

# wine\_visualizations

October 17, 2017

## 1 Plotting with Matplotlib

Use Matplotlib to create bar charts that visualize the conclusions you made with groupby and query.

```
In [168]: # Import necessary packages and load 'winequality_edited.csv'
import pandas as pd
import matplotlib.pyplot as plt
% matplotlib inline

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

```
Out[168]:
```

	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	acidity_levels
0	8.8	6	white	High
1	9.5	6	white	Medium
2	10.1	6	white	Medium
3	9.9	6	white	ModeratelyHigh
4	9.9	6	white	ModeratelyHigh

### 1.0.1 #1: Do wines with higher alcoholic content receive better ratings?

Create a bar chart with one bar for low alcohol and one bar for high alcohol wine samples. This first one is filled out for you.

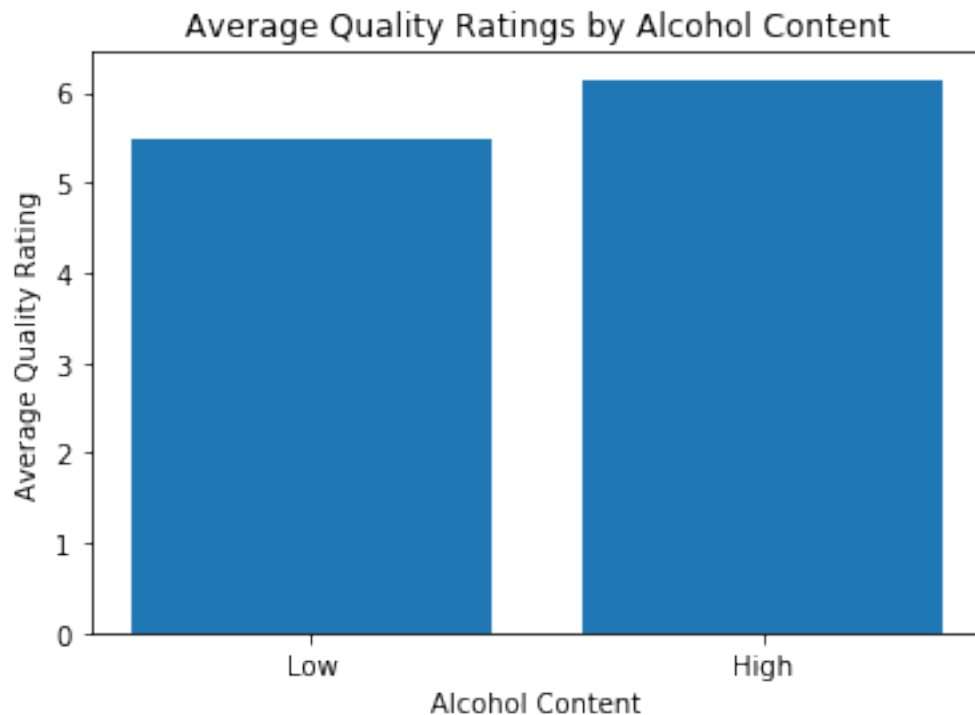
```

In [169]: # Use query to select each group and get its mean quality
median = df['alcohol'].median()
low = df.query('alcohol < {}'.format(median))
high = df.query('alcohol >= {}'.format(median))

mean_quality_low = low['quality'].mean()
mean_quality_high = high['quality'].mean()

In [170]: # Create a bar chart with proper labels
locations = [1, 2]
heights = [mean_quality_low, mean_quality_high]
labels = ['Low', 'High']
plt.bar(locations, heights, tick_label=labels)
plt.title('Average Quality Ratings by Alcohol Content')
plt.xlabel('Alcohol Content')
plt.ylabel('Average Quality Rating');

```



### 1.0.2 #2: Do sweeter wines receive higher ratings?

Create a bar chart with one bar for low residual sugar and one bar for high residual sugar wine samples.

