Lab1
Yiqiao Jin
UID: 305107551

1.
Unit of observation is one person. 22 different variables are recorded.

Numerical variables:
Apgar5, BirthWeight, Gestation, Fage, Mage, Feduc, Meduc, TotPreg, Visits, Marital, Racemom, Racedad, Gained

Categorical variables:
BMonth, BDay, DOW, Gender, Premie, LowBirth, Hispmom, Hispdad, Smokes

2.

New_2009Births

Smokes	false	1744
	true	192
Column Summary		1936

S1 = count( )

There are 192 smokers and 9.9% of the sample are smoking mothers

3.

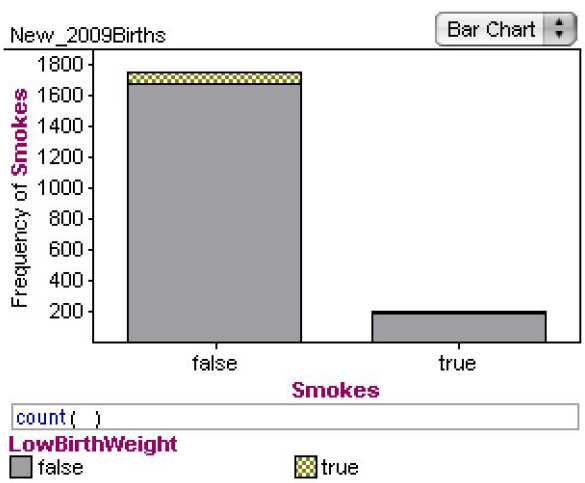
New_2009Births

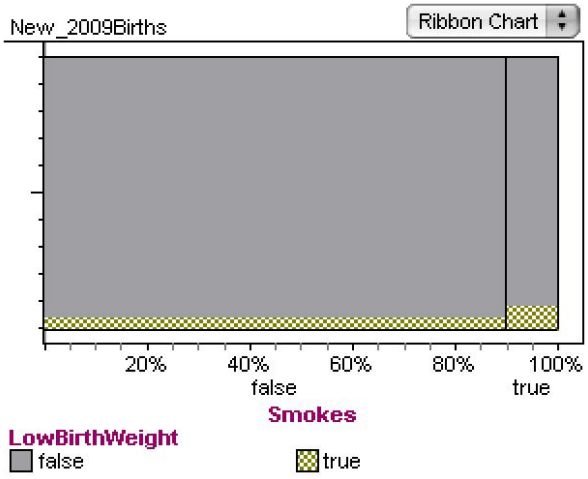
		LowBirthWeight		Row Summary
		false	true	
Smokes	false	1673 0.959289	71 0.040711	1744 1
	true	177 0.921875	15 0.078125	192 1
Column Summary		1850 0.955579	86 0.0444215	1936 1

S1 = count( )
S2 = rowproportion

It is possible that low birth weight is associated with mothers' smoking habits. The two-way table suggests that 7.81% of the samples with smoking mothers have low birth weight, whereas only 4.07% of the samples with non-smoking mothers have low birth weight.

4.





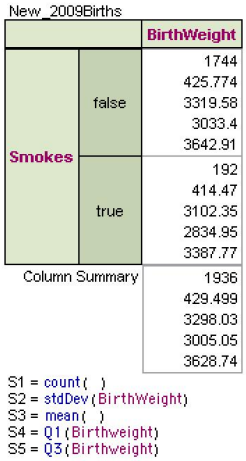
Ribbon chart better illustrates the possible association between smoking and low birth weight because it shows the percentage, or the relative ratio, of people with low birth weight among the two groups. The bar chart only displays the absolute numbers, which is unsuitable for comparing percentages.

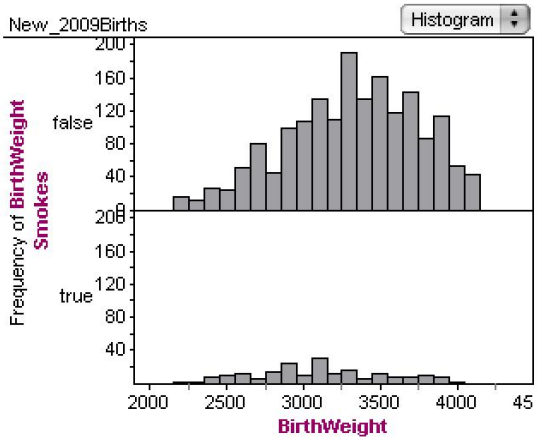
5.

I dropped the numerical variable BirthWeight to replace the categorical variable in Question 3. After that, Fathom automatically adds the mean() function to the table.

I added Q1, Q3, Standard Deviation and found that the group of people with smoker mothers have lower mean, Q1, Q3 than the group of people with non-smoker mothers. Also the smoker group has a smaller standard deviation.

The new table is shown below:





6.

The graph shows that there is a correlation between babies with lower weights and mothers' smoking habit. The graph displays a median baby weight of about 3250 for people with non-smoker mothers and 3100 for people with smoker mothers. The result is expected.

Summary:

Concepts covered in the lab include numerical and categorical data; types of charts such as pie chart and bar chart; graphs such as histogram and dot plot. The spread and shapes of graphs including unimodal, bimodal and multimodal are also covered. I have seen these concepts in the textbook, lectures as well as homework. With the help of these concepts, I was able to create two way tables and charts to find correlations between data(e.g. correlation between smoking habit of mothers and low birth weight). Additionally, I was also able to interpret graphs such as histograms in this lab and identify the center, the median, the modes and the shape of graphs. They helped me determine that there is a higher percentage of babies with low birth weights who have smoker mothers than those with non-smoker mothers.