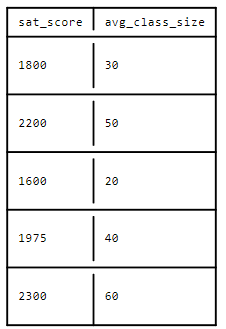
Correlations tell us how closely related two columns are. We'll be using the [r value](https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient), also called [Pearson's correlation coefficient](https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient), which measures how closely two sequences of numbers are correlated.

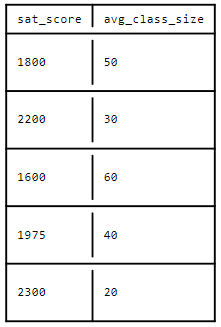
An r value falls between -1 and 1. The value tells us whether two columns are positively correlated, not correlated, or negatively correlated. The closer to 1 the r value is, the stronger the positive correlation between the two columns. The closer to -1 the r value is, the stronger the negative correlation (i.e., the more "opposite" the columns are). The closer to 0, the weaker the correlation. To learn more about r values, see the [statistics course](https://www.dataquest.io/course/probability-statistics-beginner).

The columns in the following diagram have a strong positive correlation -- when the value in class\_size is high, the corresponding value in sat\_score is also high, and vice versa:



The r value for the columns in the diagram above is .99.

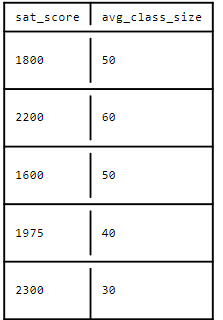
The columns in the following diagram have a strong negative correlation -- when the value in class\_size is high, the corresponding value in sat\_score is low, and when the value in sat\_score is high, the value in class\_size is low:



The r value for the columns in the diagram above is -.99.

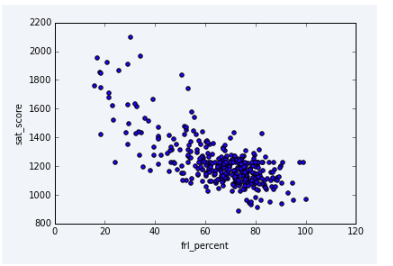
In the next diagram, the columns aren't correlated

-- class\_size and sat\_score don't have any strong pattern in their values:



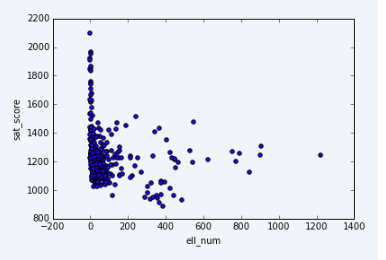
The r value for the columns in the diagram above is -.02, which is very close to 0.

In general, r values above .25 or below -.25 are enough to qualify a correlation as interesting. An r value isn't perfect, and doesn't indicate that there's a correlation -- just the possiblity of one. To really assess whether or not a correlation exists, we need to look at the data using a scatterplot to see its "shape." For example, here's a scatterplot with a very strong negative r value of -.73:



Notice how in the image above, all of the points appear to fall along a line. This pattern indicates a correlation.

Here's a scatterplot with an r value of .15, which indicates a weak correlation:



Notice how the data points in the image go in several directions, and there's no clear linear relationship. We'll explore correlations in greater detail later on in the statistics content. For now, this quick primer should be enough to get us through this project.

Because we're interested in exploring the fairness of the SAT, a strong positive or negative correlation between a demographic factor like race or gender and SAT score would be an interesting result meriting investigation. If men tended to score higher on the SAT, for example, that would indicate that the SAT is potentially unfair to women, and vice-versa.

We can use the pandas [pandas.DataFrame.corr()](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.corr.html) method to find correlations between columns in a dataframe. The method returns a new dataframe where the index for each column and row is the name of a column in the original data set.

instructions

* Use the [pandas.DataFrame.corr()](http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.corr.html) method on the combined dataframe to find all possible correlations. Assign the result to correlations.
* Filter correlations so that it only shows correlations for the column sat\_score.
* Display all of the rows in correlations and look them over.

# find all possible corellations in the combined dataset

correlations = combined.corr()

# Filter correlations so that it only shows correlations for the column sat\_score.

corellations = correlations["sat\_score"]

print(corellations)