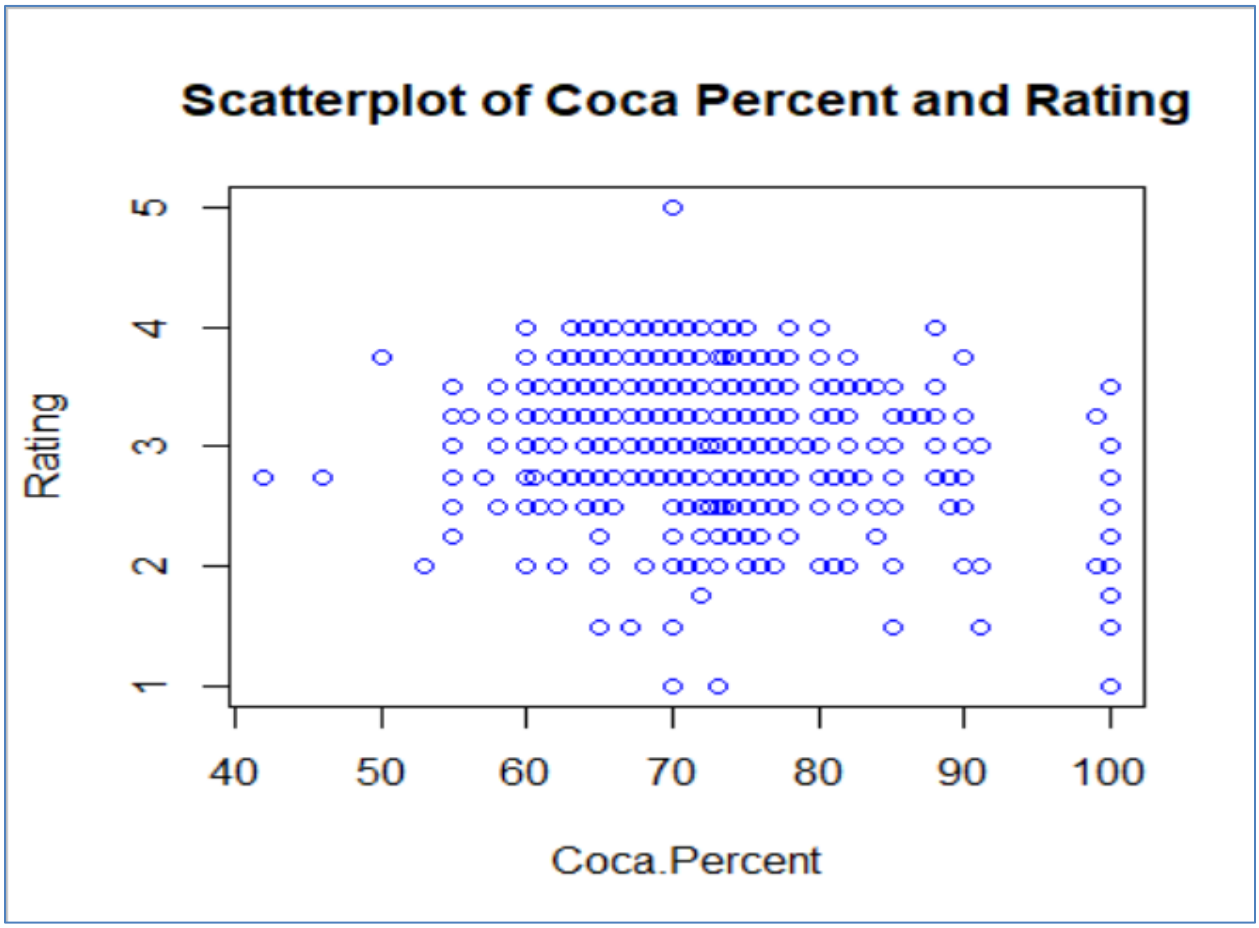
Step 7: Scatterplots and correlation

1. For at least one pair of quantitative columns/variables:

*** Plot a scatterplot of the data in the two columns/variables.**



*** Compute the correlation between the columns/variables.**

Correlation = -0.165

P-value = < 0.0001

*** Interpretation of the scatterplot and correlation. (ex. how closely are the two variables related? If there is a relationship, does it appear linear?)**

P-value of correlation coefficient shows that there is significant linear relationship between two variables, but the relationship is very week. Scatter plot is also suggesting a very week linear relationship between two variables.

Step 8: Confidence Intervals

1. Choose at least 2 quantitative columns, and do the following for each column:

The two chosen columns are “Rating” and the “Cocoa percent”

*** Compute the 95% confidence interval for the mean**.

1. Cocoa Percent

Mean = 71.698

The 95% confidence interval is estimated to be [70.098, 73.298]

2. Rating

Mean = 3.186

The 95% confidence interval is estimated to be [3.065, 3.307]

*** Interpretation of the confidence intervals based on the code. (ex. are the confidence intervals large or small? How much should we trust our estimates of the means?)**

Cocoa percentage confidence interval shows that we are 95% confident that the average cocoa percentage will lie between 70.098 and 73.298. Rating confidence interval can be interpreted as, we are 95% confident that the average rating will lie between 3.065 and 3.307.

The intervals are precise and trustable in a way because they are containing the true value of mean already computed.