```

In [171]: # Use query to select each group and get its mean quality
sugar_med = df['residual_sugar'].median()

```

```

low_sugar = df.query('residual_sugar < {}'.format(sugar_med))
high_sugar = df.query('residual_sugar >= {}'.format(sugar_med))

low_sugar_quality = low_sugar['quality'].mean()
high_sugar_quality = high_sugar['quality'].mean()

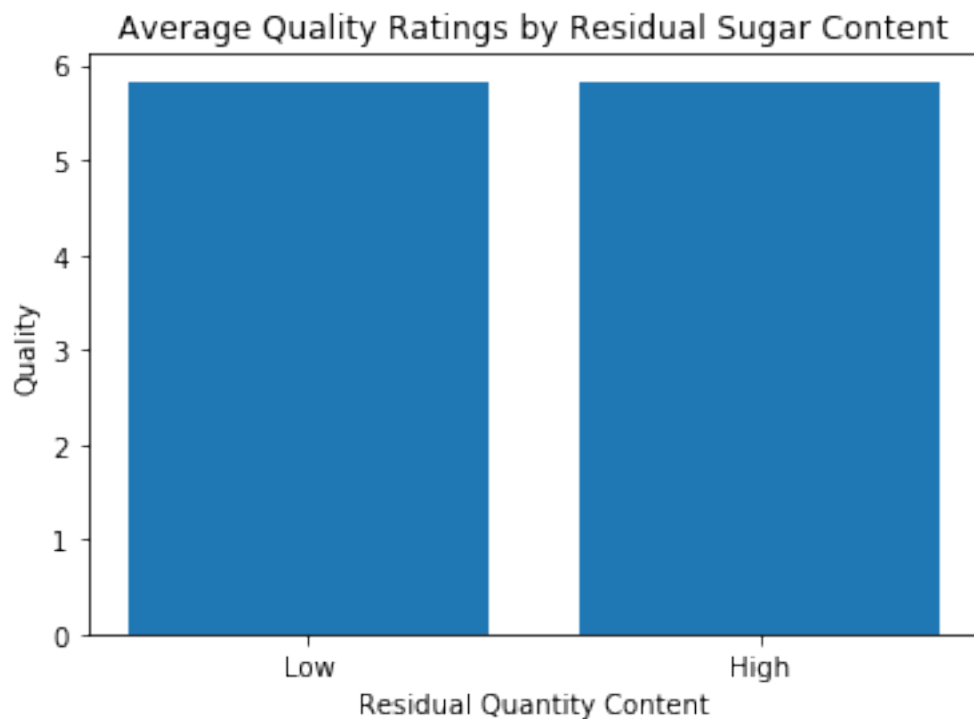
```

In [172]: *# Create a bar chart with proper labels*

```

locations = [1,2]
data = [low_sugar_quality, high_sugar_quality]
labels = ['Low', 'High']
plt.bar(locations, data, tick_label=labels)
plt.title("Average Quality Ratings by Residual Sugar Content")
plt.xlabel("Residual Quantity Content")
plt.ylabel("Quality");

```



### 1.0.3 #3: What level of acidity receives the highest average rating?

Create a bar chart with a bar for each of the four acidity levels.

```

In [173]: # Use groupby to get the mean quality for each acidity level
acidities = df.groupby('acidity_levels').mean()

acidity_lows = df.query('acidity_levels=="Low"')['quality'].mean()

```

```

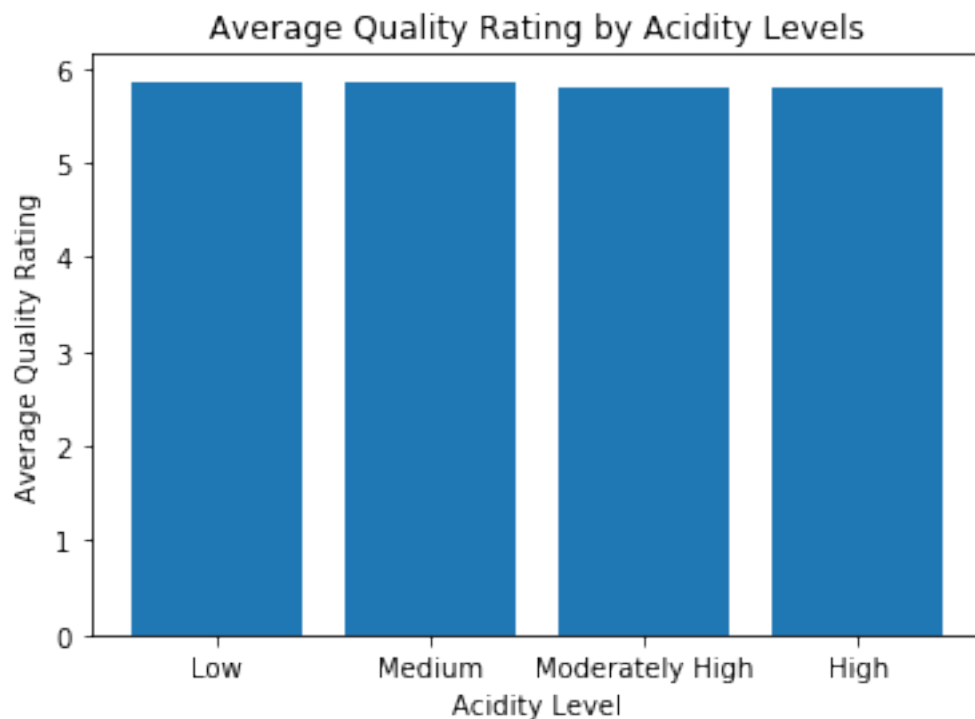
acidity_meds = df.query('acidity_levels=="Medium"')['quality'].mean()
acidity_modhigh = df.query('acidity_levels=="ModeratelyHigh"')['quality'].mean()
acidity_high = df.query('acidity_levels=="High"')['quality'].mean()

```

