

Team 82 Final EDA

Team 82 project examines factors affecting students test scores in grades 4 and 8.

We examine parent social and economic demographics, school funding and community factors

```
In [1]: import pandas as pd
import numpy as np
import os
import json
import matplotlib.pyplot as plt
import seaborn as sns
import math

In [2]: dfs= ('df_1011','df_1112','df_1213','df_1314','df_1415','df_1516','df_1617','df_1718')
df1616= ('df_1617','df_1718')

In [3]: math_df_201011 = pd.read_csv('./data_sets/clean/math_scores_sy1011.csv')
math_df_201112 = pd.read_csv('./data_sets/clean/math_scores_sy1112.csv')
math_df_201213 = pd.read_csv('./data_sets/clean/math_scores_sy1213.csv')
math_df_201314 = pd.read_csv('./data_sets/clean/math_scores_sy1314.csv')
math_df_201415 = pd.read_csv('./data_sets/clean/math_scores_sy1415.csv')
math_df_201516 = pd.read_csv('./data_sets/clean/math_scores_sy1516.csv')
math_df_201617 = pd.read_csv('./data_sets/clean/math_scores_sy1617.csv')
math_df_201718 = pd.read_csv('./data_sets/clean/math_scores_sy1718.csv')

r1a_df_201011 = pd.read_csv('./data_sets/clean/r1a_scores_sy1011.csv')
r1a_df_201112 = pd.read_csv('./data_sets/clean/r1a_scores_sy1112.csv')
r1a_df_201213 = pd.read_csv('./data_sets/clean/r1a_scores_sy1213.csv')
r1a_df_201314 = pd.read_csv('./data_sets/clean/r1a_scores_sy1314.csv')
r1a_df_201415 = pd.read_csv('./data_sets/clean/r1a_scores_sy1415.csv')
r1a_df_201516 = pd.read_csv('./data_sets/clean/r1a_scores_sy1516.csv')
r1a_df_201617 = pd.read_csv('./data_sets/clean/r1a_scores_sy1617.csv')
r1a_df_201718 = pd.read_csv('./data_sets/clean/r1a_scores_sy1718.csv')

parent_social_econ_by_district_df = pd.read_csv('./data_sets/clean/teacher_parentdy_econ_by_district_df.csv')

/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (20,24,216) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (19,21,25) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (18,20,22,24,26,38,42,44,46,50,124,128,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (9,10,12,14,16,18,20,22,24,26,28,30,32,33,34,36,40,42,46,52,54,56,58,60,62,64,66,70,72,74,76,78,80,82,84,86,88,90,92,94,96,98,100,102,104,106,108,110,112,114,116,118,120,122,124,126,128,130,132,134,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (18,20,22,24,26,38,42,44,46,50,124,128,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (18,20,22,24,26,38,42,44,46,50,124,128,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (18,20,22,24,26,38,42,44,46,50,124,128,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
/opt/cuda/lib/python3.6/site-packages/ipython/core/interactiveshell.py:3020: DtypeWarning: Columns (18,20,22,24,26,38,42,44,46,50,124,128,136,138,140,142,144,146,148,150,152,154,156,158,160,162,164,166,168,170,172,174,176,178,180,182,184,186,188,190,192,194,196,198,200,202,204,206,208,210,212,214,216,218,220,222,224,226,228,230,232,234,236,238,240,242,244,246,248,250,252,254,256) have mixed types. Specify dtype option on import or set low_memory=False.
```

