - NAME	MLTC - PACE - PLAN - NAME	null	2009-01-01
COMPREHENSIVE - CARE - MGMT - CORP	COMPREHENSIVE - CARE - MGMT - CORP	null	2009-01-01
		null	2009-01-01
EDDY - SENIOR - CARE	EDDY - SENIOR - CARE	null	2009-01-01
		null	2009-01-01
		null	2009-01-01
INDEPENDENT - LIVING - FOR - SENIORS	INDEPENDENT - LIVING - FOR - SENIORS	null	2009-01-01
		null	2009-01-01
PACE - CNY	PACE - CNY	null	2009-01-01
		null	2009-01-01

The bottom section of the table, which is not highlighted, contains the following data:

AMERIGROUP	AMERIGROUP	PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
CCM - SELECT	CCM - SELECT	PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
ELANT	ELANT	PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
GUILDNET	GUILDNET	PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
		PARTIAL - CAPITATION	2009-01-01
HHH - CHOICES	HHH - CHOICES	PARTIAL - CAPITATION	2009-01-01

# Visibility: Row Removal

## OpenRefine

74 matching rows (6636 total)

Show as: rows records    Show: 5 10 25 50 rows

All	Column 1	plan name	plan type	report date	Column 2	Column 3
4.	NYS JANUARY, 2009	NYS JANUARY, 2009		2009-01-01T00:00:00Z		
63.	NYS JANUARY, 2010	NYS JANUARY, 2010	PARTIAL CAPITATION	2010-01-01T00:00:00Z		
129.	NYS JANUARY, 2011	NYS JANUARY, 2011	PARTIAL CAPITATION	2011-01-01T00:00:00Z		
204.	NYS JANUARY, 2012	NYS JANUARY, 2012	PARTIAL CAPITATION	2012-01-01T00:00:00Z		
279.	NYS JANUARY 2013	NYS JANUARY 2013	PARTIAL CAPITATION	2013-01-01T00:00:00Z		
410.	NYS JANUARY 2014	NYS JANUARY 2014	PARTIAL CAPITATION	2014-01-01T00:00:00Z		
607.	NYS FEBRUARY, 2008	NYS FEBRUARY, 2008	PARTIAL CAPITATION	2008-02-01T00:00:00Z		
663.	NYS FEBRUARY, 2009	NYS FEBRUARY, 2009	PARTIAL CAPITATION	2009-02-01T00:00:00Z		
722.	NYS FEBRUARY, 2010	NYS FEBRUARY, 2010	PARTIAL CAPITATION	2010-02-01T00:00:00Z		
788.	NYS FEBRUARY, 2011	NYS FEBRUARY, 2011	PARTIAL CAPITATION	2011-02-01T00:00:00Z		
862.	NYS FEBRUARY, 2012	NYS FEBRUARY, 2012	PARTIAL CAPITATION	2012-02-01T00:00:00Z		
937.	NYS, FEBRUARY 2013	NYS, FEBRUARY 2013	PARTIAL CAPITATION	2013-02-01T00:00:00Z		
1080.	NYS, FEBRUARY 2014	NYS, FEBRUARY 2014	PARTIAL CAPITATION	2014-02-01T00:00:00Z		
1278.	NYS MARCH, 2008	NYS MARCH, 2008	PARTIAL CAPITATION	2008-03-01T00:00:00Z		
1334.	NYS MARCH, 2009	NYS MARCH, 2009	PARTIAL CAPITATION	2009-03-01T00:00:00Z		
1395.	NYS MARCH, 2010	NYS MARCH, 2010	PARTIAL CAPITATION	2010-03-01T00:00:00Z		
1462.	NYS MARCH, 2011	NYS MARCH, 2011	PARTIAL CAPITATION	2011-03-01T00:00:00Z		
1536.	NYS MARCH, 2012	NYS MARCH, 2012	PARTIAL CAPITATION	2012-03-01T00:00:00Z		
1611.	NYS, MARCH 2013	NYS, MARCH 2013	PARTIAL CAPITATION	2013-03-01T00:00:00Z		
1754.	NYS APRIL, 2008	NYS APRIL, 2008	PARTIAL CAPITATION	2008-04-01T00:00:00Z		
1811.	NYS APRIL, 2009	NYS APRIL, 2009	PARTIAL CAPITATION	2009-04-01T00:00:00Z		
1874.	NYS APRIL, 2010	NYS APRIL, 2010	PARTIAL CAPITATION	2010-04-01T00:00:00Z		
1941.	NYS APRIL, 2011	NYS APRIL, 2011	PARTIAL CAPITATION	2011-04-01T00:00:00Z		