```

In [174]: # Create a bar chart with proper labels
locations = [1,2,3,4]
labels = ["Low", "Medium", "Moderately High", "High"]
data = [acidity_lows, acidity_meds, acidity_modhigh, acidity_high]
plt.bar(locations, data, tick_label=labels)
plt.title("Average Quality Rating by Acidity Levels")
plt.xlabel("Acidity Level")
plt.ylabel("Average Quality Rating");

```



#### 1.0.4 Bonus: Create a line plot for the data in #3

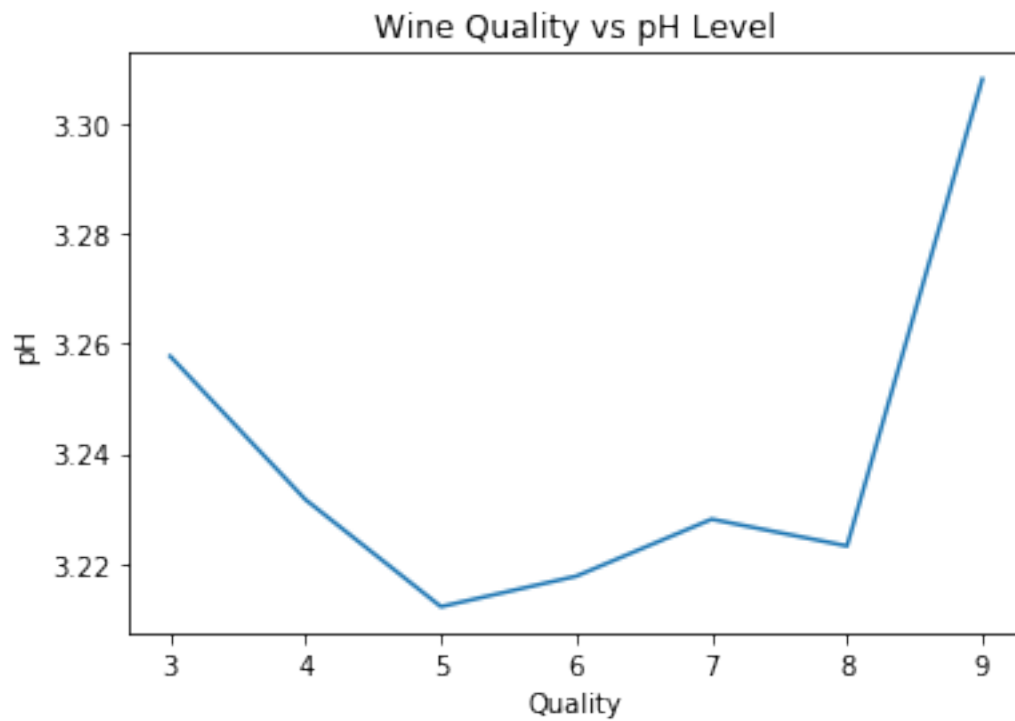
You can use pyplot's `plot` function for this.

```

In [191]: #x = df.groupby('quality').nunique()
y = df.groupby('quality')['pH'].mean()
x = df.groupby('quality')['quality'].mean()
plt.plot(x,y);
plt.title("Wine Quality vs pH Level")
plt.xlabel("Quality")
plt.ylabel("pH")

```

Out[191]: Text(0,0.5, 'pH')



Compare this with the bar chart. How might showing this visual instead of the bar chart affect someone's conclusion about this data?