ncaa_region_optimizer

June 10, 2019

1 Genetic Algorithms for Region Paritioning

We will be using some Python modules installed by pip rather than Anaconda, so I must adjust the import path.

```
In [1]: import sys
        sys.path.insert( 1, '/usr/local/lib/python3.7/site-packages' )
```

1.1 Import the data

```
In [25]: import pandas
         data_filename = 'wrestling-schools-data.csv'
         data_table = pandas.read_csv( data_filename )
         len( data_table )
```

```
Out[25]: 106
```

1.1.1 Preview it

```
In [26]: data_table[:5]
```

```
Out[26]:    UniqueID           College/University Name                  Street  \
         0         1                     Adrian College          10 S Madison St
         1         2  Alfred State College  (add 2018)  10 Upper College Drive
         2         3                      Alma College        614 W Superior St
         3         4                          Augsburg      2211 Riverside Ave
         4         5                   Augustana (IL)              639 38th St

                 City  State   Latitude  Longitude  Power-1  Power-2  NCAA Asgt  \
         0       Adrian     MI  41.899337 -84.044547   2.4514    2.927          3
         1       Alfred     NY  42.254334 -77.789646   0.0000    0.000          0
         2         Alma     MI  43.380011 -84.655654   5.1091    5.941          3
         3  Minneapolis     MN  44.963541 -93.267835   9.6340    8.890          2
         4  Rock Island     IL  41.470591 -90.583733   0.0000    0.301          1

            ND Asgt  ND Asgt2  ND Asgt3
```

```
0      2.0       1.0       5.0
1      NaN       NaN       NaN
2      6.0       1.0       5.0
3      6.0       5.0       3.0
4      2.0       4.0       3.0
```

1.1.2 Drop schools we don't want in this analysis

Some schools were dropped for various domain-specific reasons. See paper.

```
In [4]: data_table = pandas.concat( [ data_table[:31], data_table[32:61], data_table[62:85], da
        num_schools = len( data_table )
        num_schools
```

```
Out[4]: 103
```

1.1.3 Make it easy to fetch desired rows/columns

```
In [5]: def school ( key ):
            column = 'UniqueID' if type( key ) == int else 'College/University Name'
            return data_table[data_table[column] == key].iloc[0]
        ( SCH_ID, SCH_NAME, SCH_ADDR, SCH_CITY, SCH_STATE, SCH_LAT, SCH_LNG, SCH_POW1, SCH_POW2
          SCH_NCAA, SCH_ND1, SCH_ND2, SCH_ND3 ) = list( data_table.columns.values )
        def all_ids ():
            return list( data_table['UniqueID'] )
        def index_to_id ( index ):
            return all_ids()[index]
        # print( school( 2 )[SCH_NAME] )
        # print( school( 'Augsburg' )[SCH_ID] )
        # print( school( 50 )[SCH_LAT], get_school( 50 )[SCH_LNG] )
```

1.2 Map distance tools

Import tools that can compute distance on the (curved) surface of the earth.

```
In [6]: from geopy.distance import great_circle
        def school_latlng ( school ):
            return ( school[SCH_LAT], school[SCH_LNG] )
```

Now pre-compute the distance between any two pair of schools and cache it in a matrix, because we'll be asking these distance questions a million times below, and this cache will speed it up a lot.

```
In [7]: school_locations = [ school_latlng( school( index_to_id( i ) ) )
                             for i in range( num_schools ) ]
        distance_matrix = [ [ great_circle( school_locations[i], school_locations[j] ).miles
                              for i in range( num_schools ) ] for j in range( num_schools ) ]
        def distance_lookup ( school_index1, school_index2 ):
            return distance_matrix[school_index1][school_index2]
```

1.3 Utilities for partitions

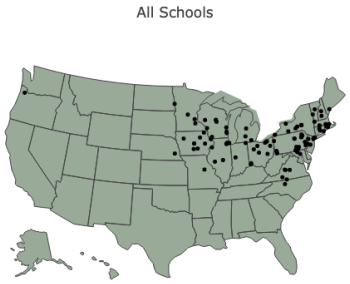
```
In [8]: num_parts_in_partition = 6
        def indices_for_part_in_partition ( part_index, partition ):
            return [ i for i in range( len( partition ) ) if partition[i] == part_index ]
        def schools_in_part_in_partition ( part_index, partition ):
            return [ school( index_to_id( i ) )
                     for i in indices_for_part_in_partition( part_index, partition ) ]
        def size_of_part_in_partition ( part_index, partition ):
            return len( indices_for_part_in_partition( part_index, partition ) )
        def random_partition ():
            import random
            return [ random.randint( 0, num_parts_in_partition ) for i in range( num_schools )
```

```
In [9]: import statistics
        def print_partition ( partition ):
            for part_index in range( num_parts_in_partition ):
                schools = schools_in_part_in_partition( part_index, partition )
                powers = [ school[SCH_POW2] for school in schools ]
                print( 'Region {:1d}, {:2d} schools, mean power {:7.5f} (stdev {:7.5f}):'.forma
                    part_index + 1, size_of_part_in_partition( part_index, partition ),
                    statistics.mean( powers ), statistics.stdev( powers ) ) )
                print( '------------------------------------------------------------' )
                centroid = (
                    statistics.mean( [ school[SCH_LAT] for school in schools ] ),
                    statistics.mean( [ school[SCH_LNG] for school in schools ] ),
                )
                print( '    Centroid: {:7.3f} lat, {:7.3f} lon'.format(
                    centroid[0], centroid[1] ) )
                latlngs = [ school_latlng( school ) for school in schools ]
                print( '    Mean distance to centroid: {:8.3f} miles'.format(
                    statistics.mean( [ great_circle( centroid, latlng ).miles for latlng in la
                ) )
                for s in schools:
                    print( '        {:30.30s} {:30.30s} {:>7.1f} miles'.format(
                        s[SCH_NAME],
                        '{}, {}, {}'.format( s[SCH_ADDR], s[SCH_CITY], s[SCH_STATE] ),
                        great_circle( school_latlng( s ), centroid ).miles
                    ) )
                print()
        # print_partition( random_partition() )
```

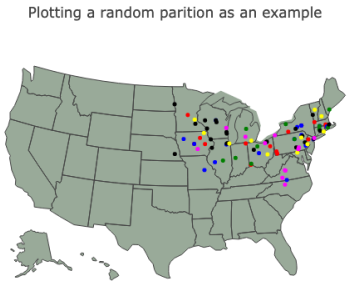
1.4 Import map-drawing tools

```
In [10]: from plotly_for_usa_points import usa_map # this works only if you've done conda inst
         # also one needs all the extensions mentioned here:
         # https://github.com/plotly/plotly.py#jupyterlab-support-python-35

In [11]: usa_map( { 'black' : [ school_latlng( school( id ) ) for id in all_ids() ] }, 'All Sc
```




```
In [12]: def partition_map ( schools_partition, title = 'Partition of All Schools' ):
             colors = [ 'red', 'blue', 'green', 'black', 'yellow', 'magenta', 'cyan', 'white',
             def points_in_part ( part_index ):
                 return [ school_latlng( s )
                          for s in schools_in_part_in_partition( part_index, schools_partition
             return usa_map( {
                 colors[i] : points_in_part( i ) for i in range( max( schools_partition ) )
             }, title )
         partition_map( random_partition(), 'Plotting a random parition as an example' )
```



1.5 Components of the Objective Function

First, we will want to experiment with the range of the various components of the objective function, to see how we should rescale them to match each other.

```
In [13]: def range_experiment ( func, num_tries=100 ):
             data = [ func( random_partition() ) for i in range( num_tries ) ]
             return min( data ), max( data )
```

1.5.1 Component 1: Variance of size of parts in the partition

```
In [14]: def part_size_variance ( partition ):
             return statistics.variance( [
                 size_of_part_in_partition( part, partition )
                 for part in range( num_parts_in_partition )
             ] )
         # print( range_experiment( part_size_variance ) ) # gives a max in the 50s
         # plus we want size variance to be bad, so we need a -1 multiplier, so:
         def obj_fn_component_1 ( partition ):
             return part_size_variance( partition ) * -1.0 / 50
```

1.5.2 Component 2: Total distance between schools in each part of the partition

```
In [15]: def total_distance_in_one_part ( part_index, partition ):
             indices = indices_for_part_in_partition( part_index, partition )
             return sum( [ distance_lookup( i, j )
                           for i in indices for j in indices if i < j ] )
         def total_distance_of_all_parts ( partition ):
             return sum( [ total_distance_in_one_part( index, partition )
                           for index in range( num_parts_in_partition ) ] )
         # print( range_experiment( total_distance_of_all_parts ) )
         # plus we want travel distance to be bad, so we need a -1 multiplier, so:
         def obj_fn_component_2 ( partition ):
             return total_distance_of_all_parts( partition ) * -1.0 / 50000
```

1.5.3 Component 3: Variance of mean powers of each part in partition

```
In [16]: def mean_power_of_part ( part_index, partition ):
             powers = [ s[SCH_POW2] for s in schools_in_part_in_partition( part_index, partitio
             if len( powers ) > 0:
                 return statistics.mean( powers )
             else:
                 return 0
         def part_power_variance ( partition ):
             return statistics.variance( [
                 mean_power_of_part( part, partition )
                 for part in range( num_parts_in_partition )
             ] )
         # print( range_experiment( part_power_variance ) )
```

```
        # plus we want power variance to be bad, so we need a -1 multiplier, so:
        def obj_fn_component_3 ( partition ):
            return part_power_variance( partition ) * -1.0 / 5
```

1.5.4 Objective function: sum of 3 components

```
In [17]: def objective_function ( partition ):
             return obj_fn_component_1( partition ) \
                  + obj_fn_component_2( partition ) \
                  + obj_fn_component_3( partition )
```

1.6 Solving the problem

```
In [18]: num_generations = 10000
         def progress_bar ( name="Progress", size=num_generations ):
             from tqdm import tqdm_notebook
             bar = tqdm_notebook( range( size ), desc=name )
             def step ( *args ):
                 bar.update( 1 )
                 bar.display()
             return step
         from ga_for_partitions import optimize_partition
         best, fitness_curve = optimize_partition(
             objective_function = objective_function,
             initial_pool = [ random_partition() for i in range( num_parts_in_partition ) ],
             size_of_partition = num_parts_in_partition,
             prob_mutate = 0.1,
             num_generations = num_generations,
             progress_callback = progress_bar()
         )
```

```
HBox(children=(IntProgress(value=0, description='Progress', max=10000, style=ProgressStyle(des
```

```
After 10000 generations: max score =  -1.6606   100% done, 15:56/15:56 (00:00)
```

```
In [22]: print_partition( best )
```

```
Region 1, 13 schools, mean power 2.65169 (stdev 4.79868):
----------------------------------------------------------
    Centroid:  39.249 lat, -78.329 lon
    Mean distance to centroid:  137.454 miles
        Alfred State College  (add 201 10 Upper College Drive, Alfred   209.5 miles
        Averett University (add 2017)  420 W Main St, Danville, VA      193.4 miles
        Elizabethtown                  1 Alpha Dr, Elizabethtown, PA    111.2 miles
        Ferrum College                 215 Ferrum Mountain Rd, Ferrum   184.5 miles
```

```
        Gettysburg                    300 N Washington St, Gettysbur    70.8 miles
        Greensboro College            815 W Market St, Greensboro, N   233.9 miles
        Johns Hopkins                 400 N. Charles Street, Baltimo    92.0 miles
        McDaniel                      2 College Hill, Westminster, M    74.2 miles
        Messiah                       1 College Ave, Mechanicsburg,     96.7 miles
        Thiel                         College Ave, Greenville, PA      184.1 miles
        Washington and Jefferson      60 S Lincoln St, Washington, P   120.3 miles
        Washington and Lee            204 W Washington St,, Lexingto   117.9 miles
        York (PA)                     443 Country Club Rd, York, PA     98.4 miles

Region 2, 12 schools, mean power 3.00458 (stdev 6.42396):
---------------------------------------------------------
    Centroid:  43.408 lat, -91.255 lon
    Mean distance to centroid:  158.209 miles
        Buena Vista                   610 W 4th St, Storm Lake, IA     206.1 miles
        Chicago                       5801 S Ellis Ave, Chicago , IL   211.3 miles
        Lakeland                      W3718 South Dr, Plymouth, WI     166.3 miles
        MacMurray College             447 E College Ave, Jacksonvill   259.6 miles
        Simpson                       701 N C St,, Indianola, IA       184.1 miles
        St. Johns (MN)                2850 Abbey Plaza, Collegeville   215.6 miles
        St. Olaf                      1520 St Olaf Ave, Northfield,    119.6 miles
        Wartburg                      100 Wartburg Blvd, Waverly, IA    77.4 miles
        Wisconsin-Eau Claire          105 Garfield Ave, Eau Claire,     98.3 miles
        Wisconsin-Oshkosh             800 Algoma Blvd, Oshkosh, WI     122.8 miles
        Wisconsin-Stevens Point       100 Main St, Stevens Point, WI   114.5 miles
        Wisconsin-Whitewater          800 W Main St, Whitewater, WI    122.8 miles

Region 3, 13 schools, mean power 1.84731 (stdev 3.25232):
---------------------------------------------------------
    Centroid:  41.180 lat, -83.685 lon
    Mean distance to centroid:  132.526 miles
        Adrian College                10 S Madison St, Adrian, MI       53.1 miles
        Baldwin Wallace               275 Eastland Rd, Berea, OH        95.4 miles
        Case Western Reserve          10900 Euclid Ave, Cleveland, O   110.1 miles
        Heidelberg                    310 E Market St, Tiffin, OH       26.6 miles
        John Carroll                  1 John Carroll Boulevard, Univ   113.6 miles
        Luther                        700 College Dr, Decorah, IA      439.3 miles
        Mount St. Joseph              5701 Delhi Ave, Cincinnati, OH   147.2 miles
        Muskingum                     163 Stormont St, New Concord,    130.9 miles
        Ohio Wesleyan (add 2018)      61 S Sandusky St, Delaware, OH    69.5 miles
        Olivet                        320 S Main St, Olivet, MI        108.2 miles
        Penn State - Behrend (add 2017 4701 Behrend College D, Erie,   197.4 miles
        Thomas More College           333 Thomas More Pkwy, Crestvie   156.1 miles
        Trine University              1 University Ave, Angola, IN      75.5 miles

Region 4, 14 schools, mean power 1.61357 (stdev 2.04692):
---------------------------------------------------------
    Centroid:  42.157 lat, -72.218 lon
```

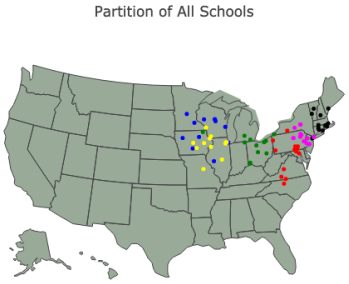
```
    Mean distance to centroid:   79.841 miles
        Bridgewater State University   131 Summer Street, Bridgewater    65.0 miles
        Castleton University          62 Alumni Dr,, Castleton, VT     111.4 miles
        Coast Guard                   31 Mohegan Ave Pkwy, New Londo    57.3 miles
        Johnson & Wales (RI)          8 Abbott Park Pl, Providence,     47.4 miles
        New England College           98 Bridge St,, Henniker, NH       73.7 miles
        New York University           383 Lafayette Street, New York   134.9 miles
        Norwich                       158 Harmon Dr, Northfield, VT    139.7 miles
        Rhode Island                  600 Mt Pleasant Ave, Providenc    44.6 miles
        Roger Williams                1 Old Ferry Road, Bristol, RI     58.8 miles
        Southern Maine                96 Falmouth St, Portland, ME     146.9 miles
        Springfield                   263 Alden Street, Springfield,    17.8 miles
        Stevens Institute Of Technolog 1 Castle Point Terrace, Hoboke   135.3 miles
        Trinity (CT)                  300 Summit St, Hartford, CT       37.4 miles
        Wesleyan (CT)                 45 Wyllys Ave, Middletown, CT     47.7 miles

Region 5, 13 schools, mean power 2.66454 (stdev 2.62010):
----------------------------------------------------------
    Centroid:  41.741 lat, -90.524 lon
    Mean distance to centroid:  109.925 miles
        Augustana (IL)                639 38th St, Rock Island, IL      18.9 miles
        Central College               812 University St, Pella , IA    125.9 miles
        Coe                           1220 First Avenue NE, Cedar Ra    61.4 miles
        Cornell College               600 1st St, Mt Vernon, IA        152.3 miles
        Dubuque                       2000 University Ave, Dubuque,     53.0 miles
        Elmhurst                      190 S Prospect Ave, Elmhurst,    133.3 miles
        Loras                         450 Alta Vista St,, Dubuque, I    52.8 miles
        Millikin                      184 W Main St, Decatur, IL       154.5 miles
        North Central (IL)            30 North Brainard Street, Nape   121.8 miles
         Westminster (add 2017)       501 Westminster Ave, Fulton, M   213.5 miles
        Wheaton (IL)                  501 College Ave, Wheaton, IL     124.7 miles
        Wisconsin-La Crosse           1725 State St, La Crosse , WI    148.2 miles
        Wisconsin-Platteville         1 University Plaza, Plattville    68.6 miles

Region 6, 15 schools, mean power 1.56467 (stdev 1.94103):
----------------------------------------------------------
    Centroid:  41.405 lat, -75.596 lon
    Mean distance to centroid:   63.167 miles
        Centenary (NJ)                400 Jefferson St, Hackettstown    55.3 miles
        Delaware Valley               700 E Butler Ave, Doylestown,     79.2 miles
        Ithaca                        953 Danby Rd, Ithaca, NY          84.3 miles
        Keystone College              1 College Rd, La Plume, PA        13.5 miles
        King's (PA)                   133 N River St, Wilkes-Barre,     18.2 miles
        Lycoming                      700 College Pl, Williamsport,     75.5 miles
        Merchant Marine               300 Steamboat Rd, Kings Point,   104.8 miles
        Muhlenberg                    2400 W Chew St, Allentown, PA     56.2 miles
        Pennsylvania College (add 2017 1 College Ave, Williamsport, P    74.9 miles
        Rochester Institute of Technol Lomb Memorial Dr, Rochester, N    87.7 miles
```

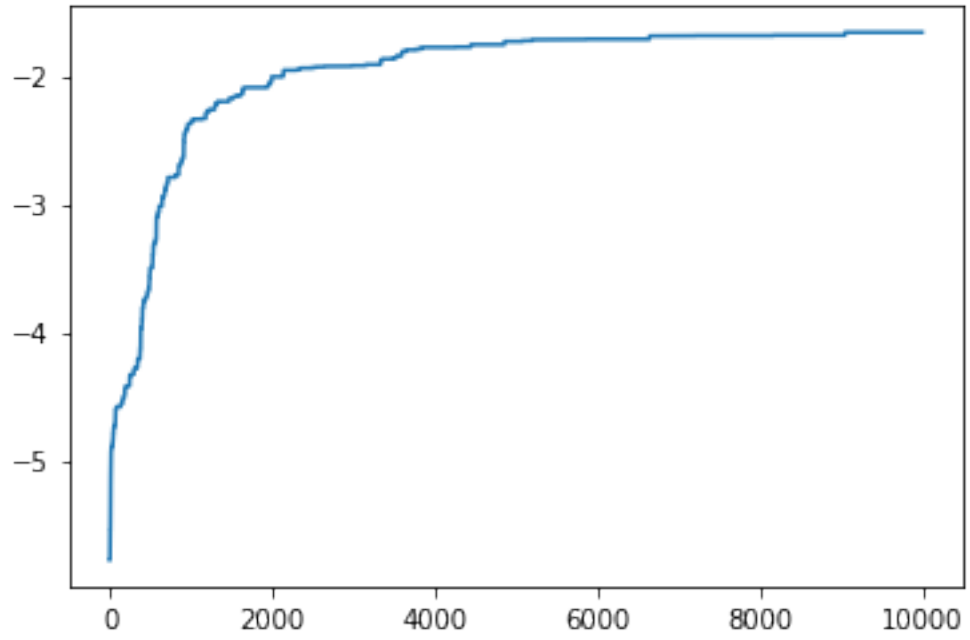
```
Scranton                        800 Linden St, Scranton, PA        3.5 miles
SUNY-Oneonta                    08 Ravine Pkwy, Oneonta, NY       87.7 miles
SUNY-Oswego                     7060 New York 104, Oswego, NY     87.7 miles
The College of New Jersey       2000 Pennington Rd, Ewing Town   100.1 miles
Wilkes                          84 W South St, Wilkes-Barre, P    18.8 miles
```

In [23]:
```
partition_map( best )
```



In [24]:
```
from matplotlib import pyplot as plt
%matplotlib inline
plt.plot( range( len( fitness_curve ) ), fitness_curve )
```

Out[24]: [<matplotlib.lines.Line2D at 0x112b4d6d8>]



In [ ]: