eda_visuals

October 18, 2017

1 EDA with Visuals

Create visualizations to answer the quiz questions below this notebook.

In [1]:
```
# Load dataset
import pandas as pd

df = pd.read_csv('winequality_edited.csv')

df.head()
```

Out[1]:

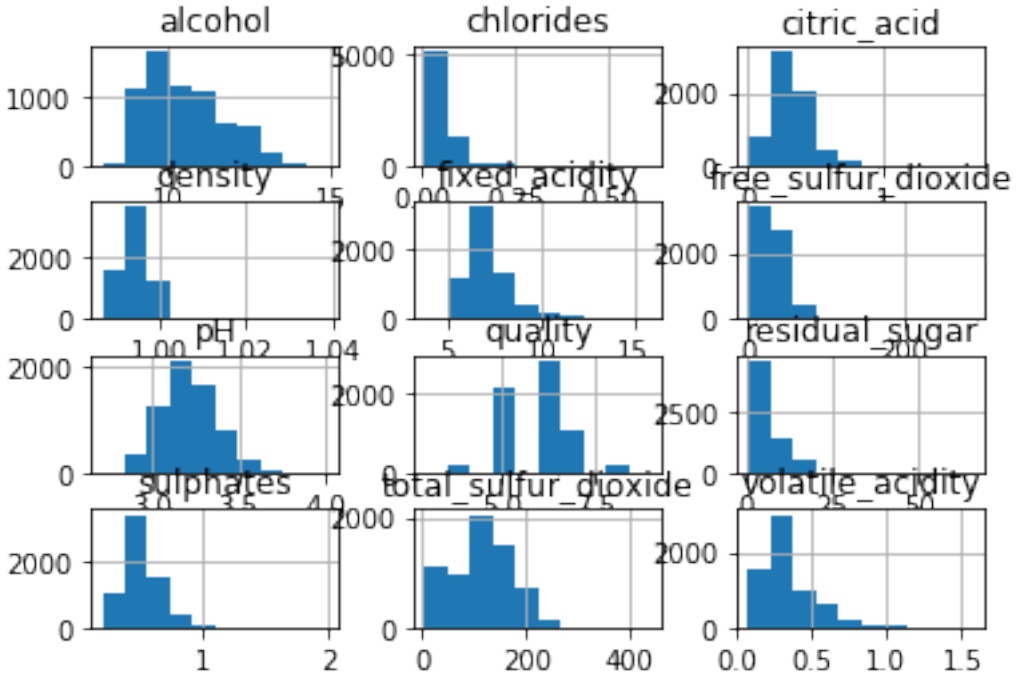
	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide	density	pH	sulphates	alcohol	quality	color
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	RED
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8	5	RED
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8	5	RED
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8	6	RED
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4	5	RED

In [2]:
```
df.shape
```

Out[2]:
```
(6497, 13)
```

1.0.1 Histograms for Various Features

In [4]:
```
% matplotlib inline
df.hist();
```



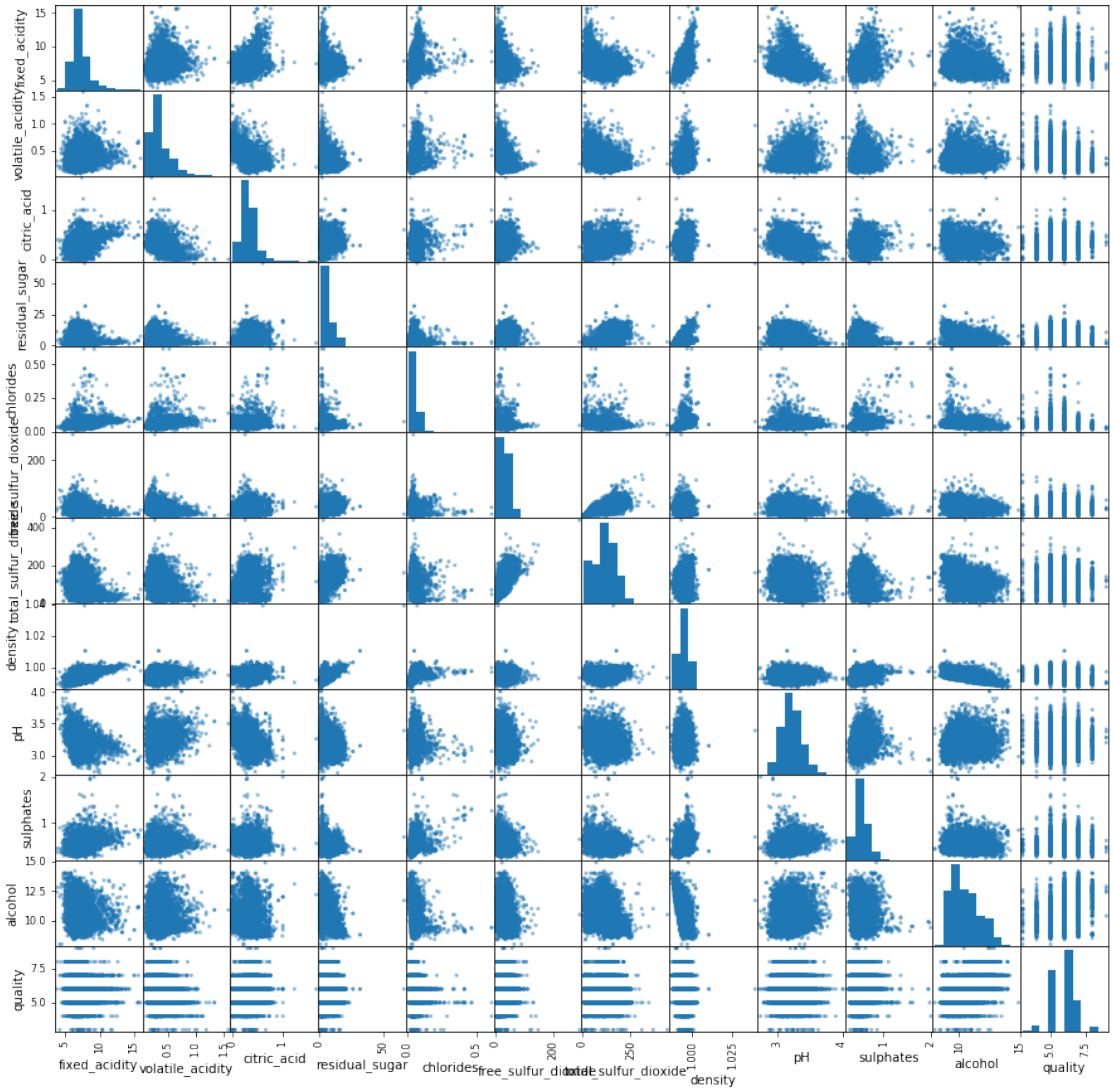
In [ ]:

In [ ]:

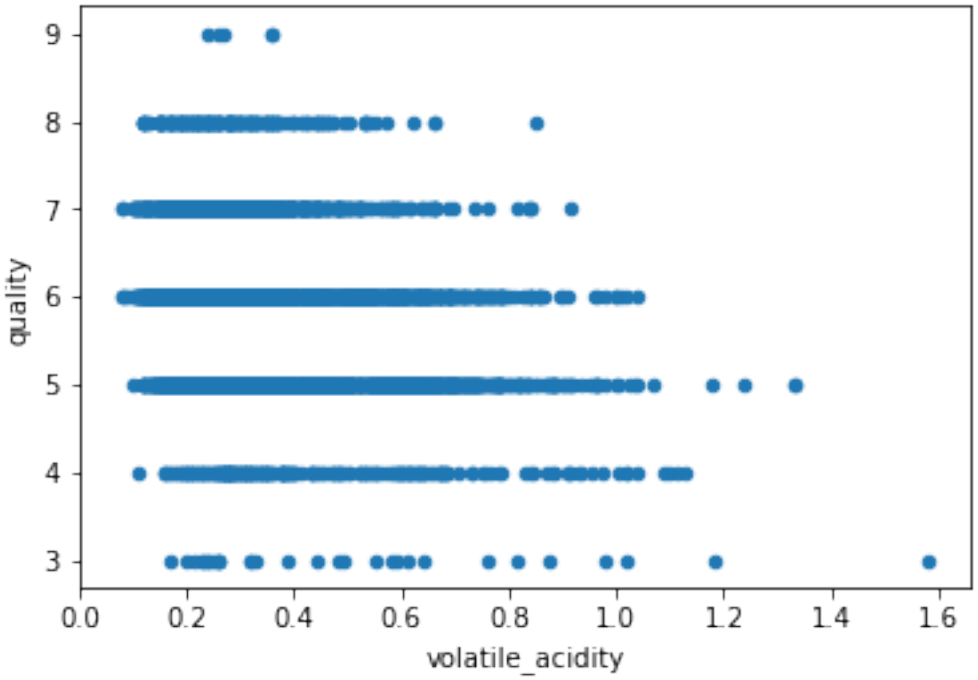
In [ ]:

