Edify:

General Idea: Plot data about public high schools in Philadelphia to help visualize aspects of school statistics, such as SAT and ACT scores.

Team Members: Angela Yu and Ana Mei

Goals:
- Load, specifically public high school, data from the "Data Used" section below
- Create a rudimentary website, centered around a Google map that will be used to visualize our data
- Allow the users to view information about different schools based on data types (scores, number of incidents, etc.)

Division of Work:
- data mining, database queries -> Angela
- plotting data/graphics -> Ana

Modules Used:
csv, MySQLdb, numpy, requests, urllib, time, nose

Data Used:
http://www.pde.state.pa.us/portal/server.pt/community/data_and_statistics/7202/sat_and_act_scores/674663 - SAT and ACT scores
http://webgui.phila.k12.pa.us/offices/a/accountability/open-data-initiative - General school information

About Edify:

We were able to find a bunch of interesting data about several schools near the Philadelphia area. The data originally came in the form of an Excel spreadsheet. We thought it would be very troublesome for someone to look through the whole file to find the data they are looking for, so we thought of a graphical way to represent these information. That's how Edify came about.

First, the data on the Excel spreadsheet was changed into csv files. Next, we created Python scripts that parse and inserted these data into MySQL database tables. The tables themselves were created with Python. We needed to get the latitude and longitude for each school in order to place markers on the map, so we sent request to Google in order to get a JSON that contained this data and inserted it into a table of all the school locations. Once the data was in, we used another Python script to query for the needed data. For SAT, the average total score was missing from the original dataset and we also added the linear regression slope of the total average score across the four years of data we used.

The purpose of this project is to simplify data viewing for the users, so it's important that we have a front end that allows direct user interaction. In response to this, we built a website with a very simplistic design and a large map. On the right side of the screen is a form that allows user to filter the type of data they want to be displayed. Besides the category on the main dropdown list, we have separated the data by year. For example, if you choose the "SAT Scores" on the menu, an extra section containing two radio buttons will show below. The radio button will have the year 2010 and 2011 as an indication of which year you want to see. Once the user made his or her choice, press the submit button and the map will be loaded with various markers. Each marker represent a school. By hovering your mouse over the marker, you are able to see the variety of information available from that school. In addition to just retrieving and displaying data, we also calculated the average and trend of each year for better data analysis.

Technical challenges mainly stemmed from the quality of the data (missing fields and fields with '*' as the value) and that the data came from different locations, which caused inconsistent naming conventions.

Notes:
1. We had access to our own server so we decided not to use Flask.
2. Pygmaps we found uses version 2 API of Google maps, which is currently depreciated. The version 2 API will only work until May 19, 2013. We wanted Edify to last longer that that, so we used Javascript for our Google Map instead and used Python as much as we could.