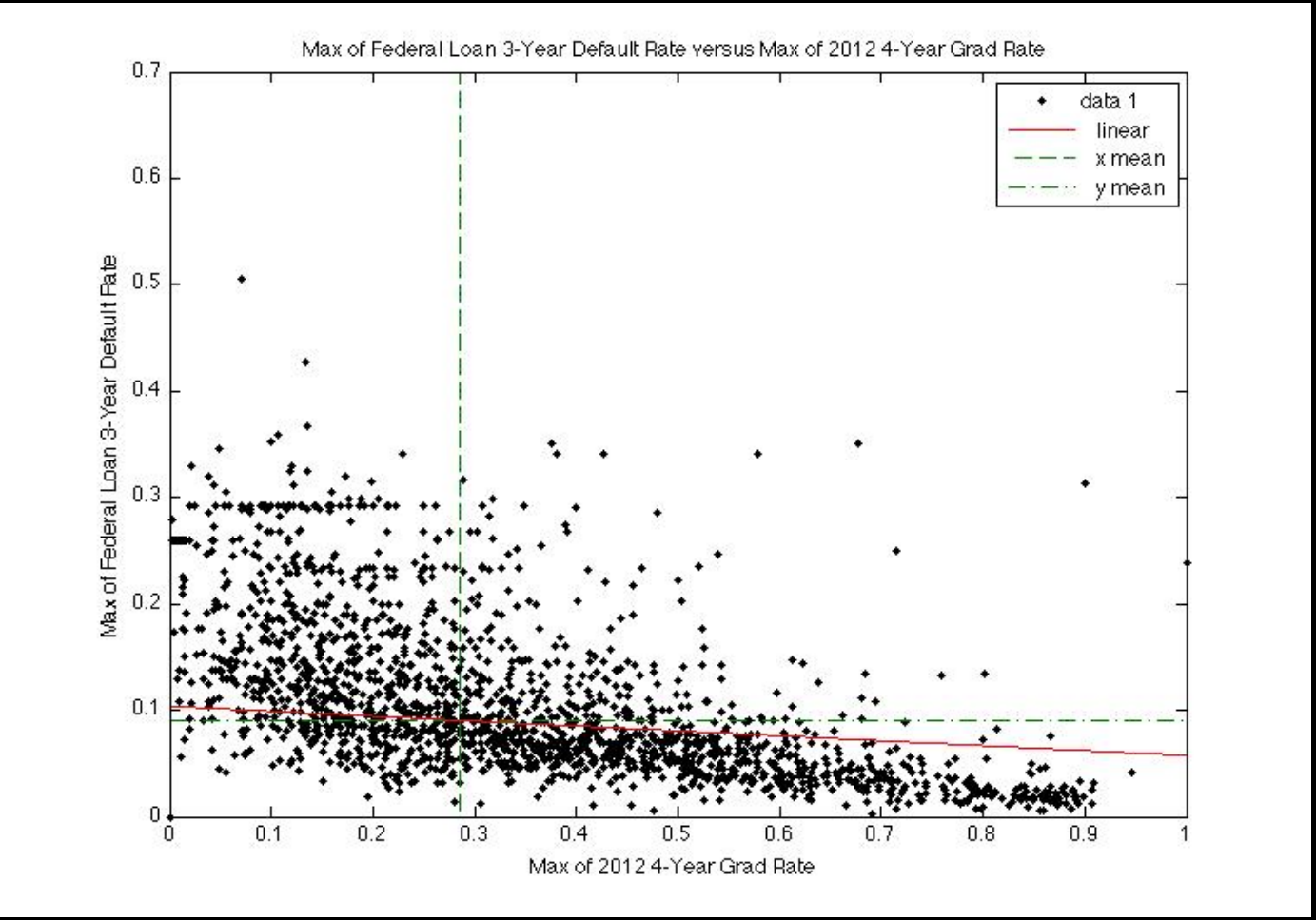
TRUITION DATA-Assignment

CHRISTIAN HANSEN

Setup

- From the excel file I made a pivot table of the parameters suggested and uploaded the maximum of the values to a csv file which I then manipulated in Matlab to plot and evaluate.

