eda_visuals

October 17, 2017

1 EDA with Visuals

Create visualizations to answer the quiz questions below this notebook.

In [6]:
```
# Load dataset
import pandas as pd
% matplotlib inline

wines_df = pd.read_csv('winequality_edited.csv')
wines_df.head()
```

Out[6]:

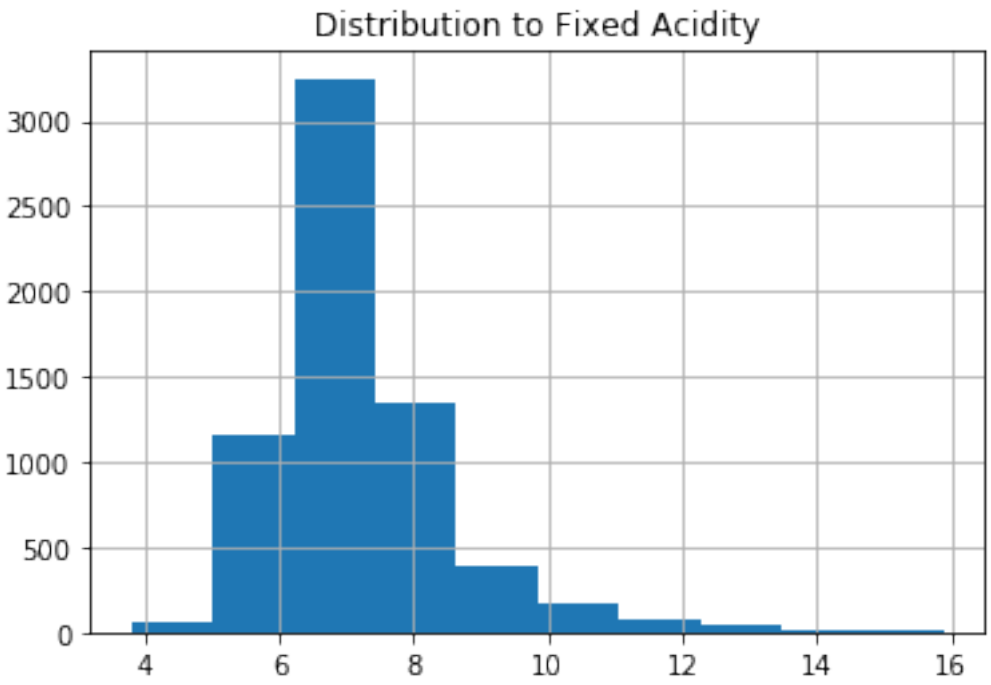
	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	\
0	7.0	0.27	0.36	20.7	0.045	
1	6.3	0.30	0.34	1.6	0.049	
2	8.1	0.28	0.40	6.9	0.050	
3	7.2	0.23	0.32	8.5	0.058	
4	7.2	0.23	0.32	8.5	0.058	

	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	\
0	45.0	170.0	1.0010	3.00	0.45	
1	14.0	132.0	0.9940	3.30	0.49	
2	30.0	97.0	0.9951	3.26	0.44	
3	47.0	186.0	0.9956	3.19	0.40	
4	47.0	186.0	0.9956	3.19	0.40	

