Compiling 2013> New York School data into similar format to latest years.

After downloading the 2012 data from the New York State Education Department (https://data.nysed.gov/downloads.php) and exporting the .mdb file into csv(s), this program gets these years into the same format as the latest years allowing for historical comparisons.

In [157]:
```
import agate
import csv
import agateexcel
```

Due to missing school values and changes in data columns, I needed two different typetesters to force the columns into their correct formatting.

In [158]:
```
tester_15 = agate.TypeTester(limit=100,force={
    'NRC_DESC': agate.Text(),
    'NRC_CODE': agate.Number(),
        'BEDSCODE': agate.Text(),
        'L1_COUNT': agate.Text(),
        'L1_PCT': agate.Text(),
        'L2_COUNT': agate.Text(),
        'L2_PCT': agate.Text(),
        'L3_COUNT': agate.Text(),
        'L3_PCT': agate.Text(),
        'L4_COUNT': agate.Text(),
        'L4_PCT': agate.Text(),
        'L2-L4_PCT': agate.Text(),
        'L3-L4_PCT': agate.Text(),
        'MEAN_SCALE_SCORE': agate.Text(),
        'COUNTY_DESC': agate.Text(),
        'TOTAL_TESTED': agate.Text(),
        'COUNTY_CODE': agate.Text(),
        'Sum_Of_SCALE_SCORE': agate.Text()
})
tester_16 = agate.TypeTester(limit=100,force={
    'NRC_DESC': agate.Text(),
    'NRC_CODE': agate.Number(),
        'BEDSCODE': agate.Text(),
        'L1_COUNT': agate.Text(),
        'L1_PCT': agate.Text(),
        'L2_COUNT': agate.Text(),
        'L2_PCT': agate.Text(),
        'L3_COUNT': agate.Text(),
        'L3_PCT': agate.Text(),
        'L4_COUNT': agate.Text(),
        'L4_PCT': agate.Text(),
        'L2-L4_PCT': agate.Text(),
        'L3-L4_PCT': agate.Text(),
        'MEAN_SCALE_SCORE': agate.Text(),
        'COUNTY_DESC': agate.Text(),
        'TOTAL_TESTED': agate.Text(),
        'COUNTY_CODE': agate.Text()
})
```

This program takes in the csv of a specific year's assesment data and its tester and returns a dictionary of BEDS (state id codes) and test data.

In [159]:
```
comparison_dict = {}

with open('raw/3-8_ELA_AND_MATH_2017.csv', 'r') as csvfile:
    reader = csv.DictReader(csvfile)
    for row in reader:
        if row['BEDSCODE'] not in comparison_dict and row['SUBGROUP_NAME'] == 'All
Students':
            comparison_dict[row['BEDSCODE']] = row
```

In [160]:
```
print(comparison_dict['400000000000'])
```

```
{'COUNTY_DESC': 'NIAGARA', 'L3-L4_PCT': '39%', 'L2_COUNT': '558', 'L1_PCT': '2
9%', 'L4_PCT': '7%', 'COUNTY_CODE': '40', 'SUBGROUP_CODE': '01', 'L3_COUNT': '55
1', 'SUBGROUP_NAME': 'All Students', 'NAME': 'NIAGARA COUNTY', 'TOTAL_TESTED': '
1722', 'L1_COUNT': '495', 'L4_COUNT': '118', 'ITEM_DESC': 'Grade 3 ELA', 'L2-L4_
PCT': '71%', 'NRC_CODE': '', 'NRC_DESC': '', 'SY_END_DATE': '06/30/2017', 'BEDSC
ODE': '400000000000', 'ITEM_SUBJECT_AREA': 'ELA', 'MEAN_SCALE_SCORE': '306', 'L2
_PCT': '32%', 'L3_PCT': '32%'}
```

