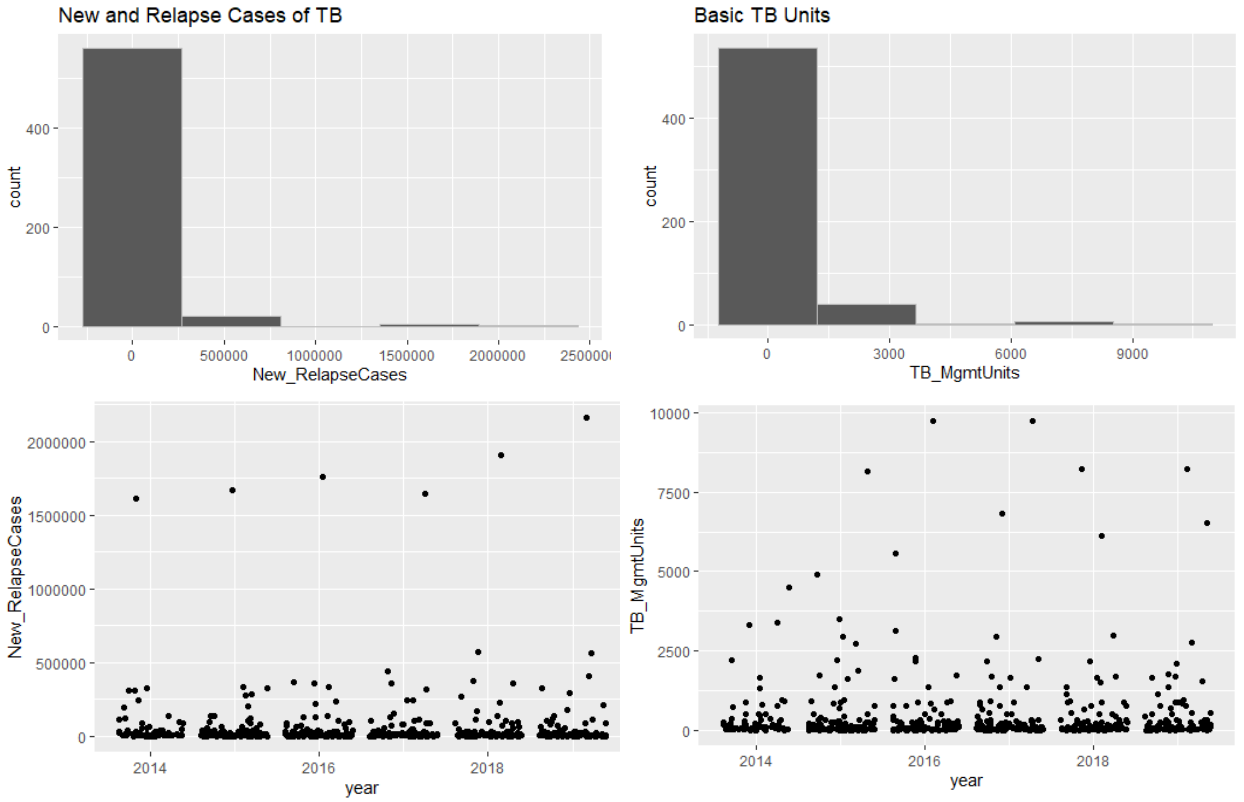
ASSIGNMENT-2

Data Summary:

Variable Name	New_Relapse Cases (Each)	TB_Mgmt Units (Each)
Description	Total of new and relapse cases and cases with unknown previous TB treatment history (Time series data 2013-2019)	Number of TB Basic Management Units in the country (Time series data 2013-2019)
Data Type	Discrete, Interval	Discrete, Interval
Observations	586	586
Mean	53840	429
Median	9082	93
Min	80	0
Max	2162323	9746
Range	2162243	9746
Standard Deviation	193212.6	1101.62



Planning & Analysis:

We test the Pearson coefficient of correlation(r) to determine whether the linear relationship in the sample data effectively models the relationship in the population.

Null Hypothesis (H_0): The correlation coefficient between number of Tuberculosis management units and, new and relapse Tuberculosis cases is not significantly different from 0.

ASSIGNMENT-2

Alternate Hypothesis (H_1): The correlation coefficient between number of Tuberculosis management units and, the new and relapse Tuberculosis cases is significantly different from 0.

Statistical Test: A two-tail t-test with significance level of 5% ($p < 0.05$ to be significant) is used for testing the correlation between the two variables.

Assumptions for this test are as follows:

Normality: The two variables have a normal distribution. We use a Shapiro-Wilk (S-W) test for confirming normality.

Null Hypothesis (H_0): The data are normally distributed.

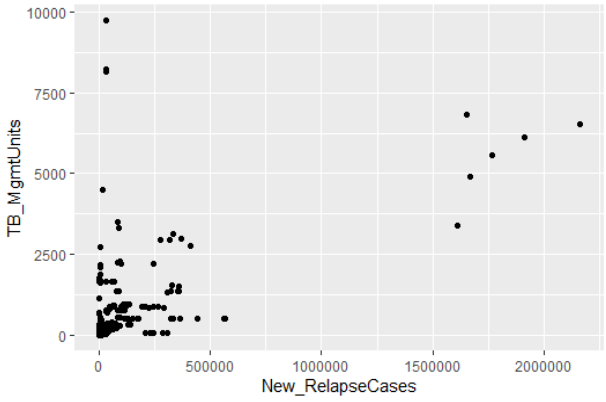
Alternate Hypothesis (H_1): The data are not normally distributed.

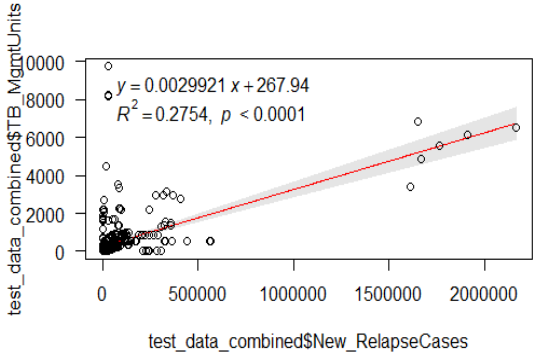
Both variables fail the normality test, hence we conclude that both variables are not normally distributed.

Test Results (from R console):

```
Pearson's product-moment correlation

data:  New_Relapse Cases and TB_Mgmt Units
t = 14.898, df = 584, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4634880 0.5810797
sample estimates:
     cor
0.524783
```





Conclusion:

The p-value from t-test of Pearson coefficient of correlation is less than 5%, which indicates that the results are significant and we can statistically reject the null hypothesis. Hence, there exists a correlation between the number of basic Tuberculosis management units and the number of new and relapsed Tuberculosis cases in a country. There is a moderate positive correlation between these two variables with a mean value of r = + 0.525 and the population correlation coefficient lies between r = [0.463,0.581] with 95 % confidence. However, there's a caveat that the results from the test are not reliable as the data violated the test assumption and can be avoided by gathering more reliable data.