

# EDA

*Sri Santhosh Hari*

We first visualize how the variable of importance, **Bankruptcy Rate**, has changed over the years.

It can be observed from Figure 1 that the bankruptcy rates have generally increased since 1987 with significant increase in later half of 2000's, which can be attributed to the global financial crisis/recession.

Looking at the monthly changes in Bankruptcy Rates in Figure 2, one can infer that March, the last month of financial year, tends to have higher bankruptcy rates compared to other months.

We then visualize the changes observed in external variables (House Price Index, Population, Unemployment Rate) to asses if multivariate approach is suitable as opposed to univariate modeling. Figure 3 shows the trend in each of the external variable (scaled) along with Bankruptcy rate. It can be seen that HPI has a clear relationship with Bankruptcy Rate, as the trends observed in both the variables appear to be similar, with some time-difference. Population appears to linearly increase over the years but without valleys, may not be an ideal variable to include in the model. We will evaluate models with population included and excluded.

Figure 4 helps us visualize the the relationship of HPI and 18-month-lagged verison of HPI with Bankruptcy Rate.

Figure 5 visually captures the relationship between the Bankruptcy Rate and other external variables available.

Table 1: Correlation (in %) of Bankruptcy Rate with External Variables

	Bankruptcy Rate
Unemployment Rate	-31.69
Population	89.84
House Price Index	68.97
Lagged House Price Index	75.08

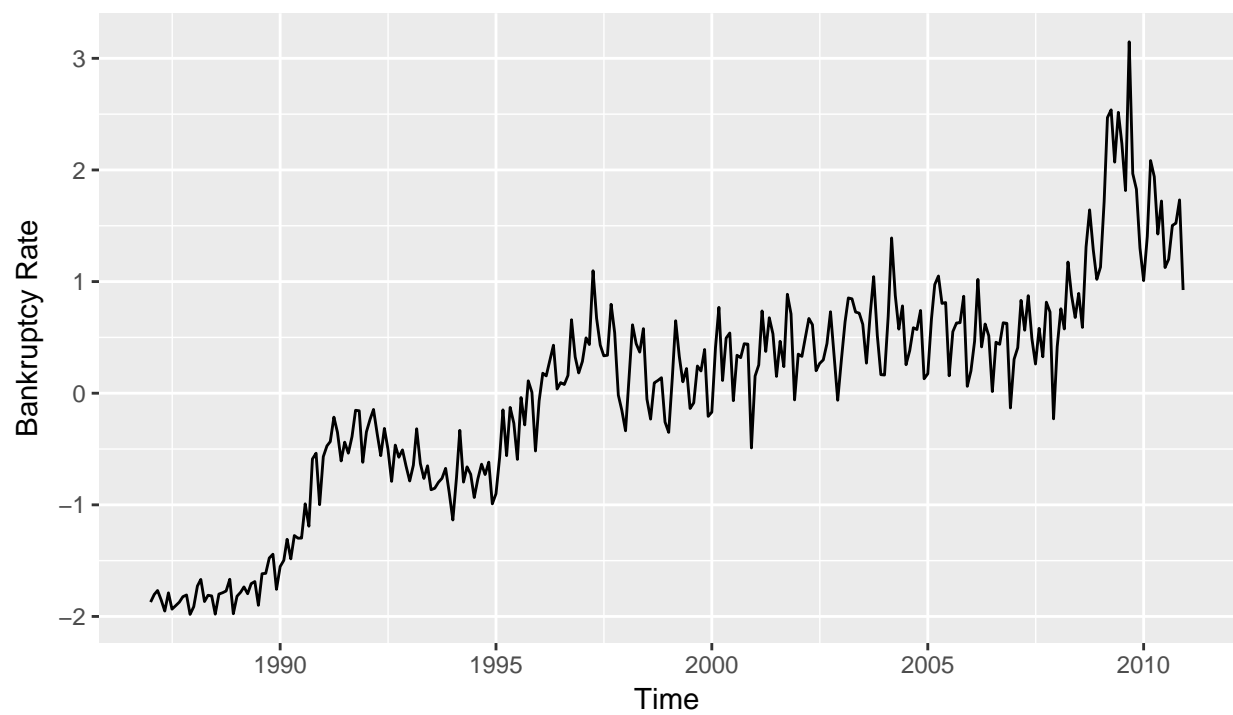


Figure 1: Monthly Canadian Bankruptcy Rates (1987 - 2010)

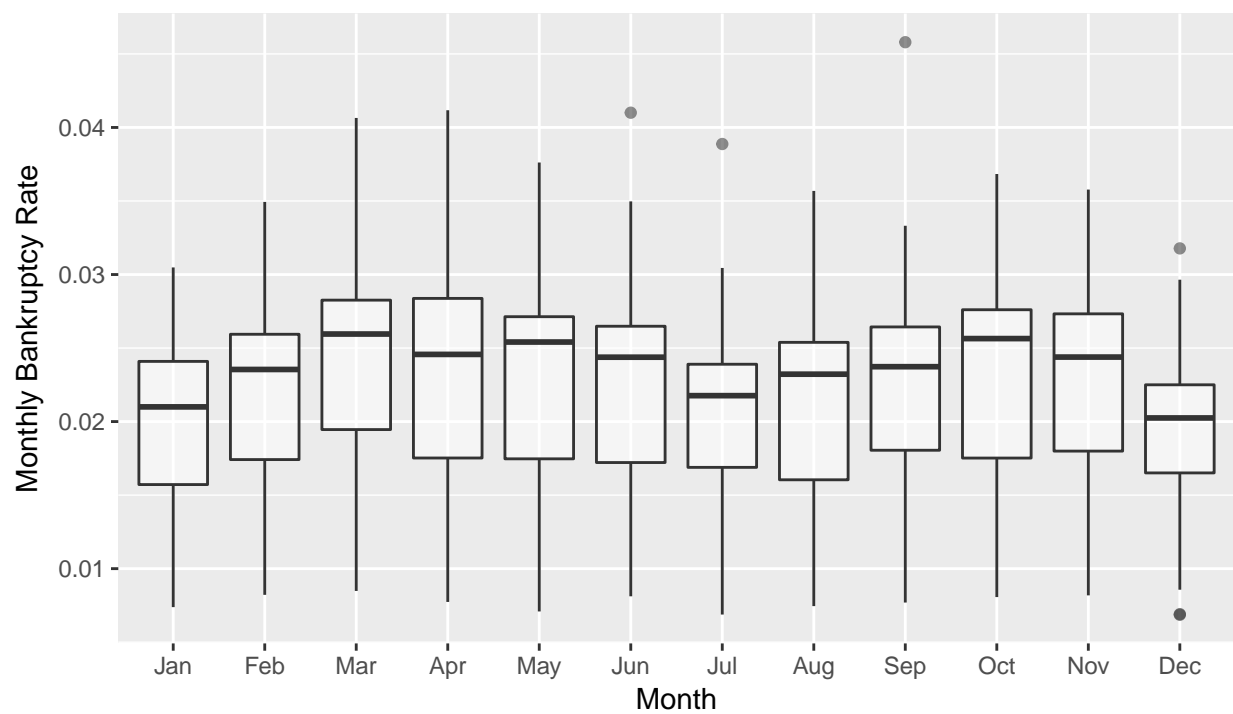


Figure 2: Bankruptcy Rates across Months

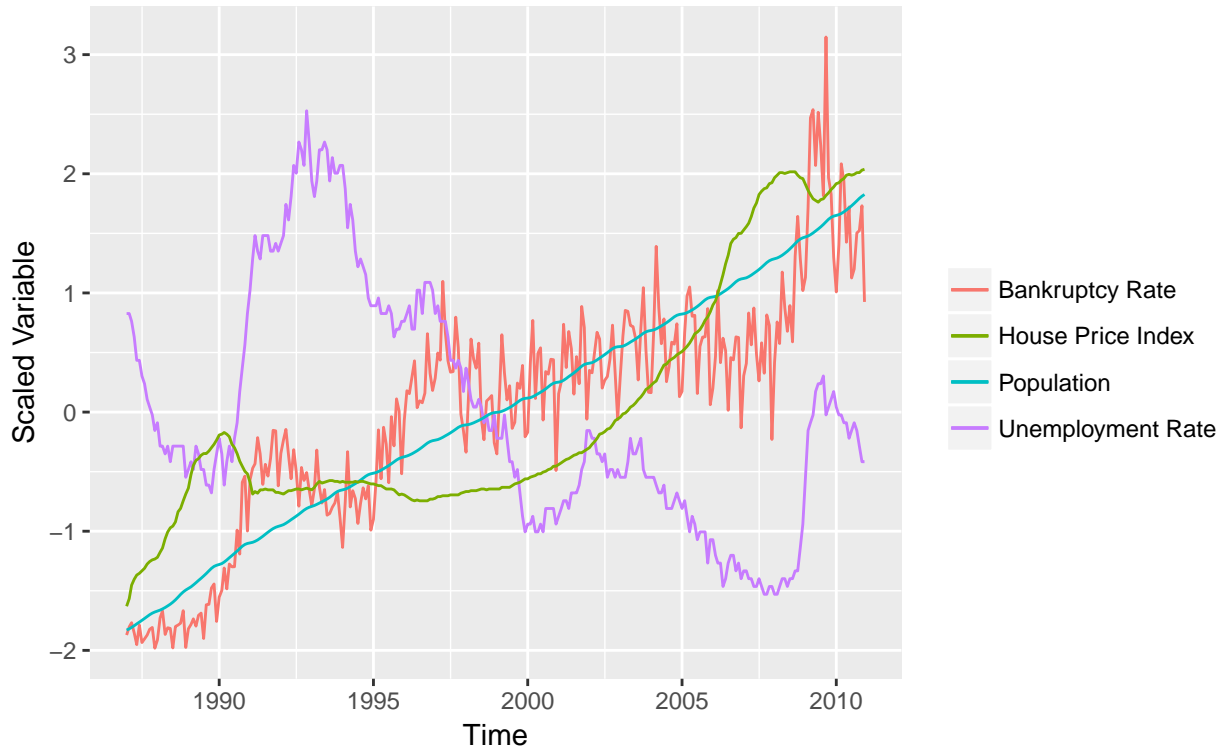