1.0.2 Scatterplots of Quality Against Various Features

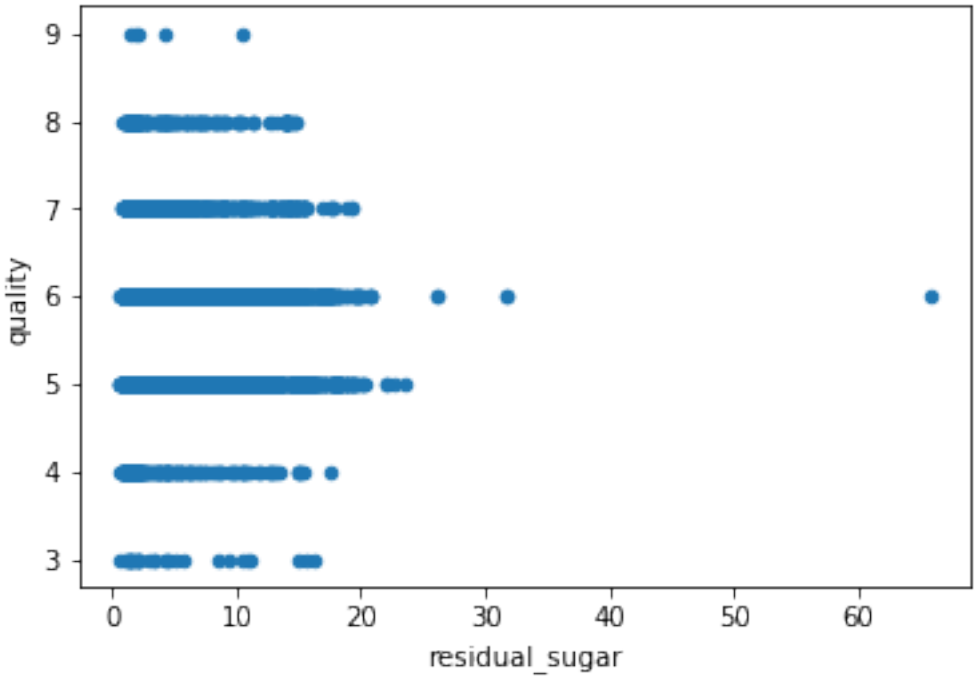
In [5]:
```
pd.plotting.scatter_matrix(df, figsize=(15,15));
```