In [169]:

```
unmatched = {}

def mani_table(table, subgroup, year_folder, clean_info0, clean_info1):
    year_key0 = year_folder[13:15]
    year_key1 = year_folder[15:17]
    for row in table.rows:
        if row['YEAR'] == '2012' and year_folder == 'school_scores2012':
            clean_dict = {}
            geoid = row['ENTITY_CD']
            if 'County' in row[1]:
                #County codes have 0001400000 where current use 140000000
                geoid = row['ENTITY_CD'][4:] + row['ENTITY_CD'][0:4]
            if geoid not in comparison_dict:
                unmatched[geoid] = row['ENTITY_NAME']
            else:
                subject = subgroup[:-1]
                grade = subgroup[-1]
                current_info = comparison_dict[geoid]
                clean_dict['COUNTY_DESC'] = current_info['COUNTY_DESC']
                clean_dict['COUNTY_CODE'] = current_info['COUNTY_CODE']
                clean_dict['SUBGROUP_NAME'] = row['SUBGROUP_NAME']
                clean_dict['BEDSCODE'] = geoid
                clean_dict['NAME'] = current_info['NAME']
                clean_dict['TOTAL_TESTED'] = row['NUM_TESTED']
                clean_dict['L1_COUNT'] = row['LEVEL1_COUNT']
                clean_dict['L2_COUNT'] = row['LEVEL2_COUNT']
                clean_dict['L3_COUNT'] = row['LEVEL3_COUNT']
                clean_dict['L4_COUNT'] = row['LEVEL4_COUNT']
                clean_dict['ITEM_DESC'] = "Grade {0} {1}".format(grade, subject)
                if subject == 'Math':
                    clean_dict['ITEM_SUBJECT_AREA'] = 'Mathematics'
                else:
                    clean_dict['ITEM_SUBJECT_AREA'] = subject
                clean_dict['MEAN_SCALE_SCORE'] = row['MEAN_SCORE']
                clean_info0.append(clean_dict)
        elif year_folder != 'school_scores2012':
            clean_dict = {}
            geoid = row['ENTITY_CD']
            if 'County' in row[1]:
                #County codes have 0001400000 where current use 140000000
                geoid = row['ENTITY_CD'][4:] + row['ENTITY_CD'][0:4]
            if geoid not in comparison_dict:
                unmatched[geoid] = row['ENTITY_NAME']
            else:
                subject = subgroup[:-1]
                grade = subgroup[-1]
                current_info = comparison_dict[geoid]
                clean_dict['COUNTY_DESC'] = current_info['COUNTY_DESC']
                clean_dict['COUNTY_CODE'] = current_info['COUNTY_CODE']
                clean_dict['SUBGROUP_NAME'] = row['SUBGROUP_NAME']
                clean_dict['BEDSCODE'] = geoid
                clean_dict['NAME'] = current_info['NAME']
                clean_dict['TOTAL_TESTED'] = row['NUM_TESTED']
                clean_dict['L1_COUNT'] = row['LEVEL1_COUNT']
                clean_dict['L2_COUNT'] = row['LEVEL2_COUNT']
                clean_dict['L3_COUNT'] = row['LEVEL3_COUNT']
                clean_dict['L4_COUNT'] = row['LEVEL4_COUNT']
                clean_dict['ITEM_DESC'] = "Grade {0} {1}".format(grade, subject)
                if subject == 'Math':
                    clean_dict['ITEM_SUBJECT_AREA'] = 'Mathematics'
                else:
                    clean_dict['ITEM_SUBJECT_AREA'] = subject
                clean_dict['MEAN_SCALE_SCORE'] = row['MEAN_SCORE']
                if row['YEAR'][2:] == year_key0:
```

```
Accessing school_scores1011
Grabbing ELA3 Subgroup Results

/Users/lheinle/anaconda/envs/py35/lib/python3.5/xml/etree/ElementTree.py:1237: R
esourceWarning: unclosed file <_io.BufferedReader name='raw/school_scores1011/EL
A6 Subgroup Results.xlsx'>
```

```
90042
clean_info0=0;clean_info1=0
Grabbing ELA4 Subgroup Results
88574
clean_info0=42449;clean_info1=43244
Grabbing ELA5 Subgroup Results
86815
clean_info0=84281;clean_info1=85662
Grabbing ELA6 Subgroup Results
67491
clean_info0=125145;clean_info1=127254
Grabbing ELA7 Subgroup Results
60036
clean_info0=156615;clean_info1=159665
Grabbing ELA8 Subgroup Results
59280
clean_info0=184502;clean_info1=188401
Grabbing Math3 Subgroup Results
90188
clean_info0=211953;clean_info1=216473
Grabbing Math4 Subgroup Results
88675
clean_info0=254461;clean_info1=259798
Grabbing Math5 Subgroup Results
86970
clean_info0=296350;clean_info1=302247
Grabbing Math6 Subgroup Results
67654
clean_info0=337289;clean_info1=343908
Grabbing Math7 Subgroup Results
60171
clean_info0=368829;clean_info1=376402
Grabbing Math8 Subgroup Results
59437
clean_info0=396789;clean_info1=405178
raw/3-8_ELA_AND_MATH_2010.csv
raw/3-8_ELA_AND_MATH_2011.csv
dict_keys(['COUNTY_DESC', 'ITEM_DESC', 'L4_COUNT', 'ITEM_SUBJECT_AREA', 'L2_COUN
T', 'BEDSCODE', 'SUBGROUP_NAME', 'NAME', 'COUNTY_CODE', 'MEAN_SCALE_SCORE', 'TOT
AL_TESTED', 'L1_COUNT', 'L3_COUNT'])
dict_keys(['COUNTY_DESC', 'ITEM_DESC', 'L4_COUNT', 'ITEM_SUBJECT_AREA', 'L2_COUN
T', 'BEDSCODE', 'SUBGROUP_NAME', 'NAME', 'COUNTY_CODE', 'MEAN_SCALE_SCORE', 'TOT
AL_TESTED', 'L1_COUNT', 'L3_COUNT'])
Accessing school_scores0809
Grabbing ELA3 Subgroup Results
88926
clean_info0=0;clean_info1=0
Grabbing ELA4 Subgroup Results
87504
clean_info0=40983;clean_info1=41756
Grabbing ELA5 Subgroup Results
86299
clean_info0=81486;clean_info1=82827
Grabbing ELA6 Subgroup Results
67139
clean_info0=121036;clean_info1=122972
Grabbing ELA7 Subgroup Results
59937
clean_info0=151234;clean_info1=153538
Grabbing ELA8 Subgroup Results
59335
clean_info0=177850;clean_info1=180748
Grabbing Math3 Subgroup Results
89078
```

