Significance of Macroeconomic Factors to the Level of Nonperforming Loans in Kenya

Cheryl Akinyi

14 May 2022

Significance of Macroeconomic Factors to the Level of on Perfoming Loans in Kenya

A loan is considered non performing if it has not been serviced for 90 days. In Kenyan banks, non performing loans have been a persistent problem. This is because; the main goal of every banking institution is to operate profitably in order to maintain stability and sustainable growth. However the existence of high levels of non performing loans negatively affects the level of private investment, impair a bank's ability to settle its liabilities when they fall due and constrain the scope of the bank credit to borrowers. External and internal economic environments are viewed as critical drivers for non performing loans. In this regard, the main goal of this project is to investigate the link between non performing loans and macroeconomic factors, and establish the extent to which these factors affect the level of non performing loans in Kenya.

R Packages

For this analysis using R programming, we will make use of the following R packages:

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("broom")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
install.packages("ggpubr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.1'
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.6     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(dplyr)
library(tidyr)
library(readr)
library(broom)
library(ggpubr)
```

The Data

We apply the principles on the Non Performing Loans data set that is available on the World Bank and the Central Bank of Kenya (CBK) data websites. The given data are around non performing loans and macroeconomic factors in Kenya between 2006 and 2015. Measured variables include Gross Domestic Product (GDP), Exchange Rate and Interest Rate.

We combined the CSV files into one Excel file to level up our productivity and efficiency in data transformation and management, plus remove any missing values in Excel.

The following commands reading our data and display the first 6 observations.

```
non_performing_loans <- read.csv("/cloud/project/Non Performing Loans Dataset.csv")
head(non_performing_loans)
```

```
##   Year   NPL         GDP Exchange_Rate Interest_Rate
## 1 2006 19.35 25825524821         72.10         13.64
## 2 2007 10.23 31958195182         67.32         13.33
## 3 2008  9.01 35895153328         69.19         14.02
## 4 2009  8.00 42347217913         77.35         14.80
## 5 2010  6.29 45405587557         79.23         14.36
## 6 2011  4.43 46869457318         88.81         15.05
```

The Problem

With this data, it is possible to answer many interesting questions. Examples include:

- **Do macroeconomic factors have significance influence on the level of non performing loans?**
- Does gross domestic product have an explanatory power as a determinant in the level of non performing loans?
- Is effective exchange rate a relevant explanatory variable for the level of non performing loans?
- Is there significance in the relationship between interest rates and the level of non performing loans?

Additionally to these concrete questions, the possibilities for exploration, sandbox style data analysis are nearly limitless. Here, we will focus on the first bolded question.

Data Cleaning

We will check out the amount of missing values in the data or any column names that need to be adjusted. We rewrite the column names in GDP and change the GDP values into a percentage to match the rest of the variables.

Because the variables are quantitative, running the code produces a numeric summary of the data for the independent variables (GDP, exchange rate and interest rate) and the dependent variable (non performing loans):

```
NPL_loans <- non_performing_loans %>%
  select(Year, NPL, GDP, Exchange_Rate, Interest_Rate)%>%
  rename(GDPUSD = GDP)%>%
  mutate(GDPUSD = GDPUSD / 1E9)
as_tibble(NPL_loans)
```

```
## # A tibble: 15 x 5
##     Year   NPL GDPUSD Exchange_Rate Interest_Rate
##    <int> <dbl>  <dbl>         <dbl>         <dbl>
##  1  2006 19.4    25.8          72.1          13.6
##  2  2007 10.2    32.0          67.3          13.3
##  3  2008  9.01   35.9          69.2          14.0
##  4  2009  8      42.3          77.4          14.8
##  5  2010  6.29   45.4          79.2          14.4
##  6  2011  4.43   46.9          88.8          15.0
##  7  2012  4.59   56.4          84.5          19.6
##  8  2013  5.05   61.7          86.1          17.3
##  9  2014  5.46   68.3          87.9          16.5
## 10  2015  5.99   70.1          98.2          16.2
## 11  2016 11.7    74.8         102.           16.6
## 12  2017  9.95   82.0         103.           13.7
## 13  2018 11.7    92.2         101.           13.1
## 14  2019 12.0   101.          102.           12.4
## 15  2020 14.1   101.          106.           12
```

