Where did you find this data set? What do the data represent, how were they gathered or produced?

I found this data set from the following link: http://lib.stat.cmu.edu/DASL/Datafiles/Colleges.html. The site has a large amount of data sets that are available. The data sets cover a wide range of topics including sports, education, economics, nature, and nutrition. This data set is data collected from 50 universities around the country. The 50 universities are the top 25 liberal arts colleges and the top 25 research universities. There are 7 variables for each school. They are as follows:

1) School type (Liberal Arts or University)
2) Median combined Math and Verbal SAT score of student
3) % of Applicants accepted
4) Money spent per student in dollars
5) % of students in the top 10% of their high school graduating class
6) % of faculty at the institution that have PhD degrees
7) % of students at institution who eventually graduate

The 50 colleges are listed below:

Liberal Arts Schools	**Research Universities**
Amherst	Harvard
Swarthmore	Stanford
Williams	Yale
Bowdoin	Princeton
Wellesley	Cal Tech
Pomona	MIT
Wesleyan	Duke
Middlebury	Dartmouth
Smith	Cornell
Davidson	Columbia
Vassar	University of Chicago
Carleton	Brown
Claremont McKenna	University of Pennsylvania
Oberlin	Berkeley
Washington & Lee	John Hopkins
Grinnell	Rice
Mount Holyoke	UCLA
Colby	University of Virginia

Hamilton	Georgetown
Bates	UNC
Haverford	University of Michigan
Colgate	Carnegie Mellon
Bryn Mamr	Northwestern
Occidental	Washington University
Barnard	University of Rochester

The website did not say how the data was obtained or the school year in which the data is from. Information like the type of school, SAT scores, percent of students admitted, and percent of students in top 10% of high school class can easily be found from www.collegeboard.com. I did not try to find the data for the other variables but I would imagine the information could be found on the internet.

Why would this data be of interest to students in a class you are teaching?

This data may be of interest to my students because some of them may be planning on going to college in the future. The colleges are all extremely competitive colleges so I would not expect most of them to be interested in these specific colleges but it could spark their interest about other colleges. This data may be especially usual for an AP Statistics class because those students are more likely to want to attend college.

Depending on the area I am teaching in, many of my students may not be planning on going to college so this data may not be of much interest to them. In this case, the emphasis could be on if schools are being efficient in the way they spend their money. We could focus on the percent of students who graduate and the amount of money spent per student to determine if schools are spending their money in an efficient manner.

What statistical concept could be illustrated with this data? What other questions might you have students investigate with this data set?