Math Analysis for School Year 2010 - 2011

# Math scores by school district									
math_df_201011									
	STNAM	FIPST	LEAID	leannm0	ALL_MTH04numvalid_1011	ALL_MTH04pctprof_1011	ECD_MTH04numvalid_1011	ECD_MTH04pctprof_1011	
0	ALABAMA	1	100002	Alabama Youth Services		NaN		NaN	
1	ALABAMA	1	100005	Albertville City		320.0		89.0	
2	ALABAMA	1	100006	Marshall County		422.0		91.0	
3	ALABAMA	1	100007	Hoover City		1036.0		86.0	
4	ALABAMA	1	100008	Madison City		606.0		89.0	
5	ALABAMA	1	100009	Al Just Deaf And Blind					
6	ALABAMA	1	100011	Leeds City		113.0		77.0	
7	ALABAMA	1	100012	Boaz City		161.0		77.0	
8	ALABAMA	1	100013	Trussville City		320.0		90.0	
9	ALABAMA	1	100018	Alabama School of Fine Arts		NaN		NaN	
10	ALABAMA	1	100030	Alexander City		237.0		87.0	
11	ALABAMA	1	100033	Glenwood Mental Health Learning Tree Inc		NaN		NaN	
12	ALABAMA	1	100035	Leaning Tree		5.0		NaN	
13	ALABAMA	1	100042	Laurel Oaks Behavioral Health		NaN		NaN	
14	ALABAMA	1	100060	Andalusia City		120.0		87.0	
15	ALABAMA	1	100090	Anniston City		194.0		67.0	
16	ALABAMA	1	100100	Arab City		171.0		2.0	
17	ALABAMA	1	100120	Athens City		238.0		82.0	
18	ALABAMA	1	100180	Attalla City		83.0		82.0	
19	ALABAMA	1	100183	Alabama Classical School		NaN		NaN	
20	ALABAMA	1	100185	Saraland City		148.0		82.0	
21	ALABAMA	1	100186	Father Purcell Center		2.0		NaN	
22	ALABAMA	1	100187	Father Walters Ctr		2.0		NaN	
23	ALABAMA	1	100210	Auburn City		562.0		91.0	
24	ALABAMA	1	100240	Autauga County		79.0		91.0	
25	ALABAMA	1	100270	Baldwin County		2263.0		86.0	
26	ALABAMA	1	100300	Barbour County		92.0		52.0	
27	ALABAMA	1	100330	Bessemer City		374.0		80.0	
28	ALABAMA	1	100360	Bibb County		280.0		72.0	
29	ALABAMA	1	100390	Birmingham City		2075.0		73.0	
...	
19989	BUREAU OF INDIAN AFFAIRS	59	5900169	Tohono O'odham High School		NaN		NaN	
19990	BUREAU OF INDIAN AFFAIRS	59	5900170	Toneale Day School		22.0		30.0	
19991	BUREAU OF INDIAN AFFAIRS	59	5900171	Rough Rock Community School		22.0		10.0	
19992	BUREAU OF INDIAN AFFAIRS	59	5900172	Red Water Elementary School		12.0		25.0	
19993	BUREAU OF INDIAN AFFAIRS	59	5900173	Little Wound School		42.0		15.0	
19994	BUREAU OF INDIAN AFFAIRS	59	5900174	Taos Day School		17.0		50.0	
19995	BUREAU OF INDIAN AFFAIRS	59	5900175	Keams Canyon Elementary School		8.0		25.0	
19996	BUREAU OF INDIAN AFFAIRS	59	5900176	Lummi Tribal School System		14.0		25.0	
19997	BUREAU OF INDIAN AFFAIRS	59	5900177	Wist'oozi'brolta		28.0		30.0	
19998	BUREAU OF INDIAN AFFAIRS	59	5900178	Two Eagle River School		NaN		NaN	
19999	BUREAU OF INDIAN AFFAIRS	59	5900179	Chi Chi'ah Comm School		9.0		25.0	
16000	BUREAU OF INDIAN AFFAIRS	59	5900180	Kaibeto School		34.0		24.0	
16001	BUREAU OF INDIAN AFFAIRS	59	5900181	Sitting Bull School		18.0		10.0	
16002	BUREAU OF INDIAN AFFAIRS	59	5900182	Greasewood Springs Community School		21.0		10.0	
16003	BUREAU OF INDIAN AFFAIRS	59	5900183	Wounded Knee District School		15.0		25.0	
16004	BUREAU OF INDIAN AFFAIRS	59	5900184	Laguna Elementary School		39.0		15.0	
16005	BUREAU OF INDIAN AFFAIRS	59	5900185	Wa He Lut Indian School		14.0		25.0	
16006	BUREAU OF INDIAN AFFAIRS	59	5900186	Casa Blanca Community School		37.0		34.0	
16007	BUREAU OF INDIAN AFFAIRS	59	5900187	Hannahville Indian School		11.0		25.0	
16008	BUREAU OF INDIAN AFFAIRS	59	5900188	Beclitcho Day School		8.0		25.0	
16009	BUREAU OF INDIAN AFFAIRS	59	5900189	Mandaree Day School		18.0		10.0	
16010	BUREAU OF INDIAN AFFAIRS	59	5900190	Tiospaye Tribal School		15.0		25.0	
16011	BUREAU OF INDIAN AFFAIRS	59	5900191	To'hajilee Day School		22.0		10.0	
16012	BUREAU OF INDIAN AFFAIRS	59	5900192	Gila Crossing Community School		48.0		44.0	
16013	BUREAU OF INDIAN AFFAIRS	59	5900193	Saba Dabal Boarding School		9.0		25.0	
16014	BUREAU OF INDIAN AFFAIRS	59	5900194	Bogwe Chitto Elementary School		18.0		50.0	
16015	BUREAU OF INDIAN AFFAIRS	59	5900195	Skunk City Community School		25.0		30.0	
16016	BUREAU OF INDIAN AFFAIRS	59	5900196	Meskwaiki Settlement School		17.0		70.0	
16017	BUREAU OF INDIAN AFFAIRS	59	5900197	Noll School		NaN		NaN	
16018	PUERTO RICO DEPARTMENT OF EDUCATION	72	7200303	PUERTO RICO DEPARTMENT OF EDUCATION		37732.0		52.0	
16019 rows x 16 columns									

```
In [5]: math_df_201011_clean = math_df_201011.dropna()
math_df_201112_clean = math_df_201112.dropna()
math_df_201213_clean = math_df_201213.dropna()
math_df_201314_clean = math_df_201314.dropna()
math_df_201415_clean = math_df_201415.dropna()
math_df_201516_clean = math_df_201516.dropna()
math_df_201617_clean = math_df_201617.dropna()
math_df_201718_clean = math_df_201718.dropna()

In [6]: math_states_2011 = math_df_201011_clean.groupby(by=["STNAM"]).mean()
math_states_2012 = math_df_201112_clean.groupby(by=["STNAM"]).mean()
math_states_2013 = math_df_201213_clean.groupby(by=["STNAM"]).mean()
math_states_2014 = math_df_201314_clean.groupby(by=["STNAM"]).mean()

In [7]: math_states_2011
```

