Inferential Statistics
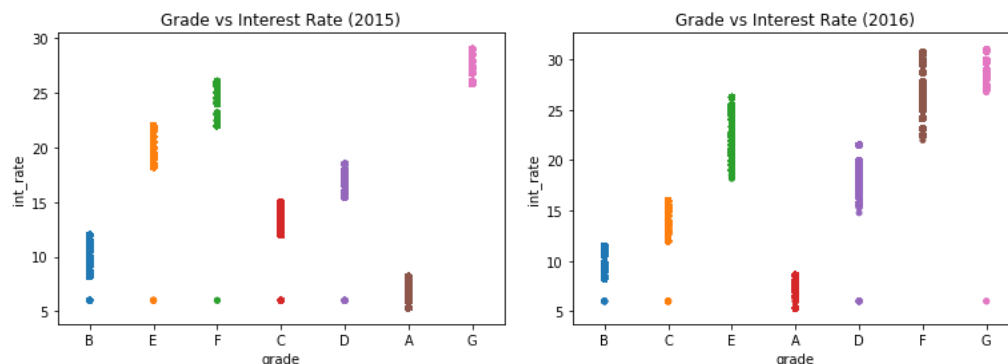
1. Is there any significant change in the interest rate for the year 2015 and 2016 for grade E loans ?

Null Hypothesis: There is no difference in the interest rates for grade E loans in 2015 and 2016.
Alternate Hypothesis : There is difference in the interest rates for grade E loans in 2015 and 2016.

```
In [18]:  #Difference in the means
          print("Difference in the interest rates:", int_rate_2015.mean() - int_rate_2016.mean())

          #t-test
          stats.ttest_ind(int_rate_2015, int_rate_2016)

          Difference in the interest rates: -2.98085786549305

Out[18]:  Ttest_indResult(statistic=-179.11915103758525, pvalue=0.0)
```
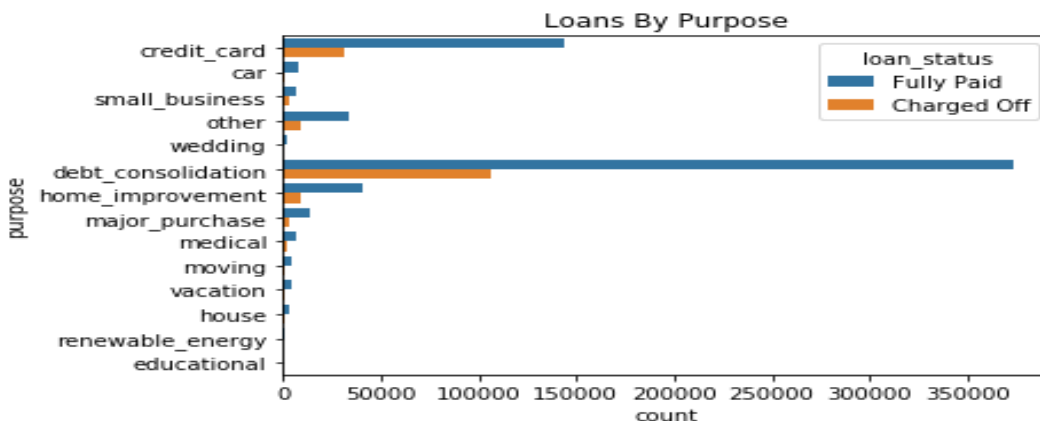
p-value < 0.05, so we can reject the null hypothesis in favor to alternate hypothesis. So there is a significant difference in the interest rates for the years 2015 and 2016 for grade E loans. There is an increase of 2.98 interest rate for grade E loans.

2. Does loan purpose has the significant impact on the charged off rate?

From the above, we can see that the loans with debt consolidation as purpose have highest number of loans issued and charged off loans as well. It is evident from the graph that there is relationship between the purpose and loan status.

Null Hypothesis: Loans Purpose has no significant association with the loan status
Alternate Hypothesis: Loans Purpose has no significant association with the loan status

The chi-square test of independence is a statistical test used to determine whether two categorical variables are independent of each other or not.

```python
In [20]: from scipy.stats import chi2_contingency
         def chisq_of_df_cols(df, c1, c2):
             groupsizes = df.groupby([c1, c2]).size()
             ctsum = groupsizes.unstack(c1)
             print(ctsum)
             # fillna(0) is necessary to remove any NAs which will cause exceptions
             return(chi2_contingency(ctsum.fillna(0)))
         print(chisq_of_df_cols(filtered_loans_df, 'loan_status' , 'purpose'))
```

```
loan_status         Charged Off  Fully Paid
purpose
car                        1320        7386
credit_card               31914      143734
debt_consolidation       106717      373614
educational                  56         270
home_improvement           9312       40252
house                       895        3015
major_purchase             3222       13809
medical                    2004        6486
moving                     1383        4163
other                      9567       33185
renewable_energy            150         445
small_business             3103        7024
vacation                   1007        3871
wedding                     277        1996
(2461.2539968957035, 0.0, 13, array([[  1.83674736e+03,   6.86925264e+03],
       [  3.70573167e+04,   1.38590683e+05],
       [  1.01337778e+05,   3.78993222e+05],
       [  6.87778127e+01,   2.57222187e+02],
       [  1.04567592e+04,   3.91072408e+04],
       [  8.24911803e+02,   3.08508820e+03],
       [  3.59311328e+03,   1.34378867e+04],
       [  1.79117678e+03,   6.69882322e+03],
       [  1.17006672e+03,   4.37593328e+03],
       [  9.01959831e+03,   3.37324017e+04],
       [  1.25530057e+02,   4.69469943e+02],
       [  2.13654267e+03,   7.99045733e+03],
       [  1.02913549e+03,   3.84886451e+03],
       [  4.79545915e+02,   1.79345408e+03]]]))
```

The p-value is 0.0 , we reject the null hypothesis in the favor of alternate hypothesis. So, there is a statistical significant association between loan purpose and the loan status.