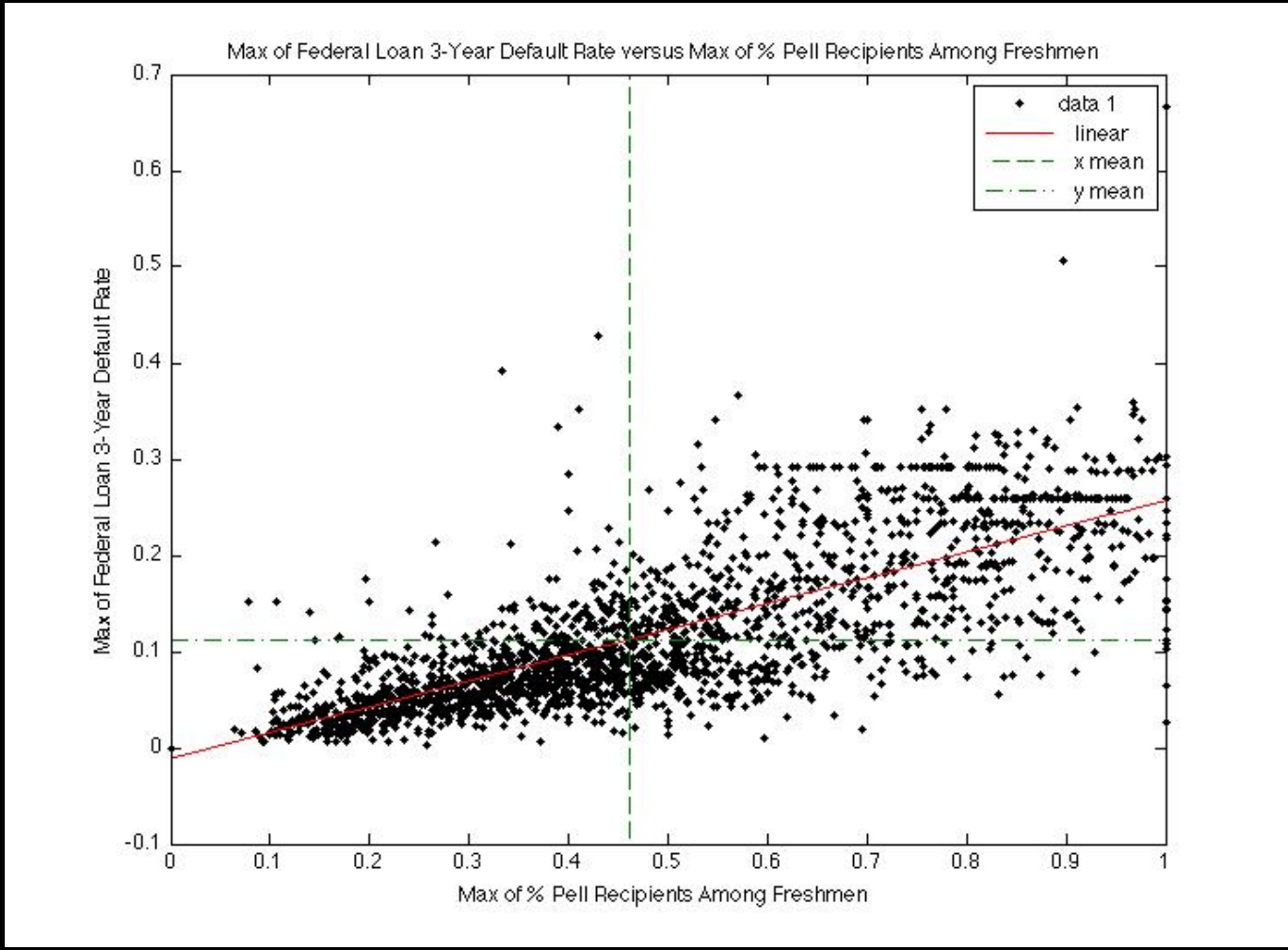
2012 4-year graduation rate



As the graduation rate increases the default rate decreases

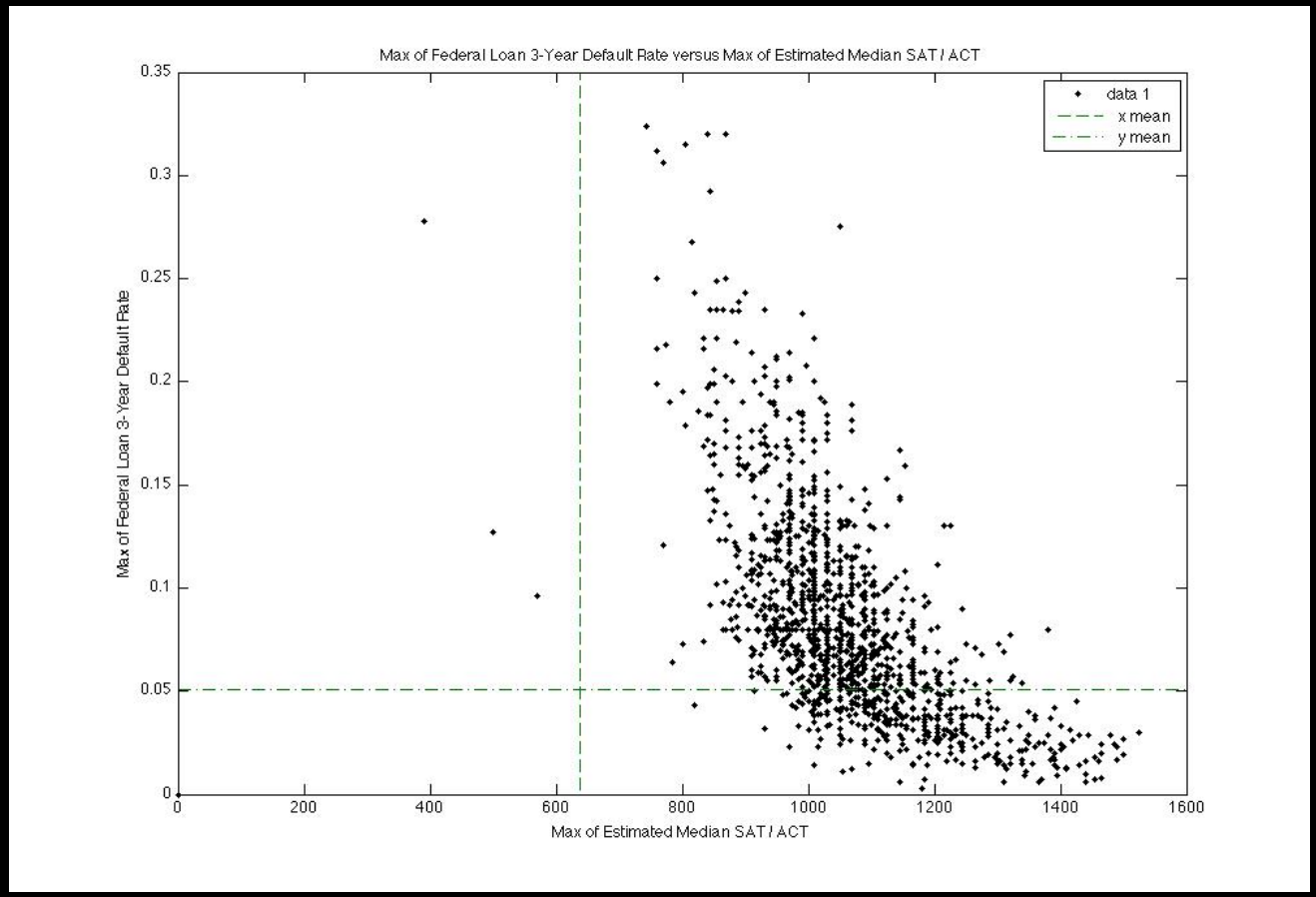
Xmean = 0.2854, ymean = 0.09048

Pell Grant



As the pell grant percentage increases the default rate increases. That's troubling.
xmean = 0.4617, ymean = 0.1127 true

Median SAT/ACT scores

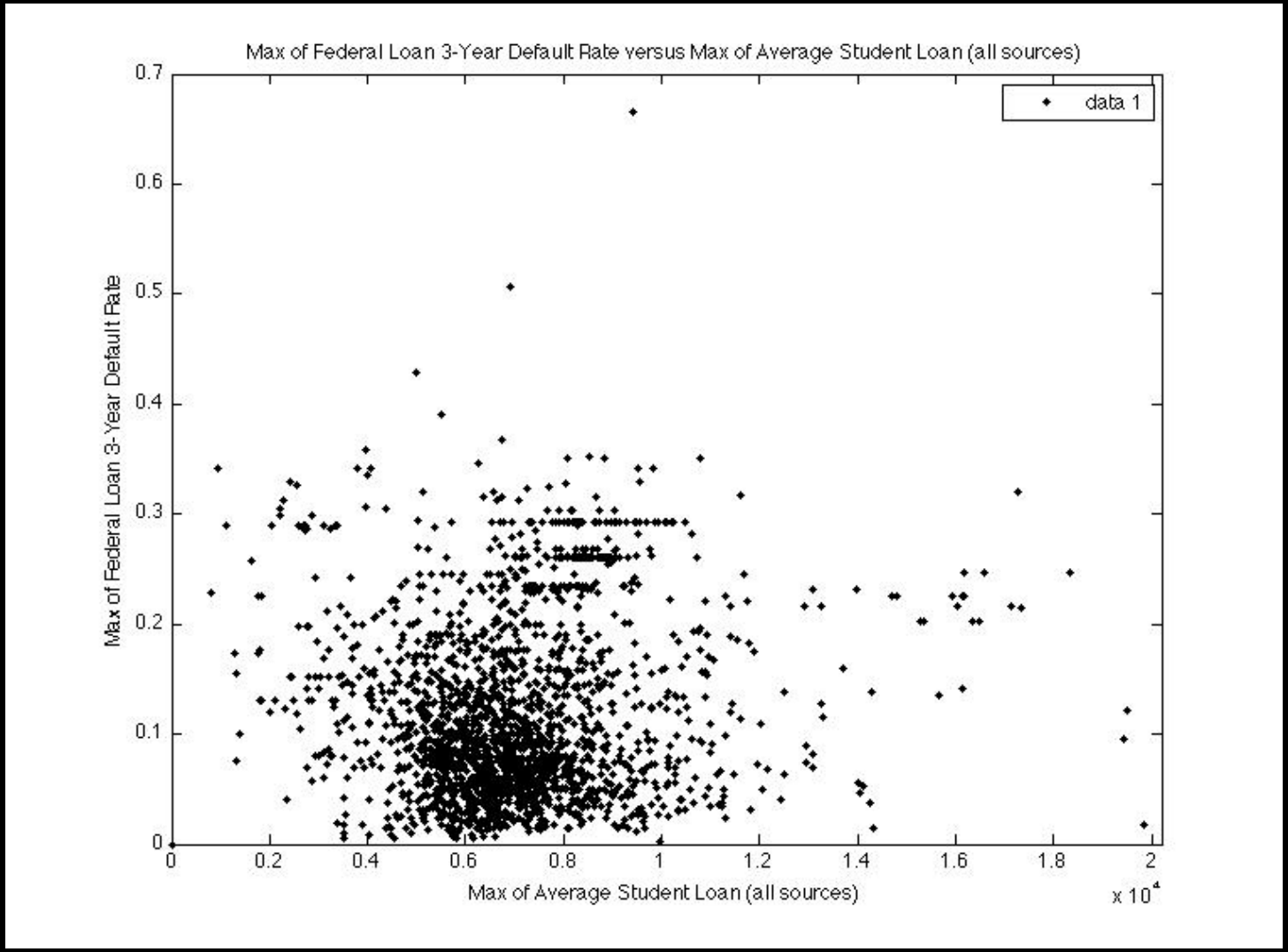


As SAT/ACT scores increase the default rate decreases

Mean is thrown off by the zeros and outliers. There also appears noise, the data is very uniform at certain x values.

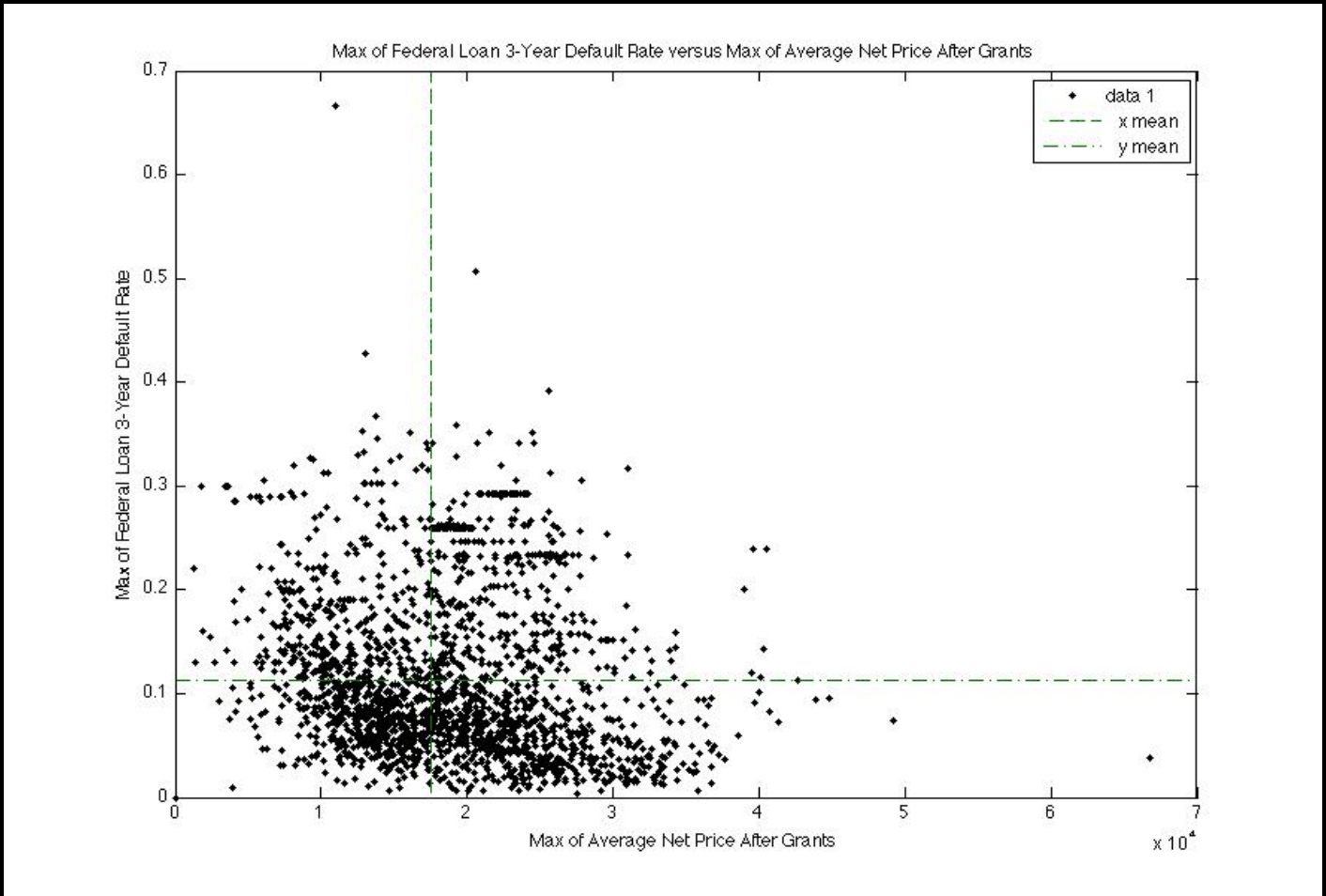
Xmean = 638.5, ymean = 0.05054

Average of student loans



There is a high clustering or is a random distribution of student loan debt and default rate, but a percentage of default rate occurs around ~$6,000.

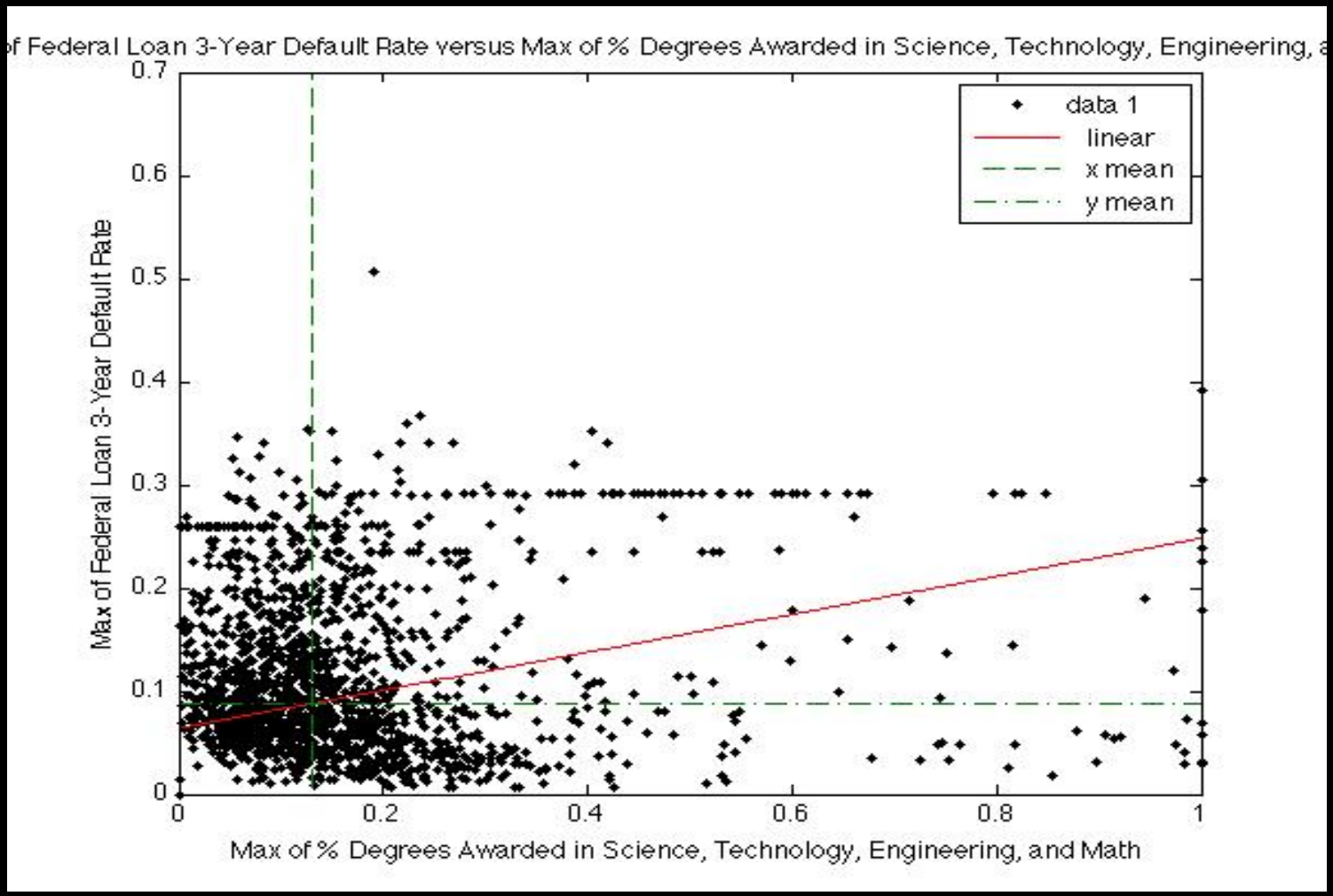
Average net price after graduation



This shows a more random distribution of default rate correlation to net price. Net price doesn't seem to indicate a clear trend in the data.

Xmean = 1.753e+04, ymean = 0.1125

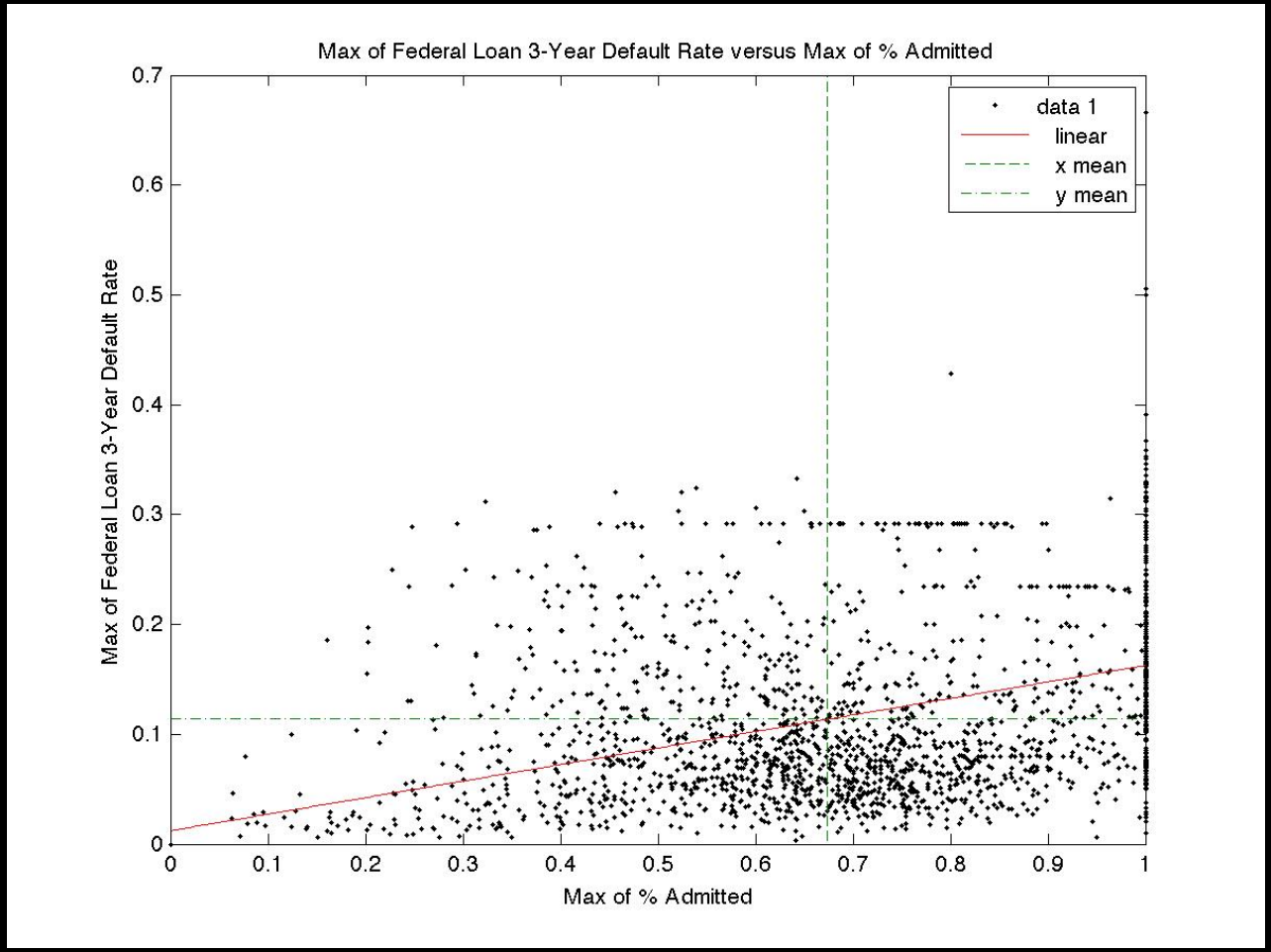
Stem Degree Awarded Rate



A non-uniform distribution centered on the mean value of the data.
There is a higher density of points for lower STEM award percentage. There is not a high correlation between the data.

Xmean = 0.1306, ymean = 0.08797

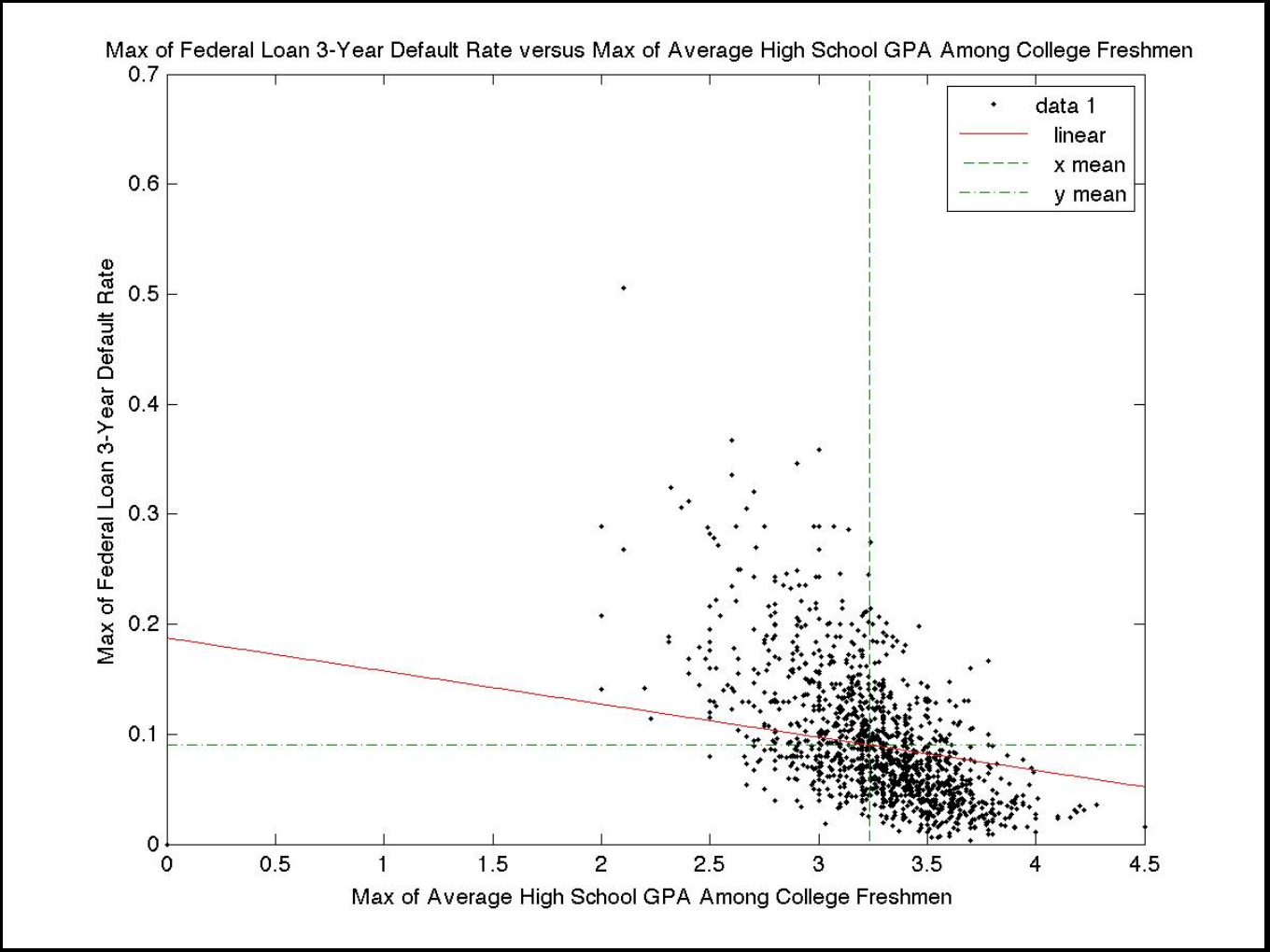
Percentage of admitted students



As admission increased, federal loan defaults increased. This would seem to have Underlying features that aren't being addressed, as the admissions increases so does that Number of students taking out loans.

Xmean = 0.6741, ymean = 0.1137

High School GPA among freshman



As GPA increases the default rate appears to decrease. A logical correlation appears to exist..

Xmean = 3.234,ymean = 0.09025

Final Analysis

- From what can be seen in the data there are correlations between 4 year grad rates, Pell grant percentages, median SAT/ACT scores, percentage of admissions, and high school GPA. Although, admissions is possibly more complicated than it would seem.
- The net price after graduation doesn't correlate with the default rate, as well the average student loan amount, and STEM degrees awarded.
- The non-correlating data would imply that more analysis is needed.
- The correlating datasets would imply that we should look at the interconnected or latent features of the data to make a stronger analysis of trends.
- The financial data as well as college major is less conclusive towards finding correlations in default rates, but the academic ability and testing seems to show a clearer correlation.
- My final analysis would be that as academic ability increases the 3-year student loan default rate appears to decrease.

Notes

- I'd like to have done this in python, but had issues with the .csv and .dat files and found Matlab's interface more convenient for that manipulation.
- I could look at the data more in depth using machine learning algorithms and do more complicated regression, which I'm more familiar with in python than Matlab and so decided to pursue a qualitative analysis for the time being.