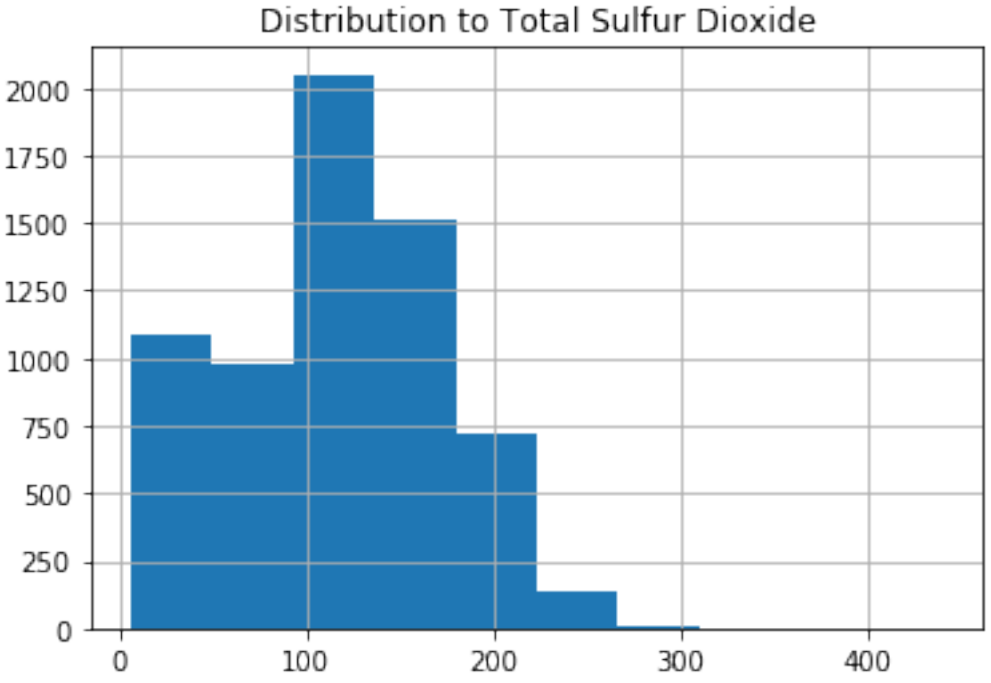
	alcohol	quality	color
0	8.8	6	white
1	9.5	6	white
2	10.1	6	white
3	9.9	6	white
4	9.9	6	white

1.0.1 Histograms for Various Features

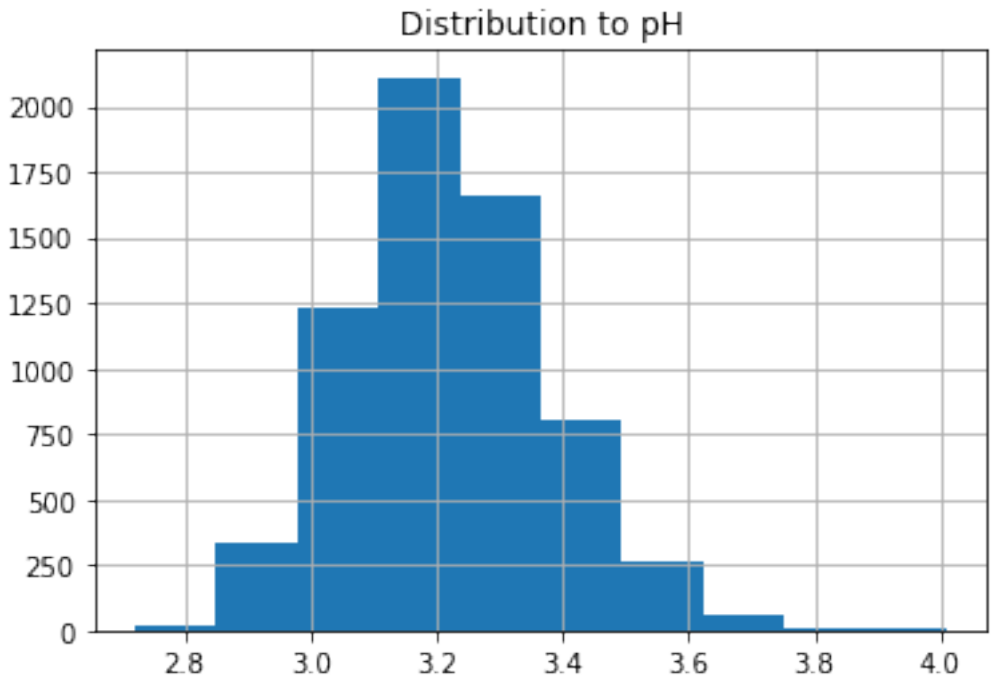
In [43]:
```
wines_df['fixed_acidity'].hist().set_title('Distribution to Fixed Acidity');
```



