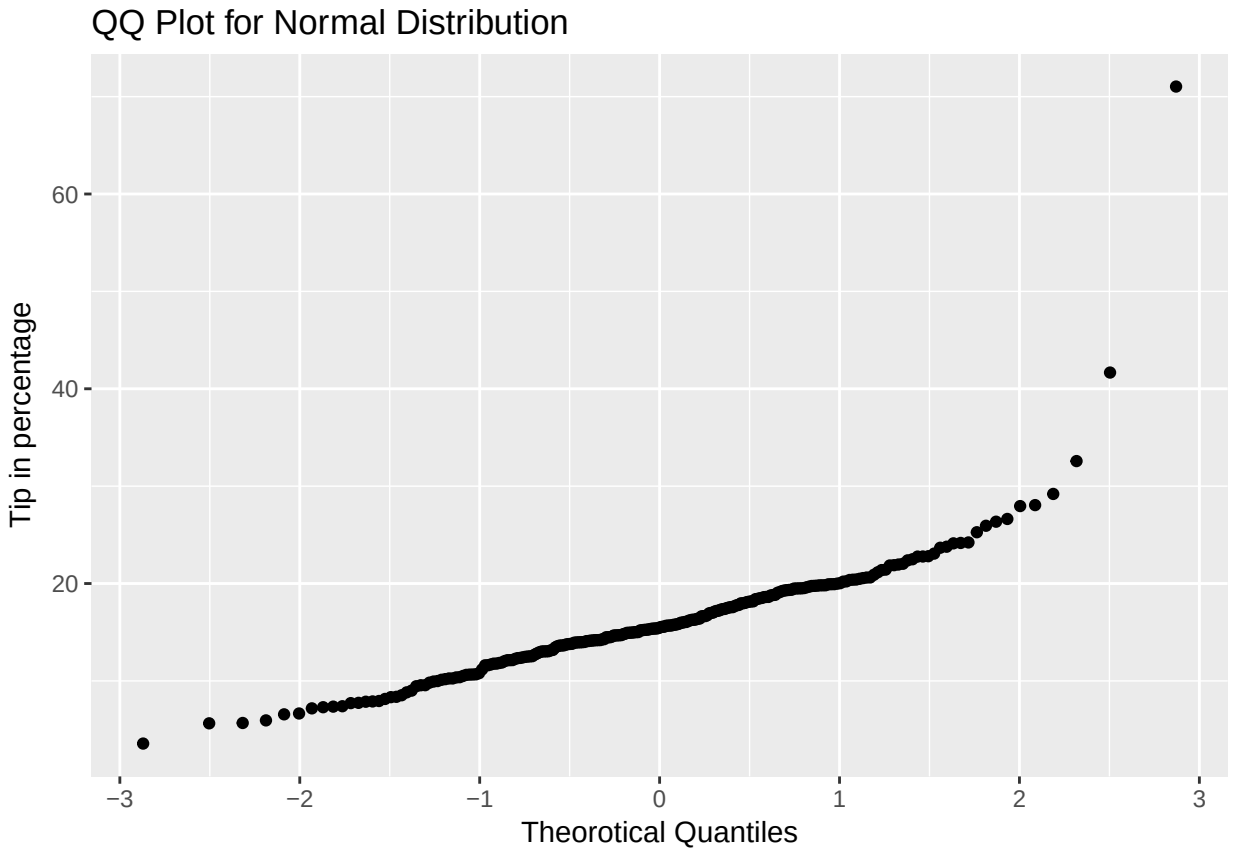
Assignment-1(EDA)