In [170]:
```
print(clean_info[1359])
print(comparison_dict['490000000000'])
```

```
{'COUNTY_DESC': 'RENSSELAER', 'ITEM_DESC': 'Grade 3 ELA', 'L4_COUNT': '2', 'ITEM
_SUBJECT_AREA': 'ELA', 'L2_COUNT': '71', 'BEDSCODE': '490000000000', 'SUBGROUP_N
AME': 'Students with Disabilities', 'NAME': 'RENSSELAER COUNTY', 'COUNTY_CODE':
'49', 'MEAN_SCALE_SCORE': '643', 'TOTAL_TESTED': '216', 'L1_COUNT': '104', 'L3_C
OUNT': '39'}
{'COUNTY_DESC': 'RENSSELAER', 'L3-L4_PCT': '44%', 'L2_COUNT': '421', 'L1_PCT': '
25%', 'L4_PCT': '8%', 'COUNTY_CODE': '49', 'SUBGROUP_CODE': '01', 'L3_COUNT': '4
93', 'SUBGROUP_NAME': 'All Students', 'NAME': 'RENSSELAER COUNTY', 'TOTAL_TESTED
': '1374', 'L1_COUNT': '349', 'L4_COUNT': '111', 'ITEM_DESC': 'Grade 3 ELA', 'L2
-L4_PCT': '75%', 'NRC_CODE': '', 'NRC_DESC': '', 'SY_END_DATE': '06/30/2017', 'B
EDSCODE': '490000000000', 'ITEM_SUBJECT_AREA': 'ELA', 'MEAN_SCALE_SCORE': '309',
'L2_PCT': '31%', 'L3_PCT': '36%'}
```

