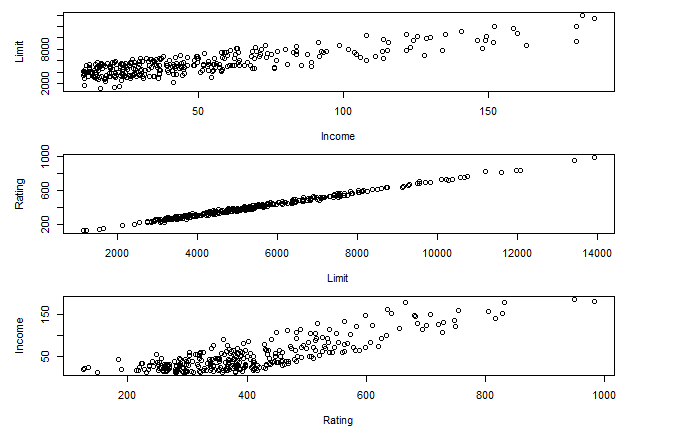
David Lowe

EDA #1

1. Credit card companies gain revenue as their customers fail to pay back the full amount of what they owe and accrue interest. This presents a challenge because debtors can sometimes build so much debt, that they can no longer pay the credit card company what they owe. Those responsible clients who pay their debt off before it every accrues interest are also of no use to credit card companies, because there is no money to be made. Credit card companies find the most value in those customers who have a moderate amount of credit card balance to be paid. This would imply that they are not yet at risk of bankruptcy, but they provide payments augmented by the interest rates of the credit card.

Understanding what factors or demographics influence the amount of credit card balance a potential customer will have can inform decisions about who the company should issue credit cards to. A credit card company will want to find which segment of the population will continue to pay their balance, but will also accrue interest. This can help in maximizing revenue, and tailoring credit cards for specific demographics. This will by our purpose in analyzing the given data.

1. The first noticeable trait of the data is that there are several different variables collected with the credit card balance. This means that there may be several different factors that affect the credit card balance of a customer. In plotting the data there is a linear relationship between many of the explanatory variables and credit card balance, such as the credit limit and credit rating. There are also various categorical variables, such as gender, race, and marital status. These are more interpretable as factors. Age is represented by year counts, and shouldn’t be treated as a continuous variable. A couple of things to note is that consumer credit rating and credit limit are very heavily correlated, and both are correlated with income (see graphic to the right). Including all three of these variables may affect our model unintentionally.
2. I recommend we use multiple linear regression (MLR) to build this model, though we will need to account for a few of the nuances of the data stated above. MLR would give us the ability to see how each of the variables affect the credit card balance of a consumer. Once we have found estimates of the effects of each variable, we will have a model that we can use to predict the balance of a consumer, given their other demographics. This could help us find a target range of consumers that amount a moderate credit card balance, and will produce revenue for the company.
3. Something I don’t know how to deal with is treating age as a count variable, unless I treat it as a factor. Since it is not continuous, we need to treat it differently than the rest of the variables, but it has too many values to be used as a factor.