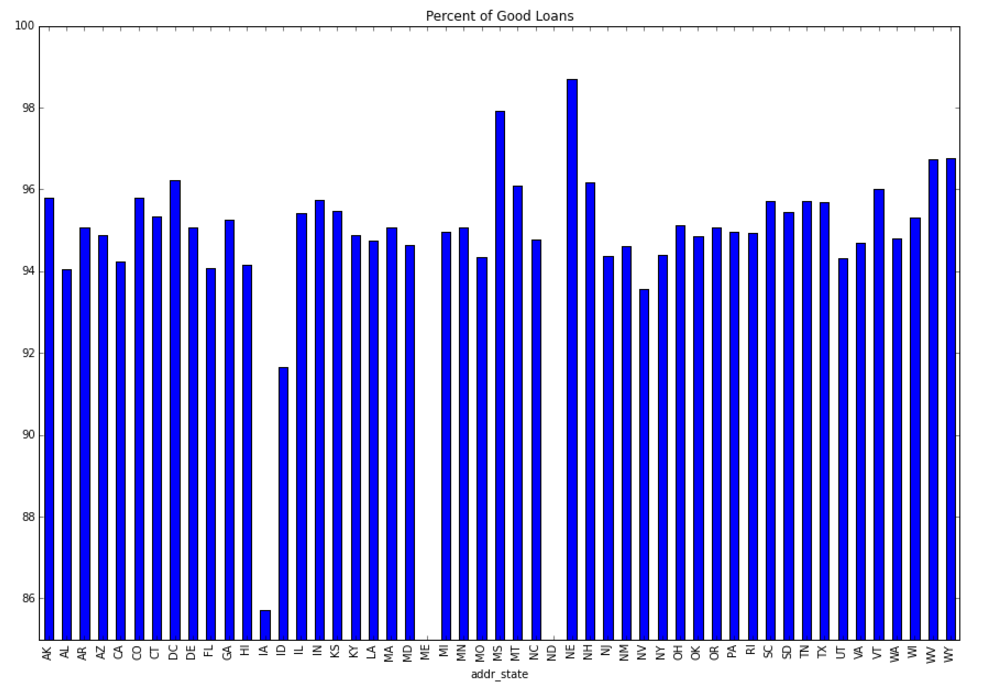
1. Having made these plots, what are some insights you get from them? Do you see any correlations? Is there a hypothesis you would like to investigate further? What other questions do they lead you to ask?
2. By now you’ve asked a bunch of questions, and found some neat insights. Is there an interesting narrative, a way of presenting the insights using text and plots from the above, that tells a compelling story? As you work out this story, what are some other trends/relationships you think will make it more complete?

I am exploring the Lending Club’s loan dataset with the objective of looking for any trends in loan defaults in order to create a machine learning model to predict loan defaults. As a first step, I looked at some basic statistics and found several avenues to explore.

The first, and somewhat obvious, dimension to consider was the average FICO score of the loan recipient. The average FICO score of recipients that are in default is 692 while the average score of those who are current is 697. This suggests that there is no significant difference in default rates based on the FICO score of applicants who receive loans based on the Lending Club’s current selection criteria.

Next, I analyzed the percent of loans that are in default against those that are current on a state by state basis. The results are summarize below:



Note that the results for IA, ID, ME, and ND are based on too few samples to be significant. Excluding those states, there appears to be some difference in default rates by state that may be worth exploring.

One possibility is that these differences in default rates between states is simply not statistically significant. The other is that the difference is taken into account when assigning ratings to the loans. These possibilities are avenues for further exploration.