	DELAWARE	10.0	1.000745e+06	640.666667	66.000000
	DISTRICT OF COLUMBIA	11.0	1.100330e+06	3378.000000	46.000000
	FLORIDA	12.0	1.201014e+06	3983.291667	74.479167
	GEORGIA	13.0	1.302368e+06	1762.645833	81.708333
	HAWAII	15.0	1.500030e+06	13492.000000	61.000000
	IDAHO	16.0	1.601595e+06	1172.666667	85.000000
	ILLINOIS	17.0	1.721418e+06	1346.866667	84.400000
	INDIANA	18.0	1.805895e+06	839.296296	76.000000
	IOWA	19.0	1.916215e+06	698.666667	74.500000
	KANSAS	20.0	2.008843e+06	959.733333	83.600000
	KENTUCKY	21.0	2.102909e+06	609.333333	76.500000
	LOUISIANA	22.0	2.200839e+06	1128.877195	69.487179
	MAINE	23.0	2.314670e+06	238.000000	67.000000
	MARYLAND	24.0	2.400348e+06	3819.800000	91.466667
	MASSACHUSETTS	25.0	2.506938e+06	922.950000	34.750000
	MICHIGAN	26.0	2.621225e+06	482.500000	61.437500
	MINNESOTA	27.0	2.723353e+06	1248.181818	83.181818
	MISSISSIPPI	28.0	2.802482e+06	518.600000	62.266667
	MISSOURI	29.0	2.920111e+06	643.406250	46.812500
	MONTANA	30.0	3.011220e+06	849.500000	73.000000
	NEBRASKA	31.0	3.158413e+06	1528.600000	65.400000
	NEVADA	32.0	3.200210e+06	1007.300000	72.333333
	NEW HAMPSHIRE	33.0	3.304590e+06	1020.000000	59.000000
	NEW JERSEY	34.0	3.408198e+06	901.625000	61.000000
	NEW MEXICO	35.0	3.501083e+06	1454.000000	43.333333
	NEW YORK	36.0	3.609020e+06	1556.982143	58.928571
	NORTH CAROLINA	37.0	3.702164e+06	2703.526316	80.947368
	OHIO	39.0	3.904527e+06	958.428571	68.809524
	OREGON	41.0	4.106883e+06	754.125000	67.833333
	PENNSYLVANIA	42.0	4.215874e+06	1261.350000	74.800000
	PUERTO RICO	72.0	7.200030e+06	3773.200000	52.000000
	SOUTH CAROLINA	45.0	4.502493e+06	1517.714286	78.095238
	SOUTH DAKOTA	46.0	4.663045e+06	1260.000000	78.000000
	TENNESSEE	47.0	4.702629e+06	2697.833333	43.083333
	TEXAS	48.0	4.823379e+06	2019.383838	85.474747
	UTAH	49.0	4.900614e+06	219.187500	79.375000
	VIRGINIA	51.0	5.102195e+06	2037.812500	86.812500
	WASHINGTON	53.0	5.304895e+06	885.433962	55.603774
	WEST VIRGINIA	54.0	5.400729e+06	709.300000	45.000000
	WISCONSIN	55.0	5.508017e+06	1689.444444	75.000000
	WYOMING	56.0	5.602990e+06	760.500000	81.000000
8	# MTH04 ~ Math Grade 4 MTH08 ~ Math Grade 8 # ECD ~ Economically Disadvantaged HOM ~ Homeless math_states_2011.columns				
9	Index(['FIPST','LEAID','ALL_MTH04numvalid_1011','ALL_MTH04pctprof_1011', 'ECD_MTH04numvalid_1011','ECD_MTH04pctprof_1011', 'HOM_MTH04numvalid_1011','HOM_MTH04pctprof_1011', 'ALL_MTH08numvalid_1011','ALL_MTH08pctprof_1011', 'ECD_MTH08numvalid_1011','ECD_MTH08pctprof_1011', 'HOM_MTH08numvalid_1011','HOM_MTH08pctprof_1011'], dtype='object')				
9	print (math_states_2011['ALL_MTH04numvalid_1011','ALL_MTH04pctprof_1011', 'ECD_MTH04numvalid_1011','ECD_MTH04pctprof_1011', 'HOM_MTH04numvalid_1011','HOM_MTH04pctprof_1011', 'ALL_MTH08numvalid_1011','ALL_MTH08pctprof_1011', 'ECD_MTH08numvalid_1011','ECD_MTH08pctprof_1011', 'HOM_MTH08numvalid_1011','HOM_MTH08pctprof_1011']				
	count	ALL_MTH04numvalid_1011	ALL_MTH04pctprof_1011	ECD_MTH04numvalid_1011	ECD_MTH04pctprof_1011
	mean	49.000000	49.000000	49.000000	49.000000
	std	2507.321656	68.438518	1587.364669	1587.364669
	min	5618.352658	13.908299	4119.035556	4119.035556
	25%	38.800000	34.750000	38.800000	38.800000
	50%	839.296296	61.000000	402.041667	402.041667
	75%	1172.666667	69.487179	687.375000	687.375000
	max	37732.000000	91.466667	2816.000000	2816.000000
	count	ECD_MTH04pctprof_1011	HOM_MTH04numvalid_1011	HOM_MTH04pctprof_1011	ALL_MTH08numvalid_1011
	mean	49.000000	49.000000	49.000000	49.000000
	std	61.520052	58.026897	44.528995	44.528995
	min	14.807518	116.242878	10.330968	10.330968
	25%	26.850000	6.000000	17.300000	17.300000
	50%	62.000000	24.062500	27.000000	27.000000
	75%	63.923077	33.047619	42.285714	42.285714
	max	73.600000	53.111111	47.937500	47.937500
		85.000000	823.000000	62.062500	62.062500
	count	ALL_MTH08numvalid_1011	ALL_MTH08pctprof_1011	ECD_MTH08numvalid_1011	ECD_MTH08pctprof_1011
	mean	49.000000	49.000000	49.000000	49.000000
	std	2389.067924	60.422706	1390.749615	1390.749615
	min	5570.598002	16.311394	3735.671966	3735.671966
	25%	35.400000	9.000000	35.400000	35.400000
	50%	774.370370	49.482143	687.333333	687.333333
	75%	1086.500000	64.000000	612.282051	612.282051
	max	1635.883333	72.777778	900.911111	900.911111
		37650.000000	86.708333	25919.000000	25919.000000
	count	ECD_MTH08pctprof_1011	HOM_MTH08numvalid_1011	HOM_MTH08pctprof_1011	ALL_MTH04numvalid_1011
	mean	49.000000	49.000000	49.000000	49.000000
	std	61.311983	62.229522	39.484141	39.484141
	min	16.079022	109.552137	8.867172	8.867172
	25%	8.000000	7.000000	6.000000	6.000000
	50%	40.000000	19.481481	27.000000	27.000000
	75%	53.266667	25.500000	32.571429	32.571429
	max	60.666667	39.222222	40.416667	40.416667
		82.520833	779.000000	50.000000	50.000000
10	#df.sort_values(by='coll1')				


```

'pct_Below PovtLvl_Age_18_64', 'pctmoe_Below PovtLvl_Age_18_64',
'pct_Below PovtLvl_Age_gte_65', 'pctmoe_Below PovtLvl_Age_gte_65',
'LEA_NAME', 'ST_NAME', 'LEA_CWTFTEST', 'LEA_CWTFTEST', 'GeoId_y',
'Geography_y', 'Year_y', 'Iteration_y', 'num_Educational_Attain_POP',
'num_Educational_Attain_POP_L9th', 'pct_Educational_Attain_POP_L9th',
'num_Educational_Attain_POP_9th-12th',
'pct_Educational_Attain_POP_9th-12th',
'num_Educational_Attain_POP_HS_GRAD',
'pct_Educational_Attain_POP_HS_GRAD',
'num_Educational_Attain_POP_AssocDeg',
'pct_Educational_Attain_POP_AssocDeg',
'num_Educational_Attain_POP_BacDeg',
'pct_Educational_Attain_POP_BacDeg',
'num_Educational_Attain_POP_GradProf',
'pct_Educational_Attain_POP_GradProf',
'PDP02.5_3Test',
'pct_Educational_Attain_POP_HS_Grad_higher', 'PDP02.5_3Best',
'pct_Educational_Attain_BS_Deg_higher',
dtype='object')

