Data

Source, scraping method, cleanup, etc.

Snippet of Sample Data

```
In [2]:  import numpy as np
         import pandas as pd

         df = pd.read_csv("data/collegedata_normalized.csv", index_col=0)
         df.head(10)
```

Out[2]:

	studentID	classrank	admissionstest	AP	averageAP	SATsubject	GPA
0	S50C3UECT8	NaN	0.838909	7	1.067927	0.323654	-0.18810
1	GBWZQQRBEV	NaN	0.666993	7	0.661638	-0.440897	0.493067
2	MXXLWO1HQ2	NaN	0.208552	0	NaN	0.323654	0.395752
3	5KSL7C8SLZ	NaN	1.297350	7	0.864783	1.088204	0.103824
4	RQWLNGGZ49	NaN	0.323162	1	-0.354087	-0.440897	0.541710
5	A7WDLWR2VM	NaN	0.323162	0	NaN	-0.440897	0.784989
6	CECT9K8GDY	NaN	0.724298	0	NaN	-0.440897	0.298443
7	9PM1B51CXG	NaN	-2.542098	3	-0.828092	-0.440897	-0.67465
8	G14LB2SV7O	NaN	-0.020669	7	-0.354087	-1.969998	0.493067
9	PSBRN09QGH	NaN	-0.135280	4	0.001416	-0.440897	0.784989

10 rows × 34 columns

```
In [ ]:
```