```
summary(NPL_loans)
```

```
##       Year           NPL             GDPUSD       Exchange_Rate
##  Min.   :2006   Min.   : 4.430   Min.   : 25.83   Min.   : 67.32
##  1st Qu.:2010   1st Qu.: 5.725   1st Qu.: 43.88   1st Qu.: 78.29
##  Median :2013   Median : 9.010   Median : 61.67   Median : 87.92
##  Mean   :2013   Mean   : 9.190   Mean   : 62.36   Mean   : 88.36
##  3rd Qu.:2016   3rd Qu.:11.675   3rd Qu.: 78.43   3rd Qu.:101.40
##  Max.   :2020   Max.   :19.350   Max.   :101.01   Max.   :106.45
##  Interest_Rate
##  Min.   :12.00
##  1st Qu.:13.48
##  Median :14.36
##  Mean   :14.84
##  3rd Qu.:16.34
##  Max.   :19.65
```

We see that there are no missing values but GDP is in millions while the other variables are in percentages. We use the pipe operator %>% to rename the GDP column and change its values into a percentage. Because the variables are quantitative, running the code summary() provides a summary of the data for the independent variables (GDP, Exchange Rate and Interest Rate) and the dependent variable (NPL).

Explorative Analyses

Our main variables of interest are:

- *Year*, which conveniently is already in the column Year.
- *NPL*. This will indicate the Bank non performing loans to total gross loans (%)
- *GDP*, the monetary value of a;ll finished goods and services produced within a country's borders in a given period of time, (current US$).

- *Exchange Rate* between currencies, the rate at which one currency will be exchanged for another. It is also regarded as the value of one country's currency in terns of another currency.
- *Interest Rate*, the amount paid by an individual to a financial institution as the cost of borrowing. The money lender takes a risk that the borrower may not pay back the loan, thus interest provides a certain compensation for bearing risk coupled with the risk of default and inflation.

Let's have an explanatory look at all our variables of interest to answer our question **Does our data meet the main assumptions for linear regression?**

1. Independence of observation i.e. no correlation

We use the cor() function to test the relationship between the independent variables and make sure they are not too highly correlated.

```
cor(NPL_loans$GDPUSD, NPL_loans$Exchange_Rate)
```

```
## [1] 0.9368294
```

The output is 0.9368294. The correlation between GDP and exchange rate is high (a correlation of r = 0.9 suggests a strong, positive association between two variables).So we can not include both parameters in our model.

```
cor(NPL_loans$GDPUSD, NPL_loans$Interest_Rate)
```

```
## [1] -0.2101594
```

When we run this code, the output is -0.2101594. The correlation between GDP and interest rate is small (a correlation of r = -0.2 suggest a weak, negative association). So we can include one parameter in our model because one variabe (GDP) has a strong positive correlation with another variable (exchange rate).

```
cor(NPL_loans$Exchange_Rate, NPL_loans$Interest_Rate)
```

```
## [1] -0.1082939
```

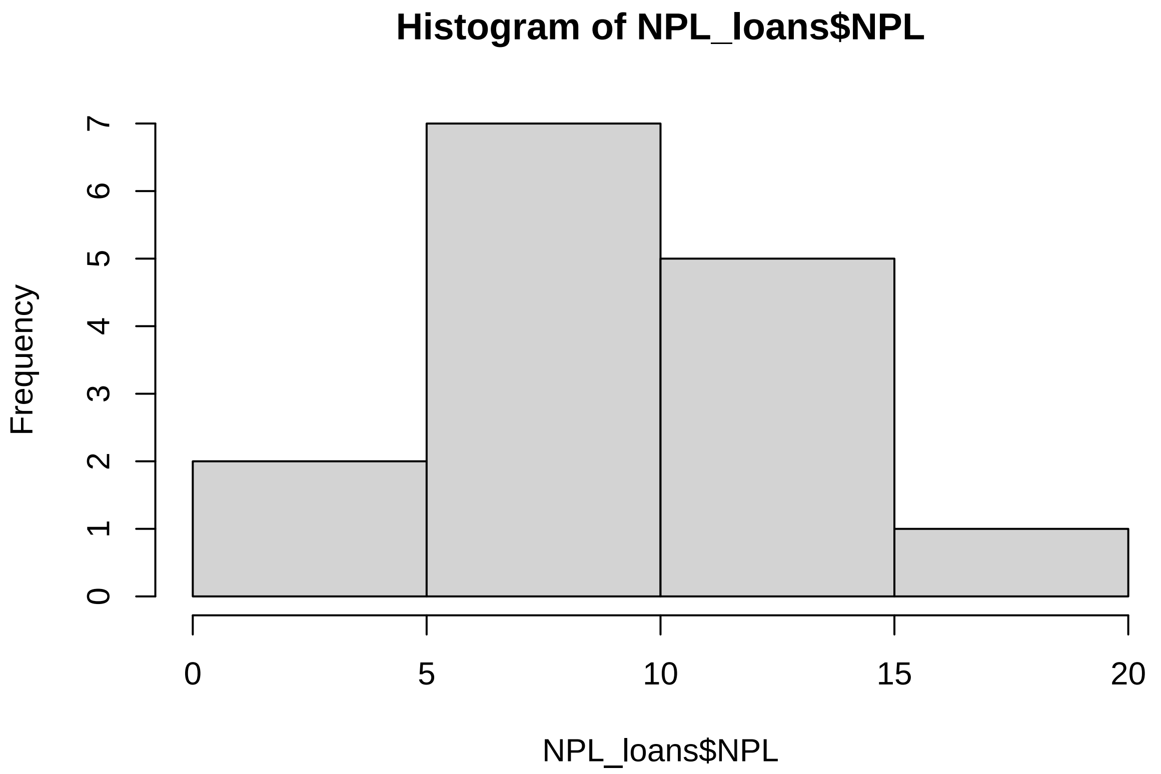
When we run this code, the output is -0.1082939. The correlation between exchange rate and interest rate is small (a correlation of r = -0.1 suggest a weak, negative association). So we can include both parameters in our model.

It is important to note that there may be a non-linear association between two continuous variables, but computation of a correlation coefficient does not detect this. Therefore, it is always important to evaluate the data carefully before computing a correlation coefficient. Graphical displays are particularly useful to explore associations between variables.

2. Normality

We use the hist() function to test whether the dependent variable follows a normal distribution.

```
hist(NPL_loans$NPL)
```

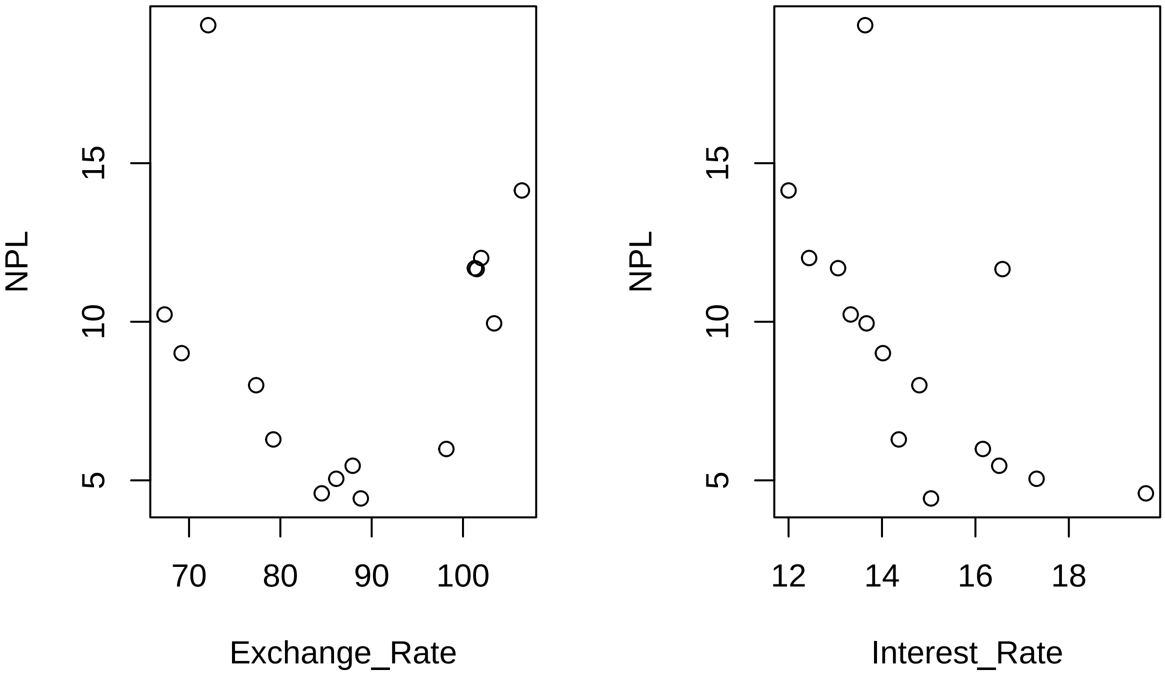


The distribution of observations is bell-shaped, so we can proceed with the linear regression.

3. Linearity

We can check linearity using scatter plots: one for GDP and NPL, one for Exchange Rate and NPL and one for Interest Rate and NPL

```
par(mfrow = c(1,2))
plot(NPL ~ Exchange_Rate, data = NPL_loans)
plot(NPL ~ Interest_Rate, data = NPL_loans)
```



4. Homoscedasticity

We will check this after creating the model.

Perform the linear regression analysis

Now that we've determined the data meets the assumptions, we can perform a linear regressions analysis to evaluate the relationship between the independent and dependent variables. Let's see if there's a linear relationship between GDP, Exchange Rate, Interest Rate and non performing loans. To test the relationship, we first fit a linear model with non performing loans as the dependent variable and GDP, exchange rate and interest rate as the independent variables.

```
NPL_data <- lm(NPL ~ Exchange_Rate + Interest_Rate, data = NPL_loans)
summary(NPL_data)
```

```
##
## Call:
## lm(formula = NPL ~ Exchange_Rate + Interest_Rate, data = NPL_loans)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.4852 -1.3613 -0.9148  0.7158  8.6067
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)   28.4611741  9.6093101   2.962   0.0119 *
## Exchange_Rate  0.0003732  0.0698843   0.005   0.9958
## Interest_Rate -1.3009354  0.4495127  -2.894   0.0135 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.456 on 12 degrees of freedom
## Multiple R-squared:  0.414,  Adjusted R-squared:  0.3164
## F-statistic: 4.239 on 2 and 12 DF,  p-value: 0.04048
```

Our data fits the model:

NPL = 28.4611747 + 0.0003732(Exchange_Rate) - 1.3009354(Interest_Rate)

The coefficients of the independent variable shows the proportion of each factor that adds up to the level of non performing loans in that period. The initial figure is the y-intercept of the model, the value of non performing loans if all independent variables are zero.

The estimated effect of exchange rate on non performing loans is 0.0003732 and interest rate is -1.3009354.

This means that for every 1% increase in interest rate, there is a correlated 1.3% decrease respectively in the level of non performing loans. Meanwhile, for every 1% increase in exchange rate, there is a 0.0003% increase in the level of non performing loans.

The standard errors and the t-statistics for these regression coefficients are small. The p-values reflect these small errors and t-statistics. The R-squared is 0.414 which means that 41.4% of the total variation in non performing loans in Kenya can be accounted for by the independent variables; exchange rate and interest rate.

The p-values measure the quality of the fit of the model to the data, the smaller the p-value the stronger the linear relationship between the variables.

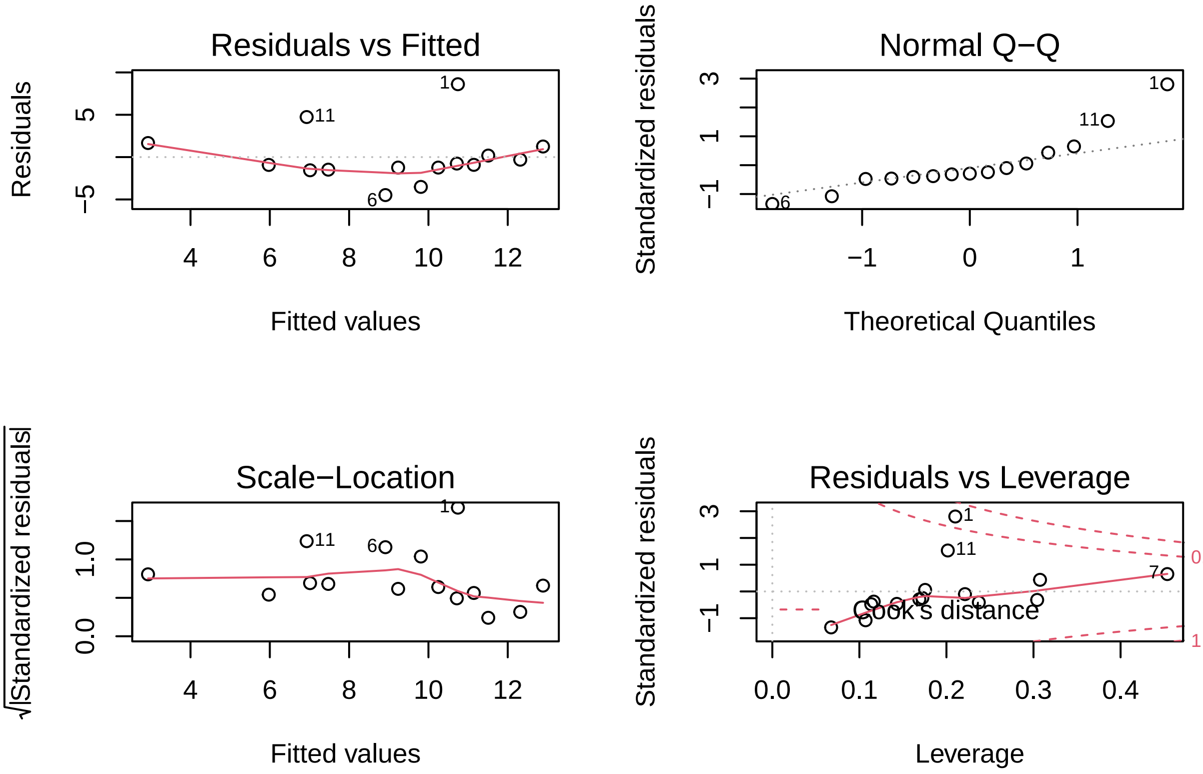
From our data, only interest rate has a p-value less than 0.05 which is our significant level of confidence. This means that it is the only variable with significant predictive capacity in our model.Based on our findings, the model is:

NPL = 28.4611747 - 1.3009354(Interest_Rate)

Check for homoscedasticity

Before proceeding with data visualization, we should make sure that our model fit the homoscedasticity assumption of the linear model. We should check if our model is actually a good fit for the data, and that we don't have large variation in the model error, by running this code:

```
par(mfrow=c(2,2))
plot(NPL_data)
```



```
par(mfrow=c(1,1))
```

Residuals vs Fitted plot shows if residuals have non-linear patterns. Notice how equally spreads residuals are around a horizontal line without distinct patterns, that is a good indication we do not have non linear relationships.The most important thing to look for is that the red lines representing the mean of the residuals are all basically horizontal and centered around zero. This means there are no outliers or biases in the data that would make a linear regression invalid.

In the Normal Q-Q plot in the top right, we can see the residuals are lined well on the straight dashed line.Of course they wouldn't be a perfect smooth line as observation numbered 1 looks a little off.

The Scale location plot shows if residuals are spread equally along the ranges of predictors. This is how we check the assumption of equal variance (homoscedasticity). The residuals appear randomly spread.

Residuals vs Leverage helps us find influential cases if any. Unlike other plots, this time patters are not relevant. We watch out for outlying values at the upper right corner or at the lower right corner. Those spots are the places where cases can be influential against a regression line. Look for cases outside of a dashed line, Cook's distance. When cases are outside of the Cook's distance (meaning they have high Cook's distance scores), the cases are influential to the regression results. The regression results will be altered if we exclude those cases.

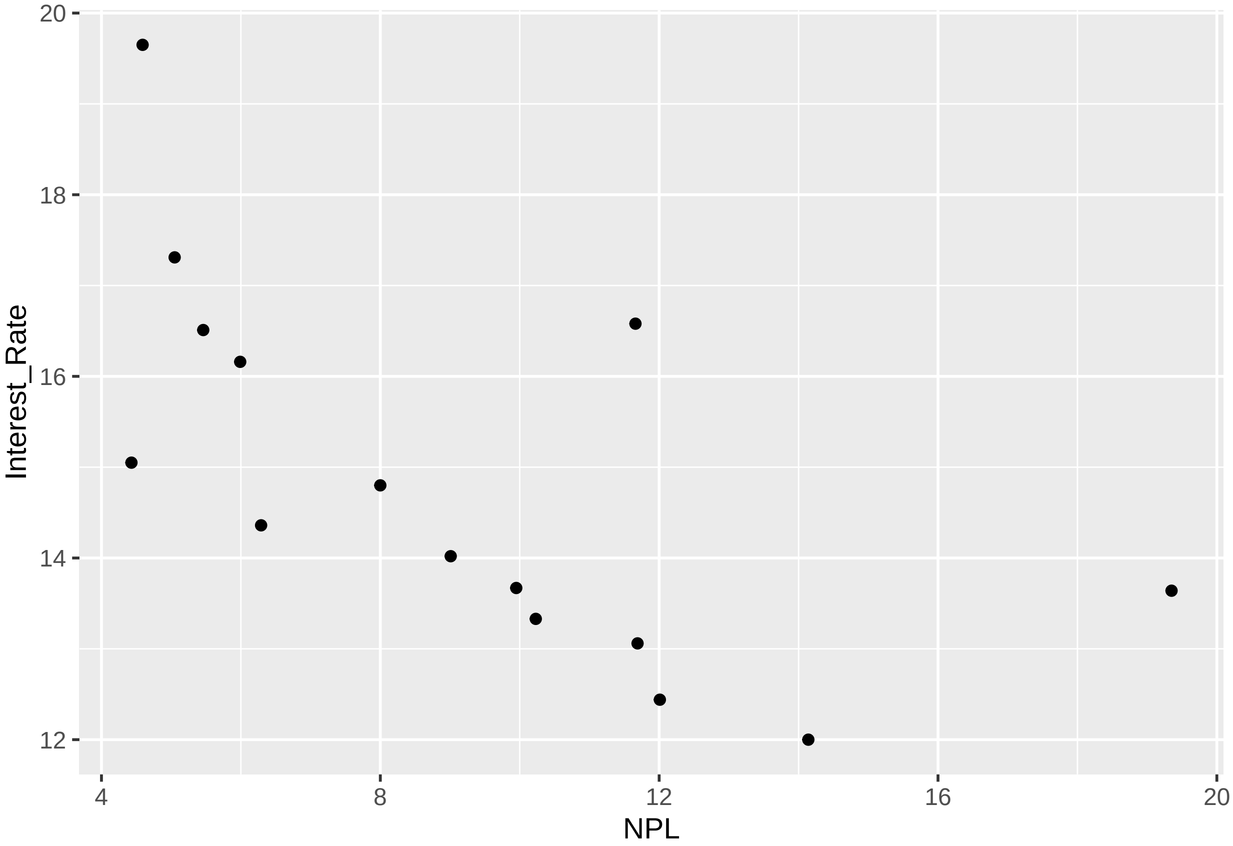
Based on these residuals, we can say that our model meets the assumption of homoscedasticity.

Visualize the results with a graph

Next, we can plot the data and the regression line from our linear regression model.

Plot the data points on a graph

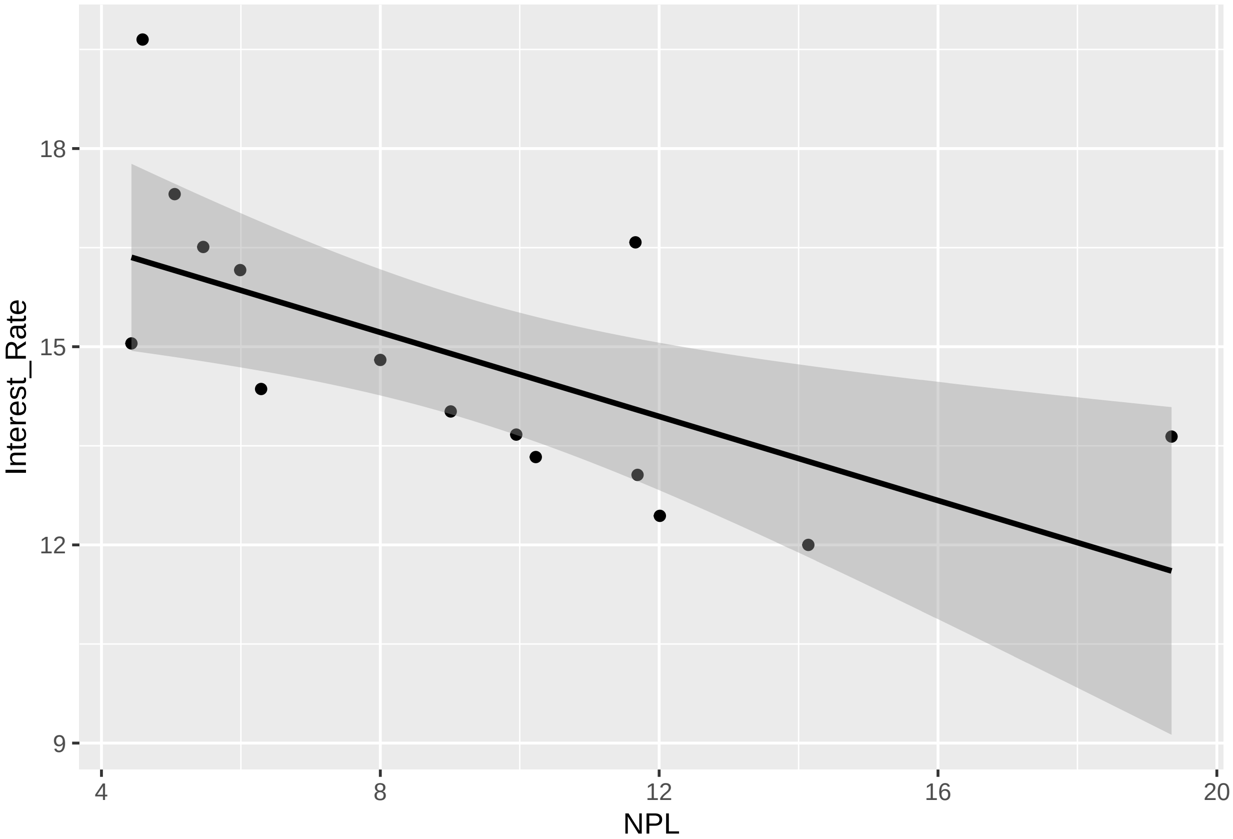
```
NPL_graph<-ggplot(NPL_data, aes(x=NPL, y=Interest_Rate)) + geom_point()
NPL_graph
```



Add the linear regression line to the plotted data

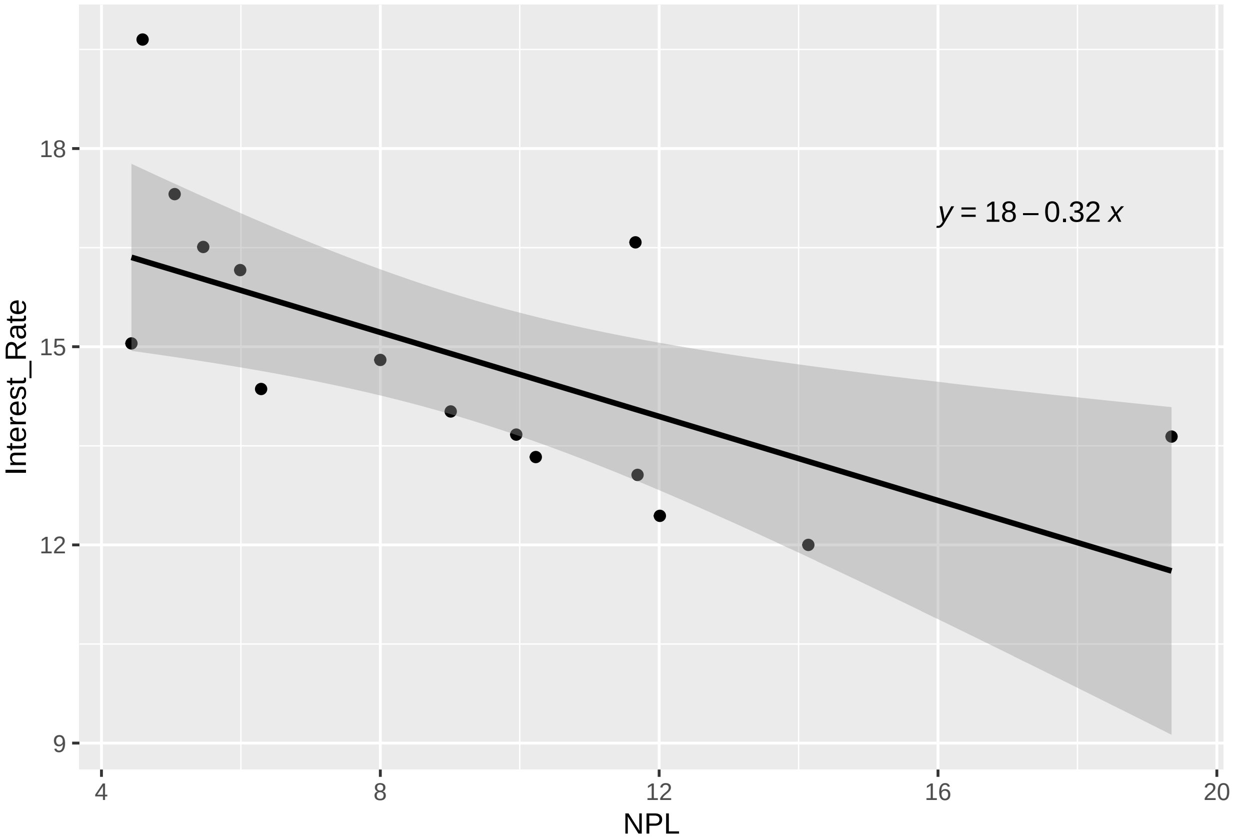
Add the regression line using geom_smooth() and typing in 'lm' as your method for creating the line. This will add the line of the linear regression as well as the standard error of the estimate (in this case +/- 0.01) as a light grey stripe surrounding the line:

```
NPL_graph <- NPL_graph + geom_smooth(method="lm", formula = y ~ x, col="black")
NPL_graph
```



Add the equation for the regression line

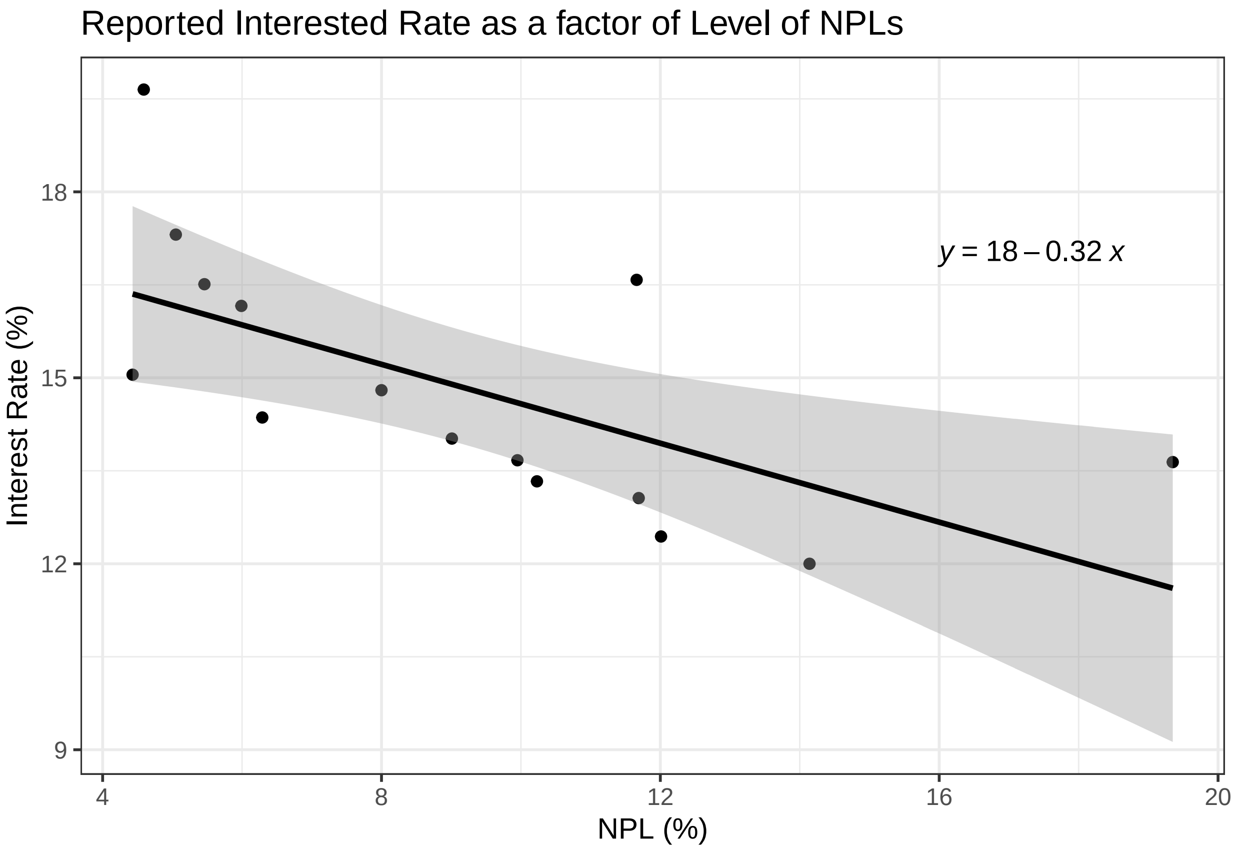
```
NPL_graph <- NPL_graph + stat_regline_equation(label.x = 16, label.y = 17)
NPL_graph
```



Make the graph ready for publication

We can add some style parameters using theme_bw() and making custom labels using labs().

```
NPL_graph +
  theme_bw() +
  labs(title = "Reported Interested Rate as a factor of Level of NPLs",
      x = "NPL (%)",
      y = "Interest Rate (%)")
```



Conclusion of Findings

Based on our findings, there is a significant relationship between interest rate and level of non performing loans ($p < 0.05$, R-Squared = 0.41). Specifically we found a 1.30%% decrease (± 0.45) in the level of non performing loans for every 1% increase in interest rates.

The model can be used to determine the level of non performing loans experienced in Kenya and also determine am approximation of future values.

Although it has significant capacity, interest rate is the only macroeconomic factor of those tested that is significant in forecasting the level of non performing loans in Kenya.