```

```

In [15]: print (parent_social_econ_by_district_df_clean[['pct_Below PovtLvl_All_Ages', 'pctmoe_Below PovtLvl_All_Ages', 'pctmoe_Below PovtLvl_Age_gte_18',
'pct_Below PovtLvl_Age_gte_65', 'pctmoe_Below PovtLvl_Age_gte_18',
'LEA_NAME', 'ST_NAME', 'LEA_CWTFTEST', 'LEA_CWTFTEST', 'GeoId_y',
'Geography_y', 'Year_y', 'Iteration_y', 'num_Educational_Attain_POP_L9th',
'num_Educational_Attain_POP_9th-12th',
'pct_Educational_Attain_POP_9th-12th',
'num_Educational_Attain_POP_HS_GRAD',
'pct_Educational_Attain_POP_HS_GRAD',
'num_Educational_Attain_POP_AssocDeg',
'pct_Educational_Attain_POP_AssocDeg',
'num_Educational_Attain_POP_BacDeg',
'pct_Educational_Attain_POP_BacDeg',
'num_Educational_Attain_POP_GradProf',
'pct_Educational_Attain_POP_GradProf',
'PDP02.5_3Test',
'pct_Educational_Attain_POP_HS_Grad_higher', 'PDP02.5_3Best',
'pct_Educational_Attain_BS_Deg_higher']]).describe())

pct_Below PovtLvl_All_Ages      pctmoe_Below PovtLvl_Age_gte_18      \
count      5920.000000      5920.000000
mean      12.172010      4.677044
std      8.719912      3.066878
min      0.000000      0.200000
25%      5.500000      2.400000
50%      8.719912      4.000000
75%      17.000000      6.100000
max      62.500000      25.900000

pct_Below PovtLvl_Age_gte_18      pctmoe_Below PovtLvl_Age_gte_18      \
count      5920.000000      5920.000000
mean      12.168193      4.675642
std      8.719912      3.066471
min      0.000000      0.200000
25%      5.500000      2.400000
50%      8.719912      4.000000
75%      17.000000      6.100000
max      62.500000      25.900000

pct_Below PovtLvl_Age_18_64      pctmoe_Below PovtLvl_Age_18_64      \
count      5920.000000      5920.000000
mean      12.185270      4.677061
std      8.745468      3.095060
min      0.000000      0.200000
25%      5.500000      2.400000
50%      8.719912      4.000000
75%      16.925000      6.200000
max      63.600000      26.000000

pct_Below PovtLvl_Age_gte_65      pctmoe_Below PovtLvl_Age_gte_65      \
count      5920.000000      5920.000000
mean      9.897382      67.957973
std      25.047672      27.840805
min      0.000000      0.900000
25%      0.000000      46.200000
50%      0.000000      4.000000
75%      0.000000      100.000000
max      100.000000      100.000000

LEA_CWTFTEST      LEA_CWTFTEST      ...      \
count      5920.000000      5920.000000      ...
mean      0.909723      0.024822      ...
std      0.116968      0.014627      ...
min      0.000000      0.003000      ...
25%      0.823000      0.012000      ...
50%      0.880000      0.024000      ...
75%      0.980000      0.035000      ...
max      1.424000      0.089000      ...

num_Educational_Attain_POP_AssocDeg      \
count      5920.000000
mean      11.633041
std      2074.923142
min      0.000000
25%      120.000000
50%      290.000000
75%      676.250000
max      98245.000000

pc_Educational_Attain_POP_AssocDeg      num_Educational_Attain_POP_GradProf      \
count      5920.000000      5920.000000
mean      10.488901      6.579603e+03
std      4.918035      5333.390163
min      0.000000      0.000000
25%      7.000000      165.000000
50%      87.000000      245.000000
75%      13.000000      1480.000000
max      38.000000      235730.000000

pc_Educational_Attain_POP_BacDeg      num_Educational_Attain_POP_GradProf      \
count      5920.000000      5920.000000
mean      19.474324      6.779444
std      9.966001      3.066471
min      0.000000      0.000000
25%      12.000000      0.000000
50%      18.000000      255.000000
75%      26.000000      180210.000000
max      60.000000      180210.000000

pc_Educational_Attain_POP_GradProf      PDP02.5_3Test      \
count      5920.000000      5.920000e+03
mean      11.633041      6.579603e+03
std      10.301973      2.089708e+04
min      0.000000      2.000000e+01
25%      5.500000      1.035000e+03
50%      8.719912      1.335000e+03
75%      15.000000      2.575000e+03
max      68.000000      6.116250e+03

pct_Educational_Attain_POP_HS_Grad_higher      PDP02.5_3Best      \
count      5920.000000      5920.000000
mean      89.778041      2764.778178
std      9.006337      9160.017904
min      16.000000      0.000000
25%      87.000000      245.000000
50%      92.000000      762.500000
75%      96.000000      2450.000000
max      100.000000      415940.000000

pct_Educational_Attain_BS_Deg_higher      \
count      5920.000000
mean      18.777103
std      0.000000
min      0.000000
25%      0.000000
50%      0.000000
75%      0.000000
max      97.000000

[8 rows x 28 columns]