In [171]:
```
print(len(unmatched.keys()))
print(unmatched)
```

619
{'280210030006': 'MILBURN ELEMENTARY SCHOOL', '660412020003': 'CLARK ACADEMY', '
321000011237': 'MARIE CURIE HIGH SCH-NURSING, MEDICINE & APPLIED HLTH PROF', '32
1100011270': 'ACADEMY FOR SCHOLARSHIP AND ENTRENEURSHIP', '571502060003': 'CANIS
TEO-GREENWOOD MIDDLE SCHOOL', '140600010103': 'GROVER CLEVELAND HIGH SCHOOL', '4
01301040004': 'BARKER MIDDLE SCHOOL', '310200860905': 'ROSS GLOBAL ACADEMY CHART
ER SCHOOL', '001800000000': 'Similar Schools Group #18', '491302060003': 'GEORGE
WASHINGTON SCHOOL', '651201060002': 'SODUS INTERMEDIATE SCHOOL', '007400000000':
'Similar Schools Group #74', '660401030002': 'W L MORSE SCHOOL', '180300010004':
'ROBERT MORRIS SCHOOL', '251601060005': 'CHITTENANGO HIGH SCHOOL', '460801060004
': 'CLEVELAND ELEMENTARY SCHOOL', '321200011270': 'ACADEMY FOR SCHOLARSHIP AND E
NTRENEURSHIP', '331700860951': 'FAHARI ACADEMY CHARTER SCHOOL', '091101060006':
'PERU MIDDLE SCHOOL', '331300010103': 'SATELLITE THREE', '261600010036': 'SCHOOL
36-HENRY W LONGFELLOW', '530600010020': 'CENTRAL PARK MIDDLE SCHOOL', '580235060
006': 'BELLPORT SENIOR HIGH SCHOOL', '670201060001': 'ATTICA SENIOR HIGH SCHOOL
', '070600010014': 'PINE CITY SCHOOL', '342800010030': 'PS 30', '331600010385':
'SCHOOL OF BUSINESS FINANCE & ENTREPRENEURSHIP', '070600010007': 'FASSETT ELEMEN
TARY SCHOOL', '042400010010': 'IVERS J NORTON ELEMENTARY SCHOOL', '131601060007
': 'ARLINGTON MIDDLE SCHOOL', '310500860963': 'NEW YORK FRENCH-AMERICAN CHARTER
SCHOOL', '142500010010': 'TONAWANDA MIDDLE SCHOOL', '310500010172': 'POWELL MIDD
LE SCHOOL FOR LAW AND SOCIAL JUSTICE', '580503030001': 'CONNETQUOT ELEMENTARY SC
HOOL', '331900010166': 'JHS 166 GEORGE GERSHWIN', '332300010634': 'GENERAL D CHA
PPIE JAMES MIDDLE SCHOOL OF SCIENCE', '051901040005': 'UNION SPRINGS MIDDLE SCHO
OL HIGH SCHOOL', '210402060001': 'FRANKFORT SCHUYLER CENTRAL HIGH SCHOOL', '3314
00010019': 'PS 19 ROBERTO CLEMENTE', '332200011468': 'KINGSBOROUGH EARLY COLLEGE
SCHOOL', '421101060003': 'MARCELLUS HIGH SCHOOL', '310500010410': 'URBAN ASSEMBL
Y INSTITUTE FOR NEW TECHNOLOGIES', '560501040005': 'SOUTH SENECA MIDDLE SCHOOL',
'002700000000': 'Similar Schools Group #27', '280211030010': 'OCEANSIDE SENIOR H
IGH SCHOOL', '500201060003': 'NORTH GARNERVILLE ELEMENTARY SCHOOL', '47160104000
2': 'OTEGO ELEMENTARY SCHOOL', '580205060001': 'GATELOT AVENUE SCHOOL', '1426010
30021': 'THOMAS JEFFERSON ELEMENTARY SCHOOL', '010605060006': 'LATHAM RIDGE SCHO
OL', '140600010120': 'CAMPUS WEST SCHOOL', '280410030004': 'MEADOW DRIVE SCHOOL