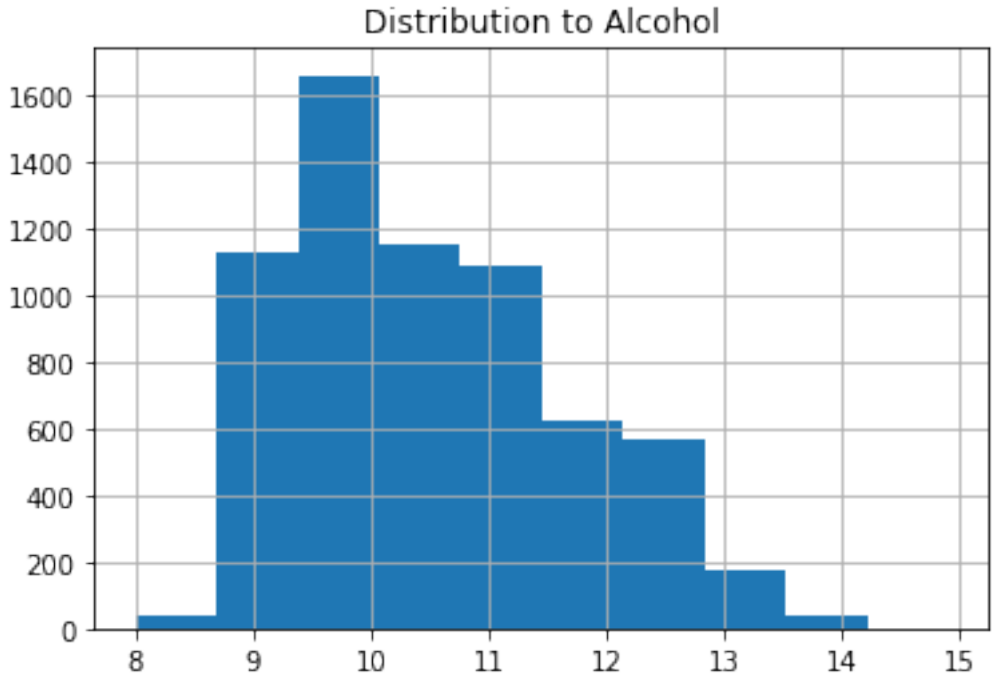

```
In [44]: wines_df['total_sulfur_dioxide'].hist().set_title('Distribution to Total Sulfur Dioxide
```



