Automatic Grading for Essays: Milestone#3
Stephen Turban, Noah Yonack

**Github:** https://github.com/noahyonack/AutomaticEssayGrading

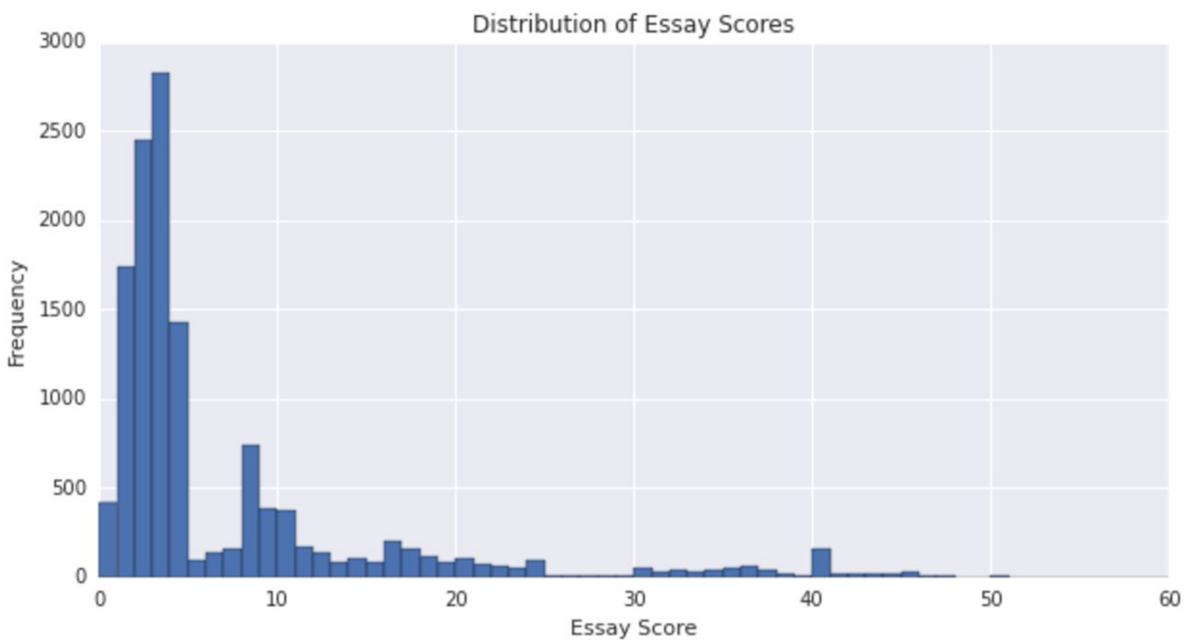
For milestone 3, we analyzed a dataset of roughy 13,000 essays, each of which came from one of 8 different essay prompts. Each essay was rated independently by two different judges, with the grading scale determined by the prompt (e.g. some prompts had scales from 1 to 5, others from 1 to 6, etc.).

Given our ultimate goal of teaching a model how to automatically score an ungraded essay, we looked at two different features of the dataset:

1. Length of each essay (and how that affects essay score)
2. Number of unique words per essay (and how that affects score)

Why are these metrics important? Finding a relationship between a given feature and the score of an essay can help us determine which features to pay attention to and which to ignore when training our model!

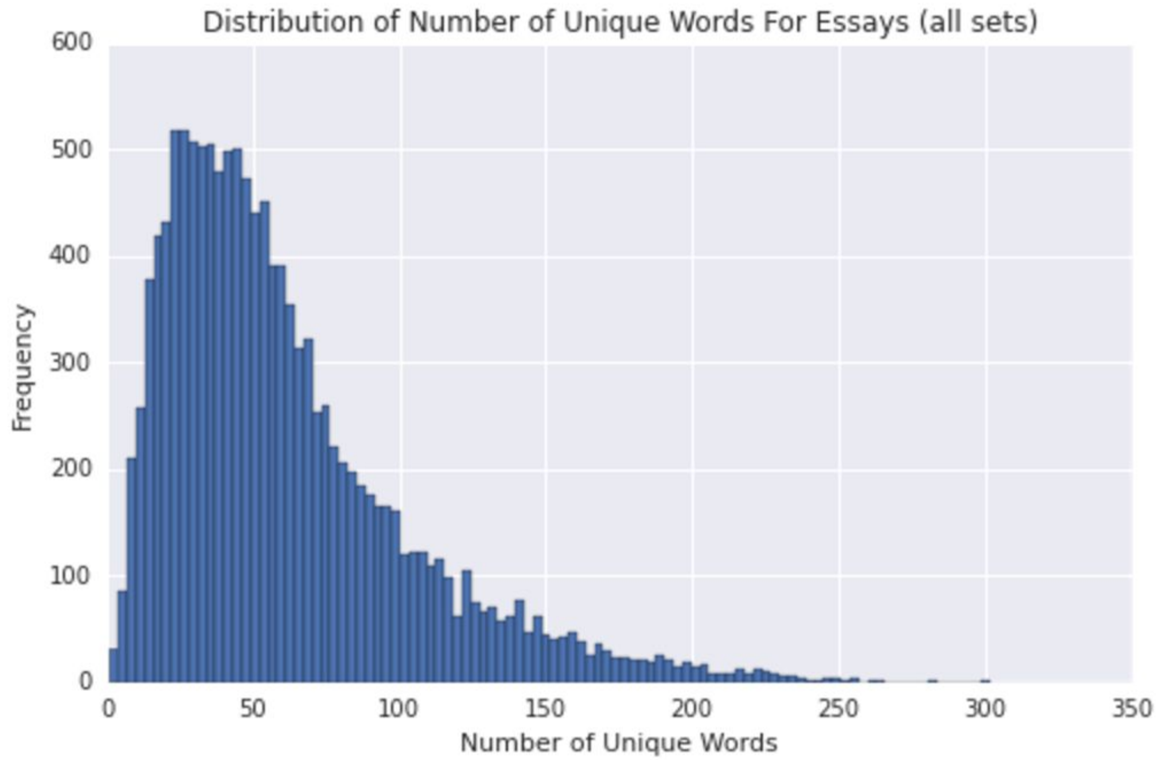
The first step we took to analyze our data was plot a histogram of essay scores. This is a helpful plot to look at because it can give us an idea as to the prior distribution of essay scores, which we can then use as either a prior in our model or to extract more features. The histogram is below:




The first thing to notice is that the range of essay scores is huge — from 0 to 60. This is because essays are scored based on scales that vary by set. This is an important thing to notice because it means that we'll need to use the essay set as a parameter in our model so that we reduce the variability in our predictions. When we partition essays by set and plot score histograms, we see that different sets have differing distributions (e.g. essay sets 7 and 8 have distributions that look somewhat normal, whereas essay set 6 has a negatively skewed distribution).

We then looked at scatterplots of essay length vs. essay score (see *Visuals*), which we partitioned by set. We saw that most essay sets have scores which are moderately correlated with essay length (e.g. essay set 8), whereas other essay sets have weaker, though still non-zero, correlations (e.g. essay set 3).

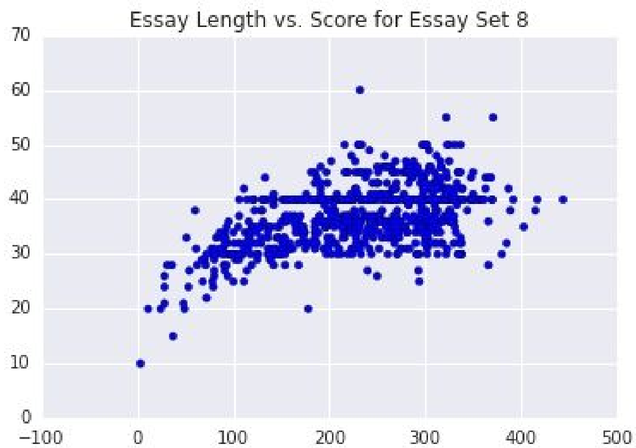
We also looked at relationships between essay score and the number of unique words per essay. Plotting a histogram of unique words frequency shows that most students use between 0 and ~100 unique words per essay, though there are certainly some students that use many more unique words. What does this histogram tell us? We might find during correlation analysis that these more eager students (i.e. those in the range from [100, 350]) do better on average than their more unoriginal counterparts.

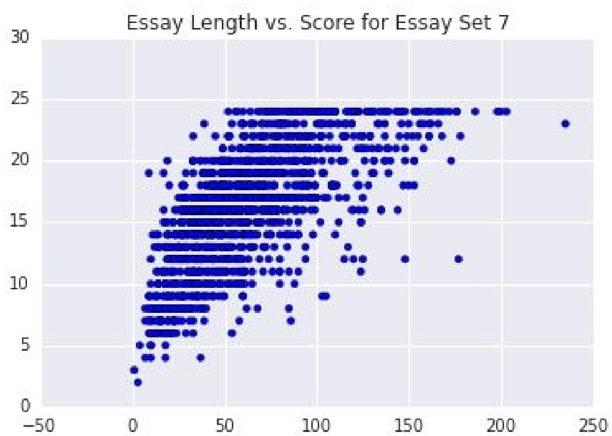


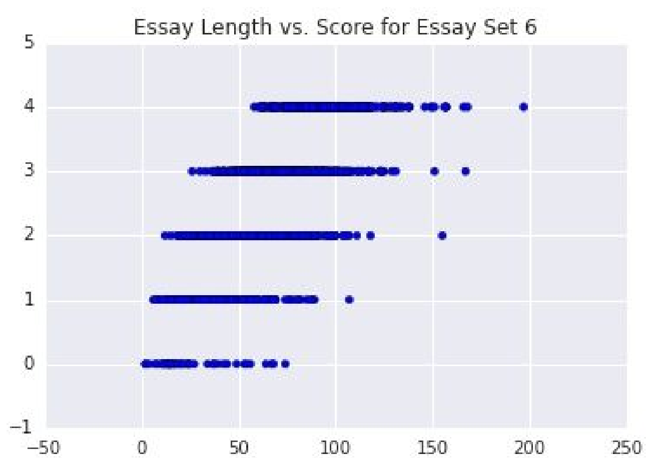

To explore this question, we plotted scatterplots of essay score vs. number of unique words (we choose not to include these plots in this write up because they are so similar to the essay-length scatterplots). We find that these scatterplots are not all too different from those that plot essay score against essay size (meaning that the number of unique words is indeed moderately correlated with essay score for a handful of essay sets in the same way that essay size is moderately correlated). Why are these scatterplots the same? Well, it seems as if the size of one's essay serves as a proxy for the number of unique words in that essay, which matches intuition; in other words, the longer the essay, the higher the number of unique words. Because of the high correlation between essay size and number of unique words, we probably only need to use one of the two measures, and this can be determined at a later stage when we need predictors that fit certain assumptions (i.e. if we decide to run a regression, we might choose whichever predictor is distributed normally, which is an assumption of linear regression).

## Visuals

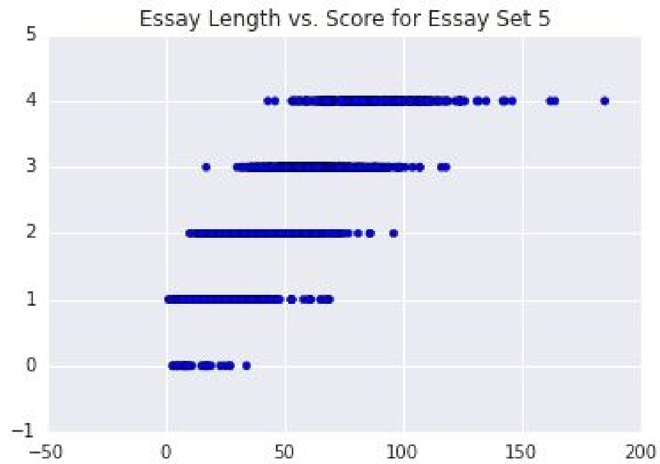
*Scatterplots of essay length vs. score, partitioned by essay set.* Note that these plots are essentially the same as those of number of unique words vs. score, partitioned by dataset.

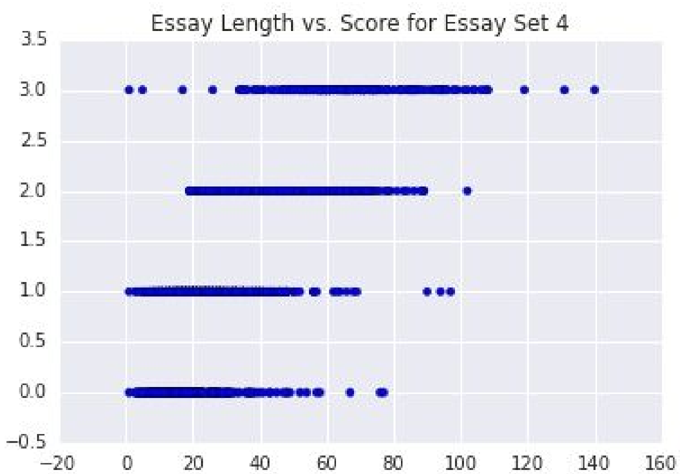


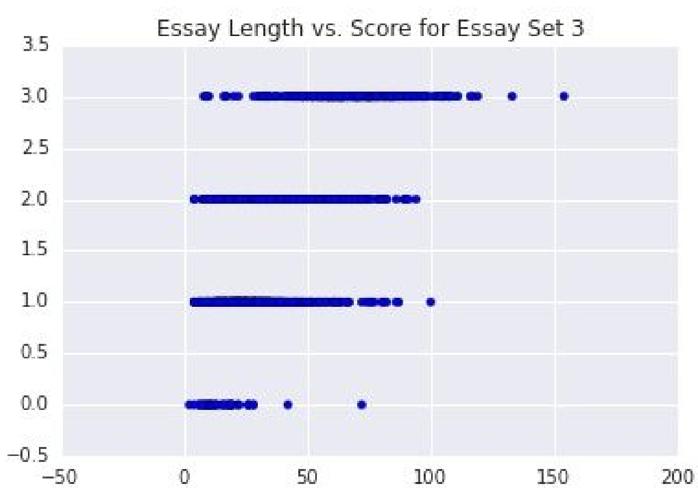



Automatic Grading for Essays: Milestone#3
Stephen Turban, Noah Yonack



Automatic Grading for Essays: Milestone#3
Stephen Turban, Noah Yonack