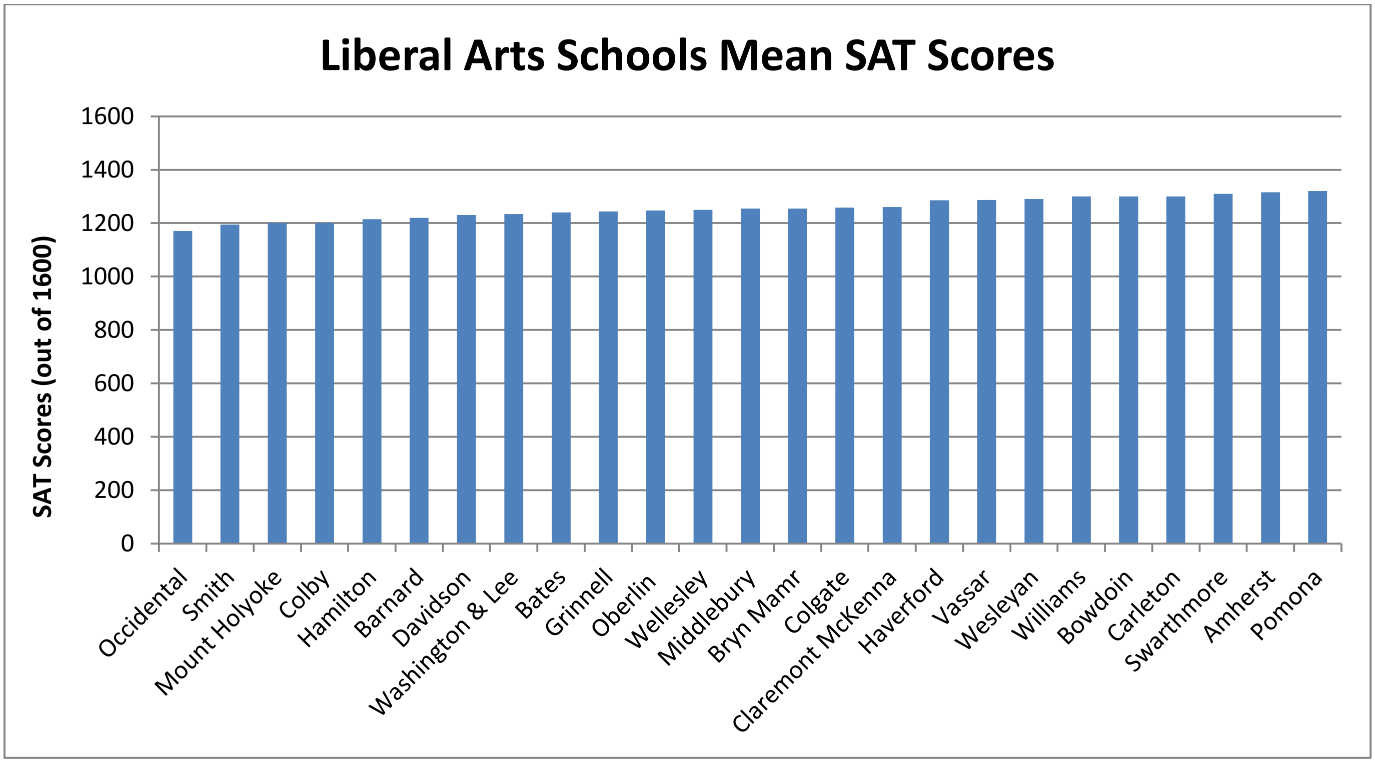
This data set could be used to illustrate basic descriptive statistics like mean, median, and standard deviation. The data could also be divided into two categories (Liberal Arts or Research University). Using this data, students could determine if one of the categories skewed the data. Students could also explore correlation between two of the variables. For example, students could see if there was a correlation between amount of money spent per student and percent of students who graduate. Another example would be to look for a correlation between SAT scores and the percent of students who graduate. Dot plots, histograms, and scatter plots could also be created from this data.