Figure 3: Trends observed in variables across years

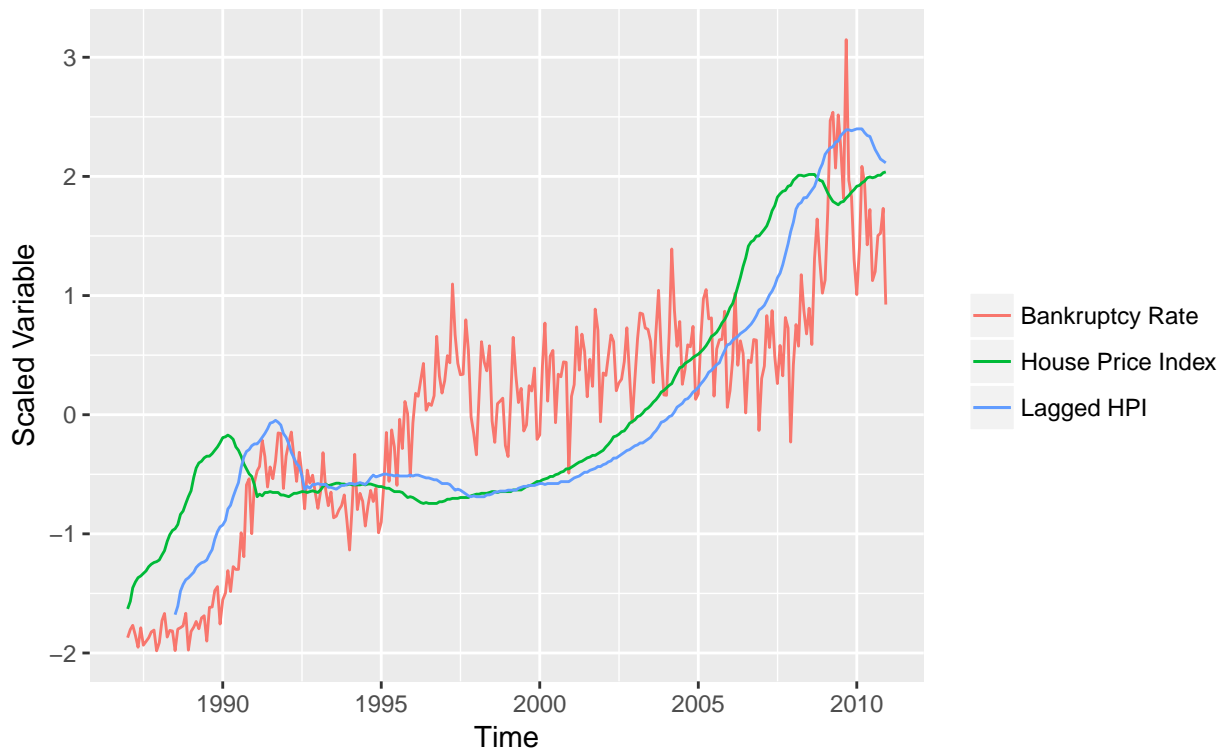


Figure 4: Bankruptcy Rates and House Price Index changes across years

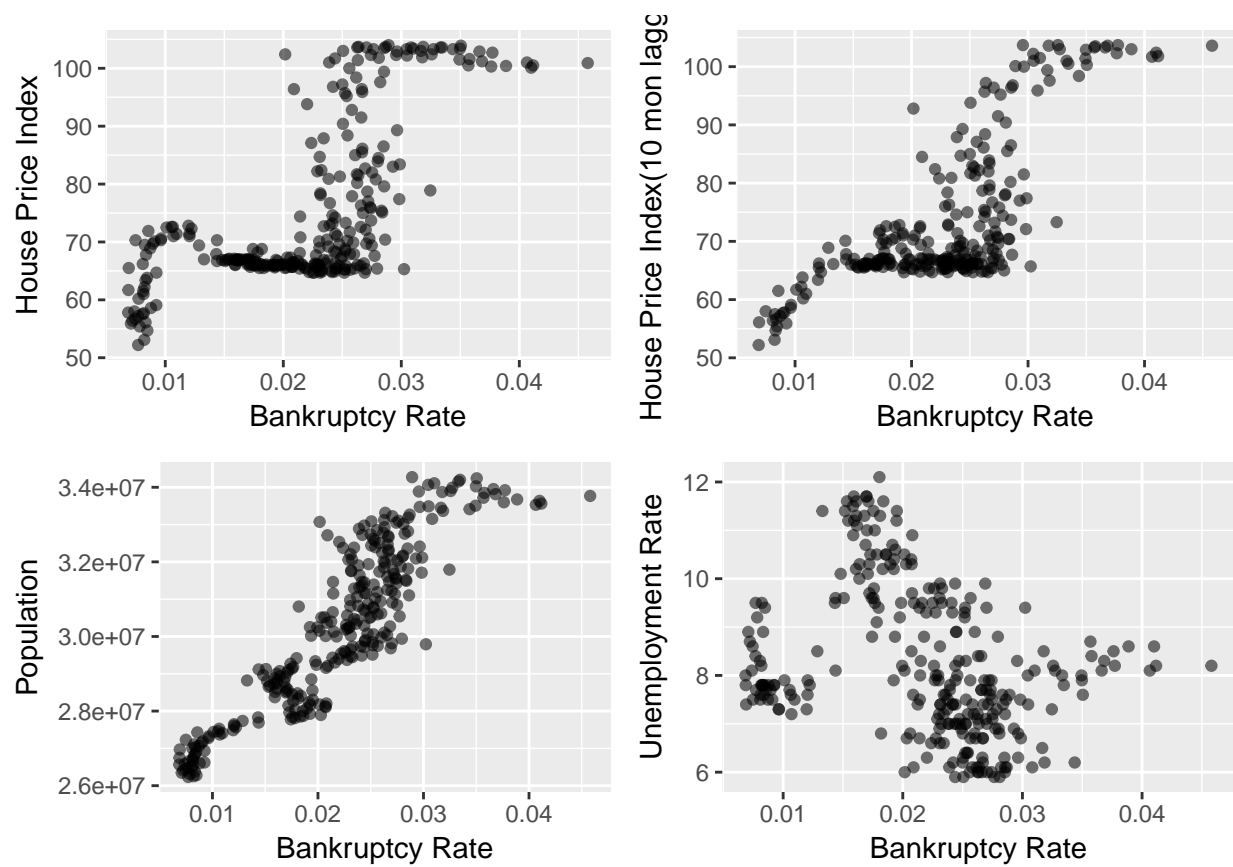


Figure 5: Bankruptcy vs External Variables