In [45]: `wines_df['pH'].hist().set_title('Distribution to pH');`

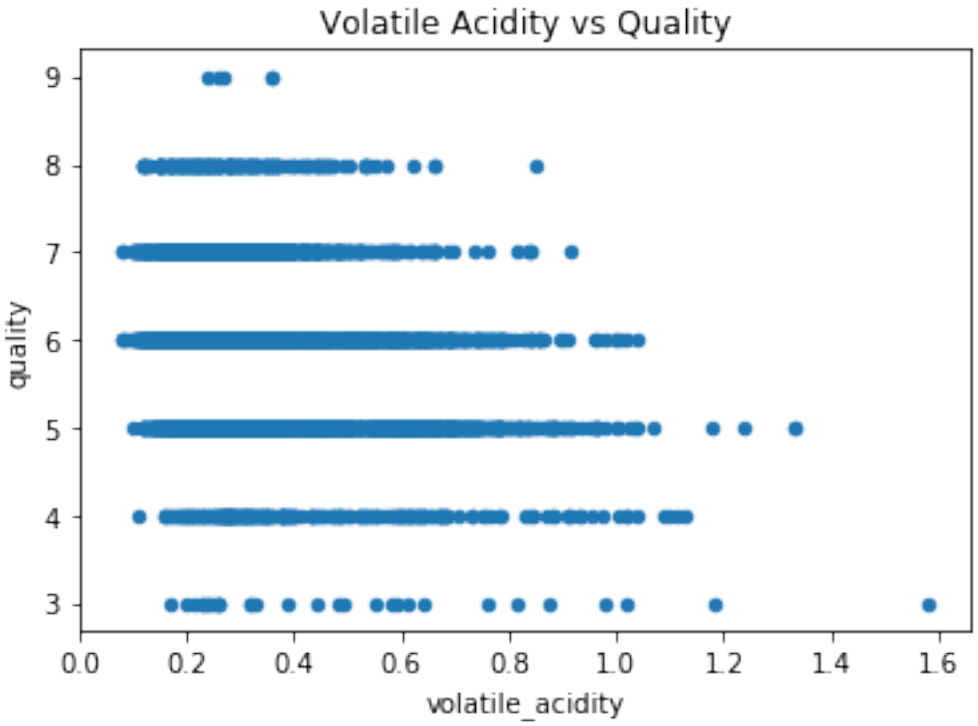


In [46]: `wines_df['alcohol'].hist().set_title('Distribution to Alcohol');`

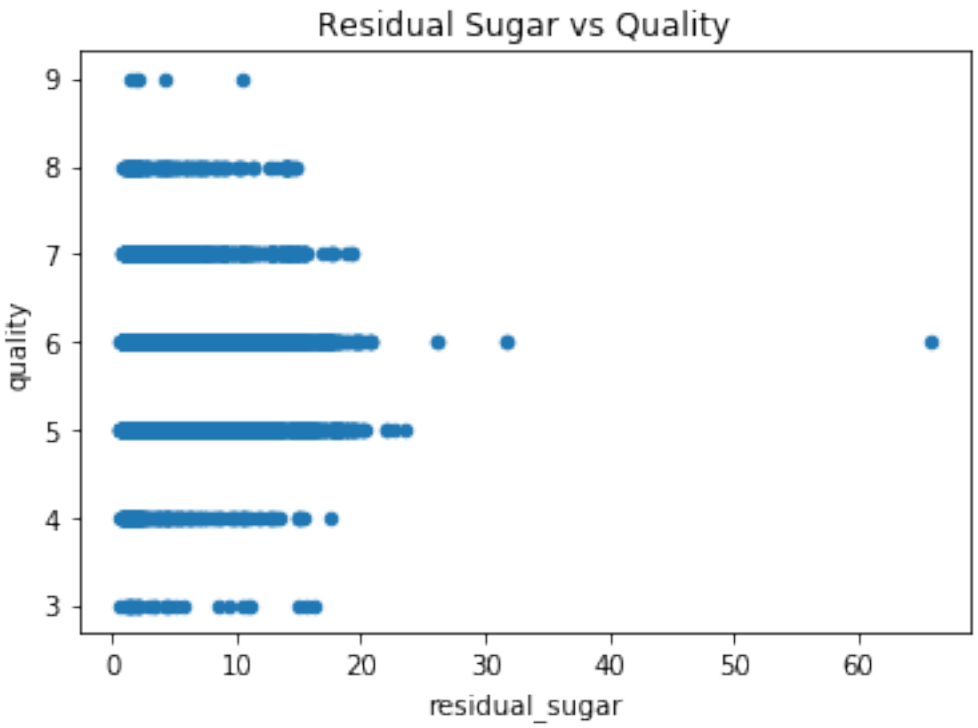


1.0.2 Scatterplots of Quality Against Various Features

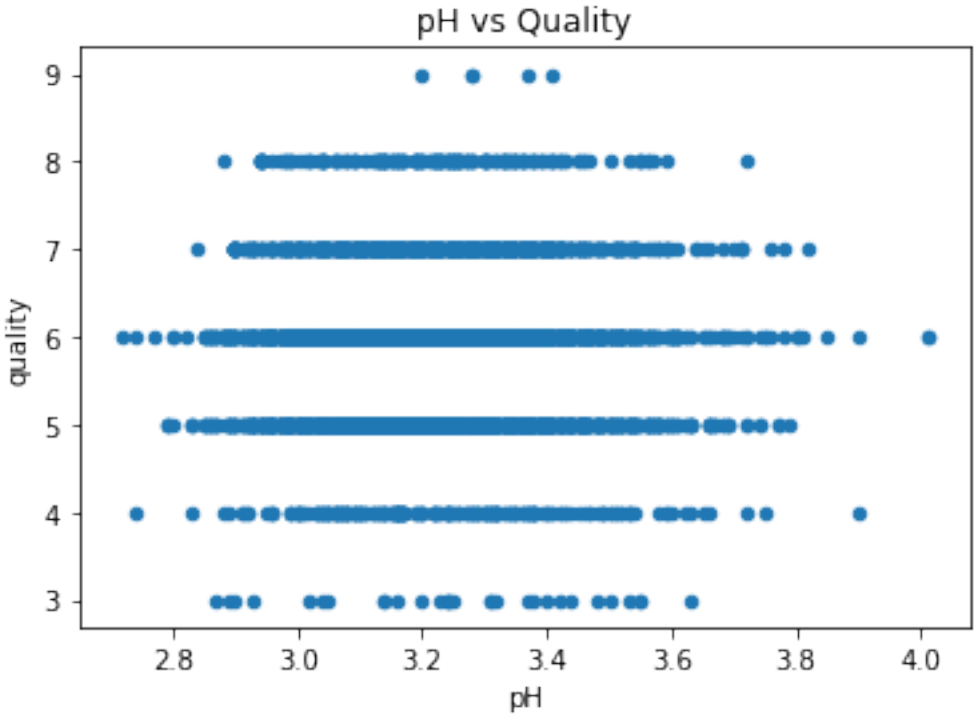
In [38]:
```
wines_df.plot(x='volatile_acidity', y='quality', kind='scatter',
              title='Volatile Acidity vs Quality');