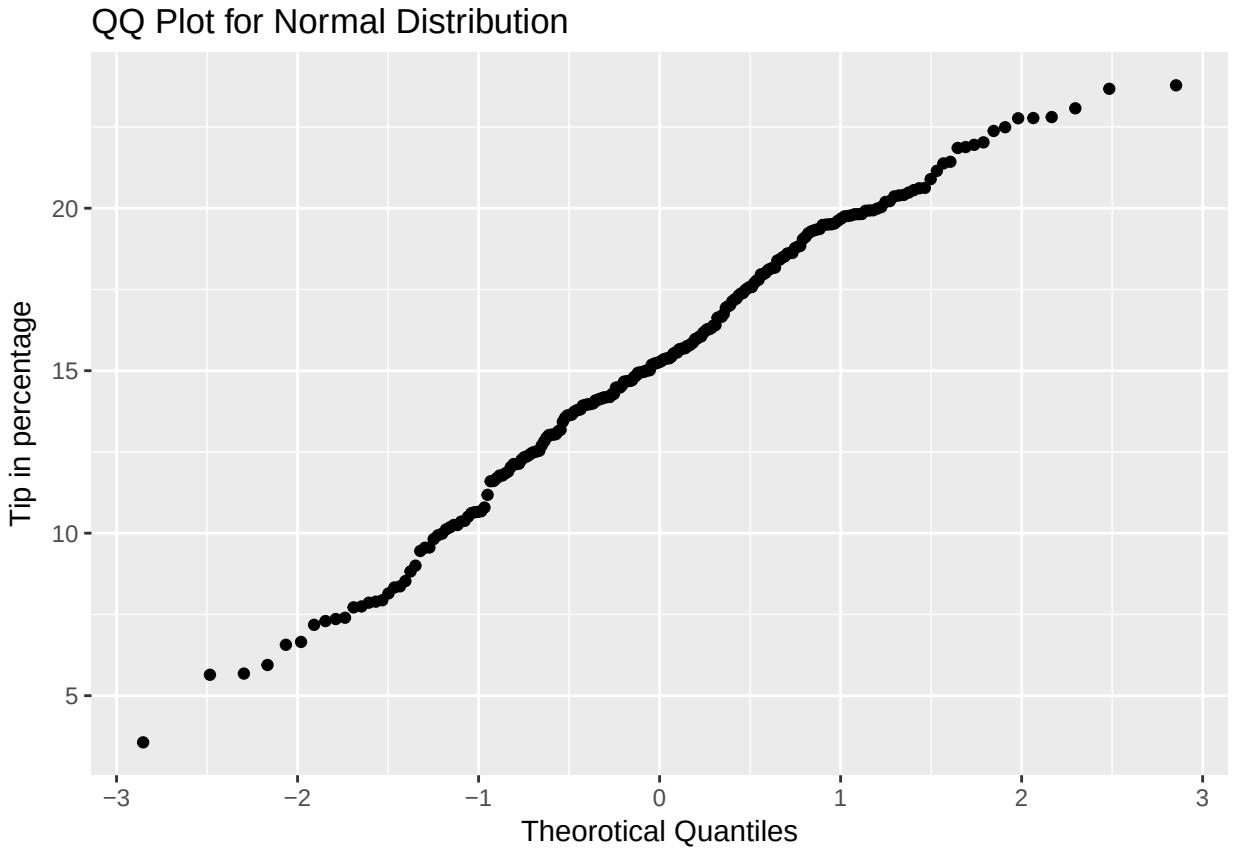
Dwipam

1/18/2017

Question 1:

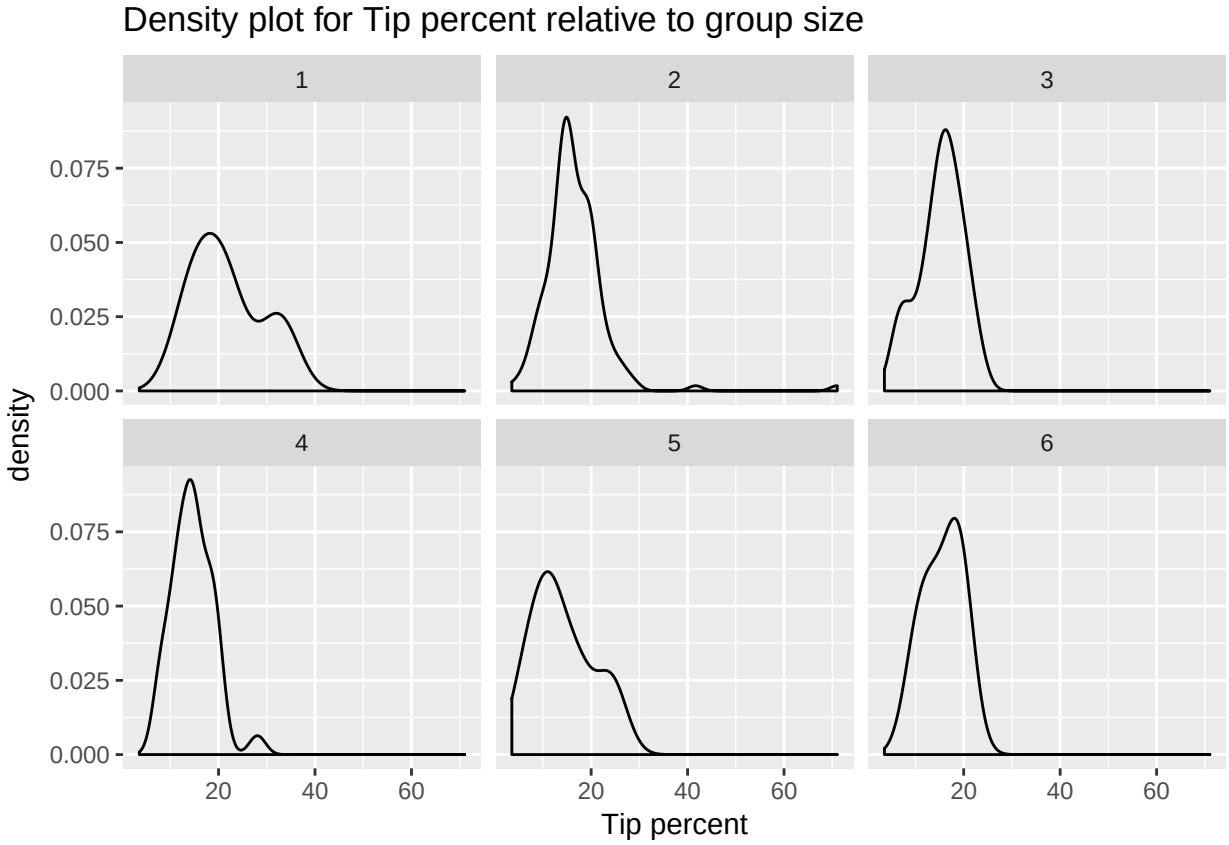


Looking at the Normal QQ plot, we can fairly say that tips percentage does not come from the normal ditribution. Lets remove the outlier. (*As consulted with professor)



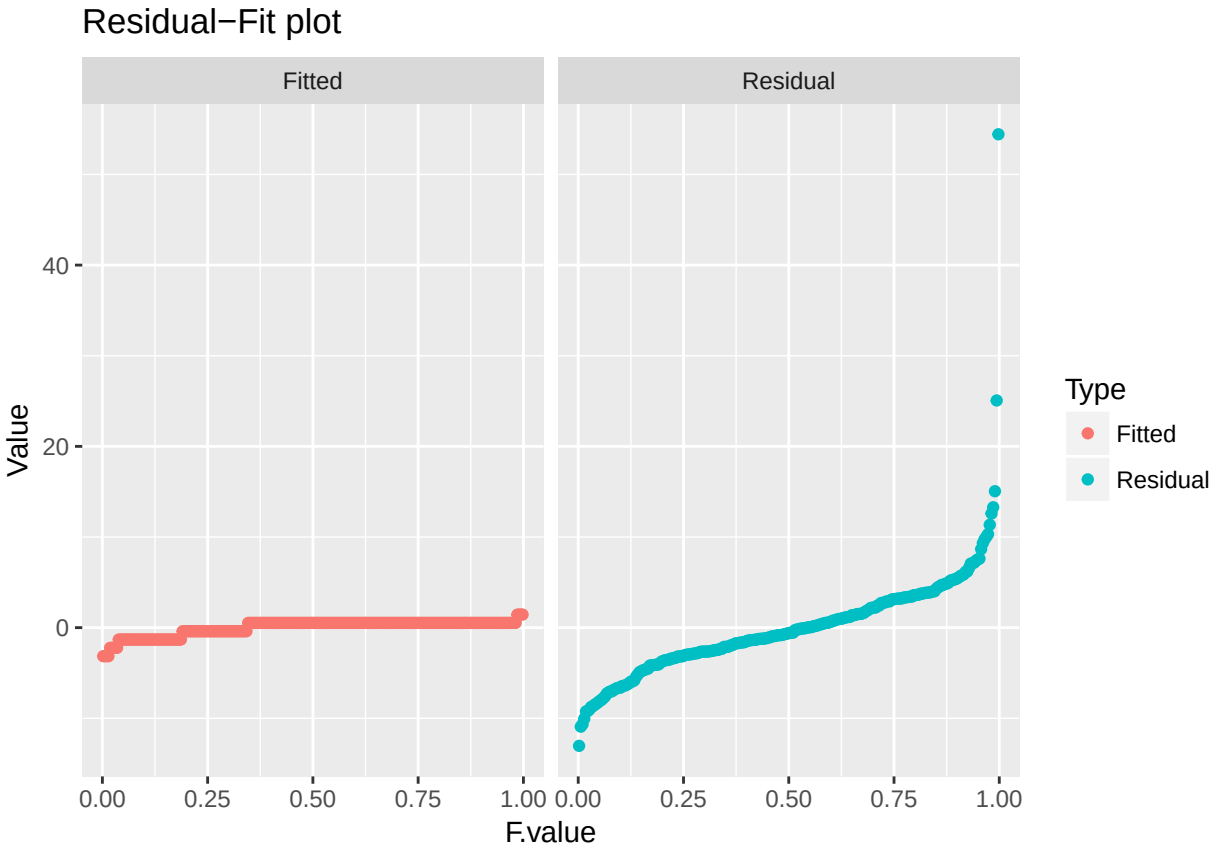
Now after removing the outliers. it does not have stright line for QQ normal plot. Hence, we can confirm that percentage tipped does not have Normal distribution. When we check the Shapiro-Wilk normality test, we get P-value = 0.033. Hence With 95% confidence interval we can reject nul hypothesis and say that data does not follow normal distribution.

Question 2:



Though they differ in sizes, looking at the density plot, it seems that the distribution varies relative to group size. It's fairly common that there are only few occasions where group of 6 people attend the restaurant. One common thing about all of them is that, all of them has mean between 14%-17%, except group size 1. We cannot say anything about distribution of percentage tipped on Group size 1,5 and 6 as the sample size is very low that does not meet basic requirements of statistical analysis. For group size 2 there is a spike at the peak and same for 3 and 4. The width of peak for group size 2 is broader compared to 3 and 4. Group size 4 has second small peak on x axis starting at ~23% and ending at ~30%. There are some extremities found in group size 2 and 4.

Question 3:



Looking at the residual fit plot, there is huge amount of variance still left within the residuals, and only small variance is being captured by the model as indicated in Fitted plot. Also we have pretty low R-square for this model i.e. 0.02.