```

```

In [17]: df = parent_social_econ_by_district_states_df

# reset index only once
df.reset_index(level=0, inplace=True)
df['ST_NAME'] = df['ST_NAME'].str.upper()
df2 = math_states_2011
df2.reset_index(level=0, inplace=True)
# merge test results and parent data
df3 = df.merge(df2, left_on='ST_NAME', right_on='STNAME')

In [18]: df3

Out[18]:
   ST_NAME  LEAID_x  Iteration_x  pct_Below PovtLvl_All_Ages  pctmoe_Below PovtLvl_Age_gte_18  pct_Below PovtLvl_Age_gte_65  PovtLvl_Age_gte_18  PovtLvl_Age_gte_65
0  ALABAMA  1.016377e+05      2020      18.400000      5.970370      18.383704
1  ALASKA  2.003937e+05      2020      14.681395      5.065116      14.688372
2  ARIZONA  4.049697e+05      2020      20.117021      6.129078      20.105574
3  ARKANSAS  5.068456e+05      2020      17.702727      8.444091      17.690465
4  CALIFORNIA  6.211498e+05      2020      13.941816      5.924072      13.939615
5  CONNECTICUT  8.042628e+05      2020      10.395312      5.747656      10.395204
6  COLORADO  9.026891e+05      2020      6.324841      3.580255      6.324204
7  DELAWARE  1.000764e+06      2020      11.600000      3.387500      11.600000
8  DISTRICT OF COLUMBIA  1.100303e+06      2020      16.400000      1.000000      16.400000
9  FLORIDA  1.201026e+06      2020      16.598507      3.782090      16.598507
10 GEORGIA  1.302836e+06      2020      19.677586      5.881034      19.668391
11 HAWAII  1.500032e+06      2020      7.700000      0.500000      7.700000
12 IDAHO  1.601712e+06      2020      12.454652      6.471739      12.478726
13 ILLINOIS  1.721564e+06      2020      10.571411      6.122822      10.566380
14 INDIANA  1.806528e+06      2020      11.027526      4.963763      11.024042
15 IOWA  1.916556e+06      2020      8.022760      4.907903      8.025836
16 KANSAS  2.007427e+06      2020      9.932830      6.162264      9.924151
17 KENTUCKY  2.102877e+06      2020      20.523125      6.856875      20.518750
18 LOUISIANA  2.201077e+06      2020      19.904348      4.808696      19.886957
19 MAINE  2.317757e+06      2020      10.525000      6.367241      10.525862
20 MARYLAND  2.400375e+06      2020      9.029167      1.983333      9.029167
21 MASSACHUSETTS  2.507280e+06      2020      5.531522      3.839855      5.528261
22 MICHIGAN  2.619355e+06      2020      12.799409      5.266732      12.797441
23 MINNESOTA  2.718913e+06      2020      8.032121      4.026970      8.031515
24 MISSISSIPPI  2.802553e+06      2020      23.103759      7.600752      23.107519
25 MISSOURI  2.918444e+06      2020      13.890758      7.568483      13.884123
26 MONTANA  3.010825e+06      2020      12.244828      9.247701      12.234483
27 NEBRASKA  3.137500e+06      2020      8.549315      5.937900      8.552511
28 NEVADA  3.200272e+06      2020      12.298357      4.835714      12.285714
29 NEW HAMPSHIRE  3.304832e+06      2020      4.950427      4.632248      4.592182
30 NEW JERSEY  3.409452e+06      2020      6.822600      4.663400      6.821200
31 NEW MEXICO  3.501917e+06      2020      22.335294      7.364706      22.329412
32 NEW YORK  3.616626e+06      2020      9.539360      4.871200      9.537920
33 NORTH CAROLINA  3.702700e+06      2020      17.236752      4.464957      17.237607
34 OHIO  3.904984e+06      2020      11.260267      5.256427      11.257262
35 OREGON  4.106814e+06      2020      12.743939      6.638636      12.740152
36 PENNSYLVANIA  4.214613e+06      2020      10.471630      4.255734      10.471227
37 SOUTH CAROLINA  4.502354e+06      2020      10.167424      5.736250      10.381250
38 SOUTH DAKOTA  4.638635e+06      2020      10.167424      5.855303      10.154545
39 TENNESSEE  4.702133e+06      2020      17.120438      5.570803      17.121898
40 TEXAS  4.826378e+06      2020      13.959647      6.735598      13.958560
41 UTAH  4.900604e+06      2020      9.979487      3.546154      9.974359
42 VIRGINIA  5.102095e+06      2020      12.931818      4.840152      12.918939
43 WASHINGTON  5.309095e+06      2020      11.066197      5.548826      11.064789
44 WEST VIRGINIA  5.400840e+06      2020      18.138182      5.552727      18.140000
45 WISCONSIN  5.508451e+06      2020      8.425604      4.263527      8.425587
46 WYOMING  5.603443e+06      2020      9.258537      5.585837      9.258537

47 rows x 48 columns

```

```

In [19]: df3.columns

Out[19]: Index(['ST_NAME', 'LEAID_x', 'Iteration_x', 'pct_Below PovtLvl_All_Ages', 'pctmoe_Below PovtLvl_Age_gte_18', 'pctmoe_Below PovtLvl_Age_gte_65', 'PovtLvl_Age_gte_18', 'PovtLvl_Age_gte_65', 'LEA_NAME', 'ST_NAME', 'LEA_CWTFTEST', 'LEA_CWTFTEST', 'GeoId_y', 'Geography_y', 'Year_y', 'Iteration_y', 'num_Educational_Attain_POP', 'num_Educational_Attain_POP_L9th', 'pct_Educational_Attain_POP_L9th', 'num_Educational_Attain_POP_9th-12th', 'pct_Educational_Attain_POP_9th-12th', 'num_Educational_Attain_POP_HS_GRAD', 'pct_Educational_Attain_POP_HS_GRAD', 'num_Educational_Attain_POP_AssocDeg', 'pct_Educational_Attain_POP_AssocDeg', 'num_Educational_Attain_POP_BacDeg', 'pct_Educational_Attain_POP_BacDeg', 'num_Educational_Attain_POP_GradProf', 'pct_Educational_Attain_POP_GradProf', 'PDP02.5_3Test', 'pct_Educational_Attain_POP_HS_Grad_higher', 'PDP02.5_3Best', 'pct_Educational_Attain_BS_Deg_higher', 'STNAME', 'FIPS', 'LEAID_y', 'ALC_MTH04numvalid_1011', 'ALC_MTH04pctprof_1011', 'HOM_MTH04numvalid_1011', 'HOM_MTH04pctprof_1011', 'ALC_MTH08numvalid_1011', 'ALC_MTH08pctprof_1011', 'HOM_MTH08numvalid_1011', 'HOM_MTH08pctprof_1011', 'HOM_MTH08numvalid_1011', 'HOM_MTH08pctprof_1011'],
dtype='object')

```

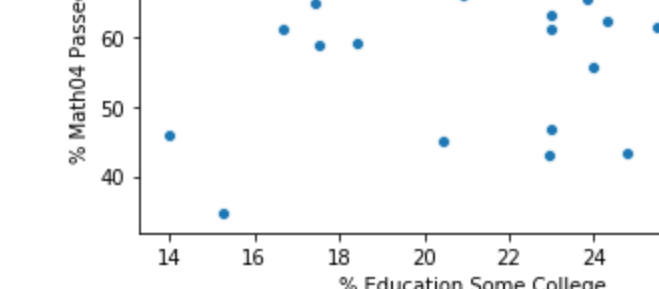
```

In [20]: Parent_social_econ_df_sy201011=df3
Education_dict={'pc_Educational_Attain_POP_L9th': "% Education Lower than 9th Grade",
'pc_Educational_Attain_POP_9th-12th': "% Education 9th -12 Grade",
'pc_Educational_Attain_POP_HS_GRAD': "% Education High School Diploma",
'pc_Educational_Attain_POP_SomeColl': "% Education Some College",
'pc_Educational_Attain_POP_AssocDeg': "% Education Associate Degree",
'pc_Educational_Attain_POP_BacDeg': "% Education Bachelor Degree",
'pc_Educational_Attain_POP_GradProf': "% Education Graduate/Professional Degree",
'pct_Educational_Attain_POP_HS_Grad_higher': "% Education HS Diploma and Higher",
'pct_Educational_Attain_BS_Deg_higher': "% Education BS Degree and Higher"}

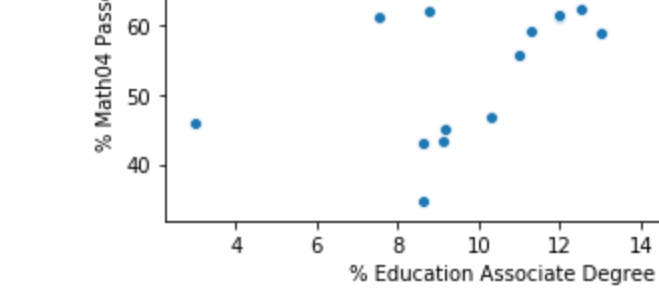
In [21]: for key, value in Education_dict.items():
sns.scatterplot(x=key, y = 'ALL_MTH04pctprof_1011', data = Parent_social_econ_df,
plt.ylabel("% Math04 Passed 2010-2011")
xlabel = "(0) format(value)"
plt.xlabel(xlabel)
title = "% Scatterplot of Grade 4 Math % Pass vs (0) format(value)"
plt.title(title)
plt.show()

```

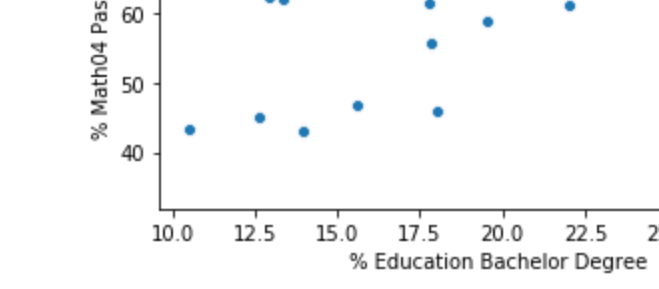
Scatterplot of Grade 4 Math % Pass vs % Education Lower than 9th Grade



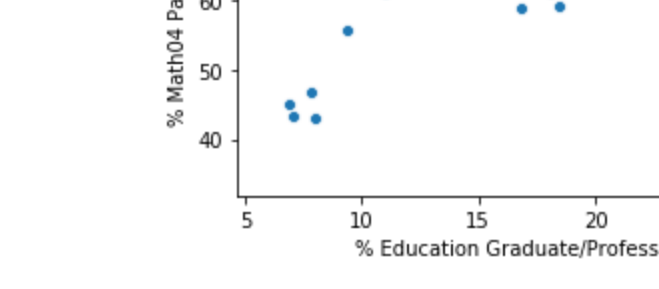
Scatterplot of Grade 4 Math % Pass vs % Education 9th-12 Grade



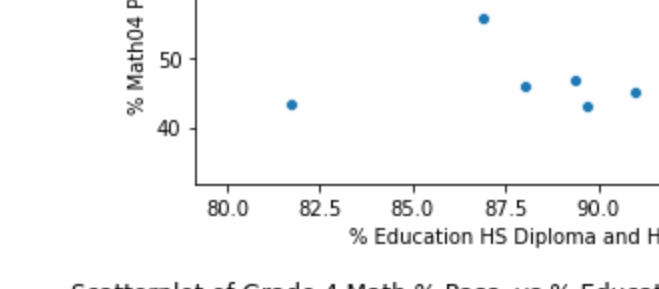
Scatterplot of Grade 4 Math % Pass vs % Education High School Diploma



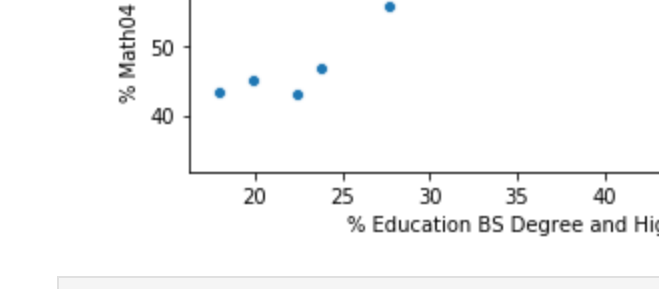
Scatterplot of Grade 4 Math % Pass vs % Education Some College



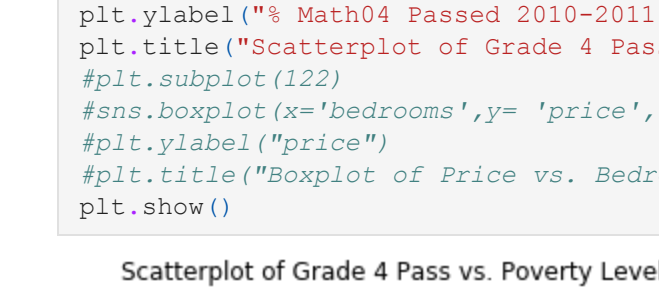
Scatterplot of Grade 4 Math % Pass vs % Education Associate Degree



Scatterplot of Grade 4 Math % Pass vs % Education Bachelor Degree



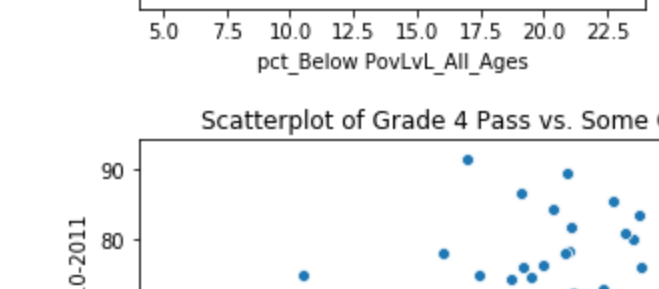
Scatterplot of Grade 4 Math % Pass vs % Education Graduate/Professional Degree



Scatterplot of Grade 4 Math % Pass vs % Education HS Diploma and Higher



Scatterplot of Grade 4 Math % Pass vs % Education BS Degree and Higher



```

In [22]: plt.figure(figsize=(10,8))
plt.subplot(121)

sns.scatterplot(x='pct_Below PovtLvl_All_Ages',y = 'ALL_MTH04pctprof_1011', data = df3,
plt.ylabel("% Math04 Passed 2010-2011")
plt.title("Scatterplot of Grade 4 Pass vs. Poverty Level")
plt.show()

sns.scatterplot(x='pc_Educational_Attain_POP_SomeColl',y = 'ALL_MTH04pctprof_1011', data = df3,
plt.ylabel("% Math04 Passed 2010-2011")
plt.title("Scatterplot of Grade 4 Pass vs. Some College")
plt.show()

#sns.boxplot(x='bedrooms',y='price', data = listings)
plt.ylabel("price")
plt.title("Boxplot of Price vs. Bedrooms")
plt.show()

In [23]: # Create a correlation matrix
corr = Parent_social_econ_df_sy201011.corr()
pos_cor = corr['ALL_MTH04pctprof_1011'] > 0
neg_cor = corr['ALL_MTH04pctprof_1011'] < 0

In [24]: corr['ALL_MTH04pctprof_1011'][pos_cor].sort_values(ascending = False)

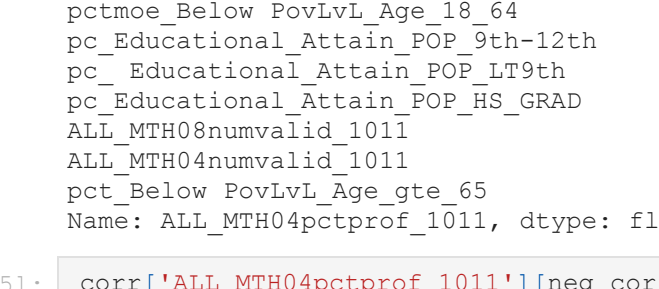
Out[24]:
ALL_MTH04pctprof_1011      1.000000
ECD_MTH04pctprof_1011      0.958014
pctmoe_Below PovtLvl_Age_gte_65      0.664670
HOM_MTH08pctprof_1011      0.816835
HOM_MTH04pctprof_1011      0.696511
HOM_MTH08pctprof_1011      0.555276
pc_Educational_Attain_POP_SomeColl      0.408007
LEA_CWTFTEST      0.372201
pc_Educational_Attain_POP_AssocDeg      0.204845
pctmoe_Below PovtLvl_Age_gte_18      0.095094
pctmoe_Below PovtLvl_Age_18_64      0.096583
pc_Educational_Attain_POP_9th-12th      0.053208
pc_Educational_Attain_POP_L9th      0.046383
pc_Educational_Attain_POP_HS_GRAD      0.041005
ALL_MTH08numvalid_1011      0.015891
ALL_MTH04numvalid_1011      0.008353
pct_Below PovtLvl_Age_gte_65      0.007468
Name: ALL_MTH04pctprof_1011, dtype: float64

In [25]: corr['ALL_MTH04pctprof_1011'][neg_cor].sort_values()

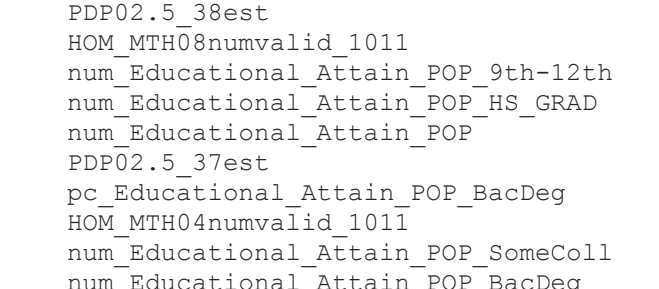
Out[25]:
pc_Educational_Attain_POP_GradProf      -0.373750
LEA_CWTFTEST      -0.296637
pctmoe_Below PovtLvl_Age_gte_18      -0.266330
num_Educational_Attain_POP_GradProf      -0.151693
FIFST      -0.126006
LEAID_x      -0.125706
LEAID_y      -0.125540
num_Educational_Attain_POP_L9th      -0.114629
PDP02.5_3Best      -0.107713
HOM_MTH08numvalid_1011      -0.103350
num_Educational_Attain_POP_9th-12th      -0.099718
num_Educational_Attain_POP_HS_GRAD      -0.098726
num_Educational_Attain_POP      -0.094095
PDP02.5_3Test      -0.092600
pc_Educational_Attain_POP_BacDeg      -0.086043
HOM_MTH04numvalid_1011      -0.077489
num_Educational_Attain_POP_SomeColl      -0.074466
num_Educational_Attain_POP_BacDeg      -0.071297
num_Educational_Attain_POP_AssocDeg      -0.058828
pctmoe_Below PovtLvl_Age_gte_18      -0.052364
pctmoe_Below PovtLvl_Age_18_64      -0.050780
ECD_MTH08numvalid_1011      -0.047288
pct_Below PovtLvl_Age_gte_18      -0.038706
pct_Below PovtLvl_Age_18_64      -0.038656
pct_Below PovtLvl_All_Ages      -0.038509
Name: ALL_MTH04pctprof_1011, dtype: float64

```

Scatterplot of Grade 4 Pass vs. Poverty Level



Scatterplot of Grade 4 Pass vs. Some College



```

In [23]: # Create a correlation matrix
corr = Parent_social_econ_df_sy201011.corr()
pos_cor = corr['ALL_MTH04pctprof_1011'] > 0
neg_cor = corr['ALL_MTH04pctprof_1011'] < 0

In [24]: corr['ALL_MTH04pctprof_1011'][pos_cor].sort_values(ascending = False)

Out[24]:
ALL_MTH04pctprof_1011      1.000000
ECD_MTH04pctprof_1011      0.958014
pctmoe_Below PovtLvl_Age_gte_65      0.664670
HOM_MTH08pctprof_1011      0.816835
HOM_MTH04pctprof_1011      0.696511
HOM_MTH08pctprof_1011      0.555276
pc_Educational_Attain_POP_SomeColl      0.408007
LEA_CWTFTEST      0.372201
pc_Educational_Attain_POP_AssocDeg      0.204845
pctmoe_Below PovtLvl_Age_gte_18      0.095094
pctmoe_Below PovtLvl_Age_18_64      0.096583
pc_Educational_Attain_POP_9th-12th      0.053208
pc_Educational_Attain_POP_L9th      0.046383
pc_Educational_Attain_POP_HS_GRAD      0.041005
ALL_MTH08numvalid_1011      0.015891
ALL_MTH04numvalid_1011      0.008353
pct_Below PovtLvl_Age_gte_18      0.007468
pct_Below PovtLvl_Age_18_64      0.007468
Name: ALL_MTH04pctprof_1011, dtype: float64

In [25]: corr['ALL_MTH04pctprof_1011'][neg_cor].sort_values()

Out[25]:
pc_Educational_Attain_POP_GradProf      -0.373750
LEA_CWTFTEST      -0.296637
pctmoe_Below PovtLvl_Age_gte_18      -0.266330
num_Educational_Attain_POP_GradProf      -0.151693
FIFST      -0.126006
LEAID_x      -0.125706
LEAID_y      -0.125540
num_Educational_Attain_POP_L9th      -0.114629
PDP02.5_3Best      -0.107713
HOM_MTH08numvalid_1011      -0.103350
num_Educational_Attain_POP_9th-12th      -0.099718
num_Educational_Attain_POP_HS_GRAD      -0.098726
num_Educational_Attain_POP      -0.094095
PDP02.5_3Test      -0.092600
pc_Educational_Attain_POP_BacDeg      -0.086043
HOM_MTH04numvalid_1011      -0.077489
num_Educational_Attain_POP_SomeColl      -0.074466
num_Educational_Attain_POP_BacDeg      -0.071297
num_Educational_Attain_POP_AssocDeg      -0.058828
pctmoe_Below PovtLvl_Age_gte_18      -0.052364
pctmoe_Below PovtLvl_Age_18_64      -0.050780
ECD_MTH08numvalid_1011      -0.047288
pct_Below PovtLvl_Age_gte_18      -0.038706
pct_Below PovtLvl_Age_18_64      -0.038656
pct_Below PovtLvl_All_Ages      -0.038509
Name: ALL_MTH04pctprof_1011, dtype: float64

```

Scatterplot of Grade 4 Pass vs. Some College



Scatterplot of Grade 4 Pass vs. Poverty Level



Scatterplot of Grade 4 Pass vs. Some College



Scatterplot of Grade 4 Pass vs. Poverty Level



Scatterplot of Grade 4 Pass vs. Some College



Scatterplot of Grade 4 Pass vs. Poverty Level



Scatterplot of Grade 4 Pass vs. Some College