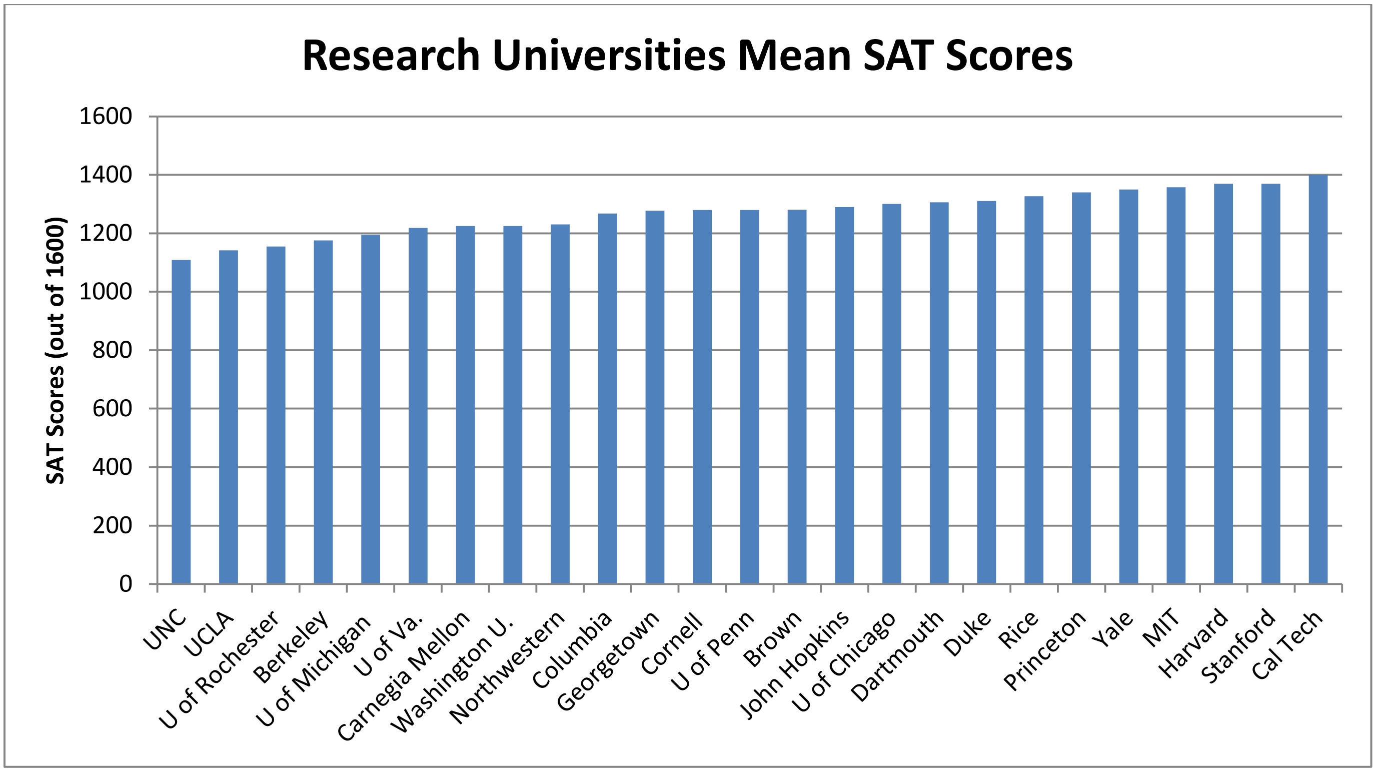
Students could also look for correlation between percent of professors with a PhD and percent of students who graduate or between SAT scores and the percent of student admitted to the school. The students could look at these correlations for the entire data set and for Research Universities and Liberal Arts colleges. They could compare the results from the different categories and make hypotheses as to why there are differences between the two types of universities. These particular schools would not be of interest to most students because they are the top schools in the country. Students could use www.collegeboard.com to research information about schools they are interested. The students could then compare the information about these schools they researched. If the students are interested in attending college, as a part of a project, they could write a report on their findings that persuades their parents to let them attend the school they want to attend.

The first aspect of the data we will look at is the average SAT scores for the schools:

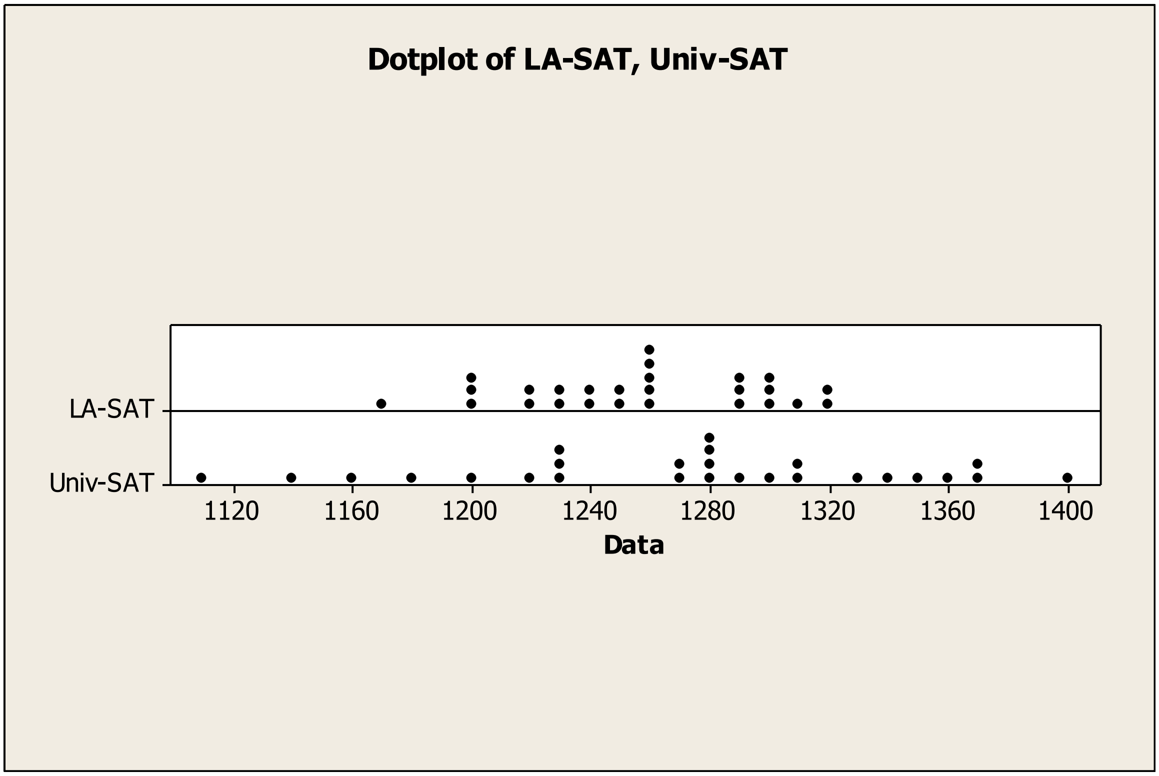
	Mean	Median	Standard Deviation	Minimum	Maximum	Range
All Schools	1263.24	1264	61.685	1109	1400	291
Liberal Arts Schools	1255.2	1255	41.49	1170	1320	150
Research Universities	1271.28	1280	76.89	1109	1400	291

The following are a bar graph of the SAT scores for each school. Because there are so many schools in the data set, they are divided into two groups. The first graph is for the Liberal Arts schools and the second is for Research Universities. (Both graphs were created using Microsoft Excel)





The average scores of all of the schools was about 1263 (out of 1600) and the mean for research universities was higher than the overall mean while the liberal arts average was below the overall mean. The bar graphs with each of the school's scores show that the scores are fairly consistent in both groups of schools. I did find it interesting that although the research universities had a high mean than the liberal arts, the overall minimum and maximum scores were both research universities. This may indicate that the liberal arts schools are more similar in their academic difficulty but the research universities are some of the most difficult schools in the country. This is also shown in the range of the two groups. The range for the liberal arts schools is 150 while the range for the research universities is 291. Dot plots can also be used to show that the scores for liberal arts schools have less variance. As seen in the box plots below, the scores for liberal arts schools are more grouped together than the research schools. The following dot plot was created used MiniTab 15.



Both groups of schools, and the overall data, tend to be symmetric. This is seen in the fact that the median and mean for each group are very close together. The mean and median scores for the liberal art schools and the overall data are nearly the exact same score. The median for the research schools is slightly higher than the mean for the group. This would indicate that the research universities SAT scores are slightly left skewed.

In doing some more research at www.collegeboard.com I also found the average score for all of the students in the graduating class of 2011 who took the SAT to be 1011. Even students at University of North Carolina, the school with the lowest SAT score in this data set, had an average SAT score that was approximately 100 points above the national average. It should be noted that the average of 1011 is for the year 2011 but the data set did not indicate the year that the data it from.

Another aspect of the data we will analyze is the relationship between the amount of money a school spends per student and the percent of students who graduate from the school. First we will give summary of the data for each of the variables. The data for each individual school can be found at the link on the main page for the entire data set. The following table is a summary for the dollars spent per student.

	Mean	Median	Standard Deviation	Minimum	Maximum	Range
All Schools	$30,157.70	$24,994.50	$15,126.30	$17,520.00	$102,262.00	$84,742.00
Liberal Arts Schools	$21,753.60	$20,377.00	$3,454.53	$17,520.00	$27,879.00	$10,359.00
Research Universities	$38,561.80	$37,237.00	$17,551.80	$19,365.00	$102,262.00	$82,897.00

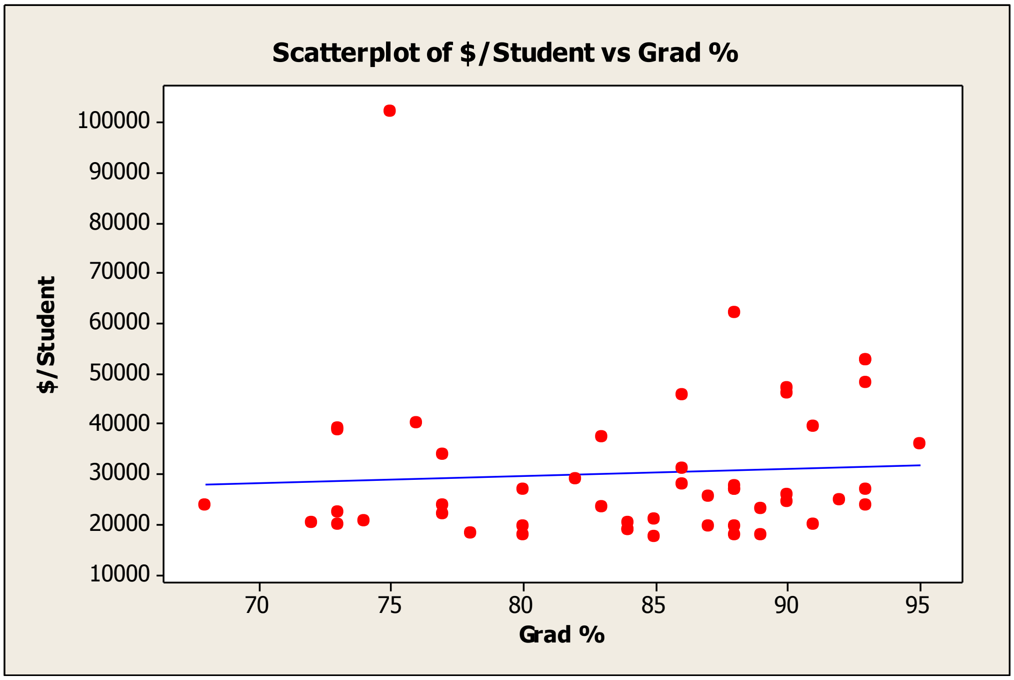
Overall, the research universities spent a large amount of dollars more per student than liberal arts schools. However, this does not necessarily mean that more money was spent on each student. The money may be spent in ways that do not directly affect or benefit the students. It should also be noted that the highest amount spent per student was $102,262.00 by Cal Tech and the next highest was Stanford at only $61,921.00. It is also interesting to note that Cal Tech and Stanford also had the highest SAT scores. The mean and median for the amount of money research universities spend per student are close together, indicating this data is fairly symmetric but the high standard deviation of $17,551.80 shows the data varies. The amount spent per student at liberal arts schools has a fairly small standard deviation at only $3,454.53 so this data does not vary nearly as must as the research universities. Additionally, research universities spend about $17,000 more per student than liberal arts schools.

The following table is a summary of the percent of student who graduated from the university. Keep in mind that all of the values are a percent.

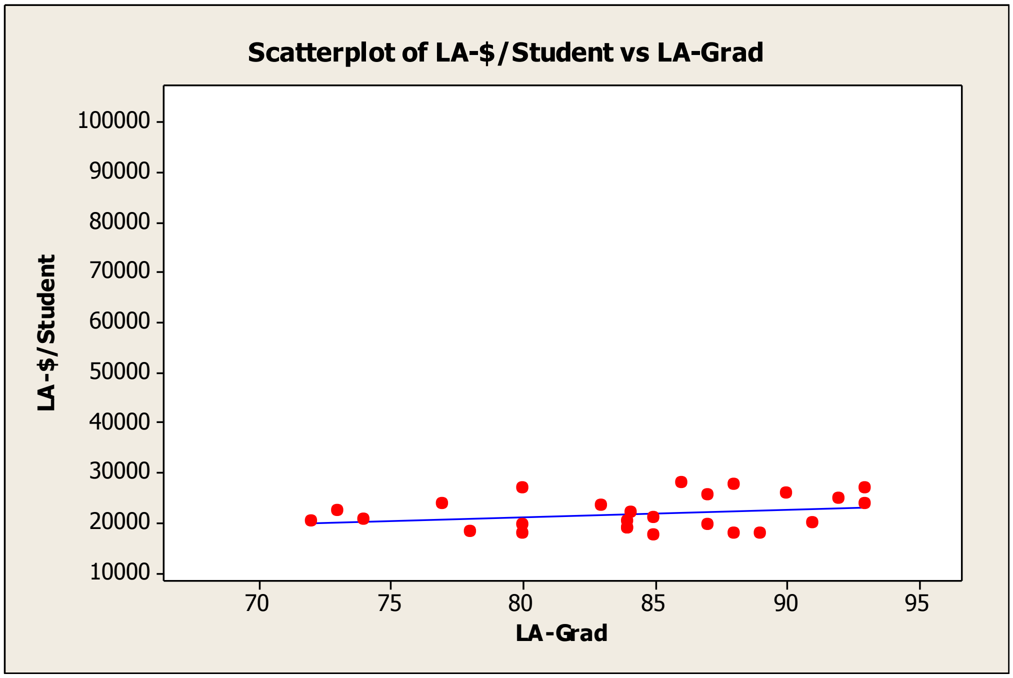
	Mean	Median	Standard Deviation	Minimum	Maximum	Range
All Schools	84.16	86	6.96	68	95	27
Liberal Arts Schools	84.12	85	6.09	72	93	21
Research Universities	84.20	88	7.86	68	95	27

The mean, median, and standard deviation is nearly identical for all three of the groups. The only difference in groups is the range. The range for the liberal arts schools is slightly smaller than the overall range and the range for research universities.

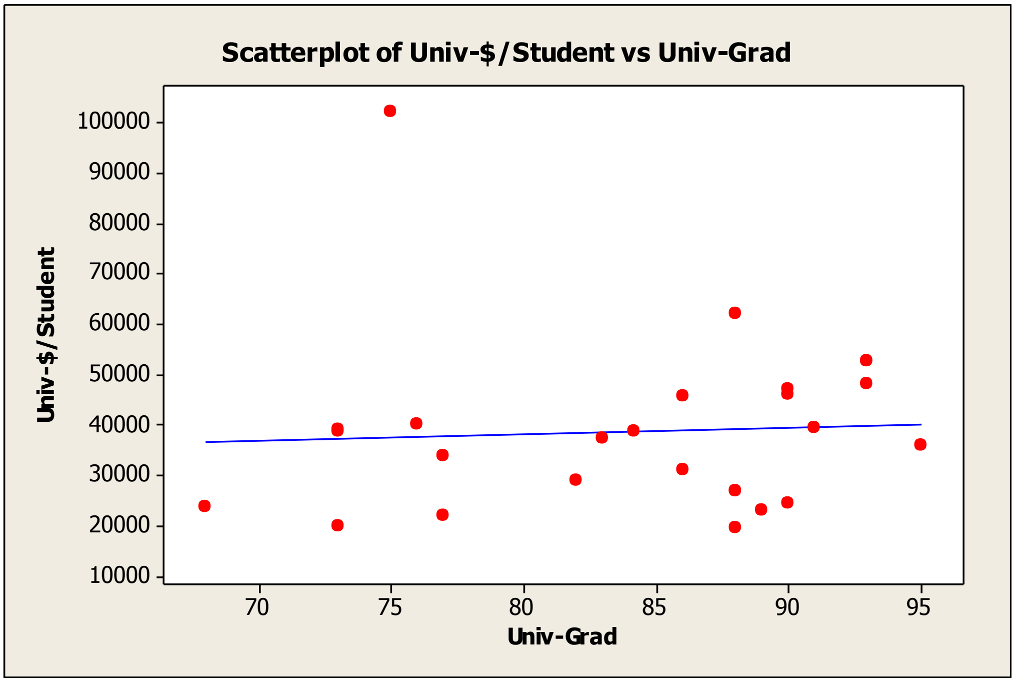
We will now look at the relationships between these two variables. The percent of students who graduated are along the x-axis and the amount of money spent per student is along the y-axis. The scatter plot was made using MiniTab 15. The blue line in each of the graphs is the linear regression equation. The following is a scatter plot of the data for all of the schools.



The following is the scatter plot for the liberal arts schools.



The following is the scatter plot of the research universities.



Using MiniTab 15, we found the Pearson correlation coefficient for each of the categories. This data is found in the following table:

	Pearson correlation coefficient
All Schools	0.067
Liberal Arts Schools	0.264
Research University Schools	0.059

Overall, there appears to be no linear correlation between the amount of money a school spends per student and the percent of students who eventually graduate. The liberal arts schools are slightly more linearly correlated than then research universities.