## Dataprep

Preview

rbc	column2	rbc	plan name	rbc	plan type	report date	rbc	column3	#	column4
NYS JANUARY, 2009	NYS JANUARY, 2009	NYS JANUARY, 2009	NULL	PARTIAL CAPITATION	2009-01-01					
NYS JANUARY, 2010	NYS JANUARY, 2010	NYS JANUARY, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-01-01					
NYS JANUARY, 2011	NYS JANUARY, 2011	NYS JANUARY, 2011	PARTIAL CAPITATION	PARTIAL CAPITATION	2011-01-01					
NYS JANUARY, 2012	NYS JANUARY, 2012	NYS JANUARY, 2012	PARTIAL CAPITATION	PARTIAL CAPITATION	2012-01-01					
NYS, JANUARY 2013	NYS, JANUARY 2013	NYS, JANUARY 2013	PARTIAL CAPITATION	PARTIAL CAPITATION	2013-01-01					
NYS, JANUARY 2014	NYS, JANUARY 2014	NYS, JANUARY 2014	PARTIAL CAPITATION	PARTIAL CAPITATION	2014-01-01					
NYS FEBRUARY, 2008	NYS FEBRUARY, 2008	NYS FEBRUARY, 2008	PARTIAL CAPITATION	PARTIAL CAPITATION	2008-02-01					
NYS FEBRUARY, 2009	NYS FEBRUARY, 2009	NYS FEBRUARY, 2009	PARTIAL CAPITATION	PARTIAL CAPITATION	2009-02-01					
NYS FEBRUARY, 2010	NYS FEBRUARY, 2010	NYS FEBRUARY, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-02-01					
NYS FEBRUARY, 2011	NYS FEBRUARY, 2011	NYS FEBRUARY, 2011	PARTIAL CAPITATION	PARTIAL CAPITATION	2011-02-01					
NYS FEBRUARY, 2012	NYS FEBRUARY, 2012	NYS FEBRUARY, 2012	PARTIAL CAPITATION	PARTIAL CAPITATION	2012-02-01					
NYS, FEBRUARY 2013	NYS, FEBRUARY 2013	NYS, FEBRUARY 2013	PARTIAL CAPITATION	PARTIAL CAPITATION	2013-02-01					
NYS, FEBRUARY 2014	NYS, FEBRUARY 2014	NYS, FEBRUARY 2014	PARTIAL CAPITATION	PARTIAL CAPITATION	2014-02-01					
NYS MARCH, 2008	NYS MARCH, 2008	NYS MARCH, 2008	PARTIAL CAPITATION	PARTIAL CAPITATION	2008-03-01					
NYS MARCH, 2009	NYS MARCH, 2009	NYS MARCH, 2009	PARTIAL CAPITATION	PARTIAL CAPITATION	2009-03-01					
NYS MARCH, 2010	NYS MARCH, 2010	NYS MARCH, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-03-01					
NYS MARCH, 2011	NYS MARCH, 2011	NYS MARCH, 2011	PARTIAL CAPITATION	PARTIAL CAPITATION	2011-03-01					
NYS MARCH, 2012	NYS MARCH, 2012	NYS MARCH, 2012	PARTIAL CAPITATION	PARTIAL CAPITATION	2012-03-01					
NYS, MARCH 2013	NYS, MARCH 2013	NYS, MARCH 2013	PARTIAL CAPITATION	PARTIAL CAPITATION	2013-03-01					
NYS APRIL, 2008	NYS APRIL, 2008	NYS APRIL, 2008	PARTIAL CAPITATION	PARTIAL CAPITATION	2008-04-01					
NYS APRIL, 2009	NYS APRIL, 2009	NYS APRIL, 2009	PARTIAL CAPITATION	PARTIAL CAPITATION	2009-04-01					
NYS APRIL, 2010	NYS APRIL, 2010	NYS APRIL, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-04-01					
NYS APRIL, 2011	NYS APRIL, 2011	NYS APRIL, 2011	PARTIAL CAPITATION	PARTIAL CAPITATION	2011-04-01					
NYS APRIL, 2012	NYS APRIL, 2012	NYS APRIL, 2012	PARTIAL CAPITATION	PARTIAL CAPITATION	2012-04-01					
NYS, APRIL 2013	NYS, APRIL 2013	NYS, APRIL 2013	PARTIAL CAPITATION	PARTIAL CAPITATION	2013-04-01					
NYS MAY, 2008	NYS MAY, 2008	NYS MAY, 2008	PARTIAL CAPITATION	PARTIAL CAPITATION	2008-05-01					
NYS MAY, 2009	NYS MAY, 2009	NYS MAY, 2009	PARTIAL CAPITATION	PARTIAL CAPITATION	2009-05-01					
NYS MAY, 2010	NYS MAY, 2010	NYS MAY, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-05-01					
NYS MAY, 2011	NYS MAY, 2011	NYS MAY, 2011	PARTIAL CAPITATION	PARTIAL CAPITATION	2011-05-01					
NYS, MAY 2012	NYS, MAY 2012	NYS, MAY 2012	PARTIAL CAPITATION	PARTIAL CAPITATION	2012-05-01					
NYS, MAY 2013	NYS, MAY 2013	NYS, MAY 2013	PARTIAL CAPITATION	PARTIAL CAPITATION	2013-05-01					
NYS JUNE, 2008	NYS JUNE, 2008	NYS JUNE, 2008	PARTIAL CAPITATION	PARTIAL CAPITATION	2008-06-01					
NYS JUNE, 2009	NYS JUNE, 2009	NYS JUNE, 2009	PARTIAL CAPITATION	PARTIAL CAPITATION	2009-06-01					
NYS JUNE, 2010	NYS JUNE, 2010	NYS JUNE, 2010	PARTIAL CAPITATION	PARTIAL CAPITATION	2010-06-01					

6 Columns 74 Rows 3 Data Types    Show only affected

# Visibility: Row Removal

## OpenRefine

6562 matching rows (6636 total)

Show as: rows records    Show: 5 10 25 50 rows

All	Column 1	plan name	plan type	report date	Column 2
1.	MANAGED LONG TERM CARE ENROLLMENT	MANAGED LONG TERM CARE ENROLLMENT			
2.	BY PLAN, COUNTY, AND PROGRAM	BY PLAN, COUNTY, AND PROGRAM			
3.	BY PLAN, COUNTY, AND PROGRAM	BY PLAN, COUNTY, AND PROGRAM			
5.		NYS JANUARY, 2009		2009-01-01T00:00:00Z	
6.	MLTC PACE PLAN NAME	MLTC PACE PLAN NAME	PACE	2009-01-01T00:00:00Z	COUNTY
7.	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01T00:00:00Z	NEW YORK
8.	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01T00:00:00Z	WESTCHESTER
9.	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01T00:00:00Z	Total
10.	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	ALBANY
11.	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	SCHENECTADY
12.	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	SCHOHARIE
13.	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01T00:00:00Z	Total
14.	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	PACE	2009-01-01T00:00:00Z	MONROE
15.	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	PACE	2009-01-01T00:00:00Z	Total
16.	PACE CNY	PACE CNY	PACE	2009-01-01T00:00:00Z	CHAUTAUQUA
17.	PACE CNY	PACE CNY	PACE	2009-01-01T00:00:00Z	ONONDAGA
18.	PACE CNY	PACE CNY	PACE	2009-01-01T00:00:00Z	Total
19.	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	PACE	2009-01-01T00:00:00Z	
20.	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	PACE	2009-01-01T00:00:00Z	
21.	MLTC PARTIAL CAPITATION PLAN NAME	MLTC PARTIAL CAPITATION PLAN NAME	PARTIAL CAPITATION	2009-01-01T00:00:00Z	
22.	AMERIGROUP	AMERIGROUP	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK
23.	AMERIGROUP	AMERIGROUP	PARTIAL CAPITATION	2009-01-01T00:00:00Z	Total
24.	CCM SELECT	CCM SELECT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	NEW YORK
25.	CCM SELECT	CCM SELECT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	WESTCHESTER
26.	CCM SELECT	CCM SELECT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	Total
27.	ELANT	ELANT	PARTIAL CAPITATION	2009-01-01T00:00:00Z	ORANGE

## Dataprep

Preview

rbc	column2	rbc	plan name	rbc	plan type	report date	rbc	column3	#	column4
179 Categories	MANAGED LONG TERM CARE ENROLLMENT	179 Categories	MANAGED LONG TERM CARE ENROLLMENT	2 Categories	null	Jan 2008 - Feb 2014	51 Categories		0 - 120.86k	
MANAGED LONG TERM CARE ENROLLMENT	BY PLAN, COUNTY, AND PROGRAM	MANAGED LONG TERM CARE ENROLLMENT	BY PLAN, COUNTY, AND PROGRAM	null	null	null	null			
NYS JANUARY, - 2009	NYS JANUARY, - 2009	NYS JANUARY, - 2009	NYS JANUARY, - 2009	2009-01-01	2009-01-01	2009-01-01	2009-01-01			
MLTC PACE PLAN NAME	MLTC PACE PLAN NAME	MLTC PACE PLAN NAME	MLTC PACE PLAN NAME	PACE	2009-01-01	2009-01-01	2009-01-01	COUNTY		TOTAL
COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01	2009-01-01	2009-01-01	NEW YORK		ENROLLMENT
COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	COMPREHENSIVE CARE MGMT CORP	PACE	2009-01-01	2009-01-01	2009-01-01	WESTCHESTER		2184
EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01	2009-01-01	2009-01-01	Total		170
EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01	2009-01-01	2009-01-01	ALBANY		2354
EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01	2009-01-01	2009-01-01	SCHENECTADY		1
EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	EDDY SENIOR CARE	PACE	2009-01-01	2009-01-01	2009-01-01	SCHOHARIE		1
INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	PACE	2009-01-01	2009-01-01	2009-01-01	Total		110
INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	INDEPENDENT LIVING FOR SENIORS	PACE	2009-01-01	2009-01-01	2009-01-01	MONROE		246
PACE CNY	PACE CNY	PACE CNY	PACE CNY	PACE	2009-01-01	2009-01-01	2009-01-01	Total		246
PACE CNY	PACE CNY	PACE CNY	PACE CNY	PACE	2009-01-01	2009-01-01	2009-01-01	CHAUTAUQUA		1
PACE CNY	PACE CNY	PACE CNY	PACE CNY	PACE	2009-01-01	2009-01-01	2009-01-01	ONONDAGA		317
Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	PACE	2009-01-01	2009-01-01	2009-01-01	Total		318
Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	Total MLTC PACE Enrollment	PACE	2009-01-01	2009-01-01	2009-01-01			3028
MLTC PARTIAL CAPITATION PLAN NAME	MLTC PARTIAL CAPITATION PLAN NAME	MLTC PARTIAL CAPITATION PLAN NAME	MLTC PARTIAL CAPITATION PLAN NAME	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	NEW YORK		431
AMERIGROUP	AMERIGROUP	AMERIGROUP	AMERIGROUP	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	Total		431
CCM SELECT	CCM SELECT	CCM SELECT	CCM SELECT	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	NEW YORK		1040
CCM SELECT	CCM SELECT	CCM SELECT	CCM SELECT	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	WESTCHESTER		17
ELANT	ELANT	ELANT	ELANT	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	Total		1057
ELANT	ELANT	ELANT	ELANT	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	ORANGE		104
ELANT	ELANT	ELANT	ELANT	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	ROCKLAND		34
GUILDNET	GUILDNET	GUILDNET	GUILDNET	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	Total		138
GUILDNET	GUILDNET	GUILDNET	GUILDNET	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	NASSAU		519
GUILDNET	GUILDNET	GUILDNET	GUILDNET	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	NEW YORK		5810
GUILDNET	GUILDNET	GUILDNET	GUILDNET	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	SUFFOLK		217
HHH CHOICES	HHH CHOICES	HHH CHOICES	HHH CHOICES	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	Total		6546
HHH CHOICES	HHH CHOICES	HHH CHOICES	HHH CHOICES	PARTIAL CAPITATION	2009-01-01	2009-01-01	2009-01-01	NEW YORK		706

6 Columns, 6,657 Rows, 3 Data Types

Show only affected

# Conclusion

- Visualization assists in reducing error-proneness
- OpenRefine and Dataprep share a lot of the same core functionality, but Dataprep's features are richer.
- Between-table operations are still difficult, *mentally hard operations*, when confined to a within-table context.
  - See report for details