', '331800011500': 'CANARSIE HIGH SCHOOL', '070600010015': 'RIVERSIDE SCHOOL', '
401501060001': 'W H STEVENSON ELEMENTARY SCHOOL', '261600010069': 'SCHOOL WITHOU
T WALLS', '332300010055': 'IS 55 OCEAN HILL BROWNSVILLE INTERMEDIATE SCHOOL', '2
51400010001': 'SYLVAN-VERONA BEACH ELEMENTARY SCHOOL', '310600010164': 'I.S. 164
EDWARD W. STITT SCHOOL', '140203060010': 'WILLIAMSVILLE NORTH HIGH SCHOOL', '030
701060004': 'PORT DICKINSON ELEMENTARY SCHOOL', '580405060015': 'CHESTNUT HILL E
LEMENTARY SCHOOL', '140600010039': 'DR MARTIN LUTHER KING, JR MULTICULTURAL INST
ITUTE', '231101040004': 'CONSTABLEVILLE ELEMENTARY SCHOOL', '005100000000': 'Sim
ilar Schools Group #51', '571000010014': 'NORTHSIDE BLODGETT MIDDLE SCHOOL', '18
0300010003': 'JACKSON SCHOOL', '500401060009': 'SUFFERN SENIOR HIGH SCHOOL', '32
0800011282': "YOUNG WOMEN'S LEADERSHIP SCHOOL-BRONX CAMPUS", '261600010061': 'EA
ST HIGH SCHOOL', '441600010015': 'NORTH JUNIOR HIGH SCHOOL', '310400010101': 'PS
101 ANDREW DRAPER SCHOOL', '671201060003': 'PERRY MIDDLE SCHOOL', '006200000000
': 'Similar Schools Group #62', '650902040001': 'RUBEN A CIRILLO HIGH SCHOOL', '
081401040002': 'OTSELIC VALLEY ELEMENTARY SCHOOL', '320900011250': 'EXIMIUS COLL
EGE PREPARATORY ACADEMY', '006600000000': 'Similar Schools Group #66', '12050104
0003': 'ABRAHAM KELLOGG ELEMENTARY SCHOOL', '000300000000': 'Similar Schools Gro
up #3', '140600860857': 'SANKOFA CHARTER SCHOOL', '004300000000': 'Similar Schoo
ls Group #43', '320800860962': 'METROPOLITAN LIGHTHOUSE CHARTER SCHOOL', '002400
000000': 'Similar Schools Group #24', '043001040001': 'RANDOLPH MIDDLE SCHOOL',
'511101060006': 'EAST SIDE ELEMENTARY SCHOOL', '332300011493': 'BROOKLYN COLLEGI
ATE-A COLLEGE BOARD SCHOOL', '661500010004': 'WOODSIDE SCHOOL', '591201040004':
'TRI-VALLEY MIDDLE SCHOOL', '591401060001': 'CORNELIUS DUGGAN SCHOOL', '24170104
0005': 'YORK MIDDLE SCHOOL', '490101040003': 'STEPHENTOWN ELEMENTARY SCHOOL', '2
80215030007': 'LAWRENCE SENIOR HIGH SCHOOL', '190301040002': 'DURHAM ELEMENTARY
SCHOOL', '500402060018': 'RAMAPO HIGH SCHOOL', '003600000000': 'Similar Schools
Group #36', '280210030009': 'SHUBERT ELEMENTARY SCHOOL', '321000011324': 'BRONX
EARLY COLLEGE ACADEMY FOR TEACHING AND LEARNING', '310500010195': 'IS 195 ROBERT
O CLEMENTE', '662300010050': 'RIVERSIDE HIGH SCHOOL', '042400010005': 'BOARDMANV
ILLE ELEMENTARY SCHOOL', '310200010414': 'NYC MUSEUM SCHOOL', '002800000000': 'S
imilar Schools Group #28', '321100010322': 'ASPIRE PREPARATORY MIDDLE SCHOOL', '
261600010076': 'BIOSCIENCE & HEALTH CAREER HS AT FRANKLIN', '007000000000': 'Sim

In [ ]: