Project 2 by Weitng Lin & Linlin Zhu

In [2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
df=pd.read_csv('Most_recent.csv')
```

```
/Users/xiuronglin/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3020: DtypeWa
rning: Columns (6,9,31,1608,1619,1620,1621,1622,1623,1624,1625,1626,1627,1628,1629,1688,1689,1690,169
1,1692,1703,1704,1725,1726,1727,1728,1729,1743,1815,1816,1817,1818,1823,1824,1830,1831,1879,1880,1881,
1882,1883,1884,1885,1886,1887,1888,1889,1890,1891,1892,1893,1894,1895,1896,1897,1898,1909,1910,1911,19
12,1913,1957,1958,1959,1960,1961,1962,1963,1964,1965,1966,1967,1968,1969,1970,1971,1972,1973,1974,197
5,1976) have mixed types. Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

In [4]:
```
df.shape
```

Out[4]:
```
(7058, 1977)
```

In [5]:
```
df.columns
```

Out[5]:
```
Index(['UNITID', 'OPEID', 'OPEID6', 'INSTNM', 'CITY', 'STABBR', 'ZIP',
       'ACCREDAGENCY', 'INSTURL', 'NPCURL',
       ...
       'OMAWDP8_NOTFIRSTTIME_POOLED_SUPP', 'OMENRUP_NOTFIRSTTIME_POOLED_SUPP',
       'OMENRYP_FULLTIME_POOLED_SUPP', 'OMENRAP_FULLTIME_POOLED_SUPP',
       'OMAWDP8_FULLTIME_POOLED_SUPP', 'OMENRUP_FULLTIME_POOLED_SUPP',
       'OMENRYP_PARTTIME_POOLED_SUPP', 'OMENRAP_PARTTIME_POOLED_SUPP',
       'OMAWDP8_PARTTIME_POOLED_SUPP', 'OMENRUP_PARTTIME_POOLED_SUPP'],
      dtype='object', length=1977)
```

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7058 entries, 0 to 7057
Columns: 1977 entries, UNITID to OMENRUP_PARTTIME_POOLED_SUPP
dtypes: float64(617), int64(14), object(1346)
memory usage: 106.5+ MB
```

In [7]:
```
df.head()
```

Out[7]:

	UNITID	OPEID	OPEID6	INSTNM	CITY	STABBR	ZIP	ACCREDAGENCY	INSTURL	
0	100654	100200	1002	Alabama A & M University	Normal	AL	35762	Southern Association of Colleges and Schools C...	www.aamu.edu/	https://galileo.aamu.ed
1	100663	105200	1052	University of Alabama at Birmingham	Birmingham	AL	35294-0110	Southern Association of Colleges and Schools C...	www.uab.edu	uab.studentaidc
2	100690	2503400	25034	Amridge University	Montgomery	AL	36117-3553	Southern Association of Colleges and Schools C...	www.amridgeuniversity.edu	www2.an
3	100706	105500	1055	University of Alabama in Huntsville	Huntsville	AL	35899	Southern Association of Colleges and Schools C...	www.uah.edu	
4	100724	100500	1005	Alabama State University	Montgomery	AL	36104-0271	Southern Association of Colleges and Schools C...	www.alasu.edu	aic

5 rows × 1977 columns

In [8]:
```
print('The data has {0} rows and {1} columns'.format(df.shape[0],df.shape[1]))
```

```
The data has 7058 rows and 1977 columns
```

1.What is the most costly college? What is the cheapest?

In [9]:
```
df['total_tution'] = df['COSTT4_A'].fillna(0) + df['COSTT4_P'].fillna(0)
df = df[df['total_tution'] != 0]
```

In [10]:
```
df.groupby('INSTNM')['total_tution'].mean().sort_values(ascending =False)
```

Out[10]:
```
INSTNM
L3 Commercial Training Solutions Airline Academy                  105745.0
Aviator College of Aeronautical Science and Technology             93704.0
University of Chicago                                              72717.0
Jewish Theological Seminary of America                             72120.0
Columbia University in the City of New York                        71972.0
Harvey Mudd College                                                71917.0
Northwestern University                                            70317.0
Washington University in St Louis                                  69754.0
University of Southern California                                  69547.0
Dartmouth College                                                  69474.0
Drexel University                                                  69462.0
Occidental College                                                 69442.0
Haverford College                                                  69387.0
Claremont McKenna College                                          69385.0
Southern Methodist University                                      69358.0
University of Pennsylvania                                         69340.0
Georgetown University                                              69313.0
Duke University                                                    69169.0
Landmark College                                                   69020.0
Amherst College                                                    68986.0
Yale University                                                    68950.0
Trinity College                                                    68940.0
Oberlin College                                                    68670.0
Bard College                                                       68528.0
Barnard College                                                    68512.0
Scripps College                                                    68464.0
Williams College                                                   68430.0
Bennington College                                                 68420.0
Bard College at Simon's Rock                                       68409.0
Pepperdine University                                              68352.0
                                                                     ...
Wytheville Community College                                        7084.0
Instituto Tecnologico de Puerto Rico-Recinto de Guayama             7061.0
Lake Career and Technical Center                                    7050.0
Gordon Cooper Technology Center                                     7044.0
Baldwin Park Adult & Community Education                            7040.0
Southeastern Baptist College                                        6950.0
Chattahoochee Valley Community College                              6854.0
South Texas College                                                 6787.0
American Technical Institute                                        6782.0
```

```
Carver Career Center                                              6562.0
Lincoln Trail College                                             6551.0
Hacienda La Puente Adult Education                                6509.0
Virginia Beach City Public Schools School of Practical Nursing    6407.0
Indian River State College                                        6276.0
Cloyd's Barber School 2 Inc                                       6255.0
Palau Community College                                           6085.0
Mineral County Vocational Technical Center                        5957.0
Colegio Universitario de San Juan                                 5950.0
Putnam Career and Technical Center                                5925.0
Kiamichi Technology Center-Durant                                 5894.0
D A Dorsey Technical College                                      5854.0
Aparicio-Levy Technical College                                   5705.0
J F Ingram State Technical College                                5496.0
Escuela De Troqueleria Y Herramentaje                             5481.0
Cleveland Community College                                       5185.0
Wes Watkins Technology Center                                     5180.0
Instituto Tecnologico de Puerto Rico-Recinto de Manati            5025.0
Instituto Tecnologico de Puerto Rico-Recinto de San Juan          4007.0
Instituto Tecnologico de Puerto Rico-Recinto de Ponce             3930.0
C. Alexander School of Cosmetology                                3522.0
Name: total_tution, Length: 5751, dtype: float64
```

In [11]:
```
df.groupby('INSTNM')['total_tution'].mean().sort_values(ascending =True)
```

Out[11]:
```
INSTNM
C. Alexander School of Cosmetology                                   3522.0
Instituto Tecnologico de Puerto Rico-Recinto de Ponce                3930.0
Instituto Tecnologico de Puerto Rico-Recinto de San Juan             4007.0
Instituto Tecnologico de Puerto Rico-Recinto de Manati               5025.0
Wes Watkins Technology Center                                        5180.0
Cleveland Community College                                          5185.0
Escuela De Troqueleria Y Herramentaje                                5481.0
J F Ingram State Technical College                                   5496.0
Aparicio-Levy Technical College                                      5705.0
D A Dorsey Technical College                                         5854.0
Kiamichi Technology Center-Durant                                    5894.0
Putnam Career and Technical Center                                   5925.0
Colegio Universitario de San Juan                                    5950.0
Mineral County Vocational Technical Center                           5957.0
Palau Community College                                              6085.0
Cloyd's Barber School 2 Inc                                          6255.0
Indian River State College                                           6276.0
Virginia Beach City Public Schools School of Practical Nursing       6407.0
Hacienda La Puente Adult Education                                   6509.0
Lincoln Trail College                                                6551.0
Carver Career Center                                                 6562.0
American Technical Institute                                         6782.0
South Texas College                                                  6787.0
Chattahoochee Valley Community College                               6854.0
Southeastern Baptist College                                         6950.0
Baldwin Park Adult & Community Education                             7040.0
Gordon Cooper Technology Center                                      7044.0
Lake Career and Technical Center                                     7050.0
Instituto Tecnologico de Puerto Rico-Recinto de Guayama              7061.0
Wytheville Community College                                         7084.0
                                                                     ...
Pepperdine University                                               68352.0
Bard College at Simon's Rock                                        68409.0
Bennington College                                                  68420.0
Williams College                                                    68430.0
Scripps College                                                     68464.0
Barnard College                                                     68512.0
Bard College                                                        68528.0
Oberlin College                                                     68670.0
Trinity College                                                     68940.0
```

```
Yale University                                              68950.0
Amherst College                                              68986.0
Landmark College                                             69020.0
Duke University                                              69169.0
Georgetown University                                        69313.0
University of Pennsylvania                                   69340.0
Southern Methodist University                                69358.0
Claremont McKenna College                                    69385.0
Haverford College                                            69387.0
Occidental College                                           69442.0
Drexel University                                            69462.0
Dartmouth College                                            69474.0
University of Southern California                            69547.0
Washington University in St Louis                            69754.0
Northwestern University                                      70317.0
Harvey Mudd College                                          71917.0
Columbia University in the City of New York                  71972.0
Jewish Theological Seminary of America                       72120.0
University of Chicago                                        72717.0
Aviator College of Aeronautical Science and Technology       93704.0
L3 Commercial Training Solutions Airline Academy            105745.0
Name: total_tution, Length: 5751, dtype: float64
```

In [12]:
```
#the most costly college is L3 Commercial Training Solutions Airline Academy
#the cheapest college is C. Alexander School of Cosmetology
```

2.What is the average cost for college for colleges in different parts of the US

In [13]:
```
df.groupby('STABBR')['total_tution'].mean()
```

Out[13]:
```
STABBR
AK    20202.666667
AL    20661.987179
AR    18277.480000
AS     7400.000000
AZ    21326.971429
CA    25254.285971
CO    23154.380435
CT    26631.231884
DC    36920.666667
DE    24394.266667
FL    23675.617089
FM     9554.000000
GA    23428.047297
GU    12339.000000
HI    20745.285714
IA    25967.620253
ID    19609.424242
IL    23245.472477
IN    27260.939394
KS    21792.555556
KY    22480.465909
LA    20873.833333
MA    35062.945946
MD    25328.820513
ME    28090.529412
MH     8750.000000
MI    22736.019737
MN    24886.870968
MO    22620.850340
MP     8734.000000
MS    19143.018868
MT    17119.900000
NC    22907.287425
ND    17174.206897
NE    24390.181818
NH    28653.166667
NJ    25148.864286
NM    18100.279070
NV    22693.571429
NY    27910.160000
```

```
OH    23277.090909
OK    17511.097345
OR    23797.000000
PA    29006.827044
PR    11653.033058
PW     6085.000000
RI    34898.300000
SC    24322.868687
SD    21214.333333
TN    22750.207792
TX    21572.531335
UT    20058.442623
VA    25169.522388
VI    16786.000000
VT    39868.857143
WA    22176.948454
WI    24552.855422
WV    18101.865672
WY    14714.200000
Name: total_tution, dtype: float64
```

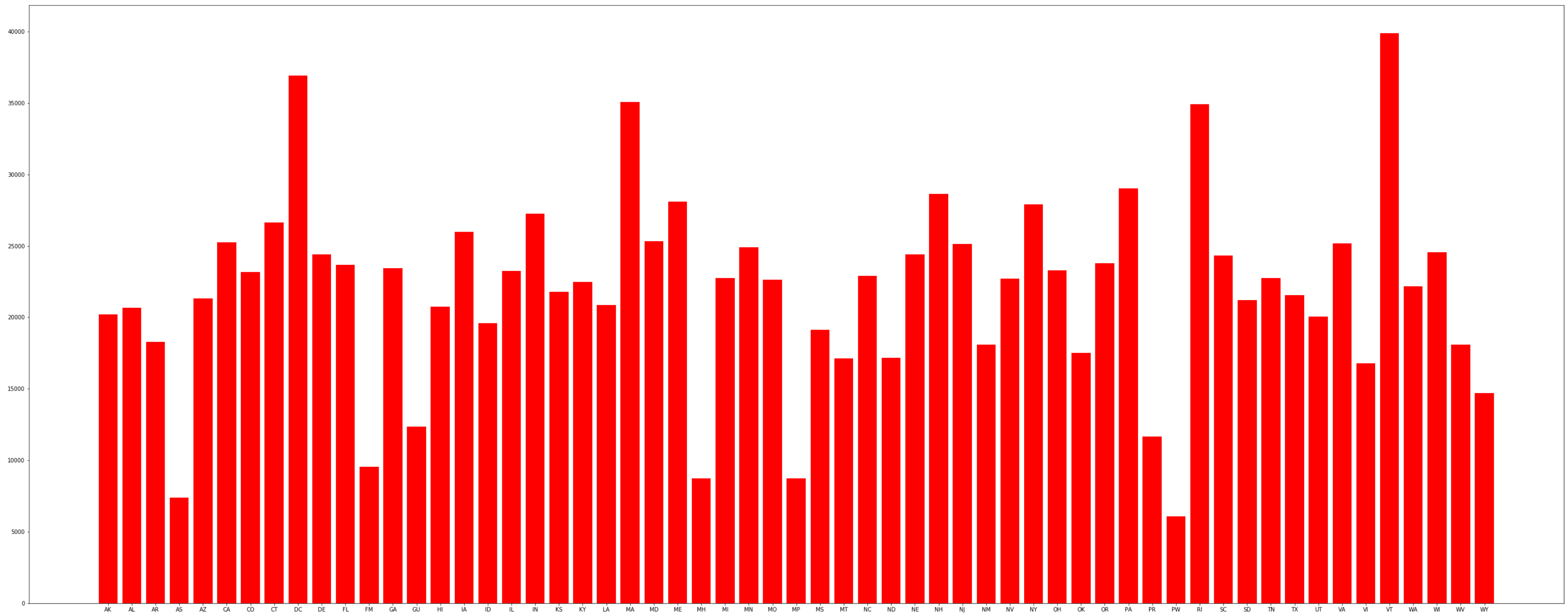
In [14]:
```
#What is the average cost for college for colleges in different parts of the US
df.groupby('STABBR')['total_tution'].mean()
```

Out[14]:
```
STABBR
AK    20202.666667
AL    20661.987179
AR    18277.480000
AS     7400.000000
AZ    21326.971429
CA    25254.285971
CO    23154.380435
CT    26631.231884
DC    36920.666667
DE    24394.266667
FL    23675.617089
FM     9554.000000
GA    23428.047297
GU    12339.000000
HI    20745.285714
IA    25967.620253
ID    19609.424242
IL    23245.472477
IN    27260.939394
KS    21792.555556
KY    22480.465909
LA    20873.833333
MA    35062.945946
MD    25328.820513
ME    28090.529412
MH     8750.000000
MI    22736.019737
MN    24886.870968
MO    22620.850340
MP     8734.000000
MS    19143.018868
MT    17119.900000
NC    22907.287425
ND    17174.206897
NE    24390.181818
NH    28653.166667
NJ    25148.864286
NM    18100.279070
NV    22693.571429
```

```
NY    27910.160000
OH    23277.090909
OK    17511.097345
OR    23797.000000
PA    29006.827044
PR    11653.033058
PW     6085.000000
RI    34898.300000
SC    24322.868687
SD    21214.333333
TN    22750.207792
TX    21572.531335
UT    20058.442623
VA    25169.522388
VI    16786.000000
VT    39868.857143
WA    22176.948454
WI    24552.855422
WV    18101.865672
WY    14714.200000
Name: total_tution, dtype: float64
```

In [15]:
```
state = df.groupby(['STABBR']).mean()
plt.figure(figsize=(50,20))
plt.bar(state.index,state['total_tution'],color='red')
```

Out[15]: <BarContainer object of 59 artists>



3.What is the average cost of religious vs. secular institutions?

In [16]:
```
# Create a new dataset indecate the different religions
religiou= df.groupby('RELAFFIL').mean()
# Create a new dataset indecate the doesn't have religions
secular = df[df['RELAFFIL'].isnull()]

rel = religiou['total_tution'].mean()
sec = secular['total_tution'].mean()
```

In [17]:
```
rel
```

Out[17]:
```
35304.21875642725
```

In [18]:
```
sec
```

Out[18]:
```
21895.61946729338
```

In [19]:
```
#"The average cost of all religious institutions is $35304.21875642725
#The average cost of all secular institutions is $21895.61946729338
```

4.What percent of colleges have an open admission policy?

In [20]:
```
# Open admission is represented by the column OPENADMP
# Colleges with an open admission policy are represented by 1.0
adm = df[df['OPENADMP'].notnull()]
openadm = adm.groupby('OPENADMP').size()
openrate = float(openadm[1.0]) / len(df['OPENADMP']) * 100
openrate = round(openrate,2)
```

In [21]:
```
print(openrate)
```

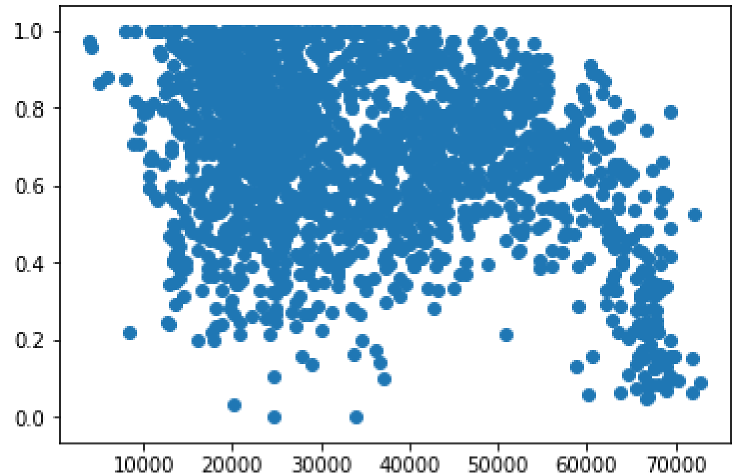
```
65.76
```

In [22]:
```
#65.76 percent of colleges have an open admission policy
```

5.What is the correlation (scatterplot) between admission rates and college cost?

In [23]:
```
plt.scatter(df['total_tution'], df['ADM_RATE'])
```

Out[23]: <matplotlib.collections.PathCollection at 0x1a3033f5f8>



6.What is the correlation between SAT scores and admission rates? Are there any outliers?

In [24]:
```
corr=df.corr()
```

In [25]:
```
corr
```

Out[25]:

	UNITID	OPEID	OPEID6	SCH_DEG	HCM2	MAIN	NUMBRANCH	PREDDEG	HIGHDEG	CONT
UNITID	1.000000	0.404391	0.703476	-0.403578	0.060837	-0.399365	0.218942	-0.416574	-0.411594	0.48
OPEID	0.404391	1.000000	0.506002	-0.222312	0.032336	-0.122262	0.208191	-0.226693	-0.225232	0.31
OPEID6	0.703476	0.506002	1.000000	-0.564070	0.075663	-0.067748	-0.009493	-0.565021	-0.565996	0.55
SCH_DEG	-0.403578	-0.222312	-0.564070	1.000000	-0.005817	0.147829	0.035107	0.946966	0.899707	-0.40
HCM2	0.060837	0.032336	0.075663	-0.005817	1.000000	-0.083521	0.117531	0.000228	-0.009318	0.09
MAIN	-0.399365	-0.122262	-0.067748	0.147829	-0.083521	1.000000	-0.541109	0.167519	0.137340	-0.31
NUMBRANCH	0.218942	0.208191	-0.009493	0.035107	0.117531	-0.541109	1.000000	0.014370	0.051479	0.22
PREDDEG	-0.416574	-0.226693	-0.565021	0.946966	0.000228	0.167519	0.014370	1.000000	0.902380	-0.40
HIGHDEG	-0.411594	-0.225232	-0.565996	0.899707	-0.009318	0.137340	0.051479	0.902380	1.000000	-0.38
CONTROL	0.482515	0.315501	0.555806	-0.401863	0.093042	-0.311192	0.221663	-0.401687	-0.382893	1.00
ST_FIPS	0.141743	-0.012795	-0.039005	0.030508	-0.045248	-0.014274	-0.014349	0.032446	0.025705	-0.03
REGION	0.097989	0.102737	0.163336	-0.117991	0.058573	-0.051692	0.010553	-0.114964	-0.072736	0.05
LOCALE	-0.141363	-0.089669	-0.151339	0.040276	-0.058035	0.113280	-0.098476	0.024273	-0.009425	-0.33
LOCALE2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
LATITUDE	-0.145927	-0.096769	-0.183972	0.108464	-0.030577	0.094079	-0.049475	0.107136	0.074661	-0.10
LONGITUDE	0.005157	-0.057375	-0.072721	0.075764	-0.055800	0.016974	-0.000509	0.073179	0.051464	-0.02
CCBASIC	-0.307206	-0.162033	-0.431689	0.799136	0.016306	0.055523	0.107842	0.800780	0.883162	-0.25
CCUGPROF	-0.396347	-0.220180	-0.528604	0.858298	-0.033482	0.143404	-0.011756	0.863897	0.914610	-0.33
CCSIZSET	-0.458946	-0.274698	-0.601761	0.865236	-0.037810	0.199471	-0.024481	0.870958	0.924810	-0.46
HBCU	-0.097495	-0.063716	-0.118361	0.135732	0.026512	0.059568	-0.039517	0.132983	0.123301	-0.09
PBI	-0.038697	-0.024972	-0.039826	0.007795	0.015751	-0.007941	0.004172	0.006734	0.041477	-0.09
ANNHI	-0.037743	-0.015005	-0.024652	0.034803	0.011080	0.025239	-0.014402	0.034233	0.033730	-0.06
TRIBAL	0.000620	0.023418	0.053823	0.002351	0.008343	0.035074	-0.023150	0.002701	0.014235	-0.07
AANAPII	-0.089836	-0.045586	-0.079923	0.084640	-0.019117	0.044170	-0.031741	0.086222	0.096905	-0.14

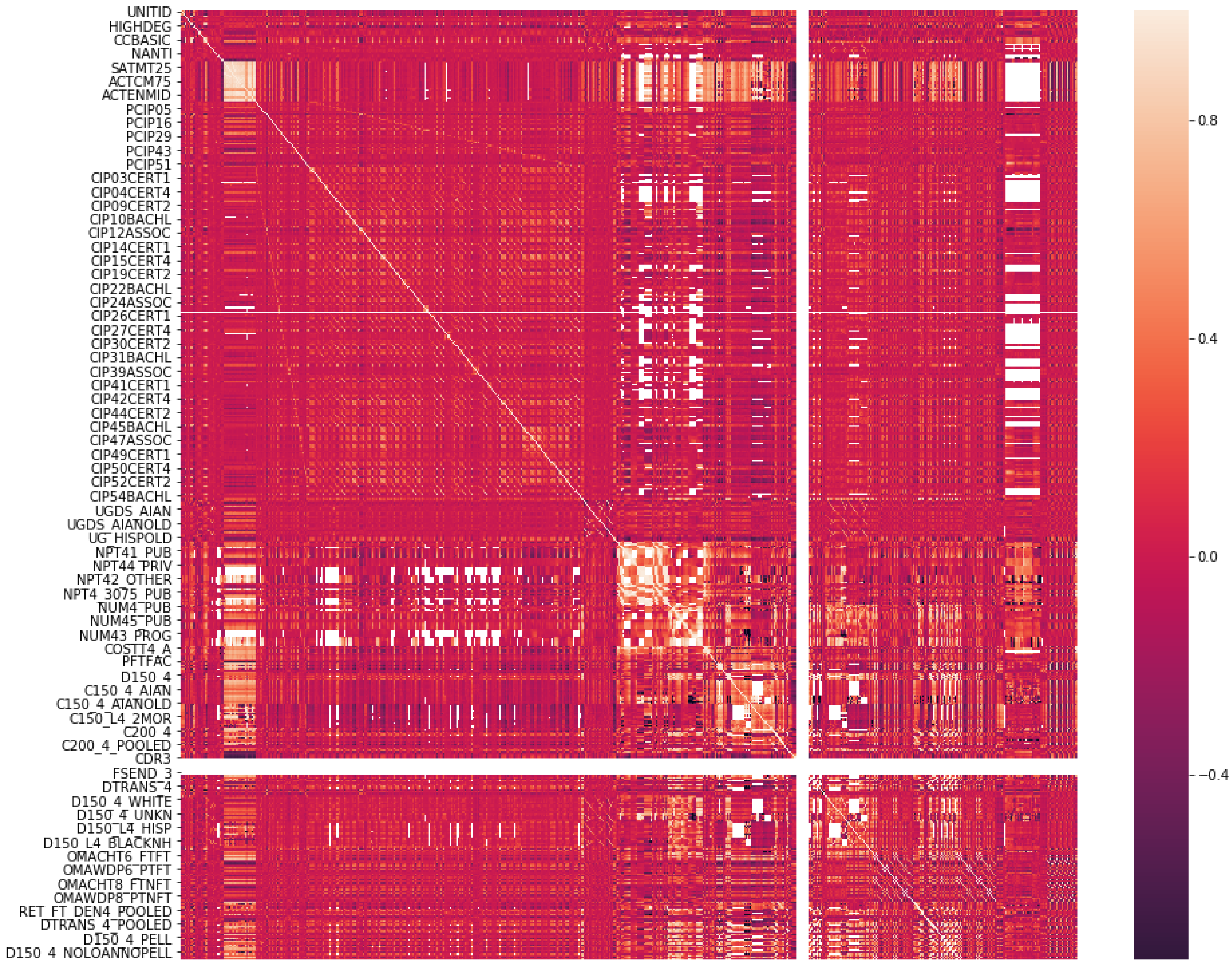
	UNITID	OPEID	OPEID6	SCH_DEG	HCM2	MAIN	NUMBRANCH	PREDDEG	HIGHDEG	CONT
NANTI	-0.033184	-0.023565	-0.042107	0.035105	0.011604	0.017849	-0.016543	0.028701	0.024563	-0.07
MENONLY	-0.037816	-0.009893	-0.012054	0.123918	-0.012195	0.044226	-0.029641	0.103872	0.082402	-0.01
WOMENONLY	-0.061770	-0.042622	-0.079336	0.102761	-0.010350	0.037536	-0.024790	0.104683	0.101058	-0.01
RELAFFIL	0.075427	0.142742	0.143405	-0.008700	0.013851	-0.009468	0.000099	0.030766	-0.121881	0.08
ADM_RATE	0.114609	0.130341	0.130341	-0.080637	0.005388	-0.084669	0.108254	-0.117781	-0.071292	-0.01
ADM_RATE_ALL	0.119586	0.119228	0.138030	-0.083913	0.005739	-0.042408	0.054345	-0.121068	-0.076443	-0.00
...	...	...	...	...	...	...	...	...	...	
MTHCMP3	-0.014976	-0.122831	-0.122832	0.229917	-0.020988	-0.064633	0.145573	0.276222	0.462006	-0.15
MTHCMP4	-0.045400	-0.183845	-0.183846	0.251932	0.012391	-0.065158	0.161554	0.330341	0.462249	-0.15
MTHCMP5	-0.102229	-0.219415	-0.219415	0.193211	-0.043913	-0.121937	0.135052	0.196352	0.412997	-0.12
MTHCMP6	-0.059024	-0.207428	-0.207429	0.247293	-0.030993	-0.136376	0.203557	0.211246	0.398564	-0.03
POOLYRSOM_ALL	0.174177	0.055468	0.181240	-0.001949	-0.012183	-0.021549	0.053552	-0.010157	-0.017620	0.09
POOLYRSOM_FIRSTTIME	0.206620	0.070644	0.238201	0.051038	0.007070	-0.027526	0.021351	0.010869	0.027892	0.11
POOLYRSOM_NOTFIRSTTIME	0.269315	0.093720	0.326309	-0.138725	-0.007928	-0.086831	-0.001980	-0.140837	-0.144221	0.25
POOLYRSOM_FULLTIME	0.243266	0.081396	0.264131	-0.011362	0.033879	-0.042287	0.075467	-0.021640	-0.010135	0.11
POOLYRSOM_PARTTIME	0.243661	0.178441	0.294154	-0.002022	-0.039130	-0.098215	-0.003833	0.006762	0.008099	0.46
OMENRYP_ALL	-0.043783	-0.006866	-0.007344	-0.043389	-0.006825	-0.001302	-0.014770	-0.064118	-0.068836	-0.14
OMENRAP_ALL	-0.297737	-0.179910	-0.331517	0.222794	-0.070372	0.225292	-0.121971	0.205087	0.078577	-0.45
OMAWDP8_ALL	0.066657	-0.011503	0.057823	0.138330	-0.086340	0.016372	-0.111412	0.161615	0.277867	0.33
OMENRUP_ALL	0.156753	0.145911	0.184810	-0.301127	0.143901	-0.183121	0.209670	-0.310492	-0.338762	0.00
OMENRYP_FIRSTTIME	-0.044914	-0.008823	-0.003797	-0.046365	-0.005683	0.002046	-0.024727	-0.067568	-0.073269	-0.16
OMENRAP_FIRSTTIME	-0.281993	-0.179957	-0.339375	0.315836	-0.062710	0.218100	-0.130393	0.302094	0.200274	-0.39
OMAWDP8_FIRSTTIME	0.041181	-0.048217	0.046538	0.100998	-0.101397	0.039903	-0.146648	0.118887	0.228282	0.30
OMENRUP_FIRSTTIME	0.162912	0.175201	0.191207	-0.312919	0.145876	-0.192484	0.241291	-0.319065	-0.356216	-0.00
OMENRYP_NOTFIRSTTIME	-0.042107	-0.006955	-0.020443	-0.021044	-0.022192	-0.018316	-0.005107	-0.043984	-0.035645	-0.12
OMENRAP_NOTFIRSTTIME	-0.279043	-0.157296	-0.287827	0.038551	-0.082179	0.194366	-0.102678	0.016648	-0.115459	-0.47

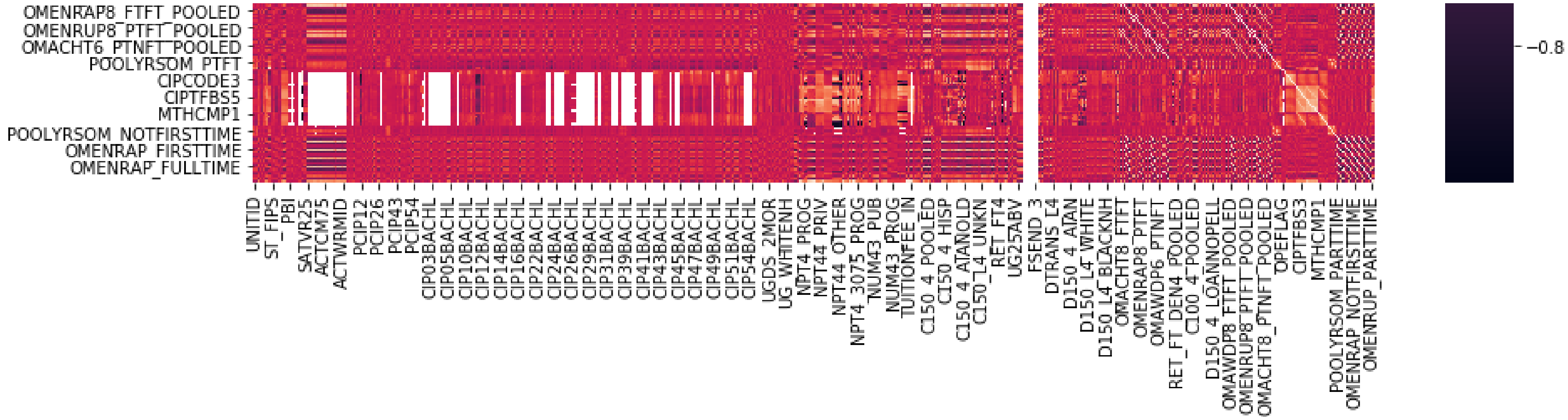
	UNITID	OPEID	OPEID6	SCH_DEG	HCM2	MAIN	NUMBRANCH	PREDDEG	HIGHDEG	CONT
OMAWDP8_NOTFIRSTTIME	0.075317	0.004635	0.049882	0.173103	-0.059508	0.013698	-0.090166	0.200870	0.304076	0.30
OMENRUP_NOTFIRSTTIME	0.145377	0.120772	0.176782	-0.217223	0.134129	-0.166226	0.181294	-0.226727	-0.236672	0.06
OMENRYP_FULLTIME	-0.042959	-0.008380	-0.017061	-0.019971	-0.000542	-0.002759	-0.007416	-0.042180	-0.059235	-0.17
OMENRAP_FULLTIME	-0.289722	-0.179799	-0.332451	0.239914	-0.064084	0.218658	-0.109640	0.223789	0.099332	-0.43
OMAWDP8_FULLTIME	0.063716	-0.024328	0.051463	0.106959	-0.063627	0.001227	-0.071317	0.129895	0.250215	0.32
OMENRUP_FULLTIME	0.165000	0.166552	0.206570	-0.293151	0.116856	-0.170410	0.161486	-0.303184	-0.330512	0.02
OMENRYP_PARTTIME	-0.027792	0.022581	0.028471	-0.098456	-0.017747	-0.013196	-0.024178	-0.106725	-0.062875	0.00
OMENRAP_PARTTIME	-0.235909	-0.251556	-0.287530	0.073843	-0.091773	0.159551	-0.080814	0.076461	-0.005540	-0.35
OMAWDP8_PARTTIME	0.035847	0.028747	0.027627	0.178739	-0.075054	0.027880	-0.046486	0.196673	0.286352	0.18
OMENRUP_PARTTIME	0.149171	0.154013	0.180179	-0.194677	0.142310	-0.140696	0.109423	-0.210647	-0.244499	0.09
total_tution	-0.135507	-0.123956	-0.231899	0.472730	-0.006195	0.014980	0.026661	0.476438	0.483530	0.16

632 rows × 632 columns

In [26]:
```
plt.figure(figsize=(15,15))
sns.heatmap(corr)
```

Out[26]: <matplotlib.axes._subplots.AxesSubplot at 0x1a2f51bf98>





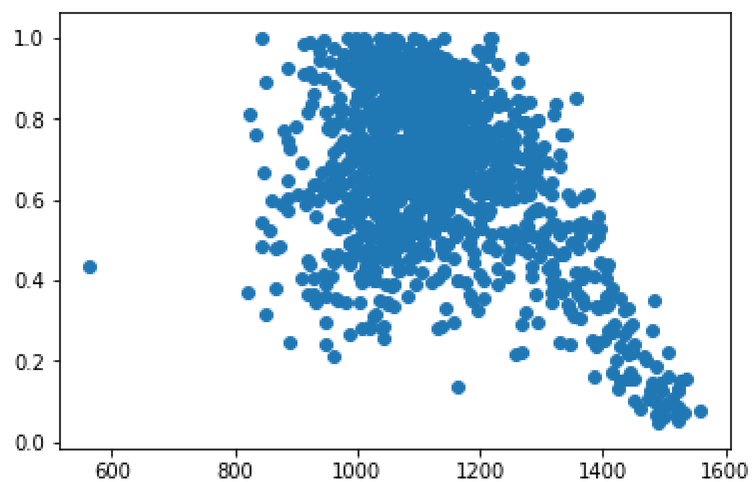
In [27]:
```python
df['SAT_AVG'].corr(df['ADM_RATE'])
```

Out[27]: -0.4106203354294415

In [28]:
```python
#the correlation between SAT scores
#        and admission rates is -0.4106203354294415 which means it's weak correlation
```

In [29]:
```python
plt.scatter(df['SAT_AVG'], df['ADM_RATE'])
```

Out[29]: <matplotlib.collections.PathCollection at 0x1a28ea7b70>



In [30]:
```
#yes there is a outliar
```

In [ ]:
```
```

In [ ]:
```
```