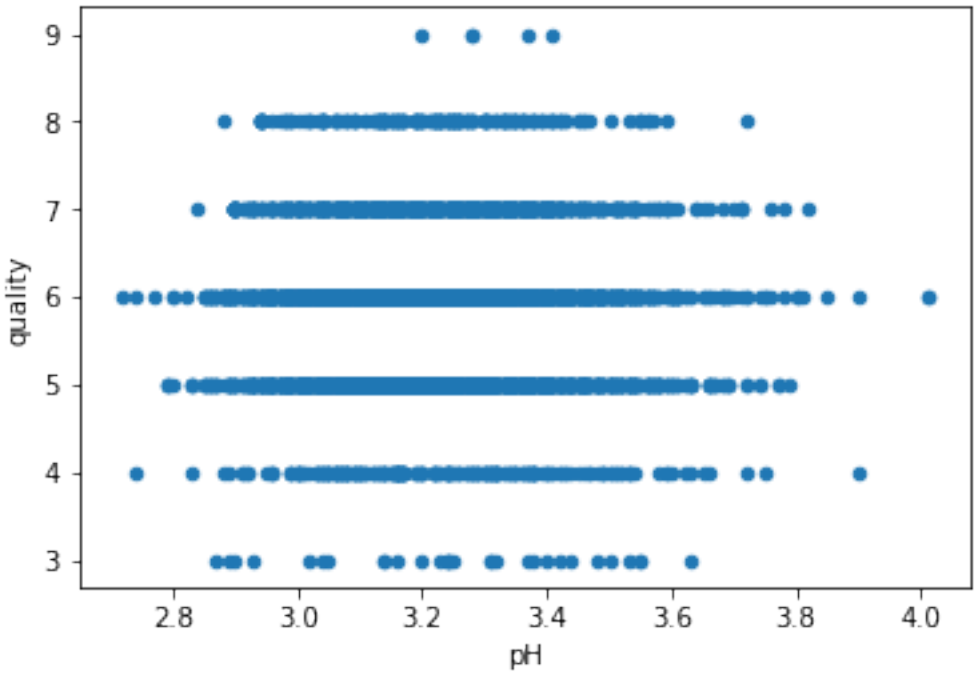

```
In [6]: df.plot(x='volatile_acidity', y='quality', kind='scatter');
```



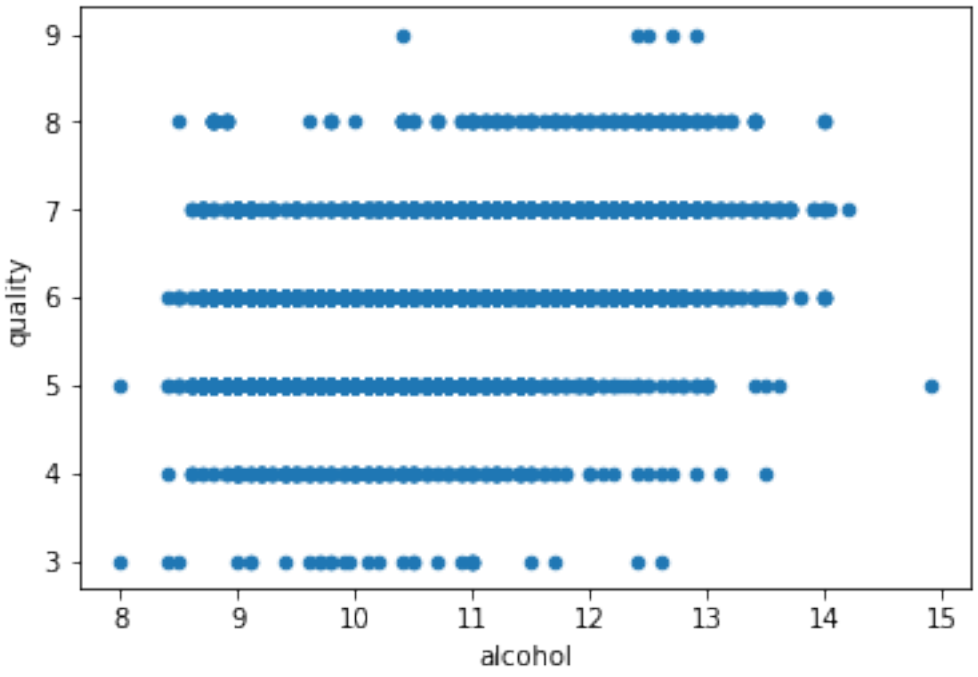
In [7]: df.plot(x='residual_sugar', y='quality', kind='scatter');



In [8]: df.plot(x='pH', y='quality', kind='scatter');



In [11]: df.plot(x='alcohol', y='quality', kind='scatter');



```
In [ ]:
```