Analyzing and Visualizing the Data

Syntax

- Find correlations between columns in a dataframe:

In [ ]:
```
combined.corr()
```

- Specify a plot type using Dataframe.plot():

In [ ]:
```
combined.plot.scatter(x='total_enrollment', y='sat_score')
```

- Convert a Pandas series to list:

In [ ]:
```
longitudes = combined["lon"].tolist()
```

Concepts

- An r value measures how closely two sequences of numbers are correlated.
- An r value ranges for `-1` to `1`.
- An r value closer to `-1` tells us the two columns are negatively correlated while an r value closer to `1` tells us the columns are positively correlated.
- The r value is also known as Pearson's correlation coefficient.

Resources

- R value (https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient)
- pandas.DataFrame.plot() (http://pandas.pydata.org/pandas-docs/stable/generated/pandas.DataFrame.plot.html)
- Correlation (https://www.mathsisfun.com/data/correlation.html)
- Guess the Correlation (http://guessthecorrelation.com/)