```



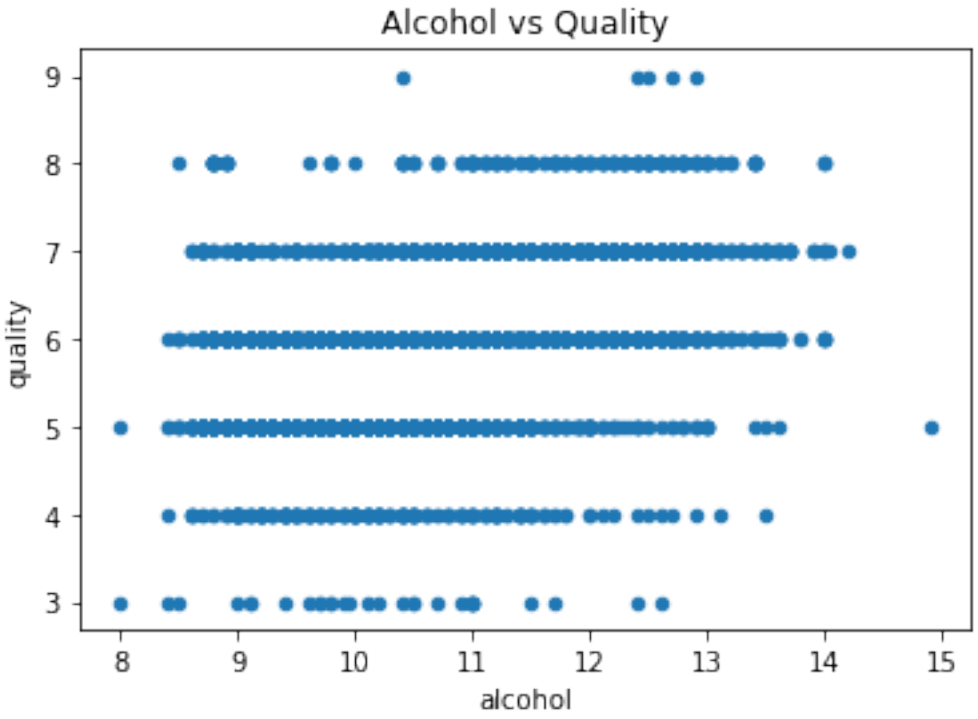
In [47]:
```
wines_df.plot(x='residual_sugar', y='quality', kind='scatter',
              title='Residual Sugar vs Quality');
```




```
In [48]: wines_df.plot(x='pH', y='quality', kind='scatter',
                       title='pH vs Quality');
```



In [49]:
```
wines_df.plot(x='alcohol', y='quality', kind='scatter',
              title='Alcohol vs Quality');
```



In [ ]: