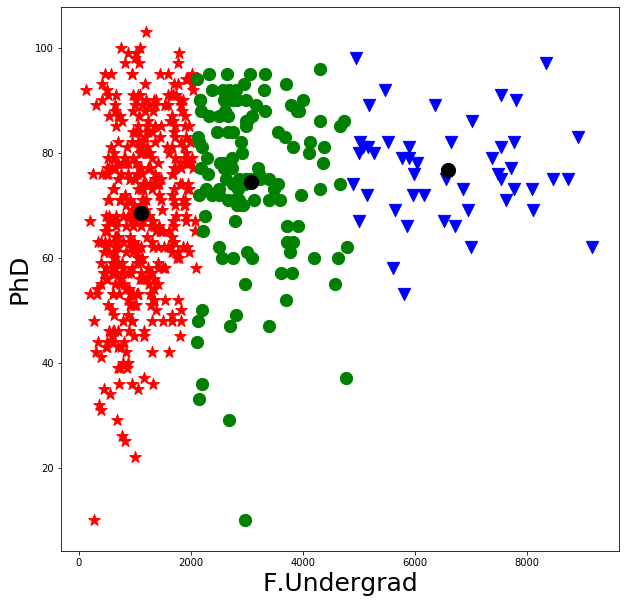
**Case study report**

This is report on case study on a dataset on US colleges. It has 17 features and 777 data point for all the colleges in the United States. Some important features mentioned are applicants, enrolees, top 10 SAT scorers in the college etc

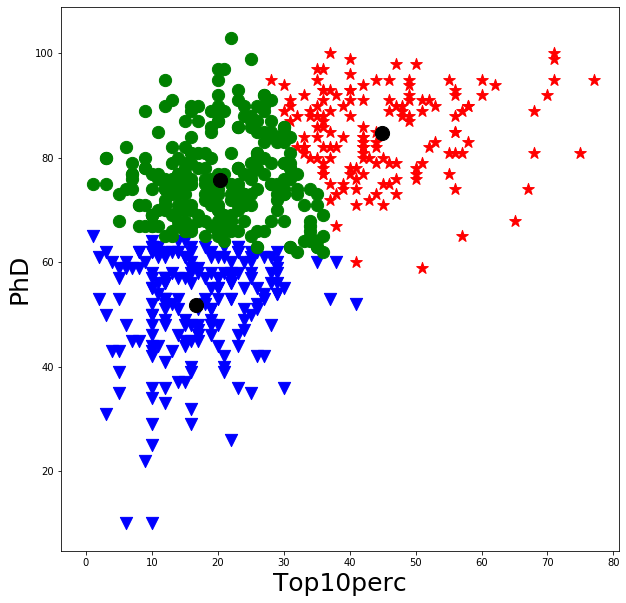
Here I’ve tried to maximize research output of the college, that is increasing number of PhD’s of a college so as to get more research grants to a college.

**Undergraduates vs PhDs**

****

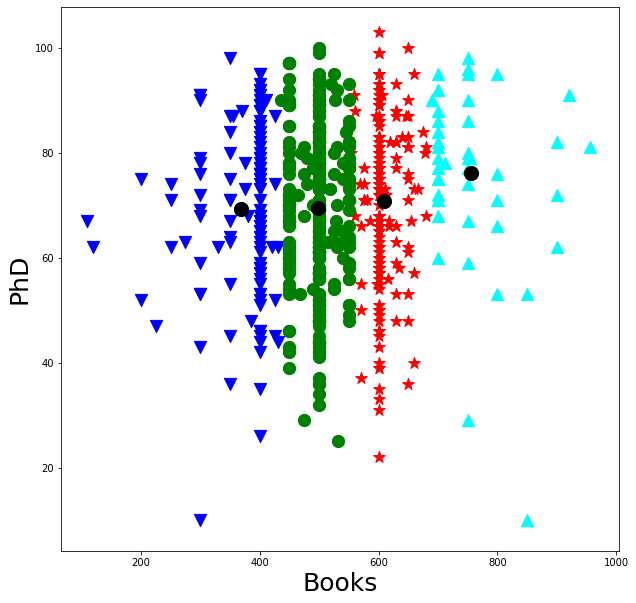
Here by looking at the centroid of clusters colleges having more than 2000 undergrad student do a little better at research output. It’s the same for postgrad students.

**Top 10 SAT scores vs PhDs**

****

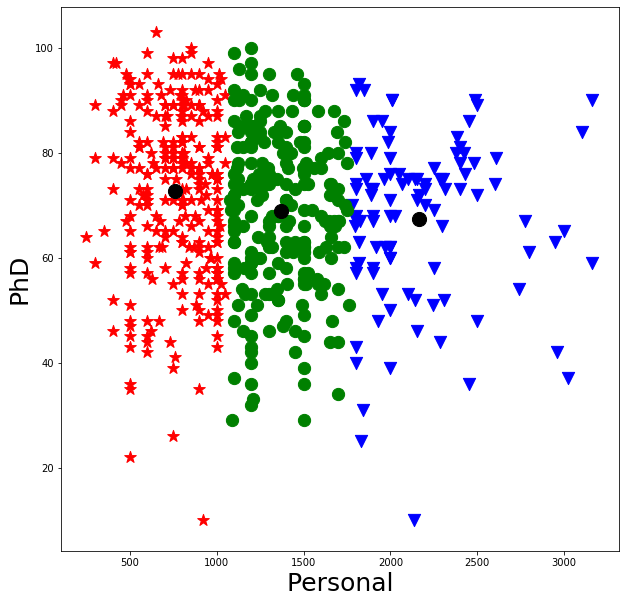
Here we can see that colleges getting students with top scores in Standard aptitude test get more research output.

**Number of books vs PhDs**

****

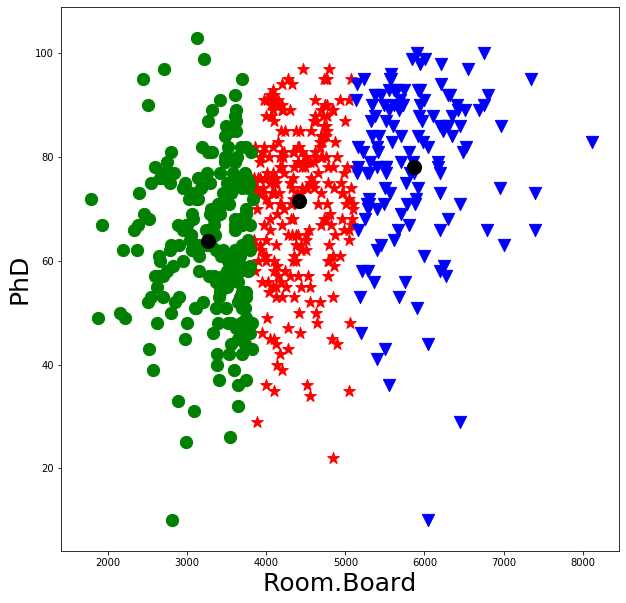
Here book seem to making no difference having more books, but have more than 700 books in their curriculum had little better research output.

**Personal vs PhDs**

****

Here the most counter intuitive inference seems to be that colleges that have the number of faculty members to less than 1000 have more research output.

**Number of boarding rooms vs PhD**

****

Here we can see that colleges having more than 5000 boarding rooms will do better an putting out more research.

**Conclusion:**

By using K-means cluster analysis we’ve come to the conclusions above.