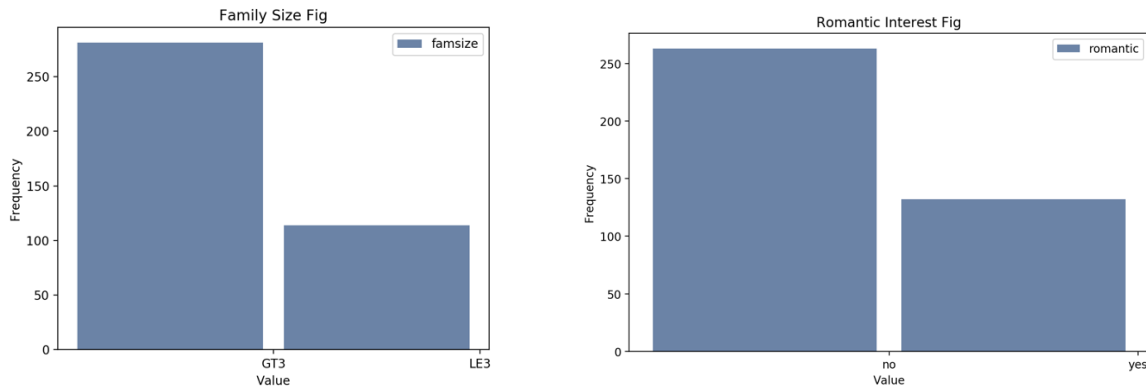# Assignment1 Analysis Report
- Xiaowei Wan

Plots:

Histogram shows above.
Test this data set is good sample or not, I am using one sample t test to test it against the general population.

Using python scipy stats to calculate one simple t test statistics and p-value of family size and romantic data. Null hypothesis is m = 0. P-value is less than 0.05, the result is significant. The same for romantic. The third result is two sample test between family size and romantic. The result is significant as well since p-value = 0.

```
family size: t-statistic = 31.164 pvalue = 0.0000
romantic: t-statistic = 14.062 pvalue = 0.0000
t-statistic = 11.447 pvalue = 0.0000
```

Assignment 2 – Probability Mass Functions
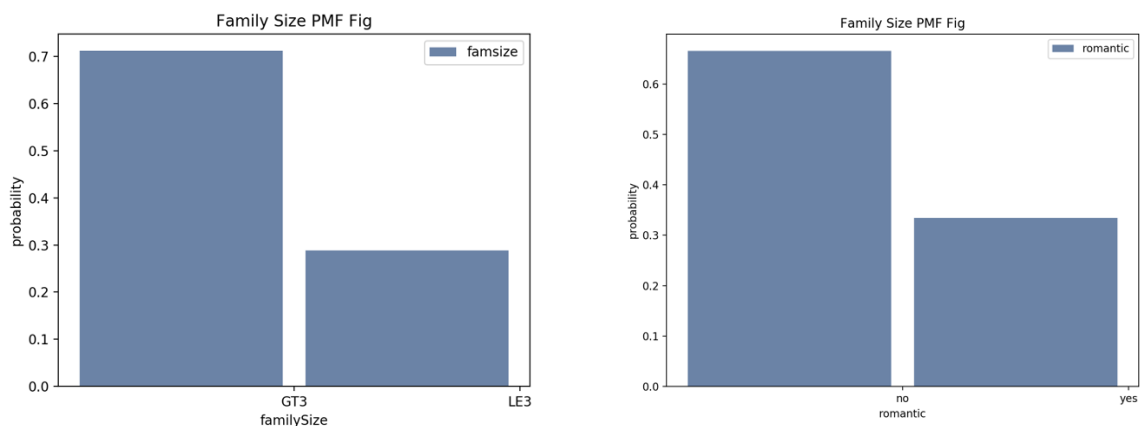
PMF graph of family size shows that probability of family member less than 3 is around 30%. Student is in relationship probability is around 33%.

Python code:

```python
10
11    famList = list()
12    romanList = list()
13    data = open("student-mat.csv", "r")
14    # read famsize and romantic data into list
15    for line in data:
16      spLine = line.split(",")
17      famList.append(spLine[4])
18      romanList.append(spLine[22])
19    # remove first element
20    famList = famList[1:]
21    romanList = romanList[1:]
22
23    # calculate percentage of family size have three or less family members
24    famLE3 = famList.count("LE3")/float(len(famList));
25    print('family has three or less members percentage=', "{:.2f}".format(famLE3))
26    # calculate student in relationship percentage
27    romanticY = romanList.count("yes")/float(len(romanList))
28    print('student in relationship percentage=', "{:.2f}".format(romanticY))
29
30    famSizeHist = thinkstats2.Hist(famList, label='famsize')
31    romanList = thinkstats2.Hist(romanList, label='romantic')
32
33    # plot familiy size histogram
34    thinkplot.Hist(famSizeHist)
35    thinkplot.Show(xlabel='Value', ylabel='Frequency', title='Family Size Fig')
36    # plot romantic interest histogram
37    thinkplot.Hist(romanList)
38    thinkplot.Show(xlabel='Value', ylabel='Frequency', title='Romantic Interest Fig')
39
40    # Use One Sample T Test to valuate whether this data set is a good sample or not.
41    # Our null hypothesis is that: true_mu = 0
42    famList = map(lambda x: 1 if x == 'GT3' else 0, famList)
43    romanList = map(lambda x: 1 if x == 'yes' else 0, romanList)
44    true_mu = 0
45
46    print('family size: t-statistic = %6.3f pvalue = %6.4f' %  stats.ttest_1samp(famList, true_mu))
47    print('romantic: t-statistic = %6.3f pvalue = %6.4f' %  stats.ttest_1samp(romanList, true_mu))
48    print('t-statistic = %6.3f pvalue = %6.4f' %  stats.ttest_ind(famList, romanList, equal_var=False))
49
50    # Probability Mass Functions.
51    famPmf = thinkstats2.Pmf(famList)
52    romanPmf = thinkstats2.Pmf(romanList)
53
54    famHist = thinkstats2.Hist(famPmf, label='famsize')
55    romanHist = thinkstats2.Hist(romanPmf, label='romantic')
56
57    # Plot family size Pmf
58    thinkplot.Hist(famHist)
59    thinkplot.Show(xlabel='familySize', ylabel='probability', title='Family Size PMF Fig')
60
61    # Plot romantic Pmf
62    thinkplot.Hist(romanHist)
63    thinkplot.Show(xlabel='romantic', ylabel='probability', title='Family Size PMF Fig')
```